(19) Евразийское патентное ведомство

(21) 202292526 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.12.29

(22) Дата подачи заявки
2021.03.04

(51) Int. Cl. ***C12N 15/09*** (2006.01)
C12N 15/87 (2006.01)
C12N 9/1276 (2006.01)

(54) УЛУЧШЕННЫЕ СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА

(31) **62/985,264; 63/035,674**
(32) **2020.03.04; 2020.06.05**
(33) **US**
(86) **PCT/US2021/020943**
(87) **WO 2021/178717 2021.09.10**

(71) Заявитель:
ФЛЭГШИП ПАЙОНИРИНГ ИННОВЕЙШНЗ VI, ЭлЭлСи (US)

(72) Изобретатель:
Кливер Стефен, Штайнберг Барретт Итан, Рубенс Якоб Розенблум, Ситорик Роберт Джеймс, Саломон Уильям Эдвард, Ван Цзы Цзюнь (US)

(74) Представитель:
Медведев В.Н. (RU)

(57) Раскрыты способы и композиции для модулирования генома-мишени.

202292526 A1

202292526 A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-575561EA/061

УЛУЧШЕННЫЕ СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА

РОДСТВЕННЫЕ ЗАЯВКИ

1. Настоящая заявка заявляет приоритет по заявке на патент США под регистрационным номером 62/985264, поданной 4 марта 2020 года, и заявке на патент США под регистрационным номером 63/035674, поданной 5 июня 2020 года, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

ПРЕДПОСЫЛКИ ИЗОБРЕТЕНИЯ

Интеграция представляющей интерес нуклеиновой кислоты в геном происходит с низкой частотой и небольшой сайт-специфичностью при отсутствии специализированного белка, предназначенного для способствования событию вставки. Некоторые существующие подходы, такие как CRISPR/Cas9, являются более подходящими для редактирования небольших участков и являются менее эффективными при интеграции более длинных последовательностей. Для других существующих подходов, таких как Cre/loxP, требуется первая стадия вставки сайта loxP в геном, а затем вторая стадия вставки последовательности, представляющей интерес, в сайт loxP. В данной области техники имеется потребность в улучшенных белках для осуществления вставки последовательностей, представляющих интерес, в геном.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Настоящее изобретение относится к новым композициям, системам и способам для изменения генома в одном или нескольких местоположениях в клетке-, ткани- или субъекте-хозяине *in vivo* или *in vitro*. В частности, в настоящем изобретении представлены композиции, системы и способы для введения экзогенных генетических элементов в геном хозяина. В настоящем изобретении также предусмотрены системы для изменения последовательности геномной ДНК, представляющей интерес, например, путем вставки, удаления или замены одного или нескольких нуклеотидов в/из последовательности, представляющей интерес.

Признаки композиций или способов могут включать один или несколько из следующих пронумерованных вариантов осуществления.

1. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

2. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

3. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует терапевтический полипептид или кодирует полипептид млекопитающего (например, человека) или его фрагмент или вариант.

4. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует терапевтическую некодирующую РНК (например, miRNA).

5. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит регуляторную последовательность (например, промотор, энхансер, сайт связывания эндогенного регуляторного компонента, например, сайт связывания miRNA), например, которая изменяет экспрессию эндогенного гена или некодирующей РНК.

6. Система согласно любому из предыдущих вариантов осуществления, где регуляторная последовательность приводит к повышению уровня экспрессии эндогенного гена или некодирующей РНК.

7. Система согласно любому из предыдущих вариантов осуществления, где регуляторная последовательность приводит к понижению уровня экспрессии эндогенного гена или некодирующей РНК.

8. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), (ii) необязательно

последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью;

где:

(i) полипептид содержит гетерологичный нацеливающий домен (например, в DBD или эндонуклеазном домене), который специфично связывается с последовательностью, содержащейся в сайте-мишени; и/или

(ii) РНК-матрица содержит гетерологичную последовательность гомологии, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в сайте-мишени.

9. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

10. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

11. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) домен связывания ДНК-мишени, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

12. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1 или таблице X) и (ii) домен связывания ДНК-мишени, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

13. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы из белка, отличного от ретротранспозазы, например, из ретровируса, например, домен обратной транскриптазы, указанный в таблице Z1 или Z2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним; и (ii) эндонуклеазный домен и/или домен связывания ДНК-мишени; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

14. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы из белка, отличного от ретротранспозазы, например, из ретровируса, например, домен обратной транскриптазы, указанный в таблице Z1 или Z2, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от него; и (ii) эндонуклеазный домен и/или домен связывания ДНК-мишени; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

15. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице Z1 или Z2) и (ii) эндонуклеазный домен и/или домен связывания ДНК-мишени; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с 5'-UTR или 3'-UTR последовательности элемента из таблицы 10 или таблицы X.

16. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице Z1 или Z2) и (ii) эндонуклеазный домен и/или домен связывания ДНК-мишени; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы), или (ii) нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

17. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице Z1, таблице Z2 или таблице X) и (ii) эндонуклеазный домен и/или домен связывания ДНК-мишени;

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта; и

(c) интеин.

18. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит интеин.

19. Система согласно любому из предыдущих вариантов осуществления, где интеин представляет собой расщепленный интеин.

20. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность (например, спейсер CRISPR), которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

21. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где домен RT имеет аминокислотную последовательность из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

22. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (etRNA) (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где система способна производить вставку в сайт-мишень размером по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов.

23. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где последовательность гетерологичного объекта имеет длину по меньшей мере 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, или 1000 нуклеотидов.

24. Система согласно любому из предыдущих вариантов осуществления, где справедливо одно или несколько из следующего: домен RT является гетерологичным по отношению к DBD; DBD является гетерологичным по отношению к эндонуклеазному домену; или домен RT является гетерологичным по отношению к эндонуклеазному домену.

25. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где система способна производить делецию в сайте-мишени размером по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов.

26. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где (a)(ii) и/или (a)(iii) содержит молекулу TALE; молекулу с "цинковыми пальцами" или молекулу CRISPR/Cas, выбранную из таблицы 1, или их функциональный вариант (например, мутантную форму).

27. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность (например, спейсер CRISPR), которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

где эндонуклеазный домен, например, никазный домен, разрезает обе нити ДНК сайта-мишени, и где разрезы отделены друг от друга по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 или 30 нуклеотидами.

28. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) последовательность, которая специфично связывает домен RT, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

29. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит последовательность, которая связывает (a)(ii) и/или (a)(iii).

30. Система для модификации ДНК, содержащая:

(a) первый полипептид или нуклеиновую кислоту, кодирующую первый полипептид, где первый полипептид содержит (i) домен обратной транскриптазы (RT) и (ii) необязательно ДНК-связывающий домен,

(b) второй полипептид или нуклеиновую кислоту, кодирующую второй полипептид, где второй полипептид содержит (i) ДНК-связывающий домен (DBD); (ii) эндонуклеазный домен, например, никазный домен; и

(c) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает второй полипептид (например, которая связывает (b)(i) и/или (b)(ii)), (ii) необязательно последовательность, которая связывает первый полипептид (например, которая специфично связывает домен RT), (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

31. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT) и (ii) ДНК-связывающий домен (DBD); а также (iii) эндонуклеазный домен, например, никазный домен;

(b) первую РНК-матрицу (или ДНК, кодирующую РНК), содержащую (например, в направлении от 5' к 3') (i) последовательность, которая связывает полипептид (например, которая связывает (a)(ii) и/или (a)(iii)), и (ii) последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени) (например, где первая РНК содержит gRNA);

(c) вторую РНК-матрицу (или ДНК, кодирующую РНК), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает полипептид (например, которая специфично связывает домен RT), (ii) последовательность гетерологичного объекта и (iii) 3'-концевой домен гомологии.

32. Система согласно любому из предыдущих вариантов осуществления, где вторая РНК-матрица содержит (i).

33. Система согласно любому из предыдущих вариантов осуществления, где первая РНК-матрица содержит первый конъюгирующий домен, а вторая РНК-матрица содержит второй конъюгирующий домен.

34. Система согласно любому из предыдущих вариантов осуществления, где первый и второй конъюгирующие домены способны гибридизироваться друг с другом, например в жестких условиях.

35. Система согласно любому из предыдущих вариантов осуществления, где ассоциация первого конъюгирующего домена и второго конъюгирующего домена обеспечивает совместную локализацию первой РНК-матрицы и второй РНК-матрицы.

36. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (i).

37. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (ii).

38. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (i) и (ii).

39. Система для модификации ДНК, содержащая:

(a) первый полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит домен обратной транскриптазы (RT), где домен RT имеет последовательность из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и необязательно ДНК-связывающий домен (DBD) (например, первый DBD); и

(b) второй полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) DBD (например, второй DBD) и (ii) эндонуклеазный домен, например, никазный домен.

40. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая первый полипептид, и нуклеиновая кислота, кодирующая второй полипептид, являются двумя отдельными нуклеиновыми кислотами.

41. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая первый полипептид, и нуклеиновая кислота, кодирующая второй полипептид, являются частью одной и той же молекулы нуклеиновой кислоты, например, присутствуют в одном и том же векторе.

42. Система согласно любому из предыдущих вариантов осуществления, обладающая одной или несколькими (например, 1, 2, 3, 4, 5, 6 или всеми) следующими характеристиками:

i. последовательность гетерологичного объекта кодирует белок, например, фермент (например, лизосомальный фермент) или фактор свертывания крови (например, фактор I, II, V, VII, X, XI, XII или XIII);

ii. последовательность гетерологичного объекта содержит тканеспецифический промотор или энхансер;

iii. последовательность гетерологичного объекта кодирует полипептид из более чем 50, 100, 150, 200, 250, 300, 400, 500 или 1000 аминокислот и необязательно вплоть до 7500 аминокислот;

iv. последовательность гетерологичного объекта кодирует фрагмент гена млекопитающего, но не кодирует полный ген млекопитающего, например, кодирует один или несколько экзонов, но не кодирует полноразмерный белок;

v. последовательность гетерологичного объекта кодирует один или несколько интронов;

vi. последовательность гетерологичного объекта является отличной от GFP, например, является отличной от флуоресцентного белка или является отличной от репортерного белка; или

vii. последовательность гетерологичного объекта содержит только некодирующие последовательности, например, регуляторные элементы.

43. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

44. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

45. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

46. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

47. Система согласно любому из предыдущих вариантов осуществления, где полипептид обладает активностью при 37°C, составляющей не менее 70%, 75%, 80%, 85%, 90% или 95% от его активности при 25°C при в остальном аналогичных условиях.

48. Система согласно любому из предыдущих вариантов осуществления, где полипептид получен из гомойотермного организма, например, птицы или млекопитающего.

49. Система согласно любому из предыдущих вариантов осуществления, где полипептид получен из одной или нескольких из ретротранспозазы CRE, ретротранспозазы NeSL, ретротранспозазы R4, ретротранспозазы R2, ретротранспозазы Hero, ретротранспозазы L1, ретротранспозазы RTE, ретротранспозазы I, ретротранспозазы Jockey, ретротранспозазы CR1, ретротранспозазы Rex1, ретротранспозазы RandI/Dualen, ретротранспозазы Penelope или Penelope-подобной ретротранспозазы, ретротранспозазы Tx1, ретротранспозазы RTEX, ретротранспозазы Crack, ретротранспозазы Nimb, ретротранспозазы Proto1, ретротранспозазы Proto2, ретротранспозазы RTETP, ретротранспозазы L2, ретротранспозазы Tad1, ретротранспозазы Loa, ретротранспозазы Ingi, ретротранспозазы Outcast, ретротранспозазы R1, ретротранспозазы Daphne, ретротранспозазы L2A, ретротранспозазы L2B, ретротранспозазы Ambal, ретротранспозазы Vingi и/или ретротранспозазы Kiri.

50. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит эндонуклеазный домен из подвижного элемента, например, эндонуклеазы, подобной рестрикционной эндонуклеазе (RLE), эндонуклеазы, подобной апуриновой/апиримидиновой эндонуклеазе (APE), эндонуклеазы GIY-YIG.

51. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен является интактным.

52. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен является инактивированным.

53. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеаза вносит однонитевые разрывы в ДНК.

54. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеаза производит двухнитевой разрыв.

55. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит последовательность из таблицы 3A, или 3B, или 10 (например, одну или обе из 5'-нетранслируемой области из столбца 6 таблицы 3A или 3B и 3'-нетранслируемой области из столбца 7 таблицы 3A или 3B) или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

56. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит последовательность из таблицы 3A, или 3B, или 11 (например, одну или обе из 5'-нетранслируемой области из столбца 6 таблицы 3A или 3B и 3'-нетранслируемой области из столбца 7 таблицы 3A или 3B) или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

57. Система согласно любому из предыдущих вариантов осуществления, где система обладает одной или несколькими (например, 1, 2, 3 или всеми) следующими характеристиками:

i. нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или нуклеиновая кислота, кодирующая РНК-матрицу, являются отдельными нуклеиновыми кислотами;

ii. РНК-матрица не кодирует активную обратную транскриптазу, например, содержит инактивированную мутантную обратную транскриптазу, например, как описано в примерах 1-2, или не содержит последовательность обратной транскриптазы;

iii. РНК-матрица не кодирует активную эндонуклеазу, например, содержит инактивированную эндонуклеазу или не содержит эндонуклеазу; или

iv. РНК-матрица содержит одну или несколько химических модификаций.

58. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) содержит (i) последовательность 5'-UTR, которая связывает полипептид, (ii) последовательность 3'-UTR, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) промотор, функционально связанный с последовательностью гетерологичного объекта,

где промотор расположен между 5'-нетранслируемой последовательностью, которая связывает полипептид, и гетерологичной последовательностью, или

где промотор расположен между 3'-нетранслируемой последовательностью, которая связывает полипептид, и гетерологичной последовательностью.

59. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) содержит (i) последовательность 5'-UTR, которая связывает полипептид, (ii) последовательность 3'-UTR, которая связывает полипептид, и (iii) последовательность гетерологичного объекта, и

где последовательность гетерологичного объекта содержит открытую рамку считывания (или последовательность, обратно комплементарную ей) в ориентации 5'-3' в РНК-матрице; или

где последовательность гетерологичного объекта содержит открытую рамку считывания (или последовательность, обратно комплементарную ей) в ориентации 3'-5' в РНК-матрице.

60. Система согласно любому из предыдущих вариантов осуществления, где последовательность 5'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

61. Система согласно любому из предыдущих вариантов осуществления, где последовательность 5'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

62. Система согласно любому из предыдущих вариантов осуществления, где последовательность 5'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется существенным структурным сходством (например, существенным

сходством вторичной структуры) с последовательностью 5'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

63. Система согласно любому из предыдущих вариантов осуществления, где последовательность 5'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR из столбца 6 таблицы 3A, или 3B, или 10.

64. Система согласно любому из предыдущих вариантов осуществления, где последовательность 3'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

65. Система согласно любому из предыдущих вариантов осуществления, где последовательность 3'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется существенным структурным сходством (например, существенным сходством вторичной структуры) с последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

66. Система согласно любому из предыдущих вариантов осуществления, где последовательность 3'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 3'-UTR из таблицы 10 или столбца 7 таблицы 3A или 3B.

67. Система согласно любому из предыдущих вариантов осуществления, где последовательность 3'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

68. Система согласно любому из предыдущих вариантов осуществления, где 3'-UTR РНК-матрицы (или ДНК, кодирующей РНК-матрицу) фланкирована доменом гомологии, например, как описано в данном документе, например, доменом гомологии, имеющим по меньшей мере 5, 10, 20, 50 или 100 оснований, характеризующихся по меньшей мере 80% (например, по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 100%) идентичностью с нитью ДНК-мишени, где необязательно домен гомологии содержит последовательность в соответствии с 3'-плечом гомологии из таблицы 11 или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

69. Система согласно любому из предыдущих вариантов осуществления, где 5'-UTR фланкирована доменом гомологии, например, как описано в данном документе, например, доменом гомологии, имеющим по меньшей мере 10, 20, 50 или 100 оснований,

характеризующихся по меньшей мере 80% (например, по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 100%) идентичностью с нитью ДНК-мишени, где необязательно домен гомологии содержит последовательность в соответствии с 5'-плечом гомологии из таблицы 11 или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

70. Система согласно любому из предыдущих вариантов осуществления, где по меньшей мере один из домена обратной транскриптазы, эндонуклеазного домена или домена связывания с ДНК-мишенью является гетерологичным, например, по отношению к другим доменам.

71. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен является гетерологичным по отношению к домену обратной транскриптазы и/или домену связывания ДНК-мишени.

72. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа апуриновой/апиримидиновой эндонуклеазы (APE), и (ii) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену не содержащего LTR ретротранспозона APE-типа.

73. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы ретротранспозона типа Penelope-подобного элемента (PLE), и (ii) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену ретротранспозона PLE-типа, например, где ретротранспозон PLE-типа содержит эндонуклеазу GIY-YIG.

74. Система согласно любому из предыдущих вариантов осуществления, где ретротранспозон PLE-типа не содержит функциональный эндонуклеазный домен.

75. Система согласно любому из предыдущих вариантов осуществления, где ретротранспозаза PLE-типа содержит Penelope-подобный элемент, в котором в естественных условиях отсутствует эндонуклеазный домен (например, элемент Athena, например, как описано в Gladyshev and Arkhipova PNAS 104, 9352-9357 (2007)).

76. Система согласно любому из предыдущих вариантов осуществления, где ретротранспозаза PLE-типа, в которой отсутствует функциональный эндонуклеазный домен, слита с Cas9, например, где ретротранспозаза PLE-типа, слитая с Cas9, имеет функцию DBD и/или эндонуклеазы (например, никазы).

77. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную

(например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа эндонуклеазы, подобной ферментам рестрикции (RLE), (ii) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену не содержащего LTR ретротранспозона RLE-типа, и (iii) домен связывания ДНК-мишени, гетерологичный по отношению к (i) и/или (ii) (например, гетерологичный ДНК-связывающий домен типа "цинкового пальца").

78. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа эндонуклеазы, подобной ферментам рестрикции (RLE), и (ii) гетерологичный эндонуклеазный домен и/или гетерологичный ДНК-связывающий домен.

79. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа апуриновой/апиримидиновой эндонуклеазы (APE), (ii) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену не содержащего LTR ретротранспозона APE-типа, и (iii) домен связывания ДНК-мишени, гетерологичный по отношению к (i) и/или ii) (например, гетерологичный ДНК-связывающий домен типа "цинкового пальца").

80. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа апуриновой/апиримидиновой эндонуклеазы (APE), и (ii) гетерологичный эндонуклеазный домен и/или гетерологичный ДНК-связывающий домен.

81. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену обратной транскриптазы ретротранспозона типа Penelope-подобного элемента (PLE), (ii) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену ретротранспозона PLE-типа, и (iii) домен связывания ДНК-мишени, гетерологичный по отношению к (i) и/или ii) (например, гетерологичный ДНК-связывающий домен типа "цинкового пальца").

82. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) домену

обратной транскриптазы ретротранспозона типа Penelope-подобного элемента (PLE), и (ii) гетерологичный эндонуклеазный домен и/или гетерологичный ДНК-связывающий домен.

83. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (iii) промотор, функционально связанный с последовательностью гетерологичного объекта.

84. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит (iii) ДНК-связывающий домен.

85. Система согласно любому из предыдущих вариантов осуществления, где ДНК-связывающий домен обладает эндонуклеазной активностью.

86. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен или эндонуклеазная активность приводят к образованию двухнитевых разрывов в ДНК.

87. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен или эндонуклеазная активность приводят к внесению однонитевых разрывов в ДНК.

88. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) последовательности из столбца 7 таблицы 3A или 3B.

89. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) последовательности элемента, указанного в таблице 10, таблице 11 или таблице X.

90. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или нуклеиновая кислота, кодирующая РНК-матрицу, связаны ковалентной связью, например, являются частью слитой нуклеиновой кислоты.

91. Система согласно любому из предыдущих вариантов осуществления, где слитая нуклеиновая кислота содержит РНК.

92. Система согласно любому из предыдущих вариантов осуществления, где слитая нуклеиновая кислота содержит ДНК.

93. Система согласно любому из предыдущих вариантов осуществления, где (b) содержит РНК-матрицу.

94. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит сигнал ядерной локализации.

95. Система согласно любому из предыдущих вариантов осуществления, где (a) содержит РНК, кодирующую полипептид.

96. Система согласно любому из предыдущих вариантов осуществления, где РНК согласно (a) и РНК согласно (b) представляют собой отдельные молекулы РНК.

97. Система согласно любому из предыдущих вариантов осуществления, где РНК согласно (a) и РНК согласно (b) присутствуют в соотношении, составляющем от 100:1 до 10:1, от 10:1 до 5:1, от 5:1 до 2:1, от 2:1 до 1:1, от 1:1 до 1:2, от 1:2 до 1:5, от 1:5 до 1:10 или от 1:10 до 1:100.

98. Система согласно любому из предыдущих вариантов осуществления, где РНК из (a) не содержит сигнал ядерной локализации.

99. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит сигнал ядерной локализации и/или сигнал ядрышковой локализации.

100. Система согласно любому из предыдущих вариантов осуществления, где (a) содержит РНК, которая кодирует: (i) полипептид и (ii) сигнал ядерной локализации и/или сигнал ядрышковой локализации.

101. Система согласно любому из предыдущих вариантов осуществления, где РНК содержит последовательность псевдоузла, например, с 5'-стороны от последовательности гетерологичного объекта.

102. Система согласно любому из предыдущих вариантов осуществления, где РНК содержит последовательность структуры "стебель-петля" или спираль с 5'-стороны от последовательности псевдоузла.

103. Система согласно любому из предыдущих вариантов осуществления, где РНК содержит одну или несколько (например, 2, 3 или больше) последовательностей структуры "стебель-петля" или спиралей с 3'-стороны от последовательности псевдоузла, например, с 3'-стороны от последовательности псевдоузла и с 5'-стороны от последовательности гетерологичного объекта.

104. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица, содержащая псевдоузел, обладает каталитической активностью, например, активностью расщепления РНК, например, активностью расщепления РНК в цис-положении.

105. Система согласно любому из предыдущих вариантов осуществления, где РНК содержит по меньшей мере одну последовательность структуры "стебель-петля" или спираль, например, с 3'-стороны от последовательности гетерологичного объекта, например, 1, 2, 3, 4, 5 или более последовательностей структур "стебель-петля", последовательностей шпилек или спиралей.

106. Система согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы имеет аминокислотную последовательность домена обратной транскриптазы элемента, указанного на фигуре 10, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

107. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен имеет аминокислотную последовательность эндонуклеазного домена элемента, указанного на фигуре 10, или последовательность, характеризующуюся

по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

108. Система согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы имеет аминокислотную последовательность домена обратной транскриптазы R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi или RTE_Ele2, представленную в данном документе, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

109. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен имеет аминокислотную последовательность эндонуклеазного домена R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi или RTE_Ele2, представленную в данном документе, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

110. Система согласно любому из предыдущих вариантов осуществления, где полипептид имеет аминокислотную последовательность R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi или RTE_Ele2, представленную в данном документе, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

111. Система согласно любому из предыдущих вариантов осуществления, где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с 5'-UTR или R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi или RTE_Ele2, представленными в данном документе.

112. Система согласно любому из предыдущих вариантов осуществления, где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с 3'-UTR или R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-

54_AAe, Tad1-65B_BG, TART-1_DWi или RTE_Ele2, представленными в данном документе.

113. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, содержит кодирующую последовательность, которая является кодон-оптимизированной для экспрессии в клетках человека.

114. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит кодирующую последовательность, которая является кодон-оптимизированной для экспрессии в клетках человека.

115. Система согласно любому из предыдущих вариантов осуществления, где система содержит одну или несколько молекул кольцевых РНК (circRNA).

116. Система согласно любому из предыдущих вариантов осуществления, где circRNA кодирует полипептид Gene Writer.

117. Система согласно любому из предыдущих вариантов осуществления, где circRNA содержит РНК-матрицу.

118. Система согласно любому из предыдущих вариантов осуществления, где circRNA доставляется в клетку-хозяина.

119. Система согласно любому из предыдущих вариантов осуществления, где circRNA способна линеаризоваться, например, в клетке-хозяине, например, в ядре клетки-хозяина.

120. Система согласно любому из предыдущих вариантов осуществления, где circRNA содержит сайт расщепления.

121. Система согласно любому из предыдущих вариантов осуществления, где circRNA дополнительно содержит второй сайт расщепления.

122. Система согласно любому из предыдущих вариантов осуществления, где сайт расщепления может быть расщеплен рибозимом, например, рибозимом, содержащимся в circRNA (например, посредством саморасщепления).

123. Система согласно любому из предыдущих вариантов осуществления, где circRNA содержит последовательность рибозима.

124. Система согласно любому из предыдущих вариантов осуществления, где последовательность рибозима способна к саморасщеплению, например, в клетке-хозяине, например, в ядре клетки-хозяина.

125. Система согласно любому из предыдущих вариантов осуществления, где рибозим представляет собой индуцируемый рибозим.

126. Система согласно любому из предыдущих вариантов осуществления, где рибозим представляет собой рибозим, отвечающий на белки, например, рибозим, отвечающий на ядерный белок, например, белок, взаимодействующий с геномом, например, эпигенетический модификатор, например, EZH2.

127. Система согласно любому из предыдущих вариантов осуществления, где рибозим представляет собой рибозим, отвечающий на нуклеиновые кислоты.

128. Система согласно любому из предыдущих вариантов осуществления, где каталитическая активность (например, аутокаталитическая активность) рибозима активируется в присутствии молекулы нуклеиновой кислоты-мишени (например, молекулы РНК, например, мРНК, miRNA, ncRNA, lncRNA, тРНК, snRNA или mtRNA).

129. Система согласно любому из предыдущих вариантов осуществления, где рибозим отвечает на белок-мишень (например, белок оболочки MS2).

130. Система согласно любому из предыдущих вариантов осуществления, где белок-мишень локализован в цитоплазме или локализован в ядре (например, представляет собой эпигенетический модификатор или фактор транскрипции).

131. Система согласно любому из предыдущих вариантов осуществления, где рибозим содержит последовательность рибозима из ретротранспозона B2 или ALU или последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней.

132. Система согласно любому из предыдущих вариантов осуществления, где рибозим содержит последовательность рибозима типа "головка молотка" из вируса кольцевой пятнистости табака или последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней.

133. Система согласно любому из предыдущих вариантов осуществления, где рибозим содержит последовательность рибозима из вируса гепатита дельта (HDV) или последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней.

134. Система согласно любому из предыдущих вариантов осуществления, где рибозим активируется компонентом, экспрессируемым в клетке-мишени или ткани-мишени.

135. Система согласно любому из предыдущих вариантов осуществления, где рибозим активируется компонентом, экспрессируемым в субклеточном компартменте-мишени (например, в ядре, ядрышке, цитоплазме или митохондриях).

136. Система согласно любому из предыдущих вариантов осуществления, где рибозим содержится в кольцевой РНК или линейной РНК.

137. Система, содержащая первую кольцевую РНК, кодирующую полипептид системы Gene Writing; и

вторую кольцевую РНК, содержащую РНК-матрицу системы Gene Writing.

138. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица, например, 5'-UTR, содержит рибозим, который расщепляет РНК-матрицу (например, в 5'-UTR).

139. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит рибозим, который является гетерологичным по отношению к (a)(i) (домену обратной транскриптазы), (a)(ii) (эндонуклеазному домену), (b)(i) (последовательности РНК-матрице, которая связывает полипептид) или их комбинации.

140. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим способен расщеплять РНК, содержащую рибозим, например, с 5'-стороны от рибозима, с 3'-стороны от рибозима или внутри рибозима.

141. Липидная наночастица (LNP), содержащая систему, полипептид (или РНК, кодирующую его), молекулу нуклеиновой кислоты или ДНК, кодирующую систему или полипептид, согласно любому из предыдущих вариантов осуществления.

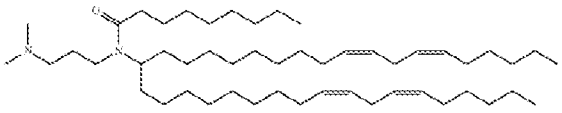
142. Система, содержащая первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы Gene Writing (например, описанной в данном документе); и

вторую липидную наночастицу, содержащую молекулу нуклеиновой кислоты системы Gene Writing (например, описанной в данном документе).

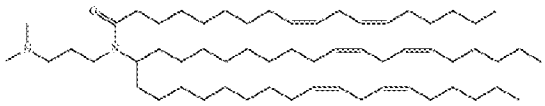
143. Система или полипептид согласно любому из предыдущих вариантов осуществления, где система, молекула нуклеиновой кислоты, полипептид и/или ДНК, кодирующая его, составлены в виде липидной наночастицы (LNP).

144. LNP согласно любому из предыдущих вариантов осуществления, содержащая катионный липид.

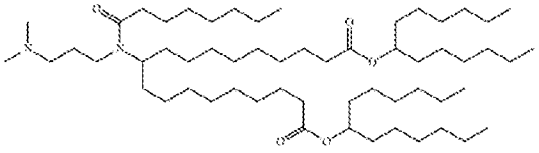
145. LNP согласно любому из предыдущих вариантов осуществления, где катионный липид имеет следующую структуру:



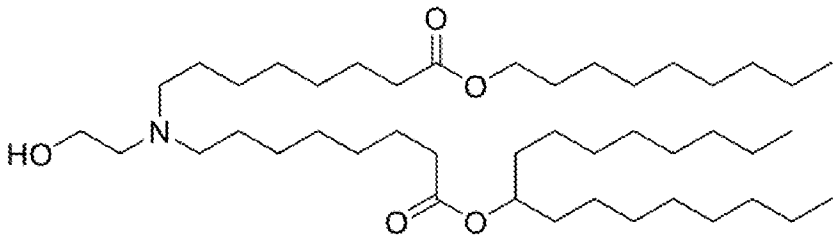
(i),



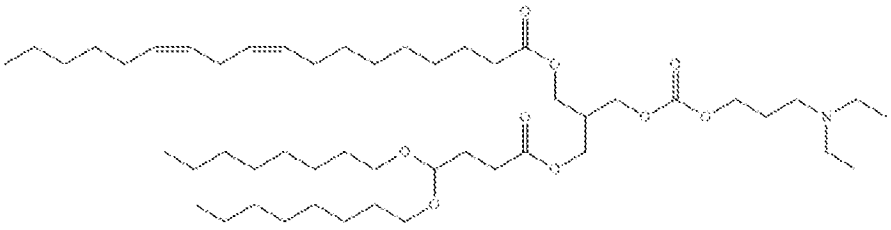
(ii),



(iii),



(vii) или



(ix).

146. LNP согласно любому из предыдущих вариантов осуществления, дополнительно содержащая один или несколько нейтральных липидов, например, DSPC, DPPC, DMPC, DOPC, POPC, DOPE, SM, стероид, например, холестерин, и/или один или несколько липидов, конъюгированных с полимером, например, пегилированный липид, например, PEG-DAG, PEG-PE, PEG-S-DAG, PEG-Cer или PEG-диалкилоксипропилкарбамат.

147. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где система, полипептид и/или ДНК, кодирующая его, составлены в виде липидной наночастицы (LNP).

148. Система или полипептид согласно любому из предыдущих вариантов осуществления, где липидная наночастица (или состав, содержащий множество липидных наночастиц) не содержит реакционноспособных примесей (например, альдегидов) или содержит реакционноспособные примеси (например, альдегиды) в количестве, меньшем, чем предварительно выбранный уровень.

149. Система или полипептид согласно любому из предыдущих вариантов осуществления, где липидная наночастица (или состав, содержащий множество липидных наночастиц) не содержит альдегидов или содержит альдегиды в количестве, меньшем, чем предварительно выбранный уровень.

150. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где липидная наночастица содержится в составе, содержащем множество липидных наночастиц.

151. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием одного или нескольких липидных реагентов, характеризующихся общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%.

152. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием одного или нескольких липидных реагентов, характеризующихся общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 3%.

153. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием одного или нескольких липидных реагентов, содержащих менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

154. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием одного или нескольких липидных реагентов, содержащих менее 0,3% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

155. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием одного или нескольких липидных реагентов, содержащих менее 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

156. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%.

157. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 3%.

158. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц содержит менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

159. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц содержит менее 0,3% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

160. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц содержит менее 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

161. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%.

162. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 3%.

163. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, содержат менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

164. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, содержат менее 0,3% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

165. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, содержат менее 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

166. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов) определяется с помощью жидкостной хроматографии (LC), например, в сочетании с тандемной масс-спектрометрией (MS/MS), например, в соответствии со способом, описанным в примере 7.

167. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяется посредством выявления одной или нескольких химических модификаций молекулы нуклеиновой кислоты (например, описанной в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах.

168. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество разновидностей альдегидов определяется посредством выявления одной или нескольких химических модификаций нуклеотида или нуклеозида (например, рибонуклеотида или рибонуклеозида, например, содержащихся в молекуле нуклеиновой кислоты или выделенных из нее, например, как описано в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах, например, как описано в примере 8.

169. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида выявляются посредством определения присутствия одного или нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 8.

170. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), характеризующуюся по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с последовательностью полипептида, кодируемого последовательностью элемента из

таблицы 10, таблицы 11 или таблицы X, или ее доменом обратной транскриптазы или эндонуклеазным доменом.

171. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных отличий от последовательности полипептида, кодируемого последовательностью элемента из таблицы 10, таблицы 11 или таблицы X, или ее домена обратной транскриптазы или эндонуклеазного домена.

172. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), характеризующуюся по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с последовательностью полипептида, приведенного в таблице 3A или 3B, или ее доменом обратной транскриптазы, эндонуклеазным доменом или ДНК-связывающим доменом.

173. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных отличий от последовательности полипептида, приведенного в таблице 3A или 3B, или ее домена обратной транскриптазы, эндонуклеазного домена или ДНК-связывающего домена.

174. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), характеризующуюся по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с аминокислотной последовательностью из столбца 7 таблицы 3A или 3B или ее доменом обратной транскриптазы, эндонуклеазным доменом или ДНК-связывающим доменом.

175. Любая вышеперечисленная система, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных отличий от аминокислотной последовательности из столбца 7 таблицы 3A или 3B или ее домена обратной транскриптазы, эндонуклеазного домена или ДНК-связывающего домена.

176. Любая вышеперечисленная система, где РНК-матрица содержит последовательность элемента из таблицы 10 или таблицы X (например, одну или обе из 5'-UTR из таблицы X или 10 и 3'-UTR из таблицы X или 10) или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

177. Любая вышеперечисленная система, где РНК-матрица содержит последовательность элемента из таблицы 10 или таблицы X (например, одну или обе из 5'-

UTR из таблицы X или 10 и 3'-UTR из таблицы X или 10) или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

178. Любая вышеперечисленная система, где РНК-матрица содержит последовательность из таблицы 3A или 3B (например, одну или обе из 5'-UTR из столбца 5 таблицы 3A или 3B и 3'-UTR из столбца 6 таблицы 3A или 3B) или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

179. Любая вышеперечисленная система, где РНК-матрица содержит последовательность из таблицы 3A или 3B (например, одну или обе из 5'-UTR из столбца 5 таблицы 3A или 3B и 3'-UTR из столбца 6 таблицы 3A или 3B) или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

180. Любая вышеперечисленная система, где РНК-матрица содержит последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

181. Любая вышеперечисленная система, где РНК-матрица содержит последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

182. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит последовательность из приблизительно 100-125 п. о. из 3'-UTR из таблицы 10 или 11 или из столбца 6 таблицы 3A или 3B, например, где последовательность содержит нуклеотиды 1-100, 101-200 или 201-325 из 3'-UTR из столбца 6 таблицы 3A или 3B, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

183. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит последовательность из приблизительно 100-125 п. о. из 3'-UTR из таблицы 10 или 11 или из столбца 6 таблицы 3A или 3B, например, где последовательность содержит нуклеотиды 1-100, 101-200 или 201-325 из 3'-UTR из столбца 6 таблицы 3A или 3B, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

184. Любая вышеперечисленная система, где (a) содержит РНК, и (b) содержит РНК.

185. Любая вышеперечисленная система, которая содержит только РНК или которая содержит РНК в большем количестве, чем ДНК, в соответствии с соотношением РНК:ДНК, составляющим по меньшей мере 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1 или 100:1.

186. Любая вышеперечисленная система, которая не содержит ДНК или которая не содержит более чем 10%, 5%, 4%, 3%, 2% или 1% ДНК по массе или по молярному количеству.

187. Любая вышеперечисленная система, которая способна модифицировать ДНК посредством вставки последовательности гетерологичного объекта без промежуточной ДНК-зависимой полимеризации РНК согласно (b).

188. Любая вышеперечисленная система, которая способна модифицировать ДНК посредством вставки последовательности гетерологичного объекта с помощью обратной транскрипции с праймированием мишенью.

189. Любая вышеперечисленная система, которая способна модифицировать ДНК посредством вставки последовательности гетерологичного объекта в присутствии ингибитора пути репарации ДНК (например, SCR7, являющегося ингибитором PARP) или в линии клеток, дефектной по пути репарации ДНК (например, линии клеток, дефектной по пути эксцизионной репарации нуклеотидов или пути репарации, направляемой гомологией).

190. Любая вышеперечисленная система, которая не вызывает образование двухнитевых разрывов на выявляемом уровне в клетке-мишени.

191. Любая вышеперечисленная система, которая способна модифицировать ДНК с помощью активности обратной транскриптазы и необязательно в отсутствие активности гомологичной рекомбинации.

192. Любая вышеперечисленная система, где РНК-матрицу подвергали обработке с целью ослабления вторичной структуры, например, нагревали, например, до температуры, при которой ослабляется вторичная структура, например, до по меньшей мере 70, 75, 80, 85, 90 или 95□□°C.

193. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрицу затем охлаждали, например, до температуры, которая обеспечивает возможность существования вторичной структуры, например, до температуры, которая меньше или равняется 37, 30, 25 или 20°C.

194. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) содержит первый домен гомологии, содержащий по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 5'-конце РНК-матрицы, и второй домен гомологии, содержащий по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 3'-конце РНК-матрицы.

195. Система согласно любому из предыдущих вариантов осуществления, где (a) и (b) являются частью одной и той же нуклеиновой кислоты.

196. Система согласно любому из предыдущих вариантов осуществления, где (a) и (b) представляют собой отдельные нуклеиновые кислоты.

197. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени (например, где нить ДНК-мишени представляет собой последовательность ДНК человека), на 5'-конце РНК-матрицы.

198. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени (например, где нить ДНК-мишени представляет собой последовательность ДНК человека), на 3'-конце РНК-матрицы.

199. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит активный домен РНКазы H.

200. Система согласно любому из предыдущих вариантов осуществления, где полипептид не содержит активный домен РНКазы H.

201. Система согласно любому из предыдущих вариантов осуществления, где эндогенный домен РНКазы H полипептида является инактивированным.

202. Система согласно любому из предыдущих вариантов осуществления, где полипептид транспозаза содержит мутацию, приводящую к инактивации и/или делеции сигнала ядрышковой локализации.

203. Система согласно любому из предыдущих вариантов осуществления, где полипептид не содержит функциональный сигнал ядрышковой локализации, например, не содержит сигнал ядрышковой локализации.

204. Система согласно любому из предыдущих вариантов осуществления, где активность сигнала ядрышковой локализации снижена на по меньшей мере 50%, 60%, 70%, 80%, 90%, 95% или 99%.

205. Система согласно любому из предыдущих вариантов осуществления, где полипептид содержит сигнал ядерной локализации (NLS), например, эндогенный NLS или экзогенный NLS.

206. Система согласно любому из предыдущих вариантов осуществления, где:

полипептид содержит (i) первый домен связывания ДНК-мишени, например, содержащий первый домен "Zn-палец", (ii) домен обратной транскриптазы, (iii) эндонуклеазный домен и (iv) второй домен связывания ДНК-мишени, например, содержащий второй домен "Zn-палец", гетерологичный по отношению к первому домену связывания ДНК-мишени; и

где (a) связывается с меньшим количеством последовательностей ДНК-мишеней в клетке-мишени, чем аналогичный полипептид, который содержит только первый домен связывания ДНК-мишени, например, где наличие второго домена связывания ДНК-мишени в полипептиде с первым ДНК-связывающим доменом обеспечивает более точную специфичность полипептида в отношении последовательности-мишени по сравнению со специфичностью полипептида в отношении последовательности-мишени у полипептида, содержащего только первый домен связывания ДНК-мишени.

207. Система согласно любому из предыдущих вариантов осуществления, где (iii) содержит (iv).

208. Система согласно любому из предыдущих вариантов осуществления, где второй домен связывания ДНК-мишени связывается с последовательностью геномной ДНК, которая находится на расстоянии менее 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 9, 8,

7, 6, 5, 4, 3, 2 или 1 нуклеотида от геномной последовательности, с которой связывается первый домен связывания ДНК-мишени.

209. Система согласно любому из предыдущих вариантов осуществления, где второй домен связывания ДНК-мишени связывается с последовательностью геномной ДНК, которая находится на расстоянии 1-100, 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 1-5, 5-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30, 5-20, 5-10, 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-100, 50-90, 50-80, 50-70, 50-60, 60-100, 60-90, 60-80, 60-70, 70-100, 70-90, 70-80, 80-100, 80-90, или 90-100 нуклеотидов от геномной последовательности, с которой связывается первый домен связывания ДНК-мишени.

210. Система согласно любому из предыдущих вариантов осуществления, где первый или второй домен связывания ДНК-мишени содержит белок CRISPR/Cas, эффекторный домен TAL, домен "Zn-палец" или мегануклеазный домен.

211. Система согласно любому из предыдущих вариантов осуществления, где система способна разрезать первую нить ДНК-мишени по меньшей мере дважды (например, дважды), и

где разрезы необязательно находятся на расстоянии по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 или 200 нуклеотидов друг от друга (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов друг от друга).

212. Система согласно любому из предыдущих вариантов осуществления, где система способна разрезать первую нить и вторую нить ДНК-мишени, и

где расстояние между разрезами является таким же, как расстояние между разрезами, произведенными доменом обратной транскриптазы, например, доменом обратной транскриптазы в случае, когда он расположен в его эндогенном полипептиде.

213. Система согласно любому из предыдущих вариантов осуществления, где разрезы расположены на расстоянии 1-500, 1-400, 1-300, 1-200, 1-100, 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 1-5, 5-500, 5-400, 5-300, 5-200, 5-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30, 5-20, 5-10, 10-500, 10-400, 10-300, 10-200, 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-500, 20-400, 20-300, 20-200, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-500, 30-400, 30-300, 30-200, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-500, 40-400, 40-300, 40-200, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-500, 50-400, 50-300, 50-200, 50-100, 50-90, 50-80, 50-70, 50-60, 60-500, 60-400, 60-300, 60-200, 60-100, 60-90, 60-80, 60-70, 70-500, 70-400, 70-300, 70-200, 70-100, 70-90, 70-80, 80-500, 80-400, 80-300, 80-200, 80-100, 80-90, 90-500, 90-400, 90-300, 90-200, 90-100, 100-500, 100-400, 100-300, 100-200, 200-500, 200-400, 200-300, 300-500, 300-400, или 400-500 нуклеотидов друг от друга.

214. Система согласно любому из предыдущих вариантов осуществления, где расстояние между разрезами является таким же, как расстояние между разрезами, произведенными доменом обратной транскриптазы, например, доменом обратной транскриптазы в случае, когда он расположен в его эндогенном полипептиде.

215. Система согласно любому из предыдущих вариантов осуществления, где оба разреза произведены одним и тем же эндонуклеазным доменом (например, белком CRISPR/Cas, например, управляемым множеством gRNA, например, расположенными в РНК-матрице).

216. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит второй эндонуклеазный домен.

217. Система согласно любому из предыдущих вариантов осуществления, где:

i) первый эндонуклеазный домен (например, никаза) разрезает подлежащую редактированию нить ДНК-мишени, и второй эндонуклеазный домен (например, никаза) разрезает нередактируемую нить ДНК-мишени, или

ii) первый эндонуклеазный домен (например, никаза) производит один из двух разрезов в подлежащей редактированию нити ДНК-мишени, и второй эндонуклеазный домен (например, никаза) производит другой разрез в подлежащей редактированию нити ДНК-мишени.

218. Система согласно любому из предыдущих вариантов осуществления, где (a), (b) или (a) и (b) дополнительно содержат 5'-UTR и/или 3'-UTR, функционально связанные с последовательностью, кодирующей полипептид, последовательностью гетерологичного объекта (например, кодирующей последовательностью, содержащейся в последовательности гетерологичного объекта) или ими обеими.

219. Система согласно любому из предыдущих вариантов осуществления, где 5'-UTR и/или 3'-UTR обеспечивают увеличение экспрессии функционально связанной(связанных) последовательности(последовательностей) на по меньшей мере 10%, 20%, 30%, 40%, 50%, 70%, 70%, 80%, 90% или 100% по сравнению с в остальном аналогичной нуклеиновой кислотой, содержащей эндогенную(эндогенные) UTR, ассоциированную(ассоциированные) с последовательностью гетерологичного объекта, или минимальную 5'-UTR и минимальную 3'-UTR.

220. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) содержит (i) последовательность, которая связывает полипептид, (ii) последовательность гетерологичного объекта и (iii) рибозим, который является гетерологичным по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации.

221. Система согласно любому из предыдущих вариантов осуществления, где (a), (b) или (a) и (b) содержат интрон, который обеспечивает увеличение экспрессии полипептида, последовательности гетерологичного объекта (например, кодирующей последовательности, расположенной в последовательности гетерологичного объекта) или их обоих.

222. Клетка-хозяин (например, клетка млекопитающего, например, клетка человека), содержащая любую вышеперечисленную систему.

223. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение любой вышеперечисленной системы в клетку, ткань или субъекту,

где система осуществляет обратную транскрипцию последовательности РНК-матрицы в нить ДНК-мишени, за счет чего обеспечивается модификация нити ДНК-мишени.

224. Способ согласно любому из предыдущих вариантов осуществления, где клетка, ткань или субъект являются клеткой, тканью млекопитающего (например, человека) или субъектом-млекопитающим (например, человеком).

225. Способ согласно любому из предыдущих вариантов осуществления, где ткань представляет собой ткань печени, легкого, кожи, системы крови, иммунной системы или мышц.

226. Способ согласно любому из предыдущих вариантов осуществления, где клетка представляет собой фибробласт.

227. Способ согласно любому из предыдущих вариантов осуществления, где клетка представляет собой первичную клетку.

228. Способ согласно любому из предыдущих вариантов осуществления, где клетка не является иммортализованной.

229. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

230. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

231. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

232. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

233. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR или последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

234. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице

3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где последовательность РНК-матрицы, которая связывает полипептид, характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы), или (ii) нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

235. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой (например, ДНК или мРНК), кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы из белка, отличного от ретротранспозазы, например, из ретровируса, например, домен обратной транскриптазы, указанный в таблице Z1 или Z2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним; и (ii) эндонуклеазный домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

236. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой (например, ДНК или мРНК), кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы из белка, отличного от ретротранспозазы, например, из ретровируса, например, домен обратной транскриптазы, указанный в таблице Z1 или Z2, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от него; и (ii) эндонуклеазный домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

237. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице

3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен;

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта; и

(c) интеином.

238. Способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит интеин.

239. Система согласно любому из предыдущих вариантов осуществления, где интеин представляет собой расщепленный интеин.

240. Способ согласно любому из предыдущих вариантов осуществления, где полипептид не содержит домен связывания ДНК-мишени.

241. Способ согласно любому из предыдущих вариантов осуществления, где полипептид получен из обратной транскриптазы, кодируемой ретротранспозоном APE-типа.

242. Способ согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

243. Способ согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

244. Способ согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

245. Способ согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит домен связывания ДНК-мишени, например, имеющий аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

246. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) РНК, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или

таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей, содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты.

247. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) РНК, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей, содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты.

248. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) РНК, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей, содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты.

249. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с:

(a) РНК, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей, содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты.

250. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований экзогенной последовательности ДНК в геном клетки млекопитающего.

251. Способ согласно любому из предыдущих вариантов осуществления, который приводит к удалению по меньшей мере 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК из генома клетки млекопитающего.

252. Способ согласно любому из предыдущих вариантов осуществления, который приводит к изменению по меньшей мере 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК в геноме клетки млекопитающего.

253. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению последовательности, кодирующей белок, в геном клетки млекопитающего.

254. Способ согласно любому из предыдущих вариантов осуществления, который приводит к удалению последовательности, кодирующей белок, из генома клетки млекопитающего.

255. Способ согласно любому из предыдущих вариантов осуществления, который приводит к изменению последовательности, кодирующей белок, в геноме клетки млекопитающего.

256. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению в геном клетки млекопитающего некодирующей последовательности, например, кодирующей некодирующую РНК, например, miRNA.

257. Способ согласно любому из предыдущих вариантов осуществления, который приводит к удалению из генома клетки млекопитающего некодирующей последовательности, например, кодирующей некодирующую РНК, например, miRNA.

258. Способ согласно любому из предыдущих вариантов осуществления, который приводит к изменению в геноме клетки млекопитающего некодирующей последовательности, например, кодирующей некодирующую РНК, например, miRNA.

259. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению в геном клетки млекопитающего регуляторной последовательности, например, промотора, энхансера, сайта связывания miRNA.

260. Способ согласно любому из предыдущих вариантов осуществления, который приводит к удалению из генома клетки млекопитающего регуляторной последовательности, например, промотора, энхансера, сайта связывания miRNA.

261. Способ согласно любому из предыдущих вариантов осуществления, который приводит к изменению в геноме клетки млекопитающего регуляторной последовательности, например, промотора, энхансера, сайта связывания miRNA.

262. Способ согласно любому из предыдущих вариантов осуществления, где добавление, удаление или изменение регуляторной последовательности в геноме клетки млекопитающего приводит к повышению экспрессии кодирующей или некодирующей последовательности в геноме клетки млекопитающего.

263. Способ согласно любому из предыдущих вариантов осуществления, где добавление, удаление или изменение регуляторной последовательности в геноме клетки млекопитающего приводит к понижению экспрессии кодирующей или некодирующей последовательности в геноме клетки млекопитающего.

264. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего; и

где первая РНК кодирует полипептид, кодируемый последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, где полипептид управляет вставкой РНК-матрицы в геном.

265. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего; и

где первая РНК кодирует полипептид, кодируемый последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, где полипептид управляет вставкой РНК-матрицы в геном.

266. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего; и

где первая РНК кодирует полипептид элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним, где полипептид управляет вставкой РНК-матрицы в геном.

267. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего; и

где первая РНК кодирует полипептид элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от него, где полипептид управляет вставкой РНК-матрицы в геном.

268. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность, и (i) фланкирующую последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X, и/или (ii) фланкирующую последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего.

269. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с композицией на основе РНК, где композиция на основе РНК содержит:

(a) первую РНК, которая управляет вставкой РНК-матрицы в геном, и

(b) РНК-матрицу, содержащую гетерологичную последовательность, и (i) фланкирующую последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы), и/или (ii) фланкирующую последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы),

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где композиции на основе (a) и (b) не содержат более чем 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты,

где способ приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований последовательности ДНК (например, экзогенной ДНК) в геном клетки млекопитающего.

270. Способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит последовательность, которая связывает полипептид; где необязательно последовательность, которая связывает полипептид, характеризуется:

(a) по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR или последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10 или таблицы X; или

(b) по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы), или (ii) нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы X (например, содержащими область связывания ретротранспозазы).

271. Способ согласно любому из предыдущих вариантов осуществления, где по меньшей мере 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК добавляются в геном клетки млекопитающего без доставки ДНК в клетку.

272. Способ согласно любому из предыдущих вариантов осуществления, где по меньшей мере 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК добавляются в геном клетки млекопитающего,

где способ не включает приведение клетки млекопитающего в контакт с ДНК, или где способ включает приведение клетки млекопитающего в контакт с композицией, содержащей менее 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02% или 0,01% ДНК по массе или по молярному количеству нуклеиновой кислоты.

273. Способ согласно любому из предыдущих вариантов осуществления, где по меньшей мере 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК добавляются в геном клетки млекопитающего, где в клетку млекопитающего доставляется только РНК.

274. Способ согласно любому из предыдущих вариантов осуществления, где по меньшей мере 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК добавляются в геном клетки млекопитающего, где в клетку млекопитающего доставляются РНК и белок.

275. Способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица служит в качестве матрицы для вставки экзогенной ДНК.

276. Способ согласно любому из предыдущих вариантов осуществления, который не включает ДНК-зависимую полимеризацию РНК для экзогенной ДНК.

277. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению по меньшей мере 1, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 или 5000 пар оснований ДНК в геном клетки млекопитающего.

278. Способы согласно любому из предыдущих вариантов осуществления, где РНК согласно (a) и РНК согласно (b) связаны ковалентной связью, например, являются частью одного и того же транскрипта.

279. Способ согласно любому из предыдущих вариантов осуществления, где РНК согласно (a) и РНК согласно (b) представляют собой отдельные РНК.

280. Способ согласно любому из предыдущих вариантов осуществления, который не включает приведение клетки млекопитающего в контакт с ДНК-матрицей.

281. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

282. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК, например, как

описано в примере 14 документа PCT/US2019/048607, включенного в данный документ посредством ссылки во всей своей полноте.

283. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

284. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен, где один, два или три из (i), (ii) и/или (iii) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта;

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

285. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR или последовательностью 3'-UTR последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X, и (ii) последовательность гетерологичного объекта,

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

286. Способ модификации генома клетки человека, включающий приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (например, указанный в таблице 3B, таблице 10, таблице 11, таблице X или таблице Z1 или Z2), (ii) эндонуклеазный домен и необязательно (iii) ДНК-связывающий домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (i) последовательность, которая связывает полипептид, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) частью последовательности элемента из таблицы 3B, таблицы 10 или таблицы X, состоящей из нуклеотидов, расположенных с 5'-стороны относительно старт-кодона, или (ii) частью последовательности элемента из таблицы 3B, таблицы 10 или таблицы X, состоящей из нуклеотидов, расположенных с 3'-стороны относительно стоп-кодона,

где способ приводит к вставке последовательности гетерологичного объекта в геном клетки человека,

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

287. Способ обеспечения добавления экзогенной кодирующей области в геном клетки (например, клетки млекопитающего), включающий приведение клетки в контакт с: (i) РНК, содержащей некодирующую нить экзогенной кодирующей области, где необязательно РНК не содержит кодирующую нить экзогенной кодирующей области, и (ii) полипептидом, содержащим последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, где необязательно доставка включает невирусную доставку.

288. Способ обеспечения добавления экзогенной кодирующей области в геном клетки (например, клетки млекопитающего), включающий приведение клетки в контакт с: (i) РНК, содержащей некодирующую нить экзогенной кодирующей области, где необязательно РНК не содержит кодирующую нить экзогенной кодирующей области, и (ii) полипептидом, содержащим последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, где необязательно доставка включает невирусную доставку.

289. Способ обеспечения экспрессии представляющего интерес полипептида в клетке (например, клетке млекопитающего), включающий приведение клетки в контакт с: (i) РНК, где РНК содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, которая будет кодировать представляющий интерес полипептид, где необязательно РНК не содержит кодирующую нить, кодирующую представляющий интерес полипептид, и (ii) полипептидом ретротранспозазой, содержащим аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, где необязательно доставка включает невирусную доставку.

290. Способ обеспечения экспрессии представляющего интерес полипептида в клетке (например, клетке млекопитающего), включающий приведение клетки в контакт с: (i) РНК, где РНК содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, которая будет кодировать представляющий интерес полипептид, где необязательно РНК не содержит кодирующую нить, кодирующую представляющий интерес полипептид, и (ii) полипептидом ретротранспозазой, содержащим аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, где необязательно доставка включает невирусную доставку.

291. Способ обеспечения экспрессии представляющего интерес полипептида в клетке (например, клетке млекопитающего), включающий приведение клетки в контакт с: (i) РНК, где РНК содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, которая будет кодировать представляющий интерес

полипептид, где необязательно РНК не содержит кодирующую нить, кодирующую представляющий интерес полипептид, и (ii) полипептидом ретротранспозазой, содержащим аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, где необязательно доставка включает невирусную доставку.

292. Способ обеспечения экспрессии представляющего интерес полипептида в клетке (например, клетке млекопитающего), включающий приведение клетки в контакт с: (i) РНК, где РНК содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, которая будет кодировать представляющий интерес полипептид, где необязательно РНК не содержит кодирующую нить, кодирующую представляющий интерес полипептид, и (ii) полипептидом ретротранспозазой, содержащим аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, где необязательно доставка включает невирусную доставку.

293. Способ согласно любому из предыдущих вариантов осуществления, где последовательность, вставляемая в геном млекопитающего, представляет собой последовательность, являющуюся экзогенной по отношению к геному млекопитающего.

294. Способ согласно любому из предыдущих вариантов осуществления, где экзогенная последовательность, вставляемая в геном млекопитающего, в природе не встречается где-либо в геноме млекопитающего.

295. Способ согласно любому из предыдущих вариантов осуществления, где экзогенная последовательность, вставляемая в геном млекопитающего, в природе встречается где-либо в геноме млекопитающего.

296. Способ согласно любому из предыдущих вариантов осуществления, который работает независимо от ДНК-матрицы.

297. Способ согласно любому из предыдущих вариантов осуществления, где клетка является частью ткани.

298. Способ согласно любому из предыдущих вариантов осуществления, где клетка млекопитающего является эуплоидной, не является иммортализованной, является частью организма, представляет собой первичную клетку, является неделящейся, представляет собой гепатоцит или получена от субъекта, у которого имеется генетическое заболевание.

299. Способ согласно любому из предыдущих вариантов осуществления, где клетка млекопитающего находится в организме субъекта, у которого не имеется заболевание, например, для дополнения генома субъекта.

300. Способ согласно любому из предыдущих вариантов осуществления, где приведение в контакт включает приведение клетки в контакт с плазмидой, вирусом, вирусоподобной частицей, виросомой, липосомой, везикулой, экзосомой, фузосомой или липидной наночастицей.

301. Способ согласно любому из предыдущих вариантов осуществления, где приведение в контакт включает применение невирусной доставки.

302. Способ согласно любому из предыдущих вариантов осуществления, который включает приведение клетки в контакт с РНК-матрицей (или ДНК, кодирующей РНК-матрицу), где РНК-матрица содержит некодирующую нить экзогенной кодирующей области, где РНК-матрица необязательно не содержит кодирующую нить экзогенной кодирующей области, где доставка необязательно включает невирусную доставку, за счет чего обеспечивается добавление экзогенной кодирующей области в геном клетки.

303. Способ согласно любому из предыдущих вариантов осуществления, который включает приведение клетки в контакт с РНК-матрицей (или ДНК, кодирующей РНК-матрицу), где РНК-матрица содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, кодирующей полипептид, где РНК-матрица необязательно не содержит кодирующую нить, кодирующую полипептид, где доставка необязательно включает невирусную доставку, за счет чего обеспечивается добавление экзогенной кодирующей области в геном клетки и экспрессия полипептида в клетке.

304. Способ согласно любому из предыдущих вариантов осуществления, где приведение в контакт включает введение (a) и (b) субъекту, например, внутривенно.

305. Способ согласно любому из предыдущих вариантов осуществления, где приведение в контакт включает введение дозы (a) и (b) субъекту по меньшей мере дважды.

306. Способ согласно любому из предыдущих вариантов осуществления, где полипептид осуществляет обратную транскрипцию последовательности РНК-матрицы в нить ДНК-мишени, за счет чего обеспечивается модификация нити ДНК-мишени.

307. Способ согласно любому из предыдущих вариантов осуществления, где (a) и (b) вводят по отдельности.

308. Способ согласно любому из предыдущих вариантов осуществления, где (a) и (b) вводят вместе.

309. Способ согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота согласно (a) не интегрируется в геном клетки-хозяина.

310. Способ согласно любому из предыдущих вариантов осуществления, где ткань представляет собой печень, легкое, кожу, мышечную ткань (например, скелетную мышцу), глаз или ткань глаза или центральную нервную систему.

311. Способ согласно любому из предыдущих вариантов осуществления, где клетка представляет собой гемопоэтическую стволовую клетку (HSC), Т-клетку или естественную клетку-киллера (NK).

312. Любой вышеперечисленный способ, где последовательность, которая связывает полипептид, обладает одной или несколькими из следующих характеристик:

(a) располагается на 3'-конце РНК-матрицы;

(b) располагается на 5'-конце РНК-матрицы;

(b) представляет собой некодирующую последовательность;

(c) представляет собой структурированную РНК;

(d) образует по меньшей мере 1 структуру шпилечной петли и/или

(e) представляет собой направляющую РНК.

313. Любой вышеперечисленный способ, где РНК-матрица дополнительно содержит последовательность, содержащую по меньшей мере 20 нуклеотидов, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) нити ДНК-мишени.

314. Любой вышеперечисленный способ, где РНК-матрица дополнительно содержит последовательность, содержащую по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 нуклеотидов, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) нити ДНК-мишени.

315. Любой вышеперечисленный способ, где последовательность, содержащая по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 нуклеотидов или приблизительно 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100 или 2-100 нуклеотидов, на по меньшей мере 80% идентичная нити ДНК-мишени, располагается на 3'-конце РНК-матрицы.

316. Любой вышеперечисленный способ, где РНК-матрица дополнительно содержит последовательность, содержащую по меньшей мере 100 нуклеотидов, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) нити ДНК-мишени, например, на 3'-конце РНК-матрицы.

317. Способ согласно любому из предыдущих вариантов осуществления, где сайт в нити ДНК-мишени, по отношению к которому последовательность обладает по меньшей мере 80% идентичностью, располагается проксимально по отношению к сайту-мишени (например, в пределах приблизительно 0-10, 10-20, 20-30, 30-50 или 50-100 нуклеотидов от него) в нити ДНК-мишени, который распознается (например, связывается и/или расщепляется) полипептидом, содержащим эндонуклеазу.

318. Любой вышеперечисленный способ, где РНК-мишень содержит домен гомологии, который содержит последовательность в соответствии с 3'-плечом гомологии из таблицы 11 или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

319. Любой вышеперечисленный способ, где РНК-мишень содержит домен гомологии, который содержит последовательность в соответствии с 5'-плечом гомологии из таблицы 11 или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

320. Любой вышеперечисленный способ, где последовательность, содержащая по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 нуклеотидов или приблизительно 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100 или 2-100 нуклеотидов, на по меньшей мере 80% идентичная нити ДНК-мишени, располагается на 3'-конце РНК-матрицы;

где сайт в нити ДНК-мишени, по отношению к которому последовательность обладает по меньшей мере 80% идентичностью, необязательно располагается проксимально по отношению к сайту-мишени (например, в пределах приблизительно 0-10, 10-20 или 20-30 нуклеотидов от него) в нити ДНК-мишени, который распознается (например, связывается и/или расщепляется) полипептидом, содержащим эндонуклеазу.

321. Способ согласно любому из предыдущих вариантов осуществления, где сайт-мишень представляет собой сайт в человеческом геноме, характеризующийся наиболее высокой идентичностью по отношению к нативному сайту-мишени для полипептида, содержащего эндонуклеазу, например, где сайт-мишень в человеческом геноме содержит по меньшей мере приблизительно 16, 17, 18, 19 или 20 нуклеотидов, идентичных нативному сайту-мишени.

322. Любой вышеперечисленный способ, где РНК-матрица содержит по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 оснований, на 100% идентичных нити ДНК-мишени.

323. Любой вышеперечисленный способ, где по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 оснований, на 100% идентичных нити ДНК-мишени, располагаются на 3'-конце РНК-матрицы.

324. Любой вышеперечисленный способ, где по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 оснований, на 100% идентичных нити ДНК-мишени, располагаются на 5'-конце РНК-матрицы.

325. Любой вышеперечисленный способ, где РНК-матрица содержит по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 оснований, на 100% идентичных нити ДНК-мишени, на 5'-конце РНК-матрицы и по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 оснований, на 100% идентичных нити ДНК-мишени, на 3'-конце РНК-матрицы.

326. Любой вышеперечисленный способ, где длина последовательности гетерологичного объекта составляет от 50 до 50000 пар оснований (например, от 50 до 40000 п. о., от 500 до 30000 п. о., от 500 до 20000 п. о., от 100 до 15000 п. о., от 500 до 10000 п. о., от 50 до 10000 п. о., от 50 до 5000 п. о.).

327. Любой вышеперечисленный способ, где длина последовательности гетерологичного объекта составляет по меньшей мере 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600 или 700 п. о.

328. Любой вышеперечисленный способ, где длина последовательности гетерологичного объекта составляет по меньшей мере 715, 750, 800, 950, 1000, 2000, 3000 или 4000 п. о.

329. Любой вышеперечисленный способ, где длина последовательности гетерологичного объекта составляет менее 5000, 10000, 15000, 20000, 30000 или 40000 п. о.

330. Любой вышеперечисленный способ, где длина последовательности гетерологичного объекта составляет менее 700, 600, 500, 400, 300, 200, 150 или 100 п. о.

331. Любой вышеперечисленный способ, где последовательность гетерологичного объекта содержит одно или несколько из следующего:

(a) открытую рамку считывания, например, последовательность, кодирующую полипептид, например, фермент (например, лизосомальный фермент), мембранный белок, фактор свертывания крови, экзон, внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок), внеклеточный белок, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок, запасной белок, иммунорецепторный белок (например, химерный антигенный рецепторный белок (CAR), Т-клеточный рецептор, В-клеточный рецептор) или антитело;

(b) некодирующую и/или регуляторную последовательность, например, последовательность, которая связывает модулятор транскрипции, например, промотор, энхансер, инсулятор;

(c) акцепторный сайт сплайсинга;

(d) поли(А)-сайт;

(e) сайт эпигенетической модификации или

(f) единицу экспрессии гена.

332. Любой вышеперечисленный способ, где ДНК-мишень представляет собой геномный сайт "безопасная гавань" (GSH).

333. Любой вышеперечисленный способ, где ДНК-мишень представляет собой геномный сайт Natural Harbor™.

334. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в геном при среднем числе копий, составляющем по меньшей мере 0,01, 0,025, 0,05, 0,075, 0,1, 0,15, 0,2, 0,25, 0,3, 0,4, 0,5, 0,75, 1, 1,25, 1,5, 1,75, 2, 2,5, 3, 4, или 5 копий на геном.

335. Любой вышеперечисленный способ, который приводит к тому, что приблизительно 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90% интегрированных элементов в сайте-мишени в геноме не являются усеченными согласно измерению с помощью анализа, описанного в данном документе, например, анализа из примера 6 заявки согласно PCT № PCT/US2019/048607.

336. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта только в один сайт-мишень в геноме клетки.

337. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень в клетке, где вставленная гетерологичная последовательность содержит менее 10%, 5%, 2%, 1%, 0,5%, 0,2% или 0,1% мутаций (например, SNP или одну или несколько делеций, например, усечений или внутренних делеций) по сравнению с гетерологичной последовательностью до вставки, например, согласно измерению с помощью анализа из примера 12 заявки согласно PCT № PCT/US2019/048607.

338. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень во множестве клеток, где

менее 10%, 5%, 2% или 1% копий вставленной гетерологичной последовательности содержат мутацию (например, SNP или делецию, например, усечение или внутреннюю делецию), например, согласно измерению с помощью анализа из примера 12 заявки согласно PCT № PCT/US2019/048607.

339. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в геном клетки-мишени, и где клетка-мишень не демонстрирует повышение уровня экспрессии p53 или демонстрирует повышение уровня экспрессии p53 на менее чем 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии p53 измеряют по уровню белка p53, например, в соответствии со способом, описанным в примере 30, или по уровню p53, фосфорилированного по Ser15 и Ser20.

340. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в геном клетки-мишени, и где клетка-мишень не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 50%, 25%, 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК.

341. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень (например, при числе копий, составляющем 1 вставку или более одной вставки) в приблизительно 1-80% клеток в популяции клеток, приведенной в контакт с системой, например, в приблизительно 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70% или 70-80% клеток, например, согласно измерению с помощью ddPCR отдельных клеток, например, как описано в примере 17.

342. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень при числе копий, составляющем 1 вставку, в приблизительно 1-80% клеток в популяции клеток, приведенной в контакт с системой, например, в приблизительно 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70% или 70-80% клеток, например, согласно измерению с помощью выделения колоний и ddPCR, например, как описано в примере 18.

343. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень (целевые вставки) с более высокой частотой, чем для вставки в сайт, не являющийся мишенью (нецелевые вставки), в популяции клеток, где соотношение целевых вставок и нецелевых вставок составляет более чем 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 200:1, 500:1 или 1000:1, например, с использованием анализа из примера 11 заявки согласно PCT № PCT/US2019/048607.

344. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в присутствии ингибитора пути репарации ДНК (например, SCR7, являющегося ингибитором PARP) или в линии клеток, дефектной

по пути репарации ДНК (например, линии клеток, дефектной по пути эксцизионной репарации нуклеотидов или пути репарации, направляемой гомологией).

345. Способ согласно любому из предыдущих вариантов осуществления, где клетка характеризуется уменьшенной активностью Rad51-опосредованного пути репарации, уменьшенной экспрессией Rad51 или компонента Rad51-опосредованного пути репарации или не содержит функциональный Rad51-опосредованный путь репарации, например, не содержит функциональный ген Rad51, например, содержит мутацию (например, делецию), приводящую к инактивации одной или обеих копий гена Rad51 или другого гена в Rad51-опосредованном пути репарации.

346. Любая вышеперечисленная система, составленная в виде фармацевтической композиции.

347. Любая вышеперечисленная система, находящаяся в фармацевтически приемлемом носителе (например, везикуле, липосоме, природном или синтетическом липидном бислое, липидной наночастице, экзосоме).

348. Любой вышеперечисленный способ, который приводит к множеству (например, по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9 или 10) вставок последовательности гетерологичного объекта в геном клетки-мишени.

349. Способ согласно любому из предыдущих вариантов осуществления, где множество вставок происходит одновременно или последовательно.

350. Любой вышеперечисленный способ, который приводит к множеству (например, по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9 или 10) делеций последовательности гетерологичного объекта в геноме клетки-мишени.

351. Способ согласно любому из предыдущих вариантов осуществления, где множество делеций происходит одновременно или последовательно.

352. Любой вышеперечисленный способ, который приводит к множеству (например, по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9 или 10) замен оснований последовательности гетерологичного объекта в геноме клетки-мишени.

353. Способ согласно любому из предыдущих вариантов осуществления, где множество замен оснований происходит одновременно или последовательно.

354. Любой вышеперечисленный способ, включающий приведение клетки в контакт с множеством различных РНК-матриц, каждая из которых содержит последовательность гетерологичного объекта.

355. Способ согласно любому из предыдущих вариантов осуществления, где различные РНК-матрицы содержат по меньшей мере две (например, по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9 или 10) различные последовательности гетерологичных объектов.

356. Способ согласно любому из предыдущих вариантов осуществления, где каждая из по меньшей мере двух различных последовательностей гетерологичных объектов содержит разную полезную нагрузку.

357. Способ согласно любому из предыдущих вариантов осуществления, где каждая из по меньшей мере двух различных последовательностей гетерологичных объектов содержит одинаковую полезную нагрузку.

358. Любой вышеперечисленный вариант осуществления, в котором клетка-мишень системы Gene Writing была ранее модифицирована в одном или нескольких локусах.

359. Любой вышеперечисленный вариант осуществления, в котором ранее отредактированная клетка представляет собой Т-клетку.

360. Любой вышеперечисленный вариант осуществления, где одна или несколько предыдущих модификаций выбраны из нокаутов генов, например, эндогенного TCR (например, TRAC, TRBC), HLA класса I (B2M), PD1, CD52, CTLA-4, TIM-3, LAG-3 или DGK.

361. Любой вышеперечисленный вариант осуществления, где последовательность гетерологичного объекта содержит TCR или CAR.

362. Способ получения системы для модификации генома клетки млекопитающего, включающий:

(a) получение РНК-матрицы, содержащей (i) последовательность, которая связывает полипептид, содержащий домен обратной транскриптазы и эндонуклеазный домен, при этом последовательность характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с последовательностью 5'-UTR или последовательностью 3'-UTR последовательности элемента из таблицы X, и (ii) последовательность гетерологичного объекта.

363. Способ согласно любому из предыдущих вариантов осуществления, дополнительно включающий:

(b) обработку РНК-матрицы с целью ослабления вторичной структуры, например, нагревание РНК-матрицы, например, до по меньшей мере 70, 75, 80, 85, 90 или 95°C, и/или c) последующее охлаждение РНК-матрицы, например, до температуры, которая обеспечивает возможность существования вторичной структуры, например, до температуры, которая меньше или равняется 37, 30, 25 или 20°C.

364. Способ получения системы для модификации ДНК (например, описанной в данном документе), при этом способ включает:

(a) получение нуклеиновой кислоты-матрицы (например, РНК- или ДНК-матрицы), содержащей гетерологичную последовательность гомологии, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в молекуле ДНК-мишени, и/или

(b) получение полипептида системы (например, содержащего ДНК-связывающий домен (DBD) и/или эндонуклеазный домен), содержащего гетерологичный нацеливающий домен, который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени.

365. Способ согласно любому из предыдущих вариантов осуществления, где:

(a) включает введение в нуклеиновую кислоту-матрицу (например, РНК- или ДНК-матрицу) гетерологичной последовательности гомологии, характеризующейся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в молекуле ДНК-мишени, и/или

(b) включает введение в полипептид системы (например, содержащий ДНК-связывающий домен (DBD) и/или эндонуклеазный домен) гетерологичного нацеливающего домена, который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени.

366. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (a) включает вставку последовательности гомологии в нуклеиновую кислоту-матрицу.

367. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (a) включает замену сегмента нуклеиновой кислоты-матрицы последовательностью гомологии.

368. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (a) включает осуществление мутации одного или нескольких нуклеотидов (например, по меньшей мере 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов) нуклеиновой кислоты-матрицы, в результате чего образуется сегмент нуклеиновой кислоты-матрицы, имеющий последовательность из последовательности гомологии.

369. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (b) включает вставку аминокислотной последовательности нацеливающего домена в аминокислотную последовательность полипептида.

370. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (b) включает вставку последовательности нуклеиновой кислоты, кодирующей нацеливающий домен, в последовательность, кодирующую полипептид, содержащуюся в молекуле нуклеиновой кислоты.

371. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (b) включает замену по меньшей мере части полипептида нацеливающим доменом.

372. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (a) включает осуществление мутации одной или нескольких аминокислот (например, по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500 или больше аминокислот) полипептида.

373. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает приведение клетки в контакт с:

(a) полипептидом или нуклеиновой кислотой, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицей (или ДНК, кодирующей РНК-матрицу), содержащей (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью;

где:

(i) полипептид содержит гетерологичный нацеливающий домен (например, в DBD или эндонуклеазном домене), который специфично связывается с последовательностью, содержащейся в сайте-мишени в геномной ДНК или прилегающей к нему; и/или

(ii) РНК-матрица содержит гетерологичную последовательность гомологии, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в сайте-мишени в геномной ДНК или прилегающей к нему;

за счет чего обеспечивается модификация сайта-мишени в геномной ДНК в клетке.

374. Способ получения системы для модификации генома клетки млекопитающего, включающий:

(a) получение РНК-матрицы, содержащей (i) последовательность, которая связывает полипептид, содержащий домен обратной транскриптазы и эндонуклеазный домен, при этом последовательность характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью с: (i) нуклеотидами, расположенными с 5'-стороны относительно старт-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы), или (ii) нуклеотидами, расположенными с 3'-стороны относительно стоп-кодона последовательности элемента из таблицы 3B, таблицы 10 или таблицы X (например, содержащими область связывания ретротранспозазы).

375. Способ согласно любому из предыдущих вариантов осуществления, дополнительно включающий:

(b) обработку РНК-матрицы с целью ослабления вторичной структуры, например, нагревание РНК-матрицы, например, до по меньшей мере 70, 75, 80, 85, 90 или 95°C, и/или

(c) последующее охлаждение РНК-матрицы, например, до температуры, которая обеспечивает возможность существования вторичной структуры, например, до температуры, которая меньше или равняется 37, 30, 25 или 20°C.

376. Способ согласно любому из предыдущих вариантов осуществления, где система представляет собой систему согласно любому из предыдущих вариантов осуществления.

377. Способ согласно любому из предыдущих вариантов осуществления, который дополнительно включает приведение РНК-матрицы в контакт с полипептидом, который содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, или с нуклеиновой кислотой (например, РНК), кодирующей полипептид.

378. Способ согласно любому из предыдущих вариантов осуществления, который дополнительно включает приведение РНК-матрицы в контакт с клеткой.

379. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует терапевтический полипептид.

380. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует полипептид млекопитающего (например, человека) или его фрагмент или вариант.

381. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует фермент (например, лизосомальный фермент), фактор свертывания крови (например, фактор I, II, V, VII, X, XI, XII или XIII), мембранный белок, экзон, внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок), внеклеточный белок, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок, запасной белок, иммунорецепторный белок (например, химерный антигенный рецепторный белок (CAR), Т-клеточный рецептор, В-клеточный рецептор) или антитело.

382. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит тканеспецифический промотор или энхансер.

383. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует полипептид из более чем 250, 300, 400, 500 или 1000 аминокислот и необязательно вплоть до 1300 аминокислот.

384. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует фрагмент гена млекопитающего, но не кодирует полный ген млекопитающего, например, кодирует один или несколько экзонов, но не кодирует полноразмерный белок.

385. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует один или несколько интронов.

386. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта является отличной от GFP, например, является отличной от флуоресцентного белка или является отличной от репортерного белка.

387. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид обладает активностью при 37°C, составляющей не менее 70%, 75%, 80%, 85%, 90% или 95% от его активности при 25°C при в остальном аналогичных условиях.

388. Система или способ согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или

нуклеиновая кислота, кодирующая РНК-матрицу, являются отдельными нуклеиновыми кислотами.

389. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица не кодирует активную обратную транскриптазу, например, содержит инактивированную мутантную обратную транскриптазу, например, как описано в примерах 1 или 2, или не содержит последовательность обратной транскриптазы.

390. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит одну или несколько химических модификаций.

391. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта расположена между промотором и последовательностью, которая связывает полипептид.

392. Система или способ согласно любому из предыдущих вариантов осуществления, где промотор расположен между последовательностью гетерологичного объекта и последовательностью, которая связывает полипептид.

393. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит открытую рамку считывания (или последовательность, обратно комплементарную ей) в ориентации 5'-3' в РНК-матрице.

394. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит открытую рамку считывания (или последовательность, обратно комплементарную ей) в ориентации 3'-5' в РНК-матрице.

395. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит (a) домен обратной транскриптазы и (b) эндонуклеазный домен, где по меньшей мере один из (a) или (b) является гетерологичным.

396. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит (a) домен связывания ДНК-мишени, (b) домен обратной транскриптазы и (c) эндонуклеазный домен, где по меньшей мере один из (a), (b) или (c) является гетерологичным.

397. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

398. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность, кодируемую последовательностью элемента из

таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

399. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

400. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

401. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы, кодируемый первой последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и (b) эндонуклеазный домен, кодируемый второй последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; где первая последовательность и вторая последовательность выбраны из элементов в разных строках таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

402. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы, кодируемый первой последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и (b) эндонуклеазный домен, кодируемый второй последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; где первая последовательность и вторая последовательность выбраны из элементов в разных строках таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

403. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы первой последовательности элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательности, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и (b) эндонуклеазный домен второй последовательности элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательности, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; где первая последовательность и вторая последовательность выбраны из элементов в разных строках таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

404. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы первой последовательности элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательности, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и (b) эндонуклеазный домен второй последовательности элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательности, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; где первая последовательность и вторая последовательность выбраны из элементов в разных строках таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

405. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени, кодируемый последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

406. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени, кодируемый последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

407. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним.

408. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от него.

409. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен; где (a) имеет аминокислотную последовательность, указанную в таблице 3B, таблице 10, таблице 11, или таблице Z1 или Z2, или таблице X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и где необязательно (b) имеет аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

410. По сути чистый полипептид, содержащий (a) домен обратной транскриптазы и (b) гетерологичный эндонуклеазный домен; где (a) имеет аминокислотную последовательность, указанную в таблице 3B, таблице 10, таблице 11, или таблице Z1 или

Z2, или таблице X, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и где необязательно (b) имеет аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

411. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени, содержащий аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

412. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, дополнительно содержащий домен связывания ДНК-мишени, содержащий аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

413. По сути чистый полипептид, содержащий (a) домен связывания ДНК-мишени, кодируемый первой последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, (b) домен обратной транскриптазы, кодируемый второй последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, и (c) эндонуклеазный домен, кодируемый третьей последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X; где:

(i) первая последовательность и вторая последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(ii) первая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(iii) вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X; или

(iv) первая последовательность, вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

414. По сути чистый полипептид, содержащий (a) домен связывания ДНК-мишени первой последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, (b) домен обратной транскриптазы второй последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, и (c) эндонуклеазный домен третьей последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X; где:

(i) первая последовательность и вторая последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(ii) первая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(iii) вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X; или

(iv) первая последовательность, вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

415. По сути чистый полипептид, содержащий (a) домен связывания ДНК-мишени, например, содержащий первую аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, (b) домен обратной транскриптазы, содержащий вторую аминокислотную последовательность, указанную в таблице Z1 или Z2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и (c) эндонуклеазный домен, содержащий третью аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней;

где необязательно элементы первой аминокислотной последовательности и третьей аминокислотной последовательности выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

416. По сути чистый полипептид, содержащий (a) домен связывания ДНК-мишени, например, содержащий первую аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, (b) домен обратной транскриптазы, содержащий вторую аминокислотную последовательность, указанную в таблице Z1 или Z2, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и (c) эндонуклеазный домен, содержащий третью аминокислотную последовательность элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, или последовательность, имеющую не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее;

где необязательно элементы первой аминокислотной последовательности и третьей аминокислотной последовательности выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

417. По сути чистый полипептид, содержащий (a) домен связывания ДНК-мишени, содержащий первую аминокислотную последовательность, (b) домен обратной транскриптазы, содержащий вторую аминокислотную последовательность, и (c) эндонуклеазный домен, содержащий третью аминокислотную последовательность; где каждая из первой, второй и третьей аминокислотных последовательностей кодируется последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, или содержит аминокислотную последовательность, указанную в таблице Z1, или аминокислотную последовательность домена, указанного в таблице Z2.

418. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где первая аминокислотная последовательность кодируется последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X.

419. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где первая аминокислотная последовательность содержит аминокислотную последовательность, указанную в таблице Z1 или Z2.

420. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где вторая аминокислотная последовательность кодируется последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X.

421. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где вторая аминокислотная последовательность содержит аминокислотную последовательность, указанную в таблице Z1 или Z2.

422. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где третья аминокислотная последовательность кодируется последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X.

423. По сути чистый полипептид согласно любому из предыдущих вариантов осуществления, где третья аминокислотная последовательность содержит аминокислотную последовательность, указанную в таблице Z1 или Z2.

424. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен обратной транскриптазы и (b) эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и где по меньшей мере один из (a) или (b) является гетерологичным по отношению к другому.

425. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен обратной транскриптазы и (b) эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и где по меньшей мере один из (a) или (b) является гетерологичным по отношению к другому.

426. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид представляет собой слитый белок, содержащий (a) домен обратной транскриптазы, кодируемый первой последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X, и (b) эндонуклеазный домен, кодируемый второй последовательностью элемента, указанного в таблице 3B, таблице 10, таблице 11 или таблице X; где первая последовательность и вторая последовательность выбраны из элементов в разных строках таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

427. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен связывания ДНК-мишени, (b) домен обратной транскриптазы и (c) эндонуклеазный домен, где по меньшей мере один (например, 1, 2 или все) из (a), (b) или (c) содержит аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и где по меньшей мере один из (a), (b) или (c) является гетерологичным по отношению к другому.

428. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен связывания ДНК-мишени, (b) домен обратной транскриптазы и (c) эндонуклеазный домен, где по меньшей мере один (например, 1, 2 или все) из (a), (b) или (c) содержит аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и где по меньшей мере один из (a), (b) или (c) является гетерологичным по отношению к другому.

429. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид представляет собой слитый белок, содержащий (a) домен связывания ДНК-мишени, кодируемый первой последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, (b) домен обратной транскриптазы, кодируемый второй последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, и (c) эндонуклеазный домен, кодируемый третьей последовательностью, указанной в таблице 3B, таблице 10, таблице 11 или таблице X; где:

(i) первая последовательность и вторая последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(ii) первая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(iii) вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X; или

(iv) первая последовательность, вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

430. Любой полипептид согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы характеризуется по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с доменом обратной транскриптазы, кодируемым последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

431. Любой полипептид согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен характеризуется по меньшей мере 80% идентичностью, например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью, с

эндонуклеазным доменом, кодируемым последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

432. Любой полипептид согласно любому из предыдущих вариантов осуществления, где ДНК-связывающий домен характеризуется по меньшей мере 80% идентичностью, например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью, с ДНК-связывающим доменом, кодируемым последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

433. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен обратной транскриптазы и (b) эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и где по меньшей мере один из (a) или (b) является гетерологичным по отношению к другому.

434. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен обратной транскриптазы и (b) эндонуклеазный домен, где один или оба из (a) или (b) имеют аминокислотную последовательность последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательности, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и где по меньшей мере один из (a) или (b) является гетерологичным по отношению к другому.

435. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид представляет собой слитый белок, содержащий (a) домен обратной транскриптазы первой последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, и (b) эндонуклеазный домен второй последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X; где первая последовательность и вторая последовательности выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

436. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен связывания ДНК-мишени, (b) домен обратной транскриптазы и (c) эндонуклеазный домен, где по меньшей мере один (например, 1, 2 или все) из (a), (b) или (c) содержит аминокислотную последовательность последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, и где по меньшей мере один из (a), (b) или (c) является гетерологичным по отношению к другому.

437. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (a) домен связывания ДНК-мишени, (b) домен обратной транскриптазы и (c) эндонуклеазный домен, где по меньшей мере один (например, 1, 2 или все) из (a), (b) или (c) содержит аминокислотную последовательность последовательности

элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательности, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее, и где по меньшей мере один из (a), (b) или (c) является гетерологичным по отношению к другому.

438. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид представляет собой слитый белок, содержащий (a) домен связывания ДНК-мишени первой последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, (b) домен обратной транскриптазы второй последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X, и (c) эндонуклеазный домен третьей последовательности, указанной в таблице 3B, таблице 10, таблице 11 или таблице X; где:

(i) первая последовательность и вторая последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(ii) первая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X;

(iii) вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X; или

(iv) первая последовательность, вторая последовательность и третья последовательность выбраны из разных строк таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

439. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен, где домен RT имеет последовательность из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

440. Любой полипептид согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы характеризуется по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с доменом обратной транскриптазы последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

441. Любой полипептид согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен характеризуется по меньшей мере 80% идентичностью, например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью, с эндонуклеазным доменом последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

442. Любой полипептид согласно любому из предыдущих вариантов осуществления, где ДНК-связывающий домен характеризуется по меньшей мере 80% идентичностью, например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью, с ДНК-связывающим доменом последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

443. Нуклеиновая кислота, кодирующая полипептид согласно любому из предыдущих вариантов осуществления.

444. Вектор, содержащий нуклеиновую кислоту согласно любому из предыдущих вариантов осуществления.

445. Клетка-хозяин, содержащая нуклеиновую кислоту согласно любому из предыдущих вариантов осуществления.

446. Клетка-хозяин, содержащая полипептид согласно любому из предыдущих вариантов осуществления.

447. Клетка-хозяин, содержащая вектор согласно любому из предыдущих вариантов осуществления.

448. Клетка-хозяин (например, клетка человека), содержащая:

последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид) в сайте-мишени в хромосоме и одно или оба из:

(a) одной или обеих из нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 5 таблицы 3A или 3B) с одной стороны от последовательности гетерологичного объекта (например, выше нее) и нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 6 таблицы 3A или 3B) с другой стороны от последовательности гетерологичного объекта (например, ниже нее) и/или

(b) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

449. Клетка-хозяин (например, клетка человека), содержащая:

последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид) в сайте-мишени в хромосоме и одно или оба из:

(a) одной или обеих из нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 5 таблицы 3A или 3B) с одной стороны от последовательности гетерологичного объекта (например, выше нее) и нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 6 таблицы 3A или 3B) с другой стороны от последовательности гетерологичного объекта (например, ниже нее) и/или

(b) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где

один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

450. Клетка-хозяин (например, клетка человека), содержащая:

последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид) в сайте-мишени в хромосоме и одно или оба из:

(a) одной или обеих из нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 5 таблицы 3A или 3B) с одной стороны от последовательности гетерологичного объекта (например, выше нее) и нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 6 таблицы 3A или 3B) с другой стороны от последовательности гетерологичного объекта (например, ниже нее) и/или

(b) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательности, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

451. Клетка-хозяин (например, клетка человека), содержащая:

последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид) в сайте-мишени в хромосоме и одно или оба из:

(a) одной или обеих из нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из столбца 5 таблицы 10 или таблицы 3A или 3B) с одной стороны от последовательности гетерологичного объекта (например, выше нее) и нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из столбца 6 таблицы 10 или таблицы 3A или 3B) с другой стороны от последовательности гетерологичного объекта (например, ниже нее) и/или

(b) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательности, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее.

452. Клетка-хозяин согласно любому из предыдущих вариантов осуществления, содержащая: (i) последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид) в сайте-мишени в

хромосоме, где локус-мишень представляет собой сайт Natural Harbor™, например, сайт из таблицы 4, приведенной в данном документе.

453. Клетка-хозяин согласно любому из предыдущих вариантов осуществления, которая дополнительно содержит (ii) одну или обе из 5'-нетранслируемой области последовательности гетерологичного объекта и 3'-нетранслируемой области последовательности гетерологичного объекта.

454. Клетка-хозяин согласно любому из предыдущих вариантов осуществления, которая дополнительно содержит (ii) одну или обе из нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 5 таблицы 3A или 3B) с одной стороны от последовательности гетерологичного объекта (например, выше нее) и нетранслируемой области (например, нетранслируемой последовательности ретротранспозона, например, последовательности из таблицы 10 или столбца 6 таблицы 3A или 3B) с другой стороны от последовательности гетерологичного объекта (например, ниже нее).

455. Клетка-хозяин согласно любому из предыдущих вариантов осуществления, которая содержит последовательность гетерологичного объекта только в сайте-мишени.

456. Фармацевтическая композиция, содержащая любую вышеперечисленную систему, нуклеиновую кислоту, полипептид или вектор и фармацевтически приемлемый наполнитель или носитель.

457. Фармацевтическая композиция согласно любому из предыдущих вариантов осуществления, где фармацевтически приемлемый наполнитель или носитель выбран из вектора (например, вирусного или плазмидного вектора), везикулы (например, липосомы, экзосомы, природного или синтетического липидного бислоя), фузосомы, липидной наночастицы.

458. Полипептид согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит последовательность ядерной локализации.

459. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

460. РНК-матрица согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (i).

461. РНК-матрица согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит (ii).

462. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая (например, в направлении от 5' к 3') (i) последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) последовательность, которая

специфично связывает домен RT полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

463. РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен RT содержит последовательность, выбранную из таблицы 3B, таблицы 10, таблицы 11 или таблицы X, или последовательность, которая характеризуется по меньшей мере 70%, 75%, 80% 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

464. РНК-матрица согласно любому из предыдущих вариантов осуществления, дополнительно содержащая (v) последовательность, которая связывает эндонуклеазный и/или ДНК-связывающий домен полипептида (например, того же полипептида, который содержит домен RT).

465. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность из (ii) специфично связывает домен RT из таблицы 3B, таблицы 10, таблицы 11 или таблицы X или последовательность домена RT, которая характеризуется по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ним.

466. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, представляет собой последовательность из таблицы 10, таблицы 11 или таблицы X, 3A или 3B или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

467. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида, (i) последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

468. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (iii) последовательность гетерологичного объекта, (iv) 3'-концевой домен гомологии, (i) последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), и (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида.

469. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где РНК-матрица, первая РНК-матрица или вторая РНК-матрица содержит последовательность, которая специфично связывает домен RT.

470. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, расположена между (i) и (ii).

471. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, расположена между (ii) и (iii).

472. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, расположена между (iii) и (iv).

473. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, расположена между (iv) и (i).

474. Система или РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, расположена между (i) и (iii).

475. Система для модификации ДНК, содержащая:

(a) первую РНК-матрицу (или ДНК, кодирующую первую РНК-матрицу), содержащую (i) последовательность, которая связывает эндонуклеазный домен, например, никазный домен, и/или ДНК-связывающий домен (DBD) полипептида, и (ii) последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени) (например, где первая РНК содержит gRNA);

(b) вторую РНК-матрицу (или ДНК, кодирующую вторую РНК-матрицу), содержащую (i) последовательность, которая специфично связывает домен обратной транскриптазы (RT) полипептида (например, полипептида из (a)), (ii) последовательность, связывающую сайт-мишень (TSBS), и (iii) последовательность матрицы для RT.

476. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая первую РНК-матрицу, и нуклеиновая кислота, кодирующая вторую РНК-матрицу, являются двумя отдельными нуклеиновыми кислотами.

477. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая первую РНК-матрицу, и нуклеиновая кислота, кодирующая вторую РНК-матрицу, являются частью одной и той же молекулы нуклеиновой кислоты, например, присутствуют в одном и том же векторе.

478. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение любой вышеперечисленной системы в клетку, ткань или субъекту, за счет чего обеспечивается модификация нити ДНК-мишени.

479. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X выбрана из ретротранспозаз Vingi-1 EE, BovB, AviRTE_Brh, Penelope_SM и Utopia_Dyak.

480. Любой вышеперечисленный вариант осуществления, где РНК-матрица содержит 5'-UTR и 3'-UTR из одной и той же последовательности элемента из таблицы 3B, таблицы 10 или таблицы X.

481. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X содержит ретротранспозазу Vingi-1 EE, и где РНК-матрица содержит 5'-UTR и 3'-UTR Vingi-1EE.

482. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X принадлежит к кладе эндонуклеаз, подобных рестрикционной эндонуклеазе (RLE).

483. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X принадлежит к кладе эндонуклеаз, подобных апуриновой эндонуклеазе (APE).

484. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X принадлежит к кладе Penelope-подобных элементов (PLE), где необязательно последовательность элемента из таблицы X содержит домен GIY-YIG (например, эндонуклеазный домен GIY-YIG).

485. Любой вышеперечисленный вариант осуществления, где последовательность элемента из таблицы X принадлежит к кладе, выбранной из клад CRE, NeSL, R4, R2, Hero, L1, RTE (например, AviRTE_Brh или BovB), I, Jockey, CR1, Rex1, RandI/Dualen, Penelope (например, Penelope_SM), Tx1, RTEX, Crack, Nimb, Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi (например, Vingi-1_EE) и Kiri.

486. Любой вышеперечисленный вариант осуществления, где домен связывания ДНК-мишени является гетерологичным по отношению к одному или нескольким другим доменам полипептида (например, домену обратной транскриптазы и/или эндонуклеазному домену).

487. Любой вышеперечисленный вариант осуществления, где гетерологичный домен связывания ДНК-мишени содержит Cas9, никазу Cas9, dCas9, "цинковый палец" или домен TAL.

488. Любой вышеперечисленный вариант осуществления, где гетерологичный домен связывания ДНК-мишени содержит домен Cas в соответствии с таблицей 9 или таблицей 37 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичности с ним.

489. Любой вышеперечисленный вариант осуществления, где гетерологичный домен связывания ДНК-мишени содержит N-концевой домен dCas9.

490. Любой вышеперечисленный вариант осуществления, где система дополнительно содержит направляющую РНК (например, U6-управляемую gRNA), содержащую по меньшей мере 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 или 25 нуклеотидов, являющихся гомологичными по отношению к последовательности ДНК-мишени.

491. Любой вышеперечисленный вариант осуществления, где РНК-матрица дополнительно содержит область gRNA, содержащую по меньшей мере 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 или 25 нуклеотидов, являющихся гомологичными по отношению к последовательности ДНК-мишени.

492. Любой вышеперечисленный вариант осуществления, где последовательность gRNA находится на 5'-конце матрицы.

493. Любой вышеперечисленный вариант осуществления, где последовательность gRNA находится на 3'-конце матрицы.

494. Любой вышеперечисленный вариант осуществления, где последовательность gRNA содержит каркас, способный обеспечивать привлечение Cas9.

495. Любой вышеперечисленный вариант осуществления, где последовательность gRNA содержит домен гомологии, например, описанный в данный документе.

496. Любой вышеперечисленный вариант осуществления, где эндонуклеазный домен является гетерологичным по отношению к одному или нескольким другим доменам полипептида (например, домену обратной транскриптазы и/или домену связывания ДНК-мишени).

497. Любой вышеперечисленный вариант осуществления, где гетерологичный эндонуклеазный домен содержит Cas9, никазу Cas9 или домен FokI.

498. Любой вышеперечисленный вариант осуществления, где полипептид содержит домен РНКазы H.

499. Любой вышеперечисленный вариант осуществления, где полипептид не содержит домен РНКазы H или содержит инактивированный домен РНКазы H.

500. Любой вышеперечисленный вариант осуществления, где нуклеиновая кислота, кодирующая полипептид, дополнительно содержит вторую открытую рамку считывания.

501. Любой вышеперечисленный вариант осуществления, где нуклеиновая кислота, кодирующая полипептид, содержит последовательность 2A, например, расположенную между последовательностью ORF1 и последовательностью ORF2, где необязательно последовательность 2A выбрана из T2A (EGRGSLLTCGDVEENPGP), P2A (ATNFSLLKQAGDVEENPGP), E2A (QCTNYALLKLAGDVESNPGP) или F2A (VKQTLNFDLLKLAGDVESNPGP).

502. Любой вышеперечисленный вариант осуществления, где полипептид содержит интеин.

503. Любой вышеперечисленный вариант осуществления, где система содержит интеин (например, содержащийся во втором полипептиде).

504. Любой вышеперечисленный вариант осуществления, где полипептид кодируется двумя или более отдельными открытыми рамками считывания, каждая из которых кодирует фрагмент полипептида.

505. Любой вышеперечисленный вариант осуществления, где интеин (например, транс-сплайсинговый интеин) соединяет два или более фрагмента полипептида с образованием полипептида.

506. Любой вышеперечисленный вариант осуществления, где система содержит: (i) первый фрагмент полипептида, содержащий по меньшей мере один из домена обратной транскриптазы, эндонуклеазного домена и домена связывания ДНК-мишени, и (ii) второй фрагмент полипептида, содержащий по меньшей мере один из домена обратной транскриптазы, эндонуклеазного домена и домена связывания ДНК-мишени, где первый фрагмент полипептида не содержит тот же тип домена, что и второй фрагмент полипептида.

507. Любой вышеперечисленный вариант осуществления, где:

(a) первый фрагмент полипептида содержит домен обратной транскриптазы, а второй фрагмент полипептида содержит эндонуклеазный домен, необязательно где первый фрагмент полипептида дополнительно содержит домен связывания ДНК-мишени или где второй фрагмент полипептида дополнительно содержит домен связывания ДНК-мишени;

(b) первый фрагмент полипептида содержит домен обратной транскриптазы, а второй фрагмент полипептида содержит домен связывания ДНК-мишени, необязательно где первый фрагмент полипептида дополнительно содержит эндонуклеазный домен или где второй фрагмент полипептида дополнительно содержит эндонуклеазный домен; или

(a) первый фрагмент полипептида содержит эндонуклеазный домен, а второй фрагмент полипептида содержит домен связывания ДНК-мишени, необязательно где первый фрагмент полипептида дополнительно содержит домен обратной транскриптазы, или где второй фрагмент полипептида дополнительно содержит домен обратной транскриптазы.

508. Любой вышеперечисленный вариант осуществления, где интеин соединяет первый фрагмент полипептида со вторым полипептидом с образованием полипептида.

509. Любой вышеперечисленный вариант осуществления, где интеин индуцирует слияние:

(i) домена обратной транскриптазы с эндонуклеазным доменом,

(ii) домена обратной транскриптазы с доменом связывания ДНК-мишени или

(iii) эндонуклеазного домена с доменом связывания ДНК-мишени.

510. Любой вышеперечисленный вариант осуществления, где интеин является гетерологичным по отношению к одному или нескольким (например, 1, 2 или всем) из домена обратной транскриптазы, эндонуклеазного домена и домена связывания ДНК-мишени.

511. Любой вышеперечисленный вариант осуществления, где интеин представляет собой расщепленный интеин.

512. Любой из предыдущих вариантов осуществления, где ДНК, кодирующая полипептид, содержит плазмиду, миникольцо, ДНК Doggybone (dbDNA) или ceDNA.

513. Любой из предыдущих вариантов осуществления, где РНК, кодирующая полипептид, содержит одно или несколько из следующего: область кэпирования, поли(А)-хвост и/или химическую модификацию, например, один или несколько химически модифицированных нуклеотидов.

514. Любой из предыдущих вариантов осуществления, где РНК, кодирующая полипептид, содержит circRNA.

515. Любой из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, содержится в вирусе (например, AAV, аденовирусе или лентивирусе, например, лентивирусе, дефектном по интеграции).

516. Любой из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, содержится в наночастице (например, липидной наночастице), везикуле или фузосоме.

517. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью, кодируемой последовательностью элемента из таблицы X.

518. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с доменом обратной транскриптазы аминокислотной последовательности, кодируемой последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

519. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью, кодируемой последовательностью элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

520. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

521. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с доменом обратной транскриптазы аминокислотной последовательности последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

522. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью последовательности элемента из таблицы 3B, таблицы 10, таблицы 11 или таблицы X.

523. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

524. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100%

идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

525. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

526. Любой вышеперечисленный вариант осуществления, где полипептид, домен обратной транскриптазы или ретротранспозаза содержат линкер, содержащий аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

527. Любой вышеперечисленный вариант осуществления, где полипептид содержит ДНК-связывающий домен, ковалентно присоединенный к остальной части полипептида посредством линкера, например, линкера, содержащего по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400, или 500 аминокислот.

528. Любой вышеперечисленный вариант осуществления, где линкер присоединен к остальной части полипептида в положении в ДНК-связывающем домене, РНК-связывающем домене, домене обратной транскриптазы или эндонуклеазном домене (например, как показано на любой из фигур 17A-17F в заявке согласно PCT № PCT/US2019/048607).

529. Любой вышеперечисленный вариант осуществления, где линкер присоединен к остальной части полипептида в положении с N-концевой стороны от области альфа-спирали полипептида, например, в положении, соответствующем варианту v1, как описано в примере 26 в заявке согласно PCT № PCT/US2019/048607.

530. Любой вышеперечисленный вариант осуществления, где линкер присоединен к остальной части полипептида в положении с C-концевой стороны от области альфа-спирали полипептида, например, перед РНК-связывающим мотивом (например, в положении -1 относительно РНК-связывающего мотива), например, в положении, соответствующем варианту v2, как описано в примере 26 в заявке согласно PCT № PCT/US2019/048607.

531. Любой вышеперечисленный вариант осуществления, где линкер присоединен к остальной части полипептида в положении с C-концевой стороны от области статистического клубка полипептида, например, с N-концевой стороны относительно ДНК-связывающего мотива (например, ДНК-связывающего мотива c-Myb), например, в положении, соответствующем варианту v3, как описано в примере 26 в заявке согласно PCT № PCT/US2019/048607.

532. Любой вышеперечисленный вариант осуществления, где линкер содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

533. Любой вышеперечисленный вариант осуществления, где полинуклеотидная последовательность, содержащая по меньшей мере приблизительно 500, 1000, 2000, 3000, 3500, 3600, 3700, 3800, 3900 или 4000 смежных нуклеотидов с 5'-конца последовательности РНК-матрицы, интегрирована в геном клетки-мишени.

534. Любой вышеперечисленный вариант осуществления, где полинуклеотидная последовательность, содержащая по меньшей мере приблизительно 500, 1000, 2000, 2500, 2600, 2700, 2800, 2900 или 3000 смежных нуклеотидов с 3'-конца последовательности РНК-матрицы, интегрирована в геном клетки-мишени.

535. Любой вышеперечисленный вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9 или 1,0 интегрированного элемента/геном.

536. Любой вышеперечисленный вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,01, 0,02, 0,03, 0,04, 0,05, 0,75 или 0,1 интегрированного элемента/геном.

537. Любой вышеперечисленный вариант осуществления, где полипептид содержит функциональный эндонуклеазный домен (например, где эндонуклеазный домен не содержит мутацию, приводящую к устранению эндонуклеазной активности, например, как описано в данном документе).

538. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень не приводит к изменению (например, повышению) уровней белков p53 и/или p21, уровня фосфорилирования H2AX (например, гамма-H2AX), уровня фосфорилирования ATM, уровня фосфорилирования ATR, уровня фосфорилирования Chk1, уровня фосфорилирования Chk2 и/или уровня фосфорилирования p53.

539. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня белка p53 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня белка p53, индуцируемого путем введения сайт-специфической

нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

540. Любой вышеперечисленный вариант осуществления, где уровень белка p53 определяется в соответствии со способом, описанным в примере 30.

541. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования p53 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования p53, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

542. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня белка p21 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня белка p53, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

543. Любой вышеперечисленный вариант осуществления, где уровень белка p21 определяется в соответствии со способом, описанным в примере 30.

544. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования H2AX в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования H2AX, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

545. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования ATM в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования ATM, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

546. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования ATR в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования ATR, индуцируемого путем

введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

547. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования Chk1 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования Chk1, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

548. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования Chk2 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, или 90% от уровня фосфорилирования Chk2, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

549. Любой вышеперечисленный вариант осуществления, где домен связывания ДНК-мишени распознает конкретную последовательность ДНК-мишени.

550. Любой вышеперечисленный вариант осуществления, где домен связывания ДНК-мишени связывается с множеством (например, случайных) последовательностей ДНК-мишени.

551. Любой вышеперечисленный вариант осуществления, где РНК-матрица содержит направляющую РНК (например, U6-управляемую gRNA).

552. Любой вышеперечисленный вариант осуществления, где домен связывания ДНК-мишени содержит один или несколько из ДНК-связывающего домена ретротранспозазы, описанной в данном документе (например, ретротранспозазы элемента из таблицы X, 10, 11, 3A или 3B), Cas9, никазы Cas9, dCas9, "цинкового пальца", TAL, мегануклеазы и/или фактора транскрипции.

553. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит домен обратной транскриптазы ретротранспозазы, описанной в данном документе (например, ретротранспозазы элемента из таблицы X, 10, 11, Z1, Z2, 3A или 3B).

554. Любой вышеперечисленный вариант осуществления, где эндонуклеазный домен содержит эндонуклеазный домен ретротранспозазы, описанной в данном документе (например, ретротранспозазы элемента из таблицы X, 10, 11, 3A или 3B), Cas9, никазы Cas9, фермента рестрикции II типа (например, FokI), резольвазы для структур Холлидея, эндонуклеазный домен RLE, эндонуклеазный домен APE или эндонуклеазный домен GIY-YIG.

555. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий

домен (DBD) и (iii) эндонуклеазный домен; где DBD и/или эндонуклеазный домен содержат гетерологичный нацеливающий домен, который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени (например, геномной ДНК).

556. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая нацеливающий домен (например, гетерологичный нацеливающий домен), который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени (например, геномной ДНК), последовательность, которая специфично связывается с доменом RT полипептида, и последовательность гетерологичного объекта.

557. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где полипептид содержит гетерологичный нацеливающий домен, который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени (например, геномной ДНК).

558. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичный нацеливающий домен связывается с другой последовательностью нуклеиновой кислоты, нежели немодифицированный полипептид.

559. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где полипептид не содержит функциональный эндогенный нацеливающий домен (например, полипептид не содержит эндогенный нацеливающий домен).

560. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичный нацеливающий домен содержит "цинковый палец" (например, "цинковый палец", который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени).

561. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичный нацеливающий домен содержит домен Cas (например, домен Cas9 или его мутантную форму или вариант, например, домен Cas9, который специфично связывается с последовательностью, содержащейся в молекуле ДНК-мишени).

562. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен Cas ассоциирован с направляющей РНК (gRNA).

563. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичный нацеливающий домен содержит эндонуклеазный домен (например, гетерологичный эндонуклеазный домен).

564. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен содержит домен Cas (например, Cas9 или его мутантную форму или вариант).

565. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен Cas ассоциирован с направляющей РНК (gRNA).

566. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен содержит домен Fok1.

567. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит по меньшей мере

одну (например, одну или две) гетерологичную последовательность гомологии, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в молекуле ДНК-мишени (например, геномной ДНК).

568. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где одна из по меньшей мере одной гетерологичной последовательности гомологии расположена на расстоянии приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от 5'-конца молекулы нуклеиновой кислоты-матрицы.

569. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где одна из по меньшей мере одной гетерологичной последовательности гомологии расположена на расстоянии приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от 3'-конца молекулы нуклеиновой кислоты-матрицы.

570. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии связывается в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов от сайта однонитевого разрыва (например, создаваемого с помощью никазы, например, эндонуклеазного домена, например, как описано в данном документе) в молекуле ДНК-мишени.

571. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется менее чем 50%, 40%, 30%, 20%, 10%, 5%, 4%, 3%, 2% или 1% идентичностью последовательности с последовательностью нуклеиновой кислоты, комплементарной эндогенной последовательности гомологии немодифицированной формы РНК-матрицы.

572. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью молекулы ДНК-мишени, которая отличается от последовательности, связываемой эндогенной последовательностью гомологии (например, замененной гетерологичной последовательностью гомологии).

573. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии содержит последовательность (например, на ее 3'-конце), характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, расположенной с 5'-стороны от сайта однонитевого разрыва молекулы ДНК-мишени (например, сайта, подвергаемого однонитевому разрыву никазой, например, эндонуклеазным доменом, как описано в данном документе).

574. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии содержит

последовательность (например, на ее 5'-конце), подходящую для инициации обратной транскрипции с праймированием мишенью (TPRT).

575. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, расположенной в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов от сайта-мишени для вставки (например, с 3'-стороны относительно него), например, для последовательности гетерологичного объекта (например, как описано в данном документе), в молекуле ДНК-мишени.

576. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит направляющую РНК (gRNA), например, как описано в данном документе.

577. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит спейсерную последовательность gRNA (например, на расстоянии 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от ее 5'-конца).

578. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая (например, в направлении от 5' к 3') (i) последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), (ii) последовательность, которая специфично связывает домен RT полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

579. РНК-матрица согласно любому из предыдущих вариантов осуществления, дополнительно содержащая (v) последовательность, которая связывает эндонуклеазный и/или ДНК-связывающий домен полипептида (например, того же полипептида, который содержит домен RT).

580. РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен RT содержит последовательность, выбранную из таблицы 3B, 10, 11 или X, или последовательность, которая характеризуется по меньшей мере 70%, 75%, 80% 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

581. РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен RT содержит последовательность, выбранную из таблицы 3B, 10, 11 или X, где домен RT дополнительно содержит ряд замен по сравнению с природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

582. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность из (ii) специфично связывается с доменом RT.

583. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, представляет собой последовательность, например, последовательность UTR, из таблицы 3B или 10 или

последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

584. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида, (i) последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

585. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (iii) последовательность гетерологичного объекта, (iv) 3'-концевой домен гомологии с мишенью, (i) последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), и (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида.

586. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК системы (например, РНК-матрица, РНК, кодирующая полипептид из (a), или РНК, экспрессирующаяся из последовательности гетерологичного объекта, интегрированной в ДНК-мишень) содержит сайт связывания с микроРНК, например, в 3'-UTR.

587. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где сайт связывания с микроРНК распознается с помощью miRNA, которая присутствует в типе клеток, не являющихся мишенями, но не присутствует (или присутствует на сниженном уровне по сравнению с клетками, не являющимися мишенями) в типе клеток, являющихся мишенями.

588. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где miRNA представляет собой miR-142, и/или где клетка, не являющаяся мишенью, представляет собой клетку Купфера или клетку крови, например, иммунную клетку.

144. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где miRNA представляет собой miR-182 или miR-183, и/или где клетка, не являющаяся мишенью, представляет собой ганглиозный нейрон дорсального корешка.

588. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система содержит первый сайт связывания с miRNA, который распознается с помощью первой miRNA (например, miR-142), и система дополнительно содержит второй сайт связывания с miRNA, который распознается с помощью второй miRNA (например, miR-182 или miR-183), где первый сайт связывания с miRNA и второй сайт связывания с miRNA расположены на одной и той же РНК или на разных РНК одной системы.

589. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 2,

3 или 4 сайта связывания с miRNA, например, где сайты связывания с miRNA распознаются одной и той же или разными miRNA.

590. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК, кодирующая полипептид из (a), содержит по меньшей мере 2, 3 или 4 сайта связывания с miRNA, например, где сайты связывания с miRNA распознаются одной и той же или разными miRNA.

591. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК, экспрессирующаяся из последовательности гетерологичного объекта, интегрированной в ДНК-мишень, содержит по меньшей мере 2, 3 или 4 сайта связывания с miRNA, например, где сайты связывания с miRNA распознаются одной и той же или разными miRNA.

Определения

Домен. Термин "домен", используемый в данном документе, относится к структуре биомолекулы, которая участвует в осуществлении определенной функции биомолекулы. Домен может содержать непрерывную область (например, непрерывную последовательность) или отдельные несмежные области (например, несмежные последовательности) биомолекулы. Примеры белковых доменов включают без ограничения эндонуклеазный домен, ДНК-связывающий домен, домен обратной транскрипции; примером домена нуклеиновой кислоты является регуляторный домен, такой как домен связывания с фактором транскрипции.

Экзогенный. Используемый в данном документе термин "экзогенный" при использовании его в отношении биомолекулы (такой как последовательность нуклеиновой кислоты или полипептид) означает, что биомолекула была введена в геном, клетку или организм хозяина посредством манипуляций, осуществляемых человеком. Например, нуклеиновая кислота, которую добавляют в существующий геном, клетку, ткань или субъекту с применением методик рекомбинантной ДНК или других способов, является экзогенной по отношению к существующей последовательности нуклеиновой кислоты, клетке, ткани или субъекту.

Геномный сайт "безопасная гавань" (сайт GSH). Геномный сайт "безопасная гавань" представляет собой сайт в геноме хозяина, способный служить местом для интеграции нового генетического материала, например, таким образом, что вставленный генетический элемент не вызывает значительных изменений генома хозяина, представляющих риск для клетки- или организма-хозяина. Сайт GSH, как правило, соответствует 1, 2, 3, 4, 5, 6, 7, 8 или 9 из следующих критериев: (i) расположен на расстоянии > 300 т. о. от гена, связанного с раком; (ii) находится на расстоянии > 300 т. о. от miRNA/другой функциональной малой РНК; (iii) находится на расстоянии > 50 т. о. от 5'-конца гена; (iv) находится на расстоянии > 50 т. о. от точки начала репликации; (v) удален на расстояние > 50 т. о. от любого ультраконсервативного элемента; (vi) обладает низкой транскрипционной активностью (т. е. мРНК отсутствует на расстоянии +/- 25 т. о.); (vii) не располагается в области вариации числа копий; (viii) находится в открытом хроматине

и/или (ix) является уникальным и имеет 1 копию в человеческом геноме. Примеры сайтов GSH в человеческом геноме, которые удовлетворяют некоторым или всем из данных критериев, включают (i) сайт 1 для аденоассоциированного вируса (AAVS1) - встречающийся в природе сайт интеграции AAV-вируса в хромосоме 19; (ii) ген хемокинового (с мотивом C-C) рецептора 5 (CCR5) - ген хемокинового рецептора, известного как корецептор HIV-1; (iii) человеческий ортолог мышиного локуса Rosa26; (iv) локус рДНК. Дополнительные сайты GSH известны и описаны, например, в Pellenz et al., электронная публикация от 20 августа 2018 года (https://doi.org/10.1101/396390).

Гетерологичный. Термин "гетерологичный" при использовании для описания первого элемента по отношению ко второму элементу означает, что первый элемент и второй элемент не существуют в природе в описываемом расположении. Например, гетерологичные полипептид, молекула нуклеиновой кислоты, конструкция или последовательность относятся к (a) полипептиду, молекуле нуклеиновой кислоты или части последовательности полипептида или молекулы нуклеиновой кислоты, не являющимся нативными по отношению к клетке, в которой они экспрессируются, (b) полипептиду, или молекуле нуклеиновой кислоты, или части полипептида или молекулы нуклеиновой кислоты, которые были изменены или подвергнуты мутации по сравнению с их нативным состоянием, или (c) полипептиду или молекуле нуклеиновой кислоты с измененной экспрессией по сравнению с нативными уровнями экспрессии при аналогичных условиях. Например, гетерологичная регуляторная последовательность (например, промотор, энхансер) может применяться для регуляции экспрессии гена или молекулы нуклеиновой кислоты таким образом, чтобы она отличалась от экспрессии гена или молекулы нуклеиновой кислоты, обычно наблюдающейся в природе. В другом примере гетерологичные домен полипептида или последовательность нуклеиновой кислоты (например, ДНК-связывающий домен полипептида или нуклеиновая кислота, кодирующая ДНК-связывающий домен полипептида) могут быть расположены относительно других доменов или могут представлять собой отличную последовательность или происходить из отличного источника по сравнению с другими доменами или частями полипептида или кодирующей его нуклеиновой кислоты. В определенных вариантах осуществления гетерологичная молекула нуклеиновой кислоты может существовать в нативном геноме клетки-хозяина, но может характеризоваться измененным уровнем экспрессии, или иметь отличную последовательность, или и то, и другое. В других вариантах осуществления гетерологичные молекулы нуклеиновых кислот могут не являться эндогенными по отношению к клетке-хозяину или геному хозяина, но вместо этого быть введенными в клетку-хозяина с помощью трансформации (например, трансфекции, электропорации), при этом добавленная молекула может интегрироваться в геном хозяина или может существовать в виде внехромосомного генетического материала как транзиентно (например, мРНК), так и полустабильно в течение более чем одного поколения (например, эписомный вирусный вектор, плазмида или другой самореплицирующийся вектор). В некоторых вариантах осуществления домен является гетерологичным по отношению к

другому домену, если первый домен в естественных условиях не содержится в том же полипептиде, что и другой домен (например, в случае слияния двух доменов разных белков одного и того же организма).

Мутация или подвергнутый мутации. Термин "подвергнутый мутации" при использовании в отношении последовательностей нуклеиновой кислоты означает, что нуклеотиды в последовательности нуклеиновой кислоты могут быть вставлены, удалены или изменены по сравнению с эталонной (например, нативной) последовательностью нуклеиновой кислоты. В локусе может быть произведено одиночное изменение (точечная мутация), или несколько нуклеотидов в одном локусе могут быть вставлены, удалены или изменены. Кроме того, одно или несколько изменений могут быть произведены в любом количестве локусов в последовательности нуклеиновой кислоты. Последовательность нуклеиновой кислоты может быть подвергнута мутации с помощью любого способа, известного из уровня техники. В некоторых вариантах осуществления мутация имеет место в естественных условиях. В некоторых вариантах осуществления желаемая мутация может быть получена с помощью описанной в данном документе системы.

Молекула нуклеиновой кислоты. Термин "молекула нуклеиновой кислоты" относится к молекулам как РНК, так и ДНК, включающим без ограничения кДНК, геномную ДНК и мРНК, а также включает синтетические молекулы нуклеиновых кислот, такие как те, которые синтезируют химическим путем или получают рекомбинантным способом, такие как РНК-матрицы, описанные в данном документе. Молекула нуклеиновой кислоты может быть двухнитевой или однонитевой, кольцевой или линейной. В случае, если молекула нуклеиновой кислоты является однонитевой, она может представлять собой смысловую нить или антисмысловую нить. Если не указано иное, и в качестве примера для всех последовательностей, описанных в данном документе, в общем формате выражения "SEQ ID NO:", "нуклеиновая кислота, содержащая SEQ ID NO: 1" относятся к нуклеиновой кислоте, по меньшей мере часть которой имеет (i) последовательность под SEQ ID NO: 1 либо (ii) последовательность, комплементарную SEQ ID NO: 1. Выбор между двумя вариантами обусловлен контекстом, в котором используется выражение "SEQ ID NO: 1". Например, в случае, когда нуклеиновая кислота применяется в качестве зонда, выбор между двумя вариантами обусловлен требованием того, чтобы зонд был комплементарен требуемой мишени. Последовательности нуклеиновой кислоты согласно настоящему изобретению могут быть модифицированы химически или биохимически или могут содержать неприродные или дериватизированные нуклеотидные основания, что без труда будет понятно специалистам в данной области. Такие модификации включают, например, мечение, метилирование, замену одного или нескольких встречающихся в природе нуклеотидов аналогом, модификации межнуклеотидных связей, такие как незаряженные связи (например, метилфосфонатные, фосфотриэфирные, фосфорамидатные, карбаматные и т. д.), заряженные связи (например, фосфотиоатные, фосфодитиоатные и т. д.), подвешенные компоненты (например, полипептиды), интеркаляторы (например, акридин, псорален и т. д.), хелаторы, алкиляторы и модифицированные связи (например, альфа-

аномерные нуклеиновые кислоты и т. д.). Также включены синтетические молекулы, которые имитируют полинуклеотиды в их способности связываться с указанной последовательностью посредством образования водородных связей и других химических взаимодействий. Такие молекулы известны из уровня техники и включают, например, те молекулы, в которых фосфатные связи заменены пептидными связями в остове молекулы. Другие модификации могут включать, например, аналоги, в которых рибозное кольцо содержит мостиковый компонент, или другие структуры, такие как модификации, обнаруживаемые в "запертых" нуклеиновых кислотах.

Единица экспрессии гена. Единица экспрессии гена представляет собой последовательность нуклеиновой кислоты, содержащую по меньшей мере одну регуляторную последовательность нуклеиновой кислоты, функционально связанную с по меньшей мере одной эффекторной последовательностью. Первая последовательность нуклеиновой кислоты функционально связана со второй последовательностью нуклеиновой кислоты, если первая последовательность нуклеиновой кислоты помещена в функциональную взаимосвязь со второй последовательностью нуклеиновой кислоты. Например, промотор или энхансер функционально связан с кодирующей последовательностью в случае, если промотор или энхансер влияет на транскрипцию или экспрессию кодирующей последовательности. Функционально связанные последовательности ДНК могут быть смежными или несмежными. При необходимости соединения двух областей, кодирующих белок, функционально связанные последовательности могут находиться в одной и той же рамке считывания.

Хозяин. Термины "геном хозяина" или "клетка-хозяин", используемые в данном документе, относятся к клетке и/или ее геному, в которые был введен белок и/или генетический материал. Следует понимать, что такие термины предназначены для обозначения не только конкретных рассматриваемых клетки и/или генома, но и потомства такой клетки и/или генома потомства такой клетки. Поскольку в последующих поколениях могут происходить определенные модификации как в результате мутации, так и в результате влияния окружающей среды, такое потомство может фактически не являться идентичным родительской клетке, но по-прежнему быть включено в объем термина "клетка-хозяин", используемого в данном документе. Геном хозяина или клетка-хозяин могут представлять собой выделенную клетку или линию клеток, выращиваемую в культуре, или геномный материал, выделенный из такой клетки или линии клеток, или могут представлять собой клетку-хозяина или геном хозяина, которые образуют живую ткань или организм. В некоторых случаях клетка-хозяин может представлять собой клетку животного или клетку растения, например, как описано в данном документе. В определенных случаях клетка-хозяин может представлять собой клетку крупного рогатого скота, клетку лошади, клетку свиньи, клетку козы, клетку овцы, клетку курицы или клетку индейки. В определенных случаях клетка-хозяин может представлять собой клетку кукурузы, клетку сои, клетку пшеницы или клетку риса.

Псевдоузел. Термин "последовательность псевдоузла", используемый в данном документе, относится к нуклеиновой кислоте (например, РНК), имеющей последовательность с самокомплементарностью, подходящей для образования структуры псевдоузла, например, содержащей первый сегмент, второй сегмент между первым сегментом и третьим сегментом, где третий сегмент является комплементарным первому сегменту, и четвертый сегмент, где четвертый сегмент является комплементарным второму сегменту. Псевдоузел необязательно может иметь дополнительную вторичную структуру, например, структуру "стебель-петля", расположенную во втором сегменте, структуру "стебель-петля", расположенную между вторым сегментом и третьим сегментом, последовательность перед первым сегментом или последовательность после четвертого сегмента. Псевдоузел может содержать дополнительную последовательность между первым и вторым сегментами, между вторым и третьим сегментами или между третьим и четвертым сегментами. В некоторых вариантах осуществления сегменты расположены в направлении от 5' к 3' в следующем порядке: первый, второй, третий и четвертый. В некоторых вариантах осуществления первый и третий сегменты содержат пять пар оснований, характеризующихся абсолютной комплементарностью. В некоторых вариантах осуществления второй и четвертый сегменты содержат 10 пар оснований, необязательно содержащих одно или несколько (например, два) выпетливаний. В некоторых вариантах осуществления второй сегмент содержит один или несколько неспаренных нуклеотидов, например, образующих петлю. В некоторых вариантах осуществления третий сегмент содержит один или несколько неспаренных нуклеотидов, например, образующих петлю.

Последовательность структуры "стебель-петля". Используемый в данном документе термин "последовательность структуры "стебель-петля" относится к последовательности нуклеиновой кислоты (например, последовательности РНК), характеризующейся самокомплементарностью, достаточной для образования структуры "стебель-петля", например, имеющей "стебель", содержащий по меньшей мере две (например, 3, 4, 5, 6, 7, 8, 9 или 10) пары оснований, и "петлю", содержащую по меньшей мере три (например, четыре) пары оснований. "Стебель" может содержать ошибки спаривания или выпетливания.

КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

Файл патента или заявки на патент содержит по меньшей мере одно графическое изображение, выполненное в цвете. Копии данного патента или публикации заявки на патент с цветным(цветными) графическим(графическими) изображением(изображениями) будут предоставлены Ведомством по запросу и при уплате соответствующей пошлины.

ФИГ. 1 представляет собой схематическое изображение системы редактирования генома Gene Writing™.

ФИГ. 2 представляет собой схематическое изображение структуры полипептида редактора генома Gene Writer™.

ФИГ. 3 представляет собой схематическое изображение структуры иллюстративных РНК-матриц Gene Writer™.

ФИГ. 4A и 4B представляют собой серию схем, на которых показаны примеры конфигураций Gene Writer с использованием доменов, полученных из различных источников. Gene Writer, описанные в данном документе, могут содержать или не содержать все изображенные домены. Например, в некоторых случаях Gene Writer может не иметь РНК-связывающего домена или может иметь отдельные домены, которые выполняют функции нескольких доменов, например, домен Cas9 для ДНК-связывающей и эндонуклеазной активности. Иллюстративные домены, которые могут быть включены в полипептид Gene Writer, включают ДНК-связывающие домены (например, включающие ДНК-связывающий домен элемента последовательности, указанной в любой из таблиц X, Y, Z1, Z2, 3A или 3B; "цинковый палец"; домен TAL; Cas9; dCas9; никазу Cas9; фактор транскрипции или мегануклеазу), РНК-связывающие домены (например, включающие РНК-связывающий домен белка B-бокса, белка оболочки MS2, dCas или элемент последовательности, указанной в любой из таблиц X, Y, Z1, Z2, 3A или 3B), домены обратной транскриптазы (например, включающие домен обратной транскриптазы элемента последовательности, указанной в любой из таблиц X, Y, 3A или 3B; другие ретротранспозазы (например, как указано в таблице Z1); пептид, содержащий домен обратной транскриптазы (например, как указано в таблице Z2)) и/или эндонуклеазный домен (например, включающий эндонуклеазный домен элемента последовательности, указанной в любой из таблиц X, Y, 3A или 3B; Cas9; никазу Cas9; фермент рестрикции (например, фермент рестрикции II типа, например, FokI), мегануклеазу, резольвазу для структур Холлидея, ретротранспозазу RLE; ретротранспозазу APE или ретротранспозазу GIY-YIG). Иллюстративные полипептиды GeneWriter, содержащие иллюстративные комбинации таких доменов, показаны на нижней панели.

ФИГ. 5 представляет собой схему, на которой показаны модули иллюстративной РНК-матрицы Gene Writer. Отдельные модули иллюстративной матрицы можно комбинировать, переупорядочивать и/или исключать, например, для получения матрицы системы Gene Writer. A=5'-плечо гомологии; B=рибозим; C=5'-UTR; D=последовательность гетерологичного объекта; E=3'-UTR; F=3'-плечо гомологии.

ФИГ. 6 представляет собой таблицу, в которой перечислены модули иллюстративной РНК-матрицы Gene Writer. Отдельные модули можно комбинировать, переупорядочивать и/или исключать, например, для получения матрицы системы Gene Writer. A=5'-плечо гомологии; B=рибозим; C=5'-UTR; D=последовательность гетерологичного объекта; E=3'-UTR; F=3'-плечо гомологии.

ФИГ. 7A и 7B представляют собой схемы, на которых показан иллюстративный процесс осуществления однонитевого разрыва второй нити. (A) Никаза Cas9 слита с белком Gene Writer. Белок Gene Writer вводит однонитевой разрыв в нить ДНК посредством своего EN-домена (обозначается как *), и слитая никаза Cas9 вводит однонитевые разрывы либо в верхнюю, либо в нижнюю нити ДНК (обозначается как X). (B) Gene Writer нацеливается на ДНК посредством своего ДНК-связывающего домена и вводит однонитевой разрыв в ДНК с помощью своего EN-домена (*). Затем никазу Cas9 применяют для создания второго

однонитевого разрыва (X) на верхней либо на нижней нитях выше или ниже однонитевого разрыва, введенного с помощью EN.

ФИГ. 8 представляет разработанную конструкцию для скрининга опосредованной ретротранспозонами интеграции в клетки человека. Плазмиду с управляющим элементом, содержащую кассету экспрессии ретротранспозазы (управляющий элемент), вводили путем трансфекции вместе с плазмидой с матрицей, содержащей зависимую от ретротранспозона репортерную кассету. При том, что экспрессия из плазмиды с матрицей приводит к образованию нефункционального GFP вследствие прерывания антисмыслового интрона, транскрипция молекулы-матрицы из плазмиды с матрицей приводит к образованию РНК с интроном, удаляемым путем сплайсинга, которая затем может быть подвергнута обратной транскрипции и интегрирована системой. Таким образом, экспрессия репортерной кассеты будет иметь место только из интегрированной репортерной кассеты (интегрированная gDNA, внизу), а не из плазмиды с матрицей. HA=плечо гомологии там, где это применимо; CMV=промотор CMV млекопитающих; HiBit=метка HiBit для количественного определения экспрессии белка; T7=промотор T7 РНК-полимеразы; UTR=нетранслируемая последовательность, например, нативные UTR ретротранспозона; pA=поли(A)-сигнал; SD-SA используется для обозначения донорных сайтов сплайсинга и акцепторных сайтов сплайсинга антисмыслового интрона в последовательности, кодирующей GFP. HA=плечо гомологии там, где это применимо (например, см. пример 6); CMV=промотор CMV млекопитающих; HiBit=метка HiBit для количественного определения экспрессии белка; T7=промотор T7 РНК-полимеразы; UTR=нетранслируемая последовательность, например, нативные UTR ретротранспозона; pA=поли(A)-сигнал; SD-SA используется для обозначения донорных сайтов сплайсинга и акцепторных сайтов сплайсинга антисмыслового интрона в последовательности, кодирующей GFP

На ФИГ. 9 раскрыт скрининг ретротранспозонов-кандидатов и идентифицированы 25 кандидатов, которые интегрируют полезную нагрузку в транс-положении в клетки человека. В общей сложности 163 системы ретротранспозонов анализировали на активность в клетках человека, как описано в примере 4. Интеграция, измеренная с помощью ddPCR, показана как копии/геном для каждой системы управляющий элемент/матрица ретротранспозона. Высота каждого столбца указывает на среднее значение для двух повторностей. После дополнительной оптимизации в примерах 4 и 5 конструкции с более высокой активностью дополнительно выделены на фигуре 10.

ФИГ. 10. Конфигурации Gene Writing с высокой активностью, основанные на хитах ретротранспозонов из таблицы 3B и дополнительно улучшенные в примерах 7, 8 и 9. В случае тестирования нескольких конфигураций данной системы, например, с альтернативной последовательностью, кодирующей ретротранспозазу (пример 8), или добавлением плеч гомологии (пример 9), показана только наиболее эффективная конфигурация. Для систем, улучшенных по сравнению с исходной конфигурацией, указанной в таблице 3B и оцененной в примере 7, улучшения, описанные в примере 5 (фигура 11) и примере 6 (фигура 12), подробно описаны в таблице 11.

ФИГ. 11. Консенсусная последовательность ретротранспозона может восстановить или улучшить активность интеграции в транс-положении в клетках человека. Эффективность интеграции, измеренная посредством показателя копии/геном с помощью ddPCR, как описано в примере 5, показана для каждой системы управляющий элемент/матрица ретротранспозона. Высота каждого столбца указывает на среднее значение для двух повторностей. Незакрашенные кружки обозначают каждую повторность, а столбцы представляют исходные последовательности (светло-серые) или новую последовательность белка, полученную с помощью консенсусного подхода (темно-серые).

На ФИГ. 12 показано, что последовательности плеч гомологии, созданные на основе консенсусного мотива, могут восстанавливать или улучшать активность интеграции в транс-положении ретротранспозона в клетках человека. Эффективность интеграции, измеренная посредством показателя копии/геном с помощью ddPCR (пример 6), показана для каждой системы управляющий элемент/матрица Gene Writer. Высота каждого столбца указывает на среднее значение для двух повторностей. Закрашенные кружки и белые столбцы представляют последовательность матрицы без плеч гомологии, а незакрашенные кружки и заштрихованные столбцы обозначают конструкции, содержащие плечи гомологии.

На ФИГ. 13 A и B показан анализ люциферазной активности для первичных клеток. LNP, составленные в соответствии с примером 11, анализировали в отношении доставки груза в первичные гепатоциты человека (A) и мыши (B) в соответствии с примером 12. В результате анализа люциферазы была выявлена дозозависимая люциферазная активность в лизатах клеток, что указывает на успешную доставку РНК в клетки и экспрессию люциферазы светлячка из мРНК-груза.

На ФИГ. 14 раскрыта LNP-опосредованная доставка РНК-груза в печень мыши. LNP, содержащие мРНК люциферазы светлячка, составляли и доставляли мышам iv, а образцы печени собирали и анализировали в отношении люциферазной активности через 6, 24 и 48 часов после введения. Активность репортерного гена в различных составах соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003. Экспрессия РНК была транзиентной, и уровни ферментов возвращались к уровню, близкому к фоновому уровню при обработке средой-носителем, через 48 часов после введения.

ПОДРОБНОЕ ОПИСАНИЕ

Настоящее изобретение относится к композициям, системам и способам для нацеливания на последовательность ДНК, ее редактирования, модификации или произведения манипуляций с ней (например, вставки гетерологичного объекта в виде последовательности ДНК в сайт-мишень генома млекопитающего) в одном или нескольких местоположениях в последовательности ДНК в клетке, ткани или у субъекта, например, *in vivo* или *in vitro*. Объект в виде последовательности ДНК может включать в себя, например, кодирующую последовательность, регуляторную последовательность, единицу экспрессии гена.

Более конкретно, в настоящем изобретении предусмотрены системы на основе ретротранспозона для вставки последовательности, представляющей интерес, в геном. Настоящее изобретение частично основывается на биоинформатическом анализе для идентификации последовательностей ретротранспозазы и ассоциированных с ней 5'-UTR и 3'-UTR из разнообразных организмов (см. таблицы 3B, 10, 11 и X). Дополнительные примеры элементов ретротранспозона указаны, например, в таблицах 1 и 2 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления системы, описанные в данном документе, могут обладать рядом преимуществ по сравнению с разнообразными более ранними системами. Например, в настоящем изобретении описаны ретротранспозазы, способные к вставке длинных последовательностей гетерологичной нуклеиновой кислоты в геном. Кроме того, ретротранспозазы, описанные в данном документе, могут осуществлять вставку гетерологичной нуклеиновой кислоты в эндогенный сайт в геноме, такой как локус рДНК. В этом заключается отличие от систем Cre/loxP, для которых требуется первая стадия вставки экзогенного сайта loxP перед второй стадией вставки последовательности, представляющей интерес, в сайт loxP.

Редакторы генома Gene Writer™

Не содержащие длинных концевых повторов (LTR) ретротранспозоны представляют собой тип мобильных генетических элементов, широко распространенных в геномах эукариот. Они включают, например, ретротранспозоны типа апуриновой/апиримидиновой эндонуклеазы (APE), ретротранспозоны типа эндонуклеазы, подобной ферментам рестрикции (RLE), и ретротранспозоны типа Penelope-подобного элемента (PLE). Ретротранспозоны класса APE состоят из двух функциональных доменов: эндонуклеазного/ДНК-связывающего домена и домена обратной транскриптазы. Примеры ретротранспозонов класса APE можно найти, например, в таблице 1 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте, включая перечень последовательностей и последовательности, упомянутые в таблице 1 в ней. Ретротранспозоны класса RLE состоят из трех функциональных доменов: ДНК-связывающего домена, домена обратной транскрипции и эндонуклеазного домена. Примеры ретротранспозонов класса RLE можно найти, например, в таблице 2 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте, включая перечень последовательностей и последовательности, упомянутые в таблице 2 в ней. Домен обратной транскриптазы не содержащего LTR ретротранспозона функционирует путем связывания с последовательностью РНК-матрицы и осуществления ее обратной транскрипции в ДНК-мишень генома хозяина. Последовательность РНК-матрицы содержит 3'-нетранслируемую область, которая специфично связывается с ретротранспозазой, и вариабельную 5'-область, как правило, имеющую открытую(открытые) рамку(рамки) считывания ("ORF"), кодирующую(кодирующие) белки ретротранспозазы. Последовательность РНК-матрицы

может также содержать 5'-нетранслируемую область, которая специфично связывается с ретротранспозазой. Penelope-подобные элементы (PLE) отличаются как от содержащих LTR ретротранспозонов, так и от не содержащих LTR ретротранспозонов. PLE обычно содержат домен обратной транскриптазы, отличный от домена элементов APE и RLE, но сходный с таковым у теломераз и интронов группы II, и необязательный эндонуклеазный домен GIY-YIG.

Другие иллюстративные классы ретротранспозонов включают без ограничения ретротранспозоны CRE, NeSL, R4, R2, Hero, L1, RTE (например, AviRTE_Brh или BovB), I, Jockey, CR1, Rex1, RandI/Dualen, Penelope или Penelope-подобные (PLE) (например, Penelope_SM), Tx1, RTEX, Crack, Nimb, Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi (например, Vingi-1_EE) и Kiri.

Как описано в данном документе, элементы таких ретротранспозонов могут быть разбиты на функциональные модули и/или модифицированы для нацеливания на последовательность ДНК-мишени, ее редактирования, модификации или произведения манипуляций с ней, например, для вставки объекта в виде (например, гетерологичной) последовательности нуклеиновой кислоты в геном-мишень, например, геном млекопитающего, посредством обратной транскрипции. Такие разбитые на модули и модифицированные нуклеиновые кислоты, композиции на основе полипептидов и системы описаны в данном документе и называются редакторами генов Gene Writer™. Система редактирования генов Gene Writer™ содержит: (A) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и либо (x) эндонуклеазный домен, который обладает функциональным свойством связывания с ДНК, либо (y) эндонуклеазный домен и отдельный ДНК-связывающий домен; и (B) РНК-матрицу, содержащую (i) последовательность, которая связывает полипептид, и (ii) гетерологичную вставочную последовательность. Например, белок редактора генома Gene Writer может содержать ДНК-связывающий домен, домен обратной транскриптазы и эндонуклеазный домен. В других вариантах осуществления белок редактора генома Gene Writer может содержать домен обратной транскриптазы и эндонуклеазный домен. В определенных вариантах осуществления элементы полипептида редактора генов Gene Writer™ могут быть получены из последовательностей ретротранспозонов, например, ретротранспозонов APE-типа, RLE-типа или PLE-типа или их частей или доменов. В некоторых вариантах осуществления не содержащий LTR ретротранспозон RLE-типа относится к кладе R2, NeSL, HERO, R4 или CRE. В некоторых вариантах осуществления редактор генома Gene Writer получен из элемента R4 X4_Line, который обнаруживается в человеческом геноме. В некоторых вариантах осуществления не содержащий LTR ретротранспозон APE-типа относится к кладе R1 или Tx1. В некоторых вариантах осуществления редактор генома Gene Writer получен из элемента Tx1 Mare6, который обнаруживается в человеческом геноме. Элемент системы редактирования генов Gene Writer™, представляющий собой РНК-матрицу, как правило, является гетерологичным по отношению к элементу, представляющему собой полипептид, и предоставляет

последовательность объекта для вставки (обратной транскрипции) в геном хозяина. В некоторых вариантах осуществления белок редактора генома Gene Writer способен к осуществлению обратной транскрипции с праймированием мишенью.

В некоторых вариантах осуществления редактор генома Gene Writer объединен со вторым полипептидом. В некоторых вариантах осуществления второй полипептид получен из не содержащего LTR ретротранспозона APE-типа. В некоторых вариантах осуществления второй полипептид содержит мотив, подобный "цинковому суставу". В некоторых вариантах осуществления второй полипептид является гомологом белков Gag.

В некоторых вариантах осуществления редактор генома Gene Writer содержит последовательность ретротранспозазы элемента, указанного в таблице X. В таблице X представлены серии последовательностей нуклеиновых кислот и аминокислот (указанные по названию гена в Repbase и видам) с соответствующими номерами доступа в GenBank, если они доступны. Нуклеиновая кислота и аминокислотные последовательности элементов согласно Repbase, указанные в таблице X, включены в данный документ посредством ссылки во всей своей полноте. Последовательности элементов согласно GenBank, указанные в таблице X, также включены в данный документ посредством ссылки во всей своей полноте. Последовательности нуклеиновых кислот из таблицы X включают в некоторых случаях последовательность, кодирующую полипептид (например, открытую рамку считывания) (например, кодирующую белок последовательность из соответствующей записи в Repbase, включенную в данный документ посредством ссылки). Последовательности нуклеиновых кислот из таблицы X включают в некоторых случаях последовательность 5'-UTR (например, последовательность 5'-UTR из соответствующей записи в Repbase, включенную в данный документ посредством ссылки). Последовательности нуклеиновых кислот из таблицы X включают в некоторых случаях последовательность 3'-UTR (например, последовательность 3'-UTR из соответствующей записи в Repbase, включенную в данный документ посредством ссылки).

В некоторых вариантах осуществления открытая рамка считывания (ORF) для аминокислотной последовательности аннотирована в Repbase. В некоторых вариантах осуществления ORF для аминокислотной последовательности не аннотирована в Repbase. Если ORF не аннотирована, ее может идентифицировать специалист в данной области, например, посредством выполнения одной или нескольких операций трансляций (например, трансляции всей рамки) и необязательно сравнения результата трансляции с последовательностью обратной транскриптазы или консенсусным мотивом. В некоторых вариантах осуществления аминокислотная последовательность из таблицы X, например, используемая в данном документе, представляет собой аминокислотную последовательность, указанную в соответствующей записи Repbase, или аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты в соответствующей записи Repbase. В некоторых вариантах осуществления аминокислотная последовательность, кодируемая элементом из таблицы X, представляет собой аминокислотную последовательность, кодируемую полноразмерной последовательностью

элемента, указанного в таблице X, или последовательностью, характеризующейся по меньшей мере 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней. В некоторых вариантах осуществления полноразмерная последовательность элемента, указанного в таблице X, может содержать одну или несколько (например, все) из 5'-UTR, последовательности, кодирующей полипептид, или 3'-UTR ретротранспозона, как описано в данном документе. В некоторых вариантах осуществления аминокислотная последовательность из таблицы X представляет собой аминокислотную последовательность, кодируемую полноразмерной последовательностью элемента, указанного в таблице X, или последовательностью, характеризующейся по меньшей мере 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней. В некоторых вариантах осуществления 5'-UTR элемента из таблицы X содержит 5'-UTR полноразмерной последовательности элемента, указанного в таблице X, или последовательности, характеризующейся по меньшей мере 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней. В некоторых вариантах осуществления 3'-UTR элемента из таблицы X содержит 3'-UTR полноразмерной последовательности элемента, указанного в таблице X, или последовательности, характеризующейся по меньшей мере 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней.

В таблице X также указаны организмы-хозяева, из которых были получены последовательности нуклеиновых кислот, и список доменов, присутствующих в полипептиде, кодируемом открытой рамкой считывания последовательности нуклеиновой кислоты. Домены, перечисленные в таблице X, указаны как идентификаторы доменов, которые соответствуют записям доменов в InterPro, перечисленным в таблице Y. Таким образом, последовательность из Repbase, указанная в таблице X, может кодировать полипептид, содержащий один или несколько доменов, указанных в таблице X по их идентификаторам доменов. Конкретный домен или тип домена, связанный с каждым идентификатором домена, более подробно описан в таблице Y. В некоторых вариантах осуществления представляющий интерес домен (например, как указано в таблице Y) может быть идентифицирован в последовательности нуклеиновой кислоты (например, как указано в таблице X) посредством выполнения операций трансляции всей рамки и идентификации аминокислотных последовательностей, кодирующих представляющий интерес домен.

Инструменты обнаружения ретротранспозонов

В результате повторной мобилизации с течением времени подвижные элементы в геномной ДНК часто существуют в виде тандемных или диспергированных повторов (Jurka Curr Opin Struct Biol 8, 333-337 (1998)). Инструменты, способные распознавать такие повторы, можно использовать для идентификации новых элементов из геномной ДНК и для наполнения баз данных, например, Repbase (Jurka et al Cytogenet Genome Res 110, 462-467 (2005)). Одним из таких инструментов для идентификации повторов, которые могут содержать подвижные элементы, является RepeatFinder (Volfovsky et al. Genome Biol 2 (2001)), который анализирует повторяющуюся структуру геномных последовательностей.

Повторы могут быть дополнительно собраны и проанализированы с использованием дополнительных инструментов, например, Censor (Kohany et al. BMC Bioinformatics 7, 474 (2006)). Пакет Censor берет геномные повторы и аннотирует их, используя различные подходы BLAST, по отношению к известным подвижным элементам. Трансляция всей рамки может быть использована для создания ORF для сравнения.

Другие иллюстартивные способы идентификации подвижных элементов включают RepeatModeler2, который автоматизирует обнаружение и аннотирование подвижных элементов в последовательностях генома (Flynn et al bioRxiv (2019)). В дополнение к выполнению этого с помощью доступных пакетов, таких как Censor, можно выполнять трансляцию всей рамки данных генома или последовательности и проводить аннотацию с помощью инструмента для белковых доменов, такого как InterProScan, который помечает домены данной аминокислотной последовательности с использованием базы данных InterPro (Mitchell et al. Nucleic Acids Res 47, D351-360 (2019)), что обеспечивает возможность идентификации потенциальных белков, содержащих домены, связанные с известными подвижными элементами (например, домены элементов, указанных в таблице X).

Ретротранспозоны можно дополнительно классифицировать в соответствии с доменом обратной транскриптазы с использованием такого инструмента, как RTclass1 (Kapitonov et al Gene 448, 207-213 (2009)).

Полипептидный компонент системы редактирования генов Gene Writer

Домен RT

В определенных аспектах настоящего изобретения в основе домена обратной транскриптазы системы Gene Writer лежит домен обратной транскриптазы не содержащего LTR ретротранспозона APE-типа или RLE-типа или ретротранспозона PLE-типа. Домен обратной транскриптазы дикого типа ретротранспозона APE-типа, RLE-типа или PLE-типа может использоваться в системе Gene Writer или может быть модифицирован (например, посредством вставки, делеции или замены одного или нескольких остатков) с изменением активности обратной транскриптазы в отношении последовательностей ДНК-мишеней. В некоторых вариантах осуществления обратная транскриптаза изменена по сравнению с ее природной последовательностью таким образом, что она характеризуется измененной частотой использования кодонов, например, улучшенной для применения в клетках человека. В некоторых вариантах осуществления домен обратной транскриптазы представляет собой гетерологичную обратную транскриптазу из другого ретровируса, ретрона, ретроэлемента, генерирующего разнообразие, ретроплазмиды, интрона группы II, содержащего LTR ретротранспозона, не содержащего LTR ретротранспозона или другого источника, например, как показано в таблице Z1, или содержит домен, указанный в таблице Z2. В некоторых вариантах осуществления система Gene Writer содержит полипептид, который содержит домен обратной транскриптазы не содержащего LTR ретротранспозона RLE-типа из клады R2, NeSL, HERO, R4 или CRE, не содержащего LTR ретротранспозона APE-типа из клады L1, RTE, I, Jockey, CR1, Rex1, RandI/Dualen, T1, RTEX, Crack, Nimb,

Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi или Kiri или ретротранспозона PLE-типа. В определенных вариантах осуществления система Gene Writer содержит полипептид, который содержит домен обратной транскриптазы из ретротранспозона, указанного в таблице 10, таблице 11, таблице X, таблице Z1, таблице Z2 или таблице 3A или 3B. В вариантах осуществления аминокислотная последовательность домена обратной транскриптазы системы Gene Writer является на по меньшей мере приблизительно 50%, на по меньшей мере приблизительно 60%, на по меньшей мере приблизительно 70%, на по меньшей мере приблизительно 80%, на по меньшей мере приблизительно 85%, на по меньшей мере приблизительно 90%, на по меньшей мере приблизительно 95%, на по меньшей мере приблизительно 96%, на по меньшей мере приблизительно 97%, на по меньшей мере приблизительно 98%, на по меньшей мере приблизительно 99% идентичной аминокислотной последовательности домена обратной транскриптазы ретротранспозона, последовательность ДНК которого указана в таблице 10, таблице 11, таблице X, таблице Z1, таблице Z2 или таблице 3A или 3B. Домены обратной транскрипции могут быть идентифицированы, например, на основании гомологии по отношению к другим известным доменам обратной транскрипции с помощью стандартных инструментов, таких как средство поиска основного локального выравнивания (BLAST). В некоторых вариантах осуществления домены обратной транскриптазы являются модифицированными, например, посредством сайт-специфической мутации. В некоторых вариантах осуществления домен обратной транскриптазы сконструирован таким образом, что он связывает гетерологичную РНК-матрицу.

В некоторых вариантах осуществления полипептид (например, домен RT) содержит РНК-связывающий домен, например, который специфично связывается с последовательностью РНК. В некоторых вариантах осуществления РНК-матрица содержит последовательность РНК, которая специфично связана с РНК-связывающим доменом.

В некоторых вариантах осуществления домен RT демонстрирует повышенную строгость инициации обратной транскрипции с праймированием мишенью (TPRT), например, по сравнению с эндогенным доменом RT. В некоторых вариантах осуществления домен RT инициирует TPRT, когда 3 нуклеотида в сайте-мишени непосредственно перед однонитевым разрывом первой нити, например, геномная ДНК, праймирующая РНК-матрицу, характеризуются по меньшей мере 66% или 100% комплементарностью по отношению к 3 гомологичным нуклеотидам в РНК-матрице. В некоторых вариантах осуществления домен RT инициирует TPRT при наличии несовпадения менее 5 нуклеотидов (например, несовпадения менее 1, 2, 3, 4 или 5 нуклеотидов) между гомологичной РНК-матрицей и ДНК-мишенью, праймирующей обратную транскрипцию. В некоторых вариантах осуществления домен RT модифицирован таким образом, что строгость в отношении несовпадений при праймировании реакции TPRT увеличивается, например, где домен RT не допускает каких-либо несовпадений или допускает меньшее количество несовпадений в области праймирования по сравнению с доменом RT дикого типа (например, немодифицированным). В некоторых вариантах осуществления домен RT

содержит домен RT HIV-1. В вариантах осуществления домен RT HIV-1 инициирует более низкие уровни синтеза даже при несовпадении трех нуклеотидов по сравнению с альтернативным доменом RT (например, как описано в Jamburuthugoda и Eickbush J Mol Biol 407(5):661-672 (2011); включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен RT образует димер (например, гетеродимер или гомодимер). В некоторых вариантах осуществления домен RT является мономерным. В некоторых вариантах осуществления домен RT, например, домен RT ретровируса, в естественных условиях функционирует как мономер или как димер (например, гетеродимер или гомодимер). В некоторых вариантах осуществления домен RT в естественных условиях функционирует как мономер, например, получен из вируса, в котором он функционирует как мономер. Иллюстративные мономерные домены RT, их вирусные источники и ассоциированные с ними сигнатуры RT можно найти в таблице 30, при этом описание сигнатур доменов находится в таблице 32. В некоторых вариантах осуществления домен RT системы, описанной в данном документе, содержит аминокислотную последовательность из таблицы 30 или ее функциональный фрагмент или вариант или последовательность, характеризующуюся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ней. В вариантах осуществления домен RT выбран из домена RT вируса лейкоза мышей (MLV; иногда называемого как MoMLV) (например, P03355), эндогенного ретровируса свиньи (PERV) (например, UniProt Q4VFZ2), вируса опухоли молочной железы мышей (MMTV) (например, UniProt P03365), вируса обезьян Мейзона-Пфайзера (MPMV) (например, UniProt P07572), вируса лейкоза крупного рогатого скота (BLV) (например, UniProt P03361), вируса Т-клеточного лейкоза человека 1 типа (HTLV-1) (например, UniProt P03362), вируса пенистости человека (HFV) (например, UniProt P14350), вируса пенистости обезьян (SFV) (например, UniProt P23074) или вируса пенистости/синцитиального вируса крупного рогатого скота (BFV/BSV) (например, UniProt O41894) или его функционального фрагмента или варианта (например, аминокислотной последовательности, характеризующейся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ним). В некоторых вариантах осуществления домен RT является димерным в своем естественном функционировании. Иллюстративные димерные домены RT, их вирусные источники и ассоциированные с ними сигнатуры RT можно найти в таблице 31, при этом описание сигнатур доменов находится в таблице 32. В некоторых вариантах осуществления домен RT системы, описанной в данном документе, содержит аминокислотную последовательность из таблицы 31 или ее функциональный фрагмент или вариант или последовательность, характеризующуюся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ней. В некоторых вариантах осуществления домен RT получен из вируса, в котором он функционирует как димер. В вариантах осуществления домен RT выбран из домена RT вируса саркомы/лейкоза птиц (ASLV) (например, UniProt A0A142BKH1), вируса саркомы Рауса (RSV) (например, UniProt P03354), вируса миелобластоза птиц (AMV) (например, UniProt Q83133), вируса иммунодефицита человека

I типа (HIV-1) (например, UniProt P03369), вируса иммунодефицита человека II типа (HIV-2) (например, UniProt P15833), вируса иммунодефицита обезьян (SIV) (например, UniProt P05896), вируса иммунодефицита крупного рогатого скота (BIV) (например, UniProt P19560), вируса инфекционной анемии лошадей (EIAV) (например, UniProt P03371) или вируса иммунодефицита кошек (FIV) (например, UniProt P16088) (Herschhorn and Hizi Cell Mol Life Sci 67(16):2717-2747 (2010)) или его функционального фрагмента или варианта (например, аминокислотной последовательности, характеризующейся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ним). Природные гетеродимерные домены RT могут в некоторых вариантах осуществления также функционировать как гомодимеры. В некоторых вариантах осуществления димерные домены RT экспрессируются в виде слитых белков, например, в виде гомодимерных слитых белков или гетеродимерных слитых белков. В некоторых вариантах осуществления функцию RT в системе выполняют несколько доменов RT (например, как описано в данном документе). В дополнительных вариантах осуществления несколько доменов RT слиты или разделены, например, могут находиться в одном и том же полипептиде или в разных полипептидах.

В некоторых вариантах осуществления Gene Writer, описанный в данном документе, содержит интегразный домен, например, где интегразный домен может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит интегразный домен. В некоторых вариантах осуществления в домене RT (например, описанном в данном документе) отсутствует интегразный домен или он содержит интегразный домен, который был инактивирован путем мутации или подвергнут делеции. В некоторых вариантах осуществления Gene Writer, описанный в данном документе, содержит домен РНКазы H, например, где домен РНКазы H может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы H, например, эндогенный домен РНКазы H или гетерологичный домен РНКазы H. В некоторых вариантах осуществления в домене RT (например, описанном в данном документе) отсутствует домен РНКазы H. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы H, который был добавлен, подвергнут делеции, мутации или заменен гетерологичным доменом РНКазы H. В некоторых вариантах осуществления в результате мутации домена РНКазы H образуется полипептид, демонстрирующий более низкую активность РНКазы, например, как определено способами, описанными в Kotewicz et al. Nucleic Acids Res 16(1):265-277 (1988) (включенном в данный документ посредством ссылки во всей своей полноте), например, ниже на по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% или 90% по сравнению с в остальном аналогичным доменом без мутации. В некоторых вариантах осуществления активность РНКазы H устранена.

В некоторых вариантах осуществления домен RT подвергнут мутации для повышения точности по сравнению с в остальном аналогичным доменом без мутации. Например, в некоторых вариантах осуществления мотив YADD или YMDD в домене RT (например, в обратной транскриптазе) заменен на YVDD. В вариантах осуществления

замена YADD, или YMDD, или YVDD приводит к более точной активности ретровирусной обратной транскриптазы (например, как описано в Jamburuthugoda and Eickbush J Mol Biol 2011; включенном в данный документ посредством ссылки во всей своей полноте).

Разнообразие обратных транскриптаз (например, содержащих домены RT) было описано без ограничения теми, которые используются прокариотами (Zimmerly et al. *Microbiol Spectr* 3(2):MDNA3-0058-2014 (2015); Lampson B.C. (2007) Prokaryotic Reverse Transcriptases. B: Polaina J., MacCabe A.P. (eds) Industrial Enzymes. Springer, Dordrecht), вирусами (Herschhorn et al. *Cell Mol Life Sci* 67(16):2717-2747 (2010); Menéndez-Arias et al. *Virus Res* 234:153-176 (2017)), и мобильными элементами (Eickbush et al. *Virus Res* 134(1-2):221-234 (2008); Craig et al. *Mobile DNA III* 3rd Ed. DOI:10.1128/9781555819217 (2015)), каждый из которых включен в данный документ посредством ссылки.

Таблица 30. Иллюстративные мономерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	**Номер доступа**	**Организм**	**Последовательность**	**Сигнатуры RT**
Q4VFZ2_9GAMR - только остатки	Q4VFZ2	Эндогенный ретровирус свиней	MGATGQQQYPWTTRRTVDLGVG RVTHSFLVIPECPAPLLGRDLLTK MGAQISFEQGKPEVSANNKPITVL TLQLDDEYRLYSPLVKPDQNIQF WLEQFPQAWAETAGMGLAKQVP PQVIQLKASATPVSVRQYPLSKEA QEGIRPHVQRLIQQGILVPVQSPW NTPLLPVRKPGTNDYRPVQDLRE VNKRVQDIHPTVPNPYNLLCALPP QRSWYTVLDLKDAFFCLRLHPTS QPLFAFEWRDPGTGRTGQLTWTR LPQGFKNSPTIFDEALHRDLANFRI QHPQVTLLQYVDDLLLAGATKQ DCLEGTKALLLELSDLGYRASAK KAQICRREVTYLGYSLRDGQRWL TEARKKTVVQIPAPTTAKQVREFL GTAGFCRLWIPGFATLAAPLYPLT KEKGEFSWAPEHQKAFDAIKKAL LSAPALALPDVTKPFTLYVDERK GVARGVLTQTLGPWRRPVAYLSK KLDPVASGWPVCLKAIAAVAILV	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			KDADKLTLGQNITVIAPHALENIV RQPPDRWMTNARMTHYQSLLLT ERVTFAPPAALNPATLLPEETDEP VTHDCHQLLIEETGVRKDLTDIPL TGEVLTWFTDGSSYVVEGKRMA GAAVVDGTRTIWASSLPEGTSAQ KAELMALTQALRLAEGKSINIYTD SRYAFATAHVHGAIYKQRGLLTS AGREIKNKEEILSLLEALHLPKRL AIIHCPGHQKAKDPISRGNQMAD RVAKQAAQGVNLLPMIETPKAPE PGRQYTLEDWQEIKKIDQFSETPE GTCYTSDGKEILPHKEGLEYVQQI HRLTHLGTKHLQQLVRTSPYHVL RLPGVADSVVKHCVPCQLVNANP SRIPPGKRLRGSHPGAHWEVDFTE VKPAKYGNKYLLVFVDTFSGWV EAYPTKKETSTVVAKKILEEIFPRF GIPKVIGSDNGPAFVAQVSQGLAK ILGIDWKLHCAYRPQSSGQVERM NRTIKETLTKLTAETGVNDWIALL PFVLFRVRNTPGQFGLTPYELLYG GPPPLVEIASVHSADVLLSQPLFSR LKALEWVRQRAWRQLREAYSGG GDLQIPHRFQVGDSVYVRRHRAG NLETRWKGPYHVLLTTPTAVKVE GISTWIHASHVKPAPPPDSGWKAE KTENPLKLRLHRVVPYSVNNFSS	

POL_SFV1 - только остатки	P23074	Вирус пенистости обезьян 1 типа	MDPLQLLQPLEAEIKGTKLKAHW DSGATITCVPEAFLEDERPIQTMLI KTIHGEKQQDVYYLTFKVQGRKV EAEVLASPYDYILLNPSDVPWLM KKPLQLTVLVPLHEYQERLLQQT ALPKEQKELLQKLFLKYDALWQH WENQVGHRRIKPHNIATGTLAPR PQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGK WRMVLDYREVNKTIPLIAAQNQH SAGILSSIYRGKYKTTLDLTNGFW AHPITPESYWLTAFTWQGKQYCW TRLPQGFLNSPALFTADVVDLLKE IPNVQAYVDDIYISHDDPQEHLEQ LEKIFSILLNAGYVVSLKKSEIAQR EVEFLGFNITKEGRGLTDTFKQKL LNITPPKDLKQLQSILGLLNFARNF IPNYSELVKPLYTIVANANGKFIS WTEDNSNQLQHIISVLNQADNLE ERNPETRLIIKVNSSPSAGYIRYYN EGSKRPIMYVNYIFSKAEAKFTQT EKLLTTMHKGLIKAMDLAMGQEI LVYSPIVSMTKIQRTPLPERKALP VRWITWMTYLEDPRIQFHYDKSL PELQQIPNVTEDVIAKTKHPSEFA MVFYTDGSAIKHPDVNKSHSAGM GIAQVQFIPEYKIVHQWSIPLGDH TAQLAEIAAVEFACKKALKISGPV LIVTDSFYVAESANKELPYWKSN GFLNNKKKPLRHVSKWKSIAECL QLKPDIIIMHEKGHQQPMTTLHTE GNNLADKLATQGSYVVHCNTTPS LDAELDQLLQGHYPPGYPKQYKY TLEENKLIVERPNGIRIVPPKADRE KIISTAHNIAHTGRDATFLKVSSK	IPR043502, SSF56672, IPR000477, PF00078

			YWWPNLRKDVVKSIRQCKQCLV TNATNLTSPPILRPVKPLKPFDKFY IDYIGPLPPSNGYLHVLVVVDSMT GFVWLYPTKAPSTSATVKALNML TSIAIPKVLHSDQGAAFTSSTFAD WAKEKGIQLEFSTPYHPQSSGKVE RKNSDIKRLLTKLLIGRPAKWYDL LPVVQLALNNSYSPSSKYTPHQLL FGVDSNTPFANSDTLDLSREEELS LLQEIRSSLHQPTSPPASSRSWSPS VGQLVQERVARPASLRPRWHKPT AILEVVNPRTVIILDHLGNRRTVS VDNLKLTAYQDNGTSNDSGTMA LMEEDESSTSST	

POL_MPMV - только остатки	P07572	Вирус обезьян Мейзона-Пфайзера	MGQELSQHERYVEQLKQALKTR GVKVKYADLLKFFDFVKDTCPWF PQEGTIDIKRWRRVGDCFQDYYN TFGPEKVPVTAFSYWNLIKELIDK KEVNPQVMAAVAQTEEILKSNSQ TDLTKTSQNPDLDLISLDSDDEGA KSSSLQDKGLSSTKKPKRFPVLLT AQTSKDPEDPNPSEVDWDGLEDE AAKYHNPDWPPFLTRPPPYNKAT PSAPTVMAVVNPKEELKEKIAQL EEQIKLEELHQALISKLQKLKTGN ETVTHPDTAGGLSRTPHWPGQHIP KGKCCASREKEEQIPKDIFPVTET VDGQGQAWRHHNGFDFAVIKEL KTAASQYGATAPYTLAIVESVAD NWLTPTDWNTLVRAVLSGGDHL LWKSEFFENCRDTAKRNQQAGN GWDFDMLTGSGNYSSTDAQMQY DPGLFAQIQAAATKAWRKLPVKG DPGASLTGVKQGPDEPFADFVHR LITTAGRIFGSAEAGVDYVKQLAY ENANPACQAAIRPYRKKTDLTGYI RLCSDIGPSYQQGLAMAAAFSGQ TVKDFLNNKNKEKGGCCFKCGK KGHFAKNCHEHAHNNAEPKVPG LCPRCKRGKHWANECKSKTDNQ GNPIPPHQGNRVEGPAPGPETSLW GSQLCSSQQKQPISKLTRATPGSA GLDLCSTSHTVLTPEMGPQALSTG IYGPLPPNTFGLILGRSSITMKGLQ VYPGVIDNDYTGEIKIMAKAVNNI VTVSQGNRIAQLILLPLIETDNKV QQPYRGQGSFGSSDIYWVQPITCQ KPSLTLWLDDKMFTGLIDTGADV TIIKLEDWPPNWPITDTLTNLRGIG	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			QSNNPKQSSKYLTWRDKENNSGL IKPFVIPNLPVNLWGRDLLSQMKI MMCSPNDIVTAQMLAQGYSPGK GLGKKENGILHPIPNQGQSNKKGF GNFLTAAIDILAPQQCAEPITWKS DEPVWVDQWPLTNDKLAAAQQL VQEQLEAGHITESSSPWNTPIFVIK KKSGKWRLLQDLRAVNATMVLM GALQPGLPSPVAIPQGYLKIIIDLK DCFFSIPLHPSDQKRFAFSLPSTNF KEPMQRFQWKVLPQGMANSPTL CQKYVATAIHKVRHAWKQMYIIH YMDDILIAGKDGQQVLQCFDQLK QELTAAGLHIAPEKVQLQDPYTY LGFELNGPKITNQKAVIRKDKLQT LNDFQKLLGDINWLRPYLKLTTG DLKPLFDTLKGDSDPNSHRSLSKE ALASLEKVETAIAEQFVTHINYSL PLIFLIFNTALTPTGLFWQDNPIM WIHLPASPKKVLLPYYDAIADLIIL GRDHSKKYFGIEPSTIIQPYSKSQI DWLMQNTEMWPIACASFVGILDN HYPPNKLIQFCKLHTFVFPQIISKT PLNNALLVFTDGSSTGMAAYTLT DTTIKFQTNLNSAQLVELQALIAV LSAFPNQPLNIYTDSAYLAHSIPLL ETVAQIKHISETAKLFLQCQQLIY NRSIPFYIGHVRAHSGLPGPIAQG NQRADLATKIVASNINTNLESAQN AHTLHHLNAQTLRLMFNIPREQA RQIVKQCPICVTYLPVPHLGVNPR GLFPNMIWQMDVTHYSEFGNLK YIHVSIDTFSGFLLATLQTGETTKH VITHLLHCFSIIGLPKQIKTDNGPG YTSKNFQEFCSTLQIKHITGIPYNP	

			QGQGIVERAHLSLKTTIEKIKKGE WYPRKGTPRNILNHALFILNFLNL DDQNKSAADRFWHNNPKKQFAM VKWKDPLDNTWHGPDPVLIWGR GSVCVYSQTYDAARWLPERLVR QVSNNNQSRE	

POL_MMTVB - только остатки	P03365	Вирус опухоли молочной железы мышей	MGVSGSKGQKLFVSVLQRLLSER GLHVKESSAIEFYQFLIKVSPWFPE EGGLNLQDWKRVGREMKRYAAE HGTDSIPKQAYPIWLQLREILTEQS DLVLLSAEAKSVTEEELEEGLTGL LSTSSQEKTYGTRGTAYAEIDTEV DKLSEHIYDEPYEEKEKADKNEE KDHVRKIKKVVQRKENSEGKRKE KDSKAFLATDWNDDDLSPEDWD DLEEQAAHYHDDDELILPVKRKV VKKKPQALRRKPLPPVGFAGAMA EAREKGDLTFTFPVVFMGESDED DTPVWEPLPLKTLKELQSAVRTM GPSAPYTLQVVDMVASQWLTPSD WHQTARATLSPGDYVLWRTEYE EKSKEMVQKAAGKRKGKVSLDM LLGTGQFLSPSSQIKLSKDVLKDV TTNAVLAWRAIPPPGVKKTVLAG LKQGNEESYETFISRLEEAVYRM MPRGEGSDILIKQLAWENANSLC QDLIRPIRKTGTIQDYIRACLDASP AVVQGMAYAAAMRGQKYSTFV KQTYGGGKGGQGAEGPVCFSCG KTGHIRKDCKDEKGSKRAPPGLC PRCKKGYHWKSECKSKFDKDGN PLPPLETNAENSKNLVKGQSPSPA QKGDGVKGSGLNPEAPPFTIHDLP RGTPGSAGLDLSSQKDLILSLEDG VSLVPTLVKGTLPEGTTGLIIGRSS NYKKGLEVLPGVIDSDFQGEIKV MVKAAKNAVIIHKGERIAQLLLLP YLKLPNPVIKEERGSEGFGSTSHV HWVQEISDSRPMLHIYLNGRRFL GLLDTGADKTCIAGRDWPANWPI HQTESSLQGLGMACGVARSSQPL	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			RWQHEDKSGIIHPFVIPTLPFTLW GRDIMKDIKVRLMTDSPDDSQDL MIGAIESNLFADQISWKSDQPVWL NQWPLKQEKLQALQQLVTEQLQ LGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQP GLPSPVAVPKGWEIIIIDLQDCFFN IKLHPEDCKRFAFSVPSPNFKRPY QRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDD ILLAHPSRSIVDEILTSMIQALNKH GLVVSTEKIQKYDNLKYLGTHIQ GDSVSYQKLQIRTDKLRTLNDFQ KLLGNINWIRPFLKLTTGELKPLF EILNGDSNPISTRKLTPEACKALQL MNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLP HISPKVITPYDIFCTQLIIKGRHRSK ELFSKDPDYIVVPYTKVQFDLLLQ EKEDWPISLLGFLGEVHFHLPKDP LLTFTLQTAIIFPHMTSTTPLEKGI VIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQP FNLYTDSKYVTGLFPEIETATLSPR TKIYTELKHLQRLIHKRQEKFYIG HIRGHTGLPGPLAQGNAYADSLT RILTALESAQESHALHHQNAAAL RFQFHITREQAREIVKLCPNCPDW GHAPQLGVNPRGLKPRVLWQMD VTHVSEFGKLKYVHVTVDTYSHF TFATARTGEATKDVLQHLAQSFA YMGIPQKIKTDNAPAYVSRSIQEF LARWKISHVTGIPYNPQGQAIVER THQNIKAQLNKLQKAGKYYTPHH LLAHALFVLNHVNMDNQGHTAA	

			ERHWGPISADPKPMVMWKDLLT GSWKGPDVLITAGRGYACVFPQD AETPIWVPDRFIRPFTERKEATPTP GTAEKTPPRDEKDQQESPKNESSP HQREDGLATSAGVDLRSGGGP	

POL_MLVMS - только остатки	P03355	Вирус лейкоза мышей Молони	MGQTVTTPLSLTLGHWKDVERIA HNQSVDVKKRRWVTFCSAEWPT FNVGWPRDGTFNRDLITQVKIKV FSPGPHGHPDQVPYIVTWEALAF DPPPWVKPFVHPKPPPPLPPSAPSL PLEPPRSTPPRSSLYPALTPSLGAK PKPQVLSDSGGPLIDLLTEDPPPYR DPRPPPSDRDGNGGEATPAGEAP DPSPMASRLRGRREPPVADSTTSQ AFPLRAGGNGQLQYWPFSSSDLY NWKNNNPSFSEDPGKLTALIESVL ITHQPTWDDCQQLLGTLLTGEEK QRVLLEARKAVRGDDGRPTQLPN EVDAAFPLERPDWDYTTQAGRN HLVHYRQLLLAGLQNAGRSPTNL AKVKGITQGPNESPSAFLERLKEA YRRYTPYDPEDPGQETNVSMSFI WQSAPDIGRKLERLEDLKNKTLG DLVREAEKIFNKRETPEEREERIRR ETEEKEERRRTEDEQKEKERDRR RHREMSKLLATVVSGQKQDRQG GERRRSQLDRDQCAYCKEKGHW AKDCPKKPRGPRGPRPQTSLLTLD DQGGQGQEPPPEPRITLKVGGQPV TFLVDTGAQHSVLTQNPGPLSDK SAWVQGATGGKRYRWTTDRKV HLATGKVTHSFLHVPDCPYPLLG RDLLTKLKAQIHFEGSGAQVMGP MGQPLQVLTLNIEDEHRLHETSKE PDVSLGSTWLSDFPQAWAETGG MGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYR PVQDLREVNKRVEDIHPTVPNPY NLLSGLPPSHQWYTVLDLKDAFF	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			CLRLHPTSQPLFAFEWRDPEMGIS GQLTWTRLPQGFKNSPTLFDEAL HRDLADFRIQHPDLILLQYVDDLL LAATSELDCQQGTRALLQTLGNL GYRASAKKAQICQKQVKYLGYLL KEGQRWLTEARKETVMGQPTPK TPRQLREFLGTAGFCRLWIPGFAE MAAPLYPLTKTGTLFNWGPDQQ KAYQEIKQALLTAPALGLPDLTKP FELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPC LRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSN ARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILA EAHGTRPDLTDQPLPDADHTWYT DGSSLLQEGQRKAGAAVTTETEV IWAKALPAGTSAQRAELIALTQAL KMAEGKKLNVYTDSRYAFATAHI HGEIYRRRGLLTSEGKEIKNKDEI LALLKALFLPKRLSIIHCPGHQKG HSAEARGNRMADQAARKAAITET PDTSTLLIENSSPYTSEHFHYTVTD IKDLTKLGAIYDKTKKYWVYQGK PVMPDQFTFELLDFLHQLTHLSFS KMKALLERSHSPYYMLNRDRTLK NITETCKACAQVNASKSAVKQGT RVRGHRPGTHWEIDFTEIKPGLYG YKYLLVFIDTFSGWIEAFPTKKET AKVVTKKLLEEIFPRFGMPQVLGT DNGPAFVSKVSQTVADLLGIDWK LHCAYRPQSSGQVERMNRTIKET LTKLTLATGSRDWVLLLPLALYR ARNTPGPHGLTPYEILYGAPPPLV NFPDPDMTRVTNSPSLQAHLQAL	

			YLVQHEVWRPLAAAYQEQLDRP VVPHPYRVGDTVWVRRHQTKNL EPRWKGPYTVLLTTPTALKVDGI AAWIHAAHVKAADPGGGPSSRLT WRVQRSQNPLKIRLTREAP	

POL_HTL1A - только остатки	P03362	Вирус Т-клеточного лейкоза человека 1 типа	MGQIFSRSASPIPRPPRGLAAHHW LNFLQAAYRLEPGPSSYDFHQLK KFLKIALETPARICPINYSLLASLLP KGYPGRVNEILHILIQTQAQIPSRP APPPPSSPTHDPPDSDPQIPPPYVE PTAPQVLPVMHPHGAPPNHRPWQ MKDLQAIKQEVSQAAPGSPQFMQ TIRLAVQQFDPTAKDLQDLLQYL CSSLVASLHHQQLDSLISEAETRGI TGYNPLAGPLRVQANNPQQQGLR REYQQLWLAAFAALPGSAKDPS WASILQGLEEPYHAFVERLNIALD NGLPEGTPKDPILRSLAYSNANKE CQKLLQARGHTNSPLGDMLRAC QTWTPKDKTKVLVVQPKKPPPNQ PCFRCGKAGHWSRDCTQPRPPPG PCPLCQDPTHWKRDCPRLKPTIPE PEPEEDALLLDLPADIPHPKNLHR GGGLTSPPTLQQVLPNQDPASILP VIPLDPARRPVIKAQVDTQTSHPK TIEALLDTGADMTVLPIALFSSNTP LKNTSVLGAGGQTQDHFKLTSLP VLIRLPFRTTPIVLTSCLVDTKNN WAIIGRDALQQCQGVLYLPEAKR PPVILPIQAPAVLGLEHLPRPPQIS QFPLNPERLQALQHLVRKALEAG HIEPYTGPGNNPVFPVKKANGTW RFIHDLRATNSLTIDLSSSSPGPPD LSSLPTTLAHLQTIDLRDAFFQIPL PKQFQPYFAFTVPQQCNYGPGTR YAWKVLPQGFKNSPTLFEMQLAH ILQPIRQAFPQCTILQYMDDILLAS PSHEDLLLLSEATMASLISHGLPV SENKTQQTPGTIKFLGQIISPNHLT YDAVPTVPIRSRWALPELQALLGE	IPR043502, SSF56672, IPR000477, PF00078

			IQWVSKGTPTLRQPLHSLYCALQ RHTDPRDQIYLNPSQVQSLVQLR QALSQNCRSRLVQTLPLLGAIMLT LTGTTTVVFQSKEQWPLVWLHAP LPHTSQCPWGQLLASAVLLLDKY TLQSYGLLCQTIHHNISTQTFNQFI QTSDHPSVPILLHHSHRFKNLGAQ TGELWNTFLKTAAPLAPVKALMP VFTLSPVIINTAPCLFSDGSTSRAA YILWDKQILSQRSFPLPPPHKSAQ RAELLGLLHGLSSARSWRCLNIFL DSKYLYHYLRTLALGTFQGRSSQ APFQALLPRLLSRKVVYLHHVRS HTNLPDPISRLNALTDALLITPVLQ LSPAELHSFTHCGQTALTLQGATT TEASNILRSCHACRGGNPQHQMP RGHIRRGLLPNHIWQGDITHFKYK NTLYRLHVWVDTFSGAISATQKR KETSSEAISSLLQAIAHLGKPSYIN TDNGPAYISQDFLNMCTSLAIRHT THVPYNPTSSGLVERSNGILKTLL YKYFTDKPDLPMDNALSIALWTI NHLNVLTNCHKTRWQLHHSPRL QPIPETRSLSNKQTHWYYFKLPGL NSRQWKGPQEALQEAAGAALIPV SASSAQWIPWRLLKRAACPRPVG GPADPKEKDLQHHG	

POL_FOAMV - только остатки	P14350	Спумаретровирус человека	MNPLQLLQPLPAEIKGTKLLAHW DSGATITCIPESFLEDEQPIKKTLIK TIHGEKQQNVYYVTFKVKGRKVE AEVIASPYEYILLSPTDVPWLTQQ PLQLTILVPLQEYQEKILSKTALPE DQKQQLKTLFVKYDNLWQHWEN QVGHRKIRPHNIATGDYPPRPQKQ YPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRM VLDYREVNKTIPLTAAQNQHSAGI LATIVRQKYKTTLDLANGFWAHP ITPESYWLTAFTWQGKQYCWTRL PQGFLNSPALFTADVVDLLKEIPN VQVYVDDIYLSHDDPKEHVQQLE KVFQILLQAGYVVSLKKSEIGQKT VEFLGFNITKEGRGLTDTFKTKLL NITPPKDLKQLQSILGLLNFARNFI PNFAELVQPLYNLIASAKGKYIEW SEENTKQLNMVIEALNTASNLEER LPEQRLVIKVNTSPSAGYVRYYNE TGKKPIMYLNYVFSKAELKFSML EKLLTTMHKALIKAMDLAMGQEI LVYSPIVSMTKIQKTPLPERKALPI RWITWMTYLEDPRIQFHYDKTLP ELKHIPDVYTSSQSPVKHPSQYEG VFYTDGSAIKSPDPTKSNNAGMGI VHATYKPEYQVLNQWSIPLGNHT AQMAEIAAVEFACKKALKIPGPV LVITDSFYVAESANKELPYWKSN GFVNNKKKPLKHISKWKSIAECLS MKPDITIQHEKGISLQIPVFILKGN ALADKLATQGSYVVNCNTKKPN LDAELDQLLQGHYIKGYPKQYTY FLEDGKVKVSRPEGVKIIPPQSDR QKIVLQAHNLAHTGREATLLKIA	IPR043502, SSF56672, IPR000477, PF00078

			NLYWWPNMRKDVVKQLGRCQQ CLITNASNKASGPILRPDRPQKPFD KFFIDYIGPLPPSQGYLYVLVVVD GMTGFTWLYPTKAPSTSATVKSL NVLTSIAIPKVIHSDQGAAFTSSTF AEWAKERGIHLEFSTPYHPQSGSK VERKNSDIKRLLTKLLVGRPTKW YDLLPVVQLALNNTYSPVLKYTP HQLLFGIDSNTPFANQDTLDLTRE EELSLLQEIRTSLYHPSTPPASSRS WSPVVGQLVQERVARPASLRPR WHKPSTVLKVLNPRTVVILDHLG NNRTVSIDNLKPTSHQNGTTNDT ATMDHLEKNE	

POL_BLVJ - только остатки	P03361	Вирус лейкоза крупного рогатого скота	MGNSPSYNPPAGISPSDWLNLLQS AQRLNPRPSPSDFTDLKNYIHWFH KTQKKPWTFTSGGPTSCPPGRFGR VPLVLATLNEVLSNEGGAPGASA PEEQPPPYDPPAILPIISEGNRNRH RAWALRELQDIKKEIENKAPGSQ VWIQTLRLAILQADPTPADLEQLC QYIASPVDQTAHMTSLTAAIAAA EAANTLQGFNPKTGTLTQQSAQP NAGDLRSQYQNLWLQAGKNLPT RPSAPWSTIVQGPAESSVEFVNRL QISLADNLPDGVPKEPIIDSLSYAN ANRECQQILQGRGPVAAVGQKLQ ACAQWAPKNKQPALLVHTPGPK MPGPRQPAPKRPPPGPCYRCLKE GHWARDCPTKATGPPPGPCPICK DPSHWKRDCPTLKSKNKLIEGGL SAPQTITPITDSLSEAELECLLSIPL ARSRPSVAVYLSGPWLQPSQNQA LMLVDTGAENTVLPQNWLVRDY PRIPAAVLGAGGVSRNRYNWLQG PLTLALKPEGPFITIPKILVDTSDK WQILGRDVPSRLQASISIPEEVRPP VVGVLDTPPSHIGLEHLPPPPEVP QFPLNLERLQALQDLVHRSLEAG YISPWDGPGNNPVFPVRKPNGAW RFVHDLRATNALTKPIPALSPGPP DLTAIPTHPPHIICLDLKDAFFQIPV EDRFRFYLSFTLPSPGGLQPHRRF AWRVLPQGFINSPALFERALQEPL RQVSAAFSQSLLVSYMDDILYASP TEEQRSQCYQALAARLRDLGFQV ASEKTSQTPSPVPFLGQMVHEQIV TYQSLPTLQISSPISLHQLQAVLGD LQWVSRGTPTTRRPLQLLYSSLKR	IPR043502, SSF56672, IPR000477, PF00078

			HHDPRAIIQLSPEQLQGIAELRQAL SHNARSRYNEQEPLLAYVHLTRA GSTLVLFQKGAQFPLAYFQTPLTD NQASPWGLLLLLGCQYLQTQALS SYAKPILKYYHNLPKTSLDNWIQS SEDPRVQELLQLWPQISSQGIQPP GPWKTLITRAEVFLTPQFSPDPIPA ALCLFSDGATGRGAYCLWKDHL LDFQAVPAPESAQKGELAGLLAG LAAAPPEPVNIWVDSKYLYSLLRT LVLGAWLQPDPVPSYALLYKSLL RHPAIVVGHVRSHSSASHPIASLN NYVDQLLPLETPEQWHKLTHCNS RALSRWPNPRISAWDPRSPATLCE TCQKLNPTGGGKMRTIQRGWAP NHIWQADITHYKYKQFTYALHVF VDTYSGATHASAKRGLTTQTTIE GLLEAIVHLGRPKKLNTDQGANY TSKTFVRFCQQFGVSLSHHVPYNP TSSGLDERTNGLLKLLLSKYHLDE PHLPMTQALSRALWTHNQINLLPI LKTRWELHHSPPLAVISEGGETPK GSDKLFLYLLPGQNNRRWLGPLP ALVEASGGALLATDPPVWVPWR LLKAFKCLKNDGPEDAHNRSSDG	

O41894_9RETR - только остатки	O41894	Вирус пенистости крупного рогатого скота	MPALRPLQVEIKGNHLKGYWDSG AEITCVPAIYIIEEQPVGKKLITTIH NEKEHDVYYVEMKIEKRKVQCE VIATALDYVLVAPVDIPWYKPGP LELTIKIDVESQKHTLITESTLSPQ GQMRLKKLLDQYQALWQCWEN QVGHRRIEPHKIATGALKPRPQKQ YHINPRAKADIQIVIDDLLRQGVL RQQNSEMNTPVYPVPKADGRWR MVLDYREVNKVTPLVATQNCHS ASILNTLYRGPYKSTLDLANGFW AHPIKPEDYWITAFTWGGKTYCW TVLPQGFLNSPALFTADVVDILKD IPNVQVYVDDVYVSSATEQEHLDI LETIFNRLSTAGYIVSLKKSKLAK ETVEFLGFSISQNGRGLTDSYKQK LMDLQPPTTLRQLQSILGLINFAR NFLPNFAELVAPLYQLIPKAKGQC IPWTMDHTTQLKTIIQALNSTENL EERRPDVDLIMKVHISNTAGYIRF YNHGGQKPIAYNNALFTSTELKFT PTEKIMATIHKGLLKALDLSLGKE IHVYSAIASMTKLQKTPLSERKAL SIRWLKWQTYFEDPRIKFHHDAT LPDLQNLPVPQQDTGKEMTILPLL HYEAIFYTDGSAIRSPKPNKTHSA GMGIIQAKFEPDFRIVHLWSFPLG DHTAQYAEIAAFEFAIRRATGIRG PVLIVTDSNYVAKSYNEELPYWE SNGFVNNKKKTLKHISKWKAIAE CKNLKADIHVIHEPGHQPAEASPH AQGNALADKQAVSGSYKVFSNEL KPSLDAELEQVLSTGRPNPQGYPN KYEYKLVNGLCYVDRRGEEGLKI IPPKADRVKLCQLAHDGPGSAHL	IPR043502, SSF56672, IPR000477, PF00078

			GRSALLLKLQQKYWWPRMHIDA SRIVLNCTVCAQTNSTNQKPRPPL VIPHDTKPFQVWYMDYIGPLPPSN GYQHALVIVDAGTGFTWIYPTKA QTANATVKALTHLTGTAVPKVLH SDQGPAFTSSILADWAKDRGIQLE HSAPYHPQSSGKVERKNSEIKRLL TKLLAGRPTKWYPLIPIVQLALNN TPNTRQKYTPHQLMYGADCNLPF ENLDTLDLTREEQLAVLKEVRDG LLDLYPSPSQTTARSWTPSPGLLV QERVARPAQLRPKWRKPTPIKKV LNERTVIIDHLGQDKVVSIDNLKP AAHQKLAQTPDSAEICPSATPCPP NTSLWYDLDTGTWTCQRCGYQC PDKYHQPQCTWSCEDRCGHRWK ECGNCIPQDGSSDDASAVAAVEI		

POL_MLVBM - только остатки	Q7SVK7	Вирус лейкоза мышей	MGQTVTTPLSLTLEHWGDVQRIA SNQSVGVKKRRWVTFCSAEWPTF GVGWPQDGTFNLDIILQVKSKVFS PGPHGHPDQVPYIVTWEAIAYEPP PWVKPFVSPKLSLSPTAPILPSGPS TQPPPRSALYPAFTPSIKPRPSKPQ VLSDDGGPLIDLLTEDPPPYGEQG PSSPDGDGDREEATSTSEIPAPSPM VSRLRGKRDPPAADSTTSRAFPLR LGGNGQLQYWPFSSSDLYNWKN NNPSFSEDPGKLTALIESVLTTHQP TWDDCQQLLGTLLTGEEKQRVLL EARKAVRGNDGRPTQLPNEVNSA FPLERPDWDYTTPEGRNHLVLYR QLLLAGLQNAGRSPTNLAKVKGI TQGPNESPSAFLERLKEAYRRYTP YDPEDPGQETNVSMSFIWQSAPAI GRKLERLEDLKSKTLGDLVREAE KIFNKRETPEEREERIRRETEEKEE RRRAGDEQREKERDRRRQREMS KLLATVVTGQRQDRQGGERRRPQ LDKDQCAYCKEKGHWAKDCPKK PRGPRGPRPQTSLLTLDDQGGQG QEPPPEPRITLTVGGQPVTFLVDT GAQHSVLTQNPGPLSDRSAWVQG ATGGKRYRWTTDRKVHLATGKV THSFLHVPDCPYPLLGRDLLTKLK AQIHFEGSGAQVVGPKGQPLQVL TLGIEDEYRLHETSTEPDVSLGST WLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIQQYPMSHEAR LGIKPHIQRLLDQGILVPCQSPWN TPLLPVKKPGTNDYRPVQDLREV NKRVEDIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLRLHPTSQ	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			PLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRI QHPDLILLQYVDDILLAATSELDC QQGTRALLQTLGDLGYRASAKK AQICQKQVKYLGYLLREGQRWLT EARKETVMGQPVPKTPRQLREFL GTAGFCRLWIPGFAEMAAPLYPL TKTGTLFSWGPDQQKAYQEIKQA LLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLS KKLDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAM LLDTDRVQFGPVVALNPATLLPLP EEGAPHDCLEILAETHGTRPDLTD QPIPDADHTWYTDGSSFLQEGQR KAGAAVTTETEVIWAGALPAGTS AQRAELIALTQALKMAEGKRLNV YTDSRYAFATAHIHGEIYRRRGLL TSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLAD QAAREAAIKTPPDTSTLLIEDSTPY TPAYFHYTETDLKKLRDLGATYN QSKGYWVFQGKPVMPDQFVFEL LDSLHRLTHLGYQKMKALLDRGE SPYYMLNRDKTLQYVADSCTVC AQVNASKAKIGAGVRVRGHRPGT HWEIDFTEVKPGLYGYKYLLVFV DTFSGWVEAFPTKRETARVVSKK LLEEIFPRFGMPQVLGSDNGPAFT SQVSQSVADLLGIDWKLHCAYRP QSSGQVERINRTIKETLTKLTLAA GTRDWVLLLPLALYRARNTPGPH GLTPYEILYGAPPPLVNFHDPDMS ELTNSPSLQAHLQALQTVQREIW	

			KPLAEAYRDRLDQPVIPHPFRIGD SVWVRRHQTKNLEPRWKGPYTV LLTTPTALKVDGISAWIHAAHVK AATTPPIKPSWRVQRSQNPLKIRL TRGAP	

Таблица 31. Иллюстративные димерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	**Номер доступа**	**Организм**	**Последовательность**	**Сигнатуры RT**
Q83133_AVIMA	Q83133	Вирус, ассоциированный с миелобластозом птиц, 1 типа	RATVLTVALHLAIPLKWKPNHTPVW IDQWPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYRLL HDLRAVNAKLVPFGAVQQGAPVLS ALPRGWPLMVLDLKDCFFSIPLAEQ DREAFAFTLPSVNNQAPARRFQWKV LPQGMTCSPTICQLIVGQILEPLRLKH PSLRMLHYMDDLLLAASSHDGLEA AGEEVISTLERAGFTISPDKVQREPG VQYLGYKLGSTYVAPVGLVAEPRIA TLWDVQKLVGSLQSVRPALGIPPRL MGPFYEQLRGSDPNEAREWNLDMK MAWREIVQLSTTAALERWDPALPLE GAVARCEQGAIGVLGQGLSTHPRPC LWLFSTQPTKAFTAWLEVLTLLITKL RASAVRTFGKEVDILLLPACFREDLP LPEGILLALRGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDA SSSTHKGVVVWREGPRWEIKEIADL GASVQQLEARAVAMALLLWPTTPT NVVTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHVRSH SEVPGFFTEGNDVADSQATFQAYPL REAKDLHTALHIGPRALSKACNISM QQAREVVQTCPHCNSAPALEAGVNP RGLGPLQIWQTDFTLEPRMAPRSWL AVTVDTASSAIVVTQHGRVTSVAAQ HHWATAIAVLGRPKAIKTDNGSCFT SKSTREWLARWGIAHTTGIPGNSQG QAMVERANRLLKDKIRVLAEGDGF MKRIPTSKQGELLAKAMYALNHFER	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			GENTKTPIQKHWRPTVLTEGPPVKIR IETGEWEKGWNVLVWGRGYAAVK NRDTDKVIWVPSRKVKPDITQKDEV TKKDEASPLFAGISDWAPWEGEQEG LQEETASNKQERPGEDTPAANES	
POL_ SIVM 1	P05896	Вирус иммуноде фицита обезьян	MGARNSVLSGKKADELEKIRLRPGG KKKYMLKHVVWAANELDRFGLAES LLENKEGCQKILSVLAPLVPTGSENL KSLYNTVCVIWCIHAEEKVKHTEEA KQIVQRHLVMETGTAETMPKTSRPT APFSGRGGNYPVQQIGGNYTHLPLS PRTLNAWVKLIEEKKFGAEVVSGFQ ALSEGCLPYDINQMLNCVGDHQAA MQIIRDIINEEAADWDLQHPQQAPQ QGQLREPSGSDIAGTTSTVEEQIQW MYRQQNPIPVGNIYRRWIQLGLQKC VRMYNPTNILDVKQGPKEPFQSYVD RFYKSLRAEQTDPAVKNWMTQTLLI QNANPDCKLVLKGLGTNPTLEEMLT ACQGVGGPGQKARLMAEALKEALA PAPIPFAAAQQKGPRKPIKCWNCGK EGHSARQCRAPRRQGCWKCGKMD HVMAKCPNRQAGFFRPWPLGKEAP QFPHGSSASGADANCSPRRTSCGSA KELHALGQAAERKQREALQGGDRG FAAPQFSLWRRPVVTAHIEGQPVEV LLDTGADDSIVTGIELGPHYTPKIVG GIGGFINTKEYKNVEIEVLGKRIKGTI MTGDTPINIFGRNLLTALGMSLNLPI AKVEPVKSPLKPGKDGPKLKQWPLS KEKIVALREICEKMEKDGQLEEAPPT NPYNTPTFAIKKKDKNKWRMLIDFR ELNRVTQDFTEVQLGIPHPAGLAKR KRITVLDIGDAYFSIPLDEEFRQYTAF TLPSVNNAEPGKRYIYKVLPQGWKG	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			SPAIFQYTMRHVLEPFRKANPDVTL VQYMDDILIASDRTDLEHDRVVLQL KELLNSIGFSSPEEKFQKDPPFQWMG YELWPTKWKLQKIELPQRETWTVN DIQKLVGVLNWAAQIYPGIKTKHLC RLIRGKMTLTEEVQWTEMAEAEYEE NKIILSQEQEGCYYQESKPLEATVIK SQDNQWSYKIHQEDKILKVGKFAKI KNTHTNGVRLLAHVIQKIGKEAIVI WGQVPKFHLPVEKDVWEQWWTDY WQVTWIPEWDFISTPPLVRLVFNLV KDPIEGEETYYVDGSCSKQSKEGKA GYITDRGKDKVKVLEQTTNQQAELE AFLMALTDSGPKANIIVDSQYVMGII TGCPTESESRLVNQIIEEMIKKTEIYV AWVPAHKGIGGNQEIDHLVSQGIRQ VLFLEKIEPAQEEHSKYHSNIKELVF KFGLPRLVAKQIVDTCDKCHQKGEA IHGQVNSDLGTWQMDCTHLEGKIVI VAVHVASGFIEAEVIPQETGRQTALF LLKLASRWPITHLHTDNGANFASQE VKMVAWWAGIEHTFGVPYNPQSQG VVEAMNHHLKNQIDRIREQANSVET IVLMAVHCMNFKRRGGIGDMTPAE RLINMITTEQEIQFQQSKNSKFKNFR VYYREGRDQLWKGPGELLWKGEG AVILKVGTDIKVVPRRKAKIIKDYGG GKEMDSSSHMEDTGEAREVA	

POL_RSVP	P03354	Вирус саркомы Рауса	MEAVIKVISSACKTYCGKTSPSKKEI GAMLSLLQKEGLLMSPSDLYSPGSW DPITAALSQRAMILGKSGELKTWGL VLGALKAAREEQVTSEQAKFWLGL GGGRVSPPGPECIEKPATERRIDKGE EVGETTVQRDAKMAPEETATPKTV GTSCYHCGTAIGCNCATASAPPPPY VGSGLYPSLAGVGEQQGQGGDTPPG AEQSRAEPGHAGQAPGPALTDWAR VREELASTGPPVVAMPVVIKTEGPA WTPLEPKLITRLADTVRTKGLRSPIT MAEVEALMSSPLLPHDVTNLMRVIL GPAPYALWMDAWGVQLQTVIAAAT RDPRHPANGQGRGERTNLNRLKGL ADGMVGNPQGQAALLRPGELVAIT ASALQAFREVARLAEPAGPWADIM QGPSESFVDFANRLIKAVEGSDLPPS ARAPVIIDCFRQKSQPDIQQLIRTAPS TLTTPGEIIKYVLDRQKTAPLTDQGI AAAMSSAIQPLIMAVVNRERDGQTG SGGRARGLCYTCGSPGHYQAQCPK KRKSGNSRERCQLCNGMGHNAKQC RKRDGNQGQRPGKGLSSGPWPGPEP PAVSLAMTMEHKDRPLVRVILTNTG SHPVKQRSVYITALLDSGADITIISEE DWPTDWPVMEAANPQIHGIGGGIPM RKSRDMIELGVINRDGSLERPLLLFP AVAMVRGSILGRDCLQGLGLRLTNL IGRATVLTVALHLAIPLKWKPDHTP VWIDQWPLPEGKLVALTQLVEKELQ LGHIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGAPV LSALPRGWPLMVLDLKDCFFSIPLAE QDREAFAFTLPSVNNQAPARRFQWK VLPQGMTCSPTICQLVVGQVLEPLR	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			LKHPSLCMLHYMDDLLLAASSHDG LEAAGEEVISTLERAGFTISPDKVQR EPGVQYLGYKLGSTYVAPVGLVAEP RIATLWDVQKLVGSLQWLRPALGIP PRLMGPFYEQLRGSDPNEAREWNLD MKMAWREIVRLSTTAALERWDPAL PLEGAVARCEQGAIGVLGQGLSTHP RPCLWLFSTQPTKAFTAWLEVLTLLI TKLRASAVRTFGKEVDILLLPACFRE DLPLPEGILLALKGFAGKIRSSDTPSI FDIARPLHVSLKVRVTDHPVPGPTVF TDASSSTHKGVVVWREGPRWEIKEI ADLGASVQQLEARAVAMALLLWPT TPTNVVTDSAFVAKMLLKMGQEGV PSTAAAFILEDALSQRSAMAAVLHV RSHSEVPGFFTEGNDVADSQATFQA YPLREAKDLHTALHIGPRALSKACNI SMQQAREVVQTCPHCNSAPALEAG VNPRGLGPLQIWQTDFTLEPRMAPR SWLAVTVDTASSAIVVTQHGRVTSV AVQHHWATAIAVLGRPKAIKTDNGS CFTSKSTREWLARWGIAHTTGIPGNS QGQAMVERANRLLKDRIRVLAEGD GFMKRIPTSKQGELLAKAMYALNHF ERGENTKTPIQKHWRPTVLTEGPPV KIRIETGEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPDITQKD EVTKKDEASPLFAGISDWIPWEDEQ EGLQGETASNKQERPGEDTLAANES	

POL_HV2D2	P15833	Вирус иммунодефицита человека 2 типа	MGARGSVLSGKKTDELEKVRLRPG GKKKYMLKHVVWAVNELDRFGLA ESLLESKEGCQKILKVLAPLVPTGSE NLKSLFNIVCVIFCLHAEEKVKDTEE AKKIAQRHLAADTEKMPATNKPTAP PSGGNYPVQQLAGNYVHLPLSPRTL NAWVKLVEEKKFGAEVVPGFQALS EGCTPYDINQMLNCVGEHQAAMQII REIINEEAADWDQQHPSPGPMPAGQ LRDPRGSDIAGTTSTVEEQIQWMYR AQNPVPVGNIYRRWIQLGLQKCVR MYNPTNILDIKQGPKEPFQSYVDRFY KSLRAEQTDPAVKNWMTQTLLIQN ANPDCKLVLKGLGMNPTLEEMLTA CQGIGGPGQKARLMAEALKEALTPA PIPFAAVQQKAGKRGTVTCWNCGK QGHTARQCRAPRRQGCWKCGKTGH IMSKCPERQAGFLRVRTLGKEASQL PHDPSASGSDTICTPDEPSRGHDTSG GDTICAPCRSSSGDAEKLHADGETTE REPRETLQGGDRGFAAPQFSLWRRP VVKACIEGQSVEVLLDTGVDDSIVA GIELGSNYTPKIVGGIGGFINTKEYK DVEIEVVGKRVRATIMTGDTPINIFG RNILNTLGMTLNFPVAKVEPVKVEL KPGKDGPKIRQWPLSREKILALKEIC EKMEKEGQLEEAPPTNPYNTPTFAIK KKDKNKWRMLIDFRELNKVTQDFT EVNWVFPTRQVAEKRRITVIDVGDA YFSIPLDPNFRQYTAFTLPSVNNAEP GKRYIYKVLPQGWKGSQSICQYSMR KVLDPFRKANSDVIIIQYMDDILIASD RSDLEHDRVVSQLKELLNDMGFSTP EEKFQKDPPFKWMGYELWPKKWKL QKIQLPEKEVWTVNAIQKLVGVLN	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			WAAQLFPGIKTRHICKLIRGKMTLTE EVQWTELAEAELQENKIILEQEQEGS YYKERVPLEATVQKNLANQWTYKI HQGNKVLKVGKYAKVKNTHTNGV RLLAHVVQKIGKEALVIWGEIPVFHL PVERETWDQWWTDYWQVTWIPEW DFVSTPPLIRLAYNLVKDPLEGRETY YTDGSCNRTSKEGKAGYVTDRGKD KVKVLEQTTNQQAELEAFALALTDS EPQVNIIVDSQYVMGIIAAQPTETESP IVAKIIEEMIKKEAVYVGWVPAHKG LGGNQEVDHLVSQGIRQVLFLEKIEP AQEEHEKYHGNVKELVHKFGIPQLV AKQIVNSCDKCQQKGEAIHGQVNA DLGTWQMDCTHLEGKIIIVAVHVAS GFIEAEVIPQETGRQTALFLLKLASR WPITHLHTDNGANFTSPSVKMVAW WVGIEQTFGVPYNPQSQGVVEAMN HHLKNQIDRLRDQAVSIETVVLMAT HCMNFKRRGGIGDMTPAERLVNMIT TEQEIQFFQAKNLKFQNFQVYYREG RDQLWKGPGELLWKGEGAVIIKVG TEIKVVPRRKAKIIRHYGGGKGLDCS ADMEDTRQAREMAQSD	

POL_HV1A2	P03369	Вирус иммунодефицита человека 1 типа	MGARASVLSGGELDKWEKIRLRPG GKKKYKLKHIVWASRELERFAVNP GLLETSEGCRQILGQLQPSLQTGSEE LRSLYNTVATLYCVHQRIDVKDTKE ALEKIEEEQNKSKKKAQQAAAAAG TGNSSQVSQNYPIVQNLQGQMVHQ AISPRTLNAWVKVVEEKAFSPEVIPM FSALSEGATPQDLNTMLNTVGGHQA AMQMLKETINEEAAEWDRVHPVHA GPIAPGQMREPRGSDIAGTTSTLQEQ IGWMTNNPPIPVGEIYKRWIILGLNKI VRMYSPTSILDIRQGPKEPFRDYVDR FYKTLRAEQASQDVKNWMTETLLV QNANPDCKTILKALGPAATLEEMMT ACQGVGGPGHKARVLAEAMSQVTN PANIMMQRGNFRNQRKTVKCFNCG KEGHIAKNCRAPRKKGCWRCGREG HQMKDCTERQANFLREDLAFLQGK AREFSSEQTRANSPTRRELQVWGGE NNSLSEAGADRQGTVSFNFPQITLW QRPLVTIRIGGQLKEALLDTGADDT VLEEMNLPGKWKPKMIGGIGGFIKV RQYDQIPVEICGHKAIGTVLVGPTPV NIIGRNLLTQIGCTLNFPISPIETVPVK LKPGMDGPKVKQWPLTEEKIKALVE ICTEMEKEGKISKIGPENPYNTPVFAI KKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDV GDAYFSVPLDKDFRKYTAFTIPSINN ETPGIRYQYNVLPQGWKGSPAIFQSS MTKILEPFRKQNPDIVIYQYMDDLY VGSDLEIGQHRTKIEELRQHLLRWG FTTPDKKHQKEPPFLWMGYELHPDK WTVQPIMLPEKDSWTVNDIQKLVG KLNWASQIYAGIKVKQLCKLLRGTK	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659

			ALTEVIPLTEEAELELAENREILKEPV HEVYYDPSKDLVAEIQKQGQGQWT YQIYQEPFKNLKTGKYARMRGAHT NDVKQLTEAVQKVSTESIVIWGKIPK FKLPIQKETWEAWWMEYWQATWIP EWEFVNTPPLVKLWYQLEKEPIVGA ETFYVDGAANRETKLGKAGYVTDR GRQKVVSIADTTNQKTELQAIHLAL QDSGLEVNIVTDSQYALGIIQAQPDK SESELVSQIIEQLIKKEKVYLAWVPA HKGIGGNEQVDKLVSAGIRKVLFLN GIDKAQEEHEKYHSNWRAMASDFN LPPVVAKEIVASCDKCQLKGEAMHG QVDCSPGIWQLDCTHLEGKIILVAV HVASGYIEAEVIPAETGQETAYFLLK LAGRWPVKTIHTDNGSNFTSTTVKA ACWWAGIKQEFGIPYNPQSQGVVES MNNELKKIIGQVRDQAEHLKTAVQ MAVFIHNFKRKGGIGGYSAGERIVDI IATDIQTKELQKQITKIQNFRVYYRD NKDPLWKGPAKLLWKGEGAVVIQD NSDIKVVPRRKAKIIRDYGKQMAGD DCVASRQDED	

POL_FIVPE	P16088	Вирус иммунодефицита кошек	KEFGKLEGGASCSPSESNAASSNAICTSNGGETIGFVNYNKVGTTTTLEKRPEILIFVNGYPIKFLLDTGADITILNRRDFQVKNSIENGRQNMIGVGGGKRGTNYINVHLEIRDENYKTQCIFGNVCVLEDNSLIQPLLGRDNMIKFNIRLVMAQISDKIPVVKVKMKDPNKGPQIKQWPLTNEKIEALTEIVERLEKEGKVKRADSNNPWNTPVFAIKKKSGKWRMLIDFRELNKLTEKGAEVQLGLPHPAGLQIKKQVTVLDIGDAYFTIPLDPDYAPYTAFTLPRKNNAGPGRRFVWCSLPQGWILSPLIYQSTLDNIIQPFIRQNPQLDIYQYMDDIYIGSNLSKKEHKEKVEELRKLLLWWGFETPEDKLQEEPPYTWMGYELHPLTWTIQQKQLDIPEQPTLNELQKLAGKINWASQAIPDLSIKALTNMMRGNQNLNSTRQWTKEARLEVQKAKKAIEEQVQLGYYDPSKELYAKLSLVGPHQISYQVYQKDPEKILWYGKMSRQKKKAENTCDIALRACYKIREESIIRIGKEPRYEIPTSREAWESNLINSPYLKAPPPEVEYIHAALNIKRALSMIKDAPIPGAETWYIDGGRKLGKAAKAAYWTDTGKWRVMDLEGSNQKAEIQALLLALKAGSEEMNIITDSQYVINIILQQPDMMEGIWQEVLEELEKKTAIFIDWVPGHKGIPGNEEVDKLCQTMMIIEGDGILDKRSEDAGYDLLAAKEIHLLPGEVKVIPTGVKLMLPKGYWGLIIGKSSIGSKGLDVLGGVIDEGYRGEIGVIMINVSRKSITLMERQKIAQLIILPCKHEVLEQGKVVMDSERGDNGYGSTGVFSSWVDRIEEAEINHEKFHSDPQYLRTEFNLPKMVAEEIRRKCPVCRIIG	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			EQVGGQLKIGPGIWQMDCTHFDGKI ILVGIHVESGYIWAQIISQETADCTV KAVLQLLSAHNVTELQTDNGPNFKN QKMEGVLNYMGVKHKFGIPGNPQS QALVENVNHTLKVWIQKFLPETTSL DNALSLAVHSLNFKRRGRIGGMAPY ELLAQQESLRIQDYFSAIPQKLQAQ WIYYKDQKDKKWKGPMRVEYWGQ GSVLLKDEEKGYFLIPRRHIRRVPEP CALPEGDE	

POL_EIAVY	P03371	Вирус инфекционной анемии лошадей	TAWTFLKAMQKCSKKREARGSREA PETNFPDTTEESAQQICCTRDSSDSK SVPRSERNKKGIQCQGEGSSRGSQPG QFVGVTYNLEKRPTTIVLINDTPLNV LLDTGADTSVLTTAHYNRLKYRGR KYQGTGIIGVGGNVETFSTPVTIKKK GRHIKTRMLVADIPVTILGRDILQDL GAKLVLAQLSKEIKFRKIELKEGTM GPKIPQWPLTKEKLEGAKETVQRLL SEGKISEASDNNPYNSPIFVIKKRSGK WRLLQDLRELNKTVQVGTEISRGLP HPGGLIKCKHMTVLDIGDAYFTIPLD PEFRPYTAFTIPSINHQEPDKRYVWK CLPQGFVLSPYIYQKTLQEILQPFRER YPEVQLYQYMDDLFVGSNGSKKQH KELIIELRAILQKGFETPDDKLQEVPP YSWLGYQLCPENWKVQKMQLDMV KNPTLNDVQKLMGNITWMSSGVPG LTVKHIAATTKGCLELNQKVIWTEE AQKELEENNEKIKNAQGLQYYNPEE EMLCEVEITKNYEATYVIKQSQGIL WAGKKIMKANKGWSTVKNLMLLL QHVATESITRVGKCPTFKVPFTKEQV MWEMQKGWYYSWLPEIVYTHQVV HDDWRMKLVEEPTSGITIYTDGGKQ NGEGIAAYVTSNGRTKQKRLGPVTH QVAERMAIQMALEDTRDKQVNIVT DSYYCWKNITEGLGLEGPQNPWWPI IQNIREKEIVYFAWVPGHKGIYGNQL ADEAAKIKEEIMLAYQGTQIKEKRD EDAGFDLCVPYDIMIPVSDTKIIPTDV KIQVPPNSFGWVTGKSSMAKQGLLI NGGIIDEGYTGEIQVICTNIGKSNIKLI EGQKFAQLIILQHHSNSRQPWDENKI SQRGDKGFGSTGVFWVENIQEAQDE	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			HENWHTSPKILARNYKIPLTVAKQIT QECPHCTKQGSGPAGCVMRSPNHW QADCTHLDNKIILHFVESNSGYIHAT LLSKENALCTSLAILEWARLFSPKSL HTDNGTNFVAEPVVNLLKFLKIAHT TGIPYHPESQGIVERANRTLKEKIQS HRDNTQTLEAALQLALITCNKGRES MGGQTPWEVFITNQAQVIHEKLLLQ QAQSSKKFCFYKIPGEHDWKGPTRV LWKGDGAVVVNDEGKGIIAVPLTRT KLLIKPN	

POL_BIV29	P19560	Вирус иммунодефицита крупного рогатого скота	MKRRELEKKLRKVRVTPQQDKYYTIGNLQWAIRMINLMGIKCVCDEECSAAEVALIITQFSALDLENSPIRGKEEVAIKNTLKVFWSLLAGYKPESTETALGYWEAFTYREREARADKEGEIKSIYPSLTQNTQNKKQTSNQTNTQSLPAITTQDGTPRFDPDLMKQLKIWSDATERNGVDLHAVNILGVITANLVQEEIKLLLNSTPKWRLDVQLIESKVREKENAHRTWKQHHPEAPKTDEIIGKGLSSAEQATLISVECRETFRQWVLQAAMEVAQAKHATPGPINIHQGPKEPYTDFINRLVAALEGMAAPETTKEYLLQHLSIDHANEDCQSILRPLGPNTPMEKKLEACRVVGSQKSKMQFLVAAMKEMGIQSPIPAVLPHTPEAYASQTSGPEDGRRCYGCGKTGHLKRNCKQQKCYHCGKPGHQARNCRSKNREVLLCPLWAEEPTTEQFSPEQHEFCDPICTPSYIRLDKQPFIKVFIGGRWVKGLVDTGADEVVLKNIHWDRIKGYPGTPIKQIGVNGVNVAKRKTHVEWRFKDKTGIIDVLFSDTPVNLFGRSLLRSIVTCFTLLVHTEKIEPLPVKVRGPGPKVPQWPLTKEKYQALKEIVKDLLAEGKISEAAWDNPYNTPVFVIKKKGTGRWRMLMDFRELNKITVKGQEFSTGLPYPPGIKECEHLTAIDIKDAYFTIPLHEDFRPFTAFSVVPVNREGPIERFQWNVLPQGWVCSPAIYQTTTQKIIENIKKSHPDVMLYQYMDDLLIGSNRDDHKQIVQEIRDKLGSYGFKTPDEKVQEERVKWIGFELTPKKWRFQPRQLKIKNPLTVNELQQLVGNCVWVQPEVKIPLYPLTDLLRDKTNLQEKIQLTPEAIKCVEEFNLKLKDPEWK	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661

			DRIREGAELVIKIQMVPRGIVFDLLQ DGNPIWGGVKGLNYDHSNKIKKILR TMNELNRTVVIMTGREASFLLPGSSE DWEAALQKEESLTQIFPVKFYRHSC RWTSICGPVRENLTTYYTDGGKKGK TAAAVYWCEGRTKSKVFPGTNQQA ELKAICMALLDGPPKMNIITDSRYAY EGMREEPETWAREGIWLEIAKILPFK QYVGVGWVPAHKGIGGNTEADEGV KKALEQMAPCSPPEAILLKPGEKQN LETGIYMQGLRPQSFLPRADLPVAIT GTMVDSELQLQLLNIGTEHIRIQKDE VFMTCFLENIPSATEDHERWHTSPDI LVRQFHLPKRIAKEIVARCQECKRTT TSPVRGTNPRGRFLWQMDNTHWNK TIIWVAVETNSGLVEAQVIPEETALQ VALCILQLIQRYTVLHLHSDNGPCFT AHRIENLCKYLGITKTTGIPYNPQSQ GVVERAHRDLKDRLAAYQGDCETV EAALSLALVSLNKKRGGIGGHTPYEI YLESEHTKYQDQLEQQFSKQKIEKW CYVRNRRKEWKGPYKVLWDGDGA AVIEEEGKTALYPHRHMRFIPPPDSD IQDGSS	

A0A142BKH1_ALV	A0A142BKH1	Вирус лейкоза и саркомы птиц	TVALHLAIPLKWKPDHTPVWIDQWP LPEGKLVALTQLVEKELQLGHIEPSL SCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRG WPLMVLDLKDCFFSIPLAEQDREAF AFTLPSVNNQAPARRFQWKVLPQG MTCSPTICQLVVGQVLEPLRLKHPSL RMLHYMDDLLLAASSHDGLEAAGE EVISTLERAGFTISPDKIQREPGVQYL GYKLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPPRLMGPF YEQLRGSDPNEAREWNLDMKMAW REIVQLSTTAALERWDPALPLEGAV ARCEQGAIGVLGQGLSTHPRPCLWL FSTQPTKAFTAWLEVLTLLITKLRAS AVRTFGKEVDVLLLPACFREDLPLPE GILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSS THKGVVVWREGPRWEIKEIADLGAS VQQLEARAVAMALLLWPTTPTNVV TDSAFVAKMLLKMGQEGVPSTAAA FILEDALSQRSAMAAVLHVRSHSEV PGFFTEGNDVADSQATFQAYPLREA KDLHTALHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAGVNPRGL GPLQIWQTDFTLEPRMAPRSWLAVT VATASSAIVVTQHGRVTSVAARHH WATAIAVLGRPKAIKTDNGSCFTSK STREWLARWGIAHTTGIPGNSQGQA MVERANRLLKDKIRVLAEGDGFMK RIPTGKQGELLAKAMYALNHFERGE NTKTPIQKHWRPTVLTEGPPVKIRIE TGEWEKGWNVLVWGRGYAAVKNR DTDKIIWVPSRKVKPDITQKDELTKK DEASPLFAGISDWAPWKGEQEGL	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

Таблица 32. Описания сигнатур InterPro, присутствующих в обратных транскриптазах из таблицы 30 (мономерных вирусных RT) и из таблицы 31 (димерных вирусных RT)

Сигнатура	База данных	Краткое название	Описание
cd01645	CDD	*RT_Rtv*	RT_Rtv: обратные транскриптазы (RT) ретровирусов (Rtv). RT катализируют превращение однонитевой РНК в двухнитевую вирусную ДНК для интеграции в хромосомы хозяина. Белки этого подсемейства содержат длинные концевые повторы (LTR) и представляют собой многофункциональные ферменты с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). Вирусный РНК-геном входит в цитоплазму в качестве части нуклеопротеинового комплекса, и в цитоплазме организуется процесс обратной транскрипции с образованием линейного ДНК-дуплекса посредством сложного ряда стадий. Этот ДНК-дуплекс является коллинеарным к своей РНК-матрице, но содержит концевые дупликации, известные как LTR, которые отсутствуют в вирусной РНК. Было высказано предположение, что для создания LTR необходимы два специализированных

			переключателя матрицы, известных в качестве реакций переноса нити или "прыжков". [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]
cd03715	CDD	*RT_ZFREV_ подобные*	RT_ZFREV_подобные: подсемейство обратных транскриптаз (RT), обнаруженное в последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа,

			содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ позволяет предположить, что ZFERV принадлежит к отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
PF00078	Pfam	*RVT_1*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. [PMID: 1698615]
IPR000477	InterPro	*RT_домен*	Применение РНК-матрицы для получения ДНК с целью интеграции в геном хозяина и использования клетки-хозяина является стратегией, применяемой при репликации ретроидных элементов, таких как ретровирусы и бактериальные ретроны. Фермент, катализирующий полимеризацию, представляет собой РНК-направленную ДНК-полимеразу или обратную транскриптазу (RT) (2.7.7.49). Обратная транскриптаза

			встречается в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II [PMID: 12758069], бактериальные msDNA, гепаднавирусы и каулимовирусы. Ретровирусная обратная транскриптаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Обнаружение ретроэлементов у прокариот поднимает интригующие вопросы, касающиеся их роли у бактерий, происхождения и эволюции обратных транскриптаз, а также того, являются ли бактериальные обратные транскриптазы старшими, чем эукариотические обратные транскриптазы [PMID: 8828137]. Было определено несколько кристаллических структур домена обратной транскриптазы (RT) [PMID: 1377403].

IPR043502	InterPro	*Суперсемейство ДНК/РНК-полимераз*	Эта запись представляет суперсемейство ДНК/РНК-полимераз, которое включает ДНК-полимеразу I, обратную транскриптазу, РНК-полимеразу T7, ДНК-полимеразу, обходящую повреждения (Y-семейство), РНК-зависимую РНК-полимеразу и РНК-зависимую РНК-полимеразу dsRNA-фага. Данные ферменты имеют сходную белковую укладку в своем активном центре, которая напоминает субдомен "ладонь" полимераз в форме правой руки. [PMID: 26931141]
SSF56672	Суперсемейство	*ДНК/РНК-полимеразы*	Данное суперсемейство включает ДНК-полимеразы и РНК-полимеразы.
PF06817	Pfam	*RVT_большой палец*	Данный домен известен в качестве домена "большой палец". Он состоит из четырехспирального пучка [PMID: 1377403].
IPR010661	InterPro	*RVT_большой палец*	Данный домен известен в качестве домена "большой палец". Он состоит из четырехспирального пучка. Обратная транскриптаза превращает вирусный РНК-геном в двухнитевую вирусную ДНК. Обратная транскриптаза часто встречается в полипротеине с интеразой, рибонуклеазой H и/или протеазой, которая отщепляется до того, как фермент начинает действовать. Влияние антиретровирусного лечения на

			первые 400 аминокислот обратной транскриптазы HIV является хорошим. Однако мало что известно о влиянии антиретровирусных препаратов на С-концевые домены Pol, которые включают домен "большой палец", соединительный домен и домен РНКазы H. Существующие данные свидетельствуют о том, что это могут быть весьма консервативные домены. [PMID: 1377403, PMID: 18335052]
PF06815	Pfam	*RVT_ соединительный*	Данный домен известен в качестве соединительного домена. Данный домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
IPR010659	InterPro	*RVT_ соединительный*	Данный домен известен в качестве соединительного домена. Данный домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
cd03715	CDD	*RT_ZFREV_ подобные*	RT_ZFREV_подобные: подсемейство обратных транскриптаз (RT), обнаруженное в последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая

			ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ позволяет предположить, что ZFERV принадлежит к отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]

Эндонуклеазный домен

В некоторых вариантах осуществления полипептид содержит эндонуклеазный домен (например, гетерологичный эндонуклеазный домен). В определенных вариантах осуществления эндонуклеазный/ДНК-связывающий домен ретротранспозона APE-типа, эндонуклеазный домен ретротранспозона RLE-типа или эндонуклеазный домен ретротранспозона PLE-типа может использоваться или может быть модифицирован (например, посредством вставки, делеции или замены одного или нескольких остатков) в системе Gene Writer, описанной в данном документе. В некоторых вариантах осуществления эндонуклеазный домен или эндонуклеазный/ДНК-связывающий домен изменен по сравнению с его природной последовательностью таким образом, что он

характеризуется измененной частотой использования кодонов, например, улучшенной для применения в клетках человека. В некоторых вариантах осуществления эндонуклеазный элемент представляет собой гетерологичный эндонуклеазный элемент, такой как нуклеаза Fok1, Cas9, никаза Cas9, эндонуклеаза, подобная ферменту рестрикции II типа (нуклеаза RLE-типа), или другая эндонуклеаза RLE-типа (также известная как REL). В некоторых вариантах осуществления гетерологичный эндонуклеазный домен расщепляет обе нити ДНК и образует двухнитевые разрывы. В некоторых вариантах осуществления гетерологичная эндонуклеаза обладает никазной активностью и не образует двухнитевые разрывы. Аминокислотная последовательность эндонуклеазного домена системы Gene Writer, описанной в данном документе, может являться на по меньшей мере приблизительно 50%, на по меньшей мере приблизительно 60%, на по меньшей мере приблизительно 70%, на по меньшей мере приблизительно 80%, на по меньшей мере приблизительно 85%, на по меньшей мере приблизительно 90%, на по меньшей мере приблизительно 95%, на по меньшей мере приблизительно 96%, на по меньшей мере приблизительно 97%, на по меньшей мере приблизительно 98%, на по меньшей мере приблизительно 99% идентичной аминокислотной последовательности эндонуклеазного домена ретротранспозона, последовательность ДНК которого указана в таблице X, Z1, Z2, 3A или 3B. Эндонуклеазные домены могут быть идентифицированы, например, на основании гомологии по отношению к другим известным эндонуклеазным доменам с помощью инструментов, таких как средство поиска основного локального выравнивания (BLAST). В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой Cas9 или никазу Cas9 или их функциональный фрагмент. В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой Fok1 или ее функциональный фрагмент. В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой резольвазу для структур Холлидея или ее гомолог, такой как фермент, разрешающий структуры Холлидея, из Sulfolobus solfataricus -- Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой эндонуклеазу из большого фрагмента белка сплайсосомы, такого как Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). Например, полипептид Gene Writer, описанный в данном документе, может содержать домен обратной транскриптазы из ретротранспозона APE-, RLE- или PLE-типа и эндонуклеазный домен, который содержит Fok1 или его функциональный фрагмент. В еще нескольких других вариантах осуществления гомологичные эндонуклеазные домены модифицированы, например, посредством сайт-специфической мутации, с изменением активности ДНК-эндонуклеазы. В еще нескольких других вариантах осуществления домены эндонуклеазы модифицированы с устранением какой-либо скрытой специфичности по отношению к последовательности ДНК.

В некоторых вариантах осуществления полипептид Gene Writer обладает функцией расщепления сайта-мишени ДНК посредством эндонуклеазного домена. В некоторых вариантах осуществления эндонуклеазный домен также является ДНК-связывающим

доменом. В некоторых вариантах осуществления эндонуклеазный домен также представляет собой домен связывания нуклеиновой кислоты-матрицы (например, РНК-матрицы). Например, в некоторых вариантах осуществления полипептид содержит эндонуклеазный домен, ассоциированный с CRISPR, который связывает РНК-матрицу, содержащую gRNA, связывает последовательность ДНК-мишени (например, комплементарную части gRNA) и разрезает последовательность ДНК-мишени. В определенных вариантах осуществления эндонуклеазный/ДНК-связывающий домен ретротранспозона APE-типа или эндонуклеазный домен ретротранспозона RLE-типа может использоваться или может быть модифицирован (например, посредством вставки, делеции или замены одного или нескольких остатков) в системе Gene Writer, описанной в данном документе. В некоторых вариантах осуществления эндонуклеазный домен или эндонуклеазный/ДНК-связывающий домен изменен по сравнению с его природной последовательностью таким образом, что он характеризуется измененной частотой использования кодонов, например, улучшенной для применения в клетках человека. В некоторых вариантах осуществления эндонуклеазный элемент представляет собой гетерологичный эндонуклеазный элемент, такой как нуклеаза Fok1, эндонуклеаза, подобная ферменту рестрикции 1 II типа (нуклеаза RLE-типа), или другая эндонуклеаза RLE-типа (также известная как REL). В некоторых вариантах осуществления гетерологичная эндонуклеаза обладает никазной активностью и не образует двухнитевые разрывы. Аминокислотная последовательность эндонуклеазного домена системы Gene Writer, описанной в данном документе, может являться на по меньшей мере приблизительно 50%, на по меньшей мере приблизительно 60%, на по меньшей мере приблизительно 70%, на по меньшей мере приблизительно 80%, на по меньшей мере приблизительно 85%, на по меньшей мере приблизительно 90%, на по меньшей мере приблизительно 95%, на по меньшей мере приблизительно 96%, на по меньшей мере приблизительно 97%, на по меньшей мере приблизительно 98%, на по меньшей мере приблизительно 99% идентичной аминокислотной последовательности эндонуклеазного домена ретротранспозона, последовательность ДНК которого указана в таблице 1, 2, 3A или 3B. Эндонуклеазные домены могут быть идентифицированы, например, на основании гомологии по отношению к другим известным эндонуклеазным доменам с помощью таких инструментов, как средство поиска основного локального выравнивания (BLAST). В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой Fok1 или ее функциональный фрагмент. В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой резольвазу для структур Холлидея или ее гомолог, такой как фермент, разрешающий структуры Холлидея, из Sulfolobus solfataricus -- Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой эндонуклеазу из большого фрагмента белка сплайсосомы, такого как Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). В определенных вариантах осуществления гетерологичная эндонуклеаза получена из белка, ассоциированного с CRISPR, например, Cas9. В определенных вариантах

осуществления гетерологичная эндонуклеаза сконструирована таким образом, что она обладает только активностью расщепления однонитевой ДНК, например, только никазной активностью, например, является никазой Cas9. Например, полипептид Gene Writer, описанный в данном документе, может содержать домен обратной транскриптазы из ретротранспозона APE- или RLE-типа и эндонуклеазный домен, который содержит Fok1 или его функциональный фрагмент. В еще нескольких других вариантах осуществления гомологичные эндонуклеазные домены модифицированы, например, посредством сайт-специфической мутации, с изменением активности ДНК-эндонуклеазы. В еще нескольких других вариантах осуществления эндонуклеазные домены модифицированы с устранением какой-либо скрытой специфичности по отношению к последовательности ДНК.

В некоторых вариантах осуществления эндонуклеазный домен обладает никазной активностью и не образует двухнитевые разрывы. В некоторых вариантах осуществления эндонуклеазный домен образует однонитевые разрывы с большей частотой, чем двухнитевые разрывы, например, по меньшей мере 90%, 95%, 96%, 97%, 98% или 99% разрывов являются однонитевыми разрывами, или менее 10%, 5%, 4%, 3%, 2% или 1% разрывов являются двухнитевыми разрывами. В некоторых вариантах осуществления эндонуклеаза практически не образует двухнитевых разрывов. В некоторых вариантах осуществления эндонуклеаза не образует двухнитевые разрывы на выявляемых уровнях.

В некоторых вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию; например, в некоторых вариантах осуществления эндонуклеазный домен разрезает сайт-мишень геномной ДНК рядом с сайтом изменения в нити, которая будет удлинена записывающим доменом. В некоторых вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию, и не обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, не подлежащей редактированию. Например, в случае, если полипептид содержит эндонуклеазный домен, ассоциированный с CRISPR, обладающий никазной активностью, и не образующий двухнитевых разрывов, то в некоторых вариантах осуществления указанный эндонуклеазный домен, ассоциированный с CRISPR, вносит однонитевые разрывы в ДНК сайта-мишени в нити, содержащей PAM-сайт (например, и не вносит однонитевые разрывы в ДНК сайта-мишени в нити, не содержащей PAM-сайт).

В некоторых других вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию, и нити, не подлежащей редактированию. Не желая ограничиваться какой-либо теорией, полагают, что после полимеризации (например, обратной транскрипции) записывающего домена (например, домена RT) полипептида, описанного в данном документе, из последовательности гетерологичного объекта нуклеиновой кислоты-матрицы (например, РНК-матрицы), механизм репарации клеточной ДНК должен исправить однонитевой разрыв в нити ДНК,

подлежащей редактированию. ДНК сайта-мишени теперь содержит две разные последовательности для нити ДНК, подлежащей редактированию: одна соответствует исходной геномной ДНК, и вторая соответствует той, которая полимеризована из последовательности гетерологичного объекта. Считается, что две разные последовательности уравновешиваются друг с другом, где сначала первая из них гибридизируется с нередактируемой нитью, а затем другая, и то, какую из них механизм репарации клеточной ДНК включает в свой исправленный сайт-мишень, считается случайным. Не желая ограничиваться какой-либо теорией, полагают, что введение дополнительного однонитевого разрыва в нередактируемую нить может привести к тому, что механизм репарации клеточной ДНК будет чаще использовать последовательность на основе последовательности гетерологичного объекта, чем исходную геномную последовательность. В некоторых вариантах осуществления дополнительный однонитевой разрыв расположен на расстоянии по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, или 150 нуклеотидов с 5'- или 3'-стороны от модификации сайта-мишени (например, вставки, делеции или замены) или однонитевого разрыва на подлежащей редактированию нити.

В качестве альтернативы или дополнительно, не желая ограничиваться какой-либо теорией, полагают, что дополнительный однонитевой разрыв в нередактируемой нити может способствовать синтезу второй нити. В некоторых вариантах осуществления, где полипептид Gene Writer вставил или заменил часть редактируемой нити, необходим синтез новой последовательности, соответствующей вставке/замене в нередактируемой нити.

В некоторых вариантах осуществления полипептид содержит один домен, обладающий эндонуклеазной активностью (например, один эндонуклеазный домен), и указанный домен вносит однонитевой разрыв как в нити, подлежащей редактированию, так и в нити, не подлежащей редактированию. Например, в таком варианте осуществления эндонуклеазный домен может являться эндонуклеазным доменом, ассоциированным с CRISPR, и нуклеиновая кислота-матрица (например, РНК-матрица) содержит gRNA, которая направляет внесение однонитевого разрыва в нити, подлежащей редактированию, и дополнительную gRNA, которая направляет внесение однонитевого разрыва в нити, не подлежащей редактированию. В некоторых вариантах осуществления полипептид содержит множество доменов, обладающих эндонуклеазной активностью, и первый эндонуклеазный домен вносит однонитевой разрыв в нить, подлежащую редактированию, и второй эндонуклеазный домен вносит однонитевой разрыв в нить, не подлежащую редактированию (необязательно первый эндонуклеазный домен не вносит (например, не может вносить) однонитевой разрыв в нить, не подлежащую редактированию, а второй эндонуклеазный домен не вносит (например, не может вносить) однонитевой разрыв в нить, подлежащую редактированию).

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в первую нить и вторую нить. В некоторых вариантах осуществления однонитевые разрывы в первой и второй нитях происходят в одном и том же положении в

сайте-мишени, но в противоположных нитях. В некоторых вариантах осуществления однонитевой разрыв второй нити происходит в смещенном местоположении, например, выше или ниже первого однонитевого разрыва. В некоторых вариантах осуществления эндонуклеазный домен образует делецию сайта-мишени, если однонитевой разрыв второй нити находится выше однонитевого разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен образует дупликацию сайта-мишени, если однонитевой разрыв второй нити находится ниже однонитевого разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен не образует дупликации и/или делеции, если однонитевые разрывы первой и второй нитей находятся в одном и том же положении в сайте-мишени (например, как описано в Gladyshev and Arkhipova Gene 2009, включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления эндонуклеазный домен обладает измененной активностью, зависящей от конформации белка или статуса связывания с РНК, например, что способствует внесению однонитевого разрыва в первую или вторую нить (например, как описано в Christensen et al. PNAS 2006; включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен содержит хоминг-эндонуклеазу или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу из семейств LAGLIDADG, GIY-YIG, HNH, His-Cys-бокса или PD-(D/E) XK или ее функциональный фрагмент или вариант, например, которая содержит консервативные аминокислотные мотивы, например, которые указаны в названиях семейств. В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу или ее фрагмент, выбранный, например, из I-SmaMI (UniProt F7WD42), I-SceI (UniProt P03882), I-AniI (UniProt P03880), I-DmoI (UniProt P21505), I-CreI (UniProt P05725), I-TevI (UniProt P13299), I-OnuI (UniProt Q4VWW5) или I-BmoI (UniProt Q9ANR6). В некоторых вариантах осуществления мегануклеаза является мономерной в естественных условиях, например, I-SceI, I-TevI, или димерной, например, I-CreI, в своей функциональной форме. Например, мегануклеазы LAGLIDADG с одной копией мотива LAGLIDADG обычно образуют гомодимеры, тогда как представители с двумя копиями мотива LAGLIDADG обычно встречаются в виде мономеров. В некоторых вариантах осуществления мегануклеаза, которая обычно образуется в виде димера, экспрессируется в виде слитой молекулы, например, две субъединицы экспрессируются в виде одной ORF и необязательно соединены линкером, например, в случае слитого димера I-CreI (Rodriguez-Fornes et al. Gene Therapy 2020; включенный в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления мегануклеаза или ее функциональный фрагмент изменены для способствования никазной активности в отношении одной нити двухнитевой молекулы ДНК, например, I-SceI (K122I и/или K223I) (Niu et al. J Mol Biol 2008), I-AniI (K227M) (McConnell Smith et al. PNAS 2009), I-DmoI (Q42A и/или K120M) (Molina et al. J

Biol Chem 2015). В некоторых вариантах осуществления мегануклеаза или ее функциональный фрагмент, обладающие таким предпочтением для однонитевого расщепления, применяются в качестве эндонуклеазного домена, например, с никазной активностью. В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу или ее функциональный фрагмент, которые в естественных условиях нацеливаются или сконструированы для нацеливания на сайт "безопасная гавань", например, I-CreI, нацеливающаяся на сайт SH6 (Rodriguez-Fornes et al., выше). В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу или ее функциональный фрагмент с толерантным к последовательности каталитическим доменом, например, I-TevI, распознающим минимальный мотив CNNNG (Kleinstiver et al. PNAS 2012). В некоторых вариантах осуществления каталитический домен, толерантный к последовательности-мишени, слит с ДНК-связывающим доменом, например, для управления активностью, например, посредством слияния I-TevI с: (i) "цинковыми пальцами" для создания Tev-ZFE (Kleinstiver et al. PNAS 2012), (ii) другими мегануклеазами для создания MegaTev (Wolfs et al. Nucleic Acids Res 2014) и/или (iii) Cas9 для создания TevCas9 (Wolfs et al. PNAS 2016).

В некоторых вариантах осуществления эндонуклеазный домен содержит рестрикционный фермент, например, рестрикционный фермент типа IIS или типа IIP. В некоторых вариантах осуществления эндонуклеазный домен содержит рестрикционный фермент типа IIS, например, FokI, или его фрагмент или вариант. В некоторых вариантах осуществления эндонуклеазный домен содержит рестрикционный фермент типа IIP, например, PvuII, или его фрагмент или вариант. В некоторых вариантах осуществления димерный рестрикционный фермент экспрессируется в виде слитой молекулы, так что он функционирует как одноцепочечный, например, слитый димер FokI (Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

Применение дополнительных эндонуклеазных доменов описано, например, в Guha and Edgell Int J Mol Sci 18(22):2565 (2017), которая включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит белок Cas9 *Streptococcus pyogenes* (SpCas9) или его функциональный фрагмент или вариант. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит модифицированный SpCas9. В вариантах осуществления модифицированный SpCas9 содержит модификацию, которая изменяет специфичность мотива, прилегающего к протоспейсеру (PAM). В вариантах осуществления PAM характеризуется специфичностью в отношении последовательности нуклеиновой кислоты 5'-NGT-3'. В вариантах осуществления модифицированный SpCas9 содержит одну или несколько аминокислотных замен, например, в одном или нескольких из положений L1111, D1135, G1218, E1219, A1322, R1335, например, выбранных из L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. В вариантах осуществления модифицированный SpCas9 содержит

аминокислотную замену T1337R и одну или несколько дополнительных аминокислотных замен, например, выбранных из L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q и T1337M, или соответствующие им аминокислотные замены. В вариантах осуществления модифицированный SpCas9 содержит: (i) одну или несколько аминокислотных замен, выбранных из D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q и T1337; и (ii) одну или несколько аминокислотных замен, выбранных из L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337R, T1337K, T1337N, T1337N, T1337Q и T1337M, или соответствующие им аминокислотные замены.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит домен Cas, например, домен Cas9. В вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит нуклеазоактивный домен Cas, Cas-никазный домен (nCas) или нуклеазонеактивный домен Cas (dCas). В вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит нуклеазоактивный домен Cas9, Cas9-никазный домен (nCas9) или нуклеазонеактивный домен Cas9 (dCas9). В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит домен Cas9 из Cas9 (например, dCas9 и nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas9 (например, dCas9 и nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas9 из S. pyogenes или S. thermophilus или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит последовательность Cas9, например, описанную в Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; включенном в данный документ посредством ссылки. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит HNH-нуклеазный субдомен и/или субдомен RuvC1 из Cas, например, Cas9, например, описанной в данном документе, или ее вариант. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит полипептид Cas (например, фермент) или его функциональный фрагмент. В вариантах осуществления полипептид Cas (например, фермент) выбран из Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (например, Csn1 или Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2,

Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, эффекторных белков Cas II типа, эффекторных белков Cas V типа, эффекторных белков Cas VI типа, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, сверхточного варианта Cas9 (HypaCas9), их гомологов, их модифицированных или сконструированных вариантов и/или их функциональных фрагментов. В вариантах осуществления Cas9 содержит одну или несколько замен, например, выбранных из H840A, D10A, P475A, W476A, N477A, D1125A, W1126A и D1127A. В вариантах осуществления Cas9 содержит одну или несколько мутаций в положениях, выбранных из D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987, например, одну или несколько замен, выбранных из D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A и/или D986A. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит последовательность Cas (например, Cas9) из Corynebacterium ulcerans, Corynebacterium diphtheria, Spiroplasma syrphidicola, Prevotella intermedia, Spiroplasma taiwanense, Streptococcus iniae, Belliella baltica, Psychroflexus torquis, Streptococcus thermophilus, Listeria innocua, Campylobacter jejuni, Neisseria meningitidis, Streptococcus pyogenes или Staphylococcus aureus или ее фрагмент или вариант.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит домен Cpf1, например, содержащий одну или несколько замен, например, в положениях D917, E1006A, D1255 или любой их комбинации, например, выбранных из D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A и D917A/E1006A/D1255A.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, как описано в данном документе) содержит spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK или spCas9-LRVSQL.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит аминокислотную последовательность, указанную в таблице 37 ниже, или аминокислотную последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит аминокислотную последовательность, которая характеризуется не более чем 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, или 50 различиями (например, мутациями) по сравнению с любой из аминокислотных последовательностей, описанных в данном документе.

Таблица 37. Каждая из эталонных последовательностей включена посредством ссылки во всей своей полноте

Название	**Аминокислотная последовательность или эталонная последовательность**
Cas9 Streptococcus pyogenes	
Иллюстративный линкер	SGSETPGTSESATPES
Иллюстративный линкерный мотив	$(SGGS)_n$
Иллюстративный линкерный мотив	$(GGGS)_n$
Иллюстративный линкерный мотив	$(GGGGS)_n$
Иллюстративный линкерный мотив	$(G)_n$
Иллюстративный линкерный мотив	$(EAAAK)_n$
Иллюстративный линкерный мотив	$(GGS)_n$
Иллюстративный линкерный мотив	$(XP)_n$
Cas9 из Streptococcus pyogenes	Эталонная последовательность в NCBI: NC_002737.2 и эталонная последовательность в UniProt: Q99ZW2
Cas9 из Corynebacterium ulcerans	Эталонные последовательности в NCBI: NC_015683.1, NC_017317.1
Cas9 из Corynebacterium diphtheria	Эталонные последовательности в NCBI: NC_016782.1, NC_016786.1
Cas9 из Spiroplasma syrphidicola	Эталонная последовательность в NCBI: NC_021284.1
Cas9 из Prevotella intermedia	Эталонная последовательность в NCBI: NC_017861.1
Cas9 из Spiroplasma taiwanense	Эталонная последовательность в NCBI: NC_021846.1

Cas9 из Streptococcus iniae	Эталонная последовательность в NCBI: NC_021314.1
Cas9 из Belliella baltica	Эталонная последовательность в NCBI: NC_018010.1
Cas9 из Psychroflexus torquis	Эталонная последовательность в NCBI: NC_018721.1
Cas9 из Streptococcus thermophilus	Эталонная последовательность в NCBI: YP_820832.1
Cas9 из Listeria innocua	Эталонная последовательность в NCBI: NP_472073.1
Cas9 из Campylobacter jejuni	Эталонная последовательность в NCBI: YP_002344900.1
Cas9 из Neisseria meningitidis	Эталонная последовательность в NCBI: YP_002342100.1
dCas9 (D10A и H840A)	
Каталитически неактивная Cas9 (dCas9)	
Никаза Cas9 (nCas9)	
Каталитически активная Cas9	
CasY	((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1 белок CasY, ассоциированный с CRISPR [некультивируемая бактерия из группы Parcubacteria])
CasX	uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53
CasX	>tr|F0NH53|F0NH53_SULIR белок, ассоциированный с CRISPR, Casx OS=Sulfolobus islandicus (штамм REY15A) GN=SiRe_0771 PE=4 SV=1
CasX Deltaproteobacteria	
Cas12b/C2c1	((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG эндонуклеаза C2c1, ассоциированная с CRISPR OS=Alicyclobacillus acidoterrestris (штамм ATCC 49025/DSM 3922/ CIP 106132/NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1)
BhCas12b (Bacillus hisashii)	Эталонная последовательность в NCBI: WP_095142515
BvCas12b (Bacillus sp. V3-13)	Эталонная последовательность в NCBI: WP_101661451.1

Cpf1 Francisella novicida дикого типа	
Cpf1 D917A Francisella novicida	
Cpf1 E1006A Francisella novicida	
Cpf1 D1255A Francisella novicida	
Cpf1 D917A/E1006A Francisella novicida	
Cpf1 D917A/D1255A Francisella novicida	
Cpf1 E1006A/D1255A Francisella novicida	
Cpf1 D917A/E1006A Francisella novicida	
SaCas9	
SaCas9n	
PAM-связывающий SpCas9	
PAM-связывающий SpCas9n	
PAM-связывающий SpCas9-EQR	
PAM-связывающий SpCas9-VQR	
PAM-связывающий SpCas9-VRER	
PAM-связывающий SpCas9-VRQR	
SpyMacCas9	

В некоторых вариантах осуществления полипептид Gene Writing имеет эндонуклеазный домен, содержащий никазу Cas9, например, Cas9 H840A. В вариантах осуществления Cas9 H840A имеет следующую аминокислотную последовательность.

Никаза Cas9 (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR
KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII
HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD.

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, например, RT M-MLV дикого типа, например, содержащий следующую последовательность.

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS
IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI
SGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQ
GTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK
TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTA
PALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR
MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD
RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSS
LLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSR
YAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLI.

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, например, RT M-MLV, например, содержащий следующую последовательность:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL
DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG
QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQ
ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWP
PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL
LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT
DGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVY
TDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH
SAEARGNRMADQAARKAAITETPDTSTLL

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, содержащий последовательность из аминокислот 659-1329 NP_057933. В вариантах осуществления полипептид Gene Writing дополнительно содержит одну дополнительную аминокислоту на N-конце последовательности из аминокислот 659-1329 NP_057933, например, как показано ниже:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQD**
LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF
EWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDL
LLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLT
EARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPD
QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN
ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP
LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK
MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLS
IIHCPGHQKGHSAEARGNRMADQAARKAA

Сердцевинная RT (жирный шрифт), аннотирована согласно вышеуказанному

РНКаза H (подчеркнуто), аннотирована согласно вышеуказанному

В вариантах осуществления полипептид Gene Writing дополнительно содержит одну дополнительную аминокислоту на С-конце последовательности из аминокислот 659-1329 NP_057933. В вариантах осуществления полипептид Gene Writing содержит домен РНКазы H1 (например, аминокислоты 1178-1318 NP_057933).

В некоторых вариантах осуществления домен ретровирусной обратной транскриптазы, например, RT M-MLV, может содержать одну или несколько мутаций

относительно последовательности дикого типа, которые могут улучшать характеристики RT, например, термостабильность, процессивность и/или связывание с матрицей. В некоторых вариантах осуществления домен RT M-MLV содержит, по сравнению с приведенной выше последовательностью M-MLV (WT), одну или несколько мутаций, например, выбранных из D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, например, комбинацию мутаций, таких как D200N, L603W и T330P, необязательно дополнительно включая T306K и W313F. В некоторых вариантах осуществления RT M-MLV, используемая в данном документе, содержит мутации D200N, L603W, T330P, T306K и W313F. В вариантах осуществления мутантная RT M-MLV содержит следующую аминокислотную последовательность.

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS
IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI
SGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQ
GTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK
TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTA
PALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR
MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD
RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSS
LLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSR
YAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLI.

В некоторых вариантах осуществления полипептид Gene Writer может содержать линкер, например, пептидный линкер, например, линкер, описанный в таблице 7. В некоторых вариантах осуществления полипептид Gene Writer содержит гибкий линкер между эндонуклеазой и доменом RT, например, линкер, содержащий аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGSS. В некоторых вариантах осуществления домен RT полипептида Gene Writer может быть расположен с C-концевой стороны от эндонуклеазного домена. В некоторых вариантах осуществления домен RT полипептида Gene Writer может быть расположен с N-концевой стороны от эндонуклеазного домена.

Таблица 7. Иллюстративные линкерные последовательности

Аминокислотная последовательность
GGS
GGSGGS
GGSGGSGGS
GGSGGSGGSGGS

GGSGGSGGSGGSGGS
GGSGGSGGSGGSGGSGGS
GGGGS
GGGGSGGGGS
GGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS
GGG
GGGG
GGGGG
GGGGGG
GGGGGGG
GGGGGGGG
GSS
GSSGSS
GSSGSSGSS
GSSGSSGSSGSS
GSSGSSGSSGSSGSS
GSSGSSGSSGSSGSSGSS
EAAAK
EAAAKEAAAK
EAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK
PAP
PAPAP
PAPAPAP
PAPAPAPAP
PAPAPAPAPAP
PAPAPAPAPAPAP
GGSGGG
GGGGGS

GGSGSS
GSSGGS
GGSEAAAK
EAAAKGGS
GGSPAP
PAPGGS
GGGGSS
GSSGGG
GGGEAAAK
EAAAKGGG
GGGPAP
PAPGGG
GSSEAAAK
EAAAKGSS
GSSPAP
PAPGSS
EAAAKPAP
PAPEAAAK
GGSGGGGSS
GGSGSSGGG
GGGGGSGSS
GGGGSSGGS
GSSGGSGGG
GSSGGGGGS
GGSGGGEAAAK
GGSEAAAKGGG
GGGGGSEAAAK
GGGEAAAKGGS
EAAAKGGSGGG
EAAAKGGGGGS
GGSGGGPAP
GGSPAPGGG
GGGGGSPAP
GGGPAPGGS

PAPGGSGGG
PAPGGGGGS
GGSGSSEAAAK
GGSEAAAKGSS
GSSGGSEAAAK
GSSEAAAKGGS
EAAAKGGSGSS
EAAAKGSSGGS
GGSGSSPAP
GGSPAPGSS
GSSGGSPAP
GSSPAPGGS
PAPGGSGSS
PAPGSSGGS
GGSEAAAKPAP
GGSPAPEAAAK
EAAAKGGSPAP
EAAAKPAPGGS
PAPGGSEAAAK
PAPEAAAKGGS
GGGGSSEAAAK
GGGEAAAKGSS
GSSGGGEAAAK
GSSEAAAKGGG
EAAAKGGGGSS
EAAAKGSSGGG
GGGGSSPAP
GGGPAPGSS
GSSGGGPAP
GSSPAPGGG
PAPGGGGSS
PAPGSSGGG
GGGEAAAKPAP
GGGPAPEAAAK

EAAAKGGGPAP
EAAAKPAPGGG
PAPGGGEAAAK
PAPEAAAKGGG
GSSEAAAKPAP
GSSPAPEAAAK
EAAAKGSSPAP
EAAAKPAPGSS
PAPGSSEAAAK
PAPEAAAKGSS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA
GGGGSEAAAKGGGGS
EAAAKGGGGSEAAAK
SGSETPGTSESATPES
GSAGSAAGSGEF
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS

В некоторых вариантах осуществления полипептид Gene Writer содержит последовательность dCas9, содержащую мутацию D10A и/или H840A, например, следующую последовательность:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG
MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES
EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN

GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

В некоторых вариантах осуществления молекула РНК-матрицы для применения в системе содержит в направлении от 5' к 3' (1) спейсер gRNA; (2) каркас gRNA; (3) последовательность гетерологичного объекта; (4) 3'-концевой домен гомологии. В некоторых вариантах осуществления:

(1) Представляет собой спейсер Cas9 размером ~ 18-22 нуклеотида, например, 20 нуклеотидов

(2) Представляет собой каркас gRNA, содержащий одну или несколько шпилечных петель, например, 1, 2 или 3 петли для ассоциации матрицы с никазным доменом Cas9. В некоторых вариантах осуществления каркас gRNA в направлении от 5' к 3' несет последовательность
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
AAGTGGGACCGAGTCGGTCC.

(3) В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину, например, 7-74, например, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70 или 70-80 нуклеотидов или 80-90 нуклеотидов. В некоторых вариантах осуществления первое (наиболее близкое к 5'-концу) основание последовательности не представляет собой С.

(4) В некоторых вариантах осуществления 3'-концевой домен гомологии, который связывает праймирующую последовательность-мишень после того, как происходит однонитевой разрез, составляет, например, 3-20 нуклеотидов, например, 7-15 нуклеотидов, например, 12-14 нуклеотидов. В некоторых вариантах осуществления 3'-концевой домен гомологии характеризуется 40-60% содержанием GC.

Вторая gRNA, ассоциированная с системой, может способствовать обеспечению полной интеграции. В некоторых вариантах осуществления вторая gRNA может нацеливаться на местоположение, которое находится на расстоянии 0-200 нуклеотидов от однонитевого разрыва первой нити, например, 0-50, 50-100, 100-200 нуклеотидов от однонитевого разрыва первой нити. В некоторых вариантах осуществления вторая gRNA может связываться со своей последовательностью-мишенью только после осуществления редактирования, например, gRNA связывает последовательность, присутствующую в последовательности гетерологичного объекта, но не в исходной последовательности-мишени.

В некоторых вариантах осуществления описанная в данном документе система Gene Writing применяется для осуществления редактирования в клетках HEK293, K562, U2OS или HeLa. В некоторых вариантах осуществления система Gene Writing применяется для

осуществления редактирования в первичных клетках, например, первичных кортикальных нейронах мышей E18.5.

В некоторых вариантах осуществления домен обратной транскриптазы или RT (например, описанный в данном документе) содержит последовательность RT MoMLV или ее вариант. В вариантах осуществления последовательность RT MoMLV содержит одну или несколько мутаций, выбранных из D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S и K103L. В вариантах осуществления последовательность RT MoMLV содержит комбинацию мутаций, таких как D200N, L603W и T330P, необязательно дополнительно включая T306K и/или W313F.

В некоторых вариантах осуществления эндонуклеазный домен (например, описанный в данном документе) содержит nCAS9, например, содержащую мутацию H840A.

В некоторых вариантах осуществления последовательность гетерологичного объекта (например, из системы, описанной в данном документе) имеет длину приблизительно 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000 или больше нуклеотидов.

В некоторых вариантах осуществления домен RT и эндонуклеазный домен соединены гибким линкером, например, содержащим аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGSS.

В некоторых вариантах осуществления эндонуклеазный домен является N-концевым по отношению к домену RT. В некоторых вариантах осуществления эндонуклеазный домен является C-концевым по отношению к домену RT.

В некоторых вариантах осуществления система осуществляет включение последовательности гетерологичного объекта в сайт-мишень с помощью TPRT, например, как описано в данном документе.

В некоторых вариантах осуществления система или способ, описанные в данном документе, включают фермент или систему CRISPR, нацеленную на ДНК, описанную в публикациях заявок на патент США № 20200063126, 20190002889 или 20190002875 (каждая из которых включена в данный документ посредством ссылки во всей своей полноте), или их функциональный фрагмент или вариант. Например, в некоторых вариантах осуществления полипептид Gene Writer или эндонуклеаза Cas, описанные в данном документе, содержат последовательность полипептида согласно любой из заявок, упомянутых в данном абзаце, и в некоторых вариантах осуществления РНК-матрица или направляющая РНК содержит последовательность нуклеиновой кислоты согласно любой из заявок, упомянутых в данном абзаце.

Домен, связывающий нуклеиновую кислоту-матрицу

Полипептид Gene Writer, как правило, содержит области, способные ассоциировать с нуклеиновой кислотой-матрицей Gene Writer (например, РНК-матрицей). В некоторых вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу,

представляет собой РНК-связывающий домен. В некоторых вариантах осуществления РНК-связывающий домен представляет собой модульный домен, который может ассоциировать с молекулами РНК, содержащими специфические сигнатуры, например, структурные мотивы, например, вторичные структуры, присутствующие в 3'-UTR в ретротранспозонах, не содержащих LTR. В других вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу (например, РНК-связывающий домен), содержится в домене обратной транскрипции, например, компонент, полученный из обратной транскриптазы, имеет известную сигнатуру для предпочтения РНК, например, вторичные структуры, присутствующие в 3'-UTR в ретротранспозонах, не содержащих LTR. В других вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу (например, РНК-связывающий домен), содержится в ДНК-связывающем домене. Например, в некоторых вариантах осуществления ДНК-связывающий домен представляет собой белок, ассоциированный с CRISPR, который распознает структуру нуклеиновой кислоты-матрицы (например, РНК-матрицы), содержащей gRNA. В некоторых вариантах осуществления gRNA представляет собой короткую синтетическую РНК, состоящую из каркасной последовательности, которая участвует в связывании белка, ассоциированного с CRISPR, и определяемой пользователем нацеливающей последовательности из ~ 20 нуклеотидов для геномной мишени. Структура полной gRNA была описана в Nishimasu et al. Cell 156, P935-949 (2014). gRNA (также называемая sgRNA или единичная направляющая RNA) состоит из последовательностей, полученных из crRNA и tracrRNA, соединенных искусственной тетрапетлей. Последовательность crRNA можно разделить на направляющую (20 нуклеотидов) и повторяющуюся (12 нуклеотидов) области, тогда как последовательность tracrRNA можно разделить на антиповтор (14 нуклеотидов) и три структуры "стебель-петля" tracrRNA (Nishimasu et al. Cell 156, P935-949 (2014)). На практике последовательности направляющих РНК, как правило, разрабатывают таким образом, чтобы они имели длину 17-24 нуклеотида (например, 19, 20 или 21 нуклеотид) и были комплементарными по отношению к последовательности нуклеиновой кислоты-мишени. Специализированные генераторы и алгоритмы для создания gRNA являются коммерчески доступными для применения в разработке эффективных направляющих РНК. В некоторых вариантах осуществления gRNA содержит два компонента РНК из нативной системы CRISPR, например, crRNA и tracrRNA. Как хорошо известно из уровня техники, gRNA может также содержать химерную единичную направляющую РНК (sgRNA), содержащую последовательность как из tracrRNA (для связывания нуклеазы), так и из по меньшей мере одной crRNA (для направления нуклеазы к последовательности, предназначенной для редактирования/связывания). Также было продемонстрировано, что химически модифицированные sgRNA являются эффективными при использования с белками, ассоциированными с CRISPR; см., например, Hendel et al. (2015) Nature Biotechnol. 985-991. В некоторых вариантах осуществления gRNA содержит последовательность нуклеиновой кислоты, которая является комплементарной последовательности ДНК, ассоциированной с геном-мишенью. В некоторых вариантах осуществления полипептид

содержит ДНК-связывающий домен, содержащий белок, ассоциированный с CRISPR, который ассоциирует с gRNA, что позволяет ДНК-связывающему домену связываться с геномной последовательностью ДНК-мишени. В некоторых вариантах осуществления gRNA содержится в нуклеиновой кислоте-матрице (например, РНК-матрице), поэтому ДНК-связывающий домен также является доменом, связывающим нуклеиновую кислоту-матрицу. В некоторых вариантах осуществления полипептид обладает функцией связывания РНК в нескольких доменах, например, может связывать структуру gRNA в ДНК-связывающем домене, ассоциированном с CRISPR, и структуру 3'-UTR в домене обратной транскрипции, полученном из ретротранспозона, не содержащего LTR.

В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) содержит 3'-концевой домен гомологии с мишенью. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью расположен с 3'-стороны от последовательности гетерологичного объекта и комплементарен последовательности, прилегающей к сайту, подлежащему модификации с помощью системы, описанной в данном документе, или содержит не более 1, 2, 3, 4 или 5 несовпадений с последовательностью, комплементарной последовательности, прилегающей к сайту, подлежащему модификации с помощью системы/Gene Writer™. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью отжигается с сайтом-мишенью, что обеспечивает наличие сайта связывания и гидроксильной группы на 3'-конце для инициации TPRT с помощью полипептида Gene Writer. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью имеет длину 3-5, 5-10, 10-30, 10-25, 10-20, 10-19, 10-18, 10-17, 10-16, 10-15, 10-14, 10-13, 10-12, 10-11, 11-30, 11-25, 11-20, 11-19, 11-18, 11-17, 11-16, 11-15, 11-14, 11-13, 11-12, 12-30, 12-25, 12-20, 12-19, 12-18, 12-17, 12-16, 12-15, 12-14, 12-13, 13-30, 13-25, 13-20, 13-19, 13-18, 13-17, 13-16, 13-15, 13-14, 14-30, 14-25, 14-20, 14-19, 14-18, 14-17, 14-16, 14-15, 15-30, 15-25, 15-20, 15-19, 15-18, 15-17, 15-16, 16-30, 16-25, 16-20, 16-19, 16-18, 16-17, 17-30, 17-25, 17-20, 17-19, 17-18, 18-30, 18-25, 18-20, 18-19, 19-30, 19-25, 19-20, 20-30, 20-25, или 25-30 нуклеотидов, например, имеет длину 10-17, 12-16 или 12-14 нуклеотидов.

В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, может содержать gRNA (например, pegRNA). В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, может связываться с полипептидом Gene Writer™ посредством взаимодействия gRNA-части нуклеиновой кислоты-матрицы с доменом, связывающим нуклеиновую кислоту-матрицу, например, с РНК-связывающим доменом (например, с гетерологичным РНК-связывающим доменом). В некоторых вариантах осуществления гетерологичный РНК-связывающий домен представляет собой белок CRISPR/Cas, например, Cas9.

В некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, характеризуется частично раскрученной лентовидной структурой gRNA, связанной с ДНК-мишенью (например, как описано в Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484). Не желая

ограничиваться какой-либо теорией, полагают, что эта неканоническая структура облегчается вращением каждого шестого нуклеотида из гибрида РНК-ДНК. Таким образом, в некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, может допускать увеличенное количество несовпадений с сайтом-мишенью с некоторым интервалом, например, по каждому шестому основанию. В некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, характеризующуюся гомологией с сайтом-мишенью, может иметь неоднозначные положения с регулярными интервалами, например, по каждому шестому основанию, которые не требуют образования пары оснований с сайтом-мишенью.

gRNA с индуцируемой активностью

В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, содержит направляющую РНК (gRNA) с индуцируемой активностью. Индуцируемая активность может быть достигнута нуклеиновой кислотой-матрицей, например, РНК-матрицей, дополнительно содержащей (в дополнение к gRNA) блокирующий домен, где последовательность части блокирующего домена или его всего является по меньшей мере частично комплементарной части gRNA или ей всей. Таким образом, блокирующий домен способен гибридизироваться или по сути гибридизироваться с частью gRNA или с ней всей. В некоторых вариантах осуществления блокирующий домен и gRNA с индуцируемой активностью расположены в нуклеиновой кислоте-матрице, например, РНК-матрице, так что gRNA может принимать первую конформацию, в которой блокирующий домен гибридизирован или по сути гибридизирован с gRNA, и вторую конформацию, в которой блокирующий домен не гибридизирован или по сути не гибридизирован с gRNA. В некоторых вариантах осуществления в первой конформации gRNA не способна связываться с полипептидом Gene Writer (например, с доменом, связывающим нуклеиновую кислоту-матрицу, ДНК-связывающим доменом или эндонуклеазным доменом (например, белком CRISPR/Cas)) или связывается со значительно сниженной аффинностью по сравнению с в остальном аналогичной РНК-матрицей без блокирующего домена. В некоторых вариантах осуществления во второй конформации gRNA способна связываться с полипептидом Gene Writer (например, с доменом, связывающим нуклеиновую кислоту-матрицу, ДНК-связывающим доменом или эндонуклеазным доменом (например, белком CRISPR/Cas)). В некоторых вариантах осуществления то, находится ли gRNA в первой или во второй конформации, может влиять на активность связывания ДНК или эндонуклеазную активность полипептида Gene Writer (например, белка CRISPR/Cas, который содержит полипептид Gene Writer). В некоторых вариантах осуществления гибридизация gRNA с блокирующим доменом может быть нарушена с помощью открывающей молекулы. В некоторых вариантах осуществления открывающая молекула содержит средство, которое связывается с частью gRNA или с ней всей частью или с блокирующим доменом и ингибирует гибридизацию gRNA с блокирующим доменом. В некоторых вариантах осуществления открывающая молекула

содержит нуклеиновую кислоту, например, содержащую последовательность, частично или полностью комплементарную gRNA, блокирующему домену или им обоим. С помощью выбора или разработки подходящей открывающей молекулы, предоставление открывающей молекулы может способствовать изменению конформации gRNA таким образом, чтобы она могла ассоциировать с белком CRISPR/Cas и обеспечивать функции, ассоциированные с белком CRISPR/Cas (например, активность связывания ДНК и/или эндонуклеазную активность). Не желая ограничиваться какой-либо теорией, полагают, что предоставление открывающей молекулы в выбранное время и/или в выбранном местоположении может обеспечить пространственный и временной контроль активности gRNA, белка CRISPR/Cas или системы Gene Writer, содержащей их. В некоторых вариантах осуществления Gene Writer может содержать белок Cas, указанный в таблице 9 или таблице 37, или его функциональный фрагмент или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% идентичностью с ними.

Таблица 9. Белки CRISPR/Cas, виды и мутации

Вариант	**Исходный хозяин**	**Последовательность белка**	**Мутация никазы**
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEID EEENPIRLIDLGVRVFERAEVPKTGDSLAM ARRLARSVRRLTRRRAHRLLRARRLLKRE GVLQAADFDENGLIKSLPNTPWQLRAAAL DRKLTPLEWSAVLLHLIKHRGYLSQRKNE GETADKELGALLKGVANNAHALQTGDFRT PAELALNKFEKESGHIRNQRGDYSHTFSRK DLQAELILLFEKQKEFGNPHVSGGLKEGIET LLMTQRPALSGDAVQKMLGHCTFEPAEPK AAKNTYTAERFIWLTKLNNLRILEQGSERP LTDTERATLMDEPYRKSKLTYAQARKLLG LEDTAFFKGLRYGKDNAEASTLMEMKAY HAISRALEKEGLKDKKSPLNLSSELQDEIGT AFSLFKTDEDITGRLKDRVQPEILEALLKHI SFDKFVQISLKALRRIVPLMEQGKRYDEAC AEIYGDHYGKKNTEEKIYLPPIPADEIRNPV VLRALSQARKVINGVVRRYGSPARIHIETA REVGKSFKDRKEIEKRQEENRKDREKAAA KFREYFPNFVGEPKSKDILKLRLYEQQHGK CLYSGKEINLVRLNEKGYVEIDHALPFSRT	N611A

		WDDSFNNKVLVLGSENQNKGNQTPYEYF NGKDNSREWQEFKARVETSRFPRSKKQRIL LQKFDEDGFKECNLNDTRYVNRFLCQFVA DHILLTGKGKRRVFASNGQITNLLRGFWGL RKVRAENDRHHALDAVVVACSTVAMQQK ITRFVRYKEMNAFDGKTIDKETGKVLHQK THFPQPWEFFAQEVMIRVFGKPDGKPEFEE ADTPEKLRTLLAEKLSSRPEAVHEYVTPLF VSRAPNRKMSGAHKDTLRSAKRFVKHNEK ISVKRVWLTEIKLADLENMVNYKNGREIEL YEALKARLEAYGGNAKQAFDPKDNPFYK KGGQLVKAVRVEKTQESGVLLNKKNAYTI ADNGDMVRVDVFCKVDKKGKNQYFIVPIY AWQVAENILPDIDCKGYRIDDSYTFCFSLH KYDLIAFQKDEKSKVEFAYYINCDSSNGRF YLAWHDKGSKEQQFRISTQNLVLIQKYQV NELGKEIRPCRLKKRPPVR	
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEES SPIRLIDVGVRTFERAEVAKTGESLALSRRL ARSSRRLIKRRAERLKKAKRLLKAEKILHSI DEKLPINVWQLRVKGLKEKLERQEWAAVL LHLSKHRGYLSQRKNEGKSDNKELGALLS GIASNHQMLQSSEYRTPAEIAVKKFQVEEG HIRNQRGSYTHTFSRLDLLAEMELLFQRQA ELGNSYTSTTLLENLTALLMWQKPALAGD AILKMLGKCTFEPSEYKAAKNSYSAERFV WLTKLNNLRILENGTERALNDNERFALLEQ PYEKSKLTYAQVRAMLALSDNAIFKGVRY LGEDKKTVESKTTLIEMKFYHQIRKTLGSA ELKKEWNELKGNSDLLDEIGTAFSLYKTD DDICRYLEGKLPERVLNALLENLNFDKFIQ LSLKALHQILPLMLQGQRYDEAVSAIYGDH YGKKSTETTRLLPTIPADEIRNPVVLRTLTQ ARKVINAVVRLYGSPARIHIETAREVGKSY QDRKKLEKQQEDNRKQRESAVKKFKEMFP	N605A

		HFVGEPKGKDILKMRLYELQQAKCLYSGK SLELHRLLEKGYVEVDHALPFSRTWDDSF NNKVLVLANENQNKGNLTPYEWLDGKNN SERWQHFVVRVQTSGFSYAKKQRILNHKL DEKGFIERNLNDTRYVARFLCNFIADNMLL VGKGKRNVFASNGQITALLRHRWGLQKV REQNDRHHALDAVVVACSTVAMQQKITRF VRYNEGNVFSGERIDRETGEIIPLHFPSPWA FFKENVEIRIFSENPKLELENRLPDYPQYNH EWVQPLFVSRMPTRKMTGQGHMETVKSA KRLNEGLSVLKVPLTQLKLSDLERMVNRD REIALYESLKARLEQFGNDPAKAFAEPFYK KGGALVKAVRLEQTQKSGVLVRDGNGVA DNASMVRVDVFTKGGKYFLVPIYTWQVA KGILPNRAATQGKDENDWDIMDEMATFQF SLCQNDLIKLVTKKKTIFGYFNGLNRATSNI NIKEHDLDKSKGKLGIYLEVGVKLAISLEK YQVDELGKNIRPCRPTKRQHVR	
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKEN	N580A

		AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRELIN DTLYSTRKDDKGNTLIVNNLNGLYDKDND KLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR NKVVKLSLKPYRFDVYLDNGVYKFVTVK NLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYNNDLIKINGELYRVIGVNNDLL NRIEVNMIDITYREYLENMNDKRPPRIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIK KG	
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD	N580A

		AQKMINEMQKRNRQTNERIEEIIRTTGKEN AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRKLI NDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR NKVVKLSLKPYRFDVYLDNGVYKFVTVK NLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNNDLL NRIEVNMIDITYREYLENMNDKRPPHIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIK KG	
SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKT REVIDAGVRLFPEADSENNSNRRSKRGARR LKRRRIHRLNRVKDLLADYQMIDLNNVPK STDPYTIRVKGLREPLTKEEFAIALLHIAKR RGLHNISVSMGDEEQDNELSTKQQLQKNA QQLQDKYVCELQLERLTNINKVRGEKNRF KTEDFVKEVKQLCETQRQYHNIDDQFIQQ YIDLVSTRREYFEGPGNGSPYGWDGDLLK WYEKLMGRCTYFPEELRSVKYAYSADLFN ALNDLNNLVVTRDDNPKLEYYEKYHIIEN VFKQKKNPTLKQIAKEIGVQDYDIRGYRIT KSGKPQFTSFKLYHDLKNIFEQAKYLEDVE MLDEIAKILTIYQDEISIKKALDQLPELLTES EKSQIAQLTGYTGTHRLSLKCIHIVIDELWE SPENQMEIFTRLNLKPKKVEMSEIDSIPTTL	N588A

		VDEFILSPVVKRAFIQSIKVINAVINRFGLPE DIIIELAREKNSKDRRKFINKLQKQNEATRK KIEQLLAKYGNTNAKYMIEKIKLHDMQEG KCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVS FDNSLNNKVLVKQSENSKKGNRTPYQYLS SNESKISYNQFKQHILNLSKAKDRISKKKR DMLLEERDINKFEVQKEFINRNLVDTRYAT RELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANA DFLFKTHKALRRTDKILEQPGLEVNDTTVK VDTEEKYQELFETPKQVKNIKQFRDFKYSH RVDKKPNRQLINDTLYSTREIDGETYVVQT LKDLYAKDNEKVKKLFTERPQKILMYQHD PKTFEKLMTILNQYAEAKNPLAAYYEDKG EYVTKYAKKGNGPAIHKIKYIDKKLGSYLD VSNKYPETQNKLVKLSLKSFRFDIYKCEQG YKMVSIGYLDVLKKDNYYYIPKDKYEAEK QKKKIKESDLFVGSFYYNDLIMYEDELFRV IGVNSDINNLVELNMVDITYKDFCEVNNVT GEKRIKKTIGKRVVLIEKYTTDILGNLYKTP LPKKPQLIFKRGEL	
SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKT REVIDAGVRLFPEADSENNSNRRSKRGARR LKRRRIHRLNRVKDLLADYQMIDLNNVPK STDPYTIRVKGLREPLTKEEFAIALLHIAKR RGLHNISVSMGDEEQDNELSTKQQLQKNA QQLQDKYVCELQLERLTNINKVRGEKNRF KTEDFVKEVKQLCETQRQYHNIDDQFIQQ YIDLVSTRREYFEGPGNGSPYGWDGDLLK WYEKLMGRCTYFPEELRSVKYAYSADLFN ALNDLNNLVVTRDDNPKLEYYEKYHIIEN VFKQKKNPTLKQIAKEIGVQDYDIRGYRIT KSGKPQFTSFKLYHDLKNIFEQAKYLEDVE MLDEIAKILTIYQDEISIKKALDQLPELLTES EKSQIAQLTGYTGTHRLSLKCIHIVIDELWE	N588A

		SPENQMEIFTRLNLKPKKVEMSEIDSIPTTL VDEFILSPVVKRAFIQSIKVINAVINRFGLPE DIIIELAREKNSKDRRKFINKLQKQNEATRK KIEQLLAKYGNTNAKYMIEKIKLHDMQEG KCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVS FDNSLNNKVLVKQSENSKKGNRTPYQYLS SNESKISYNQFKQHILNLSKAKDRISKKKR DMLLEERDINKFEVQKEFINRNLVDTRYAT RELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANA DFLFKTHKALRRTDKILEQPGLEVNDTTVK VDTEEKYQELFETPKQVKNIKQFRDFKYSH RVDKKPNRKLINDTLYSTREIDGETYVVQT LKDLYAKDNEKVKKLFTERPQKILMYQHD PKTFEKLMTILNQYAEAKNPLAAYYEDKG EYVTKYAKKGNGPAIHKIKYIDKKLGSYLD VSNKYPETQNKLVKLSLKSFRFDIYKCEQG YKMVSIGYLDVLKKDNYYYIPKDKYEAEK QKKKIKESDLFVGSFYKNDLIMYEDELFRV IGVNSDINNLVELNMVDITYKDFCEVNNVT GEKHIKKTIGKRVVLIEKYTTDILGNLYKTP LPKKPQLIFKRGEL	
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPS KKFKVLGNTNRKSIKKNLMGALLFDSGET AEATRLKRTARRRYTRRKNRIRYLQEIFAN EMAKLDDSFFQRLEESFLVEEDKKNERHPI FGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAEN SDVAKLFYQLIQTYNQLFEESPLDEIEVDA KGILSARLSKSKRLEKLIAVFPNEKKNGLFG NIIALALGLTPNFKSNFDLTEDAKLQLSKDT YDDDLDELLGQIGDQYADLFSAAKNLSDAI LLSDILRSNSEVTKAPLSASMVKRYDEHHQ DLALLKTLVRQQFPEKYAEIFKDDTKNGY AGYVGADKKLRKRSGKLATEEEFYKFIKPI	N872A

		LEKMDGAEELLAKLNRDDLLRKQRTFDNG SIPHQIHLKELHAILRRQEEFYPFLKENREKI EKILTFRIPYYVGPLARGNSRFAWLTRKSEE AITPWNFEEVVDKGASAQSFIERMTNFDEQ LPNKKVLPKHSLLYEYFTVYNELTKVKYV TERMRKPEFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEIIGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTIL DFLKSDGFSNRNFMQLIHDDSLTFKEEIEK AQVSGQGDSLHEQIADLAGSPAIKKGILQT VKIVDELVKVMGHKPENIVIEMARENQTTT KGLQQSRERKKRIEEGIKELESQILKENPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS VENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIK RQLVETRQITKHVARILDSRMNTKRDKND KPIREVKVITLKSKLVSDFRKDFQLYKVRDI NNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKR FFYSNIMNFFKTEVKLANGEIRKRPLIETNG ETGEVVWNKEKDFATVRKVLAMPQVNIV KKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKA KKLKSVKVLVGITIMEKGSYEKDPIGFLEA KGYKDIKKELIFKLPKYSLFELENGRRRML ASAKELQKANELVLPQHLVRLLYYTQNISA TTGSNNLGYIEQHREEFKEIFEKIIDFSEKYI LKNKVNSNLKSSFDEQFAVSDSILLSNSFVS LLKYTSFGASGGFTFLDLDVKQGRLRYQT VTEVLDATLIYQSITGLYETRTDLSQLGGD	
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA	N863A

		EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG	

		GFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN	N863A

		SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESIRPKRNSDKLIARKKDWDPKKYGG FVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASARFLQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPRAFKYFDTTIDRKVYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA ERTRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV	N863A

		GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESIRPKRNSDKLIARKKDWDPKKYGG FLWPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAKQLQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTRLGAPRAFKYFDTTIDPKQYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKN SRIFPAAQAENNLVRRTNRQGRRLARRKK HRRVRLNRLFEESGLITDFTKISINLNPYQL RVKGLTDELSNEELFIALKNMVKHRGISYL	N622A

		DDASDDGNSSVGDYAQIVKENSKQLETKT PGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRY RTSGETLDNIFGILIGKCTFYPDEFRAAKAS YTAQEFNLLNDLNNLTVPTETKKLSKEQK NQIINYVKNEKAMGPAKLFKYIAKLLSCDV ADIKGYRIDKSGKAEIHTFEAYRKMKTLET LDIEQMDRETLDKLAYVLTLNTEREGIQEA LEHEFADGSFSQKQVDELVQFRKANSSIFG KGWHNFSVKLMMELIPELYETSEEQMTILT RLGKQKTTSSSNKTKYIDEKLLTEEIYNPV VAKSVRQAIKIVNAAIKEYGDFDNIVIEMA RETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLW HQQGERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTP YQALDSMDDAWSFRELKAFVRESKTLSNK KKEYLLTEEDISKFDVRKKFIERNLVDTRY ASRVVLNALQEHFRAHKIDTKVSVVRGQF TSQLRRHWGIEKTRDTYHHHAVDALIIAAS SQLNLWKKQKNTLVSYSEDQLLDIETGELI SDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKD KADETYVLGKIKDIYTQDGYDAFMKIYKK DKSKFLMYRHDPQTFEKVIEPILENYPNKQI NEKGKEVPCNPFLKYKEEHGYIRKYSKKG NGPEIKSLKYYDSKLGNHIDITPKDSNNKV VLQSVSPWRADVYFNKTTGKYEILGLKYA DLQFEKGTGTYKISQEKYNDIKKKEGVDSD SEFKFTLYKNDLLLVKDTETKEQQLFRFLS RTMPKQKHYVELKPYDKQKFEGGEALIKV LGNVANSGQCKKGLGKSNISIYKVRTDVL GNQHIIKNEGDKPKLDF	

BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIG VRTFEKAENPKNGEALAVPRREARSSRRRL RRKKHRIERLKHMFVRNGLAVDIQHLEQT LRSQNEIDVWQLRVDGLDRMLTQKEWLR VLIHLAQRRGFQSNRKTDGSSEDGQVLVN VTENDRLMEEKDYRTVAEMMVKDEKFSD HKRNKNGNYHGVVSRSSLLVEIHTLFETQR QHHNSLASKDFELEYVNIWSAQRPVATKD QIEKMIGTCTFLPKEKRAPKASWHFQYFML LQTINHIRITNVQGTRSLNKEEIEQVVNMAL TKSKVSYHDTRKILDLSEEYQFVGLDYGKE DEKKKVESKETIIKLDDYHKLNKIFNEVEL AKGETWEADDYDTVAYALTFFKDDEDIRD YLQNKYKDSKNRLVKNLANKEYTNELIGK VSTLSFRKVGHLSLKALRKIIPFLEQGMTY DKACQAAGFDFQGISKKKRSVVLPVIDQIS NPVVNRALTQTRKVINALIKKYGSPETIHIE TARELSKTFDERKNITKDYKENRDKNEHA KKHLSELGIINPTGLDIVKYKLWCEQQGRC MYSNQPISFERLKESGYTEVDHIIPYSRSMN DSYNNRVLVMTRENREKGNQTPFEYMGN DTQRWYEFEQRVTTNPQIKKEKRQNLLLK GFTNRRELEMLERNLNDTRYITKYLSHFIST NLEFSPSDKKKKVVNTSGRITSHLRSRWGL EKNRGQNDLHHAMDAIVIAVTSDSFIQQVT NYYKRKERRELNGDDKFPLPWKFFREEVI ARLSPNPKEQIEALPNHFYSEDELADLQPIF VSRMPKRSITGEAHQAQFRRVVGKTKEGK NITAKKTALVDISYDKNGDFNMYGRETDP ATYEAIKERYLEFGGNVKKAFSTDLHKPK KDGTKGPLIKSVRIMENKTLVHPVNKGKG VVYNSSIVRTDVFQRKEKYYLLPVYVTDV TKGKLPNKVIVAKKGYHDWIEVDDSFTFLF SLYPNDLIFIRQNPKKKISLKKRIESHSISDS KEVQEIHAYYKGVDSSTAAIEFIIHDGSYYA	N607A

		KGVGVQNLDCFEKYQVDILGNYFKVKGEK RLELETSDSNHKGKDVNSIKSTSR	
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKEN AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRKLI NDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR NKVVKLSLKPYRFDVYLDNGVYKFVTVK NLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNSDKN	N580A

		NLIEVNMIDITYREYLENMNDKRPPHIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIK KG	
cCas9-v17	Staphylococcu s aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKEN AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRKLI NDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR NKVVKLSLKPYRFDVYLDNGVYKFVTVK NLDVIKKENYYEVNSKCYEEAKKLKKISN	N580A

		QAEFIASFYKNDLIKINGELYRVIGVNNSTR NIVELNMIDITYREYLENMNDKRPPHIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIK KG	
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKEN AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRKLI NDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR NKVVKLSLKPYRFDVYLDNGVYKFVTVK	N580A

		NLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNSDDR NIIELNMIDITYREYLENMNDKRPPHIIKTIA SKTQSIKKYSTDILGNLYEVKSKKHPQIIKK G	
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDA GVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLK QIAKEILVNEEDIKGYRVTSTGKPEFTNLKV YHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTG THNLSLKAINLILDELWHTNDNQIAIFNRLK LVPKKVDLSQQKEIPTTLVDDFILSPVVKRS FIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKEN AKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEENSKKGNRTPFQYLSSSDSKISYETFKK HILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRV NNLDVKVKSINGGFTSFLRRKWKFKKERN KGYKHHAEDALIIANADFIFKEWKKLDKA KKVMENQMFEEKQAESMPEIETEQEYKEIF ITPHQIKHIKDFKDYKYSHRVDKKPNRKLI NDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSR	N580A

		NKVVKLSLKPYRFDVYLDNGVYKFVTVK NLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNNNRL NKIELNMIDITYREYLENMNDKRPPHIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIK KG	
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIK TLSLVSHIHDSGLDPDEIKSAVTRLASSGIA RRTRRLYRRKRRRLQQLDKFIQRQGWPVIE LEDYSDPLYPWKVRAELAASYIADEKERG EKLSVALRHIARHRGWRNPYAKVSSLYLP DGPSDAFKAIREEIKRASGQPVPETATVGQ MVTLCELGTLKLRGEGGVLSARLQQSDYA REIQEICRMQEIGQELYRKIIDVVFAAESPK GSASSRVGKDPLQPGKNRALKASDAFQRY RIAALIGNLRVRVDGEKRILSVEEKNLVFD HLVNLTPKKEPEWVTIAEILGIDRGQLIGTA TMTDDGERAGARPPTHDTNRSIVNSRIAPL VDWWKTASALEQHAMVKALSNAEVDDF DSPEGAKVQAFFADLDDDVHAKLDSLHLP VGRAAYSEDTLVRLTRRMLSDGVDLYTAR LQEFGIEPSWTPPTPRIGEPVGNPAVDRVLK TVSRWLESATKTWGAPERVIIEHVREGFVT EKRAREMDGDMRRRAARNAKLFQEMQEK LNVQGKPSRADLWRYQSVQRQNCQCAYC GSPITFSNSEMDHIVPRAGQGSTNTRENLV AVCHRCNQSKGNTPFAIWAKNTSIEGVSV KEAVERTRHWVTDTGMRSTDFKKFTKAV VERFQRATMDEEIDARSMESVAWMANEL RSRVAQHFASHGTTVRVYRGSLTAEARRA SGISGKLKFFDGVGKSRLDRRHHAIDAAVI AFTSDYVAETLAVRSNLKQSQAHRQEAPQ WREFTGKDAEHRAAWRVWCQKMEKLSA LLTEDLRDDRVVVMSNVRLRLGNGSAHKE TIGKLSKVKLSSQLSVSDIDKASSEALWCA	H573A (альтернативная)

		LTREPGFDPKEGLPANPERHIRVNGTHVYA GDNIGLFPVSAGSIALRGGYAELGSSFHHA RVYKITSGKKPAFAMLRVYTIDLLPYRNQD LFSVELKPQTMSMRQAEKKLRDALATGNA EYLGWLVVDDELVVDTSKIATDQVKAVEA ELGTIRRWRVDGFFSPSKLRLRPLQMSKEG IKKESAPELSKIIDRPGWLPAVNKLFSDGNV TVVRRDSLGRVRLESTAHLPVTWKVQ	
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRI FTKVENPKTGESLALPRRLARSARKRLARR KARLNHLKHLIANEFKLNYEDYQSFDESLA KAYKGSLISPYELRFRALNELLSKQDFARVI LHIAKRRGYDDIKNSDDKEKGAILKAIKQN EEKLANYQSVGEYLYKEYFQKFKENSKEF TNVRNKKESYERCIAQSFLKDELKLIFKKQ REFGFSFSKKFEEEVLSVAFYKRALKDFSH LVGNCSFFTDEKRAPKNSPLAFMFVALTRII NLLNNLKNTEGILYTKDDLNALLNEVLKN GTLTYKQTKKLLGLSDDYEFKGEKGTYFIE FKKYKEFIKALGEHNLSQDDLNEIAKDITLI KDEIKLKKALAKYDLNQNQIDSLSKLEFKD HLNISFKALKLVTPLMLEGKKYDEACNEL NLKVAINEDKKDFLPAFNETYYKDEVTNP VVLRAIKEYRKVLNALLKKYGKVHKINIEL AREVGKNHSQRAKIEKEQNENYKAKKDAE LECEKLGLKINSKNILKLRLFKEQKEFCAYS GEKIKISDLQDEKMLEIDHIYPYSRSFDDSY MNKVLVFTKQNQEKLNQTPFEAFGNDSAK WQKIEVLAKNLPTKKQKRILDKNYKDKEQ KNFKDRNLNDTRYIARLVLNYTKDYLDFL PLSDDENTKLNDTQKGSKVHVEAKSGMLT SALRHTWGFSAKDRNNHLHHAIDAVIIAY ANNSIVKAFSDFKKEQESNSAELYAKKISE LDYKNKRKFFEPFSGFRQKVLDKIDEIFVSK PERKKPSGALHEETFRKEEEFYQSYGGKEG	N582A

		VLKALELGKIRKVNGKIVKNGDMFRVDIF KHKKTNKFYAVPIYTMDFALKVLPNKAVA RSKKGEIKDWILMDENYEFCFSLYKDSLILI QTKDMQEPEFVYYNAFTSSTVSLIVSKHDN KFETLSKNQKILFKNANEKEVIAKSIGIQNL KVFEKYIVSALGEVTKAEFRQREDFKK	
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLG VRIFDRAENPQTGESLALPRRLARSARRRL RRRKHRLERIRRLVIREGILTKEELDKLFEE KHEIDVWQLRVEALDRKLNNDELARVLLH LAKRRGFKSNRKSERSNKENSTMLKHIEEN RAILSSYRTVGEMIVKDPKFALHKRNKGEN YTNTIARDDLEREIRLIFSKQREFGNMSCTE EFENEYITIWASQRPVASKDDIEKKVGFCTF EPKEKRAPKATYTFQSFIAWEHINKLRLISP SGARGLTDEERRLLYEQAFQKNKITYHDIR TLLHLPDDTYFKGIVYDRGESRKQNENIRF LELDAYHQIRKAVDKVYGKGKSSSFLPIDF DTFGYALTLFKDDADIHSYLRNEYEQNGK RMPNLANKVYDNELIEELLNLSFTKFGHLS LKALRSILPYMEQGEVYSSACERAGYTFTG PKKKQKTMLLPNIPPIANPVVMRALTQARK VVNAIIKKYGSPVSIHIELARDLSQTFDERR KTKKEQDENRKKNETAIRQLMEYGLTLNP TGHDIVKFKLWSEQNGRCAYSLQPIEIERLL EPGYVEVDHVIPYSRSLDDSYTNKVLVLTR ENREKGNRIPAEYLGVGTERWQQFETFVLT NKQFSKKKRDRLLRLHYDENEETEFKNRN LNDTRYISRFFANFIREHLKFAESDDKQKV YTVNGRVTAHLRSRWEFNKNREESDLHHA VDAVIVACTTPSDIAKVTAFYQRREQNKEL AKKTEPHFPQPWPHFADELRARLSKHPKES IKALNLGNYDDQKLESLQPVFVSRMPKRS VTGAAHQETLRRYVGIDERSGKIQTVVKT KLSEIKLDASGHFPMYGKESDPRTYEAIRQ	N605A

		RLLEHNNDPKKAFQEPLYKPKKNGEPGPVI RTVKIIDTKNQVIPLNDGKTVAYNSNIVRV DVFEKDGKYYCVPVYTMDIMKGILPNKAI EPNKPYSEWKEMTEDYTFRFSLYPNDLIRIE LPREKTVKTAAGEEINVKDVFVYYKTIDSA NGGLELISHDHRFSLRGVGSRTLKRFEKYQ VDVLGNIYKVRGEKRVGLASSAHSKPGKTI RPLQSTRD	
iSpyMacCas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRKLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLKREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ	N863A

		NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEIQTV GQNGGLFDDNPKSPLEVTPSKLVPLKKELN PKKYGGYQKPTTAYPVLLITDTKQLIPISV MNKKQFEQNPVKFLRDRGYQQVGKNDFIK LPKYTLVDIGDGIKRLWASSKEIHKGNQLV VSKKSQILLYHAHHLDSDLSNDYLQNHNQ QFDVLFNEIISFSKKCKLGKEHIQKIENVYS NKKNSASIEELAESFIKLLGFTQLGATSPFN FLGVKLNQKQYKGKKDYILPCTEGTLIRQS ITGLYETRVDLSKIGEDSGGSGGSKRTADG SEFES	
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEID EEENPIRLIDLGVRVFERAEVPKTGDSLAM ARRLARSVRRLTRRRAHRLLRTRRLLKRE GVLQAANFDENGLIKSLPNTPWQLRAAAL DRKLTPLEWSAVLLHLIKHRGYLSQRKNE GETADKELGALLKGVAGNAHALQTGDFRT PAELALNKFEKESGHIRNQRSDYSHTFSRK DLQAELILLFEKQKEFGNPHVSGGLKEGIET LLMTQRPALSGDAVQKMLGHCTFEPAEPK AAKNTYTAERFIWLTKLNNLRILEQGSERP LTDTERATLMDEPYRKSKLTYAQARKLLG LEDTAFFKGLRYGKDNAEASTLMEMKAY HAISRALEKEGLKDKKSPLNLSPELQDEIGT AFSLFKTDEDITGRLKDRIQPEILEALLKHIS FDKFVQISLKALRRIVPLMEQGKRYDEACA	N611A

		EIYGDHYGKKNTEEKIYLPPIPADEIRNPVV LRALSQARKVINGVVRRYGSPARIHIETAR EVGKSFKDRKEIEKRQEENRKDREKAAAK FREYFPNFVGEPKSKDILKLRLYEQQHGKC LYSGKEINLGRLNEKGYVEIDHALPFSRTW DDSFNNKVLVLGSENQNKGNQTPYEYFNG KDNSREWQEFKARVETSRFPRSKKQRILLQ KFDEDGFKERNLNDTRYVNRFLCQFVADR MRLTGKGKKRVFASNGQITNLLRGFWGLR KVRAENDRHHALDAVVVACSTVAMQQKI TRFVRYKEMNAFDGKTIDKETGEVLHQKT HFPQPWEFFAQEVMIRVFGKPDGKPEFEEA DTLEKLRTLLAEKLSSRPEAVHEYVTPLFV SRAPNRKMSGQGHMETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEA LKARLEAHKDDPAKAFAEPFYKYDKAGNR TQQVKAVRVEQVQKTGVWVRNHNGIADN ATMVRVDVFEKGDKYYLVPIYSWQVAKGI LPDRAVVQGKDEEDWQLIDDSFNFKFSLHP NDLVEVITKKARMFGYFASCHRGTGNINIR IHDLDHKIGKNGILEGIGVKTALSFQKYQID ELGKEIRPCRLKKRPPVR	
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPS KKFKVLGNTNRKSIKKNLMGALLFDSGET AEATRLKRTARRRYTRRKNRIRYLQEIFAN EMAKLDDSFFQRLEESFLVEEDKKNERHPI FGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAEN SDVAKLFYQLIQTYNQLFEESPLDEIEVDA KGILSARLSKSKRLEKLIAVFPNEKKNGLFG NIIALALGLTPNFKSNFDLTEDAKLQLSKDT YDDDLDELLGQIGDQYADLFSAAKNLSDAI LLSDILRSNSEVTKAPLSASMVKRYDEHHQ DLALLKTLVRQQFPEKYAEIFKDDTKNGY AGYVGIGIKHRKRTTKLATQEEFYKFIKPIL	N872A

		EKMDGAEELLAKLNRDDLLRKQRTFDNGS IPHQIHLKELHAILRRQEEFYPFLKENREKIE KILTFRIPYYVGPLARGNSRFAWLTRKSEE AITPWNFEEVVDKGASAQSFIERMTNFDEQ LPNKKVLPKHSLLYEYFTVYNELTKVKYV TERMRKPEFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEIIGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTIL DFLKSDGFSNRNFMQLIHDDSLTFKEEIEK AQVSGQGDSLHEQIADLAGSPAIKKGILQT VKIVDELVKVMGHKPENIVIEMARENQTTT KGLQQSRERKKRIEEGIKELESQILKENPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS VENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIK RQLVETRQITKHVARILDSRMNTKRDKND KPIREVKVITLKSKLVSDFRKDFQLYKVRDI NNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKR FFYSNIMNFFKTEVKLANGEIRKRPLIETNG ETGEVVWNKEKDFATVRKVLAMPQVNIV KKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKA KKLKSVKVLVGITIMEKGSYEKDPIGFLEA KGYKDIKKELIFKLPKYSLFELENGRRRML ASATELQKANELVLPQHLVRLLYYTQNISA TTGSNNLGYIEQHREEFKEIFEKIIDFSEKYI LKNKVNSNLKSSFDEQFAVSDSILLSNSFVS LLKYTSFGASGGFTFLDLDVKQGRLRYQT VTEVLDATLIYQSITGLYETRTDLSQLGGD	
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPS KKFKVLGNTNRKSIKKNLMGALLFDSGET	N872A

		AEATRLKRTARRRYTRRKNRIRYLQEIFAN EMAKLDDSFFQRLEESFLVEEDKKNERHPI FGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAEN SDVAKLFYQLIQTYNQLFEESPLDEIEVDA KGILSARLSKSKRLEKLIAVFPNEKKNGLFG NIIALALGLTPNFKSNFDLTEDAKLQLSKDT YDDDLDELLGQIGDQYADLFSAAKNLSDAI LLSDILRSNSEVTKAPLSASMVKRYDEHHQ DLALLKTLVRQQFPEKYAEIFKDDTKNGY AGYVGADKKLRKRSGKLATEEEFYKFIKPI LEKMDGAEELLAKLNRDDLLRKQRTFDNG SIPHQIHLKELHAILRRQEEFYPFLKENREKI EKILTFRIPYYVGPLARGNSRFAWLTRKSEE AITPWNFEEVVDKGASAQSFIERMTNFDEQ LPNKKVLPKHSLLYEYFTVYNELTKVKYV TERMRKPEFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEIIGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTIL DFLKSDGFSNANFMQLIHDDSLTFKEEIEK AQVSGQGDSLHEQIADLAGSPAIKKGILQT VKIVDELVKVMGHKPENIVIEMARENQTTT KGLQQSRERKKRIEEGIKELESQILKENPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS VENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIK RQLVETRQITKHVARILDSRMNTKRDKND KPIREVKVITLKSKLVSDFRKDFQLYKVRDI NNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKR FFYSNIMNFFKTEVKLANGEIRKRPLIETNG ETGEVVWNKEKDFATVRKVLAMPQVNIV	

		KKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKA KKLKSVKVLVGITIMEKGSYEKDPIGFLEA KGYKDIKKELIFKLPKYSLFELENGRRRML ASAKELQKANELVLPQHLVRLLYYTQNISA TTGSNNLGYIEQHREEFKEIFEKIIDFSEKYI LKNKVNSNLKSSFDEQFAVSDSILLSNSFVS LLKYTSFGASGGFTFLDLDVKQGRLRYQT VTEVLDATLIYQSITGLYETRTDLSQLGGD	
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMVKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQGDFYPFLKDNREKIEKILTFRIPYY VGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRLRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGGHKPENIVIEMARENQTTQKGQKN	N863A

		SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKGNSDKLIARKKDWDPKKYGG FNSPTAAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIGFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGVLHKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGVPAAFKYFDTTIDKKRYTSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMVKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQGDFYPFLKDNREKIEKILTFRIPYY	N863A

		VGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRLRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGGHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKGNSDKLIARKKDWDPKKYGG FNSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIGFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASASVLHKG NELALPSKYVNFLYLASHYEKLKGSSEDN KQKQLFVEQHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGASAAFKYFDTTIGRKLYTSTKEVL DATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF	N863A

		GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMVKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQGDFYPFLKDNREKIEKILTFRIPYY VGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRLRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGGHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKGNSDKLIARKKDWDPKKYGG FNSPTVAYSVLVVAKVEKGKSKKLKSVKE	

		LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGVLQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTINRKQYNTTKEVLD ATLIRQSITGLYETRIDLSQLGGD	
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY	N863A

		DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK	N863A

		HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASARELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD AQLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTFKQKQYRSTKEVL DATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN	N863A

		SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FLWPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAKQLQKG	

		NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKQYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR	N863A

		KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKQYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDY	N863A

		FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASARELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKEYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG	N863A

		NLIALSLGLTPNFKSNFDLAEDTKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKLYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEKV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGDQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFIQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIKKGILQTVKVVDELV KVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGVLQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD	

		ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDTKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKLYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEKV VDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGDQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLK IIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFIQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIKKGILQTVKVVDELV KVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKV	N863A

		ITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESIRPKRNSDKLIARKKDWDPKKYGG FVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASARFLQKG NELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPRAFKYFDTTIDRKVYRSTKEVLD ATLIHQSITGLYETRIDLSQLGGD	
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKN SRIFPAAQAENNLVRRTNRQGRRLARRKK HRRVRLNRLFEESGLITDFTKISINLNPYQL RVKGLTDELSNEELFIALKNMVKHRGISYL DDASDDGNSSVGDYAQIVKENSKQLETKT PGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRY RTSGETLDNIFGILIGKCTFYPDEFRAAKAS YTAQEFNLLNDLNNLTVPTETKKLSKEQK NQIINYVKNEKAMGPAKLFKYIAKLLSCDV ADIKGYRIDKSGKAEIHTFEAYRKMKTLET LDIEQMDRETLDKLAYVLTLNTEREGIQEA LEHEFADGSFSQKQVDELVQFRKANSSIFG KGWHNFSVKLMMELIPELYETSEEQMTILT RLGKQKTTSSSNKTKYIDEKLLTEEIYNPV VAKSVRQAIKIVNAAIKEYGDFDNIVIEMA RETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLW HQQGERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTP	N622A

		YQALDSMDDAWSFRELKAFVRESKTLSNK KKEYLLTEEDISKFDVRKKFIERNLVDTRY ASRVVLNALQEHFRAHKIDTKVSVVRGQF TSQLRRHWGIEKTRDTYHHHAVDALIIAAS SQLNLWKKQKNTLVSYSEEQLLDIETGELI SDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKD KKDETYVLGKIKDIYTQDGYDAFMKIYKK DKSKFLMYRHDPQTFEKVIEPILENYPNKQ MNEKGKEVPCNPFLKYKEEHGYIRKYSKK GNGPEIKSLKYYDSKLLGNPIDITPENSKNK VVLQSLKPWRTDVYFNKATGKYEILGLKY ADLQFEKGTGTYKISQEKYNDIKKKEGVDS DSEFKFTLYKNDLLLVKDTETKEQQLFRFL SRTLPKQKHYVELKPYDKQKFEGGEALIK VLGNVANGGQCIKGLAKSNISIYKVRTDVL GNQHIIKNEGDKPKLDF	
St1Cas9-LMG1831	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKN SRIFPAAQAENNLVRRTNRQGRRLARRKK HRRVRLNRLFEESGLITDFTKISINLNPYQL RVKGLTDELSNEELFIALKNMVKHRGISYL DDASDDGNSSVGDYAQIVKENSKQLETKT PGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRY RTSGETLDNIFGILIGKCTFYPDEFRAAKAS YTAQEFNLLNDLNNLTVPTETKKLSKEQK NQIINYVKNEKAMGPAKLFKYIAKLLSCDV ADIKGYRIDKSGKAEIHTFEAYRKMKTLET LDIEQMDRETLDKLAYVLTLNTEREGIQEA LEHEFADGSFSQKQVDELVQFRKANSSIFG KGWHNFSVKLMMELIPELYETSEEQMTILT RLGKQKTTSSSNKTKYIDEKLLTEEIYNPV VAKSVRQAIKIVNAAIKEYGDFDNIVIEMA RETNEDDEKKAIQKIQKANKDEKDAAMLK	N622A

		AANQYNGKAELPHSVFHGHKQLATKIRLW HQQGERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTP YQALDSMDDAWSFRELKAFVRESKTLSNK KKEYLLTEEDISKFDVRKKFIERNLVDTRY ASRVVLNALQEHFRAHKIDTKVSVVRGQF TSQLRRHWGIEKTRDTYHHHAVDALIIAAS SQLNLWKKQKNTLVSYSEEQLLDIETGELI SDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKD KKDETYVLGKIKDIYTQDGYDAFMKIYKK DKSKFLMYRHDPQTFEKVIEPILENYPNKQ MNEKGKEVPCNPFLKYKEEHGYIRKYSKK GNGPEIKSLKYYDSKLLGNPIDITPENSKNK VVLQSLKPWRTDVYFNKNTGKYEILGLKY ADLQFEKKTGTYKISQEKYNGIMKEEGVD SDSEFKFTLYKNDLLLVKDTETKEQQLFRF LSRTMPNVKYYVELKPYSKDKFEKNESLIE ILGSADKSGRCIKGLGKSNISIYKVRTDVLG NQHIIKNEGDKPKLDF	
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKN SRIFPAAQAENNLVRRTNRQGRRLARRKK HRRVRLNRLFEESGLITDFTKISINLNPYQL RVKGLTDELSNEELFIALKNMVKHRGISYL DDASDDGNSSVGDYAQIVKENSKQLETKT PGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRY RTSGETLDNIFGILIGKCTFYPDEFRAAKAS YTAQEFNLLNDLNNLTVPTETKKLSKEQK NQIINYVKNEKAMGPAKLFKYIAKLLSCDV ADIKGYRIDKSGKAEIHTFEAYRKMKTLET LDIEQMDRETLDKLAYVLTLNTEREGIQEA LEHEFADGSFSQKQVDELVQFRKANSSIFG KGWHNFSVKLMMELIPELYETSEEQMTILT	N622A

		RLGKQKTTSSSNKTKYIDEKLLTEEIYNPV VAKSVRQAIKIVNAAIKEYGDFDNIVIEMA RETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLW HQQGERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTP YQALDSMDDAWSFRELKAFVRESKTLSNK KKEYLLTEEDISKFDVRKKFIERNLVDTRY ASRVVLNALQEHFRAHKIDTKVSVVRGQF TSQLRRHWGIEKTRDTYHHHAVDALIIAAS SQLNLWKKQKNTLVSYSEDQLLDIETGELI SDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKD KADETYVLGKIKDIYTQDGYDAFMKIYKK DKSKFLMYRHDPQTFEKVIEPILENYPNKQI NEKGKEVPCNPFLKYKEEHGYIRKYSKKG NGPEIKSLKYYDSKLGNHIDITPKDSNNKV VLQSLKPWRTDVYFNKNTGKYEILGLKYS DMQFEKGTGKYSISKEQYENIKVREGVDE NSEFKFTLYKNDLLLLKDSENGEQILLRFTS RNDTSKHYVELKPYNRQKFEGSEYLIKSLG TVAKGGQCIKGLGKSNISIYKVRTDVLGNQ HIIKNEGDKPKLDF	
St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKN SRIFPAAQAENNLVRRTNRQGRRLARRKK HRRVRLNRLFEESGLITDFTKISINLNPYQL RVKGLTDELSNEELFIALKNMVKHRGISYL DDASDDGNSSVGDYAQIVKENSKQLETKT PGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRY RTSGETLDNIFGILIGKCTFYPDEFRAAKAS YTAQEFNLLNDLNNLTVPTETKKLSKEQK NQIINYVKNEKAMGPAKLFKYIAKLLSCDV ADIKGYRIDKSGKAEIHTFEAYRKMKTLET	N622A

		LDIEQMDRETLDKLAYVLTLNTEREGIQEA LEHEFADGSFSQKQVDELVQFRKANSSIFG KGWHNFSVKLMMELIPELYETSEEQMTILT RLGKQKTTSSSNKTKYIDEKLLTEEIYNPV VAKSVRQAIKIVNAAIKEYGDFDNIVIEMA RETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLW HQQGERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTP YQALDSMDDAWSFRELKAFVRESKTLSNK KKEYLLTEEDISKFDVRKKFIERNLVDTRY ASRVVLNALQEHFRAHKIDTKVSVVRGQF TSQLRRHWGIEKTRDTYHHHAVDALIIAAS SQLNLWKKQKNTLVSYSEDQLLDIETGELI SDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKD KADETYVLGKIKDIYTQDGYDAFMKIYKK DKSKFLMYRHDPQTFEKVIEPILENYPNKQI NEKGKEVPCNPFLKYKEEHGYIRKYSKKG NGPEIKSLKYYDSKLGNHIDITPKDSNNKV VLQSLKPWRTDVYFNKNTGKYEILGLKYS DMQFEKGTGKYSISKEQYENIKVREGVDE NSEFKFTLYKNDLLLLKDSENGEQILLRFTS RNDTSKHYVELKPYNRQKFEGSEYLIKSLG TVVKGGRCIKGLGKSNISIYKVRTDVLGNQ HIIKNEGDKPKLDF	

В таблице 9B представлены параметры для определения необходимых компонентов для конструирования gRNA и/или РНК-матриц в применении вариантов Cas, перечисленных в таблице 3A, для Gene Writing. Ярус указывает на предпочтительные варианты Cas, если они доступны для использования в данном локусе. Место разреза указывает на валидированные или предсказанные требования для мотива, прилегающего к протоспейсеру (PAM), валидированное или предсказанное местоположение сайта разреза (относительно наиболее вышерасположенного основания сайта PAM). gRNA для данного фермента может быть собрана посредством соединения последовательностей crRNA, тетрапетли и tracrRNA и дополнительного добавления спейсера на 5'-конце с длиной в пределах от спейсера (min) до спейсера (max), которая соответствует протоспейсеру в

сайте-мишени. Дополнительно, предсказанное местоположение однонитевого разреза ssDNA в мишени является важным для разработки 3'-концевой области РНК-матрицы, которую необходимо отжечь с последовательностью непосредственно с 5'-стороны от однонитевого разреза с целью инициации обратной транскрипции с праймированием мишенью.

Таблица 9B. Параметры для определения необходимых компонентов для конструирования gRNA и/или РНК-матриц в применении вариантов Cas, перечисленных в таблице 9A, для Gene Writing

Вариант	PAM	Разрез	Ярус	Спейсер (min)	Спейсер (max)	crRNA	Тетрапетля	tracrRNA
Nme2 Cas9	NNN NCC	-3	1	22	24	GTTGTAG CTCCCTT TCTCATT TCG	GAA A	CGAAATGAGAA CCGTTGCTACA ATAAGGCCGTC TGAAAAGATGT GCCGCAACGCT CTGCCCCTTAA AGCTTCTGCTTT AAGGGGCATCG TTTA
PpnCa s9	NNN NRTT		1	21	24	GTTGTAG CTCCCTT TTTCATT TCGC	GAA A	GCGAAATGAAA AACGTTGTTAC AATAAGAGATG AATTTCTCGCA AAGCTCTGCCT CTTGAAATTTC GGTTTCAAGAG GCATCTTTTT
SauCa s9	NNG RR;N NGRR T	-3	1	21	23	GTTTTAG TACTCTG	GAA A	CAGAATCTACT AAAACAAGGCA AAATGCCGTGT TTATCTCGTCA ACTTGTTGGCG AGA

SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAGTACTCTGTAAT	GAAA	ATTACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9	NNGG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT
SauriCas9-KKH	NNRG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9	NGG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-NG	NG (NGG=NGA=NGT	-3	1	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAA

	>NGC)							AAAGTGGCACC GAGTCGGTGC
SpyCas9-SpRY	NRN>NYN	-3	1	20	20	GTTTTAG AGCTA	GAAA	TAGCAAGTTAA AATAAGGCTAG TCCGTTATCAA CTTGAAAAAGT GGCACCGAGTC GGTGC
St1Cas9	NNAGAAW>NNAGGAW=NNGGAAW	-3	1	20	20	GTCTTTG TACTCTG	GTAC	CAGAAGCTACA AAGATAAGGCT TCATGCCGAAA TCAACACCCTG TCATTTTATGG CAGGGTGTTTT
BlatCas9	NNNNCNAA>NNNNCNDD>NNNNC	-3	1	19	23	GCTATAG TTCCTTA CT	GAAA	GGTAAGTTGCT ATAGTAAGGGC AACAGACCCGA GGCGTTGGGGA TCGCCTAGCCC GTGTTTACGGG CTCTCCCCATA TTCAAAATAAT GACAGACGAGC ACCTTGGAGCA TTTATCTCCGA GGTGCT
cCas9-v16	NNVACT;NNVATGM;NNVATT;NNV	-3	2	21	21	GUCUUAG UACUCUG	GAAA	CAGAAUCUACU AAGACAAGGCA AAAUGCCGUGU UUAUCUCGUCA ACUUGUUGGCG AGAUUUUUUU

	GCT; NNV GTG; NNV GTT							
cCas9-v17	NNV RRN	-3	2	21	21	GUCUUAG UACUCUG	GAA A	CAGAAUCUACU AAGACAAGGCA AAAUGCCGUGU UUAUCUCGUCA ACUUGUUGGCG AGAUUUUUUU
cCas9-v21	NNV ACT; NNV ATG M;NN VATT ;NNV GCT; NNV GTG; NNV GTT	-3	2	21	21	GUCUUAG UACUCUG	GAA A	CAGAAUCUACU AAGACAAGGCA AAAUGCCGUGU UUAUCUCGUCA ACUUGUUGGCG AGAUUUUUUU
cCas9-v42	NNV RRN	-3	2	21	21	GUCUUAG UACUCUG	GAA A	CAGAAUCUACU AAGACAAGGCA AAAUGCCGUGU UUAUCUCGUCA ACUUGUUGGCG AGAUUUUUUU
CdiCas9	NNR HHH Y;NN RAA AY		2	22	22	ACUGGGG UUCAG	GAA A	CUGAACCUCAG UAAGCAUUGGC UCGUUUCCAAU GUUGAUUGCUC CGCCGGUGCUC

								CUUAUUUUUA AGGGCGCCGGC
CjeCas9	NNNNRYAC	-3	2	21	23	GTTTTAGTCCCT	GAAA	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTTT
GeoCas9	NNNNCRAA		2	21	23	GUCAUAGUUCCCCUGA	GAAA	UCAGGGUUACUAUGAUAAGGGCUUUCUGCCUAAGGCAGACUGACCCGCGGCGUUGGGGAUCGCCUGUCGCCCGCUUUUGGCGGGCAUUCCCCAUCCUU
iSpyMacCas9	NAAN	-3	2	19	21	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
NmeCas9	NNNNGAYT;NNNNGYTT;NNNNGAYA;NNNNGTCT	-3	2	20	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAG

								TCCGTTATCAA CTTGAAAAAGT GGCACCGAGTC GGTGC
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAG AGCTA	GAA A	TAGCAAGTTAA AATAAGGCTAG TCCGTTATCAA CTTGAAAAAGT GGCACCGAGTC GGTGC
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAG AGCTATG CTG	GAA A	CAGCATAGCAA GTTTAAATAAG GCTAGTCCGTT ATCAACTTGAA AAAGTGGCACC GAGTCGGTGC
SpyCas9-3var-NRTH	NRTH	-3	2	20	20	GTTTAAG AGCTATG CTG	GAA A	CAGCATAGCAA GTTTAAATAAG GCTAGTCCGTT ATCAACTTGAA AAAGTGGCACC GAGTCGGTGC
SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAG AGCTATG CTG	GAA A	CAGCATAGCAA GTTTAAATAAG GCTAGTCCGTT ATCAACTTGAA AAAGTGGCACC GAGTCGGTGC
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAG AGCTA	GAA A	TAGCAAGTTAA AATAAGGCTAG TCCGTTATCAA CTTGAAAAAGT GGCACCGAGTC GGTGC

SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas	NG;GAA;GAT	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAA

								AAAGTGGCACC GAGTCGGTGC
St1Cas9-CNRZ1066	NNA CAA	-3	2	20	20	GTCTTTG TACTCTG	GTAC	CAGAAGCTACA AAGATAAGGCT TCATGCCGAAA TCAACACCCTG TCATTTTATGG CAGGGTGTTTT
St1Cas9-LMG1831	NNG CAA	-3	2	20	20	GTCTTTG TACTCTG	GTAC	CAGAAGCTACA AAGATAAGGCT TCATGCCGAAA TCAACACCCTG TCATTTTATGG CAGGGTGTTTT
St1Cas9-MTH17CL396	NNA AAA	-3	2	20	20	GTCTTTG TACTCTG	GTAC	CAGAAGCTACA AAGATAAGGCT TCATGCCGAAA TCAACACCCTG TCATTTTATGG CAGGGTGTTTT
St1Cas9-TH1477	NNG AAA	-3	2	20	20	GTCTTTG TACTCTG	GTAC	CAGAAGCTACA AAGATAAGGCT TCATGCCGAAA TCAACACCCTG TCATTTTATGG CAGGGTGTTTT

В некоторых вариантах осуществления открывающая молекула является экзогенной по отношению к клетке, содержащей полипептид Gene Writer и/или нуклеиновую кислоту-матрицу. В некоторых вариантах осуществления открывающая молекула содержит эндогенное средство (например, эндогенное по отношению к клетке, содержащей полипептид Gene Writer и/или нуклеиновую кислоту-матрицу, содержащую gRNA и блокирующий домен). Например, индуцируемая gRNA, блокирующий домен и открывающая молекула могут быть выбраны таким образом, чтобы открывающая молекула представляла собой эндогенное средство, экспрессируемое в клетке или ткани, являющихся мишенями, например, за счет чего обеспечивается активность системы Gene Writer в клетке или ткани, являющихся мишенями. В качестве дополнительного примера, индуцируемая

gRNA, блокирующий домен и открывающая молекула могут быть выбраны таким образом, чтобы открывающая молекула отсутствовала или практически не экспрессировалась в одной или нескольких клетках или тканях, не являющихся мишенями, например, за счет чего обеспечивается то, что активность системы Gene Writer не проявляется или практически не проявляется в одной или нескольких клетках или тканях, не являющихся мишенями, или проявляется на сниженном уровне по сравнению с клеткой или тканью, являющимися мишенями. Иллюстративные блокирующие домены, открывающие молекулы и их пути применения описаны в публикации заявки согласно PCT WO2020044039A1, которая включена в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, может содержать одну или несколько UTR (например, из ретротранспозона типа R2) и gRNA. В некоторых вариантах осуществления UTR облегчает взаимодействие нуклеиновой кислоты-матрицы (например, РНК-матрицы) с записывающим доменом, например, доменом обратной транскриптазы, полипептида Gene Writer. В некоторых вариантах осуществления gRNA облегчает взаимодействие с доменом, связывающим нуклеиновую кислоту-матрицу (например, РНК-связывающим доменом) полипептида. В некоторых вариантах осуществления gRNA направляет полипептид к соответствующей последовательности-мишени, например, в геноме клетки-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица может содержать только мотив связывания обратной транскриптазы (например, 3'-UTR из R2), и gRNA может быть предоставлена в качестве второй молекулы нуклеиновой кислоты (например, второй молекулы РНК) для распознавания сайта-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица, содержащая RT-связывающий мотив, может существовать в той же молекуле, что и gRNA, но процессироваться с образованием двух молекул РНК благодаря активности расщепления (например, рибозима).

2. В некоторых вариантах осуществления РНК-матрица может быть специально разработана для коррекции заданной мутации в геномной ДНК клетки-мишени (например, ex vivo или in vivo, например, в ткани или органе, являющимися мишенями, например, у субъекта). Например, мутация может представлять собой мутацию, ассоциированную с заболеванием, по сравнению с последовательностью дикого типа. Не желая ограничиваться какой-либо теорией, полагают, что наборы эмпирических параметров помогают обеспечить оптимальные исходные конструкции РНК-матриц или их частей in silico. В качестве неограничивающего иллюстративного примера для выбранной мутации можно использовать следующие параметры разработки. В некоторых вариантах осуществления разработка начинается с получения примерно 500 п. о. (например, вплоть до 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 или 700 п. о. и необязательно по меньшей мере 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 или 650 п. о.) фланкирующей последовательности с каждой стороны от мутации, которые служат в качестве области-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица содержит gRNA. Методика разработки gRNA известна специалистам в данной

области. В некоторых вариантах осуществления gRNA содержит последовательность (например, спейсер CRISPR), которая связывает сайт-мишень. В некоторых вариантах осуществления последовательность (например, спейсер CRISPR), которая связывает сайт-мишень, для применения при нацеливании нуклеиновой кислоты-матрицы на область-мишень выбирают посредством рассмотрения конкретного используемого полипептида Gene Writer (например, эндонуклеазного домена или записывающего домена, например, содержащего домен CRISPR/Cas) (например, для Cas9, мотив, прилегающий к протоспейсеру (PAM), NGG непосредственно с 3'-стороны от gRNA-связывающей области размером в 20 нуклеотидов). В некоторых вариантах осуществления спейсер CRISPR выбирают посредством ранжирования сперва в зависимости от того, будет ли разрушен PAM в результате редактирования, индуцированного Gene Writing. В некоторых вариантах осуществления разрушение PAM может привести к увеличению эффективности редактирования. В некоторых вариантах осуществления PAM можно разрушить также посредством введения (например, в качестве части или в качестве дополнения другой модификации сайта-мишени в геномной ДНК) молчащей мутации (например, мутации, при которой не изменяется аминокислотный остаток, кодируемый последовательностью нуклеиновой кислоты-мишени, если таковой имеется) в сайт-мишень во время использования Gene Writing. В некоторых вариантах осуществления спейсер CRISPR выбирают посредством ранжирования последовательностей по близости их соответствующего геномного сайта к желаемому месту редактирования. В некоторых вариантах осуществления gRNA содержит каркас gRNA. В некоторых вариантах осуществления используемый каркас gRNA может представлять собой стандартный каркас (например, для Cas9, 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC-3') или может содержать одну или несколько нуклеотидных замен. В некоторых вариантах осуществления последовательность гетерологичного объекта характеризуется по меньшей мере 90% идентичностью, например, по меньшей мере 90%, 95%, 98%, 99% или 100% идентичностью, или содержит не более 1, 2, 3, 4 или 5 положений неидентичности по отношению к сайту-мишени с 3'-стороны от однонитевого разрыва в первой нити (например, непосредственно с 3'-стороны от однонитевого разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от однонитевого разрыва в первой нити), за исключением любой вставки, замены или делеции, которые могут быть записаны в сайт-мишень с помощью Gene Writer. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью характеризуется по меньшей мере 90% идентичностью, например, по меньшей мере 90%, 95%, 98%, 99% или 100% идентичностью, или содержит не более 1, 2, 3, 4 или 5 положений неидентичности по отношению к сайту-мишени с 5'-стороны от однонитевого разрыва в первой нити (например, непосредственно с 5'-стороны от однонитевого разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от однонитевого разрыва в первой нити).

В некоторых вариантах осуществления РНК-матрица может содержать последовательность gRNA, например, для направления GeneWriter к сайту-мишени, представляющему интерес. В некоторых вариантах осуществления РНК-матрица содержит (например, в направлении от 5'- к 3'-концу): (i) необязательно последовательность (например, спейсер CRISPR), которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, описанный в данном документе (например, полипептид Gene Writer или Cas), (iii) последовательность гетерологичного объекта и (iv) 5'-концевой домен гомологии и/или 3'-концевой домен гомологии.

В некоторых вариантах осуществления молекула нуклеиновой кислоты-матрицы содержит 5'-концевой домен гомологии и/или 3'-концевой домен гомологии. В некоторых вариантах осуществления 5'-концевой домен гомологии содержит последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с последовательностью нуклеиновой кислоты, содержащейся в молекуле нуклеиновой кислоты-мишени. В вариантах осуществления последовательность нуклеиновой кислоты в молекуле нуклеиновой кислоты-мишени расположена в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов от сайта-мишени для вставки (например, с 5'-стороны относительно него), например, для последовательности гетерологичного объекта, например, содержащейся в молекуле нуклеиновой кислоты-матрицы.

В некоторых вариантах осуществления 3'-концевой домен гомологии содержит последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с последовательностью нуклеиновой кислоты, содержащейся в молекуле нуклеиновой кислоты-мишени. В вариантах осуществления последовательность нуклеиновой кислоты в молекуле нуклеиновой кислоты-мишени расположена в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов от сайта-мишени для вставки (например, с 3'-стороны относительно него), например, для последовательности гетерологичного объекта, например, содержащейся в молекуле нуклеиновой кислоты-матрицы. В некоторых вариантах осуществления 5'-концевой домен гомологии является гетерологичным по отношению к остальной части молекулы нуклеиновой кислоты-матрицы. В некоторых вариантах осуществления 3'-концевой домен гомологии является гетерологичным по отношению к остальной части молекулы нуклеиновой кислоты-матрицы.

В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) содержит 3'-концевой домен гомологии с мишенью. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью расположен с 3'-стороны от последовательности гетерологичного объекта и комплементарен последовательности, прилегающей к сайту, подлежащему модификации с помощью системы, описанной в

данном документе, или содержит не более 1, 2, 3, 4 или 5 несовпадений с последовательностью, комплементарной последовательности, прилегающей к сайту, подлежащему модификации с помощью системы/Gene Writer™. В некоторых вариантах осуществления 3'-концевой домен гомологии связывается в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов от сайта однонитевого разрыва в молекуле нуклеиновой кислоты-мишени. В некоторых вариантах осуществления связывание 3'-концевого домена гомологии с молекулой нуклеиновой кислоты-мишени позволяет инициировать обратную транскрипцию с праймированием мишенью (TPRT), например, при этом 3'-концевой домен гомологии выступает в качестве праймера для TPRT.

Способы и композиции для модифицированной РНК (например, gRNA или РНК-матрицы)

В некоторых вариантах осуществления РНК-компонент системы (например, РНК-матрица или gRNA, например, описанные в данном документе) содержит одну или несколько модификаций нуклеотидов. В некоторых вариантах осуществления паттерн модификаций gRNA может значительно влиять на активность *in vivo* по сравнению с немодифицированными направляющими или направляющими с модифицированными концами (например, как показано на фигуре 1D в Finn et al. *Cell Rep* 22(9):2227-2235 (2018); включенном в данный документ посредством ссылки во всей своей полноте). Не желая ограничиваться какой-либо теорией, полагают, что данный процесс может быть обусловлен по меньшей мере частично стабилизацией РНК, обеспечиваемой модификациями. Неограничивающие примеры таких модификаций могут включать 2'-O-метил (2'-O-Me), 2'-O-(2-метоксиэтил) (2'-O-MOE), 2'-фтор (2'-F), фосфотиоатную (PS) связь между нуклеотидами, замены G-C и перевернутые связи, лишенные оснований, между нуклеотидами и их эквивалентами.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывает сайт-мишень) или направляющая РНК содержит 5'-концевую область. В некоторых вариантах осуществления РНК-матрица или направляющая РНК не содержит 5'-концевую область. В некоторых вариантах осуществления 5'-концевая область содержит спейсерную область CRISPR, например, как описано в отношении sgRNA в Briner AE et al, Molecular Cell 56: 333-339 (2014) (включенном в данный документ посредством ссылки во всей своей полноте; применимо в данном документе, например, ко всем направляющим РНК). В некоторых вариантах осуществления 5'-концевая область содержит модификацию 5'-конца. В некоторых вариантах осуществления 5'-концевая область со спейсерной областью или без нее может быть ассоциирована с crRNA, trRNA, sgRNA и/или dgRNA. Спейсерная область CRISPR может в некоторых случаях содержать направляющую область, направляющий домен или нацеливающий домен. В некоторых вариантах осуществления домен-мишень или последовательность-мишень могут содержать последовательность нуклеиновой кислоты, к которой направляющая область/домен направляет нуклеазу для расщепления. В некоторых вариантах осуществления белок spyCas9 может быть направлен направляющей областью/доменом к последовательности-

мишени молекулы нуклеиновой кислоты-мишени с помощью нуклеотидов, присутствующих в спейсерной области CRISPR.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывает сайт-мишень) или направляющая РНК, например, описанная в данном документе, содержит любую из последовательностей, показанных в таблице 4 документа WO2018107028A1, включенного в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления, где последовательность содержит направляющую и/или спейсерную область, композиция может содержать или не содержать данную область. В некоторых вариантах осуществления направляющая РНК содержит одну или несколько модификаций любой из последовательностей, показанных в таблице 4 документа WO2018107028A1, например, идентифицированных в нем с помощью SEQ ID NO. В вариантах осуществления нуклеотиды могут быть одинаковыми или разными, и/или показанный паттерн модификаций может быть таким же, как паттерн модификаций направляющей последовательности, показанной в таблице 4 документа WO2018107028A1, или сходным с ним. В некоторых вариантах осуществления паттерн модификаций включает относительное положение и отличительные особенности модификаций gRNA или области gRNA (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области). В некоторых вариантах осуществления паттерн модификаций содержит по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% модификаций в любой из последовательностей, показанных в столбце "Последовательности" таблицы 4 в документе WO2018107028A1, и/или в одной или нескольких областях последовательности. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным паттерну модификаций любой из последовательностей, показанных в столбце "Последовательности" таблицы 4 в документе WO2018107028A1. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в одной или нескольких областях последовательности, показанной в таблице 4 в документе WO2018107028A1, например, в 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2 и/или 3'-концевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным паттерну модификаций последовательности в 5'-концевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в нижней стеблевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в области выпетливания. В некоторых вариантах осуществления паттерн

модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в верхней стеблевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в связующей области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в шпилечной области 1. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в шпилечной области 2. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в 3'-концевой области. В некоторых вариантах осуществления паттерн модификаций отличается от паттерна модификаций последовательности из таблицы 4 документа WO 2018107028A1 или области (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области) такой последовательности, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам. В некоторых вариантах осуществления gRNA содержит модификации, которые отличаются от модификаций последовательности из таблицы 4 документа WO 2018107028A1, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам. В некоторых вариантах осуществления паттерн модификаций отличается от паттерна модификаций последовательности из таблицы 4 документа WO 2018107028A1 или области (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области) такой последовательности, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывается с сайтом-мишенью) или gRNA содержит 2'-O-метил- (2'-O-Me-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит 2'-O-(2-метоксиэтил-) (2'-O-MOE-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит 2'-фтор- (2'-F-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит фосфотиоатную (PS) связь между нуклеотидами. В некоторых вариантах осуществления gRNA содержит модификацию 5'-конца, модификацию 3'-конца или модификации 5'- и 3'-концов. В некоторых вариантах осуществления модификация 5'-конца включает в себя фосфотиоатную (PS) связь между нуклеотидами. В некоторых вариантах осуществления модификация 5'-конца включает в себя 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модифицированный нуклеотид. В некоторых вариантах осуществления модификация 5'-конца включает в себя по меньшей мере одну фосфотиоатную (PS) связь и один или несколько 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модифицированных нуклеотидов. Модификация конца может включать в себя

фосфотиоатную (PS), 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модификацию. Эквивалентные модификации конца также охватываются вариантами осуществления, описанными в данном документе. В некоторых вариантах осуществления РНК-матрица или gRNA содержит модификацию конца в комбинации с модификацией одной или нескольких областей РНК-матрицы или gRNA. Дополнительные иллюстративные модификации и способы для защиты РНК, например, gRNA, и их формулы описаны в документе WO2018126176A1, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления для введения модификаций (например, модификаций 2'-OMe-РНК, 2'-F-РНК и PS) в РНК-матрицу или направляющую РНК используются структурно-ориентированные и систематические подходы, например, описанные в Mir et al. *Nat Commun* 9:2641 (2018) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления включение 2'-F-РНК приводит к повышению термической и нуклеазной стабильности дуплексов РНК:РНК или РНК:ДНК, например, минимально препятствуя излому плоскости сахарного кольца в конформации C3'-эндо. В некоторых вариантах осуществления 2'-F может лучше переноситься, чем 2'-OMe, в положениях, где 2'-OH важен для стабильности дуплекса РНК:ДНК. В некоторых вариантах осуществления crRNA содержит одну или несколько модификаций, которые не приводят к снижению активности Cas9, например, C10, C20 или C21 (полностью модифицированные), например, как описано в дополнительной таблице 1 в Mir et al. *Nat Commun* 9:2641 (2018), включенном в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления tracrRNA содержит одну или несколько модификаций, которые не приводят к снижению активности Cas9, например, T2, T6, T7 или T8 (полностью модифицированные) из дополнительной таблицы 1 в Mir et al. *Nat Commun* 9:2641 (2018). В некоторых вариантах осуществления crRNA, содержащая одну или несколько модификаций (например, описанных в данном документе), может быть спарена с tracrRNA, содержащей одну или несколько модификаций, например, C20 и T2. В некоторых вариантах осуществления gRNA содержит химерное соединение, например, из crRNA и tracrRNA (например, в Jinek et al. *Science* 337(6096):816-821 (2012)). В вариантах осуществления модификации из crRNA и tracrRNA картируются на химерное соединение с одной направляющей молекулой, например, для получения модифицированной gRNA с повышенной стабильностью.

В некоторых вариантах осуществления молекулы gRNA можно модифицировать посредством добавления или удаления встречающихся в природе структурных компонентов, например, шпилек. В некоторых вариантах осуществления gRNA может содержать gRNA с одним или несколькими удаленными 3'-шпилечными элементами, например, как описано в документе WO2018106727, включенном в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления gRNA может содержать дополнительную шпилечную структуру, например, дополнительную шпилечную структуру в спейсерной области, которая, как было показано, приводит к

повышению специфичности системы CRISPR-Cas согласно идеям Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019). Дополнительные модификации, в том числе примеры укороченной gRNA и специфичных модификаций, приводящих к улучшению активности in vivo, можно найти в документе US20190316121, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления структурно-ориентированный и систематический подходы (например, описанные в Mir et al. *Nat Commun* 9:2641 (2018); включенном в данный документ посредством ссылки во всей своей полноте) используются для поиска модификаций для РНК-матрицы. В вариантах осуществления модификации идентифицируют по включению или исключению направляющей области РНК-матрицы. В некоторых вариантах осуществления структуру полипептида, связанного с РНК-матрицей, используют для определения не контактирующих с белком нуклеотидов в РНК, которые затем могут быть выбраны для модификаций, например, с более низким риском разрушения ассоциации РНК с полипептидом. Вторичные структуры в РНК-матрице также можно предсказать in silico с помощью программных инструментов, например, инструмента RNAstructure, доступного по адресу rna.urmc.rochester.edu/RNAstructureWeb (Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); включенный в данный документ посредством ссылки во всей своей полноте), например, для определения вторичных структур для выбора модификаций, например, шпилек, стеблей и/или выпетливаний.

Предполагается, что может быть полезно использовать кольцевые и/или линейные состояния РНК во время составления, доставки или реакции Gene Writing в клетке-мишени. Таким образом, в некоторых вариантах осуществления любого из аспектов, описанных в данном документе, система Gene Writing содержит одну или несколько кольцевых РНК (circRNA). В некоторых вариантах осуществления любого из аспектов, описанных в данном документе, система Gene Writing содержит одну или несколько линейных РНК. В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, нуклеиновая кислота-матрица, молекула нуклеиновой кислоты, кодирующая полипептид Gene Writer, или и то, и другое), представляет собой circRNA. В некоторых вариантах осуществления кольцевая молекула РНК кодирует полипептид Gene Writer. В некоторых вариантах осуществления молекула circRNA, кодирующая полипептид Gene Writer, доставляется в клетку-хозяина. В некоторых вариантах осуществления кольцевая молекула РНК кодирует рекомбиназу, например, описанную в данном документе. В некоторых вариантах осуществления молекула circRNA, кодирующая рекомбиназу, доставляется в клетку-хозяина. В некоторых вариантах осуществления молекула circRNA, кодирующая полипептид Gene Writer, линеаризуется (например, в клетке-хозяине, например, в ядре клетки-хозяина) перед трансляцией.

Было обнаружено, что кольцевые РНК (circRNA) в естественных условиях встречаются в клетках и выполняют разнообразные функции, включающие как не связанные с кодированием, так и связанные с кодированием белка роли, в клетках человека. Было показано, что circRNA может быть сконструирована путем включения

самовырезающегося путем сплайсинга интрона в молекулу РНК (или ДНК, кодирующую молекулу РНК), что приводит в результате к циркуляризации РНК, и сконструированная таким образом circRNA может характеризоваться повышенными продуцированием белка и стабильностью (Wesselhoeft et al. *Nature Communications* 2018). В некоторых вариантах осуществления полипептид Gene Writer™ кодируется в виде circRNA. В некоторых вариантах осуществления нуклеиновая кислота-матрица представляет собой ДНК, такую как dsDNA или ssDNA.

В некоторых вариантах осуществления circRNA содержит одну или несколько последовательностей рибозима. В некоторых вариантах осуществления последовательность рибозима активируется для саморасщепления, например, в клетке-хозяине, например, что таким образом приводит к линеаризации circRNA. В некоторых вариантах осуществления рибозим активируется, когда концентрация магния достигает уровня, достаточного для расщепления, например, в клетке-хозяине. В некоторых вариантах осуществления circRNA поддерживается в среде с низким содержанием магния перед доставкой в клетку-хозяина. В некоторых вариантах осуществления рибозим представляет собой рибозим, отвечающий на белки. В некоторых вариантах осуществления рибозим представляет собой рибозим, отвечающий на нуклеиновые кислоты. В некоторых вариантах осуществления circRNA содержит сайт расщепления. В некоторых вариантах осуществления circRNA содержит второй сайт расщепления.

В некоторых вариантах осуществления circRNA линеаризуется в ядре клетки-мишени. В некоторых вариантах осуществления линеаризация circRNA в ядре клетки включает компоненты, присутствующие в ядре клетки, например, для активации события расщепления. Например, ретротранспозоны B2 и ALU содержат саморасщепляющиеся рибозимы, активность которых усиливается при взаимодействии с белком Polycomb EZH2 (Hernandez et al. PNAS 117(1):415-425 (2020)). Таким образом, в некоторых вариантах осуществления рибозим, например, рибозим из элемента B2 или ALU, который отвечает на ядерный элемент, например, ядерный белок, например, белок, взаимодействующий с геномом, например, эпигенетический модификатор, например, EZH2, включен в circRNA, например, системы Gene Writing. В некоторых вариантах осуществления ядерная локализация circRNA приводит к увеличению автокаталитической активности рибозима и линеаризации circRNA.

В некоторых вариантах осуществления рибозим гетерологичен одному или нескольким другим компонентам системы Gene Writing. В некоторых вариантах осуществления индуцируемый рибозим (например, в circRNA, описанной в данном документе) создается синтетически, например, с использованием конструкции аптамера, отвечающего на белковый лиганд. Была описана система для использования сателлитной РНК рибозима типа "головки молотка" вируса кольцевой пятнистости табака с аптамером белка оболочки MS2 (Kennedy et al. Nucleic Acids Res 42(19):12306-12321 (2014), включенный в данный документ посредством ссылки во всей своей полноте), что приводит к активации активности рибозима в присутствии белка оболочки MS2. В вариантах

осуществления такая система отвечает на белковый лиганд, локализованный в цитоплазме или ядре. В некоторых вариантах осуществления белковый лиганд не представляет собой MS2. Способы получения РНК-аптамеров для лигандов-мишеней были описаны, например, на основе систематической эволюции лигандов экспоненциальным обогащением (SELEX) (Tuerk and Gold, Science 249(4968):505-510 (1990); Ellington and Szostak, Nature 346(6287):818-822 (1990); при этом способы из каждого из них включены в данный документ посредством ссылки), и в некоторых случаях им способствовала разработка in silico (Bell et al. PNAS 117(15):8486-8493, способы из которого включены в данный документ посредством ссылки). Таким образом, в некоторых вариантах осуществления аптамер для лиганда-мишени создается и включается в синтетическую рибозимную систему, например, для запуска опосредованного рибозимом расщепления и линеаризации circRNA, например, в присутствии белкового лиганда. В некоторых вариантах осуществления линеаризация circRNA запускается в цитоплазме, например, с использованием аптамера, который ассоциирует с лигандом в цитоплазме. В некоторых вариантах осуществления линеаризация circRNA запускается в ядре, например, с использованием аптамера, который ассоциирует с лигандом в ядре. В вариантах осуществления лиганд в ядре предусматривает эпигенетический модификатор или фактор транскрипции. В некоторых вариантах осуществления лиганд, который запускает линеаризацию, присутствует в целевых клетках на более высоких уровнях, чем в нецелевых клетках.

Дополнительно предполагается, что для линеаризации circRNA можно использовать рибозимную систему, отвечающую на нуклеиновые кислоты. Например, биосенсоры, которые воспринимают определенные молекулы нуклеиновой кислоты-мишени для запуска активации рибозима, описаны, например, в Penchovsky (Biotechnology Advances 32(5):1015-1027 (2014), включенном в данный документ посредством ссылки). С помощью данных способов рибозим естественным образом сворачивается в неактивное состояние и активируется только в присутствии определенной молекулы нуклеиновой кислоты-мишени (например, молекулы РНК). В некоторых вариантах осуществления circRNA системы Gene Writing содержит рибозим, отвечающий на нуклеиновые кислоты, который активируется в присутствии определенной нуклеиновой кислоты-мишени, например, РНК, например, mRNA, miRNA, направляющей РНК, gRNA, sgRNA, ncRNA, lncRNA, тРНК, snRNA или mtRNA. В некоторых вариантах осуществления нуклеиновая кислота, которая запускает линеаризацию, присутствует в целевых клетках на более высоких уровнях, чем в нецелевых клетках.

В некоторых вариантах осуществления любого из аспектов в данном документе система Gene Writing включает один или несколько рибозимов с индуцируемой специфичностью в отношении ткани-мишени или клетке-мишени, представляющим интерес, например, рибозим, который активируется лигандом или нуклеиновой кислотой, присутствующими на более высоких уровнях в ткани-мишени или клетке-мишени, представляющих интерес. В некоторых вариантах осуществления система Gene Writing

включает рибозим с индуцируемой специфичностью в отношении субклеточного компартмента, например, ядра, ядрышка, цитоплазмы или митохондрий. В некоторых вариантах осуществления рибозим, который активируется лигандом или нуклеиновой кислотой, присутствует на более высоких уровнях в целевом субклеточном компартменте. В некоторых вариантах осуществления РНК-компонент системы Gene Writing представлен в виде circRNA, например, которая активируется путем линеаризации. В некоторых вариантах осуществления линеаризация circRNA, кодирующей полипептид Gene Writing, активирует молекулу для трансляции. В некоторых вариантах осуществления сигнал, который активирует circRNA-компонент системы Gene Writing, присутствует на более высоких уровнях в целевых клетках или тканях, например, так что система специфично активируется в данных клетках.

В некоторых вариантах осуществления РНК-компонент системы Gene Writing представлен в виде circRNA, которая инактивируется путем линеаризации. В некоторых вариантах осуществления circRNA, кодирующая полипептид Gene Writer, инактивируется путем расщепления и разрушения. В некоторых вариантах осуществления circRNA, кодирующая полипептид Gene Writing, инактивируется путем расщепления, при котором сигнал трансляции отделяется от последовательности, кодирующей полипептид. В некоторых вариантах осуществления сигнал, который инактивирует circRNA-компонент системы Gene Writing, присутствует на более высоких уровнях в нецелевых клетках или тканях, так что система специфично инактивируется в данных клетках.

Дополнительно в данный документ включены композиции и способы сборки полных или частичных молекул РНК-матриц (например, молекул РНК-матриц для Gene Writing, необязательно содержащих gRNA, или отдельных молекул gRNA). В некоторых вариантах осуществления молекулы РНК могут быть собраны посредством соединения двух или более (например, двух, трех, четырех, пяти, шести, семи, восьми, девяти, десяти или более) сегментов РНК друг с другом. В одном аспекте в настоящем изобретении предусмотрены способы получения молекул нуклеиновых кислот, при этом способы включают приведение двух или более линейных сегментов РНК в контакт друг с другом в условиях, которые позволяют 5'-концу первого сегмента РНК ковалентно связываться с 3'-концом второго сегмента РНК. В некоторых вариантах осуществления присоединенная молекула может контактировать с третьим сегментом РНК в условиях, позволяющих 5'-концу присоединенной молекулы ковалентно связываться с 3'-концом третьего сегмента РНК. В вариантах осуществления способ дополнительно включает присоединение четвертого, пятого или дополнительных сегментов РНК к удлиненной молекуле. Данная форма сборки может в некоторых случаях обеспечивать быструю и эффективную сборку молекул РНК.

В настоящем изобретении также предусмотрены композиции и способы для соединения (например, ковалентного соединения) молекул crRNA и молекул tracrRNA. В некоторых вариантах осуществления молекулы направляющей РНК со специфичностью в отношении различных сайтов-мишеней могут быть получены с использованием одной

молекулы/сегмента tracrRNA, соединенной с молекулой/сегментом crRNA, специфичной в отношении сайта-мишени (например, как показано на фиг. 10 в US20160102322A1; включенном в данный документ посредством ссылки во всей своей полноте). Например, на фиг. 10 документа US20160102322A1 показаны четыре пробирки с различными молекулами crRNA, где молекула 3 crRNA соединена с молекулой tracrRNA с образованием молекулы направляющей РНК, что таким образом изображает иллюстративное соединение двух сегментов РНК с образованием молекулы РНК-продукта.

В настоящем изобретении также предусмотрены композиции и способы для получения молекул РНК-матриц со специфичностью в отношении полипептида Gene Writer и/или геномного сайта-мишени. В одном аспекте способ включает: (1) идентификацию сайта-мишени и желаемой его модификации, (2) получение сегментов РНК, включая вышерасположенный сегмент гомологии, сегмент последовательности гетерологичного объекта, мотив связывания полипептида Gene Writer и сегмент gRNA, и/или (3) соединение четырех или более сегментов в по меньшей мере одну молекулу, например, в одну молекулу РНК. В некоторых вариантах осуществления все или некоторые сегменты РНК-матрицы, содержащиеся в (2), собраны в молекулу РНК-матрицы, например, один, два, три или четыре из перечисленных компонентов. В некоторых вариантах осуществления сегменты, содержащиеся в (2), могут быть получены в дополнительно сегментированных молекулах, например, разделенных на по меньшей мере 2, на по меньшей мере 3, на по меньшей мере 4 или на по меньшей мере 5 или более подсегментов, например, которые впоследствии собраны, например, одним или несколькими способами, описанными в данном документе.

В некоторых вариантах осуществления сегменты РНК могут быть получены с помощью химического синтеза. В некоторых вариантах осуществления сегменты РНК могут быть получены с помощью транскрипции нуклеиновой кислоты-матрицы in vitro, например, с помощью предоставления РНК-полимеразы для воздействия на когнатный промотор ДНК-матрицы для получения РНК-транскрипта. В некоторых вариантах осуществления транскрипцию in vitro проводят с использованием, например, РНК-полимеразы T7, T3 или SP6 или ее производного, действующего на ДНК, например, dsDNA, ssDNA, линейную ДНК, плазмидную ДНК, ампликон линейной ДНК, линеаризованную плазмидную ДНК, например, кодирующую сегмент РНК, например, под транскрипционным контролем когнатного промотора, например, промотора T7, T3 или SP6. В некоторых вариантах осуществления используется комбинация химического синтеза и транскрипции in vitro для получения сегментов РНК для сборки. В вариантах осуществления gRNA, вышерасположенный сегмент гомологии с мишенью и сегмент, связывающий полипептид Gene Writer, получают с помощью химического синтеза, и сегмент последовательности гетерологичного объекта получают с помощью транскрипции in vitro. Не желая ограничиваться какой-либо теорией, полагают, что для получения более длинных молекул РНК может лучше подходить транскрипция in vitro. В некоторых вариантах осуществления температура реакции для транскрипции in vitro может быть снижена, например, до менее 37°C (например, между 0-10°C, 10-20°C или 20-30°C) для

получения более высокой доли полноразмерных транскриптов (Krieg *Nucleic Acids Res* 18:6463 (1990)). В некоторых вариантах осуществления протокол для улучшенного синтеза длинных транскриптов используется для синтеза длинной РНК-матрицы, например, РНК-матрицы размером более 5 т. о., например, с использованием T7 RiboMAX Express, который может образовывать транскрипты размером 27 т. о. in vitro. (Thiel et al. J Gen Virol 82(6):1273-1281 (2001)). В некоторых вариантах осуществления модификации молекул РНК, описанные в данном документе, могут быть включены во время синтеза сегментов РНК (например, посредством включения модифицированных нуклеотидов или альтернативных связывающих химических объектов), после синтеза сегментов РНК с помощью химических или ферментативных процессов, после сборки одного или нескольких сегментов РНК или при их комбинации.

В некоторых вариантах осуществления мРНК системы (например, мРНК, кодирующую полипептид Gene Writer) синтезируют in vitro с использованием ДНК-зависимой транскрипции РНК, опосредованной Т7-полимеразой, из линеаризованной ДНК-матрицы, где UTP необязательно замещен 1-метилпсевдо-UTP. В некоторых вариантах осуществления транскрипт содержит 5'- и 3'-UTR, например, GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC и UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA, или их функциональные фрагменты или варианты и необязательно содержит поли(А)-хвост, который может быть закодирован в ДНК-матрице или ферментативно добавлен после транскрипции. В некоторых вариантах осуществления донорная метильная группа, например, S-аденозилметионин, добавляется к метилированной кэпированной РНК со структурой кэп 0 для получения структуры кэп 1, которая повышает эффективность трансляции мРНК (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

В некоторых вариантах осуществления транскрипт от промотора Т7 начинается с мотива GGG. В некоторых вариантах осуществления транскрипт от промотора Т7 не начинается с мотива GGG. Было показано, что мотив GGG в сайте начала транскрипции, несмотря на обеспечение лучшего выхода, может приводить к синтезу под действием RNAP Т7 лэддера поли(G)-продуктов в результате проскальзывания транскрипта по трем остаткам С в нити матрицы от +1 до +3 (Imburgio et al. Biochemistry 39(34):10419-10430 (2000). Для настройки уровней транскрипции и изменения нуклеотидов в сайте начала транскрипции таким образом, чтобы они соответствовали альтернативным 5'-UTR, в идеях Davidson et al. *Pac Symp Biocomput* 433-443 (2010) описаны варианты промотора Т7 и способы их обнаружения, которые соответствуют обоим этим признакам.

В некоторых вариантах осуществления сегменты РНК могут быть соединены друг с другом посредством ковалентной связи. В некоторых вариантах осуществления РНК-лигаза, например, РНК-лигаза Т4, может использоваться для соединения двух или более сегментов РНК друг с другом. При использовании реагента, такого как РНК-лигаза, 5'-конец, как правило, связывают с 3'-концом. В некоторых вариантах осуществления, если

два сегмента соединены, то могут быть образованы две возможные линейные конструкции (т. е. (1) 5'-сегмент 1-сегмент 2-3' и (2) 5'-сегмент 2-сегмент 1-3′). В некоторых вариантах осуществления также может иметь место внутримолекулярная циркуляризация. Обе эти проблемы можно решить, например, блокируя один 5'-конец или один 3'-конец, чтобы РНК-лигаза не могла лигировать этот конец с другим концом. В вариантах осуществления, если конструкция 5'-сегмент 1-сегмент 2-3' является желаемой, то размещение блокирующей группы либо на 5'-конце сегмента 1, либо на 3'-конце сегмента 2 может привести к образованию только правильного линейного продукта лигирования и/или предотвратить внутримолекулярную циркуляризацию. Композиции и способы для ковалентного соединения двух сегментов нуклеиновой кислоты (например, РНК) раскрыты, например, в US20160102322A1 (включенном в данный документ посредством ссылки во всей своей полноте), наряду со способами, включающими применение РНК-лигазы для направленного лигирования двух однонитевых сегментов РНК друг с другом.

Одним из примеров блокатора концов, который можно применять в сочетании, например, с РНК-лигазой Т4, является средство для дидезокси-терминации. РНК-лигаза Т4, как правило, катализирует АТР-зависимое лигирование фосфодиэфирных связей между 5'-фосфатным и 3'-гидроксильным концами. В некоторых вариантах осуществления при применении РНК-лигазы Т4 на концах, подвергаемых лигированию, должны присутствовать подходящие концы. Один из способов блокирования РНК-лигазы Т4 на конце предусматривает отсутствие правильного формата конца. Обычно концы сегментов РНК с 5-гидроксилом или 3'-фосфатом не будут выступать в качестве субстратов для РНК-лигазы Т4.

Дополнительные иллюстративные способы, которые можно применять для соединения сегментов РНК, включают клик-химию (например, описанную в патентах США №№ 7375234 и 7070941 и публикации заявки на патент США № 2013/0046084, полное раскрытие которых включено в данный документ посредством ссылки). Например, одна иллюстративная реакция клик-химии происходит между алкиновой группой и азидной группой (см. фиг. 11 документа US20160102322A1, который включен в данный документ посредством ссылки во всей своей полноте). Любая клик-реакция потенциально может быть использована для связывания сегментов РНК (например, Cu-катализируемое азид-алкиновое циклоприсоединение, промотируемое напряжением азид-алкиновое циклоприсоединение, лигирование по Штаудингеру, лигирование тетразином, фотоиндуцированная тетразол-алкеновая реакция, тиол-еновая реакция, использование сложных NHS-эфиров, эпоксидов, изоцианатов и реакция альдегид-аминоокси). В некоторых вариантах осуществления лигирование молекул РНК с использованием реакции клик-химии является преимущественным, поскольку реакции клик-химии являются быстрыми, модульными, эффективными, часто не приводят к получению токсичных побочных продуктов, могут проводиться с использованием воды в качестве растворителя и/или могут быть настроены так, чтобы они были стереоспецифическими.

В некоторых вариантах осуществления сегменты РНК могут быть соединены с использованием азид-алкинового циклоприсоединения Хьюсгена - реакции, которая, как правило, представляет собой 1,3-диполярное циклоприсоединение между азидом и концевым или внутренним алкином с получением 1,2,3-триазола для лигирования сегментов РНК. Не желая ограничиваться какой-либо теорией, полагают, что одним из преимуществ данного способа лигирования может быть то, что эта реакция может быть инициирована посредством добавления необходимых ионов Cu(I). Другие иллюстративные механизмы, с помощью которых сегменты РНК могут быть соединены, включают без ограничения использование реакций присоединения галогенов (F-, Br-, I-) и алкинов, карбонилов/сульфгидрилов/малеимида и связей между карбоксильными и аминогруппами. Например, одна молекула РНК может быть модифицирована с помощью тиола в положении 3' (с использованием дисульфидамидита и универсальной подложки или подложки, модифицированной дисульфидом), и другая молекула РНК может быть модифицирована акридитом в положении 5' (с использованием акрилового фосфорамидита), тогда две РНК молекулы могут быть соединены с помощью реакции присоединения Михаэля. Такая стратегия также может быть применена для поэтапного соединения нескольких молекул РНК. Также предусмотрены способы связывания более чем двух (например, трех, четырех, пяти, шести и т. д.) молекул РНК друг с другом. Не желая ограничиваться какой-либо теорией, полагают, что это может быть полезно в случае, если желаемая молекула РНК длиннее, чем приблизительно 40 нуклеотидов, например, так что снижается эффективность химического синтеза, например, как указано в документе US20160102322A1 (включенном в данный документ посредством ссылки во всей своей полноте).

В качестве иллюстрации, длина tracrRNA, как правило, составляет около 80 нуклеотидов. Такие молекулы РНК могут быть получены, например, с помощью таких процессов, как транскрипция in vitro или химический синтез. В некоторых вариантах осуществления, когда для получения таких молекул РНК используется химический синтез, они могут быть получены в качестве единого продукта синтеза или посредством связывания двух или более синтезированных сегментов РНК друг с другом. В вариантах осуществления, когда три или более сегмента РНК соединены друг с другом, можно применять разные способы для связывания отдельных сегментов вместе. Также сегменты РНК могут быть соединены друг с другом в одном сосуде (например, контейнере, колбе, лунке, пробирке, планшете или другом резервуаре) все одновременно, или в одном сосуде в разное время, или в разных сосудах в разное время. В неограничивающем примере для сборки сегментов 1, 2 и 3 РНК в порядке номеров сегменты 1 и 2 РНК могут быть сначала соединены друг с другом 5'-конец к 3'-концу. Затем продукт реакции можно очистить от компонентов реакционной смеси (например, с помощью хроматографии), затем поместить во второй сосуд для соединения 3'-конца с 5'-концом сегмента 3 РНК. Затем конечный продукт реакции можно соединить с 5'-концом сегмента 3 РНК.

В другом неограничивающем примере сегмент 1 РНК (приблизительно 30 нуклеотидов) представляет собой последовательность распознавания локуса-мишени

crRNA и часть шпилечной области 1. Сегмент 2 РНК (приблизительно 35 нуклеотидов) содержит оставшуюся часть шпилечной области 1 и часть линейной tracrRNA между шпилечной областью 1 и шпилечной областью 2. Сегмент 3 РНК (приблизительно 35 нуклеотидов) содержит оставшуюся часть линейной tracrRNA между шпилечной областью 1 и шпилечной областью 2 и всю шпилечную область 2. В данном примере сегменты 2 и 3 РНК связаны 5'-конец к 3'-концу с помощью клик-химии. Дополнительно, оба из 5'- и 3'-концов продукта реакции являются фосфорилированными. Затем продукт реакции приводят в контакт с сегментом 1 РНК, имеющим 3'-концевую гидроксильную группу, и с РНК-лигазой Т4 для получения молекулы направляющей РНК.

Для соединения сегментов РНК в соответствии со способом по настоящему изобретению можно использовать ряд дополнительных связывающих химических соединений. Некоторые из этих химических соединений представлены в таблице 6 документа US20160102322A1, который включен в данный документ посредством ссылки во всей своей полноте.

ДНК-связывающий домен

В определенных аспектах ДНК-связывающий домен полипептида Gene Writer, описанного в данном документе, выбран, разработан или сконструирован для связывания с требуемой последовательностью ДНК-мишени хозяина. В определенных вариантах осуществления ДНК-связывающий домен сконструированного ретротранспозона представляет собой ДНК-связывающий белок или домен, гетерологичный по отношению к нативной последовательности ретротранспозона. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент представляет собой элемент "цинковый палец" или TAL-эффекторный элемент, например, полипептид "цинковый палец" или TAL или их функциональный фрагмент. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент представляет собой ДНК-связывающий элемент, направляемый последовательностью, такой как Cas9, Cpf1 или другой белок, связанный с CRISPR, который был подвергнут изменению таким образом, чтобы у него отсутствовала эндонуклеазная активность. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент сохраняет эндонуклеазную активность. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент замещает эндонуклеазный элемент полипептида. В конкретных вариантах осуществления гетерологичный ДНК-связывающий домен может представлять собой любые один или несколько из Cas9 (например, Cas9, никазы Cas9, dCas9), домена TAL, домена "цинковый палец" (ZF), домена Myb, их комбинаций или их множеств. В определенных вариантах осуществления гетерологичный ДНК-связывающий домен представляет собой ДНК-связывающий домен ретротранспозона, описанного в таблице X, таблице Z1, таблице Z2 или таблице 3A или 3B. В некоторых вариантах осуществления ДНК-связывающие домены могут быть идентифицированы на основании гомологии по отношению к другим известным ДНК-связывающим доменам с помощью инструментов, таких как средство поиска основного локального выравнивания (BLAST). В еще нескольких других вариантах осуществления

ДНК-связывающие домены являются модифицированными, например, посредством сайт-специфической мутации, обеспечивающей увеличение или уменьшение количества ДНК-связывающих элементов (например, числа и/или специфичности "цинковых пальцев") и т. д., с изменением специфичности и сродства связывания с ДНК. В некоторых вариантах осуществления ДНК-связывающий домен изменен по сравнению с его природной последовательностью таким образом, что он характеризуется измененной частотой использования кодонов, например, улучшенной для применения в клетках человека.

В некоторых вариантах осуществления ДНК-связывающий домен содержит мегануклеазный домен (например, как описано в данном документе, например, в разделе "Эндонуклеазный домен") или его функциональный фрагмент. В некоторых вариантах осуществления мегануклеазный домен обладает эндонуклеазной активностью, например, активностью двухнитевого расщепления и/или никазной активностью. В других вариантах осуществления мегануклеазный домен характеризуется сниженной активностью, например, у него отсутствует эндонуклеазная активность, например, мегануклеаза является каталитически неактивной. В некоторых вариантах осуществления в качестве ДНК-связывающего домена используется каталитически неактивная мегануклеаза, например, описанная в Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012), включенном в данный документ посредством ссылки во всей своей полноте. В вариантах осуществления ДНК-связывающий домен содержит одну или несколько модификаций по сравнению с ДНК-связывающим доменом дикого типа, например, модификацию посредством направленной эволюции, например, непрерывной эволюции с помощью фага (PACE).

В определенных аспектах настоящего изобретения связывающий сайт ДНК хозяина, в который производится интеграция с помощью системы Gene Writer, может находиться в гене, в интроне, в экзоне, в ORF, за пределами кодирующей области любого из генов, в регуляторной области гена или за пределами регуляторной области гена. В других аспектах сконструированный ретротранспозон может связываться с одной или с более чем одной последовательностью ДНК хозяина. В других аспектах сконструированный ретротранспозон может иметь низкую специфичность в отношении последовательности, например, связываться с несколькими последовательностями или не иметь предпочтения в отношении последовательности.

В некоторых вариантах осуществления для редактирования локуса-мишени в нескольких аллелях используется система Gene Writing. В некоторых вариантах осуществления система Gene Writing предназначена для редактирования конкретного аллеля. Например, полипептид Gene Writing может быть направлен на конкретную последовательность, которая присутствует только в одном аллеле, например, содержит РНК-матрицу, гомологичную аллелю-мишени, например, gRNA или домен отжига, но не второму когнатному аллелю. В некоторых вариантах осуществления система Gene Writing может изменять аллель, специфичный для гаплотипа. В некоторых вариантах осуществления система Gene Writing, которая нацелена на конкретный аллель,

предпочтительно нацелена на этот аллель, например, обладает по меньшей мере 2-, 4-, 6-, 8- или 10-кратным предпочтением к аллелю-мишени.

В определенных вариантах осуществления РНК системы редактирования генов Gene Writer™ дополнительно содержит последовательность внутриклеточной локализации, например, последовательность ядерной локализации. Последовательность ядерной локализации может представлять собой последовательность РНК, способствующую импорту РНК в ядро. В определенных вариантах осуществления сигнал ядерной локализации расположен в РНК-матрице. В определенных вариантах осуществления полипептид ретротранспозаза кодируется первой РНК, а РНК-матрица представляет собой вторую, отдельную РНК, и сигнал ядерной локализации расположен в РНК-матрице, а не в РНК, кодирующей полипептид ретротранспозазу. Не желая ограничиваться какой-либо теорией, полагают, что в некоторых вариантах осуществления РНК, кодирующая ретротранспозазу, нацеливается главным образом на цитоплазму для способствования ее трансляции, тогда как РНК-матрица нацеливается главным образом на ядро для способствования ее ретротранспозиции в геном. В некоторых вариантах осуществления сигнал ядерной локализации расположен на 3'-конце, 5'-конце или во внутренней области РНК-матрицы. В некоторых вариантах осуществления сигнал ядерной локализации расположен с 3'-стороны от гетерологичной последовательности (например, расположен с 3'-стороны от гетерологичной последовательности непосредственно рядом с ней) или расположен с 5'-стороны от гетерологичной последовательности (например, расположен с 5'-стороны от гетерологичной последовательности непосредственно рядом с ней). В некоторых вариантах осуществления сигнал ядерной локализации находится за пределами 5'-UTR или за пределами 3'-UTR РНК-матрицы. В некоторых вариантах осуществления сигнал ядерной локализации находится между 5'-UTR и 3'-UTR, где сигнал ядерной локализации необязательно не транскрибируется вместе с трансгеном (например, сигнал ядерной локализации находится в антисмысловой ориентации или располагается ниже сигнала терминации транскрипции или сигнала полиаденилирования). В некоторых вариантах осуществления последовательность ядерной локализации находится внутри интрона. В некоторых вариантах осуществления множество одинаковых или разных сигналов ядерной локализации находятся в РНК, например, в РНК-матрице. В некоторых вариантах осуществления длина сигнала ядерной локализации составляет менее 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 или 1000 п. о. Можно использовать разнообразные последовательности ядерной локализации РНК. Например, в Lubelsky и Ulitsky, *Nature* 555 (107-111), 2018 описаны последовательности РНК, которые управляют локализацией РНК в ядре. В некоторых вариантах осуществления сигнал ядерной локализации представляет собой сигнал ядерной локализации РНК, полученный из SINE (SIRLOIN). В некоторых вариантах осуществления сигнал ядерной локализации связывается с белком, обогащенным в ядре. В некоторых вариантах осуществления сигнал ядерной локализации связывается с белком HNRNPK. В некоторых вариантах осуществления сигнал ядерной локализации богат пиримидиновыми остатками, например,

представляет собой C/T-богатую, C/U-богатую, C-богатую, T-богатую или U-богатую область. В некоторых вариантах осуществления сигнал ядерной локализации получен из длинной некодирующей РНК. В некоторых вариантах осуществления сигнал ядерной локализации получен из длинной некодирующей РНК MALAT1 или представляет собой M-область MALAT1 длиной 600 нуклеотидов (описанную в Miyagawa et al., *RNA* 18, (738-751), 2012). В некоторых вариантах осуществления сигнал ядерной локализации получен из длинной некодирующей РНК BORG или представляет собой мотив AGCCC (описанный в Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014)). В некоторых вариантах осуществления последовательность ядерной локализации описана в Shukla et al., *The EMBO Journal* e98452 (2018). В некоторых вариантах осуществления сигнал ядерной локализации получен из не содержащего LTR ретротранспозона, ретротранспозона, содержащего LTR, ретровируса или эндогенного ретровируса.

В некоторых вариантах осуществления полипептид, описанный в данном документе, содержит одну или несколько (например, 2, 3, 4, 5) последовательностей, нацеленных на ядро, например, последовательность ядерной локализации (NLS), например, как описано выше. В некоторых вариантах осуществления NLS представляет собой двухкомпонентную NLS. В некоторых вариантах осуществления NLS облегчает импорт белка, содержащего NLS, в ядро клетки. В некоторых вариантах осуществления NLS слита с N-концом полипептида Gene Writer, описанного в данном документе. В некоторых вариантах осуществления NLS слита с C-концом полипептида Gene Writer. В некоторых вариантах осуществления NLS слита с N-концом или C-концом домена Cas. В некоторых вариантах осуществления линкерная последовательность расположена между NLS и соседним доменом полипептида Gene Writer.

В некоторых вариантах осуществления NLS содержит аминокислотную последовательность MDSLLMNRRKFLYQFKNVRWAKGRRETYLC, PKKRKVEGADKRTADGSEFESPKKKRKV, RKSGKIAAIWKRPRKPKKKRKV KRTADGSEFESPKKKRKV, KKTELQTTNAENKTKKL, или KRGINDRNFWRGENGRKTR, KRPAATKKAGQAKKKK, или их функциональный фрагмент или вариант. Примеры последовательностей NLS также описаны в документе PCT/EP2000/011690, содержание которого включено в данный документ посредством ссылки для раскрытия иллюстративных последовательностей ядерной локализации. В некоторых вариантах осуществления NLS содержит аминокислотную последовательность, раскрытую в таблице 8. NLS из этой таблицы можно использовать с одной или несколькими копиями в полипептиде в одном или нескольких местоположениях полипептида, например, 1, 2, 3 или больше копий NLS в N-концевом домене, между пептидными доменами, в C-концевом домене или в комбинации местоположений, чтобы улучшить субклеточную локализацию в ядре. В одном полипептиде можно использовать несколько уникальных последовательностей. Последовательности могут быть в естественных условиях однокомпонентными или двухкомпонентными, например, иметь один или два отрезка из основных аминокислот, или могут использоваться в качестве химерных двухкомпонентных

последовательностей. Ссылки на последовательности соответствуют номерам доступа UniProt, за исключением тех случаев, когда они указаны как SeqNLS для последовательностей, полученных с использованием алгоритма предсказания субклеточной локализации (Lin et al BMC Bioinformat 13:157 (2012), включенный в данный документ посредством ссылки во всей своей полноте).

Таблица 8. Иллюстративные сигналы ядерной локализации для применения в системах Gene Writing

Последовательность	**Ссылки на последовательности**
AHFKISGEKRPSTDPGKKAKNPKKK KKKDP	Q76IQ7
AHRAKKMSKTHA	P21827
ASPEYVNLPINGNG	SeqNLS
CTKRPRW	O88622, Q86W56, Q9QYM2, O02776
DKAKRVSRNKSEKKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15
EEQLRRRKNSRLNNTG	G5EFF5
EVLKVIRTGKRKKKAWKRMVTKVC	SeqNLS
HHHHHHHHHHHHQPH	Q63934, G3V7L5, Q12837
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0
HKRTKK	Q2R2D5
IINGRKLKLKKSRRRSSQTSNNSFTSR RS	SeqNLS
KAEQERRK	Q8LH59
KEKRKRREELFIEQKKRK	SeqNLS
KKGKDEWFSRGKKP	P30999
KKGPSVQKRKKT	Q6ZN17
KKKTVINDLLHYKKEK	SeqNLS, P32354
KKNGGKGKNKPSAKIKK	SeqNLS
KKPKWDDFKKKKK	Q15397, Q8BKS9, Q562C7
KKRKKD	SeqNLS, Q91Z62, Q1A730, Q969P5, Q2KHT6, Q9CPU7
KKRRKRRRK	SeqNLS

KKRRRRARK	Q9UMS6, D4A702, Q91YE8
KKSKRGR	Q9UBS0
KKSRKRGS	B4FG96
KKSTALSRELGKIMRRR	SeqNLS, P32354
KKSYQDPEIIAHSRPRK	Q9U7C9
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2
KKVSIAGQSGKLWRWKR	Q6YUL8
KKYENVVIKRSPRKRGRPRK	SeqNLS
KNKKRK	SeqNLS
KPKKKR	SeqNLS
KRAMKDDSHGNSTSPKRRK	Q0E671
KRANSNLVAAYEKAKKK	P23508
KRASEDTTSGSPPKKSSAGPKR	Q9BZZ5, Q5R644
KRFKRRWMVRKMKTKK	SeqNLS
KRGLNSSFETSPKKVK	Q8IV63
KRGNSSIGPNDLSKRKQRKK	SeqNLS
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS
KRKGKLKNKGSKRKK	O15381
KRRRRRREKRKR	Q96GM8
KRSNDRTYSPEEEKQRRA	Q91ZF2
KRTVATNGDASGAHRAKKMSK	SeqNLS
KRVYNKGEDEQEHLPKGKKR	SeqNLS
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524
KVNFLDMSLDDIIIYKELE	Q9P127
KVQHRIAKKTTRRRR	Q9DXE6
LSPSLSPL	Q9Y261, P32182, P35583
MDSLLMNRRKFLYQFKNVRWAKGR RETYLC	Q9GZX7
MPQNEYIELHRKRYGYRLDYHEKK RKKESREAHERSKKAKKMIGLKAKL YHK	SeqNLS
MVQLRPRASR	SeqNLS
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40

PDTKRAKLDSSETTMVKKK	SeqNLS
PEKRTKI	SeqNLS
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5
PGKMDKGEHRQERRDRPY	Q01844, Q61545
PKKGDKYDKTD	Q45FA5
PKKKSRK	O35914, Q01954
PKKNKPE	Q22663
PKKRAKV	P04295, P89438
PKPKKLKVE	P55263, P55262, P55264, Q64640
PKRGRGR	Q9FYS5, Q43386
PKRRLVDDA	P0C797
PKRRRTY	SeqNLS
PLFKRR	A8X6H4, Q9TXJ0
PLRKAKR	Q86WB0, Q5R8V9
PPAKRKCIF	Q6AZ28, O75928, Q8C5D8
PPARRRRL	Q8NAG6
PPKKKRKV	Q3L6L5, P03070, P14999, P03071
PPNKRMKVKH	Q8BN78
PPRIYPQLPSAPT	P0C799
PQRSPFPKSSVKR	SeqNLS
PRPRKVPR	P0C799
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9
PRRVRLK	Q58DJ0, P56477, Q13568
PSRKRPR	Q62315, Q5F363, Q92833
PSSKKRKV	SeqNLS
PTKKRVK	P07664
QRPGPYDRP	SeqNLS
RGKGGKGLGKGGAKRHRK	SeqNLS
RKAGKGGGGHKTTKKRSAKDEKVP	B4FG96
RKIKLKRAK	A1L3G9
RKIKRKRAK	B9X187
RKKEAPGPREELRSRGR	O35126, P54258, Q5IS70, P54259
RKKRKGK	SeqNLS, Q29243, Q62165, Q28685, O18738, Q9TSZ6, Q14118

RKKRRQRRR	P04326, P69697, P69698, P05907, P20879, P04613, P19553, P0C1J9, P20893, P12506, P04612, Q73370, P0C1K0, P05906, P35965, P04609, P04610, P04614, P04608, P05905
RKKSIPLSIKNLKRKHKRKKNKITR	Q9C0C9
RKLVKPKNTKMKTKLRTNPY	Q14190
RKRLILSDKGQLDWKK	SeqNLS, Q91Z62, Q1A730, Q2KHT6, Q9CPU7
RKRLKSK	Q13309
RKRRVRDNM	Q8QPH4, Q809M7, A8C8X1, Q2VNC5, Q38SQ0, O89749, Q6DNQ9, Q809L9, Q0A429, Q20NV3, P16509, P16505, Q6DNQ5, P16506, Q6XT06, P26118, Q2ICQ2, Q2RCG8, Q0A2D0, Q0A2H9, Q9IQ46, Q809M3, Q6J847, Q6J856, B4URE4, A4GCM7, Q0A440, P26120, P16511,
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6
RKRTPRVDGQTGENDMNKRRRK	O94851
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270
RLRFRKPKSK	P69469
RQQRKR	Q14980
RRDLNSSFETSPKKVK	Q8K3G5
RRDRAKLR	Q9SLB8
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922
RRGRKRKAEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1
RRKRSKSEDMDSVESKRRR	Q7TT18
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8
RRPKGKTLQKRKPK	Q6ZN17
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0
RRRGKNKVAAQNCRK	SeqNLS
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9
RRRQKQKGGASRRR	SeqNLS
RRRREGPRARRRR	P08313, P10231

RRTIRLKLVYDKCDRSCKIQKKNRN KCQYCRFHKCLSVGMSHNAIRFGR MPRSEKAKLKAE	SeqNLS
RRVPQRKEVSRCRKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5
RVGGRRQAVECIEDLLNEPGQPLDL SCKRPRP	P03255
RVVKLRIAP	P52639, Q8JMN0
RVVRRR	P70278
SKRKTKISRKTR	Q5RAY1, O00443
SYVKTVPNRTRTYIKL	P21935
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9
TLSPASSPSSVSCPVIPASTDESPGSA LNI	SeqNLS
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9
SPKKKRKVE	
KRTAD GSEFE SPKKKRKVE	
PAAKRVKLD	
PKKKRKV	
MDSLLMNRRKFLYQFKNVRWAKGR RETYLC	
SPKKKRKVEAS	
MAPKKKRKVGIHRGVP	

В некоторых вариантах осуществления NLS представляет собой двухкомпонентную NLS. Двухкомпонентная NLS, как правило, содержит два кластера основных аминокислот, разделенных спейсерной последовательностью (длина которой может составлять, например, приблизительно 10 аминокислот). В однокомпонентной NLS, как правило, отсутствует спейсер. Примером двухкомпонентной NLS является NLS нуклеоплазмина, содержащая последовательность KR[PAATKKAGQA]KKKK, где спейсер заключен в скобки. Другая иллюстративная двухкомпонентная NLS содержит последовательность PKKKRKVEGADKRTADGSEFESPKKKRKV. Иллюстративные NLS описаны в международной заявке WO 2020051561, которая включена в данный документ посредством ссылки во всей своей полноте, в том числе для их раскрытия в отношении последовательностей ядерной локализации.

В определенных вариантах осуществления полипептид системы редактирования генов Gene Writer™ дополнительно содержит последовательность внутриклеточной локализации, например, последовательность ядерной локализации и/или

последовательность ядрышковой локализации. Последовательность ядерной локализации и/или последовательность ядрышковой локализации может представлять собой аминокислотные последовательности, которые способствуют импорту белка в ядро и/или ядрышко, где он может способствовать интеграции гетерологичной последовательности в геном. В определенных вариантах осуществления полипептид системы редактирования генов Gene Writer (например, ретротранспозаза, например, полипептид в соответствии с любой из таблиц 1, 2, 3A или 3B в данном документе) дополнительно содержит последовательность ядрышковой локализации. В определенных вариантах осуществления полипептид ретротранспозаза кодируется первой РНК, а РНК-матрица представляет собой вторую, отдельную РНК, и сигнал ядрышковой локализации кодируется РНК, кодирующей полипептид ретротранспозазу, а не РНК-матрицей. В некоторых вариантах осуществления сигнал ядрышковой локализации расположен на N-конце, C-конце или во внутренней области полипептида. В некоторых вариантах осуществления используется множество одинаковых или разных сигналов ядрышковой локализации. В некоторых вариантах осуществления сигнал ядерной локализации имеет длину, составляющую менее 5, 10, 25, 50, 75 или 100 аминокислот. Можно использовать разнообразные сигналы ядрышковой локализации полипептида. Например, в Yang et al., *Journal of Biomedical Science* 22, 33 (2015), описан сигнал ядерной локализации, который также функционирует в качестве сигнала ядрышковой локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может также представлять собой сигнал ядерной локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может перекрываться с сигналом ядерной локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может содержать отрезок из основных остатков. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть богат аргининовыми и лизиновыми остатками. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, обогащенного в ядрышке. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, обогащенного в локусах рибосомной РНК. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, который связывает рРНК. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из MSP58. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой однокомпонентный мотив. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой двухкомпонентный мотив. В некоторых вариантах осуществления сигнал ядрышковой локализации может состоять из нескольких однокомпонентных или двухкомпонентных мотивов. В некоторых вариантах осуществления сигнал ядрышковой локализации может состоять из смеси однокомпонентных и двухкомпонентных мотивов. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой два двухкомпонентных мотива. В некоторых вариантах осуществления мотив ядрышковой локализации может представлять собой KRASSQALGTIPKRRSSSRFIKRKK (SEQ ID NO:

1530). В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из киназы, индуцирующей ядерный фактор κB. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой мотив RKKRKKK (SEQ ID NO: 1531) (описанный в Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004).

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК, кодирующая полипептид Gene Writer, или ДНК, кодирующая РНК), содержит сайт связывания с микроРНК. В некоторых вариантах осуществления сайт связывания с микроРНК используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, сайт связывания с микроРНК может быть выбран на основании того, что он распознается miRNA, которая присутствует в типе клеток, не являющихся мишенями, но которая отсутствует (или присутствует на сниженном уровне по сравнению с клетками, не являющимися мишенями) в типе клеток, являющихся мишенями. Таким образом, если РНК, кодирующая полипептид Gene Writer, присутствует в клетке, не являющейся мишенью, то она будет связываться с miRNA, а если РНК, кодирующая полипептид Gene Writer, присутствует в клетке-мишени, то она не будет связываться с miRNA (или будет связываться, но на сниженных уровнях по сравнению с клеткой, не являющейся мишенью). Не желая ограничиваться какой-либо теорией, полагают, что связывание miRNA с РНК, кодирующей полипептид GeneWriter, может приводить к снижению продуцирования полипептида GeneWriter, например, посредством разрушения мРНК, кодирующей полипептид, или посредством препятствования трансляции. Соответственно, вставка последовательности гетерологичного объекта в геном клеток-мишеней будет происходить более эффективно, чем в геном клеток, не являющихся мишенями. Система, имеющая сайт связывания с микроРНК в РНК, кодирующей полипептид Gene Writer (или кодируемый в ДНК, кодирующей РНК), может также применяться в комбинации с РНК-матрицей, которая регулируется вторым сайтом связывания с микроРНК, например, как описано в данном документе в разделе под названием "Компонент системы редактирования генов Gene Writer™, представляющий собой РНК-матрицу".

По номерам доступа, представленным в таблицах X, Z1, Z2, 3A или 3B, могут быть определены последовательности нуклеиновых кислот и соответствующих полипептидов для каждого ретротранспозона и его доменов, например, с помощью стандартных инструментов для анализа последовательностей, таких как средство поиска основного локального выравнивания (BLAST) или CD-Search для анализа консервативных доменов. Другие инструменты для анализа последовательностей известны и могут быть найдены, например, на https://molbiol-tools.ca, например, на https://molbiol-tools.ca/Motifs.htm. SEQ ID NO: 113-1015 соответствуют первым 903 строкам таблицы 2.

В таблицах X, 3A и 3B в данном документе представлены последовательности иллюстративных транспозонов, в том числе аминокислотная последовательность ретротранспозазы, а также последовательности 5'- и 3'-нетранслируемых областей,

обеспечивающие возможность связывания ретротранспозазы с РНК-матрицей. В некоторых вариантах осуществления 5'-UTR из любой из таблиц X, Z1, Z2, 3A или 3B обеспечивает возможность связывания ретротранспозазы с РНК-матрицей. В некоторых вариантах осуществления 3'-UTR из любой из таблиц X, 3A или 3B обеспечивает возможность связывания ретротранспозазы с РНК-матрицей. Таким образом, в некоторых вариантах осуществления полипептид для применения в любой из систем, описанных в данном документе, может представлять собой полипептид из любой из таблиц X, 3A и 3B в данном документе или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним. В некоторых вариантах осуществления система дополнительно содержит одну или обе из 5'- или 3'-нетранслируемой области из любой из таблиц X, 3A и 3B в данном документе (или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними), например, из того же ретротранспозона, что и полипептид, упомянутый в предыдущем предложении, как указано в той же строке той же таблицы. В некоторых вариантах осуществления система содержит одну или обе из 5'- или 3'-нетранслируемой области из любой из таблиц X, 3A и 3B в данном документе, например, сегмент полной последовательности ретротранспозона, который кодирует РНК, способную связываться с ретротранспозазой, и/или подпоследовательность, указанную в столбце под названием "Предсказанная 5'-UTR" или "Предсказанная 3'-UTR".

В некоторых вариантах осуществления полипептид для применения в любой из систем, описанных в данном документе, может представлять собой молекулярную реконструкцию или предковую реконструкцию на основе выровненной полипептидной последовательности нескольких ретротранспозонов. В определенных вариантах осуществления 5'- или 3'-нетранслируемая область для применения в любой из систем, описанных в данном документе, может представлять собой молекулярную реконструкцию на основе выровненной 5'- или 3'-нетранслируемой области нескольких ретротранспозонов. По номерам доступа, представленным в данном документе, могут быть выровнены последовательности полипептидов или нуклеиновых кислот, например, с помощью стандартных инструментов для анализа последовательностей, таких как средство поиска основного локального выравнивания (BLAST) или CD-Search для анализа консервативных доменов. Молекулярные реконструкции можно создавать на основе консенсусных последовательностей, например, используя подходы, описанные в *Ivics et al., Cell 1997, 501-510; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99.* В некоторых вариантах осуществления ретротранспозон, из которого получены 5'- или 3'-нетранслируемая область или полипептид, представляет собой молодой или недавно активированный мобильный элемент, как оценивается с помощью филогенетических методов, таких как методы, описанные в Boissinot et al., Molecular Biology and Evolution 2000, 915-928.

3. **Таблица Z1. Иллюстративные домены обратной транскриптазы из разных типов источников**

Источники включают интрон группы II, не содержащий LTR ретротранспозон, ретровирус, содержащий LTR ретротранспозон, ретроэлемент, генерирующий разнообразие, ретрон, теломеразу, ретроплазмиду и эволюционировавшую ДНК-полимеразу. Также включены ассоциированные сигнатуры RT из баз данных InterPro, Pfam и CD. Хотя эволюционировавшая полимераза RTX может выполнять РНК-зависимую полимеризацию ДНК, сигнатуры RT не были идентифицированы с помощью InterProScan, поэтому вместо них включены сигнатуры полимераз.

Белок	**Тип**	**Номер доступа**	**UniProt**	**Последовательность**	**Сигнатуры RT**
Marathon RT	Интрон группы II	CBK92290.1	D4JMT6	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC	IPR000477, PF00078, cd01651
TGIRT, trt	Интрон группы II	AAT72329.1	Q6DKY2	MALLERILADRNLITALKRVEANQGAPGIGDVSTDQLRDIYRAHWSTIRAQLLAGTYRPAPVRRVGIPKGPGGTRQ	IPR000477, PF00078, cd01651

				LGITPVVDRLIQQIALQELT PIFDPDFSPSSFGFRPGRNA HDAVRQAQGYIQEYGRYV VDMDLKEFFDRVNHDLIM SRVARKVDKKRVLKLIRY ALQAGVMIEGVKVQTEEG TQPGGPLSPLLANILLDDLD KELEKRGLKFCYRADDCNI YVSKLRAGQRVKQSIQRFL EKTLKLKVNEEKSVADRP WKRAFGLFSFTPERKARIR LAPRSIQRLKQRIRQLTNPN WSISMPREIHRVNQYVGM WIGYFRLVTEPSVLQTIEG WIRRRLRLCWQLQWKRVR TRIRELRALGLKETAVMEI ANRTKGAWRTTKPQTLHQ ALGKYTWTAQGLKTSLQR YFELRQG	
LtrA	Интрон группы II	AAB06503.1	P0A3U0	MKPTMAILERISKNSQENID EVFTRLYRYLLRPDIYYVA YQNLYSNKGASTKGILDDT ADGFSEEKIKKIIQSLKDGT YYPQPVRRMYIAKKNSKK MRPLGIPTFTDKLIQEAVRII LESIYEPVFEDVSHGFRPQR SCHTALKTIKREFGGARWF VEGDIKGCFDNIDHVTLIGL INLKIKDMKMSQLIYKFLK AGYLENWQYHKTYSGTPQ GGILSPLLANIYLHELDKFV LQLKMKFDRESPERITPEY RELHNEIKRISHRLKKLEGE EKAKVLLEYQEKRKRLPTL PCTSQTNKVLKYVRYADD	IPR000477, PF00078, cd01651

				FIISVKGSKEDCQWIKEQLK LFIHNKLKMELSEEKTLITH SSQPARFLGYDIRVRRSGTI KRSGKVKKRTLNGSVELLI PLQDKIRQFIFDKKIAIQKK DSSWFPVHRKYLIRSTDLEI ITIYNSELRGICNYYGLASN FNQLNYFAYLMEYSCLKTI ASKHKGTLSKTISMFKDGS GSWGIPYEIKQGKQRRYFA NFSECKSPYQFTDEISQAPV LYGYARNTLENRLKAKCC ELCGTSDENTSYEIHHVNK VKNLKGKEKWEMAMIAK QRKTLVVCFHCHRHVIHK HK	
R2Bm	Не содержащий LTR ретротранспозон	AAB59214.1	V9H052	MMASTALSLMGRCNPDGC TRGKHVTAAPMDGPRGPS SLAGTFGWGLAIPAGEPCG RVCSPATVGFFPVAKKSNK ENRPEASGLPLESERTGDN PTVRGSAGADPVGQDAPG WTCQFCERTFSTNRGLGVH KRRAHPVETNTDAAPMMV KRRWHGEEIDLLARTEARL LAERGQCSGGDLFGALPGF GRTLEAIKGQRRREPYRAL VQAHLARFGSQPGPSSGGC SAEPDFRRASGAEEAGEER CAEDAAAYDPSAVGQMSP DAARVLSELLEGAGRRRA CRAMRPKTAGRRNDLHDD RTASAHKTSRQKRRAEYA RVQELYKKCRSRAAAEVID GACGGVGHSLEEMETYWR	IPR000477, PF00078, cd01650

				PILERVSDAPGPTPEALHAL GRAEWHGGNRDYTQLWK PISVEEIKASRFDWRTSPGP DGIRSGQWRAVPVHLKAE MFNAWMARGEIPEILRQCR TVFVPKVERPGGPGEYRPIS IASIPLRHFHSILARRLLACC PPDARQRGFICADGTLENS AVLDAVLGDSRKKLRECH VAVLDFAKAFDTVSHEAL VELLRLRGMPEQFCGYIAH LYDTASTTLAVNNEMSSPV KVGRGVRQGDPLSPILFNV VMDLILASLPERVGYRLEM ELVSALAYADDLVLLAGS KVGMQESISAVDCVGRQM GLRLNCRKSAVLSMIPDGH RKKHHYLTERTFNIGGKPL RQVSCVERWRYLGVDFEA SGCVTLEHSISSALNNISRA PLKPQQRLEILRAHLIPRFQ HGFVLGNISDDRLRMLDV QIRKAVGQWLRLPADVPK AYYHAAVQDGGLAIPSVR ATIPDLIVRRFGGLDSSPWS VARAAAKSDKIRKKLRWA WKQLRRFSRVDSTTQRPSV RLFWREHLHASVDGRELR ESTRTPTSTKWIRERCAQIT GRDFVQFVHTHINALPSRIR GSRGRRGGGESSLTCRAGC KVRETTAHILQQCHRTHGG RILRHNKIVSFVAKAMEEN KWTVELEPRLRTSVGLRKP DIIASRDGVGVIVDVQVVS	

				GQRSLDELHREKRNKYGN HGELVELVAGRLGLPKAEC VRATSCTISWRGVWSLTSY KELRSIIGLREPTLQIVPILA LRGSHMNWTRFNQMTSV MGGGVG	
LINE-1	Не содержащий LTR ретротранспозон	AAC51271.1	O00370	MTGSNSHITILTLNVNGLNS PIKRHRLASWIKSQDPSVC CIQETHLTCRDTHRLKIKG WRKIYQANGKQKKAGVAI LVSDKTDFKPTKIKRDKEG HYIMVKGSIQQEELTILNIY APNTGAPRFIKQVLSDLQR DLDSHTLIMGDFNTPLSILD RSTRQKVNKDTQELNSAL HQTDLIDIYRTLHPKSTEYT FFSAPHHTYSKIDHIVGSKA LLSKCKRTEIITNYLSDHSA IKLELRIKNLTQSRSTTWKL NNLLLNDYWVHNEMKAEI KMFFETNENKDTTYQNLW DAFKAVCRGKFIALNAYK RKQERSKIDTLTSQLKELE KQEQTHSKASRRQEITKIR AELKEIETQKTLQKINESRS WFFERINKIDRPLARLIKKK REKNQIDTIKNDKGDITTDP TEIQTTIREYYKHLYANKL ENLEEMDTFLDTYTLPRLN QEEVESLNRPITGSEIVAIIN SLPTKKSPGPDGFTAEFYQ RYKEELVPFLLKLFQSIEKE GILPNSFYEASIILIPKPGRD TTKKENFRPISLMNIDAKIL NKILANRIQQHIKKLIHHDQ	IPR000477, PF00078, cd01650

				VGFIPGMQGWFNIRKSINVI QHINRAKDKNHVIISIDAEK AFDKIQQPFMLKTLNKLGI DGMYLKIIRAIYDKPTANII LNGQKLEAFPLKTGTRQGC PLSPLLFNIVLEVLARAIRQ EKEIKGIQLGKEEVKLSLFA DDMIVYLENPIVSAQNLLK LISNFSKVSGYKINVQKSQ AFLYNNNRQTESQIMGELP FTIASKRIKYLGIQLTRDVK DLFKENYKPLLKEIKEDTN KWKNIPCSWVGRINIVKM AILPKVIYRFNAIPIKLPMTF FTELEKTTLKFIWNQKRARI AKSILSQKNKAGGITLPDFK LYYKATVTKTAWYWYQN RDIDQWNRTEPSEIMPHIY NYLIFDKPEKNKQWGKDS LLNKWCWENWLAICRKLK LDPFLTPYTKINSRWIKDLN VKPKTIKTLEENLGITIQDIG VGKDFMSKTPKAMATKDK IDKWDLIKLKSFCTAKETTI RVNRQPTTWEKIFATYSSD KGLISRIYNELKQIYKKKTN NPIKKWAKDMNRHFSKEDI YAAKKHMKKCSSSLAIRE MQIKTTMRYHLTPVRMAII KKSGNNRCWRGCGEIGTL VHCWWDCKLVQPLWKSV WRFLRDLELEIPFDPAIPLL GIYPKDYKSCCYKDTCTR MFIAALFTIAKTWNQPNCP TMIDWIKKMWHIYTMEYY	

				AAIKNDEFISFVGTWMKLE TIILSKLSQEQKTKHRIFSLI GGN	
Penelope	Не содержащий LTR ретротранспозон	AAL14979.1	Q95VB5	MERSPEPSININGRHAVCTA TNMSYAKIKTKYKDSKRTI NKFQLTLVKLTKLKSSLKF LLKCRKSNLIPNFIKNLTQH LTILTTDNKTHPDITRTLTR HTHFYHTKILNLLIKHKHN LLQEQTKHMQKAKTNIEQ LMTTDDAKAFFESERNIEN KITTTLKKRQETKHDKLRD QRNLALADNNTQREWFVN KTKIEFPPNVVALLAKGPK FALPISKRDFPLLKYIADGE ELVQTIKEKETQESARTKFS LLVKEHKTKNNQNSRDRAI LDTVEQTRKLLKENINIKIL SSDKGNKTVAMDEDEYKN KMTNILDDLCAYRTLRLDP TSRLQTKNNTFVAQLFKM GLISKDERNKMTTTTAVPP RIYGLPKIHKEGTPLRPICSS IGSPSYGLCKYIIQILKNLT MDSRYNIKNAVDFKDRVN NSQIREEETLVSFDVVSLFP SIPIELALDTIRQKWTKLEE HTNIPKQLFMDIVRFCIEEN RYFKYEDKIYTQLKGMPM GSPASPVIADILMEELLDKI TDKLKIKPRLLTKYVDDLF AITNKIDVENILKELNSFHK QIKFTMELEKDGKLPFLDSI VSRMDNTLKIKWYRKPIAS GRILNFNSNHPKSMIINTAL	IPR000477, PF00078, cd00304

				GCMNRMMKISDTIYHKEIE HEIKELLTKNDFPPNIIKTLL KRRQIERKKPTEPAKIYKSL IYVPRLSERLTNSDCYNKQ DIKVAHKPTNTLQKFFNKI KSKIPMIEKSNVVYQIPCGG DNNNKCNSVYIGTTKSKLK TRISQHKSDFKLRHQNNIQ KTALMTHCIRSNHTPNFDE TTILQQEQHYNKRHTLEML HIINTPTYKRLNYKTDTEN CAHLYRHLLNSQTTSVTIST SKSADV	
RT M-MLV	Ретровирус	ADS42990.1	P03355[660-1330]	TLNIEDEHRLHETSKEPDVS LGSTWLSDFPQAWAETGG MGLAVRQAPLIIPLKATSTP VSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLR EVNKRVEDIHPTVPNPYNL LSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQG FKNSPTLFDEALHRDLADF RIQHPDLILLQYVDDLLLA ATSELDCQQGTRALLQTLG NLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEAR KETVMGQPTPKTPRQLREF LGTAGFCRLWIPGFAEMAA PLYPLTKTGTLFNWGPDQQ KAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSK KLDPVAAGWPPCLRMVAA	IPR000477, PF00078, cd03715

				IAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWL SNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPD LTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETE VIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYT DSRYAFATAHIHGEIYRRR GLLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKG HSAEARGNRMADQAARK AAITETPDTSTLL	
RT RSV	Ретровир ус	AAC825 61.1	P03354[709- 1567]	TVALHLAIPLKWKPDHTPV WIDQWPLPEGKLVALTQL VEKELQLGHIEPSLSCWNT PVFVIRKASGSYRLLHDLR AVNAKLVPFGAVQQGAPV LSALPRGWPLMVLDLKDC FFSIPLAEQDREAFAFTLPS VNNQAPARRFQWKVLPQG MTCSPTICQLVVGQVLEPL RLKHPSLCMLHYMDDLLL AASSHDGLEAAGEEVISTL ERAGFTISPDKVQREPGVQ YLGYKLGSTYVAPVGLVA EPRIATLWDVQKLVGSLQ WLRPALGIPPRLMGPFYEQ LRGSDPNEAREWNLDMKM AWREIVRLSTTAALERWDP ALPLEGAVARCEQGAIGVL GQGLSTHPRPCLWLFSTQP TKAFTAWLEVLTLLITKLR ASAVRTFGKEVDILLLPAC	IPR000477, PF00078, cd01645

				FREDLPLPEGILLALKGFAG KIRSSDTPSIFDIARPLHVSL KVRVTDHPVPGPTVFTDAS SSTHKGVVVWREGPRWEI KEIADLGASVQQLEARAVA MALLLWPTTPTNVVTDSAF VAKMLLKMGQEGVPSTAA AFILEDALSQRSAMAAVLH VRSHSEVPGFFTEGNDVAD SQATFQAYPLREAKDLHTA LHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAG VNPRGLGPLQIWQTDFTLE PRMAPRSWLAVTVDTASS AIVVTQHGRVTSVAVQHH WATAIAVLGRPKAIKTDNG SCFTSKSTREWLARWGIAH TTGIPGNSQGQAMVERAN RLLKDRIRVLAEGDGFMKR IPTSKQGELLAKAMYALNH FERGENTKTPIQKHWRPTV LTEGPPVKIRIETGEWEKG WNVLVWGRGYAAVKNRD TDKVIWVPSRKVKPDITQK DEVTKKDEASPLFAG	
RT AMV	Ретровирус	HW606680.1	-	TVALHLAIPLKWKPNHTPV WIDQWPLPEGKLVALTQL VEKELQLGHIEPSLSCWNT PVFVIRKASGSYRLLHDLR AVNAKLVPFGAVQQGAPV LSALPRGWPLMVLDLKDC FFSIPLAEQDREAFAFTLPS VNNQAPARRFQWKVLPQG MTCSPTICQLIVGQILEPLR LKHPSLRMLHYMDDLLLA	IPR000477, PF00078, cd01645

				ASSHDGLEAAGEEVISTLE RAGFTISPDKVQREPGVQY LGYKLGSTYVAPVGLVAEP RIATLWDVQKLVGSLQWL RPALGIPPRLMGPFYEQLR GSDPNEAREWNLDMKMA WREIVQLSTTAALERWDPA LPLEGAVARCEQGAIGVLG QGLSTHPRPCLWLFSTQPT KAFTAWLEVLTLLITKLRA SAVRTFGKEVDILLLPACFR EDLPLPEGILLALRGFAGKI RSSDTPSIFDIARPLHVSLK VRVTDHPVPGPTVFTDASS STHKGVVVWREGPRWEIK EIADLGASVQQLEARAVA MALLLWPTTPTNVVTDSAF VAKMLLKMGQEGVPSTAA AFILEDALSQRSAMAAVLH VRSHSEVPGFFTEGNDVAD SQATFQAY	
RT HIV	Ретровирус	AAB50259.1	P04585[588-1147]	PISPIETVPVKLKPGMDGPK VKQWPLTEEKIKALVEICT EMEKEGKISKIGPENPYNTP VFAIKKKDSTKWRKLVDF RELNKRTQDFWEVQLGIPH PAGLKKKKSVTVLDVGDA YFSVPLDEDFRKYTAFTIPS INNETPGIRYQYNVLPQGW KGSPAIFQSSMTKILEPFRK QNPDIVIYQYMDDLYVGSD LEIGQHRTKIEELRQHLLR WGLTTPDKKHQKEPPFLW MGYELHPDKWTVQPIVLPE KDSWTVNDIQKLVGKLNW	IPR000477, PF00078, cd01645

				ASQIYPGIKVRQLCKLLRG TKALTEVIPLTEEAELELAE NREILKEPVHGVYYDPSKD LIAEIQKQGQGQWTYQIYQ EPFKNLKTGKYARMRGAH TNDVKQLTEAVQKITTESI VIWGKTPKFKLPIQKETWE TWWTEYWQATWIPEWEF VNTPPLVKLWYQLEKEPIV GAETFYVDGAANRETKLG KAGYVTNRGRQKVVTLTD TTNQKTELQAIYLALQDSG LEVNIVTDSQYALGIIQAQP DQSESELVNQIIEQLIKKEK VYLAWVPAHKGIGGNEQV DKLVSAGIRKVL	
Ty1	Содержащий LTR ретротранспозон	AAA66938.1	Q07163-1[1218-1755]	AVKAVKSIKPIRTTLRYDE AITYNKDIKEKEKYIEAYH KEVNQLLKMKTWDTDEY YDRKEIDPKRVINSMFIFNK KRDGTHKARFVARGDIQH PDTYDSGMQSNTVHHYAL MTSLSLALDNNYYITQLDIS SAYLYADIKEELYIRPPPHL GMNDKLIRLKKSLYGLKQS GANWYETIKSYLIQQCGME EVRGWSCVFKNSQVTICLF VDDMVLFSKNLNSNKRIIE KLKMQYDTKIINLGESDEEI QYDILGLEIKYQRGKYMKL GMENSLTEKIPKLNVPLNP KGRKLSAPGQPGLYIDQDE LEIDEDEYKEKVHEMQKLI GLASYVGYKFRFDLLYYIN TLAQHILFPSRQVLDMTYE	IPR013103, PF07727

				LIQFMWDTRDKQLIWHKN KPTEPDNKLVAISDASYGN QPYYKSQIGNIYLLNGKVI GGKSTKASLTCTSTTEAEIH AISESVPLLNNLSYLIQELN KKPIIKGLLTDSRSTISIIKST NEEKFRNRFFGTKAMRLR DEVSGNNLYVYYIETKKNI ADVMTKPLPIKTFKLLTNK WIH	
Brt	Ретроэлемент, генерирующий разнообразие	NP_958675.1	Q775D8	MGKRHRNLIDQITTWENLL DAYRKTSHGKRRTWGYLE FKEYDLANLLALQAELKA GNYERGPYREFLVYEPKPR LISALEFKDRLVQHALCNIV APIFEAGLLPYTYACRPDK GTHAGVCHVQAELRRTRA THFLKSDFSKFFPSIDRAAL YAMIDKKIHCAATRRLLRV VLPDEGVGIPIGSLTSQLFA NVYGGAVDRLLHDELKQR HWARYMDDIVVLGDDPEE LRAVFYRLRDFASERLGLK ISHWQVAPVSRGINFLGYRI WPTHKLLRKSSVKRAKRK VANFIKHGEDESLQRFLAS WSGHAQWADTHNLFTWM EEQYGIACH	IPR000477, PF00078, cd01646
RT86	Ретрон	AAA61471.1	P23070	MKSAEYLNTFRLRNLGLPV MNNLHDMSKATRISVETLR LLIYTADFRYRIYTVEKKGP EKRMRTIYQPSRELKALQG WVLRNILDKLSSSPFSIGFE KHQSILNNATPHIGANFILN IDLEDFFPSLTANKVFGVFH	IPR000477, PF00078, cd03487

				SLGYNRLISSVLTKICCYKN LLPQGAPSSPKLANLICSKL DYRIQGYAGSRGLIYTRYA DDLTLSAQSMKKVVKARD FLFSIIPSEGLVINSKKTCISG PRSQRKVTGLVISQEKVGI GREKYKEIRAKIHHIFCGKS SEIEHVRGWLSFILSVDSKS HRRLITYISKLEKKYGKNPL NKAKT	
TERT	Теломераза	AAG23289.1	O14746	MPRAPRCRAVRSLLRSHYR EVLPLATFVRRLGPQGWRL VQRGDPAAFRALVAQCLV CVPWDARPPPAAPSFRQVS CLKELVARVLQRLCERGA KNVLAFGFALLDGARGGPP EAFTTSVRSYLPNTVTDAL RGSGAWGLLLRRVGDDVL VHLLARCALFVLVAPSCAY QVCGPPLYQLGAATQARPP PHASGPRRRLGCERAWNH SVREAGVPLGLPAPGARRR GGSASRSLPLPKRPRRGAA PEPERTPVGQGSWAHPGRT RGPSDRGFCVVSPARPAEE ATSLEGALSGTRHSHPSVG RQHHAGPPSTSRPPRPWDT PCPPVYAETKHFLYSSGDK EQLRPSFLLSSLRPSLTGAR RLVETIFLGSRPWMPGTPR RLPRLPQRYWQMRPLFLEL LGNHAQCPYGVLLKTHCP LRAAVTPAAGVCAREKPQ GSVAAPEEEDTDPRRLVQL LRQHSSPWQVYGFVRACL	IPR000477, PF00078, cd01648

				RRLVPPGLWGSRHNERRFL RNTKKFISLGKHAKLSLQE LTWKMSVRDCAWLRRSPG VGCVPAAEHRLREEILAKF LHWLMSVYVVELLRSFFY VTETTFQKNRLFFYRKSVW SKLQSIGIRQHLKRVQLREL SEAEVRQHREARPALLTSR LRFIPKPDGLRPIVNMDYV VGARTFRREKRAERLTSRV KALFSVLNYERARRPGLLG ASVLGLDDIHRAWRTFVLR VRAQDPPPELYFVKVDVTG AYDTIPQDRLTEVIASIIKPQ NTYCVRRYAVVQKAAHG HVRKAFKSHVSTLTDLQPY MRQFVAHLQETSPLRDAV VIEQSSSLNEASSGLFDVFL RFMCHHAVRIRGKSYVQC QGIPQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLRLV DDFLLVTPHLTHAKTFLRT LVRGVPEYGCVVNLRKTV VNFPVEDEALGGTAFVQM PAHGLFPWCGLLLDTRTLE VQSDYSSYARTSIRASLTFN RGFKAGRNMRRKLFGVLR LKCHSLFLDLQVNSLQTVC TNIYKILLLQAYRFHACVL QLPFHQQVWKNPTFFLRVI SDTASLCYSILKAKNAGMS LGAKGAAGPLPSEAVQWL CHQAFLLKLTRHRVTYVPL LGSLRTAQTQLSRKLPGTT	

				LTALEAAANPALPSDFKTIL D	
RT Maurice ville	Ретропла змида	NC_001 570.1	Q36578	MPNHRLPNCVSYLGENHE LSWLHGMFGLLKRSNPQT GGILGWLNTGPNGFVKYM MNLMGHARDKGDAKEYW RLGRSLMKNEAFQVQAFN HVCKHWYLDYKPHKIAKL LKEVREMVEIQPVCIDYKR VYIPKANGKQRPLGVPTVP WRVYLHMWNVLLVWYRI PEQDNQHAYFPKRGVFTA WRALWPKLDSQNIYEFDL KNFFPSVDLAYLKDKLMES GIPQDISEYLTVLNRSLVVL TSEDKIPEPHRDVIFNSDGT PNPNLPKDVQGRILKDPDF VEILRRRGFTDIATNGVPQ GASTSCGLATYNVKELFKR YDELIMYADDGILCRQDPS TPDFSVEEAGVVQEPAKSG WIKQNGEFKKSVKFLGLEF IPANIPPLGEGEVKDYPRLR GATRNGSKMELSTELQFLC YLSYKLRIKVLRDLYIQVL GYLPSVPLLRYRSLAEAINE LSPKRITIGQFITSSFEEFTA WSPLKRMGFFFSSPAGPTIL SSIFNNSTNLQEPSDSRLLY RKGSWVNIRFAAYLYSKLS EEKHGLVPKFLEKLREINF ALDKVDVTEIDSKLSRLMK FSVSAAYDEVGTLALKSLF KFRNSERESIKASFKQLREN GKIAEFSEARRLWFEILKLI	cd00304

				RLDLFNASSLACDDLLSHL QDRRSIKKWGSSDVLYLKS QRLMRTNKKQLQLDFEKK KNSLKKKLIKRRAKELRDT FKGKENKEA	
RTX	Сконструированная полимераза	QFN49000.1	-	MILDTDYITEDGKPVIRIFK KENGEFKIEYDRTFEPYLY ALLKDDSAIEEVKKITAER HGTVVTVKRVEKVQKKFL GRPVEVWKLYFTHPQDVP AIMDKIREHPAVIDIYEYDI PFAIRYLIDKGLVPMEGDE ELKLLAFDIETLYHEGEEFA EGPILMISYADEEGARVIT WKNVDLPYVDVVSTEREM IKRFLRVVKEKDPDVLITY NGDNFDFAYLKKRCEKLGI NFALGRDGSEPKIQRMGDR FAVEVKGRIHFDLYPVIRRT INLPTYTLEAVYEAVFGQP KEKVYAEEITTAWETGENL ERVARYSMEDAKVTYELG KEFLPMEAQLSRLIGQSLW DVSRSSTGNLVEWFLLRKA YERNELAPNKPDEKELARR HQSHEGGYIKEPERGLWEN IVYLDFRSLYPSIIITHNVSP DTLNREGCKEYDVAPQVG HRFCKDFPGFIPSLLGDLLE ERQKIKKRMKATIDPIERKL LDYRQRAIKILANSLYGYY GYARARWYCKECAESVIA WGREYLTMTIKEIEEKYGF KVIYSDTDGFFATIPGADAE TVKKKAMEFLKYINAKLP	IPR006134, PF00136, cd05536

				GALELEYEGFYKRGLFVTK KKYAVIDEEGKITTRGLEIV RRDWSEIAKETQARVLEAL LKDGDVEKAVRIVKEVTE KLSKYEVPPEKLVIHKQITR DLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKG SGRIVDRAIPFDEFDPTKHK YDAEYYIEKQVLPAVERIL RAFGYRKEDLRYQKTRQV GLSARLKPKGTLEGSSHHH HHH	

4. **Таблица Z2. Описания сигнатур InterPro, присутствующих в обратных транскриптазах из таблицы Z1, по состоянию на 4 марта 2020 г.**

Сигнатура	**База данных**	**Краткое название**	**Описание**
cd00304	CDD	*RT_подобные*	RT_подобные: семейство, подобное обратной транскриптазе (RT, РНК-зависимой ДНК-полимеразе). Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные

			интроны и подвижные элементы, в которых отсутствуют LTR. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]
cd01645	CDD	*RT_Rtv*	RT_Rtv: обратные транскриптазы (RT) ретровирусов (Rtv). RT катализируют превращение однонитевой РНК в двухнитевую вирусную ДНК для интеграции в хромосомы хозяина. Белки этого подсемейства содержат длинные концевые повторы (LTR) и представляют собой многофункциональные ферменты с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). Вирусный РНК-геном входит в цитоплазму в качестве части нуклеопротеинового комплекса, и в цитоплазме организуется процесс обратной транскрипции с образованием линейного ДНК-дуплекса посредством сложного ряда стадий. Этот ДНК-дуплекс является коллинеарным к своей РНК-матрице, но содержит концевые дупликации, известные как LTR, которые отсутствуют в вирусной РНК. Было высказано предположение, что для создания LTR необходимы два специализированных переключателя матрицы, известных в качестве реакций переноса нити или "прыжков". [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]

cd01646	CDD	*RT_бактериальный_ретрон_I*	RT_бактериальный_ретрон_I: обратные транскриптазы (RT) в бактериальных ретротранспозонах или ретронах. Полимеразная реакция этого фермента приводит к образованию уникального комплекса РНК-ДНК, называемого msDNA (многокопийная однонитевая (ss) ДНК), в котором небольшая ssDNA ответвляется от небольшой молекулы ssRNA посредством 2'-5'-фосфодиэфирной связи. RT бактериального ретрона продуцирует кДНК, соответствующую лишь небольшой части генома ретрона. [PMID: 1698615, PMID: 16093702, PMID: 8828137]
cd01648	CDD	*TERT*	TERT: теломеразная обратная транскриптаза (TERT). Теломераза представляет собой рибонуклеопротеин (RNP), который синтезирует повторы теломерной ДНК. Субъединица теломеразной РНК обеспечивает матрицу для синтеза этих повторов. Каталитическая субъединица RNP известна в качестве теломеразной обратной транскриптазы (TERT). Домен обратной транскриптазы (RT) расположен в С-концевой области полипептида TERT. Единичные аминокислотные замены в этой области приводят к укорочению теломер и старению. Теломераза представляет собой фермент, который в определенных клетках поддерживает физические концы хромосом (теломеры) во время репликации. В соматических клетках репликация отстающей нити требует постоянного присутствия РНК-праймера на примерно 200 нуклеотидов выше по ходу транскрипции,

			который является комплементарным матричной нити. Поскольку существует область ДНК, расположенная на расстоянии менее 200 пар оснований от конца хромосомы, где это является невозможным, то хромосома постоянно укорачивается. Однако избыток повторяющейся ДНК на концах хромосом защищает от разрушения кодирующую ген ДНК. Теломераза обычно не экспрессируется в соматических клетках. Было высказано предположение, что экзогенная TERT может увеличить продолжительность жизни клетки или даже привести к ее иммортализации. Однако недавние исследования показали, что ряд онкогенов может индуцировать теломеразную активность. В противоположность этому, онкоген c-myc может быть активирован в клетках человека, иммортализованных с помощью TERT. Сравнение последовательностей относит теломеразные белки к семейству RT, но выявляет признаки, отличающие их от родственных белков ретровирусов и ретротранспозонов. [PMID: 9110970, PMID: 9288757, PMID: 9389643, PMID: 9671703, PMID: 9671704, PMID: 10333526, PMID: 11250070, PMID: 15363846, PMID: 16416120, PMID: 16649103, PMID: 16793225, PMID: 10860859, PMID: 9252327, PMID: 11602347, PMID: 1698615, PMID: 8828137, PMID: 10866187]
cd01650	CDD	*RT_nLTR_подобные*	RT_nLTR: Ретротранспозон, не содержащий LTR (длинных концевых повторов), и ретровирусная обратная транскриптаза (RT),

			не содержащая LTR. Это подсемейство содержит как ретротранспозоны, не содержащие LTR, так и ретровирусные RT, не содержащие LTR. RT катализируют превращение однонитевой РНК в двухнитевую ДНК для интеграции в хромосомы хозяина. RT представляет собой многофункциональный фермент с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). [PMID: 1698615, PMID: 10605110, PMID: 10628860, PMID: 11734649, PMID: 12117499, PMID: 12777502, PMID: 14871946, PMID: 15939396, PMID: 16271150, PMID: 16356661, PMID: 2463954, PMID: 3040362, PMID: 3656436, PMID: 7512193, PMID: 7534829, PMID: 7659515, PMID: 8524653, PMID: 9190061, PMID: 9218812, PMID: 9332379, PMID: 9364772, PMID: 8828137]
cd01651	CDD	*RT_интрон_G2*	RT_интрон_G2: обратные транскриптазы (RT), происходящие из интрона группы II. RT транскрибирует ДНК, используя РНК в качестве матрицы. Белки данного подсемейства обнаружены в бактериальных и митохондриальных интронах группы II. Их наиболее вероятным предком был ретротранспозонный элемент как с gag-подобными, так и с pol-подобными генами. Данное подсемейство белков, по-видимому, захватило последовательности RT из подвижных элементов, в которых отсутствуют длинные концевые повторы (LTR). [PMID: 1698615, PMID: 8828137, PMID: 12403467, PMID: 11058141, PMID:

			11054545, PMID: 10760141, PMID: 10488235, PMID: 9680217, PMID: 9491607, PMID: 7994604, PMID: 7823908, PMID: 3129199, PMID: 2531370, PMID: 2476655]
cd03487	CDD	*RT_бактериальный_ретрон_II*	RT_бактериальный_ретрон_II: обратные транскриптазы (RT) в бактериальных ретротранспозонах или ретронах. Полимеразная реакция этого фермента приводит к образованию уникального комплекса РНК-ДНК, называемого msDNA (многокопийная однонитевая (ss) ДНК), в котором небольшая ssDNA ответвляется от небольшой молекулы ssRNA посредством 2'-5'-фосфодиэфирной связи. RT бактериального ретрона продуцирует кДНК, соответствующую лишь небольшой части генома ретрона. [PMID: 1698615, PMID: 8828137, PMID: 11292805, PMID: 9281493, PMID: 2465092, PMID: 1722556, PMID: 1701261, PMID: 1689062]
cd03715	CDD	*RT_ZFREV_подобные*	RT_ZFREV_подобные: подсемейство обратных транскриптаз (RT), обнаруженное в последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-

			вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ позволяет предположить, что ZFERV принадлежит к отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
cd05536	CDD	*POLBc_B3*	Каталитический домен подсемейства B3 ДНК-полимераз B-типа. Белки архей, участвующие в репликации ДНК, являются сходными с белками эукариот. Некоторые представители архей также имеют несколько ДНК-полимераз семейства B (B1, B2 и B3). До сих пор не была(были) установлена(установлены) какая-либо(какие-либо) конкретная(конкретные) функция(функции) для различных представителей ДНК-полимераз B-типа архей. Филогенетический анализ ДНК-полимераз семейства B эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства B архей и эукариот. Структурное сравнение термостабильной ДНК-полимеразы B-типа с ее мезостабильным гомологом позволяет предположить наличие нескольких приспособлений к высокой температуре,

			таких как более короткие петли, дисульфидные мостики и усиление электростатического взаимодействия в областях контакта субдоменов. [PMID: 10997874, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321]
cd05780	CDD	*ДНК_полимеразы B_Kod1_подобные_экзо*	3'-5'-экзонуклеазный домен ДНК-полимераз семейства B архей со сходством с Kod1 Pyrococcus kodakaraensis, включая полимеразы из Desulfurococcus (Pol D. Tok) и Thermococcus gorgonarius (Pol Tgo). Kod1, Pol D. Tok и Pol Tgo представляют собой термостабильные ферменты, обладающие как полимеразной, так и 3'-5'-экзонуклеазной активностью. Они представляют собой ДНК-полимеразы семейства B. Их амино-концы содержат DnaQ-подобный 3'-5'-экзонуклеазный домен DEDDy-типа, который содержит три мотива последовательности, называемые ExoI, ExoII и ExoIII, со специфическим паттерном YX(3)D в ExoIII. Данные мотивы образуют кластер вокруг активного центра и участвуют в связывании металлов и катализе. Экзонуклеазный домен полимераз семейства B содержит бета-шпилечную структуру, которая играет важную роль в переключении активного центра в случае неправильного включения нуклеотидов. Представители этого подсемейства проявляют сходство с эукариотическими ДНК-полимеразами, участвующими в репликации ДНК. Некоторые археи имеют несколько ДНК-полимераз семейства B. Филогенетический анализ ДНК-полимераз

			семейства B эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства B архей и эукариот. [PMID: 18355915, PMID: 16019029, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321, PMID: 9098062, PMID: 12459442, PMID: 16230118, PMID: 11988770, PMID: 11222749, PMID: 17098747, PMID: 8594362, PMID: 9729885]
PF00078	Pfam	*RVT_1*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. [PMID: 1698615]
PF00136	Pfam	*ДНК_полимераза_B*	Данная область ДНК-полимеразы семейства B, по-видимому, состоит из более чем одного структурного домена, возможно, предусматривающего формы активности элонгации, связывания ДНК и связывания dNTP. [PMID: 9757117, PMID: 8679562]
PF07727	Pfam	*RVT_2*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Данная запись Pfam включает обратные

			транскриптазы, не распознаваемые моделью Pfam:PF00078. [PMID: 1698615]
IPR000477	InterPro	*RT_домен*	Применение РНК-матрицы для получения ДНК с целью интеграции в геном хозяина и использования клетки-хозяина является стратегией, применяемой при репликации ретроидных элементов, таких как ретровирусы и бактериальные ретроны. Фермент, катализирующий полимеризацию, представляет собой РНК-направленную ДНК-полимеразу или обратную транскриптазу (RT) (2.7.7.49). Обратная транскриптаза встречается в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II [PMID: 12758069], бактериальные msDNA, гепаднавирусы и каулимовирусы. Ретровирусная обратная транскриптаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Обнаружение ретроэлементов у прокариот поднимает интригующие вопросы, касающиеся их роли у бактерий, происхождения и эволюции обратных транскриптаз, а также того, являются ли бактериальные обратные транскриптазы старшими, чем эукариотические обратные транскриптазы [PMID: 8828137]. Было определено несколько кристаллических структур домена

			обратной транскриптазы (RT) [PMID: 1377403].
IPR006134	InterPro	ДНК-направленная_ДНК_полимераза_B_с несколькими_доменами	ДНК представляет собой биологическую информацию, которая указывает клеткам, как существовать упорядоченно: таким образом, точная репликация является одним из наиболее важных событий в жизненном цикле клетки. Эту функцию выполняют ДНК-направленные ДНК-полимеразы (2.7.7.7) посредством добавления остатков нуклеотидтрифосфата (dNTP) к 5'-концу растущей цепи ДНК, используя в качестве матрицы комплементарную цепь ДНК. Небольшие молекулы РНК обычно применяются в качестве праймеров для элонгации цепи, хотя концевые белки также могут применяться для синтеза цепи ДНК de novo. Несмотря на то, что существует 2 разных способа праймирования, они опосредуются 2 очень сходными классами полимераз, A и B, со сходными способами элонгации цепи. Ряд ДНК-полимераз был сгруппирован под обозначением ДНК-полимераз семейства B. Шесть областей сходства (пронумерованных от I до VI) обнаружены во всех или в подмножестве полимераз семейства B. Наиболее консервативная область (I) включает консервативный тетрапептид с двумя остатками аспартата. Было высказано предположение, что он может участвовать в связывании иона магния. Все последовательности в семействе B содержат характерный мотив DTDS и имеют несколько функциональных доменов,

			включая 5'-3'-домен элонгации, 3'-5'-экзонуклеазный домен [PMID: 8679562], ДНК-связывающий домен и домены связывания как для dNTP, так и для пирофосфата [PMID: 9757117]. Данный домен ДНК-полимеразы семейства B, по-видимому, характеризуется более чем одной формой активности, возможно, включающей элонгацию, связывание ДНК и связывание dNTP [PMID: 9757117].
IPR013103	InterPro	*RVT_2*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эта запись включает обратные транскриптазы, не распознаваемые IPR000477 [PMID: 1698615].

В **таблице 3A и 3B** (ниже) показаны иллюстративные белки Gene Writer и ассоциированные с ними последовательности из множества ретротранспозаз, идентифицированные с помощью интеллектуального анализа данных. В столбце 1 указано семейство, к которому принадлежит ретротранспозон. В столбце 2 приведено название элемента. В столбце 3 указан номер доступа, если таковой имеется. В столбце 4 приведен организм, в котором обнаружена ретротранспозаза. В столбце 5 приведена предсказанная 5'-нетранслируемая область, а в столбце 6 приведена предсказанная 3'-нетранслируемая область; обе из них являются сегментами, которые, как предсказывается, обеспечивают возможность связывания РНК-матрицы с ретротранспозазой из столбца 7. (Понятно, что в столбцах 5-6 показана последовательность ДНК, и что последовательность РНК в соответствии с любым из столбцов 5-6 обычно содержит урацил, а не тимидин.) В столбце 7 приведена предсказанная аминокислотная последовательность ретротранспозазы. В столбце 8 таблицы 3B приведен предсказанный домен RT, присутствующий на основании анализа последовательности, в столбце 9 приведено положение старт-кодона, а в столбце 10 приведено положение стоп-кодона.

Таблица 3A.

Семейство	**Элемент**	**Номер доступа**	**Организм**	**Предсказанная 5'-UTR**	**Предсказанная 3'-UTR**	**Предсказанная аминокислотная последовательность**
R2	R2-1_TG	.	Taeniopygia guttata	GTCTAGTTACAACTGG GCATCGCTGCAGAGA TCGCACCTCCTCGTGG TCCCGCTGGTAGCCCT TCGAAGGGTGACTAA GTCGATCTCTGCCCCA GGTACGGAGCCGTTG GGACTCACCAGTCCA ACGTAACTCCTGCCTA AATTCGGTGAAACAA ATTCCTCGGTAAAAA GCCCC (SEQ ID NO: 1140)	TTCAGGTTATTTAGATG CTTAGTTTTTGTACCTT TCTTGTTTTGTTTAGGA TTTTGATAGTGTTAGTA TTTTTATATTTTTGTAC GATTGCATAATGTTCTT TTTTATACAGTTCTGTT TTAATAAAATAGACGA TAGCTAGAGACGTTAG GGCAGCCACAAGCCAG TTAGGTAGCGGATAGT AGGTAGGAACAGACTT TTACTATTTCATAACGC GTCAATTACCACCTGAT TTGGACCAATTCACGG GATTTGTCCAAGGTGG ACGGGCCACCTTTACTT AACCCGGAAAAGGAAC ATATATAATTTATGTGT	MASCPKPGPPVSAGAMSLE SGLTTHSVLAIERGPNSLAN SGSDFGGGGLGLPLRLLRVS VGTQTSRSDWVDLVSWSHP GPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKF PPNLAENDLALLVNLEFYRS DLHVYECVHFAAHWEGLSG LPEVYEQLAPQPCVGETLHS SLPRDSELFVPEEGSSEKESE DAPKTSPPTPGKHGLEQTGE EKVMVTVPDKNPPCPCCGT RVNSVLNLIEHLKVSHGKRG VCFRCAKCGKENSNYHSVV CHFPKCRGPETEKAPAGEWI CEVCNRDFTTKIGLGQHKRL AHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEELLIRL EAQFEGNKNINKLIAEHITTK

					GTTCGATAAA (SEQ ID NO: 1263)	TAKQISDKRRLLSRKPAEEP REEPGTCHHTRRAAASLRTE PEMSHHAQAEDRDNGPGRR PLPGRAAAGGRTMDEIRRHP DKGNGQQRPTKQKSEEQLQ AYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQT AQDCFGCLESISQIRTATRD KKDTVTREKHPKKPFQKW MKDRAIKKGNYLRFQRLFY LDRGKLAKIILDDIECLSCDI PLSEIYSVFKTRWETTGSFKS LGDFKTYGKADNTAFRELIT AKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTM EIFNLWLTTGKIPDMVRGCR TVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSK ACPLNPRQRGFIRAAGCSEN LKLLQTIIWSAKREHRPLGV VFVDIAKAFDTVSHQHIIHA LQQREVDPHIVGLVSNMYE NISTYITTKRNTHTDKIQIRV

						GVKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSIT AMAFADDLVLLSDSWENM NTNISILETFCNLTGLKTQGQ KCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKY LGLQFDPWIGIARSGLSTKL DFWLQRIDQAPLKPLQKTDI LKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLP PCTCDAILYSSTRDGGLGITK LAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHK KLWIQAGGDRENIPSIWEAP PSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWT KLASQGRGIVNFERDKISNH WIQYYRRIPHRKLLTALQLR ANVYPTREFLARGRQDQYI KACRHCDADIESCAHIIGNC PVTQDARIKRHNYICELLLE EAKKKDWVVFKEPHIRDSN KELYKPDLIFVKDARALVV

						DVTVRYEAAKSSLEEAAAE KVRKYKHLETEVRHLTNAK DVTFVGFPLGARGKWHQDN FKLLTELGLSKSRQVKMAET FSTVALFSSVDIVHMFASRA RKSMVM (SEQ ID NO: 1016)
R2	R2-1_Gfo	.	Geospiza fortis	AGACTTAAGTGAGTTT GGTTACAACTGGGCA TAGCTGCAGAGACCG CGCCTCCTCGCGGCCC CGCTGGTAAGCCCTTA ACAGGGTGACTAA (SEQ ID NO: 1141)	GGTAGATAATCTTTGTA TAGTGGGGGGGGATCT CATGTACCGGGTTTCTT TTATTTGATTTTCAATA AAACAGACGGTAGCTA GGTTCGCAAGGCAGCC ACAAGCCAAAGATAGG TAGGGTGCTCATAGTG AGTAGGGACAGTGCCT TTTGATTCACAACGCGT CAATACCATCTGACAC GGATACCCTTACCGGA CTTGTCATGATCTCCCA GACTTGTCCAAGGTGG ACGGGCCACCTTTACTT AACCCGGAAAAGGAAC ATATATTAATTATATGT	VGLCPSPGVDGTHQPNDSFQ NFGETNFSVQVARLVTRNL APRSVRGNGFGSGMATHPV PADESGHESDPFLVGRSCGQ PARLTRQSVGTQTSRDDILP SKTTKLTENELDLLVNFSLE LYRSDLQGFVQEGIHFSVNR EVLEGFPEVYEQPAPQPAVG DDLNTSLPPDNNICVLEKGS SEAVEDGTPEVAHPVPETQG KESPNNIVMVTLPNKNPPCP CCRVRLHSVLALIEHLKGSH GKKRACFRCVKCGRENFNY HSTVCHIAKCKGPKVEKAP VGEWICEVCGRDFTTKIGLG QHKRLAHPLVRNQERIDAS QPKETSNRGAHKRCWTKEE

					GTTCGGAAAA (SEQ ID NO: 1264)	EEMLIKLEVQFEGHRNINKLI AEHLTTKTSKQISDKRRLLP RKQLTDLSKGVAGQKVLDP GLSHQPQLGVVDNGLGGGH LPGGPAAEGRTIEPLGHHLD KDNGHREIADQHKAGRLQA HYRKKIRKRLSEGMISNFPE VFEQLLDCQEAQPLINQAAQ DCFGCLDSASQIRKALRKQN TQKDQGDQPKRPAQKWMK KRAVKRGHFLRFQKLFHLD RGKLAKIILDDVECLSCDIPP SEIYSVFKARWETPGQFAGL GDFEINRKANNKAFRDLITA KEILKNVREMTKGSAPGPD GIALGDIRKMDPEYTRTAEL FNLWLTSGEIPDMVRGCRT VLIPKSSKPERLKDINNWRPI TIGSILLRLFSRIITARLTKAC PLNPRQRSFISAAGCSENLKL LQTIIRTAKNEHRPLGVVFV DIAKAFDTVSHQHIIHVLQR RRVDPHIIGLVKNMYKDIST

						VITTKKNTYTDKIQIQVGVK QGDPLSPLLFNLAMDPLLCK LEEHGKGFHRGQSKITAMA FADDLVLLSDSWEDMNANI KILETFCDLTGLKTQGQKCH GFYIKPTKDSYTVNNCAAW TINGTPLNMINPGESEKYLG LQFDPWVGIAKTSLPEKLDF WLERIDRAPLKPFQKLDILK TYTIPRLTYVADHSEMKAG ALEALDRTIRSAVKDWLHLP SSTCDAILYTSMKDGGLGVT KLVGLIPSVQARRLHRIAQS PEETMKDFLEKAQMEKMYE KLWVQAGGKRKRMPSIWE ALPEVVPSIDTATTSEWEAP NPKSKYPRPCNWRRKEFKK WTKLIAQGWGIRCFKGDKIS NNWIRHYRYIPHRKLLTAIQ LRASVYPTREFLARGREDNC VKSCRHCEAAEESCAHIIGM CPVVRDARIKRHNRICERLM EEAGKRDWTVFQEPHIRDV

						TKELYKPDLIFVKEGLALVV DVTIRFESTKTTLEEAAAEK VNKYKHLETEVRNLTNAKD VIFMGFPLGARGQWYNKNF ELLDTLGLPRSRQDIIAKTLS TDALISSVDIIHMFASRGRRQ HA (SEQ ID NO: 1386)
R2	R2-1_ZA	.	Zonotrichia albicollis	CGACTTGAGAAGGTC TGGTTACAACTGGGC ATAGCTGCAGAGATC GCGCCTCCTCGTGGCC CCGCTGGTAAGCCCTT AACAGGGTGACTAAG TCGATCTCTGCCCCAG TCCAGGAGCCGCTGG GTTTCACCAGCCCAGC GATTCCTTCCAAATTC GGTGA (SEQ ID NO: 1142)	GTAGTCACATTGCACTT TCTGTAACTTGCACTGG GTGTGGGATGTGGGCC TGGGGTGTGGGTTATG GGGTATATATGTGGGA TATTCTGGTGGGAATGT CCATTCACTGTATGCCT ATCTTTTTAATAAAAAG ACGGTAGCTAGGTTCG CGAAGCAGCCACAAGC CAATAGCCAGTTAGGT AGCTCATAGTGGGTAG GTGACAGGAACCTTTG ACTCAGAACGCGTCCA TTAACATCTAGAACGG ACCAAACTTCGGACAT	NKFLGKSRVAYCLKPGPPVS DRGKEFGSGLTTHPEPESES GHDPTVPNPGPSLGAGEGA QPLPLLRVSVGTQTCEEDFIT SRPTKLPGIESELGPLVKFSL EVYRSDLKGDVQFEGIHFPD NWGVLEGFPEVYEQLAPQP NGGDELNHSLPGDREGDVL EKDSSEKEKEAAPEALPSVQ RARSEQLPDNIVKVTVPDKN PPCPCCGVRLNSVLALIEHL KGSHGRRRVCFRCAKCGRE NFNHHSTVCHYAKCKGPQI ERPPVGEWICEVCGRDFTTK IGLGQHKRHMHAMVRNQE RIDASQPKETSNRGAHKRC

					GCACCGATTAACCGGA TTTGTCCAAGGTGGAC GGGCCACCTTTACTTAA CCCGGAAAGGGAACAT ATATAGTTATATGTGTT CGTAATA (SEQ ID NO: 1265)	WTKEEEELLMKLEVQFENH KNINKLIAEQLTTKTAKQISD KRRMLLKKGRGTTGNLETE PGMSHQSQAKVKDNGLGG DHLPGGPVVDKGTIGKPGQ HLDTDNSHQITAGKKKGGG LQARYRRRIMKRLAAGTINI FPKVFKELINDQEARPLINQT TEDCFGLLDSACQIRTALRE KGKSQEERPRKQYQKWMK KRAIKRGDYLRFQRLFHLDR GKLARIILDNTESLSCDISPSE IYSVFKARWETPGHFNGLG DFEIKGKANNKAFRDFITAK EIEKNVREMSKGSAPGPDGI ALGDIKKMDPGYSRTAELF NLWLTAGDIPDMVRGCRTV LIPKSTTPERLKDINNWRPITI GSILLRLFSRIITARMTKACP LNPRQRGFISAPGCSENLKL LQSIIRTAKNEHKPLGVIFVD IAKAFDTVSHQHIIHVLQQR RVDPHIVGLVNNMYKDIST

						YVTTKKNTHTDKIQIRVGVK QGDPLSPLLFNLAMDPLLCK LEESGKGFHRGQSSITAMAF ADDLVLLSDSWENMKENIKI LETFCNLTGLKTQGQKCHG FYIKPTKDSYTINNCPAWTIN GTPLNMINPGESEKYLGLQI DPWTGVAKYDLSTKLKIWL ESIDRAPLKPLQKLDILKTYT IPRLTYLADHSEMKAGALEA LDQQIRTAVKDWLHLPSCT CDAILYVSTRDGGLGVTKL AGLIPSVQARRLHRIAQSPD ETMKDFLEKAQMEKMYEK LWVQAGGKKKGMPSIWEA LPMTVPPTNTGNLSEWEAP NPKSKYPKPCDWRRKELKK WTKLESQGRGVKNFRNDTI SNDWIQYYRRIPHRKLLTAI QLRANVYPTREFLARGRGD NYVKFCRHCEADLETCGHII GFCPVTKDARIKRHNRICDR LCEEAAKREWVVFKEPHLR

						DATTELFKPDVIFVKEDRAL VVDVTVRYESAKTTLEAAA MEKVDKYKHLEAEVKELTN AKDVVFMGFPLGARGKFYK GNFNLLETLGLPKTRQLSVA KTLSTYALMSSVDIVHMFAS RSRKPNV (SEQ ID NO: 1387)
R2	R2Dr	AB0971 26	Danio rerio	AATCCCCCCTACCCAA TCCCCCCGTCGTGACC TCCAGGCCAGGAATC ACGAGCGTACGACAG TGGCCATCCGGCAAT GACAATAGCGTGACT AACGACAATGAGTCA GATCCATGACCCTTGG AGTGGGTTAACCTCCG CCTCTTTAAAAAC (SEQ ID NO: 1143)	AAATCCCAGCGGGATA CAGCAAGAAGGTATCG GATCTAATAAGGTTGA GCGAGGAGAGGGTGGA GATCCTTTGGGGGGGG TCGGGCTAAGTTCCCCT CTCGGGTCCTCCCACGG TGACGCTCTACCCCTCC CTCCTCGCTCGTAGAAC CCAACGGTGAACACGG TTGGCAGGATGAAGTG ACGTGAGGGGTAAGAC ATGCGTACGTGAGCGC GCATTTTTGCTGTTCTC TGGACTGGGTTTCGTCC CCCTCACAACCATCACT	MESTAKGKSYWMARRPVE GATEGSLGRVPFVTRDPKRK PEAKRTLTHGLGLRECSVVL TRLIEGRRGRDHTPSGWNA QRGMPNDESSVEEPNGPIPS NPIPTGTQALPEPMADGEQG EHPGVVVTLPLRDLNCPLCG GSASTAVKVQRHLAFRHGT VPVRFSCESCGKTSPGCHSV LCHIPKCRGPTGEPPEKVVK CEGCSRTFGTRRACSIHEMH VHSEIRNRKRIAQDRQEKGT STDGEGRAGVERADAGEGP SGEGIPPKRPRRARTPREPSE PPANPPILSPQPDLPPGGLRD LLREVASGWVRAARDGGTV

					TACACTATAGGGGCAC	IDSVLAAWLDGNDRLPELV
					AGCGGCTCCTACCTCCC	DAATQRTLQGLPAGRLARR
					TCCCTATGACCCCCCCT	PATFVAPNRRRGRWGRRLK
					TCCCATACCGATCCATG	LLAKRRAYHDCQIRFRKDP
					GCTGTTCTAGTCTGGAC	ARLAANILDGKSETSCPINE
					CGAGGGTCGGACGGGG	QAIHEHFRNKWANPSPFGGL
					CATTTGAAGGTAGCTG	GRFGTENRANNAHLLGPISK
					GAATCCTCCGCTGCTGC	SEVQTSLRNASNASTPGPDG
					GAGCCTGAGGTCGATG	VGKRDISNWDPECETLTQLF
					GTTAGAGGTGAAATAC	NMWWFTGVIPSRLKKSRTV
					TTGGGAGGAGACACAG	LLPKSSDPGAEMEIGNWRPI
					CCTCCGGAGAGCCCCT	TIGSMVLRLFTRVINTRLTEA
					CCCGGGTGGTCATCAT	CPLHPRQRGFRRSPGCSENL
					GGCAACCGGGTGAAAC	EVLECLLRHSKEKRSQLAVV
					CTTACGGTTTCACTTAC	FVDFAQAFDTVSHEHMLSV
					GAAACAGCACCATAAC	LEQMNVDPHMVNLIREIYT
					AGCGCCGTAATAGCGC	NSCTSVELGRKEGPDIPVRV
					ACCGGTGTGACTACTGT	GVKQGDPLSPLLFNLALDPL
					CCAGTGCTGATATTCTC	IQSLERTGKGCEAEGHKVTA
					ATCTGGAGAATACAAC	LAFADDLALVAGSWEGMA
					ACGGGTAATGGCAGAG	HNLALVDEFCLTTGLTVQPK
					TATTCAAAACCCAAAT	KCHSFMVRPCRGAFTVNDC
					GTTTACGATCGACCAA	PPWVLGGKALQLTNIENSIK

					CGGAGTCGTTCCCTTGC	YLGVKVNPWAGIEKPDLTV
					ATCTAGGCCGGACCCG	ALDRWCKRIGKSLLKPSQK
					AAACTGCCGTAATTGC	VYILNQFAIPRLFYLADHGG
					CCGTCCCCAAGGTAGC	AGDVMLQNLDGTIRKAVKK
					CTCTTAGAAAACCGAA	WLHLPPSTCNGLLYARNCN
					GCCCGGTCGGGGCGGT	GGLGICKLTRHIPSMQARRM
					GGTTGCGGCGGCGCTG	FRLANSSDPLMKAMMRGSR
					CGGGGGCCTGCTGCTC	VEQKFKKAWMRAGGEESA
					GGGCGGCGTCGGTGTG	LPRVFGANQYQEGEEVAND
					CCGCGGTGGTTGCGGT	LVPRCPMPSDWRLEEFQHW
					GGTGCGGCGGGGATCT	MGLPIQGVGIAGFFRNRVAN
					CGGTCCTTGCGGTGCCG	GWLRKPAGFKERHYIAALQ
					CTGTGCCGCCGCGGTC	LRACVYPTLEFQQRGRSKA
					GCGTCGGTGGCGCTGG	GAACRRCSSRLESSSHILGK
					GGTGGTGGCCCGAGTG	CPAVQGARIRRHNKICDLLK
					GCGTCGGCGTGCCACT	AEAETRGWEVRREWAFRTP
					GCCCATAGTCGCCCGC	AGELRRLDLVLILGDEALVI
					GGGGGCGACCGATCTG	DVTVRYEFAPDTLQNAGKD
					GAGGGGCGAGGGGGCT	KVSYYGPHKEAIARELGVR
					CGCGGGACTTTAACGA	RVDIHGFPLGARGLWLASNS
					GAAACGGAACGCAACT	KVLELMGLSRERVKVFSRLL
					TCTCGCATCGCTCCCGG	SRRVLLYSIDIMRTFYATLQ
					GACTTTCCCCCCTCGTT	(SEQ ID NO: 1388)

					CAGCCGAGGGATGCCA AAAGGCATGAAAGGTA AGTACCATACCGGTCC GCAAAACTCTCTTCTGA CTCGGTTCTCTGTTGGT TTTCTAGAGTAACAAC GAGGTGGAGGAGAGGG ACATGGCAGGGACTCC CATTCGTGCCAGCGGG TGGGGACAGATCGAAG GAACGGTTCGAGGGCG TAACAGACGAGAGGGA ATCCGGTCACACATTG ATGCCATGCCTAAATA GGCGAGGTTTGTATTTC TACTTTGTGGGTTCAGT ATAGTCGGAGCATATG GTCGGTTGTCCCGTTGT TTTCACGGCGGGCAAG CGACTATCATGATAAA GTAGAATGGGAGACGG GCTCCCTGACAAACCC GGAAAGGCGCCCCCCC	

					GTGGTTCGTAGCAGCT GACGGATCACGCTCGA AGAAAAATGAGTGAGA GGGGACGCCGCAACCA C (SEQ ID NO: 1266)	
R2	R2-1_GA	.	Gasterosteus aculeatus	CATATTGGGGTCTCAG GAGGAGACACAGGGT CTGTTGCGGCTCCGGT AAACGGTACCGGAGT CGGTTAAGCATCGTTT GGGCCCGCCTCCACGT GGTGGTCCGCGGTAA CACCAATAGGGTGGC TAAGAGGCCCAGTAA TTTCCCCGAATTGTCT TCCCCCCCGCGCGGG GGGGACCCCCCTTTAG TGTCGGAGCGGTCGC GCCTCCGCGTTTGGGG TGTCGCAGGCGTGAG CCTTCGTCCCCTTAAG TTCAGACGGTCCCGGC TTCTTGCCGGGCCAAC	GGAGGGGAGTAGGTCT CTACTCTGACCCGAAG GGCCCCCCCGTTTCAGA CCTGATTCTAGGCTACC TGTGCCTAATTGGGGG GGTCCCAAAGAGATGT TGTCTGTTGTAGAAGG GTTTGCGCCACTGACTG CACGGAAGGGTGGGCC TCGACAGGTAGGGGTT ACATGACTCCGTGCTGC TCAGCAGACCCGCGCC TCTGAGACCGGGTAGG GCTACTTGAACAAGCG ACGCCCTGGTGTATGTC CGTATCCTAACCTGGTT TGGGAAAGCCGATACC GGCAATGCCCGCCACA	MLRGGVGTPPAGGAGAVGP GMASPGGCSVRFSPGGRRL LGHRTGGLSPSVSWRLKRLS VSLRRWSGPGLLGADGAGG GAAVASPRGTQVLGSGAGR RWLGHGSRGSSPSAARGLR RLTVRLKRLSGGLLSPKACR DAEEGSSSSPGFRNPKGLGG RGLTPLGSRRFCRLTVSLNR WRGSLVKLNASSRASGRRT PVKPACDSRAGRGSEHAEG GGVSAAPMVLRSRRKLTFS VDGDSNSGDRARSGSVSAA RPGHLLVDGESASSRSGPAG DARLAGPSTRSRRKGCLPPV DFENPKKRTRLMAKMTNGN PTSHVPCPAPCSNGHEGGGR VAVIEGRLPELSGSRISGIQP

				CCCCGGTGCAGCGTTC	GGTGTCGCGCACCCCA	ALPVETSFVGQSTGRGADG
				TCCCATGTTGGATCGG	CGGGATGACGTATGGG	DANANSSPPSPNLGGSVGM
				CACCCAGCCCCGGGT	CCCCGGGGGACCTCAT	VPAVRDGTPPLGRPGEDHSR
				GCCATGCGAGTTCAG	GGATACTCCACTGGAC	ECAGGNTPLWMLEDSFRCD
				ACATTTTGTTTATGTA	TTGCACAATCCTGGTGT	YCPREFGTRAGRSLHMRRA
				TCGTCTGCGTGGTTGA	ACTGGATGCAGCGACG	HLAEYDGAGFCWGERLSEF
				CTTGCTAAGCTCATTT	TTGGTGACATAAGCAA	AATRLWSTEETKKLAVFCE
				CCTCCTCTCACTGCGT	TCGCTAAGTCGGGGTA	RGVPSPSECRAIAASLGAGK
				CCCCCCAGGTGCTGAT	GGGGAGGTGGGGACCT	THHQVRSKCRLVFEAIRRRE
				CGGTTGAAGAGGATT	CGGCACGGCTGTAGGA	LLEVAAATERLEKSARRKQP
				CGTCGTTGACCTCGGC	ACGGGTGTATGGGCTC	AVPPAPVHGVRGVLRGLLG
				GGTGAATTTGGGATTG	CGGCAGCCGTCGTCAC	KRVPREGGTTGSTSARIVRR
				TATTATACAGGTAGGT	TCCCATACAACACAGG	DDCRQGAVASASLNLIRRLG
				ATAGAGGGCGTGCGG	GGCTGCATCCTGGTGG	RKATGRSGRRRVLGRPPRM
				(SEQ ID NO: 1144)	CCGGTGCTAGTTGGTTC	DVRRSVRMRRMRRFLYRLA
					TGGAAGCCCGCCCGGG	RLGWAKLAMFVLDGQMGA
					CTGGTTCGCAGAAGCA	SCPVPLVEVSAVFRERWSIV
					GGGTGCGCCCAGGGTA	RAFLGLGQFGGFGTADNAG
					GGTTTGGTATATCTGGG	FGKLIDPAEVRAHLQSIKNR
					TCCGGTGCGATACCTAT	SSPGPDGITKVALSKWDPEG
					CGATGGGCAGCGAGGG	IKLAHMYSTWLVSAGIPKVF
					CCGCCTCGTGACGCGCT	KKCRTTLIPKTGDVSLHGDV
					GTGTGGAGCTGGAGCC	GQWRPITIASLVLRLYSRILT

					GGCCTGGGTATGAACA	ERMTVACPSHPRQRGFIASP
					GTTCTTGCGGATGTGGC	GCSENLMLLEGCMSLSKAG
					GTAGCTAGATAGTACC	NGSLAVVFVDFAKAFDTVS
					CGTGGTTGTGGGCGTG	HEHLLSVLVQKGLDQHMVE
					GTGTCGACCAAATGTT	LIKDSYENSVTKVHCQEGCS
					GTCCTGTGTGCACATAG	TDIAMKVGVKQGDSMSPLL
					GCCAAGGGTTACGTGG	FNLALDPLIQQLEREGRGFP
					GTGGCAGTCAGAAGCA	VNGKSITAMAFADDLAIVSD
					CCCGCACCTGGAAGTG	SWEGMRANLDILVDFCELT
					ATTGCCCCGGGATCCC	GMRTQPSKCHGFLIEKSGSR
					GGCTCTCTGTGAAGAG	SYKVNRCEPWLLNDTALHM
					CTACCTTGAGGAAAGG	VGPKESIKYLGVQVNPWTGI
					TGTTCCGCTGGAACTCA	FAEDTVAKLRQWVVAISKT
					AGACCCTACAGTAGGG	PLRPLDKVSLLCQFAVPRVIF
					GATATCAACTGGCTTTG	VADHCMLSAKALTEMDRSI
					AGGTGCTGTGATTCCG	RQAVKRWLHLARCTTNGLL
					GAACCAGGGCGAGGGC	YSRKSSGGLGIPKLSMIVPA
					GAGTACTTAGAGCATG	MQARRLLGLSRSKDETVRW
					TCCAAAAGCCCGGGGA	MFLETTDHVAFERAWLRAG
					ACGTTCCGGGGGCCTG	GSPDEVPELGPDLVEGSPAE
					CTTGGGTCGTTGGACCC	GNADPVSTVRPRKRIVPCD
					ACATCCGTAAAACGAT	WRQVEFDRWAGQLVQGKG
					GGATCTCGCGTCGGCG	IRTFEADKISNCWLYDYPPN

					CTCGGGAGAACTTCCC GCATGAACGCTGATTG CATGTGAGAACGCCCC CACGGCGGCGGGGCAG GCGCTCCCCCTGGGTGT AAGGCTCGGGGGGGTC ACGGCTCCGCTCTAAA AG (SEQ ID NO: 1267)	KLKPGDFTAAVQLRANVYP TRELAGRGRTDTIDVCCRHC GEAPETCWHILALCPKVKRC RIQRHHKVCQVLVAEAERH GWEVEREKRWMLPSGECV APDLICWLDELALIVDVTVR YEFDEESLERARIEKECKYR PLIPVIRASRVQTKKVTVYG FPLGARGKWPAKNELLLAD LGLSKARTRSFAKLLSRRVL LHSLDVMRTFMR (SEQ ID NO: 1389)
R2	R2_BM	AB0768 41	Bombyx mori	GGGCGATACGCATAA TTTTAATTTCCCGATT GAAATCCAGTCGTCTT AATCTGGTGACCAGT GGCGCGGTCACCAGT ATAGTGCACAGGACG TGAATGGCTCCGAGG CTGGCGGAGTCACTC ACTATAAGTGTGAGA GACGATGTCCTGTGCC AAGTATACGTCCAAC	GCCTTGCACAGTAGTCC AGCGGTAAGGGTGTAG ATCAGGCCCGTCTGTTT CTTCCCCGGAGCTCGCT CCCTTGGCTTCCCTTAT ATTTAACATCAGAAAC AGACATTAAACATCTA CTGATCCAATTTCGCCG GCGTACGGCCACGATC GGGAGGGTGGGAATCT CGGGGATCTTCCGATCC	MMASTALSLMGRCNPDGCT RGKHVTAAPMDGPRGPSSL AGTFGWGLAIPAGEPCGRV CSPATVGFFPVAKKSNKENR PEASGLPLESERTGDNPTVR GSAGADPVGQDAPGWTCQF CERTFSTNRGLGVHKRRAH PVETNTDAAPMMVKRRWH GEEIDLLARTEARLLAERGQ CSGGDLFGALPGFGRTLEAI KGQRRREPYRALVQAHLAR

				CCTAACGGGTTAAGT	TAATCCATGATGATTAC	FGSQPGPSSGGCSAEPDFRR
				GAAATTAGTTGCTCAT	GACCTGAGTCACTAAA	ASGAEEAVEERCAEDAAAY
				AACAGGGACGGTGTA	GACGATGGCATGATGA	DPSAVGQMSPDAARVLSEL
				CCTGTTTGCTCGTGGC	TCCGGCGATGAAAA	LEGAGRRRACRAMRPKTAG
				TGGCTATCGAATGGA	(SEQ ID NO: 1268)	RRNDLHDDRTASAHKTSRQ
				CGGGACCAATACACC		KRRAEYARVQELYKKCRSR
				CCCCTGTTAGTAATGG		AAAEVIDGACGGVGHSLEE
				GGTAAGAGAGAGCGG		METYWRPILERVSDAPGPTP
				TCTGAAACTATGGCCG		EALHALGRAEWHGGNRDY
				AAATCACGACGCCCC		TQLWKPISVEEIKASRFDWR
				ACTCCTACCCATAACC		TSPGPDGIRSGQWRAVPVHL
				TGCACGTGGTACCGCC		KAEMFNAWMARGEIPEILR
				GCACATTGACCGATA		QCRTVFVPKVERPGGPGEY
				CGGGAGGAGGGGCAG		RPILIASIPLRHFHSILARRLL
				CACTTGAATCACGTAG		ACCPPDARQRGFICADGTLE
				TCTTGGTGTAGCCATT		NSAVLDAVLGDSRKKLREC
				GCGGGACTACAGCCC		HVAVLDFAKAFDTVSHEAL
				TCGTAAGTGCCGCCTT		VELLRLRGMPEQFCGYIAHL
				AGAACGCAACGGGGC		YDTASTTLAVNNEMSSPVK
				AATAGGTGGGCCGGG		VGRGVRQGDPLSPILFNVV
				GCGCTAGCGGGGGGG		MDLILASLPERVGYRLEMEL
				AGTAATCTCCCCTGTT		VSALAYADDLVLLAGSKVG
				GGCGTGCACCGCACT		MQESISAVDCVGKQMGLRL

				GCTCCCACTGGGGGC AGTGTCATCCGGAAA CAGGTGGGCCGGGGC GCCACCAGGGGGGAG CAATCCCTCCTG (SEQ ID NO: 1145)		NCRKSAVLSMIPDGHRKKH HYLTERTFNIGGKPLRQVSC VERWRYLGVDFEASGCVTL EHSISSALNNISRAPLKPQQR LEILRAHLIPRFQHGFVLGNI SDDRLRMLDVQIRKAVGQW LRLPADVPKAYYHAAVQDG GLAIPSVRATIPDLIVRRFGG LDSSPWSVARAAAKSDKIR KKLRWAWKQLRRFSRVDST TQRPSVRLFWREHLHASVD GRELRESTRTPTSTKWIRER CAQITGRDFVQFVHTHINAL PSRIRGSRGRRGGGESSLTC RAGCKVRETTAHILQQCHR THGGRILRHNKIVSFVAKAM EENKWTVELEPRLRTSVGLR KPDIIASRDGVGVIVDVQVV SGQRSLDELHREKRNKYGN HGELVELVAGRLGLPKAEC VRATSCTISWRGVWSLTSY KELRSIIGLREPTLQIVPILAL

						RGSHMNWTRFNQMTSVMG GGVG (SEQ ID NO: 1390)
R2	R8Hm-A	.	Hydra vulgaris	TTCAAGTGGATGAAG CTGGGAAGGTAATCT GTAGTTGGTTGAGTTG GTTGCAGATTACTGCT GTCGATTTTGCTTTCT ATTGAAAGCCTGTCTC TACGGGTCCTGAAGCT TGAATTTTGGTAGCTA TAGTTTTGTGGGAGGA AAGTGGAATTTTGTAC CATCTTTTGTCTCTCG TATCTACTATAGTAAA TCCGGTCATGCAGCCT CTACGCGGCGCAACT AGAAACTTGGATCAG TGATCAAGGCTAATG CATGCCGGGTCTCCTC AGATTAGGAGTATAA TACAAATCTGACTTCA TCACTAAGAGGCTAT GGGGCTAACGATCCT	TAAATGCCAAAAGTTG CTTGGGCTAAATGATA CGTACGCTAGAAAAAG CGACTTGCTGCACGGA TGACGGTTCATCAGAG CCCGATATGTGCATGTC AAGGCGGCAGGGAGAA TCACTAGTGTAGCTGTT CTTTCCATTACGACTTA CGCGGTTAACGTGGCA CGATAGATTTACACCA GGAAATAATACGTGAA GGGTTCCACCATATACT GGAGTTTAGATCTATG AGGGAAACATTTGTAA TAAGTCAGTCTGGTAA CCTGGCGCCGCTGTTGA GTCAAATTAACTATGTC AATACTCATTAAGTTAT CGACTTTGATATGGCAT GGGGTGATTCCGCGTT	MNLLIVTSSIKESDVPSSGKG GVAVNNITAGASGKDTCVII HPGTDGIWCCTECVEIHNSG KDLKRHLAKRHPSVTISGYK CNLCPFVSERQLSVGTHLRY CRGVKEVVKREFACASCSFS SDTFSGLQVHMQRKHIAEW NDQLKEKTEFAWTDRELRE LAEKELTTPSFRYNKIFYAA LGTSRTYDAVRKIRYNDRY KSAIAEMRSQIADAAAAAQ ERDVERGLVSAHSDRGKEM LPVVETKSDIQVNNDIKKDI ELTPNSRQKQTNLALARPAV IEVEEDLGRQDVKQYLASLR QDDYTSPAERSIFAYCREET NWSATKRQVLKISRTTRGLR QPKKVRPFEFPEGFKPNRNM RKWRKYRFLQECYREKRAE TVSKILDGTFIDEPEEEIRPEL EEVQRMYIDRLEKRTQLDTT

				ATAGTCTCG (SEQ ID NO: 1146)	ATATCAAAGTCAAACA TGATGATTGCAATGAG AAACTACCACGCTTGG TCACGTTTGTGAGGAG AACATCTCATTCAAGCC TCCCGGATGTCGGCAC CCGCTGACATCTTCTGG CTTATGAAAATTTTCAT TAATTTTTGTAAGTCAT GGGCGGCTTGAAAGC (SEQ ID NO: 1269)	KIVQTDEVFCLQSYGRITIGE VRDALGASKKDSASGPDGL LLQDVRRLGPLLLCNIFNM WYLHGIPVEENRCRTILLYK SGDRHLASNYRPVTIGNML NRLYAKIWDKRIRKNVRLH VRQKAFIPVDGCFENVKTIQ CVLQSYRKRKLEHNVVFIDL AKAFDTVLHDSIRKALWRK GVPSGVVKVVDSLYAGAVT SISVGKTKTRSICINSGVKQG CPLSPLLFNLILDELAERIEA TGCGLDLDGHVLSSMAFAD DYVLLAKDSVEMNELIRVC STFFKEKGLSVNPGKCQSLR VLPVKEKKRSMKVLVRPHR WWRIKDQDVDIPSMTYDSL GKYLGVSIDPTGKIALPIEEW KNWMTKLKECKLKPEQKV KILKEVVCSRVNYVLRMSE CGISELRSWTRFVRNWAKNI IHLPTWCSSDWIHSIKGLGIP DVSKGIVIQRMRASEKMSTS

						EDGIVRVVGARLVQKNRVL WEKAGFEGIELKAARRHCE VERLNNIGNITNGVALKTIA AVSSVNRYWMIEDNLKSGN KILVWKAMAGAIPTKINLSR GVADQTLKKCRRCGLTAET DGHILAGCHTSSDAYSKRH NMLCDKLAKELKLNGGPNR RVWRERTCFTSTGRRYRPDI IVKDDSKITVIDMTCPYEKSE GHLIQCESAKVTKYEPLKLD KYWTRELEGANGIVAEKVE LMGLAIGAIGTIMRSTLRKL CELKSGRIVRRLQMIACNNS AQIIKGHLSRATRRNLR (SEQ ID NO: 1391)
R2	R8Hm-B	.	Hydra vulgaris	CTTGGGGTCACTGACA CATTTTTCGGTAGCCA TAGTTTTTTGAGAGGA AGAGTGGAAGTTTTTC CATGAGTCGTCTCTCG TATAAACTGTGGTAA ATCCGGCCATCCAGCC	ATGCCCGAGGTAGTTG GGATAATGATGCACAA GCTCGTAAGGCGACTT GCTGCACGTATGCCGCT AAACGCTTAGCTCGAT GAGTGCATGTCAAGAC GGTCGGGAGTATGATC	MSNRITIGDVPSVGKGGLTV NKQTAGADGAEACVVIHPG AKGIWSSPACLRKFTIGKEL RAHLAQIHKLAPSAVRYRC NKCPYEGDVQLSVGTHLRY CKGIAGVVEEKKQFACAICN FSSDTFSGLQVHKQRKHVV

				TCTACGCGGCGCAACT	AGTGGAGCTGACTTTCC	EWNEQLKEKTEFAWTDREL
				AGAAACTTGGATCAG	AGACAACTCACGCGGA	RELAVKEVTIPFSVVNTETF
				TGATCAAGGCTAATG	TTCGCGTGCGGTGGAT	AVLDITTRTKDAVRKIRYTD
				GATGACGGGACTCCA	ACAACACCTGGTATAA	RYKSILAEVRAQVNAVAEE
				TGGATAAGGAGATAT	CATATGAAGGGTTCCA	APQASDESQITLLVNTGRGA
				AAAGATCTTATTTGAA	TCTAGTACAGGGATAA	ELQPAVINITDSIELVTDVNE
				CGCATCTTAAGGGGTT	CGATCCATGGGAGCAA	VEMVTSNSTNEEQPINAPVE
				ATGGGGCTAACACCC	ACTAATTAGTTGGAGG	PAVIEADLGRQDAKLYLASL
				CCTTAATTCTGGTGCA	TAATCCAACGCCGCTGT	RQSDCTNASDRWTLAYCRG
				CATTTATTGACCGTT	TGAGTCAGTTTTTAACC	EVDWCKTKSRLFKVSRHAR
				(SEQ ID NO: 1147)	GCCAGTCAACTCTTGTA	GLRQPQRVENWEFPEGFRP
					GGTTATCGGTCTTCGGC	NRNLRKWRKYSFLQSCYRT
					AGACCTTGGACCGCCT	KKKETVSKILDGTFKDTPEE
					AGCGCCGGCCAACAGT	EIRPELEEVQRVYVDRLEVR
					TTGTCGTCGACTAACAT	TQLDTTRTVHIDERFDLVSY
					GATGATTTGCGAGAGA	GRITIREVQDAISASKKDASG
					AACCCACGCTTTGTCAC	GPDGLLLQDVKKASPRQLCI
					TTATGTGAGGATAAAA	IFNMWYLHGIPVVENRCRTI
					TCTCTTGTCCATATGAT	LLHKGGEKHLTSNYRPVTIG
					CCTTTGAAGGGAACAG	NMLNRVYAKIWDRRIRKNL
					CGCTTTGAGCTTGCTCG	QLHVRQKAFVPLDGCFENV
					GCGTTGGCACCTTTAGT	KTIQCILQSYRRSRREHNVV
					CTGTAATATTTTCTTGA	FVDLAKAFDTILHDSIEKAL

					TATTATGGACGAAAAA GGTAGTATGGTTGCA (SEQ ID NO: 1270)	LRKGIPRSVIKVVDSLYAGA VTSITVGKTKTRPICINSGVK QGCPLSPLLFNLVIDELAERL EATGCGLDLEGHVISSMAFA DDYVLLAKDSVEMNVLMN VCNTFFEEKGLAVNPAKCQ SLRVLPVKGKRSMKVLTRT HRWWKINNQDVEIPSMTYE SVGKYLGVMIDPAGKIALPI EEWKLWLTRLRECKLKPDQ KVKVLKEVVCARANYVLR MSGCGICELRKWSRFVRGW VKSIIHFPAWCNSEWMHSSK GLGIPDVVSGIVIQRMRAAE KMAKSTDGVVRVVGARIVQ TNRVLWKRAGLAGIELDAA RKFCEVKRVNKIGNQTNGG ALKTIAESSVSRHWLLEKNI RPGNKILVWKAMAGVIPTKI NLSRGVADQTLKKCRCCGL TAETDCHILAGCPTSRDAYS KRHNLLCDKLAKELRLNGG PSRRVWRERMCLSGNGRRY

						KPDIVVKDDGVITVIDMACP YEKSERHLSQCEDAKVAKY EPLRLDRSWTQELEGNNGR SANEISVVGIAVGAIGTITRK TQRILSKLKLAKVGRPLQIIA CNESAQIIRRHLSGSRLRNLR (SEQ ID NO: 1392)
R2	R9Av	GQ398057	Adineta vaga	GAAATAGTTTGCAAT GGTAGGTGTATGGCG CCTCTGTGTCTCTCTT TCGCTGGATATAGTTT GACGATTTTGTACCAG GTATCTGTTTCTTGTG AGTTCAGCACCAGTTT GAACAGGCTTAGCGA TAGACCTTCGAACTTG AAACACTGTTGTGAA GCTGGCTGGGCCCCTG CAGATTTTCTCGATTA GAACGTGAGTGTTAC GTCCAGAATGACCCA CCAGTGGTTAGTTCTA CGTTGCCCTGGAAAG	ACTAGTCTCCTTCTTCT ATTAGTCAGTCTAATTA ATTTTTCTTACATTCTA CATCTAGTTCCATTATT AAATTGGTATGATCAG TGCTATCTCTGCTACAC TCAATGCTTAATCGTAT GTTATTGACAGTCTGAC ACTTGATTACTCTTACG ACATATGCACTGTTTGC TTCAGAGAAACCACTG TTCATATAGTGAAGTTC CTCAGTTTTCTGTTGAT ATATTCTTCTTTCATTC TCGCTTCTCCTTTTCTA CTGTGTTCTTTTTATCA	MNLPIREHAVSVHNINKFNY LCQLCSKSYDTINSVKAHYV ACRRQKNASSTTAVPTNVIN NNQLAINTNQVISRNPLQCV ECLMKQVDFYAKDTKALVT HMRTKHAAAYEESKKVATR RVAWSPDEDQILAELEVKL KKIQKGQLLSRLVVEYNKC ADKSKAPSRSKDAIRTRRQQ HDYKLLLRSLQSQQPPVGSE DSDSDISSSNNNPLTTTHNV TPTPDSSNVVLLIQKIRESVD SIVKITNLKLNTNMLNAASA FINQNNNMDPLELSMRGIEE DVKAIRDKELQKPTRNVPSS TTSRKPTRNAKRLEKSKKY

				GAGAAAAGTTGAGCT AAAATCGCACGGCCT AGTTGTTTATCAAATA GGCACGGTGAGGAAC TCTTCTATGTACCCTG ACTAAAGTACTCACTT GTGCGCTGGGTTTGCT CCCCCTCGCATTGACT TATCTGATCGCACTAC CCACCAAACGAAACA TAAACTTAGCTCGTGG TATCAGTCCACAGCGT GTGCAGTCGGATTCA GGGGAGCGTGTTAGT GACAAGCAGGATAAT ATTAACATAGTTAATG TTAAGGCGTTCAACAT TCCTTATCCAATTGGA AGAGTTGACTGTGAA GTTTGTCATGAAGACA TTGGACAA (SEQ ID NO: 1148)	GTTTTTTGTGGAAAAAT TGAGAATAAATAAAGT (SEQ ID NO: 1271)	GYYQHLYYNNKKKLVAEIL DGETSGAKPPPMNLVEDYY RNIWSRSTIDDSPVNNIKTV NSDSIFAPISRDEIKLALSNT KKDSAAGPDAVTIKEAKAII DNLYVAYNIWLGVQGIPEQ LKLNKTILIPKGNSDLSLLKN WRPITISSIILRVYNRLLAYR MNKIFKTNDKQVGFKPVNG CGINISWLHSLLKHARLNKN SIYACLVDVSKAFDSVSHQS IVRALTMNGAPSLLVKLIMD QYTNVNTVITCSGSISNKINI SSGVKQGDPLSSLLFNLVID ELFDVIKDQYGYTIDNIGTT NARCFADDLTLISSSRMGM NKLLELTTKFFKERGLNVNP SKCMSIGMSKGYKGKKSKIE SEPLFSITDAQIPMLGYIDKT TRYLGVNFTSIGAIDAKRIK KDLQDTLDKLEHLKLKAQC KMDLLRTYMIPRFMFQLIHT ELYPKLLIKMDILIRKLAKRI

						LHLPISTSSEFFYLPFKEGGL QLTSLKEAVGLAKIKLHKKI MSSNDPMLCYLIESQRSRIV EHFMKDLKLGDSLTLNEMN NIKECFMKEKRISFAQKIHG VGFEVFSSSPLTNQWINGEIK TMTTKTYINSIKLRTNTLETR VTTSRGLNIIKTCRRCHVAD ESLMHVLQCCSSTKGLRYSR HHKICAKVANKLVMNGYG VFREKSYPDPNNSGSYLRPD IIAVKNGHVIVLDVTVVYEV TGATFINAYQTKINKYNAIM VQIEQMFNCVNGELHGLVIG SRGSIHHSQLHIWHQMGFSS IELKYVAIGCMEDSLRIMSTF SKAIT (SEQ ID NO: 1393)
R2	R2Ol	LC349444	Oryzias latipes	CGCACAGGGGACACA GAGCCTGCCCAAGTA CCGCTCCCGAGGGAG CGGGAAACGGGGGGG TGACTATCCCCTGGGG TCCGGCGAGAGCGCT	GGGGGACAGCTGGGAG TCTCGGCATGATTACAA ATCTTGCGCTGCACTCG GATGTCGTCCCCGTGAC GGACACATTAATCCGG AAAGCGAGTGGTGACT	MGTDTVYVGQDYPSGLSKR VPARLVAGPMLRERSCHAH VFRAGHMWNWRTSLPSGR WDQPALEKSRVLTRSVATA TDPEITSYPGKSVSTSTQVQE EDWCSRESGWISPGLAPEEP

				GGTCTACGGACCAGG GGTGGCTGTGGGCAG GCTGCTCCTCAGGCCA GTTGATTAGTTACGCA TGGGCTGTACCTCCAC GTGGTCCCGCTGGTAA CGACTTGTCGGCTAAA TCAGCCCGCCCACCAT CTGGGATATGGTTGAC CGTCTAACCCCAGTAC TCAGGTCACAAACAA A (SEQ ID NO: 1149)	CGCCTCAAG (SEQ ID NO: 1272)	SVVSEITASMVATMRVATEE VVLEPQPEQVVTILPEHGRN VPPGLAEQDTASPIEVSVLLP DLAENCPLCGVPSGGLRLLG KHFAVRHAGVPVTYECRKC AWRSPNSHSISCHVPKCRGR ARMPSGDPGIACDLCEARFA TEVGVAQHKRHVHPVEWN KVRLERRGARGGGIKATKL WSVAEVETLIRLIREHGDSG ATYQLIADELGRGKTAEQV RSKKRLLRIDTASNSPDDAE VEEERLESLAVRSSSRSPPSL VATRVREAVARGESEGGEEI RAIAALIRDVDQNPCLIETSA SDIISKLGRRVDGPKRPRPV VREQTQEKGWVRRLARRKR EYREAQYLYSRDQARLAAQ ILDGAASQECALPVDQVYG AFREKWETVGQFHGLGEFR TGARADNWEFYSPILAAEV KENLMRMANGTAPGPDRIS KKALLDWDPRGEQLARLYT

						TWLIGGVIPRVFKECRTKLL PKSSDPVELQDIGGWRPVTI GSMVTRLFSRILTMRLTRAC PINPRQRGFLASSSGCAENLL IFDEIVRRSRRDGGPLAVVF VDFARAFDSISHEHILCVLEE GGLDRHVIGLIRNSYVDCVT RVGCVEGMTPPIQMKVGVK QGDPMSPLLFNLAMDPLIHK LETAGTGLKWGDLSIATLAF ADDLVLVSDSEEGMGRSLGI LEKFCQLTGLRVQPRKCHGF FMDKGVVNGCGTWEICGSP IHMIPPGESVRYLGVQVGPG RGVMEPDLIPTVHTWIERISE APLKPSQRMRVLNSFALPRII YQADLGKVTVTKLAQIDGI VRKAVKKWLHLSPSTCNGL LYSRNRDGGLGLLKLERLIP SVRTKRIYRMSRSPDIWTRR MTSHSVSKSDWEMLWVQA GGERGSAPVMGAVEAAPTD VERSPDYPDWRREENLAWS

						ALRVQGVGADQFRGDRTSS SWIAEPASVGFAQRHWLAA LALRAGVYPTREFLARGKE KSGAACRRCPARLESCSHIL GQCPFVQANRIARHNKVCV LLATEAERFGWTVIREFRLE DAAGGLKIPDLVCKKADTV LIVDVTVRYEMDGETLKRA ASEKVKHYLPVGQQITDKV GGRCFKVMGFPVGARGKW PASNNTVLAELGVPAGRMR TFARLVSRRTLLYSLDILRDF MREPAGRGTRVALIPAATG AAN (SEQ ID NO: 1394)
R2	R2_LP	AF01581 4	Limulus polyphemus	TGGGAGGAGACCCAA ACTATCCTAGGATGG GGCGGAACCGACCAT ATGAGCCATATTAAC ATTGCCCACACTATCC TCTGGAGGTACCTCCT CGTGGTACGGCTGGA TATAGGTAAATCCTGT AACCAAATCCTCCAA	ATTTTGTCTCTTTCCCC AATGATGTCTACTAGC ACGCTGCCGAAGCTAG ATAGATTGAGGAATCT GCGTAATCTGTAATGAT TACGCCTCATGGGCATC TATCGGTAGCGTCGAC CCTGACGTTAAATTGG	GIDGYMFGYARASGSTSVSI QSSSMTEGETNERATPRASD SSSVSIQSSCVTEGECLPPTD NCNPSVENQLPCVTEGRFER VGSLVTVRLPFRKVACDLCS KEFLTYSKFAVHQANFHNSE TQACCTYCGKSDGNHHSIA CHVPKCPWRRTVTFAANLS NFLCDLCNDSFKTKSGLSQH

				CCCGTGAAGGAGAAC ACTAAAACCCATATA GTGGCCTCGCCAACC ACTATATGTCCAACGG CAGGAGAAGCTATCT CCCGGATGGGAAGGA AAACCCTAAACCGTG ATGGGAACTTACCGG CCCCATCAGCTATTGG GTACCCGGTAGGGAC TTGCAACCCTACCCTG TATTTGCATTTTATAG GGAACCGGTCGGCCC TATATCAGAGTAGAC CGTTTATTAAATATGG GTGAAAATATTAACA GTAAAAGCTATGGTTT GGCGTCCGTGTGGTGC CAGGGCGGCGGCCAA ACCCGAGCTACTTGGC ACCAACTGGGGATGG TAGCTTCCGAGCGATT CCCTGGCGACGTGGG	GTAATAAGAAATATCG A (SEQ ID NO: 1273)	KRHKHPCSRNAERILSLGVR TPSARPRQVVWSEEETRTLR EVEVVYSGQKNINVLCAGH LPGKTSKQVSDKRRDLHRIR SSNVHGTPTTQSRGDPVEQV EEYEELDWEGMHPFPDPDS KFCSYLDQLRDQKGLTEPV WQEIEIVAQEWVENLAHVQ SSWNHERTTKQVPENNTPA RRPFKRRLHRVERYKRFQR MYDLQRKRLAEEILDGREA VTCNLKKEEIKDHYDQVYG VSNDRVSLDDCPRPPGANN TDLLKPFTPTEVMDSLQGM KNGAPGPDKITLPFLQKRLK NGIHVSLANVFNLWQFSGRI PECMKSNRSVLIPKGKSNLR DVRNWRPITISSIVLRLYTRI LARRLERAVQINPRQRGFVP QAGCRDNIFLLQSAMRRAK RKGTLALGLLDLSKAFDTV GHKHLLTSLERFAVHPHFVR IVEDMYSGCSTSFRVGSQST

				ACCGATCGACGATGG		RPIVLMRGVKQGDPMSPILF
				AGTCCAAACATCCGG		NIALDPLLRQLEEESRGFMF
				AATAGAGGAATTGAG		REGQAPVSSLAYADDMALL
				AAATACCTATTCCACC		AKDHASLQSMLGTVDKFCS
				ACCGGCTCACATACCC		GNGLGLNIAKSAGLLIRGAN
				AAGGTGAACCCGGTG		KTFTVNDCPSWLVNGETLP
				CAACTAGAGTACAAC		MIGPEQTYRYLGASICPWTG
				CTATCTGTGGCGGTAG		INSGPVKPTLEKWIANITESP
				GTGCCGAACCACTCA		LKPHQRVDILCKYALPRLFY
				GGTGACGGGCTTGTTT		QLELGTLNFKELKELDSMV
				ATTGATGTCTCCCTAC		KQAVKRWCHLPACTADGLL
				GAGACACGAATTGTG		YSRHRDGGLAVVKLESLVP
				ACAAATCCACTCCGGT		CLKIKTNLRLVHSTDPVISSL
				GGACAATTACCCGAT		AESDGLVGAIEGIAQKAGLP
				CTATGAACCTGTTACC		IPTPDQRSGTYHSNWRDME
				GATATTAGACAAGAA		RRSWERLALHGQGVELFKG
				AATAAAGAACTGACA		SRSANHWLPRPVGMKPHH
				ACGCCTAGAGCTTCA		WVKCLAMRANVYPTKRGL
				GGCAGCATGTCTGTA		SRGNLSKNKDSAKCRGCTS
				AGTATCCAGTCATCGA		MRETLCHLSGQCPKLKSMRI
				GCGTGACTGAGGGCG		RRHNKICEHLIAEASFKGWK
				AAATTGATAATAACTC		VLQEPTLVTDNGERRRPDLI
						FHRDDKAVVVDVTVRYEIS

				TGAAACTGA (SEQ ID NO: 1150)		KDTLREAYASKVRRYGCLT EQIKDLTGATSVVFHGFPMG ARGAWFPESSDVMADLNIR SKYFEEFLCRRTILYTLDLL WKSNNEQYLERLAP (SEQ ID NO: 1395)
NeSL	NeSL-1	Z82058	Caenorhabditis elegans	GCTCACTTTCTATCGT GTTAACCGTACGTTTA CACTCCCAGTGAGTGT AATAAAGGTTATTCG ATAGAGGGTGTCTCCC TCTTTCTTGGGTAATT CTTCGGCGGTCCGGG GTCTCTCCCTCGTCTT TTTTTTAAACTTTTCTT TCTCATCCACTCTTTT GCTCCTTTTTACTAAC TCTTGTACTCTATAGT CTTTTCTCATCCCCCA TCCGCCGTTGGGCAA AGTTTATTTACTTTGT TAAATCCATATTTTAT CTCTCTCACCCGTACA	CCTCCAGGGCACGCCG CACGCCAAAAGTCCTG GCATAACTCTGCAAAT AACATCAAACGTCAAT CAACTCCACAAACTCTC CACTCTCTTCAAGTCTT CTCGGTGCTTCCAACAC CACAATGGTGAAAGCT CCTTCACCTTTTCCCTC CAAAATTCTTCCCATGT GGGGAAGTCCTGTTCTT GTAAGCTCTCCGGAGG CTGCAAGAGCAGAAGA AATTCTTCTTTCTGACA AGGTCAGAAGGAAGTC CTGTTCTTGAGGCGTCC ATCCCGGGCGTCATAG	MLRRKGRHRMVMVNSVK WQPSAHAEAIGTGKSWAPQ RSQASEHGWQSNAMFDPPN RILFARDSWSLNQSTHLQNQ RSGSGLGIRPGQVRNNMVG GGPHRAGDPKRRVELVSIQG SEVTVRTIYPSDEIFSCYSKS CDIKTKAGYGPEDLKHLTR HIKNEHGLKARWAYQCGLC NEKSDPSVSEGHKWMEAH MVAVHQSSAEKRIKSYQKC TGARVAEQLQAAAPSLTVP GKHKSGSRDAAKDSMTPTK DDDPKTRIYQTRSVVKKSTQ KTAEPTDEGSRGPKYASIFQ KSVKARKSLALLCELSSPKP MNPLPTNELTLKEGNSRELA

				GAAAGCGTCTCCTTCT	GAGAGATCAGATGCAC	KEEAPSEGIDDIVIIDLDESEE
				CAAACGCTTTTCTGTA	CTTCTAGCAGGAGCTA	SPPRRKRFNTWCLDHESSRE
				CTTTTTCTTATATTTTC	GAAGGGCTGCCCTGTC	AWLDDTAIFWYISYLCRGST
				ATTAACATATTTTTCC	TTGAGATCCCCACGGG	KYSALDPCLWSMYKVKGSR
				TGTTTATACTAACCTA	GGTCAATAGACGGGAG	YILDRLESSITYFFPICEEDH
				ACCTCCATTGTCAATT	GGGCTGCTGGCTTTCTC	WTLLVLKDNSYYYANSLHQ
				ACTAACTAACTTGTAC	TTTTTAAGAGGAAGCA	EPRGPVRDFINDSKRARKEF
				AACGGATTTCG (SEQ	CCAATCCGGAGATCCTT	KVQVPLQRDSFNCGVHICL
				ID NO: 1151)	AGGGGTCAAAGGATTA	MTNSIMAGGKWHSEEDVRN
					AAAGGCAGCAGGTCCA	FRKRLKKTLQEEGYELYSV
					ATTCTCCTCACTGACTT	NSLGIPFQAPTTEQMDYKET
					CGGTCAGAGAGGAGTC	RCKRSYASVLTQISPPAKRP
					CCGCCTTGGAGACCTCC	DCKPDNNIFVPTKDCAAEG
					CCGGGGAGGTTGCTGA	NPQEKGRNESPEEINTEHIVV
					AGAGGCGGAAGCTCCT	AGKPANNISPRCRSTSEMLF
					TCTAGCAAGAGCTAGA	EMVKATTSSGRSSLGTMTQ
					GGGAGTTCCCAGTCCT	DEFIRTSTIAEAVPLMSIKLPP
					GAAACCCTTGCGGTTG	MELPRKILPPIPPRKPTQTNG
					ATGATGGAATGGAAGA	GQKGKQQRVPTGKPDTLNA
					GTACTTCGGTACTGCTC	KVRNWFNNQLESYAMEGRS
					GTTGCTCTCTCTGCGTT	FQRLEWLTEVLTASIQKAAA
					TTACTGCCGAGGGCCG	GDEGIVDIICKRNPPLEVAK
					GATTTGCTCGAATCGCG	GEMCTQTENKRKTTNNAAR

					AAAGGTCTCAATCGAC	IADPIQSSKGAGDVKASYW
					CATTCAAGATGACGGC	KERARTYNRIIGSKEELCKIP
					TTATCTAAGGTCCGAA	IDQLEDFFKKSTSRTNVQESI
					AGCAGTTGGGAGAGTA	MKEKSSKIPALKIGNWMEK
					ACGTGTTCTCCTACCTT	KFIGKEVAFALRKTKDTAQ
					TCAAGTTGAATGGTCGT	GADGLRYHHLQWFDPSGEL
					TTTACTGTTTGGGATAG	LAKVYNECQRHRKIPKHWK
					CTGACTTGATGCTAGTA	EAETILLFKNGDQSKPENWR
					CGCTTCATCTGTGGATG	PISLMPVIYKLYSSLWNRRIR
					ACGCTCCCCAAGCAGT	AVPNVLSKCQRGFQEREGC
					CAAGTAGACTTGAAAG	NESLAILRTAIDVAKGKRRN
					GTGCCCTCGCCCTAGTT	LAVAWLDLTNAFGSIPHELI
					AGCTCTTAGACCTTATG	EYALTAYGFPQMVVDVVK
					GGTCGCCATGGTTGTG	DMYQGASMRVKNATEKSD
					GACGGGTATGCTTGCC	RIPIMSGVKQGDPISPTLFNIC
					GGAGCCGAGTCGTGTT	LETVIRRHLESANGHQCLKT
					TCTTAGAACCAACCTCG	RIKVLAFADDMAILTDSPDQ
					ACGAGGCGAAAGCTTG	LQRELSKLDNDCTPLNLIFK
					CACAAGTTAGCACAAT	PAKCASLVIQKGVVRSASIK
					TGTGGTAGGGCCGACT	LKGNAIRCLDENTTYKYLG
					AGAAAATGAGTCCCTT	VQTGSAARISAMDLLEKVT
					AGGGGGTTACGCCTTG	KELECVVKSDLTPPQKLDCL
					GCGAAAGTGAGGACAA	KTFTLSKLTYMYGNSIPLITE

					TTGGCATTGACGGGTG	IKMFANIVIRGVKVMHRIPV
					CTTCGGCACTAGGCAA	RGSPLEYIHLPVKDGGLGVA
					AGGCGCCACCACACTG	CPKTTCMITFLVSTLKKLWS
					TCCAATCTCTAAAAAGT	DDEYIKTLFTSLAEEVVKKE
					TCACATTCATCGAAGA	SKKSTVTMDDIADYLNVEE
					ACTACCGGAACCAACC	RINRSEFGYNSITRLRDVMR
					ACACATGTGTTGAAAC	NLAITGDSPLYRLKMVVKN
					CTACACGGTGGAAGGG	GKIALLVQATSESMERIYTE
					AAAGGAAAGCTTCGCT	EDAKKLQRSLKDQVNKALK
					GGAACGAAAAGAACGG	HRFNTTKVVKSKVVRVVQQ
					ATAGGTTCCCCTTCTTG	HPASNRFVTKGGNLSLACH
					ATGGCTGTGAGGCTTA	RFVHKARLNLLACNYNNYD
					GGATGGACGGGAAGGC	KSKSKVCRRCGKDLETQWH
					CGTGAGGCCTCAGGCG	ILQNCPFGFSKKITERHDAVL
					GGTAACTCGGCCAGAC	HKVKTLIESGGKKNWTMKI
					GCTAGTTGATCTTCGGA	DEELPGFSRLRPDICLKSPDE
					TCACGACAGCCCTGGC	KQIILADVACPYEHGVEAME
					TAAGAGGAACCCTGGA	RSWQAKIDKYETGFAHLRK
					TGGAGTGTGAAGGATG	SGTKLTVLPIIIGSLGSWWKP
					GGCGGGTAGGGGGTTA	TGDSLKELGIKGSVINSAIPE
					AGCCTGTTGACAGACC	LCATVLEHSKNTYWNHIFG
					ACCGACTGCAGTCACA	EAYIPNPMRNGHAKPAGNG
					AAATCAGTGATTATGC	

					GGGTGGACCAATCTGT TGGCGGGTGTTTCCCTC TACCTGACCCCGCAAT ATGGTATGTACGATCCT CGGATCTAAAATTCAT AATGGCCCACCACAAC CATAAACCTCCCTAGC AGCTGGTGGTCCCGAT AATTCGGGTTCTTGCCA CTACTGCGACCCAGGC TCGCC (SEQ ID NO: 1274)	WKKERLQKAPVRPTN (SEQ ID NO: 1396)
CRE	Cnl1	.	Cryptococcus neoformans	CCCTCTTAATACCCCA TAACACATAACAACC CCCTAATCAACGTTCT CTGCACCTTAAACACC ACCAAC (SEQ ID NO: 1152)	TGAGGAAGAGGAGGTT GGATTATTTTTCTTTT CTTTAATAAGTTGTTTA TTTAAGTAGTTTCTTTC ATTCGGGCAACCCACA CGACAACCCAATAAAT TAAACAACGAAAAATG CAACCTCTATAACCC (SEQ ID NO: 1275)	MSLQRAKNARGDPGRCNLC SADYRDLKDHLNKQHSTHF FVPSDLRGSSLVACPRCGTP CSAGTGLSRHQSRYCGLTAP RIRRNRVGNSTNTSRCPPSN TAASPIVPSPSPERPSPPQPAE VVASLEPLSEAEEVLEVAQV DAETVDTLEGTRRAPESVPR SAEEGSTRVRELNMTAPEEE HRGEEESSHTNPTAPAGLEN AVSSTLGPSPGTLPSLLPSQE

						CANERFLYLAHLPVRSKPLP NNLVTDFMDAAERCALAYI AQPSDSTLLAFLALPKVGLT QALAPEQPLRPSTFLKQFPHI PWPEQPPARRPPSNIRPDTTK QVIKLVENGRLGAAERVLEE DASVAELDQGVIDQLITKHP KGPSCPFGNAVGPTPGKAPD IDTIQKALDSFKPDTAPGVS GWSVPLLKTAAKREPVKQF LQLLCAAIANNTAPGRSMLR TSRLIPLKKDDGSIRPIAVGE LIYRLCAKALIISHFQPDFLL PFQLGVKSIGGVEPIVRLTER VLEGSAGAEFSFLASLDASN AFNRVDRAEMAAAVKTHA PTLWRTCKWAYGDSSDLVC GDKILQSSQGVRQGDPFGPL FFSITLRPTLNALSQSLGPST QALAYLDDIYLFSNDSQVLS KTTQFLADKQHIIKLNEKKC KLISFDEIRQEGFKMLGTMV GGKEKRAEFLEGRIRKEMA

						KVGKLKDLPHQHALLLLRF CIQQNLRHLQRSLRSDDLVD LWERLDTMLWEEVKRMRM RQREDTAEEEALGRSLTKLP ARLGGLGLLSFKDVAPLAY RSAAEASDTLLDNLGLLSSP EEPPTPIPQRTRCAELWESQ QEAILHNLGDTERKRLTENA SRLGRSWLSVIPYLQPLRLS NVEIASGLHDRTLVGSSIPVC RFCGSDSPLGHDELCRARNP WTQRRHNAINRVIYQHLKQI QGATVEIEPHTLSGQRRNDL RVRGSSALAFTDYDLKVYS LGDRDARSTVTPCAPNGKL ADFCLDRCVNWLDKVGQV VSKNAPKVTGGVFKPIILST GGLMSRSTADEWKDWRDA MPVGGFEKMEKRIGVELVK ARARTLVL (SEQ ID NO: 1397)

CRE	CRE-12_CCri	.	Chondrus crispus	ACGCCCCCTATCCATTTCTGCCAGCCTCCCATCGGCTCGCCGTCTCCGCAACCCCTCTTCCTCGGCTGTACCAGTTCCGCTCCCACAACCTCCCTCGCCACA (SEQ ID NO: 1153)	TAAGTCCTTGACGCCTGCCCCGTGATACAGCATCGGTACCCCTAGCATTTGAATAAAAAA (SEQ ID NO: 1276)	MSQPNISSAETPLSQLPTPVPTPPSPSNPSLSLPTVRDLLLCPIRSSHVYSSIPSSCLHSFTMLLIKTVRAASATMTPTESHRAFIHLHILPIAVLRRSFRGETGWRSRTGQHHALRQRIRRASSGRHWAALWHEALAAHQVDLDYRTRHSRRYQASATSRHRIGRAMRLAADAQYGRAMSALKAKPLPDLHAAATRDTLTALHPPPASPVQPLSPTDLPPVPEITEGQVLRAARALNPTSAAGPDHLSPRILQLLARTTISPEAGVTGLSALTNLVRRLARGDIPDRTAPLLAAATLIPLQPRPHKIRPIAVGQALRRLVTKVLLPPAIQDTRDHLLPEQLANSVASGMDAIVHDTRMLMHRHGRNPDYIMVSVDARNAFNTFSRQSLLDRLPLQTPSLARFLNLIYGRTVPDLVLPSSPRFLMKSQEGTQQGDPASMLLFSLAIQPLLRRLTRECRL

						DLNRWYADDGTLVGPISEVI KALRILRDDGPQSGFHVNIN KCRAYWPTVMPEKLSELLRI FPLHVECGEGGVALLGAPL GTDAFVRRHLMNKVQSCHA SLSLLDEIPDARTRFHLHRVT GSVCKVEHVFRLTPPHLSLP AATKFDEQQIAAYSRLNDV AVSTSMATQIGLPFRLGGHG FTPLSPFIHASYAASLIEAAP VRVKGPHNPSESFYRRMAR RHIVHVLGALNPEVRTRGIL GTHSPLGPFEPEALLSRPERV HHTLIQAMQGATSRLYWEH TAWDLDPLPRNHSAASVRR RARYNSLRAPGAASFLCSHP SLTSRVPSAVWSCMLRRHL DTPVYCDSIRPLICSHCCKP MDARGDHAAICRHGFGVVH RHNTVRNLLARHAFRAAGL CCDLEVPSLLPNTANRPADI LVQPAPPPSGALPDRPTAYD VTVRSPYCRSTMSLAAKGL

						AGAAEAADLDKLRVHSRTV RDAFHLQPDSPLPLLDWHFV PLAFDTLGATSSRTMAVLEY LAHRIANRTYSSYGTAKIRL LQRISFAVWSSLASATLSRM PYHGAALSSPAQV (SEQ ID NO: 1398)
CRE	CRE-13_CCri	.	Chondrus crispus	CNCCAGCCAMCGATC CCGCCGCCACTCGCM GCCCGGCCGTCTCGAC CGCCACCTCCCCGASG CCCCAGCCCATC (SEQ ID NO: 1154)	TAATTCACCTTCATATC TGCTAGTGTCTCTGTAA GCGCACCCCTCATGCAT TGATAAAATTACCCCCC A (SEQ ID NO: 1277)	MAXXPXISPPGAPPAPLRYR MLQCPPPLPKXXXXPVPHP MSSPIRXRLPHRXMRGPPSX TPPRDMHRPHGTPGPHSHR XCGRPPXHCTHASXQPRXA XHXLQXPKLRSPPPHPHVSP LILCXGPLPTPMTQPRMKRA LSXSAKAPPTKRPSASQGPA ASSHDXPRTPPPXPPRPPPYR FPPPTLDQHXFALSXAYPHP XPRRPPSPXRXLRHSFPPRFG XQTFSSIPGPRLHSTVLLLIR LVRAATAANTPETTTLXSCT FTCSRLPFFERPSXAXLAGG PRAVNFMLSACXYGERVXD ESGXSYXXKXHCITSHPPRP

						RRYSRQHSRNHHTXFLHLH LLPTAXLREAFRGEXGWRSS RGQLHALRLXIRRACTGRE WGLLXXEALDAHSXRTEW QHTHARRPSPPVSPSARAAR AMRLASQAQYGRAMRTFT NPPLADLNDPATMERLQAL HPTPTVPVVPLPPSAQPRPPE VTXEAVXRAVRRLNPNSAA GPDRMSPKLLHLLAHTPISP EAGVTGLSALTNLVSRLAR GSLPPCTIPLASAATLLPLQP RPGKIRPIAIGQALRRLVTKX LLPAAIDDCRDHLAPEQXAN GIPNGIDAIVHDARMLVRRH GNDPHYXMVSIDASNAFNN FSRQQVLDQLPTRAPSLSRY LDMVYARAPSPLVLPSXPPT ILHSRXGSQQGDPASMLLFS LALQPLTRLISRECXLXMNR WYADDGTIIGRIDEVXKALD IITKEGPRFQFFLNPSKTRVF WPSRQXDLLSPLMTVGPLR

						VIDEGGVXLLGAPIGSPSXM AQYIREXLNTCKTALAHLD HIPEARMRFHLHRVSASACR LQHLFRLVPPDFAXPFAQQF DRDQLXAYXRFNSVTMSPRI VPKYGCXFXTXATASPHWH LPYTPXTLLASSIPLQHGYK VPTFPPSLSISVLHEARCGSF FEIYLHSXNPHTSRCDYVAK NRAQIRLXFSHGGHGLTSLA STIHASYAASLIDTAPARLQ GPHFPAVSQYQRFARGPLRV VLRNLPSFVQPAHFSMTEXD LGCLEPXALLARPERIHTFLL QAQYSAAASSYWQXPLWES FPNPGDHSAASLRKRVRYNS LLAPGATSFLTAHPAATSRV HNATWSTMLRRHLDAPVTN DSISPLRCXHCSKPMDARGD HAXIXSHGFGTLHRHNTVR NVLARQLFRVAGLAYSLEV PFLIPNTAARPADILVQPPPP APGLPPDXPTAYDVTICSPFR

						RGMLYHAARHRGGAADAA SVRKXKALERTIRXALLIED DNXPPPLDWHFQPLSFDALG APSQSTVHVIEDHAKLMAL RNSCTIATAKSRIQQRLSFAI WSSAAAAILSRLPTHAADIS YPIEV (SEQ ID NO: 1399)
CRE	CRE-1_ACas	.	Acanthamoeba castellanii	TAACCCTAACCCTCTC CCTCGGCCCCTCTACC CTAAAGCGCCCTAATC GACCGGCGACGCCCT AATCGCTACCCTCTAC GCCCTAATCGACTTTG GCGCCAAAGCGACTT TCCCCGGCCGATTTTC TTCCTGCCTTTTTCTTT TCTCTCCAAGCGACGC GCCTTTTACTTTGCCG CCGTTCTGTTTTTTCTT TTCTCTTTGCACTTCG CTTCTACTTCACACCT CCTCCTCCTCCTTCTC GACCCGCGCGGCCTC	TAAGCCGCGCGACGAG GACGGCCAGGACGACC AGGACGACGGCGACGG CGACCACCTAGCACCG CACGCGCCACGACATA TTGTCGCGCGCTGTACA GGCGGCTAGGTCGAGC CCAGCCGACCGTTCTG AGCCTCAGTCGGCTTG AGCCCCCGGCTTCCCA AGGCCTACCGGGGCGG CTCTTTTTCGCCCTGGT TTTTGCCGGCCTGTTTT TTCTCTCCCCCTTTTCC CCCCTTTTCCATTTGTA CTTAGTTTTTCCTTCGG	MATTTISRSPSSSSSSSSARS RASASTSASVASIPRLFRDG RFHCPLAHCQTRTSTWQDL SAHLTRMHDGDVPRDVAA ACGIVQCLHEGCRKWFRGA AGLASHRGKARHAPPPAPR AALAVAAVPRADSRGRTPA PTPSVAPPXAGPPPRAAPRA APSPLPCPPALPHPPPSASPP TSSVTSPCSPPTTPPSQPSPDL FSGFANAPTTPSPPSTPXSSP AGSPIPAARRFVLPVATPYP APAPRANRPKLSPVARPFVP KARAGAIPEASSPVTPQDRA VSRREDAAAAPSSAPGLGLA DEHEDDDTYGGDTIALTAP

				GAGCGACTTGCTGCA	CCGCGGCAGCTTGTTGC	HAPRETRAPFEFEACFLEEE
				GCGGCTCCCGGCCTCC	CCGGCATAGTGTTAAT	APATAGDLPPYARAFLACPS
				CCCACGCGGCCTGCTA	ATGTTTAAAAAACGTG	ARLQEIPRRLKSAWQAAAK
				CTCCCGCTTTCTAGAC	TAAATAAATAACTGTTT	TIAEAALDCHTAGDTQGYN
				GCCCCCGGTCTTGCTC	AACCCTAACCCTAACC	AHLRLFIELPARGLAVPTNC
				TCAGTCTCCCGCATCG	CTAA (SEQ ID NO: 1278)	RGAARTKLQRERLLDIAAG
				AAGCGGTAGTCGGGG		RIPAIPDPPCDAPGADDALR
				TACGTGCTCAAGTGAC		GFPVSGTTAGDVSNDDDSG
				TCAAGCCTCTTTTCAG		GVHDRPAATASARQAKRLV
				CCTCGGCGCTCTCTCA		EQGLSSRALRALERGEPAVA
				ATCCGCCTCAGTCTTA		SADTLGRLEALHPPNPTDRG
				GCCTTTCAAGTTGCTC		LWPGAPKAAIPRVTAKHLA
				GATTACGCTCTCGAAT		QVAKELPRGSAPGPSGWTFE
				CGCTCTCTCTCTCAGT		LVQAAIDRQPTGTVAAFLID
				CTCAGTCTCAGTCTCA		MAQRALRGTLHWRGLLTAS
				ATCTCGATCTTGCCTT		RLVALKKPDGGVRPIAVGE
				CGCCTTCGTCTCGACG		ALYRVIGRLVLKADRVMSS
				CCTTGCTCTCGWAATC		ADATQYVGRHQYGVAYPG
				GCTGCCACTACGTGCC		GVEAPVHAVRELHDSGQLR
				AGCTTTTTCGTGCCTT		AVVSLDWRNAFNSLDRVHT
				GTCTTCGTGTCGACCC		ALLIADRAPALARLYEWSY
				GGACCGTTTGCAAGC		REDSVLVLPRAFEKAGLPAS
				CCTCGCCTTCGTACCC		LLSQAGVRQGDVLGPLFFAI

				CGCTCTCGTAGCCGTT CTCATCGCTGAAGCGT TCTACGCGCTGGCAGC AAGCCTCGGCCCTAG CTTGTAGCGCCGCCGG TGGCCGCTCGCCAAC (SEQ ID NO: 1155)		GAAPVLDEIDAIPYVTPRAY LDDIFVTIPHGVTDAATKAA VAATFATAEREGAAAGLRL NRCKSAVWAADAEALLPPH AAGAREDVESCAPVREGLKI LGAPVGSPAFVAKSLDGIIK RAIGTLDLVADAELPLQHKL VLLRQCVAQIPTFWARAVP DAGPALAVWDTALLRRTGA LVGLDVRDGSLQADIARLPV RLGGLGLRSMKDTAPRAFV ASILFAAALANTRRSELTCS ASTARRLRAALPELARTDAC NDEAAWRRSIARGVFPDVD KLGTTQLQRVLQGMADSKS AHRTRRQVPFLFAAVFEDA ATPGSGAWLAAIPSDPTLVL PDAELAEAVRIKLLTTTANA AGVCPACHKTGIDPSHAYTC VSLSHLRTARHDVVVRRVE LACKTEKPVREHVLAIPPVA PTDNNNNGDEDGSPVTTAD DNADGHAVATKRRPETRAS

						ARAAAAAATAAAAAAIIND NSLLSDDDDDDDHDDNCHG EERGEGERNVTCPGHYTATP FAADDTLDNSDEDNEDNAH EDDDEDGKDDNDDDVYNN CNSSSSDGDEGGDDLDYEY SDQSVTRSVDAATGESPNPE RPTTPTRALLRADLWLPATS TAVDVMVAAACRRSRAKA FDRAVSRKAAKYGPAVADG SIAKVVPFVVSPFGVLSRPA KAFLKRAMGDTTAAKQAK ARLRLAVAAVRGTARLSYA WGACAALIVGGN (SEQ ID NO: 1400)
CRE	Cre-1_FCy	.	Fragilariopsis cylindrus	ATCAATCTAATACTGA AGGCAATACCAAACT CAACCCGAAATCAAA ATCGTTAGAATCAATA TACGACCCCCGCTGCT GTACATGTCCAGCCG GATCTCGTTGTAAAGA AGATTGCAGCTGTAA	TAGCACCACCATCTATT CATATCCACACACTGA CCACCTCCACCTTCACA ACTCCACTCTCAATTCC CCTGACTACTAAGAAA TATTTCATGGTGGTTAC ATTGGAGGTATCCCAA CACCAAGCACAAAATG	MAPLPWNAATSSPPSPVPLT NDKKKDSTLPTATSKNLSK NNNNKNNNTNRINNIKNND NTNDGSNKINLKLPPAAVKI TNPYKNKKKNKKKNNAGK SNPKTNQNPNSSPLSDNDDD DTDSSNITINRRLKFGTDDL APPNPPSNTNTIGTATAATA

				AAAAGTTGGACTTCTT TGTTCTTCCTGTGAAG AAGTTGATTGTGGCTG TTCAAATTCTTTTCAT AATAAAGAATTA (SEQ ID NO: 1156)	AACCCACTAACCCTCTC ATCCTATCCACGGGGA CCACCTTTGAGCAGAA CACCATTCATATTACAA CCTTTAGCTAGATTAAG ATAATTATTTAGTACAT ATTTTATACTATTAAAA AAAAAAAAAAAA (SEQ ID NO: 1279)	ATATTTATAATATTATNTTT TTTTTNNTTGDNLASNINNN NNNNNSGSNNSNTNNINNT DGNGSNNRPPPRVYTVDPRS DLPGAEISAANKMLDEVYG DHVHDNPGSHLSGLISSSQD QLWQGYFRRLIPHNQSLYD CPKGKLGKDITNEYSNLFEA IMNGKCNMEKLLVFPVVVL QRRHGVTKNADVKRRLLSR LTAWKEGKFKYLVEDTHRD LIAKQSKARGDTTPAHRAK VYSSKLMRGHLQSAVNYIT DREGGGILYPYDVDEKSGH TVSRVLQDKHPSMRDPGPT AMPAYESVPELPTLEITADT VEIVAGKLSGGAGLSGVDSI QLKHLLLHHGQASQRLRNV CAKFGRWLANEHPPWASYR AMLANRLIALDKMPGIRPV GIGDTWRRFFAKLVLAVSM SYATDCCGSDQLCAGLRAG VDGAIHGLSAMWREMESEE

						NTGFVLIDADNAFNEVSRIN MLWTIRHEWPAGARFAFNC YRHHSLLVVRNPGGKPFTFF SKEGVTQGDPFAMIAYGVA LLPLIRKLKELNVLLVQSWY ADDASAAGKFDEILRLFQDL LRMGPDFGYFPNASKSILITH PDNVVAAHHFFNETHGLGF KISTGSRFLGGFIGDTTSRDE YVSTKIADWIHGTKELAAV ARLKYPHAAYTGITKCLQH KWSFTQRVIPGIDDLFQPLE DELTNNLLPALFGDPPSTMD DKLRLLTALPVKHAGLALP NPVTSSATNYKNSTLMSSHL LLAVQGKINFSLQDHRDTCQ SSLSASRELRQTENDSSLTNL LAALPPAAAGQPSTTRAIKR AGETGLWLTTIPNHINGNIL GCDEFIDAIRLRYQKVPHNL PAKCDGCGSAFDVGHALQC KSGGLIIRRHDELNLELASLA KMALRESAIRAEPEINPSASI

						MDSPTTITAIDTNGDRGDLLI KGFWDNGMDAIIDVRITDT DAKSYRTRDPKKVLQSQEK EKKKKYLDQCLLQRRAFTP FVVSVDGLIGYEASNVLKQL SKRLADKWNKPYSVTCGIV RSRISIACARASNQCLRGSRI PFKTMSRQIQWEDGAGAGL YRIVR (SEQ ID NO: 1401)
CRE	Cre-1_HM	.	Hydra vulgaris	TTTCTAATGTTACGTG ATATGATATGGTTAGT TCATGGTTAGTTTATG TTTATGCTTAGTTTAT GGAAAATCGTTTATTT ATGGCACAATATTGTT TGCTGTTTTTAAATTT ATGTAACGTGTGCATT TGATGTATATTCTTGA ACTTTTTAATCTGAAT TTTTACTTGGTTTAAT ACGTTTATTATATTCT TCGATTGAGCAATTTA TCCTATCAAAGCAATT	TAACTTGTATTTTTAAA TTGTTTTATTAGTTT (SEQ ID NO: 1280)	MNMVSICKRCDRSFTTLKG LNIHKGQCKIFVSNTNKQIN NVVNNELTTPNKNKVEINTI LNCDEISVEHYSTNTPYLPKI NICESIIDPNDYLWGHMPFSF LLNHVNTIYDEIVFYHKNLF KVPSGKGGKMFIEELTFWL KQFNNRTKLNGIAMKCFMI VPSLMLQKPSIRSKAKEHAE CLVRRITLWRNGNFSELMRE IRYIQSKINTSKKKRTFEDISR IFAKLMMEGKVAAALKVLD RESSGILQCSESVLKELKSKH PDETPVQDNCLLYGPLQNTP

				TATCCTTCGATTCGAG		ECLFDSIDEISIFNSALQTKGS
				CAATTTATCCTTCGAT		AGPSGMDADLYRRVLCSKC
				TCGAGCAATTTATCCT		FGPSCKTLREEIATFTKNIAT
				TCGATTGAGCAATTTA		KSYQPDIVQPYIACRLIPLDK
				TCCTATCAAAATTAGC		NPGIRPIGIGEVLRRIVGKTIS
				ATATATACTGCAATTT		HHCQKEIKEAAGPLQTCAG
				TCAAATAATCTACGA		HGAGAEAAIHAMQKIFHQE
				AATAAGTTCACTTACT		DTDGVLLIDARNAFNCLNRS
				GAAAATCATTAAGTA		VALHNIQITCPILAMYLVNT
				AAAGAAGAAAGGAAG		YRKPAKLFIYGGETIFSKEGT
				AAAAAATAAAAATAA		TQGDPLAMPWYSLSTVTIIN
				AAAGTAGTAAATCCTT		TLKLVIPDVKQVWLADDAT
				TCATAACAATAATCAT		AAGKLQSLKKWYKCLEDV
				TCTATTATTAAATTTA		GGLYGYYVNQSKCWLIVKS
				AAGGAATATTTTGGTT		DNQAEEAKLIFGNSINITTQG
				TTGTACTAAATCATGC		KRHLGAALGSEAYKKVYCE
				GTTCATATTTCACCGA		DLVSKWSKELNNLCEIATTQ
				AGAAGGGGGCTGCTA		PQAAYSAFIKGYRSKFTYFL
				TATTTTTGTTTGAAGT		RTIEAFENFVTPVEKILSEKL
				TGTTTATCTTAAAACT		LPVLFGTDCSIIKENRDLLAL
				TTAAACTTGTGTTCAA		NPSEGGLGICNLITEAKEQH
				CCAACCGTAAACATT		TASKKITNLHIKSILDQSDV
				AGTTCGCTGTTCGCTC		MKEKDDFGKTFSEIKTKTN

				AAATTATCTACAATAT AAAATTTATCAATCTT TTTTCGTTACGGTAAA CAATAAACAATAAAA TAACTATAGTTATTTT ATTGTTTACCGCATAT TGTTTAACTATAGTTA AACAAAGTATTTGTTT ATGGAACATTACCAG TATCTCTTGTTAAGGT AAACAACAAAACATA GACGGCATCTCTTTTT AAGGTAATTAAGTAT ACGGCTAATAATAAA AATATACAGCTAATA ATAAAATCTTCA (SEQ ID NO: 1157)		MDKSKKKKEEVKKIHAGLP ENLKLLVEQACDKGASSWL NTLPIKEQHLDLNKEEFKDA LRLRYNVPLANLPSYCACGE KFDELHAMSCKKGGFVCNR HDNIRDLLTVCLNKVCTDV QAEPHLIPLTNEKFNFKTAN TNDEARLDIKAKGFWRKGE TAFFDVRVTHVNSKSSKKQP TKHIFRRHEDAKKREYLERV LEVEHGTFTPLIFGTNGGFG DECKRFTALLAQKLSLKMG ERYGAVINWLRTRLSMEITR ASLLCLRGSRTPFRHYNTDD VGLENVQCGLI (SEQ ID NO: 1402)
CRE	CRE-1_LSa	.	Lactuca sativa	ACATTAAATTAGAGA GGTTGATGTTTCAATG GAAGAAGATGAAATT CCAAGAAGCTATTTTT GTTGCCCACCAAGTGT TTGATAAAATGTCCAA	TGAACTATATTTTATAT ATTAAAAAAA (SEQ ID NO: 1281)	MASSSTSSSDICLCPFRSFHC CPNGEVGSKGIXRMISHIKR HHLLTEDRKCVLREALSSDV GLFMAVEETLKAFGQWMC GKCMTLHALSRYCHHPDGR VXFVTGADGSSRYIVGILKP

				ACTAATTTTTCTCTTG TTGCAGCTTTATTGTT CAAGATAATGTAGTTT GCTTAGTTTGAGCGTT CCTTGTGCACACCAAC AGTGTGTTGGTGTGCC ATTTCCTTTCCTTCCTT TTTAACTATTGCTTCA TAGCTTAAGCTTCATC TCGAGGCTTGTTCTCT TGT (SEQ ID NO: 1158)		STKESVTNALGGLVFDVGLL DRVFKEPITTVKSIPHSCRLA FSQALKTALYKVIAQPGSVD AWICLLLLPRCTLQVFRPKN RQECRSGNRKSLQQSSILKS LDTWGKEDGIRKLVQNMLD NPEVGAMGQGGGILQKEST SSNTNIRQCLRKVADGHFTA AVKVLCSSGVAPYNGDTIK ALEDKHPFRPPPSMPSPIISEP PLVADFDCVFGCIKSFPKGT SCGRDGLRAQHXLDALCGE GSAIATDLIRAITSVVNLWL AGRCPTILAEFVASAPLTPLI KPDNGIRPIAVGTIWRRLVS KVAMKGVGKEMAKYLNDF QFGVGVSGGAEVVLHSANR VLSEHHADGSLAMLTVDFS NAFNLVDRSALLHEVKRMC PSISLWVNFLYGQAARLYIG DQHIWSATGVQQGDPLGPL LFALVLHPLVHKIRDNCKLL LHAWYLDDGTVIGDSEEVA

						RVLNIIRVNGPGLGLELNIKK TEIFWPSCDGRKLRADLFPT DIGRPSLGVKLLGGAVSRDA GFISGLAMKRAVNAVDLMG LLPQLCDPQSELLLLRSCMG IAKLFFGLRTCQPVHIEEAAL FFDKGLRRSIEDMVVCGGPF FGDIQWRLASLPIRFGGLGL YSAYEVSSYAFVASRAQSW ALQDHILRDSGICGMDSDYL CAMTRLRDTIPGFDCSGFTN KDTPPKSQKALACALFSKIV KDMEVDFDMTVRQKAVFE CLRAPHAQDFLLTIPIDGLG QHMSPVEYRTILRYRLMIPL FPIDEICPVCRKACLDTFGEH AVHCRELPGFKYRHDVVRD VLFDACRRAGISAKKEAPVN FLTDPQDGRSTLRPADILVF GWVGGKHACVDLTGVSPL VGLRSGGFTAGHAALKAAA CKVAKHENACIENQHVFVP FAFDTFGFLAPEAVELLNRV

						QRVMHSNVISPRSTDVVFKR ISFAIQKGLAAQLVARLPSID MY (SEQ ID NO: 1403)
CRE	Cre-1_MB	.	Monosiga brevicollis	CATCTTGGCGTGAACC ACGTTGTCAGACAAA ATCTGCAACCCCGCTC TTTGCGGCCCGCGTTT TGGCGGCGCCCTCGCT CCCACCGTGTCCGCTC GCTTGCTCGCTTGCTT GCCCCGCGGAC (SEQ ID NO: 1159)	TAGGTAGGCACCGTCT CGGGGGTCCCTCTGTG GGGATCCCTGTGTGCA CCTGTCGCTCCCTAGGT GGTTCCTCGTTGTGTCT TTTGATGGCTTGACTTG TATTTTTGTTTTAATTTT GCTTTAATTTTTGCTGT ATTTGTGTGGTATTTTT GCTGAATTTTTGTAAGG TCCTTTGTATGATGTCT TTGTCTTCTGTGGTCGG TTGTTTTCCTCAATCCG ACGTTGTGTCTCGTTGG ATGTGAGCGTGCCGTG GTGTTCTTTGTGTTTGT GCTGTGATGGCTTGTAG TTGTGATGTGTGACTGC CTTTTTGGGTGTCTTGT GTTTGAAATGGCCGTAT	MATESGGEDSWTQVRGAK RPSAESPPSNTTTSPSQTHRS AKHTKHGSARHDRNHVFPD PMTTPLRPHARHSVPTARAS SHVPSTSPAAGATESSARAV VPAAEPVTRTSNGGGEQHPI IGNTSNASPRTPRTPSSPRSF AQVAAAMPAAATATSSAPM TEDLSASVPSEPNGSGEQQP SPESTGQTHHSIPNTPSDFLT MSSDESDSPPRSTALRAPTPI APPAHDGDGDTNGSATPEPL VQSPTPAQMVLPYPSGTQQT HSDPSPPSASPPATTILPAAIS HPVEHSEHANSAPLGEVSES ETHNTAGEHSESEQDVLLSD PAPPIAANVLDAQRKVLLKT SGHRQLLACPFGLCKCKGPR LDRKAWVNHVLREHPYDE QATDLVKQVMEAKLVAQC

					CTCTGGTTATACTTGGT CGTTTTGTACGATTTTT GTTTCTATGTGCGTGAT TCTTCGCGCTTGTACTT CTTGGCATGATAGAAG CCAATGAATGTGTCTTG TTCTCTTGTGTTGTTTT GCGTGCCGTCGTGATTT TGATGTCGGGGTTGCA CAGCTTTGCTTTCAGCT CTGAGGTTCAAACACC TAATTTA (SEQ ID NO: 1282)	NKCHLFFEAAGISQHRSRCG ANLKRATEALFHAAGHDLL EIMRGAWPQQCVGSRISVCE LLKLARHPLMQRSRYPSNA TETKLMAATLSQLYWSAVH SDYTAEEREMCWALILALPS MLLSAPSTALSTIDLRNMFH DRLRWLVTGQLGRVVDAM RKAVARKQSRRGQLNAGA GAHPNDAVDQSLRSLVRDP DLADEAWANHVTNRLNRG QIAKAFDADKARAVIGNSEV QAVRDLLVPPGLTPYIASTP ASTSTLAPATAVSSPTVSFTK GELPKALAATKGVTDPYGW SGELLASIYRIKEHFSQVLGP RQGSTSDPTAPSDGDAPQGP TTATGGPQVALNKIFHHIAN NTVPESIRHALCSINYTILEK ANGKFRPVGTDSIFNKVVNR ALLEQQQPHIAHLLQASPEL AVGVKDGISAAVGMAFGEL QACESTPGWTMLSLDFKSA

						FNYTDRARLHEIVADKVPGL LRAFERHYARPTTHCIVDKF FKVIDIDVGQGIVQGNELSPF FFALYSCEVLGLLDATTDYR CKVIKYLDDIVLMGPAEDV AADVEIVKARAESAGLHLQ PSKSRFYMPRHHSASITAIKS VLPDAVRETANTGMTVLGT PIGRREWMKKQLNDKAKHI AGKLNDMLTTGVSLQALLT AMQYVPSLINHLYTLPPSLT SGLSELLNRACKDTFVKAFF AKVNLSAPAGAEGHDVTLE QLLEARLFTRANTGGFGLH DLVERGPVAYVCNMAKLAT RYPRVYDRLLEDASRAADF EAHVQRAGFQMATVKDAA TQRPAEIIALRSKAALDDLM AKCALDLQQAYLASREWG VSTVLTMRGRDKLRRLSDT TFAIAVVSMMGFGLHELINV KPTDKCPLCSSKTPQPRLTR EHLLTCRPIKRHNALRDEMG

						RLLRYATLSHVWVEKSGYN ANGQSCRIDLHCRNPFPGGA LGPALPDLGIDVTVRTAQPP TTSQACIKVGAALRRAEKEK RDYYTGFNHGKTLIVPAAM TTTGGFASSFVDLLGQLARC AEARGVYQPGLDEAFVPRW KGRFAALVHQMNADHIQRH FGGVCLRSS (SEQ ID NO: 1404)
CRE	CRE- 2_HMa	.	Hydra vulgaris	AATTTAAAAAAAAAA AATCGTTTATTTATGG CATAATACTGTTTGTA ATTTTTGAAAATTCGT GCAACAACTGCAGTT AAATTGAAGAGCTGA AATTTAAGATCTGAGC TTTTCAATCAGAGTTT TTTACCCTAAAACATT AAATTTTATCATAACA AAAATCGTTCTAATAT TATTAAACTTAAAGA AATTCGTTCTTATATC	TGAGCTCTTATAAATTT ATATTATAGCATTTTGT TTTA (SEQ ID NO: 1283)	MSSCKVTIPHVCPYCKVELK TICGINRHILKCKKNPLQIPS LQKTNTSLTLEPNTKVIPSIT KQNDIIIASTSSNNLAFNQKK DYTLTPTYSRKTTPVSILSSM KMTPISITSHIVRRKLPELPS QTTNHLFNENFINVPFLPEIM NHLPVPNNNVMWGVYSYQ QFKLFVDSTYDEIVNYRRNI FNIPSGKAGKEFIEELTFWLR KFNSTSSLNSIALKVTMILPN LLLQKPSAKSKSKEHTLCLT RRIDLWKKGDTSLLLKEVR

				AAATCTTATTTCAGTG TTTCACAGACGAAGG GTTTTACTAGATTTTT ATTTTTTCAACTTTTG AATTTGTTTATTATAA AACTGTAAACTAGTGT GCAACCAACCGTAAA AAWTAGTTAGCTGTT CACCCAAAATATTATT CAGTATGAAAATATTT AATCTTCTTTATTCGC AGTAAACAATAAAAT ATCTAGTTAAACAAA ATATTTCTTAATAATA AAAAACAAAAACTTT TTCTTAACAAGTACA (SEQ ID NO: 1160)		NIQKKFVNSKXKRSMDDISR IFAKLIMEGKITAALKFLEKE ASSGILPLSDNTLKDLKSKH PEPSRVEDYSLLFGPIDLIPK CFFDCIDEQLVMKAAFATK GSAGPSGMDADIYRRILCSK NFIKEGKELRKEIAKMTQNL LTETYEPTFLEAFTACRLIPL DKNPGIRPIGVGEVLRRIIGK VISWSFNSEIKEAAGPLQTC AGHGAGAEAAVHAMKEIFD NVQTDAILLIDAKNAFNCM NRQVALHNIQIICPLISIYLIN TYRNPSRLFVAGGKEISSQE GTTQGDPLAMPWYSCNTTII IEHLLVNYPQVKQVWLADD AAASGSIANLHSWYQHLIDE GCKHGYYVNQSKCWLIVKS PSLAENAGIVFGKSVNITTEG QRHLGSVIGSQNFKNKYCTE KVAKWLTELKQLCKVAETQ PQAAFIAFTKGFRSKFTYFLR TIPKFEQYLAPVDEILSHLLL

						PTLFGKDTPFEDHIRKLFTLT PRDGGLGIPILVEEAPHQFLS SVKLTKNLVQQIIDQDKILK TKNSSGNVLEDLEKILTTDR LKHRKEKIIAVDSMQPDSML RNIQQTRSECASTWLNALPL ENQGFVLNKEEFRDALCLR YNFDLKNIPRICECGEPFNVT HALSCKKGGFISSRHDNIRN LFTTLLKRVCINVQSEPHLIP LDNENFYFHTANKSNQARL DIKANGFWRNGQTAFFDVR VTHVNSMSNKNLDIAAIFRK HEKEKKREYGERVREVEHG SLTPLVFGTNGGMGKECHR FVRRLAEKLAEKQNEKYSV VMTWLRTKLSFEILRSTILCL RGSRTPWTKKNDFEIGVDFK MDALEARI (SEQ ID NO: 1405)
CRE	MoTeR 1	JQ74748 7	Magnaporthe oryzae	CCCGAACCCGAACCC AAACCCAAACCCAAA CCCAAACCCAAACCC	TAATAGGTAACGTCCCT ATTTTTGTCTTTGGTTTT GTTTTTATCTTTGTTTTT	MVCPTCNGVYADYNDHIRK KHPDERYTALQLQPLGLTPC PICKTACKNDLGVKTHLSKI

				AAACCCAAACCCAAA	GTTTTTGTTTTCGTTTTT	HKISGASKISTQPRIRTENTD
				CCCGGAGGGTTCCCA	GTTTTTGTTTTCGTTTTT	NTNSVPTSSFNPVLPEIQTLT
				AGTCGCCTAAACCCG	GTTTTTTTTTTTGTTTTT	PGLNNSRWADNPRKRRADT
				AAGGGTTTAGGATATT	GTTTTTGTTTTTGCCTTT	PSPTRGRNTRPRRFSYTDIDL
				ATTTCGTTTATTAGAA	GTTTTTGTTTTTATCTTT	TNDEPADNPRANNPRVNNP
				TTGGATAATTATTTAC	ATTTTTGTTTTTGTTTTT	RVNNEPPSSPNSLPSISEFHT
				CCCTGTTGGACAGGG	ACTTTGTTTTATTTGTTT	PGTLPLTNSNISLKDQHDKIT
				GGGTTGCAGGGGTTA	TATATTTACCTTTTGAT	GPILQKPLIQKLIEYSKIPIPE
				AATTAAGGTTTTTTAT	TTTTTCTATTTTTCCCAC	HHLHARQAKIFADAANRIA
				TATTTATGCGCCGTTT	CCTTATTATTATAACCC	KNFIQSPTEKTLFNLLILPRIF
				ATTTGTTTACCCCCCC	CAACCTACTAATATTTT	GIGLINGKVTKIMQNFPSQIP
				AAATATTATAAAAGC	TTCTTTTTTCTTTTTTCT	PIPKIDFPSEKTDSDPVLNAK
				GCGTTCCATCCTCTTA	TTTTACGGTTTTATTTT	KLLEKGYIGRAAKAIIDPTPV
				GGAAAAGCGAAGCTT	CCCGTTTGTTTTTTCTA	APETPESLNILREKHPIGQNN
				TTCCTTGTAAAAGTCG	TTTTATTTGTACGACAA	PFNTKSQPISGRQITEKAILL
				CTAGACTTTTACTATA	AACCCTTAGCAAATAA	AISSIGREKAPGLSGWTRSLL
				AAAGTCGCTAGACTTT	GCTTAGAATATAATAA	DAAIKIPTQNDVIPALRLLTD
				TATACCAATCTTTTAA	AGCGCGAATTAAAA	MIRQGTAPGRELLCASRLIG
				CAAAAAGCGTAGCTT	(SEQ ID NO: 1284)	LSKPDGGVRPIAVGDLLYKI
				TTTGTTGCCAATCTAT		AFKAILNTLWSPNCLLPYQL
				TAAAAAAAGCGGAGC		GVNSIGGVEPAIFTLEEAIMG
				TTTTTTTAACTTTTTCT		PNINGIKSITSLDLKNAFNSV
				TTTTTTTTTTTTTTTCT		SRAAIASSVAKYAPTFYRST

				TTTTTTTTTTTTTTTT		CWAYNQPSILITENGSVLAS
				CTTTTTTTTTTTTTTT		AQGIRQGDPLGPLLFSLAFR
				TTTTTTATATATATTA		PTLETIQKSLPYTYIAAYLDD
				TTATTATTATTATTAG		VYILSKTPVKDKIAKIIEKSPF
				CGGTGGGGCTATTTAT		TLNSAKTTETDIDTLKTNGL
				GCGCTTTAATTTGTGC		KTLGSFIGPTELRKEFLQNKI
				GGGGCTATTTATGCGC		QNFESSINALKKLPKQYGLLI
				TTTAATTTGTGCGGGG		LRKSTQLLLRHLLRTLNSQD
				CTATTAATGCGCTTTA		LWELWEKTDKLIADFVINLT
				ACTTTACAAATTTTAT		VTKRKKRPITDFVTPLITLPI
				TTATGCGCTTTAATTG		KDGGFGLLRHNGIAQDIYFA
				CTGCGGGCCTGTTAAT		AKDLTTEIRHKIQRISNDFPQ
				GCGCTTTAATTTACAA		NQSPTATEILHLLHNGVLAD
				ATTTCATTAATGCGCT		CKNGLTNAQLNALTENASY
				TTAACTTTTATATTTA		LGRKWLNILPIQKSNRLTDW
				CTAATGCGTTATTTAT		EMAEAVRLRLLAPVKPLTH
				ATAATTGCTATTATTA		PCNHCGNRTNINHEDVCKG
				TCGTTGCTATTATTAT		AVRKYTARHDQINRSFVNSL
				TATTGCTATTATTATC		KSRPEIDVEIEPDLNNENNV
				GTTATTATTATTGCAA		NNANTTTENPTPSPNGQNDT
				TTTTATTATATAAACC		GCLFTTPIRSGTRNGQNGLR
				CTCGTTTGTCCCTCGA		ADFAVINGVSKYYYDVQIV
				TTTATCCCGTTTCTTTT		AINKDSGNTNPLNTLADAA

				CCATCCCATCGCGCGT TTTCGTAAGCTTTGGT TTTCGTAGGATTTGCT TTCGTAGGCTTTGCTT TCGTAGGCTTTCGTCA GCTTTTACCTGCTTTT ATTTTTTCTTTTTCTTT TTATTCCCCCCCCTTT TTTTTACCTGGTTTAT TAGCGGTTTACCTGCT TTTATTACCTGGTTCC CCTTTACCTGTTTTAT TAGCGGTTTACCTGCT TTTATTACCTGGTTCC CCTTTACCTACTTTAT AAGCGGTTTACCTGCT TTTATTACCTGGTTCC CCTTTACCTGTTTTAT TAGCGGTTTACCTGCT TTTATTACCTGGTTCC CCTTTACCTGTTTTAT TAGCGGTTTACCAGCT TTTATTACCTGGTTCC		NNKRRKYQFLDPFFHPIIISA GGLMEKDTAQAYKQIQKLI GPVAAHWLDTSISLILLRSRT TAAISIAKNRPRA (SEQ ID NO: 1406)

				CCTTTACCTACTTTAT TAGCGGTTTACCCGTT TCTATTAGTGGGCATT TATTTCCCGTTTTTATT AGCAGTTAAATTTACC CTTTTAAGGTTATTTA CCTGCTTTTATTCACA GGGCACCCCTGTTTTT ACTAGCAGTTAAATTT ACCTTTTTAAGGTTAT TTACCTGCTTTTATTC ACAGGGCACCCCTGTT TTTACCAGCAGTTAAA TTTACCTTTTTAAGGT TATTTACCTGCTTTTA TTAACAACCCTTTATT TTTTCCTATTAACGGG TATTTATTTACCTGTT TTATTGGAATTCACCC GTTGGACGGC (SEQ ID NO: 1161)		
HERO	HERO-2_BF	.	Branchiostoma floridae	TTTTCAGTCTGGCTCA GCCAGTGACCGCCGG	TGATTAAAGACCCGAA ACACCCAATGACCCCG	MNAVCVCGKVCKNQRGLRI HQTKMACLRRVQAEHRSGA

				GAAAGTCCGGCTGAC TACCACGAATAGGGT GGTGACAGCTGGATA GACAGACGACAGCTC GGAAAGACGGCATTG GGGCAGTATGGGTTG GCACCCCTAACTGCAT CTCCCCTAGGAGAGC ATCCCGCAACACGCT ACAAAGAACCACAAA GAGCAATACCCCCAG GGATGCCCGAGAGGG GGGGAGGATGAGCAT CCCATTCGGACGGTCC AATCGGTATTGACCCC AGCAAACGGAGAATC GACA (SEQ ID NO: 1162)	GGTTCATCACTGATGAT GTGTCCCTGTTCGCACT ACCAGAGTGTATTCTA GAG (SEQ ID NO: 1285)	VATTVEPVLSASAPGQTEED QGPEAPHSARNLRATPAPPQ GRKSDHHRVKWPAANSKE WSQFDEDVDMILESVSRGST DQKLQSMCTVIMSMGAERF GTIGQRKPTDTMKPNRREV KIRQLRQELKSLRRSFKASTS GEERAALAELTHHLREKLRT LRRAEWHKKKGKERARKRS AFITNPFGFTKRLLGQKRSG NLTCPVEEINLHLSNTFSDAS RDVDLGPCPLLVTSPEPEVH FDISEPTLKEVRETVKAARSS SAPGPSGVVYKVYKHCPRL VVRLWRILKVVWRRGKVA ADWRQAEGVWIPKEEESSK VDQFRLISLLSVEGKIFFKIV AQRLIKYLLDNQYIDTSVQK GGVPGVPGCLEHTGVVTQLI REAKENRGDLAVLWLDLAN AYGSIPHKLVETALTRHHVP ESIQNLILDYYSNFWLRAGS STATSAWQRLEKGIITGCTIS

						VPLFALAMNMIVKGAEAGC RGPVSRSGTRQPPIRAFMDD LTVMTATVPVCRWLLQGLE RLITWARMSFKPAKSRSLVL KKGKVAERFRFTLGGTQIPT VSEKPVKSLGKVFNSSLKDT ASVQQTRSDLTTWLEGIDKT GLPGSFKAWMFQHGVLPRV LWPLLVYEVPMTMVEQLER TISRFLRKWLGLPRSLSNIAL YGRSTKLQLPLSGLTEEFKV TRAREVLMYRDSSDSKVSS AGIHVRTGRKWKAQEAVD QAEARLRHSVLVGSVAVGR AGLGSCPKPRYDKVSGKEK RLLIQDEIRAGEEEDRRCRM VGMRKQGAWTRWEHADSR KVTWPELCRAEPSRIKFLISS VYDVLPSPANLHVWGLAET PSCQLCQRRGTLEHILSCCP KALGEGRYRWRHDQVLRV LADTVSNAIQSSRSQQPPKK SIVFVRAGEKTRQQPTSAGG

						LLSTARDWQLLVDLGRQLK FPEHIVATSLRPDMVLVSES TRQVVLLELTVPWEERISEA NERKRAKYAELVVQSQSNG WRARCVPVEVGCRGFAGQS LAYVLKLLGVRGFRLRKSIR DILEAAEKASRWLWFRRGE PWKPHGHRSGNDQPRLGRP GEGVW (SEQ ID NO: 1407)
HERO	HERO-2_DR	.	Danio rerio	TTCAAGCCTGGCGCA GCCAGTGACTCCTAG GAATAGACTAGGTGG CAACCAAGAATAGTT TGGTCGACTACTGGA GAGACAGTTGACGGC ACGGAAAGACGGCAC TTGGGACAGTATGGG TTAGCACCCCAGCCTG TGTCTTTCGTGAGAGA GAACCCAAACAAGCT ACGGAAAGCCCCACA GAGATATACCCCCAG GAGATCCCGAGAGGG	TGATCAACCCCGGCTG GGTCACCTGGGTGAGA GTGTATGATGTTGAGA GACCCGAAACACTCAA TGATCCCAGGATACAT CACTGATGATGTGTCCC AAATGCATCCATGAGA TGTTTCTTGCATAA (SEQ ID NO: 1286)	MTHANEQTTNKIYVTCICGK LCKNHWGLKIHQARMKCLE QESKVQRTGPEPGETQEEPG PEATHRAKSLHVPEPQTPSE VVQQRIKWPPASKGSEWLQ FDEDVSNIIQAIAKGDADSR LKTMTTIIFSYALERFGCIEK GKTKPTTPYTMNRRATQIH HLRQELRSLKKLYKKATDE EKQPLAELKNILRKKLMILR RAEWHRRRGRERARKRAAF ITNPFGFTKQLLGDKRSGRL ECSIEEVNRFIEETVSDPLRE QELEPNKALISPTPPAREFSL

				GGGGAGGATGAGATC TCCAATCGGACGGAT CAAAGGTTA (SEQ ID NO: 1163)		RGPSLKEVKEIIKASRSASTP GPSGIPYLVYKRCPGLLLHL WKILKVIWQRGRVAEQWRC AEGVWIPKEENSKNINQFRII SLLSVEGKVFFSIVSRRLTEF LLENNYIDPSVQKGGIPGAP GCLEHTGVVTQLIREAHENR GDLVVLWLDLANAYGSIPH KLVELALHRHHVPSKIKDLI LDYYNNFKMRVTSGSETSS WHRIGKGIITGCTISVILFAL AMNMVVKSAEVECRGPLTK SGVRQPPIRAYMDDLTITTT TVPGSRWILQGLERLIAWAR MSFKPSKSRSMVLKKGKVV DKFHFSISGSVIPTITEQPVKS LGKLFDSSLKDSAAIQKSKK ELGAWLAKVDKSGLPGRFK AWIYQHSILPRVLWPLLIYA VPMSTVESLERKISGFLRKW LGLPRSLTSAALYGTSNTLQ LPFSGLTEEFMVVRTREALQ YRDSRDGKVSSACIEVRTGR

						KWNAGKAVEVAESRLQQK ALVGTVATGRAGLGYFPKT LVSQVKGKERHHLLQGEVR ASVEEERVSRVVGLRQQGA WTRWNTLQRRITWANILQA DFQRVRFLVQAVYDVLPSPS NLHVWGKNETPSCLLCSGR GSLEHLLSSCPKALADGRYR WRHDQVLKAIAASLASAINT SKNHRAPRKAVHFIKAGEKP RALPQLTTGLLHKASDWQL EVDLGKQLRFPHHIAATRLR PDIIAISEASRQLIILELTVPW EERIEEANERKRAKYQELVE ECRERGWRTYYEPIEIGCRG FAGRSLCKVLSRLGITGVAK KRAIRSASEAAEKATRWLWI KRADPWTAVGTQVGT (SEQ ID NO: 1408)
HERO	HERO-3_BF	.	Branchiostoma floridae	CTGACCAGCAGACGG GAAGCCCGCGACCAA CTAGTCTCCGCAAATA TTGCACACAGGGCGA	TAGAAACCCACAAGGC TGAGAAATGTAGAGCA TCTGTATGGACAATATT GATGATTGAAATGTTGT	MALPAVRSGPASTWTLLITL VIVAAKGTDGFMSFKLPLLS TDTWSGYNNDVKTLLGPLH HELATNEMSPKLAGEGFSDI

				CCCTATGGAGCTGATT CAGTCAAATTTCCTCT GAGATATACCGATAA CTATCTACAGAAACTG CACAGTTAGTTTGGAA AGAGCTTTTCTACTGA AAGACAGCAAAATCC GCCACTTTAGACGAG CGTCAAGACTGCCCTC CCCATAACCAAT (SEQ ID NO: 1164)	GATTTTAGATCAAATTT AGAAATATGAAAACCG AACTAAACTAAATATA ATGTTTTTTTTAAAGTA ATGATAAGCAATACCC ACATTGTGCAATACTAT CTATGTTATGTCCTTTG TCCCCCCTGCATGTTTG GTCAATAATGACCATC GTGTCCTGGGCTCCGTG TACCTTTCTTTACTATG AATAAAGAATGATTTT ACTAC (SEQ ID NO: 1287)	MCDFMASKPEFSHTTEESHS EGYISHEPQSLAQVKRLKNK LRKKAFRADATPEDRKAFR DAIKTYSFMKRQQKRKETT KSAAHQEKEYHKNFWKFA GKCAKGQLDIPPVKPAFSVY YANEYYKNKYSHPTRVDFN KLLWFPHLPVEEQLPANSFD MSPVRPKDIKAVLSKRCATS APGPDGIMYGHLKHLPACH LFLSTLFSKLLESGDPPTSWS SGNVSLIHKDGSPEAAENFR MICLTSCVSKIFHQILSERWA KYMTCNDLIDPETQKAFLTG INGCVEHVQVMREILAHAK KNRRTVHITWFDLADAFGS VEHELIYYQMERNGFPPIITT YIKNLYSRLKGKVKGPGWE SDPFPFGRGVFQGDNLSPIIF LTVFQPILQHLKGVEQQHGY NLNDKHYVTLPFADDFCLIT TNKRQHQKLITQISSNTKSM NLKLKPRKCKSMSIVSGKPS

						DISFTIDGDPVKTTKDAPEKF LGGYITFLSKTKETYDILAKT IETTVENINKSAIRNEYKLRV YMEYAFPSWRYMLMVHDL TDTQLQKLDSIHTKAIKTWL RMQPSATNAILYNTRGLNFK SISDLYLEAHALAYSRSVLK ADEKVKHALQAKLDRESQ WTRKMQKWGIGKCHTIHQ QAIHVAKDSEWTSVRKHVK QQVTDMRHDVWTKHQENL LQQGQMLQLLEEEKCDLTW RSAMYNLPRGILSFAVRASI DALPTLCNLTTWGKRNTDK CKLCGNRETLHHVLNHCGV ALQQGRYTFRHNSVLKHITD TIIESIDTSRINATIYADIQGY TTNGGTIPVHTIPTTQKPDLII YLPEQKTLHIHELTVPFEKNI KTSHDRKVNKYSTLAADLE TAGISATLTCFEVGSRGLVT PENKTRLRTLFKIVKAKPPK TLFTDISRIAMLSSYAIWNSR

						HEPYWESETLL (SEQ ID NO: 1409)
HERO	HERO Dr	.	Danio rerio	AAAGCAGTAGAG (SEQ ID NO: 1165)	TAGCATGCCACTTGGA CACAGGCCGGGGTCTG ATCAGCCTCGGTCGGG TCGCCTGGAGGAGGGT GTCTGTTGCAAGACCC GAAACACCCTGTGAGC CCAGGAAACAACACTG ATGATGTGTCCAAGGTT GTGCATCAGGAGATGT TTCTGTAAC (SEQ ID NO: 1288)	MTTHRAEVTTSGKTQEEPGP EATHSAQSLLVSPTPAAGRS PATQSCPQVTAAHNSPQSPQ SQQVAVTRSDCVPLAQPRIQ WPQSSKKAEWLQFDKDVN QILEVTGKGGVDQRLSTMT TLIVNIAAERFGTVTPKPTPS TYTPSHRVKEIKRLRKELKL LKRQYKAAGEVERAGLEDL RGILRKQLVNLCRAEYHRK RRRERARKRAAFLANPFKLT KQLLGQKRTGKLTCSKEAIN NHLKATYSDPNREQPLGPC GALLTPPEPTSEFNMKEPCR SEVEEVVRRARSSSAPGPSG VPYKVYKNCPKLLHRLWKA LKVIWRRGKIAQPWRYAEG VYIPKEEKSENIDQFRVISLL SVESKIFFSIVAKRLSNFLLS NKYIDTSMQKGGIPGVPGCL EHTGVVTQLIREAREGRGDL

						AVLWLDLTNAYGSIPHKLV EVALEKHHVPQKVKDLIIDY YSKFSLRVSSGQLTSDWHQL EVGIITGCTISVTLFALAMN MMVKAAETECRGPLSKSGV RQPPIRAFMDDLTVTTTSVP GARWILQGLERLVAWARMS FKPAKSRSLVLRKGKVRDEF RFRLGQHQIPSVTERPVKSL GKAFNCSLNDRDSIRETSTA MEAWLKAVDKSGLPGRFK AWVYQHGILPRLLWPLLIYE VPMTVVEGFEQKVSSYLRR WLGLPRSLSNIALYGNTNKL KLPFGSVREEFIVARTREHL QYSGSRDAKVSGAGIVIRTG RKWRAAEAVEQAETRLKH KAILGAVAQGRAGLGSLAA TRYDSASGRERQRLVQEEV RASVEEERTSRAVAMRQQG AWMKWEQAMERNVTWKD IWTWNPLRIRFLIQGVYDVL PSPSNLYIWGRVETPACPLC

						SKPGTLEHILSSCSKALGEGR YRWRHDQVLKSIAEAISKGI KDSRYRQATAKVIQFIKEGQ RPERTAKNCSAGLLSTARD WVMTVDLERQLKIPPHITQS TLRPDIILVSEATKQLILLELT VPWEERMEEAQERKRGKY QELVEQCRANGWRTRCMP VEVGSRGFASYTLSKAYGTL GITGTNRRRALSNNVEAAEK ASRWLWLKRGEQWGQ (SEQ ID NO: 1410)
HERO	HERO Fr	.	Takifugu rubripes	AGACTAGGTGACAAC CAAGAACAGTTWGGT CGACTACTGGAAAGA CAGTTGGCAGCTCGG AAAGACGGCACCCGG GACAGTATGGGTTAG CACCCCAGCCTGTATC TTTCGCGAGAAGGAA CCCAAACAAGCTACG GAAAGCCCTACAGAG AAACACCCCCAGGAG	TGATCACCCCGGCTGG GTCGCCTGGGCGAGGG TGTATGATGTCGTGAG ACCCGAAACACCCTAT GAACCCAGGATACATC CTGACGATGTGTCCCA GTGCATCCAGGAGATG TAKCTTTAAGT (SEQ ID NO: 1289)	MTPAMEMTTTVTCICSKLC KNQRGLKIHQARMKCLERE VEVQRTGPGPGETQEEPGQE ATHRSQSLHVPEPPNPNRVV QQQRIKWPPANRRSEWLQF DEDVSNIIQATAKGDVDSRL QAISTIIVSYGSERFGRIEKG NTETTSYTMNRRSFKIHQLR KELRTLKKQFKRAXDGDKQ ALKELYNILRKKLKTLRRAE WHRRRGRERARKRAAFIAN

				ATCCCGAGAGGGGGG GAGGATGAGATCTCC AATCGGACGGACCTA ACGTTA (SEQ ID NO: 1166)		PFRFSKQLLGDKRSGRLECS REEVNRFLQNTMSDPLRGQ DLGPNRALISPAPPSAEFKLA EPSLKEVEEVIKAARSASSPG PSGVPYLVYKRCPEILRHLW KALKVIWRRGRVADQWRC AEGLWIPKEEDSKNINQFRTI SLLSVEGKVFFSIVSRRLTEF LLKNNYIDTSVQKGGIPGVP GCLEHNGVVTQLIREAHESK GELAVLWLDLTNAYGSIPH KLVELALHLHHVPSKIKDLI LDYYNNFRLRVTSGSVTSD WHRLEKGIITGCTISVVLFVL AMNMVVKAAEVECRGPLS RSGVRQPPIRAYMDDLTVTT TSVPGCRWILQGLERLILWA RMSFKPTKSRSMVLKKGKV VDKFRFSISGTVIPSITEQPVK SLGKLFDSSLKDTAAIQKST EELGGWLTKVDKSGLPGRF KAWIYQYSILPRVLWPLLVY AVPVTTVESFERKISSFLRR

						WLGLPRSLNSAALYGTSNTL QLPFSGLTEEFKVARTREAL QYRDSRDCKVSSAGIEVKTG RKWKAEKAVXVAESRLRQ KALVGAVATGRTGLGYFPK TQVSHARGKERNHLLQEEV RAGVEEERVGRAVGLRQQG AWTRWESALQRKVTWSNI MQADFHRVRFLVAAVYDA LPSPANLHAWGKSETPTCSL CSGRGSLEHLLSSCPKSLAD GRYRWRHDQVLKAVAESIA LAISTXKHHHAPKKAISFIKA GERPRAGPQITTGLLHTAXD WQLHVDLGKQLIFPQHIATT SLRPDMIIISEASKHLIMLELT VPWEERIEEANERKRAKYQ ELVEECRGRGWRTFYEPIEV GCRGFAGRSLCKAFGRLGV TGTAKKRAIKXASEAAERA TRWXWLKRADPWVATGTQ AGS (SEQ ID NO: 1411)

HERO	HERO Tn	.	Tetraodon nigroviridis	AGATTGGTCTGGCTAA GCCAGTGACGTCCAG GAACAGACTGGCTGA CGACCACGAATAGAG TGGTGACAGCTTGGAT AGACAGCTGACAGCA GGGAAAGACGGCAAC CGGGGCAGGAAGGGC TAGCAACCCAGCCTG CATCTTCCGTGAGGAA GAACCCAAAACTTGC TACGAAGAGCCCGAA GCAAAGATACCCCCA GGGGAGCCCGAGAGG GGGGGAGAATGAGCT CCCCAAACGGACGGA TAAC (SEQ ID NO: 1167)	TGATCACTCCCAGTCGG GTCGCCTGGGTGAGGG GGTCTGATGTTGAAAG ACCCGAAACCCCCGAT GACCCCAGGTACTATC ACTGACGATGTGTCCA AGACATGCATCAATAG GTGTATTTAGAAATC (SEQ ID NO: 1290)	MATTQASVKPTAVATCVCG KICKNPRGLKIHQTKMGCLA SVQPEQRARFSLSESREVPA RAEPYGPQQPHSPEALGETQ EERGQESPHSAQNLRAQVA QAPDNPQHHRRVKWPPASK VSEWQQLDEDLEGILESTAK GGVDRKLQTMTTLVISFATE RYGTMEKRAAPEKYTKNRR AEKISQLRQELRVLKKQFKG ASEDQKPGLAELRCTLRKKL LTLRRAEWHRRRAKERAKK RAAFLANPFGFTKQLLGQK RSAHLECAKEEVDSYLHDTF SDAERENSLGECRVLISPPEP ACSFNTKAPTWKEIQTVVR AARNNSAPGPNGVPYLVYK RCPKLLARLWKILRVIWRRG KVAHQWRWAEGVWVPKEE KSTLIEQFRTISLLNVEGKIFF SILSHRLSDFLLKNQYIDSSV QKGGIPGVPGCLEHCGVVT QLIREAREGRGSLAVLWLDL

						ANAYGSIPHKLVEMALARH HVPGPIKTLIMDYYDSFHLR VTSGSVTSEWHRLEKGIITG CTISVIIFALAMNMLAKSAEP ECRGPITKSGIRQPPIRAFMD DLTVTTTSVPGCRWILQGLE RLMTWARMRFKPGKSRSLV LKAGKVTDRFRFYLGGTQIP SVSEKPVKSLGKMFDGSLK DAASIRETNDQLGHWLTLV DKSGLPGKFKAWVYQHGIL PRILWPLLVYEFPISTVEGLE RRVSSCLRRWLGLPRSLSSN ALYGNNNKLTLPFSSLAEEF MVTRAREVLQYRESKDPKV ALAGIEVRTGRRWRAQEAV DQAESRLHHKELVGAVATG RAGLGTTPTTHLSRLKGKER RDQVQLEVRASIEEQRASQ WVGLRQQGAWTRWEEAM ARKISWPELWRAEPLRIRFLI QSVYDVLPSPSNLFLWGKV ESPSCPLCQGRGTLEHILSSC

						PKALGEGRYRWRHDQVLK AIAESISSAMEYSKRLPLPGR GVRFVRAGEQPPPQPRAQPG LLATARDWQLRVDLGKQLK FPENIVETNLRPDIVLHSQSS KQVILLELTVPWEERMEEA YERKAGKYAELVEDCRRAG WRSRCLPIEVGGRGFAGKSL CKAFSLLGITGMRRRKAICA ASEAAERASRWLWIQRDKP WTSASWTQAGN (SEQ ID NO: 1412)
NeSL	LIN9_SM	.	Schmidtea mediterranea	AAACGACATCATGAA CGCTTGGCCGCAACA ATCCAGTTATCCCTGC GGTAACATTGTGGAA CTCATAAGACAAGTA CTAAAAGAAGAATTA GAAAAATTAGAAGAA AAAATTGAAAATAAT TTATTTATAAAATTTA AAAATTTAAATAAATT TAAAAATTTAAATTTA	TAAAATGGCAAAAAGA TATTTCAAGATGAATTG TGGACTCATCTAAAAA ATGACCACCTTGAGTCC AAATATGCCTAGCTATC ATGGTTGCTGATGGAA ACAGTAAGGCACCTGA TAGCTAACTTTTCACTG TGAATATCTTCAGATAT TCACAGTGACACGAAA GGACACCACTAGTAAA	MMDSRQLNTPKIRKYQNPK MTNDIMKSYNYAVLSDVTP QETTQTTTHLNVDIDNETTQ PKQPLTKSGKPKSKPIAVSY KFKDATFIWDTTPQTNPPRD CTKLIDKTRPRKTIFKKSAFQ SYLKKELSNETFVEVKTFLM ATHKYRFKDENSRLLAYRII NRYVMETANEFKETEFDMA RFAKFFTIPENWLKHLKPYS TATETSPADRIKVQKLVDLT

				AATTTAAATGAAGAT AAAAATTTATTTAATC CAATAAATAATCAAG AAAATCAAGAAA (SEQ ID NO: 1168)	AACCACTAGTTTTTTCT GACACCTCTTGCTACAA ACTCTGTAAAAATCAA AAGGATCGATAGGCCG CGCTTTCACGGTCTGTA TTCGTACTGAAAATCA AGATCAAGGAAGCTTT TCCCCTTTTAGTCAACA CCAGGTTTCTGTCCTAG TTGAGCTTCCCTTGGGA CATCTGCGTTACCATTT GACAGATGTACCGCCC CAGTCAAACTCCCCAC CTGACACTGTCCTCAAA ACAGTTCAATTGCATCC GAAGATCGCAATTTTTT CACTAAAATAAATTAA CAAAAGTTAATTATACT GCTTCATTGAGTAAGTA GAAAAACAATC (SEQ ID NO: 1291)	CRYPFKTQEEQTSVANFLHF FTQRSIIGISRDYKFQKFIPFM ARKNTRPETTSTMVTTSPTE QNRLPMVIITPLEEPKSEHRR PEKRGASNDTIVLSDEEFPLL KRRTLPTRKSKNPTGAGNVP TETECTDEVKFILNNEYQIEC KECGKVWENVRNGLNHLR QKHDFPNRTDVMVSCVRCE VPIKGAECVNHIKNHKKDD KEESEAGSLVANTQDIPNES SLSQAAIEVYLRNILKMKEN QERNIQYLEPSTANFLINRNL RAFYQNVKIEKLIGWEQVIW LIHWNKCHWIVYLANCDSK TSVILDSDNQMTLQQRCNIK AKFDKFLEGTFEEKTVLGTL ERKVPQQPNNFDCGIYVIQY ISDFLKDPQRIDYHTPDSKRI RKEIGELILEEMKNPASKIKN PNKEIQSLLQKFRLLQINVN DVFHWFAAEYQKSLPKIRT KRDGKLNKLSCSYQIQRLFG

						LAPKRAVKEIYFQETSTADL ETRVLNEHFKKDESTMKEC KIKNGNHYQDWITKAQIDN KEILEALKNSTDSAPGEDNIP LRQWIIWNNDGVLFDMFNY IKRTHDIPDMWKNYTTTLLI KPGKSQESNIPANWRPISILP TSYRIFMKVLNKRVLEWAN RGELISKWQKAVDKANGCD EHSYVIQALIEKANRSYYKN EQCHLAFLDLADAFGSIPFQ VIWHTLKNMGMDEETINLL KEIYKDCSTKYKCGKNESEK IKITKGVRQGCPLSMTLFSLC IQYLIQGIAEKKKGATIAGQE VCILAYADDLVIVANTAKD MQMLLTTIENLAKQADLIFK PAKCGYYRDPRDKKSMMKI YGKEISIVDEKNVYTYLGVR IGDTKKKDLNVRFEEVKKK TTAIFKSKLRSDQKLEAYNIF CQSKFVYILQGEDIAKTKIET YDEEIKKMIKEDILKLQDKS

						PFTDFVIYSPREKGGLGITKII DEQTIQTINRTAKLLNSSHR AIRAIIYEELIQVANLRGEKEI NTIEEALKWLEGTNKYKKN SNAKTTWITRVREAFQTLEK KHKIKVRFVPKENCIGYKIK CDTQEKIVELDNSKELSKSL HWMIKEAYYKEWKALKCQ GYIISLKTSEFMEWKMPRGL PDPDWRFLTKVKANMLDV NMKQANQGGRLGSTKCRK CEDKESASHVINHCASGNW SRVEKHNQVQNELAKELTK RNISFEKDSIPKETKESLRPD LVIRLKDKIMIVDIKCPFDEE SAIESARNKNIDKYRELAKEI QAKTGLQTTVSTFVVCSLGT WDKRNNELLRQMGIRYEES KEMRINMIQKAIHGSRKTYD HHRNFNNG (SEQ ID NO: 1413)
NeSL	NeSL- 1_CBre	.	Caenorhabdit is brenneri	AAGGACGCTGGTTTA AGGCCGAATTCGTTCG	TAAACCCACACGAGAM CTACGACGCCATAAGA	MPLXISDCVHLVSAEGDTM NGRSTCGPLSRSSSVVSRSR

				TTCTTTTTCTGGCGGT	TCAGGCATGTACGGAT	SSPSPSVPPHPSPSIGPDTGLS
				CTTGCTTTGAGCTTGG	GTGAATGAGACTGATG	AGIIGTSRGCSLWLPEVDNA
				TTTCCGATCCT (SEQ	AACGGAATGAGCACGT	LSQWLRKGLERDHEVLVCG
				ID NO: 1169)	GCCCATAAGATCGGGT	FEAAKPLSLSKARLLRKTPR
					ATKAAAGAWCAGAGA	NTGVVRHILEFDGRLVHTNC
					CGATCCCTAMCATCGG	NETECVLSTLXSXXAVEVV
					GAAAACACGAGTTATA	RISLKCEPREPCEPKCVLSIL
					CTGCTTCACTGAMCTC	CSDKIVXISFECETREPFPFFX
					GCTAAGCTCTCATAATG	DRKFREPIPFVFERMYDPRD
					ACCGAACTTGTTCGCA	PIPSFICWMYDLRQRMTPGT
					ACTGCCTCCTAACCGG	LPXNPLSXENKDSWGRPAVI
					GCGGGTGTGAGAAGGG	KNEIRSMRSYLEENVKENRL
					AGGTCGCCTTGAGGCG	NLLRRLRGGGEGKKMIRKL
					GACGCAATGAGGGATG	VAEKKSDTEAVCRILYPLDD
					TGTGCAGGTTCCCCCTC	RYECFVDGCETTSTMGYGS
					TTGAGATCCGAAAGTC	SDLKYMTTHIKKEHGVKVQ
					TAAAAGTACTAGACCG	WTYECSLCNKQAPFMGGAA
					AAAGATCGAGGACGGA	SKWVTAHMATKHTETVKL
					CGGGATGGCCGCGAGG	KLKPSISTTAKVAAKLDEIA
					CACACGGCGGGTAACA	VSLPKPRQVRVLRDPDEVKE
					CAGCCAGATAACCTAG	KVAKPTLASTREEVKRNAL
					TAGATCTTCGGATCTCG	RNMAPLVELSSQNQLTGAE
					TCGGCCTGGAGATATG	RPEETSEAMRLEECRTPEKI

					TGGAACCCTGGGAAAG GAGAAAGTTGTTTGTTG GGCTGGCAAGAGTGAA GTTTGAATGTGAACCA CCGTCATGCAACCACT AAACCAGTGGCGATGC GGGTGGAGTCATCACA GGAAAATGTTTCTGTTG CTTGACTTATCAGTGTT TGATATCGCCCTCAGGC ACAAGTATGAAGGCCC CCACCCACATAAACTC CCTAGCAACTGGTAGT CCAGCAAGCGCTGGTW CTTGCTACTATTGCGCC CCAGGCTCGCCC (SEQ ID NO: 1292)	AELEGKIQTRTVTKKLSALK ESMEKRTREEKVGKPSLAPI HEEVKKTARRSLAPLVEPST FTHLTGASRLQAVRDAFSK ANKDAAAKRRSSLAKPARL SEIMNTTFTKETVNETKEPV NDTDESIATIQPQVRVYRFN TWCLDHETTREAWLTGEVV DWFMGKVTEKKDQYRVFD SLVWSMYKFHGVGYVLDL MRDPLTYFLPICEHDHWVLL VIDEKGIWYGDSKGAEPCRE IAKFIEETKRERRMFPVPVPL QRDGVNCGVHICLMVKSIV NGEPWYTEEEVKVFRRNVK RGLKEFGFELYSERIVYVGD DSIKVNDEHDDDVVFLSEET NNTTFTIEQAEDPAEEDAQH LESPVKPVKLMELKIPKIEIK KKEIRRKPKQQIEKKRKVPT GKPDELLVRVRLWLEREVQ SYFDSGKRFQRLEWILDVLT AAIHKATAGDEQAIERIEKR

						SPPLEVEEGEMSTQTEPKKR ERKEKESGCEMKASHKEMY FKNRSKAFNVIIGKDSKQCEI PIETLQKFFEGTTAETNVPAE VLKEMGSRLPKLEALDWME ANFIESEVSDAMKKTKDTAP GVDGLRYHHLKWFDPEYK MLTLLYNECKNHRKIPSHW KEAETILLYKGGDETRPDN WRPISLMPTIYKLYSSLWNR RIRSVGGVMSKCQRGFQER EGCNESIGILRTAIDVAKGK RRNLSVAWLDLTNAFGSVP HELIKSTLESYGFPEMVTEIV MDMYRGASIRIKSKNEKSEQ IVIKSGVKQGDPISPTLFNMC LENVIRRHLDSASGHRCIKT KVKVLAFADDMAILAENRD QLQTELNKLDKECESLNLIF KPVKCASLIIERGMVNKNAE VVLRGKPIRNLDENGSYKYL GVHTGIATRVSTMQLLESVT KEMDLVNQSGMAPFQKLDC

						LKTFVLPKLTYMYANAIPKL TELKVFANLTMRMVKEIHEI PIKGSPLEYVQLPPSQGGLG VACPKITALITFLVNVMKKL WSSDSYIRKLYRDYLDEVA ETETGMEEMTKEDIAKYLS GDVPIDKKAFGYNTFTRVR DVCNSLTXIXGAPLHKLKIV ERDGDFAILVQATKEGMEKI FTCAQEKKLQQLLKAEVNT ALAHRFFTEKPVKSAVMSV MRQYPQSNAFVKNGKNVSI AVHSWIHKARLNALHCNFN TYGENKSKVCRRCGKDVET QLHILQXCEYGLPKLINERH DAVLHVVRNLIRKGSKKDW KLKIDETVSSCNQLRPDIYM CSPDGKEVIMADVTCPYESG MQAMQESWNRKVTKYEGG FSHFXKMGKKFTVLPIVVGS LGTWWKPTTNSLVQLGIEK XTIRRVIPELCSMTMEYSKD VYWNHIFGDTFRKPPMRFG

						VEKPKGNSWKKEGSEPKGA ASSD (SEQ ID NO: 1414)
NeSL	NeSL- 1_CJap	.	Caenorhabdit is japonica	GCGCCCCGGGTTACAT TGTCGGGGCCACCTTT CTCTTGGAGTAGAGTA CAGTCTACTAATTTTT TGATAAGCTAGTCGG GTCCGAACCACTAGA GTTTGCTTGAAAATGC GTCAAACCAGCATTTT AGAACTCGCCCAAAA GTTCGGCCCCGACCCC CAAACAAATGGGACC TTCTTGACGATTTTCC CTGAAAATCGGAGGA TGGAATGGTCCCCTAT TCTTGTAAATAGKACT GTGCAATACCCCTTCG TCATCTGTGGGGAAC AGATGACACGTGACG TCATCCGTGTAGACGT CACGTTTTCCCGTGCC TGCGGGAGCCCCCAA	TAAAAGCCAAAAGCCA CGGAGCATCGGGAAAG AAAAATGGAAAAGGAC TGAAAACGAGACTGAA AAATCCCAAACAAAAC AAATCCAAAACAAACT GAAAAAAAAAAAAAA ACAAAACAAAAACTGG ACAGACACTGGAAACA GTGTCAGGCAAAGTCG CCGATTATACTGTTCCA CGCCTTAAAAGTCCCG AAATGGCGCAAAACAA CCTGAATCTATCTGAAA GTGCTCCAAACCACGC ACAACTCGGAGAAAAT CAGGGACAAGTTGCTT CACGCAACGGGCTGGG ACAGGTACCCCCTCCTG AAACCGCGAGGTTGAG GATGGACGGGAAGGCC	WRRPAPKQTKNSSLHHLGH EVKRIARLKPGIFEFHAKPK NSSLHHLGHGVKRXARLKP GIFEFHAKXKNSSLHHLGHE VKRIARLKPGIFEFHAKPKN SSLHHLGHEVRRNSRLKPGI FGFYQKSKNSSLHHLGHEV RRIARLKPGILEFHAKNRIKS GLKVTFLSDLXAHAGALAC SRFLASTLKTEHCRQKSFKP VGFLLHFLKNSSINEVASLR NVKKXFLEFFSGKPIGGMAS FSRTKITFFKLCLKNFVLSAE NPPIIRQKTNQNKASXVQIA RGGHLSDCLPSQKMAGVLG RLFLSVQSTLSHRPFDTLLRS DDDKRGRKTIKLQFFIKENL VTPXVARDVKILXKQTKNN SGNSDSNSETKNFSKNKVSR QNGPLIGGGNHKKIGENQIT RTLEIESKSDDNKVLVLRILY

				TCGAGCAATTTTTGCT	GCGAGGCTTATGGCGG	PTNDWYKCYSQWCQHKSL
				CTTTTGAGTGTCTGGA	GTAACTCGGTTGGTGTG	VGYGAHDLKYLTDHIKSTH
				ACGCTTGAAACCCCA	CTAGTAGATGATTTATA	SKKVEWSYQCSICDAKAEG
				GACAAATCAGGCCCA	TCCGACAGCCCCAACT	TGTKAARWITAHMPKVHGI
				GTCGTCGGAAAATTTC	AAGAGGAATCCTGGGA	EATHRIKQNSEKTTNVKTAN
				TTTTTGAAATTTTTTG	AAGGAAAACTTGAAAA	SLQEMALSLQKPKNGPKKV
				GCGCCTGCGAAAAAA	AGTTTTTACAGGGCTGG	VMATSTTPEKKISELESKIQT
				ATTTTTTAACCGCCAC	TAATAGTTCAGCACAA	REVAKQLSALKESAQKNQQ
				AAACCCCCGGGAGGC	TTGTAGTCTACTGTCTT	GNKTKNVKSSLKTIAENTNE
				GCGGWTAGGGATATC	GCAACCACAACAAACC	TKKISARKSLINYLKPEDVL
				GATGTCATCGACTCGT	AGTGGTTCTGCGGGTA	NHIPKEPKPASAKXGLQELT
				CGGTGATCTTTGATTT	GATCAAACTATAATTTG	GAQRLQETRRRFMAGNRRD
				TCTCTCTGCGTCTCCT	TGTGTTTTCTTTTACTT	SIARRESLSLGKISNSFKIELK
				ATTTTGGAACAGTCTC	GACCCGGGCAACACAT	NAPEKTTLKKPAVTQKQNT
				GACCAAAAAACCGGG	TATACCACGTCCACAA	SQNVSSSTVVKENKTGNDVI
				CCTGGCAACCCACCG	GGACGAATTCATAATG	TIDDTETVKRKINTWCLDHE
				AATCCGGATGTCGGA	GCCCCTCCCTAAATAA	STENAWMADDIIFWYIQKQI
				GGGATTTGGCAAGAA	ACTCCCTAGCAACTGGT	EISLDNKKFKVIDPLIWTTYR
				ATGTTGGAAATAACG	GGTCCGGCGAAGCCGG	IYGVECVQDELVGFEKYFFP
				AAATTTCGTTATTTTC	TTCTTGCCACTATTGCG	ICENGHWVLLIIDDKRVWYS
				AGCACAATTGTCAAA	CCCCAGGCTCGCCC	DSLADKPIEVIEDLINKLNRT
				CCGGCAAGAAAACTG	(SEQ ID NO: 1293)	QGKFNQTVPKQKDGFNCGV
				GATGGACAAGACACA		HVCLVAKSVITENFWYTEK

				CAATTTACCGGAAATT GTGCTTGTTACGTCGA ATTTCCCAATTTTGAA AAAATTCCTCGTTCCA CTGGTCGGGACGCGA GGTCAGACGATCTGC ACGTCTGAAACCCAA AATCTTCGGATTTTAT GCAGTAG (SEQ ID NO: 1170)		DVNDFRKTVKLWLFSEGFE LYSEPYKQIQNKNISVNSEK NQISDNEKNWGDKTQTVNE STLKERDEDIFLLRPHISVGV ALKTEDEKNQKAENLKAPQ KLKAIRRLKILKTCLKKLTA VKGKPEETERAAIPNLMAIK LKTPPKVEPVRRNPEKGENY XKSQPNKKRQIPTGKPDELV KKVREWFEIQFQAYFEDGK SFQRLEWXTGLLTAAIHKAS AGDEQAVGKIIKRCPPLEIEE GEMATQTETKQKPKNQKST KGANSSSSIREAYAENRART FNKIIGKDDKCEIPIEKIEKFF ENTTSNTNVPTETLARITSDL PKLEIGSWIEEEFREKEVAEA LKKTKDTAPGVDGLRYHHL SWFDPKXKLLTKLYNECRE HKKIPGHWKEAETVLLYKG GDETQAENWRPISLMPTICK LYSSLWNKRIKSVTGVLSKC QRGFQEREGCNESIAILRTAI

						EAAKGTKKSLSIAWLDLTN AFGSVPHESIEATLIAYGFPG MVTEVIKDMYNGASIRVKT KNEKSKQILIKSGVKQGDPIS PTLFNICLESVIXRHLKSADG HKCIXSNIKLLAFADDMAIL SDSKTKLQQELQKMDDDCT PLNLIFKPAKCASLIIEWGKV QKDQKIKLKGQFIRSLAEQD TYKYLGVQTGIETRVSAMQ LMKKTVSELDKINCSALAX WQKLDAVKTFVLPKMTYM YANTVPKLSELKEFANITMR AIKVMQNIPVKGSPLEYVQL PIGKGGLGVACPKTTALITY LVSTMKKLWSTDDYIRKLH TDYLKMVAIKETKTKEVTL EDLASYLSDDKTVCKKAVG YNSFTRVREICKTLSKNKGA LLSQLKIIAKDGKLAILVQA XKDGKTKIFTHDHVKTLQK XLKKEINEALLHRFTTEKRV KSEVVRVVQEYPQCNSFVR

						DGGKVSIGAHRFVHKARLN LLACNYNTWQDAATKQCR RCGYEKETQWHILSSCPKSM GGKITERHDSVLKTVKEMIQ TGSLKNWKLKLDHELPGST RLRPDIYLRSPNGSEIILGDV TIPYEHGIEAMQTAWQKKIE KYEEGFKYLRSTGKKLTIVPI VVGALGSWWKPTTDSLVSL GIDKNTVKRAIPEICSTVLEY SKNIYWNHIFGDSYQKVPM FFGGEKPKGQSWKKVKPPE GKTASNHEPPG (SEQ ID NO: 1415)
NeSL	NeSL-1_CRem	.	Caenorhabditis remanei	CGCGAACCAGTCAT (SEQ ID NO: 1171)	TAGCCGATCGTAAAAG AAACCGAGCCGTAACA ACAAGCAAAGTAAACA AAAGAAAAATCAATAA AAAGGAAGGTTGACCT CAGACCCCGAGGAGGG AAGAGAGACACCSAGA AAAAGAGAGACGCAGA GAAAAGGAGAGACACC	MTVFIDRGIGERGQMAVCSL HRYFSFSPFSPIPPYVNNGSF GENGCGTDKSLLPVIEVVVR EVKINWSENILVVECLIMVK SGERVVVKRQNLEKVIQNL ARINSTLFSNLGNQIFCVVPR IKDSTNKEQGYRKEKQXKF HVSFRSIKSQVPPYLRGGGD VMEDTEIRGIRKLEPEAQLD

					TCTCATAAGGAGAGGT	SSKPLICRVLYPTQGYMYKC
					AGGTCAATCCAAATGT	FYPKCKGHSNGSTDLRSLKK
					AAACAGAAAAAACCAG	HMVDKHFTNIEFAYKCATC
					TGGGGAGGAAAGAAAG	MFLTTGKSATALKSIKAHM
					ACTGATTTCACCCACTA	ASHHKVTMEPGKKSLVQKL
					AAATGAATTTGGAAAC	NARLEEAAPSLPMPRNRSKV
					AGAATTTGGAAGAGAA	IQLTPEKSISELEKKKQTRSV
					AAGAGAAAGGGAAACC	AKQLSTLKESAQKKEEEVKI
					TAAAGAAAATAGTTCT	AEVKKREPRLSIIPESNVRRS
					CTTGCCAAAATTCTGTA	LAAGLEQCINPEQSVAQRIR
					GAGGAATACTTTGTCA	EKREEYAKASREAAAKRRS
					AAACATGATAGAAACC	SLAMKPARLPDKENEITLQE
					AGTAATCTGGTACGAA	TKKIDDPIVIDLEKECILTTV
					AGACAAGTAAGACCTG	LQVPRNQFNSWCLEHETTID
					AACTGACAAGAAGGAA	AWLTDEVIHMYMCTITENR
					GTCAGAAAGAAATACC	KYFMAIDPVLWPVYVRNGA
					GCTCACAAAGCCTGTG	EDLLRRTSCPGTFFFPICESN
					ATCGATTCTCTTACCTA	HWVLLVIEHDVYWYLDPKG
					CTGAACTTGTTCTCTTG	EEPKGNVEILLESMKRKRQY
					GCCTCGTAACCGGCTA	YEFPPPSQRDNVNCGVHVC
					AAGGGAGAAGGAATGT	LMAKSIVDECGYNWYSEED
					TAATTGGAGATAGACA	VRSFRTNMKDILKSKGYELC
					TAAAGATAGGTGGAGT	PEPYNRQNLLKTEKQKEVIL

					GAAGGTCCTGTTCTTGA	EEMIDSFVVEDDMTFTVHR
					AACTAGGAGGAATGTG	DSDHGDDEVEHLKTIEQEPE
					GAAAGAGCAGAAGGCC	NEISEIENVEGSVDSVIPKLM
					GCGAGGCTTTAGACGG	EMRVQTPPVINEKRGKKRV
					GTAACTCAGTCAGTTGC	SAKEKPRKQKEKEQKVPTG
					TAGTGGTCTTCGGATCC	KPDELVKRVRVWFEKEFKS
					AACGGCTTCGGACATA	YVEDGKSFQRLEWXTDVLT
					GTGAGGAACCCTGGGT	AAIQKASAGDEKAVELIEKR
					ACGGAGAAGAAATGGA	CPPLEXEEGEMCTQTEKKK
					AAAGAGATAGGGCGGG	KPKSGKGNGGQESMKSLMA
					CAAAGGCTAAGTTCAT	SYSENRAKTYNRIIGKHSKQ
					ACACTGTCATGCAACC	CEIPIAKVQKFFEGTTAETNV
					ACTAAACCAGTGGGAT	PKETLKEMCSRLPKVEVGT
					CTGCGGGTGAATCACTT	WIEGEFSESEVTEALKKTKD
					TCGAAAAGAAGTGAAT	TAPGVDGLRYHHLKWFDPE
					GGACGTGCTGATGTCT	LKMLSQIYNECREHRKIPKH
					GACTTTAAAGAAGTCT	WKEAETILLYKGGDESKXD
					GAAATTAAAAAAACAG	NWRPISLMPTIYKLYSSLWN
					ATATAAAGGCCCCTCA	RRIRAVKGVMSKCQRGFQE
					CTATAAACTCCACAGC	REGCNESIGILRTAIDVAKG
					AACAGGTGGTCCGGCG	KKRNIAVAWLDLTNAFGSV
					AGGCCGGTTCTTGCCAC	PHELIKETLESYGFPEIVVDV
						VEDMYRDASIRVTTRTEKSD

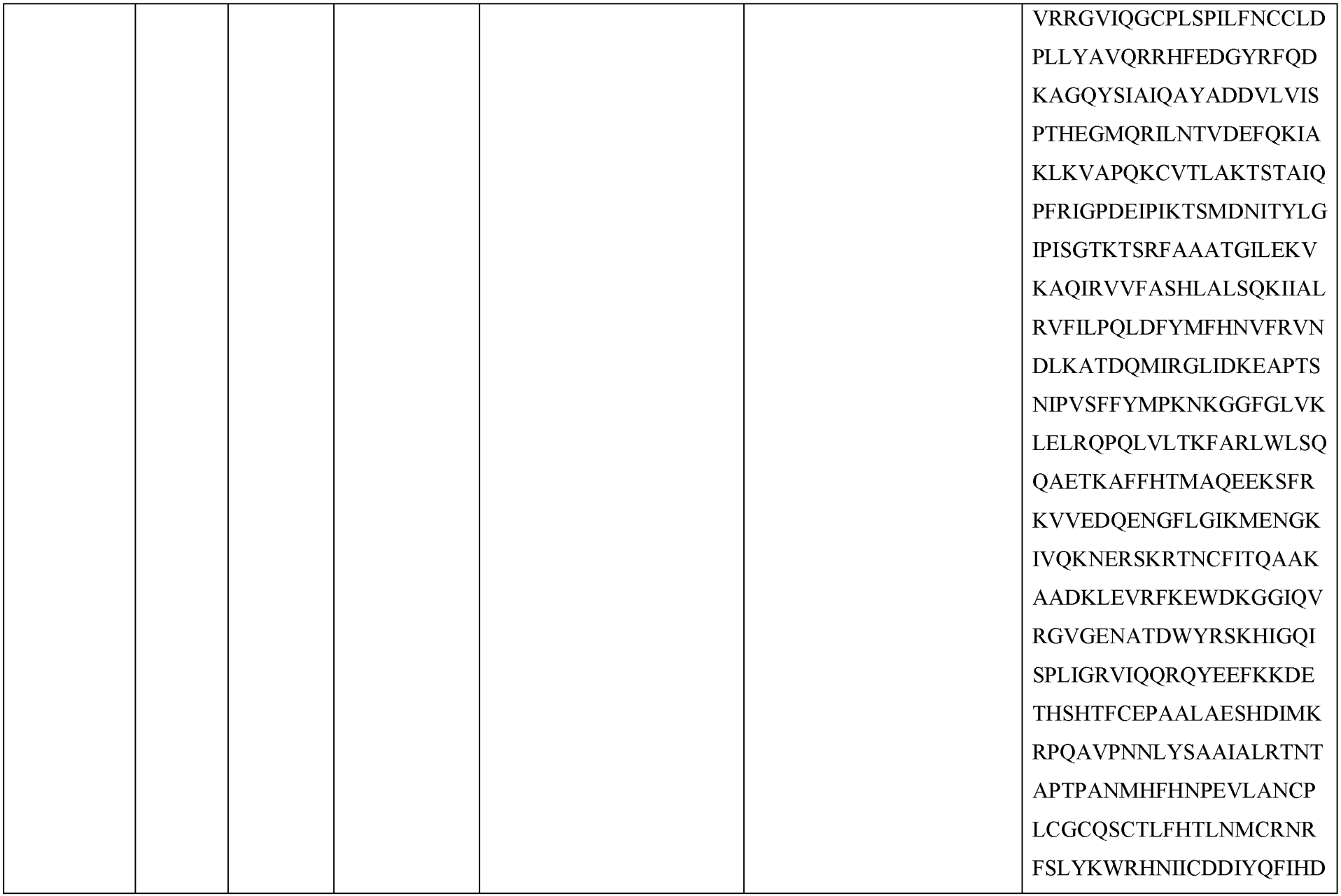
					CATTGCACCCCAGGCTC GTC (SEQ ID NO: 1294)	QIMIKSGVKQGDPISPTLFN MCLESVIRRHLDRSVGHRCL KTKIKVLAFADDMAVLAES SEQLQKELTAMDADCSALN LLFKPAKCASLILEKGIVNRL NEVVLRGKPIRNLMENETY KYLGVQTGTETRVSIMDHIT EVSREIDLVNMSQLAMHQK LDILKAFILPKMTYMYQNTT PKLSELKVFANLVMRSVKEF HNIPLKGSPLEYVQLPVGKG GLGVACPKNTALLTFLVTIM KKLWSSDSYIRKLYTDYLEE VAKVEIGKFEVNLNDLAEFL SDERAVDSKLFGFNAFTRVR EVVRSLCKNKDSPLHSLKIIE REGKLAISVQATEESIEKIFT EDQEKKLMYLLKGELNTAL QHRFFTQKVFKSEVMRVVQ QHPQSNSFVRNGGKMSFSA QRFVHPGRLNQLPCNYNTW AKGRTKLCRRCAKNENETQ SHILQVCDYSIGNIIKERHDA

						VLYKFRELIKRGSKGHWLE RTDRTVPNTGSQLKPDLYLE SPDGKHVILADVTVPYERGI EGMQKAWNEKINKYTDGY KEIFRRQGKSLVVLPLVVGS LGTWWKPTEESLIKLGVEKT TVRRIIPETCGMVAEYSKNC YWRHIYGEKYVQTPMINGG KKPEGNDWKKCEKGIEVPK VTN (SEQ ID NO: 1416)
NeSL	NeSL-1_TV	.	Trichomonas vaginalis	GGGTGAGTAGTCTAG TGGT (SEQ ID NO: 1172)	TAAGAAGAGATAAGAC GAGTGAGAAGAACAGA AGCATAGTAGGATTGG CAGAGCTTAAGCGATG TCACTCGGTACGAAAC GTGTACCAAACACCGG ATTCCGTGCTAGGAATC ACAAGCCAAAATAAAA GAGACACCACGAAAAT TACTCACCCTCCCTCAA ACAGATAATAATATTA ACCTCCCATCCATCAGT CCGTATGGTCTGATAAC	MIPVLGTGGPEKLPLQSYVY CGNTAITDSFTPTAKTILKPE EQNLDIVLKNIAALNPENYS DLIRSLSKMEFRLDYPKEIEN YWISEKLFSQSIASLPISLLV ASMFSPEDRDLSTEPFHCNA DGCNFHCDNCERMVEHIRE HHNTDPMINTFETTEDTFRRI TAIKIDKTGIEELNPLKYRCS YCDELFTEAEDHAIHMISHL TEKLSPDISFFFNDILRLYKTI DKPTVQNLFPETQVAIFDTL EETNRFRLIVGREAIETIEEA

					AGACTAGCACCACATC CATGATACACTCATTGG AGTGAAAACCACCAAC AACAAATCCACCTAGA CCAAATCCTGCCCCACC TCCACCCAAGTAGCTC GCTTCGCTCGCTCACCT AAAACTTTGCTCGCTCG CTTCGCTCGCTCGTCTT AACCCTTTCCGAATAA ACACTTACAATTCCCGG CTCGCCCCATTTTTT (SEQ ID NO: 1295)	FPPSPPGTDRKPSIIITDTCQL RFVPCMDEPPKGDLGILTLL LRDFSAHNIPIKSLNNKELIA DKDIDYSPDFVEGALANAEE HDTTNSQNNNGRYINSAEK LTEFLIQCEDYLTNIKTLEDL ERFYTTIKDYRVNKEVIAED TPIFVYFLVEEGKLPKPGLRC PLESYEGHEDKAFESLRKLC DHFKGEIAKTSFDPKVHTIDI WVEFLAQAYGTGTFVYKDE NGNIDLDTHVFKCPYADCS YTNNDRSKLMDHMKTKKH AKNVYIERYGFFWGIVIEGV NRPKGIVYPTLKDIKEHACR KCPEAGCNTYVTELSDIKEH LKKKHKSTTAGVDGEIAHT DATYCWITKEELDALHAER ARERAEQVDNTPVQQIINAD NNEENNENQEDNGNNEEAD ALDPPNNTTETEDEAVHAVI INPPATEEEEVAIIAEARRNIP ELQQAEERGCVTPKMTSLV

						RLKLLKGGGELFNKKLTPLA TRYAATGNTEADKIKVDYL TLKCNAALREMIYTNNHSES KFMTAENGEDTAPPPRISED TRDRIQKAANEIKGTLIKVV KHISHARCLKDSTRDDEHN KFVEMIAKIKNDLRDNKFEQ YNIEEIFQGPISDQSILNIVNT EDNNEFIKKMDYINRILGTP QDASPYARKKLQACFADNP TKTLRNIILADKVPQQSLKPS EYLDYYGPQWANEAEGYE NFLHHDYALPERYGQVFAN DFLDFMTNESKIIEVIRNKNH LSAHGLDGIPNSVYMLFPVS AAKFLSILFRSIIISGHIPDCW KLSKTVMLFKKDDPSLAKN WRPIGITSCTYRIFMTLVNK ALQMIPMFHAMQKGFVRGA TLSEHIAVANEVLCQSTRTQ SEMFQTAIDFTNAFGTVPHQ LIFDSLEAKKVPDSIINLLKD LYKGARTAIYTRHAHSEIVP

						VRRGVIQGCPLSPILFNCCLD PLLYAVQRRHFEDGYRFQD KAGQYSIAIQAYADDVLVIS PTHEGMQRILNTVDEFQKIA KLKVAPQKCVTLAKTSTAIQ PFRIGPDEIPIKTSMDNITYLG IPISGTKTSRFAAATGILEKV KAQIRVVFASHLALSQKIIAL RVFILPQLDFYMFHNVFRVN DLKATDQMIRGLIDKEAPTS NIPVSFFYMPKNKGGFGLVK LELRQPQLVLTKFARLWLSQ QAETKAFFHTMAQEEKSFR KVVEDQENGFLGIKMENGK IVQKNERSKRTNCFITQAAK AADKLEVRFKEWDKGGIQV RGVGENATDWYRSKHIGQI SPLIGRVIQQRQYEEFKKDE THSHTFCEPAALAESHDIMK RPQAVPNNLYSAAIALRTNT APTPANMHFHNPEVLANCP LCGCQSCTLFHTLNMCRNR FSLYKWRHNIICDDIYQFIHD



						HYPGVTIKCSARITSDGYQT TGPELDDTVKDLLPDLVVY DEANKMIKIIEVTCPYGTDN NVGNSLDAAYDKKVNKYK SLAEQTERLFNWTTTLSIIVV SSLGVIPLRTKLDALRISPAD HIQLLKRLSMHAIAASACIV FEKVPEFFGMRCRPLPGRVT APNAAIPPNNNENNNDTDH GQENQQATSEEQPTNNGNA QEDNGQGEQINNSTEQTISV DQIIEEDAENNAIEQALDQP DEDEFLN (SEQ ID NO: 1417)
NeSL	NeSL-2_CBre	.	Caenorhabditis brenneri	GACTCGCCTTGGGGA AGGTWTTTCAGGGGK SAATTGCCGMAGGCA AGGCAGCCCCCSMMT AGCTTACAAAGTAAG TACMCATTTTCATTTC TTGTGAATTCTTTAAA CATATTTTTCTTGTTTT TTGATTTCTTTTTTCTC TACCTTCCCCCAATTC	TAAACCGGCTCCTCTGG GAGGAGGTATGTCAGA GGACATTCTCCGTGGG CGGATGGGAGGAGTAG GGTAACGACCCGTCAT TCTGGATGCCTAAACC ACCACAATCTGTCAAG GCAAAGTGCCCCAAAA GCACACGCGTGGATCG GTTTGGATGCCGACTG	MRYHXSNXPAXRTSDNXW RSIXKDVRRPDPSTIEEKSRY NRSIGIPDSLKXRSSAVRSXS SXPPSGPQDVRLXNSPSLDD RRRLVDCETTLGSYREWTD KPMMGKMTYAAVTKRAPP RPQTGGARLSTNLLADEMEI KYRDTNDIRLVIDLPNPHLIK CPLCKSCISARGRGANALKY MKRHIADAHHLNADFVYKC

				TTCCCCTCATCTTGTG	AGCCAGAGGGCAAAGT	SRCQEHEPENVCGAKWIVN
				TATACATCCCCCTCCT	CGAAGGCCGGTAGGCT	HLKRVHGYTLEDAVSTAKP
				CCAACCAATCAATAC	CCCGGCGGGTTGTCCGT	STRQQIANAFNDSAPFIDAR
				ATTGACCTCTCTCTTC	CATAGTCAGTGGTGCG	KTSDVPEKKSREAGLEKFLA
				TGTCAAAAAATCAAT	CCTACACCCAACTGCTA	PTKSEDTREKTPPSTRKSSES
				ACTAGTATATTGTCCC	TGACACACAAGGACAA	SEASIQSTIQETLSESSDTLTV
				TTGTATAGTATTATTT	CCCAAAATAAATAAGC	QEIINISSEDEMDEEPPKRRV
				GACGTCGTCTTTGTAT	CAAGGCGGCGTTAGCT	NVWALIHENGKDAWIDSDL
				TAGGAGTAGGTAACA	TCGAGCTAACAAGCTC	MVIFLESRARGYESCSIIDPL
				AWCTGTGTATGGCTT	CCCGAGAGGATGGTTG	NFICTDMSYLTTIVRRRMEE
				CAAAAAGCATGCACA	CCACAGGGCACCATCC	GYKKIIFPLCANDHWTLVTI
				AACWCCTGTCAAAAA	TGGGGAACGACCCGAT	TGSTATFYDPMGNEPTETVK
				GTAWTTCCCATCMTG	CTTTCGGATGCCCAACC	KMIDELDLEMQLAPSNSPRQ
				TGAATAGCTCAACGA	ACCGCCAATCTGTCAG	RDSWNCGVFVMKMAEAYI
				CWKGAAGMCCAATG	GCAACGTGCCCCAAAA	KDTQWDLTDVDTDVKTFRR
				AT (SEQ ID NO: 1173)	GCACACGTGCGGAGCG	SLLTELKAKFNIFAEDIQTYR
					GTTGGATGCCGACTGA	PPSRKALTRNSQSPVVVCHK
					GCCAGAGGGCAAAGTC	CSRPATPIQDVSRMEVEEAP
					GTAGGCCGGTAGGCTC	VLVPTPEEPPQEWTFVGKNR
					CCGGCGGGTTCTCCGTC	KRGVTSRTPNTSPEAKRPAF
					GTAGTCAGTGGTGTGC	PPVPLKPSANRWHFPEEETE
					CTACACCTAACTGCTAT	KMEVSSADEVKNSTPPKPPK
					GACAAGCGTATAGGAG	IPNLLAMKIASPVPLKRGNPS

					GCCCGGAAAAACAAGC	KKHGKGHMMNTARKGPTK
					CAAGGCGGCGTTAGCT	KEMPKGEPANLIVKIRSWFD
					GAGAGCTAACAAGCTT	EQLKMYKDEGSNLQRLTWL
					CTCGTGGATGGGTGCC	SDSLTAAIGKAFNGNKYIVD
					AGAGGGCACCATCCTG	QIIKRNPPPLVEKGAMSTQT
					GTGGGTGGATGGGGGG	SRKRDEFKPRERMAQEPNEP
					AGCTTGGGAACGTCCC	LRIQYAKNRQKTFFKIIGKQS
					GATCCTTCGGATGCCCA	EQCTINIETVEQHFRKTLKAP
					GACCACCGCAATCTGT	VVSENAIKTVCGSIKKVLMP
					CAAGGCACCGTGCTCC	KTIEDPISSVEVKSILTKVKD
					AAAAGCACACGCGCGG	TSPGTDGVKYSNLRWFDPE
					GTTGGTTTGGATGCCGA	GERLAKLFEECRKHREIPSH
					CTGAGCCAGAGGGCAA	WKEAETILLPKDCSDEEKKK
					AGTCGTAGGCCGGTAG	PENWRPIALMATIYKLYSAV
					GCTCCCGGCGGGCTCTC	WSRRISGVQGVISPCQRGFQ
					CGTCATAGTCAGTGGT	SLDGCNESIGILRMCIDTASV
					GTGCCTTCACCCAACTG	LNRNLSCSWLDLTNAFGSVP
					CTATGACATGCGTACA	HELIRRSLESFGYPQSVIQIV
					GGAGGCCCGGAAAAAT	TDMYKGATMKVKTADQKT
					AAGCCAAGGCGGCGTT	QSIKIEAGVKQGDPISPTLFNI
					AGCATAGGGCTAACAA	CLEGIIRMHQMREKGYDCV
					GCTTCTCGTGGATGGGT	GHKVRCLAFADDLAILTNN
					GCCAGAGGGCACCATC	KDEMQEVIDKLDADCRSVS

					CTGGTGGGTGGATGGG	LIFKPRKCASLTIVRGAVDK
					GGGAGCTTGGGAACGT	YAKIRINGDAIRTMADRDTY
					CCCGATCGTTCGGATGC	RYLGVKTGVGGRASETEALI
					CCAACCACCGCAATCT	QVVKELQKVHETDLAPHQK
					GCCAGGCAACGTGCTT	LDILKTFLLPRLQHLYRNAT
					CGGAWGGTCATTGGTT	PKLSELREFENVVMKSVKR
					CTAGACTTGTAATAGA	YHNIPIKGSPVEYVQIPVKK
					CCATTGGCCGGAAGAG	GGLGVLSPRLTCLITFLTSTL
					CACACGCGCGGTTGGT	CKLWSDDPFISSIHKDALSRI
					TGGATGCCGACCGAGC	TVKAMGLTTQSATIKETCEY
					CTAGAGGGTGCAAACC	LNTRKAVTKGGYSLFCRMN
					TGAAGGGCGAGGTCGA	ESLRTLSVIQGAPLKSMEFIP
					AGGCCGTGAGGCTCCC	VNNEIGIAVQATKDSEIKVF
					GGCGGGAAACTCCGTC	TKADSLKLMSKLKDLVRSA
					ATAGTTAGTGGTGTGCC	MLKRFLEEKSVKSRVTQVL
					TACACCCGACGACTAT	QHHPQSNRFVRDGRNCSIAA
					GACACATAGGAGGAAT	QRFVHPARLNLLSCNANTY
					CCTGATCTGATATGATC	DVNHPKGCRRCQADFESQQ
					ATGTATATAGGGAGGG	HILQNCHYSLAGGITQRHDR
					CGAAGGTAAATAGTCA	VMNRILQEIGNGRKAHYKI
					GKGTCAAAGTCCACGT	MVDMETGATRERPDIIMEER
					GGCAGCTACTCCCCAG	DGPEVLLADVTVPYENGVQ
					CATAGTAGTGATGCGA	AVERAWDKKIEKYKHFLDY

					GTGGAWCCAACTTTGA CACTGATGTTCCCTGAG CCTGACCCATCTGCACA AATCCAACAGTGTATG ATGGCCCACACACTGA GGACGAGTATCACTTG TGATACTCAGAGGTGT CCCCCATGATCAACCA ATATCACAGCTAGCGG ACCTACCGTGAGGTAG ACCCCCGCCGCTGTAG CAGGCTCGCCTC (SEQ ID NO: 1296)	YRKIGKKATILPLVVGSLGT YWPDTSHSLKMLGLSDGQI RNVIPEICQIALESSKNIYWK HILGDSYKTVEGLFCQRNNK EVRFEGKGEKHHVSQRFQP LKCEKVRTMKSTKEEGRSR SNAKKGPNWRRSKSESDGR SVSKGRYWRDPSNKPPHSK MTQSALAKR (SEQ ID NO: 1418)
NeSL	NeSL-2_CRem	.	Caenorhabditis remanei	CCAACTCTCATCGTAT TAACCTACGGTATTCA CTCCTAGTGAGTGTAA TAAAGGTTAATTACGT TTTCTCTTGCMAGAGA AAAAGAAAATTCGAA TCCTTTTTGTGTAACT CACAAACTGACAGAG ACCTATCGAATTTCCT TTGTTTCGTATATAGG	TGAATACCGTCAGATA AGCCCCCAACATAAAA ATAAAAGTCGGCGTTA GCTAACCACTAAACCG GCTCCTCATTGGGGGA GAGTATCATTCCGGTGC TCTCCGTTTGGGCGGTA GGGAGGAGTTGGGTAG CGACCCGGAAGTATGG ATGCCCAACCACCGCA	MTNVYLKPVNDNQTNKTG DNSRNTMSNSQCEMTWKPV ARTYAQAASTNPADDKTVT VLGCKYNLLKLGNTPQTSK RSPPKPSRGGARISSVYTLTD ELEITHREEGKITFAIDLPNK NNILCPLCRECTQTRGRGSS FTKHMKLHVKEKHQLDATF IYKCSMCNEYEPEKKCGTK WIQTHLQKVHNYKYDESAI

				AATAGTCACTCTGGAC	ATCTGATCTGGCATTGT	VVPVPPNTRQQIANELNNAA
				CACGAAGTGGACAGT	GTTTCGGATGGTCTCTG	PFVDIRKPKAAAVEEKKTEN
				TGTCGGCGGACTTCCA	TCTCTAGATCTGAAATA	GALLKFLTKSNKDDQVKSPS
				GAGTGGAGAGAAAAG	GAGCTCTGGCCTGAAG	XDIPDAESPEKETQALTIDPK
				GTGTGAAGAGAGGAG	AACACACGCGCGGACC	GNNSPSKSSIRSSQSSASSVC
				GTCTAGAAACACTTCG	GGTTGGATGCCGACTC	QEIQEIITLSEDEDPKGARPK
				GCTGTCTAGGACCAGT	GATCTGGAGGGTGCAA	PGINVWSLINETGKDAYIDT
				TCCTGAGTGGAAAGA	ACCTGAAAGGGAAAGT	DIMMAFLKMRVENCDSVNII
				GGAAGGTCTAGAAAC	TGAAGGCCGTGAGGCT	DPLNYQFPARVDLVPLIQRN
				ACTTCGGCTGTCTAGG	CCTGGCGGGAAACTCC	LEDGKKRVVFPICADEHWT
				ACCAGTTCGTGAGATC	GTCATAGTCAGTGGTGT	LLTISNGIAAFYDPTGSRMSS
				TCTCGTGGAGAGTTGA	GCCAACACCCGACGAC	YIEELVNELGLIIPKEQDEQP
				AAACAGTCAGCTGAG	TATGACATAGTTGGAG	RQRDSYNCGVFVMKMAEA
				GCTACTGTATTTCTTG	GAATCCTGATCTGATA	FIQDTEWEMEEVEEDVKNF
				ATAGCCCCGCCCCCA	ATAATCATTGTTCATAT	RRNLLEELKPNYEIFAEKIKY
				ATCCCCCTCCCCCCCC	AAGGGAGGGGGATGGT	YNSPGKSFAQSRPTSRSSQC
				CCTCGACAGATTTTTC	AAATACCCAGGGTCCG	AVCPTCSRSATPMMDVGNM
				TGTTTGACCTCCTGGA	AAACCATCAAAGCAGC	EVDPVPQQQETPKSREPEQD
				ATTTGCGAGGAGTGC	TACTGACCAGCATAGT	EGWKVVGKARKRGVVTER
				GCGAGAATTTTCGAAT	AGTGATGAACACATAG	SPNISPEAKRQFTGPEIKVVS
				TCTTCGCGCGTTTTCT	ACCCTGGGGTTCCCTGA	PGKFHPLVGETEEMEVTCDS
				CGAAATTTTCCAGAA	ACTCGACCCATCTGCAC	PPTKEPTTEPKVTPSLPAMKI
				GATTCGAGCGGAGAA	AAACCCACTTTGTACA	ASPEVTKKQTSKKKGKYGK

				TCTTCGAGAAAGTGA GCTGAATTTCGCGCGA ATTTTCCGCGATTTTC AAATTATCGATTTTTG TCGGAAAATTTATTTT CTGGCAAAATTTGATT GAGTTCACGCGGGAG AGAAGGAATTGTTGG AAAAGGGTATTGATTT TTGTGGCGGAGGAAA CTCCCACTGAATCAAT AACTCTCAAAGGAGA ACTCATCGAACAACCT CGGGTGACCTGAATCT TGGGCGAAATTTTCGC ATTGACACAAGATAA MACAAATTACTGTKG AAAATAAATCAGAAC AAACTGTCAAAAAGA GAGACAAAAAGTATT GATTAACAACATC (SEQ ID NO: 1174)	AATGAACCAAACTGAT GAAGAGTTTAATGATTT CTTACATCACAGCTAGC GGACCTACCGTGAGGT AGACTCCCGCCGCTGT AGCAGGCTCGCCATTG (SEQ ID NO: 1297)	KKQXTKKAQPPKGEPTKKA QPKGEPAKLIEQVRTWFDK QMKSYQEQGSNIQTLTWIA DSLTAAIFKANSGNKYLVD KITARCPPPLLNEGEMATQT SRRTEAVKPKDRFVKESNEP LRIQYAKNRAKTFNVIIGKH SARCEIDINVVENHFRQTLK AQPVTEEALNTVCSGIKKAK VDPSIEGPISSGEVKAILAKIK DTSPGTDGVKYSDLKWFDP EGERLALLFDECRQHGKIPS HWKEAETVLLPKDCTEEER KKPENWRPISLMATVYKLY SSVWNRRISSVKGVISDCQR GFQAIDGCNESIGILRMCIDT ATVLNRNLSCSWLDLTNAF GSVPHELIRRSLAAFGYPES VINIISDMYNGSSMRVKTAE QKTQNIMIEAGVKQGDPISP TLFNICLEGIIRRHQTRKTGY NCVGNDVRCLAFADDLAIL TNNQDEMQDVLNQLDKDC

						RSVALIFKPKKCASLTIKKGS VDQYARIKIHGMPIRTMSDG DTYKYLGVQTGNGGRASES ESLTQIAAELQMVHDTDLAP NQKLDVLKAFILPRLQHMY RNATPKLTELKEFENTVMKS VKMYHNIPIKGSPLEYVQIP VKNGGLGVMSPRFTCLITFL ASTLFKLWSDDEYISSIHKK ALSRITAKVMGLKTQKATL QEQCEYLNTKKAITKGGYSL FSRMNEAIRTLSVNLGAPLK SMQFIPENGEIALEVQASENS QIKVFSKADSMKLVTKLKD LVKSAMLKNFLENKKVKSK VVQVLQHHPQSNKFVNDGK NXSISSQKFVHPARLSQLVC NGNSYSKDLPKNCRWCGYE CESQAHILQHCTYSLSSGITQ RHDRVLNRILXEVIKGRKNN DYYDIMVDTEPGPTRERPDII MIQKDGPEVLLADVTVPYE NGVVAIEAAWDWKMEKYS

						HFIDYFARLGKRAVILPLVV GSLGTYWPDTSNSLRMLGL SDGQIRNLIPDISMIALESSK QIYWRHIFGDSYRIVSDLYC RKDQQEIRFGDEPMENVQV SDRFQPFKTREREKKSEEEK KRRSKSKKGKTWRGSKKQT DSRQSGKSNQNQGFQRSVG QGVSR (SEQ ID NO: 1419)
NeSL	NeSL-4_CRem	chrUn	Caenorhabditis remanei	CCCTTTTCTATCGTAT TAACTACGATAACCG CTCATTTGAGTGTAAA AAAGGTTCCCCCCTCC TCGCCTGCCTTACCCA CGCATCTCTGCCTCTG GGAAGGCGGAGGGTC AACTTGCGGGTCTGTG GATTTCCTTTCCTATC CACCGCCCATATTCTC TGTCGAAAGCCTACCT AGATCAGCCGGGAGT TTTTCCTATCCCATTC AGGCGATCGCTCAAG	TAACATGCCTTGGAAG GCACCACGCCAAAAGT CCTGGCAACTGATTTGA ATAATGTATAAAAGTA ACTGGAACCAAATGCC CGATAGGTAGGGCGGG AGAAAATGACCTAGAA AACACAAAGTCCCAAG CCCCCGGATTCGAAAG ACCTATAGGAAGTCAG TGAATAGAGAGAAATA TCAAACAAATCTCACC CATTCACAAGGACTTA CTGGTCGAGTAGAAAA	MTKTEWSWRHRSRSRSVGI VVKIDTSDYANVRVHVAAD LSNEDGHTSHNNGIILPIPMK PSVDRFCQIQYPPRGYYVPH PQSQKGHDAKPSRHWNEEA QPPYYHNNNHGRRGRSAKP SGRRPPRKPILQEESLAAHPQ IPGDTASAVPLYSDVVNNEN KSQGKPPQGSHRRSGRPGTK PSVPVGEAEQETNSRPIAPEP IVKFKHDKHGWTTVQGSHS SGRPVPKPSVPVVSEANRFQ LLQEGDFPPLTTSESSQEEIK VPNYQRIVSPIPLPSEEDSKL

				GCTGTTTTATCGACAC	CAAGCCAAAACATCAA	PTKSNYRAPKGRKSRNYKK
				TCCTTCTTGACAAGTA	GCACGACGCAAAAAGG	PQQQNPKKYQQRLPYQPKV
				TTTATTTCTTGACAAA	GGTAACTTTGGGCAAC	NNAPTDRMAPEQLKGGGGK
				TTCTATTTTTCCTTTTA	TAATTAACGGATACCTC	TAHNDIEEMEIEEDTDEKIIQ
				TCGATTTTCTCTTATTT	CGTGTATCAGGCAAAG	VKRIKIVNKLTPHHFVCMM
				ATCGATTCTTGTGAAA	CCGCCACCAACAGCAA	TYPTDNIYRCFVKGCTATSQ
				AT (SEQ ID NO: 1175)	ATTACTGCCCGATAGGT	GGWGAEDLKYLTVHIRQEH
					AGGGCGTGAGAAAATG	KIKVEWTYECGICGDLSGGA
					ACCTACAACCTCCAAG	GKHISKWIKPHMRKKHNRD
					ACCCGAGCCCACGGAA	APTNFKMGSRSSGKPKITEL
					TCGAAAGACCTATAGG	LEESAPSCSNPRRKTLNQKK
					AAGTCAGTGAATTGAT	TAIITQVTPEKLKTGYQTRS
					GGAAATACAAAACCAA	VTKALSVLKESRQKELEVLR
					ATTTCTTCCATTCACAA	EEEKANAKQKSKLHPFFTK
					GGACTTACTGGTCGAG	APHIDGVKPTVRRELSKMIT
					TAGAGCACAAGCCAAA	PGGEHKGTKIPMVHTKRGLI
					ATATCAAGTATGACGC	QKINRKAKKAKPMHLDESTI
					AAAAATGGGTAACCTT	IEASQLDVITIDDDDEDDNM
					GGGCATCCAATCAACG	TPMRRRFNTWCLDHETTQE
					GATACCTCTGCGTATCA	AWLTDDVINWYLKDLCFGN
					GGCAAAGTCGCCACCA	EQYMLVDPLVWLIYKMGG
					AACTGTACTACTCCGA	MAGVEQRFKSKKTCLFPICE
					AAAAACCAAGAAACAT	ADHWILLVFDETNLCYANS

					GATTTTCCCACTCCGTT	LGSQPNGQVKNFIQQLNRKL
					AAAGCATCTCAACCAA	CSFEKEVPLQKDSVNCGVH
					GCTAAAGCGGTAAGGT	VCLIAKSIVNGQFWYDDSD
					TATCATGTCAAAAGGT	VRTFRTNAKAALKAQGYEL
					GTAGCTACAGCAACCT	FSEAPKQIENPDSSHREDIKE
					AAAGCCCGAAAGGTAG	NSMEMCSESLMIVATPQRSE
					GGCCGTATAAAAAGAC	APMELVDTEPSDLESPKSDR
					CTACACCCTCCAAGAC	VVYEDCITALSDVSEPRMTP
					CTAAACCCACGAACTC	EKSETPEVPVVEERDLDWPK
					GAACGACCTACAGGAA	LESPKSDRVVYEDCITDLSD
					GTCCGTGAATGGAGAG	VSEQRMTPEKCETPEAPLVV
					AAATATCTCACCAAAT	ECVELERLPKDLPVTDRSTV
					CTCTTCCATTCACAAAG	VAIPEAVKLEEKSEVVIPRL
					GCTAACTGGTCAAGTA	MELSYTVPPEPSPVVEYTQP
					GAGCACAAGCTAAGCC	YTHTHTKPKVKATCQMGK
					TCCAAGCACGAAGTGA	KRKVPTGKPDELIQIVRQWF
					TATGGGTAATTTAGGC	EKEFNDYVTEGRNFQRLEW
					AACCAATCAACGGATA	LTNLLTAAIQKASAGDEETI
					CCTCCGTGTATCAGGCA	EKIRKRCPPPEVRENEMSTQ
					AAGTCGCCACAAACAC	TSQRQKPTTTNQKKRSRNTT
					TGTACTACTCCGTTACT	QSDTQANTYWRNRAKTYN
					CCCAAACACATGGATC	QIIGQDFKQCDIPIAILEEFYK
					TCCTTCTCTCACCAAAA	KTTSVTNVPQETLVKVTSRL

					AGCTTTATAACCAAGCT	PRLDIGKWIEDPFTEQEVFG
					AACGGTGGAAAGGACA	ALKKTKDTAPGTDGLRYYH
					TCATGTCACGAGGAGT	LQWFDPDCKMLSSIYNECQ
					AGCTACAGTAACCTCTC	HHLKIPAQWKEAETILLFKS
					TCTTGAGACTGCAAAG	GDESKPDNWRPISLMPTIYK
					TCGAGGATGGATTGGG	LYSSLWNRRIRTVKGIMSKC
					AAGGCCGCGAGGCAAA	QRGFQEREGCNESIGILRSAI
					AGGCGGGTAACTCGGC	DVAKGKRSHLSVAWLDLTN
					CAGACGCTAGTGATCTT	AFGSVPHELIESTLSAYGFPE
					CGGATCCGACAGCCCT	MVVHIVKDMYKDASIRVKN
					GGCCTTAGAGGAACCC	RTEKSEQIMIKSGVKQGDPIS
					TGGGATAAGGAGCACG	PTLFNMCLETVIRRHLKESS
					ACGGGAAGGATGTTCC	GHKCIDTRIKLLAFADDMA
					GCAAGGATTTCCCTTCC	VLAESKEQLQKELTEMDED
					CATTAGTCAGGGCTGG	CTPLNLIFKPAKCASLIIEFG
					CAGTTGGTAATATAGC	KVRTHEQIMLKREPIRNLND
					CTTTCTACACACCACCG	DGTYKYLGVHTGADARTSE
					TCTTGCACCCACTAAAC	EELIISVTKEVDLVNRSALTP
					CAGTGGGATATGCGGG	PQKLDCLKTFTLPKMTYMY
					TGGACTCAATGTAGAA	ANAIPKLTELSAFANMVMR
					AGGTGTTCCCACTGCCT	GVKIIHYIPVRGSPLEYIQIPT
					GACTCGCCAACTTTATA	GKGGLGVPCPRITALITFLVS
					TGTCTTGTCAACATAAT	TMKKLWSDDEYIRKLYNSY

					GGCCCCTCACTATAAA CTCCCTAGCAACTGGTG GTCCGGCGAAGCCGGT TCTTGCCACTATTGCGC CCCAGGCTCGCC (SEQ ID NO: 1298)	LKKVVEAETGIVEVSTKDLA EYLSNKVPSRKHEFGYNCYS RIREVCNGLALNQAAPLYKL EFIEQDNELAVVVQPTEESK ERIFTKDHVKKLQSLLKASV NDALLHRFLTTKPVKSEVV QVLQQHPQSNSFVRMGGKV SISVHVWIHRSRLNQLTCNY NIFDPKQPKNCRRCGYKNET QWHILQDCTYGWAKLIRER HDAVHHKVVTMICAGAKK NWGRKIDQELPGFTSLRPDI CLTSPDGKEVIFADVCVPYS RTRNIEFAWKEKIRKYTEGY SHLVAQGIKVTVLPIAIGSLG TWWTPTNESLYQLGISKSDI RSAIPLLCSTVMEYSKNAY WNHIYGNSYTSVPLRYGHQ KPDGDDWKKELSCEPVLAL QQ (SEQ ID NO: 1420)
NeSL	NeSL-4_SM	.	Schmidtea mediterranea	TTAAATCATTTTTAAA TGTGTTTGAATATCTT AAATTATCAAATCATA	TGAGTGTGCTACGAGG CAGCGCTGGTAATTGC ATCGGCGTTGCAGATTT	MNVDLDATIKSIGMNTKETT YPNSQLRVETTPCTSTTIMH ASCNTTSTISYSPLPSAVSLP

				TTAATATCAATGCTAA	GTGTACGATAGATAAA	ESPASSITITTTDDNCDIIETP
				AAAAAAATCGTGCKC	AACCAATAGTAATAAA	YPLPQTNGDLSEILKDIEAN
				ATCAGGCGCACGAAA	TGCTGAGCCTAGCTCGC	KDTTMSNKVLDCDSDSGDD
				ATAATGGACACAACT	ATATCTAAGCCGAAAG	RDMIIENDRESDMDLFSQSL
				CGTCGACCTGCTGTCG	GCAGCATATATATGAG	LNTNQSDERREKNLTENAPT
				ACTCACAGAGAACCT	ACAATTTAAAAAAAAA	EITTEKSYFDIISKASDNTTS
				CAATTTGGAAGAATG	(SEQ ID NO: 1299)	KKLLNVKNELTAGLPPMPP
				GGAAGCCTATAATGC		VTNTAKFIRNVRPEDIADPT
				TACAATTCCGCCAACC		LYRLDSRGKLGCRTQYKKP
				CCTATTTGAATGACAG		GCGDIAVYDYEAIVEHAAFI
				ATAGTCAAATATCAA		HTIPFNEQNNVDCQPCHPKK
				AAAATATACAAACTG		GKDVHTIVLIKYADIFNHIEA
				CTGTCAAGCGTGACTC		HSHVVQTAITDNMKTYLRL
				ACTTCCTTCCAATCGA		TKENXFYCSYRNNKKKNKC
				AAAATAGGAAKATGT		KKAFNLESNMMDITEHMKT
				AAGAAACATGAAAGT		HTGYSFDXNLNILCYCGIWK
				CAAGCTGAAAAACCA		PFTELIAHIKTEHLQEYINSIP
				ATAATATGTCCTAAAA		NKENIHNTTTIVSPLNFAGIL
				TAAAACAATTTGAAA		ASGETQNIPDEEIIKPRDLPE
				ATATGCAAAAAATAC		NLAFNRNIENELSWSQHLVK
				CTATAAAATCACAGC		AYIFSYAVKTSTIFINPYTCN
				CGAATAAATTCCCATC		ALIQCNYKTFFETFPFKDFA
				CGTTCTAAGCAGAAA		KWNEIVLPIHNNTSSWSFFF

				CCGCTACGAACTACTG		LNKKKRVAMIIDPTADDSHT
				CAAGAATCGGATCAA		LHFELATDILRTILNVQNIFE
				GTATATTAATTTCCCC		DLNFPLTEVEYPVCHEANLS
				CCMGGGGGAAATTAA		AFXVCHFLKCLMSDLPIDIP
				TATACTTGTTAKAAAA		DIDHMKETMRPIIRKYNCAK
				TTAATTTTTTAATAAA		FPESDVRNYRVLIEDLIYQL
				AATAAATAAATCGAA		NLDTITCEEILCEIERINGRLN
				TAAATATAAAATAAA		PKRYFKESKPKTDIIHLQKK
				AATAAATCAAATTAA		KSAELLCVKRLKFQISQKTEI
				ACTTTTATTAACAATA		GKIWENDDVDHRPPMARFL
				AAATCGCAGTAAGTA		KTFASQDCPVSNTSSINLPY
				AATTTCCACTGTTATT		YMDTDTDXCTDCENLSHIM
				AAATTTAAAACAAAA		KNLDSSAPGMDLITGGDWK
				TTCCTTTAAAAATGCC		KISPKHELITAICNCILRNKV
				TCTCTTTTTCAGTAAT		CPEKWKLFRTVLILKPGKMS
				AACACCTTTTCTTGCT		ESFRANSWRPLAIMDTAYRI
				TTTATTACTATTTCTT		FTTLLNNRLLQWIRNGNLIS
				GTGTACTGTACAAATC		PNQKAIGIPDGCAEHNATLH
				GAGCACAGTTATTGC		FAIDRAKRCKTELHIVWLDI
				AAATAGGACATAGAA		ADXFGSLPHDLIWYTLANM
				ATTCCTTTTTAAGTAA		GLKNETLTLIKELYKDVKTI
				ATTTAAATCCATGAGA		FDCQGTLSEPVPITKGVKQG
				AATAAAATAAAATCC		CPLSMTLFCLSIDYILKSILTN

				TTTTGATTCAAAGTTT CTATGTTGCTTTCTAA TAGAATGGTGTAAGC ATTAATGGGTCTTGAT TTTTATAAATTAAATA TATTTAATCTATTAAA TTAATATGTTTTTATT AATTATTAATTTTTAT AGTGGGGGGAAATTA ATATACTTGATCCCAA GAATCAACTGATGAT GAAGAATATGTTATTT CAAAATACATACAAG AAGCTGGAAAAAACA AATCAATCGCTACA (SEQ ID NO: 1176)		YPFLLHDLNISILAYADDLV LLSDSYLEIKKSLESTVELAA FANLKFKPSKSGYLSINNVN SDILKLHLYNEEIPTISENNK YRYLGVDFSYKRNQDVDGR LGSALALTRSLFKSYLHPAQ KLNAYKTFIHSKLIFSLRNCV IGHRILDCDRNRVTQGREKQ LGFDQEIKALLKTMIGDKFQ AXNNYFPYTHCKLGGLGITS AIDEYLIQSITGITRLFHSSNL SFRKMLITELAHSRGGKNFE AGLKWLNCEVNKAFPNTSF FVKFQKSALALKRKFCICVN LKFVEDNFSLEMTYKKRTS YVNHQNLSTLSKELHDFVG LYYAEQXCQMRVQGHIATA IGDSITAKYLIASDILNDAQY YFLVRARNNLLNLNYNAYR LKYNIGTKCRLCHLDEETQA HXFNHCRAKPNARRVKHEN VLVSIVAFLEKIGFEIDVEKS PKYISIPTKLKPDMVIRSKRN

						KDIHVLDLKVPYDSGEGFEK AREDNYVKYKDLAIEIGKAF NQKATISAVVIGCLGTWDK KNNAALSKIGLTKTEIISLAR IACPNAVIACYHIYREHVSFT KSAMALPFSLA (SEQ ID NO: 1421)
NeSL	R5	AY216701	Girardia tigrina	GTAGGTAACTATGACT GCAAAATAATAATTCT ACACCTATTGTTGATA ACTCATCTCGTGCGCA AACGGAGCATGTTATT TCTAATCATTTCGTCA CACAGGATTCTTCTAA TTCTGATAGTAATATT ATAGATAGAGATAGG AACCTTGTTGATTTAG ATGCGTCAATAACTTC TCCTACTATTATACAG CCAGAGGATAGTAAG ATATCTGAGGATGAG GACTTCATCTTAGTCA ATAGGAAAAAGAGCA	TGATCCGTGTGTTTGTG TCGTATGATTGTTTCCG TGTGTGTCTATATTTTT CTTTTTTATACTTTCAA TTACCTCGTTGTAATGT TATAACTTCATATGGAA TATATGTAATTTAGTTT AGTTTAGTTAGTTTAGT TTAGTTTAGTTTAGTTT AGTTAGTTTAGTTAGTT TAGTTAGT (SEQ ID NO: 1300)	TTGRNLGQWSCYSRSIQQSN YSFKLSSTEVGELVEQSPAP LQSPQFSNNYNNLNINNNLY YSLNTFNQSNNLCCLVNIEF FPTQHLLGDIVNSGCINYMN NYNNFDNINLYINSNVLSYN NYNHSFLASPYTTNITEHADI NMHVQEVNMQQDNNTQHA ITQQVSLQATSLQHTLDEMI VQFNTAVRLKKKHKVAKIF RGHNHRKDLPTLPAREQYK TKPKLAIREVLHRKTTATSS PSENAIKAFFSSYSRPAELFT GQELLESSWFPVHPEDDFEF RIPGRDQIAKYIKFASKSAA GLDWITYEDIKLGDPSGEIL

				AAAATAAGAAAAAAT		QPIFEYIVQNNICPSEGKASR
				CTAAGAAAACAACTG		TIMIPKPGKSDYSDPSSWRPI
				AAAATAAAAATGAAA		TITSAVYRLLMKYLTWELY
				TTCCTATTCAAAAGAG		NWILLNQMLSRSQKSLGKFE
				TAAAGATAAGAAAAA		GCHDHNAMLNMLIQDVRR
				GAAGTCTAAAATTAA		QTNPSNPINKNKRLYIVFLDF
				TACCGAAAAACTAAC		TNAFGSVPLDTLMYVPQRF
				TGAAAATATTACTACT		GLGTSALTLIKNLYLDNYTN
				TCTGAAATACCACTTG		VTCGESKIENVKLNKGVKQ
				AAATTGCTCCTTCCAT		GCPLSMLLFNIFINIIIRAIEA
				ACCTTTACCTTCAGCA		MPDVHGYPLGDMDIRILAY
				AGTACCTCGGGTTCTC		ADDIALISDSHKDLQEMVYK
				AACAACCGGCCAATC		AEYIGRILGLLFNPSKCALM
				CTCCAGAAGACGCTA		DIPHDKKRTPPILVNGEMIK
				CTCTAAGTGATACGG		CVGKADPYKYLGTFRSWFR
				ATCTCTTCCTTACACA		KLDIKELLQMMMDETKLITE
				GGATGATCCCGATAG		SNLHPHQKIHAYETFIHSQLP
				TCTTATTCTTTCTGGA		FHLRHSRIPFSDFITNRKTNK
				AGTACTCAACCAACCT		TTNNSNDSEKSIQKAYDPES
				TTGTTGACCTCAACCC		GQLFLNTFALPSGCAKDFFY
				TTCACAGCAATCGGA		ITKDAGGPQLTSGLDEYLIQ
				ACTTCCTTCAAATACT		SIMYIFRLLGSEDPTLNSAIK
				GACAGCCAAAGATTT		HDLISHLNLKGFVNINFSQAI

				GAGGCGGGTGAAACA CCCAAAATCATAACTT CTTACAGGGATGACCT TTTCTACTCTACAGTC CTTCACTACAACTCAG ATACAGGTTACGGTAT AAGTGTTGACAATGG GGAGCAGAGGTTTCG AATTCTTGCTAGGAAT CTTGTCAGGAAAACC AAGGATAAGTTCCCCT CTTTATATGCTGGACA AGTAATTAGACACAC AGTCTTCTTCAATCAC TTCAACCAGGCATACT ACGCCAATAATATAA CTGATAGTAAAGGTA ATCTAATTGAGTTTTC TGATGATAAGCCTTTT CAAAGTATACCGACT GACCCAAAAACTGAA CTAGAGCAAATTAGG AGAGAGAGACAACAT		SIFNSNFTDRTDHFSHLSRTE WARLQLARKKLKSTLAIQT NVCLINGHLVLTLSLENNVL LIDSKEKGDVKKIHASLMGF LRLAHLIRLQKHGWSKLLFS ATTHHEILNKRILNGHVPYK IWYFIHRARLGLLPTKLFSVS NLCRKCGGKKETMSHALVN CPMMQTLINERHDALEISLV QILSSKFQGTVIRQKTYVNE LRPDITMESDTQYYLVEVKC PFDTKMSFELRTQQTTDKY NIIIEILEDVHPGKEVRLVTFI VGTLGSWGPQNSDFLRDLG FSKDEIDQVKTRLMLQNINS SCEQWKRFVQYAPTITPGPIP DAESEDDQGTSDNGPTAAT VQGPVIGDEEEELQIYDSGL DESSDDEPDPDDAELLFTIDI EQYLNSVITD (SEQ ID NO: 1422)

				CTAGTTGATAGAGCTC TTAGACATAATCAGTT ACGGGAAACTTATATT TTAAATAAACTTAATA ATAATAATGGGGGGG GTGGCGAACATTTGA AAAGGAAAAAGATCA AAGTCAATACGGATG ATGTCTCCAGCAATGA TGGAGACAGAAAACA TAG (SEQ ID NO: 1177)		
NeSL	Utopia-1B_CPB	.	Chrysemys picta bellii	GTTTAATTCCTTCTGA TGGACATCTGCAACA CCCGTCCTGAAG (SEQ ID NO: 1178)	TAACCGAGACCGCCGA CCAGGGAAATAACCCA CTTCCTTCCCTGACGAA CCAAGGGACGCACCCC ACCCATGTACTTATTCG CTACACCGATACTGACT TGGACTCCTTATACATT CCATGGGTGGCGTACC CGAGCCCACTTATCCAC TGACACTTTAAAAACTC TTGCACCCCAATCTGGG TCTATGCCGGTTATGCG	MESPAXIFEKIDAALXIYSAA AXLXXNSLSLSPXXAXXSX XAAPASSTPQKTQXKPIPXT TLGASRKXRTTXKDEXIXX WXKKAPVDTSXGRXSTRRT ALRDLTSRSXNIXXALQEED PRRTPPXSRDQDAERRPAAP EKAATRGAPPTIQDQDADR CPAGRDATGGAPRRPRTRM LXAAPLGRMPPEEPPPTTRD QDADRRPAAPERDAPEGTTS STPDPETTYHPPVRRRAAPR

					ATATGTATGTATCTCTT CATCCTTGCAACCGATA CCTGTAATCCCTCATAA CCCAAGCCTGACCCCA GATGTACAGTACCTTCC CTCTTAACTCGTGTATA TTTAATTTTAAACATTA ACTTTAATAAAATTTTT AAA (SEQ ID NO: 1301)	GTHSXAXDLDAARCPSGQR DIVASESSTPPGATSPPQASL PDXEESPAESAGTTEVRPTE GEAGEDDCIYLQYPXPTGLL LCPFCLPXHGVQTLGALSKH VRKAHNKRIAFRCSRCDAPF ETQKKCKXHXATCKGPLTT AKVNPTDTLRVPTPTPTDGP ASAPQPASPEPQXVRGDQPP TEGSATPASRTDDATKRTSP ASRIPTLDPAVRGITATSQVS DLTRCLSDLIKTIRHNTDTRR XSAPPQVTSCRPAVGATSTA PQAARRDPANGGASRSPQIP RPDPAPGRPNTSSKVTQRDS DRQKPHAPPRTPQPDTTRRR TRTIPSASKHDRAPTKPNTG VSRTPLPPGRSSAASETPRA APPPPHQDPRLKTHLNTAPQ SEGQQGHRLSPQHLNPRRQ RSRRNDGREQRVATPWQSA WMEELAKAEDFETFDTLMD RLTAELSAEITARRREPQEAS

						RATRRFPAPTRNNTAREGRR GDVGRRYDPAAASRIQKLY RTNRTKAMREILDGTSSYCA IQPERLYSYFKDVFDHEAQT NLQRPECLLPLPRINLTEDLE RDFSPQEVQARLMRTKNTA PGKDGIRYHLLKKRDPGCL VLAAIFTKCKQFHRVPRSW KKSMTVLIHKKGERDDPGN WRPISLCSTIYKLYASCLAA RITDWSVCGGAVSSVQKGF MSCEGCYEHNFLLQTAIQEA RRSKRQCAVAWLDLTNAFG SIPHHHIFATLGEFGMPETFI QILRDLYKDCTTTIRATDGE TDAIPIRRGVKQGCPLSPIIFN LAMEPLIRAISSGPTGFDLHG KKISILAYADDLALVADSSE SLQQMLDVTSQAAEWMGL RFNPKKCASLHVDGGARAL VRPSRFLIQGEPMASLEEGE VYQHLGTPTGVRVRQTPED TIAEILRDAAQIDSSLLAPWQ

						KINALNTFLIPRISFVLRGSA VAKVPLNKADSTIRQLVKK WLYLPQRASTDIIYISHRQG GANVPRMGDLCDVAVMTH AFRLLTCPDPTVRSIAQEAV RDVVRKRIARAPSEQDIATY LSGSLEAEFGREGGDLSSLW SRARNASRRLGKRIGCCWK WCEERRELGILVPRIKTPDH TIVTPTARAMLERTLKDAIR CHYAENLKRKPDQGKVFEV SSKWDASNHFLPGGSFTRFA DWRFVHRARLNCVPLNGAI RHGNRDKRCRKCGYANETL PHVLCGCKQHSGAWRHRH NAIQNRLVKAIPPSLGKITLD SAIPGTDSRLRPDIVVTDAEK KKVLMVDVTVPFENRSPAF HEARARKALKYTPLAETLR AQGYEVQIHALIVGALGSW DPHNEPVLRACGVGRRYAR LMRQLMVSDTIRWSRDIYTE

						HITGHRQYHTE (SEQ ID NO: 1423)
NeSL	Utopia-1_ACa	.	Acanthamoeba castellanii	CCCGTCAAGGGTGCTC CACGAGATCCCTGTCG CTAGCCGACCGGTTTT ACCACCCCACCCCGCC CGGACAACCACGGAC CCTGCTCCGCAGCAG GACCCCACGCACG (SEQ ID NO: 1179)	TAACAACCATGTATGG TGAACCACACCTCTCTC GATCTTGTATTCTGTGA TTGGACATCAGAGTTCC TGCGAAGGGATACACT CTGCCAATCTCGTGGGT TGTAATAAATCCACAC CTTCAACA (SEQ ID NO: 1302)	MAAKSVACPHDGCANKYA SEASLRRHIKNKHATDEEGD ETSHSCPHCHRPFSTARGLS VHIGKSHRQAPPEPTRPPPAP APADPGLDPDPGPTVTPPSR DDEDREEPDDDPVEIADLSC PHCAQALPSAHGLANHLRA CKDHRVPAPGAPRSGPPSSR YWTAVEHHRYVEAMARFA DHPDLLARAAAHIGTRTYK QVDSHRTKVIAAEREGRPVR TLDPTMDWRMRPYCASTTA RWLAEQGRSPVAPRSPCPEP HAPPPAAALLYIPATPPAPTP RAPVAPPKLAPPAESTVPAT PDGNPEAPAPPFSAPGPPTPK ALPPPPPSRRNLRPHLVPKD AWQGVADAVAPAASRLLRT PLAHLSTEQWATFEAALAG LEATLHHAARSAEAVPTRC ASRAREDAERQLREARKTR

						EIFGKAAALYAAGKDPTATI ERIPPEVRLHLPTPGSAEWP ARAAAARRVIRRAVARADR LRKRMGILDSDRDLQRLFN ANQKKAVRQILAPSTKAPRC QLDPAAVEEAYIQTLAKPPPI DPSPPWKNSVQWPRPPTAA DDGGSPFSVAEVRAQLRRLP NGSAPGIDGIPYEAYKRTKL DATLAHVFEVVRLNARLPA RWDVARTVLLYKKGDPND TGNWRPISLQVTIYKIFTAAL SKRLISWAGKHNTFSASQKG FLPAEGCHEHAFVLRSVLDD ARRHKQNVYLAWYDLRNA FGSVSHDLIAWCAAMLGLP RYLRDAIGAIYRHSALFVQV GDQETTGVIPMRCGVKQGC PLSPLLFNLCVEPALRCLRRT TGYKFYGTSITVEGQAYAD DLLTAAPSAYHAARQVATIE EWANWAGVSFVVQALSLD APAGKCAALAINFEGGLMH

						SIDPALKVQGAAIPAMSRNN VYRYLGVHVGLTDALGQA NELLEKASRDARTICASGLE PWQKVVAIKTFILSRLPFFFH NGKIQRGRCQQFDRELRENL RAALRLPVCTTNAFFHSRVA SGGLGILPIAEEQQVYLAAH VFKLLTSPDLSIRAIARHQLA EVTHARHTTPVQDGEASPFF GWLMRGQEVASTTPSGDVS SIWFAAAGAYSRMGWSVRD ALHPTLTVGPGVQFEGRFQR ANVIPALRASAFSRHAVEWS ALRTQGRAAAYQHAVHPAT HHWVHNSAGLTTKEYRFAI KCRLGLLPTRAAPHHRNGPT ACRACSYARETANHVLGHC PATKAEVIARHNRICRALAQ AAEASWTSVLEDVPIPGVDS PLRPDIYCSRPGQCAIIEVAV SYEDAFNASMEGRAKQKTD KYAGLAATVEEQLRLQTRH AAFVVGFSGVVLPASVTAT

						ATSLDLPPKTWNVLLKRCV AASIKGSYTAWRRFRRSTP (SEQ ID NO: 1424)
NeSL	Utopia-1_AEc	.	Acromyrmex echinatior	GGTGCACAACGGATG CATCATACGTGTACCG GAGCATACGGGCTGT CACGGCGGCTGCATG CGCGATCTAGCTCGG AGATTTTATTTATTTA TTTATTAATTTATTTA TTTATTCATCGAGTGT GAGTGTTCGCGTTTTG CCGAGAAGCGATTTTC GTTAAGTGATACGCG CCGCGTTCATAGGTTA G (SEQ ID NO: 1180)	TAAATTATTTTGTCTTT GTCTTGGCCCCCCCTTT TTAAACCAAGCAGGAG AGAGTGGCCCAATGCC CAACTATTATATATTAA CTATTTACTGTGATATT TATTATTTGACTGTTGG GCGGGCCCCTCTCTGCT GGTTTTATTTATATATA TTTTTTACTCGCGTACT TTTTGTACTACTCTATT TTTCTTTTTATTTTAGCT ATGCTATTTTTATCTCT TTCTTTGTCTCTATTTTC TTTCTTTTTTTCTTTCCT TTTCTTTTCTTTCTTTTA TTCTTCTTTTATTTATCT TTTTTTCTTTTCTGTTGT GGGGCCCTGACCGTCC GAGTGTGAATGCCGCG	VCSVRGCRREDSRRFYKFKF PLNFVKVPKTIVIGSAFQKSS VSARSQNHSRSTRVPKTRQP RTSNTIGRYTAASANNYLTV IITGNYTVFAQWICYRECTW LLSKFVNFFLTIIGYFFQLRL VVIYEGPVILDTFSNCGSSLF MRGQXSKALLVRLNRSALA MADPQVHYIDYPLPPRVKC VKCFGAEGAGKVKGEYSDP PHLAKHLKKCHPGDTLNYK CSICDLRGTGKYPLRDVKAH YAECHVSPAVDAAGPSTRG SLGECSGAGQPTASRAAKA TTRLAETVGGTDKRRAATS GSRQLTLPFAATPSPSTAAG EARAPRSXSTTPTSRSPSYA AVTAGPPSMRSTTTSTTARS KTVAKGAAPNTTTTTTARR SGEAAATRKPPTTATVSKPR

					AAAAACAATATTATGT TTTATACGAGTGTGCAT GTGCGTGATATATTTAT CTATTTTATTTTATTTAT TTATTATAATTTATTGC CGCGCGCGCTCCTCCG GGACTTTTATTCGTTGA CAATACTGTGATATTTT TCTGCKCAGGCTGGGG GGGCTTGCCCCCCAGC CCCTTAGTTTTAATTGC CTATGCGGGGGGGGCT TTTGTCCCCCGCAAATG TATATATATATATATTT AGCGCGCGGCTTAGCC GCTTTTGTTTGTATTAC CCCAGAGGGGAATTGT CCCTCTGGGGAAAAAA AATGATTGGAAAAATA AAGTGAGCTAA (SEQ ID NO: 1303)	VLSVETVRLPVDDIQRAGVQ NAAKPARAPSRPPQRTSPEA GGPRTTGAKEKCGEGAYKK LPANSGNPISTRTRRATSVPV EKSEGTARRERVSPHPPPKGI DIILSSTSEEEGTPYQPGGVG RLRLRRKKVTGPPPKMTPRE GVVTRARRSTSAPVEKSALD ARLTALDRTSSRATGNPTSQ IAGGLYTSRGQPERTPPARL PSLSPTTRGSPSGSLGEIRTPI SPATSLPATLTTCTVTTTTCG SPITSTGFTGGVGRLITPPSLP QTNILPTIGEEGTSPCVAVVT THPRPTGEDAPCEAPQPVSD HQRQSIGEPRRDTDTHLACD VATGNAHHLHGDDDHLWK PHNIHGLHRWRGEADNTXE PPPNEHPPDHRGGRNVTVR GGRHHPSLGGRSTSPLILPRP TTPEPERGQEERRLEGAAQP PTTPVVEGDNQWDGQWTV SVRRRARRQQLNDTSPSNSE

						SPPTAGPSRSPRIAPLSALIA ASTSRHETSLNLNCTNGNIC MDRTPPRNILPVXAERRRET SPQDRVEGDIGYGAGKVSA EHPSAPVNVRGVMSRGRAT ASSIVPPRANRGEGGRQHHS RRRPDAPVGQPSRDHPAPAT VARQRRRERVAARDALLDR AKDVATIADLEAFAASVAA FFGEDASATGAAARARDRS VRSREAGARRGVKGGERPE REGAGRPGSAPADPGASGE ARGDWVREAKRLQALYRA NRRKAVREVLQGPADQCQV PKRQVQEYFERLYSGGEDL AGAGVEAERPDPSSPREVSA VLGPLAEREVDRRLRRMNN SAPGPDGVSYRDLRGADRG ARLLTALYNICLRLEAVPAS WKTSNTVLIHKKGDRGMLE NWRPLALGDTVPKLFAALL ADRLTDWAVTRGKLCSAQK GFLRDEGCYEHNFVLQEVL

						THAKRSKRQAVVAWLDLSN AFGSIPHATIRRALIRSAVPR GLIAIWDSMYDGCTTRVRT AEGHTAPIPIRSGVRQGCPLS PIIFNLAIDSVVRVAAEXND GYSLHGNTWSALAYADDIA LLAQTPEGMERMLASVEAE AASVGLRFNPAKCATLHVG AGNGGRVLPTSFQIQGETIN PLAQGESYTHLGVPTGFSVD QTPYAAVGDIVSDLRAVDR SLLAPWQKIEMLGTFILSRL DFLLRGARVFKGPLTAVDL NIRRHVKSWLNLPQRASAE GVYMPPRWGGCGLLPLSDL ADVLTVAHAYRMLTVRDG AVRELAWESLRGVVGRRIG HAPSCEDIASFLSGSLDGRM RGGGEASLWSSARNAALRQ SERLSLRWRWVEATEEMTL ECRGPRGAAIKIPPEARGQV VNRLRSAVAEHYASRLLSKP DQGKVFEVSSRSRVSNHFIR

						GGSFTRFADWRFIHKARLD VLPLNGARRWEANDKRCRR CGEVSETLPHVLCHCGIHSA AIQLRHDAVLHRLWKATRL PGVVRVNQRVEGVSDELGA LRPDLVVRHEPSKSVVICDV TVPFENRWTAFEDARARKIA KYSPLAEELQRRGYRVVVT AFVVGALGSWDPRNEAVLR LLRVGNQYAAMMRRLIVSD TIRWSRDIYVEHVSGTRQYL APSRPSGDLATPPRAVRRRW LAEERSAQDAARRGSDSVS VA (SEQ ID NO: 1425)
NeSL	Utopia-1_AMi	.	Alligator mississippiensis	TGCTGGAAAGACGGA GAACCGCTTCCTTTTT CCCTGCGCCTGGCCTG GTATTGCAGTACCTCC AGGATTAGCGCCAAC TAGTCCGGCAGACTGT CGGAATACAGCAATA GAAAGWGAGCTGACT AGCAGCTTGCTTTCCT	TGAACCCCCCCTCTGCA CCAGATGGACCTTCACT TCGAGAGGATTCTTCA GCAATGGACGACCCCG CTCCACCCGAAGAGGA CCCCCGCGATGAGACT CTATATGGACTGAGAC ACTTTTTCTTCGAACCA CTTCCTCCACCATTGCG	CHHAGLRPGTPNRTRRPDQ TAPLPDPRGHPMPPNRRGSR SRPEEPSRREPPXPRACQGL RVWSPPQQRMPTPWQTLWL EELSRATTFKAFEASVARLT EELSAAARPGQPRGGNNRP ATRRDHRLQPQRRPRRQRY DPAAASRIQKLYRANRPKA VREILEGPSAFCQVPRETLFN

				TCCTCCGGTGCAGCAT	GACCATTGTAACGGGT	YFSRVFNPPAEAAAPRPATV
				GGGTTCTCGTCAGTC	TTGTGTGTATCTATCTT	EALTPVPPAEGFEDAFTPQE
				MTGACGGGCTAGGGA	CTTTCTCTCTCAGCGTC	VEARLKRTRDTAPGRDGIRY
				AGGCGGTGCTGCCAG	GCGAACCCCCTCCCTCC	SLLKKRDPGCLVLSVLFNRC
				TACGTCCGAAAGAGT	CCTTCCCCTCCCCCTCC	REFRRTPTTWKRAMTVLIH
				GCCGGTTGCGCAAGC	CCCCCACCCCCGGGCTT	KKGDPTDPGNWRPIALCSTV
				GACCGCGCCACTCAG	AGTTGGCTAACATTGTA	AKLYASCLAARITDWAVTG
				GTGAGTAGCCAAGGG	TCTCCTGTAACCTAGTT	GAVSRSQKGFMSTEGCYEH
				TCTTACAGTTCACCGG	GCGTTCCCCTCCTCACC	NFTLQMALDNARRTRKQCA
				ACCCGAWAACGCGAA	CCCATCCCTCTATTGTT	VAWLDISNAFGSVPHRHIFG
				AACCCCAACTCGGGC	AGTCCCTCGCTCGGGC	TLRELGLPDGVIDLVRELYH
				TAGTAGCCGAAGACC	GCTCTGTATTTCCCTAC	GCTTTVRATDGETAEIPIRSG
				TGGGTCCCCCCCMGG	CGGCTTTGTCATCTTTT	VRQGCPLSPIIFNLAMEPLLR
				TCAGAGTAGGCGAAC	TTGGATTCACAATCCTA	AVAGGPGGLDLYGQKLSVL
				GCCWGKGCTCAGAGG	AACATCTACTAATAAA	AYADDLVLLAPDATQLQQM
				ACGGAACGCGGAAAA	AGTCAATC (SEQ ID NO:	LDVTSEAARWMGLRFNVA
				CACCCCCAGGTCCCA	1304)	KCASLHIDGRQKSRVLDSTL
				AGGACGCCCTGATCC		TIQGQAMRHLRDGEAYCHL
				ACTGACAAGAACGCT		GTPTGHRAKQTPEETINGIV
				CGAGGCACGCCAGGA		QDAHKLDSSLLAPWQKIDA
				GACCCCCAGCTAGGG		VNTFLIPRVAFVLRGSAVPK
				TGGACCGCCGACTGC		TPLKKADAEIRRLLKKWLH
				AGGTCCGGAGGACCC		LPLRASNEVLHIPYRQGGAN

				TCCCAGGAGGGTGGA CCAGCGAACCCAAGT TGGCGACGAACCCTG ACGCACCCCCCACGA TGTCAGGACCCCGAC AGGCGGCGGTGGACC ACTGACCATCGACCG ACCCCCAGAGGCAGA GAGACTCTCAGAGCC CGGAACCCCGGCTGA CGAGAGCCGCCTCCC GGCGGAGGACCCCGG AGCCTGAGGATGCCC CCCGGATGACGGCGG AGCGCCCCGAGCGAC AGCGGACCCCTCCGG ACCCCCACGGCCCCTC GGTGACGATGGCGGG CCCCGAACGACGACG ACCCCCGGACCCCGG CGGTCCCGAGGACGC CCCCCCCGAGGGTCTC CCCACGCTGGTGGAG		VPRMGDLCDIAVVTHAFRL LTCPDXTVSIIAASALEETAR KRIGRQPTRRDLATFLSGSL EGEFSRDGGDFASLWSRAR NATRRLGKRIGCAWTWTEE RRELGVSLQPAPHADRVTV TPRTRTFLERFLKDAVRNKY AGDLRAKPDQGKVFDVTSK WDSSNHFMPSGSFTRFADW RFLHRARLNCLPLNGAVRF GHRDKRCRRCGYVAETLPH VLCSCKPHARAWQLCHNAV QDRLVRAIPAAAGEISVNRT VPGCESQMRPDIVITNEEAK KVVIVDVTIPFENRRQAFTD ARARKREKYAPLADILRGR GYDVTVDALIVGTLGAWDP SNESVLHACRVSRRYAKLM RCLMVSDTIRWSRDIYVEHI TGHRQYTDPTRRTAAGPDP EGTA (SEQ ID NO: 1426)

				GAGCCCCGGACCCCC CCGACACCGGACCCC CCCACGGACGACCCA GGCGAAGGCGTAGAC ATGACAGCACTCACG TTCCTCCCCTTCCCCC TCCCGGCGAAGCTGTT CTGCCCGACCTGCCAC CCGCCAAGACAGTAC AGGTCGCACGGCGAC ATGAACAAGCACCTA CGGCGCTTCCACCAGC TGCGCCTAGCCTTCTA CTGCGCCCTCTGCGGC ACCGAGTACGAGGCC CTGAAGCTCCTGAAG AACCACCAGAAGGGA TGCGAGGGCCACGGA GCCGAGAGGAGACCC GGCACGCTGGTGAGG TCCGCTGCCCCGGCCC GCCGGACCCAGGCCG CGGTGCGAAGGCCCG		

				CCAGACTGGCCACCC CGCCGACAACCCCAC CGGACCAGACCTCCA GGGACCACCCGACGG AGAGACCTGCCCCAG TGA (SEQ ID NO: 1181)		
NeSL	Utopia-1_CMy	.	Chelonia mydas	CTCTTCTTATGAATAC TTGCAACACCTGCACT GAAGATGGATTCTCC GGCTGCTATTTTTGAA AAACTGATGCTGCTTT GAAGGTGTATTCTGCT GCTGCTACCTTGGAAG GAAATTCTCTCTCTGC TCCTGAGACATCCCCA GCTGCACCGTGTACCA CCACCACCACTGCTGC TGCTCCACAGAAGGTT TCTCGGACA (SEQ ID NO: 1182)	TGAGCCGGTACGACAT CGTGCATCAACTATGA GAAAGGGACTGAGAGA CTTTTTCCATTGGACCA TATGAACTGGAACCAT AAACTCACTGAACATT AAATCTCACCAAATGA GGGTAAATCCATCCTC ATCATCGTATCCACTCA TTATACTCCACACCTGA ACATAGCCATTATATG AACAACATACCCCCAT ATCTCAATGTCTGTACT TTGACCCGTTAACCTTT TACCCCCAATCGGGGA TATTGCAGATTATGTAT TCCTTACGCCACCCGAT	MTTKKVLGASTTLQTSSTK GKNSGCSKDPLRDAVPGRS WILRPACRDITTRRNIPPAPQ QQQPPMESPPTLQLQDALRR PSPTPAAAQVADAGGALAA LHTIKRGISVDWTSISPKXXQ RXTSASPDACPASETTQRDX RXLLDARPAGPLDPTRPHQD EPASDTADAAGTPLLQGNE DTIYLQYPLAADMLICPICSP PQSFHLLGVVTRHLKRCHSK RVAFSCALCSLPFETQKQCK MHQVACRKCLKGTTQSPAP APSPPAARRPAAPEPQRRKX TSQAAVKKPAPVARPAERD AAIEKVPAASGNITQVLASR RPVSPSHVAKXISMLRRLSA

					CCTAAACCGAATTTCGC ACCCCTTGATAATCTGT ACCTTATTCCCTGATAA CCAGAAACTTCTATGCT TAAACTCTGTACCGTTT TTTTTTATTTCAACATC ATCTTAATAAAATTATT AAA (SEQ ID NO: 1305)	ASPPVQHVPVPRRISAPPRIA ARDPVAGRASAAPQTALRT PAAGGASTTPQTALRTPTAG GASAMPQTTLPXPRRPDWR NQPRSHSKAPGLHRQTDQH GPQVHSAGHCLREISRSSSN RLGSSHSAAATHRRTGGVP ATPEPDRVSPTTSNAXIPPEIP PQHPTEGNPDPRDRRQADH TAGSEPAPDEVEDXEGQRP MVRAATPWQTAWTEELQA AASFDDFDLLVDRLTRELSA EIAPRRSSNQENAPPAHRTP APNHNTTTRGARSRDASRR YDPAAASRIQKLYRANRSK AMREILDGPSPYCTIPSERLY SYFKDVFDRIARNDAQRPEC LRPLPRVDEAGVLETDXTPK EVMARLSKTKNTAPGKDGI PYSLLKKRDPGCLVLATLFN QCKRFCRTPSSWKKAMTVL VYKKGERDDPSNWRPISLCS TMYKLYASCLASRITEWSVS

						GGAISSIQKGFMSCEGCYEH NFVLQTTIETARRARRQCAV AWLDLANAFGSMPHHHIFA TLQEFGMPENFLRVIREVYE GCSTTIRSVEGETAEIPIRSG VKQGCPLSPIIFNLAMEPLLR AISNGTDGFNLHGERVSVLA YADDLVLTADDPESLQGML DATSRAADWMGLRFNAKK CATLHIDGSKRDSVQTTGFQ IQGEPVIPLAEGQAYQHLGT PTGFRVRQTPEDTIQEILQDA AKIDASLLAPWQKINALNTF LIPRISFVLRGSAVAKVPLNK ADKIVRQLVKKWLFLPQRA SNELVYIAHRHGGANVPRM GDLCDIAVITHAFRLLTCPD AMVRNIAANALHDATKKRI GRAPSNQDIATFLSGSLDGE FGRDGRDIASLWSRARNAT RRLGKRIGCRWEWCEERQE LGVLVPQIRSNDNTIVTPSAR GMLERTLKAAIHSLYVETLK

						RKPDQGKAFELTSKWDASQ PLPRRGRLHPFRRLAVHPPC PAQLRPAQRSRPPREPRQAL QEVRLLQRDPAPRPVQLQAP LQSLAAAPQCHPEPPGESHR TAPGGGRRELRHPRYPASGT PANHFLAGGGFTRFADWRFI HRARLNCVPLNGAVRHGNR DKRCRKCGYSNETLPHVLC SCKPHSRAWQLRHNAIQNR LVKAIAPRLGEVAVNCAIPG TDSQLRPDVVVTDEAQKKII LVDVTVSFENRTPAFREARA RKLEKYAPLADTLRAKGYE VQMDALIVGALGAWDPCNE RVLRTCGIGRRYARLMRRL MVSDTIRWSRDIYIEHITGHR QYQEV (SEQ ID NO: 1427)
NeSL	Utopia-1_CPB	.	Chrysemys picta bellii	TTTTTTCTGATGCTTG ACTGCAAACACCCAT CCAGAAGATGGAATC TCCTGCAGCCATTTTT GAAAAAATTGATGCT	TGAGCCAGAGTGACAT CGTTCTCCCACTACGAG AAAGGGACCAAGTGAC CTTCTCCGTTGGATCAT ATGAACTGGAACCATA	MTQDQDADCCPAGKDATR GAPPMTQDQDADRCPAAPE RDAPEGTTSSTPDPKTTYHP AVRRRAARRGMHLRAQDL DAARCPSGQRDNVASESSAP

				GCTTTAAAGATATACT	AACTCCCTGAACATTA	PRATSPPQASLPDPEESPGES
				CCATTCTCCTWKTTTG	AATCTCACCAAATGAG	AGTTEIRPTEGEAGEEDRIYL
				KAAGAAAACTCTTTTT	GGTCAATCCATCCTCAT	QYPLPTGLLLCPFCLPVHGV
				CAGCTTCAGCTATTCT	CATCATATCCACTCATT	QTLAALSKHVRKTYNKRIAF
				GTCATCGGCTGCTGCT	ATAMTCCACACCCGAA	RCSRCDLPFETQKKCKFHQA
				GTTCCTGCTTCCCAGA	CACAGCCACTCTATGA	TCRGPPTTAKVNPTDILRVP
				AAGCTCAGCMAAAAC	ACTTCATACCCTCATAT	TLTPTDDLASAPQPASPESQ
				CTATCCTGAAGACCW	CTCAATGTCTGTACTTT	QIRGDQPPTEGSVTPASRTD
				CCCTTGGTGCCTCACG	GACCCATCAACCTTTTA	DATKRTSPVSRIPTLDPAVR
				GAAGACCCGGASCAC	CCCCCAATCGGGGATA	GTTATSQVNNLTRRLSDLIK
				CTGCAAGAACCAAAA	TTGCAGATTATGTATTC	TIRHNTDTRRCSAPPQVTSC
				CATTAGGAGCTGGCT	CTCATGCCACCTGATCT	RPAVGATSIVPQAARRDPAN
				GAAGAAACCCCCCGT	TAAACCAAACTTTGCA	GGASRSPQIPQPDPAPGRPN
				GGATACCTCWGCAGG	CCCTCGATAATCTGTAT	TSSKVTQRASDRQKPHAPPR
				GAGACCTGGSTCCAG	GTTATTCCCTGATAACC	THQPDAARRRTRTIPSASKH
				MAGGACAKCTCTTCG	AGAAACTTCTATGCTCA	DRAPTKPSTGASRTPLPPGR
				GGACCTCMCATCSAG	AACTCTGTTCACTATTT	SSAASETPRAALPTTPGPPPQ
				GAGCAAGAATATCTC	TTTTTAACATCATCTTA	DPPEHRSTVRGTTRPQTVPA
				AACAGCTCTTCAGGA	ATAAAATTTTTAAATCT	APEPAETTQQEERRPRARVA
				GGGGGACCCCCGGAG	GTT (SEQ ID NO: 1306)	TPWQSAWMEELAKAEDFEN
				AACCCTGCCCGCTTCC		FDTLMDRLTAELSAEITARR
				CAGAACCAGGATGCT		REPQEAARATRRFPAPSRNN
				GATCGCCGCCCCACC		TAREGRRGDVGRRYDPAAA

				GGGAAGGATGCCACC GCAGGAGCCCCCCCA (SEQ ID NO: 1183)		SRIQKLYRMNRTKAMREILD GTSSYCAIQPERLYSYFKDV FDHEAQTNLRRPECLSPLPRI DLTEDLERDFSPQEVQARLS RTKNTAPGKDGIRYPLLKKR DPGCLVLAAIFNKCKQFHRV PRSWKKSMTVLIHKKGXRD DPGNWRPISLCSTIYKLYAS CLAARITDWSVCGGAVSSV QKGFMSCEGCYEHNFLLQT AIQEARRSKRQCAVAWLDL TNAFGSIPHHHIFATLGEFG MPETFIQILRDLYKDCTTTIR ATDGETDAIPIRRGVKQGCP LSPIIFNLAMEPLIRAISSGPT GFDLHGKKLSILAYADDLVL TADDPESLQGMLDATSRAT DWMGLRFNAKKCATLHIDG SKRDSVQTTGFQIQGEPVIPL AEGQAYQHLGTPTGFRVRQ TPEDTIQEILQDAAKIDASLL APWQKINALNTFLIPRISFTL RGSAVAKVPLNKADKIIRKL

						VKKWLFLPQRASNELVYIA HRHGGANVPRMGDLCDVA VITHAFRLLTCPDATVRNIA ANALRDATEKRIGRAPSNQ DIATFLSGSLDGEFGRDGRDI ASLWSRTRNATRRLGKRIGC RWEWCEERQELGIRVPQIRS DDNTIVTPTARGLLERTLKA AIRSLYVETLKRKPDQGKAF ELTSKWDASNHFLDGGGFT RFADWRFIHRARLNCVPLN GAVRHGNRDKRCRKCGYP NETLPHVLCSCKPHSRAWQ LRHNAIQNRLVKAIAPRLGE ISVNCTIAGTDSQLRPDVVV TDEAQKKIILVDVTVSFENR TPAFREARARKLEKYAPLA DTLRAKGYEVQMDALIVGA LGAWDPCNERVLRTCGIGR RYARLMRRLMVSDAIRWSR DIYIEHITGHRQYQEA (SEQ ID NO: 1428)

NeSL	Utopia-1_DYak	.	Drosophila yakuba	AAAGTGTAGTTCTTTT CTGTTTTAGTGTAGTG GGAAGTCTGTTTCTTT TTATTATGTTTTTTAC GAAAAAGTCCTGGTC TTTGAAATTCATTGTC TAAATTTTAAATAAAA TTATAAAATTTAAAAA GAAAATTAATTAAAG AAGCGATGAAATATC TCTGAAATTCAATCAA TCAATTAATCATGGCG TCTCAGCGAGTGCAC GTATTTGCCTACCCCT TCGTGGGACCATTCCG GTGCTCCGTATGCATG GATGCGTCCGGGATG CATCCCACTAGGTCGC TGGGCGAATACGGCA CATACGCTGCGGCAT ATCAGCACATAACCC GGCGCCACCCACAAG TGGTTATTACATACCG	TAAAAAATTAAAATGC CTTAAAAATAAATAAA TATATCAAAATTTAAA AAAAAAAACGAGGAAC AAATAAACACAAATTC TGAAAGATTTATATAAT TTAAAAWATAAATCGA AAATAAATGTTGAAAA CAAAAAAAAAATAATA ATAATAATAAAAACAC AATAACACTCACCCGG CCTGCCCCAGAGGCAG GTAAACATTTACTGGCC ATATGGCTTTTTTTTTA A (SEQ ID NO: 1307)	YAPGYEAAQSPCGREPPRD HHRRPRDACGSSHSPEPCLT TPRLLPETVSAEPCDDESQR TRYASPHKQARTLHDAEPR DASREHAPSCAEPRCHRCQ WTHWKDCCPHSTNTTDGPE GTDRCADTITSPATAACPQR SPCPLGSSNGCDETAPEKRQ PAADLVHTAPFAVLVRAGP FADLVRAGPFADHHQDDDP LPHRSGSLGPLCSKQKDPRK THQHRHSGQAGNQTHTDIP RAAPSRRAAICLMANAAAT REDLLRAATSLSEMAAANQ PTRSPTGGGEPTSQGRRGPQ ALADAAKRIQQIYRTNIPRA MRKVLRTLLTAVFSACLRT GHVPDLCKKSRTVLIHKKG DRTDLSNWRPLSMGDTIPKL FAAVMADRLTAFLTNGGRL SEEQKGFLQHEGCHEHNFV LGQVLEESRRQGKDLVMG WLDLSNAFGSIPHATIMDAV

				TTGCCGGGTCTGTGGC GCTGA (SEQ ID NO: 1184)		AGMGIPSRIRTIIHQLATGAA TTAKTIDGMSEEIPIEAGVRQ GCPASPILFNIAIERVLRKIKT VNAGYLLYGSRISPLAYADD LVLIASSPEEMRSLLRAADD AAIEAGLHFNPKKCATLHLT GKKSSRRAVQTGFLVRGTPI PAMTEGDAYEYLGIPLGLK KNQTPRAAMEAIVGDIAKID DSLLAPWQKIDAARTFVAP KLDFVLRSGATLRAPLRHLD TVIKKHIKKWLYLPQRASAE VVYTPLKKGGAGILPSSILA DVLTIAQAHRMVSCPGEVV SRIASEGLREAVKRKINREPS GDEMAHFLSGSTLSGETASF GDAGFWSRVRMATKRQAV HLGVRWAWRGGELLVESR GQRNRPVATDSNSRSQLIQR LRCAAQDEFLTILINKPDQG KVAKLSTLTPVSNAFIRDGS FTRFADWRFIHRARLGVLPL NGAIRWGSGDKRCRVCGYQ

						LESVPHVLCHCMHHSNAMQ QRHNAVMDRLAKAGSRLG TPRVNCRVEGVAEDMAALR PDLVWRDERSRKIVIVDVTV PFENGAEAFDNARGEKEEK YRPLAEALRAMGYQVKLEA FIVGALGSWDPKNERVLKTL GVSRFYAGLMRRLMVADTI RWSRDIYVEHVSGIRQFTLP SGAPSN (SEQ ID NO: 1429)
NeSL	Utopia-1_Gav	.	Gavialis gangeticus	CGCTGGAAAGACGGA GAACCGCTTCTTTTTC CTGCGCCCGGCCTGGT ATTGCACTTCCTCCAG GACCAGCGCCAACCT AGTCCGGCAGACTGC CGGAATAATAGCCTC AGAAAGAGAGCTGGC TAGCAGCCCTCTTTTC TTTCCTCCGGTGCAGC GTGGGTTCTTGTCAGT CCTGATGGGCTAGGG AAGGCGGTGCCGCCA	TGAACCGCCCCCCCTCC GCGCCAGACGGACCTT CACTTCACTCCGAGAG GATTCTTCGACCACGG ACGACCCCGCTCCACC CGAAGAGGACCCCCGC GATGAGACTCTATACG GACTGAGGCACTTCCTT CGAACCACTTCCTCCAC CATTGCGGACCATTGTA ACGGGTTTGTGTGTATC TATCTCCTTTCTCTCTC AGCGTCGCGAACCCCC	MSGPRQAAADPRPSTDPRR QRDSQSPEPRLTRAASRRRT PDPEDAPRTTAEHPERRRTP PDPRGPSATTAGPERRRPPD PGGPEDDPPEGLPTLVEEPR TPPTPDPPDGRPRRGCRRGS AHVPPLPPPCEAAVPDLPPA KAVQVAQRHEQTPTALPPA APSVLLLPLRHRVRGPEAPE EPPQGMPGPRGREETRHAG EVRRPTTRAAARRPARPAAP PATPPDQTSGDRPTERPAPA TPPRRSAPRDPRPDVTPRPD

				GTACGTCCGAAAGAG	TCCCCCACCCCCCACCC	GPPPGPPGPPDAPDPPRIPEP
				CGCCGGTTGCGCGAG	CCGGGCTTAGTTGGCTA	PGEPEPPGALQLPSVPGSPG
				CGACCGCGCCGCTCA	ACATTGTATCTCCTGTA	AETSAQQRMPTPRQALWLE
				GGCGAGTAGCCCAAG	ACCTAGTCGCGTTCCCC	ELSRATAFEAFEASVARLTE
				GGTCTTACGGTTCGCC	TCCTCACCCCCATCCCT	ELSAAARPGQPRRGADNGP
				GGACCCGATAACGCG	CTATTGTTAGTCCCTCG	TTRRDHRPQPQRRPRRQRY
				AAAGCCCCGACTCGG	CTCGGGCGATCTGTATT	DPAAASRIQKLYRANRPKA
				GCCAGTAGCCGAAGA	TCCCTATCGGCTTTGTC	AREILEGPSAFCQVPRETLFN
				CCNTGGGCCTCCCTCC	ATCTTTTTTCTGGATTC	YFSRVFNPPAEAAAPRPATV
				CCAGGTCGGAGTAGG	CCGATCCTAAACATTTA	EALTPVPPAEGFEEAFTPRE
				CGAACGCCCGTGCTC	CTAATAAAAGTCAATC	VEARLKRTRDTAPGRDGIRY
				GGAGGACGGAACGTG	TGTTCTTT (SEQ ID NO:	GLLKKRDPGCLVLSVLFNRC
				GACAAAACACCCCCA	1308)	REFRRTPAAWKRAMTVLIH
				GGTCCCAATGACGCC		KKGDPTDPGNWRPIALCSTV
				CTGATCCACTGACAA		AKLYASCLAARITDWAVTG
				GAACGCTCGAGGCAC		GAVSRSQKGFMSTEGCYEH
				NCCAGGAGACCCCCA		NFTLQMALDNARRTRKQCA
				GCTAGGGCAGACCGC		VAWLDISNAFGSVPHRRIFG
				CGACCACGGGTCGCG		TLRELGLPDGVIDLVRELYH
				GAGGACCCTCCCAGG		GCTTTVRATDGETAEIPIRSG
				AGGGTGGACCAGCGA		VRQGCPLSPIIFNLAMEPLLR
				ACCCGAGTCGGCGAC		AVAGGPGGLDLYGQKLSVL
				GAACCCCGACGCACC		AYADDLVLLAPDATQLQQM

				CCCCCCGCG (SEQ ID NO: 1185)		LDVTSEAARWMGLRFNVA KCASLHIDGRQKSRVLDSTL TIQGQAMRHLRDGEAYCHL GTPTGHRAKQTPEETINGIV QDAHKLDSSLLAPWQKIDA ANTFLIPRVAFVLRGSAVPK TPLKKADAEIRRLLKKWLH LPLRASNEVLHIPYRQGGAN VPRMGDLCDIAVVTHAFRL LTCPDATVSIIAASALEETAR KRIARQPTGRDLATFLSGSL EGEFGRDGGDFASLWSRAR NATRRLGKRIGCAWTWTEE CRELGVSLQPAPHADRVTV TPRTRTFLERFLKDAVRNKY AGDLRAKPDQGKVFDVTSK WDASNHFMPSGSFTRFADW RFLHRARLNCLPLNGAVRF GHRDKRCRRCGYAAETLPH VLCSCKPHARAWQLRHNAV QDRLVRAIPAAAGEISVNRT VPGCESQMRPDIVITNEEAK KVVIVDVTIPFENRRQAFTD

						ARARKREKYAPLADTLRGR GYDVTVDALIVGTLGAWDP SNESVLRACRVSRRYAKLM RCLMVSDTIRWSRDIYVEHI TGHRQYSDPTRRAAAGPDP EGTA (SEQ ID NO: 1430)
NeSL	Utopia-1_LV	AGCV01358106	Lytechinus variegatus	ATCTACTATC (SEQ ID NO: 1186)	TGAATAGCATTTATATT GTGTTCCAAACAACAT ACTCATTATTATATCTA AACATTTTTTTTCTGT TCCTGACAATCTACGTA AAGTCTGCTAACCAAC TGGCATGATGAAATAA GATAAAATCCCCTTAC ACATTAATTTCTTGTCA CATCATAATGCTTTGTC AAAGCAATGTCCTACA TAATATCTCGATGTCAC CCCAATTAATTTTACAT CCTTCGGTAACCTTTAT ACCGTTGGATCAACAT ATATGATTTGTAAAACT GTTATTTCTGAGTTTTT	MSCPREGSDHLGPDPETPAL HQGSDIRVTSSRLRGSRGKS SRQPSSRHQVPASEASATAQ QTAANECQVCGSSFATSSGL RRHMARLHRAASADPEGAA PASITEIFDYPLPSRWKCSAC SENFFNQQTLKRHQTRHHP ATTFAYAFRCSSCRSEFDSA RRAANHWQVHKKERSQLS GTEPQASSQARVSMAHSPPP LPNTSWAELASNPAEIPSFV WESPPKNRPSVEEFGSSLPT DVTMMSQSPPPQVQSSPVP ALTPLSPAATASSSPPGAAR QLTPPTQTNTPVTQRARLQP EADVVPELPPSVTEHPVSDA QHWVDAVSSASDWSEFEAV

					TCTATGCTAATAAA (SEQ ID NO: 1309)	CDQFVIHAVAVSRPNLARPQ QQDRQRSGDHPPRQQRGQH RPTFDVREASRIQKLYRTSK KRAIRHILKEKSPSFSGSESD VLDFFREVYSAKEVDEEAV GKLASSLFDVPQGDDSATSL SLPTSAKEIGARLSRMTNSA PGKDRLEYRHIRRADGSFSIS EAIFNKCLAEGRIPAPWKTA STILLHKAGPTDDPANFRPIA LQSCLYKLFMAVLADRLTK WACENQYLSPEQKSARPCE GCFEHSFLLSAALKDCRRNQ KTICIGWLDLRNAFGSIPHPV IKIVLSSLGVPDSLVTLLMD AYNGASTSFTLTGGQTDTVP IRSGVKQGCPMSPILFNLAIE LIIRAVKKNASDNHLGVTVQ GKNLSILAYADDLVLLSRDT EGLQSLLQVAGSSASTLQM QFKPQKCATLTLDCKRGTN VRQSAHHIQGAAIPSLTEEE RYRYLGVPIGLPRLTSLQESS

						RKLSSDIETISSSLLAPWQKL DAIKTFVIPILQYTLRATEYL KSDLKPLRAAIIKHVKKICH LPVRSSNAFVFASRPSGGLA FVDPGVDADILVVTQAVRT LASDDDTVRAVALGQLTSV VHRTVHSAPSDDCIDKFLSG SSEGPLANSGNSGQASSLWS RTRAASRRLKIRIVGASSGDI KVESGGRAIPSKKVTAGLRS DHHNEMSEKLRSLPDQGKV ARALSLDSFANATSWLTSGS FIRFCDWRFIHRARLNCLPT NAAVRRWKQNANTKCRRC DHQLETLPHIINNCRPNMVPI RRRHNSIQERLVKAIHYGDI YQDQHVPGDPNPRERPDITV VEGNKVTIVDITIPFDNGPDA LSTAANAKVMKYDTLRQEL ASRGMDVEVHAFVIGSLGS WHGDNERVLGRLGISRRYR TLMRRLCCIDAIKGSRDIYIE

						HVTGHRQY (SEQ ID NO: 1431)
NeSL	Utopia-1_NVit	.	Nasonia vitripennis	CCATTCCTTCGTACGGGTTTTCGTGCCGGCATAGCCGGGTGGGAGACTCGCGCGGGGGAGGTCATATCTCACCACCATCCTCGAGCTTTTTGCTGCACCTGA (SEQ ID NO: 1187)	TAGTGGGGCCATAACACCTAGGCCCCACAGTGTGGCGATGTCCATGTGTGTCGTCCTTACTTATTTATTTATTTGTCCTGTCGAGCTGTTTAATCTATTCATATGTATGTGTTGTGTGTGTCGTCCCTCGTAGCAGTTTTATTCCGTCCAACCAGAGGTCGACCAATATTAAAATAAGCATGGCTTGAAGCAGGCCAAGCGCCGTGTTCTAACCCCGTTTTAGGGGAAGTTACTTAACCTAAAAATACAACTTTTCC (SEQ ID NO: 1310)	SGXTGREVKCITVNVLMEQQPHTKAIREGDFIVILLPQSDDETLCCPLCVGRGRYSGKTRVECLNRHVKEVHPDLTTTFRCWGCGFAAPGDKKYPRKIVTQHCATCVPEVSSAPSGRVDGERRVNTRRRLGIAAATEASPVRRTRRNGLASPPVEQNISQSAAPPEPARVPQHPEIVALGESADDEVFRSPVNSPPRDWRAAAPQQAASSSPXAVPGITAATPSNTTRTGNGSAXSILAEHPIPAPPPTNTTEANGRADIPRSGRAPPPGXQAARRRAPTTEQRRIVGLLEAATGREQLEEATTQAMLFLARLTGRRPEPRNAIRPGXRQRHPAQGDVQAQAPDRIXEAKKLQRLYRTSKKRAVQKILAGPXMNCQIDKNTITAHFVELAARRDGGEDWPDVFDREEPTAASGE

						ALCTPITREEVFRRLKGRNN TSPGPDGITYRDLAKAXPGA HVLAALYNXIWRIEATPAL WGVSNTTLIYKKGDAMDIS NWRPISLGDTVPKLFAAILA DRIKRWAVANGRYSASQKG FLEFEGCYEHNFVLQEAIRE AKGGRKELVVAWLDLASAF TSVPHSSILQALEGHGLPSK ARNIISSLYTGMTTRFHTAE GPTDPILIQSGVRQGCPLSPD VFNLTLEVVLREIQRTGEGY TIEGRRISHLAYADDVAILA DSPAGMRRLLFAAERGARA VGLTFNPAKCATLHIAGRGE EAVRPTEFSVQGTPVRALAS GEAYEHLGIPTGYQVRQTPI NTLRDLLADIGSIDRSLLAX WQKLDAVGTFLLPRLDFTM QGAHIDKGFLTEADKIIKKA AKSWLSLPQRASAELVFLPP SQGGGGLLTVAHSYKMLYS SDVTVSTIAGSTLRRTVSERL

						KKRASNIDIARFLSGDLDLP RSTSPSTFWTKVRSAALRIK TKLGLRWSWCQGGEVLLM ACGDPRAPGTRVSPQTKHL VTTSLRRCLNRHYAESLLAK KDQGKVFEVTRRSGQSNHF LRSGSFTRFCDWRFIHRARL DVLPLNAAKRWQRGMDKR CRRCGSDLETLPHVLSHCGP HSAARQKRHNNIQDRLVKA ASRCPGTISVNQTVVGVRGP DAALRPDIVVRDDVNRRVTI VDVAVPFENRLEAFDGVRE AKIAKYTPLARQLTDSGYTV TVEAFVVGALGAWDPRNER VLSLLSISRYYAILMRRLMV SDTIRWSRDIYVEHVSGIRQ YRE (SEQ ID NO: 1432)
NeSL	Utopia-1_PCa	.	Phytophthora capsici	GCCCGCCGGTGGAGT AGCCATGTTGGCCACC ACCGCCCAAGTCTCCG CCGCAGCTGCGACTG CTGCTGCTCATCGAGT	TAGACGGCACAGTTCT GGCCCACGTAGGCCGA AAGGGCCCCACCCATG TAGGGAACCGCCCTCG GGAAATCCATTCCGGT	MVVSRITARLEATPAPRWDP PLPRRVIASRIADRLVPATAP CRSALNAAFPSPSRDTVTES FTQEDRQLEPLTRHVDEETK DSELPGRAPTVLDEESKDND

				CGCAGTAGTCGCAGC	GTTCGACTGAAGAGAT	ATAGEWLLRFDGACRANPG
				TGCACAAGCACCACC	GCTCCTTCGCCTTGACG	PGGAGAVLFNPSGAASWTC
				TCCTGCCGGACGCGCC	GAGGTACATCTCGACA	SHFMPGATETNNTAEYTAFL
				GCCGTGGAGCACCAC	GTCGAACTTCAACTCGC	LGARAAADHGATLLRVQGD
				GCGCGCGCTGAGCCG	AACATATCCGATACAG	SQLVLRQVKGIYGAKSTRLR
				TACCAAGACCAAGGN	TTACAAACCACAGTTA	RLRDAVRAELARVGQFSLH
				TTCCAGGCTCGCGCGC	GATATCAGATAGGAAC	HIDRQDNAHADRLANRALD
				GCGTGGMGCTCCCAG	CTTCCTTTAGGAAGCTA	MKSTLVECATHPGRNACTT
				CCAGCGACGCGTTCA	ACGGGTACACTGGATG	TLTTSAAAESPASPPPVGAR
				GCAGCAGGGTCGGCT	GTAAATACACATACAT	DTPMADAGEERLADVDDGE
				GCGGCAGCTCGACGC	TTC (SEQ ID NO: 1311)	VYAAMRLGPGEVPERRPRL
				TGCACGGGGACGGCA		RLRQLSDEELEAASEMVERL
				GCGACGGCAGGCACC		GAALSAKITDAEDWASAEG
				AAGACCACGACGACC		YITALPYMLYDKLQSYSQAP
				AAGCCGTCGCTGCCGT		RGPQQPVLTRSPRGDDRPAS
				CCCCATGGACGTCGA		SEPNASSTTGGVASEHQPRR
				CCAAGGTGCTCGGTG		RRRRGRRKGRRQRRNPRRS
				GCCGACAGCGGATCA		GREGATGGHQQHKKHKPRP
				MCACCCGCTGTCGCC		PRETQHHREHRLDEALDEL
				GCCACCGGAGCTCCG		HALERTDPHNRPAIAKARRR
				AGTCGGCGGCAAGCG		VGRIRSAINQQLLRHKFDTD
				CCGCCGCCTGAACGA		EKACVDGILSTARAERAAR
				CGGCGACGACGAAGA		AATPSPPASGAPTTTVSAPG

				CATCCGCGAGCTGGC		AIVTNDDGTCPIPSDKLWRH
				CGAGCTTCTGCTGCCC		FDAVNTPRLDFDAEAPGSA
				GACGAGGAGGAGGCC		AFRAAMDHLPAATRLLDLL
				GACGACCACAAACCA		KEAPSTDEIETQLQHVKASS
				GCGCCCAGGTTACCC		SPGLDGVGYDVYKRFTIQLL
				GCGACCAGCGCTCAT		PVLRAAFRCCWLYKKVPQS
				CCGGCCTCCGTCCTCG		WKLGVVRLLHKKGPREDPA
				CTGTGTACGCGCACA		NWRPICLQQAIYKLYTGILA
				ACGCGCAGCGCTTCA		RRLTRWMDANDRHAPGQK
				ACTGCACGTTGTGCGT		GFRAVNGCGEHNFLAATLID
				GTACACGGCTGCCAG		NARRKHRPLYEVWYDFRNA
				CTTCGCTGCTCTTACG		FGSVPFALLWDSLQRLGVPP
				CGACACAGGGACTCT		DYVDMCKGLYNQASFVVG
				CGGCACCGGCGCGTG		NAVDGSTAPVEQRVGVFQG
				ACCTTCCTGGACAGGT		CPLSPQLFNAAISPLLYALRR
				TCTCGGCGGGTTGCGC		LPDTGVQLSSVDRPGASAY
				GTGCGGCAAACCTTTT		ADDLKIFSGTKAGITQQHEL
				GCCTCGAGGCTGGCC		VATFLRWTGMQANPAKCRS
				GCAGCAAGACACGCA		MGVRRNTNGAVEADNVHL
				CAAACGTGCGCCAGC		ELDDTPIPSMTHMQSYTYLG
				CTCAGCACCACACTG		IGDGFDHVRRRVELAPKLKT
				GTCGCGGTTTCGACGA		LKHDVTALVESGLAPWQVV
				CAGGAGGAGCATCAA		KAVKVYLYPRVEYALRHLR

				GCCACACTGTCGTCGG AGCCAACACCACCGT CGCCACGGCGGTCAC CGCCGAACCCCCCCTG CTCCACCATCAAGCCT CGGAACTCACTGTGCC CCCCCCCACGTGTGAG TTCCCCCGACGTCGAT GTGCAGCTGCACAGT CCGCCACAGGAAGAT CAGCACGAGGACGCC ACCCANCACCCGGAG AGCACGCAACACCAA CCTCCTGAGGCAACCC GCTGGGGTTCGCCGCT CGCGCCCACGCTCGTT GCCTCCAGGATTGCTC AGCGACTCGGCGAGC TGGAACCTCCACGCTG GGGCCCACCATTACCC CGCGCG (SEQ ID NO: 1188)		PDDQHLESFDLHLRRGLRHL LRLPKNATNEFFYAPVSRGG LGLLPLVELHAALQIAHGW QTLHSPDPAIRRVAREQLYQ IADARHRLDKDHWPHRREE LCELLLNGELGTSAHAPPKR RNGDIGSLWVDVRKNLKTF GLKVATAPANQETGVPAQP LQLRVPHHAEWLDHGNVLR HVKLHIKNLHWQTWCALSD QGKTARVHGGVGSAFLTRP RGMWESDYRFAVAARLNV VDTVNTLSRRRLRAHDRCR YPACRWKETLAHVLNHCPG TMDAVRGRHDDALKEIEHT LRASSGDRRELRVNQTVPGL PGPPLRPDIQVYNHDKRTVA VVDLAVAFDEQPSEDPESSG LAKAVQIKKAKYAGIKEHL ENQGWKVHLSAIVYGSLGS VAASNHKVYTEHLGLLKRD AKRLDRQLSSACIQSSRRIW NFHCAKHRARQHEHQAPPS

						QATRGRRVTETGGNPSRTD RR (SEQ ID NO: 1433)
NeSL	Utopia-1_PI	.	Phytophthora infestans	AGCTCGGCCTCGCGG	TAAGCTGGTCATCATG	MLADPAALAAGLARAPPPP
				CTGCCTTCCCAGGCGC	ACCGACAGGGCACTAC	SAPQDPSPAFPAGPAGQNPR
				CGCCGACTTCGCGCTC	CCAGGTAGGGAACCGC	AAAPARVEVHTVVAPPGRA
				TGGCGCGGCCCACAC	CCTTAAAAAACCCAGG	GGMLPDPGLVEEPIQATYAH
				GCCGCCGCCGAGCCT	AAGACACAAACACCCT	DAAQFECALCPYVAESMAV
				CCAAGCGCGCCCGTT	CCACTTAGTGACATAC	LVQHRRSAHRGTRFKDIFTS
				GGCTTTCGCAGACGC	ATATTTTAGCCTAGATT	GCQCSLVFYARIVAASHAV
				AGGGCTCGCGGCGAC	TCAGTTACGGAGAGGT	ACARRNQRAVPPAPTPVAPT
				GGCCCTCGCCAGCCCC	TACTAACTGGTAAATA	RPEATPQPTGYLAAAMTAA
				CAAGACCCCCCCTAC	CGAACACATATTCTGTT	AAAASSDTVVAAATNMQSA
				GATGTGGCACCACCC	CTAATCAGTGTGAAAA	VPAAAKTTGLQLVPPELEPA
				GGCAGGGCGGCCGGC	CTGGTTTTCGCCTTTTG	LPQRASCHAGKRRRLNADE
				AGGCTGCCCGACTCG	GCGGACTTTTTCACTCG	AVTPCTPTARVSPQTEVAM
				GTATCGCCGGGTGCTA	CATTTTTGGGCAATCGT	APHDAPQDDTVLQREAAEP
				CACTCTCAGCCGCCAC	CTGCGGCTAGCTTGCTA	QPDPAATPGAQVQRVEDTT
				AGCTCGGGCCTTGGC	GCGGCGGACGAGCGGT	AAQDDTVQQDHDADTAQV
				GGTCCGCCATTGGCCC	CTCCGGGGGCGTTCAC	SPPRRTPTRWGPRPSSTQEPS
				TTGGAGCTCGACAGC	CTTTCCCCCGCGAGGCC	PMTGEPAATLAARRPLTPAA
				GACAGCAGCGACGAC	AACTACACCGATCTTCT	TGTRATRWGPCHRAIGAAAI
				GAGGACGCTCAAGAC	CTACACTTTTCTAATTC	ARLVTGLSTEPAQPQRRQPP
				CCCCACGCCGCCGCCC	GCCTCCGTCTTCGGTCT	PPQEPPSQPEPLAAAATAAA

				CAGAACCCCCAGAAG	TCGGTTGTCGGGCTTTT	DIAATVAADIAAAAANAAM
				ACGTCGCGAGTGTGCT	TTCTTTTTGACCAATCA	DVDGGPAADETWLLRFDGA
				TGCCCCACCCGGCAG	GAGCGCGCCATGCGCC	CRRNPGPGGAGAALFAPSG
				GGCAGGCAGC (SEQ ID	TCTTCTGGCCAATCAGA	AVVWTCSHYMPSRSETNNT
				NO: 1189)	GACCGGGCCCTGTCCTC	AEYTALLLGVQSAVHHGAS
					GGACAGCGAGGCCTCC	HLEVEGDSSLVIAQVKGTFA
					ACGGCCAGCCAATCGA	CRNARLRQLRNRVRHALRS
					GTCTCGGCAGCGACGC	VDTHKLRHIDRQANAHADR
					GTCTTTCTATAGCGCAG	LANRALDQRRTSSECGTHGS
					CTGACGAGGCCGATCT	CMDSCLAVPTALAAQETPP
					GGCGGCCCCCGATTGG	AAPPSTSATPAEGNAMDDIA
					TCCGACTTTCGGCCAAT	AEIAARDEGETFPVLPIGPGS
					CAGCGACGACGAGGGG	APERQPRLRLRQLSDEERDA
					GCAGGGGTTTACACTTT	AADALQELADTMASKIEDA
					TGCCCCCGTTTCGACTT	DSWTSGEGYISSIPERIREVL
					CAACTTCAGGCCAAAA	QPYATAPPQPGRSRRQQRR
					TGGCGATTTGGACCCTC	RPPRVTRNQREHRLDEALD
					CACGCGCCGTGCCACT	DMAATQQATPRDQRAVRR
					GCTCGGCACCGGCGGC	ARRRVGRVRASMAQQELRH
					GATTCAGCGGGTGCAA	EFAKDESKCVAKILKTASTE
					CTTCGGGCACGTGTGC	TAAEDEHPETCPIDAATLHA
					AACACATGCAGCGCCC	HFTGVNAPRTDFDYDATSG
					ATTGCACGCCAAGCGG	REFRAAMSDLPPATVEIDAF

					CATCGCGGGACGACGC CTCGGCCGCCCAAGCG CAGCCCCGCCCTTCCAG CACGACCTCGCGCCGTT TGGCGGATCGCCATCA AGACGTGCGAGAGCCA GGCGGGGTCGGGCAAA ATATACTTACTCTAAGT ATGCCCGAATCCCTGCC CTCTCAGGCTGAACGC GGCCCCATACTTGATCT AAGTATGGGAGGATCC CTGGCCTCTCAGGCTGT ACGCGAGACCC (SEQ ID NO: 1312)	DAELTIDEVEDQLTRAAKTS SPGHDGIDYGIYSRFAAQLV PLLHAVFQFCWRHRRVPRL WKVGIVRLIHKKGDPRQPT NWRPICLQPTIYKLYSGLLA HRLSRWLEGNDRLPMAQKG FRAFNGCHEHNFMATTLLD QTRRQHRKLYQVWYDLRN AFGSLPQQLMWRVLRHLGV DSGFIDRCRDIYRDSAFVVA NAADGATDPVRQEVGVYQ GCPLSPLLFVAALVPLVRRL EKLDGVGVPLADGVRPCTT AYADDLKVFSDSAAGIRKC HDTVAGFLAWTGLRANPGK CASLAVTTNARGNPTRDSS MRLEVHDAAITTLSLHESYR YLGVGDGYDHVRHRLQLEP KLKQLKREAVALLTSGLAP WQVVRALKVYVYPKVEYA LRHLRPLQSQLQAFDRVVSK GLRHLLSLPRSATSEVLYAP TSSGGLGLQPLVELHRALQL

						AHAWQMLHSKDPAIQAVA RAQACQVVRKRYRLQEDH WRGRDDELVRSFLNSELAA SPHAEVLRRNGDIASLWSDV QRWLRIYHLRFEHCDETEA HGPLSFRVPHHNKWLTHKT VLRHVKLHLKIRHQTRWKG MVDQGKTVRVHGGVGAKF MTTGAGLSDDDYRFGVKGR LNQVDTNSVLKRKRLRAHT TCRDPTCSSAETLAHVLNHC ESNMDAIRQRHDDALEQIGS KIRGALDRAKSPTELRLNQT VPEYTGAALRPDIVLRNVAA KTMVIADLAVTFEDQAARA RHSSLQLSHDHKTLKYQPIV AELQHKGWRVQTAAIVYGT LGSVQPSNFKAYTEKFKLHK REARQLDLQLSSHCIRASHRI WGWHCRQHRDRQRSGTAS RASRGSGGAPRRTSQAPAR R (SEQ ID NO: 1434)

NeSL	Utopia-1_PMi	.	Patiria miniata	CTGATGTGGATACCTT GGAATTACTCAACCGT GTCGGAGTCTTTTGTC TTTTGCGCCCAACACC TCATGGATACCATGCT TGTCGCTGGAGCGAC GTTACAAGCGTGAGG GCGCCCTCCATGCCGG ACAGCTGGTCTGTGCC (SEQ ID NO: 1190)	GATTAGCGAACACTAA TATCCTGCCMTAGACG TGATTGCTAATCCGCAA ACCAACCGGATCTACA GCCTGAACACTGAACT TTAATCTTCACCCATGT CACATCTGGACACTAG GTTTTTGCTCTGTTTGT GTTTTCCTGCCTTTWCM TTGGAWCTTTCGCMCT GGAATTTATTTGTCGCT TGGATTATTTTTTTTCT CACAATTTGGATCTATT TTCGTTTGTTCACTTCG AACTCTAGCTGCCCTTT CTTCGGACACTGAACTT TAATCTTCGCCATGGCT GTCAGTCGCCGGTTCAC TTGCTGCGGTGGGATCC TGTTGTGATAATCCCCG TGCATTGCCCATGGATT TATTTCCGCCTTAGTTG TTCCTAACCTTGGATTT	MCLKSFSSTSGLRRHMARL HRQPSPDASTPSTMTEVFPY PLPKVWPCVVCRENFYHNQ TLKRHQKNFHPNVDLTTVY QCSVCGQEFVTGRKASFHF KVHRRMSASAIPTPSAMPSS PMDLIRGLVGEPLPPSPARTP PPLARYISPAPRSSFSPPWNP SPPPRSPTPLPRPLTPPPRSPS PPPRSPTPPPPVTLTTAPVTE PAVPVALTTAQVTEPSAPAV HTAAPVTLSNAPVTEPATPA TDPATPVTRLHSPVTHISCSI SFTASHAPYSCAAPTSPSVY ACSPRRRQCSSTIAAVCNSE ASSGNPCLLALPVHRHHLPD TSPQRPGLLFHHPGIPPHRPG RPPHCHGHSLHRHDHRARR PGRQHHLPRSPSPPPRSPSPP SRSPSPPPRSPSPPPRPSTPPP RSPSPTPRSPSPPPRPPILPPRS PDSTHRSLTSHARSPSRPATP PIPVQPRHLRPSTPAHPGVD

					ATTGCTGTGGGTGATGC	NAVPPSQQSAIDVWLAELSR
					CCGGGTTTTGTTTACAT	SADFESFEDVCDRFVEFAAA
					CGGGATCCCGCTGCGG	EGRNNGRPARPAHQPPRDR
					CTCGGTGTTCCATGCGA	GNQGPRPQRPPRPHRPGLGP
					CTGGCAGCCCCTTTGTT	EFDAQEASRLQKLYRTSKK
					TACTCTSGACTCTATTC	RAIRTILTGSDVRYSGYRGP
					ATTGTTGTATTTCTTCC	HPALMSDTAATEVLTSGFLS
					ACACTGGCCAGTGATC	LEVFSAREVDTDTIATDTSL
					ACACTGCTGTGTTTCCC	LFPNSAQARESGQDLLRPVT
					GGGAAGATATCCTCTG	QREVSLRLGRMSNSAPGKD
					CGGTTTTCACGCTCTGG	RLEYRHIRQVDGAFRVTLEI
					GTGTCTCCCCGGGCAA	FNRCLRESRVPSSWKTATTV
					CGCACTGGTTGCTTGCT	LIHKKGDATDPANFRPIALQ
					GCGCCATCACCCTTTTC	SCLYKLLMAILSDRVTTWA
					GTTTATATTCATTTTCA	LDNDLISSSQKSARPGEGCY
					GTCTGCCGTTATCTTGG	EHTFLLSTVVKDARRNQKN
					CCAGCGCTCATTCTTTT	MYAAWLDLQNAFGSIPHDA
					GTGATGGCCGTGGACT	MFTVLTSIGAPEGLVSLVRD
					GACCCTCTGCGGTTTTC	VYTDASTDFVTPTGRTAAVP
					TGCGGTCASCACTCTCG	IHSGVKQGCPISPVLFNLTLE
					GTGAATGCTGTGCCAC	LIIRAVNASATRDRSAPVVH
					CATTTTTCATTTGTTTA	GQAVPILAYADDLVILSRSS
					CTTTTTCAGCTACAATT	DGLQSLLTTASIMATKIQLK

					ATCCTGGCCAGCTTTCA	FKPAKCASLSLECRRGTKVR
					CTCTTTTGTGATGGCCG	PLEFNVQDKIIPALTEEQHYR
					TGGACGGACCCTCTGCT	YLGVPIGLYRTDDSLETLVA
					GGTTTTCACGCTCCGGG	KMTDDIQRIDSSLLAPWQKL
					TTTGTCTGCGGTCAGCA	DAIRTFVQPCLAYTLRAGDC
					CCCTTGGTGAGTGCTGG	AKKHLKRLRGQLVKTARKV
					TCGTTTGTTTGCTCATT	CNLPTRATTNYIFADRRAGG
					TTGCTTAGTTCACCATT	LGFIDPNVDADIQIITQAVRM
					ATCTTTMTTCTCTTTTG	LSSPDDITRAIATGQLSSVVH
					TWTGGTTTTCCTAGCG	RTIHRAPTQEETDEFLSASM
					GTTGTCTGGGAGTTGA	EGDFANSGNSGQASSLWSR
					GCTGCAGTTGTCTGGTC	ARAAARRLKVTISGSLSGSV
					TTGGTTTTACCCCCATT	ITKSTENREMAAKSITTALR
					TGTTTTCTTTTAACGCG	AQSRAHYTHRLLSLPDQGK
					GGGCGTATTGCCTTGAC	VGQSLNQDQYMNSSSWMSS
					CGGCCGTCTCAGCTTTT	GSYILFCDWRFIHRARLNTL
					CTCCTAGAGCAACCTTC	PTNATAQRWKPNTSPACRR
					CGTTCATCCAACTTTTA	CQHPQETLPHILNHCPPNMV
					GTTTTCTCAGTTCTTGG	PIRRRHNLVQQRIVSAVRHG
					CCATTCCGGTTGGTTAA	RVFVDQHVPEDPNPRERPDI
					TTTTTATTTATACTTAA	TVVEGDKVTIIDVCCPFDNG
					TTTTATGTTTACATTTT	RDALMTAAAAKETKYADL
					CTGGTTGGAGACCATTT	KQALVAAGKDVEVFGFAVG

					TAGCTTGTTTTAATAGC TTTTCTTCTTTAATTAA TACCCTCTGCCATTGAG GGTTTTTATTACTATTA ATTTTGTTTACTCTTTG TAACTTGTTTGATTGAA TATTTTAATAAACCAC (SEQ ID NO: 1313)	SLGSWLPSNERALRRLGIAK RFRTLMRKLLRIDAIKGSRD VYIEHMCGHRQYT (SEQ ID NO: 1435)
NeSL	Utopia-1_PS	.	Phytophthora sojae	AGACGAGCAACGCGC TGGGGCCCAAGACCT GGCACGAACGACACT GCCCAGGCTGTGAAC GACGAGCACGCTGCT AACCCGGCAGCGCAC CGGCCTCTGGGCTCCG CTGCACCGGTTACCGG TGCAACACGGTGGGG TCCACGTCACGGCGC GATCGGGGCAGCGGC TGTAGCTCGCCTGCTC ACAGGCCTACCCACG GCACCAGCACCAGCT ACGCGCCGGCCTGCTT	TGAAGCTGCACAAGCG CGAGGCTCGACAGCTG GACCTTCAGCTGTCGA GCCACTGCATCCGCGC CAGCCACCGCACCTGG GGCTGGTACTGCCGGC GCCACCGCGAAGGACA ACGGAGCGGCAACGCC TCGCGAGCGCCGCGTG GGTCTGGGGGGGGCCC GCGGCGCACATCGCAG GCTCGGGCACGGCGGT AAGCTGGTCATTTGACC AACAGGGCACTACCCA GGTAGGGAACCGCCCT	MDVDGGPAMPEPWVLRFD GACRRNPGPGGAGAALFKP CGTVAWTCSHYMPNSSETN NTAEYTALLLGVQSAVHHG ASHLEIEGDSHLVVAQVKGT FACRNPRLRQLRNRVRHAL RAVTSLTLKHIDRKANAHA DRLANRALDLKRSLAECGE HQGAMESCLHMNPAAQRQ REQPAPPARPACAPTRAESA SDHDEDIDAEIAARDGGEAF PTLPIGPGTAPARQPRLRLRQ LTEDEQEAAASALQAMAEE LACKIEDADSWTSGDGYISA IPSRIRQLLQPFTAAQPHPRP

				CGGCTCGGCGCTGCCC AGACCCGCCCGCTCCC CCCGCGGCCACGACG ACAGCCCCCGATGCG (SEQ ID NO: 1191)	TCAAAAACCCAGGAAG ACACAAACACCCTCCC TTTAGTGACATGCATAT TTTAGCCTACATTTCAG TTACGGAGAGGTTACT AACTGGTAAACACGAA CACACAT (SEQ ID NO: 1314)	PLQQQRQRPPRVTRTQREHR LDEALDEMAAVQQERPTSR SAVRRARRRVGRIRASMRQ QQLRHDFARNESKCVEDILR AASAETAAEEHPETCPIDSG TLHEHFTAVNSPRINFLPDE ACGALFREAMADVGTPQER RSALTDELTMDEVEDQLMQ AATNSSPGHDGVGYDIYKK FAAQLVPLLHAAFQSCWRH HRVPALWKVGFVRLIHKKG DPNDPANWRPICLQTAIYKL YSGLLARRLSAYLEANGLLL MAQKGFRAYNGCHEHNFV ATTLLDQTRRMRRRLYQVW YDLRNAFGSVHQDMLWYV LRLLGVERAFVERCDDIYED SYFVVGNAADGATEPVRQE VGVYQGCPLSPLLFIAALVP LLRALEKLDGVGVALADGV RPCTTAYADDLKVFSDSAA GITRCHAVVEKFLEWTVLQ ANPGKCAFLAVTRNARGNP

						AHDKDMKLSLHDEEVSSIKL HDSYRYLGVGDGFDHVRHR LQLEPKLQQIKREAVALMQS GLAPWQVVKALKTYVYPK VEYALRHLRPLQSQLQGFD RVVAKGLRHLLRLPRSATN EVLYAPTSSGGLGLQPLVEM HRALQIAHAWQMLHSKDPA IREVARAQVWQVARKRHRL REEHWRERDDELVRCFLNS ELAASPHAEALRRHGDIGSL WSDVQRWLRIYHLSLVVQD DRNGLDPLGLRVPHHAKWL DHKSVLRHVKLHLKIRHQT RWKGLADQGKTVRAHGGV GAKFMSTWAGLSDDDYRFG VKARLNQIDTNAVLKRKRL RSHKTCRDPTCSSAETLAHV LNHCESNMDAIRQRHDDAL EQIGSKIRNALKRGKSTAEL RLNQTVPEYTGAALRPDIVL RIVAAKKMVIADLAVTFEEH AAGARHSSLQLSHDHKTLK

						YQPIVAELQLKGWQVQTAA IVYGSLGSVQPSNSTPTRKS (SEQ ID NO: 1436)
NeSL	Utopia-1_PU	ADOS01001321	Pythium ultimum	GCGGTGTACGCGCAC AACGCCGCGCTCTTCG AGTGCACGTTGTGCGC GCACACCGCGCGGGA TCTCGCCGCGCTCCAG CAGCATCGGCGCTCC GCGCACCGCAGCGTC CGCTTTGTGGATCACT TCCACAGCGGATGCG CGTGCGGCGTGAGCTT CCACTCGCGTGCGGC GGCAACCAAGCACGC GCGCGAATGTCCAGA GAGCGCGTTCTCGGTC GCCGCCGCCGCGCGC ACTGCAGCGGCCAAC ACCGCAGGTATGTCTC TCGGCGCCGACGAAC GCGACCACCTCGCGTC CGTCGGCGCCTTGCAT	TGATGCGGGTCATATTG ACCGAAAGGGCACCAT CCACGTAGGACACCGC CCTCAAAAACCCAGTT CAGTTTATTGACACCCT CCACTTAGTGACATGC ATATTCAAACCGATAC ATATTCGTTAGGAGAG GTTACTAACTGGTAATA TATCACCATTTC (SEQ ID NO: 1315)	MGTQSARERGAPSAPHSHT LGPRTPPRPPACSKHGELES AAGGRDGQCSDGAERERDA ERDIRANERDCNGDGDGDD ADSDSDDRNDARRRSRRPR ATATTTTSAPTTTTTTTSAT TSATTPATDSSPWVLRFDGA CRRNPGPGGAGAALFEPGG AVVWTVSHYLPGSETNNTA EYSAMLLGVRSAIHHGATR LRVEGDSHLALSQVRGTFA CTNRRLRKLRNRVQAALRE LGDYRLVHIDRQANAHADR LANRALDLRKTKVDCGPHA TTTDACVQPAEILAPTARLS SSSSSSSSSSDEPMPGLEEP AADDETDADAEADIAMRDG GEIFPTLQIGPGSAPAQQPRL RLRQLSDDESEAAARTLEHF ANDMASKIADADDWRSGE

				GATGTTGCATCCCCCG		GYISAIPVRLRELLAPYAVPI
				CTTTTGGCAACATTTT		RSPPRNASSRPPRPQSRPPRP
				GCCGGTTGCGTTTGCG		PRVTRHQREHRLDEALDDL
				ACCGCGGCAGACGCA		AAAQRSTSTDQRSIRNARRR
				TCAAGCGCCACCGTG		VGRIRTAQAQSDLRSQFATN
				ATCGCAGACGCAGCC		ERACVESILRAAKPDGTEPQ
				ATGCAGCACAGTGCT		ASAGTCPIDRATLHAHFAGV
				GTGCCCTCTGCTGCCG		NTPRERFDFDDALGADFRA
				CCCAATCCCCTCGGCG		ALDVLPPPDQAADAFADELS
				TGCGCACGTCCCCCCC		LGEVEDQLDRVVASSSPGLD
				GTGCCGCGCGCCACC		GVGYDVFKRFRLQLLPLLH
				ACGACACCATCCGCG		AAYQCCWRHRRVPATWKV
				CTGCGGATTGGTGGC		GLVRLLHKKGDRAEPNNWR
				AAACGCCGCCGCCTG		PICLQQAIYKVYSGLLARRL
				AACGACGACGGCGAC		SRWLEANERFTTAQKGFREF
				AACGAAAACAGCGAC		NGCHEHNFVASSLLDQTRR
				GGCCGCGACGCCGAC		LHRKLYAVWYDLRNAFGS
				ATCGAGATGCGCGCC		MPQPLMWRVLARLGVDTA
				GACGACACCGACGCA		FLQRCEDIYADSFFVVGNAA
				CCAGCGCCGACCAAC		DGATDPVRQEVGVYQGCPL
				CCCGCGACCAGTGCG		SPLLFISALIPLLRALQRLPG
				GCTGCAACGCCCGCG		VGVPLADGVRPCTTAYADD
				CGCACAGCACCAGCA		LKVFSDSAAGIQQCHGTVA

				CCAACGGATGCCGCG ACCGCGCGCCGCCGC CACACGCG (SEQ ID NO: 1192)		RFLRWTGLRANASKCALLP VTTTARGNPAIDDTLQLELH GDAIARLTLQSSYAYLGVG DGFDHVQHRVQLAPKLAEL KRDAVALLRSGLAPWQVLK AIKVYLYPRIEYALRHLRPL QSQLEGFDRAVAKGFRHLL RLPANATNELLYAPVSSGGL GLLPLVELHKALQIAHGWQ MLHSKDAAVQAIARAQVRQ VVQKRYTLDADHWQGRDD ELVQLFLNSELAASPHATIK RRNGDIGSLWSDVQRHLKT LQLRLETREPTADAPDSPNG LLHLRVPHHRKWLSHKTVL RHMKLHIRLCHKHKWQSMS DQGRTVRAHGQAGSHFVSR GVGLWDADYRFALQARLN QLDTNSTLKRRRQRTNATC RAPNCSRTETLAHVLNHCET NMDVIRQRHDGALEQIGAAI NAAIKGRRTDTEVRLNQTVP EFNGPAWRPDIQVRDARSK

						TMVIADLAITFEDQPNDQSA SSSLQHSREHKIAKYQPIAA ALERQGWRVHTSAIVYGSL RSVHPSNFTVYTELLGLLKR DARRLNTTLSCHCIRSSRRV WNWHCGQHRARQHQRCQE GRAHGSGGNQRAEGGTATT (SEQ ID NO: 1437)
NeSL	Utopia-1_SM	.	Strigamia maritima	GGAGTGTTCTTTTCGG AGACGCCGCCTACTTT AGAGGAGAGAATCCC CACGGGCATCCTCATT TGATCTGATCCATCGA GTATCTGCGAATAGTC GGCGCACTCCTTTTGC CATGATCCCGGGGGGT CTCATGGTAAAAAGG TTTGTGGCACGGCTTA GTTGACGCCCCTCTTC CACGTCACTCGGCCTG CATCGATCGACTGCTC TCTCCTCTTCCTCCCT CCCTCCTCTTCACGCT	TGATGGGAGAGTGAGG AATCTTCTCCACTGTGC AAAACCATACAGTCAG AAGATGCTAACTACTA GTTTGATACCCTGTGCC CCCTGCAATGTCCCGCG TGTCGTACCCAAGCCC GGCTGGCATTGAGACA CATTAGGCTCTCGCTCC CCCGTATACTCTCATAA TTTCGTGTACGCTAATC CTACCCTACCCCTCCCT TTGACCACTCACCCAAC CATGTGTATAGCTGTGC TGGTGATCCCGGGGCG	MATVRLKYPYPPEGILCGPC AANTNAPQTRPYSDKSGLA KHLKLYHKATLVVECRHCG HEESDLRKMKKHISTNHPV AAAAAPTVPPRLGPTAPPPP RVILRPRFIPRPRTPSPSSSSS SASSPASSRRSVSLPPASPPV SSASSPAARSGRNSPDSQGT APVTPIGTVRNSPAGSPALS YSTASPIASTITTPRHLSPASP ALSAGPGSLGASPPVSPTAA TVPPAPPATVPAVMAATVPF VAATTVPSVGSSTVPQRPAG PRRPPPFPIDDWIGRIARVSS LPELDAVSRLLEDEVVKRRP

				CTCTTCACGTGACCCC TTCCCATCCCGCCCCT CGGCTTTTGGCAAAG ATCTGTGTGTCCTCCA AAGCACCCATCTACC ATTTGCTCGAGTTGCG ATTGGTCGAAGCTGCC ACGCCACTCGTCTACT CTGCCTCTCTGACCCC TTCCCCCCTCTCTCCC TCTGCCGCCCTCGCCT CTCGGGTCCCTTCCCA TCTCCCTCCTCCCATC CACGTGCTCCCTCCTC TCGTCACGTGATCGTT GCGGCTCGAC (SEQ ID NO: 1193)	GTTATTCACTGGTTATC ATATCATTCTAAAATGA TCTTTGATCTCTCAATT AACAACTAACTTATTTC TGTTTGTTTCATTGTTT CACCTCGTAAGAGGAA GTTCATTGTGCGATAAA TCAA (SEQ ID NO: 1316)	PDPNARPASLHPTRRPPPPST RPRPCHGTGGTSVSLAALTS SCIREDHRGSLPLLCTSGWD SPWPLSPSPSSHCPSSNPCPS SSTPPSSLLGPPRHSHLTWR GSGSTTPSHRHCSRARYHHL LWRLPYPSSPRLPYPPSALA RYPSVPPALDDPLPSLSTIGS EGSPECHLCRNWTPSQGCSR MKWSRDAPLTPTPDPPHSIR HAALPLHLLALVPAMGPAE LQSLWRRSRPRAFAKITEGA SPFCALPVGTVHGHFLQVH QATAHLPTPVPLPPLRPPRSS DPLVTPISPGEVLDRLRRAT DTAPGPDTIIYSEWRAIDPTG RLLSSLFQKVQTFGAPTRWK ESTTTLIHKGGDHTAMSSW RPIALLSTVAKIYGSILSHRL TTWAVQNGRLSLSQKGFLP FRGCLDQNYLVQSCLQDAR RNKKTLSLAFLDLKNAFGSI PHLTIRHSLEWLGLAPSSIDI

						LEASFLGSSTRVRTETGLTPP ISLDTGVVQGAPLSPILFNLA IEPLLRTVPSAHSGFSLHGHV VSVVAYADDLAILAPSTPAL QSQLDAISGMADWAGLSFN PAKCATVTLTGKDNSRDTLS LQGSPVPSISDGDAYKHLGV PTGTTTFPSGTDAIKKMTTD LQAIDHSDLAPAQKLDALRT FIMPQLSFHLSHGSVPKAPLT QLDKKIKRAAKHWLFLPQR ASNEILYMSHLHGGQSLLPL SVLADIGQVTHAVALLQSR DPAVADLALRTCREVASKR AKKTVNGPELAQYLSGSTD GIYCTPTSDIPSLWTTARAAT RRLSSTLPLTWTSPLPSGVPF LSINGSPLSPFRVQSTLTNAI RHNHLSTLIAKRDQGNSYRT SHDPDPSNYWVKGGDFLRF CDWRFIHRARLNLLPVNGA RRWDANSIKTCRRCGAPNE TLAHVLNVCPVGLPEMKKR

						HDAIHARIKKALRPSPHTVV HHDRTVPGCGPLRPDILRISE RDKSVAIVDIHVPFDNGTDA VERAHETKRAKYELIRRHYE HQGYRVTFDSLVVTALGRL WRGSEAALQALQISSQYSKL LRKLLVADAIHGSRNVYAH HMTGMVM (SEQ ID NO: 1438)
NeSL	Utopia-1_SP	AAGJ02140537	Strongylocentrotus purpuratus	AAGGCTCAAACCAGG CTGCCAACCAAGCTG CCAGCTTAGGCACCA ACCAAGCTGCCAGTC CAGGCTCCAACCAGG CTGCCCACCAATCTGC CAGCTTAGGCTCCAAC CAAGCTGCCCACCTA GCTGCCAGCTTAGGC ACCAACCAAGCTGCC AGCCGAGGATCCAAC CAGGCTGCCCACCTA GCTGCCAGCTTAGGCT CCAACCAAGCCGCCA	TGACGAAATGTTCAAT ATATGACATTCTAATGT TCATGTATTTTTTGTTT GCTTGACAAAGCTAGA TGAATATATTCCCTGTC CTACTACATACATCTCG ATGTGCCCTTGCTAGAC GATCCATTGGTAACCAT GATATCAATGGATTAA CACATGATCTGTAAAC AATATATGATTTACACA ATGTATTTATTGTTTCA ATAAATCTGTTCTTTTC ACTTTAATACATGAAA	MSHSITEVFDYPLPSRWKCT VCLENFFNQQTLKRHQARH HQTTSFLYVFRCSACQAEFD SARKASNHWQSHKRKPILS QPAVNEIPSSGLDPSPPRSRP PVEVIGSSFPDDVSMLSEPST PSTSLQMDPEVVHPPSRSISF SPMHLSPTQPASPIQIGVEVS FNSSSSLQMDPEVVQPPSPSI SFSPMHLSPTQPASPIQIGVE VSFNSSGSLQMDPEVVQPPS PSISFSPMHLSPTQPVSPIQIG IEVSFTSSSPLQMDPEVVQPP SPSMSYSPMHLSLTQPDSPIP

				GCCCAGGCTCCAACC	CATGACTGTGCCTTCTC	VDIDVIPAAEVPLPDIEIPPSP
				AGGTTGCCACCCGAA	CAACTGGAGACCTACA	DRHPAAEVPLPDIEIPPSPDR
				AGTCTGCCAGTTTAGG	TAATTTGTTAAAATGAT	HPVAEVPLPDIEIPPSPDRHP
				CACCAACCAAGCTGC	ATAAATATTTCGAAGA	QSPPRPVMMEQPVHTPPPA
				CAGCCGAGGCTCCAA	TGAAATTATTATTAATA	DTQQANGPQHWVTVLANA
				CCAGGCTGCCACTCG	A (SEQ ID NO: 1317)	TNWEDFGRVCVEFANHAVE
				AGGCTCCAACCAGGC		AARSRQDAPQVRPAAQRQP
				TTTCACCCGAAGCATT		RRPTRPRQPTFDVREASRLQ
				ACCCAGGCTGCCCGTT		KLYKRSKKRAVRHILRDDA
				TAGGCACAAACCAGG		PSFSGSNEQLLDYFKEIYAPP
				CTGCCAGCCGAGGCT		EIDENRAQQLAESLFTDLEE
				CCAACCAAGCAACCA		AKESAAALMSPISQQEISTRL
				GCCCAGGTTCCAATCA		SRMSNSAPGKDRLEYRHIRQ
				AGCTGCCAGCCGAGG		ADGACRVTHIMFNRCLQEH
				CTCCAAGTAAGCTGCC		RIPSAWKEATTILIHKSGTTD
				AGCCGAGGCTCCAAC		DPANFRPIALQSCLYKLFMG
				CAGGCTGCCACTCGA		ILSDRMTQWACNHNLLSPE
				GGCTCCAACCAGGCTT		QKSARPCEGCHEHTFLLSSV
				TCACCCGAAGCATTAC		IKDTKRNQKTANIAWLDLR
				CCAGGCTGCCCGTTTA		NAFGSIPHQAIHAVLTTIGAP
				GGCACAAACCAAGCT		VSLVMLLKDTYTGASTSFLS
				GCCAGCCGAGGCTCC		TSGETDPIQIQSGVKQGCPM
				AACCAAGCAACCAGC		SAILFNLTIELIIRAVKKKAT

				CCAGGTTCCAATCAA		DDGLGLVVHGQRLSIMAYA
				GCTGCCAGCCGAGGC		DDLVLMSKTPEGLDAILSVA
				TCCAACCAAGCTGCC		SEQAETLRLAFKPTKCASLS
				AGCCGAGGCTCCAAC		LSCRHGTSVLPREYTVQGHL
				CAGGCTGCCCACCAA		MPALDEEEQYRYLGVPFGL
				GCTGCCAACTTAGGCT		PRFTNLKDLIGKLKGNIETIA
				CCAACCAAGCTGCCA		SSLLAPWQKLDAIKTFVQPG
				GCCCAGGCTCCAACC		LSFVLRAADYLKSDLRSLKS
				AGGTTGCCACCCGAA		AITTNVKKICQLPLRAANAY
				TCTCTGCCAGTTTAGG		IFAAKESGGLAFIDPNVDADI
				CACCAACCAAGCTGC		QVITQAVRVLSSDDEVVQTI
				CAGCCTAGGCTCCAA		ATSQLKSVVHRTIHAVPTEE
				CCAGGCTGCCACTCG		DIDNYLSGSNEGLLANSGNS
				AGGCTCCAACCAGGC		GQASSLWSRTRSAARRLHL
				TTTCACCCGAAGCATC		TLRATTSGTVVVNQQADID
				ACCCAGGCTGCCCTTT		HTRDILPASITRGLRLIQRTT
				TAGGCACAAATCAAG		NAEKLKSLPDQGKVARSLS
				CTGCCAGCCGAGGCT		NDPFANGSSWHATGKFIRFC
				CCAACCAAGCAACCA		DWRFIHRARLNCLPTNVAT
				GCCCAGGTTCCAATG		KRWKANANGKNGHQQETL
				AAGCTACCAGCCGAG		PHVLNHCLPNMVPIRRRHD
				GCTCCAACCAAGCTGT		NIQQRLVTAIRHGDVFVNQH
				CAGCCGAGGCTCCAA		VPGDPNPRERPDITVIEGNK

				CCAGGCTGCCCACCA AGCTGCCAGCTTAGGT ACCAACCAAGCTGCC AGCCCAGGCTCCAAC CAGGTTGCCACCCGA AACTCTGCCAGTTTAG GCACCAACCAAGCTG CCAGCCGAGGCTCCA ACCAGGCTGCCACTC GAGGCTCCAACCAGG CTTTCACCCCAAGCAT CAACCAGGCTGCCCG TTTAGGCACAAACCA AGCTGCCAGCCGAGG CTCCAACCAGGCCTCC AACCAAGCTACCAGC CGAAGCTCTGCCAGA TTAGGCACCAACCAA GCTGCCACCGTTAGA GGCCCAGAGCCCACC AATCCCATAACGTGTG AGATATGTGAAGCTTC CTTCCACACCTCCGCC		VTVIDISVPFDNGPNACTTA AQAKVEKYSALRQALRDM GRDVEVHGFIVGALGTWHQ GNERALGRLGVSRWYRTLM RKLCCIDAIQASRDIWVEHV TGHRQYE (SEQ ID NO: 1439)

				GGTCTCCGTCGCCACA CGGCCAGGCTTCATCG CACCACTGCTGAACA CACTGACGACAGC (SEQ ID NO: 1194)		
NeSL	Utopia-1_TSP	.	Trichinella spiralis	TTTCTGGTATGAATCC CAAGCGGATTCGTTAC GAAATTTGCATAAGTT TTTGAAAAAATAGGC ATTTGGTCGAGTGCTC GCACCACCATTTGTCG CGGGTCGTCCTGATAT TGCACTACATTCAGGA ACGGCCTATTCCCTTC GGGGAATTGTGTTTTA GGAATTGGAATCGGT TTGGTTACGATCGGTC GAGTGGTTCGTGAGA TCGAGTGACAGCCGG GTGGCAGCGACA (SEQ ID NO: 1195)	TGATCCGTCCCGAACC AACGGAACCACATTGC CGCATGACTTCGATTTC GCTTTTGCTCTTTTTGT ATTTAATTTTGCTATTA ACAATTCAGTTTGTTAA CTGTTTTGTATTCATTT GAAGATCCAAATAAAA C (SEQ ID NO: 1318)	MCSAKTPALKSGRRRGKEV NYEGQIVRVERRRGSRSSTS ATDLGTRMVTRGRKKLME ASVREAGHHGGESASTVDV DVVESKKITGKTARRNRRAP SGDGKRRESCGAECGQAVC GNAVADRSEASSPRTPNVSK SGRDKCGQPTIKASTPSPPK RKPTTSSSPRTPCLSKRGARS KIPSTPDTPSTSGGSGKQRVL VSPLLRTEKLPDLEVLQRTE EQVTVRATFPIAQAVVCPLG CEKPYTAVRPDGQFAHQTL TRHFMRVHNCHSVQWHYR CRNCNTDFLPADHRYPLRV VNTHVRSCVSRWEITRKLGE SEDLHGVRCDLCDYVGVSK RAVGLHRRRHANENIMQNT

						GTAAQIEALSKQVGEIRVAG DYSQFKFGKRVRQYVAPTQ RRDGLDEEEVHEAEEEEVPA ESRTILGEPSTATAIGAEEISA TGPVRADTAAQQMICRIGQ WCVWPQDYHSIPAPQCWTD TLMDLMIEQIVLQRYPDGA GVSVMSCSAVSAAIHHEISA EFAAQVMSSHDASLYCIIPV NVRNHWQMIVLDVAERVV HYYCSLREHNTVVLSSLLSL VELSGKHTGCTSWKIETHD GAPVQTNAFDCGPFSCLFLK HLLHGIDMNFGDRESAALR TDLKFMIDAVSTPVVPATDK LKKKPDGSATQLTQFQQKF LSASDNWQSPDVDLQAVYD EVVESIVSGHNEPSSRNTSRL QKKSPGKGGKGQVRRRSAV TRDPAWLKSASAVQKAFNS APARTVNAILRRPNACPSFT ATQVADHYFNLRPAVTSLA PEVIDILPPPATDHSMLVAEL

						SESEVWEKMQKAPNSAPGA DRITIRMVRMADPGAMILTR FYRACLLRKWVPLQWKQSV CKLLYKDGDKERLANWRPI ALEPVLQRVLSAVVASRVT NWARANGLISLEAQKGFQP ADGTSEHNFVMEVAIQEAR RTNAQLAISWLDISNAFGTV SHQLLFSLLERYGLDPTFTSF IQNLYKDATIVVKGANGTH VTARWSVGVRQGDPCSGIL FCLFVEPLLRSVLPSLPCEAE TTAVNVLGQPITALAYADDI ALFAPSIGVMQQQLCKIQG MASAMGFRFNPKKCASLYL NRAVVNAATFTISGEEIPAL VHGDTFRYLGVAAGLGKPQ TPFSLLRENLREAELIFRSKL APWQKMDAYRTYVLPRLTF QLMIAKFNNIKQSAGQYDR AILRLVKRCFQLPVETSTDFI RAPRQCGGLGVPSLRELYAT AKVSRALKMLWSPCRVVSS

						LAASQLQRVASAYFAKRLR DVEAADLSTFMNAARSTPL DRSGYPTCLWMDARKQMS YLTKVAGVDCYFLVGEAGT SFFIRNGLGQTVSVLSPLRK NKVMSVLGGAIQTRHLDAW LQCKRQGKTASCIVLDRSSS RFITTGRYTSFAAMRFALPA RLDLLPCRARSSMRSYQNC RRCGYDRETLPHILQHCRQF SAPAYQARHDAVQGRLETV MRRRFPNLRVNRALPEIGSN KRPDLVVVDEEKRLVILLDI AIVFENTAAAFVDARTRKW AHYEKEILAYRLRGYSVTY DAIVVGALGTWDPKNDAIL KRIGVVSQRYLRLMKVLVV SEMLEHSSRIYRKHLGLRDL LPDTGTKRRPVGTTETDPPG GDLRQKKRNTISARASGGK CLERRFTSPVGTPSQRGELQ CQPCPGPRRPALAGIAPNPPS

						LQPRKPPPRQHQKPVTKSTA H (SEQ ID NO: 1440)
NeSL	Utopia-2_CMy	.	Chelonia mydas	GGCAGAAACTGCACS TTCTAGAAGACTCACT GCCTATCCTGAGGAA GACTACCGCTTTGGAG ATGGATTCTACTGCCG CTGCTTCAAAAGAAA ATCTTCATGCTGCTTC GGAGGCTCCAGGACA GASTGAGAAGATGAT CGCTCCCTCGCCGATT CCACAGAAACCTTCA ACTGCCGCTCGGACC GCTGCTGCTCCACGGA GCGCCCATCGGGGAA CAGCCTCTAAGAGAC CCTCAGCGCTTCCAGG ACATCGCAGATGAGC AGCATCGATTGGAAA AGCAGCGCCTCCCTG AAGAAAACCCGTGGA GATGCTGCTGCAGGG	TGAGCTGGAGTGCCGA TGAGAAGCGCAGTCGG GAAAGTAACTGAAATA CTTTCCTCATGGATTGT ATTTTCTAAATGGACAA CCTACCTAATTCTCAAT TACTGAGGGACAATCT CCACTCATTGATATATT TTGCTTTCCACAACCAA ATCTCTGTACAACTTTT CATGAGTGATGTACCC GAGTACTTGGATTCTAA TATCTAAACTGTATTGT TAAATCTATTCACCTAA ATTTGGGTTATTGCTGA TTATGTACTCTATGTAT CATATGACTTTTAAAAA CAAACTTTGTATTTGTG GATAATCTAAGCACTA TACCCAGATGTACAGA CACTCTTTTCCCAACCT	MLQLRLPTPQTLRLLHPSQL PQSHSTKRWSNIYEERARAP KDTTIERVSAASKIPKLDPA KRRIGAPLQLMQGNSISRQL SASSQYVQHNAVVRRVSAP PHTGNFTGSCRRKLALPHRP PSETQPAESYLHCWTTQRDP TLADQHGLQDHPTGLPPEGS HCIKDPRQNSSTEGQRRSGN SQARRIPKRASTAASKTVLP KRTSAALKSVREDTALVLE DPAKWSSQHREGXRQQANP TAVFQPEPAEIEQQPXVRAA TPWQAAWMEELARTASFX DFDLLVDRLTKDLSAEIVSG RKGTQENTPTAHRQNQNNM REARRRNISRCYDPAAASRI QKLYRSNRPKAMREILDGPS SYCAIPSERLFLYFKGVFDR VAQNDMQRPECLXPXPRVD YAEDXEQDFTSWEVEARLT

				AWATCTTGCACCTCG AGGACGGCCTCCCAG GACATCGCAGCCAGG ACAAACATCATCTCGC CTGCTCTTCAGGASAA GGATGCCMSAAGAAC TTMTCCCACCTCCTCM ACTGCCCAGGATCCTS ATGCTGGTCGTCGTCC TGCTGTGCTGGAAGM KACCCCAACTGGCGA GACCTCAGAAACCAC CCAGMWGGACCKCA ACATCTCACTAGATGC CTGCCCAGCCGAACCT CTCCATGCWACTCTCC CAGAGCAACSAGAAC CATCCAGMGAATCCG CTGATATGACTGAAG CCMATCCAACAGAGG GAGAAGGAAAGGAGA ATGACTGCATCTATCT CCAGTATCCCCTCCCT	ATGTATTATATTTTTTT AACATTAGCTTTAATAA AATTTTTAAA (SEQ ID NO: 1319)	KTKNTAPGKDGIRYNFLKK RDPGCLVLTAIFNKCKQFRR TPSSWKKSMMVLVYKKGK QDNPNTRRPISLCSTMYKLY ASCLAARITDWSVNGGAISS IQKGFMSCKGCYEHNFVLQ TAIHMARRAWRQCAIAWLD LANAFGSMPHQHIFDMLREF GMPENFLQLVRELYEGCTTT ICSMEGETPEIPIRSGVKQGC PLSPIVFNLAMEPLIRAISSGL GGFDLYDNRVNILAYADDL VLIADNPESLQQMLDITSQA ANWMGLRFNARKCASLHID GSRRDSVQATSFQIQGEPMI FLEDGQAYQHLGTPTGFRV QQTPEDTIAEILRDVARIDSS LLAPWQKINALNTFLIPRISF VLRGSAMVKVPLNKADNTI RQLVKKWMFLPQRASNELV YISHRQGGANVPRMGDLCD VAVITHAFRLLTCPDAMVR NIAESALQDAVKKRIARTPS

				ACGGACACGCTCCTCT KCCCCTTCCGCTATCC GAGGGTTCCAGTACA TTGGCAGTCTCAGCA MACACCTCAAGAGAA TCCATASCAAGCGGAT CACCTTCCGGTGTGCC CTCTSCGACCTGCCTT TCGAGACGCAGATGA AATGTAAGTCTCATCA AGTCACCTGCAAAGG ACATCTCGAACTGGA AGAGTCCAACTTTACC AGTCTATGTTGCCGCC ACCCCATCTCTGCTCC GAAAGCAGAAACACC AC (SEQ ID NO: 1196)		NQDVATYLSGSLEGEFGRD GGDFASLWTRARNATRRLE KRIGCHWTWCEERQELGVL VPQVKNTDHTIITPRARTML ERTLKDAIRCQYVENLKRKP DQGKAFEVTCKWDASNHFL PGGSFTRFADWRFIHRARLN CVPLNGAVRHGNRDKRCRK CGYANETLPHVLCSCKPHSR AWQLRHNAIQDRLARAIPPP VGKVAVNSAIPGTDSQLRPD IVITNEDRKKIIMVDVTVPFE NRTPAFHDARARKVEKYAP LAETLRAKGYQVQTHALIV GALGAWDPSNERVLRECGI GQRYARLMRQLMVSDAIR WSRDIYIEHITGHRQYQEG (SEQ ID NO: 1441)
NeSL	Utopia-2_PCa	.	Phytophthora capsici	ACCGCCCAAGTCTCCA CCGCAGCTGCGACTG CTGCTGCTCATCGAGC CGCAGTAGTCGCAGC TGTACAAGCACCACCT	TAAGCTGGTCATTTGAC CGACAGGGCACTACCC AGGTAGGGAACCGCCC TTCAGAAACCCAGGAA GACACAAACACCCTCC	MQDMEEELLLDVEMETETT EPQTSTAXDATTTTDRPTRW GPHPRAVAAAAIAQLVTGE XAXPALPSRQDRRPAPRSHA PTRSRWGPRHQAVGAAAIA

				CCTACCGGACGCGCC	CTTTAGTGACATACATA	SLATGLPASAAPVSRATKHG
				GTCGTGGAGCACCAC	TTTTAGGCTACATTTCA	EGRRRLQTRWGPRVSIPRAA
				GCGCGCGCTGAGCCG	GTTACGGAGAGGTTAC	RRPGSRWGPPRAAGASGQL
				CNTCAAGACCAAGAA	TAACTGGTAAATAAAA	PASASGATGQLPEHVEAITT
				TTCCAGGCTCGCGCGC	AGCACTTT (SEQ ID NO:	TPRVASDADEGPTPPDPWIL
				GCGTGGCGCTTCAAG	1320)	RFDGACRRNPGPGGAGAAL
				ATGGCGACGCGACCA		FKPSGAVVWTCSHYMPSSN
				GCAGGAGTGGCCGCT		ETNNTAEYTALLLGVQSAV
				CTACTGACGGCGGAG		HHGATRLDIEGDSSLVIAQV
				ACGAGGCCAAGACCG		KGTFACRNAKLRQLRNRVR
				AGGACAGCGACGCGG		HALRSVEKYTLRHIDRKAN
				ACGCAGCCAAGCGCC		AHADRLANRALDRRSSSSEC
				AAGCCGCTCAAGCAG		EPHGSCMERCCGTDTTPAV
				CACCGAGACCCCATA		QGPTPQAAAAVPVQVWPQ
				CCCMGGACACCGCGC		WQRQTMVAWTTSHGGRCR
				CGTACGACGGCCACA		DCSTRCGRSLPSLAHRPRLS
				GGGACTCGGTGCTGG		PRRQPRLRLRQLSDEERDXA
				CTGTGTACGCACACA		ADALQELSDVMASKIVDAD
				ACGCACCTGCATTCAC		SWDTGEGYISSIPERIREVLQ
				CTGCGCGTTGTGTGTG		PYTTRPPRPGHQQQQRRRPP
				TACACAGCACGCAAC		RVTRNQREHRLDEALDDMQ
				TTCGCCGAGCTGACCA		ATQQAAPRDQRAIHRARRR
				AGCATCGCCATGCAG		VGRVRASMAKQELRQAFAK

				CGCATCGTCACACCCG		DESKCVSKILAGASAETAAE
				CTTCGTGGATCACTTC		EHVDECPIDAATLHAHFTGT
				CACAGCGGGTGCACG		NAPRTDFDYDAACGQEFRG
				TGTGGCATCGGCTTCC		ALDSMQPPTVATDAFEEELT
				AGTCGCGCGCGGCAG		IDEVEDQLTRAAKTSSPGHD
				CTACGCGACACGCTC		GIGYDIYSRFAAQLVPLLHA
				AAGCCTGTGCAGACA		AYQFCWLHRRVPALWKLGI
				GCACACACGCCACCG		VRLIHKKGDPMQPTNWRPIC
				TAGCTGCCTCGCGCGA		LQPAIYKIYSGLLARRLSRW
				CCCGGCMNCCASTGCS		MEQNQRLPMAQKGFRAFN
				SCCSGNGCCGGMAGC		GCHEHNFVATTLLDQTRRS
				GASGAGGAGGACCNC		HRRLYQVWYDLRNAFGSLP
				GCACCCCCCGGTCCTC		QQLMWSVLRHLGVDASFIA
				TSSTCGCGGCAGCATC		RCKNIYQDSAFVVANAVDG
				TGCAGCTGCAGCAGC		ATDPVRQEVGVYQGCPLSP
				CGCATCAAGCGCCAC		LLFISALVPLIRRLEKLDGVG
				TCCAGCTGCAGATACT		VPLAEGVRPCATAYADDIK
				GCCACCACGCAGAGC		VFSDSAAGIRKCHDAVTRFL
				GCCGTGCCCATCGCTA		EWTGLRANPGKCASLAVTT
				CTCCTGGGCCCCAGTA		NARGNPVRDDGVHLELQGE
				CGCCCCCCACGTGCTG		VIAPLSLHDSYRYLGVGDGF
				GAGCCACCTCCAGAG		DHVRHRLQLEPKLQQIKREA
				CTCCGAGTTTCCGGCA		VALMQSGLAGWQVVKALK

				AACGCCGMCGCCTCA ACACGCCGATCGACC TGCAGCCGCTGGACG TGGACGCGCTG (SEQ ID NO: 1197)		TFVYPKVEYALRHLRPLQSQ LQGFDRAVVRGLRHLLRLP QSATTEFFYTPTSGGGLGLQ SLVEMHQALQVAHAWQML HSKDAAVVAVAKEQVCQV ARKRYRLQEEHWRGRGDEL VRLFLNSELAASPFADCLRR NGDIGSLWTDVQRTLRLHH LSLTAQDDRDGQDPLALRV PHHTKWLDHKTVLRHVKLH MKIRHQTRWKGLVDQGKT VRVHGGLGAKFVSTGAGLS DDAYRFGVKARLNQVDTN AVLKRKRLRSSKTCRDPTCS SAETLAHALNHCASNMDAI RQRHDDALEQIGSKIRGALE RAKSTTELRLNQTVPEYTGA ALRPDIVLRNVAAKKMVIA DLAVTFEDHAAGARHSSLQ LSHDHKTLKYQPIVAELRVQ GWQVQTAAIVYGSLGSVQP SNFKTYTEKLKLHKREARQ LDLQLSSHCIQASHRIWGWH

						CRRHREGQRSGNTSRASRGS GGTPRRTSQVRARR (SEQ ID NO: 1442)
NeSL	Utopia-2_PI	.	Phytophthora infestans	GCTCGGCCTCGCGGCT GCCTTCCCAGGCGCCG CCGACTTCGCGCTCTG GCGCGGCCCACACGC CGCCGCCGAGCCTCC AAGCGCGCCCGTTGG CTTCCGCAGACGCAG GGCTCGCGGCGACGG CCCTCGCCAGCCCCCA AGACCCCCCCTACGAT GTGGCACCACCCGGC AGGGCGGCCGGCAGG CTGCCCGACTCGGTAT CGCCGGGTGCTACACT CTCAGCCGCCACAGCT CGGGCCTTGGCGGTCC GCCATTGGCCCTTGGA GCTCGACAGCGACAG CAGCGACGACGAGGA CGCTCAAGACCCCCA	TAGGCGGAAACCAGGC CCAAGACGGCCGACAG GGCCCCACCCAGGTAG GGAACCGCCCTAGAAA CCCATTTCGGTGGTCGA CTCGAAGGCCTTACCTA TTTTTTCCTTAGACATT CAATTAGGTAGCGACC AAATTACAAATTTGGT AACGAGTAAGCCAAAT GGTAATACACAAAACT TTTCTGTTCTAATCAGT GTGAAAACTGGTTTTCG CCTTTTGGCGGACTTTT TCACTCGCATTTTTGGG CAATCGTCTGCGGCTA GCTTGCTAGCGGCGGA CGAGCGGTCTCCGGGG GCGTTCACCTTTCCCCC GCGAGGCCAACTACAC	MLADPAALAAGLARAPPPP SAPQDPSPAFPAGPAGQNPR AAAPARVEVHTVVAPPGRA GGMLPDPGLVDSSPAAATA ATPAPVAATATTARAAARV AVEHHAHAEPNQEHLPMAR VLVEPMQVDECSSCDRSTLT ADDGSGDDVAAPSSMLSND VAAPMDVDSGTSCPPTLQQ PLQRPRALHVGSKRRRLDA DDGEEAHQLQEEEEAGIHAP ALRLSAASAQPASVLAVYT HNASRFDCTLCAYTAGSFAS LLTHRNSRHRRTAFLDRFSA GCACGVPFASRLAAARHAQ ACASLSSAPSAEASSAAGTS SPTADGADSTVSAVAHAEP GLPHHNDTELTASPPLVSSS DVEVQATKTEATDNRWGAP LPRVLVASRIAGRLAQVPPP

				CGCCGCCGCCCCAGG	CGATCTTCTCTACACTT	RWGPPLPRTTIAARIATRLA
				ACCCCCCGCAGACGT	TTCTAATTCGCCTCCGT	ATPAPRWDPPLPRSLVVSRI
				CGCGAGTGTGCTTGCC	CTTCGGTCTTCGGCTGT	AARLLPALPDAPACEEEAKD
				CCACCCGGCAAGGCA	CGGATTTTTTCTTTTT	SDTMDWAPTWTNEETKESE
				GGCAGC (SEQ ID NO:	GACCAATCAGAGCGCG	PHDEAPGQVDEETIDDADGE
				1198)	CCATGCGACTCTTCTGG	WLLRFDGACRANPGPGGAG
					CCAATCAGAGACCGGG	AALFKPSGPVVWTCSHYDP
					CCCTGTCCTCGGACAGC	STTATNNTAEYTALLLGAR
					GAGGCCTCCACGGCCA	AAADHGVTKLRIEGDSTLVI
					GCCAATCAAGTCTCGG	QQVRGIFATRSTRLRALRNK
					CAGCGACGCGTCTTTCT	VKLELARVGSFSLHHIDRQA
					ATAGCGCAGCTGACGA	NGHADRLANAGLDRRRTKL
					GGCCGATCTGGCGGCC	ECSVHPDGRGCTNTSVATA
					CCCGATTGGTCCGACTT	APTAPAAPLPSARPPASTAA
					TCGGCCAATCAGCGAC	PSPDDDHSDQGDIDDGEVY
					GACGAGGGGGCAGGGG	AAMCISPDAVPHRRPRLRLR
					TTTACACTTTTGCCCCC	RLTDEESEEAGNVVERLAAS
					GTTTCGACTTCAACTTC	LAAKIADAPDWETAEGYITA
					AGGCCAAAATGGCGAT	LPYALYDKLQPYSQSQHQPP
					TTCGACCCTCCACGCGC	RQQQQQQRQRPRQQQQTRQ
					CGTGCCACTGCTCGGC	RRQRRCKRGGGSQHRQRKT
					ACCGGCGGCGATTCAG	RRRRPPRVTRHHREHRIDEA
					CGGGTGCAACTTCGGG	LDDLHALESRRPQDRTAISK

					CACGTGTGCAACACAT GCAGCGCCCATTGCAC GCCAAGCGGCATCGCG GGACGACGCCTCGGCC GCTCAAGCGCAGCCCC GCCCTTCCAGCACGAC CTCGCGCCGTTTGGCGG ATCGCCATCAAGACGT GCGAGAGCCAGGCGGG GTCGGGCAAAATATAC TTACTCTAAGTATGCCC GAATCCCTGCCCTCTCA GGCTGAACGCGGCCCC ATACTTGATCTAAGTAT GGGAGGGATCCCTGGC CTCTCAGGCTGTACGCG AGACCCGTACGGCCGA ATCCCCTGGCCTCTCAG CCTGTACGCGGGGC (SEQ ID NO: 1321)	ARRRVGRIRSALDQHQLRH RFDTDEKACVDGILAAARD KDRAASVTTTAQTAAPPHS APASAPSSAVDDGICPIPGDL LHAFFTDVNTPRTEFDADSPI GARFREALAQLPAAIAATEL LMEPPSPDEVEDQLQRVRGT SSPGLDGVGYDVYKTFTQQ LLPALHAAFSRCWTDQRVP QSWKLGVVRLLFKKGDRQD PANWRPICLQQAVYKLYAG ILAHRFTRWLDANTRHADA QKGFRAVNGCGEHNFLAAT LTDNARRRRRELHVVWYDI KNAFGSVPHELLWEVLRRM GVPAQFIACCQGIYDAAAFT VGNAADGTTAPIQLRLGVF QGCPLSPHLFTAVISPLLHAL KRLPGTGVQLSAVDRPGAS AYADDLKVFSDTKDGITRQ HQLVTDFLRWTGMVANPSK CSTMSVQRDNRGVLKTANL TLQLDGAQIPALGMTEAYA

						YLGIGDGFDHVRRRVELAP KLRELKADTTALMQSGLAP WQVVKALKVYIYPRVEYAL RHLRPFQQQLQGFDRHLAR GLRHLLRLPTSATTEFLYAP TSRGGLGLLPLTEVHGALQI AHAWQTLHSPDPAIRRIARV QLRQVADARHRLDAEHWK ERGEELCERLLNSQLGTSAH APPKRRNCDIGSLWVDVQR HLRSLGLQLQTAPADTHTG APAQPLQLRVPHHDKWLTH KDVLRHVKLHIKNNHWHR WTSMRDQGKTARAHGGEG SGFLTQPRGMWEADYRFAV AGRLNQVDTYSVLKRRRLR SHDRCRQPGCHRAETLAHV LNHCPGTMDAVRGRHDGA LKRIERELHASATDRRDRVE LRVNQTVPSLAGPALRPDLQ LYNHTKKTVAVVDLAVAFE EQASDDASSSALSLIASHKR AKYDRIKRHLERQGWKVHL

						SALVYGSLGAVASGNYQVY TTHLGLLKRDAKRLDRQLS VECIQSSRRIWNLHCSQHRT RQHQARPSQGPRGSRATET GGTPSQTSRR (SEQ ID NO: 1443)
NeSL	Utopia- 2_PR	.	Phytophthora ramorum	TCAAGCCCCGCCGCC AAGCCCAGCTGCGGC TGTTGCCGCCCCTCCA GCAGCAGCAGTCGCA GCTGAACCTACAGCC CCTCCTGCTGATCGCG CCGCCGTGGAGCCCC GCGCGCGCGCCGAGC CGCCCCAAGAACAAG CACCCCCAGCTAGCG CGCGCGTGGAGCCC (SEQ ID NO: 1199)	TGAGCACCTTGGGTTGC TCAAGCGTGATGCGAA GCGGCTGGACCGGCAG CTCTCGGTGGCGTGCAT CCAGTCCAGCCGCCGC ATCTGGAACCTGCACT GCAGCCAGCACCGCGC GCGCCAGCACCAAGCA CCAGGGGGAAGTCGGG CGGCGGAGACCGGGGG GACTCCGCCGCGCACC GGCCGCCGCTAGACGG CACACAGGCCCACAGC GGCCGACAGGGCCACA CCCAGGTAGGGAACCG CCCTCAAACCCCGCCG GTACATTATGGTCCGAC	MSGDVVSSDGSSRTTDASG DGDDGAGSSDAAGDVGVV AMDVDQGARRQQPPWQRV GGKRRRLNDVDDEDTRELA ELLLEEEDEAGDHAPAPRLS AASARPASVLSVYAHNAQR FQCTLCTYTAASFASLKRHR DSRHRRTAFLDRFSAGCAC GAPFASRLAAANHAHACAS LNRTLSVAATPAAGELSPTA GAANATVKAATVTPDSPRQ DPPELAASPPLASSPDVAVQ AADMQAPTRWDPPLPRTLV ATRVASRLTDLTPPRWGPPL PRATVVSRIAARLEAAPTPR WGPPLPRVVVASRIAERLAP PELAADDETKDGEEDQSFTE

					ACCTATGAGGTGCAAC CTGTACACAAGTTACA CACCACATAGCGACTA CCAGGTATTTACTACCT GGAAGCCAAGGATTAA CCGGTCGGTAATACAC ATAACTTT (SEQ ID NO: 1322)	PVAAARSXGGEDANGEWLL RFDGACRANPGPGGAGAAL FKPSGPVVWTCSHYMPSSSE TNNTAEYTALLLGMRAAAD HGATRVHVEGDSTLVIQQV RGIFATRSTRLRGLRKSVKA EMARMEHVTLHHIDRQANG HADRLANAALDRRKTKLEC GLHPDGQGCSSTAATTAVPS VVPDRPPSSTAAAPTPSAEP DETEQGDIDDGEVYAAMCI GPDSIPERRPRLRLRQLSETE EEEAGAIVERLAATLAGKIT DASDWATAEGYITALPYTL YDKLQPFAQHRHQPRPQHR QQPQRDPPLGTHDGDHGQP STSRSRRRRRRAKDRLRRRP PRITRHHREHRLDEALDDLR AVEHASPHDRPAVARARRR VGRVNSAIAQQQLRHKFDK DEKACVDGILAAARASRGL ATPSASASRHPPPVPSTAAD DGSCPIPSDELHAFFTAVNTP

						AGTFEPMAPVGAPFRSAVA HLPAATSQPELLSDAPTTDDI EDQLQRARGSSSPGLDGVG YDIYKAFAAQLLPALHAAF ACCWRHKQVPQSWKVGVV RLLFKKGERTEPANWRPICL QQAIYKLYAGVLARRLTRW LDANGRHADTQKGFRAMN GCGEHNFLAATLVDQARRK RRELHVVWYDFANAFGSVP HDLLWEALERQGVPSPFIAC CRGLYADAVFTVGNAADGT TAPIALRVGVFQGCPLSPHL FTAAIAPLLHALKRLPDTGV QLSRVDCLGASAYADDLKIF SGTEGGTKRQHALVADFLR WTGMRANPAKCCTMSVQR DTRGVLKACNLGLQLDGAP IPALTMSASYAYLGIGDGFD HVRRRIELAPKLQELKHDAT ALLQSGLAPWQVVKAVKV YLYPRVEYALRHLRPFHQQ LEGFDRHLVRGLRHLLRLPA

						NATTAFFYAPVSRGGLGLLP LTELHAALQVAHGWQMLN SKDPAIRRIARVQLRQIADA RHRIDAQAWQDREEELAQL LLNSQLGASTGAPPKRRNG DIGSLWVDVQRHLRHLSLK LETAPACAETGTAAAMLQL RVPHHDKWLDHKTVLRHV KLHYKNKHWARWAAMXD QGKTARTHGGAGSGFLTRP RGMWEADYRFAVAARLNQ LDTHSVLKRRRLRXHDRCR QPGCTQGGDAGARAQPLRR HHGRGPRPPRRRPQAHRAR AARVVAGRPGPRRAPGQPD GAVARRPRATARPPAVQPH QEDGGGGRPGRGVRGAGER RPGELGAGTHRRTQAREVC RRQATPRAPRVEGPPLGARV RLARRGAGRQPQGAY (SEQ ID NO: 1444)
NeSL	Utopia-2_PU	.	Pythium ultimum	TGACTGGTGTTTGATC ACGATCAATGAGGTG	TAAGCGGGGGGTCCAG ACCCCACAAGAGAGAA	MDYDDSEFFDAICIPDEDAD VLDDGDEGDEGGNDDESSE

				ATTAACATGAGCCGG	GCAGGAATCATGGTCC	PLPLAITNAPSAPLHATMLC
				AGCAGGCCCCTTACA	GCATGGACCAGTAGGG	GTVTQPWLLRFDGACRRNP
				CGCTGGTGCTGTAATG	CACGCTCCACAAAGGT	GPAGAGATLTRPNGIILWTH
				GTTCAGGAATGCTCTT	TATCGCCCTCAAACCCA	YRYIPDKTATNNVAEYEALL
				ATGAGTAACTCCACA	TCACACGAAGGATCTA	DGLRCAAHHGVKHLRIEGD
				GTATAATTTTTGTTAG	AAAAGAAAGCAACAAT	SNLVIEQVKGIFACSTSLRPR
				CGGGGGTGAGCGCGT	CGAAATAGTAAATAGT	RDQVREILRHFETYSFRHIDR
				GCGCCGCCCCCACCTT	AAATAGCTTAGAAAGG	ALNRQADRLANQALDLLKT
				TCTCTTTTCGTTTGTAT	TCAACATGCGAAATGC	VSVCALSQTRVQDDTGAAH
				TTGGTTCCTCTGGAGC	ATGAGGAACGTAAGAT	GCWHWTPPDASPTDDASTSI
				GACTTCCGTGGCTTTT	GGTAATACATTTTTATA	LTQDVPVPMDIDDYDPDEP
				GTCGCGCGCTATCGCG	(SEQ ID NO: 1323)	MNAADDPVSINAEREGGTV
				CCCGGCATGAGGATC		YPVLRLGPNVVPERQKRLQI
				GTCGCTCCTGACGCCA		PWLPPREMQKLEKKIEVLGE
				TGTGCGCCAGTATATC		TFASRIRDAPDWFSAEGYIT
				GGAGCCCCGTGCTCC		ALTSELATLIRQSTAATTGP
				GCTTGGGTGTGTGTTC		NAARPCERTISKEKRRARRT
				CGCCACGACGCCACG		TPLQRALAEAKHELQIIQPD
				CGATTTGCGTGCACGG		ASRKSVRKAARRVKRISQA
				CATGCGCTGGCGACTT		QQRHDLRRLFSTNERRCVE
				CCGCTCAGTGCGCGA		KILRDPPVGPSSTSSSLPATD
				ACTCGCGGTGCATCGT		DDRCTIDPADLFAYFQTQAT
				CGACGTGCGCACCGG		APTNFDFDDEGGELFRSVLD

				TCTCTGGCCTTCCAGG		ELPRADQEVHLLEDEITRDEI
				ACTGGTATGAGTCGTC		EDQLSRISKSTAPGLDGITNA
				GTGCGCGTGCAGAGC		VYVRFKLQLLDALQAAFNA
				GCTCTTCACCGCCCGT		CWRYNRVPSMWKAAFVRLI
				CTCGACGCGTTCGTTC		YKKGNRAVPSNWRPICLQQ
				ATTCGAGCCGTTGCGA		TVYKLYTAILASRLQRWMD
				CCACAATGCGAACCA		ANARFTMSQKGFRAFNGCH
				ATCAGCGAACGCCGT		EHNFVATCLHDQTRRLRKK
				CCCCTCGTCCGCCGCC		LAIVWYDLRNAFGSLPHEYL
				TCGACAACGAGAGCG		WRVLARLGMPPQFVARVRQ
				GACGCGGATGCGACG		LYADASFTVESRDGTTDPVQ
				GCCCCCCCTCCCGCTA		LERGVYQGCPLSPYLFIAALI
				TCTCTCTTCCGCGGAC		PLVRALHKLKDQHGIVLAP
				GGGCTTCCCCCGCCCC		GVTDCVSAYADDIKIFARSG
				TGTCCTAGTTGCGCCG		TGAKALHEIVVRFLSWTNM
				CTCTCCTTCTGGATGC		AANPAKCALMVTDGARGG
				GGCAACTTGGACGAC		DDTDASMTLSIEGETIPRLTG
				CCACGTGCACGCGTG		KEGYVYLGVEDGLAHERRA
				CGCTGCCGCCAGTGC		TCLRDSLKAASADVVRLLRS
				GACGGCGTTGATCCA		DLAPWQIVRAIKSHVLSRFD
				GCGTCATCTACACGCC		YVLRHLRPFLSLFDGFDKML
				GTTCCTCCCGTATCGA		VRGIKRLCQLPQTATSEFLFS
				CAGCAACCGATCTCA		PTSAGGLGFLPLKELFAALQI

				ACGGCGATGATAGCC AACCTCCAGACCCGC GTGCCAACATCCAAC AGCTTCCAGCCATCCA TTCGACAACGACAAC GACAACGACAACGAC GACGACCCGCTGCTC GACGACGTGCTC (SEQ ID NO: 1200)		VHALQMLHSKDANVRAIAR HQALQVVRKRYALQSDHW SDREEELLEEFFNGTLERSPF ALAKKVSGDIASLWTDVRV NLTKYGLKFGEAHGRRLQP LVSHTDKQLAPQQWASAIK THMRLRHLKRWTTLVDQG KTARMHERIGSAFLTRPSGV YDASYAFAVRARLNQVDTR SALKRKRIVNNSRCRVSGCS ELETLAHVLNHCRFGSDSIR ARHAETLLLIKTTMERELTR PGRQHQRLLVDATVPEARD PVPSNDAAESNIGAMAPSIS HLRPDIQLYDNKTMEAVIID LAVAFEDQSTDDAASSSSFA RVKGVKTKKYEVIKQFLEY KGYTVHVAALVYGSLGSVD TGNFAVYTERLGLRKGAVR RLECSLSARHINFAHRMWR RHAIAHTTGLRLIGTNSVQQ QGVQRAPAEKQQHQRPVQR PSRAQAPRDQPSQQQSQQ

						SFQQQSQQSQQSQQSQQSQ QSRQRHAPTPTPVPVPVPVP VSTPTLTPTPTPTRRPKPAPS STQPQQGAPAQRRQQREQK KQPACRRRHATAVRETPPA APTAARTATPTARPTATTRS TSTTRSTATATSTTRSSAPTS RPSAPRPRSSAPRPRSSAPTT RSAAPTTRSAAPTAIATTSV YKSRRTTNAISGATTRRSAS SKRAPMQPRTALTPTQQQQ RQ (SEQ ID NO: 1445)
NeSL	Utopia-3_PI	.	Phytophthora infestans	GTACGGCCGAATCCCT GGCCTCTCAGCCTGTA CGCGGGGCTATACTTG GTCTAAGT (SEQ ID NO: 1201)	TAGACGGCAAAGTTCT GGCCCGCGTAGGCCGA AAGGGCCCCGCCCAGG TAGGGTAACGCCCTCG GGAAAACCATTCTGGT GTTTGGCTGTTTTTCAA CAGTCGAACTTCAACTC GGAGCATATCAGATAC ACTTACTCACACATTTA GATATCAGATAGGGAA CCTTTATTAGGGAGATA	MPRSLASEPVHSSASRLPSQ APPTSRSGAAHTPPPSLQAR PLASADAGLAATALASPQDP PYDVAPPGRAAGRLPDSVSP GATLSAATARALAVRHWPL ELDSDSSDDEDAQDPHAAA PGPPADVASVLAPPGRAGS MLADPAALAAGLARAPPPP SAPQDPSPALPAGPAGQNPR AAAPARVEVHTVVAPPGRA GGMLPDPGLVDSSPAAATA

					ACGGGTACACCGGATG	ATPAPVAATATTARVAVEH
					GTAAATATACAAAACC	HAHAEPNQEHLPMARVLVE
					TTCTCTGTTCTAATCAG	PMQVDECSSCDRSTLTADD
					TGTGAAAACTGGTTTTC	GSGDDVAAPSSMLSNDVAA
					GCCTTTTGGCGGACTTT	PMDVDSGTSCPPTLQQPLQR
					TTCACTCGCATTTTTGG	PRALHVGSKRRRLNADDGE
					GCAATCGTCTGCGGCT	EAHQLQEEEEAGIHAPALRL
					AGCTTGCTAGCGGCGG	SAASAQPASVLAVYTHNAS
					ACGAGCGGTCTCCGGG	RFDCTLCAYTAGSFASLKTH
					GGCGTTCACCTTTCCCC	RNSRHRRTAFLDRFSAGCAC
					CGCGAGGCCAACTACA	GVPFASRLAAARHAQACAS
					CCGATCTTCTCTACACT	LSSAPLAEASSAAGASSHTV
					TTTCTAATTCGCCTCCG	DGADSTVSAAGHAEPDLPR
					TCTTCGGTCTTCGGTTG	HNATELTASPPLVSSTDVEV
					TCGGGCTTTTTTCTTTTT	QATETEATENRWGTPLPRV
					GACCAATCAGAGCGCG	LVASRIAGRLAQVPPPRWGP
					CCATGCGACTCTTCTGG	PLPRTTIAGRIATRLAATPAP
					CCAATCAGAGACCGGG	RWSPPLPRSLVASRIAGRLLP
					CCCTGTCCTCGGACAGC	ALPDAPACEDEAKDSDEMD
					GAGGCCTCCACGGCCA	WEASEPHVEAPGPVDEETID
					GCCAATCAAGTCTCGG	DADGEWLLRFDGACRANPG
					CAGCGACGCGTCTTTCT	PGGAGAALFKPSGPVVWTC
					ATAGCGCAGCTGACGA	SHYDPSTTATNNTAEYTALL

					GGCCGATCTGGCGGCC	LGARAAADHGVTKLRVEGD
					CCCGATTGGTCCGACTT	STLVIQQVRGIFATRSTRLRA
					TCGGCCAATCAGCGAC	LRNKVKLELARVGSFSLHHI
					GACGAGGGGGCAGGGG	DRQANGHADRLANAGLDR
					TTTACACTTTTGCCCCC	RRTQLECSVHPDGRGCTNTS
					GTTTCGGCTTCAACTTC	VATAAPTASAAPSTPTRPPA
					AGGCCAAAATGGCGAT	TTAAPFHSDQGHIDEDDERR
					TTGGACCCTCCACGCGC	ADIDDGEIYAPMTLGPDEVP
					CGTGCCACTGCTCGGC	ARRPRLRLRQLSDEELEAAG
					ACCGGCGGCGATTCAG	AIVERLSASLSAKITDAEDW
					CGGGTGCAACTTCGGG	GTAEGYITALPHLLYDKLLP
					CACGTGTGCAACACAT	YSRTAPRHQRPPRPSRNQQD
					GCAGCGCCCATTGCAC	HPQPRRDQPQRNVDEQQHA
					GCCAAGCGGCATCGCG	ESQQGEDQRQQQPPTRRRR
					GGACGACGCCTCGGCC	RRGKRRGRRQRRHPRQPGQ
					GCTCAAGCGCAGCCCC	STRASQQSRQRRPRPPRVTR
					GCCCTTCCAGCACGAC	HHREHRIDEALDELHTLERA
					CTCGCGCCGTTTGGCGG	RPQDRSAIDKARRRVGRVR
					ATCGCCATCAAGACGT	GAINQHLLRHRFDTDEKAC
					GCGAGAGCCAGGCGGG	VADILEKAHAARAARTAQA
					GTCGGGCAAAATATAC	AGAATSTGGAATAPTQQAA
					TTACTCTAAGTATGCCC	TSALGDADDGTCPILADEL
					GAATCCCTGCCCTCTCA	WQYFTGTNTPRWEFNPATP

					GGCTGTACGCGGCCCC ATACTTGACCTAAGTAT GGGAGGATCCCTGGCC TCTCAGGCTGTACGCGR GGACCAAGTACAGCCG AATCCCTGGCCTCTCAG CCTGTACGCGGGGCTA TACTTGGTCTAAGTATG CCCCGGTCGCTGGCCTC TGAGCCCGTACGC (SEQ ID NO: 1324)	VGEAFRTAMARLPPATRLR ELLTEAPTADEIETQLQHVR GSSSPGLDGIGYDVYQRFAQ QLLPVLTASFKRCWTAKMV PQSWQVGVVRLLYKKGAH DDPANWRPICLQQAIYKLYT GVLARRLVRWLDVNDRHA PGQKGFRAVNGCGEHNFLA ATLIDQARRKRRSLYEVWY DFRNAFGSVPFQLLWDSLQ RLGAPADFIDMCKGLYHQA AFVIGNAADGPTAAIRQQVG VFQGCPLSPQLFNVAISPLLF ALRRLPETGVQLSGDDRVG VSAYADDLKTFSSTKAGAT KQHELVAAFLAWTGMKAN AAKCSSMGVRRNSNGATEA DNLDLALDGTPIPSMTHMQS YTYLGIGDGFDHVHRRIELA PKLKTLKQDTTALLESGLAP WQVVKAVKVYLYPRVEYA LRHLRPEDQLLESFDLHLRA GLRHLLRLPKNANNDFFYSP

						VSRGGLGLLPLVELHAALQI AHGWQMLNSTDPATRRIAR EQLHQIADARHRLDKAHWK ERGDELCQLFLNLDLGTSAH APPKRRNCDIGSLWVDVRK NLQAFGLKLETAPADAESGT PALPLQLRVPHHEKWLTHR DVLRHVKQHLKNKHWRAW CAFQDQGRTARAHGGVGSS FITRPRGMWESDYRFAVAA RLNMVDTSATLARRRLRAH DRCRYPGCRWKESLEHVLN HCPGTMDAVRGRHDGVLRE IEHALRAPSGARRELRVNQT VPGLPGPALRPDIQVYNHDQ RTVAVVDLAVAFDRQDRDD PETSGLAKAAAEKKAKYTGI QRHLERQGWKVHLSALVY GSLGSVAPNNYKVYTEHLG LLKRDAKRLDRTLSVACIQS SRRIWNLHCAKHRARQHQT PSQSRGRRVTETGGAPSRTD RR (SEQ ID NO: 1446)

NeSL	Utopia-3_PR	.	Phytophthora ramorum	TGCGCGGCAGACCAA GACGCGCAGCCAACA ACAGACCGTGCAGCG TGCGGGTGGAAAGCG CCGCCGCCTGAACGCT GGTGACGATGAAGAC CAGCGAGAGCTGGCC GAGCTCCTGCTCGTGG ACGAGGACAAGGCTG GCGCCGAACACCCCG CGCTCAGGCTGCCCAC GGCCAGCGCTCATCC GGCCTCCGTCCTCTCC GTGTACGCGCACGCT GCAACTCGCTTCGACT GCACGCTGTGCACGT ACACGGCTGCCAGCC TCGCTTCGCTCAAGCG CCACCGCTCGTCTCGG CACCGACGCACGGCC TTCCTCGACAAGTTCT TGGCGGGCTGCGCGT GCGGCACGCCCTTCGC	TAGACGGCAACACTCT GGCCCGCGTCGGCCGA AAGGGCCCCACCCACG TAGGGAACCGCCCTCG GGAAACCCAGTCTGGT GTTCGGCCAAGAAATG CACCACCACCACGGCG GAGGTGCATTTCGACA GTCGAACTTCAACCCG CCACATATCGGATATA GTTACAGCTCTAGTTAG ACATCGGATAGGAACT TCTTAGAAAATTAACG GGTATACCGGATGGTA AATAAAATAAAAACTT C (SEQ ID NO: 1325)	MQDQVDAEQQARNRWGPP LPRPLVASRVAARLGEVPPP RWGPPLPRGVVVSRIAARLE AVPVPRGGPPLPRSFVATRI ADRLAPPSPDLSLLDEEMKE SEPPDPTHHSADEDSTDAET ADAVMEPAFVSDPPTATPRE WRLQFDGACRGGPNPGGAG ALLYNPEGAVVWTGSHYMP GAKETNNSAEYTALLIGARA AADHGARQLRIEGDSLLVIR QVKGLYATKSTRLRQLRNA VRHELARVGQHSLHHIDRQ GNAFADRLANRALDLKSDK VECKEHPVAGACTTCMGSP SAGPPATPPPTTADIEMADA GSDDELRADIDDGEVYAPM RLEPGVIPTRRSRLRLRQLTD DEMEAAGEVVERLSAGLSA KIADADDWETAEGYITALPY MLYDKLQQYTQVRHGTARS PAPHPQRRDVQGQVETHRE PRHETIGQPDQPGEPSPTRRR

				ATCGAGGTTGGCCGC AGCCAGACACGCGCA AGCGTGCGCCAACCT CTGCACCACCTCGGCG ACGACTTCGACGGCA GCAAAGGCATCAAGC CCCACTGCTGCCGGA GGCAGACCCACCGTC CGTGCAGTGGTCACC GCCGCGCCCGACCTG CCCCGCCAGTATCCCT CGGAGCTCGTTGCGTC CCCCCCGCAGCCGAG CTCCACCAACGTTGCA (SEQ ID NO: 1202)		RRGKRKGRRQRRHPRRTNC GGGGRQQRKQRHPRPPRGT RHHREHRIDEAIDELHALER ARPQARPAIAKARRRVGRIR SAIDQQLLRHRFDTAEKECV DGILAAARTARDARTTVRA AAATGTTATPETAVTSGTEQ QDDNGTCPIPSEVLWRHFDS VNTPQRDFDPEAPEGAAFRS AMARLPAATRFMELLKEEP STDGIEVQLQHASSTSSPGL DGVGYDVYKRFASQLLPVL KAAFKCCWTHKQVPQSWK LGVVRLLYKKGDREDPAN WRPICLQQAIYKIYTGVLAR RLTRWQDANDRHAPGQKG FRPVNGCGEHNFLAAMLID HARRKHRPLYEVWYDFRNA FGSVPLGLLWDALERTGVP AEYIAAVQGLYDHAAFMVG NAVDGSTAPILQRVGVFQG CPLSPPLFSAAISPLLHALQL LPSSGVQLSGDDRPGVSAYA

						DDLKTFSGTKAGVTEQHEL VAMFLRWTGMAANPAKCR SMGVRRNGNGAIEADHLEL ALDDTPIPTLTHLQSYTYLGI GDGFDHVRRRVALAPKLKL LKQDATALMESGLAPWQV VKAVKGYLYPRVEYALRHL RPDDQLLESFDLHLRRGLRH LLRLPKSANNDFVYAPVSRG GLGFLPLVELHAALQIAHG WQMINSPDPAIRRIAREQLH QVADARHRLDKDHWKQRG DELCELLLNGELGTSAHAPP KRRNGDIGSLWVDVRKNLK AFGLKLATAPADPESGAPAK PLQLCVPHHAEWLDHRNVL RHVKQHMKNKRWRAWCS HVDQGRTARAHGGVGSGFL TRPRGMWESDYRFAVAARL NMLDTVNVLARRRLRAHDR CRHPGCRWKETLAHVLNHC PGTMDSIRGRHDDALKEIER TLHASSGDRQGRTELRTNQT

						VPGLAGPALRPDLQVYNHD QRTVAVVDLAIAFDEQPRD DPESSGLAKAAAEKKAKYA GIKRHLERQGWKVHLSALV YGSLGSVAPSNYKVYTEHL GLLKRDAKRLDRQLSVACI QSSRRIWNLHCAQHRARQH QDQPAPRGRRVTETGGTPSR TDRR (SEQ ID NO: 1447)
NeSL	Utopia-4_PI	AATU01001281.1	Phytophthora infestans	AGCTCGGCCTCGCGG CTGCCTTCCCAGGCGC CGCCGACTTCGCGCTC TGGCGCGGCCCACAC GCCGCCGCCGAGCCT CCAAGCGCGCCCGTT GGCTTTCGCAGACGC AGGGCTCGCGGCGAC GGCCCTCGCCAGCCCC CAAGACCCCCCCTAC GATGTGGCACCACCC GGCAGGGCGGCCGGC AGGCTGCCCGACTCG GTATCGCCGGGTGCTA	TAAACGGGTCACTTGA CCGACAGGGCACCACC CAGGTAGGGAACCGCC CTTTAAAACCCAGGAA GACACAAACACCCTCC ACATAGTGACATACAT ATTTTAGCCTAGATTTC AGTTACGGAGAGGTTA CTAACTGGTACATAAA ATTACACATTCTGTTCT AATCAGTGTGAAAACT GGTTTTCGCCTTTTGGC GGACTTTTTCACTCGCA TTTTTGGGCAATCGTCT	MLADPAALAAGLARAPPPP SAPQDPSPAFPAGPAGQNPR AAAPARVEVHTVVAPPGRA GGMLPDPGLVEEPIQATYAH DAAQFECALCPYVAESMAV LVQHRRSAHRGTRFKDIFTS GCQCSLVFYARIVAASHAV ACARRNQRAVPPAPTPVAPT RPEATPQPTGYLAAAMTAA AAAASSDTVVAAATNMQSA VPAAAKTTGLQLVPPELEPA LPQRASCHAGKRRRLNADE AVTPCTPTARVSPQTEVAM APHDAPQDDTVLQREAAEP

				CACTCTCAGCCGCTAC	GCGGCTAGCTTGCTAG	QPDPAATQGAQVQRVEDTT
				AGCTCGGGCCTTGGC	CGGCGGACGAGCGGTC	AAQDDTVQQDHDADTAQV
				GGTCCGCTATTGGCCC	TCCGGGGGCGTACACC	SPPRRTPTRWGPRPSSTQEPS
				TTGGAGCTCGACGGC	TTTCCCCCGCGAGGCCA	PMTGEPAATLAARRPLTPAA
				GACAGCAGCGACGAC	ACTACACCGATCTTCTC	TGTRATRWGPCHRAIGAAAI
				GAGGACGCTCAAGAC	TACACTTTTCTAATTCG	ARLVTGLPTEPAQPQRRQPP
				CCCCACGCCGCCGCCC	CCTTCGTCTTCGGTCTT	PPQEPPLQPEPQAAAATVAA
				CAGAACCCCCAGAAG	CGGCTGTCGGATTTTTT	DIAATVAADIAAAAANAAM
				ACGTCGCGAGTGTGCT	TCTTTTTGACCAATCAG	DVDGGPAADETWLLRFDGA
				TGCCCCACCCGGCAG	AGTGCGCCATGCGACT	CRRNPGPGGAGAALFAPSG
				GGCAGGCAGC (SEQ ID	CTTCTGGCCAATCAGA	AVVWTCSHFMPSRSETNNT
				NO: 1203)	GACCGGGCCCTGTCCTC	AEYTALLLGAQSAVHHGAK
					GGACAGCGAGGCCTCC	RLNIEGDSHLILSQVRGAFA
					ACGGCCAGCCAATCAA	CNNKRLRSLRNRVQASLRQ
					GTCTCGGCAGCGACGC	LDWYRLQHIDRKANQHADR
					GTCTTTCTATAGCGCAG	LANRALDLRRTVTECGPHA
					CTGACGAGGCCGATCT	ETRNRCFQTPQPLVEPGETH
					GGCGGCCCCCGATTGG	CVPGSDEVLAANTAMEDAT
					TCCGACTTTCGGCCAAT	AVPTEDDEAEVAARDGGEV
					CAGCGACGACGAGGGG	FPTIAIGPDSAPAKQPRLRLK
					GCAGGAGTTTACACTTT	KLDEDDFDAAAAAVTRVSE
					TGCCCCCGTTTCGACAT	ELASKIVDAGDWTSGEGYIS
					CAACTTCAGGCCAAAA	AIPERLRAALRPFALPTQPAR

					TGGCGATTTCGACCCTC	PQPREPRMQQPPRRPPRVTR
					CACGCGCCGTGCCACT	DHLEHRLDEALDTMENVQR
					GCTCGGCACCGGCGGC	STPQNQKAVRRARRRVGRL
					GATTCAGCGGGTGCAA	RSAMDRTRLRKKFATHERE
					CTTCGGGCACGTGTGC	CVAEILRRASTEEAANPSQE
					AACACATGCAGCGCCC	KCPIDRATLHEYFTATSTQR
					ATTGCACGCCAAGCGG	TPFDYDSAKGTEFRTFLEVM
					CATCGCGGGACGACGC	STPSHETSALTAEPTLDEIED
					CTCGGCCGCTCAAGCG	QLAHVKAGSSPGHDGVGYD
					CAGCCCCGCCCTTCCAG	VYRRFQVQLLPLLHAAFRFC
					CACGACCTCGCGCCGTT	WRHRRVPALWKVGFVRLL
					TGGCGGATCGCCATCA	HKKGDPQQPNNWRPICLQT
					AGACGTGCGAGAGCCA	AIYKLYSGLLARRLSKFLEA
					GGCGGGGTCGGGCAAA	NELLPMAQKGFRAFNGCHE
					ATATACTTACTCTAAGT	HNFVATTLLDQTRRMHRRL
					ATGCCCGAATCCCTGCC	YQVWYDLRNAFGSLPQQL
					CTCTCAGGCTGTACGCG	MWGVLRQLGVTEEFVARCS
					GCCCCATACTTGATCTA	GIYEDSYFVVGNASDGATEP
					AGTATGGGAGGATCCC	VRQEVGVYQGCPLSPLLFIT
					TGGCCTCTCAGGCTGTA	ALVPLLRALENQDGVGVPL
					CGCGAGACCCGTACGG	ADGVRPCATAYADDIKVFC
					CCGAATCCCTGGCTTCT	DSATGIQRCHALVTRFLEWT
					CAGCCTGTACGCGGGG	GLQANPAKCAFLPVTRSQHS

					CTATACTTGGTCTAAGT ATGCCCCGGTCGCTGG CCTCTGAGCCCGTACAC A (SEQ ID NO: 1326)	NPTRDRDIELRIHGEAIATLG LQESYRYLGVGDGFDHVRH RLQLEPKLKQIKREAVALLH SELVPWQILKALKVYIYPKV EYALRHLRPLKSQLQGFDSA IVRGLRHLLRLPENSHDGLF FSPTSAGGLGLLSLVELHEA LQVAHAWQMLHSKDPAIRA IARTQVGQVARKRFKLVEE HWRGREDDLAQRFLNTELA ASPHATETRRNGDIGSLWN DVRDTLQTLGLKFAAGDEE EAPGLLQLRVPHHTKWLSH STVLRHVKLHMKLRRMDT WKSKVSQGTTVREHGGVGS RFITAGAGLSDAEYRFAIAP RAHLIDTNSTLKRRRLRAND TCRAPGCSYTEPPAHILNKC SPNMDAIRKRHDDALERIAD ALRRKVEKSGGRLEVAINKT VPEYDGAALRPDIVLRNTET KRAIIADLAITHENQPTDATT SSALQQSRDNKITKYQTVAA

						AMMRAGWRVRVTGIVYGS LGSVLPSNFKVYTELLALLK RDARRLNRQLSSHCIRASAR IWSAHCRRHRERQRSGNAS RASRGSGGAPRRTSQASAR R (SEQ ID NO: 1448)
NeSL	Utopia-4_PR	.	Phytophthora ramorum	CCAGGAATCACCCCC GCCGCCCCCAAGCGC TCCGCCGCCGAGCCC AGCTGCGGCTGCTGCC GCCCCACTGGTGACG GCAGTCGCAGCTGCC CCTGCCACCTCTCCTG TTGGCCGCGCCGCCGT GGAGCCCCGCGCGCG CGCCGAGCCGCCCCA AGAACAAGCACCCCC AGCTAGCGCGCGCGT GGAGCCC (SEQ ID NO: 1204)	TAGACGGCACACAGGC CCACAGCGGCCGACAG GGCCACACCCAGGTAG GGAACCGCCTTCAAAC CCGGCCGGTACATTAT GGTCCGACACCTATGA GGTGCAACCGGTACAC AAGTTACACACCACAT AGCGACTACCAGGTAT TTACTACCTGGAAGCC AAGGATTAACCGGTCG GTAATACACATAACTTT (SEQ ID NO: 1327)	MSGDVVSSDGSSRTTDASG DGDDGAGSSDAAGDVGVV AMDVDQGARRQQPPWQRV GGKRRRLNDVDDEDTRELA ELLLEEEDEAGDHAPAPRLS AASARPASVLSVYAHNAQR FQCTLCTYTAASFASLKRHR DSRHRRTAFLDRFSAGCAC GVPFASRLAAANHAHACDS LNRTFSVAAAPAAGELSPTA GAANATVKAATVTPDSPRQ DPPKLAATPPLASSALVVDP DHAEQQARERWGPPLPRTL VAGRVAARLSEVPAPRWGP PLPRGVVAFRIGHRVLPPEM TSDEETKDDSSVQDGDRQD YPVAAMDVDSGMSGEWLL

						RFDGACRANPGPGGAGAAL SQPDGSVVWTCSHYMPSSS ETNNTAEYTALLLGTRAAA DHGTTTLRVEGDSTLVIQQV RGIFATRSVTLRHLRDQVKL ELARVGRFSLHHIDRQANA HADRLANRALDLRRTVSEC GVHPDGNGCTPTAIDDRPLA PTQQPPDAPPPPAADIEME DPDDEDLADIDDGEVYAAM RVGPNATPQRRRRGRSGTA KKHRRQRPPRVTRHHREHR LDEALDDLHAVERSTPSDRT TVRRARRRVGRVNSAIEQQ RLRHRFDTDEKACVTDILAK ACATREAARTTASGGDPPA GPATPAAGSADDGTCPILGE ELWRFFDSVNTPRQEFAPDA PVGAAFRSALARLPAATSCK ELLTAAPSAGEVEDQLQHV RGASSPGLDGVGYDVYQHF AAQLLPALTAAFKACWTAK RVPQSWKLGVVRLLHKKGA

						REDPANWRPICLQQAIYKLY TGLLARRLVRWLDANDRHA PGQKGFRAVNGCGEHNFLA ATLVDQARRKRRTLFEVWY DFRNAFGSVPFALLWDALA RLGVPDDYVTMCKGLYESA AFVVGNAIDGTTDPIALRVG VFQGCPLSPQLFNAAISPLLF ALQRLPATGVQLSGDDCPG ASAYADDLKIFSGTEDGIKR QHALVADFLRWTGMAANP NKCCTMSVQRDGRGVLKTD DLQLDLAGTPIPALSMSASY TYLGIGDGFDHVRRRVELAP ALKQLKDDATTLLQSGLAP WQVVKAVKTYLYPRVEYA LRHLRPFQQQLEGFDRHLAR GLRHLLRLPGNATAECFYAP VSRGGLGLLPLTELHAALQV AHGWQLLNSKDPAIRRIARV QLRQIADARHRIDSRAWEG RDEELCELLLNSQLGTSPDA PPKRRNGDIGSLWVDVQRH

						LRTLGLKFATAPACADAGS AATTLQLRVPHHDKWLDHR TVLRHVKLHVKHRHWSKW AAMRDQGKTARAHGGAGS GFLTRPRGMWEADYRFAVA ARLNQLDTHSVLKRRRLRA HDHCRQPGCSRAETLAHVL NHCAGTMDAVRGRHDDAL KHIERALHASSPGGQDRVEL RVNQTVPSLAGPALRPDLQL YNHTKKMVAVVDLAVAFE EQASDDPESSALARIAAHKR AKYAGVKRHLERQGWKVH LSALVYGSLGAVPAGNHKV LTEHLGLLKRDAKRLDRQL SVACIQSSRRIWNLHCSQHR ARQHQAPGGSRAAETGGTP PRTGRR (SEQ ID NO: 1449)
NeSL	Utopia-5_PI	.	Phytophthora infestans	CTCAAGCCTAGCAGC GGCTACCGCAGCTACT CCAGCTCCGGTAGCTG CTACTGCTACAACTGC TCGCGCTGCTGCTCGC	TGACGCACCGTGACAT AGTGCGGCACGTGAAG ATGCACATGAAGCTCC GACACTGGGCCAAGTG GGCGGCCATGCGCGAC	MARVLVEPMQVDECSSCDR STLTADDGSGDDVAAPSSLN SNDVAAPMDVDSGTDCPPA LQQPPQRPRALHVGSKRRRL DADDEEEARQLQEEEEAGIH

				GTCGCCGTGGAGCAC	CAAGGCAAGACAGCTC	APALRLSAASAQPASVLAV
				CACGCGCACGCTGAA	GTGCACATGGTGGGGT	YTHNASRFDCTLYAYTAGS
				CCGAACCAAGAACAT	TGGTAGTGGCTTCCTCA	FASLKTHRNSRHRRTAFLDR
				CTACCG (SEQ ID NO:	CACGGCCGCGAGGCCT	FSAGCACGVPFASRLAAAR
				1205)	GTGGGAAGCCGACTAC	HAQACASLSSAPLAEASSAA
					CGGTTCGCGGTGGCCG	GASSHTVDEADSTVSAAGH
					GCCGCTTAAACCAGGT	TEPDLPRHNATELTASPPLV
					AGACACGCACAGTGTC	SSPDVEVQAPETEATENRW
					CTCAAGCGCCGGCGCC	GTPLPRVLVASRIAGRLAQV
					TCCGAGCACATGACAG	PPPRWGPPLPRTTIAGRIATR
					GTGCAGACACCCAGGA	LAATPAPRWDPPLPRSLVVS
					TGCACGCGCTCCGAGA	RIAARLLPALPDAPACEEEA
					CGCTGGCGCATGTGCTT	KDSDTMDWAPTWTNEETK
					AACCACTGCGACGGAA	DSEPHDEAPGQVDEETIDDA
					CCATGGACGCAGTCCG	DGEWLLRFDGACRANPGPG
					TGGCCGGCCATGACGC	GAGAALFKPSGPVVWTCSH
					CGCACTCAAGATTATTG	YDPSTTATNNTAEYTALLLG
					AGCGTGCGCTCCTCGC	ARAAADHGVTKLRVEGDST
					ATCGTCGGCCGACCAG	LVIQQVRGIFATRSTRLRAL
					CAGGACCGTGCTGAGC	RNKVKLELARVGSFSLHHID
					TCCGCGTGAACCAGAC	RQANGHADRLANAGLDRR
					CGTGCCGTCACTCGCCG	RTKLECSVHPDGRGCTNTSV
					GCCCCGCGCTACGGCC	ATAAPTAPAAPLPPARPPAT

					CGACCTTCAGCTCTACA	TAAPSHDDDHSVQGDIDDG
					ACCACACCAAGAAGAC	EVYAAMCIGPDAVPHRRPR
					GGTGGCGGTGGTCGAC	LRLRHLTDEESEEAGDVVER
					CTGGCCGTGGCGTTGA	LAASLAAKIADAPDWETAE
					GGAGCAGGCGAGTGAC	GYITALPYALYDKLQPYSQA
					GACGCGAGTAGCTCGG	QPQPPSQQQQQQQQRPRQQ
					CACTGTCCCGGATCGCC	QQTRQRRQRRGKRGGGSQR
					AACCACAAGCGAGCCA	RQRKTRRRRPPRVTRHHRE
					AGTACGACCACATCAA	HRIDEALDDLHAIESRRPQD
					GCTACACCTCGAGCGC	RTAISKARRRVGRIRSALDQ
					CAAGGATGGAAGGTAC	HQLRHRFDTDEKVCVDAIL
					ACCTCTCGGCACTCGTG	AGARASQGATTAPPSATTDP
					TACGGGTCGCTTGGGG	PAPMDDSRCPIPGDDLWRFF
					CGGTCGCTAGTGGCAA	DSVNTPRRSFDAEAPDGAAF
					CTACCAGGTGTACACC	REAMACLPAATRAQELLTE
					ACACACCTGGGGCTAC	APTVDEVEDQIQHARASSSP
					TCAAGCGCGATGCAAA	GLDGVGYDIYKQFAAHLLP
					GCGGCTGGACCGGCAG	ALHTAFVCCWNHKRVPQS
					CTGTCTGCCTAATGCAT	WKLGVVRLLHKKGDRQDP
					CCAGTCCAGCCGCCGC	ANWRPICLQQTIYKLYAGIL
					ATCTAGAATCTACACTG	SRRFVRWLDANARHAEAQK
					CAGCCAGCACCGGACT	GFRAMNGCGEHNFLAATLV
					CGCCAACACCAGGCGA	DHARRKRKELHVVWYDLA

					GGCCCAGCCAAGGACC AAGAGGCAGCCGGGCG ACGGAGACCGGGGGGA CTCCGTCCCAGACAAG CCGCCGCTAGGCGGAA ACCAGGCCCAAGACGG CCGACAGGGCCCCACC CAGGTAGGGAACCGCC CTAGAAACCCATTTCG GTGGTCGACTCGCAAG CCTTACCTATATTTTAG ACGTAGCGACCAAATT ACAAATTTGGTAACGA GTAAGCCAAATGGTAA TACACAAAACTTT (SEQ ID NO: 1328)	NAFGSVPHDLLWETLARQG VPPTFVDCCRGIYSDAAFTIG NAADGTTAPIRLRVGVFQG CPLSPHLFTAAIAPLLHALKR LPVTGVQLTGVDRPGAAAY ADDLKTFSSSVDGIKRQHEL VATFLRWTGMAANLSKCSA MSVQRDSRGVLKTGDLCLK LNDAEIPALSMTASYAYLGI GDGFDHVRRRLELAPMMK QLKHDATALMQSGLASWQ VVKAVKVYLYPRIEYALRH LRPFKQQLEAFDEHLRRGLR HLLRLPTNATSAFFSAPTSR GGLGLLPLTELHAALQIAHG WQILNSPDGATQRIAREQLR EIPDARHRLDTAHWRNRDA ELCELLLNSQLGQSSHAPPK RRNCDIGSLWIDIRRQLGTL GLKFETAPGRRSHQPARPAI AAFACRTTTSG (SEQ ID NO: 1450)

NeSL	Utopia-5_PR	.	Phytophthora ramorum	CCAGGCATCACCCCC GCCGCCCCCAAGCGC TCCGCCGCCGAGCGC AGCTGCGGCTGCTGCC GCCCCACTGGTGACG GCAGTCGCAGCTGCC CCTGCCACCTCTCCTG TTGGCCGCGCCGCCGT GGAGCCCCGCGCGCG CGCCGAGCCGCCCCA AGAACAAGCACCCCC AGCTAGCGCGCGCGC GGAGCCC (SEQ ID NO: 1206)	TGATGCCCGCCAACTA CAAGGTGCTTACTGAG CACCTTGGGCTGCTCAA GCGTGATGCGAAGCGG CTGGACCGGCAGCTGT CGGTGGCGTGCATCCA GTCCAGCCGCCGCATCT GGAACCTGCACTGCGC GCAGCATCGAGCACGG CAGCACCAGGGCCAAG CGCCAAGGGGCAGTCG GGCGGCGGAGACCGGG GGGACTCCGCCACAGA GCGGCCGCCGCTAAGC GGACATCGGGCCCGTA GCGGCCGACAGGGCCA CACCCATGTAGGGAAC CGCCCTCTAAACCCGCC CGGTACATTATGGTCCG ACACCTATGAGGTGCA ACCGGTACACAAGTTA CACACCACATAGCGAC TACCAGGTATTTACTAC	MSGDVVSSDGSSRTTDVSG DGDDGADGAGSSDAAGDV GVVAMDVDQGARRQRPPW QRVGGKRRRLNDVDDEDTR ELAELLLEEEDEVGAQAPAL RLFAASAHPASVLSVYAHN AQRFVCTLCAYTAASFASLK RHRDSRHRRVSFVDKFSAG CACGTPFGSRLAAARHAQA CASLSIPRTVTAPAAAGDLS PTATGANATASAAATSPDLP RPASPELAASPPQTSPFDVAI QADAAEQTAWTRWDPPLTR AAVAARVASRLAVVPAPRW GPPLSRTLVASRIAARLDAQ TSRWGPPLPPAMVASRIASR LAAMPAPRWGPPLLRTVIAS RIADRLLPPELAADEETKDD DVHMDNAASVDVDEESEV ADVVMTDHDGEWLLRFDG ACRANPGPGGAGAALFKPS GPVVWACSRYMPSSSATNN TAEYTALLLGARAAADHGA

					CTGGAAGCCAAGGATT AACCGGTCGGTAATAC ACATAACTTT (SEQ ID NO: 1329)	THLRVEGDSTLVIQQVRGIF AARSTRMRALRNQVQSELA RVGSFSLHHIDRQDNAHAD RLANRALDLRRTVIECGIHC DGVGCTATTTEVQSSSAPEI PTRPVADDHDEHEVVDVVD VCGVCGVCGDRGTCGVCD VSGDIDDGEVYAAMRTGPD AVPARRPRLRLRKLTDEEQE EAGTLAERLGATLAAKIAD ARDWESAEGYITALPYLLY DKLLPYSQGPARSLPVRQHQ RQQQQPDGQFQRPTQSRSA ARRQRRQRHRARRRPPRVT RHHREHRLDEALDDLHAVE RATPSDRRSIRRARRRVGRV NSAVEQQRLRHHFDTNEKG CVEILLAKARAQRSTTVART AVGEPNSGAAEDDGTCPIPS ERLHRHFTEVNTPGSSFDAM APVGAPFRAALAHLPAATE ASELLTEAPTPDEIEDQLQR AKGTTSPGLDGVGYDVYKA

						FSTQLLPVLHAAFQCCWQH HRVPQSRKQGIVRLLYKKGP REDPANWRPICLQQVIYKTY AGVLARRFTRWLAANGRH ADAQKGFRTVNGCGEHNFL ASTLIDHARRSRRELHMVW YDLKNAFGSVPQELLWEVL QRMGVPPAFVEVCQGLYQD AAFTVGNAADGPTDLVRQL VGVFQGCPLSPHLFTAAISPL LHALDRLKDTGVRLSADDR PGASAYADDLKIFSGTADGV KRQHALVADFLRWTGMVA NPNKCCTMSVQRDGRGVLK ACDLELQLDGARIPSLIMNA SYAYLGTGDGFDHVRRRIEL VPALMQLKDDATALLQSGL APCQVVKAVKTYLFPRVEY ALRHLRPFQQQLEGFNRHL VCGLRHLLRLPVSATTSFFF VPVSRGGLGLLPLTELHAAP ADRRPAPPPRPRPLEGAGGE NMRAADQLAARDVGPRPTQ

						APQRRHRLVVGRRPAPPPRT RPQARDRAGVRGDRHRGG DAAASRAAPREVAGPPHGP AAGEAAHEEQALAAVGRDE GPRQDRPHPWWCRERLPHA ASRPMGDRLPLCSGGSAQP AGHAQRAEAPAPPRARPVS TAGLLPCRDTGTRAESLRRH HGRGPRPPRRCAQDHRARA HRVVRAARTAPSSGSTRPCP RSPAPRCGPTSSSSTTPRRR WRWSTWPWRSRSRRATTPR ALRWRASPRTSEPSTPASSG TSSAKGGRSTSRRSCTARW AR (SEQ ID NO: 1451)
NeSL	YURE Ci	.	Ciona intestinalis	ATCAACCCACTACTAC ACCCTCTACAAAGAA CCCACTACAGCAAGA TCTTGGCGACCAACTA CACCTGCAGCCTCCCG TCGACTCTTCACCACC TGCGCACCTCCCTCCG ACTCCAAGCGGCAGA	TAGCCAAAAGATTTGG TGTTCTTGGCGAGCTGC AGTCCCGGCAGAGTAG GCGAGATCGTGAGCTC TACACCGCTTCACGATT TGACGAGTTCTTCTTGA GTCGATCCCCAACGGC AACGCACTCCAAATTT	MATSSSSVSSGNVQTEVRCV YHGKGDLLLECPVAHCPSIH PTVATITKHLKKHHTPQFEQ ITTKNLTITYTCSQCSFSTTG LTQHHISKHYKTCKGVGAV QEGNKGRFCCPACGTRWAL LCKARHHFNNVHFEYDTPPI AAFSGTPYKLKKRKFTIINK

				CAGTGGCACCACCAG	AACGACCGCATGGCGA	ALTYSCPLPLNQLLCPLWSC
				CACTCCAAGGGTCCA	CGCCACTGATCTCCCAG	SLTILNKPLSSVQQETAHGD
				CAGCCAGAAGAGTGC	TCATCCGTGCTCCGAGG	GSQGQSYVPTQLRQVLRAR
				ACCCCCCACTGCTAGG	GCACTTAAGGGGAACC	CHCGNPPIGKGHWASCQGK
				AAGAGCTAGGCCATT	GATTCACTGATGCTGCA	RPLSSPKGGRSSPTPPANLTL
				GCCCGCTGTGGGCCTT	AGTTGGCGGGGGCGCA	HFLNYLPFQLPSQSSSPQSST
				GTGCTGGCTGTCTGCC	CTCACCAATGGAAGAT	LDPTACKARVPIPSFLRGDC
				GCTGATTGGCACGTTA	GAGGGAGCCCAACGAT	EVTFFIIPSVNFYRPYLSYPL
				GGGAGTGTGCGCTGA	GCCCCGAACTACCGCA	RMFWRNRTSSGHCSLHRVV
				GCACGATAGGGAGGC	CCGTCCCCAGCACCGT	RGFVRERLVPHPRSKSPART
				TGCTGAAGCCGAAGA	GTCCTTGTGAGTAATTA	PLEFLCEFRLAGVFPDPGKV
				ACCCGCGCCGGGGAC	TTTTCACTAATCCAGGG	ASLRPVPAPLTLCLSPPVAGP
				TTTAGGACTTTAACTA	ACGGGGCCATGAACTG	MISCEDHSAPPSVRSSSPIPN
				AAATCTGTAACACTA	AACTATGTCTGCACGCT	SPASVSSVEAHLSDLLDKVS
				AAATTGGAATACTGG	GCCCCGCCTAGCCTCG	SGELRPLSPTLPSSGFFGPLL
				ATTACTATTGGATTAC	GCCACAAATAAATAGT	PPTPPPRPTPSAEKASPSGLS
				TACTGGACAATAACCT	CAAGCCGGGGCGTACT	YLPCREVKIASIARPSPASQR
				GATCTACAACAACCA	ACCAATTGATTCGGATT	VGCDADRTGPSLNPNYQQT
				ATTGGATTACAGAAC	TGGCGGGGCCCATCAC	SPPSTPSFSPIVRPPKFPRSGA
				CAAATTACAATCTCAA	GCGCCTCCGCCCCCAC	KVNSKSKPPGVRPRRAKPIE
				CTACAAACAAAAAGT	ACACAAAAACACTTTT	PGTESASPVDVDTISSSVQEP
				GAGTCCCCGGCCGGG	AAACTCTTCGGTTCCCC	CTPENRTPEFFYERKWLVSI
				TTTCTATATTACCCAA	AACCACTACAACAAAG	LNIHEREGSNFFQFNRDLEY

				TTACTGTTTTTACAAT	CGAGCGGCCCCTCTTA	WTQLLSGSQKGGRAKRASY
				TTTAAAATTTTATAGT	GATCCAATTTTAAAATT	NRGAANQAMKNRDSGRKD
				ATTTTACTAATTTTTC	TTAAATCAGTGCACTCA	FDPRPVAGHSSGGGTELGSR
				CGCCCCGCTAGCACTT	ACTTTTTTACGTGTTGT	PRYPKGARLRADFWRDMK
				AAATTGGCACCCCCCC	GTTTTTTGTATTTTTTCC	GTVRKLLDGSNGERRCGIPL
				CCCCCSTCCAAAAAAA	CACAGATTTTGTATTTT	DIIERKFRQVSMPGWIDHRR
				AATAGATAACCCTCC	TATATTATATTTTATAT	YAAGASPSLVTQAETDVAIT
				ACCACCTAAACCCCG	ACACAACACTATTTTTA	SEEVEAVLSGLNVQSAPGSD
				GTCACCCCCTCAGAAC	TACACTACCTTGCACTG	GLSYRFWKGLDPSGRLLSCL
				TAATTGGAACCCGGA	TCCCACTTTTTGTAATT	FEIVRRHGRIPGAWPTCSVIL
				TCTTTTCCCTCTATCA	ATCACCTTTTACCTTTT	LCKDAQGDVQDVGNWRPIT
				AATTTCCGGTTCAGTT	ATGCCGCTCGCTGAGCT	ICRTLYKLYAAVIARRIQTW
				TGGCTTAATTACCTAA	CCTTGCACGGATGACC	AKQGGVLSRLQKGFMPVEG
				AACCCGTCTTTATTTT	AGACAAACTTTTATAA	VFEHVFMLDTVLSDAKLRR
				ATCCCTTTATTGCCCC	AATTATAACATTGTTTT	KNLLAVFLDVRNAFGSVRH
				TTAAACCAAATTATTC	AATTGTCGCGGGGTAT	ECLLKVLRHFDAPHYLVEL
				AAATTGAAATCCGGC	CAGTGGCGCCCCCTTGC	VRDIYTGATCRVRSSVGETG
				CAAATTGCCTTGCTGA	GGCCGTGAAAGCCCTT	DIPCDRGVRQGYPLSGILFN
				ACCACCATTTTTGTTT	TTCACCTTAATTCGCCC	LVTEVLIPGLSAGNDGYRM
				TTCTTATAAAAGTAAA	CTTAGTCCAAATTTTTC	ACLGGKLTQVLAYADDLVV
				TTTTTCAACAACCCAA	CCCATTGCCTGTAAAA	VTENRDQMLRQLGVCEEFG
				AATCATAAAATTAAA	GTGCGTTGCCGTCGACT	RWAGLAFNQRKCGLIGWRT
				ACTGTTATCTGATTAG	CGAACTCGTCCTTTTTA	LRGGRRVALEDPLLLNGVEI

				TCTAAATTCTGTTATT	CCGCCTCTCTGTTACTG	PLLRPGEHYKYLGAMTGVM
				GACAAAAACCTCCAT	AACTATGACCAGCTTG	SVPRTGSQLIKDFRARLQRL
				ACCAACCTTAATCTTA	GCTGTTGAAGTCGGCTT	FTSFLTPHQKLIALKRFLLPS
				TCATTATCTTCACCCA	TAATTCGCCGGCTTCAC	LSFHLRVRPIARSELIALDRR
				GACTGCTCAATCTAGT	ATTATATTTTTTTGTTG	VRECLRVAFRLTKPSCQAVF
				TCTTTTTTCAATCAAT	GGTAATCTGTTTTTTTT	HTPTDMDGLGVPSVCSESSI
				AAGGAGGTGTTTTCTT	CATATATAAAACCATTC	LTIAQGFKVLTSPDGTVSAT
				TGGTCTGTTGTTTTAT	GGCTTAAAATCACCAA	ASARVKLYAAKFGGLTEAG
				AGTTTACCGTTATAAT	TCCCCATCATCCACCCT	PSDWARYLSGDDVNGNSTR
				AGTCTGTTGTCCTATA	GGTCACCCATTGAAAC	KPGANLPSGLWTRVRCASR
				GTTTTCCGTGGTGTAT	ATCTCTTTTTAGTCAAT	QLGAVWRVCPENGITVRVR
				TTCTGGTCTACTATAG	TATTTTTTCAGATTACC	NSVITSRDRRKLIRSFHDCSN
				TTGGTGCCTCTAAAAC	CCTCCCAAATTGTCATA	QQWKEQWMQHPNQEKTAA
				ATTATTATACCGAAGA	TAGTTTAACACCCCGTT	AHMAYADANRWVKQPSVM
				GTGCTAGCGATTTCTA	TGCCAAGTTGGTCTTTT	EPHTYFFALRARLNLLPTRV
				TTATTAACTCACCCAC	CCCCATCCCCTTCTGTC	SRAIYSRDQHPDILCRRCGA
				CAGTACTGCTGTGCAT	CTTCCGGTAATCCCATA	SVESLPHVLNHCPPNMSIILG
				CCACGCAGCACCACG	ACATTTTATTCTTTAAT	RHNLVLQEVLNAVDKTQFK
				CCTGCCATCACGTACA	TCGGCCCCATAAATAG	EISVDRTVPEHMSETGEALR
				GTGCGCTTGCTGCCTT	TGTAATCTTAAGACGA	PDIVARRNDGSVVVVDVAC
				GTGTTTTGTTTTGCAG	AGTCCCCAGAGGCCCG	PFDQKANFDEAAKRKLLKY
				ACACCACTGGACGAT	GTCCCCCCCTTCTTTGT	DKLCCNIAASTGKPVECHSI
				AG (SEQ ID NO: 1207)	CATGGACCGGGCCAGC	VVGSLGSLAEGLSTSLRALG

					CCCCCTGTTACCACAAC AACCCACCATTTTATTC TTTCTTTCTTTTTATTAG TATTTATTTATATTGCC GCCCAACTGTCGTGTCG TGGCGCAGGGGGGTTC CCTCTGTCGGCCCAGTG GACGACCGTCTAAAAA ACAGGCACAGGCAGCA GGCACTTATCACCCGGT ACCTCCGGGTACCGGA GGCTGGTTTTCAGCGCA CTGTACGTGATGGCCA ATTTTATTCATTGCATT TTATCCGCGTCGTGGTG TTTGCGTGGATGCATTA ATAAAAGATGAAATTC C (SEQ ID NO: 1330)	ITDFARSKLVACHQG (SEQ ID NO: 1452)
R2	PERER E-9	BN0008 00	Schistosoma mansoni	ATCTCACGTTTTAATT TATTTTGAACTACTG CAGTCTGAGTGCTTCT AACGACCCGAAGGCT CAGAAACTACCCACTT	TAACGGCTGAACGAAT AGCCCCCTTCACTCTTA GACATTCCCCCACTGTT GTTGCTTATCTTCATGC TCTTGTGTTAATTGACT	MPVSTGAETDITSSLPIPASSI VSPNYTLPDSSSTCLICFAIFP THNILLSHATAIHHISCPPTP VQDGSQQMSCVLCAAAFSS NRGLTQHIRHRHISEYNELIR

				CTTGAACTGCTACTTT	GCTCTCTTCTGGGTTGA	QRIAVQPTSRIWSPFDDASLL
				TTGCTGTTTATCCACA	CGTCTGATTGTCTCTCT	SIANHEAHRFPTKNDLCQHI
				ACAACAGTTGTGATTC	CTCTTTCCATATTGCTT	STILTRRTAEAVKRRLLHLQ
				TATTCTCCANATATTC	GCTCTGCCCGCTTACTT	WSRSPTAITTSSNNHTITDIP
				CTTGTGCTTTTGTCAA	CCAATAGTTGTCATATT	NTEARYIFPVDLDEHPPLSD
				CATTATTCTATACCAA	ATGTCTTTGTTTACTTG	ATTPNASTHPLPELLVILTPL
				CTGTACCACCTACTTC	CCATGTCTAACGACAA	PSPTRLQNISESQTSHESNKN
				TTCATCTCACGTTTTA	TTACTTTATCTACCTTA	SMHTPPTYACDPDETLGATP
				ATTCTGGTCTTATTTT	GTTTGTCCTCTTGGTTT	SSTIPSCFHSYQDPLAEQRGK
				CTCATCATTAGTCACG	CGATTGCCTTCATATGT	LLRASASLLQSSCTRIRSSSL
				GAGAGGGCCTATGAA	TCATGGCGGAATCTGA	LAFLQNESTLMDEEHVSTFL
				CGGTCCGTGACGCGA	TGTTTATAATGACTATT	NSHAEFVFPRTWTPSRPKHP
				AATTCTATCCGCGATT	CCTATTACCACCACTAC	SHAPANVSRKKRRKIEYAHI
				TCGACCTCTCCTGCTA	AACTACTATTATTATTT	QRLFHHRPKDASNTVLDGR
				GTGGTCCCCGAAGTA	TCATTACTATTAACATT	WRNPYVANHSMIPDFDCFW
				CGGTTCCTCTGGCCTG	ATTATAAACATTATTAC	TTVFTKTNSPDSREITPIIPMT
				TCAGTTGTGTTAAAAC	TATTATTATTATTACTA	PSLIDPILPSDVTWALKEMH
				TATATAATAACG (SEQ	TTATTACTTCTACAATT	GTAGGIDRLTSYDLMRFGK
				ID NO: 1208)	AATATTATGGCTACTCC	NGLAGYLNMLLALAYLPTN
					TCTCAGCACACCAATA	LSTARVTFVPKSSSPVSPEDF
					AAATATCAATCAAACA	RPISVAPVATRCLHKILAKR
					TCTCAATTATATCCACC	WMPLFPQERLQFAFLNRDG
					TATTAAACTCTCTCTAT	CFEAVNLLHSVIRHVHTRHT

					TTCCCCTGAGTTATAAA CTTACAATTCAGTCTAA CCGAATATCTCTCTTTT ACAAATCTTAAGTATGT AATTTTGTGCCAAACCC ATTTGGGTCTGTACAAT TTGATACTTAAAAATA AATGTTATTAGCC (SEQ ID NO: 1331)	GASFALLDISRAFDTVSHDSI IRAAKRYGAPELLCRYLNN YYRRSTSCVNRTELHPTCGV KQGDPLSPLLFIMVLDEVLE GLDPMTHLTVDGESLNYIA YADDLVVFAPNAELLQRKL DRISILLHEAGWSVNPEKSR TLDLISGGHSKITALSQTEFTI AGMRIPPLSAADTFDYLGIK FNFKGRCPVAHIDLLNNYLT EISCAPLKPQQRMKILKDNL LPRLLYPLTLGIVHLKTLKS MDRNIHTAIRKWLRLPSDTP LAYFHSPVAAGGLGILHLSS SVPFHRRKRLETLLSSPNRLL HKLPTSPTLASYSHLSQLPV RIGHETVTSREEASNSWVRR LHSSCDGKGLLLAPLSTESH AWLRYPQSIFPSVYINAVKL RGGLLSTKVRRSRGGRVTN GLNCRGGCAHHETIHHILQH CALTHDIRCKRHNELCNLV AKKLRRQKIHFLQEPCIPLEK

						TYCKPDFIIIRDSIAYVLDVT VSDDGNTHASRLLKISKYGN ERTVASIKRFLTSSGYIITSVR QTPVVLTFRGILDRASSQSL RRLCFSSRDLGDLCLSAIQG SIKIYNTYMRGT (SEQ ID NO: 1453)
R2	R2-1_BTe	.	Bombus terrestris	TCAATAGTTACTCGGG GGCAGGCGGGATATT GGTCTTGCCTTGCCCA AGTCACACTCCTACCT CCTCGTGGTACCGCCG GTAACACGCGCACGT CCACATCAGCGAGGG GCGTACTCCCCCGGAT GTGGCGGCGCGTGGC TAAACGGAGTGTGGC GACGAAGGAGCGAAA GACTAACAACTATAA CGGTCTTCCGTAACGG CTACTTGGAGCCGTGA ATAATGGAGCCTATAT TAAACCCTGGAACTC	TGAAAACACGATAACG ATTATGAATCAAAATA AGAAAGTAAACATCCC AGAAATTGTCTACGTCT TATTTGTTATCTATTTA TTTGTTTA (SEQ ID NO: 1332)	IAKFDNNTNSASDAAPLSPG GAVADLSASEGTTDNDQAM SPAMSLXTVPLVGNRVACP XCEKREANLFFLNLSDLDRH LTQHHPDAPIXWSCIDCAKC FPKLHGARCHIPKCGGASSQ ARTGEFQCEACPMSFGSRRG LSTHERHAHPAVRNIKRRGA DPPEENTKSWKVEXVARLK GLWEIFKDHKYPNKEISKFL TTKTVDXXKYQRKKLNLIG XESPQEATSLATEGGCDLVS SGNASFGSPVGRNENEEELI HEWKLSLKNEINKPTEVPPIL KEVYNRLMLIWEEHQDDRD SLTESLDHFIRTALYELINKI

				GTTCCTTCGTTCTGTT GACGACTGGAACGGC AAAGGACATGATTTG GATAACAATTGGAAA CTTAATATCGAATTCA CAATTAA (SEQ ID NO: 1209)		NKNQTDLKTKRAAKTKSPK NNRNSRKRFSYARCQELFHE CPRRLADAVVNNDQAYLEP ARQPPGSEEVRGLYEKLWG QVGSTYVPAPVTRVPKLSLS EIFPPIAAEDVGERIGKIRKK AAAGPDGLQRDHLTIPGLPII MAKIYNILVYCSYFPSAWKE NRTTLIPKINKPCSLVENWR PITISPILGRIFSSIIDGRIRRGT VLNMRQKGFTSENGCKINIE LLNSALNYSKRNSGGIFTIV DISKVFDTVPHAALKPCLAK KGVPALIVDLIDEMYKNVK TTIKTKDGGVEIMIRRGVKQ GDPLSPLLFNLCLEPLLEEIE EQASGINVSEHRKVSVLAFA DDIVLLGADAREAQHQINVL TDYLQSLMMNLSIEKCQTFE VVAKKDTWFIKEPGLKIGN QIMPTVDPDEAFKYLGAKIG PWKGVHCGVIVPELLSVVK RVRKLSLKPGQKVELLTKYI

						FPRYIYHLLVSPPSDTVLKLL DSEVRQEVKTILHLVPSTAT GFFYTPKACGGIGIPRFEHIIK LGTLKSAIKIANSIDPAVAGL IDDAAIKKLKQTANSLRINW PASLEDIEKARKRLRKEHISQ WADLKCQGQGVPDFIKNKT GNLWLEDHSLLKPSRLIDAL RLRTNTFGTRSVLARADKNI DVTCRRCRAQPETLGHILGL CQHTKGLRIKRHDEVKSLLE GRLKSKKNNEVFVEPTIKAG GSLFKPDLVIKNGERVLVVD VTVRYENKNYLALAEKEKI EKYRPCLRALKEIFNAKGGE ILPVVLGSRGTITPNTEKVLK RLGIANNDIKTILLNVLRSSI ELCNIFIDD (SEQ ID NO: 1454)
R2	R2-1_Crp	.	Crocodylus porosus	AGGCGTCTCCTTTAAG GGCAACGGTCTGGTT ACGCGGTCGCAGCAG KCTTGCKCCAGGTACC	TGAATCCCACTCTGGG GACCCCCAAAAATTAG AAAAACCCAAAACAGT TGTGTTTAAGTGTGTTC	VPPGAEARGRYHHPRXEXA RQGEPPSXRVFLVXLPDSNP PCPICGDHVXXXSVLALHCV EGHXWAXVQYQCTHCGILC

				TCCWCGTGGTTCCCG	TTGTTTGTCCCTTTGGC	HIPRCQGRVXEXTGKDXXC
				CCGGGTGCCGMAGMC	TTCACCTCCAAGTTGCG	PECPASFDEKAGLSQHKRHT
				CCAGGSCTGTCGGTAG	ATCCCCCATCTCCCCTG	VTXSXERVAGXLLRAXLRH
				CTCGATCCTGGCACAG	CGCTGTCTTTCTGAATG	GCWSVEEEETLTRLDAMFX
				TAMGGCCAGGGGAGT	ACCAGTGGTGTTGAGG	GARNINQLIAAEXVSKMXK
				KCTTCCTTGCTGCWG	CTGGTGTGACCTCGGTC	QISDKWRXLXLXPEQTTXG
				GTGCCCCACAAGCGT	ACCTCCAAGGCCAAGT	GXAESASVVXXESMTPEME
				KTGGCAGSMCMCCTG	GCCCTGGCCCCGAGTA	AQSPAXPPGKIRKIFTGQDG
				CTTCWTCGCAAMAAT	GGACCAGGTGGCCCAG	HAGGXAWENQEDFHWTRW
				MASAGKGTSMTCAGT	CTTGCTGGGCACCCGTC	ARRWLKRGQXLSDKVQEVL
				AGTCGGCCCCGCCGCT	ACCGCCCAGGGCAAAT	GXWVEGQPRIXAWVDXVSL
				AGCCAAAACTGTTCG	GGAAGGGATCTATCCT	DVLTLFLGVPPGPQRAPSKK
				CCACSCAGTTACAGAT	GACCACTACCAGGCTA	GPXEGGKPTSWMNKRAIKW
				GGTGCCTGTGCTGACC	AGTGTGGTGCTGCCTA	GTFLRYQHLFGANRKLLAA
				KGKKGCCCCGTGGGC	GCCTGCCGTAAGGTCA	XILDGAXRNQXTLLLEEVX
				TCGGWGGTGCAGGGG	AGCGCCCTGCTGCCGCT	QXYXGKWEAEPPFEGLGRF
				WGGCCGGCCCATGGC	CGGGTAKCAGTCCTCG	GXXRDVDSFAFEALITXEEA
				TGGGCCAGACSTGGG	TTCACTCGTCCCTCCTA	VKHMMXMAXXSAPGPDKL
				CCKTGGAGCCCGCTCC	GTACCCTCCGCMTCTG	TLRDLRRADPEGDALAELFS
				CAMCCCAGAGTTCCC	CGCTTCTGCTATCCACT	LWXITGVVPDRLKEXQXVLI
				CCTCAAMGCCGCCAG	GTGGCCGGWGATGCCG	PKAVDFEKLRQLGNWRPITI
				GKCAGMAMCAKCCG	AGGWTGSWMCAWCCT	XSIVLQLXSRVLTARLTAAC
				GGGAGKGMCACGGCC	CGACACCTCCAGGGCC	PIXPHQQGFISAPXCAENLK

				CCGGCCGTGA (SEQ ID NO: 1210)	AAGCGCCTTGGCCMCA	XPELIXRKVKXDRRPLGVAF
					GGTAGGACCTGGCACC	VDXARAFDSVSHDXISWVL
					TGCCCAGGGGGCCAGA	KAKGVDQHIVNLIEDSYQK
					CACTGCCTGCGGCAAG	VTMRVQVFSGSTPPISIKXG
					GGAAAGGGAGCCGTCC	VKQGDPMSPLLFNIAMDPLI
					CTGACCGTTACCGGGCT	XKLKTVRQGVKVGSASLTT
					TGATGGTGCTGGCTAG	LAFADGLXLLXDSWEGMQ
					CCCGCCGTATGTCAAG	HNITTSXTPGRACNTTSRHP
					CACTCCACAGCTGCTCG	RGLLQPHGPTSATXKMXGV
					AGTTCGCTGGCTTCACC	LLESXMRLLYGEQLRGLED
					TTCATCCCWCCTAGTG	XRPXXHDAXARRADTISGL
					CCTTCCGCCTCTGCGCT	EXRSLGWDXQTRFGYATXL
					ATTTTCGTCCCGACTCG	LAREDGDCXAQTNAEALC
					TACCTCCCCACCTCTGC	WXSXPFPGCAXRPXYANXG
					GCTACMGCTATCCCTG	WVASEALDSMSRRXVKEW
					MAAGGACCAAGTGGGC	FHLPACTDXLLXSRHRDGG
					AGGGGSGTTCGCCCCCC	LGLLRLARXXLAAXVRRPIR
					GCCCGCAGGAGGCTCG	VATSSDEVTRKVSYACGISD
					GCGTATCCGTGGCKTC	EVERLXLAXGGDXSNVPRF
					WTGCCTCCACCGTCTT	EDPXAPKSXXVQGPHEAAQ
					GTGCCGCTAGAGGGGT	ETPRVVRTQAIPWPSNWRA
					ACCTCMGAGACCGGCG	EEHSKWAQLSCQGERVELF
					CAACACGACCTTGACG	CNDPVSNGWINSRGQLAER

					STTAGACAGTAGGGTG MAACAKCCCTGCTGCA GGCSTGAAGGGCCAAA CGGCTGTGCCATGAGA GGGGAACCTTGAAGAC CGGGSCAGTCAGCCAG TTAGTCAGTTGGGCGA AACAATCCCAGCTGCA GGCCCAMMAGGGCTGT CAGGTGAGGGGGTATC CCCAWCCACCCCCCGC CGCCGACTACGGAGGC AKGAAGTCCCTAGTGA CTTCKGACCCCCACGTC TTGTGCCGGGAGAGGG GAACCTTGAAGATCGG GGCAAGCCGCACTTGA TAGTTAGCCAGTCKGG TGAAACAATCCCAGCT GCGGGTCCGAAAGGGC CGACTWCCAGGCGAGG GGGGCCTGCGGAAAMC CCCCTCCATGGTACGG	LWIMALKLRSNIYPTREFLG RGQAGTNIGCRHCTHPRETL GHILGICPAMQEARILRHNK LCKILAAEGKNCEWTVYYE PHLHNAAGELRKPDLIFVRD GTALVVDITVWYEGGPATL LSTTAEKATKYLDLNTQIQE LTGAEQVTYFGFPIGARGK WHADNWRVLSELGLSNSRK ERVTRLLSWRALLGSVDMV NIFVSKHRQENLLDEHCTPA EQVVSSYAS (SEQ ID NO: 1455)

					AGKTCTGGCGTCCTMA CCGACWCCTTGCCACC AACGTCTTGTGCCGGG AGAGGGGAACCTTGAA GATCGGGSCAAGCTGC ACTTGATGGTTAGTCAG TCGGGTGAAATAATCC CAGCWGCCCCCGCTGT GACTGCTAAGMCWGGT CCCCAAGGGGCATGAG GCATSTGCGCTGAGCCG GSAGGGGTGACACMCG GCGATCGGCGCAGCAC AKASTGAAGGGAGGCA CTTGCTGAGACTGCTTC TGAGGCCCCAGACTTG GGGTGGTGCAGCCTTG TCTGGGGTATGGTACA GCACCCTACTGCTCCCT TTGGKCAGCAGAATTC GTCCCGACCTCTTACCC ACCCGAGTCTGCGCCTT GTTCCGCTATCCTGCAT	

					CTCCGATCCACCTCGCT GTCTCCCCGCTGCGCTG CTTTTCTCTCAAGTGGG TTAAATCTTGTCATGAT TACCTCCCACGTTTCCG CTCAAGGGCAATGCCC AAMATGACGGGGATCG CTGGTGCATGGCAGTC ATGAGACCATCCGGAC CCTCCGGTGGTCGCTAT AGTCATTTTKTGTTGCA TGGGGCATSCTGAGTCA CTTAACCGAAAGACTC WAAATAACTCAAAAGA GGKAMCCTCTGSGGTT CGGTAAA (SEQ ID NO: 1333)	
R2	R2- 1_DWi	.	Drosophila willistoni	GAAGCTGGGTCGGAT GAGCGCAGAAGGGGT GTTCTTTGGAACACTG TAATTCATAAGTCGTA AGTCTGATCAAGTCG ACTCGAAACCTCCTCG	TAGATGTACTAACCTCT AGCTTTTCTTATACTTT TGCCTGCTACCTTGGCA TTACATCTAAAAAGGT ACAAACATCGCATTGT CATAAAGAGGTGGTTT	FERRSNSWGYRPLEPRSVGT ESNNNSPRSNITITSATSRPG DQPREAIAVVNLAGEIPCAV CGRLFNTRRGLGVHMSHQH KDELDTQRQREDVKLRWSE EEAWMMARKEVELEASGN

				TGGTGTTTCCTGGGTG CTGTTGAGTTCCTAGT CTCTAGGTTCTTTTCA GTAGCTAA (SEQ ID NO: 1211)	TAGTACGTAGGCGCTG TGGGACTTCATTGTCCC GGTGATGCAGTGAATC GTGCATACGAGATTGT CCAGTAGTTGGTTGCTC GTATCTTTAGAAGATTT CCTTCCTCGGCGATCAA AAAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 1334)	LRFPNKKLAEVFTHRSSEAI KCFRKRGEYKAKLEQIRGQS TPTPEALDSITSQPRPSLLER NHQVSSSEAQPINPSEEQSN WEIMRILQGYRPVECSPRWR AQVLQTIVDRAQAVGKETT LQCLSNYLLEVFPLPNEPHTI GRSNLRRPRTRRQLRQQEY AQVQRRWDKNTGRCIKSLL DGTDESVMPNQEIMEPYWK QVMTNPSTCSCDNTRFRME HSLETVWSAITPRDLRENKL KLSSAPGPDGITPRTARSVPL GIMLRIMNLILWCGKIPFSTR LARTIFIPKTVTANRPQDFRP ITVPSVLVRQLNAVLASRLA SKVNWDPRQRGFLPTDGCA DNATLVDLILREHHKRWKS CYLATVDVSKAFDSVSHQAI IKTLQAYGAPTNFVSFIEEQY KGGGTSLNGAGWSSEVFIPA RGVKQGDPLSPLLFNLIIDRL LRSYPREIGAKVGNTMTSAA

						AFADDLVLFAETPMGLQTL LDTTVGFLASVGLSLNADK CFTVSIKGQAKQKCTVVERR SFCVGERECPSLKRTEEWKY LGIRFTADGRARYSPADDLG PKLLRLTRAPLKPQQKLFAL RTVLIPQLYHQLTLGSVMIG VLRKCDRLVRQFVRRWLDL PLDVPVAYFHAPHTCGGLGI PSIRWIAPMLRLKRLSNIKW PHLEQSEVASSFIDDELQRA RDRLKAENVQLCSRPEIDSY FANRLYMSVDGCGLREAGH YGPQHGWVSQPTRLLTGKE YLHGVKLRINALPSKSRTTR GRHELERRCRAGCDAPETT NHILQKCYRTHGRRVARHN SVVNAVKRGLERKGCVVHV EPSLQCDSGLNKPDLVGIRQ NHIYVIDVQVVTDGHSLDQ AHQRKVERYDRADIRSQMR RFFGATGEIEFHSVTLNWRG IWSGQSVKRLIAKDLLIAED

						TKLISVRAVNGGVTSFKYFM YCAGYTRS (SEQ ID NO: 1456)
R2	R2-1_Gav	.	Gavialis gangeticus	AGGCATCTCCTTKAAG GGTAATGGTCTGGTTA CATGGTCATAGCAGG TTTGTGTCAGGTACCT CCCAGTGGTTCCCGCC GGGTGSCAMAGCCCC AGGGCTGTCGGTAGC TCGATCCTGGTACAGT ACGGCCAAGGGAGTT CTTCCTTGCTGTCGGG TGCCTCGCAAGCACKT GGCAGCCCCAATCGC TTCATTGCGAAAAAC ACAAACGTCCTAAGG GGATGATCAGCTAGT CAGTTCTGCCGCTAGC CAAAACTGTTTGCCAC CCAGTTACAGATAGC GTCTGTGCTGA (SEQ ID NO: 1212)	TGAGAAGTTGCGAGTT CTTATGCAAGTTGAATA CCACTCTKGKGACCCC AAAAAAWWAAACCCC AAAACAGTTGTGTTTA AGTGTGTTCTTGTTCGT CCCTTTGGCTTCACCTC SAAGTTGCGATCCCCCC ATCTCCCCTGCGCTGCC TTTCAGAACGGCCGGT GGTGTCGAGGCTGGCG CGACCTCGGTCACCTCC AAGGCCAAGTGCCCTG GCCCCGAGTAGGACTG AGTGGCCCAGCTCGCT GGGCACCCGTCACCAT CTGGGGCAAATGGAAG GGATCTGTCCTGACCAC TACCAGGCTAAGTGTG GTGCGGCCTAGCCTGC	PAAPRAWGAVEAGPWPGRT RAVEPAPSPESSPSEAARAA PAGEGHGPGHESPSVQRPEA DTTAPGVSAPTREGEPPSTR VFLVRLPDSNPPCPICRDHV GKPSALALHCVESHAWADV QYQCTHCKKVSANKHSILC HIPCCQGRVPEWTGKDWAC PECPASFNKKVGLSQHKRH VHPVTRNVERVAGSLSRAG LRPQTRRGCWSVEEEETLTC LDAMFRGARNINQLIAAEM VTKMPKQISDKRRQLGLCPE QTTLGGDAESTSVVEEESMT PEMETQSPINPPGKIRKILAQ RARRWLKKGQGLSDKVRE VLGAWVEGQPRIHAWVDSV SLDVLTLFLGVPSGPQRAPN KKRPKEGGKPTSWMNKCA VKWGTFLRYQHLFGANRKL

					CGTAAGGTCAAGCGCC	LVAIVLDGADRNQCTLLLEE
					CTGCTGCCACTCAGGTA	VFQAYREKWGLEEVLRAYR
					TCAGTCCTCGTTCACTT	GKWEVESSFEGLGRFGVRR
					GTCCCTCCTAGTACCCT	DADNFAFKALITPEEVVKH
					CTGCCTCTGCTCTTTTG	MMAMASKSAPGPDKLTLRD
					CTATCCACTATGGCCAG	LRRADPEGDALAELFSLWLI
					TGATGTTGAGGTTGGTG	TGTVPDGLKECRSVLIPKTV
					CATCCTTGGTCACCTCC	DREKLGQLGNWRPITIGSIV
					AGGGCCAAGCGCCTTG	LRLFSRVLTARLAAACPINP
					GCCACAGGTAGGACCT	RQRGFIAAPGCAENLKVLEL
					GGCACCTGCCCAGGGG	LLRKRKRDRQPLGVVFVDL
					GCCAGACACTGCCTGT	ARAFDSVSHDHISWVLKAK
					GGCAAGGGAAAGGGAG	GVDEHIVNLIEDSYQKVTTR
					CCGTCCCTGACCGTTAC	VQVFNGVTPPISIKTGVKQG
					CAGGCTTGAGATGGTG	DPMSPLLFNIAMDPLIAKLE
					CTGGCTAGCCCACCAT	TDGQGVKVGSASLTTLAFA
					ATGTCAAGCACTCCAC	DDLVLLSDSWEGMLKNISIL
					AGCTGCTTGAGTTTGTT	EDFCNLTGLRVQPKKCQGFF
					GGCTTCACCTTCATCCC	LNPTCDSFTVNNCEAWKIA
					ACCTAGTGTCTTCTGCC	GREITMLGPGESTRYLGLNV
					TCTGCACTATTTTCATC	GPWVGIDKPDLGTQLSSWL
					CCAACTCGTACCTCCCC	ERIGTAPLKPMQKLSLLVQY
					ATCTCTGCGCTCCTGCT	AIPRLNYQADYAGIGRVALE

					ATCCCTGCAAGGACCA AGTAGGCAGGGGGGTT CATCCCCCTACCTGCAG GAGACTCAGCATATCC ATGACTTCTTGCCTCCA CCGTCTTGTGGCGCTAG AGGGGTACCTCAGAGA CCGGCACAACATGACC TTGACGGTTAGACAGT AGGGTCAAACAACCCT GCTGCAGGCCCAAAGG GCCAACAGCTGTGCCA CGAGAGGGGAACCTTG AAGACTGGGGCAGTCT GACCATGCTGGTTAGTC AGTTGGGTGAAATAAT CCCAGCTGCAGGCCCA AAAGGGCTGACAGTCA GGTGAGGGGGTATCTC CATCTGCTCCCCACTGC CAACTACGGAGGCATG AAGTCCGTAGTGACTTC TGACCCCCACGTCTTGT	ALDSMNRRKVKEWFHLPAC TSDGLLHSRHRDGGLGLPRL AKAIPEAQVRRLIRVATSSD EVTRKVSYACGISDEVERL WLARGGDMSSVPRFEDPEA PRSPGVQGPCEAAQEIPSVV RKLAIPRPSNWRSKKHSKW AQLSCQGEGMELFCNDPVS NGWNNSRGQLAEHLQIVAL KLRSNIYPTREFLGRSQAST NVGCRHCTHPHETLGHILGI CPAVQEARIIRHNKLCKILA AEGKKCEWTVYYELQLLNA AGELCKPDLIFVRDGTXLVV NVTVGYEGGPAXLLSTAAE KATKYLDXNAQIQELTGAE QVTYFGFPIGARGKWHADN XRVLSELGLSNSRKERVARL LXWRALLGSVDMVNIFASK HRQENLSDXALSPS (SEQ ID NO: 1457)

					GCCATGAGAAGGGAAC CTTGAAGATTGGGACA AACCGCACTTGAAAGT TACTCAGCCGGGTGAA AATAAGTCCCAGTTGC GGGCCCCTCGGGGCTG ACAGTCAGGTGAGGAG GGCTGCAAAGCCCATC TCCTGACTCCAGAGGC CTGGCGTCCTAACCGA CTTCTTGCCACCAATGT CTTGCGCCAGGAGAGG GCAACCTTGAAGATCG GGGCAAGCCGCACTTG ATAGTTAGCCAGTCGA GTGAAACAATCTCAGC TGCGGGTCCGAAAGGA CTGACTTCCAGGCGAG GGGGGGGCCTGCGGAA AACCCCCTCCATGGTAC GGAGGTCTGGCATCCT AACCGACACCTTGCCA CCAATGTCTTGTGCCAG	

					GAGAGGGGAACCTTGA AGACTGGGGCAAGCCG CAGTTGATGGTTAGTCA GTCGGGTGAAATAATC CCGGCTGCACCCTGCTG TGACTGCTAAGCCCGG TCCCCAAGGGGCATGA GGCATGTGCGCTGAGA CGGGAGGGGTGACATC TGGCGATCAGCACAGC ACAGACTGAAGGGAGG CACTTGCCGAGAATGC TTCTGAGGCCCCAGACT TGGGGTGGTGCAGCTTT GTCTCGTGTATAGTACA GCACCCTACTGCTCCCT TTGGGCAGCAGAATTT GTCCTGACCTCTTACCC ACCCGAGTCTGCGCTTT TGTTCCACCTCGCTGTC TCCCTGCTGTGCTGTTT TTCTCTCAAGTGGGTTA AATCTCAACATGATTAT	

					CTCCCACGTTTCCGCTC AAGGGCAATGCCCAAC ATGACGGAGATCGTTG GTGCATGGTAGTCACG AGACCATCCGGACCCT CCAGTGGTCGCTATAGT CATTTTGTGTTGCATGG GGCATGCTGAGTCACTT AACCGAAAGACTGTAA ATAACTCAAAAGAGGT ACCCTCCGGGGTTCGGT AAA (SEQ ID NO: 1335)	
R2	R2-1_IS	.	Ixodes scapularis	GTTCCAAAGGAAGGC ACTCCTTTGGTTCGTG ATGAGATGTTCATGGT GCTTGCCTAGCTGGAG AAATCCGACTCACAC CTGCACGTGGTCCCTG CCGCCTGCCAGTATGC CGAGGAAACGGGTGC AACTTAATCCGTGGAT ACTGGTAGCAACGTG AGCAACGGTACGGTC	TAGTGTGACGGAGTCC TCAAGCCCCCACAAGT GCCTGCCAGGTGGCAG GAAAGGGCAACTACTG GTGAGCGACCCAAGCA AGGCGGAGCCAAGACC AAGCTGGAGCCAAGAG CAACTCCAGGAGGCAG GGGTGGATATCAAGAG CAACCCCAAGGGACAC AGACCACGGGCAACTA	MQCTSRLADAPRFARVGVE GEGVGASGNGTDAQLWYG CTGCDEAFSSLRGLRIHAAQ KKHGNQDGLLRLPAGRPRK RRVGKSTTAGASDRVTTDP VPAPVPESPGLLPGLPGPSLP GCSDLPPGVLPGGWSASPGP LSWPPSLDAGPLPGPSRVSP GPSRPSPGKPTGPPSLDAGPL PGPSRVSPGPSRPSPGKPPGT PEPLPGSPGGRRGVSPGQPG

				CTTCGCGGACCACCCT GGGCGTTCGGGTTGCC AGCCCGTTCGCCCGA AATATCTTGGCCCTGA AACTAAAAGAAAA (SEQ ID NO: 1213)	CTGGTGAGCGCCCAAG ACAGGGGTGGATATTA AGAACAGCCCCACAAA GTGTTACCTATATTAAC AATAAAGTTGAAGCCT CAACCACGCATTGCGG GTTAGATGGCGTGGCTT GGCCCGCCGCCATGAT GAGCTGGAACCCTCCA CCTGGTGGGCCGCACG AGACCACCGGCTCTTTC TACTAAGGCCGGTCTCC GTGACTGCGGTTGGGA TAAACTCCAAGCACTG AGCGGTAAAAAAAAAAA AAAAAAAAAAAAAAAA AA (SEQ ID NO: 1336)	SRTDPSSSAGAGHFVCPQCS RAFSSKIGMSQHQKHAHLEE YNAGINITRTKARWDPEETY LLARLEATLNPDHKNINQTL HAALPRGSCRTLESIKAHRK QAAYRDLVTSLRSARESSEA QHVPDRPLETPEPQTPANPQ RDSKQAVIEALQSLIGRAPP GSFQGARLWDIARQATRGT NILPLLNSYLRDVFTLPTKPT RKKPAVRPARSRRKQKKQE YARTQDLFRKKQSDCARAV LDGPTSSSVPGTGAFLQTWR EIMTGPSPALEAPPLPTRGEV DLFFPATAQEIQSAEIAVNSA AGPDGFSARLLKSVPALLLR VMVNLLLLVRRVPAALRDA RTTFIPKVPDAVDPSQFRPIT VASVLQRLLHRILAKRALEA IPLNFRQRAFQPVDGCAENI WLLSTALNEARTRRRPLHM ASVDLTKAFDRVTTDAILRG ARRAGLSGEFIGYLKELYTT

						SRTLLQFQGESLLVEPTTGV RQGDPLSPILFNLVLDEYLSS LDPDISFVSGDLRLDAMAFA DDLIVFASTPAGLQDRLDAL VEFFDPRGLRVNVKKSFTLS LQPGRDKKVKVVCDQIFTIG GTPLPASKVATPWRYLGMT FTPQGSINKGTSEQLDLLLTR TSKAPLKPQQRLVVLRNYLL PRLYHRLVLGPWSAALLLK MDTTIRGAIRRWMDLPHDT PLGFFHAPVTEGGLGINSLR ASIPAMVLQRLDGLHFSTHP GAEVAIQLPFLTGLHRRAEA AAQYQGQRLLSKADVHRM WSARLHGSCDGRPLRESKR VPAAHRWAAEGTRLLSGRD FISITKLKINALPTLERTSRGQ HKDIQCRAGCQAVESLGHV LQACHRGHRGRIRRHDNIAR YVCGRLTQIGWAVKWEPHY SVAGRTLKPDIVAHRGAETV VLDAQVVGTSMRLGFHHAQ

						KKEKYSLPDLLHQVCEGRR DAARVSTITLNFRGVWAPES AQDLKSLGLTDNDLKLLTV RCLQGGAQCFRLHRRMTTV VKATGDEANALPAHSGLPP TQLGGRTLGPSAHNQSARTT (SEQ ID NO: 1458)
R2	R2-1_MLe	.	Mnemiopsis leidyi	TGGGGGCCCCTTGGA CTTGCTCCCTGGGGCA GGACACCAGTGAAAG GAGATCCTCAAGACA GGACAGAGAGACAGG CACACCAACCCTTCGA ACCTGAGGAGACACC CAGATCTGGTTACCCC ATTCCTGTAACCATGG TTGCTCTCCGTGCGCT GTTAAGGAAACCCAG CCTAGTACCCTCGGGG AAAGGTTGCGTAGTA CTTAGCAGTGTGCGG ATCAAACCTCTACCGG TCTCTCTAGCGATGAA	TAGTGACGAGAAAAGT CGCTTTATCCTTACATA ACAGTGTGATAGTCAT CCTTATCACAACTGTCC TGGCGAACAAAATGGC AGAGGGATAGTACTCG GTCCAACCAGAAGGAA GCCCACACGATGCCAA ACTTGCTGTAACCCACG TGAGCGGAAAACATCC TCCGAGTATAGTATGAT GGAAGGATACAGGATG TGGACCGCTCTAGGCG GCGGACGGTGTAGACG GCGACCTTACCATAGCT GCGACTGCTGTGCGAT	MSNTSHSKLNLKMDNKLKT SLETPSGVRADSIITRVRTSS NRGEHSNGVTYPRCEQGVA PLDTHGGICDAPPQVTVPAT ETDKQKKCEYCEFTYLKPR QIGTHMRKRHPQEWNDIKR TKFLSEKRQKRWLDEDFELL CIGQEEYLVLSSIGKQGKGIN QYIQTKYFPTLSTDAIKSQR KSRRFSEYSEKRSRELQPCN TSSDPEELPNEAVTENSPLSF DPLDRDVVKKISSKDHGDQI LLVQEHLINGRYQEANTLA KAIFEKLSGKFPNLKTGDHR PGKQQTARKVGKKRVRGSG KKLSPSKQNRRELYAIVQKQ

				AGTTTCTCCGACTGGA	CCGGAACGGGCCTTTCT	WRTKKRSKVINQILTGNLNK
				ACTTGAGGGACTGGC	CACTTACCTCAAGCACC	EQSYTHTPDQLAQFWSTLFG
				TAGCCCAGCTAGCCTG	GTGTCGGATCGCGGGG	RVSPRDDRPINHRRSVIPELD
				TAGAGCAATGCGTAT	TGGGCGCGGGATTCCT	KPLSVEEVEAALKGAKDAA
				ACGATGCCTTGGCGA	GCTGGGAAGTTGTCTG	TGIDGVPISHLKHLGSAALTI
				CAATGGCGACCGCTG	CACCCAAGTCGTCAGC	LYNGLYVTGSIPDPWKRAR
				CTTAGCAGACGGAGG	TAACTGTCGGTACTTAT	TILIPKSNPPASPGDYRPISISS
				TTAGTGAAAGGGCGA	ACGCACCCAGAGCCAT	YFYRIYTSAISKRLASAVSLD
				CTTGCTGTTCATAGTC	GCTCTTACCAGGCCACT	DRQKGFIKEDGIRDNLSLIDT
				ACGTGAGTCGTCTAG	TGACGGCGCTACACCG	LINETKAGSKSLFMTFMDVK
				AAACTGCACCGATGC	GTCTGTAGGAGGGTAT	KAFDSVSHYAIARSLEWAG
				TGCCCTCTGTTCCAGG	CTCAGCGACTTGGACA	VPDGMRSVIADLYQDCTTDI
				AGAAGGACAGTGGAG	AAATGGATCTTGACTG	CGRSVKVTRGVKQGDPLSS
				GACTAAATCGTAGCG	CCTGAGGGCGCTTATC	TLFNLVIEMVMSNVPERLGI
				CGAGCGGTGTT (SEQ	GAGTGACCCAGAGTAA	QFQGHRLFYLAFADDLVLL
				ID NO: 1214)	GCTGGTAGGAAGAATC	TRGPTANQKLVSLVHEQLA
					TTGCAGTTGAGAGGGC	RVGLELHPGKCKSIAIMADP
					TAGTAGGGCCAGCCTC	KRKTTFVDQGSSVLIGGEPV
					TGCTCGTCAACTGTCCT	SSLGPQEWYKYLGIKLGSGG
					AGGCGTTAACGTGTGC	MPQGIYRDQLADLLAKTDS
					CTCCCTATGAGGTAACC	APLKPQQRLYILRSHILPKFN
					GTGAAGTATTACTTGTT	HRLMFERVTCQTLEGLDKLI
					TCCCTGTAGAGACCAT	RTHVRKWLKLPKDTPGPAF

					ATAACTAGCTGAGAAA CGCCCTGCGGGTTAAT ATACAGTTACCGCCGTC CTGCATCCCCGAGTTGT GACCAGCGTACAGCCA TGTACCACGGCACTTCC AGCATTCTCTGGTCTAA GAATGTATCAGCTGGC CCGCCGAAAGATAGAG CGACCCCGCCTCTATCA CTAGAGTAAGCAGGCA CGCGAAATATATGCAG GACGCCTTGGCTCAGT AGCTCTGGTACTGAGT ATAGCATGGCTGAACG CCCCTTTAAGTTCGGTG GCCCAGAGACTCTCCG AGACTTCCCAGTCCTGG AGAGGGTACACCTTTA CCAGAACTTTCTGTGGT CGGTTG (SEQ ID NO: 1337)	YADKGSGGLGLITLRYRVPL LKLRRHKKMADSPDPVIRLI PNAEPTISLLARWTKMCSLY GKQYQHKSELSKIIRDKYWT MCDGKGLRTEVPPDTAKKT LSLLFEDRTPLKPGQLIGAIG VRLNTLGTPARNNRAKGYS PEANICDKCPGNRQATLGHI SQTCPATHGRRVKRHDKIV NRIAKALKERGSVKNILTEP HLRHDKLPLRKPDLIVHTEK SVEIIDVQVVADQGISRHED EDQQKKIVKYDVDGYKRAA YKMLGIDYGSIPCNVSAFTIT WRGNLAPHSLKLASRLQFSP VLKYIVADSLVDTWGAFLI WGKTS (SEQ ID NO: 1459)

R2	R2-1_PM	.	Petromyzon marinus	CTATTAATGGGATGA AGAAGGGGGACACGA GTTTGTGTGTGCATCC AGTTTCCATGGTGCAT GCAGGAGTGGTGGTT TAAATGGCGAGACTC TACAGGGCTTCCATGG CTACACGGGATGCAA GGCATCAGACATTTTG GCACAGGCAATCCTTT TGGTCTCTACCGCAAT CATGTCTTAGACCTCA GTAGCGACCACTACA ACCACAGTGGTGACT GCTGTTGAGTGAAGG ACGACTGAGCGCTGG ATAACAACTTTCTTGC GTGGCCCAACATCGA AGCAACCACTTCGGA GCTGGCACAAGGCAA GAGGGCAGCCCAAGG TGTGAATCATCTCAAC TTCACTGCAGGAAGA	TAATTTAAGGTAAAAT CGTGGGATTGTTTTGAT GGCAATCTGCCTAGTC GCGGCCTTCCATTTTGG GTAGGCAGCAGACCCA TCTATATAACAAACTAC TTTGCCTTTCATAGGGG TACCCGACCCTACCAA CTTTCGGGGAAGTAAA AGAAA (SEQ ID NO: 1338)	MNERLTDELTTEFILSDMFL WDYPCTDQNKCYPCNLVFL DHRTWSSHMARVHPHANK TYKCRICNRTADSIHKIASH YGRTCKSLIGKTNAITTTIDE TLFSCLHCSRGFTTKTGLGV HTRRTHPTEHEAILQQNTPG RKVRWGEEEVEIMAHKEAQ QKDEDINMNQLIQNSVMPH RTLEAIKGKRRNIKYKELVR TLKETTYKVENQCLVNLVL PTTSEITTTPSEGDQPAIRAE KEQSPTAAEDLQVIINDLKS QNFSHNQALLLLNSHVEKFL NRSKPIKRKDHVNQQEIDEN RHRRQSKQTKYRRYQYLYH TNKKALLDEITSDRSGPSIYP TEESIRGTFVTLFESNSPPDNI PSKLKNDQSCIDIVKAITLDE LIKTLAIMKDKSPGQDNITLS DLRTLPIKYLLDILNIILYIQD IPQIWKQHRTRLIPKTKEELE KPSNWRPITISSIVIRLLHKIL

				AATGCTGTGCAAGGA		SYRLGQQLKLNYRQKAFLP
				TGAGTGTGAACGACA		VDGCFENSALLHFIIHNARQ
				CCAACGGGATTGTTGC		KHENTQIVSIDLSKAFDSVS
				TGACCAGGAGGTGCC		HESIIRALNRFNLSKESITYLT
				AACCAAATTTGAATG		NIYKCNLTDIVFGSTIMRNIN
				GATTGACTTTGGGCCT		LKRGVKQGDPLSPLLFNMI
				GGTTTCTCCTGCGTGT		MDELLDNLPTYIGVNVGNQ
				ATTGCACGGAAAAAC		KVNSMMFADDLILFAETEC
				AAGTGGCTACACGTG		GMNKLLDITTKFLDDRHLKI
				TGGCCGTCGTGTCCTG		NINKCNSLRFIKYGKQKTFS
				GGGTTTCGCAACACA		VATTSSYFINNEPINPVSYVK
				ACTCCACAAGATCGA		GFKYLGIEFDPRGKRSISCNL
				CAACTATGAGGATGA		LAAMLNKLTRAPLKPEKKV
				CAATGTACTTAAAGA		YLINNNLIPRIIHQLVLGKVT
				ACAAAGAGACTGACG		KGLLMSLDSEIRKTVKLLLR
				CCAAAGGGGATTTAA		LPHDTPDSFFYTSVSNGGMG
				ACCGCCAAATCGTAC		IRNLCDSVALSIINRHNKLIT
				ATTGGGTCTCACTACA		SDDLVIRALSQQSYTIATLK
				ATTTTTTTACGTGTAT		QAHIIAGSKFPSKSLNQNKW
				TTATTTTCCTAAGTGT		SNKLYQTTDGRGLVYCQSQ
				CTGTACTTGCCATTCT		TENNSWITGNHRTIKSYNYI
				TCGCTGCTTTTTCTGC		DMVKLRINALPTKSRCNRG
				ATTAATTGCATATCGT		TLETKQCRFKCRSINNQISEE

				ATGCAAATAAGCGAA TTAACCACCACCGTGC AACTATATGCAGATGT TACAGCTGAGCCCTCT ATCATACCGGTGTACT AATCTGGTATGGTGTT GGCATGCTATGCTTGC GTAACGACCTTTGCTG ATTGGTTCAGTCGGCT GATGGTGGGTTCAGG CGAAACATTTGTATAT TGGTTTAATCAAACCG AAACACTAAAATTTTG AACACAGTTTTCCATT ACACCAGTTGTATTGC TAGAAGTGCAAATCG AAGGAGTCAATTTTG ACCGACGATTAGCTG CCGATGTGCGGTGAA AAAGCTGATCACAAT AGCATACACTTGGGC CGACAACCCCGTGTG CTATAAACGTAAGTC		TLAHILQKCDRSHYSRIARH DSLVQFLATAAQKLNWEVI KEPTLPSDTNKAKPDLILVR DSHVLIVDVAVPWESRSLA HAYDFKVKKYATDKKMQA YLKTIYPEKEIRTEALIISARG GWCALNNMVTKKVGLSSA WVKLALIKVMEGSVKIWRS WSKG (SEQ ID NO: 1460)

				GCGAATTATAAAGAA AACAAACCGGACGGA CTACTCGGTGACGAA CTAACATCGCTC (SEQ ID NO: 1215)		
R2	R2-1_SM	.	Schmidtea mediterranea	CAGTGCTATTCGAATG TCAATGTGAAGAAAT TCAACTAAGCTCTGGT TAACGGCGGGAGTAA CTATGACTCTCTTAAG GAATTAAGAATTTACC TGCCKTAKTAAAART GAAATCMGTTGTTCA TWGCAAGTGGTATTG TACACCTTCCCGCGGT GCTAGTCGTTTAAAAC TAAGTTACAAACCAC GAGGGGCGTCCTGAC GGACTGSAAAAGCAT TGAGRGTCMTGAAGA GAGGCTCTTATTGTAC GAATCTCTTCAACGAT CGAAGTCTGGACCGA	GGCCACGCGCGTCGTC CTTGTTTGATCACTAGT GGATCAACCTTCGACTC CCCGGAACTGTGGGAG TGGCGGAAGAAAGGCC AGAGGATGTCCTGAAA CCATATATTTATTTATA GAAGTTTTACTTCATCC TATTTACGTATTTCAGT ATGAAAATGAGTAAAG TTCTCGACTCGATGAGT TGGGGGCAACCATTGG GGGTCCTGAAGAGAGG CTCTCACTGTAAAAAAT CTCTTCGTGTCTGTTTA TTCCTAGGCACNTGCTG CATTATGAAGCGGWGA AAGTAAAGTTAGAGCT	MKKVLNNETEKLPGSNLTF MCGFCDREFDTARGRGVHE SRGHLVERDAAVQSRVKAV VSKKYYYSNEEDVALAKM QLXHADLAKSEXLEAMYLA LGKGRTREAIEQHIRKSLRY KGVLEEQRKLLETARGNVR QNNVGVPASNATKNLQRFL ESLPLGTNRREERLDRIIRSN SIESQRLELIHYCNDMCQDF VQLDCQXNPINAIRRRNPKR LSKKQLKRAKFSALQRLWIR DRKAAAQLVLKDKLDSLLS NKEDSKDLGSYWQQVFERE SELDRRPIPQVVENEELNSPV LEKEVEWAVKNIKKSTAAG PDGLTALALKKIPYSELVKL FNIILLVGFLPDVLKNSRTILI

				TATGAGAACTAATAC	GAGAGATAGGTACTTG	PEVDNPQGGGDYRPISINSV
				ATTAGTTGACAGGTG	CTGCATTATGAAGCGG	LTRTLNKILAKRVSEGDFGI
				AAAAATACTGTTGATT	AGAAAGGCCTCGAATA	NGQKGFKSVDGCLENLATV
				ACTTAGTTCTCAGTCA	AATAGGGTGTTAGAGT	ESILADARMKKNKLAVVFL
				TGTGGTATATTGCCAG	TATTGATGGAGAGTAT	DMSKAFDSVNHESIVRAGEI
				TCAATTACTACATWA	ACTAGTAAGCTTAAGC	KGYPKLLMTYVKECYNDAT
				ATATTAGTGTGGCTCT	TGCGCCTCGCGCGGTG	TNVAGVTAKFNRGVKQGDP
				CAAAGGAACACGATT	CCCAAAAATATACTTA	LSPALFNNVIDLAIERVSGTG
				GRTCGGCAGTCCAAT	ATGAGAGCAATAACTC	IGYNMGGKKYSVVAYADD
				GCGCGACTGGCGGGC	AAGGNGAGTTTAATTC	LVLFGESREGLQIALTALLEE
				TTGTTGTTTGCATTTG	ATATGCGCATGCGGCA	LKLNGLTPNPAKSASLTFER
				TTACCGGCTACTTGAA	CCAAGGTGCTGAATGG	SGPHWFASTDTVTALGDQIP
				AAGGTTATATATAGC	CATCGATTAAACCTCTC	AMGNIETYKYLGIKFNSCGV
				AGACGCTTAAAGCGC	CTGTTGTAGAAGCAGG	VKGSLPGIYTKKLELISKAPL
				GACTGTAATTTACATY	TCATAAATGGAGGRGG	KPQQRLAMLTDFLIPGVLHQ
				TCATTGCCCAGTATTT	GCAACCACTGAAACTT	AVFGQTNAGDLRSLDKRTR
				GTCTTTTGTCAGATTT	ATGAGCCAGAAGAAGC	RAVRSWCHLPSDTSTAFIHA
				AGCAAAATTTCATATT	TTAACTACAWAAGTTT	KAKDGGMGIPSIRAEVQFGK
				TTGTTAATTACCTTAA	TAGGCAATTACTGAAC	LDRFGKLPNVKDERSKVLA
				CTGGTTAAACGATCCC	GGAGTTAACTGTTAGTT	DNAHIKKKMLEKLGVGIPIK
				ATAATTGCTTGCAATT	AACACTACCATGTAGTT	GVRCKNKLEFYNKMREELI
				ATTATAAAGTAATTCA	GTTTATAAAGCAAATA	KSNDGIGLKEASLVPSANTW
				GGTAAAAATTACATA	TCAGGTTTCAGTCTATA	LKLSDLHMSGRTFVGCLKT

				TCTGGCTGATCCTGCC	TACTAAAAGTATTTTTT	RGNLMATVTRTSRGGQNPG
				AGTAGTCATTTTACTT	GATACCGTGGTATATA	IELNCKKGCQYQGSLNHIVQ
				CCGCCGCGCTATAAA	GGCAACTAGTTAGGAA	KCPVVKGLRIKRHDEVVKY
				ACAGTTTAAAAACTG	ATAGTAAGGGATGACG	VEEITKKAGWSATMEPIIPFE
				AATAGGAATCAAAAA	CATTGTCTCTCTTTATG	GSHRKPDLVLVRGDLGKVV
				GAACATGGCAAGCGA	GTACTGAGGAAACTTA	DIQIVSDHCGLDEKNSCKIG
				CTATATGTAACTGGGC	TCGACTCGCGAGGGAT	KYDNDIIRNYVRGLGPSRVE
				ATTCAACATTCCCTAT	GAAACCCGTAAACCGA	VAAITLNWRGVWSRDSFNLI
				TACATATGGTGGTGCC	TCGATYTAGCCTATAA	KRLGMTEMDAKIISMRVLA
				TGGGGTCTGTTTTATA	GTACCAGCGACAGTTA	STAKMFKTCKKVLEPVCRT
				TAATGGGTACCCGGG	AACCATCTTACGCGAG	KTADCDGYGPEETSARPCH
				AAGTGGATCTGTATCA	GGGTAAAACCTGAGGA	ELNLKESSGT* (SEQ ID NO:
				CCAGTCATGGTGCCAT	CCGATTATGGTATAACT	1461)
				ATCTTTKGATAAAGAT	TCTCAAGATTAGCACA	
				ACAGTTTAAAACTGC	AAATGCGAGTGCAACT	
				GATGATACTAATAGA	TGAGGAGGAGGATTTG	
				GATCCTCTTAGACCTT	AGTGTTAATTCATAATG	
				CGTAAAGAAGTGGGG	TACTAATCTAATTAAAC	
				ATTGATGACATTAGCA	TGTGACGGGAATTGCA	
				TTGGAAGAATTAAAT	GCTTCGGCTGTAATTAC	
				CTCCAAGGAAATGGA	TTTGAGGCCTATCACGG	
				GTAACTTCAATGAAGT	ATTGTAAGGAACATAT	
				CCCACAACCCCGTTGA	TGACACCGTAAGTCTA	

				AGGGCTGGGTTCGAG TATCGAGAGAAAACT CTAAATTCTCTTCGGT TMTGTCCAACGGAGG GGACATTACTGTAAA ATATCCTCTAAAAACA ACT (SEQ ID NO: 1216)	ACGTGTTCCCGATTTCC AACCAGGTCATATGAA GGGCTGCCCTTGATAA GGCGGATTTGACCCAA TTCTTCATATGAGAGGC TTATTCCAGCCTTCCCG TAGTACCGTGAGGTTTT CCCGCCTCGAACGGAA CAATGTTGCAGGGTAA TTAAGTACATCGGGCT ATATMGCGATATTTAA CGTTTTA (SEQ ID NO: 1339)	
R2	R2-1_SP	.	Strongylocentrotus purpuratus	TCTCGCGACGCGTTCT TCTGCCTGATGAAGTC ACGTCAGGTAATAGA CTTAGAAGGTTGATG AGCGTTCCTCTCCTGG ACCGGGGGTGAGGAT GTGTTCTACTGAATCA TCGTTCCGGTTGTGAG GTCCGCTGCAAATAG GCCTTGGGGTGGTCTA	TAAACCTTGCCTCCCCG GGCCCCCCTCAGTGACT AAGACAACTTTCACCG TAATAATCATATATTTG TATACCATGTATTAATC TAGGTAACAATTGAAA GTAACATTGAACCGTA TTCATACTTTAATGAGT ACAATAGTGAAGGGAA CTGTATATTACATACCT	MENSFAWEGTSSAEGRTTV EDSPSSSDDFVSNVGFKVAK ADPTVWEEANMSEDNTIIED PPSSSDDFANNVGFKVTRAD PTAWEEASTSTETEDLPSSS NFIDNVETQIDMAGPTAWE DADSNEDNIDEGTPNNINNN LAIVRGRADAYACSCCERNF ISLKAIGTHLKETHNKKVVF ECAKCQHTFVKAHGLACHV

				CCTCCGGGTCGCTACT	CGAGATAGAGGTTTTT	PKCKEDADTPMLNRLLHGC
				CCTTTTGAGCTAGTCG	GTACCTTAAGGGTTCGT	GECGLAFNTRRGLSQHERH
				ATCCAGGTGAGAGTC	GAGAATCCATACTGCA	RHPSACLSTRRRSRLDGIAR
				GGGGAAGCCCACTTA	TAAAGGGGTTAACTTG	KKSLRNRRDIWTNDEIRLLK
				GGTGGGCCAGCTAAG	TAACTCAACCCCAGGG	QLMIQYEHAKKINIKIAEHF
				CAGATCACCCCCCCA	AGACAAGCCGACAAAT	NHKNAKQVMHKRRSLREK
				GCACGACAGTGCGCT	TCGGCATTACTATGTGA	DMALGAPHDAPPPLAEEPIIE
				GTAATATAGCCTGTTG	GGGTCATAAGTGTTAA	VVEGAREELEQAPVDVILPD
				AGAGTGCACCCATTTA	AGGACCCATTGTATAT	LEALTVNDGRGGGSPVLTE
				TACTGTTATAAAAGTA	AACTTGTAAATGTACA	GGESTRDMEENGGTDTRSPS
				TAAATGGTTCTTAGCT	GAGAATTCGAGTTGAG	PREERAGSTPWERGWQPRV
				AACCTATCCAATGATT	TTCGAAGAATAAACAA	DRGRGEYKGYGGERGDRHT
				TTTATTGTTATGTATA	ACAAAAAAGAAAACAA	VSLSQRGESGVDPLTPGGVV
				ATACTGACCACTTTAG	AGAGAATTCGTTCGTCC	EDYDDSYLEDYFPGWDEDE
				TCTAGTGTCATGTAAG	ATAGAAGCGGAAAAAG	HMHIIGRLDLSDESEQGEAA
				GATCCGAGACAATAG	CGCAAAAAGGCGTCCT	VSPRLGSFGDLLEEVNAME
				CCTATGTCTCTAGTCT	AACCTCACGGTCTAAC	GKDNLSEALAETLGLVLHE
				AGCAATAACACAAGA	CCAGAAACATTTAAGG	GHRVEYIKEKMNINVKQMA
				GGGATAATCCACCTCC	GGGGAAGGAGCTCATT	TEILAHGANKGNPKRRKEA
				CTACTCCAGCCACCAC	ACTCCAATCCAAAACC	VAAKRNPGTRLDRAQRDNQ
				CTTTACCTTTCTTATCT	ACGTTCCCTAGCCCAA	AKAKAKEKKRIFSETQTQY
				TCCATATAGATGCCAA	AAGCGCGCTTCAATGA	KKNPHRLVEKLLDGKGDER
				CCAGACGGGTCGATG	GGGAGCCGATATTGGA	CSVSLEVIQRTYMNRFSRES

				GGCAGTGAGAAACAA GAATGGGGATAATTT GGAATTTGACATTCTT TCCTAAAGCAGACCT GAAGGGTTAAGAGTC ACAAGCGGAGTCCGG ACTGTCTTTATAAACA GACAATATTTTCTTAT TTTACCACCTATATGG GGATATTTCCCAACTC GTAATATAGGGCCCA CTGTCAAGTGGCTGGT ATAAGTTACCCTGTGG GTAACAAATAAATTC AAACAAAGA (SEQ ID NO: 1217)	GGACAGAGTTCATCAA AAGCCACCTGTAAGTA GGCCCATTTCTGCCAAG GACATGCGCAAAAGGA AGCAGATTATCAAACC AGTCAAACAAGCACAA ACATTGGGGGGATGGG ATAACCCCGGAAAGAG AGGGATCTTTAGATAG TGGATGGAAGGGGTGG GACGTTGAATCCAACC ATGCCGTTTTTATGTTC CCGATAAAGAAGGATA AGGTCACTCCAGCCTG ACACACAAAGTGGGGT AAAAGAACTCCGCTCG TACGGACTCCAAATAG AA (SEQ ID NO: 1340)	KEVDIGAYVDPETVEDNQGI VDPISKAEIERAISTTKKGSA PGPDGVTYDALKAYGNCQL YLLIMYNTWLAMGKVPSEA KTYRSILIPKGQGDPMDINN FRPLTLANVISRLYSKILTRR LDGAVSVCPRQRGFTHKASI EDNTLILRELIMKSKRNKEC LAVVLLDLAKAFDTVSHDLI IKALRRHRVHEHLISVIMDL YEGGTTSFTTDEGTTCPIAIR SEVKQGDSLSPVLFNLALDP LLATLEQRGKGVEIGGHTFV SLAYADDTALVSSSHLDMT ANLDITVEYLNATGLSLNVR KCQGFLLTPINKSFLVNEAE SWVVEREAIPWVEPGDTAK YLGVQVGPWSRPWPSIQPVI KRLTAYCESIDKAALKPRQR IHILTTYIAPRIAFEIAEGGYS TLVDCRGGIQYTRIREVDMT IRNYVRKWLFLPACLSNSFL YTRRGEGGLGLVSFYDYVP

						TERMRKLVRVCDSEDPVIA GAAASLGLRERAAKISAQTG LPVPVKPKGAHNAWRKVQ KKKWKAQPTQGKGVSCYQ HRLGNKWLGAPSFLTENDYI WAIKLRTNLVPTREAMGRGI IGRNQVECRHCHTTIETMGH ISGYCQMVLDIRLIRHNRICK ALIKAATATGLRVTEEPRIV GTDGKNYLPDLIFSAGAGEP CYVVDPTVVWDDDPKNLRE AWRGKVRKYTPIIPAVEAM LHPSSVQIFGFVCGARGTWC PMNDDIAKIVGLKNSGISRT LQIVLCDTIRMVKAFMAR (SEQ ID NO: 1462)
R2	R2-1_SSa	AGKD01072455	Salmo salar	AATCTTTAACCCCGGA CTCTTGGGGTTCTTAC GACTCTGTATGAGGA ACAGTCGAAGAGAGG GCGCTACCAATCCAA GTATATGTCCCAAGA GGGCTGGGACAGGGT	TAGATCCGTTTGTTATG ATTGGAGGGAGCCTGC CGAGTGGTATGAGCGC TCCAACTATTGAACCCA TATGATTCCCGAGGCCT GGCCAGACGCCTAGAT GCCTGCCACAATTGAA	MSGKRIVEMSGCDEKICQN KHCLKRRWAWISGPKGETS PPRKRGTCENVSFQDKSHAS DPDPLKAPEAREDAGSVAP QWVGEIKTPSLTSRDGVSEV VLPPQPVHAEGVSPASDSKD KATKITLLISLPVCDLRCGRC

				GGAAGAGTGCACCTC	CGCAGCCCTAGCTTGCT	ERPLETVGKAVRHFAVAHP
				GCGATCTGGGGCAGG	AGGAGATCCATAGGAA	TVSVVFKCQKCEKSSKNSHS
				GAAGGAATGGAGAGA	CTGGCCTATGGGGCGT	ISCHIPKCKGMTETRTDVEG
				AGTCGAAGAAGGCTT	CATGACGGTTGAAGTT	DHGCDHCQEKFTTAMGLTQ
				GTAGAGAAGGGGCTC	CCTCCATAGCGTGCTTG	HKRHRHIVQYCKEKEGEMT
				TCCTAGATCCTAACCT	GGAGGGGACGACAATG	ARRKGEVEAVKWSEWEESE
				GTATGACGCCCGTAA	ACGAGTCATGACGTAC	VARLSDGLAGLKMINRRIAD
				AACGGTGACCCCAGT	CGAGAGAACCCCAACC	SLGTGKTAEQVRQKRRRMR
				AGCGAATAAAGGAGG	CAGGTTGGGGGAGAGA	PEKVRCDKPKEAKDKSNLIK
				CAGGTGACAATAGAG	GCCAGCAAGAGCGGAG	MLSIPSATPTPQTGLKGFLLG
				GGCAGGGCCGACTTC	ATGCTTGGTATACCAA	ELNGVATKGEVQIGGVTLSL
				CCAGGTTTACATTGTT	GCTAGCAGAGAGAGGG	RGVEQDSALLNTSALELQRL
				GTACTTGTCAACATAA	TTGAAGAGGATGACTA	LGGRAGSANPLSLQRERETT
				AGAGGTGTCTCAATA	CTGGGCTCAGAGTCAT	LPSERRKTKQGEYRRVQKM
				GTTTGAATCAACAAG	CTCACCCTAAAAGGCG	FRSNEKKIAKYILDGNGDGE
				GGAGAGGAATACCGA	GTGGGGCATCGGTTGA	AASPPLEIALAFKSRWEEVE
				CCTGCTCCCTTGGGGC	ACACCTACCCATACCG	TFHGLGQFYSRGEADGVVF
				GGGGGTACTGGTCTTA	GGATGGGAGGTGGTAG	RSLISMSEVCENLGAIKNNT
				GCCCGGTTCCCCGCAA	GCCGAAAAAGAACAGG	AAGPDGITKPALLEWDPTG
				GTTTCCTTTGCCTGGG	AAGATGGTGGAGTAAG	AKLAAIFSIWLTSGTLPGPFK
				ATGTGCCTGACTGGCT	TTGAGAGCGGTTGCTC	KCRTTLIPKTDDPILLTQVAG
				CCATCCCCTTTCCCCA	GGGAAGTTATGTTGTG	WRPLTIGSVVLRLYSRILTH
				TTAGGCACGGCTAGA	ATAACTCCATTAAGGC	RLERACPINPRQRGFISSPGC

				TGACGCACCGATGGG	CGGTGGGCATGGTGCG	SENLMILGGLIKRSWAKGER
				CGGGTGTGTAGGTCG	GATAATGGAAACTATA	LAVVLVDFARAFDSVSHSHI
				CTACCGAAGGGGACT	AAAACAATAAAAAGAA	LEILRQRGLDEHIIGIVGDSY
				GGGGGTGTCCGGTGA	AGACCAAAAAAATGTT	TDVTTTITVSGEQSPPIDMRV
				ACCAGGACTTCCCAA	CTGTTATGATGCCTTAC	GVKQGDPMSPLLFNLALDP
				AATGGTCTCACATTTT	ACATGTCTGGGAGACC	MIDTLERYGLGYRMGEQQI
				TAAGCGGCTTGAGTAT	CCATAAGGGTCTCCCCT	TALAFADDLVLVSDSWEGM
				CGCCCAGTATCCTCGC	TATACTTCACTGGGAA	ACNIRILEEFCRLTGLRIQPR
				GCGGCACTGGGAACC	ACCCCATAAGGGTATC	KCHGFLIQKIQRARSVNLCK
				CAGTCAACCGCTCTGT	CCCCTATATTTACTGGG	PWIVCGEELHMVGPEESVSY
				GCCCCGGCGCAGGCG	AGACCCCATAAGGGTC	LGMKVSPWHGIMEPDPVER
				GGGGTTTAATGTCTCC	TCCCCCTATAGATGTAG	LCNWISSIGRSPLKPSQKVR
				CCGGCTTCACCGGCGC	AGCGTAAGGGGTCTCC	MLNVYAAPRMTYQADHGG
				TTCGGCGACGACGCA	AAAGTACCGGCCGATA	LGPIVLNVLDGMIRKAVKV
				GAGGAGCACCCGGAG	TGGCCTTATGGCAAACT	WLHLPLCTCDGLLYSRCQD
				GCCCCCATGAACTTAA	CTGGTGGTAGGGACAA	GGLGIVKLACQIPSIQARRV
				ACCAACCTATCTTGAA	GGAGGTAAGGGCAGTG	YRLWHSKEAITRVVTRRTV
				ATATGGCCTCTCGTTC	CCAACCCCTACTTGATC	EAEEYRGMWLRAGGSEAGL
				GGGTGAAGGGCAGGT	GGGACCATCCAGGGAA	PPLEDREEGAVQCTDTAGSV
				GGGAAGAGAGGGCTG	TGCCATCCTCCCGCGAA	KPKNPVIPDWRRAEFLKWQ
				CCTCACGATAAACAC	GGTGATGTGGTGAGGT	NLTAQGVGVQVFGGDKNSN
				CTAGTCAATAGCCAGT	AAGGGGGGAGCCCGTC	HWMANPETLGSKERHYIAG
				CGGGAAAAATGTGGA	TTCGAGTTTCCCCAACC	LQLRANVYPTREALSRGRPD

				ATGTTAGGACAGGGA	CCTACCCACAGGTGAG	LPKVCRQCLAGTESCAHILG
				GGTAAGGGAGGCGGC	AGGAGGAGAAGAGGA	QCPAVKDSRIRRHHKLCDLL
				TTTTTCGTAAGGCTCC	ATCTGTCCCCAACGGG	ASEAESAGWTVIREMCCRT
				TTCAACCCCTACCTGT	AGGAGGGTGAGGTGTA	RAGALRRPDLVFVKTGFAL
				AGTCCACCTATTGCAG	AGGGGGAGACCTTCTA	VVDVTVRYEMAYDTLMGA
				GTGTTGACAACATGC	GTAGGGTCTTCTCAGTC	AAEKVARYTPITPYVAMTL
				AAGATGACCTGCCTC	GCCTGACGTCCTGACTG	KARRVKVFGFPLGARGKWP
				GTTACGGGTCGCGTAT	TGGGGTGGATCAGTAC	GSNDRLLKAMGVGGGRRK
				CATTGCTACAGGTCGT	CCTACAGGTGAGACCG	QLAKLFSRRALLYSLDVLRD
				GTGCCGCTTCTAAGAG	GTGAGGTAAGGGTGTG	FYRAEGETGDLDDESVDDH
				GATAGTAAGGAGAGG	GCCCTCTTGAGGGCTGC	L (SEQ ID NO: 1463)
				TTATAGGGAGGTCCTG	GCCAACCCCTACTCGA	
				TTAGGGCTTCCTCAAC	GGTAACCTGAGGGAGT	
				CCCTCTCTATGCGATT	GGTGGAATGGCGGCAT	
				CCTTACAGGAGTGGA	GTTAGTGCTGGGACTTG	
				TCGAGAAGTCCCGGA	ATTGCGAGGGTTTAAT	
				CGTAATACACCCTGG	GAGAGTGGCCTGCTGA	
				AGGTAAGGGAGTGGC	GAGCAACACTTGTGGT	
				CTTCTAGTAGGGCTGC	GCTTAAAGCGGGGCGG	
				TTCAACCCCTCTGATG	CCCATGACCACCGTGA	
				GGAGTGTACCGGGAA	GATAGGACACTGCACA	
				CCTCGACTTGTAAGCA	GTGCAGCCATGAGGTT	
				CAGGTTAGTATGGGA	CCTGGAGGATGATGCG	

				GCAGAAGGGGGAGCC GTAATGGGCTTCTCTT TCACCTGCTTACATAA TACCTGTGGTGCATGT ATCTAGGTCTTGGCGG GAGAGTACTGAGAGA CAAGGTTGAGACCCC AAGATTGGGTCTCCCT AGCCTCTATAGCTGCG ACTCTTAGCGGGGAT ATGGAGTAACATGTA CCAAGGGAGTGAATA ATAAAGGAATTGACG GGGTACAAGTGACTG TTGGCCCGAATCCTAG CCACTTGATGACCAG GGATATATTACAGACT GAGAGCGAATCTAGA GACGTGAGAATAAAG TGAGAATTGGTTGAG CGAATACAGAGGAAG (SEQ ID NO: 1218)	ATGAGGTGGGGGCCTC ATCAGCCCCTCCTGGCA GGGCGTCGGCCAGGGA AACTAAATGTCTCTAGC ATGTCAGTGCAGTGAG GTAAGGGGAGAGCACT CTAGTAGGGCTCTTCCA ACCCCTACCTGTAGGTC ACCTGGTCCAGGTGTC GATGATGTGAAAACAA GAGCTACTTTGGTACCG GTCTGTTGCAAAAAGG GTTCTGCAGAGGACGA CGGCTATCCCTATCGGG AGGGAATAGTCGGTCC CAGGTAGTGGAAAATG GGGCTTTCCACTGAGC ATGAAAATGTGGTAGA GGTTGCGTCCAACCCA ATGATTTGCAGCAGAG CTCTTGGACACGAAGT CTGTATAGTCCCATGCA GGCAGCCAACCAGAGA	

					ATGGTGGCAAGACCCC AGCTCCGTATGGGAGG GGAGGGCCAAGATATA CGGAACGGCTGCTAAA GCGTTCTGCCGGTGTCA GTCTAATCACAGACAG CTGTGACGAAACAAAG TATGGGTTCCGACATGC TTGGTCAGCTCTTAGCC GCAAGGCTTAAATCGA ACGCAGCCCGCCGAGA GTGAACATTAAACGGG GATGGAATGTGTCTAG CGGTTACGTACTACCA GGGCTCAGGTTCGCCT GAGCCGAGGCTCTACA CGTCATGGTGGGAGTT CTCCCCACGCTCGTGAG GGCATGTAGTGGGATG GCATGTGGCGGACCAT CAGCTGGCACTACCAG GCCTCGGGCTTGCCCG AGTGCGGGACCTCACA	

					CATTGTAGGTGTGCTTG TCCCCCCTACGTTCGAA GACTTGAGGCGGAGAA TACTCATAGGCCCCAC GGCAAAGGGACACAAC ACGGAGGCTTGTGTCC GACGAGCCGTGGACTC CTATAGACAGCCCGGG ATATCACTGGGCACGC TCATACTGAAGAAATT CGATGAACCGGGCCTA CCGGAGCAAATGCACT CTAATCGCCTTTGTGGG CGACTGTGGCCCCCTCA TGCGAGTGAGGAATAT CATAAACTGCAATGGT TCAAAAAGTGATTCCT ATGGCTCGTCGGGGAG GGCTGACTGGGGCAAG CAAATGATTGAAAGGG GAAGAACCTTTTTCAAC TGTTTCTTGCCAAGCCC GGTTGATGGTGGCGCT	

					AGTAATTGCGACGGGA AAATGCGGTTTAAGTCT CCGAAGTAGTGCGTAG CACCGGATGTCGACCG GGTGTAAAAGCCCTTC GTAAAGTCCCTGGGGA GGTCAGTCCTGGGGCT ACTGATGCGCAGTATG TAATTCGCAGAATAGG GCCATCGATACCGCCT GCGTGACTCGACTGGG TTTCCACTTGAGGATAT CCGACCGTAGCGTGCA CCCTCTTGTAGTTGCGC CGGAAACGGCTGTGTT CCCTCACGTATGTGAG GAAACTCAACAATGTG AGTGGGTAAACGGCGG GACGAACTATGGCTCT CGT (SEQ ID NO: 1341)	
R2	R2-1_TCas	.	Tribolium castaneum	AGTCATAGAGCCAGA ACCTCCTCGTGGTCCC GCTGGGCACAGGGAT	TGAAAAGAGTGGCAGT TGTGGAGACTTCCTTCG GACGTCGGGTCGGATT	MSRRPGKSNEPPVRSRAMG LTTLSGTKTSNSGAQGPSTS APMQNMAGGFVCDCGRSY

				TAATTTTTCTGTGGCA AATTTGACTGGCTTCA GAGAGCGTTTTTCGAA GTGGWCTGTGTGACT GCGTTCCCCCCTTAGT TGCTATWTCCGCTKM GATTAACATCTCACCT CGACGTWTAAGATCA TT (SEQ ID NO: 1219)	TTCGGACGCCAGGGTA CCTCCACCGCTGGGTTC ACAAACTAGGCGAACA TCTGCCGATACCCTCTT TAGGTCATAGGACCAC ATGTCTCTGCACGAGAT TAACCCA (SEQ ID NO: 1342)	ALKTSLARHKKECGKNNAE CRWCGTRFNTLAGTRQHER KAHFVQYQSDLAKALPQPE SELMEKIAIVEARSXNGIFYK EMMASTGLTHQQVRSRREK PEYKGFLERARRSLAQTNIR AGSISPASTXAGSLESASPKA GCSSSASPGPTTRSRAPTKG VPXRSSNSARIVVEAQVHTR APPNTGETEVALRESRRTVP RLGXNPSRPCGISPLMAIAID EDSVLGGLRVQAGPSPTAV HSVEAFPGTSSMTPMETDRV HNKSGIDPILEHNGTRQVRR EESSTREDPVEQWSPNYPKT PVTMPNITTTADAXXTSYNR TPQTLPGNRRRRSRSLPPVQ RKSASDXLESVDSLGPWAV FLQDQVDAGSLSGNDSLAD LVRVALTKSDRGVLNDAVN RYLAQRAESLRIRKRGSKGK RKSKTGRHYGQTTSGSGQR AALFKKHQDLFLKNRRGLA

						ETILSGKEDFGPRPEPPVTSV EEFYGGIFESPSPPDNEPFEV RATGVEDPPPLTSPWTKRTL RSPCYWRRRPPPTYITMDEI KAARAGWQISAPGSDQIPVA AVKTMSELELAILFNIILFRN VQPSAWGVLRTTLVPKDGD LRNPANWRPITISSALQRLL HRVLAARLSKLVSLSSSQRG FTEIDGTLANALILHEYLQY RRQTGRTYXVVSLDVRKAF DTVSHCSVSRALGRFGIPSVI REYILATFGAQTTIKCGSVTT RPIRMLRGVRQGDPLSPVLF NLVMDELLEKVNEKYEGGS LQSGERCAIMAFADDLILIA DRDQDVPAMFDDVSTFLER RGMSVNPAKCRALIAGAVS GRSVVRTGSSYKIHNTPIPN VDALDAFKYLGLEFGHKGV ERPTIHNLSVWLNNLRRAPL KPDQKCLFIRQYVIPRLLYG MQNPQVTSRVLREADRLIRR

						HLKTYYHLNVHTPDSLIHAS VSDGGLGIMELRKAIPRIFLG RLVKLLNKNKDSVLSSVLQS NRVRTLMGKLSTMAGEVPE STFWRNQIASGPLSKGLEQA AEDSASRLWISEKPSGWSGR DHVRAVQLRTGNLPTKAIPS VPVGQRRCRHGCACDESISH VLQMCPLTHADRIRRHDEV VKKVARHCTSRGWTVEVEP HIRSRCGRLFKPDLAVHQPG GAIVIADVQVSWDSESLTVP YERKRAKYDVPQFHQAAQH AWPGKALTFAPVIVGARGI WPRINNDRSAALQIPPVVRR ACVNSVVKWGSSIHATFMR SVWANRLNPRPLRA (SEQ ID NO: 1464)
R2	R2-1_TGut	.	Tinamus guttatus	CTGGGGACCGTGGTT ACAACCCGGGCTTAG CTGCAGAGACAGTAC CTCCCCGTGGTTCCCG CCGGACCCCGTAACA	TAGGGGGCTTGGCATTT CTCATTGCCTGCTCCTG AAAGGATATGGGTCCT GCGTCGCGTGGTAGGC AGACCCATTCGTCCGA	MGSWIVNFVSVATQTGEFP VDTARRAPVPVTSYPESECH XPLPLTFCNSDVTIWGGVRP EPVDCLGDLPEXYDALPGV AGPREXVGGSPPGEGVRSPG

				TCGGGTGACTGAATCT GTCTCTGCCCCGGGAG TAGTTCCTCCTTGCCC TATTGACCAGCGGTCG CCGGCTGCTCAATAGT ATTCTAGGCGTGAAAT ATAGCGATAGTCCTA GTGGTTGTCTTACTGG GCCATAGCCCCTTGCT TCAGGGGTCATTCGCG AAGTCTCTCAGGAGA ACTGGGGGTGGTGTTC TTCTGGGTATAGCTAA ACCCCCTAGACTGTGT CCGATCC (SEQ ID NO: 1220)	GTAGGGGGCTTGGCAG TNTCCATTGCCTGTGCC CGAAAGGACGTGGGTC ATCTGGTCTGTCTGCCT ACACCTCTCTAGACTTG TAACATCTAGTCTGTCA ACAAGATCAAAATTCT TCACACAGACGACCGA GCTTGCTCAGTCTTCCT GTACCCGCAGAATTTTG CTCTTGCTCTCCTTTGG CTGTGTCCTGGACGTGG GACTATTCCATCTCGTC CCAAATGCCGCGTCCA ATTATACCGGATTTGAC AAAGCGGACGGCCCGC TTTATAAGCCGGAAAA GGTGCCTTGTAAAATTG CAAGGTTCATTAAATA G (SEQ ID NO: 1343)	IASPSGTAVQHDFGSPILVPG AEAAEVSTPVVKVPQDHPA CPCCGTRVVKVTALSEHLR RAHGRKRVLFQCSRCGRMN EKHHSIACHFPKCRGPPVEE GPLGAPEWCCEECGQKFNT KSGLSQHKRSVHPLTRNVER IEAARPKGKGKRGAHKGCW TEAEVAQLIELEGRFKNQRFI NKLIAEHLPSKSAKQISDKR RQLAAATKTSSPEKRVTSST SGESSPEVEKVEGIKREYRR RVGEWLCAGSLXDQTSFQK ILEDVESGSEIVTGPLEELAS FARGKLAAARVRHHRKHPA EAVPAREEQRWMKRRVGR RGLYLRFQRLFALDRRKLA GIILDDVESIKCPLPMEEVAD VFRRRWEEVAPFTGSGSFRS LGKADNGAFKPMISAKEVM KNVXEMSRRSAXGPDGLSL RDLMKIDPQGSRMAELFNL WLLAGRVPDQVKAGRTVLI

						PKSADPGKIGNIDNWRPITIG SVXLRMFSRILSARLRRACPI NRRQRGFIAAPGCSENLKLL QALIKSAKRDHRTLGVVFV DLAKAFDSVNHQHIFQVLV QKGVDGHIIDILRDLYTNAG TYLESGSQRSGFIKILRGVKQ GDPLSPILFNLALDPLLCRLE DRGLGYKYGDQQIXSLAFA DDLALLSDSWEGMQQSIRV VEEFCQRTGLRVQAPKCHG FLIRPTKESYTINDCDPWTIA DMQLDMIDPGSSEKYLGLGI DPWIGLSRPELSEVLTRWVK NIGGAPLKPLQKVDILRSYA LPRLLFIADHAGLSATCLHS LDLSIRSAVKGWLHLPPSTC DAIIYVSYKDGGLGLPRLAS LIPNVQARRLVRIAQSEDDVI RSVVLQEGIQEEIRKVWISA GGRPEKVPSVTGEFPVMEA QAADEALSEWERRAPRTIYP IPCKWRKREMENWTNLKSQ

						GHGIRNFENDRISNDWLLHY GRIPHRKLITAIQLRANVYPT REFLARGLGEGAPRGCRHCP AEWESCSHIIGYCPAVQEAR IKRHNDICGVLAEEARKLG WVIFIEPHLRDNTNELFKPD LVLVKGSCAKVVDVTIRYES GLTTLSDAAAEKARKYQHL AGEVRALTSATTVDFLGFPI GARGKWYVGNNGLLSDLG FSTSRVVRIARALSKKALLSS VDIIHIFASRARQAQTSE (SEQ ID NO: 1465)
R2	R2-1_TSP	.	Trichinella spiralis	CTCCTGACTAACCTGA TTTCGTCCGTGCGGCG GCGTTTTCTTTTCGCT CTCCGCTCGTCGAAAT TTGCTGTAGTTGATTC GCTTTTCTTTGCGTTTT CTTCTACTTTCGCAGT TTTTTCTGCATTGCCA CG (SEQ ID NO: 1221)	TGAGGTTTTTGTTTTCT TTTTTCCTTTTACCATTC TTGTTCCATTGTTGTTA TTTGCTTTAATCCTGTA TTTTACCGCCGGCAATT CCATTGTTATTATTACT GTTACTGTTATTATTGT TACTATTGTTTTTACTT TTACTTACTACTGTTAT TATACTTTAATTCGTTA	MSNRLANTAAAGGVPEKTS GTLDIPGQPSSSGEKRAISYP GPFGCNSCSFTSTTWLSLEL HFKSVHNIRDFVFLCSKCKK SWPSINSVASHYPRCKGSVK AAVVPTSLANTCTTCGSSFG TFSGLQLHRKRAHPDVFAAS CSKKTKARWSNDEFTLLAR LEAGLDPACKNINQVLAERL MEYNITRGVEMIKGQRRKD

					ACTTACGTTATTGTTAC CACTACTTACTTTGCTC TCTCGCAAACGTTCGTT GTTGTTTCTTTTGGACC AGGTTTAGAGAAATCG CACGCACAGCGGAACT GGACCGCTTAAGCCAG AAATAGTAAAGTAACA A (SEQ ID NO: 1344)	QYKALVRQLRSNSETQQCV GLAGSMDSNVPANDTSSSV ASEVSITYPEYGAVMSCDLI KEATGMAIVDINELQSNLRK AFLSGRKLPMKFHGARETA QKKMANPRVAKFKRFQRLF RSNRRKLASHIFDKASLEQF GGSIDEASDHLEKFLSRPRLE SDSYSVISGDKSIGVAHPILA EEVELELKASRPTAVGPDGI ALEDIKKLNTYDIASLFNLW LKAGDLPASVKASRTIFLPK SDGTTDISNCRPITIASAMYR LFSRIITRRLAARLELNVRQK AFRPEMNGVFENSAILYALI KDAKVRSREICVTTLDLAKA FDTVPHSRILRALRKNNVDP ESVDLISKMLTGTTYAEIKG LQGKLIPIRNGVRQGDPLSPL LFSLFIDEIIGRLQACGPAYD FHGEKICILAFADDLTLVAD SAAGMKILLKAACDFLEESG MSLNAEKCRTLCITRSPRSR

						KTFVNPAAKFIISDWKTGISS EIPSLCATDTFRFLGHTFDGE GKIHIDTEEIRSMLKSVKSAP LKPEQKVALIRSHLLPRLQF LFSTAEADSRKAWLIDSIIRG CVKEILHSVKAGMCTDIFYI PSRDGGMGFTSLGEFSLFSR QKALAKMAGSSDPLSKRVA EFFIERWNIARDPKVIEAARR VYQKKRYQRFFQTYQSGG WNEFSGNTIGNAWLTNGRA RGRNFIMAVKFRSNTAATR AENLRGRPGTKECRFCKSAT ETLAHICQRCPANHGLVIQR HDAVVTFLGEVARKEGYQV MIEPKVSTPVGALKPDLLLI KADTAFIVDVGIAWEGGRPL KLVNKMKCDKYKTAIPAILE TFHVGHAETYGVILGSRGC WLKSNDKALASIGLNITRK MKEHLSWLTFEIIFITQISRIY NSFMKK (SEQ ID NO: 1466)

R2	R2-1_TUr	Каркас_6	Tetranychus urticae	CTCTCTTATTTTAACA TATTCGATGTACTCGT ACATTGAATATGCTTT TATTTTTTTTCAAAGT TTTTTGGGTGCATACC CCTGGAAAATTCTGA GATGTATAAATCTCCC ATCAGCTTTGGCTGAA ACGTTGGCTAAGTTTT GTAGGTTGTTTGCCCC CTACTACTTAGTCGCA AATGGTATTTGCTAAC AGTTGTTAAATTGTTA CATTTACAAGTCCTAT CCAGTGCCTCCTCGTG GCGCTACCCGGTAAC ACTTAGAGTAATCTGA GTGGCTAAACTGGAA GGGCGGAAAATGCAA ACAGGCGGTTGGTAG ATGCTTCGGCATTTTG CCAAAAATCCACGGC TTTTTAGCCCAACAAC	TGATGTATCCCTTCAAT ATATTGTAATCCTCATT CGTCCCTATCCTTTCAT TTGATAAAAGAAACTT TGTTGCTCCTTTTAATA GTTGGTCCCTCCTGTCC CTTTTCTGGAACCTGGT TGTATCGATTATTGAAA GTTGCAATAAACGGAT TTAA (SEQ ID NO: 1345)	MCILGGLTSHSREGGLSRGS SQLKTVKPQNEEDNGTTQL KAGSADSFPRPSGDLNPEEP LSIDICPVCFRQMKSYLGVR VHMQKMHLEEYNASIPDPV VSHTRWSDEEAAQLAFTEA KIEVDKLLPRGKGINKFLLE LLPGRTLESIKSHRKRQSHK DLVRKYVKEFVDTLAADND DDTIICQDNGDIFNDPIVGAT DSQSETETVADPAEFKTFIEL ADDPTKPKVVAKLRNLIKD KPKSEILGSDILVRILRRTLH GLPVEDELDQYLEVYFTGKI KQRRSKTQTALSKKQIKQR DYGRLQELYSRSRKRCANEI LNPTSMSGGFGHQELSEFWT KTFGPDEQPTLGEVEIIPKEN CWWDIFSPISSDEIKASYPSI GKAAGPDNFSAYQLRKVPV WHLECLYNIFAFYKDIPSRL KDAKTILIPKKDNAESPGDF RPITLSSIITRHFHKILATRVN

				ATCAGGGTGATGGAC CCGCCAGCTTGTGGTC AGGATCCCATCCATG AATAAAGCATGGCTC TGCTTCTGGTGCATCC TCAACGGGATCGGCTT CGGCTGGATGTAAGT CTTGCGGAGGC (SEQ ID NO: 1222)		NFVRFHPMQRGFIQSDGCLE NTALIQTVIREAKVRRKQVH ITFCDVRKAFDSVRYDSIIAA IAKKGAPGSFIMYLSNLYRG NKTTLLTAGGETRITPTRGV RQGDPLSPILFNCVMDQILT ALPSRTGFTLSAGDESVNVN CLAFADDIILISKTKNGHQEL LDVTQRILKENGLDLNPDKC CSLSLIPHSKTKKIKVVRADF VVNGVKVRSMSIGDSTCYL GVSINVTGQVAPVKMYQAL CEKLDSAAIKPHQRLYILKH FVITKMFHPLILSTIAAHKIK NLDLISRRYVRKWLHLPHD CGSGMIHAKVSDGGLGVPL LFRTIADLKVRRKEKLQVHE NPIFRILAKLSTVSKELENCK KIASKTTDIQEKTFKEMLAT YDGLSLKEARAVPEVHKWV DSYDKRYKFAGRDFVQVIQ ARFNALPTRSRVWRGRGAD EKSLRCRAGCNARETLNHV

						SQSCFRTHRVRTARHDKILD FICERLDVVGVKYVREKPIS FPGKKLIPDLIVENTDQALV LDLQIVGDNSELPLDERGKN KVIKYNCSEMQELYKRKKK TLAVKALTLHYKGLMAPET SNILRSFGFKSKDLEKMAYM ALFGTVAAWGIFNRSTETM RSVANWPRPEEL (SEQ ID NO: 1467)
R2	R2-2_DWi	.	Drosophila willistoni	GAAGCTGGGAAGCTG GGTCGGATGAGCGCA GAAGGGGTGTTCTTCG GAGCACTGTAATTCAT AAGTCGTAAGTCTGAT CAAGTCGACTCGGAA CCTCTTCGTGGTGTTT CCTGGGTGCTGTTGAG TTCCTAGTCTCTAGGT TCTCTCCAGTAGCTAA (SEQ ID NO: 1223)	TAGATGTACTAACCTCT AGTTTCTCTATACTTTT GCCTGCTACCTTGGCAT TACATCTAAAAAGGTA CAAACATCGCATTGGC AAAAAGAGGTGGTTTT AGTACATAGGCGCTGT GGGACTTCATTGTCCCG ATGATGCAGCGAATCG TGCATACGAGATTGTCC AGTAGTTGGTTGCTCGT ATCTTTAGAAGATTTCC TTCCTCGGCGATCAAA	FERRSNSWGYQNLEPSNVG QDMNTVPRINNTTTTPATSR PGDQPREAIAVVNLAGEIPC AVCGRLFNTRRGFGVHMSH QHKDELDTQRQREDVKLR WSEEEAWMMARKEVELEA SGNLRFPNKKLAEVFTHRSS EAIKCFRKRGEYKAKLEQIR GQSTPTPEALDSITSQPRPSL LERNHQVSSSEAQPINPSEEQ SNWEIMRILQGYRPVECSPR WRAQVLQTIVDRAQAVGKE TTLQCLSNYLLEVFPLPNEP

					ANAAAAAAAAAAAAA AAAAA (SEQ ID NO: 1346)	HTIGRSNLRRPRTRRQLRQQ EYAQVQRRWDKNTGRCIKS LLDGTDESVMPNQEIMEPY WKQVMTNPSTCSCENTRFR MEHSLETVWSAITPRDLREN KLKLSSAPGPDGITPRTARS VPLGIMLRIMNLILWCGKIPF STRLARTIFIPKTVTANRPQD FRPITVPSVLVRQLNAVLAS RLASKVNWDPRQRGFLPTD GCADNATLVDLILREHHKR WKSCYLATVDVSKAFDLVS HQAIIKTLQAYGAPTNFVSFI EEQYKGGGTSLNGAGWSSE VFIPARGVKQGDPLSPLLFN LIIDRLLRSYPREIGAKVGNT MTSAAAFADDLVLFAETPM GQTLLDTTLGFLASVGLSLN ADKCFTVSIKGQAKQKCTV VERRSFCVGERECPSLKRTE EWKYLGIRFTADGRAQYSP ADDLGPKLLRLTRAPLKPQQ KLFAHRTVLIPQLYHQLTLG

						SVMIGVLGKCDRLVRQFVR RWLDLPLDVPVAYFHAPHT CGGLGIPSIRWIAPMLRLKR LSNIKWPHLEQSEVASSFIDD ELQRARDRLKAENVQRCSR PEIDSYFANRLYMSVDGCGL REAGHYGPQHGWVSQPTRL LTGKEYLHGVKLRINALPSK SRTTRGRHELERRCRAGCD APETTNHILQKCYRTHGRRV ARHNSVVNAVKRGLERKGC VAHVEPSLQCDSGLNKPDL VGIRQNHIYVIDVQVVTDGH SLDQAHQRKVERYDRADIR SQMRRFFGVTGEIEFHSVTL NWRGIWSGQSVKRLIAKDL LIAEDTKLISVRAVNGGVTS FKYFMYCAGYTRS (SEQ ID NO: 1468)
R2	R2-2_PM	.	Petromyzon marinus	CGGTGCGTTCCCTTGG GTAAGGAACACGAGT CTTAGTGGCCTTGACC TCCACGTGGTCCCGCT	TGAGAATAATGAGGTG CTAACCTCCCTGGGCCT GACCAAACCCAGAACA CATCACTGGCCAAGAT	RPQTKLMTDKLKFSSQLAR GLAKQRAMDGARVGDPPIT VRPTETDLCNTEGSWGRRP MKLLFVSVSTQTQNEDALW

				GGTAACATCATCTCTT GATGATGGCTAACAA GGCTAATGCACCCATT CCATCTCCTATCTCGC ATGGAGGCCGCTATG CGTGATTACTAG (SEQ ID NO: 1224)	GATTTCCCGCAGCACGT TGCTTTTCTCTCTCGAT ACCCGAGAATGTTCTG CGGAACTACGCAGTCT ATGAACAGTCACAGAC AACCTCTGATCCAAG (SEQ ID NO: 1347)	ASDVAKPMASRSALKMTSIP SMTFHNSSLEKEEEMNYDF YEQIKSLVESDDSSDDFTED DEDVEESFLDISAEEPVLGKF PIDTKGTITVVLPSLEYICVIC KQHMGKASELVAHFNIKHR DIPLVFKCAKCDKTNSNHRS IACHAPKCGGIKLTEESLPM VCECCQARFATLSGLSQHK RHAHPVTRNEERIKDGIKGT SQRGVHRSCWSLKEVEQLA LLELQFQGKKNINKIIAEALG TKTNKQVSDKRRDLSKKTG APMSDSLHFSSRPLETLSPPP NVTTGTSSILAQAAERLTNE NSGTLEKPAMEAIKAWLNG EGQHDALVETATALMLCPM RLVKNKGKRSKPENDIIKPRI LPTRSWMKKRAEKRGSFMK HQKLFFKNRSLLASLVLDGT ERHECRIPNADVYRFYCEK WEKVLPFNGLGQFKSSGVA NNEYFEPLISVEEVQTAIRAI

						KPTSAAGPDGLTRAAICAAD PEGRTLTALFNAWMITGIIPK ELKKNRTILIPKVMDDEKLK ELGNWRPITIGSMILRLFSRI MTARLARACPLNPRQRGFIA ASGCSENLKVLQDLMRHAK KLHRPLAVMFIDIAKAFDSV SHAHILWVLRHKKVDEHVV GIIQNAYDRCTTSFKSNGEST REISIRVGVKQGDPMSPLLF NLAMDPLICTLESHGVGYSI DTDHVTALAFADDLVLVSE SWVGMAANLAILESFCGLS GLEVQARKCQGFMISPTKDS YTVNNCDPWTIKNKDVHMI QPDESTKYLGLKICPWTGIIR SDLHVQLKTRISKIDEAPLKP TQKVELLNAYALPRLLYPA DHSDCKQSTLRVLDQEIIKA VKGWLHLPASTCDGLLYAR ARDGGLAILKLENAIPSVQV RRLQRIANSSDAIARNIASSQ GVEEEYRSLWVRAGGDSEA

						IPTFFLRGSESKEPVYPRPCD WRKRESRRRCEKPVQGRGI VNFAQDRISNAWLGPRCGF KQCFFIAALQLRANIYPTRES INRGRDGASRSCRKCSARLE SLSHILGQCPAVQKFKDCAT QTEAEMVPHGPAGNALPGW SLSRTFLVNVPQYKNSRIAR RNKISDILADEAARLGWWV YKEPRFTSEAGELRKPALVF AKGEEALVIDVTVRFELSRK TSSEAASHQVAYYTPPCDQ VKVLTKASNVTFFGFQVGA RGKVAP (SEQ ID NO: 1469)
R2	R2-2_SMed	.	Schmidtea mediterranea	AGTCATAGGGTGAAC TGCAATTCTGACACGA TGACCGAGCTGTGTCA GTTTGCAGCTAGTCGC TAAAGACTCGATCAG TCCGCCAAGTGAGGT GGCCGGGTATCTGCA GCACTAGAGCCACTG GTATCAAGAGCAGAG	TAGAAGGGAAACAAAG GAAAAACGAAATGACT GGAAACTATGAAGGAT ATAGCTGAAAGCCGCA AGGAAGGCTAAGTCCT GAAACCGATCTACATC TTCGATCCCAAGAGGA ACTGTGGGTTAAGCTTG AGCCGACGGAAAAAGC	RKELVTIKNLFEESGATAPA PVPLEVAVEVHQSSSVPEIT DESTTTQEGSYSEPPIHRCEN CGREFRTRAGVQQHRRKAH TNEFMEEKEKAAPTKKLRW TDEEKEILIESEIKIIKEGSLK EQHEINKILASRMPGRSQDG IAKIRQKQEHKAEIQRRLHG TVTTNETRGNRTSEITEPIRS

				ATACGCGAGTGGAAG TTGAGTACGACTACCT TCACGGGGTCCTCCTG ATAACCACAGTGGAC TGTGGGAACTAAATG TGTGCTCAGCGTTCCC TACTTTCTCGTAGGGT AAAGGGTATGATAAC CCAGAGAATATCCCA TGGGAGATATCCATG GAAAAAGCACCACGT TAGACAATCCGATGG TCTAACTCGGCTCCGA GGGGCTAACTATCCC AAAGGGCTTAA (SEQ ID NO: 1225)	GAATGCATGTTAGACG ACGAGGTACAGTCACC TCCTCGTGGTATTTGGC GGGCAATGCTCACTAA ATTAACTGTGAGTAGCT GAGAACTGTATGTGTA TCATGAAAAAAAAA (SEQ ID NO: 1348)	LPINTKTWSEDEMKRMLAE EVKLRTKNEKDINKKLAEIF PNRTMGSIKSKRTKDKDYQ DLVKLTMQTISENPDNETDF NTSNTENNSTDAEKEVKNY LNMLLLTINEEEWLTSTLKE AATLALQGKKTEASEKLNE YASKTLFPGLKITNQTRKRE KKISKRETRRQEYAEIQKLY KKNISSAAEKAINGKWSIKP EEEYHNNKDLIKAWKPILEA PPFSDCRPIENIKEMDYALM EISTAEIFLAIRAMGKTAPGP DGIKYSKLKKNIQSMAILFN TCLLTSFLPLPLKIARTILIPK QENPGILDYRPLTIASVVTR VFHSILAKKLDNNAQLSQRQ KGFRKCDGVAENIVILETILT NSRSEKRPLCMAFVDLRKA FDSVGHESIIRGAKRVGVPP MLLEYISSSYQNASTNLFGEI LNSRRGVRQGDPLSPILFNF VIDEALENLNRNIGYLLKEE

						KVSCLAFADDIVLIAETKGG LENHIEKLLEKLNGAGLELN ASKCATLMVMKNGKEKST YISTKAIKIKENDIPTMKATE TYKYLGLQMGFKAREQNA NEVITEGLENITRAPLKPQQR IHILRDFLIPRLIHKLVLGRV AKKSLKRIDQNIRKKVRNW LHLPKDTTAAFIHADAGDG GLGVPALEHTIPLLKRERITN LRKSNDPVTKECLRMEYTK QVLGKWSRPTKIGETLATN KSQLKEAFRKQMLITLDGK GLKDHHETPTIHKWIRRGEN MTGKQFITAVKIRGNLVATK SRNSRGRPEQEKLCEAQCGR PDSLGHILQGCWRTHGMRV ERHNNICRRIKAIMKGKESE VVEEPRLQTNEGLRKPDLLI CHKGKIIICDAQVVADSSNC SLESENQRKIDYYKKDSVVS EARKLIGRVDEDIIIMAVTFN WRGAISKTSIRDLDMLLDIK

						SKEVIKMSRKIIRDNSIMVE MHRNRTEKRR (SEQ ID NO: 1470)
R2	R2-2_TCas	.	Tribolium castaneum	TGGAAGACCCCGCCC ATGAGGCTTGGAGAG TGTGATCCTGATCACA CTTGAAAAGTTATGCT GAGTACGTCGTGAGA GTCGGTAACTGTCCCA GGATGGTCTGGGATA GGCTAAACCTCAGCA GGGGAAAGTTGTAGG GGCCTGCCACCCCTAC ACTTTTATAGATATGG CATTCGATACCTCAAA TAGAGCCTCGGACTTG GAGGAGCATGGTTCC CCTCCTCCTCGTACTA GACCTGGAACCAACG GTCTTGACAACCCCAT TGGACCTACGGGAGC GGACCATGCTCATGG ACATGGATTCCGAAG	TGATGCTCCTTTGGTTT TACCATCTGTGGGGGC ATCGGTCCTCACGGTTT CCTCGGGTTTCCTATTG TTTTTCCTAAACCCGAC AAGGAGCCCTTTGGCC CTCCTCCTTAAACACCT CTCCTTCATCCTSTTAG TCCATTCCGGCTAAAAT GATGAAGACCGAGGAG TGTCACTCTCTTGGCGG GGTTAACCCGTCCAAG TGTAAATGTGACCTCGC CATTCGGGCTCTGATA (SEQ ID NO: 1349)	MKSRSFRRIGDCAAGSSRRG VRLTGKAGREGRFAASPHLS PRYLAGSVSGNVPSVPPGPG LGAGAPAFAAGRNADGGPA QNPCPYCARSFTTANGRGL HIRRAHPDEANNAIDIERIHA RWSHEETAMMARLEAGAIQ RGGVRFMNQFLVPRMPGRT LEAVKSKRRDATYKALVQR FLQAPQINLPELRDGDAPRQ PDPQQENPPEPPSFDGAIRGA VADLVGGVDWQRLGFQGD RLCNIARRACDGGDVSGQL LGWLRDVFPVKRVSTRGDQ SDLDVDGALVSRRTARTRE YARVQELYRKEPKACLARIL GDRREGANRAPNRDPAFIDF WRGVFSEASAEVEGWAEEV SDHGELARRVWDPISVEEV GRSRVRNGAAPGPDGIAVS

				ACGAAGCGGGGGAAC		VWNKLPPEAAALLFNVLLL
				ACGGACCCCCCGCCG		GRCLPAELTRTRTVFIPKTD
				ATAATGCTCACTTAAC		APRTPADYRPISIASVVARHF
				GTCAGGCGAACCCAT		HRVLSARVQRIPDLFTKYQR
				CGAAATCATCTTGATG		GFLSGVDGIADNLSVLDTML
				TTACCCTTTCAAAGCA		TMSRRCCKHLHLAALDVSK
				GGTCATGCGGCATAT		AFDTVSHFAIVRACRSIFGSA
				GTCTCAATGCCGGAA		ETVLEEGGRRHFVQVRXGV
				AGGGTGGCTCCCGGG		RQXDPLSPLLFNLVLDRALK
				CGTAACGGACATCTAT		RLSTDVGFRLTDATKVTAL
				CGCTGATGGAASCTAG		AFADDVVLCATTARGLQTN
				CTCGCTCTTGAAGAAC		LDVLEAELRLAGLLLNPNKC
				GGCGGACGGGGACTT		QALSLVASGRDHKVKLVTK
				GGAATCGTGGTGTGG		PTFKVGQNTIHQVDASSIWK
				TTCTGATGTAAGTCCT		YLGIQFRGSGMCGCGSEGV
				GAAATTATGGCGTGA		AAGLKRITCAPLKPQQRMH
				TGGCCCGCCCGCCCG		LLRVFFLPKFYHAWTFGRLN
				ACCGGAGGGACTTAG		AGVLRRLDVVVRTSVRTWL
				AACCCCCTTCCGCGAG		RLPHDIPVGYFHAPTKSGGL
				GGTCCTGTCTGTAGGT		GIPQLSRFIPFLRLKRFDRLG
				CCWCCATCTCCGTAA		RSAVDYVRECAFTDIADRKI
				AACGAGTTGGAGGAA		RWCRERLSGIVDQVAGGRD
						ALDAYWTAQLHQSVDGRA

				ACCGCAGACGGGG (SEQ ID NO: 1226)		LRESASVASSTQWLRCSTRA IPASDWLHYTAVHIGALPSR VRTSRGRRGGQDVSCRGGC LLDETPAHCIQVCHRTHGGR VLRHDAIAKRISADLMELG WIVTREVSFRTTAGVFRPDM VAVKEGVTVILDVQIVSPAP TLDEAHRRKVAKYRDRADL ARYLAEAAVARGRAPPANI RFASATISWRGVWSAESVGS LRELGLSARHFDRYTTMAL CGSWRNWVRFNASTASRM GRGRGDASPRRHENQQ (SEQ ID NO: 1471)
R2	R2-7_MR	.	Megachile rotundata	TCTAGTTAGCAAGCG GCCCCCTCTAA (SEQ ID NO: 1227)	TAAATGTCGAACCGAA TTTTGGGTAACGTGCAC CCCACCATCCTTAATCG GCAGCACGCAATAAAG CCGTGGGCAGTGGTTTT AGTGGGTAGTCATTAG GAGTCCCACAGTACCC AGCGAACATCTTAGTG GGTCTGCGTAAACGCA	SGPATSTFGETKSRLCEPTSA LGCRPGAVVIQWAQIHKEK RKRIVGWPLGHLGSPTSLKL RHPRLQAKRIVPVLAELMQ CLCARHVSGRSPQKSAWAF TLEERTSACGVDAMFVCST CQRSFATKIGLGVHVRRAH VEVANAAISVERVKDRWSE EERRIMAAVEVRGVLSGAR

					TTTCCACTGCCTATCCT CCGGGAAAAAAAAAAAA AAAAAAAAAAAAAAA AAAAAAAAAAAAAAA AAAA (SEQ ID NO: 1350)	FINEYIMSHLQTSRTLESVKG TRKNPKYKELVATLLEEART SVREESPRSAVNDSATQPSG PSDTRSLRTEHLFTESTEPFE HRIRELIGDLEGVTDFRAELL VSIAEQQLQGDEVAESLTR WLGEVFKPENQQQQVQRKR RRQRKAPVSGQLPKWRERR RDYAAMQTLFHRNPSLAAG RVLDGKNESRPPDLPEMTAF WEPILTEQSAEHRAVGPASE KSELCSVWGPVEKEELLSSV PPLDTAVGPDGVTARQWRA VLPAVRALLYNIILKRGSFPA SMLESRTVFLPKKQHSVNPA DFRPISIASVVVRQLHKILAM RLRRTNLVDERQRCMDDGC AENITVLASLLDDARHGLKE LHLVSLDCAKAFDSVSHHAI DATLKECGLPAGFVQYISRT YSDSSTRLEVGRNRSEPIKT NRGVRQGDPLSTLIFCLCFD RVARTLSPHIGYDLNNTRIST

						LLYADDAFLVSTTAPGMNIL LRSVEESAGEVGLSFNTSKC SALSLIPSGKEKKMKVGTTP TFKTSQGFITQITPSQEWRYL GVDFQYSGPKKASRSLKIEL ERISKAPLKPQQRLLILRVYL LPRYYHHLVLSRTTLGHLRG LDLQVRAAVRRWLSLPRDIP IAYFHTTAKEGGLGLPAFET SIPCLMLARLRSMETSTCKA ARAAVQGFWVQKRIHWAT AALTKNGEALTCKADVDR WWASRLHKSVDGRELRECS GVGSSSTWVNSALNITGRD YVQYHHVRINSLPTRIRTSR GVRREGMEVTCRAGCQVTE TAAHVIQSCHRTHGGRILRH NAVCKVLASGLRDKGWEV REEPKLRTRQGLRKPDIVAI KDGVARVIDAQVVSGSGPL DEAHETKRKYYSDNGDVTA AIARECNIAPSNVAYSSCTIS WRGVWSPRSAADLLQVGLS

						KKLLGFITLRVLRGSHLNWT RWNKMTTMRVHHQRTGIG (SEQ ID NO: 1472)
R2	R2Ame l	.	Apis mellifera	TGGTAATCAAATGCCT CGCTATTTTAGTAGCG GTAGCGCTCCGCCCGC GCAGGAACCATTGAC GCCGCCGTAGTGTGG GTGATTTTATATCCAA CCAATCACGTCAACTA CGATCATTTGTAATCA CCGACGGTACTTGGTA GGGGTACCACATGGG CATTCTTGCTCATTCC ACAACGCCGCCTCCAT CATGGCAACAATTTA AAATATATATAAATTC TTAAGGTTTGACCGTA TTCATATATATATATA TATATTTAATATTACA ACCATAAATCTTATAT CGAGCCTTCTATTTGG TCTCAAAAGCAATAC	TGATCGTTAAAAGTAA AAATCTATTTATTTATT TTTATTCCTATATTATA ACACATTATTTATTTAT TTACTTATTGTTTTAAA GATGACGAAGCCGCAA GGCCAATCCAAATTTA ACAAAAGAACGAGACT ACTGGTCGACATTAAA AAGACGAAGCAGCTGC CAGCTGATAAACAACA GAGCCCGTCTCGGCCTT TACACCGAGCGGTGCA AGTCCTGACGTACTATT GTACGTCTAGGGCGCG GGGCAGATTCTACCGT GTAGAATCTGGGGCGA CGCCTCCGCGAGGCAC TCCCTGGACAACGTAC GCTAAAGCGTACGGCT	MSSNEEGASDTGAPGPGVP VADVSAADGRATYDDHGM STDYEKQTIELPLNGQIQCL WCHIEGRNQRFLQESQYLK HKDTQHPKGEIIWRCAACQ KEFEKLHGCRCHLPKCKGR KEAKGVAKFKCDSCEESFLT QRGLSMHELHRHPAIRNLK RTQGTSRGNTRPINRASVWS KEETDLLIKLNERYKHLKQP NVALKEYFPDKTLKQISDKR RLLPVQEPEDVATTDETGPP PSDSSEESIYESATEDEGGGD MQQTAPNDSWKEPFIQSIRT NHLEEEDSLRKVEEAIERMA MNEGVTEQEVGTLLEQFVD SLTQSPTTERKGSRRKSQKT TKRKTTHNNRKKFLYAKHQ ELYKKSPRRLLELALSGESSS GREVVNLPEADSVGPLYKSL

				GTTGTCAGATCTTGTA	AAGTGCGCCTCCCGAA	WGQIGPEKTHRNQPMCNNI
				GAACATCAGGAGTGA	AGGGTCCCCGTTCCTAA	DMSEIWTPIALESLVEKFKKI
				GCGGTGCGCTGTGGT	TTTTTCCGAGCCCGCGG	KSDTAAGADQIKKFHLRKK
				ATCCGTGCTTTGTGCC	GCAGATCTCGTGGCAG	GALHVFAKLCNLLMLHRIY
				GCGGCGACAAACCAA	TGACGCTAGAAAGTTA	PAQWKTNRTTLIPKPGKSAE
				TACGCTGCTGCCTGTC	AGTCCGCGGACATATA	EVENWRPITIGSLLGRIYSA
				GCAAAGCAATACGCT	AAATTACAGCCTTAAA	MIDRKLRSKIKQHIRQKGFT
				GCTGATTCTCGGATGC	TAATGAACCCCACGAA	QEDGCKNNIAILSSALTKMK
				GGGTGTCGACGGTCA	GGAGGTATCCTCGAAA	EDSGGIITIIDISKAFDTVPHG
				CGCAAAGCGATACGC	TTCCGCCACGATCCTTC	EISQSLMNKGVPSPICEYIQK
				TGGTGGGGTTTCAAA	TGATCGTAGGCGCAAA	MYIGCKTIIYCRDKKTLPVDI
				ACAATACGGCGCTGG	ACA (SEQ ID NO: 1351)	LRGVKQGDPLSPLLFNLIIDP
				TGCTAAAAAGCATTAT		IIGTLDETTEGIKLENENISVL
				GCCGCTAACGGCTGG		AFADDLVLLAKDKETADKQ
				ATTGTCGATCGCCGCT		NRLINEYLDDLKMKVSAEK
				GCGGGGGCTAGTGGC		CTTFEIKRQNKTWFLGDPQL
				GCACCCAGAGAGGTG		TLGQQRIPYADPEAAIKYLG
				CGACGCGCAAGCATT		TNFNPWRGLCKTSIKEIIDAA
				GGTTCTGTGCGAAGC		RTVKQLKLKPHQKINLIRTY
				GGAGTTCTTGAGAGT		LLPRYIHKLVANPPPLGTLD
				AATGGTTGCTGGGGG		LIDKELKTIIKEILHLHPSTTD
				CACAAAGCGCAACAT		GLIYTDKSHGGLGIQRVANI
				ATAGCCTCTTATGCCT		VKLAKLKHSILMTRSEDNA

				CAAGTCGTAGTTCGTA CCTCCACGTGGTCCCG CTGGAATGCCTATCGA CTCCTCCCCGGAGGAT CATAGAGTTCGAAAC CGGCTACGGCGAGGC AAGGGCGGTGAGGTG CACACCGATGGGGAG CAGCGACCCCACCTA CCCTTAGCTAAGAGA GCAGGCGATCCGCCA ACTGTCAGCACGAAA TAAACTAATCATATGT ATACGAGGGAGAATT TACAACGGGTACCTTG TGCCCGAACCGCCTGT AGGTATCACCTACAG GTGTTAAAATGAATCT GATAGCTGGCGGATC GTCGACCCTCTTTGAT GGCTCTGCGCCAACG ACTGGAAAGAATAGG AACGGAAGTCTAATG		VKIALNGQEGMVKRYATSI GLQWPCGIEEIEETRKKLKR ADTNKWKTLISQGQGIKEFF GDKTGNAWLYNPEMLRPSR YLDALKLRTNTYGTKAALH RAKRDIDINCRRCGVQVETL GHILGLCTHTKNKRIKRHDE ICDLIAKNVSKEYVIFREPEV EVNGDRRKPDMVIKDHDKV YVVDVTVRYENNDSLNKAY KEKENKYKETAEIMRRDLK AKESRVLPVVIGSRGAVPRA TIENLKVLGLQTKHALTASL IALRSSIEMANEFLDYDHTT (SEQ ID NO: 1473)

				GAAGGAAAGTGTCGG GAGCACTATAAATTCC CAAAGAAGAAAAGAA AAGAAAAAAAATAAA AAACCCAAATTAA (SEQ ID NO: 1228)		
R2	R2B_N Vi	.	Nasonia vitripennis	GACTAGACTATGGGTT CAGTCAGTCCCAAAT AGCCGATCCTGGCGC GTCCGGCAGTAATGC CACGTATGAGTCGGTT ACCCATCTCTAAACGC GTAGAGGTGGGGAGC TAAAGGCCAGGCGGT TTACCCGACGTCGAAT TTCTCCAGGTCTGTGT CAGTCGACGGAATAA AGGTACTACAACATCT ACTATCTATCGGGATC GGAAGACGCCTTACA GCGTTTTCCGATTTTT GCTCTTTGAGCATTTT TCTTCAAATTGCGATA	TGACCTGAACAAAACG TGTTGTCTTGTCTTGTC TAAAACTATTTATTCGA AATAAGGGGAGGCTAA CTGCCTGCAAGTTGAA CGCGAAAGTTAGACCT TCCCACCTAAAGCCCA AAAGTGATCGGGGAAT GAATCCGCGGGTGACC CCAGAGTTGGGTAAAC CCTTGAAACGTTGGAG AAGCGGAAGAGAGTCC CGCCACCGAGCATCGA GTGCTGCGGCGCCCGA ATGAAACCGATCGCGG ATGGTGCAAGTCGTAG GACGGGGCACGACCTA	TFAPTHPMVRSGPCRKTKRP GSDYRESLIMDSGNNVASEP RGAVDVTSAAPIGAELNAEP CEGRNQRREAALSAQTRRR NXARRARNAQQADEPGDDE EIETHGPLTIRTXEPMEIVAI AKNPQACPKCLQGGTQLLC MGSWELSRHINKEHPSVDV TWVCGACQRRCTTLRSWSC HVLHCKGRQEPKDLPFKCE HCSLSFDSQIGLSQHERHVH PEVRNDKRAAEANKPKGKS GRRPSIWSDEDLLLIRELESE YHGARNINEKIAEHFPDRTG RQVSDARRRKDYAALRGRG GPQGPAEGVEAIEEVDEGEI PEGEELVATDGAALESGPPE

				ACCGACCCGATCACG	AGCCTCTGTCACGGCG	NGGSAPAEQVNAPALESSSQ
				CGGGGCTTTGACAAA	GCGAAGCCAGGAATCA	QDRECSPAVGSDEQIEDSSD
				GCAATGCGTGGTCGG	CCATGCAAAGGTGTGA	DDEFSDALGEISLPEPLSVER
				TAAGATGGTTGCAATC	ACTGGGGCGGATACCT	TTISPPPRDDWKGPMRWEIC
				TTTTCCACCTCGTTTC	CCACGGGGTTTCCCTGG	NASEEAGSYANWVTGLQEL
				TTTTACGGAACGAAA	GCATCGCGCGAGCGAT	VRNNALSEIGLDSLYDQLIQI
				GCAATGCGTGTGGGG	GGCCAAAGTCCGCTTTC	MRHPSDDNEQDRLQLNARG
				AACGTTAAAAACTCC	TCAGCTACAAAACAAA	PPRRGHRKNRRRRRLTAAD
				CTTCATGCATCCCAGG	AATGGTATGAGACTTC	RKRFAFARCQDLWNNNPKK
				ATTTATCCTGCTTACT	GTTAACACTAATTTTTC	LAELVIANDLSILQRRQAPG
				GCAAAGCAATGCGTG	CGAGCCTAGCAGGCTC	RTETQTLYNELWGRVGPNIE
				TGGAGCGACTTTACCA	CCTTGACAACGCTTATG	APRRTEDPIPVSRIFTPITPQEI
				CGAGTCGCTCCACCGC	AATCTGGAAAAGGACA	MGRIRRIKNDSAAGPDGVT
				AAAGCAATGCGTATC	CAAAGTGGAAAAAGCG	KDDLRGRGVSIALSKLFNSIL
				GCGCAAAAGCAATGC	CTGATGGTGGACAAAA	LAGYYPKAWRENRTTLLPK
				GTGTGGGGGACTTGTC	GTCAGTTGAGACTTGAT	PEKDPADVKNWRPITISSMV
				AAAGATCCCCCGCCG	ATCAGTTGTTTTGACTA	SRVYSGLLDQRVRAVIKQC
				CAAAGCAATGCGTGT	AGAATTTTATTATCGTT	DRQKGFTEENGCFSNIQLLD
				CGGCACCACGTAGAG	GACTTTTAAATATTTTA	DAVSNAKKAGGVITILDVSK
				CAAAGCGTGTAGGCA	TTATTGACTGTTAATAT	AFDTVPHAVIQGCLEKKGIP
				GACTTTGTCAAAAGTA	ACTGACTTGGGACCAA	ETVAAYISSMYRDCSTAIRT
				GTTCTGCCGCAAAGC	GTCATCTCTGTTACCCG	RSGDVKIGMKRGVKQGDPL
				AATGCGTGTGGAGAT	GTACCGGTTCCTGTCAT	SPLIFNLVLEPLLERLQETSG

				CTTCGCCGGTGAAAG CAATACGTGTGGGCG AACTTAG (SEQ ID NO: 1229)	CAAACCGGAAAGTCCG TCCCACGTAATGTGGTA GACGCAGGAG (SEQ ID NO: 1352)	VEIEGMNLSCAAFADDIVCF ANTAPEAGRQLRMVADYLG RLDMSLSVSKCIAVEYVPHR KTWYTKNPGLEVNGNAVPS ISPSETFKYLGAKVSPWKGL LEGFESDAFREVISRVQRLPL KPMQKVDLLQMYIFPRYTY GLITSPPAKAVLKTIDRIIRTR IKEILHLPESVSSSFLYTPRK QGGLGLLEVEKMVLIAALR NGLRARQSHDPVTRAAMNS NAADDRLKSYADALRLHWP LTTKELDTYKYQLRLSYAQ KWAEQKWQGQGVEEFAQD PVGNSWLQRYDLLPASRYID AIKLRTNTYPTRALMKIIDG RVDSSCRKCQGSSETLGHIL GRCRYTKDKRISRHNEIKDL LKARLAKNHQVMDEPQITV RGQRFKPDLVVKTNEGRVH VIDVTVRYEHRTYLDEGRTE KIGKYRQILSTLRRDLHSNA EEVIPIVIGSRGAIPRETRKAL

						SKLGIGKSDWLTISLIALRSS LEIVNAFMDD (SEQ ID NO: 1474)
R2	R2Ci-B	AB0971 22	Ciona intestinalis	CGACGGTGAACCACC TTGTCGCGGTGTAAGA GCTTTAGTGTCTCGAA CAAGAAATAGCTTGT GTGCTGTCCTTCTGGG CGGTGCACATACTTCT TAACCTCCCGAGGCC ATGCCGGCGGGGGCT TTAGCCCCCGGCAGGT TTTACCATGCCGGACG GGTTCGAGAGGTAGA GGCCAAACTAAGAGT TCACCAGCAGACTTCG CACGCGGCTGGCCAC TGGCCGAAGTTTAAA CAACAGGGCCGCATC TTCCCAAACTCAATAT ATGGTGTTAAGTGAA CCGTGCCG (SEQ ID NO: 1230)	TGACAGTAATATGAAA ACATCACATCTGACCG GCACAGAATCACCATG CCGTAATGCACCCAAC TAAGGATTCCAATGGG TAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAA AAAAAAAAAAA (SEQ ID NO: 1353)	MGEWPWVSWSLTVLVEKW RPFTILQPYPMPGQLRVDVY LPRKTSYLMDKNIYENTTSP GGGPLCGEKTHRSDVIIPPPG FAPSTDTASNTLGENVDASA TTSSANPLSQEPGWCESCSK LFKSQRGLRVHQRSKHPELY HSQNQPLPRSKARWSDEEM VIFAREEIANRKIRFINQHLH KVFPHRTLESIKGLRGKNVR YARIMADLEAEMTSQPEAA TSLCTETSENLASSNVLPQT RGWAENLVENIDTAHLANL GPLSQFEPGKPSSSTKEAINT EYNDWISKWLPSGAAHRER RANPPSTKLNARATRRLQYS RIQNLYKLNRSACAQEVLSG AWKVQSGELNLKEVQPFWE KMFRKESAKDRRKPKPTGE VLWGLMEPLTIAEVGSTLKS

						TTPSAPGPDKLTLDGVKRIPI AELVSHYNLWLYAGYQPEG LREGITTLIPKIKGTRDPAKL RPITVSSFICRIFHRCLAQRM ETSLPLGERQKAFRKVDGIC HNIWSLRSLIHNSKDNLKEL NITFLDVRKAFDSISHKSLGI AAARLGLPPPLITYISNLYPN CSTKLKVNGKISKPIEVRRG VRQGDPLSPLLFNAVMDWA LSELDPRVGVQIGEQRINHL AFADDIILVSSTKIGMVSSIN TLSRHLAKSGLEISAGKEGK SASMAIVVDGKKKMWTVD PLPRFKVNSQKIPALSITQQY KYLGINIDAQGARNDAARIL TEGLAELSRAPLKPQQRLYL LRVHLLPKLQHGLVLSSCAK RALTYLDKSVRSAIRRWLTL PKDTPTAFYHAKACDGGLGI TRLEHTIPILKRNRMMKLTL SEDPVIMELVKLTYFTNLLH KYSNVKLLNSWPVTDKDSL

						ARAEASMLHTSVDGRGLSN CSDVPRQSDWVTNGASLLS GRDFIGAIKVRGNLLPTKVS AARGRQREITCDCCRRPESL GHILQTCPRTWGPRISRHDS LLKRVRNQACLKNWTPIIEP SIPTNIGLRRPDLVLAKGNIA FLVDATVVADNANMQLQH EAKVEKYNNSDIKEWIKVH CPGVDEVRVTSLTANWRGC LYGGSASFLTEDLGLPKAEL SLLSAKINEKGYYLWCAHY RGTARLWNRPLRS (SEQ ID NO: 1475)
R2	R2C_NGi	.	Nasonia girault i	CGGGTTCCCCCGACTT CGGCTTGCCGTGGTCT GGGGCTCACTGCTTTT TGTGGAGTCATGGTTA CATGGTGACCCTGGTT CCTCGCACCCCCGCTG GAAACTATCTGGGGA GGCCATGATTGGGTA ACGATAAAGGTCCTG	TAGCGGACTGGACTGT CTGGAGGAGTGTTTAA CTCGGGTTCTCATGGGA ACCCGACAACGTTGTT ATCTTGTATGACAATTC ATAAAAAAAAAAAAAA AAAAAAAAAAAAAA (SEQ ID NO: 1354)	WVTSPRRPRYVGPQKKKAS DGNDGRAAARAEPTNPGGP DRADDDEGDVKFWCEFPGC DRFFMTRSGRGLHHKKGHP DWNDQRNLAGKQHRKEIW SEEERLLLAKKEAELAISGA RFINVELRDFTARSLDAIKG QRKRPDYKILVEKFVRELRV RGIRQGVASRSQQARAMAV

				GTCGTGTCCTCCTGAG ATAGGCTGAATGGGT CACTAAGTGGCACCT AA (SEQ ID NO: 1231)		AGAPAATSSGAPPVATQPPP SGRVLRSQVVEAPAMEIPVA ESEGDSSGDELFEDVEPVRL SDLPPDRFTIYFAGLEIPGTE DIYAHRLHTICLMTTWRTKE EVRLELGLFLKDLFPSKGSQ ERPERTNLPDPRNRIERRRG EYKKCQDLWRRNKSTCVQR ILKEDLSQGECLPRELMEPF WNATFTQNPGTAPVLPPPTE VYSSVWEPIRPENIKGNYPP QNTAAGIDGLTVGDLKGVS REMLARIFNLFMWCGKLPE HLCASRTILLPKKPGAKVPG EFRPITVTSVLIRTFHKVLAE RLKVVPLDPRQRGFRESDGC AENVMLLDMTIRYHHERRR KMFLALLDMAKAFDSVSFE SMREVLTTKGIPTPFIEYFMT HLEDSFTVLQHGNWQSGKI HPTCGVKQGDPLSPPIFNFIM DEMLKRLPKEIGVNLDGLFV NAMAFADDLSLVANTEQGL

						QILIDEATSFLGLCGLRANPN KCVTLAIKTIPKEKKTAIDPS SHFRIGNAVIPSLKRTDEWV YLGIKFNSNGRLISDAKPKLI KDLELLTKAPLKPQQRLWA LKVIVIPGILYRGTLGSSTAG YLRSLDCVIRAYVRRWLRLP GDCPNGYFHAAVADGGLG VHPIRYKAMVDRLARLRKL EKSAYITGPEAARYLQRQVS IAENRLRDGANRIMSDASML REFLRELLYKSFDGRPLENS SKVPGQHRWVEEPTRFLSG ADYMNCIRARIAALPTAARC ARGRLKDKHCRAGCGNVET LNHVLQFCHRTHGTRIGRH DAVVKYVVGGLKKRGYAV KEEPKIVLQDVVYKPDMVA TKEGKTLILDAQVLGDQRD MRLAHEDKLRKYGAPEFKR KIRSETGSATIKSLSVTLSWR GLWGPDSVKGLLEEGVILK KDLKILSTRVLIGALAGWRR

						FNERTSMATSGRREEVTTR MVRRWKRRERVGVG (SEQ ID NO: 1476)
R2	R2La	JN937617	Lepidurus arcticus	GGGGTAGCAATTGAT CGATTCCCGCTTCCTC GTGGCGCTACCCTGG GTAATACTATGAGGA ATTGATCACACCGTAG CGAACGTCATCAGTC ACAGCTGCACGAATC CAGATAGAAATATAA CAGACGAGTAATTCTT TTAAAAGCTCGTCAGT AATCTTCCCAG (SEQ ID NO: 1232)	TGATAATCGCTCCATCC TGCAACTAATTATGAAT GCAAATCTGTTAAGTG ACATTAGTGATACTTAC CTGATACTTACCCTGGT ATTTATTTGACCTATAC TTACCCTGGTATCTACC TGACATATATTTATCTA ACCACCTACCTATGATG ACTCCCGCGGAAACTC TCACTTACCTTATTACC CACTTGGTCTTTTATTT TCTCGTTCCTTATTACT TTGTTCCTTTGGTGTAG GGTTCTCTGGTTTTTGG AACGGCTTCCTTAGCCG GAATTTTGTCTGATGTA TCTTGCTTGTGTCCTTG AAATATACGACCCAGG CTTGCGTCATTTAGGCT	MSGKSSKPRTVSSGSSSQET PPSGSNACDICGKCFMKPVG LSRVHPSQYHARLEKNQPK AKKFRWTDEDLYFLAKKEA ELLHLGSIKFVNKELAEFFPE KSVDQIRGQRRSETYKQQV LSIHSELLKLQTVADSPPPSR IPAKEVSAWLDFFLALPKTK NKFSEDKLDQLIRTAQDGTL ILDDLDLYLREVLVQPTSQG EKQAKLLPPPKSSREKRDRE YARAQNLYRKNKTACVNAI LDGNKKCENKIPDIDDFWKT IFESHSPPDAEPVCYVVDEEP TNIWSWISFFEMNHNYPDSS TSPGPDGVTARMLRSIPARV LNKLLNLLLFIEDLPAVFKC HRTVLIPKIDNPTSPGEFRPIT ISSIVVRQLNKIIAARVSEGV PINPRQKAFRQIDGCAENVF

					CTGGGAAA (SEQ ID NO: 1355)	LLDFILRDAKTKIKSLSLATV DIKKAFDSVSHHSIFRAIRGA RCPENLVNYIQNSYSGCTTQ ISVGGSISASKIPMNRGVKQ GDPLSPVLFNLVINEIIRKLP ASIGYPINSELSINCIAYADD LILVTNTREGLKLLLGLLNE ELPKRGLELNASKCFGLSLT ALGKLKKTHLCTSDQLDLH GTLIKNLTAEESWVYLGVPF SHIGRSKSFSPDLEALLNKLQ KSPLKLQQKLFALRVYLIPR LLHGLVLSRVAIGELKIMDK LILKHLRVWLRLPKDTPLGF FYSPVKLGGLGIKNLRTNVL KCRKQRIERMLVSPDDVVR LVAESEIFLKETDKLKDLLTI NGMCLDXRNVPRTGKNNKF WSERLYTSFDGKTLAYSEYF TQGGGWIREDKILQPAHVFA ECIKLRINALPTKSRVAHGR PTKDRSCRAGCLDVQKVPTI ETINHIAQVCPRTHGARIKR

						HDRLVQFLSLNLRKNPKRN VLVEYNFRTVAGIRKPDIIVI EDTRAVILDVQVVGDSSNLE MEYLEKSRKYSNDANFINA LQKLYPTVTNLTFHAVTFNN RGLIAKSTVAALRMLGVPPR CIMILCVISLEKTLEVWRMF NQSTASARK (SEQ ID NO: 1477)
R2	R2LcA	.	Lepidurus couesii	TTTGGGGTAGCAATTG ATCGATTCCCGCCTCC TCGTGGCGCTACCCCG GGATAGCCTCAAAGA AATTTGACGGTAAAG CAAAGAGGAATTGAT CACCCAAGGCAGTAC ATCGGCCTTCCTGCAG GAGCTCTGATAAAGA TATTAGTGAGTTATTC TGTTGAAGCTCGCTAT TTCATTCCCCTG (SEQ ID NO: 1233)	TAGTGCTTGAGTGATGC CTATCCTTTCTTTGATT AACTCTTACCATATACT TACCAGTTCTTACCCGT ACTTACCCTGTATACTT ACCTGTGTGCGTACCTG TGTACTTGTCCTTTAGC CGCCTTGTGTTTTTACC ATTGGTACTTACCTTGT GTGGTTGCCCGATACTT ACCTTGTATTTGCCTTG TAATTCTGCATGATATT TATTGTGTAGGTTCCTG ATGCTTACCTGATTTGT	MSEESRPKQTASKRGAAVE KTMMSGTYVCTLCGRSFEK SVGLSLHTNRMHPEAYNKL KEAKKPVLKKARWSEEEVF LLAQKEAELSFIGGIKFMNIE LHKIFPERELEGIKGQRKNPT YKAQVVSLLAEIRESKANDS SSSSSSSSSCDSASLGISNWL EFLLALPKTSNQFQEGRLDR LISDALRGVDVLENLDAYLL EVFAKPMAQNPCPKPPPPAK NSRERRDREYSRVQNFYKK NRSACINSILDGNTRSQNVIP GLTKFWTETFEKNSPPDDEA

					CCCCCTCATCATCTTTA GTTTCGTTCTATTTCAC TCCATTATGGAGTTCCG TTTGTTTTTTGGTGGAG GTACAGCACCCTTTAA GCTGGAATTGAGTGAG TTTATGTACTTTGGATG GTTGTAATAAACTACCC GGAGGCAT (SEQ ID NO: 1356)	PDQFVADEPRDMYKWITFY EMSQDYLDSSTAPGVDGFS AKQLRSMSPRVLNKILNLLL LSENLPNSFKMHKTVLIPKID DPKSPGDFRPITISPVLARLL NKILAARLSKLVPISQRQKA FLPVDGCGENIFLLDYILRSS KKSSKSVAMAVLDVKKAFD SVSHHSILRALNEAKCPINFI NFVRNSYDGCTTKLTCGGT SFPDSVRMNRGVKQGDPLS PVLFNLIIDSAIRKLPDSIGYV IRDGLKINCLAYADDLILVA SSRAGLKTLLNIVAEHLSLR GLDLNAAKCHGLSIIASGKA KTTYVSAADSLDLDGQPIKN LGVLDTWTYLGIPFSHLGRA EKVSPDLTNLLNKLQKAPLK LQQKLYAVRNFVIPRALHGL ILSKTNLKELNTLDRAIRVFL RTLLYLPKDTPLGFFHSPIKS GGLGITCFRTSVLKCRLQRI ARMRSSCDGVIQAVAESDIF

						ADEYAKLRDLIRINGNVLDT TESIKRYWAQRLHSSVDGK TLAYMDYFPQGNLWMSED KVSQRSYVFADCVKLRINAI PTRVRVSRGRPNKEMCCRA KCFDSQRMPAFESLNHITQV CPRTHGSRIQRHDKIAKFLF KNLNNCPSRSVLYEPHFVTV DGLRKPDIIIYDDSHMVVLD VQVVSDSANLEKEFECKAK KYANDVALRSAMLIKYPFIK SFSFVAATYNNRGLIAKSSV QVLRQLGLSPRSIMVSILICL EGTLETWRIFNQSTMNAH (SEQ ID NO: 1478)
R2	R2LcB	JN937619	Lepidurus couesii	TTTTGGGGTAGCAATT GATCGATTCCCGCCTC CTCGTGGCGCTACCCT GGGATAACCTCAAAG AAATTTGACGGTAAA GCTAAGAGGAATTGA TCACACCGTGACGAA TATCATCAGTCACAGC	TGATAATCGCTCCATCC TGCAACTAATTATGAAT GCAAATCTGTTAAGTG ACATTAGTGATACTTAC CTGATACTTACCCTGGT ATTTATTTGACCTATAC TTACCCTGGTATCTACC TGACATATATTTATCTA	MSGKSSKPRTVSSGSSSQET PPSGSNACDICGKCFMKPVG LSLHMSKVHPTQYHARLEK NQPKAKKFRWTDEDLYFLA KKEAELLLLGGIKFMNKELA EFFPEKSVDQIKGQRRSETY KQQVVSIHSELLKLQAVADS PPPSRIPAKEVSAWLDFLLA

				TGCACGAATCCAGAT AGATATATAACAGGC GAGTAATTCTTTTCGA AGCTCGTCAGTAATCT TCCCAG (SEQ ID NO: 1234)	ACCACCTACCTATGATG ACTCCCGCGGAAACTC TCACTTACCTTATTACC CACTTGGTCTTTTATTT TCTCGTTCCTTATTACT TTGTTCCTTTGGTGTAG GGTTCTCTGGTTTTTGG AACGGCTTCCTTAGCCG GAATTTTGTCTGATGTA TCTTGCTTGTGTCCTTG AAATATACGACCCAGG CTTGCGTCATTTAGGCT CTGGGAAA (SEQ ID NO: 1357)	LPKTKNKFSEDKLDQLIRTA QEGTPVLNDLDLYLREVLV QPTRQGERQAKPLPPPKSSR EKRDREYARVQNFYRKNKT ACVNAILDGNKKCENKIPDI DEFWKAIFESQSPPDAEPVS YVVDEEPKNIWSWISFFEMN RNYPDTSTSPGPDGVTARM LRSIPARVLNKLLNLLLFIED LPAVFKCHRTVLIPKVDNPA LPGEFRPITISSIIVRQLNKIIA ARVSEGVPINPRQKAFRQID GCAENVFLLDFILRDAKTKI KSLSLATVDIKKAFDSVSHH SIFRAIRGARCPENLVNYIQN SYSGCTTQISVGGSISTTKIL MNRGVKQGDPLSPVLFNLVI NEIIRKLPASIGYPINSELSIN CIAYADDLILVANTREGLKL LLNLLNEELPKRGLELNASK CFGLSLTALGKLKKTHLCTS DQLDLHGTLIKNLTAEESW VYLGVPFSHIGRSKSFSPDLE

						ALLNKLQKSPLKLQQKLFAL RVYLIPRLLHGLVLSRVAIG ELKIMDKLILKHLRVWLRLP KDTPLGFFYSPVKLGGLGIK NLRTNVLKCRKQRIERMLV SPDDVVRLVAESEIFLKETD KLKDLLTINGMCLDXRNVP RTGKNNKFWSERLYTSFDG KTLAYSEYFTQGGGWIRED KILQPAHVFAECIKLRINALP TKSRVAHGRPTKDRSCRAG CLDVQKVPAIETINHIAQVC PRTHGARIKRHDRLVQFLSL NLRKNPKRNVLVEYNFRTV AGIRKPDIIVIEDTRAAILDV QVVGDSSNLEMEYLEKSRK YSNDATLSMRINALQKLYPT VTSLTFHAVTFNNRGLIAKS TVAALRMLGVPPRCIMILCV ISLEKTLEVWRMFNQSTASA RK (SEQ ID NO: 1479)
R2	R2Nvec-A	.	Nematostella vectensis	GGTTGGGGCCTTCTCG TGGCGGAGTCGTGAG	TGATGGTGGGTTACTCG CCTCTGTGTAACAGGC	MLRGTGNMNDKRDGSATA DPTSALLGAVGDGSLVCNL

				TAAGGGGTATAGGGG TAAGGGACACCACGG ACCGAGAACGGTTAC CGCTCAAGGCGAGTG GTGGAAGGCATAAAA TCGTAACGCCGCCCTC CGACCTGCTCCTGAAA CTAATGCCAACCAACT GACTGTGGGGCTAAC CTCCCCAGAGTCAGG (SEQ ID NO: 1235)	AAATGAAAGCTGCGCA AGCAGTCGATGAGCCA AAGCCGCACAGCCCCC GACTGGGGTACAGGGC AGCCCTGGGCTATGCC CGAAGTCTTTTGACAG GTCCAAACTTCACCCTT GCCGCCAGTAGGGCAC CAGGGCCAGGTGCAGG TGGGCGCTTTGTTCTAT TTGGTTTCGCTTAATTT TTGTTAAATTTTTCCCG TGCGCCCACCCTTTTTA ACCCTTTTAACACCAAT TTTTTTTACAACCCTCT ACTCAATAATCCAAAT GAATAAAAGCGGCATA ACAGGTGAACAC (SEQ ID NO: 1358)	CGLACKSRGGLSIHRRSKHA TVYHAERQPAPRAKARWTN DEMILVARKQIASEKSRCSA VVEGMREAVPHRTFDAVKS LKTKNRNYTRILEQIRAECS EEEVIESGVLKDRTENVCVQ TTSNVPGSAGRAASVELEGN IQVGHQLAQKTMAGNNSRK QPANHTNWAEFNIEEGNITL RKSKRKANGMPDATHRPGP PTVDSLKHPVCLLQGAADK RDEPHTVEQLYYNIEEGMPL AEEQQWSEKLFDAIDSSLLS VEVELGRIVPGCPDEETRQLI DREFLDFIHSYSREKPPQRGL AKSKPPPKGPKSLRRQQYRQ LQRLWDKNRSAAAEQALTG KWQEVRTAAGVPLSLMEVP WREIFETPSTTDVREPAPAG PVLWQLLRPVTIAEVEDAIS SKKSASGPDGVPCAALQTM GAASLAAHFNLWLLAGTQP KRLTECRTIFVPKEVNTHLP

						LHHRPITIGSVVVRLFHQILG KPMEAVLPLGSGQRGFRKG DGICQNIWLLHTLIRRSTDLL RPLKLVFLDVKKAFDSVSHE SLLIAAKRLGVPGPLLTYINE LYSRSETVFEVGGESSGSVK VSQGVKQGDPLSSTLFNCVI DWAVSDLDPHIGVLLGESR VSFLAYADDLVLLSETEAAL TSQLNSIEKSLAHCGLKLST GDSGKSASLNIVIDGKAKR WVVNPTPFLRASGGEIRSLV ANETYKYLGINIGAQGVKA AEYNAFKEALDNLSRAPLKP QQRLFLLKTYLLPQLHHSLV LSRTTGKLLNSLDALVRKA VRGWLKLPHDTHRAFFYAH QADGGLNVPSLYHLIPLLRR SRYERLTRVEDPEIREVSRT DYFKRVLGAAAAATTVAGH RIDSKVTLRLAWREALYAS ADGRGLSQCPLVPEVHSWV TDVSGLQTGSQYISAVRLRG

						ALLPTAVRKSRGRGGVNSN CDCCGRGQPEFLGHVLQTC PRTWGSRISRHNHVLSLIAK ACRSRRWQVLEEPIIQTPAQ LLKPDLVIWNHQAAYVVDV SVPGDNTPLSTCHNRKVAY YSGESVREWVRSKTGHNPT VSSVVINWRGAMAKESYRL LTKDLRLAKTLPRLLVLRVL EGGHGIWLNFHRSTFAVGV T (SEQ ID NO: 1480)
R2	R2Sm-A	.	Schistosoma mansoni	ATGTTTTAATTTATTT TTGAACTACTACTGTC TGAGTGCTTCTTACAA CCTGAAGGCTCAGAA ACTACCCACTTTTTGC TGTTTATCCACAACAA CAGTTGTGAATCTATT CTCCAAATATTCCTTG TGCTTTTGTCAACATT ATTCTATACCAACTGT ACCACCTACTTCTTCA TCTCACGTTTTAATTC	TAGCCCCCTTCACTCTT AGACATTCCCCCACTGT TGTTGCTTATCTTCATG TTTTTGTGTTAATTGAC TGCTCTCTTCTGGGTTG ATGTCTGATTGTCTCTC TCTCTTTCCATATTGCT TGCTCTCCCCGCTTACT TCCAATAGTTGTCATAT TATGTCTTTGTTTACTT GCCATGTCTAACGACA ATTACTTTATCTACCTT	MPVSTGAETDITSSLPIPASSI VSPNYTLPDSSSTCLICFAIFP THNILLSHATAIHHISCPPTP VQDGSQQMSCVLCAAAFSS NRGLTQHIRHRHISEYNELIR QRIAVQPTSRIWSPFDDASLL SIANHEAHRFPTKNDLYQHI STVLTRRTAEAVKRRLLHLQ WSRSPTAITTSSNNHTTTDIP NTEARYIFPVDLDEHPPLSD ATTPDASTHPLPELLVILTPL PSPTRLQNISESQTSHESNRN

				TGGTCTAATTTTCTCA TCATTAGTCACGGAG AGGGCCTATGAACGG TCCGTGACGCGAAATT CAATCCACGAATTCGT CCTCTTCTGCTAGTGG TCCCCGAAATACGGTT CCTCTGGCCTGTCAGT TGTGTTAAAACTATAT AATAACG (SEQ ID NO: 1236)	AGTTGGTCCTCTTGGTT TGGTTGCCTTCATGTGT TCATGGCGGAATCTGA TGTTTATAATGACTATT CCTACTACCACCATTAC AACTATTATTATTATCA CTATTATTAACATTATT ATTACTTCTACAATTAG TATTATGGCTACTCCTT TCAGCACACCAATAAA ATCTCAATCAAACATCT CACTTATTAAACTCTCT ATTTCCCCTTCGTTATA AACTTACAATTCAGTTT AACCGAATATCTCTCTT TTACAAATCTTAAGTAT GTAATTTTGTGCCAAGC CCATTTGGGTCTGTACA ATTTGATACTTAAAAAT AAATGTTAT (SEQ ID NO: 1359)	SMHTPPTYACDSDESLGVTP SSTIPSCFHSYRDPLAEQRSK LLRASASLLQSSCTRIRSSSL LAFLQNASTLMDEEHVSTFL NSHGEFVFPRTWTPSRPKHP SHAPANVSRKKRRKIEYAHI QTLFHHRPKDAANTVLDGR WRNPYVANHSMIPDFDCFW TTVFTKTNSPDSREITPIIPMT PSLIDPILPSDVTWALKEMH GTAGGIDRLTSYDLMRFGK NGLAGYLNMLLALAYLPTN LSTARVTFVPKSSSPVSPEDF RPISVAPVATRCLHKILAKR WMPLFPQERLQFAFLNRDG CFEAVNLLHSVIRHVHTRHA GASFALLDISRAFDTVSHDSI IRAAKRYGAPELLCRYLNN YYRRSTSCVNRTELHPTCGV KQGDPLSPLLFIMVLDELLE GLDPMTHLTVDGESLNYIA YADDLVVFAPNAELLQRKL DRISLLLHEAGWSINPEKSRT

						LDLISGGHSKITALSQTEFTI AGMRIPPLSAADTFDYLGIK SNFKGRCPVAHIDLLNNYLT EISCAPLKPQQRMKILKDNL LPRLLYPLTLGIVHLKTLKS MDRNIHTAIRKWLRLPSDTP LAYFHSPVAAGGLGILHLSS SVPFHRRKRLETLLSSPNRLL HKLPTSPTLASYSHLSQLPV RIGHETVTSREEASNSWVRR LHSSCDGKGLLLAPLSTESH AWLRYPQSIFPSVYINAVKL RGGLLSTKVRRSRGGRVTN GLNCRGGCAHHETIHHILQH CALTHDIRCKRHNELCNLV AKKLRRQKIHFLQEPCIPLEK TYCKPDFIIIRDSIAYVLDVT VSDDGNTHASRLLKISKYGN ERTVASIKRFLTSSGYIITSVR QTPVLTFRGILERASSQSLRR LCFSSRDLGDLCLSAIQGSIK IYNTYMRGTQRLNE (SEQ ID NO: 1481)

R2	R2Tc	EU854578	Triops cancriformis	TTTTTGGTCTGGCATTTGATCGTTTCCGCCTCCTCGTGGCGCCAGACTGGGTAAGCTGATTTATCAGTGAGCTAAGAGAAACGATCACCGCAGGAGTCCATCTACCTGCGCTGCACGTGATTTCATTGTGCACTTGGCGAGTTATCCCTTGTGGAGCTCGCTCGTTAGCTTTTGAG (SEQ ID NO: 1237)	TAGATGACTGCCCTACCCCTTTGCTGCCGAAGAACTACTGAAGACTATTGAACTAACCAACTGTGTTAAGTAAGAACTAATGCCTCTTTTTCCCTGCATGTATCCCCTGATCAGTGACTTATTTTCTTTTCCTGTTGCGCCCTTTGTTTAGTTATTTCCTTTAATTACTAGAATTATCTTTTCGTTCTCCGTCTAATTGCTTTTTCGGTGTAGGAACGGCTACCTAAAGCTGGAAGTGGGAAGTGTTTTCAATGTACTTTGTGATTATAGAAATATATGACCCGAGGTGCATTGTTTGGCATTTCCTCGAGAAA (SEQ ID NO: 1360)	MSQKRRPEKAVPDEGATAHDVAQPDKSKCSVCGETFKGPASVTMHMVKKHPVEFNELKMAKKPVPKKVRWSEEEIFQLARTEAELTLQGVRFINVELQKIFPAREIEGIKGQRKLAKYKELVKDQLDEIGRAPNPPEQEIGEDVPSPFKAWLELLLALPKTPNDFLEHKLDNIIVQALKEDVNSDQVFNDLNSYLKLILEPSGRAKSVPGEIIHGDPSGSAKTSVTKAPKPATVSSSRKKRRDAEFARIQRLYRKNRTSCINTILDGNTREHEAPKNMEGFWREIFERESPDDPDDPDIFLEEEASDIWKYISFYEMCNLYPPPSTAPGPDGFSSKDLRRMTPRVLNKILNLLLHLRDLPQILKSHRTVLIPKTDLPTKPGDFRPITISNILVRHLNKILANRVSHLIPINERQKAFLPIDGCAENIFTLDFILHHARTKIKSLSMAILDISKAFDSVSHHSI

						FRALREARCPIGFIKFIENCY GGCFTKLFCGGVKYPSEVS MNRGVKQGDPLSPVLFNLVI DGLIRQIPSALGFNVSDQVK VSCIAYADDLILIATTRAGLK TLLDLTNSYLAKRGLSLNPD KCSALSIVASGKQKLVYIAS SEHFDLAGQKMRNLNVGDS WRYLGIQFSHLGRAEKVTP DLTCLINRLQKAPLKLQQKL YALRIYLIPRLIHGLTLSKTN LGELKTLDKLIRKYIRAWLH LPDDTPMGYFYTPLKAGGL GLPSLRLVILNNRLERILRM KASQDIIVRTIAESETLGVEI RKLHDLLSIDGTILDTSVKIH SFWAERLYSSYDGKCLCNS ANFPPGNKWIGEDSLNQRSH IFADCLKLRINALPTRSRTAR GRPLKDKPCRAGCRNSDGV KVIETLNHITQVCERTHGAR VKRHDRLVDFAVKGLQRPH RVVLKEPHYKTVNGVRKPD

						IVIKIPDHTYICDFQVVSDTS CLELEFRKKALKYAEDKGL CDQLTRDHPGELSFTAITFN TRGLIAKSSVTALRKLGMPP RSIMTLQKICMEGSLEIWRIF NQTTAMARN (SEQ ID NO: 1482)
R2	R2_DAn	.	Drosophila ananassae	AGAATATGGATTTGAT TGTGCAGAGGGGGTG CTATACCGTAACTCGT AAGCCATGCAATCAG ATCAAGTCGACTCAA AACCTCCTCGTGGTAT TCTCTGGGTGCCAGTA TTTACTGGTAGCTGA (SEQ ID NO: 1238)	TAGCCAATGCACGGGT TCCAGATTAAGCTTGCT GCCGAAGCATACCATC AAAATCGGCATAAAAT TCGCTTAATAAAGGAG GTGGTTTTAGTACGTAG GCGTCCCGGGACTTGTC TCGGGATGAATCGTGC ATGCGTATAATTGGGA TCGATAACAAATACCA ACTAAGTTATTACTAAT ATATCGAAATACATAA ATATCCCGTCCTTACGT ATCTTTGAAGATTTCCA TCCTCAGCGAACAAAA	FERRKDPWGYRPPGTLKQIG ATENNEPRNLNRFVRGESTA SSLESTQFGTSAEVNLAGRV PCTICEMTFSSKRGLGVHMS HRHKDDLDAQRLRVDKKA RWSEEETLMMARKEVELAA SGVRFLNKKLAEIFTHRSAD AISSYRKRSEYKAKLEQIRG QSVPTPEAEEINTTQRRPSNS EQNRRVPRSEGGPIAPTEQT NNEILRVLQGLAPVVCLPR WRAEVLQNIVDNAQVSGQE TTLQSLSSYLMEIFPPRNEPH ILTRPRTEPRNMRQRRRQQY ARVQRNWDKHPGRCIKSLL EEDDESVMPNQEVMEPYWR

					AAAAAAAAAA (SEQ ID NO: 1361)	RVMTQPSSSSIKRDMFNME HSLERVWSAVNQRDLRATK VKLSSSPGPDGITPKTARSVP EGIMLRIMNLILWCGNLPYSI RLARTIFIPKKATANQPQDY RPISVPSVIVRQLNAILASRL SAAINWDTRQRGFLPTDGC ADNTTIVDLVLREHHKRFKS CYIGTLDVSKAFDAVAHEA VYNTLASYGAPKGFINYLRK AYEGGGTMLAGNGWVSEA FIPARGVKQGDPLSPILFNLV IDRLLRSLPSEIGAKVGNAM TNAAAFADDIVLFAETPMGL QKLLDTTVCFLSSVGLTLNT DKCFTVSIKGQAKQKCTVV ERRSFLIGGRECPSLKRTDE WKYLGIKFTAEGRARYDPA EDLGPKLLRLTRAPLKPQQK LFALRTVLIPQLYHKLTLGS VTIGVLKKFDKLVRYTARK WLGLPVDVPVSFFHAPHKS GGLGLPSLRWTAPMLRLKR

						LSNIKWPHLERSEVASSFVE EEMRRARDRLQAGSEELLT RSQVDSYLANRLHMSVDGC GLREAERFAPQHGWVSQPT RLLTGKEYTDGIKLRINALPS RSRTTRGRHELERRCRAGC DAPETTNHILQQCYRTHGRR IARHNGVVNFLKRGLERRG CVVHVEPSLQGETGLNKPD LVAIRQNRIYVIDTQIVTDGH SLDQAHQRKVGKYDTPDIR TNLRRSFGAFDIEFHSATVN WRGIWSGQSVKRLIASDLLS SGDSNIISVRVISGGLWSWR QFMYLSGYTRDWT (SEQ ID NO: 1483)
R2	R2_DM	X51967	Drosophila melanogaster	TTGGGGATCATGGGG TATTTGAGAGCAGAG GGGGAGTATTCTTCTG TAATTCGTAAGTCATA TCATATGATGTGCGGA AGGGGAATTTTACTCT GTAACTCACAAGTCTC	TAGCTAAATCGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC AATAAAGGCATAAACA TCGCAAAATAATGGTT ATAATTAAATGGCTAT GAGGATGGTTTTAGTA	MTTRPSVDIFPEDQYEPNAA ATLSRVPCTVCGRSFNSKRG LGVHMRSRHPDELDEERRR VDIKARWSDEEKWMMARK EVELTANGCKHINKQLAVY FANRSVEAIKKLRQRGDYK EKIEQIRGQSALAPEVANLTI

				TCCTTTACTCAAGTCG ACTCAAAACCTCCTCG TGGTGGTCCCGGTAAT GCTAAACTCGTTTAGC AGCTAATTTGAGCGG AAAAACTTTTCCGATG GGCTGGTTCCCCAGA GGAAATTTATTCATAT TGGAACTACAAGCAC AAATAACGAGCCTCG GATACCTTTACACAAT CTG (SEQ ID NO: 1239)	CGTAGGCGTTGCGGAA CTTCGGTTCATATAGAG CAATGAATCGTGCATG CTAGGAAAACTGACCA CACACAGTGTTGGCAG ACCTAGTATCTTTCGAA GATTTCCATACCTCCGC GATCAAAAAAAAAAAA AAAAAAAAAA (SEQ ID NO: 1362)	RRRPSRSEQDHQVTTSETTPI TPFEQSNREILRTLRGYSPVE CHSKWRAQELQTIIDRAHLE GKETTLQCLSLYLLGIFPAQ GVRHTLTRPPRRPRNRRESR RQQYAVVQRNWDKHKGRC IKSLLNGTDESVMPSQEIMV PYWREVMTQPSPSSCSGEVI QMDHSLERVWSAITEQDLR ASRVSLSSSPGPDGITPKSAR EVPSGIMLRIMNLILWCGNL PHSIRLARTVFIPKTVTAKRP QDFRPISVPSVLVRQLNAILA TRLNSSINWDPRQRGFLPTD GCADNATIVDLVLRHSHKH FRSCYIANLDVSKAFDSLSH ASIYDTLRAYGAPKGFVDY VQNTYEGGGTSLNGDGWSS EEFVPARGVKQGDPLSPILF NLVMDRLLRTLPSEIGAKVG NAITNAAAFADDLVLFAETR MGLQVLLDKTLDFLSIVGLK LNADKCFTVGIKGQPKQKC

						TVLEAQSFYVGSSEIPSLKRT DEWKYLGINFTATGRVRCN PAEDIGPKLQRLTKAPLKPQ QRLFALRTVLIPQLYHKLAL GSVAIGVLRKTDKLIRYYVR RWLNLPLDVPIAFVHAPPKS GGLGIPSLRWVAPMLRLRRL SNIKWPHLTQNEVASSFLEA EKQRARDRLLAEQNELLSRP AIEKYWANKLYLSVDGSGL REGGHYGPQHGWVSQPTRL LTGKEYMDGIRLRINALPTK SRTTRGRHELERQCRAGCD APETTNHIMQKCYRSHGRR VARHNCVVNRIKRGLEERG CVVIVEPSLQCESGLNKPDL VALRQNHIDVIDTQIVTDGH SMDDAHQRKINRYDRPDIRT ELRRRFEAAGDIEFHSATLN WRGIWSGQSVKRLIAKGLLS KYDSHIISVQVMRGSLGCFK QFMYLSGFSRDWT (SEQ ID NO: 1484)

R2	R2_DPe	.	Drosophila persimilis	AAGATATGGATCTGAATAATAGCGTAGAAGGGGAGTCATTCCGTAATTCGTAAATCGTAAAAATCAGATCAAGTTGATTCAAGACCTCCTCGTGGTATCTTCTGGATGCTATTAGACTGA (SEQ ID NO: 1240)	TAGCCTATACACTATGTTGGAGAGAAGACGCTTGCTACCTAGGCAAAATGTGAAATTAGGTATAAACATCGTGGTTGTAAAACTTGAGGTGGGTTTTTAGTACGTATGCGTGATTACTTCGTAATCATGAATCGTGCATGCTAGTGGGGTTTGGCCTCCACTAGTATCTTTGAAGATTTTCCTTCCTCAGCGATCAAAAAAAA (SEQ ID NO: 1363)	SSFGLIVTNLNSETVLWGCQPLGQFSLIGTNMQNTTPRIINTNSLTNQIPTVSSLGAQSEHSAQVNPNSGYQCTICESSFRSKSGLGVHMSRRHKDEFDQLRLRTDRKAQWSEEELSMMARKEIELAANGERYLNKKLAEVFTNRSVDAIKKCRQRERYKTKIEQLKGQAVPLPEALESETIQRRPSIRERDLLVTPPNTLGTTPTELSNSEILAVLQGYPPVVCNDQWRVEVLQSIVDGAQASGKEITLQRLSTYLMEVFPSQNDRPIQTRPPRRPRNRRQGRRQQYALTQRNWDKHKGRCIKAILDGTEGTATMPSQGIMGSYWRQVMTQTSPTYSGTNTTFRTEHPLEGVWSPITLGDLRVHRVSLTKSPGPDGITPRTVRSIPSGVMLRIMNLILWCGKLPVSIRQARTIFIPKVGNASRPQDFRPITVQSVMVRILNAILASRLTSSVDWDPRQR

						GFLPTDGCADNTTIVDLILR DHHKRCKSLYIATLDISKAF DSVSHAAVSATLTAYGAPK EFVDYVQNSYEVCGTTLNG DGWRSEEFIPARGVRQGDPL SPIIFNLIIDQLLRSYPNEIGA TIGDHTTNAAAFADDIVLFA ETRLGLQTMLDTTVDFLSSV GLTLNSDKCFTVGIKGQPKQ KCTVVIPETFRIGSRSCPALK RTDEWKYLGITFTAQGRTR YSPADDLGPKLLRLTRSPLK PQQKLFALRTVLIPQLYHKL TLGSVMIGVLRKCDILVRST VRKWLGLPLDVSTAFFHAP HTYGGLGIPSVRWVAPMLR MKRLSNIKWAHLAQSEAAS SFLTDELNKARGRTLAGLNE LTSRTEIETYWANRLYMSV DGRGLREAGLFRPQHGWVC QPTRLLTGQDYRNSIKLRIN ALPSRSRTTRGRNELERQCR AGCDAPETTNHILQNCYRTH

						GRRVARHNCVVNNLKRILE EKGHTVHVEPSLQLETSVSK PDLVCIRDNHACVIDAQIITD GLFLDDVHHRKVEKYKRPE VISALRREFGVSGNVEVLSA TLNWRGIWSNQSVRRLIAK GLISSGDSNVISARVVTGGL YCFRQFMYLAGYTRDWT (SEQ ID NO: 1485)
R2	R2_DPs	.	Drosophila pseudoobscura	CAATTGGAAAGATAT GGGTCTGAATAATAG CGTAGAAGGGGAGTC ATTCCGTAATTCGTAA ATCGTAAAAATCAGA TCAAGTTGATTCAAGA CCTCCTCGTGGTATCT TCTGGATGCTATTAGA CTGA (SEQ ID NO: 1241)	TAGCCTATACACTATGT TGGAGAGAAGACGCTT GCTACCTAGGCATAAT GTGAAATTAGGTATAA ACATCGTGGTTGTAAA ACTTGAGGTGGGTTTTT AGTACGTATGCGTGATT ACTTCGTAATCATGAAT CGTGCATGCTAGTGGG GTTTGGCCTCCACTAGT ATCTTTGAAGATTTTCC TTCCTCAGCGATCAAA AAAAAAAAAAAAAAA	SSFGLIVTNLNSETVLWGCQ PLGQFSLIGTNMQNTTPRIIN TNSLTNQIPTVSSLGAQSEHS AQVNPNSGYQCTICESSFRS KSGLGVHMSRRHKDEFDQL RLRTDRKAQWSEEELSMMA RKEIELAANGERYLNKKLAE VFTNRSVDAIKKCRQRERY KTKIEQLKGQAVPLPEALES ETIQRRPSIRERDLLVTPPNT LGTTPTELSNREILAVLQGY PPVVCNDQWRVEVLQSIVD GAQASGKEITLQRLSTYLME VFPSQNDRPIQTRPPRRPRN

					AAAAAA (SEQ ID NO: 1364)	RRQGRRQQYALTQRNWDK HKGRCIKAILDGTEGTATMP SQGIMGSYWRQVMTQTSPT YSGTNTTFRTEHPLEGVWSP ITLGDLRVHRVSLTKSPGPD GITPRTVRSIPSGVMLRIMNL ILWCGKLPVSIRQARTIFIPK VGNASRPQDFRPITVQSVMV RILNAILASRLTSSVDWDPR QRGFLPTDGCADNTTIVDLI LRDHHKRCKSLYIATLDISK AFDSVSHAAVSATLTAYGA PKEFVDYVQNSYEVCGTTL NGDGWRSEEFIPARGVRQG DPLSPIIFNLIIDQLLRSYPNEI GATIGDHTTNAAAFADDIVL FAETRLGLQTMLDTTVDFLS SVGLTLNSDKCFTVGIKGQP KQKCTVVIPETFRIGSRSCPA LKRTDEWKYLGITFTAQGR TRYSPADDLGPKLLRLTRSP LKPQQKLFALRTVLIPQLYH KLTLGSVMIGVLRKCDILVR

						STVRKWLGLPLDVSTAFFH APHIYGGLGIPSVRWVAPML RMKRLSNIKWAHLAQSEAA SSFLTDELNKARGRTLAGLN ELTSRSEIETYWANRLYMSV DGRGLREAGLFRPQHGWVC QPTRLLTGQDYRNGIKLRIN ALPSRSRTTRGRNELERQCR AGCDAPETTNHILQNCYRTH GRRVARHNCVVNNLKRILE EKGHTVHVEPSLQLETSVSK PDLVCIRDNHACVIDAQIITD GLFLDDVHHRKVEKYKRPE VISALRREFGVSGNVEVLSA TLNWRGIWSNQSVRRLIAK GLISSGDSNVISARVVTGGL YCFRQFMYLAGYTRDWT (SEQ ID NO: 1486)
R2	R2_DSe	.	Drosophila sechellia	GGGATCAGGGGTAAT TGCGAGCAGAGGGGG AGTATTTTTCTGTAAT TCGTAAGTCATATCAT ATGGTGTGCGGAAGG	TAGCTAAAACGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC AATAAAGGCATAAACA TCGCAAATAATGGTAA	FERQNFSDGLVPQRKFIHIGT TNRNNEPRSNLRNLMTTRPS VDIFPEDQYEPNAAATLSRV PCTVCGRSFNSKRGLGVHM RSRHPDELDEERRRVDIKAR

				GGAATTTTACTCTGTA ACTCACAAGTCTCTCC TTTACTCAAGTCGACT CAAAACCTCCTCGTGG TGGTCCCCGGTAATGC TAAACTTGTTTAGCAG CTAA (SEQ ID NO: 1242)	TATATAAATTGGCTATG AGGATGGTTTTAGTAC GTAGGCGTTGCGGAAC TTCGGTTCAGATAGAG CAATGAATCGTGCATG CTAGGAAACTGAAGTG TTGACAGACCTAGTATC TTTCGATAGATTTCCAT ACCTCCGCGATCAAAA AAAAAAAAAAAAAAA AA (SEQ ID NO: 1365)	WSEEEKWMMARKEVELTA NGHKHINKQLAVYFANRSV EAIKKLRQRGDYKEKIEQIR RQSALVPEVANLTIRRRPSR SEQNHQVTTSETTPITPFEQS NREILRTLRGYSPVECHSKW RAQELQTIIDRAELEGKETTL QCLSLYLLGIFPAQGVRHTL TRPPRRPRNRRESRRQQYAV VQRNWDKHKGRCIKSLLNG TDESVMPSQEVMVPYWREV MTQPSPSSCSREVIQMDHSL ERVWSAITEHDLRASRISLSS SPGPDGITPKTAREVPSGIML RIMNLILWCGNLPHSIRLAR TVFIPKTVTAKRPQDFRPISV PSVLVRQLNAILATRLNSSIN WDPRQRGFLPTDGCADNAT IVDLVLRHSHKHFRSCYIAN LDVSKAFDSLSHASIYDTLR AYGAPKGFVDYVQNTYEGG GTSLNGDGWSSEEFVPARG VKQGDPLSPILFNLVMDRLL

						RNLPSEIGARVGNAITNAAA FADDLVLFAETRMGLQVLL DRTLDFLSLVGLKLNADKCF TVGIKGQPKQKCTVLEAQSF YVGSREIPSLKRTDEWKYLG INFTATGRVRCNPAEDIGPK LQRLTKAPLKPQQRMFALR TVLIPQLYHKLALGSVAIGIL RKTDKLIRYYVRRWLNLPL DIPIAFIHAPPKSGGLGIPSLR WVAPMLRLRRLSNIKWPHL TQNEVASSFLEAEKQRARD RLLAEQNELLSRPAIEKYWA NKLYLSVDGSGLREAGHWG PQHGWVNQPTRLLTGKEYI DGIRLRINALPTKSRTTRGR HELERQCRAGCDAPETTNHI MQKCYRSHGRRIARHNCVV NRIKRGLEERGCVVIVEPSL QCESGLNKPDLVALRQNHID VIDIQIVTDGHSMDDAHQRK INRYDRPDIRTELRRRFEAA GDIEFHSATLNWRGIWSGQS

						VKRLIAKGLLSKYDSHIISVQ VMRGSLGCFKQFMYLSGFS RDWT (SEQ ID NO: 1487)
R2	R2_DSi	.	Drosophila simulans	GGGATCTGGGGTAAT TGCGAGCAGAGGGGG AGTATTTTTCTGTAAT TCGTAAGTCATATCAT ATGGTGTGCGGAAGG GGAATTTTACTCTGTA ACTCACAAGTCTCTCC TTTACTCAAGTCGACT CAAAACCTCCTCGTGG TGGTCCCCGGTAATGC TAA (SEQ ID NO: 1243)	TAGCTAAAACGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC AATAAAGGCATAAACA TCGCAAAATAATGGTT ATATATAAATGGCTAT GAGGATGGTTTTAGTA CGTAGGCGTTGCGGAA CTTCGGTTCAGATAGA GCAATGAATCGTGCAT GCTAGGAAAACTGACC ACACGCAGTGTTGGCA GCCCTAGTATCTTTCGA TAGATTTCCATACCTCC GCGATCAAAAAAAAAA AAAAAAAAAAAAA (SEQ ID NO: 1366)	TCLAANLSGKNFSDGLVTQ RKFTHIGTTNTNNEPRISLHN LMTTRPSVDIFPEDQYEPNA AATLSRVPCTVCGRSFNSKR GLGVHMRSRHPDELDEERR RVDIKARWSEEEKWMMAR KEVELTANGHKHMNKQLA VYFANRSVEAIKKLRQRGD YKEKIEQIRGQSALVPEVAN LTIRRRPSRSEQNHQVTTSET TPITPFEQSNREILRTLRGYSP VECHSKWRAQELQTIIDRAE LEGKETTLQCLSLYLLGIFPA QGVRHTLTRPPRRPRNRRES RRQQYAVVQRNWDKHKGR CIKSLLNGTDESVMPSQEVM VPYWREVMTQPSPSSCSGE VIQMDHSLERVWSAITEHDL RASRISLSSSPGPDGITPKSA REVPSGIMLRIMNLILWCGN

						LPHSIRLARTVFIPKTVTAKR PQDFRPISVPSVLVRQLNAIL ATRLNSSINWDPRQRGFLPT DGCADNATIVDLVLRHSHK HFRSCYIANLDVSKAFDSLS HASIYDTLRAYGAPKGFVD YVQNTYEGGGTSLNGDGWS SEEFVPARGVKQGDPLSPILF NLVMDRLLRNLPSEIGAKV GNAITNAAAFADDLVLFAET RMGLQVLLDKTLDFLSLVG LKLNADKCFTVGIKGQPKQ KCTVLEAQSFYVGSREIPSL KRTDEWKYLGINFTATGRV RCNPAEDIGPKLQRLTKAPL KPQQRMFALRTVLIPQLYHK LALGSVAIGVLRKTDKLIRY YVRRWLNLPLDVPIAFIHAP PKSGGLGIPSLRWVAPMLRL RRLSNIKWPHLTQNEVASSF LEAEKQRARDRLLAEQNEL LSRPAIEKYWANKLYLSVD GSGLREAGHWGPQHGWVN

						QPTRLLTGKEYIDGIRLRINA LPTKSRTTRGRHELERQCRA GCDAPETTNHIMQKCYRSH GRRVARHNCVVNRIKRGLE ERGCVVIVEPSLQCESGLNK PDLVALRQDHIDVIDIQIVTD GHSMDDAHQRKINRYDRPD IRTELRRRFEAAGDIEFHSAT LNWRGIWSGQSVKRLIAKG LLSKYDSHIISVQVMRGSLG CFKQFMYLSGFSRDWT (SEQ ID NO: 1488)
R2	R2_DYa	.	Drosophila yakuba	GGGAACATGGGGTAA AGGTGAGTAGAGGGG GAGTATTTTTTATACT CTGCAACTCATAAGTC TTGCCTTTACTCAAGT CGACTCAAAACCTCCT CGTGGTGTTTCCCGGT AATGTTAAACTTGTTT AGCAGCTAA (SEQ ID NO: 1244)	TAGCTTAAAACGTTTGG TTCACATACATCTGCCT GCTGCCTTGGCACAAT ATCAAAAAGGCATAAA CATCGCACATATTGGTT ATTTACGGCTATGAGG ATGGTTTTAGTACGTAG GCGTTGCGGAACTTCG GTTCGGATAGAGCAAT GAATCGTGCATGCTAG GAACTGACCAAATAAC	FERRIFPKGLVPLTKDNHIGT TNLQNEPRIFTNDLLTTRPSV DHVPEDQYEPNAAATLSRV PCTVCDRSFNSKRGLGVHM RSRHPDELDEERRRVDIKAR WSEEEKWMMARKEVELMA NGFKHINKQLAVYFANRSV EAIKKLRQRGDYKEKIEQIR GQSALAPEVANLTIRRRPSR SEQDHQVPTSEASPITPLEQS NREILRTLRGYSPVVCPSKW

					AGCAGCCCTAGTATCTT TCGAAGATTTCCATACC TTTGCGATCAAAAAAA AAAAAAAAAAAA (SEQ ID NO: 1367)	RAQELQTIIDRAEFEGKETTL QCLSLYLQGIFPVQGVRHTL TRPPRRPRNRRESRRQQYAV IQRNWDKHKGRCIKSLLNG TDESVMPSREFMEPYWREV MTQPSPSSCNGEVIRTDHSL ETVWSAITEQDLRASRVSLS SSPGPDGVTPKTAREVPSGI MLRIMNLILWCGNLPHSIRL ARTIFIPKTVTAKRPQDFRPI SVPSVLVRQLNAILATRLTSS IDWDPRQRGFSPTDGCADN ATIVDLVLRHSHKYFKSCYI ANLDVSKAFDSLSHAAIYGT LRAYGAPKGFVDYVQKTYE GGGISLNGEGWCSEEFVPAR GVKQGDPLSPILFNLVIDRLL RALPSEIGTKVGNAMINAAA FADDLVLFAETRMGLQTLL DKTVDFLSTVGLKLNADKC FTVGIKGQPKQKCTVLEAQS FCVGSREIPTLKRTDEWKYL GIHFTASGRVRCNPAEDIGP

						KLQRLSEAPLKPQQRLFALR TVLIPQLYHKLSLGSVTIGVL RKTDKLIRFYVRRWLNLPSD VPIAFVHAPPKCGGLGIPSLR WVAPMLRLRRLSNIKWPHL VQSEEASSFIEAEKQRARGR LIAEQNELLSRPAIEKYWAN RLYLSVDGGGLREAGHYGP QHGWVSQPTRLLTGKEYLD GIRLRINALPTKSRTTRGRHE LERQCRAGCDAPETTNHIM QKCYRSHGRRVARHNCVV NRIKRGLEERGCVVIAEPSL QCESGLNKPDLVVLRQNHID VIDVQVVTDGHSMDEAHQR KINRYDRPDIRTELRRRFEA AGDIEFHSATLNWRGIWSG QSVKRLIAKGLLSKYDSHIIS VQVMRGSLGCFRQFMYLSG FSRDWT (SEQ ID NO: 1489)
R2	R2_KF	GU949558	Kalotermes flavicollis	GAGAGGTCACTGTTG CTGATCAGCTGGACA CCTAGCTGACTGCGTG	TAATGTCCCTTTTGGCT TGCCCCCACCTGCTTAA AGGAACTGGCAGGAAA	MEVPLTSSLGSGATQAPGTP ELLGEHTVERPGLDQGHSY GLLMDDVELPVRLPFFGPLI

				CATGCGTGCACGCCG CGCGCCGCCTCCTCGT GGTACCGCTGGTAAC GCAGGGTCACTGAGA CTACCTGCGGCTAAA GGCGCGCGCGAGGGA TTGTAAGTCCACCACC TCCACGTGGTTCCCCG CGGGCAACGGCATAG TTCATCTGCCGGCCAA GCGTGCCGTTCCCTCG ATCCTCCTCTGATTAG GGATAGGAGGGGGGC GGTGGCTCCCGCCGA CGACCTGGCAAACAC CTTGGACACGCTAAG ATAATAGGCCGTCCTC CGGGCCGGCCGAATC ATAAGCACACA (SEQ ID NO: 1245)	GAGAGTGATCCGTGCC ATAGAAATATGGTTAT CCGGGGCAAGTCACTA GCAATATGGGACTTCTC CGGGTCCGTGCGGTCCT TCCAACATGAGCTGGA CGTAGTCCACTCTATGA CTTGAACGATACGGGG GCGTATCTCCCCCGGA AGAGGTCGCCTAGGCG ACTTAGAA (SEQ ID NO: 1368)	CPGCRTLLTSEETISSHHRRV HPDARTRWVCYGCDSPFMT YRAIKCHLPKCSGRKVVTG DHICNGCTKRFESQRGLSLH KRRAHPGLRNEEMLEPPVR AERRPNAHKSSIWSIDEIRIL EQYEAAYVGDLHINMKIAA HLPFKTNKQVSNYRNDRRK KSRTATDASQQGLGPNDGN RGIVPSGQSSPLFLEGSDAEG DEDVFNVLVPPTLGGLEPAG QVHSLSEGETSPLVGEADPC FMGGTPSAGEASGSTLLGPD PTPADGYSLVRKDLQLSVQT SPLLAVGSVGTESVQFERGV LSCGTPPEFLHPEQFAHCAN NDPVLNASEEQVHAPLGEE ANDLPDNNHPSELGVDPEDP TCSPATEQVQPSSEEEADDP FAQFKAWRRRVASYALKIE TGVLPAQVDDLLRRLRDGD TQSKVTCEEVEEVVLSLTRT ILGGTAPKKRVEGRTKWTY

						KSRTNHEARKRIMYARCQD LYRRRPQRPVERAVGYQAE ESLLDNQDERPSHGAFETFY TGLWGKSGQCNITMPPGVP RHTGHVLREVTPKDIYSRLR KLKKDYAPGPDGVTKLKVQ SMGAYPSLLAKVYNLVMLT GYFSSCWKEHKTSLIPKDRG SPMDVSNWRPITIGSLLSRIY TGLIERRLRTVSDIHQRQVG FMPVNGCAANLFIFDECIRQ AKKEGTIVGSLIDVAKAFDT VPHEAILRALSSQGVDEHTM AHIRDMYSGIRTRINGKGSDI PLVRGVKQGDPLSPMLFNM VMDPLIRDLQRKGFRIGGHE IGALAFADDIVLLADSIDGA QDHVDQVGRYMNKLGMTL NPRKSSSFLITAMRKTWICR DPGLSIGETKVPGARPSSAL KYLGVNYTLSEGLESGALID KLMQAVNRARGLALKPLQK VNLILERIIPKFLYGIILGGPS

						LTRLHAADKCVRMAVKEIL HLHPSTTDHVLYARKKDGG MGIPRLAHLVRLASLRSGLA LLASGDVAVQAAGMAGDL EGRCKKVANDLRLNWPVTL RDVVRASNKFKSQESKDWE RLASQGHGVKDFRNDRLGN CWLYDPTVLSSSRYTDALRL RTNTFGVNVALRRADKDLE VNCRRCHGKPETLGHVLGE CVAGKGMRIQRHDKMAAF VATKCEEKGYQTTREQLFSI EQGKLKPDLVVIDGERALIV DVTVRFESGNALSRGASEKI EKYQPLADYFVSQGAVREA NVLPIVVGSRGAITQATLKS LATLGLDVERVGKYLAICA VASSVEIACMHLDYT (SEQ ID NO: 1490)
R2	R2_RL	GU949555	Reticulitermes lucifugus	CTCACTGCTGTCATGC CATTGTTAATGCGTTG GGTGATGGCGGGTGA TGGGAGACGAGTTAC	TGAGTTAGATATATATG ACGACTGATTCTTCCCT AACTAAAATATAATTT GGAATGTGCATTACTTA	MMADYNNSVDHALEDNTR LIFARDAVLARVCGPFDNLE CGLCGVLLTSLQGVREHCH RAHHNLDLTFQCTKCDKGF

				AGCAGAGCTGGCTTC	TACTTTATTGTATTTAT	SSYRGICCHFSKCIGARISVS
				ACGGGGGGGGCTGTA	TGACATCCGTAGTAGTC	EGPLSCSECEREFDSKRALST
				GTTGCCCGAACCAGTC	TGTTTAGTGGATTTAGA	HERHMHPGIRNAKRLKDFN
				TGCTTCTGAGTGCCTC	TCGTTAACCATTCTGTG	PRGGGGKTIHGNTKWTEEE
				CACGTGGCCTCGCTGG	ACGTCGCTGATGTCTAT	VQLLVSLSKRFEGYKSINKEI
				AAACGTCGGAGCTGC	ATGTACCAGTGGCAGA	SLILTSKTCKQISDKRRYLNL
				TTACGGCTAACCCGTA	ATGCTGACTAAGAACA	HNGNGGLAAAEAVLEFCED
				GACAACTCCGGCGGC	AAATAATTTTAAACAA	SHPEVTESDGAVLSEIMDEE
				CAAACTCAGAAAACG	TAGACAGCACTAAGAA	CHQSSVTMRSSIVHGDIGRE
				GCCTTACTACCAGGGT	ATTCTGCAAACTGGAA	VQGKELVRIPPDNSVMGNC
				CATCCCTCGCGGCGGC	CGTGGCCCACGGTAGG	VVLLRKLATDKRSDLDLSK
				GCTAAGACCTTACTGT	CCAATATACCGGAAAG	DKELRIDIEKATKANRESAD
				ATGTACTACGGACTAC	GGAAAATGACACATCC	GRVIQSVAGNEIDPDTFQWK
				AGTTGAGTAGCGCGG	CCCCTTAAA (SEQ ID	ELLLGQVRGFPRVDENSELF
				GAGAGCACTTGATGT	NO: 1369)	DLDDKLTKELSSDSPVWND
				GAGGGTAGCAACTTG		NCELIVSDLCQVLCKKKYEL
				TTGCTACTTCGACTGT		GRQHHVRKGKRHRGIHHKR
				CTCCCTGAAGATTTCC		EKFRECQKIFRKSPRKLAEY
				GAAGGGGTGCCGCAA		LYRDKDLSHISKDASTPQGI
				GTCCATGGGGGCGGT		EQYYSQLWGEPELLESNTIE
				CCTGCTGTGGGAGTTA		EKLPSSSLFDCLPPITPEEVE
				TCTTGCACCCGGGAA		GRIHKIRPSSAPGLDGVRKIH
				GCCCTCAGTGGATAA		LVGKGITLVLVKLYNLLFLT

				ATACTCCAAAATGGCT		GGYPECWKRNRTVFIPKIGK
				TAGCCACCTCACGTGG		DLSEVGGWRPLTIGSLLARM
				AGAGAGGGGAATTAG		YSAFLERRIRRVTSLSLSQRG
				GGGTTTACTCCTAAGG		FTNIQGCHVNLTILKEGIRQA
				GCCGATGCCTCCAATC		KVKNGGVIVSVDIEKAFDTI
				GGATTTCCCCGGGCA		PHSVIFSRLASQGVPPLLRKII
				ACGGCTATTATCAGCC		SNMYKDVYTVIEGQCIPIKR
				GGCTAAGTTCGGACC		GVKQGDPLSPLLFNIAIDPVL
				CTCATCTACTGATGGG		RSLEEFQGGLPLGNSAIKILA
				ATACTCTCTCTCCCCC		FADDIILGASSAGQAQQMV
				CGCCTGCGTGATTAGT		DMLGIGLTSCGLGVSHRKCF
				TGGGAATGCACGCGG		GFQIVNKNKTWTIVDPMITL
				GCGGGTGATTCTGAC		NGSSLPFSGPEDRLPYLGVDI
				AAGCCCAGCCCAAAA		NPWDRKSRYDAGQRLISAA
				GTTCTTACCACGACGT		KRGSQLSLKPQQKINLITAFL
				ACCCGGGGGTGTCAC		LPKFLYILIEDPPSPAYLKSID
				CCGAACTAATATATTT		HDLRQIYKNILHLPNCVSTA
				TCCGCCGATCGGTTCG		FMYSPKRDGGLGLPRLSCLV
				CTGTGGGGGGATGTC		PLAHLKAGIKLGSLQDSLVR
				GGGGCCTGGTGGCCG		EITTSDRFVRTMGSVAHSLW
				GGAACCGTCAGCCGC		ASWSLTLQDIYKLKSALKRR
				TCTTTGAGTGTCCTGC		EAKAWESCVSQGQGAAQFR
				GTAACCCGGAGGCGG		GDSIGNNWLHNPGTYRPGQ

				TGACGTCGAAGGGCT AGGCTAGCACGGTCC AGACCCCTGAACGCC TGGGGAGACGGGCCC ACCACCTGAGTAGGA GTGGGTCCTTTACCTC ATTTGAGGTGTCTCCT CGTTCTTGTAGGGGGC GAAGGACGAGGAATG CATCCGTCCCTCCAGG ATGCTGGTGGTTTCCG TCTGGTGGGCTCATGC TGAAACGCGCAGGTG CCGATTTGTCGAAGA GGACATATGGGGTAA CCCATAGAACCTAGG CAGGAGTAATCCCTTA GCTGGGGGGGTCAGC TGGTGGGTCCTGTCAA TTTATCCTCCCCTCCA TGCCAGAGCCGGTCC GAGGTTAGAGACGGA CTCATTTTCTCCTTTTT		YIEALKLRANLTGVRVNLK RSGYNVPITCRFCKDIPETQA HVLGLCPKTKGMRIQRHDSI VNRVRDKLKTKSPVALMHE QNFTVEEGQVFKPDIVTILG EVGYVIDVTVRYDDRDYIK DASVEKIRKYEALKGYLKD LYPQLNKVEVLPLVFGSRGA VPGSTVHNMGLLGFTKREM VHISRKVIADSLIISNFLEVY (SEQ ID NO: 1491)

				ATATGTC (SEQ ID NO: 1246)		
R2	R2_RU	GU949554	Reticulitermes urbis	CACACTGCTGTCATGC CATTGTTAATTCGTTG GGTGATGGCGGGTGA TGGGAGACGAGTTAC AGCAGAGCTGGCTTC ACGGGGGGGCTGTAG TGGCCCGAACCAGTCT GCTTCTGAGTGCCCCC AAGTGGCCTCGCTGG AAACGTCGGAGCTGC TTACGGTTAACCCGTA GACAACTCCGGCGGC CAAACTCAGAAAACG GCCTTACTACCAGGGC CATCCCTTGCGGTTGC GCTAAGACCTTACTGT ATGTACTACGGACTAC AGTTGAGTAGCGCGG GAGAGCACTCGATGT GAGGGTAGCAACTTG TTGCTTCTTCGACTGT	TGAGTTAGATATATATA ACGACTGATTCTTCCCT AACTAAAATATAATTT GGAATGTGCATTACTTA TACTTTATTGTATTTAT TGACATCCGTAGTAGTC TGTTTAGTGGATTTAGA TCGTTAACCATTCTGTG ACGTCGCTGATGTCTAC ATGTACCAGTGGCAGA ATGCTGACTAAGAACA AAATAATTTTAAACAA TAGACAGCACTAAGAA ATTCTGCAAACTGGAA CGTGGCCCACGGTAGG CCAATATACCGGAAAG GGAAAATGACAAATCC CCCCTTAAA (SEQ ID NO: 1370)	MMADYNNSVDHALEDDTR FIFARDSVLARVCGHFDNLK CELCGVLLTSLQGVREHCH RSHHNLDLTFQCTKCDKGFS SYRGICCHFSKCRGARISVSE GPLSCSECERKFDSKRALST HERHMHPGIRNAKRLKDFN PRGGGKTIHGNTKWTEEEV QLLVSLSKRFEGYKSINKEIS LILTSKTCKQISDKRRYLNL HNGNGGLAAAEAVLVFCDD SHLEVTDSDGAVLSEIMDEE YYQSSLTMRSSIVHGDIGRE VQGKDLVRIPPDNSVMGNC VVLLRKLATEKRSDLDLSK DKELRIDIEKATKANRESAD GRVIQSVADNEIDPDTFQWK ELLLGQVRGFPRVDENSELF DLDDKLTSELSSDSPVWND NCELIVSDLCHVLCKNKYEL GRQHHVRKGKRHRGIHHKR

				CTCCCTGAAGAGTTCC		EKFRECQKIFRKSPRKLAEY
				GAAGAGGTGCCGCAA		LYRDKDLSHISKDVSTPQGI
				GTCCATGGGGGCGAC		EQYYSQLWGKPELLESNTTE
				CCGGCTGTGGGAGTT		EMLPSSSLFDCLPPITPEEVE
				ATCCTGCACCCGGGA		GRIHKIRPSSAPGLDGVGKIH
				AGCCCTCAGTGGATA		LVGKGITLVLAKLYNLLFLT
				AATACTCCAAAATGG		GGYPECWKRNRTVFIPKIGK
				CTTAGCCACCTCACGT		DLSEVGGWRPLTIGSLLARM
				GGAGAGAGGGGAATT		YSAFLERRIRRVTSLSSSQRG
				AGGGGTTTACTCCTAA		FTNIQGCHVNLTILKEGIRQA
				GGGCCGATGCCTCCA		KVKNGGVIVSVDIEKAFDTI
				ATCGGATTTCCCAGGG		PHSVIFSRLASQGVPPLLRKII
				CAACGGCTATTATCAG		SNMYKDVYTVIEGQCIPIKR
				CCGGCTAAGTTCGGA		GVKQGDPLFPLLFNIAIDPVL
				CCCTCATCTACCGATG		RSLEEFQGGLPLGNSAIKILA
				GGATACTCTCTCTCCC		FDDDIILGASSAGQAQQMV
				CCCGCCTGCGTGATTA		DMLGIGLTSCGLGVSHRKCF
				ATTGGGAATGCACGA		GFQIVNKNKTWAIVDPMITL
				GGGCGGGTGATTCTG		NGSSLPFSGPEDRLPYLGVD
				ACAAGCCCAGCCCAA		TNPWDRKSRYDAGQRLISA
				AAGTTCTTACCACGAC		AKRGSQLSLKPQQKINLITTF
				GTACCCGGGGGTGTC		LLPKFLYILIEDPPSPAYLKSI
				ACCCGAACTAATATAT		DHDLRQIYKNILHLPNCVST

				TTTCCACCGATCGGTT CGCTGTGGGGGGATG TCGGGGCCTGGTGGC CAGGAACCGTCAGCC GCTCTTTGAGTGTCCT GCGTAACCCGGAGGC GGTGACGTCGAAGGG CTAGGCTAGCACGGT CCAGACCCCTGAACG CCTGGGGAGACGGGC CCACCACCTGTGTAGG AGTGGGTCCTTTACCT CATTTGAGGTGTCTCC TCGGTCTAGTAGGGG GCGAAGGACGAGGGA TGCATCCGTCCCTCCA GGATGCTGGTGGTTTC CGTCTGGTGGGCTCAT GCTGAAACCCGCAGG TGCCGATTTGTCGAAG AGGACATATGGGGTA ACCCATAGAACCTAG GCAGGAGTAATCCCTT		AFMYSPKRDGGLGLPRLSCL VPLAHIKAGIKLGSLQDSLV REITTSDRFVRTMGSVSHSIG ASWPLTLQDIYKLKSALKRR EAKAWESCVSQGQGAAQFR GDSIGNNWLHNPGTFRPGQ YIEALKLRANSTGVRVNLKR SGYNVPITCRFCKDIPETQA HVLGLCPKTKGMRILRHDSI VNRVRDKLKTKSPVALMHE QNFTVEEGQVFKPDIVTILG EVGYVIDVTVRYEDRDYIK DASVEKIRKYEALKGYLKD LYPQLNKVEVLPLVFGSRGA VPGSTVHNMGLLGFTKREM VHISRKVITDSLIIISNFLEVY (SEQ ID NO: 1492)

				AGCTGGGGGGGTCAG CTGGTGGGTCCTGTCA ATTTATCCTCCCCTCC ATGCCAGAGCCGGTC CGAGGTTAGAGACGG ACTCATTTTCTCCTTTT TATATGTC (SEQ ID NO: 1247)		
R2	RaR2	FJ461304	Rhynchosciara americana	CAGAACGTGGAGAAA CGGAATAACTACCCA GATCCGTTGGTTAACC GGTGGCAAAGTTAAT CAAGGTTGCCATAGG CTTAATAACCCTATGG AAATGTTTCCACACAC CTCCACGTGGTGCCTG CCGGAAATTGTTCTAG GGTGAACAAGCTAAG TTGTGAGAAACGGGC TCCACCACAATATGG AGCCTGCCAGGGCGC GAGACTCAGGACTCT CCATGTACAAAGTGG	TAGAAATGTGTGCGAT AAGGTGTGAATAGAAG GGTTCACCAAGAGGGA GACCTAGTTTGGACCTC AGAAATGGGGTCATAG GAGTGATAGGTTGTAA AGCCGTTGGGGAATCC GGCTACACATGGTATCT CAGGAGCCATTCATGC GCTGATCTCATTAAGGC GTAATAAACTGTGAAA CAGATCCTGATAATGC CGTGCTACCAAATGAT GTAACGAGGCGGAAAT AAAATTAATCTGGGGC	MSNYNETNTSGGDNPRMAT QTTGSLSSGPINQHTCELCC RTFGTRAGLGQHVRKTHPIE SNQSINVERKKRRWSPEEIR RMANMEAQATINNIKHLTQ YLATYLPQRTLNAIKGRRRD AEYKELVTGIIANLRSNSSTQ QTNQVCNESEMSQRSKILQS IRESVRDLRSRRNKYAKALQ ELGEAALCGKMLNEEQLIH CIKSMFNTAKCPKGPRFRKT ATHSGTNKQQRQQRYARVQ KLYKMNRKVAAKMVLEET DKIQIKLPDHDPMFKFWESE FKEGEGMPERMPKDLKESP

				TTAGTTGCAAAAAGA GTGCGCCTAGCATGA CTGATAATTTTTCACT GAAATAACGTTGAAC TTTATCTGTGTCATGT GCACAACACTATGGT GTCTGATCAAGCACC ATCAGTGGTGGACCT GCTAATGTATTAGTAG AACGTGTCCAGGCGA TAATGCACACACGGC CTCCGGGCCCATCGCT TTTTTTGAGATTCCCT AGAAACTTCAGTGTGT GCGACAACTGTATAA CCCATAAGGGATGGA CAAAGGTTATACTAG GGGGTAAAAACCCTA ATCGGCTAATGGCAA ATGGGATGTAGAAAT GCCAAAGATTACTCG CACCGAATAATGGTG GCCGAAAAGCGGGTA	GTTCTGCGGAATGACT ACTAAATATAGCGATG CTATATATACAAACGA CTGATGGTAACACCGG CCTTA (SEQ ID NO: 1371)	DLKAIWDPVTEEEVRKAKV ANNTAAGPDGIQPKSWNRIS LKYKTLIYNLLLYYEKVPHK LKVSRTVFIPKKKDGSSDPG EFRPLTICSVVLRGFNKILVQ RLVSLYKYDERQTAYLPIDG VGTNIHVLAAILNDSNTKLS ELHVALLDITKAFNRLHHTS IIKSLVGKGFPYGFITFIRRM YTGLQTMMQFEGHCKMTQ VNRGVYQGDPLSGPIFLLAI EKGLQALDKEVGYDIGDVR VNAGAYADDTDLVAGTRL GLQDNINRFSSTIKQVGLEV NPRKSMTLSLVPSGKEKKM KVETGKPFRANDVPLKELSI NDFWRYLGISYTNEGPERLS LTIEQDLERLTKAPLKPQQRI HMLNAYVIPKYQDKLVLSK TTAKGLKRTDRQIRQYVRR WLKLPHDVPIAYLHAPVKS GGLNIPCLQYWIPLLRVNRV NKITESQRSVLAAVGKTALL

				ATCGAATGAAATGGT AATGCTGTAGCGGAA ACATGATCACATTCTG TGACAGTAAACCATT AGACCTAGGGGGAAC TATGATTAACAAGAT ACCAGCTTACATGGA AGCAATGAAATGTAA GTCACAGTAGTGATA AGTGGTGAAGAGTCT TGTAATCACCGTAACT AGGCCAGGTTCTGGG GATGCCATGAACTTA GGGGGAGTATGGTTA GCAGATCTACCAGCT AACACTATTACTGATA ATATGTAAGCCGCAG TAGCGCTAAGTGGTGT ACAGATTTGCAACCA CCGTAACTAAGTTCTG TTTCGATGGACTAGGG GGAATCATGATTAAC AAGATACCAACTTAC		TSTVYKCNQSLATLGGNPT MLAYRTYWEKELYAKVDG KDLQNARDDKASTRWNGM LHSDISGEDYLNYHKLRTNS VPTKVRTARGRPQKETSCR GGCKSTETLQHVVQQCHRT HGGRTLRHDRIVGLLQHELR RDYNVLAKQELKTGIGLRK PDLVLIKDDTAHIVDVQVAR CSKLNESHVRKRSKYDKKEI EVEVKSRYRVSKVMYEACT ISYKGIWDKQSVMSMRRLG VSEYCLFKIVTSTLRGTWLC WKRFNMITSVRS (SEQ ID NO: 1493)

				ATGGAAGTTATGAAA TGTAAGTCACAGTAGT GATAAGTGGTGAAGA GTCTTGTAATCACCGT AACTAGGCCAGATGT AGTCAAAGCACATGT TTAGGGGGAACAAGG TTAACATTGGTAAAA GACCAATGCAACCTC CGTAACTAAACGTGA AGGCACAAAACTAAA AGTCCAGTGGATGAC AGGTGAGGTCATCAC TGGACCCAAATGTTTT AAGCTCATCATAACA ACACGGTGAAAAATC CAGCATTTATTTGCCT GATTGAGTAGCTTCCA CACTATTCCAAAGCCG AACCTATCTGGGTTTT TCTTGAAAGGCCGTAT AGGGCATATGTCGAG AAATAAGTCCAAGGT		

				GAGGTAGTGTGGCCC TGTACCCAGGGGTAA GGTACTATACGGTCG AGTGGCTCAGTAGGC CTAACTAGCCACTGA GTCACTATAATGACTA GTT (SEQ ID NO: 1248)		
R2	YURE-2_Cis	.	Ciona savignyi	CCAAAATTACTTCCAG CACCTCCACAGCAGA CGAACGAAGAAAGAA GACTTACGAAATAAG AATAAGCAGTTTAAA GACGAAGCAGACGAA CACCACTCCACCAAC GACGCTCTGCAGCTAC ACCACCACCATCTCGC CGCAACAAGAAGAAT TTTCCGCTGCTCTCCA CTCTGCTCCAACACCA CCTCTCCTGCTCTGTG GACTGCTGCCTTGCTG CTGGACCAACCTCTAC CCGAAGGAACCCTTC	TAAGATCCGCGGCTGT GGCGCCGAACGAGCAC CTGCCCATTCTTCTTGT AGGGACTTTTTCACCCT CACTCCCCCCAATAGTT TTTTTTCGTTTTTTCGTT TTTTCACCCCCACCCCA CTCGCCTCTGGGCTGCA CATCCCACACGTAGGG ACCTGTTTATATTATTT GCCTTTTATATGTACCA CTTTTTAAATATATTTT TGTACCCCACAAGATG CTTTTCGCCAAAAAAA AAAAATTTTTGTATCAC ATTTTTATATTTTGTAA	MAGHKITMSEGKLLEVAVR YGGVRNVSYECPVPDCTKT FSQANNLIRHLNNFGNTKHR AHNFTYFFTCEKCKIQIHSNT KHNISNHYKQCCATGGGPS CETGQYFCPACEQAGLGNL ESALRHFQSSHPEFNLPPRSQ FSKSHPNSYTLSLKPKDHLM KILYSGPLTPGQLVCPIKICL RSSAARLFHDVSKLRKHML VDHNRTLVYETTCGKCLRP VDTSKNMRKTTSHFEKCSG ESFISSPSPIPQKTYKLDLPST STPPPRKSPKLQPYKPIRTFK NPLTKSSQSKSDNPPKPTPFF SPRTLERSASWPALSEVVDP

				GAACCCAGCAGCAAG	AACACAGATTTTTATAA	LPKLKEKHPSLPCALDKCPP
				GTACGTGTCACCACCT	ACTTTGCACTATTTTTA	SPRIKPSTLVPPCHTANNSPK
				CTCCACAGAGCCAAG	TATAAACTTCGCACTTA	PTSPESPSTLKPLPRPIRPSKP
				GCCAGAGTGGATAGA	TTTAAAATGAATCGCAT	LEDWLTVRSVGPDREIVLNI
				GCAGCGGCCTCTACA	CTTTTTTATATACACCA	GPRPRPGPAAGSRTTSPPST
				ACCAAGTTCTCCACTC	ACACAAACAGGATGTG	APAKRVAANPIAAPLSGEPG
				CGACGACAAAACACC	CAGCTCAGGGGAACCA	ATLDCGQTGRKVQPPKKRP
				TGCCTCGTGGCCAGA	ATCCTGCGTCCCTCCTA	TESAGSLPPPAEPATDLLTG
				GTCCTGCCGAAGAAG	GCGGCGGGAGGGGCGC	REGLARLVEEYHLSGDFGAF
				AATCTTCGACCCCAAC	CCACCTACCCCCACGCT	CRDLERWTALSSTNRRPKPR
				ACCGTCTCGGCAGGG	CCTCTTGAGCACCAAC	RGRYNRGAAARATRNRGR
				CTCCAAGCAACGACA	AGGGACTCCCTTCCGG	DDRQDPQDRDDQGGPGPVT
				TCAGCCACGAGTGCC	AGCCCCTGCACCCTCA	CGRPQRYKRAAALRSAFGR
				CACCGGAGTGGACGC	ACTTTTCTTATTTTTAA	DMKATVRRIIDGERGDARC
				GGCGAACCCGAGGAC	AAAAAAAAAATCATAT	EIDPKTIEGRFRDELSPPVRE
				CGCTGTCGCCAAGAA	ATATTGATCTTGACGAC	GPECSLPPWMAEAQAGEHA
				CAATAACATCACCGC	GGGGGCTACATTCAGC	PSNDSQPGDAYDGPITALEV
				CTGCAGCAGCTACCA	CCCCAAAAACCCACCC	EMVLSTLNVGSAPGSDGLS
				AACAAAGGTTAGTCT	ACCATCCCCAACGAGT	YGFWRALDPKGLVLSELFE
				CCTACCTCACCTACAA	GCCGGGGCATTGAAGA	VCRIERRVPGPWKSSRVTLI
				ACTCGTCTGAATAGAC	GCTCCGGCACAATTAG	CKDAEGDLDDLGNWRPISIC
				GCCCCCGCGTGGGAG	CACTTAGCTTATTTATT	QTVYKIYAAVLARRLQSWA
				CTAACCTTGGTAGCTA	ATTTTTTGTCAACATTT	LDGGVISRSQKGFMPFEGVY

				GCTGCAGTGCCCTGCG	TTGTTTTTTCAAATTTTT	EHVFLLDSVVADARATRRS
				ACAGCGGCCTCGAAG	TCACCCCTCACCCCCAC	LAVCWLDLRNAFGSVDHTT
				AGCTGCAGAGCGTCA	CCTAATAGGTCCCTCGG	IVEALSRFGAPAGLVEMISDI
				GCCTGCTTCGACCTCG	GCTTGGGCCCCTTTTTC	YTGGSCRIRTRAGFTPDIPVG
				CTTGTTCTCATTTCAA	GTGCTCGAGAAGCGTC	RGVRQGCPLSGIIFNLVMEV
				CCTACTCCGTCCTGTG	ACATCGCCCCACTGAC	LLRGVEANNACGYRLSCAG
				GAATCAAAAGAGCCC	CACGACCTTCCCCGAC	GASVRVLAYADDVALVGSS
				CACTAACAATTACATT	ATTGGAGTCCTTGGCGT	RAEXKIQLGVCERFAAWAG
				CATAAAATCTAGCAA	CTCCCAGGTCGAAACA	FSFNNKKCAAMVLKHQRGG
				GAACGAAGAGAAGCG	GTCCCAAGTGATAGCA	RRLLDSAPLRLCGEEVAILG
				ACCAACTTTTAATCCA	CCTAATGCTCGACTTGT	PDSFYKYLGAHTGYGRQTG
				TAACTTTTAGATCTTT	TTCGGCCTGGGCCGCC	GQLVDRVERQVVRLFTSFLT
				TTATTTATTACTGTTTT	GAGGATTCCCAGAACG	PTQKLSALKRIVLPAMSFHL
				TAAGCCCTAAGCATAT	ACCATTCTTCTAAATAA	RVRPCAEGHLRRLDNTVRR
				TGCCTTTTTTTAGATC	TATTTATATTTCAGAAT	CVKTALRLPKGSCRAFFHTS
				TTATAATTATAAAAAT	AAAACTATATATATATC	PDAGGLGITSVVAECDILTV
				AGATTCAAAGTTAAC	GTTGGCGGGACTTGTCC	TQAFKMLSSPDHLVSLVAK
				ACCACCAGGCCGCTA	CGCCTCGATACCGAGT	GRLGMHAARMGRSETASAC
				CAGAGCATTTTATTTA	GCTGCAGAGCGGCAAA	AMADYLSGDSVMGHXSWK
				ATCAATTTGTTACCGA	ATAAAGAAGAAACCGA	TGYRMPADLWTATRAASRR
				CCTCCTGCTGCTTCTT	CGTCGCTCTGCAGCCA	LSLRFSPQPQGEFGLESGTFK
				TTTACTTTCTCCAGAC	AGGACCACCCAAACTC	IAPRERRSLTRRLHHRQNLW
				TACTACCCGGATACA	AAGCCAGCACCGTCGA	WRNQWAALPNQGKTVAAH

				ACCCTTGGAAACGAG AGGA (SEQ ID NO: 1249)	CAACCAACATCCTCAA GTCGGCGGTTGCTGGA ACAACTCATAACATCTT CAANATAAATTATCAC CCTGTGCAGCAGGAGG CCGTGCTTTTAAAACTA CTCTGTAGTGGCTCATG ATAATATTTCGCTCCTT TTTTGCCCCGTGTAAAC TTAGTNGATGCGAATA AAATCAGTTGAATCA (SEQ ID NO: 1372)	SAYAASNNWVKGPSSLAPQ ALFFGLKARLNQMPTRSVK ACYSRAPNYDKSCRRCGAE VETLPHVLNHCPKSMKSILE RHDSVLAEVLAAIPRGTFAS VDVDRTSREHFRRVGEALR PDIVARRHDGSVVVADVTC PFESCASALDTAAARKIEKY DQLCANLRQLYRKPVESHA LVVGSLGSWGRTNNTALAA LGIRGAVRSRLAKQLVNLSV EGSHNIWLRWSGGIPKDLVR (SEQ ID NO: 1494)
R4	Dong	.	Bombyx mori	GCTAGCTCCCTAAAAT CCTACCTTACGTCCGA GGCGAACATCTGTCC ACGTGGGGAGCGGAA ACGCGTACTATCGAA ACTTACGCGGCTAAC AAGGTAAAGGTAACC CATTAATATGGAGAC AAGACTAAAAAGAAA ATTGAGAGGGCCGCT	TAAAAACTAGCATAAT TATTAACTCATAACTAA TGTATATTACTTGGCCA AAAGCCCGTATATACA GTTCCACCGGCTCTGTC GACAGACTGAACTGAG AAAGGGGAAACATATG GAAATAATAATAA (SEQ ID NO: 1373)	MLRRGRIFLPASTKAGKTRG RMKWSREVNLFIMRTYYYV TKLETDLTIYRKKLHEHFSL KYPNVIISQQRISDQKRAIER NKLLSQETLDRLKEEVRKQL EDEQTNNVENEKLNSETYS HEYTTLTPQTILTKKTQQHT NIISSTQTSHSSTQTESITLLL ENEVDILNTNPTEGATQTQE VKDKFETNLTMYSGMDPKA

				TCCCGGGGGCGATCG CCGGGGCACACCTGG AGCTGGCGCTGGGTG TTCCAGCATAGGATCC GTCAGCGGCGGAGAG TTGAGCAGGCGGGCT CTCGCCGGGGATTTAC AACCCGAAAATGCTA CAGCCGCCAACAGAA ATCACGATGTAGAAA GTAGGAGCAATAGCC CATGTGAACCTTACAG CCCGAGTACCGGTTCA TACAACCCCTCGGTAC AATCATCACCATCATC CTCGGGTCATAGAGG CTCACCAACGTCGACT ATGG (SEQ ID NO: 1250)		RPPLPKLKYSSKLNELIRLFN NDILVDYISPDTQLSDVHTL TYCTAVTISEQLKYKIIAIEG NARHKKNFKPPWQQRLEKD IAKLRADIGKLTQYINNNRS KKVVQSVEQIFKNTKIHTSH ENGNKKSQEFLDTLKQKLA LKAHRLKRYNNSQKRKNEN TIFLTNEKLFYRNLIKPKTDR DNSNIDIPTAEQLEMYWARL WENSAKHNDKANWITEEKE RWDTIEEMQFDDVTEEEITTI TARLHNWKSPGIDKIHNFWF KKLICLHKTIAKNLTDIISGN QSIPEFIATGITYMIPKGDFSI EASQYRPITCLPTIYKILTTVI TKKINSHIEHNNILAEEQKG CRRGHMGCKEQLIIDSTIMK HATTKNRNLHCTYIDYKKA FDSIPHSWLIQVLEIYKINPIII SFLRNIMTHWQTTLKLKNPP NFVTTRQIAIKKGIYQGDSLS PLWFCLALNPLSHQLHNDR

						AGYRIKQQDNTETIISHLIYM DDIKLYAKNDKEMKKLIDT TTIFSNDISMQFGLDKCKTV HIIKGKVQPGDYTIDDTNTIT AMEPSDLYKYLGFQQLKGL DHITIKQSLTSEYKKRINAIC KTKLSGKHLIKALNTYAIPIL TYSFGIIKWSKTDIEQIERITR TTLTKHNNLHPKSAIERLTIK RQDGGRGMIDIWHLWRKQI HSLKTFFYIKSDLSEIHRAIA QNDNNYTPLNLKQKELIDN TENLRNRNPQKDMEENWK KKALHGRHPHDLSQSHIDSK ASNMWLKTGSLFPETEGFLI AIQDQVINTKNYRKYIIKDPT IRDDKCRKCNTQPETIQHITG ACSTLTQTDYTHRHNQLANI IHQQLALKHKLIQNTNTPYY NYKPQTVLENDSCKLYYDR AILTDRTIHYNRPDITLQDKN NKVTYIIDIAVPNTHNIQKTF TEKMTKYTELKEEIVRIWKQ

						KKAYIVPIIISTTGVVPNHIH NSLKLLDLKDNIFISLQKAAI LNTCRIVRKFMQLEENQTY YTQ (SEQ ID NO: 1495)
R4	DongA G	AB0971 27	Anopheles gambiae	GAAGGCTAACCACAA TA (SEQ ID NO: 1251)	TAACATCCGGTGCAAA CTCATTAACATTAAGA AAAGAGAGAGGAGAA ATGAGAATGAGATTCA TTCACCTTTGGCATTTG AATAGCCCGGGGTAGG TGAAAAGTTCCCAGCA TATTGCTGAGAAGTGA CAAAATTCGGATAATA ATAATAATAATAATAA TAATAATAATAATAAT AATAATAATAATAATA TGCATAATAATA (SEQ ID NO: 1374)	METRSMRKRTTRLPEEGAPT GAGPGTGDRASIQRLEDEM VQERSFSQRALPVPRTQNRN GSPINHQGNAASANVAVAD RQQSLILAGGRRQRIMWTRE MNHYVIRCYYVYTRMETD MPGRVKMLGMFNDRFPRFA HQLDLSKLYIRQRAIILPEEL EFIKLEVRREFGEEEAGWRE SSRISARLNTIDQNTSRASED RDLDEPTAPGLSVDIQHQM ATAVTQFHGTDPLSRHRLPK LHYSYRLKTAVSIINQDVLP QYLDSVGSIEDLQLIVYSAA VAVVRTLWLRTYPQGDSEG RPCSKAEKPAWMRRLENRI NATRTKIGRMQEYQRGNSS MKVVRQIAEMVKPKELRDL TDANITEVLDIHLQRLSALA

						KRLRRYAECSKRKEQNRMF NINEREFYNWIRNDKPNFRE GLPDIGDFTQFWANLCEKPV QHNSEGMRLAEDERFSDGIE DMPVLVVNAQDIREATQYT RNGAAPGPDFVYNFWYKKL ITIHEQIAACFNTVLEDSRKL PKFITGGVTYFLPKDQNTKN PAKYRPLTCLSNLNKVLSSV ITQKVKDHCDTNNVMTEEQ TGRRKNTQGCKDQVIIDAVI VGQAAKKQRNLDMAYIDY KKAYDSVPHSYLLKVLQLY KVDGNVIKLMQHAMGMWS TSLHVTDGKVVLRSRSLNIR RGIFQGDTFSTLWFCLAMNP LSRTLNQQCNFGYLLKSEEI STRITHTFFMDDLKLFAETV QKMHHLLKNVQGFSNDIKM EFGIGKCRSIHLHRGQVLDA DSFRANEQEEIRHMVQGET YKFLGFLQLRGIHYAVIKKE LQDKFLHRVSCILKSFLSVG

						NKVKAINTFAVALLTYSFGV MKWSNTDLEALERTIRVVS TKHQMRHPKASVERVILPR KIGGVGIIDIQALCISQIHQLR SYFVESQNRHELYRTVYKA DHGLSALHLAQQDYQLNCN IKTVDGKGATWKQKELHGT HTHQLNLEHIDKVSSSTWLV RCDLFCETEGFMVAIQDRVI ATWNYRRCILREDVEDRCR KCNSGGESIEHVIAGCPVLA GSAYLDRHNDVAKIVHQQL ALRHKLVERFLPCYRYLPDP VQENDCIKLYWDREIITDILI RANRPDILVYEKRKKRATID IDIAVTLDHNVQTTFSTKVM KYHDLAEELKQTWYLEDIRI VPVIISATGIVPMALLRSLDE LELQRELPRIQKAVILRTCST LRRFLNPYN (SEQ ID NO: 1496)

R4	R4-1_BX	CADV01008175	Bursaphelenchus xylophilus	GGGATCCTGGGTTCCT	TAAGAAAAGCATGAAA	MTCNNAVVFPPADGNPAGT
				ACTACCTCGCTCCACC	TAATAAGAAATCAGAT	ADRNFAIRFPSSEPPGPSGIR
				TCCTCGCGATGGATCC	AAGAATAACAAGAATA	PSEPLDGRTGIGDVEHAQAG
				TGGGGAAGTCTCCGG	CTAATAAGTATATCATG	NGGFLVDVLEYKEAHRYGS
				ACTGAGCTAAGAGAG	TAACTATGACAAAAAG	KCEFCYVQTKGTVCSKPRT
				CGTTAAAGTAGAGGG	AACGCACCAATAAGAA	DAWLKCEILFLLHHAYTAN
				TGACGGCGTAAGTAC	CATGCTTGAGTGGCCA	QNKSIELAESAFRRAGITRRS
				CTCCAAGTTGCGGTGG	GCTCTGCAGGCAAAAG	KATIAKRWSLIQRGKGTDY
				AGCGGAACATCTACT	TCGAATTTGGAACAGC	KEYWDEYFEKFRYECNPTPI
				CTTCGGAGAGAGGGG	CGGTAATGGAAGACCT	VRRKRNRLAAGLQSPSSVPN
				AAGCTCTATGGCGGC	GCAACAAACGTGGGGT	GYEFERKRTCETPLDTKASS
				GTTAGAAAGGTTGGA	AGCAGGCAATATGTAA	LPLICNLLTGIVGVENVEEN
				CTACGGCAACGCCAG	CTATGACAGACCAAAA	MSVECTEPKELSGTANSSVP
				GGAGATGGGGAAGGT	CTCCGAAACTCTGGTA	GLAEGVYERRHNNVNEPAA
				TCATCAGGTGATACTA	ATGAGCCCGTGCCCCC	GCPQDVPVANNLIDSPTTND
				GTTCGCTACTGTCATT	CAAGCATGTGGTCTCGT	RLEAEFKAQLDRAERSYMR
				CGATGTATCCGGAAC	TCGATGTAGTTAGGAA	RRLPRLKNLSPDERMWIGTT
				CATACTCGCCAAGTTG	CAGTTCTCTTTAACCCG	VERLRLETVSEPVCEQWRL
				TGAACTATGTGAAAG	TGATGATTACGCCCTGT	ANAGLYAAIRSIAVMRPLDA
				TCTGGATCCAATCCAA	CTTAAATGGCAGGTGC	AREAHKTWLLNMKMTERK
				GACCACGGGGCGCAA	CACCAAATACCGAACA	LRQQIGWVETTRRTKNEAR
				TTAAAAGGTGTGAAG	CTCGTTGAGGTATGGTG	TERQEIVYRKVAKLRRERFP
				CAGCTTGCTGGTGATC	GTCCGAATGTGAAGCT	EMDLDSVSVHLKRKLELLK

				ACCACGGTGGTACCT ACACCGGCGGGGGAA CATCTTTAATGCCGAG ATGACCGCGCGAATA ATAGGGAGCCGGCAA AAGCCGAGCGTAAGT GGAAAGGATACAGAA ATTTGCTCAAAAGTAT ACCAACCCGCCACTTA TAAAAGACCGAGAAA AGGGTACCACTAGAG CTCATTATAAAACATC T (SEQ ID NO: 1252)	GGGAGTACAATTTGGT ACGAGAGCACCAGCGC CCCCGATCTAAGTGAT GACGCATGCGTCGGAA CAATGAAGACGGCTGG CAAACATTCAGGAGTC GCAAA (SEQ ID NO: 1375)	GRIQVRTAERLRRDTREAAG PYGKTALRGQGFAPNVKDA TQYWSGLAQPSGQKCSENS AILSDWKELVECNLSSLPDQ MEPLVVQGISRASPWKSPGP DGIFNYYWRQDFIVDWLKQ LMLDSLRTGHYPWKLSSGR TVLLYKDGDPTKAENYRPIT CLNGCFKMINSVVSEVILKR VENTIALPIEQMALRRKVW ACVESQIWDQIKQRKLSDRT QKCKVAWVDFSKAYDSLN HDAIKFVIGVLKLPTGINNY LLDSMQNWSTHLELKSSGK VVRGPSYPIKRGVLQGDSLS PTLFVVVTSIIVRHIKTIESSD IQMYMDDIKLYGKDQETLT RLIKELQTVSNKLGLCMNLK KCAILGDDLPEEINGIEHLKE SYKYLGVPQREITQVRATM AALEKKILTEVDTSLGAAEL SYRQRISRVNSKIAPLVRFV VQSMLVTPRDVLKVYNRLG

						GIDVEIRRRLVKYEIRYKKS NVARLYLDRKVGGIGFVNL CRIMVEAVAARAVYCRLAP SFNEFQDFLAEQNTSPITAA QTILDKCGINIELSTSTLGDV KKIVRNHYHELWLTAWKNT GLYKRWENDHVDIKRSSLW INRGNLSANNARIGIGIQDNS IFCRGFVGNKCDTKYCRLCG DGIESVSHIVTGCPTHRTNL YIERHDCVARNVYAYLAIR YGIPVPHYTQRVKTIEKNGD QSVELYWNYKFPCTRALEA CRPDIVLIDKVSKRTHIIEVA VSWRGRLQEMVDRKVYKY TVNGEYEADGSSRGWNIVR ELNDQYGFPVEVYTLVIGAG GEILPCTVKDVERLTGGAAT DNLIERMERSAVLGSCRIIKR HLAL (SEQ ID NO: 1497)
R4	R4-1_HG	.	Heterodera glycines	TGGCGATACTCGGAA CCTCCGGGGAGCCTG GTAGGAGTTGGCCTA	TGAGGACTCANAATTG ACAATACACCTCAGA (SEQ ID NO: 1376)	MISCDLERETLTQMALFRAR SDKTPTHAGIPAPDEVREGG CGQNRTNPAAPRGKAAAIQ

				CAGGTCGCGAAAGTC		RQNGITIPIXACAQSGLVRTQ
				CCTAGGTGCTGCACG		RVQQWSAVEESALKDVVVR
				GGTTGCGCTAATCCGA		NTDDRGLINWAKGVLPEWQ
				GGGCGCTGGGTTACCT		RLCQLNPTMYMARSSPSLSN
				TCCCATCGGCCAAAA		KWASLRRTHVGPGCPSKEG
				ACGTCTGGGCCTTCTT		SGPSQDLSDVKIQPARLAHD
				AGCCGCGGGTCCAGT		TVAELPQRTVPCGTDGHGVI
				ATTTCTGTTGAGCCTG		DSDETETALAEVSRSSPFGE
				ACAGTTCTTCCCGGAT		REPLDLGATERITRKRLRNA
				ATGGCGAAAGATTAC		VRDVVPPRKRRVPSTPSRKE
				AGGGCGGTATTTCGTG		QDLVPEVDGPAPTDVLTHPP
				AAACCTAAAAAWGGT		TESEPEPMLDPLSLVQLVRP
				CGGGCCGAATGGCAC		QLGRAMGWAAEEMELGNV
				GGACAGACTCACTTC		VMDVELKREFNREVRRVGR
				GGAGTGAGCTCGGGG		TPPDQMYKRGAGPPLPQKR
				GCATCCGTGTGTTACC		EPERVALLEQLIAARVERGI
				CCGCTGCACCACGCC		NRGLDWFLELNVAVFAAAR
				GAAGCTGTCATAGCG		VLSRRERVETLADRLHINDS
				AGCCCGAAGGGGAAT		ATLSEVSRRRAKAERKLRC
				GGCCATGGAGACTCC		AREQPWMSRRIRXLGVRVE
				AGCCTCACCCTGTAAC		RLKQLADLVRQRIAGRGNR
				TCGAACCTAAGTCCA		SSYEGPRRRFRLRPSLRSVTE
				GGCCCCTTTCTGGTGT		APVNPPLNGNEVYTFWHSL

				TGGCTCGCACTGGTTA		WAQSLRANTDDCQLREFKN
				GGAACACGACAGTCT		QLSAARHTDLTSVGTSSLVQ
				CGTGTAATCCCACACG		MFSAALRKMKKGKAPGPD
				CACCGAAAGCCCTAG		GIRAAWWGVFRRIAPYVAT
				CCCTCGTAGGCGAAG		WVVRVIRGAEPVANWICNG
				GCTGCGTTGCTGGTTT		LTVLLPKSSDNADPSNYRPIT
				CAGAACTGTGAGCWA		CLNTCYKLFTAVIAQITASY
				GTGGGGTTCGGATGG		VDVLGGLPRQQVALRKGV
				CCGAGTGTACCAACCT		WGTSVSLMIDALTVADARR
				GCTTGCTAACAGGTA		AKRPLGVCWFDFKKAFDSV
				GCATAGAGTAATATG		PHNLIRWILRVIGLPPVILSVI
				CTAGTAAGCAGGGCA		VSVMDQWATRLKIGGKVM
				CGAGAAAGGGCCGTA		PKTIPVRTGVFQGDTLSPLLF
				AAGGCTCCGGCTATG		CLSVWPISFALDQFPQYQFR
				AAGGACCTTGCGACC		CANHLQQGFSVGHVFYMD
				ACGCGTGTGTCTCCCA		DLKCYCPDREVLTAVIQQV
				CGTGCGGATTCTTGAA		QKSASALGLTIHYKKSAWL
				GCCAGAGTCTTGCACT		DQDGGKSGKAVLGVPXLVG
				GCGCGCAGGATGGAG		TYKYLGMHERFMIVSKDSL
				CCTGTGCAACTCCTCC		ESVRGKFMGRLKTLWTSKL
				CTCGCTGATCGCAGG		TFGQAMLGTKSXCMPVVRY
				AGTGGAGGATCACCA		VLQNLFLPKSEFNQTRLVLR
				CTCTTTTTACCTTGCT		EWDRQIRDLLDECNIRQVFR

				AGCTTGGGGTACCAC CTTGAGCTGGGGCCG GCCTTGCTAGCTTGGG GTACGACCCTTTGAGC TGGGGCTGGCCTTACT AGCTTGGCGCGCCAC CCTTGGAGCTGGGGT GGCGCAAGATCACTT GTATACGGTCTAACCA ATACATTTGAAAAGC GATCAAAGCGAA (SEQ ID NO: 1253)		SKTELYVSREEGGWGLPSM EDALEEEVVTKLAILVARQE TEPLFRVCEALERKRCPTPLS LGLQILKDWGVGVELQGRT LLLNGNTVGPSQATRKLTGE LVLRREAERLSRWRSKVKP GCGMTGGAWRDVPGIDVH LSNRWLVKGALSPTVVSNS LAIRANTVILRGSGGGYTKG TLLRCRGCGNTGETRRHIVS ACSLGRQKGAASRRHDNVC RILVRAICHKLNIEPPNSANF PHVVVLEGSGAKMWIDFPF VVPHKIRHTRPDIVVLFEWN GVRRLSVIEVAVSDVANMQ TQHIRKSHRYGTNSTEPFVA GVTPTYRNDCLAAQLRAKF KAQQVDVIPIIVGTTGETLD GEFGRIRKGLPMLTKLQMPR LWSEIQRAVILGSYRILVEHL ALPKGGA (SEQ ID NO: 1498)
R4	R4-1_PH	.	Parhyale hawaiensis	CCGACCGCCAGCGGG ATAACTGGCAAACCC	TGAGCCTTAGGTCGCG GGATGTGACCCGGCGC	MKMSHNRDTPSNGVKGTSV RLGTSLVRSPVGEAGAVRER

				TGTCTCGACCACCGGC CCGTGAATCCATCGG GGCGTATGAGTCTGA CACAGGGGGGTGTTT AAGGTGACCCTGTTGC GAGGAAATGCGCAGC AAAAGCCGGATGAGC CTTAGAACATCGAAG GCCAACGACAACCTG CCGAAAGACTGGCGA CTAATCCAGTCAACCT CCTGTAGGTCACCGGC TGGTCATGCTGAATCT CAGCTTAACCAGGCG TACGAACTTACAGTTG GAGGGTCGAGCACCC CTGATGGCTGAAAAG GACCATCAAAGTCGA AGGTAGCCAACGAGA AGACCAACGGCTGAT TCAGGCGGAAGAGTC AACTCGTTGAATGCGT TCGACAGCTTGGGGT	CAGAGTGTAGAGCTGA ACATCGCTCAACCGAT CCAATTTGGGTCGTGA AATCCCCTCGATAATA ATAATAATAA (SEQ ID NO: 1377)	GTHPSESVSQDSDASVNATG EGSVREQAPLSPPGAEEATV PTQRRTRHKWSREDRVVLW ECFVASKREGPGYLKRLKQ LWDERGIPGNFPQASLSGQI RQICSKNLLSEEERLQIAAR MEAQVASPSADEPARQVPT RPVTPPRSPPVEPARRPSIPSE ETPDLGAVPSEIDSADPNRSP SRGPRHLPAHNMSQSESEDD VTDPDVSDQQRSDSLEPRDL LRNSSVESTPGHPNQELSDT LLSNYVPSEIDSDDPNQSPRR GPRHLPAHDMSLSDSMDEE TEPDLSDQQRSDLLELRDLL RNSSVETTPKGHPSLRHLPE PKIRAAAFRVNSVIGKIHTN NITETNALIKAGADLAVRILE VQPRPQRTQRKKDPPWKHR LEKNIAEIRKHLSWISEWRR GNLHDEEKKTLLESRYRCLE VGLTNLEDTLKQRLSAKRS KVRRFEARVAGFHQNQLFN

				AGATGGAACTCCTAA GCCCTGAAAGGCAGT CCATCTTCGCAGACGC TAAGGTGCCCCGCCG CCTGAGGGTTATCAG GCCCCGCCGCCTGAG GGTTACCAGGCAACC (SEQ ID NO: 1254)		TNQKRLYQTLRGEETSSDSP NAEESIRFWSDIWSKEVRHN NTAEWLHDVKEKNVAADP DLTITSQQLKKQLSKTKNW KAPGPDMVQGYWIKTFTSL HSRIAAQLNHCLQRGTVPT WMTTGKTVLIQKDKAKGTE VSNYRPITCLPLMWKVLTGI IYERVYQHLDSKKLLPDEQK GCRRNTRGTKDHLLVDKLL TKDARSKKKNLSMAWVDY KKAFDMVPHSWILECLDIYG IAGNIRNLIATTMPNWKTQL TSANKHLGEVSIKRGIFQGD SLSPLLFVLTMIPLSETLNKA GQGYNYSRTMKLNHLLYM DDLKLYAKSKDQVEQLLNI VHQYSQDIKMQFGVSKCGV LNIERGEVTASEGITIEEGTIK DIEEAGYKYLGVMEYNTILH RTMKDSIRKEYLTRLRLILK SHLNGGNTIKAINTWAVPV VRYSAGIINWTKKDCTDMDI

						KTRKLMTIYRALHPRSCVDR LYINRREGGRGLISVEDCVE AEKRALSQHFRESDDPWAR CLVEAKLLKETETADQFKER RRLDRTNKWKSMKMSGQY LEAVQDKIVPDSWNWLLRG ELKRETEGTILAAQEQALRT RYIQNKIDKRNVPSTCRICRS SDETINHVISECGVLAQKEY KRRHDKVARHLHWTLLRIH NFPVSERWYEHEPAPVVEN EAVQIYWDKRMETDRVLHA NRPDIVVKDKQEKSAKLIDI SIPFDSRIVDKEAEKKEKYR DLAIELQRLWQMKVDVVPV VIGALGAMSKNLKTALREL KCGHLHPGTLQKSALLGTA HIIRKVL (SEQ ID NO: 1499)
R4	R4-2_BX	CADV01009048	Bursaphelenchus xylophilus	GAGGATCCTGGGTTCC TACTACCCTGCTCCAT CTCCTCGCGATGGATC CTTGGGGAAGTCTCCG GACTGAGCTAAGAGA	TAACAAGTGTAATAAA AACCACCCATGCGTAG TAAACCGATCAATTATC TAGCAAAATCGCAGGT CAGAAGACCAAAGAAC	MVFNNCKPKHLCPAIRPTGQ QETNGGSEGTAEPTAGPSRP AVSEDAAQPVPLFEEGEYIR AHRDKTCPYCEVLWIGARS SKARSDSWPLCQILYLMKR

				GCGTTAAAGTAGAGG	CGACCCAGAGGAATAG	NDDLRGQRTRYPLLESSLRA
				GTGGCGGCGTAGTGA	GACCAGAGCTGAAACT	AGIARTKFAIIKCIRNVLRDR
				CTTCCAAGTTGCGATG	CTCAGATACGCCAACG	YVPNGPYSEHWKIYRANSG
				GGGCGGAACATCTAC	GTCCTAATAAAACGTC	EVPQGATITKGKRSARVAGL
				TCTTCTGAGAGAGGG	GTTAAGTAAAGCATCG	PSPSQSGHHTKRIQAGTGIET
				AAAGCCCTATGGCGG	TTAAGTACAAAACAAA	ETTVTETNTTPEVSHEHRDP
				CGGTAGAAAGGTTGG	AGCACTGTAAACGCGA	CGEPETSAANVDKVTELTED
				GCTACGGCAACACTT	GGCCCCCTCTTTGCCAA	GSETRGTANVANGGVSVSD
				GCCATGATCAGATTCG	ACTCCGGTAATCCTCGT	PGRKRQSSSQNRGNIETTNP
				ATCAAAATTAGCCTCT	AGTACGGTGCTTTTTCC	ELVGMWEDMFGVQLDGAM
				GGGGCTGGCAACCCT	CGCTTCACGAATCAGA	RTTERPRLPKLKHLSEPERL
				ACAACGGATTGTAAA	ACGCTGCCAGATCTTGT	WIRAKLEQAWLQCVSYDVE
				CTGAACTATGCTATGC	CCGACATGGGCCGTAG	QQWLNANAVLYAAIRSVAA
				AAAATGAAAATAAAA	GGTTTGGGAGTACCAG	SRPCKEAREAQKTWLDNKK
				AATGGGGGCTTTACA	CGTGGGGCGGAGAGCG	KDEAKLRRLIGRISSVHSMP
				ATCTAAGGTGTTGGCA	TACCTGGGTACACTGC	KGDRTPREKKLVKNITKLK
				GATCACGAAAACTGC	ATAATCGGGTCTCAGA	NTHYPDMDWGGLLNHFKV
				CCGTCGATGAGGGTG	AAACCACCTATGGTTTA	KLSQLKEKISVRVAEHKRKV
				AGATAACATCCTACG	TTATTCTGTCTCCCATC	NRNAAGQYGKSVAGSAGL
				GAACAGCCCCTGCTG	TGCAGGGTAGCTTTTCG	APDVVSATAYWSGLAQPGP
				GACCAAACCAAATCA	TTGGGCCATAGGAGCC	KKFKASSPIFQTWKDDVAK
				TCCACAAATTGGAGG	TAGGGGCAAGAGTGCA	NLNTEPVLLYPIIKECIRKPSP
				GTTTTCTTGGTAGTTT	TGTAGTCTTCAACGGCC	WKAPGPDGIYNYYWQQEFV

				CCCTTGGCACGTCTTG	ATGCCAGGGAAACTTG	AQWIQTLVKRTLDIGRFPTA
				TTTCATAAGCCAGAAT	TGAGAGGTGAGGGATA	LMCGRTVLLFKSGDKSMPQ
				AAAAACGATACCATA	ACTAGCATCAGATAAT	NYRPITCLNGCFKITNAVLT
				CAGACATGAGGCTGG	ATCAGTCATGAAATTA	KVILQRVQDTCALPREQMA
				GTACCTGCCAGCCGC	GTAACAACCAACGTTC	LKPKVWSCMEAQLRDQAL
				GACACGGAAAACCGG	ACCGTCGTTGGCAAAA	QSEIGDDCKTAWIDFSKAYD
				TAGGTGCAACCGGAG	CACCGACTAACGATGC	SLDHDALRFVIQTIALPAGM
				ACAGCTAGGGGAAAG	TAGTTAGAAAGAGTCG	EEYLLKSLDSWRTQLVLSD
				AAAATAGTAAAGTGT	GGTCTTCCCAAAGTTAG	AGKVVSGKPYPIKRGVLQG
				CGAAAACAAAACAGG	GTGCTTGCACCGAAGC	DSLSPALFVLTTSPIVAHLQR
				TAGCCCCTGGCTAGA	CGATCCGCTCTACCCAC	TCPTGRIQLYMDDIKLYGKT
				GGGAATGGGACATTG	AGCTCTGCCCAGCGTT	ESDLCMLIKETQRVANKLGL
				TCCGATTAGGTTGCCT	(SEQ ID NO: 1378)	NINLKKCALFGKSIKQSIAGF
				TGACCCAATGAAAGC		DPLGDRTYKYLGIPQRDVA
				CAACCGGGTTTACATT		DIKQAYDELKAKTVQTIGET
				ATCGGTCTACCAAGG		MACDYLTTRQVINRLNSKIP
				GCAATGACCAGAGAA		PVVRFVTQSALCSAPMTRGL
				ACTGCGACTATGCCGT		YNKITELDNVSRAELRKVLI
				AACCCTTCGCAGATTG		YKATNVSRFYLATKEGGFG
				CCGATGAGAACCATC		YASLQQVFVEAVVSRAIYCL
				GATCGTAAGTCGAAG		RAPSLCDIREFILSKFDPVKV
				CCAAGCGGATTGACC		ARIALARSKIDMDIERMDM
				AGTGGAGGGTTATCC		ASATRTIRQHYQAKWKTLF

				CGACAACAGC (SEQ ID NO: 1255)		QQSKLYQKWVQHKIDIPNSS RWLQRGEISPRNCRIAVAVQ DNTLLCRGFVGSKDPNKQC RLCNAGIETASHIVTECSTHR VHMYIERHDSVARNIYAVL AKNCGFWIPHYSQKIPTVKI TKSYELYWNYKFPCTQALE ACRPDIVLIDRAKKRILVVE VAVSYVTRLEQMTQRKLYK YGVNGEYQADGETRGWNIC RELVQKYNMRIDLCIVVIGA CGEILPCMVKEIEKISKVSGR QLLERCQRSAVLGTVRTVR RHLAN (SEQ ID NO: 1500)
R4	R4-2_HG	ABLA01000389	Heterodera glycines	TGTGGCGATACTCGG AACCTCCGGGGAGCC TGGTAGGAGTTGGCCT ACAGGTCGCGAAAGT CCCTAGGTGCTGCACG GGTTGCGCTAATCCGA GCCCCTTCCGGTGTTG GCTAGCACTGGAGGG GGGGGGGTCCAAAGA	TAGTCGTAGCCCAGAT GTCGACAGCCCCTACA CCGGTAGATGGAAGGA GGTATGTATCTAAGCCC ACGGCAAGCCACCAGT GGAAACGGTGACCACC TCTGTTCGCGGAAATGC CCCGTAGAGTATATGA AGGTCGGAGCATTGAA	MRKSFLQHIPELSSHIAMSVP ARNYPKMCSLAQGSSGTLS HNGKGVAMHRCPSDDCAG KDPPQRGSQKGNLRSVRWT PSEEKAVFEYWSRLEQHAM LNGSEARGTCAITRSQFLIH WDGERESRSLSDGVPEYPM RTERAYYERVRLLRQRGWQ WDCANECLVIGQCAEPCRK

				ATACGTGAAGCGGAC	GCACAAGTGAGAACCC	PNVVAIKADKGMKRSLVKG
				CCTGTGTTTTCTCTGA	TGGAAGTATGGTGGTG	KLLSLPHVMGEINQVSVQVE
				CCTCATACGGGTATAT	AGCTGGTAACCTCGAA	VPLPSVPASVPQVEGVESKG
				CCCGATGATATCTATA	TTCCTCCTATGGGTGCT	FTETEPSNKPSLEGNPAEEGL
				CAGTGTTCATTTTCAT	TGCGCCCGTAGGTCATT	RKPERVNVPVHGIISDSERK
				TTCTTAACCTGCTTTT	TGTGTATGTAGGGATG	DLKDRFWSAYKTAKRSVGF
				TCCTCAAGAGAAATA	GAAGGAATCTCGATAG	RPALKIEPNRVNRAQWEVL
				TGACCACATTCGTTTG	CGCAAATCGGGATACC	DSCVVEVLKKRETSNGYRG
				CCATTTCGAAGGTGCG	GCACTGGCGATCATGC	CVLRHLNVAVYAAGYVLAE
				ACTTTTGAGGGAGTTC	TGCCAGTGGGCCCTGCT	GNKERRQVIRRQSAEWLLR
				TCAGGGACATCAAGC	GAGTTGGGCGGGTTAG	QKSEINNIRRHIGWITDELTR
				TGTTCATGGACGGCTT	CAACCCGTCCAGGGAC	RRTGKNPTSRQLSNFAWLQ
				GACGGCTCGTGGCGG	GCAGATGCCAAAAGCC	RRYQVLGKPVRETRDLEVQ
				CACGCAGCCGCGCGG	TGGTGACACTCTAAAG	RERLVSRLRLAQDRINSSMD
				GCCAGTTAGGGCTGG	GGAGTCTACGGCGACG	REERVRKRMLPLRRKLEEPL
				TGCCAATGAGGCAGC	AGAACATCCCCTAAGT	GDSKLDTKQARTFWASLIGE
				GAAGAAGTCTCTGAA	C (SEQ ID NO: 1379)	RKEFGKIPELENWAEEVRSK
				ACGCCAAAAGCAGCG		VTDGQGFASDHVDQTVWK
				GGAGAAGAGAGAGAA		KILGKARPLKAPGPDGIPNL
				AGCGGGGACGGGTTA		LWKRLPSANQALFKWLMGI
				GAGATAATAGCATGA		KRKQLSVPSWLTKGRVVLL
				TTGACGGCGAAGCAA		PKGGDPVDPANYRPIACLNT
				TACTTTGACTTCCAGG		QYKLVTGMVTAWVSEHLTT

				TACGGAGGACGGCGG		YSILPIEQRAMVSGTWGCTH
				CAAAAAGCAAAAAGG		AMVIDRAITSYAEATGLPLY
				GGACGAAACCGATGG		VGFVDFAKAFDSVSQPWIR
				CATTGCCAAAGGGGC		YALKVAGVHKRIRCLIGILM
				CGAAACCATTGAGCT		KCWSVRYEVFKSGRVLRSA
				GGACGCTGATGGGCC		PLAVKNGVLQGDTLSPLLFC
				GTCTGCCGCGCAAAA		LSVAVVSSAVGSLFDFEVTIP
				GGAGGGGGGAGCCCC		GRGVMQQQNHLFYMDDFK
				TCCGTCTGGGAAGAA		GFAPSEASLTRMLVTLERTA
				GGTGCGGCGCCATTC		SALGLKINKRKCALVHPRER
				GTCCAACAAACCGGA		ENEETGSDIPVLGLRDTYKY
				ATAGATCTTGATGATC		LGIEERFGIVFEDAWDRVRT
				TCAGCAATTGATCATA		KMFERMRTLLCTEHTFGEL
				TTGATCTCTTTTTCCA		RAAFASTIAPVARYLFLNVI
				GTTGTTGCGATAAATT		VGGPSWSETLTKAKDMDLR
				ATCGTGCATTATTTCT		IRRLLWERRDNEPGWRFKH
				TCGATTTCTCCAAAGC		CSADRLYLRVQYGGLGFVS
				TTAACTCCTTTCCTAA		VEDTLSESIIYCWAYVQCRP
				TACCTACTCATGTATA		ELELARELFGTLNRSARSGI
				CGTTAGTAGGCATGTT		KQSIAKGARKVFRSYALLSK
				TTATGCAGGTAAAGA		NSAQRVSDLDGDASPGFRV
				TAGACCTGGTGCCCCC		GEMIFMEPTRGARAIVKILR
				GGGCGACTTGGGATG		KENDSRRLAAWKGRPMGG

				TGGATGATGGTTGGC GGAGAGTTCTGATGA CGCCGTAGTTCCGGA GGAACAGTTCCTACTA GTGCCAGCCTGGGCT AGTGGAGCTGGTCTG GCGAGTGCTTTGCTCG TCAGATAAGGGGTGG TTGGGGTGGTACCTCT GGAATGATGTGTCGG ATAAGCACTCTGCCTT ATAAAGCTGTCGTTCT GCCCAGGTTTCTCCTA ACCAGGTTGAACTTGT AATGAGCCTTTGGGTA TCTGGTCGGGGTCTGC GCGGAGATAGCTCGT GAGGAGCTGTTATTAC ATTATTGGTTGAACGC TTTTGGGTCGTACTGC ACAGATAAAT (SEQ ID NO: 1256)		RVVSLPELDQVHSYHWLIR ARIGRRSFRDCIAAQEGQLK ARELMCPHINAKAKWCRRC GDGRVETEQHILSGCAWSR TGTMLDRHNGVVRQVHTA LCRKYGLPVSSHVVPLHAVI ENEHAKILYDVALHTSPAGV LPREDGSTSYTGLRSTRPDM VIFDKKARTILIVEISVPWRE NLVKQELIKWRKYAINSMIE PLELAEAEIPGPNLKHALGL AYGTSFPTVKVVPIVVGSCG EVLPNITKRLSELGIPKRGIPS LLESIQRAAIIGSGHVIRAHL SVPRSESET (SEQ ID NO: 1501)

R4	R4-2_SRa	CACX01002006	Strongyloides ratti	CGTCAGAAGAGCAGGTGTTTTTCAAAGCAAAGACTTATTCTACGAAGGGGAAAGATGATCAAACATGCAGATTTGGCTGCAATGAAATAGAATCAAACTACCATGTGGTGACATCTTGCAAATACCATTCATACACAAGAAGACACGATATGGTTGTGTACAATATATTAAAAAGACTAAAAGAAAAGTATAAGTTAGATGGTGATTTACAATATGGTAGATCAGTATTTAATGGGAAGAAGGAAAAAATTAGGATCAGATCTGGCCATAAGTTTTTGACTGAACAACCGTTAGTAAACAATAAACCGGATATTGTTATAACGATGAAAGAAGGGAAAAAAAGATAACGTACATA	TAATTAGATTTATATGTGCCACTTTACCTCAAAGATATATTAATACTAACTTAATTATAATTTATTATGAATTATAATATAATAAGTTTTAAAATAAAA(SEQ ID NO: 1380)	MQKFSVPKDSSQIFLVDSILNKHICSTKNKVKDVIKRRSIIKTLICAAGLTLRKLVCGKLGNNKYNSKINQLWKKERKIINCIEDLKHLIETNKRRHNFGKRLRNAKVSPSEMLKDYYNKLRYIKNEITACLDEHKKAILRTKFKLTPSIKIISNIQNHNDEEAELPKEEEFVKYYKELFTNKDGDDKETPHLDNWLKKFSKTLIVDWTINDKEILEALKYCGNFKAPGSDMVMKVCYKWFKSAQNYLIRWIKSTWYGEYTINKKDTNAVTFMIWKRDGKPKNDVKSYRPISCLNCDFKLLNKLIANKIYESIEKILPINQMAVIKNKHGTCEALLLYKSLVQSMKFRRTKDVKEIWCSWIDFSKCYDSISHKCLKKMIQSIKAPPIIHKLILDGIDSWNISICNGKNISKTKIPVKSGILQGKVASSLYFVLLTGEISYALNKEEQVPIETITPSNTLKINHI

				CTAGAGATGTCGATA		SFIDDYQLYATSQKKVEKLT
				CCACATATCTAGAACT		IKLREIAEEMNLKLNPQKCG
				TGAAGATGCAAGAAA		IYGTDDLGKRLMLKESSLNF
				GAATAAAGTATAAAA		PYTSEYKYLGLVENSLDLKD
				AATATTGTGTGAGCTC		INIQLFKDKILSKYSTIFESRL
				CATGGTAAAAATAAC		TTHQKRKVFNSTISPCAAYY
				GAATGATAATGTTGAT		LGNLITNKCSIQELLNECKKF
				TCGATAGCTCGTGATT		DQMVRNQLVNQNIKKLQVS
				TTAATCTACTAAATTT		NSRIYLPKEYNSLGLNEIEIE
				AATGGAAAGAAAAGA		VAANIIRKACYIKKRETLRG
				AAGGTGTAAGATAAA		VDKLYIAMSKNGHRNTLSD
				GTTTGGATCATTTGTA		ALYITKKYSNFQINWNIMG
				TTTGGCTGTTATGGAG		MVKDQNNILLDAKKIIENIK
				AATATGTCTTGACTGA		EKRRNLWLEHWKKGNMTY
				ATCTGCTCTAAGAACA		ANEAIKKEFHLPDLNIDSKY
				CAGAAAATCCTGATT		LMLCYAGSEEQIIYNGHVSL
				GAATTGGGATTCTCCA		VNQSSPSSRLCRKCNKLEET
				AAAAAGAGATGGATA		SYHVASVCEFHKKNLHLMR
				GCTTGATCAAAGAGT		HNSAVYHIITELCRIMKVKC
				GTTCTTACAGTTAAAT		TLRYPEASGIIKSGNMKIAA
				GAATGAGACAGCAAG		GVKYTFGTAKIYHNKPDLV
				AATTATTATGAAACAT		WYTPEVIYVIEVSISSLKNAK
				CTAGAAGGCGAAAAG		SQMKMKTARYAVNSTKKL

				AATCAAGATGAAATC AGCCATTAATAACTTT TATTTACCCAATTATT TATTGCGTTAACTTGA TTATATGATATTTATT GTATATTTATTGTTTA ATTTAGTTTCATAGTA CAAATAAAAGAATAT ACTACGATTCTTTACT TTTAAGTTCTACGAGA ACGTTGTTTTAGAATA TTTTAATAATATTTCT ACTACAATTAAGTAAT TAGAAGAAATCAACG AAAGCAGCTAAACTT ACTCGCAAAATTCGTT GATCGAGGCTGGAAT GGCACCAACTAATAT ACTAACCAACAATAG AAAAAAAAAGAAGAA CTTCAGTGGACTTCAA ATAAAATAAAATTATT GGTAGAATTGTATGAT		ENFAALNNLKKGENFVEILS HKANFKRVHFMPLVFCTFG EIPKETMKYLEKLNFSNEKI KTIASPIARYTGRTLKAHFT N (SEQ ID NO: 1502)

				AAAACTGAGAAAATT TGACTAAAACAAAAA GATTAGAGCAAATAT GTCATCATTTTCCTAA TCATACAATTAAAGC GATGATGACGAAATT GAGAGAATTTGAAAG AGAAAAAAGGGAAAA TTGTGAAATTAAAATG AGGGAAAATGAGGAA AAACCTGAGGAAAAA TTAAATTTTGACAACT ATGAAGAAGCAAAAT TGAAAAGAGGAATTA AGTGTAAAAAAGAAG TAAAACCAATTGTTAT TGAAAACAAGGATTA GGAATTTTTGAAAAC GGAAAAATTATTTCTC AAGTTTGTCAACACA ATCAATTACCAAAGA AAGAAGGGTAAATCA		

				ACCAGA (SEQ ID NO: 1257)		
R4	R4-3_BX	CADV01008832	Bursaphelenchus xylophilus	GGGTCCTCGGTTCTTA CTACCGTGCTCCACCT CCTCGCGATGGACCCT GGGGTAGGCCTCCGG GCTGAGCTAAGCAGA GCATTAAAGTAAAGA GTGACGGCGCAGTTG CTTCCAAGTTGCGGTG GGGCGGAACATCTAC TCTTCTGAGAGAGGG GAAGCCCTATGGCGG CGTTAGAAAGGTTGG ACTGCGGCAACACTA GCCATGATCAGATTCG ATCAAAATAGCCTCTG GGGCTGGCGACCCTA TAACGGATTGTAAACT GAACTATGCTAACCTG TCAGTAAAGACAGAA CGGGGGCTTTGCAATC TAAGGTGTTGGCAGA	TGAGACCACCCATGCG CAGAGTATCCGAATCA GTGAAAGTCCAAGTTT CAGGACAGAAACGTCA GATAAGTCCAAGAGAA ACGAGAAAACAAGTTC AAGTATGCAAGAGTTA ATCAATAAGAGAGTAC CGTAAATGTATGACCC CCCCCTTTGCCAAGTCG ACAACTGTCATGCAGG TGTCTCTCTTTTCACCC GCCATATGGACCAAAC GCTATCCAGCCTCGCTC AGAAGAGCCTTAGGGC TGGGGAGTACCACATG TGGCGGAAACTGAATC TGGATGCGATGCATAC CGGGAGCGCAGCGAAA TCACTTAACGCTGGTGC ACCTCCTGCTATCGTAG	MISRSQADRPVEGQPVTAM SFHNLEPNNLYPENLRPTGS QDANRGVADIAEEVTGPSG LVTNEEAARAPPLFVEGEYK RAHCGGGKCHYCRVLWIGA RSSKARTDSWNLCEILFLIN KCMELGNVRRIYSPLESSLK EAGINRTRHAIVKCRLAVM RDRFVDNAPYSEHWRLYNA CAENRAVVVPMDSATTVKK RTARQAGLESPSQIGVAGKR VHEAETGTDRINAVIETNTT PLEDIDLSPETPEGLAELPST VEIMELTEDGSRSRGTANDA DGGVSISDPLRNRPSSSQESR NVPEQVDPDGELVWESLYG AQLRGAMRTTDRPRLPKLT KFSAAEQLWIKSKVEKARLE CVSYGIEQQWLRASAVLYA TIKTVAACRPYNKAREAHK VWLENKRAEEKRVRRIIGRI

				CCACTAAAACTGCCCT	TACTTCCTAGATAGATG	ETVRTMPKGKRTDKQIRLA
				TTGATGAGGGTGAGA	AGTAGGGTGGGCTAAA	RKINRLKRVSFPEMDWHGF
				TAACATCCTACGGAA	GGTAGTCGTCTTCAATG	LNHFKAKLDLLKKLISVRVA
				CAGCCCCTGCTGATCC	GCAATACCGAGGGGTC	EHERKISRKIAGTYGKSVSG
				AAACAAAATCATCCA	TCGTGACAGGTGAGGG	QSGFTPDVVAATTFWSGLA
				CAAATCGGAGGGTTTT	ATAACTAGTATCAGAT	QPGPKKFKKNSLIFQTWKDS
				CTTGGTAGTTTCCCTT	AATATCAGTCATGAAA	VVENMNTEPVLLHPLIIECM
				GGCACGTCTTGTTTCA	TTAGCAACAACCAACG	NKPSPFKATGPDGIFNSYWR
				AAAGCCAGAATATAA	TCACCGTCGTTGGCAA	QGFIANWVKSLIQRTIQTGE
				AACAATACCGAAGTA	AACACCGACTAACGAT	FPASLMCGRTVLLYKNGDT
				CAGAATGGCTGGAAC	GCTAGTTAGAAAGAGT	AKPENYRPITCLNGCFKMTN
				CTAACAGCCATGGCA	CGGGTCTTCCCAAAGTT	AVITKVIVQRVQDTCALPGE
				CAGAAAACCGGTGGG	AGGTGCTTGCACCGAA	QMALKPKVWACMEAQLRD
				TGCACACCGGAGACA	GCCGATCCGCTCTACCC	QALQSEIGNDCKTAWIDFSK
				GCAGGGGAAAGATTG	ACAGCTCTGCCCAGCG	AYDSLDHDAIRFVIETLALP
				CAAAGTGCCGTAAAT	TT (SEQ ID NO: 1381)	DGMEKYLLKSLESWKTKLV
				AAAGATGGTAGCTCT		LSNRGKVATGRPYKIKRGV
				CTGACCTGAGGGAAT		LQGDSLSPALFVIATSPIVSH
				GGGACACTGTCCGATT		LKRVCPSGRIQLYMDDIKLY
				AATGCCTTATCCCGAA		GKSETELRMLIKEVQKVAN
				GACGGTCAAGGTTCT		KLGLQMNLKKCSTYGAGLT
				ATCTTATCGGTCTCCA		ESIAGFDPLGDRAYKYLGVP
				AAGGGCTCTGGCCAT		QRSVADTNLAFGELEGKVIR

				AGAAGCTGCGAGGAA GCCGTAACCCTACGC AGATTGCCGCCGTGC (SEQ ID NO: 1258)		SIEETMACEYLTMRQVVTRL NSVIGPLVRFVAQSVLTSQA KVSWIYNKISDLDSKIRAKL AQTGLRYKKSNVARLYLSK SKNGIGLVNVQQVLVEALV SRAIYCLRAPSLVEIREHILT AEFDPVGAARTVLRRSRIQL EIERVEMASAISAIKTNYQA RWMTKFTQSKLYQKWVHH DIDLANSNLWLERGEISPQN ARIAVAAQDNTLLCRGFVG NRESEKQCRMCNMGIETCS HILTECSYHRAHMYIERHDS VARNIYAVLAKDHGLWIPH YSQPVSSVTKTPTCELYWN YKFPCTRALEACRPDIVLIDR AKRTILIVEVAVSYVTRLKQ MVSRKVYKYGVNGEKGAD GESRGWNMIRELSEVYNMK VNLCAVVIGASGEVLPCTVK AIQSISSKTSSRQLLERCQRS AVLGSTRVVKRHLAEFH (SEQ ID NO: 1503)

R4	R4-4_BX	.	Bursaphelenchus xylophilus	TGCCAGCGGTGTTGAT TAGGTCCAAGTTCTTT GGCCAAAGATCCGCC CTCGGTTTAGCAGTAC CGAACGAGTATACCTT CAAGTGGTGGCACTG AATTAGACTGAATACT CTGAACTGTAGACTTT TGTGCAACTGTGTATG GTGTGGAAGACTTGTT TGTACCACTATCAGCT TTATTGGGGCTCGTTA CTGTTTCATACAGGTA GATGTCCCCTTTAGAG ATTTCCCTGCAGTTTG CGCTCCAAGTCGCTAG CCTCTTGCGTGTAGTC AAAGGAATACATTCG CCGTCGGTGACAGGG CTATACCCGGCGACTA CGGACTTGTTTATTAC GTAGTGCAGCCTCGTT TAAGACGAATGTGAA	TAATTAGGCAGTGCTCC TAGCGGAGTGCCGTGA AGTGGTGTCAGTACCC GTCGTGTGGAAAGCCC AGGAGGGTTAGTACCG ACAGTGGGAAACCCGC TGCACGCAACCTAAAG ACATTTGCCCTTCGGGG GAGAGGTATGAGACCA CACTAGTCATGGTTGCT TGCGCAAGCATGACTC ATTCATTGTAAGTTCGT ATTATGAGGCCTCGCG GAATGCGAAAGTCCAA GAGCGCCAACTTGTCTC TTTGGGAGGTTCCTCCG GGTTCCTTAGGGTTGGC GTACCATGCTTCGTGTG AGACCTAGGCCGCTTG GACCTAGAGTTAAGCT GTTGCATGTTTGAGGAT GCCTTAGGCGGACGCC ATG (SEQ ID NO: 1382)	MEILWEDLRLKIEDRYGVTL PQRSASSLKNQYPKVILRGL PDSGLPWAGVQVNDTGQV VVVDHAEAATLRGSSPAAV DGEAEEPVVPPLPAAEVVES AADAAVPDPQSEIVADQGV ETRPVENPPANSRETETEPV EVEPYLEGQYKFFVSKILGK SMWRKPIKYPRRVPETLWR QANELIERSIRQGEVSIQSLN CMVYAAGCAVKSSLDKKD QEAKRRESEWYACRKAEIK ALERYLNFIDLELKRRSASR PLTSRQRQNLGVLITKYGRA RVRSGVRLSELQAMLRDAL VGIRKCMAKRSADKKRKQG KFVPIQRYLEPSSAEPRLSPD TVRAYWNDIVGSSQQSTSDS TIQDWSSNLSVPSQELNASKI MGWWRAAVSKSKPNKAAG PDGIPGVLWKRFRSASEWV CTWLYRLLQKRRIITPRWLS VGRVVLLPKKGPLEDPANY

				AAGAAGGTGTGATTA		RPIACLNTVYKLITSVVEMA
				CTAAGCGTTATGAGTC		VREQIQACPGLVPYEQIANR
				GGGTTATCTGGAAACT		KGVWGCTHASIVDRMITGA
				CCGCCCCGCCGCAAT		SREGKGGGFPDLRVLFYDC
				GGCTTCAAGGGTCAA		KKAFDSVNRDHMFAVLRVA
				ACTGCTAACATTTTAA		NVNVKVVHLLHTLSQQWC
				ACCAATATTTGGCAGC		VRYELRRNNRVERSSPLRVK
				AGCAGACTATGATGT		RGLLQGDTLSPTWFCLCMA
				CAGAGTGCCGGGGGT		PISASIKTLNPGPTLRPNMGN
				CTATCGATATAAGACT		GRNRGQVAIQVSHVFYMDD
				GAAAGCGATAGACGT		LKVYCPRVADQRRMEQNIP
				GGAGTGAAAGGATTC		QLFGEIGLSINASKSAAAAA
				CCTTTGTTAAGAGAAT		VGRYVESELPVLGTKDEYK
				CGGTAGAATTCACTTT		YLGIESGFVVNEVAALDRM
				TTACTTATTCAAGAAC		QAVLLNRVEAILSVKEHTVG
				TTAACAGCAACAAGC		QRRDAIRAKAIPGGAYILGHI
				ACTCGCGAGGATTAC		ILSDLDPRGAAERMRRLDIEI
				CGCCCCAGATTCGGTC		RRLVKSAGILHDKCSTARIH
				GGCGGTACTTCACCTG		LSCEQGGLAWPSMERAYYV
				CTTTCTTCCACTTTCG		AVAYSASYLLTSQDETISRA
				GAATCTGGCATCCTGG		RDYFVSGRLSNKFTVYKHL
				CTTTCAGTGTGGTGAT		TSIVDSLGLSVELPDPNGLPT
				GGCCGGCTTGAGTTTT		GQPSVLARTIARAIDAKLEA

				CTTGAGTCGTGCGAGT GCCTCATCTGGGACGT CCGGACCGATTGATG GAGTCTGCAGTGGAC GAGGACTTGATGGAC CGTAACCATAGTATAT CCCTCATGCGTCTTCT CGACTCGAGGGGGTA GCTTGCACTACCCACC CTTCTCTTCTCCGATT GGGATTTAGACCTAG CCCTCTGGTGTGTCTC GACCCGCGATATCAG ATTCCTGAATCGACTG TGAGAAATGTCTACG CGCAAAGATCGACCC ATTGCCACCCGGCTAT GTGGATCGGGCTCTTG ACTGCTTATCTCCGGC TTTAATCGCTTGAGGA AAGGGGGGTGTCGCC CGAAAGGGTTACGCG ATCATCCGATTCCTAC		QWKETLLTYQRAGRVERAD PTVVDHANSYHWLRKAWIN EKAYQHAVSVMEGTLLEGV NPHGVLTMCRACKAPSASIA HIITGCAELRKSHMKVRHDG VTRWLYNALTEVDGSLPKF HYTQQIPAEMRGERLTVRY DSDIVTPNKPRHNRPDLVVF DSTRKVIYIVEVSVTWLSVL QKQYDNKLNRYAVNSNHEF SESIPYPPGVNLANEIRVLYP QFTGGVKVFPMIISPTGEVH MQFVPHLAELLENPNIPRILE KIQRSVVLGTDYIIRSYFAM (SEQ ID NO: 1504)

				CGTAAAAACGTAGTT GAAGTAGGATTGAAC CTGAGTACCAAGTGA AAAGAGTGGCTATAA TGCTCCATGGTATACC CTAGGGATAACCGTG TCGAGCGCACTGCTCA ATACCCTGATTTGTTA GTGTAGGTATGTCTAG TGGCTGCCTAGGCAG ATTCGCCTATTCACTA CGTAAAATCTGGGTG AGTACCATAAGAACC TCCTGTAGGGCCAGG AGTCAAACTAGTCCA GGTTCGGTATGCTTCG GTATACTCTCTCACGG GCTAGTCACCTAAGA GTTAAGAACCGCCTTT TTCTGCACTGTGAATA GAAAAAGAAGGGCGG GAGAATATACGGGCG AGGTAAGCACGTCGG		

				AACGGGGTGTGCCCA ACTCGTACCGTACGCC CGGACAAGGACCATC TTTCGTACCCCGTTCA CCCGGCACAAGTCCG ATTGTCTCTCCCGAGA GGCGTCGGCGGAGGT TGGCTAACGCCGTCTC TCCAACCATCGGTTTG GGTTTAAGTAACGCCC CAGTGGCCGGTAGCC AAACTGATGGAGGCA GAAATGCCGGTCCGTT GCTAAATGCGGGAAC GGACCAAAACGCCGG TGTGGTGGATACAGG TGGAAGAGTCTTTTGG TCTACGCAAGAAAAG ACCAGGCTCAAGTAC GAATACGACCGTCTTC GTGAACGACGCCGTC GAGACAAAGTTCGAA ATCCTACGGTCGATAA		

				GGCC (SEQ ID NO: 1259)		
R4	R4_AL	U29445	Ascaris lumbricoides	GGGGCCGGTGGGTTT ACTCACTTCTGACCCA CCACCAACGGAACGA GGGAAAGCAGAGCTG GGGCCCTCTTCCGATT GGCATGGAACCGACC TCCACGTGGTGGCCCT GGGCAACGGAATTCA AGAGAGGATTTAATC CTCTCTATCATTTGCA AGATGGATGAGATCG AGGTATCCGGCAAAC AGGTTCCAAGTGAGC ACCTTTCCCATAGCTG GGAATATGGGTTAGG CGTCCTCTGACATATA AGAGGAATCAGACTC GTTCGCGCCCGGTCAT TAACATCGATCAGCG GGAGGGCCGGACTGA AGTAAATTTCCTGTTG	TAGTCGCTAAGGGGTC CGGAAATGGTCCGGTC CTGCGCTACCCGGTTCT GGTAGCACGTTCAAGC GCTCAATCGCCTGCCTT GTAGGCAGTCCATCTGT GGAAGTCGCGCTCTTG ATACAGATGTGGACGG ATGGAAGCAGATGATA GAGCCGGTGACGGCCC TACTAGCCAAACGC (SEQ ID NO: 1383)	MPCSTNSFFERGTPEPHREPI SGTDSSESLGMGTHRSPRLN DDEVINGPKGHESDPVHVV RAPRTLHPRRLELPIGVNNL GEASQLRQDSAIAEEAQLES TENHDGRRPPLRGGRKLWS EKEIATLRRLCEAYGNRQVC WKEVQRKFADFHEERTVAA LATKWGALKRPRAPMVGAP PTPDHDPERGPAGEGDGGTT SQENVPTDDPIPANGPTEGK ESDVRPAVACRCTEPEEQL MESDVRPPAVVRLADPEQH TMKSGVKPVALDGSADLEE RPKEKDIEQMGVDFEGEPRF RAFRKAFYGYFRWAVNSFD REPVKRVRRDCPKVFYAYA DYLIATGSSKALGPNQSRIG RLNGLVYAAARTIHQFWRE EVGHRQQGEKGWYTKTKA TREDLQMLISMMESELARR

				GCCCGAGTGCAGGTG GAGCTCGGACCCGAA AAACGATCCCTAAGA GGACCACAACCCGAA GGGATGGACGCAGTC GCCCCGGCACGTTTGG TGTTGGTTATCCTGGA GTGTTGTGGGACGAA TAGCT (SEQ ID NO: 1260)		KEKRKPGAKELENIHKLVA RLGTRSTSGIVRRLEMTRQR LKLLEDRISLHEQEKRRKRL RKQFAETPSLKLLTKGAKDR GDTMVTMKSVMDFWRPIIG RRVTSNPDQLQVLRDWRDE QKKAYPADLDLEKADLEEK YEGAIRRIQPWKAPGPDGLH AHWWKALPSAKRLLGELV VDWLTTGKVTTGWMCRGR TILIPKKGDRGDPSNYRPITC LNTCYKVLTSVMNSVILSHL SRGEALPMNQRAMRKREW GCTHAMVLDRAMVMDAM AQKKHSLSVAWLDYRKAY DSVSHEYIRWAINSVNIPRSV QLTLKRLMSDWETRFESTQ CRPKLRSDKMKVLNGIFQG DSLSPTLFVLCIAPISYALNK GVGQCQSSSGWSAGYGFEI GHQFYMDDLKLYARTPAM LDSQIQVVSEVSEAMGLHLN LSKCAKAHYAPHGAGGAQE

						AVEGAEGSRKGEIPILGLRST YKYLGVEQRLLPMEVALKE FEDKFMDRAETIFASELTWG QMATAYNTIAIAGLRYVYS NTNGASPKLLEALKRAATL DTRIRDLLRRHKCRFRNSFV ERLYIPRECGGYGLKSVEDT LRESILATWSYIATNPHLAG QQYFFERLAARGKRTPMAD GVKILLDLGVEPQVDLKRRT VTVDGIVFEDPTKLHRYLVG KLLKARTEARIRRWKEASL AGRLVNDTSIDMRLSCLWM KKGFVSARNLRDALAVQEG SLLTRACPALKGKGGQEVC RCCHAAPETAEHITSACRY WLPSLYVERHDSVARNLYY VICCRYGITPVHYSNRVSPLS ENSQCRVLWNMDMQTRTP MKHRKPDIVVFDLKREKIL MFEVSIAHASGLLKQREIKIN RYTVNSEELPDETITPYPPGP NLAADLAATYGWQVEFAPV

						VVGTCGEHVPAVKEDLQRT LDLKPHQVEALLERISRSAVI GTARVVRAHLACS (SEQ ID NO: 1505)
R4	R4_Hmel	.	Heliconius melpomene	ATAAATAATAATAAT AATAATAATAATAAG CCCCCTAAAATCCAAC CATACGTCCGAGTCG AACATCTGATTCTCGT GGGGGGCGGACACGT GAAAATAA (SEQ ID NO: 1261)	TAACTTATTGTCAGAAT TCCTTACTAGTAATAAT AATTATCGCTGAAAAT CTCCACCCAAATATTGC TTGGCTATATGCTCGCA ATTTTTGGTTAACGTAC CCCAATGATTTGGGAG AACAAAAATGGTAAAA CTATAATAATAATAATT ATATTAATAT (SEQ ID NO: 1384)	ITYTANMALVTLFMENMEN KRYNLRPLPGGRRGASGAN AGCHSMRTVGDGGLSRRVP LEKNVTAEQSSSPLTSSSSHS PVSSIPSPSSTRTLLNSPNSSP TSSHSSLVIRSADVVQEALA NYPAPTAGSIRARKKWTDI MNRYIWRTYLIITKCETTLL NNYLEPLHQEFSSKFPEMQV TRQRIGDQRRAIIRNKLLSD DTLAQILIEVKELLQIGDQPL TQNNIHSTQLSHSNTRIKWS NELNEEIVKCYFEVTLLEVN KTSYRKNLYSLFISRNPHLS HLTEQRIADQRRLIFMNKSV HNDRIIELKREVEIKLANSNS LTKNITESNSPSSQTNEINDS AYVQSNLQPVEPLDQHCINR HNLIEKHYVEQEFNNALIQF

						NNTNPETRPYIPRQKSSRKFS QIVSFLNSEVLPKHLNNELD FNALHNIIYTAXYTASLCNG TKFSFIDNYRPRNSKPSWQR RLESRIDKYRLQIGRLTQYIS GNRNRKILKTVEEIKTQYKI HSHHEEPNTELPHFLDTLKQ KLNATSNRLRRYLTCTKRK QQNNTFVNNEKHFYRTLSST NQNTTTQLXEHPTENNLQQ YWANIWETSIEHNADAEWL NKIPDXEINXMKFKDISIETF NQIIQRTHNWKAPGTDNIHN YWYKKLTCTHSLLLKHINQ FIQSPCTLPLFITNGITYMLPK GLDPTNPANYRPITCLQTIY KIITACITDIIYKHIDQNNILA EQQKGCRKNSQGCKEQLTI DAIVMKQAHNKNXNTMYID YRKAFDSVPHSWLLYILKK YKIHPILITFLSSVMLSWKTR LKLINNNETLITDWIKIQRGI FQGDALSPLWFCLALNPLSE

						LLNNTNTGFKLKHNNTYHII SHLMYMDDIKLYASNNKEL KILADLTQSFSTDIRMEFGIE KCKVHSIKRGKSQQNTYILN TGEQIESMDENSTYKYLGFQ QAKQIQQKQTKIELTNKFKF RLNQILRSQLNSRNIIKAINT YAIPILTYSFAIINWSQTDLS NLQRIINTHMTTHRKHHPKS CIQRLTISRLDGGRGLIDIRN LHNNLVTKFRNYFYAKAEIS ELHKFIVNIDNKYTPLNLND RNIQLNQTLITKQQKIEAWS LKSLHGRHLADLSQTHVDK VASNEWLRRGDLFPETEAF MMAIQDQVIDTRNYQKHIIK RPNMVNDLCRRCYSSPETIQ HITGACKTIVQTDYKHRHD QVAAIIHQHLAFKHSLITQA QKTPYYKYSPQAILESTNFK LYWDRTIITDKTVHYNRPDI LLHDKVKXSVYLIDIAIPNT HNLASTFSNKIDKYTDLTIEL

						KSQWKVQSVTTVPIVLSTTG VVPHTLHTSLETLGIHRLSYI LLQKAAILNTCRIVRKFLSSN N (SEQ ID NO: 1506)
R4	Rex6	.	Takifugu rubripes	TTCTATGCGCCTTATG CGACTGGATAGGCCA GTGGTTTACGCCGCTG ACTTTGGTGCGGAAG GTTGTCGGTTCGAATC CAGGCGAGCCCTTAG GCAAGGCTCCTTACGC ATATATGCCTACACCT CGGT (SEQ ID NO: 1262)	TAGAGGACCCGAGTCT GAAGGAAGGAGGCACC GCCCAGGAGGGCGAGG AAGAGATTTTTTTTTAT ATATATATATATATATA TA (SEQ ID NO: 1385)	MSGTXTDRVIPARTSPGSTR SASGVGEPGPPDVKLATGTR HSWSRAENVVLMECYYGSN PSERGYMQRMWEKWVLRN PTSSLTKKQLLAQCSNIRNK KLLSQLEIDEARRCASPTVQI CYGKGEPGRQVSXGVISSSP PNIEIGYKAPMTDGLGTRAA DLRERIMKSWGNSTTSLPRL THKVPDQSLLEDMNTALSTI PTTTITETNQLMYAAATVIL QMLGYKMKSMNSQKEQMA PWRRRLEAKIMATRREVSLL TELSRGVNLRTEXPKKYNK LSTTEALETAKQRLTALATR LKRYTREVEARRINKVFSTN PAKVYSQWQGNKMTTDPPR AETEQYWKSIWEKEATHNT XAQWLQDLQTEHSQLPEQD

						PVVITLADIQTRVSKMKSWT APGPDKIHAYWLKKLTALH ERLAAQMNQLLTSGNHPEW LTQGRTVLIMKDPQKGTIPS NYRPITCLSTTWKLLSGIIAA KISRHMDQYMSRAQKGIGN NTRGAKHQLLVDRAIAQDC RTRHTNLCTAWIDYKKAYD SMPHTWILECLKLYNINRTL REFIQNSMKLWNTTLEANS KPIARVSIRCGIYQGDALSPL LFCIGLNPLSQIITKSGYGYQ FRSGTTVSHLLYMDDIKLYA KNERDIDSLIHLTRIYSKDIG MSFGLDKCGRMISRRGKVIA TDGVELPEGNITDVQDSYKY LGIPQANGNHEEAARRSATA KYLQRLRQVLKSQLNGKNK IQAINTYALPVIRYPAGIIPW PLEEIQATDIKTRKLNGKHKI QAINTYALPVIRYPAGIIPWP LEEIQATDIKTRKLLTMHGG FHPKSSVLRLYTKRKEGGRG

						LVSVRTTVQEETTSLREYIK KMAPTDRLLSECLRQQKPT KEEEPEGLSWKDKPLHGMY HRQIEEVADIEKTYQWLEKA GLKDSTEALLMAAQEQALS TRAIEARVYHTRQDPRCRLC GDAPETVQHITAGCKMLAG KAYMERHNQVAGIVYRNIC TEYGLEVPGSRWETPPKVLE NKQAKILWDFQIQTDKMVV ANQPDIVVVDKHQKTVVVI DVAIPSDSNIRKKEHEKLEK YQGLKEEMERMWGMKATV VPVVIGTLGAVTPKLSRWLQ QIPGTTSEISVQKSAVLGTAK ILRRTLRLPGLW (SEQ ID NO: 1507)

Таблица 3B.

Семейство	Элемент	Номер доступа	Организм	Предсказанная 5'-UTR	Предсказанная 3'-UTR	Предсказанная аминокислотная последовательность	RT	Начало	Конец
	Crack-3_SP	GCF_000002235.5	Strongylocentrotus purpuratus	AAGGTGGTTCAGTTCCTCCGTTTAGAAACAGGGGCTTAAAATTTGCCCATTTGAATATTCATAGCTTAGTTGCTAAGATTGACGAACTTAGATTACTTCTGAGAAATGCATCATTTGATGTTATTGGTTTAAATGAAACATTTTGCGACTCC	TGAATGGCTCAGCCGTACATACACTACGTGCATGTTCTGTAATATTTACATGTCTTATTCACATAATTGCGACTTAATTAATATTTTCCTATTGACATTACTGCTTACTAGGTTAATGATATATGTTAATTTTGCAAACGAGTATTATTTTTGTAACTGTTCTGTATAAATGTTATAAACGATGTAAATATGTTCATTCAGGGCCCCATGGAAGACCACTACTTGCTGGTGAATGGGCTACCCTGTCAAAAGATCAGAAATAAATAAATAAA	MIDDSELYIDDYILLRRDRTRHGGGVAVYISKRFNYKRRDDLISSTLECLWFELHYPNKSPILVGEFYRPPNSHVDFFEELDGNLDSALSEDMSCIILGDFNCDCMPHQLDLNAEKLMFYTNMYGLSQLIDTATRVTNTSSTLIDLIFVSDKELYNEWGIFETSISDHYLVYTVRDFSAKIHNKSEYAEFRSFKNLDEEKFYSELRSTPWQVIKDIDNVDVAWDMWLTLFNNIINKHIPVCKKRIRKKACPWISDEIVKLMTERDKVLKAAKANKSDDLWDTYKYLKNTVIHLMRKRKREYNADQISQNHNNSKQMWRCLRNLVPKSLPISPNSINVDGANVSDNANIAEAFNDFFISNATKITKDLRRNKSQLRVDEDNNIGPTHSQFNLNLITEDFVLKQIDNLSIDKATGEDHISCKLLKMTKNVIAESLCDIINKSLSTGVVPREWKKARVVPIFKSGDISSLNNYRPISI	cd01650	452	710

						LPIVSKIIERAVHQQLSEFLDVNKLLHP NQSGFRPKHSTATALAKLVNQWSLN IDNKEISGVAFIDLRKAFDTVDHELLL HKLACIGCSNRTIYWFKSYLGDRQQV TQFKGAKSHPSTIKLGVPQGSILGPLL FSIYVNSLPECTPEGIIDMYADDTTLT VSANDATVVEEKLTVALNRVMAWIK ENRLVLNTEKTCVMLIGSPANLKKIES FTISLNGNLIKRVQFTKCLGLLIDEELK WSKQIENVSKLTQKNISMIKRAKSFLP QKSLKLLYNSLVLPRLDYCSVVWSNR FHSHTTKLQKIQKRAARIILNKTYDTP SADLFTSLKWMTLDKRFELSRVMMI YKCVHNLAPSYLQVGLINPNDVHEH HTRQIDSGCLRVPRFHTDCYKCSPIVS SIFEWNKLDNSLRTASSVNSFKRLFKQ SCNL			
	CR1-2_IS	GCF_002892825.2	Ixodes scapularis	TT	TGATCAGTTGAAGCG CTGATGTAGTATTTG CTTTTTGCTCTCACAC CTGGTGATGTCTCTG	MRQHNNILNDQGNALDLCFSNTDGI HVARSELSMVPEDSYHPALDVVLSIF DPPVLSPSNGPSPHTSYKYSAGDYTG MYHHLSGVDWASILESSDINGQVYLL	cd01650	316	567

					TTTTGTGCTTTTGATT	TTVVQEAMDMYIPLKTSKKSRFPIWF			
					TCCCTTGTTAATCTGA	SRELRTLLKRKDFLHRRFKKSGLPKWE			
					GTGTTTTTCATACTTG	VEFKICRKRCKKILNRDRQAHSDSVES			
					TTAATATTTTTATTCT	DLVKNPKKFWRYAKSRLGKNIESEIVL			
					CAGACTGGAGAACTA	NQHSDSAKGMSELFADHFASVYINN			
					CGTTCTCCTTCTGCAC	NNCSSSILPHSDVFGDSLGSVVVDEE			
					TCGTGTGTATGCGTG	CVRLCIKKMKPTLSAGHDGIPAAIVKA			
					TTTGTATGTGTGTGTT	YHELLTPILCTIFNNSLTSGIFPDDWKL			
					TTTGCTCATTTTTTTTT	AVVVPVLKSGNSGDVLNYRPISLLSSF			
					AGTGTGCCAGCACTA	SKLFEMVIHQFLSFKFRSIVIPNQHGF			
					AGACCGATTGGTTGT	MSGRSTSTNLVSFMSAASEVVCNRG			
					TCCTGGACACTTTAA	QLDVFYFDLSKAFDVVNHNILLYKLSS			
					TAAACATTGATTGAT	YGVCGSVYNLLKSYLSDRHAYVRVNN			
					TGATTGATTGA	TTSSLYKASSGVPQGSVLGPLLFNIFV			
						NDVSRVILNSSFLQYADDIKLYKKISTL			
						EDCIALQNDAYSFGRWCLDNDLRLN			
						HSKTKIMTYSRKTHDIFFPYSYHGELL			
						SRVSELRDLGVVFDSTLRFDTHVIKVV			
						QSALRTLGAVSRITKEFKSPSAFFTLYC			
						SLVRSQLEYASVVWNGISKTNSSSVE			
						RVQNKFISIFKYRYLENGDPCLNGVC			

						NSQLVNLLSLQQRRVKADLLFLFKTLH GSINSPSLLSEVSLRVPRVPTRLQSSFY IARHFSNLNPIQRSAECYNRYSDTIDIF NSCGSKFGNAVQIFLLQEKG			
	Jockey-1_DSi	GCA_004382185.1	Drosophila simulans	TCGAGCTTTATT GCAAGACCAACC AAGCAGATATTC TTCTCATATCCGA AACCCATTTTTCA GACAGGTCCTAC TTCACCTTAAGG GGGTATGA	TAAAATTAAATTATAT ACATACATTATTATAT ACATCATACA	MPRVSHPPEWFSILIRSSIKFDTQPPI QEDWVQCARISVSTTHGDITLAAAYC PPRFRATDDCFSSLLDSLGPKFVIGGD FNAKHTWWGSRVCNPKGRALCRQV LSRNLNCHGTGTPTYWPTDPAKTPDI LDFAISGGLDPARIQTREVEDLLSDHS AIAVSVYASAMLRSAGKKLIFKTSDISK YHHWLNNSTVTNPPLDTPELIDAAV VNLTSQIQEAATYAAPRFSNQTRDRS RDVHFWNPRVNELKSEKRRLRRVWF LSRNPRDKTALNRATKELKETLVRLRK EALDSFVGNLAPGDPHHNLWNVTK RIKRPLKRIPPIRRQDNSWCRTETERA NAFAEHLEEVFTPANRCSPNEAAETA RLLSEPLFDSESIPQATEEEISNLIAAM GNHKAPGDDAINAIALKLLPPISKSLIT TIINKCLQLGHFPSSWKRAEVVMILKP	cd01650	438	693

						GKHEANPASYRPISLLPVLSKILERVFL TRVLPVIEEAGLIPDHQFGFRRRHGTP EQGHRVAQYILDAFENKKYCLAVML DIKQAFDRVWHPGLLLKLKNGLPQP YFKFLKSFLEGRRFAVRCGEARSDVKE ARAGVPQGSVLGPLLYNAYTADLPVI PSQYIMAATYADDTAFLTTADTPAEA SQTMQEQLDLLNTWHTRWNISVNS DKSTATTFATRRGTCPPVSFNGFPIPE VANPNYLGFTLDRRLTWRPHLLKKRK QAENRVREFYWLMGRQSKLPTSTKL LIYKAIIRPIWTYGIQLWGTASKSNINII QTFENKTLRTVTNAHPFHDNATIHEV LSFPWVKDEIKKSSRQYMQRLQDHP NNLAQDLLDTSQRTRRLHRSHPLDLS GIRHAN			
	CR1-4_MUn	GCF_012275295.1	Melopsittacus undulatus		TGATGAAGGGAGTG GAACACCTCCCTTAT GAGGAGAGGCTGAG GGAGCTGGGTCTCTT TAGCTTGGAGAAGA	MAFDTVSHSILIDKLRKCGLDDQVVR WIKNWLKGRRQRVVVNGAESSWRS VTSGVPQGSVLGPVLFNIFINDLDEGT ECTLSKFADDTKLGGVADTPEGCVAI QRDLDRLESWAGRNLMKFNKGKCR	SSF56672	2	126

					GGAGACTGAGGGGT GACCTCATCAGTGTT TACAAATATGTAAAG GGTGGGTGTCAGGA TGATGGAGCTAGGCT TTTTTCAGTGATATCC AGTGATAGGACAAG GGGCAATGGGTGTA AACTGGAGCATAGG AGGTTCCACGTTAAC ATCAGGAAGAACTTC TTTACTGTAAGAGTG ACAGAGCACTGGAAC AGGTTGCCCAGGGG GGTTGTGGAGTCTCC TACGTTGGAGATATT CAAGGCCCGCCTGGA CAAGTTCCTGTGTGA TGTACTCTAGGTTAC CCTGCTCTTGCAGGG GGGTTGGACTAGAT	VLHLGKNNPMYQYRLGVDLLERVKG KGPGGPGGAEPQR			

					GATCTTTTGAGGTCC CTTCCAACCCTTGGG ATTCTGTGATTCTGT GA				
	Tad1- 65B_BG	GCA_90051 9115.1	Blumeria graminis	AT	TAAGTCCTGGCGAAA TAGAGGTTTTTCAGG ACAAAGAGTTGTGAG GAAGGAGGGAATGC GCAGGGCGCTTAATT TTCTTGGAACTTATTT CTTTATTTATGTTTTA CGTATAGATTTGACA AGCAAACTATGTACT TAGGTTGTAGTGTGT CCCGAGGGAGCATA GACCCGTATCACGCC TATTGGGTGTAGGGC GTAAGCCTGATACGA AATGAAATCAAATCA AATCAAA	MKIAECGVPQGSPLSPLLFLLYIAILLQ DENKNTKFGYADDIAILSIGNTASEAV AEAQKEVDKLVQLASENRIDFDSTKS ELLVIGGGPKKKLDTSGLSVQIQGHSV PTSPYVRWLGVWLDSQLNFKQHVQ EWCGKAQRITQFMRQTNSVQRSAA PGPMIKVVQACVLSTALYGAEAWW PGLTRITTQRGKKVCTGVGWHTDLL DSTILKAVRAALPVWRTTPNMALHR ESGITPSAVILQQRNILAAARIQRLDD WHPLVFRSREDPKDTISRLGLCICQK NRSFKFPEWYLTRLEISAKNVPMAEG PGKLYFPEAESIRKNTGSDFQDEADY AKQWIASLSAETVCAYSDSSSCGPAR SAWGYTLYKGGQVIGSDSGPLLGAE VYDAEIFGAQKALERALGKAGNCPIK VLLDNTTAVRALRTGQTKSSQEAVEV	SSF5 6672	6	148

						FTRLARQARSVEIRWIPAHAGVMGN EDTDRLAKEALRNLPDETDTSPSEMR DRPKYTFAALTRFVRQRCDEVVESW WRDHRPSRYEELDLQMKRKRPPELA LPRWAHHRLITARTGHGDFIKYHERF GHEFENCECVCGREKRPWHFAECRI ALQQWRNLKHEPPPRGREMLAEKG WERFLQFLTVTKCYERQTLHRED			
	RTE-1_EL	GCF_011004845.1	Esox lucius	CTCCTTGAACTG CCACCTTAACGT GGTGGAGGGGT TTGAGTACCCGA GTGACCCTAGGA GCTATGTTGTCT GGGGCTATATGC CCCTGGTAGGGT CTCCCAAGGCAA ACAGGTCCTAGG CGATGGGTCAGA CTAAGAGCGGTT CAAAACCCCTTA	TAAGCGGAAGATGATGCAATGCAATGC	MQPDPRTRKLALGTWNVTSLAGKEP EMVREVERFRLEVVGITSTHGLGSGT TLLERGWTLHHSGVAHGERRRAGVG LLIAPQLCRHVLEFTPVNERVVSLRLR VGDRSLTVVCAYGPNGSAEYPTFLES LGGVLESAPTGDSIVLLGDFNAHVGN DSDTWRGVIGRNGPPDLNPSGVQLL DFCASHSLSITNTMFKHKGVHQCTW HQDTLGRRSMIDFVVVSSDLRPYVLD TRVKRGAELSTDHHLVVSWIRWRGR KLDRLGRPKRTVRVCWERLAESPVRE IFNSHLRQSFDWIPREVGDIESEWTM FSTAIVEAAARSCGRKVSGACRGGNP	cd01650	529	784

				ATGAGAAATTGT GCCGTGACGTCG CCCGGCATGGCG CAGCCGGGGCCC CACCCTGGAGCC AGGCCCGAGGA GCGACATGGGG CCGCCCTCCCGT GGGCTCACCACC CACAGGAGGGA CCATAAGGGGCC GGTGCGAGAGG ATCGGGCGGCA GTCGAAGGCGG GGGCCTGA		RTRWWTPEVRDAVKLKKESYQAWL ACGTPEAADGYRQAKRTAARVVVEA KTRAWEEFGEAMEKDYRLASKRFW QTIRRLRRGKQCPTNAVYSRGGQLLT STGDVVGRWKEYFEDLLNPAVTSSIE EAEDEGSEVDSSITRAEVTEVVKKLLG GKAPGVDEIRPEYLKSLDVVGLSWLT RLCNIAWQSGTVPLGWQTGVVVPLY KKGDRRVCSNYRGITLLSLPGKVYAR VLERRIRPIVEPRIQEEQCGFRPGRGT LDQLYTLYRVMEGSWEFAQPVHMC FVDLEKAFDCVPRGILWRVLREYGVL GPLLRAVRSLYDRSRSLVRIAGSKSDL FPVHVGLRQGCPLSPVLFVIFMDRISR RSQGPEGVRFGDHTISSLLFADDVVV LAPSSQDLQHALGRFAAECEAVGMK ISTSKSEAMVLSRKRVACPLQVGGEC LPQVEEFKYLGVLFTSEGRMEREIDRR IGAASAVMRSMYRSVVVKKELSRKA KLSIYRSIYVPTLTYGHELWVMTERTR SRIQAAEMSFLRRVAGRSLRDRVRSS			

						VTREELRVEPLLLHIERGQLRWLGHLF RMPPERLPGKVFRSRPTGKRPRGRP RTRWRDYVSRLAWERLGVPPEELEE VSGEREVWASLLRLLPPRPGPG			
	ORTE-3_AAe	GCF_002204515.2	Aedes aegypti	CATACGTTGAGT GACTACCGTTGA GAGCGGACGTGT CTGTGCTTGCGT GCTGTTACAGTG GTTCGAGTGAAA AGATCGTTTTTG CGATTGTCGTAG GAGGAATTGTGC TGGGGTTTTTGG GTTTCCGGAAGT GCCGTGTTGGAG CAGTTGGGAGGT TACCGTGCTGGG TTTCCGACGGAT TGGACTGGATCC GTGATTTGGCCA	TGATGCGATTCAGAA CGACTGGAGTTGAA GAAGAGAGGAGAGA AGATCACCATATTGT GCATTGCCAGAATAA AGGTATTGTAAGTAC AAATCAATGTTTTAA ATATGTAAAATGGAA GTCAAACAAGTATAC GTTATACTGGGAAGG TTTAATTTAATGGTA AGTGAAATTATTACA AAACATAATCAATAG CACCAATCGTAATTTT CATATATTATATATCA AACAGGACACCCAAT TGAAGCAATATGGAT	MMTMTFNGIRRLGWFLSVLSLVKQR LLNSSHESNSLSKLILKQICVKLNGVNF TQEDVDELNSIILSKIEGWEQDLNKFI EFFDNIQFEVFDIAEMANHCLLLAVS QIFECEVRVSNDKDDCKIIGEGYKNS WFGVLEIGMESSERVEKVTRALKGSS KIKLGKCGKDRQVNGMAGKGNEPLI VKEECLWYTTIGNWIIEIESKGEDDGK LYSGLMMNVIQKMNEGEIGEQEM MEMLKQIDLFFEMYGEILNRLEEEV MINGNALYEIWKIVSKLAVALNREIIV LMRGGEEIVLGVTERKRKFASFWLVA EFQEVGVRYYSIKSSKLIQEWKENMSI KPAIDVFEKRAEEVSYNLRFKESNQQ AVIKRIDGDGSCLIRAMLDQMGRID QNYYMSDYNVTKIRNNIASYINNSRE RFNVYLYEHIEEVGTIERVIENIREKSY	cd01650	1078	1338

				GCAACTCGGCTG	GAGGAGAGTGGAAT	WLGEEAIITVAELFEKQVEVYTNQGL			
				GGTGAAATAGGC	TGCCAACTCCTTATC	SVRIGVRKDNESALRIFYTGNHYDSVI			
				TTTTCGGAGAGT	GCTAACATACTGTTG	SIAEEEAGEIEAEMININEAEGNKMQ			
				AGTGAGAAAAC	ATAGAGGGATGGTC	AGGESRAVKSNKVREVNESSSHSKKI			
				GAAAAAGCAGCT	CTTAAACCCCACTCAT	TQYGKQSTNTTGDSNNHEQSKKKVE			
				GGTAGTGGAGT	TGAGACATCCTTTTG	KQLRKDSEIKDKENKALKIVSWNVRG			
				GCGGAGGAGGT	TATGATGAAATGACC	CRRKVKRDEIDDILVANDVMIAALQE			
				AGAAGCCTAGTG	GGTAGATCGGTGACC	VNTIGDDLSTNNYSWQVKDNHTNK			
				GAGGAGTGCGG	AGTATCTAGGTTATA	ARGLAIVVRKNQRIRILQAKECRKGI			
				AGGGGGTGAAA	ATTTTTGGAATGCCTT	MWIKLTINGVKLIVINVHAPNSNQKN			
				GTGCGGTGAGG	AGGTAAACCTTAAAT	FLSYLNEIVTRESDRKSLIVLGDFNAQI			
				GCTGTAGTCGGT	AATAGCAAAGAACTG	GSEDLQPEDRVWIGTKLGHRYCNDN			
				AAATTCCGCATTT	AAACAATCAACTGGA	GELFKYFLHLSKLKNVSTKIGKSIEITW			
				CAGATTGCTGAT	TAACATCATATAAAA	RTKDKSSQIDHVLVPQYTSFRIRYIKG			
				TGTTAGTGTCCT	CGGACAGTCTACTGT	CWTDLKTDHVMLTIGLVFTNNPVLK			
				GAAGAGAAGAC	CTAACGGCTGTGCCA	KSNKDTARINVDLLKIPEKQKCFREKL			
				TAAATTCCTTTAA	TGCAACAATGAGTGG	SSIPTKTDIQTDTLDKCYEEMASRLRI			
				TTGGGGCAGCTG	TTACCCTTACATACAT	AAYNTLKVSTVPNTPRRKYALDRLRD			
				GCGTGGTTAAGG	ACATACATACAT	ALKVSKKHPDMINYKYRVANRRHEY			
				GATTGTTCCGGG		QIAVKEHREKEITDFFKNLNDYDAGV			
				GAAGCGGAAAG		RVRKTYNYLKDFVKRKKQKKAMISSR			

				GGAGTGTGAAGT		VWEDILKTCEGPEVEFCAEHDYTPLP			
				GTTTGGCCAGAG		NPPTNDEIKSILFSACNGKSPGSDRVV			
				GAAGTTGGCGG		AEYIKAADEETLERIIEIIKAVYIRNDMP			
				AAAAAAAAATTG		KEWKKSTQIPIPKKSGASEADHFRRIS			
				GAACGCCGAAG		LCNVVYKIYAKWIAKQLRAYAGEPGL			
				AAAGAAAAAAGT		HQAAFTANRSTDDHIFVARRMMEEY			
				TTTGAGGCGAGA		WNSGQDLFILALDISKAFDNVKLQSI			
				GTGGCGGGGGG		GKVLIDLGVPSRLVDRVLSCVKDEVTR			
				GTGTTGCGGTCG		IMWENQYSAEVKRGKGIKQGCPLSP			
				AAAGCTTCAACG		LLFNYLIQDVLKKVVAEIPELKLMGIN			
				GAAGAAAACAC		QLILPLILVFADDILIIARTKAELEKILHE			
				GAGTGGTGTGA		LELQLAKVGLEINRDKSQVLIRYPNDK			
				GTGAGAAGACCT		AQHPKEILLNNKIFKVSDSLAYLGICLT			
				TGATCGGTTTTG		ATLNRKLTCKQRCLNAVKSSRIIVEFC			
				GCGGTAAAGCTC		KTFKPSWELGKLMYKTVIAPSLLYGTK			
				GTTGCAGGAGTT		VSVMTKKSRLKIGNYEKLILRNIYQYC			
				CGGAAAAAAAAA		KKPKDLKFNARKLLDGKTINRRIRVGR			
				AAAAGAAAATTG		ISYFGHVRRREKNHPLKLAYKLKFDKK			
				GAAGCGAGAGT		KEGRPSLTWLKSLDQDFERYQNMNR			
				GAGTGCCACGTA		SSWEELSKDREKIKKKLDEIYKNEESEI			
				GAAGCGAAGTG					

				GAGAAAAGTGG CATCGCACACTG TGAGAGGAGGA AGTGGAGCATCG AAATAGGCAAGC GGCTGAGTAGCG ACAGAGAGAGA GTAAGTGAGAAA GCGCATCTGTAG CGGGGAGGCAG AAATAAGAAGCA GGGCTGCTGGTC GATCCGTGGCGA ATGCGGTGGAG AGCACGAAGGA AGATTCCAACGC AGCGAGACTGA GTGCACCGGTTT CATTGGTGGCAG TATTGGTGAGTC ATTCACTTTCTCG		SDGESSDGDQNSRYKHWKWKRRNE K			

				TTCCAATAATTTT TCCAAGTATCTC GAAGTTTGGTCC TC					
	Proto2-1_OL	GCA_004348135.1	Oryzias latipes	CA	TAATTTATATCTTTTT ACTTTTTATTATGTAT ACTGTATTTGTATCTT TTTTCTATGGACCAC ATCTGAGTCTGAAAT AAAAGTGTGTT	MVVKWGNVLSEPFYVSNGVQQGGI LSPFLFNVYIHDLSTNLNLCKTGCVVG CQTFNHLMYADDLVIFSPYSAGLQTL LKVCSDYGRDFDIKYNPSKSKVMIVR TKDDRNVNFPVFKLCGTTLEQTSSIKY LGHFITDDWRDDRDIQRQYCKLYAQ ANMLMRKFSACSLQVKCSLFRAYCT PLYTAPLWWNYGRGTFHKFLVAYND AFRLLLQVPRWYRASQLFVDAHIPTC EALLRKLTYSLMERLDKSKNSIIMSLT DPKRSSCRYTSALRKHWIQRLYSF	SSF56672	13	208
	Penelope-2_Sln	GCA_010367695.1	Solenopsis invicta	AAAGATCAATGA CCAGCTAATTAA AGACTGGTTACG AACCACGGGAA GATTCGTTAGGG AGCATCCTAATA	TAGGTTCATCACCCC TTCTCCTTCTATTCCT TTCGTTCGGGCCACC GAGGCTATCTCGTTT CGGTCTATATTATTTG TATCTATTCTCACCGA	MEGMLSDESTYVKLSKDPIQKLTKEA RSLLSIWKDKGFIEQHIYRSLLTTDGVL PRAYGLPKIHKPNNPLRIIISSINSPLHS LAYFLHNIIYNSIPEADSFVKNSFHLVN KLNGLHVESDHVLVSLDVVSLFTNVP KNLAVESVTTRWDLISKNTTVPVNEF	cd00304	189	269

				TTTTGTTTACCAG AGCGGACAAGG GTAATGTAACAG TTGCCCTTGACA AGAATGATTATA TCTCGAAG	GATGTAACATCCCAT CTCAAGGCTGTGACA CGAGTTTTGGTGCTT AGAGGTAATTTGTAT TTGATTTTCGCTTTGA CTTATCTCCTCTTTTT CCCTTTTTTGTTGAAT TTACATTACTTTATAG AAGTGTATTTTTTACA ATTTTATATCTAATGA CTTGTACAGCTTCCAT TTTTATTCTTATTTATT TATTTATTGAATTTTT TACTTATATTCCGCTT ATTCTTCTTTTTTAAC TATTTAATTTTCATTG TTGTATGCCACAAGT TTTATAGTCATCATTT TAACATCTTTCTCTTA TTTATTTATTTATTTG TTAGCAACCAATCGA	LKALNLIIDSTYFKFNNTCYKQIFGLP MGSPLSPILADLVMQDLECKALKLLS FHLPFYCRYVDDIILAAPANSLSGILDT FNSMHDRLQFTMEREIEGKISFLDV MLINVNNRIIFDQYRKPTFSGRYLNFF SHHPLCHKKGVIFGLVDKLLLLSHPQF HCKNFINLVNILLENNYPLNFIFSMIN SRLKHLFYKDDTYVEDKRVRRFFTIPYI ESVSEKFSSFSCKFDCSVAFTIPSKLNK FITTGKDHTERLSCNDVVYKINCKDC DSSYVGQTKRQLKTRVSEHKADIRKA SSPSVISLHRIGENHEFDWENVKILDK EPSYVKRSISEMLHIKNQMHGLNKQS DTELLSDAYLPILERIPPS			

					TGTTGTTTCACAAAA CTCTTGTCTCGGCAC ACGCTCTGTGA				
	PENELOPE	GCA_007989325.2	Группа virilis	GGTCGCCAGAGC CATCAATAAATA TCAACGGAAGGC ACGCCGTATGCA CAGCAACCAAC	TAAAATAATGTATGT AAAATGTTCGAAATA ATGTTTAATTTATTGT ATTATAATTGTTAATT GTTTTTTGTATCTTGG TGTTAGTGCCCTGAA GACGGTTTGCCGATG TGCAACCGAAATATA TCGGAAGAGAATTGA ATAAAATTGTTTTTCA TTGTTTGTTTTAACAA ACACGGACCTCGAGC CAGCCAACAAATAAA TATTGAAATATGGAA A	MSYAKIKTKYKDSKRTINKFQLTLVKL TKLKSSLKFLLKCRKSNLIPNFIKNLTQ HLTILTTDNKTHPDITRTLTRHTHFYH TKILNLLIKHKHNLLQEQTKHMEKAK TNIEQLMTTDDAKAFFESERNIENKIT TTLKKRQETKHDKLRDQRNLALADN NTQREWFVNKTKIEFPPNVVALLAKG PKFALPISKRDFPLLKYIADGEELVQTI KEKETQESARTKFSLLVKEHKTKNNQ NSRDRAILDTVEQTRKLLKENINIKILS SDKGNKTVAMDEDEYKNKMTNILD DLCAYRTLRLDPTSRLQTKNNTFVAQ LFKMGLISKDERNKMTTTTAVPPRIY GLPKIHKEGTPLRPICSSIGSPSYGLCK YIIQILKNLTMDSRYNIKNAVDFKDRV NNSQIREEETLVSFDVVSLFPSIPIELAL DTIRQKWTKLEEHTNIPKQLFMDIVR FCIQENRYFKYEDKIYTQLKGMPMGS	cd00304	470	558

						PASPVIADILMEELLDKITDKLKIKPRLL TKYVDDLFAITNKIDVENILKELNSFHK QIKFTMELEKDGKLPFLDSIVSRMDN TLKIKWYRKPIASGRILNFNSNHPKS MIINTALGCMNRMMKISDTIYHKEIE HEIKELLTKNDFPPNIIKTLLKRRQIERK KPTEPAKIYKSLIYVPRLSERLTNSDCY NKQDIKVAHKPTNTLQKFFNKIKSKIP MIEKSNVVYQIPCGGDNNNKCNSVY IGTTKSKLKTRISQHKSDFKLRHQNNI QKTALMTHCIRSNHTPNFDETTILQQ EQHYNKRHTLEMLHIINTPTYKRLNY KTDTENCAHLYRHLLNSQTTSVTISTS KSADV			
	l-3_SPur	GCF_000002235.5	Strongylocentrotus purpuratus	GGTAAAAAGTCG CCTCATATAACTA ATAAGTCCATAT CAGACCTATTGA TTGTACAATGGA ACGCACAAGGCT TGACGGCGCATG	TAGCTTTTAAAGCTA TCAGGGATCCCCTTA ACTTATGTCGATCCT GTAATAATATATAAA ATGAAGTATTTGCGG TTTTTCCATGAACAAT TGCAACTAACTCCCA	MCIQETWFNDFNVLSIPNYVCVCKN RPDGRRGGLVMYLKNDITFQAMDT PRNAEYQKVDIFCKNGVISIFNYYNPC KKIELIDLNNMIRNVPGKYLILGDFNS HNPLWGSNKLDTNGRTVESFIENND LVLLNDGSPTRLDPYYGTYSNLDLSLV SKELARNTEWKVLSNNFGSDHNLVC	cd01650	462	717

				GCGCAAACTTTT TATCTTCCCTTAT TGATAACAATTT GTTAGGTATTCA CCCAGATATA	TCTCGCCTTTTATGTG CATTTTATGCAATTTT AATATTGCAGCTTTTT TTTTTTTTTTTTTTTTA TAAAACCATATCACT ATTAATTTCATTTCAC CAAGGAAAGTCGCTC GTGGGCTTGATTTCC GAAGCCCACGCCAAA GCTATACACGAAATT GACTGGCGCAAATAA TCTTTGAGATTGAAG AGCGCCATCTAAACG AAGAACCAACCAACC	LNFKHVYNRPGGENMGQRWSFRKV NWEEYREEIEEALRKEGYENLGAELNI QGRYNVIVKVIKETSNKCIPRVQIAKR RKVNPWWNNACKKAVQERNRLKN RLRRYYDLNDVKKYLKKKAVAQKVIR RAQYDYWLEYCRRMNRFTDIGKVW EQVKRMRDSTMNKHIFHNLFVDGNI LKESEEKANAMATYFKSVDESLSVTK NENENILYNNSSTNDDDCINDDFNM SEMGNVLKEMKNSSSGKDDIYLIMIK NLPNKALVYLLNLYNDIWKISLFPHQ WYEVIQIPILKSGKDPTQVTSYRPISLI STFSKIMEKMIKNRLTWFLEKNRLLN PIQSGFRKYRNVKDHLVRLETDIHISL KSKKYTVAVFLDLEKAYDSLWKTGLL RKLSILGIGGKMLLFIKAFLLSRISNVRI NNYMSRHYTLENGIPQGSSLSPLLFTL MLNDLKINVSDVKLSCYADDIAIWCS NKDLELARRKLQEALHELEIWCWTW KLNVSPTKSAVLVFTNKQKKDILLQIK GHNLPICEQYKFLGVRFDSNLTWSVH			

						IDSIIDKCKKRINLLRCVSGTPWGSKYE ILMMIYKGLILSILDYGSEMYDSACISL TKKLDSIQYRALKIITGTIHNVSLEALQ VLTGELPLRYRRRLLNDNFRVYLLSSTL NHPTKEVLGDDYDLNNLEWKIGEGP FKLRSFRDEVVVEGQTNLLPKIPEWR LARPQVSFEHHGKVQKKEMPPFEIKS IVLDSIGNSWNGFLHIYTDGSKLEDG KTGAAFYVSHYNVVKFFRLSHVCIMR AELVAILMALEWINTFIESSSVVIFSDA LSALQMIDSYFIRCAIVNEILSIINILQF KDVCVAFEWIPSHCDIRGNEIVDRAA KKGAMKRCIDIVIPNTRSELAASRLIH YNKLWQAEWDNAEHGRFLYALHSD VSKPIHRQGLSRREECLIQQLRVGNCS LNYYLFQINQHDTGLCEVCQEDETIK HFIFYCTKYDKDRRLLKERIKLNNPNL KQLLMSNPIFTRELLGYIYDTNRFR			
	L2-3_XT	GCF_000004195.4	Xenopus tropicalis	TATCCCTCTGCTG TATACTGCCATAT AACCCACTATATT	TAGCAATACCCTGTG CCCCACCTCTCACAAC TCTGGTCATGCCCAT	MFYPALLLFALSCFLKHLQPHPPSSQK GITITASLLSSPSLSAHEVFKLLCSITPV NSFSSHKTKTYKSRSHLAFLSFLLLAAG	cd01650	702	964

				GCAGCCTCTTTAT	TCCCACACCTTGTGTC	DISPNPGPYPIPVLLGTRSPNPPCTPL			
				TAATGGTAATTC	TCAACCCTTTCTCCTT	KYRTLTPIPKASIQISCALWNSRSVCN			
				TCATCTGTGTATC	GTAGATTGTAAGCTC	KLTSIHDLFIAKSLNLFAITETWLSPTD			
				CCTCTGCTGTATA	TTTTGGGCAGGGCCC	TASPAALSFGGLHLTHTPRPGNRAG			
				CTGCCATATAAC	TCTTCACCTCTTGTAT	GGVGLLLSPRYSFKVLPPPPSLSFSSFE			
				CCACTATATTGC	CGGTTATTGATTGCT	VHSIRLFSPVSLRIAVIYRPPGPTSQFF			
				AGCCTCTTTATTA	TTATATGTTACTCTGT	DNFSAWLPYFLHIPTSWTDLTILDSFS			
				ATGGTAATTCTG	ATGTCCAATGTATGT	PPAIHLKSFHLLLLFHFPHLRFTLLGSS			
				TCATCTGTGTATC	AACCCACTTATTGTA	PLFRFALRSYTDLLDRPHNSLTTFLPG			
				CCTCTGCTGTATA	CAGCGCTGCGGAATA	FLTSFIYRPPGPTSQFFDNFSAWLPYF			
				CTGCCATATAAC	TGTTGGCGCTTTATA	LSSDSPSIVLGDFNIPVNNPNAPATSK			
				CCACTATATTGC	AATAAATGTTAATGT	LLSLTSSFGLSLCSDSPSHSNGNSLDLI			
				AGCCTCTTTATTA	AATGTAATGTAAT	FTKCCSTSNFTNSPFPLSDHNLLTFQL			
				ATGGTAATTCTC		SLTPSTPSAVPQTYTYRDLQSLNTSHL			
				ATCTGTGTATCCC		SSSFDSLHSNILTVSCPNQATSVYYSTL			
				TCTGCTGTATACT		TTALNELAPAKTKRMRPKPLQPWHT			
				GCCATATAACCC		AHTKSLQKQSRTLERRWRKSRSESDF			
				ACTATATTGCAG		CSYKSALLSYNTTLCQAKQTYFTALIN			
				CCTCTTTATTAAT		SLSKKPAQLFSTFNSLLSPPPPPPCASV			
				GGTAATTCTCTC		TAQAIAEHFKNKIDTIRSDIAQLNPHN			
				ATCTTGTGTATCT		HSPPSLHAPQSLLGSFSPVTEEEVSKL			

				GCTATGCCAATA		LSSSHLTTCPLDPIPTKLLRNANPCLIK			
				ACTCATTCTATTA		ALTHLFNLSLSTGIFPSQLKHALVTPIL			
				GGCCTACCTTGA		KKPSLDPSNPTNLRPISLLPFISKLLERL			
				TTTAGTACTTGCC		VYNRLTLFLSDNNLLDPLQSGFRQQH			
				ATAGCTGTGTCA		STETALTRLTNDLLSAKAKKHYSLLILL			
				AACTCACATCTG		DLSAAFDTVDHPLLLQTLRSLGLRDTA			
				TGTCCACTCCTCA		LSWFSSYLSDRSFSVSYNGESSSPLPLS			
				GTGCCAATATTT		VGVPQGSVLGPLLFSLYTSSLGKLISSF			
				CACAAGTGCTAG		GFQYHLYADDTQIYLSSPDLNPELLTR			
				GAAAATTGCAGT		VSSCLSAISTWMSQRYLKLNLSKTELV			
				TAAACAAGGCAG		LFPPSNAHIVPEVSITVNNSTITPSSQA			
				TTCATGGAACTA		RCLGVILDSALSFTPHIQSLIKSCHFHL			
				ACTCATATCTCTT		RNISKIRSFITQDAAKILIHSLIISRLDYC			
				CTCACTTCTAATG		NSLLIGLPLQRLSPLQSIMNTAARLIHL			
				GTAATTCTCATCT		SNRSSSAMPLCQSLHWLPLPSRIKFKL			
				GTGTATCCCTCT		MTLTFKAVHNSAPPYISELISRYHPTR			
				GCTGTATACTGC		LLRSSTDLLLNSSLIPSSHARIQDFARA			
				CATATAACCCAC		APLLWNSLPQTVRLSPNLSAFKRSLKT			
				TATATTGCAGCC		HLFREAYPNLV			
				TCTTTATTAATGG					
				TAATTGTAACTC					

				ATCTTGTATCTGC TATGCCAATAAC TTTTTCTGTTAGG CCTACCTTGATTT AGTACTTGTGAT AGCTGTGTCAAA CTCACATCTGTGT CCACTCCTCAGT GAGCAAGAGAT ATTTTTACAGCCC CAAAACTTGGTT CTGGTGCTTGTA AAAGTCTCCCCC TCCCCACACAGC CTGCTTACTTTGA TCCCTGATTGAA CTGATTAATTAG TCAATTAGTTTCT GGCTTTCCCCAA TCAGCTGCATAT AAATTCAGATGC					

				AGCCTTGGGGAT GGCACTAGTGTC AGGGGCAGGCA CTGGTGCTACCA TTTCCTAAGTGCC AACATTTCACAA GTGCTAGGAAGA AGGGAATGGCA GTTAAACAAGGC AGTTGGACAAGG CATTGACAAGGC TACAAACTATCTC CTATCAACACTG AACAACGTGGAA TCGCAAGAAAAC AACTTTGCTTATT CATGGTAGTTTG TTATTTCCTATCC CAAA					
	L1-1_VVi	GCA_011037095.1	Vitis vinifera	GGAGTGTTCAGT GAATTATTGTGG	TAAGGGTGTGGATA GGGTGGCTCTTCGTT	MKLRVLSWNVRGANDSEKRKVIKALI KSQKVDVVCLQETKMQEMSRMIVR	cd01650	496	763

				GACTAGAAAGA	CCTTGGCAGGCTGCT	SLGVGRCLDWKVLNSRGSSGGVLVF			
				GTGCGGGGTTGC	AGGGAGCGTGGGGT	WDKRVLHLLEAEVGNFSVSCRFKNCE			
				TTCAGATTGCTA	GTTTTTCTTTCTTTTTC	DDFCWLFSGVYGPSLMKEREDFWAE			
				ACTGAAGTTTGG	TTTTCTCCTTCTTGTA	LGAVRGLWSDPWCVAGDFNVVRFP			
				GTTTGTTGGTCC	CTTTTAGCCACTGCTT	VESSRGGRLSASMRRFSEIMEDLELR			
				AGTGAGGGGAG	CGGTGGTTTTTAATA	DLPLQGGSFTWKGGLNNQSHSRLDR			
				AGGGGCTTATTA	CAATCTCTTTATTTAT	FLISNEWEDHFSGSVQYVLPKPTSDH			
				AGATGGATGTT	CAAAAAAAAAAA	FPILLDGGGVRSGPMPFRFENMWLK			
						EEGFKEKMQGWWVGLNFSGSASYV			
						LVSKLKALKSLLRDWNKLEFGKVEVN			
						KALALSQVDFWDKMELSRTLSVQEV			
						DARRGAKEDFKKWALMEEISWRQK			
						SREVWLREGDRNTKFFHKMANAHR			
						RGNMLSRVKINGVWLTKENEVKEGV			
						VNEFKAMLSSAGGWRPSLRGLSFERL			
						EAVDAASLEEPFSEQEVMEALKGFCG			
						DKAPGPDGFSMAFWQSSWEFVKEE			
						VLGFFREFHNHGRFVKSLNATFIVLIP			
						KKGGAEDLRDFRPISLVGGLYKWLAK			
						VLANRLKRVVGKVVSKAQNAFVQGR			
						QILDAVLVANEVLDSVLKNKEEVVMC			

						KLDIEKAYDHVEWSFLFSVMRKMGF GEKWIRWMKWCISTVSFSVLVNGSS SGFFQSSRGLRQGDPLSPYLFVLVME AFSSLLRKAVAGGFVSACKARSRGGE GVNVSHLLFADDTLVFCGASKVQLLH LNWILMWFEAMSGLRINLDKSELIPV GCVNNVEELAAAIGCKVGSLPTSYLG LPLGAQYRSRAVWDGMEERMRKKL ARWKSQYISKGGRITLIRSTLANMPIY FMSMLSMPRKVRLRLERIQREFLWG GGALERKIHLVKWELVCLEKDNGGL GVKSLSILNKALLCKWSWRFAMERE AFWNQVIRGKYGEEQGGWSSKEAR GETHGVGLWKTLRKEWEVVKSRLVF VVGNGKRIKFWKDIWCGDEPLCVSF PSLFALAVSKDAWVKDVWRCNEGG GSWSPLFSRPFNDWELEEVCSFFVAL NRKQIQQGVDDRVIWRETNCGKFSV KSLYKSLVSGHPISFPSSAIWKVTVQP RVSFFGWEATWGKALTLDQLQRRG WALANRCYLCQRHEESIDHILLHCDK			

						VRTLWVLLYSMFGVQWVLPATVKET LSGWNGSFVGKKRKGVWKASPLCLF WTVWKTRNKVAFEEEELSIQRLKASF VYFLWSETKRSIKDGPSTLVDFVGWV ASR			
	R2-2_SMed	GCA_002600895.1	Schmidtea mediterranea	GTCATAGGGTGA ACTGCAATTCTG ACACGATGACCG AGCTGTGTCAGT TTGCAGCTAGTC GCTAAAGACTCG ATCAGTCCGCCA AGTGAGGTGGCC GGGTATCTGCAG CACTAGAGCCAC TGGTATCAAGAG CAGAGATACGCG AGTGGAAGTTGA GTACGACTACCT TCACGGGGTCCT CCTGATAACCAC	TAGAAGGGAAACAA AGGAAAAACGAAAT GACTGGTAACTATGA AGGATATAGCTGAAA GCCGAAAGGAAGGC TAAGTCCTGAAACCG ATCTACATCTTCGATC CCAAGAGGAACTGT GGGTTAAGCTTGAGC CGACGGAAAAAGCG AATGCATGTTAGACG ACGAGGTACAGTCAC CTCCTCGTGGTATTT GGCGGGCAATGCTC ACTAAATTAACTGTG AGTAGCTGAGAACTG	MRKELVTIKNLIEESGATAPAPVPLEV AVEVHQSSSVPEITDQSTTTQEGSYSE PPIHRCENCGREFRTRAGVQQHRRK AHTNEFMEEKEKAAPTKKLRWTDEE KEILIESEIKIIKEGSLKEQHEINKILASR MPGRSQDGIAKIRQKQEHKAEIQRRL HGTVTTDETRGSRTSEITEPIRSLPINT KNWSEDELKRMLAEEVKLRNKNEKD INKKLAEIFPNRTMGSIKSKRTKDKEY QDLVKLTMQTISENPDNETDFNTSN TENNSTDAEKEVKNYLNMLLLSNNEE EWLTSTLKEAATLALQGKKTEASEKL NEYASKTLFPGLKITNQTRKREKKISKR ETRRQEYAEIQKLYKKNISSAAEKAIN GKWSIKPEEEYHNNKDLIKAWKPILE APPFSDCRPIENIKEMDYALMEISTAE	cd01650	474	730

				AGTGGACTGTGG GAACTAAATGTG TGCTCAGCGTTC CCTACTTTCTCGT AGGGTAAAGGG TAAGATAACCCA GAGAATATCCCA TGGGAGATATCC ATGGAAAAAGCA CCACGTTAGACA ATCCGATGGTCT AACTCGGCTCCG AGGGGCTAACTA TCCCAAAGGGCT TAA	TATGTGTATCATGAA AAAAAAA	IFLAIRAMGKTAPGPDGIKYSKLKKNI QSMAILFNTCLLTSFLPLPLKIARTILIP KQENPGILDYRPLTIASVVTRVFHSILA KKLDNNAQLSQRQKGFRKCDGVAE NIVILETILTNSRSEKRPLCMAFVDLRK AFDSVGHESIIRGAKRVGVPPMLLEYI SSSYQNASTNLFGEILNSRRGVRQGD PLSPILFNFVIDEALENLNRNIGYLLKE EKVSCLAFADDIVLIAETKGGLENHIE KLLEKLNGAGLELNASKCATLMVMK NGKEKSTYISTKVIKIKENDIPTMKATE TYKYLGLQMGFKAREQNANEVITEGL ENITRAPLKPQQRIHILRDFLIPRLIHKL VLGRVAKKSLKRIDQNIRKKVRNWLH LPKDTTAAFIHADAGDGGLGVPALEH TIPLLKRERITNLRKSNDPVTKECLRM EYTKQVLGKWSRPTKIGETLATNKSQ LKEAFRKQMLITLDGKGLKDHHETPTI HKWIRRGENMTGKQFITAVKIRGNL VATKSRNSRGRPEQEKLCEAQCGRP DSLGHILQGCWRTHGMRVERHNNI			

						CRRIKAIMKGKESEVVEEPRLQTNEG LRKPDLLICHKGKIIICDAQVVADSSN CSLESENQRKIDYYKKDSVVAEARKLI GRDDEDIIIMAVTFNWRGAISKTSIRD LDMLLDIKSKEVIKMSRKIIRDNSIMV EMHRNRTEKRR			
	Utopia-1_NVit	GCF_009193385.2	Nasonia vitripennis	CCATTCCTTCGTACGGGTTTTCGTGCCGGCATAGCCGGGTGGGAGACTCGCGCGGGGGAGGTCATATCTCACCACCATCTTCGAGCTTTTTGCTGCACCTGA	TAGTGGGGCCATAACACCTAGGCCCCACAGTGTGGCGATGTCCATGTGTGTCGTCCTTATTTATTTATTTATTTGTCCTGTCGAGCTGTTTAATTTATTCACATGTATGTGTTGTGTGTGTCGTCCCTCGTAGCAGTTTTATTCCGTCCAACCAGAGGTCGACCAAGATTAAAATAAGCATGGCTTGAAGCAGGCCAAGCGCCATGTTCTAACCCCGTTTTAGGGGAA	MSGLTGREVKCITVNVLMEQQPHTKAIREGDFIVILLPQSDDETLCCPLCVGRGRYSGKTRVECLNRHVKEVHPDLTTTFRCWGCGFAAPGDKKYPRKIVTQHCATCVPEVSSAPSGRVDGERRVNTRRRLGIAAATEASPVRRTRRNGLASPPVEQDISQSAAPPEPARVPQHPEIVALGESADDEVFRSPVNSPPRDWRAAAPQQAASSSPPAVPGITAATPSNTTRTGNGSANSILAEHPIPTPPPTNTTEANGRADIPRSGRTPPPGQQAARRRAPTTEQRRIVGLLEAATGREQLEEATTQAMLFLARLTGRRPEPRNAIRPGPRQRHPAQEDVQAQAPDRIQEAKKLQRLYRTSKKRAVQKILAGPNMNCQIDKNTITAHF	cd01650	477	740

					GTTACTTAACCTAAA AATACAACTTTTCC	VELAARRDGGEDWPDVFDREEPTAA SGEALCTPITREEVFRRLKGRNNTSPG PDGITYRDLAKADPGAHVLAALYNSI WRIEATPALWGVSNTTLIYKKGDAM DISNWRPISLGDTVPKLFAAILADRIK RWAVANGRYSASQKGFLEFEGCYEH NFVLQEAIREAKGGRKELVVAWLDLA SAFTSVPHSSILQALEGHGLPSKARNII SSLYTGMTTRFHTAEGPTDPILIQSGV RQGCPLSPDVFNLTLEVVLREIQRTGE GYTIEGRRTSHLAYADDVAILADSPA GMRRLLFAAERGARAVGLTFNPAKC ATLHIAGRGEEAVRPTEFSVQGTPVR ALASEEAYEHLGIPTGYQVRQTPINTL RDLLADIGSIDRSLLAPWQKLDAVGT FLLPRLDFIMQGAHIDKGFLTEADKIIK KAAKSWLSLPQRASAELVFLPPSQGG GGLLTVAHSYKMLYSSDVTVSTIAGS TLRRTVSERLKKRASNIDIARFLSGDLD LPRSTSPSTFWTKVRSAALRIKTKLGL RWSWCQEGEVLLMACGDPRAPGTR			

						VSPQTKHLVTTSLRRCLNRHYAESLLA KKDQGKVFEVTRRSGQSNHFLRSGSF TRFCDWRFIHRARLDVLPLNAAKRW QRGMDKRCRRCGSDLETLPHVLSHC GPHSAARQKRHNNIQDRLVKAASRC PGTISVNQTVVGVRGPDAALRPDIVV RDDVNRRVTIVDVAVPFENRLEAFD GVREAKIAKYTPLARQLTDSGYTVTVE AFVVGALGAWDPRNERVLSLLSISRY YAILMRRLMVSDTIRWSRDIYVEHVS GIRQYRE			
	LIN9_SM	GCA_00260 0895.1	Schmidtea mediterran ea	AAACGACATCAT GAACGCTTGGCC GCAACAATCCAG TTATCCCTGCGG TAACATTGTGGA ACTCATAAGACA AGTACTAAAAGA AGAATTAGAAAA ATTAGAAGAAAA AATTGAAAATAA	TAAAAGGGCAAAAA GATATTTCAAGATGA ATTGTGGACTCATCT AAAAAATGACCACCT CGAGCCCAAATAAGC CTAGCTATCATGGTT GCTGATGGAAACAGT AAGGCACCTGATAGC TAACTCTTCACCGTG GGTGTCTTCAGACAC	MMDSRQLNTPKIRKYQNPKMTNDI MKSYNYAVLSDVTPQETTQQTTHLN VDIDKETTQPKQPLTKSGKPKSKPIAV SYKFKDATFIWDTTPQTNPPKDCTKLI DKTRPRKTIFKKSAFQTYLKKELSNETF VEVKTFLMATHKYRFKDENSRLLAYRI INRYVMETANEFKETEFDMARFAKFF TIPENWLKHLKPYSTATETSPADRIKV QKLVDLTCRYPFKTQEEQTSVANFLH FFTQRSIIGISRDYKFQKCIPFMARKT	cd01 650	793	105 7

				TTTATTTATAAAA	TCACGGTGACAAGAA	NGQETTSTINTTPPTEENRLPMVIITP			
				TTTAAAAATTTAA	AGGACACCACTAGTA	LEEPKIEHRRPEKRGASNDTIVLSDEE			
				ATTTAAATTTAAA	AAAACCACTAGTTTT	FPLLKRRTLPTRKSKNPTGAGNVPTET			
				TGAAGATAAAAA	CTGACACCTCTTGCT	ECTDEVKFILNNEYQIECKECGKVWE			
				TTTATTTAATCCA	ACAAACTCTGTAAAA	NVRNGMNHLRQKHDFPNRTDVMV			
				ATAAATAATCAA	ATCAAAAGGATCGAT	SCVRCEVPIKGAECVNHIKNHKKDDK			
				GAAAATCAAGAA	AGGCCGCGCTTTCAC	EESQAGSLVANTQDIPNESSLSQAAIE			
				A	GGTCTGTATTCGTAC	VYLRNILKMKENQERNIQYLEPSTAN			
					TGAAAATCAAGATCA	FLINRNLRAFYQNVKIEKLIGWEQVI			
					AGGAAGCTTTTCCCA	WLIHWNKCHWIVYLANCDSKTSVIL			
					TTTTAGTCAACACCA	DSDNQMTLQQRCNIKAKFDKFLEGT			
					GGTTTCTGTCCTAGTT	FEEKAVLGTLERNVPQQPNNFDCGV			
					GAGCTTCCCTTGGGA	YVIQYISDFLKDPQRMDYHTPDSKRIR			
					CATCTGCGTTACCATT	KEIGELILEEMKNPASKIKNPNKEIQSL			
					TGACAGATGTACCGC	LQKFRLLQINVNDVFHWFAEEYQKSL			
					CCCAGTCAAACTCCC	PKIRTKRDGKLNKLSCSYQIQRLFGLA			
					CACCTGTCCTCAAAA	PKRAVKEIYFQETSTADLETRVLNEHF			
					CAGTTCAATTGCATC	KKDESMMKECKIKNGNHYQDWITK			
					CAAAGATCGCAATTT	AQIDNKEILEALKNSTDSAPGEDNIPL			
					TTTCACTAAAAAAAA	RQWIIWNNDGVLFDMFNYIKRTHDI			
					TTAACAAAAGTTAAT	PEMWKNYTTTLLIKPGKSQESNIPAN			

					TATACTGCTTCATTGA GTAAGTAGAAAAACA ATC	WRPISILPTSYRIFMKVLNKRVLDWA NRGELISKWQKAVDKANGCDEHSYV IQALIEKANRSYYKNEQCNLAFLDLAD AFGSIPFQVIWHTLKNMGMDEKTIN LLKEIYKDCSTKYKCGKNESEKIKITKG VRQGCPLSMTLFSLCIQYLIQGIAEKK KGATIAGQEVCILAYADDLVIVANTA KDMQMLLTTIENLAKQADLIFKPAKC GYYRDPRDKKSMMKIYGKEISIVDEK NVYTYLGVRIGDTKKKDLNVRFEEVK KKTTAIFKSKLRSDQKLEAYNIFCQSKF VYILQGEDIAKTKIETYDEEIKKMIKEDI LKLQDKSPFTDFVIYSPREKGGLGITKII DEQIIQTINRTAKLLNSSHRAIRAIIYEE LIQVANLRGEKEINSIEEALKWLEGTN KYKKNSNAKTTWITRVREAFQTLEKK HKIKVRFVPKENGIGYKIKCDTQEKVE ELDNSKELSKSLHWMIKEAYFKEWKT LKCQGYILSLKTSEFMEWKMPRGLPD PDWRFLTKVRANMLDINMKQANQ GGRLGNTKCRKCEDKESASHVINHCA			

						SGNWSRVEKHNQVQNELAKELTKR NISFEKDCIPKETKGTLRPDLVIRLKDKI MIVDIKCPFDEESAIESARNKNIDKYR ELAKEIQAKTVLPTTISTFVVCSLGTW DKRNNELLRQMGIRYEESKEMRINM IHKAIHGSRKTYDQHRNFNNG			
	R2-1_TG	GCA_012069535.1	Taeniopygia guttata	GTCTAGTTACAA CTGGGCATCGCT GCAGAGATCGCA CCTCCTCGTGGT CCCGCTGGTAGC CCTTCGAAGGGT GACTAAGTCGAT CTCTGCCCCAGG TACGGAGCCGTT GGGACTCACCAG TCCAACGTAACT CCTGCCTAAACT CGGTGAAACAAA TTCCTCGGTAAA AAGCCCC	TAATTCAGGTTATTTA GATGCTTAGTTTATG TACCTTTCTTGATTTG TCTAGGATTTTGATA GTGTTATTATTTTTAT ATTTTTGTACGATTGC ATAATGTTCTTTTTTA TACAGCTCTGTTTTAA TAAAATAGACGATAC CTAGAGACGTTAGG GCAGCCACAAGCCAG TTAGGTAGCGGATAG TAGGTAGGAACAGA CTTTTACTATTTCATA ACGCGTCAATTACCA	MASCPKPGPPVSAGAMSSESGLTTH SVLAIERGPNSLANSGSDFGGGGLGL PVCLLRVSVGTQTSRSDWVDLVSWS HPGPASKSQRVDLVSLFPKHRVDLLS KNDQVDLVAQFLPSKFPPNLAENDL ALLVNLEFYRSDLHVYECVHFAADWE GLSGLPEVYEQLAPQPCVGETLHSSL PRDSEPFVPEEGSSEKESEDAPKTSPP TPGKHGLEQTGEEKVMVTVPDKNPP CPCCGTRVNSVLNLIEHLKVSHGKRG VCFRCAKCGKENSNYHSVVCHFPKCR GPETEKAPAGEWICEVCNRDFTTKIG LGQHKRLAHPAVRNQERIVASQPKE TSNRGAHKRCWTKEEEELLIRLEAQF EGNKNINKLIAEHITTKTAKQISDKTRL	cd01650	690	956

					CCTGATTTGGACCAA TTCACGGGATTTGTC CAAGGTGGACGGGC CACCTTTACTTAACCA GGAAAAGGAACATA TATAATTTATGTGTGT TCGATAAA	LSRKPAEETREEPGVCHRTRRAAASL RTEPEMSHHAQAEARDNGPGRRPL PGRAAAGGRTMDEIRRHPDKGNGQ QRPTKQKSEEQLQAYYKKTLEERLSA GALNTFPRAFKQVMEGRDIQLVINQ TAQDCFGCLESISQIRTATRDKKDTVT REKRPKKPFQKWMKDRAIKKGNYLR FQRLFYLDRGKLAKIILDDIECLSCDIPP SEIYSVFKTRWETTGNFKSLGDFKTYG KADNTAFRELITAKEIEKNVQEMSKG SAPGPDGITLGDVVKMDPEFSRTMEI FNLWLTTGKIPDMVRGCRTVLIPKSS KPDRLKDINNWRPITIGSILLRLFSRIV TARLSKACPLNPRQRGFIRAAGCSEN LKLLQTIIWSAKREHRPLGVVFVDIAK AFDTVSHQHIIHALQQREVDPHIVGL VSNMYENISTYITTKRNTHTDKIQIRV GVKQGDPMSPLLFNLAMDPLLCKLE ESGKGYHRGQSSITAMAFADDLVLLS DSWENMNTNISILETFCNLTGLQTQ GQKCHGFYIKPTKDSYTINDCAAWTI			

						NGTPLNMIDPGESEKYLGLQFDPWI GIARSGLSTKLDFWLQRIDQAPLKPL QKTDILKTYTIPRLIYIADHSEVKTALLE TLDQKIRTAVKEWLHLPPCTCDAILYS STRDGGLGITKLAGLIPSVQARRLHRI AQSSDDTMKCFMEKEKMEQLHKKL WIQAGGDRENIPSIWEAPPSSEPPN NVSTNSEWEAPTQKDKFPKPCNWR KNEFKKWTKLASQGRGIANFERDKIS NHWIQYYRSIPHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCDADIE SCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYK PDLIFVKDARALVVDVTVRYEAAKSSL EEAAAEKVRKYKHLETEVRQLTNAKD VTFVGFPLGARGKWHQDNFKLLTEL GLSKSRQVKMAETFSTVALFSSVDIV HMFASRARQSMVM			
	ORTE-1_AAe	GCF_002204515.2	Aedes aegypti	ACGGATTAATGGTGTGAATGTGAACGGGAGGTTTCG	TAGTAAATAATTGAAGAAGTATCAAAAAAAAAAAAAAACATACAC	MDLFNFDINDFQTAVVQKIPAGSFITAIMLQINKINSEFKNHQLNEMEFCCLIADFIRCNDHLFSKMWEEDIKEVVHKI	cd01650	1122	1384

				AAAGACAAATGC	AAAAACCTACAAAAT	FSMESGVEVHMAQAIVELFQTKICFY			
				ATGAACGGAGCT	TTACACACACGCATA	QKKREVEVIQPNSCSNTNMLTEINILF			
				AGCGAGTGATAA	AACATATTGGAAAAT	TFEDGTEKFYSIQSITDKKSSGKKESGY			
				ATCGGGAGCGG	AACACATATGAGTAT	NEQPRNELYKVKEYESGREMMVKKI			
				GGAGACGAAGT	GCACAAATACGAATT	KGDGNCFIRAIIDQMNKSEDYYRCNL			
				GAGAGGTTGGT	ATACATATAAAAAGA	SDIIELRNKIANHMNKNRERYEMYLIE			
				GCGAGAGAGAC	TACGAAAACAAATAT	SLEENGFLKFIEDIRKPGVYLGHESMV			
				GAAGTGAAAGG	ACATACACAAAAACA	AVTEIFDKCILLYQTDSETLRIGEEKDR			
				TGGGTGATAGTA	TATGATAACTCACAC	EEDVLRLLYTGENGVKNHFDSIRIVKG			
				AAGTAGGTGAG	GAAACACTAATACAC	QTEENTGTEKTSVEEVKEKESYTSMH			
				ATGTGTGAGAGG	TCAATCACTCATAAA	GKHNDDSEEISTIRYRKTQDTPITGKQ			
				TAAATATATAGA	CACATACAAAATAAT	IKETGKLVQEEGKQWQWNKERNNQ			
				CACTGGAGTGAA	CAATAATGAAAATTA	KKSDDKKAVEKNMDERGNETRKITR			
				GTGACCATTGGA	CTCTGAAAATACACA	EKHKPINKDRAITKDFITDNRKEEILDL			
				TTATTAGTAGTA	AAACATAGATACTAA	EEKRRRQRSKEREDRLEQLATQIKTN			
				GTGTAGGTCAAT	AATAAACACACACAT	EERSMLINGSQRQGWEENDSQERLD			
				CAAGAAGTGGTG	AAATTACACATAAAC	REKQAKTDKNTRGRIQMDVQGKTEE			
				AAGGTGGTAGTG	ACATATACAGTATTG	PLSIVNDEQVRKENNTKDVKLGDKKE			
				ACGATGACGGTT	AACAGAACATTTGCA	NPRISTKQRNDRKRAKNERVDEEGV			
				AGGATGCCATAG	ACTTTTTCGGTTTTCC	QENSDTMDGGTKKHRNHVAEERGE			
				ATAATATATAAA	ATACGGAATGGCCG	TRKGQVIESEESRATGVKATGEIGEEN			

				GCGACTTGGTGG	GTTTTGGTAGGTTTG	EEVKESIMLIATLNIRGCSKLEKRQEID			
				TGTTGAAGAAAT	ATCTCGGTTGTTTGT	NMLESFNIDVAALQEVNVKAELITTR			
				GAGAAGGGTGT	GGGCCGAATTGAGT	NYEWRIGGQGQNKSRGLAILIRKDK			
				AGAGAAGAGCG	GAGATTTTCACGGAG	GIDVMEWKEIGKYGLSISLLSKGRKIIL			
				GAAAAGAAAGG	CATTGGATTTCGGCA	INVHGPNKNGYIFYSSLGKLISQDHLR			
				TGAATTGAAACA	CGTGTTTTTGAATGG	SNLIMVGDWNAQIGKDSVMPEDAE			
				CAGGGGTGTTCA	TTTTTCCGATTATGG	CIGNRLGFSNNNDNGEEFKMFLTIHK			
				ACTGAATTGGGC	GTTTGGATGTAGTGG	IKNMSTVIGKNTDITWKSGKKESQID			
				TACGTGACAGGA	CGGTTTGACTGAAGT	HVLKPREGKIEIRFIKGYWTGINTDHK			
				TTGGAGCGATAA	GAAATTTGAAATTTA	MIVTGIRTREKENKEKGKKKIPIDASV			
				GAGAAAAATAA	ATACCCCTACACAAA	LKYDTIKEKYQEALKKYEIRQEKGNTIE			
				GTGAAAGGTATT	AACTGTCTTCAGCAA	QDFKAVAGKMKKAANEVLRSARAPL			
				AATTATTTAAAA	ACTTGTTCATAGCGT	TPIRKAALNKLKTAINLSNKHPDIFPYR			
				CTAAGGTAAGAC	CAAAATAGACAACTT	WKLKDRRMEFQAAIRSHNERKIKKFY			
				GACTGATGACTG	TGCTGAAAATATCAT	KELNDFDVAVRIKKSYQFLKGFLKRKK			
				GAATGCATTCTA	CAAGCTATTCTGTTA	HKNVYIPMSQWNEALKESEGPEIQL			
				ATACACTGATGA	ACGTGTTGAGTTATG	MDEHDTCPVTVPPKEEEMMQIVQQ			
				ATTAAGGTAAGA	TCAATAAAAAAAAAA	QCNGKSAGADGIRMELLKYADKQTQ			
				CTAGATATAAGC	TCGTAATAAGGGTCG	QELADIWKRVWIENHLPEDMEKTIQ			
				AAATATATATTAT	CTTTGGTCAAACCAT	VPIPKNAKAKGVQDFRRISLCNVAYK			
				GAGTTTTCTTTTT	TAATGCTTTTGAACTC	PYAKWIKQRLREFTGEPDINQAAFTE			

				CTTTATCAATTAT AGCATAACATTT ATAACTACACAA AGAGTACAAA	GTGGTTGGGGGGAT CAGTCGGAGGCATGT GTTGAGGTTCGGGTT GTGCCTGGTGTTCTG TGGGGGTTTCATTCG GATCGGTCCGTGGAT AGCTGGGGTCTGGCT TGCCGGGGTTGGTCG TTTCGTGTGGGGACC GGGGAGGTTGCGAG TGTTTTGTCAAATACT GTAAATACATACAAG ACAAAATAACACAAG CACATGTATATATAT ATGAGAACACATGCA GAAATAGACATTGAC TTATAACACAAATTA AAGCTGTTTTAAAAA GATAATTACTCAATTT TAATAATAAACACAT AGCATTAGTACAAGA	GRSTDDHMFVTRRILEEYWNAGNPL YVAALDISKAFDNVSLLQLRSILAKLN VPSHLIDRVLQCIRKEKIRVRWQNQY TTECNRGKGVKQGCPLSPFLFNIVM QDVVRKVKEKVPELNLMNTGSLKLPL ILIFADDILIITKDEEGLKRILIALEESLRE VGLEINNDKSQILIRAPNATKRPPNTI NLNGRNYDVKTSMRYLGAWLTDTL NRPATTKQRCVNAAKVSKLVVEFCKK FKPTWEIGKLIYNTVIAPSILYGTKTAT LTKRSRVQLGKYEKQILKDIWNQCRK TDERKFNVRKELKGKTINRRVRVNRIS YYGHIQRREETHPIKMAMKLKFNKKK HGRPSFTWKKSLEQDFNRYNGITEEE WSQIAQDRDKIKKKAEEIYRNNESEIS DGEDSTEQDM			

					ATGACATGCAGTCAT CAGTCGTTTGAGTTG GTTAAAAGAAAATAA AATTGTTCGATTTAA AAAATGCAAGGAATT TGAATATCATATAAT TAACTAGAAATAATA ACTACACTTGTTTACT TTCGAGGAATGATTA GAATATCTCGCGTTC GGTATCGTGGCTGTG GTGATCGGTTTCACT TCGTCGATTTTCCTTT TAACTAATTACTTGG CCGGGCGACTGTTTT TGACTGCTATTTAGA TTGAGATTGTCCATT GTCATGCTAAATCGT AAAGTGTTACCAAAA TCTATTGAACTTCGT GTATTAACCCTTACTC			

					GACGAAGAGAGATC GATCACTGCGGGTAC GCCTGTATGATGCTG AACGCAAGTTATGAA TATAGAAAACTCTTC TCTTCACTTTTGCATG GGATGTTTTAGATTA GCAGTGTTTGTTCGT TTTACTGGTGATTTTA GTGATGTCATTGTGT TGTGTGGGTTTTTCTT TTATAGCCAGTAGAT TGTTGCAGAGCGCCC AAAGGGGTTCAGATT ATGATCATAGTGGAA TTTATTATTTTCTTGT TATAAATAATATAAG AATATAAAGAATAAT TACAATTATCTTCATT GTACAATGCAAATCA TAAACTGAACTCCTT			

					AATAATAAATGTCTA GTTATTTTTTATATGT TAACCAATTAGAAGT TTTGAAGTAGTACAA TAAAGTAGATTAGTA AATGTGTATATGCCG AAAAAAAGTCTTAAT GCCAACCAAAGTTAT AACAAGGTTCAGTCG TAGCATTAAAATTCA GAATAACTTGTCATG CCGACTGAATGTTGT TTATAAAAATCAAAG AAGAATAATCATAAT AATAAAAATAAACAT GGTGGAGGGAAATG ATAGCAACGAGAAG CATCATAATAATTCTA TGAGAAACTTATAGC AGGTTATAAATAAAT ATGATGATTATTTCTA			

					CTGTTAACTTATTATT CAAATCTTTAAAAAT AGTTATAGAATGAAA CATTGCTATTGCCTG ATATGAAATTAAAAT AATTGAAAAAACATT AATAAAATCAATAAT TAGTTTTGAATTCAA AAAAAAAAATAAATA AATACAAATATTAAA AATAATAAAGATACA TAATGATTTGGATGA TGTGAGGGAGGATG GGAGGTCGTTGCTAT CATTTGTCTGTATGTA TAAACATAAAATAAA TGTTAAAATAAACAA GTATGTATAATTTTA GTCGCTATTAACTGA TGGATATACTGGTAA GAATACATATTCGTC			

					TTGTTTGCGAGCAGT GAACAGTAAATTATA CATCTATTGCAGTTTA ACCTGATGAAACTAC AGAGTATAATGACCG TTACATAACCGAATT ATATTAACACTAGTA TGTTAAATTCAAATTT CTAATGAATTAAATT TCATTTCTATTAAATT TATAGAGAACAAAAT AGGAATAATCGATGG ATCGATGTTGGAGCT GCAAACAGACTAAAG CTGCATATTGCACAC GGTCCTCTATGGAGT TATATTTCATAATTTC TTCTCTCTATATATAA AACATTTTAGGGGCA GCAGCAGGAAGGTG TGAGCATTTAAATAG			

					TGTTAAAATTGAAGC TCTGAACCGTATTTCT CTTAGGTACGAAAAG CGTTAAAAACAAAAC GTAGGATGTACGGA GTATCATAATGGTTA TATTATATTGTAGTTT ATTGTACGAGACCTT TTAGAACAATAAATC GTCAATTGGGAAATT CGGATATGAAGTGGT GTTGCACGAAAGTAA CATCAGTAGCAACAG TGTTGTCGTTTCATAC AAGGGACTAGTGTGT AATGTTATTTTATGTT ATATTGTGTAGCGTT GTACTATATCTACTGT TCATAGGTAATAAAT AGGTATACAAATAAT TCCTTTGGCGCATAA			

					AATAACAGATAATTC TTATTCCTTATAGCTA CTAGTCCAGTGGAGG TGCACGAATTACTTA TTTATAATTGTTTTGT TATACTAATTCATTG GCAGAAATTCCAGAA GCGTCATGGAGTCAT CGGAGATACTAACAT AATCACTGCTATTAA TATTTCAGCAGTTCTA TTCCGAAAACTGTAA TTCGTCCTTACTCTGG AGACTACCTAACGAG GTTCAATTCATCACA ATTTAAGAATGGAAA TTCTGAAACATCTAT ATTCATATCATATACA ACTATAAATTCCATTC TATATATTTTATTATA GTTATTTCATTTATTA			

					TATGCTGGGTATTGA GAGGGTGCATTAGG GCATCAGAGAAGTCA CAGATGGACGATGTA GTAGGGTGTCAGCC GAAATCTAGGAGAAT GTGCCGTAACATCCT TACACTACATGCTGC TACTTATAAATGACTT TAATGTAATTATCTTT GTTTACTACAGTAGG CTATAGCGAATGAAA CGAAGCGTCGGATGT ATTATTATTTTGTTTG TACTCCATGAACATA TGGAAGCTGGAATTT TGTTTTTGATGTATCC GGAAACTTTATTAAC ATATCACTGATTTTTT TATTTTTATTCAAATT ATTTGTTTATTCTTAT			

					CATTCTTCTGTCGACC AGCCTAGGAGCATCT GTAGCACAAATATTC AAAAGAAAATTATAT GTCGGCCGATATAAT ACTGCATGCTGTTTC ACTGCGTCATAATTT ATTGTAACTACCATTT TATTTTCATATTTTAT TATCATATTATTATAT ATATATACATTATAAT CTTATCTACTATATAA CATATCAATATAATT GAATTTTTCATGATTT GGGAAGGGAGGGA GCATCAGGGCATCAT TGAAATTACAACTGG ACAATGAAGTGGAGT GCCAATCGGAGTCAG GAGGAAAAGTGGTA GTCGTTCACGTCTAG			

					TACTTTTCTCTCCAAC AGTTGTGAAAGATTT GACGCGCCCTTCTTC TAACAAACCTTTCCG TGGAAAGCTCGACAA GTAATGCAAATTCTA ATACTCAATCTGTTG GATCATATGGCTGGA AGGAGATTCTCTATT TCCCGCGGGTTTCGC GTAAGTGTTTCTACTT ACGGGTCCGCCACCC CCCTTTTGGCCCTTCA TACAGCGATATGTTG AATGCAGATTGGTAA TAAAATTTGACCTTA CTGTTGAGTGGAGCC GCATGCAGAGCAAG ACTCAAGCAAGGATT GGTGGCTACCTACAT ACATACATACA				

	CRE-1B_FoCa	GCF_002217175.1	Folsomia candida	TGCATCATTGCCAAA	TAAGTGTATTTCATACTTTATCAATGTTTATGTTAATAAATAATT	MPLFNRPPQCRQLCPNPLLSANNATIQVLDSPVSTHNETSRSTQLDAHRSNFVGGSQIAKVESSDLDRSSPFNGVFYRIETGKIIAQPSSSPLPPSSSPLLPSSSPLLPLSSSSSSTPLPLSSSSSPPPPSSSSSPLLPLSSSSSPPPPSSSSSPLLPLSSSSSSTPLPLSSSSSPPPPSSSSSPLLPLSSSSSSTPLPLSSSSSPPPPSSSSSPRPHPGPRKKRNKQKKRTKPRKLPPPTSSQPLPHTTEGLNPVFEDLFCPVSGCPGPSSIGTDEWILCNACLIWYHTSCLFLQAIPEGDWFCANCPQIDTQIGPQINSNTSSPANQSYLGTHIQLTQSIPFTNPRASQPISTLSTPLSNFISSKETCTSLTLEWLNNHIVPGGVIEKAILHDAYRRDSLGIFKPINDAEMGKLIKNSFSNIRTKRIRTELHGTSKYHYVGIQFKSEHSADSVFDFQTIISSLRQNCPTLKQCPKACRIPVAQVLKSAVDKCISENTPEAWHLLLTFPYITLQCESKANIKNGQSLNSKIRNNLSIFESLDKNTYQDYLITLSSKNSAHDKPGSTIKDIGPKVISKVSEGNIRGAIRLLTSDDIIAPNSSETLE	SSF56672	674	871

						TLVEKHPPQPLDNFTVLDEPNIPPFTT TVEQVISAIFSFPSGSAGGLDALRPQIL KDLVSQHTGDVGTQLLESMTKFCNF MLAGKVPLDICPFIYGASLTALKKKSG GIRPIAVGCTYRRIVSKICVKVMNHVF QDYFSPVQVGFNVKNGAEAGAHAA RRFFSFPHKSPTAFLKIDFKNAFNMV DRGKLLSSVHELVPTLYPYINQCYRFS SSLVWNSSLIESCRGVQQGDPLGPPL FCLILHPVVSNLGSEFNMWYLDDGTL GGDPTTVLDDFKHIISCANEVGLELNF DKCEVSVLGLQTNTQCILESFQSVAP GISCMTQENAELLGAPLTYESIEHVFR KKINAFSNMAGRLNSIPNHIAYYLLKH SLAIPRLVYLLRAAPIFKSPHLLEEFDSL LHDTLEQITNCQINPLSLEEIALPVKLG GLGIKQSKTLATPCFLASSFATTDLVSL ILPSYVTALSDSLVTEAQDLWKSISCD ELPRSTISSIQSSWEIPIFNQVCSRILDS STTPINTARILASRSKESGAWLQALPS RNLGTLLDDNSFRIAIGLRLGLKLCFP			

						HKCVCGTQIDELGLHGLSCKKSAGRH SRHTQCNDIIKRALISADIPSIREPLGC CRTDGKRPDGLTLIPYKQGRSLIWDF TCTDTFAPSYLSHTLRHAGTAAKLAET KKHKLYSELEKQYIFVPVAVETSGVW GMEGLKFIQDVGRRMKDVTGPMST TFLIQRISLAVQRGNAASIMGAVPSQ RGLEELFYILNDRSFYLQKSCKGPNGD LGRLSGSSKGFPNPPTSGFSSTPQAN			
	L2-22_NV	GCF_000209225.1	Nematostella vectensis	AAACGGGAGATC TAATCTCATCCTT GGCCTCAGTGAG ACGTTTTTGGAC GAATCATGGACG GATACCGAGTTG GCTATAGATGGT TACTCGCTGGTT AGAGCTGATCGC ACATACAACTCT GGTGACGGGCT GCTAGTTTTCATA	TAAAACCGGATATCC ATTCTCATGTTAATTG TTATTGCATGTTAATT AATGTTTGTCTGTTTG CACTTGTTTCTATGTT GACGTCTGCCTGTTT TTGTAGTTAGTTTTGT CGTTAGTTTTTAGTTG CTTTATCAAGATGTA GTATTAGTGATGTCT TGTATAGTTTAATTTC CAAGTTTCCCCTTTGT	MFSHASSAISQQVMVLGDLNFDLLE RPLPAPTKQYLSIFDSLGYQQLITKSTR PISNSLLDHIFVSSSDFVIEADTISNTAS DHLPIFVAWKLKAGYTRPTGHKVLRF RSRKNFCIESFMHDLSLVPWSTLDIYD DVNDALDQWYKLFNDILTVHAPIKE HRIKRPLKLDWFNAEISAAIKARNCLH RIATSLKTDTSWSEYRRARNHVVHLIR KAKKTFYNDAIKNNLENPKNLWKIIR NISPAKHVNLPNRLTVNDTLYEDTFDI ANFFNEHFSNISSQVCNDAVPEFPDR SVLTNFINFKHQPGSYFSIPPISEDFVL	cd01650	374	636

				CCCAGCCATATTT CCTTCAAAAGGC GCGTGGATCTTG AGTTTGAAGGTA TTGAAGCCATTT GGCTTGAGCTAT TATTCCACCGATC GAGCCCGTGCTT TGCTTGTTTTATT TACCGTCGTCCTT CTTCTACCTCTGC ATACATAGCTCTT TTGGACGAA	ATTTCCGTAAATGTCT AACATAGTTTTTTTTA TTATTATCATAGGTA CTTAGATATAGGGAG GGCCATATATTATAA AACTGTAAAAGGTTA ATATGAAACCCTCGT TAAATAAAGTTTATT GTATTGTATTG	NSPSRLDFGKAAGLDELKVFFLKSAAP EIFRSLTKILNLSISSCVFPDPWKSAKV PPLFKDGSLFDRTNFRPISVLCILSKVL ERHIHTSFYYFLTSQDLLTDFQFGFRW FRSCELATINLSDSILKNMDSGLPSG MLLIDLKKAFALVDHNTLLLKLGLYGC DELTLKWFKSYLSGRTQKTCFKGVLS NALPINVGVPQGSILGPLFFLLFINDLP LFLSSDLVTDTTMFADDTTFLISGASV PDIELKLNKLATDVHKWATLNRMVL NTKKTKSMLIGSHQKVSNLAESSLSVT LNSCPIEQVNIARVLGVVFDNHFTCE PHIELFCKYINSRISLLRSIAPFLTPAGAI HYYNACIHSQFPYCASVWGNCSQTL QLLLLRSQKRAARTILSADFSVPSVSLF SRLNWLYICDLIKQRKLWTLFSILINPD CPLCLRSKFSFVSHERRTRSSTPNNLV LPQPRTDAGKRTFSFSAISLFNLLLPSL KQLAFSRFLNLNQLPNVTSLAKKKVR DIFHSHLNSVSHLEQLFCHACRFYIDC RCYFS			

	Crack-5_AAe	GCF_002204515.2	Aedes aegypti	GGTATTCCAACGGGAATGTCTTCTTACGTAAGGATGAAAGTTCTAAGGCAATCCATGTGAGTGATGTTAGTGTTCTATCATCGATTGTGTAATATTTCCTGATTAAGTTAA	TGAGATGTCCATATTGAATTACTGAAATGTAATGAGTTAATGTTAGTATTAAGTAGAAAATAAGAAACATTGTTAGATTATATGCTTGCTCCCTCGAAATATTTTACTTTTCGAGTGGCTGGACACAGTTAATTATGTAACTTTTAAAGCTGCGTTGTTAAATAAAGCAAAAAAAAATAAAA	MITLIKINYKAELINELNINFCLEEIKLDKENLNILYLNTRSCRNKEEELYQLINEFDGSVHVVVFAETWLYKNEIFNLLGYDAYHGCRDGRGGGVSIFVLSDYRSQLMLVINSEDNNFLMVELIDQNVKIMGVYNPGRSFNLFFEQLEQILHNYKRIIVCFDFNINLLDSSGELIRQYTYKIQSLGYIILNKLDARYATRISNTIATVIAHIITDIFEYDLQFGIKDTERLISDHRTLTLSMRIENHKKNKPDHKKICVNYENIIRSTSQNTFEICQQFSDIVVVLNEIMTKNRTEILSRKKAMVIRKPYINNALIAEIQNKNELYRQYRTAPGGVLKDELYILYARKRNMLKNKIKFAKEQYYGRKMHKNINNYKKTWSTLREIIFQNGSENITPQALTLKENGKIIDSAAQIAKCLNEYFINIGWNIQQNGTGNDNFLIYLNNMPETRFSFSVVETEKAKNVINSLKSGVAVGLNKISTKCIQKCVTLIAPKLTEIVNNVILSSQFEDCLKIAKIVPLYKSGDKFSKSNYRPVSIIPALSKIPECIMNEQIKHHLLQNELLNKNQF	cd01650	515	769

						GFLPKKSTETASLELLNFIVKGLDDGE YVGCIFIDLRKAFDCISHELLIAKLSFYG FTPSAVSLLKSYLANRKQQCQVNGVL SEPLLIKSGVPQGSIIGPILFNIFINDLF NIPLKGKLQCYADDAVLKYRAKDLETL QSMMQHDLQLLPSWFSANKMQIN NKKTNFIIFSLKTSNVSLDLFINNERLS QVTEVDYLGLAIDKNLNWKSHITLVK NKITPFIFALRRVRGCLSTKSCWQLYN AFILPHLSYMNTVWGSAAASHLNVL KVLQNRVIKTIRKLPFLFPSASLYSPEV LPINLLHKYNCLILVYKIKYKLVESNIRL TEFHDLHSHSTRARQNFYVSRPRTEL SLKHFIYSGITQFNALPNHLKLEIHLSS YKKKLKKYLFENN			
	CR1-54_AAe	GCF_002204515.2	Aedes aegypti	A	TAAGTAAATCAGTCT GTAGGAAATTTATCC AAGACGTTGCAAATA AATAAATAAATAAA	MWLRPDVLNSELCSNYTIFRCDRSEQ TSSFQRGGGVLIAVNASFRCTSVSLN DCDSLEQVVVSVNLPMSTIYLCGIYIR PTSPPEAYSMHATAIQSICNFSSETDA IIVVGDYNLPGLSWISDEELNSYLPCN ASTEAETTFVEAVMATGLYHVNSIRN	cd01650	470	721

						VNNRILDLAFVSDVCDIEVLEAPSYLL RMDAHHKSFVLRVQIWETRNNSSSD SNDETDFDFSLCDFESLNASLSMINW TNELANKGTDEATAHFYDKLYDVLHS NVPRKRLRPKQKGKLPWWSPELRRL RNGVRKARKRYYRSKSHVVRQNLHVI ESRYNKCKVFTFRNYINSIESNLKNNP NSFWSFVKSRKSKNRIPEQIRFQNRT SANVDESANLFADFFSSVNSTISPTLS TATRHSIPTHDITLSFLNFTEQDVKTTL TSLNVKKGAGADRLPPSYWKECAES MKLPISLIFNRSLADRKFPAVWKTAS VVPIHKSGSINSVDNYRGISILCCIAKV FESMMHMVLYNATRHLISDSQHGF VKKLSTVSNLMCYTNFLSSEIERRQQ VDAVYFDFSKAFDKVPHDLAIKKLEQ MGLPDWITEWLRSYLSGRKAFVRIG DTQSRGFDITSGVPQGSVLGSLIFVLFI NDLSHRLKSCKLFYADDLKIYRAITSVI DCYALPSDINELLLWCSENGMQLNN SKCKCITFTRSLSSIYFNYEINGSIIDRV			

						TTINDLGVTIDSKLKFSKHVRVTTAKA FAVLGFIRRNSQAFQDIYTLKTLYCSL VRTILEYAACVWSPYCTTQVLHVEKV QRSFLRYALRSLPWNDPVNLPDYESR CMLLNMETLSFRRTKLKQMFLFDLIT HNIDCSDLLAEVGFLAPSRYLRSRQLL APRTHRTAYGQNSPLTSCIRSFSIVND VFDFNLSRVEFKRRISSIR			
	CR1-1_CBre	GCA_000143925.2	Caenorhabditis brenneri	TTTCGAAGACAG ACGAGATGCGAT CAAGACCCTCAC TACATTCAACTCA ATCAAAAGCCGC GTGACCGGTTGT CACAACGCCACC ATTCGTCAGGAC ATGTCCAAACCG GAGCTGTCAAAA TTTCGGGACTCG TGGAAGCAGGCC ATCAAGAACAAC	TAACATGTAACTATTT GTTTGATGTATAAGA GGCCTATTGGTCTTT CACTCAATAAACTCA TCATCATCA	MNCFSIKNKLPQLDLVSSTNNFDVICL SETKLDDSFPDSLLSLNNKFSVVRKDR NKHGGGVAILITKKVKFCPIEIPKHFSS PEISGVDILLRGTTVRVIVAYHPSHTSS LERLVQSLEFLLCSQKHCVVLGDFNM PHINWETLTACDSPCKTFLAFVTNNG LTQHVLSPTRLNPDNILDLCLSNTKIV RDVQVGELFSDHKIISVTLKLSNSAKK VVKRVQLFRKANFQLINYNLSMINW DVEMSCMSLEEKYEFFVSVLQHQIAL HVPCKNINVLAKKHGPKILSLQKEKLR IWRAAGNSDNYKKISLILKEELIQEDQ MNTEKSLTYGTSKKFFKFINSRYKSND	cd01650	460	710

				AACGAAGCTGGA CAAGGGATCTGG ACTGTGAGAAAC CTCGAGCTGGTC AGAATTCCCTAC AAGGACAATGAA ACTCCTCACCCTT GGACAGTGCGA AACAGAACCAAG TAAAGTTTTTCTC CCTTGCCCAT		EIGVIKDSQNIPVNSNILKAELLSDAFS KVFTVDNSSSTLFPSRTNNVLESMTF EPSVIEKVLLKLVPKCNTTPDEIPAIFLK KVGTSIAIPLSIIFSESFRTSQVPRLWKT AIVKPLHKKGPRTDPNNYRPISLTSSV CKVMEKLVRKNLTDYLNSQQLFSRA QFGFRSNMSTESQLTSYFSTLISNHSL RKPTYSVYVDFKKAFDTVSFEKLTCRL MSYGIKGQLLDWIVCFLTNRTQRVN VNGVLSQEKFVLSGVPQGSVLGPLLF LLFINDIGDKFQSHFLLYADDLKLFSTN PSTIQTDLGLLSEWCDKWQMSVAPN KCENIIFNLTKSPRTLNFDFTIKGSKIPR TKKIRDLGLILSSDLPFSEHIDTILRRSH QRINIFFNILRRASFETFIKCYTIYVRPIL EYGSIVFNPVLKEQIKRIESVQKSFVYR VFKKFHIPYLNYFDALSVCNLTSLELRR LHADLVFLYKVIVSKETLLYPPFFTLPSL PNIRKHKFYFRSNSKNSNKFFSQFISN RSLNCWNDLSPSVFPPTPHSINFISNI SYCNFQKYLTLNPHNF			

	LINE2H_CE	GCA_004526295.1	Caenorhabditis elegans	AGCTGGTTAATTGCATGTAGGCTAATGCTCGTTCCATTGCATCGCTTGAAAAACTTGAATTTGTCCATAAGTTCGTAGCCTCTAATAGTTTAGACATCTTGTTTGTTACCGAAACATTTCTAAGTAATGCTTTTACTGATTTCTCATGCTCTTCACCTCACTATACCTGCATAAGATCTGATCGGCTTGAGTGCCATCCAAAAAATGCGGCGGAGGTACTGCTGTGTTTTTTCATAAAAAATGGGCTCTTACGCCTGT	TGATCTCATTTCCTCACTCATTTGTACATGATGTCATATTTTTTTATTTATTTTTTATTATTTTCTCATCAACTTTTCTTGGAAATCATTTGTATCATCTGACCTCGGACATGTTCCAATCGTGTCACCTTATCATTCACATTTCTTACCTCATCTCATTTTTATTCCTCATTTTGCTCTTAATTTTCAAGGCAAGAAATAGTGAATAAATAAATTAA	MKFHDLAQMVNFPTRFSKLNSGTFLDLIFTNNSGVLQNLKPGTPFINSDHISIFFSLSLPQPTPAHSAPPKSFMNFRKCDFFNLNRYMVSYNWPKQLSFFAPPDSKLNHFLNIFNELISDFTPQYPTLRVQSPKVTGRKLYRKLRKHHPDTESSVLKKRVKSRLKFLKKKNMAFENKLIHSKNSKSLFSYIRNRIKPSSSLTHFSVDNNNVDKPIEMAEVLVKEFSKSVVTHDPPFPTFPSCESTPSPVDTSPLTIASTILKLKPKIGYSLDNINFFILKKCCDSISIPLSLIFKEFITTNSFPSCWKTSIVIPVHKKGNSSDPKNYRPIALTHPLSRLFERIVLNPTRKTFSHKISKFQFGFLNNRSCITAHLNASSNYHKILSVPNQFVDIILFDFAKAFDSVPHELLCKLRNFGFDDNICLWFKAFISDRVSKVKIGDSLSNSSFKNSSGVLQGTVTGPFLFLVYINDLLDLFPPDVHVIAFADDLKLFSNNCQSLRSSIKIIENWCKIWQLKLAENKTKVLHVGKKNPKYKYYVNGQNIECCSKACDLGIWVDDQLAYEKHI	cd01650	309	557

				CTCAATTGCACCT TCCTTTTATCTCT CGCATTTTTGTCA AATTCTGGCATT AGACCATAAGCA ATCCAAAACTAG ATTTATACTGATA TACCGCCCCCAT CCACTACCGCAA GTAGAACCGCTG AGCTCTTTAAAA ATCTAAACCAGC TTCTCGACTGCCC CTCCTATACTTAT GTCCTCCTCGGT GATTTCAATTTTG CAAACACCAAGT GGAATTCAAATT TTCTTCCCACGG GCTCTGAGTATT TCTATAATTGG		LVKVNAAMFKCRQILKNFRSLNMMF YFKLFNTYIRPILEYGCEIYHPKSASLTK KLEQPLRYFSRHVFKKCKTQYDSYEH RLNQSNQQSLKHRRVFLILKTFHNIIT GKYHFPNLALYLKNSKSPRFPYQMTA IGKTNNNSFLHTHLTIWNLVTKHFPV PVSGYTFSKRISCLPLESYV			

	RTE-1_CRem	GCA_010183535.1	Caenorhabditis remanei	CTTCTTCTGTACTGAGCCTCTTCGCGGCGAAGAAGATCAACAGATGTCACACTTCTACAAACCCGGCGAGTTCGTCACGCCGGGAGTGAAACATCTGGCCAAGAAGTACAACGTCTCACGCAGGTCAGAGGTAGCGTCGAATACTCTCTGTGACCGGGCGGCTTTAGCATTGGTTTAGCAGACCATAGTAAAGTAATCAGAGACACCCAAACCGCTCAAAACAGCACCAAAACCTGCCTCAGTCCTCCGCAC	TGAGCGAAGAACGGATCGACTAAGTAACTAAGTAAGTAAGTAA	MDVYHSKRTKLTRPKTRKTDLRICTFNCRSVATDNRLAELLEETNRIKYDIIGLSETKRSAEVHMTNRDGTGVILGKRNDSSVSGGVGFIINKSLMPKIKEIKIVNHRIGYLTLQVNKNQYTIIQGYAPTADYEEEDHSNFYENLEDVYKSCKSRYKMVIGDFNARIGERKGNEVFIGPHSMENRNDSGERLATFCEVNRVFHMNSQFIKPTHRRWTYISPDKQHRHELDHILANGKYITDVSVVPSFTNGSDHRMLRANVHINAKQAKFEQVKRRKPPRRVLDPTAALLATENLDSCEDQDIDKEYDTLVHVLKKAQDAAVTIPTNHSRNRLKDSTRLLLSKRRFTDRSDPNFKTLSKECRQAVKRDHENFAKDRLLNAANQRKSLRKVARDINEYQSYIPCLLSGETGEKLTSREKMEIEVRRFYSNLFATKQSKATTIIPQQTKALPPFLPEEIQHSLNSFQNGKAAGEDKLSADFLKSCHSSVHKLLANKFSRYLQEGKVPTKWKSSKTTLIFKKGDKENLENYRPICLLPVLYKVFTKCILNRIRKSL	cd01650	511	779

				TCGGCCTCTGCA ACTGTTACGGTA GATTCAGTAGCA CGTCGAGAGAG GCGATCTAATCC CGGAGTT		EEAQPVEQAGFRRSFSTIDHIHSVQR LLEVGREYQIPITMVFIDFKKAFDTIEP TALWESLKTQGIDSGYIKLLKECYNDC STTITPFYNPVNIPITRGVRQGDPISP NLFSACLESAFRRMSWPHLKEDKDD YDNSPGIRINGRNLTHLRFADDIILISK TPQIAEKMLQELVNSCETVGLEINAS KTKVLRNKFASSHQIHIRKNNTMSAIE DVEEYVYLGRLLNTKNDLEPEIHRRRR SAWAALNNIKNTTDALTCPKIRAQLF DSIVLPALTYGSETWTFTKALSERVRIT HAALERKLVGITLTEQREKNLHREDIR KMSQVKDPLVFITKRKLSWAGHVMR RNDNRWTTLLQEWIPRNEKRPVGRP PMRWADSLKKEISVRQGTQLIEPWS TIAKDRKKWIAVIRAHTN			
	NeSL-2_CRem	GCA_010183535.1	Caenorhabditis remanei	CCAACTCTCATC GTATTAACCTAC GGTATTCCTCCTA GTGAGTGTAATA AAGGTTAATTAC	TGAATACCGTCAGAT AAGCCCCCAACATAA AAATAAAAGTCGGCG TTAGCTAATCACTAA ACCGGCTCCTCATTG	MTNVYLKPVNDNQTNKTGDNSRNT MSNSQCEMTWKPVARTYAQAASTN PADDKTVTVLGCKYNLLKLGNTPQTT KRSPPKPSRGGARISSVYTLTDELEITH REEGKITFAIDLPNKNNILCPLCRECT	cd01650	895	1156

				GTTTTCTCTTACC	GGGGAGAGTATCATT	QTRGRGSSFTKHMKLHVKEKHQLKA			
				AAAGAAAAAGA	CCGGTGCTCTCCGTT	TFIYKCSMCNEYEPEKKCGTKWIQTH			
				AATTTCGAATCCT	TGGGCGGTAGGGAG	LQKVHNYKYDESAIVVPVPPNTRQQI			
				TTTTGTGTAACTC	GAGTTGGGTAGCGA	ANELNNAAPFVDIRKPKANVVEEKKT			
				ACAAACTGACAG	CCCGGAAGTATGGAT	ENGALLKFLTKSNKDEQEHSQSNDSP			
				AGACCTATCGAA	GCCCAACCACCGCAA	NVESPEKANQALTIDPKGNNSPSKSS			
				TTTCTTTTGTTTC	TCTGATCTGGCAATG	MRSSQSSASSICQEIQEIITLSEDEEPK			
				GTATATAGGAAT	TGTTTCGGATGGTCT	AARPKPGINVWSLINEKGKDAYIDTDI			
				AGTCACTCTGGA	CTGTCTCTAGATCTG	MMAFLKMRVENYDSVNIIDPLNYQF			
				CCACGAAGTGGA	AAATAGAGCTCTGGC	PARVDLVPLIQRNLEDGKKRVVFPICA			
				CAGTTGTTGGCG	CTGAAGAACACACGC	DEHWTLLTISNGIAAFYDPTGSRMSS			
				GACTTCCAAAGT	GCGGACCGGTTGGA	YIEELVNELGLIIPKEQDEQPRQRDSY			
				TGAGAGAAAAG	TGCCGACTTGATCTG	NCGVFVMKMAEAFIQDTEWEMEEV			
				GTGTGAAGAGA	GAGGGTGCAAACCT	EEDVKNFRRNLLEELKPNYEIFAEKIKY			
				GGAGGTCTAGAA	GAAAGGGAAAGTTG	YNSPGKSFAQSRPTSRSSQCAVCPTC			
				ACACTTCGGCTG	AAGGCCGTGAGGCT	SRSATPMMDVGNMEVDPVPQQQE			
				TCTAGGACCAGT	CCTGGCGGGAAACTC	TPKSREPEQDEGWKVVGNTRKRGV			
				TCCTGAGTGGAA	CGTCATAGTCAGTGG	VTERSPNISPEAKRQFTGPEIKVVSPG			
				AGAGGAAGGTCT	TGTGCCAACACCCGA	KFHPLVGETEEMEVTCDSPPTKEPTT			
				AGAAACACTTCG	CGACTATGACATAGT	EPKVTPSLPAMRIVSPEVTKKQTSKKK			
				GCTGTCTAGGAC	TGGAGGAATCCTGAT	GKYGKKKQQTKKAQPPKGEPTKKAQ			

				CAGTTCGTGAGA	CTGATAATAATCATT	PKGEPAKLIEQVRTWFDKQMKSYQE			
				TCTCTTGTGAAG	GTTCATATAAGGGAG	QGSNIQTLTWIADSLTAAIFKANNGN			
				AGTTGAAAACAG	GGGGATGGTAAATA	KYLVDKITARCPPPLLNEGEMATQTS			
				TTAGCTGAGGCT	CCCAGGGTCCGAAAC	RRTEAVKPKDRFVKESNEPLRIQYAK			
				ACTGTATTTCTTG	CATCAAAGCAGCTAC	NRAKTFNVIIGKHSARCEIDINVVENH			
				ATAGCCCCGCCC	TGACCAGCATAGTAG	FRQTLKAQPVTEKALNTVCSGIKKAE			
				CCAATCCCCCTCC	TGATGAACACATAGA	VDPSIEGPISSGEVKAILAKIKDTSPGT			
				CCCCCCCCTCGA	CCCTGGGGTTCCCTG	DGVKYSDLKWFDPEGERLALLFDECR			
				CAGAATTTTCTGT	AACTCGACCCATCTG	QHGKIPSHWKEAETVLLPKDCTEEER			
				TTGACCTCCTGG	CACAAACCCACTTTG	KKPENWRPISLMATVYKLYSSVWNR			
				AATTTGCGAGGA	TACAAATGAACCAAA	RISSVKGVISDCQRGFQAIDGCNESIG			
				GTGCGCGAGAAT	CTGATGAAGAGTTTA	ILRMCIDTATVLNRNLSCSWLDLTNA			
				TTTCGAATTCTTC	ATGATTTCTTACATCA	FGSVPHELIRRSLAAFGYPESVINIISD			
				GCGCGTTTTCTC	CAGCTAGCGGACCTA	MYNGSSMRVKTAEQKTQNIMIEAG			
				GAAATTTTCCAG	CCGTGAGGTAGACTC	VKQGDPISPTLFNICLEGIIRRHQTRKT			
				AAGATTCGAGCG	CCGCCGCTGTAGCAG	GYNCVGNKVRCLAFADDLAILTNNK			
				GAGAATCTTCGA	GCTCGCCATTG	DEMQDVLNQLDKDCRSVALIFKPKK			
				GAAAGTGAGCC		CASLTIKKGSVDQYARIKIHGTPIRTM			
				GAATTTCGCGCG		SDGDTYKYLGVQTGNGGRASESESLT			
				AATTTTTCGCGAT		QIAAELQMVHDTDLAPNQKLDVLKA			
				TTTCAAATTATCA		FILPRLQHMYRNATPKLTELKEFENTV			

				ATTTTTGTCGGA AAATTTATTTTCT GGCAAAATTTGA TTTAGTTCACGC GGGAGAGAAGG ATTTGGTGGAAA AGGGTATTGATT TTTGTGGCGGAG GAAACTCCCACT GAATCAATAACT CTCAAAGGAGAA CTCATCGAACAA CCTCGGGTGACC TGAATCTTGGGC GAAATTTTCGCA TTGACACAAGAT AACACAAATTAC TGTTGAAAACAA ATCAGAACAAAC TGTCAAAAAGAA AGACAAAAAGTA		MKSVKMYHNIPIKGSPLEYVQIPVKN GGLGVMSPRFTCLITFLASTLFKLWSD DEYISSIHKKALSRITAKVMGLKTQTA TLQEQCDYLNTKKAITKGGYNLFSRM NEASRTLSVNLGAPLKSMQFIPENGE IALEVQASENSQIKVFSKADSMKLVTK LKDLVKSAMLKNFLENKKVKSQVVQ VLQHHPQSNKFVNDGKNISISSQKFV HPARLRLLVCNGNSYGKDHPKDCRR CGYECESQAHILQHCTYNFSTGITQR HDRVLNRILHEVIKGRKNNDYYDIMV DTEPGPTRERPDIIMIQKDGPEVLLA DVTVPYENGVVAIEAAWEWKIEKYS HFIEYFARLGKRAVILPLVVGSLGTYW PDTSNSLRMLGLSDGQIRNLIPDISMI ALESSKQIYWRHIFGDSYRIVTELYCTK NKQEARFGDEPMENVQVSDRFQPF KTREREKKSEEEKKRRSKSKKGKTWR GSKKQTDSRQSGKSYQNQGFQRSV GQGGSR			

				TTGATTAACAAC ATC					
	CR1-2_CRem	GCA_010183535.1	Caenorhabditis remanei		TGATTTTTTTTTTTCTT TCTTTATTCTCCCTCC CCCTTGCCCAACACC TATTTTATAATTCCTT ATTCCCGGTTTCCTTT CTCAGGTTTAGTCTC GTGAGGGACTGTACC TGACCTCCAGTGACA TTTTCTGTATATTTAA ATAGTTAAATAAAAT ATCAA	MSAQSLALIDHLDEYLPLKSCILAETLI THLLLESSIFNHPFSLSSSLKSLPERPFP TRILPHLQEPPLLNSDHFSVDFSLITTV SNHSSASLPTSPTFRNEHRLNYKKCDF KSLNADIANVDWDLVLSSISSPTDKFA LFSKMLSDFIISHTPLSTPPKPHRTSFSI RQLKRARQRYSKLLKLPNPPADQIFRL RTLIASTSKRIKSNMVSEERKILSAPNS RSARLLIKKRTRTSLKIPPLYVNNELISS NAAKSSIFSSVFFSNYNCSPSSSLIPVS NNSSNNISPSELFLPWIIENTLRKVPP RCGFTTHHANFYIIRNCATTLAIPLSIIY SDSFLKSHVPASWKHAIIIPIPKKGSLT SPENYRPISLTDPFSRVCERILCNYIKIN LFRNISRFQHGFLSKRSCSSSLVHSISE YRLLLSSHKSLDVVYFDFRKAFDQVD HQFLISKLDSFGIPSNIISWLTDFLSDR TFSVKIDDFVGTPSASIPSGVPQGSVS GPLLFLVFINDLLLKLADISSLCIAAFAD	cd01650	350	599

						DIKLYSHDPLALQMGIDLVESWASDN SLPLAHSKTALLRLGPRNSSHPYTIGG SLIDSVDSVRDLGLLIEPNLKFTRHINR AVALALLRSKQLLKSFKSNSPQFYIFLF KTYVLPLVEYCSVVYSPPPSSKLAHKLE TPLRFFSRKILQRCNTPYCSYSDRLSQL DLFSMRHRRLKAQLLLLYNLIIGASYF PSLETHVRLSSSSRRPMSLICINPNCS NFFSTVIPIWNALTINVPHFLCPSEFN SLLVNNIARF			
	Penelope-6_CoFl	GCF_000648655.2	Copidosoma floridanum	ACTCCTGTAGCCAACAACAACAATGATTCAAAGTGGCTGTTTAATCTTTCTAATGTTACTATACCAGAAGAAGTCTATAATATTTTGAAA	TAGAGCAACAGATTTACTGTAATTTAAATTTGCGAGTTCGTTGACTTGTACTGGGGTGGGGTGGCTGTAGCGTTGTGTGACTTAAATATACTCATGTACAGTGTCACACTTACACTGACCGATAGTATATAAGCACGATCGTAACTTAAGAAGTGCGTATTTTTC	MIRETSYFTTNMNTHSKFEFFKLLETSLHFVPTVTDKIRTNLLKMFDKFLKSTIAQKFDNQKIINDINKTALFLKHNKELLVTKVDKGNPTVVLDRSEHINSVEEQLNNDKYYTELLKNLLNELIKSVTTFVKRWYDLGIFKDCEFDIRSSEISKNDLSKCYGLVKIHMAGNPIRIIVSAVNSPTYNLSKFFTKFLNDNIPKPNSCIKNSFHIKEKLNSFIIPQHHTIISLDVVSLFTNTSESIILKSIKNNWKYLRNKINLSSQEFCEGVSLILNSTVFQFNNKFYKQIFGSAMRSPV	SSF56672	203	380

					GATCTACCATCTATC GTGATGCCTATAGTC GCTATTGTCGTGGCT ATGTGAGTGTTCATT ATTGTGCGATTTTCTT GTAAAGTATTGATGA CTTGAAATATAATTG TTTCATATTTTTATAT TTTTATAATTGTTATT TAGAAAGTTGTACCT GATGAGGGCCTTGTA ACGTGGTTTACTGTA CGGGCCGGAACGTT GTTTACAGCAATAAA CGATTTTGATGGCAT CACGAAGAGAGTCAT TATTTCTTGATAAAA ATTATCTGCAATCAC GAGAGCCACTGACG AACTA	SPVLCDLALRCLEEDCLKKLSFFVPFYV RYVDDILICVRNNSIEETITVFNNHDR HLQLTVETPVDNEISFLDLRLIFESDGS IITNWYRKPTHSGRYLNYNSHHPLSQ KIAIVYTLVDYSIKLSDSKFHKENLDISK KILFNNQYPLNFINKYIKKRLTILSINTP LNRVVSNNRGLMILPYFQGFDSIIKRY LYKNNINVIFKCIDKFSKFIRLGKDRTK HTDRCNAVYRIVCNDCDSCYIGQSSR RLKTRINEHMNDCKHRREKSALVDHI AKTNHSFNFDEVKILDTESNRFKRLFS EMLCIDSYRNNINRMHDTQNLKLCY KNVITEHT			

	RTEX-15_CGi	GCA_902806645.1	Crassostrea gigas	T	TGAGTTTCACTTTGTATAACTTTTCCTAACTCTATAACCTTTATATTAATATATGTTTACCTCTGATCAATTTATATGTATATTTTAATGTTATTTTTGTCCTCATGTACCATATGTTTGAAATGGTTTTGAGCGAATAAATGAACTTGAACT	MAVTCRNCSTVDYVLASANVFEILEDLDIHDFCNLFSDVHNPVSFSLSTCYLSRNISTGKNHSGVKLWCPDKADQFRSNIDVHAVKSIENYLNEISNSPGDLCSDNINKVVKDICTVFTDAARDTFGPSHFSPCIEKHNKPWFNRECKSARKQYHLAKSMHNKSKTDDNKNKLKEKSKIYKKAMDNAVKTYNTGITRKLMQLRSTNTKDFWNILNSSKRDSKSPVNIKDMFDFFKNINEGSLSPDSDTLHEPQVDGSSNELLNCEINESEITDAVKKLKNGKAPGYDNIVNEHIASTLNTFLPIYKLFFNAVFDSGLVPEEWLVGIIKPIYKNKGDPTSPENYRPITLLSCLGKLFTSILSKRLELYSEEIELIHNNQAGFRKGFSTLDHILTLQFLSNTLMKCKKKLFCAFIDFKQAFDTVWRNGLWCKLLNYGINGKCFKYITNMYKGIKSLIKMNGTVSDFFSCNVGVRQGENLSPFLFSLYINDLENYLVEKNIAGLQSFTTSIENELCVYIIMCIYIKLLVLLYADDTVILAESGDDLQNALNEFECYCKEWK	cd01650	324	600

						LNVNIDKTKVLIFSKGRMCKRVFYFN DRVLEIVKEFKYLGIIFSRAGSFLKAKK HLCDQAQKAMYGVIRKIRQFNLPVT CQLELFDKMVVPVLLYGCEVWGFEK LDSIERVHLKFLKHILCLKSSTPNYMVY GETGRFPLSVIVFTRMVSYWCKLISST NTSITNILYRYLLNQTENTTSTNPWID CIRHVFNSCGLSNVWHSQFLNIVNQ KWVSTAVKQRLQDQFIQKWHADIN NSSKGVIYKIFKTNFGYEKYLDILPLKF RKILGKFRTSNHHLPVELGRWEGIPL NERFCPLCNTKRIADEFHYILECDAIS QVRNKFIDKKFSARPNVLKFSYLMTT CNMKLLRKLCIFITKIYEVVCPP			
	CR1-24_CQ	GCF_000209185.1	Culex quinquefasciatus		TAAATCATGTTTTGTT TAAATTTATGTTTTGA TAACTTGACTATATG TATGTTTGTAAAGTT CTTATTAGTTTTAAGC AACCAAACTTCATTG AGACCACCGGGTCA	MEDPRLFVTVVPSHSRNYAFCTSHPS RPGPVFGPRRGVFQNRSVGKYPYAE NDLVSLPDARLTSSNVSTNESIRQDSA YRSQDAASQQSNDNADPGRPPTCY MEDLRLSVTVVPPHSRNYAFCTSHPS RPGPVFGPRRGVFQNRSVGKYPYAE NDLVSLPDARLTSSNVSTNESIRQDSA	cd01650	918	1167

					GATGATTTCAAATAA ACAAATC	YRSQDAASQQSNDNADPGRPPTCY MEDLRLSVTVVPPHSRKYAFCSSHPS RPGPVFGPRRGVFQNHLVGKYPYAE NDLVSLPDAHLTSSNVSTNESLRPDS AYRSQDAASQQSTDNADPGRPPIRI MEDPDPPDTVVPPHCCDSSRSGHQS RPGPVFGSREGVFQTPNAGKYQAAF QVGYALDPDPRLLSSDVQTTPVHSD DPTTPVHHDSRRSSAGQPVIFEPIRFY YQNVRGLNTKVDDFFLSCCDCDHDV ILLTETWLDGRVTSAQLFGNNFSVYH TYRNPANSKRKVGGGVLIAVRNSLDS VECVQAVAENLEQMFVTIRTPSKRIF VGCWYLPPEKRFETELMEKHLSCIENI RRIAGPHDVIVVAGDYNQGSLVWKH LSGNRSYVDPQLTRTRTEKERLTHGV MLDGMANCNMHQCNLVKNDRNRI LDLIFVSDDDCTVVAAADEPLVPLDA PHPALVFDIDASRTVFVQVDNNDRV LNYKRTNFDELQRTLADIDWTPVTSA MDVDTAVREFSNLLNFHLAQHTPQF			

						RTPPKPEWSNARLKRLKKERAAALKT FTLRRCRMTKLAFNAASSRYRRYNRY RYNVYIRKTQSQLRRNPKKFWTFVKS KYKERGLPATMALGDSTATTNDEKCE LFARHFSSVFREPEIISSTHHLDAVPRD LVDINTFVISRDMIEKATKKLKSSFSPG PDGIPSAVLKRCANQLIEPLLVIFNLSL RQAIFPTAWKKSFMFPIFKKGSKRSIE NYRGITSLCSGSKLLEIIVSDVFLFNCRS YVAPEQYGFMPKRSVNTNLLEFTSVC IDKIAEGRQVYAIYTDLKAAFDRIDHD IALQKFSKLGFSPMLCRWLESYLKGRII QVKIGASVSLELTNGSGVAQGSNLGS IVFSGFFNDANFLFSGNCKLSYADDFK LFIPIDTLADCYVLQSRLDKFSEWCGH NRLELSSAKCYVISFNRKQSRELFTFD YHLGSHVLERLEVVKDLGVLLDLKLDF HLQLSSVIDKANRKLGFMFKAAQEFD DPLCLRTLYCALVRSHLETSAVVWAP YHQNWIDRIERVQRKFVWFALRRLP WNNPAQLPSYETRCNLLGIETLQNRR			

						TISRAVFAAKVITSEIDSPNLLFQLNAR VQPRNLRPTPGFLSRPLVRTVYAENS PIRALATAFNEFYHLFDFHEPVSRFLE RLRTELRDRSRSSRTAENPQQRASRT RRSSR			
	Crack-10_CQ	GCF_000209185.1	Culex quinquefasciatus	C	TAATTATTTTGTTAGT TAATTCTCCGATTCAT TCCCTCGTTTGTATTT GTATCCATCTTGAAC ACATTTATCTTGGTA GTTTTCTTTTATTTAT TTTCAGCCGCTAATC GCCAGAAACCGAACT CCTTAAAAGGTCTAC GCCGATGGAGTTCTG GAGCCAGCCGCATCA ACTTTGTCGTTACATT TTTATTTAATTTTCAA TGACATGTAATCTAC TTAATTTGTATTGCAC ATTGAAAAGAAACAG	MLDNRDDVLGLFLDLSKAFDTVDRKI LIDKLRYAGVRGIALDLFTSYLSNRAQ FVCIDGIHSLLTLVNVGVPQGSVLGPL FFIIYLNDFSLLPLKGDLRLFADDSSLFY NNKSTNMNDLNLRDDLTIVIEYFRLN KLTLNINKTNIINIKNSSRSIPNSLSLTT SKFPDLEVVSDTKYLGIILDNRLNWSA HINALILKLHKITGIIFKIKHKLPQKVLFL IYHSLFHSILSYVTAVWGNACGLLINK LQVAQNRILKIILNLPIRSHTVDLYVKN NILPVRGIYVFQVCSFIYLCLNNKTHS NTKFVRSNHQYGTRYHDLLNRPNVT TVPGERSINFNGAQLYNYFNNRFGKS SSLNIFKNQLKQCLSRPDIIEKLLKSVNI FA	IPR00047	1	182

					TTTTTTCAGTCAGCAC TTTGCTGGCTTTTTCT GTCAAATAAAAACAA TCAATCAATCAATCA ATCAA				
	Crack-10_CQ	GCF_000209185.1	Culex quinquefasciatus	C	TAATTATTTTGTTAGT TAATTCTCCGATTCAT TCCCTCGTTTGTATTT GTATCCATCTTGAAC ACATTTATCTTGGTA GTTTTCTTTTATTTAT TTTCAGCCGCTAATC GCCAGAAACCGAACT CCTTAAAAGGTCTAC GCCGATGGAGTTCTG GAGCCAGCCGCATCA ACTTTGTCGTTACATT TTTATTTAATTTTCAA TGACATGTAATCTAC TTAATTTGTATTGCAC ATTGAAAAGAAACAG	MLDNRDDVLGLFLDLSKAFDTVDRKI LIDKLRYAGVRGIALDLFTSYLSNRAQ FVCIDGIHSLLTLVNVGVPQGSVLGPL FFIIYLNDFSLLPLKGDLRLFADDSSLFY NNKSTNMNDLNLRDDLTIVIEYFRLN KLTLNINKTNIINIKNSSRSIPNSLSLTT SKFPDLEVVSDTKYLGIILDNRLNWSA HINALILKLHKITGIIFKIKHKLPQKVLFL IYHSLFHSILSYVTAVWGNACGLLINK LQVAQNRILKIILNLPIRSHTVDLYVKN NILPVRGIYVFQVCSFIYLCLNNKTHS NTKFVRSNHQYGTRYHDLLNRPNVT TVPGERSINFNGAQLYNYFNNRFGKS SSLNIFKNQLKQCLSRPDIIEKLLKSVNI FA	IPR00047	7	181

					TTTTTTCAGTCAGCAC TTTGCTGGCTTTTTCT GTCAAATAAAAACAA TCAATCAATCAATCA ATCAA				
	L2B-6_CQ	GCF_000209185.1	Culex quinquefasciatus	AATTGG	TGAAGTGGCATCCCA TAGGTGTGGCTGTGA AGGAGAGCATGTCAT GACGGTCGGCTTCTC TACA	MLAIEIHHQQFNGIFGGVYHSPSSSD RDFLDSFENWLQQVMADDKTNVIS GDFNICWNNEGYSRDLKNIADAVGL EQKVLEHTRISVRSRTTIDLVFANVED CTVKVLADWRITDHETLGINVEGVCK INMPEKKTEYTCWRKYSKERLQNTLL CDETLWRNNLPVDEKAQQFSRALAT AVGQLTETRMRRSADTKRWFTPHLQ AVKLKRDEAYKIFKTTNCAEHWERYK QFRNEYVRELQKAKNSSVQQEIQSCS GDSKKLWRCLKDLIKPGGTPEKEIIFD DKTEPCSDEETANRLNHYFIKSVKEIH DSIPTVNQPQQDPLPTVRTLDISEFSQ FQPISMEKLRKIVAAMKNCAGVENIS KRVLEDAMDVVGDKLLDIINSSLNLG VFPQEWKKSVVVPIPKVPNSTRAEDR	cd01650	397	652

						RPINMLPIYEKVLETVVREQLVEYVDR AGLLIEEQSGFRKQHSCESALNLLMLK WKQSIENNKFILTVFVDLKRAFETIDR RKLLEVLRRNRIKGDVWRWFKSYLEY RTQITRYNNSVSEAEAVELGVPQGSV LGPLLFILYINDLTKALRRACVNLFADD TVIYVAGDKLDECYEIMNNELAVFAD WLKWKKLKLNVSKTKYMVVTTRRSE SNCTISVDGELVERVNAIKYLGVMLD EKLSFAEHVDYTIRKAARKLGVLCRIN RYLSFDNKIMVYKTLIAPHFDYCASILF LATNQQRKRMQIVQSKAMRMILRC DRLTPRILMLDSLQWMSIRQRVEYST LVFIFKVVNGLAPQYLTKTVQCGRDV HRHFTRQAGDIRLLNFKKSCTQNSLF YRGYNLFNMLPESTRATTNLREFKKL CKPFVRQRPLE			
	RTE-3_CQ	GCF_000209185.1	Culex quinquefasciatus	TGTGTGTGCGCC GTGCACATCGGT ATTGTTTTTCGTA CCGTTCCCGCGA	TAGCGCCAATAAAAG TAAAGTAAGTAA	MVLRRDRGLVQATRTRRKTPVQEAH DASRTNRHGTGHHMRSHDWKLGT WNCRSLKFDGSIRILSDILRVRKFSIVA LQEVGWIGAEEVQAYPRIGCTIYQSR	cd01650	569	829

				AAGTGCAAAATT		GENKRLGTAFIVLGEMRDRVIGWTPL			
				TACACGAAAATA		TDRMCVLRIKGRFFNISIINVHSPHSG			
				CGCGTTAGTGGG		SEDDDKDAFYEQLNWTYNSCPKHDV			
				GTGTGAAATTCA		KIVIGDFNAQVGQEEEFRPVIGKFSA			
				TTGCATCCGTGTT		HVRTNENGLRLIDFATSKNMAVRSTC			
				CTGAACCCGTCT		FQHNLRDKYTWRSPQGTESQIDHVV			
				CCTGGCCCGGGA		IDGRHFSDIIDVRTYRGANVDSDHYL			
				AAGTGATAGTCC		VMVKMRQRLSLAKSVRYRRPPRLDL			
				GTGTTTTACGAC		ERLKLPEVASRYAHSLEAALPGEGELL			
				CGCGAGTGAGAC		EAPLEDCWRSVKAAITNAAESTIGFV			
				AAATCAATCACC		ERGRRNDWFDEECRAILEEKNAARR			
				AGATTTGACCGA		AMLQYNLRDYEEAYGQKRRQQHQL			
				ATATCGCGTTCC		FRAKVRHQEELEFEDMEQLHRSNET			
				GCCACGGGCAAT		RKFYKKLNGSRNGFTPRVEMCRDKN			
				TGAAAAAAAGTA		GAILTNEREVIDRWKQHFDEHLNSAE			
				CAAAGAACTGCT		AEAGVQGGRREDFIGTAGEGEEPVP			
				AAGGCATCGTCG		TMREVKDAIKKLKNNKAAGKDGIGA			
				TCGTCGTCGCGA		ELIKMGPEKLASCLHRLIVRVWESEQL			
				TCGATGGTCGTT		PEEWKEGVICPIYKKGDKLDCENYRAI			
				GTATTGTGTCGC		TILNAAYKVFSQILFSRLSPIAEGFVGS			
				GCGGCGTGTGG		YQAGFVMGRSTTDQIFTVRQILQKCR			

				AAGTGTACCGCG		EYQVPTHHLFIDFKAAYDSVDREELW			
				AAGACCATACTG		KIMDENGFPGKLIRLIKMTMDGARC			
				ACGCGTGTCGTC		CVKISGAESDSFTSLGGLRQGDGISCL			
				GTCGTCCTCGTC		CFNVVLEGVMRRAGFNMRGTIFSKS			
				ACTGTGTCTGTT		NQFICFADDMDIVGRTFKAVADAYT			
				GCTACGAGAGTC		DLKREAEKVGLRVNVAKTKYLLAGGT			
				GTGTATAGAAGA		ESLRARIGPSVTIDGDEFEVVEEFVYL			
				GAAGCCGTTGGT		GSLVTSDNSCSREIRRRIIAGSRAYFGL			
				GCACGGTGGGG		HKSLRSRKFSLHTKCSIYKSLIRPVVLY			
				TGAATTGCTAGG		GHETWTMLEEDLRALSVFERRVLRTI			
				AAAGCTACAGAA		FGGVYENDGWRRRMNHELAQLYNE			
				AAAAAAAGTGCT		PSIRKVAKAGRLQWAGHVARMPER			
				GCAAAGTTTTTT		ADHLSQRNQKINPAKLVFVSEPVGTR			
				GGGTGAAATTAA		RRGVQRARWVDQVESDLESVGAPR			
				AAAGAAGGAAA		NWRNAAMDRACWRRIVQQAKLMV			
				AATCAAGAAGCG					
				CAACCCTGCTGC					
				ATACACAGCCCC					
				CCCCCCCCCCCCT					
				CTAGTTCCGTTTC					
				TGGGCGTCTGCA					

				CCCCATGTTAGG GGCGGCCCAAAC GGACTAGGTGGT CGATTCCTTCTAT CCGTGAGCGGCT GTTCCCCACGTT AGGGGCGGCTC GAAAGAAGTGG CCGACACCCCCC CCCCCCCCCCCCC ACCCTTCTTTGAG CGTCTGTTCCCCA GGTTAGGGGCG GCTCAAAGAAAC CGGTGTCCTGCT CTATCGTCGAGG TAAGCGTCTGTT CTCCAGGTTAGG GGCGGCTTACAG CAGATAGAGTTC GGACCCCCCACC					

				CCCTCCCCCCATC GAGCGTCTGTTC CCCAGGTTAGGG GCGGCTCGAAAC AGCGTCTGTACC CCAGGTTAGGGG CGGCTGAGTAAA AGTCCTTGTGTC GGCGTGGGACT GTAAACAGTACC GGCACG					
	Kiri-7_CQ	GCF_000209185.1	Culex quinquefasciatus		TAGATCCAGATTAGA TTTTAGTTTTTTTTTTC TCACCCGTTCAGAGA ATTGACAATTTCTTTG GCTCCAAATGTACAA TCTCCTTATCCCTTCC TTGCCCTGATACAAA GTGTAAACTAACAGG ATATCGTTACCAAAT	MKNDPAGDVDADELNAFFSSGHRQ LQAVDRSDNSSEPCHRTTADHGDNG FAFRHTDVNEISRKILEVQTNATGTD DIPISFVKLLCPFVLPMLAHLFNHIIDT SSFPAAWKKAIVTPIPKSSSPIQPKDF RPISVLPAVSKVLEKVLLGQISEHLNA AEPPLLAQNQSGYRKGYSTTTALTKV VHDVYSNLDENRCTVMVLVDFSLAF NCVNHQILRKKLNTEFKFTRSACDLIT SFLSGRSQIVRFGNSMSAALDVPDGT	cd01650	111	371

					AAATTACAATTACAA TTACAATTACA	PQGSCLSALLFSLYINSLPRNLKCEWQ LYADDLQVYLSGPIAEVDRIMRDVNE DLAAIADWARVNQLFPNPKKTQAIV FNKTGTVTPQENIVFCDSIIPLSSQVIN LGLHMDCNLTWKAQVNDVVRKVYH TLRTFRRFGSVLSLQTRRKLVQAVIVP FFTYCDTVYFPGLSAALREQLHRGFKS ALRFVHNLRRRDTTVALRNSIMGHDL PDNYQLRVCCFMKKAFEGTLPDYIM QHLQRGRQERTAGFIIPRHTTSSGKS VLVHGASCWNSLPMAIKREARFNPF KRAVITHMQN			
	Rex1- 20_DRe	GCA_00869 2375.1	Danio rerio	ACGTGGC	ATTGTGCTGAACTTG AATTAGATATTGGAT ATGTACAAGTGTTTA CTCGTTGTTTTTGGA ACCAAGGTGTGCAGT ATGAAATGTAATCTT ATTATTGTTTTTATAT GTATGTAACTTATTTA TCTATTTTTAGTATGG	MMAPYLWQPVAAALDVFNFVYLLY FLNIFLQWLRTLYLHTFQFESIRKHWII FIFILAFLIGKSDAVNSPAIADAVIYSRE QLLALRHTSTFSVPTMEVPTELKRRR RGCRAGVKRRLKKKRYKPSLPVIIMG NVRSLGNKTEELTALVRTQREYRECSF MCFTETWLNDKTPDSVVALDRFHLV RADRCAKQTGKNKGGGLALFVNER WCRPAHITTKEVVCNKDIELLAVGM	cd01 650	578	828

					AGAGCTCTGCTGTGA CATGTGAATTTCCCTC CGGGGATTAATAAAG TCAAACTAAACTAAA CTAAACTAA	RPYYLPREFSYVIAITVYIPPSANASNA CEHINTVTASIQLKYPHSLIVISGDFNH VSLQSSLHTFTQYVDCGTRDCKTLD MLYANVKDAYASTPLPPLGRSDHNLI YIRSVYVPAVRKQPPTIRTVQVWSDS VYEALMDCFDSTDWEVLCKSHDEDI DSLTDCITDYITFVLRTQYPQKRLNAT QTISHGLYNFCVENTVPTKKVKCYSN NKPWVTPELKALLNEKKRAFLAGDRE EIKRVQKELKIKIKEGKVIINLKWNINY NKIMLEMYGVIVCTKLEVTIEQVRNG LRRIKKKKSAGPDGISAGVLKACADQI CWVVHYIFNLSLRLERVPILWKTSCV VPVPKKPNPTEFSHYRPIALTAQLMK VFERLVLNYIKKLLIASEDRLQFAYKTG VGVDDAIIYLLHKSLSHLETSGNTVRIT FFDFSSAFNTIQPALLKVKLENAGLQK GMVSWLTDYITNRPQYVRMHNCVS QVIMSSTGVPQGTVLAPYLFTFYTSD FQYSSKNCHLQKFSDDSAIVGYIRNH DESEYRLLLKSFADWCDMNCLNLNT			

						SKTKEIIVDFSRHPSQAVPVIIQESEIEI VSSYRYLGVYLNNKLDWSENTIQVYK KGQSRIHLLRRLRSFGVCQTLLKTFYD SVVSSVILYAVTCWGGGLLEKEKNKL NKLIKRAGSVVGGALPTIEEIAQDRM LDKFVSIMNHDSHPLYDTVQACASSF STRLIQPRCYRERYRKSFLPTVIRLYNQ SHTR			
	L2-6_DRe	GCA_00869 2375.1	Danio rerio		GAAAAGAGATCATAT CTCCCCTGTCCTTGCC TCATTGCACTGGCTA CCTATAAAATACAGG GTTGATTTTAAAATTC TTCTTTTCACTTATAA AGCTTTGCACAATTC TGCCCCGGCCTATGT CAGCGAGCTTATTAG GCCATATACAGCTTG CAGATTCTTAAGATC ACATGATCAGTTTCTT TTGTCTGTCCCTCGGT	MLTQPVSFSRVFHENSALEFSNSYTN FLSNANASLPSCSDTEQLAETFNSAC DFALNEVAPLKSKRSRVSSIPQPWLN SSTRTLRQECRRAERRWKKDGLQVS YQILKDCLNSYQDSVSEAKSKYFSNLI AKSASKPKVLFKTINSVLNPTMCSVP VVNLETCIKFSDFFAQKIQDIKSHLFPT DIDHSDSVPPSNSFTNFQPVSSSQLL NTISQMKLTFCPSDVLPCRLFKEVLDT ICPSLLALINSSLINGVFPDVFKHAVVQ PLLKRPNLDPAVPDHYRPISKLSFISKV LEKIVLSQLTSYLDIFYILDPFQSGFRAK HSTESALLKVTNDILLSMDSGSSAILIL	cd01 650	260	521

					CTCGCTGCAAGTCAA AGGGTGATCGAGCTT TCAGTGTTGCGGCCC CAAGGCTCTGGAATA GTCTCCCTCTGAATAT CAGAATCTCTCCTTCT TTGGATGCTTTTAAA TCTAACTTAAAGACA CATTTTTACTCTCTTG CCTTTGACTGATGCA TGCATGTTATTTTTGT ATTATTTATATTTTAG TATTACTTTTAAATAT ATTTTATTGTATTTTT AATACCAGTTATTTTT AGTTTTATTGTCTAGT TTATTTTATTGTAAAG CACTTTGGATCAACT ACGGTTGTGTTAAAT TGTGCTCTATAAATA	LDLSAAFDTVDHNILLRRLECVVGVQ GLALQWFSSYLKGRSFSVNIGSFTSPP KPIHCGVPQGSILGPLLFSLYMLPLSSI LLKHKISYHCYADDLQLYFPIKLGKDLS LENLFLCLTDIKTWMTNNSLQLNENK TEVLILGPAATNNSIKTQLGALNDNL HNHAKNLGVIFDPLLQFDKHVNAVV KSSFFQLRLVAKVKRFLSKKRFRNRNP CFNLLSSRLLQFFVYWPTSVNTLSATV GPKCSRQINNRN			

					AAGTTTGCCTTGCCTTGCCTTGCCT				
	Rex1-23_DRe	GCA_008692375.1	Danio rerio		TGTATGGATGTTGTATAGTAGTATTTTTTTATTTTTATTTTTTAATGTGTACATGGAACTTCGCCTGTGTATGGTACCTTGTTGCAATTTAGTTTCCCTAACGGGATAATAAAGCTGAGTCTGAGTCTGAGTCTGAGTCT	MLDLCYGSIKGAMKAVAGPPVGFSDHNVVHLLPVYKSVLRREKVRKREIKVWSEDNSLALQGCFDCTDWSVFKETSESIDELTEVVCDYISFCTDMIIPHKTVIDFPNNKPWLSNDTKSLLQSKHKYFKQGDMSSVKQLKKAVRVQINKDKIRYKDKIENQLSSNNLRSAWQGVKAMAGCKEDRGGSEINLNGFSSAREFAEELNTFYLRFDEDDFKSENDNLKQHLSCSNSVNISRIQVKGLLQRTNTKKSPGPDNIGGRVLHTCAEQLSEIFHYIFSQSLMLQRVPEIWKKSVIVPIAKTKFPRELNDFRPVALTSLVMKCFERFIKGEILLQSKDKFDPLQFAFRQSRGVEDANLTILNYLFNHLENPKTHARLLFIDFSSAFNTMKPHLLVEKLISLFNLDLNICGWILDFLVGRQQCVRVNGALSDFQFCSTGSPQGCCLSPLLFIMYTNDCRSIHRDCHIIKYADDSVIVSLLHDQDLGHGPAVQEFTSW	cd01650	292	545

						CDRFSLQLNASKTKDMVIDYRKASPT PNITNIKGLDIEIVDTHKYLGVVIDNM LTFQPNTQAVYKKVQQRLFFLRKLRS FQVSTSMMTLFYKQFIESVLSFCIVA WYGNLSVSSKNRLSSLVKVAGKIIGF NQTGLMVIYQNHVIRRAQGILCTPD HPLYSVFQLLPSGRRFRVPICRTNRTR NSFIMVAIRSLNESRGEGDVSIY			
	REX1-7_DR	GCA_008692375.1	Danio rerio	TGCGCAGGCGCG AGCTTGCTGCAG CATGTTTTGACT GCTCTGTCTTGTC TTAGTGTTTTAG AGTTTAATTTGTT GTCTTTATTTTTA TTTTTTTGTTTTA GTTTATTAGTGTT AGTAAGCGTGCC CTCCTTAAACAT GAGGGTAGAGA GAGCCATCTCTG	TAATTTATGCAATATC TTGCTGCTCGTGCCA TATATTAATTAATTAT TACGGGTTTTAATGC GTGGAAGTCTGTGCA TTATGTTCTATGATGT GCAATAGTCTATTAG GTGCAATATTTGTAC GTTTTTTTACTGGGTC TTAAATATTTTTAGTG CAATATCTTTGTTTTT TAAATGAGTGCTGAA TGACTTTACTACTTGA	MTWLHECISDSVVDVAGFSSVRADR TRRESGKSKGGGLAVFVNERWCHPN HIVIKERICSPDIELLAVGLRPYYLPREF SHVIAIVVYIPPSANATTACDVIHSAT ARLQTMHQDAFVLISGDFNHVSLNR TLSTFSQYVECPTRENKTLDLLYANVE EAYSCIAIPPLGRSDHNLVYLQSTYLPL VKGSQLLLGQGSGQKRPVQPCKTA WRLRTGMLCACHMVWTLITSQTVSL SISTSVWIVLYHEIHTLLSYCTMKSIRC FPNNKPWVTKDIKALLNEKKVAFRSG DKEAAKEVQRRLSVRLKEGKERYKKK LEQELKQNNSRYVWRGMKNITGFKT	cd01650	491	742

				TGTTTTTTTGGTT	ATTATTGTATGTGTA	SSPRTDGDVDRANQFNMFFNRFDS			
				ACTGCTCTTTGTT	ATTGTTTTATGTGTAA	RPLSSSDTALTAPSASAPPNSVTSNCS			
				TGAGCTTGCCGC	TATTATTGGTTTTGGT	ILPYFTPSPSSSFTIVPTEASIASPPFIGV			
				ACGCTGCTAATG	TAATATTATTTTCATT	SDSYKPPMIITEDQVRRELCRLHPAKA			
				CTAACACAGCTA	ACTTGATATATGACT	PGPDGLTTQVLKLCASQLSGVLLRIFN			
				GCGGTTGGTCCG	TATTAGTATTAGTTAC	LSLSIKKVPVMWKTSCLVPVPKKARP			
				GTTTTTGTTTATA	CACTGATTCTTTTAAG	SVPSDYRPVALTSHSMKVFERLVLGT			
				CCCCGGACCAGC	TGTAATATCTTATATA	LRPLVRSAQDSLQFAYQAKIGVEDAII			
				TGATCTCGTTGA	ATAGTTTGTATACTTA	YLLHRAYSHLDKPSASLRITFFDFTSAF			
				GGCAGGATGCTT	TGTATTGTGGCTGTT	NTIQPARLGSKLSAMQVHAPLVAWI			
				CGAGCACAGTGC	TTCTTTTATAGATTGG	MDYLTDRPQYVRLQGNKSDTVLCST			
				GGTCTCTCGGGT	TTGGATTTATGTGAC	GAPQGTVLSPFLFTLYTSDFQYNTEG			
				ATCCCCGCGGGT	TTATTTCACTTACTTA	CHLQKFSDDSAIVGCIKGGDDLEYRM			
				TACGGAGGGGG	ATTTATTTATTTTGTC	AVDSFVDWCELNQLQLNIQKTKELV			
				AAGACTAAGAGA	TTGTGGTTTTGTGTTT	VDLRRSRRSPVTPLSIRGVDVEIVQDY			
				GGCTGCCGTGGA	CTGTGTTCTGTGTGTT	KYLGVYIDNKLDWSKNSLVTYKKGQS			
				GGACAAGGTGC	ATGGCAGTCATTACA	RLYFLRKLRSFRVCNTMLRMFYESVV			
				GAGGGAAAAGA	TAAGCAATTTCCTGC	ASAIFYAVVCWGGSVKVADMKRFNK			
				ACGAGGAATGG	GGGATAAATAAAGTT	LIRKAGSVVGEELDNMETVAEKRTLC			
				GAAATTGTCTGA	TTCTAAGTCTAAGTCT	RLRSIMHNVDHPLHNVVDELRSTFSH			
				TGAGCAGCTGAG	AAGTCTAA				

				AAGACGGTATAA ACCTAACCTGCC GTCTGTTGTTAT GGGCAACGTGA GGTCACTGGAAA ATAAGATGGACG AGCTTACGGCAC GGGTCAGATGCA CAGAGAATTTCG TGAGTGTAGTTT GATGTGTTTTACT GA		RLTSLKCSTERHRKSFLPTAIRLHNVSL SHLPLQ			
	Tx1-49_DR	GCA_008692375.1	Danio rerio	C	TAAAAGGCTTTTGTA AGACCTATAGCACTC AACTAATTAACAAAA TTTTAATTATTTATTG AGACATATTCGCATA TTAATATAAATGAGT AAATAAATAAAATGA ATAGTATAAATGAAT	MYYMQEKEKGGVLIGLDLEKAFDRV EHDFLFAVLKEFGFGENFRKWIQIFLY RYFNLYLVLNVFFFLTECLQPTRSIRQ GCPLTALLYTLEAEPLGLAIKSDRSIKGI QIENNDNREAVYQYADDTTSLGILER YCKASGAKVNVGKSLCMRVGRVKEI TQHIPFKEEKEHTKMLGIWVGMDSN KADILNWERVLNGIEKRLRFWKARNL TLKGKYSSKFDVKRKVFILNALMLSKV	SSF56672	8	159

					AAAATGGGTATTAAACAAGATATAAT	WYVLSVTALPEVYYEKINDVVLKFLWGNGRAKIAYYTLIGEVKEGGLGLLDSFLKMKTLTIKIIQRFLEGKMGVIWKSVLDYYLKKCGHFNMGCNVLLMKMKKRMIAGLPKFYVEVLESGLRIMRGIFRKCVDYTKSKGADSRTTTFTELNNFKTILSEGNVLYFKEWWDGGLRQVKDVLYEVIKGFVPLKVIVEDGYKRVVKAKLQRQCHELKKAIPKDCIVIIEKKKSWAMKMNYQRPILRRKMERSMSLCTVKSFYAIFR			
	L2-42_DRe	GCA_008692375.1	Danio rerio	GACGTGTGCAAAAG	TAATCTTTATTGTGGTTGTTGCTGTAATACTTTTATTCCTATTATTACTTTTATTGTTATCTTTTACTCTTTTAGAGTGTGTGCTATGGTTTTGTTCAGCACTTTGGTCAGTGCCAAACTGAAGTAAATGTGCTTTATAA	MAPVCLSACRLPRQFLNVCSATILSVLFLLLFMSCVDALISYDRQALLNIRGLLENQPRDCSITHDYCYSSFPFPTAAFVRRLPCCLIYQRKRRRRRGNRGGVRVRMRKEISISLALHTACSFVPTYGHRGLHPARRSWDVRYTCHLPIFPDKRQLAEFGAPPRVRIRRRGVNLQHIKSLDYVHQSDLQVISHINMALLNARSVSNKTFILNDFFTSHNLDFLFLTETWIGAGESSVFGELCPLNCSFISTPRCIGKGGGVAMVF	cd01650	673	932

					ATAAACTTTACTTACT TACTTACTT	RNWFRCRTISVGDFTTFESQCALIESD TTPFLCVLVYRPPKPDRGFIKEFSEFVS HIITSYDRLLILGDFNIHVCCPGKPLVT EFMHVLDSFGFTQHIVGATHALGHT LDLIMSYGFSIENIVIEDASFSDHKPIV FNTTLSYFSGTTKSEGFHSRCINSLTA NQFSTMFNSNKVSTAILAAAEHSSSP DDLISLFNSSCSDILDCVAPLKYKYPK MKSQPWLDDLSRSLRQICRKAERRW RKDHLVVSLEIFRDAQVNFQKAAKKA RATYFSDIINKHSHTPKILFDTINAIISP RTSCKIEQSSEICERFKRFFTEKVADIR SSFAPLCPSSCVDSTVSSATFCNFEPV SLSELRGLVFQTKTSTSPCDPISSNLLK ETFNTIGSSIQAIINSCLVSGTVPAFLK HAVVQPLLKKPNADINCLNNFRPISKL PFMSKVLERAVLSQLQPFLKSSHKLEP LQSGFTMHHSTETALLKVWNDLLLS VDSGSNAVLVLLDLSSAFDTVDHTILL SRLEHWVGIKGSALQWFQSYLKSRSF SVSVGQFMSTSAPLLCGVPQGSILGP			

						MLFNLYMLPLGNIIRKHNISFHFYAD DSQLYVPLKSQNSLQSLINCLEDIKCW MAKNFLQLNNDKTEVILFGPSKTRM SYIGGLGSLAPYVKSQVKNLGVIFDSE LCLEKQISSVVKNSYYQLRIISRLKSILS FQDLEKVIHAFITSRLDYCNSLYVGLP QSSLARLQLVQNAAARLLTGTKKHA HISPVLASLHWLPVNFRIQFKILLIVFK ALNGQAPSYISELINLQSTPRSLRSSN KILLQVPRSRLKLKGDRAFAVAAPRL WNQLPPDITSASSLSIFKSRLKTHFYSI AFPGF			
	RTE-1_DR	GCA_00869 2375.1	Danio rerio	GGTTCCAAGGTT GTCATAGCCGGG GGCGGTAATGG GGATAAGCTCCC ACTATCTATAAA GTACCCCTATAC GGCGTGCGTCTC AAATAGCCTCTG ACAACCAAGTCC	TAGGATACGTCATCC ATGGTCAACACTGAC CGACGGAGGCCTACT A	MLTGLSEDLRGIDDSRKTAVINNELLR LNVDIAALQETRLADSGTLKEKDYTFY WQGRAPDEPRQHGVAFAVKNNLLS MVEPGRNGTERLLTLRLNTTTGPLTL VSVYAPTLNATLETKDKFYGNLTSVIN NIPDKEQLVLLGDFNARVGANHESW PSCLGKFGIGKMNENGQRLLELCAFH NLCIANSYFQTKPQHKVSWRHPRSK HWHQLDLILVRRSAINCVLHVRSYHS	cd01 650	526	803

				AGCTCCTGGCCT		ADCDTDHSLVCCKIRLNPKRFHRLRK			
				TCAAGTGTGGTT		QGNPRINVSRMSQPDLTQKFAETLV			
				TAGCTACCAAAC		KELDTAQTGDSALEMWESLRNTMQ			
				CCGGCGGAACTG		RTALATFGKRTTKTHDWFEAKASTMI			
				TTTTCACTGACA		PCIEAKRAALTEYKRSPTQKNLQILRST			
				GGAGAAGGGGC		RSKAQHIARYCANEYWQELSNDIQK			
				GCAGGCGGGTC		AAIAGNIRGMYDGIKKALGPTQCKTA			
				ACTGGCGCCTTA		PLKSSTGEIISDKQQQMERWVEHYS			
				CAACCAGTTGCT		DLYSRQNTVSSAALDVIKCLPTMEEL			
				TCGGGGAGATG		DEEPTAEELRKAIDKLASGKAPGSDGI			
				ATATTCGTTAGC		PPDLLKQCKDSLLHPLHKALCQCWKE			
				CTGGGAAGGCA		GAVPQDMRDAKIITLYKNKGERSDC			
				GATCATCTAGGG		NNYRGISLLSIVGKVFAKVILARLQKLA			
				GAAGGCAACCCT		ERVYPESQCGFRAERSTVDMIFSLRQ			
				GTTTTCAAACCTC		LQEKCREQQMPLYISFIDLTKAFDLVS			
				CGCTGCCTTGCG		RDGLFKILPKIGCPPKLQSLIESFHTDM			
				GCTATATCCATTC		KGTIQFNGSCSEPFSISSGVKQGCVLA			
				ATGGAAAAGGCT		PTLFGIFFALLLRHAFGSATEGIYLRTR			
				TCAGGAGTAAAC		SDGKLFNLSRLKAKTKVRETLIRDMLF			
				CTCGAGGAAAAA		ADDAAVTTHTQEELQSLMTRFSMAC			
				TCCGGAGTCGGA		KDFGLTISLKKTNVLSQDTATPPTITV			

				GTCCCTAAGGCA GTTTAACGCTGT TTGCAGCCTCAC TCTGGCAACTCC TGCGACGGCGCC GATACCAAACTG TAGCAGCCCTGC TGTTCCTTTGGAT TTGTCGACAACG TGGAGAGGGGG GACCCGCTACAT GGGCAACAGCCT GTCCTCCATAAT ACATTGCCCTGG CTAGTATCCGAT CTCGCACACCCT GGAGAGGACAC TCCAGCCTCGCT AGCACTCTGGCG TGGATACAACAC GGGGAGCAGTA		DDYQLDVVHQFTYLGSTITDNLSLDA ELDRRIGKAASTLARLTTRVWTNHRL TTATKMAVYNACVISTLLYGSETWTT YARQERRLNTFHLRGLRRILGITWQD KVSNVKVLTRAGLPSMYTMLRQRRL RWLGHVCRMEDGRIPKDILYGELSSG KRTTGRPFLRFKDVVKRDMKALDINT KSWEDLAADRLKWRCTLTKQLKSGE KNMMRASEDKRVHRKVQSVSQGAT YQCDCCGKKCLSRIGLFSHQRRCLRQ SVMPQYKS			

				GTTTACCGGTTA TAAGCCACAGCT CGGTTGGCGTAG AGCAAGGCGCCA GGGACTGCTCCC GACGGTGGGAG GGATCTTCGGGT CCCACTGGACAG TCACCGCCCGCC TCAAGCTGGGCA GCCCCCAGTCAA TTAGGTACTGCC CCGCCACAGTCT GCTTTCCTCATTG GGTGCATGGGG AATAGGAGCATT TCGAACAGCAGA CTGCAACCATCG CACCAGACAATA AAATAACACAGA GAAAGAGACCA					

				GCTCTAAAACTG GGATGCTGGAAT GTCCGTACA					
	Rex1-26_DRe	GCA_008692375.1	Danio rerio		GATGCAGGACTGCCA GACTCAAGAACTCTT TCTTCCCATCAGCCAT CAGACTTCTAAACAG ATAACCAACGCACAT AACAGTCTATTTTTCT GCTCAGTACAACACT ACATACTCCATCCCAT TTTGCACTATTTTATT TTTTATTCTTTTTTTTT GCACATAATCCAATT GCACTAGGCACTTTT TTGCATAATAAGCAC AATAAAAAACAACAA AAACAAAAAACACTG TACACTGTTTACATCT GTTTCATTACTCAAAT CAGGTTTACATATTG	MSWRPSWQTCMLAPHINVVFLLFF VCFLSFPRCIHAMHYQHIADKLAGSC NLRQTFRIYQFYSTGLTTHCPLPQPLQ RRRRKRHKRGKRGGLHARLKSRANR PPLPSILLANVRSLDNKLDELRARITSQ REVRECCALIFTETWLSEKVPGTAVQL QTHSLHRGDRTTASGKAKGGGVCVF INNSWCGDVQTVHKHCSPDVEFLLL KCRPYYLPREFTAVFIAAVYIPPRANA TAALGKLFNVFNAQEMAHPDAVIIA AGDFNQCNLRTVLPKYHQHVSIPTRE NNTLDNVYSNIRGAYRAAPRPHFGQ SDHISLFLYPAYRQRLKQTNPITKLVKI WNPQTESTLQDCFALTDWDVFKTA ATKEDLSVNVQDYAEYVTGYISTCVE NIIPTIQVKKFPNQKPWINSKVRHML NARSLAFTSGNENEYKSAKYGLRSAIK EAKRQYKEKLDNTYSTASAGQMWR	cd01650	581	834

					TTTACTTTCATAGATA TTTATATACTTATATT TTACATTCAATCTTCG GTCACTTCTGTGTAT ATATGTGTATTTATGT GTATGTTGTGTATGA TGTGTATGTTGTGTG TTTGACTTCACTGTG GACGGCAAAGTAAG AATCTCATTGTACAG GGAGACGTGTTTCCT TACTGTGCACATGAC AATAAACAGTTGAAT TGAATTGAATTGAAT T	GLQHITDYKSTTNVTISASESLPDDLN TFYARFESSSINTEQRHTQTQTSQPAL PPPVVTPAAVHKALRKINPRKAAGPD NIPGQALKACASELVDVFTSIFNLSLS QSSVPTCFKTTTIVPLPKKSPLTCLNDY RPIALTPIIAKCFERVVLPLIQSSIPDTL DPLQYAYRTNRSTSDAIAAALHTSLSH LEDKDTYIRMLFIDYSSAFNTVIPHNL THKLSELGLHPTLCDWLLDFLTGRLQ SVRIGNKSSSTILTNIDTPQRCVLSPILY TLFTHDCVASHKNNTIIKFADDTAVIG CITGGDEAAYRKEGASLVTWCDNNN LTLNTDKTKEMIVDMRKERKTHQPLF IRKLEVERVSSFKYLGVHISEDLTWTL NTTQLVKKAQQRLYFLRKLRKLGLSSK ILSNFYSCVVESILTNCITVWYGNATV KDRKRLPRVVRTAEKIIRSPLPSLQTIY HHRVHKRTASILKDPTHPQHGYSHSY PQGGGIGV			
	RTE- 2_DAn	GCF_00328 5975.2	Drosophila ananassae	GGGGAGAATGG GGGAGGTTATAT	TAGGGTTGTGATGGC CATTTATATATTATAT	MRWIGQGCLRKKNCDIYYSCHPERH EFGCGFVVGARLRRRVSHFLPINERL	cd01 650	490	743

				AAGGTAAAGTGT		ATIRITAKFFNISLICAHAPTEEKDDVT			
				CTGATTGCACTC		KDAFYAELDRAYGRCPSHDIKILLGDF			
				AACTTGGTGAAT		NAKVGREDIFGATVGQFSLHETTSSN			
				GCCAGCATGATC		GLRLIGFAAAHNMVVRSTGFRHLDIH			
				CTTATCACTGCTC		KASWLSPDRLTRNQIDHVVIDARHAS			
				TCCCCAAGCCGA		NVLDVRSCRGPNIDSDHYLVAAKIRT			
				CGCACTACCTCC		RLCVAKVARRGLLRKLDIGKLQSQRT			
				ACGTGGTTCTCC		RDAFSTQVTGLLSRATPTPEDISNM			
				CTGGATTGTCAC		WALISHSLRTSAEEVIGFQRTPKRNP			
				TCAACACCCGTT		WYDQECRDATAAKDAAYGRTLQAG			
				GTTTGACAAACT		ATRAVVEVYRSRRRDEKRLFRRKKRE			
				AAAGCGCGCGAC		QERRECESIESSRDRNEARNFYQRVK			
				GAGAGTCCTAGC		RLTQGFKTGGVACKDDDGNLITDAEI			
				TCTTCAAACTAG		VLRLWRDHFSSLLAGSGHDDYGEYD			
				GTTGTGGCGTAG		PQTPIYGTDVDVPTPSHAEVKDAIQR			
				GACTTGAAATCC		LKNNKTAGADGLPAELFKAAGDMLV			
				ACCCACGGAAAA		GSMHKLISKIWLTESMPDDWNHSMI			
				CAAATCTTCAGT		CPILKKGDATLRTNYRGISLLPVAYKVL			
				GAATGTAGCAGA		TSVLCENLKPHAEALIGPYQCGFRPG			
				TGAAGCCTCGGA		KSTIDQIFTLRQILEKSYENQIDTYHLF			
				ACGGAACCAATC		VDYKAAFDSPRRDRLYAAMSELGIPA			

				TAAAGTTGACGA CCATGGCACACG TAAAAAGGACAA TGATTTGAGAGT ATGCACTTGGAA TGTACGCACATT GTACAGACCAGG TGCTGCTCAACA GCTAGCAGATAC CCTCAATAAATG TAGGGCTGACAT CACTGCTATCCA GGAA		KLIRLCMMTLSNTISSVRVGNDQSET FYTKCGLRQGDSLSCMLFNILMESIIR KAGVHRTGTILTNRCVQLLGYADDIDI IGRTKRDVTAAFGAIEKESAKVGLAV NMDKTKFMVCSSRESRRLDSQLSAG SHSFESVKEFIYLGSAISSTNDVSLEIKR RITLANRCYFGLSRHFNSRALSRPTKIT LYRTLILPVLLYGAECWTVVQSDAAAL GVFERKILRKIFGPICVGDAYRTRWN HELYELYGDVDIVSRVKIQRLRWLGHI ARMEEDAPARKVFDAVIVGKRRRGR PRIRWQDQVMEALSTSGVTNWRAR AQDRKAWRQIVQQAVTR			
	Jockey-14_DAn	GCF_003285975.2	Drosophila ananassae	AG	TGATACCACACCTGA CCTAGTCTCCATTACT ACCTCACAGATGGCC ATAACTACAGAACCC GACGCCATGACACAG CAACGCCCCGCCAAT AGCCGTGCAGCAACA GCCCACTAATACGAC	MQDIDDAIEIFNRNIYHAAKIATRHIK TSNRNPTGQFLLDRKLLNLQAIKKYTK QIKMARQTPEAAKNYKAANERFLRAI KEAKTKRTNNILNKIDPLDRYRMQKL WKITNHFKRQPEPNWPLKLYTGDGT SAKDWYSWTKSSEEKTEAFTRHLEER FSPKLTNSESFRAHVISKKEEICTQKLL MTTPCWMHRENYGHNSDKTNLPFL	cd01650	268	540

					GCACAGACGCCCGAT	PITPNEVLAEVKLLQNKKAPGPDQID			
					CCAGCCGCCACTGCT	NKLIKLLPHKATLHLTHIYNCMLRYSY			
					CCGATTACAAAAACA	FPKQWKRAAITMIKKPGKTPHEVSSY			
					AGCGACTGCGTCCAG	RPISLLPGFSKIFETLLMHRMFKIQDFE			
					GCAGAAACCCCGCTA	RAIPSHQFGFRKQHSTEQQLLRVTQH			
					CGCAGCAACGGCACA	ILSAYEQKMYCSAVFIDVSEAFDRVW			
					TCAACACGACGCTCA	HDGLLLKLAKLLPHSLYKILQSYLTDRS			
					GCCACCCGGACCAGC	FTMNSTNGIPSRTGKISAGVPQGSVL			
					CGTCACCACTCTGGC	GPILYTVFTSDMPLPEVISSQHTYNPS			
					CACCAGACCAAACCC	YFENMMLSTYADDTIILCTKREPQLAI			
					CCGACGACGCCCAGC	QTTESFLKYITDWATSWCITLNPNKTI			
					AGAAACTCCGCCGCG	HVLFTKRSVANLTGNKSPKMKGILIK			
					CAGCATCGCCTCATC	NEPKHRYLGLILDRKLLLHHHINQLQN			
					AGTATGACGCACAAC	RIRNVAAGLNWIIGTKSKLSPQSKILLY			
					CACCTGCATCAGCCG	KQIIAPIWQYALPLWGALTSNTQFQR			
					TCACTACTCTGGCCTC	IEAAQSKLIRKILKAGRYTRNQVLRDA			
					CAGACCAAGCTACAG	YNIKSINEIYSEASNRLATSLASHPNTE			
					CAACCAGGCAGCAAA	ARKILMDPVTPNRLEMPRYLDQLNR			
					CCCGCCACGCAGCAA	FFLPLRQQLQQQNQLNTHQPALPTLI			
					CGGCTTGCCAACACG	SYEQQLLSQQASIKARQQDEEREQN			
					CACACATGATCCTGG	KRLKIEQDPHYRPSIGWVNAMSYKLR			

					AACAGCCGCTGCCGA CTTGGCCTCCAAAAC CAACAACGGCGTCCA GGCAGCTACATCGCA GCGCTGCAGCAACG GCCAGCCAACACGTC GCACAAGAACCTAGA AAAGCAGCCACCATT CTGACCACCAGAACC AACAATGGCATCCAG GCAACCACCCCGCAC AAAGCACAGGCCCAC CAACCACTGAAGAAA ACACCTGAAAGCGCC GCCAATGGATTGACC ACCAAACTCCCGACG ACGCCCAGCAGTAAC CCCGCCGCGCAGCAA CGGCTCACCAATACG ACACACTGATACCTG GACCAACCGCCACTG	HGITTRWEMADKIKYEPEYIQRLIMS PLLPSEIPPPTPDQELDARLSTLKTQII NDLDLNLNYETNLAEGDSRTRFLPNS SSPILQAPQIPCINLVEDHPLGNPAIA DIWKSTPEI			

					CCCTGGCCAACAGAC CAAACGACGGCATCC AGGCAGCAACTCCGC CACGCAGCAAAGGCT TGTCAA				
	R2_DPe	GCF_003286085.1	Drosophila persimilis	AAGATATGGGTC TGAATAATAGCG TAGAAGGGGAG TCATTCCGTAATT CGTAAATCGTAA AAATCAGATCAA GTTGATTCAAGA CCTCCTCGTGGT ATCTTCTGGATG CTATTAGACTGA	TAGCCTATACACTAT GTTGGAGAGAAGAC GCTTGCTACCTAGGC AAAATGTGAAATTAG GTATAAACATCGTGG TTGTAAAACTTGAGG TGGGTTTTTAGTACG TATGCGTGATTACTT CGTAATCATGAATCG TGCATGCTAGTGGGG TTTGGCCTCCACTAG TATCTTTGAAGATTTT CCTTCCTCAGCGATC AAAAAAAA	MSSFGLIVTNLNSETVLWGCQPLGQ FSLIGTNMQNTTPRIINTNSLTNQIPT VSSLGAQSEHSAQVNPNSGYQCTICE SSFRSKSGLGVHMSRRHKDEFDQLRL RTDRKAQWSEEELSMMARKEIELAA NGERYLNKKLAEVFTNRSVDAIKKCR QRERYKTKIEQLKGQAVPLPEALESET IQRRPSIRERDLLVTPPNTLGTTPTELS NSEILAVLQGYPPVVCNDQWRVEVL QSIVDGAQASGKEITLQRLSTYLMEV FPSQNDRPIQTRPPRRPRNRRQGRR QQYALTQRNWDKHKGRCIKAILDGT EGTATMPSQGIMGSYWRQVMTQT SPTYSGTNTTFRTEHPLEGVWSPITLG DLRVHRVSLTKSPGPDGITPRTVRSIP SGVMLRIMNLILWCGKLPVSIRQART	cd01650	411	673

						IFIPKVGNASRPQDFRPITVQSVMVRI LNAILASRLTSSVDWDPRQRGFLPTD GCADNTTIVDLILRDHHKRCKSLYIAT LDISKAFDSVSHAAVSATLTAYGAPKE FVDYVQNSYEVCGTTLNGDGWRSEE FIPARGVRQGDPLSPIIFNLIIDQLLRSY PNEIGATIGDHTTNAAAFADDIVLFA ETRLGLQTMLDTTVDFLSSVGLTLNS DKCFTVGIKGQPKQKCTVVIPETFRIG SRSCPALKRTDEWKYLGITFTAQGRT RYSPADDLGPKLLRLTRSPLKPQQKLF ALRTVLIPQLYHKLTLGSVMIGVLRKC DILVRSTVRKWLGLPLDVSTAFFHAP HTYGGLGIPSVRWVAPMLRMKRLSN IKWAHLAQSEAASSFLTDELNKARGR TLAGLNELTSRTEIETYWANRLYMSV DGRGLREAGLFRPQHGWVCQPTRLL TGQDYRNSIKLRINALPSRSRTTRGRN ELERQCRAGCDAPETTNHILQNCYRT HGRRVARHNCVVNNLKRILEEKGHT VHVEPSLQLETSVSKPDLVCIRDNHA			

						CVIDAQIITDGLFLDDVHHRKVEKYKR PEVISALRREFGVSGNVEVLSATLNW RGIWSNQSVRRLIAKGLISSGDSNVIS ARVVTGGLYCFRQFMYLAGYTRDWT			
	TART-1_DWi	GCF_000005925.1	Drosophila willistoni	AATATTGGCTTA TTTTACTTAAATA TTTGGTCTCTTTA TGTTTTTCCCCAC TAAAAGCTAATT GCTCTAAAACTT CTAGCCTATAAA GCTAACTCAACT AAAAGCTGACTG CTCTAAAACTTTT AGCCTTTAAAGC CAAGCCTTGCTC GCAATTGATGCG CCAACAGCCAAG CTAAAAAACTAG TGTTACTCAGAC CCAAAATATGGG	TGAGGAAAAAATCA GGGACAACCACTTGT CATTCAAAGGCTATC CTGCCCCCCTTGGGC CAAACGAACAAATTC AATACACGGAAAATG TACTACTCGCTCTAA ACATGGCTAGTCTGG AATCTGAGTTGCTCA AATTAATTGATAACA GCACGAATAGCACTG ATGCTCTAAAAAAAA AGGAAAACTGATACT GCACCCATGATGGCC CTAACTAACAATCAC TCTAACACCAATACT AATACTTCTGCAAAC	MNILNKEISLSKNKHLNKKLCKIDSSD RYRPHKLWQLTNKLKRQPLPNWPLR IPSGVDKKHAWTRSPEGKMELFTDQ LEKRFSPNLTNSSLDRAKINEETNAILA QKMLMEYACKTHHSIASSSANFDSS NSPFVKPITLSELKKEIDSLINKKAPGY DRINNALIKVLPLKAIIFLLKIYNSILRFG HFPLVWKKAIVVMVKKPKKTISDASS YRPISLLPAFSKILESLIAKRLLKFNRLS QSIPNHQFGFRKFHGAEHQLARVTQ FLQRAFEDKLYGSPIFIDITEAFGKVW HCGLRNKLIKLLPINLYKVLDSYLKDRT FSVRGNHGIMSRDGKILAGIPQGSVL GPLLYNIYSSDMPIPKFIIEHETLIEYYP TYKHGDTLLSTYADDTIFISTDKNALV AMNKTENYLHKFSTWSNKWCFTVN VSKTVHVLFSRRSTSPAVKAATVKFN	cd01650	195	471

				TCTCCCTACTTTT	ATTCACTTACCTAACC	NKPIKNTSSCRYLGVLVDSKLLFKTHIA			
				TTCTCCAGTGTG	ATAAAGAAATATGAT	KLCAQAKSTYTRYKWLGGAKSHLSTR			
				GAGGAAAACCA	TTTAGGCGGCCCCTG	CKVTIYKQILVPIWTYAIPIWGSLISNT			
				GTTTTTTCCCCGA	CAAAGCCAGAATCCT	LFKKLTGTHNKLLRKLTNANIYCPNSEI			
				CATTAATCGCCA	GCCCACTATTCAGCT	QEKHNAKPLETIYNDACLRFFNSMSD			
				CGTTTTTGTAGCC	GCCGTCCTGGCCTGG	HETKEIEKLYLKPFIPKRLPWPRYSKQI			
				GCCTTAGCACCA	CCAGAGGGCCAGTCC	DKLVASRTIKKAQTPPLMPPLNPSVT			
				AATTCTTATTCGC	TCGACAGCTGTGCCT	VTNIKLRIESKLMEFFSKAPALIADNC			
				CAAGTTTATTATT	CCAGTAGACGCGCCT	HQTYNRITEKQDVLRAKEASIARKNR			
				AAAAAATTGCTA	CCTGCACTCTCTGCCT	QKIQSGH			
				CCTACTTGACTAC	CGTTCTCAGCGTCTG				
				TTATGCTAAAAG	CACCATCTCGTCCATC				
				TAACATAATCTTA	GTCCTGTCCAGATTT				
				AATCTATCATACT	CCGCTCGCCTGATGC				
				ACAAACTATCTT	ACCAATACCTGAAAA				
				ACCTAAATCGCT	TAAAACAAATCTAAA				
				AAAAACTAAATG	AATAAAACCAACTAA				
				TTAATCTAATGTC	AAAGTACAAATACTC				
				TTATTGTATATTT	ACCTGAAAATACACA				
				TTTATTTAGTATT	ACAGCAAACAGTAGC				
				AATCTAGTTTTAA	TCCTTGTCCAGCGTC				

				GATAATTTTAATC	GTATGCATCGTCTCA				
				CCCCCGCCACTG	TCTGCGTCGTCCAGA				
				CTGCTGTAACGG	ATCTCGTCAAGCTAG				
				CCTTTTAAATTTA	TCCGTCTGCAGCTCC				
				AATTTTTGCTATA	TGTATCTGAAAATAT				
				TACCCTTGTAGG	ATTAAAATTATTTTCT				
				TGAAAGGGCAA	AAATCAGCAATAAAA				
				AAAATAAAAAGC	CACAAAAACTCACCT				
				GGCCTCAATCGA	TAAAATCTACGTCAG				
				AAGACTGAGGCC	AGAGCAGCAGCATC				
				AGGCGGATCAAA	GTCTGCGAATCCTCG				
				CGCCTTACCTTAA	TCCAGCATCTGGCCA				
				AAAGTGGGGTA	ATCATCTGGTCCATCT				
				ATTCAAGGTCAT	CGTCCTTCACTAGCG				
				GTAAACGCGGCA	CCTGCACCTGAAAAA				
				GAACTCTCCTTG	AAACACAAATCGTCA				
				AAGTGATCGGCT	GAAAAATCTCTCAAT				
				GACAAGAAACTG	AATCACGTAAACATA				
				GACATCCTGGCG	CCTGCAAATTCTCTTC				
				ACGGGTGGTCCT	AGTCACCTGCAGCAA				
				ACTCATTTCCCAC	TACCAGCTGGTCCTT				

				ACGACCGGCGAA	CTGGCCCTCTACCTA				
				AAAATCCCTCAG	AGGCTGCATCTGAAA				
				CCATTGACTTTG	AAAAACAAAATATCA				
				GAGTTTATTCTG	ACAAAGAATCAAGCT				
				GAATAATCGAGG	AATTTACCTAAATAC				
				ACAGATTAACTT	ATATCTAAAATCCAG				
				CACAAACTATAT	CCGCAGCCGCTAGCA				
				GCGAACTATCCT	GTTTCAACTGGAGCT				
				CTGACCACTTAC	CCTCCTGCATCTTCTG				
				CTGTTTATGTACA	CTGCACCACGTGGCC				
				CATCACGATGGA	CGTCATCTGGTCCTA				
				CCAGCACCCAAA	AACGCCATAAAGCTG				
				CTTACTTCCTGAC	CTCCTACACCTGGAA				
				AATCGAAGACTT	AAAATATAAAAATGC				
				TTGCCTGTAGGA	TAAACTATGCTACTA				
				TCTAACATTAAA	AATCAACAAATTACT				
				CTGTTCCAAAAC	CACCTGCCATGCAAA				
				ACCCTCAGCGAC	TGCCGCCAGTAGCAT				
				TCCCTCCACCTAA	CGCGTGCAAATCTTC				
				ATATGGAAATCG	GTGCTGAATCTGGCC				
				AATCTGGATTAG	AATCTCCTGGTCCAG				

				ACATTGAGGACG	CTCGTCCTTCACTAGC				
				CGATTAATATCTT	GCCTGCACCTGGAAA				
				CACAAACAATAG	AGAACACAAATCGTC				
				TCACCATGCTGC	AGAAAAATCAATCAA				
				AGCCCTGGCCAC	TAATCAGGTGAGCTT				
				TAAACCTCTACG	ACCTGCAAATCCTCTT				
				AAGGGACCTGCC	CAGTCACCTGCAGCA				
				AAATACGAATCG	ACACCAGCTGGTCCT				
				TACCGAATCGGA	TCTGGCCCTCTACCTA				
				GGAGCCATACCT	AGGCTGCACCTGAAA				
				ATAAATAAAAAA	AAAAACAAAAAATCA				
				TTGCACAAATTA	ACAAAGAATCAAGTT				
				TTAAAACAAAAA	AAATTAACTAAATAC				
				AATTAAATAAAA	ATATCTAAAATCGAG				
				TTGATATAATGC	CCGCAGCCGCTAGCA				
				CTCTCACGAAAC	GTTCCAACTGGAGCT				
				TTTGAACGTAAA	CCTCCTGCATCTTCTG				
				AAAACATTAA	CTGCACCACGTGGCC				
					CGTCATCTGGTCCTA				
					AACGCCATAAAGCTG				
					CTCCTGCACCTGGAA				

					AAAATATAAAAATGC TAAACTATGCTACTA AATCAACAAATTACT CACCTGCCATGCAAA TGCCGCCAGCCGC				
	CRE- 1_FoCa	GCF_00221 7175.1	Folsomia candida	GGTAACTTACAC CGGAGATACTAC TTATCTATCGTTG GAAGTCCCTCTA CTTTCACTCTTGA ATTCCGCCCCCG TCTATATATTTTA AGCGTGGAAGG AAATTGCCAGTA TTGTTGGATACA AGTCGTCAGACC ACG	TAAGTGTATTTCATA GTTTATCAATGTTTAT GTTAATAATAAATAA TAC	MPLFKSLDIWKYVNLTLLPPLPPSSSS PPLPPSSSPLPPSSSPLPPSSSPLPPSSS PLPPSSSPLPPSSSPLPPSSSPLPPSSSP LPPSSSPLPPSSSPLPPSSSPLPPSSSPL PPSSSPLPPPSSSSSRPHPGPRKKRNK QKKRTGPRKCPPPTSSRPLPHTTEGL NPVSDDLFCPVSGCPGPSSSATDEWI ECSACLLWYHTSCLLLQAIPEDDWFC ANCPQIHTQIGPQVNSNTPSPANQS TLGTHSQIIQSSPCTNPRASQPISTLST PLSNFISSKATCTSLTLEWLNNNIVPG GIIEKNLLHEAYRQDSLGIFKPLNDAEI GKLIKNSFSNIRTKRLRTEFHGASKYH YVGIQLKSEQVPETVFDFKSIISSLRQN CPTLKQCPKACRIPVAQSIKSAIDKCIT ENTPEAWHLLLTFPYVTLQCDSRANI	SSF5 6672	619	816

						KNGQSLNNKIRNNLSIFESLDKNTYQ DYLISLSSKNSVFDKSGSKINDIGPKVI SKVSEGNIRGAIRLLTSDDVVAPNSSA TLDTLVEKHPQQSLDNFTVLDEPNIP PFTTTTEQVISAIFSFPSGSAGGLDALR PQILKDLVSQHTGDVGTQLLESLTKFC NFMLAGKVPLDICPFIYGASLTALKKK SGGIRPIAVGCTYRRIVSKICVKVMNH VFQDYFSPVQVGFNVKNGAEAGAH AARRFFSFPHQSPTAFLKIDFKNAFN MVDRGKLLSSVHELVPTLYPYINQCY RFSSSLVWNSSLIDSCRGVQQGDPLG PPLFCLTLHPVISNLGSEFNMWYLDD GTLGGNPTTVLDDFQHIISCANEVGL ELNFDKCEVSVLGPQTNTQCILESFQS VAPGISCMTQENAELLGAPLTLESIEH VFRKKIHAFTNMAGRLHSIPSHIAYYL LKHSLAIPRLVYLLRAAPIFKSPHLLEEF DSLLHNALEQITNCQINPLSLEEIALPI KLGGLGIRQSKTLATPCFLASSFATTD LVSLILPSYVTALSDSLVTEAQDLWKSI			

						SCEELPRSTITSFQSSWESPIFNYVSSR ILESSTTPINTARILASRSKESGAWLQA LPSRNLGTLLDDNSFRIAIGLRLGLNLC FPHKCVCGAQVDELGLHGLSCKKSA GRHSRHTQANDIIKRALISADIPSIREP LGCCRTDGKRPDGLTLIPYKQGRSLI WDFTCTDTFAPSYLSHTLRHAGTAAK LAETKKHKLYSELKTNYIFVPVAIETSG VWGTEGLKFIQDVGRRMKDVTGPM STTFLIQRISLAVQRGNAASIMGAVPS QRGLEELFYILNDRSFYLQKSCKGPNG DLGRLSGSSKGFPNPPTSGFSSTPQA N			
	L1-16_GMa	GCA_012273815.1	Glycine max	GTTTTGGGAAGT AATCCCAAGGCC CAGCAGAGTGAC AAATCAAAGGAG ATGATGAGAAGT AATACCAAGGCC CAGCAGACTGAG GAACCTTTCGTG	GC	MIILSYNSRGLGRGVKWAAVRRLIVK HKVDLVCLQETKRENFNKSICQAIWG DSTGHWDFVPSVQAAGGLLCMWN NSTFVVERKEKGRSFLMLEGTCCSNN QKLIFVNVYAPCDLVGKKALWDELR QLKASNPSGMWCFLGDFNSIRSSQE RISLSQRRADPYDIAAFNQWIDDMEL QEIKCYGNSFTWIRPNGCVKSRLDRF	cd01650	497	762

				GACTCCAACCCA GCTTCTAAAGGT CACCCTCTGTCTT CTTTATACGAGG ACCCTATACAGC TGCAATGTGACA TCTTGAAGGAAA TGGGCTTGTCTT ATGAGGAAGAA GATAGCAAGGTT AAGGGGCTGCTG CTGGATATGGAG AATAAAGATGAT TTGATGGCAGCA GAGAAGGGAAT TAAATCTCAGCA ATT		LVSQNWLSLWPESCQTVLQRNLSDH CPSILQTNMVDWGPKPFRVFDWWL QQKGYQKMVREAWKKDQQGGWG GIALKNKLKNLKSVIKCWSKAEGNINP KKVLNIQQKLNEVEDLAAHRNLTDQE IKDRTLLQQELWNASNAFESLLRQKS RARWLKEGDYNTAYFHKYMNFRRA YNNIPGIFIYGEWIQQPDAVKNEAAN FFHNRFTEQMHSRPTLDGVHFKSISQ RQREQLTVPFSDQEIKQAVWSCGGD KCPGPDGLNFNFIKEFWDVLNPEFRR FVDEFYAHESFPRGSNASFVALIPKKS HPQSLNDYRPISLIGCMYKVIAKLLSN RLRSVMDELIDERQSAFIKERHILHGIL ILNEVVEEAMRSKKPAMIFKVDFEKA YDSVSWSFLDYMLLRLGFCQKWRR WISACLHSATISVLINGSPSKEFNPSR GLRQGDPLAPLLFNIIGEGLTGMMRE AVHKNLYRSYLVGKKREPVNILQYAD DTVFVGEVDWQNVHAIKALLRGFEL ASGLKINFAKSQFGIIGGGVNWGLAA			

						ANILNCSQMNYPFHYLGIPIGANPSSL MVWEPLITKFKAKLSKWAQKDISMG GKVTLINSVLNALPIYLLSFFRIPQRIV QKLISLQRKFLWGGDNTHNKIPWVK WADICLPKNEGGLGIKDISKFNVALM GRWIWAFASGQQQLWVRVLISKYG GWPEFQHDTDKRGFSHWWRDLRKI FHQSDHNIFKHQLTWKVGCGETIKF WTDKWLGEDYTLEQKYNQLFLISRQ QNSLISNMGEFNHDSWEWNLRWR RNLFDHESDL			
	Nimb-3_LG	GCF_000327385.1	Lottia gigantea	GCCTCGAACGGA ACAGGATTAGAC TCGTATCTTTAGT ACTGTCGCA	TAAAGTAGCACATCA GAAGAGCGCCCCGCT CCTCTGTAGGTGTGG CGATCATTGGGGACT TAGTTCTCGATAGAT CGCCAAGAAGAAAC ACAGAAGAAGAACG CAGTATCAAGACTAC AGAAGAAGATCAAG AAGAAACCATGGGA	MDRSLKILQHNLNGLRARLTELIKFLS TARYDVLCFQETKLYPGVKDTILPGYR LSRLDRPGAHSKCFGGGLAIYVRTELS YQTLVLHSTEFEVNGVSLSLADGSKIA IVNVYLPPGKNVIEPSSYRNLCEQISH RNVLFCGDFNSKHNAWGSYKQDER GEGLLQFIEDNDMCILNDGRPTLIHN SSGVMSHLDLTIATSNIASNCEWQVS STETLGSDHLLIDITYRGQIKREGTVGK PKWRIAKADWIKFTENCSQWDVKDI	cd01650	506	765

					GCAACCCGTTCCCTA CTCAACGTCTGCGAT ATGGAATTTCTCCGA AGCGGACACCATCCA CCAACAAGAAGAAG A	QDADVDVYANNVTMVLMEAAKDAI PRTSGKNRACPKPWWSDACQNAVK LRKKALRRVQRNLNTETKRTYLEATE NARKVISEARTSHWHDFCQTLNQHT QIRDVWAKIKSMNGTKSTREIPVVTY GDRVGESDRQKADLMAESFADASSD KHHSPEFMNRKTNFAVDHSNVLTHI DENDPLNLPFTSDELQKALSSKKDSA AGEDEVKYSMLKKLPPEASKILLDFFN ESWTNGHVPTKWKTASIIPILKAGKD PNKTESYRPVSLTSTLCKILETMVTHR LRYYLESNNLLDKNQSGFRKLRSTAD HIVRLQHAMSKTLNSSGFLCAVFVDL TKAFDLVWTDGLMHKLRQLGIGGR MHNWIASFLADRKIRVTIGKSLSEEFS LENGTPQGSVISPTLFSVMVNDIRNP KYLSNKCQIAMFADDSTFWGNGTLY TNLQMRIQTCLDRFAKWANDWGFR ISKAKTVGMLVCKDPQLTDMRLVLD GERIQFVSETKLLGVTFDRKLSWEPHI KNITNRCQIVLNLMRNISGNEWGAN			

						VVTLMMIYRALIRPILDYGSQAFMTA RPMLLHKLDVIQHKALRLATGAGIGIK TESLLQDVGEVPLQIRREELALRYWSK LKHLGESSPTYLETARFRHAKKTISFG QSVRDSVRKYGLDRLPIKPLVFSQVA PWTVSPPIVDTSLASRLDKKRDDPCIA RMIALTHMDARWSESIRIYTDGSHN PDDGTTGLGIYVPSSSTRISKRLPNNC SVYTSEMLAILYAVKYSAGHYHSNYV VLSDSLSSLTSILNDSSNSQAELVDEIR QACSLVHSQGKSIELAWIPSHVGIQG NEIADELARIGLGLSEIDRIDYSLSEIRS MITKSIKSTWEATLIDANAFRFRLNDS ISSKVFLHHPNRQICRTVARLRMMGE VRNHKYELKLCPCGEAGTAEHLLIHCE QLRELRKSLVEIMKRYRLEYTLKNILSH TGEGMSTICETVYNFLSESRMIT			
	CR1- 6_NVi	GCF_00919 3385.2	Nasonia vitripennis	T	TCAGATATATGTATC TCCGTATTATTATTAT TTTTGTACATTCGAAT CTAATGAACTATCTG	MTYPREGCTGPRDGLSKQEELTAGYS GFSNAWTLTRIIMSHRAWTVELFLLN CTIGPIDPSTSLCTTFLSKATTSRCHYW SVPLLVLMLMKSLDAVFPGPQETALF	cd01 650	711	962

					TCACTTGTTGTATATT TTTCAAAACTTGTATT TATTATAGTATCTTTC TAAAGGACGCGTCA GCCGTTAAATTATTA AATCAATAAATATAT AAATAT	LVRFGSSDDASRVLKNRKLLPSSVEAT ADHTAYQRDLYRKLKLKASLHNTSHP DSLKKVVYVKGTPTLVDLRRTSTAPLS EKPSKNLQASETFKKLAGILKPVNKLS LYYQNAYSMYGKLHLFKAKLFQLSILP DIVVITETWLQDCVLSSELGMQQFD VFRCDRDLSSHGVQRGGGVLIGVKKF LQAELVHVDNVIEQLIIRISLPSHKLIV AASYLRPDSSVQVFSSHVSNLEDANI KFPDHRMVLLGDFNLHHITWSSDTS AYSPMAYLSPALRSAADIISGYSTSMD LVQHFKHHPKKGYSLDLLFAQKDFISP MDFNEDILPCDDHHVPGMFSCNISV ASSASCNLKFCKRKFYSADFEAIRARL SEEDWATVLYSFSDDPDSGIARFYEIL NGVIDQHVPITKQSLSSYPKWYDSELI NAICKKKIAHSTWKSSNSKRDEIEFKR LRAVCVRLSRSKYRAYISSVELGLKRN MRTFWSFIKNLRNDSSIPSKMYHGPT RAASDVEVSNLFSGYFSSVYRTDSLRE FESNDEPLVSLSNIVLSIETLRSIVANLD			

						DNINGGPDGIPPYFIKRCWSLLERPIV HIFESMLRSGYFPTFWKYSFVVPIYKN GDRHSVANYRPISTLSCIPKILDAYLAS ELSNSLLSKVAQQQHGFISGRSTLTNL LVFNDFLSESLCNHEQTDAVYTDMSK AFDSVNHSRLISKLFNFGIRGKLLDLLR SYLTNRSQAVRVNGAVSSPVDVPSG VLQGSHLGPLLFCLYVNDLVPKLSAA QVLLYADDIKLFMRVRSADDSVKLQR DLDTLVAWSVENGLSLNANKCSVISF SRSKSLLLFNYCINGSRLQRVDVVKDL GVIFDSSLTFITANQQTDVVVSRSLR MLGFIKRSTVDFSDPSAIIYLYKTLVLP HFLYCPQIWRPHSQYLINKLESVRHSF LRYIAFKMGKPLHRFCHDYSEIAEGC DILTVDSTFYLHDCVMTFKILRGLTNC ERLSDLFVSRTVAYPLRNLRDLREGTF SSNLGFYSSINRMKRQWNSIPNNVQ SKESVGQFKSSIRSLSLAY			
	Daphne-15_NVe	GCF_000209225.1	Nematostella vectensis	ACATAAACTCTG AAACTACATCT	TAGATATGATTCTAC TTTATTTCATCAAATC	MESLFIEIINEKSKNTLIGVIYRPPHSN VNDFLNEMELLTTKLNTENKKCYILR	cd01650	412	668

					AGCCACATGTAAATT TAATAATAAGGATTC CTTAGTCAAATCAAG AACATTCATATATATT TTAGTAGAGACGCTA CTTAGTCTGCGACAG GGGTCTACACCTCAT AAGCATTGTGCTTTT AGTAGACTCCTCATC ATCACTATTTTTTGCA TTATTATCATTATACT ATTATTTGGTGAAAT AAACTTGATTGATTG ATTGA	DFNLNLINIEYHELTRQFLDIMYSSSM IPTISKPTRITSHSATLIDNIFTNNINM QDSFSGLLCADISDHIPIFHITKIGKTK QNKEKRNMYTYKRIMNDSCIDNFLG EIDQINWDNVVLKENVNDSYNSFLQ NITYIYETCFPLKKFRNKKINSKPWITN GLLKSIHKKHKLYKKSQELPSESNFNK FKQYRNKLNSLLRIAKRDHFTDSLEKS KMDAKNTWKILNSLLKKNIKKPNYPD IFKNNNQEISNPKDIADKFADYFTNIG PNLSKNIENTNIDIKSYLKGNYSNSLLF RLTTPDEIKEITLSLKNKNSCGFDEINIR VIKAAISILSNPLSTLINQSIMAGIVPD KLKIAKVIPLFKSGEKELFSNYRPISILP CFSKIYERAIYNQLISFINKYNILNNNQ YGFRDKHSTVMAVCDFIEKISSSIDKG YNTIGIFLDLSKAFDTINHNILLQKLQF YGIRGIPLLWFQNYLSNRKQFVNYKD TKSEEYDITCGVPQGSILGPLLFLIYIND VPNATELFSFTLFADDTNLIFKSKSAE DMDEILNSELFKLTEWFKTNKLSLNT			

						KKTNYIIFRNSKRKTKHTTKIEINNEPI QMVNETKFLGLIIDSQLNWKSHINHV RTKITKNVGIMNKLRFYLPKFSLKTLYF TLVHPYIHYGNINWASNYDTSLKRIFL LQKRCVRIITNSHYLSHTEPLFKEERLL TVYDINTLETALFMYKFDNNQLPGTF DNFFKRNNTIHSYNMRNSCDYRTTL RKTRRGQFSISYRGPKIWSNLTCQAK QCRSLNMFKKIVKDNILSQKAAIDL			
	L2-8_OL	GCA_00434 8135.1	Oryzias latipes	T	TAGTGGGACTTGTTC ATTTTATTAATTTAAT TCTTTTATTGTTTTTT GCTTTTATTGTTCAAT TGATTTTATTGTTTTA TTGTGATTGTTTCTTT ATTGCATCATATTGTT TGTGATTTTATTGGA TTTTATGTTAAGCACT TTGGTTGCCTTGGCT TTTATAAGTGCTATA	MRKYNVSFHCYADDIQIYLPLKSGGA EGLDRLQNCLADIKSWMELNFLCLN KNKTEMFVFGKSDLLTGCCSAVGPLD PSDAPFIRNLGVNLDRHFKFDKQVSL VVQSSFYHLRLISKVKPYLPPHALQQV IHALIFSRLDYCNSLYSGMDQSLLRRL QLVQNSAARLLTGTRRCEHITPILRSL HWLPVSFRISFKILLLVFKCLNGLAPQ YLSELLHVYRPCRVLRSEDQMLLVVP RTRLVTRGDRAFSVAGPKLWNALPY SVRSARTVSNFKTLLKTHLFTQAFNSS ERI	IPR0 0047 7	3	89

					CAAATAATAAATTGA TTGATTGA				
	Rex1- 5_OL	GCA_00434 8135.1	Oryzias latipes	G	TAACCTTTGCACATTT TTTTTATTATCTCTTT ATTTAGTTTTTTCTAT TTTGCACATTTCTTTT GCACATCTCTTTACCT GCACTGTTCTGTAAA TAGCTCTTTAAATTAT AATACCTTGATGTCT GCATTTATCCTACATT TTCTTGTTCTTATTTTT ATTTATATTATTCTTA CTCTTTTTTTCCCTCT GCTGTCAGTTGTTAT GGTAACGAACAAATT TCCCACGTGTGGGAT CAATAAAGTATTTCT GATTCTGATTCTGAT	MKLNQHKAAGPDGVSPRTLRTCAH QLSAVLQHLFNLSLRLERVPLLWKTSC LVPVPKKTSPSHPGDFRPVALTSHVM KVLERLVLAQVRPVVSSSLDPLQFAYR PRLGVDDAVIHLLQRTHSHLDGGGG MVRITFFDFSSAFNTIRPSLLGDKLRL MGAGASTVSWITDYLTDRPQFVRLG GVLSDVVLSGVGAPQGTVLSPFLFTL YTSDFQHNSESCHLQKFSDDSVVVG CISGGREEEYRRLVDGFVEWAGENH LLLNVDKTREMVIDFRRKRTVLQPLCI LGRDVEAVEEYKYLGVTIDSRLSWRS HSTAVYKKACSRLYFLRKLRSFNVCSK MLGIFYQSVVASTLFSSVVCWGSSIA VSDTNRLNKLIRKAGSVIGCKQETFEA VMERRSLNKLLSIMDNPEHPLHPVLD KQRSSFSRRLRQLRSKTARYQKSFIPN TITLFNNSQCRR	cd01 650	51	301

	RTE-15_PSi	GCF_000230535.1	Pelodiscus sinensis	CACTACACCTTTTTTGTCAAGGAAAACATAGTGATGAGAAACGCCTGCATGGAGTTGGTTTTGCCATTCGCAACTCGCTGCTCAAGTCAATAGAC	TAATGCTGCATCCCACCGTCTTTTGAGACATATGGAGGCCATGA	MTPSAESKRIMTLRLFTESGYINIISAYAPTLTTNSADKNHFYESLQKVVDSISSDEQLFILGDFNARVGIDTSSWPICLGHHGIGKMNENGQRLLEFCSINELCITNTFFPGKECHKVSWHHPRSGHWHQLDFVLVQRKDLNCVKSTRTYHSADCDTDHSLIVSKTKLKAKKFHRTTLNKKPKINCMNTRNSAKCEDFAKAFAKKWEENAHVESVEETWQTLCSAIYEAAREAFGEKKSSNKDWLEENADRLMPLIKQKNEAYLEYIKNPSDSAKAKLQHAKGLLQREIKHCINQYWIDICSNIQVASDTGNARAMYEGLRKALGPNITKIAPLKTKEGISITDKSKQMARWIEHYSELYSEERNIMSETRDFIPTMELMQELDAVPSKQELIDAIDALSSGKAPGTDCIPAEVLKHNKDCLIPPLYDLLLKCWTKGEIPQDMKDANIVTLYKNKGDKGDCNNYRGISLLSITGKAFAKVILRRLQLLANSIYPESQCGFCTGRSTIDMIFSLCQLQEKCREQNMPLYISFVDLTKTFDTVSRQGLYLL	cd01650	445	722

						LERIGCPPKLLQLIIAFHEGMRATIQF DGYTLDSFQVNTGVKQGCVLAPTLF GIFFSVLLSCAFKNSHGSIYLHTRSDGS LFNLARLRAKSKIKKVLIRELLFAADAA IVAHNETQLQDLMAQLSRTCQQFAL TISIKKTVVLAQGNTTSPCIMLNNNPL EVVLKFCYLGSTITSNLSLDAEINIHIGK AATTFSRLTKHAWANSKLTIRTKILIY QACVLSILLYGSETWTTYASQEKQLSS FHLHCLHKILNIKWQDKVPNEKVLER AGMSSMGAILHNRRLRWLGHMKR MDNNCIPKNILYGELAEGTRPKGRPK LCFKDYCKATLVKFQIDIDSWENLAV DRNTWHQHLRQGQKHYDKTLAEN QNKKRMKRKEQMITTNTTTAWSCIY CSRTCLSHIGLLSHTKHCKKRPV			
	Utopia-1_PS	GCA_009848525.1	Phytophthora sojae	AGACGAGCAAC GCGCTGGGGCCC AAGACCTGGCAT GAACGACACTGC CCAGGCTGTGAA	TGAAGCTACACAAGC GCGAGGCTCGACAG CTGGACCTTCAGCTG TCGAGCCACTGCATC CGCGCCAGCCACCGC	MDVDGGPAMPEPWVLRFDGACRR NPGPGGAGAALFKPCGTVAWTCSH YMPNSSETNNTAEYTALLLGVQSAV HHGASHLEIEGDSHLVVAQVKGTFA CRNPRLRQLRNRVRHALRAVTSLTLK	cd01650	499	763

				CGACGAGCACGC TGCTAACCCGGC AGCGCACCGGCC TCTGGGCTCCGC TGCACCGGTCAC CGGTGCAACACG GTGGGGTCCACG TCACGGCGCGAT CGGGGCAGCGG CTGTAGCTCGCC TGCTCACAGGCC TACCCACGGCAC CAGCACCAGCGA CGCGCCGGCCTG CTTCGGCTCGGC GCTGCCCAGACC CGCCCGCTTCCC CCGCGGCCACGA CGACAGCCCCCG ATGCG	ATCTGGGGCTGGCAC TGCCGGCGCCACCGC GAAGGACAACGGAG CGGCAACGCCTCGCG AACGCCGCGTGGGTC TGGGGGGGCCCCGC GGCGCACATCGCAG GCTCGGGCACAGCG GTAAGCTGGTCATTT GACCAACAGGGCACT ACCCAGGTAGGGAA CCGCCCTTCAAAAAC CCAGGAAGACACAA ACACCCTCCCTTTAGT GACATGCATATTTTA GCCTACATTTCAGTT ACGGAGAGGTTACTA ACTGGTAAACACGAA CACACAT	HIDRKANAHADRLANRALDLKRSLAE CGEHQGAMESCLHMNPAAQRQRE QPAPPARPACAPTRAEPASDHDEDI DAEIAARDGGEAFPTLPIGPGTAPAR QPRLRLRQLTEDEQEAAASALQAMA EELACKIEDADSWTSGDGYISAIPGRI RQVLQPFTAAQPHPRPTLQQQRQRP PRVTRTQREHRLDEALDEMAAVQQE RPTSRSAVRRARRRVGRIRASMRQQ QLRHDFARNESKCVEDILRAASAETA AEEHPETCPIDSGTLHEHFTAVNSPRI NFLPDEACGALFREAMADVGTPQER RSALTDELTVDEVEDQLMQAATNSS PGHDGVGYDIYKKFAAQLVPLLHAAF QFCWRHHRVPALWKVGFVRLIHKK GDPTDPANWRPICLQTAIYKLYSGLL ARRLSTYLEANGLLPMAQKGFRAYN GCHEHNFVATTLLDQTRRMRRRLYQ VWYDLRNAFGSVHQDMLWYVLRLL GVERAFVERCDDIYEDSYFVVGNAAD GATEPVRQEVGVYQGCPLSPLLFIAA			

						LVPLLRALEKLDGVGVALADGVRPCT TAYADDLKVFSDSAAGIARCHAVVEQ FLEWTGLQANPGKCAFLAVTRNARG NPAHDKDMKLSLHDEEVSSIKLHDSY RYLGVGDGFDHVRHRLQLEPKLQQIK REAVALMQSGLATWQVVKALKTYVY PKVEYALRHLRPLQSQLQSFDRVVAK GLRHLLRLPRSATNEVLYAPTSSGGLG LQPLVEMHRALQIAHAWQMLHSKD PAIREVARAQVCQVARKRHRLREEH WRGRDDELVRCFLNSELAASPHAEA LRRHGDIGSLWSDVQRWLRIYHLSLV VQDDRNGLDPLGLRVPHHAKWLDH KSVLRHVKLHLKIRHQTRWKGLADQ GKTVRAHGGVGAKFVSTGAGLSDDD YRFGVKARLNQIDTNAVLKRKRLRSH KTCRDPTCSSAETLAHVLNHCEPNM DAIRQRHDDALEQIGSKIRNALKRGE STAELRLNQTVPEYTGAALRPDIVLRN VAAKKMVIADLAVTFEEHAAGARHS			

						SLQLSHDHKTLKYQPIVAELQLKGWQ VQTAAIVYGSLGSVQPSNSTPTRRS			
	Utopia-2_PS	GCA_009848525.1	Phytophthora sojae		TAGGAACCTAGCCGA TTGCAGCAGCACCTG CGCCCCGCCAGCACG CACAAGCGCCCCAGC CAAGACACCAGGGG CAGTCGGGTGACGG AGACCGGGGGGACT CCGTCGCGCTCCGGC CGCCGCTAGACGGCA ACCAGGCCCATTGCG GCCGACAGGGCCAC ACCCATGTAGGGAAC CGCCCTCGGAAACCC TGCACCGGCTCACAT GGTCCGACCCCCCGA TTGGTTCAACCCGTA CACCAACTACACACC AATTAGGGACTCCCA GGTAGTCACTACCAG	MDVDQGRVRPRQRRPSPRRPSERRP DGKRRRLNSEDDADAQKPAEHLQLA DEDGAENHTPVHTPSAASAHQAPVL AVYVHNAERFTCTLCTFTAASFSSLQR HRTTRHRRAVLQDKFLAGCACGTPIA SRPAAAKHAQACAGRSTALPATAPP VAGTSSPTDAAATATVAAATVTPDSP QPATPELAAPPPLASPTHPEVVEVQE QPSQGRWGPALPRALVAERVAARLG AIPAPRWEPPLPRELVSTRIAARLGEV PAPRWGPPLPRRMVASRIAARLLPPE PTADGETKEDTPADGADLTARDANQ ENNPAAAPDRTATAAAMDMDSGA DGEWLLRFDGACRANPGPGGAGAA LFKPDGSVVWTCSHYMPSSSETNNT AEYSALLLGSRAAADHGATSLRVEGD STLVIQQVRGIFAARNAILRRLRDQVK VELARVGSFSLHHIDRQANAHADRLA NRALDLRRTMMECGAHTAGDGCTN	cd01650	912	1175

					GACGCAACGGACTAA CCGGCAGGTAATACA TTAACCCTT	TTLDPMPVLPPQPTPRAPPPTPAADT NIEMQEPEDEDLADIDDGEVYAAMR IGPDAVPQRRPRLRLRQLTEAEDDAA REMVERLGATLAAKITDASDWESAE GYITALPHLLYDKLQPYSQAQRRPEPR PQRQRQQHQGSQCPRQQRQQGEQ RRQRPLHDQGQDRRSLEGDAHGDN ERDGDASQQDSARRRQCRRRGRSG AAKRRRRRRPPRVTRHHREHRLDEAL DDLHAVEQASPGDRTAVRRARRRVG RVNSAIDQQHLRHRFDTDEKACVEDI LAKARAQQETPRTAASSAGTASDPA APAAGATDDGTCPIPGDELWRYFDG VNTPSQEFAPDAPAGAAFRSAMTHL PAASRFRELLTEAPTQDDIESQLQHA HGTSSPGLDGVGYDVYQHFAAQLLP VLATAFQKCWEAKRVPQSWKLGVV RLLHKKGAREDPANWRPICLQQAIYK LYAGVLARRLVRWLDANDRHAPGQ KGFRAVNGCGEHNFLAATLVDQARR KRRQLYEVWYDFRNAFGSVPFALLW			

						DALARLGVPEDYVTMCKGLYESAAF VVGNAADGTTAPISLRVGVFQGCPLS PQLFNAAISPLLFALKRLPATGVPLSG DDCPGASAYADDLKIFSGTEDGIKRQ HALVADFLRWTGMAANPSKCSAMS IQRDRRGVLKADDLKLELDGSPIPALS MSASYTYLGVGDGFDHTRRRVELAP TLKQLKDDATALLQSGLAPWQVVKA VKTYLYPRVEYALRHLRPFQQQLAGF DRHLARGLRHLLRLPASATAQCFYAP VSRGGLGLLPLTELHAALQVAHGWQ MLNSKDPAIQRIARVQLRQIADARHR LDTQAWAGREEELCELLLNSQLGTSP DAPPKRRNGDIGSLWVDVQRHLRTL GLQFATAPACADTGTAAATLQLRVP HHDKWLDHRTVLRHVKLHIKHRHW TKWAAMRDQGRTARAHGGVGSSFL TRPRGMCEADYRFAVAARLNQLDTH SVLKRRRLRAHDRCRQPGCSRVETLA HVLNHCPGTMDAVRGRHDDALKTIE RALIASSSGDRQDRVELRVNQTVPSL			

						AGPALRPDLQLYNHTKKTVAVVDLA VAFEEQVGEDPDSSALARIAAHKRAK YDRIKRHLERQGWKVHLSALVYGSLG AVAGGNRSVYTEHLGLLKRDAKRLD WQKKEEDREDKCSSWPHEGWVKAP GDRTAIG			
	L1-2_SP	GCF_000002235.5	Strongylocentrotus purpuratus	TTATAGTAAGTT GTATACTAAAGA AGATATTGATGT TAAAATGCAGAA GATTATATTTTTT CTCAAAATATTTC ATCTCTCTCTGAC GATGAAAATAAA ATATGTGAAGGT TTATTAACACAC AGAGAATGTAAA CAAGCTGTATTC GCA	ATAAATAACTGTGTA TTATTGATGTTCCATT GATTTGCTAAATTGT AAGCAAAATTATGTT ACTGTTGAATATTGA ATAAACACCTCTAAG AGGCAAAGAGAAAA A	MKKNKTPGSDGITAEFYQVFWNEIC DILVAALNECYANEKMSDTQRKGIITL IYKKGDHRELKNWRPITLLNSDYKIIAT VLAKRFQKVIGKIIHENQSGYIRGRLA ACNVRLTKDIIEYMTRNGKNGALMM VDFSKAFDVIDVTFIKKCLSKFNFGPV FQKWVTILYTNISSSVIVNGWLSETFS VLRGIRQGCPLSALLFVMAAELMAN KVRENKEIKGLDIDDKGIDLKLLQYAD DTLFYVSDINSFENILIELEVFGIVAGP KLNREKTVLIWIGTACQRWSLSCFGL HWTDEPVKYLGHHIAPDLQCALKID WENKLEKMQKLLDNWRKRNLTLFG RLTIIKSLAISQVIHLMIVDTIPSTFLIKL NSLIFKFLWKSKIEKVKTFRLKWLGMF	cd01650	47	306

						LDNTTGNWKDMCKHWFDCLGGLH LLLNCNYTDYTLNLLKLRNVPHFYIEIL HAWYLINKYTSAECQTQKHVDENIL WYNQDIVFEKKMLFYEDWYKSGIIYL NDIFKNGTFVSVNELFIRLKTRRSKQF LIFDYTKLQKSIPKIWINDKHRVPLVFK DHNFSIPHFVFGKSVKPVSLLSSKQW YNMISFEQIKINKCCLYWEDVLQKDV KWSSVFNRNLIGIKDFRLREFNFKLLY NMLPVRNNLFKWGLSYDDLCLKCKK TEDVLHAFISCKANETFFAYLKDLVLR AFNLRIEIDIVLLLKMNSQTEIDLIMTI AFWSIYKMIIFRNKYGYDNRDLSLIYL FKREMTNRIEINEHLKVQNMKFPAN LLSYI			
	RTE-3B_SP	GCF_000002235.5	Strongylocentrotus purpuratus	GGGACAACGGC GTCTCTCAGAAC TGGTGTTCACAT GGATATATTCAT GGTCACCACACT GCGGTGAGGAA	AAGAAGAAGA	MESHNSKSKRNGVKGTSRNRGASLG RPSHVRQVGPDRHSASVRNRNIMRI GTWNVRTLYQKGKLDNVVQEMNR MRINILGISETRWTGSGEFTVEKHRM IYSGGIRHERGVGVLLDPFTANCVLGF WAVSDRVLLVKLQGKPFNMAIIVVY	cd01650	549	830

				GGCAACGGGAA ACCACCTTCACTA TTCTTCCCAAGAT ATACC		APTAESTDEDIDIFMNSQITMTAKSQ CKSNEIVIAMGDMNAKVGRGRYGG TVGPYGIGDRNERGNRWVEWCEAK GMVIVNTWFKEHPRRIYTWKSPGDR ARNQIDYITINERFKRGVKQAKTYPG ADCRSDHIPVVATLQCKLKKMTKAKS IQKVDFEQLQNRDVRLRYTVKVQNR FDSLTDQGEEVTWESMRDILVETAEE TLPKKEKKSKQRWMTEEILSKMSKR QKIKDRDGIEYRELDREIHDHCTKAKE EWLNKQCAEIERSFNANDPRVFKKV NDLSGKKLRSSKLGCIKAKDGSVLVEK EEILNRWTEYIGELFDDVRKEPPHCA GSIEGPRILLSEVRSAISMMKRNKAA GPDGVVMEMVEALENYGVMKLTEV INKIYDDGKFPDDLCKSIFITLPKKPGA VECGQHRIISLMSHITKIILRVLLQRAR NHLTPEISNVQFGFVKDAGTRNAVFV LRSIAERAVEMQKDVFLCFIDYTKAFD KVRHDQLFEDLTKIGLDTKDVKLLQN LYWEQTACVRVETELSDYINIKEGGS			

						ARMCHVSRSLQLLQVTTSTLKRGVR QGCVMSPDLFNYYSELILRNLEQERG LKVGGHNITNIRYADDTILIAESEKDL QRLLDVVVRESDKRGLSLNYKKTECM VISKKGGIPCSLKINNNQIKQATSVNY LGTTITDDARGEQEIKRRIALAKSTFSK LENILKNKKLKMKTRLRVLNCYVYPVL MYGCEAWTLTSDLRKRLESCEVWFL RRMLRISWTERVSNDEVFAMAGTRR KLLNNIRGRQLSFLGHVIRKNGLENLA LTGRIEGRRSRGRRRILWMSSLRSGIS ECGIDLGENELLHRIWDRELWHSMV ANVSGYGT			
	CR1-25_SP	GCF_000002235.5	Strongylocentrotus purpuratus	A	TAGGTGTTGTCTATT GTCTCATGTTATTACA CAACCTCCTCCCTCCA CAAAATAGCAGTAAT CACGGAAGAATCCTG GATCGATGAACAGG CTTAGACCTACAACA TCGTGTTTTAAAGTA	MSGETRYIYDKADYDAMYDKIMEID WNNEFWEMSMEKSWSFFKTTLLDIF NRYTPKKTLNVSQRKKHWMNRETLE VVREKHRHFRKWQEMKSEKNREKYS SAEIEQQRRCYSRSSNQARWQCRKA VSSYEKYIASGVKKNARAFWSYVNSKI TTRESVCNLTKPDGSKTETDQEKAEV LQDFFSSVFTQEDLQDIPLPLDVEFDT	cd01650	270	526

					AAGTAAGTAAGTAAAGTAA	PLDDIQITSGDVFELLRRLKPNKSPGPDKIHPRLLKELADVLALPISLLFQKSLESKQLPQDWKSANVTPIFKKGNKSDSKNYRPVSLTSIICKMMENLVRKQLLDHLMTNDLISVHQHGFVVGRSCITNLLEVLDQWTQILDDGGTVDAIYLDFMKAFDTVPHHRLLLKLESYGVRGNVLAWLESFLSGRQQRVVVRGKQSKWGKVLSGVPQGSVLGPCLFLLFVNDLPKVVNSNISMFADDTKLYVRADTADLRKQLQEDLSRMEEWAEKWQLCFHPDKCTVLRLGKLPAEASTDYCMKKGTDSVTLSVTKVEKDLGVLVDSQLSFKDHISQTVLKANRVLGVIRRTFMYLNTDTMPFLFKGLVRPILEYGQAAWSPYRLGEQRMLESVQRRATKIIPGLRNLTYQERLTKLKLPTLVHRRKRGDMIDVYKYVNGLYTVQEMPFKMSTEDRTRGHSRKLEKKYSRLDIRKYFFSNRVVELWNGLPENVVAAGSVNSFKNRLDSHWSKSVPVSLLYEPV			

	Crack-8_SP	GCF_000002235.5	Strongylocentrotus purpuratus		TAAATATAACCCAAG TACTCGCAAATAAGT AGCCACTATCACAAA CTTCAATATTTAGTAA ATTACCAATTGCAGA TCACTGTATTTCTCAT AAATATCCTACAAAT TACAAATAATTATGT TTAGACACACTGTCA CCTCAATATAATATTT TTTCTTCTTCAAAATT TGTTTATGTATGATTT CTTGATGACTTGATTT TTTTGGATTTTTTTTG TTTACTTGCTGAGTTT AAATTTACATTTGAG ATTTAATCTGTAATTC TTTCTGGTATGTTTAC TACTCACCATGTGAT GATGTTTTCATGTAT ATGTATGTCTTATATA	MIISKPITRLFNLSLRLGVVPQEWKQA RVTPIHKGGDKDNVNNYRPISILPTIM KVLEKAVHSQLHDHLLEFDLLNAQQS GFRKGYSTNTLLTYVSDNLHRNIDSG KLTGIVFLDIRKAFDSVDHSKLLEKLYT YGVRGNTLSWFESYLRGRYQCTLVQ RARSDLQVLQSGVPQGSILGPLLFTLF MNDLPNSVEHSKVMLYADDTAIMFS ALNPKEIESTMNSDLDRVKSWLCHN KLSLNTDKTNFLLVGTPQKLRHIEAPI DIKIDGLQIRQVEHVKYLGVWLDQHL NWNEHMKHTSSKIASRLALLRRVRKY LTIDSAKLLANALVLPYFDYCCTSWSN CNQGLRDTLIKLHKQMARIVLQMNV LTPSSEMFSKLKWMNMEKRWQFQK CKLMFDIVNGYSPSYLQDMFTQSAA THRYSTRIADTRGLVIPKVKTENGRRA FSFEGAMTWNNLPSELRNAQTKQSF CTSYFRNM	cd01650	26	284

					TCTATTTTATGTTATG TAGGGCCCCCAAGAA TAACAGTATTTTAACT ACTGGTGGGGCTACC CTGTTGAAATAAAAC GAAATAAATAAATAA ATAAATA				
	Daphne-19_SP	GCF_000002235.5	Strongylocentrotus purpuratus	ACCAGAAGGTGT TAGCTGCATGGA CTCGTGATGGTA TGATTATTGCCTC AATTCAAGCCTC TGGTGGAAAATC GACCTCTAAAGT TATACGTTCAGA TATCGACCTCGA AAACCTTTGATA GTTATGTCACGT TAGCGATATTAT GAACGATCGAGC TGACATTTTAAT	TAACTCACCAGGCTC TTTACCCCCCCCCCTT GTTTTCCCTCTTTTTG TCATGCTTCCAATCGT ACAACAATTAATTTA CCTTTTATTTACATTA TACTGAGGGATGGG CTCTAGCATAAGGAC CACACTCCTTTTTAGC TCATCCCTTCACTCCC ATTATGTTATGTTATT TTGTTGTTGTTGAAT GTATGATATTTTATGT GTGATTATTCCTCATT	MPGSTVNSIINDQSFVEFLSPRNGEN DFYGPSHGSYTCNEFNDLVNDSNDG LFSILHINCRSLSKNYDNISTFLSSLTVQ FTAIALTEIWIKEYNSYYDIHNYTLVLK GRQNKKGGGIGFYVRNDVKFNRRLD LDINTPYIESLFIEVVTANISAILGVVYR PPDQPIQSFLESVENVITTVNTENRIF YMIGDYNINLFKVNCQSHVKDFIEMF MAVGMYPLVHGPTRVVNNSISLIDNI FTNNLENCSSGTILTDISDHFPIFCIYS QSTQTVTETCITMKRQFSAKNINTFIT NIGNTHWILKENDANSSYDCFLKQF QHVYEICFPMKCQSKKKKRFDKPWF NDELRQLNRKKSLYYKKYLNNRTAYR	cd01650	550	804

				GACTTATATTTCT CCCTGGAGTGAG AAGTTTTTTACAT TTTATCTTTTTGG ATTTTTCTAATAC TCGAACTTTTAA ATGTAAAATACC TTTTGTATTTCTA ATTTGTTTTTTGT ATACTGCTTGT	TAATTGTATATGTTCT TGTATTTTTAATGATG AGTGTAAATAAACGA ACTTGAACTTGAACT TG	LDAYKKCRNKLTNRIKYVKYKYFKQSF VDAKGDTKQTWKIINGLLGTKSKKNE IPSTLQYNNETIDNDTDICESFNEYFIN IGPKLTSTLDIPTGDFKQYLKKMNLLL FFYFFFTPVTPREIINIVDKLKSNKCSG FDSVDVKMVKRVIHLICYPLCAIFNSS LETGIVPDQLKIAQIIPIFKSGTKTEMS NYRPISLLPVFSKILERAVYNRLYSFLN LHNILNNSQFGFRTGHSTSSASVDFT CQVSQVFDKSEVMMGLFLDLSKAFD TLDHNIILAKLFNYGIRGVLLNWFKSY LSNRKQFVKIKGNTSSMQTVVCGVP QGSILGPLLFILSINDMCNISDKLKYILF ADDTSVFMSHKNIQSLQQDFGNEIN KLLHWLSLNKLILNVNKTNYMVFTKQ QKIDFNEVDLQVNNHEFKRVSVNKFL GIEIDSKLTWNEHTGAVCNKISKNTGI LNKLRFLPKCVLRMLYNTLILPHLYYCS EVWAMTSQTNLNRILKMQKKAIRIIS NSHYRAHSKPIFYKLKLLNVFDICEYHI AVFMYLCRNESLPCALLKYFNLNSTF			

						HDYETRNASNFHLPQIKTTLMKKSISF DGPKIWNSIPINIRNCRSLQTFKRKYK MYLIESYVL			
	Daphne-26_SPur	GCF_000002235.5	Strongylocentrotus purpuratus	CCAAACCGTTTTC AGTCATAGGAGT CACAGAAACGTG GCTTAATGATAC ATCTCCATCTACT TTATGTAATATTG ACAATTATTCGA TTGTACGTTCTG ATAGAAAAACCG GAAGGGGTGGT GGAGCTGCT	TGACAAATCTCTTAA ATTAACACATGATAT CATATATGTTTGTCA GCATTTCCTTGATATA GCAGGAAAATCCCCC TCCCATGCACCCTCTA TCCCATTTTTTTTTTTT TTGGTTGCAAGTCAC TGTTTATTTTTTGTTA TTTTCCCCCAATTATG TTCTGTAATTTACTTT ACTTTACAATCGTTAT GATTTTAAATGTTTCT GGGCTGCAAGCGTAT CAAGCTGTCACGCTT AGGCTGTAGCCCTTC TGTATCATCACCTCTT CGAGCGTATGTATCT	MFISNLLTYKIRSDLSLKSDGTDSICIEI NMAQRKNILICTIYRPPHTNVELFLDE FDRMLHKVNLENKDVYLMGDFNIDL LQEDNTFHNMLRSNSFNAMIDKPTR LANNSSTLIDNIFTNVTNNYIGSGLLYS EISDHLPIFLICNHVRNYPEDEKNQA MIRKITAEKIAMFNDELLKVTWNDVV ACTDTNKAYALFWNLFSSLLNKHIPM TKKRSARNRIKQPWITKGIFCSIKKRN RLYKYSLRHPGEESLNNYKSYRNKLTT IIRQSRQMYFAKQFENANGNASSTW NVINQALKKQKSKSKCHITKDNNTM VDSKEIANCFNAYFVGVGPNQANTIA SNGVDFSQFLQTPCNKSLFLKPTNLIE ILNIARSLKISYANGHDGISTYLLKQVI GNIITPLVHIFNLSLTTGTFPEAFKIAKV VPIYKKDDSSQVSNYRPVSILSSLSKVL ERLIYNRLYCFLTENDLLNQDQYGFRK	cd01650	424	681

					ATGATTTACTGAATG TGGAAATGTAATTTA ACTTATGTTTTATACT CATGTAAAGACATTG TCACATTATTGTTATA ACCTATGATGATTCA GTAACCAATAAATAT TTCAACTATTTCAACT ATTTCAACTA	YHSTDLALAQLYDKVSNALASRKHVI GVFMDLSKAFDTLDHKVLSSKLHYYG VRGVALDWFNDYLSGRKQYVSYESV DSSLLPMKCGVPQGSILGPLLFLLYVN DISNSTTSLQYIIFADDTNVFCSNSDF RSLQRILNTELPKLSLWFRSNKLSLNLK KTNYIHFQSRKNATDKPIPEIVLDGEY LEKKQNTKFLGTIINENLKWSDHVHY VSKRVSRTVGLLYKLKHYVPLNILYML YNSLILPYMSYCNILWADSKVYTDKM LLLQKKAIRVCTNAGYRDHTDPLFAR LKTLKVTDINFLQKALFMFRLHNDSLP SHFPPMFCLNRNIHTYPTRQSNKYHL CNPKTTMAMKSIRHSGPDIWNSLSV DIRNLKSVYTFKKSVKSILLSKY			
	CR1-30_SP	GCF_000002235.5	Strongylocentrotus purpuratus	GTGAGGTGATG GCTACAGGGAGC AGACATGGTCTG GTTTTGGTGATA TAAAATCCACTT AATCTTGCACTTT	TAAGCTTACTAGCTT GGTTTTTACTTGCACT GGACTTAGCATATTT TACTGTAAATTGACA AGTGCTTTGTCCACA CACCATTCCAGCAAT	MKNYFLPPSSSRSHPLLLAILSIVCALK LLKSGNLYTQIADHRSASLPFFDLHAT LCEPRSLHQPTRRRTYTADIDNCKLLL FISSCNKWNRNTKLCPLDFRSTSLSHL TILLLLSGQVETNPGPSTSAPETFPCAI CGDEVRDNDPALLCDHCDCWCHISC	cd01650	746	998

				TAACTGCAGATC	TCATGACACTCCTTGT	VGISPDSYDILTKKSRSFAWVCCQCSS			
				GTGTATTTTATTT	AGTCACTTGCCTCAC	TNIGSSSSLQGLNLSQTNRFSPLHHDS			
				ACCTGATTTTGCT	CTACCATCCCAGTAT	SNEEILPSINPLTSTPTSSPGGRRQPRF			
				TCTTGAACATATC	AGCTTCCTGAGTGAG	HPPSRNRKKAFKVMVVNCNGLKGN			
				TTCAAGCTTTTGA	CTGCCAAAGGGATTA	AKKSSFHAAIAHHTPDFVFGCESKLD			
				TATGATTATTTTG	CCCAACCAAGTACCA	DSLSSYSIFPSNYSIYRKDRNIHGGGVF			
				TTCTTTGATACCT	AGTAA	IAIIDTLIVTECPEFDSDCEIQWCNVQL			
				TCTGGAGGTAGA		ANAKPLYIGSYYCPPNNRQQGLEGLH			
				ATATTGTGATAG		QSLSKIMTRHRRSHPNILIAGDFNHP			
				TTTTACAGCCATT		DINWDDQSTTNPATSTCHQKLLDILL			
				TTCTGCCAAGGA		HNSLSQTVREVTRPSSNNILDLVVTS			
				ATTTGAACACAC		NPALVENVCVKSGISDHNIVTFTLAA			
				ACCTGTAGTCCC		NPKISVKPLRKIYQFHKADEQQLRDA			
				ATATGGAGGATC		AAEFATEFLQSDPERRSVETNWSIIVN			
				CTGAGTTTGCTT		FLNRCLSELIPSKMSKGRKHLPWITPP			
				ATTTTGCCCAGC		LKRQMRKRDSLFKKAHSQSCSTAWT			
				CTTCTTCATACCT		AYRQFRNQVAKNVHRARCEYVNNII			
				ACACTACAGGCA		GASLKDNPKSFWSYIKTCRSENIGIPT			
				ACAATAACTGTA		LRTPSKLCASALDKAETLNDYFQSVFT			
				ATGTATATTTTCT		PHQLQPLVQKGPSPFPTIGHLHVHRT			
				GACTGAATTATG		GVEKRLQQLDPSKASGPDELPPKLLKL			

				CTTATATTATTCT TATGATATAGAA GCAATTACTCAC TGGGATTTAAAT CATTTACCAGAT GATTTAGCTAAC CGTCTACTACTAC TAATAGCTAACC TACATAAAATGG GTGTAGCTGAGT GAGTGAGACTGC TGAGCTTTTCTG ACTTCCCTTTGTT TGTTCTGAATAC GTATTGAATATT GAATACTGTCCT GCCAAATTTAAT CACCGGTATTCA TCACCGTTCGAC TTGCAAACGCCA TC		VAREIASPLSFLFQQSYNSGIVPAQW KQALVTPIHKSGEKCDPSNYRPISLTC VCCKVMEHIILSHVSKHVAVNNILTD AQHGFRRGLSTTTQLTTAVHDWSSIL QQRSQADIIFLDFKKAFDRVPHLLLST KLQYYGISGDSLQWIMSLLTDRQQA VVVDGSKSTWRDVTSGVPQGSVIGP TLFLLYINDIQDNIQSTMRLFADDSTL YREIRCPEDHQILKDDLQQLSRWSSN WMMDFNVKKCAILTITRKRTPSIREY HLSNGLIPRVKEYKYLGVTVAADLRW NKHCQTIRHKASRTLGLLRRTLSPCTR DVKSRAYTALVRPQLEYASEAWNPH SVTVTKMLEQVQRTAARFVYRDYRY TTSPSALVAALGWDTLHTRRILDQCT LLYKIHHRLVSIPAPTIVIPATYFGRHD HNLKYVIPVATIDSFKFSYYPRAIRIWN HLPGSAVNATGITNFKEAALPSIRMM QPPVGSHII			

	L2-1_SP	GCF_000002235.5	Strongylocentrotus purpuratus	AG	TACCTGTTCATAAGAGAATAATCTATAAAGTTTCACTATTTGTTTATAAATCCATGTATAACCTTATCCCATCCTATATTTCTAGTCTGTTTTGTCCCATCCAGCAATCAAGATATCCATTACGTTCAAGGTCAGACCCCCTTCTTCTGCAAATCCCCCGCATCAATTCAAAGAAGGGTCAACAAAATATAATTTTTTCTGGTCCCACTATTTGGAATAGTCTCCCTTTAAGCATTCGTTCTTGCACCTCAGTCGCATTATTTAAGAAACATTTAAAAACCCATTTGTTTTCTTTGTAATTGT	MHGHTLDLVITRCADNILHNCMVLDKGISDHYVISCDLNLQKSDAKRTLITSRNYRLINQESFGNELVAVLQSISQCGDVNEQVEFINKAICDVLDRHCPVSVRMKRDRPCGPWYDDNIHLLKRKKRKSERKWRKSNLRADRDIFVCHKDELNAAIVKSKSSYFTNLLLEANCKTMYATLHLLLGSSKKVLPYCDDQTDLANKFAGFFDDKISKIRTVLDEREQGLQTSGIVQSMDKHQCETGFSTFVDVSPSDIKMIISKSANKFCSLDPMPTWLVKDNIEVLVSPFVSLINSSLRSGVFPDKFKDAIISPCLKKSTLDSLVLANYRPVSNLAFLSKVLETIVSKQLVKHLDRCDLHEVLQSAYKANHSTESALLRVKHDLLLAIDKREGVIMVMLDLKSAFDTIDHSVLLRRLHSDFFLSGTVLDWFQSYLEGRSFAVSIGNHLSHKLSLKYGVPQGSVLGPLLFTMYLRPLGDILRNNGVEFHLYADDVQLYYSFNPKSSSSFQNAILLIEKCVSEVNVWMIKNKLLMNEQKTESLLIISSRLRHYLPDH	cd01650	306	575

						VHLLLGDAFVEPSESVRTLGVVLDSSL TMSQQVTSLCKSLNFHLYNISRVRRM LTDEACHHAIRSLVLSRLDNANSLLINI NVSDIDRLQRIQNRAARIISNARKYD HVSPFLNNYTG			
	CR1Y_SP	GCF_000002235.5	Strongylocentrotus purpuratus	TGTGTGTGTGTC GCGGGAGTCGG ATGTACACTGCA AATGCTTCCTGG AAACTGCCTACA TCGACACCCTGG GATGTCTTTATCT GGAATTTCTTCAT GAATTTAGGAAT TTCACCCTTCCTT CTCATGGCAGAT TGAATTTTATGTT CTTTTGGACCAA TTCTGTCAACTG AAATCTAAGCTA GATTTTGCCACTC	CAAGACCCTCCAGGA CCTGCCCATGCACAT AGCTACCAAGTCCTC CGTTCAAGAACTGAG CAGCATAAGCAGTCC TTCTTTCCCAGGACT GTGAGAGTGGAACA ACCTCCATCAGAAAC AGTATGGCCTCATCC GTCAATGCCTTCCGT CAATGCCTGACGGGA GGCAGCAACTAAGTC CGGGAACTTTTTACT TGCACCATGTAAATA GCATTGCACCATGCA CCTTGCTCTCGTAAG	MAPAQLCGSATSVEYTISAGQIYSFP VPSVATSAAKICHLNHTDIEPCTTNPD SKDEGESDDSLCLGSPKFSSSPRTLKK KPINPTKPKAKVLTILNVNCQSIRAKK QSFAILIHQHNPDVVIETESWLTEDIR SSEVFPTDHYNIARRDRPNDLHGGVF LAIRKDLLASREIELETDCELIWCKISM PGAKTFHVAMKLRAHLVQDQHARS QDFPCCDEMSLTELERSLSRLSPTHSV VIGGDFNLPGWDWRNQQVGNCRY PNLHRRLGSLLDDHGLTQVVTEPTRY QNTLDLLLVSNPTVVNEVTVVPGVSD HDCPLVKLNVRPTRPDEEESFTLQQS SVGEVWDQPGQMRRKVLLYSKAQ WEKFGESIEAVRDEIEAKVTYASVNEL WLIFKTGIEKGINDFIPQKLTPHKANL	cd01650	621	854

				ATTTCATTAAGA	CACACACACACTCAT	PWVTNGIRRLIKKRDRLSNECRHLRK			
				GTGCCATCAAAA	CAGTCTGCAAGAATC	REGLIPYVLEERLRDLKRTLQFALRKAY			
				TCTGTAGTCCCA	TTCAACTGTGAAGGA	WNYVESIFTVEDDDEGPGKYNCMKR			
				CAGCAAGACTCT	GGTAGACTTAACCCA	FWQFIKNNRNDSAGIAELKSDGVLFS			
				GATGTCTACAGC	AAGAAGAAGAAGAA	EPAEKADALNRQFKKVFTQEIPIPPHL			
				TTTCACATGGGC		LPDTSEHESMPDIVITLKGIEMMLKKL			
				TACAGTCATACA		KPHKAPGPDGITPRILKQLSRVIAPILR			
				TGTACAGACATG		IIYRLSYDTGEIPDDWHKANVVPVYK			
				TACAGGGAACCA		KGDRTDPSNYRPISLTCICCKMMEHI			
				CCTAGCTTCTAGT		FASSIMRHGDEKDILCPLQHGFQAKK			
				TCTACACTGTAC		SCELQLTRLCHLGFVTDLLNNMELSQ			
				AACTGGCAACCA		QTDVLITDFSKAFDKVGHQRLLKKME			
				TGGTCAATACTG		YYGVRGKSIDWISSFLPNRKQQVVLD			
				TAATGTATTATTA		GHTSPTVEVQSGVPQGSVLGPCLFLF			
				TTATTACATTCCT		YINDLPNQLSTKVRLFADDTVLYHTIQ			
				TTCTCCACTATGA		ETHEDAELLQQDLDQLCGWSSKWM			
				GTTGATTGAATT		MALNPKKCQVITVSRKKSQVCPSVFP			
				AACATGCCCAAG		QWRNHWSLCQLPNTLALLSLLTSTG			
				AAACCTAAATTT		TQHVSATCKKANNTLAFLRRNLRVNS			
				CCATGTGGTTCA		TQLKTTAYQALVRPLLEYGSAVWDP			
				TGTGGGAAAAAT		HTIRCVNQLEMVQRRAARYVLNRYH			

				GTGGCTTGGAAG TCAAGGGCATTC AGTGTGAAGGTG AATGTGAACAGT GGTTCCATGCAG AGTGTATTAACT TTAACTCTACAG ACTACAGTGCCC TAG		HTSSVSDMLSQLGWTSLQKRREHQR LTLLYKMHYSLVDFDIHHPTSNRQQS			
	CR1-7_SP	GCF_000002235.5	Strongylocentrotus purpuratus		TGATGTTACTTTTAAT AATTACTGTCGATGT TCAAACTGAATGTTT CTTTTAGTTTGCTGTT GTTACCCTGTGAATT GCCATTGCGTACGGG TAGTGAAAATGTTAT ACAAATATTGCTAAG TGTATGAATATTGTTT ATTTTTCTTCATAATA AGTTTCAACTTAATTA AATTACGTTTTTTTAA	MDFNLNLLQCDENTSCQDFLDLMLS KSFAPTITKPTRITEYSATLIDNIFVSGT PIVKSGVIVTDISDHFPIYTVISNVISHS NPNTNKCSGIRYLSDPDLRRLKEALTI TDWYAVYNCDDVDTSFDNFMEILM LTLNSFPTRQQNCSNYKKHPRLAWIS KSLLKSINRKNTLYYKYKKNHTVKSKL KYSRYRNILTTSLRLAKKNYFTKQFNV NKGDIKNTWRVINNVLKSRRVSTHIS NIMINENRIQDPLVIAEEFNNYFANIG PNLAQNIPVSDLHFQYFLETKNTNSM FCTPILEHEVIDIVKNMKTKKSAGND	cd01650	356	610

					TATACACAAGACTAT GTCATTGTTGTAACTT TCGGAGCCTATGACC TACAAGCATTGCTTT CTCTTTAGGCTCCATC ATATTTCTTCCCTTAT TCAACTTAGTCTTATG TATCATTCTTTTTTTTT TTTTGTATAAACGCA TCCTAAAATGAGTTC AAAATGTAACATGCT ATATGTAATGATGTA CTATACAACATGTGA GAAATATGAAAATAA ACTGAAACTGAAACT GAAA	GITNFILKEIIGEIVSPLTHLLNLSIMNG IVPKKMKIAKVIPIFKKGNAQDIGNYR PISLLTSFSKILEKLMYSRTMQFLIDFDI LSDTQFGFRQKHSTTHALLMFLDKVA HAIDDMAHTVGIFLDFSKAFDTINHD ILLYKLSHYGIRGKALEWFRNYLTDRK QYVSINGHDSYQRSVSCGVPQGSLLG PLLFVIYINDFRNSSEIMSFILFADDTNI FYSHRDPQYLLRTVNSELKRIQAWIN ANKLSLNVSKTHFMLFSNCIKSLPGNI LIDNTNLNQAESTKFLGIYIDSDLSWKI HVNYLCKLISRNTGILNKLKHEFPTHIL LSLYNTLILPYLNYGILAWGNGTKAQL EKLLIIQKRAIRNINHAGFLSHTSQLFF NNKVLRIGDVYVYSVGVFMYQLTINE LPAVICALFVRNNTIHSYSTRQRNLFH LPLTRTLFAQKNITFTGPKLWNDLPIE FTESTSLYSFKRKLKCHLLSNYALMVE EA			

	Hero-2_SPur	GCF_000002235.5	Strongylocentrotus purpuratus	ATATATCTTCACTCTCGATATATCGGAACTCTCGATTTTTCCTACCCTACAACTACCGGCGACGAATCACCTAACAACCGCA	TTTATATATACATTAATTTTTCATATTTGTTTTGAACTTGTAACCATTCTGTACCAAAATGACAATATTTAACGTAATACGCAATTTATCTCGATTCAATTATACACTATTCATTTATGTATCATGTATGGGGATCTTCTAGCTTCCTCACCATAATGTTACAATCTTCTTCTTTTACCCTGTTCTGTACTATACCACTTCTGTTTTCCTCTCCTTTTCTCCTTTTCTCCTTTTCACATCCTTTTCCCTCTCTTTATCCCGGTGCTCCTTGTTATCACATATTAT	MPPYVLIGSLLLFFTFQANCLRFNVPNLRATEWQPHNAAIKDRLTLLHHDLGAGIVTPKSAADDLTSILTEYFSGQDDFVDKKLSDTYIEHEPKSLIQARKCKNVLRKRAFGKNGTAEDRKAFRAAVKTVSYLRKQDKQKKCDKSRSHMEKRYRNNFYEFSKAACSGSLNESPTKPLFSEDVANNYFGDKYSTPNVINIGSLNWFPYVNINTPDRFEFDMTPIKPKDVKNVLRRKKAMSAPGNDGLMYGMIKHLDSVHHFLATLYNKILDSGDPPESWSESKVTLIYKSGDTDKPENFRMISLTSCISKVFHQIIAERSMSYLTENNFIDKTVQKAFIDGVNGCIEHNIVLHEILSDAKCKKRSLHTTFFDLADAFGSVSHQLITSSMERYAIPTPVARYVATLYSRLNGRVRGPGWITKPFEFKRGVFQGDPLSPTIFLMCFNPVLEYLKEQQDQFGYKWGDRKIITTPYADDFNLITGNKLQHQKLMNGINDRLKSMGLTLKPRKCSSLSLTSGKPNIVDFTIGSDVISSVHGHPVKFLGSLITFSMKTS	cd01650	280	545

						EIFEYLKGKIESTLENINKCNIRNEYKLC ILTRYALPSLRYMLTVHDVHASHLKSL DIILDRYIKMWLGIPSRGANLAVVHD PDGLNIPSISDLYRTCHCLAYSRSRVK GDDVVNHALDRRLERERGWSRQTSF SLEAHEIHSKIPDSVSWNQTKTVIKKA VQDVCRENWKRVIEPLLVQGEFMKV VHMENTDVTWKSIIYNLPQSVMKFS VNAAIDSLPTNKNLYRWGKRISDKCN LCNCTGTLHHVLNNCERMLDRYTWR HNNILRIMLKHIKNTSNNSTKVYADIE GESINGGTIPPNIIRTQHKPDICHIDNK NKDIIIIELTVPFETNIDKAHQRKQDKY ASLTHDITSTGWKSKLICIEIGSRGLITP QNKYRLKEILKITNSTRYPESRDEILKA VLFSSYVIFQGRREPAWQFDQYL			
	Crack-2_SP	GCF_000002235.5	Strongyloce ntrotus purpuratus	GGTCTGTTGTAT TACGATATTTCTG ACCATTTACCGA TATTTATTATTTC TAAAGTTCACCTT	TTTCAACATGCAGTA GGCAATTGTCAAATT TCGGTCACATTGCAA CTCAGTTTTGCAAATT TTGTTTTCTATTTCGT	MHRKQTKHNIESLNFDLAKEDWVEV LCVSDVNVAYEKFLEIVLHYYHKNIPL VKIKKGKKAKQPWITKGILRSIVARNK LYKEALTSQCNDKLKKYKTYRNHLTSL IRLSKKLYYSNNIESNKSNNNLLWKTV	cd01650	263	519

				GTTAAAAAAATT ACTAAAAATAAT TGTGCA	ATTTCTTAATTTCCTT TATTGTAAAGGCTGT ATTTGTTTCTACTCGA CTATACCACACAGAG CTACGTACACTTCCA CTATTTGTTACAAAC ACATACAAACACGCA CGCTCGCGTGCACGT GCACACACACACACA CACACCCACACACGC ACACACGCACACACA CAATATATATATATAT ATATACACTTGCACA ACACACTCACCAACT CATACACAAAGATCG CTTGAACACTTTTTCA TTTTAAGTTTTACCTT AAGATTTTTTTTTGTT TTGTTCTTGTTTTTTTT TAGGCCGCCATCCTT GTCAAGCTTTGCTTA	KELLGKKKTESNKCFLIDEQQITDPIKI SKAFNEYYIDVGPNLAANIDSGDMD YNKSLNFPRENSLFLTPTNDYEIINVV QHLNSTKSSGHDSISVHLLKQIIFNMV TPLTYIFNLSISTGVFPHLLKIAKVIPIYK KDDSSVISNYRPISLLPSISKILEKIIYKRL NTFLQKNNSLIPNQFGFRIGHSTDYAI LHMYDKIIEAFANSEHIVGVFMDLSK AFDTIDHKILIHKLQYYGVRGQALSW FTNYLQDRQQYVSFQNHESTKLTIKC GVPQGSILGPLLFIIYINDIIETSSLLNFII FADDTSISYSHKNFDTLITRLNCELMKI SRWFKSNKLSLNLSKTNFIHFKQTNV HDMRNCNIEIDGLPLIEKQSVKFLGVS IDSHLNFNEHIRNIHTSISRNIGILYKLK QILSERSMIILYNSLILSHVIYCNIVWG NCSLTRINSLFLLQKRAIRMITNSHYIS HTEPLFQRLKTLKIQDIHIFQTCIFMYK YSHNQLPSLFHNIYSLNSNIHTYPTRC STNYHLENPKILLAQKSVRHHGPDV			

					TGATTGCGGTCTCCT GCACTTTTCCACCAG AATAGTTAGGATAAT ACTTTAAATTTATGAT AATATGTTATGGAAA TAATAAGTTTCGTTAT TATAATTATGTACAT ATATTATGTAAAAAA ACATTTTGACATGAA GAAATTATTGTTTGTT TATGTTGTATTATACC TTACATGTATATTATG TTATTTGTTATTCAGT TATATTTTATTTGGAA AAAATGCAGAAATAA AATCAAATCAAAATC AAA	WNSLPNSIRHCSSLYSFKASLKKTLISK YLHN			
	RTE- 11_SP	GCF_00000 2235.5	Strongyloce ntrotus purpuratus	ACAAATGGATAC AAGATGCAAAAA AGTGGAAACCTC ACGTCAGGCTAC	GTCAAGTAAGTCAAT TA	MPAKPCRESIRIILQRRIMKKTEEVLSE AQAGFRPGRSTMDQLFTLRQIMEKR LERNQSLLCCFIDYEKAFDSVWQEGL WKAMNFFGYSSKYIHLLQALYQQSSS	cd01 650	12	235

				CAAATCCAAGGC AGTCTACACAAC AATCAAACAAAT CACTAAGAAATC TAGCATCAGAAT GCAGGCAGTGA AAAGCAAAGATG GAGAAATCCTTA CCGAGATGAAAG ATGTTAAAGAGA GATGGAAAGAA AGCTTTGAGGAA CTGTACAATCAC CAGAACCAAACA AATGGAGAAGC ACTACTCAGATG GCCAGCCTCGAA CCGGTTCCAGAC ATTCTGCGAGAT GAAGTAGAATTT GCTGTACGAAAA		AVRVNGELTDWFKTTVGVRQGCALS PQLFNILLEAVMLYAIHDTKIGVRVQ GQLINNLRFADDIVLIADNDRDLQTIV DLVHQSSSNFGLKINITKTEVQAISKQ PQLLNINVGGEQLKQVDKFTYLGGTI SQTGSCSEEVKHRIGKAIGAFQALKKI WASQDIRCSTKVELYKVLVLSILLYGA ETWTLKKVDENRLHTFEMACLRRIM GVTRFDRLRNTHIRTQLNMEETIIDR VATKRLRYFGHINRMNSKRYPHILLN GNIHGKRPRGRPAKRWTDCIKADCK NRQVDSLTKATRLTEDRKVWQAIVK QKPSYIPVMVWTA			

				CTCGCAGAAGAG AAAGGATTTGAC AGCATCACAGGA GAGGAACTCAAA GCTACTGGGCCA ACAGGAATCACC ATACTTCACAAA ATATGGCGGGAA GAAAGTTTTCCG GCTGACTGGGGA AAAGCCATAATA ACACCAATCTAC AAGAAAAAACAA ATCAACTGTGGA AACTACAGGGGT ATTAG					
	REX1-2_XT	GCF_000004195.4	Xenopus tropicalis	GATTCAAG	TGACATTCTGCCCAA ACCAACACACACTCC CCATAACTACTGGAC TCTTCGCAGTGTCAC TTTAAGCAAAGCCAC	MAPVRMAAVAQCPPQLCFFLMFVV LFFSFFATCNALISYDRQTLLDIRQNS QSGPSQTEFFMNRGDLPWLPWQKK HRRKRGKRAGILVRLRNRQHRAPLPS ILLANVQSLENKLDELRARITFHRDIK	cd01650	564	818

					TTTAAATCCTCACTGC	NCNIFVFTETWLVPSIPDSAIVPEGLSI			
					ACAATTTCAAATATT	HRRDRTIDSGKCRGGGVCVLTNKRW			
					GCATTGCATTTTATTA	CTDVQTVSSGCSPDLEYLMIRCRPFYL			
					TTGTAAATACTTGTA	PREFTAIFCTAVYIPPHADTTRALDELF			
					CCTTTATTGCACACTT	DAIDRQETLHPEAAFIVAGDFNKAN			
					TTATTATATTTAATAT	MKKVLPKFYQHVNFPTRGENILDHV			
					TTGTATTTTTTTTTTT	YTPYAHAYKARPRPAFGKSDHSSVLL			
					TTGTTTTTTGTTTTGT	LPAYRQKLKREQPMMRSVQRWTDQ			
					TTGTCTATGTAAAGT	SDSTLRHCFNITDWEVFRTAADNNIN			
					TGGGAGGAACACGG	NYAEYVMDFIHKCINDVVPRVNVRT			
					GTCAAAAAAATGTCA	YPNQKPWVVNGEVRAALRARTAAFKS			
					TTACAGTAATGATGC	GDPSEYKSSRYALRKAIKTAKGQYREK			
					TTCATTGTTATTTGTA	VESCYSNSNIRNMWSGLKTITDYKGK			
					TATGACAATAAACTT	SCRVGEVTASLPDELNTFYARFDKHLS			
					GAAACTTGAAACT	QELLLEEDGRPYQISQADVCKHFRKV			
						NTRKAPGPDNIPGRVFKACAHELAD			
						VFTDIFNLSLLQSAVPTCFKRTTIIPVP			
						KKMNVSCLNDYRPVALTSIAMKCFER			
						IVKAHITSLLPASLDPLQFAYRSNRSTD			
						DAIAIALHTALSHLDQRNTYVRMLFID			
						YSSAFNTIVPSRLVMKLRDLNIGSSLC			

						SWILDFLTNRPQVVRIGNITSSTLTLST GAPQGCVLSPLLYTLFTHDCTATHSS NTIIKFADDTTIIGCISDGDESAYRAEV RALTSWCRDNNLLLNVSKTKELIVDY RRLQGGGHTPIHIEGAEVERVSCFRFL GINISEDLSWSHHVGVITKAARQRLF FLRRLRRFGMDSRILTNFYRCTIESILS GCITTWYGSCNALDRKALQRVVRSA ERITRTELPAMQDLYRQRCRRKMQRI LSDPSHPSHRLFTLLPSGRRYRSIQTRT SRYRDSFYPQAIRLLNC			
	LINE1-38_ZM	GCA_902373975.1	Zea mays	GATGGCCCTCGT CTCTCCCGCCCCC GCCCTCTCGGCT CCCGACGCCCCT CCCCCTCCCCAG CACCAGCACGCG CCGGTTCCCATTC CGGGGCCGGCG GCCAACTCCCTT GCTGCCGTCGCC	TGAGCTCAGGATGTG GTGTAGTGCCGGGG CAAAAGGGTTGAACC TGCTTTCAACTCTTAT GTTTATGCTTGACCC AGGGGGATAGTCTA GTGGCTTGTTCAGAG TGGTTGTGTGTGGTT GCTGTGTTGCGTCTT TTGGTATGTTGTGTC	MLGVSAAQGNGGPASSRRKRWRKK KRLEREDEPTLEIVAPPPPVQMLPAS DATACVIDWSVQMSIAEDNLKTAAV ITVVCSRDISANEIVSVLAPRLEMEEG GLVLRQLSQSSFLAVLPSRVQLDTLVD SFPIIREENFMLSSRRWTRFGGAGGN SLKHLVELELRGLPVHAWETSVVQQL INPHACISQVLPDTLEVRSLEAFRCLA WCSDPERIPATKDLWITEPDQASAAE GKKALVYPVQIRWVPADLSLLTGDPP	cd01650	1115	1378

				GCAGCACCCTCT	TTGTGATGTGTGTGT	LLIGGDDQDKSDTPPSRRRCSDNPPR			
				AGCCCTGGCGCG	CGTGTCAGGTGGTAT	TQHDAINDVPGNKTHTQRRPVHDRL			
				TCGGCTTCTATCC	GTTGTGTGCGGCGTG	GPLRRSDSPRELSHPRVQIDFEARSSQ			
				CCTCCTTGTCTCC	TGTGTGTGTGTGTGT	PCVLSDAVADSCVGNPPTPVPDGVQ			
				TCCCGTGTCCCAT	TGTACCTTCTTCTATC	CLAAGDVLLRSEEGCEDAEVIPMTP			
				GTAGCCAGGGG	TATCTATTAATGGAA	MAVSSPCPSSGGRRRLLLKTATKQAK			
				GCGCTCCAAGTC	TGATACGCAGATCTC	FRNPPIAFFRRRRRAPRTPEGATLPPC			
				TCTAAGGTGGGG	CTGCGTGTTCGAGAA	SSGDRGAGSSLGALPPAPAAGVAASL			
				TGACCTCAGTCC	AAAAAAA	LPHGSPPLVESHGSPGQQDFLRRVSK			
				GGCCTCTTCGGA		AVGAVLPLPKAVGGGRRRRQPTSLP			
				TGGCGTGACGTC		DAIRRSRRVAGLGVEQVQHIPKARRT			
				GCTCGACTCCCG		IMKSILAEPDSEHFSQQNLDAYAKLF			
				GCCTTCCTTCAA		QAPLSRAQVEALAALFGWVMPDDE			
				GGACGTGCTGGT		GEVVPPGTATLVCLQPTVAYYSNCCP			
				GAGACCAGCGA		MDPSLILFWNVRGLNSAARQDVVRS			
				ATCCTTTGCCGCC		LVDSLRVDVVCLQETKLVNCDRRTVL			
				CAGCCGTGTGGT		SMLGADFDNNFICLPSVGASGGVLIT			
				AGTCCAAGAGTC		WRAKLGRIVASRVDSFSVSVKFNPAD			
				AGCCCTGGCAGC		GVAWWLTCVYGPQGHDRKIAFINEL			
				TGGAACGTCCTC		RAIRAACTGPWLLGGDFNLIFRDEDK			
				GGCGCTGGGGA		NNININRPMMGRFRRCINDLSIKDID			

				AGGTGAACCATC		LLGRKYTWSGANNVQVKLDRVFCSS			
				GGCCGCCTTGCT		DWEEKFPNYLLQSSASEDSDHCPLLL			
				CTCACCCCCTGTC		GLKDNHTGKRRFHFESFWPRFDGFL			
				TACAAAGCGGCG		EAVSNTWGSVQQSSCPLKTLSLKFKA			
				GAGTCCCAGGCA		TTKGLQSWSDRKVGNFKLQLELAREI			
				GGCTCCCCACTC		LLQLEVAKESRILNVLETWLFGRLKKC			
				TACACCGGCTTT		SLSLASLIRSMARTRSRIAWLKDGDA			
				CCGCCCAGCTCC		NTRLFHSWSRYRKRKNFIAKIQDGEQ			
				AAGTTTTGATGA		LVTSHSDKAEVLRSFYSDLIGTGQQRS			
				TCAAGGATGGAC		RTVNLDTLEMQHTDLQGLEQPISEEE			
				AACGGTGGAATC		VWNTIKQLPPDKAPGPDGFTGRFYK			
				CAAACAAGGTAG		VCWSIIKKDIMAAIMAIWNKDFRNF			
				AAGACGCCGGTT		RYLNTAFITLLPKKEDAILAKDFRPISLI			
				GAGGGCTTCACG		HSFAKLITKIMASRLSGYMDKLVSRN			
				ACGCCCGAGTCG		QSAFIKGRFIQDNFMLVQQTARLLHA			
				AGTCCCTGCTGA		QKQPRILLKLDISKAFDSVSWSFLLEVL			
				GCTTGTTGGAAG		EKVGFGPIWRDVISGLLMTSSTQILLN			
				GTGCTTCAACTG		GVPGEFIQHLRGLRQGDSLSPLLFIIV			
				CTTGGGCTGCGA		MDVLNGLIVKAADEGLLQPLSTRSIQ			
				TCACAGGGCTGC		HRISLYADDVVIFLRPAASDLHLVTRIL			
				GGGGTGTGACA		ELFGDASGLKTNVQKSSVAPIHCSPV			

				GTGAGCCTCGGT GCTTCCTCTGCTC GGAGCTTGGCCA CAAGTCTTCCTCC TCCTGCCCCAAG ACCCCGGTGTGG AAGCGGCTCGG GCGCCCCAGGCC TGTGAAGTCCCG GTCTGGGAGCG GCTTTCACCGGC AGTTGAGGGGA GAGTCCAACGAG CGGCAGGCCACC GTTCGGTTTGGC AGCGTATCGCTA CGCCCGAAGCTG TTCCGATCGCCC CTTCTGTCCTGGT GGGAAGC		DMEMVTQIFPCKVETLPLKYLGLPLSL RKLSKEQLQPLIDRLADLLPNWKADL MTRAGRAIQVQHVLTATVIYHAMAL DLPKWAIKAIDKIRRGFLWKGRKDVL GGHCLVAWGKVTRPKELGGLGISDL QKLNWALRVRWLWLEKTDPSKSWA DLPLQSSSILKSILSTALITEIGDGRNTL FWRDRWLMGLRIEDLTPHIFAIVPTR ITNRRTVAEGIVEMAWIRDLRGTITW DVLSDFLTLSEVIADYSINVGVPDKHI WKFSPTGQYSSKSAYDMLFLGSTSLG AYERIWESWAPSKCSFFLWLVMHNR CWTADRLARRGLPHPSVCPLCDQEE ETIQHLLISCVFARQFWYSLLSRSGLP NLSPDNASCSLDDWWSRAVEKIPRD LQSGFNSLVILGAWSLWRHRNDCVF NGEIAKPGECSHPRWR			

	LINE1-24_ZM	GCA_902373975.1	Zea mays	CCTGCATGCTTG AGTTCTCCACTG CTATGGCTCGGG AGGAGGCGGCT CTTCGCCGGGCT TTGTTCGTCTCTA TTGTTGGCAGCA GACCGGAGATCA GGGAGCTGAAG TGCGGGATGAA GTGGCTCGGGCT TTTGGAATTGAC GTTAATTCAATAT CGATTCATCAGT CCTTGCCAGAAG ATTTCCTATTGTT GTTACCGGATGA AAGTGCTGATAG AGTCTTCAGCGG AGGAAAATTGTT TAGGGGCCCTCT	TAATTCTCCAGGGGG ACTCTTCCAAGCTTTA TCTCCTTAACTCGGC GTATGTGACTCTCTT ACCAAAAAAGCTGAG GCTGTGGAGGTAAA AGACTTCAGACCGAT TAGTTTAATTCACAG TTTTGCCAAGATTCTC ACCAAGATTCTGGCC AATAGGCTTGCTGCA ATCCTCCCTACTTTGG TCTCCCCCAATCAGA GTGCATTCATCAAGG GAAGAAGTATTCATG ACAATTTTATGATGG TTAGAGAAACTGCAA AAGCGATTCACAGGC AGAAAATTCCTAGAA TCTTGCTGAAGTTAG ACATTGGCAAGGCCT	MNLSKILIWNVRGLNRKARRDAVRMMVESTRPDIVCLQETK KESISRFMVFSLLGNVFDQFTFLPAIGTRGGILLAWKGSVCSA LQTRVDEFSISVQFNQNDSLPWWFTGVYGPQSDEKKLQFL QELRMVRSVCVGPWAVGGDFNLIYRAEDKNNNNLDRAM MGRFQRIINDLQLVELPLLGRKFTWSNERSSPTLVRLDRFFYS RDWEDLFPDCLMQSSSSSISDHCPLVVGLHEFTQGKRRFHFE SFWTKLEGFMTEVENSWCQPIAVGCPLQRMADKLKRLSKH LQSWSQKRVGHVKRQLVMTNEILHRLEVARDSRDLSYHEE WLRRKLKHHSLGLASLERTMARHRSGLNWLKEGDANTAYF HHHSRYRKRKNFIGKLKVEDRIVVDQQEKKHAVWNFYNNLL GVAASREFTLNLGVFHRPASNLEVLDQIFSEEEVWNTIKSIPS DKAPGPDGFTGRFYKMCLDGCQRGFLGSCQ

				TTTCAATCTTCAA	TCGACTCAGTTTCGT	
				TTCAGACGGTGG	GGCCTTTCCTCCTTGA	
				TCGCGGTTGGCT	AGTTCTGGAACATGT	
				CATGCTGAGGCG	GGGATTTGGTCTATT	
				GCTAGGTTGCCA	CTGGCGTAATTTAAT	
				GCACTGGTGGAA	TGCTAAGCTATTGTG	
				GTCGAGGTTCAT	TTCCTCCTCTACTCGA	
				GGCATCCCAGCC	GTGCTAGTTAATGGG	
				CACGCATGGGAT	GAACCAGGAGATCTC	
				CGTTCTACGGTG	ATTTGTCACCAGCAA	
				GAGTATCTGCTC	GGCCTGCGGCAGGG	
				CGTGACTCTTGC	AGATCCTCTCTCCCCA	
				ATTATCACAGAC	ATGCTCTTTATATTAG	
				ATTCATCCGAGC	TAATGGATGTGCTGA	
				ACGTCATTGAAG	ACTCTCTTTTTCAGAA	
				AATGATCTGTCTT	AGCAAATGAGTTGG	
				CCTTCAGGCTTC	GCTCTTGGGACCTTT	
				GTGCATGGTGTT	ATTAAGAAGGGGAG	
				TGACACAGATTT	GGGGCCAGCGGGTC	
				GTTCCCTCGGCC	TCTCTTTACGCCGAT	
				GATGAAGCTCCT	GATGTTGTATTATTT	

				CATTGTTGAGCC	GTTCAGCCTAACAGA	
				GGGGACGGATG	AATGAGCTGAGTGTG	
				TGCAGGAGAAA	GTCAAGGAGCTGTTG	
				AGATGTCTTTCCT	AGGGTGTTTGGCAA	
				ATGATATTGAGG	GGCATCAGGGCTGAT	
				TGGCTGTTACTG	TACAAACATAAATAA	
				CTGTGATGGTTC	GTGCAGTATGACACC	
				CGACGAATTTGA	TATTCAGTGCAACGA	
				CTCCAGCTCATTC	GCATATCATTGCTGC	
				CCCTGGCATGGA	TGCTCGACAGATGTT	
				CGGGAGGCATG	ACTCTGCAATGTTGT	
				AGGATGGCGGC	TCCTTTTCCATGTAAA	
				GGGAGGCATTCG	TATCTAGGTCTTCCTC	
				GAATCCACGTCG	TCTCTGTTAAAAAGC	
				GGCATCGCGCCC	TCCCTAGAAGATCCT	
				AATTTTCCAGCG	TTCTGGAGCTTATTG	
				TCTGGGCCCAAC	ATAAGGTGGCAGCA	
				GTTCGTGGTGCC	AAGCTTCCTGGCTGG	
				CAGCGCTGCGGC	AAAGCTGCTAATCTT	
				AGCGGGCCCGG	AACCTGGCCGGAAG	
				CTGCAGGACGAG	GGCAACTTTGGTTAA	

				GCGCGCGGTCCT	AGCTGTGCTCACTGC	
				TTGGACGAGGA	TATTCCCATTTATCAT	
				GGTGGCTGCAGC	CTAATTGCTTCAATGT	
				CCCCCTGTTGCT	CCAAGGTGGGTAATC	
				GCGGGCCTGTGC	AAGGCAATAGACAA	
				GGCGGCTGGCG	GATTCGCAGGGGTTT	
				GTTATGGCGGGA	TCTTTGGAAAGGAAG	
				GTCGAGGAGGT	AACTGACATTAGAGG	
				GGCCGCACAGG	AGGACATTGTGTCGT	
				AACCGAGGGCCT	TGGGTGGTCCAAGGT	
				CGCCGTCCAGCG	CTGTCGTCCCCTCTCT	
				CACTGCTGGGGT	CTTGGTGGGTTGGGA	
				GCTGGCGGAGG	ATTCACAATTTGGAA	
				AGGTTGGGCCAC	GTTATGGGATGGTCT	
				TGGAATGGGCCA	TTAAATTTGAGATGG	
				GGGAGGTGTTG	CTTTGGCTCCAGAAA	
				GCCCAGATGTCA	ACTGAACCAGCTAGG	
				CGGCAGCAGATG	CCTTGGGCAGATTTT	
				GGTCCCGGGACA	AAAATAAAGGTGCAC	
				GTTTCGCCCCGC	AGATGCTTCAGCTCT	
				TCACTCCTGTGC	CTTCGCAGCCTCTGT	

				AACAAGATGATG	GTGTTCGGTGGTTGG	
				GGCAGGGCTGT	TAATGGGGCAAACAC	
				GTTGGCTGGGTC	TCTCTTCTGGACAGA	
				GCTGAAATCGCT	TAAATGGCTTCAGGG	
				TGGGTCGGGCTC	CCAATCTTTCTCAGAT	
				TCCCCGACAGAG	TGCCCCCGCTCTAAT	
				GTTTGTCGGGTT	AAATCGGGTGCCCGC	
				GCAGAGGCTGAC	GAATATAAGAGGCG	
				CAAATTTCGCCC	CAGATCTGTCCAAGC	
				CCTTCAATGTCG	CGCCTTAACAGCCAA	
				GCTTGGTCCCCC	TGGTTGGGTGCATGA	
				GGCTGACCTCAA	TATCTCTGGAAGCCT	
				CTCAAAATCTTTG	TACGGCGCAGGTGCT	
				GTTGACTGGCGG	GTTTGAATTTTTATTG	
				ACAGAACCTCCC	GTTTGGGATTTACTC	
				TGAATGTTTATG	CAAGAGGTGCAACT	
				CTAGGAGGCCAG	GTACCCTAATGTGGA	
				CTGATTACAGAT	GGATCAACATCGGTG	
				TGAAAAAGCAGA	GTTGCCATCCTCCTC	
				CTTCAGATACCG	GGGGCTGTATTCCGC	
				CACCGGTTCAAT	CAGCTCTGCTTATTTG	

				CCCTTCGGCCAG	CGCTTCCATGGTGGT	
				ATGAGGTGTCGG	GCAACAACCTTTGAG	
				CACCTCTGGTAC	CCTGCTACAAGAATC	
				CTGCTGGTATTG	TGGAAGACTTGGGCC	
				GAGAGGCGACC	CCCCCGCGCTGCAAG	
				CCCGTTATTGGT	TATTTCATTTGGCTTG	
				GGCTAATATCCA	CATCTCTCAACCGAT	
				AGAACATTGACT	GTTGGACGGCTGATC	
				GCTTGCTTCCGC	GACTGGCACGTCATG	
				AACCTCCAATTC	GAATGGATCATCCAG	
				AAAAAAGAAGA	ATAGTTGTCCCTTGT	
				ACTAAACAGCTT	GTGATCAGCAGGAT	
				CCGCCAGATTTT	GAGTCAGTGCATCAC	
				GTCCCACGTCGC	TTGCTGGTGGATTGT	
				AGTAGTCGCTTG	GTTTTTACGAGGGAA	
				TCAAAGAAAAGG	GTTTGGTTTACAACA	
				GAAGGCTTAAAT	CTAAGCAGGCTAGGT	
				TCAACAGTCAGA	TTGCAACATCTTGCTC	
				CAAGTCCAGGCG	CCACGTCTGATGCTA	
				GAGCTCATTTGA	CAGGGTTTCAAGATT	
				AGTTGAACGTCA	GGTGGAGTGAGGCA	

				CAAACCAAGGTG	GAAAGACAAGTGTCT	
				GCTGTGACTGAT	AAGCATAAGAAAAA	
				GAAATGCTGGAG	AGGTTTCAACTCTATT	
				GAATTTGGCAGC	GTAATTTTAGTAGCT	
				TGTTTGACAAAC	TGGCAGATTTGGAAG	
				CCCTTTCCAACTC	CATAGAAATGCGTGT	
				TCACATCAAGGC	GTGTTTGACAATGCT	
				CCTGGCAGCTTT	TCCCCAAACATTCAC	
				GTTTGGATGGTC	GCAATTTTGCTGTCC	
				GGTCCCAGATAA	ATCCATGATGAGGCT	
				TGTGCAAGAATG	TTATTATGGGGTCTT	
				CTTGGTGTTCCTT	GGAGGGGCTGCTGC	
				TGCTAGGGAGAA	TCTTAGGCAGGTTTG	
				GCATTGAGTGTG	GTTGTAGTTTTTTTAT	
				CTTGCTGCTGTA	TTTTGGTTTTTTTAAT	
				GCCTCACTTAGT	TTTTTGGGGTGCTTC	
				GGTTGTATTTCT	CTGGGCACCTTTGTA	
				GTTATCA	CTGGGTCTTTTCTAG	
					ACCCTTTTTCTTATCT	
					TCTTAATATATTTGAT	
					GCGTGGTTCTCCCAC	

					GTGTTCGAGAAAAAAAAA				
	LINE1-41_ZM	GCA_902373975.1	Zea mays	G	TAATTACGTAACCTCTATGTACCCTTCCTGCGGTGTCTCTTGGGCACTGCTTCACTCTCTTCTTTATTAATAGAAACGTGCACAATCTCGTGCCGTTCGTTCAAAAAAAAAA	MADVMAVMNSLHDQRFLNFDLLNRANIVLLPKHDGAENLSSFRPISLIHSVAKLFMKLLSMRLAPTMRDIISKCQTAFIKGRSIHDNFLFVRNMARRYHRNGTPMLLVKLDISKAFDSVRWDYLLALLAHLGFPPKWRDWVASLLATCSSQVLLNGIPGQSITHGRGLRQGDPLSPFLFVLAIDPLQRLLHLATEAGILSRVAKARTRLRIAMYADDAIIFVKPEKRELDSFATLLHQFGEATGLRTNIQKSSVVPIKCAGLNLDEILAGFPASRSQFPIKYLGIPLSVERLKKVDFQYLLDKVSSKLTCWQGRNFTFAGRLVLVKSVLSAQPVHTLTTINVPKVVLEKIDKFRKRFLWAGNENLTGGKCKVNWPTVARPLDLGGLGVLDIHRFARALRLRWLWREWEVPEAPWAGLQTPCNDTDKLLFAAATRIAIGDGAKTSFWHCAWVNGRRPKDIAPDIFAASSDCRLSVREALTDSTWVNRIDLHAIHT	cd01650	23	286

						ADHLKQFVELWSFAHSVSLVQNTED RIDWIYNASGEYSAASAYRIQFMGHT CSNLKALIWKPWGPSKCKLFAWLIIR NRVWTSDRLAARGWPHSPSCPLCKR HPETAHHIIAGCRYSRRIWDALASWL SIGPLAPATWPDSTSIKQWWFMVGS VAVPKKGTRSLLLLVVWHIWLERNN RTFQRQERSVADLLAAIKAEARLWG VAGAKHLLALLPML			
	LINE1-35_ZM	GCA_902373975.1	Zea mays	ACAAAACATGGT ACTGCGCTACGA CGGAGCCGACG GGTTGCAGGGG CAGGGGTGGAA TTTAACATGCAG GATCTGGAGAAC AGAGCCACAAAA CAAGTAATGCGG ACTCTGAGAGTC ATCACAGAGAAT GCAAGCATCAAC	TAGCTCTGGTTTTGTT TTCAGGATCTGTTTG CTCTTGTTTTTGTTTT TTTCCTCTTCTTTTTTC TTTGGTTTTTTTATCC CCTCCGGGGAGGTCT TTTGTACTATGGTCTA TTTTGGACCCTCTTTT TCCTCTACTTAATATA ATGATGCGCAGTTCT CCTGCGCTTTCGAGA AAAAAAAA	MDPSKIICWNVRGLNSADRQNSVRN LVLSSKADVVCLQETKMTVISTRTILA ALGSDFCQFSFLPSCGASGGVLVAW KRHLGVTGLSRIDRYSISIQFCNAHNR CAWWLTCVYGPQSNSDKMLFLQEL RDVRAACTGPWLIAGDFNLIYKASDK NNSNINRAMMGRFRKLVEDLSLNEV PLFGRKYTWSNQQDSPVLVKLDRVL CSVEWEDNFPNVLLQSAASEGSDHC PLLLGIRDNKLGKRRFCFEAFWPKLEG FQDVIQEVWDSVAAVNCPFMTLNLK LKATARKLQSWSGKHVGHVRTQLAL	cd01650	498	760

				CAGCAAGCTGTT AAGGATTATGAG AATCTTTTTCAGC ATCCTTTAACTCA AAGTCATGTTGA AGCGTTAGCGGC TCTCTTTGGATG GGAGACTCCGG GGAATCTCTTCC CGTAGTGTGTTT TTCTTTCTGGGTT GTAATTTATAATT GTTGTA		AKEILHRLEIAQDCRALSLAEIWLSNLL KKHSLFLSSLQRTIARSRSRISWLKEG DANTKLFHMHSRHRKRRNFVAQLV NGDQILTNHDDKAVLVDNFYARLIGE CGDREQTINLEALHLPTFNLAHLDEP VTELELWNTIKGLPNDKAPGPDGFTG RFYKSCWNIIKMDLMAVVAAVWNR KFDNFCLLNSAFITLIPKHDGAVHVKD FRPISLVHSAAKLITKLLANRLSQKLND MVSPIQSAFIKGRFIQDNFMLVQQT ARLFHQQKRPRLLLKLDITKAFDSVS WPFLIEVMKYLGFGQIWRDIICGLLSS SSTQVLLNGIPGGHIEYRRGLRQGDP LSPMLFILVMDVLGLLFSKAEHAGLLL DLSPRRPLHRVSMYADDVALFLHPSA ADIAITLDILHLFGEASGLRNNEQKSN VYPIQCGGVELEMVQDLLPCGISTFP CKYLGLPLSLKNLTKVQVQPIVDRIAD QLPGWKADLMTRAGRKVQVQHVLT GMLIYLAMAVNIPSWAIKAIDKIRRSF LWRGRKEARGGHCLVAWGRVCRPL			

						QLGGLGISSLKELIWALRMRWLWLQ KTEPNRPWCSLPIKVPDKARAFFSMV LTSKVGNGANTLFWTDKWIHGQKIA DLVPRLFDIIPKKITNRRTVQEALTDRR WIADITGSLSVGVLLDYLHLWNLISDF ELLPDIEDKHTFSIAANGTYSAKAAYD GLFIGAVQFGHYERVWKTWAPSKCH FFLWLSVLNKCWTADRLAKRGLDHP ERCPLCDQESESLDHILVSCVFAREF WFRLMCIFRLQNLAPTPSTVSFTGW WEMIDNGSGDMIMKGVNSLIALGA WIIWNHRNRIVFDGISPSVSAALCQA REEQQLWEMAGAKGLSFLAATIRGT			
	LINE1-2_ZM	GCA_902373975.1	Zea mays	GGACCCTGTTGA TGAGCCAGGGAC CGTGGAGCAGCT GGCGCCCAGTAC TGAGGATTCTAT AGCTGTCACCTC TATTCCTTCCGCG AATTTACCTCTGC	TAGCAGTTTTCCTGCT GGTAGTCGTTTTTGG TCGGGTTGTTACCTT GTGTGTGTGTGTGGT GCGTTGTGTTATTGC CGGAGGGGTCTCGG CCTAGGGGCCTTTGT AATTAGGGGTTGACA	MNLSNILVWNARGLNNKARRDYVR DTILSSKADIICLQETKVASFSSHLLLSV CGSDFDKFLTLPAVGTRGGILIAWKG AVCQVISSRIDNFSVSVQFAGMDGM NWWFTGVYGPQEDDRKIQFLQELLD FRTLCSGPWLLSGDFNMIYQAADKN NANLNRALMGRFRRFLDDCLLKEIPL HGRKFTWSNERSSPTLVRLDRVFCCN	cd01650	496	759

				TTACTTGGGAAG	AGTCTTGTCTCTCCTA	DWEDFFPNSLLQSASSGVSDHCPLIL			
				AGGATAGACTTG	TGTTCTTTTTTTCTTCT	SLNVQTHKKRRFHFESYWPKLPGFLE			
				GTGCTGCTGTGG	TAATATAATGACACG	AVSHSWNAPVSTACPVEGIFLKLQRL			
				TTCCAGACATTG	CAGCTCTCCTGCGTTT	SKGLQKWSHHKVGNVKLQLATAKEI			
				TGTCTAACCAGG	GTTCAAAAAAAAAAA	LHRLEIARDNRVLSPGEEWLRKKLKL			
				GTATTCCTCACCC		HCLGLASLERTIARLRSRILYLQEGDA			
				TGTTGAGAGGCA		NTSFFHQQARYRKKKNFIAKFQVEER			
				TAGGTCTGTTGA		IVTSQEEKQQAALDFYDNLLGTAEHR			
				GCCGCAGGATGC		DYTLDLHSLGIQQHDLGDLELVFSME			
				TGCCATTAATCG		EVWSVVKDLPLDKAPGPDGFTGRFY			
				CAGAGGTGGCCT		KVCWDIIKEDMMGALLAVQCGHVS			
				CCCTTCCAACCTT		KLKLLNTAFITLLPKKVDALLVKDYRPI			
				CAGGTATATTTT		SLVHSFAKLVTKVLAARLAPHLPSMV			
				AGGCGAAAGCG		SINQSAFIKGRSIQDNFLLVQQLTRSL			
				TCACTCTAAGAA		HHSKEPHILLKLDISKAFDSISWPFLLD			
				ATGTTCTGCGCC		MLQHLGFGRSWCNLISLLLCTSSTRIL			
				GGATGCGCCTGC		VNGEPGEYILHHRGLRQGDPLSPMLF			
				TGTTTCTTTCAAT		ILVMEALNAMVSYAFREHMLQPIAN			
				GACCTGGAGAAA		QHAKHCISFYADDVVLFMRPSHQDIS			
				ATCATCAAGCCA		AIISILDIFGHATGLKTNISKSSVTPIQC			
				GTGGTACAGCTG		GANELTTISALLPCATKNFPCTYLGIPL			

				CTGCCCATGCCG		SIRRLANSDLLPLVDKVADKLPGWKA			
				GTTGCCAAGATT		NLLSRAGRLVMVKTVLSSVPIYLMLAL			
				AGAAGCAGTAA		ELPKWVFKAIDKRRRGFLWKGQGDA			
				GACTGTTGTTGC		NGGNCLVSWEAVQRPLKYGGLGILN			
				TGTGCCTCGTCG		LEMFGWALRARWLWLQKTDASRP			
				GAGTAGGAGGG		WAGLPIRVHRNAKALVDVAITSVVGS			
				TGGCTGGTGTCA		GESVKFWSDRWLLGKTVAEHCPTLI			
				AGCCTTGTTCTCC		QVISRRALKRRTVAEGLTNRQWVSDI			
				AGGTCCGATCAT		RGGLSVTVLVEYLQLWNLVDGVVLQ			
				CACTGAAGCCCA		PDIADQHIWRLSAHGTYCSKSAYDAL			
				AAAGAAAGTCAT		FVGSIPFGPWRRVWKTWAPLRCKFF			
				CAGGAGTTTGGG		VWLAIKNRVWTADRLAKRGLPHPVA			
				TCTCGCAGCCAC		CPLCDQAEETIQHILVSCVFARQIWTS			
				GGAAGAAGCCAT		MLHNLGLLAIVPQHGCTRFSNWWC			
				TGATCAAAAAGC		QSIKKVEKSLRKGLNSLIILVAWEIWK			
				TCAAGACAACTA		HRNACVFEGVVPCTQRVQSAVIEEG			
				TTTTAAGCTGTTC		NVWCLAGASALQDLLLRQLPRGP			
				GAGGGGCGCCT					
				GACGGACTCTCA					
				TCTTGCTGCACTT					
				GCTGCTATTTTTG					

				GCTGGCAGGCTG AGCGATTTGATG ATTCCAGCAGCC CAGTTCTCCTGG CAGGCTGTTAAT TCCTGCTGTCCTT GGTCGTGTTTAT TTGTTTTCA					
	LINE1-18_ZM	GCA_902373975.1	Zea mays	CCAAAAAGGAAC TTATTGATTCAGT TTATCTAAAAAA CTTCTGTCCTAAT CAGTTTGATTCTT TTGCTTTTGTTCC TTCGGTGGGGAA CTCGGGGGTTCA ATTATCATCTGG AAAAGTTCAAAA TAACAGGAAATG TCATCTTTCAAAA TGACTATGCTCA	TAACGTTCCTTTTTCT ATGGAGGTCATTATT CTCATGACATGGAGC ATTTGGAAAAGTCGA AATGAATGGTTGTTT GAAAACAAAGACCCT TCAGTGCATCACTGT ACGAATGAATTCAGT AAGGAACTACGGTTG GTTATTCACAGAGCG CGAGGGAAATACGA CATCTCCATTCCAAG CTGGTTGGATCAGTG	MPSDKLWLLVGDFNLIRRPEDRNKA GGDINLMLKFNEAISNLDLVEIPLHGL SYTWSNRQREPLLQRLDWFFISQEW SIVYPDTCAKTMPRDISDHVPCLVSFK SKIPKPKIFRFENFWLEFEGFMNIFQT TWLGLPCLPDKAKNLTAKFKIVRKVLK DWQRSLPKIDKTVRDIKLLIEFIDNIEE HRDLSIEEWNFRDIMQQKVAELLNIQ KIYWKQRASIKWITEGDICSRFFHAH ATVKHRRNTIASLTDENGIIHFEHDHK SNLLWEAFKSRLGSSDFLGIDFDLSEL LIRNEDLQGLDSPFTKQEIDNIIKSLPS DKSPGPDGFNTNFIKKCWHIIAQDFY	cd01650	365	629

				GTCAGTGGAATT CATATCCAATTTA TCGCGTGCTCTT GGATTATTACAA ACATTTATGCAC CTTGTACTCCAG GGAAAATAGATT TTCTCAACTGGTT CTATAACATTTCT	GCAGTCATAAGATTT GTTTTGCTCTGACCA ATCTTTGTAACTTTTC AATTAGATCTTACCA GCTAACCTCTAACTCT TGTAATCAAGTCGAG TTGTACCTTTTCATAT TTAATAAAATTTTACA GTAGGGGTTTCCCCT CCTGATCTTTCAAAA AAAAA	DLCEKFYQEEVCLRSINGSFIVLIPKKE NPQKVGDYRPISLLNNNMKILTKLLA NRLQSFMTRLVHRNQYGFIKGRTIQ DCLAWAYEYIHLCHSSKKEIIVLKLDFE KAFDTLEHELIMQVMTHKGFGSKWI NWVRNILHSGTSSVLLNGVPGKSFHC KRGVRQGDPLSPLLFVLAADLLQSIIN KARHLNLLQLPLPDICEQDFPIVQYAD DTLLIMEACPKQLFFLKAILNSFASSTG LRVNYNKSNMYPINVNPEKMEILSRT LNCQIGAMSFTYLGLPLGLAKPRIQQ FLPLIQRIEKRLSCSSNFLSQAGRLELV NSVFSALPTFFMCTLKIPKSIIKQIDIYR KHCLWRGNDTSSKKPPLVAWSTVTQ RKVNGGLGVLRLETQNDALLLKFFHK FFNSYDLPWVNIIWNKYYTPNRLPGH RRIGSFWWKSMLKLLHNFKGLAHPII GNGRSILFWEDQWNKSIPRQQYPEL FSFTKSTKLSIAEVKEQDHLFRIFQLPIS EQAYEQYLELQVVWEQIVLSNVKDR WRYIWDSDSYSTQKAYRHLMGQTQ			

						VHSIYKCLWKSKCQPKHKVFYWLWL KNRLNTRNMLRRKNMTLESYSCENC LWQREETLYHLFLRCNFAKACWNSIG IPPTQNCSSRGGLGKPQTTT			
	LINE1-28_ZM	GCA_902373975.1	Zea mays	GGCGGTCATTTC GGCGGACCGTTC GGCTCCCCGCCC TCCCGCCTTCCCT GCATGCCCTTCC GCCTCCCCCGCC CATCGCTCCGTT AACGGCGCCCGC CCCACCCTCCCC GCACCACCTAGC CCCTCGCCGGCG CTCACCCCGACT CCGGCGACCACT TCCTCCTCCCCCT CCCGTGTGGCGA CTTTGTCTCCCGC GGCTTGTCCGTT	TAATGTTTTTCCTATC CGATGTTCTGAGGCG GATCTTGCTTGTGTTC AGGAGCTCCTTCCAT GTTCCATTTCGGAAT TTCCCTGTCAGTACCT TGGCCTCCCGCTTTCT CTTGGAAAGCTTAAG AGAAGTCAAATTCAA GCAATTGTCGATAAA ACTGCAGCACAGTTA CCTGGGTGGAAAGCT GATCTGATGACAAAG GCTGGTAGAGCCATT TATGTTCAATTTGTG ATGACATCAAGAATG ATTTATGCTGCTCTG	MVVFSMFFQMDPSKIVCWNVRGLN GSARQDVVRCLVDSVKADVVCLQET KMSVFSRLFVFRLLGSEFQNFVFTPSV GASGGILVAWKNNISSCGSRIDDHCV SVNFSLETGSSWWLSCVYAPQGNHE KIQFLQDLRLIRTHCSGPWLVVGDFN MILCDEDKSNSNVDRAMMGRFRR WKDDLALLELPLIGRRFTWSNGHAC PTLVRLDRMFCTADWEDLFPDCLLQ SAATQDSDHCPLLLGLCDINKRKGRF HFQAFWVKLEDFQDVVANAWSSIP AGQCPLVSLSKKLQATAKALQRWSD NKVGDLKLQLEMARELLHRLEMAHD SRTLSPSEDWLRRTLKKHILVLSSLSRS IARLRSRINWLKEGDAHTSLFHAQARI RKKKNFIGSLSTSEGLILTSHEDKADEL FHFYDALLGTALGRENSLNLDALEMP	cd01650	506	704

				CTCCCCCGCTGG	GCGTTGGACATCCCT	SFNLNELDVPFSEEEVWGTVKSMPS			
				GCGTTCCAAGGC	TCTTGGGCTTTTAAA	DKAPGPDGFTGRFYKVCWVTIKADV			
				GCAGCGGTGGG	GCAATTGATAAGCTT	MAAISCVWARKFRNMGLLNSAFITLL			
				CGTCTTCCTGCCT	AGACGAGGCTTTCTG	PKVEGATQVKDFRPISLVHSFAKLVTK			
				CACCAGCCCGTC	TGGAGAGGGAGGAA	MMANRLARHLNDMVSPNQTAFIKK			
				CTCATCGGCCTT	GGAGGTGAATGGTG	RFILDNFMLVQHTARFLHRQRQARIL			
				GGGCGTTGGTGC	GCCATTGTCTGCTGA	FKLDISKAFDSVSWAFLLEVLQKLGFG			
				CGTTGGAAGCGG	CTTGGCCCAAAGTCA	QIWRDIICGLLASSSTRILLNGVPGDYI			
				TGCCTCGCAGAG	CTAGACCCAAGGACC	AHQRGLRQGDPLSPMLFILVMDTLN			
				GACTTTCAAGGA	TGGGGGGTCTGGGT	LLIQRASDEGLLHSLSTNSLQHRFSLY			
				CGTGCTGCTGCT	ATTAGTGATCTTCGG	ADDAVVFLKPAAFRHKSDSGNPSDF			
				CCCTCCGCCCCT	ATCATGAGCTGGGCA	WGCVRSKDQRSEE			
				GCCGTCTGCCCG	CTGCGGGCTAGGTG				
				GCAGTCCCTCCT	GCCTTGGCTGCAAAA				
				CGACCTCCGCAC	ATCTGATAGGACTAA				
				AGACCGAAGTTT	GCCATGGTCGGACTT				
				GCGGTGGTGCAA	CCAGATTCAAACTTG				
				CGTCCCTGGAAC	CGACATTGTGCAAAG				
				GGTGCGCAGGCC	CTTGTTGGCTGCGGC				
				TGTTCCTGACCG	TGTTATCACTCATTTG				
				GGAGGGCTGGG	GGAGATGGGTCTGA				

				TTAGGGTCGAAC	TACTCTGTTTTGGAA				
				CTCGTCGACGTC	GGATCGTTGGCTGAA				
				AACTCCCCCCTTC	TGGCAAATGTATTAA				
				GCCACCTCCCCG	GGAGGTGGCACCTG				
				TCGATCGATCCC	CTGTCTTCGGTATGG				
				TGAGGATCTCCG	TGCCCAAGCGTATTG				
				GGTAGGTGTTTT	CTTCTAGACGCCTGG				
				AACTGTCTGGCC	TCAAGGATGCCTTGC				
				TCCTCTCATCGG	CTGATCTCAATTGGT				
				GCGGCGGATTGC	TGGCTGATCTGAGAG				
				CGACGTTCGGTG	GGGCTATTTCTGTTA				
				CGCTGCTTCAAG	CAGCCCTTGCTAATC				
				TGTTGGGGGTTC	TCATGGACCTTTGTG				
				GGACATCGTGCA	ACTTACTCGAGGGGA				
				GTTCAGTGCAAG	TGGTTCTGCAGCCGG				
				GTTCGTGCTGCG	ATGTTCCTGATGTGC				
				CAGGTGATCAAA	ATACCTGGAGGTTCA				
				GCTTCAGTTCGG	GTGAATCGGGTGAAT				
				TGTGGGATCGTC	ACTCGGCTAGTTCAG				
				TTGGCCCTAAGC	CTTACAGGGCTTTGC				
				TGGTTCACAAGC	TTCTTGGCTCTGTTAG				

				CTTTGGACTCTCC	CTTTGAGCCTGCTGA				
				TAGACGCGCCAT	ACGCATTTGGAAATC				
				GGAGTGGCGCA	CTGGGCTCCAGGAAA				
				GGGTCTCTTCGC	GTGTAAATTCTTCCTC				
				CGATCCCGGCCA	TGGCTTGTGGAGCAC				
				TCCAGAGGTCCG	GATAGATGTTGGACG				
				GTGTGGCTGATC	GCTGACAGACTTGCT				
				CTACCTTGGGTG	AAGCGTGGCCTGGAT				
				GCACGAAGAAAC	CACCCAGAGTCCTGC				
				GCCGCCATCGTA	CCTTTATGCGACCAG				
				CGCGGGGTCGA	ATGGAAGAAACTATC				
				AGGCCTAAGACC	AACCATTTGCTGGTT				
				GCCCCCGTTGCA	CAGTGCAGTTTTGCA				
				CCTTTGGAGGAA	AAACAGTTTTGGTTC				
				TCCAGTGAAGTT	AGCCTTCTGCGGTCG				
				GCACCTCCACCC	GGTGGTTTCCAGGAG				
				GGTTCCTCTTCCC	CTCTGCCCTTCCCATG				
				TAGAGCCGGTCA	ATGACGTCTGTTTTG				
				GGGTGCATGCCC	AAACCTGGTGGAGG				
				CTCAACCGGTTG	AATAGTGCTGCCAGG				
				TTCCTGATATGG	GTGGGCATGCCTTTA				

				TGGAAGCCACAC	TCCAAGGGATTCAAT				
				ATCTCTGCATTTT	TCTCTTGTCATTTTGG				
				GGACCGTGCGG	GTGCATGGACTTTGT				
				ATGCGCTTGCTA	GGAAGCATAGAAAT				
				GAGAAGAACTGC	AGGTGTGTTTTTGAT				
				GGTTGCGACGGT	AACTGTAACCCTAGC				
				GTGCTCTGTTCG	GCTGCGGTGGTGTTG				
				TTGTTGTCACTG	GCAGCTGCAAGGGA				
				GTTCAAGACCGA	GGAGTTGATGGGTT				
				GGGTGGCTGCTG	GGCGTATGGCTGGA				
				AGACTCTTGCTG	GCCAAAGCGCTCTCT				
				ATGAGGTGGCTG	TTCTTTCATTTAGAAG				
				CTGGTTTCGGGC	AGGTGTCTTAGGTGT				
				TGGATGCGTCGT	CTTAGTTGTTTGGGG				
				CCTTCTCCGTTCA	AGTGTTTTTTGGTTTT				
				TCGTTCGTGCCC	GGTTTGCCAAATGTA				
				AGAAGATTTTGT	ACACCTTTGTACTGG				
				CTTAATCCTGAAT	GGTTTTTTCTCCCCAT				
				TCGGAAGAGTTG	TTTATTCTTCTTCTTA				
				GCCAATCAGGTG	ATACAAAAATACGCA				
				TACAACAATGGT					

				GCTGTTTTCAATT CTCTTTCTGGGTC TTTCAAGTTCCTG AGATGGTCTAGA CTTGCCCATGCC GAGGTGGTGGC CTTTCCATCTCCT GTTCTTGTGGAG TTCAAAGGGATT CCAATTCATGCTT GGGATTTGGCTA CTGCTCAGAATT TGTTGAGGGATC ACTGCACTGATG TGGAGCTCCATC CTGACACGGTGA ATCGACACGACT TCTCTTCCTTCAA GGTTTTTGGATG GTGCCGACGTCC ACACCTCATTCCA	GCTCTCCTGCGTATTC AAGAAAAAAA				

				AATTTGGTGGAG CTCCTTGTTCCGG AACCGTTGATGA TGACGCCTGGCT CTACCTCTCTTGA GAGAAGATTGCT CGCTTACCAAGT GACTGCCCTTGT GTCTCTGATCTCT CCGCCGGTGCAT AGCGCTGCTGCA CCTCCGTCGCCG GCGTTTGATGAT GATCAGGACAAC GGGCCTCGTCGG CGCAGGCTCAAC AGTTCAGTTCAG ATGGGGCATGCT GAGTCCGGGGA AGCACTTGCAGC GCGAAGGCCCGT					

				CCGTGAAAGGCT GGGTGTTGCCCC GGTCTCTGTTCG ACATGTGGAGGC TGGATTGTCCGT GAAAGACTCGGC GCCTTCCGTTGC CTTGGTTACTGG GGACACGCATAC TGATGTTGTGGA GCCGGTGATTGG CTGCATGGATGC TGGGTTGAGTCT GCCCCTGGAGAA TGTTCATGGGCC TGGCACTGAGAC TTTGGAGATTCC AGGTGCTTGCGA TGGCCCGGGTGG TGAAGTCTTCTT GTCCATGACTGA					

				TGTTGACTCTTCT GTTCCGGCGGTG GAAACTTTGAGT GCATTTTCATCCC TCCCGAGTGGCC CTGCCATCGCTA CCCGTGTGGGGA CTGTGTTGAAGC CGTCCATGAAGG CTGTTGGGGCAG GGCTCAGGAGGT GGCCGTGTTCCC CAATCAAATATTT CAGAAGGAAGT GCAGAGGGGCT GCTGCTTCACAA TCCCCTCTCTCGG CTGCTCAGTCTG CGACCACTGCTG ATTTCATCGCTA GAGTGTCCAGGC					

				CGGCCGCTCGGG TCCTGTGTGGGC CGCAGAAGCGG CAGCGTCGCCCC AGGGCACCTTGT GAGGGGGCTCTA CCTCGGCGGAGC AGGCGCTTGGCT GGAGCAGATGTT GAAAAGGTAATG GCCCCAGCACAA GTTTCGAAAGGA AAATGGAGGATT GCTAATGAGCTC GGCCTTCTTGGT GGGGATCCATCT GAATCGCTGTCC CTGGAGGTTCTC GATAACTATGCC AAGGTTTTCGGT GAGCGGCTGACT					

				GATCTGCAAGTT AAAGCTCTTGCG GCTCTCTTCAACT GGACTGTCCCAG AAGACTTGCAGG GTGCAGTGGTGC AGGCCTGATCGC CTCTGTGTGTATT G					
	L2-2_SSa	GCF_00023 3375.1	Salmo salar	TACATTACAACG GCTTGATTTGTTT GATCTGAGCAAT TCTTCTTAGCTAG CTACATAGCCGT CTTTGTATCAAA GATAATTGTGTA GTTTAGAGTAAT TAGAGTAATTAT CGAGGCTAGCTA TCTTCGTCCTCCT AACGTAGTCAAC	TAGGATAAAGTAATC CTTCTAACCCCCCCCC CCCTTAAAAGATTTA GATGCACTATTGTAA AGTGGTTGTTCCACT GGATATTATAAGGTG AATGCACCAATTTGT AAGTCGCTCTGGATA AGAGCGTCTGCTAAA TGACTTAAATGTAAA TGT	MCLSIPVLSPLHRPYKRSTPRGRCHS NLVVPARTTHVEFQVSGSLWNCRSA ANKAEFISAYATLQSLDFLALTETWIT TDNTATPTALSSSAYVFSHTPRASSQ RGGGTGILISPKWTFSLSPLTHLSISSF EFHAVTVTSPFKLNILIIYRPPGSLGEFI NELDALISSFPEDGSPLTVLGDFNLPT STFDSFLSASFFPLLSSFDLTLSPSPPTH KAGNTLDLIFTRCCSSTNLIATPLQVS DHYLVSFSLSLSSNTSHSAPTRMVLRR PNLRSLSPATLSSSILSSLPSAQTFSNLS PDSASSTLLSSLSASFDFLCPLSSRPAR	cd01 650	519	780

				ACCGCTGCTAGC		SSPPAPWLDDSLRAHRTGLRAAERK			
				TAGCTAGTCATC		WRKTRLPADLASFHSLLSTFSSSVSAA			
				ACTGCTAGCTAG		KATFYHSKLQASASNPRKLFATFSSLL			
				CCAACTTCTACC		NPPPPPPPSSLSADDFVNHFEKKVDD			
				GACTAGCAGCAC		IRSSFAKSNDTAGPAHTALPCALTSFS			
				TGTAGAAACTAT		PLSPDEISRLVTAGRPTTCPLDPIPSSL			
				TACATTACAACG		LQTISGDLLPYLTSLINSSLTAGYVPSV			
				TAACGACTTGAT		FKRARVAPLLKKPTLDPSDVNNYRPV			
				TAGTGTGGTGTT		SLLSFLSKTLERAVLGQLSCYLSQNDLL			
				AGTGTTAGTTAG		DPNQSGFKTGHSTETALLCVTEALRT			
				CCAGCTACATAG		AKANSLSSALILLDLSAAFDTVNHQILL			
				TTGTCTTTGCTGT		STLSELGISGAAHAWIASYLTGRSYQV			
				CTCTGTATCTAA		AWRESVSAPRALTTGVPQGSVLGPLL			
				GATAATTGTGTA		FSLYTKSLGSVISSHGLSYHCYADDTQ			
				GTTTAGAGTAAT		LIFSFPPSDNQVANRISACLADISVW			
				TATCGAGGTGAG		MTDHHLKLNLGKTELLFLPGKDCPFH			
				CTAGCCAGCCGC		DLAITVDNSLVSSSQSAKSLGVTLDNT			
				GACGCGGCGCCA		LSFSTNIKAVTRSCRFMLYNIRRVRPC			
				GACTTAGTCAAC		LTQEAAQVLIQALVISRLDYCNSLLAG			
				ACACCTAGTCAT		LPACAIKPLQLIQNAAARLVFNLPKFS			
				CATTAACCCACT		HVTPLLRSLHWLPVEARIRYKTMVLA			

				GCTAGCTAGCCA ACCGTTACCGAC TAGCAGCGCTGT AGATACTAATAC TTTACAACGGAA CGATTTGACTAG TGCAGTGTTAGC TAGCTAGCTACT AGTTGTCTTTGTC ATAGCTTGATAA TTGTGTAGTTTA GAGTAATTATCG AGGTTAGCTAGC CAGCTATTTCCGT CCCCCGCGACGC CATTTTTCCTAAC CCAGCCAACTAT TACCGACGAGCA GCATTGTAGAAA CTAAATACATTA CAAAGGAACGTC		YGAVRGTAPPYLQALIRPYTQTRALR SSTSGLLASLPLRKHSSRSAQSKLFAAL APQWWNKLPHDARTAESITTFRRHL KPHLFKEYLG			

				TTGATTAGTGTT ATGTTAGCTAGC TAGCTACAAAGT TGCTTTGTATCAT GACAAGGTGTAG TACTGAAACTAT CGAGGTTACCTA GCCAGCTACACG TTCAAACAAAGT CAACAACGCAGC CACTGCTAGCTA GCCTACTCCACC AGCCAGCAGTAC TGTATCATTTTCG TCATTTTAGTCAA TAAGATTTTTGC AACGTAAGCTTA ACTTTCTGAACAT TCGAGACGTGTA GTCCACTTGTCAT TCCAATCTCCTTT					

				GCATTAGCGTAG CCTCTTCTGTAGC TTGTCTACT					
	Dong-1_MMa	GCA_000484575.1	Mesobuthus martensii	TAACCCCTCCTG AATCCTGCTAAG CTGCCTCAAAGT GGCGGAGGGGT GTACGTGGGCG GAATAGGCGATG CAAGTCAGAAAT TATTTTCTAGTAG GGTCACCCTGGG CTAGAAGGCTAT TCCATTATGCTCA ACTAA	TAACTCACAATTAAA CATCACAAGTAGCTA TGTCTTGACTCTTGTC CGACATAGCACTTTG CCCCGGAGTTATTCC GTGCTCACATTGATG TACCATAATATAATA ATAATAATAAATACA	MLAGVPTLVVFFQYAMAQSRKRKHT GSDVSRVNKVRVCELGRGSHSESLLQ SPLATGGESSTNPSLATAIQIPRVGNG SPHQSPTSQSVSLETQPETTKAGKPR QRMKWTTDMNEFIMRTYYTITKLET DITGYRRELYDKFTSRYPNITVTEQRIA DQRRTIIHRNLIPHSTLESIKTQVAQQI SQHETDSHTQIEHDVEQTPEITYIDD NTQPINNTDTDLINKLTDTLQETLTKF TGTDPTQRPRIPKLHNSQQLNNTLHC MNTHILPKLINAETTLIEIHTLIYAAAV TVTTHLGLKTLPNKIGTNTKHHLKPA WQKRLEKQINEIRTELGRLTQYTKNN TSTKLMKSIRKIFRKLKTHTRHENHNT HIKEYIDTLKQKLAAKSQRLKRYKTSLA RKLDNTLFTNNEKTFYRNINTKDPKQ TQNREIPTQQSLTSFWSALWAVPVT HNQRASWIAEEETRTLDVPEMPYTTI	cd01650	534	805

						TTEELTNTIRYTHNWKAPGVDGIQNF WYKHFTCLHDLLAKHLTHLIHEPKDT PSFLTHGITFMKPKSTNTGNPSQYRPI TCLPTLYKILTSCITNKIYAHIEENNILTE QQKGCRRSHMGCKEQLIIDTTILKQV HRQHRHLHAAYIDYKKAFDSIPHSWL IQTLQLYKINTKYITFLQNIMQHWRTT LHITINNNHVATEPIHIRRGIFQGDSLS PLWFCLALNPLSNKLNTSNCGYDIRQ NQTTHYIHSHLFYMDDIKLYASTQTQ LQQLLQITHTFSKDICMEFGLDKCKTL HIEKGVLKKGGFTLENGDTIESMEEG EVYKYLGFQQARLTEQKKTKQTLSKL YTNRLNNILKTHLNGKNTVKAINTFAI PILTYSFGVIQWTRTELESLQRKTRTTL TKYQYHHPKSAIERTTLPRKEGGRGLI DIPHLHDRQIYNLREFFHKKQHTSQL HKAIVLADRNITPLNLQNNTPCQIDN TAHIQEKKTSWCKKELHGRHPHDLQ KHFIDAKASNSWLVRGDLFTETEGFII AIQDQVINTKNYRKYILKDSTLTNDTC			

						RKCHIAPETIQHITSGCKLLAQTDYLH RHNQVANIIHQKLATKHKLINTLTPYY KYIPDTILENSTHKMYYDRTILTDKTIH HNRPDITLIDKQNKVTYIIDISIPDTHN IQDAFTQKISKYTELSQEIKRLWGMQ KVVTVPIILSSTGVIPHTLHKSLKTLDL HPKTYITLQKASLLNTCNIVRKFLSTDT DS			
	SR1	GCF_000237925.1	Schistosoma mansoni		TGATTTACCAAATTCT TTTTTTCCCTCTATTA TCGTTCATATACCTA GGTTCTTGCCTGGAG GTATTGGTGATCCAC TGCTACCAGACACGG AAGCCCGTTAAGCGA AAGCTTCTTCTATTCC GTCCTCAACCATTTG AACCATTTG	MEFLVDDLSILAPLGKSDHAVLSFSFV SKTELRYPTSNKRWNFKRLNVLALQD YLQQVDWDVHPQLEVDAHWDFLLH TILCATEHSVPKMVPKSYKQPPIIKNR TRRLLSRKRHCWAEYKRTDNNGAYR QYKHIRNICTKAIREDRLQFQTKLIDKF VSNPKSLFSYAASLRQGKTGVSQLLG PNGPTNNDSDAANLLAEQYSRTFQL THINHTDESFTCTCTGLSEVDLSADLV LRKLQHLRKDTSPGPDMVHSAVLRE AASILATPLSVMFAHSLSRGKLPEIWK LAHITPIFKGGRRSEPSSYRPVALLSIPS KIMESLVYDGILEYLSSSKFFSPQQHG	cd01650	288	537

						FRKGHSCMTNLLTAVDRWTTILDRK GKVDVIYLDFSKAFDRVNHACLIKKLG RLGIKPPLIDWLSSYLENRHFKVRVNF TLSQAMECPSGVPQGSILGPLLFLIYI NDLPQQVSSDLLLFADDVKLWREIRN HNDILVLQEDLTRLQSWADDNGLTF NTSKCKVVHLRHVADYSYNLGNSPLE VSQVEKDLGVLVPYDLKSYANCDKN ASQANLALVTLKRIFGQFDGRTFHIIF NSFIRPHLEYGNIVFPPSLQKDKDTLE RIQRRATKSVRGLKFKPYEERLKSLNL YPLEYRRLRGDLLMTYSILNTSGHPLK HLLKLSHNTNLRGNTQKLETLHSRTD CRHNFYSVRVVKCWNSLPTELVQAT SQESFKRKLDLFLRTKDNILL			
	CR1_AC_1	GCF_000090745.1	Anolis carolinensis	TTGATCCCAGAA GGGCAGTTGGTT CTCGCAGAAGCA GAGGCAGCAGG AGGACATTTTTG CTCCCTGGCTTCA	TGAATGCGATTTCCT GCTTCTTAGCAGGGG GTTGGACTAGATGGC CCATGTGGTCTCTTCC AACTCTACTATTCTAT	MQQKQNIPITCTSCGMFRFFTQKLD NYICSKCKQMTLMEQRIQQLEHRIKT LKDIQALELFLDTAQHAAVDQQPAS HQHHQDHDQEPYGEINSKVDNPQA WKEVTHRRRRRTRQPPQNSSAQLHL HNRFEILTSLTYNQETHLVEDHSFLDT	cd01650	1008	1259

				GGTAGGATTTGG	GATTCTATGATTCTAT	TQWTTLDQCAGDSPDGDSASPQSH			
				CTAAGATTTGGC	GATTCTATG	LCNHANAPSPIIDQEQQEHPLGETRS			
				TAAGATTTTAAA		NRNIPWESNGLSDVSQWIVIDECTG			
				CTGAGTGTGGGC		DVEEEDNTLHLHDSLQQEHSSGDIHT			
				TTGGCTCCTGTG		VLHKRDPVNPQRKQVLVVGDSLLRG			
				GTCAAAGTCTAA		TEAIISRPDGMARETCCLPGAKIHHIT			
				CTGTTGTGTGAC		QRLSRLLKPHHPPHLMLIHVGTNDTA			
				TGTTGTGTTGAA		RHTFQKITNDFRALGTKLKLYNVHVIF			
				AGAGAGCCAGG		SSLLPVVGHGSTRAGKIVQVNNWLR			
				TACTTAAGTTGA		KWCQEEHFGFLDHGLLFQEDGLLAS			
				TTGACAGCAGGC		DGVHLTQVGKHLFAHRLTNLIRRTLN			
				ATTGTCCCCTGTT		PWSTLPRGWTTGKRWGASHTSRKT			
				AATTGAGTTTCT		SFCTQTHKPHQAHFKLDPLGEGNNS			
				GGGAAAGCTTTA		LANTILPTTTGNRRKAKRRAAQTQQ			
				TTTGCCAGTGTG		GPSTESTIIPNKQFEGRSQGLTCLYTN			
				AGTTTCTGCCAG		AQSMGNKQDELQLLAQHHTYDVIGI			
				AGCCTGATCCCA		TETWWDDSHHWNLTIEGYNLFHRN			
				GAAGGGCAGTT		RTKGRGGGVALYVKNSYVAEEMQD			
				GGTTCTCGCAGA		CNPGNQLESIWIRIKGTGTQKDLVVG			
				AGCAGAGGCAG		VYYRPPSQDEGLDEAFCQQLTKQAQ			
				CAGGAGGACATT		RRDIVVMGDFNYPDICWKTNSAKST			

				TTTGCTCCCTGGC		KSNKFLTCLADNFMVQKVEEATRGS			
				TTCAGGTAGGAT		ATLDLILTNVEDLINTVEVVGSLGASD			
				TTGGCTAAGATT		HVLLQFAIQRNAETKTSQTRILDFKRA			
				TGGCTAAGATTT		DFQKMKELLSGIPWTPILKNKGVKDG			
				TAAACTGAGTGT		WEFFKSEILKAQMQTVPTKKKNKTSA			
				GGGCTTGGCTCC		KKPEWMSKELLTELKLKSDMHKKWK			
				TGTGGTCAAAGT		RGEITKEEFKRIANTCREKVRKAKAQN			
				CTAACTGTTGTG		ELRLARDIKNNKKGFFAYVGRKRKKK			
				TGACTGTTGTGT		EAIGPLQGEDGVMATGDREKAELLN			
				TGAAAGAGAGCC		AFFASVFSEKESHLQPQQHGMDEGL			
				AGGTACTTAAGT		GEIQPQIGKQVVQEHLASLNEFKSPG			
				TGATTGACAGCA		PDQLHPRVLKELAEVISEPLAIIFESSW			
				GGCATTGTCCCC		RTGEVPADWRRANVVPIFKKGKKND			
				TGTTAATTGAGT		PNNYRPVSLTSIPGKILEKIIKEVVCEHL			
				TTCTGGGAAAGC		ETNAVIANSQHGFTKNKSCQTNLISF			
				TTTATTTGCCAGT		FDRVTSWVDTGNAVDVAYLDFSKAF			
				GTGAGTTTCTGC		DKVPHDLLANKLVKCGLDKTTVRWIC			
				CAGAGCCTGATC		NWLSERTQRVLTNASSSSWKEVTSG			
				CCAGAAGGGCA		VPQGSVLGPVLFNIFINDLDEGLEGTII			
				GTTGGTTCTCGC		KFADDTKLGGIANTPEDRSRIQNDLD			
				AGAAGCAGAGG		RLERWAKTNKMKFNRDKCKILHFGR			

				CAGCAGGAGGA CATTTTTGCTCCC TGGCTTCAGGTA GGATTTGGCTAA GATTTGGCTAAG ATTTTAAACTGA GTGTGGGCTTGG CTCCTGTGGTCA AAGTCTAACTGT TGTGTGACTGTT GTGTTGAAAGAG AGCCAGGTACTT AAGTTGATTGAC AGCAGGCATTGT CCCCTGTTAATTG AGTTTCTGGGAA AGCTTTATTTGCC AGTGTGAGTTTC TGCCAGAGCCTG ATCCCAGAAGGG CAGTTGGTTCTC		KNGNQRYRMGDAWLDSSVCEKDLG VLVDNKLNMSQQCDAAAKKANGILA CINRGIASRSREVMLPLYSALVRPHLE YCVQFWAPQLKGDVDKLESVQRRAT KMIKGLENKPYEERLKELGMFSLQKR RLRGDMIAMYKYVKGSHREEGASLF SAALQTRTRNNGFKLQERRFHLNIRK NFLTVRAVRQWNSLPGAVVEAPSLE AFKQRLDGHLSGVL			

				GCAGAGGCAGC AGGAGGCACCTT TACTAACTATCCT TTGCAGTACATA CAGGAAACAAA GCAACATAAACA AATACACAAAGA GGTGGTTTGTTG CAGTTTGCACAT AAAGTGCGTGCA TATTACTTAACTG TACAAAGATCCC ACTTGGCCACCA CGCTCACCAAGA G					
	Penelope-2_PSi	GCF_000230535.1	Pelodiscus sinensis	AGAAACTCGCAC AGATACAGACAG ACATCATCTTTCT ATCCAAGTGCAA GAAA	TAACAAGTCTTTGAA CTTTCTCACCCATTAT CAAGACAGTTTCCCC AATTATCACCTGTAA TACCATTAACTCACA AACATCCCACTCTCCC	MDIVPKGLKVNNPLRSTYFTDYSERL CLTLSKKLRNHLISILYKKQKTIKNDISN LDRLISRQTSTPTDTLQDFCSTRQEIYY THFYSLQQRKDNKFSNLLHTTGYNSN CLNPPDNIVNLSSYKLNPAEESVLSRG LSFCPTSPTNWIQFCGDLEAFFRRLRL	IPR00047	305	558

					TACCTCTAATATCATC	REYFQTTTEHQSDLPDPPYQHQRKN			
					AATTCACAGACACTT	NSIWTPPDGRNYNLDFYIHSFRNNA			
					ACCTTCCTTCCTCCCC	QTDIIRKRQQATHNLSRAEHHAIQSL			
					CTTCCCACCCCCCCGC	KNNLDIVIKPADKGGAVVIMNKADY			
					CTCCTCCTTCTGTTCT	NQEATRQLSNTTFYKPLPSDPTLDYQ			
					GCAATGTGATTTGTC	KKLHCLLKSLPTATREQILTESPSDPRP			
					CTTTTCATATTTGTTC	GLFYLLPKIHKPGHPGRPIISGIGTLTT			
					ATTTTTTTTAATTGTA	GLSSYVDTLLKPFVTNTPSYLRDTTDF			
					TCCTTTGGTATATATG	LRKLQNIDNLPNNTILATMDVEALYT			
					GTTGTGACTACTTTCT	NIPHEDGLQAIRNTIPEDTTANLIADL			
					TCCACTATTTGATCTG	CNFVLTHNYFRFENNLYLQISGTAMG			
					AGGAAGTGGGTCTG	TRMAPQYANIFMADLEQRFLNSRPL			
					GCCCACGAAAGCTCA	SPLLYLRYIDDIFMIWTHGQETLEIFH			
					TCATCTAATAAACCAT	RDFNNLHPTINLSLDHSTREIHFLDTT			
					CTTGTTAGTCTTTAAA	VQINNGKLDTTLYRKPTDSYSYLHASS			
					GTGCTACATTGTCCT	SHPEHTTRSIVYSQALRYNRICSNPTD			
					GCATTTTGCTTCAACT	RDQKLQDLYQAFINLNYPPRETKKQI			
					ACCCCAGACTAACAC	ERARRIPRNHLLQDRPKKTNNRTPLVI			
					GGCTACATTTCTATC	TYNPQLKPVQHIINKLQPILEQDTILKE			
						ALGDRPIVSYRQPPNLKMILTNHHRT			
						YHTNTNPGTFPCNKPRCQLCPHIHSA			

						DTIIGPNQVSYKIKNTYSCASRNIIYAI MCRKCPSAMYIGQTSQTLRQRINAH KTDIRQDHKEKTVSCHFNQKGHSLN DLATCILLQRPFTSALERESSELSFMLK FDTYQPGT			
	Penelope-2_PSi	GCF_000230535.1	Pelodiscus sinensis	AGAAACTCGCAC AGATACAGACAG ACATCATCTTTCT ATCCAAGTGCAA GAAA	TAACAAGTCTTTGAA CTTTCTCACCCATTAT CAAGACAGTTTCCCC AATTATCACCTGTAA TACCATTAACTCACA AACATCCCACTCTCCC TACCTCTAATATCATC AATTCACAGACACTT ACCTTCCTTCCTCCCC CTTCCCACCCCCCCGC CTCCTCCTTCTGTTCT GCAATGTGATTTGTC CTTTTCATATTTGTTC ATTTTTTTTAATTGTA TCCTTTGGTATATATG GTTGTGACTACTTTCT	MDIVPKGLKVNNPLRSTYFTDYSERL CLTLSKKLRNHLISILYKKQKTIKNDISN LDRLISRQTSTPTDTLQDFCSTRQEIYY THFYSLQQRKDNKFSNLLHTTGYNSN CLNPPDNIVNLSSYKLNPAEESVLSRG LSFCPTSPTNWIQFCGDLEAFFRRLRL REYFQTTTEHQSDLPDPPYQHQRKN NSIWTPPDGRNYNLDFYIHSFRNNA QTDIIRKRQQATHNLSRAEHHAIQSL KNNLDIVIKPADKGGAVVIMNKADY NQEATRQLSNTTFYKPLPSDPTLDYQ KKLHCLLKSLPTATREQILTESPSDPRP GLFYLLPKIHKPGHPGRPIISGIGTLTT GLSSYVDTLLKPFVTNTPSYLRDTTDF LRKLQNIDNLPNNTILATMDVEALYT NIPHEDGLQAIRNTIPEDTTANLIADL	IPR00047	327	504

					TCCACTATTTGATCTG AGGAAGTGGGTCTG GCCCACGAAAGCTCA TCATCTAATAAACCAT CTTGTTAGTCTTTAAA GTGCTACATTGTCCT GCATTTTGCTTCAACT ACCCCAGACTAACAC GGCTACATTTCTATC	CNFVLTHNYFRFENNLYLQISGTAMG TRMAPQYANIFMADLEQRFLNSRPL SPLLYLRYIDDIFMIWTHGQETLEIFH RDFNNLHPTINLSLDHSTREIHFLDTT VQINNGKLDTTLYRKPTDSYSYLHASS SHPEHTTRSIVYSQALRYNRICSNPTD RDQKLQDLYQAFINLNYPPRETKKQI ERARRIPRNHLLQDRPKKTNNRTPLVI TYNPQLKPVQHIINKLQPILEQDTILKE ALGDRPIVSYRQPPNLKMILTNHHRT YHTNTNPGTFPCNKPRCQLCPHIHSA DTIIGPNQVSYKIKNTYSCASRNIIYAI MCRKCPSAMYIGQTSQTLRQRINAH KTDIRQDHKEKTVSCHFNQKGHSLN DLATCILLQRPFTSALERESSELSFMLK FDTYQPGT			
	REX3	GCF_00277 5205.1	Xiphophoru s maculatus	C	TAAGCGGAAGAAGA TGGATGGATGGATG GATGGATGGATGGA TGGATGGATGGATG	MTLGRSSMIDFVIVSSDLRPYVLDTR VKRGAELSTDHYLVVSWLRRWGTW SLSGPCSVPPLSRRPIGAVAARLSVPV AAATLEPAGGHLLVRDAVRLKKESYR AFLACGTPEAADGYRRAKRHAARAV	cd01 650	277	532

						AEAKTRAWEEFGEAMEKDFRTASRR FWSTIRRLRGGKRCSTNTVYSGDGVL LTSTRDVVGRWAEYFEDLLNPTNMP STEEAEPGDSGLGSPISGDEVAEVVK KLLGGKAPGVDEIRPEFLKALDVVGLC WLTRLCNIAWTSGAVPLDWQTGVV VPLFKKGDRRVCSNYRGVTLLSLPGK VYSGVLERRVRRIVEPRIQEEQCGFRP GRGTLDQLYTLGRVLEGAWEFAQPV YMCFVDLEKAFDRVPRGALWGVLRE YGVPGPLIRAVRSLYDRCQSLVRIAGS KSGSFPVRVGLRQGCPLSPILFITFMD RISRRSQGVEGIRFGGLRISSLLFADD VVLLASSDRDLQLSLERFAAECEAAG MGISASKSEAMVLSRKRVECLLRVRG GVLPQVEEFKYLGILFTNGGRREREID RRIGAASAVKRALYRSVVVKRELSQK AKLSIYRSIYVPTLIYGHELWVMTERT RSRIQAAEMGFLRRVAGLSLRDRVRS SVIREGLRVEPLLLHIERSQLRWLGHL VRMPPGRLPGEVFRARPTGRRPRGR			

						PRTRWRDYVSRLAWERLGIPPEELEE VAGEREVWASLLKLLPPRPDLG			
	CR1-3_IS	GCF_00289 2825.2	Ixodes scapularis	TCTATTGATTGCT TAGTCATATCTG AAACGAACATAA AGGACTGCGACG AAGCATTATTCC AAAATAAAAGTT TTGAAACGATTA CAAATAACAGGA AGAACAAGCGA GGAGGTGGAAT C	TGAACAAGTGTAAAA AAAGGGTTTGTTATA AGTGACTAGAGCTCT GTCGCTGCCCTACGC ACAAGCCGCTTAGGC TTATGTGGGGTATGA ATCTCTGTAAAATTG TATGTGGAAAAGATG AAATAAAGATTATTA TTATTAT	MVMVNKKWNIKNSKKITFSRAEAVQ IDIENEKEQIQILAVYRAPQGNIAEFID DIDNWIKVENSKNLIITGDINIDITKQS NLVDKYLNCLAEHGLQSMIKDITREEI LGLQLTRSCIDHICIRSKKYCKAHVIAE KLADHYFVGIAIPELGIPESERDQVTT YINNAKVDKYVREINWESIANEEDPEI IYNKIKNAFTDIYKKSSEQVTVKKRKT GNQWVTKEIDQLIRKKNKAWRLLKS SPESSQRKEEYRSIRNKVTSIIKKEKRN YFWKKFEDTKDPSQCWSVINQVLGR NTRLSIDQTIKKNFGTYCNDEKIANDF NNEFIQSIELLCNHGYQNETETRARIG NSMYLPEMTEEDLVEIVSNMKTNKA VGIDSIRPRDIKENVNIIKTVLLALYNK SLAIGKVPSGLKTAVVRPIYKKGDKKT LLNYRPVCILPVLNYIMEKYVHKALNG FINKYNLLTKTQYGFKTNSSTITLLEDF SDFVNKKLDEGNYVLALFIDFSRAFET	cd01 650	415	680

						VRHETLIDRLFSLGVRGKLLDWFKDYL SGRTQVVKIASSFSEKRLVKQGVPQG TVLGPILFNIYINDIVNVHLNSVLFQFA DDTVILYADKDYTTAENKIQEDVYKLV KWYKENQIIINTKKTELMCFRSPQRR TQMEGNIKIHSDSCKNCNCEPLKLTT SVKYLGLYIDEHLKWNQHVEHICKRL RSVAAQLYHLRSCLPQPILKMIYHSLG KSVLVYGITIYGHCAKYLQDKINSILNK VVKNLYVDENRPLHYMYQHLDIQKL SDLLELNIIIQHQFSDKYKTKYIPIRNLR PKNLFQEPKVKTKYGKAVRASYVPRV FNKLPPRIIEITKISELKRELKQWFINRK ELQLAEEKSTRGWNLNCVVLCVLFCR ELWSKGVYMSCFHFFFLFARN			
	RTE-1_PM	GCF_010993605.1	Petromyzon marinus	GCCTGTGCGTGT GTGTGTGTGTGT GGCAAAACAACC CCCTACCTCCTCG CGGTGCCGGACG GGCAACTCTGTC	CG	MTSPGQTPIGGARRLNLYPRGSSRN NEKPLRVATWNVRTLTAEERLPLLCR ELDRYRLDLCGIQEVRWTGSGSCHHE GWTVLYSGQQTAKQQGVALLLNKR MKGAWERGGEVWEAVSPRLLWAR LPINRRRGMVVVVTYAPTEDEHPWR	cd01650	578	856

				GTGGACAGTGCT		DAGKADESDAFYDLLSQVLAKVAPR			
				GTGGGGGGAGG		DRVIMLGDFNARVGHDAGTWSGVI			
				TCCAATTTTGATT		GGHGPDQLNNNGRRLLDFCAAHGL			
				GGCTCCTCCCCTC		VITNTLFRHRRIHTTSWMHAGSKQE			
				ATGGCATATGTG		HLLDYIVTDQRMRAHVLDTRTYRGA			
				TCCACCACAGGG		DLPSDHRLVICKMRLPFFHSGGGCTP			
				GCTAAACGGGG		RTPFQPKVAWSQLSDESCKREFHHRL			
				GCTCTGGACGGG		RLGLASSPTPPDVEGKWARFRTVTH			
				TGCGAGGGGCT		AAAMAALGTRSRPPQRPWLTEEVFR			
				GGTAACCCCTCC		MVEKKRQAHCHRLQNPTSENEEAYR			
				ACCTTCAATTAAT		SLRSEVRRAVRRCKDGWWSRVAEE			
				TACCACGAAACC		MESASVARDHKSLFNSIKKVTCSATPI			
				ACGAATTAAACT		TIIRGKNGDPISDPQEQVCRFAEHFCE			
				GAAATATGTAAC		VLGEGKSLSQENIVAQSEDDPAEAAV			
				TAATAAGCTTGT		PQILEEVTPCEWLAEVPSVEEVLKAIR			
				GCCGGAGGCCCC		SLRPGKAAGRDDLPSEMLREGGVCA			
				GGGGGGCCAGG		QTALYDIITTVWTTGRVPQDWKDAL			
				CGTCT		VVPLFKKGDRTDCNNYRGISLLSVPG			
						KVLARIIQHRLARHAEERLAEPQCGFR			
						KGRSCTDAIFSVRLLQQRCREFQQDL			
						HLCFIDLVKAYDTISRPGLWVVLRAFG			

						VPDPLLTIIKDLHDGERARVKLRGRES EPFPVHLGVRQGCPLAPTLFNIYFDH VVREAFNSCTGGISIWYRYNGRLIDA RVRSRDGSLLLNHVMYADDLVIAAHS EEELETLLLRLERATKRWGLTISPTKTK HLPGYFVPSDTPPPQLPIGDGAVVER VESFPYLGSVLMTGSSLTAEVSLRISR AAISFHRLRNAVWKLPGLSLRTKLQV FSATVIPSLLYACETWTPLMEHLRRLE TFRLACLRSILGLTRLDCVRSSQILERS GQGPIGELLRQARLRWLGHVARMPS HRMPKQLLFGRIEGAKRARHGLEKR WSDVVQEDVELSGLADDWYQRCQD RPLWRRLVKDATGQLEAVALQQQR RAEERRSQQRAERRVAKAQQVGTV GGTGGASASHLSQSTRGTWVCPVTS CQRAFDTS			
	Perere_Smed	GCA_002600895.1	Schmidtea mediterranea		TTTACGTATGTTAAAT CATGTTTGTTTTTTTA ATATTATTAAAATAAT TGAAAATTTTAAATT	MIAVLSKYIVRKVFYWRYRLIKSLPNL FNYLFINHAAAVPIFKRFMKVYNSLKR KELDISFLETCLEEKMIPSFIYKMKLSD CLSSKDVQLIHRRTLRRMLESEKIRRT	cd00304	474	565

					GCATTAATTTTTAAA	SIDKKLTNIGLMLNNATNEFCLQVCV			
					AAATTAGAATATAAG	EWCIKNSIVQIKSKENAHIKKLINLRKI			
					TAAAAAATTTAAAAT	FGISKFSVENTVINLSKIVLNDDEERVL			
					TTAAAATTAATTTAAA	KFGPNHCIKNRLDETKLCAKIESLYFHI			
					GTGAATTGAATATAT	ARSMKGINGNYVKSSLSASTMEYIDK			
					TTTAAAAATATTTATT	NKKQNIRDFMTLKNLAKKNINICKFD			
					TATTAAAAAATTAAA	KGNGIVILDHDVYVNKLNDLLNSPQF			
					TTAATAAAAAATAAT	VKLPKGRKNGKLPPLKDEEFVQKML			
					TTGTTTAAATTAAAT	NGFIEKKLIDEEVGKSIRPIGSQPAKLY			
					GTTAATTCATAATATT	GLPKIHKDGVPMRPVLSMVGTAQYK			
					TTTAGTTGGTTTTGA	IAKFLDGLLKPLMRSEFECKDSFEFVS			
					GTACCTGAAG	CIAKLQKRSMNDVMVSFDVCSLFTN			
						VPLVETIDLCCALWNENDSEHHILDR			
						RAFRKLLEFATSNVNFLFNDEWYQQI			
						DGVAMGSPLAPTMASIFLASLEKKIA			
						SFQFEKPFVYKRYVDDIFLIFENQKHV			
						EPFLQFMNSLHKNIVFTCETERKSSIA			
						FLDLLIQRKDKQYETEIYRKPTDTGLYT			
						SPESFCEFKYKRNMVKGLIYRSWALS			
						STFANSVKSVDKLVELLIKNGYSKSFLL			
						LMIKETVDKLIKNPKDPCCGKSECNLS			

						DGCELSVDLGDCKNGYKKIEPKYVLVL PYSEGFTNYKRKLGKLIGNLEYKIVSNS CKVRNMFINKSKTPVGLCSDLVYQFT CNGCNATYIGETSRHLCTRVLEHCRL KGLTNISEHNRGCKSDIGMSDFRILLR SFNSYWERVICEALLIRSLDPKINVQS AVTTNVLNVFK			
	L1-3_MD	GCA_000002295.1	Monodelphis domestica	GACAGAGTGAA ACAACTTTCAAG AAAAGAAAAACC CAAAAGGTAGGC AGAGAACATCTG GGGCACTGGGG TGAGAGGAGAG CAGAGCCAGCAA GAGGCATAGCAA GAAGTGAAGAG CAAGCCACATAC CCCACCACACCC CACATCTGCTGT CCCAGGTTCATG	TAATTATTAAGAACT ATTTTGCCCAATTATA TGGCAATAAATATGA CAATCTAGGTGAAAT GGGTGAATATTTACA AAAATATAAATTGCC TAGATTAACAAAAGA GGAAATAGAATACTT AAATAATCCCATCTC AGAAAAAAGAAATT GAACAAGCCATCAAG GAACTTCCTAAGAAA AAATCCCCAGGGCCA GATGGATTCACAAGT	MRKKLKELKEEIDSKTILVGDLNLPLSELDKSNQKINKKEVREV NEMLEKLELIDIWRKLNRDKKEYTFFSAVHGTYTKIDHVLGH RNIANKCRKAEIINATFSDHNVIKIIINKGPWKGKFKINWKLN NLILQNWWVKKEIIETIMDFIKENDNEETTYQNLWDTAKAV LRGKFISLSAYINKIEREEINELGLQLKKLEKEQIKNPQIKTKLEIL KIKGEINKIESKRTIELINKTRSWYFEKTNKIDKVLVNLIKKRKK RIKLTVSKMKRVISPLMKRKLR			

				AACCTAAGCCCA	GAATTCTATCAAACA	
				AGCCAAATCAGA	TTTAAAGAACAACTA	
				CCCTGAGATCCT	ATCCCAATACTATAC	
				GCTGGGCAAATA	AAATTATTTGACAAA	
				GTGGTACAAGCC	ATAAGCAAAGAAGG	
				AACCCATACCCC	AGTCTTACCAAATTCT	
				TTGAAATTTCAA	TTTTATGACACAAAT	
				ACCTGTGGAGAA	ATGGTACTGATTCCA	
				GTCCAAGGCAAT	AAGCCAGGCAGACC	
				CTGGGCTGAAGT	AAAAACAGAGAAAG	
				GGGCTGATCCAT	AAAACTACAGACCAA	
				GTGGGCCCAGA	TCTCCTTAATGAACAT	
				GCAAAGCCAGG	AGATGCAAAAAATCT	
				AATCCCAGTCTG	TAAATAAAATACTAG	
				GGCTTGGGATAA	CAAAAAGACTACAGC	
				GGACATAATTCA	AAGTTATCACAAGGA	
				CAGTTTGGGGTG	TTATTCACTATGACCA	
				AGGTGATAGTAC	GGTGGGATTTATACC	
				TCAAGGGGAGCC	AGGAATGCAAGGAT	
				TCACTGGGATCT	GGTTTAATATTAGGA	
				TTTTGCCTAAAG	AAACCATCCACATAA	

				CTCAGGGCAGAG	TTGACCATATCAATA	
				CTCCAAGACAGG	ATCAAACCAACAGAA	
				GCAATCAAGGGT	ATCACATGATTATCTC	
				GTGCCTGATTGG	AATAGATGCAGAAAA	
				ATCAAGTACACA	AGCCTTTGACAAAAT	
				GTGAGACCAAAA	ACAACACCCATTCCT	
				AACTTGAGCCTA	ATTGAAAACACTAGA	
				AGCCTGGGGCTA	AAGTATAGGAATAGA	
				GAATTTGATCAG	AGGGCCATTCCTCAA	
				CCCCTGACCTGA	AATAATAAACAGTAT	
				TCAAGATCAGAG	ATATTTAAAACCATC	
				TGAGATCAAAAG	AGCAAGTATCATCTG	
				CTTACACCCAAG	CAATGGGGACAAGTT	
				CCTGAGGCAAGT	AGAAGCCTTCCCAAT	
				CTTTAAGCTATCA	AAGATCAGGAGTGA	
				GCATCTCAAGGG	AGCAAGGATGCCCAT	
				TCAATTGAATCA	TATCACCTCTATTATT	
				GCAGGGGGAGG	TAATATTGTACTAGA	
				GGGCTCTCAGAG	AATGCTAGCAGTAGC	
				CTCTCAGCCCTCA	AATTAGAGAAGAAA	
				GACCATCAGGTA	AAGAAATTGAAGGG	

				AACTGGTAAAAA	ATTAAAGTAGGCAAT	
				CCCAGAAAATAA	GAGGAAACTAAACTA	
				GCTGAGAATAAC	TCACTCTTTGCAGAT	
				ATCAAGGAAGG	GATATGATGGTATAC	
				ACTGAAGTTTAA	TTAAAGAATCCTAGA	
				GAGGGTACCCCA	AAATCAACTAAAAGG	
				ACTTCCCAGAAA	CTAGTGGAAATAATT	
				ACAGAGCCTACC	AACAACTTTAGCAAA	
				TATAATTAGATA	GTTGCAGGGTACAA	
				GGAAAAATGAG	GATAAACCCACATAA	
				CAAACAATCAAA	ATCATCAGCATTTCTA	
				AAAGCACCTAAC	TATATTTCCAACAAA	
				TCTAAAGAATTTT	ACTCAGCAGCAGGA	
				TTATGGAGACAA	GTTAGAAAGAGAAA	
				AGAGCAAGGTAC	CTCCATTTAAAATCAC	
				AGACACAGAAG	TCTAGACAATATAAA	
				GGGACAGTGAA	ATATTTGAATCTATCT	
				AGCAAAGGAAA	GCCAAGACAAACACA	
				CACATGCAAAGT	GGAATTATATGAACA	
				CCAAAAGAAAAA	CAACTACAAAACACT	
				TATGGATTGGAC	TTTCACACAATTAAA	

				ACAGATTCTGGA	ACTAGATCTAAATAA	
				AGAGCTCAAAAA	TTGGAAAAACATTAA	
				AGACTTCAAAAA	TTGCTCATGGGTAGG	
				CCAATTAAGAGA	ATGAGCTAATATAAT	
				GGCAGAGGAAA	AAAAATGACAATCCT	
				AGTGGGGAAGA	ACCCAAATTAATCTA	
				GAAATGAAAGTG	CTTATTCAGTGCCAT	
				ATGCAAGAAGAA	ACCTATCAAACTACC	
				AAGAAAGTGATG	AAAAAACTTTTTTATA	
				CAAGAAGAAATG	GAATTAGAAAAAAAT	
				AAAAGTGATGCA	ATAACAAAGTTCATC	
				AGAAGAAATGG	TGGAAGAACAAAAG	
				GCCAATTGAAAA	ATCAAGAATATCAAG	
				AGGAGAACCAA		
				AAACTGACAGAG		
				GAAAATCAGGTC		
				TTAAAAATCAGA		
				ATTGACCATCTA		
				GAAACTAATGAT		
				TTCATGAGAAAT		
				CAAGAAACAATA		

				AAGCAGAATCAA AGGAATGAAAA AAACAGAGGAA AACATGAAATAT CTTATTGAAAAA ACAACTGACCTA GAAAATAGATCT AGGAGAGACAA TTTGAGAATTATT GGACTACCTGAA AGCCATGATCAA GAAAAAAACCTT GGACATCATATT ACAAGAAATTAT CAAAGAAAACTG CCCAGAGATCCT TGAACAAGAAAA TAAAATTGAAAT TGAAAGAATTCA CAGATCACCTCC AGAAAGAAATCC		

				TCAAATGACAAC TTCCAGGAATGT AATAGCTAAATT CAAGAGCCACCA AGCCAAGGAGA AAATACTTCAAA CAGCCAGAAAGA AACCATTCAAAT ACCATGGAGCTA CAATCAGGATCA CACAGGACCTAG TAGCTTCTACATT AAAGGATCAAAA GGCATGGAATAT GATATTCAGGAA GGCAAAAGATTT GGGTTTACAACC CAGAGTCACCTA TCCAGCAAAACT GAGTATATTCTTT CAGGGGAAAAA		

				ATGGACATTCAA TAAGATAGAAGA TTTCCAAGCATTC CTGAGGAATAAG CCAGACCTAAAC AAAAAATTTGAA GTCCAAACATAA GACACAAGAGA AGCCCAAAAAGG TAAATAAGAAAG AGAAAATTTAAG GGACTCATTAAG GTCAAAATGTTT ATATGTCTACAT GGAAAGAGGAT TTCTGTAATTCTC AAAAATTACTCTT AGTAGTAGAGAA GATAGAAGGAAT TTACTCAGAAAG AGGGTGGAGAA		

				GTAAGCTGATTA TGATGATATGAT GTATATGATGAT ATATATGTAAAT GTGATGATATGG AATGATATGTAT GTATGATATGAT ATGTATGCATAT GAATGATATGTA AAAAATCAATCA AGGGCTGAAAG AGAGGGTTGCAC TGAGAGAAAAG GGAAGGGAGAG ATAGAATGGGGT AAATTATATAAC ATAAAGAGGCAT GAAAAATCATTA TAGAGAGGGGA GGATAAGGGTG GTGATGGGCAAT		

				GCTTAAACTTTAC TCTCATTGTAACT GGCTCAGAGAG GGAAAAACAAGT ATACTCAGTGGG GTATAGAATTCT ATCTTACCCTACA GAGAAGTAGAA GGGGAATAAGA CAAGGGAAGGG GGAGGTAATTGG AGGGAGGGGAA GTGGGGGGGTG ACAAAAAGCAAA ATACTGGTGAGG AGGAACAGAGT GAAAGGGAATA GAGCAGGATCTA AAGGGGATAATA TGATGGAGGGC AATACACAGTTA		

				GTAATCATAACA CTGAATGTGAAT GGGATGAACTCA CCCATAAAATGG AAGTGGATAGCA GAGTGGATTAAA AACCAGAATCCT ACTATATGTTGTT TACAAGAAACAC ATTTGAGGCAGG GTGACACACACA GATTCAAGGTAA AAGGCTGGAGC AGAATTTATTAT GCATCATCTAAA GTAAAAAAAAGC AGGAGTAGCAAT CATGATACCAGA CAAAGCTAAAGT AGAAATTGATCT GATTAAAAGAGA		

				TAAGGAAGGAA ATTACATTTTGCT AAAAGGTACTAT AAACAATGAAGT AATATCATTACTA AACATTTATGCA CCAAATGGTATA GCATCTAGATTT CTAA					
	Poseidon-1_PM	GCF_010993605.1	Petromyzon marinus		TGATTGATTGTTCAC GCTTACAGTGGCTTT CAGACAGCATACATA AAAAATTTCGTCTCC CCCCTCCTCTTTTCAA AAAATAGATGGACTA GTCCTGCTGGGCAAA CTGCCAAATCGTTGC GCGTGACGCAAACGT TTTTTTTAAATAAATA AATATGCCGTTTTCA AATAAATAAATTTAT	MNRLYGEGFHITLRSFNKLRDKKAQL IESLTFLKRCRDSGVIPIAFTLTSATVTS SNAKRILFKTSLALISEKIRLIRTELDQL DKTLYKSHLSLTAIMSTSDWEHVDN MTASSAATMSTKCKENLISKFNLLFN KRPDPHTKKLDVNKVVINRSNSSISD GLKSVLAKGLNFAVTPKHIPVNDIIAQ VDVALRGLGHIAADVARSEIAATLKN AKMPASNLSKDERFELYKLKKRNDIV VLSADKGNATVIMDKIEYNNKIKELLD DGSYTKIKRDPTNSLVTQTNKIISDST ALIDQSKQFSLKPSAARPPRLYGLPKI	cd00304	440	524

					GCAAACCGACACACC TGATGATGACCGTTT GAGTTATGGTCGAAA CGTCGTGAGAATTTT TATTGTTTTATATATA TTTTGTTATTTTGTTG TTTCTCTTCAACGTGA CTTTACCGAAAAAAC TAAGAATATGAATTA CTGCAGTCACGAAAG CTTTAAATCAAAA	HKVDTPLRPIVSTISSPTYKLARFLSEIL RPFTGKTTSAVQNSTHLVSTIKGISLSP DDLLVSFDVVSLFTKVPVPDTIAIIEDL VHQGLNTEVPPLVRHCLTNTYFTWN GTFYKQTEGTPMGSPLSPVIANIFME KFETDAIQTATLRPTLWKRYVDDTFVI WPHGRPALDQFFDHINSIHPAIRFTM EVEIDGKIPFLDVLIKKRPDGSMSHSV YRKATHTDAYLKPDSHHHPSQKNSLI KTLYHRAQCISDQEHISSELKHVHQAL RANGYSDYTIRKAITQTMKKNPLEEP FKTITLPYIAGITKKISKILNKHKIQVRF NTLHKIRDLVPSVKDVIPDIQYPGVYK LSCLCGSSYIGETKRHISDRIREHKADL KHGRTNTSAVADHCSNSVGSHAIDF SKTEVLCKPSGFHERFVHEAIHIYKHR DNFNREDGFYLSNHWKDLINHFSS			
	Poseidon-1E_PM	GCF_010993605.1	Petromyzon marinus	AGCTTAGCTGAT TGGAGAATATAG AGTTTATTGATA ACATGAAGGTAA	TAATTAATTGTATTTT ACGATGACCTTCAGA CCGCAGCTCTTTCGC CAGCATTGCCTCTTA	MNHCSHESFKSKFNRLYGEGFHTSLR SFNRLRDKKAQLIESLTFLKRCRDSKVI PTAFTLTSAVATSLNAKRILARTSFALI SEKIRLTRTELDQLDKTLYKSHLKLTAT	cd00304	453	537

				TTTATATTTTAGT	ATACCTCTCACACCAC	MSTSDWEHVDNMTTSSAATMSNK			
				TTTTATTGGTTCA	GGTTAATGGTATGGT	CRDNLISKFNHLSSKKSDPHTEKIDVN			
				CTATATAGTACT	TTACAACATGGAGTA	KVVINRSNSSISDGLKSVLAKGLNFAV			
				GTGAATAGCTCA	TACTAGGGCTTCGCA	TPRHLPVNDIIAHVDVALKGLSHISAD			
				GATGTGTAGGCC	GCTGTCTCTCTCCCCC	VARSEIAATLKNAKLPPSNISKDERFEL			
				TTGACTTCAACAT	TCCTCTGTTCCGACCA	SQLKKRDDIVVLSADKGNATVIMDKT			
				AGTCATTTCTGT	TGGACGAACTCGTCC	EYNNKIQELLNDGSYTKIKRDPTNSLV			
				GACCAGTCACTA	GTCTGAACGGGGCCC	TKTNKIISDSTALIDENKQFSLKPSAAR			
				GCAAATCACCCA	GTAGCTATCCGTGGT	PPRLYGLPKIHKEDTPLRPIVSTISSPTY			
				GTTATTTCAATGC	ACGAAATTTGAGGGT	KLARFLADILRPFTGKTTSAVLNSTHL			
				TAAATTCGCCTA	AAGAATCCACCCATT	VSTIKDISLSPNDLLVSFDVVSLFTKVP			
				GTTATTTCAATGC	TTTTTCGCCATATAGC	VPDTIAIIEDLVHQGLHTDVPPLVRHC			
				TAAATTCTCCTGT	ATTGTGATATATAAC	LTNTYFTWNGTFYKQTEGTPMGSPL			
				TTGATTTGTACTA	GACTTATGGTTGGTA	SPVVANVFMEKFETDAIQTATLRPTL			
				AGTGCCTCGGTG	GCTTAGCTGATTGGA	WKRYVDDTFVIWPHGQPALDQFRD			
				AATGTGTATGCT	GAATATAGAGTTTAT	HINSIHPAIRFTMEVEKDGKLPFLDVLI			
				CTGACTTCAACG	TGATAACATGAAGGT	KRRHDGSMSHSVYRKATHTDAYLKP			
				TAACTATTTCTGT	AATTTATATTTTAGTT	DSHHHPSQKNCLIKTLYHRAQCISDQ			
				GAGCAGTTACAA	TTTTATTGGTTCACTA	EHITSELKHVHQALRANGYSDYTIRKA			
				GCTAATCACTCA	TATAGTACTGTGAAT	ITPVMKKDPLVEPLKTITLPYIAGVTKK			
				GCTATTTCAATG	AGCTCAGATGTGTAG	ISKILNKHKIQVRFNTVHKIRDLVPSVK			

				CTAAATTCTATGC	GCCTTGACTTCAACA	DALPDIQYPGVYKLKCLCGSSYIGETK			
				TAATTCCATTGTC	TAGTCATTTCTGTGA	RHISDRIREHKADLKHGKTNSSAVAD			
				AAGGGGTGCCTC	CCAGTCACTAGCAAA	HCYNSVGSHDIDFSKTEVLCKPTGYH			
				TGCGGATGTGTA	TCACCCAGTTATTTCA	ERFVHEAIHIYKHRDNFNREDGFRLS			
				TGCCTTAACTTA	ATGCTAAATTCGCCT	NHWKDVLNHFCS			
				GCTATTTCAATG	AGTTATTTCAATGCT				
				CTAAATTCTATGC	AAATTCTCCTGTTTGA				
				TAATTCCTTTAAC	TTTGTACTAAGTGCC				
				AATGGGTGCCTC	TCGGTGAATGTGTAT				
				TGCGGATGTGTA	GCTCTGACTTCAACG				
				TGCTTTAACTTTT	TAACTATTTCTGTGA				
				AACATAGCTATT	GCAGTTACAAGCTAA				
				TCTGTGACCAGT	TCACTCAGCTATTTCA				
				CACAAGCTAATC	ATGCTAAATTCTATG				
				ATTCAGCCATTTC	CTAATTCCATTGTCAA				
				AATGCTAAATTTT	GGGGTGCCTCTGCG				
				ATGCTAATTCCTT	GATGTGTATGCCTTA				
				TGTCAAGGCTCT	ACTTAGCTATTTCAAT				
				TTGTTCCTCTCAC	GCTAAATTCTATGCT				
				ACAGCTTAAAAG	AATTCCTTTAACAAT				
				CCGTGATCTGAA	GGGTGCCTCTGCGGA				

				CGTGGCTCTCTC ACTCCAGCGCTG TCCTACCAGACA GTCTCGCTCCCTT CACCTGATGAAG GCCGCTTGTGTT GTGGCCGAAACG TCGTGAGAATAT TATTGTGTTATTT TATTTTATCAATT TATTATTTTATTC TCATTCTACGTG ACTTTACCGAAA GAACCAAAGAA G	TGTGTATGCTTTAACT TTTAACATAGCTATTT CTGTGACCAGTCACA AGCTAATCATTCAGC CATTTCAATGCTAAA TTTTATGCTAATTCCT TTGTCAAGGCTCTTT GTTTCCTCTCACACA GCTTAAAAGCCGTGA TCTGAACGTGGCTCT CTCACTCCAGCGCTG TCCTACCAGACAGTC TCGCTCCCTTCACCTG ATGAAGGCCGCTTGT GTTGTGGCCGAAACG TCGTGAGAATATTAT TGTGTTATTTTATTTT ATCAATTTATTATTTT ATTCTCATTCTACGTG ACTTTACCGAAAGAA CCAAAGAAGATGAAT			

					CATTGCAGTCACGAA AGCTTTAAATCGAAA				
	Poseidon-2B_PM	GCF_010993605.1	Petromyzon marinus	T	TAAACTAATTTTCAAT GGTACAGTAAATTAT TTTTCAAGTAGTGGT ACCCCAAACTTTCGC AGCTGCTTATAAATC TCCCCTCTTTCACACC ACTCAATAGGTAAAC TCACTCCCCTCCCTCC CCCCCCCCGCGCCAC ACGCCAACTCCACGC ACGTACTCCACGCAC TCACTCTCTCACTCAG TCAATCAATCCAATC AATCTGCTAATTTATT TATGCCACTTTCTTTT TCACTCCACCCATTCG TTGCTCTCACTCAGTC CATTTACTAGGTGCG CTCGGTGACTGGATT	MELDTLDRTLLQAHLMLTQKMSPD WNHVDNMTASSAESLSVKCKANLIS KFNHLSKKNPVPHITKLDVNKIVINQS STPISDNLKTVLARGLNFAVTPKHIPV DDIISQVDVALKGVGHIQADVARAEI ATILKNAKKPISNISKNERSELANLKNR SDIVILSADKGNATVIMDKDSYNNKI QNLLNDGSYSKIKNDPTNSLAIKTNKII TSSTSLIDPTKRPSLKPSAARPPRFYGL PKIHKEGTPLRPIVSTINSPTYRLARFLS DTLLPFTGKTASAVLNSAQLVSMVKD VVLSPDDILVSFDVVSLFTKVPVPDTIE IIENLVQHGLDPDTPALVRHCLNNTY FTWNGSFYKQNEGTPMGSPLSPVIA NIFMEKFESDAIQTSTLRPSLWKRYV DDTFVIWPHGLPALNLFLHHINSRHP AIKFTMEVEKDGKLPFLDVLFERRPD GSLGHSVYRKPTHTDSYLKPDSHHHP AQKNSLIKTLYHRAHCISDQHHLTSEL	cd00304	361	445

					TACTGCTCTGACTGT CTTAACTAGCTGATTT CACTGAGTGCACTCA GTGAATGTAATCACT CACTAACTTCACTAAT TCTACTAATTCCTGTC AGTCCACTCAGTCTA CTAACTCACTCAATTC AATCTAAATTCAATTC TAATTCCTCAAGGGG TACCTCTGCGGATGT GTCATGCTACATACC CGCTTATGGTCCATC ATGGCCCATTCACCT CCCAGTTTCATCATG CTAAATTATATGCTA ATTCCTTTGTCAAGG CGCTTTGTTTCTTCTC ACACAGCTTAAAGGT CGTGATCTGAGCGTG GCTTTTTTCACTCCAG	KHVHLALKRNGYSDRTIRQAVSPSIQ KQPLDTPVKTITLPYIAGVTKKISKILN KHKIQVRFNTVHKIRDLIPSVKDVIPD LQYPGVYKLSCTCGDCYIGETKRHVS DRIREHKSDLKHGKTNSSAVADHCYN AVGDHDIIFSDTKVLCKPTGFYQRFV HEAIQIYKHRHNFNREDGFKLSNHW KDVLNRTKH			

					CGCTGTCCTCCAGAC AGCTCGGTCCCTTCA CCTGATGAAGGCCGC TTGTGTTGTGGCCGA AACGTCGTGATTTTT GTTATTTTATTTTTTA CCTACGTGACTTTAC CCAATAGAACCAAAG AGTATGGATCCTTGC AGTCACGAAAGCTTT AAATCTAGAA				
	Poseidon-1B_PM	GCF_010993605.1	Petromyzon marinus	CA	TAACTGGCTGTGATG TTTCCATCCTACCTAC CGATGTTAGCAGTTA TTTTCGACTCTCTGAG TTACATTCACTAACAC AGCCAACCGTTAGCT GTGTTAGTAAATGGC GGCTTGCCACTGCCT CGTCCATGCCCTGGC CGAGCTTGGCCCCCT	MGSPLSPVVANVFMEKFETDAIQAS TLRPSLWKRYVDDTFVIWPHGLPAL NQFLNHINSIHPAIKFTMEVEKDGKL PFLDVLIEKRPDGSMGHSVYRKATHT DAYLKPDSHHHPSQKNSLIKTLHHRA RCISDHHHLASELEHVHQALKMNGY SDYTIRKAITPTVGKEPSEDPIKTITLPY IAGVTKNISKIMSKHKIQVRFNTVHKI RDLIPSVKDVIPDMQYPGVYKLSCVC GSSYIGETKRHVSDRVREHKADLKHG	cd00304	1	83

					CCGATCTTACATTTAC ATGCAAATTTCTCTCC TACTAGCCCCGCCCA CTCGCTCCTACATGA CCCTTGGCTGCTACA CCCATCTCTAAACCTT TATAAACAGCCTGCC TCAGCAGCTTTCCCTC TTTTTTCACCTGATGA CGATTGCTTGTGTTG CGATCGAAACGTCGT GATTTTTAATTTTTAT TATTTTTTCTATCTAC GTGACTTTACCGAAA GACCCAAAGAATATG AAT	RTNTSAVADHCFNAVGSHEIDFSKTK VLCKPTGYQQRFVHEAIQIYKHRHNF NREDGYKLSNHWKDVINQFKS			
	L2- 24_ACar	GCF_00009 0745.1	Anolis carolinensis		GATACGGTCGATATA ATCTGATGTATTCGC TGGATTAATGTGTAA TTGAATACAGACTGG CTCTGACTTTTGGCCC	MPIKMVRPRRLMDPDGFLRSLGDLP ASSRGDPVDALVTRYNSEMARAIDTI APERPLSLRRAKPAPWFTEELAVMKR ERRGLECVWRKLRSESDRARAKALRK AYSVALRAAKKSFQTARIAAAKNRPS	cd01 650	257	517

					AATGCTACTGGCCCT GTTGTAATGGCCACA CAGTGCTTTACTGTTT TAAATGGGGTATAAT ATTGGTTTTTAAGTG TGTTTTTACGATCGTT TGTTTTACTATATTTT GTATTATATGTGGCA TTAATCTTGCCATTAT GTAAGACCGCTCTGG GTCCCCCTGGGGGA GAAGAGCGGTATATA AATTAAATAAATAAA TAAATAAATAAATAA ATAA	ELFRVVGELLNGTPQTGPPDDLAENC AAFSHHFADKVAQIRSELDATVDAVS EEVARAPVCPILMDSFQLVQPDDVD KILGEVRATTCALDPCPSWLVKEARG GLVEWVRVVINASLREGIFPSCLKQA VVKPLLKKPSLDPTNWNNYRPISNLP FLGKVLERVVASQLQGFLDDTDFLDP SQSGFRPGHSTETALVALVDDLRREL DRGSVTLLVLLDISAAFDTIDHGILLGR LLGMGLGGTVLQWLQSFLEGRSQM VKLGDTCSDPWPLTCGVPQGSILSP MLFNIYMKPLGEVIRSFGVRCHLYAD DTQLYYSFPPASKEAPQILNQCLAAVS DWMRANKLKLNPDKTEVLLVSRKSD RGIGWQPVLDGVALPLKAQVRSLGV LLDSSLTLEAQVSAVAGRAFAQLKLV RQLRPYLVKSDLATVVHALVTSRLDY CNALYVGLPLKTARKLQLVQRSAARL LTGANYRERTTPLFKELHWLPVIFRA QFKVQVITYKALNGLGPTYLSDRISPY EPARRLRSSGEALLSVPPPSQVRLVGT			

						RERAFSAVAPRLWNSLPREIRQAPTL LSFRKQLKTWMWDQAFGPS			
	Jockey-8_DPer	GCF_003286085.1	Drosophila persimilis	GGCTAATGCAGC ACACAAATTTAC TTGCCTGTGAGA TCGCAACGCACC CATACATTCCAA GAAGGCTGGCCA GGCCAAGGTATG CCGCAATGATAG AAGAGCACAATC CAAACAGCAACA GAAGCCTGACAG CCATCCAAACAA TACTGCTCCAAG TCAGACGCGAGA ACTGGCTGCGCA ATCCATTTGCAA CACACAATATGG AAATCAGCCGCA GTAGCGCTAATG	C	MSASLPPANAGGDGNWFDMKRDY QFSSQDEHNRNNIPHISLAPTREDSIQ MCTIAAKIKNHQTINIGAIYCPPAHE WIDTQFKQVFDSLGSRFIVAGDWNA KSQWWGNARSCRRGAALLKCIQEN DQSILATGSPTHFPSNRRHMPSAIDF AVYKGIRHNLIHIETSYDLGSDHNMQ IIKLLLRPEYGKKKILLPKNANIRRFQA HLEQRIHLETELHNGDDIDDAVAILEK NIYIAAELASQTSSNQHTVYGPQNFH LDSQTLKAIKIKRILCKINQKDRCSNPLI ARTINQLTNRISKRIKLQKDKQMNRL LHDIDPTDRYSMQKLWRLTNTIKRQP IPNLPVRRHVAEGQQAPADPWCKTT EEKAEAFAQHLEARFTPKITAALEDRE EINAVLQENGPDEAAQIIRPVTIQELK SEIKKLKTKKAPGLDSIDNRVIKSLPTS AIEYLLLICNSALQFGHFRKSWKCAKIS MLRKPGKPSDRLESYRPISQLTGFSKIL	cd01650	470	731

				CAGCTGCTACGC AGGCAAATGACG GCCCCAGCAGAG AACCTGATGATG GACGGCGGCAG CTCCAGAATCCA GATGATGGCGGC GGCTGAAAAGA AACTAAAAACCT TTTTATTCTTCCA CAGATACGAAGC CACTTCACCTGC ACAGAAATTGGT AATCCACATCCA GAATGACGATGC ACTGGACCTCGG CCAGCTTCAGCC CTGTTGATGCTG TCCTGCAGTCAC AGTTTAGCGCCG ACACTCGCAAGG		ERLILKRLFEEESFAMAIPNHQFGFRS QHSAEQQIGRVTQFILRSYEAQRHCS AVYMDIQEAFDRVWHQGLLCKVKTII PDNLYILLTSYLSDRSFHVVCCNNITSS LKHITAGVPQGSVLGPVLYNLYTADM PQPPRHANSTQMLATYADDAVYLVS ALSPKQANENVERLLADALSWSKKW GITISLAKTAHVLHTLSPVQQHTPAPR LAGQLLAKKSVMPYLGIHMDTKLNF NKHVAAIVTKTRARIKKLDWIIGRNSK LPLPPKVLLYKQLIAPIWQYALPVWG PLTSSMQMAKLQTIQSKALRAMTNA PRYVSNNALHADLQVATVEEVYNRL CRNCYT			

				ATTCCACTTCGCA GCCGACGACGA GACCAATAGGCG TAATGATGTTCTT CACGGCTGCATA CTCCACCTGAAA TAATTAAAAACA AAAACAATGTAA ACAAAGTAGCAA TAAGTTTAGTGT TACCCTGGCATTT AGCCAAATTATC GATAGCTACTAA AAAAAAATAAAT ACCAAATAAGCC AATACCGTAAGA AATGTTACCCTG CCAATTTACAGC CACTTCTGCAACT ACATGTAATGTA GATAAGCAGTTG					

				GCGCGAAACAAT TTGTAAAATATG CTAATTATAAGA GTATAAACAAAT AAAACGGTTCTC TCTTC					
	R4-1_PH	GCA_00158 7735.2	Parhyale hawaiensis	CCGACCGCCAGC GGGATAACTGGC AAACCCTGTCTC GACCACCGGCCC GTGAATCCATCG GGGCGTATGAGT CAGACACAGGG GGGTGTTTAAGG TGACCCTGTTGC GAGGAAATGCG CAGCAAAAGCCG GATGAGCCTTAG AACATCGAAGGC CAACGACAACCT GCCGAAAGACTG	TGAGCCTTAGGTCGC GGGATGTGACCCGG CGCCAGAGTGTAGA GCTGAACATCGCTCA ACCGATCCAATTTGG GTCGTGAAATCCCCT CGATAATAATAATAA TAA	MKMSHNRDTPSNGVKGTSVRLGTSL VRSPVGEAGAVRERGTHPSESVSQD SDASVNATGEGSVREQAPLSPPGAEE ATVPTQRRTRHKWSKEDRVVLWECF VASKREGPGYLKRLKQLWDERGIPG NFPQASLSGQIRQICSKNLLSEEERLQI AARMEAQVASPSADEPARQVPTRPV TPPRSPPVEPARRPSIPSEETPDLGAV PSEIDSADPNRSPSRGPRHLPAHNMS QSESEDDVTDPDVSDQQRSDSLEPR DLLRNSSVESTPGHPNQELSDTLLSNY VPSEIDSDDPNQSPRRGPRHLPAHD MSLSDSMDEETEPDLSDQQRSDLLEL RDLLRNSSVETTPKGHPSLRHLPEPKI RAAAFRVNSVIGKIHTNNITETNALIK	cd01 650	616	879

				GCGGCTAATCCA		AGADLAVRILEVQPRPQRTQRKKDP			
				GTCAACCTCCTG		PWKHRLEKNIAEIRKHLSWISEWRRG			
				TAGGTCACCGGC		NLHDEEKKTLLESRYRCLEVGLTNLED			
				TGGTCATGCTGA		TLKQRLSAKRSKVRRFEARVAGFHQN			
				ATCTCAGCTTAA		QLFNTNQKRLYQTLRGEETSSDSPNA			
				CCAGGCGTACGA		EESIRFWSDIWSKEVRHNNAAEWLH			
				ACTTACAGTTGG		DVKEKNVAADPDLTITSQQLKKQLSK			
				AGGGTCGAGCAC		TKNWKAPGPDMVQGYWIKTFTSLH			
				CCCTGATGGCTG		SRIAAQLNHCLQRGTVPTWMTTGKT			
				AAAAGGACCATC		VLIQKDKAKGTEVSNYRPITCLPLMW			
				AAAGTCGAAGGT		KVLTGIIYERVYQHLDSKKLLPDEQKG			
				AGCCAACGAGAA		CRRNTRGTKDHLLVDKLLTKDARSKK			
				GACCAACGGCTG		KNLSMAWVDYKKAFDMVPHSWILE			
				ATTCAGGCGGAA		CLDIYGIAGNIRNLIATTMPNWKTQL			
				GAGTCAACTCGT		TSANKHLGEVSIRRGIFQGDSLSPLLF			
				TGAATGCGTTCG		VLTMIPLSETLNKAGQGYNYSRTMKL			
				ACAGCTTGGGGT		NHLLYMDDLKLYAKSKDQVEQLLNIV			
				AGATGGAACTCC		HQYSQDIKMQFGISKCGVLNIERGEV			
				TAAGCCCTGAAA		TASEGIIIEEGTIKDIEETGYKYLGVME			
				GGCAGTCCATCT		YNTILHRTMKDSIRKEYLTRLRLILKSH			
				TCGCAGACGCTA		LNGGNTIKAINTWAVPVVRYSAGIIN			

				AGGTGCCCCGCC GCCTGAGGGTTA TCAGGCCCCGCC GCCTGAGGGTTA CCAGGCAACC		WTKKDCTDMDIKTRKLMTIYRALHP RSCVDRLYINRREGGRGLISVEDCVEA EKRALSQHFRESDDPWARCLIEAKLL KETETADQFKERRRLDRTNKWKSMK MSGQYLEAVQDKIVPDSWNWLLRG ELKRETEGTILAAQEQALRTRYIQNKI DKRNVPSTCRICRSSDETINHVISECG VLAQKEYKRRHDKVARHLHWTLLRIH NFPVSERWYEHEPAPVVENEAVQIY WDKRMETDRVLHANRPDIVVKDKQ EKSAKLIDVSIPFDSRIIDKEAEKKEKYR DLAIELQRLWQMKVDVVPVVIGALG AMSKNLKTALRELKCGHLHPGTLQKS ALLGTAHIIRKVL			
	Proto2-4_CTe	GCA_000328365.1	Capitella teleta	GATGAATTGTAT AATTGTGTTGAT TATGATCGTGAT TCA	TGATTAAGCTACCTTT ATGTTGCAGTTTTGA TTTGTTTGTTCGTTTG TTTGATTTTCTTTCAC CTACTATAATCCTGTA TATTTATTTATACGAA	MNELIAFFKISVQNVCDQGSCYSSHQ FTAQGVRSAVSKLKDKSDCKDNFMS NHVIHGGDTLFVHLSFLFNCMISHTC DPKQLRMSVLIPIPKNRRKSLSVSTNY RSIALSSIVGKVFDRMILTEHVNIFSTS NLQYGFKPKHSTVQCTFVLQNVINHY KSRGSPCYVMLLDASKAFDRVQYVRL	cd01650	85	341

					TAAATTATTATTATTATTA	FYLLRERGLCARIFMLLALLYTNQSMCVRWKNCISDNFSCSNGIKQGGVLSPTLFCIYFDELLIRLKHCGSGCHLGHLFTGALSYADDVALLAPTLTAAKVMLNVCEQFANNFNVLFNADKSALLVCGRPIRKEVNLRLNGRIIQQNSTAVHLGTTIGDNSSTRNVKKAASDLYCLTNVLMCKFGHCSIDVLSALFFFFFYCSSFYGCPLWDLRNIEPILTAWRKCLKRIWKLNIRTRTKYINNLPKGNLISALLIRSLSFYMQCLSSSNAIVKLASNYCFYGCMYSVISENVHLCSVLSSVPLRFIFDDFENVKSIIQRGMYDEATIAESLALKDLCYLRDGLMFLNFEITSEELLCAFNHFTI			
	R4-1_AC	GCF_000090745.1	Anolis carolinensis	CCCCCCCGCAGGTGCCACCTTGACGTGGTGGGGGGGCTTAGCTGCTCCAATGATGACTGAGGGCTGTGCTGGCGGTAGTGTAA	TAGACACTTGGGAAGTGTTCGACTTGTGGTTTTGTGAAACGAAATCCAGCATATCTATCTTGTTTGCTGTGTCATACAACGTCGTTGTGTTGAT	MENQTQTKSILARSSPRIKRTAVDAADSGHPSTSGLRNHQNPNEQSQKRQKYTMPENRTIMKCYYKSEPQRRGYQKRMHQLWKQEYPDSQITEPRLADQRRFIIRNKVFSEVELEEIQKICKVNYHQTTAQTAAETPATLGMVEQIEPEESVELLQEFEEPALETSVEPPGTLTARQQELK	cd01650	465	731

				CTACCGGCAGGT CCAACCAAGCCA GAGAGGCCTCAG CTGAGGAGCACA ACCAAGAGCACC TAACCCATTAGC ATT		DKIMAHAAANAIRKRLPTLKTVPKRH LAPLMKDVNAALSTVQITSIEQTNQY AYSAAVIVTEELGLLQPRQPQRKSTG KPKWKVRLELKIKKLRSDASNLKNMK EKKLKNDKIKQYLIRKYWLNTRKIEEA LEIVKEQITATARKIERYEARIIQYRQN QLFQSDQRRFYQSLNQTTDTVTIKPE KTATTKFWKELWENNKNYNKNAGW IKEFEGKFSQNKMELMEITTEMISKR VQKVKNWTSPGSDQLHGFWLKHLIS LHGKMAQQFNEMLQKGSISEWLTT GRTYLIQKDPAKGAAPGNYRPITCLPT MFKLLTGIIADRIQDYLEEKNILPDEQ KGNKRKSRGTKDQLLIDKMILENCKS RKANLHMTWIDYKKAFDSLPHSWIIK CLDAIGICKNVGTFIENMMEHWKTE LFVGNESYGLVNIRRGIFQGDSLSPLL FIIAMIPLSTILQKTNLGYQTSKNSHKI SHLMYMDDLKLYGKTETEIQSLTNTV RIFSTDINMEFGLDKCSTVALKKGKIIE SEGINMPNGQTIKCHQPEAYKYLGIL			

						QLDNIKHEHVKTVVSKEYTQRVRKILK SKLNGGNTIKAINTWAIPVIRYTAGII NWTQVELDNLDRKTRKLMTIHHSLH PRSDVDRLYLPRRSGGRGLLQVKQA VKEEEHALAEYVKQSEEPALIEVKNQ KLLKTQQTKNQYKKTALQTRADSWH NKTLHGKFLDKIEGKADKEKTWLWLT NGTLKKETEGLILAAQEQAIRTNAIKA KIEKSADDPKCRLCKETDETIDHILSCC KKIAQTDYKQRHNYVAQMIHWNLCL KYHLPAAKNWWDHKPAKVLENEHA KILWDFRIQTDKVLEHNTPDITVVEK NKVWIIDVAIPGDSRIDEKQQEKLSRY QDLKIELQRLWQKPVQVVPVVMGTL GAVPKDLSRHLETIDIDKITICQLQKAT LLGSARIIRKYITQS			
	CR1-66_AAe	GCF_002204515.2	Aedes aegypti	ATCGCTGTTCGTACCTGA	TAGATTTTTATGTTTTTTTTTCATTAAGACAAACAGATGTCAGATGAATAACAACTTCAATAAATAAATAAATAAA	MYSKRFQPSNETLAISHDVQPGCTPVSLVEASYPLVTVEPILPATSSHPGPVYEYGEEVFQTLFAGKYSCDSSNFSPVTSVASSDIPTRKSDTKSAAPDDFASDNPANRLPNVITDDLNNNQSCHALRFYYQ	cd01650	638	885

					TAAATAAATAAATAA AT	NVRGLRTKVDSFFLAVSECDYDVIVLT ETWLDDRFQNAQLFGSKIIVYRTDRN AANSSKRRGGGVLIAVSTHLNSFIDPT LTSIALEQLWIKISTPSRTISVGVIYLPP DRRDNMDYIRDHLNSMNSVLSNLAL NDCVLQFGDYNQPNLLWEASDDGF LRVNSQSILSTAGSTLVDGFHFHGLT QLNGIPNRNSRFLDLVLGNDEARYSC NVTEAIEPLSDLDVDHPALDVILNLPD SITFEDCSDNSQMYDFRQADYAGIH AAISRIDWQPIDDANNIDDAVTYFNH VVQDIIAQNVPLCRPKPKPPWSNSHL RQLKRNRAKSQRHYCRNRSQFTKQE FNRASNAYRTYNRLLYRRYVVRTQTS LRRNPKRFWSFVNSKRKEVGLPRSVH LDGQTATSEQEKCFLFAQHFQKTFH DSPATADQIAAAIEYTPRDVFDFRSFT IDEQTIQAAFRKLKMSHAPGPDGIPS CILKNCSHSLATPLAKLFSRSLQEGKFP DEWKVSSMFPVYKKGDKKDVHNYR GITSLCSCSKLFEILINDALFNSCKSFIST			

						AQHVFFPKRSVTTNLTEFVSLCLLTLD SGGQIDTIYTDLKAAFDRVDHGILLAK LDKLGVSHSLVRWFESYLTNRLLYVKI GSSQSESFTNLSGVPQGSNLGPLLFLL FINELSQLLPPGCRLFYADDVKIYMIID STADCELLQSRVNIMANWCSSNFLTL SVNKCNVISFHRRQKPIIFDYRICDTVL NRVDQVKDLGITLDQEMNFKPHYN DVIAKANRQLGFIFKLSDDFRDPLCLK GLYCSLVRSILEFSVVVWCPYQSCWI ARLESVQRRFLRYAFRHLPWRDPLNL PAYEDRCRLLGIETLEQRRRRAQAVF AAKLIVGDIDSPELLRQLNIYAPEQVIR QRNFLQLVPCNRLYGMNEPLRASAD ALNEAFTAFDFNLPLISFIRRLSLS			
	DongAa	GCF_00220 4515.2	Aedes aegypti	GGTGGGAGCTCA GGCAACATAACT CCTGGGCGTCCA TGAAAAACCACC ATCTTGAGTAGG TGGGAGCTTGAG	TGAGAAGCTCATCCA GCAGACTCATTAGGA TAAGTTAAACCGACC AATGAGATCCACAGA GCCTAATCCCCTTTG GCATACAGATTGCCC	MKANPAMETRSMRNRRMRSPEEG APTGAGPGTGDRASGQRLEEDNVQ ERPSSQRAQPVPRTRNRNGSSRNHQ GHAAPTDVAVADRRQSLTLAGGRRQ RIMWTREMNMYVIRCYYVCTRMET DMSGRSGMLEMFNERFPRFARQLD	cd01 650	530	802

				GTAAAATATCCT GGGCGTCCATGT AAAACCACCCTC GGCGAGCAGCG GCGCTTGAACCA AGCTATTAAGCA CTGTAATTGGAG GAAATGCCAATG CAGAACCCGTAG CTTCCGGGA	GGGGTAGGTGAAAG TTTCCAGCATTTTTTG CTGAGAAGTGCCAAA ACTCTATAATAATAAT AA	LNKLYTRQRAIMSHNMLTAAEVEYIK LEVQREIGEEGTRSSDTSRRSSVGLDA PITSESRDSPAPTAPGPTADMQRQQ LRDELALHMGTAVTQFRGTDPMSR HRIPKLRYSYRLTSAVSILNQDILPQYL EAVENLEDLQFIVYSAAIAVVKTLGLR TRPQGEDGTRPNAQKPVWMRRLES RIATLRSKVGRLTQYKQGNRSTRLVR HVAEIVRPAELRDLREVDITEILDTHV QRLSALAKRLRRYGECSKRKEQNRMF NINEREFYNRIRNDKVDFGEGLPEIGD VTQFWANLWENPTEHNGDGMWL AEEERQCDGIGDMTAVVVTAQDIRE ATRYTRNWAAPGPDFVHNFWYKKL TTIHGRMAECFNTVLGDPTQLPEFIT RGVTFLLPKDQHTADPAKYRPITCLSS LYKVLSSVIARRVQAHCDANNVMTE EQKGCRRNTQGCKDQVIVDAVIVGQ ATQKQRNLSMAYIDYKKAYDSVPHS YLLKVLQLYKVDGNVIRLMQHAMG MWSTSLHITDGTAVLRSRTLSIRRGIF			

						QGDTFSPLWFCLAMNPLSKALNQCN YGYQLKSGERSTRVTHTFYMDDLKLF AESVQRLHQLLQLVTTFSNDIRMEFG IDKCRSIHLRRGQVMDASCFRVNEQE EIRNMVEGETYKYLGFLQLRGIRHTM IKKELQEKFLSRVNCILKSFLSAGNKVK AINTFAVPLLTYSFGVVKWTKTDLEAI ERAVRVAFTKHRMRHPKSSIERVTLP RAAGGRGVTDIQALCVSQIQQLRAYF VESQNRHEIYRTVCEADHGFSALHLA QEDYQLNCDIKTVDEMIAMWKQKE LHGTHPHQLELEHIDKVASNTWLVR GDLFSETEGFMVAIQDRVIATKNYRH YILHEDVEDRCRKCNSVGETIEHVVA GCSVLAGSAYLDRHNEVAKIVHQQL ALKHNLVDRFVPYYKYLPDPVLENSCI KLYWDREIITDVLIRANRPDIVVYDKR MKRVTLIDIAVPLDHNVQSTFSNKIA KYHDLAEELKQMWHLEDVRIVPVVL SATGIVPKSLLRSLDELELKKDLHSIQK AVILGTCSIVRRFLNHHN			

	RTE-1_Alp	GCF_000816755.2	Arachis ipaensis	ACTCCGCCTATGTTACACTTGCGGTACATAGCCGGTCCCAAGCCCGGATAAAGGAGGAGGGTTGTGTTAGGTCTTCGGCAACCAACATAAAAATATAGCCGAACCCCCATGAC	TAGTGGGA	MNQRHYCAKARSLPGSNAPYGSSTVSNEQEPLHRCPDVVLNEQGFSRFRERTRVNKLVHKVKGKGRSDRRLRFGTWNIGTLTGKSMEVVDTMTRRKINIMCLQETKWVGAKARELDTSGFKLWYTGKVKNRNGVGIIVDKQWKKDVVDVKRVGDRIISIKLVVEGGAFHVISAYAPQVGSDEQHKIRFWEDLESLVQGIPLGDKIFLGGDLNGHVGREVTGYGSIHGGHGFGVINAEGKTILDFSSTFDLLIANTCFKKRDEHLITYKSGMTSSQIDFFLLRRVDRKFCINCKIIPGESLTTQHRVLVMDFRVEQKLRKRHHTKNPRTRWWRMKGEEQRSFLRRVGEEAKWDGNGSAEEMWREMAEVIRRTAKESFGESKGIGPRDKESWWWNASIQEKIKIKRECFKEWSLCRNADNWEKYKAAKKETKVAVSEARTRAYEGLYQSLGTKEGEKGIYRIAKSRERRTRDLDQVKCIKDKDGEVLAQEEKINERWKSYFYELFNEGQKTLPSLGRLCTREEDQNFDYYRRIRDFEVKEA	cd01650	589	836

						LKQMKNGRAVGPDNIPIEVWKGLGE KGINWLTKLFNEILRSKKMPDEWRKS TLVPIYKNKGDIQSCGNYRGIKLMSH TMKLWERVIERRLRKETQVTENQFG FMPGRSTTEAIYLLRRMMERYRSNK RDLHMVFIDLEKAYDRVPREVLWKV LEKRRVRIAYIRAIKDMYDGATTSVKT QGGVTEEFPIGIGLHQGSSLSPYLFTL VLEVLTEHIQEPVPWCMLFADDIVL MGESREDLNKKLELWREALEVYGLRI SRSKTEYMECKFSLRRENPNIEVKIGE NILRKVKSFKYLGCIIQDNGEIEQDVN HRIQAGWSKWRSASGFICDKKVPLKL KGKFYRTAIRPAMLYGTECWAAKGE HEHKLSVAEMKMLRWMSGHTRLDK IRNEDIRERVGVAPIVEKMVESRLRW FGHVRRRPIEHPVRRVDEMEDGQRA KGRGRPKKTIHEVVKRDLHVNGLSVD MIHDRAQWRRLIHVADPT			
	Ingi-2_AC	GCF_00000 2075.1	Aplysia californica	CCGGAGTGGGG CGACCATCCCCG	TGATGCGATCGAGAA GAAAAAGAAGAA	MTLEEGNVRYGARPVRTHQAGAVP GPKPRRGPPQRLEGLPSSWSARRRG	cd01 650	548	804

				ACCCGGCGGGTC		RKATGNTNTKQVKKSLNIMHWNAE			
				CCAGGTGGCGG		GISNKKTELENFLKDNEIHISCIQETHL			
				ATAGGGAACGG		QEEASFKVRGYQCFRNDRQGRKKGG			
				CCACTGTTATAG		VLTLVRNNLNTVEIKKYTGEAEYIHNR			
				AGGTTAGCTGTG		ITTSRGNLDLVNYYCPNDTNLVLETIE			
				AATATATTGAAT		VPDTDFIIAGDFNSQSQSWGYNRLD			
				AAGCAGCCCCGG		RRGEEIEAWQDENHLLLVNDPLDQP			
				ACCAACTGGCGA		TFYSRAWHSTTTPDLALCTENLHRHL			
				TGCTCTTACCAG		SRKVETQLGGSDHRPVLLTIVESLPED			
				GACGGGGTGAG		SPKRPRWNYKKADWRLFSIRTNELTK			
				GTGGTAGGTGGC		NIAVFGKNINNAIREFNENILKSAKETI			
				ACTACCACGTCA		PRGVRKNYTPYWTPELQKTHNDLIKA			
				TTAAATGCCCTA		REEAETHPGRENNNKLQRCKAHYLK			
				GGAGTCCTCTGA		TKLECTRKSWRDKTHKLNLEKDTTKL			
				CGGGAGCATCAT		WRLVKGLSDEGNKNDKISLEVEGEM			
				GAGGAAAAGGA		EMEKRAAHTFVKAYAQESNIKIPRER			
				CCGCATTAAGTC		QKEVRQEERRKRSTNNIPEAMEKDIT			
				CTGGCTACGGGG		MEELKKAINQLKQRKSTGPDEITNEM			
				GGGATCCGGTAA		LQHLGNTALQTLLDIFNLSWKIGQVP			
				CAATTCCTCCCTG		QCWKEATMILILKPGKSKAQPLSYRPI			
				TGCTCCCAGGGT		SLISCVCKTMERIIACRMQWYLETQH			

				AGGTCCTTGGAC CTGGAGCTGGAG TGCTGAAATCGG AGGTATTGGGCC GCAAGGCGCACT CTAGAGAACCAC AATACCACGTAA TACTGCAATAT		IIVPEQAGFRQFRSTEDQTIYLSQVIED SFQAQKVMLATFVDLQKAFDKVWK EGLLVKLLRSGIQGNMYQWTKSYLH NRRARVLVDGYQGRKMLLRQGVPQ GGVMSPLLFILFINDLMNELPKGIHA ALYADDLVLWCSEEFVTTATYRMQL ALDKLTTWTKTWCVTANKEMSSATL FTLSQAEPKHLTLDKTPLKFENQQKYL GIIFDKRLTWKQHIQNAENKARKKLN IMRKLAGTQWRANEKTLRQVYLGAI RPHLEYGSSAWMTAAQTHKYTLEKV QNQALRIITGAMRSTPIEKMEQKTSI VALQRRWECKALVQYTKATAMEDHI LHTRTERSNTARLKRSNFIKESRKIQRS LQDHLPATVNMLNHTPEAPSWEKK NTNISIHTSIPMFTTKDDHSDLSKKSL TLSMLEERYPREAWIRIYTDGSATDA VKRGGAGVYVQFPNRRRQAEAVPT GRHCTNYRAETEALVQAANIIKNVVD PQSQVVFLTDAMSVLQAINSGKLPRL RKEIEDIPCLRIMLQWIPSHCGVLGNE			

						QADRMAKLGAQKEQPDNSVNYTEIK TIINTLFKAPQTPDGYHELSRQEQVIIF RLRTQHTRLRQLMYKCFKIVQSPKCQ CGEADQTVEHILQHCKTHRALRQET WPTPLTLEEKLYEPPDELTRTAQFMT DTGLQV			
	JAM1	GCF_00220 4515.2	Aedes aegypti	TTGGCGCGGTAA ATGGCTGGGCAT GGCGTACCATTG GTACCTCGCGTA CCTGCAGGAATA AAATAGACCCCT TTGTGCGGTCCT TAGCCTCTTGCCC AGCAACTCCTAT CCCTACCTCCTCG TGGTACTGGCCG GGGTACGAGTAA CCTTGGGGAAGA TCGGGTAACCAA CCCCCGGTGGGA	TAATTGATGTAAAGC CAAATAAGTAAGTAA GTAAGTAAG	MGDMQRRVIGWWPINERMCKLRIK GRFFNFSIINVHSPHSGSTDDDKDAF YAQLEREYDRCPSHDVKIIIGDLNAQV GQEEEFRPTIGKFSAHRLTNENGLRLI DFAASKNMAIRSTYFQHSLPYRYTW RSPQQTESQIDHVLIDGRHFSDITDV RTYRGANIDSDHYLVMVKLRPKLSVI NNVRYRRPPRYNLERLKQPDVANAY AQHLEAALPDEGELDRAPLEDCWRT VKAAINDAAESVVGYVERSSRNDWF DEECQEVLEEKNAARAAMLQHGTR QNVERYRLKRKQQTRLFRDKKRRLEE VECQEMELLYRSQETRKFYQKLNTSR KGFVPRAEMCRDKDGSILTDGREVIE RWKQHYDEHLNGAENTGTEGQDSE	cd01 650	457	716

				ACTTTGGTCGTA		GDGYVSTADSGNQPAPTMGEVKDAI			
				TGCTGACAGGGA		QQLKNNKAAGKDGIGAELIKMGPDR			
				AGGGGGGGTTT		LAACLHRLIVRIWETEQLPEEWKQGV			
				GCTTTTGCAAAC		ICPIYKKGDKLECENYRAITILNAAYKV			
				CTGGAGCGTCTG		LSQILFRRLSPIANEFVGSYQAGFIDG			
				TACTCCACGTTA		RSTTDQIFSVRQILQKCREYQVPTHHL			
				GGAGCGGCTCAC		FIDFKAAYDSIDRIELWKIMDENSFPG			
				AACAGCGTCTGT		KLTRLIRATMDGVQNCVKISGEHSSS			
				TCCCCATGTCAG		FESRRGLRQGDGLSCLLFNIALEGVM			
				GGGCGGCTGATC		RRAGLNSRGTIFTRSGQFVCFADDM			
				ATCGTCCGAGTG		DIIGRKFETVADLFTRLKREATRVGLM			
				CCAGAGAAGGA		VNASKTKYMLVGGTERDRTRLGSSV			
				CTCTAAGCTAAA		TIDGDTFEVVDEFVYLGSLLTADNNV			
				CTGCGCACTATG		SREIRRRIISGSRAYYGLQKKLRSRKIH			
				GCCCTCCGAACA		PRTKCTMYKTLIRPVVLYGHEAWTM			
				TTTAGGGGGAAT		LEEDLQALGVFERRVLRTIFGGVQEN			
				GGTCCTCCGGAA		GVWRRRMNHELAQLYGEPSIVKVAK			
				ATCTAGGGGGTT		AGRIRWAGHVARMPDNNPVKMVF			
				GGTGTCAGGCCC		ATNPVGTRRRGAQRARWIDQVHQD			
				TGCAAGCCAACC		LESVGHSRGWREAAMNRGNWRNI			
				GTAAAAACACAT		VGEALSR			

				CAGCACAGGAAC GTCAACGAGAGA ATACGGACCGGA ACAATCGGCAAA GACCACAGCGAC GAAAATGGACTA GCGATTGGAAAC TCGGTACGTGGA ACTGCAAATCTC TCAACTTCATTG GAAGTACTCGCA TACTCTCCGATGT ACTGAAGACCCG CGGTTTCGACAT CGTAGCGCTGCA GGAGGTGTGCTG GACAGGAGCATT GGTGCGAACGTT TAGAGGTAATCA TACCATCTACCA GAGCTGCGGCAA					

				CACACGCGAGCT GGGAACAGCTTT CATAGTG					
	RTE_Ele2	GCF_00220 4515.2	Aedes aegypti	TGGGTTGTAAAA TGGTGAGTCGAC AACCAGGAAGG AGCGTCCAACAT AGCTCTGGTCCT CGCAAGTCCCTA CCTCACGCTTCCT CGGGTCTAACGA TGACAAAGACCG CCAGCTAAGGGT TGCGTACTTAGC TGGTAGTGCAAC CTGGGCACTGTT GTCCTTCTGACAT CAGCTAGAGTGA GGGGGTGCGTCC CGAGCGTCTGTT CACCAGGAGGTG	TAGATTCAACGCAAT AGCAAGGAATTGTTG CCCATCAAGTATCAA GTAAGTAAGT	MGSIPNFSPTTKNGERRNNERYFRQ PTWQRNKDYEWKLGTWNVRTLNEP GRASLLARELQKVGVGVAAIQEVRW PRSGEREFRAVDPIANTTFKYNIYHSG GDKAEHGVGFIVIGKQMKRVMKWK PISERICVLRIRGKFFNYSLINIYAPTND KPDDVKDAFYEGLDKAYGECPKHDV KIVIGDANAQVGREDFFRPIIGKESLH SVTNDNGLRLVTFAAARGMAICSTYF ARKDIRKHTWRHPNGDICNQIDHILV DGRHFSDVIDVRSFRGPNIDSDHYLV VSKIRARLSTVANSRPQRTMRFNIQR FSADGVTAEYHQKLDERISEVNVSEN LNNLWDAIHGVVSETAREVIGTARRR PRSGWFDEECQRVTDEKNVARRRM LVSGTRLNRERYREARSAEKRIHRRKK KEFDENVIAEAQNCMEQNDMRRFY ETVNGVRRKSAPSPVMCNDREGNLL	cd01 650	561	821

				CGGCTCAAACAG CGTCTGTTCTGG TATCCAGCGGCT GAGTATGAAATG TTGTATCACGTC AACTAACCTAAG GTGGCAGCCCCA TCAACGTGATGT AGGTAGCGCAAC CCCGGTAA		TDKTMVAARWKEHFETLLNGGNDR ASENRMNIDNDGQAVEPPSLDEVKK AIKELKNNKAAGKDELPAELLKHGSE QLHRIIHHILLRIWEEEELPASWMDG LICPLFKKGHRLECANYRGITLLNSAYK IMSRILFNRLRPLEESFVGEYQAGFRE GRSTSDQMFTLRQILDKFREYNLQTH HLFIDFKAAYDSVKRNELWEIMMEH GFPTKLIRLIRATLEGSKSSIRVADEIST SFVTLDGLKQGDALSNLLFNIALEGAI RRACVHRNGTIITRSYMLLGFADDIDI IGIDRRAVEEAYVPFKRETTRIGLIINTT KTKYMVAGRHRGPIGDVGSEVVIGG DIFEVVEEFVYLGTLVTCDNDVTREVK RRIAAASRAFYGLRNQLKSRSLQTKTK LALYKTLILPVALYGHESWTLKEVDRR AFGVFERKVLRTILGGKLDNGIWRRR MNHELYQVYNEMDIVKRIKHGRLR WAGHVARMSEERQAKQIFSREPGR GRRLRGRPRTRWLFAVEEDLRSLNV QGDWKRLAQDRVQWRRLLHSA			

	RTE_Ele2B_AAe	GCF_002204515.2	Aedes aegypti	TGGGCTGTGAAATGGTGAGCCGACAACTAGGAAGGAGCGTCCAACATAGCTCTGGTCCTCACAAGTCCCTACCTCACGCTTCCACGGGTCAAATGATGACAAAGACCGCCAGCTAAGGGTTGCGTACTTAGCTGGTAGTGCAGCCTGGGCACTGTTGTCCTTCTGACATCAGCTAGAGTGAGGAGGTGCGTCTCGAGCGTCTGTCCACCAAGGAGGTGCGGCTCAAACAGCGTCTGTTCTGGCATCCAGCGGC	TAGATTCAACGTAGCGGATTGTAGCCCATCAAGTATCAAGTAAGTA	MGSISNFLPTTKNGERRNNERYFRQPTWQRNKDYDWKLGTWNVRTLNEPGRVSLLARELQKVGVSVAAIQEVRWPRTGEREFRAVDPVANTAFKYNIYHSGSDKAEHGVGFIVIGKQMKRVMRWKPINERICVLRIRGKFFNYSLINIYAPTNDKPDDTKDAFYDCLDKTYGECPKHDVKVVIGDANAQVGREDFFRPIIGTESLHSATNDNGLRLITFAAARGMAISSTYFARKDIRKHTWMHPNGELCNQIDHVLVDGRHFSDVIDVRTFRGPNIDSDHYLVVSKIRARLSTVANTRTQRPLRFNIQRLSTEGVADEYRRKLDERIRGYNVGENLNNLWESIHGAMSTAAREVIGTAQRRPRNGWFDEECQNVTDEKNVARSRMLVSGTRQNRERYKEARAAEKRIHRRKKRQYEEDIITEAQNSIEQNDMRRFYETVNGVRRKTAPSPVMCNDREGNLLTDKILVAARWKEHFEQLLNGENTSASENRLIISNDGQAVEPPTLDEVRRAIKELKNCKAAGKDELPAELFKHGSEQL	cd01650	561	821

				TGAGTATGAAAT GCTGTATCACGT CAGCTACACCTA AGGTGGCAGCCC CACCATCGTGAT GTAGGTATCGCG ACCCTGGTAA		YEILHRTLLRIWEEEELPASWLEGLICP LYKKGHRLECANYRGITLLNSAYKIMS RVLFNRLRPLEESFVGEYQAGFREGR STTDQMFTLRQILDKFREYNLQTHHL FIDFKAAYDSVKRNELWQIMVEHGF PAKLIRLIRATLDGSKSSVRVADETSTS FVTLDGLKQGDALSNLLFNIALEGAIR RAGVQRSGTIVTRSHMLLGFADDIDII GIDRRAVEEAFVPFQRETARIGLTVN TSKTKYMIAGRQRGSNRDVGSEVVL GGEKFEVVEEFVYLGTLVTCDNDVTR EVKRRLAAANRAFYGLRNQLKSRSLQ TKTKFALYKTLILPVALYGHESWTLKE VDRKAFGVFERKVLRTILGGKQENGI WRRRMNHELYQVYKEVDIIKLIKHGR LRWAGHVVRMPEERQAKIIFSREPG RGRRLRGRPRTRWLFAVEEDLRALN VQGDWKRLAQDRAQWRRIIHSA			
	RTE_Ele5	GCF_00220 4515.2	Aedes aegypti	ACCGCTAGCGCA TGACGGCTCACC ATAGTAGCATGT	TGATGCTACGCTGTA GCCATCAAGGTATCA AGGTAGGTA	MHDVGSATLVRWPIEDSSPTKKGKS RANGLILRQQTRQRIKDNDWKVGS WNVRTLNEPARVGLLARELRNVGVN	cd01 650	570	830

				CCGAAAGAACTT GAAACTATTGAG AACCAGAGCTAC AGGTCCAAAGCC CTGATAAAGGTG GATGGTTGTGAT GGCTATGACCGT GGTCATCATGTA AAGAACAAGCG GACGATGCTGTA TCGTGTCAGCAT CGAGGAGGCAG CCCCTCTAG		VAAIQEIRWPRTGEREFRAVDPIANT SFKYHIYYSGGDRAERGVGFIVIGKQ MKRIIRWKPVSDRICVLRMKGKFFNY SLISIYAPTNDKPDDMKDEFYESLDKA YGECPKQDVKIVIGDANAQIGKEDFF RPVIGTESLHSVTNDNGLRLVTFAAA RGMAISSTYFARKNIRKHTWRHPSG DACSQIDHVLVDGRHFSDVIDVRTFR GPNIDSDHYLVVAKIRARLSSVTSSAT HRTLRFNIQRLSTDGVAAQYRQQLD EKLGRVNVAGDVDSLWDPIHEAVTT TAREVIGTGQRRRRNDWFDEECQRV TDMKNVARSRMLVAGTRQNRERYR AARAEEKRIHRRKKRQHEESVVADA QESMNRNDMRRFYATVNGARHNT VPVPAMCNDREGNLLTDKTAVAAR WKEHFQQLLNGENGNVARNRMNI DDDDQAVDPPTIGEVKKAISELKNCK AAGKDEIPVELLKHGSEQLYQSIHRVL LKVWEDEELPTGWLDGLIRPIYKKGH RLECANYRGITLLNSAYKILSRILFNRL			

						RPLEESFVGEYQAGFREGRSTTDQM FSLRMILDKFREYNLQTHHLFIDFKAA YDSVKRNELWQIMSENGFPAKLIRLI RATLDGSKSSVRIADEVSTSFVTLDGL KQGDALSNLLFNIALEGAIRRSGVQR NGTIITRSHMLLGFADDIDLIGIDRRS VEEAFVPLKRETARIGLTINSTKTKYM VAGRDRVRHGGVSAEVVFDGDVFE VVEEFVYLGTLVTCDNDVSREVKRRV AAANRAFYGLRNQLRSRSLQTETKFA LYKTLILPVALYGHEAWTLKESDRKAL GVFERKVLRTILGGKLENGVWRRRM NHELYQVYKDANIINRVKYGRLQWA GHLVRMSEERIAKIIFSREPGRGRRLR GRPRIRWLYAVEEDLATLNVRGNWR SFAQDRRRWSSTIRPAMA			
	RTE_Ele2 C_AAe	GCF_00220 4515.2	Aedes aegypti	GGGCTGCAAAAT GGTGAGTCGACA ATCAGGAAGGA GCGTCCAACACA GCTCTGGTCCTC	TAGCCCATCAAGTAT CAAGTAAGTA	MGSIPNSLSTTKNGERRRNERYFRQP TWQRNKDYDWKLGTWNVRTLNEP GRVSLLARELRKVGVCVAAIQEVRWL RSGEREFRAVDPIANTAFKYNIYHSG GDKAEHGVGFIVIGKQMKRVIRWKP	cd01 650	561	821

				ACAAGTTCCTAC		INERICVLRIRGKFFNYSLINVYAPTND			
				CTCGCGCTTCCA		KPDDVKDAFYECLDKTYGECPKHDVK			
				CGGGTCAAATGA		IVIGDANAQVGKEDFFRPIIGKESLHS			
				TGACAAAGACCG		VTNDNGLRLVNFAATRGMAISSTYFA			
				CCAGCTAAGGGT		RKDIRKHTWQHPNGELCNQIDHVLV			
				TGCGTACTTAGC		DGRHFSDVIDVRTFRGPNIDSDHYLV			
				TGGTAGTGCAGC		VSKIRARLSTVSSSRNRQSLRFNIQRLS			
				CTGGGCACTGTT		ADGVAADYHQKLDERISEIDESVNLS			
				GTCCTTCTGACAT		DLWESVHGAVSTVAREVVGTAQRRP			
				CAGCTAGAGTGA		RNGWFDEECQRMTNEKNLARSRML			
				GGAGGTGCGTTC		VSGTRQSRERYKEARAAEKRIHRRKK			
				TGAGCGTCTGTA		KEHEEAVIAQAQEAMEQNNMRRFY			
				CACCAGGAGGTG		ESVNGVRRKTAPSPAMCNDREGNLL			
				CGGCTCAAACAG		TDKTMVAARWKEYFESLLNGDNGS			
				CGTCTGTTCTGG		GSGSRIQIDDDGQAVEPPTLDEVKKA			
				TATCCAGCGGCT		INGLKNNKAAGKDELPAELLKHGSEQ			
				GAGTATGAAACG		LHELLHRIISRIWEEEQMPTSWLEGLI			
				CTGTATCACGTC		SPLYKKGHRLECANYRGITLLNSAYKI			
				AGCTACACCTAA		MSGILFNRLRPYEESFVGEYQAGFRE			
				GATGGCAGTCCC		GRSTTDQMFTLRQILDKFREYNLQTH			
				ATCAACGTGATG		HLFVDFKAAYDSVKRNELWQIMSEH			

				TAGGTAGCGCGA CCCCGGTAA		GFPAKLIRLIRATLDGSKSSVRVVDEIS SSFVTLDGLKQGDALSNLLFNIALEGA IRRAGVQRSGTIITRSHMLLGFADDID IIGIDRRAVEEAFVPFKRETARIGLTIN TTKTKYMIAGGQRGSGRVSGSEMVL GGDKFEVVEEFVYLGTLVTCDNDVTR EVKRRIAAASRAFYGLRNQLKSRSLQT KTKLALYKTLILPVVLYGHESWTLKEV DRRAFGVFERKVLRTILGGKLENGIW RRRMNHELYQVYKEVDIVKRIKHGRL RWAGHVARMPEERQAKIIFNREPGR GRRLRGRPRTRWLFAVEEDLRALNV QGDWKRLAQDRAQWRRLIHSAQIH RSEL			
	JAM1C_A Ae	GCF_00220 4515.2	Aedes aegypti	GGGTGGTAAATG ACTGGACAATGC GTACCATTGGTA CTTCGCGTACCT GCAGGTATAAAA TAGACCCCATTT GTGGTCCTTAGC	TGAAGGACGTAGAG CCAGCAAAAGTAAAG TAAGTA	MKSSVQEAYNVNSDRNNRHRPRHR KRTNDWKLGSWNCKSLNFLGSTRILS ELLRVRKFDIVALQEVCWKGSTVHTY RDGYTIYQSCGNRHELGTAFIVMGE MQRRVIGWWPIDNRMCKLRIKGRF FNISIINVHSPHLASDDDDKDAFYAQL EREYDGCPSHDVKIVIGDLNAQVGQE	cd01 650	550	810

				CTCTTGTCCAGTA		EEFRPIIGKFSAHQLTNENGLRLIDFA			
				ACTCCTATCCCTA		ASKNMAIRSTYFQHSLPYRYTWRSP			
				CCTCCCCGTGGT		QQTESQIDHVLIDGRHFSDIIDVRTYR			
				GCCGCCTGGGAT		GANIDSDHYLVTVKVRQRLSVVNNIR			
				ACGAGTAACCGT		YRRPPRYNLERLKQPEVATEYAQSLE			
				AGGGAAGATCG		AALPEEGELTEAPLEDCWSSLKAAIN			
				GGTAACCAACCC		NAAEGAIGFVEANRRNGWFDEECQ			
				TGGTGGAACCTT		TVLDEKNAARAMMLQQGTRQNVER			
				GGTCGTATGCTG		YKQKRRQQTHLFRDKKRRLEELECEE			
				ACAGGGAAGGG		MEQLYRSQETRKFYKKLNASRKGFVP			
				GGGCTCCTCTCTT		RAEMCRDKDGGILTDEREVIERWKQ			
				CTGAGGGTGTAG		HYDEHLNGAEEEDQDSRRNGFISTA			
				CTTATCAGAGCG		DEGDVPTPTIGEVKDAIKQLKNNKAA			
				TCTGTTCTCCATG		GKDGIGAELIKMGPDRLATCLHRLIA			
				TTAGGGGCGGCT		RIWDTEQLPEEWKEGIIYPIYKKGDKL			
				CAAAACAGCGTC		ECENYRAITILNAAYKVLSQIIFRRLSPL			
				TGTTCTCCATGTT		ASRFVGSYQAGFVDGRSTTDQIFML			
				AGGAGCGGCTG		RQILQKCREYQVPTHHLFIDFKAAYD			
				ATCATCGTCCTA		TIDREELWKIMDENGFPGKLTRLIKAT			
				GTGCCAGCGTGG		MDSVQCCVKISGALSDPFETRKGLRQ			
				GACTCTAAACAG		GDGLSCLLFNIALEGVMKRAGFNMR			

				TGCTGTGCACGA TGATCCTCCGGC GAGACAGGGGG TTGGTGCAGGCC TTACAAGCCAGC CGTAA		GTIFNKSSQFICFADDVDIVGRTFQVV AEQYTRLKREADRVGLKVNTSKTKYL LAGGTERDRARIGRRVTIDGDEFEVV DEFVYLGSLITSDNNCSREIRRRIIAGS RAYYGLHKTLRSGKLHFRTKCTMYKT LIRPVVLYGHETWTMLEEDLQALGVF ERRVLRTIFGGVCENGVWRRRMNH ELAQLYGEPSITKVAKAGRVRWAGH VVRMPDNNPAKMVFNSNPAGTRRR GAQRARWFDQVEQDLGSVGRSRN WRLAAMDRVSWRNIVAQVMS			
	JAM1B_AAe	GCF_002204515.2	Aedes aegypti	GGTAAATGGCTG GGCATGGCGTAC CATTGGTACCTC GCGTACCTGAAG GAATAAAATAGA CCCCTTTGTGCG GTCCTTAGCCTCT TGCCCAGCAACT CCTATCCCTACCT CCTCGCGGTACT	TAGATGTTAACTAAA TAAATGAAT	MKNQATNNQRENTNRDNRRRPQR RKGTSDWKLGTWNCKSLNFIGSTRIL ADVLKDRGFGIVALQEVCWKGSMV RTFRGNHTIYQSCGNTHELGTAFIVM GDMQRRVIGWWPINERMCKLRIKG RFFNFSIINVHSPHSGSTDDDKDAFYA QLEREYDSCPSHDVKIIIGDLNAQVG QEEEFRPTIGKFSAHRLTNENGLRLID FAASKNMAIRSTYFQHSLPYRYTWRS PQQTESQIDHVLIDGRHFSDIIDVRTY	cd01650	553	795

				GGCCGGGGTAC		RGANIDSDHYLVMVKLRPKLSVVNN			
				GAGTAACCTTAG		VRYRRPPRYDLERLKQPDVAAAYAQ			
				GGAAGATCGGG		QLEAALPEEGELDEAPLEDCWRTVKA			
				TAACCAACCCCC		AINDAAESNVGYVGRSRRNDWFDEE			
				GGTGGGAACTTT		CQEVLKEKNAARAVMLQQGTRQNV			
				GGTCGTATGCTG		ERYKRKRQQQTRLFREKKRRLEETEC			
				ACAGGGAAGGG		EEMEQLCRSQETRRFYQKLNASRNG			
				GGGGTTTGCTTT		FVPRAEMCRDKDGSILTDEREVIERW			
				TGCTTTTGCTTCT		KQHFDEHLNGAESTGNEGRDNGGN			
				GCAAACCTTGAG		AFVSTAEDGNQPAPTLREVKDAIHQL			
				CGTCTGTACTCC		KNNKAAGKDGIGAELIKMGPERLAIC			
				ATGTTAGGAGCG		LHRLIGTIWETEQLPEEWKEGVICPIY			
				GCTCACAACAGC		KKGDKLDCENFRAITILNAAYKVLSQII			
				GTCTGTTCCCCAT		FRRLSPVVNEFVGSYQAGFVDGRSTT			
				GTCAGGGGCGG		DQIFTVRQILQKCREYQVPTPLGIMD			
				CTGATCATCGTC		ENSFPGKLTRLIRATMEGVQNCVKVS			
				CGAGTGCCAGAG		GEHSSSFGSHRGLRQGDGLSCLLFNI			
				AAGGACTCTAAG		ALEGVMRRAGLNSRGTIFTRSSQFVC			
				CTAAACTGCGCA		FADDMDIVGRTFEKVADLYTRLKREA			
				CTATGGTCCTCC		AKVGLVVNAAKTKYMLAGGAERDR			
				GAACATTTAGGG		ARLGSSVTIDGDTFEVVDEFVYLGSLL			

				GGAATGGTCCTC CGGAAATCTAGG GGGTTGGTGTCA GGCCCTGCAAGC CAGCCGTAA		TADNNVSREIRRRIISGSRAYYGLQKK LRSKKIHARTKCTMYKTLIRPVVLYGH ETWTMLEEDLQALGVFERRVLRTIFG GVQENGVWRRRMNHELAQLYGEPS IQKVAKAGRIRWAGHVARMPDSNP AKMVFASDPVGTRRRGAQRARWA DQVRIDLASVGQNRGWRDAATNRV LWREIVDSVLSV			
	Lian-Aa1	GCF_00220 4515.2	Aedes aegypti	TTAGTGGCTGCT TGGGCACGTCAG GAGTTAATCATC AATATTAACCTCC TTTTCCCGAGCA GCCTAGATAGCC GCGTAGTGTCGG TAGCGGTTGTTT CAACTGGCTAAG AATTAACACTAC GGACTGCCTGTT CCGGTGGTAAAA GTCCACCTCACA	TAGGCTCTCTTTACG CTCATGCGTTTTGCA GTGCCCTTTTTAGGG CGCTGTTCGAACCCA TTGTGGTATGGAGCT ACATGCTCTCATTTCG CTTATGCGATCTTCCC TCTTCAAGGGACCCC ACTCCTATTTCCTCCC ATCTTTCCCTTCCCTT TCCTCTCCCATCGGG TAGATGATGAAATAG GCTCAAATATGGCGA	MQNGAAQVALVQEPYFRRGNFYLG NLVDPVFATFSKLEMANSRSMPRAC VLVNKAIVATLISELTTRDVCAVTIDVS VGDLNRKYVYCSVYLPHDEPSPTDDF KRVVVHCVTKGLPLIVGSDANAHHII WGSSDINLRGSSLMEYLSSTDLGLLNI GNRPTFMVSNREEVLDITLCSNRISH ELTNWHVSDEESLSDHRYIFFEHSNV TAQTLRFRNPRSTNWELYIELVATKF HGYSPSIENPSDLDDAVDTTTSYIME AFEEACPLRSVKTTRGTPWWNSDLT RLRKQCRRSWNRRRSAGSESFKSAR KAYKKALRSAERSGWKNLCTNVSSLS	cd01 650	468	725

				GGTGACCCCTAA	TGGCACAAATCTCCC	EVSRLNKILAKSKDFQVNEIRLPNGDF			
				TTTATGGTGTGA	AACTGGTGGGGAAC	TSSDEEVLECLFNTHFPGCVDIASTDE			
				TGCGTACCGTGC	GTGCCTTTGGAGCCG	PNVFSCSYESLASARSIVTTESIQWAL			
				CTAAGAATGAAT	GCCTTCTGATACCTG	NSFAPFKSPGADGIYPVLLQKGFEFIK			
				GGTTAGGGGGG	ATACCTGA	HVLKKLLVSSFATGYIPKSWRDITVKFI			
				TCTAAATAAAAC		PKGGRASYEEAKSFRPISLTSFLLKCLE			
				CTAACCGCAAAC		RIIDHHIRDVYLANMPLHVNQHAYQ			
				GGAGCCTGTGG		SGKSTVTLLHKVVYDIEKAFAQKQSCL			
				GGTACCAGGGC		GVFLDIEGAFDNVSFDAILEAARNHG			
				GCCCTCCACAGT		LPTMITNWIHQMLKNRHLFSTLRQA			
				ATTGAGTCCTTCC		AIRKLSVCGCPQGGVLSPLLWNLVAD			
				TGTGCTACCCGG		TLLRQLNNCGFPTYGFADDYLALIVG			
				AGCAATGGTGCA		MCISTLFDLMQSALQVVESWCRQYG			
				GGTGACCTTGTG		LSVNPNKTSIVLFTERRNRDGIRPLRLF			
				TTTCTCCGAGAT		GTEINVTDQVKYVGVILDSKLSWTPH			
				AATCGGCTGCCC		IDFRVKKACMAFGQCRRTFGKTWGL			
				TTCTTCAGTCTCT		KPKYIKWIYTTVVRPILAYGCLVWWQ			
				ATCTTGAGGCTT		KGEVRTIQSKLGHLQRMCLMAMSG			
				AATAAGGGTGG		AFSTTPTAALEALFDVAPLHIYLKQEA			
				GATTATGAATAT		LSCSYRLWVLDLLEKNPVNRRSTHTSL			
				GTTGACATTTAA		FPLLVNWDKIVLAPSDLTIACNFPYRT			

				TTTAAATTATCAC CTATATGGATTC GCATTATGCGTT TTACACAGTGTA TTCTGTGCTCTTT CGCCTTTGGCGA CTTTAAATAGAT ACCGATCTGGTT TTTTCGTTGCTGG TCTTTTTCGTGCT GAGATTGCAAAA AGCTTAATCCTG CCTAGTTTGGGT AGTGGCTACGGT TAGGTTAGCTCA GATCAATCTTCA GCATAAAAGAAC AGCAACGATCAA TCTTTGCAGACTC		FTTQFPSREEWTSGYLERSISNNIVCY TDGSLLEGRAGAGVYSRELRLNQFYS LGRNCTVFQAEIFALMCGVQSALQQ RVMGKVIYFCSDSQAAIKALASANSR SKLVIACRTQIEELNSVNSVNLVWVP GHSSIAGNELADELARDGASHDFIGP EPAIPISKCWVKLQINSWAATQHKQY WNSLESCRQTKLYITEPSPKVAKYLTN LSKQNCSLLVRALTGHCRLNYHMANI QRADSFVCDSCDSDYGTSYHLICNCP VFAQMRFQLLGKHLLSETEYRSLNLQ DILLFLTRCGNEL			
	RTE_Ele4	GCF_00220 4515.2	Aedes aegypti	GGGCTGCAAAAC GGTGAGTCGACA	TGACTGGTAAGGTAA GTA	MPHQRDRPSVSWDVKQSWHDGAP ARQECWRRPNKPPVKNLIANNIGDN	cd01 650	588	848

				ACTAGGAAGGA		TTRNNRQRPTRRNKDYDWKLGTWN			
				GCGTTCAACACA		CKSLGFAGCDRIIYDELHPRNFDIVAL			
				GCTCTGGTCCTC		QELCWTGQKVWKSGHRAATFYQSC			
				ACAAGTTCCTAC		GTTNELGTGFIVLGKMRQRVIGWQP			
				CTCACGCTTCCAC		INARMCKMRIKGRFFNYSIINVHCPH			
				GGGTCAAACGAT		EGRPDDEKETFYAQLEQIYDGCSPRD			
				GACAAAGACCGC		VKIVVGDMNAQVGREEMYRPVIGR			
				CAGCTAAGGGTT		NSLHAVSNDNGQRCVNFAASRGMV			
				GCGTACTTAGCT		VRSTFFPRKDIHKTTWRSPDNLTENQ			
				GGTAGTGCAGCC		IDHVLIDGKFFSDITNVRTYRSANIDS			
				TGGGCACTGTTG		DHYLVAVCMRSKLSTVTNSRRSRTPR			
				TCCTTCTGACATC		LNIERLRDSEVAQEYAQQLEVALPTEE			
				AGCTAGATTGAG		QLGAATLEDGWRHIRSAIGSTATAAL			
				GAGGTACGTCTC		GPAAPNQRNDWYDGECEQLKNEKN			
				GAGCGTCTGTTC		AAWARMLQHRTRANEARYKQARN			
				ACCAGGAGGTGC		RQNSIFRRKKRHQEERDREAMEELFR			
				GGCTCAAACAGC		AKDTRKFYEKLNRSRKGFVPQADMC			
				GTCTGTTCTGGT		RDCDGNLLTNEREVVERWRQHYDE			
				ATCCAGCGGCTG		HLNGDVASSGGGTEITLGARADDERL			
				AGCAAGAAATGC		PPPDLQEVETEIRRLKNNKAAGVDQL			
				TGAATCGCGCAC		PSELLKYGGDALARALHWVISKIWEE			

				AGCTAAATCCAA GGTGGTAG		EVLPEEWMEGVVCPIYKKGDKLDCA NYRAITILSAAYKVLSQILCRRLSPIARE FVGQYQAGFMGERATTDQIFAIRQV LQKCREYNVPTHHLFIDFKSAYDTIDR EQLWQIMHEYGFPDKLIRLIKATMER VMCVVRVSGTLSSPFESRRGLRQGD GLSCLLFNIALEGVIRRAGIDTSGTIFR KSVQLLGFADDIDIVARNFETMADTY IRLRAEARRIGLNINVSKTKYMIARGS REDTARPPPRVHIDGDEIEVVEEFVYL GSLVTADNDTSREIQRRIVAGNRAYF GLRRTLRSSKVRHRTKLTIYKTLIRPVL LYGHETWTLRAEDQRALGVFERKVL RTIYGGVQMEDGTWRRRMNHELH QLLGEPTIVHLAKIGRLRWAGHVTR MSDTNPVKMVLENNPTGTRRRGAQ RARWVDQVEDDLRTLRRMRGWRR AAMDRVEWRRLLRTAEATQALA			
	CR1- 10_AMi	GCF_00028 1125.3	Alligator mississippie nsis	CT	TGACCCCCAGTTGTC TTTCCTGCCTAGTGC AGGGGGCTGGACCC	MGQGFKLRLLGDGETLIGACPGTNT QESPVDRDTRRATLAGAEGIVLPQH DRKSRVSVGGLRCLYTNARSMGNKQ	cd01 650	539	790

					GATGATCTTCCAAGG TCCCTTCCAGCCTTAC AATCTATGAATCTAT GAAT	EELALLLSSGHYDLVGITETWWDSSH DWSVAIEGYTLHRRDRVGKKGGGVA LYVKEQYTSLEADIGSKEGQLETLWV KLWGGGGERDLTVGVYYRPPNQGE ELDLEFSRELTEAAHARDMVVMGDF NYPDISWEEHSAKSDQSRSFLTCIEEL FLTQEVHGPTRGNALLDLVLAKGDDL VTDLNIEGSLGDSDHETIMFTVHHKA DKSDSRIEVLDFKNANFNKLRSLVGE VLWDQGPKVKGVHEEWLFLKDTILE AQRKSIPMQRKGSRRAGKPSWLNRE IMVLLRKKKEAYKQWKLGTVPKEDYT TMAHTCREQIRKAKAATELRLATGIK DNKKSFFRYVGNRRKTNGSVGPLLN NSGQLVTDTQEKAELLNDYFASVFH QTKGKIMPGRVQDEHGRNDHLPTIA VELVYDQLEKLDIYKSAGPDELHPRVL KELAEVIAEPLAKLFQNSWCSGEVPE DWKRANVVPIFKKGRREDPGNYRPIS LTSIPGKILEKFIKESICDKLAEGKILND SQHGFIAGRSCLTNLISFYDQVTHHL			

						DKREEVDIVYLDFQKAFDLVSHDVLM RKLEDCGLNCETVRWVGNWLQGRT QRVLVNGSVSSWREVASGVPQGSVL GPVLFNIFINDLDVGVKSSLAKFADDT KLWGSVVTLEDRLQIQADLDRLASW ADQNQMKFNVEKCKVLHLGTNNPQ HTYRLGGTKLTSTMNERDLGVIIDHS MNMSHQCDAVASRANNTLACINRC ISSKTKEVILPLYSALVRPQLEYCIQFW APHFNKDVEKLERVQRRATHMIRDL QGKPYEERLRDLGLFSLKKRRLRGDLV AAYHYIRGVHQGLGEQLFTRAPLGKT RSNGHKLLEDRFRLNIRKNFFTVRVSR LWNKLPPEVVQSPTLEIFKRRLDSHLA GVT			
	CR1- 2_ASi	GCF_00045 5745.1	Alligator sinensis	CCTGACCGTATT ATGAGGGACTTC GAGGCTCTGGG GACCAGGCTGAA GGAGACGGGGG CGCAGGTGGTTT	TGAGTCCCGTTTTTGT TTTTAGGGTGAGTGG TTTTGTTTGTGTTTTT GGGTTGTTCTCCCGC CCTGGCGGGGGGTC GGGCTAGAAGATCTA	MGNKQEELALRIASSNPDVVGLTET WWDSSHDWAVNIRGYRLYRRDREG RKGGGVALYVKEHYTSSASSMGSGE GRAEVLWVKIQGGQGERDLTVGVYY RPPNQEEELDREFSGQLAETLKARDV VVMGDFNYPDICWEEQSAGTDRSR	cd01 650	470	720

				TTTCCTCTCTCCT	TTAGGTCCCTTCCGA	RFLAELQDLHLIQEVHSPTRGDALLDL			
				TCCGGTGAGGG	CCCTGAAACTATGAA	VLAAGDDLVRGLQVLDHLGNSDHRV			
				GGCGGGGACAG	ACTATGAAA	LEFTIQRRVSRARSKVVALDFRRANFS			
				AGGCATGGGACC		EMRRLVGEELGSLRAEELSVQEEWSF			
				TGCATTGGTGAG		LKETILGAQGVSIPTRSKAGRSVQKPP			
				GTCAACCAGCGC		WLTKGVRECLLAKRVAYARWKGGVI			
				TTACGCAGCTGG		SKEEYTSSARACRRAVRKAKAEIELGL			
				TGCCGCCGGGCA		ASRIKDNKKSFFKYIGRMKKAPGNVG			
				GGCCTTGGGTTC		PLQDMRGNLVVASEDKADLFNQFFT			
				CTTGACAACGAC		SVFLCRDRDFPAVIEAGREGIAFRPQV			
				CCATGCTTTCGG		GEDRVRELLEGLDAFKSAGPDALHPR			
				GAGGGGGGATT		VLRELAGVIAGPLARLYERSWCLGQV			
				GCTGGGGCGGG		PDDWKVANVVPIFKKGKREDPGNYR			
				ATGGGCTTCATC		PVSLTSILGKLFEKIIKEHICDGPASEM			
				TCTCCCCTAATG		MLRGNQHGFIRGRSCQTNLIAFYDQ			
				GCAAGCGTATCT		VTKAMDAGVAVDVVFLDFSKAFDTV			
				TCTCTTCCAGGTT		SHPILINKLGDCGIDAYTVRWIADWL			
				GGCTGATCTCGT		KGRTQRVVVDGSYSTWGEVGSGVP			
				ACGGCGAGCTTT		QGSVLGPALFNFFISDLDVGVESNLFK			
				AAACTAGACTCG		FADDTKIWGEVSTLAGRERLQKDLDR			
				CCGGGGGAGGG		LQGWANRNRMCFNTDKCRVLHLGS			

				GGCTGCAGGAA GACGAGGAGAC GCCATCAGGGCA AGGGGTGCAAA GGTAAGTAATGG GGGTCTTGCCAT CAGGGCAAGGG GTGCAAAGGTAA GTAATGGGGGTC TTGCCATCAGGG CAAGGGGTGCA AAGGTAAGTAAT GGGGGTCTTGCC ATCAGGGCAAG GGGCTGCAGGA AGATGGGGGGT CGCCATCAGGGC AAGGGGCTGCA AAGGTAAGTAAT GGGGGTCTTGCC ATCAGGGCAAG		RNQQHTYTMGSSLLESTEAERDLGVI IDSKMNMGRQCEVMVGRANRTLSC IHRCISSRAREVILPLYATLVRPQLEYC VQFCAPHFRRDVDNMERVQRRATR MIRGQQGRPYSERLRDLNLFSLHKRR LRGDLVTVYKLTRGNQKGLGETLFPL APAGITRNNGHKLMERRYRLDIRKNY FTVRAARSWNQLPREVVLAPTLGIFK KRLDTYLAGLI			

				GGGTGCAAAGG TAAGTAACAGGG GTCCTGATGGTC TGGCTAATCAGG GGGTGGGGGGG CACACAGTCTTG CAAGGCCTTAAA TGCCTGTACACC AATGCTCGTAGC					
	L2-18_ACar	GCF_000090745.1	Anolis carolinensis	GAGAGGCCTGCT GTTGTGCTGCAG AAGAGGCCAGA AACATCTCCAGC CAGCCTGCCTGT TGCTGCCTCTGG GACTCCTTGTGA GTAGACTGCAGC CCCAGGCCCTTT CCATCTGGGTGT GACTCAGCACTT CCTAGGAAGGAC	TAAGAATATATGACC TCAACTGACAGACCC AGTCACATGAACATG GAAAGCGCCTTAAAT GTATATTTAAATGTA TAATTATGTATATTTT AATGGCTATTCTTAA TTTTATTGTAATTTTT AATCTTGATGGATTT TTAAATTTGATGAAA ACTGTTTTTAATGTGT ATGGTTATACTGTAA	MLLMHPLHFTTAFALCIVGGGWGRL EGYGDCRVGEGNGCERMPIYNRKLV NQITSFGPGEEKGKGQCDRGGGCSD GWGAPIEVVSGRGRYGKGRPNHLIR PRVFGRTLIVPKRSPDIANLGSQDGG PSRLKVVLFNARSVNGKTSFIQDLILD EHVDLACITETWLDEMGGVNLTQLC PPGYSVQHQPRSGGRGGGVAVVYK DSIPLTRCPVPQSTMFECVHLRVGDR DRLGILLVYRPPRCTTVSLPELAEVVS GLALESRRLIVLGDFNVHAETNLSGA AQDFMAAMATMGLSQLVTGPTHSA	cd01650	610	870

				CATCTATAAAGC	GCCGCCCTGAGTCCC	GHTLDLVFCQGWEEGGGVEELTIAPL			
				CTGAGTCCACAA	CTGATGGGTGAGAA	PWTDHHLIRFRLTAPPNLCRGGGPK			
				GGCCCTGGCAGG	GGGCGGGGTAGAAG	MVRPRRLMDPEGFLTTLGEFPASSA			
				GCTGAGAGGCCT	TGATGTAATAAATAA	GDPVDALVSLWNGEMTRAIDTIAPE			
				GCTGTTGTGCTG	ATAAATAAA	RPLSSNRAKPAPWFTEELAAMKRKK			
				CAGAAGAGGCC		RELECVWRSGYNESDRTRLGSYLRAY			
				AGAAACATCTCC		AAAIKAARRAHLAANIASAQNRPAAL			
				AGCCAGCCTGCC		FRVIRGLLTPTNQGGSPDNSAARCEA			
				TGTTGCTGCCTCT		FARFFADKVALIRSDFDTILTAVSEDV			
				GGGACTCCTTGT		TSAPACPVLLDSFQLVELEDVDKILGE			
				GAGTAGACTGCA		ARPTTCILDPCPSWLVREARGGLAE			
				GCCCCAGGCCCT		WVKVVVNASLREGKFPASLKMAVIK			
				TTCCATCTGGGT		PLLKKPSLDPTQFGNFRPISNLPYLGK			
				GTGACTCAGCAC		VLEHVVASQLQGFLVDTDFLDPAQS			
				TTCCTAGGAAGG		GFRPGHGTETALVALVDDLRRELDRG			
				ACCATCTATAAA		SVSLLVLLDLSAAFDTVDHGILLGRLA			
				GCCTGAGTCCAC		GMGLGGTVLQWLHSFLEGRSQMVS			
				AAGGCCCTGGCA		LGDACSAPQPLTCGVPQGSVLSPML			
				GGGCTGAGAGG		FNIYMKPLGEIIRSFGVRCHLYADDV			
				CCTGCTGTTGTG		QLCHSFPPVTKEAVQVLNRCLAAVSD			
				CTGCAGAAGAG		WMRANKLKLNPDKTEVLLVSRKAEQ			

				GCCAGAAACATC TCCAGCCAGCCT GCCTGTTGCTGC CTCTGGGACTCC TTGTGAGTAGAC TGCAGCCCCAGG CCCTTTCCATCTG GGTGTGACTCAG CACTTCCTAGGA AGGACCATCTAT AAAGCCTGAGTC CACAAGGCCCTG GCAGGGCTGAG AGGCCTGCTGTT GTGCTGCAGAAG AGGCCAGAAACA TCTCCAGCCAGT CTGCCTGTTGCT GCCTCTGGACTC CTTGCCTACTGCC CCCCTCCCTTTTT		GTGLQPVLDGVTLPLKAQVRSLGVIL DSSLSLEPQVSAVARGAFAQLKLVRQ LRPYLGKPDLATVVHALVTSRLDYCN ALYVGLPLKTVRKLQLVQREAARLITG AAYRERTTPLLRQLHWLPISYRAQFK VLALAYKALNGSGPAYLSERVSRYDP PRSLRSSDEALLAVPSASQVRLAGTR DRAFSVVAPRLWNALPIEVRQAPSLL SFRKRTKTWLWDQAFEHESAAKMR			

				CCCAGCGCTGGT GAGTGGACATAA GAAGGGAGGGA AGACTTTTAAAA CTCCAAAGCCAT CCGGAGGAAGA AGGCAGAGGAA CTGACTTTAAAA TCATTCTGTCTCT TGCTGCAGGCCT ACTGCCCTCCTC GCTTTCCCCAGC ATTGGCTATTGA ACTAAGAGTCTG TCTTACTGAGGT CCACTCATTATTT ATATATTGCTGCT ATAACAATCTTTA AATGCTGTTATA ATAACTATTATG CACTGTGAAGCC					

				TGGTCCCCCTGA CTAGTTTACTAA GTCTATCATCTCT AATTGGTCCTGC CCTGGCACCTTA TAATACCATATT GCACCATATTGA ATTAGTTCACCAT CATAATCCCCACT GCCTGCCTTGCA CTTTAGTCAGCTC AGACCCCTATAC CT					
	RTE-1_AC_1	GCF_000090745.1	Anolis carolinensis	CCCCCTCGATGG ATTGTCACCTTGC CGTGGTGAGGG GGCTTGCGTGTT CCGATGAACCTG TGGGCACAACCA CTGGAGTGATGC ACTCCCAGGAGT	TAAGTAA	MLTKTSRKIFYCQLDTILSEIPKEDKIIL LGDFNARVGRDSDLWPGIIGKDGVG NSNSNGILLLTKCGEHNLVITNTLFRQ KNKLKTSWKHPRSKHWHLLDYVITR ARDRRDVLLTRAMTGTDDCWTDHR LIRSTMAIKIVPKRRLQGRKTRRKMN TQALQEPSKRALLQTTLKDHLPTEHP ENVEEHWNKLKTSIITACEESIGYLTK	cd01650	417	698

				GGCCGCAGGGG		KHQDWFDDNDKEIQQLIDNKRKAF			
				AGGTTCCAGACC		QTWQRDTNCAAKKKIYASAKAEVQR			
				AAGCACAATCCG		RTRELKNIWWTKKAEEIQHLADTHD			
				AAGACCCAAAGA		AQGFFKATKIIYGPRNHGIQPLRSSDG			
				CCTCAACGGCGG		TKILKDKTSIALRWKEHYQNLLNRSSN			
				AGCAGGCGGAG		VAEETLSQIPQQQTRDELAALPSLEEV			
				GATAACATGGTA		SNAISQQKNNKASGPDGIPAEIFKEG			
				CATGTTACAACG		GPELIQQLHQLIEKVWMTEKIPADFK			
				GCTGTGAAGGCG		DATIITLFKKGDRTDCGNYRGISLLTSA			
				GAAGAAGGCTG		GKILARILANRLLPVSEDTLPESQNGF			
				CAACAGACTGAG		RPSRGTVDMIFTARQLQEKCREQNQ			
				AAGCCACTGTCG		PLYMAFIDLAKAFDTVNRSALWTILQ			
				TTGTGTTAACCA		KIGCPDKFVNILRLLHDDMMATVLDS			
				CACCACTGCTGG		NGSQSDPFKVESGVKQGCVIAPTLFSI			
				AACCTCACTCTGT		FIAMILHLVDGKLPTGVEIIYRTDGKLF			
				GAAGACTGTGTG		NLSRLRAKTKVTTTSVIELQYADDNV			
				TTGACCGGCCGT		VCAHSEEDLQATLNTFAEAYEKLGLSL			
				GCACCGACCTCC		NIEKTKVLFQQAPANPSAKPGIQLNG			
				ACACATTAAAAA		VTLENVDHFRYLGSHLSTKVNIDTEIQ			
				AAAATCACGCAC		HRLSSASAAFFRMKQRVFDDRDIRRE			
				AGGCGTCTTCCA		TKVLVYKAIVLPTLLYACETWTVYRRH			

				ACAAAAATAAAA		TKLLERFHQRCLRKILQISWEDRRTNV			
				ACAAACCTACAA		SVLEEAKTTSIEAMLLRHQLRWTGHV			
				AAGTCCCATGGC		VRMPDHRLPKQLLYSELKNGKRNVG			
				GATCAGCGAGTG		GQEKRFKDGLKANLKNCGIDTENWE			
				GCGACGGGGGC		ALALERSNWRSAVTSSAAEFEEARTE			
				AGGACTGTGAAA		GLREKRAKRKERQANPDRDRLPPGN			
				TCTGGAAGCCCC		RCPHCGRICGSRIGLFSHLRTHAQDT			
				TAGTCACAGACT		KVGRPSSSNYEGSPK			
				GGCACATGGGC					
				GGTGGATATGGA					
				CTCAGTCGTTCTA					
				CCTCAAGGACCG					
				AGGCAGTTGAGT					
				AGTCCGGCAGCT					
				GTATCCATGACT					
				GAGCAGCCCTTT					
				TTAGGATCCGCT					
				CTGCTCACCCCA					
				CACGGGGAAGG					
				GGCTAGAAAAG					
				GTGCCCTAAACA					

				TAGTCTGCCTCTC CTACCCTGACTG GACTGCCGCGTC CAGAGGGGTCAC CACTCTGCGGCC AAAAAAGAAAAA TGAACTTTGGAA CATGGAACGTAC GGACATTGTTAG ATAACACTGACA GCGAACGCCCCG AACGCAGGACTG CTCTCATTGCAA GGGAGCTGGGA CGCTTTAAGATC GACATAGCAGCC CTTCAGGAGACC CGGAGAGCAGG AGAGGGACAGC TGAAGGAAGAA AAGGGAGGCTA					

				CACCTTCTTCTGG AAGGGACTGCCT GAAGAAGAGCG AAGAATACACGG AGTTGGCTTTGC TATCAGAAACGA CCTGGTGAAGCA CCTGACTGAAGC ACCCACTGGCAT CAACGAACGACT TTCAACCCTCCG AATTAACCTTGC CAAAAACCAACA GGCAACCATCAT ATGCGCCTATGC ACCAACTCTAG					
	Vingi-1_Acar	GCF_000090745.1	Anolis carolinensis	GGGGGACACGG AAAGAGCCTCCC CGAAGATTGAGT GAATTCAGTCGG GCGTCCCCTGGG	TAGTTGCTTGTGATTT CTTTTCTTTTTTATTTT ATTTCCATTATTTGAA ATGTATTTGTTGTAG	MDEYQRSLSRPLLTIMSINIEGLSLAK EELLAKMSEDISCDILCIQETHRDITM RRPKILGMQLAVERPHRQYGSAIFVR SGVAISATSLTEVNNIEILSVELDSCTV SSLYKPPGADFYFTPPTSCHNHEAHF	cd01650	469	724

				CAACGTTTCTTGT AAGCGGCCGATC TTTCCACCCCAAA AGCATTGGATGA	CAATGCTTTTGACAC GAAATAAATAAA	VVGDFNSHSCVWGYDEDDRNGEAV LTWADNSRMSLLHDSKLPPSFNSGR WKRGYNPDLIFVKESISHQCTKRVLN PIPNTQHRPICCVAYAAVRPKSVPFR RRYNFNKANWTKFTETLEAAISDIEPS IENYDLFVEAVKRSSRLSIPRGCRTSYL PGLNEESLNQLQEYLRLFQENPYSDG TIAAGQKLSTALANAKKDRWIELLEN LDMSKSSRKAWQLLRRLDSDPLVNP GHANVTPDQIAHQLIQNGKTNCSRIK MKINRVPELETHQLSSPLNLKELREAI KRCKTGKAPGLDDLMMEQIKHLGPK AENWLLKFYNQCLAHKQIPRAWRKT KIIAILKPGKDASNARNYRPISLLCHLY KVYERMLLNRLGPVIEPKLIAQQAGF RPGKNCTGQILHLTEHIEEGYEKGCIT GTVFVDLTAAYDTVQHRKMLHKVYH ITRDFDFTKTVQTLLENRSFYVEFQGQ KSRWRRQKNGLPQGSVLAPTLFNIFT NDQPQPPLTKSFIYADDLGLTTQAKD FETVEKQLTNALKDLSSYYKENHLKPN			

						PAKTQVCAFHLRNREANRKLKVTWE GQELEHCFHPKYLGVTLDRTLTYRKH CMNTKHKVAARNNILRKLTGSAWG ADPQVIRTSALALSFSTAEYACPVWH KSAHAKQVDIALNETCRIITGCLKPTP VDKLYKLAGIAPPDVRREVAANGERK KVEHCESHPLHDYHPPPTRLKSRKGF MRTTTPLDVPPATARVSLWAAKPGN SNWMAPQEGLPPGANQEWATWKS LNRLRSGVGRSKDNLARWHYLEESST LCDCGAEQTTQHMYACPQCPASCTE EELFKATDNAVAVARFWSKTI			
	RTE-1_AG	GCA_001542645.1	Anopheles gambiae	TGCTCTGTAATG GGATGGGATCCG AAGCCCCATTGA GGGATAATACAG GCTCTCCCATCCA ACTCCTATTCCGA CACGTCCTCGTC GTGCAGAGTGGT AACATGTGTCAC	TAGCGCCGGATAAGT AAGTAAGTAAG	MGSWNVRTLSEAGALKKLDDALATL SMDLVALQEIRWLGNGVHNRRGKH CYDIYYSCHDRHRVLGTGFAVGPRLK PAIMDFKAINDRLCTLRMRGKFFNISL INVHAPTEDKEEEEKDLFYGRLARIVD ACPRHDLIIILGDFNAKVGREPMYRQ YTGCHSLHEHSNDNGSRLVQFAAAN NLVVGSTKFARKKIHKITWAHPGGES FNQIDHVLISRRRQSSLLNVRTYRGA	cd01650	521	783

				CACATATCCAAG		NIDSDHYLVGLVIRCRIARPRANGGG			
				CTTGGGTACACG		ENTQARLNTDSLRDIAVQQEFKTALE			
				GGCTTGACCCAA		ESLLPEDRYETTSERWNALKTKIINCA			
				CCCCTTGGGCGG		RNILPPRRGNTKSGWFDDECRQVTE			
				ATGGTGGCACAT		RKNTAYRAMQQRHRTRACAEEYSRL			
				GGCGAACCAGG		RREEKRVHRSKKHALEEQNMRELEQ			
				AGGGGGGTGGG		TREAYGPTRKFYQAIAGHRNNVVPK			
				TATCCCGGGAAA		VTCCRNKDGDLVSNQPEVLSRWAQY			
				CTGGGTGCCTGA		FDELLNDQFSEQLEAPLADNVMLLPP			
				ACGTCGGGGGG		SIEETRKAIRRLKNNKAPGTDGIAAEL			
				GCAACCCGACGC		VKNGGARLENEIHQIVTEVWDSESM			
				TAAACAAAACGG		PCDWNLGIIYPIYKKGDRLDCNNYRG			
				TCACGTGGGCGC		ITVLNTAYKIFSLILQDRLVPHVEEIVG			
				AGGGCCAGTCAC		NYQRGFRNGKSTTDQIFTMRQILEK			
				CCAGTCCCATGA		MAEYRNDTYHLFIDFKAAYDSIARVK			
				ATCAACGACTAC		LYDAMSSFGIPAKLIRLVRMTMTNVT			
				GGAAAGCACCA		CQVRVDGKLSGPFATTKGLRQGDGL			
				GAAATATTCGAA		ACLLFNLALERAIRDSRVETTGTIFYKS			
				ACGGACTTAACG		TQILAYADDIDIIGLRLSYVAEAYQGIE			
				ATACCGACCCTA		QAAENLGLQINEAKTKLMVATSADL			
				CGCAATGCCCCC		PINNPNLRRRDVQIGERTFEVVPEFTY			

				GGACTACTCTGAAA		LGSKVSNDNSMEVELRARMLAANRSFYSLKKQFTSKNLSRRTKLGLYSTYIVPVLTYASETWTLSKSDEALLAAFERKMLRRILGPVCVEGQWRSRYNDELYEMYGDLTVVQRIKLARLRWAGHVVRMETDDPARKVFLGRPQGQRRRGRPKLRWQDGVEASAIKAGITDWQTKARDRERFRTLLRQAKTAKRL			
	RTE-5_AC	GCF_000002075.1	Aplysia californica	TGGCTGGAGCTAATGCTCCTGGTAGGGCCACCCATGCCAGACAGGTCGACGGGTAGAGGTCAGACATAGCGTGATCCACCGGCCCTCCAGGTTTCGGAGGTTCAGCTCAGGGCTAGCGACCCAGACTGGTAAAAAGCATATGTTACGGAAACAGC	TAGTAAGTAAATAAGTAA	MNAGSESRQEAVCRKIEDLNTAKTKVIIGLWNVKTLYGTGKSAQVTAEMRRYGLHVLGISESQWTGSGWLRTATGETLIYFGREDNQHREVVATILNKGLEKCLMEWKPINSRMMTLRLKGKHTNTSIIQCYAPTNDSEDESKDYFYEELQRGLTGIPRHDLIVVMGDLNVKVGEDNTNYERTMGQHGMGTRNENGERLVDFCALNNLVIGGTLFPHRTIHNLTWNSPNGRDKNQIDHLMIGGMLRRSLQDVKVRRGADVGSYHHLVTSRIKFKLKKTFTREATRTQFNIQKLSDGNVTSAFRLELKNRFQALQDLEETTPNTDEVNTQWKL	cd01650	557	811

				AACAAAGAGAAA CCATATTTCTGTG TGTGTTGGACCT TCAGAGTCGCTT GTGACTCGT		IVEAYTESSKTCLGVKKKNASNEWIQ QGTWNIIEERREIRKRELAAKSIRLQE KYAEDYKEVCQKVKRLVREDKREAM ENLATETEEAAAKGEQGNVYKITRIIC GKYHGGMSGPIKGKIGRLLTIEKEQE ARWPEHFREVLNQPPPEDTPDIQEA TLNLDISSDPPQRHEVVKAVKSLKEG KSPGMDNLNAELFKVDPELTTTILQP LFKTIRKGEQITEDWRTGVIVKVPKKR ALGDLNNWTGITLLSVASKILAKIIINR MSDAVDICLRKEQARLRKERRCTDQI FALRNITEQCTERQRQLYVNVVDFEK VFDGVHRESLWCILRSYGIPGKIVEVI KSFYQSFKHVVGNRTTAFEVKTGVR QGCNMSAIMFNLAIDWVMRKTTED TSRGIRWTLFSTFEDLDLVDDLAQLS HTNQHMKEKTNPLHKYERHIGLRISQ KKTEMMTLNVSSPTLIMENNAHLPIT DKFTHLGSINKTDGGTKEDIMNRLGK ERSVHRSVINIWKSEQFSIQTTLKQYN SCVLSTLLYGFECWRKTGHDRAKQSS			

						FHTTSLRQILRIFWPDNISNTDLLQK WRQEDRTTIIKRRRWRLIGHVLRREP DSVVKTAFHWTPEGHRKRGRPKVT WRRKVESEMKSWNHSWGTLTKLG QNRQSWRSFVAARNSRWRNG			
	BovB	GCF_002263795.1	Bos taurus	GCGCAGGCGGCT GCGACGGGCGC ACTAAGCGCGGC CGAGAGGAGCT ACCCCACGTCCG AGGTCAGGGGC AGAAGCCGGGA GGACCCCATGCC CGAAGGGCGGC GGCCAAGAGGA GTTACCCCACGT CCGAGGTCAGG GGCAGCGGCCG AGAGTGCCAGAC TGCGACGGCGCA GGAACGGCCGA	TGAACTGAACTGA	MKLSHARGATQDGRVMVERSDRM WSTGEGNGKPLQYSCLENPMNSMK RQNDRILKEELPRSVGAQYATGDQW RNNSRKNEGMEPKQKQYPAVDVTG DRSKVRCCKEQYCIGTWNVRSMNQ GKLEVVKQEMARVNVDILGISELKWT GMGEFNSDDHYIYYCGQESLRRNGV AIMVNKRVRNAVLGCNLKNDRMISV RFQGKPFNITVIQVYAPTSNAEEAEV ERFYEDLQDLLELTPKKDVLFIIGDWN AKVGSQETPGVTGKFGLGIRNEAGQ RLIEFCQENALVITNTLFQQHKRRLYT WTSPDGQHRNQIDYILCSQRWRSSI QSAKTRPGADCGSDHELLIAKFRLKLK KVGKTTRPFRYDLNQIPYDYTVEVRN RFKGLDLIDRVPDELWNEVRDIVQET	cd01650	610	861

				GAGGAGCTACCC		GIKTIPMEKKCKKAKWLSGEALQIAV			
				CACGTCCGAGGT		KRREVKSKGEKERYKHLNAEFQRIAR			
				CAGGGGCGGCG		RDKKAFLSDQCKEIEENNRMGKTRDL			
				GCCGAGAGGAG		FKKIRDTKGTFHAKMGSIKDRNGMD			
				ATACCCAGCGTC		LTEAEDIKKRWQEYTEELYKKDLHDP			
				CGAGGTCAGGG		DNHDGVITDLEPDILECEVKWALESIT			
				GCGGCGACGAG		TNKASGGDGIPVELFQILKDDAVKVL			
				AGGAGTTACCCC		HSICQQIWKTQQWPQDWKRSVFIPI			
				GCGTCCGAGGTC		PKKGNAKECSNYRTIALISHASKVMLK			
				AGGGGCGGCGG		ILQARLQQYVNRELPDVQAGFRKGR			
				CCGGGAGGAGC		GTRDQIANIRWIMEKAREFQKNIYFC			
				TACCCCACGCCC		FIDYAKAFDCVDHNKLWKILKEMGIP			
				CTAAGCCCGAGG		DHLTCLLRNLYAGQEATVRTGHGTT			
				CCAGGGGCGGC		DWFQIGKGVRQGCILSPCLFNLYAEYI			
				GGCCGGGAGGA		MRNAGLEEAQAGIKIAGRNINNLRY			
				GCAACCCCACGC		ADDTTLMAESEEELKSLLMKVKVESE			
				CCGAGGCCAGG		KVGLKLNIQKTKIMASGPITSWEIDGE			
				GGCGGCGGCCG		TVETVSDFIFLGSKITADGDCSHEIKRR			
				GGAGGACCAACC		LLLGRKVMTNLDSIFKSRDITLPTKVRL			
				CCACGTCCAAGG		VKAMVFPVVMYGCESWTVKKAERR			
				AGCCGTGGCTGC		RIDAFELWCWRRLLRVPWTARRSNQ			

				GCGGGCGCAGG AGGGCCTAGAG GAGCTATCCCAC GTTGAAGGTCAG GAAGGGCGGCG GTGAGGAGATAC CCCTCGTCCAAG GTAAGGAGCAGT GGCTGCGCTTTG CTGGAGCAGCCG TGAAGAGATACC CCACGCCCAAGG TAAGAGAAACCC AAGTAAGACGGT AGGTGTTGCAAG AGGGCATCAGA GGGCAGACACAC TGAAACCATACT CACAGAAAACTA GTCAATCTAATC ACACTAGGACCA		SILKEISPGISLEGMMLKLKLQYFGHL MRRVDSLEKTLMLGGIGGRRRRGRQ RMRWLDGITDSMDVSLSELRELVMD REAWRAAIHGVAKSRTRLSD			

				CAGCCTTGTCTA ACTCA					
	L1-1_Cho	GCA_00016 4785.2	Choloepus hoffmanni	GAGGCGGGGCA AGATGGCGGAG TGGTGAGGTGTA TGTTTTAGTTACT CCTCCAGGGCAG TAGGTAGAAAGC CAGGAACTGCGT GGACTGGACACC GCAGAGCAATTT GACTTTGGGCAT ACTTCATACAAC ACTCATGAACAC GTGGAACAGCTG AGATCAGCGAAA TCTGTAAGTTTTT GTGGCCAGGGG ACCCGCGCCCCT CCCTGCCAGGCT CAGTCCCGTGGG	CAATAGTACAGACAG CCCAATTATAAAATG GGCAAAAGATATGA AAAGACAGTTCTCTG AAGAGGAAATACAA ATGGCCAAGAAACAC ATGAAAAAATGTTCA GCTTCACTAGCTATT AGAGAGATGCAAATT AAGACCACAATGAGA TACCATCTCACACCA ATTAGAATGGCTGCC ATTAAACAAACAGGA AACTACAAATGCTGG AGGGGATGTGGAGA AATTGGAACTCTTAT TCATTGTTGGTGGGA CTGTATAATGGTTCA GCCACTCTGGAAGTC	MADSRNAFTVITLNVNGLNSPIKRYR FAEWIKKNEPSICCIQETHLRHRDTKK LKVKGWKKIFHASYSQKKAGVAILISD KIDFKCKDVMRDKEGHYILIKGAIQQ EEITIINVYAPNQGATKYMRQTLAKLK EAIDVSTIIVGDFNTSLSPIDRSTRQKT NKEIENLNNLINELDLTDIYRTLHPKSP GYTFFSSAHGTFSRIDHMLGHKTSLN KFKKKIEIIQSTFSDHNGIQLEVNNHQ RLRKFTNTWRLNNTLLNNQWVKEEI AREIAKYIETNENENTTYQNLWDAAK AVLRGKFIALNAYIKKEERAKIKELMD QLKKLENEQQTNPKPSRRKEITRIKAE INDIENKKTIERINNTKSWFFEKINKID KPLARLTKSKREKTHINKIMNEKGDIT ADPEEIKKIIRGYYEQLYANKLDNVEE MDNFLETYEQPRLTREEIEDLNQPITS KEIQSVIKKLPTNKCPGPDGFTGEFYQ TFQKELTPILLKLFQNIEENGTLPNSFY	cd01 650	512	774

				AGGAGGGGCCG	AGTCTGGCAGTTCCT	EANINLIPKPGKDATKKENYRPISLMN			
				TCAGTGCTGGGA	TAGAAAACTAGATAT	IDAKILNKILANRIQRHIKKIIHHDQVG			
				ACGAGGAGGGA	AGAGTTACCCTTCGA	FIPGMQGWFNIRKSINVLQHINKSKE			
				GAACTGCAGTTG	TCCAGCGATTGCACT	KNQMIISIDAEKAFDKIQHPFLIKTLQ			
				CAGCTCTTATTG	TCTCGGTATATACCC	KVGIEGNFLNMIKSIYEKPTASIVLNG			
				GAAACTCATTCT	GGAAGATCTGAAAG	ERLKAFPLRSGTRQGCPLSPLLFNIVLE			
				ACTGATCCAAAC	CAGTGACACGAACAG	VLARAIRQDKEIKGIQIGKEEVKLSLFA			
				TCCAAACATAGA	ATATCTGCACGCCAA	DDMILYLENPEKSTIQLLELINKFSKVA			
				TAGACTGAGACC	TGTTCATAGCAGCAT	GYKINAHKSVMFLYARNEQTEETLKK			
				AGACACCAGAGA	TATTCACAATTGCCA	KIPFSIATKKIKYLGINLTKDVKDLYKEN			
				ATCTGGGAGCAG	AGAGATGGAAACAA	YITLLKEIEGDLKRWKNIPCSWIGRLN			
				CCAGCCCAGCAG	CCCAAATGTCCTTCA	VIKMSILPKLIYRFNAIPIKIPTTYFADLE			
				AGAGGAGATAG	ACAGATGAGTGGATA	KLVIKFIWKGKMPRIAKDTLKKKNEV			
				GGATAGCAAAAA	AATAAAATGTGGTAT	GGLTLPDFEAYYKATVVKTAWYWHK			
				CAGCAAAGAAAA	ATACACACGATGGAA	DRHIDQWNRIENSEIDPQIYGRLIFDK			
				ACCCAAAAATAA	TACTACGCGGCAGTA	APKVTELGHNSLFNKWGWESWISIS			
				AAGCAGAGACTT	AGAAGGAACGATGT	KRMKEDPYLTPYTKINSKWIKDLNIKD			
				TTTGGAGTTCTG	CGTGAAACATATGAC	STIKLLEDNVGKHLQDLVLGGHFLDF			
				GTGAACATAGAA	AACATGGATGAACCT	TPKAQATKEKIDKWELLKLRSFCTSKE			
				AGGGGAAGGGC	TGAAGACATAATGCT	FVKKVKRQPTQWEKIFGNHVSDKRLI			
				AGAGCTCAGGCA	GAGCGAAATAAGCC	SCIYKQRNPTTQ			

				GAGTCAGGCTCA	AGGCACAAAAAGAG				
				TATGCAAATCCA	AAATATTATATGCTA				
				GAGGGCGAGTTT	CCACTAATGTGAACA				
				CCTTGTCTGAAA	TTGAAAAATGTAAAA				
				AGCTGGTTTCTCT	CAAATGGTTTATAAT				
				GCTCTCTGGATC	GTAGAATGTAGGGG				
				GTTCCTTAATGG	AACTAGCGATAGAGA				
				CCCTAATCTCCTT	GCAATTAAGGAAGG				
				GTCTCTTAGCATT	GGGAACGATAATCCA				
				TCAATAGCCCAT	ATAAGAACAGATAAG				
				TAGATCTGCAAG	CTATCGTGGGTAAAT				
				GAGGGCCCTTCT	TTAACGTTCTGGGAA				
				TTTTCTTCTTTTTT	TGCCCAGGAATGACT				
				TTTTTTTTAATCT	ATGGTCTGTTAATTTC				
				TTTTCTCTTTTTCT	TGATGGGTATAGTAG				
				AAAACAATTACT	GAACAAGTTCACAGA				
				CTAAGAAGCCCA	AATGTTGCTATATTA				
				ATACAGAAAGCC	GGTTACTTTCTTGGG				
				TCAAAGACTTGC	GTAGAGTAGGAACAT				
				AATTTGGGCAGG	GTTGGAAGTAAAGTA				
				TCAAGACAAGAG	GTTATCTTAGGTTAG				

				CAGAACTAAGAG	TTGTCTTTTTCTTACT				
				AGCTCTGAGACA	CCCTTGTTATGAGTTT				
				CAAGGCAATAAG	GTTTGAAATGTTTTTT				
				TCCAGTGGCTGA	TATTGTATGTTTTTTT				
				GAAAATTCACTA	TAAAATTTTTTTTATA				
				AACACCACAACT	TAGTTGATTTAAAAA				
				TCCCAAGAAAAG	AAAAGTTAATTTAAA				
				GGGGGCGTCCCC	AAAAAAAAAAACAAT				
				TCACAGCCATCTT	GAAAAAAATATGCAG				
				CCTGGTGGGCAG	AGCCCCCTTGAGGAG				
				GGAACACTCCTG	CTGGTGGAGAATGC				
				CCCGGCGCCAGC	AGGGGTATTGGGCTT				
				CCCACAGCCCAG	CCCCACCTCGATGGT				
				AGCTGCCCCAGA	TGCTGATGTGCTCAC				
				CAACCCAGTGTG	AGACATAGGGGACT				
				ACGGGAAGTGTT	GGTGGTTTGATGGGT				
				TCCAACAACACA	TGAGCCCTCTACCAC				
				CGCACACACCAC	AGGACTTGCCCTTGG				
				AAAATTGGGCAT	GAAGACTGTTGCTGC				
				GGACATTAGCCT	AAAGGAGAGGCTAG				
				TTCCTGCACCCTC	GCCTCCCTATAATTGT				

				AGCTGATTGTCC	GCCTAAGAGCCTCCT				
				CAGAGTTGGGAA	CCCGAATGCCTCTTT				
				GGTGGAGCAGT	GTTGCTCAGATGTGG				
				GTGAAAAGGGG	CCCTCTCTCTCTAGCT				
				GAAATTAACACG	AAGCCAACTTGGCAG				
				CCCCATTCAGCC	GTGAAATCACTGCCC				
				ATCCTTTCAGCA	TCCCCCCTACGTGGG				
				GACTGGGAACCT	ATCAGACACCCAGGG				
				CCCTACACAGCC	GAGTGAATCTCCCTG				
				CTGCAGCCCAGA	GCAACGTGGAATATG				
				ACCTCCCTGGGG	ACTCCCGGGGAGGA				
				GGACGGCACTCA	ATGTAGACCTGGCAT				
				CCTGTGACATAG	CGTGGGATGGAGAA				
				CACAGTCATCCC	CATCTTCTTGACCAA				
				TCAACAGAGGAC	AAGGGGGATGTGAA				
				CTGGGGGTGCAC	AGGAAATGAAATAA				
				GGCCTGGAAGA	GCTTCAGTGGCAGAG				
				GGGACCCACTCG	AGATTCCAAAAGGAG				
				CAAGTCTCAGGG	CCGAGAGGTCACTCT				
				GCCATACGCCAA	GGTGGGCACTCTTAC				
				TACCAAGGACTT	GCACAATATAGACAA				

				GTGGGTCAGTGG	CCCTTTTTAGGTTCTA				
				CAGAGACAAACT	ATGAATTGGAGTAGC				
				GTGGCAGGACTG	TAGCAGTAAATACCT				
				AACTGAAGGATT	GAAACTATCAAACTA				
				AGACTATTGCAG	CAACCCAGAACCCAT				
				CAGCTTTAAATCT	GAATCTTGAAGACGA				
				CCAGGAACACCA	TTGTATAAAAATGTA				
				GGGAGATTTGAT	GCTTATGAGGGGTG				
				TGTTAGAGCCGC	ACAATGTGATTGGGA				
				CCCCCCTCCCTGA	AAGCCATATGGACCA				
				CCGCCCAGACAC	CACTCCCCTTTGTCTA				
				ACGCCCCATATA	GTTTATGGATGGATG				
				CAGGGCGGGCA	AGTAGAAAAATGGG				
				ACACCAACTACA	GGAAGGAAACAAAA				
				CACGCAAGCTTG	AAACAAACAACAAAG				
				GTACACCAATTG	GCACCCAGTGTTCTT				
				GACCCCACAAGA	TTTTACTTTAATTGCT				
				CTCACACCCCCA	CTTTTTCACTTTAATT				
				CTCACCACAAAG	ATTATTCTTGTTATTT				
				ACAAAGTGGGG	TTGTGTGTGTGGTAA				
				GAGAACTGGCTT	TGAAGGTGTCAGGG				

				GAGGGGAATAG GTGGCTCGTGGA TGCCACCTGCTG GTTAGTTAGAGA AAGTGTACTCCA CGAAGCTGTAGA TCTGACAAATTA GAGATAAGGGTT TCAATCGGTCTA CAAATCCTAAAA GAACCCTATCAA GTTAAGCAAATG CCAAGAGGCCAA AAACAACAGAAA ATTTTAAAGCAT ATGAAAAAACCA GACGATATGGAT AACCCAAGCCCA AGCACCCAAATC AAAAGATCAGAG GAGACACAGTAC	ATTGATTTTGGTGAT GAATGCACAACTATG TAATGGTACTGTGAA CAATTGAATGTACGC TTTGTTTTGTATGACT GCATGGTATGTGAAT ATATCTCAATAAAAA TGAATTAAAAAAAAA AA				

				TTGGAGCAATTA ATCAAAGAACTA AAGATGAACAAT GAGACCATGGCA CAGGATATAAAG GACATGAAGAA GAGCATGGCACA GGATATAAAGGA CATCAAGAAGAC CCTAGAAGAGCA TAAAGAAGAAAT TGCAAGAGTAAA TTAAAAAAATAG ATGATCTTATGG AAATTAAAGAAA CTGTTGACCAAA TTAAAAAGATTC TGGATACTCATA GTACAAGACTAG AGGAAGCTGAAC AACGAATCAGTG					

				ACCTGGAAGATG ACAGAATGGAAA ATGAAAGAACAA AAGAAAGAATTG GGGAAAAAAATT GAAAAAATCGAA ATGGACCTCAGG GATATGATAGAT AATATAAAACGT CCAAATATAAGA CTCATTGGTGTTC CAGAAGGGGAA GAGAAGGGTAA AGGTCTAGGAAG AGTATTCAAAGA AATTGTTGGGGA AAACTTCCCAAA TCTTCTAAACACC ATAAATACACAA ATCATAAATGCC CAGCGAACTCCA					

				AATAGAATAAAT CCAAATAAACCC ACTCCGAGACAT ATTCTGATCACA CTGTCAAATACG GAAGAGAAGGA GCAAGTTCTGAA AGCAGCAAGAG AAAAGCAATTCA CCACATACAAAG GAAACAGCATAA GACTAAGTAGTG ACTACTCAGCAG CCACCATGGAGG CGAGAAGGCAG TGGCATGACATA TTTAAAATTCTGA GTAGAGAAAAAT TGCCAACCAAGA ATTCTTTATCCAG CAAAGCTCTCCTT					

				CAAATTTGAGGG AGAGCTTAAATT TTTCACAGACAA ACAAATGCTGAG AGAATTTGCTAA CAAGAGACCTGC CCTACTTGAGAT ACTAAAGGGAGC CCTACAGACAGA GAAACAAAGAAA GGAGAGAGAGA TACATGGAGAAA GGTTCAGTACTA AAGAGATTCAGT ATGGGTACATTA AAGGACATTAAT AGAGAGAGGGA AAAAATATATAT CTGACAAACATA AACCAAAGGATA AG					

	R2NS-1_CSi	GCA_003604175.1	Clonorchis sinensis	TAGTCGGCGAGCTGA	TGAATGTACATTCTT GCCATTGAATCTCAC ACCAGACCCAGTATG ACGGACACTTTGTGC CTGATGTGCGAGTCT TGACTGTGCTAGCGC TTACCGCGCCTTGAA GAGCATTCAGCATTG TTTGTCCTTTCTTCGG TTGTTAGACTTTACAC GCATGTTTCCTTACCA AAATTCTTACGTACG TTGGGATTCATCCTA TCTGACTGGAACTGT TGGTTGCATGACTTC GAATGAGACATTTCT TTCTTTTATCTCATAT TCTCCTGTAACCCTTT CGCATTCATCGCTTG CATTCACTTTTTATGT CTGTGACCATGCTCT	MTTSSWCRLGSLDASLGELRVQQLQ TGLTTLFGFNVLVTFDNVHYKTSGAS APVPTSTQERLVGLTCEECGKWCKSK AGLVAHHRVHDNDSVGTNMVAQL ACADCSRLFPTKIGLSQHRRHAHPTQ HNADKLSRVKHSGARWSQQESQSLL RLANNLYPSCETQTALFARLEQYFPG RSAISIKTRLRVLNWQAQQDESSSGG PDQTIGQIAAYSSEADDYSVWFKQTV DCAVSLLESHADSSLASVDLLAFARGL QSGIMTPEQVLSLLDLHASRTFPHTW KTVSRRRRQLAHRMPVNRKQIRRAN YAAIQTLYHQRRKDAASAVLDGSWK DLYKGNCGLPPDAEQYWKQVLSAPK HVDSRPSRVVVPSDWSLIEPITGEEV GRTVRSMGNSSPGLDKLTPRMLRRF NANVLAGYFNLLLLAGGCPPHLCRAR ITLVPKVPNPTSPDQLRPISVSSILVRC FHKVLADRWSRRLQLPSLQFAFLHRD GCLEATSLLHALLRHSSATASNLSLAF VDISKAFDSVSHDTIVRSAEAFGAPSP	cd01650	431	689

					TCAAAAATAAACGAT A	LVRYIAQSYENAVAVFPSSEVHCHRG VRQGDPLSPLLFIMAMDEVLGLSMP QLGYQFHDTLVDGFAFADDWVVCA ESQARLKEKLEAAAVELGRAGMKINA RKTKAMVICGDRKHRATAVSVEPFCF AEELITPLGPTDTVTYLGIPFTFKGKGV FNHRQHLLKLLDEVTRAPLKPHQRM EITRNYLIPRLTYSLVLGQVHRNTLKRL DNYIRQSIRGWLRLPKDTPISYIHAGK QHGGLGIPSLSATIPMQRRVRMEKLL STQCRVLRNVVNDSAFGKVVRDLSLP IRVHGSCVNTKEELVAAWGDSLHNS VDGRGLRELVASPLSNRWLVFPERVF PRIFIRGIQLRCNLLRTRVRSARHGHG GQTILCRGNCGQPESLVHILQSCWIT HDARCARHNRVARELAKRLRRLGYT VFEELRAPTSTSFIKPDLIAVRERRATV IDVSIVSDGRGVTVWNEKKQKYGAD EFSLAIISALRAIGCDVDFLVHQPMIIS YRGICFPQSAKAVIGLGLSKVTVSDLC LLAIVGSLRTYDTFMRGTWR			

	L2-2_CCa	GCA_00401 1595.1	Cyprinus carpio	CAAAAGGTAAAG GGGGAGGTGTT GCTTCAATTTATA ACAATGTTTTCA GGATTTCTCAGA GGGCAGGCTTCA AGTATAACTCGT TTGAAGTATTGG TGCTTAATATAA CATTATCCAGAG AAACAAATGTTA ATGATAAATCCC CTGTG	TAACATACTAATACA CTTCTAATATCCAAAT CCGTTAAAGGATTTT TAGGCTGCATTAATT AGGTAAACCGGAACC GGGAACACTTCCCAT AACACCCGATGTACT TGCTACATCGTTAGA AGAATGGCATCTACG CTCATATTAGTCTGTT TCTCTCTTATTCCGAG GTCACCGTAGCCACC AGATCCAGTCTGTAT CCAAGTCAGAGGGTC ACTGCAGTCACCCGG ATCCAGTATGTATCC AGCCCAGATGGTGG ATCAGCACCTAGAAA GGACCTCTACAGCCC TGAAAGACAGCGGA GACCAGGACAACTAG	MFVLATVYRPPGHHTDFIKEFADFTS ELVLAADKVLIVGDFNIHVDNEKDAL GSAFIDILNSIGVRQHVSGPTHCRNH TLDLILSHGIDVDGVEIMQPSDDISDH YLVLCKLHIAKTVNSTSCYKYGRTITST TKDCFVSNLPDVSQFLSISKTSEQLDD VTETMDSLFSSILNTVAPLRLRKVKEN SLTPWCNEHTRTLKRAARKMERSWR KTKLEVFRIAWRESNLSYRKALKTAKS DYFSSLLEENKHNPRYLFNTVAKLTKN KASTSVDISQHHSSNDFMNYFTSKVD TIRDKIVTMQPSATVSHQTVHYRSPE EQFHSFSTIGEEELYKLVKSSKPTTCM LDPIPSKLLKEVLPEVIDPLLTIINSSLLL GYVPKTFKLAVIKPLIKKPQLDPKELV NYRPISNLPFLSKILEKVVSSQLYSFLEK NGICEDFQSGFRPYHSTETALLRVTN DLLLSSDRGCISLLVLLDLSAAFDTIDH TILLHRLEHFVGINGSALAWFKSYLYD RHQFVAVNEEVSYRSQVQYGVPQGS VLGPLLFTLYMLPLGDIIRKHGVSFHC	cd01 650	385	648

					AGCCCCAGAGACAG	YADDTQLYISSRPGETYQFEKLTECIV			
					ATCCCCTGTAAAGAC	DIKNWMTSNFLLLNSEKTEVLIIGPKS			
					CTTGTCTCAGACGAC	SACNDLERCLSLDGCSVNSSSSVRNL			
					CACCGGGACAAGACC	GVLFDSNLSLESHVSSICKTAFFHLKNI			
					ACAGGAAACAGATG	SKLRPMLSMSNAEMLIHAFMTSRLD			
					ATTCTTCTGCACAATC	YCNALLGGCSARLVNKLQLVQNAAA			
					TGACTTTGCTGCAGC	RVLTRTRKYDHISPVLSTLHWLPIKHRI			
					CTGGAATTGAACTGC	DFKILLITYKALNGLAPQYLNELLLHYS			
					TGGTTTCGTCTGGTC	PPRPLRSQNSGNLIIPRISKSTAGGRSF			
					AGAGGAGAACTGGC	SYLAPKLWNNLPNIVREADTLLQFKS			
					CCCCCGACTGAGCCT	RLKTHLFNLAYT			
					GGTTTCTCCCAAGGT				
					TTTTTTCTCCATTCTG				
					TCACCGATGGAGTTT				
					TGGTTCCTTGCCGCT				
					GTCGCCTCTGGCTTG				
					CTTAGTTGGGGACAC				
					TTCATTTACAGCGAT				
					ATCGTTGACTTGATT				
					GCAAATTATTGCACA				
					GATACTATTTAAACT				

					GAACTGAGCTGCATG ATGACATCACTGAAT TCAATGATGAACTGC CTTTAACTGTCATTTT GCATTATTGACACAC TGTTTTCCTAATTAAT GTTGTTCAGTTGCTTT GACGCAATCTTTTTT GTTTAAAGCGCTATA TAAATAAAGATGACT TGACTTGACTTGAC				
	L2-2_DRe	GCA_008692375.1	Danio rerio	GCAGGAGAAGC ACTTTAGCAGCA TCTAGAACAGCA GCCTGTAAGTAC ATTTAAGATTTGT TTCAGTTGTTGT GTATGGTTTGAG GACTTGTTTCCA GCTGTTTGTGTA AAGTTGTAGGAC	TAATCTGCAATTGCC TCTTTGAATATCACAC TAATTGTACACAAAA AAAAAAAAAAAAAA AAAAAAAATAAAAAA ACTACTAACACTTCCC TTCTTAGACTTTACAG ACCTGAAACTTGCCT TTAGTACTTATTCATT GTTGCTCTTAGTTGT	MCFLIPVVTNTRKTREVRCRRNPHNL RSIHVSTISQLSLSVGLWNCQSAVNK ADFITSIATYSDYNLMALTETWLRPED TATHATLSANFSFSHTPRQTGRGGGT GLLISKEWKFTLIPSLPTISSFEFHAVTII HPFYINVVVIYRPPGKLGHFLDELDVL LSSFSNFATPLLVLGDFNIYVDKPQAA DFQTLLASFDLKRAPTSATHKSGNQL DLIYTRHCFTDQTIVTPLQISDHFLLSL NIHITPEPPHTPTLVTFRRNLRSLSPNR	cd01650	508	768

				ATTTAAACTTGCT	GTAAATTGCTTCCTT	LSTIVSDSLPPSRKLTALDSNSATNTLC			
				TTCAGTTGTGTAT	GTCCTCATTTGTAAG	STLASCLDRLCPLASRPARASPPAPW			
				AATACTAGAAGT	TCGCTTTGGATAAAA	LSDALREHRSKLRAAERIWRKTKNPA			
				TGTTTAGCCACT	GCGTCTGCTAAATGA	HLLTYQTLLSSFSAEVTSAKQTYYRLKI			
				GTTTCCTTGGTTA	CTAAATGTAAATGTA	NNATNPRLLFKTFSSLLYPPPPPASSTL			
				CTATAAGAGCTT	AATGT	TTDDFATFFCTKTAKISAQFAAPTTNT			
				GTGTAGCGAACG		QDTTPTPHTLTSFSQLSESEVSKLVLSS			
				CAGACGCGGTTC		HATTCPLDPIPSHLLQAISPAVIPTLTH			
				GCGTCGTCCGCC		IINTSLDSGLFPTTFKQARVTPLLKKPN			
				TCAGTTTCGGCC		LDHTLLENYRPVSLLPFMAKILEKVVF			
				CTTGTTTTGACTC		NQVLDFLTQNNLMDNKQSGFKKGH			
				GGGAGGCGTGT		STETALLSVVEDLRLAKADSKSSVLILL			
				CCAAACGCCAGG		DLSAAFDTVNHQILLSTLESLGVAGTV			
				TAACCACTATAT		IQWFRSYLSDRSFRVSWRGEVSNLQ			
				AGTGAGCACGGT		HLNTGVPQGSVLGPLLFSIYTSSLGPVI			
				AGCATTAGTCGG		QRHGFSYHCYADDTQLYLSFHPDDPS			
				CAGGAGAAGCA		VPARISACLLDISHWMKDHHLQLNL			
				CTTTAGCAGCAT		AKTEMLVVSANPTLHHNFSIQMDGA			
				CTAGAACAGCAG		TITASKMVKSLGVTIDDQLNFSDHISR			
				CCTGTAAGTACA		TARSCRFALYNIRKIRPFLSEHAAQLLV			
				TTTAAGATTTGTT		QALVLSKLDYCNSLLAGLPANSIKPLQ			

				TCAGTTGTTGTG TATGGTTTGAGG ACTTGTTTCCAGC TGTTTGTGTAAA GTTGTAGGACAT TTAAACTTGCTTT CAGTTGTGTATA ATACTAGAAGTT GTTTAGCCACTG TTTCCTTGGTTAC TATAAGAGCTTG TGTAGCGAACGC AGACGCGGTTCG CGTCGTCCGCCT CAGTTTCGGCCC TTGTTTTGACTCG GGAGGCGTGTCC AAACGCCAGGTA ACCACTATATAG TGAGCACGGTAG CATTAGTCGGCA		LLQNAAARVVFNEPKRAHVTPLLVRL HWLPVAARIKFKTLMFAYKVTSGLAP SYLHSLLQIYVPSRNLRSVNERRLVVP SQRGKKSLSRTLTLNLPSWWNELPN CIRTAESLAIFKKRLKTQLFSLHFTS			

				GGAGAAGCACTT TAGCAGCATCTA GAACAGCAGCCT GTAAGTACATTT AAGATTTGTTTC AGTTGTTGTGTA TGGTTTGAGGAC TTGTTTCCAGCTG TTTGTGTAAAGT TGTAGGACATTT AAACTTGCTTTCA GTTGTGTATAAT ACTAGAAGTTGT TTAGCCACTGTTT CCTTGGTTACTAT AAGAGCTTGTGT AGCGAACGCAG ACGCGGTTCGCG TCGTCCGCCTCA GTTTCGGCCCTT GTTTTGACTCTCG					

				AGGCGTGTCCAG CTGAATTCAATC AGCTAGTGCTTT GGGGTTATATAA ACAACTAGTTCA CCGCGGCAGCG GTCGCGGCAGCC TCGTGTGAAGAC CGACGAGGGTA AAGACCATCGAC TCTACCTGCGCG ACTCCACCGAGC AAAGACACCGAC AAAGCACTTGAG TACTTTACTGTAT TGTTTTACTTTAC ACTTATTTTTTGT TGTCAGTGCACT TTTATT					

	RTE-1_DCa	GCF_001605985.2	Dendrobium catenatum	GTTGCCAGGTCCCAAGTCCTGGATGAGAAAAGTGGAGGGTTGCGTTAGGTAGTCGACAGCCAGCGTAAAACTTGTCGAACAATAATGAC	TAGTGGGATAAAGGCTTGATGTTGTTGTTGTTG	MNPLRNCVNARASPGSDALYGLNTVTSVQGPLHRRPGAVLNRQGFPHQGRMWAKKLVHETKIRIGTWNIGSLTGKLMELIDIMIHRKLNILCIQETKWVGEKAKEIDSLGFKLWYSGRSRTRNGVGVIIDKSLKDNVIKINRIGDRIIALKIVVDDKIVNIISAYAPQVGLDMSSKLNFWNDLEELIQNIPLEERVFIGGDLNGHVGREVDDFKSIHGGHGFGTTNEGGKTILDFCLAYDFLIANTLFKKRNEHLITYKNGYSTTQIDYFLIRKLNRLDCKDCKVIPSENLCTQHRLLVLDFKIKGHKIKNKVKQFSRIRWGHLKNINKNIFESKILDLKILSVHEDANIMWEKVASDIKNIAKDVLGVSKGKAPNSKNSWWWNPNVQEKIKNKKICYKNWFNFKNSDNWEIYKVARKDSKKAICDAKTLAFSDLYNRLDTKEGERYMYKLAKQRDRSSKDLVQVKCIKDEDQKILFLEDDIQQRWMRYFAKLFNEDCENKIDYSDLSTLNENKDFCFYRRISKTEVREALKKMKGGKAVGPDDIPIEVW	cd01650	581	829

						KCLGDEGISWLTKLFNTILKTKKMPDE WRMSVLIPIFKNKGDAQDCANYRGI KLMCHTLKLWERVMEHRLRRDTNVS QNQFGFMPGRSTMEAIHLLRGLME KYRENKEDLHMIFIDLEKAYDKVPRE VLWRVLEKKGVNNAYIQIIKDMYAG AITCVQTQGGLTKYFPISVGLHQGSA LSPYLFALVMDVLTRHLQEDVPWCM LFADDILLVDKTREGVEGKLELWRSTL ESKGFRLSRSKTEYMECNFSSNRPSE GIVTLGDQVINKSTRFRYLGSIVQSDG EIDGDIISRIQVGWLKWRNASGLLCD RNVPLKLKGKFYKMVVRPAMLYGAE CWPLKEKHNTKLSVAEMRMLRWM SGFTLRDRIRNEHIREKVGVAPVEDKI RESRLRWFGHIKRRPSDDPVRRVEVL DLTYVKKGRGRPKKTWLENIRNDLSL LDLNENLTFNRTQWRKRIHVADPT			
	Daphne-2_LMi	GCA_000516895.1	Locusta migratoria	TTTCACCAGAAT ATTGGGGGACTA AAGAATAAAATA	TAAAGGAATTTCTGA ATGGCAACTCCTTCT ACACTTTAGATGAAT	MCIMSEISTSNNKIKTIWNIITRETGQ PRTQEDSITIKVNEKITNKKSEVENTF NNYFLNVVENIGSKCSLEEARQLMEQ	cd01650	153	443

				GATGAGCTTTTG GTTTGTTTAGAT GATATAGAAACT GATAATGTAATA GATATACTATGC CTGTCTGAGCAT CACATTGTCACT GGTATGGAAAAT GTGAATGTCAGT GGATATAAATTA GCTGCAAATGTA AGTAGAGATAAT AAGGTAAGAGG AGGAGTTGCCAT ATATGTCAAAAG CTATCACAGTGT AAAAAATTTAGA AACAAAAAAGTT TTGTGTAGAGCA ACATATTGAAGC ATGTGCCTGTCA	TTTTAGATAAAAATT AGTAATTTAGGAAAA AAATTAAAAATAAAA AAAAAATAAAATAAA ATATTAAAATATTGA GTGTCAAGTAATAAT ATGTATAGACACATT TCACTAAACTGACAC GTTCCACATCATTAC GAAATGTCGTATTTA TGATCTATGGAACAA GTATTAATCTAATCTA ATCTAA	GVPRQFDTMEIPPTSQSEIRKLINSLK SKSSHGIDGISSRVLKACSQEICRILSHI CNSSLEQGIFPDRLKYAIVKPLHKKGD TLDVNNYRPISLLTALSKILEKVMYSR VASHINKNKVLTKCQFGFQKGFSTEN AIYIFTNEILNALDNRKSPVGIFCDLSK AFDCVNHGILLDKLKYCGMSGKVHK WFKSYLTGRVQKVDISNSHNIQKQPS PQTGELSSMGCHRVRYWDLCCSYTK TAESSNWGTIKHGVPQGSVLGPLLFL IYINDLPFYIHKDAKLVLFADDTSIAITP ESQELTDEIVNEVFQRIIRWFSVNGLS LNFEKTQYIQFHTINGITPLINIDFNQK SVAKVEYSKFLGLCIDEGLNWKKHIED LLKRLSSATYAIRVIANFGDVQLSKLAY YAYFHSLLSYGIIFWGNSSFSKKVFIAQ KRVIRIIAGAHPRSSCRHLFKDLGIFTIY IYIYSLMKFVINNPTEFKCNSSVHSYN TRRKGDLHYSRLNLALSQKGVNYSAT KVFGHLPSSIKILTDSQLAFKKN			

				GCTTAAACTCAA TGATGATTCTCTT ATAATTGTAACA GTATATAGGTCC CCTTCAGGAAAC TTTCATTATTTCT GGAAAACTTGGA TGCCTTGTTGTG CTATCTGTCAGA TAGGGGGAAGC AAATTATTATTTG TGGTGATTTCAA TGTTGATTCACT GAAAGAGTGTAA TAGGAAGAATGA CCTGGAAGTCTT ACTTGGCTCTTTC AATTTGACATCT GTTATTGATTTTC CTACTCGGGTAG TAAAGGACAGCA					

				GTACATTGATAG ATAATTTTTTATA GACCAAGATAAG TTTAACAACATA AATGCTTGTCCT GTTGAGAATGGT CTTTCTGATCATG GTGCACAGCTAG TAACAGTATATG ACATAGCTCCTTT CAGTAATTCGAA ACAGTCCTCCAG AGTTGTGCGCTC AATTAATGATTT AACAACTAGAAA TTTCAAAGAAAG TCTACAGCAGTT AGACTGGGATGA GGTTTATAAGGA ACCTGATGCTAA TTTAAAATATAA					

				CTTATTTCATGAT ACATTTGTAAGG GTATTTGAAAAC TGTTTCCCCAAG AAATTAGTTAAA TATAATTGTAAG AAACCATGCAAA AAACCATGGCTT ACTAAAGGAATA AAAATATCTTGT AACCGAAAAAG GGAACTATATCT ATCAGCAAGAAA GAGTAATAACCC AGAAACAATCAA ATATTATAAAAA CTACTGTGCTAC ATTAAGAAAAGT TATTAAAAAATC CAGAAGC					

	RTE-2_LMi	GCA_000516895.1	Locusta migratoria	CCACCCCCATGA ACCACAGACATT GCCGAGGTGGG GTGGCTTGCGTG CCTCAACGATAC AGATAGCCGTAC CGTAGGTGCAAC CACAGCGGAGG GGTATCTGTGGA GAGGCCAGACA AACGTGTGGTTC CTGAAGAGGGG CAGCAGCCTTTT CAGTAGTTACAG GGGCAACAGTCT GGATGATTGACT GAGCTGGCCTTG TAACATCAGCTA ACATGGCTTTGC CATGCAGTTACT GCGAACGGCTGA	TGAAGACCACAACA	MMASSWVKYSGGKIVPHSDLRAGT TQEDVVIKKNKTGILRIGAWNVLTLN RVGKLENLKREMDRLKLDMVGISEV RWQDEQDFWSGEYRVINTKSVGGN AGVGLIMNKKIGMRVSYYEQHSERII VVKIDTKPTPTTVVQVYMPTSSADDE DIEEMYEEIKEIIQKVKGDENLIVMGD WNSVVGKGREGNVVGEYGLGKRNE RGDRLVEFCAEHNLMITNTWFKHHE RRLYTWKRPGDTGRYQIDFIMVRNR FRNQVLNCKTYPGADIDSDHNLLVM NCRLKLKRLQKGRKVRRWNLDKLKE PEVVESFRGSIRQGLAETGESNTVEEE WVALRDRVVKAAEDHIGRRTRPNRN PWITQDILNLIDERRKYKNSASETGKR EYRRLKNEIDRKCRLAKQEWLEDKCR TVEECITRGKIDTAYRKIKEIFGDKKSS CMNIKNSDGKPVLSKEGKAERWKEYI EGLYKGNKLEGSILEMEDEVEEDEIG DTILREEFDRALKDLSRNKAPGVDDIP SELLISLGEPVMKKLFHLVCKMYETGE	cd01650	548	807

				AAGCAAGGGGA AACTACAGTCGT TACTTTTCCCGAG GACATGCAGCTC TACTGTATGAAT ATATG		IPSDFKKNIVIPIPKKASADKCENYRTIS LISHGCKILTRIIYRRMEKLVESDLGED QFGFRRNVGTREAILTLRLILEDRLRK GKPTFLAFVDLEKAFDNVDWNTLFQI LKVAGVKYRERRVIFNLYKNQTAVIRV EGHERDAEVGKGVRQGCSLSPMLFN LYIEQAVKESKEKFGEGIKVQGEEIKTL RFADDIVILSETAKDLEEQLNGMDSV LKGRYKMNINKSKTKVMECSVRKSG DAEGIRLGNETIKVVDEFCYLGSKITS DGRSKEDIKCRMAMGRKAFLKKRNL LTSNINLNVRKSFLKAFVWSVALYGSE TWTIRSSDKKMIEAFEMWCYRRMM KIRWVDHVTNEEVLNRIGEKRSFWHI LTKRRDRLVGHILRHQGIVNLALEGSV WGKNRRGRPRDEYSKQIQKDVGCRT YSQMKRLAQDRVAWRAASNQSAD			
	RTE- 1B_LMi	GCA_00051 6895.1	Locusta migratoria	CAAAATAGTGAA GCAAGACCTACA AAAAATCCTTGG ATATCACAGAAA	TGAAGACCACAACAA CAACA	MAGVLEAKCKSVEECIKGKSRYRLQE NKRTFGTRKQLYEYQELDGKPVLSKE GKAERWKEYIEQLYKGNELRDDVVG REEEVDQDEIGDTILRDEFDRALKDLG	cd01 650	148	420

				TTTAATTAGAAA ATATAAAAATGC AGTATGAAGGCA AAATGAGTACAG ACGACTAAAAAT GAGATTGACAGG AAGTCAAAATGG CTAA		RNKAPGVDDIPSELLKSLGEPAMTRL FELVCRMYETGEIPSDFKKNVIIPIPKK SGADRCENYRTISLISHGCKILTRIIYRR MEKLVEADLGEDQFGFRRNVGTREA ILTLRLILEDRLRKGKPTFLAFVDLEKAF DNVDWNTLFQILKSAGVKYRERRVIY NLYKSQTAVIRLEGHEREASVGKGVR QGCSLSPMLFNLYMSRRPFSDVIQFV HEQAVKESKEKFGIGIKVQGEEIKTLR FADDIVILSETAKDLEEQLNGMDSVL KRAYKMNINKSKTKIMECGLRKSGDA EGIRLGNETMKVVDEFCYLGSKITCD GRSTEDIKCRMAMGRKAFLKKRNLLT SNINLSVRKSFLKAFVWSVALYGSET WTISSADRKRIEAFEMWCYRRMLKIR WVDRITNEEVLNRIGEKRNFWHTLT KRRDRLVGHILRHRGIVNLALEGSVW GKNRRGRPRDAYSKQIQVDVGCRSY EEMKRLAQDRVGWRAASNQSPD			
	Daphne-1_LMi	GCA_000516895.1	Locusta migratoria	TAACTAAAACCT CTAGGACAGCAA	TGATGCATGCACTCA TACTATATTAGTATTA	MQLLVLGIDSKQIMKTAKSTYRRVVN PPKIENFRKLLKDMNWKDVYSAHG	cd01650	288	549

				TTGATAATATTTT TATAGACAAATC AAAGGAACAAAA TCATATCATAAA ACCAATAATAAA TGGCCTATCAGA TCATGAT	TTATATTATCAGCAA AAAAAACCTAAAAAT TATGTATCTGTATTTT ATTATAAGTGTATGT ATCTGTATTTATATTA TTAGTATTATCTATTA TTCAGCTTGTAAATTT TGTTTCTTGACATGTT CCACATCCACGAGGA TCTCCTCAACATGGA TCTACGGAACGAATA AACTAATCTAATCTA ATCTA	MNEKYNVFINNVSIMFENCFPLKVTR IKQKSIIKPWITQGIKISCKTKRKMYLS AKSSCNDDDLTKYKEYCKILKKVIQAS KQMHYEKKIATSGNQIKTIWDIVKEE TGRTRKEQEQVALRVDETLVTDGHS VANLFNKYFISVTDSMGLSGSVNNAL EYLKLAFTNNFTYMNMSLTSPQEITSI IKSLKTKHSSGYDEISTKLIKACSCELSTI LSYLCNQSIITGTFPDWLKYADIKPLFK KGDKEIPSNYRPISLLPAFSKILEKVMY KQLLNHLTTNDILSRTQFGFLKGSDIE KAIYTYSENVLNSLDNKLQAAGIFCDL SKAFDCVNHNILLSKLEFYGVTGNAA KWFKSYLTNRKQRVSVQVTSELSQQ SSSDWEEITCGVPQGSILGPLLFLVYI NDLSSVTLPEAEFVLFADDTSIAIKSTS SIVIERSANDIFMNINKWFKANSLSLN FEKTHYMQFRTCKRFPDSICIKYEEEQ IEEVDSIKFLGLQLDNKFSWEEHTTEL QKRLNKSVFSIRVLSDIGDIKITKLAYF AYFHSIMSYGIIFWGNSSSQSKVFRV			

						QKRVIRIICGVNSRTSCRSLFKELGILTT ASQYIYSLMKFVINNISLFPTNSSVHTY NTRNKNDLHKDLKALTLVQKGVHYS GTHIFNNLPSSIKNLATNKDQFKRNLK DFLVANSFYSISEFLNCTK			
	CR1-2B_MUn	GCF_012275295.1	Melopsittacus undulatus	C	TAGGTTACCCTGCTC TTGCAGGGGGGTTG GACTAGATGATCTTT TGAGGTCCCTTCCAA CCCTTGGGATTCTGT GATTCTGTGA	MQIPVAPEGVRTKEEFALVDEDWVR DQLCNLDIYKSMGPDGMHPRVLREL AEVIARPLSIIFGKSWETGEVPEDWK MANVTPVYKKGKKEDPGNYRPVSLT SVPGKVMEQLILDSISRHIKDEGVIKN SQHGFMRGKSCLTNLIAFYEEVTRW RDDGRAVDVVFLDFSKAFDTVSHSILI DKLRKCGLDDQVVRWIKNWLKGRR QRVVVNGAESSWRSVTSGVPQGSVL GPVLFNIFINDLDEGTECTLSKFADDT KLGGVADTPEGCVAIQRDLDRLESW AGRNLMKFNKGKCRVLHLGKNNPM YQYRLGVDLLERSEGERDLGVLVEGR LRELGLFSLEKRRLRGDLISVYKYVKG GCQDDGARLFSVISSDRTRGNGCKLE	cd01650	77	327

						HRRFHVNIRKNFFTVRVTEHWNRLP RGVVESPTLEIFKARLDKFLCDVL			
	RTE- 1_MD	GCA_00000 2295.1	Monodelph is domestica	GGGTGTATGGG GTGCTCAGGGAG GGGTAGTATCTC TGGTATGGAGG GCTTGTCGTGCC CTCCTAGGGCAG CTCTCCAGCCTCT GACCCCCACCTG ACACCCAGCTCT CACTTGTGGCTC CCAGTAGCTGCT AGCATGTGGCAG CGGCCACACCCC GGGCAACGGCTT CGACAGGCCGGC TAAACCTTGTGA GGGTAGCCATCG GGTCATCGACCC CTGGTGAACCAG	TGAAACTGCACAAAG ACAATAGTCATTCTC GATCACCGAGAGACT ACCACT	MDSTAHPNQGRGLEKVSQTLPALQT PGQHTAAGGSSPLSGRNQRKNTKKL LLGAWNIRTLLDRENTPRPERRTALIG KELARYNIDIAALSETRLPEEGSLSEPT TGYTFFWKGRASNEDRIHGVGLAIKT SLLKQLPDLPVGISERLMKIRLPLSKDR YATIISAYAPTLTSTEETIEQFYSDLSAV LHSVPTNDKLILLGDFNARVGQDHER WKGVLGKHGVGKMNNNGLLLLSKC SEFELTITNTVFRMANKYKTTWMHP RSKQWHLIDYIIVRRRDIQDVKITRA MRGAECWTDHRLVRATLQMRIAPR HPKRAQTVRAFYNVSRLRDPSYLQTF QSCLDDKLSAKGPLTGSSTEKWNQF RDAVKETSKAVLGPKQRNHQDWFD ENNTAIEDLLSKKNKAFMEWQNNP NSAPKKDRFKSLQATAQREIRKMQD RWWEKKAEEIQRFADMKNYKQFFS ALKTVYGPLKPTTTPLLSSDGDTLIKDK	cd01 650	588	870

				GGCTTTGCTCAC		KGISNRWKEHFSQLLNRPSSVDQSAL			
				CCAGCATGTGAA		DQIPQNRTIEQLDVPPSIEEVQKAIKQ			
				GACTGCTTCGGC		MSAGKAPGKDGIPTEVYKALNGKAL			
				TGAACAGACGGA		QAFHIVLTSIWEEEDMPPELRDASIV			
				AGAAACCAATAA		ALYKNKGSRAACDNYRGISLLSTAGKI			
				GAAGGTTCAACG		LARVILNRLLSSVSEQNLPESQCGFRP			
				GCTGAGAGGGC		DRSTIDMVFTVRQMQEKCLEQNLSL			
				GACGCAGCAAA		YIVFIDLTKAFDTVNRDALWVILSKLG			
				GCACTGTGGAGT		CPAKFVKLIQLFHVDMTGEVLSGGET			
				GCTTAGGGCGTG		SDRFNISNGVKQGCVLAPVLFNLFFT			
				TTGGAGCACAAA		QVLRHAVMDLDLGVYIKYRLDGSLFD			
				GGACAACACGGC		LRRLTAKTKTTERLILEALFADDCALM			
				CATCCAATGCAG		AHQENHLQTIVDRFSTATKLFGLTISL			
				CTGAGGAAGTCT		SKTEVLFQPAPGRPTNQPCITIDGTQL			
				CCAGATGTAACA		SNVNTFKYLGSTIANDGSLDHEINARI			
				ATTTTTCGTGCCA		QKASQALGRLRCKVLQHRGVSTATKL			
				CTGGACCCAGGC		KVYNAVVLSSLLYGCETWTLYRKHMK			
				TTCCAACGCCGA		QLEQFHQRSLRSIMRIRWQDRITNQ			
				GAGAGTGGGAC		EVLDRANSTSIEVMVLKTQLRWSGH			
				TGTCTCTGTGCAT		VIRMDPQRIPRQVFYGELSAGLRKQG			
				CGGCTTTTCCACT		RPKKRFKDQLKSNLKWAGITPKQLEL			

				TAAATCTCTTTCA CGCACAAGTATC TTTGTGCACACTC ATCTATCCTAACC CCGTCCACCCTCT TCAAGACCTGCG GCGATGGGGGA GTGGCGACGCAA CAGGTGGAGGT GACCACTGGCAG TTGTAGTCACGA TCCTGCACGTAG GCGGCCCACGGA CCAGTGGTCGCT CGGCCCTGTGGG CAGCAGGGACGT TCGGCAGCATCC TGGGCGACTGAG CAGCCCTCTCTA G		AASDRSSWRTHINHAATTFEDERRR RLAAARERRHQATTAPPVTTGVPCP MCHKLCASAFGLQSHMRVHRR			

	RTE-1_NVi	GCF_009193385.2	Nasonia vitripennis	ATGGCGGAGTACTAGCATCGGGTAACTTCACGTACCTGTATGGGAGTGGGGGATCTCTAGGCACCGGATTTCTCGTAAGCAAAAGCATCATACATTCGGTTAAAAGTTTCAAATCCGTCAATGATAGGCTCTCGTACATCATTATCGAAGGTGAATGGTATAGATATGTATTTATTAATGTACACTGTCCTACGGAGGACAAAGAAGAAGCTAAGGATCTCTATTACGAAACTTTAGAGCAGGTAATCGA	AAAAAA	MTNRIAYGGTSKKTVKEAANKVLGKKKKPKSKPWFDEECELWFERRKKAKLDSLQNRSDRTVEEYSNVRKQTSAIYRNKKREYQKNLIRRIETNSKENNPREMYRGINAIRKGFRSRAQLMKDENGDLVTNDNELLSLWKNYFDKLLNVHENSEELGDEIHTAELHVEEPSYQEVEAAIKKLKNNKAAGNDSIPAELLKYGGVELTFKIYKLVCAIWKNETIPENWKESIIIPIFKKGDKTDCNNYRGISLLATCYKVLSNVIQARLTPFAEDIVGDYQCGFRRNRSTSDQMFTIRQLLEKKWEFCETIHQLFIDFKKAYDSIKRSKMYQILVLLGVPKKLVRLIQICLNGSTGKVRVGGNVSEPFMIRDGLKQGDGLSTVLFNLTLEYAVRKMQVSQLGATLNGTTQILGYADDLDILGDCRETVARNAEILIKAVEYTGLEVSESKTKYMIVDKLGICRGEEDLRVGNFTFEKVSEFRYLGTTINDRNEINVEINKRLHSGNACFYAVSNLLKSRLLSKNVKIRIYRTIILPVVLYGCETWALTKQADNRF	cd01650	228	486

				CCAGTTCGCGTC TTACGACACAAG AATAGTATTAGG CGATTTCAATGC TAAAATAGGTAG GGAGGAAATGTT TAGGCCTACTAT AGGGAAGGAAA GCCTGCACGAAG CCAGCAATGATA ACGGTATTAGGG TCATAAATTTCGC GGCGGCAAAAG ATCTTATAATCAA AACTACGTGTTTT AAGCACAAGGAC ATACACAAGGCA ACGTGGACATCG CCGGACGGGGC CACACAGAACCA AATTGATCATTTC		RVFENKVLRKIYGPKKDEETGEWRRL HNDELHNLYASPNINRIIKSRRLGWA GHVARMGDDRTAARVMKGRPMVT RPLGRPRRRWEDNVKADLVEIGRVG VDRRGASWVGLTQDRAAWKACVD EAMNFRVPNAT			

				CTCATTGAAAAA AGACGTAAGGG CTTACAGAGGGG CAGATAGCGACT CGGACCACTTCC TAGTAGTAGCCA AATTAAGAGCTA GACTAGTAGCGA ATCAAAATAGTA AGCGAGCAAACA AGGTAGAAAGCT TCGATATTGAGA AGCTACGAGATA GAACAGAGCGA ATTAGGTACCAG ATAGAAATTAAT AACAGGTTTCAG GCACTTGAAGAA GCAAACACATCG CCGGAGGGGA					

	RTE-2_NS	GCF_000393655.1	Nicotiana sylvestris	TACTGTCAGGGC TACACCAGTGGT AGCGGTGACAGC CCTTTTTGGTTTT TGGGTCGTGGTA TTTTCTGTGTTTT CAGGGAAATTCT CGAAGTTGTTGG AGACAAGTTGGG CGCACGAGCACT AGCAAAGGAGA GCAACCTGGACG AGCGTCAGCAAA GGGCGGTGGGA AGCGGTGCGTGA GCGTACAACTAC ATCGAGTACAGC AAATCAACACAG GTTTGGTTCTCG GGAGTTGGTACC TCCCGTTCTACTG	TAGGGGAAGTTGTG AGATATTTCTACAGC GCACTAGAGTGAGA CTAGCCAGTTAGGAG TTAGTCTTAGGATGC TACTGGTCAGCTACT GATGCAGGGCTTTAT CTGTTGGATATTAGT ATACCTTCCATCTTTT TCGTATTTCCTATATT TCTTATATTGCTGTTA TTTTGTTTTTATGTAT TTTATGTTATGTTATG TATTTTATGTTATTTT ATTATGAGTCTATTG ATAGTACTAATATAT CGTCTCTTGTTGCTTT CTTGAGCCGAGGGTC TCCTGGAAACAGCCT CTCTGCCCCTCGGGG TAGGGGTAAGGTCT	MTIKLVVGECTLNVVSAYAPQVGLDE EIKRRFWEGLDDIVRSIPPSERLFIGGD FNGHIGSSAGGYTEVHGGFGFGERN GGGTSLLDFAKAFDLVIANSSFPKREE HLVTYQSSVAKTQIDYLLLRRCDRRLC EDCKVIPGETLATQHRLLVMDIGIMIR RKKRSVRGRPRIRWGALTKDKAQELE GRLSAMGAWRSSGDANTMWSTTA DCIRKAAREVLGISSGRTGGHKGDW WWNAVVQGKVEAKKAAYLRLVGST GEEEKRANSERYKVARKEAKMAVTE AKTAAFARLYEELGNKGGEKKLFRLA KVRERTARDLDQVRCIKDDDDKVLM GDDQIKRRWQTYFHKLLNEEGDQDI VLGELRNADSPHELSYCRDIEVDEVM EAMRKMRRGRATGPDEIPVELWRC VGRAGLEWLTGLFNVIFKTNRMPEE WRWSTMVPLYKNKGDIQSCNNYRG IKLLSHTMKVWERVVEMRVRRTVSIS DNQFGFMPGRSTTEAIHLIRRMVEQ YRDKKKDLHMVFIDLEKAYDRVPREV	cd01650	438	687

				TTTCCCTTTTGGT GTTAGCTAAATT GATGTTACTTTA GTTCACAATAGT TCAATCATTCAAC TAGTGTACTTGT AGTCACTTGTTTC CTTCTAGTTTGAT TCGTTGTCCGTT GTGCACTAGCGA GTCTTCTTTAGAT TTGTCGTAGGTG TAGTGGCTGCAG TCTGGGATGATA GATTAAGGGCAT GTCCTCGGGTGG GGCAGAGTGTG GGGCCGGGGTC AGGGTGTGGGA CGGGAGGCAGG GGAGGTAGAGG	GCGTACATATTACCC TCCCCAGACCCCACT TGTGGGATTATACTG GGTTGTTGTTGTTGT	LWSCLEVKGVPVAYIRVIKDMYDGAK TRVRTVGGDSEHFPVVTGLHQGSAF SPFLFALVMDAITHHIQGEVPWCML FADDIVLIDETRGGVNERLEVWRHAL ESKGFRLSRTKTEYLECKFGAEPTEAG VEVRLDSQVIPKRGSFKYLGSVIQGTG EIDEDVTHRIGVGWMKWRLATGVL CDKKVPPLLKGKFYRAVVRPAMLYGT ECWPVKNSHIQKMKVAEMRMLRW MCGHTRMDKIRNEDIREKVGVAPM EDKMREARLRWFGHIQRRSTDAPVR RCERLAVVGTRRGRGRPKKYWGEVI RQDMARLRITEDMALDRELWRSSIK VVG			

				GGGCAAGGGAG CCTATAGGTTGA GAATCGGGTCAT GGAACATAGGTT CGCTAACGAGTA AGTCTATAGAGT TGGCGAAGATCC TTCAGAAGAGGA AGATTAATATAG CGTGTGTCCAGG AGACTAGGTGG GTTGGATCGAGG GCGAAAAACGC GGATGGGTATAA GTTGTGGTACTC TGGAGTCCTGAG GGGTAAGAATG GAGTGGGTATCC TGGTAGATAGCC ATCTTAGAGAGT CGGTGGTAGAG					

				GTCAGGCGGGT GAATGATAGACT A					
	RTE-1_OAf	GCA_004365145.1	Orycteropus afer	ACTCGTCTGTCA GTTTGTCATACT GTGGTGGCTTGC GTGTTGCTGTGA GGCTGGAGGCT	TAACAACAACAACAA	MPLVFQMPAGLHMVHRFQWSFQT KTDSEEGRGHPLLKELARENLMNSSE TLSDTLLENFMNSPETLSDTVPEDEPL RLEGTQNMTGKELPHQSRTNLNDM DKVKLSGPPFADVARLKLRRNSCKHP LIIGTWNVRSMNLGKLEVVKNEMECI KIDILGISELKWTGFGHFESDNHMIYY AGNDKLKRNGVAFIITKNVSRSVLKY NAVSDRIISICLQGRPVNTTIIQIYAPT TKAKDEEIEDFYQLLQSEIDQTCNQD VLIITGDWNAKVGNKEESPVVGKFGL GDRNNAGDHMIEFCKTNDLFIANTF FQQHKWWLYTWTSPDGVHRNQID YICGKRRWKSSISSAKTRPGADCGTD HQLLICKFKLKLKTIKTSPREPKYDLEYI PPEFRDHLKNRFDALNTNDQKPEEL WDDIKNIIHEESKRSLKRQERKKRPK WMSEETLKLALEWRVAKANGRNDKI	cd01650	628	879

						KELNRRFQRAAREDKANYYNEMCKD LELENQKGRTCSAFLKLKELRKKFKPQ VAILKDSMGKILNDAGNIKRRWKEYT ESLYQKELVDVHPFQEVAYDQEPMV LKEEVEAALKELAKNKAPGIDGIPIEM FQQMDAPLEALTQLCQEIWKTATW PTDWKRSIFVPIPKKGDPTECGNYQTI SLISHASKILLKIIQKWLQQYIDRELPE MQAGFRRGCGTRDIIADVRWILAQS KEYQKDVYLCFIDYAKAFDCVDHNKL WITLRRMGIPEHLIVLMRNLYIDQEA VIQTEQGDTAWFKIKKGVCQGCILSP YLFNLYAEQIIREAGLYEEEHGIRIGGR LINNLRYADDTTLLAESEDDLKHLLMK IEDYSLQYGLHLNIKKTKILTTGPISSITI NGEKIEVVKDFILLGSTINTNGSSSQDI RRCIALGKSAAKDLFQVLKSKDVSLKT KVCLTQAMVFSIASYACESWTMNKE DRRRVDAFELWCWRRMLKVPWTA KRTNKSVLEEVQPECSLEARMARLHL TYFGHVVRRDHSLEKDVMLGKVEGQ			

						RKRGRPSTRWIDTVAATMGSSIARIV RLAQDQAVFHSAVHRVAMSQRRLD GT			
	RTE-2_OL	GCA_00434 8135.1	Oryzias latipes	ATCTAGCCACAG GGGTTGCAAGCC AGTTACTTCTGT GCCGGTCCCAAG CCCGGATAAATA GAGAGGGTTGC GTCAGGAAGGG CATCCGGCGTAA AAAATTGCCAAA ATAACC	TGATGGGAAAAGCC GAAAGAGAAAGAAG AA	MRIIHNTLDIPDRSRPGLTTTATDAV NLQGVGGNLTTVGRRKRGGRRVRG QREKGKGRNIGLRIGTLNVGTMTGK GRELADMMERRKVDVLCVQETRWK GSKARSIGGGYKLFYHGVDRKRNGV GVILKGEFVNSVLEVKRVSDRMMSLK LEIEGVMVNVVSGYAPQVGCELEVK ERFWSELDEVIESIPRGERVVIGADFN GHVGEGNRGDEEVMGRFGVKERNL EGQMVVDFAKRMEMAVVNTYFQK REEHRVTYRSGGRSTQVDYILCRRGH LREVNDCKVVVGESVARQHRMVVC KMTLEVRKKKRGKTEKKTKWWKLQ NEETCEEFRQKLRQVLGGQDELPDD WETTAEIIRETGRKVLGVSSGKRKDG KDTWWWNEEVQECVQRKRLAKRK WDVERTEEGRQEYKEAQRRVKREVA KAKQKAYDELYDRLDTKEGEKDLYRL	cd01 650	572	818

						ARQRDRDGKDVQQIRVIKDRDGKVL TTQESVQKRWKEYFEELMNEENDRE RREEDVVVVEQEVAEIGKDEVRKALK RMKSGKAVGPDDVPVEVWKCLGET AVEFLTSLFNRILESEKMPEEWRRSVL VPIFKNKGDTQNCSNYRGIKLMSHT MKLWERVVEARLRKKVEICEQQYGF MPRKSTTDAIFALRMLMEKYRDGQK ELHCVFVDLEKAYDRVPREELWYCM RSSGVAEKYVRVVQDMYERSMTVV RCAVGQTEEFKVEVGLHQGSALSPFL FAMLMDRLTDEVRQESPWTMMFA DDIVICSESREQVEEQLERWRFALERR GMKVSRSKTEYMCLNERDQGRSVRL QGAEVKKVQEFKYLGSTVQCDGECG KEVKRRVQAGWSGWRKVSGVLCDR RVPARLKGKVYKTVVRPALLYGLETV AVRQRQEAEMEVAEMKMLRFSLGV TRLDRIRNEYIRGTAHVACVSDKVRE ARLRWFGHVQRRDSGYIGRRMLEM			

						ELPGRRARGRPKRRYMDVLTEDMKL ANVRVEDVHDRVRWKRMIRCGDP			
	RTE-1_OL	GCA_00434 8135.1	Oryzias latipes	TGTGAACCGCCA CCTTAACGTGGT GGGGGAGTTTG AGAGCTCGGGTG ATCCCAGGAGCT AAGCTGTCTGGG GCTTAAGCCCCT GGTAGGGTCTCC CATGGCAGACAG GTCCTGGGTGAC GAGCCAGACAAA GAGCGGTTCAGA AACCCCTG	TAAGCGGAGGAAGA TGGATGGATGGATG GA	MKRNNEDPITSPGLASPGPHSGARP RVGARRRAPGGRGSSHGPRSGTARR SDVGPLSHGPTTRRRPQKGPVQCGL GGSRGRVPRRPKPRSWNLAFGTWN VTSLGGKEPELVREVERYRLDIVGLTS THSLGSETQSLERGWTLYHSGVAQG ERRRAGVGLLVSPQLSRQVLEFTPVD ERVMSLRLRVGDRSLTVVSAYGPNSS SEYQAFLVSLEGVLESAPTGDSIVFLG DFNAHVGNDSGTWRGVIGRNAPPD LNPSGVLLLDFCARHSLSITNTMFEHK GVHQYTWHQDTLGRRSMIDFVVVS GDLRPCVLDTRVKRGAELSTDHYLVV SWIRWRGRRPERLGRPKRTVRVCWE RLAEPSVREVFNSHIRENFKQVPREV GDIESEWTMFSTSIVSAAARSCGRKV SGACRGGNPRTRWWTPEVRDAVKL KKESYQAWLACGTPEAADRYRQSKR AAARTVAEAKTRSWEEFGEAMEKDY	cd01 650	613	868

						RLASKKFWQTIRRLRRGKQFSTGTVF SAGGELLTSTGDIVGRWKEYFEDLLN PTDTPSFEEAESEDFGVGLSITRAEVT EVVNELLGGKAPGVDEVRPEHLKSLD AVGVSWLTRLCSIAWRSGTVPLDWQ TGVVVPLFKKGDRRVCSNHRGITLLSL PGKVYARVLERRVRPIVEPRIQEQQC GFRPGRGTVDQLYTLSRVLEGLWEFA HPVHMCFVDLEKAFDRVPRGILWGV LWEYGVRGALLRAVRSLYDRSRSLVRI AGSKSDLFPVHVGLRQGCPLSPVLFIV FMDRISRRSQGPEGIWFGDHRISSLL FADDVVLLASSNRDLQRALGRFAAEC EAAGMRVSTAKSEAMVLDRRKVVC HLSVGGVSLPQVEEFKYLGVLFTSEGR SEREIDRRIGAASVVMRSLYRSVVVK RELSRKAKLSIYRSIYVPVLTYGHELW VMTERTRSRIQAAEMSFLRRVAGRSL RDRVRSSVTRGELGVEPLLLHIERSQL RWLGHLVRMPPGRLPGEVFRACPTG RRPRGRPRTRWRDYVSRLAWERLGV			

						PPEELEEVAGVREVWASLLSLLPPRP GPG			
	L2-6_OL	GCA_00434 8135.1	Oryzias latipes	AATTCACACATC AGTCACAAAACC TACGACATGCAC ACATCAGATTTT ACCCACAACCAT GAAGATGGCTCT TTTAAATGTCAG ATCACTTTTAAAT AAGTCTTTTATCA TTAATGATCTGA TTTAGACCATGA TTTAACCTGTTTG TTTTTAACTGAAA CCTGGCTTGGCT CTGCAGCTCTAC TGAGTTCTACCTT TTTAA	TGATTTTAACTGCTA ATTTATTCTATTAGTT TTAGTATAGGCTATT TTATGTATTTATTTTA ATTTTATGCATCTTAT TTTCAATTTTATGTTT TTTTTTTCTTTTTATGT TTTTAATTTTAGCATC TTATGATTTTAGCCCC AGTGTTTCTTGAGGG GATCTTTTCCTCCAG GGATTGCCGGCTTGG TCAGCGGTTGTTGCT GCAGTTGTTGCCCTT GCTGGGGCGGCAGG GGTCTAATTTTACAG CTTCTTGGCTGTGAA GGGGACCCTGGTGTT GTCCCGGTCTGGGTG	MFFFSIRGGKKGGGTASIAHKTILSEN ILLSKYCSFEHHAFLFGDPPVLTVTVYR PPTPPSCFIEEFSEFLSSLHIRYNRILITG DFNLHVDKNDDIYACEFLNLLNSMDF QQHVTQPTHNRGHTLDLVITYGLST GVSSVVDLAISDHYCVFFNVTSEIQQE ASVRTVRKRYLTPEVAASFIDIFKQTP AQILPAPCDFIVDHFNNRLQSSIDAV APFKTKTLRTTPKMPWKTDDIKGLKQ NCRRAERRYRKTKLIIHKDILKEQLKIY NNSVKRARTSYFSKIISENKNNPKFLF KTIDILINKDFNKSSMPSSNAACEDFA DHFMGKINAIRSNISSMQYTDLNRPP ALFLPEETLGSFVLVDAEMLGRVVSQ VNTTTCLLDPIPTSLFKTFYGFFESELL NIVNCSLQTGVFPSACKTAVVRPLLKK SQLDPKNLDNYRPVSNLPFLSKINIEK SVATQLHEFLNAQKILEKYQSGFRKN HSTETALLKIVNDLRLNYDSQKLSVLV	cd01650	421	681

					GGCACTGTGGCGAT	LLDLSAAFDTVDHQILLDRLRSLVGLS			
					GTTCCGGTGGCCGG	GNVLKWFYSYLTDRHFYVSMDTCSS			
					GCTGGCTCCTGGTAT	GIHEISCGVPQGSILGPILFNLYMLPLG			
					GAAAGGATTCCCAGA	EVIRRHGVDFHSYADDTQLYIAVSPD			
					TACTGTTTCCTCGCA	DLEPVNTLLNCILDIKSWMAENFLQL			
					GCAGAGCTAAGCCAT	NQDKTEVLIIGSEDKREMILPHLHNFK			
					ACTTGTTTATTTATTT	PSQCVRNLGVLFDSELNFIPHIKNVVK			
					ATTTATTTATTTATTT	TGFYHLKNIARVRPFLSLASTEVLMH			
					ATTTAAGTTACATAG	AFIFSRLDYCNALLSGLPKKSISNLQLL			
					TCATATTTCATTGTAC	QNSAARVLTRTRGREHITPVLKSLHW			
					GTGGGAGCCATGGG	LPVRFRIDFKILLMVYKCFYGLGPSYL			
					TCATGTGGTTTAATG	NDILLKYEPSRTLRSSGAGLLVVPKVR			
					GAGGGGAGTGGGAA	TKTYGEASFCHYGPRLWNGLPENLR			
					GGGTGAAGGGTGGG	AAETVEVFKKRLKTHLFNLAFS			
					TTGGGTTTTTATTGTA				
					ATGTTGTGTGTTTACT				
					GTTTTTAGTGTTAAA				
					GCGCTTCGAGCTGCA				
					CAAAGTGCATGAGAA				
					GCGCTATTTTATAAA				

					TTTGATTTGATTTGAT TTGA				
	L2-2_OL	GCA_00434 8135.1	Oryzias latipes		TGAAAGGTCCTATCT GCACTTTTCTTTCCTT TCCTTCCTTTATTTTT AAATCATTTTTCAATT TTTTTCTTTTCTTTTAA AGTAAATTTTACGTT GATTATTTCTATGAT GTATTGTGATTTTAAT GCATTTTTCTGTTTTG TGAAGCACCTTGAAT TACTTTGTGTACGAA TTGTGCTATACAAAT AAACTTGCCTTGCCTT GCCTTGCCTT	MSSFEYVSFKIELKQTPPLLCIVMYKP PQHSQSFIDDFTEMLSVVCTDFDGLV ITGDFNVHVDNVNDRNAKELNAVLE TFDLIQHVSCPTHSRGHTLDLLITKGV NIYNVSVVDVALSDHFCVFFDLSVAP KPPPGPAVVQRRQINESTRAQFVEM INPENASGANVDEMLNSVTSSILNVL DAIAPMKVKLRKHRQKAPWRNDDL VRAQKQQCRRAERKWRKSKLQVHY GMYKGELYAYNQILCRTRERYFSEIIG SCSSNSRVLFTTVNRLTNPPTQLPIEL VCTPKCNEFAVFFNDKVQGIKKAINS TTHITIERPPTHSKLTHFVPVTDQTVQ ETITRLSSSTCCLDVLPTRFLKSVLNSV LPSITQIVNMSLQSGIFPEALKTAVIKP LLKKSGLDPTVLNNYRPISNLPFLGKV LEKVVYQQLTDFLLLNNSFDVFQSGF RPHHSTETALIKVTNDIRLNTDDGKVS VLILLDLSAAFDTVDHGILLHRLQDW	cd01 650	387	648

						VGISGSALSWLKSYLEDRKYFVEIGSC VSDYMALTCGVPQGSILGPLLFNLYM LPLGQLIRSNNISYHNYADDTQIYVSL TTGEYGPVDSLSHCLQQISAWMQN NFLQLNSNKTEVIIFGPQKQREIVSSYL HSLSLKPSNQVRNLGVIMDSDLNFNS HIKSVTSAAFYHLKNIAKIKNIVSKPDL EILIHAFVSSRLDYCNGLLTGLSKRAVK QLQYIQNAAARVLTRTRKYDHISPVL RSLHWLPVPQRIDFKAALLVYKSLHG RAPKYISDMLVPYEPSRTLRTSGAGLL LIPRVRTKQGESAFQYSAAKIWNSLP EVVRQASSVFMFKSRLKTFLFSRVYD			
	CR1-1_PH	GCA_00158 7735.2	Parhyale hawaiensis	GCGTGGCCTGCG CGTTATCAGGCC GCCCCCATTGTC CGGCCACCGGAC GCCTCTTTGTTTG TATGAGCTGGCT TTGGGTTTGGGG CTCTGGGTAGCC	TAGCAGCGTTTTGGG CTTCCCGGTGTTGCA AGATGCTTCATTTGG GTTGTATTACCTGTCC GTGGGAGCTCAAGC GTGAAGTTTTAATAA TAATAATAA	MLYTIFLVYGILCFFFVYFCIYLYITVYL CFLFFLFLLLCGDVESNPGPGRARGCR LLYCNIRGLHANLAELDFVSRGVDVV CCSETLVAGRRHDAELALAGFQSPFR RLCGSGPGFRGMAVYVRSGFCAYRQ SVHECACHEIIVVRVCGRLNNYYLFSL YRSPATDDSLYDCLLTAMASIQSTDP KAAFVFVGDVNAHHRDWLGSASPT	cd01 650	536	795

				CCTGAGTTGGAG		DCHGVAALDFCTLSSCVQLVRGSTHI			
				GAAATGCCACTC		AGNCLDLVMTDVPDLMTVTVGSPIG			
				TAAAGGCGGATG		SSDHSHLVVSLDLNQVVPVVDTRRV			
				AGACCTCTTTTG		VFVKSRANWVAITRAVRTLPWRQIIH			
				GTCCAACGGCAT		SEDPVSELNDLTVSILERFVPKRTILVR			
				CCCTGAGAGGAG		SRDKPWFDDQCRLAFEAKQAAYRA			
				ATGAACAGGCTG		WRRSRDRTLWQTYVDRRAEAKRVY			
				GTGCCATAGCTG		EEAQRRLRQRSRESLLSIDHPHRWW			
				GGTACTCAGTCT		SELKGSVFGAEPSLPPLVGPGGGIITD			
				TCCGCAGTCTGC		PLARAELLSAHFDGKQSRDVIALPHG			
				TCCTCTCTCTCTT		CHPEPRLTSLAFRSGAIKVLLEGLDPY			
				CCATCAGGTAGT		GGVDPVGMFPLFYKQLADVLAPKLA			
				ACCACCTGATGC		VIFRRLIRLGNFPRCWRTGNITPIPKG			
				TCTTTTCCTTTCC		PVSPYVANYRPITLTPILSKVFEKLIAG			
				TTTCTCTTCTTCTT		KLGRFAEVTGLLPAGQFAYRIGLGCC			
				CTTTTCGGCCAA		DALLSVSHHLQSALDCRSEARLVQLD			
				AGCCTGTACTGG		FSAAFDRVNHRGLLYKLESFGVGGRV			
				TGATGGCAGGGT		LSIIRDFVSERTQSVSVDGVLSASVGV			
				CTGGGGCGCGTA		VSGVPQGSVLGPLLFVLYTSDMFSSL			
				GATAGCTGGCCG		ENTLINYADDSTLMAVIPAPRLRDAV			
				CTTTCGTAGCTAT		AQSLNRDLSRISAWCSAWSMKLNAS			

				CGTTACTGTCGA TTTGTTTTTGCTT TATT		KTKSMIISRSRTLVPQHPQLEIDDTLL QESSSLEILGVVFDEKLTFEPHIRRLVS RASTKIGLLRKVNSVFGDSQVARRCF YAFLLPVLEYCSPVWASAADTHLRLL DRLVSSASRLCADNDLVNLSHRRRVA ELCMFYKVYNNERHWLYSSLPALKVF GRETRAACGAHSFTLEAVRCRTNQF CRCFVPWSAKVWNLLPASAFGRIGL QAFKSAVNGFLLDFL			
	RTE-1_PSi	GCF_00023 0535.1	Pelodiscus sinensis	TGGTACGCTAAG AGCACGGCAAG ACATGAAAGATG CCCTGGTCAACC ACTGCGCCCAGC CCATCTCCAACC GTCTCGACTTTA GTCCTGCCATTG GAGTAAGATGG AATCTGGGAAGA GAGAGTGAGGC TGACTATGCGCA	TAGATGTCAAGGTCC TCTTCGAATACGATG GACGAACAACAACAA C	METGGGYTFFWRGRSSDERRESGV GFAVKNHLVRKLASLPKGVNDRLMT LHLPLPRGKQATLISAYAPTMTNPDE VKDKFYEDLDALLSSVKRTDRLILLGD FNARVGSDHTAWDGVIGKNGIGKC NSNGLLLLKTCAAHDLIITNTIFRLPTR KKTSWMHPRSKHWHLIDYVIVRRKD RQDVRVTKAMPSADCWTDHRLIISK LSFRIKPKRRPQGNKAVMKKINVSRL RNPNTATLLAEDLGNQLLELQLEENA EKDWERFRDIVHSSALKVLGPSHRKQ QDWFDENDEEIKAILAEKHRLHRAH	cd01 650	488	770

				CCTCTCCCTCACT		QNDPSSTAKKNAFNNIRRIVQNKLRK			
				TTAATCCAATTCA		MQDSWLSAKADEIQKFSDRNDAKRF			
				CGCGCAAGTCTT		YEALRSLYGPQPSGSSPLLDSDGATLI			
				GACATCTCGACA		TEKALILQRWAEHFQSVLNRPSSINN			
				TCTCCTAAGTCCT		EAIDRMPQADINHSLDVSPSETETLQ			
				GTGGCGACGGG		AISQLSSGKAPGSDAIPAEVYKGGGA			
				CGAGCGACGAA		VLVNKLTQLFQSFWNQGSIPQELKD			
				GCAGCAGGTGTG		ASIVHLYKRKGNRQVCDNHRGISLLSI			
				GATACACTGGGA		AGKILARVLLNRLINHLEQGLLPETQC			
				GCTGTAGCCGCG		GFRKERGTVDMIFAARQLQEKCQEQ			
				GACCTGCACACA		NRELYTTFVDLTKAFDTVSREGLWRI			
				GGCGGCTCAGGT		MSKFGCPDRFILMVRQFHDGMTAR			
				CTAATGGTCGTC		VLDDGEASEAFPVTNGVKQGCVLAP			
				TTCTTGCTGGTG		TLFSIMFSAMLSDAFQNSSLGIRLRYR			
				GAAGCAACAGCT		TDGKLFNLRRLQAVTKIKETVLRDLLF			
				GGATTCGGCGTC		ADDCALNAGSEQEMQASMDKFAAA			
				TACCTGAGTGAC		CDNFGLLINTKKTEVMHQPAPKAQH			
				TGAGCAGCCCTC		QVPTITVKGQKLQAVDQFTYLGSTLS			
				TTTAGGATTGCA		RSVTIDIEVNCRVAKASSAFGRLLTKV			
				CTGCTCACCTCC		WDRRGISLATKLKVYRAVVLTTLMYA			
				GTAGTAAGAGG		SETWTVYRRHARQLNHFHTICLRRLL			

				AAGGGGCTAGA AAAGGTGCCCTA AACATTGTTTGCT TCACAGAACCCT GGCCAGCTTACC GCGGCTGGAGG GGATCCCATCTC ATGCGGTCGAAC AATAAAATTTCG AACAAAATGCAA GGTGACACCACT CACTATTGGCAC ATGGAATGTACG CACGCTTCAGGA CAACCCTTCAGC AGACCGCCCAGA AAGAAGAACAG CCCTGGTCGCCA GAGAGCTCACAA GATTTAACATCG ATATTGCGGCAC		RIRWQDKVPDTEVLSRAGLPSVYTLL MKAQTRWAGHVARMPDHRIPKQLF YGELFQGKRSHGGQKKRYKDTLKASL KSLDIDTTSWETLAQDRSTWHTLIYQ GCQTSEARRTTIAQEKRALRKARVAK AATVVPTHVCPTCGRTFRARIGLISHL RTHHEQSITC			

				TGTGTGAAACTC GTCTAGCAAATG AAGGTCAGCTC					
	Poseidon-3_PM	GCF_010993605.1	Petromyzon marinus	ATATTGGTCGCG AAAGCCTACGTG TTAAACTGACTA ACCTCTATGGGG AGGATACTTTTCT TAAAACCAGAAA ATTGGAGAAGCT TCGCATTAAGCG G	TAATACACATTACATT TTGATTTAAAGAATT ACTAATCTACAACTC ATTACTAATCATAACC AATGACTGATTACTA ATCGATTATATAATT ACATTACCCCCCCCTC CGCCTCTAATCACTG CACCTGAATGAGTTA AGTGATGACGTCACC CTGCTGCTCTATATA AAACCGGGAGCAAG CGAGGGCTTTACCAT TCTGGCGAATTGCCC GATGATGCTGGAAGT ATTTACCAGCGAAAC GTCGTACTCAAACTG TAAATGTGGATTAAC	MKLISSLAFLKRCRDSKVIPNAVKITH HIKSSKANQILHKASELLLIERIQHNRS ELSYISESLKQLHLDISNSVSMHYWNII DNITQQSADNVFKLTKTNQINKFSRL HNKQIPTENQKCKNKTVLNLTNTVLT GDMVSILSKGLNFAVTPRTIPVDEIIA EVESAVRLLPHNEAEMTRHEVAKIIE NSKVPKVNISSNEAKALRELRVNKDI QILAADKGRTTVVMSRADYDNKIRN LLEESTYIKIKRDPTNTLERKTTAIIANS LIPKNEQYLIKPRASKPPRLYGLPKIHKI GIPLRPIVSTIGSATYQLAKYITKQIQPY IGKTSSFIKNSEHLVNVIKDIKLTPDDIL VSFDVISLFTKVPVDESLKLLDTLVKNT GLQKDIPKLAGHCLNNTYFLWRGDF FKQKEGAAMGSPLSPAIANLYMENF EIRALASSTLKPSFWFRYVDDTLVVW PHGKDTLNDFLNHLNAQNPAIQFTM	cd00304	416	499

					TCTTCAACACACCGG TGGGCTGAAGTAGT GAATTATTTGTCTTTT GTCAACGTGACACTT TTTTTGCTGATTGGCC GTGTCGGGTGGTGG AGGGTTACGACACAT TTATAAATAAACGCG ATCTAACCCAAAAAC CTTTATTATGAATAAT ATTGGTCGCGAAAGC CTACGTGTTAAACT	EMEVNNTLPFLDILLGHGLNGKITHS VYRKQTHTDAYLNAESHHHPAQKSS LIKTLVHRAKTVSDGESLVREIKHVKD ALRQNGYNNAQINRALNAKSRSQSN TTESDQIVKTIKIPYISGVTDKLARVLQ KQNITVRFGTVHKIQKLVSCTKDALPK LSTEGVYRINCSCGSCYIGQTGRNISV RLKEHQRNIKLSQTQLSAVAEHACQ QGHSIDFDNVAVVCKTNYYWQRLIR ESIEIERNKNNFNRSDSYRLSAAWKRI IK			
	Poseidon-3B_PM	GCF_010993605.1	Petromyzon marinus		TAATACACATTACATT TTGATTTAAAGCTTTC GTGACTGCAGTGATT CATGTTCTTCGCTCTT TCGGTGGAGTCGCGT TGGGGGGAAGCAAT AAAATAATAGAATAT ATATGTTTGTTGTTTT GTTGTTTCCCTTCGGC	MVDEIIAEVESAVRLLPHNEAEMTRH EVAKIIENSKVPKVNISSNEAKALRELR VNKDIQILAADKGRTTVVMSRADYD NKIRNLLEESTYIKIKRDPTNTLERKTT AIIANSLIPKNEQYLIKPRASKPPRLYG LPKIHKIGIPLRPIVSTIGSATYQLAKYIT KQIQPYIGKTSSFIKNSEHLVNVIKDIK LTPDDILVSFDVISLFTKVPVDESLKLL DTLVKNTGLQKDIPKLAGHCLNNTYF	cd00304	260	343

					GTGGCTTTGCCGAGG	LWRGDFFKQKEGAAMGSPLSPAIAN			
					GAGCTAAGAATATGA	LYMENFEIRALASSTLKPSFWFRYVD			
					ATGCACATTACAATT	DTLVVWPHGKDTLNDFLNHLNAQN			
					CCTTATAGGGGAAAC	PAIQFTMEMEVNNTLPFLDILLGHGL			
					ACATACTAATTAACG	NGKITHSVYRKQTHTDAYLNAESHH			
					GATTACTAACGTAGC	HPAQKSSLIKTLVHRAKTVSDGESLVR			
					ATATAATCACTAACA	EIKHVKDALRQNGYNNAQINRALNA			
					CACATTACAACTAATT	KSRSQSNTTESDQIVKTIKIPYISGVTD			
					ACTAATCTACAACTC	KLARVLQKQNITVRFGTVHKIQKLVS			
					ATTACTAATCATAACC	CTKDALPKLSTEGVYRINCSCGSCYIG			
					AATGACTGATTACTA	QTGRNISVRLKEHQRNIKLSQTQLSA			
					ATCGATTATATAATT	VAEHACQQGHSIDFDNVAVVCKTNY			
					ACATTACCCCCCCCTC	YWQRLIRESIEIERNKNNFNRSDSYRL			
					CGCCTCTAATCACTG	SAAWKRIIK			
					CACCTGATATGAGTT				
					AAGTGATGACGTCAC				
					CCTGCTGCTCTATATA				
					AAACCGGGAGCAAG				
					CGAGGGCTTTACCAT				
					TCTGGCGAATTGCCC				
					GATGATGCTGGAAGT				

					ATTTACCAGCGAAAC GTCGTACTCAAACTG TAAATGTTGTGGATT AACTCTTCAACACAC CGGTGGGCTGAAGT AGTGAATTATTTGTC TTTTGTCAACGTGAC ACTTTTTTTGCTGATT GGCCGTGTCGGGTG GTGGAGGGTTACGA CACATTTATAAATAA CGCGATCTAACCCAA AAACCTTTATTATGA ATAATATTGGTCGCG AAAGCCTACGTGTTA AA				
	Rex1- 2_PM	GCF_01099 3605.1	Petromyzon marinus	AGATGGCGCCGA TGATGGCAGCCT CGGCCGGTTGGT GCGCCCTGTTTT GTCCAGTTTTGTC	TGTTTTTAACATTTGA TCTACATAACGACAC CTTGCCGATATTAAC TATGTTACCCTATTTG TATTTTTTATTTTTAA	MALMMVASAGRCALFCLVLSVNSVF CQDFRSSLTRDELLNIRETTPVDLIPTF LLPAAELVGILVNAALGFAQATKRRP RGKRSGALVRLRRRGIRTVLPSFFLSN VRSLYNKMDELQLLMVKNKDFLSAS	cd01 650	574	826

				GTTTTGTCTTGA	ATTATATCTTAAAGCT	ALCFTETWLCDLIPDTSLQLAGFHLVR			
				GAGGGGCCTGCT	ATTTTATATTCGATAT	SDRDIALSGKTKGGGICFYINSGWCV			
				GTTCAACACACT	TTATTTAATTTTATAT	DVTVILQHCSPLLESIFITCKPFYSPREF			
				CGTGCTGGTGAG	ACTTCAGCCCCTTGG	ASLILVGVYLPPCPQVKEAQRMLADQ			
				AGAGGCCTGCTG	TGTTAGTTAGACTTTT	ILSVERANPDSLVIVLGDFNKGNLSHE			
				TTCATCACACCC	AACCCTTATTATAGT	LPKYIQFIKCPTREGNTLDHCYTTVSR			
				GTGCTGGTGAGA	AAGACTTGTTTAAAA	AYHAVPRAALGHSDHAMVHLIPAYR			
				GAGGCCTGCTGT	CTTATTTAAGATTTTA	QKLKLCKPAVRTSKQWTSEAVEDLRA			
				TGAGGGGCCTGC	GATTCCTATATGTTG	CLDCTDWDMFTTATNSLDELTEAVT			
				TGTTCATCACACC	AATGTATGCACCATC	SYISFCEDCCIPTRTRVNFNNNKPWF			
				CGTGCTGGTGAG	CTGCCAAAGTAAATT	SAKLRRLRLEKEEAFRSGDRDRFKESK			
				AGAGGCCTGCTG	CCGTGTCGGTGTGAA	NRFSKAVREAKRLYSERLKHQFSTND			
				TTCAGCACACCC	CTTACATGGCGAATA	AASVWRGLRQITNYKPRAPHTTNDS			
				GTGTTGAGGGGC	AAACCCTTTCTGATTC	RLANDLNEFYCRFERQLDCPEPPLPT			
				CTGCTGTTCAGC	TGATTCTGA	QEAPSPTNTTETTISFQEGEVNRLFKR			
				ACACCCGTGCTG		QNPRKAAGPDSVSPATLRHCADQLS			
				GTGAGAGAGGC		PVFTNIFNTSLETCHVPACLKSSTIIPV			
				CTGCTGTTCAAC		PKKLRPTGLNDYRPVALTSVVMKSFE			
				ACACCCGTGCTG		RLVLAHLKAITDPLLDPLQFAYRANRS			
				GTGAGAGGGGC		VDDAVNMALHFTLQHLDAPASYARI			
				CTGCTGTTCAGC		LFVDFSSAFNTIIPALLQDKLFQLNVPY			

				ACACCCGTGCTG GTGAGAGGGGC CTGCTGTTCAGC ACACCCGTGCTG GTGAGAGAGGC CTGCTGTTCAGC ACACCCGTGTTG AGGGGCCTGCTG TTCAGCACACCC GTGCTGGTGAGA GGGGCCTGCTGT TCATCACACCCG TGTTGGTGAGAG GGGCCTGCTGTT CAGCACACCCGT GCTGGTGAGAG AGGCCTGCTGTT CAGCACACCCGT GTTGAGGGGCCT GCTGTTCAGCAC ACCCGTGCTGGT		TTCKWITDFLSDRKQCVKLGTQVSGS RSISTGTPQGCVLSPLLFSLYTNSCTSN HLSVKLLKFADDTTLIGLISGGDESDY RWEADNLVTWCSQNNLELNALKTV EMVVDFRKNTAPLTPITLLDSPVKIVE SFRFLGTILSRDLKWELNISSLTKKAQ QRMYFLRQLKKFNLPKTMMVQFYT AIIESILTSSITVWYAAATAKDKSRLQR IIHTAEKVIACNLPTLEDLHTTRSMRR ARKIMADFSHPGHSLFQLLPSGRRLR SIRTKTSRHKNSFFPSAAGLLNKAKGL H			

				GAGAGGGGCCT GCTGTTGAGGGG CCTGCTGTTCAG CACACCCGTGCT GGTGAGAGAGG CCTGCTGTTCAG CACACCCGTGTT GGTGAGAGGGG CCTGCTGTTCAG CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCAA CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCAG CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCAA CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCAA					

				CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCAA CACACCCGTGCT GGTGAGAGGGG CCTGCTGTTCATC ACACCCGTGCTG GTGAGAGAGGC CTACTGTTCATCA CACCCGTGCTGG TGAGAGGGGCG TGCTGTTCATCAC ACCCGTGTTGGT GAGAGGGGCCT GCTGTTCATCACT CCCATGCTGGTG AGAGAGGCCTAC TGTTCATCACACC CGTGTTGGTGAG AGGGGCCTGCTG TTCACCACTACCC					

				ATGCTGGTGAGA GGGGCCTACTGT TCAACACACCAG TGCTGGTGAGAG AGTCCTGCTGTT CAGCACACCCGT GCTGGTGCGAAG					
	RTE1_Pca	GCA_00402 6925.2	Procavia capensis	GGTAAGCAGGG TAACACCATCAG GTAAGGAGGGA GTGAGCAGCCCC CACTTGTCTGTCA GTGGAGGCTTGC TTGTTGCTGTGA TGCAGGAAGCTA TGCCACTGGTAT TTAGGTACCAGC AGGGTCAACCAT GGTAGACAAGTT TCAGCTGAGCTT CCAGACTAAGAC	TCTGCTGCAAAGGAC CTCTTCAAAGTATTCA AAAGTAAGGATGTCA CCTTGAAGACTAAAA TGCGCCTGACCCAAG CCATGGTATTTTCAAT CACATCATATGCATG TGAAAGCTGGACCAT GAACAAAGAAGACA GAAGAAGAATTGAT GCCTTTGAATTGTGG TGTTGGAGAAGAAT GTTGAATGTACCATG GACCGCCAAAAGAAC	MNSSQTLSDMVPESEPPRLEGTQKT TGEELPPQSRVNLNDVDGVRPLGSSF ADVARLKMRRNSCKHPLIIGTWNVR SMNLGKLEIVKNEMECINIDILGISELK WTGLGHFESDNHIVYYAGNENLKRN GVAFIIKKNISRSVLKYNAVSDRIISIHL QGRPVNTTIIQIYAPTTKAKDEEVEDF YQLLQSEIDQTCNQDALIITGDWNAK VGNKEGSVVGKYGLGDRNNAGDH MIEFCKTNDFFIANTFFNQHKWRLYT WTSPDGIHRNQIDYICGKRRWKSSIS AVRTRPGADCGTDHQLLICKFKLKLK KIRTSPREPKYDLEYIPPEFKDHLKNRF DVLITNDRRPDELWNDIKDIIHEESKR	cd01 650	569	819

				AGACTAGGAAG	AAACAGATCTGTTCT	SLKRQERKKRPNWMSEETLKLALECR			
				AAGGACCCAGCA	GGAAGAAGTACAAC	AAKAKGRNGEVKDLNRRFQRAARED			
				GTCTACTTCTGA	CAGAGTGCTCCTTGG	KRKYYNDMCKELEMENQKGRTHSAF			
				AAGGAATTAGCC	AAGCAAAGATGGCA	LKLKELKKKFKPRVAIVKDSTGKLLND			
				AGGGAAAACCTT	AGACTGCATCTTACA	SESIKRRWKEYTESLYQKELVDVQPFE			
					TACTTTGGACATGTT	EVAYDQEPMVLKEEVEAALKALAKN			
					GTCAGGAAAGATCA	KAPGIDGISIEIFQQTDAALEVLTHLC			
					GTCCCTGGAGAAGG	QEIWRTATWPTDWKRSIFMPIPKKG			
					ACATCATGCTTGGTA	DPTECGNYRTISLISHASKILLKIIQKRL			
					GAGTGGAGGGTCAG	QQYIDKELPEIQAGFRRGHGTRDIIAD			
					CGAAAAAGAGGAAG	VRWILAESKEYQQDVYLCFIDYAKAF			
					ACCCTCGATGAGATG	DCVDHTKLWLTLQRMGIPEHLIVLLR			
					GATTGACACAGTGGC	NLYMDQEAVVRTEQGDTAWFKVRK			
					CGCAACAATGGGCTT	GVHQGCILSPYLFNLYAEQIIREAGLY			
					AAGCATAAAAACGAT	EEEQGIRIGGRLINNLCYADDTTLLAG			
					TGTGGACATGGCGCA	NEEDLRHLLMKIKDHSLQYGLHLNIR			
					AGACCGGGCAGTGTT	KTKILTTGPVSNIVINGERIEVVKDFILL			
					TCGTTCTGTTGTTCAT	GSTINALGSSSQEIKRHIALGK			
					AGGGTCACTATGAGT				
					CGGAACCAACTCGAT				

					GGCACCTAACAACAA CAA				
	Crack- 1_SSa	GCF_00023 3375.1	Salmo salar	TCCAAGATGGCG TAGCAGTCAGAC GTCTTTGTCCTCC TCTTGTCGTGTCC CGTGTATATATTT TTATATCTTTTTT CTTCGCATATCTT TTTAATATTTTTC TTAACCTCAACTT CAAAACACTCTC CTGCAACCCGCC TCACCCAATGTG GAATAGATATGC TATTTTATTTACC TCGAACTGGAAT CCACAACAGAAG CTAGCCAGCTAA CTAGCTACTAGC TAGTAGTCAGCT	TAGCCCATCTATAATT AGCCCAAACAACTAC CTCATCCCCATACTGT TATTTATTTATTTTGC TCCTTTGCACCCCAGT ATCTCTACTTGCACAT TCATCTTCTGCACATC TACCATTCCAGTGTTT ACTTGCTATTATTTCG CCACCATGGCCTATT TATTGCCTTACCCCTA TCTTACTCATTTGCTC ACATTGTATATATAA CTTTTTCTACTGTATT ATTGACTGTATGTTT GTTTACTCCATGTGT AACTCTGTGTTGTTG TATGTGTCGAACTGC TTTGCTTTATCTTGGC	MITRLSMPLLNVNMPCPLLFWLVFIG LFHCRASSPAHYTLSNLSVPPPTHAM TSPGFNDVSRDNISLIITQCLGLPPLYS HPTIPLSVHYTLNLFYRPQKPAPFTLC SGRARRPILIAFSRTLILLLLCSSGDVEV NPGPAVPSSTPTPQALSFDDFCNRKS LGFMHVNIRSLLPKFVLFTALAHSAN PDVLAVSESWLRKTTKNSEISIPNYNI FRQDRTAKGGGVAIYCRDSLQSSVLL SRSVPKQFELLLLKIHLSKNKSLTVAAC YRPPSAPSCALDTICELIAPHLSSELVLL GDLNWDMLNTPAILQSKLDALNLTQ IINEPTRYNPKAVNTGTLIDIILTNLPSK YTSAVFNQDLSDHCLIACIRNGSAVK RPPLITVKRSLKHFSEQAFLIDLALVS WKDIDLIPSVEDAWLFFKSAFLTILNK HAPFKKFRTRNRYSPWFSPDLTALNQ HKNILWRSALASNSPRDMQLFREVR NQYTQAVRKAKASFFKQKFASCNTN	cd01 650	647	905

				AGCCACGGTTAG	CAGGTCGCAGTTGTA	SKKFWDTVKSMENKSTSSQLPTALRI			
				CGGTCATCAGCT	AATGAGAACTTGTTC	GNSVTTDKSTIIENFNKHFSTAGHAF			
				ATCCTTTAGCTCG	TCAACTTGCCTACCT	HLATPTPVNSTAPPTATRPSLPHFSFS			
				AAAAGCTATCGC	GGTTAAATAAAGGTG	QIQSADVLKELQNLDPYKSAGLDNLD			
				CAGTTTGTACAG	AAATAAAAAATAAAT	PFFLKLSAEIVATPITSLFNLSFVSSEIPK			
				CGCGACTCAGAC	AAA	DWKAAAVIPLFKGGDTLDPNCYRPIS			
				CAGAGCATACCG		ILPCLSKVFESQVNKQITDHFESHRTF			
				GACCTATTTTTCT		SAMQSGFRAGHGCTSATLKVLNDIIT			
				CCATATCCCCGG		AIDKKQYCAAVFIDLAKAFDSVNHHIL			
				ATTTCTACCGCA		IGRLNSLGFSNDCLAWFTNYFSDRVQ			
				AGCTCTGGACAT		CVKSEGLLSGPLAVSMGVPQGSILGP			
				TTACACCTGGAT		TLFSVYINDVALAAGESLIHLYADDTIL			
				CATTGCAGCTAG		YTSGPSLDTVLTNLQTSFNAIQLSFRG			
				CTAGCTGCATCC		LQLLLNTSKTKCMLFNRSLPAPARPSS			
				GAGTGACTACTA		ITTLDGSDLEYVDNYKYLGVWLDCKL			
				TTGGCTAACGTT		SFQTHIKHLQSKVKSRIGFLFRNKASF			
				TCGGTCCCGAAG		THAAKHTLVKLTILPILDFGDVIYKIAS			
				CAAGCACCAATT		NTLLNKLDAVYHSAIRFVTKAPYTTH			
				ATCCTGGAGCTA		HCDLYALVGWPSLHTRRQTHWLQVI			
				GCCCATGCTAGG		YKTLLGKVPPYLSSLVTIAAPTCSTRSS			
				CCCATCTCCCGG		RYISLVTPKANSSFGRLSFQFSAAND			

				CTAGCTGAAGAG TTCCATCAGCCA CTCCTTGGGCTA CAATACCTATTTT GCCAACTGACCT GGACCCGTTTTA CTGCCGTGCGGA GCCCCACCGAGC CTTCACGACTGG ACTACCGACGTT ATCTGCCCGAGG GAGTTATCCAAC TGGCCCCTCCGT CGCGACGTAACC TGAACGCCCATC TGCTGGCCTGCT AACCGCGGATCG TTAGCTGTCTAG AGCATATCGGAC TGCTATCTGAAT AGGTCTATCGGA		WNELQKSLKLETLISLTSFKHQLSEQL TDHCTCT			

				CAATTTCTTGGG CCACTATACCTAT TTTGCCAATTGG ACTGGTCCCCTCT ACCACACGGAAC CCTACTAATCCAT CACGACTGGTCT ACCGACGGAAAC GCACGAGGTGG CTAAAACAGACT TCTTCCGTCGCG ACGTCCCTCTAA GGCCCTTCTGCC AGCTTGCTACCT ATGGCCCGCTAG CTGTCTGAATCG CCGTGTCCCCAG CCAGCCCAACTA CTCACTGGACCC CT					

	Rex1-2_SSa	GCF_000233375.1	Salmo salar	CTTCCGGCGCCGACAGAG	TAACATTGAGTGGCT GCTGCCAACATACTG ACTCAACTCCAGCCA CTTTAATAATGGAAA AATTGATGTAATCAA TTTATCACTTAGCCAC TTTATATTATTTATAT TAGATAATGTTTACA TACCCTACATTACTCA TCTCATATGTATATAC TGTACGCTATACCAT CTACTGCATCTTGCC ATCTTGATGTAATTA GATGTATCACTAGCC ACTTTAAACAATGCC ACTTTATATAATGTTT TCATACCCTACATTAC TCATCTCATATGTATA TACTGTACTCTATACC ATCTACTGCATCTTGC CATCCTGATGTAATG	MAASLRVPRKLCSILFFYVLFLTLLPQV ILGFITYSREELLDIRATSTHQHYDQEY DFPEADPLFGPPPRTMDRIPAGDPK QRRRRRRGRRSGLLVRLRRRAHRAPL PSILLANVQSLDNKVDEIRARVAFQR DIRDCNVLCFTETWLTRDTLSESVQP PGFFTHRADRNKHLSGKKKGGGVCL MINETWCDHNNIQELKSFCSPDLEFL TIKCRPHYLPREFSSIIITVVYIPPQADT STALKELHWTLCKLETTYPEAAFIVAG DFNKANLKTRLPKFYQHIDCATRAGK TLDHCYSNFRDAYKALPRPPFGKADH DSILLLPAYRQKLKQEAPALRSVQRW SDQSDSTLQDCFDHVDWDMFRIAA NNNIDEYADSVCEFINKCIGDVVPTA SIKTFPNQKPWIDGSIRAKLNARTTAF NQGKVTGNMTEYKQCSYSLRKAIKQ AKRQYRDKVESQFNGSDTRGMWQ GLQSITDYKRKTSPVADHDALLPDRL NNFFARFEDNTVPLTRPATKTCGLSF TAANVSKTFKRVNPRKAAGPDGIPG	cd01650	579	832

					TATCACTAGCCACCTT AAACAATGCCGCTTT ATATGTTTTCATACCC TACATACTCATCTCAT ATGTATATACTGTAC TCCATACCATCTACTG CATCTTGCCTATGCC GTTCGGCCATCACTC ATTTATATATTTTTAT GTACATATTCGTATTC ATTCCTTTACACTTGT GTGTATAAGGTAGTT GTTGTGGAATTGTTA GGTTATATTACTTGTT AGATATTACTGCATG GTCGGAACTAGAAG CACAAGCATTTCGCT ACACTTGCATTAACA CCTGCTAACCATGTG TATGTGACAAATACA TTTGATTTGAT	RVLRACADQLAGVFTDIFNQSLSQSA VPTCFKRATIVPVPKKAKVTELNDYRP VALTSVIMKCFERLVKDHITSTLPDTL DPLQFAYRPNRSTDDAIAITLHTALTH LDKRNTYVRMLFIDYSSAFNTIVPSKL VIKLETLGLDPALCNWVLDFLTGRPQ VVRVGNNISTPLILNTGSPQGCVLSPL LYSLFTHDCVAMHASNSIIKFADDTT VVGLITNNDETAYREEVRALGVWCQ ENNLTLNVNKTKEMIVDFRKQQREQ PPIYIDGTVVERVESFKFLGVHITDKLK WSTHTDSVVKKAQQRLFNLRRLKKF GLSPKTLTNFYRCTIESILSGCITAWYG SCSAHNRKALQRVVRSAQRITGCKLP ALQDTYTTRCHRKAKKIIKDNNHPSH CLFTPLSSRRRGQYRCIKAGTERLKNS FYLKAIRLLNSHH			

	RTEX-1_SSa	GCF_000233375.1	Salmo salar	CTATCAGTTAGC ATGTGGAACATT CAGGGCCTAAAC TCATCAACCTTTG GACTGAAGAGTT TAGCACTGGAGT TCAACAAAAATC TTAAAGATGTTG ACGTCATCATTCT GCAGGAGACAT GGTGTAAGGCTG ACATTGTCACTC ACTGTCCCACAG GCTACAGAGAG GTCATTGTGCCA TCACAGAAACAC AGCTCTGTCAAT AGAGGCAGAGA CTCTGGAGGATT GATCATTTGGTA CAAATCCGAACT	TGACAGACCAATCAA CCTGCACATGTCCTCT ACTGTATGCTTATTGT TATTGTTGAATGTAT GGTTATTTTGACCCTT GGTTATTGTTGTTACT GTTGTCCCGTTGACA ATATTTGATTCTCATT TTTATTTATTTTTATA TTGTAAATATCCAAA ATAAGCTTTGGCAAT ATGTACATTGTTACG TCATGCCAATAAAGC GAATTGAATTGAATT GAATTGAGAGA	MCSFAQYISPPQNPHITQRRSSPPRIP ILLRGDISPPQNPHITQRRYPPLRIPILL RGDIPPSESPYYSEEIFPTLEEETCHFQ AQGNVLICGDTNARTGTLPDLTSTRG DSFITGHTVSNCLHLPHRNNSDSTVN KNGRDLLQLCRSLGLYFVNGRLRGDS LGRFTYCSPLGHSTVDYMITDIDPFSL SSFTVKPLTPLSDHSQITLFLKRTDME TTTHSQPSKLYNIRNSYRWAQNSTEE YQKATWNQNIQTLLDNFLDTTFTHS KEGINLAVKNINYIFRQTAKEAQLKLIK NKTKKITDDNWFDADCKIIRKKLRTLS NQKHRDPNNGELRLHYCETLKLYKRT LRTKKAQYNSKQLTLIEESINTNNFW QNWKKLKKSKQEELAIQNGDIWTTH FKTLYNTVQIDTNAEQRQIHEKLNGL EKAIKDNQNPLDSPITDQELYKKLQAL KFKKACGPDGILNEMLKLTSAKFQLAI LKLFNLILSVGYFPDIWNQGLITPIFKN GDKFDPNNYRGICVNSNLGKVFCSII NVLSKSQIGFIPKHRTTDHIYTLHTLID	cd01650	499	800

				ACAAAATCTAAT TGATCCCCTCAA AATTGGCAAATA TCACATTTGGTTA AAACTGAAAAAA GAACTTGTACTG ACAGAAAAAG		KHVHQNNTKIYACFIDFQKAFDSIWH TGLFYKVIESGVGGKTYDIIKSMYTGN TCSIKIGKKRTEFFNQGRGLRQGCNLS PALFNIYINELATILEKSSAPGVSLHNS EVKCLLFADDLCLLSPTAHGLQQSLDL LEQYCQTWALAVNPKNTCPREDPGP VRPVLQDLCLLSPTAHGLQQSLDLLE QYCQTWALAVNPKKTKIMIFQRRSR SQGIRPKFSIGTKYIEYCTHYNYLGLKIS STGHLNEAVNELREKARRAFYAIKKQI QIEIPIKIWLKLIEYVIEPIALYGSEVWG PLAKQDFIKWDKHPIETLHAEFCKILL HVQRKTTNNACRAELGQYPLIIKTQK RAIKFWKHLKYSDPLSYHYQALQCQE LSKEKSPLIQLVLGLSSQTCSTNTLKPQ DQNIQSIRINQITTQSKQNYIAYWET QAQTQSKMQCYLALNRQYTVANYLT MVTDQNLRKTLTKYRLSEHSLAIEKG RHRKTWLPVEERLCNHCTTAEPETEL HFLTKCQKYKTIRECHFPKFETLIQGFK			

						DLSDEDRLPVLLGEDAESCGLAAHYIA ACHKMRDSV			
	Crack-1B_SSa	GCF_000233375.1	Salmo salar	CCAAGATGGCGT AGCAGTCAGACG TCTTTGTCCTGTC GTGTCCCGTGTA TATATATTTTTAC ATCTTTTTCTTCG CATATCTTTTTAA AATATTTTCCTAA AACTCAACTTCA AAACACTCTCCT GCAACCCGCCTC ACCCAATGTGGC GTGGATCTTTTTT TTTACTAAAGTAT TTCTATTTACCTC AGATCTGGAATC CCACAACAGAAG CTAGCCAGCTAA CTAGCTACTAGC	TAGCCCATCTATAATT AGCCCATCTAACTAA CTACCTCATCCCCATA CTGTATTTATTTATTT ATTTTGCTCCTTTGCA CCCCAGTATCTCTACT TGCACATTCATCTTCT GCACATCTACCATTC CAGTGTTTAATTGCT ATATTGTAATTACTTT GCCACCATGGCCTAT TTATTGCCTTAACTTA CCTCATTTGCACACA CTGTATATAGACTTTT TGTTTTCTTTTGTTCT ACTGTATTATTGACT GTATGTTTTGTTTACT CCATGTGTAACTCTG TGTTGTTGTATGTGT	MITRLSMPLLNVNMPCPLLFWLVFIG LFHCRASSPAHYTLSNHLVPPPTHAM TSPGLNVSRDNISLIITQCLGLPPLYSH PTIPLSVHYTLKLFYRPQKPAPFTLCSG RHRRPILIAFSRTLILLLLCSSGDVEVNP GPAVPSSTPTPQALSFVDFCNRKSLG FMHVNIRSLLPKFVLFTALAHSANPD VLAVSESWLRKTTKNSEISIPNYNIFR QDRTAKGGGVAIYCRDSLQSSVLLSR SVPKQFELLLLKIHLSKNKSLTVAACYR PPSAPSCALDTICELIAPHLSSELVLLG DLNWDMLNTPAILQSKLDALNLTQII NEPTRYNPKSVNTGTLIDIILTNLPSKY TSAVFNQDLSDHCLIACIRNGSAVKR PPLITVKRSLKHFSEQAFLIDLAGVSW KDIDLIPSVEDAWLFFKSAFLTILNKH APFKKFRTRNRYSPWFSPDLTALNQH KNILWRSALASNSPRDMQLFREVRN QYTQAVRKAKASFFKQKFASCSTNSK	cd01650	646	904

				TATCAGTCAGCT	CGAACTGCTTTGCTTT	KFWDTVKSMENKSTSSQLPTALRIGN			
				AACCACGGCTAG	ATCTTGGCCAGGTCG	SVTTDKSTIIENFNKHFSTAGHAFHLA			
				CGGTCATCAGCT	CAGTTGTAAATGAGA	TPTPVNSTAPPTATRPSLPHFSFSQIQ			
				AACCTTTAGCTC	ACTTGTTCTCAACTTG	SADVLKELQNLDPYKSAGLDNLDPFF			
				GGAAAGCTATCG	CCTACCTGGTTAAAT	LKLSAEIVATPITSLFNLSFVSSEIPKD			
				CCAGTTTGTACA	AAAGGTGAAATAAA	WKAAAVIPLFKGGDTLDPNCYRPISIL			
				ACGCGACTCAGA	AAAA	PCLSKVFESQVNKQITDHFESHRTFSA			
				CCAGAGCATACC		MQSGFRAGHGCTSATLKVLNDIITAI			
				GGACCTATTATT		DKRQYCAAVFIDLAKAFDSVNHHILIG			
				ATTTTTATTCTTT		RLNSLGFSNDCLAWFTNYFSDRVQC			
				TTCTCTCCATATC		VKSEGLLSGPLAVSMGVPQGSILGPT			
				CCCGGATTTCTA		LFSVYINDVALAAGDSLIHLYADDTILY			
				CCGCAAGCTCTG		TSGPSLDTVLTNLQTSFNAIQLSFRGL			
				GACATTTTCACCT		QLLLNTSKTKCMLFNRSLPAPARPSSI			
				GGATCTTCGCAG		TTLDGSDLEYVDNYKYLGVWLDCKLS			
				CTAGCTAGCTGC		FQTHIKHLQSKVKSRIGFLFRNKASFT			
				TATCCCGAGTGA		HAAKHTLVKLTILPILDFGDVIYKIASN			
				CTACTCCTGGCT		TLLNKLDAVYHSAIRFVTKAPYTTHHC			
				AACGTTTCCGTC		DLYALVGWPSLHTRRQTHWLQVIYK			
				CCGGAGCAAGCA		TLLGKVPPYLSSLVTIAAPTCSTRSSRYI			
				CCAATTAGCCTG		SLVTPKANSSFGRLSFQFSAANDWN			

				AAGCTAGCCCGG CCAGGGCTCCTG AGCTACCACCGA AGCCCACTCCTG GGCTACAATATC CGGACCCCTTTT ACTGCCGGTACG GGGCACGGAGC CCCGCCGATCCT CCACGACTGGAC TACCGACGTTAT CTGCCCGAGGAT TCCAACTGGCCC CTCCGTCGCGAC GTAACCTGAACG CCCATTCTGCTAA CCGCGGCCTACT AGCTATCTAGAG CTACTTGGAACC CTACTAATTCCAC GACTGGTCTATC		ELQKSLKLETLISLTSFKHQLSEQLTDH CTCT			

				GACGTCACCGCA CGAAGAGGCAA AAACAGACTTAC CCCCATCGCGAC GTCCCCCAAAGG CTAACTTGCTTGC CCCGGTCTGTTA ACTGCTAGCTTG CTTGCCCCGGTC TGCTAACTGCTA GCTTGCCTGCCC CGGTCTACTAAC TGCTAGCCCCGG TCTGCCAACTGC TTGCTTGCTAACC CGGTCTGCTAAC TGCTAGCTTGCC AGCCCCGGTCTG CTAACTGCTAGC TTGTTTAGCCCC GGCCTACTAACT					

				GCTAGCTTGCTA GCATCGGCCTGC TAGCTGTCTGAA TCGCCGTGTCCC CAGCCAGCCCAA CCACTCACTGGA CCCAT					
	Rex1-3_SSa	GCF_000233375.1	Salmo salar	AG	TAATCAAATGGCTAC CCAGACTATTTGCAC CCCCCCCCCCCCACG CTGCTGCTACTCTCTG TTATTATCTATGCATA GTCACTTTAATAACTC TACCTACATGTACAT	MAPEGMAAVLSALNQPCYFVCFFAL FVTCFVHNVAATVSYDRKELLDIRTAI THLKLDKEFFFNESDGRDILQTPDQA QIPVIRWRRKRRFRGKRSGCLVRIRR RVANLPLPSVLLANVQSLENKWDELK ARISYQRDIKNCNILCFTESWLNDDIK NIQLAGYTLYRQDRTAASGKTRGGGL CIFVNNSWCTISKEVSRFCSPEVEYLM ISCRPHYLPREFSSVFFVAVYIPPQTDA GTKTALNELYTAISKQENAHPEAALLV AGDFNAGKLKSVLPNFYQHVKCATR GKKTLDHLYSTHRDAYKALPRPPFGK SDHNSILLIPAYKQKLKQEAPVTRSIKK WSDEADAKLQDCFASTDWNMFRDS	cd01650	581	833

						SDGIEEYTTSVTGFINKCIDDVVPTVT VRTYPNQKPWITGNIRTELKGRAAAF KERDSNPEAYKKSRYALRRTIKQAKR QYRTKIESYYTGSDARRMWQGLQTI TDYKGKHSRELPSDTSLPDELNNFYA RFEASNTETCMRASAVPDDCVITLSA ADVSKTFKQVNIHKAAGPDGLPGRV LRACADQLASVFTDIFNLSLSESVIPTC FKQTTIVPVPKNTKVTCLNDYRPVALT SVAMKCFERLVMAHINTIIPETLDPLQ FAYRPNRSTDDAISIALHTALSHLDKR NTYVRMLFIDYSSAFNTIVPSKLITKLR TLGLNTSLCNWILDFLTGRPQVVRVG NNTSATLILNTGAPQGCVLSPLLYSLF THDCTARHDSNTIIKFADDTTVVGLIT DNDETAYREEVRDLTVWCKDNNLSL NVIKTKEMIVDYRKRRTEHAPILIDGA VVEQVESFKFLGVHITNKLTWSKHTK TVVKRARQNLFPLRRLKRFGMGPQIL KRFYSCTIESILTGCITAWYGNCSASD RKALQRVVRTAQYITGAKLPAIQDLY			

						TRRCQRKALKIVKDSSHPSHRLFSLLP HGKRYRSAKSRSKRLLNSFYPQAIRLL NS			
	L2-2B_SSa	GCF_000233375.1	Salmo salar	GTTCCGGAACTC CAATTTGAGCAG CAGCAGCTGAAA AATGAGCTCCTA CAAATATTGAAA TTTACATGGTTTT TAGAGTGAAGTA CATCGGAAAAAA GCAAAAGAAAAC TGCAAGACTGAA TAGGAGAAAAA ACAAGCAACAAA CAAAGAAGATAG GCTTTTGAAAAA TAACTATTTTTCA TAAAATTGTAGT TATCTTAGCTAG CTGAATTGTTTAC	TAGGATAAAGTAATC CTTCTAACCCCCCCCC CCCCCTTAAAAGATT TAGATGCACTATTGT AAAGTGGTTGTTCCA CTGGATATCATAAGG TGAATGCACCAATTT GTAAGTCGCTCTGGA TAAGAGCGTCTGCTA AATGACTTAAATGTA AATGTAAATGTAA	MCLSIPVLSSLHRPYKRFTPRGRCHSN LVVPARTTHVEFQVSGSLWNCRSAA NKAEFISAYATLQSLDFLALTETWITT DNTATPTALSSSAHVFSHTPRASSQR GGGTGILISPKWTFSLSPLTHLSISSFE FHAVTVTSPFKLNILIIYRPPGSLGEFIN ELDALISSFPEDGSPLTVLGDFNLPTST FDSFLSASFFPLLSSFDLTLSPSPPTHK AGNTLDLIFTRCCSSTNLIATPLQISDH YLVSFSLLLSSNTSHSAPTRMVLRRPN LRSLSPATLSSSILSSLPSAQTFSNLSPD SASSTLLSSLSASFDFLCPLSSRPARSSP PAPWLDDSLRAHRTGLRAAERKWRK TRLPADLASFHSLLSTFSSSVSAAKATF YHSKFQASASNPRKLFATFSSLLNPPP PPPPPPSLRMTSSTILKRRLTISDPRLL SQTTPLVLLTLPYPVLEALCYLLLPPESS SPSPPSSLSADDFVNHFEKKVDDIRSS	cd01650	577	838

				CAGTTGCTAAGC AGTTGCTAGGGA CTCTTTTGGAAG AAGCTAGCTAGC TAACGAAGAACA ACAAAGAATATC TAGTTAACAGAA GAAAAGAAAGA GAAAATCAGGAC AGAAGAAAAAG GATATCAAAGTG AAACAGTTCTCT AAAGACACAGG AGACAATAATAC AAGAAACAACAC TTCTGTAGCTTGT CAACT		FAKSNDTAGPAHTALPCALTSFSPLSP DEISRLVTAGRPTTCPLDPIPSSLLQTIS GDLLPFLTSLINSSLTAGYVPSVFKRAR VAPLLKKPTLDPSDVNNYRPVSLLSFL SKTLERAVLGQLSCYLSQNDLLDPNQ SGFKTGHSTETALLCVTEALRTAKANS LSSALILLDLSAAFDTVNHQILLSTLSEL GISGAAHAWIASYLTGRSYQVAWRE SVSAPCALTTGVPQGSVLGPLLFSLYT KSLGSVISSHGLSYHCYADDTQLIFSFP PSDNQAANRISACLSDISVWMTDHQ LKLNLGKTELLFLPGKDCPFHDLAITV DNSLVSSSQSAKNLGVILDNTLSFSTN IKAVTRSCRFMLYNIRRVRPCLTQEAA QVLIQALVISRLDYCNSLLAGLPACAIK PLQLIQNAAARLVFNLPKFSHVTPLLR SLHWLPVEARIRYKTMVIAYGAVKGT APPYLQALIRPYTQTRALRSSTSGLLA PLPLRKHSSRSAQSKLFAALAPQWW NKLPHDARTAESITTFRRHLKPHLFKE YLG			

	SR2	GCF_000237925.1	Schistosoma mansoni	CCCAAATGCCCTGGTACGGCCGAGAGTGGGGAGAGTCCACTCTCCCTCTCGAAATGCTCTCACATGGCCACGCGTATATAGCCTCTGCCAGGGAAGTCCTACTCACTGCCTTCTCGTGGCGGGGGTGTTGTTTACGAAAGTGAGAGGACGAAAAGCGAATGTCCGGCGCTTTAACCGGGTTGGTGGACACGGAGGGTCCACCTAGGGGAGTTGGAAAACCCTGATTCCAAACCAATGGTGCACATGGGC	AGAGTGGTTCTAACTTTCTTAACTGAAAGAGTCTTCTGGTTGTACATTTCAGTCCGCCTTATCTTTTCATCCCTTCTCTTCCTACTTTCATTATTTTGTGTGGCGCATATGTATCTGGTGCCCTTTTGTACCAATATATATGTGTTTAAATAAATAAATAAATAAA	MLPPIMNTEAVFLKPRTPFKLASFNVRTLMQIRQQIGLAMSLEGLNVDVCCLSETRIQDSSEVLQIRSPSVASKSLFHVRLSGDPVASSSGLAGVGVALSARAEAALIDWIPINSRLCAVRLESSIKVRRNRREKRCLFVISAYAPTDCSPDAIKDEFYHQLTVLLQKARSTDIVVLAGDLNAQVGRLGTEESRLGGRWGLVGRRTDNGDRLLQLCTDHNLFLASTNFRHSHRRCATWRPPSASQAWTQIDHIAISYRWRGCVQDCRSFWSTYLESDHALVCANLTLLFSGRRIDHHQRIDVSKLAATSVASKYRAELASRLATIPPKSIDEHWLHLHDAMKMASKAACGFAKRPAYKHWVSSGSLQLMEARRSTPGDREFDHKRRLLRNEIGQSLRKDREAWWSERANEMEAAAASGNCRKLFQLIRATGSKKSGVSETIYEDDGMPITNISRRLGRWAEFFEGQFNWPAAPATSTSLSCPPWPVTTDPPNEAEVRKELQLLKRYKSPGPDDLPPALFKDGGDFLTKELTVLFAKVWEQESVPT	cd01650	549	808

				TCCAGTATCCTG ATGGAACGAATG GCGAATGAACCT GTCATTGGTCAC CGGCTACCATGG GACTGCATCTCC TCACGATGCTCC ACTGCCTTGTGG ATCAGACCTTTA GGTCAAAGGCTC GGGGTGTGGCCC CCTAAGAAAACC ACCTGCTTCGGT TTGGGCACCCGG GCAGTATCACAG CCCACACACAAA TAAATGAAATGA TGAATTCTATTG ACGATACTATTCT CGCTCATACTAG AAGCAAGACAAT		SWNESIVVPIFKKGSRCSCNNYRGISL LPIASKLLASIILRRLFKTRERLTREEQA GFRSGRGCIDHIFTLRQMLEHRHTYR RPTIVVFLDIRAAFDSLDRTVLWDCLL KKGVPEKFINILKALYTNTSGRVRAYN HLSPLFHSSSGVRQGCPISPFLFNFAI DDILETALMDVSNGGVDMLPGERLL DLEYADDIVLLCDNAQGMQSALNQL AISVRRYGMCFAPSKCKVLLQDWQD SHPVLTLDGEQIEVVEKFVYLGSYISA GGGVSDEIDARIMKARAAYANLGHL WRLRDVSLAVKGRIYNASVRAVLLYA CETWPLRVEDVRRLSVFDHRCLRRIA DIQWQHHVSNAEVRHRVFGRRDDN SIGVTILKHRLRWLGHVLRMSSQRLP RRALFADSGTGWKKRRGGQCMTW CRGMKESCKGLASVGPSRLPGWGPR DGATQWLETLSDMAQNRSQWRSC CNLLLLST			

				TTACATTATTATT ATTATTATTACCA TTATTATTATTAT TATTATTATTACT ATTATTTACTACG TCGCTAATTCACC CCCTTTATCCCCA ATTTGTGTAACCT TACTTTTCAACTT ATGCAGCAGCTC GCCCATGGTGGA AAATCTGTCCTAT CTAAACGATCTC CAGTCACTGTCT GGTTGAAAGTGA					
	LIN2_SM	GCA_002600895.1	Schmidtea mediterranea	AACGTT	TACTACCACAAGACG GGTATTTTGGTAAAT CTTGAAAAATTTCGA TTTAATTAATGTTTGA GTTAAAATTGCCATT TGACTTGTGTATGTG	MKIEIGEFAALFQRFFEQLNAKSKQEE EFRNESVPGSRVPSEDMPVFQINERF HALERSNELLNGKMDSIVNTLNYLISG LNGNFDNNNKLNCICKTNHVKEPLSI VKTNLVSELPSALLVQNKVPCLKLSEE DLIRKIPTKIVPSKKVSKKTIEHTSRVNF	cd01650	1533	1797

					ATTTATTGTGTTAATT	SLFVRVLDSNERQQFFKLFPNLFALLK			
					AATGAATTCCACCCT	SKESEKKSEVLNIELTKISKKDNPLNIA			
					CTTCCAGGGAAAGGT	GFKNFFFRNNNICVTIDFYNSSKVKSK			
					TTAACTTTGCATAAA	ITKNLKIKTSLINSKIVKNYNTNKFIPAS			
					ATTAAACTCTCTGTCT	LSIKKFLFVLSDKLAETTNISNVTVREE			
					ACACTTAGTAGCAAA	QINKIITLLKIKNSLPDYLVKSQIKYKDG			
					TTAATCTAGAATGAA	DFLMLDLLNSCEIEGELGQKITELSRS			
					GAAGCCTGATGGTAT	QIKELFNSFPEIQELFFIEKKANSGTFF			
					TTTTCCCAAATCTGCT	HRKKSIDENPEKLINLINKFKIEKTFPQ			
					GAAAACTTAGTTCTG	QISFVGGKFVKIDKLNDSQFPQVESES			
					AAGGATGGAAATAA	DIMFNLGLFSLFGLSEYDNILNSLLCQ			
					AAGCAGATAAAGATT	PKSPEKLVDIIFHLNYLNLKGNFDTST			
					CAATATTCGATAAAG	NGTYENYVQNYIVNCLLMHSILDLET			
					AAATGGACCTCAACA	CAFLNLDNIELWAPKLIELIPSSSHYSS			
					ATCAATGTCAATTAT	SCKEKLLKLKQWKLSGEIDRIHNIICPE			
					CAATAATAAATCTGG	TIIEINNFLGVIDQSYTTKSDQLKFLTSL			
					GCAGATGTCAGAAG	GEKINNIEIDNDVDDLVKAKRMLYFII			
					GCGATAATCAAATGC	RDKKRFLNKDNFNYRIPFNCRKICTLK			
					TGAATTCCAATGACC	MSDFITKLIQSNELFEEREAVVDNNID			
					GAATCTACTTTGATG	VSPNENLNILIVEPVNPTPDTVIVGQ			
					TCTTAGATTTGTGATC	MDTFPTDIDETEPVTIDKPLKKITSKQ			

					AGCTTCCACTACAAA	AVKPKAQKSKREVPASVSISVPISEET			
					GGTGCTGCATCAAAA	DADSTIINYFFPERDPKIWFTDDDIDY			
					GCCAATGGAGATTCA	YLESHIGNPNFAPVKCFIVEILSSNCKD			
					ATGCCGGATGAAAAT	EIFPMPDNLFKAQIILCPLNINESHWIL			
					GCCGATGAAAGAGA	FVYCKLTKISYFVDPILRKRHNLINKSS			
					ATCAGAGCCGAGGA	ALKITMALNNLFDLKAAVTFHPHDNL			
					CAATTTAGAGAAAAT	VYQDNNYDCGPYICAYAMLISQEWQ			
					TCGATTGCAACTGCG	TFPNNFIETIRKNVHRFQTSKFQNPPK			
					TCAAGAGTTCCTCAA	VSGGGITTLPIFKRFQNVTRPPKVALP			
					CAATAAGATGAAATA	YKASLQNENSCVPSLIKEWFSNKNYIT			
					TCGAAACTCAATTGG	VLSPELTLNLIAHNNQFIRENVTFRQF			
					ATCTTTTGCTTTTGGA	TRVKYVIAVIPNNPHWCLYIFSPVFEH			
					GAGTCTCAGAACGGC	SYILNFQGRVLNERLVSIGESMTDYLN			
					AAGCTCCATGTAGAG	KFFISTAKVNFIVNHNQTHFSFRSEDP			
					CTGTGTCCGATATGA	HFFASAFIEFLERFFFGNILSFTKINETL			
					AAAATTCAAAATTGT	KSFKLQITNEICKNLNLINFSINDEIITK			
					TATATTAATAAATTTT	YFKNLNLPSNFIFLNSAMCTAIMDDC			
					TACTATAGTACTACT	TRYLNDFLNYDNIIKAQVIFAIVAPPKF			
					GAGTGATTTTGTTAA	REILFIIDYNTDEYYLLNPTTLLGNDNY			
					GTCTACGACTTAGTC	LSTCKILRNKINEIRESPNRVLNLGECP			
					GTAGACGTAGTTGTA	HDIRGSGLVSKILFCSYMSNYANDLPL			

					GGCGTTGTTGGGCGC	VGLNLRQIGNVVSSFLPIVSEIKKVKSL			
					TGATGGGAGCTGAT	ENNIINQNHKESENLKERKNKIDQLLI			
					GGGCGATGATGGGT	KVFKMNVDEIVDVILDQISIKTVPTHK			
					GCGGATGGGCGATG	PYLGNKERVESQINKSNLIKNFNSDM			
					ATGGGTGCTGATGG	KTTVNTIINDVDVDVRPKYENIITNFE			
					GTGCGGATGGGTGC	HKIPLDRTWDKVFKFCNRSDKKLKLK			
					GGATGGGTGCGGAT	YLSNSEVLFELKKADNTSPGVDGIKYR			
					CGGTGCGGATCGGT	DLVMLDPEGKLLTFLFNKIISKKVIPDS			
					GCGGATGGGTGCGG	WKSFKTLLIPKPDKGDKYDNVSSWRP			
					ATGGGTATGGATGTG	IALLSVIYKLFASCLANRLTFWINKNNL			
					TGTGGATGGGTATG	LHIGQKGGSRHDGCVEHNSILSSALE			
					GATAGGTACGGATG	HSKYSKNSPLAIAWLDIKDAFGSVPH			
					AGTGTGGTTCTAGAA	AYMWSLLRFVGVDEEFVTILERMYS			
					TGTAGGTAGGTGTG	DTSTFYSCGPILTPNIAIRQGVKQGCPI			
					GAAGTTGGTAGGTGT	SMILFALAINPVLEAVSRSKCEPFMIG			
					GTGGAAGGTTCTAGA	ESAVKILAYADDIALVSKSAYDLQRVT			
					AGGTAGGTGTGTGG	NIAVAIASEIGFEYRPEKCGYIELPLIET			
					AAGGTTCTAGAAGGT	ESQILINDIKIKKLASSEFYQYLGVPVG			
					AGGTGTGTGGAATTT	EETDQSPYDILKKVVSDTRKIADSDLL			
					GGTAGGTGTGTTGGT	GWQKLKAYKIFLHSRLVFPFRTREIKT			
					CCACAACTACGTGAC	SALAKSNANNNLTVNITSQLRNCFRK			

					ATAATAACAGTTAGG TAATGTGTTGGTCCA CGTGACATAATAAAG TTAGGTAATAAGTGT GTCTCCGAGTAGTGA CTGACCAGTAAGTAC GTGACATAACAATGG TGACATAATAAGTGT GTCTCCGTGTAGTGA CTGACCAGTGGTCAA GTGATGTAATCATGA GGTCACTGACCAAAG GTCAATTACTAGGTA GCAACTGAGTGGAC AAAGGTCAATCTCCA GGTAGCAACAGCGC GTGCAGAGGTCAAAT GTTACGTAATAATTA CGTAAGCCAAGGGG GGTTGTATCTGACGG	MMSLPNHSEVCYFYNSTENGGANC VDLLDEYHTQTITHFFRLFTSSCEFAR KVNLDSLNFVAGPRLGIINPTLQERFD WINGKEPKPRHSGKKTRFQRARIAIA FFKKEHDINIMFIVTKGQPSLFITSEKR GTFILTPDLRKTTSKVLHMALCDSYLS KWEKSCTSNFIASAIKLSPKINKAIYRG DISEFAWHFIHRARTNTLSINAKPHN KGNKRLCRLCHTEDETMSHIIQSCKIH LTLGSERHNDCLDLISSHLSKNSNLIVV VDHVCSLVPESKERVDLMITDMVRK KIFLVDMKCPCDTINNFALVDLLNLQ KYDSLKMQIQDAKPDFKVELDTCIIGS LGSFPPKTPELLLKIGVDRTHLKGLLKD CALSNISHSARRWH			

					GCTCACATCCTACTACT				
	Tx1-2_SPur	GCF_000002235.5	Strongylocentrotus purpuratus	CAATCCCTTGATATGTGTATTGTGGAAAAGAACTGTTGTGAGAACTTTTGTTTTTGCATTAACGAATTCTCCAAACGTCGAAAATTCAGGAATGTAAAGTGAAGAAAAAAATTCCCCGCCATATCGTACTTGGAGGTATAGATTTTAAGACTCAAGAATGTAAGCCTTGTGACATTCCATGAGATGTAATCTGGTTGGAAGTACCATGCAATTTAAAGCACCAGTCA	TAATGGATGTGGGTATTATTATTATTATTAGTATTATTAATTCTCTCTCGTGTACAGGTGGACAAACCGGAACAGGTTGCATGCAGTGTGTGTCGGCGTGAAGGGGGTATGATGAGATGGGTGGGCTAGGGGGGGGGGGGGGTAAGAGTTTGTGTGGTGTGAGTTGAGGGAACAGGTTGCATGCAGTGTGTGTTTGTGTTTGTATGGTTGTTTATGTGTGTGAAAGGGGGTGTGGTGTGGGTGGGCTAGAGGGGAGGGGGGGGGGTGTGAGAGT	MSVNLKLLSINVRGLRDRVKRKKIFQWCKMSANMIFLQETFSTEDIEYSWKGDWEGSMFFSHGSNHSRGVIVMFKPGITFVVEKTLIDEGGRYVIIKGKCQGEKLILGSVYFPTRNKENAQCDFLSDLNKIICSVFSQEYSLIIGGDFNVIMNHNLDYMGASNISKSKYRDSLLDFLDKLDLIDVWRKRHPNKKEFTFRQKQPLVQSRLDYIFISSKVDKLIRKSEILPSITPDHAGVRIEFNNLVDSFVYGKSYWKFNSSLCNEKDFVDGVKKKISEIEETWRNQITSKIKFWDFMKMKLREYIMKFSRERAKLRRVHIAKLEKEIKDLEYRLTINIDCHKSIQDEIDMKKMELEKLFDYSRQGLKVRSRTEWSEEGEKNSKFFEQLLKSNKRKSVIKEIYNENKETVRDGKGILKIIREFYENLYACKNGEITEDIENMFLNELPKLSKENKEYCEGKITLAECYQAVKEMKWNKSPGNDGFSTEFYFTFWPILGKTIVEAFN	cd01650	520	774

				AAACTCATGAAT	GTTGTGATAGGATGG	ESFETKMLSNSQRQGVITLIEKEGKDP			
				GTAAAGAGAAA	TAATATGTAATTTTAA	LYMKNYRPITLLNVDYKILSKVLATRV			
				ATAGTACCCGCC	ACCTTCCTCTACTGTT	KEVLEDIIHYDQVGYVKDRNIGEAVRL			
				AATTTGTACTAAT	ATTTTGATTATTCACT	IDDTFFQSLYQPVGFLIAVDFEKAFDC			
				ATAGGCATAGGT	GTTACATGCAAGACT	ISHKLLFKALEQFGFGKVLCSWVKTLY			
				TTGCCTTATTGCA	TTTTTTTCCCCTTTTTT	TNVSSCVMNGGHSTGYFGINRGVR			
				TTCCATGAGATG	TTTCTCTATGTTCATT	QGDPLSPYLFVIVIELLARYIRKDESIKG			
				TCTAGCTTGTTAC	GATTGATGTAACTTA	VGFGNKEIKQVLYADDITLFLKDMDSI			
				TTAATCTTTAAAA	TGAAATGTCTTGTAT	NKVKKVFKNFEKLSGLKINMDKTNF			
				TGTGTCATTAGA	CAAAGATC	MWLGEESDRPGLPFFGHRVQHIKIL			
				GGAATGTGAACA		GVYFARHSKIKDDLNYKEILSKIKRLIG			
				CTGTACAACGCA		WWKQRDLTMFGKIKLLKIYVFSKLNY			
				AATTATAGAACG		VSSLMTVPQNIFEEIEKISFNFIWGGR			
				GGTTCCTTTTAAA		DRIKRKILYQDYCAGGLRVTNFKIFIKT			
				GATATTTACAAA		QRIMWIKRLLYGEKYSGWKVTFDHF			
				TGGAGAGGGGG		FRSLGGRIIFLCNYDINKMDLKGIPSFY			
				AGTGGTTTGGTG		KDMLKAWQELDKCRHFGENKNNPII			
				TGCATGTGTGTG		FSNKCICLSNKTFFDVELFQKGILQAS			
				TGTGTGTGTAAC		DIIVRGHLKPILHFLNLGFETKSLLRIQK			
				TTACATACATAC		IFSAIRNDWIQSDFMEVDTENFEIEL			
				ATGTACATCGCA		MIFGRLQKLCDIKSRRLYDYFIEDLQK			

				TTTCTCTTTCCTTT TCCATTTGGTGA GCAAAGTAATCC AACTAATCAATA GCCATGAACACG TGAGAGGCTTCC TCAGCTATGATT GGTCCGCGAGA GCTGGAGGGCG AAAGTGTTTCTTT TGTTCTTTTCGTC ACCGGCTTCTCG TGGTCACCAACG CCGGACATGCTG ATTGGCTGAATG GTCTTTGCGGCA GCAGGAGGGCG GGACGTGTTTTG AGACAAGTCTCG TCTCCAGCTGGT CCTAGACATTTC		EYTLHVTDGNCQFDLGMTELKETFTR MKCAMLCNKHREFQYKQLQGAVYT KRELFKFGFESDHFCSFCRLEVETYQH LFLDCLKVKRLWQEVIQKLHLIEIDIDD WKPIFMGISGNTARIFVINCIIFLVKYII FISRKEGTLPSVQKTVHMLKNYRDEE YKIAFKKGKLGMHLQKWEQIDAFLL			

				AGTTGATCGAGA TAAATGGCATCT GAAATATACAAC ATAATTGTCGAT TTGTTTCTGCTCT CCCCCCCCCCCCC TTTAATAGTAAA TATCCATGTAATC TTACAATCTTAGC AGGAGGATTAAC GTTAGTGCCTTG TGAAGCAAGCAT AATTAGATCTCG TCACCGGCTGCT TGTGGTCACTGT CTATGACGGCAG GTGGGCGGGAC GTTGTTCTTTTAT TTAGAAAAGTCT CGTCCCCAGCCA GTTTTAGACACC					

				AGCGAGGTCTTT AAATCAACTATA AATTTTATATCTA GACTCTGTTTTG AGAGAAATTAAA AGGGGAAAATA AATAATGCCCTTT TACGGTTACCTT AAGCCGGTCGTT AACAGTGAAAAA AAGAAAGAAAA AAAGAAGAGGA ATAGTTCAAAAG TTAATTAATTCGT TTTCTTAAAATAA AAATAAAATAAA AGATGGGAGGG GAAGAAACTCCT TATATACTGTAA AGTTTATAATGG TATCAAAAGCCA					

				TTTGGTGGGGTA CAAGCTGTTGAA GCTTAAAGTATG CACGGGGTTTAT TTATTACGTTCGT TTATTACGTTTTT TTTCTTTTAAAAT ATTATATATAGC CGTTTGAGCCTG GGGCTAAAGTAT CATTAGTTACAC ATTATCTGAATCC GTTAATTATGAG GTCAAGTTATAT CGTCTGGAGTTA TGATAAATTGCT CTGTTTTATTTTC GGTGTTGGTTTT TTTTTTTTTTTTT TTTGTCACAGTA AGAAGACTGTAT					

				TAATACTTTGTGT AAAAGGCTGCCA CGCCATGATTTA GAATAAGTGATG GGAGGGGTTTCG AACATGCCAGCA TATATTATGCTTA TGTGTATGAATG TGGTTGATAGTG TTGAAGGGTGG GGTGATGGTATG TGTGGTATGTTTT CCGGTTCCGATG TCTGCCTGTACTT GTTTAATTTGCA ATCTAACTTGAA AGCGTTTGTTTTT CCTAGAGTTGGA AATGCATCAAGC TTCCCATCTACTG TGAAACTGTCAC					

				CACTATGTACATT TCTAAAGGTTCTT TGTTACAGGTAT AGATTCATAGTT CTTG					
	Rex1-2_OL	GCA_004348135.1	Oryzias latipes		TGACTGGAGTCTGGA AAAACAAACTGCACC TTTGCTCATCCTAATT GTTCGCTACTGTCATT GTCACTTTTGCACATT TGCACTTTTAGAACT GCCACTCGACATCTT GTGTTTTTGTGTGTAT GTT	MQVNKALRRINTRKATGPDNIPGRA LRACANELTDILTSIFNLSLSQRVVPSC FKTTTIVPLPKKSPPTCLNDYRPVALTP IIMKCFERVMLTHIQSSIPDTLDPLQY AYRPNRSTSDAIAAALHTSLSHLENKD SYIRMLFIDYSSAFNTVVPHKLTHKLF SLGIHPIICDWLLDFLTGRPQSVRIGN RTSASIITNTGTPQGCVLSPILYTLFTH DCVASHKDNIILKFADDTAVIGRITGG DEAAYRREVASLVTWCEDNNLTLNT DKTKEMIVDMRKERRNPHQPLLIRGL EVERVSSFRYLGVHLSQDLTWTLNTS QLVKKAHQRLYFLRRLRRFGMSPKILS NFYSCMVESILTSCITVWYGSSTAAD RKRLQRVVKTASKITRTPLPSLQSIYKR	cd01650	55	309

						RVCRRAASIIKDPTHPQRGLFTLLPSG RRYRSVKCRTTRLKNSFFPSAIRLLNS			
	RTE-1_DaCa	GCF_001625215.1	Daucus carota	CATGAGGTGATA GTCATGAATCCA AATATAATAGAT AAATAAGATATT TTTTTAAGGTGC AGTTCTGCTCCCT AACTAATAGTAG AAATAAAGTAAT GGTTATAAGGGT TCTCTAAAACAT GACCCAAGTTAA TAAATGTCTCTA ACCAAAAAATGT CTACATTTTTAAT GTGTTTTTTATAC GCAATAAATTAT AACTAATTGAAC ATCACTAAAATT ATAGACACAATT	TGAGGGTTATGGTGT ATGTTTAAAGGGAG GGTTCAGCTCAGTCA TTACCAGATGCATTA GCGCCAATCAAAAAT GTTTGGGTAAGTCCC ATGTGAAGAAGGCC ATGGGCTAAGGTATT ATATTCTTATTATAGC ACACACTATTTTATTT GTTGTGCACTAAGCC GGGGGTCTTCACGG AAACAGCCTCTCTAC TTTAAGGGTAGGGG CAAGGCTGCGTACAT CTTACCCTCCTCAGAC GCTGCTCACTGAGTA TGTTTGGTTTGGTTT GGTT	MANTIESMTFWGCLDDLVRSITLDQ FLFVGGDFNGHIGARADGYQGVHG GFGYGVRNDNGSTLLEFATAHDLVIV NSCFRKRDDNLITFRSGGHATQIDYLL IRSRDFRFCSYCKVFPVEACASQHRLL VMDLSLDGPPMEGLRVVTPRILWRN LHGSKVAEFKDMLEGFFRLEDDVDQ MWIRMADLIRGVARRTLGVASGNV KEQRESWWWNDDVQAKVQFKKGC FLEFMSCPEGPDRHIKRELFKLAKRM AKQAVAEAKTKAYQDMYRRLGTKEG VNEIFKLAKARNKRSQDIGAVRYIKDE GDRVLLYDGDIIARWGRYFSELFNAA RGREIVFEQENVNVSHNDVGMSQDI TMAEVRVALGMMGKGKTVGLDEIPI EVWQCLGEQGVRWLTALFNVIFRTS RMPSEWRLSVVVPIYKNKGDAQSCS NYRGIKLLSHTMKLWERVIECRIRRIV TISVNQFGFMPGRSTIEAIYLIRCLME	cd01650	413	662

				GGGATAAAAAGT TAGGCACATTTG AAAACTTTACAA ATATAAGGACAA ATCCGACAGTGG CTTAAAAATCCA AATTTGGACCAA TTCTAGCCTAAAT CTTCTCAACGGG GGCCTGAATTAC GGTCCAAATTTA GGCCACTTAG		KYRECRKDLHMVFIDLEKAYDSVPRV VLWRCLAARGVPEVYIRTIQDMYSA VGTCVRTPVGDTQYFPVEVGLHQGS VLSPLLFIIVLDVITRDIQAPVPWCML FADDIVLIAESRSEVNVKLEQWHMSL EGYGLRLSRSKTEYLWANFSEEIHEAD VAVCIAEGRVPQTNTFKYLGSIIQSNG DISADVTHRISTGWLRWWAATGVLC DKNVPLKLKGKFYRVAIRPSLLYGSEC WPLRKAQERRLETAEMRMLRRICGN TMTDHIPNDTFRRLLGVESISKKIREG RLRWYGHVRRKCNSAPVRRVEHISV RGKRKRGRPRRTWADQLSLDMGAL NLTGDMTLDKNDWRRRIRVVETRSR GSQRL			
	Penelope-2_LG	GCF_000327385.1	Lottia gigantea	TACGATATATAG ACACAGCTTTAT GGAACAACAATC ACTGGTTAAAGT TTTATGCAACGTT TCGACGTGTTTC	TAATTATTCATATTCA AACAATTTCCAACTTT AAGCATAGTTAGTTA TATGAACAATGTAAT TTCTATTAATAAAAG ACTCCAATACCAGCA	MRSTMLSNEFLTKSVINLSQKNLTDE QINLLSRGLKFSPVTKLKNHDDFIVNI ERGLHQLAPGGHINYLRHQVTEILAK NKIQPNNLTKVQRDALTDLRQDNKIA ILPADKGKATVIMDKKEYQETVTELLS DAKTYRRLRKNPTSIYKDRLDKKLNEL	cd00304	316	397

				CCCCGTCGTCCTC	ATAGTTGTTTATTTAA	ESSNQITNQLKFKLRSSRHPRSPYARA			
				AGGCACATAATG	TCTTCAATTAATCTTT	TIKIHKNPIKARLLICSRDSVFYNIAQH			
				AGAAGAAGTGAT	TTTAATTCATTTACTT	LSRILAPLGKTADSFIKDSEEFVTKLKEL			
				GGGCAAACTATA	TATTTCAATTATTTTA	QITGKMYSYDVVDLFTNVPCKDVLSII			
				TATACATTAATG	CTTAATCAGCCTGTA	RDKISKLETPLDTTLTTDSIVDLIQMC			
				CTTATTATTAAGA	TGTTATTTTAAGAATT	MSSTYFTWGDEIYEQIHGLPMGSPLS			
				ACACACAAGAGA	ATATACCTCTTATTGT	PLITEIYMTAFEEKTMETSPISPICWFR			
				TAAATGGTTTATT	ACCTGTTAAAATTTA	KVDDTFVMLRAEDDPSALLLNAQHK			
				GATACATTACAT	ATGTAATGTATCAAT	RINFTMEAEEDDSIPFLDVRIKRLGGQ			
				TAAATTTTAACA	AAACCATTTATCTCTT	LKTCVYRKPTHTDQYVNYHSNTPQH			
				GGTACAATAAGA	GTGTGTTCTTAATAA	VKRAAITALVRRAKRICDPEYFQQEL			
				GGTATATAATTC	TAAGCATTAATGTAT	QHIKKVFCLHNNYPPAFVHQAFRSVL			
				TTAAAATAACAT	ATATAGTTTGCCCAT	MAKPTNQQPNPNRPSTRLAIPFIGKA			
				ACAGGCTGATTA	CACTTCTTCTCATTAT	SYQIKKILKKSVGIETVFSGNSTLKDLL			
				AGTAAAATAATT	GTGCCTGAGGACGA	RASGKLNPSNNQKPKGCVYQVECSC			
				GAAATAAAGTAA	CGGGGGAAACACGT	GSSYIGETKRPLDVRLKDHQDKPGNE			
				ATGAATTAAAAA	CGAAACGTTGCATAA	GTSAVAEHLAMGRKTGHKVNFDNT			
				AGATTAATTGAA	AACTTTAACCAGTGA	NILTINKHIDKTRRLHEAIHIRRYQPDL			
				GATTAAATAAAC	TTGTTGTTCCATAAA	NRDEGVRLSHAYDKLINL			
				AACTATTGCTGG	GCTGTGTCTATATAT				
				TATTGGAGTCTTT	CGTA				

				TATTAATAGAAA TTACATTGTTCAT ATAACTAACTAT GCTTAAAGTTGG AAATTGTTTATTA AATACATTTATG AATATGATGATA AGTTTATCATAT GCATGCGATAAA CGAACACCTTCA TCTCTGTTAAGG TCAGGCTGATAT CTTCTGATGTGG ATAGCTTCATGA AGTCTCCTGGTTT TATCTATGTGCTT ATTGATTGTCAG TATGTTGGTGTT GTCAAAGTTGAC TTTATGTCCTGTT TTCCTTCCCATAG					

				CAAGATGTTCAG CCACAGCAGATG TTCCTTCATTGCC GGGCTTGTCTTG ATGGTCTTTGAG TCGAACATCAAG GGGTCTCTTTGTT TCCCCTATATAA GAACTTCCACAG GAGCACTCAACT TGGTAGACACAT CCCTTAGGTTTTT GGTTGTTGGAGG GATTCAATTTACC ACTGGCTCTGAG CAAATCTTTAAG CGTTGAGTTGCC AGAGAACACTGT TTCAATGCCGAC TGACTTTTTCAG GATCTTCTTGATT					

				TGGTATGATGCC TTTCCGATGAAT GGGATAGCTAAG CGGGTTGATGGT CTGTTGGGATTA GGCTGTTGGTTG GTTGGTTTTGCC ATTAATACGGAT CGAAAAGCTTGA TGTACAAAGGCT GGAGGGTAGTT GTTGTGCAAGCA AAATACTTTCTTG ATGTGTTGTAAT TCTTGTTGAAAG TATTCTGGGTCA CAAATCCTTTTTG CTCTTCTGACCA GCGCTGTGATGG CTGCCCTTTTGAC ATGTTGAGGGGT					

				GTTGGAATGGTA GTTCACATACTG ATCTGTATGTGT AGGTTTCCTATA CACGCATGTCTT CAACTGACCGCC TAACCGCTTAATT CTCACATCCAGA AAGGGTATAGA GTCATCTTCTTCT GCTTCCATGGTG AAATTGATCCTCT TATGCTGAGCAT TAAGATGCTGAA GTAGAGCTGATG GGTCATCTTCAG CTCTAAGCATGA CAAAGGTGTCAT CAACTTTCCGGA ACCAGCAGATTG GTGAGATCGGA					

				GAGGTTTCCATG GTCTTTTCCTCAA AGGCTGTCATAT AGATTTCAGTAA TCAATGGTGACA AGGGGCTTCCCA TAGGTAGACCGT GGATTTGTTCAT AGATCTCATCTCC CCATGTGAAGTA GGTAGAGGACAT ACACATCTGA					
	Crack-23_SP	GCF_000002235.5	Strongyloce ntrotus purpuratus	TCCGTTTCTGCTT AATAAATGTGTT AATATATTTCCA GAGTTTTGTCTG GGGCTGTCGTAT TATCAAGCAATA TGCTTACAACGA CAGCCCTTCTTCC ACATATAATTTTA	GTCTGGAACGCTCTT GTTCATTGCCGTGCA ATATGAATGCTGCAA TCTCCTTTAACATTTT ATTTTAATTAGCACA CGCCAATAATGAATC TTGCTTACTTATTTTA TCAATATGCTAAACA ACATTATCATTGATG	MIPELCSTGLKICHININSLPHKMDEL KILLGNDPFDIVAISETHCDSTITDLDIS LDNFCVTRKDRSRHGGGIALYVKNNI SYEVISQLDNELEVIFVKVKQKQTHPL YLGVVYRPPNSPNEFFDKLSEIIHSISE NAGEIIIIGDFNCDLLPEICTPFSRALLS LMDTNLLHQLISVPTRVTPHTKSLIDH VFSSIPEEHQHTGVVKTHISDHYLIFT VFGYARAKQKNDDNIIEYRCFRHFSN	cd01650	516	775

				TAATGTGTCATA	TTTTTTTTTTTAAACC	DNFVTMLQNLPFNENLNDKDPNHA			
				ATGTTTGATAAT	TCAATCAAGGAATCT	WQCWVDLLISAVNKHAPVKSKRIRS			
				CCATGAATGCAG	TTGAAATTGTATTTCA	KLCPWITGDLIEAMNRRDWLYTMAT			
				AAGAAATCAAAT	TGTACTTAAGCAATA	KPSNPDSSEFWSEYKSLRNEITCMIR			
				CAAAATCAAATC	CCATGTTATGAATTG	NAKNDYINALLHENGGTPNAMWKT			
				AAAATTCATGTA	ATTATGATACTAACA	LKLISSSKKESNLKLEVDGEEVIDAEKI			
				TATTGCATACTTT	TATGCATGTTATTATA	ANHFNNYFVDAVDTMFNPLNFEND			
				AACTTCTGGCCA	TTTATGTGCATAATAT	VNNSTIDRASEPSVHSDTVDTVPPVN			
				ATCAGCGTAGTT	GTCAACTATTTGTTTT	SFDLPVITDEFVMKEIERIPIKKATGPD			
				GGCCTTTTATTGT	TATAATTAATAGTTTA	DVSIKLLKFACNSQNVIQSLTYILNSSL			
				CAACGATAACAT	AAGTATTAAATGGGA	DQGIFFDEWKIARVQPIYKAGSKLSV			
				CATCCCTGGACT	AAGTATTAATAGTAT	ENYRPVSLLSIPSKILEKAVNTEFQEYL			
				ACATTTGGCGGC	ATATGGTATAACGTA	KVNKILTERQFGFRPNHSCETALLCM			
				AGATGGCGTGCA	ATGGTGACGGCAAA	VDLWAQNVDKGNLNGVAYIDMRKA			
				TCTTAACAACCAT	CTTTTTTAACATTGAA	FDAVNHSTLLMKLKNVGCTDRAIKW			
				GGAATACGCGTA	ATATATTGCAATAAT	FRSYLGGRSQFVSIKGKKSSTRTINYG			
				ATGGCGCGAAAC	TTGTTTTGATATTTAT	VPQGSVLAPLLFSIFINDLPTSIHSAD			
				CTCATCACTCGTC	ACCATACTTTGTTTTA	MFLFADDATMSVQGKTLSEIQTKLNL			
				TCCGAGATCAGT	ACTTTAACCTAATACT	ALDEVYSWTQKNKLLLNTTKTKVMVI			
				CAAGCGTCACTA	ATGTGTTATTGCGTA	GSKQKLQTLNENNLNVRIKTTPIECVT			
				GAAGTAGCCCGC	CAAATGTCAACTACA	QYKCLGVIVDSNLLFSKHVEKVALQM			

				CGTCTGACCCAC	TGTATATATATATATA	KQKLGILRRLKGTFNKHYLSMIYWGYI			
				TGCCCAATCAGC	TTTCAATTGTGGGTTT	LPHAMYCCTVWTNRSLHNYETINKL			
				GCCCGCGATCTC	TTGTTGTTGTTTGTTT	HKRAAYIISGCSWSTPSHEVLLQLNW			
				CGAATACAATGA	ACAATAGTTTTAATT	STLEALYSRAMGIMAFKCVHHLAPAL			
				GGACCGATCCAT	ATAAACTGGGAAAGT	LANKFILHDDVAQRMTRNSHELKLES			
				TCTCGAATCGGC	ATTTAATTGTATACAT	NRCNTEFFKRSFVNSAITLWNNLPIET			
				AAGACTCCGAAA	TTTATAATTACTCAAT	RFSPNLNVFKSNLSH			
				GCTCCTCGCACG	GGTGACGGTAAACTT				
				CCATGGCTGCTG	TTTTTAACATTGAAAT				
				AGAATGATAACT	ATTCGCAACAATTTT				
				CAGCTTTTCGGA	ATAAATATCTATAGC				
				GACGCCAACATC	ATACCTTATCTTAATT				
				GCAAATTCCGCA	TGAACTTCGTATACT				
				AAAAGCCGAAG	ATGAAATGTACTATA				
				GGGAGAGTTTTC	TCGCCTTTTTATTGCA				
				CCCCGAGACTGG	TGTCCGAATTATCTCT				
				CTCGACTGTCTCC	CATATTATAATGTGC				
				AGACGGCCCGCA	ATTTATTGTTGTTTTG				
				CGCTCCTGAATC	ATTGACATTATTATTG				
				GTCAAGCTCAGC	AACCATGTATGTTGT				
				AATAGATGATAG	AAAGGGCCCTGATG				

				CGATAATTTAAG TCTTCACGCGGA GGATCTTAATGT AAACGCTACAAA TGTAATAAACAT CAATGAGAAACA AACGCAAATTGA AAAAACTACAGC TTCCAAGGATGA AAGTGAACTAAA TGAAGGACACAA TGTGACAAATAC ACAAGGATATTA TGAACTTGTTGA TGGC	AAAACCAGCTCATGC TGATCGGGCTACCCT TTTTAAAGATGTGAA ATAAATAAATAAATA AATAAA				
	HEROTn	GCA_000180735.1	Tetraodon nigroviridis	AGATTGGTCTGG CTAAGCCAGTGA CGTCCAGGAACA GACTGGCTGACG ACCACGAATAGA GTGGTGACAGCT	TGATCACTCCCAGTC GGGTCGCCTGGGTG AGGGGGTCTGATGTT GAAAGACCCGAAACC CCCGATGACCCCAGG TACTATCACTGACGA	MATTQASVKPTAVATCVCGKICKNP RGLKIHQTKMGCLASVQPEQRARFSL SESREVPARAEPYGPQQPHSPEALGE TQEERGQESPHSAQNLRAQVAQAP DNPQHHRRVKWPPASKVSEWQQL DEDLEGILESTAKGGVDRKLQTMTTL	cd01650	362	627

				TGGATAGACAGC TGACAGCAGGG AAAGACGGCAAC CGGGGCAGGAA GGGCTAGCAACC CAGCCTGCATCT TCCGTGAGGAAG AACCCAAAACTT GCTACGAAGAGC CCGAAGCAAAGA TACCCCCAGGGG AGCCCGAGAGG GGGGGAGAATG AGCTCCCCAAAC GGACGGATAAC	TGTGTCCAAGACATG CATCAATAGGTGTAT TTAGAAATC	VISFATERYGTMEKRAAPEKYTKNRR AEKISQLRQELRVLKKQFKGASEDQK PGLAELRCTLRKKLLTLRRAEWHRRR AKERAKKRAAFLANPFGFTKQLLGQK RSGHLECAKEEVDSYLHDTFSDAERE NSLGECRVLISPPEPACSFNTKAPTW KEIQTVVRAARNNSAPGPNGVPYLV YKRCPKLLARLWKILRVIWRRGKVAH QWRWAEGVWVPKEEKSTLIEQFRTI SLLNVEGKIFFSILSHRLSDFLLKNQYID SSVQKGGIPGVPGCLEHCGVVTQLIR EAREGRGSLAVLWLDLANAYGSIPHK LVEMALARHHVPGPIKTLIMDYYDSF HLRVTSGSVTSEWHRLEKGIITGCTIS VIIFALAMNMLAKSAEPECRGPITKSG IRQPPIRAFMDDLTVTTTSVPGCRWI LQGLERLMTWARMRFKPGKSRSLVL KAGKVTDRFRFYLGGTQIPSVSEKPV KSLGKMFDGSLKDAASIRETNDQLG HWLTLVDKSGLPGKFKAWVYQHGIL PRILWPLLVYEFPISTVEGLERRVSSCL			

						RRWLGLPRSLSSNALYGNNNKLTLPF SSLAEEFMVTRAREVLQYRESKDPKV ALAGIEVRTGRRWRAQEAVDQAESR LHHKELVGAVATGRAGLGTTPTTHLS RLKGKERRDQVQLEVRASIEEQRASQ WVGLRQQGAWTRWEEAMARKISW PELWRAEPLRIRFLIQSVYDVLPSPSN LFLWGKVESPSCPLCQGRGTLEHILSS CPKALGEGRYRWRHDQVLKAIAESIS SAMEYSKRLPLPGRGVRFVRAGEQP PPQPRAQPGLLATARDWQLRVDLGK QLKFPENIVETNLRPDIVLHSQSSKQV ILLELTVPWEERMEEAYERKAGKYAE LVEDCRRAGWRSRCLPIEVGGRGFA GKSLCKAFSLLGITGMRRRKAICAASE AAERASRWLWIQRDKPWTSASWTQ AGN			
	CR1-65_AAe	GCF_002204515.2	Aedes aegypti	C	TGACTGTATGTTTAG CTTAAGTTATTTATTT ATACATTCATTAAGA CAACTTCGTGTCAGA	MNDHLSANMHRTLSHSGRTTSSTM EAPTPPSTVESLQPALISRPDPVVGRG GGVFQPVLSGKYNKITDSLLPDISPPS SLRIYYQNIRGLRSKIDSFFLAISELDYD	cd01650	586	838

					TGGATTAAATCAGAA ATACAAATACAAAT	VIVLSETWLDDCILSSQLFGSNYSVFR TDRSALNSDKGRGGGVLIAVSSRISCS VDSAPVSNQLEQLWVKIVLPRNNVSI GVIYLPPDKKNDIDTIQRHIESIGAVHS HLHENDMAMLFGDYNLSDVHWTS DANDNVFVDLTRTRLNPASSTLLDGF CFHGLSQVNTIVNSGGRTLDLVLLND VVLPNHSIHEAVEALTTVDVHHPPIEL VVSCPLPVVYETVSDNTGFDFKKANF ESLNVALRSMDWDRLDSIEDVNEAV AFFTNSVMQAVAEHVPPRRPPSKPP WSNSRLRLLKRRRSAALRFYRSHRSQ QSKLHFNRSSSEYRSYNAYLFAQYKSR TEQNLRSNPKQFWSFVNTKRNEDGL PTSMYLDELSADSASDKCELFAMQFK RAFNDYASPASQIDRALQDTPQDVFS YDMFHVSEEEVIGAIGKLKFSYNPGP DGIPPALLKRCSSTLLAPLTKLLNHSLR QQVFPACWKKSFLFPIHKKGDKRCVS NYRGITSLCACSKVFEIIVNDSLFESCK RYISTDQHGFFPKRSVSSNLVEFSTLCI			

						RAIDAGKQVDAVYLDLKAAFDRVDH RILLQKLRKCGVAINFVDWFRSYLTER SLCVRIGESESASFTNISGVPQGSNLG PLLFTLFINDVSLLLPTGNRIFYADDAK IYMIVESLDDCYRLQHLLNLFEAWCT RNCLTLSIEKCQVISYGRKRSPIRFPYE LSGTTLERVKCVRDLGVTLDEKLTFNC HLNDVISRANKQLGFVLKVSSGFKDP LTLKALYCALVRPILEFAAIVWCPFQN YWISRIESVQRKFVRHALKNLPWRDP ENLPPYHERCRLLGIDTLENRRHVAQ TMFVAKVLRNEIDSPSLLADVNIYAPE RTLRRRQFISLGSRNTRYGQHDPVRF MALKFNEAYHLFDFNVTITTLRNHFI QYFRRN			
	I-4_DAn	GCF_00328 5975.2	Drosophila ananassae	TT	TAAAATTAAAATAAA TATACTGTTAACTTA GCTAATTATAAAAAT AAATAAATAATTAAA ATACGTAGTCGAAGG CCTCGTAGCTAGCAC	MKINQANWDKFATHTNNNNNLYPP STNINKETAQINKIIMKSAHLAIPQTLP HRHTHTVPWWNKDLADLQEKKKQT WKALKRDTTQENIIKYRKTNALFKRA VKIRKKEATSKFTSTIQPNTPTTKIWN NIRRFCGLNPVKQIHCITNNNVTTND	cd01 650	260	518

					TACACCCCAAACTGT TAGTCATAGTTTTGT AAACTGTACTATCAT ATCCAATAAATAAAT AAATAAATAT	KQKIANLLAQHWSQLSKDSNFSQEFT SHKYNTTTLNHIPSKSAMEIDKPITKLE LTIALRSLKGITPGINRISYPMLKNSPL AIQNRILSHFNHIFENYIPQSYKTSLIIPI LKAQKDKTDIASYRPISLNCCVGKTLD KIIANRLWWYLNSNKLLHPHQYGFKK GKSTSDNLLYADHIITKSLNTRKHTSLI SLDFARAFDRVGIHSIVDQLIEWKCS HKIINYIINFTKNRKIAVRIGSSISNFYPL ENGIPQGSPISVVLFLIAFNKLANILTIY KEIKFSAYADDFLLLVNLKNQKNTTIN LTHIFRKIYDWGSYSGASLSQTKCQHL HICRKHRCNCTITTQNSNIKNVTSLKIL GLTINNRYNWSSHINSLIPSLSNKHNII KCLASHKFNCNTSTLLNITKSILISKIDY ALHIYGYTPKNFLNKLKTTLNAALRTV LGAFRTTPTHNLLYEANIPTIEQQRDH LTTKLFTNLTRSKDTPLHKHIQPKTNN KKWTSTLERTIENCQLHNIPFKPTKLY KTKPPWQLPQNIINTSLHNYKKTNTS SDTFQKLFFEIKETYKNHTFIYTDGSKI			

						NLNTAFAVTTETDVLRYGILPPSSSILT GEMIGILEAIKISTTKRGKSVICSDSLG AIEAIQNPLNLNKYPTYIRSLLQTKQP KIKILWIPGHSNIKGNDLADRTAREAT KMPLIYTPNENITDINKMLKNILIQEK ETLINKCNTWYKNINTNNTHISALSKI QQEEISRIDQIKFIRLRLGHTSLTHKHL LDPNTQRNCYRCQSSTPTNMEHILDS CPAISQNKQNLLKNRPATEYLKNPTA ENIKVIIKFLKETNLYNEI			
	CR1-83_HM	GCF_000004095.1	Hydra vulgaris	ATGGCGGCGTAA AGATGGCAGCCT AGTTGAGCACGT GTTTTGAAATAA TTTTTAATA	ACATTTTAAATACTAA CTTTTTTTATGTTTGT ATGTTAGTGTGTTAC AGCTGTTAAAGTTAC AAACTCTACGTAGTT TGGACTCATCACTCTT TAAGTGTACAGCACT ATTATATTATATT	MVVSKDTTQAINTAIEVLRKELIERIIK VEKENERLNTENESLKKQLAEQCNQ NNNKPLFSSLIKKTSEGAPPISDSETRL LNAISAEQKEKIKKEKNVLVFGLKESK RKSKDEIKEDDDESVGEVFDAIGLDKS CIVKHFRLNSRDKNRPGPLVIELVDPS YQKQVLSVARDLNKVEAYKNKVFINP DLTYRQRASLKLLLGERNSLNKIEEGQ KFDISICYQKRYAAENNRGRSKVRHF DLLSETICCRKLKSKNTKKINAPSSLISK LSEITCLYMNATSLNNKFDELIEVINS	cd01650	788	1054

						NKANVVMICETWWNNESVTNINGY TLFRRDRGVGRGGGVCIYVQNNIKAY EVSEKCLNDPNIEQIWCVIELGSEKVL CGSMYRTGISNISNCINITKSINFAHS MNLLNKYTGILICGDFNFNAIKWTED YYGTLTSETDTPAKIFLECLNNCFMH QNVIKPTFQNNIGNETNVLDLIIAENK NRVYKLVHLPPLGGINHGHHIIKFNYS HEECIDNIVLIKEKFNFNRGKFDKLNN YFKILNWDNEFKGLEINKCYEKWLSLY EEGCKLFIPIIKSGSIQINTKPLWINREL KALSKEKRNVWFECRRQKFRDNCLL AKYKELNKVYKKKSESKNTRFIKKKVN QRIKDFEYDLALNSKFNPKKVYSYINN KSKVKDNIRAIQLKNGSIETDLKVIAN TLNGYFASVFNKPQTLSAPIIDIETQKI CKNPDFNASVVEEYLNKLDTNKATG MDRVHPKVLKECKSTLAKPISLLFNKS FETGKLPKLWLCANIVPLFKNGDKLN PCNYRPTCIFNICSLTHATIDLPVSLTS VVCKVMEKIIKDKMMKHLVENKLIN			

						KNQHGFVNNKSCVTNLMESLDFITSS IDNGLDVIVLFLDLAKAFDSVDHERLL LKLNALGFKSKLFDWCQGFLKNRLQR VVIGNNKSEWEEVVSGVPQGSVLGP LLFVVYINDISNKIKSSCKLFADDMKIL RAIKNDNDVIELQQDIDLLMEWSNE WLMKFNQQKCKVMIIGNSQHKVLM NNTALAYTSMEKDLGVYISDDLEWK HHVNKAVNKANQKLGQIKHTFKYLD EKTMKLLFIALVRPYLEYAAPIWNPYR QYDKDKLECVQQRASRIKILKGYSYEE RLKKLDLLSLENRRRRGDLIQMFKYLK GYDIINFYTKPEMLDNGYKTRGHNSK LRRQYTGNIKRHNFFTNRVVNDWN NLEQETIDAVSINQFKKKIG			
	CR1-22_SP	GCF_000002235.5	Strongylocentrotus purpuratus	TATTGTGAATGA ATTTAATGACTTT TTTGTAAATATTG GACCGAATTTAG CAGATGAATTGC CGAATACAGATG	ATAAATATAATTCAT AATTCATATTTTACAT TATTAGTTTGTATATT GTAAGACATTGTTGA GTACTGATTCTTGTTT TTTTTAAAGTATCGAT	MIPVQLIRDSAPYISHPLAYIFNLSLET GKVPDEIKIAKVTPLHKKGPREDPGN YRPISVLPVIGKVLEKLINIRLIQFLEKN DIFYYHQYGFRKKYSTKLSLINLINSTL KSIDQGKVTLGMFIDFKKAFDTINHNI LLGKLSSYGIRGLPLEWVKDYLSNRFQ	cd01650	37	295

				TAGCATATGGTG ATTTTTTGGGGC AGAATATAGAGC AATCTTTATTTTG GAAACCAATAAC TGCAGTTGAAGT TAAGAGTCAAGT TCTGCTACTTGAT ACCAAAAAGGCT TGTGGTCACGAT AAC	TGTAAAAGTAAATTC ATTATTTTTTGTTTTG TTACGTTCTCTAAACT GATTTGTACCAAAAT ACTTTGTTTAGCTGTT TAAGGGGTTAGGCTC GACTAGCACATGTGC TATTATCTGGCCCCTT TTACCAACCAATAGT TTCTCTTCCTGTATAT GTAATTATTGTATAA AATGTATGAAATGTA TTGCTGTTGGTAAAT AAAA	FVQCGQTLSERKKIRCGVPQGSVLGP TLFLIYINDLPTVTDYFNFRLFADDSNL FHTFSADQGEIDVNEVNNNIQKIHK WCTANKLTINSMKTNFMLICSRRKKI GINGKLKIAGEEIKEVDVATFVGVHID RHLSWKSHIRFVNQCIRKKVGVLFRL RHFVPSRTLILLYKAFIQPHISYGIEVW GSTYKTSLRSILYSQKMAVRAITFSEW RTPSDPLFKRLRILNVYNLHDLSISTF MFDLVNGNLPHILIEYCDIIEHSYSTR QKEDGQLRLPKLRTTQGQFSLSFVGS EFWNRLPVDVRLTRSRNRFRNSVTS HLLNKS			
	Daphne-17_SP	GCF_000002235.5	Strongylocentrotus purpuratus	TCCATTTAAGTTC TTCTTTACATTTT GTACATAGTTGT TGTTGTTGTTTGC TTTTTACTACTTT GTTTTAAGTGTA GAATTCATATTCC	TGATATTAGCTACAG ATTTTAACGGTTGTT AAATTCATTATATTTT ACATGTAATTGTTGT TTGTTTCACTTTTTAT TACAGCTTAACTTTTT ATATTTGATTAATTG	MLCRARFVQCLNEKHVSVSTLREFLS PSDPQSTNNFTQSHYTPDEFNTFVT QKSLDKDVNNKLSFININCRSLQHNF DALSVLMMSLRMQVSIIGVSETWLR DENSQYSLCNLDNYVFVGNNRRLGR GGGVGIFIRKDINFKRRTDLDINESFLE SIFIEIERFNSSILICVMYRPPSQPVDDF	cd01650	557	812

				TTTGCTTTTTTGG TTTTCAA	GTCAGCCGCATACAA GGACCACCTCCTTTCT GGCTGCTACCAGCAT TAGCAATTTCCTTTGT ATTATTCATGTTTGTA AATTGATTCTGTGTA TTTCTGTGTTTATCTA TGTTAGTAATTGTAA TTGTGTGAAACTTTG TATACTGCTTTTGTTT TACCCTGTATGTTTTG TATTGCTTGGCAAAA TAAATTGAATTGAAT TGAATTGAA	LMSLVNVFDVVNVEKKLFYLMGDFN IDLLNVGKVQSVNDFLNILMLHSMIP LINEPTRVTDRSSSLLDNIFTNDSCYIY SGILFTDISDHFPVYCICDTDIPKSSCET DFYKRHLNENNIRMYFLQLLYSTEWY LSSDVNSSYNNFVFKFMKVYNLCFPV KLVRMKYKRKNKPWITNEIIKLSKKKC RLYKLLIKNPNVYRKEVYRKIRNELNN KIKIEKKNYLKKSFKDSTGNMKKTWDI INKAMGKKCKMTNIKEINNNDTVITD KNDISNNFNEYFVDVGYNLNCKFDCF DDSECIKYIGMNDKSMFFKLVTSSEIE DIVKKIKNNTSAGYDDVDVKIVKKCIY LLSSHLCTIFNQCLEQGVFPDKMKIAK VIPVFKNGVSETMSNYRPISVLTVFSK ILEKCMYNRLYAFLTKCNIINDNQYGF RAGHSTSCAIIDFIHKVASAIDNNEV MLGLFLDLSKAFDTLDHHILLVKLYQY GIRGIGLDLFKSYLLHRHQIVSIANVKS NYQEIKCGVPQGSILGPLLFILYINDLP SVSRILKSILFADDTSIFLSHQNIQSLQ			

						DIFNQEIKLINTWFSVNKLSINISKTHY MIFTNKKCDTDSIIIRLCNTEIKQVSSL KFLGITIDNKLKWTDHINAVCNVLSK NIGVMYKLKSFPKCILKMIYNSLILPFL YYGILAWGNAHETHIRKVILLQKRAIR TICHTSYLAHSHPLFKEMKILNFTDMY NYQCAIFMFQCYLQILPKSLLKLFILNC NIHHYSTRNAENFHLPLVHTTSYKRT VIFNGPCLWSTIPRNIRELKFVNTFKR KYKKYLMDVY			
	R2NS-1_SMed	GCA_002600895.1	Schmidtea mediterranea	CCTTTTTACGGTA TAATTATAAACA TTTATCATAACAT ACTTTCTGTATCA ACTATATATATG AGTATAAATCTG GATCTTCTGAGA AGGACTCATATA TATAAGATAATC TATAAAACTGGA AGGTAATGACGA	TGAGTAGTGGTATAT TTATAAATATAAGTA GCGAATTTATAATAT TTATCTATACTATCTT TATATATTTCAAGTG GATCTTTTATGACTA ATTAAAAATACTTTTT TATATTTTTAACTTTT ATTACTTTATTTTCTA TTTTTATATTCTATGA TTCTTATAATGTAATT	MAIITGEGDCHLLKKSLLKLKTMKNK TNKLNVNKGNDLDVTFPCDFCGLSFP TKIGRGVHISKAHRVERDAKLSGHAS AKKAYYWSTEEETRMAKIHLQEESRG DSNHVLRAMHSACSTGRTLESIRLH MKAKRYVKIYTELRNEKQIQSSGVENI LNNISDASPLTLPSETETLTSYLLNLPLK KGVSKWREQDLDFIIHSVSDSEEQKR LLDVYCKTVTEGYRKGKTGPLKKSTKV SPHHATTTKSKKFLKKKLKRELYGKVQ QLWIKDRCEAAKLVLSGKVSTLICEED	cd01650	385	643

				TGATTCAGAGGT	TTATTGGGAAAGAGA	RSARQGEYWKKIFERESEPDIRPVPE			
				CCAGTGACTATT	GTAAACGGGGGGCA	VYADPMLVRPISEDEIKKVVKDIKLKT			
				CCTCGAGAGAGG	ATGGGGTTAGTCTCA	SSGPDNISTEVLREIPVLVLVKLFNLIIL			
				AACTAGTTACGC	GTATGGATCCGGACC	LSHLPSEFKSSRTTLIPKVETPKDPGDY			
				AGTTCTGAATCG	CTCCGGAATCGGGAT	RPISVSSVFVRALNKILASRVSNGNKG			
				GCAATCACTAAC	TCAATGACTGGGGCA	AHAQKGFKALDGCLENTVLLNCILGD			
				TCAACAGGCTGA	GTAGATGCGGGATG	AKINNKSIALSFIDMKKAFDSVNHDSI			
				TGTCAGGATATG	GAAATTAAGAGTACT	LRAATIAGYPKLLLGYIKESYNGADTE			
				GAAAGGGGTCTA	TTGGCATTAGGGTTT	VEKIKATFKRGVKQGDPLSPVLFNNIV			
				ATAGTGGTCTTA	CTGGGACTGTGAGTG	ELAIKKAYESGIGYRMNGTVFPIMAY			
				AAACTCTCTTTTG	GTGAAACCGTAAAAA	ADDIILLSESKEGLQDLLDIVLTELRHN			
				AAAGTGACTTCT	GGGAGTAAGAATCC	GLEANADKCASIIAKKDGKQKKWVIS			
				TATTGGACACCT	GGGCTATTACCTGTA	SDSVLAAGKTMPCMNANSMYKYLGI			
				TTTTATTGTATAA	CGGCTAGAGAAATCT	YFGPLGITTERLTKIYSNQLYNIGIAPLK			
				TTATAAACATTA	AGTTCAGCCTTGATG	PQQRLEMINDFLLPGILHEAVLGNTN			
				ATTATAAATCAT	CCGTATTTTTTTGGG	VGELKDLDKMTRVAIRKWLHLPKDT			
				ATCATACTTTCTG	AAGTTTATTCGCACT	CSSFIYTGVKQGGLGVRSILAEVQFGK			
				TATCAACTATATA	GCCGTGGTTTGCGAC	ISRLESLNKSKDTRIQALTTCEVVKKAF			
				TATGAGTATAAA	CAAGTTTGATGGGAA	RQWSKPIRVGDSIATSKVQFYEANRE			
				TCTGGATCTTCTG	GACACAAATAGATTT	LLLSKCDGKGLREAGKIPFAHDWLRA			
				AGAAGGACTCAT	AGTTTTGGTTCTTTTG	KGKLMKGRSFINSIKVRGNLFGTPVR			

				ATATATAGGATA	CCATTACTGAAATCT	LNRLNKSGIPLIKCKMGCLMPCSLNHI			
				ATCTATAAATCT	ACAACGAAGCCACCG	AQTCPKVHGLRVERHNKVMNYIADV			
				GGCAGGTAGTG	GCTATAAAGAGTGGT	ARKGEWKTLVEPIIPYGATIRKPDLVV			
				ACGATGATTCAA	ACTCTTGTGGTATAC	IKDKKAMVLDIQVVSDVGNLNERNTL			
				AGGTCCAGTGAC	CTTTTAAGCACGGTA	KSEVYNDKIIADWVKNAGADSVEFG			
				TATTCCTCGAGA	AGGAAGTGGAATGG	AITLNWRGLWASDSWETVSKIGVNK			
				GAGGAACTAGTT	ATGGGTTGCTCATTC	SDAMIISLKVLNWTARMYCAWNKG			
				ACGCAGTGCTGA	CGGACTGCGCGTTGG				
				ATCGGAAATCAC	AGCAATTATATTGAT				
				TAACTCAACAGT	CGGGTGATACTAAAA				
				CTGATGTCAGGA	AGCCGGGAAATATCA				
				TATAGAAAGGG	ATTACTAGTTTTCTGG				
				GTCTAATAGTAG	AATACTGCTGCTCAC				
				TCTTAAACCTCTC	TTACAGTTTCTGGATT				
				TTTTGAGAGTGA	ACAAATTTACCCATA				
				CTTCTTATTAGAC	GGTAATTGATATATA				
				ACCTTTTTATTGT	GCTGGCTGTAAGGG				
				ATAATTATAACC	GTATGGGACCCTTAC				
				ATTTATCATATCA	AGGTATCATTATTAA				
				TACTTTCTGTATC	AGGCTTGGATGATCA				
				AACTATATATAT	GATGGGTATATAAG				

				GAGTATAAATCT	GGAGGGCTATCAAG				
				GGATCTTCTGAG	TTACTGGTTTTCTGAA				
				AAGGACTCATAT	TGATATTTGGATTAC				
				ATATAGGATAAT	ATTTCTATCAAGGTA				
				CTATAAAACTGG	ACAGCGACGTCTATA				
				CAGGTAGTGACG	ACTTATATATCCATAC				
				ATGATTCAAAGG	GAAATGGAACGATCA				
				TCCAGTGATTATT	GATGGATAAACAAG				
				CCTCGAGAGAGG	GGATGGATTATCAAG				
				AACTAGTTACGC	TTACTGGGTTTTCTG				
				AGTGCTGAATCG	GATGATATTCGGATT				
				GCAATCACTAAC	ACATTTATCTTAAAG				
				TCAACAGTCTGA	GTAACAGCGAAGTCT				
				TGTCAGGATATG	TGAACTTAACTTTCCA				
				GAAAGGGGTTTA	TACGTCTAGGGAAAT				
				ATAGTGGTCTTA	TCGCTAATAGAAATA				
				AAACTCTCTTTTG	TTAGTGATTAATAAC				
				AGAGTGACTTCT	CACACGGGCTAACAA				
				TATTAAACACCTT	CCTCTAAGGTTGGGC				
				TTTAATGTATAAT	AAGTGAGTTGGATTC				
					ATCCAGCTTGCGCGT				

				TATAATTAAATAT AAGCAG	ATTATATTTTCGATCC GATGTCACCTGAGGT TACGGGGAATTAAAC CACAGCTTTTCTTATA TTACTGAATCAGGAT ATAAGGATTGCATAT TAAAACGTTGGTGGG TTAATGAACTATAAC TGAGGAGTAATCAAA TAGTAATGATTTGAA TGCAAAAGGTTTCAT TTGTTACTGATTAACT TGCGGGTTTCATTTA AATATCTTATTTTATA CATGGGCAAGAATAA CAATGTGAAGTAATT CAGCTCTAGTAATCT AATGATAAAGCAATT ACCAAGGTATAACGC CCTATTTTGTAGGGT TGGCCAAATATAACT				

					TGATAATAAATGCTG ATTATTGATCTACTA GGTATCAGGGTGCA ATGCCTTACGAAGGT AAGGTTGGACAAATG TAACTTGGTAACATG TATAACTATAAGACC CCCGTTCAGGATAGT GGTGATTGATAGTGG TGATGTAGTTAATAC AGGCGTGTATGAATC TGGTGAAGGTGGGA AAAACTTCATAACTC CGGTTATGATTTGAA GGTCAGGTTTGGATT GGGTCTGATCGGCTA ATCAAGGTAGTGATG AACTCAAAGTAGGG GTCTATGGAGAGGG AGCACTTGATAAGAT AATTCACGATAAGAT			

					GGTTAGTAAAAACTC GCGATCCCTGCGTTG GAAAGTTTTGAAGTG CTGAAGGTTGGTTAG CAATAATCAACAATG CTCGTTAGCGAAGGG AGAATAATTCCGGGC TATCATTCAAAACTG TATTTTATCAATGAAT TATT				
	Vingi-2_SPur	GCF_000002235.5	Strongylocentrotus purpuratus	AGGCTCTGTATA AGCTTGGCTTGT TAGAGCTCTTGT GAGTTGTGTGCC GCGGGGTTTCAT ATCCTTAAACTG CTGGCGAACAGA GCAACAAGACCA AGCTCTGTAATT GTTAGGTTTCTCC AATGGTAGGTTT	TAGTGACTCGATAAG AAGAAGAA	MATNDNDDQFVAEAQGRTHGSLSG PALSIISSNIEGLSGAKQELLADLCITTN CDILCLQETHRGPTRVRPRISGMTLIA ERPHDQYGSAIFARNGLAVDKSYLSD INNIEVLSIDLMGITVSSVYKPPQEPFD LPGRATDGRMNVVIGDFNSHSISWG YREMNEDGELVETWADANQLSLIHD AKLPASFNSARWRRGYNPDLIFTSTRI ANQCEKVILDPIPHSQHRPIGLKVNA VITPRAVPFRRRFNLRKANWEKFAKD FDSYIQDLPATPQNYDTFVELMKKSS	cd01650	487	742

				GTCCAATAATGC		RRNIPRGCRTSYIQGLTDHSKVMYED			
				CGAGTTTGTGCC		YIEHFQMDPFSEETLESGENLTKEIGE			
				ACGGAACCTAGT		SRRRKWQEFIQSTDMTHSSRKAWKT			
				CGGACATTTCGA		IRILGNDYTKAQPRSEVTADQVAHQL			
				TTATATTCTTTTT		LLNGRGNPDHHPRRARLPKVPDALS			
				ATGTCATACGTC		SQDDPGFTRPFTMDDLLAAIKAMKN			
				TTTATCTTTTTTTT		NKAAGLDDILCEQIKQLRPGALNWLL			
				TTTTAAATATTTT		KMFNECLGTNNIPKLWRKSKVIALLK			
				TTATAGTTTAGTC		PGKDAASPKSYRPISLLCHTYKLFERLI			
				TACCTTGCTGAT		LNRIAPFVDEHLIPEQAGFRPGKSCTG			
				ACCTCCTCAAAA		QLLNLTQFIEDGYESKLITGAAFVDLS			
				TATATACGATAG		AAYDTVNHRILTRKIFELTKDVRLTGLI			
				ACTATCATTAATT		QTLLSNRRYFVVLNGNRSRWRRQKN			
				CTGCTGCATTAC		GLPQGSVLAPLLFNVYTNDQPIHPNT			
				CATATAGCTTACT		RSFLYADDLCIATQNQSFVKLEESLSD			
				TAGAGGTTACTG		ALAGLIPYYATNHLRANPDKTQISAFH			
				CAGATCAATCAC		LKNRDANHQLRISWYGKRLKHTPNP			
				CTGTACTGTATCT		VYLGVTLDRSLTYKNHIANTKAKVGA			
				AGAGTATAGTTA		RNSILKKLANTNWGTDARTIRTTALA			
				AGTGTAGGAGA		LCFSAAEYASPVWSRSAHAPKIDPAL			
				ACAGTAAAAACA		NSSCRAITGCLRPTKVEDIYLLCGIAPP			

				GTGGGGCCATTG AAGGAGCGTTCA CGTCGCGGTCCG TCAGATGAAAAT CTGCGAATCCTT CAACCACACTAC TCAAGTTTCTTTT AGTAACTACCAC AACCAACACC		HVRRLVLSQKEKDKQENNPLHPLFEQ QPVGKRLKSRHSFLHSVEPLNTSPHN QRMTLWSNHLQTTPHKLSLSPKESLA TGSSETWPTWLSLNRLRTGYGRCRVL MQRWGLGDEQTACECGQEQTMEH LLVCPLLPEPCTQDDLEAVNCRAKAC VDHWTGLV			
	I-1_DPe	GCF_00328 6085.1	Drosophila persimilis	ACTCTGATATCG ACTCAGAAATAG ACAGTAAACACC CTATCAAATTCCC ATAAAAACAGCA CAACCTCTAACA TTGTAAAAATTCT ACCACTAACTAC CTCTAAGCGCAG ACTAGCAGCAAT TAAGGCTGCGGA CAAATTAAAAAG	TAACTCTAATATAAA ACCTATGGAATACAT CTCACATATAATAAA CATAATATAATAAGC ATAACATAATAACAA CATAGCCGAAGGTCG TAGTAACCAGTGCTA TCATTAATAAGCTAG GTTAATAGTTTTAACT CTAACCTCGCTGTAA ATAAATAAATAAA	MTLTIIQWNLRGYINNYNELQIIIKQY QPHIIALQETHLFNIHNIPIPINFSLYAK NYSPNFGGAAVLIHNSIESKEIHLQTD FDATAIEVESKKKFKIISAYISPSQNFN ESNLHNVLSNNSSSLILGDFNSWSKL WGSPKNNARGNTVAKFLNNTNFILL NDKSPTHFTTHNTYTHIDLSFCSPDLF QYANWKIIDDLHGSDHFPILITLFNN NKQNKTQNNPIFKINSANWDAFKDK TDYFNTKNSPSNNINKESSCLNKILLQ SAHLTIPQTSIKNHPRNVPWWNNGL QTLNNERKQAWNKLKKHITTTNIIRF	cd01 650	491	749

				CTCAAGCAGCAA AAAAACAAGACA GAAGAACCCCAA TACTGACGCCTT GAGTAGCAACAA CTGCTCAATGGA AATAGAATCCGA CTAAAACTAACA GATCAAGTAAAA ACAACCCAAATA ACCAGCACTCAA GAGATAAAAACT CAAAAAA		RKAAAKFKKEIKSAKNKSLSSFTSEIQP QTKTSQTWSNIRRFCGLNNMKNIHTI IHPTTLQPTSDNPSIANSFAEYWSDQ SKDSNFSSMFTNNKYSISTYQSSPPSQ SAQKIESDITSIELSSALSSLKGHTPGID RISYPMIANTSPAVKDRILKHYNNILN NYIPQAYKIGLVIPIAKPGYIKTSITSYR PISLNSCFAKVLDKIIAKRLWWFVTTN KHIHTNQVGFKRGKSVTDNLAYLDHL ITDALSHKQHLSIVSLDFNKAFDKIGL HAITKQLQEWNIGPKILNYIINFMSKR KIKCRVANNLSSTFPLHNGIPQGSPIS VIIFLIAYNKLANLITIQRNLTFTAYADD FHIIVKFKKNQTQENFNLDDLFQKIDE WCSFSGSSLSFDKCKHLHICRKHNCK CTITTRNLNIPSVDHLKILGLTFDSKYR WTRHIEDLASSLSNKINIIKCLSSNKFN CNTHTLISITKAIIVSKIDYGLVFYGNSS ITWLKKIQTILNCAVRIALGAFRTTPIT NLLMEANINTVADRRDFLQAKLIKNL YFPNNSPLHKISKSLQSRIHFNNKIPSA			

						ITYSLQHCKKQNISLLIKPEIKTKLPFW KFQPNTINTELEKLDKNNTHDSIYQQ HFLTIKNQFNDHTFLFTDGSKTISTTSF AITTEKNTLKVGLLQNFSSIYSAEIIAIF EAIQIVKSMRGKYMICSDSLSSLQAIG NPDNFNYYESKIRSLLMQLHPKIKLL WIPGHKNIKGNELADQKAKSASTMP LTYTHNNNISDINKHLKTKMLKDKLIK LEQTSNWYQFINPNQITADKFTKRLD QNTTSRIKQTKIIRLRLGHTRITHQHYL DPTKINTCTHCLNNVPVTIEHLLKDCP SSQLTDIRFKIFANNNPLTFLSNPTKTE IDLIIKYLKETNLYPLI			
	LOA-2_DSer	GCA_008042835.1	Drosophila serrata	TGGCTGGCAAAA ATCCTCCAGCTG AACCCTCCACGA GGGAGCCGGCT AGGCAATGGGC ACTGTATTACCG CGTACAATGGTA ATGCAGTGGCGC	TAAGATTCTCCAGGA CCATGTGAAAAACAA CATCCGGATAAATAA GGGCCATACTGGCCT ATGATTGGCCCCCTC AGGAGGGCCGTCCG GTTTAACCTAACCTA ACCT	MSLRVLQLNLHKSKVASAELLIAMEQ GSADVALVQEPWIASGNAVAGLKSS NYNAFYTPSVSRNRTAVLVRKGLQA NLMSHYSSDDLTVVLLESKKKRLLVAS CYMAHDRPAPPEELRSLVEPGPKDQ QLIVGADANAHHSVWGSPDINDRG ESLLDFILSTNLYIANVGEEHTFVGPTS SNVLDITLVAGQSTMVSDWKVLERP	SSF56672	436	801

				ATGCTCTGTCCG		SFSDHRYIQFQCEFDHSPKPTVFRNP			
				GGGTAATGCAGT		RNTNWGKFQDIISRKLATSPGTVKSD			
				GTCTAGCTAAGG		EDIEKSLETLSKELLGAYHASCPISRSK			
				CTTCGATTCAATT		KRTRPPWWNKDLSLQRDDLREFFKL			
				TCTCCCGCCTAA		AKQAKDDIINEEYKILLRNYKKEIRRA			
				ACGCCATGGTTG		QRSSWRNFCSNIETAPETARLRKLLSK			
				GAGGTGCCTGGC		QPTIQSQLKRDDGQWTEGSEEALTA			
				GGCAGGCATAAA		LMHGHFPGCTKITNTHPAISFWIGNL			
				ACGCAAGTGGGT		LSCRRVQSDWGTASMVKGAHRGTP			
				GGTTCCCCGGAA		QGGVLSPLLWNLVVDDLIRRFERKAP			
				ATAAACCAATTTT		KITAYADDISIIITGVCPSTLSSIMESTL			
				TCACAGGATGGA		REICEWTDKVGLSINADKTDLILFTKR			
				GACAGGAGAGG		YKVPPWIPPKINTTRLTPKTQVKYLGT			
				CAACCGGGGCTG		VLDSKLAWKPNVLERVKKATIALYAS			
				GGATGTCAGCCC		KKMLSSTWGLSPALMHWVYISVVRP			
				AAACGTCGGTGG		TLLYGVLVWWQAMEKRTYKKLMER			
				AAGGCGTGACTG		TQRQALLCISGALRSTPTKALETIIGID			
				CTACCACCGGCG		PLDIHAQLTAGKAAQRLVALGSMSTK			
				GGCACGAAGGG		GYGHSSIGRTMSSVSDYMTPRTCPD			
				GAGCAATCCTCT		VKTSTFMGPEDWKTVQDHARYLNIY			
				CTGCGTGTCACC		TDGSKMEGGVGAGIYCTDPETRLSYK			

				AGCGGTCACACC TCTGATTTGACG GGAGCAGGTCA AGTCAGTTGTAG CGCTAATGCCAA ATAACGACGAAA AGAAGGCTAAGC CGCACGGGTGCC GTTGTCTATACA GACTCGGACCTG GAGAGCGTGCA GCTTTCGAAAGA AGAGGAAGAAA GCCTTCTTCGCTC GCCAGAACCAGG CACCAGCTCCCC GCGAATAGACG GGCCGAAGCGTT CAAGCGAGGGG AAGAAGGCTAA GGCGTAGTATAA		LPSHCSIFQAEVFAIRKAAELAQEIARP CEAVNLFVDSQAAIRSMQSSAVSSK NVLASRDALDILSTTTSVRIYWVPSHQ GIEGNEIADELAKEGVGLAIDRTENVP VSLRTLQSELERRADVRAKSRWRGNS ATCKTSKIMCKERNDKLSHYVLHLSR KDCRLLVGLLTGHCLAAAHAVKLGIT NNGNCRKCDESEATETLEHLICKCPA LTQARVKHLGAPVLASLADASKRSPG ELLAFAKTTAILKDFNTA			

				AGCTGCCCTCCG AATCCAGAGCCG GCTCCGGACAAA AGCTGCCCTAAC CGAAGAGGAGA ACGTAAAGGTCA ACTGACAGGATC GATGAGAGCAG GCTGCACTTCGC CAAGCTTCCTAG CTTGGCATCAAC GGCCGGGTTCGC CAACAAAGTGTA GGTGATGCTTGC CACCAAGCGGCA ACGCTCGGCCGA GAGTGCCATAAA GGACATTCCGGC GAGCAAGCGGG GACTCGTGCGCA AGGAAGCAGGC					

				CCCCCCTGGAAG GCGAAAACCAAA TCCAGATTAACA AGAAGGCAGAG GACATCCTGTAT CCCAGACACGGG AAGATGGCATGG GTAGCGTCTACC TGCGCCTCAAAA AGCGCCATCCTG GGGATAAGGAT GACCACACCCTG AAGGCAGGCGA GATGGAAAAGG ACCTAGGTCTGG AAACCATTTTTG ACGCCGCCCAGG AACTAACGCTGG ATGAGTCCGAAG AGGAGGAAGAC GGTGACCTGACG					

				ATAGTAGTCAAC CCGTCGAGTGGA GTTGAGCCCAGT ACCC					

В таблице 10 представлен перечень систем ретротранспозаз, не содержащих LTR, с аннотацией сигнатур доменов обратной транскриптазы.

Представлена база данных белков ретротранспозаз и соответствующих 5'-UTR и 3'-UTR ретротранспозонов для использования в новых системах Gene Writing. Домены обратной транскриптазы в белках, описанных в данном документе, были идентифицированы с использованием консервативных сигнатур RT и аннотированы для указания присутствия и местоположения доменов RT в полипептидных последовательностях. В некоторых вариантах осуществления система или способ, описанные в данном документе, предусматривают полипептид, имеющий аминокислотную последовательность в соответствии с таблицей 10 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или его функциональный фрагмент. В некоторых вариантах осуществления система или способ, описанные в данном документе, предусматривают домен (например, домен обратной транскриптазы), имеющий аминокислотную последовательность, соответствующую домену (например, домену обратной транскриптазы) из таблицы 10, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или его функциональный фрагмент. В некоторых вариантах осуществления система или способ, описанные в данном документе, предусматривают РНК-матрицу, содержащую последовательность, соответствующую одной или обеим из предсказанной 5'-UTR и предсказанной 3'-UTR из таблицы 10, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или ее функциональный фрагмент.

Семейство	Элемент	Номер доступа	Организм	Предсказанная 5'-UTR	Предсказанная 3'-UTR	Предсказанная аминокислотная последовательность	RT	Начало	Конец
Penelope	Penelope-2_ADi	.	Acropora digitifera	TTTCATCTCTCCCACCGACGCAGCACCACAGTTTCTTTAGAAACTAGAAATTCATTTACTACTTAACCGCTGCGACCTCATTCATTCCTCGCTACCGCGACGCTGAAACTCACG	TAATTTATTCACAACTTCATGTGACCATATCTTCACCAATGGCAAAGCTCCTCTACACTCTCATATAAACCACAACAACCCACAATTCCTCTATTCGCTCTGACGAAGGGCTAACGCTCGAAACGTCAGCTTTCTAAATCTTTCACGGTGGTAATTCAACCTTTATCAACTCGTTTGATAAAACCAAATTTTTGT	MPNLKTLTKELRQEIVSYLYPLPSISSNLAFRLAIHIFTKTRFSSHINFISRCLQHKVIPKGFRSNFHASIFSNVSNFNLKYLHEIQRTQNTFSHNIMRSTIRAMCSKRNELNKQIIECRSQLSNVCPSVLLRSIHAKIQALNSKLFHHLEQTKKHKLENLTNSPKNRTEPLENQNHHTVVTIPENLPLSNAEKSVLSKGLNFVPISKKSDEFTTRQDVEKFLRRVQLKAFFHNKEDKSDNTEKDAFETLTAKKSKWTPPEGQFASIDYFIKKCRHDVHKLKSNCNTKLSNLSKEEWTALINLKNRKDLVIKAADKGGATVVWRTDLYQQEAI	IPR000477	371	610

						RQLSDPTFYTKVNKDLT PANQKIVKDTIQELIRKQ ELPVTAQNLIITTPRTSCI YFKPKIHKPNNPGRPIVS ACSCPTELISSYLDKVMT PIVKSLPSYIKDSNHALE TFGNFNFSGENKIIFTMD ITSLYTVIPNNEGLQALK YFFNQRPIKKPSSETLLR LAELVLTLNCFSFGDNY YKQINGVAMGTKMGPS YANLFVGFIENKFFSNY HGPKPDLYKRYIDDCVG ATSSTKEELNLFINSVNS FHPALKYTWEISENSLAF LDIKLSINGNGLSTSVHY KPTDSHNYLLHSSSHPQ HVKNAIPFSQFLRLRRLC SDDTDFNNKCEEMCQFF KKRGYPDSAVTTGKHR AQEIDRETALQTSQNEET DRIPFTLTYHPQNLAIKN VILKNFKILSNDPETKHIF			

						SLPPLISFKRDKNLGNFL VRSAFKFNNKPGTFTCK RTRCKTCPFISNTVNISG PNRSVKVTDHFTCISSNV IYCITCTLCKKIYIGETGR RLADRFREHLRDAEQNN TDASKPVARHFNLPNHS HHNMTICGLYLHHGNTE SRKNLEQKLIFQLGTLSP HGINERLSFH			
Penelope	Penelope-1_ADi	.	Acropora digitifera	GCCTCTTCGCTT GAATCTCTGCTG GCCCACAGACG TGCAAAAAAAG CCGCTCGTTCCC GCGGTATTATTA ACCACAATCTC ACCGNCGATAA CCTCAATCAAC ACATCAATAAC ACTAATGCCAG TACTAATGACA TACTATCTAATG	TAATATTGTTACCT CTTATCATGTTATTT CATTATTTTTTCACC TTATACTTATCTTTA GTTTACACTGTCAA TTATTTTTACTTATT TCCGGTTATTACAA AATATGCAACATAA AAATCACGTACCAA TGTAAAATTCTCAT TGTACCTGAAGAAG GCTGGTTTGGCCAG CCGAAATATAGTAC	MRLAEYFFDENSNNRTA NEHDTPFHNKSTWNPPN DRERALNTFLDAVKLDI TTTKPKPTQDNLTVTER QAIGQLKQRQDIIIKPAD KGSGTVVMDKTWYIDE CNRQLTDTKFYQRLDED ITADIQKRVTFYVNRMH KDKLINDKTKQYLIQSD VKPGRFYILPKVHKPGN PGRPIVSSNSHPTERLSH FVDYHLQPLVHKLPSFV KDTNDFLNKLLTIGNLP	IPR000477	143	383

				TTACACATAAC GCCTCTACCTCA ACTGTCACACG CAAACGCTCTC GCCGCGAAACC AAACATGCCAA CATTCATTTGGA TCATTCTTCTGT TATCAACCTATC CAGTTGCTCTTT ATCTACCGACG AGATATCTATCC TTTCACGTGGAC TCACTTTCTGCC CTACTCCACGA CATATCAACTG GCCGGAAGTTT CAGCCGACATT TACGACTTTTCT CGGCGC	ACCTACAAAAAAG ATCCTTCTACGTTG TGTCGACTTCTGCT T	ANSLLVTLDVSSLYTNIP HNEGINACERFLRTSSHK TIPTSTLCDLIRMILTMN NFSFNDNHYLQIHGTAM GTKMAPSYANLFLGFFE ANALKNAPFQPHTWLR YIDDIFMIWTEGLDNLKI FIDYLNNIHSTIKFTSSHS STNIPFLDVNVSLTNDGN ISTDLYTKPTDKHQHLL YSSCHPLHTKKAIPFSLA LRLRRICSTDATFHTRTA QLATYLLKRGYNRNFV NKQIRRAADIPRQLTLQT KDINKPKRIPFITTFNPSL PHISHIIKKHFNLLLSSNR CKSVFQHPPVVAFRRSP NLRDLLVTAKLSPNSTN PQLPCGSFRCGKNCATC PYISHGLTTYTFFSTGET RPIKFNFTCETKNLIYMI QCNRCNLQYIGETKRRL KDRFNEHRRTIDNPNNK			

						SKPTTAAEHFLSSPNHTA NDMILIPIEKIFSNRDSIR KAREAFLIQKGRTIDPDG LNIREETY			
Penelope	Penelope-4_ADi	.	Acropora digitifera	TTTATGAT	TGATTTACTCTCAC AGTACTGCCCAACG TATTCTACGATACA CGTACTGACCTTAG ACCTGTAGACGGAT CGAAACGCACCAAT CACTGTTTAGTCTT TCCAGCCAATTACA TCTAGGCTCAACTG ACAGTCGACCAATA ACATCACGAATAAG TTGACCAATTACAT CTACGACCGGAGTT TTTATAGTATCTAC TGACGTTACACAAA TCACTTGACTCTGA AGATGACTTCCGCT CAGGTTGTCGAAAC GTCAGTCAATGTCA	MTPGFKPFTILRSLGRQS ERLANNYISYASNVRRIR QQLWFNHRCKDLRLVPS GLRLKSPLNTQEAIQIVK ATCRRLIRARINDCHRRL NHYNNRLQQRLDKLKQ LIPTNLLDVIRDIADQRA KKTTEQYRIKTELKLTRL QRTKNKKRHKTDENWV RNISSRPLDKTETQVLSY GLKHSVTPKRIPTETIVSS VEAVLSRQRDLSEPTKD NIRSRIASTIQSASLTDNN LTKDERQALKRLRNDN DILILPADKGRVTVVMD KTDYHDKMDELVNDKQ TYEVLKRDPTPALQRKL NSKLLQLKKADAIDIRR YNRLRCPVPQPAKLYGL	cd00304	459	541

					TCTCAAACAGTCCT TCTCAGGACTACAC TCACCCGGACGATC GTACTTTAC	PKLHKPNVPMRPIVSFC GSPTYELSKYLTTILKPL TDESRHKLQSTETFIDAI KTAQVPDDHKLVSFDV KSLFTSIPLQLALDCTET AINNSTLQLPLLTNDLM DLLNLCLTSTYFQYNGQ HYKQLHGTAMGSPVSV VVAEIVMQNIEEQALAS YKRTIPLWLRYVDDTFT TLHKDEIDDFHEHLNRQ NAHIQFTKEIEDNGKIPF LDCLVIRDNNRLQTTVY RKPTHTDRLLDESSYNPS SHKATTIRTLTRRALLVC NSHDSLADERKYLDNVF SKNNYNCDFVTRNTYRT EPNETNTNLTPTTTVTIP YIKGTSEIIARILQPYNIR VAHRPITTLRKLLTNVK DKDQPRDRQGAVYKIK CCDCQATYIGETGRNLN TRLTEHRRATRNGDINN			

						NIAEHHLQTNHRIDWDS ATCVTYNTNYYQRIVLE SWFTNLEQTPINRCLQLP APYKRLIDDINRQTTH			
I	I_Ele 14	.	Aedes aegypti	CAGTTGACAGC CGACCGCCGTC GCGTACAGTCG CCTTTCGCATTG TTCCGTACTGTT AATATCAATTA ACAGCGCATCA GCCAACTCGAT CGTTGTCTGACG GGGTGCTCCAC AGCGGGGCGAG TGTGCATTGTTT TATTCCGGATCA GTGCAGCACCC AGTGGTGTGGT GAGAAGTGAGG TTAAATAGACT CGGTAGCGTGG TTCGAGTACTGT	TGATCCTGCCCAAT ATGACGACGACATT GATGCCAGCCCACG GATACGAAACTGTT TTACTAAATTGTAC TATGTATAGATGTA AGCTTGAAATGTTA GATTGACAGTTATC TCCTTTTCATCGGTT CAGTCTTCTGCTGA CCGTATTTCTCCAG GGTCTCAGCCTTCT GCTGAGCCTTCTCC CGTGTTGAACTAGC ATAATGTTAAAAAA CACGTTAATAAAGA TGAAAAAAAAAA	MPGPALGGAPLESATRT LCPPNGFSQTYFSTAAPS PLPVVGSTPGPTLGGALS GSVSSTFLGSTNSKETSL SSFPVPALPTSHNFILQW NLNGYRSRLCDLELIIQN NQPWILALQETNNISAD EMGRTLGGQYAWTVKR LANQRHSAAIGILKGIPH KILDLPSDLPIVGIQLLGS PSVSVACAYLPGGNIPNL HGQIQNCLQALPEPRIFL GDLNSHHPVWGSPKAD KRGSTLINAFEEEDMLIL NDGSPTFFNGHFSSEIDV TAVSLSFARNIQWSVSS DLHGSDHFPIRIALTTGA SIPKNTRRPRWKYDKAD WAEYDLLIRSALRNSPP	cd0165 0	575	833

				ATAKTATTATTT		DNMADFISTIVESAKATI			
				CGGGCGCGAGT		PRTSSTPGRKALRWWSD			
				GTATGTGTTAGT		ETKTAVKARRKALRQLR			
				AACTTGTGAGT		RLPLDHPDRQSTLENYR			
				GGTATAAATAA		KLHLECRNIIGDAKKAT			
				AATTATCTTCAA		WESFLESMNASQTTSEL			
				ATTAAAAAGCG		WNRVNALNGKRKTTPL			
				GTTCAGTGGTTT		TLQVQGNAISDPEAVAS			
				TCAACTGTAGT		ELGKYFASLSAIESYETH			
				GTGACTGACCA		FXNSVKPSTSAINNFIIPP			
				ACCAAAACAGG		NTADPEMEQPFSARELS			
				TGAGTCTTCCAC		FALSRSNGRSAGPDEVG			
				GAGTGCCTCCTT		YPLLKNLPSVGKLALLD			
				CCGGCCCGCCG		LYNKIWANNDFPPAWK			
				GCGCCATGCTG		ESLVVPIPKANIPTRDPT			
				GCGAACGGCCA		KFRPISLTCCAGKVMER			
				TCCCCTCCACCC		MVNRRLRYHLEANGLL			
				AGGGGACCCAG		DFRQHAFRQGYGTSTYF			
				GGGGCCCCGGA		AGLSDVLKEAYDKGNH			
				GGAGGATTTGG		AEVISLDISKAFNRTWTP			
				CAAAGTAGGCA		LVLEQLAAWGIGGRTLH			
				ACAGAGTTCTC		FVRNFLADRTFQVVIGN			
				GGCGACTACAC		TKSSTFPEETGVPQGSVL			

				CGGTGCATTGCT		AVTLFLIAMNGVFSRLP			
				ACCGGAGTTTA		KNIYVFVYADDIVIVVC			
				TGGACCGTGAA		GSTPTMTRIRAQTAVKS			
				GGTACTGCAGG		VAKWASDNGFQLSASK			
				ACAACTGCAGT		SIRCHICPSGHRITGPDITI			
				ATCTAAGAATG		DGQPIPLRKTVRILGVIV			
				GAAGCGACTGC		DRALSFRQHFDTVKSAC			
				TGGAACCATGC		RSRLNLIKSISRPHRSNN			
				CCCAGGATCCG		REIRFRVARAIVDSRLVY			
				TTTCTTCTTCGG		GLELTSIAMDRLVEVLSP			
				CTATCCGTCGA		IYNSYIRIISGLLPSTPSDS			
				AAAGGCCATTG		ACVEAGLLPFRIFIRSSIC			
				GAGGTCCCATTT		CKTAAFLSKTAGEDRVF			
				TCGGGGCCTTC		LLDEGNRALSTAANLSL			
				AAGGAAAATAA		PPVTRVHWLGDRSWRS			
				AGGCATGTCCT		VPPKIDNKIKNSFSAGSN			
				ACGTGTTGAAG		SAALRRSVAELLQTSYS			
				GTGCGCAGTCA		DFALRYSDGSLTSAGVG			
				AGACCAGGTMA		IGVAGDIPDVSMSLPSQC			
				ATAAGCTCCTCC		SVFSAEAAAAFIAATTPS			
				GCATGAGCCAA		DRSILVLTDSASVISALQ			
				CTGCAAGATAA		SDSPTHPWIQAILKYALP			
				MACCCCAATCC		DTVFTWIPGHCGVPGNE			

				GGATCGCTGAG		MADHLARSGLSGQRYTS			
				CACCCGCGGCT		EVPFMDLKRWIKSTFRQ			
				MAATCAAAGGG		HWEDSWYRTRTLFLRKI			
				AATGTGTGGTG		KNSTTTWSDLPILKDQRI			
				ACCAACCATGA		LSRLRTGHTLISHNMGG			
				CACAATCGGAC		GPFHKECESCHIPASVDH			
				TATCGGATGAC		VLCACPMYEHLRQIHGL			
				TACCTGCTTCAA		SDNIGEVLRDDATTIAAL			
				CAGCTCTCCTCC		LSFLRDANLYSSI			
				CAGGGAATCAA					
				GGGMSTGCGAA					
				GGATCACGCGA					
				CGCGGACCGGC					
				ATGGCCAAAGC					
				CGTGGTAAACA					
				CCGCCACCATC					
				GTACTCACCATC					
				AAGGGCACCGT					
				TATTCCGGAGC					
				ATGTCGTTTTCG					
				GTTGGTCACGG					
				TGCAAGACGAC					
				GTCTTTTCTACC					

				CATCTCCCATGC					
				TATGCTTCCGCT					
				GCTGGGAGTAT					
				GGGCATACCGG					
				TAAGCGTTGCA					
				CAAGCCCCCAG					
				CAGCATCTGCG					
				GACGCTGCTGC					
				GAGGTACACCC					
				GGAAGAACAAA					
				CCCCCGATCCTA					
				CAAACACCCAG					
				GAAGGAGCCGC					
				CAGTGCCCACG					
				TTGCGTCGAAA					
				ACGCGGTACCA					
				GTGCACCGCAG					
				AGCCGTTGCTG					
				CAAGACCAGTG					
				CAAAACAGGAG					
				ACAGCACTCAG					
				TCTCAAGCCGC					
				AAGTGTCCGGT					

				GTACWWGAAG					
				GAACAGGCCAT					
				ACAACACCTTC					
				GCGTCGACGTG					
				GACATTCCGTAT					
				CCGCAGGCTCG					
				ACGTGAATACG					
				AAGCCCAACAA					
				GCGAGCAAGTC					
				GAAGTGGTGCC					
				AACTTCTGCCG					
				GTGTGGTCAGT					
				GCCAGCAAGGA					
				CGCTGAAATCG					
				ACAGTCTGAAG					
				GCCATGGTGGC					
				TGCAAAACTCC					
				GCGGACGGAGA					
				AGAAAAGAGGA					
				TGGCTGAAATG					
				GAACGGGCTCT					
				GCAGAGCCGCA					
				CCGTCAGCGAC					

				CGCCTCGAGGC					
				GGTTAAAGAAC					
				ACGGCACCATC					
				GGAGAGCTGGT					
				CAAGCAAGTGG					
				CAACCCTAACT					
				GAAACGGTTCA					
				GAACCTCCAAA					
				ATGCTCTAGCA					
				GTAAGGACAAA					
				CAAATCCCGGA					
				TCAAAGTCGCA					
				CCATCGCTGAG					
				CTTCTAGGCAG					
				GCAATCAGCTG					
				CCGTCCAACGA					
				CAAACCCGCGA					
				CAAGCTTTGGCT					
				TACCGCGCAAG					
				ASGATCAAACG					
				CAAACCGCCCC					
				AAGTCGCCAGA					
				GGTTCCCGTCAT					

				GAAACCGCCCA					
				GTCTGAACCCC					
				AAAGTGGACGT					
				ACAGGTGTCCA					
				AATGGATTTCA					
				AATATAACACC					
				GGACCTCAAGA					
				AACTCTCACAG					
				GCGATGAAACC					
				AACGATGGAAA					
				TGCGTCTGATCA					
				TAGCATGATGT					
				CAGCCACATCC					
				GAAGCCACTGG					
				AATCTCCAACA					
				CGTCTCACGTTT					
				CGAACCCTACG					
				AAAAGAATGCA					
				CTCAACCTCGG					
				ATCACAGTTGC					
				CCATCCGGCGA					
				TTCAAATGCCTC					
				ATCCACCAACT					

				CGAAGCGGAAA AACAAGCACCG CAAAAGAGGTG GTAGAACCAAA CAGTAAAACAG TAATTTACGCTC AGTCCTCTTCAA GCACATGGTCT AAACAATTCGA CAATATTTTCTT CCCGTCATCTTC ACTCCGCAAAT TAGCGTTGAGA CAGTCAAATCA CAAGAGTCGAC TGAGCAAATAC CGATGAACACA CTACAGCCTCC GAACGAAGTAC TCACGGACAAC TCAGCGACAAC AGGGGTCACGG ACAGCCGGGGC					

				CCCGTCAGTGC					
				GGACGCTACAC					
				CCCAACCGGAA					
				CTGGCGGACAA					
				CCCCCGGCATC					
				CGGGCGTAACT					
				GTTGTTATCGGG					
				ACGCAAAAGGG					
				ACACAGTGCCT					
				ATCCCCCGAAA					
				GTCCCAACTAG					
				CTGTATGGGTA					
				CGCCAGGGCCC					
				CGTCGGTGCGG					
				AAGTTTTGGAA					
				GAACCGGAACC					
				GGCGGACAACC					
				TCTGGCGCCCAT					
				ATACGGTCTGG					
				AAGAGGGACGC					
				AAAAGGGATCA					
				ACATCCTATTCC					
				CCTKAGGACAG					

				CGTGGGAAGCA AGCCACAGGAA ACCTTTTTTCTG CCCCTGCCAAC ACACAGAGGGT GAAGGTCGGGA CGCCCAATACC CCGAAGGACAA CGAGGGAAGCA AGCCTCCGTCG AGTCGAAACCC CCCCGGAASAG AAAACCGCCGG ACCACTACACC TCCAACTGGCA CAACCACCCCG GTCAGCGACAA GCTGAATCCCC CGACTTCCCGG GCAGCCTTGGG ACCCGGAAGCA AGCGACGCCAG GTTGTATGACCT					

				GTTACACAACG					
				CTCGTTACAATC					
				GCTCAGCCCGC					
				CGACTTCAGCC					
				CGGGTATTCCG					
				GAACCGTTTCC					
				AGACTTCAACT					
				CCAAACGCGGA					
				AGCATACGAAA					
				ACCGATCGCCA					
				TCCGCGATAAT					
				KATAGCTCAAC					
				CCGCTGCTGGC					
				CCCTCAAGCCA					
				AACAGCCTCCC					
				CAGATCCCTATC					
				AGGATCGCCCT					
				GGGCACCTCGG					
				ACTCAACAATC					
				CGAGCGGTAAG					
				GCCGACCCTTTC					
				CCGCTGGGCCT					
				CCACCCGGTGA					

				ATCTTCGTCCGC CACAACAACCT ATGTCAACCCA TTTTTCGACGCG GCTCCCAGTCTT CTTCGCTCTCAA ACTCACAGCAG CACTCCTTGATA AGGTTTCCTAGC AGCCCTGTTGG AAACGACTGTC TTGCCGCAGCTC CCGAAGTTCCG GCAGTAGTAGG TCCC					
RTE	RTE_ Ele3	.	Aedes aegypti	GGGCTGCAAAA TGGTGAGTCGA CGGAGTCCTCA CAAGTTCCTACC TCATGCTTCCAC GGGTCGACGGA TGACAAATGAC CGCCAGCTAAG	TAAGTAAGTAAGTA AGTA	MKLEILGLSEVRWPNFG EHRIPSGQILLYSGLRGE HAPRHRGVGFLLSAQAY AALMKWEPINERIIVARF RTRVRNLTIVQCYAPTD AAELQDKENFYSQLNAV VDKIPKGDIKIYMGDFN AKIGSDNSSYEPVMGRH	cd0165 0	503	762

				AGTTGCGTGCTT		GLGEMSENGELLAEFCG			
				AGCTGGTAGTG		NNDMVIGGSLFPHRPVH			
				CAGCCTGGGTC		KVTWVSRDGFTENQIDH			
				CTGTAGTCCTTC		ICISRKWKRSLLDVRNK			
				TGACTTCAGCTA		RSADIASDHHLLIGEVRL			
				GATTGAGGAGG		RIARIRRQEERVGRRFNT			
				TACGTCCACCG		RRLEDATVKRSFVEELE			
				AGCGTCTGTTCA		TXAANISEGGSVEEQWT			
				CCAAGGAGGTG		AIKNAFITTSENNLGELR			
				CGGCTCAAACA		TRRKQWITDETWRKIEE			
				GCGTCTGTCTGG		RREAKAAIERAKTRGAK			
				CATCCAGCGGC		ALSRQRYSALEREVKRS			
				TGAATAAGAAA		CRRDKRAWADSLADEG			
				TGCTCTTCCCCG		EKAAATGDIRLLYDISRR			
				GAAGCTACACC		LSGAKMNTTMPVKDTT			
				TAAGATGGCAG		GQLLTDPTDQLKRWFEH			
				CCCCATCACGG		FEQLFQVSTRPPTPQHDP			
				CGGATAGGGAA		PRVRRISRVNTEAPSLHE			
				CCTTAGGCTAA		IETAIQSMKSNKAPGVD			
				CAACCTACTGTC		RISAEMLKADPAISAQLL			
				CCCGAAACATC		HQLFRNIWDTATFPADW			
				AAAAAATGTTC		MQGVLVKVPKKGDLTV			
				AAGAAACCGAT		CDNWRGIMLLCIVLKVL			

				AGAGAAAGATT		CKVILNRIQEKIDATLRR			
				ACGAACTGATT		QQAGFRAGRSCVDHIVT			
				TAACGGCAACG		LRIILEQVNEFQESLYMV			
				ACTTTTAGCGCG		FIDYEKAFDRLNHENM			
				AAACAACGGAC		WGALRRKGVPEKIIGLIE			
				ACGAATTGGAA		AQYEAFTCRVLHNGVLS			
				CATGGAACGTA		EPIRVVAGVRQGCILSPL			
				CTAACCCTAGC		LFLIVIDEILIGAIDREQN			
				CCAGCAGGGTA		RGLLWQPITMEHLNDLD			
				AACTGGCACAA		LADDIAILAQRRSDMQS			
				CTCGCCAACGA		KLDDLAERSSAAGLNIN			
				GGCGTGTCGC		VNKTKSLDVNTAGPSSF			
						TVAGQTVENVENFQYL			
						GSQLASDGGTKIDIGARI			
						KKARAAFASLRNIWKNS			
						QISQRTKIRIFNSNVKSV			
						LLYASETWCVSAENTQR			
						LQVFINRCLRYIIRAWWP			
						HNWISNADLHRRCHQQP			
						IQTEIRQRKWGWVGHTL			
						RRGGNEICKQALEWNPA			
						GHRSRGRPRGSWRRSLN			
						KEIKEVDGNLTWPQVKA			

						IAGNRPGWKSFTSALCT RMGAQDP			
Penelope	Penelope_Ele2	.	Aedes aegypti	AGTTAGGGATT TGTATAATTATT ATTATCCCCCTT TTCAATTTTTGA CCAAGACATGA CAGGTATACTC CTGTCGAAATTT TTTCCTAGGTAG TTTGATTGTGTA GGTACTTATCTG GTGTTAGTATTT AAGGTAGAGAG ATGCGACGATT TCTTCAGTGGAT ACGAGTAACTG ACACTGAAGAA GACTGCAAGTG GTAGTCGAAAT ACGCGTATCTGT CAAAGATAAGC ATTTAGGAGCG	TGAATACATAGGTA GATAGTTAGGGATT TGTATAATTATTAT TATCCCCCTTTTCA ATTTTTGACCAAGA CATGACAGGTATAC TCCTGTCGAAAATT TTTTCCTAGGTAGT TTGATTGTGTAGGT ACTTATCTGGTGTT AGTATTTAAGGTAG AGAGATGCGACGA TTTCTTCAGTGGAT ACGAGTAACTGACA CTGAAGAAGACTGC AAGTGGTAGTCGAA ATACGCGTATCTGT CAAAGATAAGCATT TAGGAGCGGAATTA AAAGGTACACGGT ACTCCATCTACTTT	MASTLSESNTFIKLKKSI AGLKKDIAFLSECRKLN LTPKSHKIKVRGQVPQDI IKKMESEYLRSSIKKLHS KLNSQTLKCYSLHLKLA KQYPWEFPLFLTKVIQA EECEARRKKDIHSKKLAI LKNKQHNKVVTALPKV HSIEGFVVNRSTQQFSEE QLTLLNKGLAYAVATKP DIEQTIIDVETAITHTIPIE FQESARNDTEESLKGNK FQLRGQNSDEDKVLSEL RDKSVYYVKADKGNAV VILDKDDYDNQMNEKIT TGPYRHFRVDPLPTMVR HVEKTLKECKPLLGDNT SRLKVSNPVLPRIKGLPK IHKPGNEMREIISAEGSP THKLAKWLVKEFQSMP KPFESRSVNNTQEFASQL	cd00304	445	525

				GAATTAAAAGG	AAAGTTTCTCTATT	QASGYIQDDEIMVSFDV			
				TACACGGTACT	CGAGATTCTGCTCA	KALFPSVPVKEAINLLEE			
				CCATCTACTTTA	GAGGATTCGAACAT	WLLNQHNESNWKNKVR			
				AAGTTTCTCTAT	TCATTAAAGAG	YYIKLTRLCMEENYFSF			
				TTGAGATTCTGC		RGNFYKQTKGAPMGNP			
				TCAGAGGATTC		LSPFLSELFMANFESILK			
				GAACATTCATT		KKDLLPQRWWRYVDDV			
				AAAGAGTTGTA		FSIIKKDYLPKILEAINSIH			
						KDIKFTHEEEKDNKLPFL			
						DLLITRESSHFEFEIYRKP			
						TNTMRVIPSTSNHTYQH			
						KMAAFHHMIHRMQTLP			
						LSETGKSKELDYIFETAR			
						LNGYRNSTIQAVVDKRA			
						RDKYRKSMTTLTAEKED			
						LKRVAVNFDTKITQPLA			
						KKFRKYGVDLVFSSRSS			
						QLKSRLGSTKDKINKLN			
						RAGVYKISCPHCNKIYIG			
						QTKRTLEIRFKEHIAEVK			
						KADKELEKGLAYDFKSK			
						VAEHVFQDGHQMTSDNI			
						DIVRSVSSSWKLDVAES			

						LEIYKQKPSLLLNRDQG NGQSRLFKFVPKKYKRA			
CR1	CR1-61_AAe	.	Aedes aegypti	TT	TAATGTATCCTGTT AAGTGTTGTCCAAG AGTTGATGCTTAGT TTAGTTTAGTTTTA ATTTTATTCATTAA GACCATGATGTCGG ATGAATCTTCAATA CAAATACAAATACA AATACAAATAC	LDNFAFRGLKQRNLLRN HLDRTLDLIFTNDAITDC SSVDEACQPIVDIDPHHP AFVLTVRSPTIVPFVEVS DEEVRDFGKVDYETLHV LLSRVDWSTVYQCNDV DNAVSIFNGKLHECFTH CVPLAMPRRKPAWSNA RLLYLRRNQSKMQRRFS KRKNDCNKLAFNRASRS YRYYNRYLYGKYVRRV QDNLRRHPKQFWSFVNS KRNETGLPSFLSLNGHS ASSAEGKCELFASWFAS VFDTSSQSQADPDAASQ SLPTDIIDVDIFEVSHMM VEEALKKLKPSISPGPDG IPACILKHCKAVLIAPLK AIFNLSLRCQQFPKAWK MSYMRPIFKKGSRTEVS NYRGITSLPAGSKCFEII	cd01650	332	579

						MNDVLLQACKTYISQNQ HGFFPQRSVISNLCEYTS FCIKEMDKGGQVDSIYT DIKAAFDTVDHSILLAKL NRLGASSRICSWLYSYL TDRQLLVKIGKAISAPFT PHCGVPQGSNLGPLLFV LFFNDVSHVLSNCYLLF YADDLKIYLIVRSINDCR QLQEQLDLFVSWCTLNH LSISVSKCCIISFRRMKSP VLYNYAINGTILNRVDN VKDLGVLLDHQLTFKM HYSETVDKANRQLGFIF KIASEFDDPLCLKSLYCS LVRSILEFASVIWSPYDA VWIARIESIQRRFVRCAL RSLPWSDPLRLPPYEHR CALLGIETLEERRKMNQ AVFGAKVLRSEIDSPALL RQLNIYAPQRVLRSRQL LYQPPRRTMYGSNNPID			

						TISRRFQDNYNLFDFNIS TNTFKNRLRNRQV			
CR1	CR1-81_AAe	.	Aedes aegypti	T	TAGTTTGGGCCCCG TATCACTCTGTACA CATCAATCGACTTG AACGAGTTCAAAA AAGTTTTATTAGGT ATGCACTCCAGAGA CTACCTTGGCGAAA TAGGATGCTGCTGC CGCCATACGAGCAT CGCTGTGCACTCGT CGACTTACCTACAC TACGCAACAGGAG AACTCTTCTAAAGC AGCTTTTCATTTTTG ATCTGCTCGAAAAT AACATCGACTGCTC TGGTCTGTTAGGCA AATTAAGTTTCAAC ACCCCTGTGAGGCT CCTACGGAGGACTG CACTATTCCGACAC	SQHQRGGGVLIAVKKAL QSNVILINDANQLEQISV RITYGTMSLVVCTVYLP PNTEPALYQQHAASVNQ LCKLLGTQDEIVVLGDY NLPYLRWRFDDDIGAFL PINASSEQEISLVEDIVAT GLMQMNHLLNSNERLL DLAFVSNASVCELLEPPL PLMKVDRHHTPFVLTLD VIQTPLENRKNETFFDFR QCNYDELNAEIAQINWR DILAFGSLDDAVTRFYSE LNVIFERHVPRRSRACRP DNKQPWWNAQLRNLRN RLRKARTRYFRVRNRLN RDAVRDLESQFNELNVS CFQSYIDRLESNMKDDP KQFWSYLRSRTTSRGFP QLINYDGKTATTPDDAV KLFSSFFQSVLSITPPPSS	cd01650	441	690

					GCTGCTCATCGAAC GTTATATGGGCAAA ATAATCCGCTTGAT GTATGTTGTGAAAT GTTTAATAATGTGT ATCACTTGTACGAT TTTAACATAGCTAA GGAAACATTTAGAT TAAGGATTCTACGG AATTGAAGTCTGTG CGATGTATTTTTAT TTAAATCGAAGACC AAATAAATCAAATC A	ASYLSALKTFDLNLQPTS FTQHEVYKKLCSIDGSK GPGSDGLQPSFIKACSAS LAEPAALLFNRSLAESVF PAKWKEALITPIHKAGN VHDVKNYRGISILSCLPK VFESMLLDFLYPAVRHII TEDQHGFVRKRSTTTNL MSYVSTLIDKIEKRQQID AVYVDFTKAFDRVPHQL AVEKFQRMGFPDWLTK WVLAYLINRSASVRLGT TLSDPFEITSGVPQGSHL GPLLFILFVNDLCSELKS SKVMYADDLKIFRVINS PVDCCALQMDIDKILDW CSRNGMDVNIAKCNVIT FTRKRSPTVFEYSMHGA TIERVRLVKDLGISLDCK LRFVEQISSVIAKAYAVL GIIRRNTKDFRDVYCLK ALYIYHLFAAFWNMEL			

Crack	Crack-13_AAe	.	Aedes aegypti	CA	TAATATGCATTTTT TCTAATGCCACTTC CGTATCTGTTCTTT GTCATTCCACGCCA CCACCACCCTTCGC CACCACCCACCGCC AGCCACCAATCACC ACCACCCACTGCTA GCCACCCGCCGCCA GCCACCCGCCGCCA TCGCCACCATTCAT CACTTCACCCATCG CTTGCCACCAACTT CAGTAAAGATCAAT GCTCATAAATATTT ATAATGTGAATTGA AAGTTTGTGAAAAT TCTAAACTATAGCG CTTCCTTCAAAGAG CTACGCTCATTGGA AGTGCAGTTATCGC TCAATCGTTGAATT ATACTTATTTTTTTG	HGGLAVFVRNELQMDV CRNIVIEGFHHIHCRFKF EANAINLHAVYRPPSFD VRRFLHEIEGMISLSKNN EKCVLVGDLNIPLNITTN NIVGEYIRLLETYNMLPT NTVVSRQSSNNILDHVV CSTSMIEDVFNETIFTDV SDHCLICSSINMKCHTSE STLHKRIIDHSRLNELFE RSILNIPQNLSVNERLNY VITHYNEILNDCSRLVTQ RVKLKGSCPWMTLDLW KLIKIKDKLLKKKRNNP NDNHTTELFAHVSKILQ DKKAKTKRDYYEGLLS GNNQKAAWKIIKNVMG NQKSNSKPTAIRNGAQLI TDPQQVCSSFNDHFCAV GPKLSSTINTNRNINRFG TLGSLQSTIFLKPSTRNEI ILLINRLDSKKSAGPDGIP VSFIKHHHSFFASLLMET	cd01650	420	676

					AAAAGATGAGAAG	FNEIIVSGEFPDCLKIARV			
					GTTTTATGCCTATG	IPVFKSGDPKDMNNYRPI			
					GGAGAAGTGGCTTA	STLSIIDKLLEKLLVSRIL			
					AAGTAGCTACACTC	EFSTKHHLIYGQQYGFR			
					TCACGGGCTTTTCC	QGSSTLTACHELVDEIY			
					CTTCTCCAAGAGAA	DALDNKQIVGALFIDLK			
					AAAAAAGCAGTAA	KAFDTVDHDLLLRKLES			
					ATAAATAAATAAA	YGIRGIAKDLLKSYLTDR			
						HQYVSIGEHKSQLRLVT			
						TGVPQGSNLGPILFLLFV			
						NDVARLNLHGKLRLFA			
						DDTAVFYRGTNCDTILN			
						QIKMDLEVLLEYFGENV			
						LSLNLNKTKYMLIHTPR			
						RKIPAHGPISIGGHNLDK			
						VYEHEFLGLTIDSVMSW			
						AAHINKLKSKVSSLCGIL			
						RKISSFMPQTCLKQIYFA			
						LVHSRLQYVTAVWGSA			
						SKSHLRELQVLQNRCLK			
						IVLRKPLLYPTVNLYSNR			
						NDSLLPIKALYEYQVLV			
						QMQKIVKGTSTHHNTVL			

						HMSQQSRSSRQANHIFL FRPNSEFGKKKITYFGSK LYNALPENCKSFQNMFH FKRFVRNALKQKVEQFV L			
CR1	CR1-121_AAe	.	Aedes aegypti	AA	TGATTTTTATTGTGT ACTTGTTTTTATAT ATCTTGACCTATTT TGTCGTAAATTTAA GTAACCACACATCA TTGGGACTAATTGT CTGTTGGTGTAAAC AAATAAACAAAAA AAA	VNDVVNENGRCLDLCF VSTQDVAPSIAIAPSALV KNVSHHPPLVLLLADSE PCEPLTISPVSYDFNRAN QQGIVEFLASVNWLEVL DTNDVENAALTFTHILG HVIERHVPKKVHGTAGP PWQTQELRRLKTSKRAA LRLYSKRGTQSHRNHYV SVNNKYKSTSRRCFRRY QQHIQQNLKTKPKSFWR YVNQQRKESGLPSTVKL DDEVASDRNQACRLFA KKFSSVFCDEMLSEEYV ALAASNVPLVMETLNAF SVDDGMIARAATQLKSS INPGPDGIPSSFVKSYIEQ LLLPIRHVFALSLTSGTFP	cd01650	316	566

						SVWKNAIIFPVHKKGDR KDVSNYRGISSLCAISKL FELVVMEPLIAHCKQFIS NEQHGFIPGRSTTTNLLS LTSHVSDSFTERVQTDVI YTDLTAAFDKINHLITIA KLERLGISGLMLRWFKS YLTERQLTVGLEGHFSE TFTATSGIPQGSHLGPVIF LLYFNDVNYVIKGPRLS YADDLKMYLRVRSICDA ITLQNDLDAFAGWCSLN RMVVNPEKCNVISFSRR KEPVYFNYQLFGKILQR VDHVKDLGVILDAQLTF KHHMSYVIAKASRTLGF IFRIAKEFTDVYCLKSLY CSLVRATLEYCSTVWHP HYQNGVQRIESVQRRFI RFALRRLPWRDPHRLPS YHSRCQLIHLETLQSRRD INRAMIVADLLQGRIDCP AILGQVDISARHRNLRA			

						NSMLRPPLRRTNFGQNV AINGLQRVFNRVAPAFD FHLPRHTVRRNFAEILTR RIV			
Kiri	Kiri-9_AAe	.	Aedes aegypti	ATGAACTGTTCC ACTCCTAAAAG TTATATTGGGA AGCTACTGTTTC CCCTTCTGCTGC TGTTGTTGCTGT GTGCTGTTGATG CTACTGCTGTTG CTGCTGGGAGG CGTTGCTGTCGA ATAAGATTTTTC TGCTGCTCTCGA TTGTGGTGAATT AGATCAAATTA GACTTTTTCGTT TTCGGCATATTA GCAGACATTAT TTAGCTGTTTTT TCTTTGTAATGT	TAGAACATTCAGGA AAATCGAATTCAGG AACATAATTCGAGT TAGTTAAATAGTTT TAATGGAATCACCA CAGTGTAACATTTC AAGAAGGAGTTTCC TTACGTTACATAAA TTCAAATAAAGAAA TAAATAAAATAAAT AAATA	NICHINVQSLIARNFTKF HELKLTFIDSKVDIICFTE TWLNHTISNTMIGIEGYK LIRNDRNRHGGGICIYLR NSLLYKVVSMSQSSSND FYGTEYLNIEIKVGFDKV LLGVIYNPPNIDCIDVLH DILENCTINYDCSFFVGD FNTDILKFSSRSRRFGEM LDAMSYECMNSEPTYFH QSGCSLLDLFITDSPDFV AKLDQISMPGVSNHDLV FCSLRLSLNRVEDIVQYR DYVNFDSSALLDGFYNI EWNNLFCMEDPNEILNFI NYHILKLHDDFIPLRRKK KNRTPWFNSDISHAFINR NLAYDQWKRSRTEIDRN TYKRLRNFANDLVRQA	cd01650	467	726

				GCTTTTCTCTAA TGTCAAATGAA CTTCAGACTAA AAGTTAATCAC AATGAACATAT TAGTAGTTAGT AGTTTTGCTTTT CCTGATTTCTAC ACTGCATGTTGC TTGGGTTCGATG GGTAGACTGCT TTTGAAGCCTAT ACTGTTCAACAT TTAGATATGGA TGACTTTTCCCT TGATAATGCTTC TCCGAACAATG TAATTCCACGTG TTGTTCTTAATG CTGCATTAAAA CATGACTGTTTG A		KINHDKQNFNVDLPSKQ LWKNIKKLGISNDSSFLD QDCLNANQINDYFISNYI NDSSSVPIPFSSDGFKFN VFHEEDIVNSIFSIKSNAV GLDNIPIRFIKILLPIALPV YKFLFDSIIKTSVFPRAW KNSKVIPIKKKGNSASLS NLRPISILSSLSKVFEKLL KIQITKFITDNDLLHPLQS GFRQNHSTNSALIKVHD DIARVIDRRGIAVLLLIDF AKAFDRVSHSKLVKKLS SKFSFSHNAVKLIESYLN ERKQAVFYNGVLSDFSFI ESGVPQGSVLGPLLFSLF LNDLPASLEFCSIHMFAD DVQVYLCASDDVDMND FARKINHDLNNILRWSQ NNLLAVNPSKTKAMLIC KLKNRPLPPSIYFDGKRV QFYDKLENLGVIFTSNLC WDAFINGQCGKIYGTLK			

						KLNLTTKHLDTAMKIKL FKSLILPHFIYCDFLFSNA TASALNRLRVALNACVR YVFNLTRYSHVSHLQDV LLGCQFSKFYSFRACLQ VYIIIKRKTPRYLYDKLQ FMRNTRTMSLVIPNHSS AYYGQSLFVRGINNWN RLSPTLKSTANIYGFKGG LLRELERLQ			
CR1	CR1_ Ele32	.	Aedes aegypti	GG	TAGGATTTAGTTTC AATTCATGTAGACA AAAGTCAGATGGAT AATATACAAATATA CAAATATACAAATA T	LRTKIDALYCNAIDCNY DVIILIETGLDAAINSVQL FGDNFNVFRCDRNAKNS TKSSFGGVLIAVHRRFTS SSISITTGESLEQVCVSVL VGNQRILLCAIYIPPDKS KDVNVMDEHIESVRELR NLSSGNDYLLVFGDYNQ PELCWKRCGDEIRPTNS GSISAASSSLIDGMDFCN LHQMNIERNHLNRTLDL VFTSLELSPEITCSVAPLL PTDQHHPPLELDLMLPN	cd0165 0	497	748

						GVNNVPTVSNATGELDF TKIDFEALSLFLSETVWH VLLDGLTADNMAETFCS VIANWLNTNVPTRRPAV SPVWGTPLLRNLKRRRN ALQRRLRRHRSLENIRSF HRACDEYRELNSSLYKS YVLRVQSNLQRNPKQF WNYVNSKRKSSTIPMSV FLNDSVSSSPLEKCELFA AHFSSVFDAKPASPSEV DSAVSDVPRDLLDLDIF QITPAMVSQAAKKLKCT FSPGPDGIPSAVYCRCMS ALAEPLCRIFNQSFVDST FPAIWKQSFLVPVHKKG EKRNVINYRGITSLSAAS KLFEMIVSDVILAQAKG YISIDQHGFMPGRSVTTN LLDFSTTCFEQLENGAQ VDVIYTDLKAAFDTINH DILLAKISKLGASNRLSS WLLSYLTGRTLRVKVDS			

						CMSSSFTSMSGVPQGSN LGPLLFLLFFNDVMLLL GHGCRLAYADDLKLYF VVRTVEDCVRLQALLD MFVDWCRRNRMTLSVP KCIVVSFYRITRPIMHNY VVNDVVLKRADSFCDL GVLLDHKLTFNLHRSNV IAKANRQLGFIAKISRDF TDPHCLKALYCSLIRPIL ESAAVVWTPYQLSWSL RLESVQRKFMRLALRNL PWRDPLNLPPYPDRCRL LDLDTLDRRRKTQQVAF IAKLMNGEVDCSRLLSM LNFRVAPRVLRGGALLQ QRFHRTAFGSNEPFTSMI RLFSMFEDLYDFDGSSN QFVSRVRSSDLI			
CR1	CR1_ Ele41	.	Aedes aegypti	CTCCCAGCCTGC ACTTCCAGGCA AGTATGCGTGTT ATACCAGACGT	TAATGTTGTTTTAT GTTGTTGTATTTTTA TCGTTGATTGTAAT TTTGTTAGTAACTG	MTPKCNPVMSSNGQQP GRSVESLMEAPGPSDPV ERNAVTVLHSRPGPAVG CGTGVFQSATSGEYSSII	cd0165 0	607	857

				TCTTTSCGTGAT GAATTGCCAAA TTTCAGCTCTCT ATCCGCCGCTG ATGTAACATCA	CTTCATATTTGTTA AACTAGAATAAGAT ATACATAAGTTATA CATCATTGGGACCT GTTTTGTCTGTTGA TGTCATCCAATAAA TAAATAAA	DNAPSPSIVLNSRTCNAR KMPISVNAKKSDDLLIY YQNAGGMNCDVDKYLL ATSDDCYDIIVLTETWL DTRTISTQVFTPEYEVFR CDRNPRNSRKSTGGGVL IAVKKKLKANIIENDLW CSTEQLWVHVQFFGHSL FLCGIYIPPDRARDETLIE THLQSVTSVIEIAKPMDE IVVLGDFNLPGISWISSG NGFLHPDMNHSHVHSA ASRLLDGYSTATLRQIN HVTNENLRSLDLCFVSA QDVAPPISNAPSSLVKQV NHHPPLVLKLIVQHYSTS ESIAQVTYDFTKADHEGI LEFLAAVEWTEVLDTRD VDNAVSTLAHILGHVIE RHVPKKTLASKGLPWQT RELRTMKTAKRTALRRY TKNPTPRMRDQYVRLN AAYKRTSRHCFNRYQQ			

						QVQRDLTAKPKSFWRY VNEQRKVSGLPPSLFLN GDVASEPQSVCQLFAQK FSSTFCDEVIPEVLVTQA ASNVPLVNDSWATLTV DDDMISRAASQLKSTMK PGPDGIPSAFLKRHIDSL LVPIRHVFNLSLISGAFPS LWKNASVFPVHKKGER KDVNNYRGISSLCAISKL FELVVMDPLMTHCKNH LSDDQHGFISGRSTATNL IRLTSHIADSFAEGAQTD AIYTDLTAAFDKINHAIA TAKLERFGISGNMLRWL HSYLIGRKLTVTIEGFQS NEFLATSGIPQGSHLGPL IFLIYFNDVHYTLKGPRL SFADDLKIYRRIHSDYDA MCLQQDLHAFASWCTL NRMTVNPGKCSVISFVRI RRPVVFRYELYNTVIQR VDSVKDLGVILDPQLTF			

						KHHVAYVVAKASRTLG FIFRVAKDFTSVYCLKSL YCSLVRSTLEYCSVVWN PYYQNGSDRIESIQHRFI RFALRQLPWRDPLRLPS YRSRCQLIDLETLQLRRD IARAMVVADTLQGRIDC PVLLQSLDLNVRPRVLR NNTMLRLPFRRTNFGQQ SAYIGIKRVFNRVASVFD FHLSRQTIRRKFKTVFLS FLDE			
CR1	CR1-109_AAe	.	Aedes aegypti		TAATTTTTGTAATC TTGGTTTGAATATT TTAGTGTTTGTATTT TGTTTAAGAAACAA GGGTCATTGGGGTA ATGTTTTACCTGTT GACTAATAAATAAA TAAATAAA	AAPASSNILMYYQNVGG INSSLAEYNRAVSDSCY DVYAFSETWLNANTLSN QLFDSSYIVYRQDRSAS NSNKRTGGGVLFAVRSN FKSRVLNSPNGSSIEQLW VAITTVDATLYLCVIYIP PDRVNDESLIEKHTDSLD WVVSQLKPRDSIMIIGDF NLSGISWQYDXSGSLFP VASRSTISLTSRKLLDAY	cd01650	518	768

						STACLRQTVGIMNENSR			
						VLDLCFVSEELRDASTV			
						MQAPSPLVKTCRHHPPIL			
						LKMEICPQKRFHDVSES			
						VFYDFRNADFNGMNDF			
						FENVDWDDVLRDSDAN			
						LAASTISAILLYGIDQYV			
						PVKVKREPSKPAWSNSE			
						LKNLKRVKKAALRRHS			
						KFRTDSTRARYMQANTE			
						YKQLNDCLYNAHQDHL			
						QSILKANPRSFWKHVNE			
						QRKESGLPPTMSDGLHE			
						ADSKEDIADMFRTQFSN			
						VFTNEHVDPDIVNDAIR			
						DVPRLSTFVQQIIITDDM			
						VASASKELKSSTGSGPD			
						GIPSLVLKRCAETIAAPL			
						ARVFNLSLASGVFPNCW			
						KQSYVFPVFKKGSKQLV			
						SNYRGIAALSATSKLFEV			
						IILQKLVQSYAHYISPDQ			
						HGFMPKRSTTTNLACFT			

						SFILRLMEAGQQVDAIY TDLSAAFDKMNHQIAVA KFDKLGMGNSMLLWLQ SYLTGRNMSVKIGDHVS SPFRVWSGVPQGSHLGP FLFLLYMNDVNFKLKCL KLSYADDLKLYWTIKQP EDAVFLQHELEAFAEWC QINRMSLNVSKCSVISFS RKHAPFHFGYFLAGTQL ERVSTVKDLGILLDSKM TFKDHVAYAVSKASSQL GFLFRYAKKFRDVYCLK ALYCSIVRPTLEYSSIIWS PYYQNGIKRVEAIQRKF VRFALRHLRWRDPLNLP SYGSRCQLIKLESLVSRR NVAKACFVGDLLQGNID CSTLLSCLDINVRRRNLR THSFINIPSARTNYGLQE PMRSMSRVFNMCYHVF DFNVSRETNKSSYRRVL C			

CR1	CR1-48_AAe	.	Aedes aegypti	AATTCTGGCAC CACTGCGTGTTT TCGTTTGTGTTG TTTATAACGTCG CCCTAAAATCCT GTGAATATTCGT GCAATCAACTC TATAAATCAGG GTTTATTGCTAA ACTGCCTCGTTC CGCTTCTGTTAG CCGCAAGCATC AGGTTTTCGCAT CACTTTTGGCAA GGTAGGACAGT GATTTGTGATCT AAATCCGAGCA GGCTAGAGAGT GGAAGCATCCG CCATCGCCCGA TAACATATTTCC CACCTCGCCGCT CTCCACCTTCCA	TAACAGTCTGGCGG ATTAAGGTATTGAT CCAAGACGTAGCA ATTAAATAAATAAA TAAATAAA	MRTKTNEFRTKMETCD YDIIVLTETWLRPDVVN AELASNYCIFRCDRNDN TSTLQRGGGVLIAVKST LRCSAVELTDYIHLEQV VVAVRLPDTTIYVCGIYI RPNSHFDVYDSHASAVQ EICNLSSTRDSIVVVGDY NLPHLTWHFDDDCNSYL PRNASSESEIMFTQSMIG SGLHQINSLRNMNGRIL DLAFVSEXSFVELIDPPS SLLKVDSHHKPFVLRIEI ETRTRAPVEEADVDDFD FKRCDFAALNDMLVAV DWNETFGDAPLDRITVL FYDKVFDIIRHHVPRRRR CSAVHSKLPWWTPELRH LRNIVRKARKRYFRNKS HDNRKNLKSLETRYSVC QENAFRSYTNRMESNLK RDPKSFWTYIKKLKNIN RIPEELSYNGTTAETAEA	cd01650	495	745

				ACCGATCGAAA		SANLFADFFSSVHNNHS			
				GCTGCAACAAG		PTFSSSARQNIQSFNVSL			
				ATCAAAAATCA		PMITISQHEVSSALKKLD			
				ACCCGCCCCCA		VSKGAGTDCLPPTFLKE			
				TTTGCCCCANCG		CADSLKTPISVIFNRSLRE			
				CCCCACCAACG		QVFPEIWKVASITPIHKS			
				CCCTACTAYTGT		GSVHSVENYRGISILCCI			
				TTGCTGCACCG		AKVFEEIIHRSLYSAVRPI			
				AACGAAATCAA		ISDVQHGFVKKRSTVSN			
				AACAAGAAATC		LMSFTNVLSNSVEKRRQ			
				GTCCTGTAATAC		VDAIYFDFAKAFDKVPH			
				TCCGCTTCATAT		ELVISKLKHIGLPDWITE			
				TGGAAATAAAC		WLRSYLTDRKAHVKVG			
				TACTTACGCTGC		RARSHSYDISSGVPQGSV			
				CTCGTCTTTCAT		LGPLIFILFINDLAVRLKS			
				TCCAACGCTCC		GKLMYADDLKIYREISS			
				GAAAGGGACCA		TLDCCALQSDVNELISW			
				ACGAACGATCA		CTENGMELNIDKCKTITF			
				TCATACAGCCTC		TRRQTYVTHDYAIGNNII			
				AATCGCCGTCC		ERVFSIRDLGVIIDSKLKF			
				AAAGCCAACCC		TEHISTVTAKGFAVLGFI			
				GTTGCTACCAA		RRNSQSFRDVYTLKALY			
				GGATTAATCGC		CSLVRSVMEYAVCVWS			

				CAAACTCTGCA		PYHATHMLRIEKVQRCF			
				TCCGTAGTATCA		IHYCLRQLPWNDPTNLP			
				TCCATTCACGTA		DYVSRCRLIDLETLTSRR			
				AGGAAATTCTG		TKIQRLFVYDLLSNNIDC			
				TTTTGGTTCTGC		SELTNELRFYAPTRQLRE			
				ACTTTGTTTATG		RQLLVIRPHRTTYGQNC			
				TATTGAACTGTG		PMSSCLRAFNDVGCYFD			
				AAGCTCACCGT		FNVSKCTFKHRIKYLN			
				TGACAAGCCAC					
				CGTTTGTGACA					
				GAGAACACACG					
				CGCAGCTCGTG					
				GTCATTCCGAG					
				AATCAATAGTT					
				CACAGTATATC					
				ACCCCTTCGTAA					
				CTGGGAGCTTG					
				TTTCTCGCAGCT					
				AAAATGCAGGG					
				TGTTTGTGAAA					
				AATGCGCTTCC					
				GAACTATCCGC					
				TGATTTTTGGAG					

				CTGTGGTGGATT TTGTTCTGGTCG CTTTTGCTTCAG GTGTACCGGAA TGAGTCCTGAT ATCCAAGCTGC CGTTGCGAATA ACGGGTGCCTT CACTGGATGTG TAACGCCTGTA GCGGAATAATG ATTAAAGCTCG CTTTACTAAGGC AATCACATCAG TGAATGCTGCCT ATGAAGGCATA ATCGAATCAAT GAAAATCCGAA ATTCGAGACAA CATTTTGTCCGA AATCCGCTCGG AATTGCAGAAT AACCTTAAACG					

				AATTCCGGAGA TTGTTGCAAAC ACACCGATTCG AAATATTCAAC GTCGCACATTTG ATTCGGCTGTCC GTAATCCTGCA AAACGTCTTCGT GCAAATGATGA TGTGCCTCTCGA CCCGCCTAAAA AACTGGTCTGT GGCACTGATGG CCACAGTACTA GTTCTTCAGACG TCATCGCAGTTC CTGAACGAACG GATGAAAACAC GTTTTGGCTTTA CTTATCAGGAG TGTCCCCGAAA GCGCCTGATGA GAAGGTTATCG					

				AAATGGTCAAG					
				AATAGGCTCGM					
				AACAGACGACT					
				TAACTGTCGTTA					
				AACTTGTTCCTC					
				GTGGAAAGGAT					
				ACAAACAATCT					
				CACGTTTGTATC					
				CTTCAAAATAG					
				GAATGAAGCTT					
				GATCTTAGGGC					
				AAAAGCAATGA					
				CTTCATCAACCT					
				GGCCAGTTGGA					
				ATTCGTTTTCGC					
				GAATTCGAGAA					
				CCACAGTTCCTC					
				AAGGGCGGGGT					
				TCTGGAATCCCT					
				CGGAGAACCAG					
				CCCGAACCACA					
				GATCGTAACAG					
				TATTGAACTAGT					

				ATCCCCATCGTG TCCCATAGATCT TCATTCTCCTCG TGATGAGACCG TTTTAAGCCATT TGACAATCTACT ACCAAAACGTG CGGGGT					
CR1	CR1-64B_AAe	.	Aedes aegypti		TAGCTTAAGAAACA GTCTGTGGAATGTA ACAATTCGAGACGG TGACAATAAATAAA TAAATAAATAAATA AA	EQSIVCVKLQKSSIYICGI YLRPNSQPVLYSSHSAAI QQLSERISRFDSIVVVGD YNLPQLIWQTDDDINGL LPSNASSEQEVTLLETM VASGFQQINNIVNSNGR LLDLAFVNDNNDVELLE PPTPLLRMDNHHKPFVL RIDVNDSCTQVGDSNHA DDFDFRYCNFDELKSAI ATVEWATLFIGKDTDET VCTFYEVLHRILDEHVP RKRRRRTHPFKHPWWT SELQHLRNIVRKSRRRYF QSRTVENRDKLRLLEAS	cd01650	413	664

						YIDCQTASFRRYVARME TTAKEDPSAFWTFIRNR KRVNKFPAEMAFRDTV ANSPEGIANLFADFLENV HCTNPPTFSPDSLRNCPT FDLNLEPFDFAQQDVLS ALQKLDTSKGPGTDNLP PLLFKECADSFQIPLTIIF NKSLQSRTFPMLWKTAS VIPIFKAGSTRGVENYRG ISILCCLAKIFEELVHNVL YTASQPLISQFQHGFVK KRSTTTNLMTFTSFLSTE IENKHQVDAIYFDFSKAF DKVPHDLAISKLRHLGF PNWIADWLRSYLTNRK AFVNINGTYSRVISVTSG VPQGSVLGPLIFVLFIND LCFRLKSGKLFYADDLK IYRTIASHLDCCALQAD VNELELWCQENGMELNI KKCKSIAFSRRQSRTEFD YKIGSEPLERVESIRDLG			

						VIIDTKVRFNDHISVINA KAFAALGFVRRSTNDFN DIYALKSLYCSLVRSILE YAACVWAPHHTTQIVRL EKVQRSFIRYALRQLPW SDPVNLPDYPARCMLIN LEMLDARRNNLQRLFVF DLIMGNIDCPALLEDVQ FYAPLRQLRERDLLQIRR HRTSYGFNNPLSRCFRLF NCASALFDFNVSKYVFK NRLKDF			
CR1	CR1-8_AMi	.	Alligator mississippiensis	TA	TGATCTCAGCAGTA TTCCTGCCCAGAGC AGGGGGGTTGGACT CGATGATCTTTCGA GGTCCCTTCCAGCC CTTAATTTCTATGA TTCTATGATT	GIYYRPPHQEEELDLEFS RELMEAACSRDMVVMG DLNYPDICWEEHLARSD RSCRFLTCVEELYLVQE VNGPTRGKALLDLVLAK GDDLVGDLRIEGNLGDS DHEMIEFTIRRRAGKSAS RAEVLDFRKADFNELRK IVGETLWEEGLEGKGVQ EGWSFLKGTILRAQKRSI PVCRKGGKGAGKPSWL	cd01650	375	626

						TRGVKECLKKKKEAYG QWKQGVVTKEEYTSIAR ICRERIRKARAVTEIRLV TEIKDNKKSFFRYIGSRR KTKGNVGPLQNGTGQLI TDSQEKAELFNEYFASV FLHANGGRPPGRILDGC GRGDCQPMVSVELVRE HLERLNVFKSAGPDGLH PRVLRELAGVIAGPLAK VFEDSWCSGEVPEDWK RANVVPIFKKGRREDPG NYRPVSLTSIPGKILEKII KKSICERLAEGMMLKDS QHGFVAGRSCLTNLVSF YDQVTNRLDVGDAVDV IYLDFKKAFDMVSHDAL IKKLGDCGLAESTVRWV GNWLRGRTQRVLVDGS VSSWREVASGVPQGSVL GPVLFNIFINDLDEDVES TLAKFADDTKLWGKAV TLGDRMRIQADLDKLER			

						WAERNRMQFNKDKCRV LHLGRSNQRHEYRLGGD VLASTMAERDLGVMVD HRMNMSRQCDAVVNRA NSILGCISRCVASRAKEV ILPLYSALVRPQLEYCVQ FWAPHFKKDVENLERV QRRATRMIKGLEDKPYE ERLRDLGLFSLGKRRLR GDLVAVYKYVRGEYRE LGEKLFTRAPQGRTRSN GHKLLEDGFRLDVRKNF FTVRVTRIWNRLPKEVV QSPTLEVFKRRLDRHLA GII			
CR1	CR1-9D_AMi	.	Alligator mississippiensis	TCAGAACAAGA CACATGCTCGG GTGGGATGGGC TTCACCTCTCCC CCAAAGGTAAG CATGTGTTCTCT TTCAGGTTGGCT GA	TGACCCCAGGGTTG GTGACTCTTTTCTT GCCTGCCAGGGGCA GGGGGTCAGACTA GATGGCCCACTGAG GTCCCTTCCAACTC TATGAATCTATGA	SPPAGFKLGSSGEGEDE YGGSHGTPNQETPTRSA QVCSKDPIRQQFKGGPG TATRGIRCIYTNARSMG NKQEDLELLLASNNFDL VGLTETWWDSSHDWVV GIEGYRLYRRDRVGKKG GGVALYVKEQYTSSLIK	cd01650	536	787

						LGSEGGQAETLWVRIQG ARGERDLVVGVYYRPP HQDDELDQEFSKQLEEA VRSRDMVVMGDLNYPD ICWEEQSARSNRSRRFLT CIQDLHLTQEVRCPTRG NALLDLVLATGDDLVG DLRIQGNLGDSDHQMIK FTIQRRAGIIPSRARVLNF KKADFNRLRGLVSRVLG AKNLDALRVQDGWTFL KETILWAQSESIPWQRK GGKGARKPLWLTRGIQD TLRAKKEAYRKWKQGV ASKEEYTSLARTCRESV RKAKATLELRLASGIKD NKKSFFRYVGSKKRVRG SIGPIQDGQGGLVTDSRD KAELFNEFFASVFLNGN LDEPSTYITDTNKGNISR PTVSADTVKRHLEGLDV FKSPGPEGLHPWVMKEL AGVIADPLARIYERSWC			

						SGQVPEDWKRANVVPIF KKGRKEDPGNYRPVSLT SIPGKILEKIIKEHICGSPL ANVMRRGNQHGFLAGR SCLTNLVSFYDKATNCL DAGVEVDVIYLDFKKAF DKVSHTILINKLRGFALD DYTVDWVANWLKGRT QRVVVDGSVSTWKDVG SGVPQGSVLGPVLFNIFI SDLDEGVGSTLSKFADD TKLWGAANTLERRNQIQ ADLDRLGKWAESNRMQ FNKDKCKVLHLGRKNC HHTYRLGGDLLSSTEAE KDLGVTVDSRMNMSRQ CDEAVGRANRTLSCIRR CFTNRSREVILPLYAALV RPQLEYCVQFWMPQFK RDADNLERVQRRATRLI RGLQGRPYEERLRDLNL FSLSKRRLRGDLVAAYK FIRGEQHEVGDDLFTRA			

						PPGMTRNNGHKLLERRF RLDIRRKFFTVRVARIW NGLPREVVLAPSLGVFK RRLDRHLAGIT			
RTE	RTE-1_ATr	.	Amborella trichopoda	TAATTAGTGCTT ATGCACCACAA GTAGGATTAGA TGATAGAATTA AGAGACAATTT TGGGAGGACAT AGATGGAATAG TCCAAGACATA CCAAATGGAGA GAGGACCTTTTT AGGAGGGGACT TAAATGGTCAT GTAGGGAAGGA TAGTAGAGGAT ATGAAGGGGTA CATGGGGGATA TGGGTTTGGGG AAAGAAATGAG	TAATTGGGAAAAG GCTATAATGATGAT	MGETILDFAMAYDLAIA NTYFKKRDEHLITYKSG SHKSQIDLFLVKKVDRPS CKDCKVIPGESLTTQHR LVVMDVCIKTWKRRGE SNGCPRTRWWNLKNEKI DLFRDKLVEEGKWNSE HDVNVIWNEMADCIKK VAKEVLGESRGKRNFDK ETWWWSDEVQKAISEK RSCYKTWQCNRNDQNL EEYKNAKKNAKQIVSKA KLKAYDNLYNKLGTRE GEKDVFKLAKIREKKCR DLNHIKCIKSEDQRVLV KDDEVNERWRVYFQKL LNDEHNGNIVVEGTVGT SEGGGHRYIRRIRMSEV KEALKRMKIGKALGPDG	cd01650	353	603

						IPIEVWKCMGEIGIDWLT KLFNKVLRTKKMPDEW RKSIVVPIYKNKGDIQNC ANYRGIKLMSHTMKLW ERVIEQRLRHETSVSENQ FGFMPGRSTTEAIFLLRQ LIEKFRAKGKDLHMVFI DLEKAYDRVPRDLIWRV LEKKGVSRHYIDLVRDM YDGALTSVRTTSGETSEF SVTIGLHQGSALSPYLFV LVMDDLTRHLQDEVPW CMLFADDIVLIDETRNG VNTKLELWRDALESKGF KISRTKTEYMECKFSNN RSRDEEVVKIDGQEVQK RPHFRYLGSIIQENGEIEE DVAHRIRAGWAKWRCA TGILCDRRIPMKLKGKF YRTAVRPALLYGAECW ADKKQYTHRMSVAEMR MLRWMSGKTRRDRIRN ESIRENLGVAPIGDKMRE			

						SRLRWAGHVWRRPSTTP VRRCELVQVEGLKRAR GRPKRTWLEVVRKDMG TYGLTEGMAFNRAEWR TRIHVADPN			
L2	UnaL2	.	Anguilla japonica	CAGTGTGCATCT GATTGTGTCGTC GCTTCTGCCGTC CCCCGCGATTC AGATAAGCGTA TCTTAACTTGAT TTGTCTCTGCTG TTGCTAGTTAGA GAACATAGTTG TGCGTAATTTAG ATAATCTTTTTT AAACGTGTCTTT ACTGTTGCTAGT CAGCGAACTTA GTTGTGCGTTAG CTGAGAATCTCT GTAGTTGACTC ACTGTTGTTAGT	TAACCACCACCATT CTGTATATTTCACT CTTTAATTTAAAAA AAAAACAAAAAAA AAAAAAAAAACCC CCCCCCCCTTTTTC ATGTCACTTGTATT TGTCTTTGTCCTAA TACTGTAGCTTACT CTTCTGCCTAGTTG GCTTTGCACAGGTT AGGTTAGAATAGTG TTCACTGTGTGAAC TGTGTTCTTAGCTA GAAATAGCTGTACA AAATAAGTATTATA CCTTTCTGAACTTG TGTTCAGCAGATGC	MFINPIPLCALPRPKWSI HRRRNLSNMIYPPLSTPQ EFTVTGGLWNCQSAVK KADFITAFAKTKSLDFLA LTETWITPDNTATPAALS AGYSFSHTPRASGRGGG TGLLISLAWKFCVLPSPN PTSSSFESHAITVTHPSKF HIIVLYRPPGPLGSFLDEL DTLLSSLPEDGTPAILLG DFNLSPETSQLSRVASLL QSFGLSLSSSPPTHKAGN QLDLVFTRSCSIPQLTVT PLHVSDHFFLSFPLSYLH PLNSSNSSPTVTFRRNLS TLSQSSLISTALSTLPPPA SFSNLPVDLATSTFNSCL SQSLDSLCPLVSKPARSS	cd01650	513	773

				TAGTTAAACGC GTTAGTGAAAC TGTGTGTGGGG GTTGGTGTTTAA CTGCCCGGTATT GTTGAGCTAATT TCAAGTAGCTTC ACCTGGTGCTTA TCTGCCTTAATG AAGGTGATGCA AGCACGTAATT GTCACCCGGTA TTTATAGCTCCA GCGGAGGCTGC CATAGGCAGCC TCGTCGTCAGTT TGTG	CTACGACCATGATA TGCACTTTTGTACG TCGCTTTGGATAAA AGCGTCTGCGAAAT AAATGTAATGTAAT GTAATGTAATG	PPCPWLTESLRSSRTSLR AAERKWRKSRSSDDLTA YHTLLAAFTSALTSAKA SFFQSKISACVSNPRKLF STFSSLLTPPPPPPPSSLS ADDFLAAFEGKVATIRN SFAHPVSPPPPRDRPTAT LTSFEMLEDSDVLQLITS HRATTCPLDPIPSKVLQS ISSEFLPYLSSIVNSSLSS GHVPSDFKMARVTPLLK KPTSDPSDVTNYRPVSLL PFLSKTLERAVLKQLTSF LHQNNLLDPHQSGFRSG HSTETALLGVTEALATA RASRLSSVLILLDLSAAF DTVNHKILLSTLAGMGI AGTALSWFASYLADRTY QVTWKGSASASHPLVTG VPQGSVLGPLLFSLYTRS LGSVINSHGFSYHCYAD DTQLFFSFPPSVTQVNDK ISSCLADISTWMASHHL			

						KLNLNKTELLFFPCKTSL LRELSITVDGTTVTASHS AKSLGVVLDDQLDLKE HIKATSRSCRFLLYNIRRI RPYLTTHSTQLLVQATV TSRLDYCNSLLASLPAC AIQPLQMIQNAAARLIY NLPKFSHVTPLLRSLHW LPVAARIRFKTLTLAYTA ANRTAPIYLQDMIRFYV PARPLRSAEAGRLVTPPT RPKGSQSFSTLAPQWWN ELPVPLRTSPSLPIFRRGL KTHLFRLYLD			
R4	R4-1B_AC	.	Anolis carolinensis	ATGMTGTTGAG AGCTGTGCTGG CGGTAGTGTAA ACTACCAGCAG GTTCAACCAAG CCAGAGTGGTC TCAACTGAGAA GTAGAACTAAG	TAGACACTTGGGAA GTGTCCGACGTGTG ATCCAATACAACAG CCAGCAGAGTGTCT GCTGTGGACTCATC TTGTTGTGTTTCAA	MVKQTLKKSILARSSPRI IRTTASATPGHLSTSGLQ EPGPIAEQPGHGPVGQLE AKQLQKWQKFSMPQNR TVMNCYYRSEPQKRGY QKRMYQLWKLEYPDST VTEQRLADQRRYITKKK VLSEVELEEIKSNCETEE QTTVETAAEDLMTSEPT	cd01650	476	742

				TGCACCCAACC CATTTAAAT		EHMEPEENAGIFQDDEH EEPVAIAYTTPGTLTMK QQELKENFLTHAALKEG RQRLPTLKTVPRKHLAP LMNDVNAALATVEITT MEETNQLAYSTAVIVTQ ELGIQQLKPARKMIGKP KWKIRLELKIRSLRSDAS NLKNMKEQKLKNVKTK NYLIKKYWLNTRKIEEA LEIVKQQITATARKIERY EARVIQYKQNQLFQSDQ RRFYQSLNQQVAVKNE KPEKTATNAFWKDLWE TEKGYNKKAEWIKDFEN KFSENKMRTLEITTEMIT ERVKKVKNWTSPGNDQ LHGFWLKYLSSLHSKM AEQFNEMLQTGNISKWI TTGKTYLIQKDPAKGAIP GNYRPITCLPTMFKLLTG IIADQIQDYLEENNILPVE QKGNKRRSRGTKDQLLI			

						DKMILENCKSRKTNLYT TWIDYKKAFDSLPHCWII KCLDVIGVCENIRKFTEN AMKQWKTELFVGNESY GTVNIRRGIFQGDSLSPL LFIIAMVPLSVILQKTNL GYQTARNSPKISHLLYM DDLKLYGKSETEIQSLTN TVRIFSTDISMEFGLDKC ATVALKKGKITHSEGIE MPNGQTIKCNQPEAYKY LGILQLDNIKHGHVKNV VSKEYIQRVRKVLKSKL NGGNTIKAINTWAIPVIR YTAGIVNWTQAELDDL DRKTRKLMTIHYSLHPR SDVDRLYLPRKSGGRGL LQVKQTVEEEKHALAD YVKGSQEPTLREVNSSK LLQVQKTKSEYHKNTIQ SRRENWRKKALHGQFL GKIESQIDKEKTWLWLT NGTLKKETEGLILAAQE			

						QAIRTNAIKAKIEKSTTD PKCRLCKEADETIDHILS CCKKIAQTDYKQRHNTV AQMIHWNLCHKYHLPA TKNWWDHKPEKVTENE HVKLLWDFRIQTDRVLE HNTPDLTIVLKDKVWIV DVAIPGDSRIEEKQLEKL TRYEDLKIELQRLWHKP VKVVPVVIGTLGAVPKD LGLHLNTIGADKITICQL QKATLLGSARIIRRYITQ S			
L2	L2-19_ACar	.	Anolis carolinensis	ATGCATTACAG AAACCTGGATA CGAGAAGGAGA TAATATTCCCTT GCATGAACTAA CACCGTCAGGT TTCCAGATGTTT CAAAGATCTCG AAAGGGGGGTC GAGGGGGGTGG	TAAATCAGATCTAA TTTGCCAAATAGTC ACTGTTGAATGTTT TTAGTTGAATGTTT TTATGTTGAATGTT TTTATATTGTGGAA TGATATTTTAAATT GTAAGTCGCTCGGA GCACTCTGGTGGAG AGCGACTAATTAAG	MGIECIGLTLAIRESVTIL LVYRAPGAPNENIPKLL DLVADWALKFPSLIVLG DFNVHADRLLSSSIARDL VQSMETIGFSLMVTGPT HRAGHTLDLIFSIGIQVT LSSVLEIPWSDHYILNTQ IKSSPPQHPETEQIFARPR KLMDPIRLSSVLKELDY TADSLHERVENWNKNL	cd01650	388	648

				GAATAGTGGCT CGCGATCATTTC TCCATTAGAACT ATTCCTACCCCT GATATT	AAATAAAGTGAAG TGAAGTGAAGTGA AGTGAA	STALDIVAPKRAFRPRRN RSPWFTDALRQMKLER KRLERKWRKKPDETSKL ALETHQKIYEAATREAK KEYNSSLIASARSRPAQI FKIVRNIINPTHPAKDMA SQDITAETFQSYFINKINL LREGLPPTEDTLRELEAF WPSSDKLLELFVPLEME DLEKTLAAARPTTCSLD PCPSRLVKECWEGLRDP LRDIINSSLEQGIFPECLK EAVVHPLLKKPSLDAAIP SNYRPVSNLPFLSKVIER TVAGQLQQFLDETAGLD PFQSGFRSGHGTETVLV AITDEIRRQLDRGGSALL VLLDLTAAFDTVDHDLL IHRLAMSGVRGLALXWF NSFLRDRRQRVEWSGQS SESSCLCCGVPQGAILSP LLFNIYVRPLAGLVRSFG IECYQYADDTQLLLRLE			

						PGTTTIPENFKLCLEAVM SWLRASRLKVNPAKTEI LWPGQPPELIQSLPDFDG EVLVPSATVKSLGVVLD SSLTMEAQITAISKQAFF HLRQARKLASYLSTEAL ATVIHATVTTRLDYCNA LYAGLPKTKTQKIRMVQ NAAARLLAGSSIRSHITPI LKQLHWLPIEHRISYKILI LTFRARNGQGPLYLRDR LIPFFHQWSPRSSQENLL YVPGPREVHLEATRRRA FSTYAPILWNSLPPYVRA MSELRPFVKALKTWLFG CAFK			
Penelope	Poseidon_Ac	.	Anoliscarolinensis		GCTTTTCAATGCTA ATTAAAGTGATTAA CTACACAACATTCA CACTGACCTCTCTC ACCCTAGACTTTCC ACAGATATATATTA ACCTCTTTGCTTAG	MPAIDVGETSGENASGT WPHSPKDIQQPCDPGHE SLRQHIEHLYGEKLFQD TRRLEKLRTRKAHLLCS LTFLLRCRDTDTIPQFLA AKRTFKTPQAHRIYNRL ERNLLRERIHTIRKELAN	cd00304	470	554

					TTTTCTCCATACCTC ACAACCTCTGAGG	TDKELLHLHINISQKMN TQDWDKIDSLTYKKME KDMVLHTKRQKQKFDK CRKQQPKPELDKSRTIIN LTDRQLTEDQVSILEKG GNFAVTTTRIPVENIIAN VESAIYQLPEEEAEEVR METARILRNAKLPPSNIT RKERQAIKDLNSDPEIIIL PADKGNATVIMETKQY KEKIRQLLDPTIYKKLKQ DPTNKITRKTNTLIKNSSI NFDIRQQLCKSEALPPRL YGLPKIHKDSIPLRPIVSA IGSPTYNLAKFLATQLQT HIGLTAHYIKDSTHFIEKI SNLNLSTKDILISFDVVS LFTKVPVADTLTLIKQNF PEDITALFHHCLTTSYFQ WDTGFYEQKDGVAMGS PLSPVVANFYMEYFEKQ ALETAPKKPTVWFRYVD DTFTIWSHGEEELSKFLD			

						HLNSIHPNIQFTMEKEKE GKLPFLDVLVIRKPNQQ LGHTVYRKPTHTDRYLH KNSNHHPSQKRSTIKAL TDRAQRICEPHLLQGEL NHLNWALQANGYSTTDI RRAARPRTSHESQDKDP PRGKVFLPYIKGTTDRIG KLMKKHNLQTIYRPTKK IQQMLRSAKDKRDPLSS AGVYRIPCSCGQVYIGTT KRSAQTRVKEHERHCRL IQPEKSAIAEHLMNQPG HRILFENTKMLDHSNNY HVRLHREAIEIHKHVDN FNRKEETMKMNKIWLP VLQNSKIKTVNKKQYSE NRGFPDMNQPRAVNDS KQRMPQRQEEDSR			
Vingi	Vingi-2_Acar	.	Anolis carolinensis	GGGGGACATGA GAGAAGCCTCC ACGCAGGATCG TGACACATCCG	TAGTATTAATGCCA AGTTTTTGTTTTTTT TTCTTAAAAAAAAT CTCTATGTTTGCAA	MTIISINIEGMSAAKEQL LYELCRDKKCDVLCVQE THRDKQQKRPKVPGMIL VAERPHAQYGSAVFVKP	cd01650	459	714

				GGCGTCCCCTG GGCAACGTCTTT GTAGACGGCAG ATTCTCTCACAA CCAGAAGCGAC GTGCAACTTGC AACATGCAAGC AAACAAACCAA CCAACGGACCT TTTCAGGACCA ACT	ATCCATTACAACTT GTACCCTCGGTTTC GCTTCTGACACGAG AAATAAATAA	GLQVSSVHNTEENNIEVI TVELNNITITSVYKPPNE EFCFSSPENFDSHQRTVV IGDFNSHSHVWGYAQED KNGEAVLSWAELHKLSL IHDSKLPPSYNSGRWHR GYNPDLLFVSDSIVQQCS KSVGPPIPNTQHRPIVCQ LFPTIRPLEVRFKRRFNF RKADWSKFSDMLDDKII SVAPIPEEYERFIKIVKTC SRASIPRGCRTHYLQGIT PDTAALLEDYYRLHNED PFSEQTLQTAQSILSSISE AKKEQWIKLITEVDMGR SSQKAWRLLRRLSNDPS QTNTHANIKADQIAHQL LKNGKPKLSTKVGRKPI VRQPDNENNHLHEPFTS TELDMALNKCKNGKAA GLDDLRMEQIKNFGPKA RCWLLKMMNNCIASCQI PKIWRKARVIAILKPGKD			

						RNDPKSYRPISLLCHLYK VLERLILNRIMEKIDPCLI PEQAGFRKGKSCTSQVL NLTQHIEDGFERQQITGA VFIDLSAAYDTVNHRLL LRKTYNITKDYHLTRFIG NLLQNRSFFVEFQGQRS RWRKQKNGLPQGSVLA PSMFNIYTNDQPLPEGTE SFIYADDRAITAQAGSFE MVEQKLSEALGALTAY YRENQLIPNPSKTQTCAF HLKNRQASRALRITWEG IPLEHCSTPKYLGVTLDR ALTYKKHCFTIKQKVGA RNNIVRKLTGTTWGSQP DTVKTSALALCYSAAEY ACPVWNTSHHVKTVDV ALNETCRIITGCLRPTPLE KLYCLAGIAPPDICREVA ANNERTKALTSPAHPLF GYQPARQRLKSRNSFLR STEILAGTPQQARVQKW			

						QAKTRNLNQWLTPDER LPPGHTEDWATWKALN RLRSGTTRCRANLKKW GYKVESATCDCGEEQTT DHLLQCILSPATCTMED LLTATPEALQVASYWAK DI			
CR1	CR1-9_AG	.	Anopheles gambiae	TTGACAGTTGA CAGCTGACGAC WGAGTGTGTTT TCTCACTGTGCT CCGCGTAAATC CGTGAATTATTT TGTCGTTTCGGC CTWATTTTCGTC AAACTGCCGCT TAGCTGTGACTT AAGCTGTCCCT GGTAGTCCTGC CTTCAATCGTTC GTCACAACCAT TCCTGACTTTGC TGTATTTCCTGC	TAAATGTTCTGTTC GTTATCTAATGTAA TGTATTGTAAACAA ATTTTGACTCGAGA GGGCTTCATAGTCC ATCGATTAATAAAC TAAACTAAACTAAA CTAAA	MEGVCCDCSVPLESLEC GIMCSFCDAVYHSACTK SPFTVIEEVKRTASLHWS CVGCSNALGNPHSKAV KTTGMQVGFQAALTAA VEVMKAALVPPVMHEIR EGFASFAAAHQAPSEHF NVPPPDALPSGKRRRLF RDVVASTEAIVVDNAPL YPVNTDTDNTHQRNRPS LPPIITGTDTTTTSIPTVS QLPRTNYMWLHLSRLA KTVTVDQVVSMVKSQL DTTDVIAFSLLKAGTTIS SVSSLTFKVRIIAALRDK ALTAGSWPVGLGVREFI	cd01650	909	1158

				GTTTGTTCCTGA GTCCTGTGGCAT CTATCTGCAAC GGCAACTAAA		SLPQRPSHHRPDLTNTNT PIPVRHSPMHPVQLSPD MNCTLISPAPQKQLQPH QTPLKCTKLLFNTFFLSN NHTTVSSKGTVPLNLTS HLFTRNNNTSTTRSLTD DLAPTYTRIPLNVRSTSIT VYYQNVRGLRSKADEF RLTVLEADYDVVVLTET WLDPSLPTALLFGDAHC VYRCDRNAANSSLSRGG GVLIAVSSALRSHALPVL AQTLEFACICIQLPTYRL YIAAAYLPPNHSTDEGKI NALLDTVNSICDTLGPN DRFVLVGDFNQPALSWS PAQSNHETPFVFFEPHTS SVRSAQFVDGLHQNALY QLNNNTNSMGRILDLAY GNWSSAATSSLIRVSEHP LLPIDCYHPPLEFDLEITA SSMNHATTVTEIKLNYA RADLTKLKRLICSFNQSF			

						ECSNYTSIDHATNDFSEF MRAALRECAPIKKPHRG PPWGDRTLHALKKAKR AAYDYLRANRSTPCRLY YNGAHALYRRYNRVCY RRYIRQTERSLRKYPRRF WSFADNMRKSTGLPGTI RYKDNCAHSPLDICELF ATRFEDSFISNITPPEAVA SALVNTPADAIHVTLPIV TETLVTQKIERLKSSYSE GPDGIPATILKRCAVSVA PVLTKIFNESLRTGTFPSL WKVSWLTPIHKKGSKNE AVNYRGITSLAACAKVF ELIIYEPLMASAGKYIST NQHGFMPKRSTTTNLM QFVSGCYKSIDDGMQVD AIYTDIKAAFDSVSHNIL LAKLDRLGLSPPLVKW MKSYLCGRSYAVRMGS HQSRSIGASSGVPQGSNL GPLLFLLYISDLCTALPD			

						NSCLLYADDAKIYREIRE PEDHLQLQATLTEFVSW CKRNALSVCIDKCAIISF SRSRAPALFNYTLDGQN LERVNCVKDLGVLLDA KLSFEEQIDQVVASGNR LLGLVINMTRELRDPMC FKTLYCALIRPLLEYASI VWWPASTRALARLESIQ RKATRFALRDWPRRLD YRTRCLLLGIPPLAERVE HTRLAFITGILNGTIDCPE LLSRIHLYVPARILRRRP MLAVAETRTTFGSRNPL LCMCRLLNSANDIYEPG MTITELTSLLSVRNAFNN NNI			
CR1	CR1-10_AG	NT_078268.2	Anopheles gambiae		TCCTTTATAATCCTT TTCCCTGATACTCT TCCCTATCCTCTTTT TTCTCATCCTCTTTT TTCCGACTATCATA TCTAATATTATTCA	SPRIPRIFGLQCSPGFSST TAAAIPSQAINRCDPRTR ITSVYPYRFSAANFPPVD EPAATTAQSNSTSTPITP LHHAAERFETLPSVASPI TVSAPLPLHASFPHPIDD	cd01650	634	882

					CTCTTCTTTTCTTAC	LRFYYQNTRGLRTKLDD			
					GTTTGTATTTTATCC	LRLASADTEYHVLILTET			
					TAATATCTGTCTTT	WLSENIPSSLAFDNNFVI			
					CATGTGCTCAAAAG	YRCDRNSRNSSRSRGGG			
					TTCGCTTTAGTATT	TLIAVSRKLAACEISPAS			
					AGGAATTTAGATAG	DSLEQCWIRLKLPSCSLF			
					GTTAAGTTAGGAAT	IGSIYLPPTLSADASTLNT			
					TAGTTAGATTTAGG	LDASIDEIISRMKKNDRL			
					AATTGTGAAACCAT	LLFGDFNIPSITWHASPD			
					CAGGCTGACAATTT	PAPPFLPLLVPASSAHSS			
					GAATAAACTTGAAA	SLFIDIVRYHSLSQLSSVS			
					TGAAATGAAA	NFRQVQLDLIFTFPKNQS			
						NATSTCTYVSQAPVPILQ			
						IDPHHPPLELSLKITGSLT			
						PSNRSQHSSSRLTTAPSL			
						DFRKTDFNLLNTSLSNID			
						WSQLYSCKDVNLAVSTF			
						TSLFLSTLPQCCPKRNPP			
						RSPPWSDATLKSLKRDK			
						SRYQRAYWSTRSVNAL			
						RTYKYAAAAYRLYNRT			
						RYNGYILILQTRLRWYP			
						RAFWKFMNARRKGSSL			

						PSTMHLESVYASSPEDIC			
						SLFRNHFSNIFSPPSQIPN			
						TASQCLSYTPQDVIDVSC			
						FGFDEHSVSEALSKLKPS			
						FSAGPDGIPPSLLKKTAE			
						VIAPLLSFIFNLSLRLCCF			
						PSSWKSGWLCPIFKKGN			
						SADVSNYRGIVLLPACS			
						KVFEIAVHSVISFQVKSW			
						ISVNQHGFMPSRSTVTN			
						LMQFASHVIPSLDSGLQ			
						VDTIYTDFKAAFDTVPH			
						HLLLLKLKKLGFSNMLI			
						DWLYSFLNSRSIRVKTD			
						TTFSSSFCPSSGVPQGSV			
						LSPLLFVLFVSDVHAILP			
						SEGHLLYADDLKIYLTIS			
						THSDCLRLQEFLNSFSL			
						WCSYNQLSLCPEKCSVIS			
						FCHNRNPIHYEYSLSNSV			
						LSRTSLIRDLGVLWDDK			
						LSLNEHLESVVASASRTL			
						GLICRLTRDFRDPLCLKS			

						LYCSLVRSTLEYASVVV GPVRSGWILRFEKVQKR FTRFAFRRMAGPNAAMP SYESRCNLLGLESLEVRR SQARVLFIGRMLNGLID APTLLEKIGLYAPSRNLR ARSALNMVSRQTTFGSN EPLLLCVRTFNTSTDPPH PRL			
CR1	CR1-8_AG	.	Anopheles gambiae	TATATATCTTCA ATCGAAAGGCA TCCCCGAAAAT TTTGGCGCTTCA TGGACAGCAGA CGACGCTCCAA GCACAATCATG GGCACGTCACC CGAAGATATTT GCAACATTTTCG CTGAACGATTTT CCG	TAATTTTATTTTCA AAATACATTTTCTA TTCTTGTGCACCTG TTCTCCCCTATCGA TTTATGTACCTTAC ATAGTCTATAAGCA CTAATTTAAGTACA TTCATGTAGGATCT TCAAGATCCCGATG GATTAATCAATAAA CATATTAAACATAT TAAACA	MPSPIVDQTVISRATSYT PLNMIDYTIPTFDIRTVE KCLKSLKPSTAPGPDGIP SYIISKCYQSLAPVLTNIY NTSLSSGIFPRLWKTSW MVPIFKKGDRSCASNYR GITSLCAVSKAFELLLYG PMLSATSNYIMDNQHGF VPKRSTLTNLTEFVSFCK RNLDSGGQVDAIYTDLK AAFDLISHDILIAKLQKL GFSEQIVKWLHSYLTER SYKIKVNDHTSREVLGT SGVPQGSNLGPLLFILYI	cd01650	92	340

						NDVGQLLAEHRFLLFAD DVKLFAPINDTNDCINLQ HAINLFSIWCRDNAMVL CIEKCRVLSFYRSRSCTR FNYQIDNISVERTDTFRD LGIILDTRLTFNDHLENI VSRGNQLLGWIIRSTQGF RNPMTIKTIYCAYVRSV LEYGSIVSDPCTEQWSN RIEAIQRKATRYAVRLLP WRQGDVLPSYHSRCLLL GIQSLKKRREIAKCLFIS GLLNHHIDAPTLLATIEF NAPSRNLRTRQFISVPRY RTRFGQSDPMSAMCRTF INNFDLFDFNIPQNVFRD RLRTQLT			
RTE	RTE-1_AC	.	Aplysia californica	TTAGCCGTGCA GTCTCGGCATTG AGAAGAATAGG ACTGACGTGGT GCTAGCGTGTC GTACGAGGCGA	TGATGATGATGAT	MMQGGESRKEAAGQKP PPSSKSKSLRIGSWNVRT LYEAGKTAQVLQEMHR YQLHILGISETHLNQSGQ KRMKEGELFVYSGNDS DGPHREGVGMIISKTAQ	cd01650	562	819

				CTAAATGGCGA		KTFRGWEAHGPRIIMAS			
				ACCAGGTAGTG		FGTKKELKVRDGRKKVI			
				ACTCCTGAAAC		NMNIIQVYAPTNDAEVE			
				AGAGTCATCCT		NKDDFYNRLQNVIDKLP			
				GCCTCCAGGTC		GKDINIVMGDLNAKVGS			
				GGGGGTTGAGT		DNSHNEEIMGRHGLGEA			
				CACGAGGCTAA		NDNGERFQNFCTFNNLV			
				CAACCTCTCTCG		IGGTIFPHKRIHKATWVS			
				TAAAAAGAACC		PDGTTENQIDHFCISRKF			
				TGTTACAGAAA		RRSLEDVRVLRGADVAS			
				CACATACAAGC		DHHLILGKMKLKLRRYA			
				AAAAACCATTA		QVAAGNRLKYEVSLLK			
				TCTCTGCTTGGC		DPVKKNEFKLELKNRFQ			
				TGATCGTGGAT		ALQDLDECEVNEHWDK			
				CTTCTCCAAATG		LREAVNATCQSTLGPKD			
				GAGAGTTT		NTRNEWISQDSLDKVKR			
						RRELKAQVNNSRTRTEK			
						LAAQQEYTEVNREVKNS			
						IKKDKNTYLNGLAEKAE			
						KAATDGHLRIVHQTTRT			
						LSGRFSKPTLPVKDKEG			
						KTIFNPEGQLNRWREHF			
						DQLLNRPPPENQPEILPA			

						RNDLPIDTEPPSREEIAA AIKTLKSHKAAGPDHIPP EALKAEKDTFVEILYALF EKVWQKEEIPSDWKEG HLIKLPKKGDLSNCNNY RGITLLSIPGKVFNRILLE RIKLATDEKLRGNQAGF RKNRSCVDQIATLRIIVE QSLEWNSPLVVNFIDYE KAFDSVDRETLWKLMR HFGIPEKIVKLVKATYD GTSCQVFHEGQLSEPFR VQTGVRQGCLLSPFLFIL AIDWVMKEATRGKRNG IQWIPWQQLDDLDFADD LALLSHTHAQMQEKTKE LNSLSRSVGLRIHPGKSK VVKLGKVSNLAVTVED TPLEEVDSFTYLGSVIDG KGGTEADVKARTAKAR SAFKQLSKVWKASNISL RTKIRLFNSNVKSVLLY GCETWKTTKTVLNKVQ			

						TFINRCMRSILRIKWQDR IRNEELWERTGQKDIEV EIGLRKWRWMGHTLRK PSGNITRQSLQWNPQGA RGRGRPRETWRRCVER DMREMKLTWNELGRM AQDRDKWKELVCGLYP ARGERH			
CR1	CR1-1_AC	.	Aplysia californica	AATTGAACAAA TTATGGAGGAA CTGGGAGTGAC TCGCAAACCCTT TGATATCAGGA GATTATTTGGTA GGAACAGAACA GAGGATAGCAG AGACAGCGGTA ATGACCGGGGA AGCGAGAAAAG GCCAAGACCTA CACTACTGAAA TTTGAAACTGA AGAAGCAAGTG	TGATTTGTCTAGCC TGCACGCAATGCAT TATTGTCGTTCATT ACCAGATAGTTGAA CAGGCTCTACTTAG AGCCTCAAACAACT AAATCGTCAATGTC AATGTCAA	MSLKSMVEIYQPDIIAVT EVKPKYVRFKIQESELSL DGYDYELFHNLEEEGRG LVLYVRKALKPGLCDDL DNDFCEKRFVKCVLKD GVSLLIGLAIYRSQHSPT AEEDARKLNKLFELVSL RKETHKIIMGDFNFRQID WETETSNAGEKHIATKF LKATKDAFFIQHQKTPT RYRKGETPNILDLVFTNT EQLVTDIKVEAGIGKSD HLCLLMTLSLSAGTAPK VTRRNFRKTDVEKLRED LGAYDWESELANKNVE	cd01650	479	737

				CTAATACGATG		DTWSVIKSKIEGAIGAST			
				AGGAAATTTAT		PLMNISGRKNKRYMDK			
				TGAGGCAAGAA		DTLEAVRKKHRAYRKW			
				AAGACGCAGCA		HRTGKTTDEIAYNKANN			
				AAGGACAAAAA		KARKACRKANRNYEKK			
				TGAAGGAGAAG		LAIEAKKNPKVFYSYAN			
				AGGACAAGTTA		SKIKSKTGIADLNKDDG			
				TACTTGACTATT		TKTKSDTEKAELLNKFF			
				GGGATGCGAAG		QSIFTNEDTGPLPDFEEY			
				AGATATKACAA		DYNTELQDFNIETVTVK			
				AGCAGGAGAGA		ELLVSLDVNKATGPDGI			
				GAGGAGGACTA		PPRVLSQAADILALPLRI			
				TGCCCTCTACCA		LFRRTLNDGVLPVEWKT			
				GGAGTTGAAGG		AFVSPIFKKGNKAAVNN			
				CTAAGAAAGAG		YRPVSLTCVICKVMEKL			
				AGGTAAAAAAA		VRGHIMDHIVNNMLISK			
				CTTGGGGGACA		DQHGFVTGRSCITQLLE			
				ATTTCGCTCACT		ALDDWTSVMDEGGSVD			
				GGGTGAGGAGA		VIYMDYQKAFDSVPHRR			
				GACGGCAAAAT		LIQKLSACGVKGRVLKW			
				TCGGAACATTG		VRDFLTSRKQRVVINGS			
				GGAAATACCCA		RSEEAEVTSGIPQGSVLG			
				ACGGAACGTCA		PLLFVIYINDLPRGLKTT			

				GAAAGAGAGGG		AKMFADDTKLYVRSDT			
				AGACAATACAA		VNGPQDLQDDLDTLQE			
				GAATAAAGCCA		WSQRWLLRFHPQKCCV			
				AAGACAGCAAA		LKVGKDNDNEYFMDQA			
				GAATGTAAATA		VDGSTSRHKLNVTKREK			
				GTTTTGGCCATT		DLGVVVDNKLTFKEHV			
				GTTCGCCTTTTT		AQMTSKANRIVGLIRRSF			
				CATATTTTAATT		DYLSEDMFVLLYKSLVR			
				ATTGCTCATCTC		PLLEYAHCVWQPFEKGL			
				GTTCCTGACGT		CSEIEQVQRRATKMLSH			
				WCAGAATAGAA		LKEKPYSERLKKLKLPSL			
				GTTTCACTTCAT		EHRRLRGDMIETYKYIK			
				TTTCTTCTTTTT		GYYITERPAFTKATTTQL			
				CACCTACATTTC		RGNSEKLQKQRSRLCVR			
				ATTCCAATCGG		ANFLENRVNTIWNNLPE			
				AATGCACATAC		SVVSAPNVDTFKRRLDE			
				TTTAGAGAATT		YWKDLPTIYNPTCQN			
				GTACTTGCTTGT					
				ATACAAATGCA					
				GATCAATTGTTG					
				AACAAAAGAA					

Tx1	Tx1-1_AC	.	Aplysia californica	TCAAGCGACAA TCCGTGAGGCT AACGCTGACAC ACAGGCGGAGA CACAGGCAGAG ACACACACCCT GCGTGCCGTAA GGCAAACAAAT ATGGACTCTTTT TTCCCGCAGCCT GACCGCTCCCG CCCCCCAGTAC GCTCACCCCGA AGCCAGGACAC AAGGTCTCCAA GTGTCAAGACG AGGAGACATAA AAGACAGGTGA GCGACGAGCCA GCTCGGTCACC GAGCGTGAAAN CTCGAAAACAG GAGGTGAGTGG	G	MSIRIASLNCRGLNKKL KRRAIFRKCTQYDIICLQ ETYVTNTTAREWSLEWS GDFFYSEGTNKSKGQIIL VNKHFKYKTVELCIKTE RILGLSFSXDDVSYVAIN VYGPVVKNDKLDFIDKL YSVLKDEKADHTIICGDF NVVHNNTLDIVTGEKHD DMSTQKFRDWIANCQIV DTWRGVHGEEKDFTWS RPTPFTARRLDYIFCNTE LFPFVIDSKHAFVYGTD HKAVITTVRTEKFKRGP SYWKFNNSLLKDKEYV NSMEKCIDEYLERTLDE TDKTIRWELLKVMIKSH TIDYCIRKCRALKLRENL ITEELLKLNDQVVKNPQ DLTLLKQFESKKKELEIY QLHKARGAQIRSRQKWI EEGEKNTKYFLRMEQM RGGSNVITSLQDDKNNN	cd01650	507	772

				CACGAGAAGAA		LYGHVDIVDSIKNHFERL			
				GGCAAGACACG		YKKDHNIETNRVDDFM			
				GTAGATTATTTT		QGCSFPVLDDVDKEICD			
				GATTATGAGGA		SELSIVELDAIVNKLNIDS			
				GCCAGTGACTA		APGIDGITTPFYKFFWSK			
				CCGAAACGTGT		LRKPLFTNFQQCLETGEL			
				GAGCACGAAGA		SPTQRRGVISLLHKGKN			
				GGTCGGTGCGT		LRRDNIKNWRPITLTNT			
				GAGTGGTGTGT		DYKIFSKILAIRLQRVMP			
				ACACGTTTCGGT		VLINENQTGFLKGRSISE			
				AGCCATTGGTA		HIRTLDDIIYVASEKNLP			
				ACGACTTGGAC		GMVVSLDFQRAFDTVE			
				TGGTAAAAGGA		KGTILATLKKFNFGEGFI			
				CTTTGAGTTTGG		NMVGVLMANTECCVQN			
				TTTTGCGTCTGA		GGWLSEWLNTERGIKQ			
				CTGTTTTGG		GCCTSPLLFVLVAELMSI			
						KLRNNSSISGMSTILSGEI			
						INFCKVLQYADDTTLILK			
						DQEDLASALLDIDSFGSI			
						SGLKLNKDKSQGMWIG			
						SDRTRTEKLGGISWVPN			
						GNTIKILGIHFSATEEASR			
						IELNWIDKIENMKNIMK			

						MWMKRNPSLYGKVIICK TFLLSQISYIIQSLSLPER VLNTIDTLFFKFLWQKR HSNKKAREKVKRTVMC RSIEEGGLGVISVKDQQK VFLLNWLKRMATVTEG RPNNIPNRISDLAFTEVG GCDYFFSATANCKKTGK LCHASKFWESVMDTWIE MASIKKEPSSTIEILMEPL FNNRNIKYKGKSLYLKR WMNSGIKYVYHLYDRRI LKSKEELSHITGTYAAFD FDYYALMNALPTQWKQ DVSDNNTEELLVNTREV INTNGKKTVLEILSKSNR EIRQLLKYRRHAEICGRN FWMRKMNIDITCKYNIA HRSTKESRLRLLHFKILH NIFPSNILLNRLGIKNTEL CEHCGVKDVIEHMFIHC TLLKNFWKYVFHTICLK			

						TSSRFPMTDCNILFGFSE RGQG			
Hero	HERO-1_AC	.	Aplysia californica	GTAGAGAGGAA GAGCCCGAAAC AGCGACGCATT CTCTCCTGATGA TCTCCATGCTAG AGTCCGGGATA CGGAGCGTTCG GCGGTGAGCAG TGCATACGGCT GTTTAGACTAC GTACCTACCAG TAGCAGGTGTA TGCTACGGACA TCCTGGAAAGA CAGCACGACAA CCCGGAGTGAC GGGTGGACATC GGGAAAGACCG AGGACGTCGGG AAAGACCGACG GTGGCAGGGGC	TAAAGTTGGCAGCA GAGGAAAACAGTC GAACTAACATGGA GACCTCACATCGGT GAGGGGGCTCGGT GCGTCAGAAAAGCT GTAGAGAGGAAGA GCCCGAAACAGCG ACGCATTCTCTCCT GATGATCTCCATGC TAGAGTCCGGGATA CGGAGCGTTCGACG GTGAGCAGTGCATA CGGCTGTTTAGACT ACGTAC	MDNDLSVKRTMQKENQ ENSFMCEQCYRKFDTW RGLRVHQGKVCLKKRR QCRSSDCKTRSKSSQDE NHSGAISASVDSPFTLSQ QEMEQQTEPGLRKPRIL WPAANEKAKYKTLEEE VQKKLKEVKKSSNRGD AKEILSAFATIIYDSAIKE FGICEGKKKTEKKSGLS RRQRVLAQLRKQKRDL QKQKKYAPPEEIEGLRIL FEDLKKKSRDLQRHERR CTRRKERKREREQFLKN PYEAAKKLFTEARSGKL ECTKEELDAYVKQTYSD SRREEPLPYMRGLKRPA APVVKYYLGPIREKEVD EFVKKARAKSAPGGDG VSYKVFKYCDTLRHTLF EMLRGLWKDKEIVKDW	cd01650	355	614

				TAACTACCCAG		CRAEGVYLPKEQNAKEI			
				GTCACTGCAAA		GQFRPISIINVTCKIFMGI			
				TNTAAGTTACGT		LAKRTVAYLQSNGYVD			
				CAACCANTAGA		ETVQKAGVPGIPGCIEHA			
				AAAGACAGCCC		FTIWDTIQEAKKNKESL			
				CGGGATCCCCC		NVVWLDLANAYGSVPH			
				GAGAGGGGGGA		KLLMKAMDFFYVPQEV			
				GTAGAGGATGG		QNIMREYYSNFQMRFST			
				ATCCACAAGGA		ANFTTEWHRLEIGIAAG			
				CGGATTCA		CSISVIWFILVMEMLLRS			
						AECCEEEAKVRSPKKAF			
						MDDVTLLTRDGDTMQS			
						VLNRLDKLITWARMKF			
						KAKKSRSLTINKGKQKQ			
						LKFTIAGEQMPTVKEQP			
						VKSLGRWYSGTLSDRSQ			
						GVEIMKQAENGLKAIDQ			
						TKLPGKHKIWCLQFALY			
						PRLAWPLTMYEVALSRV			
						EMIERKCNSYIRKWLGL			
						PRTINTSSLYRHRGALQL			
						PLTSIVEIYKAGKIRTVL			
						MLRESRDKEISGKPPDV			

						KTARKWKAEVAANEIIS ALEHGDIVGPAQPDRLG LGNGDFRPFRKMSQRDR RKAATRHMQKMEAERR EIHLIQCSQQGQMTRWE EHVVERKVGWSDIWKW NTSRLSFLLRSTYDVLPS PVNLVRWKVREDNKCR CGKLGTMKHILSNCNLA LNRYTWRHNEVLKVLT KMATQQVEEGKYAPKI QKHGPTKIAFVPQGSKV AKRKKANEAMKDQSGG TWEVAADLKGFERFFPIP TTKKPDIVLWEEKDKEV HLVELTVPHEDNIYAAH ERKENRYEALVEQCEEA GWKATHFPVEVGCRGF VATSARKWMRVAGLTA KKSNELTNSLQETAEKA SHWIWLKREDSSWNES			

Ingi	Ingi-6_AC	.	Aplysia californica	CCCCCCCCAAA TAGACCCACCT CTTCGTGGTGG GGGGGGGGATT AGTACTCTTCGA ATAGGAAAAGA AAAAGAGGAAG CTAACGGGTCA AGGGGTGACCC CGAAAAGAGAC ATATTACTTTAC TATGTTTCTTAC TCTTTCTATTTA TTGTTTTTATTG TTTTACTTTATT ACATTTATTAAC TTTTCTTTTCTC CTCTCTCCATCG ACAACCTTTCTT TGACCTGGAGT TGGGACGGGTA TTGGTGGGGTG GAAAGCAGAGG	TAGAGACACTTGAG AGTGGGACTGATGT GAACAAAGATATAT TGCACTTGATTTGG AGTCTACATAATTT GATAAGTAGGCACC ACGTCAGAGTAGTC CTACAATGGATTCC TGCCCACACTGGGA TACCGGGAAATGA AAGAGCTGATGCTC TGGCAAAGAAAGG AGCTAGCCTCCCTC AGCCTGAAATTCCT ACGAATTACACCAC CTGCTGCCAAATGA TCAAGTCAAACCTC AAGGAAGAGTGGA TGAATGATTGGGCC ACAGGTACGACAG GAAGGAGGATGTA TGGCCACATGAATA AACCACTGGTCAAA	MLFTNDRIPYELRRRLG NQGLPQVLVGNQAGLV DGYWKKKKIRKQRKWK NKKKTPPTTGNGKEKLI DFNILQANVCGLDKKKT QLGKILYERKVHIALLQE TLHSSCDINITGYTPYPC KCEGCQGIITYIRNDIQG EVELLNWHPTDAHKVT LWYGNSKLILYNIYCPPS STLNFLEQSTIFKKTVIA GDLNGHSPLWGYADQN NTGKRIEELCXSSNLILL QDEDSTKTLLHRRHGTL HRPDLTLVSADLHNYCS EEVLDCISSDHRPVLTKI SILKHRQRKRRTRWNFK KANWDLFRKTAEDTVK LDDLDNKSIDALNSELTS TILDAAKKSIPQGCRTNY KPFWNERLEKATKTKN DARHEYEKDLKNMDKK IRYKKAVAEAKLVTKES	cd01650	550	806

				TTCATTGTCAGC	GATCCAATCAATGA	KKEAWTQKCEGLNLHQ			
				ACTGTACTGGA	GTTGAGGAGGGGA	GGREAWSLLNNLTGEK			
				TGTGACTTTACG	GATCAATCCATAAT	RKENPKPLHTETEVLTSE			
				TAGGGAATACC	CTTCCAACTTAGAA	LKKAEHFNKHFASINKS			
				CAGCTACATAC	CACAACATGTACCC	SKRNDLDRGLLSVLKGK			
				ATGTATTAAGC	CTCAACCAACATCT	ERDLMTPSIFQDFLTKRE			
				ATGCACTTATCG	CAACCGAATTGGAG	LDCAIKKLKKRKSPGPD			
				GAGAGCCCTAT	TAAAGCAGTCTGCT	KVHNEMLQNLGEKAKL			
				GACTGCGGAGA	GCCTGCCCACTATG	LLLKLYNRTWTEGNLPK			
				AGGTGCAGAGA	TGACTACCCCTCAG	AWKLATINPILKKGKKA			
				CCATTTGCGTCT	AGACAGTGGAACA	NDPKNYRPISLTSCIGKL			
				TGCTTGTTGACA	TCATCTGTTCTTCTG	CERILNSRLYWWLESSG			
				ACATTAACCCTC	TAGAAAACTAACTG	LISQCQAGFRSKGRTED			
				TGGGATCTGCC	ATTTAAGGGGATGT	QLFRLTQKIIDGFQRGEQ			
				ATGCCTTCCCCT	TTTCTTCCCAAGCT	TTAVFVDLQQAYDRVW			
				GCCCCCATCTCT	ACCCCACACATCTA	RKGLLLKLQNMGIKGN			
				CGGATGAGAGG	ATTGCCTCTACTCC	MYSWIKSFLTDRLIQTQ			
				ACTTGGTTGCTC	TCAAGACAACAATT	MDETLSSKAVLEEGLPQ			
				TGGGAGGGGGG	AGAACAAACATGC	GSSISCTLFLVFLNDVSD			
				GGGGGGGAATC	AACTATTACCGAAT	VLKTEKALFADDLVLW			
				TAGCTTCCAATT	GGCATCATGCAGAA	HTGNSTTISRRRLQDDL			
				GGTATTATTATT	GAGTTAACGCTCAG	NSLGTYCRLWKMKVNC			
				ACTATTAGTATT	ATGCTACTGGATCG	EKTVYSVFTRSYNLANA			

				ATTATTATTATA TTATAATTTTTG TATTTCAGCAAC ACACAGCAACT ATGCTCTTAATA GCGTCTTTCTCT TCTTTATCTCCT CAAACAACTTC TCAATATAATA CAACAGACACC AACATACCCAT TCATTTTTCTTA CAATCTCTCTGT GATGGAGTTGG GTTTTTGTTTGT GGCATTGCACC TACAGCCACTG CTTTATCTCCTA CCCCGGGATGG AAGGAAAATCA CAGAACCACCA GTTTTAG	CTAACAACAACAAC AA	EIPLKIDGTNLKKENNPV YLGVQLDSKLNLKQHT NNLRKKAMKRLNLIKRL ASTNWGSDKNTLRGLY LGYTRAVFDYNLVLQNI CSKTTKNSMDTIQNQAL RLISGGMRSSPTAACEIH TNIEPLETRRKRAALEVY ERSKRLENNHPNRVLVD KWKPNQRLKNTKSVLD TVLNLQEKHHLPERREP LERVPSTLPPHLTIAKPEI RKTLLDKSNKNTYPVAL KHSALETIDSYPPTWIHS YTDGSAFKATINAGYGS VINLPSGIKKKKSLTPAG PSAPISLPNNRLSPTQQT TSTTCLTLTHLLPQTSLS LQTAFQH			

RTE	RTE-6_AC	.	Aplysia californica	GTAAAGCTGGG GCGACCATCCC CGACCCACCGC GTCCACAGGTG GCGGATAGTGG AACGGCCTACA TTATGTAGGGT AGCTGCGAATA ACCAATAAGCA GTCGCGGACCA ACTGGCGACAG TCTTGGAGAGC GAGGTGGGACA GGGTGGTGGCA ACCCTGTCTCTA TAAAATACCTC ATGTGCTCAGT AAAGAGACTGT CAGGCGAGAGG GAGGCGCCGAA ACCTGCCACCT GCCCCCACTGC CTTGTGGGTTAG	TGATGATGAT	MDHITSINKDXDVKGSS ERTASSGRVSHAQQVTP DRHQDTAHKRKNQTRNI ATWNVRTXLEKGKLDNI KKEMKRMNVNIMGLCE VRWTGAGRILSDSYNIIY SGGDKHERGVAVIMDM PTSKTLKGYLTISDRVM LVRLKGSPFDTTIIQAYA PTSESTEEEIEKFYDEINQ AKTHCKSQDIVIVMGDF NAKVGATRHEDIVGPYG LGXRNERGEKLIEWAIL NDMIIGNTWYEQPPRRL WTWKSPGDSARNQIDYI MIKKRYKNGMLNVKTR PGADCYSDHVPVVGKFR IRLKKTHMSKNKNMKL NMALLQSNKEMQDKFRI TVQNKFGPLEEIQDIEKQ WEQLKTSINEAAKDVIP ERTTKAKKKWMTEEILN LMEERRTNKNNXEKYV	cd01650	549	806

				TCCTTTTGGACA AAGGCTAGCGG AGTAAACCGTA CAAAAATCCAG CAGAAGGCAAG GGGAAACCACT GCTGTATTCTTT CCCTAGAACAT T		NLHKEIRHKCNLAKETW LKEECEEMENHSNANSK DLHRKIKEVCHTRTRGS TGCLKSKDGNIIMDKDK ILERWTEYIGDLYNDKR DEQRRVENNNEGPPIMK EEVIKALRKMKNGKAA GPDNITTEMIKALEDFGI TTLTKLLNNIYATGHIPS DLSKSIFIALPKKPNANE CETHRTISLMSHITKILLR IIMMRIRNKILPEIAQEQC GFIEGKGTVNAIYVIRTTI ERALEVQKDLHLCFIDY TKAFDKVRHDKIIDMLE VLNIDGKDLRIIKNMYW DQTAAMRIENDISEYIQI KRGVRQGCVLSPDLFSL YSENIMRHIEEMPGISIG GHNINNLRYADDTVLIA EDEEHLQNLLNVIVQES DNMGLSLNIKKTETMVI SRXTPIPTCNIKINDTTLK			

						QVQHFKYLGTIISSDGRC HTEIKSRICQAKATFHK MKQILTNINLSLETRKRV LKSYIEPILLYGCEAWTI NKQLERTIVATELWFIRR MLRIPWTARKTNEEVLV EANYRKQLITNIRKRQA NFFGHTMRKEALENIVT TGKFDGKRGRGRQREKI LDKMTTWLGAARTTDTI AAXKDRXKWRSMIAHA SRQGTG			
Ingi	Ingi-7_AC	.	Aplysia californica	GGGCTGTAAGG GTTGAGGGTGA AACCCCAGACG CTCTCTGGCAA GCCAGCCTCAA CCAAGTCAACT CCCCGTCCTTGT CGAGCAGCCTG GGGACCCCTGC TCTGTGCATAG GAGCAGGGGTT	TAAAGGGATTTGAA AACAACAACAA	MHTSLHNPXTNKMPLR QMKLYELGCRLGYQGP RSLPSPGGGSDKSATGT HTAGIGATSNTQTLNILQ LNISSIQNKKMELAKTLS DLRIHVALLQETQLPKRI ELHLTGYTMYHCKCTK CQGIATLIRNDITAEVSH PETDDETDMQEVXVQQ SDGTTYTIYNVYCPPRX TLQLNFPNTDHKHTILA	cd01650	545	802

				GCAGACGGAGC		GDFNGHSPLWGYSDTD			
				TCTGGCCAGGA		XTGTVLEELLRSSNLISL			
				GGTACCATAGC		QNDKTPPTLLHRASGTL			
				CCTGTGAGGTG		SRPDLTLVSADLHPHCS			
				CGANGTGAGCA		MTVLDDIGSDHRPILLQI			
				ACGACCCTTCG		SSNRKREPTNPIRRSWN			
				GCTCGTAGAAG		YKKANWELYQQKSESR			
				GGTCCGTACAC		LLDVCTTNPIDECYKDIT			
				ATCCACCATTAT		TALKTSAKEAIPRGNPK			
				GGTAGGTCTGG		KFRPYWDKELEQLTKAR			
				CGTCACAGCAG		KKARRRAEKNPTAENR			
				GCCGACCTGTTC		KHYNYLTAKVKLKTKS			
				ATTCAGGAACT		AKRQTWREHCGELDLR			
				CACTGTCGACA		KDGRKAWNLLTNLSGT			
				TCTTCTTCCCTC		KRRKNPAPMKPTEKGET			
				CCTCTCGTGGG		CRAEVMNKYFASINKKG			
				AGCTCTACAGA		NQNKLSHAIKQLTKRAE			
				CGGGGCATGGT		RTPKANLEVFEKEFTLTE			
				ACCCACAGACT		LKAALKKCKKRKAPGP			
				GCCAGATGACT		DHVTNEMILNLTDSSLSI			
				CATCGTGAGAT		LLGFFNRTWNEGQLPTE			
				GTCTCCGCGGA		WKKANITPILKKDKPAS			
				CCCAGAAGGCC		QPSSYRPISLTSCLGKVV			

				TTGTAGCTGTAA		ERMVNDRLYWWMEST			
				GGAATAGGCTG		GTISTSQAGFRKGKRTM			
				CCTGGAAACAG		DHLIRFAQKTSDAFQDK			
				ATGACGTCCCT		DNVVAVFIDLRQAYDRV			
				ACAGATATCCT		WRQGLFYKMQKLGIAG			
				GGCAAATGCTC		RMYSWIKNFLSCRTIQTS			
				GCTGACCCATG		CDSKISSQRTLEEGLPQG			
				ACGTGACGATC		SALSCTLFLLFMNDLPD			
				CTCTGTCTGAGT		VISCQKALYADDLLXWS			
				AGATTTGGCCA		KGSDMRALNQAINADLT			
				CCCCTCGGATTA		TISSYCQLWKMEINTDK			
				TTCTGCCCCATG		TVFGLFSLKNKVTASQV			
				AACGGAAATGT		NLKINGTTIRHEDYPKYL			
				TAATAGCAGAT		GVQLDRKLSLKEHLGEL			
				AGCCAGGAGAC		TKKATSRLNLLKHLSSL			
				AGCAGGTGGCG		NWGADKATLRQLYIGY			
				GGGATTGGGCA		VRSVIDYSLPLQSIASQSS			
				AACGCAAGCTA		LAQLDRVQNQALRFVSG			
				CGCTTGGTGAT		AMKTTPTSACEIHTNIEP			
				GTAGACGACTC		LDLRRKSATLETYERFV			
				ACTGTGGAAGG		RLEDTNENRKMADEWE			
				AACGCGTTACC		PRTRIQKTSFMREATKLL			
				CCTTACTCAATG		EEFDPPENREPITNISSEP			

				GGACCCCTTACT		PYWQTPEITIRTRLLDPS			
				GGACTGCTCAG		CDKTTPEEILRTSALETV			
				ACGATCACTTGT		DSYAETPIHAYTDGSAT			
				CATTGGTTGCAC		NATRNAGYGSTIKYRNN			
				TTTGTGGCTCCC		LGEKKQTQLFGPCGTYS			
				ATCTCTCAGAA		SNYDAERIAIVKTLEQIA			
				CTAGAGTGTGC		QDFFEGAVTPGDVVVFS			
				CTACACCCCCTA		DSMSVLQAVEGGGDYR			
				TGGTTGTCTACT		GDGNILSLVKQMGTSYG			
				GTCTAAGAAAT		VRVTLQWVPGHSGVPG			
				CCCCGTGAAGG		NETADHLSKTGASLPQT			
				TTGATGGAAAA		DTTASLQTCRGIIQAKSR			
				CAGGACAGCGA		DEWLNRWANAETGRSL			
				ACATTGCACCTT		FGYMDKPNPKDPSNLLC			
				TGTTCTCTTGTG		RADQTLIFRARTGHLHT			
				GCAAGATGTTT		NKHINRINPMWEPHCRH			
				GGGCACAAGCC		CDYHEETVEHLLLHCPN			
				AGTGTCAGGAC		LAVLRRRLLHTPTDLQS			
				AAGACTACCAC		LLYSGVDQMTRTCGYL			
				CTACCTGAGAT		REALRCQERNARS			
				CCTACAGGG					

RTE	RTE-4_AC	.	Aplysia californica	CTCCCCTGCTGGTGGGGGGCCCCGTAGTTTTACTATGGGGGAGCAAAGAATACACCTGCCGTGATACTGCGTGTCGTAAGAGGCGACTAAAAGTGCATAGGTCTGGTCGCCCGAAAAGTGCCAGTTTGCGGAGGGCCAACTACCCAACCCAGTAAAAAACAGTAAGTTACAGAAACACGGACAAACACAGCTTTTTATTTTTCGATCTTGGGAGACGGAGGAGCTTCTGCATCAGCAGGAACA	TGAGAAAGGCAATGATGAT	MKDSGQSRKDSSDPKPSLSNKSLRIASWNVRTLYEAGKSSQVVQEMKRYRLHILGVSETHWIQSGQKQLSSGEQVLYSGREDKYHAEGVGIILSKYAQKSLRGWEPHGERIIMASFTTRSKSINMNIVQIYAPTNDASDEDKDAFYSQLQGVIDKLPRKDVNIIMGDANAKIGQDNVGYERIMGQHGLGQMNENGERFANICAFNSFVIGGSVFPHKRIHKATWVSPDGVTENQIDHFCISRKFRRSLEDVKVQRGADVGSDHHLLLAKLKLRLKKHGTSTGIKRTKFQVSLLDGHKKEEFQLQLRNRFQPLADLGGEDTEVDTHWTKVKEVFTSTCQEVLGGKKWEHKEWISQHSLDLIRIRRQLKAEANASRTRGSRADALEQYRTTAKDV	cd01650	551	808

						KKSIRKDKEEFMNRLAE KAEEAAVGGHMRILYQ TTRTLAGKYGKPEVPVK DQEGKSIFGKEAQSKRW VEHFNSLLNRPPPSNPPN ILPARNELPICCDAPSKQ EILDAIKKLNKGKAAGP DQIPPEALKADANVTAD ILLPLFTRIWNESKFPKD WKEGHLVKLPKKGDLS NCSNYRGITLLSIPGKVF NRIVLERIKTGTDPRLRD EQAGFRKNRSTTDQIAT LRIIVEQSLEWNSQLIVN FLDYEKAFDSIDRATLW KIMRHHGIPQKLVNLIKS MYEGTSCRVVHDGQLT DSFDIKTGVRQGCLLSPF LFILAIDWLMKETTRGH RNGIQWTPWTQLDDLDF ADDLALLSHTQEQMQA KTSVLNSLSESVGLRIHP GKSKVLRTDAQRKEQV			

						TIEGKALENVESFCYLGS IIDREGGTGAEIKSRIGK AQAAFISLDRIWKTREIS LRTKLRIFNSNVKSVLLY GCEAWNASQTCIKRIQV FLNRCLRKLLRIRWTDR VTNVEVLRKTRQIPVRD EVGRRRWRWIGHTLRK SGSCITKKVLDWNPQGK RSRGRPRGTWRRITEIDV QKSGKTWNNIKRLAQD REKWGVFVRGLYPGAG			
RTE	RTE-3_AC	.	Aplysia californica	GTCAGGTCAGC CTTAGTGCTGCT ACCATAATGAC AGTAGATGGTA CTACCTGCAGCT GTTCCTACTGCC TTGTAGGTGCAT ACGGGAGTATG AAAGGTTAAGG GAGTAAACCCT AAACAAATCCG	TAAGCTAAATTGAC CGTCTGCATAGACG AAAAGAGCCATAT ATATA	MSPRVTFDDGSDYSVAL ISYRPSQTGQPPVSKVLI RHVCLLLWVYGARTKT ERQTLNPAPGKKRVKLT PWKLGTWNVRTMTTGL DTTNLEKTEDIRKTAIID RELTRLNVDITALQETRL ANSGTIRERNYTFFWKG LGEDERRIHGVGFAVHN RLLNRISTPQGISERITVM EMETKTGKAHFISAYAP	cd01650	605	868

				GAGGGAGGGCC		TLTTENEKKDQFYEELQ			
				CCTAAGGCAGT		DAINNVPKADQLFLMGD			
				CGACAGGGAGG		FNARVGTENKVWPTCL			
				AACGATCCGAA		GPFGLGKINENGQRLLEF			
				CTATCTTCTGGC		CTNNNLCITNTFYSGKD			
				AGCCCCTGCAA		RHKVSWCHPRSGHWHQ			
				CTTTACTGGTTC		IDFIIVRKKDANMVKQT			
				CAAATGTATTG		RSFYSADCNTDHLLVIA			
				ACCTCGTCTCTC		KAKIVTKKTPRAEKQKK			
				CTTTGGACTAA		PKINIRNTKDPALKAKFQ			
				ATCAGCTAAGT		EHVKEELQHLDSSSPDN			
				CGAGAGGGGGT		MWNSMRSTIYNSASITF			
				TCGTGCAAATT		GIERPCKEDWVEASAST			
				GGGCAAGGCAC		LAPLLEKKRKILIEHNTN			
				GTTCTCCATATA		PTESTKQQLRAAKSNLQ			
				GTCAGGCCCAG		RETRKCANQYWDKLCS			
				GCATTTAGCTTA		EIQQASDTGNIRKMYENI			
				GTCAAGATTTC		RVAIGPTANKVAPLKSA			
				AAGATTTCTGG		TGEQLTDKKKQMERWV			
				GAGGACGCTGC		EHYSDLYSEERHVDRQL			
				AAAAGAGCAAA		HEAIPQLSEMSELDTTPT			
				TACAGCGGATC		EEELSSAIDELACRKAAG			
				AGACCCACAGC		NDNIPAEVLKENKETLLP			

				TCAACCACAGC TACAACAGCAT TTATATGATTAT GGAGAGGACAC TTCATCCTTGTT GTGACGAGCAA GGGAAAAAACA CGGGAGGTGGC AGTTTACGGGTT ATAAGACATCC TTATTTGGCGAA ACAGG		YLYKLLTTCWLKAEIPH DMRDAKITTLYKNKGD KGDCNNYRGISLLSITGK AFARIILKRLQKLAERVL PESQCGFRSNRSTMDMI FSIRQLQEKCREQRLPLH TAFVDLTKAFDTVSRSG LYMVLGKIGCPPILLQLII SFHEDMKATIQFDEGCV LAPTLFSTYFAVLLQHAF KESPGDVFLHWRTDGSL FNLARLKAKTKINKTTV RDLLFADDAALVAHSES TLQTMMDRLSTACKVFS LVISVKKTVILTQEGAEQ SHIWLNDKELETVDKFC YLGSTISSTTSLDYEINIRI GKAATTFGRLTKRAWE NKRLTIKTKIKIYEACVL STLLYGSETWTLYSRQE KRLDIFHLRCCRKILGIS WRDKVTNVEVLARANL QSMTTILCLKRLRWLGH			

						VKRMDDNRIPKQLLYGE LAQGKRQRGRPKLRYK DTCKNSLSKLGVDVSIW EEIAEDXKTWRKXIKER TKMLENSLRNNQAEKR QRRKENSGNTELRSTTL TCKHCDRICASNIGRISH ERSCMKTTLRPS			
RTE	RTE-7_AC	.	Aplysia californica	GGTGTATGCAC CGTGCAACGAC TCCTACTCAGAC GAAGAAAAGGA CTCTTTCTTCGA ACAACTCTCGG ACTTAATTAGC AACATCCCAGA CGAGGAAGACC TTTATGTA	TAAAGTAA	MGDFNGRVGERRTPWT KHLGPHSDHRTPCNYNG NHVLELCAEHDLFITNTF FQHRASQIYTWYRWNNI MVSSQIDFILTRTRMRITI TDSRAIPNAGLDTDHRPI ITTLTTKKERKPKKHKR QQERLNMHKLGDDQVQ AQIRSTLNEKLGSINSTIL TVEEAWDTFKTTLLDTM KEVCGTKKTRGKHRKA TAWWNEEVKDAIKEKK RLYNVWVKSKNEEDYI KYRLARRHSKKVVITSK EKSWTQYGEKLSETCKT	cd01650	392	640

						SPREFYKKVKTMRTRDE PYVPTTVINDTNGEALH EEKEIIKRWEEYFQELLN PSGVQAQNTQCPFTANY PEHSEPTILESEVREALK TSPKNKAAGDDGITTEAI LACGETGIHWLTIIFQKA WKERKVPADWQNAIVV PIWKKKGSKKDCNTYR GISLLSHAGKMYAKILE RRTRAKTEHLLSDAQFG FRKGRGCTDAIFALRQL CERALEYDKDLHLVFVD QEKAFDRVNRNKLWKIL EQYDIKGQLLDNIRAIYA NSRSAVRTTSGTSDWFP VTSGVRQGCNLSPLLFVI YMDQITKEANPDPEALN ELMFADDLAMINNKTQL QEHINQLNASCEKYDMK ISISKTEVMTISRRPGKV DININGSQLKQSREFKYL GSIFTENGKLDREIETRC			

						QRANAVSYQLGPLLKHP NIPMSTKAKLINAIFLPTL TYQCQTWTLTKALERKL VTCEMKCLRRAVNKTR RDKIKNEVIRDTVGATP VLPHIEKQKVKWFGHLT RMPSNHPALRAYNIKYS GWRARGRPRRRWSESV ADTLKDHGMSLLQATR LAADRHLCLPATPQGTS GRKK			
Tad1	Tad1-23_BG	.	Blumeria graminis	ACATCCACATTC CGAAAAGTCGA CCCCATAGATTT TTATCCATAAA GTTGACAGCAA ATCGTAATCGT GATGACGAGTA TTCCTACGAGA GGAATTAATGG TGGAACGCCTT GCGCGGGCTCT AGAGGTGGTAG	TGAAGGTTGACAGA CTTCTCCCAGTAGC AAGGAAGTGGACG GTATCCGCACTGGA CATGACAGTGATAG TGCACTTCTTGAAG GTAACACAACATCA AA	MLGAHDLNLILPKGTIT WKSRGSQSTLDLAFVSN ALESTVMACKPAFDLEA SSDHIPICTQLHIEPPNRE EQQPRPQWKKADWEEV NRRLALKLSRLSTDHLE LNSPESIDQRVSAITQAIK ETIPTARPSTFAKPYWTK ECSETVKEARKARRLWT KLGSEESWVEYQKATN RKKSQIKKAKMIGWRAI VSEATNDPAKLWKLAK	cd01650	332	595

				CGTCGCGTCAA		WAKKSPEERSRLPQLPDI			
				AGGCGCCACAA		KDKDGNMFTEDTEKADI			
				GACACAGACGC		LAQHFFPQPVQANLEDI			
				GTCCCTGAAAA		DGTRYPEELTTVSSIITAS			
				GATTCAAGGAC		EIQDTMANLASEKAAGP			
				GCAACTAAAGG		DGISNWMLKNCRITLGK			
				CGAAGAGCTCT		VLEKLFNACISIGYHPKG			
				CCAGTAGAGCG		FKESVTIVLRKPQKESYD			
				TACAGGAGAAT		NPKAYRPIALLNTLGKV			
				GATTGAAGAGG		LEKLVANRLSKAAEAHH			
				CCATTCAGGTTA		LLPDEQMGARQLRSTIS			
				TCGAAGGCCTT		AVELITEQVHSIWGKNK			
				CAGAAAGCGAA		RRVASLLSLDISGAFDNV			
				AGTCACCCCCC		SHKRLIHNMRAKGIPRW			
				CCATCCCCGCCC		TTQYIESFLEHRTTSMNL			
				AGACAGGCCAT		GSFKGEQFPTSTGIPQGS			
				AGTGATGATAA		SLSPILFLFFVSTLLPELQ			
				TTCTGCAATGGC		SPSSTAVGFVDDTNILT			
				TAAGATCCTAC		WSNSTEDNCRNLERLHG			
				AGGAGATAAAG		ICEKWADRHGVKFAPEK			
				GACATCAAAGC		YQLMHFSRARKRHNMQ			
				CTCCATAGCCC		ATVQIQGHLAKPQPSLR			
				AAAATAGCCCA		VLGIHLDPKLNWGAHV			

				CCAATGCAAGC		KSIQLKVEPQIQALGRLA			
				CCCAGGCCGCT		QSTWGATFHKAKLVYN			
				CCTGGGCAAAA		AVVRPAVTYGSPIWAEP			
				GTCGCAAGCAA		GMAGKIPERVIKPLRSIQ			
				GGGAGAGTCGT		RRCLKIVTGAYNSTSSK			
				ACGGAACTACT		VLEHETSTLPMEIYLKQR			
				ATCCGAATACA		RVQHAGLSKSLPVQDTI			
				GGATGAACAGG		SSACRKIKQPRAGRQDA			
				AAAAGAGAGAA		RAVTRIMDIADWGKIGA			
				TAGCCAAGCTC		NIXTKERQKAAGKIAAF			
				TCGAGCGAAGA		KEWERSWTRLAARQPT			
				GTTAGTCAAAA		NHHATADPKIWNAANIF			
				AAATCGGAGTA		TDDKTGRTRMDQRGTPS			
				CAGCAAATAGT		SLHKHLSRAQSSIAMQIR			
				AGGTGCTAGAC		SEHIGFNSYLHRRKVPG			
				AAATGACAAAT		VDSPRCQCGYPSQNVKH			
				GGTCAAGTGAA		MVMACTSWAKGRGEIL			
				GGTCTACTTCGC		RKAENRSFEAMMNNPK			
				CGGTGAAGGAA		DLARITRWIQQEGRLEQ			
				CGAAAGAATTA		FRLTLEVEATMRAREEG			
				TGGAGCGTCAC					
				AAGGACTGGAC					
				AACGAAGCTGG					

				CTGCTTCTGCGA AAGTTGCCGGC CCTGTTTACCAA GTTCTAGTGCAT GACATGCCGCT CTCCTTTGAGCC ACAAAATGCAG AACATCTGAAG CTGCTACAAAA AGCCAACCATG TTATATACAAG GCATGGTAATM CAAAGAGCCGC ATGGCTTAAGA AATCTCAGCAG CCAGGGAAAAA GTCTGGATCTTA ATCCTGTGGATT GATCAGGCCGT ACACGCAGATA TGGCGATAAAA AAGGGCTTGAT GTGGCAATATG					

				AGCTCAAGTCA					
				ACAGAAATATT					
				CAGGAGTGGCT					
				TCCGCATGATG					
				CAATGCTTCAAT					
				TGCCAAAAGTA					
				CGGACATATTG					
				CGAAAATCTGC					
				TCCGCAGACTC					
				GAAGTGTGGCC					
				ACTGTGCAGGA					
				GAGCACAACAC					
				TCGAAACTGTA					
				CAGGAAAACAA					
				GACGTCAGATG					
				CAGCAACTGTG					
				GAAAAGGCCAC					
				AAAACATGGGA					
				CCAGAGCTGTC					
				CAGTTAGGGCA					
				GCTGCAAAAAA					
				TAGGGCTGTCC					
				AAAACAGGACC					

				CAAGACCCAGG					
				AATGTTCTTGAC					
				CGACGAAAAAT					
				CGACAAGAAAT					
				CAAGATGGCGA					
				ATGGCAGATTG					
				CAGGAAGCAGG					
				AAAAGAAGAGC					
				AGTAATGACTG					
				CACCACTTGTTG					
				GGGCCGATGGT					
				AGTATAGATGG					
				TCGTAGGCGTC					
				CAGGCAGACCA					
				CCAAAGATCCC					
				ATCAGCTGTCCC					
				CACAGTCGCAG					
				GGCTCTTCGCA					
				AATCTCAGCGA					
				GGTGCCATCAA					
				AAATGACACCC					
				CGGCCTGTGAC					
				ACGGGGAAATC					

				AAGCAGATGAG					
				ACTGAGAGCGA					
				AATGGATTCAT					
				GAAGGACAAAT					
				CAATCCTTTCAA					
				TCATACAGTAC					
				AACGTCAACAG					
				GAGTAAGAATA					
				AAGTGCAGCAC					
				GATTTCCTCCAA					
				GCTTTGAATCCA					
				CTGAAGCACCA					
				CATAGTGGCCC					
				TGCAAGAACCA					
				TGGCGTCACCC					
				CACMGAACAAC					
				ACGACCGTCAC					
				CCACCCTGCCTA					
				CCACCTGGTCTT					
				CCCAAGTGGCC					
				ACAAAGGCAGG					
				ACATGCATCTAT					
				ATCAGCAAAAG					

				TCTGGCTATTAG CAAATGGAGAA GAGAAAGAGTG CCAGAGGAATC GCAAGGAGACA TCACATCKGTC AGCCTCGATAC CGACCTTGGTA AAATTTGGATA CATAGCGTGTA CAATCCACCCC CATTACACCAC AGCAGCCTAGA AGAAGGAACAC TCAAGTGGCTC CCGCAAATCCT ATCTCAAGAGG GGGAACACATT CTAGTCGGAGA CTTTAATGTGCA CCATCCGCATTG GGGGGGACAGT CAGTACTTAGG					

				CACCATAAGAT CGCAGACAGCT TCATCGAA					
CR1	CR1-7_CBre	.	Caenorhabditis brenneri		TGATTTCTCTTGTCT TTCCCTTTCTTCACT TGTCGCTTTTGTTTG ACCTCCCATTCGCT CCATTTTAGGCATA TCCCCCGTTTCTCTT TCTCCCCCGGTGAG GCCTCGTGAGGGGC ACTTGCCAGGTTCT TCTGATGTATCATA TTGATTTTCCCTCT TTTGGAAATAAATA AATAAATAAA	LLQKLINFGIPSSFVSWF AVFLADRTFSVKINDSIG DTSPTIPSGVPQGSVSGP LLFLIFINDLLLSLATIPLL HIAAFADDIKJFSNDPFV VQTGIDTVVGWAKSNSL PLAHSKTALLHLGSLNPF FPYTIAGNPIIPTSSVRDL GLLTEPNLKFSAHINRTV ALALLRSKQLLKSFKSSS PQFYVVLFKCYVLPLIEY CSVVYCPPSTSKLSLKLE YPLRFFTKKVFHRCNISY TSYPDRLAHLDLFSLRH RRLKAQLQLLYNFLSGS SYFPNLDSHVRFSSSPRR PMNLVNVNPRVKDFFFI TVPIWNAITCHCPTFLRP SEFNSLLQCSIQRF	SSF56672	45	118

RTE	RTE-1	AF054983	Caenorhabditis elegans	CTCTTTCTGTACTAAGCCTTTTCGTGGCGAAAGGGATCAACAGATAACCACTTCAACAAACTCGATGATTTCGTCACATCGAGAGTGGCTTGGTTTGGAACCAAGTGCAAATTATCTGGCCAAGAAGTACACCGTCTCACGCAGGCCAGAGGTAGTGTCGAATACTCTCTGTCCGGGCGGCTCTAG	TGAAGAACGGATCGTCTAAGTATCTAAGTATCTAAG	IGLAGHYRVSRDTTFALNSNPSCLSPPHPASAKGKTVPGARRKRRSNLGVQTICQSDLEIKKPRDKTRRTDYRVCTFNCRSLSSDDRLAELLEETRRIQFDVIGLCETKRAAEAHLIHHDGTGVFLGKRNESSVSGGVGFIVRSTLLPKIVEVRFLSHRIGLLTFKVSRKFNCTVIQVYAPTADSDLEEICDFYDSVEDVFRECRSKYKLVIGDFNARMGCRQNNERYIGPHAMEPRNDTGELLATFCETNRLWHTNSMFKKPMHKRWTFVSPDGNHRHEIDHILANGKFVTDTTVLPSFTNGSDHRLLRCNLHFNNCLAKLEQVRRRKPPKRVLDPAIAYAISATVTVQSDPDLDTDYDNLIQSLKELQDQAIVRPANHSSNRLSEETRKLLNKRRF	cd01650	562	831

						MDRNDPQFKSISDKCRE AVQKDHEAFASTRLLSA ANQKKSLKRVARDINEY KSVIPCLKSTSTGERITSR VKMEQEIEKFYTELFKS AVSNSQTSSIPATATPPPF LPEEIRHVLRSFPNGKAA GQDKISADFLKSCHDNV IDLITDRFNRYLHSRNVP KPWKTSKTTLIFKKGDR ENLENYRPICLLPVLYKV FTKCLLNRMRRSLDEAQ PVEQAGFRRSFSTIDHIH SLQRLLEVGREYQIPLTL VFIDFKKAFDSVEHQAI WKSLDEQGADGAYIDLL KECYKNCTTNFTPFHRP VAVPVTKGVRQGDPISP NLFSACLEHVFRKLSWIE LKGEAEDYDTIPGMRVN GRNLTNLRFADDIVLIAN HPNIASKMLQELVQKCS EVGLEINTGKTKVLRNR			

						FADPSKVYFGSPSPTTQL DDVDEYIYLGRQINAQN NLMPEIHRRRRAAWAAF NGIKNTTDSITDKKIRAN LFDSIVLPALTYGSEAWT FTKALSERVRITHASLER RLVGITLTQQRERDLHR EDIRTMSLVRDPLNFVK KRKLGWAGHVARRKDG RWTTLMTEWRPYGWKR PVGRPPMRWTDSLRKEI TTRDADGEVITPWSTIAK DRKQWLAVIRRNTTNS			
Vingi	Vingi-2_CE	.	Caenorhabditis elegans	C	TGAAAACGGCGCTGCGCCGCCTGTTGCACAACAACAACAA	NMKQPAKRTVMTCLDQ RAAFDKVPIKKLCVDLH ESLNEAKISKWLTNYLS KRIIKANFGNHSSKWCT LLGGVPQGSVLSPNLFSF YMRNMPIHPKTEITVYA DDTTILAQDTNPFKAGE LIQEHLDLLYDFFNERK MSISAEKSTTTLFTCDTK EFKTKIPVNWNGAEIPTT	SSF56672	8	150

						NQVRLLGVHLNTMCGS KDHVAETRKKMMVSTN ILKALSGTKWGCSKELLI STAKALVKPIAIYGAPA WSQLLSETNLNKLEAAY RNTLRTCCGLTSDTPIEH IYAEAQMMPLKEEFELA KQQTYIATTKSGTHPNR NLESRRNTERRSNRPREP PLNLQEATKTHLSRLKG TKAEAQKANHTRFVREF LNSAPPNRILGATPPPLN SSEESSLPKATRTELARL RCGHSLLIPRYKARIEGT PIGTCGCGDNEGGIDHY LSCGRSHPISKEKLWTSP MEAARALGLAYKPYDP GGGATTSV			
CR1	LINE2G_CE	.	Caenorhabditis elegans	AGTGACGGGTC TCAGTAATCAA TACTCTGTCTTC CGTACGGATCG TGCTAAAGGTA	TGATCTAATGTTAC CTTATCCTCACTTCT ATCCATCATTTATT CTTGCCTTTTATAC ATCGATATTTGTGT	MPTRKSKILDVALSNSVI SNLNIMAPIGSSDHNTIK FTVQLLQNISVPKKMIPD FSKLDWPCCATYFENIE WGKVFDTHSHVNDIYEC	cd01650	299	552

				AACGAGGTGGA	TTTTTGCTTCTTCTC	FLGHIRTMFSQAVPTMP			
				GGTGTTTGTAAT	CAATGTAGCCATTA	MPETSKPVPQYLRSFEQ			
				GTGGTAAAAGT	TTTTCTTCTCATCTC	MVKNQFSKTIMSKTMN			
				AGAAATCCCAT	TTTTTTATTTTTGTT	DFIEYIYLARKYRKKLR			
				GTGAAACAGTA	CCGTTTTATTTATGT	KYLSNLEVKLFKKKGNS			
				GTTTCAGTTGAA	AATTTATTTATGTA	KFSKHAKTLLKPRNCSV			
				AACGCTTAACG	ACCACTTGTGATGG	PALIDNDENEIVSELDKA			
				CTTATGAAATTC	TTTGTATCTATTTAT	NLLASYFEAQYTNQNNT			
				TCGTAAATGAC	GTGTGTTTTTTTTCT	NFALDDDFVSDQPIPWV			
				CTGAACATTGA	ACCTCACGATATCT	TDAEVLKVIMASKKSCS			
				CGTAGTTCCCTT	TACTATCGTAATAA	VTSDQVPFTFIKILSPLISS			
				TAGAATCATAA	ATTAAATTAAATTA	VLAQIFNLSMMRGQVPH			
				ACATATATCGT	AATTAAAACTT	AWTESIVIPLDKITNANN			
				GCTCCTTCTTGC		PTDFRPISLTSHLCRVYE			
				ACTTATCAAAA		RCILAKIELFLIDNKFWS			
				TACTGTTAGATT		QNQHGFRPKKSTVTNM			
				AATTAGTAAAA		LESLNDWSKSIDLGEQID			
				TTGAAGACCTT		IIYLDFSKAFDRLPHDLL			
				GTGGAACATCA		LDKLVSLRMNKNLLIWL			
				AGGCAAAGTTG		NEFLSNRSFRVRIGQTLS			
				TAGTTGTGGGA		SRKLATCGVPQGAVLSP			
				GATCTTAATATT		LLFGIYVNDISSILPEKVA			
				GGAGGTATCGA		CKQFADDTKLYVSTPKT			

				CTGGGGCCTCA ACGACTCTAAA ACTAAATTAGG TTCTTTATTAGT AGATCTGTGCA AAAACTATAAT CTGGAACAATA TGTCAAA		ESENNLQSALDAVVDW TRGSKLTLNQSETVHVT VGKRRTDFKYHLDGYPI ERKAITRDLGFLISEKLD FSEHWRKCINLAKFQLA NMFNQYSTSNKKLMILL YKTFIRPRLEYGTVVSSP TKKSDEKAIESVQNAFT RRLYSRIKKRYINPTDKD YKTAMERNELFDLQTLS TRRKIIDRIQVIRMNNGK VDLKTSEFFEQQETHTRS KKKYIWSTGKSKLRRHF FVNRTLATMKQH			
CR1	CR1-15_CJap	.	Caenorhabditis japonica		TAACTTAATGTAAAAATTGACTTGTATTAGTGGCTTTGCCCTTCAATCAATAAATAAAA	STKLDLXXXEDJXFNEX HLNKSSPKAFFNFINSRX KXXQEIXXXRDXXXXQ VTSDYMKAELLSNCFSN AFTEDKNTNTDFPYRTN HFINSLCFEPFIVESILSKL TPKCNTTPDGIPAIFLKR TCTSIALPLSIIFRNSYNQ XVLPSFWKTALVKPLHK	cd01650	151	401

						KGSRSDPNNYRPISLTSA VGKVMEKMVRKSIILFL DDFKLLSNSQYGFRSKM GTESQLLSYQNKLVTNY LSKQTTFSVYIDFRKAFD TVGTKKLLIKLERYGIRN HLLRWLENFLSNRTQRV IINGVMSEEKSVLSGVPQ GSVLGPLLFLLFVNDIGD NFNSDYLLYADDLKIFST DKSSIQRDLDSLSTWCE NWQMSVAPNKCETIAFY HSKRGMDKDTKFTIGGA TVPITSRIRDLGMFFTSD LSFSNHIDTVLRKAHQR VNIFFNVLRHAGFDTFIK CFIIYVRPLIEYGSIIFSPT LKDLVKSVESVQKSFLY RCFKKFRIEYISYFDSLS YCKLESLEHRRLIIDLTF LYKLIVSKEIFIHNFTCIT RSPLPSLRRHPYIVRSTL RNSSKLSHQFFSNRIITC			

						WNSLPSYLFPKLICSETF KSALKCLDLSKFLTLNM HNY			
RTE	RTE-1_CJap	.	Caenorhabditis japonica	TGCTTGCTGGAT ACCTCCTCGCG GTGCAAACTGG GCAATGCTGTG CTCCTACGACA GGGGAAATTGG ACACCACTGTT AGGCACAGTAG CCAAAATCCAC TCTTTGCCACGG TAGTGTCCGAT ACTCGTGAAGA AGGGGCTTCAC GCCGGCCTAGT AAGCCTGCTGA AGTACAAACCT AAACTTACTAA AAAGCTGAATC AAAAGAATGTA AAGGACCCGTC	TGAGGTCGAACCTC ACAAGTATCACAAG TAAGTAAGTAA	MFGRSTTDTEITEASYGP KNRRKGNDFHIGSLNVR SISDQSKAEAFDLLAVNS GCKVIALQETKRKEGEW VLSSGAELICTRREAREG GLAFWIHKDWLDSLVET KRESDRISTASFDFHGSH LHVINVYAPTSAANETQ SDKFYKDLKKVYTRLSK WDSSIIICGDFNAVVGCS SDDLPFVGGHGTGQRSD NGDRLVNFAFSSRVSILN TIYPKRPNRKWTWISPN GKTKNEIDFILTNRREIA VDCEVYCDYLNCSDHR LLRARFRSTAKKEKSRK FAKTKRAGGERSWDPV RFKNTITEMLTENPPTCY EELTESIRSAASTATHAT HQLSRISEATKKMMLSR	cd01650	531	773

				CAGGAATTCAC CAGCTCGGCCA CGGTGCTCGCC GAGTTGTCTTGG TCGGAAGAAA		FQMIHAGQTKNADLAKI NNEVRRSLENDLDKWR EVKSLEAAVKGRSLKKT KRALTLKSSPFTNLERPD GSTTYNRPEIRKLVLTHF SNLYAATTREPRSRYNH PDEPEILQHEVEHAIVTS KTNISPGPDQITMLQLKL GIESIAPHLTAALNQVLT NGRTPEDWKTVKISLLP KTTKPKKVKDYRPVALS SIVSKLFTKILTRRITAKS EDYLEESQAGFRRGRGC ADNIQVVAQMWEKCTE FKIPLVAVFLDFTCAFDN VNWTKISEVLNNLQIGR NVIRALNNSNSSAIGELN VLNKKMRFKIKRGVRQ GDSSSPLLFALALQAILD ELDPAPCEDDKTGIDING ESIHRLEFADDVVLFANS VKEAEDRANRIAMGCPK YGLNINPSKTQILMNKY			

						VTRSDIDILNTRIEISQKV KYLGRTFADDGSLTEEV ARRIRAGWAAFNAIRRE LLQSPQKIRIRLLTTTVIP AMTYGCETWTSKEADT KRLQRAVSNMFEIAGCD PPDMEKQVLRRKLKWA GHVSRMKYDRWAKIAS EWDPRGSARPRGRPPKR WADDITDSIKIFIHNEAL RGNLGHGRRRIGILEAR RAWSTVGRDRVTWRSL VRSIQDGN			
CR1	CR1-6_CJap	.	Caenorhabditis japonica	GGCGGTCGGCC TATTTTTCTCCG ACCGTCCATGTC GCGTGAAGAAT TGGTGCAATAC CACGACAACCG CCGCAAGAACT	TGACCACTGATGAC CGAACGTTTACGCM AAATCTATCCTCTC CTCTCTCCCAAAAT KTKTTWWTTTCAC GTGTACTCTAACCA CATATCCCTGCAAA TTCCTCTATTCTTGC CACTCCGTTTTCTTT GAWTCCTCTTSTAC	MPTNHSQRTLMPRRCK YERTRQRIFSRIQKTNFA QHRXFNFFXHSPDLILSI LSANXRSLMANKLEFQY IVNTTSPDILCIQETWLSS KFGNASFLDPNSNYHVF RADRCATSRGGGVCTLV KKEFPXTIITAETKNSAY ELITVDIDVXVTPLRIINI YRPPSCSFENTITLLNKV	cd01650	528	785

					ACCTCTTCTGTATTT	SDLSNHVGQJVILGDMNI			
					ATTACTTTTCACGT	NGINWNSAVPTSQTRLG			
					ACCGCAATGGGCGT	SLLCDFCNQNNLTQYIR			
					ATACAAATAAATAA	KATRKENLLDVVFSNSR			
					ATAAATAAATAAA	LSDVRILSPIGSSDHNSIN			
						FKLDILQRPPEKKILFYD			
						YKNIKWPKCLDYLGSID			
						WGRLLHPGNSISTMYQN			
						FIDHMREAFELSVPKIYV			
						EPTIRNTPDYLKSFYKLV			
						QLKLERVKASRGRXDFL			
						DYIKTAKKFRKKLLKFR			
						RNEEVKLLESSGVSRFSK			
						KAKYLLKPKSLRIPTLTN			
						PLGVLSVTNAEKAEILA			
						DSFESQYVNRPLLNFELP			
						SNFKSPQEIPWVTDAEM			
						LSMMLKCKNSCSQTPDN			
						VPYKFLKLIAHLISSPLA			
						QLCNLSMMRXECPSLW			
						KESIVIPLNKKEKPMAPT			
						DFRPISLTSQLCRIYERCL			
						LKKILQHLDTIKFWNND			

						QHGFRPKKSTVTCMLEA VNDWTESIDEGAQVDVI YLDYAKAFDRVPHDILL GKLVDANLNPNLIKWIR SYLSNRXFSVKVNNESS TPKHAQCGVPQGAVLSP ILFGIFVNEIPSILPSGIRC KQFADDIKIYTAIPKSDTI EDHETLMQSAIDAVCN WSENSKLALNDSKTLVI SLGKNPRTTKYQINNNPI RRETSIRDLGFIITPMLDF TGHWKKAICKAKFSVFQ IFNHYCSNNPRLMTLLY KIFIRPILEYGTEISNPTR KLDCKAVESVQNAFTRR LYSRIKGRYLTPNDPDY LTAPQRNSLFNLSSLESR RQTSDRKLILKMISGLVD IDTSPFFTYEKNTRTRAK NKFKWNHCKTKIRRNFF VNRTLSSFKL			

CR1	CR1-12_CJap	.	Caenorhabditis japonica	TT	TAAATTAATTCATTTCAATTGTAGGATTGAAACATATTTAATCTTTTTTTTTCTATTTCTATTTGTTTTTCTTCATACTTATTGTACATTTTTATTGAATTTTTCCCCCCTTCATTTACAAAATTGTACTGTTTTATCTCTTTCTCTTTCACTATAGCGCAATTGTAATCTAGCTTAAATTATCTCCAACATGACTTTGTTTGATTATTATCGAAGGATCCTCTTGCGGGGTGATGTCGAGTAATTTTATTTATTCATTGTGCCATTCCCACCTCTTTCTCTGATCTCGAAATGTTTTTTTTCTCAACCCG	LINILNELLTELCPRIPERPXSKVCPLLRKLRRLRRYKSSSLSHADIANRLKRYRALASRYRKWVLKHETKLINESNCTSIRKFVSKRLKSKQSIPALLDNGKAVTSDFVKCDLFSXSFMKNFACPPEEYSTPTPSHYGSDSLSSFSSPSSSRCLFLPYIVEKVLQHLPPKCGFSPHLANYLVLKKCATPLALPLSIIFSQSYCESEVPAAWKLAVVIPVHKKGNPACTDNYRPISLTDPIARVMERVICNYIRSELAYKFSPYQHGFLHFRSCTTALASSTAKYRQILKKSKTLDIIFFDFSKAFDQVSHSILLDKLAKFGVDPLTCSWFENFLCDRSFTLKVNNTFGDKLYPVYSGVPQGSVSGPLLFIIFINDLLFSLPPPIHYSCFADDIKIFSDSPEILQS	cd01650	215	460

					TGAATAAATAAATA AATAAATA	GIDTVVQWSIRNCLPLAP TKTSLLHLGDKNRMLK YFIHGIEIQHSSCVRDLG LITDSSLTFKIHISRIAAL ANLRANQLLKCFSSQSP QFYAHLFNTYVRPLLEY CSIVYSPSPSSSLAGKLE NPLRTYSRRVLQRCNISF SSYENRLEILNLTSTRHN RLKSQLLFLYRIVTNNSH FPNLSDFVSFVKSNRRP MLLVRNDTCTLHFFART VPIWNSLLRNCSTFLSPIE FSNLLNVNISRF			
CR1	CR1-16_CJap	.	Caenorhabditis japonica	TGCCTCTCAGAC TTTAGAACTTTT CAAATCCCTTAC CGCACTTCTAG	TAAATTTTGTGCTT GTGATATCATCCCA CCCGTCTTTCTCTC ATCTATTTGACCGC GGATTTTTCCAATT GGTCATCTTACTTA TTTATTTTCTCTCCC CTTATTTTTATACC GTTGTCCTCTTTTCC	IFXTIILLYWAISIFPQSTG DLLPDSPEFLVDXMTSF HLSQLVTFPTRTSRLGK KNFLDLLFSSNTSIIKNL KAGSPFILSDHCSVLFSL SLATSTIPKPPTEDPQHL HFRKCDFNKLNCHLFSY NWPKQLSYFATADSKLY HFLKIFNELIEQYTPISTK	cd01650	343	590

					ATGTGCCATATACA TATTTTATGCGAG AAATGAATAAACAT ATTATTATTATTA	RTVHSKITGKKLYRKLR KRHPNSLSLSLKKRVKS RLKNIRKSFESQEKKVIA SKNPRSLYSFIKSRIKTKS TLTHFEVDKKVIQKPSD MADLLVNIFSKNFNTPPP PFPALPSPKPLNTSIDTSP FGIYQVIAKIKPKIGYSK DNINNFIIKNCADSISIPLS LIFGEFLLSRTFPDCWKIS TVIPVFKKGNIHDPLNYR PIALTHPLARLFEKVVIG PIRRTFSPKLSKKQFGFL NNRSCTLAHLHASSIFHS VLSYPKSYLDVILIDFRK AFDSVPHELLLCKLSNF GFDTKLCEWFREFLTGR SSIVKIDNNYSSIKFKNSS GVLQGTVTGPFLFLVYI NDLLESFPDDVHVVAFA DDIKIFGNNRSSLQKSIKI IEEWCSIWQLKLAENKT QVLHLGVHNPRVEYHIG			

						GITIDKCTSVRDLGIWVD EKLQFRKHINIIVNSCMY KVRQILNSFRSSNLNFYF KLFNTYIRPVLEYGCEVF NPCSVSMSTSLEKPLRFF SRAVFKRCRVNYLSYAD RLLQANQQTLKHRRVLL ILKTYHNIISGWYHFPNIS QYIKLAHTPRFPYLLQPR GKWTPNNFLHKYLSLW NKIAKKFPKTVSGATFH LHISALPLDSYL			
Vingi	Vingi-1_CJ	.	Caenorhabditis japonica	TA	TGACCGAAGATCGT CAAGGAATGAGTA GAGTTGCACAACAA CAACAACAA	RPAPPPPSQHTTKCRAKP AKTRILSWNVNGWSKR RSELATIIAKEKPDIVCL QETKLDKTNLGYTPTGY KVALASRNKFGGGVCT MVRDDWIFTISKTTSKK TAEILSITIDKLMVTNVY FPPRSSDSPTAEVREDV MEILDMASEKHIITGDW NAHAKTWHSHKTDERG AMIEDELDTRGVHVVLN	cd01650	502	758

						EDSFTRAASGVLSSPDITI STSDLAACSSWRTCGVL NTDHLPVIIDTMAEIPKL VRPXRIMANFKKANWT DFSDSIETKVACYAGPR DAYSLENFLSATITRAAS KHIPHGKFTPTREHTGDT GEIAELTRERERIQKRNP LDSNLKALNDKIEAARR ELRRLXWHQKIEDMED KRKIDCNALWRVIKGLK PKTKSTGSIRLPNGTTTT NPKTIANAMASSLVAVG KGRDCREERIAWRRTKR RLKKLRPQLPASSISHAE VKLAILKGKPSKALGPD GICHLHLKHVPDNCINL MAELFNASIDENKVPDS WKKANVFMLPKPGKDP SEIKSFRPVSLLSPVAKV LERIILERIRREIESPKDQ HAFKKNHSTTTAITEATS CIVGGLNMEKPACRTIM			

						TCLDMKAAFDTISIAKLL CGLEDAIADKKVCGWL GNYLHKRRINVQYDDV KSSWLTLRGGVPQGAVL SPNLFNFFIRDVPAIPDTK IITYADDMTLLVQDPLLS KAAEKSQDALDSLQKW FKDKHLDISAGKSTVTIF TADTKEFKYDPGLKWE GSVIPMQNRTRLLGVEL DTMLRMTGHVEQVSAR LAKANNILRALAGAKW GSSKETMLRTFKAIIKPL ATYAGPAWHQLMSETQ KTKLERQYVGGLKACC GLTKDTPRQLVYHETRM LPLSEELTLGSEQLAIAAI KTPGHPTSDLSRRRPKER TCGNRPRTPPLGCLEDR GDEWRRMRGDTRAVQK RNHTSFVKSYLSNHGIHP ILGRQPPALSEEESTLPR NTRVELARLRAERSLLL			

						EKYKAKVENRPVVCCIK CNDDVGDLKHFLKCYP VKPLPMSKLWKDPVAA ATALGLAVTPFDPGGDA DS			
Vingi	Vingi-2_CJap	.	Caenorhabditis japonica	CACTAAAATAA AGGACCTGAAC GTGTACATTAA AACGAAACAAA CACACACTTGC ACGAAATGCGA CCAAGTCACTA ACCGCCACCGC CACACCWCTAG CAAGACCAAAG ACCAGCAAWGG GACGWGCAGAC TCAGGCCAATC CCACCACGWCG CGCTCAATACA CAATCAAACGT GTGAAGATTCT GACATGGAACG	GAA	MPPGNEDRYRTRGIHSN WLQDRYGPEEQARRRS VXNGQARLGLHDHSIHD EESGLLITNVYFPPRSSD SPXAEVREEVLQILDME DRKHILTGDWNAHSSA WHSHKTDERGSMIEDEL DSRGNHVILNENTFTRA ATGILSSPDITMATADIA SRATWHTCGVVNTDHL PIIIAMDSEIPKREKPKRV IANYRKADWKGFKTDV ETRIRSYVGPRDAYSLEK LISKAILEAAGKHIPSGRF VPTRECQGDTKEMDDLL KKRDEIQKLNPLDGKIA EINDVIEKSRRELRKERW LNKVEDMEIKRKTDSNC	cd01650	442	698

				TAAATGGCTGG		LWSLLKGLKPKARSSGP			
				GCCAGAAGAAA		VNLPNGTTTTNAKAIAN			
				CAGTGWGCTCG		AMAKNLVTVGEGRENK			
				GTGCTATCATTA		DEKMAWRRAKRNLVKS			
				GCAAGGAGAAG		KTTLATTITHAEVKLAIT			
				CCGGACGTTKT		NGKSSRAMGPDGLCHIH			
						LKNLPENGIHALAELFN			
						ATISENKIPDSWKKANV			
						FMLPKPGKDLTEIKSYRP			
						VSLLSPVAKVLERVILER			
						IKPEIQSPTSQHAFKKAH			
						STTTAIIEVTNRIVGGLN			
						MPKPANRTIMTCLDLKA			
						AFDTVSIARLLCELDGAI			
						QDKAITRWLGNYLYKR			
						KIRTQFGEEKSKWRTLR			
						GGVPQGAVLSPALFNFY			
						IREVPTIPGAEIIGYADDT			
						TILVQDPILQKAADTSQN			
						ALDXMKTWFDERHLDIS			
						AGKSSVTIFTADTKEFNF			
						DPGLTWDGTPIPMQNRT			
						RLLGVEFDTMMKMNGQ			

						VAKVSERLAKSNNILRA LSGASWGSSKETMLKTF KAIIKPLATYAGPAWHQ LMSDTQMAKLERQYIG GLKACCGLTKDTPKELV YYETKMMPLKEELELGS EQFAISAMRTPGHPAKD LHTRRPA			
RTE	RTE-2_CJap	.	Caenorhabditis japonica	CTTCTACACGTA ATGAGCCTCTCC GTGGCGTAGAA GATCAACACCT GGATCTTTTCTT CAACTCCGGCG GATTTCGTCACC CCGGCAAAGAT TCAAGTGGCTA ATCATTACACA CTGCCTCGCTCA GGTCGGATGTA GTGTCGAATAC TTCCTCCCTGCG CACATTCCTCGG	TGATCGAAGAACG GATCGACTAAGTAA CTAAGTAAGTAAGT AA	MMHRDGTGVFLGKRNT TSVSGGVGFIVNSTLLPK IQEIKFLNHRIAYLTFQV DRRLSCTVFQVYAPVAD SDPEELADFYESLEEAYT ACRSRYKLVVGDFNARI GPRQHTERFIGPHSMEPR NESGELLATLCESNRLW HMNSQFGKPSHRRWTHI SPNKLHKHEIDHILANG KFVTDVSVVPSITTGSDH RLLRGNLHFNTKIARFE QVKRRKPPKRVLDPAVA HAITATATITPSADIDLD YINLIAMLKELQDQAVT	cd01650	447	720

				TTTAGCAGGCC		QPSNHSTNRLSGTTRQL			
				GTTGAAGAGTA		LTKRRFMDRXDPNFKSL			
				GGCATACCCGC		SNECREAVKKDHEEFAK			
				TACGAACAGCA		DRLLQAAQNKKSMKKV			
				CCAACAACTGC		ARDIQEYKTFIPCLKSST			
				CTCAGTCCTCCG		SGTRITSRTEMEKEIQQF			
				CATCCGGCCTCT		YSKLFRSNRGPPVIASPI			
				ACTCTTGTTACG		GMEIPPPFLPEEVRAAFR			
				GCAGATTCAAG		SFPNGKAAGDDKITADF			
				AGTATGTCGGA		LKSCHDNVILXLTDRFT			
				AGAGGCGATCT		KYLQSGKVPEKWKTSN			
				AATCACGGAGT		TTLIFKKGDKEDLENYR			
				ATCGAACGTCT		PICLLPVLYKAFTKCVLN			
				GCCACACTGGA		RIRTTLEEAQPVEQAGFR			
				CGGAACAAGAA		RSFSTIDHIHSIQRILEVG			
				CACCCGGCCAA		REYQQPLTLVFIDFHKAF			
				AAACAAGAAGA		DSVEPQAIWESLKSQGV			
				ATCGACTACAG		QPAYIALLQQCYTNCTT			
				AGTGTGTACGTT		TITPFHKKIEIPITRGVRQ			
				CAACTGCAGAT		GDPISPNLFSACLESVFN			
				CGGTTGCCACC		KMSWQHLRGDEDDEND			
				GATGACCGACT		KANLLPGIRINGQNLTHL			
				TGCAGAGCTGC		RFADDIVLVASNPKAAS			

				TAGAGGAGACA AAAACGATCAA GTACGACATTA TCGGACTGTCG GAAACGAAGAG AGCAGCTGAAA CACAC		LMIQELVEMCGKVGLR MNTNKTKVMRNKFASQ SPVNVTLNNVTTCIEEV NEYVYLGRLLNHNNELE PELHRRRRAAWAAFNNI KNTTDALSCPRIRAQLFD SIVLPALTYGSEVWTFTQ ALSERVRVTHAALERRL VGISLSEQRQRNLHREDI RAQSEVRDPLLVIKKKK LGWAGHIMRRNDGRWT RLVQEWYPIGEKRPVGR PRMRWSDSLKEQISLFD GNHLKTHWSTIAKDRM AWKAVIRDHIR			
CR1	CR1-13_CJap	.	Caenorhabditis japonica	G	TAACTACTGCTGAA CGGACGTCACCAAG ATCCTTTATCACTT CTACCCCACCTTTT TTTTTCTGCTCTCAT TCTTGATCTGTTTCT ACCCAACTCCTTAT TGCTCTCATTATTTC	VLVNRHEIQFLIDRAKPD ILCLQETWLSEKTPSPTL VNPPSRYQIFRTDRAHE KRGGGVCTIVKSEIPAK VITSETKNMAYEILALDI NISTTPIRLINVYRPPTCS YANTLTLLTKISDMTNX NGQIIILGDFNIKNIKWIP	cd01650	470	724

					CCTGCACATACCTT TAATGTATCTGTGT ACACAATATCAACA CGTCACATTGGCGT AAATAAATAAATA AATAAATAAATAA ATAAA	AGPMTKDRLGTILSDFC KNNNLKQFVNXPTRKN NILGIVFSNTKLENVEISX PIGSSDHRKITFTLDILEK PTKHLIEIWDFKGINWK VCAAELNLIDWPTIIDPV KTISEIYNEFLTIMRRICE LAVPKKRINKELYNVPE YLVTFHKTIENKFDKLL KSRSWADYALYIELSRK FRKKLLKFRRNQELLTL KKSGNSHFTHIAKNILKP KFSMVPTFVDQSGHHAT SDHEKAEVLAVNFESQY VNKSNTSFTLPPDYKSD QEIPWVTDAEMLKFIQK CKNSCSDTPDKVPFKFL KLIAPFIASPLSQICNLSM MRGEVPEKWQQSIITPIN KIAKPQVPTDFRPISITSQ ICRIYERCILSKLLPFLDK IRFWNNNQHGFRPKRST VSCMLESLNDWTNSIER			

						GSQVDVIYFDYAKAFDR VPHDLLLDKLVSIKLNG NLLKWLASYLSNRTFTV KVTKNQSSTKMAQCGV PQGAVLSPILFGIYVNEIP SKLPMGVHCKQFADDIK LYSEIPKDSDPQNNKLQ QAIDAICQWSETKKLAL NNSKTILISLGKTTKNCD YVINGSPIGRKQLTRDLG FMISPKLDFSEHWHKAT NQAKYALSLIFKKYCSD NSRLLTLLYKIFIRPILEY GTVISNPHKKSDIKSVES VQNAFTRRLYSRQKGRY IRADDPEYLSAPERNSLY GLSSLENRRKYFDRKFL TKMIAGNVDIDTRPFFTF FQNTRTRAKTKFIWARS KTKIRQEFFTNRTLTAIT L			

CR1	CR1-17_CJap	.	Caenorhabditis japonica	AT	TAATTTTTTGTCTACATAGATTTTCCTAACATCTCTTTTTAACTTCTTTCGTTTGTTTCTAAATCTTCCGTCATATCGTTTCTTATACGTGTGTAATAGTTTGTTATTTATATATGTGCTGATGAATAAGTGGCAGCTGCCCTTCAATCAATAAACAAATACAACAAATA	KILVSGSSKNFFKFINQRMKESSDVGMIMDGKTXVYDNYRKAELFAESFGKVYSTDDGQYTTCDLRTSAHLSEIIFEPYAVELELSKLASKVNTSPDEIPAIFLKKICTSIALPLSIIFNTSMRTGFLPDLWKTAIVKPLFKKGSRSDANNYRPISLTSTICKTMERMIRKEMMKHIYGNKLLSNAQYGFRPHRGTEAQLLHYISCIQSNMEKKKDMSSVYVDFRKAFDTVSLPKLQIKMESYGICGRLLNWLSSFLTDRHQSVVLNEVFSPRMVVKSGVPQGSVLGPVLFLLYVNDIGDSLECDTLLYADDLKIFSTDSAALQRDLDRISNWCRTWQMDVAPNKCEFITFNYSKKNSALSFPLMIGNFNLQRTTEIRDLGVIFSNNLSFKSHISVIVR	cd01650	134	384

						KAHQRINILFNIIRLSTID VFIKCYVIYVRPILEYGS VVFSPISKELVKTIESLQ KSFIYRCYKKFGIVYNSY FEALEQNNLETLERRRL MADLVFIYKWLVSREINI GHDLFISSANLLPLRRHN YYIRCQLTGATRIGSQFL SNRIIRCWNSLPVKTFPV KFSSTAFKNRLQNLSLN NFLLLHFHNF			
CR1	CR1-8_CRem	.	Caenorhabditis remanei		TAGAGCTTATTCTT ATTCTGTACTCCCA AACCACCTTTCCCC GCATTACCACATTG CTCAATTCTTAATG CCTTCTTTTATTCCC CTCATACCGGTCAT TCTTTTTTTTTGTTG CCCCACATCTCTCA CTTTAACCCACCCT TTGAATGGTCTCGT GAGGGACTCATTCT	VIPIPKKGSLSSPENXRPI SLTDPFARLFERVICEYI KLHFAHKFSQNQHGFLA YRSCTSSLVHSISCYKSS LSSNNSLDVIFFDFKKAF DKVNHKLLLQKLALFGI PHLFIEWFSNFLSGRTFSI KIEDFTDTSITQIPSGVPQ GSVSGPLLFLIFINDLLLD LALIPSLQVSAFADDIKI YSSNPVAVQKGIDLIET WASSNSLPLAHTKTSLL	cd01650	1	246

					CAGCCTCCCGCCCC ATATATTGATAATT TTTTATAAGAATAA AATACAAATACAA AATACA	RLGSKNISFPYFIAGQPIE TSKSVRDLGLITDSTLKF KSHINKTIASALLRTKQL LKSFKSTSPQFYIFLFNC YVLPIIEYCSVVYSPPPAS KLSLSLETPLRFFTRKIFQ RCNITYSSYSDRLAQLNL FSLRHRRLKAQLLLLYK FLSGTSYFPHLDSYIRFSS STRRPMNLICIKPKCSDF FSHTIPIWNAITSQSSYFL SPSEFNTLISSSITRY			
CR1	CR1-5_CRem	.	Caenorhabditis remanei		TAAATTATAACTGT TTCTTTTTTCTTAT TTGTATAAGTGGTG AAAGTCCTTCAAAT AAATAAATAAATC	VQKGPNSLAVDNSTKQ VKFLSVGYANCFSVKNK LAQLELVTITNNFDIVCL TETKLDNTFPDSLLSLSN NFSVVRKDRNKHGGGV AILISKSIRFLPIEIPSSLLS SEIAGVDIMAGGVSIRIV VGYHPSHHSKLDGMISC LEFLLSTRKNCVILGDFN MPHISWPSLTASDSHCK KFLSFVTKNGLTQHIASP	cd01650	484	735

						TRLKPDNILDLCFTNTDI LRDVRVGDLFSDHKLIH VTLSVKNRTKKLTKEVK LFRKADYASINYLLSNT DWVLRFSNLNADGMYE NLLSILHELIASYVPVKAI NTLSKRHSAEILKLQKA KLNVWRKEGNSTNYKN ISADLKVALIKEERRVND EKLTNGSVKDFFKFINSR YKDNQEIGTLKNDSGAP INCDSEKVELFSDSFSKV FTEDNNVQPHFDKRTEE LVSSPDFEPYIIEHTLSKL TPKLNTTPDGIPALFLKN VCTAIALPLSIIFRESFRT SIVPTAWKTAIVKPLHK KGSRANPNNYRPISLTSS VGKVMEKLVRKQLTNY LNSNRLLSNCQYGFRSS MSTEAQLLSYQADILTN YICKKTTHSVYIDFKKAF DKVSLSKLKIKLRSYGIH			

						DDFFNWLCAFLTNRTQR VCINNVFSSDRSVLSGVP QGSVLGPLLFLLFINDIG DAFESNYLLYADDLKLF STNADCIKKDLVRLSVW CDNWQMGVAPEKCEVI SFNHSKKHSNSASLNFSI NDAIIPQTKIIRDLGIIFNS DLNFSNHLDVTLRKAHQ RVNIFFNVLRHADFEIFI KCFKIYVRPLLEYGSTVF SPTSKEQVRLIESVQKTFI FRVFRKFNIPYISYFESIK HCDILSLEHRRIIIDLVFM YKILVSKEVRIYNDVHIS LPRFSNNLRRHPYYIKSK LSNSTKNTCQFFTNRIISC WNKLPPHIFPSFPNSDVF RCNISFNFVEPFLILKHSN F			
RTE	AviRTE_CAn	.	Calyptea nna	CCCTCCACGGG CACCCCCTTGTC GTGGGGGAGGA	G	MLDTADSGRPERRSALI AHELSRLNIDIAALSEVH LHEEGSLKEHGASYTLY	cd01650	531	809

				GCTTGCGTGCCC		WSGKPKTERHLSGVGF			
				TTGTGAGGTTG		MIKNSIASKLENLXTGHS			
				GGAGCTCTGCT		DRIISLRLPLHNKQHVVL			
				GGCGGTGGCAT		FSIYAPTLQADHAEKVK			
				ATGCCTCCAGT		FYTDLRRLTQKVPAGDK			
				AGGGTCTCCCA		IIILGDFNARVGKNYEA			
				TGCCAGACAGG		WKGVLGKHGVGNCND			
				TCACAACTGAA		NGRLLLEFCAEQQLTITN			
				GGGTCAGACAA		TIFQQKDSLKTTWMHPR			
				AGTGTGTCCAT		SKHWHLXDYVLVRQRN			
				GGGCAGGATGG		VGDVRHTRAMPSAECH			
				GCTCAGTAGCC		TDHRLVRCKLTLYFKPK			
				TTCAGGCAACC		PKRGGIPRRRLKVNDLQ			
				ATCCTAGGAGA		KDTVRNSFQVNLQARLL			
				AGGACAACTCT		DHPIDPSPEALWQHIKNS			
				AACTCCAAACC		ILQSSEESLGFSSKKNKD			
				CGGGCAGATGG		WFDENNQEIQELLTKKR			
				AGCTCGTTTAGC		TAHQAHLAQPSCHVRK			
				CCTGTAAGGCC		AAFHLACSKLQQKLRDI			
				ATCCATCTAAG		QNKWWINLAEKTELCA			
				AGAAGGATACT		VTGDHRGLYEVLKAVY			
				CTAACTAAACC		GPTHQVQSPLLSADGQM			
				TACGTCCTGAG		LLTDKTSILNRWSEHFQT			

				GACTTCGCTGTC		LFSASCVVQGSAIQHITQ			
				ACTGTCCAAGC		QPVKTELDIAPTLGETLK			
				TTGCTGGGCCCT		AIQQLKTGKAAGVDGIP			
				GGCAGATGAAC		SEVWKHGGQALHAKFH			
				CTTAGGAGTAA		ELLVSCWEQGKLPPDLR			
				AGGGTGGGGCC		DAVIITLYKKKGEKSNCS			
				AGTTCCGTGCA		NYQGITLLSIAGKILARIL			
				CGCTGCGCCTC		LNRLVPAIAEEVLPESQC			
				ACCTAAAAAAG		GFRANRSTTDMVFVLRQ			
				CCACAGCGCAG		LQEKCREQNKGLYVTFI			
				GCTCGAAGGGT		DLTKAFDTVSRKGLWQI			
				GAACCCACACT		LERLGCPPKFLKMIILLH			
				TGTAAAGCCCC		EDQQGQVRYGDALSESF			
				TGCAGCGACAG		PIINGVKQGCVLTPTLFTI			
				GCGAGGGWCGA		FFSTMLQRATADLDEEN			
				AACGGCAGGGG		GIYIRYXTDGSLFNLRRL			
				AAAGGTGCACT		KAHTKTLNHHVPELLFA			
				GGAAGCCGCAG		DDAALVAHTEAALQRIT			
				ACCCAACCCTG		SCFAEAADLFGLEVXLK			
				CATGCAGGCGG		KTEVLYQPAPQEDFHHP			
				CTCAGGACATT		HITIGESELKSVQQFSYL			
				GGTCGCTCGAG		GCIISSDAKINKEIDNRLA			
				ACTGACCCAGG		KAYRAFGRLHKRVWSN			

				AGATGACAGTC		KHLKKSTKISVYRAIVLS			
				TTGTTCGGCAGC		TLLYGSESWVIYRHHLR			
				ATCCTGAGTGA		LLERFHQCCLHTILNIHW			
				CCAAGCAGCCT		TDCVTNVPVLEQAGVTS			
				TTTTCAAGGAG		IEAMILRTQLCWAGHVS			
				AGCACTGCTTG		RMEDHHLPKIVLXGELA			
				CTCCACATGGA		TGCRNRGAPKRRYKDSL			
				GAGGGGCCTAG		KQSLSFGHTDCHQXSTL			
				AAAAGGTGGCC		ASSRETWRHTIHNAAAS			
				TAACCAAAGCT		FXNTHRISLEEKRQHRK			
				CATCTCCACCCC		NRALAVPSQETFCCAFC			
				CCACCCGGTTG		XHTACPTLALSAISVLAA			
				GCTASCCACGGT		SVGRALLKSSFVKPSH			
				CAACGGGCATC					
				TTCTGATGCAGT					
				CGAACAAAGAC					
				AAAGAGACAAA					
				GGCATACACCT					
				GCCTCTAAAGG					
				TGTGTTCAGATT					
				GACACTCGCAT					
				GTTGGAACATC					
				AGAACC					

Daphne	Daphne-17_CTe	.	Capitella teleta	CACGCGGCGGTGCAACCGCACCAGCGAAAGACGACG	TAGGGCAACTACAA GATCTTATTCCTAT GTAAAACACCCTAG CTACTCTTACACTT TTTCTTCTCTTTCCT ATTTGATTAGTCCT CCTAGTTCTTCTAG CTCTATCCCCGCTC TTTATTCTTTAACC GTTTTGTGTCTCAC TTATATGTTTACCC TGTTTTCGGTACTTT CTTTAACAATTTCT ATTTCTTTTGTACTT TTATAACTTTACTG TACTATGTTTGTAT TATTTTTTGTATTGT ATTACCATCATGAT AATCATTCAATTCT CACTCATTTGGTGG CATACATATGTTAC GTTCAGTTATGAAC TTCGTATGTCTCCT	MTSLTTMAAIATTHART DWMTNQPNMAERTDN DSLVIGTYTATVHLDST DNDALLRSHYNLNQVL QTNQQEQDNTPINFSPY YDHEHLIQHLAQPTNNV KFKILSLNCQSLNAKFD QLMILIAQLNHNXVIFDA ICLQETWIKESELPPVFNI KGYQAIHQARHCSGHG GLLIFLKDSYKFKIIPFEK SAVWESQFINIDFENGN KDSLLVGNIYRPPHKVK DDLDLFSNQFTRNLLQC SKYKNIAIAGDFNINLLD VNSTTSPAAHDFPSSIISN DFLPKITLPTRISESSSTLI DNFFCKLKNDSTSSGVVI HKISDHQAYFTCLPIRSN KIHNIEKIKMRRVSQESII AIGNDIRNATGRLDLSTE TDPNVNYEALMETLHDS MSQHIPIKEVKFNRHKH	cd01650	615	877

					GACCCTTGTCGAAT AAATAATCTATCTA TCTATCT	KGSSWITAGIVRSIRERD KLYCKLIKAKPNSNKRE SIKLTLQLYTRILQNIIRK AKSSYYSMHFAQLKND MKKTWKAINDILHRNK NSLNRNYPEQFKINDQLI SDKNIIAEEFNSFFTSVG KILASRIAHNSTNTATDY LSGNVTSRFVLQKVSES DIERIIAQLKSVSSCGYD MVTPKLLKALSPFLIMPL TIIVNQAISTGIFPDSLKR AKVVPIYKKRCPFIKKD DETVLTNYRPISLLPAIS KIFEKVIHNQLQHYFTSN DLLYHSQHGFRPKHSTE LAALQLLDQITTDLDAN KIPIAVYLDLSKAFDTIN HNILLKKLNYYGITGTA YEMMKSYLTNRKQYVS FADCNSRDIMIDTGVPQ GSILGPLLFIIYINDLHKS CSNFLPIVYADDTTLYKT			

						INSPNSINSINQELEHVNL WLKIXKLSLNASKTRYM VFHKPRKRLPTLNISIDN ENVERCISFNYLGIHLNQ NLTWNDHIKRTSLKIAR ATGRINCLKHFLPKKILL TLYNSLILPHLTYGILAW GAKSNDIAKLQKKSVRII TASKYNAHTEPLFKAHH LLKIEDIYKVYQLNFYFK WKNGNLPKYFNSFSCNT NRDTHSYHTRTTSTVSA RVNHEFAKHSLRHSLIIL LNQTPNIITDKITTHSYH GFSSYVKNYYINTYTNV CYNSTCYICRNQQQLL			
CR1	CR1-6_CTe	.	Capitella teleta	GAACAGGAGAA AGATGCCATGA GAAAGAGAAAT GTCATCATACA CGGACTGCCAG AAAAGAAGACC AACAACAGAGA	TAAAATGTATCAAA ATAAAATAAAATA AAAGGAGCTGATC GTAAGAGGCTATTT AAGCCTGAAGATCA GAAGAGGACCTGT AAGTACCTGTAAGT	MTMYIKDHIDYTMLDIG EDTCEVQAARLLMEQG QATIFASVYRSPNSSKEN SDNINAFIRNFSNERAYT LIAGDFNYPTIGWKTMT GRTEADIEFIEALRDGFLI QHVDRPTRGRGSNNPSL	cd01650	423	675

				GGAGAGAAAGA		LDLVLTKEHTATPTVEY			
				AAGAAGATGAG		HQPLGKSDHSILDIRIEM			
				CAACACAATCG		KQSETFIKTIPNYNRGNF			
				AGAGGATTATG		DRMRRVLSDVDWEEGL			
				GACGCATTGGA		GREETTQKKWELFHETI			
				ACTGAACCTAG		TNAVNRCVPKIKIKRKV			
				AAATTGAAAAC		LKNIPVDNKTLTKIRRKK			
				TTCTACCGACTG		RLWKQYLETGNQTIYHD			
				GGAAAANTACC		YCRTRNQVRQLTRKNM			
				CAATGAAGACG		RELEKNIAKEAKTNPKK			
				CCGACAGGGAA		FWAYANQKTKYRETIPQ			
				CATACTACTAG		LLKTGNPNDGMLTENN			
				CAGACCCCTGA		QEKAQTLATFFNSVFSE			
				AGTTGTGCTTCA		KPNDHWNLSPSSYPTIEQ			
				ATAGTGAAGAC		DLIITEASVLHHLSKLKK			
				AAGGTGAAGAT		TKSPGPDGIHPRILIEIKEI			
				GTTCATGGGGA		IAKPLSIIYQHSLDQGQL			
				GGCTGAACAAA		PDTWKTANITAIHKKGD			
				TTAAGAGATGC		KRIAGNYRPVSLTSVPC			
				AGAGGAAGATA		KIMEKIIRKSIMDHLIRN			
				TCAAACGAATT		NILSNKQYGFITGRSTLL			
				AGAGTTCAACA		QLLTVLDQWTEILDRGS			
				CGACATGAATT		EIDVIFLDFKKAFDRVPH			

				TACAAGAAAGA		QRLIDKLEYYGINGKILT			
				GCAGAGATTAG		WISAFLINRKQRVTIHEA			
				AGAACTGGTAC		TSAWAEVKSGIPQGSVL			
				AGGAAGCAAAA		GPTLFIIFINSLPDVTSSPT			
				AACTTAACATT		FLFADDAKIFREITKPTD			
				AACAGAGAGTG		SLLLQEDINNMYHWTLE			
				GAGACTTTGTA		SLLTFNPDKCSSMTISRS			
				CACGTCGTGAG		GTNHRQYQMNNSPLNK			
				AGGAAAAAAGA		TIIERDLGVILDNRLTFD			
				TAATCAGAGTG		HHIQAKIKKSNSIMALIR			
				AAAGCCAGAAA		RTYTYLDNHTFLLLYKS			
				GAACGGTCAGA		FVRPHLEYANQVWKPH			
				GCAGGAGAGAG		LQKHITAIENVQRRATRI			
				ACCCAAGAGGA		LPGMKDLNYEQRLRKL			
				CTCAACATCTG		KLPTLQYRRIRGDLIETY			
				GATCACGAACA		KIATEYYDTQITQNFFQF			
				TAGACACCCTT		NSRTSRGHPLKLYHQPA			
				ACAATAGCAAA		RTELRKNSFKIRIAQLWN			
				ATTTAACGAATT		QLPSEIVASPSIATFERRL			
				AAAAACCATAA		DHHLKNHPKHNPFRPP			
				TAAAGGAATCG					
				CCACCGGATAT					
				CATAGCCCTAC					

				AAGAAGTTAAA CCGAAAAACGC ACGTACAAACT GGGAGGCAAAG AATTATATAAT ACCAGGATACA ATATAGAAGAA AGGATAAGAGA AGATGGAAGAG GA					
L2	L2-6_CTe	.	Capitella teleta	T	TAAAAAGAAGTTA AAGACTTTTTTTAT TCAATAATGCTTTT AATAACGCAGAAC AGTTTTAAGTGGAA ATGAACTTGGCTTG TTTTTATGTTTCTAT GATCTTTTATCCAA ATACATTTTCATAT GTTTTATTTTGTGCT TTGAAACTTTTATG TATTTTAAGCACTT GAATATAATAACAT	QLIKDKKAAYYKKLFSE ANHKETFQLLKGTTKPP AKVLRQEQEDVTLAFEF QQFFNTKIRKIHDTLQTD ENNSPHSHDKRTSASFT AFRTHTQSQVEKIIKSAP SKTCELDPMPTWLLKEQ NVLDQVLPWVTGITNSS LELSHVPKDHKTANVKP RAKKDGAAYFDLSSYRP VSNLQFLSKVLERAVAS QLTEYLEKETLMDPLQS AYRRGHSTETAVLKVK	cd01650	150	418

					TTATTTATTTATTTA TTTATTTATTTATTT ATTTGTTTATTTA	NDIDLAIDEGDGIILVLL DLSAAFDTLDHAILLKR LEKCLGIKGSALDWIESY LRNRPQRVKVGESYSSD SILQTGVPQGSVLGPLLF LLYVLPLQEVISRFRIDR HHYADDTQLYGRIRLNS IKTIDEDVREMEASIQDV RIWMKKNKLKLNEKKT EALIITSKEMKKHAEGIK LKIGGDTVTATSKVKNL GGWFDEHMKMQEQVN RTTKAGYFHLRTINRIRK NLDTPTCAKAINATVTS RLDYHNALISGARQCIM KPMQRLQNHSARVITMT DKREHIRPVLEDLHWLP VKERSDYKVLTFVHNAL HDELAPAYMKEMYRVY HPNRALRSGEDRWTLTV GTAARHEGHRSPXIHGA IMWNTLPEALRSEASHP RF			

Nimb	Nimb-2_CTe	.	Capitella teleta	GCGGCGGAGATTAAATAGATATGGTGTCTTAGATCCCGATGCCATTGATCCTTCTCTTGAGGAGGACATGGCCGTGGATGTGCCTCCTTCCACTTCCCTGACGAGCCCAGGGAAGATGCAATCCACAAAGGAGGGACCCAGCCCTGAAAAGGGCCCCTCCTAATAGACAAG	CAA	MTILQWNCRGFRANFQELKTLVAQKDATVLCLQEIKVKDSDSCSLNGFNTLKKCTLGIDGRAVGGAAVLIKKDTPHQHIPLTTNLQAVAVRVTLHRLLTVCSIYLPPGRSHGDLVELDRLVQQLPPPLLLLGDFNAHSELWGSEEPRPSGRVVEDFIAGNDLSILNTGSQTYLHPASGSFTVIDLSLCSPSAHIDFTWEVDTDQHGSDHFPIFISSECRPSLSHPKWVFKRADWTAFGSLCEIEITEDIFEDVDDPLDIFTTLLIKIAHETIPKSCSSTKCRKPWYTDECKKAIRLRRAALQKFRHHPNDDNLLSYRKAQAVVRRVIKASKRASWKNYISKLTIRSSVKKTWDMVRKISGKYKGHQIHHLTTRSGDTATDKESIAEALADEFALNSSTSHY	cd01650	489	746

						TDKFQKYKDKAERKVL NFNSNNTESYNLPFTLLE LLMSLKSSHDTSVGPDDI HYQLLKHLPINCKMILL NLFNRIWVGEPFPSAWR NAIVIAIPKPGKDISLPTN YRPIALTSCLCKTMERM VNGRLVYFLEYNNLLSP SQSGFRSQRSTADHLVS LETYIRDGFAHGDHVLS VFFDLEKAYDTTWKCGI ISDLFNMGLRGHLPVFIR NFLSHRQFSVRLGSVDSS PHPQEMGVPQGCVLSVT LFSIKINSITNAVGADIHK CLYVDDFTISYRSRYLPT AERKIQLALNGLQKWA DESGFRFSPTKTVSVCFS GRRGVFPDPLLTLNGQK IPIVEKTKFLGVIFDKKL NFKEHIQFLRGKCQKVL NLLRVISKMNWGADRE VLLRLFRSLLRSRLDYG			

						AIVYGSARISYLERLKPI QNQALRLCLGAFRTSPV TSLHTEANEMPMAIRQK LLSVQFALRVCSDTTNP AHQCIFNYNNDRFYLSK PNAIRPLALRIKEDLQTIC PDIKAITPNQLFNTPYWL LRPPELDISLIHFGKKTA NPNYTLRNEFYNLMDK YPDHKVIFTDGSKSDSA VACSATADNLRIQIRLPD SASIFSAELLAIYQVLTLL ECSANDQQQFLIATDSLS SLQAIGNFNIKHPYVFKI LEKCTLLHKKGIDLVMA WCPSHVGVMGNERADL LAKEALSFTTCTIRIPSSD FKPITHEFYKEKWQEQW SSEQENKLYCIQPTLGK WAKSSREIRREEIVLARA RIGHSHLTHGYLLRREM PPVCIPCQSILSVKHIFIEC VDFDILRRRFFNVSSMK			

CR1	CR1-13_CTe	.	Capitella teleta		TGATTTTACCCTAATTTCCTTGACACATATTCTTTGTTGCCATTGTATTTTTTTAAGCTGGCAGTTCTTATAGGTTTTCCTGTTCTGCCATGCTGTTGCATGCATTGATTTTATTGTACATATTATTTTTGCATGCATGAAGCTTAATAAATAAAAAAATA	LDCPQIPDPHFTVENVLSVLASLDTNTAFSPQDIPPIILKKCRSLLAPSLCIIFNVSIRCKTVPLEWKKARVVPVFKKGDKSAVSNYRPISILCIVSKVMERCVFNHMYPLLSSFIHPLQHGFCKGRSCTSQLLDVYNRIGSILDKGGQIDVVFLDFSKAFDCVSHKLLLYKLKHFYGVDGNLLEWIRSYLSERSQSVIVEGKSSTWLGVPSGVPQGSILGPLLFLLFINDMPYKTLHCTTALFADDSKCFKSINTTHDCELLQRDLNAMHEWSLIWQMSFNPSKCKVLSITRSHAPIHFNYAMNGVILDHVNVFKDLGVFIDQTLSFSTHIESTILKCNKMCGIVKRTIGFNAPVNVKSSLFMSLCRSHLDFASPVWSPHLKSKVKQIESVQRSMTRFILNN	cd01650	69	322

						ADLSYIERCQFLEILPLSF RREITDLLFLFKYFNGA MNVDFTSIVAKRCFNTR LRSGPFIELLTPTPRTDSF INFYFNRVAHMWNILPL EIRSCLSVSTFKSKLLSFY SNKMCFYDVNNSCTWT STCRCNGFYHT			
L2	L2-9_CTe	.	Capitella teleta	AGATCAGGTAG CACCCTCCGGT AGCACCTAAAA TTTGCTCCGACC ATTTTGCTGTTT TCCGATATCTTC CACTCCGCGGA TTGCAGCGAAT CTTGGTCAAGA C	TGATGTTTACATGT TCCCCCTTTTTATTT TCAAATTGTATTAG CGTTCTGAGCATTT TTATGGAAGCGACG CTTTATAAAATAAA TTTTCATTTTCATTT TCATTTTCATTTTC	MWKVIFVICVLSQIKAK TGSGETRPGFTICLPQPR HASPTSMPSLCLEFIRAL AVKAPARLPTDLYKRLN GLRISRRRLSKRGTKGY GRGVIPVQVSALRSADR YVNTARRDASLPTSLVTI EPMTIASSSSCVLLNCRSI CNKASEIRDFISDRRMDF LCLTETWLTGTDEDKVV SAALLPDDYAMVHRPR DARGGGVAFIHRRSITLS QIRLGDYSSFEVLDALV CTARRLRMCVVYRPPPS RRNNLTVAQFMSEFADF	cd01650	620	878

						LSDIVSLPGEPLIVGDFNF HLDDCGDADASTFSDLL TSFGLKQHVRGPTHEKG HTLDLVITKTDTALIWSS VSRDLCISDHYPVFTTLA IPKVLNPTREISYRRIKAI DKEQLDRLIEASSLSSPP ENASLDENVTNYQRTLT GILDTIAPLRSRTVRIRV NADWYNDEIRAEKQRR RQAERRWRKSKLPADR ELFKVAKNRTQSMIDEA KSLHFREKVSACSDDPK QLFRVVDTLLGNSTESPL PVHDSRVELSNRFSDFFV EKIATIQSAIPSVDNPPRP QAPLPQSQLATFQPLDV ADTCALISGAAPKSCEL DPLPSPLVKDLPSLRPIIT RIINQSLATGVFPMEYKS AVVRPLLKKPSLDRDSL KNYRPVSNLSFVSKLIER AVASQLASHLRNNDLM			

						EPRQSAYRPGHSTETAL LCIHNEIVCAIGEGKAVL LVLLDLSAAFDTVHHPT MLSSLERLGVTGTPLKW FESYLSERSQAVQIGDA KSDPAAINCGVPQGSVL GPILFTIYTSSLGALLREQ GVSYHLYADDSQLLLTC DPRDPAPAVARMEACIR SVQAWMSCHSLKMNGD KTEILLITSKPMRKVLPD VQLCIDGHQLQPSTTAR NLGVTFNRHASMESHIT SVCRLSYMQLRRLRRIK KLLPRDVLESLVHAFVT SRLDYCNSLYLGIPQTQL DRLQRIQNSAARLVTDS GAREHITPVLQDLHWLP VKERVIFKVLLIVFKSVH NLTPQYLSEMITVHIPTR HLRSSNQNFLTVPFTRSS FIKNSAFSVAGPTIFNSLP			

						INIRQAESVPIFKRLLKTH LFHNYFLT			
Rex 1	Rex1-1_CTe	.	Capitella teleta	GGTGTACAAAC TTCGTCATCAAA CACACTTCGTGC ACAGCTGATGT TGAGATCCTAT GTGTGCAGTGT CGACCATTTTAT CTACCGCGTGA GATCAGCTGTG TCGTCTTCATCG TGATCTACATCC CTCCCAGTGGC GATGCCAACAG AGCCACGGAAG CGATAGCGTCT GTTGCTTTGGAC ATACAACAAAC AAAGCCAGAGG CAGCTATCGTC ATCACAGGAGA CTTTAATGGAG	TAATGTTATTATCA TCATTAATGGCTTA ATTGTCATATTCAT CATTCATTTTACTC ATTATTATCTGTTT GGCAAAATCATTCC TTCGGGCGAATAAA ATCATCTTATCTTA TCTCATCTCATCTC A	MIYCLPTYIRKLERQKP MIKTIRQWTEDAAEQLG GCFACTDWNIFEDTTTNI NEYTNVITDYIKFCEETII PTKEIKIYPNNRPWVTN DLIHAVRKRRRLWMSG HHEECKEAQKEVNTIIA QCKDNYKKKIEESFRSG NPKKTWSGLNTITGYKK KPMTRTSISNDEGWANE LNVFYTRFEKEDEDLVF DQPAIAPIRFEEEDVRRV FMGIKQDKAQGPDGLSP RLLKACAQDLAPVFTGL FNRSLEESCVPTLWKCS TIVPVPKRPTAKELNDFR PVALTSIVMKSMERLVL RRLLHQVQLSMDSNQFA YKARRCVEDASALMVH LVSEHLETPGSYARVLF VDFSSAFNTMRPSILISK	cd01650	257	510

				CTTCGTTACATG ACGCCCTCCCG ACCTACGTACA ATACGTAAATTT CAACACAAGAG GTCGCAGCTGT CTCGATCTACTG TACAGCAACAT AAAAGATGCCT ACAAGACCACA AGCCTGCCACC CATCGGCAGAT CTGACCACACG		LEAQGVAPPLCQWILSY LKNRPQQVRVNQSLSTI RHTNIGAPQGCVLSPVLF TIYTDEHRATSQETRVIK YADDTALVALLSPGKEH LYTDAVSHFIDQCDNDD LKLNSTKTKEMCIDFRT HNTIITPIIIKGQPIEQVTT YKYLGITIQSDLRWTSHI STQCKKASARLYHLRKL RQFHVDRHLQELFFTATI DSLLLFGITVWGGRCTA RDRNNISRVQRIAGKIM NTTVQPWGVSHRQRTIM RAKKIIREPLHPLHYTFQ RLPSDRRYRALKAKTNR LRQTFVPSAIRLINSQPI			
L2	L2-3_CTe	.	Capitella teleta	GACTTCAATTTT CAACAACATGT GACTGATGCAA CACATGTCAAT GGTCACACAAT CGACCTAATC	TAATGATTTTCCTTT TTACCTCTGTTTGT ACAGCGCATTTGAA TGTTCTTTGATATT GCTGAAAATGCGTT TTAAATGTAATAAA	MSLTHEHLSIKVIDVDRS VTSDHFGVVFTINCPKPA REKQKLKVRKWKNVNI SALDETIASHIPAIESAQD PFAAYNNCLTQVVDEFA PETTIEVTKRTDTPWYN	cd01650	306	560

					TGAAATGAAATGA AAT	AELRHLKTECRRLERKW RRTQLKIDQEIYRKKCR CLRRLRDRTKMTYISSK LQNATSTKDTYNILNQL LHKQSSSPLPRHDDEIEL AETFSSFFSEKIERVRSAI GAHPDPDVTIEHHLSTEI TPLTSFRPVSRDWISKTL SSTTPKSCNLDPIPTALL RKCPSAVPLLCFTINWSL ANSTMPDSHKTAHVRPR LKKTSLDHDQLKNYRPV SNLSFISKMTEKAIVAQL KDHCFVNNIDVQMQSA YRAGHSTETAMVRVQN DLLDAVDRDGGAILVLL DLSAAFDTLDHPLILHAL KYNIGVTDSALKWFESY LTNRKQLVTIGSQTSDT KNLVYGVPQGSVLGPIL FTMYTRPLSNIISNCGMN HHLYADDTQLYIAVNVR SPTSVSSVKESIQECSNQ			

						VHRWMTANFLKVNEEK TEVLVITKPSLTEIAPSSF NIIGETITSSASVRNLGV MLDNKLSMADQVRNIC QKSFFHLHMIRKIRNFIT EDAARQLVHANVMSYI DYCNALLAGSPSYLIDR LQHIQNCAARLVKCADR HSPSLPLLKELHWLPIKF RIMFKINVITFKALRGLA PPYICSLINVYTPARSLRS SNFTLLSVPKFHLQSYG YRTFAHQAPILWNRLPS NLQNQQSLDAFKAQLKT YYFNIAFSSI			
L2	L2-13_CTe	.	Capitella teleta	C	TGACATGGCTTTGT GAAGCGCATAGAA ATGCATCATGCATC ATGTTGCGCTATAT AAAATAATAAATTC AATTCAATTCAATT CAATTC	DDLNDLVSSFNRRCLDV TNAHAPLRQITLKGDGM KPWYSDAIHSARVVRRK YERQFRKSPLEVHRQIY VDQCKEVVQLIQNTKKE YFHQKFASASAKEVFRL VDNLLHKEPNHTLPTNV PLGDLPQTFNKFFYDKV	cd01650	233	497

						HQIRAELDASPTLPSLTT PQPLVAERGGAPALETF EICSAEDLRKIIQTSGSKT CMQDSIPTPLLKEGAVL EAALPAIVNIINRSITEGC VPAELKKAVIAPILKKTG LDPNNLAHYRPVSNIGFI AKVMEKVIARQLTDHLT RHNLHDQLQSAYRKGC SVETSLLYIKSDIELILDA GDAVLMILLDLSAAFDTI DHALLLTRLSEIGLEGSA LAWVKSYLMERTQVVR IGDKISESSDLDIGVPQGS VLGPLLFLIYMIPLGSIID NRGVLRHSYADDTQLY CRVNIKDPTKLQKQIRD MEECITSVRSWMKSNKL KLNDAKTEVLLISRHEPR QPVEIRVGNAVIKPAPLV KNLGAYLDCRLSMKDQ VNRTTSAVYYHLQRIAK VRPFLSDSACANAINAT			

						VTSRLDFQNALLMGLPD CQLRKLQVAQNNAARL LTRTKKREHISPVLTALH WLPIKHRIEFKVLSMIHK SFHSLSAPDYLKKLCRP YVPSRNLRSGDDPFKVV QLRTRNLYGARTLGGLG ARLWNALPMSLRKPMT FSTFKKNLKTHFFRISQ			
Daphne	Daphne-10_CTe	.	Capitellateleta	CT	TAGTATCATCCTCG ATCTACAACTAACA TATGAAACGAACTA TTTCACTCTGAATA TGCGCTGCACCACC CACTATTTTTTCATT TTTTTTTCTCTCCGC ACCAAAGCTGAATA AAATTTTAAATTGC ACAATTTGAACAGG CGACTCACAAGTAG CCATACTTAGTCGC CTATTGACATGACC CTATTCTTTTTCGCT	KMFLDLMHSHAFFXLID KPTRITATTSTLIDNIFSN VINXTHLSGVLYTDISDH LPLFTVNQFICRNNSIVT KTIGYRSFNETNKLAFQ NAPLAIDXDXLFXCHNC GAAYTLFTSNFVDCYNE SFPIKSHKINPKRNQPWL SPELKRSINRKNRLYRRY HARPSVFNEINYKSYKS MITKEIRVAKRTYYHDL LEQNRNKIXKTWDIIKE VIGIDNTKKPIASITINDS TVSEPIEIAQHFNKYFSSI	cd01650	337	594

					TCTGTATTTCCATTT	GNNLASSLPINHLDPMR			
					CTCTAGTATGTATT	YMNPNTPRSLFLSPVTSN			
					TCCATATTTCATTG	ELYTIIFNLKDASAGIDDI			
					TCGCTTCCTGCTTTC	KPIIIKEFSQLIIKPLLHIF			
					TTTACCAACCTCCC	NLSFTQGSVPCAMKKAN			
					TTTTTCTATTTTGGC	VTPIFKGGDPTSLGNYRP			
					ATCTTATATTTGCT	ISVLPVFAKILEKLMYNR			
					AATATTTCATTCTTT	VYNYLTENRILFNKQFG			
					CTGTCGTGTCAATA	FRRGYSTEMALLTALDE			
					AATGCAATTCAATT	LSSALDMKRNVIGVYLD			
					CAATTCAATTCAAT	LRKAFDTVNIDILLAKLF			
					TCAATTCAA	KYGIRGIAHQWFTSYLN			
						DRSQRVKIGCSFSEMKSI			
						SCGVPQGSTLGPLLFLLY			
						INDLPNVLTNIFPIIFADD			
						TSLFMTGVDINDMTEAF			
						NHQLSKINEWLIANKLS			
						LNIQKTYSMLFTLSPHLY			
						SSPIPITINNTEIQRVHSTR			
						FLGVIIDDRLTWANHITH			
						VSNKIAKSIGIINKVKHS			
						LKQDTLMILYYSLIYPYL			
						TYCFLLWGKASHSHINK			

						LTILQKRVIRIITNSGFRD HTRPLFNTLRIIHMDDLF TYLCSVYIFKCIKIIFPRSF LDILNYFTDYNSNAPRSS TRASLTRSIIVPLCRTTLR QKQLRFQSIKLYNNFLM PLEILDNSNSISHFRRFLK NVLS			
L2	L2-11_CTe	.	Capitella teleta		TAAATAAATAAATA TAAATTAATTAA	KRNQLTRATRRAKIKYY SDKINESASPKEIFQVTSF LLQSKREVVLPKATTMS SGSLADEFAEFFENKITA IRNGISAAPVQSSPLSDV VALHTFALTNAEEVSTLI RKSTSKSSAVDPIPTWLP KKCILTFSPILAAIVSASL VSGCVPDQLKAAVVTPL LKRPTLDPEVYKNYRPV SNLSYVSKLIERVVAERL KEHCALNNIQCKFQSAY KQNHSTETALIRVHNDIL RAVDRDGGAILVLLDLS SAFDTIDHTLLLHHLEHQ	cd01650	155	425

						MAIGDTALKWFSSYLSE RTQRVKVKESLSSPRRM LYGVPQGSVLGPVLFTL YTSPISAIIRAQQLQYHL YADDTQLYMSFTPTSAA STTSSGASMESAVSAIKR WMDDHYLKLNEDKTE MLVIQRPNPRSSQPTHRP TLTLDGNNINPSTSIRNL GVLFNSAFNLEDHVNAL CRTMFFHVNRIWRIREY LTTSATRALMQAHVISR LDYCNALFYGLPVVVRN RLQKAQNAAARVVVRA KRSDHITPILLQLHWLPV EFRVQYKICLMTYKALH GLAPAYIDELISVHRPTR RLRSSSDIRISPVQYKLK SVGGRCFSSAAPSLWNE LPQSIRLSSHLLEYKSKL KTHFFRKAFNDIYLNKS TSYHFIYILLYIAFEQFYF FLNFAIHKFEVINK			

Proto2	Proto2-1_CTe	.	Capitella teleta	GCATCAAGGAA GGTGTGCAAGG GAGGACGAATA TTGTTGATGGTC GAGTAGGGGAA TCTGAGATCTGT GATGTGTTCCAT GATGGCTATGA GCATCTGTATCA GTCTGTGCCTTC TGATCAGAGTG AACTCGAGGAA CTCCAGCGCGA ACTTTCTGCGGG AGTAGATCGCA CATGTAAAGTA GGCTTATGTTAC TGTGATCACGT GATCAGCCCAC GTGATATTGAG TTTGCCGTTCGT AAACTGAAACC AGGTAAATCTG	TA	MSDHFIHGCPELYVHLS LFLNAMLIHLDVSSVMS RSVLVPIPKNRKKSMNS ASNYRSIAISSIVGKCLD NVILVKHADVLSSCDLQ FGFKRTHSTTQCTFVLN EVVDFYMRGGSSVYVT LLDATKAFDRVHYTKLF RLLVDRSMCNATAKLLL CMYVNQSLVVRWQSVT SDSFPVTNGVKQGGVLS PMLFCVYIETLLERLRRS GIGCYIGSTFAGAIGYAD DVALLAPSLITAQKMLN VCEDFSREFFVQFNSSKS KAIVFNAPPNAVRSMRK LTLNGSPIDYVADALHL GHVVGDKSASLNMENA CRDLVVRVNMLCSNFNF VHLDTILLLFNSYCTSYY GCSFWDFSPRTINGFVV CWRKCVRRLLDLPYRTR SKXLHLIVLKPEPRHQLF	cd01650	35	295

				ATGCCTGTCCGC AGTTA		FRFQSFYVGCTTSTNHIV SLCTRLLPFSSSSVACNL RQLMSFVNINYDQLNNL IDNHSIRKRLSEFCEATT DQESLCCASSVIELMKIR EGSLISPINPTEATDLLKH MCCSND			
L2	L2-15_CTe	.	Capitella teleta	AGCACATATCT GTTGCGACCCA TCACTCCGGAC ACACTCTTGACC TGGTCATTACGC GAGAGAATGAC CTTGTATTGAAC TCAACTGTCAA GGATTCCCTGAT CTCAGACCATA AGGTGATCCAG CTCAGCATTAA CATTGCTCGTCC CCGAGCAGAAC AGAAGAGAATG A	TGATTTTATTAATA CGTCCTCATACTGT TTTTATCACTTTTGT GCTCTTAAAAGTCG CTCATTTTGTTTTTA GTCACTTTTAATTTT AGTTACTTTGTTCA CTTTTTATATCCATT CAGTGTTCATCTTC TTTTACTCAGCGCT ATTGAGCATTTTTG GATTTTAGCGCTAT AGAAATAAATTATT CTTATTCTTATTCTT A	RTGSSRHSMLSLLLPASP MPWKSSETLTTLLTSTH TTKYLPQHLTDMHLRKP ESSHYALTLLGTMMNSE HQNMSEKEGTQVENKW SRDRKETYKEQRDIYNR LLNQTKTQYYSKMVTD KARNPKDLFQLVDLLIH GKETNPVLPEHTSLDELS DRFATFFSEKIETIRLNLS NRQTDSQLLVHQEAEPE LVSSWMVFSRVTXEEVT RVIKKSPSKTCSLDPIPT WLLKACLPALCETITAII NDSLSSGYVPDSMKVAA VSPILKKTNLDQNELKH	cd01650	258	520

						YRPISNLTFVSKTLERIV AHQLNIYMAENHLHEQF QSAYRQHHSVETALVRV QNDILMNIDRGDAVLLIL LDLSAAFDTVHHDTLIR RLEGIGIGGIVLNWFRSY LKDRHQSVTIDKHSSPR RQLGQGVPQGSVLGPIL FSIYMIPLAQIIERHNMS GHYYADDTQIYISFRPSS PLADIEAKANACCSEIQQ WMARNMLQLNEGKTD VILFGTKNILRTLDQKAS ISIGGSVVLSSASVRNLG CVQDSELKLDLHVNQIC RSAYYHLRNINKIRRYLP LDVTEQLVHSFVTSRIDF CNSLLIGIKKANQQKLQ RVQNAAARCILYIKSRE HITPYLHALHWLPVEYRI QFKVLLLTFKSLNNLAP AYLTSLLHQNAVLASSR RNSSRLLSIPRSRLVGAG			

						DRAFQHAAPTLWNASPP AMRTLSDVNVFKKGLK TFLFRRAYSDYC			
Daphne	Daphne-5_CTe	.	Capitella teleta	TG	TAATTATGTTCTTTT CCATTTATATATAT ATGCATATATATAT TTGTTTATTTGTTTG TTTGTTTTTTTTTT CTCGTTTTTAAATTT TTTTGTCGCTCTCTC TTTGTGCAAGACTC GTTAAAACTATTAT TTTCTCTCTTCCTTT TCTATGTACTTTTC ATGTGGCGCAAAA GCGATCACGCTTCT GCCACTACGTTATT TTCGCACCTTTTAT GTTGTGATTGAATA AAATTATTATTATT ATAAAAAA	ISSNDPCDAYPRFLNKIT THYESSFPIIKKKPKKRD NNKWITNGLIISIKYKNK LYLDYKRHETIANELAY KKYKNRLSGIISQAKKN YYQGLIDENRHNMKNL WRTLKELIGSKKRSSLSN EFLINNKPCKDKDMISNT FNQYFVNVGPKLASIIPN VDSNPEDYLSGSYPNSLF LTPVTINEIKLALNKLKT TASAGQDKLKPNIIKLIS DEISLPLTHIINLIFEHNK VPDEFKFANITPIYKAGN SMEVQNYRPISVLPAFSK LMERLLHDRIYVFLQQN EIISDYQYGFRKXFNSEM ALAVTTDNIISSLDSQKH VMGLFLDLKKAFDTVDF NILMKKLVHYGIRGNAL	cd01650	239	495

						DLLKNYLSNRKQSVKIE NSVSSELPVTCGVPQGSI LGPLLFLIYINDLPNAXT SCNPIMYADDTNLFLSG KNIDEMENKFNTDLVAL TKYLKCNRLSLNVSKTH SMLFTLNNSLRDSSLSLS IDGSIIDTVKCINFLGVKI DNKLNFSEHITHTCNKV SKSIGIIRKISKTVNRPTL LMLYNSLILPYLTYCNMI WGKSANIHINRLFILQKK ALRLINKAPLLTHTAPLF HECSXLNIHDLYIYRSTIF IFKLSHHMLPTPVSKFFN LTPLVHGLNTRRQLNYK VIVPFCRTSLGQKSLNFN IAKLYNNILQPLSLLNLD SLEKFKKSIHCILVAKYL			
L2	L2-17_CTe	.	Capitella teleta	T	TAAATTTCACATAA TTTTATATTTCAAT ATGTTTTTCGGACT ATTCTATGTGAAGC	TLKIAQVKPRPKKPGVN TMDPKLFRPISNLPTIEK LLEKCVIRDLHLYIQQH NIIDPLQSAYRPGHSTET	cd01650	12	272

					GCATAGCAGTGTTT TACCTTTTGCGCTA TATAAATAAATAAA TTGATTGATTGATT GATTATAA	AMLRIQSDIKTSLDEGN GVLLVLLDLSAAFDTLD HNILLSRLQNSAGITGVA IDWMKSYLSSRQQFVQI EEKRSVTHDLDIGVPQG SCLGPLLFVMYTTPLGAI FRKHGLSYHQYADDTQ VYCSLPSDDQNSLKNVL TRVEKCLDEVKCWMTQ NKLKLNESKTECIIFSSK RNKQLFHQTTIRFGGSVI TPVTSVRNLGAHLDCEM SMDKQINNTIKSAYYHL RRIAKIRRHLDRSSCTQA VVAFVSSRLDYHNGLLA GITQKNINRLQKVQNHA ARLVSQTAHHCHITPVL ADLHWLPIRARIDFKLM LNAHKAVHCNDAPKYM KELCMKYTPSRTLRSSS DPHLLAVKKSKRNIGAS TFDVYSCKTWNELPRDV			

						RDIEETICFKKTLKTHLY KRCF			
CR1	CR1-10_CTe	.	Capitella teleta	GCAACCTTTTTG TGGAAGCTAAG ACCACGGAAGG CCTCGACGAGC TTCGAAACAAG ATAGCCCCTCTC ACAAGCGATCT CTTTGGGGAAG CGGCTTCATTGC GCCTGATGGAC GACAGCAAGAC CAGGCCAAAAC ACGTCATCATC AGAAATGTTCC TGCCCACTATG ACGCGGACACC ATCCTGGAGGA GGCCCGAGTCG AATACAGCAGC GCCTCATCCGTC ATTAACTTGACC	TAAATATTCTGCAA ATTTATGACCTGTC AAGAGGACCTCCTC ATGACATGTACTGA ATAAAGTTTTCTAT TGAAAAAAAAAA	MACEWDGKNSLLKLGS ENQPSVSDVSISAMQAPP VAQSPVVSTARLSTLAL PYAQVILNVPTAKASTR HPASPVHPTSKPKKENM DSHNIWITNIRGLVSNFD SFRLEIEHAHPYLVIVTE TFLSPDVPDQALLIPGYS FLRRDRRTHGGGVALY YRSSLRLRRLTSLETQDH EVIWFQASTREHDIIGCA AYWPPGSDLDITQHLQH ALDEVSIEFNTIVLLAGD LNAHHPDLSESISTNTAG RLLQRFATVNALTQVIQ APTRVTAHSSSCLDILLT NIPEEINFLGVYPGIGTSD HGVACVDLKTTSVITHT EDSATIRPRLHFNTHSTN WKTVNDHFSNINWHHR FAACNVEETWTLFKDA	cd01650	587	851

				AAGCCAGGGAG CACCGCCAGAA GGAGGCCCATA AAGCTGAGCCT TGACGACAGCA ATCAATTGGAG CGACTTCTCAAC C		YEEVLGRLSKKSRPTSR RTNHERHQHNTLNQELI RLRATQQAAWRLHSTA RTRETHQSLVTARQEFT LACANHIRRTDEATVKR MTESSNAKQWHKYCKQ LYRGSAVKEAIPIMTVG TESYITAKDKATILNQTF IAKSRASSRPRFPSLKKR TDSTLADILFNEYRVKKI LQDLNINKASGPDGIQPS TLKNCSDSLHVPLAILFT KSFTEGVVPTEWKLAAV VPVYKNKGQRTDPNQY RPISLTSCVGKVMEKIIN DALLQHLMGNELIVDNQ FGFLPHHSTTDQIALLLH ELLAATTTKKSVVACFL DLAAAFDSVPHPAILQK LPAYGVRGRLFRWIESY LSFRTQYVVVDQSSSPIA QVSSGVPQGSVIAPTLFL LFINDLCHSLSANTSLSH			

						PDCHPEALVYADDTML YRTGKDPALICPAIEEDL RIAAKWADIWGMRFNA SKTVAMYISRTTATPPPI TFAGATLEYSHEHRHLG FILDSALSFHAHVNALTR KGATEVFLLRRLSYKVK DRDLLLKIYKMYVRPHL EYASPAWAALTVTQTGL LESLQRRAMRIILALPRN HHITDADYAALNITSLR HRRNFALACFMFKLNSN RLPRKLSKYRPIPHTNPY NTRNCVLRAPYVPYPCL RLLDRSPFLFGLKILNFIS HTFHSLPDISSFKSSLRES SYYMPLAQIYY			
Daphne	Daphne-4_CTe	.	Capitella teleta	CGATGTACTCTT TGCGAATTCTTT TCTGCCAACAA TCAACAAGCCA ACTAGAATCAC AGGGTCGTCTG	TAAAATCCTTCACA ATCCCCCCTGCTAA AATTATCTATCCCT AACAGCTCATCTCC TAACGCTTTTGTGA ATATCCTGTTTGTT	MKNSIKMKNKLYCKYL KTKNPNDQKQYKSYKIT LRSILYKAERTHYDQLF CRNKNNLRKSWEVIKTV IHKQKNNNKQQYNMEV NGKKTSDLDIITKHFNEY	cd01650	195	449

				CGACTATTATTG	AGTTTGTTTATTCG	FINVGPNLSKSIPYSNIDP			
				ATAACATTTTCA	TTGTTTGTTTGTAA	ISYITQTNRYSMFTKPTN			
				TCAATAATCTA	AATTGTAAATATCC	ETEIGNIIRALKNSAPGP			
				ACAAGAGCATC	ATTGCGGGCACATA	DGIPSSILKENTAHFIQPL			
				AACTACATCCG	CCCATGTTCTGGGT	THILNLSLSQGIVPNEMK			
				GAATCCTTGCC	CAGCCCCTCTCACT	IAEIKPLFKSGNKHLVNN			
				AACTCAATCTCT	CCACTTGTGTTTTTT	YRPISLLPAFSKIFEKCM			
				GATCACTGTCCT	TCTTTTCTTTTTGAA	ASRLIEFITKHNLLYQYQ			
				GTATTCTGCTTC	TAAACACATTTTTG	FGFREKHSTNMALHLLI			
				ACACCATTCCA	AATTTGAATTGA	DKITNAFDQNKVLVGVA			
				CAATAATAAGA		LDFRKAFDTIDPSILISKL			
				GCAACCAAAAT		EKYGIRGCTLDWFRDYL			
				AACTTCATCAC		HNRKQYVHLEGIKSSPL			
				AAAAGGCTATT		SIKCGVPQGSILGPLLFLL			
				CACTGACCGAA		YINDLPQSSQLLPIIFADD			
				ATAAAAATAAA		TNVFYSGDDSEECIGIIN			
				TTCCACGACCTC		QEMINLIKWIRSNKLSLN			
				ATAAGCTCCAC		VEKTQYIVFSKRKIAPDN			
				GCAATGGGATC		NDVNINGERLNRVDSFK			
				AAGTACTCAAT		FLGVIIDDRLSWKQHIAY			
				GAAGCAAACTG		IKTKVSKSIGILRQARPV			
				CCGAACAGCAT		LNKTTLLTVYYSFIYPYL			
				TCACATCATTCT		NYCIDVWGHCSKQDLQ			

				ACACACAATTC ACCACCATTTTT GATCAAAGTTTT CCATTAACCAA AATAAAACTTG GATATAAAAAT CGAAAACCATG GCTAACAGATG CA		SLFIVQKRAIRIINYAERN AHTDQLFKSLEILPLHHL YIFSVIIFMHKFHHKRLP PVIDELFQTNASIHSINTR QRHQLHTPLLHTNISMK SIRYRGAFIWNQTFNEID TNLTINRFKTCLKNKLFT DFPIRTSPIQ			
Daphne	Daphne-15_CTe		Capitella teleta	TCCTTGACACCA GCAACAAGGTC GTGTCTATCACC AGTGAGAACGA CCTAAGGGACT ACTGACACAGA TGACTGCCTAAT ATCACGGACTC CGATCTCTCACA ATCATTCAACCT CTTTCTGTTCTT GTTCTTGCTATG TTACAACCCCC ATCCCTACCTTT	TAACGAGCAGACG ACGCCACAGCACCA GTAGCGGCCAACTG AGTGTTACACAATG GTATGACGCTAATT ATAACGCCCACACC CCACCTCTCTTCCA AAAAAGCAATGTTT TACCTTTGCACAGC CTGATCAATATCGA AATTGCCAAATTTA TGCACCTGCACTCA AAAAATAGTTTACC ACCTCGACTCTCTG	MVPDVNPSFSIRAEDCR YIAPXEIPNLCEHYTHPL RILHVNIRSLRANYDELL NLIASYEESNRALHVLLL CETALSDANTVLFGIPGF DAVYANRKSHQRGGVA IYVRKGLNFTQRPDLSIF SEGKFETIFIELSLCQNSQ PIIIGEIYRVPSLNAPETIA FFDSFTSLLHRSNASIIFG TDQNLDLGKIDEHALTR NVYNLLESEGFLPTVNV PTRIEGVTKTIIDNIYIRG SLPSARTSVVEVKIADHL	cd01650	543	800

				ACTTGCACAAA	CTATCCTAGACTCA	PIVLTVDFHMHRKKNSP			
				GCACCTATCTGC	CATGCTCAACCTCA	LVIMHRKMDETIAQRIV			
				ACGTCTTCCACT	TCAGTACGCCACAC	DELGQCDWQSNENRDID			
				CTGGTGAGTTCC	GCTCCATCCTACCA	SAFEFFLSEMTRVLDRH			
				TATCCAACTTTC	TACCGTCCTCCCAG	APAKQIKIPSKSFRREPW			
				TTCATTCCTACC	AATAACCTCTCTAG	FTKGLQRSSRRLSKLYST			
				CATCATGCGAG	CCACTGCAAAAAGT	AAKCAKNSHQHTAYVK			
				TATCTCATACCT	TTTTTATCTGAAGG	YRNVYNSAKRAAKQLY			
				TCGCCATACCTT	TATCAGAATCTGGT	YDNLFSESKNDIKRTWK			
				ACCTACCTGTAT	CCTCCCTCAACCCT	ILNDLTKKSGKDNSPVE			
				CTGACCCTAGCT	GAAACTCGCGCACT	KLQVDDVTITEPKAIAD			
				GCGCGAACAGA	TCCCCCTCGCCGCT	AFARFFADVGPKQAGLI			
				AACCTCACTAC	TCTCCAAGCAACTA	ASHVQTDDQQNRAHLG			
				GAA	AAACGAAAATATCT	VPPSHHSVFLHPVTQQEI			
					GAATGAACTCACCA	IRIVDSLKNKKSLGLDLV			
					CCTCTCTCTAACTT	STHFVKKLKHALSAPLS			
					ACACATGTCCATCA	ALVNKSFCEGAFPNLLK			
					CCATTTTTTTTCTCT	RAKTIPIYKKGERFNMD			
					CTCTCTCTCTCTCTC	NYRPISLLPIFGKIFEKAY			
					TCTCTCTCTCTCTCT	CNRLMPFLEDRGFLCDS			
					CTCCTCCTCCTCCTC	QYGFRRKRNTTYAVSEL			
					CTCCTCCTCCTCCT	YLNSINALLKNEITLATFI			
					ACTCCTCCATCTAC	DFSKAFDTISHEILLNKM			

					ATTGTTTTTGTTTTT TTTTCTTTTTGCTTG ACAAATTCCCCAAA AGTTTCAATCTGCT TTATTTGGCCCCCT GGCCCCGTACAGAC GTTTGTGTCTCGCC GGGGGGCCTCAGGT CACGTTTAATGTCC AAATTTATAATACC CTGTAATATTCTGC CCTGAAATAAACAT TCATTCATTCATTC ATTCA	PSYGIRGRALDWIRSYLS NRSMKVSIQGCLSEEVQ LSQYGVPQGSILGPILFII YANDLNACLTHTKAIFY ADDTTLFLSGKSANVLA ERMNDDLGRAKRWCEA NSLAVNIKKTKYMLIGP NALTRCTIDLSMANEKIE RVTTFKFLGITIDERLTW KAHIESLSKKLSSGLYAL RNTKRLTSRHGLLSLYH SLIHSHLTYGCLLWGSA SDNALKPLKIAQKKAIRL VYDAIV			
RandI	RandI-4	.	Chlamydomonas reinhardtii		TAACATCCTGGCGG GGTGCAGGCCTGGC CTGCACGCCCGCCT TCACATGCTCGGTA GTGGGGGCGTACA GGAGGCGGGCCTG GCCCACCTCCTCAT CTCCCCCCACTCCC TTAATCCACACCAC	KLGPTPAALAQDPPTNP TGGPPAGPNTTNSPSPPT SANPGPAARSNTATRIVP KPAGPTKRPARRTPNPTP KLRKLATGTPSAPPPAK RLHQTTIATSHKRRHTA TLQNTPTEPTNGQAQAS PSPPPDGRISILTLNTREL LSHHESILHTVTTHTPDI	cd01650	786	1074

					TTGCCTACGGCAGG GTGTCTCTATTTCT ATTTCTATTTC	LFLTETKIPTRLHYSMGS AKKGLSSHYQTYTSNAP YNPREGTRPHAGVMVG VTHRFSRPSGTTDMNAH NTDSRLAGHVCHIRIAPP HSTPVHAIGIYAPSDDAT RAAIYKYCTDQAAKTTA AKEHLLIGGDWNAVLFP SDRSTGQLNATDRAHIA FTTSAGLHPLTPPTHEHP RPHTYLQDQATAPHMSR IDDLLVSHHTHATITTTG GGQEAVHECGDALDHR ALLSSVPLAALRLYPFPT APAAPSDPDAWVNYHPP TKDQREKARIKMEETIG QEIADLLNTLHPAHLAIT NKLYPAGLTPAQAERNT PSPAEFRALPEVSAADTT SISNTLESILQKCLDVLK SYCDKKPPRASTTSPRH HTSKTVTRLLTATSKLK ERLTAALTVTNAAGATP			

						PTTSPHPAPTPPPNTTSPT PTPPTPAEADTPTAIRSQ LLNARTLLRAAKKEARK LNTKRAARAFRRLVATQ PKRAHQKIFDETDQPRG TPALTDPTTKEVHTAPE KILSSIHQHFTAQLLPING PRTGHYDIPHSEKQYPW SRPDAPDRMPALSKPTD APATLLPGIADSTTFFGC VNHLSRRKAPGPDGIPN ELLQTIPTSLQTAIHQLM TLMWVKAHIPTHWTQS DTVLLPKKGDPLLIQNK RPIALANTLYKLYTSLIT VNAALFSEQAHIFTESQE GFLRGRNTERQVQNLLH AIEDAATTGRDLFLLYID FTSAFNTIDHDKLLIVMH DLGYPTDLIEVVRDLYG KAATSVRTEHGTTPPIPI QRGTVQGDVLSPLLFIIFI EPLLRWLHVGGRGYQY			

						GCLTKAENDKHNLSSGA FADDLVTATTQISNLRIQ VEKIEKYAAWGGLKVN VSKCQLQGILHGRARAD PTLGGPRSKPHQAALAR MLEGQIRIDGTAIPYKPP HESYKYLGIHITPTLDWS AQVQYIREKIKKRGQRI ATSLATPDQKLRMIQLTI HPAVAYSLPFMPYSTHD VTSLDSLIASTAKMCYG LKRGFPTRAILQPTNALG LGVGSIRELYVRRAGKA LVQALNDGGRLGAVTR AMLHLQHSVAQKIPADQ LRTQCRFYTSLRQLSLIT DAGLELTQQGSAYHTTL HGLSALLARTVPPGADA EQALPFRLIQPLLELTSD FRELLTPDRTCVIDANTL KNTWPRAQRRHRLALN RLTLALHAETRTRAKTH NSTAPLTTAQRTLPDGH			

						TFTGLHTAHTAPAAVD WGTLNRFLRQRAPATG QTANAGETAAATTAAA TTAAAEPEEDHHAGENH APAQTRRPGTQQRKRAC IKMVREPEGRGSPPADA PAAFDWDPSLDPLTEAH CADLIHTAPSLLGKSKPQ NPAAHLRGLLLGPTPPA TFTHRHKLSQEEQDAYL AATSDTWKHFSETYRTK ALTTTNTLYGQQERVQQ VIAHTTSKGAPFYRVQW EPTVITPGSLTLYARNRY RPKPGGLTWLLTPGNEP RVLVEWEDLEEPAAILD AQDNKQELLDALNARL AVPLGPRPPPPATDTHLS HDQRQGRWTDASNDTA RGQRTLRHYVDINVEPC NPSRDIHPTQAYTVQLG TVLDPGSPACGTHLAHV YDPRGRCIGTISAEQLSL			

						LWQRYQHAQTTWPKVH AKHDQGRGFAAAIARLL LRYKPGGGSKSHPVTEK GECMLCEGLTQTLHTAI GSTTELFASPLNVSDAAT TYYTLYEEDMLFGANY DAYSQPWDGVAVACPE PNTEDGAHAVRWALAW AEHTETLTNTRNPPTTAT LLILPTQRNAPYTRYLAH PRVTLLGSISATTTHALL GQIHNDTGITLKHDLVL VSNTPGAQLAQPIRDAL GTSLYLSGMPVDPQPQL TWAHTRWVTKQVRRM TRTRPAAPRVTPPPVPAT PATLAPFAATQPLRLNP YEGFYTDGSVQKDALG QPSAWGAAAIDARTLRC YQIDVNPGQHGCNTITR CELSAIHQVLKLDEVQA LPEVTVYTDSAASLFLIR KTINQPQLLTFSKHKALL			

						DNILALLKRRSHAGRRT NLLKVKAHSGIAHNELA DQAAVKTAKGEQIPGAE WLVEESDNDPRSKYYWI RKIIKPTVEGVPREKNEG HYAHDLNRGIATLAAPH IAQGHTRETLYTGFVAN AIPDLDIKSSTRTMNLQS GRHRKAKRCMDYLYGQ LWNNKLAARFGRTVTG RRAPRAGPDTPAPCPLC RGPDSGGHILGGCHNPQ LRAMYIKRHNIAVQTIA RAISRGDCGGCYMVMD ATSKADLPDYAADNRLP PWLLPDQEGKPAGRHLH YRPDILLIPSISLAAALNP DFVVLPSEHDTIHIIEAG YTADTNHAAKQHEKAQ QHQALAADLREAGWKV QYTPQSAISLGFAGTIRK DLHPLLTSLGVTSHKAR QCCDTLHDHAVATLNHI			

						VLTRRRLERGLPPGDPG GT			
L1	L1-6_CR	.	Chlamydomonas reinhardtii	GGCGGCGGCGG AGGCGGCGGCG GCGGCGGCTGC CCTGGCGGCGG CGGGTGCGGCG GCGCCTGGTGG TGCGGCACCCG GTGCCGCGGAG GCCCCCGCTGC CCCGGCCGCCG GGCCGGCGATG GCACCCGGTGG CGCCGCCGCGG CCACCGCCCCG GCAGTGGCGAC CGACGCGGCGC CTGCGGCCGCT GCAGCCGCGGT GACGCTGCCGC CCGCGTCCGCC CCTCTGGCTCTG	TAATTTGCTCGTGG GTGGTGGGAGCGGT GTGCCTCTGGTGTG TGCGCTGTGGGTGT GGGGCGTCGACGTG GGCGTCGCCGGACT ACGTGCCCGGTGTA CCTGCGTTGACGCC TCGCCGGTTTGATT GCGCCTCGTGCGCT AGTGCGGTTTGCTG TACGTTGGTGGGGC TTTGGCGGGCTTGG GCTTAGCTTCGACT CGCTGTGCAGCTGC CGAGACACGCTTTA CGCGAAGACCTGTT GCACGGTTGTGGAA GGTCGGTAGCCAAC TGGGGGGAGGGTG GTTTCCTTTCGGGC	QWGWWRRRWRRHAAG GGGRPCRCAPRPRPHRG DLGCPAGGSAVVRRPRQ QAAGPRPLRVGSHNVRG LRSRVQHLGCPRLHALL RCWERQQLDVVCLQET HVRAEDIDLVTADLRAA ALRLHVREWTAWWAPA PTPHQRVSGGTAILIRTC LLEQGAVTFPGGVAAPS VADGAWVGRGVSLPVQ WGGHDFTLAAAYYPLA SPEQRAFTQQWVASAV GGEGNHLLAADFNFVSD VALDTASGVPRPGDPAA ATALAAACPGLIDVLRV RRPGQRVCTYIHSHGAS RLDRILCSGSLVAQVLA CAVEAGVPSDHRLVTLT LAATPEALPPSPGLPKAR LGFLDFKDLTKDYRTWL	cd01650	604	863

				GCGCTGGCGCC	CGGAGATGGTCCCT	TAALGARPGAPAELLA			
				CGGCGCTGGTG	CTTGACCTGGGTGC	WWPQLKRAAAASAARL			
				CCGGTGCCGCC	ACGGGCACCCTTGC	SRVAAVRRIAASEEEAR			
				GCATCTGCGGC	TTGGTCGGATTGTC	ALAAAAAAAAAAVEVGA			
				GGCCGCCACGG	CCTTCTTTGTAATA	ALPEAAAAAVRARCAA			
				ATGGCAGCGGC	TCGCAACCATAA	ATAAVQAAAGAARRTR			
				AGCGGAACTGG		HEWLRGGEKPCPVLTR			
				CAGCGGCAGCG		MLRPSGGPRTIARLKRP			
				GCAGCGGCCAC		DGSATSDPTVMGQVMA			
				GGCGGCGGCGG		TYWQDVCAAPPPAPAA			
				CGGAGTGGCGG		RAQVLAALQQHGRRCT			
				CGGCGGCGACG		TEEAAQLGQLAVTAAEV			
				GCAGCTGCGGC		RAALAAAPAGRAAGPD			
				GGCCGCGGCCA		GIPVELYQYYESDFSPVL			
				TGACAATGGGG		ADVFTAVGEGAGVPGG			
				ATGGTGGGCGG		FLDGVWTFFYKAGDPLE			
				CGGCGGCAGTG		PSNYRPITLTDTDYRTLA			
				GCGGCGGCGGC		RVLCRRLQPVFGRVIDPE			
				GACGGCAGCTG		QTAFLSDRRIADNVLLL			
				CGGCGGCCGCG		QLTPGLLKAAGQNSGVA			
				GCCATGA		AFLDFYKAYDTVDRSFL			
						LACLERMGVGAGFLTW			
						VSRLLTDTRGAALVNGR			

						VSGWVPLAAGVRQGCP LAPLLYLAIAQALLSWL RSKGHGVLAGLASALAS QYADDCTPFLAGMQAV PGFLADMDVFRCASGQR LNMSKVELLVLGTSAGL GVAPAAVAGPPLPPGWR VVNTAKSLGVHFGDWK TATPPVAFDGVVGVLGL IVRMPVSMFGRAAAASA YALSKVLYHLEFAGFPS PWAPVGKLLGQVAAVV DRRLSPAKYAVAPHARP VGVSLDIMQVPPALGGV GLLPLQLHVQARQAGLA VRCVLGACGLLAFVPPW TKVAEAWLRARHPAAT PLRLLVRCDPSRVLDKL VTPECKPFLRLLGAVAH LPPITLVAPPAAGDWVY HVPLWANPALCDGAQT FDVAFADLFALPGLVYV GQLVAAHDQLEAVRAA			

						HTRPGGWGTRFGEAML DLYLTSVWRPVLGWGP RRHLPSPLLGPASPLAAA ERIAAAVALLPGGWEAA ARAARLARGRAWVMPL PADAVRQAAAGALTQV VRGLGWVLDGGPPLLLL NYSVKAGTAMQLGPQV AALRSKHLLYVRDAGVS ADGALAAAGAFAGTLA RLWALKWENYHKEAL WRVAANACWSFPKNAA QRAAGVAMPLCRACGR DMCDGDRRHYFWDCV VARALRENMAMAMLGS PPDTADDCFSRADLWLV QPPAGLAPPVWDVVCL AALSALDSGRQQVIMTG LRVRRALTSAEVMRISV SVIADFWGRLQSYVSLG LRPKAWATVPAHHPFLS RGAGDCVLLVLPDDVGS PPPSP			

RandI	RandI-2		Chlamydomonas reinhardtii		TGAGCCACGGCTCCCCCACACCTCCTCGGAGTGCGTGGTGAGGTGGCAGCGCGAGGGGGATGGGCCTCCTCGCCCTAGCCACCTTACCCCGATAATCCACGCCAGGCCCCATGGGCCGGCGTTTGAATTATTATTATTATT	MSSNPMTLQIGTQPGQQQPDLTSPSHPLRGGNNTERTTTPAAENGHIPVPLSWPRQIPSTYFREQQHLNFCTIHAINNMLGAPWLRAHAVLAFCQVVHTALADTPELQATWGHHFKHNGLFHEAAINHYLRHRPAPIDTTGTRLLLRLAAPNLDPGHSQDTILGLLPECSRNMGFMVHDGRHAYAIRKEEDSWFLVESFVHIGRKPMTPDLWQAVKGDIFCLHPYHLDMPLDPRTYPTIPNLPPPPQPPSPPPPNQTKPKRGPTQRALSTMFARQTTQLPTNHSAGDNAAPTPHPGPAETTVTRPLTQATLNITRREATEPARTQNETLNITRREATEPARTQNERPATPNSAARPTRRPRHDNQLDPEPTSPTRANTTTRPRANANMHAPNTNPTQPTANNTTQPTWAQVV	cd01650	1154	1442

						ATHAEAAGNALRTVGRI TTQAPHQRPETPPSAARP TQRPRADPPDSPTTHQT VNPNTPRRTPPGPTPDPG AGAQDNTETRPDNTTW LHIATHNVRGLGTNLLS TICHLREWNADIVILTET KLGPRTRWLKDTLRQEK LSYRTFTSTKPGTEQYK RRSAGVLMAISKRYHSG SNVHTATIPPHLHGYIVH CTIRTPHSTPLHILGVYA PEDMPTRRRIYKYCEETI NTANKTDAHVIIGGDFN AVLNEGDRTGPPDEADK HHARFVSKNKLKRLPEF GGTISMSFYQTRDGTPQ PTSRIDDILICQNIHSALP TASRTPATLQVQPCGGL FDHYSLHARVPTFPIRL WPSHTATQADTQPPETP GPWRTVVLPIPKSTLEA ARTKIEHTLANPIATLAS			

						AIQPATTAIQQHLAQHT AGATNASELQHALQTDP AIQTPDLDALANQLLEIL QDGLSILVDMCPKKPPT RGVFAPRKINKRIRESHN EIKQLRTQCDALLQQQH PYANITTTTAPNTGIPLP DPTTEQDAENPPPQQPPP AHDATTTAEPQRRPRPP PATPAIANLIDKIKRAGD TLRQILKKQTQHDIQKA RDRFQHKLATRAKQAH KDIFSTQEQSRGIPALIHP QNKAHCTDKASILDAIH AYFQGQGASITGPRSGK YKPEDRTQVYPFDPSLK NSAPDTFTLHSATTPPTH ALLPSIADSTNFDSCIDH LSRNKSPGPDGIPNELLR ALPQTLRTTLHNLLTLM WVRAHTPASWTASETV LLPKPGDPLLLKNYRPIA LANTTYKLWTSLITVAIS			

						DISHDLDLFSETQEGFLR YRNTERQTLNLVHALED AGLTKQDIYVMYADFSS AFNTISHDRLLEIMWDL GMPHDLIRVVQNLYSNA RTYIRSEHGLTAPIYIERG TVQGDTLSPVLFLVFIEP LIRWLHVGGRGYSYGCL PPHLNARYNCSSAAYAD DLAVLTNTLSDLRIQCD KIHRYSEWAGLQVNHA KCRVTGILHRRAMQDK GLNGPTCNRTLRSMLEN KIHIGTTSIPYLPATEPFK YLGVQVTMTLNWGPQF AYLCKNITDKSLSLQTC QASPEQCLRIIKSCIQSVA AYSFAVMPYTENDIRTL DNMIARLARTCYRLSRG FPTRAVLTPAEMSGLGV GSLLPIYVQKATRALTL ALNDPGRLGMVTHALL ELQTRIAGTTDMHKLRH			

						ESHYYTTLKQLSIMRAY NLTLWDHGAPYKGPPN TLLTAISRSPHPISHHLVL TLTQAGLDLRNLVGKER TPEGRACLIPSDILFQHLP RNQCQHKHRVALNRLS LCLSAECDGAREAPFNQ AESYKCHLPLPQRYRAL PNTPAFAALLPRNATPG ELTTQSTMDAHLQHMT ALHPNTEAQTPQHPNGL GPRQRQTTLAHMLRTTG APPTAPTQSPTATLTQTP TGTLPTPPLAQPMPPPLT QPTPPLTTPAGPDAPNTQ TPPTVASLTTAHQQPKS KKQRPNAIKRLQQKIDA DPRHDPLPPTDMRSHTE KSGFILNNDPDENDSGPP NFVMWAPHYFANVQKD MQGTNRMPRLGIPRHHS RFPQNRTRPNPTRTIPAH LADPYFLGSQVDDTNPL			

						NPVTAYISANPPTTPPTC AYDSRPRWADTGDFMI APEPPNWPNITGATPGDS HDPRQADHDPQGQMRA EYMQERERLRAHLDDH PPQPRPFKRWMGIRLQL LTGKLPWPTQAYHVAQ AAKQLVYSLYHDIEEISH IQAGPITKAVTAPEKGAS LHTERYYEVAWKDTTIV KGAAEAYTRMKYTGTA AAHMGYGQACHKYFLR MKWDPTSEPVDKCPTLQ GWTPAISEHLRKLAEPPT PTPQPKPKDARLTPAQQ QGRWTQADVAASHVRR TTKQNTRIDTNPCNPYL DVHPTGRYTIQIPDTYTD ERTLAHVYNPTGRHIAS VTLGTLLTLKARYEHAL STHHDATSPHRSQGGFP EDLAKLLLRTRVDSKGK TCKSAPARERSLPESLKL			

						ALYNCLASTHSDPDRPV GLVERFASPLNCSGRAT KYYSQHAADSLFGASHD AYTAPFDGLSLAHPDPD PDACTKALAWATASIEH LEHQATPAITFMIIPQGK GAASRCLASPHVTILALI PPGVSASSMLDGLGAAD EAASSKQGLQLLLIANTT GVQALNSNNNLTRLQTS LAELGAVPPPTSQPPGRR QPLEWPARVTLPRDRTA PPGRPYSHRAKARAKQL TRPQAPAARGLDTCPTP MPTLSALYTTRAAFPTR SSYLYDAATIVWTDGSC IKVAAGEDGNEMNKLG ACAWSQQTEHIYVEPGG LNSTNTIQRAELSAILVA LEYIRPRLCLLGMTQVTI ASDSLVSLYLIRRALTAP QRLLLSKHKNLLENILHE LQMYADAGVPVQLLKV			

						RSHTGLHGNEEADKGA AKVAANDMEAMGQAIR LAPDEPWSNIWWPKRNS DNFYLSDLNRGVLNSLP PEAQQGFTNNTPVLEQW AEASANMCPTLSNRTLA SSTRHPWLVKQILYARY GYLFNAQLKRRYRLGG NGNCPLCHTPDSGGHIL GGCTHPKMKGAYINRH NWAVQRVAKCISKGTY GNGFMVMDAGPLADLP HYASAQRPPAWLAAER TSSPLITRLRPDILFIPNLP AHETHREGYRPPLNKAR LPIYILEIGYSSDVRHLD KREEKSQQHQELKRLLE AAGWTVHYSTNECLAL GVTGAIPCNLTKLLTSQL GVTHTAARVCGERLHT HAVNSAGGIIKMRRVTE QELKPRGG			

Daphne	Daphne-1_CSav	.	Ciona savignyi	TG	TAAATATTATACCG TAACTGTCCAATGC CTTAGCAATATTGT AACTTCATTAATTT AGATTTTCCATACC TCGATCTCTTTAAC ACATTGTAAATCAC TGTTATATATTCAT TATGCATTTTGTTT GTTTTCTTTTTTTAG GGTTTTGTTAAAGT TTTTTTGTCATTGTA ACGTTTTTGCACAA ATTTTTCACACTGC TATTCTGTTTATTTC TTATTTATTTATCTT TATATTTTAATTTA CATTTATTATAAGG TGGCACCTGCTTCG ATAACCGCATGGTT TCTCGGTGTCTCCT TGTTAAACTTCAGT TTAATTTATTTTCGT	RFFEAEXFLSDLYXNFC NLLSHKPPEYSGDCLLQ NFFSVFSEIVDKHAPLRSI SRKKRRLLSKPWITKGL LQSIKNKNSMFSSIVKNK PNLQKYSVFKVYRNKLN HLLEISKQMYYKHEFSR SKNDIGRSWKLIRELIQI KSPTMTISKSVGITASDV TSDPINVANKFNDFFVDI GKNLASKIKPPTNHVPL KPFRKSSSFFISPSTPDEV FSLINCLDVKKGTRHEDI PTKFIKLANSIISPILSEIF NTCVKQGIYPDFLKNAE VIPIFKNGSDNLCTNYRP ISLLSQFSKIFEKLLYNRL IKYLDKHKLLSNHQYGF RENCSTTLAVADITNML LSNTDKGLYSCGIFIDLS KAFDTVDHSKLLAKLEH CYGIRGVALDLFKNYLT NRSQYTFVHNVRSSIKP	cd01650	266	523

					ATTGTTTATTATTTT GTTTTCTGTGCGTTT AACAAATAAAATAT GTTTGATTGATTGA TTGATTG	VLCGVPQGSTLGPLLFL MYINDLPLVSNFVTKLF ADDTFLCMSSKCPYELQ AKVNLELIAIDDWLRSN KLSLNYSKTSYVLFTGN LKPIGINDFHLNINNNKII RSQSAKYLGVLLDEKLN WNCHILNLSQKLARYCG IMFKLRKYVDIKTLLNIY YSLVYSQIQYAIVCWGA SNKKTLYPLNVIHNKIIR AMSYSPARCHVTPLFSD LKILKLCDIYKLEMGKL MYKFHNKTLPDSFNYIF QPLYCAHNHITRQVAHH DYFVPYFHTLLGQNSIG YRGPVAWTAIDVSLKKL GWFSFKKQLKNQYLLT YIT			
R2	YURE-2_Cis	.	Ciona savignyi	CCAAAATTACTT CCAGCACCTCC ACAGCAGACGA ACGAAGAAAGA	TAAGATCCGCGGCT GTGGCGCCGAACG AGCACCTGCCCATT CTTCTTGTAGGGAC	MAGHKITMSEGKLLEVA VRYGGVRNVSYECPVPD CTKTFSQANNLIRHLNN FGNTKHRAHNFTYFFTC	cd01650	710	981

				AGACTTACGAA	TTTTTCACCCTCACT	EKCKIQIHSNTKHNISNH			
				ATAAGAATAAG	CCCCCCAATAGTTT	YKQCCATGGGPSCETGQ			
				CAGTTTAAAGA	TTTTTCGTTTTTTCG	YFCPACEQAGLGNLESA			
				CGAAGCAGACG	TTTTTTCACCCCCA	LRHFQSSHPEFNLPPRSQ			
				AACACCACTCC	CCCCACTCGCCTCT	FSKSHPNSYTLSLKPKDH			
				ACCAACGACGC	GGGCTGCACATCCC	LMKILYSGPLTPGQLVCP			
				TCTGCAGCTAC	ACACGTAGGGACCT	IKICLRSSAARLFHDVSK			
				ACCACCACCAT	GTTTATATTATTTG	LRKHMLVDHNRTLVYE			
				CTCGCCGCAAC	CCTTTTATATGTAC	TTCGKCLRPVDTSKNMR			
				AAGAAGAATTT	CACTTTTTAAATAT	KTTSHFEKCSGESFISSPS			
				TCCGCTGCTCTC	ATTTTTGTACCCCA	PIPQKTYKLDLPSTSTPPP			
				CACTCTGCTCCA	CAAGATGCTTTTCG	RKSPKLQPYKPIRTFKNP			
				ACACCACCTCTC	CCAAAAAAAAAAAA	LTKSSQSKSDNPPKPTPF			
				CTGCTCTGTGGA	ATTTTTGTATCACA	FSPRTLERSASWPALSEV			
				CTGCTGCCTTGC	TTTTTATATTTTGTA	VDPLPKLKEKHPSLPCA			
				TGCTGGACCAA	AAACACAGATTTTT	LDKCPPSPRIKPSTLVPP			
				CCTCTACCCGA	ATAAACTTTGCACT	CHTANNSPKPTSPESPST			
				AGGAACCCTTC	ATTTTTATATAAAC	LKPLPRPIRPSKPLEDWL			
				GAACCCAGCAG	TTCGCACTTATTTA	TVRSVGPDREIVLNIGPR			
				CAAGGTACGTG	AAATGAATCGCATC	PRPGPAAGSRTTSPPSTA			
				TCACCACCTCTC	TTTTTTATATACAC	PAKRVAANPIAAPLSGEP			
				CACAGAGCCAA	CAACACAAACAGG	GATLDCGQTGRKVQPPK			
				GGCCAGAGTGG	ATGTGCAGCTCAGG	KRPTESAGSLPPPAEPAT			

				ATAGAGCAGCG	GGAACCAATCCTGC	DLLTGREGLARLVEEYH			
				GCCTCTACAAC	GTCCCTCCTAGCGG	LSGDFGAFCRDLERWTA			
				CAAGTTCTCCAC	CGGGAGGGGCGCC	LSSTNRRPKPRRGRYNR			
				TCCGACGACAA	CACCTACCCCCACG	GAAARATRNRGRDDRQ			
				AACACCTGCCT	CTCCTCTTGAGCAC	DPQDRDDQGGPGPVTC			
				CGTGGCCAGAG	CAACAGGGACTCCC	GRPQRYKRAAALRSAFG			
				TCCTGCCGAAG	TTCCGGAGCCCCTG	RDMKATVRRIIDGERGD			
				AAGAATCTTCG	CACCCTCAACTTTT	ARCEIDPKTIEGRFRDEL			
				ACCCCAACACC	CTTATTTTTAAAAA	SPPVREGPECSLPPWMA			
				GTCTCGGCAGG	AAAAAAATCATATA	EAQAGEHAPSNDSQPGD			
				GCTCCAAGCAA	TATTGATCTTGACG	AYDGPITALEVEMVLST			
				CGACATCAGCC	ACGGGGGCTACATT	LNVGSAPGSDGLSYGFW			
				ACGAGTGCCCA	CAGCCCCCAAAAAC	RALDPKGLVLSELFEVC			
				CCGGAGTGGAC	CCACCCACCATCCC	RIERRVPGPWKSSRVTLI			
				GCGGCGAACCC	CAACGAGTGCCGG	CKDAEGDLDDLGNWRPI			
				GAGGACCGCTG	GGCATTGAAGAGCT	SICQTVYKIYAAVLARR			
				TCGCCAAGAAC	CCGGCACAATTAGC	LQSWALDGGVISRSQKG			
				AATAACATCAC	ACTTAGCTTATTTA	FMPFEGVYEHVFLLDSV			
				CGCCTGCAGCA	TTATTTTTTGTCAAC	VADARATRRSLAVCWL			
				GCTACCAAACA	ATTTTTGTTTTTTCA	DLRNAFGSVDHTTIVEA			
				AAGGTTAGTCT	AATTTTTTCACCCC	LSRFGAPAGLVEMISDIY			
				CCTACCTCACCT	TCACCCCCACCCTA	TGGSCRIRTRAGFTPDIP			
				ACAAACTCGTC	ATAGGTCCCTCGGG	VGRGVRQGCPLSGIIFNL			

				TGAATAGACGC	CTTGGGCCCCTTTT	VMEVLLRGVEANNACG			
				CCCCGCGTGGG	TCGTGCTCGAGAAG	YRLSCAGGASVRVLAY			
				AGCTAACCTTG	CGTCACATCGCCCC	ADDVALVGSSRAEXKIQ			
				GTAGCTAGCTG	ACTGACCACGACCT	LGVCERFAAWAGFSFNN			
				CAGTGCCCTGC	TCCCCGACATTGGA	KKCAAMVLKHQRGGRR			
				GACAGCGGCCT	GTCCTTGGCGTCTC	LLDSAPLRLCGEEVAILG			
				CGAAGAGCTGC	CCAGGTCGAAACA	PDSFYKYLGAHTGYGRQ			
				AGAGCGTCAGC	GTCCCAAGTGATAG	TGGQLVDRVERQVVRLF			
				CTGCTTCGACCT	CACCTAATGCTCGA	TSFLTPTQKLSALKRIVL			
				CGCTTGTTCTCA	CTTGTTTCGGCCTG	PAMSFHLRVRPCAEGHL			
				TTTCAACCTACT	GGCCGCCGAGGATT	RRLDNTVRRCVKTALRL			
				CCGTCCTGTGG	CCCAGAACGACCAT	PKGSCRAFFHTSPDAGG			
				AATCAAAAGAG	TCTTCTAAATAATA	LGITSVVAECDILTVTQA			
				CCCCACTAACA	TTTATATTTCAGAA	FKMLSSPDHLVSLVAKG			
				ATTACATTCATA	TAAAACTATATATA	RLGMHAARMGRSETAS			
				AAATCTAGCAA	TATCGTTGGCGGGA	ACAMADYLSGDSVMGH			
				GAACGAAGAGA	CTTGTCCCGCCTCG	XSWKTGYRMPADLWTA			
				AGCGACCAACT	ATACCGAGTGCTGC	TRAASRRLSLRFSPQPQG			
				TTTAATCCATAA	AGAGCGGCAAAAT	EFGLESGTFKIAPRERRS			
				CTTTTAGATCTT	AAAGAAGAAACCG	LTRRLHHRQNLWWRNQ			
				TTTATTTATTAC	ACGTCGCTCTGCAG	WAALPNQGKTVAAHSA			
				TGTTTTTAAGCC	CCAAGGACCACCCA	YAASNNWVKGPSSLAP			
				CTAAGCATATT	AACTCAAGCCAGCA	QALFFGLKARLNQMPTR			

				GCCTTTTTTTAG ATCTTATAATTA TAAAAATAGAT TCAAAGTTAAC ACCACCAGGCC GCTACAGAGCA TTTTATTTAATC AATTTGTTACCG ACCTCCTGCTGC TTCTTTTTACTT TCTCCAGACTAC TACCCGGATAC AACCCTTGGAA ACGAGAGGA	CCGTCGACAACCAA CATCCTCAAGTCGG CGGTTGCTGGAACA ACTCATAACATCTT CAANATAAATTATC ACCCTGTGCAGCAG GAGGCCGTGCTTTT AAAACTACTCTGTA GTGGCTCATGATAA TATTTCGCTCCTTTT TTGCCCCGTGTAAA CTTAGTNGATGCGA ATAAAATCAGTTGA ATCA	SVKACYSRAPNYDKSCR RCGAEVETLPHVLNHCP KSMKSILERHDSVLAEV LAAIPRGTFASVDVDRTS REHFRRVGEALRPDIVA RRHDGSVVVADVTCPFE SCASALDTAAARKIEKY DQLCANLRQLYRKPVES HALVVGSLGSWGRTNN TALAALGIRGAVRSRLA KQLVNLSVEGSHNIWLR WSGGIPKDLVR			
Penelope	Penelope-1_CoFl	.	Copidosoma floridanum	TCCAATGCTTTA AAACGCTGGAA GGACCTGCACA TACAACTATCA ACCAGT	TAAGCGCTGGCTTG TTGTTGTAATTTATT GTCAGTCGATAGCA CGACAGGTGACGA GTAAGTCAACCAAA CTTACGAATTCTCG CGTACCTGCGGTAA CAATTTACATCAAA ATTTATTGTATTTTG	MAQRRFLLKCRYARVFP NHIKHLYNKCSSVNFRC KPIHADFNKSTLHFCQK LLNFEISDINTHIRFLQNS ISRLKDHIASLKLDYNIR ERHYKFVNIKQNTIYRRS NHTLHKKFKKILPKTNT TTSNNDSIHNLTNIAIPK DVAEIASLGPNFNLPSTF	SSF56672	363	534

					TGTTTGAGCCTTCC AGATCACTCATTGT GTGTTTTTAATTTTA ATTTTGTAGCCTGA TGAGGCTGACGCAT ATTGTCAGCGAAAC GTTGCGCATTAATA AATCAACGCAAGA AAATTTAAATACAC TTTACTTGTATTTAT TTATTTTCTCACATT TCGCCCCCAACGGA TCACTGATCACTTT GAATTTCCAAATTC A	RERDFLCIIKNLEVGLSH LDTNDRVKTEIRDKVSQ YLVNACKQKPAISRKDR QYAKNLKSTKNFLNKNP DIFFTSADKGNVTVCLK KSDYTERMHNLLSDKST YKIISKNPLNSLQKRTSNI LKNLNSNGFLSRTYHPN ALSLTDTTLAKCYGLPKI HKVDAPLRPIVSLINSPT YFLAKVIYDELKKSVNV PISHINNSFELKNKLKNI KPTTNYRLMSLDVSSLF TNIPLELVLRSIDRRIVTI MNKCRIPVDEIRQCTEFL FNNTFFTFDNIFYQQIFG TPMGSPISPLFADLVMD DLENDCLSILKNNYNIIP SFYFRYVDDVLLFINVD HIDQILNIFNSYHPKLQF TYEMEIDKKLNFLDLTII RNSNYFNFDWFEKPTSS GRILNYFSGHPVSYKRNI			

						VFNLVDRAILLASTRFH GVNIARVKNLLIKNNYP FNFINKNIKVRLKRHKYS NHNRSTTNTYPLCKIGIP YIDNNFEKISTMLKPYNI HTYPLINNKLQHIVRLG KDTTNKHKQTNVVYKF YCKNCPASYVGETKRA AEVRISEHMKSKNSDSV VCAHQINFDHEFDWDNF KILDKESNYKKRIISEML YIKLDKHNINRKEDVLSF NRAYLPLLRKFKCPDRV HRPLPSTFAAPPI			
Penelope	Penelope-7_CoFl	.	Copidosoma floridanum	T	TAAATTAATATTTG TGTTTTTATATCCTT TATTGTAATTTGAA TTTATCGCGCGTTT TCACTGTATAGATT GAATATATATGTAA CACATGTTGGTTTC GTCTATATATGTAT GTAAAATTTGTATT	LFNNICFDAVKMAIEKF WHHIHNKINLSLTDFLN VIEFIMNNNYFQFNNNF YLQRRGSAMGSPISPVL ADIVMRDLETACLNLCP FNIPVYFRYVDDILLCVP VDKVEYVLNVFNSYSHF LNFTIEKSENNTINFLDM TIVANNDSSISTNWFRKD	SSF56672	34	146

					TGATAATTTGAATA	THSGSYLNFYSHCPISQK			
					TGTAAGGTCAGTCG	KAIVSHLVDKNYKLSSP			
					ACTATGTGCCCTCG	EYFSNNFKYIYTVLLAN			
					TTGGGTACTGACGT	NYPLDFIESCIKKRLYFL			
					TGTTTGAATCTTCG	KSINNSQNNISPVNNAIG			
					AACAAAATAATCTT	SLVIPYCRELDYKLRVFC			
					TTTGTAAGTCTCTA	SKSFINVIFKSSNKLPFNR			
					ATTTGTAACTGTAT	MYGKDILKKLNNPNVV			
					TTAAAATTGTTTAT	YKIDCSDCSNCYIGQTLR			
					TATTGTTCTTAGTTT	KLSIRLNEHKSDIKHKRE			
					TTAAAATGTTATTA	KSALYQHHLDTGHSFNF			
					GATGTGATAACGTG	SDNEINIIDIEQNYYKRLL			
					GATGTATTGTTACC	SEMIFIEKTNNFNRMLDC			
					TCAAGTTAATTTAT	DNLNKVFKRYIHLLLK			
					TTCGTGTATATTTT				
					ATAAAATTAGGACT				
					TGAAAAGGCTCCTA				
					CCCAAAGAGGAGA				
					GAAACGTCGTCAGG				
					GGATATAATAAAA				
					AATTATATACAACA				
					GCAATAGCCATTGG				
					TTTTATTATTTGATG				

					AATTTCGTCTTGCT CGCTGTTGTTGGAG TCATAAACATCATA A				
Penelope	Penelope-5_CoFl	.	Copidosoma floridanum	TC	TAACAATTTTCATA ATTTGATAATTCAT AACAATTTATATCC ATCAATCTCTTTTT ATCGGTTAGTGGCC ACTTTCCTTGATAT TGAGTTTGATTCTT TAAAGTAAAAATA AACTTAGTTTCATA AATTACCGCTGATT TTGTACTACACCAG GTTCACCACTGTTG TAGTACTACACACA CATGTTCTCACGTA GCTTTAAAAAGAAT TCCCATATATTGTA CTATATCGCAGTTG AGTAGAATAAACAT CTCACATTGACGAC	SDINIYKNYLQKELDKIE KSLIFTLGIRFFELFERFF HYKSKIGIKTFDDILTNK FKILYFRKYKHTYEKKIN INNNSIRVSKTNSNIKSK WLHNLTTINLPKEVQFCI NLGKNFNFFDFENQNET KRVYLELFKSLNKPLHK NQSISNNIKFKILDLVKT HIKNNNKSSLDILISNML NNCNKFLKNNKNIMITK SDKGGAFVIIDKQDYINK VEESLNDKKLYEVVEHN PLPSLINKTSNLLNELNG SDLFGDSGNKKVFNTIR DKCILAKCYGLVKVHKP NQPMRLIISNIGSPTYFLS KWVSEFLKNNLPPSPHN VKNSFELKNDIKNFLIPD	SSF56672	321	494

					ACTCTGCGATTATT GTTTAATATTGTAA GTAAAGATAAAACT TTTATACATTGTTA GTGCGTTTTCTGTT GTGTACATCTTTAT ATACATATTTATAT TACCATTGCTGTAT ATTAATTTTTAGTG ACTTGATAAGGACT GCCTTGGAGTAGTC GAAACGTCGTTTTT TAAAATAAAGTCAA TTTGAACAGCATAG CCCTTAGTACTTGT CATTATTTATTGTA TTATTAGCTGTTCTT GGAATCAAACAAT ATC	GFCIVSFDVVALFNNIHF ESVKVALESVWHILNLK INLEFFSFLNIIRFIMENN YFSFNNKNYRQKFGSA MGSPCSPSLADIVMRHII KMLLFLGPFEIKLIKVYV DDLLACIKISESDKVLNY LNSIDKHIQFTMETPNNN SINYLDMTLTWDNNGSI KTNWFQKPTHANLHLN YWSHSPLSQKIGVIYSLV DRAIKLSDTCYHNNNLK KVYEILKENNYPPDFIKK HAERRKKIIYYNNKNDN NNNSNKKLNDDNKNYI VLPYIKNCDSNFKSLLN KVSYKTIFKSVNKQKFL NQLGKDSIKNKDKVCSV YCIKCKNCDKCYVGQSS RVVSKRINEHKRDVKSK KNESAIFNHVQDNGHSF DFDNFKVIDNEPNFYKR			

						LISEMVFIKKYNTVNKV TDTQSLSSVFEFYINKYL			
Penelope	Penelope-10_CoFl		Copidosoma floridanum		TGACGAGTGTTGCT TGATAGGTGTGCCA CTTGATGGCCACGT TTCTGTTGGTCCGT GTTTCCTTGTACAT ACATTGACCGTGGG TGCACCCTTATGTG GTGACAAGACCTTA ACACCATATTACTA TTTTTAATATCTTCC AGTTACATGTGCTG AGTCAGGATGTGAC TTATATTGTAGATA ACTTAACTGTACAT ACATTTTGACTGAT AAAGTTTATAATAT TAAGATTATTTAAT TATTAGTATTCCAT TGCCGTGGTTATGA CTGACTAGTCTTTA TCAGTCATGCTGCG	LLKPLEAALAVNKKIEV DTSKKIRQTYLQVIKKNI FNSKKISYFDKTINNHLN ETKMFLKEHREILVTRA DKGGATVVINKTDYINK VERALGDSMYYHRLDS QKRANHLTRIINSTKKLV DVWDNRGVFDTVEIDRP DVNQLSGCNLAKAYGLI KVHKPGLPVRIIVSAIGS PTYTLDKWFSKFFNKFC KRSPSSLKNSYDLKSTLR DLIVPHNYLLVSFDVVSL YSNLPFSLIKKSVEEKIG RTLDSMHLDRKEFLDGL DFLINSNYFTFNNHTYK QTCGTPMGGCTSPWIAE IVLESLEKSVLPRCRDSC LMYRRYVDDCLLIIQSD SVDTVLGIFNDFNPFIQF TVEREDNGVIPFLDVTIT	cd00304	283	365

					TTTGGCATTTTCAT GTTCGCATCTGCAT AGGTGACTTATATT TTCATTTAATGTAA TTTGTCATAATGTT TTTTATATTATAAA TATTGTAAATTGTA CAGTTCCCTGAAGA GGGAGTCAGGATG ACTCCGAAATATTG GATCACATTGTCCA ACTAATATATGAAA TAAATAGCCATGCT CCAAGATTCCAAGT TTTTTCATTTAAATT TCCTTAATTATGGC AAAGGAGCAAGAA GACAGTAGTTTTAT TATCAAA	RIADSVCEIDLYSKPSSS GRLLNFFSHNPVSQKIAI VFSLIDKCFKLADPKFRS RNIDRARRALTMNGYPT YFIEKYIKARISIINRFNK NKNHINSLTNCEMSPPV MEGSSRLSFIILPFYPRLE KTLRSFFSKHKLKTIFQN SNKFTNIIRLGKDQTPKL KRSGIVYRIECECGLFYV GQSKRRLATRIDEHKRM VHKCDKNSVIAAHCSEF SHTFKFDEVKILDQEXN KFKREFSEMLYISYFNRN VNRKLDTNKLHDHYKS FLYELRDRMHGWRCS			
RTEX	RTEX-22_CGi	.	Crassostrea gigas	GA	TAATTTATTCATTCS AGGATATACACTTT GATGTAACATTGTG ATTTAATATGTACC	MDNPTYSPNETLNRNFL FDEINRLVFKTKSKSASG YDEIPYGVLKSQKVIETL TRLFQLILDTSIVPSVWR	cd01650	72	345

					ATAACCCAAATATA	KAIICPILKDASTDPRIPM			
					AGTTAATAATTGAT	NYRGVSLLSCISKIYSAFI			
					ACACCTTGTATAAC	NQRITSHLDAHEYLMSK			
					GCGTGATGAACTAT	TVSEKTGRVKTRIHLNS			
					AATTGTTAATATTA	VIRIKITIRSFHDLRKCFD			
					CCAAGGTGACTTAT	FIDREMMLYKLLVHDID			
					AGGCCCAATGGGCC	GKVYNSVKSIYQNSESC			
					GGGTGCCCAATATT	VRWQSVQLCEKYLPKL			
					CATGTACAAGTAGT	RIVCALNGQLTNSCNTG			
					TGTTATAAAATTTG	VKQGDNRSPTLFSIFIND			
					TATCATACGGCACA	LVQDINELGLGVDVGEK			
					ATTGTACATAATAT	KLSLLLYADDIAMIAKS			
					GCACATGTCACTAT	KAELQRMLDKLHQWCK			
					AATAAATAATTTAC	RWRVLINTNKYKCVHFR			
					TTACTTACTTA	KGKIKRSNCNFTIGSNAL			
						ETVDHYKYLGVTFNYNS			
						CFNSNAELLAKLLAKSG			
						GRALGQIIAKIHNNKYFG			
						FTSYEKLFSSCVAPILDY			
						GAGVWGYRKFQAIDNIQ			
						NRAIRYFLGVHRFTPILF			
						LYGDVGWLPSQFRRWF			
						CMLRYWNKLICFEDDRI			

						TKMAFNLDYERCRDNW CSEVKHIMNQIGIGQYY DTKCAIDLGLVESHLTN FYSEIWKDSLENVPKLR TYKTFKTDFGLENYISFN LSKAERSHMAQLRXGIL HXRDPPLRIESGRYVGEP PNERLCTLCTQQVIENEL HFVLDCPFYNQYRTEIL GEQLLSPALLNRDRQYV FTYLMSHKFRQLSKFVV KAFRIRRQFLYR			
CR1	CR1-1_CGi	.	Crassostrea gigas	GAGGAGCAATC AATCAATCCAG TCTTTTGTACAT TTATTTTCGGTG TTATTGATTTTT CATTACACCAC CTGTTTTATATA ACTCATATTTTT TTCATTTTCTTT CATATTTACTAC ATTCTAACGCA	TAAATTCCTTTTTG CTTTTGCTTTCTATA CTGTACATTAACGT TATCTACTATATAT TTGCAATTACTTTG TAAATGTATGTCTA TGAATTTAATTATT ATAAAGTCAATAAC TATAAGTAACATGT AATATTACATGTAC TATGACAATTGTAT	MGNSLEQYRAAIGNFYF SCRCVCRSVLVNFFGFSF FSKIFMFPSIGLFALAMK HLSRHIVVYNFYLQFFM FLLLLCGDIEVNPGPIMT NVLDILHLNIRSIRNKVA NLNTIVHDFDILCFTETH LDRNVTNESLMLDGFNT IYRKDRNCFGGGIMIYLS DLLRAIRRPEFETTCNESI WIEIESHTCSYFLCCVYR	cd01650	577	830

				ATACGTTTATGC	ATATTGTATGTATT	PPNSDGNFWTSLSWSIE			
				GTTATTTAACTA	ATCATTAGGAGAGG	KVGEITDNIIIIGDLNVDL			
				CCTTTGTATATT	AACTTGTAAGTTGT	LSIPQTHVIHDIMSNGNF			
				GTTATTTACCTT	AAGAACTTGTTTCC	VNNINEPTRVTNTSSTLI			
				TTGTTTTGTCAT	GATCCTTTTGTGTT	DLVLSTPAVKVHESGVL			
				TGTGCACATAC	ACTTACACAATAAA	NIDSSVSDHKAIYISFSIDI			
				ATTTTTTATGCT	ATATGTTCAAATCA	ELNRAYERNIWVYKDA			
				ATTACTTTCACC	AAA	DFEKLNSLISNCDWDITI			
				CCTACGTTTTGT		NSAPNIDAAVAQFTSTF			
				TTATCTCTTTTG		MAYVRECVPEKTITIRPR			
				TTGAACAATATT		DKPWYDSMLRKTMRIR			
				TTGTCAATTGTC		DRLRNKALRTNNDADW			
				ATATTGAGGTA		TKFRHVRNKVNNMKKH			
				TCGTCATTTTCA		AITNYYDNIESALHDSSR			
				TTTGAATACCTT		NGKIYWKLMKDLFKVK			
				CATTCAGCTGTA		ATTEIPPLQFFENGQPTL			
				AGTTTACTACTA		AFSDIEKTETLNKYFSSIS			
				GTGGTGTATCTT		NIESANRSLPNLYYMCN			
				CCGGTTTTTGGT		NFFSDIHIEVQEVFDIISIL			
				CTTATTCATTAC		PVNKAIGPDNISHRMLK			
				CCATATACGAC		YTKFSISKPLCMLFNKSL			
				ACATTATACAG		VENSFPNYWKIAHVLPL			
				ATTGCAGCCAT		FKKDDPSVASNYRPVSL			

				CATTTCCCCCCT ATTGTTATCAAA AAGTTAATCATT GACA		LSCVSKIMERIIFKHVYN FFHENNLFYKYQAGFLP GHSTVFQLLETYHSIVKS IEDGKMCCMIFCDLSKA FDRVWHEALIFKLQTYG ISGHLLQWFKSYLSDRT QRVMYKNKLSSTLCINA GVPQGSVLGPLLFLIYIN DVATNMLSFCRLYADD NSLQHCDLYTKNIEIVLN HDLKILSDWSDKWLLKF NPSKTKAVFFTKKHVEE FPKLFFQGCQLDYVSTH RHLGLLFSNDLSWSQYI NNIVNSAYKKLGLLKKL KFTLGRTNLSKMYITFIR PVLEYASVVWDGCNMF DMERLEKVQLSAARIVT GLPILASTDSLYTETGWE SLVSRRNKAKLITMFKIH TNQVPEYLSNIIPPVTKN VSIYHTRNSENYSVPISR LELYKKSFVPDTIRKWN			

						SLKLEVREATSLNIFKKN ITESIPQPPKYFSFGKRTI NILHTKLRHNCILNYDL YRRNIIESPNCICGQKED VYHFFFVCKNYVNARN NLFELLFNRLDETILINSH LLLWGSDNLSYNTNCFI FSAVQKFINESGRFNN			
RTEX	RTEX-23_CGi	.	Crassostrea gigas	GAGAGAGAGTT GTTAGACGAGA TGAAGAGAAAC ATCAAATCTTTG CAAAAAGAAAA AAAGACATTAG GCAACTAAAAK ATTCAATCGTTG ATCTGCAATGTC GCTCTATGAAA WACAATCTAGT GTTTACGGGGC TGGGTGGCGAA ACAAAGGAAGA GGACACAGGGG	TGATTATGTACACT TTGTTTTTCTTACTG TTTATTTAATATGT ATATGTTCATATGT ATAACCTATGAGAC ATTTGTCTCTTGGA ATAAAGAACTTGAA CTTGAACTTGA	TYFCSRMLYTPESXVFY KKGKVDIFDCLEEXYCS FHRKGHIYXVGDFXARC GTLDDFISDDKLGKSVE XXLQSFXDYDNNANVY CXHNADRGVNNFGKKLI NLCRTTGLKIVNGRNEN DYNGNITFYNANGTSTI DYLLTDVYNFDNIAYFS TGIFNCFSDHAPVSFCIK TNYIDGGSKALCCANTS QGNGSINQCTSVRWKEG ADLEATWSISEHYDLFLS YVENQLYTQEDVDNCV ANFSKTLCKIMKPYCIIT	cd01650	477	736

				CAAGCTGCGGG		GGLANTVIDAERCIDKP			
				ACTCTATATATT		WFNDICKSKYRDYRGAL			
				TTGAACTGGGA		HNFNVCKSDVNRIVLCA			
				ATTGACTGGAA		KKKVYKKNSPEVIELYF			
				AGTGGAGTTCT		VQRKRFIKKTVQKFKHN			
				GTAGGGTGCAT		YERYQGDRMEYLRKCN			
				CGGTTCGGTCG		PKQFFRLFKNNKKISCD			
				ATTTCAGCASAA		VSAKEFFEHFKNLFENV			
				CAGGAGCASAC		KTTTYDNTSDDYEILPC			
				CTACTGTAGCG		YEELDVCISELEVEKAIL			
				AAATTCCTTTAT		KLKSNKSNAEDGVINXF			
				GAAAAAGACAG		FSKCKEVMIPILCRLFNN			
				CTGCAGGTCTGT		IFDSGFYPDNWSKGCIVP			
				AGCTCCCGGTCT		IFKKGEMQDLNNYRGIT			
				GAAAAAAGTTG		LVSCLAKLFTSIINTRLLE			
				AATGGATCCCA		WDQRYNIITDAQFGFKP			
				GGTCGTCCATCC		GVGTIDAIFVLQNLISRT			
				GAATTTKTTCAG		LKNKKRLYCCFVDYTK			
				ACCGGKAGCTA		AFDLIDRSKLWYKLQME			
				CAGACCTKCAG		GFGGKLLRIIKSLYKNVR			
				CTATGAAAAAG		ACVKHKGTFTETFRISA			
				ACAGATTAASTS		GVLQGEILSPLLFSMYLN			
				TCCTTGAAASAG		DFERYFINSNCTSIELQMI			

				CCTATCAATTAA		NMFLIMYADDTVLLAES			
				AWKGAMGTGG		PEGLQDMLNTLHRYTDE			
				ATTCGGTATCA		WNLSVNVQKTKIVIFRN			
				ATGAACAATTC		GKCVKDNEKWEYNSKY			
				CCGCACGCTTTT		VEVVNQFNYLGMLFNY			
				GAGGAAAGACG		NGKFLSTQKHCAAQGN			
				GAAACAGCTTT		KAMFSISSKMKDLNFNV			
				ATCCTTAACAA		ETQLNVFDTYVKSVLSY			
				AAATATGTTAG		GAEIWGFHKGQDVEKV			
				ACATGAACATA		HINFCKKVLGVRKNTCN			
				TCACTTGTAGAC		SAVYYEVGRFPLEIFRKQ			
				CAATTTTGTAGT		KIFKYWFKLRNSKNCIL			
				ACTGTTGTCTGG		KACYEDMVKNNDIWIS			
				ATACGTTTTGAA		NIRKELSILGLADLYQST			
				CATAAAAGCAA		MKESVILDIIFNRMCDVF			
				ACAGGGTAA		KQKVVNDISNASKCILY			
						KHVVDHFCLQVYLKKPI			
						HVKYRKLITRLRLSSHD			
						LKIETGRYDNLTRDCRK			
						CESCNLNVIEDEFHFLLI			
						CPSLCSLRDKYIKKYYSR			
						KPSVYKVIQLLSTSNTKE			

						LCNLGKYLYYANKQRT LHVNYPN			
CR1	CR1-3_CGi	.	Crassostrea gigas	GATGATAACTC TGTATTACCTCA AGAAATATTTA TGTGTCACAGT ATGTTTGASAGT ATGTGTATTAAT GAACAAGAAAT AATTGATATACT ATCTATTCTTCC TGTAAACAAGG CCATAGGTGCA GACCTTATTAGT CACAAAATGTT GAAAGCTACTA AACTTACCATTG TCAAACCATTAT TTCTATTGTTTA ACCGCTCACTTA GAGAATCTGTTT TTCCTATTATGT GGAAACAAGCT	TAATCACAAATCTG AGTATTTTTGTTGA TAACATTTTGCTTA TTTTTATTTATTTAT CTCGTTCATCCTTG CATACATGTAACAT TAAGTAGATAACAT TAACTTTCTTATTA CTACATTTATCAAA TATCTACGTACTGG TGTAATTTATTACT GTATAAGTACAAAC TGAGACGTATTTGC AATAATATTGTATT TGTTGTTTAACATG TAGGGAGAGGGCG AATTAAGTTTTTAT AACTTGTGCCCAAT CCCATTGCAATATA TAATGCAATAAAAT ATGTTCAAATCAA	SHPERIVFKHLYNYFHSN NMFYKYQAGFLPGHST VFQLLETYDDIVRRIDEG KSCCMVFCDLSKAFDRV WHKGLLYKLKMYGIDG KMYXWLKSYLANRTQR VMYKNLFSTKAQIHAGV PQGSVLGPLLFLIYVNDV AHNMSSFCRLFADDNCL QYSSKCVKTIENCLNHD LNVLENWSKKWLLKFN PSKTKVIFFNTRKNLEPP KLIFQDSQLEYVQTHKH LGVTFSSNLSWSSHIEILL SNAYKKIGLLKKLKFKA NRKTLSLMYTSFIRPQLE YASVVWGGCHAFDSEK LEKLQLXAARIVTGLTIF ASRESLYFETGWEPLNA RRKIQRLNIMYKIHNNL VPEYLSDIVPDMRCNAS	cd01650	4	226

				AACGTTATTCCA CTCTTCAAAAA AGATGACCCTT CTGTTTTGTCAA ACTATAGACCA GTTTCGTTATTA AGTTGTGTTGGT AA		AYYTRNSQNYNIPVCRL QIYKSSFVPTVVNEWNS LPLDIRDSSSLSLFKNKIS KSLAVAPPYFSYGKRRT NILHTRLRQNCALNIDLY RCNIISSPLCSCGKIEDAH HFFFSCPKYATVREEFFN KFLRFGKIRILDSRSLLW GDDSLSVDENKYLFSIV QIYIQGTGRFE			
Penelope	Penelope-2_CGi	.	Crassostrea gigas	TTATT	TAGACCATAATTAA TCACCTAATTGTCT ACGGACTTTTCCGG TTTTTTTCCCGATTA CGACAAAGAGCAC ACGGCGGGTGTGAC CGGTCAGCAGAGG ATGCTCACTCCTCC TAGGCACCTGATCC TACCTCTATCTATTT GGAGGTCCGTGTTG CTCTGCTTTGAATT TGTATTTCGTTTTAT	MKGPKFREPRSFTWRRN FIHIMNSVEDYAKRWAK SEKEELDTLSEWVKSIRC ILKSRIKRLRGQMKTIYP SVFNKAEVRKELDRLHD QYVLVPADKASNNIVFV CKAHYINCIFKELGFDST HGNPTYTHSSLSKQEILQ NHKSVMDIFNIPNTQNEF DLPYLYWIPKLHKSPYK QRYIAGSSKCSTKPLSLL LTKILTVVKEKLQEYCT TIYSRSGVNQMWILKNS	IPR000477	147	467

					GGATTTTTGAGATG GTTGACGGTTTGTT ATTGTCATTTTTTCA TT	KELLETLRSQNLTKINSI KTYDFSTLYTTIPHDKLK SRLFDIIDSCFFNKNGSR KNAYLVIGNLKNYFVKN HSDCTHKYSEVDIKKML EFLIDNIYVVFGNQVFQQ SVGIPMGTNCAPLLADL FLYSYEAEFIQKLVREKN KSLAVAFNSTFRYIDDV LSINNCYFHTYVDSIYPS ELEIKDTTESASSVSYLDI LLEMDIDGNLTTKLYDK RDDFNFSIVNFPYLCSNI PSSPAYGVFVSQLIRYAR ACSSYEQFLRRGKLLTN KLIKQDYQQSRLKSSFR KFYGRYNDLVSKYNLPL GRMLTDVFHTNC			
RTEX	RTEX-26_CGi	.	Crassostrea gigas	TGCGCATGCGT GAAAGTATCGA CATGGCGTAAA ATTCAGCTGCG AGAATTTTCTGT	TAATATGTGTATAT TTATGTTATAATTG TATTTGCAGTGACA TATAGGCCCAATGG GCCGGGTGTTTCAT	RAILTLSCPFIKIKSTHCH RPLTLLWNDLXWLRMG IARFMKXVSVGLYXIPW KTQISQLLQQIKFEPSEDI LSVAKELVAHLDDSAQV	cd01650	831	1088

				TTATTCCAAAAT	TATCTATTTTGTAT	VAAARLKNRKIDKPGLV			
				TTWGGTWGTTG	AACATTTTGTAATT	KISFARVDQKIRVLRAK			
				TTATGTCAACA	GTTATAACACATGT	RKLGEISDYRSVFLRSSK			
				AAAACTCCGAC	CACTATAATAAATT	SHTERLIELNARTILNEM			
				AAAAGTCCCGA	ATTTACTTACTTAC	PNGKDFRITANGRIVKK			
				CTAACGTGCCA	TTAC	TDNEPNTRDESSSELPRP			
				AATAAATCTGG		FETIRIGHLNVCGWTNK			
				ACCTGGAGCTG		NHDLRVEIIKYTDCDIVS			
				ATGTTGGGAAG		INETHLLNKENLDLDGY			
				CGAAGGCTCTC		QWMGLNRSNVHVKAPK			
				GTCGGACGGTC		GSGGVGLFIKKHLCKNY			
				AAACCCCAAAG		KVDIIDNSFDGILGVKFT			
				CCAACACCCAA		NKFSDYVFVVFSCYLPP			
				GCAAGTCAAAA		ESSPWGRDASSFFSHLLT			
				TGGCGACGAAC		QIYIQYDTDSIFICGDFNA			
				AGCTCACCGAA		RISNIPDFDKDLDTIPPRL			
				CACACCCGGCA		TIDNVHNQHGSSFLEFLN			
				AGTCCGGTAGG		DCKFCVLNGRFGNKSND			
				TCAAACATTTCT		YTFQSTRGHSVVDYICV			
				CATGCCGATTAT		PHDVYNSVLRLSNNFLM			
				GAAACATTTGT		EDLAINQMTIHFNRQEVI			
				GGAAAAAATGA		PWLTIYVFLMMFTIQCS			
				CTATTATTGCTG		DFQIISCSDIVENAGLVS			

				ATGCTGTTGATA AGTTACAGAAA GGGCAATCGAC ACTACAATCAA TTGTTGAAAGC AAATTAGACAA GTTCAAAAATG AGTTCGTTCAG GGCATCAATGA AAAATTCAAAG CCATGA		MQGENCRIPDHGFLIFDF MIRDFENAHPSENTVTSS TRPIKKYKLRCIPPSFMS SVNVRGSLLNMIRIIETC RESRENIDRTYTQLCNIIT KEMECSIPKFDWSKRSR KKFRTHKPYWNNELTEL WRVMREKEKTFLKHKK GNSVLKQKSFSEYKRAQ QELDKRMRFFRRKYNL DLAINLDNFSSNNPKKF WEELRSLGPKRKNKVPT ECYTDDNDGSISSDPTY VQNVWKSEFCNLYNPST DVAQQPSNDAFLQRVKE SVNFMENNMLDPLREH NPYLNRPIELDEVKNVI MKSKNGKSSGVDEIPYE VLKYDCVIEVIHRLFML CFESHIIPSVWRQAIICPIL KDSNSDPRVPLNYRGISL LSTIYKMYSSILNHRLVS YLEDNNLLVDEQNGFRR			

						ERSCADHVFTLNSIIQNR KETFVAFIDLQKAFDTV DRDLLKFSLLNNGIDGD FYNSIKSIYGNTVSCVRV SELCTDWFSTSAGVRQG DNLSPTLFALFINDLAKE IKELRKGITVGEINLGILL YADDIAILSDSKENLQKI LDTVSYWCEKWRLHIN MKKSGIIHFRKKGPRSDF QFQINSNILQYNTDYKY LGVIFHEKMNFQNNADT LSKAGGRALGGMISKIR SIFPKFGPYKEIGISTFEK LFLNCVAPVLDYCSEVW GLQNYHSIEMVQNRALR FYLGVHRFTPLPALHGE FGWKLCYHRHCINAVRL WNRLLTMEDDRLTKRV FLWDFNQTGLSRWTSRI KSLFESMNLLHIYDNLSI CDVNEFENKLFLTYENE WLGKINQLPKLRLYKSI			

						KTNFKTENYLEMNISKS QRSILAQFRCGILPIRIET GRYKGEPLDERICNFCV LNEIEDESHFLLNCTIYS DFRRELFNNIGYNPTLV MSNLDCIFFLLSNFPRQT SKYLJESYLYRQSLLYK KRNA			
RTEX	RTEX-21_CGi	.	Crassostrea gigas	ATCACGTGATA TGTATGGAGGT GGACGTGTTTTT ACCGGAGTACG ACCTACGGGAA TTTCATACCATA TATAGCGAAGT TTAGTTAAAGG TTGCGGTAAAT GTGATACACCA TTGATACGTGA GTATTATACTCT GTTCCT	TAACTTTAGATGTT GTTCTTTTTTGATTC TTTTCGCCTTTACTT ATGTCAATTTTTCA TATTTATGTATGTA CAATGTAATAGGGT CTCCTGTAAAATAC CCACTTCTGTACTT TACTATGTTTAATT TGTAAAATCTGATG AGCTGCAAAGCTTA AAGAAATAAAGAA TGAATG	MNLSSPSPSKPPKKKKQ RTKSSTQGKMATTQQG ACTNSVYPNDNLSQSVQ PIQQVGNLTYHQLQPMQ YIQSAHAPMQTHMQHQ PQNVQSNTPQVTPQWVS DIVTQLKNMNQKLDTID TTLSSMKREIKGLSNRV VEVEKSQEFISKSLEDNE KNTSKALTELSDLKMAL DQSLQENVRITESVLDM QCRSMRDNLLFFGIEEA QSGEQENCEDIITSLCAS KLEIRSEILIERAHRLGSK KNNKPRPIVVKFNRFPH	cd01650	920	1181

						REMVRKCSYKLKDTHIS IGEQFPKEIQERRKKLLPI YKRAKENNQKAVLARD KLFIDGKLYSDDGPVQM KATPTQPKASGRRGQKH GRGKRGQENQTGLRIIA WNIHGGLENKWSERNF KSFLCTYDIICLSECWIG KDFNISDQNFQSFVFPRK SGKSQGGGTVVLIRNVY TPFISFVESLHDTIVWLKI DKSFCNHTKDIYIAFTYF PPSNSTFLKERDIDLFME LENQLSYYMSIGDIFLFG DFNSRTKTLPDYIVNDQ VHESVLDNNLQYSSDNQ LTERVNPDLGHNEYGQR LLSLCRSTGVRIVNGRHS GGHSNDYTFNGARGLST IDYLLTTYDMFDFISKFI VCNFTTFSDHAPLHIEFK CTTLLDPCDTSSHYRTR NVFRWDSSYVDSCNDFF			

						NSIEGTCNFVDFEKVTSQ			
						SDLDECMLKITDMLSKD			
						IGNFFYVSSKNSNQSKPA			
						NSRSRYSTYLSTNDKPW			
						FNENLKSLYRDYITALRT			
						FNVCKNAQNHLNLVSK			
						KRTYKKLEAKLKRSYK			
						QIEGNRLNILKKNNPRQF			
						FAKFRKKPKSVGNVPIEI			
						FFDHFKKITNVDGSFDT			
						NCEQANMEVLIPIVNKP			
						TDECVFYELDKEIDVDEI			
						RKAISNLKRGKSHGEDG			
						ILNEYFINFQQHLLPLLY			
						SLFNCILDTGFFPKSWSS			
						SIIVPVYKKGDCTNPDN			
						YRGISLVSNMAKLFTSIL			
						NNRLLDWSKTNDVITDA			
						QFGFKPNCGTRDAIFAL			
						HSIVTNTLCKRKRLYCA			
						FIDFKKAFDSIDRSKLWL			
						KLSKVGIQGKLLNVIKA			
						LYNDVKSCVVVNGHLS			

						DYFCNNVGLMQGEVLS PILFNLYVNDFEMNFLKS GCIPHELLSLNLFLLMYA DDMVLFSESVEGLQLLL NELSHYCKTWNLCINVE KSKIVVFRNKGKIKCNE KWCIGNDVLEIVEQFTY LGIIFQYNTKFTKAEKQL SDQGRKAMFAFRKNIRG MILNHETLLSLFDCYVG GIINYASEIWGSHKGNN VEKLHLDFCKQTLGVKR SSFNAAVYAELGRNTSY YKKVLFNFKILEGCVMV GIPLITKRFCSILKFWKD VLCSNNCIIQKCYEEMYI NCELHGHKNWASDVKK ILNEIGFNEVWNKQTIDS LTLKVINQRIADLAKQDI FGKIENSQKCHFYKYLV DGFYLQYYLRKSVPVKF QQCISKLRLSSHRLAIET GRYNNTHRAQRNCFSC			

						KNYVEDEFHFILVCPVYI EYRQKYIKKYFYEKPSM FKLLQLFNNENFKDMCN LGKYIYYSMKLRNDICN			
Tx1	Tx1-5_CGi	.	Crassostrea gigas	TATAATTGTATT TCATGAACATT AATATATCTCAC ATAATTTTGGTA TGATTTTTATAA TATGCTGTCATA TTAACAAGCTC AAATCATCAGA AAATCAAGGTA CTTTTTTTTTAA ACACATAGTAT GCGCAGCATAG GGTTTAAGGTTT TTAAAAAAACA GACAATAAGTT ATACATATAAA ACTCCTCACGA CCGATTCTAATT ATTTATACTATT	TAAAATGTGTTTAT AATAGCTGCAAGG AAATTGTAAGCAAA TAATGACAAAAAA ATATAAGTCTACTT TTATGTATCCTCTT GCATATGTATGAGT TTTTGATAGATAGT GTTCCCCTCTCTTTC TCCACCCTCTTTCA CCCCCCCCCCACCTC ATAACTAAAAATAC TCNAACTTTCATGG GTGATAAAAGCAA GTTTTCGAGATTGT TTGTGCATATATAT GTCTGTCTGTCGGT GTATACACCTGTCT GTACGTGTGCTTTT	MDDKVQQSYLSILTLNC QGLGDINKRRDIFKVLK SKNYNIYFLQDTHFVEK EESLIKGLWGYKAYFNS YKSNSRGVAILINNNCDI TILNEHKDSDGNFLLLD VTIESLPFLLINIYGPNSD SPAFYINVLNKIQDIYSN QHIVIGGDFNLVFDKDL DSMNYKNLNNPKARLE VLKIMETLTLKDVYREH NPTLKKFTWRKKNPIKQ ARLDFFLISESLLNLSLY VKYENSYRSDHCPVVLY VRTNEFLKGKGFWKFN NSLLKDKEYVYLIKEKIK NVKQQYACPVYNLENIE TIDDNIQFIISDQSFLDTL LMEIRGKSISYSTYKKKK	cd01650	530	791

				GTACGATTTTTT	TTATGTGACTGTGT	TTEREKQLEQEILLLEQN			
				TTTTTAAGTTGA	TTATATTATACTTA	FTVETKDELSNKQTELE			
				CAAAAAAAAGG	TATTCTAAATCTCA	LIRKNRLQGQCIRSKVK			
				TTGTTTTTGGTC	AATGAAGTGAATAT	WIEEGEKPSKYFISLESR			
				AACTTCATGCA	GTGATGTGTATGCA	NFSSKIIPKVEKEDGSTV			
				GTCAACTATTTT	CTGTAAACTAAACA	FNQFEILNELRMFYDNL			
				TTAGTGTTTGTA	TATAAGAAGATGTA	YQKRETIDSENIYEKLKL			
				AAATTAGCAAG	AATAACTTGTACAT	YDIPKLNKQESDLLEGPL			
				CGTGATGTGCTT	GAAAGTTTGAATAA	NKLEVYNFLKNMKNDK			
				TCGCAATTCCA	AAAAAAAAAAA	SPGPDGFTSEFFKFFWND			
				AACTTCTGCCTT		LGSFITRAINNTYELGHF			
				TTAACTGTTTTT		TESNKLGVITCLPKAGKP			
				TCTCCCTTTTAT		KQFLKNWRPISLLNVLY			
				TTGCCCCCTTTC		KIASGCIANRIKTVLYKL			
				TATTCTTTTCGG		INEDQTGFLKGRFIGENI			
				GAACCGTAAAC		RMIYDIMQYTEQNQIPG			
				CTTATAAAGCA		LLMLIDFEKAFDTISWKF			
				TTTATATAGAC		IYQTLDFFNFGISIKNWIQ			
						TFYSDIKSCVIQNGIASD			
						YFFPQRGCRQGDPISPYL			
						FLLCAEVLGILIRNNKEI			
						RGIEIDGVEYKLSQYAD			
						DTSIIFDGSPQSMDGILRE			

						LDYFACISGLKINFSKTK MVWIGSKKISKEVFHHS RWKLDWNNPTFDLLGV KFSVNLNEMVELNYPSK FLEVTRLIKQWKLRNLT TVGRLTVIKTLIIPKLNH LILTIPKPSVEYLQQFEKE IYIFLWGGKVHKVKKNII IQDYRYGGLKMIDYMEF VCALKSTWIRRLIHSNTK WVKLFESELGLKIETLW EKGSDFITNISNKISNKF WKEVLLDWVKIVMSYS QNDIKVLNEHIWFNPNIK INNSSIFIKNYFNCGYKFI KDLMTEEGNFVSLETLE NSNVKTNFLEYRSLKGA VFERLKALNNNKVEYGP FIPNSLSIFYKQSKGCKD MYYVLIDNKKESIRSVK KWTESGFAYTEIEWNNI FELPFKTSKESKFQWLQF QILHRILPTNEYLFKLNII			

						DSPACSFCNQDIETIEHL FVDCFHVKEIWLYFEEW LRSKFNFLMTFDKKTILF GKFKNRNMYRLQNILIL TVKQFIFASKYKNVPNL STLMLKRTVTDRIHIEKL LLLKNCKFAEYEKYWQ TICDLL			
RTE X	RTE X- 5_CG i	.	Crassost rea gigas	AAAAATCGAGC GGCWATGACAA CATCATCAACG AGATGCTCAGT GCAGATGTGTC GACCCTTGAAC CAGTGTTAAAT GTGTTATTTAAT TCGATATTTGAT AGTGGTTATTAT CCTTTAAATTGG AAATTGGGTAT AATTGTGCCGA TTTTCAAAAAA GGAGATAAAAT	TAAGCGTAAATCAA GTTTATATATAACT ACTGCTTTGTTGAT CTTTTTGATATTTTT TCATCCACTTTATA CCTTTTATATATATT TATATAACCTGCCT CCCACCTGTACCCA CAATTATTGTAAAG TACTGTTAAATTGT ATGCTATGGCAATG TATGTAATGTATAC GCCAATAAACTACT ACTACTAC	KQSRVLQGDILSPMLFSL FLEDLETHLHENXSAGIE IDQINIFLLLFADDLAIISE TPNGLQKHLNTLHTYCE RWGLTVNIDKTKVMIFQ KGRTRKGYKFCYNNKN LEIVDSFAYLGTVLSPNG LFNKHVKITAEKTLQLY NTFMYKTQKLSLDCKT MMELFDVACSSKMNYG AEVWGYTNAKDLERVH LKFCKRLLGVKTSACTP AVYCELGRKPLLLTRTL RIIKYWFKVLRSQNCLM KSIYDYDLKNMLSSWIIF	SSF566 72	4	137

				GGATCCACAGA ACTACAGGGGC ATAACCCTACA CTCTAAGCTGCC TGGCGAAGCTT TTCTCCCGTATC CTAAACAATAG ATTACAAAGAT GGGCAGAAGAA AATAGTGTGAT ATCTGATGCGC AATATGGGTTA AAGAAAGGTTG TAGTACTATTGA TGCATTATATGT TTTAAAAAGTA TCGTTGACTTTA AAATTGCAAAT GGCAGTAAGCT CTTTTGTACATT TGTAGATTTTCG GAGAGCGTTTG ACACGGTTGAT		VKNTLLSNGFGDVWYY PESVNEKMFYVYFEQRL SDCFYQNLRVEIENAHP LRLYKYLNLPVHQAEYL NHILYDKHRKALSRLRL SSHRLRIESGRYDKKCPI NERKCIYCKSNDVEDEY HFVLVCDFYTELRKLYI KKYFYVRPSMAKFIDLL NSDSIFMLKKLGKYVYE SFNKRDNFTIRRE			

				AGAAATGCATT GTATTATAAAA TGTTTAAGTTAG GTATAGAAGGT AAAATGTTACG CATAATCAAAT CTATGTATGATG ATGTGAAAAGT TGTGTACGAGT CAATGGTAATC TTACAGATTTTT TTTTGA					
RTEX	RTEX-2_CGi	.	Crassostrea gigas	GAATCATGGAA AAGGATCTGGT GTCAATAAAAG AAGCTAATGAA AGTCTACAAAA CTCTGTTACAGA CCTAAAGGCTC GTTCAATGCGC GACAATCTGAT ATTTTCTGGAAT CCCGGAGGAGA	TAGTTGTTTTCCTAT TATGATCCTTTATA CTAAAATCTAACAT AAGTAGTGATTCTA TAGTCTGTTCTTTTT TATTTATATTTTTTC CCTGTATGAAAATG TACCTGTTTGTCTG TACCTCTGACGATT TGTATTTGTATGTA TTATATATATGTTC	MKNRHKLSNRRSGGLIL AFKDTIVKYIKVLETDC KFIFLFKLSKTLLNISEDA VCGLVYIPPEGSRYSSPD CFSEIEQEIINVESNCKYL CLFGDFNARVGKMQDY YTPDDFLAHVTHCEIQSE VATRSIIFDNYQIPLDRIV RDHTINNFGYKLVEFCK NNDIYIMNSRVGNDKNI GGTTCKDSSTVDYVLAS	cd01650	506	776

				GACATGAAGAC	CTCTTGTACCAAAT	ANFFAIVESFNIKNFCDL			
				ACCGAAGCTGT	TATTTGGTCTGAGT	FSDVHNPISFSFSICINDA			
				CCTACAAGAAT	GAATAAACTGAACT	KNNIPGKNQPSVKLWCA			
				TTATCCAAAAA	TGAACTTGAA	DKADDFRNNIQMSAVCE			
				AAGTATAAACT		IENELNVLSSCIDNICKD			
				GGATTACGAAG		NISKIAKDICTVFVNTAR			
				TCCCATTCGAA		ETFGMSKHVDNFSNEKS			
				AGGGTCCATAG		SKPWFNRVCRTARRKY			
				AATGGGTAAAT		HLAKRLYNRNNSTDNK			
				GGAAGGAGTAC		ENLKVLSKQYKRTMDG			
				AGTTCCCATCCC		CIKDYKKTITKKLVNLRS			
				AGAAATATTGT		TNPKDYWKILHGSRRDE			
				CGCCAAATTTA		KCAVSIDDMFQFLKDM			
				CCTATTTTAAGG		NEGCENGESNIDNPDFIG			
				ATAGAGAATTA		TCMYDDILNGEISDEEIV			
				ATCCGTACACG		AAVKKLKNSKAAGYDLI			
				TGCCCCCCAGA		VNEHISSTLSIFLPVYKKF			
				AGCTCCACAAA		FNIIFCSGFVPDEWLIGII			
				TCCAACGTTTGG		KPIYKNKGDPTQPENYR			
				GTAAACGAACA		PISLLSCLGKLFTSILANR			
				ATTTCCCCCTGA		LEAYATETNLIYENQAG			
				AATCGAGGAAA		FRKNYSTLDHILTLQFLC			
				AGAGAAGGAAG		DIFMKRKKKLFCAFIDFK			

				CTATATCCCGTC		QAFDTVWRTGLWSKLIL			
				CTCCGCCAGGC		NGINGKCFRYIKNMYM			
				AAAGAAAGACC		GIKSLVKLNGIASNFFEC			
				ACAAGAAAGTA		NVGVRQGENLSPFLFSL			
				AAACTGGTCCG		YINDLENFLLEKNIVGLQ			
				GGACGTGCTTT		SISNEIENELLLHLKLLIL			
				ACATTGACGGA		FYADDTVIMAESTEDLQ			
				CAGGAATATAT		NALNEFYSYCTQWKLN			
				CCCTCCTATAGG		VNVNKTKIMVFSKGPTS			
				ATCCGCTCCGA		KKLFYYKGEVIENVKEF			
				GAGCATTTCCA		KYLGIVLSRTGSFIKARK			
				ACGCCCACGCC		HLCEQAQKAMYGVIRKI			
				AGCTCAGACGC		RNFNLPVNCQFDLFDKV			
				GATCCGCCAGC		VLPVLLYGCEIWGFENL			
				AATCAAACAGC		DIIERVHLKFLKTVLNLR			
				CGAGCGACAAT		STTPNFMIYGETGRYPLY			
				ACACCCCGCAA		INVYTRIITYWGKMLISP			
				AACCCGACCAA		ENKIVYTVYRYLFSQHF			
				AGACGACCGAC		NSDCKSPWIDCIQRILNM			
				AGGCTTTTAAA		CGLQYVWQNHNFVNIE			
				CGACAACGTCA		WLSATVNQRLKDQFMQ			
				AGGATCAACCC		IWSNDINCSSRGQVYKIL			
				CCGACCGAATT		KPYFGFEKYLDILPVKLR			

				GGATAGGGAGA CGCAGGAGAAC AACTCTCGCGA ACAAACAGCTA CACGTCTCTGA AATTTCTCTCGC TTAACGTATGC GGTGTCAAACA AAAACTAAATT ACCCCGAATTTT GTGATTTTATTG CAAACTATGAT ATCATTGGACT ATCTGAAACAA AAACTGATGAT TCGGACAATAT TCAAATACCCG GTTACGTCACTT AC		KKFIKFRTSNHRFPVETG RWYNIPLNERLCLLCNK GLIGDEFHYILECSALEEI RKKYINTKYWKRPNFFK FSELMTNCNCKSLRKLC VFISKIFDAVCSS			
CR1	CR1-6_CGi	.	Crassostrea gigas	CCAAATTCCAA CAACAGCACAG ACAGTATTATTA AAGATATTTGC	TGATCTTATAACTA CTAGTACACCGTTG TTTACTATATACAT GTCAATTGCATATA	MAGDTIEQYRAAIGLFY CCVRCVKKCFYLNLFQL RXLFKCLLQYLLWFGNL YKTLSMCLAQHSLVIDY	cd01650	595	851

				AGAGAAAAACT	TGTATACGTCCATT	FFQVIFMLLFLSGDIETN			
				GGGGATCCAAA	CTATCCACTGACCT	PGPCTGTQYTLDIFHLNI			
				TAACCGAAAAC	TTATTTATGTTGTGT	RSIRNKLDSLLSLVSDFD			
				GATATTTTTCGT	ACACATATTTTGTA	VLCFTESHLNSSIPDRNL			
				TCCCATCCCATT	TTATATCTTTGTATT	FIDGFNTIFRKDRNSFGG			
				GGAAACAGAAA	GGGAGAGGGCGGT	GIIVYVSNSLRVKRRTDL			
				CAACCATGGGG	CTAAGTTATGGAAC	EPTNIECIMGNQQTLNAL			
				GCATCCAAATT	TTGTGCCCAATCCC	WVEISDPTCNILLCCTYR			
				ATTTGTCGTTTC	AATTGTATTTTTGT	PPNCDNSFWRNLSWSIE			
				CAAAATTGGAA	ACAATAAAAATATG	KAYDESDKIIVVGDLNV			
				AACAAAGAACG	TTCAAATCAA	DFLNIPHTHLVKDLLSSH			
				CAGTGTACAAA		FLENTITEPTRVTQNTST			
				GCAAAGAGTAA		LIDPILVTRDVSFFDSGV			
				ACTAAAAAATT		LDVHHTISDHKATYIYV			
				CCGATATCTTTG		KSNLNLNSAYKRKVWL			
				TGACCGAGGAC		YKNADFDKLNFLITHTD			
				CTCACCAGATA		WETLIMGADDVDQAAX			
				CCGACAAGGAA		NFSETFLSHIRKSVPEKT			
				TCATACATGAA		VTIRPKDKPWFDSALRK			
				CTGTCAGCTGCT		EIRQRNRLRKIAMNSRN			
				AAACGCTTGCG		ALDMQKFKKSRNKVNN			
				TAAAGTTCACTC		MKKYAIDNYYNNLEINL			
				TTTCTGGACCAA		LDSSKNNSKLYWRLLKN			

				CGACGGGAGAA		VFDIKMSTEIPPLQFALK			
				TTTTCACTAAAC		NGNQSIAYTNTEKAEVL			
				TGTCCGAAAAT		NDYFSSISCLDDNDARLP			
				GGTCAAAAATA		EFNSECNNFICDVCISEQ			
				TTTGATCAAATC		DVIDIISIIQVNKAVGPDN			
				CATCAGTGAAC		ISHKMLKNTMYTIAKPL			
				TGCACGAAATC		CSLFNRSLRDCTFPSSWK			
				GCGCCGCTACC		ISNVLPLYKKGDSSQMS			
				GAGTGACGAAC		NYRPVSLLSCVSKIMERI			
				AACAAGGCTAA		IFRHVYNFFHDNNLFYK			
				AACGCAAATCC		YQAGFLPGHSTVYQLIE			
				CGTACTCTCTTC		TYDHILRAIDQGQSCCMI			
				TGAACCACAAA		FCDLSKAFDRVWHKGL			
				CCCATGTGAGC		LFKLHTYGITGNLFKWF			
				AATGTATAGTA		NSYLVNRKQRVMYREV			
				TAAATTAGATA		LSSIKSINAGVPQGSVLG			
				CAATGTTCAATC		PLLFLIYVNDIADNMLTF			
				AGTTATGTGTG		CRLFADDNSLQHAAINV			
				ATATTCTAAGTA		HDMESYLNRDLLTLEV			
				TCGTGTCTTCAA		WSKKWLLNFNPSKTKA			
				AAGTAGATACA		VYFSNKLTQTNTVTPTL			
				TGTATTTGCAGA		KFQGCDLEFVPTHKHLG			
				TGTATGAAAAT		VIFSEDLSWSSYINSIIAN			

				AATTTTTAGTTA		AQKKLGLMKKLKFKLN			
				TATGTACACCA		RKSLSLLYTSFIRPQLEY			
				ATCTAATAGTTG		ASDVWGGCSINDSDKLE			
				TGGTTAAATTAC		KIQLIAARIVTGLPIFASR			
				ATAGCTACATTT		ESLYFETGWETLYSRRKI			
				TTGTTTGTTATA		SRLKTMYKIDRNLLPSY			
				TAGTAGTATTAT		VMDIFPNKRANVSNYST			
				AACTGTGTAAA		RNAHNYXIPKCRLELYK			
				CCTATGAATTTA		SSFVPTVVDEWNALSTD			
				TATTTCCCATTT		VRGSDSVRIFKEKIHVSV			
				TTTCTCAAATGC		KNENXLTCRTRPNFFTY			
				AATCAATTTGGT		GDRFLNIIHTKLRHNCAL			
				TTTTACCAATAT		NNDLFRCNIINSPLCSCG			
				TGTACTTGATGC		KJEDTYHYFFTCNKYKD			
				CATAATGTTTTT		ARNVLFNDVFQIVNLNI			
				TTTTTCTACATT		VNTHVLLWGDSSISDNE			
				TTCTTGTTTTTG		NKHLFTLVYNFIKNTERF			
				TTCATTTGTTCC		N			
				TTTCTTGTTTTT					
				CACTTTGCATTC					
				TTTATTTGTTTG					
				CATACACTTTTT					
				CAAGTGTTTGCT					

				TACTGTGTTTTT ACCTTCAACTTT TGTCTTGATGTT ACGCAACGAAG GTCATGGTCGTT CAGTACATGTA TGAATTCACAC CTGTCTTTTCAA TCGCAGTGCTCT CTCATAGTTTCA ATAACGTTAAT CAACATGTCTTT CA				
L2B	L2B-1B_CQ	.	Culex quinquefasciatus	G	TAGAAGTGACCCAC GATGGTACTGTGAC GAAGAGCACGTTAT GACGGTCGGCCATC TTCATTATCGGTAC ACATCGCATGGGAT CACTTTGGGCCGCA TATGATAAATTTAT TCAAAAGTAACGCG AATCTGGGCGCGGT	CESALNLLLLKWKQCIE KGNIVLSVFVDLKRAFE TIDRTKLIRVLKKSGIRG TVLKWFSSYLENRKQVT RYNDAVSAETAVDLGV PQGSVLGPLLFILYINDIK QALKRVQVNLFADDTV LFVVGDSFDECFDIMNE ELGGFSEWLKWKKLQL NITKTKCMLVTTRQTKD	cd01650	1	195

					TTAACCCTATGGGC	CRRSVQMDGGIVERVET			
					TCATATGTGCGTGG	MKYLGVMLDEKLNFNE			
					AATAGCAAATGGTT	HIDYTIRKAARKFGVLC			
					CCAATACCCTAAAT	RINRYLTAETKVQLYNS			
					GGATAAAAATACTG	LIAPHFDYCSSILFLATQ			
					CAATCTGATTGATA	RQMKRMQVLQSKIMRLI			
					AAATGACTCGACAC	LRCDRLTPRRSMLECLQ			
					TGGATTTGTTAGAG	WMSVKQRIEYNTLVFVF			
					CGCCTTGAGATAGG	RVKERMAPQYLTEAMV			
					ACACGAGAGAAAT	RGRDIHEHDTRGADDLR			
					GGATGGGCATACAC	LQIWRKACTQNSLFYKG			
					GGAAGTAGAGTCA	FKLYNQLPEAAKMTSNI			
					AGATTACAGGACAC	NEFKRSCKEFVRQRPLE			
					TCTCAAAACACGAA				
					CGACGAATTATCGA				
					AAGATATCTGCTCG				
					TAAACCTTCCATGC				
					TACAAAAACTGTGT				
					ATGGGTAAGAGGT				
					GGGCCATCCAAGG				
					AAAAAAAAAA				

CR1	CR1-10_CQ	.	Culex quinquefasciatus	TC	TAAATTTAGTTTTAAGTKCATCATTGGGGCAAAATAAGTGCCTGTTGAGAAGAAATAAACAAC	RRADFESLSRALRGIEWSSVLDPVDIDLAVDTYTXILGRLIEQFVPKIRKVVSHHLPWQTQELRXLKTQKRAAFKVSNGSVHLRDYYLRINHSYQRLSRRCFANYQRKMQKKLKTDPKKFWKFVNESRRESGLPTSMRLADXEGDDADKICQLFAKKFAGVFCAEPITDEEVRAAAENVPRRDELLASIEIDEGTIQAAAAKLKHSCSPGPDGIPSTLLKRCSSSLSAPLLHLFRLSLSSGKFPVAWKQAYMFPVHKKGDRTNVENYRGISALCATSKLLELVVIDPVFSHCRQDLADEQHGFFPKRSTATNLLCFTESIINSFDAHSQTDAIYTDLSAAFDKINHRIAVAKLERLGFRGSLLAWFKSYLTGRSLRVRIGDALSDIIDATSGIAQGS	cd01650	250	500

						HLGPVVFLFYFNDVNYT VTGPRLVYADDLKLFAR IDNSNDAEALQRDLDRF AGWCXVNRMILNPGKC QXITFSRRHFPVTFDYRI NDTSVERVTHVKDLGVI LDTQLTFKQHLSYIIAKA SRQLGLVIRMTRKFTDIY CLKTLYCSLVRSTLEYCS TVWSPHYNNAVCRLESI QRRFVRYALRQLPWTNP LXLPPYEQRCRLIDLDTL QLRRDLARAMTSADVL LGRIDCPDLRNQIQLRVP TRQLRHTPALEVPFRRT NYSANGAINGLKRAFNK VSSVFSLDVSRTTLRFKF VSLLRSLLRM			
CR1	CR1-25_CQ	.	Culex quinquefasciatus	CACGTGGCCTCGCGGTCTTGTGATCCGAGAATTTGA	TAATGTTAGTTTTTTAAGATTATACAGCCTGAGCATTTTAGATGGAGGCGGTGAAATAAATAAATA	RALRRIFLGADKRXQPDSAXSASSSCPTCTNKTVNAESHAKVNAETNTKVNVEVNPEIYAGHGADPKGYGTNFACESQWFTYGS	cd01650	933	1184

						RDRLTNTLRSTPSLPQQS SVQHLPASSPGRAVESIM RAPRLPDATAPSPSYRTR SPASRSHQRRPGAGVGV GGGVSQTAPLGKYPSQL QVSLPDACPTYSSSLITSS STRTANASPPAQAIQSSP GRTVESIMRAPRLPDAT TSSFSYRTRSPAGCSHQS RPGAGVGVGGGVSQPV LFGKYPFQSQRVLPDAS PTHRSCPSTSTSTRTANA SPLPAQAIQSSPGRTVESI MRAPRLPDATTSSFSYR TRSPAGCSHQSRPGAGV GVGGGVSQPVLFGKYQF STPHTSPDARSDFSATVP PSTFVTLAAPGCVYHGT MEDLGSANSGMPAALT ASDPASASSERGFLNVY YQNVRGLRTKTNSVLLA LSSCDYDVIIFTETWLVP GIKNSEFAAGYDIFRCDR			

						SAETSSLQRGGGVLVAV			
						KTGRNCKDVELVGCGSL			
						EQIAVRIVLPHQALFICCI			
						YIRPASNIDVYQLHTGA			
						VQQILGMAGVRDLVLV			
						VGDYNLPALTWFFDDD			
						THSFLPINASAEQEVLLL			
						ESMLSCGLLQINNLASA			
						HGNLLDLAFVSDATAAE			
						LIDPPCPLLKPDAHHKPI			
						LLKLDYDHGGNDDLDD			
						ADEIFDFNQCDYALLNE			
						RISAVDWDTLLDTSSVD			
						ESTANFYERLNSILHEVV			
						PKKIRRMSHRYKLPWW			
						NDELRTLRNRLRKATSR			
						YMESKTAWDKVTVQNI			
						EDEYNALNQESFRSYIN			
						NTQRSLKSDPSKFWSYV			
						NSRKNANRTPTDVSYRD			
						LNSTSPAIAANFFADFFK			
						SVHSSDAIPPRDHEFEYV			
						PTHYLPLPLPVLNDEETL			

						KALSGVDGAKGPGPDGI PPSLIKACADSLTVPVKR IFNDSLRKGVYPAMWK VASITPIHKSGSTKKVEN YRPISILNCLAKVLEKIV YDRLYAVVRPIISDDQH GFMQKRSTTTNLLTFVH STVNALESGEQVDAIYID FEKAFDKVPHALTVRKL RKLGLPAWITNWLLSYL TNRKAFVNLRIARSSTFD IPSGVPQGSHLGPLIFILF VNEICALTNSNTLMYAD DLKLFRSVKNNVDCCSL QMDVNALLHWCESNG MKVHVKKCKVISFSRSR NPLRFDYTMGDGCPLDR VNSIRDLGVTVDRKMKF NEHIALTTGKAFAMLGF LKRNTAHFDDPFALKVL YASLVRSVLEYAAVVW APYHVTLAARIERVQRA FVRFALRKLPWNNPLILP			

						SYEGRCLLFRLQSLAER RTLMQRLFVFDLLRGNI DCRSIRNNVRFNEPARQ GLRGERPLLWIPRHIFDY GYNNPLDVCFRQFNEIC SCFDYVVSKTVFKTRIS			
CR1	CR1-9_CQ	.	Culex quinquefasciatus	C	TAATCTGTTTTTGT GTTTCATGTTGTTA GCATTTGTCTTTAG TTAAGTTAAGTTTT ATCATTAAGACCAC AAGTTGTCGGATGA TTCAAATAAACAAA TAAACAAATAAAC AAAAAAAAAA	GESESRGREHGRVTRPP LAASPGRAVESHVRVPR LPDTTTSSPSHRLIRPAC CSLQSRPGAGVGAGGG VSQTATLGKYXSXSQXP LPDARPTSSSCAISSTSPP VRTPNAVRAFWSSPGRA VESHVRVPHLPDTTTSSP SHRLISPACCSLQSRPGA GVGAGGGVSQTATLGK YHSSLQNPLPDARPTSSS CATSPPVRTPNAVRAFR SSPGRAVESHVRVPRLP DTTTSSPSHRLISSACCSL QSRPGAGVGAGGGVSQ TATLGKYHSSLQNPLPD ARPTSSSCATSPPVRTPN	cd01650	910	1157

						AVRAFWSSPGRAVESHV RVPRLPDTTTSSPSHRPIS PXCCSLQSRPGAGVGAG GGVSQTATLGKYPSGLQ NPLSDARPTSSSAAFPSTI NAKGTSTPCKLRIYYQN VRGLRTKIDAFFLAVTE CDYDVIVLTETWLDIEIL SPQLFGSTYRVFRNDRN ERNSDKKSGGGVLIAISS RLDCVLDPTPVCDTLEQ LWVRLTVGASQRVCIGA FYLPPGLNRSIDLIKDHL ESIGNVLSHLNLNDSFIQ FGDYNQPGLKWIPSEEG GLEIDPVQSRVPLVSSYL IDGFNLHGLSQLNPCVN VYNNCLDLVLVNEVALP NCAVLAAVDPLLPQDIA HPALEMFVDVPAPSATA EPPLNSQQSFNFRKANY AAIQEELSRIDWSFLETA PSIDAAVDAFQSILNNLF			

						TEHVPLRRPPQKPPWSN PTLRKLNRLKTAALQEF SNHRSPFFKSKFNLASKR YRGYNRTLYRRYVIRTQ RDLRRNPKNFWSFVNSK RKADGLPPTLYLDDVAA ATSEEKCALFAAHFKDA FNSRFASEAEIIEATRNTP INAIECSLTEVANCQVSA AINKLKLSYSPGPDGIPS SVIKRCSGALLSPLTCLF NLSLRQSTFPDKWKASF MFPVHKKGDKSSVKNY RGITSLSACSKVFEILVN NVLFDSCKNYISTDQHG FYPKRSVVTNLAQFVSH CLQEMDAGAQVDAVYT DLKAAFDRVDHGVLLK RLERLGVSASLVRWLKS YLTNRILRVKIGNSESDM FISLSGVPQGSNLGPLLFI LFANELSMLLSEGCRLF YADDVKIFTVVRSVLNC			

						SALQRQLNHFNEWCVR NFLTISIEKCSTISFHRKL KPVVFDYAIAGNTLARV TQVRDLGVTLDSELCFTI HLHDIVSRANRQLGFIFR IADDFRDVACLRSLYCA LVRSILEFCSVVWCPYQS TWTAKIESVQRKFTRLA LRRARNPAGWNTYEER CRVLQLETLERRRHVSQ ATFVAKVLAGEVDTPWI RAQIQIYTPARPLRQRQY LQLAHRDTNYGQHEPV RFMCSRFNVFSHLYDPNI STPQFQQRARLWYSHQL			
CR1	CR1-5B_CQ	.	Culex quinquefasciatus	ATACTACCAGAACGTCGGAGGC	TAATTTATGTATTAGTTTGTAAGGTAAAACACCATTGGGGCCCGAGCCTGTTGGTGTCCAAATAAATAAATA	MNTRALEYRLACSAISY DVIAVTETWLKNNTLSS QVFGNGYEVFRCDREQF INSQKKDGGGVAIAVRQ GLNARIVSSDEWKCVEQ VWVAIDLADHTVYICVV YIPPDQTNNASVFEINSS SVSTVSSMARPYDELVIL	cd01650	499	750

						GDFNLPRLIWRPSRDGFL YADPELSTQTLHVTEFL DNYSSALLQQINQFPNE NNVLLDLCFVSCPDIAPP IAIAPAPLVKDVRHHPPL HLSLPVHLASEFTPIVCE VRYDYRKSDVPSMLELL RSIDWENTLDKDNLDDA VQTFTNIIGYVIDRHVPK KPVTNARVPWQTSELRR VKAAKRSALKKYSKHR TLPLHNYYVRLNSCYKR MSKQCHADHEVRTQKK LKSNPKGFWKYINEQRK EIGLPSSMFLGDKTASNT EDICQLFAAKFSSMFTQE DLSQQQITAAANNVLPN GLALNRITIDRSQILEAA TKLKSSNTAGPDGVPSIV LKKCIDGLLDPLAHLFSL SLSTGSFPKLWKHAYMF PVHKKGDKRNADNYRG ISALCATSKLFELAVMDP			

						VFFFCKNLISDDQHGFM PSRSTTTNLLTFTTFVTD SFTAKSQTDAVYTDLSA AFDKLNHSIAIAKLERM GIGGVLLRWFQSYLTGR KISVKIGDWLSAVFEAFS GIAQGSHLGPLVFLLYY NDCIHSISVPRLSYADDM KIYCRIDNASDAQVLQQ QLINFAEWCKINRMVVN PTKCSVISFSRKRNPIKFE YQICGTSIPRENCVKDLG VLLDTELTYKQHISYIVS KASRQLGFVFRTAKNFT DIYCLKALYCALVRSTL EYCCSVWSPYYENSVAR IESVQRRFIRYALRSLPW RDPFRLPSYESRCELINL DTLAVRRNVCRSLLVAD VLTSRVQCPAILRGLNIF APIRTFRNSPFLRVPIRQT NYAMYSALTGLHRAFN			

						FVASLFDFNLSRDVIKRK FLLFFRR			
RTE	RTE-1_CQ	.	Culex quinquefasciatus	TAGGGCTCCTTAGGCTAGCCACTT	TAAGTAAGTAAGT	MEPESGLRATETARRKT DWRNGTTLGMKTRTRI GTWNVLTLAQAGKLET LGREAVRLKLEILGLSEV RWPDSAEHKMPTGQVL LYSGLRSENAQRGFLLSP NARSALITWKPINERIIV ARFRTRVRNLTFVQVYA PTDAADLQEKEEFYSQL SXVVNEIPKGDIRIYAGD FNAKIGSDNTDLERIMGP HGXGEMSENGELFTEFC GNQDMVIGGSLFPHRSV HKVTWVSRDGKTENQI DHICISRKWRRSLLDVR NMRSADIASDHHLVVGE IRLRVARVVRQEEKVGC RFNVQRLSNPEVXRAFV GELQARALEIPDSTVEEE WQFIKDAFAVTGENTLG MLRTQRKEWISDATWD	cd01650	554	813

						KIEERKQAKAAIVSART RATKARTRQTYAELEKA VKRSCRRDKRAWTDSL ADEGEKAANNGDMRLL YDISRRLCGARTNTRIPV KDRSGQLITDPADQLKR WFEHFEQLLQISHPRPSP QYHPQNVRRINRVNSSP PTLREIEEAVKGMKSGK APGIDRISAEMLKADAH LSARMLHRLFSKIWDTA TFPVDWMQGILVKVPK KGDLTNCDNWRGITLLC IALKVLCKVILCRIKSKID ESLRRQQAGFRSGRSCV DHIVTLRVILEQVNEFQD SLHLVFVDYEKAFDRLD HENLWEALGRKGVPERI INLIKAQYEAFVCRVVH NGVLSDPIRVVAGVRQG CILSPLLFLIVIDEVLAGA LDRVQNRGLLWQPIRME HLNDLDFADDIVLFSQR			

						GNDMQNKLDDLAACSS AAGLKVNVSKTKSMAV NTERAQGFTVAGENVEK VDTFQYLGSQIAPDGGT KLDISTRIKKARSAFAGL RTLWRSNQISLRTKLRIF NSNVKSVLLYGSETWCV SSQNTSKLQVFVNRCLR NILRAWWPRNWVSNAE LHRRCQQRPIHVEIRERK WRWIGHTLRKDAHEICR EALDWNPQGQRGRGRP RGSWRRSLHQEISAVDG TLSWRQVKAKAENRQQ WKSFNAALCSTGEQXA			
Penelope	Penelope-4_CQ	.	Culex quinquefasciatus	GAATTGTGGTTTGGTTGAGCGTTTTAGCAGAATGCATTACTGTGAATACAAAGAGACGAGTTGTTCCTTTTAATTCCGCTATATATTTTAATAT	TGATGAAGTCTGCAAGTAGTAGACGAAATACGTATCTGTCAAGATATTAAAAATATATAGCGGAATTAAAAGGAACAACTCGTCTCTTTGTATTCACA	MHSAKTLNQTTITMATKPSEHNKFIGLKKIIAGIHKDIAFLKKCVEHRVTPTSHKIRTKGHTDSKIIRRVELELVKESIKKLYGRLSIKTLECYSLHLKLAKEFPIEFEAFLTKVKVAEKCESERKRKILDNKFRQLISHIA	cd00304	469	548

				CTTGACAGATA		PTMKPVVKPLEEFVVNR			
				CGTATTTCGTCT		SSETFSEEQLTLLNKGLK			
				ACTACTTGCAG		YAVSSKPNLENIVIDIET			
				ACTTCATCAGTG		VINANSIDEKQTKPPPLL			
				TTCTGTTCTCGA		PAQINSVRMTVSKLIREA			
				CTGAAGGTTCA		TTKHQDNAEVRVVKEL			
				TCGCTGGTGTAT		KEKPVYYLKADKGNRV			
				CCGTATTTGTAG		VIMDRDEYDKELLEKLG			
				TTACTGTAGGTA		NREKYSLQRGYTLIDMV			
				CTACCTCCTCGT		KKVENTIKECAGLLKSV			
				TCTTCTTTTGTG		GKTAKSLRVPNPTLPRIK			
				CCTGTATAGGT		ALPKVHKPGNEMREIISA			
				AACAGCTTAAA		VDAPTSRIAKWLVEEFK			
				CAGCCACGTTG		SMPKPFPSRSIISSQVFTK			
				AGCCGTTACCT		ELLESGPIAEDEIMVSFD			
				GAATCCTTGTTG		VTALFPSIPVKNAIGTLR			
				AGTAAACGTTT		DWLQSQYEGEPWRQKTI			
				GTTGCCTGCCTT		QYITFVKLCMEQSYFQF			
				GAAGATTTCCA		RDCIYQQRMGASMGNP			
				AACTTTCGGCG		LSPFMSDVYMAALEHEL			
				ACAGCCAGCTT		QTKNLLPERWWRYVDD			
				CGATGGGTTAG		VFCIIKRDSLTTVLDTINS			
				TGATGTGTCTTA		ARRSIKFTYEMEVEGKL			

				AGATGACCGCG		PFLDILILREPGCPSSFSFE			
				TCGCTAGCTGTG		IYRKPTNTRRTIPATSNH			
				ATGTTATGTTTT		SFQHKMAAYHYFIHRM			
				TCCTCGATGATG		TTLPLSEAGKQKELEYIF			
				TGTTCGGTTACT		ETAEINGYSRTTIQAIIDK			
				GCTGACTTGAA		KERKLHRTSLTTLTPSAE			
				ATGGGGGACAA		PLRRAAVEFDYRITRPLR			
				ATCCTTTCCGTT		QKLAKFGIDLVFSSRNSQ			
				TTTCCTCCATGG		LESILGSTKDPIPTLGKPG			
				CAATTTTGACAT		IYEVSCGHCDMSYIGQT			
				ATGTGTGTATAT		KRLLQTRLDEHIHTYVKI			
				GCTCATCCAAC		AMEEKRKGFVPHFKSAV			
				CTGGTTTGCAGC		TEHIIEEKHNITASDAVIL			
				AATCTCTTAGTT		RHITNPSKLAVAESLEIF			
				TGTCCAATGTA		KAGNKRLLNKDSGNGS			
				GCTCATATCAC		TWLFKLLPIQAQKKNEE			
				AGTGGCCGCAG		VVPTVTTNTDTPAMNLQ			
				GATACCTCGTA		SRTEH			
				GATGCCGGGTT					
				TGCCTAAAGTC					
				GGTATGGGGTC					
				TTTCGTTGAACC					
				TAAGATGGATT					

				CCAACTGGCTG					
				TTTCTACTGCTG					
				AACACTAAATC					
				GATGCCAAATT					
				TAGCCAGTTTTT					
				GTCGTAATGGG					
				CGTGTTATCCTG					
				TAGTCGAATTCT					
				ACTGCTGCTCTT					
				CGCAGTGGCTC					
				TGCTGATGGCG					
				TAAGGGTTGTG					
				AGTGAAGTTCT					
				GTGCAGTTTTCT					
				TTCCTTCTTATC					
				GATGATTGCTTG					
				GATAGTTGTTCT					
				GCTGTAACCGTT					
				GATTTCCGCTGT					
				TTCAAATATGTA					
				CTCCAATTCTTT					
				CTGTTTTCCTGC					
				TTCGCTGAGGG					

				GTAAAGTTGTC					
				ATCCTATGTATA					
				AAATAATGATA					
				TGCTGCCATCTT					
				ATGTTGGAATG					
				AATGGTTTGAC					
				GTGGCTGGGAT					
				TGTTCTTCTGGT					
				GTTTGTTGGCTT					
				CCTGTAGATCTC					
				GAAAGAAAACG					
				AAGATGGGCAA					
				CCTGGTTCTCTT					
				AGGATCAGGAT					
				GTCTAAGAAAG					
				GTAACTTTCCTT					
				CGACTTCCATTT					
				CGTAAGTGAAC					
				TTGATGCTCCTG					
				CGTGCACTGTTT					
				ATGGTGTCCAG					
				TACCGTTGTTAG					
				TGAGTCCCTTTT					

				GATGATGCAGA AGACGTCATCT ACGTATCTCCAC CAACGTTCTGGT AGTAGGTTTTTG GTTTGTAGTTCG TGTTCCAAAGCT GCCATATAAAC ATCGCTCATGA ACGGTGACAGT GGGTTACCCAT CGATGCTCCCAT CCTCTGTTGGTA GATGCAATCCC TGAACTGGAAG TAGCTCTGTTCC ATGCATAATTTG ACGAATGTTAT ATACTGGATCG TTTTTTGTCTCC ATGGTTCTCCCT CGTATTGTGATT GCAGCCAATCC					

				CGTAGAGTTCCT ATTGCGTTTTTT ACTGGTATGCTT GGGAACAGTGC TGTTACGTCGA AGGAAACCATT ATTTCATCCTCA GCTATCGGTCCC GACTCTAAGAG CTCCTTTGTGAA CACCTGCGAAC TGATGATCGAA CGGCTCGGGAT CACTAACCCAT CGAAGCTGGCT GTCGCCGAAAG TTTGGAAATCTT CAAGGCAGGCA ACAAACGTTTA CTCAACAAGGA TTCAGGTAACG GCTCAACGTGG CTGTTTAAGCTG					

				TTACCTATACAG GCACAAAAGAA GAACGAGGAGG TAGTACCTACA GTAACTACAAA TACGGATACAC CAGCGATGAAC CTTCAGTCGAG AACAGAACACT GATGAAGTCTG CAAGTAGTAGA CGAAATACGTA TCTGTCAAGAT ATTAAAATATA TAGCGGAATTA AAAGGAACAAC TCGTCTCTTTGT ATTCACAGTA					
CR1	CR1-27_CQ	.	Culex quinquefasciatus	GGATACCATCG TAGTTGTAGGC GATTACAATCT GCCTCACCTCA ACTGGTCTTTCG	TAAGTTGTAGTCTG TGCGGCGAGTGCCG AAGATGAAATAAA TAAATAAATAAAA	ARNCFRRVYACNXDTA NLQPAELQPKTPGPAVT PVPESTESLTYDFTRCDF EAVNJALSQINWTGILAH CSLDEAVSRFYDALXDV	cd01650	297	547

				ATTCGGACATC AACGGATACCT GCCAACAAACG CATCTGCTGA		IQLHTPLRRLHPPSNRKP WWNATLRNLRNRLRKA RTRLHKRRCQPNIAEVR ELEELYESQQNAAFSEYI HRIETDVKQNPRSFWTF LKSRKKGSGIPTEISNGC VEANSVEEQASLFADFF SGVYSTSLPAPCNEVVD HIPTHDLHLPVVHFSIED VSRALDDVDVTKGPGPD SIHPSFLKNCAASLALPT TIIFNRSLAESTFPTAWK LASITPIHKAGNVHKVE NYRGISILSGFAKLFEHL VHGALYPVLKPIIANEQ HGFVKRRSTTSNLLLFTS TLLTNIEKGQQIDAVYV DLSKAFDKVPHALLLEK LRRYGLPEWIVRWIHSY LADRKAFVKVRCTSSHL LAIPSGVPQGSVIGPLLFI LFVNDLCSVLESDKLMY ADDLKVFRSVSSPLDSC			

						ALQQDIDRLLRWCTANG MEVNVQKCKIISFTRKR TPLLSSYRMGQNELERV NTIKDLGVTIDSKACFNE HIALTTAKAFATLGFLRR NAADFVDLFALKTIYCS LVRSQLEYAVQVWAPY YAVQAERIERIQRAFTRF AVRRLPWVRPQGLSSYD DNCERLKLPALASRRVL LQRIYVFDLLSCNIDCPT LRDKITLHIPARNLRNPA PFLEVPGHRTNYGYYSP LSACCRVFNDVADVFVF GMSKCVFKTRIKNRT			
CR1	CR1-23_CQ	.	Culex quinquefasciatus	A	TGACAAAGGTTTAG ACCGGTTTATTTTT AGTCAGTTTTATTT ATTTATTTATTTATT TTTTACATTGTGCT AGTTTAGTTTGWTA GCTTAAGTTTGATT GTATATATCCTCTA	ALKKLKSSKRAALKKFS KCPISVNRLNLNQISNDY RRLNKSLFLAHQSRVQN RLKQNPKGFWKFVNEQ RKESGLPSTMIRDGVEV STSSEICDAFRVQFSSVF ADNNLSDPDVAAAADS VPDSGLVCSDLQVSRED	cd01650	190	441

					TGTCATTTGGATTG TAAAATCTGTTGAT AATAAAAAAGAGG TTTTGTGCCTATTTG AGAAGGACCCATC GGTAGGCGTTCCAC TCAAACGGGCTTTT CCCTCATATAAAAC ATAAAACA	FNAACSKLKSSTSPGPD GVPAIVIKKCANSLSEPL LRIFNLSLSLGVFPNNWK ESYVFPVFKKGCKQDVK NYRGIAALCAVSKLFEVI VLEFLRHNLSHVVSDDQ HGFFPKRSTNTNLVTYIS KIQRAISDGYQVDAIYTD LSAAFDKINHRIAIAKLR RVGLHGSLLTWVNSYLT GRTMLVKISDNLSLPFSV TSGVPQGSHLGPFLFLLY INDVNFLLRCPKLCYAD DFKLYAIIKRPTDMEDL QAQVDLFADWCRLNQM VLNPSKCSTITFTRKRCP TFFSYKILGNPLTRDSFV KDLGVLLDSKLSFKNHIS YLCSKASKQLGMIFRMT KYFRDVNCLKILYLSLV RSTLEYGSVVWAPYYH NDIGRIEAIQRKFTRFAL RHLGECPNYETRCSMLH			

						LDPLSVRREAAKAFFVA DLFRSNINCPFLTSELNV NIRRRVLRSHSFLVIPGA RTNYAFNEPVSSMCRVF ERCYPFFDFHLSRPRLKH DIXNFLREAYVF			
Kiri	Kiri-5_CQ	.	Culex quinquefasciatus	TTGTTTGGATTT AGATGATAGTA GATTTGCCCAG CAGTACAGGAT TTTTTGTTGTTA TTGTTATTTTGT TTGTTTAAGSAG TTTATCTGAAAA CCAAAACAGCA CTATTATCTTTT GTTTTGCCCTGA TTGCTCTTTTCC TACGCGTGTTTT TCTCTACATAGA TTTATATATTTT T	TGAGTTTTGTTTTG GTTTTTGATTAAGC AATGCCTTGTATTA CACACTCTTTTATA ACTACATACACAGC TGATAATCAGTCTT AAATTTATTACTGA ATTCTACTTAAACA AGTAATTTGTCGTA TTATCGACATTTTT ATATCTTTAGCTTA TTTCGTTTTAGTTTT TAATTTAGTTGTTA GTTTTTAGTTTTAAT GAAACATAAGGTCC TTATCATTTAAGTC AATGTATACAAAAA	MTNTVNNVQELRFEPIN ATLKNNLSKENGFKVGH LNTCSIICHHDDVKNIFT RTGFDCIGFTETWLNSDI KDALIKFDNFEILRHDRN STTRVXGGGVALFVRPE YKTKIIKTSPRDSIEYMF CELKVVSCKLWLVLSIN PQPIKSSQLQVVVGVIYK PPTLXDFSFFFEDLGNISS TYNNIILMGDFNINILVS NPMSNNFTRLINSFSLSI VNKSFPTHFHNTGSSLL DYFICSENLXSTLFSQVS VPAVSHHDLIYTVFNLP VPVASLKTISFRNFNRID KEAVRNDFPNFNFDQLN	cd01650	519	778

					AGGTTGTACCTTCG GTATATTGTTGAAA AATAAAATACAAAT ACAAATACAAATAC AAT	GIVDCETLASNFNHKLIQ LVNKHAPLKTKTILPTSC PWYSREAYNACVDRDL AYRFWKNNRTLENKLIY NRFRNRATNLIRIAKKQ YSAKLFDPTLPSKTIWK NFKSLGFRDDPDNLDDK FKPEDVNNCFIQKIENYP TGNRQPTIIPDIVEEDQF KFRAVTEDEVHKAILRIK SDAIGYDEIPLKFVKLFL DLLVPYITLLYNKIISTNT FPTIYKLAIVKPIPKIKNP SNPNDLRPISILPSLSKGC EILLKDQIVQYVTFKKLI TPFQSGFREHHSTSTIITK VCQDINASFDSNYVTVLI LLDFTKAFDLIDHRIFID KLKEQFQFSSDACDLMK SYLSDRSQLVQINGVRS QFRSTKRGVPQGSVFGP LAYTLSVNDLPSFITHCN KQSFADDTQIYLSCPVD			

						QLPNKISLINSDLSAISRY CKDNGLVLNTSKTKAIX FSRQNTVLPHLPPIQIDN ENVEIVDECLNLGIKMN KHLTWXDHVSLIKKKV YSTLRTLNCHRHTLSIDT KIKLVHTLLMPHFIYGN VIFSKMNVDTERHLQVC FNSCIRFIYNLRRFDHVS QFQRSIFDLSLSKFYQFQ LLLFLFKLINFKEPQYLF ESLVFSRSNRTNNIIIPVH HTNFMGNSFAVRGARL WNDLPHNLKSITLFSEFK RRCKEHLKNSI			
Tx1	Tx1-11_CQ	.	Culex quinquefasciatus	TG	TAGAAACATCCAAA GTTAAATTATCTCT ACCGAGCTTATTTG TCGTCTTATTAGCA AGTGGTGAAATGTA AATTCAAGAAAAAT GTTGTAAAATAAAG TGAAAAAAAAA	LYENIFGCLIANNFVKVV AYADDLNILVRNNHEFD TVLQLVSYFSIYSKIKLN FAKSQYLRFNGCLSGPH QIKEVESLNVLGVLICQN FNKLVEINYDSYLNKLK YTLSLHQKRILNLFQKA WLLNTYVLSKIWYICQIF	IPR000477	11	83

						PPNNKHLAAIKSICGQFI WQGFIFKVPRNELYLPVI KGGLALTDVESKAKSLF IKNILYSNADGNAPLDNF MLGQINNSTLTRNMRE WIVKADELKLESHLNSS KKIYDFFIHSLNIKCKTM EDMPNLMWSNLLNNLD SNFLNSKCKTVLFCIFKD IVLNKKKLHLYGIRGTD SKFCENCGLVESVSHLV KYCQTSEEIWRWLKNIL NKRLKILVKDPDELLSLS IGDRDFKYKAALWLTV QTIVYNLEKRGFGDLNE FKQGIRIVRFNNKAIFER HFKHFLNIF			
CR1	CR1-19_CQ	.	Culex quinquefasciatus	A	TAGGAAACAGTCTG TTCAAGACGAAGAT AAATAAATAAATA AATA	IXRTQRRXFLGXFKCCR EHSGKSATPPVPTLNDA DDAEFAARVCTAIDSKD TGACFGVDWRSGRFTH GSSDRITCGTIVPTIPASF RSSSQHLPVRVLQSSPGR	cd01650	1011	1261

						AVESLMRAPRLPDATTS SPRQRLRTPASRSHQSRP GAGVGVGGGVSQPAPF GKYPSRSQHFLPDVRPTS SSCASTSATQTAIVSPPSF RAIQFSPGRAVESPMGA PRLPDATTSSPRQRLRTP ASRSHQSRPGAGVGVGG GVSQPAPFGKYPSRSQH FLPDVRPTSSSCASTSAT QTAIVSPPSFRAIQFSPGR AVESPMGAPRLPDATTS SPRQRLRTPASRSHQSRP GAGVGVGGGVSQPAPF GKYPSRSQHFLPDVRPTS SSCASTSATQTAIVSPPSF RAIQFSPGRAVESPMGA PRLPDATTSSPRQRLRTP ASRSHQSRPGAGVGVGG GVSQPAPFGKYPPRSQH FLPDVRPTSSSCAGPSAT QIANASTLSATVPSQLAN DVQPQRPGSVDRRDSSL			

						LTIYYQNVRGLRTKTNQ			
						LHLSLSSCDYDVIAFSET			
						WLNAKIVDSELSSDYTF			
						YRADRSRETSVLERGGG			
						VLIGVKKRFHPEPIHLRG			
						GEKLEQIAVRIPLPGKTL			
						FVCCVYIPPNSDATIYTQ			
						HSSCVQQLCDLAQGPDE			
						VIVVGDYNVPNLTWHF			
						DEDINGYLPSGQSTEQE			
						VALIEDVLGTGLQQMFD			
						LPNANGRLLDLALVSDS			
						SRFQLIEPPRAILKVDAH			
						HRPFILVLETQSTGTAAE			
						QSGTYYDFANCDTMAL			
						NNSIASLDWANLLLVGS			
						VDDIVERFYDKLNEIIRE			
						VVPLKHRRPHPNSHQPW			
						WNRETRNLRNRLRKAR			
						KRFLKHPTEMLGRELAD			
						RELHYEMSLEAAFRGYI			
						SGIESNLKREPKSFWAFV			
						KKRKQDKGIPQDMVYR			

						DSSASTPDDSVQLFADFF KTVYSANQPTTSNESLD KILPHELHMPLVCLTEED VRKVLASTDASKGPGPD CIPPSIVKQCAATLANPL AIIFNRSLESGVFPAKWK VASITPIHKSGNVHNVEN YRSISILSCLPKVFEKIVH EFLSRAVQPIISACQHGF MAKRSTTTNLMTFVTD VLRSMESRCQVDAVYV DFAKAFDRVPHQLAVEK LRRLGLPEWITTWLHSY LTSRSAYVKLGGSKSGL FMIPSGVPQGSHLGPLIF LLFIDDLCHWIKNDKVL YADDLKFYRTITSPLDCL ALQSDIDSLMRWCDQN GMEANAKKCQVITFDR KLRPITHAYMMGISLLD RVCSVKDLGVTLDRKM TFNEHITSVTAKAFATIG FIRRNTTAFKDIHALKAL			

						YSALVRSQLEYAAQVW APYHNVHIARLERVQRA FVRLALRTLPWRNPPEQ TSIEDRRRLLGLDTLLKR RQRMQQLFIFDLLTKRL DCQKLSQLLIVYVPPRN LRRNPSMFRLSVHHTVY GHNQPFDVCCRLFNVVS DRFVLNMTKDSFKKLIS C			
RTE	RTE-2_CQ	.	Culex quinquefasciatus	GTGTGTGTGCGT CGCGCGTTATTA TCGCGCCTCGGT GCGGCTCCGCG AAAGTGCCCTC AGAACCGCAGA ATTACCGTTCCG TCGTGGCCGAA AGTGTGATCAA TCAACGCCGAA GTGCGTCGCGA ACCGGCGTCGT GTGTTATAATTG	TGATTGGTATGGTA TGGTATG	MARRLEIRNKRQKPIEK QTTDWRLGSWNCRSLN FLGFEFALAXELQPRNF DVVALQEVCLEEEQVLD WPGRQTNSRFXLSGGKD KKLGTGFIVRGKMQDR VFGFTAXSERMCKLRIR GRFFNYSIINVHCPHEEK TDDEKEAFYATLEEVYD GCPRQDVKIIIGDMNARF GREEMFRPTIGPESLHAV TNDNGQRCIDFAASRGM VVRSTYFPRKDIHKATW	cd01650	547	807

				ACCCGCAATCG		TSPDQRTKTQIDHVLIDG			
				ACGGCTGGGCA		RFFSDVTHVRTFRGANI			
				ATAAAAGTGAA		DSDHYLVGVDMRSKLS			
				GTTATTTTTGCC		TVFNQRRSRRAPPFNTA			
				GTCCATCGTCGT		CLQNGEVAHSYAQQLE			
				CGTCGCGGGAA		ANLPGEEELGAASLEDG			
				AACAGGCAGAA		WSRIRSAIGSAAEATLGS			
				CAAAGCAAGAA		AIRVSRNDWYDDECQRI			
				GTGTGGTGTAG		TAEKKAAYDRKLHKAT			
				CAACAAGACCA		RGNVERYRQARNRQVA			
				GTGGCAGTGCC		VFKLKKRQQEDRDCAE			
				GTTGTGTCGGTT		MEQLFRANETRKFYEKV			
				GCGTATACGCA		NRSRKGFVPRADVCRDN			
				CACTAATCGAA		GGNLIVNKSEVLDRWK			
				ACCGCGCGCGG		QYFNEHLNGDEADGDG			
				TGTGCCGAGAG		VGVNLGAPAADEQFPAP			
				TGCGAATTTTTG		DLETVKREIRKLKNNRA			
				AAAAAAGAAAA		AGKDRLPGELFKYGGEK			
				CGGGCAAAAAG		LARALHCVISKIWEEEKL			
				TTTTGAACTTTT		PEEWMDGVVCPIYKKG			
				TTGCAAAAAAT		DKLDCGNYRGITLINAA			
				AAAATAAATTT		YKILSQILCRQLSPHARR			
				TATTATTGCGAA		FVGPYQAGFTGARATTD			

				ATTTGGCGCTA		QIFCLRQILEKCREYNVP			
				GTAGTGAAGCG		THHIFIDFKAAYDTVDRE			
				CATTTACGCAC		QLWQIMHENGFPDKLTR			
				GCAGCGGAGGA		LIKATLDGVMCHVRVSG			
				CGTCCTCTTCCG		XLAEPFGSRRGLRQGDG			
				GATATGAAGAC		LSCXLFNIVLEGIIRRAGI			
				CGTTAGGCTTCG		ETSGTIMNKSYQLLAFA			
				AGATAAAAGCC		DDIDIVARNLETVMDVY			
				CCGTTACGGTA		TRLKVQARRVGLGMNT			
				KCGTGAGAAAA		TKTKYMRGRGSKDVGP			
				CCGGGGCTAGG		PSLTPLTVDGDELEEVSE			
				TCTCGGTAGCCC		FVYLGSLVTADNDTSKE			
				GTAGGTGGTTCT		IRTRIFAGNRAYFGLRKT			
				CCACCCCGAGA		LTSDRVQRRTKLTMYRT			
				CTAGGTAGCGC		LIRPVVLYGHETWTMRQ			
				AGCCCTGATAA		EDERALGVFERKVLRTI			
				GGCTGCCTACC		YGGVQVSDGVWRRRM			
				GAAAAAGAGGA		NHELHAFLEEPTIATQAK			
				TCGCGA		IGRLRWAGHVARMDES			
						QPVRLLFDRAEPAGGTG			
						RRAGKPRARWGDQVNG			
						DIRKICNLGNWRVAAQD			

						RERWKRLLATARVPGA LC			
Rex 1	Rex1-1_CCa	.	Cyprinus carpio	CAAG	TGAACCCCCCCACC CCCCCTCCCTCTAA CATACGATGTCATG CACCAGTCACTTTG TGCAGCAGTGGTCT GCTCACTACCTCAT TCACCATGGAACTG ACGTCATTCTACCA CCTCATCAGTCAGT TTAAATAAAATAAC TGCTCTTGAGCCCT TTGTCACTTTGTCA CTTTTAATCAGACT GAATAAGCTATTTT TTTATGCACTAAAA ATCTTTTTATCTGC ACTGTTTGTTCACT GGTTTGCACTCTAT CTGCCATGTGCCTT GCGCTGCTTTTATT TAACCTTATTTTTAT	MAARPHAAASLCPDRT VFSCFLSCESQCFCTSSS RLSVFKRAYSREFLLDV GRAAFLELNSAHAEQLR DLGLLRRPTSSPTPTTAS RPQRRRHKRRERKQKR GKRGGIRARLTANPHKP SIPTIVLANVRSLDNKLD YIRLLRSTQRTVRDCCVF VFTETWLSDSVPDCAIQ LDQLTCYRADRARVAG GKTRGGGLCVYINDAW CRDAVVICKHCSPLVEF MIIKCRPFYLPREYTAILL VSVYIPPNNNNSNRNDA LNELYQHISEQQTAHPD AFLIIAGDFNHADLKSVF PKIHQHINFPTRGNNILD FVYTTQRGAYKALPLPH LGASDHITVMLMPAYRP LVKVAKPVQKQIQVWP	cd01650	597	850

					TTTATTTTTTCTATT	EGSSDALQDCFDTTDWN			
					ATGTCTTTTTTACAT	MFKQAATYNNXTDLQE			
					TCCCTTATTGTATA	YSETVTAYITKCIEDVTV			
					GTTTTATCTTATATT	TKTITVRANQKPWMTGE			
					TTATATTTGATCTA	VYRLLETRNAAFRAGDE			
					GATTTTAGGCTCTA	GGLRTARANLSRGIREA			
					CTGTTAGTGTTATC	KRQYSRRIAHQFSDSRD			
					TGTATGCACCGGGG	TRSLWQGIQTITDYKPPQ			
					GTCTGAGAGTAACG	RTCDSNISLLNELNTFFA			
					CAATTTCGATTCTC	RFEAQNSTTAQKTPPPPS			
					TGTATGTATGTACT	DQVMMLTPDSVRRSFSR			
					GTACATGTGGAAGA	INARKAPGPDNIPGRVLR			
					ATTGACAATAAAGC	DCAAELTDVFTDIFNISL			
					AGACTTGACTTGAC	SQAVVPTCFKATTIIPVP			
					TTGACTTGACTTG	KKPSPSCFNDYRPVALTP			
						ILMKCFERLVMHHIKSA			
						LPPSLDPFQFAYRSNRST			
						DDAISTTLHSALTHLEKK			
						DSYVRMLFIDFSSAFNTII			
						PQQLIHKLVELGLNTSLC			
						NWLLDFLTGRPQAVRV			
						GSNTSSTITLNTGAPQGC			
						VLSPLLFTLLTHDCTPSH			

						NSNLFIKFADDTTVVGLI SNRDETNYRSEVSRLAG WCSDNNLSLNVEKTKEI VVDFRRAHTQHVPLTIN GATVERVSSTKFLGVHIT EDLSWTDNTAALAKKS QQRLYFLRKLRRARAPP PIMYTFYRGTIESILTSCI TVWYGACNASCRKTLQ RIVRAAEKIIGVSLPSLQ DIYGTRLTRKALCIAGDP THPSHSFFSLLPSGRRLR SLQARTSRLKDSFIHQAV RKLNSLPNLPPRPSSASG TTEL			
RTE	RTE-2_DPl	.	Danaus plexippus	ATCCAGAGGGT GTCCTACTAAGT ATTTTTCCGATA TTTAGCCTTGGG CATACGGGTAC GCGACCCCGGA TCCTCGAGCCA CCACTTTCATAC	TAATGCTGGGATAA TTGCCAGGAAGAA GAAGA	MVAKVYGLGTARASPIN DAQAHRPTPAISGQGLP CQVRARLKTLVPQRKLR FATWNIGSLTGRCRELA DVLMRRRVQCAFLQET RWKGNKSRNIGQGYRLI YTGSPSGKAGVAVVLSE ELRNGLLEVDRRCDRLM	cd01650	561	805

				CTCTATCCTCAG TCCATGAAGGT GGTGTCCTGTAC CTCCCCTCATTC CCCTTCCTAACC TCTTTCCCCCCT TCTCTATCTTTT CCTTTCCTGGTT TTACCCGTAAA ACGGGGTTTTG GAGGTCAGACG GCAGTCGCTCC GTTAAAACCAC TCCCGCCACTCT T		RVRVLIEGVITNLISAYT PQAGCSGSEKESFWEQF EEVLRAIPAAEVIIVGGD LNGHVGRAAETFDRVH GGFGYGRRNAEGENILR TCIASDLAVVNTFFQKTP QHLITYKSGSHSTQIDYL LTRRCHISKVTNCKVIPG ESLTAQHRLLVMDYVV TPKKKVAEKRKPRIRW WLLNGTMQTSFRAEVES QNLSTNTETAQEVWDR AQSAIITAGKRVLGLSKG GRVIDKETWWWNDEVQ EVIREKKTAFKKWQQSN SPEDRLEYIAAKRASKR AVARARSDRLSPLYDTL ETAEGQKLIYKLARARD KATQDIAKCLSVKDSQG TLLCNHASVKERWRCYF KELLNTQHPCSLPTEIPP NLGLIAPITPDETRNCLR RMKNRKAVGPDDIPIEA			

						WKSLGSLGVLILTDLFN RVLNTGTMPHQWRYSYI TPIYKGRGSVQDCGSYR GVKIMSHTMKLFERMID LRLRRECTVSECQYGFQ PGSGTLDAIFAIRTLMEA YREKRRALHVAFLDLQK AFDCVPRQCIWWALRFK GIPEAYIDIIRDMYRDSV SMVRTAVGDTKPFPISV GLHQGSALSPFLFNVVL DTVSANIQDQPPWLMM YADDIALIDESRLTLERR VNLWKGTLENGGLKLN VTKTEYMACGSPDSCTI HIGPEPAVKSEKFRYLGS ILHESGGIDHDVQARISA AWAKWRDVTGVVCDR RIPTKLKGIIYKSIIRPVLL YGSECWPTLSRHTQELH VTEMKMLRWMCGVTR ADRIRNTFIRGSLGVRDV ADKLQESRLRWYGHVA			

						RRPENYVGKICLDMSVP GARPPGRPRKRWLDTV KQDMRANGLTTADAKD RAKWRSLSRKADPG			
Rex 1	Rex1- 6_DR	.	Danio rerio	CCAAGATGGCG GCGCGCATTGA CGCAGCGGCTC AGAGCTCTCCTT CTCAGTGCTTTG TTTGTTTGTTTT TATATGTCTGTT TAATTTTTTTAA CCTCAGCCAGT CAGGCGAAAAT ATTTTATAGTCG ACAAGACCTCC TCTCAGTGGGA TTACAGAGCAA AGAGACAGTTA CAACAATGTTT ATTCAGAACCA CAGCATACCTG A	TAAATTGAGGGTGT TTCTGGCCAAAATT ATTTTAATCTTTTTA ACTTGCTGCTAGGA TATGCACTTGCCAC TTTATACACTTGGA CATTAAATTGTATG CCTTCCTGGCTCAC ATTTTATTAACTTTT AAATTTGAAAAGCT GCTTGGAAATGCAC TTGCCACTTTATAT GCACATCATACGCA CTTTTAATATTGTA TTGTAACTCTTCAT TTGATTTTATATTTT TATCTCTTTTTAATT ATTGTTGTTGTTTTA GATAAATAGGACTG	GHSENTGSPWFVIPGNN GRRREQKAETRKKSGCV DDSGKPHTSHRSPASSY ECQIIANKMDELGFSWT MQGLLPLMLTESWLHP GIPDAAVELTDFTLHRH DRNHLSQKQRGGGLCIY ISEKWAKDTKAVESLLP RFGVSDSNHCSPDLEYL TVKCRPFFLPREFSSAVF ITAVYIAPDANANSALN LLHNAISSQLNKYPEAA HIIAGDFNHVDLKIVFPK LHQHVNCATRGGNLLD KVYSNIKRGYKTIPAPHI GQSDHISLFMLPVYSSLR KRDPPTLKTVTVWPEGA SLQLQDCFAHTDWSVFE SQDLEEYTTSVLSYIKHC	cd0165 0	546	799

					CACTGAAGGGAAG GGCAAATGTAAATT TCGTTGTATTTTGT ACAATGACAATAA AGATTCTGATTCTG ATATT	TDTVTTEKHIRVYANQK PWMTGGLGFCALYRHC HHRETHQSVCQPKTLDD RGVRILLKERNIAFKSGD KERYSAARANLKTGIRG AKLKYKHKIEGYLHSNN SRQVWQGIQNITNYRPN PRSVGVDNTLAEELNCF FARFEQQSAEEATLHSLS NPSHVFTVEEDDVRKVL RTVNPRKAAGPDGITGR VLKDCADELTGVFTKIF NLSLSQAIVPLCLKASTII PVPKKTTINSLNDYRPV ALTPVVMKCFEKLVRSH ILQRLPPAFDAHQFAYRE NRSTEDAIAIALHTTLTH LEHKDSYARILFIDFSSA FNTIIPKRLVSKLNHLGF PFSICLWIQDFLSNRSQR VKVEHHVSSALSLSTGS PQGCVLSPLLYTLYTHD CTPFHSSNTILKFADDIA			

						VVGLITDGDETAYRDEV NQLALWCTDNNLALNT LKTKEIIVDFRRNRPDTL PIHIHGDCVEIVSNFKYL GLHLDQNISWKTNTIAL VKKAQQRLHFLRVLKN NNLAQELLVSFYRCTIES VLSYCIPVWYASCTAQE KAALQRVVNTAHKIIGC TLPSLEELFKARCYSKAC SILKDSSHPAYHLFQMLP SGRRYRTIKAKTNRLKN STYPRAVAIMNGH			
Rex 1	REX 1-1_DR	.	Danio rerio	TCGATCCAAG	TGACTCTTTTGCCC CCCCAATACACCCC CCACTCTCCTCTAA CTTAAACTCCTCGC AATAACTGCACTGT TTAACATTTGCACA TTTAAAATTTGCAC ATTCACTGCACCAC ATTGAACTGTTTAC TTATTGAACTGTAC	MAARAQREAQGLSSFCI FAVLFLLISGLFVQNSGA FTSYTRQELLDIGLCIPDS FISNLRLIHEIARTPRAGR PARPGGSARRPRRERKQ RRGKRGGLRAKLKLTPH RLSLPSIFLANVRSLVNK MDEIRLRINHSKRLWNC NVMIFTETWLNSGIPDN AVSLTEHHTFRADRTAD	cd01650	584	838

					ATACCCACTGCATA	DSGKTRGGGLCIYINKA			
					TGGGCATTTGTAAT	WCTNSVVVGRHCSANL			
					TATGTACACACCCA	EFLMVKCRPFYLPREFTS			
					CTATACATATACAC	TIISAAYIPPDADAKLAM			
					TTGTAATCATGCTT	NELHAAISKQQTAHPEA			
					ATTTATCTGCATAC	AFIVAGDFNHSNLKTVL			
					TACTGATTATTAAT	PKFHQNILCHTRGNKTL			
					AGCAACCTGTACAT	DHVYTNIAEAYAVTPLP			
					ATATTCATATATTG	HLGQSDHLSLFLTPKYSP			
					TAACATATACACTT	LIDRVKPSVRTIKVWPA			
					GTAATCATGTTTAT	GVDSTLQDRFQHTDWG			
					CTATCCCTGTACAT	MFASQATHGSHTDIDSY			
					ATATTCATATATTG	TSSVLEYINTTIDSVTTQ			
					TAACATATACACTT	KQITTYPNQKPWMNKE			
					GTAATCATGTTTAT	VRLLLKARNTAFRSGDA			
					CTATCTGCACACTA	QAYSTSRANLKRGIKKA			
					CTGATTATTAATAG	KHCYKLKLEEHFSNSDP			
					CAACCTGCACATAT	RRMWQGIQAITNYKPSQ			
					ATTCATATATTGTA	STPTATDVSFLNELNDFY			
					AATCTGTTCATAGC	ARFDRDNKEPYTRITSST			
					TTATCCAACCTGTA	DHSPITLTSSEVHTALSRI			
					TATAATGTTCATAG	NGRKAAGPDGIPGRVLK			
					TACATCCATCTGTA	VCAEQLAGVFTDIFNLSL			

					AATATCACCATAGT TTTCTATAACTGCA CTTTATAACTTATT CCTGTATCCTGCAC TTGCTGCTATTGCA CTGCTGGTTAGACC TAAACTGCATTTCG TTGCATTGTACTTG TACATGTGTAATGA CAATAAAGTTGAAT CTAATCTAATCTAA TCT	NLAAVPTCFKTTSIVPVP KHSSPTCLNDYRPVALT PIIMKCFERLVLAHLKDS LPSTLDPHQFAYRGNRS TDDAVSIALHSVLTHLD NKNTYARMLFVDFSSAF NTVIPSKLLIKLRDLDIDT SLCNWIMDFLTNRPQNV RSGHICSTTVTLNTGVPQ GCVLSPFLYSLFTVDCRP VNRSNTIIKFADDTTVIG LISNNDETAYREEIQHLA TWCTDNNLLLNTNKTK ELIVDFRKGRTGSHDPIH INGMAVEPVSSFKFLGT HISKDLSWTTNTSSLIKK AHQRLFFLRQLKKNQLS SAVLVNFYRCTIESILTN CVTVWYGSCSVAERKA LQRVVKTAQRITGTTLP AIEDIQKKHCLRRARSIL KDTFHPAHRQFSLLPSG			

						RRFRLPRTKTSRLRNSFF PRAVSLLNSAPH			
CR1	CR1-10_DR	.	Danio rerio	ATTGTATAGCTT TGAGGCGTAAT GAATAGTGCGA CAAATAGAGTT AAACAACCTCT TTGCTTGCCATT TTAAGACTTGCT GTCTTTTTTTAG TCTCTGTGTGCA AAAACGTCTTG CCTCCTATTCAG CGTTGGCGCTGT GGCTTAGTTGGT CAAAGTGCCTG TCTGGTAAGAA ACTTGGCTGAG GCCAAG	TAATATATTCAAAA AACATTTAAAAATC TGGTTTATAGAGAA TCAGACGTGTCATC ATTCCGATTAAAAT GAACCTAGCATGTT TCCAATTGTGCCYA CCATCTTAATGATC TGTTGTTTAAATGT GTTTCAACTATTTTT ATCTCTATTACAAA ATGTTGGTGTTTTA GYTTAATTTGTAAA TATTCTCTTTTAAC ATCTAYATRACCAT GTCTATTTTAGTAC TTTTTAACTATGTTT GATWGATTGTATA ATTCTTGTATTGTTT TTTTTAGGTTGTTC AGGGACAACAGAT	MATGGTGVVITLLVYCY FTSLTAAIRVDPVRMFCS IYPLSYHNDNPHIYKDLG YKRKGNELAKTSTGYGT YALAVTMVYPSLYNKIT KKKKAMTTTQSSRTGLII AICLMLSGDIHPCPGPHH FKTPREANPQTEDKRAT LTLQVCSSCTLLCSSAQQ ANPLLSLCPTPNTLKLSV QGVECVKGAAAVPGWS DGERQGEKTHPRLCVG MSDRSFGREFAGRVCAF RGNKRPDPETQQQQQTF LNDSASTMATNSVNRKR LNPGVVKNRKHAFFNTV NNSKIVWDPKMKPKGL LGRHLNIRSILSKSEQITH LLTNSNLDYLCLSETWL YKYAPSAAFNISGYTMF RKDREGAKGGGVMIYV	cd01650	800	1056

					AAAAAATAGCCTTC KGGCTAATTCTGGT GCATTTGCAGAAAT GCTAATTAATGTAC ACTGTCCCTGCCAA ATAAACAAATAAAT AAATAAATAAAT	KDNLQCEEINWLCDYEF ECLGLNVRLSPQMSFKII VIYRPPSSDGTFYDRFEK LLKESNLNKETVILGDFN INWEDTSCRNKLKRITD KFDLVQMINGPTKITTSS STQIDLIFTNRPERIIKTY NMLAGISDHNLTLVARK LSKKRFKFSAREHEFIGIP RNRRDGFKNAVEQIDW DRLLLGTDLEKDRGSFV GTIEKTVKNSALNLNAN TIKIYYPGLTLTYHTKTN ERERFCVKKVIKSKSAS DRNMFTTLRNRVVKNIR KAKADYFLTLIEKAKGD SKTIWNQLKKVVRHDT KNKKMLEVMIDGHITNN ACKIAEAFNNFLLTQFQI WSHCVHQKFLTCALFFI NSVSNMVPLCPSKISNM CSLNSSQPIFSISNISESDA LQAIKSLKSSRTKETFGI			

						DSLMLKELSSTLLTPITK LLNRSISEAIFPCVWKSS AVIPIFKNGDHYSLSNYR PISIIPTVSKVAEKLIANQ LIKHLNTTTFSLHPMQFG FRSQHSTETANCFLIENI KHLLDKGGVVGAVILDL KKAFDTVNHKILLSKLT KFNMSPESLKWFESYLS NRSQSVRINNYHSSPLSL PTGVPQGSILGPLLFSLYI NDLPSVCQNILIQMYAD ETVIXTHGKNTTQVAEK LTQAMAHVSEWLNQSC LKLNVSKTVAMYFSKTN IVNNLTDIYVDGDKLQIV SDFKYLGVQIDSKLTFKS QIKKVCNRXKFNIRSHLS LQAAXIFMHSMILSHITY CLTTWSQANKTTLKPIES LYKQSIKILXKKPLRHHH CSILQKHKLLSWENVIK YSNICLLYKIIHGLSSPPL			

						SQFVNIRNSTHRATRGA ARGELYRPTPEKYLWSG CLFSYSCTGVEFCPWQH QRYNLL			
L2	L2-1_DL	.	Dicentrarchus labrax	T	TAGGGCTGGATCAG GTGAGTCCTGAACC ATCCCTTAGTTATG CTGCTATAGGCCTA GACTATCGGGGGAT TTCCCATGGTGCAC TGAGCTCCTCTCTT CCTCTCTCTTTCTGC ACTCATTCATGTCC CATTAATGCATGTT ACTAACTTCACTTC TTCCCCGGAGTTCT TGTGCTTTCTCGTCT CGCAGGTTCTCATG GATCCTGGTGGAGC CTCCTGCCATGGTC CTGCTTGACTGCCA TT	VRNHSINFHCYADDTQL YLSIKPDEXNQLSKLQTC LKDIKSWMSCNFLMLNS EKNEILVLGPKNLRDSLS KDIVTLDGITLASSTTVR NLGVIFDQDLSFNSHIKQ TSRTAFFHLRNITKIRHIL SQTDAEKLIHAFVTSRLD YCNSLLSGCPNKSIKTLQ LIQNAAARVLTGTRKRD HISPVLASLHWLPVKCRI EFKILLLTYKALHGQAPS YLKELIVPYYPXRTLRSL NAGLLVVPIVSKSRTGA RAFSYQAPLLWNQLPVV VREADTLXTFKSRLKTF LFDKAYS	IPR000477	5	93

CR1	CR1-3_DBi	.	Drosophila biarmipes	AGGCAATCGAT TTGAGGGGAAA AAATCACTCTA CCTCAAAGCCA CAGAAAAATAA AACCATACATG GAACTGCCAAA ATTCGACGGTA MCTACGTCGAG TTCAAAATTATT ACTACTATTCAA TGTCAAAAAAA TCAATAATTTTA TATCCGCTTTGA AAAGGAASAGT AAAGGCTTACC AAATCAMGTCG ACCATGAATAG GGGGTATGTAA TCCGTAATGCA AGCGCCATCCG CCGCAGTGTTA CGATCTCTCTTC	TGACCAAAAAGAC GATTTAAGATTGTG TTGATGGGTTTCCG CATTAGGATTATTA GGATTCCCACAATG TTTTGGAGCAATTT CTTGATCTGCAGCG AAGTGTTTGCCAAA TGCTGTATCAGCAA TTTCAGCATACGTT AACTGAGAGACAT AATTAGCGTCAAAT GTATTGACTTATGT ATTCTTATGTTACC TTGCAAACCATTTA CTTAATCTCGTTTA AAAATTATCTGATT TTGACAAAATTAAT GATAATTGTTATAA GCCGAGGTTTGATA ATTAACCAAATGCA ATCCAGTTGGGCTT TCCAGTTTCGAATC	MKAFMRQTRKALTEFR KQLTSSFEVFSIMETQFL DCKLLNESPKRKIANNN KNSLQIPSPCIAAVGIPSP ISACLNSILAVNNSAGPG TSGLQSSFAFNNTSPGSE PGASGLQPNQNTGTTAE GSSTPQQLAPGPGSLHIN PDHATTIAPAALPSSQPA ANQPLSAIPLSGVAPRSN AADLPLRGIAPNRSVFVS RLNPEATADCVKSYLCQ KLGAQPNSISVFKFNFKQ PRNISSFKITLPESMFSNA LDPTIWPEHTIMHEYIRK GASQDPNSVDAKNCKK LNRSISVAYQNVRSLLG KLKRVYSNSTLLDFDIIA ITETWLNTTVYDQEIFVN KFAIYRLDRPTRGGGVLI AVLNSHSSELIPYDNMH SIEFIAVKVTTDKCNVFL TCSYIPPGSDPAIYLHHL	cd01650	775	1027

				GACGCCCTTACT	CGAATGAATTTAGT	SAISAVSLAVGDADQLIV			
				AGTTTTTACGGA	TGTGTTCAGAAGAA	LGDFNLPHLSWLPNSDS			
				TTTCTTCTCTCG	TTATTGCACAACTA	NILFPENQHEFIDGLTNL			
				GCAACGAAAAG	AAGATGTTATGAAA	SLGQVNSIKNSRERILDLI			
				TTGGAAAAAGT	AAATAGATTCAT	FVSDVSVATVSRCDPVS			
				CATGTCCAGGTT		TPEDSYHPALYLSVHINF			
				AGATCGAAAAA		ETTNQSASQRGNTIRDN			
				CGCGGTCGCGA		YESICLTTREHYTGLNSH			
				TGGGAAGAATA		FDRLNWSTLYSFTDLDS			
				AACCAGTTAAA		AVLWFNESVCSALDTFV			
				ATTCTAGTCAG		PSYRWKLIKASLVLQAF			
				ATATGGAAATA		VLHIPLSQVTVGNSLKPP			
				TTCTTGGGAAG		WFSRHLSYIKNTKSLLK			
				CGATTAATAAA		DIKNPRFFYAEYIAECRQ			
				CGTTATCAACA		NIRSDPKKFYSFVNSKRK			
				CTTAGTGGCTG		SNVFPPCLHLGNLSESSD			
				AGGATTCCACG		INISNLFATFFQSTYSLPH			
				ACAGGAAGGAG		INNIHIPPFKISDVLNGML			
				AGCTTCCAAGC		TIDESFSAGPDKIPSCIIK			
				GGGAGCAGCCA		QCAASLCGPITFLFNLSL			
				AACAAGCCATG		QLSFFPSVWKTSFLIPLH			
				CTTTAGAAAGA		KKGSKTDVGNYRGIAKL			
				AGGCATTCGTA		SAIPKLFERLVTCQIQHM			

				GTTTCGTGACCT		CSSIISTSQHGFVKKRSPS			
				GMCGAAGACTC		TNLLEFSRLIHCSFLQKK			
				CCAATTGAAAT		QADVIFTDFSKAFDSVN			
				KTCAATTTTGTA		HSLLIYKLSLIGFPPYLLD			
				AGTCAAAGTTC		WISSYLSNRSQRVLFKN			
				TATCAATTGCCA		VVSSLVNVTSGVPQATS			
				AACTGTTTACAT		HLGPLLFILFINDLPQVM			
				TTCGCCCAAAG		SYCTTFMYSDDVKICMS			
				TCTTAAGAAAG		FSNLSSHLLIQSDLNKFQ			
				GACGAAATTTC		AWASTKLLSLNYSKCKT			
				AACAGATGTAA		MSFSRTTFHTTSYTLGNS			
				TATATGCAAAA		LLERVVKINDLGVLFDQ			
				CATAGACCGCA		KLCFNSHISVMVNKATG			
				TTAAACATATAT		VLAFIKRWAKEFNDPYT			
				TGTTTCGGTTTT		TKSLYVSLVRPIMEYCSE			
				CCCTTGCCCGGT		VWNPQYVYQQNLIESV			
				TTCGCCTTTTGT		QKQFLLFALRNLNWDPD			
				TTCTAAGAATA		RHLPPHLSRLKLIDLNTL			
				GATCGGGAGTT		KHREFNCKIF			
				ACAAAATTACG					
				ATAGTATTGAA					
				ATATTGCTTACA					
				TAGTCATTATTG					

				CTTTTCGGCGAG CCGAGCAGCTC GATTGCTCAAG AAATATATAAA CTTAATTATTTA TTTATAATGACA ATTATTATTTAT ATAAAAATTAT ACAGTGTATTG GCCAGCCATTC CATTGAACAAA CGTATTGCCGTT GTGCTCTATCCT TAAATTACGAA CTATAGGCAAG TCAAACAGTCA TTTCGCAACGAT ACTGACCGGTT GTGTTTTATTTG AGCTCTTGAAA TTACCTTTCAAT CGCGGTTACAC CGAACGTATAT					

				AATCTTAAAATT GGAGAATAACT TTCTCAACCCCC GCTTAACACTTC ACGCACCGTAT ATAAACATTAC ATCGCGCCAAT TAACATATTTAA TTGAATCGCAA TTCATCTGCTGT TGCCACAAGKG CTTCTGTTCCCT ATCGCCGTCGC GTCTGTTCTTTA CACCCGCTCGC CAATTGCTCTCT CAAGGCCACGG TCCTCTCTCCGT GCACTTCACTTT GAGCGCTTCGC CTCTCTCATCGC CTGGATATTCG AATTGGCTCTCC					

				GAATCGACGAC					
				TCTCCCATCAAG					
				TATTCTAACGGT					
				CCTCTCTCCGTT					
				TACTTCGAGTG					
				GGTCGCCTCTCT					
				CATCGCCTTGCT					
				TGCTGCGGATCT					
				TCGAATTGGCTC					
				TCCGAATCGTC					
				GCCTCTCCCATC					
				GCCGTATCTGCC					
				ACTACCTGGTGT					
				TCTGCGTGTCTC					
				TTTAATCTCCCA					
				CGGCCTGCTCTT					
				CTGCAAGATTCT					
				CTTAACTGCCTC					
				TTAACTGCATA					
				GTGTCCAAACA					
				ATCACTAAAAT					
				GAGTGTCGTCA					
				TCTGTTCCGCAA					

				AGAAATGCGAG TTCAATTGTGTT ATCATTGGAGA TAATTATTTGTC TTGTTGGCTTTG CGACAATTGCG CTCATGTAAAA TGCGCAGGAAT GGGTCAAAACG GCCGTGTAGTT GATCTGATCTCG AAGAGAGGCGG CTTAACTTGGTT CTGCCCTACCTG CCTCGTACTTAC TCGGAG					
RTE	RTE-1_DF	.	Drosophila ficusphila	ATGTAGGGCCA GACCAATAACA ATCGTGCAGTG TTACGCCCCAA CCGAAGACGCA AACGACGACAT CAAAGACGACT	TAAAGGAAAAAAA AAA	MGDFNAKIGPNFTRMRS ATGGHGIGTRNNNGDRL VEVCQMFQLIIGGSMFP HRDIHKYTWTSPSGNTR NQIDHICISRKWRKSLLD VRNKRGASIDSDHELIIG ELSIKLDRNRPRNADDS	cd01650	381	637

				TCTATACCGCCC TTACTACCACTC TCACTGACATA CCTCGAGGCAA CATCACCGTTCT C		RNHRPSPLNIHLLGETDI SARLASTLRTHLLPQQG QTWEQTCDTLRTTAEGII GTQQRRRNNWISSATW DTINKRNSLKLLADRDS NRRDEHRALCKAVKRA ARNDKRALLDRLADNA ERAANENNMRSLYQTIE QISGGYRRKTQPVKDAN GILLTSDDDQIHRWRQH FMEISNSDMPSRSDDYR SVAPANGNRRISSAPPSV REIGNAIKKLKNNKAAG MDGIPAEILQVDTQLMA EILHPHFSSAWESEHIPD SWKCGVIVKIPKKGDLS DCNNWRGITLLNTSYKV LATILNERLTEKFEPTLR DEQAGFRPHRSCVDQAN TLRIITEQAVEWRAPLYL LFIDFKKAFDSINRAAIW GALARKGVPHKVIALIK ALYDDANLAVLHNGKT			

						SAPFQTNTGVKQGCPLS PLLFNIVVDEVMREVCK HKRGITWSFTRHLEDLD YADDICLISHKLSDIQSK VDDLVNLASRVGLQINT AKTKAMRLNHASQGNII VDGNPIEYVESFPYLGTV ISSNGGVTEDVTSRLAK ARSAFGRLYRIWRSRQIR RKTKIRIFNSCVKSVLLY GSETWLTTKTIIQQLQAF TNKCLRIICGIFWPNRISN IDLWNMTGETAIRLQIR KKKWRWIGHALRKDQX SIVRTALEWNPQGSRRV GRPKATWRRTALQELSQ YNITWDTAKQTAXNRV RWRNLVEALCSQEEL			
Penelope	Penelope-1_DF	.	Drosophila ficusphila	AC	TAGGTTTAGTTTAT AACTCCACGTCACT TGGCGCAGACGTGT CAAAATGTATGTCA ATATTTAAATTGTA	IVNNKWTILKEHTXTPS KVFLDIVKFCIQDNRYL KYETKLXIXLKGMPMGS PASPVVADIVXETLLQN TTEKLRRPPRLLTKYVD	cd00304	47	129

					ACAAAAAAGTGTAT TTTGATGTTGATTG TTTTTTGTTTTCGAT TAGTTGCCCTGAAG ACGGTTGCCGAGGA GCAACCGAAATATA TTGGTCAAAAAACA TAAATTGTGTTTTT AATATCACCCAGAC CTTGAGCCGGCAAA ACACCAAATAT	DLFGIIKEDEIENTLTTLN AFNXNIKFXMELEKDNC IAYLDTVVIRRGNELKL KWYKKPIASGRIINFNSR HPKMMIRNTASSCINRM LRISDAIFHEEIINEIKTIL TNNDFPKHMIKSLLDKY LKSDNNQRPPKKERTYIS VTYVPGLSERLAQSDCY DKENILIAHKPDNTLKQI FNKTKSKIAPTDKKDVV YRIPCNGDGKNNCNSIYI GTTKSKLKTRLSQHKSD FKLRHQTNTQKTALMSH CAKQGHSPNFEETTILAQ EQHYRKRYMMEMLHII NTPTNIRMNYKTDTENC AHIYRHLLGK			
RTE	RTE-2_DF	.	Drosophila ficusphila	GGGGGGATGGGGGAGAGCATATAAGGCAAAGTGTCTGACTGCATTCAACTTGGTGA	TGATGGCCAATTATTATTAA	MKEAEEASERDRFNVDDHGKRKKDNDLRVCTWNVRTLYRPGAAQQLADTLKKCRADIAAIQEMRWIGQGCIKKKDCDIYYSCH	cd01650	551	807

				ATGCCAGCATG ATCCTTATCACT GATCTCCTCAA GCCGACGCACT ACCTCCACGTG GTTCTCCCTGGA TTGCCATTCAAC ATCCGTTGAGT GGCAAACTAAA GCGCGCGACGA GAGTCCTAGCT CTTCAAACTAG GTTGTGGCGTA GGACTTGAAAT CCACCCATGGA AAAACCAAAAT TCA		PERHEFGCGFVVGARLR RRVSHFLPINERLATIRIT AKFFNISLICAHAPTEEK DDVTKDTFYAELDRAY GRCPSHDIKILLGDFNAK VGREDIFGATVGRFSLH ETTSSNGLRLIGFAAAHN MVVRSTGFRHLDIHKAT WLSPDRLTRNQIDHVVI DARHASSVLDVRSCRGP NIDSDHYLVAAKIRTRL CVAKVARRGLLRRLDIG KLQSQQTRDAFSTQVSG LLSRTTPTPEDISDQWAL ISHSLRTSAEEVIGFQRPP TRNPWYDQECRDATAA KDAAYRRTLQAGATRA VVEVYRLRRRDEKRLFR RKKREXEKRECESIESSR CRNEARNFYQRVKRLTQ GCKTGGVACKDNDGNL VTDAEIVLKLWRDHFSS LLAGSGTGYEDYDPQTP			

						IYGTDIDVPTPSHAEVKD AIQRLKNNKSAGADGLP AELFKAAGDMLVGSMH QLISKIWRTESMPEDWN LSMICPVLKKGDATLRT NYRGISLLPVAYKVLSS VLCENLKPHAEALIGPY QCGFRPGKSTIDQIFTLR QILEKSYENQIDTHHLFV DYKAAFDSPRRDRLYAA MSELGIPAKLIRLCRMTL SNTISSVKVGNDQSETFS TKCGLRQGDSLSCMLFN ILMESIVRKAGVHRTGTI LSNRCVQLLGYADDIDII GRTKRDVTAAFGAIEKE SAKVGLAVNMDKTKFM VCSSRESRRLDSQLTADS YSFESVKEFIYLGTAITST NDVSLEIKRRVTLANRC YFGLSRHFNSRALSRQA KITLYKTLILPVLLYGAE CWTVVQSDAAALGVFE			

						RKILRKIFGPICVGDAYRI RWNHELYELYGDVDVV SRVKIHRLRWLGHIARM EEDAPARKVFDAVIVGK RRRGRPRMRWQDQVME ALSTSGAANWRRRAQD RDAWRQIVQQAVTRQG L			
R1	R1-1C_DK	.	Drosophila kikkawai	GGAACAAACCAACCAA	TAGGGGTCACAGCGCCCGCGAAGCAATGCCAGGCGGCGGTACCGCTGGGCGTCCGATTCGTATTCTTTCTGTTAATTGTTAAGCCATCCATGTTTATGTGTAGTGTATGTATATATTGTAGAGTAATAAATATAACGAATATAAAAAAAAAA	MQLIQLNLNHCRAAQDLLTQTVRERGSEVAILSEPYRIGNSRDWATDRTGKAALWLCGVGAQQMCDIKAADGFVRANVGGTWLYSCYLAPSLSLESFGRILDELSIDLRGHSNVVVGGDFNAWAIEWGSSRTNARGRAVLETFASLDVVLLNEGSRQTFSRAGAGSVIDLTFVSSALARRARWTIGEVYTASDHEAIECSLGMAGRPSGPPPLPRRAFRQDTFRPRAFASALEGFAVDEADGANVMANKLADAVE	cd01650	455	717

						DACDRSMLQRRPFRGN HRPVFWWSEEIAELRRIC HRSRRLLQRARGSPLLV ECSERYKAARKALKIAIR DSKRESFLKLCDAAEED PWGGAYKMTVKKLNA GGNAPSDPEALEGIVRA LFPDGQPVGNFPRAEPG EFRSVTEAEVLQAGRNI KPKKAPGPDGIPSRATKL ALALLPEQVAGLFNKCL QEGTFPVRWKRQRLLLL AKPGKPPGEASSYRPICL LDIIGKVFERVIATRMTE AVEAAGGLSTEQYGFRK GRSTMDAIMKVVKTAR EAIAGSRWKGGTKSYCL VVTLDIRNAFNTADWGR TLESLRMFNIPSYLLNVA LSYFSNRVLTTDTCLGSR AYKVSAGVPQGSVLGPL LWNAMYDGVLRLPMPE GTSLVGFADDVAIVVVA			

						KELATVEAHADAAVQA VESWLANAGLQLAAHK TEAVLVSSRKAVETASV RVGGTTIRSQRAIRYLGV LLDTRLCFKEHLEFVHK KASGTAGALSRMLLNTR GPKQATRKLLTTVVTSQ MLYAAPVWAEAAKVKS YVRGVEATYRLCAVRIA CAFRTISEDAALVIAGQV PLTELVRERAEIYAAAQ XREATQQTRAESKITAR RRSLENWQNRWDSSQK GRWTHTLIPNLTSWVER SHGQVDFYLTQVISGHG CFRSYLKRFGHETEDWC PVCGSGVVEDARHVLFE CHRFDHERQELEAVAGS SIXAETLVPMMLADQRV WEATATFAADVMKSLR EQERRRKEQTE			

Kiri	Kiri-5_DK	.	Drosophila kikkawai		TAGAGCCCTTTCAT TTTTCCACAAAATA GTGACAACCCATGA GCCTTCTTACTTGTT TAATAAATTAATAC CGGCAGCCTCGCAA AGGACTAAAAATTT TATATGCCCTCGAG TCCAAATGCTCAAC TCCGAGAGGCAGTT CTTTGTCAACGCTA TCCGTCTATGGAAC CAATTACCCCATGA ACTTAAAACTATTA GGAATGCAACTCAA TTTAATTCCAAAGT AAAACTATTTTTAA GAAACAACGCATA ACAAACAAATAAA CAAACAGCAAACTT CATATTAATACGTA ATTAAAGATATAAA AATAAATAAAATA	SVAPCDNIFNFVCVSQC DVLDCILKVKSNAEGLD NINPKFLKLLLPCLLPYV TYIFNSILTQSSFPTAWK LAKILPIPKQNNEYRPISI LPFLSKVFENVLAGQIQG FLDRNLLLHNKQSGFRR KRSCLTAITDITEDIRQQ LDKNHVTILTLLDHSKA FDSVNHNILCTKLKNLF NFSDSSVRLIASYLTDRL QAVVAGAVTSAVKKLT RGIPQGSVLGPLLFAIYV NDLPNVLSDCEVHLYAD DVQIYVSSPVREINVCIS TCNRMLSLVQTWAEHN GLSINPKKSKSILISRRPI DTNDLTKPIINNIPISYVD HAVNLGLTFNNTLSWG NHISKATGRVYGILRQL ANTRNFLPEHIRLLISKTI LIPTLFYGIEIFGFCDTRD TRKLVTALNSIVRYVFN	cd01650	72	326

					AAATAAAATAAAA TAAAAATAACTAAA AATAATAATATAAA ATAAAATAAATAA ATAATTAGTTCATA TAATAATTAATGTA CTAGGTATTAAGTA GAAATCCTGTCTAA TTTTAACTCAAGTG ACCAGGGCATTAAT AATAAGTATTTTAT ACTTGTAGCACTGG GCTAATTTTGTCAA AATAAATAAATAA A	LNRFDHVSHLTTKLLDL NLNSWINL			
I	I-3_DK	.	Drosophila kikkawai	A	TAATTATAACATAT TCCACACACATATC TTTGTTAATACACT GTAGAGCGGAAGG CCTCGGTAGCCATT GCTCATTATCCTTT GTTAACTTTTAGTT TATAACTATTAGTT	TIDPAEIANTLGDLWSK QAKDSNFSPEFRTNKML ITPHNTYAPTPTAIEIEKN INLIELNSAINSLKGHTPG LDRISYSMIKNTPISFKFR ILQHFNDILDSHIPQSYK LSLVLPILKPNKPKTNPN SYRPISLNPCLAKVLDKV	cd01650	111	368

					AGCATTAAATAAAT AAATAAATAAA	VANRLWWFVLSNKLIH NNQFGFKRGRSVLDSLL YVDHLATRSLSLKKHISI ISLDFAKAFDKIGIHSLIH QLEKWKLGPRIIKYVWN FMTNRKIIVKNGTSLSISF PLDNGIPQGSPISVILFLI AYNSLTNIISIPNXLXFTA YADDFHLILQHNRNNNP QVNLSTLFHQIGEWCSY SGSSLSIDKCQLLHICRK RSCRANIITGSINIPTTDH TKILGLTINKRYSWNHHI TNLKTSLSRPLNIIKHLSS RKYNCNTTSLIQIVNSILI SKIDYCLPIYGKAPTSILN CLKSTQNSAIRAALGAF RTTPINNLLIEAKINTLEN RCKLLTSKLGTKLILSHN SPIGKLSKSIAKTKRNYR YPSTLDRIIDSCKSFSLPF SPAKFYTSKKYNPGIILK ATNTSLSLYNKNTTSPIT			

						YHQLFHSIKNSLHDHKFI YTDGSKTNDTVGYSVTN NTDIIKSGLLPQYSSVISS ELIAIHEAIKICSGSPGKY AICSDSLSSIKSISNVNNY SYYPNTIRSLITKLFPKIR LIWIPSHIGIQGNERADT AAKETHKYPLIFTDNFN TKDITQLLKKHYTDKLN SLQAQTTQWYKQILLNT PANFSFKTTNLNRQQQIII NRLRLGHTQLTNAHYM NRSLTSICPFCNTSAIDIQ HILIDCTALSHVRTICFSN DNPIHLLANPTAANSNKI IYFLHKMNLIHKI			
Kiri	Kiri-1_DK	.	Drosophila kikkawai	GGCCATGCGAT GCAAAAAGCAC AACCAGCTCAC AGGTGTTTTTAC ACGACGTGGTG TGGTGTACGTG AAGCTGAGCGA	AG	MHNICATGNFSYILEME GSVDVSHSHTLLRILAK QRSGLKICHINAQSLPKK IEEFRYMFVNSGLDVICV SETWFVQDFRDSLISCD GYYVYRGDRVSHGGGV AIFVSKKLKTKFVCKHP	cd01650	505	758

				CGGAGGGGAAC CTCGATATGTGC AAAGCTCACAC GAGCTTTCTGA GCTGATCAACG ATGCTGGCGGC GCCGATAAGAG CACATTTCGAG ATTAAGAAGCT TCACTCAATTTT GTATCTCCTATG TTTACACATAAC TATAATATTTAT TAGATTTAAAA TTAATGTAGGTC GCTATGTTTTCT TTGTGTCCTTTT CAGTATGTTTCC AATGCCTTAATT T		DNSQIEYIFLEISNKFASS KILIGCIYRPRCDITYNEII SLLGDISLNYADIIVAGD LNSNMLRETHLKNEFESI GVHLVNSVFPTHFSRTC SSLLDVFFVSDVKKVYL FDQLIVPSFSKHDLIYLT YNFDLQYNDESIVFRDF KNLDYGLLMTEFDKCN WNNIFVSPTVDQQLAQF QNNISYLYENFVPVKTK VIKHSVKSWFNNDIKLAI KSRNDAYQKWKRYRTP EHYGRFKLLRKTVTKLI DSAKTNFYSSRFQTTAST SRTWKEIKSLGIGRKPAA PEQXVDIEELNTSLSKCN VPEICENNMYLRSVPRN LCDSRFRFVCVNQCEVL VNILKIKSDAVGLDDXN PKFIKVLLPKILKYITYIF NTAFTTSTFPDRWKIAKI IPVPKANKEYRPISILPYL			

						SKAFENIMASQINSFIHD NALLNNNQSGFRKHRSC TSAILKIANDIREYADNS NVTLLVLLDYSKAFDTV NHNVLCAKLVNQFKFSK TAVKLMTSYLQNRRQS VVVSNVSSSFAFTKRGV PQGSVLGPLLFTLYVND LPSMLSSCDVHMYADD VQMYVSRPLNRINECVD ICNNELVKISNWAKENG LGLNPSKSKCIIIYKKKLS LNIGPLLLNDSPIEYVER ATNLGFIFNNTLNWSDHI NTAVGKTYGLLRSLYLV RNFLPLNVRMLLAKSCL IPTLFYGIELFANCDSNS QHKLKMAFNSITRFVFN LRRFDHVTRLSYNIFELH LADWIKMRVLVYLQKII FSKLPIYLYETLRFSTSRR SNIIIQIRH			

						SSGPRASDSFNQRSKKL			
				GCTCATTTTTGT		NNSFSLHYQNVRSLLGK			
				CTCTCGCCTTAA		LSKLHVNSVSFDANVIA			
				TCCTGGTGCTTC		FTETWLNSSVSDCEIFAG			
				TGAAGATGACG		SYVIYRCDRSMRLGGGV			
				TCAAAAACTAC		LIAVKSDLVSQRIPIIGTD			
				CTGAGCAGCAA		SILEFVAVRIKLHDYFVFI			
				ACTGAATGTTTC		SCSYIPPRSDISLYLQHLA			
				TCGAGCCGAAC		QIQSIHSTLGYNDHLIVM			
				TAACCGTCTTGA	TGATGTGAGTCGGG	GDFNLPDISWMTSTDSIA			
	CR1-4_DRh	.	Drosophila rhopaloa	AGTTCAAGTTTA	TACTATGTTTATGT	LVPTTAHEFVHGLIDLSL	cd01650	514	765
CR1				AGCAGAGACGT	TGGCAGATATTTGT	GQLNSILNCSKKVLDLIF			
				GAAATATCTTC	TAACTTTAATGTAA	VSEFANAFVTRISPLSNP			
				GTTTAAAATTG	TAAATAAATAAA	EDAYHPTLEVTLEASAP			
				GACTTCCTGAG		LQSCSEIEPASFRCFRRA			
				GAACATCTTAA		KFAELNRHLSELDWSFM			
				TAAGGCTCTCG		DSNDDPLHIVEQFYGILY			
				ATGTTTCTATAT		SAFDTFVPSFPSANAPQK			
				GGCCTAAACAT		PPWFSRRLSFLKNSMSR			
				GTTATTGTTCAC		LFKKFKRTGLQVDHSM			
				GAGTATGTAAA		YLIARSNFLVHNSVCYN			
				CATGCCTAA		GYLNRCRIEFQDNPKQF			
						YAFVNSKRKSQQFPSCL			

						HLDTTSASSDADIANLFA KFFQSTYSSFSYDNTSSY PYPFLTPTVIFTPLISEDD VLQNLKTIKMSFSSGPDL VPSCVIKKCADALCKPL SKLFQISLLHSFFPEVWK ESFIIPLHKKGSKSDIKNY RGIAKLSAIPKFFEQLITA QLQHQCSSVMSPTQHGF MKRRSTTTNLLEFTSVIR SAFKNNIQTDVIFTDFSK AFDSVNHQLLLKKLTLL GFPIMLTKWISTYLSNRT QRVAFKSTISNSVHVNS GVPQGSHLGPLLFGIFIN DLPQVVLSSYVMMYAD DVKLCLPLSNPNSHDLL QADLNNLHSWCIQNLLF LNNSKCKLMSFYRKRLL LTSYSIGSNDLERINSVT DLGVLFNHNLNFSDHIG TIVSKAKGVFAFVKRWS KEFEDPYTTKHLYISLVR			

						PILEYCSCVWSPQYTNH QDLIESVQKQFLLFALRG LNWETHRHLPSYVDRLR LLNLPSLKDRRTCHGVM FLHKLITGLVNSSYLLGQ IRFSIPVRSTRYFRPIALDI CKTNFEVHDPFRVLCSQ YNELYFLIYSECSVNLIV TNILTHLSLPPPT			
R1	R1-1_DSi	.	Drosophila simulans	CACCACGGAAG ACGAGTTGAAG AGTGCCTTGGA GGCCCAGGCAG AGATCCCTGCA GCAGTCGTGGC TATCAAGAGCC TTCGGCAATCG CAGTACGGAGG GAAGACTGCGG TGATAGCAGTC CCAGCCAATCT GACGGACCCGC TGATCAAGCGT	TGAGCAAATCCTGA TCGTCCTCGCGAAG CAATGCTTCACGGC CATACCGCGGGGA GTCCTGGATTTGCC CTCNTAAATGTTAT TATGTTGTATAGTA TATAGTTTAGAATA AGGGTTCAATAAAC GAATTAAAAAAAA	MMRLIQLNLNHCTAAQ DLLVQTVRERRVELALL SEPYRTPASPDWAFDRT KKAAIWRCSRETQQITD VLSDIGFVRARVGRWW VYSVYLAPSLTLIAFSRA LDRLAADARGRTQVLIA GDFNAWSESWGSSTTN ARGRMVLEAFATLDLAL LNHGNRHTFRRAGLGSV VDLTFTSGSSYRLARWR LSEDYTGSDHLAIICDLG SPPSSQAQTPAQARIKYK TDSLDTQVFREQFLPTAS	cd01650	473	720

				GGCAAGCTGAG		RQGAELTAVELMRRLK			
				GGTAGGATGGT		AACDASMQSSRPHSRQR			
				CCCAATGCTTG		APVYWWNQEIETARRE			
				ATCAAGGAGCT		CLSARRRYQRARGAESF			
				GGAGCCACGCC		AERQSEYRARRKTLKLA			
				AAAGATGCTTC		IRESKRKCFLDLCDSADS			
				AGATGCCTGGA		DPWGSAYKVVVKKAYT			
				AGAAGGCCACA		RTPKLLDPAMLRSVAEH			
				TAGCGGCCCAC		LFPVMDGLRPADPATGD			
				TGCAGAAGCAC		HVDSDATVSSDEILELA			
				CGTGGACAGAA		KLLKDGKAPGPDGIPNR			
				GCCAGTGCTGC		ALRLALHLQPDSFAKAF			
				TTCAGATGCGG		TKCLEEGVFPSCWKIQK			
				GACCGCGGGAC		LLLLPKPGKPPEEPSSFRP			
				ACAAAGCCGCC		ICLIDGTGKLLEKLVCIR			
				GAGTGCCCTAA		LERAIADAGDLSRSQFGF			
				CGAAGCTAAGT		RKARSTVDAVNRVVEV			
				GCTTCTTGTGCG		AAQAIDGTRWKGGSKE			
				CAAGCAGAGGT		YCLMVTLDIRNAFNTAR			
				AGCCAAGCGAC		WDRILEALIGFSVPAYLV			
				CAACCACCAAA		RIIRSYFSDRVLLCESSEG			
				TGGGCACCCGG		VFRHQVTGGVPQGSVLG			
				AAGTGCCCATT		PLLWNTMYDGILRLPLV			

				GGCGGGTAAAG GAGCACCAAAG GCACTGCA		GRSEIVGFADDVALLVV DKYPGKAEESCNQNIRA IEHWLSSMGLELAPEKT EAVLISSRKAVETATVE VGCTKVSSSRAIKYLGV MIDTRLSFREHLAYANS KAAGVNRALSNIMLNSR GPKQASRRLLTSVTRAT MLYAAPVWAKALGTES YAQGLNATHRLSALRVC CAFRTVSNAAALVIAGIP PLDLLAKENEVVFNQTH GRRLSPGVKKAIRAEAR QHTLETWQHRWTNEPK GRWTQQLIPNLKPWLER KHGDITFHLTQLISGHGC FRSYLMRFGHEDADDCP WCGGGRSETAEHVLFSC VKYARERSSLETVLGGR LNADNLVPFMLQGEAE WQAVNSFAAAITTELRR AERARRVHE			

Jockey	Jockey-3_DYa	.	Drosophila yakuba	GAAACGTCAAA	TAACAAAMGTCGA	LRRLPHLPRPKPQNNYQ	cd01650	479	749
				AGGATAGATTT	CACAAACCGCTCCC	PNCPTHLSKSHRNRLYT			
				TCACAAAATTA	CAAAAATGCAGCTC	KCSHSPNPLKWHPCKYD			
				GGACAATWWT	ATTCTACAAAATAG	NCCPIYCPPVYKWTHEN			
				GGACGCAAAAA	GAATAAACAGGCT	FHKLFSHFETLLDGKSNF			
				GCAACAATAAC	ACATGGAAGATTCT	ILSGDCNSKHRCWGNYS			
				GAACGGCAGCC	TAAAATCTACACTA	SCARGRSLMQAIMANKL			
				CCCCCCTTTCAA	AAGCCATGGATGCG	LDVVATGQATHFPFDRR			
				CAAGACCCCCC	GGTTATTTCACTGA	KQPSALDFAICKGLRRRI			
				ACCTACCATAT	AACAATAATACTTA	TRTFSSDDLSSDHLPIQII			
				AGTTCCGAAAT	AGCGCCAGCAAAG	LDLEEANLQLKKVSNSIL			
				GGCCAAATCTC	TGTCTTCATCTCTAT	KRNTNMEKFKRILEDKM			
				GGAAGGAAGCC	CCTGATGACCTGCA	LLNTEIRTGQDIDDCIDIF			
				GCCAGGAAAAC	AAACCCGGACGAA	LNNIKTAANEATPKNRP			
				TACATCAAACC	GTATCCCTCCGACG	FTNPRENSRRRENNLSL			
				CGTTCCGGCAA	ACGCATCCCTGCTG	DEATTRLLDEKKELNRIL			
				AATGCGAAGGA	ACCCCGACGAAGC	RAAGTEEAKKWFKSAQ			
				CGATAGGCCAA	GCAAGTGAAGGAT	NKLAKEIRKLRTSQINNE			
				ACCAAACGGGA	GTCATCCACTAGGA	LQLIDTTDRYRMQKLW			
				TTCTCTTCACAT	CTGGCAAAGAAGT	KWTNNVKKQPQPIWPL			
				GAAGCTGCCAT	ATCTAGAAGCAGA	KKDDDCISMRTASPWTR			
				CCAAAAACGGC	GGGACACGTGAAG	TLDERAEAFASHLESRF			
				TGAATAAATGG	GATCAACGTGCGGC	KANLRNSIEDREIIKKEL			

				CGCCGACACGC	AAAAACATGATGA	DKFKSLKPRIETGTSGLK			
				GGACAAAGCGA	ATCTAAGTAGACTC	PVTLAELDLLIKSLESKK			
				CCACAAAAAAC	ATTGCAGCAGCGCA	APGADNISNKVISNLPKK			
				AGGACCAACTC	TGCATAATGTCACT	ARIYLILIYNKILNTGHFP			
				GAAGGACAACC	TACCTGAACCTTTT	KKWKHACITMILKPGKS			
				CCAAGCCGCGA	GCTGCAAATCCCTT	PSLLNSYRPISLLSGLSKI			
				CCCACAAAAAC	GAAATGTCCTGAAA	FERIILNRLYNIESFSKAIP			
				GACAAAAAACC	TAACAGCTGCAATC	NHQFGFRKGHGAEHQL			
				CGGCACAAAGA	TCCCCGAAAACCGT	ARVTQFILKAYEDKEVC			
				CACCTTGAGGC	ACTCAATGACAAAA	SATFLDITEAFDRVWHD			
				CTTCAACGACA	GACGACCCGACTTG	GLLLKLTKLIPKQLFNVL			
				AACTCCGCAAG	TCGCAATAGATAAA	ESYLTGRTFQVRINNETC			
				GAAAGGAGGGA	CTAACTTTTTGGCA	SRIGTINAGVPQGSILGP			
				TGCTATAGAAC	CTTAGAACACAAAA	VLYSIYTSDMPFPIIKGYI			
				ACACCCGCGAC	AACATACAAGGTTT	RNISLPDYHPKNIILATY			
				GAAGCTATGAC	TGATCCAGTCTCTT	ADDTIIISRSNNTRSAINI			
				AAGCCCCAGGA	TATTGGGATTCCCC	NQTYINKFCRWAKKWD			
				GGGGTGATGCT	CACGTCCTCACAGA	IAINGKKTGHVLFSLKNI			
				GGTAGCCCGCC	CCCTAGGCGTGTTG	SENVHTTPIVDGQRAIKL			
				GACAACAAGTA	CTGCCGACCCGGCA	CKHGYLGVMLDRRLTF			
				GAGCCGCCAAA	AAAGATCGAAATC	GAHLILLKGKIIAAYKRL			
				ACTAATTTCAA	GTAACCCCTGAAAT	EWLLGKNSHLPKRTKLL			
				GACAAGGACCT	TGACCTTTTTTTCC	LWKQIISPIWHYAIAIWG			

				AAAAATTAACC	CTTAGAATTAATAT	SLTSATQTKKIQTLENKL			
				GCCGATGGGAC	CAACGAAATCTTTG	IRRIIKANRYTKQSAIRE			
				CGAGGACGTGC	TTAAAGATCGAAAT	KYKIKSFEEVFDQASKR			
				ATAAGCACTTG	AATTTAATCAAAAC	YANSIADHENPLIKELLT			
				CAAAGAGGACC	ATAAAACAGAGTAT	KSFKPIRLAASRTRYVN			
				CCCTGACGACC	TATATAATAATATA	QYTKYLLPLLQNIXKPP			
				ACAGCCTAAAC	TAAWAAAAAATTC	DEPTYKSLYSYTDKEEL			
				CTGGCGCAACG	ACAATGTTTTGAAG	EIVSKIRYDYRQTLPTLL			
				AGTACTGGAAA	CCGTCTCAACGTCA	QLEDEEMEARDCKAEEE			
				TGGAAAAGAAA	AGAAGCCCACTCGA	RKEREKEERKKALEKGP			
				ATTGACACACTT	TTTTATTGACCCTT	PDRWCELLINKFSKMYK			
				ATGGAACTCTTC	GGCGCGGCMCTGC	KGMITRERVIELMLGQP			
				TTCAAATGCTTA	CGACCCGACAAAA	PTVIKLVIPDYEP			
				GAAAATAACAA	GGTCAAAATCGTGA				
				AAACCTGACCA	CCCTTGAATTTCGA				
				CAGCACACCCC	CTTTTTCCTGGTAG				
				ATGTCCTCCTAA	CAACGATTATACTG				
				ACTTCTAAGTTC	ACAATATCCCTGAA				
				CTCGCAAAAAT	AAAGAATAACAGA				
				AAGTAGCACCA	TCAATAAAAGCTTA				
				TGACATTAAAC	AAACTACGCAACCT				
				ATAAACAACAC	ACTACTCCCCTACT				
				ACATACTATCA	AGGCGCAACATAA				

				ATACCCTTAAA	AACGAAGATTTGGC				
				ATAGGGTATTG	ACCTAAAACAGTAA				
				GAACTCATGTG	AAACCTAAAATGTT				
				GAGTTACCAAT	TTGATGTCGTTTCC				
				AAAAGAGACGA	ATGWTGGGATGCC				
				ACTAGAAGCTT	CCCTCGTCCTTCTG				
				ACATAAAAAAA	AACCCTAGGTGCGC				
				GAGGGTGTCCA	CGCTGCTGACCCGG				
				GATCATGCTGG	CAAAGGGTCAATCT				
				TAASAGAAACC	CGCGACCCTTGAAT				
				AAGCTAGAACG	TCGACTTTTTCCTG				
				AAATTCTACTAT	GCAGCAACGGTAA				
				GCTTAA	ACAGACAAAATCCC				
					TGAAAAATTATAAA				
					ACAAATTATAAAAC				
					AAATTATAAAAGCA				
					TAAAACCACGACTC				
					GGCGAAATGAAWA				
					ACGACGATTTTGGC				
					ACCTAAAACATTAA				
					AAACCTAAAATGTT				
					TTKATGCCGTTTCC				
					ATGCTGGGATGCCC				

					CCCCGCCCTTMTGA ATTCCAAGCGCGGC GCTGCTGACCCGGC AAAAGGCCAAAAT CGTGACCCTTGAAT TCGACTTTTCCTGG TAGCAACGGCATAC TGACAATATCCCTG AAAAAGAACAATA AATTAA				
R4	EhRLE2	AB097128	Entamoeba histolytica	AGTAATACAAA GGAAGATAAAT AAACTAAAAGA CATGCTAATAA GAGGAAACTAT AAAAAATTAAG AAGGATATTAA GAATAAGAAGA ATGATAGAAGA AAAGATAAGGG GAGAAGTAAAT GAAGAAATAAA ACAATTAGAAT	TGAACAATAAGTTG AAAGGTATTAAACC TTAATTTAATATTT ATTCTTTATTCTT	MSTTGKKEIRKEIGREEQ PSEEDTKEFLNKIYNRGK RIKKENERIKEIIEESNRR NLKRIIITKENIENVITRIP SWKAPGIENMHGYYWK RLESVRDRLVIVFNEWL DTPENIPIEMLTGRTILIH KGGDSKKAENYRPITCL NVIMKIFTSIINRKIEEQL ANNEEKYQINKSQYGNK KRSLAAKEVNINSLLER NEREVQKKRYLESFYDI RKAYDTVNDEWLENVL	cd01650	120	376

				ATTTAAAAGAA ATGAGGAAAAA GAAGTAGACTT AAACAAAAGTG AAAGAACAAGA ATCTACAGAAA CAAAATC		RYFKIPFQITELIKSMSSR WRIRIGYRYKEEIREVKI KNGILQGDSISPLLFILQ MNIISDVIDRTNEKRLKV KHVLYMDDIKVTSETKK GMEMVHKNIIETIEVIGM EVNVNKSGVMKKGNIGI QDNMKDIPIVDSEHQYK YLGVYQFTRNDDVGCIE RISKSIEEGIKENNKGND SSMNYINRINTEIIPKFTY SASVVKLPIMKLIEIDKM IRIEVKEGKIIGAGMCTS RLYVNRNKMGDGLKSIR DEYAIKLIRSILYYTWKR GNRIGDMIGINENKNRM LNRLYKSMLKKRINKEE FRKIVEEERGKRDKNRII KYVREKFQEDYINKWR NKKVNSEILKGFEDKTM EQKKTIKVWKGINVIRK AYIQIKKMEEKARNVGV RKALMKHDMRYAICPIC			

						NKIATINHILLNCIVTKK TQMNKHDRIGEYIWRSL KIKYQLYENKEKMKMR EEENTKKLIIRKEELGKN ERRIEKDKILIVWNHEVV NRSEDKFHKRIDIYVRDL NKKEAMIIDMTIVRDKNI SKAFTDKINMYSKLHEH IRKIERLVRVIVIPVVITV SGLINKETVILLNAEGIEI PREIMTREIVISNMKDLM RYCGDHSGEYSNVEIPE EEDSQESEGREEILT			
RTE	RTE-2_EL	.	Esox lucius	CCTCTCCGGGA GCCATCACCTTA TCGTGGTGGAG AGGTTTGTGTGT TCCTATGAACCT GAGGGCTGTGT TGTCCGGGCCTT AGTGCTCCTGGT AGGGTTTCCCAT GGCAAAGTGGT	GC	MVQNNPMYKRGRGGVT LPGGSPGPPSGARPRGR ARRRAPGGRVCHGARS GIARKSNVAPPSPSYGPT THGRNRWGRVRCHMG GSEGQGPRRTRPGRQGL ALGTWNVTSLWGKEPE LVREVERYQLDLVGLTS THSLGSGTVLLDRGWTL FFSGVAQGVRRRAGVGI	cd01650	558	813

				CTCAGGGGAGG AGCCAGACTAA GA		LTSPRLSAAMLEFTPVDE RVASLRLRVVGGGISDC CLCICPEQQFGVFGLLGD PEWSPVWGSSRGLHNSA GGLQRTRGQFCDRVIGS ETACFGHSGKERGGAVN RSPSGGELGQGVGEDSG QTGKTQTGSAGELGTSG GGPRPKDLHSHLRRSFS GIPVEVGGIEPEWSMFK ASIAEAAAGSCGLKVLG ASRGGNPRTPWWTPVV REAVRLKKEAFRDMLSR RTPETVAVYRQTRRAAT SAVKEAKQRVWEEFGE AMEKDFRSAPRCFWKT VRHLRRGKRGTIQAVYS KDGTLLTSTEEVIGRWK EHFEELLNPTNTPSIVEA ELEADGEASSISMAEVTE VVKQLHSGKAPGIDEIRP EMLKALGVEGMSWMTR LFNIAWESGKVPKEWRT			

						GVVVPLFKKGDQRVCA NYRGITLLSLPGKVYSK VLERRVRQIVEPQIEEEQ CGFRPGRGTTDQLFTLA RILEGAWEYAHPVYMCF VDLEKAYDRVPREILWE VLREYGVRRSLLRAIQSL YVQSESCVRVLGSKSDS FQVGVGLRQGCALSPILF VTFMDRISRRSRGGEGL QFGGLGISSLLFADDVVL MASSVCDLQHSLDRFAA ECEAVGMRISTSKSEAM VLSRKPMECLLRVGNEA LPQVKEFKYLGVLFASE GTMEREIGRRIGAAGAV LHSLYRTVVTKRELSRK AKLSIYRSIFLPTLTYGH EGWVMTERTRSRVQAA EMGFLRRVAGFSLRDRV RSSAIREELGVEPLLLCIE RSQLRWFGHLVRMPPG RLPREVFQARPAGRRPR			

						GRPRTRWRDYISTLAWE RLGIPQSELANV			
CRE	CRE-3_FoCa	.	Folsomia candida	AATTTTTCAAAAAAAAAAATAAAACGCCCAAAAATATAGAGAATGAA	TAAAGAATTTCATTTATATATCTATTGTAGTTATGCTAATATTATGGATTACAATTATGTATTATGATAAAGTA	MMTPTSTTAANLSTHDDQHCPISTCPGASNTDQWILCDHCQKWYHTACVAVVIIPEEEWLCPTCVISVTSSQSQVTPSSLNTTTTTKKKHIIATTQYWLKSNLVEDPLGLIPKSEIVARYAHHIQTSGEGRVVDATVGKYILQLFPSTKFSRPPTRGAANAYIGIRWKDQPPLLPHAPRNWGAELSQLKNTVPTLRRVPKGARMEVAKALAKIIEEVTTTNSEESWFKLFSFPYAVVPVPTKSDKVKNLTTWVKSRVLQWSITQIVTPSRPRPQFAKPSRPQETIAKKVEAKLADGDVRGAIRLLTSDDTIAPNNQETLDALITKHPPHPEPTYFPNAPDKAEVLAPVDPKEVLIAISSFAPG	IPR000477	404	656

						SAGGLDSLRPQILKDLVI SENGEPATNLLNAMLKF LDLVLRGDVPKSICPIFF GASMTALRKKTGGIRPV AVGNTWRRLISKIAVAR MSSTLVEIFSPHQLGVGI KGGAEAAAHAARIYWN YSHDSPKAFLKLDFRNA FNEIRRDTLLQIIREKFPN FYRFVSQAYSSPSEIFFA GTPISSLCGVQQGDPLGP ALYALVIQPIVDLIDTEL NMWYLDDSTIADSPEKV LMAMQTITRMGTEVGL HLNTTKCEVGILGVVDD QTRDEIFELFRSTAPGIQ KISPESAMLLGVPLTDEA IPTILGAKTEEMEKISTRL LKLSSHSAFFLLRASISTP RLIYFLRCAPTWRRFDA LTLYDAKLKNAMEGIIN CSLSPQSWLQSSLPVNK GGLGVRHAIDVAIPCFL			

						ASIYSSLPLAERLLPPHL HGIDTAIQEGEERWKSL GSSPLPPLEIRGIQAIWEI PQQEGVINHLTTNATTL EDKARYHAMSQPNVGA WLNAIPSPQLGTHLPNET FRIAAALRLGCDICQPHK CSCGEQVSTKGYHGLSC KNSAGRHSRHAAANDVI ARALRSAGVPIIKEPAGC AREDGKRPDGLTLIPWA RGKSVVWDFTCSDTFAP SYLPQTSIRPGVAAEKAE DRKKKRYEQLMDSFLFV PIAIETTGIWGKDGLSFIK EIGSRITNITEEKRATSFL LQRMSITVQRGNVASIL GSLPKGHKLEEIFML			
Rex 1	REX 1-1_GA	.	Gasterosteus aculeatus	CTTCCGTGGAAT TACTGGACATTC TAGTCAAAGGT GCGCTCACCTTT CGTCACGCGGT	TGACTATAACACTC CACCGGTCACTCCC CCTCATACTGCACA TGCCACTTTAACTG CAATTCATCACTTT	MRRDRDFSTSSVLCFTE TWLCGSIPDSALQLAGF QLLRADRDTELSGKTKG GGICFYFNNSWCNDVTV ILQHCSPDLETFFINCKPF	cd01650	472	725

				GAGACGCCGTA GGAGAGGAAAA CGGGCCGGCTC GCTGGTGAGAC TCCGCAGGCGT GGTTCTCGCACT CCGCTGCCAGG CATCTTTCTCTC CAACGTGCGCT CACTGTGCAAC AAACTGGACGA AATTCAGCTGCT G	GTCGTCACTCGTCA CTTTGTCTCTTGTCT GTTACTTGTTCGTT AGTGCACTTTATGC TTAATATTTTCTTT TTAACTTTTTAATA TTTAAAATTTTTTA ACTTTATTCCCTTGT TTTATACTAACCCA TAGCCTTAGCCTTA TTCTACTAACCCAT TGCATTAGCATTTC ATTCCACTTTATTTT ATTACTTGTGCACT GTTGTCTTGTCTGT CTACTGTCGCGCAC TAACCGCCAAGACA AATTCCTTGTATGT TTGACATATTTTGG CTAATAAATGTTTC CTGATTCCTGATTC CTGA	YSPREFASFILVGVYMPP AGNVHEAQRTLADQIRR VERTNPDSLVIVLGDFN KGNLTHELPKYRQFIKC PTREENTLDHCYTTVSR AYHAVPRAALGHSDHV MVHLIPAYRQKLKLCKP AVREFKKWTSEALEDLR ACFDCTDWDVFRTATDS LDEFTEAVTSYIGFCEDS CVPSCTRVSYNNDKPWF TAELRTLRLQKDQAFRS GDKDLYTEAKYKFSKA VRDAKRLYSEKLQQQFS ASDSASVWRGLKHLTN YKPKTPHSMNDLRLADE LNEFYCRFERQCPDPIPH SSTNLLQSPSPPSPSPSGA HASSSISPSPPPPATTLSIL ERDVNRLFRRLNPRKAA GPDSVSPHSLKHCADQL SPVFTDIFNTSLETCHVP ACFKASTIIPVPKKPRITG			

						LNDYRPVALTSVVMKSF ERLVLTHLKSLTDPLLDP LQFAYRANRSVDDAVN MALHYILQHLDSPGTYA RILFVDFSSAFNTIIPSLL QDKLSQLHVPDSTCKWI TDFLSDRKQHVKLGKH VSASRTISTGSPQGCVLS PLLFSLYTNSCTSSHQSV KLLKFADDTTLIGLISGG DESAYRWESDHLVSWC SQNNLELNALKTVEMV VDFRRNRAPPSPITLCDS PVTIVDSFRFLGSIITQDL KWELNISSITKKAQQRLF FLRQLKKFNLPKTMMV HFYTAIIESILCSSITVWY AAATAKDKGRLQRVIRS AERVIGCNLPSLQDLFAS RSLKRAKKIAADPSHPG QKLFVPLPSGRRLRSIRT KTSRHTNSFFPSAVGLIN RARSPTA			

RTE	RTE-1_GA	.	Gasterosteus aculeatus	CCTCTCTGGGA ACCGTCACCTTA TCGTGGTGGAG AGGTTTGTGTGT CCCTATGAACCT GAGGGCTGTGT TGTCGGGAGCC TTGTGCTCCTGG TAGGGTTACCCT TGGCAAAGTGG TCTCAGGCGAG GGGCCAGACTA AGAATGGTTCA AGACCCCC	TAAGCGGTTGAAGA TGGATGGATGGATG GATGGATGGAT	MKKRKREGVTRPGGSP GPPFGARPRWRARQRAS GGRVYRGARPGTARRS DAAAQHSTPHGLTTCGG NRWGRVRCYKGGSDSG GSRRNRPGRQMLALGT WNVTSLGGKEPELVREV ERYQLDLVGLTSTHSLG SGTKLLDRGWTLFYSGV AHGVRRQAGVGILISPRL SAATLEFTPVDKRVASL RLRVVGGKTLTVVCAY APNSSSEYSAFLETLNGV LHGAPVGDSIVLLGDFN AHVGDDGDNWRGVIGR NGLPDLNPSGCLLLDFC ASHGLSITNTMFEHKDV HKCTWYQSTLGQRSMID FVIVSSDLRPHVLDTRVK RGAELSTDHHLVVSWV RGWGKTRDRPGKPKRV VRVNWERLEEAPVQKIF NSHLRRSFSHIPVEVGDI	cd01650	616	870

						EPEWSMFKTSIAEAAAV SCGLKVLGASRGGNPRT PWWTPVVREAVRLKKE AFRVMISGGTPEAVAVY RQARRAAASAVMEAKQ RVWEEFGVTMEKDFRS APKCFWKTIRHLRRGKR GTIQAVYSKDGTLLTAT EEVIGRWKEHFEELLNP TTTPSLVEAELEAEEGSS SISLVEVTEVVKQLRSGK APGIDEIRPEMLKAMGV GGLSWMTRLFNIAWKS GTVPKEWQIGVVVPLFK KGDQRVCANYRGITLLS LPGKVYSKVLERRVRPI VEPRIEEEQCGFRPGRGT TDQLFTLSRIIEGAWEYA QPVYMCFVDLEKAYDR VPREILWEVLREYGVRG SLLGAIQSLYAQSESCVR VLGSKSKAFPVGVGLRQ GCALSPILFVVFMDRISR			

						RSRGEEGLQFGGLRISSL LFADDVVLMASSVCDL QLSLERFAVECEAVGMR ISTSKSEAMVLSRKPMD CLLQVGNVSLPQVKEFK YLGVLFTSEGKMECEFG RRIGAAGAVLHSLYRTV VTKRELSRKAKLSIYRSI FVPTLTYGHEGWVMTE RTRSRVQAAEMGFLRRV AGVSLRDRVRSSAIREEL GVEPLLLCVERSQLRWF GHLVRMPPGRLPREVFQ ARPAGKRPRGRPRTRWR DYISALAWERLGIPQSEL VDVAREREVWGPLLELL PPRPDPG			
CR1	CR1-1_Gav	.	Gavialis gangeticus	AGCTTAGTTGCC TGTACACTAAC KCCAKAAGCAT GSGAAACAAAC GTACTGACCAA CAAGCAACACG	TGATCCCAGCAGTC TTTCCTGCCCACGT GCAGGGGGGCTGG ACCTGATGATCTGT AAGGTCCCTTCCAG	LDXRHXGIXXVQEGSY WKXGGVALYVKEHYTS LXVKIGCKEGXLETLWV KIRGERREEXLIVGVYY RPLTKNSLENWQRLHGI MGXFNYPDISWDEYSAR	cd01650	521	770

				ACCAGCTCCAC CTMTGA	CCCCTAACAACTAT GAAA	SDRSQSFLTCIEELYLTQ EVYEPTRGKALLDLVLA KGDDLVSNLRIXGKLGN SDHEMITFSIHCKPGKSA SRVEIPDFRKADFHKLRR LVGXGLRSNEEGNTRRV VFPQGSNPKSSKKVYSL SQERQQKSWETPLAQQG SHGPPEIKARSLQGMEV RNNYQRGVLSSCPHLKS LILGKLISTNSGDWLVRD YGPMRKGIQEEWSFLKE AILKAQRKSIPSHKKGSK RAGRPPWLNREVMDLL KLKREAYKGWKSGTTT KEEYSAAARTCREQIRT AKALTEHKLATRIKDNK KSFFRYVGSCKRNKGSI GPLINEMGRLLTDTQDK TELLNEYFMSVFHQSKG ISLLDRIQDDHGGGDHPP IVDVDLVGEHLQRLNIY KSAGPDELHPRVLKELA			

						DVIAQPMAKLFEDSWHS GEVPEDWKKANVVPIFK KGRREDPGNYRPISLTSI PGKILEKIIKETITHRLTE GSILKGSQHGFIMGRSCQ TNLISFYDQVTCHLDRG NEVDVIYLDFKKAFDLV SHDVLMVKLGKCGLDY FMVWWLGNWLCGRTQ RVLIDGTESSWQRVTSGI PQGSVLGPVLFNMFIND LDSGITSNLAKFADDTK LWGSVVTLKDRLAIQAD LDRLMKWADLNLMSFN TEKCKVLHLGRNNPQHT YRLGSALLASTTTERDL GVMTDNKMNMSHQCD AVAGRANQTLACIHRGI LSKAKDVILPLYSALVRP QLEYCVQFWVLHFKKD VEKLKRVQRRATCMIRG LETKTYEERLRDLGLFSL EKRRLRGDLVAAYKYIR			

						GVHQGLGEHLFTKAPQG RTRNNGHKLIEDRFRLDI RKNFFTSRVPRVWNRLP PEVVQSSTLAIFKKRLDA HLAGII			
CR1	CR1-7_Gav	.	Gavialis gangeticus	CC	TGACCCCAGTGCTC TTCCTGCCATGGCA GGGGGTCGGACTA GATGATCTGCTCAG GTCCCTTCCGACCC TAGAAACTATGAAA CTATGAAA	LVRGLRVLDPLGDSDHC LLEFTIQRKVAKASSQV VALDFRRANFNELRRIV GEALKTQRGGEPGVQGE WSLLKETILQAQRKVIPR RVKGGRRAYKPPWLTK NIRECLLAKKEAYSQWK GGVITKEDYSAVARGCR GMVRKAKAEMEQGLAS RIKDNKKSFFKYIGSRKK APGNVGPLQDKTGNLV VTPDDKANIFNNFFASVF LSKDQNCHLPAGPPLGP GGGSPGPRVSEDLVREL LEGLDVFKSAGPDDLHP RVLKELAGVIASPLARL YEHSWCSGEVPEDWKR ANVVPIFKKGRKEDPGN	cd01650	288	539

						YRPVSLTSILGKLFEKIIL AHIQEGPAGNGMLRGN QHGFISGRSCQTNLVAF YDQVTKSLDAGVAVDIV FLDFRKAFDTVSHPILIR KLGDCGVDTHTVQWVA NWLESRTQRVVVDGSFS TWKDVGSGVPQGSVLG PALFNIFISDLDEGVQST LFKFADDTKMWGDVGT IEGSNRLQLDLDRLQGW AEENRMGFNTDKCKVM HLGRNNRHHTYRLGNSP LVSTEVEKDLGVIIDSKM NMSRQCGEAARKANCT LSCIHRCISSRSKEVILPL YAALVRPQLEYCVQFW APHFKRDVDSIERVQRR ATRMIKGQQGRSYEERL RDLNLFSLQKRRLRGDL VACYKLVRGDQQALGE SLFPRAPQGVTRNNGHK LAEGRFKLDIRRRYFTV			

						RAARVWNQLPSEVVQA PTLGVFKRRLDDHLTGV I			
CR1	CR1-17_Gav	.	Gavialis gangeticus	ATACAGGACCT CCACCTAACTC AGGAGGTGCAC AGTCCCACCAG GGGAAATGCCC TGCTGGACCTG GTCCTTGCCACA GGCGATGACCT GGTGAGGGAAT TGCGGGTTCTCG GCCACCTGGGC GATAGCGATCA TCGCCTGCTGG AATTCACTATCC AGCGCAGGGTG TCAAGGGCTAG CAGTAAGGCGG AAGTCCTTGACT TCAGGAGGGCT GACTTCAATGA	TGACCCCAGCATCC ATTCCTGCCCAAGG CAGGGGGTCGGACT TGATGATCTGTTTA GGTCCCTTCCGACC CTAGCTACTATGAT ACTATGA	MGVQDGWSFLKETILRA QRESVPMHTKGGKSAK KPPWLSKGIQECLRAKK EAYIRWKQGAITKEDHT SVARGCRAAVRKAKAE MELGLVTRIKDNKKSFF KYIGSKKAAPGNVGPLQ DALGNLVIAANEKADLF NKFFASIFLSRDRDISPTG STEGLRGGTARPKVRDD VVRELLEGLDVFKSAGP DALHPRVLRELAGVIAG PLARLYEHSWSSGQVPD DWKKANVVPIYKKGRK EDPGNYRPVSLTSVLGKI FEKIIKEHICEGPAGEIML RGNQHGFIKGRSCQTNLI SFYDQVTKSLDEGVGVD VIFLDFKKAFDTVSHPILI KKLGDCGIDAYTVRWV	cd01650	225	476

				GTTAAGAAGAC TAGTATGAGAG GCATTGGGATC CCGAAGTTTTG GGGAG		ANWLDGRTQRVVVDGS FSTWRDVGSGVPQGSVL GPVLFNIFINDLDEGVES TLSKFADDTKMWGEVG TPEGRDRIQLDLDRLQR WADENRMEFNADKCRV LHLGRRNQQHTYRLGN TLLISTEAERDLGVIVDS RLNMSRQCEEAVSKAN RTLSCIHRCITSRSREVIL PLYAALVRPQLEYCVQF WAPHFKRDVASIERVQR RATRMIRGQQGRPYEER LEGLNLFSLHKRRLRGD LVAVYKFTRGDQQGIGD ALFPRAPHGVTRNNGHK LIESRFRLDIRRHYFMVR VARTWNQLPREVVLAPT LGVFKRRLDRHLTGVI			
L1	L1-5_GAr	.	Gossypium arboreum	GACTTCGTTTGT GTGTCTGCCTAT TATTCTTTGAAT TTTATGGCAGTC	TAGACCGAACGGTC TTGTAATAATGTAG CCCATTTAATTTGT AAGACTATGTGTAG	MKFMCWNCRGLGNPAK LRELKQLIAANNPDLIFL SETKMSANDFRRVQNK CRMQNGLAVNSEGRSG	cd01650	498	764

				TTGTGGTGTTGA	TGAGGCTAAAAATT	GLALMWRDGMNVSIQN			
				GTGTTTCTTTGT	GTTTTTTTGATATG	YSKHHIDSMVKLENNDN			
				TGTTTTTGGTAT	GATTATCCTTCGGA	IRVTGFYGHANPNLRSSS			
				TATTTTATTTTT	AATCCATGATGCTG	WDILRRVGESVREEWIV			
				TTTTGTGTCTTT	TTATTAGTGATGTT	GGDFNAILNDAEKEGGR			
				CTTCGGTGGGTT	ACTTAATAATATGT	RGVRAQMNEFKEVMDD			
				TGCGGTGCCGG	GTTGACCCTTGTGG	LALVDIKPDSGWFTWVN			
				GGGACCGACGT	TCGTTTTTGCTCAA	NRRGGVTWGGLIKERID			
				GACATGTTGCT	AAAAAA	RFISSVSVAEKFPFISTKV			
				GTTGAAGTCTG		VRQTQSDHDAIILDLWG			
				GCGTTAGGTGG		RKPKAYPTDKRLCFKFD			
				GGATTTTCAGA		VCWAGDVEAKNVIERA			
				GGGTTTCTTGGC		WNRAATDFGEKIERVRL			
				TCTGTGACGTTC		ALGPWQRKKYGNLKRD			
				GTTGCTTTGCAT		MQKLEKHIDWIIDSVSSE			
				GGCCAAGGAGT		ESGRTLKETRRRLDFLY			
				TCTTGCGACTTT		AREESYWAQRSRSIWLR			
				ATGCGCTTTAGC		EGDRNTRYFHAKATGRL			
				GGTGGCGGGCC		KKNVIEKLQNAEGQWV			
				ATTCGAGGTGG		TNSEDINKVAKDYFVKL			
				GACGGGCGATA		FRSNGQNANTQEMGFIK			
				CATACGTGGCT		DCVTRETNEWLCMDYT			
				CGTCACGATCTT		EKEVLQAIKQMDPNKAP			

				TTCTAGCATCGT		GIDGLSGSFFKHHWDIV			
				CAGACCCTGTTT		GKDTISYCLEVLNGNKN			
				TGATGGGATTG		ISSVNETMIILIPKIKNPCE			
				TTTCATTGGCAG		LTNFRPISLCRFVYKIISK			
				CAATAAGGTTTT		VLANRLKAALPSCISQN			
				GTCGTTTCTTTT		QSAFVPGRMIHDNVLIA			
				GTCTTGTTCAGC		HELLHYLQSSKNGPNKG			
				TATTTTATTGAT		LVVKLDMSKAYDRVEW			
				TTTCTTTACGTT		NFIENVMKKMGFVDKW			
				CAGTAGGTTGC		IAKIMECVRSVTYTVKC			
				ATGGGTCTTGC		NNTLSDIFVPERGLRQG			
				ATGTATCTTGCA		DPLSPYLFLFCMEAFSR			
				TGGATTGTTTCG		MLIHAQENNTLRGIRAS			
				TCTTGCACGTAT		RNGPRINHLFFADDALLF			
				TCCACTGGTGGT		VRNKRSDIECLVTMLDIF			
				ATTTCTTATTTT		SNISGQEINFDKSMILFSP			
				CTTTGACTTTTC		NTPRAQRILFSDLLGMT			
				AACGTTATTCA		VVEKLNNYLGLPIPIGKK			
				ATCTGAGGGTTT		KTEAFKEITNRMSCRINS			
				TTTTGGTTTAAC		WTKRLLSFEGKEVFIKA			
				AGGCATTATAT		VLQSIPTYAMSIFLAPKG			
				GAATGCGGAGG		IIDDIQAMLSRTWWSGK			
				ATAAGTTTTAAT		DKGGFWTMIPWKTLCK			

				GATTTTCCCTAT		PKGMGGLGFRDVRLFNL			
				TTGTTCGATTTA		ALLGRQVWRLLNNTDS			
				GTTGTTTTTGTT		LCFKVLSSKYFPDGNIFH			
				ACAGTAGCTAT		AKKVDKASFTWSSIAKA			
				TAGGTTTCTGGG		VEAFQDGFGWQVGNGE			
				TTTTCTTGGACT		KINIWADNWGMEGLNG			
				GCTGGTGTTAG		AAVISNTLNHSEKFVKD			
				GTTGCAGTTTTT		LWHLDRRSWDENKVKQ			
				TGGTACTTTCTT		AYGQDWGDKICDIPIGD			
				GCATGTTTTCCT		EVQEDKIIWFHNTHGCF			
				TTCCTTGATGTT		TSKSAYSWLLLKEMGC			
				CGGGGATGGGG		GPHRFFWRALWKLDTLP			
				GTTTGATGGGG		KVRVFTWRVGHEILPTN			
				GGTGTTTGTTCG		GKIASIRPGFNKGCPRCG			
				TAGCTTGAGTC		AEVETLLHALRDCPTSR			
				GGGGTGTTCCT		AVLSIGGWSRSFISKSYD			
				GGCTGTTGTTGT		HCIDWLEDLMRVLDKR			
				TTTGCTGTGTTT		AMADLMTILWNCWNNR			
				TTGCAGGTTCTT		NNFIFRGKEEEAQLIWER			
				GGACTGTTTCTT		ASNLSKEFRICNMLHAP			
				TGCTCTGCTCCC		LISQNIAEKRWVKPPKGF			
				GGTGGCGGTTTT		LKINFDATVGETGIGYGT			
				CGGGGTTCTGTT		VIRDEEGFVLGGGGGFK			

				TTTCTCTATTCT TGGGGTGTTTCT TGTCGAGGTTG GGGGAGGCGAC AAGAGAAAAGT GTTGATCAGTCT CCCTGGCACTC GTTTAAGGTAG CAGTCTCGTGA GTTGTCGTTTGA AA		EGRRSVEEAECVAFEQSI DVACKLNIKDHVLFETD HAGLVNRINNLANDVTII GARLKDCTVAFTFFKSA SLI			
R1	R1-1_HSal	.	Harpegnathos saltator	CGCCGCGCTCK GCGGCCGTCAC CATAACGTGCC CCGACGGAAAA TACCATGAGTTC ATGGTCGAGGC CAGGCGCCGAA TCAGACCGGCG GACCTSGGGAT AAACTCTTCCCT GAGGGTCCGCC GGGCCCTGAAC	TAGCAGGGAGGGG GGTAAGGAGGGAG GGTGGATGTCCCCT CCCTCCTTGACAGG AGACACCAGAAGG CCAACCGGCGGGG CAACCCGCCGGGG ACCGCGCCCTGGAT TCCCTTCGTGGAAC AGGGCGCGGTACA CAGGTGGCGAGGG GCTCCCGGAAGGG	MQRMSELGAGLGVVAE PYWVPVRSVNWFGSPD GSVAMVANPAAGTPPC VFRENGSSYVAVDWGPI TVVGVYLPHHRNKAGL GELAPVKRVLDRIGELV RRCHPRPVVIAGDFNAH SEGWGCSPRQRDPRGEA VIGWAAGLGLLLMNRG STSTCVRPGGQSVIDLT WATPSAAGLFQEWRVE AEGESLSDHRHIAWALR	cd01650	474	707

				GGGGCAATGCT	AGCCCTGGGGCCGG	LPKPARVLARCPGEGNF			
				GATAGAGGTGC	GGGGTTGGGGAGG	PLPKRWALNRLDTDRFR			
				CTGGCCTCAAC	GAATACGGGAGTA	AAVLAATWHLDVGPGE			
				GTCGGGGCGTC	GTTCCCCTGACTCC	WDPSPAEGAGEIVGIVA			
				CGCCGACCGGC	CCTCCTCACGGTCC	SACDAAMPQSHPRRRRA			
				TTGCGGAGGAG	GACAGGTAGACATC	TYWWTEELENLRDSSV			
				CTCCGCAAGCT	CCGCAGGGGTTTTA	QARRRYYRARRRGGDV			
				GGCGGCGGAAA	GTCGGTAAGAGTCC	AGTKEGYLRAREALKA			
				AAGGCCCCGGT	GACATAACCCCCTC	AIVEAKRAAWEQMVSS			
				TATCGCATCCA	CCCCCCAAGGGGG	LDDDPWGRPYKRVMGK			
				GCGGCCGGCAA	GGGGTATCCATGAT	IRPWAPPPTESLDRQALE			
				AGTTCGCAGCA	GGATTTCCCCTGCG	GLLDALFPRAEGGPPEIP			
				CTGCGAATTGCT	TGACCAAAAAAAA	TPPLGEGDWDEGLAVSP			
				GGGCTGGATGC		EELTAARRRLGAKGKAP			
				GACGACCGGGG		GPDGIPGRAWALALGEG			
				CCGAAGAGGTG		NLSAATCRVLSGCLREG			
				GCGGCCGCGGT		VFPPEWRRAKLVLLPKE			
				GGCCCTCGCTG		GKAPGSASAYRPICLLDE			
				GGGGATGCTCG		AGKMLERIVADRLVQHL			
				CCTGAGGAAGT		SSKGPNLHDRQYGFRPG			
				TTCCTCGGGCGT		RSTLDAIQCFRDLTRAV			
				GATGAGCGTCT		VEEEGGVLLAVSLDITN			
				CCCAGCGGGGA		AFNTLPWPEIGRALEHH			

				GTCGGCGCCAT		GVPVYLRRILSAYFGGR			
				GGTGGTGCGGT		DLAFPVEGGVQGRRAM			
				GTCCGCTGGCG		ERGVPQGSVLGPLLWDI			
				GCGGCCGCTAA		AFDRVLRTPLPPGCHVV			
				GGTCGCCGCCG		CYADDTLVVAGGRAGG			
				CGGAGCGGGTT		KPASERRRHWHRWWEP			
				CCAGTGGGCTG		SETWVSRWRPKRPRPCS			
				GACCCGCGCCG		FTAGVGGRRRRPGSSPG			
				GGGTGACGGCG		GELGGSPPPSGDGIGIGG			
				CTGCCCGCTCGC		GNHQRPGSQGGAPKDP			
				GCAACGCTGTG		GRVLSRQGSGGAAGDP			
				CTTCCGCTGCTT		GPRRWRPRPRRADDEIS			
				AGAGCGGGGGC		RPDAGQPMELRTPLQGT			
				ACGTGCGCGAG		GPQGAQGRTGCSQPDA			
				CGGTGCACCAG		HPRGPRLEGSPPVRHGG			
				CGCCGTCGACC		ALDRPVRGAHMGATAG			
				GCAGCGGTATG		GLPRRKARAAAGPEARA			
				TGCTACCGCTGC		NQGDSGVPYRVVHGGD			
				GGTAACCCCGG		HPGRRPPCGAPRRGEAP			
				CCACCGCGCCA		PLLEGEGAPRGGEPVGQ			
				AGGGGTGCCAG		GPEGSEAPGRGPDPGEV			
				GAGGCGGTGCG		EGLAVRPPGARASDGRG			
				CTGCCCGGTGT		CPPLPVRVAGRRGRGLS			

				GCGCCGAGGCG		FRLAQVLTGHGCFGRYL			
				GGCAGGCCCTC		RRIGKELTTRCHHCPALE			
				CGGTCACAAGA		DTAHHTLAACPAWAGE			
				TCGGAGGACCT		RRVLAGAVGCREGEJSL			
				GCCTGCAAACC		PRMVKAMCGSEEAWAA			
				CCCTCCCGGAA		AVSFCEAVMSQKETAE			
				GGAGAAAGGGG		WGRRPHRARGAAPGGR			
				GGAAAGAAGAA		GTRQRARGGAGRGGLP			
				GGAAGGAGGAG		DPRGPGAPSAARARHRS			
				AGAGGCAGAAG		QGGSRGRRAAFPAFPPP			
				GGGAAGGAGGA		SDRTSRDRGERCPGWG			
				CACCGGCGGGG		WARSPTNPPLPD			
				GGAGGCCCGAA					
				CCACCACCGCC					
				GGGGGAGGACC					
				GGACCCGACGA					
				AGCCTGGGGAG					
				GAGASCGGACC					
				AGTCTCCTCCCC					
				GCCGAGGGACT					
				GAGGGGGAGCC					
				GGGGCGGATCC					
				GGCCCCCCCCT					

				GGCAAGAACCC					
				CAGAGGAAAAG					
				GGGAGGAGGGG					
				GACTCCTCCCCC					
				GACCTAATAGA					
				CTGGGACTGGG					
				AGCCGGAGAGC					
				CCCGATCCAGG					
				AGCCGGAGGAA					
				CCTCTGCCCCGC					
				AGGAAGAGGAG					
				AGGGGAAGAAG					
				AGGCCAACAGG					
				GACGAGGACGA					
				CCCGGAGGAGG					
				ASGCCCAGGCC					
				GCAAAGGTGAG					
				GAGGGTGGAGC					
				CCTCGGGCACA					
				CCCCCAACAGG					
				TGTAGAGGAAG					
				GGGAGTCGGTG					
				GGAGCGCTGCC					

				CCCGACTCCCCT GCAGGTTGACT CCCCCTCGTACC CCCGAGAGGAG GAGGTCACGCA GGTAGGGACCC CCCGTGTCTCCG CGGAGACACCC GCGTTGGAGGC CACCATGGCCG AAACCCCCTCC GGGGAGACCGA GACGGCTTGCC GTCGGAAGTCT CCCAGGAGGAG GGGGTGGACGT AGGCAGCCCGG AGGAGGCGTGT AAGGCTTAGCC TTCACGACACTC CGGGCCCTACA GGCCAACCTCG GCCACGGCGCC					

				CGGGCTCAGGA CCTCTGC					
Daphne	Daphne-14_HMa	.	Hydra vulgaris	GTGTATATAAA AACCTTCAAAT ATTTTCAAAATA ACAAAAACATA CAAA	TATTTCAAGTTATT TGATATTTTTGTCG ACTGTGTGATGAAT ATTATTTAAACAAT ATATGCTGAATACT TTATATTATAATAT TGTTCAATATTGAG AATGTTATATATCC ATATTATAAAATAT TATGTTAAAACTTG CTTAACATCTGAAG AATATTCTAACGCT TATGTAATTTTTAA TATTTGTAAATTTT CTACCTTTTAATTC GTTATATAGATTGT AAAGGGCTTCATGA TAAGATCATGTGAT CTTCTGGAAGTCCT ACCAGTAATATATT CAATGTAAATTTAC	MDSTYRNDFEKISNLFQ TNTFNLDNESDPDINYFN EINGSKFECSYFYPDKIK SIFKNNDNRDYFNIIHINI RSLKKNFDNFLNMICQT ENFFNVICLTETWCTEED FKSNSNLHLTGFNTVFLE RNVNKRGGGVLFYIKEN LAYNIRSDMSVSDDNIEI LTVEIINKKSKNILLSCC YRPPTGRTEYLSNYLVH NIIEKTSQEKKKNYILGD FNLDCFQYNEKKYITKF YNHLFETGTIPIINKPTRI TKSSSSLIDNVLTNDYFN ITLKKGIIKTDVSDHFPIF FCLGADHKTNTNGKITL KKRIYSNNNLNLFKEQL LHQDWSYINFNEDINIIY KSFFNIFYKIYEANFPIRE IVLNAKDISCPWITKGIK	cd01650	570	829

					CTTGTTCATATTTTT CTGGAAAAAAAAA	KSSKIKQKLYIKYLKNKS DKHKNNYTVYKNLFEK LRKKAKQIYYSDLLIKH KHNSRRVWQIMKEITSK HKTNTNTMPNAIKFENK LITDSNEIAAEFNNFFSSI GHNLSNKIPYVDNNSVD EFISSPNSTINFSNLTYNE FEIAFKSLKRNKAIGPDD INGNIVIDSFNEIKDILFK VFKASITQGSFPNSLKIA KVTPIFKTGDHTNITNYR PISVLPVFSKILERIMYNR IYTYLIDNKLLYKNQFGF QRNSSTEHAILQVTRSIA DSFKNSQFTLGIFIDLSK AFDTVNHHILIKKLESYG IKDKTLNWFKSYLSNRK QFVYNKESSSQIIKNITC GVPQGSILGPLLFLLYIN DLYKASTLTTVNFADDT NLFQSHENIEILFNNMNN ELKRISYWFRKNKLSLN			

						TDKTKFILFHPTSKKKRI QNILPDLFIDNTIIKREKF TKFLGVIIDENLSWSQQI DSISMKVAKNIGVLYKA RYILNKKQLTQLYYSFIH CHINYANIIWGSAHKTK LKSLYRHQKHAARIINF GNRFSSSTPLLIEMKALN IYELNIFNILCFMFKCKE NLSPSVFINLYNLKPKNK YELRDSKNIQEPFCKTKI DQFCISYRGPFLWNKIVL PNFDFSFKWSYSSFKKK TKEIIFSTQNIFIYF			
CR1	CR1-53_HM	.	Hydra vulgaris	CGCGCGCCGTT AAAGATGGCAG ACGTGACTTTGT GAAAAAATAAG AATAACAAAAC ATAAACATTTTG TAATAATAACT AAATAATAAAA GATAAACTTAT	TCTCTATGACTATT ATGAAACTGTTAAT AGCAATGGCTCATA ATAGATTCAACATG GCGTTGAATCTCAA GGAGCATAAATAA ATAAATAAATTATT T	MVRKEDFEALVSRVLKL EAELQIKDKEINDLKIVV NDMKNQSKLVSNSNDV SNTWSDVVKGRKKNTS VQLDLLNVVVKESHDRE IRQKNLVVFGFKTSTNK DESVAKEDDKQELQNLF KQINVKVTIQRVFKLNT KTDKPPPVIVVLTNKEER	cd01650	757	1007

				CTTGTCAAATTT GACA		NAVLKTAKNLRNSPETN NIFINADMTEAERFKMK ALREERAKLNEANQDLK YYYGIRNERVVKLLKKL YESTQPYQQISPKKVLY XNLNNITKNPNDRHLLN SNKRLIVYYTNSTSLNN KKNDFRALIDIEKPDVFA VTETWYTPLSDVNIEGY VIYRNDRLNGKKGGGV CIYVNSSIKSYEISNIELK SDLIEQIWTWIEYGQEK VLLGCIYRPPDSSCVTNK NIINSIIAAKRAQDKKQF TSIIITGDFNYNTLKWKD SVEESHSQQAEDFVACL NDCFFEQCVTEATFQSN ENELTNILDLVIVDDSNR IFGLSHLPPLGLSSKGHHI LKWFYKMKDRIDDFSSS TKKFAYRRGNYEKINVY IKSINWADEFQNKCIDKC YDKFLYHYNQSCKNYIP			

						IVKLKNKASEVWMSNSL RKEIKYKNNYWTTNKK MNWSIAGSKEKYNKMK RNILKQINNSVKNYERN LVNQVKNEPKLLYSYIR KKQKVKNHIRALNDNIG STMYDKHSIVDIFNKYF HSVYIKEENINDCNFEVQ TNSRFGKIFVDQALVKS YLSKLDINKSSGYDEVN PYVLKVCSESFSIPLSIIY QQSLTNSDIPDLWKMAN VSPIFKNGSKLQAINYRP VSLTSIPCKVLEKIISDNL KNYLESNNLFCKQQHGF LKGKNCITNLLESLDFLT YQLMIGNSIDILYTDFQK AFDRVPHKRLISKLCGY GIEKEMLNWIETYLSNR KQRVIIGDTKSDWLEVL SGVPQGSVLGPLLFLLYI NDLPSKIKNKCELYADD NKIIAVVNNQVDSKSLQ			

						DDINNLTEWSDKWMINF NIKKCKIMHFGKHNKNF EYFIKDSKLDSTTNEKDI GLLISNTMKWNQHVNM VVNKANSRLGLLSKSFT YKDKNTMKILYCTFVRP ILEYASPIWNPYNKNDIH KLEIVQQRATKLIPELRH LPYEERLKKLQLTSLKIR RLRYDLIQYYKIFHNLD KVCWYQQPKTAISISAN GPASNIRGHKARVDIELI KRNRPRENFFTNRVAK WWNKLPEEIISAKSVNIF KDKLDKHMVIMKL			
CR1	CR1-7_HM	.	Hydra vulgaris	AAACGTTATCA TCTATGGACTA GAAGCAACCCA ATGTTCTTCTCA GACGGAAAAAC TAGAAAATGAT AAATCGCAAGT AAATAATTTACT	AACATTAAACCGAT ACAACACGCGATCT ACCGGAAGATTTAA AAAACCTTTCCGTT CATCAAAACTTACG AGCTTTGCCATTTC CTATCGTGGCCCAA TGCTATGGAACCAT	MMPIINKSTRITQHSCSAI DNIFTNTLFHSTFKSGIIR SDITDHFPVFAIFKKPKD LQKKDWYFKKRIFTTRA KNMFKLELSTVNWNFV FLETCANNSYDLFLDKF FNLYDKHFPNKIIRIKTK GLLNPWFTKSLLKSSKR	cd01650	330	590

				TAGTGCTATCG	TTTACAAGTCTGAG	KQKLYIRFINSKKERDK			
				AATGCGATAGA	TGGTAAACTTAAAG	HNYLEHKRLYERTKKIA			
				GCAAAAGTGAG	AAGCCAATACCTCT	RTTYFKNKLLELKGNCK			
				ATCTATTCGCCG	CAAAAATTTAAAAA	KTWDAIKEILGKQKIHN			
				TCTTAAATCTAA	GAACATAATCGAAC	KALPSVITVDGTQISDVK			
				AGGAACTATCG	TTATTGCCAATCAC	SVATKFNEFFVNIGPNLA			
				ATCCTTTGCTTA	GATTCTTTACATGA	SKIKHTPAHFLSYLSKSD			
				TCGAGCTTGAA	TTTTTTCTAACAAT	KSLAFKKLTSKELDDSIK			
				GACGTTAATAA	ACACGATTATTTCC	KLNKNKATGYDAISASV			
				CAAAAATATGG	AATTTTTATTATCA	IKDCYELIKDVLLDIFNK			
				TTCTTAAAGCAT	AAGGTTACTGATGA	SLKTGVFPDKMKVAIITP			
				CAATTAAATTA	AAAGACTTTATCTT	IFKSGDESLVTNYRPISV			
				AGAAAAATCAA	CTTCAGGTTCTCCG	LPVFSKILERIMYNRIFAF			
				AAAATATGAAA	GTCTTGTAACTATT	LYDNNLLYYNQYGFRS			
				ATGTGTATATA	ATAAATAATGAAAT	NYSTEHALMQLVEVISE			
				AATCGCGATCT	GTACTTGTATTTAA	ASSHNKFTLGIFLDLSKA			
				AACAGCAAGTC	TATTTCTAAATATT	FDTVNHQILLEKLKYYGI			
				AAAGATTCATT	TTACGGAAATTATC	SHEYHKWFTSYLTDRKQ			
				CTAAAAAATAA	TTTAATTGTAAATG	FISLNDGFQTNKLKISCG			
				ATTAATAGAAC	GTTGCTTGACAAAA	VPQGSILGPLLFLIYVND			
				GAAACGAAAAA	ATTGAAATAAAAA	LYKASNTMSTIMFADDT			
				AATGCGTTAAG	AAAAAA	TLFYSHSNIISLFQVANN			
				AAGTGAACAAG		ELEQISNWLIANSLSLNA			

				AAAAAATTAAA		DKTKFSLFHSKRKFQSIP			
				TTTCATTACGGT		KSLPMLYINNINIERSFT			
				ATGACAATCTT		NKLLGILIDENLSWRDHI			
				ACTCGAATTTTT		NHISNKISRSIGILYKTRQ			
				CACAAATAGCC		ILDKSTLTQLYFAFINSY			
				ACTAGCAACCA		LSYGNIVWGSTNKSKLD			
				CTAATTTAGCA		CLLKQQNHAARIIHFQN			
				ACGTCATCAAA		RFTNAKPLLKSMRALDI			
				AGTATCTCCCCC		YELNIFQNYTIYVSTSVK			
				AAAAAAGACAA		KNAEGFRKYFP			
				ATATTGACTATA					
				CTCGAATAAAA					
				TCTCAATTTATT					
				CAAGATCCACT					
				TCTCAACATAA					
				ACAAACTGAAC					
				ACTTCTTACTTT					
				GATTCGGATAA					
				AGCCGCTATTTT					
				CCTTAAAAGAT					
				GTGAATTATATT					
				TTTCTAATCAAT					
				TCTCAGTATTTC					

				ATTTAAATATTC GTAGTATTACTA AAAACTTCGAC GAACTAAAAGT CCTACTAGATA ATCTTAACTTTA ACTTTAGCTTAA TTTGTTTAACGG AAAGCTGGATT TTAGATACAAC ATTTAATCAAA ATTCTAAATTTA TATTGCCCAACT ACAAAGCCATT AGTTTTGAACGT AAGTCTTCAAA AACAGGCGGTG GTATATGTGTCT ATGTGCTTGATG AATTTACTTATC AAATTAGATAT GACCTGTCGTTC TCTGAGCCGGA					

				TTTTGAATCTTT CACTTTAGAAA TCTACAATAAG TCATCTAAAAA TATTATTGTGAC AACTACATACC GACCACCAAAC GGATGTTTAAA TAACTTTCAAA ATCACTTAAGT GATTATTTTGAA ACAGCATGCAA AAAAAATAAAG ACCACTCATTTT TATTTCTTGGAG ATTATAACGTC GACTATTTAAA ATATGGTGAAA ATAATAAAGTA CAACAATTTTAC GATAGCTTTAAT GAGTTTGGA					

CR1	CR1-5_HM	.	Hydra vulgaris	AGAAGCTGGAA TGTTTGCGTATA TTTCATACGACA AGTTGGTGGTTC GCGATGGTCTC GAAAACCCACG TAAATATTTCTT TTAACGTTACTT TATTTTTATTTT TTATTTATCAAT AGTTTTTCAAAT TTTAATTTAACT TTTAGAA	AAACAATTTAAAGG TTTAAACTAAATCA CTTAAAAATAAATT AAATCTTTAACACT TTCTTTAGCTTTTAT TAAAATGGTTTCAA ACCATTAGTATAAA AGATAACGAAAAT AAAATTGTCTTTTT CGTTCATATTTTTG ATGATATTAAATAA TCAGCCCAAATACT TATGTTTACATATA TTGCTTTTTTTTTTT TTTAATTAAACAAC ACATTTTTTTTGTGT GTTTTTATAACTTTC ATGACTTTTTGTCG TTTTTACTTTTTTAT TTTTAAAATTATTC TGCTTGAAGATTGT TAGCAACAAATTAT TTTATTGCTGAAGT	MENNLSYNINAKNIVFS DNNSDNHTETLLQSEYY SIPESTQYLCENKNKFSIL HINIRSMNKNFENFKCLL GELNHDFKIICFTETWLK SNETNSNFELRNYTSVH QMRDICVGGGISIYVHN SIDFIQRKDLNVNNTDCE ALCVEIINKLAKNFIINAI YRTPAGSLKTFKTYLRT FLTTKNILQKHVYVVGD INIDLLNHALNSEAKTFI DILLEYNLIPTINKATRV TKKSSTLLDNVITNNFH NSRFKTGIIKTDLTDHFPI FFITESVTLNNATHKSTV FMRQINESSICQFKNLLN NYVDWNLVLQSHDVNN AYDLFLNQFSKMYDKA FPLKVKLINSKSVXSPW MTKGLLKSSRKKQKLY DKYLKNKTYKNETNYK NYKNLFEKTKKQSKVN	cd01650	551	810

					TTTTGATTTGTTTTT	YYAKLLEKNKGNPQKT			
					GTTTTTGTATTTATA	WSVIRDLIGKNKIEKNNL			
					CTGTCTTTTTGTCTT	PQKLIIEGKMIYHKEVIIE			
					TTATCATTTATTGT	KLNNFFLDVGPNLAAKI			
					GCATGTAAATTTTT	PIGQKKFDSYLATTDLIM			
					GTGTGACGGGGCTC	EEPILTKSELHTAFNSLK			
					GGTGATAAGACAA	KNKSAGIDQINVNVIKSV			
					CCTGTCTTCTTCTTG	FDIIEPSLFHIFNLSLKSG			
					CTCCGGCCAGGATT	HIPDKLKIAKITPIFKSGD			
					TTCTTTTATTGTAAT	ETNISNYRPISVLPCFSKL			
					TTCTTTTTTGTAATA	IERIMYNRLYKYLSENKI			
					TTTATCAACGGCAA	LYNNQYGFKKNHSTDH			
					ATTAAATAAATAAT	AIXELVKHISNGFNSDCY			
					TAAAAAAAAA	TIGVFIDLSKAFDTVNHE			
						ILLKKLENYGVKNQSLL			
						WFKSYLTNRKQFILKES			
						KDSKNNLITCGVPQGSIL			
						GPLLFLLYINDLYLSSKV			
						LNTILFADDTNLFYSHK			
						DITVLFKIFNEELDKINE			
						WFISNRLSLNVEKTKFIL			
						FHKPSKAENIPLKLPNLL			
						INNKIIKRESTTNFLGVLL			

						DEKLSWKFHIKYIEGKIS KNIAMLYRTKPFLNNES LKNLYFSFIHSYLSYCNI SWGSTNHAKLKKIYSKQ KHACRIVFGANRKTQCE PLLRQLGALNIYKLNIHQ VMIFMFKAKLGFSPVTF QSYFNEISHKYPTKFSVN NFVVPINFLKLSSYQIQY RGPLVWKEFLPIFSKKQ NFVLNCTLQSFKDDSKR HLLNMEFDILDFKYLF			
CR1	CR1-42_HM	.	Hydra vulgaris	TGCGCGATGAA GGAAAGATGGC GGTCTWAATTC TGAGTTGAAGT TTGGTATACTTT TATTTTAWCGA GTCGAAATTTAT ATTTATCCATCT ATTATTAAATCA ATGAAAAAGAT AAAAAATTAAT	ATATTTCTACCAAC TAATATTGACATTC TTTACTATATATGT CAATTCGTAATCCT TCAGTTTTTAATTT GATATTTTTAACAT TTTGAAATTCTCTA CTTATCTATCTACA TCTAATTTCTCTGG TTTTCTATCTGCCAT TTTTAAACTTTATT	MKNKTLCIFCFKTVTDN HKAIYCDCCSSWVHAK CNQIDLLTYNLLSEDSSE WYCFNCISKNMPFGLLT DNELLTTLSCKNPQTASS NFLVSPIHSRNILKELNTI SGSAINCKFYDVVDFNK AIKPNSDIYLHLNIASLP YHIDDLCSLISTLNIIPMA IGITESNLYLNDTNITDIT INGFNIEHCPTEAKKGGA	cd01650	640	894

				AAAGTAATTAG	ATTAAACTATCTAT	LLYLQSNLNYVVRHDLK			
				TATCAAAATAA	ACTAATATTATTAA	IYLPKLLESVFVEIIKPHK			
				TTTATTCGTGAT	TTTTAGTACCAATG	TNVLIGCIYRHPSMNQN			
				CAATATTAATA	GGTTGGCATTCTTT	EFNTTFLTPLLEKLSKEQ			
				CAAAGTAACCC	ATTTATCTATCTATT	KLIFLMGDFNVDLLSYS			
				TAAAGTAGCAA	TATCTCTTTTCATTC	ESNPTSNYLDTLCSYSFN			
				ACAAAAAAACG	AAAAATTCTCTATT	PSIILPTRITTKSQTLIDNI			
				AGCCATATTTCA	TCTCAACTATTCTA	FMNFYSSDHLSGNLTISI			
				TTATCCATCAAA	CTTCGTACTAAATC	SDHTAQFVCVPGTSNQP			
				CAAGAAGAAAA	TTTGTTGTTGCCAT	NKVKISRCLKRCLKKFD			
				GAAGCAAGAAG	CACTCTTTACTAAA	PNKFIIEISKINWELIIKKD			
				AAAAGTATAAC	TTTTTGTTGCTGCTG	DNINESINNFLKSFNRVL			
				ATAAATAAAAT	TTATTGTCTTTATTT	DQCAPYKTLTKNQIKLK			
				GAGGAGAAGAA	ATGAAATTCTTCTC	SKPWISKGLLKSISVKNK			
				CCGAAAAAAAA	TTATTATAATTGTT	LYKKFVRCKNINTKNEL			
				AATAGCACGAA	ATTATTTCGTTACT	FTKFKFYRNKISNLLKSS			
				TCTATTGGAAA	ATTATTATTATTGTT	KKSYYITYFNNNINNIKN			
				GCATAAGATTT	ATAATTATTATTTTT	TWKGIKDIINIKGSLHKT			
				AAAGATACTTA	GTTGTTGTTGTTGC	PTQLNLNENIITDHATVS			
				AACCTCCTTATT	TATTATTATTCTTAT	NIFNNYFVSIREKLLNNI			
				TTTTTCTTTTGA	CACAATCATTGTTA	MPSEYTYSDFLNFPNAN			
				TGATTTTTGGTT	TTATTGTTATTATTG	SFFIDPVSEDEVSSLISNT			
				AGCCACATTTGT	TTATTATTATTTTAY	LKNDKSYGPNSIPTFFLK			

				TTATTATTTTTC	TGTTATTATTGTTAT	LTSHIISKPLSIMINNSFK			
				TGGACTATTTAT	TACTATTATTATTG	NGKFPDVFKVAQVIPIY			
				CAATATTTGTAG	TTATTATTGTTATTA	KAGSLLDFSNYRPISLLS			
				GTTATTTATTTT	CTATTATTGTTATT	NLSKLFEKAMHNRLYKF			
				CATTACTTTATA	ATTGCTATTATTAG	LDKFKCLYQHQYGFRSK			
				GAATTATAATT	TATTATTAATAACA	HSTTLALIEITEKIRRALD			
				ATATCTAAATA	GTTATTATTATTA	NKHFACGVFVDLQKAF			
				GCTAATATATCT		DTVDHSILLTKLEYYGIR			
				GTAATATAATTT		GIPLKWFTSYLKNRTQF			
				ATTCATATATAT		VSINSTKSDLQNCSNGVP			
				AAAATATTATT		QGSVLGPLLFLLFINDLN			
				ATATACCTATAT		TSFKFATAYHFADDTNM			
				AACTATCTATAT		LLVDKSLKKINKHINHD			
				ATCTATATAGCT		LGNLVQWLRSNKLSLNS			
				TTTATTAGTGAT		NKTELIIFKSNLAPIKKQL			
				TATATAAATGTC		NFRLSGQKIYPLDSIKYL			
				ATATAACTTAAT		GIKIDSNLSFQGHLNDLA			
				CTTTCAGTTTAT		MKLSRSNGMLAKVRHY			
				GTTAATTATATA		VNLETLLNIYHAIFGSHL			
				TTTACATAACTA		RYACQVWGQSHKQIILR			
				TCATTATTTACA		LSHLQNKALKIIYFQYN			
				TAACTATCACTA		NSKSNVLYLISKILKISDL			
				TTTACATAACTA		VVFLNCLFVWNYHHKN			

				TCATTATTTATC		LPLSFNNFFSITNCHYPL			
				ATTTGAACTTTA		RSISSLNLVIPKHQSAKY			
				TTTATTATTTGT		GNKSIKYQCISSWNKLPP			
				ACATTTCTCTTC		DLKSLNSISEFKIKLFNY			
				ATTTATTCTTTT		LLKKYN			
				CTTTGTGCTTGT					
				TTTAACATTTTA					
				TTTGTATATTTA					
				TTTTTATTTATT					
				CTTTTCTTCATG					
				ATTATTTTAACA					
				TTCTTCCCAAGT					
				ATATTATTTTAT					
				TATTATATTGGC					
				TTGTTAGTGTTT					
				TTGTTAGTTGTC					
				AACATCTATTTA					
				ATCTCTTCAACA					
				TCATCTACTTGA					
				TTTTGTATAACC					
				TACCTGTACTGT					
				GTTATATATACT					
				GTTCTTATGCAA					

				ATCCATTTTGCC TCACAATTACTC TCTTATTATATT ATTATATTATTA TAAGTTTATCTT ATTTATTATATC CTTAGTCACTCC TTTTTTTTCAT CACCTTCTGTGT TTATCAATAATA AAACATAATAA CATTCTGCCTAT CAAGTT				
CR1	CR1-20_HM	.	Hydra vulgaris	AACATACAAAA TGGATAACGAT AAAGTAATTAC AATGGGAATGC TTCGAGAACTC ATGTCAATACA AGGAGATACAA TTTTCAAATGCT TTAAGGAAAGT TCGGTCAGTGTC	CATATATTTTATAG TTTACAGTATTTGT TTTCAGTATTTTTTT ATTTTTTTATTTCTT CTTCTTCTTATTTTT CTTGTTTTTAATATT TTTTGTTTTTGAATT TGTTACACTAAAAA AATATGAGCTTTGT TTTTATTGAATTTTT	MSENSLYNLVSYKMISQ PRVNSKGGGVAVYILDK FEFKKNNTFSKTNKHIET LCIEVLNTSGKSFLISSLY RPPCGKQVNFLNILKXF MSQMSKQNKHVFFAGD TNMDALSYEKFANIKNF FNLLFEFNIIPTIFKPTRIT NQSSSVIDNILTNCSLNT KFESGLFIADFSDHFPVF	cd01650	484	751

				CTGGAAAAAGT	ATTTTATTTTTAAA	HIAREVLSNIDNKKCFVT			
				TGAGAACTTAG	AATTAAATTGTTTT	YRNLSIKKLKKNLNNFK			
				CCAAAGRCATT	TATAATGAGTGACT	NKLQVEQWDKVYSSTD			
				GAAGAAATAAA	ATAGGGGCTCTATG	PNIAFDHFSSTFQNIFDN			
				ACGTGGGCTTA	AAAAGGTTGTGATG	VCPKITTEIKNKEYKNP			
				CTTTTGTTGGTG	TTAAACACATCATC	WMTITLIKASKRKQKLY			
				ATCAAATTGAG	CTTACCTTCTTTGA	VKFLKSKNKEDEKIYKN			
				ACAAAGTAAGC	GTCCCTGACTGATT	YKKFFQHSMKEEKIRYF			
				AAGGTTGATAG	AGAACAGAATATA	SNQLDKHKFDIKRTWAV			
				TAAAATGGCAG	CATAATATAAAATA	INELTGREKKKHTCTCLP			
				AAATTAAAGAA	AAAAATTATGAAA	QKLISDKKNYNKYNKN			
				ACTATTGTGTAT	AAACGTTTTTTATC	MSRIFPIKKIITSTTKICQE			
				TAAATCATTCCA	ATGTTTATCTTTAC	FNKYFVNVGPSLASNIV			
				CGGAAAATTAA	GACGTGTATTTTAT	QPNKTFDNFLGEKPTTSI			
				GCATTGAAAAT	GTATTTTTTACATG	NNEKITKLEFNKAIIELK			
				AAAAKAAAAAC	ATTATATGAAAATT	RNKSCGYDGISSNVAIYI			
				TGTCGATCTTGA	GTTATTGTAACTGT	MDSISKPLFYLISSSFENS			
				AACCGACGTAA	TTAATCAGCAAAAT	IFPNKLKLAKINPILKKG			
				CAATTTAAGAA	AAAAGTWTATATA	DCNNVSNYRPISLLPVLS			
				TTGTTGGAATCA	ATAATAATAA	KVFERIIYNRVNNYFTLN			
				AAGAAGAAWAT		NLLYKNQFGFQSKCSTE			
				AACGAATCTTTT		HAIIKLADKIFKSFDSDE			
				AATGATGTTGT		LVLGVFVDLSKAFDTVD			

				ARAAAAAGTAA		HSILLSKLKYYGIIGSTFK			
				ACCTACTAATTA		WIQSYLSNRKQFVLLEK			
				AAGAAGATCTG		SGALDITCGVPQGSILGP			
				GAAATTGCGGG		LLFLIYINDMHKASQKIS			
				TCYAGTTATTAT		TIMYADDTNLFFNHPBA			
				AGAAAGAGCRC		KTLFETMNIELEKFNLW			
				ATCGTGTCGGT		FIANKLSLNCQKTSFTLF			
				AAATTTAGTGG		HKKKNNQLISLKKKQSI			
				AAAGCCTAGAA		NLPLKLPKLLINKNEIKR			
				CTATTGTCTGCA		EKCTKFLGVLFDENLSW			
				AAATATTAAAC		RNHIGLLETKMSSVIGVL			
				TGGCAGGACAA		YKSRPFLDSKSRNMLYF			
				GGAAAACATCT		SLVHSQLLYANIVWAST			
				TATCCAACGCTC		HKTKLLKLQSLQNHACK			
				GTAAACTCAAG		VINYLNRLVNPAPVMKS			
				GGGAAAAATAT		MMVLDIEKLNTLQILIFM			
				ATATATTAACG		YKYKNNLLPSVFSDIFLI			
				AAGACTATAGT		SFSEKYNLRSNTRYNYH			
				GATGAAACAAT		LGKIKSHQSNFLITNRGP			
				GCGAATTCGTA		SIWNHFQNKNIKDLKSIS			
				AAGAACTTTTTG		SFKRETKMHLLNS			
				TACAAGCYAAA					
				ATACATAGTGG					

				TAACGGAAAAT					
				ACGCGAAAGTC					
				ATTTACAGTCG					
				ATTAGTGGTAA					
				GAGAAATGATA					
				AACGAGAACGA					
				AGGTGTTGAAA					
				GYAAGTAATTT					
				ATTTATTTACGT					
				TTTAATAGTTTT					
				ATTAGATATTA					
				AATTTTAWGGT					
				TTTTATTTAAGG					
				ATAATATTAAC					
				AATGAATATRT					
				NTAAAACAATT					
				CATAAYAATGA					
				AGCAAATTTAA					
				CAAAATTAATA					
				TCTCCAATTTTT					
				GAAATTTTGAA					
				ATATAACAAAA					
				ATGATTAYAAT					

				TTTTTGACTGAA ATAGATATTAA WTGCTTAAATA ATGAATACTAT AAACCGGAAGA ATTTAAGAGAA AAAAAGCTTAA CAAAGATTTAA ACATTTTAAGTA TAAATATACGG TCAATGAACAT TTTTATTATTCT TTAAATTATCTT TTTGACGTAATA TGTATATCAGA AACTTGGGAAG ATGCAAACTTG CCC					
CR1	CR1-3_HM	.	Hydra vulgaris	AGGTCATCGCA AATGGCGGACG TATTTTTCTAAC GGTAAAAAATA AATAAAAGAAA	ATCTTATACCTTTTT ACAACTTATTATTT ATAACCCTAACGCA GTGCTTCTATTCAT CTTTTAATTTTTCTC	MKICHHCKFCNKKIKDN HRAIQCDCCSFWIHAKCI PVDKHAYNMLSNDSSE WFCPNCISNNMPFGVLS DLELKSTLSCKNPSNKIK	cd01650	634	890

				AAGGATAAGAT	ATTTCTATCGTGAT	FHEPPDYLKNLFKTMNN			
				AAACAGTTAAT	ATATTACTAATTTT	VFNPTTKCKYLDVNELN			
				TTTTCATTTTCA	AATATCAATTATTA	KSINPGTEIYLHLNISSLP			
				ATTCGCTAAAA	AATTTATTATTATA	FHINDLSSFVGTINTPPM			
				AGTTTTTTTAAA	AACTGATATGTTTC	VIGITESNLYANDSNITDI			
				TTGTTAAAAATC	TATAGTATTCTCTT	TIQGYNIEHCPTESKKGG			
				TGATTTAAATCA	ATCACTGTTGTTGT	ALLYLNSNLNYIVRSDL			
				AACGCTTAAAA	TATTATTACTATAT	QIYATKFLESIFVEVIYPL			
				ATAAAAATCAA	TTTATATCGTAATG	KSNTIFGCIYRHPSLNITE			
				ATAAAAAAAAG	AATATTATTGTAAT	FLSIHFNPLLEKLSHEAK			
				AACTTCGAATT	TGTTATTATTGTTGT	NIVLMGDFNIDLLNYRE			
				ATTAAAAAAAA	TGTTATTATTTTTGT	SQVISNYFESLCSHSLFPT			
				AGAACACTTTA	TGTTATTATTGCTA	IILPTRVTAKTKTLIDNIF			
				AACAAAGAAAA	TTATTATTGYTATT	MNSFPTDIVSGNLTISISD			
				GAACGAAGAAA	ATTATTATTATTATT	HMAQFVCIPNNPPIKKK			
				AAAAAAAAAAA	ATTATTATTATTATT	VKMFKRSFKKFDSDSFI			
				AAGAAAGAAAG	ATTATTATTATTATT	QEISDINWELLIKDDDNI			
				AAAGAGAAAGT	ATTATTATAGTATA	NESINFFLKTFEKILDRY			
				AAACAAAGAAG	TTCTTTTTATAATCG	APYKELTKKQIKLQSKP			
				AAGAAAAAAAG	CTACTATTATTTTTT	WITSGILKSISNKNILYK			
				AAGAAATAAGT	TTATCACTATAATC	KFTRSKNVNTKNILFTKF			
				AAACAAACAAG	ACTACTATTACTTT	KLYRNKISNLLRYSKKL			
				AAAGAGAAGAA	TATTATTCCTACTT	YYASFFNNNINNVKNTW			

				CAAAACCAATT	GTATTATAATTATT	KGIKEIINIKPSTSTKSFN			
				TTAAAGAGAAT	ATTATTACTATCTC	LKINENVISDHTSVANIF			
				AAAGAAGATTG	CATCTTATTTTTATT	NNYFTTIQDELLKNTVPS			
				GTTATTCATTTA	TATTTTGACAGGTG	KCAFGNFLKIPSINSFFIE			
				TTTTCGCGATTT	TTTACTTTATATTA	PVTENEVANLIKDTLKN			
				AATTTGTAATTA	GTACAAATTACTAT	NKSLGPNSLPTFLLKLVS			
				TTATTTTTTAAT	TTGTAAATATCCTT	HIISKPLCTMMNNSFKN			
				ATAATTATTTAA	GATGTAACAACTTT	GIFPEAFKVAKVIPIHKM			
				CTTATTTGTTAT	ATTCAGAATATATW	GSYLDYSNYRPISLLSNL			
				TTTATTATTATT	TGATTTGATTT	SKLFEKAMFQRLEHFLE			
				ATTATTATTTTT		KHKFIYKHQYGFRNKHS			
				TTTTTTTTTCAC		TTHSLIEITEKIRQAIDNK			
				TTTTTCACATAT		HFACGVFIDIRKAFDTVE			
				TTTTTAATTATT		HTILLEKLKHYGIRGIPFL			
				TATTTATCATTA		WFSSYLCNRTQFVSINGI			
				ATTATATTCATT		NSGLAKSFNGVPQGSVL			
				AGTTCTAGTTAA		GPLLFLIFINDLNVSLKFS			
				TAAGGTTTTCCC		TAYHFADDTNLLLINKS			
				GTTATCTATCTA		LKKLNKNMNHDLANVV			
				AGTTGTATTATC		QWLRSNKLSLNSKKTEII			
				ACATATTATTTA		IFKSAKTKINKQLNFRLS			
				TTTATTTATCAT		GQKINPVNSIKYLGIKID			
				TAGATTATCTAT		SNLSFASHLQDLALKLS			

				TTAGGAAGTTTT		RSNGILAKIRHFVNHETL			
				GTTATATATTTT		LNLYHAIFHSHLRYACQ			
				ATTATATTTAAA		VWGQSKQLAFLRLTYL			
				TTATTTTACTAT		QNKALKLIYFQHTNSNC			
				ATATAATTTTAT		SILYFLSKVLKLCDLIQL			
				TATATTTTTTGA		SNCLFVWNQNHNNLPLT			
				TCATTATTATAA		FINFFSYRENCKYILRSA			
				CTTTATCATTAT		LNFKLSVPKYRTVHYGY			
				TATATATTCTTA		ESIQHKSIQTWNNLPSQL			
				TACATTTATTTA		KSLKSFSKFKTALFNHFL			
				TTATGGATTTAT		EKYSL			
				ATTATTTTAATT					
				TTATCTATTGTT					
				TATTATTTTATC					
				TTAGTTATTAAT					
				TAATATTATAA					
				ATCTATACTCTT					
				TATTTATAATTT					
				CGATAATAATT					
				AACTACGTACT					
				ACTACTGATTAC					
				TGTTGCTATTTA					
				TCCATTTACTGT					

				TTATTGTTTGAT TATTTTCTCTTT TAATTTTCTATA TTTTCTGTTCTA TAAACGGCTTT ACCATCTATATC					
CR1	CR1-12_HM	.	Hydra vulgaris	AAAGGTCTATA ATGAAGCCTAT TTTCGGAAAAA AAGTTGCATTTA ACAGAAACTCT ATTGCAGGCGA ACGTACGGTGC GAGAATCTTAT ATCTTTGTAGGA GGGGTCTGTAA CTCAATGTCGG AAGAAGATCTA TCTAACCACATC AAAAATGAAAT AGGTATTGAGC CTATTAAGGTA GAGCTTAACCG	TCAAATACAGGAA ACACAGAATAAAA AATCCTTTACAAAC GAAGCGTTTGATGC ATGCAACACACTTG ATTTAAATTTCATA CAGTGCATGTTAAA TGGCAAAAAAAATTA AATTGACGCAACCC GTCGTAGAGCTACA AACTAATGTTATGC TAATATTGAAAGAA ACGTTCCAGTTTTG GAATTTGAAAGAA AAGAGAGAGGCGT TTAAAATCTTGATG GAGGAGAATGTTCC	MEGNIESKTFENISFCAY NCQSFKSNSIYISELVKSF DIIFLSEHWLLNIELFVLK NIALPTHKVFFHAAEKK AFGRPFGGNAFIIKKDVL LSPHIIYEDDNILAIKGKN KCHNLVFIGIYLTSCRNN EESLSKYEQQLNLITSIV KNYEDVSECIIAGDFQSY PETLYDSEIRNNKKRNK FSNHLTNFIKSNKLSFIDI ISGAGPTYTYQHKTLSNS SYIDHIAALKESSLSFINT SVIPPSPFNFSDHLPIATN IVVTHTNLSTVINSIMHK YNYIPKYAWNNEKFINI YNHNLSVAFNYYTFNEE	cd01650	542	799

				GGAAAATAAAT	TAGAACTCTGTCCC	EIEQELKQTCEKIQNAAL			
				ACAATCGTTCAT	AATAATATGTGGAA	LAVKQCFKQQICKYSKS			
				TCAAAGTTATC	AACAAACCGGCTGT	WWTTDLKKCKENLSFN			
				ATAAAAAGTAC	TTATAAGCATTTAT	YKQWQKSNYNKTQECI			
				TGAAAAAGACA	TGCTAACAACTACA	EYKRYIFSRKNFRKAVK			
				TAATCCTAAAC	AGAGAGCCCGTGTT	AAHNKKIHEKTINIENLR			
				CCTGATATTTGG	GTTTTAGCTGATTT	TTNPQRFWDNIRKIKTK			
				GATACTAATATT	TTAATTTATTGTTTG	ATTRLFTINKSQNIKDIV			
				ATCGTTAAACC	ACTATTTAAATATT	QEFRQHFQTLLNTPLITN			
				ATTCAGGCTTCA	ATTTTACCTGTAAT	NNHNPSQIPPLSKEPNSLI			
				TCGTGTTACCTC	TTATTACCGGGTGA	ITSADIINCISQLNSNKSP			
				AAGAAACGAAA	TAAAGAGAAAATA	DSCGLSAEHLKNSHNNN			
				CACCACAGCAG	TAATATAATAATAA	LLLWLAKFYNSILSNGK			
				AGTGCAGTTAA		VPNDLSTSTIIPLVKSYK			
				TACGCTACACTC		KSLNNPDNYRGISILPIFT			
				TGTACAACCAA		KLLEYLILLICPSITDSHS			
				CTCAACAAATA		HQFGFKKNSSTLHAEFL			
				TTTTCTACATCT		LSETVKHYKNNNTPLY			
				AAAGAACAACC		MCSLDANKAFDTCNWD			
				ATCCTTTGCACA		LLFEKLYFQKKLPLSVV			
				GCAACATTCTTA		HTISSLYHSGTANVSYQ			
				CTCAACTCGAA		GVTSNQFSLSQGVRQGSI			
						LSPHLYNIYTESILNEIVS			

				TACTAAATTCTA ATTCAAA		ECKVGSTINGIYTGVIAY ADDVILLSPTISGLQELIN RFQTYGVANFIKLNTEK TEFLISGKSYIXNNVIKIN GVSIYLQNKLKHLGFQW DNDTSKNKIATLQKTNL NERITQFQSVAKALIDSG IRYCQPSTIVHLYNTLTV PKLIYGLELCKYDTNFLS KIDAVGRAVLKSFFCISK YSKNYLHSFFKIEDISTIL LRNKVNLFIRLLKNQTC LXTDN			
Daphne	Daphne-10_HMa	.	Hydra vulgaris	GGGAAATTTTTC AAGAAATGTTY CAAAAACAACA AAANGATATTT TAAMATTAATA AGTGGAAATTT AAAACTGACTA ATGATCGAATC GATGGATTGCT AAAAGAAATTA	AAACTCAAAAATTA TGGTATAAAAAATA CAAATTTAGCTTGG TTTAAAAGTTACTT GACTAATAGAAAA CAATATATTTCATA TGATGGMGGAAAA ACTGAATATATGAC AATTACTTGCGGTG TTCCCCAAGGATCA	MIFLRFYIIFLNFNIALSY FLFKMADETVLTNLNFN SFESKKSILLNNYSDPDI NFYNNDEVIKNINPLYY NPNSKNIANGMDDNSFS MLHINIRSIQKNFESLKQ FLYKIKINFQIICLSETWC HDKNIENNSNFQLPNYK VIHQVRASNKEGGGLCI FIKNSLLYKVKPNLSSTT	cd01650	592	694

				CTACATTAAAA	ATTCTAGGACCACT	NDYESLCIEIINETSKNIII			
				GAAWCTTGTGA	TTTATTTCTAATTTA	HALYRPPSGSIKVFENHI			
				AWCTTTAAAAA	TATAAACGATTTAA	KKIIKNKSSINKTVYFVG			
				AAGAAAACGAA	ATAAATTTTCTAAC	DINLNILDYDKNKNIKNF			
				AACMGAAAWG	ATTTTAAACTCAAT	FNVIFQNSYIPLINKPTRI			
				CCGATCAAAAT	TTTATTTGCWGATG	TKESATSIDQIITNEFINK			
				AAAAACAAACA	ATACTAACTTATTT	KIKTGIFKTDISDHFPIFIJ			
				AGCAGTTAATA	TATTCTAATAAAGA	SQKCIYNAQNKRKKIITT			
				AAGTATGAAAA	TATTAATATTTTATT	RLINDNSIKHFHNLLSCV			
				AATACAAAAAG	TAAAACAGTAAAC	NWDTLKLNLNABKAYD			
				ACATTGAAGAA	AAAGAACTTATAAA	IFLTEFLKLYNKAFPEITK			
				AGTCTAACGTTC	AACTAACCGAATGG	FIKTKTLLNPWITKGILK			
				TATCAAGATAG	TTTAAATCAAATAA	SSKKKQRLYNKFLKKKT			
				CCAWGATAAAA	ATTGTCCTTAAATA	LKNETDYKNYKRLFESI			
				AAGTTAGCGAY	TAAATAAAACAAA	LKRSKKHYYSEQLIKHK			
				TTAGAAAAAAA	ATATACTTTATTTC	NDSQKTWQIIKEVIGKK			
				AAAGTCATCGA	ATCGTCTTCATAAA	NLYRNLLPKNLKFNNKY			
				ATGCAAATATA	AAAGAAAATATTCC	IVNKSLVAETLNQFFVNI			
				GGTGAGAATGA	ATTAAAACTTCCAA	GPTLASKIATTQLNFKSY			
				AAAAAATAAAT	ATCTTTTTATTGAT	LTSNNINVMXNYKLTEK			
				TTAGACARCTG	AACTCATTAATAAA	ELLDAVSLLKPNESLGF			
				GAAGATCGACA	AAGAGAGCAATCA	DDISSNVIIKSIKYITIPLL			
				AAGAAGAAATA	ATAAAATTTTTAGG	HIFNLSLKQGVFPEKLKI			

				ATCTTCGAATTG	CGTKATTCTTGATG	AKVIPIFKSGDPSNVANY			
				AAGGAGTTAAA	AAAATWTAACATG	RPISILPCFSKILERIMYN			
				GAAAATGATAA	GAGGGAMCACATA	RLYSFLNINNILYNKQFG			
				CGAAAGTTGGG	AGTATAATTGAAAA	FKSGHSTDHAIIHLVHDI			
				ATGATACGGAA	TAAAATTTCAAAAA	FKAFDEKKFTLGVFIDLS			
				ATAAAAATAAT	ATATWGGTATATTG	KAFDTVDHKILL			
				AAACATTTTTGA	TATAAAGCAAAAC				
				AAATAATCTCA	AGTTTTTAAATCAA				
				ATGTAAATGGA	ACTTGTTTAAAAAA				
				GTAATCATAGA	CTTATATTTTTCTTT				
				AAGAGCGCATA	TATTCATTGYTATC				
				GAACTGGAATT	TAAATTATGCAAAC				
				ATTGAAAAAAA	ATTGCTTGGTGCAG				
				AAAACCAAGAA	CACTAATGCAACAA				
				CAATTGTAATTA	AAATAAAAAAACT				
				AGTTACTAAATT	GTTTAGTAAACAAA				
				ATAAAGACAAA	AACATGCTATAAGG				
				GTAAAAATTYT	ATWATTACAAATGT				
				AAAAAATGCCA	AGATCGTTTCTCTC				
				ACAAACTAAAA	ATACTAAAACACTA				
				GGWTCTGGAAT	TTTAATAAACTTAA				
				TTATATCGACG	TATTCTAAATGTAT				
				AGGATTTTTCGC	ATCAACTAAACCTT				

				TAGAAACTACA	TATCATATTCTTATT				
				ATAATAAGAAA	TTTATGTTTAAACT				
				AAAACTTCTTG	TGATAATAAAATAT				
				AAGAAAGCAAA	CACCAATTTTATTT				
				ATGCATAGGAA	AATTCATTTTTTAA				
				AAATGGTAAGT	AAAAATAAATCATA				
				ATTCTGTTGTTA	TATACCCTACAAGA				
				TATACGATAAR	TTTTCAAAAAACAA				
				CTTATTGTTAAA	TTATATTCAATCCA				
				GAATTTATAAA	AAACACATTATTCA				
				AAAAAAATAAT	GCAACTAAATTCTC				
				GATAA	AATTGCAAATCGAG				
					GTCCTAAATTATGG				
					AATACAATATTAAA				
					TAATGAGCTTAAAA				
					CTAATTCTTCTTTA				
					AACCAATTTAAAAC				
					AAAACTTAARCAAA				
					AACTTTTGATAAAT				
					GAAAATGAATTAA				
					ATTTCTTCTAAAAA				
					GCTTTTTAAATTAG				
					CAAACAAAAATAA				

					AAATTAATCTATTA				
					ATTACTCTTTAATA				
					TTAATTTAATATTTT				
					AAACAATTGTCTTA				
					TTGAAATCTATTTA				
					TTATCTTATAAGTT				
					TTATTTACGACGAT				
					TTTTTAAGTTTTATT				
					ATTTTTTTTKAATTT				
					CTTTTTATTATTTTC				
					TTTTTTATTARTACT				
					TACTTATTTTACTA				
					ATCAATTATTATTA				
					TAAATTTTTCGAAT				
					TACWAAATAGTTAT				
					ATATTTTATTTTTTA				
					TTTTGTTAACGTTTT				
					ACTTATGAATGCAT				
					TTTGGGGCTTAGTG				
					ATAAGACGAATTTT				
					GTCTTCTTCTAGCC				
					CCAGTCATGTAAAT				
					ATTTGTTTTTATAA				

					ATTACATTGTAAAT TATTTTCAACGGCA AATAAATAATAAA AAAAAA				
Daphne	Daphne-16_HMa	.	Hydra vulgaris	GAAAGTCTTAA CTTCCAAGAAA CAACTATTATG GAAAAAATAAA TAAGATCACGA AGCGTTACGAA AAGGAGATAAA CAACTTAAATA AGAAGACACTA GATTTGGAGAA TCGATCTCGAA GAAACAACCTA AGATTAGATGG AATATTTGAAA AGCCAAATGAA AATTGGAATGA GTGCGAAAATG CAGTAAAGGAA ATGTTTAAAAA	AAATTTATGTAAAA TGTATTTTTCAAAT CTTCAGCATATTGA GTATTAAAATGTTA GTATATTGTATATT TGCATTATGTTATA TTGTATTAGTATAT TGCATATTTGCATT ATGTTAAATCTTTG CATTATGTCTTTTTT AAGGTTCCGGTGAT AAGATCCTTCTGAT CTTCTTTCGGAAAC CTAGTCTTATATTG TTAGTACGCTGAAT TGTATATACATATA TATATATATATATA TATATATATATATA TATATATATATATA	MALISDFETLVFNFFETK FFSLHENSDPDINYFHDT KIECSYYYPNELKGILFK NVNNDQIRILHVNIRSIN CNFEKFQNLLEETKYFF NIICLTETWATSNDIKKN FNFYLSHFEMISLERQIN KRGGGVLIYIHKDIKHC CRNDLSVSDGDGEILTIE IINERTKNILVSCCYRPPD GVSENLSMFLQQNVIAK GNNEKKKNFLVGDFNM NCFVYNDDNKVKNFYD TIFESGAIPLINRPTRITK NSATLLDNIITTDLFNND IKLGILKTDISDHFPIFLTI NTDQLPKQKLNKIVKKR FYTSSNIEAFKNQLSLLH WKHLNFNENANNLYEK	cd01650	569	828

				ACAACTAAAAA	TATATACTCATTTTT	FYETFISVYESNFPIVTIT			
				TTAGTAATGAA	TTTTTCTTTTCTTGT	LKTKHFDNPWITKGFKK			
				ATTATCATAGA	AGAAGTCTTAAATA	SSKTKQRLYIKYLKTKTP			
				AAGAGCTCATA	TTTTTATGTATCAG	DSEKIYKDYKNLFEKIRK			
				GAATAGGCCAA	ACTTATATTATTAT	KLKKNYYSKLINKAKND			
				CCTAAAGAAGA	ATATCACGACTTAT	SKRTWQILKEITGKQKT			
				TAAGAAACCAA	ATTATTATATATGT	CSSSLPQMLKVDNNSLH			
				GAACGATTGTC	ATCAGACTTATATT	EPQIIAHEFNKYFTEIGSI			
				CTAAAGTTATTA	ATTATATATCACGA	LSSKIQKTQTSFYDFLLP			
				AATTTTCAAGA	CTTATATTATTTTGT	FDKNISSEELSAELSFDEF			
				CAAAACAAAAA	ATCACGACTAACAT	EKAFKSLKKNKAPGADE			
				TACTTAACGCTA	TGTAAACTAAAAAA	INGNIVIDCYEQLKNILF			
				CAAAAAACTTG	AAAAAA	KIFRASIHQGIFPERLKLA			
				CGAGGAACTGG		RVTPLHKEGDRSNINNY			
				GATATATGTTA		RPISVLSIFSKILERIIFNR			
				ATGAGGATTTC		VYNYFNYNNLLYNNQY			
				GCTAAAGAAAC		GFRKGSSTEHAIIQFIHNI			
				GATGGAAAGCC		SKSFEKSQYTLGVFIDLS			
				GAAGGATGTTG		KAFDTVDHHILIQKIKYY			
				TGGGAAGAGGT		GLNNKILKWFESYLTNR			
				CAAAAAGTTGC		KQLVYSNDGYQSEPLSIT			
				GACTTGAAGGT		CGVPQGSILGPLLFLIYV			
				AAGTATGCAGT		NDLNKASKLKSIMFADD			

				TATAAAGTATG ATAAAATTGTTT CTAGAGAATTT CGAAAGTAAAT GCGCTCATTCTT AATAAGAACTC CTTTTTAATTTA TTTGTGATAATT AATC		TNLFLSHTDIYELFSTTN KELKHISNWFKANKLTL NINKTKWIIFHSCAKKRF LTNNMPQIYIDETIIKRD TVIKFLGVYLDENITWR KHIDHISTKVSKNIGILY KARNYLNKKNLTQLYY SFIHNYINYAIIAWGSTE KSKLQHLYRRQKHAIRV VNYADRFSRSKYFFDYM NVLDVYKLNIFNVLCFT FMWKNDLSLFVFNDLFS LKPINKYTLRSNSFLNEP FCKTKFNQFCIAYRAPH LWNKIVLPNFDPSTTLPV FKTKLKNFILCMDNILKF Y			
CR1	CR1-11_HM	.	Hydra vulgaris	GGCTAAAGGAT TATCATTTACTC AAATTAAAGAA TTACTGGAAATT CATGAAAACAC AATATTAAAAA	TTTTTTGTTTCTTTT TTTAACTTTATATTT AACACTGCATTAAT AAATTTTATAATTT AACAATTTATAAAC AGTTGTTTAACACT	MDSNEPVNFESTLFNFF QTNDFLLDSDSDPDLNY FSEAGALQNKCSYFYNH EIKEFLDRDNFNVIHINIR SLKKNFENFRNXIEETFN IFSIICLTETWCDSDDVN	cd01650	418	677

				TMTTTARCGAT	TTTGTTAAAGGTAT	FNSNFHLPGFNLISLARK			
				AAATTTGATAA	CTTAAAACTGTATA	TNKRGGGVJIYVKEXLX			
				AATGGAAAACA	ACAAAACCTTTAAC	FFTRLDMSISDGDKEVLT			
				AAATGTAAGTA	AAAAAGTTTAACAA	IEILTKKTKNIIISCCYRPP			
				CTTAAAAGMGA	TTGTGCTAAAGATA	SGVIESFNSFLRNDIIKKS			
				AAATAAAATTT	TCTTAAAACTGTAT	NHEKKFNYLIGDLNLNC			
				TAAAAAATGAA	AACGCATCTATGTA	FEYHVNNNIKRFYNDIFE			
				ATAGCAGATAT	TTTATAACGGATTT	NGAIPLINKPTRITQSTSS			
				TAAAAGTGCAA	CTTTTTCTTTTTTTC	LIDNIITTDVFNESLKKGI			
				TGACTTTTCATA	TTTAAATTTGTTTAT	IKSDISDHFPIFFSINTNNE			
				AATGAAAAGTA	GTTTTCAATTCTTA	XVTKTHSLCLKTIKFND			
				TGAAAAAGAGA	ATTCTTAATWATTT	KFLYDKKEIANELNKYF			
				CTAAAGAATTA	GTTATCTTCATGTA	VSVGPNLAKKIPKAINPI			
				TACGATTTTAAA	TGTTAAGTATATAT	NNLFIPNTSQFNSFELSFE			
				GAACATCTTAA	TCCTATTGAAAGTT	EFESAFKMLKPDKAVGP			
				ACACAACATGA	TATTTATTTTGAGA	DDINGNIVINSYDIIKDIL			
				AAGTAAATGAA	ATATTTTATTAATA	FKIFSCSIKQGIFPDQLKI			
				AAAGCTTTAAA	CGAAATATTAACAC	AKVIPIFKGGDLTNVSNY			
				CAATGATGAAT	GAAAATTTTTTAGC	RPISILSVFSKVLERISYN			
				TTAAAGACGAA	GGTTCTCGACGACA	KIYNHLSVNNLLYSNQY			
				CTTGCKGAACTT	AGACCTTACGGTCT	GFKKNNSTEHAILQFTRS			
				GAAGACAGAAG	TCTACGAGTWCCCG	ITDSFENSQFTLGVFIDL			
				TCGAAGAAATA	CGTTCTTTTATTTCT	AKAFDTIDHKILFKKLK			

				ATTTAAGATTTA	TAATATCATTTATA	YYGITGNVLNWLKSYLN			
				ATGGAATTGAA	TTATTATATCATTG	NRKQFVYVDVVSPVNLF			
				GAAAAGGAAAA	TTAACGAATATTTA	DITCGVPQGSILGPLLFLI			
				TGAAACGTGGG	TTAAATACAAAGTT	YVNDLHKASNLMTIMF			
				AAGAAAGCGAA	GTAATAAAGAACA	ADDTNLFLSHNNITTLFQ			
				AAAAAAATTAA	AAAAAAA	NMNIELVKISDWFKLNK			
				AGAANTCTTAA		LSLNIDKTKWIFFHPNRK			
				AAKAAAAACTT		KHLLPSEMPLLFIDNILIK			
				GGTATTACGAA		RVTVTNFLGIYIDENLSW			
				CGAAATATAAA		KNHIENLCSKISKSIGILY			
				TTGAAAGRGCT		KARSVLNKHTLIQLYYS			
				CATAGAACAGG		FIHCHINYANIAWGSTN			
				AAAAAAGAATG		KSKLXPLYRQQKHVARL			
				AAGTTGAAATG		INFKDRFTHAKPLLFEM			
				AAAAAACGAAG		KVFNIYELNIFNVLCFMF			
				AACTATTATTGT		KCKTNISPVSFHNLYXK			
				AAAATTTTTAA		KEKNKYILRNDNFIRQP			
				ACTATAAAGAC		VFQTNFGKFCISFRGAFL			
				AAAGAAACCGT		WNKIVLKNFKDFSQECN			
				TTTARAAAAAT		YRSFKQKLKKVIFSIDDI			
				ACCGATTACTC		SIYF			
				AAATTATGGAA					
				CGAAAAACTAT					

				ACATAAATGAA GATTTTAGCGA AAGAACAATGG AAATACGAAAA AATYTWTTTAA AGAAGCAAAAG ATTTAAGAGCC AAAGGTAAGTA TGCTAAAGTTGT ATATAATAAAT TAGTTACACGC GATTTTTAAAA GAATTAATTTTT TTATTTTTGGAT TTCGAATAACT AAA					
CR1	CR1-72_HM	.	Hydra vulgaris	AGTCCCTGTGTA ATAAGCAAAAC TAATAATAATA ACATTACAATTT CACTGACAGGA GGTAATCCACT GTTATTTAGGGT	CGTTTTTTATTATTT AGTGACTGAACTAA CTTGTTCATTAATT AATTTATACCGGCT GTCATAGATATTGC CTGGCGCTTTATCT	MVNLQLGDFNTYKCEQ EAIIVDLKKLIETLRTDN ANLTKRVLDLENSKSKS SSAVTGDWASKVKKTP DQMQIINTIANETKERDK RENNLVIFGLEESTLSDK VKAIEEDKQNIETIMKKL	cd01650	791	1046

				GGTAGCCCTCCT ATCCAGTAATC CAA	CAACAGCATTAATT AATAA	KVNVKIKNVIKLKSKNG YKAPYIVVLNDKLERNS VLKNAKELKKDKTYES VFINPDLTESERYKSKLL REECKKLNNDNINKANF YFGIRNDKVVKIKKQLN AKKDLNHIRCFNYETNN KINLFSGGGQNLNKLLS NKRQLDIEIVNFHERKK CDLSINKEHDTQKCNEL MIWYTNATSLNNKIDEL RLQCEIYKPDVICVTET WFKAESDVNIISYNIFRS DRKTHGGGVCIYVKQCI NQFETNNLQLNNNFIEQI WTCICFKNNKMLIGCIY RPPDISHEYNNYVLLSIN AAKMSIYQKEYSDLIITG DFNYNTIDWSDNNCPFS ATDNESQTNKFIECLEDC FLFQCVCEPTFQIANYKS TNILDLIITNDKNKIVSIT HSAPLGNVKQGHHLLK			

						CCYLMNDYKFSTEHKKS IRLYQKGNYELFSCHFN NVDWENGFKSKSINEM YDLFLYHYNYATNLFIP RINYQNRRNRNKWITPE LKHKIRTKNNIWLANKR NGWNEESTKIQYNKLKI RIKNDIKNNIKQFESTLA SKFKSNPKLLYKYINEK KSVKTHIRALEIDNGEIIN DQFAIATELNKYFHSVY VNDNCTNIIEFHNIPSNL LLNTVLITQNDVQTRLQ ALDKSKSVGFDFVHPYV LKECSSSISYPLYLIFKKS METGTLPDMWKKANV MPLFKKGSKLKASNYRP VSLTSIPCKVLERIIADCI MQHLIKNNLLSKKQHGF MKSKSCTTNLLEYLDILT DAFHNRTPVDVLYTDFK KAFDSVSHKKLLSKLLT FGISGKLLTWIVCFLANR			

						KQRVVLGKSVTDWVDV LSGVPQGSVLGPLLFLVF INDLPENFINECRLYADD NKIIAPISSQNDSQLFQN DINKLDEWSTKWKLGL NFEKCKIMHFGSNNNEY SYFMNNNNTLMEIEKSK LEKDIGVYISSDLKWAK QVKYAASKANSILSLLN NTFKYKDKDLIKTLYCT YVRPNLEFSIQAWSPYY TKDINELEKVQRRATKM IPELRHLEYKKRLEILGL TTLETRRLRCDLIQQFKI FKGIEIIDIKQNQTMNSIC SSGPAANIRGAKHRMKP ELVRNCLKRQYFFSNRV VDGWNSLPAEVVQSTT VKSFKLNLQSVDLEAIT NKIKTKKAYY			
CR1	CR1-55_HM	.	Hydra vulgaris	TTAA	TACAATTTTGTTGT AACTTGTCCAGCAC TTGATACAATTTAA	RWQPKFCIKYYNFNFLP KLKIFIQLTLIKMVRKED FEALVSRVLKLEAELLT	cd01650	787	1043

					CATTTCAGTAAAAA	KDKEITQLKILVNDLKN			
					AAAAAAAATGTTTT	QPSNVLNSSACAWSDVV			
					TGTTAATGTATACT	KGKKSNKAAQLDLVNV			
					CAGTGTTAGTGGCC	VVKESQDRDLRQKNLIV			
					TTCGGGTCCTTCTG	FGFNTSKNKNDDVAREE			
					TTGTTTTCAAATAA	DILGIQNIFNQINAKATIQ			
					ATAAATAAATAAAT	RVFKLSIKSDKPPPVIVV			
					AAATAAATATAT	LTNKEERNAVLKTAKTL			
						RNLPETNNIFINADLTVA			
						ERLKMKALREERAKLNE			
						TNKDLQYYYGIRNERVV			
						KLIKRQFISSQQDQLLFS			
						KNLLFKNSNKFIEKNNK			
						RQLINLKNQELFNHPKN			
						NYCGKWALTNIRSLGNK			
						IADFQHFVHLHSLDICCV			
						IETWLNKNLPDSLLCPRF			
						YTVLRCDRSSKGGGIAIF			
						FRSFIQFTKVDIPPEFHEL			
						EIVCADLNLNGQSFRVIG			
						YYRSGGFGKVAIDYMQS			
						SVKCFSILCSTKKNVVIL			
						GDFNLPDIDWHYYHGPD			

						SVIHNSFLHFINRYSLTQ			
						HVDQPTRENSILDLILSSS			
						SSFISNLCIEPPIGTSDHN			
						VIIFTPNLSCLYNKTEVP			
						VQQFFCWKNADYQVIN			
						KYLKSVDWSIIFQTCFNI			
						EACWHTFSTILNNIITQH			
						XPKVRVCEEHKLSHKIH			
						YPYYVKRLINNKAMAW			
						KRSKVTGKSNDKAAYK			
						SIAAQCSDSIRKYHAAKE			
						LKMIRKDNLGSFFNFVN			
						GKLKTCSKLNDIRRSDN			
						TLCHDKEEISCIFNNFFA			
						SVFTVDNGLSPKFTDRV			
						NVHSTCLSSINFSPSTVE			
						KALRALKPTTSTGLDGIP			
						NVFLKNCANSLSIPLSHL			
						FDTSFKDGAIPLDWKVA			
						SVIPIHKKGPTVDPGNYR			
						PISLTSTCCRVMERIINN			
						KLINYLLLNNLITKEQHG			
						FIRKRSTCTNILESLHDW			

						TLNLQRRIPTDIIYFDFKK AFDSVSHPKLLMKLPAY GIVGNLLNWLTNFLHMR SQVVNLNNVTSHAISVT SGVPQGSVLGPTLFLLFI NDVSHINKQLDVKLKLF ADDIKLYSVYDVRGLQS DLHTAVNXLYEWSCTW QLQIATEKSFVCTIANQS QNIAQRVYTINNHELAH VNCIRDLGVTIDRNLKF NQHINLIVHKAMSRAFLI LKSFYSRDRLLMVKAYC TYVRPMLEYCSSVWSPY TICLINKVEKVQRFFTKR IAGLWSVSYDKRLAILK LNTLEYRRVFKDLVLCY KILNGKLDTDLSNILILN SYLNNRGHNFKLYKHH CSLNSTKCFFSNRIVNV WNSLSEDVVSASSVLMF KKRLAVLLCSS			

L2	L2-1_HMa	.	Hydra vulgaris	AGGAGCATGTT	TAAAGGAATGTCAA	MLKKEISESEIKQALGQ	cd01650	64	321
				TATGAAAATGT	AGTTATGATTTTAA	MKNNKSPGSDGLTIEFY			
				TAAGGATTGGT	AAAGTCTTCATTTT	KHFFXLIKKDLIESLNNV			
				GGGAGGCTGGT	ATTTAGCCAAAAGG	LFTGEMTNSQKEAJITCI			
				AAGGAGAGAGT	AAAACAAAATACG	FKKGNKNDIGNWRPISL			
				AAAAATAATCT	AATATAAAAATCTA	TNVDYKILTKVLANRLK			
				CGAAAAAATTT	ACTGTCTTAACTCT	DVLPNIIHSNQTACIKNR			
				AGCAAAAGATT	TTACAAATTATTAA	TIYYNLSYTRDIISIAKKS			
				AAAACAAGAAA	ACAAACCAACAAA	KLDACILSIDQVKAFDR			
				AGAACAAAAAA	AGTTTAGACTATAA	VDRAFLFDSLKHFGFGKI			
				GAATTTAAATT	ATTTTATAAATGGA	FISFIKTIYNNINAKVKIN			
				AATCAAATTATT	AAGATTGTTTTGAA	GLLSEKIDILRGVRQGCP			
				AAAAAACGCAT	AGCATGAAAAATAT	LSMILYIIQAEIFSCYVRQ			
				CGAAAAAAATG	AAATGAAAAAGAT	NKNIKGIHVKESETKJQQ			
				CATTTAAATTGC	TGGAAACGTTTCTT	YADDTNFFLIGDDSIREV			
				AATATGAAAAA	TATAAATCTTCAAA	GKALKLNQKATGAKINI			
				ACTTTACGACG	AAAGTAATACTAAC	AKCQGIWLGKNIFKDKS			
				ATTTAAAAGTT	AACAAAGCGTGCG	EYLNFNWDXNTFKSLGI			
				AAACTTGAAAA	AAATTCTATTTAAA	FFSNNANFQYETHWKER			
				TATAGAAAAAG	ACAGCACACTTTTT	FNKMENSMRKWEKYKL			
				AAAAAGCCGTA	GTTGCCAACCAACG	SIKGRKTIINSLIIPQALHL			
				GGCGCGGCAAT	ATAAAATGTTTGAT	AFIAPLPSNGILNKIEAKI			
				AAGAGCAAAAA	TTTAAAATTAGAAA	NNFYWGGKICRLNKTKT			

				TTCAGTGGAGG	AGACAAACTTTGCC	QQKINEGGAGVNSLTNK			
				TTAGAAGGGGA	CACAGTGCAAACTT	LKAIQLQWISKLYNESIC			
				AAAATGCTCCC	AAAAATGAAAATA	GPWKDAMVYILNLYRN			
				ATTTTTTTTTCG	CTGCCCACATGATG	ANQSKHTFETIPTQMKL			
				GATTGGAAAAA	TATGAATGTGAAAA	LPPFYSQLLVNWHSLKQ			
				CAAAAATTAGC	AGTACAGCCACTAA	KHVLPITKIEEILNQPLFF			
				TAAACAATATG	TATTTTATGTTCTA	NSMIKTMPNKFLTPSET			
				TCTCAGAAATA	GATATTTTAGACCA	HIKNDIILVADITKTFQPG			
				AACGGCAGAGA	GCTCTACCCTACTG	FLHHKSIGLTEKELEFLT			
				CGGGCAGGTTA	AATACCCTCATAAT	KALPRKWYKKF			
				AAATCGACTCA	AACAATTTTAAATT				
				AAAAAATTTTA	TATTTTATTTGGAT				
				TTAGAATTTGA	ACGAAAAAAACGC				
				ACATTTTTATAG	AGCCCAGAGGCTTT				
				AAACTTATACA	TTGGCAGTTTATTA				
				AATTCAAAGAA	TTAGAAAGTTTTAT				
				AACGACAAACC	TTGTGAAATATATT				
				TTCTCAAGAATT	TTAACAGAATGAAA				
				TTTGTTCTTAAA	AGTCAATTCGAAAA				
				TGCTAAAGTTA	GAAAACATTCACTA				
				AWAAAGTTACT	AGAAATATTTGCTA				
				AAGGAAGAAAA	GAAAAGTTTGATAG				
				TAAA	AAAAGTCAAAAGA				

					TTATTTTTAAATGA				
					GTTACTTATGCTGA				
					GGCAACAAAATAA				
					ACAAGAAAAACAC				
					ATAAACGAAATAA				
					GAAACATTTTTAAA				
					AATGGAAGATTAA				
					ATTTGATAGCAGAT				
					GTTCGTGACTATTT				
					CGTATAATTTTCGT				
					TTTTTCTTTTTCTTT				
					TTTCTTTTGTTTTTT				
					CGATTTTTATTTTTT				
					ATTTTTTATTTATTT				
					TTTAGTTTTATTTTA				
					TACGGGTCCAATGG				
					ACATCTTTGTAAAC				
					GTTTTTTTTCGTTTT				
					TTCGTTTTTTTTTTA				
					AACCTTGTATTTAT				
					TATGAGGACGTCAG				
					TATGGCAATTAGCC				

					GGATGACGGAATA AATTACAATACCA				
CR1	CR1-6_HM	.	Hydra vulgaris	GGGAATTGGAA GATTAACAATTT AAGAGTAGATG GAATTGATGAA AATGAAGGTGA AAGCTGGGAAG AAAGCGAA	AGTAAACTAMAAT AAAAAAAAATAAA AAAAAACAGATTC AATTATATCAATGA TATATGATAATATG AATAAGAAAAGAA AGAAAAAGGAAAA AAAAAAAAAAAAA AAAAAAAAAAAGA ATAAGAAATCCAA GTGATGATGTCACA AGTGATGATGTCAT AAGTGATGATGTTA CATATTCATTTAGT GTTAATTATGAGAT TTTATTTTTCAATT TCTTTTATATTCAAT TTTTTTTTTTTTTTT TTTTTATATTAACTT ATTTCTTAAACATT TTAGCAACTGTTAA	MKVKKIFEDILSVKNVQI ERAHRTGNKKLKMPRTI VIKLLDFRDKVAILSKSS KLKGKNIYINEDFCFETT QIRKNLREKMKIERAAG KFAFISYDKLIIRDWVAK KSKNLSSTVLKMEENVI FKSSNENDIIGDNDTDSN YFNHKMLESQYYSVSES SQYLGQTKNTFSILNINI RSLNKNFENLKILLDEIK HDFKIICLTETWCKCGET NYEFELVNYTSIHQPREV NAGGGVSIFIHNSVNYIL RNDLCVNETDYESLCVE LVNNKEKNIIINTTYRPP SGNLKKFKTHLKTFLNK ITNSRKHIYLVGDFNINLI NYASNNNAKNFINTLLQ YNFIPTINKSTRITNRSST LLDNIITNNFHNNPLYTG	cd01650	670	948

					GCATCTTAACGCTT	IIQTDLSDHFPTFLATNNI			
					CTTAATTAATTCTT	LJNNISTKSTIFRRQINEN			
					ATTTTTTATTTTTTT	SLKQFKNCLKNCVDWN			
					TATTATATTAGTTA	LILQSNXANNAYDLFLX			
					TTTTCTATTTCATTT	QFCEQYEIAFPEIKKVIN			
					CTAATTGTTTCTTTT	TKSLXNPWMSKGLLKSS			
					ATATTTATATTTGT	KKKEKLYXKYLKKKTY			
					ATGTTACGGTTATG	KNEITYKNYKNIFEKTK			
					TAGAATTTTTAATG	KRSKKLFYDQLLXKTNG			
					GGGCTCGGCGATAA	NTKKTWNLIKDIIGKNK			
					GGCGAAATGTGCCT	RKKNNLPKNLLIDGELV			
					TCTTCTTGCTCCAG	YDKSIIAEELNNXFLNIG			
					CCATCAATTTATAT	PNLAAKIPTNSXPFDSYL			
					AAATGTATATGCAA	KTYDKLMDETNLKPSEL			
					TGTAAACATTTTTT	RTAFSSLKSKKSEGFDKI			
					AACGGCGAAATAA	SVDVVKSVFDIIEPSLFH			
					ATTACACTAAAAAA	VFDLSLKSGIVPDKLKIA			
					AAAAAAA	RITPIFKSGDKSNILNYRP			
						ISVLPCFSKLLERIMYNR			
						LYNYLTENNMLYCKQF			
						GFKKKQFNLRTICCIVNN			
						LVLKKNNSTXHAVVELF			
						NQISNAFNNDCFTLGVF			

						VDLSKAFDTVNHDILIK KLEXYGVKNKNLLWFE DYPTNRKQFLYYNQNN TIKHSITCGVPQGSILGPL LFILYINDLYLASNLLBLI LFADDSNLFYSHRDIKA LFKIVNEELCKVNEWFIS NKLSLNVEKTKFMLFHK PNKSENIPLKLPDLIINKT YIKRESNINFLGVILDEN LSWNSHINSIEKKISKNIX MMYKAKPFLNIXSLKKL YFSFIHSYLSYCNIAWAS TNYTKLKKLYSKQKHA CRIVFGANRNVPCEPLLS ELGALNVYKLNIHQVLL FMFKTKNELSPKIFQSYF NEIXHKYPTKFSDNNFV VPKYNLKLTSYSIQYRG PYLWKKFPXIVNKKKTL RDISLEQFKNESKRLLLF MDLNKFDFKCFF			

CR1	CR1-19_HM	.	Hydra vulgaris	CTTAAGCCTTTGAGCCGATAAGACGTGTTTTTGTTTGAGCGTAAGTTTTCAGGAATTTTAAATTTTTTTCTTTTTTTTTTTTTTTTTCCGTATTCTTTATCTTTCTTTTTTACTTATTTTCTAACTTTTTCTAACCACTACTTCTTTTTCTACTTATTGCTTTATCTAATTAGCTTATTAATTACTTTTCTACTTAAAATTATAGTAATATATTAACTATTCGTGTCTATAATACAGTTTTTTGTTGTACTTCGCATACCTAAATTAA	TGGGGTGTTTCCCCTGCTCGCTGTATGTGTACATTCTATGTAAATTTACGGCTACAACTATAATAATAATAATAAA	MTNLTPAQLKQVEDAIKKLIHAETAKWSAKISELEQVIKQQQLIIDKLIQNPNELTNTTTSSIESSWVTVASKNIKKKQPIEQLKIINTFAIEQQQKKKREKNVIIYGIKESAKLSVADKHIDDKETVQNILVTISSDAKPKSVKRIKSTSSKPGPIIVELSDVSERNPLLTAAKKLKNSDNYKNIFISPDLTEAERYLNYQLRMKRNELNSNINLKTSPFRYGIRSSGVVKLKNIQTTTKHRPLDTVYEAAESSNSKIYKPLHKHYYPSSTLLPANLKYKSKNLHNKKTNYSVNCYYTNATSLDNKMADLISISRTFDYHVIAISETWFRNDSVPNIDGYSLYRKDRQDGRRGGGVAIYITDVFNSHEVNQFYTNLLEQVWVSVKFNNKHILIGCIYRPN	cd01650	791	1046

				ATATAATTCTTA		DASLIKYVAESFKLAHSL			
				TATTTTATACCT		IKKPYDDCLIVGDFNFPN			
				ATACTTATATTT		IKWSNDGIESFNPNNPTE			
				TATTATATATAT		YDFWNIFQNSLFTQHITF			
				CTTTATTTATAT		PTFQTDYGSSNNMLDLL			
				ATTACTATTATA		FTYNSNSILNLSPSTILGN			
				AAAATAGTTAT		TTKGHFSITWRFLVSELF			
				AACTGAGATTTT		TESESQSFNYKKGDYDSI			
				TTTTTATGCTTC		SNFIASIDWEFLFLHKTA			
				AGTTCGCTTAAT		DQMFNILISNYRTACTYF			
				TTTTTTTTTTTT		IPKINKTKKKYDPWVTP			
				TTTTTTACAGAG		HIKNLIREKKSAWHINCS			
				TTATAAAAAGT		TKWENSELKIKYNRLNR			
				TTTTTTTTTTTT		SVKTEIFKAKKAHEKEII			
				TTTTTTTTTTTT		SKSQNNPKIVYKYINKQ			
				TCTTTTTCTTCG		LKIKDRIRSLKDTSGRIIE			
				CCAAAAGTTGC		ALPDIVDLLNSQFKSMFT			
				GTTTGCACCTGT		LESSENFPLLNKSGILSEC			
				TCTTTGTTTTTA		IYNENLDFSISDIIQELAD			
				GTTTTCTGAACA		LKPYKSFGSDGLHPLILK			
				ATAAGCTAAAT		ICAKSFAIPLTKIFHKSFN			
				TTTTGCGCTCTG		SGVVPTQWKLANISPIY			
				ACACAAACTAT		KKKGDKLLAENYRPISL			

				ATTAGTAAATT		TSVACKIFEKLVKKKILK			
				ATATATCCATAT		HINDNKMLSANQHGFLY			
				ATTATATATAA		HKSCVTNLIETFDIISSSI			
				AATACATATAA		ASHQFIDVIYLDFAKAFD			
				AAAAGTTATAT		SVPHNRLLLKLRSFGING			
				ATAAACATATA		KLLQWIKSFLVGRKQRV			
				CAAATAATATA		TINNTSSAWVDVLSGVP			
				AAGAAATTAAA		QGSVLGPLLFILYTNDLA			
				TATATACAAAT		KLTETSEIFADDTKLFSK			
				AAATATGATAA		TLSQSNFVDLQDRLRLV			
				CACGGAAAAAA		YEWSNTWLLKLNTSKC			
				ACACATACTGA		CVIHYGRNNPHYNYFLQ			
				ATCTTTTTAGTC		NNQGINQSLRSSTKERD			
				CCTGTGTAATA		LGVIFTSNLKWSAHIQTV			
				AGTTATTTGTCA		VSKSLQILGMLKKTFTFI			
				TAATTGTAACA		DEKMALKLYKVFIRPHL			
				AATAAATTTGC		EYGVSIWAPFLKADMAL			
				AATAATTGTAA		VESVQRKATKWAYSLR			
				CAAATAAATAA		HLQYCERLKRLNLPSLE			
				ACATAAAAACA		FRRLRCDLIQIFKLIHRFD			
				TCACTGATATA		IVNDKYLPKYINSVSRG			
				AAAAAAAAAAA		HHSKIRREYVLNCLPRY			
				AATTCATCTGTG		NHLYNRSAKYWNNLPE			

						QITLYSTVLSFKNRLDS WLWKKFGTNIYKTGCL			
CR1	CR1-69_HM	.	Hydra vulgaris	TACAAGTTACA ATTTGTGTTTGT GTTTACAAGTTC GAGTAAAATAA	CTGGTTTTTGTTTTA TAACTATGGTCTTA TATATAATTGTAAT ATATATTTTAAATT AACTAGCATTTGTA TTATGTGATTGTAG GGCACGTTGTCAGA GTCCTTCATATTTT ATGGACCTGCGTGT CCTTTAGAACATGT ATCACTGTTCTAAT AAATTTTAATCTAA TCTAATCTAATCTA ATCTAA	NLIKILIANMVNVQSTEF NVYKADQDKLIKSLIKTI DELKNCNLLLTERLNVL EKAKVVPSTSISWASVV SGNMPPKRSTEQLHLMN AVVAEAKEREKRESNV VVFGIEASKKADLLSAR EEEKNSILHMMNSINAKI EIKQIIKLKSKSDKESPFV IVLNNRNDRNSFLKAAK SLKNSTEFNKVFINPDLT VAERYKAKLLREECKQ KNLENSEPSLFYYGIRNE KVIKISKQSFYNCKNQK YFSHGSLTSLVIPKSLQK NHASSISKQRITAVNNKS IKKVICSKRSNFINSYNH LTTNSFNNFTVYLFNAR SLANKLNDFFLYVESNK PALICVCETFFNDKLPNS MVCPKNYSIYRKDRVRI	cd01650	789	1046

						GGGVAIYCRNDVKSAIV SITQTDKDIDIICVDLIFG SSKLRLITCYRPPSYSLD DYIYLESMISIISNLCESV PQFVIVGDFNLPKIDWV NYVSPIEKCHNLFLTLVN NLGMYQYVLEPTRENNI LDLVLSNIISLLSNISVEC PFSTSDHNSVFFYINTFN YYQPETNSESFYDFKNT DIIAFNSYLSKISWDFDFS FVFTTEEYWNIFSNHLR KGIDRFVPIRTIYHNKKS RSYPKYIYKMLIRKSFL WKRWSITKTTLHKKAFI QYASKCKKAISTHHKNV ELKLVKENNLSKFFSYV NKNLNNPKKIYPLKTNS NFTTTNVEIAEVFNKHF GSVFTIDDGILPKINTFVS EHIKMNFVDFSAKIVYQ KLKNLKPSTSFGPDGIPN ILLKQLAHIICIPLSYIFES			

						SFTSSSLPKQWLQATVSP IFKKGSTSDANNYRPISL TCTSCRVMESIVSSSIAD YLNINNLITPNQHGFLSK RSTCTNLLESTNDWNKA LDSNLITDVVYIDFQKAF DSVPHPKLLKKLAMYGI IDNLLHWISAFLSNRYQR VLVGNSLSNPTCILSGVP QGSVLGPTLFLLFINDLP SIVKDLNCSLMLYADDI KLYSSFKDADYSHDLAV AIYKVYLWSKKWQLKI ANNKCFTHRIAPTSFCK DINYNYQLGDHNLTWST NPKDLGVTMDSYLNYS NHISNIVRASNIRGYLILK CFKSRDPRVIVKAYTTYI RPILEYCSPVWSPNRTK MNNTVERVQRRFTKRIS NLNNVSYSSRLSILGLEL LEIRRLKQDMITCYKILN NLVCLNKSSFFLFNNYIY			

						TRGHNFKLRKLKCKLDV RKHSFSLRVVNIWNNLP SDVVNAKNIKSFKIKINSI DFEKYCSG			
CR1	CR1-81_HM	.	Hydra vulgaris	TTTTTTGGACTT TTATAGCGATC GGACATGTTAT AAGAGCTTTTC ATTAAAAAAAA AATTTACGGAA AAATTCAAATA AAGTTAACTAT ACTTTGTAACTT TTTGTGAAAAA AGTATATTTGCA TAGCGTACCAA CT	CTTTTCTTACCGAG TTGTTGACATGTGG AACAATATGACCTC GGAGGCTGTAAATG CAAATAACATTTTA ATGTTCAAAAATAA AATTAAATCTTTCG ACTTTAACAAATAT CTATTGTAAAAGGA ACAACTATTTGTTT CTTTATAGTTTTGTT ATTTTATTTTTTTTT TGTAGCGATATTTT TAAATTTCTTTGGG CACGTTGTTAGTGT CCATCTTTGTATGG GCCTTCGTGTCCTT TAGAACATGTATTT ATACTTTGTTCTAA	MVNVQSTEFNVFKAEQE KLIKALNRTIEELKNLNL LLSERVKMLENAESVKS SPLLSWASVVSGKTPVK KTVEQLNLINVVTAEKK EREKRENNVIIFGINPSK AIDKSSAIKEDKNSIQEV LNSINLKVEVKQIIKLKS KIDNQPPFVIVLNNKTDR NSILKAAKLLKNSSEFKN VYINPDLTEAERYKAKL LRDECKKNNLENSESLL YYYGIRNDRVVKIAKQL LFNNRKLSHMSHCATEL ANSSRISCEYLISKSSKPD KIICTKKCRNKPINLNLIN GNSNNGCNSNNGCSSNP YSKNSVYLLNARSLANK LHEFFMFTEANKPAIICV	cd01650	881	1138

					TAAATATCAATCAA TCAATCATCAT	CETFFTSKISDSIVCPKDY SIYRKDRIKIGGGVAIYC RNDIKSEQVQITQTDTDI DIVCIDLIFSSHNLRLITC YRPPFYSIADVAYIELMT AFIYNLCHSVSQFVIVGD FNFPKMDWLNYVSPNE KCHILFLNLVNNLGLHQ FVQEPTRENNILDLVFSN NISFLSLISVECPFSTSDH NTVHFSININDLDQSNKN IESFYDFVNADIDGFQSY LLNINWDFEFSFLFLQLK IIGTFSPIIFAMVFINLFQL ERVILIKKLGLRVFLFVF TVEDYWNVFSNHFCNGI HQFVPIRKSNLNKKIKH YPSFIGKMLNRKAFLWK RWSNTKNITHKKAYNK YASKCKKSLLTYQMRV ELKLVKENNISKFFNYV NRKLNNSKSIFPLKDKN NNNKITCSNEEIADVFNK			

						HFGSVFTNDNNYSPFINS			
						YVDDSIGIEDCTISTRSSV			
						LVVTKKSLTCLTNTLAV			
						SLQMITITLHLLIVMLMI			
						QLVLKTVLFPPEAVYKS			
						LKEIKASTSYGPDGIPNIL			
						LKKLAHLLCIPLSFIFDAS			
						FKSSSLPKQWRQAFVSP			
						VFKKGATADPNNYRPIS			
						LTCTCCRVMERMINVYV			
						IKYLGHHNLISLNQHGFL			
						NKRSTSTNLLESANDWN			
						TALDNHQITDVVYIDFQ			
						KAFDSVSHTKLLVKLAS			
						YNIKGNLLNWISAFLTD			
						RYQQVKVGNSISNSIRIV			
						SGVPQGSVLGPTLFLLY			
						VNDLTAIXNNLDCCVKL			
						YADDIKLYSSYCDMDLS			
						QDLVVALNRLLQWAET			
						WQLTIACSKCFALRIAPP			
						SLCKGIKVNYKLGEYSL			
						AWSSNPKDLGVTMDCN			

						LNYKKHIFNVIHQANSR AYLILKCFKTRDPCILVR AFTTYIRPILEYCSPVWS PHQIGLIKSVESVQQKFT KKIRSLRNLSYHNRLQL LGIESLEVRRLKNDLVTC FKIFRKLVCTDQSMFFAI DNNSHTRGHNYKISKQS CRLDIKKI			
Daphne	Daphne-78_HMa	.	Hydra vulgaris	TGGTAAAGATG GCGGAGTTATA GATAGAATTGT AAATTTATCAA ATAAAGTAATA TATCTTTACATG ACAGCAATTAA GAGGAAAAATA TTTATAGAAAA AAAGACTAAAA AGTAAAAAAAA ATATATCGGAA AATTAATTAGA AAACCAAAGTA	AGACTCTGAAATAT GTTGTTATGCTTTT ATTATTACTACTYT CTTGATACCTACTA GTTTTATTATTAAC TTTTTGTTCTCATAT TTTTATTGTTATTAT TAGTATTGATATTA ATAATATTCTTCTA TCTTTATTAGTAAA TTGTTATTTTTATTT ATAACATAATAAGT ATTATTATTGCTTTT TTAATTTTTTCTTAA	MLIRNLCKFCNKTVVDS HKGILCNYCSSWLHPKC NYINKHTYSLIEKDSTD WYCSDCINKNLPFCSLS DFELSLTLSSQTIPTNNL KSLKSPPHFKKLFNNINK ISNKSINCKYYDILELNK VLNNNCNLYLHLNIASL PYHIDGLRSLISSLTTLPI VIGISESRLNINDSSKTDV NINGFTIEHSPTEAKKGG ALLYLRSNLNYFIRSDLI VYSPKYLESIFVEIVNPL KQNIIVGSIYRHPSMNSN	cd01650	635	891

				CTTAAAGTTTAG	GTTTATTTACTACTT	EFITDHFNPLLERLSLEN			
				AGATGTGTAAG	ATACTATGTAATTG	KQVVLMGDFNMDLLNY			
				AACTTCATAGA	TTATTACCACTATT	NESKVISSYLDTLCSYSF			
				CTCAAATAAAT	ATTGCTATTATCTA	FPSIIQPTRITATSKTLIDN			
				ATAATTAAAGA	ATATTGATTGTATA	IFLNFQTPDLISGNLTISIS			
				ACAAGATACAT	ATATCTTTATTATT	DHMAQFICIPCLSPKKTK			
				TACAAAATTAA	ATTGTTAGTATTAT	EVIFRRTFQNFDNKKFIK			
				TAAACTAAGAG	TACTAATAATATTC	DISNADWQIQITNNDNV			
				AAAGTAAAGAA	CTTTTCATCGTCAT	NESLHYFFKTFNKILDQH			
				AGAATCAAGTT	CTTATTATTATTGTT	APYKKLTAKQIKLKSKP			
				AATGAAAGRAA	AATATTATTTCTAT	WITNGILKSISIKNKLYK			
				AAGATCAAGAG	AATTATTATCAATA	QYIKAKNHNIKIHIFNKF			
				TTTTTTACATTT	CTATTATTATTGTT	KWYRNNICNLLKLSKKS			
				GTTTTGATTATT	ATTATAATCACTTT	YYINFFNNNLXNVKNT			
				TTTTGCTTTAAA	TATTGTAATGTTGC	WKGIKEIINIKPSSFSSKSI			
				TATTATTTGTAA	AGGTTTTTACTTTA	NLKFNGNTISDSKAVSNI			
				ATTTAAGTATTT	GATTAGTACTTGTA	FNNYFVNTQQTLLTKIV			
				TCCAGGATTCGT	CTATTTGTAAAAGT	SPKCNFSDYLNNPNINSF			
				TTTTATTATTGA	CCTTGAAATTTTAC	FIDPVTENEVSNLIKSLQ			
				CTTTTTGATTAT	AACTTATTTCAGAA	SKKSLGINSIPTFLLKLVS			
				TATTATTAGCAG	TATATATTGAATTG	HIISKPLCTMINSSFENGL			
				TTATCATTGTTA	AATTGAATATATAT	FPDVFKIAKVIPLYKKGS			
				TTATTATTATAA	A	RTDHTNYRPISLLSNISKI			

				TTAAATATTTAT		FEKAMHNRLYNFLDKY			
				CATTATTATTAA		QCLYKHQYGFRSKHSTT			
				ATAATATAGCA		HALIEITEVIRKALDDKL			
				TTATTTATCCTA		FACGVFVDLQKAFDTVD			
				TATTTATCCATT		HSILLSKLEYYGIRGISLQ			
				ATTTACCTCTTA		WFTSYLYNRFQFVSING			
				TTACAATTTACA		NTSSLVKCTTGVPQGSV			
				TCATAATTTATC		LGPLLFLVFINDLNLSLK			
				TTATATAAATGT		YTTSYHFADDTNLLLIN			
				TATTTACCTAAT		KSLKKINKYINHDLSNLI			
				ATTACTACTTAT		QWLRSNKLSLNTSKTEII			
				CTCCTTCTTTAC		IFKSKQSKISKHLNFRLS			
				TTATTATTACTT		GQKINPVMYLKYLGIKI			
				ACCTTATTATCA		DSNLTFSQHLKDLSIKLC			
				TCAAATTATTTA		RSNGMLAKIRHYVNLET			
				AATAACATATA		LINIYHAIFGSHLRYACQ			
				TTATTTATATTA		VWGQIRTQEILRLSNLQ			
				TACTATGTCTAA		NKALKIIYFQHYQSNSD			
				AGCTACTAGAG		HLYFLSKILKIDDLVQLL			
				GTATTAATATAT		NCHFVFNYKQKCLPSTF			
				CTAAATTTATAT		TNFFTVRESCRYHLRSA			
				TCTAAATCAAG		NSYNLHVPNHSSVKYGL			
				TTAATAATATTC		KSIKYQSVKSWNSLPSPL			

				TAATTTTATCTA GCTGTAATACTT ATTCTCCTGTGT AAAGTATTTAA ATCTCTTATATC TTACTATCATTT CTTTATAACTGA TCATTCCTCAAT ATTTAAAATAG CAACCATTTCTA GATTAACATTTC TTAGTTTTAATC TCTCAGTTTCTT TATAAATATAT AGTAGATATCTT TATTTAATTATT TTTTCTCTATAC TTTGTTTAAGCA TAATAAGTC		KSIDSITNFKNCLFNYFL ARYYT			
Ingi	Ingi-1_HMi	.	Hymenolepis microstoma	GGGGTGTGCAA TAGACTTACCTC CACGTGGTGCA CCCTGGGAAAC	TAAAGCCCGAACA AAGGAGAAGCAGT GGACAGCGGCACTC TCCAACATAGCCGA	MTSLKRNSNRRNRPCQG EPQLLQILQWNAGGMSQ DKKIQIQKILDTYNIDIFT IMESNLTDDKLTYYQFP	cd01650	521	782

				GTATACTCATAC GGCAAAAAGTC TATAAATCGGC AGTTGACTTCA AACTACT	CTGGCCTAGACTCG AAGCTGTTGCAGAA TTTAGACTGCGTAC CGGACACGATTGCC TGGCGAAACATCTC CACAGAATAGGAG TATACGCCCAACCA ACATGCCCCCTATG TGACCTTCAGGAGG AAATGGAGAAGAC CCATCTGATACGGT GTCCAGCCCTACAG ACAGCGACAGAAA CCCAGAGATACTGG GAAGCCAGAAGTC AACTAAAGGGCTA ATGATAACTATTAT AAAGCCTGCCATTT GATATAATAATAAT AA	GYTLHLLPKDRQVASGI LTGVKEGLTSHCEIIKSM GSMQDICEIIRLNVWKS QNHFKMYAVYNPPQNR PNFDLLNISHKTIVLGDF NAHSTRWGYKNRNTAG KEIEDILNSSPLELIYSDE DPATYLHYNGTRTTPDL LLVSSDISELTQRKIIDDP GSGHKPVIASITINSKSM TPKMPTKVLWKFKKAD WPKFTNLLETELNASPIN YNQHPDKLCXNITNIIIK CAKKTIPRGKVKHYRVF WSKNLEELKRKREVLRN TAEQTGKTEDVQAWRR QSAVLRQAILQAKRTTF NSFISNINYQNDSQRTFK FLGNLLNNKKKPKKEPI RLNNKLLTTDSEIANSFA RFYSXKQKKNPFVRNSS RDIRKQVTKFDEIKEDTA IPHKVLTEPFSSHELNAA			

						IKQLKCKKSPGEDCIHPE FLIRMGPKAKEILLTLFN KIWETSLVPTQWKVAIVI PILKKGKDPSNFDNYRPI SLTSMLAKLMERMVNT RLTWFLETNNILGNEQA GFRPQRSTNQQVATLSQ HIKDALDARNTLTAVFI DFKSAYDLVWKENLILK LAKIGIRHNMLNWIKAFI GQRTCKVRYGNDLSKS NLLQTGLPQGAVTSCSL FNVFINDIAELVQTVTGV KCLLYADDLVLWYSAP KKNAQKRTEIALNYALK LLANWCDNNGMVINTS KTAYQTFSLAHHSINPHL MYKDTPLEQTNESTYLG VTFDTKLTWKNHIAKIA ERASNRLNVLKRLAGSL WGSARSTLNTTYKMFV QPIMLYCCETLVTASEAI LKPLEKTHNQALRLITG			

						GIKSTPIDAMLLVTGNST IRSSVEEKALTLYEKLLR IPSDTFWSSYENRPRHLK TQVGLIQKAIELKKRLQT TDKPESLSLPQNPLADNE VQFCTQLLGHSMKADTP PDQMRLLALETINANYP ADQWLQVFTDGSYVEN QANVGAGVYSELFTFFA AAGQNRSAFEGEIEAIKI ALGQLCCRDTKFTNAVI LSDSQSAIQSIGNREPPK TAEIHECKKLYQLLKEK NKTIVLQWIPGHCGIIGN ERADTLAKKGTTILQANI FLHNEDLNQERIQDSKK QRT			
CR1	CR1-1_IS	.	Ixodes scapularis	CATTTGGCCGG CAGGGTGCGTT TTTCGCCCTTTT TTCGGCATTCTT AAAGAACTCTC GGACACGGACA	TAGTCACGACCTTG TTTTGTTGCCTGCA TTTATCCTTGATGA CGAGCGGTGTACCA CTATTTAGGCCTTG TGCTGTTTTTGGAC	MECTLYYQNVRGLRTK SSEFLSGVIGNSYTIISLT ETWLEDTIPSSQYFPPCY SVFRKDRDYAATRQRY GGGVLTAVDTSLCARRR YDLELYSECVWIEVTCR	cd01650	493	745

				CCTCGGTCGAT GTAAACGCATA ATCAGGTACC	ACCCAACAAAAAT AAAGATTATTATTA	DGLNFLIGNYYFPPHLSS DVFARHFRDLEGKLDVT KFRVHIYGDFNLPGIDW LANYNLITSDVTSLKASS LFDFIYFNGFEQRNTIQN SAGNVLDLVLVSTPINYI SPITDPLTRVDHFHPPFA VSFCFPLRKQALCSYEY LLYSSGDYLNMYKYFQ CYDWKPILNNKCADSA AEALTEVVSSAINDFVPK RVSRLNKYPAWFSKELR SCLRRKLHYHRLFKRSG LDKWYAKFSECRALAK WLFKKDESAHLHGLESD MYAKPQNFWKYVPTKK KRNSSVAPMRNSSGYTC DSIEIANMFAEHFKSCFS SASVPDNNATCHGTSDF LSFIPIDEPEIRESIRNMKS KLSAGSDGIPSFIVKGCA ELLTPVLLHIFNLSIHSGI FPSIWKRSVVVPIHKAG			

						DPSTVNNYRPVSLLCSFS			
						KVFEMIMHKRLYRHFRH			
						KMNPEQHGFLKGKSVET			
						NLCSFLDYAVPLVCNRE			
						QVDVIYFDMTKAFDKV			
						NHSTLLTKLDSYGLCLT			
						YCQWFKSYLTCRTNYV			
						RVSGSMSDYFSSPSGVP			
						QGSNLGPLLFLMFVNDIS			
						LCAKNSKVLLYADDIKL			
						FRRVQSHSDCVLLQEDA			
						LQISQWCAKNYLPLNEQ			
						KTRTMSLTRKRDSLAFS			
						YTIGGSVITRVSVIRDLGI			
						FIDSSLNFNHHVLTICNA			
						SFRNLGAISRFTRKFSSP			
						VCLVRLFCCLVRTKLEY			
						ASSVWNCINLTASCCIER			
						VQKRFIGIIYDRYFDRTC			
						YYSYEQLLKKLGLCTLL			
						DRRRLRDLMFLHKVVN			
						GAVDSPLLLSTVCFHAP			
						LRSNRFQSPFYPSALNVA			

						SPLTRMQLEFNKLDQSAVDIFSDLIVFRHRLESIICT			
Vingi	Vingi-35_LMi	.	Locusta migratoria	AGGATGTCGGATGTGCCTTTAATAGCCAGGAGATCGTTCTGATTGTCTGGCTGCCAGTTAGCGCTAGTTCATTTAGTTCACGTTTTACAACGCTATCCGGGGTGCACCACGGGGAGGCGAACTAGCATTACACCCTTACAAACGCCCGACTTATAGATAGCACGCTTGCCGCAGAACATTCCAGGTGACAGTGGGGGACCAAATATGCCCAAAGAACCCTAAACAA	TGACTGTGTTAATTTATTCTTTGTGTGCTGTATGCATCTTATATACCTTATTTTATGTCTCTTTTGTATTCTTGTATGTTTTACACATTTTTGTATTTTTTGTATGTATAGTTGTATAAGACCTGCTAAATGATGTGCACTGCCATACGATAAATAAATAAATAATA	RRGIIFFISYAAPLGGSSADTLRTTALGLVYSTAEYCAPVWMNSAHTRKIDVQLNGTMRIVSGTIRSTPTSWLPILSHIPPADLRRSNALLREYRKITNNPRLPIHHDTPDLERNRLRSRHPTMKTAKELHNTDFNLIESWKERWRQACPEQCQALPCITERPPGYELPRKAWSSLNRIRTNHGRCADSLYRWGKIPSPNCDCGADRQTIRHIIEDCEMRAYDGNVSDFLTAKDAAIDYIRNLDVCI	NA	NA	NA

				CGGATTACCAC					
				AGGGCTCCGTC					
				CTAGCCCCATTG					
				CTATTCAGCATG					
				TACATCGCTGA					
				CATGCCCCAAA					
				CAACTCCAGAA					
				AATTTGAATAT					
				GCAGACGACTG					
				GGCTTAGCGTC					
				AACATAAGAGC					
				TTTGAGACAAT					
				GGAGTGACCTC					
				GTCGTCCAACTT					
				GGCGACTGGCG					
				CATATTTCCGGG					
				GCCGAGCCTAC					
				AAAACGGGACG					
				TGCTTCCATATC					
				TGTAATAAATC					
				AGCAGGCAGAC					
				GACGACGTCTA					
				CTTTGAGGGAA					

				CTCGCTCCGGC ATAACGGTATC CAAAGTACCTC GGAGTCACATT AGACAGAACCC TCACCTATAAG CGACATCTGGA GAACACTGCAG CCAAACTAG					
Vingi	Vingi-24_LMi	.	Locusta migratoria	CGACTGTTCGC GTCCGTAGCAA TCCTCAAGCTAC ACGTGTCCCCTG GGTGGTGCTAA AGAACAGTCAC TTCGAGGCACG AAGTCAGAGCC AGAAGTCACGG CAGCCAGACCA CTTGTGGACTGT CCTGGCTATCA AACACTTGGGC TGATTAATGCA	TGACTGCAACATCA GGTCTACCCAGGCC AAAAACAGACCTGT ATAAAGTGACAGAT GATGCAATTTCATG GCTGGATGGTTTAG ACATTAATATATAA GGACCATGTTCGCT GCTTATCTTTCTTA ATTCAGTTAACATA AATTGTGTAAATTT ACTGCTCATTTCAT TGTTGTTTTTCTATT ACTTATTTTAGTTTC	MIMKKKQQTPESVLTRE LGSTTRICQLNVEGLSSS KSQFLSKFLMENDVDVA VLQETHIENXEQQRSRG KIHGYDLLGAVYHRHY GLATYVRSSIENAALVSL SADCFQQIVTKIGELTVV NIYKPPQTEWPQNVIETH PHPTIYAGDFNSHHTDW KYRTTDINGEKLSNWAE ENHLHLVFDAKDKNTFK SAIHNTETNPDLSFTTRN SQSHPIPTRRKIXESFPHS QHRPILLEIGINIPLIRSLP	cd01650	481	707

					CTAACTTAATGTTC	RPRWNLQKANWDKFSN			
					CTTTATTAGTGTTC	DIDQTLRWIPARWYNYD			
					ATTCAATTGTTCAT	RFVNIIIAAAKRYIPRGY			
					CTTACTGTACGATG	RKEYIPGWNEETEELHQ			
					TATATATCTTTCAT	RFLQENDAEIADKLLRN			
					AAATTTGTATTAAT	LDVARRKKWEETTEAL			
					ATTTTATAAATATT	NFTKSSRKAWNLLRKLG			
					TTTTTCCAAATTCC	DASAPEREVPGVSPNQIA			
					ACTGCATCACACAT	SRIVSLSRAPRDKKRTTF			
					GTAATGGACTTTTT	VKRHLKSFRQSVEAPNK			
					ACTGCATACCTGTA	ISRPFTVEEVNTAISKIKK			
					TACTTCTCTGTAAT	GKAPGFDGVLPEFLIHSG			
					TAATGCACTTAAAT	RRTREWLSQFFSHIMQT			
					TTTGATATAAGCTG	GLVPPXFKKAKIISLNNIF			
					TTTGTTGTCTCTGCT	VKPGKPKDKPESYRPISL			
					TCCCACCTGGAGAA	LSTTYKLLERLILGRIEPI			
					GTGTCATTGTGTAA	LEEHLPVEQXGFRHNRG			
					TGTATCTTCATGTC	CEDQVLSLTSHIEAGFQQ			
					CAAGTGTTTATGTC	XLKTTVAFVDLTVAYDT			
					ATACGATAAATAAA	VWKDGLLYKLYRTVPC			
					TAAAT	QKLLRLFNNMLRDRTFQ			
						IIMGDKASRVRKINNGLP			
						QGSVLAPALFNLYISDM			

						PETKGRKFGYADDLAIA VQAKXMEDADTLLSQD LEVVGKFYADWRLIVST SKTETTSFHLTIGKQQNL XKYFSTIKCSPIISSQSILV SHWIAPSPSGNTWKQRQ PKSTLAITSCTNCVAAG GEHMLPLFSLNNREATK SPKVLLNNKVLPHNQFP KYLGVTLDRTLSFRKHL EATAAKVNTRNNILHKL CGSRWGAHASTLRGSAL ALVYSAAEYCSSSWLNS SHVSKLDARLHDTMRTI SGCIKSTPVQWLPVLSHI APPPLRRKGCLVRQLKRI QSDPDLPVHKDIPHLSLR RLKSRKPPLVSAQQLQD ADFNITEQWRTDWSTVN FTALWPDPTERPTGSNFP ADNGKRLTGSEHLTESV RILCTSGAYRKSSVRLW CFASDNRPYCV			

Nimb	Nimb-38_LMi	.	Locusta migratoria	CCCCTCAGGGG CTCAGCATTAAT TTGCTGAGTAC GGGCTTGGCGA ACCCGGGGTCC TGAGCCGTGGG CTGTGTCGCGCC AACAGTCCGCT ACACCGTAAGC TCTGTGCGTGCT TCAACGACCAT CGTTCGGCGCG ACGGTGGAACG TTGTACGTTTCA GAGCCCGGGGA TCTTGGCTTAAT CGCCCGGGTCG CGAGGACGGTA CAAACCTCTAT AAAAAAAAAAC CTCAATCTCAA GGTGTGCTACT GCCGAGGAGAT	TAGAAATTGGTCTT CAGGAGAGGGACC TCCGATTAAGTTAC ACCAGAAGGTACTT GGGCTTTGGAACAC TGAGTGGGTTGACT CGCCTGCCCCAAAC AAGCTCCGTGCAGT GAAGCTTTCTACAG CTCCATGGAAGTCC TCCCTGAGGGCCTC TCGGAAGGAGTCA GTTGTCCTTTGTCG CCTGCGTATTGGTC ACACGTGGATGACA CATGGCCATCTACT TCGACGGGAGGATC CTCCACAGTGTCGC TGTGGCACTTTATT GACAGTGGTCCACA TTCTGTTGGACTGT CCCCTTTTAGCTGC CTTACGACAGTCTT	MRCRSSHPSILQWNCGG FFHHLAELQQLMCHSLL CIAYRKRGFLFRGPLPYV ATGAISKIAPSVTGSLEV CASMSLPLSPVRCAPHH ESQGLWRCVPLCPCHCP QCAVPLTTNLEAVAVRI RVRQELTICSLYLPPDEV VCLADLVPWSLNFPRRF SFWGILMPTTLCGVGRR PRLVVGMPRALVAQLPP PFLLLGDFNAHHPMWG GTQTSARGRDVEMLLTQ RNLCLLNTGAPTHFSVA RGTYSAIDLSICSPSLLPH LNWNVHDDLCGSDHFPI FLSLPETQAANRLPHWS LHRADWPAYASAAVVT TPLPAGIEVAVQQVTKAI VSAAEKTVPTASASLRR KSVPWWTQEIAEAIRER RRALQRTGDTLLGHPGT SARASAHRRHPSVENLI	cd01650	568	826

				GCATGGCTGTT	TTGACCTGCCAGCT	AFKRLRARARRLIRQSK			
				GAGGTTGAACA	GCGCTTCCTTTTAT	KDSWERFVASIGSRASPS			
				GTCGCTAGCGA	GCTTGGTGACGACG	RVWTQVRRLFGYRPPTG			
				GCAACCTCTGA	CTGCTACAGCTGAT	IPSLTIQGALCTDPQAIAE			
				GGAACCTGCCG	CGTGTTTTACGTTTT	YMAEHYARCSATETYS			
				CACCTCAGTTGT	ATCCGTGCAGCTGG	ADFRTFKERVERKQLSF			
				ATAAGGCTTAC	GTTTTATCATTTGA	RTRRQELYNAPFSEWEF			
				CCAGGCAAGCG	TCTAAAGGTGGGTA	ETALAACPDTSPGPDRIH			
				GGGCTCTGCCT	CCTCAGCCTGATCT	SQFLRHLSAEFQHQLLSL			
				GGGTGGACTTA	CTCAGGCACCCATC	FNRIWADGVFPSQWRES			
				GTTCCCTAGCTG	TCTCACAGATTTTT	TVIPVLKPGADPLVADS			
				CTCGTGGGACG	ACTCTTCTTATTTTG	YRPISLTNVLCKLLERM			
				AAATGAAACGA	TGTTGTCTGACCTG	VSRRLMWILEDRGLLAP			
				GCCAAATAGCG	CTTTTACTCGTATG	QQGGFRQGRSAIDNLVS			
				TGGTGGACGTA	TTTTAGAATGTGCA	LESAIRTAFTRREHLVAV			
				AACCACCGTCT	GAGTGGCTGGCTCA	FFDLRKAYDTTWRHHIL			
				CCAAGCGTGGA	TCCTTTATTATCCTC	ATLHGWGLRGSLPIFIQN			
				AGAGTTGCTTTA	GTGATCAGCCAGCC	FLSPRSFQVRVGTCLSSP			
				ATGCAGCGACG	ACAATATTCTGCTG	HIQENGVPQGSVLSVPLF			
				TATGATCCCCG	TAACTTTTTAGTCC	LIAINGLAAAVGSSVSSS			
				AACGTTCCCTTC	AGTTATGTTTATTT	LFADDFCIYFSSATIGVA			
				CCTGGCCACCC	CCTCTTTGTGTTTTC	ERRLQGAIRKAQSWASA			
				CTGGGAGGAAC	TTTTTGGTGTTCCG	HGFHFSGSKTRVMHFCR			

				GGCGGGCTAAA	AGCTAGTGTTCCTT	RRALHLDPELYLGNELL			
				GACAGGGATGA	ATATTGTGTACCCA	RVEETHRFLGMVFDARL			
				GCTTACTCCCCA	TTGTGTGTTTCTCTT	TWQPHLRQLRQKCWKH			
				AGTATCTTGTTT	CTGTGTGATTCAGC	LNILRCLGHTAWGASRP			
				GCAGTCGAACA	GCTCGTTTTTTCAG	TLLQLYKALIQSRLDYG			
				GATGGGGAAAC	CTTGCATTACACTG	SIVYDSAAPSALQVLDP			
				TTTCATGTCGAC	TATTGTCTTTGCTC	VLHSGVRLATGAFRTSP			
				GAAGCCACTTTT	GCAACAGGGGACT	VTSLLAEAGVPPLRFRR			
				CTTCGTGAAAA	GATGACCAGCAGTC	QRLLASYAAHVHALPH			
				CCTTGAAGATA	TAGTCCCCTCTAAA	HPNYRLLFPDVALPSPRR			
				AGTTTGGTGAA	CAACAACAACAAC	RPRSGLPIAVRVRSLLSS			
				GTCTCGGATGTC		LESLPIPPTFHYPSRAPP			
				AAGATGAGGAG		WSVPVRSFGIPSYSPHFS			
				CGGTTCGGTACT		LPFPCSSLVRTRPQLRLD			
				GATCGAAACAG		LSRDSKASVAPLNLRRQ			
				CGTCATCCGCCC		FQAVLGSFHGAEIVYTD			
				AATCGAAAGCA		GSLVDGRCGYAYAHDE			
				TTGCTTGCTTGC		HIEQHSLPCGSSIFTAELI			
				GATAAGCTGGG		AIRRALEHLRACPGESY			
				TGACGTTCCGGT		VICSDSLSGLQALDQCFP			
				CAGTATAACAC		KHPLVRSIQESVTFLFSL			
				CCCACAAATCG		AALWSSFGPRDTLEFRG			
				CTAACATGATC		TNAWIGWLSRQLLQAP			

				CAGGGTCTGAT CTATCACAGGG ACCTTCTTTTAC AGTCGATGAGA GCTACGCGCCA ACCTAGAGCGA CGGGGTGTCCA CTTCGTCCGGCG CGTGCATAGGG GTCCGAAGGAT AATAGGGTAGC CACCGGTGCTTT CATCTTGGCCTT TGAGGCGATAC ATTGCCCGAAA AGGTCAAGGTC ATGATCTACCG GTGTGACGTAA AGCCTTATATCC CTCCCCCTATGC GGTGTTTCAAAT GCTGGAAATTT GGACATATGTC					

				CTCCAGATGTA					
				GCTCCAGCCCC					
				ACATGTAGGGA					
				ATGCGGGCGTC					
				CATCACACCCG					
				GACGCTCCGTG					
				TACTCCTCCACC					
				TATCTGTGTAAA					
				CTGTGGGCAGC					
				AACATCCACCC					
				TGTTCTCCAGAC					
				TGTACCGTGCTC					
				CGCAAAGAGCG					
				AAAGATTATGG					
				AATCAAGACCC					
				TGGATCGACTG					
				ACATATACGGA					
				GGCACGTAAAA					
				AATATGAACGG					
				CTTCACCCAGTT					
				TCCATGACGAC					
				GTCTTTCGCTGC					
				AGTTACATCGTC					

				GCCACCCTTCTG ATAACACCTAC CGTGCAACAAC CTTCTGTGGGCC CAAAGGGCCGT CAAAGCTCATC TGCCTCCTCAAG GTCGGGGGCAC TCCTCTCACTGC CCCCCACTACC ACCTGGGGGGC AACTCCACCCC CACCATCGGGG ACACCAGTCCC CACCTCCCAGC CGGAGAAGGAG CAGCCTCCTCCG GCTCCCCTCGCT CGGAAGGGGTC CCTCGGGGCAC TCCCTTCTGAGG CTTCTCCTAGTG CTCCAGCGATC					

				CAGCCAATGGC TGAAGGATCCA AAACCTGCTGG TAGAAGGGCTT CCCGCTCTTCTT CTACCCCTGGC GCCGCTACAGA GAAGCCCTCCC GGCCCTGCAAA CCTAAGGGCCA GCGAGAGGGAC GGAAGAAGAAG ACAAAAGACTC TTCGGTGGCTCT AAGCCACCGCT ACTCATCTCTAC ATCGGACG					
Vingi	Vingi-46_LMi	.	Locusta migratoria	GAGCGAAAATT ATAGCCTTACT MAACCAGTAGC AAACSAAGTGA TAA	TAGTTTGTTACACG TGTAATATTATATG TATCTTTGTTATTA AAACATCAAATAAC TGTAAATTTTGTGA TGTAGCCATACGCT	TKTWCMMGGFPLTLLFI CSQTSNRTVIXXEQAGF RPNRNCIDQVLSLTTFIE AGFQNKEKTSVVFIDLT AAYDTVWRPGVILKLLQ AIPSIKINHLINEMLSNRC	cd01650	26	233

					AAATAAATAAGAAATA	FQVIMGNNISSPMKLNNGLPQGSVLAPLLFSLYISDMPETKSRKFGYADDWALATRHKDLETTESILTNDLSTIGEYFRKWRLQPSATKTEVSTFHLSNKLANKELTVHFEGKVLRHNRNPKYLGVTLDRTLSYKQHLAKTAAKIKTRNNILQKLCGTSWGSSASTLRTSALGLVYPVAEYCSPAWLNSSHTKLVDAQLNHTMRLITGTIRPTPTFWLPILSHIQPPSLRRKNALVSEYKKLQKNPHLPVHEDIPSIRKARLRSRHPPLETARTLVTNSFNTDDCWRQTWQINATPEQLKMPCISTKPPGFDLPRKDWSALNRIRTNTGRCADLLYKWKKIASPSCDCGAEKQTVYHVVQECPLRAFAGDQTEFLMVTEGSLDYIHGLDISL			

				ATGTACATAGT		MMNSRDAAHRRFKAST			
				CAGAAGTGCTC		KPEHYEEYRKLRNKVK			
				AACTGAACAGC		QCIRNAKIRYARSFVCSN			
				ACATGTCCTCTT		TNPATLWKNLRSLGVG			
				CTCAATCGGAA		KATPESVFQISADELNEF			
				TTACAGTTACTC		FSAPINCQAQINNRPEES			
				ACTATCAATAT		PSHITNDDTFHLKHVTIN			
				GACACGTCATC		TTRKAIMRISSEAIGNDG			
				GTGATGGGAGA	TAAAATATGTAGAA	ISITMIKNVADSLLPVLT			
				CCTTAACATGA	CGCCTGGTTAGGTG	DIYLISLSFSANQHIACVS			
	Kiri-		Locusta	CTTGTTAAGAG	TAAGAGAGGGCCT	KALEYIVHDQITEHLCKF	cd0165		
Kiri	5_L	.	migrator	ACACTCCATCC	GAAGGCCCTAATCT	SLFDKYQSGFRKNHSTN	0	173	410
	Mi		ia	ACAATAAACTA	TGCCAGGTTAAATA	TALIKVTDDLKFAMNNR			
				AGAGATTGCTC	AATGAGTAAATAA	EATILTLLDFSKAFDTVN			
				CAACAACATGA	ATAAA	FDILLNKMRQLKFSNSA			
				ACATTCTGCCAT		LKWFESYLKNRQQSVV			
				TAGAACCTACT		CGNEKSSWRNVVSGVP			
				CACCATACGGA		QGSVLGPLLFSLYVNDIS			
				GCACACTCTTAT		SVLTFCKYHFYADDLQL			
				AGATGTGATCG		YQSTRPEDINTAIAQMN			
				CAACGAAAAAG		EDLSSVVTWAKNLGLKL			
				ATCGACAAAGT		NAKKSQVILVAHQKLISS			
				ACAGATCTAAG		ELRDQLPPIQLDGIPIPYQ			

				TCAAACGTCAG		KTVKNLGVIIDEHLNWA			
				CCCGCCTCTCAG		DNTVTVCRKTSACLYAL			
				CACACGATGCA		KKFRNIFPQDVKCKLVQ			
				GCGCCAAAAGT		SLVLPNLHYCDMIQHGT			
				TAAATCACGTT		SSENTRRLELTMNACVR			
				ATATTACTCGTA		YICNIRRYDHISASYTQL			
				GAAACGTGAAA		GWLRPDKLRDYHTLCLL			
				CAAATAGACCG		HRLLSEQHHSTSPQNLTI			
				TGAGGCTCTAA		FRAIIILTQDPTYLVPLCP			
				CAGCCAACTGC		FIKRKHLQTPSLSLLFAS			
				TCAGAAATCTC		GTSCPALFARFSLLLRLR			
				ATGGCATCAAG		KSNTSLCHPLKTLAKTH			
				TAACCAGAGAA		VYSQVSLCVIVKPMLFSP			
				CCCACGATTGA		LFLFLLFLSPNQCSSLRY			
				TGGAAAAATTT		FSSSSFFHHNLLFRYISLC			
				TTGAACTTAGA		EYYYTHFSPPFPPLPPCP			
				ATAAACTTATTT		LLLAAIIKTFPYCNTCHV			
				CCTATATGACA		SYLSNTNINCVYVLLLL			
				AGCATGCCCCC		VLIENFNFCRFSITSVILE			
				TACAATCCGTGT		M			
				AAGAAAACCCC					
				TGCTCCATGGCT					

				GACTAAAGAAT TACGTCAA					
Vingi	Vingi-16_LMi	.	Locusta migratoria	TCGTTTCAGAAC CTGGTCTAACA GTTACTTCAGTA AATATAGAAGG CATGTCACAAT CCAAGCAACAG CTGCTTCAAGA ACTGTGCCATG AACAGAAGTTT GATGTACTGTG CATACAAGAAA CTCACAGAGAT ATCCATCAAAT GCGACCGACAA TACCTGGTTTGC AGCTAGCTGCA GAAGTAACCCA TGCTAAACGTG GAGGTGCTGTTT TTGTCAAACCA GGAATACGAGT	TAAATAATTGTCTT GTAAATATGATGTA TTCACATAACATGT AAATATTGTACATT AGTTTATACATTGT AAGTGCTTCTGACA CGAATAAATAAATA AATAA	PHQNLLRPFTQAEVDLA LSKCKNGKAAGLDDIRV EQIKNFGPITKAWLMEL MNNCVQSCKIPKLWRK AKVIAIRKPGKDQNDPK SYRPISLLCHLYKLLERLI LHRLMEKLEEYTYHTG WLTGEELHISSDETYTA HRRWVSKERDHRSCIYR SDSSLRYCKQKIFLSNFR DKEADGDRKRMDYPRE SVLAPALFNIYTNDQPLP EGTQSFIYADDCAITAQA HCFETVEENLSKALREM SAYYRENQLRPNPGKTQ TCAFHLKNRQAARTLKI TWENTTLEHCTYPKYLG VTLDRALTYKRHCLNTK QKVAARNNVIRKLTGTT WGAHPNTVRTSSLALSY SAGEYASPVWFKSSHAK	SSF56672	19	256

				CTTGAGTGCAT		IVDVALNDTCRIITGCLK			
				ATAACTAAATG		PTPLNKLYSLAGIAPPDI			
				ACAATAACATC		RREVAAKSERGKSETSE			
				GAGATCATCAC		AHPLYGHQPANQRLKSR			
				CTTGGAACTTA		KSFLRSTEALVGTSQQC			
				GTAACAGTGTG		RTELWRARYTGTAGWF			
				ATCACTTCGGTG		TPCEKLPPGYTEKWSTW			
				ATAAGCGGTCC		RSLNRLRSGVTKCKSNL			
				TCCTTCCGAAA		VKWGFLVESPLCECGAT			
				ATTCTCCTTGCT		QTTSHLLQCTQCPATCT			
				CCTCTAAAAAC		PEDLLHATPNAIQVANF			
				TTCATTCAGGA		WASKV			
				AAAATATCTTG					
				GTAATATGGGA					
				TATGAAAATGT					
				GAACCCGTTAT					
				GTCTCGAGCTG					
				AAACCAACACC					
				CWCACTCTCAC					
				TCACGACAGCA					
				AACTACCTCATC					
				CTTCAAAGTGA					
				AGGTGGCAACG					

				TGACCTCTATTT GTGAGTGAAAC TAGCCAACAGT GTTACAATCAA TGTGTAATTTCA TACCCACCACA CAACACGGGTA CCAATAATGTG TCGAATTCTTCC CATCATAAGAC CCCAAGAAACT GGTTTTAGACG AAGGGATAGTT TCAAAAAGGCA GATTGGCCGAA ATTCTCCAAAAT GCTGGATGAAA AAGTTTATACA CTTAGCCTGTGC CTGAAAGTATG AACACTTTTTGA GCTCTTAATTCC TCGCGAGCATT					

				ATCCCAAGAGC					
				TGCAAACAAAC					
				TACCTGAGGGC					
				ATTACTTCAGA					
				AACAGCTACCT					
				TGCTCGATATTA					
				TACCCAATATG					
				ATAATCCCATTC					
				AGTGATGAAAC					
				CCAAAACAGCC					
				AGATCCGTCCTC					
				TCCTCAGTCTCC					
				GAGGCAAAGAG					
				AGTACAATGCG					
				AGAAACTCATG					
				GAAGACTGTGA					
				TATGAGTAGAA					
				GTAGTCACAAG					
				GCTGGAGGCTT					
				CTAAAGGCTTA					
				ATAATGACCCTT					
				CACAGGAATAT					
				GCAGGAATATT					

				CGATCTGATAA AATTCATACAT GAAAACACCTC ATTCAACCAGG TCTGGTAAGAA GACATGTGAGA GATAA					
Penelope	Penelope-2_LMi	.	Locusta migratoria	AAGTTGTTTGA ACATTCATATGC GGGAAAAGTTA TTCACATTGTTT TTACCTCTACGG GAGGAAATGTA TCGGTGTTTACG GAAGTTGGATA AACTCCGCAGC AGTAGAGTAAG GTTATGTGCTTG TCGTGTTGAAGT GTCCATATTGTT CCACGTTTGCTA AGGTTGTGCAC CATATTAATTCT	TAACAGTCACTATC GATATTTTTGATGT TTGGTCATCTTTGG TTGTGTTGACTCTG CAGTGACTTGTACT GTGTGTGTTATCTT CCTTGTTTACCCTTT GCAAACCGAGGTAT TAAATACGCTTGCA CATTGCTTCCTTGA GTGCAGTTGTGCCT TGAGAATGGCAGG GTGTCCTCCTGCCG AAATATCGGCGGTT GTCGACGACGTCAC CCGGCAGCATACCC	MIAANVSEFSWDWLDG ASWTQSDWEFQSVTAR HSAKFERFLGSESDSNLI SRRSVINLTEKTYDDAA LSVLGKGLNFAPTPKRL PVVDFISSIEQAVSKLPS DSAEEARREACRVLTGA RPPKCNITAAEASALRSL RMDPDIVILPADKGNAT VLLNKDDYIQKMQCLLS DSAYRRIGADPTKSVER KTNNLLKKSPFSQETIKS LNSYCAVPPRLYGLPKV HKENIPLRPIVSNIGAPTY RVAKHLASLLSPMVGRC EHHIKNSADFLRRLQGL	cd00304	339	423

				CCTGCCGCCAC GTATCAAGCAA CGTGCTGCTTAG CTCTTGTTCGTG AAAGGATCCAC TTTATCGTCGTC AATTAGATTTTG TTTCCAAGGAG TTATTTCGGCTT CATTTA	GTAAGTTGTTTGAA CATTCATTAC	RLNDSDILVSFDVVSLFT RVPLSDSLQLIEVKFGVE LTNLFRHVLTSTYFLFNG QYYEQTDGVAMGSPLSP IVANLFMEDFEERALVS ATLKPACFFRYVDDTFV VWPHGSENLNGFLEHLN SIHPNIQFTMEVEKNGCL PFLDVLVRRKVDGTLGH SVYRKPTHTDLYLRADS CHHPAQREGVLRTLVHR AHVISDPESLSAELSHLE VTFRQNGYSERQIKRAL RYQPSVQRVSDDSNEVA PKSMAFLPYAGSISNKIG RILRKYDVKCVFRPPSKI KALLGSVKDDLGLRKA GVYRIPCSCGMSYIGQTI RTVEDRCIEHKRYTRLQ QPNKSAVAEHCLGTGHP MEYNNTEILACTSTYWD SVIKEAVEIKLASDLINR			

						DGGFCLNSAWNPALSLV KKQRDKINAASPVD			
Vingi	Vingi-26_LMi	.	Locusta migratoria	ACGAGTACTAC TAAAAGGATTG CAAAGAATACT CACGGCCATCC CAGTAGAAAAG CCGGCTTTCGTS CCAGCCGAAGC GTGCAGACCAA GTCCTGCTCTTA CTACACATATTG AAGCAGCTTCC AAAAGAAATTA AAAACTCTGTG GCCTTTTTGACT TGTCAGCAGCC TTCGATACCGCC TGGGACAAGGC CTGCTCTATAA	TAGATGTTTGTTGT GAATACATGTAATG TAATATATGTAATG AATGTACCATAATG TAATCAGACACTGA AATGTACTTAAATG CCATACGCTAAATA AATAA	TYENNSLSKSDKLIGNMI GNRGFQVIMGNISSQKF LQNGLPQGSVLAPLLFSL YIADMPETACTKFGYAD DWXIATRHTNFDTMZET LTSDLSALAAYFRKWRL RPNATKTEATCFHLNNR GADRTLEVYFDDTRLNY NKEPSIQHTSVNGGLGL MPRRQKRLASISITEERT ELWKYTLTTHASTTTRS QVSRVTLDRTLSFKTHV TKTAAKVRTRNNLIHKL CGTTWGSTATTLRTSAL SLVYSAAEYCAPVWLNS SHVKKIDVELNNTMRIIS GTIRSTPNIWLPVLSHILP PKMRRETALLREYNKIE ANPELPIHSDIPDLRIKRL RSRHPPLSLDSLTGADIN SWRREWLQNAPQLATH	SSF56672	34	112

						QLLEKRMAPKRPTTCHL MPCITEAVPGFDQPRKI WSTLNRLRTGHGRCADS LFKWGLLSSPACDCGAP KQTIQHIREECSARAFSG DFEEFLMATQKLLNIYK D			
Penelope	Penelope-69_LMi	.	Locusta migratoria	GTATTCTGAACT CAATTGATTTAG TGAACAAGTTA AAAGACTTCCC TATCCCCCCTAG TGCCAAGCTAA TTTCTTTTGATG CCAGCAGCTTA TTCACTAATATA CCAGTTGGTGA GTGCCTAAACA TCCTCACAGACT TACTGGATAAA TCAGGAGTTAA TCCACTTGTCAC AGATGA	TAAATTGTAATGCC TTATTTTACTAGTT ATAAAATCCTTGTG TAATCACAACTGTT TTGGCTGCTTATGT GATATAAACTTAAT GTCTTATTTTACTA GTAACCATAGTTTA ATTTAACCAGTTAA AATTTTAACATATA TTTACAGATGTTTC ATTCAACTTTGTCA CACGCTTCCATGTT CCAGACATTGGACC TGTTTAAGATAATT TTTTGGTCAAGTTT	HKVSHTYLPITKLFSFQ WQGIQTVRWSSYGFPLS PILAEIFMDNFENQFLES NPLSSKVGYWFRYVDD VLCLWTGTTRQLQTFLK QLNSAHKSINFTMEFGG NCINYLDLTIDISNHCHT FKIYRKPSTTDTVIPAQS QHPVTHKHAAFHSMIHR LLSIPMSKESFQIELNTIR QIASNNGYPLPWWLPPT LVDSIMHKKQKQHVSSL LYPVINTPTSQDHSKWC TIPYLGKVSVNVAKKLK SRNFKTSFYTHNTIGHFL PNGKDNPPALEKSGVYK	SSF56672	12	120

					AATTTAACCAGTTA	ITCNCGKFYIGQSGRAIC			
					AAATTTTAACATAT	TRLKEHHRSWKLRQGN			
					ATTCACAACGCTTC	STFADHLLKEGHSYNVT			
					CATGTTTCTAGACA	HEVLHHVQKGRKMNFF			
					TTAGACCTGTTTAA	EIMEINRHMYTSPNLVL			
					GATAAGATACGTTT	NDQTQFPYSPLLSAEPN			
					TGATCAATATTTTT	HNPVQTMPCADNVNYH			
					CAGCCAATTTCTAT	LLPTCLIPLLLFHSHVIKA			
					TGTTTCACTTTATAT	LCNHNCLAAYQI			
					TTCTTAGGATAACG				
					TCATAATCTATTAT				
					ATTGTCACGTATTT				
					TATCAATGTCAGGA				
					ATTTACACCACAAC				
					TTGAATTGTAATAT				
					TTCCCGCAGGTTAG				
					TTATAAACCAAAAC				
					ACTAGTCTCCTTTG				
					AAAGTTATATATCT				
					AGCAAGTTAAGATC				
					TTTGAAACTTAAAC				
					TTGTAATGTTTACC				
					TGATGATGGCGTAA				

					CAGCCGAAAACCG GTTCGTAAACCAAT AAATTTGCAGAATA TCTCACAGTTATCT TGTGTATTTCATTC ATAATGTATCAAAT ATTCTGTCAAGAAC CGACGGAACAGTC AGTGAATAT				
Daphne	Daphne-60_LMi	.	Locusta migratoria	TCAAAATAAAA ATAGTGTAAAG ATAACAGACAA TAAGGAACTCA AGGCGAAAACT GAAGCAGTGAT GACGAAAGAAG AAATTTTCATAT TGTCAACAAGT CATGGTAAGGG TTAGCAAAAAT AATGTCGATAT CCAATCAGAAT ATAGTCAGTGG	TGATATAGACGTAC ACAGAATTTTAAAT TGTAACATTACCAT TGTGATATATTGAG TCACTGTTCTATAT TGTACTGTACAACA ATGAGATAGAGGA CTTTGACCATTSAT GTAGATGTATACAG AATTTTAATTGTAT TTTTTCTAACAAAT TTACTATTGTTACT ACTTTTCATTGTAC TGTATTGTATTAGG	MDSTAPCKKVLAKEILQ AIKTASVDTFEKLALPAI GLLPMANEKENVPTVH KEETTMATPPEETCPQPT PTSASEALSPVETPEPSV QEESGGARASQAIPGVS QETVTASTPKGGKRSTA DEAVSSVVSKRKTVPPL RLNDFLVECRKIKKPTKI CIPLLHKKKYVEMTSRIS YPDESEADSLVPPCKEK KLIFLHTNVQSLKNKLFE LQYWLEKLNCSILMINE HWLSCGELELSVPNGYE	cd01650	693	926

				AATCTCAAAAC	AAAATTTAAAACTA	LATGYCRXNKAHGGVSI			
				CTGGAGCTCGTT	AGTATATGATAATT	YINKMLQLNYETLDVNE			
				GCGAGAGGTAC	GTACCTATGATAAA	FCIESHFEVAAVLLTELEI			
				TGAAAAACAGT	CACTGCAAATTACT	IAVSVYRPPNHSEDLFIE			
				GAAACTACTGT	AGCTGTTTAGATAT	NLQKLIVSLLKKKKSQSS			
				AAAAAGTGGAG	GTATTTAAATGTAT	NFWRHLSHSSRKKNHR			
				TGACTGTGTAGT	ATGACGACACCAAT	VLIFGDINIDIRYRSSTVT			
				AATTGGTGGTG	TACATGTAAACATG	NLLDTLRSLNFYCINDKP			
				GTAATGATGTG	TTTAAAGGTGAATA	TRKNACLDNIICNFHIGR			
				TATAGAAATGA	AAATATTCTATTCT	DQFNTIKLAISDHDGLW			
				GAGCAGCCTTG	ATTC	LRLPLSHKTDTSQQRIIR			
				CCGTGAAAGCA		PINEQKLNTLKSKLTLVN			
				CGAGAGAATCA		WNELLTRDMGANESTK			
				CTAGGAAAATT		RFLEVVTKIFEETCPAYY			
				GTCACAAGCGA		KKHSKNRSQSKPGTLKQ			
				AGGTATTTTTGG		WYTAELNNLRILMLMY			
				TAAGCATTCCTC		YDRSQHSNIDRLRYKNI			
				ATAGACATGAC		KKYYRSAIKEAKCKSND			
				CGGAGACGATT		DYILNSVNKCKAAWEVI			
				CATGTGTAAAT		KTEINHQKKDQNIPISPD			
				AAAGAAATCGT		TLNKYFANPLMENNTEE			
				CACAACAAACA		DVNKAEELLPQLHNKA			
				GGAAATTCAGG		MGHFEWNHVTCRQVSV			

				AAGATCTGCAG CAGTTTTGTGCA CACTACCTTTAT TGAGGCTACTA GCTCTATCAGA GAACATTTTAG AGA		AISKLSNSRTEDHYGLSN FIIKQISQEILTPITNIINKI LTEGVYPDSLKISVIKPV HKKGSTLSPDNYRPISLV PIISKITEYCMHQQLNNY FEHNNLICSEQFGFRPTL STVKAVDNIITKVYDCFE KKTLSCATLLDLRKAFD MVSHNILIRKLQYYGIGE NALRLLQLVNVNNKKS DPLPVKKGVPQGSVLSP FLFIVYINDLPSILPCEAV LYADDTTLIAQNSDMDE LLLNVAVAIENCNNWFQ ANXLHINSSKTENIIFDL NAHKHENKSVKLLGFHI DQSLSWETHTDKLCSKL ARVIYLLHRLRGCVSKT LLLYAYFAFFSSHLTYGI LLWGNSPGSKAVFKWQ KKAIRCVAGLAPQESCR HFFKEMKIMTVPSLYIYS CLVYVKENLHKYDLRM			

						SVHNHNTRHRHAIDKPH VRLAKVHSSYKYLGPTF FNKLPNEAHSVPLNKFK NIISKWLKSKAFYTIKEF EESTIDDLEF			
Nimb	Nimb-24_LMi	.	Locusta migratoria	CCCCCCAGGGG GCCCACAGCTC TTTTGTGGGTAC GTGTGTGGCGC GCACGGGTCCC CGAGCCATTGC GGCCTTCCTTCC TGTTTCCCAGGC CGCTTTTCCTCA CCCCGGTTCCCT CCCCTCCCTAAT CCTTCTTCCCAC CCTACCCTTCTC CAACCCCGTCTC TAGGTGTTCTCG CGATGACGGCC CGGCTATTTCCA CGGTACTGCTG	TGAGTCCTCTGGGT GGTCTTTTAGATCA TTATCTATGTTCCA CTTGACGCGTCTTT CGACTACAGCTTTT CTTGTGTACGCTAT GTGGTTGGCTTTTC CCTTTTTATACCCA TGGTCGGTCAACCA CATTCATTGGCTCT GTCATTTTTATCTC GTATTTGTTACTGT GTGTCTGGCGTCTT TTGTATCGTCTCTT GTTGACTCTGTCAA TATTTCATTGACGG TGCGTTCTTTTATTC TGCGCTATTTTAAA	MSKLRDGSVLIQAETPA QSRSLLTCDHLGDIPVSI TPHKSLNRVQGVIYHRD LLLQTDEELRDNLQRRG VHFVRRVFRGPKDNRV ATGAFVLAFEGDNLPEK VKVMMYRCDVQPYIPPP MRCFKCWRFGHMSSRC NAAATCRDCGRPSHPDT ACAPPPYCVNCGGHHSP CSQDCPALHKERKIVEY KTLDRLTYSEARRKYEK LHPVSLLTSYAAATTMA PPPVATSRPALPPSVGPM GRPSTSASQTSGGTPSSA GPVLPPTSGVVTPAPQPE KGSPSLASSSGKGTCSPL PTMSPQKSDTSQRLKDP	cd01650	901	1159

				GTTTCGCAATTG	TCTAGCTCAGGCGC	RPESRLAPQSSHTPQAPP			
				TGTCTTTGTTTG	ATTTATCGACGATT	EKPPSSGRPKERREDKK			
				CCGTTTTCACCG	ATGTTTTTCTGCGT	EKKERRKAQGDAGGLG			
				ATTTCGGCCCA	GTCGCCTAGCTATG	APSPCTGHPSRDARRFG			
				GTTAGGTCCCCC	AAGTGGGGGACTG	LCRSRGCHGHGGSDFTA			
				TCGGGGTTTGA	ATGGCCATAGTAGT	RDECSCHPSSIFIVFMHR			
				CCTCCATCAAA	CTAGTCCCCCTAAA	QSGDSLILQWNCGGFFC			
				ATTTCAGAGTA	ACCCCCCATACCAA	HLAELRQLLSSFPTLCVA			
				AGTGTGAGCCA	CCAACCAACCAAC	LQETWFPGDRTPTLRGY			
				TTTGGGGAAGA		RGYFNNRTVYNRVSGG			
				ACTCCCAACCT		VCLYVHTSVYSSPVPLQ			
				AGCCCCTACGG		TALEAVAVRVRASRDITI			
				CGCGGATTCCT		CSLYLPPDADVAPADLA			
				ACCCTTTCCTTC		ALIAQLPSPFLLLGDFNA			
				CAAATCCCTCTT		HHPMWGSNVTTGRGRD			
				CCTCGCCGTAG		FELLFNQLDLCILNTGAP			
				CTCCCTCCCCCG		THFSVAHGTFSAIDLSVC			
				CTAGGTCACCA		SPGLRPLLNWWVHDDL			
				GCACGGTAGCC		CGSDHFPIFLSLPHHQRH			
				AGTCCGTGTGG		ARLPRWSFHRADWAAY			
				TGGGGCTGTTAT		TSAAMLDTPLPADATM			
				GTACCCTCATG		AVDQIISVVVSAAESAIP			
				GTTGAGCCCCCT		SSSALPRRRPVPWWSPEI			

				GAATTGCACAG		AAAIRNRRRALQQYRRH			
				GGATCACACTG		PTVDHLVAFKRLRAQTR			
				CCGATGCCTGA		RLIRQRKQACWEEYVST			
				GCTGTGACCTCC		IGPRVSSTQIWTKIRRISG			
				TCATGAATGCC		YHPPAGVAGLSIRGTIHT			
				AAGGAGTCGAT		DPTIIAEYFAAHFALNSA			
				GCTTGTCTCTGC		STAYSQGFCTLKERAER			
				GGGGCATCAGG		KQLVFRTNRQEPYNAPF			
				ACTCCCAGCAA		TEWEFQSILASCPNTSPG			
				CGGTCATCCTGC		PDRIHCQMLKHLSVNSQ			
				CAGGTGGCGTT		RHLLDFFNRLWIEGAYP			
				GCTGTGGCTGG		SQWQESIVIPLLKPGADP			
				GTGGCGCCCTT		LNVDSYRPISLTNVLCKL			
				GAGGAGAGCCC		FERLVSRRLLWVLEYRG			
				CTGTTTCGGAGT		LLAPQQGAFRQGRSAID			
				GGGTGGCATCG		NLLSLESAIRTAFARRHH			
				GGGCGGACATT		LVAIFFDLRKAYDTAWR			
				CCGCAATGAAG		HHILATLHGWGLRGSLP			
				AGGAAACAAAC		LFIQNFLLSRRFRVRVGT			
				CTCTGAAAATC		SFSSPHIQENGVPQGSVL			
				CTGGCTCTAAGT		SVPLFLVAINGLVAAVG			
				CAGGTGTCTCA		PSVSCSLYADDFCLYYS			
				AGTGTGGTAAG		TASTSVAERRLQAAIHR			

				TTTAATTTCAAT		AQNWASAHGFTFSPAKT			
				GCTGCAAGATA		RVMHFCRQRTVHPHPEL			
				CGACCCCAAAT		YLDGELLQVVDTHRFLG			
				CGTTTCCTTCCC		LLFDSRLTWLPHLRQLK			
				TAACTCCATCAT		QRCSRHLDVLRCLSHTV			
				GGGGGGAACGT		WGAARPILLRLYNALVQ			
				GGGCCCACAAA		SRLDYGSQVYGSASPSA			
				GCAGACTGAAC		LRVLDPVHHTGIRLATG			
				CATACTCACCA		AFRTSPVTSLLAEAGVA			
				CGATACCTAGTT		PLRYRRNTYSPSMLPEST			
				GTTACCAGACA		VPLGIQITDSFSRGRSFV			
				GGATGGGGAAT		CRRGDPGRDSRRSGAVT			
				CTTTCCAGGCA		SGLREPGLWLSLPICTES			
				ACTAAGCCTTCC		SRPSPPYRHPSGHWSLS			
				TTCTTCGTCGAA		YEPCHESSRRGWSRPIAI			
				CACCTCGAAGA		PAQHLLAQYAARVHCTP			
				AAGGTTTGGAG		RHPNNRLLFPGEVIRLPQ			
				AAGTTGCCGGT		RRPRSGLPIAARVQSLLS			
						TLDSFPLNRPFRSTFHAP			
						PWTIPRPRIRLDLSREPK			
						TSTSPLVLRHQFYALLRE			
						FRNSEVIYTDGSSDGDH			
						TGYAYMHGAYVEQRSL			

						PPGSSVFTAELVAIYRAL KHIRTCPGTSFLICSDSLS GLQALDQCYPRHTLVRA IQESILALDDSGRSVTFI WTPGHVGIPGNESVDQL AKQATTCRPLAIGLPEA DFRPLLRNKIRELWNAE WSALDVPNKLRTIKDTT SVWKSSQLGSRRDSVAL CRLRIGHTRLTHGHLLR REDPPVCQCGAPLTVVH LLLHCPILAPLRHSFNLP NTLSLVLGDDASVANM VLRFLRAGGIYQLL			
Daphne	Daphne-23_LMi	.	Locusta migratoria	C	TAAAATAGAATTTT GTAATGAAGAACTG TGACAAACATATAT ATGTAAATTTGTAT TACTGTGTTAAAGT GTTCTAATAATATA TGTGAAATTGCATT AATCCTATAACATT TAATACTTATCTGT	LYLKQPQQLHKVGLPIIH SLVKCLKPQEQHQKIAL SLQCKTVSTPTTSPGRHP RRSYASALNTAATTPSSP VNTEETATVAVPSEEGG GKSASFSSKFFSRKQSKE XHNITRXKTLKEVSGSSE VTKSISCNKQNKSFHIM YHNVQSLTNKLQELEIIL	cd01650	637	930

					AAAATTAATTACCT	SSDLNADVICITEHWLEK			
					GTGTTTCATGGTAT	HQIEALKIPGFHLATSFC			
					TCATTTATTAAGTG	RTTINMVGQLFLQNRIT			
					TGTTTCACGGTGTA	WWDSYFCKTEFKAMQS			
					TATTCATTTACTAA	YSDLIKEKVFEMSAIELT			
					GTGTACCAGTACTT	DKHLVIVTVYRSPDGDIE			
					AATTTTGAGTCAAT	EFLCKFEMMLNIVEPTK			
					CATATTGTGTAAAA	KKIICGDFNVDFLKSSNY			
					CCTTTCACCATTAT	KEELLYITKNFSLIPTVN			
					ATTCATTTATTACA	TPTRITKHSKTTPDQIFIQ			
					TATACCAGAACATA	GTFSSYNSKVVKMGYSD			
					AGTCATAAATTGAA	HEAVLLTLELYSEKTPPI			
					CCATATTTTATGTA	KMIECRDFSDGNIANFC			
					GGCCTAACTACGCT	NQLSQETWYEIFYETDV			
					TTTTTGTTTTGATTT	NKMFNTFHQTYKYYFDI			
					AATGTCAATCATAT	AFPLQQRKLKPAKNKC			
					TGTATAAAACCTTT	WITNGIKISSTMKRYLHS			
					TGTAAATACCTCAT	YKKQNHLSLEMQHYIKE			
					GACAACATCGATTG	YNKIYRKVIKEAKKMGN			
					CATGGACATGCTTA	DEYIRKSKNKIQATWNI			
					AAGATGAATAAAT	VRYETDTKPVAKSISLKL			
					AAATA	ENRQVSEATTVANIFND			
						FFTKIADDLIQSNFQTPN			

						LHHTVNYKSPKGSMFLK QVTGSELMLAMSSLKKL SCGFDDIPDYIIKQTGYHI VEPLMHIINHSFSTGTFP NLLKIAKVTPLHKKGAT DDVNNYRPVAQLSGFSK ISLKSSFLIDCQLQACSSA QWLLKNIFEKLFLNRLL EFVNKNSLLSKHQHGFR KSRSTTTAIYEYLNSILK AVDEKEMATGIFLDLSK AFDVIDHGILLHKLERK GIRGISNEWIKSYLTNRQ QRVVLKQDTTIGNISRRS TYTSEKTTIKYGVPQGST LGPILFLLYIDDLNDTVL AKQKTLFADDTSILITGI DQKHLQQSTDTSLLELT HWFKINKLIINTEKTVA MNFHLSHQKQQNNHTL KINNRSLKLVDDVKFLGI YLQSDLKWNTHIKALTS KLSSICYMLRILSTSCNK			

						EIVIQAYHTHFQSAIRYG VIFWGNAPCSHSVFKMQ KRALRIICNMKKQESCR PLFPSLKILNLPCLYMFE TIVFVKEYLGSTSAYQL NSSVHSYRTRNDSNIHV SYARTSAYQKSVIHIGTQ LYNHLPGNIKSIKEDSIF KNKLKSFLMDHSFYSVD EYLSD			
Vingi	Vingi-19_LMi	.	Locusta migratoria	GGGTACGACGT GTAGACAACCT CCCCGATGGTG TACCCTGGGCG CTGTAACATGG GACAGCGGCGA AAAGTCTACAC AACTATGGCGG CCGGACGCACA AATGCCTCCCG GTCCACTCAAA GGCAAATGGAG CAAGCAAGGAA	TAGTACTGTAAGAG CCTTGTATTGTTAG AATAGCTGTATAAG AACAATGTAATTAA CTAAAGAATATTTG CCTTTGTATATGCC GAACGAGTAAATA AAATAAATA	MATLVKNEILRNVDTITP PTEHVIGIKLGELTLYNV YKPPSQKWPVDVLPKA NPQQFMLAISTPNIPDGV TRDPQLKALDPPAVYAG DFNSQHTRWGYPRSSAE GIGLNDWADNQDLHLLF DAKDKATFKSKAWGTT STPDLTFVTRGATGLPV QGTRRVVEAFPRSQHHP VIVEIGLSIPVIKSPPIPRW NLRKADWNKYRCFVEK TINRIPPKPENYKRFTNL	cd01650	421	675

				CTTTAACATGA AGAGACTGAAT GTATGTCATTTC AACATAGAAGA CTACACAAGGA CGAAGGGAGAA ATCCTTGAAAA AATTTTAACAG AGAACAAGTAA GTGTCGTACTTC TACAAGAGACC CACCTCACAGA TGAAACTATCG ACCGACTCAAA ATACCGGGATT CACTAATGTGA GCTACATTGGC CACGACAAGCG TGGT		LHKAALNAIPRGSRKEY TPCWTKECETLMEEFEK CGDTETADRLINAMNEE RRSRWKEAMENLDFTH SSRKSWGLLRKLGGATI PQRPSLGTTASEISAVLK HTSKIQLKTAEKKEVSR DLSNLLVHNPTNMDIVK EVETEEITQALKLMKKG KAAGPDEILPEFLKELGT LGRRWLAKLCTECISTG ALPKIWKEAKVIAVLKP GKTGDNPKNYRPISLLC TSYKLFERILLTRLTPYIE ANLPTYQAGFRAGRNCC DQVLSLTTYTEKGFQNK MKTGLVLIDLTSAYDTV WIEGLLLKLTRIIKCKQT YTLLKNMLTGRKIKVSL HGTNSKSVVLNNGLPQG SALAPALFNLYIADLPET KSRKFAYADDLAIAYQS KNFNDLNETLNADLKVL			

						NTYYSKWHLHPNPNKT TSTIMHLNNREAHTKLQ LTMNGQQIRHENKPKYL GVKLDRTLSYNAHLEDT KQKLKSRNAILSKLAGT TWGCGASTLRTSGLALV YSVAEYCSPVWMRSAH VAKVDVQLRETMRIISG TLRATPCAWLPVLANIE PPEIRRSQLAIRTHQKILG NLDLPIHEDLRATDRLKS RTAFWKTSEELQEFNSK TAWKELWATSAPKNSN LVDDPNKKIEGFSLPRKL WTTLNRVRTGVGRCKQ LMKKWGLTDSSECECG ETQTMDHLLACEVYGY NGDLKDIHRLSDSALKW LEKICNVV			
Vingi	Vingi-22_LMi	.	Locusta migratoria	GGACTTCCCTCA TAGCCAACACC GCCCAGTCTTCA TTGA	TAATTAAACAATGT ATAAAGTGTGTAAA GATAGCGATGTATG	NWNHIPLITTTQRPRWN FRKAKWDIHPEKLDKCL GWIHPTANNDRFMGAV KSAAKAAVPRGFRKQY	cd01650	287	563

					CCATACGATAAATA AATAAA	VRGWDKECENKYQRFL ESGDPEIADDLLHSLDN ARREKWIETVESLDFKT SSRKAWSLLRKLGGAN KLHHNSTDISANRISAHI QATSRAPRDKKHTRTIR RELSFLKKNSTPLDQYSS KFTTTEVREALKSVKAG KAPGLDGVHPEFLLTVG NMPKNGLLNFIQMSQRS AEECQGRKGTWIGWSTP RIPSYCGEYAKKWLAEF YSDVMATGNIPQKMKN SKIIALLKPGKPSNQPKS YRPIALLSVLYKLLERLL YNRLSPKILESIPDEQSGF RPNRSCTDQVLSLTTHIE AGYQKSKKHQLHLSSTF PDNAYRGRLSKKQKTSI AFIDLSAAYDTVWIQGL LYKLIKIIPCSSTIRLINN MLTGRAFQVITGERTSST KVLNNGLPQGSVLAPLF			

						FNLYISDIPNTQSRKFGY ADDWAIAAQHESFEATE DILCSDLKLLCGYFRKW RLQPNPTKTEVSSFHLCN HSARRELNVIFEGKRLSH NNYPKYLGVTLDRTLSY KEHLTNTANKMRTRNNI IHKLCGTTWGSTANTLR TSAVGLVYSVAEYCAPV WINSPHTNRIDVVLNDT MRTISGTIKSTPTVWLPT LSHIPPPEMRRKNALLRE YNKITNNPQLPIHSDVAD LERRRLKSRRPALLSAK EMMNTNFEVEDLWKQH WQEHCPPKCQSFMTGE KPPGFNQPRKVWTTLNR IRTGHGRSADAFYKWG KTPSPNCDCGAEKQTVR HIVEECPFRAYKGTFNDF LAATEEAINYIKCLDVNL			

Penelope	Penelope-47_LMi	.	Locusta migratoria	AACCATTAAATTACCCATTTCAGGCTGTAATCCCTCCCTCCACAATGA	TAGTTCCTGCCTCTACCACCCTTCCCTTCCTCTCCCTGCCACTCTCTATCCTTACATTCGTCTCCCATCCCGCCACACATTTCATTCCCTTCCCTTCCTTTTCCGCACTGTCCGTCAGACAGCACTGGCGTATTCCCCCACCCATGCCACACATCCCCCCCCCCTCCTCTCCTGCTTCCCTCTCCTTGCCACCTCCTCACCCCCACCACTGGTTGCTCTCGCCTAGGATGCCATGCTGGGTAGCGCAGATCGGCCACAGCGGCCGGAGACGGCGGTCGTGTCGTGTGTGTGTGTGTGTGTGTGTGTCTCTAAC	PPPVQISSIHSTHQPNPTKPSKPGPNSLAAPHLHRKILLLCNPKYLHPIPQTDVLALQVLEQHSHTTSKKLSNLFTSYSRLGLPLSTPSTTSRHTPQPLIADKPCLADLLHLPHPQKLPHTTTQNPQPKQTVMNLSSKSLSPIEVSVLSKGLTFCPTPKFNHAGLVKDLLSFSQSLQWKHHFATNPTIQTHTNTNIEPCLTQFTPPSSQDPPTLPPNHPLLTFQNFLTSNLASPSFPKSLKMDTNLTPSERNAIYHLKTDPDLMILPADKGSTTVVLDRKDYVAEGLRQLSDSSTYTPCHNDPIPKTQQDLQSLLKSLGPSQDLSPESISLLTPPTPRTPTFYMLPKVHKPNHPGRPIVAGYGAPTERISALVDQHLQPITHNLPSYIKDTNHFLHRLSTVPVPLPPGT	IPR000477	326	566

					GACGAAGGCCTAA AGGCCGAAAGCTAT AATGTTGTTGTCTG TTATTTGTGCCTTTC TGCGGCTCATCGGA CATCTCTGTGGTGA GTAAGCAACCTATC CTTCTCAAAAATAT TGTTGA	MLVTVDATSLYTNIPNA HGLAAVEHYLSQCPTDC KPTTSFLVTLTNYILTRN YFSFEGITYKQICGTAMG TRMAPSYANLFMGHLEE SFLTTQNPKPLIWFRFID DIFIIWTEGEDTLSTFLQN LNTFSSIHFTWSYSTQRA TFLDVDLFLKDGLISTSV HIKPTNHQQYLHFDSCH PFHTKKSLPYSLATRGR RICSDKQSLSKYAKRLTE AFTDRNYPPNLVQKQIS RALSVQSPSSSPTPTVRP QRSTPLVTQYHPGLEQL NQVLRQGFDYLSSCPEM RNTLPNILPTPSYSGIPPP TKPTEIPYQISSPPPPTVV FRRPPNLRNILVHPYSIP APNPLPHGSYPCNRPRC KTCPIHHPTTTYSSPVTGI SYPIKGRATCESSHVIYK LSCNHCAAFYVGMTSTQ			

						LSIRMNGHRQTVARRQL DHPVAEHAALHSVTHFN DCYTACAIWILPTNTSFS ELRRWELSLQYILHSRNP PGLNLR			
RTE	RTE-47_LMi	.	Locusta migratoria	GCCATAGCCGG TCCCAAGCCCG GTTGAGAAAGG AGGAGGGGGAG GAGTAACACCT CGTAAAAAACC TTGGTTCACCTG GCCCGGGTGAT AGTACCTGTGA GGGCCCCTTGC CAGGGATACGG TGAAGACCTCA ACGGCAGTGAA GGCGGAG	TAATAATAATAATAA	MKIIGTAELAEEATLQLN VQVVEDDAPLMAANMV RWYRAAAGTSIELMHR ANDASDGSRSKTHGRN QSCFRWIKIEDSWAKPII VWFFGHDTAMWKGNG KTTTLLYFLNVGRRRIM MAFLGVKYSGDKLLPRS DLRVGSISGGVIKRDEHF NLRIGTWNVRTLQRAGK LENLKREMDKCEVNIM GLSEVRWTDSGEIVSGN YTMFYSGGKSFERGVAI VMGNELVQSLTKVVCH SDRLMFVKISAKPVDILI VQVYMPTSDHEDEEVE KMYDEIEDIIQSEGTGKV NTIIMGDWNSVVGQGNE	cd01650	655	900

						ANIVGQYGSKHSWTIWL GKRNDRGGMLVEFCQR NNLVVMNTWFKKRKSK LYTWKSPGDMNRYQID YILVKQRFRGSVKDAKT MPAADIDSDHNLLVADI STRMKKIQQHGRRKQR WDLEKLRENNVAVRES LEKKFNEDEIRASAEEQ WQNVKMILLNTLKSEV GRETKRARKPWITSEML EKMEERRKCKNVNERN TEDRNRGTTKKRKVRPH VPENEGTGEGENKSVRT FGIEDSRGQVVTEQVEV LKTWEEYIRELYDTEHR PHDIDIEPEEAIDEDEKGP SILTSEVEKAIKDMKSRK ATGDDDIPVDLLKDIGN DSLKALTQVINKIYETGE WPKDFLDVTMVTLEKK KQAKKCSDYRTISLISHV AKIIARLLNRRLEHKIEE			

						VMGEDQFGFRKGKGTR DAIGMMRILSERVLAVN EEICACFIDWQKAFDRV NWNILMNILKEIGINWR DRRLIRNLYLGQRVKVR LKCGETNSVEIGRGVRQ GCCLSPSLFNLYGEMLA REALKGCGNFKVGGRLI NTIKYADDLVVLAKNQR QLQHMMNRLVEMGRK YGMEINIEKSKVMRISKR EKHLKIIVDNRELENVN QFKYLGCFITKDAHCTN EIRARIAMAKAAFNKKR TLLTSKLSLALRKKLVK CYIWSIALYGAETWTMR KKEKKYLESFEMWCWR RIEKIKWTDRVTNDDVL RRVGEKKSILRTILQRKA NWIGHILRHEGLIHDVIE GKINGTAGRGRRRIQHL DDLKDGRSYIRLKEEAE			

						DREHWHHKFSVRRHGP ATRQNTT			
Penelope	Penelope-38_LMi	.	Locusta migratoria	CACTCCTAGGG AGCTACCAGTTT CCAGTTTCATCA GCTCTGTGGAG CAAGTAGTCAA CACTCTACCATC AGTGTTGCAGA AGAGATCGGAG AGAGACGCAGG GCGCTACCAGA GCAGGCCCCAA AACAAACATCT CAAGAGACGAG AGGTGGCGCTA G	TGATGCGACGGGTA GCATACCTCGCAGG CGCACACCGAATTC GAAGCCAGGAGAG TAGCGTGGCCACGA TAACATTCTGTAGT GGTCACCAAMAGC AGAGGAGATCGCC ATGAAAGTGCTAGG AACGCGCATGGCG GGCGCGGACCACCT ATGGAACGCCTGGG ACGGCCGCCACTAG AGAGCCGCAGGGC GCGGACCGACGAG GGATCACCAGCGTC AGATACCACCGCCA GACTGGAGCACATT GGGCGCGGACCGC ATAGAAAGCACCA GACGCCGCCTAGGC	EFAGRRKLGVLPQTRGT PLSFLKTDYDQKICQLLE DPAYXLLENDPTDKVDR KTRSLLKETGWPEKVVK QLRAKAPVPPRLYGLPK VHKQDVPLRPIVSNIGA ATYATASHLKKMLAPY VGKCVHHIRNSTDFVER LKNLRISNTDIMVSFDVV SLFTRVPLQDTLDLISER FDGTVLDLFKHILTSTYF LCGGQFYEQTEGVAMG SPLSPVVANLFMEKFEEE ALTTAAYQPTCFFRYVD DTFVIWPHGQNRLDEFL DHLNSRHPNIKFTMELE NEGQLPFLDVLVKRKPD GTLGHSVYRKPTHTDLY LQADSCHHPAQKNGVL RTLVHRAKTLSDSDNLA AELDHLQRVFTNNGYSS	cd00304	203	287

					ATAAATATGGGACG CGGTCGGCTTGTCG ATCAGTCTCAGGGA ACACCTGATGAAGA CGCCGAGTTACGGT GTCGAAATATCGTG TTTGGAAGACGCGG ATCACCGGCAGTAC ACCCGATTCCACGA GA	RDIRKALHPTRRPDTTGE ERRRDREGSLPTICRAYI CQDQQNLEETQNQKCFL PTKQDQQSWIISSECSQT TATRAETSARRYTPPDA QIQQEKKEEETGKVAFL PYAGPISAKISRILRKHKI KSVFCPPNKIRALIGGVK DDLGLRKPGIYQIPCECG MSYIGQTTRTVXIRCKE HQRHTRLRQATKSAIAE HCLELDHTMDYDKTRIL AQTPRFWDSIITESIEIKM GEDLINREAGYQLSTAW NPALELLQKQRKKGHPP VELQTTTRDSSSGWPPC APTTA			
Penelope	Penelope-32_LMi	.	Locusta migratoria	TG	TGACGTCACAGCCC GCAGTGCAGCTATA TAAGGACGACAGC AGCAATCACTCGAC AGTCAACCACTTGA CAATGGCCGAGGA	WNKHFARCLVNYRQIR RETSHSIQATKTEHDKSR TQWPSELRNDPLIVVLP ADKGNAAVIMERATMT RRFERCWRRRHMQKLS RDPTSKILRKTSELLKRS	IPR000477	105	351

					GATCTCGGCCGAAA	SLEDSVIKNLLPQAPRPP			
					CGTCGTGGACAGTA	TLYGLPKVHKMNTPLRP			
					AACCACATGACGCG	IVSTIGAPTYRLAKHLAN			
					GCTGGAAGCCCGA	LLTPHVGHCEHHIKNTA			
					GAAAATTTTATCAA	DFISRIKCLRLGEEDMM			
					CATAACTCGCGCGA	VSFDVVSLFTNVPVKDCI			
					AACAATGCATTCA	DLLSQLFDDTVVDLFKH			
						TLTSSYFLYGGEYFNQT			
						DGVAMGSPLAPVMANL			
						FMEHFEDIALDTAPVRPS			
						CFYRYVDDTFVVWPMD			
						VKNSVNSSTTSIASMTRS			
						SSRWRWRKRVPCRSLTS			
						SSAVKQMDVSAMVFTE			
						NPPTQTGTCTPQAIIIHRR			
						NSLFHVRCLAHGREKLG			
						EFFDHINSIHDKIKFTME			
						VEKEGALPFLDVLVSRK			
						ANGCLSHGVYRKPTHTD			
						RYLHATSHHHPSQKQSV			
						LRTLVHRAEVVSDTENL			
						QKELSHLRKVFRDNGYS			
						NSQITKAIFPKTQKQKTP			

						EEDSKKLAFLPFCGSVT GKISRLLKRHKIDSVFRA PSKIRQLMRPVKDDVGL RTPGVYKIPCGCGQFYIG QTVRTIEQRCVEHERCY RLRYPEKSAIVEHALEN RHRIAFEETSIITRTNSFW DSVIKEAIEIKIADNTLNR DGGLQLSTAWNPAIGRL KRARQARDQNIAFYGH VAEHQ			
RTE	RTE-33_LMi	.	Locusta migratoria	GCCCTAGCTAA GTCTGGCATTGC TTCCACTACTTG TGTCAGACTTCT GCTTTCTTCGTT CCTATCGGCCTC TCTTGGTCAACT CTGCTGCTTCGA CCCCGACGGTA TTAGGTTGCGA GGCCGAGGGTG CTTTCATTTTCT	TGATGATGA	MRIGTWNVRTLLKAGK LEEVKMLMEKSQLDILG ICETRWGGNGDFMSGDF RIIHSGSARSGRNGVAIIL RGKWKNNVLNTYHCSD RIIMIKIKAQPMDLCVIA VYLPTTNSKDEEVEVVY EQIDDLMKLVKEKDNLI VIGDFNASIGEDKKGLW NGKFGLGKMNSRGERL LDFCEQYEMIVTNTCFE MPKRRRYTWKSPVDGE	cd01650	533	722

				TTTCTAACACAT		RYQIYILVKKKFINQIKSS			
				GGCTGTTAACC		HSYPGYDIDSDHVLVIA			
				GAATGGCTCCC		DYNSKFKKYKRTNQHR			
				GGTATAAGGAG		WCVEKLKEKQVAADM			
				ACCAACCTTAC		KELTDNLIGNREGWEGF			
				CAAGTGCCAAC		KETVEVAAKELVGLRKL			
				GGCCAGTAAGG		EPRKPWMTKEILDLIEER			
				GGGGATGAGCG		NRLRKKDFDKYKRVKN			
				GTAAAGAAGGG		AITQKCRLEKEKWMHD			
				GCACCTCCTTGT		KHKEIQQDLSGGQTDKA			
				CCAGGGATGAG		YSKIKTMQTKTRTKTNM			
				AAATTGTCCCTA		VKNKNGDLLFELEDVCE			
				AAGGCGGAGGA		RWKEYIEELYEGEDIAN			
				ACCATTAACAT		EGEILEEENSVDIDMIGP			
				GGTCAACGGCA		PISKEEFNSALNDLSNGK			
				GGAGGATGCAG		ATGTDNIPAEILKEIGDQ			
				AAGGCAACGGG		QKSTYINFTFKEYYEDG			
				AACCACTGCAT		KEPPTDLLNSKTVIIPKK			
				TAACGATCCCA		EMLWTVKQEKTNLVSD			
				TACAGGATTCC		LEKAKAILALRMVLERR			
				CAAATAGAAAT		IGVNRKTYLTFIDLQKAF			
				ATGGCTACGTC		DMVNLEGQRLIWNLYK			
				CAATTCAAAAT		RQEAEIEVNGVKKKAKI			

				TTTCCGGAGTTT CAATTCCCCCAT TCGGATCTCCG AGGGAAATGTC TTGAGCAAGGC TGTAAGGAGTA AAGGAATAAAA GAACAG		RRGVRQGCPLSPYLFNL FIEKAIEKLKRITKGIKIN GERIHCIRFADDIVVLAD SVRKWSLCCLSKEMEF MLLKFAKILREFNLKINK KKTKTMVVGKTKEKGK VNIKLEDDVLEQVENFC YLGSTITDDNKCITDIKR RIAWQSKLSKIKGIYMY HRYKEKNSLAKQAFQN KRNLLTDNHISLESRKKF AKTFVWSVLTYGCEAW TLGKAEKKRLEAMEMW IWRRMTKTSWVDRKSN EQVLREVNENRTLLNYI GRRKSKMIGHIMRHNQF LKNVFEGKVLGKKPRGR PRATYFNNIKEDMGIKS YEELKRMTMERGTWLN RQGIAFSL			
Penelope	Penelope-	.	Locusta migratoria	TTTAACAGATG CCGGAATGAAG GTATTGTTCCTA	TAGCCTTGTTTCGA CCCATCGTTTTGCC GATTTTCTCCACCG	ASVSRIPEDQADEIRQEV AYKLRRAPPPRSNIPSSE RAAIRSLREDKDLVILPA	IPR000477	119	362

	49_L Mi			AGTTTGCTAGA ATAGTTCACTA ATTGATTCTAAT GCACAAGATTA AGAAGCTGCCA GTCTTGCGCTAG TACGTGAACGC TTCGTTTTACTC GTAGACAACTG GATTTTGTTTCC AAAGACTTATA TTTTTTGCATCT TGAGATTTCATC ACGGCTTTCGG CCGCGTCATGG GAGTGGGTTGA TAGCGCTTCTTG GGCCTTTTCCGA CTCGGCACAGA GGGGTGCTGCC TGCAAACAGTT ATCAAATTTGA GCGACTTTGCA	GCGCGCTACTAGTT CTATTGCGCCTAGA GGGCCTCTACTTTG GTTTCGCGCACGCG CGGTGGGTCGTCAG CGGTGTATAAAGGG GCCGCTGTTTCTCG TTCTGCCTCAGTTT CCGGCGAGTAAGTG TATCAGCGCACTCA CCTGAAGATGGCAG CCAGTTGGTCTGCC GAAATATCGTGTCT AGTTGACGACGGTA TCCGGCTGCACACC CGAGATGAATATTA TCA	DKGNATVLLSRGDYVS KMNLLLDDSAYRRLRN DPTERIKRKTFHFSHGTN KTEDLSLLKKSSLSEDV VKKLHPGAAVPPRLYGL PKVHKHGIPLRPIVSNLG APTYNLAKYLASVLAPH VGKCEHHILNSSHFIRRL KSVCLGPSDLLVSFDVV SLFTRVPLEESLLLISDKL DDETVKLCRHVLTSTYF LFNDAFFEQTDGVAMGS PLSPIVANLFMEDFEDM ALRTSFLKPTCFWRYVD DIFMVWPHGMDALNHF EDGKLPFLDVLVQKNSD GTLGHGVYRKPTHTDLY LHATSCHPPHQRMGVLR TLVKRAYAISDAEHLGE ELEHLKVVFKQNGYTA QQICRAFQFGPKGEPLED NYKSIAFLPFVGNISSKIS RILRRFDIKSVFMPPPKT			

				CTCAACAGACA ATTCTGGTGAC GATTTCCGTCGG CGGGTTGTGTC AATCTTACAGA TAAGGAACTGG ACAACGCTGCA TTATCGGTTTTG GAAAAAGGTTT AAATTTTGCACC CACACCTAAGG TTGTCCCTATTA CGGAATTTATTT GTGCTGTTGA		RALLGSVKDDLGLRKPG VYRIPCYCGKTYIGQTIR TVQDRCVEHHRHTRLLQ PEKSAVAEHCLNEGHK MLYDETKLIAPASRYWE CVLKESMEIRMADNIINR DKGFPLSKTWNPILSDIK QQRSSFRPAVSFN			
Vingi	Vingi-2_LMi	.	Locusta migratoria	AGAGGCTGCCG TACCTCATACAT TCCGGGC	TAGTGGTTTTGTAA TTTGTTTATATATA ATCTCTTTTGTTTTG TATTATAGTTGTAC TCTTGTCTTTGACCT TTTGTATGCAACGC TTTTGACACGAATA AATAAAT	MNKELTSKLHEYLRLFD EDPFSTDTIRSGMTLSQE LAKAKQKRWSEMVESM DMSKSSRKAWQLLKRL SNDPTTPKERFNVTPNE VAHQVLLNGKTEGRKE KIRITRHNLETNDLSAKF TLEELKAAIKMCKSGKA PGLDDILVEQIKHFGPVT	cd01650	176	431

						MGWLLRFYNNCLTRRHI PKIWRRTRIIAILKPGKPA SDAKNFRPISLLCHLYKV LERMLLNRLSPVIEPQLI PQQAGFRPKKNCTAQIL NLTQNIEDGFERGLITGA VFVDLSAAYDTVQHKIM LHKLYQLTRDYDFTSFIG TLLQNRKFYVELDGKKS RWRIQKSGLPQGSVLAP TLFNIYSNDQPQPPNTKS FIYADDLALVSQAKDFE TVERHLTYGLKTLSAYY EKNHLKPNPAKTQVCAF HLRNKDANKKLKVIWN GTELEHCHNPKYLGVTL DRALTFRKHCLNIKQKV SARNNIIRKLTNSGWGA QPTVLRTTAIALCYSAAE YACPVWYTSVHARHVD VALNETCRITTGCLRATP THKLYNLAGIAPPDIRRE VAANIERXKVEQNAAHP			

						LHGHQTPPSRLKSRKSF MKTTVNCSTSPSTARQN LWTTRNADIATWIPPLE ALPPGENLPWKTWKSLN RLRSGVGRTKDNMARW NYLGDESTLCECGSEQT TKHLYSCPQCPVSCTLE DLLQAKENAVELANFW SDCI			
Nimb	Nimb-37_LMi	.	Locusta migratoria	CCCTCAGGGGC TCAGCATTAATT TGCTGAGTACG GGCTTGGCGAA CCCGGGGTCCT GAGCTGGGGAC TGGTAAGCGCC GCCAGTCCTCTA GACACCGTAAC CTCTGGGCACG CTTCAGCGACC ACCGTGTGGCG CGGCGGTGGAA CGTTGTGTGTAC	TGATGCTTGTGGCG TGCAAATCACTATA CGCCACATTTTAAT TGAGTGTATTTTAT ATCGTGCTGAGAGG GCAGCGGCTAATCT TGGTCGTGCCCTAC CCACTGTTTTAGCT GACGACGAGGCGA GTGTGACGAACGTT TTACAATTTTGTGA ACGTTCCCGACTCT GGCCTAAGCTCTTA GGCTGGAGGTTTTA	MKKRVDGELVKSDSFIL TFNSPRLRTMLRQASCV SQTPDHVKAGFLRLNVR PYVPNPMRCFKCQRFGH TTLGCKGRLPVANVVRQ PMMGLTVLHLRXVSTV QGTTQSGVGIVQCTXRR RKYRKLRPQGVSPMQRP KRSSRPCNLLYFPPPSLQ YCKNQLLKLMPPPRLRL LVWAPTPALASARVRQ QKLQLRSPTDNADKATG TAFAEPWFHNCRACHH VTATATTPPTPPPHHSCN	cd01650	865	1123

				CGGGGAATGGG	GTGTGTTACAGGAG	HNLCLHHQATLQSLPSR			
				GATCTTGGCTTG	GGCTCGCTGATGCC	HRHRHNPTNATTSPQLQ			
				ACCGCCCGGGT	TCTAACATCAGCGA	PQPLSPPPSHLNPRRLPA			
				CGCGAGGATGG	TCAGCGTCCCTCTA	NTVTGKTVIXLSDFVEVI			
				TACAAACCTCT	TACTCCGACATTTT	PPEPSSDPAPELMELDLP			
				ATAAAAAACCC	CTTTTAAGGACTTG	QGQXSRPNTEPPPRSVSP			
				TTCAATCCCAA	TAAGGTTGGATACT	SRKKERVKIQPPRSNLRD			
				GGTGTGCTACG	TCAGTCTGCCACCT	RWPQYCSGTCRVSGRM			
				TGCTGATGGGA	TACATCGCTGATTT	WRNSVILTQGRPICICLQ			
				TGCATGGCTGTT	GCAATATGTGTTCG	ETNFKSTDNPELRGYVL			
				GAAGGGGAACA	TGTTGTTGCCAAGA	HKKDDLSGERARGGVAI			
				GTCGCTAGCGA	CCTACATTCTATTG	FTNTAYHSEPLSLTTHLQ			
				GCAACCTCTGG	CCTGTTCTTATTTAT	AVAVSVMCVTVAPLSH			
				GGATACTGCCG	TTACTGTTGTCTAA	HTFTSRCSICYVRHRLTV			
				CACCTCAGTTGT	TTCTTTTGTCAGGG	CSLYLPPHDTFDVAALE			
				ATAAGGCTTAC	CGCTAAAAACCTTG	DLLTQLPQPFILCGDFNA			
				CTAGGCATGCG	CTGTTGTGCGCCCA	HNVLWGSASTCPRGRAT			
				GGGCTCTGTCTA	AAACATCCTATCCT	ERLLLSSSAHLLNTGQPT			
				GGTGGACACTT	ATCCTA	HFCTATGSFSAIDLSLCS			
				ACTTCCCTAGCT		PALAHTTQWEVVDDLH			
				ACTCGTGGGAC		GSDHFPIWLHLPDRVEPE			
				CGAGATGGAAA		CRPPQWVISKADWTLFS			
				AATTCTAATTTA		QLAELETRDNVQEWVD			

				CCCATGAACCT		CITQTIHSAAEASMPLSL			
				CCTAGTGGGAT		GQPKRQPVPWWTAECR			
				GGGTGTACCAC		SAIRARRAALRRFKSRPT			
				TGGGGGGTACC		ADNLAAFRMARAKCRQ			
				AACACCCAACC		IIRESRKRSWQQFLSTVT			
				AAMCAAGAGGG		RSTKSSTVWETIRRISGK			
				CTMGGGTAGCC		GGRCPVAAVMGNGVLH			
				AGTCCTCCGGA		TSPRDIAQTMAIYFANVT			
				TTCAGGTTCTCA		ATAHQDPAFQRHRVVA			
				AAAAAAAGTTT		ERGNLDFRSTGHQDYNC			
				AGTTTATAGTA		PFTLWELDSALSVARDT			
				ACAGAACACAT		SPGNDRIHYSMLRHLTQ			
				GCTGCTMMACA		RNKEILLNLFNAIRPGHF			
				GAATGTGTTTAT		PDSWREAILIPLLKPGKD			
				TGTCGTGAAGA		RTCPSSYRSVALTSCVG			
				GAAAAGAGGGC		KTLERMVNRRLVWMLE			
				TCGTTTGAGAA		SRQLLSRFQCGFRRYRST			
				GGTGTCGCCCTT		FDNLASLEAAIQQAFLR			
				CTATATCCAGA		RHHLLGVFFDIEKAYDT			
				AAGGTCTCGAG		TWRHLILRQLHEWGFRG			
				GGAATTGCTGG		CLPIFIRSFLSPRYFRYRV			
				TGMTTTGAAAT		GGVLSDRFEQENGVPQG			
				CAGTAAAACGG		SVLSVTLFAIAINSITSVV			

				TTGCGTAATGGS		KSPVQCSLFVDDFSVFCS			
				ACATTGTTGGTG		SSSLATTTRQLQLTLRRL			
				GAAACATCAAC		DEWAQKTGFKFSTEKSV			
				TTCCCAACAGG		CVLFNRSRSIFNLPELRM			
				CGGCCAACCTA		RDTVLNFKDTVRFLGLT			
				CAGACGGCTAA		FDCKLTWLPHLKDLKRR			
				GTGCCTTGGGG		CLKALSILKCLSHTSWG			
				AATATCAAATC		ADRTRLLQFYRAFVRSR			
				GAGACGGAGCT		LDYGCTVYGSARPSYLK			
				GCACAACACCC		MLDVVHHEGIRLATGAF			
				TTAACTATAGC		RTSPIPSLCAEAAEPPLSI			
				AAAGGCGTTGT		RRQLLMVRQAYAALSTP			
				CACCTGTAGGG		YTPAYNTVARPPMERLF			
				ATCTGGTGGAC		RNRLRATRPFGIRAKDC			
				ATTCCAGCGGA		FLDMGVSDFTVLRRGW			
				TGAACTGAAGG		SRYPPWLLQRPSLLLDLT			
				ATGAGTGGGCT		SFKKDCTPDITFQSEFFNI			
				TCAGAAGGCAT		LNKHRDCTVVYTDGSK			
				TGTTGATGTGCA		QGDSVGCSAVFPGHIVGI			
				AAAC		RLPAAYSVFSAELHAILQ			
						ALEHIRRARGTKFLLCSD			
						SLSALQSLLVMYPAETIT			
						QLIHDQLYLLQRRGKEV			

						SFLWVPGHVGIWGNEA ADVAAKAACSDPNVTE CAIPLRDVYSVLRQRSL QLWEETWLAVTGNKLR LVKPTVRPWRTSCQPFR RFEVTLTRLRIGHSPLTH SFFIPSGGSTDL			
Penelope	Penelope-46_LMi	.	Locusta migratoria	GTTCGTAAGTTT CCCAGGAACAG AATGGAAACGC CCATTCTTGTAA AATCGAACCTC CAGCCAAACTG ACAGCTGAACT CCGTGGCTTCTC TCCGATCCTCTC ACAGTGTTCTGC CAGCGATAAGG GATGCTACCTG ACGCCGATTAT GATGGAAAACC GACAGATGTTG A	TGACGTCACTGGCG GCGGCGCGGCTATA TAAGCACGGCAGC GGCCGCGAAATGA CAGTCATCCACTTG ACAATGGCAGAGG AGTTCTCTGTCGAA AGCTCGTGGGCAGT AAACCACATGACGC GGCTTGTAACCCGA GAAGATTTCATTAA TGGATATCGCCGCG AAAACATGCACTCA	RNTNSELRKSELKDPGR YNYLGIVTIGLLQDLPNT CTLLSMHVGHCEHHIKN TADFIDRIKGRRLREEDI MVSFDVVSLFTCVPITDS IDLLAQLFDDTIVDLFKH TLTSSYFLYGGDYFNQT DGVAMGSPLAPVVANL FMEHFEEIALDTAPAKPS CFYRYVDDTFVIWPHGQ KKLLLSLRRRYLRHLAS WTEKIEEFLEHINSIHGHI KFTMEVEKDGSLPFLDV LVRRKPDGSLGHSVYRK PTHTDRYLHATSYHHPA QKQSVLKTLVHRAETVS	cd00304	126	229

						DAXNLSNELSHLRKVFK DNGYSKRQISQAVSLKA RNQKTSEEEVKKTAFLP FCGSTTGKISRLLKRHKI DTIFRPPAKIRQLMKPVK DDVGLRTPGIYKIPCECG QFYVGQTVRTIEQRRIEH ERCFRLRYPEKSAVVEH ALENGHRIAFDETSIIKR TNSFWDSVINEAVEIKIS ENTLNKDGGLQLSTAW DPVIAELKRARRTRDQNI ALYGQGTGHQ			
Vingi	Vingi-45_LMi	.	Locusta migratoria	GACGAACAACA TGTTACGGAAT AGCAATGTATG TCAGGATGAAC ATAGCTGATTAT AGAGTGGTTTCT AAGGACTCTAC AGACAACATTT ATATCTTAGCA GTGACGATATG	TAGTATTTTCTCTG GTTTGTTTGTTTGCT TGCTTTTATTATGC ATATTTGTACATAG TTTTTGTATTTGTAA TTCTTCTATATATG AAAATTTTCAAATA AGACTGTAGCATGA CCATGTAAGCCATA	SPIQSHGFKPRAGGSPGK CAYKQSGCFDGIYPEFL KNAGPKTKRWLLQLYNI ILTTGKLPTTFRAAKVIA VRKPGKEGSEASHYRPIS LLSVVYKLLERMLLNRI QPLIDKATPLHQAGFRK NRSCCDQVLALTSFIEAG YQRKLKTAVAFVDLTA AFDTVWRHGLLLKFAE	cd01650	65	319

				TGGAACTACTA	CGAATAAATAAATA	AIPCQKLVNLMDSMLSN			
				TTGTCAACATCT	AATAA	RRFQVFLGDKGSRWRRL			
				ATAAACCCCCA		NDGLPQGSVLSPTLFNL			
				TTAGTCATCTGG		YTSDMPNTSAKKIQFAD			
				CCTACGCCCCC		DLALAYQARNLEECEGT			
				ATTCAAACTTTT		LTTDLHIMGDYFKSWRL			
				TACTGACCCTGC		KPNPTKTEVNAYHLNNR			
				CATCTATGTTGG		EANRKLNVVFENVPLEH			
				AGACTTTAATA		NPYPKYLGVILDRALTF			
				GCCATCACCAT		KNHLEYVAKKVNTRIDI			
				GCTTGGGGATA		VRKLAGSGWGANADTL			
				CAAGAATAATG		RTATLALVYSTAEYCAP			
				ATATAAATGGT		VWKNSAHVRKVDVKLH			
				AACATTTTACAT		QAMRIISGTTKSTPLPWL			
				GACTGGCTGTC		PVLCNIPPPSIRREAAAIR			
				CTTGAATAATTT		EFHKSSTHKNSLLYNQL			
				GTATCTGGTCTT		RNIPTSRLKSRKPPWPM			
				TAGTGCAAAGG		QDSPLVYDVCQEWKDV			
				ACAAAGGAACG		WKKFPPANSRLITDPTK			
				TTTCGATCAGCA		RVNGFGLPRKLWTSLNR			
				AGATGGCAGAA		FRTGHGRCSACLYEWG			
				AGACACATCAC		YKDAPACDCGAPVQTM			
				CAGACCTATGTT		QHLIEECPPRKFPGGLSE			

				TTACACTACCA GTCAAGCTAAT GGAAATCAACT GGCAAACAGGA AAGTTATAGAT TCTTTTCCACAT AGTCAACATAC GCCCATTATCAT TGACTATGGCA TCCAGATACCA TTAGTAAATTCA ACTGCCAAAGC TCGTTGGAATTT TAAAAAAGTGA ACTGGCAACAG TATGCTGACAA ACTTGATGTGA CTGTTGACTCAT CCCTGTGCAATT TAATAACTATG ACAGATTTGCT AAAGCTGTCAG AGAGGTAGCAA		LHLVTTRAVEWLRDLD VQL			

				AAAGAACGATA					
				CCTAGAGGGAG					
				GAAACAAATAT					
				TCCAGGATGGT					
				CCAATGATGTG					
				ATGAGCTATAT					
				GAGGAATACAA					
				CTTACACATTCA					
				AATAAAACAGC					
				AGACGAACTTC					
				TTCAAGCTTTGA					
				ATGAGCACAGA					
				AGAAGGAAATG					
				GCAAGAAGCTA					
				CAGAGAGCCTC					
				GACTTCACCCAT					
				TCGAGTCGTAA					
				AGCGTGGCAGC					
				TATTGAAGAAA					
				CTGGGTGCAGA					
				CTAAGCCTGAG					
				AGATGCACCAT					
				AACGTCACACA					

				AATGAAGTCGC GACACGCTTATT GCAAACTCTAA AGCATCTGTGA ACAAGAACCGC ACTAGGCTGGT GAAGAACAGTT ACGATCCAAAC GCTCTCAGCTA G					
Penelope	Penelope-120_LMi	.	Locusta migratoria	AAAACGGCCTT GACTATGCAAA AATGAAGTTCG CAAACCAGAAT TGCTTTGACGTA GCTAACATAGT TAGTAATAATTC ATCAAATGCTG TGTACATGATA GTTAATCTTAAA TAATAAGCTAA AGAACCGATGC TTTAATAACCA	TGATGTTGTACTTC GTGCAGTGTCCGAA ATATTTCTTCCGAC TTGATCCAAAGCAT TTGTATTCCAAAAG CCAATTTTTAACTG TTTTGCTTATTAAA ACTTTGTTACTCGT TTATTCTCTCTACTA ATATTGTATTGCCA AAATGACGCCTCAG CCATCTCTCATGTT AAACAAAGCATTTT	MNPQPPKLRSQFKVHKL NHPIRPISNSLNSAYHDL ARFLHETLKKSFVFQNN YSVPNSHVLVHKVKDL DCGPDTKLFSLDIKNLYT NVPVVETLKIVEKNLCQ FKKLSSEQITDFMNLINV VVKYNYFEFNGKLYQQ DDGLAMGNPLAGILADI FINSLEEKFFHKFSTTTL GIVSYSRYVDDILVIYEG PTTGIDHIFTLFNGLHENI TFTCELENEARQLNFLD	SSF56672	34	231

				AATCTGACAAA GGTTGTTCCTTG GTTATTGCTAGC AATTCTGAATAT GTCTCTAAAAC CTTAGAATTTTT TAATGAAAATA ATATTTCGGAA CTGCACGAGGA CCCGACCGCAG AATTACAAAAA GAAGTCAGGTC GGCCGTCAGTG ATGCAAAATAT TTGCTTAAGCCA TTTCATAAAAA AACTCTCGTTAA T	GTAGCAATTGTTTT TTCAATACTGTTTC TCTCTGTACAGTAG CCTAGTTGTAATTG CGGCTACTAGATGG CACTACTGAGGTAA CGTCTTCCTGTTCA CAACTCCACTTAGC ACACGCCAGTCTGA TGCCTCCGTGCAGG CGTGCAGCCCATAG TGGCTAGCTTTTTC TAAAAATTTCTTTT TAAATTTTGTGACG TAAGCTTTACAGCT GAGTTGTATATTTT ATTATGTGACTTGT TTTATTTTATAATTT CTGATGAAGGCAAT CATAGTCATTGCCG AAACCAGGTTAAAT AATTTTTAATTTGC GACCGAGGGCTGC	LTLTVEENKISFNIYRKE TYTDQIVPASSMHPQCH KNAFFHSAIHRAISIPLSK EKFEEEINLIKTIAVNNE YDPAIVDNILKKKTAKR VTTLGTPIESDQKKKYFS IPFLGPISYQIQRLLRNKY NCNVAFSTNNNLKKNFI HNLKSTCSPLDNSGVYK IVCDTCSSYYIGQTGRAF KIRFKEHLLRKNGTNPQ NSSFADHLLITGHAPKAI HDLNILHTEKKGHRLDV LEELQIFKHFSRNDGLIL NEQLQLRNRSFFNGIKPL LGL			

					WTTTTTATCCTTTG ACGTTGTAACGGTC GCTGACAATAGCAG CTATGTTCAAAATT TTA				
RTE	RTE-34_LMi	.	Locusta migratoria	TAGAGAAACAG TAGGAAATGAA AAAGCCTGCAT GAGGTGTCAAA GACTATAACTTT GCC	TAGTGACTGAAATT AAGTGTAACAGATT GAGATGGCTGGGCC ATGTAGAGCGAATG TCGGAGGATCGAGC TGTTAAAAAGGTCT ACAAGGAAATCCA GGTGAAAGAAGCA AAGGGAGGCCTCG GAAACGATGGCTG GAGGATGTGGAAA ACGATTTGAGAAAG ATGGGAGTAGGAG ATGGAGGAAACGA GCAGAAAATCGTCA GGATTGGGCTGCGA TCATCAGAGAGGCC AAGGCCCTACATGG	MGNAVVVKVHGSRETQ ESTRCSPGTTFNQTDIGN HRGMHRLIAAGRLQGG DRSLYGKMEYTRDPEEG KALKTQCFQIEGKGAKP EASGGSTIKEALQETAEE VLGLEERVKRADGTMKI VKRLWKKEMRQGRKG KEENDKQDNKIYYRKF MKKRKEADRICRRKKRE LERDGLWSRGMDYGAR EEYNDEEVRKFYKKVRE LKKGFQPRTLQRRKASN HVLCRDKEGNFIGGEEQ VLDQWAKYFSELLNVTT EREDKMSLQKEKTNETE GDQPWQNVSQDERPVPI PTIEEVRKAIKSLRNNKA	cd01650	345	632

					GCTGTAGCGCCAAG GAGTAAGTAAGTA	PGNDNISAEMIKGRGEH LTRLVHQLIVEIWEKEE MPDQWNMAIICPIHKKK DKANCENYRGIALVSVV YKVLAKVIAMRFAPITE EILRRLPVRLSAGQINNR PNFHYKANNGKVLDLH QLQKKYSGDYQCGFRQ GRSTTDQIFIIRQIMEKCY EYNVDVHQLFVDFKQA YDSVKRKKLYQTLMEM EFPSKLIRLVQMTLTSTK CQVKIEGKLSKEFEVNQ GLRQGDVLSTLLFNIVLE RVMRRVEIDNPGGTLFN RMNQNLAYVDDVCMLS RRLKELEEGFDRLEKEA EQLGVNNTKTQYLFASR KNAVVKEKYIKLNRGN YERCEKFQSWGAVTEDS RQIREEIKSRIAAGNRAY WAMIKILQSRSLSRKSK VTVYRTVIRPVVMYGSE			

						TWTMTVAEEELLRRWE RKILRKIFGAVIEVGQWR IRRNKELEELYRNLT			
RTE	RTE-44_LMi	.	Locusta migratoria	ACCTGTTACGCT GTCGTAACAGG CGGAGAGGTGA TACTCCGACGT GGCGCGTCCCA GGTGGCGGATA KGGGGGTCCTA ACTGGCTTGCC AGCGGACTTGA GCGAAATAAAA TAGCTCTCGCG GACCAAACACA AACCCTCTGGA TTAACACTCCA GTTGTATCTCGA GCAATCCTCGG AAAGCTCCACC ACTCCAACCTTA CAAATKTTATG GATTATCAGTG	CAC	MTIKCKNKMYTLVNAH APTNIDNKKNPTKVQTF WEELEDTIQKIPKENIIIL LGDFNAQLGKERKYKNI IGEYPAHKRTNNNGIRL VELCQVFNLKIMTTHFK KHPKKMKTWKAPNQLL GEFQIDHIAISRREFKTIQ NIKVQKGAKIDSDHYLV KSKIKFIPNSVKNNKKPT RKNFETNNIDADTKNIFK SKLQGETWEEMKSQIM KAAEESLPTKIKRKNIW WNDKCNEALDNRQTLW QKFKSNPTQINLERFREG RKIASYTIRNEKRLFEKL SLQQIQVHFTQHKSRDY YQSFKRHLNGYEPPSLN FRDKSGKLGTNDQENCN ILAHYFKDLLNCKAPPE	cd01650	419	678

				GAAAAGAAAAC		KPEFKTTEEIPAESHPPSL			
				AAATTACCCCA		DEVRQAIKSLKNNKSSG			
				GGGAGTAAAAT		EDGITAELLKLGEEKVTL			
				CTCTCGACCTCC		HLGTIIKEIWQTEKIPED			
				GGTCGCAAGCA		WRTALIHPLHKKGDRSD			
				GGCATTTCGGA		VNNYRGISLLSVPYKVL			
				TTCTGGGGAGC		SKILLNRIEPVLDAQIGE			
				CTGCACTTTCTG		YQGGFRKGRSCSEQIINL			
				CTCTTGTTTTAC		KNIIRTRAGQPHVTTFID			
				CAAGACAACCT		FRKAYDSIDRESLIRTLE			
				GAAAAACTTAA		EFGIDRKTREIVKQTLTD			
				ATTCAAGAMGA		TKSKVKFRGKISDPFKIQ			
				AATACAGTATC		TGLRQGDGLSPILFNITL			
				GGAACACTGAA		EKIIRETLAEYTQEKVKG			
				CATTAACTCGCT		ISFGGSVKGLTIHALAFA			
				CATCAAAACAG		DDIALLSTNIXDAIKQIEI			
				GAAAACATAAA		LAKHAAKFGLQISQEKT			
				GAACTGACTCA		KFMSNKFKEVNKLHTII			
				AGCACTTGAGC		GCIERVKHFKYLGETITD			
				ATCATGATATCC		TGTEKEAVKLRVAKLRR			
				AAATTCTGGCA		AQILTMKLYNKKCLSRN			
				CTACAAGAAAC		AKLL			
				CCGCTACACTG					

				ACGAAAACTCT TTTGAAACAGG CAACTACAGAA TTTATAAAGGA AAACCAGCAGA GTATATCTTTGG AAAATGCCCCT TATTAGGTACA GCTTTTATTGTT AGAAAAAATAT AACGGAGTCCG TAATAAACTTC AATTCTACGTCC AACAGGCTCTC AACA					
CR1	CR1-2_LMi	.	Locusta migratoria	AGTCGCCGAGT GGACGTCGTCC AGTGCACACCT GTCCAGCGCCA GCACATTTTCAG TTGTAAGTGTA ACCGGCTTCTCT CTAGAAGCTAC	TAAATAGTATAAGG TGGTTCGATGAACC CTCTGCCAGGCACT TAAATGTGATTTGC AGAGTATCCCATGT AGATGTAGATGTAG A	MNRDCDCCVRMQSELV TLRSQLQAVLTSVAQLE AAIEGRRCCWPVVGTER TPVVSTDRPAPATCPDT ASSEADTPPVVAWEVAP VRGRRRKTFRGTDREAS PVRLANRFEALSVQDNV PESDGVSHPEVTTQPARS	cd01650	852	1140

				ATATTTTGACTT		GLSQRWVCLRDGCQGG			
				TTCAGTGCGCTT		EEGQRALRVHTGGVIPD			
				TCCCTTCAGTTC		VERVLPDAMKGTGCSQ			
				ATTAAAGTTATT		LQVVAHVGTNDVCRFG			
				GTGTGCTCGCGT		SEEILSGFERLSEVVKTA			
				TTTTATTTGATT		SLACKMKAELTICSIVDR			
				CATATTCACGTT		TDCGPLVQSRVEGLNQR			
				AAATATCTGGTT		LRRFCDRVGCRFLDLRH			
				AGTGTATTATTC		RVVGFRVPLNRSGVHYT			
				CGCCGTGTGAA		QEAATRVAGAVWRGLG			
				TCGCGTCGCGA		GFLGVRSPLYAGSGYTG			
				CAGTCGTCTGTC		SGGCVAWTGRFFRIKSF			
				TTTGTTTTCTGA		GEIENGQSQRVQVDHRR			
				TTAGAAGCGAC		NVDPETIGITVVNCRSCV			
				ATATAGTATTTT		GKVPELQALIESTETHIII			
				GATTTTCGTCTG		GSESWLKPEISSEEVFSK			
				CGCGTTCTCGTT		DLTVFRKDRLNKIGGGV			
				TCAGTTTGTAAA		FVAVKSSISCSEVEVDSS			
				TTTATACTGTGC		CELLWVEVVPENRKKLII			
				TCTCGTGCATAT		GSFYRPPDSDDTIAERFK			
				TAGATTCAAAT		ENLSLITKTYPTHTIIVGG			
				CCGCGTTTTAAC		DFNLPSICWQNYIFKSGG			
				AAACAGTGTAG		RHKTSSENALNAFSENY			

				TCTTCCTTTGAG CGCATCGCGAC TGTAGTTTTTGT TTTCTCGGCACT GTAATC		FEQLVHEPTRIVNGCENT LDLLATNNPELIMSIKTD TGISEHRAVVAKLNIESP KVTKNKRKVYLFRKAD RNKIHCFLKDSFHSFLTL NVGVDQMWLKFKEVVS TAIESFIPNKLIKEGGDPP WYTKQVRTLLQKQRKR HTKFKRTRNPQDWRSYT EARNLARTSMRDAYNT YHNETLSRNLTENPKRF WSYVKYTRGKTQTIPSL RDSNGDPADDSATKAEL LNIAFRNSFTKEDETNIP DFELRTEARMSNVEVDII GVVKQLKSLNKNKSSGP DCIPIRFLSDYAEAIAPYL TAIYNRSLKEGSVPKDW KVAQVTPIFKKGSRSNPL NYRPISLTSICSRVLEHIL YSNIMNYLEQNDLLTQG QHGFRKYRSCETQLALY SHEVMSAIDKGSQIDSVF			

						LDFQKAFDTVPHKRLLI KLRGYGISPQICDWIGDF LSERTQFVVIDGKASTKT EVISGVPQGSVIGPLLFLI YINDLGEICYWSSAVPY LHKRFRGNLSSSLRLFA DDAVIYRLVKSSEDQNQ LQNDLNKISLWCKNWQ LTLNNEKCEVVHMSTKR NPLNFVYKINQSKSRPHE YEKKSVKLCLQDKSIKV KAVNSTKYLGITITSDLN WKEHVDNVVGKANQRL RFIGRTLKRCNRTTKETA YTTLVRPFLEYCCPVWD PYQVGLTEDIERVQRRA ARFVLSRYRGDSVTEMI QDLGWESLQQRRLSLRR GLLTKLQSPTFSSECENI LLTPTYIGRNDHQNKIRE IRARTERFRCSFFPRAIRE WNGR			

RTE	RTE-14_LMi	.	Locusta migratoria	CTCTGAAAGTACTGGAGATTCTAGGAAACGCCAATACGGTCCTAGAATCAATGAAAGGAAAATCCTGACATTCACCTCCACGGGGTAGAGAGGGGGCAACGTTTGCATAGCAAGCGGCTAACGAATGCAGGGACCGAATATCCCCCGGGATAAGGGGATCCCCTATACCAAGTGCCAAAGGCCTCCATGAGGGCGGATGAGCGGGTATAGGATAGGGCACTCTCTTGTCCTTGGAGTGTGAAATCGCTCCTAAAG	TGAGAACGAGAGAAGAAGA	MMTGFVTKQSGGKIVPNSDLRQGTVSKCVIKDDRRQVRIGTWNIRSLSICGKLENVKKEMEEHDIDILGISEIRWPQKGDFWSGNFRIIYSGKEGSYAGVGIILRKHVGTRVKSYVQYSDRIIMLRIENDPIDTVIIQVYMPTSSHADEEVEQVYEQLEEIMEMVKDKDNLVILGDWNASVGEGEDGNTVGKYGLGRRNDRGQQLVDFCKEKELIIANTWFQQPKRRRYTWIAPGDRARYQIDYIMVRRRYRNQAKQCKSYPGADVATDHNLVMLKSNLRLRRKGQNTKGYCKYDVKKLRQDSVKERYQKQAQQKVKEIPNGTVNEKWIGLKHSILQAADREIGNIKKQATQPWISDHTLELMEERKKFKNSKSEEGIKSYKRLRNMVNREEGIKSYKRLRNMVNRE	cd01650	545	801

				GCGGAAGAACC AGCAAGGTCAA CGGCATAGAAT GGAGAAGGCAA GAGGGAAACCA CTCCATTAAAG ATCCTTTTGGAT ATCCCATGTCA AGCAGCA		AKKDKEEYIMRRCKEVE EKLEKGNIETAYELIKKQ FKKVNGRSNNILDENGK RLFENKEVGERWKRYIE KLYEDPDEISNLENEETV TKDQLGPSLTKEEYLKA VRELKNNKAPGCDALT GEMIQGLDNETHNYIFK LIQDMYETGEIAEDFKK TIIIPIPKKPGTIKCEEHRT LSLISHMSKILTKIIYNRII SKIEQNLDEDQFGFRRN RGTREAILMLRQIMDKM LDLKRPLFIAFVDLEKAF DNVNWNLLMEIVTNVG LDFRDRRIVHELYKNQR GVINIHGEIQEAHIKKGV RQGCNLSPVLFNLYIEEG IKECKESMEGIKINGVQI KMIRFADDIAIIEESEENL QKALSRMNEILPAKYNM KINTKKTEIMVCSHEQIT TDIKINDVPLKQVNVFT			

						YLGSKITPDGKSSADIKS RIAQAKLAFNNKRMLLT SRNIKEETKKAFIKCYV WSVALYGSETWTLGVC DKRRLEAFELWCWRRM QKIPWTAKMTNEEVLRR IGEKRALVKTIMKRRDK LIGHLYRHNEFIINVMEG MVEGKRGRGRPRLQYL KQIQRDIGVESYQEMKR LSENREVWRRRIAANQS SD			
Daphne	Daphne-8_LMi	.	Locusta migratoria	GATCTTTGGCGT GACGTCACAGT GCTGGATGCTG TGCACATCGCTC TCGATTTTTTAT AATTTTAGGATT AATATAGTTATT TTAATTGTTATT TCGTGATAATAT TCAATATAAAG GTGTTCTAGAA	TAAGCTAATTCATT TTTTTCTTTATTGAT TCTTTTCACTTTGTC TTAAGATTGTCGTG TGCATAGGACTACT TAAAAATTTTCTAT TACCATTCTTGTGT AAAAATTAATGTAT TGACTTGTTCCATG ACCATGGAGTCTCA CTCCTCAATTGGGT	MIEEQERLICALQVQLK NAEEELGRLRRGRGAGE WELAVGKKVARRRYSD SFTLQVANRFDQLSELS GDEPPVAVGVGSTQQTS AFGKPKSVERSLRKKKV LLLGSSHGRGVGQQLQE VLGSEYQVTSIVKPSAGL DQVTCNIGDLCRNFSKE DQVVIVGGAGNSLDRD GEYDIGGDLERIATQTG	cd01650	851	1112

				AAGGATCGTCA	CCTATGGAACAGAA	GTNVHFVQLFQRHDRPC			
				GTAATTTTAGCG	AAAACAATAAATA	LNTAVSRVNMGLEKAL			
				TTTGTAGTGCCG	AATAAATA	MTEGMDHIAVVPVETIQ			
				TATTGCTTTAAG		RSDFTRHGLHLNRYGKG			
				TGGTGTTATTTG		RLAKLIGDSVVGGGGXT			
				TGTTCGCGTGTG		HGKIPVVVGVRAAPFLD			
				CGGCACCGACG		GCTFFRLKTADRYSCLK			
				TCTTACTGCTTT		EVPLRXEPPLTKSKNISN			
				GTTTACATCTGG		KGITVFHQNIQGIRDKVN			
				TGGCGAGTGGT		EXLIDTDSDIIGISEHFLS			
				TCCAGCCGTAG		GVTIQRLPLPGYRLAAF			
				CGGTCTCCGAA		YSRSSLQCGGVAMYVK			
				AGCTAAGTTCC		KRFFIGVNRFFKNDFSSE			
				AATAAGTGTTT		SIDFSKHCTEQIFECCAG			
				GTGTTATTGTAG		VVNVNGTKLVIVVIYRS			
				TTAAATTATTAA		PDSDFKMFLPKLELVLG			
				AGTGACTAGAT		SLYRKYKNLIVCGDFNI			
				TCAGGAGGTGT		NCVGDSSRKSQLLDLFN			
				TATCACGCTAGT		SYNMMQTVFFPTRVQG			
				TGTTTATTAAAT		NSSSTIDNIFVHSSLLEGH			
				TTAGTACTTTTC		SVSRIVNGLSDHDAQLLI			
				AACAGGTGACT		LKSICSQTNVKHKYQLL			
				TGCTGTTTTTAA		RKGKVFVLKQMLNININ			

				CGTAGTTGAGG		YSEKVNPVAIETFLSLIK			
				TTGGGAAACTTT		EQEWQDVYSADNADDK			
				AAGTTACGGTA		YNAFLNTFIMLFESCFPL			
				TTTGATAACTCT		ERSKQGTSSNRQPGWLT			
				TKTACTCTCACT		KGIRVSCRTKRELYQNV			
				GTGTTTAGTTAG		RSSQNEDAVAHYKQYC			
				TTTATTAGTGAC		KVLKNVIKNAKSMWYA			
				GCTTGAATTTGA		KRIANSRDKIKTIWSVVK			
				ACTTGTTTATAC		EVSGQQQQVEDITSVRS			
				AAGCAATAGTG		RNVSVTDKLDICAIFNN			
				TAAAGTTTCATG		HFLNIVGELNKTLVSTG			
				TGGTAGTGTTAT		NHITFLENAFPRLLPEMV			
				TTGTTCTTTTCT		LHDTSIGEIESIIKSLKTK			
				GATACTATTAGT		DSCGYDGVSSRILKYCSP			
				TGTTGTTACTTG		YVSPVLSHICNFSFQNGQ			
				ATTTTGTAGAG		FPDRLKYSVVKPLYKKG			
				ACTGAGGTTTGT		ERDNVDNYRPISMPSVF			
				TGATTGCTGTAG		SKVIEKAVYSRLIDHFSS			
				TTAAGAGGTAC		HNLLSNVQFGFRSGLTT			
				TTCCTAGTTCTT		ENAIFSFLCEALDGLNKR			
				TTGTAGAAACG		LRTLGVFFDLTKAFDCV			
				TAACTTAATAA		DHKILLQKLEHYGIRGV			
				CTTAGACGTCTA		AHNWFTSYFNNRQQKVI			

				GTTTTTTAAATA TTGTGAAATTTT ACCATGAGTGA AAAATGTGGGC TCTGTCGTAGG GTCGTGCGTAG TGGGCTGCAGT GTGGGATTTGT AGTAAGTATTTT CATGGGGGGAG TGCAGTGGGGA AGACAGTGGGC ATTGTGGTGAG GTTCTCTCCTGG AACTGTAGGAT TTGCAGCAGGA ATAAG		LRSVENGYDVQSQWGTI KCGVPQGSVLGPLLFLIY INDMPSSITGDCKIFLFA DDTSLVVKDVARNIDTV SNNAVQEISSWLVENKL MLNHSKTQFLQFLTHNL TKADILITHNGHMISETE QFKFLGVHIDSKLSWKA HVQDLVPKLNSAIFTIRT ISEISDFSTRKVVYYAYF HSLMTYGIVFWGNSPDA KRIFLAQKRAVRAISGVS SRTSCRPLFSSLGILTLAS QYVFSLMSFVVNNISLFP RVSSFHSVNTRQKSNLH LHHTTLTLVQKGVQYSA ASIFNKLPQELKNLSTNP HAFKSKLKSYLMAHSFY SVEELLEKLNN			
Jockey	Jockey-7_LMi	.	Locusta migratoria		TGACCTCTCTCTGT AGATTGCCGCCGCC TCCTCCCTCCCCCT CCCGTCCCCCGCTT	PPSVVPIMAGSPHRILYH NIRSLPANKPLFMHTLSQ HCVDAFLLNETFLQPRH SVSTTPYTLHRTDGPPPL	cd01650	551	810

					GGCTCACCCCCTCC	ARGGVAIGHLRHHPVRP			
					TTTCTCCACCTCCC	QPPVNNPAEHLILTLFLP			
					CGGGAGCTCCTTCC	SLTVTLATIYIRARVPIPY			
					TTCCGGTTTCCCCC	DFISHIDNSFTTYVIAAD			
					ATTCTCCCCTCCCC	LNLHSRSAAELQRWHQF			
					CACCTCCCTTCCCC	LPTLRGDLVPLPSPTRPE			
					TTCCCCCGCGCCTT	SNTTPDAVLVSPNLWDA			
					CTCCATCCTCCCCC	CRRKSLTSLAVTTPPYSS			
					CTCCCCCTGCAGGG	PSTPAAVPAAPPVLCPLL			
					CCAGTACAGTGCTA	GPLGRLQAEVLDLIGSD			
					GTGTTTTTAAGTGT	HAPVLLTFNPRGRPGRA			
					TCAGTGCTTCGTGT	TRPVPPPRLVHDYSRAD			
					GTATAGTGTTTCGT	WDAYRRSINTSVASHPL			
					CGCCCAGTGTTTTT	SLHTPDDIVHASSFLHQT			
					TAACGTTGCGTACA	VADAVATHIPTKAIHPH			
					CGCCTGTGGGAAAC	RPPLPPQAVLLLRESRRL			
					CAGTGTTCAGTGTT	YRSFLRTRDRDTLLRHR			
					TTTTACCTCAATGT	QLQRHIRNLLNAKKRRD			
					CGCCGTCATGTTAT	WRQTCTRLNATLPINSS			
					GTCTCGTCGTCTGC	KYWKAFHRLTGSHSVP			
					AGTGAGTGTTTTTT	AYPLLHNNLPLPDNASK			
					AGCCACCGTGGCAA	AQLFADHLSNIFDLPDNP			
					TGTCTTGTTTTAAC	RFDYSRFPDVHERTDTS			

					AACCATGTATGTTC	VPPLAPGFQYLDHIPPSD			
					TGTCCACTGTCATG	LNTPITAQDIVHTLRTKR			
					TTGTCTTTTAACATT	NTAPGHDRITYRHLRAC			
					CTCCCATATTTTCA	PPSFHSALASLYNVILFT			
					TCTTGTAATTGTAA	GFYPDLWKTSRILMFLK			
					AAGCGCCATTTTAT	PDKPPSEASSYRPICLTS			
					CACCTTTAATTTTA	VFSKVLESILSRRIHEHL			
					TATTCTGTTTTGTCC	RLHQLLPLSQCGFRPSFS			
					ATTGTTTACCATTC	VDDQLLRLVHLLSVQVN			
					GCTCGGCTGTAGAG	SRRAAILVSLDIEKAYDR			
					CAGCATCATCGGCT	VWHSGLLFKLQTYALPI			
					GCTGCCAGCCTGCC	NYVRLIASFLSKRPSYVT			
					TGTATTAGGGCATT	VNNQNSRLFYPSAGVPQ			
					AAAATTCAATAAAG	GSVLSPLLYLLYSADMP			
					GAAAAAAA	RPQPPVHLLQYADDTAF			
						LALAPSLARSQHLLQSH			
						LSEFSSWCNQWLLTVNP			
						TKSQAIIIGRTTPSFRHRD			
						LYLTLHGRPIPQTKTLRY			
						LGLTLDRDLSWNPHLKT			
						VRQKALSRLNLLKCLSG			
						RTWGLNPSTILYTYRSL			
						MRPVLCCFHPAWISAPP			

						SYYESLKILERHALRLAY RIRLPSPTRILYDLIPFPH LVQFLQHIRLKYIHRKIE SHNPLVSPLLLDPGPLPR LYRHIPPSLHLHTLYVLS EGGFRHLPLPDDVLHNR PDIYPSFQL			
Penelope	Penelope-9_LMi	.	Locusta migratoria	CACACATGTAG TGATGACTGAC ACTGTTATCAAT AAGTCAAAAAG AGTATTGCTCA AGACGAGATTT CCGTCCTTACAA AAGGTGGAAAC TATGCTATAAC GCCTGTGGAAG TACAGTTGAGG ATATTATTGCGA ATGTGAAGCAG GAATAAGAACT CTCCCGAAAGA AACTGCAAATG	TAGTGATGTTTTAT CTCTGACAAGGACT ATCTCTGCTGATAA CGTGACAGCTTGGC CACGCCCACTTTTC ACAGTATATATGGC GCTCTCAGACACCC GACTCATCAGTCGG CAAGACTCAACGCA AGTAAGAAGACCTC CGAAGATGTCCGGC GCAGCTCTGGACGA AACGTTAGGAGAA ACTGAAAAGTTTTA TGGACCACGGCCTT ATAACCCGGAAGTT	MIGERKALKELNADNDI VVLPADKGNATIVLNKD DYESKILDLLTSGQYRK LKKDPTASVLKKTNTLI KSSTSIPKEDKKALLKSE AMCPRLYGLPKVHKINI PLRPIVSAINSPTYEIARY LATCLQPYIGKTDSYIKD SSHFIGKLEGLTLRPEDM LVSFDVVSLFTMIPVNEV MTFITDIFPEDLTTLFKH CLTSSQFQWGGDFYEQV DGVAMGNPLGPAIANFY MEKFEQSALGIAKKKPV LWYRYVDDTFVVWPHG KEALDEFFSHLNNINKKI	cd00304	206	298

				CAATTCGGATT GAAACCACCAG AGTACTATGTC GCAGTAACCTC TGAAAAGTAAT TTG	CTACCAGCAACTAT GATCCGGTCGTGAA AGCCTTCATGCA	QFTMEMEKDSAISFLDV LVMREVDGKLKHKVFR KATHTDRYLHKNSNHH PQQKRGVIKSLVDRAKK ICTPEHLDAELKHLKQA FKKNGYSDKEINRVLRP NNNSLRTRDQTKRWKN TVSLPFIGKVTDQIGKIL RKHNIRPVFRPTKKICQV LRSVKDKRPTLSACGVY KIPCTCGKVYIGTTKRSI ETRLKEHKSLCRLGKTE KSAVAEHALQSGSHIVR FSETSVLSTMTNYYPRQ YREAIEIYKHKENFNRKE EAMKLSDIWTVALKDR QVLSLTDCNR			
RTE	RTE-25_LMi	.	Locusta migratoria	CCCGTGTGGAG TTTGCTGGTCTT CCATCGGGCAC CTCCCCAGGTG GCGGATAGGGG AATGCTCACCA	TAGTGTAAAACCTT ATGTACTAGGTGTA TTCATTTCTGGGCG TATTAGTATGTTGG AGGTAAACCTCTGT AATGAGGACAATCT	MPRKNWNCGRRTGDEK RKMMFGCWNVQGISTKI DLLPAELDMFNIDVVVL SETKRKGKGEEELDNYV HIWSGVSKAVRAKAGV SIMIQKKWKKRITNWTFI	cd01650	536	788

				GATATGGTGGG	CAATAATAAAATAA	NERIITVEMTLFAREVVII			
				TACCGGGGAAA	AATAAA	GVYAPTNDTKDKEKDA			
				TAAAATACCCG		FWDTLRETIEKIPRRKELI			
				GGGTGGACCAA		IMGDMNGRVGIRESCKI			
				AACCAGCAACT		VGKHGEAEYNDNGERLI			
				GCTGCCTTGTAG		DICAQFDLKITNTFFKHK			
				TATGATATTGGC		DIHKYTWQQNTKELRSII			
				TTATCAAAGGC		DYIIIRQTSSFKAADVRS			
				TAAAGGAAGAA		YRGAQCGSDHYLVKMK			
				AACCTTGAATA		SFWPWKNATNDTSNINK			
				CAAATTACCTG		MNCSEKVQNVHFNIDSL			
				GTCCTCCAGGTT		QDESIRTFFKARMERTL			
				GGGGGTTGTGC		DESFEGSTEEIYEYIKTK			
				AGTGGGCCAGC		VKNVASEVLGIKENNPK			
				TCCTCACTCACA		RAAEWWSEEIETSVKEK			
				TAAAAATATAA		RNAFVQWLNDKSEGTR			
				AATGCTAAAAA		SKYKEKKNEVEKKIRLA			
				ACCTAATAAT		KNEAWERTCANVNSKL			
						GFGRAKEAWSVLKALR			
						QDTKGKSNLQLVTQKE			
						WEEYFKKLLNEDRDEYL			
						EEGTVEENEHHDDEIXIS			
						ESEVLQVLRTGKNGKSP			

						GPGNINMEFLKYGGDKI VKLILQLFNKMJHGDSV PKEMKLGYISTIFKKGDR KICSNYRGICVTNTLMRI FGKIIKNKLEKNFRTQQE QCGFTAGRSCVDHIFTL RQILEKHREKSKNVGLIF IDLEKAYDTVPRKLLWR ALHRANINTSLIKIIEQM YKDNICQVKIGNTLSQK FRTSKGLLQGCPMSPTLF KIYIDICLRTWSQKCNSM GLEIRDGVYLHHLLFAD DQVVIAQDGEDANYMC NQLAIAYKNWGLKINYQ KTEYLTNDPHELRIEGK KIKKVNTFCYLGSILETE GKSDSEINKRISSGRKVI GMLNSVLWSRNVMNRT KKIIYKSIFESAVLYGAE TWTINQKHTKKLQALE MDFWRRSARISRKEKRR NTEVIKRMEIIERIDEVM			

						DRKKLRWYGHVRRMED TRIPKLVLEWQPEGRRR RGRPVTTWIKNVQLTM NRLGAEEEDTQDRHTW RNIVNN			
Penelope	Penelope-55_LMi	.	Locusta migratoria	AAATCACAGTA AAAGTAGCTAA TGCTCTAGAAA ATAAATCTAAC TACATGTATCCA AGACACACAGA ATTTTTGTAATA CAGAACGATCA ACTCTAAAAAA TAAACCACAAG TTAGGCAAATA ATGTTATAGAA ATAAAGCTGAT AA	TAAACAATCCCACT GCCAATCATAGATA ATAGCGCAACAAA ATTATCAAAGATAA AATATCCCAACATA TTTCGTTCAATTTTA AATAAAAGCTTTGC CTCATAGATATATA TTCTCTGTGCAAAT AATGTAAGCATATT ACATCTCTGTAGAG CTGTCAAGTAAATG TCAAAATCCGTAAC TTGATAAAGATCTT GCATGACAGCACA ATATTATAAGTTAC GTGATATATAATAC AACATAACAGAAT	RRHLVLINKQDYIQKVF EFFDKNGIIKIPCDPTNK YIAQTRKIIKSCSSILSDK DKRKYIPMNPKTPQLRG HPKTHKELSPIRPVVDSI NCPTYLISKKLNTLLQK YYTYKKKYTIKNSRTLA ANINKIDIPINGKFVSFDII NLFTNIPVKETINIIELNL LSXNKLPINSTYEIINLLH LVLEQNYFKFEDNFYTQ PDGVAMGCNLASTIADI FLNTLEERFFLHQETNKII YYYRYVDDIIVLIDGNET DINNIHGLINKLHPKIKFT MELETNKTINHLDLTIQN VQNKHRFEIYRKPTTTT NIIHAHSCHPSSHKLAFF	IPR000477	124	263

					GTGGTATGTAGCATTTCTTTTACATCTGTATAAAACAAGTAGGCAGCATAGTATAAATATTATGAAAACTTATGTATTGCAGGCATAACACCTGAAGATGACTATTAAGTCGAAATAAGCTTATMGTGTAACTAATACGAATTAAAATATACAGTGTTGAAATTATATAATTGTTATTATTAAAA	HSMINRLINVPLDDAAKQKEMTILQYIAQQNEYDPNIILKLYNNKTSPRPPTTRTEITLQRETSKLNKPIYISTPYLGKISNKVQQIFRKTNTTFAFQNRHTLKHKFVTHTTKPSKFLNSGIYKISCNSCNKYYIGQTGRNFTIRYKEHTRESEKNPSTFYQHLQSENHTVDTIDNALEILHTVQKGYEMNILEEMEVCLHAIKNTNNLLNDQTDFKHKNYFDNFIHILVHDNG			
RTE	RTE-5_LMi	.	Locusta migratoria	CCCTCGGAATACCACCTTAACGTTGGTGGGGAGGTTTGTGTGTGTCAGTGATCCACTGAGCTATGCTGGCGGGAGCTGTGCTCCCGGTAGGGTCACCCATG	TAGGAGATACAGGAAATAAGTCAAGTCAAG	MNRSRMDKKDLTFSMKSDTRIGMWNVQTMWQAGRLAEISREMKRYNIQGDEKILGLSETRWTGFGEIKTQEGDTLLYSGREEEHRGGVGLLITQKTKKSLIEWYPVSERIVVARFASRVRPITVVQCYAPTEAGSEEEKEQFYGQLHRTLLKAK	cd01650	550	782

				CCAGACAGGTC		KKDIMVVMGDLNAKVG			
				AGTTGGTGAGA		NDNQGVESIMGKHGIGN			
				TTCCAGACAAA		RNDNGERLIEFCSNHRL			
				GAGTATTCCCCC		VIGGTIFPHKNCHKVTW			
				TAGGGGGCTCC		VSPDHRTENQIDHFTISK			
				TGCCTTGGGAG		TWRTSLLDVRNRRGAD			
				TATGTGGGTGG		VGSDHHLVIATLRLKIAS			
				CAACCACATGG		TRNTQQKRRIKYNWDK			
				GCCCTTAGCTA		LKNKEVQQCFSLELSNR			
				AGTCTGGCATT		FQQLREDEEQDIERNWK			
				ACTTCCACTTAC		SIESTYIDTCEKVLGKRE			
				TTGTGTCAGGCT		RGREEWITSETWEAVNQ			
				TCCTCTGCATTC		RKECKKRLNTARTQEEK			
				TTTCCTATCGGC		VQARTEFNELNKMAKR			
				CTCCCTTGGCCA		RFRRDKRAFINNLAAQA			
				ACTCTTGATTTT		EMAAQKNDSKTLYDITK			
				TCCGACCCCGA		RLTGKCFRSEKPVKAKD			
				CGGTATTAGGTT		GKLLTSQKEQLERWREH			
				TCGAGGCCCGA		FLEILNRQTVEEEVTEAA			
				GGGAGTCTTTC		TEEEQEDVDISIEPPSRIEI			
				ATTTTCCTTCTT		QRAIQQLKNGKAAGMD			
				CTACTCTTATTG		NITPEALKVDVSTSVEIL			
				TCAACTGCTGCC		FPLFKKIWEEEKIPTDWK			

				TTGTAGTTAGA		KGLLIKLPKKGDLSECD			
				AGTGGGAACTT		NWRGITLLSVPSKVLSR			
				CAAAGGCTAAG		VILERIKKRVNSRLRKVQ			
				GGAAGAAACCC		AGFREGYSCTDQINTLRI			
				CGACAGAAAAT		ILEQSTELQSSLYLLFVD			
				CAGTAGGCCTG		FEKAFDSVMRKKMWKA			
				GCTACTTAGAA		MQEFGIPKKIIKITQEMY			
				GGCAGCCCCTT		ENYTCQVLHQGMLSEPI			
				CACCAGGCTAT		PVDTGVKQGCLLSPILFL			
				GGGTGCTGAAG		MVLDLVMQKAMNRRR			
				GCTAATAACCT		GIQWGLTGRLEDLDFAD			
				ACACACCCGAG		DLCLLASSYKDMEAML			
				ATTAAACTATTA		NDLKTEASEVGLKINGK			
				CAGAAACATCA		KTKEMRCNNRNEALFFF			
						KPRDRASQAVLLPGKFS			
						HPGGRTIEDIGNRINKAK			
						GAFAMLRPIWKSRELNS			
						YTKLRIFNTNVKSVLLY			
						GCESWKTNKQLISRLQIF			
						VNRSLRNILRIWWPQVIS			
						NEELWRRTHQTPIDIEIR			
						RRKWKWIGHILRRGKD			
						NVARQALEWNPQGSRR			

						RGRPRNTWRRSVVAEG TEKGGKAWTEIKTLAQN RVRWRKFVDALCST			
Vingi	Vingi-20_LMi	.	Locusta migratoria	GGTGGTGCTAA ATGAGTAATAG CAACAAGATGA GACGGAACTGC CCGAAAAGGCC TCCCGGTCAAA GAGTGCAGAGC AAACAACAAGA AGTCCTTTTCAG GATGCACAAGT TCGTAGAATTG GACCATCTATTC GCATACTCCAA CTKAACATAGA AGGAATAAGCA GGTCGAA	TAAACCAGTGGTGA ATGCCCGAAACGA GCCTTTGGGGGATC AATAAATGACATCC ACCATGCATCACCT GAAGGGCTCAACTG GATCACCTGTTGGA CTTAGAATCTAAGA ACTTGTGTAATGTG TACATAATTTAATA TAACACTGTACATA TGATCCTGTACATA TATTGTCCTTGTCA TACATATCTTACAA TGTATACAATTAGA ACTATGTACTTGAA CCAAGCTCTGTATC ATCATTCGATAAAT AATAA	MRLSIQIIEKPAVDVIAL QETHTADDQQIKSRGKIP GYKLIASNNHSKCGLAI YVKHTINSLKQPFKMWS SNLRQAYIKNYQVVSVS SENNIFTISAKIDSATIVN VYKPPRVNXPTPPLPSQ QHPTLFLGDFNSHHNLW GYCENDSNGEALCEWM ETENLKVIFDAKDPPTFH SGCWNRGYTPDLCFTTQ LHSSHTQIRRSALRISLT VSTDLSFSRDWNMCTPC AVTPETKVEFQESRKSN VDKATDLLQSLDENRKK IWHETVESFDHTHSSRK AWSIIRKLDATNQTPVRP KTAIKLNDFAKHLISNSK SPKDKQTRREIKTQLNT KKRAAPNSSKFSNPFTEE	cd01650	461	715

						EVRDAIRNMKVGKALD SMGSILNFSHTVDQQLS NGFLTSSQTFSPLEIFHH YSRKPSRGSKRCHQKHE SWKGAGFDGIYPEFLTH CGPATIKWLSNFFSDILS TGNLPPLFKKTKLIAILK PNKDPTKVESYRPIALLS VTYKLLERLIYNRIHXTI NKXIPIEQAGFRPGRSCN DQVFALTTHIESGYEKK FVSIAAFLDLSAAYDTV WREGLLLKFVSIIQCRKL TNLLSNMLSNRLFQIYSD NDRSKVFKLNDGLPQGS VLAPLLFNLYIYDLPATS SRKFIYADDICLVTQSPS FTTAESTLTADMEILDTY FRKWCLRLNLQKTVTSA FHLRNRHSNYKPEVKLR GQTLQYCSTPKYLGVTL DRSLTFRQHLSNVAAKV KTRNNIIQKLAGTSWGC			

						SAQTLRSTALALSYSVA EYCCPTWMNSVHSRKID VQLNQTMRLITGCIHST QTAGYQFHPTSSSTRSD ALLKIWRNIQKNPQLPIH SDTSTAEHQRLKSRKPS WKTAHHLDVTNFDINEA WKSHWEESSVSTVIWLK TLQCEFQVSHSPDASGES STASEQDKGTADIYSTN GAGQHLPSVTVELPYKQ			
Nimb	Nimb-22_LMi	.	Locusta migratoria	CCCCTCAGGGG CTCAGCATTAAT TTGCTGAGTAC GGGCTTGGCGA ACCCGGGGTCC TGAGCCGTGGG CTGTGTCGCGCC AACAGTCCGCT ACACCATAAGC TCTGTGCGTGCT TCAACGACCAT CGTTCGGCGCG	TGAAGATTGGTCTC CAAGACAGGGATTT CAGACTGCAACTGA CCCGTAAGGTTCTG GGATTTGGACCTCG GAATGGGCTGCCTT GAACGAACCGAAT AAGCTGCGCGTAGT CAAGCCCCCACAGC CGTGTGGAGATCCT CCCTGCGAGCCTCT CGTAGGGAGTCTGT	GQVWRSGIFFENASGAV LIQTASPAQSRALLACD KLGDVPVSITPHKSLNM VQGLIYHRDLLSQTDEE LRANLERRGVHFVRRVY RGPKDQRVATGAFILAF EGDTLPEKVKVMIYRCD VKPYVPPPMRCFRCWKF GHMSSRCTSSPNCRECG RPSHPDAPCTPPPVCVN CGQHHPPCSPNCGVLHR ERKIMEIKTLDRLTIQRH	cd01650	993	1246

				ACGGTGGAACA	AGTCCTCTGCAGAC	VKSMSGSIRAAPSGKDD			
				TTGTACGTTTCA	TACGTATTGGMCAT	NIFCRSHSGTSFTAINGQ			
				GAGCCCGGGGA	ACGTGGTTGACACA	CSLLCGPYGPTTSTAPSM			
				TCTTGGCTTAAT	CGGCCACCTTCTAC	GGGTSSSVAPTTPPVGAI			
				CGCCCGGGTCG	GCGAGAGGACCCA	TPSPSGTSVPAPSRRRNS			
				CGAGGACGGTA	CCAAATTGCCGTTG	LLRLLSLGRGPWDTPFR			
				CAAACTCTATA	TGGTGCACTATTGA	GFSQCTGDPSQRLKGPR			
				AAAAAAACCTC	CGTGGTCCACATCC	PAGRRASRSSSNPDATTE			
				AATCTCAAGGT	TGGTGGACTGCCCA	KPSQASKPKGQREGRQK			
				GTGCTACGTGC	CTTTTAGCCGACTT	KSKKKETLVATKPQAPS			
				CGAGGAGATGC	GAGGCAGTCTTTTA	SSASEDEVEILASPEDLAI			
				ATGGCTGTTGA	CCCTGCCTTCTGAG	PGSPESRIADTGDSEQLA			
				GGTTGAACAGC	CTTCCATTGATTTT	ADVMDEPVLLYGFSARR			
				CGCTAGCGAGC	ATCTGACGATGCCT	PVLLQWNCGGFFNRLAE			
				AACCTCTGAGG	CTATGGTTGACCGT	LQQLLCRFPVLCLALQE			
				AACCTGCCGCA	GTTTTAGGTTTTATT	TWFPAQRTPALRGYRGY			
				CCTCAGTTGTAT	CGCGCAGCTGGCTT	YKNRAAYDRASGGVCI			
				AAGGCTTACCC	TTATTGGTCTATTT	YVLATVPSAAVPLHTTL			
				AGGCAAGCGGG	AGTGGTGGGGCCTA	EAVAVRIRAMQEFTICSI			
				GCTCTGCCTGG	TTTGCCCGCCTTCT	YLPPDEAVSIADIAALIA			
				GTGGACCTTTA	CGGGCTCCCACCCC	QLPPPFLLLGDFNAHHPL			
				GTTCCCTAGCTG	TCCAGAACAATAGG	WGGTQTTGRGRDVEAL			
				CTCGTGGGACC	CCGTTGTCCTTTTA	LSQRDVCLLNTGAPTHF			

				AATGAAACGAG	CGGTTTATGACTTG	CVARGTYSAIDLSVCSPS			
				CAAATAAAGCG	TTTTTCAATTAAGG	LLPHFRWKVHDDLCGS			
				TGGCCGGCGCA	ATTTTAGTGTGTGC	DHFPILVSLLHFQPAGHQ			
				AAACGGCGTCT	AGAGTGGCTGGCTC	PRWSLRRADWQRYASA			
				CTAAGCGTGGC	ATCCTGTATTATCC	AVVPVPLPADVDMAVH			
				AAAACGAGTTC	TTGTGATCAGCCAG	QVTSSIVSAAEATVPTTS			
				AATGCGTCACT	CCACACTCTTCTGC	GSPRRKSVPWWTPEIAE			
				ATGACCCTAAG	TGTCCTATTTTAGTT	AIRDRRRALQRHRRHPT			
				ACGTTTCCTTCC	TTCTGTCTATTTGGT	MENLIAFKRLRARARQL			
				CTGCCACCCCGT	CAACTCCTGTCTGT	IRQSKRDCWEQYVANIG			
				GGGAGGAACGA	CCTCTGTTATATGT	SRASSSQVWSQVRRLFG			
				AGGGCCCTTAA	CGTCATTTTACCAT	HRPPAAIPSINIQGTVCT			
				TAGATCAGAGC	TGACTGTCTCTCTG	DPTAIAEHMADHYACNS			
				CATACTCGCCA	TACCCGGAGTATAC	ATAAYSAAFRALKVRVE			
				CAATACCTTGTA	TGGACGTGGAGTAC	RRHLSFHTRRDETYNAP			
				TGTGCGCGTTCA	TATGCGACAAAAG	FSQWEFQTALAACPDTS			
				GATGGTGAATC	GGAAACTTTTAAAA	PGPDRIHSQFLRHLSAES			
				CTTCATGACCAC	GTGAACTCTTAGCT	QHQLLALFNHLWAEGA			
				TAAGCCCAGTT	GGGAACAGGGGAC	FPSQWRESIVIPRNVQCA			
				CTTTGTTGAACA	TGATAACCCAGCAG	IQSVGIPNCPCCMPRYLP			
				GCTGA	TTTAGTCCCCTTTA	GTRPDPQPVPSPPFSRVP			
					AACACCCAACAAC	TPASRPVQPSMGRGGFPI			
					AACAACAAC	TMAGKHRYPVMKPGAD			

						PQVADSYRPISLTNVLCK			
						LLERMVSRRLMWILEHR			
						GLLSPRQGAFRQGRSTV			
						DNLVALESAIRTAFSRRE			
						HLVAVFFDLRKAYDTT			
						WRHHILATLHGWGLRG			
						PLPLFIQNFLSLRSFRVR			
						VGTSLSSPRIQENGVPQG			
						SVLSVPLFLIAIDGLAAA			
						VGSSVSSSLYADDFCLY			
						FSSTGMGVAERRLQGAI			
						TKAQSWAMTHGFQFSA			
						SKTRVMHFCRRRAVHP			
						DPELYLGXDLLRVEQTH			
						RFLGMQFDARLTWVPH			
						LQQLRKKCWRHLNVLR			
						CLSHTAWGAARTSLLQL			
						YKALIQSRLDYGSLVYD			
						SAAPSALRVLDPITIVAS			
						VSWGSPPSESAGSSYWL			
						LVARIHASPDHPNHRLLF			
						PAAVLPSPRRRPRSGLPI			
						AVRVRSLLSSLASLPIPPT			

						FHYLSRTPPWFVPRLQL RLDLSSDSKASVSPMVL CRQFQALIDSFHGAEVV YTDGSLVDGRAGYAYV HGDHVEQHSLPCGTSVF TAELVAISRALTHFRDCP GESIVICSDSLSGLLALD QFMPKHHLVRSIQDAVF TLGESGRSVVFVWTPGH MAFRETNESITWPSRPPH HGP			
Daphne	Daphne-12_LMi	.	Locusta migratoria	CAAGATGGCGC CAGGTGGAAAG TACTCACGTAA CGACGAGATTT CATAGTGACGT GTGTAAAATCT GCGAAAAAAAG TTAAAAGAGGT ATCGTATGTGTT GACTGCAAGTC TTGGTATCATGA TAAATGTGTAA	TAAGATGCAAATGT ATGTTGTCAATTAT CATCAGCAGTGATA TGAACATGTAAATT TACTAAATATAATA TTTTATTTGTCTGTA AGTGTAGATGATGT ACTTAAGATTTTAT GTTTGTATTGTATA CAAATATGATGACA ACATCCATACAATT TTTATTGTATACGG	MLIKLLQKDDHTTRQPI AIQSVSTKDTQKCQNNP KRRKLCQLSLKIRLHHG WKKNLHQYQPQQIPLK KTRPNRRNERRQGPTEE TSSEVSDKIMTPFLKILK KKKRFKAVLFSVPSKHP VHKNKVQQLEVELQTL NSSLVCITEHWCKGHEL THVALKSYVLASHYSRT TLQCGGSCIYVKEGIVFK VRDDIIALSEDKHIEISAI	cd01650	634	915

				AAGTAAATCTA	ATGGATAAATAAAT	EISKLNVLKRLTVLCIYH			
				AATACATAAAA	AAATAAA	APSGDMDTFIAKLNTSIG			
				GAGATCATGAC		HGLMYSKDSLYYVFTM			
				TGGCATGCCGG		LPVVTWILLLQSLTQAL			
				CAGTGCATTGT		DMVSKPKGRILICGDLNI			
				GAATGCCACGA		NTLNADGTCITFLNTLHS			
				AAAGTGAACTA		YNISPLINTATRITKNSAS			
				GTGAATACACT		CIDHMLTDIEKDMCNVV			
				AAAGTCAAAAG		VDHLGISDHSSQFLKVNI			
				AAGAAATAATA		NVETNTKLLKYIRVFSES			
				CGTGTGCTGAC		KRSVFASQLASESWAEV			
				AGACGATTTAG		YRESTVNGKFSKFLSIFK			
				TGGCTATAAAT		FKSEEVFPKVLRVIKAQ			
				GCAAAACACCA		NKNNWVTNGIRQSSKTL			
				AGAACTTCTGA		RHLSTIKNYNNTIQNFYN			
				ATAAATATCAA		TYKKVYRKVLLAAKRY			
				GAACTAGTAAC		QNDKLIEKANNKIKAAW			
				GAAACACAAAG		DVTNRVTNRKKAGQINI			
				AAGTGGAGCAA		QFQLNGTSVNDPKKLAN			
				AATCATTGCCA		FANKYFSNVATKLQQKF			
				AGAACAAGAAA		VNNTQHPHTLNNVPHT			
				ATATTATTGTTA		MFLQPSTAEEVSKVVQN			
				ATGGACGTTCT		LKSKSSAGIDEIPVCLLK			

				GTGCGACTAAG		DCTQSISHPLAHIINESFT			
				GTGGCATGTAC		KGCFPETLKHAKVLPIH			
				CAAATAAAAAM		KKGDPESIESIENYRPISL			
				TGCAAACAAAT		LSSFSKVMETIMKNRLM			
				GATAAGTGCAG		AYLTKYNLLCKEQSGFR			
				CTCCGGCGTTA		SGKSTSSAIAQFTNAVLE			
				ATGCTTTCAAAT		ALDKGDHVTGIFLDLTK			
				AAATTGCCGGT		AFDTVDHTILLNKLEAL			
				ATAGAAGATTT		GVRGVANMWFHSYLQN			
				TGGCTCTGAAA		RVQQVEITKTNSNHTIK			
				CAGAAGTCGCG		HLSEPEYINIGVPQGSVL			
				AACGATCTCCA		GPILFLIYINDFPQSVKHG			
				GTGACAAATTA		KSVLFADDSNILITDNSQ			
				CCATCAGAATG		IKLNQKAEDTLVHVSRW			
				CAGTACAAGAA		VENNKLTLNVKKTTAST			
				ATCGAATTCTAT		STQINPQCKKNNSINFHI			
				TGCAAAACGTG		KRKPVTNFLKIKDENIEC			
				AAAAAACGGGT		VSTTKFLGMHIDSQLKW			
				TCTGATACTGGT		TEHITVLGKKIASACYAL			
				GACAGCCACAC		RIINSVCSSSCTRTAYFA			
				CAGGAAAATTT		NIHSLISYGIIFWGTNEQ			
				CCGAGAAAATA		NRQAIFKLQKRAVRIMT			
				AATGAATATAA		NSSSRAHCTELFKYMGI			

				CACACTTTCTAT		LTVPCEYILQQSSTLEGP			
				GGCCACAGGAA		LYRIIQIYGNTNCSMRIH			
				TAACAAAGCCG		FTAVIYIKENINQYPPNSS			
				AGTGCGCCTCT		IHEHETRSSRHLHLNRK			
				ATGTGAGATAC		NTTKTQNSILYNGIKLYN			
				TGAGAAACAGT		RLPQDIKNKDKIQDFRNS			
				AACTCAGTCTGT		LKKFLEKCYYTVQDYLN			
				GACTCACTAAT					
				AATGATTTTGTC					
				ATTATCATGGG					
				AGGCAGAATGA					
				TGTCTATCGAA					
				ATGAGAGTAAA					
				AATKCCACACG					
				GTGTTTAAAATC					
				TACACTGCGTG					
				ATCTTAAAGTG					
				CCAAAATCATA					
				GTTACTAGCAT					
				GCCTCACAGAT					
				ATGACCTGCAA					
				GAAAACTCAAT					
				TGTAAATATGG					

				AACTTGCAAAA ATAAACAGACA ATTCCACAAAA TGTGCAAGTCCT TCCAAAACKTA ACGTTCTGTGA GCTTCCAAATTC AAGAGATCTAT TCACAAAACAT GGACAACACTA SAACACGCTGG GAAGTCATACA TTGGTAAA				
Nimb	Nimb-26_LMi	.	Locusta migratoria	CCCCTCAGGGG CTCAGCATTAAT TTGCTGAGTAC GGGCTTGGCGA ACCCGGGGTCC TGAGCTGGGGA CTGGTAAGCGC CACCAGTCCTCT AGACACCGTAA CCTCTGGGCAC	TGAAGTGACCTGAC GCGGCTCCGAATCG GGCACTGCCCATTA ACGCACAAGTTTTT ATTGTCAAGAAGAT CCTCCTTTTTGTGAT GCCTGTGGCGTGCA ACTTACATAGCCAC ATTTTAATGAGTGC ATTTATATCGTGCT	MGVPLGGTNTHSKKRA RLGSPPDTGLCNMSMLK SNRAHSAVQNVFLVIKR KEGSFEKISPFYIQKGLE GIAGDLKSVKRLRNGTL LVETSSSQQAVNLRNAK CLGEYEIETEMHSTLNYS KGVVSCRDLVDIPNEEL KEEWASEGIVDVQNITK RVDGGLVKSESFILTFNS	cd01650	1150	1451

				GCTTCAGCGAC	GCGAGGGCAGAAG	PKLPEYVKAGFLRLKVR			
				CACCGTGTGGC	CCAATCTTAGTGGT	PYVPNPMRCFKCQRFGH			
				GCGGCGGTGGA	GATCTGCCCTCCAT	TTLACKGEATCGKCGK			
				ACGTTGTGTGTA	TTTAGCTGATGACG	AAHEGAECTPPAKCVNC			
				CCAGGGAATGG	AGGCGAGTGTGAA	SGEHPVWSRDCTVYLEE			
				GGATCTTGGCTT	AAGGTTTTAAGATT	KKVQEIKTTKRISYGEA			
				GACCACCCGGG	TTGTGAAGCGTCTG	KKIFKAMQPPSFVTSFAA			
				TCGCGAGGATG	CACTCTGGCCTAAG	VVQKPVPKVDVSTQTD			
				GATAAACCTCT	CTTTTAGGTTGGAG	VASVGTNTCTCECTCKA			
				ATAAAAAAAAC	ATTTTAGTTTGATG	AVVXEPELPHAHQKKLL			
				CTTCAATCCTAA	CGAGTGGCTGGCTG	XSTLQSFCHPPRQRPLAR			
				GGTGTGCTACG	AGTTCCTCCTTGTG	PWHLQSQRLFRRSSLTFP			
				TGCTGATGGGA	TGTCAGCCAGCCGT	LLLSXTTATRTSEKATG			
				TGCATGGCTGTT	GCTAATCTACCTAG	VNITELLSPSPAEAIGQA			
				GAGGTGGAACA	TCCTCTCTTGGTAA	LAPSKPEAIPKVITDLSTP			
				GTCGCTAGCGA	CTGTGATATGAATG	AVTNIKVAKPKKAPGKS			
				GCAACCTCTGG	TTTTTATGGCCTCA	QEQAKAGNKPSDYIEVII			
				GGATACTGCCG	TTTAACTAGCTTGT	PDLPSDSAPELMELDVP			
				CACCTCAGTTGT	CTCTTATCTTGTTTG	QGQTSRPKTEPPARSVSP			
				ATAAGGCTTAC	ACACTCGTCTCACT	TKRKERVKIQPPRWLPY			
				CTAGGCACGTG	CTGTTTTCGTCCGG	CSGTAGFRTHVEELRLL			
				GGGCTCTGTCTA	GTGCTAAAGACCTC	TRDRPMCVCLQETCFKE			
				GGTGGACATTT	GACGTCTGCGCCCT	TDSPELRGYFLYRKDDL			

				ACTTCCCTAGCT GCTCGTGGGAC CAGCATGAATA TTTCTTTGTTTT TCCCCTCCCAGT GGG	GATCCATAATCACC ACCACACCACCA	TGVRAKGGVAIFVNATY HSSPVSLTTGLQAVAVS VYLRHRLTICSLYLPPHD ALDEDALRDLLTQLPQP FILCGDFNAHNVLWGSA NTCPRGRAIERILLSSSA HLLNTGQSTHFCAATGS FSAIDLSLSSPALAHSAQ WEVAMTYMAVTTSPSG FICQTQWYPPEPHHDGC FVRQTGPCSANWLNLN NVIASRNGWTILHXSLA LFSSPCPLCSMGSGDDL HGSDHFPLWIHLPDTVV PARTPPRWVLRQADWT LFSQLAEFEQRDSVQEW VDHITQXIHRAAEASIPQ STGQLQRHPVPWWTEE CRSAIRARRAALRKFKC RPTAENLAAFRLARAKC RKVIRESKKRSWQHFLN SVNRSTPSTKVWETIRRI SGRGGRCPVAAVMANG			

						TLHTDPRDIAQTMAEYF ATVTSTAQQDPAFQRHR MGAEKGSLDFRSTTAEE YNCPFSMWELDSACLQL VIRALDMTGYITVCYST LQNGTEKSSSHFLMSFG RQVIFPTHGVKQFEGQF GLQVHYCRRIQLPFLHV GARFCLSAARDTCPGHD RIHYSMLQHLTERNREIL LSLFNVVWASGHFPDSW REAVLVPILKPGKDRTSP SSYRSIALTSCVGKTLER MVNRRLVWMLESRQLL SRFQCGFRKYRSTFDNL AFLEAAIQQAFLRRQHL LGIFFDIERAYDTTWRHL ILRQLHEWGFRGAFPFLL CPFFLHAILGIESVTCSRI VLSKRTVFLRVAFMGFS WSLPLFIVSFLSPRYFRY RVGDVLSDRFEQENGVP QGSVLSVTLFAIAINSLT			

						SVVKSPVQCSLFVDDFSI FCSSSSLATTTRQLQLTL RRLDDWAKKSGFKFSTE KSVCVLFNRSRSILNLPE LRMRDTVLHFKETVRFL GLIFDSKLTWLPHLKDL KKRCLKALSILKCLSHNS WGADRTCLLRFYRAFV RSRLDYGCTVYGSARPS YLKMLDVVHHEGIRLAT GAFRTSPIPSLCAEAAEP PLSIRQLLLGRQGYKML STPHAPAYHSVARPPMA GCFILGRGQHDPLGFARI VSHRWVYLILMFTSGLE QISSLAPSEAQHYFRLGS FKKGCTPDFTYRLNFSK FIGITILRWSTPMVQAGD SLGCSIVFPTMLPGFASH SLFCLHSRASRDLSIRNA ACSEHTVSYLLRFTRTAI TDSLYPDEKIHSSYRTNC TPSSGGVSRCHSAGSWH			

						VGIRGNEVADGLQRRPA ATTTYTICHSPAGCPLCG SAEVLSAVGRHVAAVT GNKLRSVKPTVRRGNLL PVTDA			
Vingi	Vingi-14_LMi	.	Locusta migratoria	GAGAAGCCTCC CCGCGGAACAT ATTAATACGTTT TGGCGTCTCCCG GGACGGCGGCA AATTCCTCAAA CCAGAACATCA ACACAGTCATT AAAATGCCTCA TAAAGAACCCA AAAGCTCTTTTC AGAACCA	TAACATTACAATTA CCTATTGTACATGA CATTTATCCTACAT TTTAGCTTATAAAT TTTATTACTCTCTG GACATGGCCAGCTT TAGCCGAAAAGTCT TAGCCCTACAAAAT ACATGTTCATGTAA TATCTATGAAGAGC CTAACAAACCTATC TATCCGATGTCTAA TGTACGTATAAAGT TTGATGCTTCTGAC ACGATTAAATAAAA TAAA	MLTIISVNIEGLSSSKENL LHEMCQSNKCDVLCLQ ETHRDNNHRRPNIPGMK LVAERPHAQYGSAVFTK PDLQVKSTCHTEENNIET ITVELKDYTVTSVYKPPS EIFAFTPPANFVKEKISFV IGDFNSHSHVWGYASED KSGEAVLAWAESSQLSL IHDNKLPPSFYSRRWHR GYNPDLLFVSENIAIQSC KSVCPPIPNTQHRPIMCQ LFPAIRPQNVPFRRRYNF KKANWDRFSGLLDEKIQ TINPAPEEYDLFVDTIRT CSKASIPRGYRSNYVKG LDSEAATNLDRYYTLFE RDPFGEETLAQANLVTS	cd01650	461	717

						SISEAKKKEWVKQMEEL DMGRSSNKAWRLLKRL ATDQLNLKHANVKSNQI AHQLLKNGKPVRRNRN VQNSRQIERDISQETSEIS KPFSMTELDKAISKSRN GKAAGLDDICVEQIKNF GPAAKKWLLAXMNNC MSACKIPKIWRKTRVIAI PSETWEEPDDXXSYRPIS LLCHLYKTLERMILNRL TEVVEPLLIPEQAGFRTS KSCTAQILKLVQHIEDGF EERKITGAVFIDLTAAYD TVNHRLLLQKFYTSCKD YGMTQMIGNLLQNRRFF VDFQGQRSRWRTQKNG LPQGSVLAPLLFNIYTND QPLPEGTQSFIYADDRAI TAQGNTFETVESKLSEA LEQLAIYYKENQLTPNP RKTQTCAFHLRNRQATK KLNVSWDGIPLVHCATP			

						KYLGVTLDRTLTYRDHC MNTKQKVAARNNIVRK LAGTTWGAEPSTLRTTT LALCYSAAEYACPVWY NSSHAKQVDIPLNESCRI ITGCLKPTPTDKLHCLAG IAPPSIRRKVAASKEKLK SVTSAAHPLFGQQPARR RLISRRSFLHTTECLTDT PQNTRTEMWRNKNVHL HDWLVPTEELPPGHGEK RNVWRSLNRLRSGVAR TKSNMHRWGLSDESVL CDCGQEQTTSHLMQCEL CPSTCTVEELMKATPNA LDVASFWSKIS			
Vingi	Vingi-4_LMi	.	Locusta migratoria	GGGGGTCACAG CAGAAGCGTTC CCGTTCGCGAC CCCTGTCGAGTC TGATACCAGTCT GGTATCAGTCC GCGAATTCTGCT	TGATTCAATTGTAC TATATGTTGTATTT GTCTATCTCTGTTTT TATAGTTTTAAATC TGTACTTTATATTTT ATCTCTGTATTTGA CTACTGCTTTCTGT	MRLVNEIAHEKHGSAIF VRQGLQVSGSRTIQHNN IELIGVELDNTEIISIYKPP AQTFNLNIPLLSHTRDN MHTVVLGDFNCHSTQW GYSETNEDGHLLEQWA ENNQMCLIHDPKQPPSF	cd01650	412	667

				ACAAGACGAGC	AAAAGCACTTACTT	NSGRWRKGYNPDLIFVS			
				TACATCGGATC	TGTAGTTTAAATTA	KDLSALSYKQVLNAIPR			
				AATATGCCGAA	TGTGCTGGCTTCTT	SQHRPIKLEIRPVVQPVEI			
				ACAAAAAGATA	TCAGTTACATTTGT	PFRRRFNFKRAKWEDFG			
				GCCAAAAGAGT	GATGTAGTTCGTAC	NKLDESLKLIEPKPEHYD			
				GGACTCGTTTTA	ACGAATAA	KFVQAIKDASFKTIPRGC			
				GAGCCCACCTG		RTEYIPGFTDKNKDAYE			
				ACAATTATTTCG		RYCDLYDKSPLAEETIEC			
				CTAACATCGAA		GEELLEALSDQRRQTW			
				GGAATTTCAGT		MKTVVGMDMTHSSRKA			
				ATCCAAACAAG		WRLIKKLSNDPTIRTPHV			
				ATTTAATTGCAG		NITPNQIASQLLLNGKVN			
				AAATCTGCAGA		GKNNAARSSLKVKKSKE			
				ACCAACAGCTG		SNVSDVDFSVEELEEAIS			
				CGACGTCCTGT		SLKVGKAAGLDDIRNEII			
				GCCTACAGGAG		KNLSENXKKWLLKFFNC			
				ACCCACAGGGG		IWRTAKIPRIWRHAKVIA			
				CCTCTCAAAAT		VLKPGKDPNDPKSYRPIS			
				AAGCCAAACAT		LLCSLLKLFERLIMNRIG			
				TGTAGGT		DYIDSCLIDEQAGFRPGK			
						SCTAQALNLTQFIEDGFE			
						RGLPTGVVFIDLTAAYD			
						TVNLRTLISKVHRMTNS			

						DRLAEVLRALLFNRRFF VDFQQKRSRWRNQKNG LPQGSVLSPLLFNIYTND QPIGKKTKSFIYADDRAI ACQAKTSNEIEAILTEDL TTLTEYYRRNHLKPNPS KTQVCLFHLNNREAGK VLDIIWNDVRLNNCSHP TYLGITLDRTLTYKKHIE KTKAKVQTRNGILNKLT SAKWGANAKVLRSTAL ATCFSVAEYGCPVWERS HHTKKLDVVLNNTCRKI TGCLKPTRVDYLYSLSGI APPNIRRSFLSSIERNKQT NDHRHLLYKNQPVPKRL QSRRSFMYSVEPTVASIK SLWQEQYPEALITEEALP QGSHLDWQHWKTLNRL RVGVGRSKQLMKRWGY TDSDECECGAIQTMEHL LKCPLVDASRECTQEDL			

						LKYTGTAAAYTDFWTS KI			
CR1	CR1-3B_LMi	.	Locusta migratoria	TTCTCGGTATTG TAATCATGAAT AGGGACTGTGA CTGCTGTGTGCG GATGCAGTCTG AGCTGGTGACA CTTCGCTCCCAG CTTCAAGCTGG AGGCTGCGGTC GAGGGGCGTCG TTGTTGTTGGCC GGTCGTGGGAC AACGGATCCGT TCCACCGATCG GCCTGCGGCAG CGACATGCCCT GATACTGCCAG CAATGAGGCAG ACACCCCACCG TGGTTGCGTGG GAGGTGTCCAG	TAAAAGTGGTTCGA TGAACCCTCTGCCA TGCACTTCAATGTG ACTTGCAGAGTATC TATGTAGATGTAGA TGTAG	MEPLREMAAKEGKKAN VHSVCIPGGVIPDVERVL PDAMKGTGCSQLQVVA HVGTNDVCRFGSEEILS GFERLSEVVKTASLACK MKAELTICSIVDRTDCGP LVQSRVEGLNQRLRRFC DRVGCRFLDLRHRVVGF QVPLNRSGVHYTQEAAT RVAGAVWRGLGGFFFF RELCGVDWAVFFFFRLE GLGNTQKGFQSQRVQV KDRTNTDLKTIGLTVVN CRSCVGKVPELQALIEST ESQIIIGTESWLKPEISSA EIFTKGQMVFRKDRLNK IGGGVFIAVTGGLSCSEI EVDCSCELVWVEVIPDN QNKLIIGSFYRPPDSDDSI AEMFKESLDQISKTYPK HTILIGGDFNLPSICWQK	cd01650	772	1022

				TGCGTGGCAGG		YMFKTGGKHKTSSEIVL			
				CGGCGAAGACA		NAFSENYLEQLVHEPTRI			
				TTCCGGGGGAC		VNGCENILDLLATNNPE			
				CGATCGTGAGG		LIKSIKTDTGISEHRVVV			
				CCTCCCCGGTTC		ARLNIGLSKSTKIKRKYI			
				GTCTTGCGAAC		YSKKQIKIRLTLSKKIYL			
				AGGTTTGAGGC		FKKADKNSINAFLSENL			
				GCTATCTGTTCG		QSFETNNVSVDQMWLT			
				GACAATGTTCCT		FKEIISTTIDRFIPNKLTK			
				GAGCCGGATGG		RGLTRRGTQNRLDHCY			
				TGTCGCTCATCC		KSKEKSMPNLNAQKIPK			
				TGAAGTCACCA		TGNLLQKLEIQIHTKQIN			
				CGCAGCCTGCA		KEGADPPWYTKQVRSLL			
				GATCCGGGCTTT		QKQRKKHAKFKRTKNP			
				CACAGAGGGTG		QDWQSFTETRNLARTSM			
				GGTTTGCTTGTA		REAYNSYHNETLSRNLV			
				GTTGGGAGCTC		ENPKKFWSYVKYTSGKT			
				CAACGTTAGGC		QSIPSLHDRNGDATDDSS			
				GGGTA		AKAELLNTAFRNSFTRE			
						DEANTPEFVSRTAASMR			
						NLEVDTPGVLKQLKSLN			
						KSKSPSPDCIPIRFLSEYA			
						DAIAPYLTIIYNRSLSERS			

						VPKDWKVAQVTPIFKKG NRSDPQNYRPISLTSVCS RILEHILSSNIMNYLEEN DLLSHSQHGFRKYRSCE TQLALYSHDVMSAIDKG CQIDSVFLDFRKAFDTVP HKRLLIKLRAYGISPQLC DWIRDFLSERTQYVVID GKSSSRTEVISGVPQGSV IGPLLFLIFINDLGDNLSS SLRLFADDAVIYRLVKSP EDQKQLQNDLDKISAW CKNWQLTLNNDKCEVI HMTTKRNPLNFSYTINQ SNLQAVNSTKYLGVTIT NNLNWKEHIENIVGKAN QRLRFIGRTLRRCNSSTK ETAYTTLVRPILEYCCAL WDPHQVGLTDDLEKIQR RAARFVLSRYRGESVTE MMHELGWKSLKERRYL LRRDLFAKFLSPTFSSEC EHILLAPTYIGRHDHHN			

						KIREIRSKTERFRCSFFPR AIREWNYREIS			
Penelope	Penelope-59_LMi	.	Locusta migratoria	CTTGATGAAGA AAGACGGATAT ATAAATAAAAC GCTTGACTTCTT TAAMTGCAAAT AACATTACACG TATAGACAAAG ACCCTACAAAC CGATTTCAAGCT TTGATAAAAAG AGCTATAAAAA TCTGAGTTTCTT TTAACACAAAA AGAACAAAACA GATATATAGTT GTGAATCCACA ATGCCCCAAGC TCAGAGCACAA CCAAAGGTCCA TAA	TAAAGTACAATCCA TCCTACCGTAATAT ACTTCCAAATATGC AACAAAACGCGAT AAATTATAGTAGCA CCGCAATATTATCC CTATATTACAATAT AGACACAAACAATT TATTACACTTAAAA TAGTTCGGAATGAA ATTAAAATTTATTG TACGATATTTGTAT GGTACTGTTTATGT CACACTGCATATAA CCAACGATATTGTT ACAAACAAAATTA AGCAAAGAATATTT AATAAAAACATCTT TGGATCAACAGCTC TGTAAATTACACCA AAGTAATTACAACA	NKCSDSTNCESDKQPSN LIEKKLNKIICENYTYSR QYTVKSSTDLIAKIQDIKI PDNSRFLSLDVTNLYTN VPIDETIAIIKENLLKYRK IPKQSLNNSTITIILSHNY FSFNGHVYTQNDGLAM GSCLSGTLANIFLDHLEN KFFEEQNQWKTKVVYY FRYIDDTIILFNGCIEEIED FHQRFNNMHKKITFTIEH EINNSINFLDLTISKTNSV HTFNIFRKPTTTDIIIHNS SCHPQSHKEAYFRSAIDR ITKYPLCNKNINSELNTL KDIAISNGYNPELVNNL YKSKLNSKQKSIKDEITL TKTETRKKDKFVILPYW GQISDRIKYALKKTNLKI GFYTNNTLGTNIIHKIEKI DRYSQSGVYKIECSKCD	cd00304	123	210

					AACAGCGTGACCTA AACAGTAACCAAAT ATAAATGTATATAT GAATGTAAGTAATT ACAACTTGTACAGT CAATATAACATTGC CATATTAATGTATT TTCATGTAAAACAC ACTTTGTTCTCTGT AATAGGTGGCTTAG AAAATGGGCAAAG CCCGAAACTAGTCG TCAGCATTTCAAAT AAAAGCAGCTTTGG TGCATATATAAAAA TGTCTTTTCTACAA TATTTAAGAGCTGC AAAACGCGACATAT CAAAAACATTT	KAYIGQTGRPFKIRFREH INTSENNVSAFGTHLRES KHSTGSIEENLKILHIVC KGQLMDVLEELQIYTH MTKYPEQMLNEQTEMR NKKFIENFHNTLASI			
Penelope	Penelope-6_LMi	.	Locusta migratoria		TGAGAATGTCCTCC GCAGTCGGGGACG AAACGTTGGGTTTC AATAATAAATCCAC	GKQNCKLLSVHITISNGV CSSDWDRIDRLLHDTIN KRVLTTSSRQRKKFDSL RPQPTISQLDSSRTVINLS	cd00304	354	438

					AAGACCACGGCAT AATAGCCCGGAAAT TTTTAGTAATTTTG ACCATTCCGGCCGT GAAAGTTTACATTT TACA	GHLLSDDERSVLARGGN FAVSPRVVPTEEIIANIEA GIRNLDKVTAEEIRIEAV RVLARAKPPKSNLTTGE KNAIKSLNEDDSIIILPAD KGNCTVVLNVTDYHCKI KELLDPQTYRKLSKDPT DKVLRTVTQLIKKSSIEQ CVQKSLFNSVALPPRLY GLPKVHKEKENVPPLRPI VSAIASPTYLIAKFLTTLL RPHVGRSDSYIKNSRHFI SKLNDILVNPEDILVSFD VVSLFTMVPVNEVLQQL NLLFPPDIADLFKYCLTT TYFKWNDDFYEQTDGV AMGSPLSPVIANFYMEK FEEQALKTANKKPAIWF RYVDDTFVVWRHGREE LDRFLQHLNGINSRIQFT MEKEADGKLNFLDVSVF KKGNGRLGHTVYRKPT HTDRYLHRDSNHHPQQ			

						KRGIIKTLADRARNICEP ELLDAELKHLSNALLKN GYSSTEIRRALRPTRRVE NDERVPVKSKVFLPFIKD VTDRIGKILTKRNVAVIY KPTRKIQDHLKSAKDAR KPLEKAGIYRIPCSCGDV YVGTTKRTVKKRLDEH KGNCRRGEIDRSAVAEH VFQPGDHLIQFDKTQVL AATNGYHERLYREAIEI VKHPNNFNRMEEGVKL NEIWIPVLKKKCNSLPP WDDINGRRQSATTNCAR MSKTSDVTPLRASPRKR SLASVSSEPGCASDIQKC HDPP			
Vingi	Vingi-27_LMi	.	Locusta migratoria	GGGGGCGAAGT TGTTCGCGTCCG GAAGCAATCCT CAAAGCTCCAC GTGCCCCCTGG GTGGTTCTAAA	TGACAGAGATATGT TTTATCTTATTAATT AGACACTAGTATAT CTTACTAAAGTATT ATATGTAGCTGACA TTATACCATA	MSATTTIPNGSCQGQQN CQQNDLNRAIGPVLNVC HLNIEGISRSKCDFLSKN LLSNNTDVLALQXTHAE DLKNCARGNIHGYRLVG EIGHRTYGVATYARQDL	cd01650	490	744

				CGGACAACAAT CTCTGGTAGCTG GTCCTGTCACCA CGGTGATCTCCC AGCT		DDVSIIHESCDDDIFVLA VKVSGTTIVNVYKPPNIQ WQDTVLPVFEHPSLYVG DFNSHSSHWGYSSDNEN GIALAEWAENNNLSLVF DAKDSSTFFSARWKRGY SPDLCFCSCDQTLRSLSV TRKVLGTFPRSQHRPVLI SVGIKIPLIRSFPRPRWNF NKADWEAFTSKLEDNIR FIPPTTENYDRFAGLLKA TAKACIPRGYRKTYIPG WSQECEDLYNEFISTEDS NTADLLLNSLDNARREK WVKKVQSLDFTHSSRK AWSLLRKLSTSVSEKKL PPTHPDPSDIAKRIVALS KSKRQQSDRLISYKLNE VKKCQQAPLEVSKPFCL DELHVAISDVKTGKAAG IDDIFPEFIKHAGLKTRL WLLNLFNNILKDSSLPK AFKRSKVIAVLKPGKPA			

						DRADSYRPIALLSVCYK LLERLIYNRIVPILTASIP VEQAGFKQGRNCCDQV LALTSFIETGFEKKCTTS VAFLDLSAAYDTVWKK GLLVKLYNLIPCPILLRL LNSMLSGRLIKIYLGNTE SGFRQLHDGLPQGSVLA PILFNVYTADIPATQSRK FIYADDIAIASQHRDIHV TETTLTSDLTLLRNYFA MWKLKPNPNKTEVACF HLNNRMADYKLNISFDG VQLEHSEHPKYLGVILD RSLTFKYHLHKTADKVK TRNNILQHLAGSSWGAS AKVLRTTALSLVYSAAE YCSPVWLNSAHVRDIDV QLNRAMRLISGTLMATP LQWLPILCNIVPPDLRRK EALMREYYKILSTPELPI LQDLPPQVSRLKSRKPPL KTAVQLHESGFSPKSNW			

						ARTWDSFNGXNSIWIDD PSKELQGMDLPRREWSL TNRFRTGVGRCNYWKF KWGQADDQSCDCGADI QTMTHILTNCPLRSFSGG MTELHJVSQEAVQWLKE LDLAL			
Daphne	Daphne-6_LMi	.	Locusta migratoria	AATATGATAAA CAAATTAGAAA ACACAAATGTA ACAATAGTAAA TCAACCTCGCG ACATGACGCGT AAATAAATTAA TAGAAAAAGTA AACAAAGAAAT TGATGACTATG CAAAAAATTTC GGTATGTAAAC TTGATTGATGTA AGCAAGTTAGA GAGATCTATGC ATACCAGTCAT	TAGACATGCTAAAA ACTACTTAATAAAT GTAATATAGTAATG ACTCATTTTTGTTGT GATGTGCCATATAT ACTATGTGTATATA GAAATCTGTATACT TATGTAAAAAACCT GTAAGCAATATCCT ATATCAATAATTTG ATGAAGTGGATGCT TAAATAAATAAATA AATAAAA	KEQKISKGAQWCPSAIN RPIQDQDRSSGTRSKKSN ATPHPARPQQKGERKTS KPQILEPSADLVVFHQN VQSLTNKIDEFQIMLENE LKKVSVLCLTEHWLQD DTLNMVHFPNFILGSHF CRKKRYASPQHRKQSFV VADLHKSIHLHPSFPEAF QRKIPWKYQDLSYCMK TYLNXECVLHKLSKKNR HIIICGDFNIDFSVGLETK PKTDILNLLLTFNMTPTI QTPTCVTKTSATTIDQIF ASVPMEYKARTLCTGFG DHTAQEITIKMDLPLHC	cd01650	538	802

				GGTCTTCATCTC		KTSLLTITRSFNLQNTNT			
				AATTACAGTGG		FTQYLTNESWEEVYSCR			
				GAAACAGTTCT		DTNEKFEKFLSIFQHYFE			
				TAGCTGAAGAA		LCFPLKKRKIKKSNIKSS			
				ACTGTAATTCA		WITPGILTSLREKRVLHT			
				GTAAAATTTAA		ISKQTKTTPEFDTYFKRY			
				CCACAAGAGGT		KQILNKVIKKSKKLTNFS			
				TTAATCTGGAA		YIEKPQIKCNVDYCERN			
				GCTGTACAAAG		RQNTGHTHSNINLLENG			
				TACTTTGGACA		EEITDPHRVADLFNVYFT			
				AATGGGTAGTA		EIASNLIDKQQKPASENK			
				AGGAAGGACCA		NQQKLFIRKNNFPIPNRR			
				GGTCTCCCAGTT		KGSTEYYKRIKTKIFIRH			
				GTCAAATACGT		KCATVVTKPLTHLCNSS			
				GGCCCCAGTGA		LSNGVFPEKLKLAKVKP			
				CTGAAGAAACT		LFKKGDKRNEKNYRPIS			
				AATGCTGCTGC		LLPGFSKILEKVFYRRLIS			
				AAGACCTGAAG		FIEENNFLSNTQHGFRKQ			
				TCCTGAAGACG		KSTETAIFDFVNEALSTI			
				CCAGCCCAACA		DQKEHVSALFLDLTKAF			
				CCAAAATGGTT		DMIDHNLLLKKLEYVGI			
				TTCTAGACAAA		RGLAHKWVTSYLKNRA			
				ATAA		QIVEIAHQEGSNIYTYQS			

						ERKDIKYGVPQGSILGPV LFLIFINDLETSSSSQKYI KFADDTNILIKGKTAAEL QHEIQNCTSHIEEWFTTN NLVINIEKTVTMHLHTIQ NKHPLTPTIKLFNKKLET VNSTKFLGISIQDDMKW DTHIKYISKKLNTICYGV RILARCTTKETLKNVYY AQAESLLRYGLIFXGNSP ESRYCFVAQKRIIRSIEN ATPRKSCKPLFKALQILP LPCLYILEIVTYIKKISVA QNNLVYTNKSVHTHDT RQTLDLHSEHLRTTLRQ NGPLQMGEKIVQQTSSP YQNNQRTPEVQIGCKKL FTRKIIYKINVITLSVNIC QLNLMSCFLTYNYCNYC HLLYDVDCIII			
Nimb	Nimb-like	.	Locusta migratoria	GGGGCCCACAG CTCTTTTGTGGG TACGTGTGTGG	TAGCTCTTCCTTTTG TTGTATTTTACCCT ATGTTTTTAGCTGT	MGDNFAMTSSSSSSTLP FTQKRKLDESKSKKVLP NLPKFLVVSRTDGHDFS	cd01650	987	1244

	36_L Mi			CGCACGGGGCC CCGAGCCGAAG CAGCCTTCCTTC CCGCCCCAGGC TGCTTCAGTCCC CCCCCCCGGTTT ATGCCTTCGCCC GGAGACGGGTG TCCCGTGTGAGT TTAGGTGGTCTC ATGTACCAGCTT TTTTCTTTTCTC GGTTTCTCTTCT CCTAACCTTTGC TTCCGCCTTTCC CTCCTTACTGTA CTCTTCTAGCCT TCCTCTCCTCCT TTAGTCTCCGCC TCGGCGTTTGA GACTCAACCTTT CCCACTTCCTTC TTCTTTCCTTCT	TTTCATATAGGCTT ATTATCCTTCGGTA TATCTCAGTTTTGA TTCGGGCGCTAATG ACCTCAGCAGTTGT GCGCCCTAAAACCA CAAAAAAAAA	SVNPFVIQRGVDAIAGPV KTCSRQRNGTLLLETEG AFQVQKLLRTTLLHTFP VKVEPHRTLNSSRGVVF SRSLDGLSDEELQEYLSD QGVTAVHRVMKRVNNN LTPTQTNCVLDLHRPKQ TVFLTFDKVDLPSKIKAG YEIISVRPYVPNPTRCYQ CQRFNHTRQSCPNAAKC VTCGRDAHEGDCPPPSP RCINCMGDHAANSRECP VYKDERAIQEIRVKEKV STSAARKLLASRKPAVL PAGTYSNVLSSPRPTKET GTQTCDFTFSDTVVRSA SPKIARSTSPMPSAHSKS QSTSVSQKSRATKSEAR TSKKDPVREDFLKTSSSQ PSAAHSSNRTASKKAHK KRSSSPSPPRRDSSVTPSS SRRPRPSSVSPGRSTACR SAGISPAEEAAPSDQSRK			

				TTCCTCTTTCTT		EAEQMSEPMEHDCPSVD			
				CCTCCCTGTGCG		CSSSAASQPGPQRPSRSP			
				TTCCTGAAGGC		LAFSLFFFSAFFMALLQ			
				CGCCCACGCAC		WNIRGIRSNRDELRLLLR			
				TCGCTGCGTAA		LHCPLIVGLQETKLRPTD			
				CAGGTGACTGG		HINFPQYTAVRFDLPPVE			
				GTAATGCGTAA		GIPAHGGAMLLIREDVY			
				GTCCCAGCCCC		YHPIPLQTPLQAVAVRV			
				GGGTTGGCATG		TISTFTFSLCNIYTPSSSA			
				TAGGGCCCGCA		VLRADLLHLVAQLPTPF			
				CGTACCCCCTG		LLIGDFNAHHPLWGSPA			
				GTAAAGGCCAG		SCQRGSLLADLFNHLNL			
				GCCCAGGGAGG		VCLNTGEPTFLSDTTHT			
				GGTGATTACCC		YSHLDLAICTTPLAHRFE			
				GAGCTGATACC		WYVLSDRYSSDHFPCIIS			
				TTCCGACCGTGC		LLHHTTSPCSTSWNISKA			
				CGATTGGTCCCT		DWGLYSSMAVFHDQNF			
				CCGTCCGTTTCT		PSCDSQVNYLTDVILTA			
				CGGGAGGTGTG		AEHSIPRTSTSPHRVPVP			
				ACCTGAGGTGT		WWNAACSDAIRARRRA			
				GGACAATCACC		LRLFQRRPTLENCINYKR			
				TAAGGCGGGTG		LRAKCRRVIKESKKACW			
				AGCCCTCAGTG		TAFTSSFNRSTPSSVVWG			

				AGGGTTCCCCA TTAGGAAGGAG CGCGCCATCGG AGACGCCGGTA ATC		NLRRLSGTKVHSPIPGLN IANQVLVEPVDICNALG RSFAEVSSSANYDPAFLP RKQAEEARTPSFTSPNRE SYNATFTMQELERALAL SRSSAPGPDAIHIQMLKN LSPAGKCFLLRTYNRVW TEGEVPTRWREAIVVPIP KPGKDKHLPSSYRPISLT SCVCKVMERMVNARLV WILESRHLLTNVQCGFR RHRSSIDHLVTLSTFVM NNFLRKRQTVAVFFDLE KAYDTCWRAGILRTMH TWGLRGRLPLFICAFLK DRMFRVRVGSVLSDTFR QENGVPQGSVLSVALFA IAINPIMDCLPADVSGSL FVDDFTIYCSAQRTCLLE RRLQHCLDSLYSWSVAN GFRFSHEKTVCINFWRY REFLPPSLHLGPVALPFV ETTTFLGLTLDRKLSWSP			

						HVSSLAARCTRSLNVLR VLSGTSWGADRTVLLHL YWSTVRSKLDYGSFVYS SARPSILRRLNSIHHRGL RLAIGAFYTSPVDSLYAE AAELPLTYRRDILLCRY ACRLLSMPHHPSYPSFF NDSLDSHYGLSAIALLPP GVRFRRLLRQLDFTLPST FQIGESPVPPWLQAQVR VHLELTSLPKESTSDAV YRSRFLELRAWLENSVFI YTDGSKTNEGVGCAFVI GDDGFQYRLNDHCSNFT AELFALYQAVLYIRRHR SPSYVICSDSLSAIQSLRD PFSAHPVVHRINRSLQLL AADGPRVHLLWVPGHA GVPGNEAADAAAKAAT LLPRTATYCVPSNDIRGA LSRRLISSWHNVWSSLT NNKLRTIKPLPVAWRTT SRRSRREEVILARLRIGH			

						CRFSHRHLLTGDPAPLCS CGHPLTVRHVLLSCPNF TPVRIALNLPDTLEEILA DDARAAARVLNFINLTN LSKDI			
Daphne	Daphne-57_LMi	.	Locusta migratoria	CCAAACCATCA CTGGAACAACA ACCCCACTAGA AGCAACTCCAT AATTGAAGTAC ACAGTCAACAA AAACTGAACTA CCAGCATCAGA GGCAACTGCTA TGCCTCAGCAC CACTGACAGCA GCAACAGGACC AACTGCAACAC CAACCCAACAT TATCACCAGTA ACAGCAGCTAA	TGATGTTCCTTATA TAGCTAATCAAAAT TGAGAATATAAGTT TTGTTGCAAACATT GTACTATATACCTC ATAAGTATTCCCCC TATAAATATTGTGC TATGTATCCCTTGT TTTGTTGTAAACAT TGTACTATATACCA TACCTACCTCATAG TTCATAAGTATTCC CCTTATAAATATTA TGCTATGTATCCCT TGTAACCAACTGAC GTATTCAGAAAAAT GATACTTGTATCAT TTGATGTTGAATAA	KTTYIVTTAEHRRAKDQ EDLLCVRIFYGFSTRRKA VTNQPGDSSNTQVGAKS SVTVQESQCSFRIIHQNF GSLGNKHNDLDVLAFV NKPDVLVITEHWYPKDK ISVMHPLMMNLCSSYSR MNKSGGGVAIFAKSGSN FIVSFYCVIDVSSLSVEG VIELSAINYRWERQHIIIV GVYRPPSANIDIFLTNLN QLLVEIDNKYSGKNGCI NIIITGDINIDTLSNNSTT KRLQLLLQEFNLNLLIKE PTREVGGTRTCLDNIITN MDKFTYKTAVLKLAISD HHAVEFQCVPNINLHLV KHKPPPGKMKRYETKRT	cd01650	581	845

					AGTTTATTTATTTAT TTA	INDKNIRNLTHVLQSAT WAETNCYSYENFASDFF YYFDMTCPTKSVIVNNN EKCRKVPGRKMQKSTW IDAEVTQARNMLTLFQC KMLENRENTKLKEAFN KYHEYYKNLIGDKRSKH VAHILNNGNNTAACIWN VINSFRAGKDNVINDLVI SYNGSVIECQQDIANIFN EYFSSVGKSINDRLHNN KYAFKGKASSYSIYFTPT NSNEIRNIVKTLRNSKSA GYDGLNINIVKSCIDSISG PLSTAINNILTSGSFPDSL KIAQVTPIFKKGDRLKIE NYRPISILPVLSKIIEKVIY NRIVTFLNKHNIIFENQN GFIKNKSTVTAAAQYIQ LLHSIFDEVITAIENKEY VTGIFLDLEKAFDCVNIP ILLDKLWNIGVRGIAHD LIKSYLTNRKQFVVIRTS			

						SGTYRSTVTDMKYGVP QGSVMGPLLFLIYVNDI KYCTQNSKVILYADDTSI VCKHKELDQLEILSNCL TNEIVQYFNESHLSVSTA KTAVMDFNPSRQTDFEI QLCIGKDIASKENCVKFL GITIQDTLKWDVHVDTL ASKLAKNIFAINMLSKH CNDTQVLRTAYHALMS SNISYGIEVWGATTQAN LNKILKLQKKAVRIICGA KPKDSCRSLFPKAGILTV VNMYIYKSIMLIRKLNL NLNSDTHNYNTRNKDK FHLLSHRTALYERNPLY AGQRLANTIPSCITSLPTE KFKVTLKRILITNPHYSL DEYFCFAQNKENFIPDSL TSV			
Jockey	Jockey-	.	Locusta migratoria	CAAAACGGCGGACGACTATAGGGCCATCACAGC	TAGGAGCCGTACGCACAACTAGGATCACCAAGACGACGCAA	MPPLRQMQRLTRDQNQQMRNPRNHTARRQRNGGENSTTQPLTRSERTQEQ	cd01650	621	939

	9_L Mi			GAAGTTTGGGG CGCTTGATAAG CCCTTCTATACA TTTCCTTCTGCC GCCGACAAAAC ATTGAAGATTG TTTTCCGTGGCC TGCCCAGTAATT ACAGTCTGGAA CACATAAAACA ACAATTGTCGG ACATGGGGTAT GGTTTTCGACTC CGCTCCGTGACT CGGGTGAATTC TGGACGAACGA GGAAAGAGACA CCACTGGTCTTG GTTGTCACCTGT GACATACCTCA TAACAGGGGTC TCTGGCAAGTT ACCAGGATCGC	ACCGCAAGGACAA TCAACAAGCCAGG ACACAGAATACCA GCACCCAGTTATTT TGGAGGCGTAGCTC ACCAGCTTGGACTC CACCGACCATCGGA AACTCATATGTTTC CGTCATTGCACGAG CGAGCGAGCGA	RDRTHQAREEDVATAA LSSSSTVPATPRENAPRF EDMVTLLQKLRDLIKEIP RQITAAVTAAIQAISTGP VMARSQQGLTIAAWNA NGIRTQEGEFRQFLREEA VDVCLVNETFLKPGVHV RAANYVCYRTDRTTHG GGTAVYVRASLAHHRV QLPPLQAIQESHSGSGRH IHRQDHIRRSLQTTKDCG TTCRRRNFTVTPGKVFA AGDFNAKHHIWNSRVIN NHGWHILREAQHNDAM AIGPNDPTHFPENGGQP DVIMLRSDPFPREWGTA GCYYVALIKGVPYVISA STRNVLSSDHVPVVLDV DVDAGPAKRRGLPTNST NWDRYREHLGNSLQDA PMVEDAGVEDNVRTFT KNAIQAVKRATPRLQRD PPDRSRQLPPDILAMITA			

				TGCGACGCCTG		KNRMFREWQLTRNPAT			
				TCACCACGGAA		KRAVNRMQREIKAAVK			
				AAACTCAGATC		AHRHQEWSNKLATLRV			
				AAAGCGGGGTG		DDGTGWKATQMFLRRR			
				AACCACAGTGC		LKIPPLQVGPTIVSAPED			
				TACAACTGCCA		KANAIADVFEANFTPVQ			
				AGCTGTGGGAC		DPIDPEHVRLVQERLPTF			
				ACGTCGCCCGA		LNAQKSVQKRCPGFYA			
				TTTTGCTAATTA		ASINEKLEAPTSSLTKCF			
				CCTCCGCGCTGT		DRCRQKPQRTLQQRHCI			
				GTGAAGTGCGC		MRTGHXPQAWKHAEVV			
				GGGAGAACACG		PVPKPGKDHRLPSNYRPI			
				CCAGCCGAGAT		SLLPAJSKVFERVYLKRL			
				TGCACACGGAA		MDHIRRENILPDEQCGFR			
				GAAAGAAGACG		PGHATTHQLLRLVEEAL			
				GCCCACCCAAG		DXTERREYFGAVFLDVS			
				TGCGTTGGTTGT		RAFXSVWHDGLLFKLLD			
				GGTGGGGACCA		XRPAPAARPINSVVSPRQ			
				CGTCGCCTCCTA		DVPGAHWKRPSXRPNV			
				CCGTGGATGTTC		GNTSGQYSLTFRERSXPS			
				TGTATTCCTCCG		GMTVYSSSCWTSGLPLP			
				GGCTGCTGTAA		HVRLIRSYLQDRTFQVR			
				AGAGCCGCGGA		TEELSTRRPIRAGVPQGS			

				CAACGTCCTGG GGCCACTAGGA AACCAGCGACA GTACGCCCAGG AGTAAGTT		VLGPILYILYTADMPRID RVHLALYADDTALFSRS MNARTLGTRLQRASDEL SRWATKWRLSFNGAKT QALIIARKRVPVDLPRLS VRGTAVPWGRTAKYLG VTLDRRLTWQQHVDDV IGRARGRLRDLYPFLNPT STLPPQLGVAIYTTLIRPI FEYAAVVWGNAAQTHI KRLQGLQNRILRLALHL PRDFPTEDLLRVTEIPPIQ ARFRQAAKLFYQKAARS PNALIRALGHRVQRLNA TRWPHLLRSD			
Vingi	Vingi-21_LMi	.	Locusta migratoria	GGGGGATGGTG TTATTCGCGTCC GCAGCATTCCTC AAGCTGCACGT ATCCCCCGGAT GGTGCCAAACG AATAACAATTC TGGTAGCTGGA	TAAATGAAGCAGG ACTTAAATGGCTTA AATCACTGGACCTG GCTCTGTGACGGGA GATCCTATTCACAC ATAATTTTTAATCC CCTGAACGCAAGTA CTATGTACTGTAGT	GLHQXNVDILALQETHI EEPLNNARGNIPGYRLV AAIGHRAYGVSTYVRQG LDDVTVKHNSSNDDTFI LAVKVSEITIVNVYKPPN TQWTDTVLPVFEHPALY VGDFNSHHSQWGYPAD DENGRRLTSWAENNNFS	cd01650	524	778

				CTGTCACACCA	AACAATAGTTTAAT	LIFDAKDNSTFYSARWR			
				GTGATCTCCCA	ATGGTGGCTGACAA	RGYSPDLCFCSCERSSKP			
				GCTATGTCAGCT	GAACCATACGATAA	MTARRRVLGAFPHSQHR			
				ACCCACCAAAA	TAA	PVIISVGTQIPIIRSFPRPR			
				TCCAAATGGAC		WNFSKADWETFTSKVE			
				CTTCTCAGGACC		QSIRFIPPIITTTVSRIFSW			
				AACATAATTTTA		LLLNSVSPGASGKNTCQ			
				AGAATATAGTC		DGVENATIFTRSFKTPKA			
				CAGTTCTCAGTG		LTQPTICCSLLTLPGKRS			
				TTTGTCATTTTA		GQSRLGDLYFKGRAEHQ			
				ACGCCGAGGGT		VYPSHNNYDRFANLLLA			
				TTGAGCAAATC		IAKQCIPRGFRKEYVPG			
				AAAATGTGAGT		WSRECDDLYKEFQDTES			
				GTCTAAGTAA		PDTADHLLQSLDTARKE			
						KWIKTVENLNFTHSSQR			
						AWSLLKKLSSGKPADNC			
						TVTHPNPCDIAKRIVSSS			
						KSARLPSDKRVAYKLNK			
						ILRSQFTPLEVSRPFTEEE			
						LEAAITTLQTGKAAGLD			
						KIFPEFLKHIGPKTKFWL			
						LTVYNAILATGNLPKAF			
						KRSKIVAILKPGKSPEQA			

						ESYRPIALLSVCYKLLER LIYNRISQILDASIPVEQA GFRPGRNCCDQVLSLTS FIEKGYEMKKKTSAAFI DLSGAYDTVWRKGLLL KFYNIIPCPLLLRFLNSTL SNRLIKVYLGHKESAFR QLNDGLPQGSVLAPLLF NVYIADIPATQSKKFIYA DDIALATQHTDLPTTEA TLTNDLATLQEYFRTWK LKPNPNKTEVSCFHLNN KLANQKLNIYLDGVKLT HNNHPKYLGVILDRSLT FKSHLQKTAAKVKTRTN ILQQIAGSSWGANAQVL RTTALSLVYSSAEYCAP VWLNSAHVREVDVQLN RAMRIISGTLLATPLPWL PVLCHIPPPEIRRQEALIR EYKKILSTPQLPITQYLP QNLSRLKSRKPPLRLAS QLIAEDFSPNTTWAASW			

						DXFEGRNKHLIPSPSTKV GGMNLPRREWTLLNRF RTGVGRCNYWKCKWG MTLDQSCDCGADTQTM DHIAIDCPCVLSLRGWL VFTI			
Penelope	Penelope-80_LMi	.	Locusta migratoria	GGGCGAAACGT CAGGGGAAAGT TTCATGTATGGA CCACGGCATCT CAGCCCGGAAG CTTTAAGTA	TGACTGACAGCAGC GCCCTCTGTTTTGC CTATATATACGACG CTTACTCGCGCTCT CGTTGCAGTACGCC GCTTTACCTCCGAG GATGCCTTCCGCAG ACGAGGGCGAAAC GTCAGGGGAAAGTT TCATGTATGGACCA CGGCATCTCAGCCC GGAAGCTTTAAGTA ATGACGACACCTGC CGTGAAAGCCTTCA TTCTATGA	MTTPAVKAFIYDQTPLR RGYVNIGSAVREHAKKS QQLCSLTFLLRCRDTSTV PKFLKVKRTFRSAQAHR IYSRMESAMLRERIHQT RRALAETNSKLFQLHLLI SNTMQCDDWNKIDNLT FRTMEASAEIAGNRQKR KFDSLQKNVLVGTMED NSRTVINLSDKVLTADE VLVLAKGGNFAVTPRK VPTEDIVANIEAGIRQLS QQSADVIRMETARILRT CKPPTSNLSQAQRKALR DINADENIIVLSADKGNA TVVMKSEDYHEKIMDIL DPSNYKQLRKDPTSKIL	cd00304	435	526

						KMTNQLVKASSFSSDDK KHLCKTEAYSPRLYGLP KVHKQLVPLRPIVSAIGS PTHEVAGHLASLLQPYV GGTDSHIKNSCHFIDIIKE LSVDPDDILVSFDVVSLF TMIPVNEAITCIADIFPSD IVALFKHCLTTTYFQYK NVFYEQIDGVAMGSPLS PAVANLFMENFEQRALQ TARKRPAKWYRYVDDT FVIWTHGEEELDAFLLH LNSINPRIQFTMEKETNG QLNFLDVTVIKREDGSL GHKVYRKDTHTDRYLQ KDSNHHPRQKKRRDRLS NERTDHWVIRSIGKTHT LIVIFKKILTTTHDKKRG VIKTLVDRAHKICEPAHL DDEIEHLRVTFKKNGYS DREIDRALRPRRTTRTDE ERQQPKGKVFLPFINKV TDRIGKILSKYDVETVFR			

						PTTKIKEFLRSAKDKRHP LATPGVYKIPCSCGKVYI GTTKRSVNTRLGEHKRN CRLGHTDKSAVAEHVY QEGDHEIKFNETSVIAKT SHYYARLYREAIEIFKHR NNFNKKEEGLKLDKIW MSTLQQQHDNRLISIEEV GNSRDHSTADTT			
Vingi	Vingi-15_LMi	.	Locusta migratoria	ACCAAGGTGCT GCTGCACAGCA ACAAGTATGCA AGTAAACCTAC AATGAATAGTC CTTTTCAGGACC AACCAAGCTCA GAACACCCAGC TCATGGTAACA ACTTAAGAATA TGTCACTTGAAC ATGAGTCGATTT CATCTGCTAAG ACTCAGATCTTT	TAATTGTATGTATA TATTCTTGTATACTT GTGAAAACCATACG ATAAATAAATAAAT AA	MLVILSHSHVWGYQEED RNGAILNEWMDMYNLQ LVHNAKDPGTFRSGRW QKDYSPDLCIISKSGNNN TTQRKVIKSFPHSQHRPV IIQFGTEIPLVKSVPLPRW NFQLANWEQFRKSLDNS IQWIPPLPSNYERFLGAI KVNAKRYIPRGFRKEYIP GWSEECDELYAEYQDS HSDSTADDLLRALNKNR REKWHKTTAGLNFTHSS RKAWSLMRKLGTDPRL SNRVAHVNANSVAARL	cd01650	335	588

				CACGCCTGATG		MNISKSKVEKAYTRTVK			
				CTGAATCGGTG		KELHIKRSQLEPHSDYSR			
				TGCAATTCAGG		EFSIEEVETALKMVKVN			
				AACCAACCAAT		KAPGFDGVYPEFLKAAG			
				CTACAGAAAAG		PKTKLWLTDFFNAILRT			
				AAATCTTTTGGA		GKLPKLFKQAKIIAILKP			
				TATGTTCTGGCT		GKDGTDASHYRPVSLLS			
				GGGCTATAAAT		MTYKLLERLILNRIQETI			
				GTGACAGTAGG		DQIIPAEQGGFRKHRSCC			
				CATTGCAACAT		DQVLTLTTLVETGFQRG			
				ATGTCAGATCC		LKSAAVFVDLTAAYGTV			
				ATTTCTCTACAA		WREGLMLKLIQIIPCQKL			
				AATCATCATAA		AFLVNNMLSDRRFKVFL			
				GACAAACATCA		NGQSSRWRTLNNGLPQ			
				GAGTCTGTCTTG		GSVLAPILFNLYIHDMPD			
				GCTATAGAGAT		TNSTKIQYADDLALVFQ			
				GCAGGCACACT		SNDLKHCEQVLTSDLIK			
				ATTGCGAATAT		LSDYFHKWRLQPNPNKT			
				CTACAAACCTC		EVTAFHLNNKEAQRKLH			
				CGCTTATTTGGC		VTFDGIEVPHNYNPVYL			
				CTAATCCACTGT		GITLDRSLTFKEQCLKLS			
				TAGGTCCTTCAT		KKLASRVNLLRKIAGSN			
						WGADANTLRSAALALV			

				CCCCCAGTTGTGT		YSAGEYCAPVWENSAHTTKIDVQLNEAMRIITGTVRSTPIQWLPVLANIAPQDLRRKAAKVKLIKNVKAHKNSPLYDALQDPPACRLKSRKPPWCDTESTTSFDVTSKWQERWNENPPHNGHLVQDPSKKVEGFHLPRQTWSRLNRIRTGQGRCAYNMKKWGFLPSSDCDCGAEEQTVGHIVKQCPLRAFNGDLESLHQVTESALDWLSNLDISI			
RTE	RTE-56_LMi	.	Locusta migratoria	CTTCCCAGGGGACACAGCCTTGTAGTAGGCACCTTCAGTGCCGGGCGGGGAGGGCTTGCGTGCTTCCGTGAAGTTGAGAGCTGTGCCGGCAGTAGCGTAGCTACTGGCAGGG	TAGCGCCACAAAAGAAGAAGAAG	MFYSCDDNRHEFGTGFLVHKYLKHLVMNFTPINPRLSVIRIKGRFFNYSIINGHAPTEESIEDLKDDFYDALEKAYDKCPRNDIKIVLGDFNAQLGREDMFQPVIGAHSKHGSTNDNGIRLILFAASRNMIVGSTSFPHKEIHKGTWKSPDGRTINQIDHVLIDQRHKSNLL	cd01650	427	677

				TCACCCAAGCC		DVRTFRGPNVDSDHFLV			
				GTATTGGTCTCA		MAKIRARISNAQKERNH			
				ACTGAGAAGCC		RTKMAAIKRAIDISAKET			
				AGACGAAGAGT		LGIERESIRPTWFDEECK			
				GTCCCACAACA		RVTEEKNTAYQKTLQA			
				TAGGGCAGGGA		NCTRSLLEDYREKRRIEK			
				GGGCTCTAGAG		MTHRRKKREWTKAEIEC			
				GCGTAGTGAAG		MELSKKKMKQGILKTIC			
				CCAGCTTCCCTA		SITSLIRDEEGNIISEVNR			
				GGGGAGGAAAC		LWKDGEDTLMTCESRD			
				CTTATACAAATC		EQAEDISNDCEQGLDDE			
				CCGGTCCTCCA		QPLDGPGLEIKKAKSLR			
				GGTTGGGGGTT		KLKNNKAPGTDSIPAEL			
				GGGCTTGAGGC		LKYGGATLEQSLFELITL			
				TAACAACCTCTT		IWEQERIPNEWNEGILCP			
				CCTGTAAAAAA		IHKKGDRMECRNHRGIT			
				AGTCAGGTTAT		LLNLGYKVLSNIIFDRLL			
				GAAAACCCAAA		PQVDKEIGMYQSGFRRG			
				GAGCAGAAGCG		RSTIDQIFNLRQILEKTVE			
				GACAGATCAAT		YRIGTHHLFIDFKAAYDS			
				AGACGACGACA		INRNQLFHALQEMGIHG			
				TGGCACGGATA		KLLRMVKMIVVQGNYS			
				TTGGACAACGA		DPLEIKNGLRQGDALAC			

				ATGGTTAAAAA GCTGAAAACTG ATCTTGAATTGG TACCTGGAATG TTCGATCTTTAT ATAGCCCAGGC GCTCTAAAGAA GTTATTAGACC AGCAGGGGAAG CTCAAGGAGAG CATTATTGCCAT CCAAGAAATTC GATGGACAGGG ACGGGGTATTG GAAAAGAAGGC ACT		LLFNVALEKVIREAGLQ TRGTIFYKSVQILAYADD IDIIARTVPALKEAFTALS QASKKMGLMINEDKTK YMACGKAYQPNMPSTF SVNGYTFERVDTFKYLG STVTCFNDTSTEIKIRLIS ANKAYFALSNLFQSRLL SRPTKFTIYKTLVRPVLT YASETWPLTETDARLLD GFERKVLRRIIGPVYDGE QWRKGTMQNCTLYMP MLLWRTMEERYNAELY TIYADAPVTNIIRSGRIR WAGHVARMEESEVTKK ILFGNPGGQRGRGRPRA RWLDGVEEDLRKLGCR NWRRVAYDRDGWRKL VGKAKVHPGL			
Vingi	Vingi-31_LMi	.	Locusta migratoria	GGGTGCAATGA AGGAGCCTCCC CAAAGATTGCG ACAGCAAGATA	TAGTTCCGGACACG GAAAGTAAGTAAG TA	MKTNKTPPKRSFPRPLV TISINIEGISRDKEVLLSD LCKEYQCDFLLIQETHR GITRPSPKIDGMKMILER	cd01650	473	728

				AAAAAAAAACC GCAATCGGGCG CACCCCGGGCA ACGTTGATCATC AGCGGCTAATC CTTCAACAACC GGAACGGACTT CTGAGAA		PHEQYGSAIFARPGINVT SASLTDRDDVEILTVRTQ QYSVTSVYKPPDSTFVFT EPANFSYQDINFILGDFN CQSTAWGYKETNSSGAE LEKWAEVHDLLLIHDPK LPSSFNSGRWKKGYNPD NIFVSKKLAAQSEKLIGE PIPHTQHRPIICTVSAVV KPEQVPFKRRFNFKKAN WKKFREIIDLEVTKILHN LRCMTLLSVFRMCLGRQ YARMQNTVYRWTSEDA KPLLERYQSSSTGNELM SVIAENRKAKWRDLVEN LDMKVNSRGLGNCRTST VMPQSQTINLLMSLQTS RHHTLLNGKPRQKIPQP LTRVLAEKPHLKAPFTM SELTAAISSLKTNRAPGI DELRVSKLKTLAVIPYSG CTESQQIKNLGSNTLQW LLDLMNACVERLHIPKM			

						WRKARVVALLKPGKDP NDPKNFRPVSLLCHCFKI LERLTLSRIENVLEQQIIP QQAGFRQGKSCCGQVL NLTQHIEDGFERKEITGV AFIDLTAAYDTVNHRRL LRKVYDVTKDYHLTSLI ATFLQNRRFFVCLQGKK SRWRNQRNGLPQGSVL APALYNIYTNDQPVPNT TRQFMYADDTAVAAQG TNFLDVEEKLSSTLESLS QYYEANHLRPNPSKTQV CAFHLKNAEANRELNVT WRGERLAHCPTPKYLG VTLDRTLTYKHHCNATR EKVSARNNIIRKLTNGS WGANPKVLRTSALALSV SAAEYAAPVWSSSTHAR KVDIAVNEMARIVTGCL KPTPVHKMYLIMGIAPP NIRRDIAAEAEKTKQEK DMRHPLYGHQPADRRL			

						RSRKSFLSRTRALEGSAA ENRIARWEMDLRSPHKA AKEELAPGGDLPYTVW RTLNRMRVEVPKCKTN MQRWGFLKDNENTLCL CGAVQDPQHLLICPYLE KPCTMEDLWEANDKAIS AANFWKLEV			
Daphne	Daphne-7_LMi	.	Locusta migratoria	GGCGGCTAGCG TGAAAATACGG AATGTGACTCC GCGAAAAAACG TGAATTTTAGTG TTAAAAATAAA CAGTGTGTAAA TTGTAAAGCCG AAATTCAAAGC TAAATGAACGT GTGTCCAGTTTA CAGCTTATTATC AGCGCACTGAA GATGGATATTA CGAAACTTAAA	TAATTTAAAGGTGA ACTGATGTAGTGCT GGAAACAGAGAAG TGGACTGTGTCTTT TCACTAAATTTCTA ATACTTAAATACCA CATGTTATATTTAT AAATGATAAATTTG TAAAATTTTCTTTCT TTCTTATGTTATAT AAAATTAACAACAT GACTTGTGTCTAAT ATCATTGTAAATAT GATCTACAGACAAA	MESKIDVFGDSHAKGIS QELQVISKSSTVTGMVK PGAGFKEVTKNIVKYKY GKQDSVVICAGAKDIYE NNAMKMYKQLLSLILVL SNTNIILVNFXLRHDLPE WSCVNKEIHRINEKLKS CCRLFSNVKLIDSSTLHR ECFTKHGLHLNNNGKA RLGNLIRDAIKCNSKWT VTELGWYKCPTAETPAQ TATAPRSTIQKTVATPTE QPTTEPKLTATSEPTSEV TVKTSKFNQPESSAFTTT VAEQTSVAVAPKQITSQ	cd01650	847	1111

				GATGAAATCAG	TAAATTATTATTAT	PESLAYTATTAVAVVVA			
				TGAACTAAACA	TATT	AVVTPEEGSHLQQSSHQ			
				CATCAAGATCG		KKAATFEKPGFLLGPGS			
				CGCCATGATATT		EQQQIIGKHLNDCNQKN			
				TCGACTCTGTCG		LQQGQLVQEADSNQGR			
				GAAAAGTTGGC		ESGNINRTAVSKNSNLKI			
				GGAAGTAAAGT		MHQNIQYASKKSEELEV			
				TAAAGGATGCA		LLAAEVPNIFIMTEHGLK			
				AACAAAAAGTG		DSELKVFNLEGYKNVAY			
				AAAAAGCAAGA		YCRNEHKSGGVAIFVDS			
				AAACAGTGAAA		KVGTTPLDISQIAQISTLK			
				CAAAGAAAAKT		QQQYKSSWGGKCLIIGI			
				GTAGTGCACGA		YRSPSSSIENFLLELEVLL			
				AAATTACTTCCA		QKVMKKYTFVILAGDL			
				GGTTCTTACAG		NIDMLRDNDCTRRFQDI			
				ATACGGTTTGTG		LGSYNLQLHVKSPTRCT			
				ATATCGCTGTG		ETSSTLIDHILTNIPTERQ			
				ACAAATGCAAG		KVQVINTVIADHSGQTIE			
				TAGGCCAAAAC		LKDYAKAEDKKMIKYR			
				AACATAAAAAA		QTNIANLNSLEILVKAEE			
				CAAAACGAATC		WDDVYNEHSVEKSWES			
				CGTAAAAAACA		FCDILRHHINVACPVICK			
				ATCAAAGTGCA		TVSAKQNKKWITKEIKK			

				AATAAACAAAA TSTAATACAAGT GACAGAAAAG		AREELIKLYEDYQKDKE QASQEAYRSKKKLYTR MIREAKASQYKHEIKTS TNISKTVWNLINDNRKE SQTSKDTNIILEMEGVKV TDPLQVANSLNQHYVAI VEKLLNQXTPVTKSIKSP KWVAESLYLLPTNELEV SKIIGALKNKNTEGLDGI SAKIVKHCAKYLATPLT FLINLTMKEGVFPNCLK RSKVVPVFKGGDKEDPG NYRPITITTVLSKIVERVI NDRIVNFIDKHNILNEAQ HGFRKGRSTKTAVANFI TYITEALEQRRKVSGIYL DLSKAFDCVCHKILLKK LNCYGIRGNAADIIKTFA ENRQQCVELKHKCRNK EEKVYSEFQQVKYGVA QGSILGPLLFILYVNDMP TEINEKVIMYADDTTLL VCGDTVKDLEKRATENF			

						QMAERYFTENNLLINRK KTKCTKFETKITKSSEVF TLKIGSADLAEVDCNRF LGIEVDKHLSWENHIDK VCSKISTNIFLMRKMCY TWTQKPCGLFENKHKYI LNEKNVLYVDTETLLK MYYGLIFPHLSYGIPLW GGAADVHIQRLLKLQKK AIRCCTHTETFKTSKESH QMLAKLSPRESCKNKFK EFKILTVPSLYIYMRQDT NCTKSVYIYETIVFLMV NCTAPLNSDIHNHNTRT KEQYHFFSKRLNMTDRD PINAGCRFFNKLPLSIRCI KERSAFKRKLKEHLISGC WYKVKEYLTENL			
RTE	RTE-7_LMi	.	Locusta migratoria	CAGGGACAAGC ATCTTGCGTGAC GTGGTGGCGAG CATGGGGACCC AAGCTGGCCTG	TGACAAGAAGAGG ATCAACCCATGACA AGACCTGCCGCAAG GCAGAATACAATTG ATGATGATGAT	MDRNKLDVLGMCEVR WGGCGELESGDSDCTTX VKIKMECMELEYLRLYY SXEDKNGMYGVGILIGK RLKDKVMKVEYINGRIM	cd01650	523	779

				GCATTGCTTCCC		MIRLKGKKEDLVIIQVY			
				CTTACCTGTGCC		MPTSNHSEEEVEECYEKI			
				AGGTCCCTCCTG		ENLAEREKRNACIVVMG			
				TTATCTTTCCTA		DWNAVVGEGSEGKALG			
				TCGACCACCCTC		KFGLGERNERGQRLLEF			
				GGTCAACTCTTG		CNENLMVVGNTLFDNH			
				TTCATTTCCGAC		KRRRYTWISRFDNKRHQ			
				CCCGACGGTAT		IDYIMIQQRYRNCLKSA			
				TAAGCTATGGA		KSYPGADIYSDHILVVAE			
				GTCTTCATTTTC		VLVKLKRTWKGQKREFI			
				CTCTCTCAGTGC		NIEKLKDKGICEKLENAF			
				AAGGTAACAGG		SEIIMPGQEMECIERRWD			
				AAACCACCACT		HYKNGLKRQLRKRKWV			
				GACATTATTTTC		KKAAKETLGVKERKRV			
				TCTAGACGCTC		KKEWITPEMISKMDDRR			
				ATGGGGTGACC		KWKTVNTEEGKRNYRR			
				TCTTTGTAAAAG		LNNELRRETRKARVKW			
				ATTTGGAGGTA		LTEKCDKIENLZKEGKY			
				AAGCTTCCCCTC		DLMYKEAMDLVFREKN			
				AATCGGATCTC		SNSGIKEIEAADGKMVN			
				CGGGAGGGGAA		ETQEVMRRWEEYIEWL			
				TACAACGAGGT		YDGDTRPSEDELGLEEE			
				TTCCAGACATA		NTVAEDDKGNAILTSEV			

				AGGAGAGATGG AACAGGAAGGA AGGAGAAGGAA AAAGAAGAATT ATTGAGGATAG GAACGTGGAAT GTTTTGACCCTT TTGCAGGAAGG GAAATTAGAAA ACGCTAAACAT GAA		EASVKEMKNGKAEGIDE IPVELLKSLQREGKKELI GLYNQIYMTGNWPKDF CESILIPIEKKKNSKKCEE HRTVSLISHTAKVLLRTL NRRLYGKLNAVMGEEQ FGFRRGVGTRDAIGLLR VIGERYMEKKRKVYAIFI DLEKAFDRVRWDKLMA ILRKHGVDWRDRKLITN LYLNQTARVRIGGVISEE IELGRGVRQGCCLSPTLF NIYLEDIIKESFNGSRGV CIGGRRVECIRFADDMV VMAESAKEMEQMLDDL NTKLEEYGMKINKKKTK SMVIGGRNTKCRLKIGG EEIEQVHSFKYLGSMIM DDMYCTAEIKKRIAMAK EGFKRKQRLLCGPLKKD LRKRLAKCYIWSVVLYG AETWTLRKEDERRLEAL EMWIWRRMEQVKWED			

						RIRNEEVLRRVGEERNM LEVIRKRKRNWIGHCLR RDCLLKEGIEGMVKGKR GRGKPRYQMLDNIQGN KYEEMKRLALDREKWK RINP			
Vingi	Vingi-37_LMi	.	Locusta migratoria	GGGGACACAGA AGGAGCCTCCT AGAAGATCGAA AGATCGGGCGT CCCTGGAAACG GCACACAGCCG GMTAATCCTTC TAACACCAGAA CACCAG	TAAATTTTAAAATT TATTACATGGACTC AAACTAGATGTAAA GGATGGTGTTCTGG ACACGAAAAATAA AATAAAT	MRMSVNKRSFSEPALK MTSVNIEGLSANKELLLS DICSKCKCDVLILQETHR GPTSHRPNIKGMILTLER PHEQYGSAIFTKPGLMV RSAAMTHVDDIEILTVEL KACTVTSLYKPPGADFT FREPSNFQTQPLKFITGD FNSHSTAWGYQETDSNG EKVERWAESSGLKLVH DPKLPASFNSGRWRRGY NPDNIFVTEQIAQQSFKS VEDALPRTQHRPISLTVK AVVCPVTVPFKRRFNFK KARWRLFTEELDSKIDEI LPAACSYDKFIQLIRRVS RKYIPRGCRNEYIAGLD	cd01650	524	819

						DNIEILLKKYEELYNTDP FSEDTLETGEDLLRALTT TRREKWCDLVTSLDFKQ NSRQAWSLLRNLSNDPT SSKTLHTSVTPNQIAHQL LLNGKTTHRSPRPSIVRD MENEHHSLEAPFTLEEL QAAIDKMKNNKAAGED DLRVEQIKHFGPKTLRW LTHLMKQLRQHHADSQ SVEKSKSGSPIKTRKGAN RCQKLSPSILAVPSVQGP RENGANNCVNTMQIPRV WRKARVVALLKPGKEP TDVKNFRPVSLLCHLFK ALERMVLNRISNHVDEV LSPEQAGFRPGRSCTGQI LNLTQHIEDGYERKQVT GVAFVDLSAAYDTINHS RLVHKVYQATKDFGLSR FVQCILQNRRFYVTLQG KSSRWRIQKMVLLRAV YGLWTFKICAVHTTEQA			

						FLCHTARKKQQMENPK NGLAQGSVLAPVLFNIY TNDQPMHPDVRQFIYAD DTAVAAQGSTFEEVEEK LTRALELLGNYYHQNHL KPNPAKTQVCSFHLRNK EAKTKLHVSWKGQELE HCYAPKYLGVKLDRTLS FKQHCLDTKKKVSTRNS IIRKLTNTNWGADPQVL RTTAMALSLSAAEYAAP VWKNSAHVKHIDVAVN ESARIISGCLRPTPTEKIY PLIGIAPPSIRRQVAAEVE RKKQLTDNRHPLHGHQ QQPRRLRSRRSFLWTSEP IDVTVEERRLQLWKEAT PNPILALKEQLAAGGNK SFTTWKALNRLRTGVSN CKFNLSRWGYLDGEPLC DCGNVQTDDHLLTCPNL STCCTPLDLQLANDKAI CVADYWARNI			

					TGAACAGCAACAG	MYNFNCRMESPLKVSTT			
					GTTAACCAAACAGA	PGGNSNEESTIALCNASG			
					TTTTTAAATATCTCT	PRVLGSPNICIKGESEHP			
					GGGACAAGAAGTC	WYPSKLKFNRKFFAGTL			
					AACAATAGCTGGAT	NVNSLLKIGKQKELEILL			
				CCTGGTCAACG	TCAAGAAGTTAGGA	DKHEIQILAVQETRFLDE			
				GTCTCAAAGGC	AAGATTTGGAAAG	NITETKNYRIFKGKPAIKI			
				GGAAGAGGAAT	AAACAATATAAGTG	MKGMPILGTAFCVHKNL			
				CTTAGTTCCCAA	AAAAAGATGCAAC	VDKITEFKSSSERLSLLTI			
				GGAGAAGAGCG	AGAAAGAGAAACT	KCKNKQYTFVNCHAPIN			
	RTE-			GAAGAACTAAT	TTTAAGAAGAAAGT	DDNKKNLSKVEAFWEL			
RTE	21_L	.	Locusta	GGAGAAAACCA	GTTAAAAATGGAA	LEDEISKIPKSNVVILLGD	cd0165	561	821
	Mi		migrator	TGAAAAATAAT	GGATTCCAAGGCAG	FNAQIGREKTFRKIVGEY	0		
			ia	CATTTCGGACTA	GAGGGAGAAGAAG	PAHKRTNKNGTRLIGLC			
				ACAATCCGATC	ACAGGAGTAAAGT	KTYXLKLMSTFFKKLPR			
				TTATCTTGGAAT	GGTCTGAGGAGAG	KAKTWISPNPMLGEFQL			
				TTACTTCCTCCC	AAAAAGGATGCAT	DHVAISKHNIKEIMNVK			
				TGTACA	AGTGAAAAGATGA	VTRNREFDSDHYLTKIK			
					AAGAATACTGGAG	MKLLPXKIQKRQQRLIK			
					GAAAAAGAAAGAA	YNTDRLTITKDAIKHFQE			
					GAACAAAGGAGGA	EIKISKSGKWEELSKEIN			
					AACATTGAAATTGG	RVAKNVFGATKIMKKA			
					TGCGTGGTCCTTAG	WWNETCEEAVAERTKK			

					ATGGCCGAAACGA AGAAAGAAGAAA	WKKWKCSGKTEHLDEF KIQRKETARIIKSTKNFY QMFKNHLKGYQAPSICF RDENGKVGLSNXHNCEI LRKYFEQLLNCEEPNCK LEFPDLHENTETDPPPTA EEIKKAINALKNNKASG EDSITAELIKWSQPKIISN LQLMFEEIWETEKIPED WKVALIHPLHKKGDKQ DVNNYRGISLLPVGYKIF STIILNKVEATLDSHLGE YQSGFRKGRSCSEQIFNL KSIIRHRLTNSKDIVVTFI DFKKAFDSVDRETLDKV IREFGIKSKLANLIRETLT DTKSKVKLQGEISKAFDI KTGVRQGDGLSPLLFNC VLEKIVREWKQKLKDH KLSPIRLGTKHKGIEIHC LAFADDFAILSENLTNAE IQINLLEEIANKAGLRISV EKTKFMTNIKNAPEILVT			

						DIGXIKRVNKFKYLGEM IQENGLEKFAVEERVHK MERAYGITRNIYNKRSL SKNLKIQHYNTVVKPEC LYASECLVLNYKLHKLE VLERKIIRKILGPLRNTE VWKLRSNAEVYRNIENI TETLRKRRLLFFGHLYE			
Vingi	Vingi-42_LMi	.	Locusta migratoria	GCTAATAAGCA TAAACATTGAG CGCTTGGCCAA CGATAAGCAGG AGCTGCTCTCG GAATGTGCAGA AATACAAATGT CACATTCTGTGC ATCAGAGACCC AGACACCAATT AACCTCGTCCA ACATACCAGGA ATGAAGCTTGT AATTGAGGTAT CACACAAAAAA	TGACATGGATTCGG ACACGAAAAAGTA ATGT	RNGQMLIILYLFMTQNC HHHFTVKYGNKALIRIS ASAVLIWRQXASRQFIT HSQRHNIDQLATSWPEC LRRHKTADDTSRRRFNF RKANWTDFTSTLDNCID ELEPIPGNYERFVEKVRY ISRITIPRGCRQQYIPGLS GESKGLLDQYNEMYEK DPFSNETEEVGNLLLKEL SRTRQEKWVETIERLDM TXSSKMAWNMMKKLS ADPIAPKQASNTTADQIA HQLLISGKTGTKPRSRKI HRNLESEVSLIDTPFGIK	IPR000477	299	551

				ATATGGTAGTA		DLDNSIQSLKTGKAAGV			
				CAATCTATGTA		DDICTEQIKHFGKNKTV			
				GAACCGGACCA		DPTTNKVLLNKQIPKSW			
				AATTATAACAT		RKSKVIALLKPSKDPNL			
				CAGTGGGCTCA		ASSYRPISLLCHHSKFMS			
				ACAAACATGAC		GNSELCGRKTNSRTRRF			
				AATTTAGAAAT		PPRKIMYRAGIGADRTH			
				TCTAACAGTGG		RERLRGENITGVVFVDL			
				AAACTGACAAT		SAAYDTINQRTLLAKVY			
				GCACAATAACA		NITKDYELTKVIESLLQN			
				AGTGCGTACAA		RRFYVTLEGKKSRWRTQ			
				GCCCCTAGCCA		KNGLPKAVFCPQCCSTF			
				AAAATTCATATT		TQTTNLLLVIQSTSFML			
				AATGAACACAT		MTWPLGHSSSGQKLPAV			
				AATTTTAGAAC		EERLQNALSQMSVYYKE			
				AAAATAACAAC		NSLKPNPAKTQACAFHL			
				TGCTTCATCCTG		KSKEARRKLQIKWDGFS			
				GGTGATTTTAAC		IEHTAQPKYLGVILDRSL			
				AGCCATAGCAC		TYKSHCQKTRQKTATRN			
				TACATGGGGAT		QLLKKLIGSKWGASPHV			
				ATGAAGAACTA		IRTTAEAVCFSTAEYACP			
				ACACCAATGGA		VWNRSAHAKEVNIALN			
				AGTCTTGTAG		ETCRLMTGCLKATPLGQ			

						LYKAAGCEAPATRRLAV EHTERFRQAFDDRHPLF EAPEPLRRLKSRRSFLRS VPMEPPDKYPLTQKPPP GYHLDWRSWKSLNRLR TGVAPVKTNLVKWGIID ESDVRCNCGERQNMEH LLTCKNCPRTCSLDDLW NARDNALDVARYWAD VL			
Penelope	Penelope-30_LMi	.	Locusta migratoria	TCTTAAACGAA AATGGCTAAAA ACTACGGCCAA CATAAATTTCA ACAAGGCTTGT GTAAAGAATAA ATGTATCCCGA ACTATATCAAG TTTAAGATATGT AATCCAAAGTC CAATGTAGGTA GGGCAGTGGTT GATAAAGCGCG	TAAATGAACAAGTT CAAGCAATACATAC AGCTTTTTTAAAAA CTTTGTGGACCTAT TATAACAAAAGTCG TTTATTTTAAATTTC AACCTCTGTCTTAC TGTGCAAATGATTT CGTGACACCTGAAT GTACCCACTAGCTT GTGGATGATCTTTG GAATTTAGAGGTCT TTGTTTTCATGTGC	MIWDTVKSELDEQLQVE FLNLKSRQDQKLKKLM RQIDCRVNNAQNVDNF QFYPRVVNETDISFNDEE LRLLNKGLKFNVQPSLN DKTVKELIVATKMAVD VAKLNENEKGHVGYKL ANAIKAETNKSCNSMHN NSDHKTVTSINDKLISNK AIVTKADKGNTTVIVSEE QYINKTLQFFSENGVTEI AKDPTKKVQAELKQLL KRTSFLLTEKEKFQCYV	cd00304	353	448

				GAAATTTTGGTT GAATTCCGAAA TTAAATTCCTCC ATAAGAAAAAA GATTGTATTAGT GGGCAATTACT TAAGGCGCATT TAGAAGCAGCT AGTGCGGTACA CA	ATACGTGCGCCCCA CTCGTGTTGTTGAC TCCGTGCATTTCGC TAGTGTTATGCGCC GCCAGTGCTGGGGT CTGTTTGAATCCTG ACGAAGCACAGTAT GCTGAGGACCTTTC TGCGTAGACACGGT TACAAGGGAAGTA ACCACGTTTTCTTTT ATTACGTTTTAGAT TTTTATTGTCTCAG GTTTTAAGATTCTT GTATAGATTATCTG ATGATGAGGGCTTG ACCCTTGAAACGCG TCATATTTTAGTGT TTTTACACGATCTC TGTGGTTGAGCCTA CTTGACAAAATAAT	MNPTAPRLRSQLKVHKN GHPIRAIVNSMKGPLCKI TKKLDKILKEFYVFGDN FSVSNNTTLINEIKDKPIA DTSHLMSLDIKNLYANV PVDEAVELIKANLLKHK RLSQQEIVEFIDLLQFILK YNYFAFNGKFYIQKDGL AMGSSISGTIADIFINNLE NRLFTNMQHPVDKILYY KRYVDDTLLLIDGGQTEI TEIVNKFNSLHEKIEFTV EQEKDGTLQFLDLNITK QDGRHLFKVYRKPTTTD ILIHNSSCHPNGQKLSAL RCLVDRAFKVPMNEKD RTTEFNIIRQLAKVNGYQ ENVVNNLIRKKQRAKSN KNGTXNIKRVMVPYFG NLSQKVANILKPFKVEA AFKTNNTLHHKLRNNLN QETDKYDKAGVYKIQCS SCNANYIGQTGRSFKVR			

						YKEHMDAFRLNNFDKS AVAMHIRETGHPLKRIE DDLKILHKQDKGRKLDL LEQLEITIHNAKATFH			
Nimb	Nimb-19_LMi	.	Locusta migratoria	CCCCTCAGGGG CTCAGCATTAAT TTGCTGAGTAC GGGCTTGGCGA ACCCGGGGTCC TGAGCCGTGGG CTGTGTCGCGCC AACAGTCCGCT ACACCGTAAGC TCTGTGCGTGCT TCAACGACCAT CGTTCGGCGCG ACGGTGGAACG TTGTACGTTTCA GAGCCCGGGGA TCTTGGCTTAAT CGCCCGGGTCG CGAGGACGGTA CAAACCTCTAT	TAATGGGAAGCCTG ATATTTCAGGCACC CACCATTCAAACGG TTTTAATCTTTTCGT CCTGTCAGTCCACT TTGACTGGCAGTTT AGTGTGTGCAGAAT GGCTGGCTCATCCT ATATTATCTTTGCG ATCAGCCAGCCACA TTCTTCTGCTGTGA TATTTTATGTTGTTG TCTGTTTGGTCAAC TACTTTGTGTTTTCC TTTAAATGTTGTCT CTTAGCACGCATTA TCACTCTGACCGTC TCTGCATTAGACGT GGAGTACTTTTTAC	MALPGAKPASLSVXGPN LMLRHMTPRRSLPWPRH GRNAGREIEVRLMPPXT WYVPGLTANRLPPRSPC FLWKTLRISSGRWMACQ DEIWXCXHXDVVPSPVT GTACLVRGGGWHVKM RSGXVLITTSSPAQSRAL LACDKLGDVPVSITPHK SLNMVQGIIFHKDLLLQ XDDELRKNLERRGVHFI RRVHRGPKDNRVATGA FILAFEGDSLPEKVKVMI YRCDVKPYVPPPMRCFK CWRFGHMSSRCXGNPT CRECGRAXHPDAPCAPP PVCVNCGQHHSPCSPXC TVLHKERKIMEYKTLDR LSYSEAKRKYXKLHPVS	cd01650	1172	1430

				AAAAAAACCTC	CGCTAATTAAAAGA	MKTSMPQLHHTASAISA			
				AATCTCAAGGT	TGTGCTCTCTTGGC	TALHCLVFHSGPPGXTA			
				GTGCTACGTGC	TAGGAACAAGGGA	HJCPHGLWGHLNLHRPQ			
				CGAGGAGATGC	CTGATAACCGTTGC	GPYFGSSNHHSTCGXISP			
				ATGGCTGTTGA	TGTTTAGCCCCCTT	HPPPLHLWGHQSPPLSQ			
				GGTTGAACAGT	TCAAAAACCCAACA	RHNSXLWLLSRGRGPSD			
				CGCTAGCGAGC	ACAACAACAAC	PFLLRLLLLKDTSRWLK			
				AACCTCTGAGG		EPPTAGRRASRAPQFLG			
				AACCTGCCGCA		TEPAEPSQHEKPKEQRET			
				CCTCAGTTGTAT		KEKDKKPPKRRIPWWPP			
				AAGGCTTACCC		CHHPPPPLLLRTRETTQE			
				AGGCAAGCGGG		KDSLVATVPPPPSTSSSE			
				GCTCTGCCTGG		DEIEILASPKDLAIAESSE			
				GTGGACCTTTA		AMDVEALPQKSATAED			
				GTTCCCTAGCTG		TVNPTAPSKSACPSDQE			
				CTCGTGGGACG		ALGSASKIGDSRRYCEP			
				AGAATGAAACG		YCAVEISMSQRPGGSLIL			
				AGCAAATA		QWNCGGFFRHLPELQQL			
						LCRFPAFCVALQETWFP			
						AQRTPALRGYRGYYKN			
						RTIYDRVSGGVCIYVLA			
						TVPSAAVPLHTTLEAVA			
						VRIRAMQELTICSLYLPP			

						DEMVSLTDLAALVAQLP PPFLLLGDFNAHHPLWG GAQTTGRGRDIETLLSQ QDICLLNTGAPTHFCVA RGTYSAIDLSVCTPSLLP HFQWRVHDDLCNSDHF PVLMSLLDFQPAGCPSR WSLRRADWQRYTSAAT VPVPLPADVDMAVHQV TSTIVSAAEATIPTTSGSP RRKSVPWWTPEIAEAIR DRRRALQRHRRHPTVEN LIAFKRLRARARQLIRQS KRDCWETYVASIGCRAS RLGCGPRLGAFLGTVHQ QRSPAYTSRVTYAPTRL PSLNTWQSIMLAPLQHQ NTVQTSVHTGLLGDVCR QHRLSRLASRVWSQVRR LFGYRPPAAIPSLHIQGN LCTDPTSIAEHMAEHYA RTSATSEYSADFRALKE RVEQRRLSFHTRRAESY			

						NSAFSEWEFETALAACP DTSPGPDRIHSQFIRHLS VEFQHQLLALFNRLWAE GAFPSQWRESIIIPVIKPG ADPQVADSYRPISLTNV LCKLFERMVSRRLMWIL EHRGLLSPHQGGFRQGR STVDNLVALESAIRTAFS RREHLVAVFFDLRKAYD TTWRHHILATLHGWGL RGPLPLFIQNFLSFRSFR VRVGTSLSSPRVQENGV PQGSVLSVPLFLIAINGL AEAVGSSVSSSLFADDF CLYFSSASIGVAERRLQG AIRKAQSWASTHGFQFS PSKTRVMHFCRRRAVHL DPELYLYDELLRVEETH RFLGMLFDSRLTWVPHL RQLKQKCVRYLNILRCL SHTSWGAARTSLLQLYR ALIQSRLDYGSFVYGSA APSALQVLDPVLHSGVR			

						LATGAFRTSPVASLLAE AGVLPLRFRRQQLLSSY AAHVHATPHHPNRRLLF PAAALPSPRQRPRSGLPI AVRVRSLLSSLESLPIPPT FHCPYRTAPWSLPRPQL RLDLSHASKASVAPLSL RRQFQAVLESFCGAEVV YTDGSVVDGRTGYAYV HGEHIEQHSLPCGSSVFT AELVAIYRALEHLRACP GESLVICSDSLSGLQALA KCFPKHPLVQSIQESVSR LVRSGRSVVFVWTPGHI GIPGNERVDQLAKQATT SSPLAIGLPESDLRLEFH RKILGLWNSEWANSAQP NKLRAVKLSTVAWKSSL RSSRRESVVLCRLRIGHT WLTHGHLLRREDAPQC DCGAPLTVVHILVDCPL LVALRQSFNLPDALSFM			

						LGDDAAMVERVLDFIRA AGFYHVI			
Penelope	Penelope-1_LMi	.	Locusta migratoria	TCTCCTGGCGA ATAACAACTTC GAAAAACTACG GGTATGCTGCC AGGTATCAAAA TCGACGTCAAA TTTTTCGAAGCT GTTATTCGCCAG GAGAAATTAAG GAATCCATTACT CTACGGGGAGG AAATGTATCGTT TTTGCGGAAGT CAGATAAGCTT CGTAGCAGCAG GGTCAGGTTGC AATGTGCATGT CGTCTTTGTTGG AATGTCGTGAT AATAGTGTTGTT CCAACTTTTGCC	TGATGTTTATTATC GATATTCTGACCGG TTTCATCTTTGGTC GTGTCGGCGCTGTG TGCTTGCTTGCTCG GTGTGTGTTGTCTT TCCTGTTTACTCATT GAAAACCGAGGTTT TAAAGATCGTTGCA CATTGCTTCCTCAC TGCAGTTTCGCCTT GAAAATGGCAGGG TGGTCTCCTGCCGA AATATCGGCCGTCG TCGATTTTGGTACC TGGCAGCATACCCG TAAGTTTTTCGAAG TTGTT	MDPDIVILPADKGNATV VLNKQDYLQKMQCLLS DSAYRRISADPTKSVER KTNNLLKKSSLSQETIKS LNSYCAVPPRLYGLPKV HKDGCPLRPIVSNIGAPT YRVAKHLAGLLSPLVGR CEHHIKNSVDFLRRLEG LHLADSDILVSFDVVSLF TRVPLSDSLQLIETKFGT DLTNLFRHVLTSTYFLFN DEYYEQTDGVAMGSPLS PIIANLFMEDFEDRALES AALKPACFFRYVDDTFV VWPHGSENLNYFLEHLN SIHPNIRFTMEVEKNGCL PFLDVLVRRKADGTLGH AVYRKPTHTDLYLQADS CHHPAQREGVLRTLVHR AHVISDPESLPTELSHLE VTFRQNGYGEGQIKRAL	cd00304	201	285

				AAAGTTGTGCA		RYRPHVQRASDDSTESA			
				CATATCAATTCT		PTSVAFLPYVGSTSNKIG			
				CCTGCCGCTGGT		RILRKNNVKCVFRPPSKI			
				CGCATTAAACA		KALLGSVKDDLGLRKA			
				ACGTGCAGGCT		GVYRIPCSCGMSYVGQT			
				TAGCTCTTGTCC		IRTVEDRCIEHKRHTRLQ			
				GCGAGAGGATT		QPNKSAVAEHCLDTGHP			
				CATTACACCCG		MEYNKTEILACTSNYWD			
				CCGAGAATTGG		CVIKEAIEIKLGDNLMNR			
				ATTTTGTGTCCA		DGGFCLNSAWDPALSLV			
				AGGAATTATTTC		KKQRDKTNPTSSVEN			
				GGCTTCATTTGT					
				ACATTGCCTCTG					
				AAATTTCGGAA					
				TTTTCTTGGGAC					
				TGGTTGGACGG					
				TGCTTCCTGGTC					
				ACAATCCGATT					
				GGGAATTTCAA					
				ACTGTAACGGC					
				CCGGCATTCGG					
				CGAAGTTTGAA					
				CGGTTCCGTAG					

				CTCAGGGTCTG					
				CTTTTACAACTC					
				GACGTTCTGTG					
				ATTAATCTCACA					
				GAGAAAACCTT					
				TGACGACGCAG					
				CTCTATCCGTTC					
				TGGGGAAAGGT					
				TTGAATTTTGCA					
				CCGACGCCTAA					
				ACGTTTACCTGT					
				TGTTGATTTTAT					
				TAGTTCTATCGA					
				ACAGGCTGTTTT					
				TAAACTACCTTC					
				TGACATTAAAG					
				AGGAGGTTAGG					
				AGGGAAGCTTG					
				TCGTGTGTTGAC					
				TAGGGCACGTC					
				CGCCTGTGAGT					
				AATATAACAGC					
				AGCTGAGGCAG					

				CTGCCTTACGTT CACTTAGG					
Daphne	Daphne-26B_LMi	.	Locusta migratoria	TGCATGAACAG TGTAATACTTTA AGAGCAATATT TTTGGACCTAA ATTCATTTACAG TAAGTGTCTAG CACAAGATGGA CTGCATCTAAAT AGGCTAGGCAC AAATACTTGTA GCAAAGTATTG GAGATGTATGT GTGTGATAAAA ACAAGGGAAAC TAATGTATAGG CAGGGGGTGAA AATCCAGACAA GAATACAAAAA GACTACAACCT CTGTGAACTAG AGCATGTACTA	TAGAAATGTAGGTC TATGATAAAGTGTG TAGAGAAAATGTA AATTTACTATTAGT TATTACGATGTATT TATTTATATAAAAA ATGTAATCATGCTG TGTAAAAGTGGCAA ATATCCATTTTATT CAATATGTAGAGTA ACAAAAATTCTAGC AGTGAAAAACATTA TGTGGTAAATCGTA CCAAGTAAAATGTT AATCTACTTTTAAT ATTACAACACAAGA AAAGAAATATGTTA GGCCAGAAATGGTC CAATATCTATTGTA CCCTGTACAGTTCT GGATTCACAGGACC	MLFHQNVQSLNNKLSEI KIMLSNELIDVSVLCMSE HWLRDDALQTVSILNYE LAAYTCRQQILGGGICIF VKKGLKKSFFFFLFFFFF FFEISAIELPSLRDNFNM HLQIPNGDIKTFYQNLEN TLTQTVKNKQTIIICGDF NIDFKGSAKEKTDLMNI LVSHTLKITIHSPTREIVT LSLQVYNRSNNNKPTKE QYKAEAVKTGLGDHHG QTISFHIKGIEVNKLTSTY KTRTYNDQEIEIFKNLLK SESWEQIYDTSTENISEK FDIFIEKIKYYFDVVFPV KTRRLKDTLPKKIPSQN KWITTGIRISCANKRFLY RLTKTCDTTPAFDRYFK NYKTVLRKVIKSAKKLE NDRYISNSSNKMKAVW	cd01650	503	794

				GAGATATTTAA GGAAAACAAAG TAAGTGCCAAC CAAGTAATCAG AAAAAAGTGTA TTATTATTTT	AATAAATAAATAA ATAAATAA	NIVKKETGNNKTKNKNI QLKENQQTTSKPKEVAN MFNTYFINIAEKLIEKNV IKNPSQLATKIKHIDKTIF LTPTNTKEVQQEIKLLKS KHSSGIDNIPDFIVKKCA DFLSEPLTYLINLSFSTGS FPKCLKIAKVKPLHKKG SQEEVSNYRPVSQLSAFS KJLEKLFIKDCPISQNPN MVLEKNHSTETAVFSFL YETLQALDRREHTTSIFL DLSKAFDTINHNKLLSKI QNWGLGAAHXXIKSNP HDRKQFQKAVLIEITYES NGTVSTYHSDSLVVKHG VPQGSILGPILFLLYVND LYINNTNGKLYQFADDT SILITGSTKEIFESNIAAIT KNIRDYFDANGLIINVEK PLQWIYLRHRIDHQLTQI LICKWPDNKCGKTVAM DIFTTQNRSPANPNIDIN			

						GKTIEKTECTKFLGIWIQ QNIKWDSHVDYINRKLS KTCYIMRTLKNSISYNNL KSIYFAYAHSQMKYGIV FWGNSGASMKTFRIQKK IIRIMEGVDQRTSCKPLL KRNTPTSKHVHNGKPNY YKRKIKKEILPLPSMYIM ENLIIIKEKLNSKSPCITR NETVHCHNTRQKNNVH VKQTNTALFQKGVLHPG IRLYNALPKAIKSKTDTT KFKTAVKDYLVMHCFY SVDEYLEFQNYSIS			
Penelope	Penelope-8_LMi	.	Locusta migratoria	TCACATCAGAA ACCTCTACGGG GAGGAGATGGT TGAATCGTGGG AAGACTGGATA AACTTCGCAAG AAGAAGGGTAA GCTGCTGAGCG CTCTTGCTTTTT	TGAGGATGGCGGG CAGTTGCGCCGCTG AAATATTGTGTCAA GTTGACTTTAGTAT CCGGCTGCAAACCC GAGAAGATTACA	MSDWAHRSAVDKQSSK FQRLSRPSISHNTNRRTV INFTDKELDDAMLSVLE KGLNFAPTPKNVPLTEFI CSVEQAVSKLPGDQAEE IRQEVAHTLRRAPPPRSN ISPLERAAIRSLREDQDLI ILPADKGNATVLLSRDD YLRKMDCLLDDSAYRR	cd00304	318	402

				TATTGAGATGTC		LQKDPTERIKRMTLSLL			
				GTAGTGAAAGT		KKSSLSDDVLKKLHPGA			
				ATCATACCGAA		AVPPRLYGLPKIHKQEV			
				GTTCGCCAAGTT		PLRPIVSNLGAPTYNLAK			
				TGTGCATTACAT		HLASVLSPHVGKCEHHI			
				TGATCGAAAGT		PNSMDFIQRLKSLRLSPS			
				GGCCAATAACA		DLLVSFDVVSLFTRVPLE			
				TCAAAAAACGT		ESLQLIGEKVDKETAELC			
				GCTAGTCTTGCC		RHVLTSTYFLFNDKFYE			
				CTAATAAGGGA		QTDGVAMGSPLSPIVAN			
				GCGCATTCGTTT		LFMEDFEDMALKTAAL			
				TACTAGAAGAC		KPTCFWRYVDDTFMVW			
				AATTGGATCTC		PHGEEALNRLLDHFNCM			
				GTATCCAAGGA		HPNIRFTMEVEKEGKLPF			
				CTTATTTTATTT		LDVLVYRRDDGKLGHS			
				ACATCTAGAGG		VYRKPTHTDRYLHASSC			
				TATCTTCACGGC		HPPVQRTGVLRTLTRRA			
				TGTCGGCTGCAT		YAISDAESLTQELDHLK			
				CTTGGGAATGG		VVFKQNGYTDKQIRCAF			
				GTTGATAGCTCT		QFGPSPEPLNETHKSVAF			
				TCGTGGGCT		LPFAGSISSKIGRILKKHD			
						IKSVFRPPVKTRALLGSV			
						KDDLGLRKPGVYKIPCQ			

						CGKTYIGQTSVPFRTDV LNINATHVCSSPRSRRDS LSMWENVYWSDIRTIQD RCVEHQRHTRLLQPEKS AVAEHCLDEGHRMLYN ETHMVASASRYWDCVL KESIEIRLSDNLINRDKG YLLSKSGTLCYQMLKNN GWYFRNLNNNFLIIDYR FRRILPPERCMSCAPEGS YCHFLRTRYGSSAMYK GAADCTGNSVPARLCTS TLT			
RTE	RTE-62_LMi	.	Locusta migratoria	TTGCTTGTCTCC TACCGAGCGCC TCCCCAGGTGG CGGATAGGGGG ATGCTCACCAG ATATGGTGGGT ACCGGGGAAAT AAAATACCCGG GGTGGACCAAA ACTAGCACAAG	TAATTCATAATATT TATGTTGGCCATAA AAGTAATAAATAA AAAAAATAAATAA CCTCTGTATTATTA AATTGTCATGTTTA TTTTTGAACTAAAT TGTATTAATGCATA ACAACAAATTGTAA TAACCTCTGTATTA	MKRPFREKFPPSNPIGKE KATSDSVGSQLEDSNES EHLKSPNYHQARNAKIK HKHINYIATHNINSLLQT GKLKNFTDELNRQNIIIA GLQEMRNTIEDPFEXQG YRIYNGKPGQRVMKQC PQFGTGFIVNLKVIDSIT DFKAISPRIATLTLKATN KIYTIINAHAPTNQKNFL	cd01650	596	831

				GGCACGACGGA TCCATATGTCAG TGGCGGTACCC CAACAGCGTCT GTTCCACATGTC AGTGGCGGCTG ATCAGTATCCTC TGTGGCTAACT ACCATAGTTGT AACACGGAGCT CGCTCCACTCA AGGTTGATAGC CTCTGACAGTC AATC	TTAAATTGTCATGT TATTATTTTGAACT AAATTTGTTAACAA CAAATTGTAATAAC CTTTGTAATATTGA CTAAAGTGGTCCAA TGAAGGCCATAAA ATAAATAAATA	TRTMVEVEKFWDLLDQ TINQVNKNHVKILIGDFN AQLGKERKYRDIIGKWA PHKHTNRNGQRLVEICR TYGLISKSTYFKRKPSKL MTWKHPDWTKGEWQL DHVCMDKFFHKEIQNV KVLRGMDTGSDHYMVN IKIKFTPLSKKRQRALEIK ESMIHINKKTESIGDKRK YDPHKLIKNDKYQKNTQ NTVITDNLQELIPILKQH AEESAPVNQTSKKHAW WTTECDEIHDERHQAWI KFQTHRTEENAYNLKNI RKKFSQIIRKTKRLYQKD VLSSIEENYHETNSRDYF KAFDKQLRKFNPPTLLL RNKDGRIAHNNKENAEI MAEAFNKLLNLTNLKSF KAFNNRHKHPCNNQTR KYSPPTVQEIETALKEM RNYKACGENQVFAEMW			

						KYAGTTTRTSLHTILQKI WTTEKFPEEWTTAIITPL HKKGDKSDPDNYRGISL LDCTYKIFSKILHNRIRE QLDKELGEYQGGFRPW RSCAEQIITLKLIMAYYR KRNKPLSITFIDFKKAYD SIHRPSLLKILRHFGLHL KLIRLTELTLTNTNSKIK FEGNFLDHSIHMREWYK ANPKNIKVGTQKDNIKL NCLGFADDLALLANNIQ ETKQQIRSLQEIAQCIGL KISFEKTEIMVTEPPLAN KIRIGVHNIKIVEKFKYL GEIITHNLNEKPSWQSRI NKLIKMHHVTKNTYNK KCLSIATKLKHYKTVTQ PQITYASETIFKTTNTVAI DRLLKIERRIVRTCINKK YQKDGIWRIASNETVYK EVEPVTSTMKKKRISFLG HLMRTPENRISRKIIEKL			

						WNSKSNIRWITEIKEDM MKLQITMEDLKDKTDTL KILQDKHVTLPIKINKQR LGRVISDEERRARSERM KKYWANKKKNKNKNN RRT			
Daphne	Daphne-13_LMi	.	Locusta migratoria	CAAGATGGCGT CAAAAGTTCAG TGCAGTGACAC TTCGATCTCCTC AAAGAAGCAGT TATACAACTCTT CTGCTTGTAAA AAGTGCGATAA ACGCGTAATAA AAGGTATTCTGT GCGTCACATGC AAATTTTGGAT GCATGAAAGAT GTGCTAAAGTG AACCTTAAGTTC GTAAATGACGA GTTKCCATGGTC	TGATTTTATAACTA CATTGTATGTATAT ATTATTATTTTAAA CTGACGAGTCTCCT GTACACTAAATCAA TGATTTAACTTGTA TGTTATGAGACAAT AAAATTCTGATTCT GA	MIKKNVWFQTVKSMCV QEKKRVVPDCEIDVRPGI RTNQLNNHFKNIQQAKH NSKNVTNSENYKGVFIH VGTNSLRSCSEEDLVNE TRNVIRSAKNIYPYSRIII SSILRRRSVSDKYIDKIN CNIRDQCNRLGAIFVDS NKFIRDDCIGQDGLHLN RLGSATFSKMFIDLCKII RNRGNWQSVRGVNPVQ NYTKQGKLAVSEGGEL KHVVDKDTTKEDVLKT VNRYNSSFLMHWNIHGL RATYLQNKIAKIDELQII LSEGRNIKVVCLNEHWI TEDSITVLNKLENFNVAS	cd01650	767	1046

				GTGTGGCGAAT		SFCRKNKTHGGSCILINS			
				GTTTATGTGATG		SLEYKERDDFNFLNEELI			
				AACGGAATGAA		FESCCVELEQQNIVIISIY			
				CTGTTAAACAC		RIPGNDVTKLFFDKFQCL			
				AATAAACTCGA		LQKLTKEMRKKIIIAADF			
				AAAATGAAATC		NIDTASETSISTKFNDLIQ			
				ATCAAGATACT		QYGFRLNFTEFTRVSDK			
				AAATAATGACG		SATCIDNILTNYTLNENA			
				TTGAAGAACTT		KFTLNLGISDHYALLME			
				AAAAAGGTAAT		LLKTNKENGRKCFLKRQ			
				AGTTGCGTTGA		FNPDNIHLFCCRLNDID			
				AAGAAGAAAAT		WSLQHESSCSENFDRFL			
				GCGAAACTTGC		NSFLHVFNETFPPMSAK			
				AAGTGATGGAA		ERQQVSNISTHVCQRKT			
				ATACACAAAAG		TSKLKWITTGIKISSKRK			
				TGCGAAAAGAA		RELHAELKSNKSPEFIKY			
				TGTATCACTGTG		VKSYKKLFRKVVKAAK			
				TGGAAAAAACC		EMANTKFIRKHKNKSKA			
				AAGAAAATCCA		VWTVVKSELGIKPSNQEI			
				AGTGTTTCTAAT		SKLKIQDKIIVNPAQISET			
				TCTGGAAATGA		LMNFSTKNTRQNNCKSC			
				TGCAACAGTAA		SDIRNFNEFFITVAKSDV			
				ACAAATCGGGT		NVPSYQDNVHLFGSSLS			

				AGGCCTAGCGT		KNECLIKFDTVTVEDVE			
				GGAAAAGTGAC		KTILNLKNKNSVGWDEI			
				AAACAATGCCA		PSKVIKAASNIIAYPLCKI			
				CTTATGCGGAT		INQSLEEGYFPDSLKYAE			
				ACAACAAACCA		VKPLFKKGSKEDMGNY			
				ATACAAATGGA		RPISSFPCFPKYLKYQKQ			
				CTGTCGTACAG		FGFQKGKSTIYAINEFVN			
				CACAAAAAGAG		KVTSTLDKSQKAAGIFC			
				GGACAAGCAGG		DLSKAFDSVNHSLLLYK			
				AAAAGAAAGAA		MEKYGIMGTALQWCTS			
				AAAGAGAACAC		YLTERKQRVVITSSKGK			
				TAAAAACGACT		YSSKWRLSHRGYHKAQ			
				TCTTAATAATCG		FVETIPQGVPQGSILGPIL			
				GTGATTCC		FLFYINDLPININYHSVLF			
						ADDTSVLIENKKADEIPD			
						NVLKTLENLDVWFHLN			
						GLKLNIKKLRCLKDPRK			
						PRCLVPSKWSKIKYKKT			
						QVMQFKTKQSKLNDIQI			
						KHNSQELHEADHVKFLG			
						MKLDRIYPGPHTILGNET			
						GQNLSWSSHIDYLKNKL			
						NSLAFAMSILANVTSLD			

						TRKIVYRSYFESIVRYGII FWGNASTMIRILTLQKKI IETCGAKPRESCRPLLRD LNILSVPSLYLFEIILSIIIQ NYFFIITSSTVMKQEQRQ FHAPTHRLNLYAQAPQY MGMKICNKLKNTHFRK VETESMKRKLYSTLVQK CYYSIDEFLNDDLSIEQD KTYNYVL			
Vingi	Vingi-7_LMi	.	Locusta migratoria	GGGGTACATGA AAGGAACCTCC AMGAAGATTGC AATAATAGAAG AACAAGAAATG AACAGCAATCG GGTGTACCTGG GCAACGTCAGA AG	TGACCGACCGGACA CGGAAATGAATGA	MLRSALTADPFTTGLHS NGQHSDRSFSRPFVFISI NVEGFSREKEVLVANLC QKYSCDIXLVQETHGGS TKNRPRIAGMELILERPH DKYGSAIFARPDVNITSA SLTEDNNIELLTVQTPFF CVTSVYKPPVPILASKHQ PTSTIRTRQQHRTPHSAN SVFLCYISLQTTCANFSI KTPANFNDQDVNFVIGD FNCRSTSWGYPDSDTNG TSLETWAEAHDLHLIHD	cd01650	561	816

						PSSHLHLTVAGGKEGTT QTTFCFFKGSCTMYEDC RRAHPSYATQWQVEKR VQPRQHFVSSKVAAQCT KTVGEPIPRTQHRPVICT VNPVVVPEEVPFKRRFN FKKANWQKFSEELDSKV INIPPQPEMYETFIKCVQ QVSRKTIPRGCRTQYVP GMSSDDKLLLETYQTLY EEDPFAKETIQAGEILLE TISATRKAKWSNLVENL DMXLNSRQAWKLLKNL SGDPVKSQDNFTSMTPD KVATQLLQNGKIKGRTP KEALQRVPETEHHFLRD PFSMAELEAAISSLKMH KAPGVDDLRVEQIKNFX CKTMEWLLSLMNKCVA KLHIPKMWRKSRVIAIL KPGKEPTDPKNFRPISLL CHTFKILERLVLNRISAT LEXKFIPQQAGFRAGKS			

						CSSQILNLTQHIEDGFER GEISGVAFVDLTAAYDT VNHRKLLRKVYDLTKD YQLTSVISTFLQNRRFQV SLQGKRSRWRTXKNGLP QGSVLAPALYNIYTNDQ PIPANTRLFIYADDTAVA VQGPTFKEVEKKLSLAM DELSKYYDANHLKPNPS KTQVSAFHLRNREAEHK LDITWRGEQLKHCPTPV YLGVTLDRALTYRQQCL NVKQKVSARNGILKKLV GSSWGANPHLLRTSAFA LCISAAEYAAPVWDASA HAKQVDIAVNDTARIISG CLKPTPVQKLYPIVGIAP PNIRRAIAADYERTKQET DHRHPLYGHQAARHRL KSRKSFLARTAPLEGAPE ANRLATWKSSLPADVSP KEEIAPGANLQYPVWRS LNRLRVGVPRCKSNLKK			

						WGILPASADTSCDCGAD EDASHLLVCPLLDNTCT AQDVHEGNDKAIAAAIF WKC			
Vingi	Vingi-5_LMi	.	Locusta migratoria	GGGGGACTTGA AAGGAGCCTCC CAGCAGAATGT GCTGTTAAGCT AGCATCACATT CGGGCGTCCCC TGGGCAACGGT TTTTTCTTGCCG GCCAATCCTTTC AAATCCAGAGA CATCAAGGCAT AGCTAGTGTA	TAGAAACCTTTCTT TACATCTGCATAAT TTTTCTTTTTAAATT AATATGTAGGTATT TTATGATGTCTCTG ACACGAATAAGAA	MTSQHRTFSGPIMTIISV NIEGLSLCKEELIKDLCK KNKCDVLCIQETHRDSIL NRPKIQGMRLVAERPHR QYGSAVFVKSEIQVLSTS VSDINDVEIVTVELKNCT ITSVYKPPLSNFNVNATH KHNSTKVNFVIGDFNSH STRWGYPQDDENGMAV ISWAESNKLALIHDHKLP PSFNSGRWQRGYNPDLI FATESVSQQCIKTVCPPIP KTQHRPIMCQVMTAVRP VCVPFRRRYNFRKANW LKYQEILDREVAFLPPNP DSYDKFTDAVKRSSRLSI PRGCKTEYVQGLTPETV ADLHTYTELFEKDPFDE STVAAGQQLLASISEAK	cd01650	471	726

						RNSWVNLMEELNMSINS HKAWRLLKRLSNDPTSL NNHANITADNIATQLLK NGKPENRSRRQTKNMS NTTEDETNSLSQPISIDEL NEAIAKCKNGKAAGLD DILVEQIKHFGPQTKQW MLNMFNKCMADCKIPSI WRKSKVIAILKPGKDIN DPKSYRPISLLCHMYKIF ERILQTRLTYVIEPKLIPQ QGGFRPERNCTSQILGLT QFIEDGFQKKRVTGAVFI DLTAAYDTVNTRQLIEK LYEITRDRKFTDIIRNLL QNRRFFVEFQGKRSRWR NLTNGLPQGSVLAPSLF NVYTNDQPLPVNTSSYI YADDRVIAAQGQSFEDV EKSLNEALVELDSYYCR NQLKPNPMKTQTCAFHL RNKQASRQLQVRWRGV DLEHCKNPKYLGVTLDR			

						ALTFKQHCLNTKLKVCG RNNILRRLTGSTWGARP KVVRTSALSLAYAAGEY ASPVWCRSAHAKQVDV ALNESARIVTGCLQPTPV NKIYHLAGIAPPAIRREQ AAYKERQKSIVSKWHPL HGMEPAAQRLKSRNSFL TSTEKESDYRTTVERWK DAERHPWMEPSEALAA GHNESWTTWKSLNRLR SGIARSKVNLSKWGYQN DSTLCDCGAIQTTTHMY NCPLSPSHCTFEELMSAQ PNAIDVALFWAKNV			
Daphne	Daphne-37_LMi	.	Locusta migratoria	TGGCGGCCTGT ATAGATGTACT AATAAGTGGCT CTGTGAAAAAC GATTATGTTGTA AATATAAAGTG TATTTTGTGCAG TGAAAATGTTCT	TAATAATAAATTAC AAATAAAAAGTGTT GTTTATTTTATAAC TATGTACACAGTAT GGTACTATAACTTT GTTACTGTCTTGAA CACATTGTAATACA TCTGTATGTTTATT	MYIFSDSHGRGIASIMQT MQMKFKVTGITKPGAPL CIITNSKLLALPNLVPLY VLLHKTDNLKQISHVSN MWQSNHQGGFQNVMF VEATSSVRTISPNMVCT GIIMAKQFCARKYDNIN SQRQEIHSPIPLSLVKEKE	cd01650	818	1089

				CAGTGGCATTA	GTCTGATATCTATT	VTSHTQNEEKEQETSLSS			
				GGCTGTGTTCA	TATGGTAGTGTACA	PKKESPLLTLKRKDNWK			
				GTGACTCAAAA	TATCTGATAACTAT	RSVQKEKKNHHPVLRK			
				GTGGGTACATA	AATGTAATTGTTTA	RTRNITVFPQKRITIAHLE			
				CGAAGTGCATA	TTATTCTTGACATT	KKGQLEKKRAEREEESS			
				GCATATTTTGAC	CATTTCATATTATTT	SCSIEEVLSLSAPEEETFS			
				ACTATAAACAC	TCTTCTATCTTATG	SAEKLPAKLRRGERTKK			
				AAAATGTAATT	GGATTTTCTAAGTC	DIPLKNQDFFIPAAGEKT			
				GTGGCGTAATT	ATTTATTGTTGCTA	TRYKLDLRKTKVVPRDQ			
				CATTGTGATTTG	TTTTGTTTTATCATT	LHKFQGISCTNSIGQNV			
				CAATTGGTAGC	ATTACTATTTTATCT	MFLHINVQSLKYKINEL			
				TTTAGCAAAAT	AACCTTGACGACAC	QFWLENLNCSILSVNEH			
				ATGATGCTGAG	CGATTGCATGTAAA	WLNCEELKLYVPFGFDL			
				CGGGAAGAAAA	AATGCTCAAAGGTG	ASSYYRTNVEHGGVSM			
				TGTGCTATTTAC	AATAAAGCATCTAA	LKIMTWWSVYVKNNVN			
				GATAGAACATG	TCTAATCTTA	LIYEVTDVSELCIEAKFE			
				ACATATTGATC		AAAIKFPKQEIIVGSVYR			
				GTTAAAAAGTA		PPKHNESLFVENLEKLIL			
				CTCAAACAACT		LLLKHKNYRIVIFGDINI			
				TGGAAACATTG		DIRFKNRNVNHLLNMLR			
				GCAATAACTCTT		SLNFYCINTKPTRQAACL			
				AACGAAGAACA		DNIICNFHLNKTDFDTIS			
				AAACACGAAAG		LCISDHDGLWLRQTLNY			

				AGCATCCATTTA		KILETSTEHRIIRPINDSK			
				TTAGGCCTAAA		LNQLISKLNIVKWDEIIT			
				TGGAGCTGGTG		NVDANEATNRFLLLLAE			
				TTCGAAACAGA		LFETCCPKHYRKSVKQR			
				ACGAGAAAGAA		QCSKPQNLKKTGILHITT			
				AAACATCTTGG		KSKKNWYTPHLNRIRIL			
				AAAAACAAATG		MLLYHDKAKHNNADKE			
				CTTTAGCAGTA		MYMKIKKLYRSEIKAAK			
				GAAGCAGTTGC		CKANDAYILNSTNKCKA			
				GCCTAAACAAG		AWEIIKTEMNYDHKVY			
				CAAAAACAAGA		QIPISPNTLNEHFVGPHS			
				AGTATTTCGAA		DSSTMQDVRKAEALLPE			
				GAAACAAGAAA		TSSNLTNNFCWSRVTEK			
				AAGCAAAGCTC		EVYESVAKLSNSRAEDY			
				CTACAAAATGA		YGLSNFVIKSVRKQIVKP			
				ATCCATCAACA		LTKIINKILEGGHYPECL			
				CAGAAAAAGAG		KISVVKPVYKKGEVTSP			
				AAGGAAGAG		ENYRPISLIPIISKIIEYCM			
						HKQMYQYFETNNILSAS			
						QYGFRSRLSTTKAVESIV			
						SKIHSSFEKKEIVCATLL			
						DLRKAFDMVSHNILIKK			
						TPALWSERECYMPDGVL			

						LGEQETVSQDKKKLQH YGVRGNAICLMESYLEN RKQLVKINNESSECLPVL KGVPQGSVLGPFLFIVYI NDLPTVVPCDAVLYAD DTTLITCDTNLENVQHN MAVAMEKCNTWFEAN VLQKNDSKTEEIVFNLT APSQDNKTVKLLGFHID QKLNWDTHLDKLCSRL ARVLHLLYKLRSSVSKE LLLYAYSAFFHSHLDYGI ILWGNSVGSKTVFKWQ KKAIRCIAGLSPRDSCRG YYKELNIMTVPSLYIYN CLVYVKENLNNYDLRM TIHDHSTRSGHTIDVPYA RLAKTHNCYKYLGLLYF NKLPKEAYTVPVDQFKN TIVKWFKSKVFYSAQEF LECGTKDLNFL			

RTE	RTE-26_LMi	.	Locusta migratoria	GTGGAGTTGCTAGTCTCCCATCGGGCACCTCCCCAGGTGGCGGATAGGGGAATGCTCACCAGATATGGTGGGTACCGGGGAAATAAAATACCCGGGGTGGACCAAAACTAGCACCTACTGCCTTGTAGTTAGGTATTGGATTACCAAAGGCTAAAGGAAGAAAACCTTGATCAAAAATTACCTGGCCCTCCAGGTTGGGGGTTAATGCA	TAATCTAATATAGGTGCTGGAAAAATGTACTGACATTGTAAAACCAGAATAATAAATAA	MGQLPIHWKKTLPVKKPIRLQQDGYNRRRRRLRKRTVLIGVWNVQGLQNKVGEVIKEMEARNQDIVVLPETKKKGHGIERIGPYTHFYSGVPKNQRAKRGISILMKNKFNRMVTNWEPIDENLIKININIQEHKITLIGVYGISEDEPIRKKDEFYNRLNGVIEDIGKTREIIIAGDFNSRTGSKIGNKIIGPFGENVVNDNGNRLIDICEQNSLKILNGFYQHKDIHKYTWHKDTLNQKSIIDYIIIKQQTRIKIHDIRAYRGVTLSDHYLVNSKFLFPFRGKNSEENQNCTTEDLQTEIKITKYNLNSFENESTVFLYQKRLDEKLIETEFDNTEDQYQYIISKLHESAKEALGETDEKHTXHSLYYWDEEIKKEVKLKKEKYLKWLNTRNVQDKIELN	cd01650	592	851

						AQQAKIRKKVAEAKNK AWEQSCQKIESYIGGKR NTEAWKTIKNLRRNSKE XTHIKYITHDSWEKYFQ NLLSENRRILGKPEGILK XQHILNILLMTLGKNIFK IYYQKTEEFLGNPEYNEI TNELPDNYIHLEIETVKS AIKSLKNGRATGIGGIPA ELLKFGTEKLIELLRNLF ERCLNGEDVPNDLKVG YISVIHKKGRKDDCKNY RGITVTNTFSRLYGRIIK YFLEQEYKEKEAEEQAG FRAGRSTIDHIFCLQQIIE KNKKLGRAQPIHLVFVD IEKAYDSVPLPNLWKAL LSIGINPTIIKAIQNLYKN SISKIKIGKFLSSGFQVTK GLRQGCSMSPTLYKIYT EVTLQNWKKKCQTMGI PINDRTIYSLQFADDQLII AQDYEDIEYMTRKLIQE			

						YKKAGLNVNMNKTKY MVIGGINEDLILEEGMGT ITTTEEYKYLGVKITNDG KQEKEIRSRINAGKTTIS LLNGILWDKHITTDNKIR IYKTIVRSIITYGSEVWTT KWQLKSKLLATEMDFW RRSARISRRERIRNEVIRE KMKCKNSIIEFIEHKHLK WYGHVRRMGQERLPKC ILDWTPNGRRKKGRPPT TWMQEVQSAMRKNNIP EDLWEDREEWRTIIXNL M			
RTE	RTE-38_LMi	.	Locusta migratoria	GCCATAGCCGG TCCCAAGCCCG GTTGAGAAAGG AGGAGGGGGAG GAGTAACACCT CGTTAAAAAAC CTTGGTTCACCT GGCCCTGGTGA TAGCACCTCGT	TAATTACTCAAGTT CAATCGCTCTCTTT AAAGGGAATAATC GAAATAATAATAAT AATA	MSTQFLKPVNKLTTWKS PNSTLGEFQIDHVAISNK NTKEILNVRTRKGFFESD HHLLQIKIKFQPNRKKPQ MNKTFRIDPEYLKLNKK EISQEIRQSNATNWKELA EVLKSTMHLGKPPRKRK HRWWNSSCDQAIEERTQ AWKTFSSNKTTEKWQEF	SSF56672	393	550

				GAGGGCCCCTG		LRIQKNTSKAIRREKRRY			
				CCAGGGTACGG		DINRLKEIEQEFLKNNTR			
				TGAAGACCTCA		NFYTIFKENLTGYQPPNL			
				ACGGCAGTGAA		CFRRPDGSLETNTRNNC			
				GGCGGAGATGA		QILANIFTHSSTASRRQS			
				AGATCATCGGT		NISKTPSNPDSKPPSLKE			
				ACGGCGGAACT		VKEIINSLNNKSPGEDGII			
				GGCGGAAGAGG		AEMWKLNDSTLTQATH			
				CAGCCCTAATC		RIISKVWETEQIPAEWKC			
				CACTAGCGTCT		ALIHPLHKKVQKLIQITT			
				GCCCTGGAATC		GEFHCSQLQGNFTAPSNI			
				CAGCGGTTAGT		QNFFKGPVKQTRRTGGP			
				GGTCAGCGTCT		VNRRVSGRVSRSILRRT			
				GTTCACCTAGTT		NMEPKDILQMRRCRNTV			
				GGTGCGGCTGC		VTFVDFKKAYDSIDRQT			
				TCCTAGTTCTTC		LFHTLEEFRIDRKTRSIIQ			
				TTGATCTCAACA		QALTGTTSKVKFMGETS			
				ATCAAGTCGTA		EAFHIHTGVRQGDGLSP			
				AACATGTGACC		LLFNLVLEKIVRTWEKEI			
				TTCGTAGAAGA		KGIQLGIKKEQRIRVKCL			
				TCACCAAAACA		AFADDLAILTNNREEAR			
				AGTTTCAACTTG		LALEKLHEISQKTGLQIS			
				AATAAGCATGA		YEKTQYMENKLQIVSPS			

				TGGAACAAAAG		QPSMEKWHKFLTSDIKD			
				GAAAGAATTCC		SVHGKQAADSLPITTQY			
				ACTACCCCAGG		GKVAQVSHFRYLGEMIQ			
				CATCAGTCTCTA		PSGLNSIANQERVSRLQR			
				GACTGAATTCTC		AYKLTWNHYNKKSISV			
				CTGGTCATCGG		NAKLRHYNTVVLPEALY			
				ATCTGGGGGAC		AAETTVIHGHTKIREIEK			
				CAGGAACTTCA		QERKILRKIYGSVFREGI			
				TGCGCAGAAGA		WIKRPTREIYKDMETITD			
				GTCGGAGCTTCT		TIRKRRMKFYGHISRMN			
				CAAAATCTCTA		ETRLTRQIFEIANKSKNK			
				AGACCAAAACA		TKWITETEEDIKELGITQ			
				GAAAAACTTCA		DNINNREQFRNIIKKHKL			
				TAGGAACTCTC		KQEHTENRTGKKWSER			
				AACACAAACAC		KREHSERMKRIWEERKK			
				ACTCTCAAAAG		KKS			
				TAGGTAAACTT					
				AAACAACTCAC					
				AAACACTGTGG					
				AAAAATTCCAC					
				ATAAAAATCCT					
				TGCGATTCAAG					
				AAACACGATTT					

				ACCGACGAAAA TCATTTTGTACA GGAAACTTCAG AATCTACAAAG GAAAACCTGMA ATCCAGTTGAA CAAAGGACCCA AACAATTTGGC ACAGCTTTTGTT GTGCATAAATC AATTACAGACT CAATAATAAAC TTCCATTCAACA TCAGAAAGAAT ATCCACAATCTC ATTAAAATCTG GAAACAAGGCA TACACACTGAT AAATGCACACG CTCCCACAAAT GATTACAATAG AAAAGACCCAC AGGTAATAGAG					

				GACTTCTGGGA ACTACTGGAGG AAACCTCAAAT AAAATTCCTGC ACATCATGTGA AAATTCTTTTAG GAGATTTCAAT GGACAACTCGG CAAAGAGAAAA AATTCAGGAAA ACAACGGGAAT TCATACAGCTC ACAAGAGGACA AATAAAAACGG CGAGAGACTGA TAAACTTCTGTG AAAATTTTGATC TTAAAATC					
Kiri	Kiri-1_LMi	.	Locusta migratoria	CTGGTTCACTTT AGGATAGTGCT GTTTTACTATTA ATCTCATACTTA CATTATTAATTA	TAGTATTAATCCTA TTAGAGTTAAGTCA ATACCTTTTATTTA CATCCTTTTACATG TAATCAACCTGTAT	MCPAFAEKDAATPQPAA VALPTSGIIARATTERTK LHVAHANVQSLQAHLV EFRHIFQDLGLHIILLSET WLKPSITTSSVNLDGYV	cd01650	510	773

				CGGTACCATATT	AAATTATGTTATTA	FLRHDRLNKRGGGVGA			
				AGGTTTAAATTC	ACTGTTTATCTGTA	YIRSELSPTVLHTSNNDE			
				AATTTTATATAG	AAAATCATTGTTAT	TEQVEYMFIEIKPLKEKC			
				CGTTGGTCTCAC	GTTATAGGTAACTC	LICVLYKPPKVGALDCF			
				GCATCATTGGC	ACATGTAAAGAAAT	ETALSDLESTYEHIIIAGD			
				AATTGCTCGCTC	ACTGTATGTGTAAA	TNINFLFNSPSCMKFKR			
				TACTATCGATAT	ACCCTGGTCCGATG	MLTSSNLTLLQLAPTHH			
				TGTTATCTTTAT	TAAGAGAGGGCCT	TADSHTLIDIFATRFPDR			
				TTAGTGTTGTTC	GATGGCCCTAATCT	VIRTSQTSAPGMSAHDVI			
				CGTGCAGATTA	GATCAGGCTAAATA	FMTYSLECPRKKPLLTT			
				CTTTTACTGTGC	AACAAATAAATAA	YRDLRRVNLPELTADAX			
				TTGGCGAGATT	ATAAATAAA	EIVFEPTSDTLPDVNEKL			
				CTCCCCCCCCCC		HKITNKLNQLYDKYAPI			
				TCAGCGCCCCG		KTKKVRRKPAPWLTDEL			
				CCCTCGGCGAA		KQLMNLRDTAHRTYKL			
				CGGCCGGCGAC		HPTPENHTAYKQKRNRV			
				CTTGGCTGCAG		SQAVRNAKLRYAHSLTD			
				CGCCTCCCTGTC		NINRPAVLWKNLRGLGI			
				CCTCCCCTGTCC		GKVKTAAEPTVPIEELN			
				TGAACCCACCT		QYFTTPTTTIDHQTKQTL			
				GACCTGCTTCGC		IDHITRINNANHDKFCFK			
				TCTCTCTCTCTG		HVTGSTVKKAIMRITSAS			
				CTGTCCTGCACT		TGDDGITIQMIKLIIDPLL			

				CCTCCCCCGACC CCCCGGACCGC TAGCCACCCAG CCACGCCACAT TCCA		PIITDIFNQSLATSVFPDA WKRGLVKPLPKKDIATA PSDYRPICILPALSKALE YIVHDQLTSYLTINNLLD EHQSGFRKYRSTTSALIK VTDDLKQAMDRQEATI MCFLDFSKAFDTVDFDI LLAKLSSLNLSQGVVQW FRSYLTSRLQCVMSGTE KSQWRKVLSGVPQGSV LGPLLFSLYVNDVSSVLS FCKYHMYADDLQLYLS AKPKNLKTAIENFNSDL CALSKWAQNIGLKLNPS KTQAILVGHPRLINHKY RESQPSLILNDTPIYFSPS AKSLGVIIDDKLNWAEH TAAVCRKASASLHALQR YKKFFPLDLKKKLIQTL VIPIIDYSDVILQGLSWES SRRLELIMNACVRYICNI RLFDHISPSYKQLSWLR ADKRRDFHTLCFLYSLIN			

						VHCPSYLSSTLTLLSDQH NRTTRSCQSKTLSVPIHH SAGFSKSFSVAGTRLWN NLPQYIRELNNTASFKR QLMDYLLKQLRQSLIPY GPGTSHPFLLPDILLLPLL SPGTAFLTSFCQNSPCLN QSRYIQKYILYLLISMLL LLLLHQQQQKVSQWYC			
CR1	CR1-1_LMi	.	Locusta migratoria	TAGTTATTCTCG CATATTTGATAA ATACTGTTTGAT AGCCTTGTCTGT GGCTGGTTTTCA AAACAAGTCAG AGACCTTACTA CAGTCGTTATCA GTCAGTCCAGT GCCAGCTGTTG CCAGTGTCACA TCGTAAACAGC GCTAAGTATCT AATTTAATTTGT	TAGTGGTACTCAAG GTACCCTCCGCCAC ACACCGTCGGGTGG CTTGCGGAGTATGG TGTAGATGTAGAGA TG	MVIGCFYRPPGSAAVVA EHLKENLENISSRFPDHV IVLGGDFNLPDIDWETQ TFMTGGRDKESSEIVLSA LSENYLEQLNREPTRGD NILDLLVTNRPELFETVN AEQGISDHKAVTASIISA VNRNIKKCRKIFLFGKSD KKRISEYLTAQHRSFISS TDSVEEQWSKFKAIVQY ALDEYVPSKIVRDGKEP PWYNNRIRKLLRKQREL HCKHKHSQSLVDKQKL NEAKCSVRRAMREAFN	cd01650	388	636

				GTCAAATTTAAT		EFESKVLCTDLAENPKK			
				TTGCCAGTCTAG		FWSYVKAVGGSKQNXQ			
				TTCTTCATTTTC		TLCDENGIETEDDTLKA			
				GTTAGTACTTCT		ELLNVFFQSCFTEEDCTV			
				GTTTTAAGAGG		VPSLDCRTDYKMADIEII			
				TTTATTCTCGCG		DRGIEKQLKSLKRGKAA			
				TACTTGGCAGA		GPDGIPVRFYTEYAKEL			
				TTTCGTTTGTGC		APLLAVVYRRSLEERSV			
				TACTGTTTTAAT		PKDWKMAQVIPVFKKG			
				AGTTTATTCTCG		RRTDVQNYRPISLTSVSC			
				CGTATTTGATAA		RILEHVLCSSIMTFLETR			
				GTACTGTTGGG		NLLCRNQHGFRKRRSCE			
				CGCGTCAGCAG		TQLALFVHETQRAIDTG			
				CACACAGTCAG		SQVDAVFLDFRKAFDTV			
				CAGCACCAGCA		PHRRLMNKVRAYGLSD			
				GAGTTACAGCT		QLCDWIEEFLDNRMQH			
				AAGCAGGAGCG		VILNGEKSSEVRVMSGV			
				AACTATTGTGA		PQGSVVGPLLFTIYINDL			
				CTTATTTGTGGG		ADNIGSSLRLFADDAVV			
				CTGAGCAGTGT		YREVVTMENCIEMQEDL			
				CAATTTCTATAA		HRIDAWCREWQLNLNV			
				ACTCGCGCTAG		DKCNVLRIHRKKDPLSF			
				TAACAAGATAC		SYTIAGQQLEAVNSINYL			

				TTAACTCTCTAC AGGCAATTGTT GATTTGTTAGG ATGGATAGGAT GTGTGACTGCT GCGTACGGACG CAGGAGGAGCT GGCCGCTGTTC GCGAGCAGCTG AGCGTGCTGTT GGCCGCGGTCA GCCGTCTTCAG GCTGCTGCCTCG GAGTGTTGCGG CAGTGGGGGCT CTGGTGCGTCG CATGGGACACC ACAGGTGTTAC ATGCTTCACCCA CGGTCCCTGCTG TCGGGACATCTT CGCGCGTGCCG GACGCGGTTGA		GVGIRSDLKWNDHIKSI VGKADARLRFIGRILRK CSPKTKEVGYSTLVRPQ LEYCSPVWDPYQVGLIE EVEKIQRRAARFVTGSF GNRESVTEMIDKLQWKT LQERRSVARYGLLSKFR EHTFTEESSSILLPPTYIS RRDHEDKIREIRAHTEA YRQSFFPRTIRDWNRRE NR			

				GCCACCCTCTCC					
				TCAAGGGGTGT					
				GGCGGGTTCAA					
				CGGCGTTCGCG					
				TCGCTCGGGGC					
				GAAAGGTCAAT					
				GTGGAGGCTGG					
				CCGTGCGGCGT					
				CGCCCGCTCTGC					
				CTGTGAGTGGA					
				CATGTGGCCGC					
				TCCTTCAGCAG					
				GGTCCGAGCAG					
				GCACACGGGGG					
				GAGGGGTWTAT					
				TAGTGATTGGG					
				AGCTCCAATGTT					
				AGGCGGGTGAT					
				GGAGCCCCTTA					
				GGGAAATAGCG					
				GAAAGGTCGGG					
				GAAGAAGGCCA					
				GTGTGCACTCTG					

				TTTGCCGGGGG					
				GTCTCATCCGA					
				GATGTGGAGGA					
				GGCCCTTCCGG					
				CGGCGACAGAG					
				AGCACTGGGTG					
				CACCCGACTGC					
				AAATTGTTGCTC					
				ATGTCGGCACC					
				AATGACTCCTG					
				CCGGCTGGGTT					
				CAGAGGTCTCC					
				TCAGTTCATACA					
				GGCGGTTGGCG					
				GAGTTGGTGAA					
				GGCTGAAAGCG					
				TCGCTCGCGGG					
				GTGGAATCTGA					
				GCTAACTATTTG					
				TAGTGTCGTTCC					
				CAGAACCGATC					
				GCGGTCCTCTG					
				GTTTGGAGCCG					

				AGTGGAAGGCT					
				TAAACCAGAGG					
				CTCAGACGATT					
				CTGCGGAGATC					
				TAGGCTGCAAA					
				TTTCTCGACCTC					
				CGCTATCGGGT					
				GGAGAAATGTA					
				GGGTTCCCCTG					
				AATAGGTCAGG					
				CGTGCACTACA					
				CGCAGGAAGCG					
				GCTACAAGGGT					
				AGCGGAGTACG					
				TGTGGAGTGCA					
				CATGGGGGTTTT					
				TTAGGTTAGAG					
				AATTCCCTCCCT					
				AGGCCCGACAA					
				GACGCCTCCTG					
				AGCCGCGACAA					
				GGTAGCAGTAG					
				GCAGAACACAA					

				CAGGGAATGAA AAGATCAATGT AGTAGTAGTTA ACTGCAGGAGC GTCTATAGAAA GGTCCCAGAAC TGCTCTCATTAA TAAACGGTCAC AATGCCCACAT AGTACTAGGGA CGGAAAGTTGG CTGAAACCAGA TGTAAACAGTA ATGAGATTCTA AACTCAGACTG GAATGTWTACC GCAGAGATAGG CTGSACAGTGA AGGGGGAGGCG TGTTTATAGCGA TAAAAAGTGCA CTAGTATCGAA GGAAATAGACG					

				GAGATCCGAAG TGTGAAATAAT TTGGGTGAAGG TCACGGTTAAA GCAGGCTCGAA C					
Daphne	Daphne-46_LMi	.	Locusta migratoria	AGTTTTTCTGG MTGGATCAAAT GATGTAGCGAS AAATGAAGGAA AAGACTTGTTGT TATCATTAAGG AAGAAGCTGCA GGAACTACAAA ATACAAAAATC CTCATCTTTTCA ACACCATTTCG GCATGATCTTA AATCCTGGTCAT GTGTGAATGTCT AAGTAACGAGA GTAAATGAAGA AAATTTGTAAA	TAATTCTGTATATA TGCTGTATATTTCT ATACATTTGTAACT TGACTTTGTCAATT GCTGTAACAGCTAA ACAATAAAATTTCA TTTATTTATTTA	MVSLSNYDWSHIFNNNA HDSANTFFDRFFHTFLNI FETNCPQYKCKVNPKSN SNKHRDKNPWYTAQLA NMKSRLLLYRDSYRLQK TDMAKQRYSRAQKDYK HAVNEAKKQHNLVSIES SKNKCKKAWTIINSVSK SKQKVDVVISADEFNNF CIKSVNQISENIVRPHET VADLIKCHNVNYTLKDK FLLSEVTPDVVLTIVNNF RPSDSVDIYGISCNMLKK IIGYIVYPLTKCINKCLIE GVFPDVLKLSRVVPIYK KGDRNCPSSYRPISLTPIL SKFVLESIIYYQLCDYLT	cd01650	252	527

				CACGTTAAATT		CNNIISAVQFGFRKGKST			
				GTTGCTTTTGTA		VHAIDSLIKKVLNAYEE			
				AACATAATTAA		KKFAQATFCDLSKAFDC			
				CTTAGGTGAAA		VEHISLLKKLVYYGIINS			
				TATTTCATACAG		EVDNIFESFLSNRKQIVCI			
				STCATKGCTTCC		GKQRSQLAEVKCGVPQ			
				ATCTAATCACTG		GSVLGPLLFILMINDLPS			
				GAAAGAAGATG		YINTHSFSMPMIPLFSISA			
				TGAGTAATGAA		QTLICFKLFTILYKYSLIL			
				TCATAAATCTA		YADDTTFFNISSDFDML			
				GCCATGAGTTA		QTVTNDTLSQASVWFRA			
				TTATACACAAA		NGFLLNESKTQKIVFSLR			
				TCGATAGGTCA		DLPVEDFIDSVKFLGIVI			
				GATCAATACCA		DSKLSWEPHIKYIXTRLS			
				CTTAGATTTAAG		RVVYLLRNLKNHVPAA			
				GGAAATTCTTTT		YVRSAYFAFFQSIISYGL			
				TTAGAGGTAGA		LIWGNSSHVQDILVLQK			
				AAAAGGGCTGA		KVIRIITNSDKLAHCKPL			
				GCCAACTGGTTT		FINLKILTVINLYIYTALL			
				GCCCAGATTCA		HIKSNLSEYQRRGDVHS			
				CACTGGAAGAC		HGTRNSSHLDIPYYRLSK			
				CAGATTAAAGA		SLNSYEVVGMKMFNKL			
				TAAGAATAATC		PQEWVEAPLDLFKHKLY			

				TAATAAATGTTT GTAATATTAGA GAGAAATATAG CAGTAATGATG ATTGCCCTCATA ACAACAATTAT TCATCAACTTTG TCATTTAAGATA GGCTATCTCAAT ATTGAAGGTAT GAGTGGAAAAC ACGCAATTTTA AACGATTTTGC AAATGAGAACT TATTTGATGCAT TATGCTTAAGTG AGTATGGGTGT AAAGGAAATGA AATATCTTTTCA TGTATTAGATG ATTTCCATTTAG TTAACAGTTTTA GCAGAGAGCAG		CWLAANPCYSLNEFYEC KISWYLLGECTIISLMES FMMQCFKLYYFKYLIL WKTYSDLLIFALCKVICI YALPYVK			

				CATATTCCTGGG GGGTGTGTCAA TTTACATTAAGA AAGCATTAATA AATTGTAGTAG TGAACTAGATC TAGGGTTCTTGT GCAGTGAGATA CATTTTGAAGCC ACAGGTGTATG TATGGATMATT TAAAATTGTATT GTAGCTCTTTAT AGATCACCTAG TGGTGATCCACT AATATTTTTAAG AAAAACTTGAG AATTTATTCAAT TTCTTTCTCAGA CCTCAGTGGAA GAAATATAGTT AGTGATTGGTG GTGACATCAAG					

				CTGTATTTGATG TAAATAAAGAT GCACACACTGT AAATATTTAAA ACCGTTACGAC AATTCAACTTCA ATTTTACAAACT GCAACAGTCTG CTTGTCTTGACA ACTACTTACAA ATGTAATAGAG AGTTACTGTCCT CAGATGTAGCT GTTTTTGATTTC TCGGACCATGA CCCAGTTTGGGT AAAAATAAGTA CCGATAGGCCT AATATGCAGTA TAAGGAGTTCA AAGAGCCTAAA TAGTCATCACA AGACCAATAAC					

				ACAGGAAAAAC TTTCCTAATTC					
Nim b	Nimb - 41_L Mi	.	Locusta migrator ia	CATGGGCTCAG CATTAATTGCTG AGTATGGGCTT GGCGAACCCGG GGTCCTGAGCT GTGGGCTGTGT CGCGCCAGCAG TCTGCTACACCG TAAGCTCTGTGC GTGCTTCAACG ACCATCGTTCG GCGCGACGGTG GAATGTTGTAC GTTTCAGAGCC CGGGGATCTTG GCTTAATCGCCC AGGTCGCGAGG ACAGTASAAAC CTCTATTAAAA AAAACCTCAAT CTCAAGGTGTG	TGACTGGCTCATCC ATTATTATCCTAGT GATCAGCCAGCCAC ATCTTCTGCTGTGA TTTTTAGCACAGTT GACTATTTGGTCCA CTACTTTTGTGGTTT CCCTTTTCCTTTTAT ATGTCATTTAGCAC ACACTGCCTTCCTG TACTTCTCGTGTAC TGGCTGTGGAGTGC TTCGTCTTCAACTG TAAATCACATGTGT GATCTCTTTGCTGG GAACAGGGGACTG ATAACCCAGCCGTT TAGTCCCCTTTAAA AACAACAACA	MAPTVQPSSVGPKGHQT HSASLKVGGPSFIAPPVP WWGGNTSPPPSGTQVPI PQPEKQQPSPAPLAREEK SVGMLPSPVSKSRLEDPS RSLGDAHTTGRRASRSS SAPETSEKPSQWSKPKE QREDKQKKKRSKKKSLP TALPPSVSQTSDSEDGLE MLASPEDLALTDLPESM DIDATSTKAAAKGDTEA GINSMFMTLMASQYANS NILQWNCGGYFHRLAEL RHLLCQFPVLCVALQET WFPVQRIPALRGYRGYY KNRAAYDRVSGGVCIY VLANVPSAPVPLHTALE AVAVRIRAGQEITICSLY LPPDKAVSSAELAALVA QLPPLYLLLGDFNAHNP LWGGTEITARGRDVEYL	cd0165 0	849	1106

				CTACGTGCCGA		LAQLNICLLNNGAATHF			
				GGAGATGCATG		SLAHGTYSAIDLSLAVQ			
				GCTGTTGAGGTT		AYFHTLLGGSMTTFVEA			
				GAACAGTCGCT		TISLSCRSHSHHLQAACP			
				AGGAGAACTCT		DGHSVGQTGRLTPRLLQ			
				GAGGAACCTGC		QRSRHTLQLGPWHLLSH			
				CGCACCTCAGTT		RSVPCSPSLLPHLTWRV			
				GTATAAGGCTT		HDDLCGSDHFPLLSLPQS			
				ACCCAGGCAAG		PPAGCLPRWSLRRANWP			
				CGGGGCTCTGC		AYTSAATVTSPLPDDVD			
				CTGGGTGGACC		LAVDDVTRAIVSAAEAT			
				TTTATATCCCTA		VPKSSGLLHRKSVPWWT			
				GCTGCTCGTGG		PEIAEAIRERRRALQRHR			
				GACCAAGATGA		RHPSLEHLIAFKRLRAQA			
				AACGAGCTATA		RRLIQQSKRDSWERYAA			
				AACTATGCCAA		TIGSRASPTZVWKQVRR			
				TGGCGTCTCTAA		LFGHRPLAGIPSLHIRGT			
				GCGAGGAAGAT		LSTDPAAIAQYMADIML			
				CAAATTTCAAT		VARHLHLTVRTSAPSKS			
				GCCACGAAATA		VLNSGSFPFTCAARKNT			
				TGACCCCCGAT		MRHSVSGSLRVPAPFRA			
				CGTTCCCTTCCC		SSTGRNPQSSHSGYSVH			
				TTGGTACATCAC		RPSCYRAVHGRHYARSS			

				GGGAGGAACAT		ASASYSADFRALKERVE			
				CGGGCTAAAGA		QRQLSFHMRRQEKYNA			
				TAGGGATGAGT		PFSEWEFESALAACPDTS			
				TATACTCTCCAA		PGPDSVHSQFVRHLPAX			
				AATATCTCATAT		FQQLLLALYNHLWVKG			
				GCAGCAGAACA		AFPSQWRDSIIIPVMKPG			
				GACGGGGAMAG		ADPQVAESYRPISLTNVL			
				TTTCATGACGAC		CKIFEHMVSRRLMWILE			
				AAAACCTCTTTT		DRGLLAPQQGGFRHGRS			
				CTTCGTAGAAA		AIDNLVSLESAIRTAFTR			
				ATCTGGAGGAT		REHLVAVFFDLRKAYDT			
				AAGTATGGTGA		TWRHHILATLYDWGLR			
				AGTCGCGGGAA		GPLPTFIKNFLSLRTFRV			
				TGTCAAAGATG		RVGTSLSTHHIQENGVP			
				AGGAGTGGTTC		QGSVLSVPLFLVAINGLA			
				AGTGCTGATAG		AAVGSSVCSSLFADDFC			
				AAGCTGCATCA		LYYSSASIGVTERRLQG			
				TCCACTCAGTCG		AIRKAQSWASAHGFTFS			
				AAAGCATTGCT		ASKTRVMHFCRRVVHP			
				TGCTTGCGACA		DPELYLADEQLHVAETH			
				AGCTGGGTGAT		RFLGMLFDNRLTWGPHL			
				GTCCCTGTAAG		RQLKQKCWRHLNILRCL			
				CATAACACCCC		GHTAWGAGRQTLMHLF			

				ACAAATCACTG		KALILSRLDYGSQVYDS			
				AACATGGTCCA		ASPSALRILDPILHGGVR			
				GGGTCTGATTTA		LAMGAFRTSPVPSLLAE			
				CCACAGAGACC		AGIPPLRFRRQQLLASYV			
				TTCTTCTACAGT		AHVHATRLTIQIAGFCFQ			
				CTGATGACGAG		TPSFPLHNIDPDQVFQSR			
				TTACGCACCAA		SGYVLSFPHSSPYLYLPH			
				CTTAGAGCGAC		SPANLISLLGPSLAYSST			
				GGGGCATCCAC		CNCPAIQRRRFHHIHFIN			
				TTCGTCCGGCGC		SFWTSLARSVMQKSYTL			
				GTGTATAGGGG		MAPRSWHPSPEVSPATV			
				TCTGAAGGATT		VGFICRPRSCYTPHHPNR			
				GTAGGGTTGCC		RLLFPNTVLPSPQHRPRS			
				ACCGGTGCTTTC		GLPIAVRVRTLLSSLESL			
				ATCTTGGCCTTT		PLPPAFSCQSHIPPWSIPR			
				GAGGGCGATAC		LQLHLQLSCDPKASVSP			
				ATTGCCCGAAA		HTLHQQFLDILGTFRDA			
				AGGTCAAGGTC		EVIYTDGSQVEGRVGYA			
				ATGATTTACCG		YVHGDHIEQCSLLPGCT			
				GTGTGACGTAA		VFTAELIAIIRALDHLRT			
				AGCCTTACATCC		CPRKSYVICSDSLSGLQA			
				CTCCCCCTATGT		LDQCFPKHHLVRSIQDS			
				GATGCTTCAAA		VFSLGXSGRSVAFIWTP			

				TGTTGGAAATTT		GHIGIPGNERADQLAKQ			
				GGGCATATGTC		AIASQRLDIGTPERDIRL			
				ATCCCGCTGCA		GLHHKISMAWNSEWST			
				MCTCTAGCCCC		LSTPNKLCAIKPSTAAW			
				ACATGTAGGGA		KTSLRPSRRDSVVLCCL			
				ATGCGGGTGTC		RIGHTWLTHGHLLHRED			
				CATCCCATCCG		PPNCRCGALLTVVHILL			
				GACGCTCCGTG		DCPLLAALRQSFNLPSAL			
				TACTCCCCCGCC		PSLLADDASMVDRVLGF			
				AATATGTGTCA		IRAAGFYHMFNGGYLTF			
				ACTGTGGGCAG		RRPPLNSFNLLFLLSCQS			
				CACCATCCACC		YFQLAALVCTG			
				CTGTTCACCAG					
				ACTGTACCATTC					
				TCCGCAAAGAG					
				TGTAAAATCAT					
				GGAAATTAAGA					
				CCCTGGACAGA					
				CTGACATACAC					
				TGAGGCAAGGA					
				AAAAATATGAA					
				AGACTCCACCC					
				AGTCCGCAGAC					

				AACTTCCTATGC GACAGTTACGT CGTCGCCAGCC ACCCTGGCGAC G					
RTE	RTE-40_LMi	.	Locusta migratoria	CCCAGGGGGCT CCAACTAGGAG GCGCCCGCGGG ACCCCGAGCTG AGTCTGACATT GCTTCCACTTAC TTGTGCAGGCC CCTGCTTTTATC TTTCCTGTCGGC CTCCCTTGGCCA ACTCTTGTTTCT TCGACCCCAAC AGTATTAGGTTT CGAGGCTGGAG GGTGTCTTTCAT TTTCCAAACCCT TCTTCTACAACT GGAGAGTAAAC	TAAAGTAAAGTAA GTA	QWNKVRQLFHFKGLYIS STGFERKNIYKGTWISPD GRTVNQIDHVAIEDKLK KWITNVKTARGQSVVPI TFCSKNSKGAECGSDHF LLKIDLRVGWKPKLTKK LQKVVHFDYNKLVTDE GLQDNFRINLIVLKPWQ HCYRNPGNIATDSLDAG SLWEGIRDLTETAKEVIP KRMNRKKVWFDXECEL AILERQRAKQVWIQSAR GRYEEGKERFDEARRRT QATLRRVKRNYFKNIMV EAQKDFHGNKSKELFRK VKAIKGEGKKHGFIKDR NGRVLTNNNDIAIRWME YFDELLNCEEPXETLEFE	cd01650	394	654

				TCCAAAAACAA		EPVTVDPEYPLQHSKKG			
				AAAAATCCTGG		TCNCGSGIPPPTFEEIKK			
				TCCTCCAGGTTG		NIKELKNGKSAGEDGIV			
				GGGGTTGGCAA		AEMLKSGGDSCIQNIHK			
				CAGGCTACCAA		LVRKIWQEEVIPEDWKV			
				CCTGTTCCAGA		ALICPIHKKNDKMLCDN			
				AAACTAACAAG		YRGIALLNISYKVYSKC			
				GTAAAAACGAT		VLLSRIKPLAEELIGEYQ			
				ATAACAAGGCC		GGFQPGKSTTDHIFFIKQ			
				TCGGATAGGAC		VCEKMYEYDKEIHLLFV			
				GGATGAATGGA		DFQKAYDSIHRTSLFKA			
				AATCGACCCAA		LIEFGVSKKLVRLVSLCL			
				GCAAGTGAACG		DDTRAKIKLGNRTTDSF			
				GACAGAATACT		RVTTGLRQGDSLSPILFN			
				GGACTACGGAT		LVLEKIIREMDSFNHQGI			
				TGGAACCTGGA		QVANRNITHLAYADDIV			
				ATGTTAGAAGC		LMTHTKENLKCNANDD			
				TTGCATCAAGC		HGKHSKQIWTESEPHKN			
				AGGAGCAATAC		QIHDSWKSPTPRNPGINC			
				GGAGTATGACA		GNNTYKRVKDFKYLGSI			
				AGGGAAGTAGA		ISENSSSGMEIKARIQAG			
				AAGGTACAAAA		SRCLHSIHNILTSRSISRA			
				TGGATGTAGTA		LKLRLYKTLILPVILYGC			

				GCTCTACAAGA AATTAGATGGC CGAACTGCTCA ACCGATGTGAA AACTCAACATT CTGTATGGAGTT TCAGAAGAAAC CATACACTGGG GACAGGTTTTA AGTAAGGAAAG CCTTATGTAACC AGTACTAGACT TCAGTCAGTAA ACTCAAGAATC TCTTTTTTGACT ATAGTACTGGG GGACACAAGCT TGTCCTGATCAA CTGAATTCGAA GGACGCCTTTTA TGATGACTGAG ACTGCAAGTGA CAGTGTCCAAA		QCWSTRKQELNRLAVFE RRCLRKIFGPVKDNITGE WRRRHNIELEQLYSEPNI VGLMRSRRLEWAGHVV RMSEERWPRVAMEMIPT GKRPLGRPRKRWMDGI REDLQKLEAWKNWEQT AKNRNEWRALIISSRGPL GLIA			

				GGATCACTTAA GGTGCTGTTAG GTGACTTAAAT GCCAAAATTGG AAAAGAACAAA TTTTTTTGTGGA CTCCATGATGTA AGCAATGA					
Daphne	Daphne-51_LMi	.	Locusta migratoria		TAATTTTGTACATA ATGTTAATACATTC TAGTGAATTTGTGA AAAGTGACATTTGA TTGTTATATACATG TCTTACACATGACA TAACTTGTATTAAA TTAAATATGTATAT GTTTGCTCTTGCAC ATTAGGAGATAAGT GAAAATGAATATAT AATGCTATTGACGT ATCCTATATTACAC GTACATGTATTGTA CACAATGTAATCCT	MAPIEKCSGTVLSSKRQ KISSSSCKECEKRVMKGI MCVKCNYWLHEKCAKI NLKYISDDFPWTCHACL SVETADLVSLVESKNEII KLLQNDIEALKTEILCLK EENARLLQEKQCTVGEI HQEKQQDCANTCAREIQ QTKKETRENTYARNTDV VNENISSAVDVTVNPTC KKPADKYKWKLVTYSK RKEKTTARPTQEDENEF LIIGDSLLKNVAVPNSKI DVRPGIRTQQLGKHFKN LENAKSAEPDCEPKHNY	cd01650	933	1188

					ACAGGATCAAATA AAAATTCAATTCAA TTCAA	RGVFVHVGTNSLRSSSE EEIINDTRNVIRSARSVY ARSKLVISGIIKRRSVSD KYISQINCAIKEQCDSLG AIFIDPNKFLCENCLGKD GLHLNRLGXVTLSRMFT DVCKIIRNEGNGKLRWT DGGDKSDILAEKANFKH HSKVKEIQNRIPEQMPQ HLSQQXENYXKVLHQNI QSIGNKKLEVEILLSELQ PDILCISEHGLTENQIDNL KISDYTLASYFCRSKQKS GGVSIYIKTGKLETANN KGVKLPTVKEKDFEITG VMIEGLSVFCVYRSPSG DISIFLKKICMNMKRNIII CGDINIDMGKGSKIRQD FEELLLQYHMEATISTPT RVTSHSQSVIDNIITNISR SDYESHVIQTAITDHDG QMFYLCRRHGRACETK AKIKEIRVMSEQNINTFR			

						TILSRETWTETLQAQTV NEKYDAFIATIKFHFEVA FPKVKRQIKVQQQQQWI TREVLEQRRKLQELRQV GDVEGYKTLQKTYRKTI RKAKANSINITILNSKNK IKAAWNVINTERKENAG PNKETEIRSMKIDNRIVT DGSKIANILNDNFINVVE NLKLERKSPKKKGTKVT KYSYSTMFLRPVTEDEL RKTIQKLKAKKSAGDDE ISNHVIQLVSEQHTHTIS RYSKQLFQRWNFSREIE DKVSNTLTPLVDIANSSF RDGIFPEKLKTSKVVPV YKKGNKDDPNNYRPVS LISGFSKILEMLMYNRLV DYLDKHNILVPSQHGFR MGKSTESAIASLTEYVIQ CLDEKRVVAAMFLDLS KAFDTIEHDILIKKLENY GIRGLSGEWIESYLCNRK			

						QYVSIGNSFRSETKTIKY GVPQGSVIGPLLFNLYV NDIGGEEKEKKILYADD MTVLNHGQNMQQLERS AFISANTTAQWLTENEL VINLKKTMFTVFRNKEK SVQMNLKVNDMYLEEV ESTKFLGVIVDNELKWK EHINYVCKKLSSVIFIMK QLSQYADKQLLCTVYH GLFEPYLQYGVTVWGN SNIQETKKIVHFTKESNQ DHTKKKELGDLQCGET QTSKKRKKLFTLQKKAI RIILRRKNSETCRNCFKS LGILTFTSLYIYKTIMYIT KEISLYDTNASVHTHNT RGKNEIRNIPHRLTILEK GPRYSGIKFLRNLPKNLR DNIHGNNFSRKLKDYLM EKEFYSIEEYLLN			

Vingi	Vingi-28_LMi	.	Locusta migratoria	CGTACCCTGGGCGTTAGGAGAAATCCTWACGGCAGATCCTTCCATCACMAGGACCAAAATCAAGACGTATCTCAWCCACAACGGCTTTTCTCAAAGCMAKMTCTCGMAGKAATATCACTGAATAAAGAAGTSMTACTMTCCGACAGCTGCAGTAACACAGCTTGTKATATTTTACTGCTACAGGAGACCCACAGAGGGCCACAACAGCAAAGGCCAAAAGTAAATGGA	TAGCAGAAACTAGCAGCGTCCTACAGAATTAGTCACCTCTTTCCATATTTTATGTTTCTAGTTTGTAATTTGTGTATTTTAATTGTATATTTATTGAATTTTGCATTCTTGGATTGTAATTTGCTCAATGTAATTGTACATTGTAAACATATATTTAGTGAATTTGGTGTCCCTAGATTGTAAATTGTTTATTATAATTGTATATTTATTGATCGAAAGATATGTCTTAGTCCTGCACACGTAAATAAATAAAAA	MDLVLERPHTKYGSAIFVKSGLSIQSTGLTSENDVEILTVGFQSCTVTSVYKPPNAEFKFVEPLNFQSQPIKVVAGDFNCHSTSWGYREXDKNGEDLEAWAEQMGLVLIHDPKLPSSFNSGRWRGQYNPDNIFVSNHIAQQAKKGVLGPLPHTQHRPVTCTVAAVVKPETVPFKRRFNFKRAQWEKFSIELDKEVIKLEPSPEMYDSFIQLVRTISRRNIPRGCRTSYIPGLDKDSKSLLQKYEHLFETDPFSEETTQAGEELLNAVSATRXQKWCDMVTQLDMKQNSRKAWKLLKRLNSDPTSAPPSLVNVTPDQIAHQLLLNGKTKRRTAAKKFQRYTDEEDYLGTPFTIEELEVVFKSMKNNKAAGLDDLRPEQIKMFGPVTKQWILNLLNNC	cd01650	408	681

						ISKLQIPKIWRKAKVVAL LKPGKEPTDPKHFRPVSL LCQLYKVLERLILNRISE HVDKNLINEQSGFRPGK SCCGQVLNLTQYIEDGY ERGQVTGVAFIDLSAAY DTVNHKKLMRKVYSVT KDYRLTQFIQCLLQNRR FFCNSAHSSSSVCCKIGG FFVTLHNKKSRWRTLKN GLPQGSVLAPLLYNIYT NDQPISSEVRTFIYADDT ATAAQGKTFEEVEMKLT AALEDLGKYYDDNHLK PNPGKTQVCAFHLRNKE ARKKLKIXWRGQQLLH YDTPKYLGVKLDRSLTF KDHCQETKMKVSARNSI IRKLTGSTWGSQPSVLRT SALALCFSAAEYAAPVW RNSTHAREVDTALNETG RIVTGCLRPTPVRKIYEI MGIAPPDIRRRTAAETEK			

						KKQETDPRHPLFGHEPR KPRLKSRRSFMSTTTSTA LRPGTRRTELWQNSSNE RAGENIKEEPAEGFHLPY ATWKALNRLRTGVSRC KTNLRKWGFTNEDELCE CGDVQDPQHLLVCRNLP NPCTIEDLHAANDVAIK TAEFWSLRSI			
Nimb	Nimb-40_LMi	.	Locusta migratoria	CCCCAGGGGGC CCACAGCTCTTT TGTGGGTACGT GTGTGGCGCGC ACGGGTCCCCG AGCCATTGCAG CCTTTCTTCTTA CATTCCTGGGCT GCATTACCTCCC CCAGTTCCTCTC CTCCTCCCCATC CTAATTCCTCCC CGTCCCTCTCAT CTCCCCCCCTCT	TAAGATTTTATTGC TGACGCTTGGACAG GATGGGTGGGTTTG TGGCCCTCCCTGCG AGCAGCAGGGTTTG GAGCTCTCCCCTCC CCTCCCCCTGACGA CTTCCATTTCTCCTT TTAACTTGGTTTTA ATTCACGTTGTTTC CCCCTTTTTACATTT TTAGGCTTCTCATT TAGGAGCATACACT GTTCTCCTGTATCT	MKRVKLDNGRDTAVSE GRKKNNAPSYDKRTFPS LSTPWEEGQARRLGVKP FPRYLVCTRTDGGTFTA TKPLFFVEHIEDKFGEVE TMSKMRTGSLLIKTVSA AQSAALCACDHLGDVP VSITPHQSLNMVQGVIFH RDLLLQTDEELRANLER RGVHFVRRVQRGPKDH RVDTGAFILAFEGDTLPE KVKVMIYHCCDVKPYV PPPMRCFRCMRFGHMSS RCTAVPVCGECGRPLHE	cd01650	1012	1291

				CTCGGTGGTTTC	GCTTGTCATTAATT	GNPCVPPPVCVNCHEHH			
				CTTTAAAACGG	AACGGGGATTCTAG	SSRSPECPVYKKEKKIQE			
				CACGGCTATTCC	ATTTGTCGCGACGT	YKSLDRLTYQEARRKFE			
				TATGGTTCCGAT	TCCTTGCAGAATCC	RLHPVQLSTSYASVTSFT			
				GATGGTTGTAG	CGCTTTTTCTATTG	PPPSSSHSHPILPPPPPAPP			
				CAATTTTGATCT	GTTTTAGTCCTGTT	LPPSQRPPSSSQPEKCSPS			
				TCGTCGTTTTCC	CTACTTGATTTTAA	LALAGEGAPSQGSSSRR			
				TCTGTTTTCGGC	CCCCCTTTTAGTTC	PPDRRSSASSRQPDIQAD			
				ATAGTTCGGTCC	CTCTATTTTACTCTC	NTKVARSSSVPDLTDSSS			
				CCCTCGGGGTTT	GTCACGGTCTTAGC	LSGTCPSRPAREKEKEK			
				GACCTCCACAA	CGACCACACCCTGG	RKKKQKSQDKAPPVVLE			
				AAATTCACAGT	CTTGCACTCCATAC	APPLSSQPNNPTELVSDP			
				TACTAGTGTGA	AGAGGTCATGGGA	LFMDVVPPSSVTDVDPA			
				GCCATTTGGGG	CTGATGACCTAGCT	LPSFVTCISPFAFHMDPR			
				AAGAACTCCCG	GTTTGGTCCCATCC	SVLIQWNCNGYYRRLLE			
				ACCTAGCCCCT	ACCCATCAACCAAC	LKHLIASYSAACVVLQE			
				ATGGCGTGGAT	CAACCAAC	SHFTDGHSPTLRGFRAFC			
				CCCAATCCTTTC		RNRVRPLQASGGVCTLV			
				CTTTCCCACTTC		RTDVVSSWIPLRTTLEV			
				CTTCCTTGCCAT		VAVRVHLHTPLTICNLY			
				AGCCCCCTCCCC		LPPDRPPTSAALTALLQQ			
				CGCTAGGTCCC		LPPPFLILGDFNAHHLSW			
				CAGCACGTGTA		GSAYTTSRGRLVDQFIA			

				GCCAGTCCGTG		DHDLCLLNDGSPTHFSA			
				TGGGGGGGGGG		AHGTFSSIDLSLTSPPPPF			
				GGGGGCTGTTA		SLSFAGHRMTTFATVTT			
				TGTACCCTTATG		FRCRSRHLFLHRPFTHFS			
				GCTGAGCCCCC		PPPLFPFFRWSPHDDLCD			
				TGACAACGCAG		SDHFPLILSLPSHWTRDR			
				GGATCACACTG		IPRWVFRHADWPLYTSQ			
				CTGGTACCTGA		VVFTPSPSDSIDEVISDVS			
				GCTGCAACCTC		NAITRAASIAVPQSTGPP			
				CCCATGCAGGC		RRRPVPWWNATIASAIR			
				CAAGGAGTCGA		ERRRALQHLKRHPSRAA			
				TACCCGTCCTCG		LVNFKRLRAKARYIIKQS			
				TGGGGTATCAG		KRVCWERFLSSLGSSVP			
				GGCTCCCGGCA		LSRVWAMLRSIQGCQRP			
				ATGGCCATCGT		SSIPGLALPDGVCTDPLV			
				GCCAGGTGGCC		IAEHLAAHFAKASATAS			
				TTTGCTGTGGCT		YPVAFRLKKQQAEASNL			
				GGGTGGCGCCC		CFTTHREEPYNKPFTEW			
				ATGGGGAGAGC		ELRLALSSSHDSAPGPDS			
				CCCGGATCGGA		IHNQLLKHLSPPQRQYLL			
				GTGGGTGGTAC		QVFNHIWLQGVFPSQWR			
				CAGGGCGGACG		DSIVVPVIKPGKDPSSLD			
				TTTCGCACG		SYRPISLTNVPCKLLERM			

						VARRLCWVLESRGLLSP			
						YQCGFREGRSPIDHLLRL			
						ETAIRKAFSKRQHLVAV			
						FFDLRKAYDTAWRRHIL			
						CTLHDWAYEAPYQFLSV			
						SFYPAACSVLGLRGSLPI			
						FISQFLSRRMFRVRVGSS			
						FSSPQIQENGVPQGSVLS			
						VTLFLIAINGLVSSVGSL			
						VTPALYVDDFCIWYSSH			
						SVASAERQLQGAIRRAS			
						TWALSHGFQFSPSKTRV			
						MHFCRRTTVHPDPELFL			
						GTQRLEVVAQHRFLGLL			
						FDKKLTWLPHIRQLKISC			
						MRKLNALRFLAHSSWG			
						ADRTTLIHIYRALIQSRL			
						DYGCQVYGSAAPSALKL			
						LDPVHHCGIRLATGAFR			
						TSPVDSLLAEAGIPPLQT			
						RRHQLLVSYAIAIRKFPG			
						HPTYHVLFEGDGLRTPD			
						TRPRVGLPVGLRLASLC			

						RDFNLPSLDCMRRAIPRT PPWLVPRPQIRLDLYHA PKASVSPIIFRRLLHTVLE EFQGATIFYTDGSKFAG HTGYAFTSPIGVECHSLP ESCSVFTAELTAISRALH FIKQVALDRVLICSDSMS SLQAIDRCYSSHPLVIDI HDLLADLRRTACSVAFL WVPSHVGILGNEQADQL ARVAVTPCPFPWVIPGA DLRVQMRSLLIQDWNT NWQATALSNKLRSIKST TAAWSSSLRPSRRESTVL CRLRIGHSRLTHGFLLRN EPPPHCGCGALQSVAHIL VECPLLLALRAKYGIPHS LPSVLANDSRIADRVLGF IRESGFYSRL			
RTE	RTE-68_LMi	.	Locusta migratoria	GGTAGGGTTTC ACCCTATCCCAC AATCTAAACCC ACGCAACTACC	TGAGGGAAGTTGTG TAAACGTAACCCAT AGATGGCTGTAACG	MDNYRILKGRPAIKVKK TEKTKNVVLILGTAFFIH NSITESIVEFRSPSPRLPV VTLASGNKLYTIFNAHA	cd01650	488	752

				ATGATGGCACA	AAAATAAATAAAA	PINEDNRKNPEKVEEFW			
				ACTACGGAAAC	AA	SQLEEELFRTPDKHVKIL			
				TAGTACCCCAG		MGDFNAQIGKERKFRNI			
				CCGGTTCAATCC		VGKYVAHRRTNRNGER			
				GGGGCCAAAAT		LIDVCESFGMKIMSTQFK			
				TTTGGTTCAAGT		RAPKQMKTWRSPNPML			
				AGGGCAATGGA		GEFQIDHVATTSPKEIMN			
				TTCTGGGGGCC		VRVQKGANFDSDHYLT			
				CTACACCATCAT		RVKVKPIPGNRRRKTPRI			
				GTTTGGATCAGT		QKYRIEELKISEEIVVKY			
				CAGAGGATCCT		QHELDTLNSNEWDKMA			
				TCATCCCGCAA		QKIQEAAKQTIFLAKKK			
				ACTCAGAACGA		KHPWWDDECEDALRKR			
				AGGAGAAACAT		SEAWNKWNSTKQNFEE			
				TTCTTTGGTACA		FKTQRRTTAKVFRNAKR			
				CTTAACGTGAA		KFERDQLKELEENFQKN			
				CACCCTATTGA		NTRSFYQTFKSKLRGYS			
				AAATTGGAAAA		SPNICLKMIRANRLLITQ			
				CAAAAGGAACT		HLFKNDQGKLIMNNKH			
				GTCAAAGTACT		NCEHLARYFKELLNCDP			
				TGATGAATACA		PPEKLTFQKPQNKLPDST			
				AAATAAACTTC		PPDKKEIREIIRGLKNNK			
				TGGCAGTCAAG		VSGKDSIVAELLKYAGD			

				AAACAAGATTC ACAACAGATGA AGTGTCAGCT		SFLERLERLFQKIWLTGE IPEEWRDAVIHPLHKKG DKSNVNNYRGISLLPVA YKILAIALKHRIEQHIDQ QLGEYQAGFRTGRSCAE QIFTIKTLIKTKALAADK KVIMTFVDFKKAYDSVD RDTLIEVLKEFGIDDKTT RLIQQTLTNTRSQIKFLG ELSEPFEIKTGLRQGDSL SPILFNCVLEKVIREWRK EMSQFNFPNGIRLGRKQ SGLTFECLAFADDIVFFA ENLQIATKQIEVLKETAV KAGLQISFEKTQYMNIN ATDPVWTMRTKYGDIK RVQRFKYLGEILTPNNN EKAAIEERARKMEIAFRL CQNTYKSKSLSYQAKFR HYSSVVKPECLYASETIA RKGLSDLEKKERKFLRK ILGPRKISNEFTFRLRSNQ ELYQRFENITTTMKKRR			

						LSFYGHIKRMKSERLAN KILLFQENRKTQVPWVK EVHRDLEKCGITAEEIVN RQIFRKKVAEVKDFFEE KTRKNRVFSAEERARVS QRMKEYWRKRKEENFR K			
Vingi	Vingi-3_LMi	.	Locusta migratoria	GGGGACAAGGA AGGAGCCTCCC AGAAGATTGTA CCGTACATGTA CAGTCGGGCGT CCCCTGGGCAA CGGTAAACACC GGCAAATCCTT CCACACACTGA GCTACGTCTAA ACCGACATTGT ATTCAGAATCT GA	TGAGCCTGCTATGG ACAATAATGACGTA TATGTGTGAAAATC ATGAGTATATATGA ATGTATAATTAGTA CTTAGTTTTTATTTG TAATCTAGTATATG TAGCCTATTAATTA ACTACAGATGTAGC TCAGACACGATTAA ATAAAAATAAA	MMKGPFSEPCLTAISLNI EGLSTEKEQLISNTCKLH SCDVLCLQETHRGPSQR RPRIPGMILVKECPHDQY GSALFIRPGIKVSNIRFNT VNNTEIISVIIGSITVSSIY KPPGERFAFSQHVCHPQ LVMGDFNCHSKVWGYQ DTNEDGELLESWAEEQE LNLIHDQKQPASFNSAR WKRGYNPDLIFVSKTLS GSCNKFVLNPIPKTQHRP IMVKFRPVIVPQEVSFKR RFNFKRANWKKFADILD KEIVYLPPNAENYEKFV EAVKRSARLSIPRGCRTL	cd01650	463	722

						YIPGISNKMKQLYNEYIE QYNHDPFSVETIKLGEQ LTETISKDRQESWKHML ESLDMKHSSRRAWNLL KRLSXDPAAPKDQVSVT ANQIAHQLLLNGKPTRN SKIVTSYKRPQVDEIPPV PYFTSIELETAISSLKSRK APGPDDVVNEFIQNIGPI AKEWLLQLYNFCWANK SVPKLWLKAKVVAVPK PGKSRDDAKNYRPISLL CCPFKLYERLILNKIGPQ LESQIIPQQAGFRPGKSC TGQVLCLTQYIEDGFQK GEKTGVVFVDLSAAYXT INHKILLENFSKKIFQNE GNYHLTEVVRSLISNRR YYVEFQEQRSRWRNQK NGLPQGSVLSPLLFNLYT NDQPAGPDARSFIYADD VAIACQSRSISHIERNLSD ALEVLTEYYNRNQLRPN			

						PAKTQTCLFHLNNQEAK RELQLEWNSVKLKHCPN PVYLGVTLDRTLSYKSH VEKTRAKIYTRNGIIRKL TNSHWGANPEILRSSSLA LCFAPAEYACPVWRRST HAHKLDTALNDSCRLIT GCLKPTNRDLLYMCSGI APPQIRRQTWRWLKEAD SARIEDIPCSDISQWCKIK IEERFHFEHPLNESQSTT RESMWTQKLEKKKKKN SLQCWNIKEKLPAGSQL EWRVWKSLNRLRCRMG RSKDNLFKWGYAEDEL CQCGEKQTMTHLLSCPS LPAPCTFQDLWLANDAA VKCAEHWREI			
Daphne	Daphne-28_LMi	.	Locusta migratoria	GAAGATGGCTG CCGATCAGTAT ACTCATTTTGTG CGCGCCGTTGT GAATAGTAAAA	TAATTTCCACTGTC TGTGACCGAGACAC CACTAAAACTGCCT ATTTTAGTTACTTTC ATGCATAATGTTGT	MPRKFFRPKMFTHGKPH ETVLQVPPNNRFSVFEE NTNTVDSFATVLNNERV SANTSRKIGKKCVKVNI HADSHGRGLAKILCEDX	cd01650	924	1185

				ATCGCGTGTTCA	ATGGCATAGTTTTC	DVDAYSYVKPGAPLYEI			
				AGTGTTCTCTGT	TGGGGAATCAGCCA	GKNIKDGSPMNSNEYTV			
				GCGAGAAATAC	CTTTCCAAAAAAAA	IVGGANDVYKNELSSAI			
				AATGCTATATGT	TATTCATTGCACAA	QNLNIILESNNSNAMIIL			
				GATAATTTATGC	AAGAGAGCGATCA	GIPHRHDLIYESCVNKEI			
				GTGCCGTGTCA	GGATAATGAGCGG	AKANKQFENICNLYPNT			
				CATCTTAATAA	AGTCCAACCTAGAC	QFLPIDFLQRDSFTRHGL			
				AAAACAGTGGA	ATTCTGTAGAAACC	HLSRKGKKLLCFKIAEM			
				AAAAATGAACA	TTTTTAGAAAATTG	VSANTKKGSNSQKTVQT			
				GGTAAATCCTC	GGGATCTGACAACT	SVPQQPTQSPSAQLYNIN			
				GTGTAGGTCTTT	GCATGTCAATATAT	ATTKTPTDDPILPESTNS			
				GTACAAGTCTT	CTACTCCCTACTGT	PPSVDSIEPTSVAAMPGL			
				AGTGATCAGCG	GCTTTATTGTCAAA	IVDTTVTPASHQEEDIW			
				TGGTAAACATG	AATGTGCACTTGTA	ASRIVPEPMEEASSTPED			
				GACATCTGCAT	CAAAACTAATGATG	SSQLLKIQGFFMVSSQGR			
				CAAAATGTGGA	TTACCATACTCACA	KNSVIATPPSKPNVNTKR			
				CGAAATTCTGA	ATACCAGGAGGAA	EINPINHPKTLVSVNKLT			
				GTGAATTAGCA	GCTAGATCTCCATT	VFHQNIQGLRSKLDEISI			
				ACGAAAGAAAA	ATGAATTAAAAAAC	NILDKDCIDSADVICFSE			
				AATTATCAACA	CTAATACTGTACAA	HFMKLDTDSLQLDSYNL			
				TCCTGTGTGACG	AAAGGAGTCCACTA	AALYCRQSLEKGGVVIY			
				ACATAGCAAAG	TATGAGCTGCAAAG	VKNKIVHRSLDLSDYCI			
				TTGCGACAGGA	TGTTCAATGCTCTT	DQHCEVCGVVISTNLSEI			

				ATATCATATTTT	CCACTATCCTGAAA	TVLSCYKAPAGKICIFLK			
				CAATCAACAAA	TCACTTGTCCATGA	QIVSVNSYFLQKQRILLIF			
				CAGTGTGAATA	GCCTCCCAAATTTA	FAKTKNIIVLGDFNVDY			
				CATCGCCAAAT	AACAACAGCTCAA	LIDNMGKNDLSHVMSSF			
				ATAAATTGTGCT	ACAATATCTATAGA	NLTAQVDFPTRVTANSIS			
				GAGTTTATTAG	TAAGTCCTTTTATT	RMTISSPNDNIFLDKSRN			
				GCCTAAAAAGT	CAGTTGAAGAATAT	PNTIVKRVLNGLSDHDG			
				TCACGAGTGGT	TTCAATTTGTGTTA	QLLSIEGVGPLSKNKSYS			
				AAACTACACGA	ACAATTTATTTTTT	KTFRAINSQSINTFCDLL			
				AACGGTCGCGC	AATTTACAAGATGT	QETDWDQIEHLDNVDD			
				AAGTGCCTTCA	TATTTTTGGTAGCA	KYNAFLNEFICLFEAAFP			
				AATAACCGTTTC	TGTATTGTAATTTT	KRTAKPKHETYCHKNW			
				AGTGTCCTTGA	GTCAATATACTATT	LTRGIKTSCRRKRNLYA			
				AGAKAACACTG	CTTGCTGTGAGATC	SLKSCFSLEKEKHYKLY			
				ATACTGTTGATT	TACCTTTTTTATGTA	CSILKKVIRAKSNSIKSEI			
				TGTATTGAAAA	ATTTTATTCTTTTAA	DVSSNKIKTIWGIIKRET			
				ACAAAAGTTTG	TTTTTTGTGTAATGT	GKIPITENSLEIKKMMSF			
				TCTACAAAT	AATTTTTGTAATTTT	MIMLKKNDVIHDNVEIIT			
					AATGTGTAATTTTT	DIFNSHFLSSAAQTGKTC			
					GTCTATACCTTTAA	SINEAMLFLQKSIKKQSK			
					TGTGTAATTTGTCT	LTCLLSQCMKLRLKTKQ			
					ATTGCAAGAGGGAT	INMSPVTVHEIEVIIKEM			
					TCTGTATGTTCCTTT	KSKNSVGIDGISSRILKH			

					TGTAACAATTTGAC ACGTCCTACATCCA GTGAATGTCTCGCT GTATTGGATCCATG GGATTAAATAAATA AATAAATAA	CYKPVSKILCHIFNKSIQ QGIVPERLKYAIVKPLFK KGDKTDVTNYRPISLLTI FSKILEKLVHKRFIDHLT QYKLLNSSQFGFQKGIST DQAIFSFTNQILESFNNK LSPVGIFCDLSKAFDSVN HEILLEKAKHYGFIGSAG KWIKSYLTGRKQKVML NNSEGHSVSSSWGSIGC GVPQGSVLGPLLFLIFIN DLPLCTSLKTHFTLFADD TSILINESSNYNLEESANI VLREALNWFVINGLALN ANKTQIIHFQPSKKMEDP ITIKCNGTDLLQAESTKF LGIHIDSNLKWSEHILQL HKKLSSATFAIR			
Penelope	Penelope-3_LMi	.	Locusta migratoria	AG	TAATGTAAACACCA ATCCTGGACTTCCT TTTCCACCAATAGG AAGAAATAAAGCA AACACCAATGATAA	SNRAKNICAQASKALLR ERIRHIRSSLDSSNRKLIE LHLALSSKLAEEDWEKV DRLTFQQAERTNDSVAT RQMKKFDNLQQKTTGN	cd00304	380	464

					GAGAGCTGAAACC	GQTMDTQRTVVNLSSHE			
					ACCAATCATAACTG	LSDAATQILSKGMNFAIT			
					GCAACATTTAATAT	PRRVPKEDIIESVEASLR			
					TTAAGACACCAGCT	HLPQAEAEKIRQETVRV			
					TTCCCTCATCCAGC	LAHAKPPKQNISNEERR			
					AGTAAACGTAACAC	AIIELKNNPDLVVVQSD			
					CCAGAAGATGGCC	KGNATVVMNTTDYDDR			
					GCTTGCAATAGTGG	MEKLLSEPIYKKLKGDP			
					CCGAAACGTCGGTG	TAKVIKQTQALLKKTRM			
					AAATATTTGATATA	NFDTARRLNPQVPQAPR			
					TTGACGATGCGGTC	IYGVPKIHRESFPLRPIVC			
					CAATACCCAGAAG	SIDSPTYRLAKELVPKLQ			
					AGTTTTATCAACAA	HLVGHTDSYVKDSTHFV			
						ALLEEKRISTTDLMVSFD			
						VKSLFTNVPVSDTMNIL			
						EERVPPDVCDLVQHCLT			
						TTYFRWKGEFYEQTDG			
						VAMGSPLSPVAADIFME			
						AFEETALQSAPLRPDCW			
						LRYVDDTFVIWPHGEEE			
						LQNFHQHLNQQHNKIQF			
						TMEIEKNGVLPFLDVEV			
						YRKPEGKLGHRVYRKA			

						TSTGRYLHATSHHHPTQ KKSVLHTLTKRAHRISD EQHLKAELQNIRSIFEAN GYGKRMIEQTMTSNDG NKREEQKEEQTIALLPY VKGVTERVGKILRREGI KPIFRSNNKIRDVLGSMK DKTDKLHTAGVYEIGCE CGQVYIGETGRPINTRIA EHERYIRLGQHTKSAVA EHQEECGKQIGFQEVTV LAKQPYTLKRKIREAIEI AKRPGNMNREDGYRLP ASWLPAVAALRDDTAR SAVHRGPHSPETTAAAR TQSKQHALAGARHQCY VTATTAPHSPVSTNGKH			
CR1	CR1-3_LMi	.	Locusta migratoria	CTCAGTGACCG CCGGCCAGTGC ACACGTGACAG CAGCAGCTCTA CATTGTAAGGT AACCGGCTTCTT	TGAAAGTGGTTAAT GAACCCTCTGCCAG GCACTTAAATGTGA TTTGCAGAGTATCT ATGTAGATGTAGAT G	MARNCDCCERMEAELV TLRSQLQAALASLATLE TAVGGRRCCGPLAGVQ QTPCASVVPSDRPAAVA SPLTAREGVAPQSVVDW EVSSARVRRRKTSRGTE	cd01650	871	1141

				AATAGAAGCGA		VEASPVSLANRFEALSV			
				CTTATTTTCATT		AEDVPDADGVACPVLAP			
				TCGTGTGTGCTT		LPSRSGHSQRVGLLVVG			
				TCTTTTACAGTT		SSNVRRVMEPLRKMAA			
				CTTCAAACTTTA		REGKKSNVHSVCIPGGVI			
				GTGTGCTCGCGT		PDVERVLPDAMKSTGCS			
				GTTTATTTATTA		QLQVVAHVGTNDVCRF			
				AGAATCGCGTT		GSEEILSGFERLSEVVKT			
				TAGTTGTCAGTT		ASLACKMKAELTICSIVD			
				CAGTTGTTGATA		RTDCGPLVQSRVEGLNQ			
				TTTTGATTTATT		RLRRFCDRVGCRFLDLR			
				AGCTCGCGTAT		QRVVGFRVPLNRSGVHY			
				ATATTTTTCAGA		TQEAATRVAGAVWRGL			
				TTCGCCGTTAAT		GGFLGVRSPLYSGSGYT			
				AAACAGTTTAG		GSGGCVAWTGRFFRLEG			
				TGTTAGGTTCAT		LGNTQNGFQSQRGQVR			
				ATTTTTCGAAGC		HRKKADLGTIGVTVVNC			
				TCACTGTTTATC		RSCVGKVPELQALIESTE			
				AAACTTCTGTGT		AQIVIGSESWLTPEISSAE			
				CCTCAGACGTA		IFSKGQMVFRKDRINKK			
				CTTACTTGGAAC		GGGVFIAVTNGLSCSEIE			
				ATACATAGGTTT		ADTSCELVWVEVIPDNR			
				TTAGA		NKLIIGSFYRPPDSDDAV			

						AEMFKESLNHISNRYPK HTVIIGGDFNLPSICWHN YVFKTGGKYKTSSELVL NAFSENYFEQLVHEPTRI VSGCENTLDLLATNNPEI VKSIKTDTGISEHRVVVA RLNVEIPKCTKNKRKIYL FNKADKNSLHAFLSDSL QSFKTKNASVDQLWLTF KDIISVAIERFIPNKLIKE GADPPWYTKQVRSLLQ KQRKTHAKFKKTQNPQ DWQSFTETRNLARTSMR EAYSSYHNEILSRNLAEN PKKFWSYVKYTSGKTQT IPSLRDSNGNTTDDSTAK AELLNTAFRNSFTKEDE VNTPVFNPRTAASMRNL EVDVPGVVKQLKSLNK NKSSGPDCIPTRFLSEYA DAIAPYLAIIYNRSLSERS VPKDWKVAQVTPIFKKG NRSDPQNYRPSHSPKLQ			

						AISLTSICSRILEHILCSNI MNYLEENDLLTNSQHGF RKYRSCETQLALYSHEV LSAIDKGSQIDSIFLDFRK AFDTVPHKRLLMKLRA YGISPQLCDWIHDFLSER TQFVVIDGKSSSKTEVIS GVPREALNRSDIWRSQG SVIGPLLFLVFINDLGDN LNSSLRLFADDAVVYRL VKSTEDQKQLQSDLDKI STWCENWQLTLNKEKC EVIHMTTKRNPLNFNYA INQSNLKVVNSTKYLGV TITNNLTGENTGRKFYQI PRSYNYEQLDWREHIEN IVGKANQRLRFIGRTLRR CTSSTKETAYITLVRPLL EYCCPLWDPHQVGLTED IEKVQRRAARFVLSRYG GEASRKCCTNWGGNHE GSEKGSSFCFITLRGGSV TEMLHELGWKSLKERR			

						YSLRRDLFAKFQSPTFTS ECENILLAPTYLGRHDHP NKIREIRSKTERFRCSFFP RAIREWNGREKL			
Daphne	Daphne-41_LMi	.	Locusta migratoria	T	TAATGTGACAATCA AAATTTAATTTATA ACCTACACTCTTTT ATTGTACATATTTA TGTTAAAAAGCTAA TCATGTAACAATAT ATATTTTAATATGT AACATTTATATATG TTAACTTTTGTACC ATAATGTATGCAAT GACGATGCCTTATA CAACATCACAGTTG TCCACAGGCAAAAT AAATCAATTTCAAT TTCAAT	EQQCMPQGTCSDDQAC PHKSWSVKSQPSFLTTG MNLCNYKPVQRLEVKD CGSMSFFHLNIQSVRNK LPELELFLLEFQCDILCL MSTGHTLSNEHWLKEH EHNMYTPLGYKFLTGFF RSNIEKGGVSILMKNDL NCKIRIVDLKHLASEGQI EIAAIHLPPSADTLKSHN FSCLNDKPTRQNACLDN VISNLSRDDVNLNVVKP MLSDHDGILLKLIVPKK DHRPTQPIVNRPVRQLK PINITNLTQRLLTFDWDA TVMENTTLNNAFDYPPQ LIIMLLIIRPNLFNEYCPP PPSTKVKYKEETKEKLT MVHPHLNSLRILLKIYYE	cd01650	498	740

						KSCTDSEYRHSYNKLKN IYKKDIISAKKEANGTYI SQATNKCQAAWKIIKAE AGLQVVNPPIPRADELN KNFINSGQSAIDHSNLDD NADSFLSSLETELXVDK LKFRWHAIYPKDILDVA RSLSNSRTEDIYGLSNFTI KRIIEAICIPLCQLINRMF LEGTYPDCLKQTIVLPIY KKGDRTSPQNYRPIAIVP ITSKIIETIMKNQLNDFFC NSLLNKAQYGFRSGLST VKAVQEVVHSVLHGFE NRQHSYATLVDLSKAFD TVSHGILTKKLETYGITE KELALLKSYLANRNQM VLVNSTNKSDPLKVQYG VPQGSVLGPFLFIVYVN DLPNLLPFKSVMYADDT TFITSHNDVNAVKQMNN SMMEKATQWFQENRLT LNNNKTEHILFSLGNINN			

						EQNHPVKLLGIYLDSKL CWNTHTEKLCSKLARVL YLLRKLKDCVSKELML MAYYAFFHAPYIMEYFF GETPLERKKSSSGKRKPS DAFLARTHAYGILCFFPC TLHYGILLWGNSAGAKE VFLWQKKAIRCIFGLSKL TSCKQYFVQHKILTVPSL YVLKALMHARQNVDEY ESRSDVHLYGTRNRNLL DIPWARLSKVKNSFLHM AKIFFNKLPYNIRELHEN KFKIVIETWLKEKALYN TNEFIWTENKSCY			
Penelope	Penelope-37_LMi	.	Locusta migratoria	TTAAGACGGCC ACTGCACGGAA GATTATGGAGA AGGGAGCCATG CTATTGTGAGA GAGAATTAGTT ATACAGGCAAA GGTTGGATACA	TAAAAGAATCCGTT GAAATACGCATTGC GGAAAATCTGATGA ACCGTGACAGCGGC TACCAGCTGAATAA TGCGTGGAATCCCG TCATCTCCGAGATT TGTTCTAGACGAAG	KLLLNAVIQLLRKGLHY EIPRESGIVVLKADKGNS TVLLPRRLYLEKMYGLL SDTMYRKIEHDPTARVQ RKTCTLLNSSSLAPETSS GPYSTCATEDVHTFEFFL VSSRDIKRLRPHGSVPPR LYGLPKVHKEGIPLRPIV	cd00304	246	330

				TTTCAGGAACT	ACGCCAGAAGACTC	SNIDAPTYDLAKHLASL			
				GTTCCCTTGACA	CGATAACTGCGGCA	LRPLVGKCTHHIRNSAD			
				TCGAAGATTGC	GCGGCACTATCGAA	FINRLGALHLSSSDLLVS			
				AGCCTTTCCCCC	GACAGCTGATCCTT	FDVVSLFTKVPLADSLA			
				CGGGACTGGGT	GCAGTTCCACCAGC	LIGKRFQSDITGLFRHAL			
				TGATGGAGTTA	GAGGGCGCTGCCGC	SSTYFMFNQEYFEQTDG			
				CTTCGCCAACTT	CGGGCGGTGTGGTC	VAMGSPLSPLVANFFME			
				TGATTGGGAAT	CTTCTGTCACCGTG	DFEERALESAVCKPTVF			
				TTTGGATGCAA	TCTTCGACGCATGC	WRYVDDTFIVWPHGED			
				GTTCGGGATGT	GCGGTAATACCACC	KLMEFLHHLNSIHENIRF			
				CAAAATTGATC	AGCGCGCTATATAA	TXELEKDGCLPFLDVLV			
				GCATGGTGCCA	GGCGGAGCGGAGA	KRKSDGSVGHSVYRKPT			
				CAACAAGCTTC	GCGTCTTCGTCAGT	HTDSYLHSSSCHHPSQT			
				AACACAGGAAG	CCTTGTTGGCTCAC	MSVLKTLVHRAHTVSD			
				TTACTGTACAAC	CTGAAGATGACTGG	AESLPKELAHLRTVFRD			
				GTTCCATTTTAA	CAGGTGTCCAGTTG	NGYSTRQINRAFSPKTR			
				ATTTCGGAGAG	AAATATCGTGGAAT	NREVDKEENAPAKSLAF			
				AATTTGGACGTT	GAAGTTTACGACGA	LPFVGNVSFKIGRILNNY			
				GCAACTGTCTGT	CCGGCTGCAAACCC	QVKVIFRPPSKISTCWDQ			
				TTTGGGCAAAG	GAAATCTTTTTGAA	DFDLLGSVKDDLLLRKA			
				GACTGAAGTTTT	CA	GIYKIPCECGMAYIGQTT			
				GTACCTACACC		RTWKNVAQNINVALAF			
				CAGTGACTGGG		CNPASLQLRSIVFPTDIQ			

				TTGAAGAAGCT GTACGACATTC CTATGGCACTG CTGAAGAAATG GGAAACTTGCC GTGTTCTCTTGA		WSMIKHQFWPQQHPSG TPL			
Vingi	Vingi-9_LMi	.	Locusta migratoria	CA	TAATATTTCTGTCT GTTTATAGTTCGTA GTTTACATGTATGT AAGTAATCTCTGTA TGTTATACCTTTAT AAACTTCTGTACAC ATGATGCTTCTGAC ACGAATAAAAAAA	VLCHIFKLFEKVLHDRLS KAIDNYLIPQQAGFRPG KSTTSQVINLTQLLEDGY ENRKITGVAFIDLTAAY DTVNLRLLMKKLYAITR DFKFTSIIRSLLQNRRFSV EFQGKQSRWRAQKNGL PQGSVLAPLLFNIYTNDQ PLPAGTENYIYADDRAM AVQGDSFEEVEKRLTDA LDALCQYYADSQLKPNP SKTQVCAFHLRNKQAN RKLRVSWKDTPLEHCA NPKYLGVTLDRALTYKQ HCINTKMKVSGRNNILR KITNSSWGAKPEVVRTT ALTVCFSAGEYASSVWY	cd01650	1	232

						NSAHAKQVDIALNETCR LVTGCLKPTPIDKLYHL AGIAPPAVRREAAAYSE RARAELSESHLLHNIQPA TRRLKSRRSFLRATDNLS DQPGSRVDRWKRSRER HWMEPKEELAPGHNED WTVWKSLNRLRTNTAR SRDNLMKWGYSTVSNL CDCGDVQTTQHMYNCP ECPSHCTLEDLMLATPN AVDVARFWAKCI			
RTE	RTE-24_LMi	.	Locusta migratoria	GTCTCCCACAA GGCCAACCCCT CGTCCACGTCGT GGGAGACGGGC AACGGCGCGAA GTAAGGGATTG TCTGTAAGGGT GAAGTACAGTG GAGACTATGCG AGTTCCGCGAC CGCTACGGCAG	TGATGATGATGATG A	MDRRKVKILGLSETKW KGSGKKTLRNGYILHWC GSQTETRNGVALLLHKD VAIHTDIRYVSDRIIQMK LHVDNTYITIFQVYAPQT GCTADEKEQFLENLEDLI VEDNVIIMGDLNAQVGK ERQGCENIIGPHGYGKR NPEGEQLLDLCVRNKLII KNTWYQKRESHKITRYS WDGQHKTVIDYIITDQK	cd01650	497	746

				CTGTGAAGGCC		ASQTVKVIPSECMDSDH			
				TTACAGGAACC		RLLTADLKNVRVPLVKY			
				CTGAAAAGCGG		KKMPKIKTWLLRDPEKK			
				CGGCTAACGGG		QEYQKKIKDRIPKTEKL			
				GCTCTGGTGAA		GVEEEWKLFKDTLVTEA			
				GCCGTGTTAGG		VHICGKTSAAQREKETP			
				CGTAGCTGGCC		WWNDRVKTAIKERNVL			
				AGCGGTGGATT		KKKLDREKQNRVDVRN			
				AAAGCGTGCTT		DTEIERLSQEYRNKKLA			
				TCAAGACTAAA		VKRLIQNEKDKCWDDFT			
				ACAAGCCTCGG		TKLQEDCQGNKKLLYK			
				ATGTATAAAAG		LVNDKRYERTNIVVLET			
				AAATTACTGGA		EEGKLLRTESEIRSEMRN			
				CATGACACCTCT		HFNSLLNGNVDSSDDSD			
				CCAGGGACCTC		CPAEETPEVLQPLTWLE			
				ACGAAAGTATG		VESALRNMRNGKSPGFD			
				ACGGCCAAAAA		ELSSDMMKAAGTPGLQ			
				CCTGAAGCATA		WIYRVLSVIWTNNKIPD			
				TTGGAGACSCCT		DWTKGIIIPLFKKGDRKK			
				GAAGACACAAA		CGNYRGITLLSHTLKLLE			
				AACCAACAAAG		KIIETRLRKIVEPLLEEEQ			
				AAACAAAAACT		HGFRKGRSTIDLIFAVRI			
				CAGACTTGGAA		LAEKYWEHGKNLVMVF			

				CTCTAAACATCC TCACGCTCACTG GTAAATGTGAA GAGCTTGTTGA CCTA		LDLEKAYDSVPRNKIWK CLEDLKVPNYLIDRVKM LYQTCNSSVQIGNGRSE WFETKRGVQQGSALSPL LFIIVMDRMLKSIKETAN ESRSLVFADDVVIWGET EKEVQKSLTVWNNKFK EFNLKISKTKTVEMTINR EGTSSNIYLEGTKIKSVS HFQYLGSIFSKDNTIRQEI LNRTQKASNFYMQVRN LLWDKKVPQAAKLTLY NTYFTPILIYGMETCTLL KKDYSRIQATEMRFIRT MNQSTRRDKIRNEVHRE TAGIRVPIANILGKRRLQ WYGHVMRMDSTRTAK KYFRMEVGGRRPRGRPR RRWIEAVKSDIEKNGFN WDDVQDQMLYQDRMR WRALVSHTRETGDGK			

Penelope	Penelope-54_LMi	.	Locusta migratoria	CCAGGCACAACATTACCAGACGTGAAAGGGCGCGCTGCGGTCGCTGCGTGATGATCCAGACGTCGTCGTCCTGCCGGCGGATAAGGGGAACATAACCGTTCTGCTAACAGCTGCGGAATACAAGGAGAAAATGTATAAGCTCCTAGATGACCCTGCTTACAGGAAGATCACGACGGATCCAACCAGCCGAATACAGCGCAAGACAATTGAACTTCTTCGGAGTACTTTCACAGACGAAACAGCTAAGAAACTCCGT	TGATGATGATGGCACGTTACGCTATCGAAAATTCGTTCCACAACGATCAATCCATCCGGCTGCACGCCCGAGAGATCTTCACTCA	MGIPLRPIVSTIGAPTYQLAKHLTSVLSPQVGKCEHHIKNSAQFVERIRQLRLSNSDLMVSFDVVSLFTRVPLKESLELIEKKFGPSFTKVFNFVLTSTYFLFDGEYYEQTDGVPMGSPLSPVIANMYMEEFEDKALNTAPLKPTIFYRYVDDTFVIWPHGEDELQVFLQHLNSIHSQIQFTMEVEQNGQLPFLDVLVRRKPDGTLGHSVYRKETHTDQYLHASSCHHPSQREGVLTTLVHRARVISDEESLKPELDHLKEVFQENGYSDRQIRRAFKHKSTPQPTPEEDPPAKIAYIPYAGNVSGKIGRILRRQNIKCVYRPPAKIKSLLGTVKDDLGLRKSGVYHIPCECGKAYIGQTIRTIEERCTEHKRHTELQQPTKSAIAEHCLETXHKMKYEDT	cd00304	114	198

				CCTCGTGCGGC CGCACCACCAA GACTCTAGGGC TTCCAAAGGTA CACAAG		EVLHNTSRYWDCLYKE AIEIRLRENLINKDSGFQ LSHAWNPALKFLRSRRK QNTEHSEGRAQTEVRHS SPEQELPPPPQHPGVRTR GQWRGVTAGIYSTAAGP APQSAGVT			
Penelope	Penelope-40_LMi	.	Locusta migratoria	AACACCTACAT ACCAAGAAACA AGAATAAATGT TGAACTTTACA AAATTCATCTG GAACTAGGCAA CCACATCTCTCA GACATACTCAA TGATTTAATGA ATACAGTCAAA CAAAATACAGA ACAACTAATCA AAATGGGAGAA ACAACAACAAA AAATAAACACA CTCCTCAAACAT	TAAAACACACACAC ACACACACTAAAAC GAAAGCAACATGA TTCAAAACTCCCGC CCCCACCACAGTCA GCTGTTCTTAACAC TACTCCCAGAATTC AACTGCCCACCATC TTGTTTTCATAACA TAAACATTGAGTGA AAGTGAAAGTGAC ACAGTGCCTGTCCA CAACACGAAAACA GCGCATATCCACCA AAGTAAGTGACAA AAACTTGTTATAAA	MSGTKSAEEKTCRRIRN KLKNEGALITKADKGTP WSGCPHHKGRQGNTMV IMKQDDYRNKTLTFIED NNIHKLTSDPTQRFHKN TQTLLKSITHIFPKQKIK WLKQKNPKAPTLNCLPK IHKEHIPVRPVVNFRNAP TYYLAREIHSYLQKHYV FTNNRSILNTSELIQKIKH IIIPSTATLTSFDITSMYT NIPITETINIIKQQLQQQT DTPKSHIKETTNILKLITE QNYFTFDNQFYSQEDGL PMGSPVSGLLANIFLNHI ENKIFDTIVKPKQYKIIY	cd00304	247	335

				AGCAACACTGT	CACATAGACATTTT	WHRYVDDIICLLDETPSR			
				TAAAGATAGAA	GTTATAGAAGCAGT	IQELHNDINNVHPKIKFTI			
				ACACACAACCA	AGCTACAGTAAAAT	ETENDRRLNFLDLSIHRD			
				AACCACACTTTC	CAAACCAAAACAC	KNQHKFSIYRKPTTTSTA			
				ACAAACGCTAG	AAAAACTACGTCCT	IHKHSNNPITQKQAGFQ			
				TCAACCTGACT	GTATTTTCGTAAAA	YMLHRLNKTPLSDEQYK			
				GACATACAATT	ATCCCGATATTTAC	NELQTIKQIATENGYDT			
				CACCAAACAGG	AAGCTAACCATTCA	NIIDKLNHKIKAKLKGTT			
				AAACAAACCTC	CCCCATTAAATACA	NTQKQTDQKHNTSTQNI			
				TTAGAAAAAGG	CACGTTTTGGTTTG	NPLNSHTPNTHTQEQKS			
				ACCCAAATATA	CAGATGACAAACTG	NKTTNKWHTLTYDNKA			
				ATGTCAACACA	TCAAAAACTMTCA	THRIGNILKKQGLKVAY			
				CATATTACACA	GTGACGTTTTGTGA	RTDNSIQRKLKTTNTNT			
				CAGAACAATAG	GTTGAAGAACCACA	DKHNTSGIYQLTCQDCN			
				AGAACCTTGTC	AACACAACACATGT	SVYIGQTCRNFNTRYTE			
				ACAGAAAGTGA	GGTATTAAGGTAGG	HRRALKSNSHSTFAEHLI			
				ATTCATAATAA	CCTAGCATACCTAT	NMNHKPTDINTDLKILK			
				AACAACAAGAA	AAAAATAAGAAAC	HSKSLPQKLTIEENYHIQ			
				AGAAACACAAA	GATTTATATTAGTC	KAIIERKEIINEYTALCNG			
				TTTCAATGCAG	TCACAATTATCTGT	TIFSALKELTENP			
				GCCTCACAAGA	AGTTTTCACAGTTA				
				GAATTAATTGC	ATGTACAAAAAGTA				
				AGAGGAAATTA	ACAATCATTCCATT				

				GATGCATAATA AAAGAAAACAG AAATGACATA	GTTAATTTGACAGA CAATAAACATCCCA CTGATGATGCTGCA ACGCCAGCGAAAC ATGTATGGGAGAA ACTAAAAAACATTT GTGTTTGCTCAAGG CGGACCATACCTAT AAAGCAAATTTATT TGTAAGCAAACACG				
RTE	RTE-37_LMi	.	Locusta migratoria	TATTGGTGGATT GAGTAACCCTG AAGTGCTGCTG CCGGTCCCAAC CCGGATAAAGG AGGAGGGTTGG CAGCCAAAGAA GCCGTAAAAGA AAAGCCGGGTC ACATCTCAAGG CACCTCTGAGG CTAACTACCCC AGTTGTAATCA	TAAAAAATTCAGGC ATTTGCAGGATTTG AGGACAGAACTAG ACCGAGAACGAAA TCATCGTGGACAGA AGAGCGAAGAAAG ATGCAAAGTATGAG AATGAAAGAATACT GGAAGAGGAGGAA GAACGGAAGAAAG GAAGCTGATGTGAA TTAACGTGGTCCAA	MMASSLKGSTTPGGNQS TIVKDNVSEPTDSGELTP TCKRLKSKKVSYFATFN VNSLLKIGKLKHLTDTL TNQNIAITALQETRYTDE AAFESEGFRXLKGKPGK RVMKNVPQLGTGFIINR QILGAITDFSSKSSRLSTL TFRGANKMYTMVNVHA PTNETNKTDPEKVEEFW XDLEETLSKIPTKNVIILL GDFNAQLGKERKYRNII GNYPAHNRTNXNGERL	cd01650	556	814

				CGTTGACGCCT AGGCGTCAACC SCCAATATAAAT CATTCGTTTCAG C	AGATGGCCGAAAC GAAAGAAGAAG	VEICKTFDLIMKSTAFKR LPRKQMTWVCQILCKTD DMGMPNPLLGEFQLDH VAIPKRAGKEIQNVKVL RGANLDSDHFLXKIKIKI IPKVHQPRNTIRTVKYDT GKLKTSTDFKTSLSQKD YQGWEQLKDAAIEKRR KAWADWCSHKTLENLN KFQEVRKNTAKTIRNLK RKYEQARXTQIEEDFQK NNTRDFYRTFKTNLNRY DPPTLSFKNKEGKIVYGE MENSQILAEYFEELLNCE APADTFSYNDTEPNPDS EPPTTEEIKNIINTLKNNK APGEDGIIAELFKYANEN VIIELKQIISKIWEEEKLP EGWTSALVHPIHKKRRY IPFTKKGDKTDPGNYRGI SLIPVAYKVFSKALLNR VEKQLNPQIDEYQGGFR KGRSCAEQILNLKMIMA			

						YQKQRCKKFVLTFVDFK KAYDSVDRQTLFQILKE MKLDRKSSELVRQTLNN TTSKVKFRGKLSESFKIN TGVRQGDGLSPMLFNCV LDKVIREWRQGMKGVG GIRLGCKKKGIRVDCLA FADDIVLLSESLEEAVNQ TKELQRQAAKTGLQISIE KTKYMTNITEAPGWMAI DRGKIQKVEKFKYLGE WIDTGLTEKDAMEARIN KINLAYKLTMNIYNKKN ISINAKLKHYQTVIRPEA LYAAETLNLIRTGVIKRL ELVERKILRRILGPRRTE DGQFRNRSNQELYSKIE KISNIIKKRRCEFFGHICR MDQGRLTQQISQFLLYK KSKIRWIQEVQRDLEEM GIRDQEIHNRVIFK			

Daphne	Daphne-29_LMi	.	Locusta migratoria	GAAACCGTTGCAACAAATCGAAAATTTAGGACGATTTGTAGCAGTTTTGTAAACACACATTCATTGAAGCAACCAGTTCTGCCAGGAGCACTTTACCGGACATGGACTGCACAGAAACCACCTGGGCAAGGATGTGTTGGCCAAAGAAATTCTCCATGTGATAAAAGCTGCAGGAATGAACAGTCACGTAAAAATTGCACTGCCTGCAATATGTTCATTGCCAATTACAAAMGAAAAGAAGACGGTACCAACAGC	TAGTTGCCCACCCATTCTACAATTTAGAAGAATTTCTCAACTGTAATATAATATTGTAATATTACTGACAAGCTGTTATCTCTTTTAAATTAACAATTTATGTTATCTATGTGACGTTGTCTATTGCTGTAATGGCCTAATGACAATAAAATTAATTATTATTATT	MSQSPVASSVAVVGNKAAVTVPLREDSPVASSVAAVGNKAAVPVTVPPQZASTATASTAPPNEEAAEKKKRSSREEEVVSSVAGRRKMVPPLRLKDFLVEGSKMKNLEISDLKHPIAPHQNVHSKGNSSSELKFATNLLPEHQGIKDKEFFVNDFGLDNNFDIFCFSEHWATENELTLVLDTYNIANSYCRKLHLRGGVAIYTNQSLSYSFTKLDLDYLCEEQNFEAAGIFIVSLKLVIISLYRSPKGIISVFLEKLEMLLNVFRESKWLHYDIVIGGDLNADFDVTTNKGVSLNCNDKQRSVTELKNLLRQHNLHSINSRPTRDKACLDNIFVNFKTAEDSCMNCRRLMYVTVFPFSDHDAVSVNYISRFHPDKCDIPVPKVVITRPITDDKIS	cd01650	665	908

				ATATGAAGAAG		RFRMALADREWFDQCY			
				AAACACCTATG		GDTHCSLSSDLSAKVVF			
				GCTWCATCTAA		DRFFKTFLTLFNDSIPLK			
				TAAGTAGAAAG		NAKQHSPKKCKVIDKGF			
				TACATCAGCAC		KTKVTNRQNPWYTEQL			
				CAGCACATGAA		TNLKHQLMVLYNISNAL			
						KTEHARSAYVKCRNEY			
						KKAIVEAKKAHNFRSIE			
						NSTNKCQAAWKLINSVA			
						KDPRNKSVNVSPQKFND			
						YFIKSVEEVGNAIIKPDIR			
						SSELLSKNVSRSTTNINV			
						FTFSEVSPSLILKIVKQLK			
						SSDSADIYDISCNMLKRV			
						IDCIVYPLTFCINKCISEG			
						YFPDELKVSRVVPVYKK			
						GDTESPSSYRPISIVPAFS			
						KVLECVIYQQLSVYFEN			
						LGLISESQYGFRKNLSTV			
						DAIDSVVKYLHQVFEDK			
						GFAQVTFCDLSKAFDCV			
						EHTPLLDKLEFYGIRDIS			
						LKLLKSYLQNRRQVVCI			

						GKEMSSTEQVKVGVPQ GSVLGPFLFLIMINDLPSF IKSSTVLYADDTTFLHSS NDXSSLKTCVAKTLDQA SYWFKANGFLLNETKTQ QMVFTLKDKLPSDDPSY VKFLGVYLDDKLSWGQ HVKYISGKLSRVIYLLRR LMDYVPEKYVRITYFSF FQSIITYGIILWGNSSCVQ DILILQKKAIRVITGSPHK EHCKPLFCKLKVMTVIN LYIYYVLIYTKRKLPETK LRENVHYYNTRRNKCIF TSYHRLSKSLNSYELVGI NYLISCHNMYKIYPNKH LNKRCMTG			
Vingi	Vingi-49_LMi	.	Locusta migratoria	GGGGTACGAAG AAGGAGCCTCC TAGTAGACTGT GAACCAATAAG CTGAAGGACAC AGTCGGGCGTA	TGATCTGGACACGG AAATGAATGAATG A	MQAIAEDKRSRWCNLV ENLEEITVVQSCREPRM KQNSMEIIKKLEWDPTD HKQKNNGVTANQIASQL LKNGKTGGKKKGKELY PKNKKMMASSIHIHSNR	cd01650	226	489

				CCCTGGAAACG		TGRRHSEYETEWPPQYT			
				GTGAAAAACTT		FTATELEDAIQSMKLNK			
				CACCGGCTAAT		AAGTDDLRTEQIKHFGP			
				CCTTCTAACACC		KVRNWLLQMINACIKIP			
				AGATCAAGTTA		KLWRQARVVALLKPGK			
				GAGCCAAAAGG		EIRILKVTDQCHFYAIST			
				AACATCAAAGA		RCWSGDPKTMAAGEGG			
				TGCAACGGCCT		GLTKTWQGDTDPKSYRP			
				TTATCAAGACC		VSLLCHLYKMLERMILN			
				AACCCTCGTGA		RVAQIIDGKIIKEQAGFR			
				TAATCTCTATAA		PGRSCCGQILNLTQHIED			
				ATATTGAGGGA		GFERKLITGAAFIDLTAA			
				TTTTCTAGTCAT		YDTVNHRKLLCKLYHLT			
				AAGAGTATTTT		GDCRLTMVIGSLLQNRR			
				ATTAGCAGACA		FYVSLNGKNSRWRLQK			
				TATGCAATAAG		NGLPQGSVLSPILYNVYT			
				CATAAATGTGA		NDQPIPQDTRLFVYADD			
				TGTATTATTGGT		TAVAVQGSSFDEVSQKH			
				GCAGGAAACAC		PVVLKNLLNTTMPITSIP			
				ATCGGGAAACC		KTSSGLEELAQYYDANH			
				ACCAGCCACAG		LKPNPCKTQVCAFHLRN			
				ACCTAAAATTG		RDAGVDLDVTWQGKKL			
				CAGGAATGGAG		DTVQHLNTLGYTIPWGT			

				TTAGTCATCGA		LDRTLSFKQHCRNTKAK			
				AAGACCTCATG		VCSRNNIIRRLTNYEWG			
				ACAAGTACGGG		AQPSVLRTSALALCVSA			
				AGTGCGATTTTC		AEYAAPVWDASAHTKQ			
				GTTAGACCTGG		VDVSVNETMRIVTGCLR			
				GCTCAATGTAA		PTPTGNMYLLAGIAPPRI			
				TATCAGCTGCAT		RRQVAADAERTKKEND			
				TGACAGAGATA		YRHPLNGHVTQARRLKS			
				AATAACATTGA		RNSFIARTKILDGSQEDN			
				AGTACTTACCAT		RVCKWKNEQTGLPITPK			
				CCGGGCCCAGA		EEMAPGNKLPYTVWRT			
				AGTTCTCTGTGA		LNRMRAGVPRCKTNLR			
				CGACTGTGTAC		RWGLLTNDDDALCECG			
				AAACCACCAGA		EVQDEAHLLTCRLLPEP			
				TATGGACTTTGT		CSFSDLLQANDKAVAVA			
				TTTCAATGAACC		NYWKNKT			
				TAACAATTTTAC					
				ATATGACCATA					
				TTAATTTTGTAA					
				TGGGCGACTTC					
				AACTGCCATAG					
				TGAAGCGTGGG					
				GATACAATGAG					

				ACTGATGAAAA CGGGATACAGC TTGAATCCTGG GCTGACAGTAA TGACCTGCAAC TTATTCATGACC CAAAACTCCCA GCATCGTTCAAT AGCGGAAGATG GAGGAGGGGAT ATAATCCTGAC AACATCTTTGTG AGTAGGAAGTT GGCAATGCAGT GTGTGAAGCTT GTTGATCATCCC ATCCCAAGTAC CCAACACAGAC CCATTATTTGTG TTGTGAATGCC ATCGTGAAGCC AGATGAGGTGC CCTTTCGAAGG					

				CGGTTTAATTTC AAAAAAGCAAA CTGGCAACAAT TTCAAGAGCAA TTGGATGAAGA GATTGAGAAAA TTCCTCCTCGAC CAGATGAGTAT GGTAGATTTATC GATCTTTGAAA GAGGTTTCAAG GAAAAATATCC CAAGAGTTGCC ACACCCGTACA TCCAAGGACTG ACTGGCAGCAG TAAAGCCTTGCT TGACAGGTACC AGAAGTTGTTC AATCAGGACCC TTTCAGCGAAG ATACGTTGGAG					

				GCTGCGGAAGA CCTA					
Nim b	Nimb - 20_L Mi	.	Locusta migrator ia	CCCCTCAGGGG CTCAGCATTAAT TTGCTGAGTAC GGGCTTGGCGA ACCCGGGGTCC TGAGCCGTGGG CTGTGTCGCGCC AACAGTCCGCT ACACCGTAAGC TCTGTGCGTGCT TCAACGACCAT CGTTCGGCGCG ACGGTGGAACG TTGTACGTTTCA GAGCCCGGGGA TCTTGGCTTAAT CGCCCGGGTCG CGAGGACGGTA CAAACCTCTAT AAAAAAACCTC AATCTCAAGGT	TAATGGTGGGTTCG TATGCCTGATCTTC AGGCATACACCGCT TATAGTTTTAATCT GTTGTCACTTTTGT CTTGTCAGTCCTAT TTCCGACTGGCAGT TTTAGTATGTGCAG AATGGCTGGCTCAT CCATTATTATCCTA GTGATCAGCCAGCC AACAACATCTGCTG TGATTTTWAGCATA GTTGATTATTTGTT CCACCTCTATTGTG GTTTCCTTTTTCTTT TTATGTTGTAATTTT CGCCCACACTGTCT TCCTGTACATCTTG TGTACTGGTTGCGG GGTGCTTATTTTTA	MALPGAKPASLSVXGPN LMLRHMTPRRSLPWPRH GRNAGREIEVRLMPPXT WYVPGLTANRLPPRSPC FLWKTLRISSGRWMACQ DEIWXCXHXDVVPSPVT GTACLVRGGGWHVKM RSGXVLITTSSPAQSRAL LACDKLGDVPVSITPHK SLNMVQGIIFHKDLLLQ XDDELRKNLERRGVHFI RRVHRGPKDNRVATGA FILAFEGDSLPEKVKVMI YRCDVKPYVPPPMRCFK CWRFGHMSSRCXGNPT CRECGRAXHPDAPCAPP PVCVNCGQHHSPCSPXC TVLHKERKIMEYKTLDR LSYSEAKRKYXKLHPVS MKTSMPQLHHTASAISA TALHCLVFHSGPPGXTA	cd0165 0	1160	1418

				GTGCTACGTGC	ACCGTAAATGAAAT	HJCPHGLWGHLNLHRPQ			
				CGAGGAGATGC	ACTGTGATCTCTTA	GPCFGSTTISTVGXSVPT			
				ATGGCTGTTGA	GCTGGGAACAGGG	PEPESQQPXLAPXARKG			
				GGTTGAACAGT	GACTGATGACCCAG	SLGSLPSKASPADASGW			
				CGCTAGCGAGC	TAGTTTAGTCCCCT	LKXPPTAGRRASRASSV			
				AACCTCTGAGG	TTAAACACCCAACA	PGDGSAEPSQHEKPKEX			
				AACCTGCCGCA	CCAACCAACCA	REKKKKKNKSKTKDSL			
				CCTCAGTTGTAT		VATVPPXHSTSSSEDEIEI			
				AAGGCTTACCC		LASPKDLAIAESSEAMD			
				AGGCAAGCGGG		VEALPQKSATAEDTVXP			
				GCTCTGCCTGG		TAPSKSACPSDQEALGS			
				GTGGACCTTTA		ASKIGDSRRYCXPYCAV			
				GTTCCCTAGCTG		EISMSQRPGGSLILQWNC			
				CTCGTGGGACG		GGFFRHLPELQQLLCRFP			
				AGAATGAAACG		AFCVALQETWFPAQRTP			
				AGCAAATA		ALRGYRGYYKNRTIYDR			
						VSGGVCIYVLATVPSAA			
						VPLHTTLEAVAVRIRAM			
						QELTICSLYLPPDEMVSL			
						TDLAALVAQLPPPFLLL			
						GDFNAHHPLWGGAQTT			
						GRGRDIETLLSQQDICLL			
						NTGAPTHFCVARGTYSA			

						IDLSVCTPSLLPHLSWGG			
						STMTFAGVTISPFFCHSS			
						TWTSGSHPMVSIAADW			
						RGYTSAAVVSVPLPPDID			
						TAVQQVTAAIVSAAEET			
						VPITTGSPRRKVVPWWT			
						PEIAEAIRDRRRRCNATG			
						AAPQSKIGHPGPSAALQ			
						RHRRRPTIENLIAFKRLR			
						ARARRLIRQSKQDCWER			
						YVASIGSRASPSRVWSQ			
						VRRLFGYRPPAGIPSLTV			
						DGTLCTDPAVIAECMAE			
						HYASCSATANYSASFRS			
						LKQRVERNPLSFHTRRR			
						ELYNARLRSGSLTLLSLH			
						APIPLRDQTAYTASSSVT			
						CQRTFITNCSPYLIGSGQ			
						GVFSHPSGGIVLLSLSTT			
						GGAKSSIFPHSPPGTVQC			
						SFTEWEFNTALAACPDT			
						SPGPDRIHGQFVRHLSA			
						DFHHKLLALFNRVWTG			

						GVFPSQWRDSIIIPVMKP GADPLAVDSYRPISLTN VLCKLLERMVSRRLMWI LEDRSLLAPQQGGFRQG RSTIDNLVSLESAIRTAFS RCEHLVAVFFDLRRAYD TTWRHHILATLYDWGL RGPLPIFIKNFLSLRTFRV RVGTSLSSXRVLENGVP QGSVLSVPLFLIAINGLA AAVGSSVCCSLFADDFC LYYSSPSIGVAERRLQGA IRRAQSWATAHGFTFSA SKTRVMHFCRRRAVHP DPELYLAADQLHVAETH RFLGVLFDNRLTWGPHL RQLKQKCWRHLNILRCL SHTAWGADRRTLMHLF KALIQSRLDYGSQVYDS ASPSALRILDPILHGGVR LATGAFRTSPVPSLLAEA GIPPLRYRRQQLLASYIA HVHATPHHPNRRLLFPN			

						TDLPSPQRRPRSGLPVAV RVHSLLSSLESFPLRPAF SCLSHIPPWTIPRLQLHL QLSCDPKASVSPQLLRH QFLDILGTFRDADVIYTD GSQVDGRVGYAYVHGD HTEQRSLPPGCTVFTAEL IAITRALDHLRACPRKSY LICSDSLSALHALDQCFP KHHLVRSIQVTVFSLGK SGRSVAFIWTPGHIGIPG NERADQLAKQAIACQRL DIGTPEGDIRLGLHRKIF EAWNSEWATMSTPNKL RAIKPSTAAWKTSLRPSR RDSVVLCRLRIGHTWLT HGHLLRREEPPECRCGA LLTVVHILLDCPLLAALR QSFNLPTALPSILADDAS MADCVLGFIRAAGFYHK I			

Penelope	Penelope-1_LG	.	Lottia gigantea	TTAAGTTTAACCACCTCACCCTCATCTCCAGCCTGTAAACCACGACGTCATAATAATGAGAAGGATAAATCCAAATGGTCTATTCCCACACTCACCAAACCAACAGTTATCATTGGATCTTCCAACATCAGCAACATCCCGTTCCTACATCCAGATACCCAAGCCGAATCTTACCCTGGTGCCCGAATTCAGCATATCAATACATTACTCTCCAAGATCTCTCCAACAACCATTCCGAAAAAGGTGTTATTTCATGTTGGCCTA	TGACCCCCCTCTCTATTTTCTAGCTCATTAAGTATAATATATTCTAATCCTTATATATAGATGTTTATCTTTTCAGAACCTATWAATACTCTTTTCTTTTAGAACTGGTAACAAGAAAACTTTCATATTATAGACTTTTGTTTATTGTAGATTACTCGTGGTTGTTTTTTTAACTTAATACTAACTATGCTAACTTCACCTTATCTTGTATTGAACATTTATCTCTTTATTTGTGACCTTGGCCAGTGCTTTACATAGAAGTTTTTAGGAACTTCTTCACTCTTAACACCAATAACCA	MGLPFKLNAGGNALKLGETVVNLSSRKLSPLEERVLDRGLTFIPTSKTLSHEPILSSVEHVNRTFKLNFFFRNRLGRGKRIPFIEKSGWTPPDSAIDSDILATCDKLHEEISHLPLVNEKHNLTKFERKALNSLKNDQNIIIRKSDKGSCCVIQNREDYIFELEKQLNNANHYRVLDQPIFMETAQKINNILKTLVNKQFISDSQYDYLSAKPEVRQRHIYTLPKIHKNPETEWSLANRIPKGRPIISDCASESYRVSEYIDHFLLPLACKHPSYLKDTTDFVEKIKSLNVPPNARFISLDVTSMYTNIDNKTGLDAVRKAFLRNPCVDRPDKEILELLELCLTRNDFAFNEKIYLQTSGTAMGKRFAPSYANIFMAEWELEISSISPFEPFFYGRFLDDIFIIWTEGPEKLEKFLEYLNSFHPNIKFTMEISKTEVNFLDTTVSIKDGKFETDLYRKPTDTHQYLHFDSCHPKHTKKAIPYSQFLRIRRICSKEEDFHKQSEELKQWFLNRGYPEKLIDKQIKKVKFEDRKELLKKKTTKNTEDKIPLVITYHPALKSIKEIIKKHWHILHLSPKLKEIFTEPPMVAFRRPRNLRDLLVRAKLPKLEKKK

				AATAATAAACA		SLTNFWQYFEILNNHHP			
				GGAATCAGGAG		SVKLTANISEQSMDFLD			
				AATCCATACAC		TTVFKGPSFSQTNKLDIK			
				AGAAGAATGCT		VFFKPTDTHQLLYKNSF			
				GGAACTCCTCA		HPKHTFTGIIKSQILRFFR			
				AACTTGCCAAG		ICTQKVDFDAACRLVFS			
				TCGAAATTTCCG		KLRERGYSKRFLRKIKN			
				TCTGCTACCCTT		DTVLDLLKINEPPPPINN			
				TACGCCACACA		FGAIPCRGSRCRTCEHIH			
				GATTCCTTACTC		MSHYFQNKKNGKIFEIK			
				AAAGGAGCTAC		ERLTCTDKNIIYLISCKK			
				AATTAGCTGAA		CDKQYVGETQNALRTRF			
				CCTCATTTAGTC		SGHRFDIIHSKNRPISHHF			
				AATAACCTCCA		NLPDHSIEDLLIMPIESIP			
				GAAGTTCAACA		DSDLIKPKTKRLQRESY			
				GTCTATTATATC		WMDQLQTVAPNGLNIQ			
				ATAAACCGTCT		HDNQMIPFVATFNHTAA			
				ACCAATGCCCA		KTSSIVKQYYSELQNRFP			
				TACTATTCCTCG		NVYTKRLVIAYKRNKNL			
				CTTCAATGGAG		SNFLVRGRLDPIESNNT			
				TGGTCAGTGTC					
				ATTTCGGATAA					
				CATCCACTGGA					

				TGTCCCCCACTG CAAAAAGTATT TATTTTTC					
RTE	RTE1 _MT	.	Medicago truncatula	TAATCCTGCCA AGGCCGGTCCC AAGCCCGGATA AAAGGAGGGTT GTGTTAGGCCTC GACAGCCAAAG TAAAACTTTGTC GAATCTCTATG ACATGGACCAA ATGCAAAATAG CGTGAATGCTA GGTCGTTGCCC GGAAGCAACGC GTTGTATGGCTC GAGTACAGTGT CAAATGAGCAA GAGCCGCTGCA TCACCACCTGG ATGTAGTGAAA ATTAAGCAAGG	TGAGTTGGATCCAA ATTTGGTGTATGAT AGAACACTATGGCG TCATTTGATCCATG TAGCCGACCCCACT TAGTGGGATAAGGC TTGGTTGTTGTTGT	MFSCRFFLIRKSDRKICL DCKVIPGESLTTQHRVM VMDVRVKRRAKRRSQN GAPRIKWWHLKGEKLRI FQHKILEGGFRQPHGSA NDMWERMAHEIKKVAK ETLGESRGFGPRGKESW WWNDSVQSKVRIKRDC FKDWSRCKNVETWDKY KIARKEAKKAVSEARTQ AFEGLYQSLGTKEGEKSI YKLAKGRERKTRDLDQ VKCIKDEEGRVLVQERD IKGRWKKYFHNLFNEGY EILPDSNRLDIGEEDRNY NYYRRIQEHEVREALKR MSSGKAVGPDNIPIEVW KSLGDRGIVWLTNLFNEI MRTKKMSDEWRRSTLIP IYKNKGDIQNCANYRGI	cd0165 0	317	567

				GTTCCCGCATCT		KLMSHTMKLWERVIER			
				TCGTGAACGGG		RLRKETRVTDNQFGFMP			
				TGCGGGTAAAG		GRSTMEAIYLLRRVMER			
				AAGCTAGTTCA		YRTDKKDLHLVFIDLEK			
				TGAGAATAGGA		AYDRVPREILWKALEKK			
				TTCGTTTTGGAA		GVRIAYIRAIKDMYEGA			
				CTTGAAATATA		STSVRTQDGTTEDFPITI			
				GGTACACTTAC		GLHQGSTLSPYLFTLVL			
				AGGAAAATCTA		DVLTEHIQELAPRCMLF			
				TGGAAGTAGTA		ADDVVLVGESREEVNG			
				GACACAATGAC		RLETWRQALEAYGFRLS			
				TAGGAGGAAGA		RSKTEYMECNFSGRRSR			
				TCAATTTCATGT		STLEVKVGDHIIPQVTRF			
				GCCTACAAGAA		KYLGSFVQNDGEIEADV			
				ACTAAGCGGGT		SHRIQAGWLKWRRASG			
				GGGGAAAAAGG		VLCDKKVPLKLKGKFYR			
				CGAAAGAATTA		TAVRPALLYGTECWAV			
				GATAGTTCAGG		KSQHENQVSVAEMRML			
				ATTTAAGCTTTG		RWMSGKTRHDRIRNDTI			
				GTATACTGGTG		RERVGVAPIVEKLVENR			
				AAGTTAGATCG		LRWFGHVERRPVDAVV			
				AGAAATGGGGT		RRVDQMEESQVKRGRG			
				AGGCATCTTTGT		RPRKTIRETIRKGFRGQ			

				GGATAAGGAGT					
				GGAAGAAAGAT					
				ATTGTAGATGTT					
				AAAAGGATAGG					
				AGATCGGATTA					
				TAGCCCTAAAG					
				ATTGTAGTGGA					
				ACAAGACACCT					
				TTAATGTAATTA					
				GCGCTTATGCA					
				CCTCAGGTAGG					
				GTTAGAAGCAC					
				ACCTTAAAGTG					
				AAGTTTTGGGA					
				GGAGCTAGAAG					
				GCTTAATTCAG					
				GATATCCCGTTA					
				GGAGAAAGATT					
				TTTCTAGGAGG					
				GGATTTAAATG					
				GGCATGTAGGG					
				AGTGTTTCGAG					
				AGGTTTCGAGG					

				GTTTGCATGGG GGGTATGGTCT CGGGGAGATTA ATGCGGAGGGT AAATCCATCTTA GATGTTTCATCG GCTTTTGATCTT ACTATAGCTAA TACTTGTTTCAG GAAAAGAGAGG AGCATCTTATCA CTTTCAAGAGT GGGGTCTC					
RTE	AviRTE_MUn	.	Melopsittacus undulatus	TCTCTTTCACGG ACACCCCCTTGT CGTGGGGAAAG AGCTTGCGTGC CCTTGTGAGGTT GGGAGCTATGC TGGTGGTGGCA TATGCCTCTAGT AGGGTCTCCCA TGCCAGACAGG	TAAAACCCTTCCCA AATCTTCGTTCGCG AAGTCAAGCCATG	MGIFCCGRLKTKRQRHT PASKGVSRLTLACWNIR TMLDTANSGRPERRSAL IAHELSRLNIDIAALSEV RLHEEGSLKEHGAGYTL YWSGKSKTESHLSGVGF MIKNSIVSKLENLPIGHS DRIISLRLPLHNKQHVVL FSIYAPTLQADPAEKDKF YSDLRRLTQKVPIDDKIII	cd01650	563	847

				TCTCAACTGAA		LGDFNAQVGKNFEAWK			
				GGGTCAGACAA		GVLGKHGVGSCNDNGR			
				AGTGTGTCCAT		LLLEFCAEQQLTITNTIF			
				GGGCAGGATGG		QQKDSXKTTWMHPRSK			
				GCTCACTAGCCT		HWHLIDYVLVRQRNVR			
				TTGGGCAACCA		DVCHTRVMPSAECQTD			
				TCCTAGGAGAA		HRLVRCKLNFHFKPKPK			
				GGACAACTCTA		RGGIPRRRLNVNNLQIA			
				ACTTCAAACCC		AVRDSFQANLQTRLEDH			
				AGGCAGATGGA		LIDSSPEALWQHIKNSIL			
				GCTCGTTTAGCC		QSSVESFGFSLKKNKDW			
				CTGTAAGGCCA		FDDNNQEVQELLTKKRS			
				TCCATCTAAGA		AHQAHLAQPLCHVRKA			
				GAAGGATACTC		AFRLACSKLQQKLRDIQ			
				TAATCAAACCT		NKWWLNLAEKTQLCAD			
				ATGACCTGAGG		LGDHRGFYEALKAVYG			
				ACTTTGCTGTCA		PTYQVQSPLLSADGQTL			
				CCGTCCAAGCT		LTDKISILNHWSEHFQTL			
				CGCTAGGCCCT		FSANRVVQGXGIQHILQ			
				GGCAGATAAAC		HLVKNELDIAPTMGETL			
				CTCAGGTGTAA		KAIQQVKTGKAAGVDGI			
				AGGGTGGGGCC		PPEIWKHGGQALHAKFH			
				AGTTCTGCGCAT		ELVVRCWEQGKLPSDFR			

				GCTGTGCCTCAC		DAIIITLYKKKGVKSDCS			
				CTAAAAAATCC		SYRGITLLSIAGKILARIL			
				ACTGCACAGGC		LNRLVPAIAEELLPESQC			
				CCGAAGGGTTT		GFRXNRSTTDMVFVLRQ			
				ACCCACACTTG		LQEKCREQNKGLYVTFV			
				CAAAGCCCTGT		DLTKAFDTVSRKGLWLI			
				AGCGACAGGCG		LERLGCPPKFLKMIILLH			
				AGGGACGAAAC		EDQQGQVRYGDALSEPF			
				GGCAGGTGATA		PITXGVKQGCVLAPTLFT			
				AGGTGCACTGG		IFFSMMLQRAMVDLDEQ			
				AAGYCGCAGAC		NGIYIRYRTDGSLFNLRR			
				CCAACCCTGCA		LKAHTKTLNHLVRELLF			
				TGCAGGCGGTT		ADDAALVAHTEAALQR			
				CAGGACATTGG		LTSCFAEAAELFGLEVSL			
				TCGCTCGAGAC		KKTVVLYQPAPQEDYH			
				TGACCCAGGAG		HPHITIGESELKSVQQFS			
				ATGACAGTCTT		YLGSIISSDATIDKEIDNR			
				GTTTGGCAACA		IAKAYRAFGKLHKRVW			
				TCCTGAATGAC		CNKHLKKSTKISVYRAI			
				CAAGCAGCCTT		VLSTLLYGSESWVIYRH			
				TTTTCAAGGAG		HLRLLERFHQRCLRTILN			
				AGCACTGCTTG		IHWTDYVSNVTVLEQAG			
				CTCCATATGGA		VTSIEAILLKLQLRWAG			

				GAGAGGCCTAG AAAAGGTGGCC TAACCAAAGCT CATCTCCCCCTC CACCTGGTTGG CTAGCCGCGGC CA		HVSRMKDHRLPKIVFYG ELATGCRKRSAPKKRYK DSLKQYLSLGHIDYHQW STLASNQETWRHTIHNA VVSFENTRRISLEEKRQR RKNRAPPIPPKETFRCAF CNRICLSHIGLFSHQRAC SKRG			
RTE	RTE-2_MMa	.	Mesobuthus martensii	GCCTTAGAGGA ACCGACTAAGA TCAAAACCTAG TGTCCAATCCGT ACCTTCACGTG GTGCAGCTGTT AGTACTCAATG TTTCTTTAACAT GAGGAACGGAC ACGGATGGGCT CACACCAACCC GCATTGGCCAG CGTGGTGTGTTT TGGCTAAAACC TTCAAGCATA	TGATTATGATGATG ATGA	MNSSTNLRPSVPGGWN VPADGSKQVYRKAEGA KNHRVSCHGFYLATYN CRTLSTEGKLLEIEEELK HVKWDILGLSEVRRKGE ELLELKSGNLLYYRGTN DGRCSGSGFLINKRIKDN IIEFNSSSDRVSRLVLKIS KRYEFQIIQVYAPTTSHT DEEVEEFYEEIEKIRGRK QQFTVIMGDFNARVGIQ LQGESAVGKFGEGIRNE RGQRLVEFAESENLFIM NSFFKKNPKRKWTWRN PSGGTAEIDYILASKKYL	cd01650	561	814

						IEDCTVLNKFNTGSDHR			
						MVRCKVHINVKNERKK			
						MFKGPRSGIDIGKLKAN			
						QNLFQIELKNRFAELKA			
						NQNLFQIELKNRFTELTN			
						LTEDIDIKNLNTNITNILI			
						QTATEVAGKQRQVKND			
						KLSDKTKLLLNKRREMK			
						INGNAVSKKIEYAELCK			
						TVRKCITDDIRKYNCNL			
						VKKSIEENKGIKSVKKH			
						LCCGRKRIFALKAKDGK			
						IIKNQEDILKRVEEFYSD			
						LYRDTNNSCLVQNSKSE			
						IPPILKEEVHLAIKKMKK			
						GKAPGEDGITVDILHEG			
						GEQIEDILTDLFNRCLKQ			
						KDVPEDWKNATVILLHK			
						KGEREDISNYRPISLLSV			
						LYKVFTRILADRLKRSL			
						DEAQPREQAGFRAGYST			
						MDHIFVLREIVERTNEYE			
						IPLCLAFIDYEKAFDSVM			

						TSAVLNALANTDTDSGY VDIFSKIYSQASARIRLH KYSEKIKIGKGVRQGDTI SPKLFSAVLEEVIKKMD WDKTGLKIDGEYLSHLR FADDVVLLASNPNELQN MIQSLATESKKVGLRINI SKTKVMYNNFADQQKI KLDNIVLEEVSGYIYLGQ LLTLDGDLMPEINRRIKL AWSAFGRNNSILKGKMP LCLKRKIFDQCILPVLTY GSETWTLTKRAIQRLQT TQRSMERCMLGITRRDR KRNTWVRSQTKVIDVIQ KIKKQKWTWAGHIIRIN DNRWTTKVLHWIPRDK KRPRKRPHYRWRDEIVN FTSKCWIREAVDRKNW KLLGETFILQWIDNG			
RTE	RTE-3_MMa	.	Mesobuthus	AACTCCGCCCAAGAAGCCATATAGCTGACTGGC	TAGAAATGGGAAGGAGCTTGATGATGATGATGA	MMSGTHDSDNARAFPGVGDAQSLRRPRAVRSRLGLSHQERVRLKKLARRP	cd01650	567	814

			martensi	CAGTCCCAAGC		LLVRIASLNVGTMTGRG			
			i	CTGGATAAAGG		REIVDMMERRKIDVMC			
				AGGAGGGTTGG		VQETRWKGDKTRELGE			
				CGTCAGGAAGG		GYKIFYSSASKEGRNGV			
				GCATCCGGCCG		GIITSRKMKENIIEVSRRN			
				TAAAAACACCC		DRIMRMKLDIGETIHVV			
				TGCCAAGTCTTT		SAYAPQVGTSMEMKDT			
						FWKTMDEELRQLEEKD			
						RVIIGGDLNGHIGTNKRG			
						IERIHGGHGFGDRNQEG			
						EQIIDFAVAFDFALANSF			
						FNKKEEHLVTFKSGGNS			
						TQIDYFLYRRKHLKEVK			
						NCKVIPSDHVATQHRPM			
						LMDLHLLRPAGREKVKT			
						RVKVRKTKWFKLKDTV			
						LRSRFKETALQRVSMDF			
						EDVQEWWNRNAEVIRC			
						TGREILGETTGKCWENK			
						ESWWWNQEVQAAVQE			
						KRKAKKKWDSTHAPAD			
						LEVFKVRNKQAKKKVA			
						QAKAKACQELYDDLETP			

						EGQKKIYSIAKHRSQTA KDITHIKQIKNVNGEILR NTGKIQDRWKEYFEELL NVENERKYHEDGIPNEG LTRDITKSEVKVALKQM KNGKAVGLDGIPAEAW KALGEEGLDLLWDLLG KIYREEEIPDEWRESIIVP IYKGKGDIQMCGNYRGI KLMSHTMKILERILDKR LRAEVKIGGQQFGFMPG KSTIDAIFCLRQMMEKF REKQQELHLVFIDMEKA YDRVPRQEVWRSLRLR NAPEKYVRLIQEMYRKT TTKVRTCLGETSSFEVG VGLHQGSALSPYLFNTVI DVLSEKVREEAPWNML YADDVVLCARSRAEVEE KLENWRKALEDRGLKIS RSKTEYLACNIGTQENE ELKLMETTVRKTDHFRY LGSIVDELSETGREVNNR			

						IQAGWKNWKEVSGVLC DKKIPVRLKGKVYRTVV RPAMLYGTEALPLKKSN LQRMKVAEMRMLRWM CGVTRADRIRNEHIQGT VKVTDIASKIQESRLRWF GHLMRRDEDYTGKRLM NIKVEGRRKRGRPKRRW QDIVDDDLRSKGLRREW TKDRTQWRKLIKNSDPI			
RTE	RTE-5_MMa	.	Mesobuthus martensii	GGAACGTCCAG ACACTATGGGC CCCAGGGAAAC TCGAACTGTTAC GACATGAA	TAGCCCCAACGGTT ACCCAACCATGGGA AGATGAATGAA	MKNFTFDVLGVSEVRW TGKGELSTGDFIWSGEE KHHFKGVGFLLSKKAQ KCLISYNPISSRIISARFSA TPFNISIIHVYAPTTDASD EDIESFYGEIKNTLRETP KKDFTIITGDWNAKIGSN NTGWESAMGKEGYGDR NERGERLLEFATEHNFYI CNSRFRQKPCRKWTWT SPDGIHHNMIDLILVQSK WKTSVQNCRTFQSADIQ LDHSLVLCNIKIRLKKQS	cd01650	498	754

						TTSQGTLDFRKRNSNDL			
						KQPPIKQAYEKSLNEQL			
						KKITPTMSLDEHASILES			
						AIKTAIKSSLAVTEKPRK			
						PWITENTLKVANEKRK			
						WKLKRHISEENNNIYKA			
						HCIQVKKLAKEDKENW			
						LNEQCEAAENGLKTGKT			
						HQAYKLIKMLRKPFTPK			
						LNILKDDAGKILQTKTAI			
						TNQWTKYCSTLYNGSN			
						NNQSQATDAELEEITPPA			
						DEYPCPHILHSEVEAAIR			
						TLKQRKSPGFDEITAEMI			
						KAGGPKLTASLHELCQK			
						IWLEEKIPTNWGKTILIPI			
						PKKGDLARCENYRTISLI			
						SHLEKVFLIILLRRLQHQ			
						LEPYLAEEQAGFRKDRS			
						TVQQILLVRLLAEKAKR			
						KGKKIYNCFIDFQKAFDS			
						INQQRIWAVLASFGVDR			
						KVVKLLQNIYENTRLAV			

						KTGNDMGTWFKMLIGS RQGDPLSPLLFITYLERT MDSVTKLNLGVTIQGSII SNLRFADDINLMNEDLQ ALQDQATALQASAEKSK LRINTSKTKTMVFGARNI EQQIEINNQPVENTEQFE YLGSIITWDNNCSVEIKN RIAKATGALASLKHIWK SSKITVPTKLRILSVCVFS VLLYAAETWTLKENDK KKLLAFEMRCYRRLLKI KWQDKIRNEEIRNTLNIH TTIIDLIRQRKLRLFGHIC RMDNNRJIKHVVTSXMN GKNKQGRPPRAWMDDI TDWCKCSAQEAFHRAQ NRNQWKTATRTND			
RTE	RTE-1_MMa	.	Mesobuthus martensii	TTCTTCGAGAAGAAAGAACAGTGAAAATGGACATGGATCAGCCC	TGAAATAAGAAATCTGATGATGATGATGA	MRLTVXLANSNCTIKDVSVLNQFNTGSDHRLVHAKVKXDIQCERXRLXQHGGQKVKKEEYSLKLQTDSKCSQKTKVXYGIHTMR	cd01650	175	448

				CAACGGACGGA CAAANA		QCRRNRGNSTKKSRQEN IHQNTGADRKEKEDEDQ IHYQPDRIEDEVRKSPRR RQFNIKLLKEGGVPLYK QLAGIFSDCMRTGQIPDS WKNAVIVPIHKKGDKTD LRNYRPISLLSTIYKVFTI ILTRRLEQVLDSNQPREQ AGFRKGYSTMDHIHTIN QLKEKCKEYHISFCMAFI DYEKAFDSVSTEAILTAL SNQGIETTSTQNXCCKD SGIAPRPPSVRRNLLRIIY EEATAIIRLHKDSRKIKIQ RGVRQGDSISPKLFTACL EEIFRQINGWEHKGININ GEYLNHLRFADDIVLFA QTPNELQEMLSDLKTES EKIGLQINYEKTKVMFN AEVIPQPVTIEGRTIERVE EYKYLGQVITTQGDQM REIQHRTQLGWHACGR YSNFLRDNRIPICLKRKL			

						FNMCILPVMTYGCETWT LTKKMEQKLATTQRSM ERYMLGFTRRDRKTNIW IREQTKVEDILVTVRSSK WHWAGHIARYTDNRRT KRITEWQPRQGKRYPGR QYKRWRDDIRKVIGSQ WMQIAQXRYIWKQKEA YISSSG			
RTE	RTE-4_MMa	.	Mesobuthus martensii	TGGGGAAAGCA ATGGTGAACTA CCCTGTAAAAA CCTCACCTCGA AATTCTATGGAT CAAGACCAAAG TGTTATCGATAT GGTGTTGGACG GTGGGCCCCCG GGCTGGACAAC ATCGACCATGC TACTGGGGAAG AGCAAAGG	TAGGATTGGATTGG AT	MIRSSQAENDARGANPP GGPIADVATGKKDHLSC KDLFTIGTWNVRGLNIG KLEIIKREMTRTGLDLLG ISELHWTGSGDLKSDDY TVHFSGNDHNRRRGVAF VANKRLAKYIENARCIS DRIMTIRLRGKPLGVTII QVYAPTADADKQDIVDF YTAIQNALDQTPKKDIIY IMGDFNAKVGDGEDGQ VVGRHGLGDRNDAGDR LVQFCQENKLRITNTFFK QHKRRLYTWTSPNGQH	cd01650	537	789

						RNQIDYILCQQRWRGSIT AAKTLPGADCGSDHQLL IARIKIRLCKVKKTVVAK KFDVSRIPTQYTIEVRNR FEALSLDGGQTEELWA MIKGTIIETALEYVPYKM PTKGAVWLTEKTLQIAE KRREAKAKGEIDKARRL HADFQREARIDKNNALN ARCISLEEAFRKGRTREL FATVKQVKTPFSARQSTI KDRDGNELTDQQDIKER WREYTEDLYTSIDDQQN EDQTDQGLQEPDLLESE VEWAMRQLPSNKAPGID GIPAELLRPIPRAALTAL CQQIWKSCTWPADWRR SVFVPIPKKGDTRDCAN YRTIALIPHASKILLKVIQ KRLGNVIDRELPDVQAG FRRGRGTRDHIANLRWI MEKAREYQKKLYMCFI DYTKAFDCVEHSKLWV			

						ALRELGVPEHLIRLIRSL YQNQEATVRTRYGDTD WFKIKKGTRQGCTLSPF LFNLYAEVIMRKLDLEE SEVGVKIGGRTISNLRFA DDTTLLAESEKDLKRMI MKLKNESEKMGLQLNIK KTKIMTTAAKEVHIKIN NEEIEVVDSFIFLGSKVD CNGDWRPEITRRIALGR TAMVNMDPIWRCKDVSI TTKRRLVTAIVFPIATYG CETWILTKADRRKIDAF ELWCWRRLLRIPWTAR VTNKSVLERINPDFSLEG KITRQKLTFFGHVMRAN SLEKDVMLGRMSGTRPR GRPKTRWLDTIKKDTGK NIEELRELVRDRKTWRD MTHRIAKNRTRLNG			
Loa	Loa-2_MMa	.	Mesobuthus	T	TAGCCATCATTGTG GTGGGAAAATTTCC CAGGACTGTATCTG	YDLALIQEPWHYKGQIR GLNCKEGKLFTFTSDSG PRTCVFVRNNINVSPVW	cd01650	468	649

			martensi i		AGGTCCTTCAGACA GCTCTCAATACAGT CTCTAAGTGGTGTG ATAGGGTTGGATTG TCTGTTAATCCAGC TAAGACTGTCATTG TACCTTTTACACGA AGAAGACAATTGG CGGGGCTAATTGAA CCATCATTCTTCGG CAAAAGGTTGCAGT TTTCTACCCAGGTC AAATACCTTGGTG	DFCSRDITTVRVKLDEG GVSRELIVSSAYFPYDKD ELPPTPEFXLLVEFCQKN HFDLLVGVDANSHHVV WGSKDINRRGEYLLEFL MGTNLFVLNKGNEPTFV VVNRREVIDLTLGTQNIS DLIENWHVSGEPSSSDH RYICFNLQGMVPSQVTY RDPRRTNWASFREDLKF NLGGIQSRFCSVVDLEM TVDQIQQAILLSYHQNCP VRTARSSRQVPWWNEQ LGGLRXNVRRLFNKAK RTGDWDTYRSALTCYN KAIRKAKRSSWKRYCQ GIEDIPDVSRLVKIMSSG NFCKVGSIRLPDGSMTST CKDSLKELFRVHFPGSTL IEGSSCLQVQSDLDGHV YHAKREDWGLVRGVIN QSNIRWAVSVFKPFKSA GPDGILPVFLQQGIDILA			

						SPLCRVFRACLAFGYIP WSWRQVRVSFIPKPGHS SYTEAKSYRPISLTSFLL KTMERLVDRFVRDGAL LKTPLCRSQFAYQSGKS TENALHEVVSLIERAIGY KEIALGAFLDIQGAFDST SFEQISLAASRHGVCSTII RWIGSMLKHRVVYSTLY DEEVKILTDRGCPQGGV LSPLLWCLVVNELLTLH SMGKVSSLLGMLMT			
RTE	NAV IRTE 1	.	Nasonia vitripenn is	AGAAAAGTAAG AAGGAAAGAAC TGGAGAAGAGG CAGAAGAAAGA AACGCATTTGC AAAAGGAGGGA ATGTGAAAAGA TCACCAATTGA GAAAGAGAAGA TGGAAGAGGAG GCAATCGGGTC	TAAAAGAAAATAT AATACAAATAGAG AAAAGAGCAAACG AAAGAACGWAAAG AGCAATCGAAAGG AAACNAGAGCAAT AGAGAAGAGGAAA GGAAAAAACAAAY AAGGAAAGAWAGA AAGAACCAGACGA AGGAACKAAATAG	MRAAFDKVNRREMWE MMEKLGLGVRIRNRVK ELYRGTRSEIRIGEDKIG GFELHKGVRQGCPLSPT LFNVAMADTEKELSKV QEGGIVLGKKKFWSISY ADDVVMVANNATGLKQ MLKKFKKIIERKGLELNT EKSKIMVFKNGGRRKKE DSFFWEKKEIEVVKNFE YLGYMLKENGKEEAQIK	NA	NA	NA

				GTTAAAGTGCA	AGCGCAGGGACATT	KLKEKASTVMSSMWGL			
				CTATGAATATA	TAGAGAGCAGGGG	GEELFRDNWELRMRLFD			
				ATCCTACAAAA	AGACGACGGAACA	TMVKSVMEYGAEVWG			
				AGTGGAGAAGA	GGAGGAACAGTGG	WKGKEELEKIQRKYMK			
				TGAGTATGAAA	ATGAAACAAAAGT	WVMKLDRTTPEHVLHL			
				ATGGATAAGGA	CACTAGGWAWGGA	ETKRYKLETRTRRRAIK			
				ATAGGTAGAAA	CAAGAAAACGATA	YEVKLSKAKEGTLWRE			
				GGGATAAGATG	CAGGTGGARAGAG	CWRTLEKEEEKERRRGX			
				ATCGAGAACTG	GAACGTGGACGAG	RRXNREEELNKRGWAL			
				GGAACAGAAAT	ACAGAAAAGAAAA	SEYHRRLRAGEQVWSEL			
				ATATTGGACTG	CACGTGGTGAGCCA	ERIEKDIQWQKWESEMR			
				GAGGAAAAAAT	CAGATAGAAAAGA	DSRQARDVKELIKPQGE			
				GAACAGAAAAA	GGTTAGTAAAGATA	TPEYLRRTGNTGKKGLA			
				TAGAGGAAGGG	AAACAAGTAGTGA	MTARFRMGNEARGYKY			
				CTAAAAGGTAT	AAGGAGAAGATAA	WMKEEERICRLCREEEE			
				GGAAGGAAGAC	AAAAGTGAAAAGG	TFEYIFSRCNRTGSRENN			
				TGTGAGAAGAA	AGAAGATGTACGG	WKNILNGERKNLAKLY			
				GGAGAATAAAG	AGACAAGAAGAAA	RVIWVRKEKEKEEEX			
				GAAAAGAATTA	GAAAGAGAGATGA				
				GAAGAGAATGC	GGGAACTGAGAGA				
				AAGGAGGCGAG	GGAGGTRGAAGGA				
				AGAAAAGGAGG	TGGAAGATAAAGG				
				AAAGAGAGATG	CGAGGTTTGAGAGA				

				GAAAGAAAGAA	GAAAAAAGAGAAG				
				TGGGAATGAGA	CCCTGGAAAAGAA				
				ATGCGAAAGAA	AATGCAGGATGTGA				
				AAGAAGAAAAA	GAGAGGACGTGGA				
				AAGAAATTGAG	GTTTTTAAAGGGGA				
				GAGATATTAGG	AGATGAAAAACAT				
				GAAGATAGAAA	GCAGAAAGAGATA				
				GTCTGGAAAAA	GATGAGATGAGAG				
				GGGGAAGAGAT	AAGGGCTTGCAGGT				
				AGACAAGGGTA	AAGAAAAATAGAG				
				GAGAGAGTCAA	TGGAGACGGTAAA				
				CCAGGGGAGAG	GAAAAGGCAGGAG				
				AGAGGATAAGG	ATGAGAAGAGGGA				
				GAAAGGTAACG	GAAGAGTAATAGG				
				GAAAAGGAGAG	AGAAGAGAGAATT				
				GGACAGAAAGC	TTGAGAGAGAGAA				
				AGATAGAAGAA	AGGCTGATGCGGCC				
				ATGGAATGGAG	AGTACAGATATAGG				
				GATCGAGGAAG	AAAGGGGACAGCA				
				GAGAGAGGATA	GGTGCAAGGTGTGG				
				GGAAAAAGAAT	GAAAGAGGCCTGA				
				AATATTATAAT	TAGTACACCTAAAC				
				AGCAGGGTGGA	AATTGGACGGAGG				

				AAGAAGAAAAA	AGGATTCTATAAAA				
				TGGGATAGAAA	GACTGGCTGGAGG				
				GACGGTAGAAG	ATATAATGGAAGG				
				ACTGGATGAAA	AGTGGAATGGGAG				
				AAACTCGAAGT	ATGGAAGATACAA				
				ATAGGTTACGTT	GGACGGAGGACAC				
				GAGGAAAACGG	GAAAAAGGTAATTT				
				TCATTGAACGG	TTAAGAGAAARGA				
				GAAGATCTAAA	AGACATGGAATTGA				
				GAATAGGGGCG	TATGGGAGAAGAG				
				GAATGTAGAAA	GGCGGAAATAAGA				
				CTGGGAAGAAA	GAAGGAGGCAAAA				
				AAGGAGATATC	TCAGGCTGGAGCAG				
				ATGAGAGAGAA	TGGCTATCMTTCGA				
				GGCTAGGTTAC	TGAGAGAAGGGCC				
				AAGGGTCTGAT	AAAGCGCTGATTAT				
				ATATTTATAGAC	GAGAGAGAAAAGG				
				AACGATCTAAC	AGAAGAGAAGAGG				
				ATGGAAAGAGA	ATGAGATCGAGGA				
				GAAGAAATAAA	GGGCGAGGTCTCGC				
				GAAAAACTGAG	ACAAAAAAGAATG				
				AGAATACGCGA	GGTAGCTGAAATAG				
				CAAAGGAAAGA	AAGGCGAACTGTAC				

				AAAGAAGGAAA	AAGTGGGACGAAA				
				AGAGGTCAAGA	AGAAGGAGAGAAT				
				TAGGATACAAC	TGTAAAAAAAGGA				
				AAACTATGGAT	AAAGGGATAGAAC				
				AAAAGGCGTAG	KGGGAGAAGAGGR				
				AGTATTTATGG	CAGATATAAGAAAT				
				AAGGAAAGAGA	GGAGGTCGGGGGA				
				ACAGGAATTTTT	GTCAGAGGAGGAA				
				TCGGAGAAAGA	GAGGAAGAWTATC				
				AAAGGACAGGA	AGACAAAGATGGA				
				GAGGAAGAAAA	GAGAAAGGAAGAG				
				GAGGGGAGCAA	TAGAGTTTACTCAC				
				AGGAACTTAAG	AAACACACATAAA				
				GATAATTAATT	CACACACAAACAAT				
				GGAATGTAGCG	CAAACACACACAC				
				GGAATAAAGTG	ACACGTACGACTTA				
				AGTAGATGAGA	TACACACGCACACA				
				GGACCTGGGGC	CACACACATGCACA				
				TATCTAAAAGG	AACRGACACGAAA				
				GTTTGATGTAAT	ACARGAAAAGMAG				
				TTGTCTGCAGG	GAGARCRCACAYA				
				AAACATGGCTG	CACAYACACACATR				
				GAAGATAAAGA	TACACASGAACATT				

				AGGAAATTTCT	ACCTACAGAGACAC				
				GGGAAAAGAAA	ACATAACACACCAG				
				TTAGGAGGTTA	ATTAGGTTAAAGTA				
				TGAAGTCAGAA	GGATTAAAGCAAG				
				TTAGAAATGCA	AGGTATGTAAATAT				
				AAGAAAGGGGG	ACGAATGAATAAAT				
				AGAAAGAGGGA	ATTGTAAACACGAA				
				GAGTGAAAGGG	TAGTGAAATACAAT				
				GGAATAGTGAT	AAAGTTATTATTAT				
				GGCGGTAAGAC	TATTA				
				AGGGAACGAAA					
				ACAGAGAGAAT					
				AGAGTGGATAG					
				AAGAAAAGACA					
				AACGAATTCAT					
				AGGAGCAAGAA					
				TATTAGGAAGG					
				AGAGATATATG					
				GTGGATAGGTA					
				CGGCATACATG					
				AGAGAGGAAAA					
				ACAAGAAAACT					
				TTAAGGAGATA					

				GAAGGGATGAC					
				GGAAAACGCCA					
				GAGGTGAGAAA					
				TTGATATGGTGC					
				GGAGACATGAA					
				TGCAAGAACGG					
				GATAGGAAGGA					
				GGAGGAGTCGA					
				CGAGGAAGAAA					
				ACGAAGAGAAC					
				AGGGAATCGAA					
				AGATGAAAAGA					
				TAAACAGAGAA					
				GGGGAAGAACT					
				TCTGGAAAAAA					
				CAAGAGAAATG					
				GGACACAGTAT					
				ATTGAATGGAA					
				ATACCGAAGGA					
				GACAAAACAGG					
				AGAATACACGT					
				ACATAGGAGGA					
				GCAGGATGCTC					

				AGTGATAGATT					
				ATGTGATAACA					
				AATGAGCAAGG					
				AAAAAGAGAGG					
				TACAACGCAAT					
				AGAGATAGAGC					
				TGGGAAAAACA					
				CTCGGGAAAAC					
				AGATGAAGAGA					
				AAGACAAAATC					
				ACAGAAAGAGC					
				TATATGGACTG					
				AAAAAGCAATA					
				GGTCAATTCCG					
				AGAAGAACTAA					
				ATAAAGGGGAA					
				GAGGAAACAGA					
				ATGGACAGAAC					
				TAAAAAGAAAA					
				CTAGAAAAAGC					
				AATCAGAAGGA					
				AAAAGGGAAGC					
				GGCAGAAAAAG					

				AAAGAAAAACA CGTGGTGGGAT GAAAAATGCAG ACTCAGAAAGC AGAGGTAAGGA AGTTAGGCAAG GCGATTAAGAA GGGAGCGGGAA GGAAAGAATAC GTCAAAGCAAA GAGAGACTGGG GAATATTAGTA GAAAGAAAAAA GCAGGAAGAAA TGGAAAAAAGA ATAAAAGAGGC TGGAGATAGTA AAACCGGGAGA AAATTCTGGGA GGTGGTAAAAA GTAGGAGAAGA AAAAAGAGGAC AAGATGCAGCA					

				GTAAAATAAAG					
				AAAGAGGAATG					
				GTTAACTYACTT					
				CAAAGGGCAAC					
				TAGGAGAAGAG					
				ACAGAGGAAGA					
				TTTGGGAACAA					
				GCAYACAAGGA					
				GAGGATGAAGA					
				AGGAGAKAGAG					
				AGGATAAAGAA					
				GRAAGAATAAC					
				CGAAAACGAGG					
				TGAGGAAGGCA					
				ATTAGAAAAAT					
				GAAGAAGGGAA					
				AAGCGCCAGGA					
				GCAGACGGACT					
				ACAAAATGAGG					
				TATGGATACAC					
				GGGGAGGAACA					
				GGTAATAGGAG					
				AAATAACTAAA					

				ATACTGAATAA					
				CATATGGAAAG					
				GGGGAAAAGTA					
				CCAGAGGAATG					
				GAAAACGGGAC					
				TAATAACCCCG					
				CTTTTCAAGAA					
				AGGGAAGAAGG					
				AGGAAGCAAAG					
				AACTACAGAGG					
				GATAACTCTCAT					
				GGACACAGGGT					
				ATAAATTATAC					
				GCAGAAATAAT					
				AAGGGAAAAGA					
				TGGAAAGGGAG					
				TTAGAGGGAAG					
				AGAGATGCTTG					
				ACGATACACAA					
				ATGGGTTTCAG					
				AAAAGGGAGAA					
				GGTACGGCAGA					
				CGCAATTTATAT					

				TTTGAGTAAAG CAATAGAAACT GAACTGGATAA GAAGGGAGGTA AAGTCTACGCG TTTTTCGCGGAT					
CR1	L2-3_NVi	.	Nasonia vitripennis	TGGTGAGTACA GTTGACAGCTCT GCTGATCCATC ATTTTTCTGTAC TGCTTTTCTGCT GCTTCTTGCCTT CGGGCCCTCTTC TGTTGTAGATA GTTGCACGCTCC TTTTTGCAACCT TCACGTCGATA CATAAATCGCC GAAGGCTGCAA TTTTTCTCTGTC ATCTTTTCTACT GCAAACTTTCTA CACGGCTCCTA	CTCGTACACACTCC ACGCACACGCACAC ACCTACTTTCTCGC TCATTCCACCTTCA CTATCGCCAYACTY TCACACTATACATA CTCTAAACATACCT CAATCTATTGTTAT TTCTTTTCTTACTTT GTAATGCTGTAAAC TTACAATCGTACAA TATAATGTACGCAA AATAAAACTAATCT	MEALKTQNIVSERISKIE QRLGPLEKRLKALDELP ALKTRIHNAESTITELQA QIQDLSSRSPTMQQDNG STVPNTAEICSLRSELAE VKRRQEQTSNCVVVVT GLHYTRETSLHLLAFSV VNALDPTVLRRDVASVR TMGRLDATNSSARGDG RLPPLAVTLSSSALARSI VIAKARKRKLHTSELDA TLLEEAKALSPDHQGLIN INELLPSDVHKLRTRARL EAKKRQGCRTFVREGRL YTCAATMTASVLRSSTP TPSWRLFPGFRPLPTSSN TEATXQYTHQTLSSSAQ	SSF56672	671	971

				ACCTACAAATC		ERSKVKSEYLFCEVSAK			
				TGCTTGGCTTGG		GVSPIFVGVVYRPPHAPF			
				CTTGGCTTGTTT		FQGSNFIDQLTTHMHNY			
				CCTCACTTGCAC		STKVIMGDFNSDQLSSSE			
				GGCGTCTCTTTC		DAKFIKAFIDENSLSSVP			
				TCTCCGCAATCA		YGATHHKQGSDTWLDL			
				ACATCGTGGCG		CLIDEQDRLLSYWKTNS			
				ACATCTGCTGC		PFINGHDLITATLDVQIP			
				ATCGTGGGAGA		RYVPNTYSYRNFKGISA			
				ACTTTCAAGGC		EKLRDFLSACDWSSLTS			
				GACTTCACACT		SSLDECISTLNANLTNAI			
				GATCGAGGTCG		NHLAPLRTVTPRRQRHP			
				GTACTTTCACGC		WFTTVSHVXXXXRILVA			
				TCTTTCCACTAA		LRDLVSERDRLYRRFRD			
				CGCTCCATCTAG		SRLDSDLRIYRLARDNA			
				CCGTAAGAGTC		HKQVEEARLNYYYSRLS			
				ACAACACCTAC		TLTDVAEIWRELEKLGIS			
				TCTTTCACTCCG		ATKAPSPSRFTTDELNK			
				TCGATAACTTAT		HFSSISNDPLAPAVEDYL			
				CATCCTCCTACA		LTLESLDLPEHFEFSTITE			
				AATATCTGCTTT		SDVLAAVSHFDTQARGS			
				ACCGACTGCTA		DGIPQVVISKALPVLAPL			
				GTAAGTTATCG		LSHIFNLSLSEPCFPSAW			

				ACGAAAAATGA		KLSLVRALNKVSSPTAL			
				GTGCAAACGGC		TDYRPISLLCFLSKALEW			
				GAGAAGAGTGT		LVHRQVSEYLESRLLLD			
				GGAAGTGCCTT		NFQTGFRTGHSTHVRVG			
				TTCAGCCATGC		INKKKVTLLLLFDFSKAF			
				AGCAGCTGTAA		DTVCHVRLLRKLSTFGF			
				ACAACCTATAC		SKQVIRWFASYLTGREQ			
				AGCACAAAGAC		AVVGDNSERFSPRRPAG			
				CCCAAGAAGTG		SVWSLLLHCTSTTSVSA			
				TGAATTTTGTGG		LIPMFPILSMRMTCKFTV			
				TCTGCACTACCA		NATLRSSILYQTRRAFLS			
				CGTCAAGTGCC		SASRRVRLVLAVALYIN			
				AAAACACCGTT		DIGFCLDSDVSHLIYADD			
				AACATCAAAAC		LQIYSQCHLEELDSLSNK			
				TATCGTGGTGA		MSANAERIMGWAAQNR			
				ACGACAGGCAG		LKLNVNKTKAIVLGSPY			
				ATTATAGCCTGT		YINALPKVTLINMGVPG			
				CCCTCATGTGCA		QLHLLIWGCQVSFESSV			
				GACAACAGCAA		RNLGLVLDSKLTWKEH			
				CGCCAAGGGTG		VTQVCKRAHSLMYRLY			
				CCAAACTGAGG		FFRKSTNLRLRKHLVQA			
				ACTAAATCCAC		LLFPIIDYCSLVYCDLTQ			
				CTCTGCCACAA		ELDLKLQRLVNTGIRYIY			

				CATCAGCAACG GGAGTAACTGC AACAACAGCAG CAGGCAAGAAT ACTCCCAAACA ACAGCAGCCGG TGAATAAGAAG ACGAGGTCAGC ACCACCATCAG CCAACACGTCG AGAAATCCCTC ACCAACCCGCT CGGCCGGAGTA ACAACACCTAC GCCGTCCGTTG AGGCCGCTTTG AGGGACATTCG CGGAGCACTGC TCTCGACGACA TACGCAATGCT CAA		GVRRDEXISPYRRELQW LTTAGRRKYFTACFLRK MFNSVVPSYVLAFFDFR VTLRPVRGEVTPLEIPAF ATETLRNSFHISASYLW NNLPSHIRNTSSTVTFRK LAKDHFFQLENT			

Loa	Baggins-2_Nvi	.	Nasonia vitripennis	ACACAGATAAACTTGCATCATAGCAAAGGCGCCTCGGCCGCACTGGTCAGGCGCATGGCTAAGATGCACACAGGCAT	GGGATGTCATGACGAAGGGTGGTACAATGGGTTTACAGCCTAAGTGCCAGGGATCTCGGAGGTCCCCCCTTCATAGTAATAATAATAATAA	MPNTRTLASRGEIRGLGGAGRIFRDPTCSVPRACILVKGFDALLLPARCSRDLTAIKIRLPVGEGSEREVVVASGYFPYDSQEEPPPREVQDLVEYCRQRSIPLILGCDANAHHIVWGSSDTNGRGDALLQYLVTTSLCIMNRGREPTFYNSVRSEVIDLTLCTVGMEGWVGSWRVSNEPSLSDHRYIQFEWKERCSETRAFRNPRKTDWAFFRENLRNELQSFKPSFGTTDELDHWAFELGEIVNISFQRSCPWTIPRGTWGTPWWNRELEDLRRETRRTLNRAKNTRNSIDWRIHREAQRLYKNRINAVRIKGWRDYCEDIERYPDAARLLRILAKNPEVWLEAIRLPTGEYATSEEECLKLLLEANFPGFRLSHEMGDESSGRNRQQRAAW	cd01650	459	718

						DLAAKVVTPEKVKWAI			
						RNFQPFKAPGIDGIYPAF			
						LQEGLEELVGPLVKLFR			
						ASVALAHVPEIWKTAKV			
						VFIPKTGKPSHTTVKDY			
						RPISLNSFVFKTLERLVD			
						RFIQDDILTVFPLHPNQH			
						AYRAGFSTETALHSAVW			
						RIEEQLERGEVMVGVFL			
						DIEGAFNCTSIEAVVKEA			
						GLHGMPGPLIKWLQGM			
						LTRRTVISSLGTVTVSGE			
						VSKGCAQGGVVSPTIWC			
						LVANGPLEALNGWGCY			
						AQAYADDFLILIKGSEIG			
						AAMDAMQLALRKVER			
						WCCSTGLSVNPDKVEM			
						VIFSRKYKLDAYRAPKL			
						SGVALQVKDSAKYLGIV			
						LDKKSTWEKHLTAQCD			
						KFLVALWTCRRAFGST			
						WGLSPRIILWLYRAVLV			
						PRLAYAALVWWPRAEL			

						AGARAAPEKLRGQVLR GATGAYRTTPTKALGIL VKVEPLHLTIIGMAAKA AHRLNAYGQWTQGTRH TRLPGGVGLLPEFSIKTD MMIPRYFFGRRYGVVIP TREEWKARRDSLPGTGD VWYTDGSRAETGTGSG YYCRRDGRGTFFSLGRY ATVFQTEIYAILTCAQRN IELGARDRIITICSDSQAA LRALMAHRTTSRLVWE CKVVVNQLTAHNNKVR LLWVPGHTGIRGNEIAD RLAALGAKHPPIGPEPYT GAARCLLAGEIRDWVER EHTKEWQGTQGCRQAK AVMGQDTNVGWTKCIA GGSRNNSRLLTQIVTGHI RLRYHGLKMGKEATGV CRWCEGAEETPTHVLTV CPKFAQLRHQCLGELFP			

						TYEEIRALDAGAILTFWR RAGLPA			
R2	R2B_ NVi	.	Nasonia vitripennis	GACTAGACTAT GGGTTCAGTCA GTCCCAAATAG CCGATCCTGGC GCGTCCGGCAG TAATGCCACGT ATGAGTCGGTT ACCCATCTCTAA ACGCGTAGAGG TGGGGAGCTAA AGGCCAGGCGG TTTACCCGACGT CGAATTTCTCCA GGTCTGTGTCA GTCGACGGAAT AAAGGTACTAC AACATCTACTAT CTATCGGGATC GGAAGACGCCT TACAGCGTTTTC CGATTTTTGCTC	TGACCTGAACAAAA CGTGTTGTCTTGTC TTGTCTAAAACTAT TTATTCGAAATAAG GGGAGGCTAACTGC CTGCAAGTTGAACG CGAAAGTTAGACCT TCCCACCTAAAGCC CAAAAGTGATCGG GGAATGAATCCGCG GGTGACCCCAGAGT TGGGTAAACCCTTG AAACGTTGGAGAA GCGGAAGAGAGTC CCGCCACCGAGCAT CGAGTGCTGCGGCG CCCGAATGAAACCG ATCGCGGATGGTGC AAGTCGTAGGACG GGGCACGACCTAA GCCTCTGTCACGGC	TFAPTHPMVRSGPCRKT KRPGSDYRESLIMDSGN NVASEPRGAVDVTSAAP IGAELNAEPCEGRNQRR EAALSAQTRRRNXARRA RNAQQADEPGDDEEIET HGPLTIRTXEPMEIVAIA KNPQACPKCLQGGTQLL CMGSWELSRHINKEHPS VDVTWVCGACQRRCTT LRSWSCHVLHCKGRQEP KDLPFKCEHCSLSFDSQI GLSQHERHVHPEVRNDK RAAEANKPKGKSGRRPS IWSDEDLLLIRELESEYH GARNINEKIAEHFPDRTG RQVSDARRRKDYAALR GRGGPQGPAEGVEAIEE VDEGEIPEGEELVATDG AALESGPPENGGSAPAE QVNAPALESSSQQDREC	cd01650	626	885

				TTTGAGCATTTT	GGCGAAGCCAGGA	SPAVGSDEQIEDSSDDDE			
				TCTTCAAATTGC	ATCACCATGCAAAG	FSDALGEISLPEPLSVERT			
				GATAACCGACC	GTGTGAACTGGGGC	TISPPPRDDWKGPMRWE			
				CGATCACGCGG	GGATACCTCCACGG	ICNASEEAGSYANWVTG			
				GGCTTTGACAA	GGTTTCCCTGGGCA	LQELVRNNALSEIGLDSL			
				AGCAATGCGTG	TCGCGCGAGCGATG	YDQLIQIMRHPSDDNEQ			
				GTCGGTAAGAT	GCCAAAGTCCGCTT	DRLQLNARGPPRRGHRK			
				GGTTGCAATCTT	TCTCAGCTACAAAA	NRRRRRLTAADRKRFAF			
				TTCCACCTCGTT	CAAAAATGGTATGA	ARCQDLWNNNPKKLAE			
				TCTTTTACGGAA	GACTTCGTTAACAC	LVIANDLSILQRRQAPGR			
				CGAAAGCAATG	TAATTTTTCCGAGC	TETQTLYNELWGRVGPN			
				CGTGTGGGGAA	CTAGCAGGCTCCCT	IEAPRRTEDPIPVSRIFTPI			
				CGTTAAAAACT	TGACAACGCTTATG	TPQEIMGRIRRIKNDSAA			
				CCCTTCATGCAT	AATCTGGAAAAGG	GPDGVTKDDLRGRGVSI			
				CCCAGGATTTAT	ACACAAAGTGGAA	ALSKLFNSILLAGYYPKA			
				CCTGCTTACTGC	AAAGCGCTGATGGT	WRENRTTLLPKPEKDPA			
				AAAGCAATGCG	GGACAAAAGTCAG	DVKNWRPITISSMVSRV			
				TGTGGAGCGAC	TTGAGACTTGATAT	YSGLLDQRVRAVIKQCD			
				TTTACCACGAGT	CAGTTGTTTTGACT	RQKGFTEENGCFSNIQLL			
				CGCTCCACCGC	AAGAATTTTATTAT	DDAVSNAKKAGGVITIL			
				AAAGCAATGCG	CGTTGACTTTTAAA	DVSKAFDTVPHAVIQGC			
				TATCGCGCAAA	TATTTTATTATTGA	LEKKGIPETVAAYISSMY			
				AGCAATGCGTG	CTGTTAATATACTG	RDCSTAIRTRSGDVKIG			

				TGGGGGACTTG TCAAAGATCCC CCGCCGCAAAG CAATGCGTGTC GGCACCACGTA GAGCAAAGCGT GTAGGCAGACT TTGTCAAAAGT AGTTCTGCCGC AAAGCAATGCG TGTGGAGATCTT CGCCGGTGAAA GCAATACGTGT GGGCGAACTTA G	ACTTGGGACCAAGT CATCTCTGTTACCC GGTACCGGTTCCTG TCATCAAACCGGAA AGTCCGTCCCACGT AATGTGGTAGACGC AGGAG	MKRGVKQGDPLSPLIFN LVLEPLLERLQETSGVEI EGMNLSCAAFADDIVCF ANTAPEAGRQLRMVAD YLGRLDMSLSVSKCIAV EYVPHRKTWYTKNPGL EVNGNAVPSISPSETFKY LGAKVSPWKGLLEGFES DAFREVISRVQRLPLKP MQKVDLLQMYIFPRYTY GLITSPPAKAVLKTIDRII RTRIKEILHLPESVSSSFL YTPRKQGGLGLLEVEK MVLIAALRNGLRARQSH DPVTRAAMNSNAADDR LKSYADALRLHWPLTTK ELDTYKYQLRLSYAQK WAEQKWQGQGVEEFAQ DPVGNSWLQRYDLLPAS RYIDAIKLRTNTYPTRAL MKIIDGRVDSSCRKCQG SSETLGHILGRCRYTKD KRISRHNEIKDLLKARLA			

						KNHQVMDEPQITVRGQ RFKPDLVVKTNEGRVHV IDVTVRYEHRTYLDEGR TEKIGKYRQILSTLRRDL HSNAEEVIPIVIGSRGAIP RETRKALSKLGIGKSDW LTISLIALRSSLEIVNAFM DD			
CR1	L2-5_NVi	.	Nasonia vitripennis	TGCTGTAACCTA ACCTACAACCT CTCATGGCTTTT CTCCTTGCTGCA CGGCGTCTATTT CTCCTCAAAGC GATCCACGTGG AGCCACCTGTC GGGTACACGGA ACACCACCAGC AACATCATCAA TCACAGGTCGG TCAATTCTCACC TCTCTCGATTAG CACCCCATCTG	AA	GILPHSPVLPVRMNTPPA RPRPRRRELNQQQRQQQ QIHLPRARLERMCPKRIN RHPHRLLAPEPNRLREH RLRLSRRPLRTSSRRWM TFGRSRMRPFRHKTPCL KGSLELRKGLALLQSRL KALDDLPALKTRLNNVE SSISELQAQYQELSSRSP AEQQNISIAPASEEINNL RSELAEVKRRQERSANN VVVISDLAYTQETSLQLL AFTVLAALDSTVLRRDV ATVRTMGRLDATTTNA QGDSRFPPLAVTLSSSAL	cd01650	901	1157

				GCTTCGAGAGT		ARSIIVAKARKRKLHTSE			
				ATTAAGTCTCTA		LDAALLEEARALCPDHQ			
				GCTCTGTCGATA		GLININELLPPDIHKLRIK			
				ACTTACCACCC		ARLEAKRRRGCRSFVRD			
				GCTACTATCTTA		GRIYIRSGEDSERATIIST			
				CAACACTCTCTA		DAELEAFLAHTPPIANTD			
				CGTTATCGACTG		ITHGFFSPYAANSQHGH			
				CCAGTAAGTTG		HTLRHTPIIPPKPPSKSSN			
				TCGACGGAGAC		SSTSASKASLSEGLRVCH			
				TTAACAATAAC		FNANSLTGHIEMVRLFLS			
				ATCGAAAATGA		TRSPFHVIAVTETWLSD			
				GCGGAATTGGT		KITSIPSLDDYILYRRDR			
				GTGAAGAATGA		NRNGGGVALYIHHSLTA			
				CGAGTCGCTCCT		SVISSSDGEWSGKPGKPE			
				CATGTGTGCTCA		YLFCEISAKGVSPIFVGV			
				CTGCAAAAAAA		VYRPPHSPFIQGSNFIEQL			
				CAAAAAGAGGG		TTHMHNYSTKVIMGDF			
				AAAATGAAGAC		NSDQLSSSEDANFIRAFI			
				CTGCGAGTTGT		DENSLTSVPYGATHHRQ			
				GTGACCGGCAG		DSDTWLDLCLIDEQDCL			
				TTCCACACCATA		LSHWKTDTPFINGHDLIT			
				TGTGTCGCCTTC		ATLDVQIPRHVPATYSY			
						RNYKGICAEKLRDFLSA			

				AGGTCTATAAT GGTGA		CDWSSFTTSSLDECITIL NANLTNAINHLAPLKTV TPGRKRHPWFTAALRDL LTERNRLYRRFRKSRLP WDLYFYRIARDDAHKQ VEEARLNYYYSRLSTLT DVAEIWRELENLGITTSK SPSPARFSSDALNKHFSS ISNDPQAPSVEEYLRTLE SLDLPEHFIFRAITESDVS AAVSHFNTQARGSDGIP QVVISKALPILAPLLCQIF NLSLSEARFPSAWKSSL VRALNKVNSPTALTDYR PISLLCFLSKALEWLVHK QISEYLETRLYLDNFQTG FRTGHSTQSGLIKLTDDV RLGIDRKKVTLLLLFDFS KAFDTVCHVRLLGKLSS FGFSKQVIRWLASYLSG REQAVIGDNNELSTSLPL NTGVPQGSVLGPLLFAL YINDIGFCLDSDVSHLIY			

						ADDLQIYSCHLEELDSCS VRMSANAERIMGWAAL NRLKLNVLKTKAIVLGS PYYINALPSIANTFINIGG ARVDYESSVRNLGLVLD SKLTWKEHVTQICKRAH SLMYRLYFFRKSTNFRL RQHLVQALLFPIIDYCSL AYCDLTQELDTKLQRLV NTGIRYIYGVRRDEHISP YRRELQWLTTAGRRKY FMACFLRKIFNTSVPTYI LAFFDFHVALRPVRGEV TPLDIPSFKTETL			
L1	L1-11_NN	.	Nelumbo nucifera	GGTGGGAGTAA CAGAAGAGATA GGGGWTGGTGA GGMTGTAGAAG GGTCTAAGGAG GAAGGGGAGM GAATCTGGGAA TTAATAGTGATC CTTTTCGAGCAG	TAATGAAAGTTACC GTTTCCTATAAAAA AAA	MKIISWNVRGLGSSRKR SVIREVLRKEKPTIFLIQE SKLTEVDRRWVRSVWS ARGLSWAFTQSWGASG GIITIWKNEEMDCTEELI GNFTVTIKLRRKEDGFE WALTNVYGPNSYRERD DFWXELAAIRGLWNGP WCLGGDFNVIRFLNETT	cd01650	496	761

				TTGGTAGAGGT		NPNHRVTSSMRKFNLFV			
				GGAAATTCACT		DSYELRELPMLQANFTW			
				CCAGGTTGGGA		SNFQEEPLCCKLDRFLFS			
				TCAGAAGAGGG		SDFESAWRGLAVRALPR			
				GGCTGAACTKA		PTSDHCPVRLSQENLGG			
				TTGGGACTACTT		GPCPFRFENMWLLHPDF			
				CTAAGGAAAAG		IKKVQLWWEEDQVEFR			
				GAGAGGGAGAT		WKGYKFYKKLQALKAR			
				TAATAAGAGTG		IKAWNKEEFGRVEERKE			
				GTGCTGGAACT		WIANRIQEIDGGEAERRL			
				AAAAAGGCAAT		TVGEREERWKLKQELQE			
				TAAAAAACTTA		IIRKEEISWRQKSRCKWL			
				GCAAGAATAAG		KEGDGNTKFFHTLANGR			
				AAAAAAGGACA		KRKNFVGRLEWNGEVIT			
				AGAAGTAACTC		SEEEIARAFVEHFKSIYD			
				AGTTAAAAGCT		SPRREREEIPDLDWLPGN			
				TCAAACAAGGG		RVLMDQLEKNFEVQEV			
				GGAGGGAGAGA		KDAIFSMERSKAPGPDG			
				ATTCTAAAAGA		FTMAFYQDCWDIIGGDL			
				AGAAGAAAAAA		MEVFLEFFERGKINSGA			
				GATGAGTGAAG		NATFIALIPKKVNANHPK			
				TCGAAAGACTT		DFRPISLIGSMYKVISKV			
				AGGTGTGGTGT		LAERIKKVLPSLISPNQG			

				GGATTATGAAA		AFVSGRQIIDQILIANECV			
				GAGGAGGCACA		DYRKRQGEKGLICKLDL			
				GGAAGGAAGGA		EKAYDLVDWGFLLEVL			
				GAGAAAGCTGG		RIRGFGERWCMWVKGC			
				GTTAAATTAAG		ISSCHYSILVNGVAKDFF			
				CTGAGTCGTTTG		PAKRGLRQGDPLSPFLFT			
				GGCCTTTTTGCT		LIVDVLSVQITRADARNL			
				TAGTTTTCC		LRGWRVGREEIVVTHLQ			
						FADDLILFCEGDPEELEN			
						LKVLVENFEKVSNLRVN			
						WNKSCVAGINRQEVEV			
						ASLARSLGCEVSDWPLK			
						YLGLPLGGNPRSVSFWY			
						PVLERFSKRLAGWKRNY			
						ISLGGRITLIKSTLANLPV			
						YFLSTFKLPVKIANQVEK			
						IMRDFLWEPGDTKKDHR			
						VKWEVVSRKKDKGGLG			
						IGNIRIRNAALLGKWLW			
						RFPKEQESLWHKVIRSIH			
						GMAANGWDAGLDPRAS			
						RFSPWKDIVGILPVFSPFI			
						RYDACVGNKILFWEDL			

						WIGESPLKDRFPNLYSLS			
						LKKSQFISSFRASGDSWD			
						LGLRRGLRDSEVDSAND			
						IFSLLSSVFLDHSSPDSRI			
						WAETSTGSFSCSSFFNAL			
						VLPSNPVSSFPVKEVWS			
						APVPPKVRAFCWEVVLQ			
						KINTLEVLQRKRPNWCL			
						SPHVCSLCLRNGETVNH			
						LLISCPFSALIWDWFSSL			
						FDCRWTPPTSVADLFDP			
						SFFRQPKKKGRVMWRT			
						TRAAILWAVWEERNNRI			
						FQGKEKSAPEVIEGITIK			
						AVHWLKTLDLFKDVSID			
						DLIRGWSILARSSTSSPTI			
						QAVWSPPLQCNVKLNF			
						DGSSIGNPGPSGIGGVFR			
						DAEGKILALYSGPIGLGD			
						SLRAEVLALLQGLRWA			
						KQLGFQQLEVEGDSKLV			
						VGWLQSNSLGTWSLRQ			
						HVRELLSLTRELGAQVY			

						WSRRSANALADNLAKQ GVNREVFVLCSVVSYLL FRGHLLLWKVTVVSSGF I			
L1	L1-5_NN	.	Nelumbo nucifera	GAAGTTTGGGG CAAAAGGAGAK AGAAGGAAAGG GGGAAAGTAGT GGTCAAGGGTA GAAGGGAAGTA AGAAAGAAGTG AGAATGAAGAT ACTAAGTTGGA ATGTGAGGGGG ATAGGAAGTGC GAGGAAAAGAG CAACAATAAAA GAGGTAATAAA GAGGTTCGACC CAGACATTGTA CTGTTGCAGGA GTCCAAGAGGA TGGAGTTTAAT	TAACTTAACTGTTT CTTTTTTCCTTTTCT CTCTTTTGAGGGTT TAAGTTGTATTCGG AAGATAAGAGTGTT TGTCTTCCTTGCTGT AATTATTCTTGTGA TGTTTCAATGAATT TAATGAATCTGTTA TTGATCAAAAAAAA	MDMCGCLQMSMGQMT TEREELWEELADVRGL WDGPWCVGGDFNVIRY PEETNKTNQRVTTSMRK FNRFVDAFELKEIPLAQK TFTWSNLQARPICSKLD RFLFSVDWEEHWGNLTS RALPRVTSDHWPIVLCK EERRSGGPAPFRFENMW LLHPGFTNRVREWWNE CNPLVEWEGMRFHKKL QFLKAKIKEWNILEFGRL EERKRDLMDQLESLDSK EERGEFGEEDNELRRGL RRQLEDCLHKEEISWRQ KSRCKWLQEGDRNTRY FHSLADGRRRRNFISKLV VDGIELTQQEEIAGACVE HFKSILRRPVSGRVWVE	cd01650	398	663

				AACAGGTGGGC AAGGAGTGTAT GGAAAGCCAGG GGGATTAGTTG GGTGGCTGCTC CCTCTTGGGGTG CTTCAGGGGGG ATAGTGACAAT GTGGAAGAGGG AATGTGGTTGA AGGCTTGGAGG AACTGATTGGA AGATACTCAGT TTCCATGAAGTT CAAACAGCAGG AGG		GLDWPDVGGDNRRSLE RXFDLEEIKEAVFSMDR TKAPGPDGFTMAFYQDC WDILKEDLLMVFEEFHR RGIVNPGVNSTFIALIPK KKEASHPRDFRPISLVNS LYKIIAKVLSLRLKKVLP DIIEQNQGAFVSGRQITD QILIANESVESFKKDKKK GVVCKLDLEKAYDMVD WEFLRKVMTKMGFGAR WCKWIMGCVRNCHYSV LVNGTVVDFFPATRGIR QGDPLSPFLFLLVADVL GKLIHRADSRGLIRGWR VGREDIPLTHLQFADDTI LFCEGGNVSFIRNLKLVI QLFEQMSNLKVNFGKSC VVGVNLEEGEIRESAEV LGCEVSGWPLSYLGLPL GDNPKRESFWNPVVERF EKRLVSWKKNYLSLGG RTTLIKATLANLPIYFLS			

						VFKIPXKVSSRLERLMR DFLWEPGEEKRDHRVS WKVVTRSKNYGGLGLG NIGKRNSALLGKWLWR FPLEQEKLWARIVASIHG MEQNGWDAGIRARTTS TSPWKAIHDLLPVFSPLV RYDLNSGNSIKFWEQIW VGNVPLKDQFPSLYALS DFKNEPISSFITPSSSSSGF GWNLGLRRAVRDSDLD SISSLLELLDNVFLDPSK VDSRIWTPTPTGSFSCAS FFKVLCQSTSPATSPFPV NEVWCRPIPPRVGSFCW EAYLERFNTGDLLQKKR PNTCLSPHWCVLCKADG ETGFHLLVSCQFAKSIW EWAFSLFGRRWNPPSDL QSLFSPSFFRWHMKRGR TMWRGLRAAVCWSLW EERNSRIFSEDERGAEEL RAVIHIRAVQWLKSLEL			

						FTDYTVEDMLRKWGEIA NCTQTKPVIRVSWTPPP AGFLKLNFDGSCRGNPG QSGIGGVFRDENGVISAL FSGPIGVGDSLRAEILAV MEGVQKAKEMNVRKLII EGDSEVVIGWLIKDNCG LWNYXNERRKFRWLVR EMEIRMSWVKRSANSLA DGLAKQGIEREAMFWA KL			
L1	L1- 10_N N	.	Nelumb o nucifera	GGAGATTTGTA GGGAAGATTCA GTGGCAGAATT TGTGGGAAAGT TGGGGCGTAAA CTGAAGAACCC AAATGATTTAA AAGCATTTCTAC GAAAAAAGGAG GATAGAAAGNT CGAGAAAAGAA AATTGAGCGAG	TAGTTTCATATCTTT AATAAAGTTTCTTG TTATTGATCAAAAA AA	MRIXSWNVRGLGSARK RATIKEVXKKFNPDVVL LQESKRMDFDCRRARSV WRAKGIKWAAAPSWGS SGGIVTLWKEDVLIGLEE LIGDHSVSLKFRQVEDG FEWVLTNVYGPNDYRQ RKKLWEELSDIRGLWEG PWCLGGDFNVVRLTEET NNPRQRVTTSMRSFNQF VDXFELKELPMLNNTFT WSNLQSSPICSKLDRFLF	cd0165 0	498	763

				GTTGAGAGATT AAGAAGAGATA TTGAGGGGGAT AAGAAGTTTGG GGCCAGAGGGG ACAGAAGGAAA GGGCCCAAGTA GATTTTGGTGGT GATAGGGCAGA TTCTTGCAAAAT TGTG		STDWEAHWGNLTSRVL PRLTSDHWPIVMCKEXL RGDGPAPFRFENMWLT HPGFIEKIKEWWQHYNP PVQWEGMRFHKKLQFL KSQIKEWNRLEFGRVEE RKQVLVEQIEELDMKEG RGLFGEEDRERRWSLRK QLEDLLIKEEISWRQKSR CKWLKEGDRNTGFFHS VANGRRRKNFISKLEVE GTVISHPVEIAEAFVGHF KSIFRRPLNERIWLEGLD WPTLSPENREWLERRFX LEEIKEAIFSMDRTKAPG PDGFTMAFYQDCWGIVE EDIGSVFEEFHRRGIINPG LNSTFIALVPKKKEASQP RDFRPISLVNSLYKIIAK VLSLRLKKVLPSIIAPNQ GAFVMGRQITDQILIANE SLESFRKSKKKGVVCKL DLEKAYDXVDWSFLLE			

						VLKSFGFGDSWCXWIW GCIRSCHYSILVNGTVVD FFPASRGIRQGDPLSPFL FTLIADVLGRLIQRADSR NLIKGWRVGREEIPLTHL QFADDTILFCEGGDIEWI RNLKLAIQLFERVSNLK VNLDKSSLAGVNLEERE LREYADVLGCATSDWP MAYLGLPLGDNPKKVSF WNPVVERFGKRLASWR KNYLSLGGRITLIRSTLA NLPVYFLSVFKIPTKVSN TLEKLMRDFLWEPGEEK RDHRVSWEVVTRGKKL GGLGIGNLRKRNMALLG KWLWRFPLERDKLWTK IVASIHGLEDNGWEPGIL ARTTAASPWSAIHKLLP VFSPCIRYGLNSGNKIKF WEQLWVGDSPLMDQFP SLYALSDLKNEPISSFLSP SDSPSGFGWNLGLRRAI			

						RDTDLDHILSLLAILENV FLDPXKEDTRFWVPSSS GVFSCASFFETLTQPSDPI GLSFPVKEIWSRPIPPRV GSFCWETFLERINTNDLL QKKRPNSCLLPNRCTLC KADGETGYHLMVTCSV AKSVWEWVFSLFGCVW XPPGSVQSLFSPSFFRWP KKRGKTMWRSIRAAVC WXLWEERNSRIFSDKER TIGNLLDLILLRVAHWL KTMELFSDYSVADXVR NWKEIASSALCKRTCSPS WSPPPSGFIKLNFDGSCX RNPGPSGIGGVFRNEAG AIIAMFSGPIGEGDSLRA EILAALEGIQRARELNIR CLVLEGDSEVVIRWLKG GGSGLWSYNNEKNKFK WLSRDIVVHYNWVKRT ANSFADGLAKQGVNRD			

						STFWANFLXFLVFFGGE LSLFTSPLYLSFT			
L1	L1-17_NN	.	Nelumbo nucifera	AGTGTTTGGAG AGGTAGGGGGA TCAATTGGGCG TTTCTGCCCTCA TGGGGGTCTGC AGGGGGTGTTG TGACA	TGAATGTGCAGAGT TTATTCAGGTTCAT GCCGAGTGGGCCA ACAAAAAGGGCCA AGGCTATGTGGAGA TGTATGAGGGCTGC AGTATGCTGGACTG TTTGGGAGGAGAG GAACTCACGGATAT TCTCAGAAAAGGA GAGAACCTTGGAA GTAATATTGGAGGT AATCCATATTAGGG TGATACATTGGCTG AAAACTTTGGCCTT GTTTGCAGACTTCT CAGTAGATGAGATG GTAAAAAACTGGA AGGAAATTGCAGA ATTAGCCTCAGCTA AGAAGGCATCACTC	MWKKEVLEGKEELIGQF SVSLKFRQVEDGFEWVL TNVYGPNDYRERRQLW EELADIRGLWEGPWCLG GDFNVIRFLEETNSTRQR STMSMRKFNQFVDAYG LKEIPLSHSCFTWSNLQV NPICSKLDRFLFSNDWEE HWGSLNSRTLPRVTSDH WPIVICKEEVRNAGPTPF RFENMWLGHPGFIKKVK DWWLESNPTVKWSGMR FHKKLQHLKGRIKEWNL LEFGRLEERKQMVTAQI EALDSKEGRGCFTEEDK EQRWNLRKQLEDCLLQ EEISWRQKSRCKWLKEG DRNTKFFHAMANGRRR RNYISSLEIEGATISNQSE IAGAFANHFRSIFRRPLN NRVWIEGLEWPTLKLEE	cd01650	426	691

					TCTTGCTGGAGTCC	REWLERKFXLEEIKEAV			
					CCCACCTAATGGCT	FSMDRSKAPGPDGFTMA			
					TTMTTAAACTGAAT	FYQDCWEILKEDLRIVF			
					TTTGATGGCAGCAG	DEFHSRGVINPGLNSTFI			
					TATGGGTAATCCAG	ALVPKKRNANHPRDYRP			
					GTCCATCTGGAATA	ISLVNSLYKIITKVLSIRIR			
					GGGGGAGCAATAA	KVLPTIIADNQGAFVKG			
					GAGATGAACATGG	RQITDQILIANESLEEFKR			
					AAATCTGTTAGCTA	SKKKGVVCKLDLEKAY			
					TGTTTTCAGGACCA	DMVDWSFLLEVMVNFG			
					ATTGGGGAAGGAG	FGDRWCKWVMGCIRKC			
					ACTCCTTGAGGGCT	HYSILINGTVVDFFPALR			
					GAGATTCTTGCAGC	GIRQGDPLSPFLFTMVAE			
					CCTTGAAGGGGTTC	VLGRLISRADSRDLIRG			
					GAAGAGCCAAGGA	WRVGRESVPLTHLQFAD			
					GATGCAAATTAGAA	DTILFCEGGDLNRIRNLK			
					AGCTGATCATTGAA	LTIQLFEQMSNLKVNFE			
					GGAGATTCTGAAGT	KSCLVGVNMGEGETSEC			
					GGTGGTGAACTGGT	ADILECAVSNWPLNYLG			
					TAAAAGAGGATAA	LPLGDNPQKDEFWIPVV			
					CTCAAGGATATGGA	ERFEKRLATWKKNYLSL			
					AGTACAGCAATGA	GGRITLIRSTLANLPVYF			
					AAAACGAAAATTTA	LSVFKIPAKVASRLEKQ			

					AGTGGCTGGCTAAG GATGTTGAAATTCA GTACAAATGGGTTA AGAGAAGTGCAAA CTCACTGGCAGATG GTTTAGCTAGGCAA GGGGTTAATAGAA ACTCTTTGTTTTGG GCTTTATTGTAATT GTATTTGCTGCTCA TAGACTGTTTAGGG AGCTAAGATTGTTT AGCCCCCTTGCTTT GTAATTCTTCTTAA TTTTAATGAATACA TTTCACAGTTATTG ATAAAAAAAAAAA A	MRDFLWEPGEEKKDHR VRWEMVCRSKREGGLG IGNIRSRNLALLGKWLW RFSLEKDKLWVKIVGSI HGVADNGWDSGIVART TRAAPWKAIQAILPLFSP WIRYDLHSGNLINFWEH IWIGDSPLKDQFPSLFTL STLKNEPVSSFLVPSNST SGFSWNLGFRREIRDTEF DCILSLLSILEGVFLEPTK TDSRIWSLTPSGSFTCAS FFKALTQPPSLGLSFPVK EVWCLPIPPKVGSFCWE AFLERVNTCDLLQRKRP NSCLCPHWCALCKADG ESEYHLLISCRFTKSIWE WCTSLFGCRWPPL			
RTE	RTE-7_NV	.	Nematostella vectensis	TCCTCGATGGG ATCGCCACCTTA TCGTGGTGAGG GTGTTTGTGTGT CCCAATGACCT	TAACAGGTAAC	MTRRGQSHPDAPGLTKP LVHPKHAMRIGNWNVR TLYSSGNVAQAAREMG KRDIDIMGISETHWTGQ GKMQLRDGETMVYSGR	cd01650	567	824

				CTAGAGCTATG		EDNIHREGVGILMSKNA			
				TCGGCGGGAGT		AACLIEWTPWNERMIQA			
				ATTTTACTCCTG		RFYSRHIQLTIIHIYAPTE			
				GCAGGGACACC		DADEHLKDEFYTRLQEV			
				CATGCCGAACA		LDSRNQHDMLIITGDMN			
				GGTCGAAAGGT		AKVGDQNWDYERVMG			
				AGGGGCCAGAC		KHGLGVRNDNGERLCE			
				GAAGAGTGATC		LCDLNELVITGTLFPHRT			
				CACTGGTCCTCC		IHKATWISPDGKTKNQID			
				AGGTTGGGGGT		HVLVNKRFRNSVKDTR			
				TGGGCAAAGGG		VFRSADVGSDHYLVCAT			
				TTGATAACCCTC		VKLRLRKDPKRKGNIRV			
				TCCCATAAAAA		KYDTSKLTEKEVRNTFN			
				ATAGCTTATCAC		IKLRNRFQALEQEDEEER			
				AGAAACCAGAA		GEEVVDAKEEVERDFTI			
				GCAGAGCAAAT		MKKALLEVAETVLGKP			
				TCACTTGGGAA		RKKNKPWISGESWSLID			
				GACCGTGGCTG		ERQQMNEKILSTRSERV			
				CATCTCATGAA		KGQLRKKYREKDREVK			
				AGAGATTGT		RSLKADKKKWLEEIASE			
						AEKAARSQHMKTLYGL			
						TKKLCNERPRHSTAVLD			
						KDGNLLSKKEEIQRRWT			

						EHFREVLNREQPDNPIEP EDEVEFEFGELIEEISVNE PTFGEVKDTITTLKNGKS PGIDNITAELLKADPEFS AERVHKLLVKVWRHETI PNDWKKGLIIKLPKKGN LKECKNSRGITLLSIVGK ILGRIVIDRIRNGVDKKL RKEQAGYRRGRGTTEQ VFILRNIIEQVNEWQATL YVNFIDFEKAFDSVHRES LWVIMEKYGVPEKIIRIV RLFYEDFQCAVEDQGET GEWFNIKTGVKQGCNM SGFMFLLVMDWILRRSV GKGENGIRWKFTSKLDD LDFADDIALLSSTKHHIQ EKTAKLDGEARRAGLKI NLEKTKMLRINAKQHG NIVINGKEIEDVSEFTYL GAKVCREGGGMKDLRN RLSKARSVYTRLKRIWN SKKIMRRTKIKIYKSLVL			

						SVLLYGCETWKMNKKD DKLLDVFQQKCLRRILKI SWEDHVTNEEVQRRAG VHQLSSEVKRRRWKMI GHVLRQGRENDCSVALT WTPEGRRKQGRPKTTW RRTVERERGEAGWRNW DEVRSKAADREKWKST VKALCATRHEEDR			
CR1	CR1-20_NV	.	Nematostella vectensis	GGTCAGCGGTCGACCAAG	TAATAATACCTATT GATATATATATGCA TGAAAAGATATATT TATTTTTATTTAAC ATTTACATTAACAA TTAACATATTTTTT ACTAGTTTCTGATC TTTAGGGTGCATGC CATGGATGCATGGA GCGGCTGCCAAACT AGCGGGAGTTGGCT CAAGGGTGAGGCA AAGATTCTTTGTTT CGCGCTTGGGTTGA	MAANFLTPVFIAFLAILF NQKLHYSPNLTSGINGF GLNCILANSESILTGLEA FYSHNIRANKMNVGRSS SEHKHLLLCLLLLGGDIQ LNPGPKWKFPCGSCNKP VKSNQKGIQCDNCNLW YHTRCCSINDDLYYTLA ASSCTWLCVSCGFPSFSS SLFESSSIELQNSFSSLDS LHQDTQNISSSTIRSPETS PLDSSSRNSSNTRKQSKE KLTGMILNCNGLKGTDH ITKFQALINLHDPDFILGT	cd01650	707	959

					TTCTTGTTGGTTGG	ESKLCPDISTYSIFSQNYS			
					CTGTTTCATGCATT	VFRRDRNKFGGGVFQAI			
					TATGGCATTCACTC	KSDLVCTEEPDFEVDNC			
					TATTGATGATAATT	EIIWTSLKLANRKSIYMS			
					TTAATTTCATGTAT	SYYRPPNQPSDSIDNLNE			
					ATATAATTAATTAT	SIRRVFNKVKNFPNIIIGG			
					TTAATTTTATCAAA	DFNLGDIEWNSSVPTPT			
					TATCTAATTTCAAT	NPATASLHQKFLQTIDDF			
					AATTTTCAAGTAAT	SLSQHVKTTTRPSSGKTL			
					ATATATTTATGCTT	DLLLTTYPNVIVNTSTAS			
					CATTTTCAATTTAA	GLSDHLAVIFEVNLKPSR			
					CCTAAATTATTTTT	SLKPPHKVYQYKKANFP			
					CCGATGGCCGATAA	DLCKFMSDSSTTFFSTNP			
					TTATTATTATTATAT	EKRTVEENWVLFKTSLS			
					TTTTCTTTTGTATTT	TGISQFIPQKSSRPKYKLP			
					ACATATTGTATTTT	WINTNIKREMRKKDRM			
					CTTTTTCAACTCCT	HKKAIKSKNQQHWEAF			
					AGTATAGTCTTAAC	KRQRNLVSKLIKDSHND			
					TGTTTTTTAGGAGT	YLNNIIGNSLKDNPKKF			
					ATAACTATTAAAGA	WSYVRSSKSESIGIPPLK			
					TTAAAGATTAAAGA	NNDSSVSVTDAHKAETL			
					TTAAAG	NTFFHSVFTQEQLPTPVL			
						GNSPYSSIPDLHISAEGV			

						CKQLSELNPKKACGPDE IPANVLKEVSQSISHWLA FIFQQSFDCNKIPSDWSN ALITAIHKKGLKSEPANY RPISLTSICCKIMEHIILSH MAKHLSSHNILINQQHG FREKFSCETQLISAIHDW AKTINFRGQTDVLLLDFS KAFDSVPHQRLLIKLDF YGIRGNMLNWIKAFVSN RKQSVSVNGTVSSSKSV VSGVPQGSVLGPVLFLL FINDISTSIQSNLRLFADD CVLYREVASLQDCQALQ QDLDRLSVWSKTWQLK FNVSKCYHLGITRKKIPV KFTYSMDGKCITQASST PYLGVTIADNLSWNEHC DNICKKANSTLGLLKRIL SGCSPEVKDSAYRTLVR PKLEYATSAWNPYTQCN INKIEMIQRRAARFVFND YSRYSSVSPMINSLGWE			

						SLEQRRQLNQAHMFYKI YNGLVGISFPPEVTPITR VSRQPNCAPFHQLVTLN DTYKYSFYPRTIRTWNSI NLSTIPDSLAKFKDSLLL KVQTN			
CR1	CR1-18_NV	.	Nematostella vectensis	CCAAAATGGCT GCCACGGCTTTT GTAGCTCTGTG AATCTTGTCTCT GTTTTCATGTGT CATTGCTAGTTG AAGGCGTTTGA CTGTCAAAG	TAATAGTATAACGC TATATGTGCTGGCC CTTTGCCTTAATTC AGTCTGTAATACTG CCATGGCAATTGAA TAAAACCCCTATCT ATCTATCTATCTAT CTATCTATCTATCT ATCTATCTATCTAT CTATCTATCTATC	MINNPSNVVPVYFCLPIL AFWSFLPCYFHLKAVDR GSHVATLTSEPVASTGW SYSNQELKTIRKSQPKQR LEPKLWQKLGDLGIRKP FRSRRKPRQSRHVCTGL PVVTTVDSTGHPQLSSG CSTNSSSSQTARFPVIPD TIIANMQSLLPKLDELYA AIMQFPAGIICLTETWLN ESHPDPLIHLPGYLCFRN DRTTSRGGGVCTYINSQ YPCRRLPEFEDLSIESLWI SVRPYRLPRSVSTLLIGVI YHPPSNGEMNRQLVDHI DKNCESFLSKHPDGLVV VTGDFNPSSTGIKAQYV	cd01650	612	864

						THKTGLTQIINVLTRDTG ILDWCLTNKPKYFQQPL QLSKMGSCDHYPILIKPN LSVTNPKAHKSSVWRRN LRRSNLDAFGRWITQFC WNTILSLQTAREKFERF YTVITDAVNTYLPLQRV QVCATNKPWIGTSLCSLI ARRQKALTTLGKDSRLY KELRNQVQRECKVCRER YYPNKVAALKESNISRW WQEVKCLSGSVVSNEW WHQLMDDSTPSVQILAE RFNEFLLGLTSNFAPISQ QVTCSGPVPPELLITPGK AFKALSRIKVKKSSGPDP VPSVIWKEFAPELSDVV CNLYNSSLMEGFVPQEL KESIVHPVPKCSPPKTLS EDLRPITLTSQLAKVME GFALDSLFACVKDKLDP KQFSVSGKSTVLALVYF LHCILESLERGENYVRIF			

						FADFSKGFDLVDHNVL MSELDLLGVNEYLQNWI AAFLTARPQRVKISGHTS SPVFPNGGIPQGTKLAPL LFAILVNRLASTWPYRIK YVDDTTIFEVIPRCSPSY LPFIANDICNFATERGMR LNPKKCRELVINFLQYQ PSPVNALQLMGSVIKRV ESYKILGVHISSDLSWNK HVDYVVKKANKRLYAL RLLKKSGVPVQDLIAIYC VLIRSVVEYAGPVWANI TGFLVDIVEGIQKRALRI VLPHLRYKEALETSGLQ SLADRRRELCISLMTSAK ELEPLRSVITGGVINNVH GYSLRSGSMNTYKHISV TKRFGDFVTYRFL			
L1	L1-17_NS	.	Nicotiana sylvestris	CTCTCAATAATG TGAAAACCTTC CTCACTACCGCT GCTCCTATAGA	TAATATTAATATAT GTTTATCTTTTAAA CCAAAAAAAA	MJKALHGGPWFILSHFL TVRGWEPKFKASSTQLT YSALWVRLPELPTEFYD LEILSRIGSKLGRLLKIDT	cd01650	1264	1503

				ACACACCTTCCT		CTSSTTRGRYARICIEVP			
				TTACCTCAATCT		LEKPLKTHLYIGNHRQS			
				CTTCAACACTTT		LLYEGLNLLCTYCGRFG			
				ACACWCATCAT		HPTTSCPEVVRAPNEQN			
				GTCTGACTCTCA		XRNXQXESNSTPDTIPSK			
				GCAACAMCTGC		ENEWKTVTFTKKSIHPP			
				AACAGCTTAAT		KHLNFDKGKAATMAAT			
				CAAGTTWCCGA		PAAQPLXGQSSNPTSAV			
				CCCTCCGGATAT		LGRVTSTRQVXLQGPVI			
				AGGATCTATGC		EFNNKFQXLSNAGPSDL			
				AAATCGATGAT		QTDHFYXGPIPKNJGTEP			
				ACAAGTATGCA		LLTQQXAKAHFHTPDKL			
				TACACAACCAG		PTSLPNXLPNDPKPRLSE			
				CTACTTTCCTAC		PNPHLTKSPPLVGYNTFS			
				AAACTCTACTA		QDYXTPHZPLPCNHLLT			
				ACAGATAAATA		SSTQSSSNNSEGPLHILD			
				CCTTCTGCCAW		TLXPTQNKAIXTMHATV			
				CTCTAACCAAC		TSLTSDSRLPQNPCKQH			
				CAAATACCTTCC		DEFPCEHHHEXPTTLFPS			
				TTAAACACTCTA		LPNIYXHSIQTXKPNTTL			
				GATCTAGCAGA		XAEPQXLPNPILNAETNL			
				TGAAGTACTCC		HRHEPESVSEGKYTNQQ			
				ATAACMCTACT		PVFSSPSSHNGGXSTASE			

				GACGATGACAC		FITSNLHISAHDHGPPSPI			
				ATTTATACCAAT		LVGNGTAGICSNIHSATQ			
				TACATCAACTG		QPRGGVANXKSDSFSRN			
				ATAAATCACGT		SDRRNEAGNGSLQSKSM			
				TTATATGAGCC		VHDPDSSSNGTPGFQFH			
				MTGGAAATACT		SSAMAISNGPDESDLGP			
				CTGTGATAGTTA		NSQSPTLLPTSSYGSGLC			
				AGCTWTTTGGC		TLVPATGLIQSTLPLLEP			
				AAAMGGATTAC		TESSRESHPTKQSSTITSS			
				ACACCAAATCA		ISTRTRQRYKRGTDSNSC			
				TGCATGATAAA		GDEGDKNRILVVKRDPI			
				CTGATGACTCTG		LQPQRTVRKKFILKCPPE			
				TGGAAACCAAC		MVTCSLNFRQTTNPEVG			
				AGAGGTCCTCC		KYAMVIIPTLLQLPLSLD			
				TCTCATTGATTT		PMTSSVVHPVTQAPMSQ			
				GGGTTCTGACTT		PRNPNSYMNFLVWNCR			
				TTTTCTCATTAA		GSHNPEFRRHFRSLLDN			
				GTTCCAAAGAG		HRPTLAILLETHMQDHA			
				AAGAAAAT		SLREDFNFGNMAQAPAN			
						GLIGGLVVLWNADLINV			
						TEMRISDQEIHCMVQVS			
						PTKPPWILSAIYASRHYQ			
						LRDILWNSLINLHDNTA			

						MPWLIGGDFNEILESKE KFGGLPMNNNRADKFA NFMHXCSLIDLGYQGSR FTWTNKQRNGVPYTKK WRTILERLDRFLANYDW LNLXPNTTVIHLPRTHSD HCPLLINLEPPQPIQNRIF RFETMWASHPQFQQLV HATWSNDVPLLATTDNF KRVVTRWNQNTFGNIFQ QKRKLLARLAGLQSSVH YPTSVFLQNLETQLTLEY SHILRIEEEFWQLKSRIN WLNEGDANTRFYHLSTL HRRRRNRINALQDNVGN WVATTXEIHDLIXNYFX DLFQTDQLISENMTGNT YNCEISSDDILXLXSPLH QDEILTALNSFKAYKAP GPDGLHPIFLPKILDGCRI VLRHTKPQDQMGYIQFF YQKYWTDVGDKVSSFC QDVFCSKIIPPELNKTLIC			

						LIPKVKHPTRIQQFRPISL CNTVYKLITKIIVNRLKPI ISKIIGPTQTSFQQGKRA CDNAIIVQESLNYLAKM KGKYPSMLLKLDLEKAF DRLEWSFIRRALFYFNIP PSMISLIMSCVTTTSISILI NGSKTQYFTPTRGIRQG DPLSPYLFIMCMELFSRS ISLSVDYLKWQPIRLSNK GPLLSHLFFADDIILFSKI TTKSCHAIIDVLNNFSKY SGQKINFEKSKLFFSKNC SSLIKILLLLPLTLKKEPL LVNTWDFLCSKQDPKM QISSSSLIISKINLLDKDF VITSFNIKEGTSFGKYLG FPMFQARPKNADFQFLL DNFKNKLAGWKSRFLTL AGRATLIKSTLNQLPNHI MQYIHIPAHILNKMDQY QRNFLWGTTLEKKKLHL VRWNEITRPKEAGGLGI			

						QKLRQKNHALLASLAW			
						RLFKTPQAMWANILLTK			
						YLNTSPNNIHSQIWSNVI			
						NGWHQCQLGIKWNIGK			
						GTHIKFWEDPWIGPGTTI			
						RSKIHGPLPLDQQNITISS			
						LINNKGWDWNKLPFELP			
						ITLLKVAHNITQIIQATYI			
						SKYSNKVDNIFWGLTPT			
						GHFTTRSMYSILEAQHD			
						VSTNSSTKYNWIWKLPV			
						QPKIKGFVWLLHKKRLP			
						TKDYLRHINLIQDATCQF			
						CHQEDETIRHLFLDCTN			
						AKTFWQDLGILDFINQK			
						LHYTNSFNWVHDLITCK			
						DFHFPYQMNPATYLPLS			
						LWNIWLARNNNCYKHT			
						MSTASAKLTTQQAAKFT			
						YLVAKKHTXVKKRQQN			
						LKWKPPKEHFYKLNTD			
						GAHTTSKAGIGGLIRNSK			
						AEWILGFAAAVPXDSST			

						SAELRALLQGLQLAYSH NLVPLEVEVDSQDIITLL YSNNVSFDNLTTDCRYF LDRLGKPVVLHAYREQ NRVADKLAKAGCNFDIH SVVRILEDNPDFAWSVF EQERSAVQGHVLEHFRP HSDNYDTVNRYCNSDA VGSSINIGLYNTLAPTTG ASACNRMQYRHPPSVR NT			
L1	L1-4_ME	.	Notamacropus eugenii	AAATTCCTTTTA TGATACAAATA TGGTWTTGATA CCTAAACTAGG AAGAGCCAAAA CAGA	CTTGGTACTGGCTA AGAAATAGAATGG TAGATCAGTGGAAT AGGTTAGGTACACA GAAGTCAATGACY ATAGTAATCTACTG TTTGATAAACCCAA GGACCCCAGCTTCT GGGATAAGAACTC ACTATTTGACAAAA ACTGCTGGGAAAAC TGGAAAACAGTATG	ERKLYNFPNEYRCKNFK YQKDYNNISQEYIMXKX DLYQECRXGSILGKLXA LIISXTKLTKIILSQMQKK AFDKIHPFLLKTLKSIRIN GSFLKIISSIYLKPXASIIC NGINMPFQDGNKDVHY HHYYSILYKCLQXKKKK LKELEAKKKQNYHXLQ MIYTRILENQLKNYLKLS TKKLLEIINNFSKVAGYK INPHKSSAFLYITNKTQQ	NA	NA	NA

					GCAGAAACTAGGC	QEIEREIPFKXTVDNIKY			
					ATAGACCAATACCT	LGVYLPRQTQELYEHNY			
					GACACCATATACCA	KTLFTQIKSDLNNWKNI			
					AAATAAGGTCCAA	NCSWVGVNIIKMTILPKL			
					ATGGGTACATGATC	IYLFSAIPIKLPKLFYRAR			
					TAGATATAAAGGGT	KKYQNSSGRTKGQEYQ			
					GATACTATAAACAA	GNKEMLGKVAPPXILNY			
					ATTAGGGGAGCAA	IIKQQSSKP			
					GGAATAGTTTACTT				
					GTCAGATCTATGGA				
					GAAGGGAAGAATT				
					TATGACCAAACAAG				
					AGATAGAGAACATT				
					ATGAAATGCAAAAT				
					GGATAATTTTGATT				
					ACATTAAATTGAAA				
					AGGTTTTGCACAAA				
					CAAAACCAATGCA				
					ACCAAGATTAGAA				
					GGAAAGCAGAAAA				
					CTGGGAAAGAATTT				
					TTACAGCAAGTGTC				
					TCTGATAAAGGCCT				

					CATTTCTAAAATAT				
					ATAGAGAACTGAGT				
					CAAATTTACAAGAA				
					TATAAGTCATTCCC				
					CAATTGATAAATGG				
					TCAAAGGATATGAA				
					CAGGCAGTTTTCAG				
					AGGAAGAAATTAA				
					AGCTATCTATAGTC				
					ATAAAAAAAATGCT				
					CTAAATCACTATTG				
					ATTAGAGAGATGCA				
					AATTAAAACAACTC				
					TGAGGTACCACATC				
					ACACCTATCAGATT				
					GGCTAACATGACAA				
					AACAGGAAAATGA				
					TAAATGTTGGAGAA				
					GATGTGGGAAAGTT				
					GGAACACCTAATTC				
					ATTGTTGGTGGAGC				
					TGTGAACTGATCCA				
					ACCATTCTGGAGAG				

					CAATTTGGAACTAT GCCCAAAGGGCTAT AAAAATGTGCATAC CCTTTGATCCAGCA ATACTACTTCTAGG TTGTATCCCAAAGA GATCATAAAAATGG GAAAAGGACCCAC ATGTACAAAAATAT TTATAGCAGCTCTT TTTGTGGTGGCAAA GAACTGGAAATTGA GGGGATGTCCATCA ATTGGGGAATGGCT GAACAAATTGTGGT ATATGAATGTAATG GAATATTATTGTGC TATAAGAAATGATG AACAGGTAGACTTC AGAAAAACCTGGA AAGACTTACATGAA CTGATGCTGAGTGA AGTGAGCAGAACC				

					AGGAGAACATTGTA				
					CACAGTAACAGCM				
					ACAGTGTGTGAKGA				
					CTRTTTTTGATAGA				
					CTTAGCTCTTCTCA				
					GCAATGCAAKGATC				
					CAAAACATTTCCAA				
					AGGACTCATGATGC				
					AAAATGCTATCCAC				
					ATCCAGAGAAAAA				
					CTGTGGAATCKGAA				
					TGCAGATGAAGCA				
					GACTTTTTTTTTAT				
					TCTGTGTCTTCTTTT				
					ACAACATGACTAAT				
					ATGGAAATATGTTT				
					TACATGATTGCACA				
					TGTATAACCTATAT				
					CAAATTGCTTACCA				
					TCTTAGGGAGGGGG				
					GAGGGGAGGGAGG				
					GAGAAAAATTTGG				

					AACTCAAAATTTTA AATAAAAA				
RTE	RTE- 2_M E	.	Notamac ropus eugenii	GGTTACTGCAG CAGGCAAGGAA TCCCACAAAAA ATGTACAAAAA CATCTGCAAAG	TGATGGGCTGGCCA CGTTGTTCGAATGC AAAATGTATGCTTG CCAAAAAGACTATT TTATGGAGAACTCG CATGGGGCAGGTG ATCACATGGTGGTC AGAAGAAACAATA CAAGGACACTCTCA AGGTCTCTCTCAAG AACTTTGGATTTGA CTGTGCAACATGGG AGACACTGGCACA GGACCACTCAGCAT GGTGTGCCCACATC AGAAAAGGTGCTGT ACTCTATGAGCAAA GCAGAATTGAGATA GCACAAAGTAAAC ACAGGATGTGCAA ATTTGGAATATCCA	MIPLTISTWNVLTLIDNT KSSRPERGTALVARELN RYHIQIAALSETRLANEG QLTEVGAGYTFFWSGHS EGECREAGMGFAIKTNL ISKLVCLPKGVNDRLMT MLPLAGKHHATIISAYA STMTNPDEVKEKFCEDL ETLIINVPKEDKLIILGDF NARVGSNYQTWQGVLG RNGVGNSNSNGHLLLKT CASHDLLITNTVFXLPKC NKTXWIYPRSKHWHLID YVIVRRRDRQDVRVTKA MCDAECWTDHRLILSKL NIHIHQKYCPQGKMTTR RINVNKLELFSESEQFVA DLEGKLSQHIVGNSGAE KEWAAFRDLVYSTAFA HLGQNTHKHQDWFDEN DGEIQKLLNEKRELHRIY	cd0165 0	542	824

					CCCCAAATATTCAT	QQDSSSTSKKAAFNSIKS			
					ATGGACTATCTGTG	KVQAKLREMQDSWLSK			
					CCCAACCTGTGGTA	KADEIQFYADSNNPKCF			
					GAGCATTCTGAGCT	YDSLKAIYGPKTYGASQ			
					CATATTGGTCTGAT	LLSADGATLISDKDMILE			
					CAGCCACAGTCGGA	RWAEHFHSVLNRPSSIN			
					CACACTGAAATTTC	AEAIDHLPQVEVNPSLA			
					ACTTTATCATGGTG	ELPTEEEVLRAIRLLSCG			
					ATGTCATTTTGGTC	KAPGADSIPAEIYKVGGP			
					CTCTTTGAAGATGA	LLIQKLTEIFQVIWQEEVI			
					AGGACAACAACCA	PQEFKDASIVHLYKGKG			
					ACCTATTAT	NRLSCDNHRGISLLVIAG			
						KILARVLLNRLILHLEDG			
						HIPESQCGFRKGRGTVN			
						MVFAARQLQEKCQEQN			
						RGLYTTFVDLTKAFDTV			
						SHEGLWKIMSKFGCLEK			
						FINIVHQFHDGMFAWVL			
						DNGQCSRAFPVTSGVKQ			
						GCVLAPMLFSMMFSAM			
						LTNAFNEDEHSIKVNYH			
						TDGKFFNLKRLQAKTKV			
						EGVLVHDFLFADDCALN			

						AASEAEMQQSMDQFSA ACANFGLIINTKKTQVL HQPPPHHPYVEPSVTTN GEVLNAVDKFTYLGSVL SRDVHIDNEVDTHIARA SSVFGRLRRKVWERRGI RLTTKLKVYRAIVLTSLL YACETWTVYQRHARKL NXFHLNCLRKILRITWQ DKVPDTEVLAQAELPSI QTMLQRVQL			
L2	L2-1_OL	.	Oryzias latipes	A	TGAGTTGATGTTTT TATTGGTTTTATTTT TTTATTGGTTTTATT ATTTTTATTTAATGT ATCGATTTTTATTG ATTTTATTCTGTTTT AGTTTGTTGTTTTGT TTTTATTCTGCGTTG TTAAGCACTTTGGT CAGCTTCTGCTGTT TTAAAAGCGCTTTA TAAATAAAGTTGAC	CLPPSLLVSLSSFELELFE LGCSPPVLCAVVYRPPK YNKDFLNEFSGLLAEILL KYDRVLIVGDFNIHVCC PDHTMSKDFLSLIDSFNL VQHVSEPTQERGHTLDL VLTYGLSVSNLVICDAV FSDHMPVLFDITLPCSPC RPRAPARWVRIINPSTAA LFSASFNQTSVIPQSTCSS TDELSSWLHSTCLTVLD AVAPLKSRQPKVKTEPW	cd01650	403	669

					TTGACTTGACTTGAC	LNNVTRAARRECRKAERKFKKDKLQVSLQMLKDSWRHYQATVKEAKREHLSNVILLNRHKPRVLFSTIDAALNAPKSVCVQPTLETCETFLKFFVDKVNNVRALISPPGFDPSVVVSCPVTFDQFEPVTLSVLYETVKHLKPSGSPSDAVPPRLFKEVITTLGPAVLAVVNGSLSSGVFPRNCKHATVQPLIKKPNLDPSVLSNFRPISKLPFLSKILEKIVYVQLKDFLDENGNLEVFQSGFKSLHSTESALLRVFNDILLATDAGDHVVLVLLDLTAAFDTLDHNILISRLQQLVGIRGTALQWFRSYLEDRTFCVNLNNVQSSSAPLLYGVPQGSVLGPLLFSLYLLPLGSILRKHNCAFHCYADDTQIYVPITKKDAHSISPLIKCLEELKIW			

						MALNFLHLNEEKTEVIV FGPSGNSESHPSHTLVTP ESHPIALGSLTQSVKSTV TNLGVKMDNGFNLDKQ ISSVVRSSFFQLRQLAKV KSFLSRKHFETVIHAFVT TRLDYCNSLYFGAPHSS LSRLQLVQNAAARLLTG TWKREHITPVLASLHWL PVHFRVRFKILLFVFKSL NGLAPPYLSELLHLHSPS RSLRSADQLLLDVPRTR RKLRGDRAFAVAGPSL WNALPMQVRQARSLND FKSKLKTHLFSLAFNSV			
L2	L2-4_OL	.	Oryzias latipes	AGATTTGTTTTT ATTGACAATTTT AATCTTTTCTGG AACAGAACCTC ATTCTATAAAA CTGATGGGATT CACCCCAGTCC ACTGGGCAGCA	T	MRGSAGRKANTNWLRP LKEAPKPEEAASTTTPPP SFNMALFNVRSLSNKTFI INDFIHDYNIDGLFLTET WLSTDASTVLTEASPPN FSFVFSTREGKKGGGTA SIMSNSFAVKQITLDEYP SFEHHAFVFNSPPVLCIT	IPR000477	521	582

				ATGTTTTAGCTG		VYRPPKPSSVFIEEFSEFL			
				CMAACATCAGA		ASVHTSFNMILITGDFNL			
				CACTCTGTCCAG		HIDGKSDYCTTQFLDLLS			
				ACATCACCGGC		SLDFKQHVMQPTHNRG			
				CCCCATCATGCC		HTLDLVISHGLPTSVSAV			
				CMCACCAGTCC		VDLAVSDHYCVLFNITG			
				AGACCTCCTCCT		FIKQETSVRTVKKRYLTP			
				CATCCGTGCAG		EVVEKFITSVRSYSPPVL			
				ACATCGGTCTGT		PAPCDLLVESFNSKLKSS			
				GGAAGTGACTG		LDSVAPLXSKKIRIKATP			
				ACTGGATGTGG		PWRTEEIKQLKRRCRSA			
				TATCACCCCCCC		ERLWRKHKLEINYQLYQ			
				ACATTCCCCAA		NQLKLYNQTVKQTRNS			
				AAAACTGCTTC		YFANIISVNKNNPRVLFT			
				AATCGACAGTG		TIESLINSNVNKNTMPAS			
				CTTCCTTGCCAC		ASLCEDFADHFRSKIDSV			
				CACCTACACCC		RSNIIKGQKSLFIGNDGLI			
				CCCCCAGTCTAT		LPEEALDSFDLVEAATL			
				CCCATTAAGAC		GRVFSQVNPTTCFLDPIP			
				CATTATTACAG		TSLFKSFYGSFESEILNIM			
				ACAGGTCTTCC		NSSLQTGIFPAAFKTAVV			
						KPLLKKTNLDPSVFNNY			
						RPVSNLPFLSKVLEKLVF			

						IQLNEFLLQNNIFEINQSG FRTNHSTETALLKIVND XRCA			
RTE	RTE-1_PPo	.	Papilio polytes		TAAGGGATAAGGG CAAGTGGATGATGA TGATTGTAGA	TYVRSAAGDSDRFGVSV GLHQGSALSPYLFLLVM DALTLNIQEEAPWCMLY ADDIVLVGEDECEVQSR LEEWRVKLEGAGLKISR TKTEHLFCDFGGPTSFSP IALNGVPLPTCSDFKYLG SLVQSDGGVDRDVKSRI NAGWMKWRQVTGVTC DPRMPLRLKGQIYKSIIR PVVLYGSECWATKVAD ERRLHVAEMRMLRWM CGVTRLDKVKNVYIRGS LKVAPVVEKLRSKRLA WYGHIMRRNDGDINQK VLRLNVDGYKGRGRPK KVWMDCVKEDMLKKG VDSDMTADRDVWKSST YCADPP	SSF56672	18	134

RTE	RTE-4_PPo	.	Papilio polytes	ACCGGAGGACG GGCAGCAGCGT CCCTTTCAAAA ACGGTGGAGAA GTCTGCGATCG CCAACCCGCCT GCCAAGCGTGG CGATTATGGCA AATCCCCCCAA	TGACAATGATGATG ATGA	MMATRKVRPPVPGGPA VPGYGSGQGNGGAGGA KNPRLGYGYHTRRPLAL ATLNTRTLRTDERIAELE EELSRLRWDIIGLSEVRR EGEETIILKSGNLFFFREG DQQSQGGVGFIVHKSLV NNVVKIESVSTRVAYLIL RISKRYSLKVIQVYAPTS THSDDEVETMYEDISRA MHSSKTFYTVVMGDFN AKLGIRDGDELRMGKFG YGRRNHRGHLLAGFME KEGLYMMNSFFKKQGH RKWTWMSPDGATKNEI DFIMSTKKQIFNDVSVIN AVKTGSDHRIVRGTLNI DVKLERCRLMGSTLRPA PVHIHNAETFQLELRNQF ECLGSSVDDLNDGFVHA LQSVGSKFYKTHRKSSP KKLSDHTLELMKERRD MKLHSSDDAERYRRLN	cd01650	549	822

						RQISTSIKRDIRRFNTNR			
						VKLAIERNQGSKVFSRD			
						LSIGQSQLLKLKTGEGR			
						VVSSRPELLSEVERFYGQ			
						LYATQEPVTDLVEDPRA			
						RLTRHYTEDIPDVSLDEI			
						SMALKQLKNNKAPGND			
						GITAELLRAAGEPALKV			
						LQKLFNSVILEGHTPKA			
						WSRSVVVLFFKKGDKA			
						LLKNYRPISLLSHVYKLF			
						SRVITNRLARRFDDFQPP			
						EQAGFRKGFSTVDHIHT			
						LRQVIQKTEEYNLPLCL			
						AFVDYEKAFDSVETWA			
						VLQSLQRCRIDYRYIEAL			
						RCLYENATMSVRVQSED			
						TKPIHLQRGVRQGDVISP			
						KLFTAALEDVFKVXKYC			
						HLPETLHRGVGRRFQGP			
						EILVVTYTFLHVTFRVQ			
						VIGSLLLLRILIKIGTMLE			
						DLNRASERVGLRMNMD			

						KTKVMSNVHVSPTPITIG GHTLEVVDEYVYLGQTI QLGRSNFEREVNRRIRL GWAIAKXNNRRSGSCGI SSRPKTYRSVSRQKSLTS VCCQHTGQRRGHSLRAT SGGSKSLKGQWSGQCSE SLCEIASGTMRSAAESVT DIARRIANLKWRWAGHI ARRTDGRWGRRVLEWR PRIGRRSVGRPPARWTD DLVKVAGVHWMRAAQ DRALWQSMGEAYVQQ WTTYG			
RTE	RTE-6_PPo	.	Papilio polytes	GTACGTGCGACTCGTGAKTGCTATGTATGA	TAAGGAGATGATGATGATGATGTATG	KSEHSRQVRSRRNRQVQ CGCGPAPGIGXSPYLFLL VIDALTVEIQEEAPWCM LFADDIVLVGEDALEVQ SRLERWRDKLETAGLKI SRAKTEQLFCDFGGLSSF IPIALDGAHLPVCSDFRY LGSLIQSDGYIDRDVTNR IKAGWMKWRQVTGVIC	SSF56672	25	135

						DPQIPLKLKGKIYKAVV RPVMMYGSECWAVKGT DERRLHVTEMRMLRWM CGVTRMDKIRNEYIRGS LRVAPIVEKVRSKRLAW YGHVMRRDESHVTRRT LVMSVEGYIGRGRPRKR WMDCVKGDMIKKGVM TEMTYDRVMWRTKTCC ADPTGVG			
RTE	RTE-19_Ppo	.	Papilio polytes	GGAAAAGAAGG GAGATATTAAA AAATAAAAGGA GTTTGAAGCAG GTATATGTTACT GAAGATTTTCC GAAAGAAGTAT TAGAAAAGCGA AAAGAATTACA AATAAAGGTTA AAGAGGAGCGT GCTAAAGGGAA TATAGCGTATAT	TGAAAATGTACATA CATACTTTATTAAA AAAAAAACTTTTTA TATTTAACTTTAAC TATTTATATTGTTAT TAATAAGTTGTAAA TATTGTAAATTGTC AGCGAATAAAGGC TTTATCTATCTATCT A	MMRPRSSSLSLTKPNQIK QLNTSNQKRNNPNPAGH RGVARPLPPKSMSSGSK LKSKLYLATLNVRTLME DHRLIELEHAISQIKWDII GLCEVRRSGDEIREYNK YIFYYKGEKKGRRGIGF LINKYLKDKIISFEGISDR IAVLNILNGFKHPMSIIQ AYAPTEQDTNAEKDKFY EKLTEVIQKSNKNIIVMG DFNAKTGVKSNKDEFVL GKNSCGERNENGQRLID	cd01650	581	833

				AAAATATGATA		LCFASNLKIANSFYKKE			
				AGATAGTAGTC		KSSGRGCLQTEGTEMKL			
				AAAGAAGGCCA		ISFQFLQKRKVKWTWM			
				TTTAGTCAAAG		SPNGRNRNEIDFILTRET			
				AAAAAAGGAAA		RILSDVSVINKLNFSTDH			
				AGAGATCAATC		RMLRGTLICEDKNSRKH			
				AACATCACCAG		IISNKNGKTITVPHNILED			
				AAAATCAAAGT		IEKTLQDIKKQQTIQDKY			
				TATGCCAAAAA		DFLEKTLKEVNIFGTKR			
				ACTGGTTAGCA		GKRNKLNEETKKILEER			
				ATAATTCATTGA		NELLDNRRENRHSIANIS			
				TAAAGGAAAAT		KKIQSSIRKHRKEERLKT			
				AGGAAAAACGC		LKEQIEKTGGMTKGLKQ			
				TTTCGAC		LKEYSQWIPNIKLQKKK			
						ETYRITSRRSSIAKTATEF			
						YKQLYTESGESIVASKNI			
						IEEGDSNEEEIPTILETEV			
						AGAIRSQKNDKAPGDDK			
						IPNELLKSTLSSITKALTI			
						LFNEIINSEQIPVQWTKS			
						TITLLHKKGDKQDINNY			
						RPISLMSNIYKVFSKVIL			
						GRVTQTLDANQPREQA			

						GFRADYSTMDHINVVK QIIEKCNEYGQNYYLAF VDYSKAFDSLSHNYIWK TLASQGIERKYINIIKEIY KNSTAKIQLEQVGEEFKI ERGVRQRDPLSPKIFTAV LENIFRNIDWDKFGININ GTKLNHLRFADGIVLFA ETPETLRTMLQQLDQES RRAGLSMNPMKTKLMT NAEKDIIELRHEIIEYVDE YIYLGQIISVNQHSKEID RRISNAWKRYWSLKEV VKKKSIPMSIKTKLYNTC ILPCLTYGCQTWPSTTK HNLKLANCQRHMERSM LDIKLRNRWRNVTIREK NKSKEKKTKVNDVFKKI KKLKWNWTGHIIRTNK GKWTKDIMDWYPREGK RQRGRQIKRWEDDLPKG WRRLARDRETWKKLGE AYVEGQPDTQLDAGK			

CR1	CR1-2_PH	.	Parhyale hawaiensis	CCAGTTACCCTCCGTGGCGGTCGGTTGTGACGTGTACGGTGCTCCGCTAGAGTAGATAAAATCAGAAAGCTATAGATTTTTCTCGAAAGCAATAAACTAATTAAGTGTTAAGCAGTGTGCCAGCCTGGCTATAAACTCGTACGGACCGTGTGACCAGCGGGACCAGTGACCACACCGTAGCTCGCTTAATAGGGGAAGCGTAGACGTCAGTT	TAACGCCGCCTGGACCTCGTCCGAAGTATCATCAAAGGTATCAAAGGTA	MGGNKKGRNKTQFKPVILNSTDDLDQELEPQETRGVNECTLCLTSKQSKMIQCVLCEENNFHASCLALNQSGRYHAPVCPSGHGCSVNERSTEKQHWVCTGCRGPTTNATPKVACTKCEFLFCAPCAKLRPGMIDMFNRFPKSFMWNCERCLQVSTQNRKCVASDTKESVGESPKTAGEEPLQKYAETGDHDVKEISYEIAQSTVQKRSSSGGDWISKNEKKPRICKYYAKNQCKYGISGRGCKFFHPKRCPRFMKYGRQGCNSKRECNFFHPWLCQESFRNGTCFDTTCPEIHLKGTIRAAPTAHHRHTHAERGENSGRPWQNSNRIQTNKYGENNHQYNVERDITANDRDMSVYNQDVRRDLQNMQNFLQTLYEKISSMETYME	cd01650	959	1215

						KVQHGGPRAPWNPPRV ENPTQLWGDPGTTHQLK TQTAIHHAKQNEHQELK SNKDEILSKVHVINIRGL LTKNCRKQMVLQDIIND AGSKLLAVTETHLDDAV NDAEVKIENYVLHRSDR KDRKCGGVALYVRDHL GSHTKQKFSNGVCDCIIV NIEKLRLMCVVCYRPPK CTTEKFAEMLQVINTHC KEVQSKYQLYITGDFNF PFLKWLTTDLPDETGSF VLSRTSSATRDEKDQAL SLVRIMEEFSLTQMISSP TRNENVLDLLFVDNAQL IKNIEIQDTTLSDHRLVEI TTQLGERDSHRPQKNSQ INSISSLRDLNVHDKGIN WDGLNAELQSVNWDVL LNNSIESESINETFENILL EKCMKYLSPAKRKFNKS RHHKFRKGLLKKRAKLI			

						KCQQSTKLLQRKSECKK KLDEIERQLARSHDEEK QEKEKKAIEKIKSNSKYF FAYASHHSSVKSKIGPLV NDTGSIVXDDQEMSEILS RQYQKVFSDKIPNPEIPRI DLTTNYLTDFEFTKLSIE LAIDELSGNSAAGPDDV PAKILKMCKLSISYPLYL MWRKSMDDGVVPLTYK HAIISPIPKAGNSMDPAN YRPVSLTSHVVKTFERV FRKHLVHHLESNNLLTS NQHGFRHKRSTTSQLLE HYDSILDLLCQGKSVDVI YLDFAKAFDKVDHGILL LKLFNMGIRGKCLRWIE AFLTNRTQSVIINNKVSK RQPVLSGVPQGTVLGPIL FLIMINDMGHQCKFAVP SSFADDTRILAAVQSTA DVRNVQLDLTNIFHWTS VNNMTFNESKFELLRYG			

						KNFSLTESTSYTTRDGSE IERKQILRDLGVRIQDNA TFEDHVARVVTVGKKW SGWVLRTFSSREADTML TLWKQLILPRIEYCSPLW MPSKRKDLERLEAIQRSF TAKIRGMKDLNYHERLA QLQMYSIQRRFERFTIIY VWKILEGITVNVNGRIQ ARHSNRIGRTCEVRIPIK TGGDAIATKYSNSFVVR GARLFNKMPRHIRNVTG TETAKFKHLLDACLLKV PDKPAVSGYVNCGDNSL LATPLVTDCSITSAQDG WRA			
CR1	CR1-4_PH	.	Parhyale hawaiensis	GCGTGGCCTGC GCGTTATCAGG CCGCCCCCATTG TCCGGCCACCG GACGCCGCTTT GTTTGTATGAGC TGGCTTTGGGTT	TGATGGCTCTGTCT TTGCTGGTGCGGGT ATTGGCAGCTTTAA ACGATCTGTACACA GAGCTATTCTGTGA TCTCGGATCTTCTTT TTTTTCATTCTTTTG	MVYVYLLFFACLITRYP DVEENPGPRGFGSRSCRI LCTNVRGLYGNLAELAV ASSNYDLVLCSETLVSS HRHVSELMIPNFEKPVL RLRDSCPGFRGLAAYVR KGVVVSRQSEYECPCCE	cd01650	514	772

				TGGGGCTCTGG	TTTTCTTCACCCGCT	MMVLRVCGSFRNFYLFS			
				GTAGCCCCTGA	GGGCTGGTTGCGTG	VYRSPSADDRVFDCLLD			
				GTTGGAGGAAA	GGGGTTTGTTAACC	AMARIQSSDRRSAFAFM			
				TGCCACTCTAA	TCATGGAGTTTCTC	GDFNCHHVEWLGSGRT			
				AGGCGGATGAG	CGTTCCCACGTCGC	DCHGHAAFDFASASGCE			
				ACCTTTTTGGTC	CAGTCTGGTGTGAT	QLVEGPTHLMGGVLDLL			
				CAACGGCATCC	TTAATAATAATAA	MTDVSDLVGVSVVSPIG			
				CTGAGAGGAGA		SSDHGALSATLQVRQRV			
				TGAACAGGCTG		PRVTVRREIFLKSRVNW			
				GTGCCATAGCT		DAVSGAILALPWGVVRE			
				GGGTACTCAGT		STDPGSTLDRLLSDMVR			
				CTTCCGCAGTCT		AYVPSRFITVRTGDEPW			
				GCTCCTCTCTTC		FDDACRRAHQLKREAY			
				CTTCCCTCGGGT		ARWRTSRSAVDFQLFRE			
				AGTACCACCTG		QRRRAGATYEEAEREHN			
				ATGCTCTTTTCC		LRSRELLASSVSAHKWW			
				TTTCCTTTCTCT		STLKSAVLGVSSSVPPLI			
				TCTTCCTCTTTT		GTGGDLVADPAAKADL			
				CGGCCAAAGCC		LLRYFDGKQSRDRVRIP			
				CGTACTGGTGA		LACDPEPRFSGFAFRSSE			
				TGGCAGGGTCT		VHRLLLDLDPYGGTDPS			
				GGGGCGCGTGG		GMFPFFLRRVAGVLAAP			
				ATAGCTGGCCG		LASVFRRMLRLGSFPEC			

				CTTTCGTTGCTA TCACGACTGTC GATTTGCTTTTT TGTAATTTTTTT GATGATTTCCTT TGCTT		WRVAHVTPVPKGSVSSR TSDYRPISITPVLSKVYE RLVASRLGCFLEKQGVL PPTQYGFRKGLGTCDAL LHISHVLQLALDGGHEA RLVQLDFSAAFDRVNHL GLLHKLRLMGIGGSVLA VISQFLSDRRQCVVVDG KRSEFLDLVSGVPQGSV LGPLLFIVYTADLFHLVN NLLVGYADDVTLLAVC DKPSDRLEVSESIQHDLE MIDSWCSRWGMRLNAG KTKTMLVSRSRTVGPPF PALMLDGVSLPVVDDLV ILGVNFDPKLTFEPHIRS VASRASQKIGLMRRAW RVFGSVELTRRCFNCFV LPLLEYCSVVWGSAADC HLCLLDKVVRRCAFML GGVVPCDLAHRRRVGG VCFLYKIYNCPSHPLSEF LPDPFVRVRYTRRTGAL			

						HEFALEPLRCRTGQYSR SFFSIVCGCLEWS			
CR1	CR1-3_PH	.	Parhyale hawaiensis	CAGAAAATAAT CAACATTTGACT GTGGTGTATAT ACAAATACAAG TGTAGTGCTAA AATGTGTTCGAT TAWACTCTTAT TCATTGCTCTCG AGTGCTACCGC TAGATATTTTCT CCATTGCTTTTG GTTCTTTTGGTT GGATTCTGTGGT CGAACTTAGAA GACGCTGATCC TGAGTTACTAA GTAATATTATAT TCTATCAAATTT TCCTCTATCAAT ATTTTTTCCATC TTCCTCCTCATT	TAATGGCAATAGCT CATAGAGGGCCGG GGCGTAAGTTCCGG TGTCTCTTTGGGAA TGGGCCTATTGGTT CGTCTACTGCCGCC TATCTAATCTAAT	MKQCDECHSVFPLEFFY YEEVELESCKYCCKIQA LQAEIEKLESDREAMQR GDGGEKEKAPREAHLRL SSPSPWKTVMRNKRQRR STGAILEQTNHIKVNNRF AAFNDGDDNVATSVRP KIKLKTVNGEQHRGPPQ SRDQKKSETSVQPPSHE RERNPAPIPGQRAIKTFLI GDQIIKGQYYAFKRGCN DRKVICRPQASTSDATAI LKDLRCKKMDHVIIHVG TNDLRTYSETQLQLVKD LIAAAKKKSDHVTLTSIL PRLRAGNRLYSRIRYING QIKREAEXNGLEFIDLFP SFDDDWRLFLPDGVNLN GAGHHVLGEMLLAAAA SETAALEEGLRTRPKTIIE QPRHPAPFHALPADLPA	cd01650	1136	1383

				GACTTGCCTGAT		MEQTATNVPAAVPCNP			
				CCTCGCCAAAG		QISSPSGMKVGDPKGNE			
				GGACAAGTATG		GTIITNSLDTDEEEITPTP			
				ACTCCTGAACTT		DASVQMSRLGPPPLLPSI			
				CCCGAGTGGCT		VSNLVSTNSNMSEGTDV			
				ATAAGTTTGCCT		PQPTNARQTDLDSSNTE			
				TACTATAGGTG		RGSPVSVSKEKSQSQTP			
				GGATTAAAAAT		VAKKRGQFPRTSSIPAHE			
				TGTGTGCTTCGG		GNRSTRKSGARNIPPKT			
				TCTCAACTTTTT		NSCQPPQTRITRSKANSA			
				TTTAGTTTTCAA		GNDLLSARLGGKRPIVR			
				GTAGATAACAC		QIAADVKQSDNCQAEIQ			
				CTTTTCAAGAAC		NISLAXLQTGALAATDN			
				GCTAATATTTTT		AFADLTPPIFSQQSDAPS			
				TTTTTCAAACCT		VSVTRRSNANTSVKFLF			
				TTCTAGAGACT		GNVQSITNKMNELQAIA			
				AGGAGGTACAC		ATLAYDVIALNETWINL			
				TAGCGAATATTT		SNKNLEAEVALKGYKLF			
				ATTCTCAAACTC		NIDRPTETGRGGGSIIYIK			
				CTTACGTTTTGG		EHLNPLCKGFAATKEYE			
				ATCTTCATAAA		FITTHARLGIRQSLNIML			
				ATATTATTTTTT		AYRSPRSTAETDAALYN			
				ATTTCTTTTTTT		KIENALNTPNETILMGDF			

				TCTCTCATTCAT		NLPQIDWSTLCSPAPGSK			
				TTTCTTTTCTAA		LLELLSHHHLTQHVNEP			
				CAGACTGTAGG		TRYGNILDLVFSTPDCEI			
				TTTAGTTCAGAA		VDLAILDRLGDHNMIEF			
				ATGTTCTACATG		TLKTGNEFTTNASRLKL			
				AAATGAAAAAA		PDYRRANFQGMATEIET			
				AAGAGCTTGAA		TDFETPMRDASVEQCSH			
				ACATATGATAA		ILKTSILRAAERYIPNKR			
				TTGAGTTCATTC		VRISNDPLWMNRTVKRC			
				CTAACATTGCC		IAGDKCVTKPGSVIQQR			
				AAGATAAACAG		RTGPVTSPLDESDSQTLH			
				TTCGGAGCTAT		CRRQVRYKAWKRNPTE			
				AGGCCCGATGT		ENRSSHIRACRDVKKIVR			
				TCGATCTATTGA		TTKREKEVSVAAMAKR			
				TAATTACCAATC		NPKAFYKYVNERRLKKE			
				TTAGCGGCGTG		TIGPLKRPDGXITNDLRD			
				AGGCAAAACAA		KATILNDYFTSVFTVEN			
				AGTAACACTCC		VTNIPEPAPMQVAPINTC			
				AGAGTTCCTGG		VFTEIEILQELTHLNVYK			
				CAATTTTTGTTT		STGPDNLTPRVLREMAP			
				AATTCTTGTTTC		VIVKPLLEIFNRSLRTGE			
				TCTACTCCTCCC		VPQDWKLANVIPIHKKG			
				AAAACTCTCAA		DKKSPSNYRPISLTSVVS			

				ACGACTTCCAA		KTIERLLKKRITDHLRM			
				TAGTCTTCATTT		HQLISDTQHGFRNGRSC			
				TATTTCAAGACT		LTNLLDFFSKVTDIKDEE			
				TTTAATACATCT		KGVDLVYLDFQKAFDK			
				TATTGTAGTTTT		VPHQRLMKKLEAHGIQG			
				CTTTTATTTTGA		DYLRWIRAWLTGRQQR			
				CATGTTAAAAT		VKLEGETSDWTQVSSGV			
				GCCTAAWTCGT		PQGSVLGPLLFSIYINDL			
				TGCTGCAATGT		DLNIRSKISKFADDTKLC			
				ACGTAAAAATT		HPATKRAELQQDLDHL			
				AGAACTTGATC		MEWSTTWQMSFNVDKC			
				TTTTTTTGTTCA		SVMHVGLVNPSTPLKM			
				TCTTTCTAGGAA		NGTELAAVTQQRDLGITI			
				CATTTGTGTACG		TNKLQHSQQVDKSVKM			
				GTGCAACGGGT		ASRVLGFITRNFEYKSKD			
				TATATGTTATAT		IIVPLYKALVRPHLEFAV			
				ATTATATACAGT		QFWSPHLCKDIDNMERV			
				TCCTTTGTTTAC		QRRATKLIPQLRNKPYE			
				CTCGTCAAGGA		CRLKELGLISLEQRRLRG			
				GGTTATGTTTTT		QLIETFKYLNGFNDVPTT			
				CTCGTGTTTGTC		GLFERDLNDRLRNNGQK			
				TGTCCCTTGTCT		LKHKARRTTLSQQFFPY			
				GCCTCTTGTAAG		FIVSIWNRLPSDVVSAES			

				TCACCAAGATT		VNAFKNRLDIHWERHPP			
				ACTCTAAAACT		QLRTPPPGVASSELLLQE			
				AATGAACGGAC		SQALNGQSFRGHLNPSS			
				TCTTATGAAATT		SLSRSNYTSLSSAYQRLL			
				TACAGGGATTG		IVLACIHDRLFGWDSSPP			
				TTAGACATGTTG		SFSFFFLFLPFFSSLSPVL			
				ACAGTAACAGA		YILLT			
				TGATTAACCTTT					
				GGTAAAAATCC					
				GTTCACCTGTCT					
				GTCTGCCTCTTT					
				TTTCGTGCGCCA					
				ACTCAAAAGGT					
				TCCGGACGGAT					
				TTTTATGAAGTT					
				TATATGATTTGT					
				TGGGTATAAAA					
				CGTAGAGTAAT					
				TGACCGAAATT					
				TTGGTGATATGT					
				TTGTTCGACTCT					
				GTGATGTGTTTG					
				TTCGTCTCTGTG					

				ATGATATGTTTG TTCGTCTCTTTG TTTCTCGCTGTG TTTGATCTTGAT ATTATGCTTGTT AGCACAATAAC TCGAAAAGCTG TGGATGGATCC TCATGAAATTTA CAGGACATGTA GTAATGTTAAA AGAAATAAAGG ATTAAACTTTGG TGAGGATCGGT ATGTTCGTTTGG GTGTTGAACGG ATCTTCATGAA AATTACAGGAA TTTTTTTAATTC TACAGAAAGGC ACAGCTGATTA GATTTTGGTGTT GATACAAAGAA					

				ATTCATGGTTTC AGAGCCAATTT GAATTTTTCGTC ACCGTTGCCTTG CTTGGCGGAGG TATGCGCTCTAG TTTGTAATGAAC TTGCATGTAGG GCATCTTATTTT ACTCCTTATCAA ATTATGTCCTTT CTTGCATGTAG GTTTCTTTTCCT TTAATACATTGT AACTTACTATA GCTATATCTAG GTAGCCTTATTT GCATTTTTTACT TGTTTCATAGGA ATTGCTTGGGG GCAGGATAACA CGTTAAGATATT TTCTTTGTTTTT					

				CATCCTTGTGTT TGTAATTTAACT GCCAAGTTCTTA TGTACACTGTTT AATCAGTTAAG CATATACAGTA TTTGCGATGTAC AGTATACGGAG GTGCTACTTTGC ATGAGGAAACT TATTAACAAAG TTCATTAATCTA GTTCAAACTGTT CAGGCGATTTC GCTCGAATCTTG GGCCATAGCTTT AGTAACGGCCC TATCGATGGGG CACGCGTGACG TTTTTGCATTGA TGCCCCTGTCCC GAATTCCCGGT ACCGAGTAACT					

				GTTCCATTTCTG TTTTTCTCCCTT TCCTCTTCTCCT TTCATGTCCTCT TGTGCCGGCAG CACAGGCCTAC AAT					
Vingi	Vingi-1_PH	.	Parhyale hawaiensis	CCACCCGATGC CTGTGCGTCTTC GTCGTGGGTGG ACCCTGCGGGA GCTCTCTGCGCG GCTGCGACGGG GACTTCCGCGG GCCAAACGCAG GCACTCATCGTT GGAGTGGCTAC GCGGGTCGCTG TGGCGCTCCCTT TCTTCCTCTCTT ACTCTCTTCTCC CGCTATTCTTTC TCGCTGTCTCGT	TAACTACAACAACA ACA	MPCSGLFSAASYHPDWR CPSCAPQPPRPPSPPPASP LSPHPTNAPPLIVMQFNC NGLLHCRSELSAFLHSSN IAIAALQETKLTPSSSLPV FPGYTPVRKDRPAANGG GGLLFLIRQDIPYQQATT NHLFPNDHTVEHSGVTV TLCSSSLTLINIYVPPATS CPPGYRLDSSALFADSSA DTLIMGDFNAHNTDWY SVTQDARAAARGEEISA SINNGPYCLLNGDSPTRV PRDGPGSSPDLTVISAHL ALDYSWSTTTTLSSDHL PILITAESSDSAPSGPRRT	cd01650	536	795

				GTTTTTTTCCTT		FVNFRKANWEEFTRECE			
				CACGGACAGGT		ELITSLPLPSSCSGGLRAL			
				CCTTTCTGCCGA		NRALLSAAKHHIPAGHR			
				CGACCTGCCCG		KEYIPRLTPDARDAISRR			
				CTCTTGTCTAGG		DHLRSTSPHDPSLPQLQQ			
				CAGCCTTTTCCT		EVDRLICGAAQEAWKQ			
				TACTTGTTTCCC		TIQDCDPKSNPSKFWSLL			
				CCTCCCGCTGTC		RSLSGKQSSVPRNQPISF			
				CCTTTTCTCCCC		GTHTFTHNGDIARGFCK			
				CTATCCATTCTA		QFTATSPPALATPPRLAR			
				CTAAGCGGTGA		RLLRKRHPIAHGFSPFTP			
				TCATCCACCCCA		AQVREAIVASGSSTAVG			
				ACCCCGGCCCC		PDQLSILHLRHLGPQGII			
				GCCTGTCCCTGC		YLTRLYNLSLAHADIPA			
				TCAGTCTGCTCG		LWKLAVIIAILKPSKPAD			
				CGGAACGTCAC		KGSSYRPISLLCPAAKVL			
				CCGCCGGGGGT		ERLLLPFINEAYQPAPHQ			
				GGTCTTTTCTGT		HGFRPMHSTTSALLPLA			
				GTTCCGCCTGCG		FQATRGMNKPKPHDRT			
				CGCGCTGGGTC		VALLIDFSKAFDTVPTEG			
				CAC		LLSSVSGTTLHHNVVRW			
						LAAYLRGRAACCLYNG			
						SRSQFRAVKAGVPQGSV			

						VSPSLFNLYVSDFPIPRD			
						QLCYPSVTSYADDFTVV			
						ASHRDPLTAAAALSDHS			
						ASVSNWAANKGLTISTS			
						KSNSTLWTTDTHQSTLD			
						PGITMNDSALPLERRPKI			
						LGVTFDTHFTFTPHITST			
						AARASSRLRLLKALAGT			
						RWGQCKETLLLTYRAFV			
						RPLITYAAPVWYPGARQ			
						SAVRKLQTVQNAALRT			
						ATGCLRMASVDHLHAE			
						SKXLRIDQSLDLQCRQY			
						LVSALRPEHPAHSVVTA			
						QAGPRLIRHNLATKFQP			
						DINHLLTDNIMPSPEYRS			
						AISSLHRSAVQASLDRLP			
						VNRVLGFAPPEISPDEAV			
						LSRAARSTLSQLRSGFCS			
						ALMSYRHRIGLADSDSC			
						PECHRAPHTVNHLFECP			
						AHPTALLPTSLWDSPLS			

						VTDFLCSLPSFSYLPPHA RPPPEPPPT			
CR1	CR1- 2_PSi	.	Pelodiscus sinensis	GGGTGGTGGTG GTCGGGGACTC CCTCCTAAGAG GGACTGAGCCA TCCCTCTGCCGT CCGGACCTGGA ATCTCGAGAGG TGTGCTGCTTAC CGGGAGCTCGC ATTCAGGATGT CACTGAGAGGC TTACCAAGCTG ATCAAACCTTC GGATCGCTACC CCTTCCTGCTTC TCCATGTGGGA ACTAATGATAC GGCCAAGAATG ACCTTGAGCGG GTTACTGCGGA TTATGTAGCGCT	TAGATGGGATTGGT CCTGCCTAGAGCGG AGGGCTGGACTTGA CGACCTTCTGAGGT CTCTTCCAGTTCTA TGATTCTATGATAA GGGGATGGTTGGAT GAAATAACATGATC TTGGTAACTAATTG ACCATTCATTTCCA GTTGGAAATAGGTC AATGGAGGGATGA TAGGAGTTGCTATA GGGAACTTTCTGGG TGTCTGGCTGGTGA GTCTTGCCCACATG CTCAGGGTTTAGCT GATCGCCATATTTG GGGTCAGGAAGGA ATTTTCCTCCAGGG TAGATTGGCAGAGG	MHRARSEKGGPQFRME KIVQSTGYLKCLYTNAR SLGNKQEELEALAQAKK YDLIGITETWWDDSHD WSAVMEGYRLFRKNRQ GRKGGGVALYVREHYD CSELQYKEGEKLVESLW VRFKGANSSDVVVGVC YRPPNQVDEVDEAFFEQ LSEASRSQALVLVGDFN HPDICWETNTAVHRQSR KFLENVGDNFLVQVLKD PTRGRAQLDLLLTNREE LIGDVEVGDNLGSSDHE MVDFRILTKGRKESSKIH TLDFKKADFDSLRDLME RIPWDVNMKGKGVQDS WQYFKEALLKAQKETIP TRSKRGKPGRRPDWLTG EILGVLKHKKEAYRKW KLGQMTREELKCIVREC	cd0165 0	502	753

				GGGAAGAAGGA	CCCTGGAGGTTTTT	RGVIRKAKAQMELRLA			
				TCCAAGAATTC	CGCCTTCCTCTGTA	KDVKDNKKGFYRHVNK			
				GGAGCGCAAGT	GCATTGGGCACAGA	KKVIREGVQSLMDEGGN			
				GGTGTTCTCGTC	TCACAGCTGAAGGA	LVTEDVGKAEVLNAFFA			
				CATCCTCCCTGT	CTCTCTGCATGTTG	SVFTDKVGSWTSVLSDA			
				TGAAGGGAAAG	GGGCCTTCAAAGTA	RWEEDGQPLVGKEQVR			
				GACTGGGCGGG	TTTGAAGGCTTCAA	NYLEKLNVHKSMGPDL			
				GATCGTCGAAT	TATCTGAGACATAG	MHPRVLKELANVIEEPL			
				TGAGGACGTAA	GTGAGAGGATTATT	AIIFEKSWRSGEIPDDWK			
				ATGCGTGGTTG	CTAAGAGGGGTGG	KANVVPIFKKGKKDDPG			
				CGCAGGTGGTG	GCGAGATTCTGTGG	NYRPVSLTSVPGKIMEGI			
				TCGCAGAGAGG	CCTGCACTGTGCAG	IKESILRHLNERKVIRNS			
				GCTTTGGTTTCT	GGGGTCAGACTAG	QHGFTKGKSCLTNLISFY			
				TCGACCATGGG	ATGATCATAATGGT	DEVTGSVDMGKSVDVIY			
				ACTCTGTTCCGG	CCCTTCTGACCTTA	LDFSKAFDTVSHNILASK			
				GCGCAAGGATT	AAGTCTATGAGTCT	LRECGLDKWTVRWIER			
				GTTAGGAAGAG	ATGAGTCTATGA	WLEGRAQRVVINGSMS			
				ATGGGATCCAC		GWRSVSSGVPQGSVLGP			
				CTAACGAAGAG		VLFNIFINDLDEGIDCTLS			
				AGGAAGGAGCA		KFADDTKLGGEVDTLK			
				TCTTCGCGGGC		GRDRVQSDLDKLEDWA			
				AGGCTTGCAAA		TRNLMRFNKDKCRVLH			
				CCTAGTGAGGA		LGRKNPKHSYRLGTNQL			

				GGGCTTTAAAC TAGGTTCGTCG GGGGACGGTGA CCAAAACCCTG AGGGGAGTGGG AAAGTCAGATA CCGGGAAGAA		SSSSAEKDLGVTVDEKL DMSQQCALVAKKANGI LGCIKRSIASRSRDVIIPL YSALVRPHLEYCVQFW APHYKKDADALERVQR RATKMIRGLEHMTYEER LRDLGLFSLQKRRVRGD LIAAFNFLKGGSKEDGE RLFSVVTDGRTRSNGLK LWWERSRLDIRKNYFTR RVVKHWNGLPREVVES PSLEVFKSRLDKALAGLI			
RTE	RTE-14_PSi	.	Pelodiscus sinensis	CTGGTGCCAAG CTGTACTGGCTT TCGCCTTTCCCT TGGGTTACATC AGTGGCGTGGA GAGGGGGATCC TGCTGCTGGGC ATCAGCCGGTT CTCAACTTACTT GCCCAGGCCTA CGCCCTGGAGA	TGATCCTTCGG	MIPGLTEDLRSISNPRKT AVINNELRRLRMDIVAL QETRLSSSGCLREKDFSF FWHGKPPEETREHGVGF AIRNNLVGSVIPPNVGTE RMLKIQLQTSVGMANIF SVYAPTLTSAQEDKDKF YDELSIAISEVSSHEPLFI LGDFNARVGSDHHSWPS CLGQFGTGKMNENGQR LLELCSQHDLCVTNTFF	cd01650	525	802

				GGACACTCCGG		NVKPQHKVSWRHPRSK			
				CATCGCTTTCTG		HWHQLDLVLTKRSRLGS			
				AGCAGTGACAC		VRVTRSYHSADCDTDHL			
				AACACGGGAAG		LVCSKVYIQPRRFFRSRK			
				CAGCGGTTACC		EGRPRIDTGKTRDPSKV			
				GGTATTAAGCC		CEFVQALEEALPRRDDT			
				ACTGCACTTGAT		DISKRWEQLRDAIYDTAI			
				TGGCGTAGAGT		SIFGKKTVKSADWFEAH			
				GAGGCGCCAGG		TEVLTPLIEKKRQALAA			
				GGTTGCTTCCGA		YRSSPSEENLQALRFARS			
				CGGTGGGAGAG		QVQQVARRCANDYWLQ			
				GTCTTCGGATCT		LCSKIQQAADTGNMRG			
				CATTGGGCAGC		MYDGIKQVIGPSQRKTA			
				TACCGCCCGCCT		PLKSATGEVLKDRTKQM			
				CAAGCTGGGCA		ERWVEYYSQLYARETV			
				GCCCCCAGCCA		VTERALNAIEPLPTMTEL			
				AATAGGTGCTG		DAEPTMEELSDALKALS			
				CCTCGCCACGG		SGKAPGKDGIPSEILKCA			
				TCCACTTGCTCC		NGTLVSELYGILCQCWR			
				ATTGGGTGTGT		EGSVPQDMRDANIVTVY			
				GGGGCTTAGGG		KNKGDKSDCNNYRGISL			
				AATCACCAACA		LSTVGKLFARVVLKRLH			
				GGTGGATCGTT		VLAERVYPESQCGFRAE			

				AACCTTACACC		RSTIDMVFSVRQLQEKC			
				AGGCAAAACAA		REQNKPLYIAFIDLTKAF			
				ACAAGCGACAA		DLVSRKGLFAILAKIGCP			
				AGTAATCCAGC		PTLLSIVRAFHDGMRGTI			
				CTTCAGGCTGG		VYDGSTSEPFEILSGVKQ			
				GCACATGGAAT		GCVLAPTLFGIFFAVLLK			
				GTAAGGACC		YAFGTATEGIFLRTRTDG			
						NLFKLSRLRAKTRVLTK			
						RLREFLFADDAAIAAHT			
						LQELQSLITRFADACMEF			
						GLTISLKKTQVMGQSVG			
						NEPFISVLDYELEVVHEF			
						VYLGSTISDNLSFDSEIN			
						KRIGKAATTFSRLTKRV			
						WANNKLTVHTKVQVYT			
						ACVLSTLLYASETWTLC			
						SKQEQRLNTFHMRCLRR			
						ILGISWQDKIPNSAVLER			
						AGTVSMFTLLKQRRMR			
						WLGHVSRMLDGRIPKDL			
						LFGELASGKRPKGRPKL			
						RYKDTCKRDLSALSISA			
						NTWHSLALDKYAWKQG			

						VKEALAMYETTLVKEA EDKRALRKIRARDTRAT SAYCCSQCGKDCHSRIG LHSHRRRCSNQSNRGAN P			
L2	L2-1_PSi	.	Pelodiscus sinensis	GGTTTAGACAG TTATGKGCTATT CCTKTTCCGGGT CCTGTCCCCAGC CTCAAWAACTC TGGGAACCTTA ACGCCTGKCCA TCAGCCCTGAG CGTGCTGTTGCT GAATGCCAGGT CAGTATCCCAG AAGACAGCTGT CATCCACGATTT AATTCTGAATG AAGGGGCTGAC CTGGCTTGTATC ACTGAGACCTG GCTGAGAGTGG	TGAGTCTGTGTGTG TGTGTTCCTCCTCTC GAGATGTTTTTAGC AGGTTTTATGTCCC TCTGGGTCTCTTTT GGGGTCTCGTATTT TAAATTTTTATATA TATTTGTATTGTGTT TTTAAAAATTTTGT AAGCCGCCTCGAAT ATTTTAAATAGAGA GGCGGGGTAAAAA TATTTTAAATAAA	MLLVYRPPRCTADALLE LADVVSDVALRTPRLLV LGDFNIHAEAVLSGPAR DFMDTMTTLGLSQEITA PTHVAGHTLDLVFWTK GDRVLGIRDVATLPLSW TDHYLVKFRLSADLPLH RGGGPIKMIRPQRLMDP EGFLRSLRDCSAIMANG PVDALVDTWNTEVARA VDMIAPKRPLRASRAKF APWFSGELMAAKRSVR RLERRWRRTRAESDRLR VRAELRAYSVAVLTAK KSYFSTRIASAQCRPAEL FRVVRGLLQSGTRDEME EPLAPRCDLFAKHFADKI ARIRLDLDSGCCVSGAQ	cd01650	379	640

				AGGGAGGCGTT		NVLSAPPSLFSLDQFQLL			
				GACCTTTCCCAA		QTEDVDKVLGGVRPTTC			
				TTGTGCCCGGCT		SLDPCPSWLIKSSREGMT			
				GGGTTCTCAGT		GWVQGIINASFRSGMVP			
				ACAGCATCAGC		AALKEAVIKPLLKKPSL			
				TGTGGCTTAATG		DPRDLNNYRPVANIPFL			
				GTCGGGGAGGT		GKVLERAVALQLQAFLE			
				GGGGTGGCAAT		ETDYLDPFQSGFRPGFGT			
				TGTCTATCGAG		ETALVALVDDLRREIDR			
				ATACCATCTCAC		GSASLLILLDLSAAFDTI			
				TGGCTAGGGTC		DHGILLERLAELGIGGTV			
				TTGGTCCAACA		LQWFRSYLSGRFQKVEL			
				AAGTGTCAGAT		GGCYSTPWRLCYGVPQ			
				TCGAGTGTCTGT		GSILSPMLFNIYMKPLGE			
				ACTTGACGTTG		VIRSFGGRCHQYADDTQ			
				GGTGGCCGGGA		LYFPVTSEPGVAATDLN			
				TAGGTTGGGT		RCLEAVMDWMKSNKLK			
						LNPDKTEVLWIRRPSVR			
						ECGVVPVLDGVTLPLKE			
						QVRSLGVLLDPFLTFEN			
						QISSVARSAFGQLRLIRQ			
						LRPFLNKHDLATVIHAL			
						VTSRLDYCNALYVGLPL			

						KSLRKLQLVQNAAARV LTNTRYTEHIMPVLRQL HWLPVCFRAQFKVLVM TYKALNGLGPGYLKDRL LPYEPARGLRSAGEALL RVPPLMEIRLTSTRSRAF SAFAPRLWNSLPRDIRM APTLALFRRQAKATLFT RAFIGV			
Penelope	Poseidon-2D_PM	.	Petromyzon marinus		TAGCTGCCTTATAA TTGTACAGCCATTT CCATTTCTTCCAGT GGGTTGTACTCGCT CTCGTGCAGTGCAC CACCGCTTCCCCCG AACCCCCTGCAGCA TAACAAAAGGCGG CTTAACTTAGCCCC CCCCCCCCACCCAC GTCCCTTCCTGGCT CCTCCCCCCCCCCC TTAACCTGTATATA ATCTCGGGTGTCTC	YLRDFTEKPKSMNPCSH ESFKSRINALYGEXFHLS VRAFNRQRDKKSQLIET LTFLKRCRDNKVIPIAFT LTSAAVSSPNSKRILFNA SLSLISERIRVLRFELDSL DKTLLQSHLMLTQKMSP SDWNHVDNMTAASADS LSAKCKANLIAKFNLLS KKKPEPHITKLDVNKIVI NQSNIPISDNLKTVLAKG LNFAVTPKHIPVEDIISQ VDVALKGIGHIQADVAR AEIATILKNAKKPISNISK	cd00304	464	548

					AGAGCTCTTCACTT GATACAGCTATTTT GCTGTGACTGTTCC ACCTGAAGACGGA AGCTTGTGTTGCCT CCGAAACGTCGTGA TTTAATTTAGTTGTT TTCATTTATTTTATT TTACCTACGTGACT TTACCGAAAAAAAC CAAAGAGTATGAAT CCTTGCAGTCACGA AAGCTTCAAATCTA GAA	NERSELANLKKRDDIVIL SADKGNATVIMDKNSY NNKIQDLLNDGSYSKIKS DPTNSLTIKTNKIITSSTS LIEPTKRPSLKPSAARPP RFYGLPKIHKEGTPLRPI VSTINSPTYKLARFLSDT LLPFTGKTTSAVLNSTQL VSMVKDIVLSPDDILVSF DVVSLFTKVPVPDTITIIE DLVQHGLNPDVPALVR HCLTNTYFTWNGTFYK QNEGTPMGSPLSPVIAN VFMEKFESDAIQTSTLRP SLWKRYVDDTFVIWPH GLPALNHFLDHINSRHPS IKFTMEVEKDGKLPFLD VLFERRTDGSLGHSVYR KSTHTDSYLKPDSHHHP AQKNSLIKTLYHRAHCIS DQHHLASELKHVHLAL KMNGYSDRTIRQAVSPT IQKQPPDTPIKIITLPYIAG			

						VTKKISKILNKHKDPGPV QYSTQNPTNTKIQVRFN TVHKIRDLIPSVKDVIPD LQYPGVYKLSCTCGDCY IGETKRHISDRIREHKSD LKHGRTNSSAVADHCFN TAGSHDIVFSDTKVLCK PPGYYQRFVHEAIQIYK HRHNFNREDGFKLSNH WKDVINQVKL			
Penelope	Poseidon-1F_PM	.	Petromyzon marinus		TAACCTGCTGTGAT GTTTCCATCCTTCCT GCGATTTTAGCTGA CTCTCTGAGTTGCA TTATTAGGTTGGAT GTCTTTGTATTGAC GGGCTGGATGTCCT TCTCGCTCAGTCAC CGATGCTGCGACAT ACTAGGCAAGTCAC CGATGCTGCGGCAT ACTAAGCTCGTGCA TTGACATGCTAATC	IHKEGTPLRPIVSTINSPT YKLTRFLSETLLPFTGKT TSAVLNSAHLVSLIKDIS LSPGDLLVSFDVVSLFTK VPVPDTIAIIEDLVQQGL NXDVPTLALPYQHLLHL EWYILQTRGGYPQWVLP YRQWCANFGTALPTLTS PGMVHSTNKRRVPPMG SPLSPVVANVFMEKFET DAIQTSTLRPSLWKRYV DDTFVIWPHGQPALNQF FNHINSIHPAIKFTMEVE	cd00304	156	239

					TCTCACAATACAAT AGCCACTTCCACAG CCTCAGGGAGACCA TCTCTAGCCCCGCC CTCTTACCCTGAGG CTGCACCCATCTCA GGCCTACATAAGAC CTTGTTTCAGTAGC TTTCTCTCTCTCACC TGAAGACGGTCGCT TGTGTTGCGATCGA AACGTCGTGATCTA TTAATTTTCTATCTA CGTGACTTTACCGA AAGAACCAAAGAG TATGGATTCCTGCA GTCACGAAAGCTTC AAATCAAGAT	KDGKLPFLDVLIEKRPD GSLGHSVYRKATHTDA YLKPDSHHHPAQKNSLI KTLHHRAHCISDQHHLA SELEHVHQALRMNGYS DRTIRKAVSPAMGRQPS EDPIKTITLPYIAGVTKNI SKILNKHKIQVRFNTVH KIRDLIPSVKDVIPDMQY PGVYKLSCGCGSSYIGET KRHVSDRVREHKADLK HGRTNTSAVADHCFNA VGTHEIDFSKTKVLCKPT GYHQRFVHEAIQIYKHR HNFNREDGYKLSNHWK DVINQVKS			
Penelope	Poseidon-1G_PM	.	Petromyzon marinus	AC	TAACTGGCTGTGAT GCTTCCATCCTACC TGCCGATGTTAGCA GTTATTTCCGACTC TCTGAGTTACATTA	QPPGYENQQYITDSTSLI EPTKRLSLKPSAARPPRF YGLPKIHKDGTPLRPIVS TIDSPTYKLARFLSETLL PFTGKTTSAVLNSAHLV	cd00304	155	246

					CCTTGTTAAATGTC	SLIKDISLSPGDLLVSFDV			
					TTTACATTAACAGG	VYLFTKVPVPDTIAIIEDL			
					CTAAATGTCTCACT	VQQGLNADVPALVRHC			
					AACATAGCTGCTGC	LTNTYFTWNGTFYKQVE			
					TCCCTTTTCCTCGAT	GTPMGSPLSPVVANVFM			
					GCGGCGGTGGCGCT	EKFETDAIQASILRPSLW			
					CGTTGTTTAATAGG	KRYVDDTFVIWPHGLPA			
					CTGAATGTCTCTCA	LNQFFNHINSIHPAIKFT			
					TTAACACAGCCGCT	MEVEKDGKLPFLDVLIE			
					GCTGCGGGAGCCGT	KRPDGSLGHCLQESHPH			
					CTCACATTGACATG	GCLSKTSKRDQMAHWD			
					CTAATCTCTAATCT	TVYRKATHTDAYLKPDS			
					CGAGCAATACAATG	HHHPAQKNSLIKTLHHR			
					GCTTCTTCAGCACC	ARCISDHHHLASELEHV			
					GTAGCCGTGGGGTT	HQALKMNGYSDCTIRK			
					ACATTAACAGATTG	AVTPTVGREPPEDPIKTI			
					ACATGCTAATCTCT	TLPYIAGVTKXISKIMSK			
					AGCAATACAATGGC	HKIQVRFNTVHKIRDLIP			
					CGAGGTATACCTAG	SVKDVIPDMQYPGVYKL			
					CTCTGCGATCTTAC	SCACGSSYIGETKRHVS			
					ATTTACTTGCAAAT	DRVREHKADLKHGRTN			
					TCCATCTCTAGCCC	TSAVADHCFNAVGTHEI			
					CGCCCACTTACCCT	DFSKTKVLCKPTGYQQR			

					TGGGTTGTACCCAT CTCAAACCTATATA ACCTCTGGCTGCAG TAGCTCTCTCGCTA TATTTCTTCCTGCCT CAGCAGTCCGCCTG CTGTATTTTTTCACC TGAAGACGATTGCT TGTGTTGCGATCGA AACGTCGTGATTTT ATTTTATTATTTTTC TATCTACGTGACTT TACCGAAAGACCCA AAGAGTATGAATTC CTGCAGTCACGAAA GCTTTAAGTCAA	FVHEAIQIYKHRHNFNR EDGYKLSNHWKDVINQ VKS			
Rex 1	Rex1-1_PM	.	Petromyzon marinus	G	TAATTATTAAGTAA TACTGTTTACAATT GACTGTAAAATAGG GATGTGCAATGACA AAATGTACCTTTAT GTGGGTATGTGGGT ATACTTGGGCAGTT	MAAHTVAAAQCLPGRC LVWLFMYALCLVLCTFC RASNTFSRHDLLKIGLCY ERSVTSEFLRLHKIPVEL ARSPGSPWIIIPAGRRRR RRRERRQKRGCRAGAL ARLRRQPLKPPLPSLFLT	cd01650	580	832

					TTTATATATAAAAT	NARSLANKMDELRLQTT			
					TTATTGATTATCTTT	ANNVVKDSCILLFTETW			
					TTTTATTTTATTTTT	LHSSIPDSAIELAGYTAQ			
					ATTTCTTGTTTTATT	RHDRTVDSGKSRGGGLC			
					TATTTAAATTTATG	VYVNNSWCTNTVTVDS			
					AATGATGCACTGAC	HCSPDLEYVTVKCRPIYL			
					TGGTGAGCACTTTA	PREFTVVMITAVYIPPDA			
					AATTTCGTTGTACT	NANSAIGLLHGSISIQQS			
					GGTGACAATGACA	KYPDAVQIIAGDFNHAD			
					ATAAAGGATCTATC	LKAALPKFHQHVKCATR			
					TATCTATCT	GTKTLDKVYTNIKLGYR			
						ARPLPHLGQSDHLSLLLI			
						PAYAPIKKTAPTITKTVA			
						TWPEGATQQLQDCFDRT			
						NWGVFEHQDLEVFTDS			
						VLCYIKNCIDTVTVDKRI			
						RVYPNQKPWMTREVQQ			
						LLKERNIAFKSRDKALY			
						STARANLKRGIREAKAD			
						YRRRIEERLDSNDSRQV			
						WQGVQHLTNYRANLGA			
						ADGDTALAEELNLFFAR			
						FEVTPPGTASPNPTVHSS			

						LNLTVEEHEVRRTLRAV NPRKAAGPDGVPGRVLR DCADELAGVFTRIFNLSL EQSTVPPCLKSSTIVPLP KKTHISSLNDYRPVALTP VVMKCFEKLVRSHIIATL PRSLDPHQFAYRANRST EDAVATALHAALSHLEQ QGSYVRMLFVDYSSAFN TILPHKLVVKLGDLGLP HSTCMWIHSFLSGRRQR VRVGRHTSTALSLSTGSP QGCVLSPLLYSLYTHDC APVHHSNTIVKFADDTT VVGLISGGDESAYRDEV ERLTAWCRGNNLLLNT AKTKELIIDYRRKKTDTP PLTINGDCVERVADFRFL GVHIGEGLTWSANTSEL LKKAQQRLYFLRILRKN NITQKLLVSFYRASIESIL TYCICVWYTSCTVAQRK ALQRVINTAQKIAGCPLP			

						TLEELHSSRCIKKAQNII KDSSHPGHSLFELLPSGR RFRSVKTRTNRFKHSFY PTAITTLNTAN			
Penelope	Poseidon-1D_PM	.	Petromyzon marinus	CCCAGGTCGAC GTAGCCCTSAG AGKCTTGGGAC ACATCCCGGCA GACGTTSCTAGA TCTGAAATCGCT GCTACCCTCAA AAACGCTAAGA AACCCATTAGC AACATCAGCAA GGACGAGCGTK TTGAACTCTCTC AATTAAAAAAC AGAGATGACAT AKTCATCCTCTC AGCCGAAAGGG CAATGCCACTG TCATCATGGAT AA	TAACTGACTGTGAT GTTTCCATCCTACC TGCCTATGTCAGCT GTTATTTCCGACTC TCTGTGTGTCTTTCT CTCTCTCTCTGTGC ATCTCTCTCCGATC TCTACTCCAGTGAC CCGGGTGCTGCCGC GGGGTCCACTTTCT GCCCCCCCCCCCCC CCCCCTTCTCTCTCC CCCCCTGCAGCCGA TCTCTCCAGTGCTG CCGCGGTGGTCCAC TTTACTCCAGTGAC CCGGGTGCTGCTGC GGGGTCCACTTTCT GCCCCGCCTCCTCC	TTYNNXIQXLJNDGXYT KINRDPTNRLATKXNNII TNSTSLIDSNKQRSLKPS AARPPRFYGXPKIHKEG TPLRPIVSTINSPTYKLXX FLPETLLPFTGKTTSAXL NSAHLVSLIKDISLSPDD XLVSFDVVSLFTKVPVP DTIAIIEDLVQQGLNPDV PTLVRXCLTNTYFTWNG TFYKQNEGTPMGSPLSP VVANVFMEKFETDAIQT STLRPSLWKRYVDDTFV IWPHGLPALNQFLNHINS IHPSIKFTMEVEKDGKLP FLDVLIEKRPDGSLGHSV YRKATHTDAYLKPDSH HHPAQKNSLIKTLHHRA RCISDQHHLASELEHVH	cd00304	186	270

					CCCCCCCTTCTCTC CTCCCCCCCTACTG CCGAGGCCTATATA ACCTCGTGCCACGG CAGATCTCTCGCCT ACTTCAGCAGTTTT TTCTTGCTGTGGTT GTCCCACCTGATGA CGATTGCTTGTGTT GCGATCGAAACGTC GTGATCTATTATTT TTCTACCTACGTGA CTTTACCGAAAGAA CCAAAGAGTATGG ATTCCTGCAGTCAC GAAAGCTTCAAATC AAGAT	QALRMNGYSDRTIRKAV SPAMGRQPSEDPTNTITL PYIAGVTKKISKILNKHK IQVRFNTVHKIRDLIPSV KDVIPDMQYPGVYKLSC GCGSSYIGETKRHVSDR VREHKADLKHGRTNTS AVADHCFNAVGSHEIDF SKTKVLCKPTGYYQRFV HEAIQIYKHRHNFNRED GYKLSNHWKDVINQVK S			
Vingi	Vingi-1_PMa	.	Petromyzon marinus	GGGGAAGAAG	TGACGATCATTAAG AAAAAATAAAAAC TGTTGCTTTTCACA ATGCCATACGAAAG AAGAA	MTTCSRSSLLRAKAQRM SSHRRPNARQRYPGLKS NSTESISQGQSNVNVGPS LRIMQLNIEGLSSAKCDY IAKILQFQDVDVLALEET HLDRQCPRSKIIGYTLIA	cd01650	508	762

						ASHHPKHGLATYVKND LKDTVQVVPHTSPHYIAI KINDVTVVNVYKPPGME WPCPALPQPSHPIIHIGDF NSHHTLWGYSANDRAG EQLTSWAAAEDQHLLY DAKDRGTFKSARWKKC YTPDLCFVSRDQNGNPM PSTRTILDDFPNSQHRPV VIHIGLQIPLFRCAPKPR WNLRSAEWVNFTADLE ESVGMIPPRAVNYRRFV GLIKAVAKANIARGFRK VFTPCWTPESEQLFRKF NADKKQETAKALLESLN LQRRERWRETVEALDFA KSSRKAWRLLRLMGAA ATPAYKPPQVSPDEIARQ LINNSRTSIDKSTRRQVY KELYCTLDNCQETSEFS RPFSVEEVDAGIAQLKC GKASGLDDIYPEFIKNLG PRARKWLATFFTDIHSTS			

						IVPKEWKEARIIAILKPG KDASDAQNYRPISLLSV CYKLMERLLLKRLMPTL EEIVPQEQAGFRTGRNC CDQVLALTSHIEAGFER QLKTATAFIDLSSAYDT VWRKGLLLKLAKAVPC RPTIRLITEMLSNRWFTV HLGEAASSPRRLNNGLP QGSVLAPALFNLYTSDIP ATSCRKFTYADDIALAT QARKMEDTEAVLEKDL KTLEEYFKNWRLKPNPS KSVVSAFHLNNRQASKT LNVNFCGAKLRHDEHPR YLGVTLDRTLNYHHHL KQAKGKTKARVNIIQKL AGTTWGSDASTLKISTQ ALVHSTAEYCAPVWGA SPHTNLVDAEINNALRII SGSLKSTPLPWLPVLANI APANIRRDDATTREYNK IMKNFTLPLHQDLKEPPR			

						PRLKSRKPFWIRASALA AENKDTATRWSESWAA MDVANKNIIHDPTVEVP GFSLPRRLWTTLNRART QHGRCAYTFHKWKIKES PNCDCGEVQTIKHITDEC SIHAHPGGIAEIHTASTTS LDWMRNLCIHL			
Penelope	Poseidon-2_PM	.	Petromyzon marinus	CA	TAAACTAATTTTCA ATGGTACAGTATAT TTTTTCAAGTAGTG GTATCCCAAACTTT CACAGCAGCTTATT TCCCTTCACACCAA CATTATGTAAACTC CCCCCCCCCCCCCC CCGCGCCGGTCCAT CATGTACCTCCTAC GTACATCATGCTCA TTTTTATGCTAATTC CTTTGTCAAGGCGT TTTTGTTTCCTCCCA CACAGCTTAAAGGA	SLSHTSAAVSSPNSKQIL FKASLSLISERIRLLRFEL DTLDRTLLQAHLTLTQK MSPSDWNHVDNMTASS ADSLSVKCKANLISKFN HLSKKNPVPHITKLDVN KIVINQSSTPISDNLKTVL ARGLNFAVTPKHIPVDDI ISQVDVALKGIGHIQAD VARAEIATILKNAKKPIS NISKNERSELANLKKQK KNRSDIVILSADKGNAT VIMDKDSYNNKIQNLLN DGSYSKIKNDPTNSLTIK TNKIITSSTSLIEPTKRPSL	cd00304	401	485

					CGCGACCTGGGCTT	KPSAARPPRFYGLPKIHK			
					GGCTGTCTTCACTC	EGTPLRPIVSTINSPTYKL			
					CAACGCTGTCCTCT	ARFLSDTLLPFTGKTSSA			
					GGACAGTCCGGTCC	VLNSAQLVSMVKDVVL			
					CTTCACCTGATGAA	SPDDILVSFDVVSLFTKV			
					GGCCACTTGTGTTG	PVPDTIEIIENLVQHGLDP			
					TGGCCGAAACGTCG	DTPALVRHCLTNTYFTW			
					TGATTTTTGTTGTTT	NGTFYKQNEGTPMGSPL			
					TATTTTTTACCTAC	SPVIANVFMEKFESDAIQ			
					GTGACTTTACCGAA	TSTLRPSLWKRYVDDTF			
					TAGAACCAAAGAG	VIWPHGLPALNLFLHHIN			
					TATGGATCCTTGCA	SRHPAIKFTMEVEKDGK			
					GTCACGAAAGCTTC	LPFLDVLFERRPDGSLGH			
					AAATCTAGAA	SVYRKPTHTDSYLKPDS			
						HHHPAQNNSLIKTLYHR			
						AHCISDQHHLTSELKHV			
						HLALKRNGYSDRTIRQA			
						VSPSIQKQPLDTPVKTIT			
						LPYIAGVTKKISKILNKH			
						KILVRFNTVHKIRDLIPS			
						VKDVIPDLQYPGVYKLS			
						CTCGDCYIGETKRHVSD			
						RIREHKSDLKHGRTNSS			

						AVADHCYNAVGDHDIV FSDTKVLCKPTGFYQRF VHEAIQIYKHRHNFNRE DGFKLSNHWKDVLNQT KH			
Penelope	Poseidon-1J_PM	.	Petromyzon marinus	A	TGACTTGATTTGTG TTTTCTTGTCAACCT CCACCGTCAGCAGT GTGGGTGTGTGGAG CGAGGTGCAAATG ATTGCCTGGCGGCC CTTTGATTGGTGTA GACCCACTGCGTTG GCGTTGTTTTCTCCT ACATTAACATTAGC TGCCCCGCCCATGC TGCTGACACACTTG GGCTACGCCCATCT CTAAACCTTTATAA ACAGTTTGCCTTTA GAGGCTAACCCCCC TTTTCCTGATGACG ACCGCTTGTGTTGC	KSKFHRLYGEGFHITIRS FNKLRDKKAHLIESLTFL KRCHNSKVIPISFTLTSA AVTSSNAKRILFRASLAL ISERIRLLRTELDHVDKT LLKSHLTLTEKMSTSDW EHVDNMAASSAATMSA KCKDNLISEFNLLSKKKP DPHIKKDVDKVVINLSN SPISDGLKSVLAKGLNFA VTPKHIPVDDIIAQVDVA LKGLGHIAADVTRSEIA ATLKSAKIPVSNISKDER FELSQLKKRDDIVILSAD KGNATVIMDKTVYNDKI QDLLNDGSYTKIKRDPT NSLVTKTNSIITDSASLIE QTKRFSLKPSAARPPRLY	cd00304	443	527

					GATCGAAACGTCGT	GLPKIHKEGTPLRPIVSTI			
					GGTTTTTGTTTTTAT	GSPTYKLARYLSETLQPF			
					TATATTTAATTACT	TGKTTSSVLNSAHLVSSI			
					GAAAGAACCAAGA	KDISLSPGDLLVSFDVVS			
					ATAAGTCTACATTA	LFTKVPVPDTIAIIEDLVQ			
					TAGACTGCCAGTAC	QGLNADVPALVRHCLT			
					GCTCACTCAGTTTA	NTYFTWNGTFYKQTEGT			
					ATAACCATTGCTAC	PMGSPLSPVVANIFMEN			
					AGTCTACATCATAG	FETDAIQASTLRPSLWKR			
					ACTGCCAGTACACT	YVDDTFVIWPHGRPALN			
					CACCAGCTTAAAAA	QFLNHINSIHPAIKFTME			
					CCAGGCTACAGTCT	VEKDGMLPFLDVLIKKR			
					ACATTATAGACTGC	PDGSMSHSVYRKVTHTD			
					CAGTACACTCACCA	AYLKPDSHHHPSQKNSL			
					AGGTTAATAACCAG	IKTLHHRAQCISDHEHLT			
					GCTACAGTCTACAT	SELKHVHQALKANGYS			
					TATAGACTGCCAGT	DYTIRKAITPTVRKEPSE			
					ACGCTCACCGAGGT	DPIKTITLPYIAGVTKNIS			
					TAAGAACCAGGCTA	KIMNKHKIQVRFNTVHK			
					CAGTCTACATTATA	IRDLIPSVKDKIPDMQYP			
					GACTGCCAGTACAC	GVYKLSCICGSSYIGETK			
					TCACCCAGCTTAAT	RHVSDRVREHKADLKH			
					AACCATTGCTACAG	GRTNTSAVADHYFNAA			

					TCTACATTATAGAC TGCCAGTACACTCA CCAAACTTAATAAC CAGGCTACAGTCTA CATTATAGACTGCC AGTACACTCACTCC GATTAAGAACCAG GCTACAGTCTACAT TAGTGGTTTTTTTGT TTTTATTTTTATTTT TATTCACTTTATTTT TTATTATTATTTTTC TATCTACGTGACTT TACCGAAAGAACC AAAGA	GSHEIDFSKTMVLCKPT GYQQRFVHEAIHIYRHN FNREDGYKLSNHWKDVI NHFGS			
Penelope	Coprina_Pc1	.	Phanerochaete chrysosporium	ATGATCTGTGC ATACAAAACCT TCGTACTTTTGC CGAGAAAGAGT CCCTTAAGGGG ATTCGTTGCTCC CGACCTCGGGG AGACTTTTCTCC	TAATGCGATTCGAT AACTGCTGGCCGCA ATGGTCAGAGGATA TTTACTTAGCGGCA TACACTTTGGTTCA AACCCAAAACCCCT AACCTAAGCCCTAA GCCTAAATATTCAT	MPAQDTTRRAVRATASS STGHFPSFTRALSPSQAS VVLGEQISFGTFDERCA AALAHCTLLDEVIATLP KLYQVALSPVLTDIHHD ATKRVEIAAQLERLRQH KAAGTLPSPLKGLGSTP AIQFMKGFLESQKPDFP	IPR000477	664	917

				ACACCAATCGC	TAGCGTCTCCCTCA	GKAPDIVKACAAKLLDG			
				ATTACAGCCGA	CAGGCCTAAACCCC	FIETKAAEHAWYESRLQ			
				CGAGCGACATT	CTGGAGCGGAGTTG	AMEVTRRVLTAADTHW			
				TCGCCTCTCGTC	TACCCGTTGAACGG	KDELEQRYQVLEGFVRP			
				GGCCCGCCAGT	TACAGCCAGCAGTT	PVEGQVPATVPSAEELH			
				CCCTTCGGGCCT	ATAGAGTATTTTGT	PLKFVTDPVMVAQHENI			
				GGCGGTTGCAA	TGTACAGCGTCATT	RRSIPRLASTVIMLVVNR			
				GGGGTATATAT	ATTTAATGGATGCA	ALTQQQRNKAKRDLKD			
				CCCAGTAGGAA	GTGAGCGCCCGCCT	QADVEMATDDTAVDSA			
				ACCGCTGATGA	TCGTGGCGTGGGCG	VKRLEKQVEGLVSGMK			
				GTCTTCACGAC	CTTCGGTGATCGCC	ALTNGKGKVSFSSSSRT			
				GAAAGCTCAGC	GTGATGGCCGCGCA	HVEDAHPFVSGQSPSSR			
				GCGACACAACA	CGTGTCCCACGTGT	PFSLVDHRARHEAQGVQ			
				GTGTCGGCTGC	TCGTCTCGTGGCTG	LGKRLRQRQDDGHQGP			
				CTTCGGTTTTCC	CTAATAGCGTCTGT	SRRLRCSSGEAGSPQEG			
				TGTGAAGCATT	GATCATGTGATCTT	QGEGTREGQEDPEEIVA			
				CCGCGCTCGCC	CGACCCAGGGCTGC	YVRAAPWRYDAPHSYP			
				GCTCACTACGCT	GGTTAACTCTGGAT	DVILMLPAPRAIQLLLER			
				CGCGTCTCGCG	TCGATGCGCATGTT	VPLEVKRAARFRSRVFL			
				CCCCCCAAACC	AATAGCCGGCCCGA	GPHVTVPEHLQMHVSSS			
				CTAATCAAATC	ATCCGGGTCGGCTC	LRYLMYSVPSSNVVREA			
				GAGTAAGCACT	CGAACTGACTCCGC	WEDFKDRFRWRVFFTE			
				CTTCACC	TAGGTTCGGGAGAC	KLIDSPNDDDPDFDPDY			

					TTTCGCCGGGGTGG	VVPKARKRFDDGVRTTP			
					CTGCGTGCCGGACT	AYIELGLRKGDKFVDDY			
					TCGTCCGGGCACTT	ISNVMPSVKATSRSLNL			
					TGTGGCGTTGATCC	VWLDELKDYLSANKYIV			
					GATCAGATCGAGAT	SPSDKNLGVCILQADWA			
					CTCGAGACTTCGAA	TDQTRTLWNDPSNYRKL			
					AGGTGACTTTTCGG	GSLLEADEIMEEKHERIF			
					TAGTCCGCTGCCCG	RAARLARRLGKRQLAEF			
					CTGGGGTGTTGTTG	LVSKIHGKDAKTAGYTL			
					GCCCCTTCCGCATG	GQFYGVPKVHKRPVKM			
					ACCTGTACATTCAA	RPIVPCHSVSQAPSATYV			
					AACCTTCATACTTT	SKMLKPLVSRQPHVIHG			
					CGCCGAGAAAGAG	SKHLVRQLESLQLDKHR			
					TCCCTTAAGGGGAT	KAWIVSGDIVAFYPNIPL			
					TTGTTGCTCCCGAC	ENCIRVVTQMWRNSPEA			
					CTCGGGGAGACGTT	EKLSPEERQLFLMSFMT			
					TCTCCGCGCCAATC	ANKRLVIQFGDEYAEQI			
					GCATTACAGCCGAT	RGLAMGIACSPDLANLY			
					GAGAGACATTTCGC	GSHYENPILARPDMKSK			
					CTCTCGTCGGCCCG	FAFFGRYLDDVLGIVYA			
					CCAGTCCCTTCGGG	SSAEEAMSYANEIRYDG			
					CCTGGCGGTTGCAA	VEIEWSVSEYHTPFLDLF			
					GGGGTATATATCCC	LYLDPVEKRLHWKPYSK			

					AGTAGGAAACCGCT	PLNHRERIPWASHHPKD			
					GATGAGTCTTCACG	VKKGTFIGEMSRLATLC			
					ACGAAAGCTCAGC	SKLEHYRDAISDLSNLYL			
					GCGACACAACAGT	ARGYPDDAVRSWVKAH			
					GTCGGCTGCCTTCG	FSQRWERRLGEPIVRDE			
					GTTTTCCTGTGAAG	VFVLKSHFNPAWSAFNV			
					CATTCCGCGCTCGC	HGLSEVVVESWLSSLSK			
					CGCTCGCTACGCTC	LDGANQSQDRTWHAAE			
					GCGTCTCGCGCCCC	STDGSELKIFRQTTLGQF			
					CCAAACCCTAAACC	QGEQSSLYPGESRPEGV			
					CTAAACCCTAAACC	AGRVGVANVAKPRVTY			
					CTAATACCC	DPRDVMDVDELVALSEP			
						SFSAHDLSVVWTDNALR			
						VPKSLRIRYAMAELRLD			
						ADPQVNPALHVPLGVED			
						QVGDLGPASDAESVASG			
						DDGTTNSKPYRICTAKK			
						DSGIVQEDLDIRSVGFHD			
						RAWLVSRKKIRNLGDIV			
						NSWKKDIIQAALNSDSV			
						PHMHVDEWQ			

				AGCTCGGCCTC	TAAGCTGGTCATCA	MLADPAALAAGLARAP			
				GCGGCTGCCTTC	TGACCGACAGGGC	PPPSAPQDPSPAFPAGPA			
				CCAGGCGCCGC	ACTACCCAGGTAGG	GQNPRAAAPARVEVHT			
				CGACTTCGCGCT	GAACCGCCCTTAAA	VVAPPGRAGGMLPDPGL			
				CTGGCGCGGCC	AAACCCAGGAAGA	VEEPIQATYAHDAAQFE			
				CACACGCCGCC	CACAAACACCCTCC	CALCPYVAESMAVLVQ			
				GCCGAGCCTCC	ACTTAGTGACATAC	HRRSAHRGTRFKDIFTSG			
				AAGCGCGCCCG	ATATTTTAGCCTAG	CQCSLVFYARIVAASHA			
				TTGGCTTTCGCA	ATTTCAGTTACGGA	VACARRNQRAVPPAPTP			
				GACGCAGGGCT	GAGGTTACTAACTG	VAPTRPEATPQPTGYLA			
NeSL	Utopia-1_PI	.	Phytophthora infestans	CGCGGCGACGG	GTAAATACGAACAC	AAMTAAAAAASSDTVV	cd01650	919	1187
				CCCTCGCCAGC	ATATTCTGTTCTAA	AAATNMQSAVPAAAKT			
				CCCCAAGACCC	TCAGTGTGAAAACT	TGLQLVPPELEPALPQRA			
				CCCCTACGATGT	GGTTTTCGCCTTTT	SCHAGKRRRLNADEAV			
				GGCACCACCCG	GGCGGACTTTTTCA	TPCTPTARVSPQTEVAM			
				GCAGGGCGGCC	CTCGCATTTTTGGG	APHDAPQDDTVLQREA			
				GGCAGGCTGCC	CAATCGTCTGCGGC	AEPQPDPAATPGAQVQR			
				CGACTCGGTAT	TAGCTTGCTAGCGG	VEDTTAAQDDTVQQDH			
				CGCCGGGTGCT	CGGACGAGCGGTCT	DADTAQVSPPRRTPTRW			
				ACACTCTCAGC	CCGGGGGCGTTCAC	GPRPSSTQEPSPMTGEPA			
				CGCCACAGCTC	CTTTCCCCCGCGAG	ATLAARRPLTPAATGTR			
				GGGCCTTGGCG	GCCAACTACACCGA	ATRWGPCHRAIGAAAIA			
				GTCCGCCATTG	TCTTCTCTACACTTT	RLVTGLSTEPAQPQRRQ			

				GCCCTTGGAGC	TCTAATTCGCCTCC	PPPPQEPPSQPEPLAAAA			
				TCGACAGCGAC	GTCTTCGGTCTTCG	TAAADIAATVAADIAAA			
				AGCAGCGACGA	GTTGTCGGGCTTTT	AANAAMDVDGGPAADE			
				CGAGGACGCTC	TTCTTTTTGACCAA	TWLLRFDGACRRNPGPG			
				AAGACCCCCAC	TCAGAGCGCGCCAT	GAGAALFAPSGAVVWT			
				GCCGCCGCCCC	GCGCCTCTTCTGGC	CSHYMPSRSETNNTAEY			
				AGAACCCCCAG	CAATCAGAGACCG	TALLLGVQSAVHHGASH			
				AAGACGTCGCG	GGCCCTGTCCTCGG	LEVEGDSSLVIAQVKGT			
				AGTGTGCTTGCC	ACAGCGAGGCCTCC	FACRNARLRQLRNRVRH			
				CCACCCGGCAG	ACGGCCAGCCAATC	ALRSVDTHKLRHIDRQA			
				GGCAGGCAGC	GAGTCTCGGCAGCG	NAHADRLANRALDQRR			
					ACGCGTCTTTCTAT	TSSECGTHGSCMDSCLA			
					AGCGCAGCTGACG	VPTALAAQETPPAAPPST			
					AGGCCGATCTGGCG	SATPAEGNAMDDIAAEI			
					GCCCCCGATTGGTC	AARDEGETFPVLPIGPGS			
					CGACTTTCGGCCAA	APERQPRLRLRQLSDEE			
					TCAGCGACGACGA	RDAAADALQELADTMA			
					GGGGGCAGGGGTTT	SKIEDADSWTSGEGYISS			
					ACACTTTTGCCCCC	IPERIREVLQPYATAPPQ			
					GTTTCGACTTCAAC	PGRSRRQQRRRPPRVTR			
					TTCAGGCCAAAATG	NQREHRLDEALDDMAA			
					GCGATTTGGACCCT	TQQATPRDQRAVRRAR			
					CCACGCGCCGTGCC	RRVGRVRASMAQQELR			

					ACTGCTCGGCACCG	HEFAKDESKCVAKILKT			
					GCGGCGATTCAGCG	ASTETAAEDEHPETCPID			
					GGTGCAACTTCGGG	AATLHAHFTGVNAPRTD			
					CACGTGTGCAACAC	FDYDATSGREFRAAMSD			
					ATGCAGCGCCCATT	LPPATVEIDAFDAELTID			
					GCACGCCAAGCGG	EVEDQLTRAAKTSSPGH			
					CATCGCGGGACGAC	DGIDYGIYSRFAAQLVPL			
					GCCTCGGCCGCCCA	LHAVFQFCWRHRRVPRL			
					AGCGCAGCCCCGCC	WKVGIVRLIHKKGDPRQ			
					CTTCCAGCACGACC	PTNWRPICLQPTIYKLYS			
					TCGCGCCGTTTGGC	GLLAHRLSRWLEGNDRL			
					GGATCGCCATCAAG	PMAQKGFRAFNGCHEH			
					ACGTGCGAGAGCC	NFMATTLLDQTRRQHR			
					AGGCGGGGTCGGG	KLYQVWYDLRNAFGSL			
					CAAAATATACTTAC	PQQLMWRVLRHLGVDS			
					TCTAAGTATGCCCG	GFIDRCRDIYRDSAFVVA			
					AATCCCTGCCCTCT	NAADGATDPVRQEVGV			
					CAGGCTGAACGCG	YQGCPLSPLLFVAALVP			
					GCCCCATACTTGAT	LVRRLEKLDGVGVPLAD			
					CTAAGTATGGGAGG	GVRPCTTAYADDLKVFS			
					ATCCCTGGCCTCTC	DSAAGIRKCHDTVAGFL			
					AGGCTGTACGCGAG	AWTGLRANPGKCASLA			
					ACCC	VTTNARGNPTRDSSMRL			

						EVHDAAITTLSLHESYR YLGVGDGYDHVRHRLQ LEPKLKQLKREAVALLT SGLAPWQVVRALKVYV YPKVEYALRHLRPLQSQ LQAFDRVVSKGLRHLLS LPRSATSEVLYAPTSSGG LGLQPLVELHRALQLAH AWQMLHSKDPAIQAVA RAQACQVVRKRYRLQE DHWRGRDDELVRSFLNS ELAASPHAEVLRRNGDI ASLWSDVQRWLRIYHLR FEHCDETEAHGPLSFRVP HHNKWLTHKTVLRHVK LHLKIRHQTRWKGMVD QGKTVRVHGGVGAKFM TTGAGLSDDDYRFGVK GRLNQVDTNSVLKRKRL RAHTTCRDPTCSSAETL AHVLNHCESNMDAIRQR HDDALEQIGSKIRGALD RAKSPTELRLNQTVPEY			

						TGAALRPDIVLRNVAAK TMVIADLAVTFEDQAAR ARHSSLQLSHDHKTLKY QPIVAELQHKGWRVQT AAIVYGTLGSVQPSNFK AYTEKFKLHKREARQLD LQLSSHCIRASHRIWGW HCRQHRDRQRSGTASRA SRGSGGAPRRTSQAPAR R			
NeSL	Utopia-4_PI	AATU01001281.1	Phytophthora infestans	AGCTCGGCCTC GCGGCTGCCTTC CCAGGCGCCGC CGACTTCGCGCT CTGGCGCGGCC CACACGCCGCC GCCGAGCCTCC AAGCGCGCCCG TTGGCTTTCGCA GACGCAGGGCT CGCGGCGACGG CCCTCGCCAGC CCCCAAGACCC	TAAACGGGTCACTT GACCGACAGGGCA CCACCCAGGTAGGG AACCGCCCTTTAAA ACCCAGGAAGACA CAAACACCCTCCAC ATAGTGACATACAT ATTTTAGCCTAGAT TTCAGTTACGGAGA GGTTACTAACTGGT ACATAAAATTACAC ATTCTGTTCTAATC AGTGTGAAAACTGG	MLADPAALAAGLARAP PPPSAPQDPSPAFPAGPA GQNPRAAAPARVEVHT VVAPPGRAGGMLPDPGL VEEPIQATYAHDAAQFE CALCPYVAESMAVLVQ HRRSAHRGTRFKDIFTSG CQCSLVFYARIVAASHA VACARRNQRAVPPAPTP VAPTRPEATPQPTGYLA AAMTAAAAAASSDTVV AAATNMQSAVPAAAKT TGLQLVPPELEPALPQRA	cd01650	934	1198

				CCCCTACGATGT	TTTTCGCCTTTTGGC	SCHAGKRRRLNADEAV			
				GGCACCACCCG	GGACTTTTTCACTC	TPCTPTARVSPQTEVAM			
				GCAGGGCGGCC	GCATTTTTGGGCAA	APHDAPQDDTVLQREA			
				GGCAGGCTGCC	TCGTCTGCGGCTAG	AEPQPDPAATQGAQVQR			
				CGACTCGGTAT	CTTGCTAGCGGCGG	VEDTTAAQDDTVQQDH			
				CGCCGGGTGCT	ACGAGCGGTCTCCG	DADTAQVSPPRRTPTRW			
				ACACTCTCAGC	GGGGCGTACACCTT	GPRPSSTQEPSPMTGEPA			
				CGCTACAGCTC	TCCCCCGCGAGGCC	ATLAARRPLTPAATGTR			
				GGGCCTTGGCG	AACTACACCGATCT	ATRWGPCHRAIGAAAIA			
				GTCCGCTATTGG	TCTCTACACTTTTCT	RLVTGLPTEPAQPQRRQ			
				CCCTTGGAGCTC	AATTCGCCTTCGTC	PPPPQEPPLQPEPQAAAA			
				GACGGCGACAG	TTCGGTCTTCGGCT	TVAADIAATVAADIAAA			
				CAGCGACGACG	GTCGGATTTTTTTCT	AANAAMDVDGGPAADE			
				AGGACGCTCAA	TTTTGACCAATCAG	TWLLRFDGACRRNPGPG			
				GACCCCCACGC	AGTGCGCCATGCGA	GAGAALFAPSGAVVWT			
				CGCCGCCCCAG	CTCTTCTGGCCAAT	CSHFMPSRSETNNTAEY			
				AACCCCCAGAA	CAGAGACCGGGCC	TALLLGAQSAVHHGAK			
				GACGTCGCGAG	CTGTCCTCGGACAG	RLNIEGDSHLILSQVRGA			
				TGTGCTTGCCCC	CGAGGCCTCCACGG	FACNNKRLRSLRNRVQA			
				ACCCGGCAGGG	CCAGCCAATCAAGT	SLRQLDWYRLQHIDRKA			
				CAGGCAGC	CTCGGCAGCGACGC	NQHADRLANRALDLRR			
					GTCTTTCTATAGCG	TVTECGPHAETRNRCFQ			
					CAGCTGACGAGGCC	TPQPLVEPGETHCVPGS			

					GATCTGGCGGCCCC	DEVLAANTAMEDATAV			
					CGATTGGTCCGACT	PTEDDEAEVAARDGGEV			
					TTCGGCCAATCAGC	FPTIAIGPDSAPAKQPRL			
					GACGACGAGGGGG	RLKKLDEDDFDAAAAA			
					CAGGAGTTTACACT	VTRVSEELASKIVDAGD			
					TTTGCCCCCGTTTC	WTSGEGYISAIPERLRAA			
					GACATCAACTTCAG	LRPFALPTQPARPQPREP			
					GCCAAAATGGCGAT	RMQQPPRRPPRVTRDHL			
					TTCGACCCTCCACG	EHRLDEALDTMENVQRS			
					CGCCGTGCCACTGC	TPQNQKAVRRARRRVG			
					TCGGCACCGGCGGC	RLRSAMDRTRLRKKFAT			
					GATTCAGCGGGTGC	HERECVAEILRRASTEEA			
					AACTTCGGGCACGT	ANPSQEKCPIDRATLHE			
					GTGCAACACATGCA	YFTATSTQRTPFDYDSA			
					GCGCCCATTGCACG	KGTEFRTFLEVMSTPSHE			
					CCAAGCGGCATCGC	TSALTAEPTLDEIEDQLA			
					GGGACGACGCCTCG	HVKAGSSPGHDGVGYD			
					GCCGCTCAAGCGCA	VYRRFQVQLLPLLHAAF			
					GCCCCGCCCTTCCA	RFCWRHRRVPALWKVG			
					GCACGACCTCGCGC	FVRLLHKKGDPQQPNN			
					CGTTTGGCGGATCG	WRPICLQTAIYKLYSGLL			
					CCATCAAGACGTGC	ARRLSKFLEANELLPMA			
					GAGAGCCAGGCGG	QKGFRAFNGCHEHNFV			

					GGTCGGGCAAAAT ATACTTACTCTAAG TATGCCCGAATCCC TGCCCTCTCAGGCT GTACGCGGCCCCAT ACTTGATCTAAGTA TGGGAGGATCCCTG GCCTCTCAGGCTGT ACGCGAGACCCGTA CGGCCGAATCCCTG GCTTCTCAGCCTGT ACGCGGGGCTATAC TTGGTCTAAGTATG CCCCGGTCGCTGGC CTCTGAGCCCGTAC ACA	ATTLLDQTRRMHRRLY QVWYDLRNAFGSLPQQ LMWGVLRQLGVTEEFV ARCSGIYEDSYFVVGNA SDGATEPVRQEVGVYQ GCPLSPLLFITALVPLLR ALENQDGVGVPLADGV RPCATAYADDIKVFCDS ATGIQRCHALVTRFLEW TGLQANPAKCAFLPVTR SQHSNPTRDRDIELRIHG EAIATLGLQESYRYLGV GDGFDHVRHRLQLEPKL KQIKREAVALLHSELVP WQILKALKVYIYPKVEY ALRHLRPLKSQLQGFDS AIVRGLRHLLRLPENSH DGLFFSPTSAGGLGLLSL VELHEALQVAHAWQML HSKDPAIRAIARTQVGQ VARKRFKLVEEHWRGR EDDLAQRFLNTELAASP HATETRRNGDIGSLWND			

						VRDTLQTLGLKFAAGDE EEAPGLLQLRVPHHTKW LSHSTVLRHVKLHMKLR RMDTWKSKVSQGTTVR EHGGVGSRFITAGAGLS DAEYRFAIAPRAHLIDTN STLKRRRLRANDTCRAP GCSYTEPPAHILNKCSPN MDAIRKRHDDALERIAD ALRRKVEKSGGRLEVAI NKTVPEYDGAALRPDIV LRNTETKRAIIADLAITH ENQPTDATTSSALQQSR DNKITKYQTVAAAMMR AGWRVRVTGIVYGSLGS VLPSNFKVYTELLALLK RDARRLNRQLSSHCIRA SARIWSAHCRRHRERQR SGNASRASRGSGGAPRR TSQASARR			
NeSL	Utopia-2_PI	.	Phytophthora infestans	GCTCGGCCTCG CGGCTGCCTTCC CAGGCGCCGCC	TAGGCGGAAACCA GGCCCAAGACGGC CGACAGGGCCCCAC	MLADPAALAAGLARAP PPPSAPQDPSPAFPAGPA GQNPRAAAPARVEVHT	cd01650	1026	1289

				GACTTCGCGCTC	CCAGGTAGGGAAC	VVAPPGRAGGMLPDPGL			
				TGGCGCGGCCC	CGCCCTAGAAACCC	VDSSPAAATAATPAPVA			
				ACACGCCGCCG	ATTTCGGTGGTCGA	ATATTARAAARVAVEH			
				CCGAGCCTCCA	CTCGAAGGCCTTAC	HAHAEPNQEHLPMARV			
				AGCGCGCCCGT	CTATTTTTCCTTAG	LVEPMQVDECSSCDRST			
				TGGCTTCCGCA	ACATTCAATTAGGT	LTADDGSGDDVAAPSS			
				GACGCAGGGCT	AGCGACCAAATTAC	MLSNDVAAPMDVDSGT			
				CGCGGCGACGG	AAATTTGGTAACGA	SCPPTLQQPLQRPRALH			
				CCCTCGCCAGC	GTAAGCCAAATGGT	VGSKRRRLDADDGEEA			
				CCCCAAGACCC	AATACACAAAACTT	HQLQEEEEAGIHAPALR			
				CCCCTACGATGT	TTCTGTTCTAATCA	LSAASAQPASVLAVYTH			
				GGCACCACCCG	GTGTGAAAACTGGT	NASRFDCTLCAYTAGSF			
				GCAGGGCGGCC	TTTCGCCTTTTGGC	ASLLTHRNSRHRRTAFL			
				GGCAGGCTGCC	GGACTTTTTCACTC	DRFSAGCACGVPFASRL			
				CGACTCGGTAT	GCATTTTTGGGCAA	AAARHAQACASLSSAPS			
				CGCCGGGTGCT	TCGTCTGCGGCTAG	AEASSAAGTSSPTADGA			
				ACACTCTCAGC	CTTGCTAGCGGCGG	DSTVSAVAHAEPGLPHH			
				CGCCACAGCTC	ACGAGCGGTCTCCG	NDTELTASPPLVSSSDVE			
				GGGCCTTGGCG	GGGGCGTTCACCTT	VQATKTEATDNRWGAP			
				GTCCGCCATTG	TCCCCCGCGAGGCC	LPRVLVASRIAGRLAQV			
				GCCCTTGGAGC	AACTACACCGATCT	PPPRWGPPLPRTTIAARI			
				TCGACAGCGAC	TCTCTACACTTTTCT	ATRLAATPAPRWDPPLP			
				AGCAGCGACGA	AATTCGCCTCCGTC	RSLVVSRIAARLLPALPD			

				CGAGGACGCTC	TTCGGTCTTCGGCT	APACEEEAKDSDTMDW			
				AAGACCCCCAC	GTCGGATTTTTTTCT	APTWTNEETKESEPHDE			
				GCCGCCGCCCC	TTTTGACCAATCAG	APGQVDEETIDDADGEW			
				AGGACCCCCCG	AGCGCGCCATGCGA	LLRFDGACRANPGPGGA			
				CAGACGTCGCG	CTCTTCTGGCCAAT	GAALFKPSGPVVWTCSH			
				AGTGTGCTTGCC	CAGAGACCGGGCC	YDPSTTATNNTAEYTAL			
				CCACCCGGCAA	CTGTCCTCGGACAG	LLGARAAADHGVTKLRI			
				GGCAGGCAGC	CGAGGCCTCCACGG	EGDSTLVIQQVRGIFATR			
					CCAGCCAATCAAGT	STRLRALRNKVKLELAR			
					CTCGGCAGCGACGC	VGSFSLHHIDRQANGHA			
					GTCTTTCTATAGCG	DRLANAGLDRRRTKLEC			
					CAGCTGACGAGGCC	SVHPDGRGCTNTSVATA			
					GATCTGGCGGCCCC	APTAPAAPLPSARPPAST			
					CGATTGGTCCGACT	AAPSPDDDHSDQGDIDD			
					TTCGGCCAATCAGC	GEVYAAMCISPDAVPHR			
					GACGACGAGGGGG	RPRLRLRRLTDEESEEAG			
					CAGGGGTTTACACT	NVVERLAASLAAKIADA			
					TTTGCCCCCGTTTC	PDWETAEGYITALPYAL			
					GACTTCAACTTCAG	YDKLQPYSQSQHQPPRQ			
					GCCAAAATGGCGAT	QQQQQRQRPRQQQQTR			
					TTCGACCCTCCACG	QRRQRRCKRGGGSQHR			
					CGCCGTGCCACTGC	QRKTRRRRPPRVTRHHR			
					TCGGCACCGGCGGC	EHRIDEALDDLHALESR			

					GATTCAGCGGGTGC	RPQDRTAISKARRRVGRI			
					AACTTCGGGCACGT	RSALDQHQLRHRFDTDE			
					GTGCAACACATGCA	KACVDGILAAARDKDR			
					GCGCCCATTGCACG	AASVTTTAQTAAPPHSA			
					CCAAGCGGCATCGC	PASAPSSAVDDGICPIPG			
					GGGACGACGCCTCG	DLLHAFFTDVNTPRTEF			
					GCCGCTCAAGCGCA	DADSPIGARFREALAQLP			
					GCCCCGCCCTTCCA	AAIAATELLMEPPSPDEV			
					GCACGACCTCGCGC	EDQLQRVRGTSSPGLDG			
					CGTTTGGCGGATCG	VGYDVYKTFTQQLLPAL			
					CCATCAAGACGTGC	HAAFSRCWTDQRVPQS			
					GAGAGCCAGGCGG	WKLGVVRLLFKKGDRQ			
					GGTCGGGCAAAAT	DPANWRPICLQQAVYKL			
					ATACTTACTCTAAG	YAGILAHRFTRWLDANT			
					TATGCCCGAATCCC	RHADAQKGFRAVNGCG			
					TGCCCTCTCAGGCT	EHNFLAATLTDNARRRR			
					GAACGCGGCCCCAT	RELHVVWYDIKNAFGSV			
					ACTTGATCTAAGTA	PHELLWEVLRRMGVPA			
					TGGGAGGGATCCCT	QFIACCQGIYDAAAFTV			
					GGCCTCTCAGGCTG	GNAADGTTAPIQLRLGV			
					TACGCGAGACCCGT	FQGCPLSPHLFTAVISPL			
					ACGGCCGAATCCCC	LHALKRLPGTGVQLSAV			
						DRPGASAYADDLKVFSD			

					TGGCCTCTCAGCCT GTACGCGGGGC	TKDGITRQHQLVTDFLR WTGMVANPSKCSTMSV QRDNRGVLKTANLTLQL DGAQIPALGMTEAYAYL GIGDGFDHVRRRVELAP KLRELKADTTALMQSGL APWQVVKALKVYIYPR VEYALRHLRPFQQQLQG FDRHLARGLRHLLRLPT SATTEFLYAPTSRGGLGL LPLTEVHGALQIAHAWQ TLHSPDPAIRRIARVQLR QVADARHRLDAEHWKE RGEELCERLLNSQLGTS AHAPPKRRNCDIGSLWV DVQRHLRSLGLQLQTAP ADTHTGAPAQPLQLRVP HHDKWLTHKDVLRHVK LHIKNNHWHRWTSMRD QGKTARAHGGEGSGFLT QPRGMWEADYRFAVAG RLNQVDTYSVLKRRRLR SHDRCRQPGCHRAETLA			

						HVLNHCPGTMDAVRGR HDGALKRIERELHASAT DRRDRVELRVNQTVPSL AGPALRPDLQLYNHTKK TVAVVDLAVAFEEQASD DASSSALSLIASHKRAKY DRIKRHLERQGWKVHLS ALVYGSLGAVASGNYQ VYTTHLGLLKRDAKRLD RQLSVECIQSSRRIWNLH CSQHRTRQHQARPSQGP RGSRATETGGTPSQTSR R			
I	I- 2_RP r	.	Rhodniu s prolixus	CACGTTGTCAG AGACGATCGTG AGGATGGATAT GGTGGAAGTGC CATATTTATCAG TAAAAAATTGC CTTTCGACCCAA TAGTGCTGAAT AATCCATTAAG GGGCAGAGTTC	TAGGATGCAGCTAA ACTCCATAGTTGCG ATTTTAGTTCTTAA TTACTCACACAAGT ACTATCATAACTAC ACTACTAATTACAC CACTCTAAGCTCAC ATCAAATAACGAA GTCACATTCAGAGA CCATTGTGTTGTTC	MWGGHRENLEGRALAE LIDSYNLCTLNDGSPTMI PRPGARAQPVDITLCSPD LMSELSWEVIQEPLGSD HLGILVKSFNGRQATSG ARIRHPKAGGWNLSKA DWEKYRMVLDLKFSTS ESTSCTYDYLTQEIRSAA LGSIPRFPQWLYLNPNR AIKPWWDDDCNRVKQE	cd0165 0	340	595

				AGATATGTGGG	TAGGAAGGAATGTC	SIIAYESYKRAPTLENYV			
				GCTAAAATCCT	AGTACGCAGCGCTC	NLNRKLAISRRTYKEKK			
				ATTTAGGAGCG	AAGATATCCTTATC	RIMWYEFCNSLDRQTPS			
				TCAGTATTACG	TCATGGAAGCCACG	NVIWKRIRSLKKGYSPQ			
				GTCCTTAGCACT	ACGCAGAATGTAGT	VEVTDDGHWVEEFHKS			
				TACTGCCGTCCT	AGAAGAGTGGCTG	LTPDYASQFIDLSNIPIGD			
				GGGCATAAAGC	GCATGACGAAGAA	EDLNXPFQLWELKYALK			
				CTCAGTAGAGG	AGAAGGGAAAGAG	KGGSTSPGQDGIHYQML			
				AATGGAGTAAT	GCTGCGCTATTCCT	QNLPFPALLSLLRIMNE			
				GTACTAAATCA	CAAATTGAATATTG	MWVGGAMIESWRTHLI			
				GGTTTCGGGTCC	GAAAAGATAGGGC	FPLLKPGKDPNSYQSYR			
				TCTTATTCTTGG	TGGTTTTTTTCCACC	PIALSSCVCKTYERMIKH			
				TGGCGACTTTA	AACATATTGGTAAA	RLDWWVEEYNKLAPHQ			
				ATGCGCTTCATC	TTGAGGCGCTGGGA	LGFRRHRGTLDCLAAFV			
				CA	GCCAACTGTAATTG	NEVHLNRTENKPLLAAF			
					TTCATGCCAATATG	VDIHKAYDNVNIAILKD			
					TTGGTTTTTTCCAG	KMLVMQVPGRVIVNIIN			
					CCCTCATATAAAGT	MLQQRQLIIRRGVLVTG			
					AATTGGTTAATGTG	PRTTGIGLPQGAVLSPLL			
					CAAACAAGACCACT	FNIYTSDLRDWLVPEIRI			
					ATAACAATAAATAA	VQFADDFTVYTGGKSEE			
					ATAAATAAA	NLVEIMGQALHDLNAW			
						CNHLGFRISGNKSSVTM			

						FSRKRQNILAGTIQLGPF KVSQSPIVKVLGLYVDH KLTWRPHIKRIIQSCEKA LNVLRYVAHVTYGTDP KIALLLYRNLILSKIDYG CGIYGDACQTSLSGLDK VQYKAIRICLGAMRSTP TNVLLVEAKIMPLHIRRE LLLGKFLSKRLIKFKEVS LMRVSELSEVVRSHRFW LKKKTPLMCEMFRVTEK LYSGLEVLHASPRRSLH YEVSCKESPYVHLDLYD KVKEEWGDLTYLAFIAR MREIIINPYLLFTDGSKSE QGVGCAVLDPQTGSFRL YKLPNNASIFTAEIVAIW NAVILAEQIQSNDIIVILS DSKSALQKIKRPGDVFD PIIIEIISHLINAMGRLHLV WLKGHAGIVDNEKVDV LAKRAILLGEHTPVLNL ADFEGALRKRCYDAWQ			

						KLYDLEKEIKGKYYADL QPRIPNKPWFVEVLEGE RQFYTTIMRLRFNHNLT NPHLHKIKLLDSPKCDC DSGEDGDIDHLLFRCSR YNRNRDKFLKELATHVR FPCSAKSILAIMNPDALR IIYKFIRENKIKI			
RTEX	RTEX-2_RPr	.	Rhodnius prolixus	GAATCCAGAAA TAGAGTATCCA AAGATAAATTT ATTAATAGTAA CGGTTATAAAC TTTTGGATTTTT TCGATAGCTTCA GTCTTATAGTCC TTAACGGTAGA TTTGAAAATGA TTTGGAAGGCG AGTTCACCTAC ATTAGCGAGGT TGGCTGTTCGGT AATTGACTATTC	TAGTTTATGTTTAC TAAAATTTACTTAT TTATTAAAATATAT AATATA	MVSSFRVDHQLYSDHM PIITTLLIHDKQQIYNDNT LPLLPKLNWSKQNLQKY QQSISFYLNHIVYKYDTL DAEVNYLNDIIRSSAIGN NNKKGPKVWKADWFD VECDRARRKSFKQLRLF RRTNEVEDKHLYLKYNF DFKRLCKNKKQKYYVD CARRLYNVKSSKEFWQ QVRAFQKRKPWVLGNV SPEQLFEYFCSLHNSSVT ARLVFDVSHRVDAFLDK PFSFDELELVLSKSKENK APGTDRICYEFYKFAPIV	cd01650	280	540

				CTGCATTAGTAT TAAAGGTATTG AC		FLNRLLLLFNKIFETSKIP ASFKESIIFPLYKKGNKN SASSYRGISFINVICKLYS GLILHRLNNWLESNNIL KESQAGFRSNYSTIDNLF NLTNIIKIKLSKKKQKVY AFFVDLSSAFDSIDRFSLI YKLHRTGLSTKIVTSIKN LLSGTTASVMCRNGLSP NFGTNIGVKQGCLLSPTL FSIFINDLEECLGGGVVIS DNCKIKLLAYADDVVLL ASNPISLQLMINEMESYC YRWKLKINLQKSKIMIF RNGGRPANNEVWYLGT QKVEVVNEFKYLGIILTP TLSYTKHFQNKLSKAKM SLNTIWHHFMDNTSIPFS TKIRIFNSVNRSMMCYG AQVWGWQRFVEVERLF YFFIKKLFKSPKTTPSAFI HTETGIDPIYLYTFEMHF KYLVRVLQLPDNRYPKI			

						LCHEIIKSKILWFLEIKNL ITKFNLPNDLNWQNIGN FTLWEANFNILIKSLQAD MNNTFQLKLFNSGSHPL YKNLQYNLGDNPYLSDS HPFHLIRMIFKVRGDLLP LNYKFYKNRDLDCLMC ASGEVEDLFHFLAVCPA LSNMREQWFGSRTLPAQ HVAAYLNGKDWEALAK YCLAALSRRALLT			
RTEX	RTEX-18_SK		Saccoglossus kowalevskii		TAGTACTGTAAGTT TTTGTTGTTCTGAA TGTTTTGTTTTGTTA TTGTTATTTTTGGA CTTTACTCAAGATT TCATTGGTACGTTC CAAATAAAAAATTC TATCTATCTATCTA TCTAT	LNNKVNVEDKSFSQTIV NDISSMESLDDYKHIINQ DFTNKEVLQHIKLLKSG KASGTDMISNEMLKHG APFIINSLTKMFTQVLNS GIMPLHWNSSLIIPIFKSG NPKDANNYRGIAISSCVS KLYTMMLNTRLTKYLQ ETNQIVDNQFGFKKNHR TTDNIYILKSLINKYVNC KLKKSGNYLFSCFIDFSK AFDSVWREGLLYKLLRL	cd01650	96	361

						GVGGNFYKAIKAIYKNT KYSIKLSAGITPYFNSCR GVKQGCNLSPQLFNIFV NGICDDVKNFDPVDLNG LKLNCLLYADDLLLFST TQEGLQKCIDAVQNYCD NWHLKVNINKSKIIIFNR LGKNLENKFQFYLDATK LQCVNYYNYLGIVFQTN GKFHNAMSRLQDKAQK ALFKMYKMLWISEECYS VSLAKKLFNSIIKPTALY GADVWGGVMLNIYKGS GKNKCTVDNCNFETKLF SDKLIPEKLHIKFCKRVL GVHRKASNDPCRGELG QYPLILDICVSVMKYYQ RLVCKKKTILSLMQPMK NPKLLCMTHGLTLYKN YKKNNSIIDAAYEESKTL VHDPWLNFIQKLLVITD LTNSDPHQNNFCDIFKE KLKCLYRMKWKNNLLN			

						NCKLDFYRCAKENFEFE HYLEKIKNKNHRIAYTK FRTSAHFLPIESGRHKGI LKENRLCNLHECNEVGD EKHYILYCNNCQIKEQR SRFLAALYIYAPKLEDFS MDELFPIIVSSNDFFILSIF AKFIFYILDMF			
RTEX	RTEX-5_SK	.	Saccoglossus kowalevskii	CC	TAATAGTTGTTAGT GTTATCTGTTTTAA CATTTTATTGTATTT TTTATATCTGTATTT CTTGTTATTGTGTT GTTGTTGTTTTTGAT TATTGACCATGGTT AATGTATTCCCAAT TCATTAATTTGAGT CAAATAAAATATAC TATACTATACTATA CTATACTA	CINDSLLETFVSKHDIVC LLETWLDYTDSVSVPNG YLYFRVDRKKHKNARK ASGGMIIFYKSKFRKGIN KFISKYKESLWLKLDKH FFGLENDLYINSTYIPPIG SSYYNSSDLTGEDIFNIIA EESLHYSNHGNVCILGD LNSRTVNLNDTLDDINIS DIDIPYTELMDNNRQSM DTMNNKYGQYLCDLCL TTDMIIANGRIVGDLMG KITCFQYNGVSQVDYCV VQKSFLPKINYLKINELS HISDHCSLSLAINTNYVN	cd01650	551	812

						EHSIDISLTPYKKYNWN NSSKENFVTQLNSKDSV DKMNTFLDNHFNLEESD KAISDFNDIMHFTADKA LKFTAIKSRKFSRKNKR KNKQKWFDNDLAHVRK CLKDISRLVDKFPFDQNL RMQRMSINKRYKKLVK LKKNKFKEKLLHDLMN LETKNPSEFWDIIKQIKE YNDLDTSKHPPSIASTE WTNYFSKLYNISTTDSN SVLDSEVTNFVDNMDH NATSNPDLDKEISIAEIIK AIKKTKVNKSSGPDSILN EMFKCYSPLLLACISKLF NHILSSGCFPVHWSTSYL VPIHKSGHIWDVNNYRG IALSSATSKIFNTIIHIRIQ HFLDKNNFFQDTQFGFR KNHRTTDNVYILKSLIN KYTNKVGKHYLYACFV DLRKAFDSVWREGLLFK			

						LANAGVTGKIIHLIKSMY SNTKYAIKTNQGYCKPF MSSSGVKQGCNLSPLLF NIYINDLSESLVQTEPVR LHDMSFNCLLFADDIVIL SNSGSGLQDSLNLFNDY CKKWKLTVNIKKTKVIIF NKRGRKINDIFFLGDEQI EVVSSYTYLGITFQNSGN FTVAVNNLVKKAQRAL GVLKQTLLLNNIIASPIN AIKLFRSSVMPILLYGSEI WAAFNFKPNITPELFIQK LFSDSSKIESIHIKFCKQL LGVHKRASNIAVRAELG SYPLITEICTQMYKFWLR VKNKECSNIVHCAYQSE CTINSNSWANFVSKLLEI TRLNETDHNANIFRNNL KRIYYEKIWLPTLSKKEH NKLRIYNKYKSDYEFEP YLNKINNLNHRVSVTKF RISAHKLPIEKLRYQKIPS			

						EQRFCTLCSDQSVGDEM HALFDCKLDEIVKARSY FWNTLSLYNDNILKLPK EDMFVHIMGIKDSFIANS LGKFINTILQCNT			
RTE	RTE-1_SK	.	Saccoglossus kowalevskii	AACAGTGGCTA TGCAATAAACA GTGGCTATGCA ATGAATAGTGG CTGTGCAATGA ACAGTGGCTAT GCAATAAACAG TGGCTATGCAA TGAATAGTGGC TGTGCAATGAA CAGTGGCTATG CAATAAACAGT GGCTATGCAAT GAACAGTGGCT ATGCAATAAAC AGKGGCAGAGG AGACACAGGCT GGTGGAATGA	TGAATCCGGCCACC CTCGTCTTCGGGGA CCAAACCGGATTTA AAACTCTGGACTAC TACAT	RGEGSNSKKERCIQENA KSGTEENKVRYKRLKK QAKKVVAKAMREENEK VLSELKENPNKVYKIVK AMKGDGRDVTGGRCM RGSDGKLNFSERDRSKV WKEHMEKIMNEENEWD QDVEADLVEGPVERVSR EEVVKAIRDMKTGKAA GPSEVSAEMIFASGDIGI GVIMELCQRVLDGRGM PEEWEVSVVVPIFKGKG DAKSCGAYRGVKLLEH AMKIVERVIVNRIRKLVE VDEMQFGFMPGKGTIDA VFILRRMQEEYRAKGKK LYMCFVDLEKAFDRVPR KVMEWAMRKRGIPEAM	cd01650	188	431

						VRAVMSLYEGAKTRVR VGSELSEEFEVKVGVHQ GSVLSPLIFAVVIDVVTE SAREGLMREMLYADDL VLMDDTMEGLREKFVK WKKAFESKGLKVNLGK TKVMVSGTEEKALVSK VDPCGMCGKRVKVNSV RCVKCGKWIHGRCAKIK RVTQRLARDFACGRCK NGVGRVVEQEEKLCDE VESVKEFTYLGDKVSVG GGCEAAVTARARFAWV KFRECGELLYGKRFPLE MKGIVYKSCVRSALLYG SETWCLSENELGILRRTE RAMVRVMCGVKLMDR KITEDLMQMLGLKEAV DRLAKANGVRWYGHVL RREDGHVLRRALDLEVS GPRKRGRPKKTWKMQV EEESGKVGMRMRDALN RAKWREGVRAIATSLR			

RTEX	RTEX-10_SK	.	Saccoglossus kowalevskii	AAAGCACAGATACCATATAATTAAAAAACAATATACTAG	TAATTGTAATACGTATTTTATATATACCCTTTGATTTGTTTGTAATTTTGAAACTTTTGTTGTTTTTGTTTATTTGTGTTTGTTTTTTCTGATCTGTATCCTTACCATACAGCACCCACTGATGGTGCAGTTGTGTAATAAAGTTATTATTATTATT	INKNKKNFFKNEMINKLETCDKTDQNEYWSLLKQLRCQQNSLNPDSTINSDEWLDHFSKLYNNNNNNADFKQSLKLLEDTKLSNHELNNVISYNEIKKAIKNLKNKKSTAEDSISNEMLKTKNSLLIPCIQKLFNFILESGTYPTSWKSTILVPIHKSGDRSNPSNYRAIAISSCLSKLFTGIMNCRLTNHLEKNKLMNEVQFGFRKNRRTSDAIFLLKTAVEKTLYGKKETGAKCKRLYSCFVDLKKAFDSVSRDALFLKLQKTGINGNFYNIIKSMYTDQEYKVKLNSGITKGFHPTSGVKQGCNISPTLFNIFMNDLPGIFGENVDKPIMIDTFPLNCIQYADDLAIISESVDGLQNCLNKLNDYCTKWYLNVNVRKTKVMIFNNSNRLIKNMNFKI	cd01650	150	418

						NNEEITSVKEFCYLGTIF TPNGKFKINQSNLKNKA MKALFSIRHTILNHKMIS TKLSLKLFDQLISPILLYG SEIWANEITNENNYLEQ ACNHFYRFILGVSKVTPL IGIRGEIGRIPIRIKATLSLI KYWHRLSLLPENHILKS ALKETVSSNSSWIKHVK TILGHTDLSEVNKIYKVG NIAELTKLLEEKLKNKYI QTWYQQLSHVPKNGKS GNKLRTYNLFKMEYEFE GYLTEIQDFNLRRMVTK FRLSDHNLEIELGRRKFP RMEANYRLCKKCNTGT VEDEKHFILDCCFFTQER NTLLEDCKLEMCSTTNR NLIFTCIINSSTALAKYLK NTIYKT			
RTEX	RTEX-6_SK	.	Saccoglossus	AAGTTAAATCT AGTGAGTACTT ATGGTGTAAAA	TAGCAACTCCATCA AATTAGTCTGCAGG TAGTTTTAATGCAA	IFIYVCLYPPENSVIHNR QDIDIFEKLSQDVIKYSNI GEVLLMGDFNARTGTL	cd01650	458	726

			kowalev skii	TTGATAGTAATT TAG	TGGCATATTTAGTA GCTTAGTAGTTTTA TTGTTATACTAATT TTCATGTATTTTTTC CTCCATTCTTAGTT ACATGTATTCAATG TTTTATTCGTGATTC TGGTGGATTGAACT GTACTTCCGTCTTG CTTTGTACTAAACA ATAACATTAGCTTA TGTTTAATGTGCTA TTTGTGATATTAGT TGTGTGAAGTATTA TTTCTGTTCTTGTTT TACAACCATAAAAT TACTCTCTGTAGTA GATTTTCGTGTTGA ATAAAATTATTATT ATTATTA	ADHIINDNSTYMPIPDNY VDDEYNNRNNKDQKVN SFGKKLLELNKNNNLLIL NGRTPGDLLGKFTCYNF AGGASTVDYALVSKSFL SNIQYFRVNEINNLSDHC AISLAITTRDLLSSDNNEI KCTPLPDRYLWDTKSIK RYQDALHSTEIINKLVSF NYENNNIXNSVTELTNII CNAADNSLKKIKHKISSS NKMKNTKRQKWFDDD CWNLRKKFREAAKLVE KYPKDCHIRGQFMVLRK KYNKILKIKRKQYKQNL WDSVTSQESNPKQYWK NFKIFSKNDTSISANNDII QADEWYNHFKNLNNLK NSKTCKYDSDFQNKIINE TREMETNKEINKELNEPI TIGEVKAAVKKLKLNKS SGEDKIINELIKYGQNYL MEPITAVFNNVLKSGTFP			

						KGWKGSIIVPIHKKGSRS DPDNFRGVALSSCLGKL FTATLNNRLIKYLEKNN LYSPYQFGFRRNRRTTD CLFILKSIVDKYIHSEYSK NSKSKKLYACFVDYKK AFDSMWRDGLYWKLQ KYGITGNIYNLIKNMYS GVSYKVKTSKGLTNEFD SFMGVKQGCNISPTLFSL FINDLPEKIKSANGTPPE LNSQKLNCILFADDVLIL SQTQEGMQCSIDALAKY SNSWKLSVSLNKTKLM VFNKRNHTLNKKSVFYE GNEIATVDSFNYLGIVFT PNGKFKTASNTLKEKAM KAMFPLKQSLLTLSHYK ISTALNLFDKLIVPIINYG CEVWGEESAKHSNIIESV QFAFCRFLLGVSKNSTK NAILGELGRYPLNVKIET QMLKYWLRLKSPVIPNI			

						VATAFMEINTWKTGWV ERIEEITKNNNLSELSKQ HEKKIGEVLRENYVKTW RRELFNDSRMNGATNKL TSYRKFKTTFNMENYLN DIKXLKLRRLFCKFRISD HKLLIETGRQKKIPRNER ICKLCTDDKIEDEYHFLI GCSKYDIPRNKLFKYIQG KKSDFLKMNEDEKFRLI MTSFAGIETAQFIHTSMN LRDET			
RTEX	RTEX-13_SK	.	Saccoglossus kowalevskii	AT	TAATTAATATAACT TATTAAGATACCTT TTCTGTTTTGTTTTT TCTAACCATTTCTT GTTCTTCGTTTTTAT TTGATTTTATTATGT TTTCACTWTGTCTT CATTTGTATTTTTCG TCTTGCCATTCTAT ATGTATTTGAAACT TTATTGTATTTKAA	TDSQDINTNXKHESNILN KLSTDSKELNEPISQLEIQ KAINNLKNGKSPGKDNI SNEMIKAGRHTLIIPVQK LFNMILTSGFFPTNWNT GIITPVFKKGDKSDPSNY RGITLLNTLSKLFTAVLK QRLIKYIDVNNLIVKEQI GFIPKFRTSDHIFVLKQIL AKYLEKGKKVFACFIDF KKAFDSVWHQGLFHKL	cd01650	89	353

					AAATCCAGATCCTG AACAGGATATGATT TCAATAAAGGCATT ACTTACT	AQNGITGKFYTIIRSMYS ESLSCVKTKNGLTSEFPV KRGIRQGDGLSPLLFNIY VNDIKNEIEQSNNDPPPP SLTVSQLSCMLYADDLL LLSVSESGLQNALNSLG TYCKKWKLEINTTKSKI MIFNKMGKHIKKHKFCI DKQELENVPTYSYLGIK LSASGKFYKAQEDLANR AMKACFKLKQQLFSETN IPIKIYLKMFDSVIKPVLL YCCEIWSNENINIKNLSN FNLPMEKLHSKFCKFLL KVNCKASNIACLLELGR YPLAFSSIVQTIKYWLRI SDKPCNTLIYEAFRLGK DMDSKGIKNWVSHAKN IFNITNMDHMLDTVTAD KKQTIITSLESNLKSMFK NHFKETLYNDNKGKNR KNKLRTYRMIKHNFDFE KYLEDVKNPSHRSALTK			

						LRISAHTLQIEAGRYINQ TVEERKCHYCQEKQIED ESHFLLKCSLYADLRNK FLHTTKVITTIDKSDIENT VDILKTKDKQSLIHLAK YIFECFNLRNDTHKQKIP AAL			
RTEX	RTEX-22_SK	.	Saccoglossus kowalevskii	A	TGAATTTAATGAACTACACGATTGTATTGTTGATATATTATTTCAAGCTCTCTTTCATAACCTTTATTTGTGAATGTAATCGTTATTATTTAAGGTGACTTATAAGTCCACTGGGCTGGGCAGTTCTCTGAGCTGCAGGTCACTATAATAAATTACTTACTTACTTACT	LYNXYLISIIDKGHRDIL GVRFENKLTGFTFLVFSC YLPPENSPWGRDANAFF SHLLSQLYMTDDVDTIII CGDLNARLGNLNDNIDV VDNIRNRHVIDTKINQH GHSFCEFLQESKLCILNG RFDRHQNDFTFISSRGRS VVDYVCVPQDHLSYCN NFNVITMSSLVEKYNLY DALTDRSKLPDHSLLIFD FLDRPLGSIVEQSLPNDI NIGNNRHSATVRKYCLN QTPINFLESEVAREALLH IISKIEICRETQQEIDEIYD DLCKCILNEMNSTIPNYN	cd01650	514	778

						CSSGARKRFRHHKPYW NDELYSMWRVMRERER SYTRFGGSKNIRTILRND FKAARRAFDKKLRCCER NYRNDLVSSIEKVCTEN PHQFWEHINKLGPVKRK DIPMEVYDDGGHITSNH NIVLNKWKHDFSSLYNQ SDLDGFDANFYEHILHH KLVLENDLDDPLYDSNS ELNKNFSTVEIKTVVAK AKSGKAVGVDKIPYEVL KNDSSVAILTHLFQLCFD SAKVPCIWKKALICPILK VRSSDPRVPMNYRGISIQ STLSKLYSALLNNRITRF LENNNILVDEQNGFRKD RNCEDHIYILTSIIRNRIS HDLSTFAAFIDLRKAFDS VDRDALLYKLLINGING KMYKSIRAMYSDTNSSV KLNDFKTDWFNTEQGV RQGDNLSPTLFSIFINDL			

						AVEINLKQKGITLDNNV NVSILLYADDIVLLADNE IDMQMMLSHLNSWCEK WKIRVNNVKSKMLHFR KASEARSAFQFCIGDDV LDYTDKYRYLGIVLDEF VDYSVTAGIAADSASRA LGALFVKFKTLKNLGFS TFTKLYDASVAPILDYCS GVWGFSKFEKCEMIQNR ALRFFLGTHRFSPLPAVK GDTGWVSCRDRRWAH MVRLWKRLTELDDNRL VKRVFLWDHSLNYNNW SSEIYQIMELVDLNYEHL IDVNIHALRENIARRSES SWSEDVSHKPKLRVYAT FKFEYKPEKYVLINLSRG QRSLLAQLRMGILPLRIE TGRFRHEPVSERICKFCD ANVVEDEMHFLFDCQQ YNELRHEYLDSILNEITG			

						ISREQILVRLFSGPIRKLA LFIEKAFYKRRNLLYV			
Crack	Crack-3_SSa	.	Salmo salar	CTCTCCCGAGCT TTCGCTCGCCTC CAGCACACACA CACTGTTGGGG CGAGTGACTCT GAACTTACAAT GGCTAGCCCCA AGTCGTTATATA TTTAATTTTTTC TTGTGCATTTTG CTATTTGCCTC	TGATGTATTGTTGT CTCTACCTTCTTGC CCTTTGTGCTGTTG TCTGTGCCCAATAA TGTTTGTACCATGT TTTGTGCTGCTACC ATGTTGTTGTTATG TTGTGTTGCTACCA TGCTGTGTTGTCAT GTGTTGCTGCCTTG CTATGTGTTGTCTT AGGTCTCTTATGTG TGTTGCTCTCTTGTT GTGATGTGTGTTTT GTCCTATATTTATA TTGTATTCATTTTTT AATCCCAGCCCCCG TCCCGCAGGAGGCC TTTTGCCTTTTGGTA GGCCGTCATTGTAA ATAAGACTTTGTTC	MTPACFVDYKLSLPNRG TDYVCSHTRDSDSLLVT QTFGLPSYSNHLSPALYS CYLMKLLYNMILLPSDA HSDVINQHCRALPVLCR ALIVLLLMISGNVHVHP GPSTVASPNSDLCSDICF TDFCSRKSLGFLHVNTR SLLPKMDQLKVWVHSS NPDVLVITETWLRKSVL NTDVNLSGYNLFRQDRS SKGGGVAIFTKDHLQCS VVSTKSVPKQFDLLVLSI KLSNSSLLTVAGCYRPPS APACTLPALSSLLAPYTK SEFVLLGDLNWDMLKPP DQVLKQWDSLNLSQIIT NPTRYDSKHPEKATLLD VILTNNPDRYQSGVFCN DLSDHCFTACVRNGCSV KRPVLICHRRLLKNFNE	cd01650	627	884

					TTAACTGACTTGCC TAGTTAAATAAAGG TTAAATAAA	QAFLHELASVKWYRISLI PSVEDAWTFFFDIFSGIV NKHAPIKKMRIKNRFSP WFDRDLAELLHLKNCIW RKARHTHTQADWLSFR QMRNKCTQAIRKAKVS YFKEQFSLCGSNPKKFW KTVKDLENKPSSSQLPM SLNVDDVVVTDKKHMA ELFNHHFIKSGFLFDSAM PPCPSNISSSPTPSNATIP DASPSFSPAPLQSFSLQA VTESEVLKELLKLDPKK TSGSDGLDPFFFKVAAPI IAKPISNLFNLSLLSGEVP IAWKAATVRPLFKGGDQ ADPNCYRPISILPCLSKV LEKLVNNQLTGFLDVYS ILSGMQSGFRSGYGCVT ATLKVLNDVTIALDSKQ CCAAIFIDLAKAFDTVD HSILVGRLRSIGVSEGSL AWFANYLSQRVQCIKSE			

						NLLSQPLPVTKGVPQGSI LGPTLFSIYINNIAQAVG SSLIHLYADDTVLYSAGP SLDFVLNALQQSFLSVQ QAFSTLNLVLNTSKTKV MWFGKKNAPLPTGVITT SEGLELEVVTSYKYLGV WLDGALSFSQHISKLQA KVKSRLGFLYRNRSSFTP AAKLTLIQMTILPMLDY GDIIYRSAGKGALERLD VLYHSAIRFATNAPYRT HHCTLYSSVNWSSLYTR RKTHWLMLIYKTLLGLT PPYLRYLLQPSSSTYNTR SASHILLKVPKAHTSLGR SSFQFAAASDWNELQQT LKLDSFISISSFKDSIMDT LTDSCGCFV			
Tx1	Tx1-2B_SSa	.	Salmo salar	GT	TAATAACAAGATAA ACAATGTAAAACAT ACGGGGTCTGTAAA ATGTATATAGGTTC	QRNSKGVMVLINSNFDP NVQIVQTDHQGRWIILN MLLDNKQIWLINLYGPN NDDPSFFENIYKNLSTLQ	cd01650	457	723

					AGAACTTTTGTGAA	ATLDSIIMVGDFNTVLNT			
					ATAGCACAGTTACA	SMDRKGNHTTNYHPQA			
					AATAGAAATCAAA	LKEIMNVMDILELVDIW			
					CTGGATGGACATCA	RLKYPDLVRYTWRRLN			
					GAAATAGAGGAAG	QASRLDYFLMPFSLAPK			
					GACTAAAAACAAA	VKVLIGDRMRSDHHIIGI			
					CAAAAAATAACTAT	YITLTEFPRGRGYWKFN			
					TGTAAAATAGACTG	QSLLDDKLFRTRTEEFIT			
					TGTCTGTAAAATGT	DFFRHNIGTADPLIVWD			
					GTATAAGATGAATA	TFKCAFRGHAIQYSSIKQ			
					AACTGAAGGTAAA	KQFRSKESILTKEIEGLT			
					AGCCGAAGTGTTTA	VQLDSNKNGTIEAQNKL			
					TTAGTTTACTCCAA	EEKQKEMEELIQERSSVI			
					TTGGGGGATCGGTG	YYKNKANWMEYGEKC			
					GTAGGGTTTGCGGG	TKFFFNLQYRNATKKM			
					GAATAATAATAAA	YKCYQKNVLKLVTNDG			
					GGTATATTCTTTAA	VTHDSPNXILKEEVKYF			
					AAAAGTATGTATGT	KNMFSFQAPPSPLTETNC			
					CTATATAGGTATGT	MDFFPNNNVKLTSVQK			
					GTATGTATATATGT	DSCEGQITEEELLDAIGA			
					GTATATGTATGCAT	FKDGKTPGLDGIPVEVY			
					GCGTGTATGGATAT	KTFFDILKGPLLACFNHS			
					ATATATTTACCCCA	YINGRLSDTQQEGLISLL			

					AAAATATATGGGG GATTGGAAATGATG CAGACAATTACATT GATGGAAGCAATA	LKQDPSGIYKDPVHLKN WRPLTLQCCDAKILAKC LAHRIKKVLSDIIHPNQT GFLHGRYIGDNIRQVLET IEYYEISGTPGLVFIADFE KAFDKVRLEFIYKCLEY FNFGESLIKWVKVMYSN PRCKIVNNGYISESFKLS RGVKQGCPLSAYLFIIAI EMLAVKIRSNNNIKGLEI RGLKTKVSLYADDSCFL LKPQLESLHGLIEDLDTF AILSGLKPNYDKCTILRI GSQKNANFTLPCSLPIK WSDGDVDILGIQIPKERN DLTPIHFYRKLAKIDKIL LPWKGKYLSICGKITLIN SLVISQFTYLLMVLPTPS DLLFKLYEQKIFNFIWNG KPDKIKRAYLYNEYEFG GQKLNIKALDLSLKASVI PKLYLNPNWFSSKLVRM SHPMFKKGLFPFIQITPA			

						HFRLFEKEIISKISLFFKQ ALESWLQFQFNPPEKTE QIIQQILWLNSNIVIDKK KIFEEMFKKGIILVNDIIN RTGGVMSHMQLTQTYG NVCSTQNYNQLIAALPQ KWKRQVEGEKSKELVC RPCIKEHKWLKKSVINK NIYQFHLRTKKLTAVPY KLQNSWEEIFDVPIPWH MVYELIRKTTPDSKLQIF QFKLLYKILATNRMLYI WGIQSSQLCRFCCEEAES LDHLFWYCPYVARFWS QVQEWLKNCNICLELTL QIAILGDLKSHSQSINNIII ILAKMFIFNLQSVEAMRI ERFNCFVKHHSTVEKYM ANRNPKWMMLRDRWE GLNGAEGWD			
Tx1	Tx1-2_SSa	.	Salmo salar	CTGTATCCGGTA CAAAGACAACA AAGGAAGGATG	TGAGAGATAGAAT GATGGTCAAAGATA AAAGTAAAAATTTA	MVRGEISIASYNCNGLA DYKKRRSVFTWLKEKE YNIYCLQETHSTSLDEV	cd01650	517	783

				GTAGGCGTGCC	AAAAAGTAAAAAA	AWKKEWGGEIIFCHGQR			
				TTTATCATTGTG	AACATAATAAAAA	NSKGVMILINKNFDLNV			
				GACAAACTCTTT	GTACATTTGAATGA	QIIRNDSQGRWILLNMK			
				ATAGATGGACA	CACTAAGTGGCAGT	VDEKEIWLINLYGPNQD			
				GCTATTCCGAG	GTTTTTACAACTAA	DPHFFENIYTNLLNLQAT			
				ACAGCTCCATA	TGCCGGTTTGCCTG	NDLIIMVGDYNTVLSTS			
				ACACCGTGGCT	AGGCTGATGCCGTG	MDRKGNHSTNYHHRAL			
				GTACTAAATTCT	CCGGTGTTTGTACA	KEITNIMDTLEIVDIWRL			
				ACAGGGACTTC	CGTGCATATACACA	KNPDLVRYTWRRLNQA			
				AGGTTACGAAA	CACTCTCATTCAAA	SRLDYFLVSFSLASKVK			
				AAAAGAAGGTA	TAAACAC	KVLIGDRMRSDHHLIGIH			
				TGGGAACTGTA		ITLIDFPRGRGYWKFNQS			
				AAAATATCTGA		LLEDNLFLTKTKFITEFF			
				ACATTATGAAG		QYNIGSANPLIVWDTFK			
				TGGATATGAAC		CTFRGHSIQYSSIIKKQFL			
				ACACACGTACT		AKETRLTREIHELIVQVD			
				CAATTACATGCT		SNKNDTTEIQNKLEEKQ			
				TGCTTACACATA		KELEELIQERSNVIYYKN			
				CGGACTTATAC		KANWMEYGEKCTKFFL			
				ATACACACACA		NLQYRNANKKNLQKLV			
				CACCTGATCTCG		TEDGVIYDSPNXILKEEA			
				CTCTCTTCTCTC		NYFRQMFSFPSHPLPLNE			
				ACTCTCTCTTTC		DYGKEFFPNNIKNGKLT			

				TCTCTTCTCTCT		NVQKDQCEGQITEEELF			
				CTCTTCTCTCCC		EAIKSFQSGKTPGLDGIP			
				TCTCTCTCTATT		VEVYQVFFDILKAPLLD			
				GTCGAGAACTG		CFNYSYRNGSLSGTQQE			
				TTTTATAATTTA		GLISLLLKQDPDGKYKD			
				ATGTGTTTATTG		PVYLKNWRPLTLQCCD			
				TTTGTCTATCTT		AKILAKCIALRIKRVLPGI			
				TTTTATGTATTT		VHPDQTGFLHGRYIGDN			
				GTCTACAAGTTT		IRQLLEIIEHHETYKKPG			
				ATATATGTTTAC		MVFIADFEKAFDKVRLD			
				AACAAGCAAGG		FIYKCLDFFNFGNSLIKW			
				GGAAGATATCA		VKIMYSNPRCKIVNNGY			
				GAATAGGGGCA		FSESFELSRGVKQGCPLS			
				TTGGGGAATAA		PYLFVMAIEMLAIKIRSN			
				TAACATATTGTA		NNIRGLEIQGLKTKVSM			
				CGGAATACATT		YADDSSFMLSPQARSLQ			
				TGTACTGTAGTG		CLIKDLDNFSVLSGLKPN			
				AAGCCGTCTCT		YDKCTILRIGSLKNTTFT			
				AGAGTAATGCA		LPCSLPIKWADGEVDILG			
				AGGTCTCTTTCA		IHITKDINKLSTMNFNRK			
				TGATCATGTGA		LVKIDKILQPWRGKYLSI			
				AACGTCAATTG		YGKIALINSLVISQFTHLL			
				CTATGGAAATG		MALPTPDDSFFKSYEQKI			

				CTGGGATGTGTT		FRFIWDAKPDKIKHAYL			
				TTTCAATTATGT		YNEYELGGLRLLNIKAL			
				TAAAGGTAATT		NLSLKASLIQKFYLNPK			
				ATGATGCAACT		WFSSRLIRKAHPLFKNG			
				ATGATGGTGAT		LFAFVQIAMSHFRLIENY			
				ACGATATGGAT		TFFKVSVFFKQALQSWL			
				TACAATTATCAT		QFQFHPPEKIEQILWLNS			
				ATGAATATTAA		NVLVDKIPVFMGKMFEK			
				GGCTATACGCA		GILFLNDIVNWNGRVMS			
				CTACATGTTACA		FMELSELYGKVCSIQEY			
				CACATAGACAC		NQLITALPQKWRRRVAA			
				TCAAACATGCA		GGGRELVCLPNIKDQNW			
				AATTGGCAGAG		RRNKNSINRKVYQFHLR			
				TTATTATTTATT		TRMLTTVPYRLQNSWEE			
				TGGACATCACA		IFDVPIPWYRVYELIYKT			
				CATTATAGTCTT		TQDSRLRAFQLKLLYRIL			
				ACTCTTATACAG		ATNKMLNIWGIKSSKLC			
				ACATCATACAC		RFCCEDTESIDHLFWYCP			
				ACTCTCATTATA		QVACFWSQVQEWLKM			
				TCATATCCAATA		HSIDLKLTLEIVLLGDLE			
				TCATACACATC		RPGQSITNILILLVKVFIF			
				AACCTACTCAG		NSQSVDSIRLDRLKLYV			
				ATTTGCTACACA		KHHSIVERYVLRRNPKW			

				CATAATATCTTG TCTCAATTTTCT AACATCTGTGG CATTGGCTCAC AGGCAAACAGG CAAGGGAATAA AACAAATATTG CTAGAACCTATT TCTTGAATTCTA AATATCAGACA TTTGAAAGCTCT AGTTTTGACCTT CATAATACCTCT GATGGCAGTAA CTTTACAAGCA CTAATTATGGTC AATTTAGACTG AGAATAGTATC TAGTT		VASRDRWDGLREAEGW YVELETSGSGVAV			
L1	L1-5_SSa	.	Salmo salar	TAAAAAAAATG TATTTTTTATTA TTATTATTTTTT TTGATTGGGGG	TGAATTTGCACTTT TTAACGCACTCTCA ATTGTMATATGGTA ATCTACCTTTGGGC	WFGALSYLFSDLHAMQ ATIGTTNERRAGLTFTS WNVKGLNEPIKRGKVL AHLKALSSDIIFLQETHL	cd01650	535	793

				GTACGTTCAATT	AGCATGTGCTGGCC	KNNSHNRLKCKWVGQV			
				TAKCAAAATCC	CCGCCACCCGAGCA	YHSNFSAKTRGTAILVR			
				TTGTTCCSATCT	GTTGGGAGGGGAA	KGIPFLHKTTIVDKEGRY			
				CCTAACCCACA	TAGGGTGGAATAA	VIVIGEIHSTSVTLLNIYG			
				ATATGTAATTGT	GGGGGGGATAAAG	PNIDNPSFFKRVLALIPDI			
				ACTGCTGTCTAT	GGGGGTATAAGTG	SHTNLVIGGDLNCVLDQ			
				GTCGGTTGCTG	GGGGGGTTACATCG	YLDRSSTRRTPTSYSSEF			
				ATGGTTTATGTT	ACCCTTTTTTCTTTC	LNTYIKNSNLFDIWRISN			
				TAGACCGGCAT	TACCTGTCCTTGTT	PTGREYSFYSHVHNVYT			
				TGCTTAGCAAT	CTGTATGTCTTGTTT	RIDYFLVDAKLLPYTCN			
				GTAATCTTGTGA	TGTAATATTTGTTTT	VRYHDIIISDHSPLTFSLR			
				TGCTATATCTTG	GTGCCCAGAACTCT	LGDIVPNERVWRLNPQL			
				CTTATCTCAGGG	GGGTCTTTTGGTTA	LTEPTFCEHLKDQITFFF			
				ATTGACCCAAG	ATGTCTGTTCTCTT	DTNDNTETSPALLWETL			
				ATGAMGCTTAA	ATGTTTAGATAAGA	KAYLRGCIISFQTARSRQ			
				GTGCGCAATAT	ACTGTAAACTTGTC	NRGELVELEGQIHLLDR			
				GTTTAAGTTGCA	TTTTATACCTATAC	ENASHPSMEKHKKITSL			
				ACTTATTCTGTT	ACCATTCTTGTGTA	KFEYNQILSAKIAKSFLY			
				TAGGTTTAATA	TGTGTTCAATAAAA	AKQKYFEFGDKPHKLLA			
				GATGCTAATTG	AATATTTGAACAAA	RQLRKNVSDRMIHKVKS			
				AACACTTTAATC	AAA	ASGELLSSPKDINDRFRQ			
				GCGGGAGCCTC		FYETLYTSKADPNPLIM			
				CTCAGTTCAGAT		QNFLEDCNLPALNQEDS			

				GTTAGTAGAAG		NFLNKEISLEEIRETIKSL			
				TTCTAGGATAGT		KSGKTPGPDGYPGEFYK			
				GAGAGGTGAAC		TFSNMLSPYLHKMLVQA			
				GTACAGTTGGG		NEDGALPSTLDEAFITVI			
				ATTTTATTTTTG		HKKGKDPEEVGSYRPIS			
				TTTTACCTCTTG		LLNTDQKILAKTLANRL			
				TTCGAGGTCGC		STLIGKLVHSDQTGFIPN			
				GCCGTGCTTTTT		RNSFFNLRRLFNIMYSQR			
				TGGCATCGGCC		LPNVDLAVISLDAEKAF			
				AGACAATGTGT		DQVEWPYLFKVLQKCNI			
				TTATTATTTTTA		GDGFIKWIQLLYRNPCA			
				TTTTCATGTTTT		RILTNQSLSPRFNLYRGT			
				TTTCTCTCCTAT		RQGCALSPMLFALIIEPL			
				TTGTGTATGCCT		AQAIRSHAAIHGYNTKD			
				GTTAAATGTGG		TLNKISLYADDILLYVTE			
				GTTGATGGGGG		PQDSIPAILXVINLFGTFS			
				GGGTAAATGTT		GYRINWNKSELMPIRLQ			
				GGGGAAAGTAG		NTSWLEHLPVKLSSEKF			
				GGGGACAGGGT		TYLGIVVTKQYSLLFKE			
				GGAGAGTAAAG		NFPSLMQKLKANIQFWR			
				GTGCTTGCAGG		TLPISLLGRINAIKMVFLP			
				GGGGGAGGGGT		QLLYLYQNIPVFIPKSFH			
				GGGTTGGATAC		KQLDSIINPFIWDYKTHR			

				TGCTCGGGTGT AGGTGCTACAT ATGCTGATCGT GA		IGKKHLCKSKMEGGLSL PNFIFYYWAANLRAVTF LLDDAPPPSSWLSMERE ECHPFSIGAVILSPVNLE MSLYRNNPIIHSTVRIWK QIKVHFELRPISFMLPVA RNPSFAPSNLDNTFERW GELGISTIGDLYIGGTFAS FESLRETYNLPRSNFFRY LQIRDYVRKHLPTFGNA KPSMFDGCIKICPTSDKL ISRIYDAFQSVSAPSTDAI KAKWEEELGTDISVAD WEDSLEYIHTCSINSRHR LIQFKVLHRLHYSKTKL HRIFPGTSPMCDKCQAE QGTLLHCFALCSNLYGY WCGIFGVLSEVLETSIDP DPLLIILGVSDSLNGLTN PQKQLISYSLISAKKLILL FWKKREAPSTKLWLSEL ANTVHLERIRYILNNKLS			

						TFDQIWQPFLSYLDQAAL			
Crack	Crack-10_Ssa	.	Salmo salar	ATGCTTTTTTTTGCTATCGTACCTAACATTTTAACGGTTTACAMCGACAGGAGCTTTGTATTTCATTCATTTTTTTACSAGGAACG	TAGCTGTGTATTGCCGCTTTCCATGTTGTCTGTTGTCTGTAGCTTGTGAGGTGTGGAAACACTTTGTTGTTTTTATGGATTTTGTCTTGTTGCTTTTTGTTCTATGTTGCTCTGTCTGTATGCTACATCTTGCTTGTCCTATGTTGCTCTGTCTGTATGCTATGTCTTGCTTGTCCTATGTTACTCTGCGTGTGCTCACTGTTCAATGATTGTCTATATTGTAATTGTTTTTAATAACCTGCCCAGGGACTGCGGTTGAAAATTAGCCGGCTGGCTAAAACCGGCACTTTTACTGA	MSLRXXXYFAMIVSYVAALLIISVRPCWDDTINLLPENIYTKHIGSPIHRSGLYTAYHNVLNNGQVYLAPYSTEPLRRDKHKSNIYRKLLNYLFTTLLLSGDVQLNPGPNITEPAILTGVESRGWPRPLIVAAVEVGECYXPMVSLDSPGSRIDFDSSNIPKPLYAIPDFASGTQAVLISSPHPVQAGGIVNSATELPLIKTKQNGLNPAVRKHRKCNFFQCVNHSRVIWDPRAKPKGLLGGHLNIRSVIPKSDQIQHLLTDSNLDFLCLSETWLHKNSPCAALIVPGYNVFRRDRIEGRGGGLMIYIKEHIRCKQIEWSCDNELECIGLNVTLSPQMSFTLIGMYRPPSTKSVFFDQFNNMLRECDF	cd01650	727	983

					AACGTTGATTAATG TGCACTGTCCCTGT AAAAATAAAATAA ACTCAAACTCAA	GKEVILMGDFNINYEDK SCRKTLKRITNTFDLTQL VKGPTRVTCCSKTQIDL VFSNKPERVTKSFNMVT GLSDHNLTLIARKLSKSR FKLSTVRKPDQLRIPKSE LNYFENAINGINWNDLL SYTDVEADSEVFLSTIQA TINGFLKKIKSKPGQKST LPWLNGEIWKLMKERD YALKTALKSKLEHDRRR FTMLRNKVMKEIRQAK ANFFINIIGEAKGNSKLI WENLKKLTGKDHSNTA KRLEIMVNNNLTQDAVE IAIAFNSYFIDSVRVLTQ NPSTGLLGSVLVNDAQP VFIIREVSESKVNKVISSL KNSKSKDVFGLDSTFLK NYKESLIGPITKVTNTSIG LGVFPRVWKSAIITAIFK SGDPADVSNYRPISILPV VSKVVEKCVAEQLISHL			

						NNSPFTLHSMQFGFRAK HSTETANCFLLENVKSK MDKGGVVGAVFLDLRK AFDTVNHEILITKLSKFN FSPDALRWMKSYLEGRT QCVRVSNELSPTLSYDV GVPQGSILGPLLFSLYIN DLPSVCTGSEVQMYAD DTVIYVHAKSKQQAAQE LTTVMVQVTKWLSDSC LHLNVKKTVCMFFTKR ATDATEPDVYVSGEKLQ VVSDFKYLGIILDSNLSF KKHVKKVIQITKFNLAN FRFIRNCLTTEVAKLYFK SMILPHLTYCLTSWAQA CYTTLKPIQSVYKQALK VLDRKPNSHHHCYILRK HELLSWENLVQYTDACL VFKILNGLAPPPLSIFVK QKTQTYGSRSTRSAMRG DCIVPLRKSTFSKSAFSV RASHVWNTLPSDTHNCT			

						TYHTFTKCMKTWLKVN QTCEHNP			
Rex 1	Rex1-4_SSa	.	Salmo salar	AGATGGCGCCG ACAGACATGGC AGCTCTGCTTCT AGCTCCTAAGC AACTTTGCAGT ATTTGTTTTTTT GTGTGTTATTTC TTACATTATTAG CCCAGAACGTT TTTTGTGTTATT ACATACAGCCG GAAATAA	TAGCCGGCTACCAC CCGGTTACTCAACC CTGCACCTTAGAGG CTGCTGCCCTATAT ACATAGACTGGAAT CACTGGCCACTTTA ATAATGGAACACTA GTCACTTTAATAAT GTTTACATACTGCT TTACTCATCTCATA TGTATATACTGTAT TCTATTCTACTGTA TTTTAGTCAATGCC ACTCCGACATTGCT CGTCCTAATATTTA TATATTTCTTAATTC CATTCTTTTACTTTT AGATTTGTGTGTAT TGTTGTGAATTGTT AGATACTACTGCAC TGTTGGAGCTAGGA	LLDIRAAVTHQHYDQEY DFPELDPLFVPPRAIELIP EAAPRRRRRRRGIRSGLL VRLRRRAHHPPLPSILLA NVQSLDNKVDELRARIS FQRDIRDCNILCFTESWL SPDILSPSIQPAGFSVHRA DRNKELSGKKKGGGVC FMINYSWCDCDNVQKL KSFCSPDLEYLTIKCRPY YLPREFSSVIVTAVYIPP QADTTTALKELHWTLC KLETTYPEAAFIVAGDF NKANLRKTLPKFYQHID CSTRAAKTLDHCYSPFR DAYKALPRPPFGKSDHD SILLLPSYRQKLKQEVPV LRSIQRWSDQSESMLQD CFDHADWDMFRVASEN NIDVYTDTVTEFIRKCIG DVVPTVTIKTYPNQKPW	cd01650	539	848

					ACACAAGCATTTCG CTACACCCGCAATA ACATCTGCTAAATA TGTGTATGTGACCA ATAAAATTTGATTT GATTTGAG	IDGSIRAKLKARTTAFNH GKVTGNMVEYKQCSYS LRKAIKQAKRQYRDKVE SQFNGSDTRRMWQGLQ TITDYKGKTSHVADTDV LLPDKLNTFFARFEDNT VPPTRPAPKDCGLSFSVA DVSKTFKRVNPRKAAGP DGIPSRVLRACADQLAG VFTDIFNLSLSQSAVPTC FKMSTIVPVPKKAKVTE LNDYRPVALTSVIMKCF ERLVKDHITSTLPDTLDP LQFAYRPNRSTDDAIAIA LHTALSHLDKRNTYVR MLFIDYSSAFNTIVPSKLI IKLGALGLNPALCNWVL DFLTGRPQVVKVGNNTS TSLILNTGAPQGCVLSPL LYSLFTHDCVATHASNSI IKFADDTTVVGLITNNDE TAYREEVRALAEWCQE NNLSLQRQQNEGADRG			

						LQETAEGARPYPHRRDR SGEGGKLQVPRRTHHGP WRSGARKITSPSNVNKT KELIVDFRKQQREHAPIH IDGTAVEKVESFKFLGV HITDDLKWSTHTDSVVK KAQQRLFNLRRLKKFGL APKTLTNFYRCTIESILSG CITAWYGNCTARNRRAL QRVVRSAQRITGGTLPA LQDTYSTRCHRKAKKII KDINHPSHGLFTPLSSRR RGQYRCIKAGTERLKNS FYLKAIRLLNSHH			
Rex1	Rex1-3B_SSa	.	Salmo salar	GCGCCGGAGGA GATGGCTGCCG TTTTACGGKCTC CTAACCAATTGT GCTATTATGTGG GGTTTTTTTCGC GTTATTTGTAAC TTATTTTGTACA TAA	TAGTCACTTTAACC ATACCTACATGTAC ATACTACCTCAATC AGCCTGACTAACCG GTGTCTGTATATAG CCTCGCTACTCTTA TTTTCAATGTCTTTT TACTGTTGTTTTATT TCTTTACTTACCTAT	CFCHRILRQKRASGYQD SDHSPRIRQRFFLQQQQA GLTGYSPNTXQGRHPRY SQEEATRRGRRAGCLVR TRRRRVGKLPLPSVLLA NVQSLDNKLDEVRSRIS YQRDIKNCNILCFTESWL NDDMDIQLAGYTLHRQ DRTAHSGKTRGGGLCIF	cd01650	549	801

					TATTACACACATAC	VNNSWCTKSKEVSRFCS			
					TTACCTTTTTTTCCG	PEVEYIVINCRPHYLPRE			
					CACTATTGGTTAGA	FSAILFVAVYLPPQTDAG			
					GCCTGTAAGTAAGC	TKTALSQLYKDISKQEN			
					ATTTCACTGTAAGG	AHPEAALLVAGDFNAG			
					TCTACACCTGTTGT	KLKSVLPNFYQHVKCAT			
					ATTCGGCGCACGTG	RGKKILDHLYSTHRDAY			
					ACAAATAAACTTTG	KALPRPPFGKSDHNSILL			
					ATTTGATT	IPAYKQKLKQEAPVTRSI			
						RKWSDEADAKLQDCFA			
						XTDWNMFRDSSDGIEEY			
						TTSVTGFINKCIEDVVPT			
						VTVRTYPNQKPWITGNI			
						RTELKGRAAAFKERDSN			
						PEAYKKSRYALRRTIKQ			
						AKRQYRAKIESYYTGSD			
						ARRMWQGLQTITDYKG			
						KHSRELPSDTSLPDELNH			
						FYARFEASNTEACMRAP			
						AVPDDCVITLSVADVSK			
						TFKQVNIHKAAGPDGLP			
						GRVLRACADQLAGVFT			
						DIFNMSLTESVIPTCFKQ			

						TTIVPVPKNTKATCLND YRPVALTSVAMKCFERL VMAHINTIIPETLDPLQF AYRPNRSTDDAISIALHT ALSHLDKRNTYVRMLFI DYSSAFNTIVPSKLITKL RTLGLNTSLCNWILDFLT GRPQVVRVGSNTSATLI LNTGAPQGCVLSPLLYS LFTHDCMARHDSNTIIKF ADDTTVVGLITDNDETA YREEVRDLAVWCQDNN LSLNVTKTKEMIVDYRK RRAEHAPILIDGAVVEQ VESFKFLGVHITNKLSW SKHTKTVVKRARQSLFP LRRLKRFGMGPEILKRF YSCTIESILTGCITAWYG NCSASDRKALQRVVRTA QYITGAKLPAIQDLYTRR CQRKALKIVKDPSHPSH RLFSLLPHGKRYRSAKS RTKRLLNSFYPQAIRLLN			

						RDQKAPQQLLPPSHKTP EQVIKWLPGLFALCAPPP SSTLCLSYMHSHFNYTF MYILPQLARPTSAPAHW LTGLSALCPTTRQPLFLR CCYSLFIIYA			
Crack	Crack-2B_Ssa	.	Salmo salar	GCCCACAGGAC TGGAAAACCCA CMACCKGCCCA GGTGATAGGCC CAGGCCGAGGT WGCTKTTCTGG AWAGAGCCAAG AACTTGWGAGG AACGTATATCTT CCTCAACGAGG ACTGTCCTGAA GCTGTGCGCCA GAAGAGGAAAG AGCTGATCCCA GCCATGAAAGC TGCCAGAGTGC GTGGGGACATT	TGACCTAGATAGTT TGTGTGTATGCATT GATATGTAGGCTAC GTGTGCCTTTTTAA AAATGTATGTAGTT CGTCGAGCTGTTCT TGTCTAWTGATGTT CTGTATTATGTCAT TCTGTATTATGTTTC ATGTTTTGTGTGGA CCCCAGGAAGAGT AGCTGCTGCTTTTG CAACAGCTAATGGG GATCCTAATAAAAT ACCAAATACCAAA	MFFSLALFALFHIMSISG KLPRKGLKIAHINICSLR NKVHEINNLLTSDNIHIL AISETHLDNSFDDTSVAI QGYNIYRRDRNAYGGG VAVYIQSHIPVMLREDL MSSVIEVLWLQVHLAHL KPFLLGCCYRPPSANSQ YLNNMCEMLDSVCDVN REVYFLGDLNIDWFSSS CPLKRKLLTVTSACNLV QVINQPTRVFTNTTGTRS STCIDHIFTNAVELCSKA VSVPIGCSDHNIVAISRK AKVPTAGPKIVYKRSYK RFCCDSYVDDVKNICWS DVINEEHPDAALDEFMK	cd01650	557	816

				GCKTACATCCG		LLLPIIDKHAPVKKLTVR			
				CTACGACAGGC		TVKAPWIDEELKNCMVE			
				TCATTGTCCACC		RDGAKGVANKSGCTSD			
				CTCCCTCCCAGA		WLTYCKLRNYVTKLNK			
				AGCCTGGAAGG		KKKKLYYEAKINDIKND			
				GATGAGAGAGC		GKKTLEYFKXNYGQKD			
				CAAGCCTATGG		KFNSIFHRIRWLIHHKTIR			
				GTTCGTAGCCTC		MMEKKLWSTLNXIMGR			
				AACCCCACACA		KTNSTPSFIESDGLFITKP			
				CACACACACAC		FDVANYFNDYFIGKVGK			
				ACACACCAATT		LRQEMPTTNSEQLYSCIK			
				GATTAATGGAC		KQIMKEKQCKFEFCKVS			
				TGCTGAATGT		VGEVEKLLLSINNDKPP			
						GIDNLDGKLLRMVADSI			
						ATPICHIFNLSLEESLCPQ			
						AWREAKVIPLPKSGKAA			
						FTGSNSRPISLLPALSKLL			
						EKIVFDQIQCYFSVNKLT			
						TDFQHAYREGHSTCTAL			
						TQMTDDWLKEIDNKKIV			
						GAVLLDFSAAFDIIDHNL			
						LLKKRMCYGFSTSAISWI			
						QSYLSNRTQRVFFNGSFS			

						NVKHVKCGVPQGSSLGP LLFSIFTNDLPLALNKAC VSMYADDSTIYASATTA NEVTETLNKELQSVLEW VASNKLVLNISKTKSIVF GTNHSLSSRPQLNLVMN GVAVEQVEETKLLGVTL DCKLSWSKHIDSMVVK MGRGLAVIKRCSAFLTP HSKKQVLQALVLSNLDY CPVVWSSAARKDLVKL QLAQNRAARFALNCNQ RADINTMHASLSWLRVE ERLTASLLLFIRNINVLKI PNCLHSQLTHSSDTHTY PTRHATRGLFTVPKSRT NSRKRTVLYRALIAWNF LPSHIAQINSKPGFKKQI KQHLAAQRLSPI			
Rex1	Rex1-3D_SSa	.	Salmo salar	TGGCGCCGGAG AAGAAGGCAGA CGTTTTACGTGT CCCCAACCGAT	TAGTCAAATGGCTA CCCAGACTATTTGC ATTGCCCCCCTCCT CCACACCACTGCCA	MISCRPHYLPREFSSVFF VAVYIPPQSEAGTKTAL NELYSAISKQENAHPEA ALLVAGDFNAGKLKSVL	cd01650	372	624

				TGTGTTTTTTTG	CTCTCTGTTGTCAT	PNFYQHVKCATRGKKTL			
				TTCGTTTATTTG	CTATGCATAGTCAC	DHLYSTHRDAYKALPRP			
				CGTTGTTTGTAA	TTTAATTAACTCTA	PFGKSDHNSILLIPAYKQ			
				CTTATTTTTTTA	CCTACATGTACATA	KLKQEAPVTRSIKKWSD			
				CTTATTTTGTAC	CTACCTCAACTAAC	EADAKLQDCFASTDWN			
				ATAATGTTGCC	CGGTGCCCCCGCAC	MFRDSSDGIEEYTTSVIG			
				GCTACCGTCTCT	ATTGACTCTGTACC	FINKCIDDVVPTVTVRTY			
				TATGACCGAAA	GGCACCCCCCTGTA	PNQKPWITGNIRTELKA			
				ATAACTTCTGG	TATAGTCTCGCTAT	RAAAFKERDSNPEAYKK			
				ACATCAGGACT	TGTTATTTTACTGCT	SRYALRRTIKQAKRQYR			
				GCGATTACTCA	GCTCTTTAATTACT	TKIESYYTGSDARRMWQ			
				CCACGGACTGG	TGTTACTTTTATCTC	GLQTITDYKGKHSRELPS			
				CAGAATCCTTTT	TTATTCTTATCCGT	DTSLPDELNYFYARFEA			
				TTTCCTTTAACG	ATTTTTTTAAACTG	NNTETCMRAPAVPEDCV			
				AGTCTGACGAG	CACTGTTGGTTAGG	ITLSAADVSKTFKQVNIH			
				CCCGACGCGAA	GGCTCGTAAGTAAG	KAAGPDGLPGRVLRAC			
				CGATATACTGCT	CATTTCACTGTAAG	ADQLASVFTDIFNLSLSE			
				TTCTCGGGAAC	GTCTACACCTGTTG	SVIPTCFKQTTIVPVPKN			
				AGGCCCAGATC	TATTCGGCGCATGT	TKVTCLNDYRPVALTSV			
				CCCGTGATTTGC	GACTAATACAATTT	AMKCFERLVMAHINTIIP			
				GTGAAGAGGAG	GATTTTTGATTTGA	GTLDPLQFAYRPNRSTD			
				GCGGAGAAAAA	TTTG	DAISIALHTALSHLDKRN			
				GGGGCCAGAGG		TYVRMLFIDYSSAFNTIV			

				GCGGGCTGCCT		PSKLITKLRTLGLNTSLC			
				TCTGAGAATTC		NWILDFLTGRPQVVRVG			
				GTAGGCGATCG		NNTSATLTLNTGAPQGC			
				AATAAACCCCC		VLSPLLYSLFTHDCMAR			
				ACTTCCTTCCAT		HDSNTIIKFADDTTVVGL			
				TCTGCTAGCAA		ITDNDETAYREEVRDLA			
				ACGTGCAATCTT		VWCQDNNLSLNVIKTKE			
				TGGASAATAAA		MIVDYRKKRTEHAPILID			
				ATAGATGACCT		GAAVEQVESFKFLGVHI			
				ACGCGGAAGAT		TNKLTWSKHTKTVVKR			
				TAAACTACCAA		ARQNLFPLRRLKRFGMG			
				CGGGACATTCA		PQILKRFYSCTIESILTGCI			
				AAACTGTAATA		TAWYGNCSASDRKALQ			
				TCTTATGCTTCA		RVVRTAQYITGAKLPAI			
				CGGAGTCGTGG		QDLYTRRCQRKALKIVK			
				CTGAACGACGA		DSSHPSHRLFSLLPHGKR			
				CACTATCAACA		YRSAKSRSKRLLNSFYP			
				TACAGCTGGCT		QAIRLLNI			
				GGTTATACGCT					
				GTACCGGCAGG					
				ATAGAACAGCG					
				GCGTCTGGTAA					
				GACAAGGGGCG					

				GCGGACTATGT ATTTTTGTAAAT AACAGCTGGTG CACGATATCTA AGGAAGTCTCG AGCTATTGCTCG CCTGAGGTAGA GTATCTC					
Crack	Crack-13_SSa	.	Salmo salar	ATCCAAGATGG TAGCCAATCAG TCTGTCCGTTTG TATGTGTGATCC TTGAGATCAGA GTTTTAAGAGTT TTTATGTTTTAA TGACCCTGATCC TGTTTACAGAA GCAAATACTTTT TATGTGAAGGT GGATGCTGTAT ATAACAACTAT TATTTTTAGATC TCAGTCTCTATT	TAGGCTGGATCATG TTGTTGTATTGTATT GTTGTATGTTTTAA TTATGTAATGTATT GATTGTTGCTGCCT TCTTGGCCAGGTCT CCCTTGAAAAAGAG ACTCTGGGTCTCAA TGGGCTTTTCCTGG TTAAATAAAGGTTA AATAAAAA	MYLLLCLLFTQCHPGLN RLNDQWVSELSSPYALQ FANTPTRPSCQYHQELG LSLQLHLYHSQTQLDLH LMAYFQNLKLLNYFFSV SKDSVRFIDNGAGSTKSS RIGQLRYIVMLLLLISGN VRPNPGPVDTMQSLPTP YDFKNRAGFGLLHVNV RSLFPKIDMIRIWAKTTN ADVMVLSETWLKKSVS NNSISIDGYTVFRTDRMS KGGGVAIYVKSEFNTSEI LSITKAKHFELLAVKVN VYKDSHITVVGCYRPPS	cd01650	635	891

				TTACGAAATTTA		ASGDALNSLSDVLHKLN			
				CTTGTACAAAG		DSEFIILGDLNWDMLTS			
				ATTTCGGATTGA		VSDSFKELCDSLNLAQLI			
				GAAGGCCTTTA		NAPTRPNPRAPNKSTLL			
				TTTTTTTCTACA		DIILTNTPHKYTSTGIFCN			
				CCTTTCTGGAGT		DVSDHCAIACVRNTKMT			
				TGGAGCTCGCC		KAKPRYIFKRHFRQFDE			
				GAGTCAGGAGC		QAFLHDLYHNIDRVSLIP			
				GTCTGTCTGCGC		DVDTAWDYFYMKFVSI			
				GCAAGGCCTGC		CDKHAPVKKFRISGRDN			
				TACGCGAGCGC		PWFSDNLAELIRKRNIS			
				AGGTGCGCACT		WAQARHTNAPVDWASF			
				GAAGGCTAGGA		RALRNKCTGLIRKSKSD			
				GCGGAGAAGCG		YYLNAVTENLNNPTKF			
				GTTATCGGATTG		WKLIKSVSGSNVSSGLP			
				GAAAACCTGAT		DHLMMDSNEVKDKAHI			
				GCACCGTGGGC		VGVFNKHFISAGSVFDN			
				GGTGGGAAAAT		GGAQASNVSSVTVNYDI			
				CTACGTTGTTCT		GPHVNHFNFEPVSYTEV			
				ACACAACGTCT		YKALKAIDTKKSAGPDN			
				ACTGGTCGTTGC		LDPYLLKIAAGIITEPVA			
				ACACCTGTGCT		HIFNLSLLTNSIPSIWKSA			
				GAAATACCTCT		YVLPLLKGGDPSDANNY			

				AAAAACCGGCA TT		RPISKLPILAKVYESLVN SQLKNFFIEKNILSGVQS GFRSGHSTTTAAMAVA NDIINALDKKQHCAALF VDLSKAFDSVDHELLLA RLRNIGLSEGAVNWFRN YLSDRTQCVYTDNHKSS FLEINRGVPQGSILAPVL FSIFINDLGNGMQPAKLH LYADDTVIYSCAPSLVQ AVEELQTAFQSLQASLY GLKLVLNVQKTKFMTFT RARTLPENVSIVTSGGLS IEKVSSYKYLGIWLDDK LSFKVHVDNLVTKLKLK LGFYFRNKACFPLMARK KLVQATFLSVIDYGDLL YMHAASSVLQRLDSVY HASLRFITNAKSLTHHCT LYQMVGWTSLYMRRKI HLYVFIYKALLGKLPLY LCSLVSFTTSSYHTRSAR WLLLKVPRTFTVLGKTA			

						FSSCAPDAWNSLQSMLH LDMLVPLNEFKILMGDS VTEECKCFF			
R2	PERE RE-9	BN00 0800	Schistos oma mansoni	ATCTCACGTTTT AATTTATTTTTG AACTACTGCAG TCTGAGTGCTTC TAACGACCCGA AGGCTCAGAAA CTACCCACTTCT TGAACTGCTACT TTTTGCTGTTTA TCCACAACAAC AGTTGTGATTCT ATTCTCCANATA TTCCTTGTGCTT TTGTCAACATTA TTCTATACCAAC TGTACCACCTAC TTCTTCATCTCA CGTTTTAATTCT GGTCTTATTTTC TCATCATTAGTC	TAACGGCTGAACGA ATAGCCCCCTTCAC TCTTAGACATTCCC CCACTGTTGTTGCT TATCTTCATGCTCTT GTGTTAATTGACTG CTCTCTTCTGGGTT GACGTCTGATTGTC TCTCTCTCTTTCCAT ATTGCTTGCTCTGC CCGCTTACTTCCAA TAGTTGTCATATTA TGTCTTTGTTTACTT GCCATGTCTAACGA CAATTACTTTATCT ACCTTAGTTTGTCC TCTTGGTTTCGATT GCCTTCATATGTTC ATGGCGGAATCTGA TGTTTATAATGACT	MPVSTGAETDITSSLPIP ASSIVSPNYTLPDSSSTCL ICFAIFPTHNILLSHATAI HHISCPPTPVQDGSQQM SCVLCAAAFSSNRGLTQ HIRHRHISEYNELIRQRIA VQPTSRIWSPFDDASLLS IANHEAHRFPTKNDLCQ HISTILTRRTAEAVKRRL LHLQWSRSPTAITTSSNN HTITDIPNTEARYIFPVDL DEHPPLSDATTPNASTHP LPELLVILTPLPSPTRLQN ISESQTSHESNKNSMHTP PTYACDPDETLGATPSST IPSCFHSYQDPLAEQRGK LLRASASLLQSSCTRIRS SSLLAFLQNESTLMDEE HVSTFLNSHAEFVFPRT WTPSRPKHPSHAPANVS	cd0165 0	492	750

				ACGGAGAGGGC	ATTCCTATTACCAC	RKKRRKIEYAHIQRLFH			
				CTATGAACGGT	CACTACAACTACTA	HRPKDASNTVLDGRWR			
				CCGTGACGCGA	TTATTATTTTCATTA	NPYVANHSMIPDFDCFW			
				AATTCTATCCGC	CTATTAACATTATT	TTVFTKTNSPDSREITPII			
				GATTTCGACCTC	ATAAACATTATTAC	PMTPSLIDPILPSDVTWA			
				TCCTGCTAGTGG	TATTATTATTATTA	LKEMHGTAGGIDRLTSY			
				TCCCCGAAGTA	CTATTATTACTTCT	DLMRFGKNGLAGYLNM			
				CGGTTCCTCTGG	ACAATTAATATTAT	LLALAYLPTNLSTARVT			
				CCTGTCAGTTGT	GGCTACTCCTCTCA	FVPKSSSPVSPEDFRPISV			
				GTTAAAACTAT	GCACACCAATAAA	APVATRCLHKILAKRW			
				ATAATAACG	ATATCAATCAAACA	MPLFPQERLQFAFLNRD			
					TCTCAATTATATCC	GCFEAVNLLHSVIRHVH			
					ACCTATTAAACTCT	TRHTGASFALLDISRAFD			
					CTCTATTTCCCCTG	TVSHDSIIRAAKRYGAPE			
					AGTTATAAACTTAC	LLCRYLNNYYRRSTSCV			
					AATTCAGTCTAACC	NRTELHPTCGVKQGDPL			
					GAATATCTCTCTTT	SPLLFIMVLDEVLEGLDP			
					TACAAATCTTAAGT	MTHLTVDGESLNYIAYA			
					ATGTAATTTTGTGC	DDLVVFAPNAELLQRKL			
					CAAACCCATTTGGG	DRISILLHEAGWSVNPEK			
					TCTGTACAATTTGA	SRTLDLISGGHSKITALS			
					TACTTAAAAATAAA	QTEFTIAGMRIPPLSAAD			
					TGTTATTAGCC	TFDYLGIKFNFKGRCPV			

						AHIDLLNNYLTEISCAPL KPQQRMKILKDNLLPRL LYPLTLGIVHLKTLKSM DRNIHTAIRKWLRLPSDT PLAYFHSPVAAGGLGIL HLSSSVPFHRRKRLETLL SSPNRLLHKLPTSPTLAS YSHLSQLPVRIGHETVTS REEASNSWVRRLHSSCD GKGLLLAPLSTESHAWL RYPQSIFPSVYINAVKLR GGLLSTKVRRSRGGRVT NGLNCRGGCAHHETIHH ILQHCALTHDIRCKRHN ELCNLVAKKLRRQKIHF LQEPCIPLEKTYCKPDFII IRDSIAYVLDVTVSDDG NTHASRLLKISKYGNER TVASIKRFLTSSGYIITSV RQTPVVLTFRGILDRASS QSLRRLCFSSRDLGDLCL SAIQGSIKIYNTYMRGT			

				ATGTTTTAATTT	TAGCCCCCTTCACT	MPVSTGAETDITSSLPIP			
				ATTTTTGAACTA	CTTAGACATTCCCC	ASSIVSPNYTLPDSSSTCL			
				CTACTGTCTGAG	CACTGTTGTTGCTT	ICFAIFPTHNILLSHATAI			
				TGCTTCTTACAA	ATCTTCATGTTTTTG	HHISCPPTPVQDGSQQM			
				CCTGAAGGCTC	TGTTAATTGACTGC	SCVLCAAAFSSNRGLTQ			
				AGAAACTACCC	TCTCTTCTGGGTTG	HIRHRHISEYNELIRQRIA			
				ACTTTTTGCTGT	ATGTCTGATTGTCT	VQPTSRIWSPFDDASLLS			
				TTATCCACAAC	CTCTCTCTTTCCATA	IANHEAHRFPTKNDLYQ			
				AACAGTTGTGA	TTGCTTGCTCTCCC	HISTVLTRRTAEAVKRR			
				ATCTATTCTCCA	CGCTTACTTCCAAT	LLHLQWSRSPTAITTSSN			
	R2S		Schistos	AATATTCCTTGT	AGTTGTCATATTAT	NHTTTDIPNTEARYIFPV	cd0165		
R2	m-A	.	oma	GCTTTTGTCAAC	GTCTTTGTTTACTTG	DLDEHPPLSDATTPDAST	0	492	749
			mansoni	ATTATTCTATAC	CCATGTCTAACGAC	HPLPELLVILTPLPSPTRL			
				CAACTGTACCA	AATTACTTTATCTA	QNISESQTSHESNRNSM			
				CCTACTTCTTCA	CCTTAGTTGGTCCT	HTPPTYACDSDESLGVT			
				TCTCACGTTTTA	CTTGGTTTGGTTGC	PSSTIPSCFHSYRDPLAE			
				ATTCTGGTCTAA	CTTCATGTGTTCAT	QRSKLLRASASLLQSSCT			
				TTTTCTCATCAT	GGCGGAATCTGATG	RIRSSSLLAFLQNASTLM			
				TAGTCACGGAG	TTTATAATGACTAT	DEEHVSTFLNSHGEFVFP			
				AGGGCCTATGA	TCCTACTACCACCA	RTWTPSRPKHPSHAPAN			
				ACGGTCCGTGA	TTACAACTATTATT	VSRKKRRKIEYAHIQTLF			
				CGCGAAATTCA	ATTATCACTATTAT	HHRPKDAANTVLDGRW			
				ATCCACGAATT	TAACATTATTATTA	RNPYVANHSMIPDFDCF			

				CGTCCTCTTCTG	CTTCTACAATTAGT	WTTVFTKTNSPDSREITP			
				CTAGTGGTCCCC	ATTATGGCTACTCC	IIPMTPSLIDPILPSDVTW			
				GAAATACGGTT	TTTCAGCACACCAA	ALKEMHGTAGGIDRLTS			
				CCTCTGGCCTGT	TAAAATCTCAATCA	YDLMRFGKNGLAGYLN			
				CAGTTGTGTTAA	AACATCTCACTTAT	MLLALAYLPTNLSTARV			
				AACTATATAAT	TAAACTCTCTATTT	TFVPKSSSPVSPEDFRPIS			
				AACG	CCCCTTCGTTATAA	VAPVATRCLHKILAKRW			
					ACTTACAATTCAGT	MPLFPQERLQFAFLNRD			
					TTAACCGAATATCT	GCFEAVNLLHSVIRHVH			
					CTCTTTTACAAATC	TRHAGASFALLDISRAFD			
					TTAAGTATGTAATT	TVSHDSIIRAAKRYGAPE			
					TTGTGCCAAGCCCA	LLCRYLNNYYRRSTSCV			
					TTTGGGTCTGTACA	NRTELHPTCGVKQGDPL			
					ATTTGATACTTAAA	SPLLFIMVLDELLEGLDP			
					AATAAATGTTAT	MTHLTVDGESLNYIAYA			
						DDLVVFAPNAELLQRKL			
						DRISLLLHEAGWSINPEK			
						SRTLDLISGGHSKITALS			
						QTEFTIAGMRIPPLSAAD			
						TFDYLGIKSNFKGRCPV			
						AHIDLLNNYLTEISCAPL			
						KPQQRMKILKDNLLPRL			
						LYPLTLGIVHLKTLKSM			

						DRNIHTAIRKWLRLPSDT PLAYFHSPVAAGGLGIL HLSSSVPFHRRKRLETLL SSPNRLLHKLPTSPTLAS YSHLSQLPVRIGHETVTS REEASNSWVRRLHSSCD GKGLLLAPLSTESHAWL RYPQSIFPSVYINAVKLR GGLLSTKVRRSRGGRVT NGLNCRGGCAHHETIHH ILQHCALTHDIRCKRHN ELCNLVAKKLRRQKIHF LQEPCIPLEKTYCKPDFII IRDSIAYVLDVTVSDDG NTHASRLLKISKYGNER TVASIKRFLTSSGYIITSV RQTPVLTFRGILERASSQ SLRRLCFSSRDLGDLCLS AIQGSIKIYNTYMRGTQR LNE			
Penelope	Penelope-1_SM	.	Schmidtea	TATACAAACAC CGATAAACTAT ACCTTCCCAATC	TAACTCTGATACTT TTACTTAAAGATTA AAACAATTTTATAG	MDVVEKEGKRKPYKSN LSNNEQGVLKALKTKPY VYLPSDKGGEFCVIEKSR	SSF56672	144	314

			mediterr	CCGATCCAGAA	CAAAAAAGGAAAA	YCEAAYKHLSDSNTYK			
			anea	TTAAATTCAAC	CCTTCAGGAATGAA	KISRMSPNTIETKINKVW			
				ACTTTCGAAGG	AATCAGAGTAGAAT	KDIAAKMNLKKRLTKSF			
				TTTGTCGAAAG	TGTATTTAGGTCTT	VTKNSDLPNFYHLIKTH			
				ATT	GAAGTTGTTTGTGG	KDGPSIKIRPIVSNTNGPT			
					AGGGTTGTTTGAAG	TKTSWLLSKILKPLLNN			
					GCTTTTTTGTTGCGT	VPYHLQSSDCLMARLNK			
					TTTTAAGAGGAAAG	TGKTTHSEFKYPFSLDV			
					CTTTTTCATATACA	VALYTSIDPSEAIECLRR			
					TCTTTCATAAGACT	RLENNLLIKEMNTENIIN			
					TTAAAAACTCAGGA	LLTVILENTYFTFDQNIY			
					ACTATTTATTAATT	KQIKGLPMGSSISGLLAII			
					TGCTCTGAATTTTA	FMDYLERQTLRTIQSIGF			
					TAAAAAACTTTCTT	YARYVDDIFITTKSKKEA			
					TAGTTTTAATTTTAT	EEIFHIFNEAHPDIKFEIE			
					TAAGAAATTTGAAA	YPDENGSLNLLDLRIKID			
					GACAAAAAAAAAT	VDTPTYGFYKKSARKEI			
					GACGTTAGAAGCA	FMHQRSALPYLQKVNVI			
					GTTATTTAACATAA	KNEKQRIANRCTEGKEN			
					GATATAGAAAACAT	YKQQLXFDKVLRRNGY			
					TTTAAAATTTTAAA	HKEDIERMNSKPKTPMD			
					GTTTTATATAAAAA	RTQKKKKYFYLNIPFISD			
					AATTAAAGTTTATA	TLDFKIRKELNHTGLPIRI			

					AANATTAAATTAAT	SHKSWTLRQALANNTK			
					AAATGTTAAATTAA	KENICTMENCQTPKKICF			
					ACAAAAAAGAGAG	RKNVVYCITCNACKEIYI			
					GAAAAAAATTTATT	GSTLRNLHQRIKEHLDP			
					TTAAAAATTTTAAA	AGKSSVYRHLQQCQPRK			
					TTGTATATTTTATA	ERKSISIKILAGDNIQKNL			
					AAAGATATTTAAAT	RFKEAILILDKHPSINTKE			
					TTTTAAAATTTTAG	EISEIASLICC			
					GTTTAAAATGTTTT				
					AAAATTAAATTTAA				
					ACCAATCAAAAAG				
					CAAAAAAGTNAATT				
					TTAATGAAAATCAG				
					AAGGTATTTCGGCA				
					TTTTTCTCATTGTAT				
					ATAGTAGTTCCGAA				
					AGAAACTGAAGAG				
					GATTATTTAATAAT				
					CGAAACTTTTTTCA				
					AAAATATAATTATT				
					ACTTCGCAGCAAGG				
					GCTTATTTTGTTAA				
					ATAATAATAATAAT				

Не содержащий LTR ретротранспозон	LIN10_SM	.	Schmidtea mediterranea	TTTACACATTGACATTCGATAGCACCGGGCCGAAATCCGGATCTGATAAGTTATTTTATTCAAAACAACCATTACTTTATTTTCCAGATAATAATGAATTAAATGCTGATCCTATCTTTGAGAAAAGTTTGCAAGAATTACGTCTT	TAACTTCTA	MSKNGTDPHLLELCYKAHVFLNGSKEIPLSTAVGHWREIDSTGIFDSDFSELMEQIIAKPFKTEGLQLKIENAIKTFDKFIADAVNPDKLINHKHLNFFYNQLYYNPLKNFNFSIPIKSILKNNNIPSHRNFFNNEESTPLNKKSISFKNISHSKPFSYLTALTNNLKYENIDNILISIRNCTDSFFINLNICFFELKHLIKNNFINYFNDYFYIFYLKIQKFNNKKLQKSNYYQNNIPLLFKISSLTNDFFINIENYFSKIKTKRNINLFLNDFLXILKIKINILLFKYSXFKKLNPKNKHEKISTRSSXASGRLALAVTPSSPLDLTINTTKQTTTDINTDTTNTSHKFSNNRYSLFLDQNEPLDLSIVKHPHKNTNIPSTVKNTYNPLNKQKPFHKTTTKTIFSNAQNN	cd01650	1274	1556

						IKKHNIDTSKRPHFSXPIP			
						KSCNTDIKYEIPPIEKNTP			
						KLNDIIXSNITFNPKTPNI			
						QQNISLPKFILPSNLHNN			
						TTNVDLDAESTLVNPHF			
						QLNFNLKPCPNLFIPKXD			
						LPXNLNMEIDQKLLLTN			
						TKASDETISYSPITQTQT			
						DESVLTHKLIELISINTQA			
						NRIIEQINSDIDLFSQSPH			
						SYEIPNTETKEIFTVYNN			
						LDIAKKRVEIIIEEDILKT			
						KTQTNDDIPMINFNNNYI			
						NNNSLFKNILLSNESVEN			
						LFRDISDNDSIISKSPNKK			
						INNQDFLKKTVKINKTM			
						FNLRNDGYLICIGGKCE			
						KKKIGIHNIDNIIEHRKTN			
						HSDTIDKNTTFLCLLCEG			
						SVKRKTILLNKIIQHTETF			
						HSEKHFPVTREELNNKN			
						KILELXMNTEGTLNCRY			
						TNIKKHCQNDLTYKSTY			

						AQIEKHIAKKHKRKTCNI NEIICHCGETLEHHNIIEH HNKHIGVLPAEDKLRMI PDNIPVSVKDIGNDISILS TFFCMHLLPQSQQLLYN HTLLKPSYKQNFKTFFET FPFSQYPNWNEIILPFKTS TVNWAVFYFNKLKAFA TIIDPLSPNSLQKHRKIAD TILNTIKNIQNIIYPSNINL TELNYPMCANTNESGFFI CHIIQCISINHNYITIPIISE IRKRLDNLINKVIIDRAPI DNILKYTNILNNIKILQN NNSENPDEGLAEINRIHE CIFPIKHKIKKXLSLIHHL LLLKIIKLNKKXXFNPPS TIIKNNKIELTYKLRFNF NINYKSTINKIFNNSINTL SPHPDRIILNFEREIIIINDI KFKYKLPYYPENFIHKPT TIEEITNTFKIVDLSSSPG KDQISYLDWKKIDPDNSI			

						LCSLFNNILFKSITPLKW KTFKTKLIIKPGKENSSH DLSSWRPLAILDTSYRLF ATLINRRLLEWIKLGNLI NINQKAIGCPDGCAEHN SVLLLLNERAVRMGTNI NIVWLDLADAFGSISHD LIWNVLTNMGINNNTIK LFKNMYNNCNTYYECE NNHYNFLKICIIIVTLIMN VKTTTTKTIPINNGVKQG CPMSMTLFCLAINFIITNI LKDHPFIINNYNVSILAY ADDIVLISDNRIKIAKILT DIIKYTSDVKLKFRPTKC GFYENNRNHDEKPLKIY NEIIPIIDENNIYKYLGVG FGGKSNHNIDNILEKALT DAEATFLSDLHPSQKLF AYKTSIHSRLIFPFRTCIV KHMILDCDRNRIVDNRT KQLGFDQKIKRLIKNSIG DIFQNISNNFIYAHCSMG			

						GLGIVPSIDEYLIQSTIQL TRLLMSSDVTMRDFIIGE LIECVRARFNYLENEIEM AMKWLNGDIDDEKVGK PCLFLKFKKSVYRLKIKF SIICKLTYIDNIFSLEISKN DFKIVINNENIDNASKIM HILTGEWYALEWYKMT HQGHIAETVGKSQFSKH LIKSTTLGDNQYFFLIKA RNNSLNLNSNPSNHKDK KDVNCRLCRIQWETQA HILNHCTLSQNARRNKH NHVLEIVHNFFASKNYDI DIEKSPPNVDTTLRPDLII KTTDKNKIKLSLVDIKVP YDNETNFINARKQNLDK YRNLAIETGKDNNCKAA VYTIIVGSLGSWDKENN KDLQKLGLGVNEIGEMA RMCIKAAIEASYMIFMN HVTNNK			

RTE	RTE-1_SIt	.	Setaria italica	AACTCCGCCT	TGATTTGTGGCTCT TGTTGGGTTTCAAC TCTAGCCTACCCCA ACTTGCTTGGGACT GAAAGGCTATGTTG TTGTTGTT	MLSQHSRSQALVKEEGC DRLGEPTLNLAILMEMK PKRNPLGRNPLSDAPYR NPGMVLNGQGPGRLPL GDAPCRDLGMVSSEQGS GRHFLSGALHRRPGVVK NEQGSSHLSRRVRRARK LVEPTRIRVGSWNVGSL TGKLRELVDVATRRRVN ILCVQETKWKGQKAKE VDNTGFKLWYTGTAAN RNGVGVLIDKSLKDGVV DVRRQGDRIILVRLVVG DMVLNVISAYAPQVGH DESAKRLFWEDLDSLVR TVPSSEKLFIGGDLNGHV GTTSAGFEAVHGGFGYG SRNQEGEEVLDFAVAFD LMIANTFFRKRRSHLVTF SSGQHYSQIDFILTRRED KRACLDCKVIPGECVVS QHKLVVADFRFQMRAR RDKQAKIARTKWWKLK	cd01650	641	890

						EETSEVFKERVIKEGPW KEGEDANNMWEKMAN CIRKVASEVLGVTKGGR GVDKDTWWWNEEVQR TIKEKKECYRRLFHDRS VDNIEKYKEAKKIAKRA VSVAKGRAYEDLYQRL GTKEGEKDIYRMAKVRE RKTRDFNQVKCIKDEME NILVKEDEIKHRWQEYF DKLFNGENENTTLQLDD SFDDTNRHFVRKIQESEV REALKRMKSGKAMGPD GIPIKVWRCLGDIAIVWL TKLFNHIFRSNKMPEEW RRSILVPIYKNKGDIQSC TNYRGIKLMSHTMKLW ERVIEHRLRRITRVSMNQ FGFMPGRSTMEAIFLIRQ VMERYREKKKDLHMVF IDLEKAYDKIPRNVMW WALDKHKVPTKYVGLI KDMYNNVVTRVRTSDG			

						DTDDFPIRIGLHQGSALS PYLFALVMDEVTRDIQG DIPWCMLFADDVVLVD ESRAGVNRKLELWRETL ESKGFRLSRTKTEYMRC DFGTTTHEEGDVSLESQ VVPRKDTFRYLGSMLQR DGDIDEDVSHRIKAGW MKWRQASGILYDKRVP LKLKGKFYRTAIRPAML YGAECWPTKRRHVQQIS VAEMRMLRWICGHTRR DRVRNEDIRDRLGVAPIE EKLIQHRLRWFGHVQRR PPEAPVRSGTLRRDSNV KRGRGRPKLTWEEAIKG DLKGWNIPKDLALDRSA WKTAIHVPEP			
Penelope	Penelope-5_SIn	.	Solenopsis invicta	TTTTTGTGTTCT CATCTTTTTTTG TCTGGTCTTGAT CATTTGTGGTGA TGGGTGTCATTT	C	KTLLKVETNPXTSTNNK WFVNISKTEIPPSVSNLL QLGGNFCLPPNKKTSITE FIKDIESNRCIQKERGEIR NSVIPLLHNFLHNNSSLD	SSF56672	215	444

				TTGTATTGAATA		EVEKELLRMKRATSEFC			
				GATTTTTTATTC		KSNPNVIFTRADKGNITV			
				TTATTTTAATAT		AMDRLDYINNMEQLLH			
				ATTTGAAAATA		DKETYSIIKNNPIKKVEK			
				AATTCTAGAGG		DLNTTIKRWYDXDFISK			
				ATATCCGTTTTA		QSYFSLHSSDSTIPKAYG			
				TAGAAGTAGTC		LPKVHKENFPFRIIVSSV			
				CAATAATTAATT		NTALYPIAKFLHKIINNS			
				CCAAATTTTTTT		LPFDSRQVYNSFDIAKSL			
				GTTGGAAAGTA		SGMKIREEYCLLSLDAK			
				GGATGAGATAG		SLFTNIPLDLAMDGIRKR			
				CAATAGCGCTC		WPSIENNTKIPFDEFIFAL			
				TATCCACCATGT		KFILSSTFFTFNNVFYKQ			
				TATAGATTGTTC		TFGTPMGSPLSPIIANIVL			
				CAATTTTATGGC		HDLEEKALKSINLEIPFY			
				ATATTGGGTGG		YRYVDDIVMAAPPNSLN			
				TGAGAAAAAAA		FILETFNSFHNRLQFTIEL			
				CGATAAATAAC		EQDRSLSFLDLSLKITDN			
				GACCAGAGAAA		TIIIDWFHKKTFSGRYLS			
				GTTTTTTTATGG		FFSHHPICHKIGTIYNMV			
				AACCAATCAAT		DRALLLSHPTFQQKNLE			
				GATGATTGTATT		LIIGLLLKNGYPLEFIFKY			
				GTCGGTTATTTT		IKIRIKNLFNTKMAPIITN			

				TAATGACAGGT		DQDQTKKDENTKKFIVV			
				CCAAGAAACTT		PYIRNISEKTTSLLDKNT			
				AGGCTACGGTC		FTFGYRCINKINNFIKVH			
				CTGTTCCAATTC		KDKTERDCSSNVVYKIF			
				AATGGTGAATT		CNNCDATYVGQTKRQL			
				GTAATCTATTAT		RTRAKEHINNIKLDPSKH			
				GAAAACTGTTA		SVITNHRIFFDHSFNWKD			
				AATGTCTCGAG		IKILDTEENYQKRLISEMI			
				GATAAAATTGA		HIKQQTNGLNSQNDTDL			
				GTGAGTTCGGA		LHDSYFCLLDSLKDNTNI			
				GGTGCTGCCAT					
				AACAATGTCGT					
				CGACATAACGA					
				TAATAGAACGG					
				TATTTCTAGATT					
				TATTGACTTTAA					
				CGCTTTTTCTTC					
				TAAACCGTGCA					
				ATACGATATTC					
				GCTATAATGGG					
				TGATAAGGGCG					
				ATCCCATTGGTG					

				TACCAAAAGTC TGTTTGTAA					
R1	R1- 2_SIn	.	Solenops is invicta	GGACTTTGCGG CGGGGCTCGGC CTGGTCCTTCTG AACCGGGGCCG GGTCAGCACCT ACGTGGGGGCC GGGGGGGAGTC CATCGTCGACG TGACGTGGGCC TCCCCCGGCGC CCTTAACAGGG TAAAGGGATGG CAAGTGGAGAC CGGCTCCTTGG AAGAACTGTCT GATCACAGACT TATCTCCATGGA GCTGGTCTCCCC CCCCCCCCCCCG CGGAGGTGCGT GAGCAACGCCG	TAGGCCCTCAAGCC CTGCCGGGGGCGCC CATCGCTAGGGCGC CCCTGGCGAGCCGA GTCGGCCGCCAAGG CCGACCGGGTCCGG GGGGCTCCGGAAGT ATGCTGCACGCGTT CCCGGAGCCCCCCA GTCATGGACCAGGG CAGGACACCCGGA GAGGTTTTAGTGGG TAGGCCCTCGCCGG CACGTCGCCGGCGG GGGAGAGCCCCAC ATAACCACCAGGCC TCCCCCGGGGCCTG GGGTATGCGGTAAC GCATTTCCTCTCCG TGAACAAAAAAAA	MMSRACDASMPRSRPM ARRAAYWWSAELAELR RATVAARRAHTRAKRR GMEDALNNAAAVLREA RKALRTAIARAKAAAW DELLSELDRDPWGRPYK IVMKRLRPWAPPVSETL DPQFLGEVVDKLFPTRG RDIPEKTGPPPGWTEELE VSRHEMVAAFARITGRK AKAPGPDGIHAKVWAR ASSIIGEALRGPYTKCLR EGTFPRRWKRARLVLLK KEGKPAESPSAYRPLCM LDDAGKIFERIIANRIIRH MSRNGPDLSPGQYGFRE ARSTVDAIRQVRSLSESI TGDGKVALAISLDIANA FNSIPWGEVVAALSEHF RLPPYLVAIVRDYFRDR VLEFRDKTGLQQRRDMS	cd0165 0	210	472

				CCGCAANCGGG		CGVPQGSVLGPLLWNAT			
				CCGGAAANCGA		YDKVLRGALPPGCHVV			
				TGGGTCCTGAA		GYADDTFVVAGGASWV			
				GAAACTCGACC		IAIAMGNWAVNCVVGSI			
				CCGAAGCCCTC		KSLGLRVAPEKTEAIFFH			
				GAGGTCAGTCT		DGSSRVPPRAFVRVDNT			
				CCTCGCCTCCAC		RVRVGATLKYLGLLLDG			
				TTGGCCGGAGG		RWAFVEHFDKLAQRLG			
				GAGGAGCTGAC		RRTDALLGLMPNLRGPR			
				CTCGGAGCCGA		VGARRAYAYAAMAGAL			
				GGAGGAGGCGG		YGAPVWFREALASRRIR			
				AGCGCCTATGC		DVLHAVQRRLARRIVRA			
				GAG		YRTAPSAATMVLAGLPP			
						AEFTADALAWSYARAK			
						AVRLQGGVVTPRVAAL			
						LRERARRRAMRAWRES			
						LANDPPAAGSRILEAVLP			
						HLDEWVGRPWGGLSYR			
						TTQVLTGHGCFGEYLCR			
						IRKEPTTQCHHCEADRD			
						SAQHTLEDCPAWEELRG			
						VLREEIGDDLSLRAIISQ			
						MVSRERAWVAVSSFCE			

						QVMRQKEEAETVRERRP GGRMPPGEGNDGRGGG DGGHDPPWLPPRPPRRR RPAPRPPGPAARLXKGG GGPADQPRSPPPSPQLRR RGATAPIWRGGGDPPPL RLPAPPSGHAAAGGGPL TPASPARQTRPDGPGGA TLRHNAESLPPLPRRGG GRGKGLLSPPQGRKXGR AAVSAARRGPGLRRGPD TPGLYPQQHQGKRRGR DGVPLPRPRAD			
RTE	RTE1 _Sar	.	Sorex araneus	ATCCACTACCC AGCCACCGCCA CCCTCCAGGTC GCTCTCCGGAC ACTTGGGCCGA ACCTCACGCAC GAGTGAAGCCC CGCAAAGCCAG GTAAGCAGAAC TTTGCGACCTTG	TA	MVTPLRAIGGRPKSLRPS RRARQATESIPPARQSLA SYPSRIRYANNSNNNWP HSPVTERAPNTAALRKM SKERLKKSRGRTRLPKR RKTKMIVCTFNARTLAS EASIEDLMVQAQKIKYDI IGLTETRRHQSHHAIFDT GEELFLGTCDNRGVGGV GVLVNTNLAMSIDSFEC	cd0165 0	583	837

				AGCTCTAGGTTC		LTTRIGRLRLNRCGSLPA			
				AACGAGACCTG		VSIFVVYAPTSNYDEEEI			
				GAAACAAGATC		ERFYMELEKFYKEDHTF			
				CTCTGTCTGCCT		YKIIVGDFNAKIGPRRSP			
				CCTCCAATCTCC		EELHIGTHGLEWNDQGE			
				GGCATCTGAGG		RLSEFIMSTKTIHGNSQF			
				GAGGGGTCCAA		QKAESKRWTWESPGGQ			
				CCCAGACGCCC		FHNEIDHIIFNRRFCLTD			
				CACCCTACCCA		VAVVPKFQTGSDHRLLR			
				AGATAAATACC		AKFYFTERGEKSAKFKS			
				TCACTGGCCGA		KKATPRMTTNWELFGTI			
				CCCCCACTACAT		AATWEDAVLDNIDEEY			
				CCACTACCCAG		DRLVQHLHVCAKNAES			
				CCACCGCCACC		EKATNRRLSSETLELIRQ			
				CTCCAGGTCGCT		RGLARASGNHKLTSELT			
				CTCCGGACCGC		KLCRDAIKEDLKERRAA			
				GGACACACCAT		VLASAAEAGKSIRNARL			
				AAAACACTTAA		SFANYKTKMTALRRPDG			
				TACCCATTATTC		SITSSRKAMEKIIHDFYS			
				CCGCACCATATT		DLFDSHVHLPTYQIPQD			
				TGATAGGGTTC		GYVVPNVLPSEIRHAISL			
				AAATTTGGTGT		VKTRTAPGPDKVRPEHL			
				GAACC		KNLPPVLINTLARLFTRY			

						LSECKVPSQWKTSRTVL			
						LYKKGDIHDIGNYRPICL			
						LSVVYKLFTRVILNRIGR			
						TLDEGQPCEQAGFRRGF			
						STIDHIHTVTKLIEVSREF			
						KMPLCLTFIDLKKAFDS			
						VETEAVIEALAKQGVQT			
						QYIKILRELYCGFTTRISP			
						FYKEVIIDVKRGVRQGD			
						TISPKLFSAALENIMRRL			
						EWEGMGVKIDGRQLHH			
						LRFADDIVLITPNISQAA			
						QMLADFDHECGKVGLQ			
						LNLNKTMFMKNELVPE			
						APFALNGTNISECSSYVY			
						LGREINMRNDLVPELRR			
						RKRAAWNAFKSVEEVV			
						KRTKNLRLRAHLFDSTV			
						LPALTYASETWALRKQD			
						ENAIRVSQRGIERAMLGI			
						SRLTQVREGIRSSDLRRR			
						SKIRDAVSFAKASKIRW			
						AGHVMRFRDDRWTRAV			

						TDWIPRDVRRPRGRPPT RWSD			
Vingi	Vingi-4_SPur	.	Strongylocentrotus purpuratus	GGGGCACGCAA CAAGAGTCCTA TCTCGCACCCCG GGGCAACGTGT CTGACACGGCC GGTTGTAGCAA CTACTTCGAAG ATCAAGATTAA GATTAAGATTG AGAATAATAAT AATTTAGACTTA GACTTAGAGTA ATATTAAGATT ACGATTAAGAT CAAGACCAACA TAACATA	TGAAGTGGCCGGA GTGGCAATGCCTAA ATCGAATGCGGACG AGAATGGGGCGGT GTAAGGTCAATATG GCAAGATGGGGCT ATGCCCATGAAGGA GATACTCTCTGCGA CTGCGGAGAGACTC AGACGATGGACCA CCTCTTGCAGTGCG GCGGAACGGGTTGC ACAGATGAAGACCT GTCCCTTGTTAATG AAACGGCCATTGCA GCAATCGAGCGATG GAAGAACTTAATTT AATTTAATTTACTT CTAAGTTAAGGAAC TTTACATTACTATC TTGATCTAAGGAGA	MKMPNRAASQSNSAEG PLSRPALTVLSWNAEGIS DAKEEVLAHMCKQMNC DVLCLQETHRGSDNPRP KIKGMKLAIERPHRQYG SAIFMNEKIRPDSTSYTH NNDIEVLNVTIKGLTISSI YKPPNISFSYPSTTTNRHI VIGDYNSHSTQWGYDTT GDDGALVETWAQNNDL KLIHDPKLPKSFASARW KRRYNPDLVFCSRLLAP TSRKFVAEPIPRSQHRPI GIQLTAVITPNKTPFKRR FNFAKADWPAFTEHLDH ALTHLPADPKNYESFIDI VKKVSRTHIPRGCRTTYI PGIDEESSQLLANYTTAY DTDAFSEDTINLGEELIS RISQRHRQMWQHMIETT DLRHSSRKAWATIRKLG	cd01650	476	731

					CACGAACGAACGA ACG	IDSTRAQQHYNVTADEV ANQLLKNGKAPKTTNSH RHKLPGVDQNMSLFTRP ISEEELQKGISTLHVNKA AGIDDIRSEQIKHMGSVA QKWLLDMFNNCIENTC VPKQWLRTKVVAILKPG KDSNEAKNFRPISLLCQT YKLLERLILNRIGPHVDS QLNCEQAGFRPGKSTTG QLLNLTQHIEDGFQNNLI TGAVFVDLSAAYDTVN HRRLQTKLHDITNDLKL TRFIQKMMQNRRFFVEL DGQKSRWRKQKNGLPQ GGVLAPTLYNIYTADIPS TENTRNFIYADDTCITAQ HSTFEAVELHLNTALAS LSEYYRANYLRANPSKT QVCVFHLKNREAKRELK LKWNNVQLNHCEHPVY LGVKLDRTLSFKKHIEK TIGKVESRNAIIGKLASS			

						HYGADPKTVQSAAVAL CMSSAEYACPVWSRSA HVKKLDTTLNNTYRKIT GCLKPTKVEEIYHLAGIG SPNQRRTNTCLKERMKQ ATDDRHMLHGYLQPHK RLISRHSFLDSVPPPEQN PPPSAKPVLIKGLT			
Rex1	Rex1-1_SP	.	Strongylocentrotus purpuratus	TGAATGCGCCT AGCAGGCAGCC GTCTCTGTTTTT GCTCTGAATTTT ACCNTGTTTTAC CACTGATAATCT GAGTGTATGCT GTTTTATTGTTG ATCAACTGTCAT CTTATGTACTTG TGACACTGCAT GACATTTTTGAG GTGATGTATTTA TATATTAGATAC TGGAAACTGTTT	TGATTGTATTTAAT ATACTCTGAAACAA ATTAAATTGTCAAA TTACAATTTCATAT TCCTTAGTATATAA TTAATTCATAAATT TGAGGTCTGATTTT TTACCAAGTTAGTC AATGTATTTTTTAT GACATGAAGAATG CGATGCATATCTAT ATTTATATTTTAAT ATCTCTAATGTTCG CCTTATAATATTCA TATTTGTCTTGCAA	MDTYKVYVLFYIALCIII GLRGHSSYSKHETAFVP VRAPSRAHKQLHRSGYH KTPIKYYANASATFTLPII CCGDVESNPGPFELTSC NTNKRYKYTRGELVNIN SNSCGKSLRLPVQTWKT ISELEIGKEFKTRRGCRA GVRKLKHFKHPIPIILPTA RTKLPKKYPRFPRRRCL TSVPVAPQSGCDLPSFLL SNVRSACNKFDELSAVIS YLHVDIAAITETWFRND CPLDNFSIPGYYLCTKSR TNRIGGGVAIYVRELLNP	cd01650	658	909

				TTCTAAATTTTG	TGTATCTTTGTATTT	SPADVQVPINLEVTWVK			
				ATTCAACGGAT	TATCAGCATGTACA	IRPPRLPRSISCIYCAVVY			
				GTGTACACAGT	TAATTGTTTTCCCTC	YPQPDVELETELIDHLET			
				ACATGCAAGTA	TCTGTCCATCTGTC	TIDVILTNHTDAGFVILG			
				AAGCCATTTTAC	CCCTTTCTCCCCCCT	DFNRMNIDPVLLTHSFR			
				ACTGGGCTGTT	CTCTCTCTCTCTCTC	QVVDCPTRGAHILDKIIT			
				ACACATGTACT	TCTCTCTCTCTTCTC	NIHQFYDKPQICSPIGMS			
				ATACTGCGTGG	TCTCTCTCTCTTTTC	DHNSIMWFPLKIKPKLP			
				ATGTGTGAACG	TCTCTTTCTTTCTCT	NTTRYVFNRPMTDSAIR			
				TCGTTCGTGTGT	CTCTCCTGCTTTCCC	AFGQWISSYDWPEVTED			
				GACTGATTTGCC	ATCTCTCTTTCTTTT	VDVNVRCGSFYSTLNNA			
				TATCCGTGGTGT	TTAACTCAAATTTT	INIYFPVRKVLLHFRDKP			
				GACATTCATCTA	TAAATTGAAATATT	WMTGEIKSLIHKRQIAFH			
				CATACAATGAA	GTATAAGATTAATG	QGKGILWKRLRNKVNSS			
				TAGTCTTGATAC	GTATATAGTATATA	IQKAKVKYYRSRIQKLK			
				GCACTG	CACGCTCTTCATAA	KSNPSRWYQEIRILTKGS			
					TTATGTAGATTGTA	RQSLNIPVDGVDPSERD			
					GTGTGACATCTTGA	KIANAINEHLISVSSDLQ			
					AATGTATACTTGCT	PLNLAALPAYLPSRSLPE			
					TTTTTCCAGCAATT	VYPWQVNKVLSKIKCST			
					TACTCATGTGCAAT	ANGPDNISARLIRLFSVE			
					TCAGTTACTTTTTA	LSGPLTYIINASFSQGKV			
					CTGTGATATCTCTT	PASWKCAIVVPIPKVNP			

					GAAATGTATAATTG TCTTCACCATCAAT GTTTTATTCATGCG CAATTCAGTTACTT GTAACTGCGATACA TGGAAATGAATCAA TAAATACCATTTAT CTATCTATCTATCT ATCTATCA	ARIDKLRPISLTDHFAKV IEGFIAKWMYQDIAQHI DPCQFGNQKRLSTSHCLI NILHYVFLNAEVPKTSST ILLTDFQKAFDRIDHTVA IKKLIEIGVRPAVVYWIS DFLSERKQRVRYHSDLS EWCQINAGVPQGTKLGP LIFLVMINDFKSSESIHM FKYVDDLTLVECRKTNQ ISNMQEAVECLSLWADI NKMQLNPTKCIKMDICF SRVPVNQSSIILENCVLS TVSQVKLLGVTIQNDLK WEAHVRDIEKRANAKL YMLRSLGKFGLSHEDLI TIFIGYIRPILEYATQVWS GGLTSAQKQRIERLQKR ALRIIFGKDYHSYANALT KSKLQTLESRRNDLCVT FFKKTLSCTDQFKQFIPQ RAHNRTLRRTNKIPEPRC			

						KTKRMQNSPIPDLIRLYN KSFL			
RTE	RTE-6_SP	.	Strongylocentrotus purpuratus	GGTATGCGGCT CTTTATGTAGGT AACCCCTTTTAA TTAGCTCAGGG AAAATCCACCT CCCCTGCCTTGC ATAGATGGCAT TGGAAGGTGTA GCGAGCCTCCA TTGTAGCTATAC TTTGAATTTATT CTTATTAGCAA GCTAAGAAACA CTGAA		MMRQTNKNGYRPGCRP GEQEPRPPVTNQAEVGK LNTGFKSKSRIKIIDCKA ATLISSFNVRSLKSKEKM GELTALAEEKRIDILCIQ EHHIYHENVDIQMKDM GKGWVLAMSSGTKNTV NSTVGGVGILMSPFAYK SLEYLEKVSSRILLAKFS GNPAPTVITCYSPTNCSP DNEAVNFYSLLSDVIKKI PQHNMKIICGDMNAQVS SADSKGFTFGNRTNRNG QYLLNMTFECDLIACNT RFQKAKEKLWTFTYPNK SQAQLDYILVNKKWRNS VKDCQAYNSFGPLNSDH RIVSTKMQLSLRSNKKV KAKSNLPHDWQLFKSD CSIKEAYTVEVNNRFEA LYQDIENPTANEIYSALI	cd01650	560	825

						EAHEGAATKYVPKRKG KKKRVPWENEKIREKRK CLKVAYNKKNGAPTAE HAETLDKAKKDLQDAY DEEQSNYIAQKCTEIEIG AVQQQTRKVWDIVKEIT GKXSTPVDQLRGKTXQE RLNQWKDHFSNLLGQPP PDNNNPVSKFTIAELKV AVTSLSNNKCPGVDGIP AEVWKSGALDQQLLEV CNRTLIHGEKPDLWSRSI IIPVPKKGDLSTPQNYRG IALTPIAAKVFNKMLLN RIRPXLEPLLRKNQNGFX PGRSTVXHILTLRRLVEG IKAKHLTAVLTFVDFXK AFDSINRKKMLKILXAY GIPHTIVTAVGVMYKDT TAQVRSPDGETDSFTILA GVLQGDTLAPYLFVIAL DYALRTAXEGFEDLGFT LQERKSSRYPAVMITDT			

						DFADDIALISDDLEKAQE LLXQVESAASQVGLQIN STKTEFMLRNLREGEICT ADXTKSKQVQNFQYLG SWVISKQDLNIRKAKAW TALNKLTAIWKSNLGRK LKVRFFTVETVLLYGSE AWTLTSSLEKQLDGCYT RLLRAALNVNWRDHMT NIELYRDITXISARLRDR RLRFSGHCWRSKSE			
RTE	RTE-3_SP	048B05	Strongylocentrotus purpuratus	GGAATACCATC AGAGTTGGAAC ATGGAATGTGA GAACACTGTAT CAGAGCGGCAA GCTGGAAAATG TTAAACAAGAG ATGATGAGGCT AAATGTCAACA TATTGGGACTA AATGAAACAAG ATGGCTGAATA	AGAGAGAGA	MVITNTWFKEHPRRVYT WKSPGDIVRNQIDYIAIN SRFKRAVKQSKTYPGAD CGSDHVPVICTLQCKLK KLKRPKVVKILDFKQLS KPDIRQEFTVQVKNRFE LLIDEGEENTWETMRNI LVETAEEYVPTKERKRR NKWMTEEILLMMKDRQ KIGNRQCKKYRDLDKQI RQKCNEAKESWLNEQC VEIENQFDKNSNVYQRI	cd01650	321	578

				ACGGCGAGTTT		NEISGKKSGCSNPGCIKG			
				GTGATAGATGA		KDGSILFEMKDKMSRW			
				GTACAGATTGA		TQYIGELFEDSREAMPSF			
				TTTATGCAGGA		ADEVKGEKILKSEVRWA			
				GGTGAAAAGCA		MSKMKKNKAAGPDGV			
				TGAAAGAGGTG		VTEMFEALEEYGVDRLT			
				TGGGGTTACTG		EIINKIYDDGTFPDELGR			
				CTGGATCCAGA		SIFITLPKKPGAVDCEQY			
				TATCTCAAAGA		RTISLICHITKIILRILLLR			
				GTGTGTTGGGA		ARSRITPMVGREQFGFV			
				CACTGGTGTGT		KDAGTRNAIFVVRNITE			
				AAGTGATAGGG		RAVEMQKDVYMCFIDY			
				TGCTCCTAGTGA		SKAFDKVKHSELFEELK			
				AATTACAGGGA		KLDIYRKDLRLLQSLYW			
				CACCCTTTTAAC		NQFACIRVEGECSEYINI			
				ATCTCAATCATA		RRGVRQGCVMSPDLFN			
				GTGGTTTATGCA		MYSEIILRELEDENGIQV			
				CCAACGGCAGA		GGHSITNIRYADDTVLFS			
				TAGTGCTGACG		ESEKDLQRLLNVVVKES			
				AAGTTATTGAC		ERMGLEINYTKTKCLVV			
				ACCTTTTATGAT		SKKSSPKCNLKLKEQIIK			
				TCCTTAGATAA		QVSSFNYLGSMIAEDAR			
				AGCTAAATCTC		CETEIKRRIALAKSAFSK			

				AGTGCAAATCG AATGACATTGT CATTGTCCTGGG AGACCTGAACG CTAAAGTCGGT CATGGTACAGA GGGCAAGCAGT AGGTCCGCATG GAATTGGTGAA AGGAATGAACG AGGGGACAGAT GGGTTCAGTGG TGTGAGGGGGA GAAC		LGKILTNRTMSMRTRIR VLNCYIYPVLMYGSETW TLASDTKKRLESCEMWF LRRMMKIPWTDKVSNE EVLTRAGVQRKMIREML IRQLKFMGHITRKEGLE NLALTGKIVGRRSRGRR RILWMDSLQKLLEERGV KELGVVLIQKARSRGLW NSMIAKVKRYGT			
RTE	RTE-8_SP	.	Strongylocentrotus purpuratus	TTCTCTACTATA ATCAGTGTTTTT ATCACAGCACT GAATAGACATA ATCAATAAAGT AAAGTAATATT TGGTTGGTTCAA TTAGTACTCTTA TCATAAGAGTA	AAATGAAGGCGGT CATTGTCGCTTCGC GACAGACAGCTACT ACTACTA	MRNSMKLHTGSVVFRV NKDRAVRSGLPSTVPSM LGXKQQQQKDDTSGITT MTFGQWNVRTLLDREG TGRPERRTALVAMELEK YRIDIAALSETRFSESGR LDDMGYTFFWSGKPDG VRREAGVGFAIRKEILA KLTEMPKAVNDRIMTM	cd01650	589	869

				ATTGTAATCGCC		TVPLTKKGNATLISVYA			
				TCATTCTTTTAT		PTMTSPDETKENFYNQL			
				TAAATGTTAATT		RRTLRDTPRTDKLILMG			
				GTTTTTGCCATG		DFNARVGRDYEKWQSV			
				CTTAGTTGTAAC		IGRHGIGKCNTNGELLL			
				CACTGCCTTGTG		ALCSEFELLLTNTVFKQR			
				GTTCAGGTTCTC		EEHKTTWMHPRSKHWH			
				TTTAGAATCTAA		LIDYIITRQRDRMDVHST			
				CTAAGGCTAAG		RTMRGANCLTDHQLLRS			
				GGAGCAAACCC		KIAFALRSKRRKQGATK			
				TAACAGAAAAC		LAKLNTDRLHVNTQRLN			
				CATGCCACCGT		LEQDMEKALTNWGVQE			
				GACAGCATTGA		GMTVDQTWASLSKVVY			
				GGGCGATTGAC		ETASNTLGKPERKHQD			
				GATACGGAACT		WFDPEDSELLILMKKRD			
				CCCGACATCCTT		ETHQRVLQTRSTRSTTT			
				CGACCAGTCAC		AYKEACRHLQRYTRKM			
				TGCAGGGTCTA		KTTWWEKKAEDLQRSA			
				GTGGTTCACAC		DNNDMKSFYTGLKEVW			
				CCGTCCTGGCTT		GPQTRSTVHLKSLDGNT			
				GTCCTCGCCGGT		TFSDNKRVLERWSQHFE			
				GTGAAACTGCC		TLLNQPGDIEPAARDRIN			
				ACCAGTTCCCC		QRPVVTALDDIPTRKEL			

				GAGAGAGTGCC		NLAIASLGDGKAPGVDG			
				GTTAGTAACTTG		IPSEIWKHGGATLTNSLL			
				CAAGAGCTACC		MLIQQAWAEGSVPQEW			
				GGTACTTATCCA		KDANIVTIFKKGDRTQC			
				CCTTTGCAAGC		GNYRGISLLSIAGKVFAR			
				AGCCTTCATT		ILLNRLNAHITPDILPETQ			
						CGFRNNRSTVDMIFCLR			
						QLQEKCIEHNRSLYVVF			
						VDFTKAFDTVGRTGLW			
						QLLRKYGCPEKFTGMIE			
						SLHTGMMAKVKEAGET			
						SDSFPVSNGVKQGCVLA			
						PTLFSIFLSAMLEEAFAD			
						FEEGVYIQSRQDADLFN			
						VAHFKAKTKSTQILVRE			
						LLFADDSALVAHTPEQM			
						QHVIDVFSSASKKFGLQI			
						NIKKTEVLFQPGTDHRE			
						EEEILVDGTALNQVDDF			
						TYLGSTISKDGRIDSELV			
						KRMSKASSAFGRLRTRL			
						WNNHHVSTRVKCKIYR			
						AIVLSTLLYGAESWTLY			

						KSQVKKLHAFMMRHLR AIMRISWKDKVTNKEVL ERANLPSMEDLLIRKNL RWTGHVIRMPSERLPKQ VLFSQLPAGERGIGRPRL RYKDTIKRNLKRRQIET KTWTTAAGQRAVWRTA VK			
Daphne	Daphne-11_SPu	.	Strongylocentrotus purpuratus	TAGTATGTGAC AATTATCATGAT GGAAAATTCGA TTGTCCAAATG GGTAAAGCATT TGAATCTTGAA AGAATTGTACG GTTGTCACTAAT CCTAAATTGTCA TGATTGTATTCC TTCCAGGGGAC ATGCAGTAGAA CTTTGTATCATG CAGGAATGTAT TCTTTACCCTCA	TAGTATAACAAAAC CGTGACTTATTTAT TCTTTATCATGCAA CTGCTTACTGCACC TGCACTTGCACCGC ACCCTCTTGATCAC TCAGTTTTTCGAAG AGAGTTATGTCGCC AATTCTTTTTACGA GGCCGCCATCCATC TCCAGCCTAACCGC TCACGATGGTGGTC TCCTGCGTTCTAAA TTACCTCTCTAATTT ATTGAAAATATTTC	MNSHAHTPDILVDDINE SKYYTPQEFKNSFSYKP DNISLLHINSRSLNKNFE YVENLLHSLNNFAFSVIG ISETWLHQNSPDMFNLP NYKLIRADRKGRRGGG VAFYITQNLNFKIRSDIK LTQAESLFIEINNSNVKN VIIGLIYRPPDSNCELFCD ELDFYLHKIGDENKHNFI LGDFNINFSPSSDNSNST NCMQLMHSYGFFSIINK PTRINPHSSTQIDNIFSNV HNNKTIGGILCSEVSDHL PIFLTCECKLSCSKSFDEL	cd01650	531	787

				AATGGGCTTGA	ATATTATGTTATTG	YRKESKRNIELLKQDLFF			
				GACAAGTATTC	TATGCTTCGTACAT	EGWAEVYNVNDVNTAY			
				GTTATCCTTTGA	GTTTATCGTATGAT	KHFSNKLQYYYEKNIPL			
				TCTTGACGACTG	ATAATTGTACATGA	GKMSNKRNKPRNPWIT			
				GGATTATCAAG	TTTTTTATTTAATTA	KGLLKSIQTRNRLYKLH			
				ACCGAAGTCCT	CTTTTGTAAACATT	LRNPTDTNLNTYKIYRN			
				CACTCTTTATCA	TTATTCATTGTATTT	KLTKLIRTSRRLHFSNKI			
				ATCATTATTATC	GTTATTGTTCGAAT	KKASSNTTATWKIIKEV			
				TGTATAGCATTC	TTTTGTACTTTGAA	MGKKTNPPPTDNITLNG			
				CAAAAATTTGA	TTTGAATTTGTATA	TKIHDPSLHANSFNSFFT			
				GTCTATCTGTTA	TTTTTGAACGCAGA	NIGPELASKINRTHAHFK			
				TATTGTAATTGA	AATAAAATCAATCA	DYLSNQNQESLFLNPTN			
				CTGAACTTTTTT	AACAAACAAACAA	PSEIINITQSLNNSKSHGF			
				TTGTTTTTTGAG	A	DRISMSLLKEIMYPLAKP			
				CGGGAATGTCA		LTHIFNHSLSQGVFPDLL			
				AAACCCTCTTTG		KIAKVNPILKKDDPHEIS			
				CCATTTAATATT		NYRPISLLPSISKVLEKIV			
				ATTGTACTCCGA		HNRLHKFITKHKILNSNQ			
				AATGACGTTATT		YGFRKNYSTDLALIQIYD			
				GCTATTAAGCA		KITSAIANKEHVIGIFCDL			
				TCATCTCAGAGT		SKAFDTLNHAILLSKLSH			
				CTGCTTACATGC		YGIRGQPLLWFKDYLTN			
				ATACATTTTTCT		RRQYVTFNSYDSDPLLIQ			

				CCTCCGCACAA		CGVPQGSILGPLLFLLYV			
				CAAATCTCTTTG		NDIINTSPLLSFILFADDT			
				TTTTGCAAAACT		NIFYSHKDLNSLNNTLN			
				CCTGTGTCAATC		YEINKVSDWFKSNKLSL			
				ACAGAATCCAT		NTKKTHFMYFKHHSHN			
				TGTTATAGACTG		TETPIYIKIDGMLLEKKT			
				AACTAGTTTTGA		NSKFLGVLIDETLTWND			
				TGAGCGGGAAT		HLHQVTMCISKSIGIISKL			
				GTCAAATCCCTC		KFLLPHATLFLLYNSLVL			
				TTTTCCATTTAA		PYITYCNSVWANCGSTK			
				TAATCGTACTCC		INSIFKLQKKAIRICTGSH			
				GATTTGTCATTA		FLAHTDPLFYKLKTLKIF			
				TTGTTATCAAGC		DLNTTQVAVIMFKYINN			
				ATTTTGTTTCTT		LLPPSFDNMFRFNNSVH			
				CCCAGCCCCTCT		SYSTRISGNFHLTNPKLSI			
				ATTTTGCGTGTC		THKSVRHSGPDIWNDFP			
				AAAAATACACC		DNIKSCSTIYSLKATLKR			
				CTTTACTCGGGC		QIIQSYAPHSLN			
				GGTCGGGCTCC					
				GCTCTGTTGCTC					
				CTAAGCAATTA					
				TTGTAACTGTGA					
				CCAAGATGGGT					

				CTTATTGTATTT TCATTTGTATTA TATTTTGAAGTT ACCCGCACACT TGCACTTCGTAC ACACACAAACG CACCAGCATCA CACATGCAGTT GTTCATCCTCCA AATTTTCGGTAC TTTACATAATAC GTATCGTTAA				
CR1	CR1-31_SP	.	Strongylocentrotus purpuratus	GCTGTAGTAGT CTACATGTGTAC TGACNCTACTG ATCTGTTCCTGC CAGCAAGTAAT GTATGTATTATT AGTAGTAGTTCT TCATAATTAGAT ATTGTGTA	TAAATTCCGGCAAC GTTTTTTTTTTACTG GCACTATGTATATA TACACCTCTCACCT GCACACCTGGACAT CCCCATGTCGCCAC ACACATGCACTCGC ACACATCTGCAAGA ATCTTCAGCAGTGA AGGAGGCAGATTAT TTTGAAGAAGAAG	MGRKPKYPCGVCSRAV TTRTTGVQCEGTCLKWF HAECIGLTSTEYVAFDSS SIVWICNHCGAQNLAHS CLSSLQLXPDTNRFGVL DPDGQNVISEENSNSCRP TXDSSNPSSPGSPAYSSS PQNIKQRRSKHKPRPRV LRFLNVNCQSIRAKMQP FSILVHQHNPDIVIGTES WLSDGINSNEVFLDEQY	cd01650	632	884

					AAGAAGAAGAAAT ATA	SILRRDRSNDPHGGVFIA TKTDLIVTREEELETDCE ILWCRIEMAGSKTLRVA AYYRPKEDDEVSLAELQ RSLSRLDRRHLVVLAGD FNLPGWKWDGPYVGPC AYPALHHQFGEIINDHG LTQLVKEPTRQANTLDLI LISNPTLVNSVKVVPGIS DHNCPVADIDMSPPRRY QPKRKVPMYKKADWEG FAELMQHVADYICSKSS SASVHDLWQLFKSGIEG GIQTFVPHKMLKQKQDL PWITSGIRKLIYKRDRLH KQVTRLRKQNMLVPFPL ETELQHLKRLIQAESRK AYWSHLESIFTADGPNE YEGMKRFWKFVKNNRT DRCGIHTLKTDEHTITSP QDKAEALNKQFKMAFT NETDVPADLLPETSPYDS MPDIHISTEGIVKMLKKL			

						KVRKAPGPDGIPALVLR			
						KLSRVVAPSLRAIFDVSY			
						RTSEVPDDWRLANVVPI			
						YKKGDKTDPENYRPISL			
						TSIICKMMEHIFASNIMG			
						HARRNSILRPLQHGFQA			
						KKSCELQLTGFVTDLLN			
						NMEDNNQTDIVITDFSK			
						AFDKVGHQRLLKKMEY			
						YGVRGHNIDWISSFLSN			
						RKQQVVLDGHTSSTVEV			
						QSGVPQGSVLGPCLFLF			
						YINDLPNQLSTKVRLFA			
						DDTVLYHTIQTPEDAEL			
						LQQDLDQLCGWSSKWM			
						MALNPKKCQVITVSRKK			
						TKYVHRYSLNGEILESVS			
						AAKYLGITLTADLNWTQ			
						HVSATCKKANNTLAFLR			
						RNLRVNSTRLKTTAYQA			
						LVRPLLEYGSAVWDPHT			
						IRCVNQLEMVQRRAAR			
						YVLNRYHHTSSVSDMLS			

						QLGWTSLQKRREHQRLT LLYKMHYSLVDFDIHHY IQPATVVRTTRSSRPYCY RVMQSRTEQHIHSFFPRT VRQWNALPLEVVLAPSV EAFQRQLKTGGK			
Vingi	Vingi-3_SPur	.	Strongylocentrotus purpuratus	AACTAAAGCAA CAATTTGAAGT CTATCATCGCG ATCTTATCGCCA GTAGGCCTACA TCTCTAGCCCTC TATCATCAGAA GAGTTTCCTACG CAGTTCCTGGA AGCACGTGTCC TTTTCAGGACA AGCCATCAAG	TAGATGATGTACTC CCTGATATTCTTTG CACGAACGAACGA A	MMADTSTCVLQLNAEG PYPLKARSDYPHIAKEN WVNVICLQETHCETEEQ LHIEGFDLICFIPHKKHGI ASYVKQGMKASTTGRSP VNSALQWSTFEAMGTH FCNLYKPPPVDVEIILFP KMPRNSFVCGDFNSRST AWGYPXTNKNGKLIFE WLESQELVILYDSKDNP TFYSGSXKSWTCPDLSL VSSNIRHTCSRTVLPKFP RSGHCPIIISSDIKLQLSN LPKKRWNFRKADWPKF QELTNNKLTSRLPDPSTS NVNDSYSAFCSMLHQA AKLSIPRGRRHPYIPCWD	cd01650	473	728

						EECDRLYESCAKSKTIEE REAVGNDLLJSLGEKRRI RWEEEVSEINFTHSSRKA WTTINKLTGRSSAPKQC PVTANSIASVLVNNGKW QDKTTEAKAHSRSVNSE IKNLLKSTPESTSMSEAF TLDEVETALKSIKSGKAP GNDNIHTEFXXNMSRKA TSWLQAFLSTCMTTSTIP SKWKTAKVIAILKPKKP ADEPKSYRPISLLSHIYK LLERLLLARIVDTVEEKL PTTQAGFRKGKSTTDQV VRLVNNIETAFQKKQKF GAVLVDLTAAFDTVWH RGLYLKTLQTIPDVRIVK FIMLLVQDRSFVLETSNG ECSRKRKLRNGVPQGSV LAPILFNIYVSDMPKGEC MQYSYADDTALGFDNH SFEAIKTSLEKDLKMLAE FFNRWHLKMSIPKTVSS			

						VFHLANRCANRELNISL DGSRLNFVSTPKYLGVT LDRSLTFHEHAKATAQK TQARVSLLKKLAGTGW GANFTTLRTSTLALAFST AEYAAPAWSHSTHVHK VDVVLNDAMRLISGTLT ATPVSNLPILSGIPPAQLR RDAQYTKLAQKHDRDK CLVPAPSAFEGQRLPRR HFATQAAELPLEKPPSTL ATWIAKRWSDLQWNSS TTNLKEFIPKPCAKPSGS DLPRKAWVNLNRIRTGV GRTQYFLHKIGAXSSPN CACGEIQTMDHVIDDCPI YKSPHGASGLRELDDETI CWLNKDLPI			
CR1	CR1-24_SP	.	Strongylocentrotus purpuratus		TAGAATACCCGACG CAGAACTGAGCCA GACACTTACCAAGA AGACACCCAACTAC GAGAAGGAAGAAA	TDTDRGKQSKFKRKLRD RKLQIFYTNCDSLLNKR NELLLLLQSHDVDLVLL TEVFPKNHQNPLTHRDFI LEGYDLHCNLEDGNRG	cd01650	509	757

					ATGAAATTGAACTG GAAGCCCGACTGCT GAGAACACAGCATT CTTCCAGAATCTCG GGTAAAGGTAAAG GTAA	VAIYIARDLYHLVGECD LQHSFHESVWILITLRGS DKLLIGCMYRSPSLDNH EELKRAFTQVTSINSSHV LIAGDFNYPNIDWKLGR ARSHTEQEFIETVADCYL HQHIGEHTRYREGQRSS MLDLVFTNEDGMINHID YVPGIGRSDHIGILFGLT VYTPDTHTPTPHLQFYR GQYDDMCKELDQIEWE SELENKSVEDAWEFFLV TLTEKMEKFIPKSKPKHR RKKQRWMTREALRKQS QKHRAWKEAQRTKHPV DIMRARRLAKQLQTLTK QLCRNFEKDLAMNVKS NPKAFWRYTNSKLKTSC RLGELRTPDGERISTDKG KADALNTFFATTFTDED KGHIPKLPPRHKGETLTS ISISEDDVYKKLVHLSKY KCAGPDGIHPCVLKETA			

						ASVSEPLSIIYNASLKEG KLPSAWKVSHVVPVHK KASKMDVKNYRPISLTS VPCKILESLIRDEIVAHL MENDLLCDEQHGFVPG RSCVTQLLTVMEAWTE MLDKGTPIDVVYLDFKK AFDSVPHLRLLSKLETY GICGQIKEWIADFLQGRC QRVRVNDKLSDWTKVR SGVPQGSVLGPTMFVLC INDLPDGLQGSVKIYAD DTKLYGKASTKDDLCSI QSDLDALMEWSNTWQ MHFNTQKCKVLHIGPGN VKHIYALGDDVLEATSE EKDLGILIDDELKFHNHV AYAVKKANGVLSCIRRS FECRDRSTIPKLYKGLVR PLIDYGDSIWYPRFKTDS KTIERVQKRATKMIYDIR NRPYIERLRALNLPSLEY RRRRGDMIQVFKIIHGLE			

						RIPVDTFFEHSTSNTRGH SLKIVKQRCSTKIRQDAF SQRVINDWNTLPEEVVT ATTVNAFKARLDRVWK DKRFCSPYE			
RTE	RTE-16_SP	.	Strongylocentrotus purpuratus	GACCATATTTTT GTTTTGAAATCC CTGCTTGACATT TACTTGTCTAAA CGGAAGAGAAT TTACTGCGCTTT TATAGACTATA GAAAAGCTTTT GAT	TACAGTCTTTCTTT AGGTCAAGTATAGT TTTGATGGTTTGTTT GTTCTTTTGTTCATT TTTTTTTTTTTTGTT AGTATACTTTTTGT TTTGTTTGTCTTATT ACATTATATTCCAT TATGTTTTTGTTTGT CCAATGGACGATTA TGGCGTATTTTGTT AGTCTGCATATATC TTAACAATCCAATC ACAACTCACTTACA TGTACAAAAATGAA ACTTATAATTTCTTT TTTAGTTAAGACAA TAATGTATTTAAAA	MVDRVSLWHKLLRSGV DGKVFTVIKNMYSQAKS CVIVDGQMSEAFQCDVG VRQGENLSPLLFAIYLND FEDFLSKHFKGLKFLVE NVCADILDKEVNKDLFL QLYTLLYADDTLILADT AEELQVALNALSQYCKI WKLSVNAMKTKVLIFSR GKVRKFPVFRFGEDEID VVDDFVYLGVQVNYNG SFKKAISKQVSQARKAL FSMSLKVQRLKLPLDIQ CELFDQLVLPILIYGCEV WGYEDLSQIEIFHRKFLR SALCVNKFTPDCMIYGE TGRAWLANSIKHRMVSF WLRIVKGSPSKYSVKFY	NA	NA	NA

					AAAAAAAATTGTA	FLQRLISQNVDMSTSFK			
					AAATTATTTTTGTT	WLSCIEKILNHAGLGIV			
					TTTATCTTTCTTTTT	WLKQGEGFSNSWILSTL			
					CTTCGTGTTTGTATT	KLRLNDISKQDVMATIW			
					TAGATTGTTCATGT	TNSVCLNYRIFKEDCLFE			
					GGTATTTTCCACTT	RYLVLLSKKDRITLSRFR			
					TTCATGTATGCTTG	CRSHRLPINNNRFKEAV			
					CCATATTGTATAAT	NEISIECPCCDMHESGDE			
					ATTGTATTTATGTTT	FHYVFVCPFFRKERELY			
					AGCATAGGCTAGGT	LGRKVIVKRPCTSHMLN			
					TTAGATCATTCCTA	LFSCSGIGQLKKLVLFLR			
					TTTATCGTTATACTT	IIMSYFARLYTCKSEQSN			
					TTATATTGTTGTTTT	HMSPIKRNVTTRSGRLIK			
					GTTGTATTCTGTAT	QPTILDL			
					TCTTTGTATTATTA				
					AGCTATGTTTGTAT				
					TCTTTATTATGTATC				
					ATTATCATGACATG				
					TACATGTGTTGTAC				
					TCTCATACCCCAGA				
					AGGGGGTCGAATG				
					AGTCAATAAATTCT				
					ATTCTATTCTATTC				

CR1	CR1-5_SP	.	Strongylocentrotus purpuratus		TGAACTTTGTATAA ACTCTAATAGAAGG GGCAAACAGACTTT AGTATGACTGATTT TAAATGATATTTGC ATGTTTTTAATTAC TCACTTTGAAATTA AAATAATGTTTTTT AAAACATATAGTTT GGGATTATTACTAT TGCCTATTTCCAAT GCATGAAACCTTCG GGCGTAGACGCGG AGGAGTGGTCTATT GTTAGTGAGGAGG ATGTCTGTTGTTGT TTTCTTTGTTCATCC CCTCACTTCAATAG ATAACCTACCGTGC CTGCGCCTGCCAAC ATGGTTATGTTTTA GGATCTTGTTTATT ACATCCCAGCCTTT	PSQRVNIELCCEITVSAD KLAPNMLDNFGIIVVLS VIASLPHFIAESRPALLA GVTMEFQLVQNSIDTER YSDQYRYTKDELLSLRN TSHSRLPDNVFARIKELG IDKRRGCRAGSHKWRKI AVVEGRRRRTPKPPGVR APQLIPIKTITTVSTDRQS QGNLPTIYMCNPRSLNN KLDEFSVVVHELCCDVC VVTETWFRADQPESDFA VDDYLVFAKSRDSRRGG GVAIYAKPESMPKTVNF LPPHPHHEVIWLQLRPK RLPRDISCLYVAAVYSPP DNDCVNDLIQYLVNAID DIKTRHGDAGVVILGDF NRADTSELTSTCNLRQV VRKATRGTAILDLILTDI HQWYSEPEILSPLGLGG DHNAVLWKPKLNKTNS KNKAVRKFVRPMKDSAI	cd01650	633	885

					TTAAATAGATTTAA	REFGQWISQHTWDDVM			
					TTCACCTTAGCTAA	ASSSAQEKCDKLYASLQ			
					ACTGGATCATGTTT	NAMDEYLPMKSVKMHP			
					TTTTTTACTCTTTAC	RDKPWVTPYLKSLIGQR			
					TAATTAACTGGATA	QKAFSNDRTEEWKRLR			
					AATCACTCGAGCAT	NKVQNKKTELKKNFYK			
					TATAATACATGTAC	DRVLKLKRENPSSWYK			
					CATATTTAAAGCAT	HLKIMTSSKTASNIIIPES			
					GACCTATAGTATTT	PELDDLAIADCVNDHFA			
					ATTCATTCACAAAA	NVSQDIPELDLQSLPAYL			
					AATGTGATTTCATA	PTPKPQLTVQPHEVYKE			
					GAAAATTGAACGAT	LNRIKPGKSSGPDGIPARI			
					TAATGCTAAACATA	VKEFAVELASPLCEVLN			
					GATTTATATTTTTGT	SSYHSGSVPSQWKGAIIA			
					AATTTTCCACACGC	PIPKEKAASSIDKQRPISL			
					AACACTACTATTTT	TDHFAKIAEVFIYRKLLA			
					CAGAGTTGAAATGT	DVRHQLDPKQFGARAG			
					TCTAAACCAGTTAA	RSTTHCLVEIYNFLCKRA			
					TTGCCATGTACAGT	ELPSNVTSIVSTDFSRAF			
					ATTACGCATGTATT	DRVDHTILIRKLITLEVH			
					TCTTGTTAAGTATG	PWVIVWICSFLTDRQNQ			
					TAATGTTTTAATCA	VRYHGELSSTRSMTAGV			
					TGTAAAAGTTATCA	PQGTRLGPLLFLVLIDDA			

					TTTTCTTATACTCCC	LRESEVERWKYVDDMS			
					TCCCCCTTTCTTTCT	LASAEPSNQHGQFMQN			
					CTCCTCCGCCATAC	VLVNLEEWCNQNNVRL			
					CTGTCCTTTTGTATT	NPKKCNIMRVTFSKTPLP			
					CATTATTTCACTGT	DRDFSLCNDVLQEVSAV			
					GAATTGTGCTGTAT	KLLGVILQRDLKWNDH			
					ATTGTATGCATGGG	VSVTISRASSRLYMLRR			
					TGTTGATCCGTTGC	WKEFHLPAADLVTVYV			
					GTCGTGCTATTCAT	GFVRPLLEYACPVWHSS			
					CAGGGGGGTATATG	LTCAQSAALERIQKRAL			
					ATCTTACTGTTTCG	RLIFPGTSYNTSLDLARI			
					TGCCCCACTGGTGA	VDLETRRKTLCINFATK			
					CTGGGGTGGCGCAG	ASRSEVGSDWFTENNPT			
					CGGTTCGACGCCGG	RNLRKNKRYREPRCRTN			
					TGCCAAGAGTTGCA	RYQNSPIPFMTKLLNFN			
					TTATTATCTATAGT				
					AGTTTATTATTAGT				
					AAGTGCAATTTTAC				
					CTTGTAATAATTAA				
					TTCTGTAATCCATT				
					GCGCAATTTAGTTT				
					TTAGCTACGAAGCA				
					TGGAATTAATGAAC				

					ATTTATTGTTTACA TGTTTGATTTGATTT GAATAAACCATTTA TCTATCTATCTATCT ATCTA				
Не содержащий LTR ретротранспозон	BRIDGE1_TN	.	Tetraodon nigroviridis		TGATGAAGCCTTTT GGATAGAAGGCGA AACGTTTTCCAAAA AGAAAATTGAAAA GTCCAGATGAACAC AGAATCTTTTCTCG GATAACTATGACCT GGATGACTGAGAA CCTACACAGACA	LTKTEKDLLSRGLNSAV TPEELPVVDLITATETAI RNNKLPETEAEQIRLKVS AALANAKAPASNITAQE KRALASLAKDKDITILPA DKGRCTVVLNTTDYDS KILSLLGDSNTYEKLKR DPTSTYKKKVIDLLQKL EKEQAIDKPQYYRLYPG EATPCFYGLPKIHKEGTP LRPIVSSINSVTYNISKFL ASILAPMVGNTPHHIKNS QDFAEKVSNLQLESGET MVSFDVTSLFTCIPTSEA TEAIHKRLLLDKNLQER TTLSPAQICTMLDLCLNT TYFQFRESFYRQKHGCA MGSPVSPIVANLYMEEV	cd00304	296	379

						EKRALLSFTGAAPSRWF RYVDDTWVKIQTKEVE KFTTHLNQTDTFVKFTR EDVKGDSLAFLDCEIRIE EDRNLSIEIYRKPTHTDQ YLLFDSHHPLEHKLGVI KTLQHRATKVPTTSHGR KKEQDHLTTALKTCGYP DWAFTKTRKKRDPSNEE EKDKRHSVSIPYMAGVS EKFRRILQKHNIPVQFKP SNTLRQKLVHPKDKTPR HKQSNVVYAVQCQQEC RDLYIGETKQPLHKRMA QHRRATTSGQDSAVHLH LKESGHSFEDNQVRVLA REDRWFERGVKEAIHVK MEKPSLNRGGGLRHHLS PTYNAVLHSFHKQKPKH THSFTRAGVTSPGDPAV KRKTSQTETRSTTLPTTL PTTPETSIQTTGILMQRLS			

						VDISMPMVQKGYIFKLC TEPDQN			
R2	R2-1_TGut	.	Tinamus guttatus	CTGGGGACCGT GGTTACAACCC GGGCTTAGCTG CAGAGACAGTA CCTCCCCGTGGT TCCCGCCGGAC CCCGTAACATC GGGTGACTGAA TCTGTCTCTGCC CCGGGAGTAGT TCCTCCTTGCCC TATTGACCAGC GGTCGCCGGCT GCTCAATAGTA TTCTAGGCGTG AAATATAGCGA TAGTCCTAGTG GTTGTCTTACTG GGCCATAGCCC CTTGCTTCAGGG GTCATTCGCGA	TAGGGGGCTTGGCA TTTCTCATTGCCTG CTCCTGAAAGGATA TGGGTCCTGCGTCG CGTGGTAGGCAGAC CCATTCGTCCGAGT AGGGGGCTTGGCA GTNTCCATTGCCTG TGCCCGAAAGGAC GTGGGTCATCTGGT CTGTCTGCCTACAC CTCTCTAGACTTGT AACATCTAGTCTGT CAACAAGATCAAA ATTCTTCACACAGA CGACCGAGCTTGCT CAGTCTTCCTGTAC CCGCAGAATTTTGC TCTTGCTCTCCTTTG GCTGTGTCCTGGAC GTGGGACTATTCCA	MGSWIVNFVSVATQTGE FPVDTARRAPVPVTSYP ESECHXPLPLTFCNSDVT IWGGVRPEPVDCLGDLP EXYDALPGVAGPREXV GGSPPGEGVRSPGIASPS GTAVQHDFGSPILVPGA EAAEVSTPVVKVPQDHP ACPCCGTRVVKVTALSE HLRRAHGRKRVLFQCSR CGRMNEKHHSIACHFPK CRGPPVEEGPLGAPEWC CEECGQKFNTKSGLSQH KRSVHPLTRNVERIEAA RPKGKGKRGAHKGCWT EAEVAQLIELEGRFKNQ RFINKLIAEHLPSKSAKQI SDKRRQLAAATKTSSPE KRVTSSTSGESSPEVEKV EGIKREYRRRVGEWLCA GSLXDQTSFQKILEDVES	cd01650	542	806

				AGTCTCTCAGG AGAACTGGGGG TGGTGTTCTTCT GGGTATAGCTA AACCCCCTAGA CTGTGTCCGATC C	TCTCGTCCCAAATG CCGCGTCCAATTAT ACCGGATTTGACAA AGCGGACGGCCCG CTTTATAAGCCGGA AAAGGTGCCTTGTA AAATTGCAAGGTTC ATTAAATAG	GSEIVTGPLEELASFARG KLAAARVRHHRKHPAE AVPAREEQRWMKRRVG RRGLYLRFQRLFALDRR KLAGIILDDVESIKCPLP MEEVADVFRRRWEEVA PFTGSGSFRSLGKADNG AFKPMISAKEVMKNVXE MSRRSAXGPDGLSLRDL MKIDPQGSRMAELFNL WLLAGRVPDQVKAGRT VLIPKSADPGKIGNIDNW RPITIGSVXLRMFSRILSA RLRRACPINRRQRGFIAA PGCSENLKLLQALIKSAK RDHRTLGVVFVDLAKAF DSVNHQHIFQVLVQKGV DGHIIDILRDLYTNAGTY LESGSQRSGFIKILRGVK QGDPLSPILFNLALDPLL CRLEDRGLGYKYGDQQI XSLAFADDLALLSDSWE GMQQSIRVVEEFCQRTG			

						LRVQAPKCHGFLIRPTKE SYTINDCDPWTIADMQL DMIDPGSSEKYLGLGIDP WIGLSRPELSEVLTRWV KNIGGAPLKPLQKVDILR SYALPRLLFIADHAGLSA TCLHSLDLSIRSAVKGW LHLPPSTCDAIIYVSYKD GGLGLPRLASLIPNVQA RRLVRIAQSEDDVIRSVV LQEGIQEEIRKVWISAGG RPEKVPSVTGEFPVMEA QAADEALSEWERRAPRT IYPIPCKWRKREMENWT NLKSQGHGIRNFENDRIS NDWLLHYGRIPHRKLIT AIQLRANVYPTREFLAR GLGEGAPRGCRHCPAE WESCSHIIGYCPAVQEA RIKRHNDICGVLAEEAR KLGWVIFIEPHLRDNTNE LFKPDLVLVKGSCAKVV DVTIRYESGLTTLSDAA			

						AEKARKYQHLAGEVRA LTSATTVDFLGFPIGARG KWYVGNNGLLSDLGFS TSRVVRIARALSKKALLS SVDIIHIFASRARQAQTS E			
CR1	CR1-3_TCa	.	Tribolium castaneum	ATAATCTTTAAC AAAAATTTGCTT TAAAATAAAAT GTTTGAAAAAC CGAAATTAAAC TAAATTAATTGT GTCTAACACTTT AGTAGAGGAAA AAGGAGAAGAC ATAAAAGTCAC CAAAGTCTTCA GAGTCGTTTTTA GTACTAAATTTG CTTATATGCAA GCGCTTAATTA AGGTAACAGTT TTTTGGTTTCTT	AA	MSVCYKCKSSVVNSIAC LNCKKAYHPSCLRAFKN YSAIGQDLTSISACPCCD KLKSLQNPTSPTLNISTN TIFEDLGILKQETDLIKKT LDDLKTFITESPTTVCTF DKIDEYVRKSFDDLESK LISLLKDTLNREIEKLKN SIKNEVIAEISTNLNRTN DNIKTANSSDLTLKSTNS KNVSTQSVSKTNRYKVR IFADQKGRHCIDLLETYI PVSKYVIQSVVKPYSSLQ ELLKTENNNEQLGNNDF VVIYAGNHEAQKGKIPD ENTLNELINHKFLNTNIIL IGCPYFVNRPILNNFVYE	cd01650	915	1166

				GCTTATATCTCG		LNCNFYNVAKNNKNIY			
				AAAACAGTTAC		YLDSNDKIFDQFKTDNF			
				TCCTATCAATTT		VINKAVKRCIIKNLGDLI			
				TTATCTTTTTAC		TECQDHTNRSAVEQNPV			
				TAAAATTAAAG		ATFLDTSLPSVDKSDSQS			
				CTGATAAAATTT		FLSVFLNRFVCLQPSHSK			
				CCTACAAAATA		ISISINKKTVNSFHGIYQN			
				GTTATTTGCATT		VRGLKSKSVTFFNSFLST			
				TTTCATGTAGGA		SDEFDLIVITESWLDDTV			
				CTAACCATAAG		LDGELFSDNYYVFRRDR			
				CGAGATATAAG		DLRHLGIHRGGGVLIAV			
				TTTGAAAATAA		NNKYNCRQLDISNISRRS			
				TTAATATTTAAA		SPAIDLVAIKINVTSYQH			
				AAAAGTGCTTT		LFVFVVYIPPSISFSDFDN			
				AACAGAAAATG		FFDLLNNFDEINNDNVLF			
				CCTTGTGATTTA		LGDFNTPLFSGFSKNNFS			
				AAATACACTTA		ISLKHFASFHNFNQFNNI			
				ATTTATTGGGTC		RNNRNNILDLIFSNMDC			
				CAATACTTTATT		AVSKSSFSFVPEDPFHPS			
				ATTTACTTTATT		LEFTLDIVAKTPDKFYM			
				TTTTTGGTGCTA		NRDHVAFNFKKANFDC			
				TGCTAGGCTTGT		LYTTLQVVDWSELEKFS			
				GCTAATGTGCT		DVNEACRVFYETLYNCF			

				ATAAATAAGGT		ELSVPKLFCRKNYRTFPT			
				GTTACAATTTTG		WFNTDIINCLRRKSNILK			
				TGCTGTAAATA		KYRRTKNMHFYEEFKK			
				AGGTGTTGTGC		LRSLSKSLITTAYSQYIG			
				ATTTTGTGCTAT		HMENTISKDPKKFWSFV			
				AAACGTCAGTT		QAKKRNSRIPGVLYHNG			
				TTTTAATTTTTC		NAISNPNDIVNSFAKFFK			
				GTTAATTATTTG		SVFSDMTLTSDSPSMMT			
				TGGTAATAACTT		NLNSFITVNEINEHEIIRV			
				AATAAGTGACT		LKRSKDTLTAGCDGIPSF			
				AAGGTTTAAGG		LLRDCAHIFSKPLVYIFN			
				TTATGTTAAACT		LILQTSTFPNIWKIAHVC			
				GTAGTCAATAG		PVFKTGDITQIQNYRPISI			
				CGTTTTGCATTT		LCNFSKVFESIIYDRIYSS			
				CTTTTGGACTTT		VKRFISPSQHGFVEKRST			
				CAAACATTTAT		VTNLACFTQFISDALDR			
				ACTTTACACAA		KGQVDVIYMDFKKAFD			
				AGAGTTATACT		QIDHHLLLLKLDQYGFS			
				GATATAAATGA		GSLLSLIKSYLANREQRV			
				CGTGATCAATG		KYRNYISDSYVATSGVP			
				CTCGTACTATAA		QGSNLGPLLFLLFINDIC			
				CCAATAATCCA		GSLSTCAKLLFADDLKIY			
				ATTTTGTGCTAC		TEIKSIEDCLTLQNNINA			

				TTTTTATGAACA TTAGTTTTTACA TGTAAATTTCAT ATCAGAGGCGT ACGCTTCTAGA ATTGCTGATAAT AATTATTATTCT ATCCTCTGTTCC TTTTGTGTCCTT TCCGTGGCCTAC CTGTGCTTGTTT TTTTATTTTGAG TCACTACGAAC GGCCGCGTAGT GTGAAGCGTGT GTCCTTGTCTTT ATTCCCTTGCCT GATTGTATTTAT CAACTCACTCAT A		VVKWCNENRLYLNPSK CNVMSYTKKREFLEFVY DISSVTLHRTFIIKDLGVI FDTELAFSEHIRDVTARA IKSYGFIYRNCRDFKNLS VMKTLFFSLVRSKLEYG ALIWHPIYKIHIDQLENIQ RRFLKFLVFIIDGNYPIRG YDQNLLLNRFGLQSLQF RRICIIIKFLYNLIN			
CR1	CR1-1_TCa	.	Triboliu m	ATTCTAAGCATT GTAAAAGTATT TTTTATCATATA	CTTATGTCTTAAAA TTATTTTTATGTTAT TGTTATTTCTGTGT	MALRAARKNYYGQRIQ NADNKSKCAWNIVNKIC GRSKKCCEIQIEGDPKQI	cd0165 0	148	408

			castaneu	ATATGTATGAA	ATAATACTGACGTT	VNDINQYFINSAPFVRSK			
			m	ATGGTTCAAGA	TTTCGTATCAATTTT	SNNNTISSNAKTMYXYP			
				AAACACAAGAA	GTTACTACATCTGT	VTSQEIITIIQNLKNKKSA			
				TAACCGCACGC	ATAAATTGATAGAA	GFDEISNNIIKYCNEEISA			
				AGTAAAACAAG	ATAAATTATTATTA	PLSFVIQNSLDYGIFPEQL			
				AATAGACAATG	TTATTA	KRAQVLPIYKKGDPTLIS			
				TTTTTACAAATA		NYRPISLLSSFSKIFETVIS			
				TGACCTTTAATA		SRITNFLMKFQLLNSSQH			
				ACTTTAAAATA		GYIKGKSVDSAIFSFTNA			
				GAAGTTCTTGA		LLDKLEEKKVVLGIFLDF			
				ACCACACATAT		SKAFDCLDHNILISKLKK			
				CGGATCACATG		YGIRGIMLKLITSYLNNR			
				GGTATACAGCT		YQQVTITKNSTVYTSNK			
				AACTACGAGCA		LLIKQGVPQGSILGPLFF			
				TAAAAAATAAT		VIYINDLLNILTVSDNNS			
				ACTAATAATAC		AVNFADDTNILIWGENM			
				TGATGCAGTAA		TDTMKEAKICMNKVIN			
				AAAACGTTAGA		WTKENKLVLNFEKTYGI			
				ATTTTAAATGTA		HFAICNHKNTEKPNKIVI			
				TATGCATGTAAT		SSKEIELIEKTNFLGICID			
				CGCAATGATAT		HKLSWTNHIDNLLKKLN			
				CAATGCTCAGT		SVHFTIRVLKKYINIEQL			
				GGAGCAATTTT		RIIYFANFQSLLSYGIIFW			

				ATAAATATTTTC AATGATAAATT TAACGAATGTTT TCCTCTAAAAC AAATAAAAGAA AAAAAAACAAG AAAATATTTAT AAAAATGATCC TAAAGTAAAAT ATTATAAATATT ATCTAGATTTGT TATCTGTGATAA GTAAGTGTGAC GATAAATATGT TATTCTTTATAA GTATTTGAAAA AAAATACGAC		GQSGKISKVFVAQKQV MRTMFNLKYNETCRNIF KSNNFLTVTGLYIYRLLL FFFKQNLLFNVTNDHKY DTRKVDLEYPRHKLTLT EKNSYYAGIKLYNSLPV HLRTPCNLNLFKKKIFD YILDLEPYTLNEFYNRQ VF			
NeSL	NeSL-1_TV	.	Trichomonas vaginalis	GGGTGAGTAGTCTAGTGGT	TAAGAAGAGATAAGACGAGTGAGAAGAACAGAAGCATAGTAGGATTGGCAGAGCTTAAGCGATGTCACTCGGTACGAAACG	MIPVLGTGGPEKLPLQSYVYCGNTAITDSFTPTAKTILKPEEQNLDIVLKNIAALNPENYSDLIRSLSKMEFRLDYPKEIENYWISEKLFSQSIASLPISLLVASM	cd01650	1049	1310

					TGTACCAAACACCG	FSPEDRDLSTEPFHCNAD			
					GATTCCGTGCTAGG	GCNFHCDNCERMVEHIR			
					AATCACAAGCCAA	EHHNTDPMINTFETTED			
					AATAAAAGAGACA	TFRRITAIKIDKTGIEELN			
					CCACGAAAATTACT	PLKYRCSYCDELFTEAE			
					CACCCTCCCTCAAA	DHAIHMISHLTEKLSPDI			
					CAGATAATAATATT	SFFFNDILRLYKTIDKPT			
					AACCTCCCATCCAT	VQNLFPETQVAIFDTLEE			
					CAGTCCGTATGGTC	TNRFRLIVGREAIETIEEA			
					TGATAACAGACTAG	FPPSPPGTDRKPSIIITDTC			
					CACCACATCCATGA	QLRFVPCMDEPPKGDLG			
					TACACTCATTGGAG	ILTLLLRDFSAHNIPIKSL			
					TGAAAACCACCAAC	NNKELIADKDIDYSPDFV			
					AACAAATCCACCTA	EGALANAEEHDTTNSQN			
					GACCAAATCCTGCC	NNGRYINSAEKLTEFLIQ			
					CCACCTCCACCCAA	CEDYLTNIKTLEDLERFY			
					GTAGCTCGCTTCGC	TTIKDYRVNKEVIAEDTP			
					TCGCTCACCTAAAA	IFVYFLVEEGKLPKPGLR			
					CTTTGCTCGCTCGC	CPLESYEGHEDKAFESL			
					TTCGCTCGCTCGTC	RKLCDHFKGEIAKTSFDP			
					TTAACCCTTTCCGA	KVHTIDIWVEFLAQAYG			
					ATAAACACTTACAA	TGTFVYKDENGNIDLDT			
						HVFKCPYADCSYTNNDR			

					TTCCCGGCTCGCCC CATTTTTT	SKLMDHMKTKKHAKNV YIERYGFFWGIVIEGVNR PKGIVYPTLKDIKEHACR KCPEAGCNTYVTELSDI KEHLKKKHKSTTAGVD GEIAHTDATYCWITKEE LDALHAERARERAEQV DNTPVQQIINADNNEEN NENQEDNGNNEEADAL DPPNNTTETEDEAVHAV IINPPATEEEEVAIIAEAR RNIPELQQAEERGCVTP KMTSLVRLKLLKGGGEL FNKKLTPLATRYAATGN TEADKIKVDYLTLKCNA ALREMIYTNNHSESKFM TAENGEDTAPPPRISEDT RDRIQKAANEIKGTLIKV VKHISHARCLKDSTRDD EHNKFVEMIAKIKNDLR DNKFEQYNIEEIFQGPIS DQSILNIVNTEDNNEFIK KMDYINRILGTPQDASP			

						YARKKLQACFADNPTKT LRNIILADKVPQQSLKPS EYLDYYGPQWANEAEG YENFLHHDYALPERYGQ VFANDFLDFMTNESKIIE VIRNKNHLSAHGLDGIP NSVYMLFPVSAAKFLSIL FRSIIISGHIPDCWKLSKT VMLFKKDDPSLAKNWR PIGITSCTYRIFMTLVNK ALQMIPMFHAMQKGFV RGATLSEHIAVANEVLC QSTRTQSEMFQTAIDFTN AFGTVPHQLIFDSLEAKK VPDSIINLLKDLYKGART AIYTRHAHSEIVPVRRGV IQGCPLSPILFNCCLDPLL YAVQRRHFEDGYRFQD KAGQYSIAIQAYADDVL VISPTHEGMQRILNTVDE FQKIAKLKVAPQKCVTL AKTSTAIQPFRIGPDEIPI KTSMDNITYLGIPISGTK			

						TSRFAAATGILEKVKAQI RVVFASHLALSQKIIALR VFILPQLDFYMFHNVFR VNDLKATDQMIRGLIDK EAPTSNIPVSFFYMPKNK GGFGLVKLELRQPQLVL TKFARLWLSQQAETKAF FHTMAQEEKSFRKVVED QENGFLGIKMENGKIVQ KNERSKRTNCFITQAAK AADKLEVRFKEWDKGG IQVRGVGENATDWYRS KHIGQISPLIGRVIQQRQ YEEFKKDETHSHTFCEP AALAESHDIMKRPQAVP NNLYSAAIALRTNTAPTP ANMHFHNPEVLANCPLC GCQSCTLFHTLNMCRNR FSLYKWRHNIICDDIYQF IHDHYPGVTIKCSARITS DGYQTTGPELDDTVKDL LPDLVVYDEANKMIKIIE VTCPYGTDNNVGNSLD			

						AAYDKKVNKYKSLAEQ TERLFNWTTTLSIIVVSS LGVIPLRTKLDALRISPA DHIQLLKRLSMHAIAAS ACIVFEKVPEFFGMRCRP LPGRVTAPNAAIPPNNN ENNNDTDHGQENQQAT SEEQPTNNGNAQEDNGQ GEQINNSTEQTISVDQIIE EDAENNAIEQALDQPDE DEFLN			
Ingi	Ingi	.	Trypanosoma brucei	CCCTGGCG	TAATAAACGAATAA CAAAAATCAAAAC AAAAGGATTGCTAA TTGACATCTTTTGG AGAGTCCGGGGTG GGGGGCTTCTCGCC CCATCTGCTGTATT CCGTTCAACTGCGG AGCTACAACAAAA ATTATAGAGGGTGT GTTAGGCTGAATAA AAAAGGGAGACTC	MPATSTWCQGPVPRIIG GSQEPAAFLSWGTLLCS GYGIIQHRDQQRLAGTP FFICRSLGTCQRAISSIIRT KMLLSGDVEENPGPSLR GMQWNCAGLSQGKRLA LHKTLVDERIAFCLLSET RMTPGEAACFSVAGYQ HHGIARNCKGGGVSILV REDLPVETGMAVVGRIE QVHATIHLARGTALTVT SAYIPPKHTFTATDLDTL	cd01650	550	805

					TGCCACAGTCGCCA GACCGATAGCATCT CAGGGCTCTACGGT GATGGCTGATGGCC GCGCCAGTGGGGG GAAACTCTCACGAA GGCACGAAGAAAA TTCCAAAAAAAAA AA	LTTDGAQLIGADANAHA LAWDRASPPNTKGETLT QWCIDNQFLVCNTGECT RYARHHGESTPDVTLSR NCTVYTWTSLYSPDSDH HHIFFDVIVGDDTDALSC PRLRKPMYAWLKADWR NFRLKVDELCRKIGREK NVNTLEQKLSSAIRIATK VSVPRGCRATPPHWTPE LAKLDEEIAGCGPSHRR EKLVATRKQILDRTTKK RWSTLCSRLAVSDRCSW HIVKKVYAPRPLTTPAV LVDNAAITDYRQAERFS KLYSSRARRHPDSHPPA PIKTIASEFSPITMAELRR SIKLLPSGSAAGPDCLYN EALQHLGRTALNVVLRL FNESLRTGVVPPAWKTG VIIPILKAGKKAEDLDSY RPVTLTSCLCKVMERIIA ARLRDTVESQLTPQQSG			

						FRPGCSTLEQLLHVRAA LCRPTHQYRTGAVFVDY EKAFDTVDHDKIAREMH RMKVSPHIVKWCVSFLS NRTGRVRFKEKLSRSRT FERGVPQGTVLGPIMFII VMNSLSQRLAEVPLLQH GFFADDLTLLARHTERD VINHTLQCGLNVVLQWS KEYFMSVNVAKTKCTLF GCIERHPLTLQLDGERIG ADRTPKLLGVTFQCLQG MATHAAETRRKMDFRL LQIAAISASTWGPRRQVL RAFYLALVQAHTMYGIE VWYWDASERSRDLLAA AQHKASRIIAGIPHGTRK EDSLLEANLLPLKTTTLV RSMKFMLMCESRGGCL RRSAEEVYHSKHPVRAL HSRIMRSYPHLRIEPREH PLETSTLRHSCRPLFHTQ IKPVCADDPDDVKREAS			

						EKWIARHFARRGKEPPR REHYELWTDGSVSLGEK SGAAALLYRNNTLICAP KTGAGELSCSYRAECVA LEIGLQRLLKWLPAYRS TPSRLSIFSDSLSMLTAL QTGPLAVTDPILRRLWR LLLQVQRRKIRIRLQFVF GHCGVKRNEVCDEMAK KAADLPQLRDTWIPDIIA YAKRVLRSEEVHENTHR FGITGNHFPTKHKEELTR EEETALARFRVGSSRHY GWMLRKINPSVPPQCR WCNPQHAAIGPTIQTAP TVATRTLQRTSEPTKCTE CDATYQCRSSAVTHMV NKHGFVRADALRRIKYG DATPAVDIPPEPPPVVAI VPLPSSTRVPMRPQVLH CTLCASKFAVPGRLLHH LRTIHGIGSGSCRVKRGR ENEDSXQGDGRAPAAPA			

						SQDTRKLPFQCDLCEAS FGTRSSLSLHKKFKHKSI VTEDGTVVVVQFPRKRA REGNVDVPGKGEVQCG VCQKVLSCRDSLIRHCK AFHKGEGVELKCSKSTK LCATDSPHTNTSMLVCP TCGRQCASKTGLTLHQK KMHGMKVERAVTSQRG DCEETSLHLMSCPGLKE LRVRFAVEGVCVQDVCF SKRLAQFLIAVERSRPQV KSSPKVTVIQPTLPSATS PLIAECGASRKSTGRRNS PDTLRDRGGMQSTSNRN RKKRGRE			
Rex 1	Rex1-1_XL	.	Xenopus laevis	CACCCTCCCCCA GGGAAGACACA TCTGCAAGGGC TCCCAGGGCAC TGTTAAAATCTG CACCCATCCGT GCTTTACAAGTC	TGACCTTACTATAT ACCACAACGTTACA CACCCTTGTATAAA TAAATAGCCATTGG ATACAATTGTTCTC TATGAGCACATCTG CACTTTATCATGTT	MDVGTTKVQVYNRQTL LRYSNHAYLTIKDGLLH TLRDLGLLREPDLLTAV SPGANGGRRSTRKRCTR TRKRGKRAGVRARLRT NPSRPALPSILLSNVCSIE NKMDYIRLQQTTQREFR	cd01650	563	816

				CCCTTCTGCCAA	CTTTATCATGTTCTT	DCCMFVFTETWLSERVP			
				CATCTCCCACAC	ACTACTATCATATC	DAAIQLDGLVAVRADR			
				CAGCAGCCTAA	ACAATGATACACCC	VAALSGKNRGGGLCVYI			
				AAGGGACATCC	ATCATGTCTGACGT	NTKWCRNSVLVSSHCSP			
				GCTGCCTAAAG	TTAGCATTATTTAC	LVEFVVVKCRPFYLPREI			
				GAGGAAAAACT	TTAACATATTACAT	ASIHIIGVYIPPSANTKDA			
				GGCTCTCTAGC	CTGCTGAGTGTGCA	LWELYSAITDIQNTHPD			
				ACACACAGAGC	CTGTCTTGTCCTGT	GLIIIAGDFNHANLRSVL			
				CGGAGAACAAA	CCCTGCACTGTGCT	PKFHQHVEFATRGPNTL			
				ACGGGGCCTAC	GTCCTGTCTTGCAG	DQVYTNISSAYRAEPRP			
				CTCCTCGACAG	GAGTTGCATATTTT	HFGYSDHISVMLIPAYRP			
				AAGCCGGGGAC	TGCACTTGCACTTT	LVRHSKPVLKQVKTWP			
				TCGAGTAGTAG	TTGTCTGTTGTCCG	AGAMTALQDCFECTDW			
				AGGTACCCGGG	GTACATCTATGTTG	NMFREAATYSDYINLDE			
				ACCGAGCTGGG	TGTGTAGCACCATG	YTSSVISYITKCIDDVTVS			
				GCCTGAACTGC	GTCCTAGAGGAACG	KCITTRLNQKPWFTGEV			
				AGGGCAGACTG	TTGTTTCGTTTCACT	RVLLRARDSAFREGDEV			
				CTGGCCTGGTG	GTGTACTGTACAAA	ALRTARAKLSRAIRVAK			
				ACGTGGGGCAC	CGTATATGGATGAA	RAYTQRIHSFFQENGDT			
				CTCGGGGGAGA	ATGACAATAAACTC	RRMWQGIQAITNYRSTP			
				CTGTGGGCCTA	ACTCTTGACTCTTG	SACDSDASIADALNDFY			
				GTGGCGGACAA	ACTTG	ARFEAQNNTGMRKTITL			
				GATGGCGGCGC		PSDQALRLTVADVRMTL			

				GACAGTAGATT GCAGCGGCCTC TCCGGATCACA AGGTGCTTTCTT TTTTGTGTTTTT TGTGCTTTTTTA CTTTGTTTTTTG TGAAGCT		ERVNPRKAAGPDNIPGR VLRECADQLADVFTDIF NISLNTSTVPKCFKTTTI VPVPKKSSVSCLNDYRP VALTPIIMKCFEKLVKK HIKTLLPPALDPLQFAYR QNRSTDDAIATTLHLAL SHLDKKDTYVRMLFIDF SSAFNTIIPQHLIEKLSLL GLNTSLCSWVLDFLTDR PQSVRIGNIISRTTTLSTG APQGCVLSPLLFTLLTHD CAATHRTNHIIKFADDTT VVGLISNNDESPYREEV QRLTDWCRANNLSLNV DKTKEMVVDFRKTRSD HLPLYIDGSNVEIVKSTK FLGVHLAENLTWSHNTS YLAKKAQQRLYFLRKLR KSHLPPSILKTFYRGTIES ILSSCITVWAGNCTVME RKTLQRIVKTAEKIIGVS LPSITDIYTTRCIRKATSI			

						VADPTHPSHSLFTLLPSG KRYRSIRALTTRLCNSFF PQAIRLLNTQGPDRSLSH SHSI			
L2	L2- 3_XL	.	Xenopus laevis	CCCCAGACCTG TGTGCAGCAGC TCCCAGCTTTTA CTCTCCCTATTT ATTAAGACCTCT TTAGGGGAGAG CATTTGCTGGG GTGAGCATTTG CAAGGGTGCTA TTCGAAGTGATT CATTTACAAGT GATCTGGAAGA GTTTAACAAGG AGTTATGAATT GACATCTCTCTT CCTCTTCTGTTT ATTACAGGTAA ATCTGCTAAAA CTGCTATTCCCA	TGAACTGGCACTTA TATAACAGTGTAAC AAACTGTAACCCAC AGCACCTTAATACC CCTCTGAATTGTGT CTGTATGTTACCCT CCCATTTAGACTGT AAGCCCTACGGGGT AGGGTCCTCTAGCC TTTTGTTTCCTTGAC ACTGAGCACTTAAT CTGTATTGTAATTA TATTTTATATTTATG TGAATTGTATTTCT AATAATGCACTTAT TGTTACTTTTTATTC CAATGACCCCTTGT TTGTTACTACTAAT TTATTGTTTTGCTGT	MAKLIIFAVFACMVLPK VVTSFNRPIVCQYSIYISS SLLPSTYSSSYTLYNYLQ HLCPPVTNKTRICRHFKS SSHIFSLFYLLLLAAGDIS PNPGPCSIPTYIQPRFPAN LRKSSRSSTFSNLIQIPTL PSSCLPFSCALLNTRSLT NKSTLIHDLFALIHLVFL HQKPGSHRMIPPLRLHSL MAVSLSLTPPEPLDVVG ELEFSSRCRFKPIPIPPAF SFNSFEVHAIHIFHPLSVR VAVIYRPPSASSSAAFLS DFDTWLSFFLCTDSPAII LGDFNFHIDDASQPWPS RFLRLTSSFELQQWTNTP THKDGHFLDLVFTKKLS LSGFNSEPFPLSDHNLVT	cd0165 0	664	922

				TCTCTGCTGTGG	ACAGTGCTTTGCCC	FSISHSPCPPASSPTVLVR			
				CCCTGTTGTCTG	TCAAGGAGCGCTAT	DTRDIDLPALAQSFRSSL			
				CATTTTTGCACC	ACAAATAAAAATAT	ISLKEMSDPDTLVSEYN			
				TTTATTGGTAAA	ACATACATACATAC	HILSSTIGSFAPLQAKRT			
				TCTGCTAAAACT	ATAC	RALNPRPWLNSQTRFLR			
				GCTATGCATTCC		SCTRSAERMWKKSRTQ			
				CTTCCATCTCTG		ADFIHYKFLLSCLNSALS			
				CTGTGGCCCTGT		KANKSTTVPLIHKPDSY			
				TGTCTGCATTTT		APVRDLLNVCGRNPVRK			
				TGCACCATTTAT		QTLFTINFYCPASILPCQ			
				TGGTAAATCTG		KPTRVLQYPYKQSQIQP			
				CTAAAACTGCT		TTPVFSIQYFLHPSSTVS			
				ATTGCATTCCTT		VNQTHSPQDFADFFKDK			
				CCATCTCTGCTG		VESIRNQIPTSTNVNQLL			
				TGGCCCTGTTGT		LPKPPSACLNSFPPVTMS			
				CTGCATTTTTGC		EVSKLLLSSPLTTCPLDP			
				ACCATTTATTGG		MPSSLLKHCVAELTPLL			
				TAAATCTGCTA		THIFNSSLASGTFPSPFKQ			
				AAACTGCTATTT		ACVKPILKKATLDPACL			
				CCATCTGCATTA		SNYRPVSLLPLASKLLER			
				CACCTACACTGT		IVFSRITNFLHTHDLLDP			
				TTAGTGAGTTAT		LQSGFRPGHSTETALCR			
				ATTTATGCATTG		VTNDLQVAKAKGHFSIL			

				TTCATTTGCCAC		LLLDLSSAFDTVDHSLL			
				TTCTAAACTAGT		MQILYSLGIRNQAASWL			
				CACATAATATT		SSYLSNRSFTVSYDNKTS			
				GATTATAAGTTC		SPVPLNVGVPQGSVLGP			
				TCCCTGTAGTCT		LLFSLYTLSLGDLIHSFG			
				TGTAACTTGTAC		FNYHLYADDTQIYLSTPS			
				ATTTTACGACCC		LTAETEAQISNCLLAISN			
				CCTATAAACTCC		WMNQRHLKLNLTKTELI			
				TCTCCAAACAA		IFPPKPGPTPLFTISIDGM			
				ACTTTGTAAGTC		LINPVNSARCLGVIFDSS			
				AGGTTATCTCTG		LSFSDHINTTVKTCHFFL			
				TATTCATCCTGA		RNIAKIRPFLSPDTSKTL			
				GTCTTGCTCTGC		MHALILSRLDYCNLLLT			
				TCTCACTCTCTG		GLPNSHLSPLQSVLNTA			
				TCACAGCTGGG		ARIILLSSKRVQAPPLLK			
				TGATCAGTTTGG		SLSWLPIKQRITYKLLLIT			
				TGTTTGATCAAT		FKALHSSAPNYISSLVSL			
				TAGAATCAGTT		YVPARNLRSSQSNRLVV			
				GTGATTCCCCA		PPTTTAVSRIKPFSLATP			
				GACCTGTGTGC		YIWNAIPESLRRESSFNV			
				AGCAGCTCCCA		FKTKLKDYLWEHLDNT			
				GCTTCTACTCTC					
				CCTATTTATTAA					

				GACCTCTTTAGG GGAGAGCATTT GCTGGGGTGAG CATTTGCAAGG GTGATTCGTTTC GAAGTGATTCA TTTACAAGTGAT CTGGAAGAGTT TAACAAGGAAA GCTACAAGGAG AGACAACACAG AGGACTCCATT AAGGACATAAC ACTGCTACCGG CTAGAAACTGC A					
Penelope	Penelope-5_XT	.	Xenopus tropicalis	TTGTTTTAGACT TAATTCTACTAG	TAACACCTTCATCA ATGAGAATCTTGGT ACCATTGTGATTGG ATCATACCTTCTTA CACCTGTCAGTTTC AGCTAACACCTAAC TGAACAAGTTCAAT	ILYIMTWMNENLHRHIA THFGENTLKLVREYERT ARKLADYRNHLRFNLRC RYQGITPCSLRLGSSVKG HRAKVILQKAQKQLLNE RVRQTNFSIEVLVHKFD NAKQRLTAKLPVLTLQR	cd00304	482	563

					GGAACCATTGTGAT TGGATGTCTGTGAC TCTACTACCATCAA GAGTTTAAATACCG GGGAATTCCCTACC AGTCATTTGAACTG AAGAAGCCACTCG GATGAGTGGTGAA ACGTTTTCAAGAAA AACTCAGAAAAGTC CAGTTGTTTTAGAC TTAATTCTACTAGA TACTGTATATCATG ACCTGGATGAATGA GAATCTTCATAGAC ATA	VVEFTERAQLAQHAKG KERQINKFTSLLSRTNNT QRTDKSTGRKEDENKTC QNIKDLWVKNLSDRELT EPEKDVLAKGLNFAVAP RYVPVVDFITATESSIHN NKIPVDEAENIRLKVSAA LANAKAPPSNLSLQERR ALTSLAKDSSVTILPADK GRCTVVLNTSDYHAKVS TLLNDSDTYEQLKRDPT SNYKKKVIDCLQHLEKE KVISPALYRRLYPGEATP CLYGLPKIHKEGAPLRPI VSSINCVTYNIAKYVANI LAPLVGNTVHHIQNSMD FVKKVKELKLEIDDTMV SYDVTSLFTCIPTAEAID TVRKRLKQDTTLSSRTK LTPEQVCSLLDLCLNTT YFKHKEVFYRQKHGCA MGSPVSPIVANLYMEEV EKKALDTFKGTTPSHWF			

						RYVDDTWVKIKEREVPA FTKHINSVDKNITFTRED VKDSKLAFLDCAIVIKEG RDLDIEVYRKPTHTDQY LLFDSHHPLEHKLGVIRT LHHRAETVASNAEAKEK EYKHLKGALKTCGYPD WAFIKTKSKTNRKTRPT NRREEHSRRNNIVIPYVA GTSEKLRRIFNKHRIPVF FKPSNTLRQKLVHPKDP TPKHMKSNVVYAVQCS EECSDLYIGETKQPLCKR MAQHRRANSSGQDSAV YLHLKEKGHSFEDSNVH ILDREDRWFERGVKEAI YASLEKPSLNRGGGLRH RLSPTYNGALTSLPRQFH NSSHFQSCTLKD			
Penelope	Penelope-6_XT	.	Xenopus tropicalis	A	TAACACCTGCTTTG AAGAGTTGCATGAT ACTTGGGTAATCCT TTCAAAGTAACTCC	KGHRASNILQKAQKHLL NERVRQINFTIEALKQKI EQLKQSLTKQLPDVTLE RVVQFIEHAQMAQHIKS	cd00304	413	496

					ACAGCATTCACACC	KERQISKFTSLLSRSSGL			
					TCTGTGAGTTTCAG	KRTEELTWRQKDETSKN			
					ATGTCACCTTTCTG	TKDTWVKNLSDRVLTQ			
					AAGAGTTACATAGT	PEKDVLAKGLNYAVTPQ			
					GCCATTGTGATTGG	HIPVVELITATESAINNN			
					ATTAGTGGAAGAGT	HIKVEEAEQLRLKVSAA			
					ATATATGCTGAGGA	LSSARAPPSNLTTQERRA			
					CTTCCCATACCAGT	LTSLSKDPNITILPADKG			
					CAGTTGAACTGAAG	RCTVVLNTTDYHCKMT			
					AAGCTGCTCGGATG	SLLSDSNTYEPLRRDPSS			
					AGTAGTGAAACGTC	SYKKKVIDCLQQLEKEE			
					TTCAATGATTACTA	AIDRALYHRLYPREATP			
					AACAAGTCCAGTTG	CLYGLPKIHKDGAPLRPI			
					CTTTAAATTTATTT	VSSINSVTYSIAKYLANI			
					ATACTAGATAT	LAPLVGKTVHHIQNAKE			
						FVTKIQGVKLEAEETMV			
						SYDVTSLFTCIPTTEATE			
						TVRNRLQKDNTLSSRTK			
						LSPNQVCLLLDLCLNTT			
						YFKYKDQFYRQKHGCA			
						MGSPVSPIVANLYMEEV			
						ERKALLTFNGTTPSHWF			
						RYVDDTWVKIRSNEVA			

						AFSEHINSVDNNIKFTRE DVHENKLAFLDCLISIQE GGILKTEVYRKPTHTDQ YLLFDSHHPLEHKLGVIR TLHHRAESVTTDTESKD KECKHLRGALKACGYP DWAFVKTRATKPNRNT KRNNRPEAERRRNIVIPY VAGVSEKLRRIFNKHHIP VFFKPSNTLRQKLVHPK DPTPKEKQSNVVYAVQC SEECTDLYIGETKQLLSK RMAQHRRANTTGQDSA VFLHLKDKGHSFENSNV QILDKEDRWFERGVKEA IHVKVEKPSLNRGGGLR HHLSATYNAVLTSVPRQ FQNSSHIHSCNSNK			
L2	CR1-L2-1_XT	.	Xenopus tropicalis	AGTCCTGTGTTC AGGAGTGGCCC AGCTTGCTCTCC CTATATCTTAAG AACCTCATTTGG	TAAAACATTATTTT GCCTAGTCCTGCGC TTAAGGGCAAATGC CCATACCTGGTGCA CTCTTACCTTCCAA	MSLLLLVLSFMFVLPKV TSSFSPPADCPYSITISSSL LHSTHCCSYNLYTFLKN LAPPTNTSQKHNNFKSFS HLLSLSLLLLLAAGDISP	cd01650	616	874

				GAAAGCAGCGC	TTTGTGCCTGTATG	NPGPCRIPISYRPRNASLL			
				AAGGGTAGCAG	TTACCCAACCACTT	VKPSQTANLSNLIQIPTL			
				CGCACAAGTGA	AGATTGTAAGCTCT	SSSYLPFSCALLNTRSVT			
				ATCCTTCCAAA	ACGGGGCAGGGAC	NKAALIHDLFCSKSFHLF			
				GTGAATCATTC	CTCCTTCCTACTGT	AITETWLSQNDNAIEAA			
				ACAAGTGAACT	GTCTCATACCACAT	LSYGGLSFSHTPRITGRG			
				GGAAAAGGAGT	GGCACTTATTCCCT	GGVGVLLSPHCRFKLIPI			
				TTAACAAGGAG	GTGCATTTATATAT	PPTFRFPSFEVHAVQLFH			
				TTAGGAATTGA	ATTTATTGTATTTAT	PLSVRVAVVYRPPSSKSL			
				CACCCTCTCTCT	TTATTATAACACTT	ATFLSDFESWLSFFLCSD			
				TCTGCTTATTCT	GTCCTCCCTGTGTG	TPAIILGDFNCHIDDPSQP			
				ACAGGTAATTG	TAATTTTGTATTTTG	WPSRFLHLTSSFDLRQW			
				CTGCACTATTTC	TAAGATTGTACAGC	TNTATHKDGHFLDLVFS			
				CATCTCCACTGT	GCTGCGTACCCTTG	KNLDLSAFNSEPFPLSDH			
				GGCCCTGAATA	TGGCGCTTTATAAA	HLISFSISHVSPSPSTNAQ			
				CTGCATTTTTGC	TAAAGTTATACATA	PKTLIRNTQRVDITALSN			
				TAGTTTATTGGG	CATACATACATACA	SLSSSLSSFCASSDPELLV			
				TAATTGCTGCAC	TACATACA	NTYNSILSSAINTHAPLQ			
				TATTTCCATCTT		PQHSRAHNPRPWLNPQT			
				CACTGTGGCCCT		RFLRSCARSAERIWRKS			
				GAATACTGCAT		RVKADFLHYKFLMVCL			
				TTTGCTAGTTTA		NTALSQAKQVYYNTLV			
				TTGGGTAATTTA		NNNKSNPRRLFSIFNALL			

				GCTAGTTCTATT		HPSPNESIISEHTPQDFAD			
				TCCATTCTCAAT		FFKSKVDSIRKQFSQPSD			
				TCCCAATATTGC		TGNLNLPKSPLSVFSSFN			
				ATGCCCCCTATT		LVTESEVSQLLRSSPPTT			
				CCTGTGCTCCAC		CSLDPMPSSLLKQCAPA			
				CAATTTGTTGAT		LTPALTHIFNSSLASGTFP			
				TTTATGACCCCC		SLFKQACVKPILKKATL			
				ATCGTAAACTC		DPSCLSNYRPVSLLPLAS			
				CTCTAAACAAA		KLLERLVFSRVTNFLCT			
				CTTTGTAGTTCA		HNLLDPMQSGFRPAHST			
				GTGTATCTTTGG		ETALCRVANDLQTAKA			
				GCTGATCTTCAC		KGRYSVLILLDLSSAFDT			
				TTGTGCACTTCT		VDHSVLMQILYSLGIRD			
				CTCATCCTGGG		QAASWFSSYLSNRSFSV			
				ACTCAGCTGATT		ALANKSSTPVPLSVGVP			
				GATTTAATTTGT		QGSVLGPLLFSLYTLSLG			
				TATTTAACTAGC		DLISSFGLKYHLYADDIQ			
				TGCAATCAGCT		IYLDTPALTTDVQTQIGN			
				CTGATTCCCCAG		CLLAISSWMNQRHLKLN			
				TCCTGTGTTCAG		LAKTELMVFPPKPGPSPP			
				GAGTGGCCCAG		FTITIDGMTINPVNSARC			
				CTTGCTCTCCCT		LGVIFDQSLSFSNHINNT			
				ATATCTTAAGA		AKTCRFFLRNIAKIRPFL			

				ACCTCATTTGGG AAAGCAGCGCG GGTAGCAGCGC ACAAGTGAATC CTTCCAAAGTG AATCATTCACA AGTGAACTGAA AAAGGAGTTTA ACAAGGAGTTC AACAAGAGGAG CAACAGGAGGA GCAACAGCAGA AGACCGGCCAG CCTCTGTCCATT TGAATCGGGAC CCAGTTTACTGG ATATACCTGCAT A		SQVTAKTLIHALILSRLD YCNLLLTGLPDSHLSPLQ SILNSAARILLLSPKREPV QPHLSSLAWLPVKQRIA YKILLLTFKALHSSAPHY ISSLVSLHVPGRLLRSSQ SLRLFTPSTPTALSRLKPF YLAAPYLWNSIPESLRRE HSPTLFKKKLSCYLLEH			
L1	LINE 1- 3_Z M	.	Zea mays	GGGAAACCATA ACTGTGGAACA ATTGCTTAATCC CTTTGCTAGTGT TCACCAGGTTC	TAGAGTTTTTATAG ATGAGTCAGTCTAT GCTAGTTTCGTTCA GGCCGCCGTAAGGC GTCAGTTGTACCCC	MDPEKIIVWNVRGLNSL ARQNSVRELLNCYGADI ICLQETKMAQPSSRVLLS MLGSDFTGSLALPSNGS RGGILIAWRRHLGITGAT	cd0165 0	500	762

				ATCCTGACACC	GGTCTATTTTAGAC	RTDIHSISVQFCSVEGHQ			
				CTGGAGCTAAA	CTTGTTTATTTCTTC	WWLTCVYGPQSNEAKIE			
				GGATGTGGCTG	TTTTAATATAATGA	FLQELRHIRAACQGPWM			
				TTTACCGATGTT	TACGCAGCTCTCCT	ITGDFNLILHAEDKNNSH			
				CTGCTTGGTGTT	GCGTGTTCGAGAAA	INRAMMGRFKKLVDDL			
				TAGACCCGTCCT	AAAAAAA	ALKEIPLHGRKFTWSNQ			
				TGGTGCCTGCTT		QNQPTLVKLDRVLCSID			
				CAAGGGAGCTT		WEELFPGAVLQSAASED			
				TGGATCGTCGA		SDHCPLLLGFKDCKGVK			
				ACCTCCCTTTGC		RRFHFESFWTKIEGFHEV			
				CGTTGATGGGA		VHNAWNSVPVGHCPFST			
				TGTCTTCTGGCA		FDLKLKGTAKALQGWS			
				GGAGAACGCTG		SKKVGHIETQLHLARDLI			
				ATGTATCCAATC		HQLEIAQDGRILSVLER			
				CGCATCACATTC		WLLSKLKKLALALASTK			
				TCGTCTGCCCCA		RTIARLRSRISWLKEGDA			
				CGGAGGGATGG		NSKLFHSFARYRKKKNS			
				ATCCGATCCTA		VSKLTVGNQVLTAHSDI			
				AGCTTCATGCTA		AAAVDQFYFDLIGASKD			
				TTACTTCGGGCG		RNISINLEALETAGTPFP			
				GGGATGGTTCT		NLIELDSPITEQEVCNAIN			
				GGTCACGTGCA		SLPSDKSPGPDGFTGRFY			
				GCAGGGATCCC		KICWSIIKEDLMAAVIAI			

				CTCCTTTTTCAA		WNRKFDNFHKLNSAFIT			
				GACCTCCCCATT		LIPKKDNADQVKDFRPIS			
				CGAGTGGGAGA		LVHSFAKLITKILASRLA			
				AGGGGAGAACA		SRLNDLVSPNQNAFIKG			
				TGGTCATTCGG		RFIQDNFMLVQQTAKLL			
				GACGTCTCTCA		HQQRHPCLMLKLDITKA			
				GTTCATGAGCG		FDSVSWPFLIEVMQNLG			
				TTTGGGCCCAA		FGQIWRDIISGLLLTSSTR			
				GAGATGAGAAC		VLVNGHPGSIISHRRGLR			
				ATGCAGCAGCG		QGDPLSPMLFVLVMDVL			
				GAGCATTAGCT		GQLVSKADEEGLLQPLS			
				CCGGGGTGGCA		TRQLQHRVSLYADDVV			
				GTTTCAGATTTG		LFINPVAEDMTMISDILQ			
				AGGAGAACTTT		LFGEASGLYNNNQKSSV			
				GGAATCTGATG		YPIRCHEDNLEIIQQFWP			
				GGGGGACCTTG		CEISSFPCKYLGLPLSLH			
				TCTCCTAGGCTG		KLRREQVQPIVDKIANQ			
				GTAGAGATGAC		LPGWKADLLTKAGRSV			
				TGATGACATGG		LVQSVLTGVLIYSVMAV			
				TTCTTGTTTCAC		DLPSTTLKDIDKIRRGFL			
				CGGTGAAGGCG		WRGRKNAMGGHCAVA			
				AAGCAGCTGAA		WGSVCRPRELGGLGIFSI			
				AGTTTATTTTCG		TELSWALKMRWAWLQ			

				TCAAAGATTCA AGAATGTCCTG AACCAGGAGGA AGCTGGAACTC CCAATTCTGTCA CTGTGCTTCCTC CATCCCCCACA GACGTGGCAGC TGCTGACACAG ATCCATCTGAGT ATCCTGTTTCAG ATGTTGACGCT GTGTTGGCAAC TCCGTCTAGAA CATTACCTCCCT CCCTGTCTGATG TGGCAGCAGCA CACGCGGCTCC GTCGGAGGCTA ATGCCTTAGAC GTTGATGCTGA GTTGGCTGCTCC ATCTGAGGAGG		KTEPGRPWASLPLHTPK KIREFLQVALYSEIGNGA STLFWSDRWLDGQRIGD IAPRLIETIPMKLINKRTV QEAIEDNGWINDIPGSLS VGALADFLRIWDLVAH VDLDPHKEDKHIFRLAA NGKYSTKAAYEGLFIGS VEFEPFERIWKTWAPPK CRFFLWLVAHKKCWTA DRLEKRGLDHPEKCPLC EQERETIDHLLVNCVFSR ECWFLLLRQFGLQGLAP QPTDINFMEWWHQANE AIQGSFRDGLNSLIMLGA WILWNHRNRCIFDGLSP NVANFLIHVGDERRLWE TAGAKGLTSLVASLSHV A			

				ACATTTCAGCA					
				ACAACACATGC					
				TGCGTTGGCTGC					
				TCCATCTGAGG					
				ACCCTCCAGCA					
				ACAACACAAGA					
				TTCTGTGGAGAT					
				TTGTGCTCAAG					
				AGCAATTCATC					
				AGCAAATTGTC					
				AAACCAGATCA					
				CACGCGTATTA					
				CCAATCCCAAA					
				ATCTGTCAGAC					
				GTCAGTCCTCTA					
				TCCCCCTGGCA					
				GTTCCAAGACG					
				CAGCCGACGTG					
				TGGCGGGGGTT					
				GGAGTTGAATT					
				TAAAATTCAGG					
				ATTGGGGAAAA					
				AGATCTAAGAA					

				ACAAGCTATGG TGGCTCTAAAT ATTATTGAGGA CAATGAAGGAT TCAGCCGTCAA GCTTTAGATGA CTACGCCAAGC TCTTCAAGAAT ACTCTTTCTCGT AGTCATGTGCA GGCTCTTGCTGC CCTTTTTGGGTG GTCTGTCCCAG ATAATTTAGAA TATATAACTGG GACTGGGGATA TTACCGCAACA AGCTTATTTTAG CTGGCGATGTT GCTGATATTGTT TGTCGATTTACC TCTCATTGTTTC TAC					

RTE	RTE1_ZM	.	Zea mays	AATGGCAAGGG TCTTCGCATCTG TCTCAACGGGT GCGAGGGGTGG AAGCTAGTTGA GCCGACTAGGA TCCGTGTAGGT AGTTGGAACGT AGGGTCCCTTA CAGGTAAGCTA AGAGTTAGTCG ATGCAGCGAGC AGGAGGCGTGT TAGTATCGACA AAATGGAAGGG GCAAAAAGCGA AGGAGGTGGAG AACACTGGTTT ACTTTGGTACTC AGGAACAGTGG CAAACAAAAAT GAGTAGGTATC CTGATTGATAA	TGATTAGGCCTTTT CCCTTTTAGTGCTC TCAAAAGCTTTTCC CCTCCCCTTTCTCTC TCTTTCTCTTTTTGG TGACATCTTGTTGG GTTTCAACTCTAGC CTACCCCAACTTGC TTGGGACTAAAAGG CTATGTTGATGTTG ATGTTGATGTT	MPLSESEKRKFWEDLDG MVRAVPTNEKLFIGGDL NGHVGSTNVGYELAHG GFGYGSRNQGEDILDFA VAYNLVIANTFFRKRDS HLVTFSSGHRSSQIDFVL TRREDKQACLDCKVIPG ESVVPQHNLVVADFRFR IRTHRDKQAKIARTKW WKLKGETSEVFRERVFV EGAWSEEEDANNMWV KMATCIRKVASEVFGVT KGSIGEPKDTWWWTED VQKAIKEKKECYRSLFH DRSAVNIERYKVAKKTA KRAVSEAKGRAYDDLY RRLSTKEGEKDVYKMA RIRERKTRDLNQVKCIK DEMDQLLVKGQDIKQR WQRYFDNLFNGENETM DTQLDDSFDDLNRCFVR RIQESEVKEALKRMKGG KAMGPDGIPIEVWKCLG	cd01650	409	658

				GACCCTCAAGG		DIAIVWLTKLFNHIFRSN			
				ATGGAGTGGTG		KMPDEWRRSTLVPIFKN			
				GACGTCAAAAG		KGDIQSCTNYRGIKLMS			
				GCAGGGAGATA		HTMKLWERVIEHRLRG			
				GGATTATCCTA		MTHITMNQFGFMPGRST			
				GTCAAACAGTT		MEAIFLIRQVMERYKEQ			
				TGGGGATGTAG		KKDLHMVFIDLEKAYD			
				TCTTGAACATTA		KIPRNLMWWALDKHKV			
				TAAGTGCAT		PTKYVTLIKDMYDKVVT			
						SVRTTDGDTNVFPINIGL			
						HQGSALSPYLFALVIDEV			
						TRDIQGDIPWCMLFADD			
						VVLVDESREGVNRKLEL			
						WRQTLESKGFRISRTKTE			
						YMRCDFGTTISEDGDVS			
						LGGQVVPKKDTFRYLGS			
						MLQRDGDIDEDVSHRIK			
						AGWMKWRQASGVLCD			
						KRVPQKLKGKFYRTTIR			
						PAMLYGAECWPTKRRHI			
						QQLSVAEMRMLRWICG			
						HTRLDRVRNDDIRDRLG			
						VAPIEEKLIQHRLRWFG			

						HVHRRPPEAPVHRGIIRR DNNVKRGRGRPNLTWE EAIKRDLKEWNIPMELC LDRSAWKEAIHVPEP			
L1	LINE 1- 43_Z M	.	Zea mays	GTCTAGTCAGC AGAATCCCCCA ACCCTAGTGAA ACTTGACAGAG TATTTATGCACA CTGGACTGGGT GGATTTGTTTCC AAATTCCTTGCT TCAGAGTTCTGC CTCNAATGACT CAGACCACTGC CCCCTCCTTTTG GGCCTGA	TAGCTAGCATTGTG GTAATTCGTGTTAG TTTCTTCTTTAGNC ATGGATGGGTTTCT GTATGCCTGTTACT GGCCGCCGTAGGCG CCTCTTTTGTTTTCT TCTGTAACTGTCCT TCTTCTTTAATATA ATATCGCGCAGTTC TCCTGCGCTTTCGA GAAAAAAAAAA	MTIVWQEGAGFILSPFGL SLRDFWRLFRWPGILFQ LCSVLLQRLNSKLKATA RGLQAWSAKKVGHISAS LALARVLLHQLEIAQDS RGLSPAELWFLLCLKKH SLALSSLQRTIARLRSRIG WLREGDANTKLFHMHS RFRKKKNFVPKLVSNGA VLTSHDAKADLVNEFYG NLLGQCSDREQTISLESL GIPCHDLSALDAPFTEKE VWDTILHLPSDKAPGPD GFTGKFYRVCWEIIKAD VMAVFSAVGSRRFANFQ ILNTAYITLIPKFEGADQ VKNFRPISLVHSVAKLLT KLLANRLAGRLHDMVS PNQSAFIKGRFIQDNYM	cd0165 0	263	526

						LVQQTTRYLHQQKQPR MLLKLDISKAFDSVSWP FLLEVLQHLGFGQVWR DIISGLLWTSSSQVLLNG LPGQPILHRRGLRQGDPL SPFLFILVMDTLGFLINK AAQDGLLQPLAPRALQH RVSLYADDVVIFLRPAA MDINVILDCLQLFGDAS GLRTNVQKSSVYPIRCG DQDLALLQNMLPCELAE FPCRYLGLPLAIKKLSKI QVQPIIDKIANQLSGWK AELLTKAGRKVHVQHV LTGMVIYLAMAVDLPA WALKAIDKIRRGFLWRG RRDVKGGHCHVAWGIV CRPRELGGLGISSMKEL AWSLRMRWLWLAKTEP SRPWASLAIQVPKKVKD FFSMAMQTELGNGATTF FWTDRWLSGHRIAELVP RLFGIIPKRRVSKRTVVD			

						AISNQSWISDIQGALSVG IINEFLILWDLIDSIVLRP QVEDKHFFRLAANGKFS SKEAYRGFFLGSSEFEPF HRIWKSWAPPKTKFFM WLVAHRKVWTADRLQ KRGMDHPERCPLCDQD QETLDHMLIGCVFAREF WFKLLLQVNLQFLAPQS EDSAFLEWWRTLCTKVS GIARDGLNSLVILGVWT LWKQRNGCVFDNKSPSI ADAIRRVELEVHLWEM AGAKKLSLLTAPIPGIPA S			
L1	LINE 1-15_ZM	.	Zea mays		TGATGATAGGCCTC ATTTGGAATTGTCA GGGTTTAGGGAATA GTTCTAAGGTTCAG TTTCTGAAAGAACT CATTAGAGATGAGA AAGTTGACTTTATT GGTTTGCAAGAAAC	MAKKKRRKAPWAEAPS SPRAPAPVAHGPGGVRS GLVGLPSTVGFAAQGKC APAPFASSVSWTGWSQG ISPLAPTLAHSSRVLLPA ATSDAAVSKSRLGFPPD FVGSHSFPHCAAFLSRL WSLPVSSRSDFHGGWCV	NA	NA	NA

					CCAAAAAAGTCATT	TSLRPPVPEEEWPRLGSG			
					TTAGTGATTCTTGG	VVRSSKNLGEMDPKNK			
					TTGACTTCGCTGGC	NIPNPSWDRWGQKDSA			
					TGGTAGTAAACCTT	QGSQSWNRNRNLSWRL			
					TTGCCTGGTTTTCT	KPQIGKSDPQEGLMSSSS			
					GCCCCGCCTAATGG	GDGIQSGNPILSPSSLKK			
					TAGATCTGGTGGTT	EIDLARPAFCQKCGVEG			
					TATTGGTGGGTTTT	HHARNCFNALWCDICR			
					AATACTGATGTGTT	KDTHVTARCVLPKQSKP			
					TGATGTGAGAGAA	NMPIVGMAADGLGFYLS			
					AATGAAATTGGTGA	HFAKPISKKPKRVFIGLV			
					ATTTATGATTCGCA	KVIEGLVSAEDLAKDFG			
					CCTTGATTTTTCAT	FHFPWGKTWKATKCHS			
					AGAGAAAAAAATC	GFLMQFPSQERLDEMISF			
					TCATCTGGAACTTT	PELKMKMSGAKIVVVP			
					GTTAATGTGTATGG	WSSRAKPKSRLHTAWIV			
					TGCTGCTCAGAAGG	AENVPEELQNYQSICEIG			
					AGAACAAATGCAG	SMIGAVEEVDLMSLDSE			
					ATTCTTGAGTGAGC	DIVRFKVHIKSVAMIPPV			
					TTTCGTCTTTTTGCT	VEVAVKPFLYDIYFRIESI			
					CGAGAAGTCGTGCT	SDEGWNDETINLGKRAS			
					CCTTTGCTTGTTGG	VDIHGINDPVFDKSGKK			
					TGGTGATTTTAATA	PRNDDEVGDEDPXPNFK			

					TTATTAGAAAAGCA	VGESSSQGKESFKKSEVF			
					GARGAGAAAAACA	MDSSKDKSGDGMRLED			
					AGCCTGGTGGTGTT	LSDEKVDFDPNEDDLLS			
					AATAAGTGGAGCTT	SQELEEFAKDMEEMHTD			
					CCTTTTTAACAGTA	FQAKVGEVLSIPLTKDE			
					TCATTGAGCAAAAT	KQDXVGKKSSEILMSTS			
					GGGCTTGTGGAGTT	EGIRKSSRLEKNEDIKVA			
					TGACTTGAACAATA	DKAISRAEAKDAFLNKG			
					GATTGTTCACTTGG	MNNNPFSLLNASNEDLF			
					TCTAATAATCGGAT	DISNKLGVCLGSSDSSA			
					TGACCCTACTTTTG	MANLDLIRDLEMSRKIL			
					AAAAGTTGGATAG	AIQSCGSKKNEEPIVDDI			
					ATTTCTGGCTAGTC	NEVDSSVHADSEKDIDY			
					CTGAGTGGGATCTT	DLLDVMVLRKGRKIKH			
					GCTTATAATAATAT	RKKQQKKASPKIRYTPN			
					CAATGTTCGTGGTT	KKREAENFXLPPDPF			
					TGAATAGATCCTTT				
					TCTGATCATGTTCC				
					TTTATGCATTCAAA				
					CGGATGTTATCCCT				
					ACTGGCAGTAGAGT				
					TTTCAAATATGAGT				
					TATGTTGGAAGGTT				

					AGACCAGATTTTCA				
					CAAAAAAGTGATA				
					GACAATTGGTCTAT				
					TCCTGTGAAGGCTA				
					AAGGTAGTATTGAC				
					ATTTGGAAAACGAA				
					AATTAAACGTTTGA				
					AACAAATGCTCAAA				
					GGTTGGAATATTAA				
					TGTTGAAGGGCATT				
					ACAAAAAATGAAA				
					AAAGAATTGTTGGC				
					AAAAATTGACATCT				
					TTGACAAAATCAGT				
					GAAGTGGTTGGTCT				
					TTCAGAAAATGACA				
					GGATGGAGAAACTT				
					GACCTGGAATTGAC				
					TCTCAAAAAATTAG				
					TTGATGAGGAGGG				
					AATTAAGCTCAAAC				
					AGAGAGCTCGTGAT				
					AAATTCATTTTGGA				

					TGGCGATGAGAATT CTAAATATTTTCAT CTTTTAGCCAAGTG TAAGAGAAGGAAG CTTAAAATTGTGAC TCTTACTCATGAAG ACAAAGTTGCTCAA GATGATGATGAAAT TAATCATTTGGCGA CTTCTTTTTACCAA AATCTGTTTGGTCC TTCTCAAGAATCTA GTATTTCTTTGAAT CATTTGAATATGAA AATGTTGGATGATA CTGACAGAGATTGC CTGACCAGACCGTT TGACATGGAGGAA ATTAGAGAAGTGGT GTTTTCTATGAAAC ATAATAGAGCTCCT GGTCCTGATAGCAT CCCAGCTGAATTTT				

					TTCAGGAGTTTTGG GACACGATTAAAA ATGATCTGTTTAAC CTCTTTCAAGATTT TCATGATGGCACTC TTAACATTGAGAGA CTTAATTTTGGAAT GGTGACTCTTATTC CCAAAGTTCCTAAT GCTATTGACATTAA AGCTTTCAGACCTA TTTGCTTGCTTAAT GTGTGTTACAAGAT TATTACCAAAGTTT TGACCAACAGATTG GCCAGTTGTATCAA TAATGTGATTAGTG ATTTTCAGTATGGT TTTATCAAAGGTAA ATACATCATGGATG GAGTGGTCTCTCTC CATGAGATTATTCA TGAGGTGAAAAAG				

					AAGAAGCAGAGTG				
					GGGTGATTTTTAAG				
					GTGGATTTTGAGAA				
					AGCTTATGACAAGG				
					TTAATTGGAACTTT				
					CTTCATCATATGAT				
					GATTAAAAAAGGTT				
					TTGGTGACAAATGG				
					TGTGACTGGGTTAT				
					GAGAACTGTTAGAG				
					GTGGTAAAGTGGCC				
					ATCAAGACAAATG				
					ACAGGGTTGGTCCC				
					TATTTTACCACTCA				
					CAAAGGTGTCAGAC				
					AAGGTGATCCTTTC				
					TCCCCCTGCTTTTTA				
					ATATTGCTGCTGAT				
					GGGCTTGCTTGTAT				
					GATTCAAAAAGCCA				
					AAGATGAAGGTATT				
					ATTAGGGGACTCAT				
					TCCTCATATCATCC				

					CTAATGGTTGTTGT				
					TGTTTGCAATATGC				
					TGATGATACCATTT				
					TTTTGCTTCAAGAT				
					GATCTGGAGGGAG				
					CCAGAAATCTTAAG				
					TTTATTCTTTGCTTA				
					TTTGAACAGATGTC				
					TGGTCTCAAAATCA				
					ATTTTCATAAAAGT				
					GAAATTTTCTGTTT				
					AGGTGAAGCTGTTG				
					ATAAGGAAATGTG				
					GTATGCTGACATTT				
					TTACCTGCCCTTAT				
					AATTGTCTTCCTAT				
					GAAATATTTGGGGG				
					TTCCGATTGATGAC				
					AAAAAACTTTGCAA				
					GTCCTTGTGGTTAC				
					CCACAGTAGAGAA				
					AGTTGAAAAATGG				
					GTTCTTGGCAAGGA				

					AAATTCCTTAGTTT				
					GGGTGGTCGGTTAG				
					TCTTGATTAATAGC				
					AGTCTTACCAATGT				
					CCCTTTATACATGC				
					TCTCGATGTATAAA				
					GCTCCCAAGAATAT				
					TTTGAAGAAAATGG				
					ATCTCTTTAGAAAA				
					AGGCTTTTGTGGCA				
					AGGTGGCCATACTA				
					CAAAAAAATATCAT				
					TTGGCTGATTGGAA				
					TATGGTTTGTTCTC				
					CGAGAGGTCAGGG				
					TGGCCTTGGGGTCT				
					TGGATCTGCACAAG				
					ATGAATGAGGCGCT				
					TCTTGCTAAATGGA				
					TTTGGAAGTTGGAA				
					AATACGAATGGGCT				
					WTGGCAGACCATT				
					ATCAGACATAAATA				

					TGTTAAAGGAAGGC CTATTATTTCATTA AAAAAAGACAAAG TGATTCTCACTTTT GGAAAGGTATTCTG GATGCTCGGGACAA TTTTTACAAATATT GTAAAAAAAGGTA GGCAATGGCAATA ATACCAGCTTTTGG CACAATATTTGGTG TGATTTAGAACCTC TGTCTTCTAAATAT CCTAGTTTGTTTGA TATTGCTTATGATA AAGATATCACAGTG GAAAAAGCCTTTTC TTCTGATTTTAGAG CTGTGAATTTTAGA AGAAGGATTTATGG TGCTTTGGAAGATG ATTATAATAACCTT ATTACACAGTGTAA				

					TAATATTACTATTT				
					CTGAGGAGGATGAT				
					AAATCAGTTTGGCT				
					TCTGGGTAATAAAG				
					GTTTTTCTGTTAATT				
					CTTTTTACAAAGAA				
					ATCAAATGTGCTCA				
					GATTCCTTTGCCTT				
					CTAGTTTCTTGTGG				
					AAGACTAGACTCCC				
					ACATAAAATCAAA				
					GCTTTTTTGTGGCTT				
					GTTATGCATAAAAA				
					GATTTTAACCAAGG				
					ATAATTTGTTCAAA				
					AAAAATTGGAAGG				
					GCAATCTGGATTGT				
					ACTTTTTGTGGATT				
					TTCAGAATCCATAG				
					ATCATTTGTTTTTCA				
					GTGCTCGGTTGCTA				
					GATTTACTTGGAGA				
					ATAGTCTACTCGGC				

					TCTTGGCTTAATTT				
					CTATTCCCAACAGT				
					GTGGATGATCTGTT				
					TGGCTCTTGGATTA				
					ATAGTTTTAACAAA				
					ACTGAAAAGAATCT				
					GGTCCTTTTTGGTT				
					GTGGAGCTGTCATT				
					TGGGCGATCTGGCG				
					ATGTCGCAATGATT				
					GTTGCTTTGGAACT				
					GACAGAATTGATGA				
					TCCGACTAATGTGA				
					TTTTCTCTTGTTGCT				
					ATTGGCTTGATACT				
					TGGGCTATCCGTCA				
					GAAAAAGAAGGAA				
					AGAAAATTGGTGG				
					AACAAGGAAGCCT				
					CAAAATCAGAAGG				
					ATGGCGTGTGAGAT				
					CTACAACAAGACTC				
					GTGGTTGGAAGCCA				

					GTCGACAGACGGAT CCAATGAAGATTCT GCTGTGTTTGGCTA TTTTGGTTCATCAA CTTTGTGTTTGCTGT AATCCCCCCTCGCC TCGGTAGCTTCAGG CTACTGAAAACTTT GGGCTATATCGGGG YTGAAGTTGTTTAG TTTCGTGCTACTCT AGTTTGATCTTTGC GATCTTTGTTTTTG GTCAGCTATGTGCT GATTCTGTGCTGCT TCTGGTAGTTTGGT TCGTTCAGGTTTAG TTCCGTCGTAGCTG ATAGTTGTGTGTCT TGTCGGCTTTTAAA TTCTGGTTGGCTCT ATGGTCTTGTAAGA ACTTGTGTCTCTTTC				

					TATTTCGTAATGAAATAGGGGGTTATCCCCTCTGTCAAAAAAA				

В таблице 11 представлены Gene Writer, содержащие новые элементы ретротранспозона из таблицы 10 выше, измененные для улучшения эффективности интеграции в человеческий геном. Полипептиды ретротранспозазы были улучшены путем консенсусного картирования для повторного выведения оптимальной аминокислотной последовательности, указанной в данном документе как "консенсусная оптимизированная последовательность белка", как описано в примере 5. Как описано в примере 6, молекулы-матрицы для использования с когнатными ферментами подвергали обратному картированию на геномы их хозяев, а фланкирующие геномные ДНК использовали для выявления мотивов сайтов-мишеней. При выявлении консервативные мотивы последовательности из фланкирующей геномной ДНК эндогенных случаев появления элемента выравнивали с человеческим геномом, а новые последовательности получали из человеческого генома как 5'- или 3'-"плечи гомологии человека". В некоторых вариантах осуществления РНК-матрица, описанная в данном документе, содержит один или оба из первого домена гомологии, содержащего последовательность 5'-плеча гомологии человека из таблицы 11 (или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней), и второго домена гомологии, содержащего последовательность 3'-плеча гомологии человека из таблицы 11 (или последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней).

Таблица 11. Системы ретротранспозаз с улучшенной активностью интеграции

Элемент	**Предсказанный мотив-мишень**	**Консенсусная оптимизированная последовательность белка**	**5'-плечо гомологии человека**	**3'-плечо гомологии человека**
L2-18_ACar	CCACAAGGCCCT		AGTAGACTGCAGCCCCAGGCCCTTTCCATCTGGG	taaa

	GGCAGGGCTtaaa		TGTGACTCAGCACTTCCTAGGAAGGACCATCTATAAAGCCTGAGTCCACAAGGCCCTGGCAGGGCT	
CR1_AC_1	GTGTGAGTTTCTGCCAGAGCCattctatga		ccacggcacctggcccagtgtttttaatacaaattatagcctatggatcccactgaggacctctgtttgtgagactgggtgtgagttgctgtcagagcc	attctgcaaacctccagtgaggggaaggttgagggcagtggttggtgctcacggtggtgctgctcaaaactgacaggcagacacccggcctccagcttgc
RTE-1_Aip	Tacaaggcttgttgttgttgttgttgttttttt		TA	caaggctttgttgttgttgttgttgttttttt
Dong-1_MMa	GGTTCAACGGGCCGGGCCCGGtaataataataataaaa	MKANPAMETRSMRNRRMRSPEEGAPTGAGPGTGDRASGQRLEEDNVQERPSSQRAQPVPRTRNRNGSSRNHQGHAAPTDVAVADRRQSLTLAGGRRQRIMWTREMNMYVIRCYYVCTRMETDMSGRSGMLEMFNERFPRFARQLDLNKLYTRQRAIMSHNMLTAAEVEYIKLEVQREIGEEGTRSSDTSRRSSVGLDAPITSESRDSPAPTAPGPTADMQRQQLRDELALHMGTAVTQFRGTDPMSRHRIPKLRYSYRLTSAVSILNQDILPQYL	AAGGATGGGGTTCAACGGGCCGGGCCCGG	taataataataataaaa

		EAVENLEDLQFIVYSAAIAVVKTLGLRTRPQGEDG TRPNAQKPVWMRRLESRIATLRSKVGRLTQYKQG NRSTRLVRHVAEIVRPAELRDLREVDITEILDTHVQ RLSALAKRLRRYGECSKRKEQNRMFNINEREFYNR IRNDKVDFGEGLPEIGDVTQFWANLWENPTEHDG DGMWLAEEERQCDGIGDMTAVVVTAQDIREATRY TRNWAAPGPDFVHNFWYKKLTTIHGRMAECFNTV LGDPTQLPEFITRGVTFLLPKDQHTADPAKYRPITC LSSLYKVLSSVIARRVQAHCDANNVMTEEQKGCR RNTQGCKDQVIVDAVIVGQATQKQRNLSMAYIDY KKAYDSVPHSYLLKVLQLYKVDGNVIRLMQHAM GMWSTSLHITDGTAVLRSRTLSIRRGIFQGDTFSPL WFCLAMNPLSKALNQCNYGYQLKSGERSTRVTHT FYMDDLKLFAESVQRLHQLLQLVTTFSNDIRMEFG IDKCRSIHLRRGQVMDASCFRVNEQEEIRNMVEGE TYKYLGFLQLRGIRHTMIKKELQEKFLSRVNCILKS FLSAGNKVKAINTFAVPLLTYSFGVVKWTKTDLEA IERAVRVAFTKHRMRHPKSSIERVTLPRAAGGRGV TDIQALCVSQIQQLRAYFVESQNRHEIYRTVCEAD HGFSALHLAQEDYQLNCDIKTVDEMIAMWKQKEL HGTHPHQLELEHIDKVASNTWLVRGDLFSETEGFM VAIQDRVIATKNYRHYILHEDVEDRCRKCNSVGET IEHVVAGCSVLAGSAYLDRHNEVAKIVHQQLALK		

		HNLVDRFVPYYKYLPDPVLENSCIKLYWDREIITD VLIRANRPDIVVYDKRMKRVTLIDIAVPLDHNVQS TFSNKIAKYHDLAEELKQMWHLEDVRIVPVVLSAT GIVPKSLLRSLDELELKKDLHSIQKAVILGTCSIVRR FLNHHN		
L2-2_DRe	TGAGCACGGTAGCATTAGTCGa	MCFLIPVVTNTRKTREVRCRRNPHNLRSIHVSTISQ LSLSVGLWNCQSAVNKADFITSIATYSDYNLMALT ETWLRPEDTATHATLSANFSFSHTPRQTGRGGGTG LLISKEWKFTLIPSLPTISSFEFHAVTIIHPFYINVVVI YRPPGKLGHFLDELDVLLSSFSNFATPLLVLGDFNI YVDKPQAADFQTLLASFDLKRAPTSATHKSGNQL DLIYTRHCFTDQTIVTPLQISDHFLLSLNIHITPEPPH TPTLVTFRRNLRSLSPNRLSTIVSDSLPPSRKLTALD SNSATNTLCSTLASCLDRLCPLASRPARASPPAPWL SDALREHRSKLRAAERIWRKTKNPAHLLTYQTLLS SFSAEVTSAKQTYYRLKINNATNPRLLFKTFSSLLY PPPPPASSTLTTDDFATFFCTKTAKISAQFAAPTTNT QDTTPTPHTLTSFSQLSESEVSKLVLSSHATTCPLDP IPSHLLQAISPAVIPTLTHIINTSLDSGLFPTTFKQAR VTPLLKKPNLDHTLLENYRPVSLLPFMAKILEKVVF NQVLDFLTQNNLMDNKQSGFKKGHSTETALLSVV EDLRLAKADSKSSVLILLDLSAAFDTVNHQILLSTL ESLGVAGTVIQWFRSYLSDRSFRVSWRGEVSNLQH	GTTTCGCGTCGTCCGCCTCAGTTTCGGCCCTTGTTTTGACTCGGGAGGCGTGTCCAAACGCCAGGTAACCACTATATAGTGAGCACGGTAGCATTAGTCG	a

		LNTGVPQGSVLGPLLFSIYTSSLGPVIQRHGFSYHC YADDTQLYLSFHPDDPSVPARISACLLDISHWMKD HHLQLNLAKTEMLVVSANPTLHHNFSIQMDGATIT ASKMVKSLGVTIDDQLNFSDHISRTARSCRFALYNI RKIRPFLSEHAAQLLVQALVLSKLDYCNSLLAGLP ANSIKPLQLLQNAAARVVFNEPKRAHVTPLLVRLH WLPVAARIKFKALMFAYKVTSGLAPSYLLSLLQIY VPSRNLRSVNERRLVVPSQRGKKSLSRTLTLNLPS WWNELPNCIRTAESLAIFKKRLKTQLFSLHFTS		
R4-1_AC	AAATAA TAATAA TAATAA TAATAA TAATAA TAATAA TAATAA TAATcaat aaaaataataa taatattattatt ataa	MENQTQTKSILARSSPRIKRTAVDAADSGHPSTSGL RNHQNPNEQSQKRQKYTMPENRTIMKCYYKSEPQ RRGYQKRMHQLWKQEYPDSQITEPRLADQRRFIIR NKVFSEVELEEIQKICKVDYHQTTAQTAAETPATL GMVEQIEPEESVALLQEFEEPALETSVEPPGTLTAR QQELKDKIMAHAAANAIRQRLPTLKTVPKRHLAPL MKDVNAALSTVQITSIEQTNQFAYSAAVIVTEELGL LQPRQPQRKSTGKPKWKVRLELKIKKLRSDASNLK NMKERKLKNDKIKQYLIRKYWLNTRKIEEALEIVK EQITATARKIERYEARIIQYRQNQLFQSDQRRFYQS LNQTTDTVTIKPEKTATTKFWKELWENNKNYNKN AGWIKEFEGKFSQNKMELMEITTEMISKRVQKVKN WTSPGSDQLHGFWLKHLISLHGKMAQQFNEMLQK GSISEWLTTGRTYLIQKDPAKGAAPGNYRPITCLPT	caataaaaataataataat attattattataa	'AATAATAATAATAATAATAAT caataaaaataataataata'

		MFKLLTGIIADRIQDYLEEKNILPDEQKGNKRKSRG TKDQLLIDKMILENCKSRKANLHMTWIDYKKAFD SLPHSWIIKCLDAIGICKTVGTFIENMMEHWKTELF VGNESYGLVNIRRGIFQGDSLSPLLFIIAMIPLSTILQ KTNLGYQISKNSHKISHLMYMDDLKLYGKTETEIQ SLTNTVRIFSTDINMEFGLDKCSTVALKKGKIIESEG INMPNGQTIKCHQPEAYKYLGILQLDNIKHEHVKT VVSKEYTQRVRKILKSKLNGGNTIKAINTWAIPVIR YTAGIINWTQVELDNLDRKTRKLMTIHHSLHPRSD VDRLYLPRRSGGRGLLQVKQAVKEEEHALAEYVK QSEEPALIEVKNQKLLKTQQTKNQYKKTALQTRA DSWHNKTLHGKFLDKIEGKADKEKTWLWLTNGT LKKETEGLILAAQEQAIRTNAIKAKIEKSADDPKCR LCKETDETIDHILSCCKKIAQTDYKQRHNYVAQMI HWNLCLKYHLPAAKNWWDHKPAKVLENEHAKIL WDFRIQTDKVLEHNTPDITVVEKNKVWIIDVAIPG DSRIDEKQQEKLSRYQDLKIELQRLWQKPVQVVPV VMGTLGAVPKDLSRHLETIDIDKITICQLQKATLLG SAHIIRKYITQS		
R4-1_PH	TAATAA TAATAA TAATAtaa taataat	MKMSHNRDTPSNGVKGTSVRLGTSLVRSPVGEAG AVRERGTHPSESVSQDSDASVNATGEGSVREQAPL SPPGAEEATVPTQRRTRHKWSREDRVVLWECFVA SKREGPGYLKRLKQLWDERGIPGNFPQASLSGQIR	TAATAATAATA ATAATA	taataataat

		QICSKNLLSEEERLQIAARMEAQVASPSADEPARQ VPTRPVTPPRSPPVEPARRPSIPSEETPDLGAVPSEID SADPNRSPSRGPRHLPAHNMSQSESEDDVTDPDVS DQQRSDSLEPRDLLRNSSVESTPGHPNQELSDTLLS NYVPSEIDSDDPNQSPRRGPRHLPAHDMSLSDSMD EETEPDLSDQQRSDLLELRDLLRNSSVETTPKGHPS LRHLPEPKIRAAAFRVNSVIGKIHTNNITETNALIKA GADLAVRILEVQPRPQRTQRKKDPPWKHRLEKNIA EIRKHLSWISEWRRGNLHDEEKKTLLESRYRCLEV GLTNLEDTLKQRLSAKRSKVRRFEARVAGFHQNQ LFNTNQKRLYQTLRGEETSSDSPNAEESIRFWSDIW SKEVRHNNTAEWLHDVKEKNVAADPDLTITSQQL KKQLSKTKNWKAPGPDMVQGYWIKTFTSLHSRIA AQLNHCLQRGTVPTWMTTGKTVLIQKDKAKGTEV SNYRPITCLPLMWKVLTGIIYERVYQHLDSKKLLPD EQKGCRRNTRGTKDHLLVDKLLTKDARSKKKNLS MAWVDYKKAFDMVPHSWILECLDIYGIAGNIRNLI ATTMPNWKTQLTSANKHLGEVSIKRGIFQGDSLSP LLFVLTMIPLSETLNKAGQGYNYSRTMKLNHLLY MDDLKLYAKSKDQVEQLLNIVHQYSQDIKMQFGV SKCGVLNIERGEVTASEGITIEEGTIKDIEEAGYKYL GVMEYNTILHRTMKDSIRKEYLTRLRLILKSHLNG GNTIKAINTWAVPVVRYSAGIINWTKKDCTDMDIK		

		TRKLMTIYRALHPRSCVDRLYINRREGGRGLISVED CVEAEKRALSQHFRESDDPWARCLVEAKLLKETET ADQFKERRRLDRTNKWKSMKMSGQYLEAVQDKI VPDSWNWLLRGELKRETEGTILAAQEQALRTRYIQ NKIDKRNVPSTCRICRSSDETINHVISECGVLAQKE YKRRHDKVARHLHWTLLRIHNFPVSERWYEHEPA PVVENEAVQIYWDKRMETDRVLHANRPDIVVKDK QEKSAKLIDISIPFDSRIVDKEAEKKEKYRDLAIELQ RLWQMKVDVVPVVIGALGAMSKNLKTALRELKC GHLHPGTLQKSALLGTAHIIRKVL		
Tad1-65B_BG	ATGCGG AATCAG GGATGA CCATGG ATTCTTT GGCTGT GGGTAG AGTCCTT TCTAAC CCAGCG AAGGAT AAGGGT GCGGTA	-	ATGCGGAATCA GGGATGACCAT GGATTCTTTGG CTGTGGGTAGA GTCCTTTCTAAC CCAGCGAAGGA TAAGGGTGCGG TACCGTAATGG AACGACGCGGG	tcaaatcaaatcaaatcaaatcaaatcaaatcaaat caaatcaaatcaaatcaaatcaaatcaaatcaaatc aaatcaaatcaaatcaaatcaaatcaaa

	CCGTAA TGGAAC GACGCG GGtcaaatc aaatcaaatca aatcaaatcaa atcaaatcaaa tcaaatcaaat caaatcaaatc aaatcaaatca aatcaaatcaa atcaaatcaaa tcaa			
L2-24_Acar	CTAACC TCCGCA GGGGTG GTGGAat aaa	MPIKMVRPRRLMDPNGFLMALGDVPASSVGDSVD ALVTLYNGEMARAIDTIAPERPLAMRRAKLAPWFS GELAAMKRETRRLERVWRKRRSESNLERVKAQRK TYCKAMAAVKKTYYSSRIEAAKNRPAELFRTVRG LLNNTPQTAPTDDVAEKCEAFSQFFSKKIALIRSEL DATVDAISEDVARAPACPILMDSFQSVQPEDVDKV LGEARATTCILDPCPSWLVKEARGGLVEWVRVVV NASLREGRLPAGLKQAVIKPLLKKPSLDPTNWNNF RPVSNLPFLGKVVERVVASQLQGFLEATDYLDPAQ SGFRPGHGTETALVSLVDDLRRELDRGSVSLLVLL	CTAACCTCCGC AGGGGTGGTGG A	ataaa

		DLSAAFDTVDHGILLGRLVGMGLGGTALQWLQSF LEDRTQKVLIGDSCSAPQPLSCGVPQGSVLSPLLFNI YMKPLGEIIRSFGVRCHLYADDVQLYHSFSPASKE AVQILNRCLAAVSDWMRANKLKLNPDKTEVLLVS RKAEQGIGLQPVLDGVALPLKAQVRSLGVLLDSSL SLEPQVSAVARGAFAQLKLVRQLRPYLGKPDMAT VVHALVTSRLDYCNALYVGLPSKTVRKLELVQRS AARLLTGAPYREHTTPLLKQLHWLPVSFRAQFKVL ALAYKTLNGSGPTYLSERISPYEPPRRLRSSSEALLS VPPPSQVHLAGTRDRAFSVVAPRLWNSLPREIREAP SLMSFRKQLKTWMWDQAFGPS		
LINE1-2_ZM	CTGAGG CCTTATT ATTGTCT CCTCCG GTTGAG GCTGTC ATGTCT ACTACT GCTATG TTGGGG GCCATG GCTCCG	MNLSNILVWNARGLNNKARRDYVRDTILSSKADII CLQETKVASFSSHLLLSVCGSDFDKFLTLPAVGTRG GILIAWKGAVCQVISSRIDNFSVSVQFAGMDGMN WWFTGVYGPQEDDRKIQFLQELLDFRTLCSGPWL LSGDFNMIYQAADKNNANLNRALMGRFRRFLDDC LLKEIPLHGRKFTWSNERSSPTLVRLDRVFCCNDW EDFFPNSLLQSASSGVSDHCPLILSLNVQTHKKRRF HFESYWPKLPGFLEAVSHSWNAPVSIACPVEGIFLK LQRLSKGLQKWSHHKVGNVKLQLATAKEILHRLEI ARDNRVLSPGEEWLRKKLKLHCLGLASLERTIARL RSRILYLQEGDANTSFFHQQARYRKKKNFIAKFQV EERIVTSQEEKQQAALDFYDNLLGTAEHRDYTLDL	-	-

	GGCCCT	HSLGIQQHDLGDLELVFSMEEVWSVVKDLPLDKA		
	GAGCAT	PGPDGFTGRFYKVCWDIIKEDMMGALLAVQCGHV		
	ATCCAG	SKLKLLNTTFITLLPKKVDALLVKDYRPISLVHSFA		
	GGTCCC	KLVTKVLAARLAPHLPSMVSINQSAFIKGRSIQDNF		
	TTaagtg	LLVQQLTRSLHHSKEPHILLKLDISKAFDSISWPFLL		
		DMLQHLGFGRSWCNLISLLLCTSSTRILVNGEPGEY		
		ILHHRGLRQGDPLSPMLFILVMEALNAMVSYAFRE		
		HMLQPIANQHAKHCISFYADDVVLFMRPSHQDISA		
		IISILDIFGHATGLKTNISKSSVTPIQCGANELTTISAL		
		LPCATKNFPCTYLGIPLSIRRLANSDLLPLVDKVAD		
		KLPGWKANLLSRAGRLVMVKTVLSSVPIYLMLAL		
		ELPKWVFKAIDKRRRGFLWKGQGDANGGNCLVS		
		WEAVQRPLKYGGLGILNLEMFGWALRARWLWLQ		
		KTDASRPWAGLPIRVHRNAKALVDVAITSVVGSGE		
		SVKFWSDRWLLGKTVAEHCPTLIQVISRRALKRRT		
		VAEGLTNRQWVSDIRGGLSVTVLVEYLQLWNLVD		
		GVVLQPDIADQHIWRLSAHGTYCSKSAYDALFVGS		
		IPFGPWRRVWKTWAPLRCKFFVWLAIKNRVWTAD		
		RLAKRGLPHPVACPLCDQAEETIQHILVSCVFARQI		
		WTSMLHNLGLLAIVPQHGCTRFSNWWCQSIKKVE		
		KSLRKGLNSLIILVAWEIWKHRNACVFEGVVPCTQ		
		RVQSAVIEEGNVWCLAGASALQDLLLRQLPRGP		

Системы Gene Writer, например, термостабильные системы Gene Writer

Не желая ограничиваться какой-либо теорией, полагают, что в некоторых вариантах осуществления ретротранспозазы, эволюционировавшие в холодных условиях, не могут функционировать так же хорошо при температуре тела человека. В настоящей заявке представлен ряд термостабильных систем Gene Writer, включая белки, полученные из ретротранспозаз птиц.

Термостабильность можно измерить, например, путем тестирования способности системы Gene Writer к осуществлению полимеризации ДНК *in vitro* при высокой температуре (например, 37°C) и низкой температуре (например, 25°C). Подходящие условия для анализа активности полимеризации ДНК *in vitro* (например, процессивности) описаны, например, в Bibillo and Eickbush, "High Processivity of the Reverse Transcriptase from a Non-long Terminal Repeat Retrotransposon" (2002) JBC 277, 34836-34845. В некоторых вариантах осуществления термостабильный полипептид Gene Writer характеризуется активностью, например, активностью полимеризации ДНК, при 37°C, которая составляет не менее 70%, 75%, 80%, 85%, 90% или 95% от его активности при 25°C при в остальном аналогичных условиях.

В некоторых вариантах осуществления полипептид Gene Writer (например, последовательность из таблицы 1, 2 или 3A или 3B или последовательность, характеризующаяся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней) является стабильным у субъекта, выбранного из млекопитающего (например, человека) или птицы. В некоторых вариантах осуществления полипептид Gene Writer, описанный в данном документе, является функциональным при 37°C. В некоторых вариантах осуществления полипептид Gene Writer, описанный в данном документе, характеризуется большей активностью при 37°C, чем при более низкой температуре, например, при 30°C, 25°C или 20°C. В некоторых вариантах осуществления полипептид Gene Writer, описанный в данном документе, характеризуется большей активностью в клетке человека, чем в клетке данио-рерио.

В некоторых вариантах осуществления полипептид Gene Writer является активным в клетках человека, культивируемых при 37°C, например, при использовании анализа из примера 6 или примера 7 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления анализ включает стадии: (1) введения клеток HEK293T в одну или несколько лунок диаметром 6,4 мм из расчета 10000 клеток/лунка, (2) инкубирования клеток при 37°C в течение 24 часов, (3) получения смеси для трансфекции, содержащей 0,5 мкл реагента для трансфекции FuGENE® HD и 80 нг ДНК (где ДНК представляет собой плазмиду, содержащую в указанном порядке: (a) промотор CMV, (b) 100 п. о. последовательности, гомологичные 100 п. о. выше сайта-мишени, (c) последовательность, кодирующую 5'-нетранслируемую область, которая связывается с белком Gene Writer, (d) последовательность, кодирующую белок Gene Writer, (e) последовательность, кодирующую 3'-нетранслируемую область, которая связывается с

белком Gene Writer, (f) 100 п. о. последовательности, гомологичные 100 п. о. ниже сайта-мишени, и (g) последовательность полиаденилирования BGH) в 10 мкл Opti-MEM, и инкубирования в течение 15 минут при комнатной температуре, (4) добавления смеси для трансфекции к клеткам, (5) инкубирования клеток в течение 3 дней и (6) анализа интеграции экзогенной последовательности в локус-мишень (например, рДНК) в геноме клетки, например, где одна или несколько из предыдущих стадий выполняются, как описано в примере 6 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления система Gene Writer способна производить замену в сайте-мишени по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 или больше нуклеотидов. В некоторых вариантах осуществления замена представляет собой мутацию по типу транзиции. В некоторых вариантах осуществления замена представляет собой мутацию по типу трансверсии. В некоторых вариантах осуществления в результате замены происходит превращение аденина в тимин, аденина в гуанин, аденина в цитозин, гуанина в тимин, гуанина в цитозин, гуанина в аденин, тимина в цитозин, тимина в аденин, тимина в гуанин, цитозина в аденин, цитозина в гуанин или цитозина в тимин.

В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований).

В некоторых вариантах осуществления вставка, делеция, замена или их комбинация обеспечивают повышение или снижение экспрессии (например, транскрипции или трансляции) гена. В некоторых вариантах осуществления вставка, делеция, замена или их комбинация обеспечивают повышение или снижение экспрессии (например, транскрипции или трансляции) гена посредством изменения, добавления или делеции последовательностей в промоторе или энхансере, например, последовательностей, которые связывают факторы транскрипции. В некоторых вариантах осуществления вставка, делеция, замена или их комбинация обеспечивают изменение трансляции гена (например, изменение аминокислотной последовательности), вставку или делецию стартового или стоп-кодона, изменение или восстановление рамки трансляции гена. В некоторых вариантах осуществления вставка, делеция, замена или их комбинация изменяют сплайсинг гена, например, посредством вставки, делеции или изменения акцепторного или донорного сайта сплайсинга. В некоторых вариантах осуществления вставка, делеция, замена или их комбинация изменяют период полужизни транскрипта или белка. В некоторых вариантах осуществления вставка, делеция, замена или их комбинация изменяют локализацию белка в клетке (например, обеспечивают перемещение из цитоплазмы в митохондрии, из цитоплазмы во внеклеточное пространство (например, добавляют метку секреции)). В некоторых вариантах осуществления вставка, делеция, замена или их комбинация изменяют (например, улучшают) сворачивание белка (например, для предупреждения накопления ошибочно свернутых белков). В некоторых вариантах осуществления вставка, делеция, замена или их комбинация изменяют, повышают, снижают активность гена, например, белка, кодируемого данным геном.

В некоторых вариантах осуществления полипептид Gene Writer приводит к вставке последовательности гетерологичного объекта (например, гена GFP) в локус-мишень (например, рДНК) при среднем числе копий, составляющем по меньшей мере 0,01, 0,025, 0,05, 0,075, 0,1, 0,15, 0,2, 0,25, 0,3, 0,4, 0,5, 0,75, 1, 1,25, 1,5, 1,75, 2, 2,5, 3, 4, или 5 копий на геном. В некоторых вариантах осуществления клетка, описанная в данном документе (например, клетка, содержащая гетерологичную последовательность в сайте-мишени для вставки), содержит последовательность гетерологичного объекта при среднем числе копий, составляющем по меньшей мере 0,01, 0,025, 0,05, 0,075, 0,1, 0,15, 0,2, 0,25, 0,3, 0,4, 0,5, 0,75, 1, 1,25, 1,5, 1,75, 2, 2,5, 3, 4, или 5 копий на геном.

В некоторых вариантах осуществления Gene Writer вызывает интеграцию последовательности в РНК-мишень при относительно небольшом количестве событий усечения на конце. Например, в некоторых вариантах осуществления белок Gene Writer (например, под SEQ ID NO: 1016) приводит к тому, что приблизительно 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80-90% или 86,17% интегрированных элементов в сайте-мишени не являются усеченными согласно измерению с помощью анализа, описанного в данном документе, например, анализа из примера 6 и фигуры 8 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления белок Gene Writer (например, под SEQ ID

NO: 1016) приводит к тому, что по меньшей мере приблизительно 30%, 40%, 50%, 60%, 70%, 80% или 90% интегрированных элементов в сайте-мишени не являются усеченными согласно измерению с помощью анализа, описанного в данном документе. В некоторых вариантах осуществления интегрированный элемент классифицируется как усеченный или неусеченный с помощью анализа, включающего амплификацию с использованием прямого праймера, находящегося на расстоянии 565 п. о. от конца элемента (например, последовательности ретротранспозона дикого типа), и обратного праймера, находящегося в геномной ДНК сайта-мишени для вставки, например, рДНК. В некоторых вариантах осуществления количество полноразмерных интегрированных элементов в сайте-мишени для вставки превышает количество интегрированных элементов, усеченных на 300-565 нуклеотидов в сайте-мишени для вставки, например, количество полноразмерных интегрированных элементов составляет по меньшей мере 1,1x, 1,2x, 1,5x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x или 10x от количества усеченных интегрированных элементов, или количество полноразмерных интегрированных элементов составляет по меньшей мере 1,1x-10x, 2x-10x, 3x-10x или 5x-10x от количества усеченных интегрированных элементов.

В некоторых вариантах осуществления система или способ, описанные в данном документе, приводят к вставке последовательности гетерологичного объекта только в один сайт-мишень в геноме клетки-мишени. Вставка может быть измерена, например, с использованием порогового значения выше 1%, 1,5%, 2%, 2,5%, 3%, 3,5%, 4%, 4,5%, 5%, например, как описано в примере 8 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления система или способ, описанные в данном документе, приводят к вставке последовательности гетерологичного объекта, где менее 1%, 1,5%, 2%, 2,5%, 3%, 3,5%, 4%, 4,5%, 5%, 10%, 20%, 30%, 40% или 50% вставок находятся в сайте, отличном от сайта-мишени, например, при использовании анализа, описанного в данном документе, например, анализа из примера 8 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления система или способ, описанные в данном документе, приводят к "безрубцовой" вставке последовательности гетерологичного объекта, тогда как в некоторых вариантах осуществления сайт-мишень может демонстрировать делеции или дупликации эндогенной ДНК в результате вставки гетерологичной последовательности. Механизмы различных ретротранспозонов могут приводить к различным паттернам дупликаций или делеций в геноме хозяина, происходящих во время ретротранспозиции в сайте-мишени. В некоторых вариантах осуществления система приводит к безрубцовой вставке без дупликаций или делеций в окружающей геномной ДНК. В некоторых вариантах осуществления система приводит к делеции менее 1, 2, 3, 4, 5, 10, 50 или 100 п. о. геномной ДНК выше места вставки. В некоторых вариантах осуществления система приводит к делеции менее 1, 2, 3, 4, 5, 10, 50 или 100 п. о. геномной ДНК ниже места вставки. В некоторых вариантах осуществления система приводит к дупликации менее 1, 2, 3, 4, 5, 10, 50 или 100 п. о. геномной ДНК выше

места вставки. В некоторых вариантах осуществления система приводит к дупликации менее 1, 2, 3, 4, 5, 10, 50 или 100 п. о. геномной ДНК ниже места вставки.

В некоторых вариантах осуществления Gene Writer, описанный в данном документе, или его ДНК-связывающий домен специфично связывается со своим сайтом-мишенью, например, согласно измерению с помощью анализа из примера 21 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления Gene Writer или его ДНК-связывающий домен связывается со своим сайтом-мишенью сильнее, чем с любым другим сайтом связывания в человеческом геноме. Например, в некоторых вариантах осуществления в анализе в примере 21 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте, в сайте-мишени представлено более 50%, 60%, 70%, 80%, 90% или 95% событий связывания Gene Writer или его ДНК-связывающего домена с геномной ДНК человека. В некоторых вариантах осуществления ДНК-связывающий домен Gene Writer является гетерологичным по отношению к остальной части Gene Writer, например, таким образом, что Gene Writer нацелен на другой сайт-мишень, нежели эндогенный ДНК-связывающий домен, ассоциированный с остальной частью Gene Writer.

Генно-инженерные, например, димеризованные системы Gene Writer

Некоторые не содержащие LTR ретротранспозоны используют две субъединицы для завершения ретротранспозиции (Christensen et al. PNAS 2006). В некоторых вариантах осуществления ретротранспозаза, описанная в данном документе, содержит две соединенные субъединицы в виде единого полипептида. Например, две ретротранспозазы дикого типа могут быть соединены линкером с образованием белка, "димеризованного" посредством ковалентных связей. В некоторых вариантах осуществления нуклеиновая кислота, кодирующая ретротранспозазу, кодирует две субъединицы ретротранспозазы, которые должны экспрессироваться в виде единого полипептида. В некоторых вариантах осуществления субъединицы соединены пептидным линкером, таким как тот, который был описан в данном документе в разделе под названием "Линкер" и, например, в Chen et al., Adv Drug Deliv Rev 2013. В некоторых вариантах осуществления две субъединицы в полипептиде соединены жестким линкером. В некоторых вариантах осуществления жесткий линкер состоит из мотива$(EAAAK)_n$ (SEQ ID NO: 1534). В других вариантах осуществления две субъединицы в полипептиде соединены гибким линкером. В некоторых вариантах осуществления гибкий линкер состоит из мотива $(Gly)_n$. В некоторых вариантах осуществления гибкий линкер состоит из мотива $(GGGGS)_n$ (SEQ ID NO: 1535). В некоторых вариантах осуществления жесткий или гибкий линкер состоит из 1, 2, 3, 4, 5, 10, 15 или больше аминокислот в длину для обеспечения ретротранспозиции. В некоторых вариантах осуществления линкер состоит из комбинации жестких и гибких линкерных мотивов.

Исходя из механизма, не все функции являются необходимыми из обеих субъединиц ретротранспозазы. В некоторых вариантах осуществления слитый белок может состоять из

полностью функциональной субъединицы и второй субъединицы, в которой отсутствуют один или несколько функциональных доменов. В некоторых вариантах осуществления одна субъединица может не обладать функциональными свойствами обратной транскриптазы. В некоторых вариантах осуществления в одной субъединице может отсутствовать домен обратной транскриптазы. В некоторых вариантах осуществления одна субъединица может обладать только эндонуклеазной активностью. В некоторых вариантах осуществления одна субъединица может содержать только эндонуклеазный домен. В некоторых вариантах осуществления две субъединицы, составляющие один полипептид, могут выполнять взаимодополняющие функции.

В некоторых вариантах осуществления одна субъединица может не обладать функциональными свойствами эндонуклеазы. В некоторых вариантах осуществления в одной субъединице может отсутствовать эндонуклеазный домен. В некоторых вариантах осуществления одна субъединица может обладать только активностью обратной транскриптазы. В некоторых вариантах осуществления одна субъединица может содержать только домен обратной транскриптазы. В некоторых вариантах осуществления одна субъединица может обладать только функциональным свойством ДНК-зависимого синтеза ДНК.

Линкеры

В некоторых вариантах осуществления домены композиций и систем, описанных в данном документе (например, эндонуклеазный домен и домен обратной транскриптазы полипептида или ДНК-связывающий домен и домены обратной транскриптазы полипептида), могут быть соединены линкером. Описанная в данном документе композиция, содержащая линкерный элемент, имеет общую форму S1-L-S2, где S1 и S2 могут быть одинаковыми или разными и представляют собой два доменных компонента (например, каждый из которых является доменом полипептида или нуклеиновой кислоты), связанных друг с другом с помощью линкера. В некоторых вариантах осуществления линкер может соединять два полипептида. В некоторых вариантах осуществления линкер может соединять две молекулы нуклеиновой кислоты. В некоторых вариантах осуществления линкер может соединять полипептид и молекулу нуклеиновой кислоты. Линкер может представлять собой химическую связь, например, одну или несколько ковалентных связей или нековалентных связей. Линкер может быть гибким, жестким и/или расщепляемым. В определенных вариантах осуществления линкер представляет собой пептидный линкер. Как правило, пептидный линкер имеет длину по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше аминокислот, например, имеет длину 2-50 аминокислот, имеет длину 2-30 аминокислот.

Наиболее часто используемые гибкие линкеры имеют последовательности, состоящие главным образом из отрезков из остатков Gly и Ser (линкеры "GS"). Гибкие линкеры могут быть применимы для соединения доменов, которые требуют определенной степени подвижности или взаимодействия, и могут содержать небольшие неполярные (например, Gly) или полярные (например, Ser или Thr) аминокислоты. Включение Ser или

Thr может также обеспечивать поддержание стабильности линкера в водных растворах благодаря образованию водородных связей с молекулами воды и, следовательно, уменьшение неблагоприятных взаимодействий между линкером и другими компонентами. Примеры таких линкеров включают линкеры, имеющие структуру $[GGS]_{\geq 1}$ или $[GGGS]_{\geq 1}$ (SEQ ID NO: 1536). Жесткие линкеры применимы для сохранения фиксированного расстояния между доменами и поддержания их независимых функций. Жесткие линкеры также могут быть применимы, когда пространственное разделение доменов является критически важным для сохранения стабильности или биологической активности одного или нескольких компонентов в средстве. Жесткие линкеры могут иметь альфа-спиральную структуру или Pro-богатую последовательность (XP)n, при этом X обозначает любую аминокислоту, предпочтительно Ala, Lys или Glu. Расщепляемые линкеры могут высвобождать свободные функциональные домены in vivo. В некоторых вариантах осуществления линкеры могут расщепляться в специфических условиях, таких как присутствие восстанавливающих реагентов или протеаз. Расщепляемые in vivo линкеры могут использовать обратимую природу дисульфидной связи. Один пример включает в себя последовательность, чувствительную к тромбину (например, PRS), между двумя остатками Cys. Обработка тромбином in vitro CPRSC (SEQ ID NO: 1537) приводит к расщеплению последовательности, чувствительной к тромбину, тогда как обратимая дисульфидная связь остается интактной. Такие линкеры известны и описаны, например, в Chen et al., 2013. Fusion Protein Linkers: Property, Design and Functionality. Adv Drug Deliv Rev. 65(10): 1357-1369. Расщепление линкеров in vivo в композициях, описанных в данном документе, может также осуществляться протеазами, которые экспрессируются in vivo при патологических состояниях (например, раке или воспалении), в определенных клетках или тканях или ограничены определенными клеточными компартментами. Специфичность многих протеаз обеспечивает более медленное расщепление линкера в ограниченных компартментах.

В некоторых вариантах осуществления аминокислотные линкеры представляют собой эндогенные аминокислоты (или гомологичны им), которые существуют между такими доменами в нативном полипептиде. В некоторых вариантах осуществления эндогенные аминокислоты, которые существуют между такими доменами, заменены, но длина не изменена по сравнению с естественной длиной. В некоторых вариантах осуществления к существующим в естественных условиях аминокислотным остаткам между доменами добавлены дополнительные аминокислотные остатки.

В некоторых вариантах осуществления аминокислотные линкеры разработаны вычислительным путем или подвергнуты скринингу для максимального увеличения функции белка (Anad et al., FEBS Letters, 587:19, 2013).

В дополнение к тому, что полипептид полностью кодируется одним транскриптом, он может быть получен посредством отдельной экспрессии двух или более полипептидных фрагментов, из которых восстанавливается полный фермент. В некоторых вариантах осуществления полипептид Gene Writer получают путем экспрессии в виде отдельных

субъединиц, из которых повторно собирается полный фермент посредством сконструированных белок-белковых взаимодействий. В некоторых вариантах осуществления восстановление полного фермента не предполагает ковалентное связывание между субъединицами. Пептиды также могут сливаться вместе посредством *транс*-сплайсинга интеинов (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). В некоторых вариантах осуществления полный фермент Gene Writer экспрессируется в виде отдельных субъединиц, которые предназначены для создания слитого белка за счет присутствия расщепленных интеинов (например, описанных в данном документе) в субъединицах. В некоторых вариантах осуществления полный фермент Gene Writer восстанавливается посредством образования ковалентных связей между субъединицами. В некоторых вариантах осуществления белковые субъединицы собираются повторно с помощью сконструированных партнеров по белок-белковому связыванию, например, SpyTag и SpyCatcher (Zakeri et al. PNAS 109, E690-E697 (2012)). В некоторых вариантах осуществления дополнительный домен, описанный в данном документе, например, никаза Cas9, экспрессируется в виде отдельного полипептида, который ассоциирует с полипептидом Gene Writer посредством ковалентных или нековалентных взаимодействий, описанных выше. В некоторых вариантах осуществления разделение полипептида Gene Writer на субъединицы может способствовать доставке белка за счет поддержания нуклеиновой кислоты, кодирующей каждую часть, в оптимальных пределах упаковки вирусного вектора для доставки, например, AAV (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). В некоторых вариантах осуществления полипептид Gene Writer предназначен для димеризации посредством использования ковалентных или нековалентных взаимодействий, описанных выше.

Интеины

В некоторых вариантах осуществления система Gene Writer содержит интеин. Как правило, интеин содержит полипептид, способный соединять два полипептида или фрагмента полипептида друг с другом с помощью пептидной связи. В некоторых вариантах осуществления интеин представляет собой транс-сплайсинговый интеин, который может соединять два фрагмента полипептида, например, с образованием полипептидного компонента системы, как описано в данном документе. В некоторых вариантах осуществления интеин может кодироваться в той же молекуле нуклеиновой кислоты, которая кодирует два фрагмента полипептида. В определенных вариантах осуществления интеин может транслироваться как часть более крупного полипептида, содержащего, например, по порядку первый фрагмент полипептида, интеин и второй фрагмент полипептида. В вариантах осуществления транслируемый интеин может быть способен вырезать себя из более крупного полипептида, например, что приводит к отделению присоединенных фрагментов полипептида. В вариантах осуществления вырезанный интеин может быть способен соединять два фрагмента полипептида друг с другом непосредственно с помощью пептидной связи. Иллюстративные интеины описаны в данном документе, например, в таблице X.

Компонент системы редактирования генов Gene Writer™, представляющий собой РНК-матрицу

Системы Gene Writer, описанные в данном документе, могут осуществлять транскрипцию последовательности РНК-матрицы в сайты-мишени ДНК хозяина с помощью обратной транскрипции с праймированием мишенью. Путем записывания последовательности(последовательностей) ДНК посредством обратной транскрипции последовательности РНК-матрицы непосредственно в геном хозяина система Gene Writer может осуществлять вставку последовательности объекта в геном-мишень без необходимости введения экзогенных последовательностей ДНК в клетку-хозяина (в отличие от, например, систем CRISPR), а также исключая стадию вставки экзогенной ДНК. Таким образом, система Gene Writer представляет собой платформу для использования индивидуально адаптированных последовательностей РНК-матриц, содержащих последовательности объекта, например, последовательности, предусматривающие кодирование гетерологичных генов и/или информацию о функции.

В некоторых вариантах осуществления РНК-матрица кодирует белок Gene Writer в *цис*-положении по отношению к последовательности гетерологичного объекта. Различные цис-конструкции были описаны, например, в Kuroki-Kami et al (2019) *Mobile DNA* 10:23 (включена в данный документ посредством ссылки во всей своей полноте) и могут применяться в комбинации с любым из вариантов осуществления, описанных в данном документе. Например, в некоторых вариантах осуществления РНК-матрица содержит последовательность гетерологичного объекта, последовательность, кодирующую белок Gene Writer (например, белок, содержащий (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, например, как описано в данном документе), 5'-нетранслируемую область и 3'-нетранслируемую область. Компоненты могут быть включены в различном порядке. В некоторых вариантах осуществления белок Gene Writer и последовательность гетерологичного объекта кодируются в разных направлениях (смысловом и антисмысловом), например, с использованием схемы расположения, показанной на фигуре 3A в Kuroki-Kami et al, там же. В некоторых вариантах осуществления белок Gene Writer и последовательность гетерологичного объекта кодируются в одном направлении. В некоторых вариантах осуществления нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или нуклеиновая кислота, кодирующая РНК-матрицу, связаны ковалентной связью, например, являются частью слитой нуклеиновой кислоты и/или являются частью одного и того же транскрипта. В некоторых вариантах осуществления слитая нуклеиновая кислота содержит РНК или ДНК.

Нуклеиновая кислота, кодирующая полипептид Gene Writer, может в некоторых случаях находиться с 5'-стороны от последовательности гетерологичного объекта. Например, в некоторых вариантах осуществления РНК-матрица содержит в направлении от 5' к 3' 5'-нетранслируемую область, полипептид Gene Writer, кодируемый в смысловом направлении, последовательность гетерологичного объекта, кодируемую в смысловом направлении, и 3'-нетранслируемую область. В некоторых вариантах осуществления РНК-

матрица содержит в направлении от 5' к 3' 5'-нетранслируемую область, полипептид Gene Writer, кодируемый в смысловом направлении, последовательность гетерологичного объекта, кодируемую в антисмысловом направлении, и 3'-нетранслируемую область.

В некоторых вариантах осуществления РНК дополнительно предусматривает гомологию в отношении сайта-мишени ДНК.

Понятно, что, если РНК-матрица описывается как содержащая открытую рамку считывания или последовательность, обратно комплементарную ей, в некоторых вариантах осуществления РНК-матрица должна быть преобразована в двухнитевую ДНК (например, посредством обратной транскрипции) до того, как открытая рамка считывания может быть транскрибирована и транслирована.

В определенных вариантах осуществления индивидуально адаптированную последовательность РНК-матрицы можно идентифицировать, разработать, создать и сконструировать таким образом, чтобы она содержала последовательности, изменяющие или определяющие функционирование генома хозяина, например, путем введения гетерологичной кодирующей области в геном; осуществления или обеспечения сплайсинга экзонной структуры/альтернативного сплайсинга; обеспечения нарушения функции эндогенного гена; обеспечения активации транскрипции эндогенного гена; обеспечения эпигенетической регуляции эндогенной ДНК; обеспечения повышения или снижения экспрессии функционально связанных генов и т. д. В определенных вариантах осуществления индивидуально адаптированную последовательность РНК-матрицы можно сконструировать таким образом, чтобы она содержала последовательности, кодирующие экзоны и/или трансгены, обеспечивающие сайты связывания с активаторами, репрессорами, энхансерами факторов транскрипции и т. д., а также их комбинации. В других вариантах осуществления кодирующую последовательность можно дополнительно индивидуально адаптировать с помощью акцепторных сайтов сплайсинга, поли(А)-хвостов. В определенных вариантах осуществления последовательность РНК может содержать последовательности, кодирующие последовательность РНК-матрицы, гомологичной ретротранспозазе, являться сконструированной таким образом, чтобы она содержала гетерологичные кодирующие последовательности, или представлять собой комбинации вышеперечисленного.

РНК-матрица может характеризоваться некоторой степенью гомологии с ДНК-мишенью. В некоторых вариантах осуществления РНК-матрица имеет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 или более оснований со строгой гомологией с ДНК-мишенью на 3'-конце РНК. В некоторых вариантах осуществления РНК-матрица имеет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 175, 180, или 200 или более оснований с по меньшей мере 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% или 100% гомологией с ДНК-мишенью, например, на 5'-конце РНК-матрицы. В некоторых вариантах осуществления РНК-матрица имеет 3'-нетранслируемую область, полученную из ретротранспозона, например,

ретротранспозонов, описанных в данном документе. В некоторых вариантах осуществления РНК-матрица имеет 3'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 или больше оснований с по меньшей мере 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% или 100% гомологией с 3'-последовательностью ретротранспозона, например, ретротранспозона, описанного в данном документе, например, ретротранспозона в таблице X, Z1, Z2, 3A или 3B. В некоторых вариантах осуществления РНК-матрица имеет 5'-нетранслируемую область, полученную из ретротранспозона, например, ретротранспозонов, описанных в данном документе. В некоторых вариантах осуществления РНК-матрица имеет 5'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180 или 200 или больше оснований с по меньшей мере 40%, 50%, 60%, 70%, 80%, 90%, 95% или большей гомологией с 5'-последовательностью ретротранспозона, например, ретротранспозона, описанного в данном документе, например, ретротранспозона, описанного в таблице X, Z1, Z2, 3A или 3B.

Компонент системы редактирования генома Gene Writer, представляющий собой РНК-матрицу, описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления РНК-матрица имеет 3'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 3'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer.

Компонент системы редактирования генома Gene Writer, представляющий собой РНК-матрицу, описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления РНК-матрица имеет 5'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 5'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления 5'-нетранслируемая область содержит псевдоузел, например, псевдоузел, который способен связываться с белком Gene Writer.

В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит последовательность структуры "стебель-петля". В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит шпильку. В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит спираль. В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит псевдоузел. В некоторых вариантах осуществления РНК-матрица содержит рибозим. В некоторых вариантах осуществления рибозим сходен с

рибозимом вируса гепатита дельта (HDV), например, имеет вторичную структуру, сходную с таковой у рибозима HDV, и/или обладает одной или несколькими формами активности рибозима HDV, например, активностью саморасщепления. См., например, Eickbush et al., Molecular and Cellular Biology, 2010, 3142-3150.

В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 3'-нетранслируемая область) содержит одну или несколько структур "стебель-петля" или спиралей. Иллюстративные структуры 3'-UTR R2 показаны, например, в Ruschak et al. "Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs", RNA. 2004 Jun; 10(6): 978-987, например, на фигуре 3 в ней, и в Eikbush and Eikbush, "R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements", Mobile DNA (2012) 3:10; например, на фигуре 3 в ней, при этом содержание данных статей настоящим включено посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления РНК-матрица, описанная в данном документе, содержит последовательность, которая способна связываться с белком Gene Writer, описанным в данном документе. Например, в некоторых вариантах осуществления РНК-матрица содержит последовательность РНК MS2, способную связываться с последовательностью белка оболочки MS2 в белке Gene Writer. В некоторых вариантах осуществления РНК-матрица содержит последовательность РНК, способную связываться с последовательностью B-бокса. В некоторых вариантах осуществления РНК-матрица содержит последовательность РНК (например, последовательность crRNA и/или последовательность tracrRNA, последовательность gRNA), способную связываться с доменом Cas9 в белке Gene Writer. В некоторых вариантах осуществления в дополнение к UTR или вместо нее РНК-матрица связана (например, ковалентной связью) с объектом, отличным от UTR РНК, например, белком или малой молекулой.

В некоторых вариантах осуществления РНК-матрица имеет поли(А)-хвост на 3'-конце. В некоторых вариантах осуществления РНК-матрица не имеет поли(А)-хвост на 3'-конце.

В некоторых вариантах осуществления РНК-матрица имеет 5'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 или больше оснований с по меньшей мере 40%, 50%, 60%, 70%, 80%, 90%, 95% или большей гомологией с 5'-последовательностью ретротранспозона, например, ретротранспозона, описанного в данном документе.

РНК-матрица системы обычно содержит последовательность объекта для вставки в ДНК-мишень. Последовательность объекта может быть кодирующей или некодирующей.

В некоторых вариантах осуществления в системе или способе, описанных в данном документе, предусмотрена одна РНК-матрица. В некоторых вариантах осуществления в системе или способе, описанных в данном документе, предусмотрено множество РНК-матриц.

В некоторых вариантах осуществления последовательность объекта может содержать открытую рамку считывания. В некоторых вариантах осуществления РНК-матрица имеет последовательность Козак. В некоторых вариантах осуществления РНК-матрица имеет участок внутренней посадки рибосомы. В некоторых вариантах осуществления РНК-матрица имеет сайт саморасщепляющегося пептида, такого как T2A или P2A. В некоторых вариантах осуществления РНК-матрица имеет старт-кодон. В некоторых вариантах осуществления РНК-матрица имеет акцепторный сайт сплайсинга. В некоторых вариантах осуществления РНК-матрица имеет донорный сайт сплайсинга. Иллюстративные акцепторные и донорные сайты сплайсинга описаны в WO2016044416, включенной в данный документ посредством ссылки во всей своей полноте. Иллюстративные последовательности акцепторных сайтов сплайсинга известны специалистам в данной области и включают, только в качестве примера, CTGACCCTTCTCTCTCTCCCCCAGAG (из гена HBB человека) и TTTCTCTCCCACAAG (из гена гамма-иммуноглобулина человека). В некоторых вариантах осуществления РНК-матрица имеет сайт связывания с микроРНК ниже стоп-кодона. В некоторых вариантах осуществления РНК-матрица имеет поли(A)-хвост ниже стоп-кодона открытой рамки считывания. В некоторых вариантах осуществления РНК-матрица содержит один или несколько экзонов. В некоторых вариантах осуществления РНК-матрица содержит один или несколько интронов. В некоторых вариантах осуществления РНК-матрица содержит эукариотический терминатор транскрипции. В некоторых вариантах осуществления РНК-матрица содержит энхансерный трансляционный элемент или трансляционный энхансерный элемент. В некоторых вариантах осуществления РНК содержит R-область вируса Т-клеточного лейкоза человека (HTLV-1). В некоторых вариантах осуществления РНК содержит посттранскрипционный регуляторный элемент, который усиливает ядерный экспорт, такой как элемент из вируса гепатита B (HPRE) или вируса гепатита сурков (WPRE). В некоторых вариантах осуществления в РНК-матрице последовательность гетерологичного объекта кодирует полипептид и кодируется в антисмысловом направлении по отношению к 5'- и 3'-UTR. В некоторых вариантах осуществления в РНК-матрице последовательность гетерологичного объекта кодирует полипептид и кодируется в смысловом направлении по отношению к 5'- и 3'-UTR.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК-матрица или ДНК, кодирующая РНК-матрицу), содержит сайт связывания с микроРНК. В некоторых вариантах осуществления сайт связывания с микроРНК используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, сайт связывания с микроРНК может быть выбран на основании того, что он распознается miRNA, которая присутствует в типе клеток, не являющихся мишенями, но которая отсутствует (или присутствует на сниженном уровне по сравнению с клетками, не являющимися мишенями) в типе клеток, являющихся мишенями. Таким образом, если РНК-матрица присутствует в клетке, не являющейся мишенью, то она будет связываться с miRNA, а если РНК-матрица присутствует в клетке-мишени, то она не будет

связываться с miRNA (или будет связываться, но на сниженных уровнях по сравнению с клеткой, не являющейся мишенью). Не желая ограничиваться какой-либо теорией, полагают, что связывание miRNA с РНК-матрицей может препятствовать вставке последовательности гетерологичного объекта в геном. Соответственно, вставка последовательности гетерологичного объекта в геном клеток-мишеней будет происходить более эффективно, чем в геном клеток, не являющихся мишенями. Система, имеющая сайт связывания с микроРНК в РНК-матрице (или кодирующей ее ДНК), также может применяться в комбинации с нуклеиновой кислотой, кодирующей полипептид Gene Writer, при этом экспрессия полипептида Gene Writer регулируется вторым сайтом связывания с микроРНК, например, как описано в данном документе, например, в разделе под названием "Полипептидный компонент системы редактирования генов Gene Writer".

В некоторых вариантах осуществления последовательность объекта может содержать некодирующую последовательность. Например, РНК-матрица может содержать последовательность промотора или энхансера. В некоторых вариантах осуществления РНК-матрица содержит тканеспецифический промотор или энхансер, каждый из которых может быть однонаправленным или двунаправленным. В некоторых вариантах осуществления промотор представляет собой промотор РНК-полимеразы I, промотор РНК-полимеразы II или промотор РНК-полимеразы III. В некоторых вариантах осуществления промотор содержит элемент TATA. В некоторых вариантах осуществления промотор содержит элемент распознавания B. В некоторых вариантах осуществления промотор имеет один или несколько сайтов связывания для факторов транскрипции. В некоторых вариантах осуществления некодирующая последовательность транскрибируется в антисмысловом направлении по отношению к 5'- и 3'-UTR. В некоторых вариантах осуществления некодирующая последовательность транскрибируется в смысловом направлении по отношению к 5'- и 3'-UTR.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК-матрица или ДНК, кодирующая РНК-матрицу), содержит последовательность промотора, например, последовательность тканеспецифического промотора. В некоторых вариантах осуществления тканеспецифический промотор используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, промотор может быть выбран на основании того, что он является активным в типе клеток, являющихся мишенями, но не является активным (или является активным на более низком уровне) в типе клеток, не являющихся мишенями. Таким образом, даже если промотор интегрирован в геном клетки, не являющейся мишенью, он не будет управлять экспрессией (или будет управлять экспрессией только на низком уровне) интегрированного гена. Система, содержащая последовательность тканеспецифического промотора в РНК-матрице, также может применяться в комбинации с сайтом связывания с микроРНК, например, в РНК-матрице или нуклеиновой кислоте, кодирующей белок Gene Writer, например, как описано в данном документе. Система, содержащая последовательность тканеспецифического промотора в РНК-матрице, также

может применяться в комбинации с ДНК, кодирующей полипептид Gene Writer, управляемой тканеспецифическим промотором, например, для достижения более высоких уровней белка Gene Writer в клетках-мишенях, чем в клетках, не являющихся мишенями.

В некоторых вариантах осуществления последовательность гетерологичного объекта, содержащаяся в РНК-матрице (или ДНК, кодирующей РНК-матрицу), функционально связана с по меньшей мере одной регуляторной последовательностью. В некоторых вариантах осуществления последовательность гетерологичного объекта функционально связана с тканеспецифическим промотором, так что экспрессия последовательности гетерологичного объекта, например, терапевтического белка, повышается в клетках-мишенях, как указано выше. В некоторых вариантах осуществления последовательность гетерологичного объекта функционально связана с сайтом связывания miRNA, так что экспрессия последовательности гетерологичного объекта, например, терапевтического белка, понижается в клетках с более высокими уровнями соответствующей miRNA, например, в клетках, не являющихся мишенями, как указано выше.

В некоторых вариантах осуществления РНК-матрица содержит последовательность микроРНК, последовательность siRNA, последовательность направляющей РНК, последовательность РНК, взаимодействующей с Piwi.

В некоторых вариантах осуществления РНК-матрица содержит некодирующую последовательность гетерологичного объекта, например, регуляторную последовательность. Таким образом, в некоторых вариантах осуществления интеграция последовательности гетерологичного объекта приводит к изменению экспрессии эндогенного гена. В некоторых вариантах осуществления интеграция последовательности гетерологичного объекта приводит к повышению экспрессии эндогенного гена. В некоторых вариантах осуществления интеграция последовательности гетерологичного объекта приводит к понижению экспрессии эндогенного гена.

В некоторых вариантах осуществления РНК-матрица содержит сайт, который координирует эпигенетическую модификацию. В некоторых вариантах осуществления РНК-матрица содержит элемент, который ингибирует, например, предотвращает, эпигенетический сайленсинг. В некоторых вариантах осуществления РНК-матрица содержит инсулятор хроматина. Например, РНК-матрица содержит сайт для CTCF или сайт, являющийся мишенью метилирования ДНК.

Для способствования более высокому уровню или более стабильной экспрессии гена РНК-матрица может содержать признаки, которые обеспечивают предотвращение или ингибирование сайленсинга генов. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование метилирования ДНК. В некоторых вариантах осуществления эти признаки способствуют деметилированию ДНК. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование деацетилирования гистонов. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование метилирования гистонов. В некоторых

вариантах осуществления эти признаки способствуют ацетилированию гистонов. В некоторых вариантах осуществления эти признаки способствуют деметилированию гистонов. В некоторых вариантах осуществления в РНК-матрицу могут быть включены несколько признаков для способствования одной или нескольким таким модификациям. CpG-динуклеотиды подвергаются метилированию метилтрансферазами хозяина. В некоторых вариантах осуществления РНК-матрица лишена CpG-динуклеотидов, например, не содержит CpG-нуклеотиды или содержит сниженное количество CpG-динуклеотидов по сравнению с соответствующей неизмененной последовательностью. В некоторых вариантах осуществления промотор, управляющий экспрессией трансгена из интегрированной ДНК, лишен CpG-динуклеотидов.

В некоторых вариантах осуществления РНК-матрица содержит единицу экспрессии гена, состоящую из по меньшей мере одной регуляторной области, функционально связанной с эффекторной последовательностью. Эффекторная последовательность может представлять собой последовательность, которая транскрибируется в РНК (например, кодирующую последовательность или некодирующую последовательность, такую как последовательность, кодирующая микроРНК).

В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень в эндогенном интроне. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень и таким образом выступает в качестве нового экзона. В некоторых вариантах осуществления вставка последовательности объекта в геном-мишень приводит к замене природного экзона или пропуску природного экзона.

В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень в геномном сайте "безопасная гавань", таком как AAVS1, CCR5 или ROSA26. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в локус альбумина. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в локус TRAC. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы добавляется в геном в межгенной или внутригенной области. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы добавляется в геном с 5'- или 3'-стороны в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от эндогенного активного гена. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы добавляется в геном с 5'- или 3'-стороны в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от эндогенного промотора или энхансера. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы может составлять, например, 50-50000 пар оснований (например, 50-40000 п. о., 500-30000 п. о., 500-20000 п. о., 100-15000 п. о., 500-10000 п. о., 50-10000 п. о., 50-5000 п. о.). В некоторых вариантах осуществления

последовательность гетерологичного объекта имеет длину, составляющую менее 1000, 1300, 1500, 2000, 3000, 4000, 5000 или 7500 нуклеотидов.

В некоторых вариантах осуществления геномный сайт "безопасная гавань" представляет собой сайт Natural Harbor™. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой рибосомную ДНК (рДНК). В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой 5S рДНК, 18S рДНК, 5,8S рДНК или 28S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт Mutsu в 5S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт R2, сайт R5, сайт R6, сайт R4, сайт R1, сайт R9 или сайт RT в 28S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт R8 или сайт R7 в 18S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую транспортную РНК (тРНК). В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую тРНК-Asp или тРНК-Glu. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую сплайсосомную РНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую малую ядерную РНК (snRNA), такую как snRNA U2.

Таким образом, в некоторых аспектах в настоящем изобретении предусмотрен способ обеспечения вставки последовательности гетерологичного объекта в сайт Natural Harbor™. В некоторых вариантах осуществления способ включает применение системы Gene Writer, описанной в данном документе, например, применение полипептида из любой из таблиц X, Z1, Z2, 3A или 3B или полипептида, характеризующегося сходством последовательности с ним, например, по меньшей мере 80%, 85%, 90% или 95% идентичностью с ним. В некоторых вариантах осуществления способ включает применение фермента, например, ретротранспозазы, для вставки последовательности гетерологичного объекта в сайт Natural Harbor™. В некоторых аспектах в настоящем изобретении предусмотрена клетка-хозяин, являющаяся клеткой человека, содержащая последовательность гетерологичного объекта (например, последовательность, кодирующую терапевтический полипептид), находящуюся в сайте Natural Harbor™ в геноме клетки. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт, описанный в таблице 4 ниже. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований от последовательности, показанной в таблице 4. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от последовательности, показанной в таблице 4. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в сайт, характеризующийся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью, показанной в таблице 4. В некоторых вариантах осуществления последовательность

гетерологичного объекта вставлена в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований или в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от сайта, характеризующегося по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью, показанной в таблице 4. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в ген, указанный в столбце 5 таблицы 4, или в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований или в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от данного гена.

Таблица 4. Сайты Natural Harbor™. В столбце 1 указан ретротранспозон, который осуществляет вставку в сайт Natural Harbor™. В столбце 2 указан ген в сайте Natural Harbor™. В столбцах 3 и 4 показаны иллюстративные последовательности человеческого генома с 5'- и 3'-стороны от сайта вставки (например, длиной 250 п. о.). В столбцах 5 и 6 приведены иллюстративный символ гена и соответствующий ID гена.

Сайт-мишень	Ген-мишень	5'-фланкирующая последовательность	3'-фланкирующая последовательность	Иллюстративный символ гена	Иллюстративный ID гена
R2	28S рДНК	CCGGTCCCCCCCGCCGG GTCCGCCCCCGGGGCCG CGGTTCCGCGCGGCGCC TCGCCTCGGCCGGCGCC TAGCAGCCGACTTAGAA CTGGTGCGGACCAGGG GAATCCGACTGTTTAAT TAAAACAAAGCATCGCG AAGGCCCGCGGCGGGT GTTGACGCGATGTGATT TCTGCCCAGTGCTCTGA ATGTCAAAGTGAAGAAA TTCAATGAAGCGCGGGT AAACGGCGGGAGTAAC TATGACTCTCTTAAG (SEQ ID NO: 1508)	GTAGCCAAATGCCTCGT CATCTAATTAGTGACGC GCATGAATGGATGAAC GAGATTCCCACTGTCCC TACCTACTATCCAGCGA AACCACAGCCAAGGGA ACGGGCTTGGCGGAATC AGCGGGGAAAGAAGAC CCTGTTGAGCTTGACTCT AGTCTGGCACGGTGAA GAGACATGAGAGGTGT AGAATAAGTGGGAGGC CCCCGGCGCCCCCCCGG TGTCCCCGCGAGGGGCC CGGGGCGGGGTCCGCC G (SEQ ID NO: 1513)	RNA28SN1	106632264

R4	28S рДНК	GCGGTTCCGCGCGGCGC CTCGCCTCGGCCGGCGC CTAGCAGCCGACTTAGA ACTGGTGCGGACCAGG GGAATCCGACTGTTTAA TTAAAACAAAGCATCGC GAAGGCCCGCGGCGGG TGTTGACGCGATGTGAT TTCTGCCCAGTGCTCTG AATGTCAAAGTGAAGAA ATTCAATGAAGCGCGGG TAAACGGCGGGAGTAA CTATGACTCTCTTAAGGT AGCCAAATGCCTCGTCA TCTAATTAGTGACG (SEQ ID NO: 1509)	CGCATGAATGGATGAAC GAGATTCCCACTGTCCC TACCTACTATCCAGCGA AACCACAGCCAAGGGA ACGGGCTTGGCGGAATC AGCGGGGAAAGAAGAC CCTGTTGAGCTTGACTCT AGTCTGGCACGGTGAA GAGACATGAGAGGTGT AGAATAAGTGGGAGGC CCCCGGCGCCCCCCCGG TGTCCCCGCGAGGGGCC CGGGGCGGGGTCCGCC GGCCCTGCGGGCCGCC GGTGAAATACCACTACT C (SEQ ID NO: 1514)	RNA28SN1	106632264
R5	28S рДНК	TCCCCCCCGCCGGGTCC GCCCCCGGGGCCGCGG TTCCGCGCGGCGCCTCG CCTCGGCCGGCGCCTAG CAGCCGACTTAGAACTG GTGCGGACCAGGGGAA TCCGACTGTTTAATTAAA ACAAAGCATCGCGAAG GCCCGCGGCGGGTGTT GACGCGATGTGATTTCT GCCCAGTGCTCTGAATG TCAAAGTGAAGAAATTC AATGAAGCGCGGGTAA ACGGCGGGAGTAACTAT GACTCTCTTAAGGTAG (SEQ ID NO: 1510)	CCAAATGCCTCGTCATCT AATTAGTGACGCGCATG AATGGATGAACGAGATT CCCACTGTCCCTACCTAC TATCCAGCGAAACCACA GCCAAGGGAACGGGCT TGGCGGAATCAGCGGG GAAAGAAGACCCTGTTG AGCTTGACTCTAGTCTG GCACGGTGAAGAGACA TGAGAGGTGTAGAATA AGTGGGAGGCCCCCGG CGCCCCCCCGGTGTCCC CGCGAGGGGCCCGGGG CGGGGTCCGCCGGCCC (SEQ ID NO: 1515)	RNA28SN1	106632264

R9	28S рДНК	CGGCGCGCTCGCCGGCC GAGGTGGGATCCCGAG GCCTCTCCAGTCCGCCG AGGGCGCACCACCGGC CCGTCTCGCCCGCCGCG CCGGGGAGGTGGAGCA CGAGCGCACGTGTTAGG ACCCGAAAGATGGTGAA CTATGCCTGGGCAGGGC GAAGCCAGAGGAAACT CTGGTGGAGGTCCGTAG CGGTCCTGACGTGCAAA TCGGTCGTCCGACCTGG GTATAGGGGCGAAAGA CTAATCGAACCATCTAG (SEQ ID NO: 1511)	TAGCTGGTTCCCTCCGA AGTTTCCCTCAGGATAG CTGGCGCTCTCGCAGAC CCGACGCACCCCCGCCA CGCAGTTTTATCCGGTA AAGCGAATGATTAGAG GTCTTGGGGCCGAAACG ATCTCAACCTATTCTCAA ACTTTAAATGGGTAAGA AGCCCGGCTCGCTGGCG TGGAGCCGGGCGTGGA ATGCGAGTGCCTAGTGG GCCACTTTTGGTAAGCA GAACTGGCGCTGCGGG ATGAACCGAACGCC (SEQ ID NO: 1516)	RNA28SN1	106632264
R8	18S рДНК	GCATTCGTATTGCGCCG CTAGAGGTGAAATTCTT GGACCGGCGCAAGACG GACCAGAGCGAAAGCA TTTGCCAAGAATGTTTTC ATTAATCAAGAACGAAA GTCGGAGGTTCGAAGA CGATCAGATACCGTCGT AGTTCCGACCATAAACG ATGCCGACCGGCGATGC GGCGGCGTTATTCCCAT GACCCGCCGGGCAGCTT CCGGGAAACCAAAGTCT TTGGGTTCCGGGGGGA GTATGGTTGCAAAGC (SEQ ID NO: 1512)	TGAAACTTAAAGGAATT GACGGAAGGGCACCAC CAGGAGTGGAGCCTGC GGCTTAATTTGACTCAA CACGGGAAACCTCACCC GGCCCGGACACGGACA GGATTGACAGATTGATA GCTCTTTCTCGATTCCGT GGGTGGTGGTGCATGG CCGTTCTTAGTTGGTGG AGCGATTTGTCTGGTTA ATTCCGATAACGAACGA GACTCTGGCATGCTAAC TAGTTACGCGACCCCCG AGCGGTCGGCGTCCC (SEQ ID NO: 1517)	RNA18SN1	106631781

R4-2_SRa	тРНК-Asp			TRD-GTC1-1	100189207
LIN25_SM	тРНК-Glu			TRE-CTC1-1	100189384
R1	28S рДНК	TAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGACTGTTTAATTAAAACAAAGCATCGCGAAGGCCCGCGGCGGGTGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCT (SEQ ID NO: 1518)	ACCTACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCCCCCGGCGCCCCCCCGGTGTCCCCGCGAGGGGCCCGGGGCGGGGTCCGCCGGCCCTGCGGGCCGCCGGTGAAATACCACTACTCTGATCGTTTTTTCACTGACCCGGTGAGGCGGGGGG (SEQ ID NO: 1524)	RNA28SN1	106632264
R6	28S рДНК	CCCCCCGCCGGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGGCGCCTCGCCTCGGCCGGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGACTGTTTAATTAAAACAAAGCATCGCGAAGGCCCGCGGCGGGTGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAATGAAGCGCGGGTAAACG	AAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCCCCCGGCGCCCCCCCGGTGTCCCC	RNA28SN1	106632264

		GCGGGAGTAACTATGAC TCTCTTAAGGTAGCC (SEQ ID NO: 1519)	GCGAGGGGCCCGGGGC GGGGTCCGCCGGCCCTG (SEQ ID NO: 1525)		
R7	18S рДНК	GCGCAAGACGGACCAG AGCGAAAGCATTTGCCA AGAATGTTTTCATTAATC AAGAACGAAAGTCGGA GGTTCGAAGACGATCAG ATACCGTCGTAGTTCCG ACCATAAACGATGCCGA CCGGCGATGCGGCGGC GTTATTCCCATGACCCG CCGGGCAGCTTCCGGGA AACCAAAGTCTTTGGGT TCCGGGGGGAGTATGG TTGCAAAGCTGAAACTT AAAGGAATTGACGGAA GGGCACCACCAGGAGT (SEQ ID NO: 1520)	GGAGCCTGCGGCTTAAT TTGACTCAACACGGGAA ACCTCACCCGGCCCGGA CACGGACAGGATTGACA GATTGATAGCTCTTTCTC GATTCCGTGGGTGGTGG TGCATGGCCGTTCTTAG TTGGTGGAGCGATTTGT CTGGTTAATTCCGATAA CGAACGAGACTCTGGCA TGCTAACTAGTTACGCG ACCCCCGAGCGGTCGGC GTCCCCCAACTTCTTAGA GGGACAAGTGGCGTTC AGCCACCCGAG (SEQ ID NO: 1526)	RNA18SN 1	106631 781
RT	28S рДНК	GGCCGGGCGCGACCCG CTCCGGGGACAGTGCCA GGTGGGGAGTTTGACT GGGGCGGTACACCTGTC AAACGGTAACGCAGGT GTCCTAAGGCGAGCTCA GGGAGGACAGAAACCT CCCGTGGAGCAGAAGG GCAAAAGCTCGCTTGAT CTTGATTTTCAGTACGA ATACAGACCGTGAAAGC GGGGCCTCACGATCCTT CTGACCTTTTGGGTTTTA AGCAGGAGGTGTCAGA	AACTGGCTTGTGGCGGC CAAGCGTTCATAGCGAC GTCGCTTTTTGATCCTTC GATGTCGGCTCTTCCTAT CATTGTGAAGCAGAATT CACCAAGCGTTGGATTG TTCACCCACTAATAGGG AACGTGAGCTGGGTTTA GACCGTCGTGAGACAG GTTAGTTTTACCCTACTG ATGATGTGTTGTTGCCA TGGTAATCCTGCTCAGT ACGAGAGGAACCGCAG GTTCAGACATTTGGTGT	RNA28SN 1	106632 264

		AAAGTTACCACAGGGAT (SEQ ID NO: 1521)	ATGTGCTTGGC (SEQ ID NO: 1527)		
Mutsu	5S рДНК	GTCTACGGCCATACCACCC (SEQ ID NO: 1522)	TGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCTTT (SEQ ID NO: 1528)	RNA5S1	100169751
Utopia/Keno	snRNA U2	ATCGCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTA (SEQ ID NO: 1523)	TCTGTTCTTATCAGTTTAATATCTGATACGTCCTCTATCCGAGGACAATATATTAAATGGATTTTTGGAGCAGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGCACCC (SEQ ID NO: 1529)	RNU2-1	6066

В некоторых вариантах осуществления система или способ, описанные в данном документе, приводят к вставке гетерологичной последовательности в сайт-мишень в человеческом геноме. В некоторых вариантах осуществления сайт-мишень в человеческом геноме характеризуется сходством последовательности с соответствующим сайтом-мишенью соответствующей ретротранспозазы дикого типа (например, ретротранспозазы, из которой был получен Gene Writer) в геноме организма, для которого он является нативным. Например, в некоторых вариантах осуществления идентичность между 40 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 40 нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%. В некоторых вариантах осуществления идентичность между 100 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 100 нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%. В некоторых вариантах осуществления идентичность между 500 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 500

нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%.

Компонент системы редактирования генома Gene Writer, представляющий собой нуклеиновую кислоту-матрицу (например, РНК-матрицу), описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) имеет 3'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 3'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer. Область связывания может ассоциировать с нуклеиновой кислотой-матрицей (например, РНК-матрицей) с помощью любого из полипептидных модулей. В некоторых вариантах осуществления область связывания нуклеиновой кислоты-матрицы (например, РНК-матрицы) может ассоциировать с РНК-связывающим доменом в полипептиде. В некоторых вариантах осуществления область связывания нуклеиновой кислоты-матрицы (например, РНК-матрицы) может ассоциировать с доменом обратной транскрипции полипептида (например, специфично связываться с доменом RT). Например, в случае, когда домен обратной транскрипции получен из ретротранспозона, не содержащего LTR, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать область связывания, полученную из ретротранспозона, не содержащего LTR, например, 3'-UTR из ретротранспозона, не содержащего LTR. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может ассоциировать с ДНК-связывающим доменом полипептида, например, в случае с gRNA, ассоциирующей с ДНК-связывающим доменом, полученным из Cas9. В некоторых вариантах осуществления область связывания может также обеспечивать распознавание ДНК-мишени, например, в случае с gRNA, гибридизирующейся с последовательностью ДНК-мишени и связывающей полипептид, например, домен Cas9. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может ассоциировать с несколькими компонентами полипептида, например, с ДНК-связывающим доменом и доменом обратной транскрипции. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать область gRNA, которая ассоциирует с ДНК-связывающим доменом, полученным из Cas9, и 3'-UTR из ретротранспозона, не содержащего LTR, которая ассоциирует с доменом обратной транскрипции, полученным из ретротранспозона, не содержащего LTR.

В некоторых вариантах осуществления в системе или способе, описанных в данном документе, содержится одна нуклеиновая кислота-матрица (например, РНК-матрица). В некоторых вариантах осуществления в системе или способе, описанных в данном документе, содержится множество нуклеиновых кислот-матриц (например, РНК-матриц). Например, в системе, описанной в данном документе, содержится первая РНК, содержащая (например, в направлении от 5' к 3') последовательность, которая связывает полипептид

Gene Writer (например, ДНК-связывающий домен и/или эндонуклеазный домен, например, gRNA) и последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), и вторая РНК (например, РНК-матрица), содержащая (например, в направлении от 5' к 3') необязательно последовательность, которая связывает полипептид Gene Writer (например, которая специфично связывает домен RT), последовательность гетерологичного объекта и 3'-концевой домен гомологии. В некоторых вариантах осуществления, когда в системе содержится множество нуклеиновых кислот, каждая нуклеиновая кислота содержит конъюгирующий домен. В некоторых вариантах осуществления конъюгирующий домен обеспечивает ассоциацию молекул нуклеиновых кислот, например, посредством гибридизации комплементарных последовательностей.

В некоторых вариантах осуществления нуклеиновая кислота-матрица может дополнительно содержать gRNA. В некоторых вариантах осуществления нуклеиновая кислота-матрица может связываться с полипептидом Gene Writer посредством взаимодействия gRNA-части нуклеиновой кислоты-матрицы с гетерологичным РНК-связывающим доменом. В некоторых вариантах осуществления гетерологичный РНК-связывающий домен представляет собой Cas9.

В некоторых вариантах осуществления нуклеиновая кислота-матрица может содержать одну или несколько UTR (например, 5'-UTR или 3'-UTR, например, из ретротранспозона типа R2) и gRNA. В некоторых вариантах осуществления UTR облегчает взаимодействие матрицы с доменом обратной транскриптазы полипептида. В некоторых вариантах осуществления gRNA облегчает взаимодействие со связывающим доменом полипептида. В некоторых вариантах осуществления gRNA направляет полипептид к соответствующей последовательности-мишени. В некоторых вариантах осуществления матрица может содержать только мотив связывания обратной транскриптазы (например, 3'-UTR из R2), и gRNA может быть предоставлена в качестве второй молекулы РНК для распознавания сайта-мишени. В некоторых вариантах осуществления матрица, содержащая RT-связывающий мотив, может существовать в той же молекуле, что и gRNA, но процессироваться с образованием двух молекул РНК благодаря активности расщепления (например, рибозима).

В некоторых вариантах осуществления матрица содержит одну или несколько последовательностей, способствующих ассоциации матрицы с полипептидом Gene Writer. В некоторых вариантах осуществления данные последовательности могут быть получены из UTR ретротранспозона. В некоторых вариантах осуществления UTR могут располагаться так, что они фланкируют желаемую вставочную последовательность. В некоторых вариантах осуществления последовательность с гомологией по отношению к сайту-мишени может быть расположена за пределами одной или обеих UTR. В некоторых вариантах осуществления последовательность с гомологией по отношению к сайту-мишени может отжигаться с последовательностью-мишенью для праймирования обратной транскрипции. В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть

расположены с концевой стороны от последовательности гомологии по отношению к сайту-мишени, например, таким образом, что обратная транскрипция с праймированием мишенью исключает обратную транскрипцию 5'- и/или 3'-UTR. В некоторых вариантах осуществления система Gene Writer может привести к вставке желаемой полезной нагрузки без какой-либо дополнительной последовательности (например, единицы экспрессии гена без UTR, используемой для связывания белка Gene Writer).

Нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана таким образом, чтобы приводить к вставкам, мутациям или делециям в локусе ДНК-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана таким образом, чтобы вызывать вставку в ДНК-мишень. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать гетерологичную последовательность, где обратная транскрипция приведет к вставке гетерологичной последовательности в ДНК-мишень. В других вариантах осуществления РНК-матрица может быть разработана для записи делеции в ДНК-мишень. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может соответствовать ДНК-мишени выше и ниже желаемой делеции, где обратная транскрипция будет приводить к копированию вышерасположенной и нижерасположенной последовательностей нуклеиновой кислоты-матрицы (например, РНК-матрицы) без вставочной последовательности, например, вызывая делецию вставочной последовательности. В других вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана для записи редактирования в ДНК-мишень. Например, РНК-матрица может соответствовать последовательности ДНК-мишени, за исключением одного или нескольких нуклеотидов, где обратная транскрипция приведет к копированию этих редактирований в ДНК-мишени, например, что приведет к мутациям, например, мутациям по типу транзиции или трансверсии.

В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и

необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований).

Дополнительные признаки матрицы

В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые *in silico*. В вариантах осуществления предполагается, что РНК-матрица имеет структуры с минимальной энергией, составляющей от -280 до -480 ккал/моль (например, от -280 до -300, от -300 до -350, от -350 до -400, от -400 до -450, или от -450 до -480 ккал/моль), например, при измерении структуры РНК, например, как описано в Turner and Mathews *Nucleic Acids Res* 38:D280-282 (2009) (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые *in vitro*. В вариантах осуществления РНК-матрица оптимизирована по последовательности, например, для ослабления вторичной структуры, определенной *in vitro*, например, с помощью SHAPE-MaP (например, как описано в Siegfried et al. Nat Methods 11:959-965 (2014); включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые в клетках. В вариантах осуществления РНК-матрица оптимизирована по последовательности, например, для ослабления вторичной структуры, измеренной в клетках, например, с помощью DMS-MaPseq (например, как описано в Zubradt et al. Nat Methods 14:75-82 (2017); включенном в данный документ посредством ссылки во всей своей полноте).

5. Дополнительные функциональные характеристики Gene Writer™

Gene Writer, описанный в данном документе, в некоторых случаях может характеризоваться одним или несколькими функциональными измеряемыми показателями или характеристиками. В некоторых вариантах осуществления ДНК-связывающий домен обладает одной или несколькими функциональными характеристиками, описанными ниже. В некоторых вариантах осуществления РНК-связывающий домен обладает одной или несколькими функциональными характеристиками, описанными ниже. В некоторых вариантах осуществления эндонуклеазный домен обладает одной или несколькими функциональными характеристиками, описанными ниже. В некоторых вариантах осуществления домен обратной транскриптазы обладает одной или несколькими функциональными характеристиками, описанными ниже. В некоторых вариантах

осуществления матрица (например, РНК-матрица) обладает одной или несколькими функциональными характеристиками, описанными ниже. В некоторых вариантах осуществления сайт-мишень, связанный с Gene Writer, обладает одной или несколькими функциональными характеристиками, описанными ниже.

В вариантах осуществления настоящего изобретения предусмотрена молекула нуклеиновой кислоты или система для перенацеливания, например, полипептида или молекулы нуклеиновой кислоты Gene Writer или системы, описанной в данном документе. Перенацеливание (например, полипептида или молекулы нуклеиновой кислоты Gene Writer или системы, описанной в данном документе) обычно включает: (i) направление полипептида на связывание и расщепление в сайте-мишени и/или (ii) разработку РНК-матрицы, комплементарной последовательности-мишени. В некоторых вариантах осуществления РНК-матрица характеризуется комплементарностью по отношению к последовательности-мишени с 5'-стороны от однонитевого разрыва в первой нити, например, таким образом, что 3'-конец РНК-матрицы отжигается, и 5'-конец сайта-мишени служит в качестве праймера, например, для обратной транскрипции с праймированием мишенью (TPRT). В некоторых вариантах осуществления эндонуклеазный домен полипептида и 5'-конец РНК-матрицы также модифицированы, как описано.

6. *Полипептид Gene Writer*

7. *ДНК-связывающий домен*

В некоторых вариантах осуществления ДНК-связывающий домен способен связываться с последовательностью-мишенью (например, последовательностью dsDNA-мишени) с большей аффинностью, чем эталонный ДНК-связывающий домен. В некоторых вариантах осуществления эталонный ДНК-связывающий домен представляет собой ДНК-связывающий домен из R2_BM B. mori. В некоторых вариантах осуществления ДНК-связывающий домен способен связываться с последовательностью-мишенью (например, последовательностью dsDNA-мишени) с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ).

В некоторых вариантах осуществления аффинность ДНК-связывающего домена к его последовательности-мишени (например, последовательности dsDNA-мишени) измеряют in vitro, например, с помощью термофореза, например, описанного в Asmari et al. Methods 146:107-119 (2018) (включенном в данный документ посредством ссылки во всей своей полноте).

В вариантах осуществления ДНК-связывающий домен способен связываться со своей последовательностью-мишенью (например, последовательностью dsDNA-мишени), например, с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ), в присутствии молярного избытка конкурентной dsDNA со скремблированной последовательностью, например, приблизительно 100-кратного молярного избытка.

В некоторых вариантах осуществления ДНК-связывающий домен обнаруживается ассоциированным с последовательностью-мишенью (например, последовательностью

dsDNA-мишени) чаще, чем с любой другой последовательностью в геноме клетки-мишени, например, клетки-мишени человека, например, при измерении с помощью ChIP-seq (например, в клетках HEK293T), например, как описано в He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ДНК-связывающий домен обнаруживается ассоциированным с последовательностью-мишенью (например, последовательностью dsDNA-мишени) в по меньшей мере приблизительно 5 раз или 10 раз чаще, чем с любой другой последовательностью в геноме клетки-мишени, например, при измерении с помощью ChIP-seq (например, в клетках HEK293T), например, как описано в He and Pu (2010), выше.

В некоторых вариантах осуществления полипептид Gene Writer содержит модификацию ДНК-связывающего домена, например, по сравнению с полипептидом дикого типа. В некоторых вариантах осуществления ДНК-связывающий домен предусматривает добавление, делецию, замену или модификацию аминокислотной последовательности исходного ДНК-связывающего домена. В некоторых вариантах осуществления ДНК-связывающий домен модифицирован таким образом, чтобы он содержал гетерологичный функциональный домен, который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В некоторых вариантах осуществления функциональный домен заменяет по меньшей мере часть предшествующего ДНК-связывающего домена полипептида (например, его полностью). В некоторых вариантах осуществления функциональный домен содержит "цинковый палец" (например, "цинковый палец", который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В некоторых вариантах осуществления функциональный домен содержит домен Cas (например, домен Cas, который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В вариантах осуществления домен Cas содержит Cas9 или его мутантную форму или вариант (например, как описано в данном документе). В вариантах осуществления домен Cas ассоциирован с направляющей РНК (gRNA), например, как описано в данном документе. В вариантах осуществления домен Cas направлен на последовательность нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес, посредством gRNA. В вариантах осуществления домен Cas кодируется в той же молекуле нуклеиновой кислоты (например, РНК), что и gRNA. В вариантах осуществления домен Cas кодируется в молекуле нуклеиновой кислоты (например, РНК), отличной от gRNA.

В некоторых вариантах осуществления полипептид Gene Writer содержит модификацию эндонуклеазного домена, например, по сравнению с полипептидом дикого типа. В некоторых вариантах осуществления эндонуклеазный домен предусматривает добавление, делецию, замену или модификацию аминокислотной последовательности исходного эндонуклеазного домена. В некоторых вариантах осуществления эндонуклеазный домен модифицирован таким образом, чтобы он содержал гетерологичный

функциональный домен, который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес, и/или индуцирует ее расщепление эндонуклеазой. В некоторых вариантах осуществления эндонуклеазный домен содержит "цинковый палец". В некоторых вариантах осуществления эндонуклеазный домен содержит домен Cas (например, Cas9 или его мутантную форму или вариант). В вариантах осуществления эндонуклеазный домен, содержащий домен Cas, ассоциирован с направляющей РНК (gRNA), например, как описано в данном документе. В некоторых вариантах осуществления эндонуклеазный домен модифицирован таким образом, чтобы он содержал функциональный домен, который не нацеливается на конкретную последовательность нуклеиновой кислоты-мишени (например, ДНК). В вариантах осуществления эндонуклеазный домен содержит домен Fok1.

В некоторых вариантах осуществления домен обратной транскриптазы (RT) демонстрирует повышенную строгость инициации обратной транскрипции с праймированием мишенью (TPRT), например, по сравнению с эндогенным доменом RT. В некоторых вариантах осуществления домен RT инициирует TPRT, когда 3 нуклеотида в сайте-мишени непосредственно перед однонитевым разрывом первой нити, например, геномная ДНК, праймирующая РНК-матрицу, характеризуются по меньшей мере 66% или 100% комплементарностью по отношению к 3 гомологичным нуклеотидам в РНК-матрице. В некоторых вариантах осуществления домен RT инициирует TPRT при наличии несовпадения менее 5 нуклеотидов (например, несовпадения менее 1, 2, 3, 4 или 5 нуклеотидов) между гомологичной РНК-матрицей и ДНК-мишенью, праймирующей обратную транскрипцию. В некоторых вариантах осуществления домен RT модифицирован таким образом, что строгость в отношении несовпадений при праймировании реакции TPRT увеличивается, например, где домен RT не допускает каких-либо несовпадений или допускает меньшее количество несовпадений в области праймирования по сравнению с доменом RT дикого типа (например, немодифицированным). В некоторых вариантах осуществления домен RT содержит домен RT HIV-1. В вариантах осуществления домен RT HIV-1 инициирует более низкие уровни синтеза даже при несовпадении трех нуклеотидов по сравнению с альтернативным доменом RT (например, как описано в Jamburuthugoda и Eickbush J Mol Biol 407(5):661-672 (2011); включенном в данный документ посредством ссылки во всей своей полноте).

8. *РНК-связывающий домен*

В некоторых вариантах осуществления РНК-связывающий домен способен связываться с РНК-матрицей с большей аффинностью, чем эталонный РНК-связывающий домен. В некоторых вариантах осуществления эталонный РНК-связывающий домен представляет собой РНК-связывающий домен из R2_BM B. mori. В некоторых вариантах осуществления РНК-связывающий домен способен связываться с РНК-матрицей с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ). В некоторых вариантах осуществления аффинность РНК-связывающего

домена к его РНК-матрице измеряют in vitro, например, с помощью термофореза, например, описанного в Asmari et al. Methods 146:107-119 (2018) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления аффинность РНК-связывающего домена к его РНК-матрице измеряют в клетках (например, с помощью FRET или CLIP-Seq).

В некоторых вариантах осуществления РНК-связывающий домен ассоциирован с РНК-матрицей *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной РНК. В некоторых вариантах осуществления частоту ассоциации между РНК-связывающим доменом и РНК-матрицей или скремблированной РНК измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления РНК-связывающий домен ассоциирован с РНК-матрицей в клетках (например, в клетках HEK293T) с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной РНК. В некоторых вариантах осуществления частоту ассоциации между РНК-связывающим доменом и РНК-матрицей или скремблированной РНК измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019), выше.

9. *Эндонуклеазный домен*

В некоторых вариантах осуществления эндонуклеазный домен ассоциирован с dsDNA-мишенью *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной dsDNA. В некоторых вариантах осуществления эндонуклеазный домен ассоциирован с dsDNA-мишенью *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной dsDNA, например, в клетке (например, клетке HEK293T). В некоторых вариантах осуществления частоту ассоциации между эндонуклеазным доменом и ДНК-мишенью или скремблированной ДНК измеряют с помощью ChIP-seq, например, как описано в He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен может катализировать образование однонитевого разрыва в последовательности-мишени, например, с увеличением в по меньшей мере приблизительно 5 раз или 10 раз по сравнению с последовательностью, не являющейся мишенью (например, по сравнению с любой другой геномной последовательностью в геноме клетки-мишени). В некоторых вариантах осуществления уровень образования однонитевых разрывов определяют с использованием NickSeq, например, как описано в Elacqua et al. (2019) *BioRxiv* doi.org/10.1101/867937 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК in vitro. В вариантах осуществления однонитевой разрыв приводит к доступному основанию. В вариантах осуществления доступное основание можно выявить с помощью анализа чувствительности к нуклеазе, например, описанного в

Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809 (включенном в данный документ посредством ссылки во всей своей полноте). В вариантах осуществления уровень доступных оснований (например, выявленный с помощью анализа чувствительности к нуклеазе) увеличен на по меньшей мере 10%, 50% или больше по сравнению с эталонным эндонуклеазным доменом. В некоторых вариантах осуществления эталонный эндонуклеазный домен представляет собой эндонуклеазный домен из R2_BM B. mori.

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК в клетке. В вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК в клетке HEK293T. В вариантах осуществления неисправленный однонитевой разрыв, который подвергается репликации в отсутствие Rad51, приводит к увеличению частоты NHEJ в месте однонитевого разрыва, что может быть выявлено, например, с помощью анализа ингибирования Rad51, например, описанного в Bothmer et al. (2017) *Nat Commun* 8:13905 (включенном в данный документ посредством ссылки во всей своей полноте). В вариантах осуществления частота NHEJ увеличивается выше 0-5%. В вариантах осуществления частота NHEJ увеличивается до 20-70% (например, 30%-60% или 40-50%), например, при ингибировании Rad51.

В некоторых вариантах осуществления эндонуклеазный домен высвобождает мишень после расщепления. В некоторых вариантах осуществления высвобождение мишени определяется косвенно посредством оценки множественных оборотов фермента, например, как описано в Yourik al. RNA 25(1):35-44 (2019) (включенном в данный документ посредством ссылки во всей своей полноте) и показано на фигуре 2. В некоторых вариантах осуществления k_{exp} эндонуклеазного домена составляет 1×10^{-3}-1×10^{-5} мин-1 при измерении такими способами.

В некоторых вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью (k_{cat}/K_m), составляющей более чем приблизительно 1×10^8 c^{-1} M^{-1} in vitro. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью, составляющей более чем приблизительно 1×10^5, 1×10^6, 1×10^7 или 1×10^8 c^{-1} M^{-1} in vitro. В вариантах осуществления каталитическую эффективность определяют, как описано в Chen et al. (2018) *Science* 360(6387):436-439 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью (k_{cat}/K_m) в клетках, составляющей более чем приблизительно 1×10^8 c^{-1} M^{-1}. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью в клетках, составляющей более чем приблизительно 1×10^5, 1×10^6, 1×10^7 или 1×10^8 c^{-1} M^{-1}.

10. *Домен обратной транскриптазы*

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется более низкой вероятностью частоты преждевременной терминации (P_{off}) *in vitro* по сравнению с эталонным доменом обратной транскриптазы. В некоторых вариантах осуществления эталонный домен обратной транскриптазы представляет собой

домен обратной транскриптазы из R2_BM B. mori или вирусный домен обратной транскриптазы, например, домен RT из M-MLV.

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется более низкой вероятностью частоты преждевременной терминации (P_{off}) *in vitro*, составляющей менее чем приблизительно 5×10^{-3}/нуклеотид, 5×10^{-4}/нуклеотид или 5×10^{-6}/нуклеотид, например, при измерении в РНК длиной 1094 нуклеотида. В вариантах осуществления частоту преждевременной терминации *in vitro* определяют, как описано в Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы способен осуществлять по меньшей мере приблизительно 30% или 50% полных интеграций в клетках. Процент полных интеграций может быть измерен путем деления количества событий по сути полноразмерной интеграции (например, геномных сайтов, которые составляют по меньшей мере 98% ожидаемой интегрированной последовательности) на общее количество событий интеграции (включая по сути полноразмерную и частичную) в популяции клеток. В вариантах осуществления интеграцию в клетках определяют (например, по сайту интеграции) с использованием секвенирования ампликонов с длинными ридами, например, как описано в Karst et al. (2020) *BioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте).

В вариантах осуществления количественная оценка интеграций в клетках включает подсчет доли интеграций, которые содержат по меньшей мере приблизительно 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% последовательности ДНК, соответствующей РНК-матрице (например, РНК-матрице длиной по меньшей мере 0,05, 0,1, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 3, 4 или 5 т. о., например, длиной 0,5-0,6, 0,6-0,7, 0,7-0,8, 0,8-0,9, 1,0-1,2, 1,2-1,4, 1,4-1,6, 1,6-1,8, 1,8-2,0, 2-3, 3-4 или 4-5 т. о.).

В некоторых вариантах осуществления домен обратной транскриптазы способен полимеризовать dNTP *in vitro*. В некоторых вариантах осуществления домен обратной транскриптазы способен полимеризовать dNTP *in vitro* со скоростью от 0,1 до 50 нуклеотидов/с (например, 0,1-1, 1-10 или 10-50 нуклеотидов/с). В вариантах осуществления полимеризацию dNTP с помощью домена обратной транскриптазы измеряют с помощью анализа отдельных молекул, например, описанного в Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется частотой ошибок *in vitro* (например, неправильным включением нуклеотидов), составляющей от 1×10^{-3} до 1×10^{-4} или от 1×10^{-4} до 1×10^{-5} замен/нуклеотид, например, как описано в Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления домен обратной транскриптазы характеризуется частотой

ошибок (например, неправильным включением нуклеотидов) в клетках (например, клетках HEK293T), составляющей от 1×10^{-3} до 1×10^{-4} или от 1×10^{-4} до 1×10^{-5} замен/нуклеотид, например, при секвенировании ампликонов с длинными ридами, например, как описано в Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы способен осуществлять обратную транскрипцию РНК-мишени in vitro. В некоторых вариантах осуществления обратной транскриптазе требуется праймер длиной по меньшей мере 3 нуклеотида для инициации обратной транскрипции матрицы. В некоторых вариантах осуществления обратную транскрипцию РНК-мишени определяют посредством выявления кДНК из РНК-мишени (например, при наличии праймера для ssDNA, например, который отжигается с мишенью с по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидами на 3'-конце), например, как описано в Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы выполняет обратную транскрипцию в по меньшей мере 5 или 10 раз более эффективно (например, по продуцированию кДНК), например, при преобразовании своей РНК-матрицы в кДНК, например, по сравнению с РНК-матрицей, не имеющей белок-связывающего мотива (например, 3'-UTR). В вариантах осуществления эффективность обратной транскрипции измеряют, как описано в Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы специфично связывает специфическую РНК-матрицу с большей частотой (например, в приблизительно 5 или 10 раз большей частотой), чем любую эндогенную клеточную РНК, например, при экспрессии в клетках (например, клетках HEK293T). В вариантах осуществления частоту специфичного связывания между доменом обратной транскриптазы и РНК-матрицей измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (включенном в данный документ посредством ссылки во всей своей полноте).

11. *Сайт-мишень*

В некоторых вариантах осуществления после Gene Writing сайт-мишень, окружающий интегрированную последовательность, содержит ограниченное количество вставок или делеций, например, в менее чем приблизительно 50% или 10% событий интеграции, например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления в сайте-мишени не обнаруживается несколько событий вставки, например, дупликаций по типу "голова к хвосту" или "голова к голове", например, как определено с помощью секвенирования ампликонов с длинными ридами сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903

(2020) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице. В некоторых вариантах осуществления сайт-мишень не содержит вставок, полученных из эндогенной РНК, в более чем приблизительно 1% или 10% событий, например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице.

В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице. В вариантах осуществления сайт-мишень не содержит последовательности за пределами матрицы для RT (например, каркаса gRNA, остова вектора и/или ITR), например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (включенном в данный документ посредством ссылки во всей своей полноте).

12. *Варианты Gene Writer, полученные методом направленной эволюции*

В некоторых вариантах осуществления в настоящем изобретении предусмотрены варианты Gene Writer, полученные методом направленной эволюции. В некоторых вариантах осуществления варианты, полученные методом направленной эволюции, могут быть получены посредством мутагенеза эталонного Gene Writer или одного из содержащихся в нем фрагментов или доменов. В некоторых вариантах осуществления метод направленной эволюции применяют в отношении одного или нескольких доменов (например, домена обратной транскриптазы, ДНК-связывающего домена (в том числе, например, направляемых последовательностью ДНК-связывающих элементов), РНК-связывающего домена или эндонуклеазного домена). В некоторых вариантах осуществления один или несколько таких вариантов доменов, полученных методом направленной эволюции, могут быть получены в результате применения метода направленной эволюции к ним отдельно или совместно с другими доменами. Вариант домена или домены, полученные методом направленной эволюции, могут в некоторых вариантах осуществления быть объединены с когнатным(когнатными) компонентом(компонентами), к которым не был применен метод направленной эволюции, или вариантами когнатного(когнатных) компонента(компонентов), полученными методом направленной эволюции, например, при получении которых методом направленной эволюции мог применяться либо параллельный, либо последовательный подход.

В некоторых вариантах осуществления процесс мутагенеза эталонного Gene Writer или его фрагмента или домена предусматривает мутагенез эталонного Gene Writer или его фрагмента или домена. В вариантах осуществления мутагенез предусматривает способ непрерывной эволюции (например, PACE) или способ прерывистой эволюции (например, PANCE), например, как описано в данном документе. В некоторых вариантах

осуществления Gene Writer, полученный методом направленной эволюции, или его фрагмент или домен содержат одну или несколько аминокислотных вариаций, введенных в их аминокислотную последовательность, по сравнению с аминокислотной последовательностью эталонного Gene Writer или его фрагмента или домена. В вариантах осуществления вариации аминокислотной последовательности могут предусматривать один или несколько мутантных остатков (например, консервативные замены, неконсервативные замены или их комбинацию) в аминокислотной последовательности эталонного Gene Writer, полученные, например, в результате изменения в последовательности нуклеотидов, кодирующей Gene Writer, которое приводит, например, к изменению кодона в любом конкретном положении в кодирующей последовательности, делеции одной или нескольких аминокислот (например, в случае усеченного белка), вставке одной или нескольких аминокислот или любой комбинации изложенного выше. Вариант Gene Writer, полученный методом направленной эволюции, может предусматривать наличие вариантов в одном или нескольких компонентах или доменах Gene Writer (например, варианты, введенные в домен обратной транскриптазы, эндонуклеазный домен, ДНК-связывающий домен, РНК-связывающий домен или их комбинации).

В некоторых аспектах в настоящем изобретении предусмотрены Gene Writer, системы, наборы и способы, в которых используется или содержится вариант Gene Writer, полученный методом направленной эволюции, например, в которых используется вариант Gene Writer, полученный методом направленной эволюции, или Gene Writer, который получен или который может быть получен с помощью PACE или PANCE. В вариантах осуществления эталонный Gene Writer, к которому не был применен метод направленной эволюции, представляет собой Gene Writer, раскрытый в данном документе.

Термин "непрерывная эволюция, осуществляемая с использованием фагов (PACE)", используемый в данном документе, обычно относится к методу непрерывной направленной эволюции, в котором используются фаги в качестве вирусных векторов. Примеры технологии PACE были описаны, например, в международной заявке согласно PCT № PCT/US 2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной заявке согласно PCT PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; патенте США № 9023594, выданном 5 мая 2015 г.; патенте США № 9771574, выданном 26 сентября 2017 г.; патенте США № 9394537, выданном 19 июля 2016 г.; международной заявке согласно PCT PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г.; патенте США № 10179911, выданном 15 января 2019 г.; и международной заявке согласно PCT PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., содержание каждого из которых включено в данный документ посредством ссылки во всей своей полноте.

Термин "прерывистая эволюция, осуществляемая с использованием фагов (PANCE)", используемый в данном документе, обычно относится к методу прерывистой направленной эволюции, в котором используются фаги в качестве вирусных векторов.

Примеры технологии PANCE были описаны, например, в Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017), включенном в данный документ посредством ссылки во всей своей полноте. Вкратце, PANCE представляет собой методику быстрой направленной эволюции in vivo с использованием серийных переносов в колбах фага для селекции (SP), подвергающегося действию направленной эволюции, который содержит ген, представляющий интерес, который необходимо подвергнуть действию направленной эволюции, в свежих клетках-хозяевах (например, клетках E. coli). Гены внутри клетки-хозяина могут поддерживаться в неизменном состоянии, в то время как гены, содержащиеся в SP - непрерывно эволюционировать. После обеспечения роста фага аликвоту инфицированных клеток можно использовать для трансфекции следующей колбы, содержащей E. coli в качестве хозяина. Этот процесс можно повторять и/или продолжать до тех пор, пока в результате направленной эволюции не будет получен необходимый фенотип, например, столько переносов, сколько необходимо.

Способы применения PACE и PANCE к Gene Writer могут быть легко оценены специалистом в данной области посредством ссылки, помимо прочего, на предшествующие литературные источники. Дополнительные иллюстративные способы управления непрерывной направленной эволюцией белков или систем, модифицирующих геном, например, в популяции клеток-хозяев, например, с использованием фаговых частиц, могут применяться для создания вариантов Gene Writer, полученных методом направленной эволюции, или их фрагментов или субдоменов. Неограничивающие примеры таких способов описаны в международной заявке согласно PCT PCT/US2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной заявке согласно PCT PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; патенте США № 9023594, выданном 5 мая 2015 г.; патенте США № 9771574, выданном 26 сентября 2017 г.; патенте США № 9394537, выданном 19 июля 2016 г.; международной заявке согласно PCT PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г.; патенте США № 10179911, выданном 15 января 2019 г.; международной заявке № PCT/US2019/37216, поданной 14 июня 2019 г., международной патентной публикации WO 2019/023680, опубликованной 31 января 2019 г., международной заявке согласно PCT PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., и международной патентной публикации № PCT/US2019/47996, поданной 23 августа 2019 г., каждый из которых включен в данный документ посредством ссылки во всей своей полноте.

В некоторых неограничивающих иллюстративных вариантах осуществления способ направленной эволюции в случае варианта Gene Writer, полученного методом направленной эволюции, его фрагмента или домена включает: (a) приведение популяции клеток-хозяев в контакт с популяцией вирусных векторов, содержащих ген, представляющий интерес (исходный Gene Writer или его фрагмент или домен), где (1)

клетка-хозяин поддается инфицированию вирусным вектором; (2) в клетке-хозяине экспрессируются вирусные гены, необходимые для образования вирусных частиц; (3) экспрессия по меньшей мере одного вирусного гена, необходимого для продуцирования инфекционной вирусной частицы, зависит от функции гена, представляющего интерес; и/или (4) вирусный вектор обеспечивает экспрессию белка в клетке-хозяине и может реплицироваться и упаковываться в вирусную частицу клеткой-хозяином; В некоторых вариантах осуществления способ включает (b) приведение клеток-хозяев в контакт с мутагеном, использование клеток-хозяев с мутациями, которые повышают скорость мутаций (например, за счет наличия мутационной плазмиды или некоторой модификации генома, например, ДНК-полимеразы с нарушенной функцией исправления ошибок, генов SOS, таких как UmuC, UmuD' и/или RecA, мутации которых, если они связаны с плазмидой, могут находиться под контролем индуцируемого промотора), или их комбинации. В некоторых вариантах осуществления способ включает (c) инкубирование популяции клеток-хозяев в условиях, обеспечивающих репликацию вируса и продуцирование вирусных частиц, где клетки-хозяева удаляют из популяции клеток-хозяев, а свежие, неинфицированные клетки-хозяева вводят в популяцию клеток-хозяев, за счет чего пополняется популяция клеток-хозяев и создается поток клеток-хозяев. В некоторых вариантах осуществления клетки инкубируют в условиях, позволяющих гену, представляющему интерес, приобрести мутацию. В некоторых вариантах осуществления способ дополнительно включает (d) выделение мутантного варианта вирусного вектора, кодирующего продукт гена, полученный методом направленной эволюции (например, вариант Gene Writer, полученный методом направленной эволюции, или его фрагмент или домен), из популяции клеток-хозяев.

Специалисту в данной области будет известно множество признаков, которые можно использовать в рамках вышеописанной структуры. Например, в некоторых вариантах осуществления вирусный вектор или фаг представляет собой нитевидный фаг, например, фаг M13, например, фаг M13 для селекции. В некоторых вариантах осуществления ген, необходимый для продуцирования инфекционных вирусных частиц, представляет собой ген III M13 (gIII). В вариантах осуществления фаг может не содержать функционального gIII, но при этом содержать gI, gII, gIV, gV, gVI, gVII, gVIII, gIX и gX. В некоторых вариантах осуществления образование инфекционных частиц VSV предусматривает наличие белка оболочки VSV-G. В различных вариантах осуществления могут использоваться разные ретровирусные векторы, например, векторы на основе вируса лейкоза мышей или лентивирусные векторы. В вариантах осуществления ретровирусные векторы могут быть эффективно упакованы с использованием белка оболочки VSV-G, например, в качестве замены нативного белка оболочки вируса.

В некоторых вариантах осуществления клетки-хозяева инкубируют в соответствии с подходящим количеством жизненных циклов вируса, например, по меньшей мере 10, по меньшей мере 20, по меньшей мере 30, по меньшей мере 40, по меньшей мере 50, по меньшей мере 100, по меньшей мере 200, по меньшей мере 300, по меньшей мере 400, по

меньшей мере 500, по меньшей мере 600, по меньшей мере 700, по меньшей мере 800, по меньшей мере 900, по меньшей мере 1000, по меньшей мере 1250, по меньшей мере 1500, по меньшей мере 1750, по меньшей мере 2000, по меньшей мере 2500, по меньшей мере 3000, по меньшей мере 4000, по меньшей мере 5000, по меньшей мере 7500, по меньшей мере 10000 или больше последовательных жизненных циклов вируса, который в иллюстративных и не ограничивающих примерах с фагом M13 составляет 10-20 минут на жизненный цикл вируса. Аналогичным образом, условия можно модулировать для регулирования времени, в течение которого клетка-хозяин остается в популяции клеток-хозяев, например, приблизительно 10, приблизительно 11, приблизительно 12, приблизительно 13, приблизительно 14, приблизительно 15, приблизительно 16, приблизительно 17, приблизительно 18, приблизительно 19, приблизительно 20, приблизительно 21, приблизительно 22, приблизительно 23, приблизительно 24, приблизительно 25, приблизительно 30, приблизительно 35, приблизительно 40, приблизительно 45, приблизительно 50, приблизительно 55, приблизительно 60, приблизительно 70, приблизительно 80, приблизительно 90, приблизительно 100, приблизительно 120, приблизительно 150 или приблизительно 180 минут. Популяции клеток-хозяев можно частично регулировать по плотности клеток-хозяев или в некоторых вариантах осуществления по плотности клеток-хозяев в поступающей среде, например, 10^3 клеток/мл, приблизительно 10^4 клеток/мл, приблизительно 10^5 клеток/мл, приблизительно 5-10^5 клеток/мл, приблизительно 10^6 клеток/мл, приблизительно 5-10^6 клеток/мл, приблизительно 10^7 клеток/мл, приблизительно 5-10^7 клеток/мл, приблизительно 10^8 клеток/мл, приблизительно 5-10^8 клеток/мл, приблизительно 10^9 клеток/мл, приблизительно 5-10^9 клеток/мл, приблизительно 10^{10} клеток/мл или приблизительно 5-10^{10} клеток/мл.

13. *Промоторы*

В некоторых вариантах осуществления один или несколько промоторных или энхансерных элементов функционально связаны с нуклеиновой кислотой, кодирующей белок Gene Writer, или нуклеиновой кислотой-матрицей, например, которая контролирует экспрессию последовательности гетерологичного объекта. В некоторых вариантах осуществления один или несколько промоторных или энхансерных элементов содержат специфичные для типа клеток или тканеспецифические элементы. В некоторых вариантах осуществления промотор или энхансер является таким же, как промотор или энхансер, который в естественных условиях контролирует экспрессию последовательности гетерологичного объекта, или получен из него. Например, промотор и энхансер гена орнитинтранскарбамилазы можно применять для контроля экспрессии гена орнитинтранскарбамилазы в системе или способе, предусмотренных в настоящем изобретении, для коррекции дефицита орнитинтранскарбамилазы. В некоторых вариантах осуществления промотор для применения в настоящем изобретении предназначен для гена, описанного в любой из таблиц **9-22**, например, который можно применять с аллелем эталонного гена или, в других вариантах осуществления, с гетерологичным геном. В

некоторых вариантах осуществления промотор представляет собой промотор из таблицы 33 или его функциональный фрагмент или вариант.

Иллюстративные тканеспецифические промоторы, которые являются коммерчески доступными, можно найти, например, по унифицированному указателю ресурса (например, https://www.invivogen.com/tissue-specific-promoters). В некоторых вариантах осуществления промотор представляет собой **нативный промотор** или минимальный промотор, например, состоящий из одного фрагмента из 5'-области данного гена. В некоторых вариантах осуществления нативный промотор содержит коровый промотор и его природную 5'-UTR. В некоторых вариантах осуществления 5'-UTR содержит интрон. В других вариантах осуществления они включают **составные промоторы, в которых** объединены промоторные элементы различного происхождения или которые были получены посредством сборки дистального энхансера с минимальным промотором того же происхождения.

Иллюстративные клеточно- или тканеспецифические промоторы представлены в таблицах ниже, и иллюстративные последовательности нуклеиновых кислот, кодирующие их, известны из уровня техники и могут быть легко доступны с использованием различных ресурсов, таких как база данных NCBI, в том числе RefSeq, а также Eukaryotic Promoter Database (//epd.epfl.ch//index.php).

Таблица 33. Иллюстративные клеточно- или тканеспецифические промоторы

Промотор	**Клетки-мишени**
Промотор B29	B-клетки
Промотор CD14	Моноцитарные клетки
Промотор CD43	Лейкоциты и тромбоциты
Промотор CD45	Гемопоэтические клетки
Промотор CD68	Макрофаги
Промотор гена десмина	Мышечные клетки
Промотор гена эластазы-1	Ацинарные клетки поджелудочной железы
Промотор гена эндоглина	Эндотелиальные клетки
Промотор гена фибронектина	Дифференцирующиеся клетки, рубцовая ткань
Промотор Flt-1	Эндотелиальные клетки
Промотор GFAP	Астроциты
Промотор GPIIB	Мегакариоциты
Промотор ICAM-2	Эндотелиальные клетки
Промотор INF-бета	Гемопоэтические клетки
Промотор Mb	Мышечные клетки
Промотор Nphs1	Подоциты
Промотор OG-2	Остеобласты, одонтобласты

Промотор SP-B	Легкое
Промотор Syn1	Нейроны
Промотор WASP	Гемопоэтические клетки
Промотор SV40/bAlb	Печень
Промотор SV40/bAlb	Печень
Промотор SV40/Cd3	Лейкоциты и тромбоциты
Промотор SV40/CD45	Гемопоэтические клетки
Промотор NSE/RU5'	Зрелые нейроны

Таблица 34. Дополнительные иллюстративные клеточно- или тканеспецифические промоторы

Промотор	**Описание гена**	**Специфичность гена**
APOA2	Аполипопротеин A-II	Гепатоциты (из предшественников гепатоцитов)
SERPINA1 (hAAT)	Ингибитор пептидаз SERPIN, клада A (альфа-1 антипротеиназа, антитрипсин), представитель 1 (также называемый альфа-1-антитрипсином)	Гепатоциты (из стадии дефинитивной энтодермы)
CYP3A	Цитохром P450, семейство 3, подсемейство A, полипептид	Зрелые гепатоциты
MIR122	МикроРНК 122	Гепатоциты (из клеток эмбриональной печени на ранней стадии) и энтодерма
Промоторы, специфичные для поджелудочной железы		

Промотор	Описание гена	Специфичность гена
INS	Инсулин	Бета-клетки поджелудочной железы (из стадии дефинитивной энтодермы)
IRS2	Субстрат 2 рецептора инсулина	Бета-клетки поджелудочной железы
Pdx1	поджелудочной железы и двенадцатиперстной кишки Гомеобокс 1	Поджелудочная железа (из стадии дефинитивной энтодермы)
Alx3	Гомеобокс 3, подобный Aristaless	Бета-клетки поджелудочной железы (из стадии дефинитивной энтодермы)
Ppy	Панкреатический полипептид	PP-клетки поджелудочной железы (гамма-клетки)
Промоторы, специфичные для сердца		
Промотор	**Описание гена**	**Специфичность гена**
Myh6 (aMHC)	Миозин, тяжелая цепь 6, сердечная мышца, альфа	Маркер поздней дифференцировки клеток сердечной мышцы (предсердная специфичность)
MYL2 (MLC-2v)	Миозин, легкая цепь 2, регуляторный, сердечный, медленный	Маркер поздней дифференцировки клеток сердечной мышцы (желудочковая специфичность)

ITNNI3 (cTnl)	Тропонин I 3 типа (сердечный)	Кардиомиоциты (из незрелого состояния)
ITNNI3 (cTnl)	Тропонин I 3 типа (сердечный)	Кардиомиоциты (из незрелого состояния)
NPPA (ANF)	Предшественник натрийуретического пептида A (также называемый предсердным натрийуретическим фактором)	Предсердная специфичность во взрослых клетках
Slc8a1 (Ncx1)	Семейство 8 переносчиков растворенных веществ (натрий/кальциевый обменник), представитель 1	Кардиомиоциты из ранних стадий развития
Промоторы, специфичные для ЦНС		
Промотор	**Описание гена**	**Специфичность гена**
SYN1 (hSyn)	Синапсин I	Нейроны
GFAP	Глиальный фибриллярный кислый белок	Астроциты
INA	Белок промежуточных филаментов нейронов альфа-интернексин (a-интернексин)	Нейропредшественники
NES	Нестин	Нейропредшественники и эктодерма
MOBP	Основной белок олигодендроцитов, ассоциированный с миелином	Олигодендроциты
MBP	Основной белок миелина	Олигодендроциты
THT	Тирозингидроксилаза	Дофаминергические нейроны

FOXA2 (HNF3-бета)	Бокс Forkhead A2	Дофаминергические нейроны (также используется в качестве маркера энтодермы)
Промоторы, специфичные для кожи		
Промотор	**Описание гена**	**Специфичность гена**
FLG	Филаггрин	Кератиноциты из зернистого слоя
K14	Кератин 14	Кератиноциты из зернистого и базального слоев
TGM3	Трансглютаминаза 3	Кератиноциты из зернистого слоя
Промоторы, специфичные для иммунных клеток		
Промотор	**Описание гена**	**Специфичность гена**
ITGAM (CD11B)	Интегрин, альфа-M (субъединица рецептора 3 компонента 3 системы комплемента)	Моноциты, макрофаги, гранулоциты, естественные клетки-киллеры
Промоторы, специфичные для клеток мочеполовой системы		
Промотор	**Описание гена**	**Специфичность гена**
Pbsn	Пробасин	Эпителий предстательной железы
Upk2	Уроплакин 2	Мочевой пузырь
Sbp	Спермин-связывающий белок	Предстательная железа
Fer1L4	Fer-1-подобный белок 4	Мочевой пузырь
Промоторы, специфичные для эндотелиальных клеток		
Промотор	**Описание гена**	**Специфичность гена**
ENG	Эндоглин	Эндотелиальные клетки

Промоторы, специфичные для плюрипотентных и эмбриональных клеток		
Промотор	**Описание гена**	**Специфичность гена**
Oct4 (POU5F1)	Гомеобокс 1 POU класса 5	Плюрипотентные клетки (зародышевые клетки, ES-клетки, iPS-клетки)
NANOG	Гомеобокс Nanog	Плюрипотентные клетки (ES-клетки, iPS-клетки)
Синтетический Oct4	Синтетический промотор на основе корового энхансерного элемента Oct-4	Плюрипотентные клетки (ES-клетки, iPS-клетки)
Brachyury T	Brachyury	Мезодерма
NES	Нестин	Нейропредшественники и эктодерма
SOX17	SRY (область Y, определяющая пол)-бокс 17	Энтодерма
FOXA2 (HNFJ бета)	Бокс Forkhead A2	Энтодерма (также используется в качестве маркера дофаминергических нейронов)
MIR122	МикроРНК 122	Энтодерма и гепатоциты (из клеток эмбриональной печени на ранней стадии)

В зависимости от используемой системы хозяин/вектор в векторе экспрессии можно использовать любой из ряда пригодных элементов, контролирующих транскрипцию и трансляцию, включая конститутивные и индуцируемые промоторы, энхансерные элементы транскрипции, терминаторы транскрипции и т. д. (см., например, Bitter et al. (1987) *Methods in Enzymology,* 153:516-544; включенный в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления нуклеиновая кислота, кодирующая Gene Writer или нуклеиновую кислоту-матрицу, функционально связана с контролирующим

элементом, например, элементом, контролирующим транскрипцию, таким как промотор. В некоторых вариантах осуществления элемент, контролирующий транскрипцию, может быть функциональным в эукариотической клетке, например, в клетке млекопитающего; либо в прокариотической клетке (например, бактериальной клетке или клетке археи). В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая полипептид, функционально связана с несколькими контролирующими элементами, например, которые обеспечивают экспрессию нуклеотидной последовательности, кодирующей полипептид, как в прокариотических, так и в эукариотических клетках.

Для целей иллюстрации примеры пространственно ограниченных промоторов включают без ограничения промоторы, специфичные для нейронов, промоторы, специфичные для адипоцитов, промоторы, специфичные для кардиомиоцитов, промоторы, специфичные для гладких мышц, промоторы, специфичные для фоторецепторов, и т. д. Специфичные для нейронов пространственно ограниченные промоторы включают без ограничения промотор гена нейронспецифической енолазы (NSE) (см., например, HSENO2, X51956 в EMBL); промотор гена декарбоксилазы ароматических аминокислот (AADC), промотор гена нейрофиламента (см., например, HUMNFL, L04147 в GenBank); промотор гена синапсина (см., например, HUMSYNIB, M55301 в GenBank); промотор thy-1 (см., например, Chen et al. (1987) Cell 51:7-19; и Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166); промотор гена серотонинового рецептора (см., например, S62283 в GenBank); промотор гена тирозингидроксилазы (TH) (см., например, Oh et al. (2009) Gene Ther 16:437; Sasaoka et al. (1992) Mol. Brain Res. 16:274; Boundy et al. (1998) J. Neurosci. 18:9989; и Kaneda et al. (1991) Neuron 6:583-594); промотор GnRH (см., например, Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406); промотор L7 (см., например, Oberdick et al. (1990) Science 248:223-226); промотор DNMT (см., например, Bartge et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652); промотор гена энкефалина (см., например, Comb et al. (1988) EMBO J. 17:3793-3805); промотор гена основного белка миелина (MBP); промотор гена Ca2+/кальмодулин-зависимой протеинкиназы II-альфа (CamKIIα) (см., например, Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250 и Casanova et al. (2001) Genesis 31:37); энхансер CMV/промотор гена тромбоцитарного фактора роста-β (см., например, Liu et al. (2004) Gene Therapy 11:52-60) и т. п.

Специфичные для адипоцитов пространственно ограниченные промоторы включают без ограничения промотор/энхансер гена aP2, например, область от -5,4 т. о. до +21 п. о. гена aP2 человека (см., например, Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590; и Pavjani et al. (2005) Nat. Med. 11:797); промотор гена транспортера глюкозы 4 типа (GLUT4) (см., например, Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); промотор гена транслоказы жирных кислот (FAT/CD36) (см., например, Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; и Sato et al. (2002) J. Biol. Chem. 277:15703); промотор гена стеароил-CoA-десатуразы-1 (SCD1) (Tabor et al. (1999) J. Biol. Chem. 274:20603); промотор гена лептина (см., например, Mason et al. (1998) Endocrinol. 139:1013; и Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); промотор

гена адипонектина (см., например, Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; и Chakrabarti (2010) Endocrinol. 151:2408); промотор гена адипсина (см., например, Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); промотор гена резистина (см., например, Seo et al. (2003) Molec. Endocrinol. 17:1522) и т. п.

Специфичные для кардиомиоцитов пространственно ограниченные промоторы включают без ограничения контролирующие последовательности, полученные из следующих генов: легкой цепи-2 миозина, тяжелой цепи α-миозина, AE3, сердечного тропонина C, сердечного актина и т. п. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; и Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Специфичные для гладких мышц пространственно ограниченные промоторы включают без ограничения промотор SM22α (см., например, Akyürek et al. (2000) Mol. Med. 6:983 и патент США № 7169874); промотор гена смутелина (см., например, WO 2001/018048); промотор гена α-гладкомышечного актина и т. п. Например, было показано, что область промотора SM22α длиной 0,4 т. о., внутри которой находятся два элемента CArG, опосредует экспрессию, специфичную для гладкомышечных клеток сосудов (см., например, Kim et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; и Moessler, et al. (1996) Development 122, 2415-2425).

Специфичные для фоторецепторов пространственно ограниченные промоторы включают без ограничения промотор гена родопсина; промотор гена родопсинкиназы (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); промотор гена бета-фосфодиэстеразыr (Nicoud et al. (2007) J. Gene Med. 9:1015); промотор гена пигментного ретинита (Nicoud et al. (2007), выше); энхансер гена межфоторецепторного ретиноидсвязывающего белка (IRBP) (Nicoud et al. (2007), выше); промотор гена IRBP (Yokoyama et al. (1992) Exp Eye Res. 55:225) и т. п.

Неограничивающие иллюстративные клеточноспецифические промоторы

Клеточноспецифические промоторы, известные из уровня техники, могут использоваться для управления экспрессией белка Gene Writer, например, как описано в данном документе. Неограничивающие иллюстративные клеточноспецифические промоторы млекопитающих были охарактеризованы и использовались у мышей, экспрессирующих Cre-рекомбиназу клеточноспецифическим образом. Определенные неограничивающие иллюстративные промоторы, специфичные для клеток млекопитающих, указаны в таблице 1 документа US9845481, включенного в данный документ посредством ссылки.

В некоторых вариантах осуществления клеточноспецифические промоторы представляют собой промоторы, активные в растениях. Из уровня техники известно множество иллюстративных клеточноспецифических промоторов растений. См., например, патенты США №№ 5097025; 5783393; 5880330; 5981727; 7557264; 6291666; 7132526 и 7323622; а также публикации заявок на патенты США №№ 2010/0269226; 2007/0180580;

2005/0034192 и 2005/0086712, которые включены в данный документ посредством ссылки во всей своей полноте для любых целей.

В некоторых вариантах осуществления вектор, описанный в данном документе, содержит кассету экспрессии. Термин "кассета экспрессии", используемый в данном документе, относится к конструкции на основе нуклеиновой кислоты, содержащей элементы, представляющие собой нуклеиновую кислоту, достаточные для экспрессии молекулы нуклеиновой кислоты по настоящему изобретению. Как правило, кассета экспрессии содержит молекулу нуклеиновой кислоты по настоящему изобретению, функционально связанную с промоторной последовательностью. Термин "функционально связанный" относится к ассоциации двух или более фрагментов нуклеиновой кислоты на одном фрагменте нуклеиновой кислоты, так что функция одного влияет на функцию другого. Например, промотор функционально связан с кодирующей последовательностью, если он способен влиять на экспрессию этой кодирующей последовательности (например, кодирующая последовательность находится под транскрипционным контролем промотора). Кодирующие последовательности могут быть функционально связаны с регуляторными последовательностями в смысловой или антисмысловой ориентации. В определенных вариантах осуществления промотор представляет собой гетерологичный промотор. Термин "гетерологичный промотор", используемый в данном документе, относится к промотору, для которого не обнаружена функциональная связь с данной кодирующей последовательностью в природе. В определенных вариантах осуществления кассета экспрессии может содержать дополнительные элементы, например, интрон, энхансер, сайт полиаденилирования, элемент ответа лесных сурков (WRE) и/или другие элементы, о которых известно, что они влияют на уровни экспрессии кодирующей последовательности. "Промотор", как правило, контролирует экспрессию кодирующей последовательности или функциональной РНК. В определенных вариантах осуществления промоторная последовательность содержит проксимальные и более дистальные вышерасположенные элементы и может дополнительно содержать энхансерный элемент. "Энхансер", как правило, может стимулировать активность промотора и может быть внутренним элементом промотора или гетерологичным элементом, вставленным для повышения уровня активности или тканевой специфичности промотора. В определенных вариантах осуществления промотор во всей своей полноте получен из нативного гена. В определенных вариантах осуществления промотор состоит из различных элементов, полученных из различных встречающихся в природе промоторов. В определенных вариантах осуществления промотор содержит синтетическую нуклеотидную последовательность. Специалистам в данной области будет понятно, что различные промоторы будут управлять экспрессией гена в различных тканях или типах клеток, или на различных стадиях развития, или в ответ на различные условия окружающей среды или на присутствие или отсутствие лекарственного средства или кофактора транскрипции. Повсеместно распространенные, специфичные для типа клеток, тканеспецифические, специфичные для стадии развития и условные промоторы, например, промоторы,

отвечающие на лекарственное средство (например, промоторы, отвечающие на тетрациклин), хорошо известны специалистам в данной области. Примеры промотора включают без ограничения промотор гена фосфоглицераткиназы (PKG), CAG (составную молекулу из энхансера CMV, промотора гена бета-актина курицы (CBA) и интрона бета-глобина кролика), NSE (нейронспецифической енолазы), промоторы генов синапсина или NeuN, ранний промотор SV40, промотор LTR вируса опухоли молочной железы мышей; главный поздний промотор аденовируса (MLP Ad); промотор вируса простого герпеса (HSV), промотор цитомегаловируса (CMV), такой как немедленно-ранняя промоторная область CMV (CMVIE), промотор SFFV, промотор вируса саркомы Рауса (RSV), синтетические промоторы, гибридные промоторы и т. п. Другие промоторы могут происходить от человека или от других видов, в том числе от мышей. Обычные промоторы включают, например, промотор гена немедленно-раннего ответа цитомегаловируса человека (CMV), ранний промотор SV40, длинный концевой повтор вируса саркомы Рауса, промотор гена бета-актина, промотор гена инсулина крыс, промотор гена фосфоглицераткиназы, промотор гена альфа-1-антитрипсина человека (hAAT), промотор гена транстиретина, промотор TBG и другие промоторы, специфичные для печени, промотор гена десмина и аналогичные промоторы, специфичные для мышц, промотор EF1-альфа, промотор CAG и другие конститутивные промоторы, гибридные промоторы со специфичностью для нескольких тканей, промоторы, специфичные для нейронов, такие как промотор гена синапсина и глицеральдегид-3-фосфатдегидрогеназы, все из которых являются промоторами, хорошо известными и легко доступными для специалистов в данной области, и могут использоваться для достижения экспрессии кодирующей последовательности, представляющей интерес, на высоком уровне. Кроме того, в данном документе также найдут применение последовательности, полученные из невирусных генов, таких как ген металлотионеина мыши. Такие промоторные последовательности коммерчески доступны, например, от Stratagene (Сан-Диего, Калифорния). Дополнительные иллюстративные последовательности промоторов описаны, например, в WO 2018213786A1 (включенной в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления используется энхансер гена аполипопротеина E (ApoE) или его функциональный фрагмент, например, для управления экспрессией в печени. В некоторых вариантах осуществления используются две копии энхансера ApoE или его функционального фрагмента. В некоторых вариантах осуществления энхансер ApoE или его функциональный фрагмент используют в комбинации с промотором, например, промотором гена альфа-1-антитрипсина человека (hAAT).

В некоторых вариантах осуществления регуляторные последовательности придают возможности экспрессии тканеспецифических генов. В некоторых случаях тканеспецифические регуляторные последовательности связывают тканеспецифические факторы транскрипции, которые индуцируют транскрипцию тканеспецифическим

образом. Различные тканеспецифические регуляторные последовательности (например, промоторы, энхансеры и т. д.) известны из уровня техники. Иллюстративные тканеспецифические регуляторные последовательности включают без ограничения следующие тканеспецифические промоторы: специфичный для печени промотор гена тироксинсвязывающего глобулина (TBG), промотор гена инсулина, промотор гена глюкагона, промотор гена соматостатина, промотор гена панкреатического полипептида (PPY), промотор гена синапсина-1 (Syn), промотор гена креатинкиназы (MCK), промотор гена десмина млекопитающих (DES), промотор гена тяжелой цепи α-миозина (a-MHC) или промотор гена сердечного тропонина T (cTnT). Другие иллюстративные промоторы включают промотор гена бета-актина, коровый промотор вируса гепатита B, Sandig et al., Gene Ther., 3:1002-9 (1996); промотор гена альфа-фетопротеина (AFP), Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), промотор гена костного остеокальцина (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); промотор гена костного сиалопротеина (Chen et al., J. Bone Miner. Res. 11:654-64 (1996)), промотор CD2 (Hansal et al., J. Immunol., 161:1063-8 (1998); промотор гена тяжелой цепи иммуноглобулина; промотор гена α-цепи T-клеточного рецептора, нейрональные промоторы, такие как промотор нейронспецифической енолазы (NSE) (Andersen et al. Cell. Mol. Neurobiol., 13:503-15 (1993)); промотор гена легкой цепи нейрофиламента (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)) и промотор нейронспецифического гена vgf (Piccioli et al., Neuron, 15:373-84 (1995)) и другие. Дополнительные иллюстративные промоторные последовательности описаны, например, в патенте США № 10300146 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления тканеспецифический регуляторный элемент, например, тканеспецифический промотор, выбран из элемента, о котором известно, что он функционально связан с геном, который экспрессируется на высоком уровне в данной ткани, например, как измерено с помощью секвенирования РНК или данных об экспрессии белка или их комбинации. Способы анализа тканевой специфичности по экспрессии описаны в Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014), который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления вектор, описанный в данном документе, представляет собой полицистронную экспрессионную конструкцию. Полицистронные экспрессионные конструкции включают, например, конструкции, содержащие первую кассету экспрессии, например, содержащую первый промотор и первую кодирующую последовательность нуклеиновой кислоты, и вторую кассету экспрессии, например, содержащую второй промотор и вторую кодирующую последовательность нуклеиновой кислоты. Такие полицистронные экспрессионные конструкции могут в некоторых случаях быть особенно полезными для доставки нетранслируемых продуктов генов, таких как шпилечные РНК, вместе с полипептидом, например, Gene Writer и матрицы Gene Writer. В некоторых вариантах осуществления полицистронные экспрессионные конструкции могут демонстрировать сниженные уровни экспрессии одного или нескольких включенных трансгенов, например, из-за интерференции промоторов или присутствия несовместимых

элементов, представляющих собой нуклеиновую кислоту, в непосредственной близости. Если полицистронная экспрессионная конструкция является частью вирусного вектора, присутствие самокомплементарной последовательности нуклеиновой кислоты может в некоторых случаях препятствовать образованию структур, необходимых для репродукции или упаковки вируса.

В некоторых вариантах осуществления последовательность кодирует РНК со шпилькой. В некоторых вариантах осуществления шпилечная РНК представляет собой направляющую РНК, РНК-матрицу, shRNA или микроРНК. В некоторых вариантах осуществления первый промотор представляет собой промотор РНК-полимеразы I. В некоторых вариантах осуществления первый промотор представляет собой промотор РНК-полимеразы II. В некоторых вариантах осуществления второй промотор представляет собой промотор РНК-полимеразы III. В некоторых вариантах осуществления второй промотор представляет собой промотор U6 или H1. В некоторых вариантах осуществления конструкция на основе нуклеиновой кислоты содержит структуру конструкции B1 или B2 AAV.

Не желая ограничиваться какой-либо теорией, полагают, что полицистронные экспрессионные конструкции могут не достигать оптимальных уровней экспрессии по сравнению с системами экспрессии, содержащими только один цистрон. Одной из предполагаемых причин более низких уровней экспрессии, достигаемых при использовании полицистронных экспрессионных конструкций, содержащих два или более промоторных элемента, является явление интерференции промоторов (см., например, Curtin J A, Dane A P, Swanson A, Alexander I E, Ginn S L. *Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct*. Gene Ther. 2008 March; 15(5):384-90 и Martin-Duque P, Jezzard S, Kaftansis L, Vassaux G. *Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes*. Hum Gene Ther. 2004 October; 15(10):995-1002; оба литературных источника включены в данный документ посредством ссылки для раскрытия явления интерференции промоторов). В некоторых вариантах осуществления проблему интерференции промоторов можно решить, например, путем получения полицистронных экспрессионных конструкций, содержащих только один промотор, управляющий транскрипцией нескольких кодирующих последовательностей нуклеиновых кислот, разделенных участками внутренней посадки рибосомы, или путем разделения цистронов, содержащих их собственный промотор, транскрипционными инсуляторными элементами. В некоторых вариантах осуществления экспрессия нескольких цистронов, управляемая одним промотором, может приводить к неравномерным уровням экспрессии цистронов. В некоторых вариантах осуществления промотор не может быть эффективно выделен, и элементы для выделения могут быть несовместимы с некоторыми векторами для переноса генов, например, некоторыми ретровирусными векторами.

МикроРНК

miRNA и другие малые интерферирующие нуклеиновые кислоты обычно регулируют экспрессию генов посредством расщепления/разрушения РНК-транскрипта-мишени или репрессии трансляции матричной РНК-мишени (мРНК). miRNA могут в некоторых случаях экспрессироваться нативно, обычно в виде конечных нетранслируемых РНК-продуктов длиной 19-25. miRNA обычно проявляют свою активность посредством специфичных в отношении последовательности взаимодействий с 3′-нетранслируемыми областями (UTR) мРНК-мишеней. Эти эндогенно экспрессируемые miRNA могут образовывать шпилечные предшественники, которые впоследствии процессируются в miRNA-дуплекс, а затем в зрелую однонитевую молекулу miRNA. Эта зрелая miRNA обычно направляет многобелковый комплекс miRISC, который идентифицирует 3′-UTR-области-мишени мРНК-мишеней на основе их комплементарности по отношению к зрелой miRNA. Применимые трансгенные продукты могут включать, например, miRNA или сайты связывания miRNA, которые регулируют экспрессию связанного полипептида. Неограничивающий перечень генов miRNA; продукты этих генов и их гомологов являются применимыми в качестве трансгенов или в качестве мишеней для малых интерферирующих нуклеиновых кислот (например, miRNA-губок, антисмысловых олигонуклеотидов), например, в способах, таких как способы, приведенные в US 10300146, 22:25-25:48, включенном посредством ссылки. В некоторых вариантах осуществления один или несколько сайтов связывания для одной или нескольких вышеуказанных miRNA включены в трансген, например, трансген, доставляемый вектором на основе rAAV, например, для ингибирования экспрессии трансгена в одной или нескольких тканях животного, несущего трансген. В некоторых вариантах осуществления сайт связывания может быть выбран таким образом, чтобы обеспечивать контроль экспрессии трансгена тканеспецифическим образом. Например, сайты связывания для специфичной в отношении печени miR-122 могут быть включены в трансген для ингибирования экспрессии этого трансгена в печени. Дополнительные иллюстративные последовательности miRNA описаны, например, в патенте США № 10300146 (включенном в данный документ посредством ссылки во всей своей полноте). Однако для специфичного в отношении печени Gene Writing можно использовать сверхэкспрессию miR-122 вместо использования сайтов связывания для осуществления miR-122-специфичного разрушения. Эта miRNA положительно ассоциирована с дифференцировкой и созреванием клеток печени, а также с повышенной экспрессией генов, специфичных для печени. Таким образом, в некоторых вариантах осуществления кодирующая последовательность miR-122 может быть добавлена к компоненту системы Gene Writing для усиления терапии, направленной на печень.

Ингибитор miR или ингибитор miRNA обычно представляет собой средство, которое блокирует экспрессию и/или процессинг miRNA. Примеры таких средств включают без ограничения антагонисты микроРНК, специфичные в отношении микроРНК антисмысловые последовательности, микроРНК-губки и специфичные в отношении микроРНК олигонуклеотиды (двухнитевые, шпилечные, короткие олигонуклеотиды), которые ингибируют взаимодействие miRNA с комплексом Drosha. Ингибиторы

микроРНК, например, miRNA-губки, могут экспрессироваться в клетках из трансгенов (например, как описано в работе Ebert, M. S. Nature Methods, Epub Aug. 12, 2007; включенной в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления микроРНК-губки или другие ингибиторы miR используются с AAV. МикроРНК-губки обычно специфично ингибируют miRNA посредством комплементарной гептамерной затравочной последовательности. В некоторых вариантах осуществления все семейство miRNA может быть подвергнуто сайленсингу с помощью одной последовательности губки. Другие способы сайленсинга функции miRNA (подавление мишеней miRNA) в клетках будут очевидны специалисту средней квалификации в данной области.

В некоторых вариантах осуществления miRNA, описанная в данном документе, содержит последовательность, приведенную в таблице 4 публикации согласно PCT № WO2020014209, включенной в данный документ посредством ссылки. Также в данный документ посредством ссылки включен перечень иллюстративных последовательностей miRNA из WO2020014209.

В некоторых вариантах осуществления преимущественно подвергать сайленсингу один или несколько компонентов системы Gene Writing (например, мРНК, кодирующую полипептид Gene Writer, РНК-матрицу Gene Writer или последовательность гетерологичного объекта, экспрессируемую из генома после успешного осуществления способа Gene Writing) в части клеток. В некоторых вариантах осуществления преимущественно ограничивать экспрессию компонента системы Gene Writing выбранными типами клеток в ткани, представляющей интерес.

Например, известно, что в заданной ткани, например, в печени, макрофаги и иммунные клетки, например, клетки Купфера в печени, могут участвовать в поглощении средства доставки одного или нескольких компонентов системы Gene Writing. В некоторых вариантах осуществления по меньшей мере один сайт связывания для по меньшей мере одной miRNA, экспрессируемой на высоком уровне в макрофагах и иммунных клетках, например, клетках Купфера, включен в по меньшей мере один компонент системы Gene Writing, например, нуклеиновую кислоту, кодирующую полипептид Gene Writing, или трансген. В некоторых вариантах осуществления miRNA, которая нацелена на один или несколько сайтов связывания, приведена в таблице, упоминаемой данном документе, например, miR-142, например, зрелая miRNA hsa-miR-142-5p или hsa-miR-142-3p.

В некоторых вариантах осуществления может быть предпочтительно снижение уровней Gene Writer и/или активности Gene Writer в клетках, в которых экспрессия Gene Writer или сверхэкспрессия трансгена может оказывать токсический эффект. Например, было показано, что доставка кассеты сверхэкспрессии трансгена в ганглиозные нейроны дорсального корешка может привести в результате к токсичности генной терапии (см. Hordeaux et al. Sci Transl Med 12(569):eaba9188 (2020), включенный в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA может быть включен в компонент системы

Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии данного компонента системы в нейроне, например, в ганглиозном нейроне дорсального корешка. В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA, включенный в компонент системы Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии данного компонента системы в нейроне представляет собой сайт связывания miR-182, например, зрелой miRNA hsa-miR-182-5p или hsa-miR-182-3p. В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA, включенный в компонент системы Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии данного компонента системы в нейроне представляет собой сайт связывания miR-183, например, зрелой miRNA hsa-miR-183-5p или hsa-miR-183-3p. В некоторых вариантах осуществления для усиления ограничения экспрессии одного или нескольких компонентов системы Gene Writing тканью или типом клеток, представляющими интерес, можно использовать комбинации сайтов связывания miRNA.

В таблице A5 ниже приведены иллюстративные miRNA и соответствующие экспрессирующие клетки, например, miRNA, для которых в некоторых вариантах осуществления можно включать сайты связывания (комплементарные последовательности) в трансген или нуклеиновую кислоту, кодирующие полипептид, например, для снижения экспрессии в этой нецелевой клетке.

Таблица A5. Иллюстративные miRNA из нецелевых клеток и тканей

Тип клеток, подвергнутых сайленсингу	**Название miRNA**	**Зрелая miRNA**	**Последовательность miRNA**
Клетки Купфера	miR-142	hsa-miR-142-5p	cauaaaguagaaagcacuacu
Клетки Купфера	miR-142	hsa-miR-142-3p	uguaguguuuccuacuuuaugga
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-5p	uuuggcaaugguagaacucacacu
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-3p	ugguucuagacuugccaacua
Ганглиозные нейроны дорсального корешка	miR-183	hsa-miR-183-5p	uauggcacugguagaauucacu
Ганглиозные нейроны дорсального	miR-183	hsa-miR-183-3p	gugaauuaccgaagggccauaa

корешка из таблицы X			
Гепатоциты	miR-122	hsa-miR-122-5p	uggagugugacaaugguguuug
Гепатоциты	miR-122	hsa-miR-122-3p	aacgccauuaucacacuaaaua

Системы, противодействующие CRISPR, для регуляции активности GeneWriter

Различные подходы для модулирования активности молекулы Cas могут быть применены в сочетании с системами и способами, описанными в данном документе. Например, в некоторых вариантах осуществления полипептид, описанный в данном документе (например, молекулу Cas или Gene Writer, содержащий домен Cas), можно регулировать с помощью средства, противодействующего CRISPR (например, белка, противодействующего CRISPR, или малой молекулы, противодействующей CRISPR). В некоторых вариантах осуществления молекула Cas или домен Cas содержит отвечающий интеин, такой как, например, 4-гидрокситамоксифен (4-HT)-отвечающий интеин, молекулу iCas (например, iCas9); 4-HT-отвечающий Cas (например, аллостерически регулируемый Cas9 (arC9) или мертвый Cas9 (dC9)). В системах и способах, описанных в данном документе, могут также использовать химически индуцированную систему димеризации расщепленных белковых фрагментов (например, опосредованную рапамицином димеризацию FK506-связывающего белка 12 (FKBP) и FKBP-рапамицин-связывающего домена (FRB), доменов гетеродимеризации ABI-PYL1, индуцируемых абсцизовой кислотой, и доменов гетеродимеризации GID1-GAI, индуцируемых гиббереллином); димер пептида BCL-xL и пептидов BH3, малую молекулу A385358 (A3), систему дегрона (например, дестабилизированную систему FKBP-Cas9, систему дегрона, индуцируемую ауксином (AID), или систему дегрона DHFR E. coli), аптамер или аптазим, слитый с gRNA (например, тетрациклин- и теофиллин-отвечающие биопереключатели), белки AcrIIA2 и AcrIIA4, а также BRD0539.

В некоторых вариантах осуществления интеин, отвечающий на малые молекулы (например, 4-гидрокситамоксифен (4-HT)-отвечающий интеин), встроен в определенные сайты молекулы Cas (например, Cas9). В некоторых вариантах осуществления встраивание 4HT-отвечающего интеина нарушает ферментативную активность Cas9. В некоторых вариантах осуществления молекула Cas (например, iCas9) слита с гормонсвязывающим доменом эстрогенового рецептора (ERT2). В некоторых вариантах осуществления лигандсвязывающий домен эстрогенового рецептора-α человека может быть вставлен в молекулу Cas (например, Cas9 или мертвого Cas9 (dC9)), например, в положении 231, с получением 4HT-отвечающего средства, противодействующего CRISPR-Cas9 (например, arC9 или dC9). В некоторых вариантах осуществления dCas9 может обеспечивать репрессию функции Cas9, зависимую от дозы 4-HT. В некоторых вариантах осуществления arC9 может обеспечивать контроль функции Cas9, зависимый от дозы 4-HT. В некоторых вариантах осуществления молекула Cas (например, Cas9) слита с расщепленными белковыми фрагментами. В некоторых вариантах осуществления химически

индуцированная димеризация расщепленных белковых фрагментов (например, опосредованная рапамицином димеризация FK506-связывающего белка 12 (FKBP) и FKBP-рапамицин-связывающего домена (FRB)) может индуцировать активность молекулы Cas9 на низких уровнях. В некоторых вариантах осуществления химически индуцированная система димеризации (например, домены гетеродимеризации ABI-PYL1, индуцируемые абсцизовой кислотой, и домены гетеродимеризации GID1-GAI, индуцируемые гиббереллином) может индуцировать дозозависимую и обратимую активацию/репрессию транскрипции Cas9. В некоторых вариантах осуществления индуцируемая система Cas9 (ciCas9) включает замену домена REC2 молекулы Cas (например, Cas9) пептидом BCL-xl и присоединение пептида BH3 к N- и C-концам модифицированного Cas9.BCL. В некоторых вариантах осуществления взаимодействие между пептидами BCL-xL и BH3 может удерживать Cas9 в неактивном состоянии. В некоторых вариантах осуществления малая молекула (например, A-385358 (A3)) может нарушать взаимодействие между пептидами BCL-xl и BH3 для активации Cas9. В некоторых вариантах осуществления индуцируемая система Cas9 может демонстрировать дозозависимый контроль нуклеазной активности. В некоторых вариантах осуществления система дегрона может индуцировать разрушение молекулы Cas (например, Cas9) при активации или деактивации внешним фактором (например, низкомолекулярным лигандом, светом, температурой или белком). В некоторых вариантах осуществления малая молекула BRD0539 обратимо ингибирует молекулу Cas (например, Cas9). Дополнительную информацию о белках, противодействующих CRISPR, или малых молекулах, противодействующих CRISPR, можно найти, например, в Gangopadhyay, S.A. et al. Precision control of CRISPR-Cas9 using small molecules and light, Biochemistry, 2019, Maji, B. et al. A high-throughput platform to identify small molecule inhibitors of CRISPR-Cas9, и Pawluk Anti-CRISPR: discovery, mechanism and function, Nature Reviews Microbiology volume 16, pages 12-17(2018), каждый из которых включен посредством ссылки во всей своей полноте.

14. *Самоинактивирующиеся модули для регулирования активности Gene Writer*

15. В некоторых вариантах осуществления системы Gene Writer, описанные в данном документе, содержат самоинактивирующийся модуль. Самоинактивирующийся модуль приводит к снижению экспрессии полипептида Gene Writer, матрицы Gene Writer или того и другого. Не желая ограничиваться какой-либо теорией, полагают, что самоинактивирующийся модуль обеспечивает временный период экспрессии Gene Writer перед инактивацией. Не желая ограничиваться какой-либо теорией, полагают, что активность полипептида Gene Writer в сайте-мишени приводит к введению мутации (например, замены, вставки или делеции) в ДНК, кодирующую полипептид Gene Writer или матрицу Gene Writer, что приводит к снижению экспрессии полипептида или матрицы Gene Writer. В некоторых вариантах осуществления самоинактивирующегося модуля сайт-мишень для полипептида Gene Writer включен в ДНК, кодирующую полипептид Gene Writer или матрицу Gene Writer. В некоторых вариантах осуществления одна, две, три, четыре, пять или больше копий сайта-мишени включены в ДНК, кодирующую полипептид

Gene Writer или матрицу Gene Writer. В некоторых вариантах осуществления сайт-мишень в ДНК, кодирующей полипептид Gene Writer или матрицу Gene Writer, представляет собой тот же сайт-мишень, что и сайт-мишень в геноме. В некоторых вариантах осуществления сайт мишень представляет собой сайт-мишень, отличающийся от сайта-мишени в геноме. В некоторых вариантах осуществления в сайте-мишени самоинактивирующегося модуля используется та же или другая РНК-матрица или направляющая РНК п сравнению с геномным сайтом-мишенью. В некоторых вариантах осуществления сайт-мишень модифицирован с помощью обратной транскрипции с праймированием мишенью с использованием РНК-матрицы. В некоторых вариантах осуществления сайт-мишень имеет однонитевой разрыв. Сайт-мишень может быть включен в энхансер, промотор, нетранслируемую область, экзон, интрон, открытую рамку считывания или спейсерную последовательность.

В некоторых вариантах осуществления при инактивации экспрессия снижается на 25%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99,9% или является еще более низкой по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. В некоторых вариантах осуществления система Gene Writer, содержащая самоинактивирующийся модуль, характеризуется 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% или более высоким уровнем интеграции в сайтах-мишенях, чем в нецелевых сайтах, по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. Система Gene Writer, содержащая самоинактивирующийся модуль, характеризуется 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% или более высокой эффективностью модификации сайта-мишени по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. В некоторых вариантах осуществления самоинактивирующийся модуль включен, если полипептид Gene Writer доставляется в виде ДНК, например, с помощью вирусного вектора.

Для нуклеаз были описаны самоинактивирующиеся модули. См., например, в Li et al A Self-Deleting AAV-CRISPR System for In Vivo Genome Editing, Mol Ther Methods Clin Dev. 2019 Mar 15; 12: 111-122, P. Singhal, Self-Inactivating Cas9: a method for reducing exposure while maintaining efficacy in virally delivered Cas9 applications (доступно на www.editasmedicine.com/wp-content/uploads/2019/10/aef_asgct_poster_2017_final_-_present_5-11-17_515pm1_1494537387_1494558495_1497467403.pdf), и Epstein and Schaffer Engineering a Self-Inactivating CRISPR System for AAV Vectors Targeted Genome Editing I|Volume 24, SUPPLEMENT 1, S50, May 01, 2016, *и* WO2018106693A1.

Малые молекулы

В некоторых вариантах осуществления полипептид, описанный в данный документе (например, полипептид Gene Writer), можно контролировать с помощью малой молекулы. В некоторых вариантах осуществления полипептид димеризуется посредством малой молекулы.

В некоторых вариантах осуществления полипептид можно контролировать с помощью химически индуцированной димеризации (CID) посредством малых молекул. CID обычно применяется для создания переключателей функции белка, чтобы изменить физиологические параметры клеток. Иллюстративным высокоспецифичным эффективным димеризатором является римидуцид (AP1903), который имеет две идентичные связывающие белок поверхности, расположенные по типу "хвост к хвосту", каждая из которых обладает высокой аффинностью и специфичностью в отношении мутантной формы FKBP12: FKBP12(F36V) (FKBP12v36, F_{V36} или F_V). Присоединение одного или нескольких доменов F_V к одной или нескольким сигнальным молекулам клетки, которые в обычных условиях зависят от гомодимеризации, может привести к преобразованию данного белка в контрольный римидуцид. Гомодимеризация с римидуцидом применяется в случае с индуцируемым переключателем безопасности каспазы. В основе данного молекулярного переключателя, который контролируется отдельным лигандом димеризации, лежит гетеродимеризующаяся малая молекула - рапамицин или аналоги рапамицина ("рапалоги"). Рапамицин связывается с FKBP12 и его вариантами и может индуцировать гетеродимеризацию сигнальных доменов, слитых с FKBP12, посредством связывания как с FKBP12, так и с полипептидами, содержащими FKBP-рапамицин-связывающий (FRB) домен mTOR. В некоторых вариантах осуществления настоящей заявки предусмотрены молекулярные переключатели, которые значительно расширяют возможности применения рапамицина, рапалогов и римидуцида в качестве средств для терапевтического применения.

В некоторых вариантах осуществления технологии двойного переключателя гомодимеризатор, такой как AP1903 (римидуцид), непосредственно индуцирует димеризацию или мультимеризацию полипептидов, содержащих область мультимеризации FKBP12. В других вариантах осуществления полипептид, содержащий мультимеризованный FKBP12, мультимеризуется или агрегируется посредством связывания с гетеродимеризатором, таким как рапамицин или рапалог, который также связывается с областью мультимеризации FRB или варианта FRB в химерном полипептиде, также экспрессируемом в модифицированной клетке, таком как, например, химерный антигенный рецептор. Рапамицин представляет собой природный макролид, который с высокой аффинностью (< 1 нМ) связывается с FKBP12 и совместно с ним инициирует высокоаффинное ингибирующее взаимодействие с FKBP-рапамицин-связывающим (FRB) доменом mTOR. FRB имеет небольшой размер (89 аминокислот) и поэтому может применяться в качестве белковой "метки" или "указателя" при добавлении ко многим белкам. Коэкспрессия белка, слитого с FRB, с белком, слитым с FKBP12, делает их сближение индуцируемым рапамицином (12-16). Это может служить основой для переключателя клеточной безопасности, регулируемого перорально доступным лигандом - рапамицином или производными рапамицина (рапалогами), которые не ингибируют mTOR при низкой терапевтической дозе, но вместо этого связываются с выбранными мутантными доменами FRB, слитыми с каспазой-9. (См. Sabatini D M, et al., Cell. 1994; 78(1):35-43;

Brown E J, et al., Nature. 1994; 369(6483):756-8; Chen J, et al., Proc Natl Acad Sci USA. 1995; 92(11):4947-51; и Choi J, Science. 1996; 273(5272):239-42).

В некоторых вариантах осуществления в терапевтических клетках предусмотрено два уровня контроля. В вариантах осуществления первый уровень контроля может быть настраиваемым, т. е. уровень удаления терапевтических клеток можно регулировать таким образом, чтобы он приводил к частичному удалению терапевтических клеток. В некоторых вариантах осуществления химерный антигенный полипептид содержит сайт связывания рапамицина или аналога рапамицина. В вариантах осуществления в терапевтической клетке также присутствует "суицидальный" ген, такой как, например, ген, кодирующий полипептид каспазу. При использовании данного контролируемого первого уровня в некоторых вариантах осуществления потребность в продолжении терапии может быть уравновешена необходимостью устранения или снижения уровня возникновения отрицательных побочных эффектов. В некоторых вариантах осуществления пациенту вводят аналог рапамицина рапалог, который затем связывается как с полипептидом каспазой, так и с химерным антигенным рецептором, привлекая, таким образом, полипептид каспазу в местоположение CAR и обеспечивая агрегацию полипептида каспазы. При агрегации полипептид каспаза индуцирует апоптоз. Количество рапамицина или аналога рапамицина, вводимого пациенту, может варьироваться; если желательно удаление более низкого уровня клеток посредством апоптоза для уменьшения побочных эффектов и продолжения CAR-терапии, то пациенту можно вводить более низкий уровень рапамицина или рапамицина. В некоторых вариантах осуществления второй уровень контроля может быть предназначен для достижения максимального уровня устранения клеток. Данный второй уровень может быть основан, например, на применении римидуцида или AP1903. Если есть необходимость быстро устранить до 100% терапевтических клеток, пациенту можно вводить AP1903. Мультимерный AP1903 связывается с полипептидом каспазой, что приводит к мультимеризации полипептида каспазы и апоптозу. В определенных примерах второй уровень также можно настраивать или контролировать с помощью уровня AP1903, вводимого субъекту.

В некоторых вариантах осуществления малые молекулы могут применяться для контроля генов, как описано, например, в US10584351 в 47:53-56:47 (включенном в данный документ посредством ссылки во всей своей полноте), вместе с пригодными лигандами для контрольных признаков, например, в US10584351 в 56:48 и след., а также в U10046049 в 43:27-52:20, включенных посредством ссылки, а также в описании лигандов для таких систем контроля в 52:21 и след.

Химически модифицированные нуклеиновые кислоты и концевые признаки нуклеиновых кислот

Описанная в данном документе нуклеиновая кислота (например, нуклеиновая кислота-матрица, например, РНК-матрица; или нуклеиновая кислота (например, мРНК), кодирующая Gene Writer) может содержать немодифицированные или модифицированные азотистые основания. Встречающиеся в природе РНК синтезируются из четырех основных

рибонуклеотидов: ATP, CTP, UTP и GTP, но могут содержать нуклеотиды с посттранскрипционными модификациями. Дополнительно в РНК было идентифицировано примерно сто различных модификаций нуклеозидов (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). РНК также может содержать полностью синтетические нуклеотиды, не встречающиеся в природе.

В некоторых вариантах осуществления химическая модификация представлена в PCT/US2016/032454, публикации патента США № 20090286852, международной заявке № WO/2012/019168, WO/2012/045075, WO/2012/135805, WO/2012/158736, WO/2013/039857, WO/2013/039861, WO/2013/052523, WO/2013/090648, WO/2013/096709, WO/2013/101690, WO/2013/106496, WO/2013/130161, WO/2013/151669, WO/2013/151736, WO/2013/151672, WO/2013/151664, WO/2013/151665, WO/2013/151668, WO/2013/151671, WO/2013/151667, WO/2013/151670, WO/2013/151666, WO/2013/151663, WO/2014/028429, WO/2014/081507, WO/2014/093924, WO/2014/093574, WO/2014/113089, WO/2014/144711, WO/2014/144767, WO/2014/144039, WO/2014/152540, WO/2014/152030, WO/2014/152031, WO/2014/152027, WO/2014/152211, WO/2014/158795, WO/2014/159813, WO/2014/164253, WO/2015/006747, WO/2015/034928, WO/2015/034925, WO/2015/038892, WO/2015/048744, WO/2015/051214, WO/2015/051173, WO/2015/051169, WO/2015/058069, WO/2015/085318, WO/2015/089511, WO/2015/105926, WO/2015/164674, WO/2015/196130, WO/2015/196128, WO/2015/196118, WO/2016/011226, WO/2016/011222, WO/2016/011306, WO/2016/014846, WO/2016/022914, WO/2016/036902, WO/2016/077125 или WO/2016/077123, каждая из которых включена в данный документ посредством ссылки во всей своей полноте. Понятно, что включение химически модифицированного нуклеотида в полинуклеотид может привести к встраиванию модификации в азотистое основание, остов или и в то, и в другое в зависимости от местоположения модификации в нуклеотиде. В некоторых вариантах осуществления модификация остова представлена в EP 2813570, который включен в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления модифицированный кэп представлен в публикации патента США № 20050287539, которая включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота (например, РНК, например, мРНК) содержит один или несколько из ARCA: антиинвертированного аналога кэпа (m27.3'-OGP3G), GP3G (неметилированного аналога кэпа), m7GP3G (монометилированного аналога кэпа), m32.2.7GP3G (триметилированного аналога кэпа), m5CTP (5'-метилцитидинтрифосфата), m6ATP (N6-метиладенозин-5'-трифосфата), s2UTP (2-тиоуридинтрифосфата) и Ψ (псевдоуридинтрифосфата).

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота содержит 5'-кэп, например: 7-метилгуанозиновый кэп (например, кэп O-Me-m7G); гиперметилированный аналог кэпа; аналог кэпа, полученный из NAD+ (например, как описано в Kiledjian, Trends in Cell Biology 28, 454-464 (2018)); или модифицированный,

например, биотинилированный, аналог кэпа (например, как описано в Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)).

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота содержит 3'-признак, выбранный из одного или нескольких из поли(А)-хвоста; структуры "стебель-петля" длиной 16 нуклеотидов, фланкированной неспаренными 5 нуклеотидами (например, как описано в Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)); тройной спиральной структуры (например, как описано в Brown et al., PNAS 109, 19202-19207 (2012)); структуры тРНК, Y-РНК или РНК Vault (например, как описано в Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)); включения одного или нескольких дезоксирибонуклеотидтрифосфатов (dNTP), 2'-О-метилированных NTP или фосфотиоат-NTP; химической модификации одного нуклеотида (например, окисление 3'-концевой рибозы до реакционноспособного альдегида с последующей конъюгацией альдегид-реактивного модифицированного нуклеотида) или химического лигирования с другой молекулой нуклеиновой кислоты.

В некоторых вариантах осуществления нуклеиновая кислота (например, нуклеиновая кислота-матрица) содержит один или несколько модифицированных нуклеотидов, например, выбранных из дигидроуридина, инозина, 7-метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-О-метилриботимидина, 2'-О-этилриботимидина, 2'-фторриботимидина, С5-пропинилдезоксицитидина (pdC), С5-пропинилдезоксиуридина (pdU), С5-пропинилцитидина (pC), С5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксицитидина, 5-метоксиметилдезоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксицитидина, С5-пропинил-f-цитидина (pfC), С5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, С5-пропинил-m-цитидина (pmC), С5-пропинил-f-уридина (pmU), 5-метил-m-цитидина, 5-метил-m-уридина, LNA (запертой нуклеиновой кислоты), MGB (средства, связывающегося с малой бороздкой), псевдоуридина (Ψ), 1-N-метилпсевдоуридина (1-Me-Ψ) или 5-метоксиуридина (5-MO-U).

В некоторых вариантах осуществления нуклеиновая кислота содержит модификацию остова, например, модификацию сахарной или фосфатной группы в остове. В некоторых вариантах осуществления нуклеиновая кислота содержит модификацию азотистого основания.

В некоторых вариантах осуществления нуклеиновая кислота содержит один или несколько химически модифицированных нуклеотидов из таблицы М1, одну или несколько химических модификаций остова из таблицы М2, один или несколько химически модифицированных кэпов из таблицы М3. Например, в некоторых вариантах осуществления нуклеиновая кислота содержит два или более (например, 3, 4, 5, 6, 7, 8, 9 или 10 или более) различных типов химических модификаций. Например, нуклеиновая кислота может содержать два или более (например, 3, 4, 5, 6, 7, 8, 9 или 10 или более) различных типов модифицированных азотистых оснований, например, как описано в данном документе, например, в таблице М1. В качестве альтернативы или в комбинации,

нуклеиновая кислота может содержать два или более (например, 3, 4, 5, 6, 7, 8, 9 или 10 или более) различных типов модификаций остова, например, описанных в данном документе, например, в таблице M2. В качестве альтернативы или в комбинации, нуклеиновая кислота может содержать один или несколько модифицированных кэпов, например, описанных в данном документе, например, в таблице M3. Например, в некоторых вариантах осуществления нуклеиновая кислота содержит один или несколько типов модифицированных азотистых оснований и один или несколько типов модификаций остова; один или несколько типов модифицированных азотистых оснований и один или несколько модифицированных кэпов; один или несколько типов модифицированного кэпа и один или несколько типов модификаций остова или один или несколько типов модифицированных азотистых оснований, один или несколько типов модификаций остова и один или несколько типов модифицированного кэпа.

В некоторых вариантах осуществления нуклеиновая кислота содержит одно или несколько (например, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000 или больше) модифицированных азотистых оснований. В некоторых вариантах осуществления все азотистые основания нуклеиновой кислоты являются модифицированными. В некоторых вариантах осуществления нуклеиновая кислота модифицирована в одном или нескольких (например, ., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, или больше) положениях в остове. В некоторых вариантах осуществления все положения в остове нуклеиновой кислоты являются модифицированными.

Таблица M1. Модифицированные нуклеотиды

5-азауридин	N2-метил-6-тиогуанозин
2-тио-5-азауридин	N2,N2-диметил-6-тиогуанозин
2-тиоуридин	пиридин-4-оновый рибонуклеозид
4-тиопсевдоуридин	2-тио-5-азауридин
2-тиопсевдоуридин	2-тиоуридин
5-гидроксиуридин	4-тиопсевдоуридин
3-метилуридин	2-тиопсевдоуридин
5-карбоксиметилуридин	3-метилуридин
1-карбоксиметилпсевдоуридин	1-пропинилпсевдоуридин
5-пропинилуридин	1-метил-1-деазапсевдоуридин
1-пропинилпсевдоуридин	2-тио-1-метил-1-деазапсевдоуридин
5-тауринометилуридин	4-метоксипсевдоуридин
1-тауринометилпсевдоуридин	5'-O-(1-тиофосфат)-аденозин
5-тауринометил-2-тиоуридин	5'-O-(1-тиофосфат)-цитидин
1-тауринометил-4-тиоуридин	5'-O-(1-тиофосфат)-гуанозин

5-метилуридин	5'-O-(1-тиофосфат)-уридин
1-метилпсевдоуридин	5'-O-(1-тиофосфат)-псевдоуридин
4-тио-1-метилпсевдоуридин	2'-O-метиладенозин
2-тио-1-метилпсевдоуридин	2'-O-метилцитидин
1-метил-1-деазапсевдоуридин	2'-O-метилгуанозин
2-тио-1-метил-1-деазапсевдоуридин	2'-O-метилуридин
дигидроуридин	2'-O-метилпсевдоуридин
дигидропсевдоуридин	2'-O-метилинозин
2-тиодигидроуридин	2-метиладенозин
2-тиодигидропсевдоуридин	2-метилтио-N6-метиладенозин
2-метоксиуридин	2-метилтио-N6 изопентениладенозин
2-метокси-4-тиоуридин	2-метилтио-N6-(цис-гидроксиизопентенил)аденозин
4-метоксипсевдоуридин	N6-метил-N6-треонилкарбамоиладенозин
4-метокси-2-тиопсевдоуридин	N6-гидроксинорвалилкарбамоиладенозин
5-азацитидин	2-метилтио-N6-гидроксинорвалилкарбамоиладенозин
псевдоизоцитидин	2'-O-рибозиладенозин (фосфат)
3-метилцитидин	1,2'-O-диметилинозин
N4-ацетилцитидин	5,2'-O-диметилцитидин
5-формилцитидин	N4-ацетил-2'-O-метилцитидин
N4-метилцитидин	Лизидин
5-гидроксиметилцитидин	7-метилгуанозин
1-метилпсевдоизоцитидин	N2,2'-O-диметилгуанозин
пирролоцитидин	N2,N2,2'-O-триметилгуанозин
пирролопсевдоизоцитидин	2'-O-рибозилгуанозин (фосфат)
2-тиоцитидин	Вибутозин
2-тио-5-метилцитидин	Пероксивибутозин
4-тиопсевдоизоцитидин	Гидроксивибутозин
4-тио-1-метилпсевдоизоцитидин	недомодифицированный
4-тио-1-метил-1-деазапсевдоизоцитидин	гидроксивибутозин
1-метил-1-деазапсевдоизоцитидин	метилвиозин
зебуларин	квеуозин
5-азазебуларин	
5-метилзебуларин	

5-аза-2-тиозебуларин	эпоксиквеуозин
2-тиозебуларин	галактозилквеуозин
2-метоксицитидин	маннозилквеуозин
2-метокси-5-метилцитидин	7-циано-7-деазагуанозин
4-метоксипсевдоизоцитидин	7-аминометил-7-деазагуанозин
4-метокси-1-метилпсевдоизоцитидин	археозин
2-аминопурин	5,2'-О-диметилуридин
2,6-диаминопурин	4-тиоуридин
7-деазааденин	5-метил-2-тиоуридин
7-деаза-8-азааденин	2-тио-2'-О-метилуридин
7-деаза-2-аминопурин	3-(3-амино-3-карбоксипропил)уридин
7-деаза-8-аза-2-аминопурин	5-метоксиуридин
7-деаза-2,6-диаминопурин	уридин-5-оксиуксусная кислота
7-деаза-8-аза-2,6-диаминопурин	Сложный метиловый эфир уридин-5-оксиуксусной кислоты
1-метиладенозин	5-(карбоксигидроксиметил)уридин)
N6-изопентениладенозин	Сложный метиловый эфир 5-(карбоксигидроксиметил)уридина
N6-(цис-гидроксиизопентенил)аденозин	5-метоксикарбонилметилуридин
2-метилтио-N6-(цис-гидроксиизопентенил)аденозин	5-метоксикарбонилметил-2'-О-метилуридин
N6-глицинилкарбамоиладенозин	5-метоксикарбонилметил-2-тиоуридин
N6-треонилкарбамоиладенозин	5-аминометил-2-тиоуридин
2-метилтио-N6-треонилкарбамоиладенозин	5-метиламинометилуридин
N6,N6-диметиладенозин	5-метиламинометил-2-тиоуридин
7-метиладенин	5-метиламинометил-2-селеноуридин
2-метилтиоаденин	5-карбамоилметилуридин
2-метоксиаденин	5-карбамоилметил-2'-О-метилуридин
инозин	5-карбоксиметиламинометилуридин
1-метилинозин	5-карбоксиметиламинометил-2'-О-метилуридин
виозин	5-карбоксиметиламинометил-2-тиоуридин
вибутозин	N4,2'-О-диметилцитидин
7-деазагуанозин	
7-деаза-8-азагуанозин	

6-тиогуанозин	5-карбоксиметилуридин
6-тио-7-деазагуанозин	N6,2'-O-диметиладенозин
6-тио-7-деаза-8-азагуанозин	N, N6,O-2'-триметиладенозин
7-метилгуанозин	N2,7-диметилгуанозин
6-тио-7-метилгуанозин	N2,N2,7-триметилгуанозин
7-метилинозин	3,2'-O-диметилуридин
6-метоксигуанозин	5-метилдигидроуридин
1-метилгуанозин	5-формил-2'-O-метилцитидин
N2-метилгуанозин	1,2'-O-диметилгуанозин
N2,N2-диметилгуанозин	4-деметилвиозин
8-оксогуанозин	Изовиозин
7-метил-8-оксогуанозин	N6-ацетиладенозин
1-метил-6-тиогуанозин	

Таблица M2. Модификации остова

2'-O-метилмодифицированный остов
Остов на основе пептидной нуклеиновой кислоты (PNA)
Фосфотиоатный остов
Морфолиновый остов
Карбаматный остов
Силоксановый остов
Сульфидный остов
Сульфоксидный остов
Сульфоновый остов
Формацетильный остов
Тиоформацетильный остов
Метиленформацетильный остов
Рибоацетильный остов
Алкенсодержащий остов
Сульфаматный остов
Сульфонатный остов
Сульфаниламидный остов
Метилениминовый остов
Метиленгидразиновый остов

Амидный остов

Таблица M3. Модифицированные кэпы

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m7,2'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG
m7GppppG
m7GppppA
m7GppppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

Получение композиций и систем

Известны способы разработки и конструирования конструкций на основе нуклеиновых кислот и белков или полипептидов (таких как системы, конструкции и полипептиды, описанные в данном документе). Как правило, можно применять рекомбинантные способы. См. в целом Smales & James (Eds.), *Therapeutic Proteins: Methods and Protocols* (Methods in Molecular Biology), Humana Press (2005); а также Crommelin, Sindelar & Meibohm (Eds.), *Pharmaceutical Biotechnology: Fundamentals and Applications*, Springer (2013). Способы разработки, получения, оценивания, очистки композиций на основе нуклеиновых кислот и произведения манипуляций с ними описаны в Green и Sambrook (Eds.), *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

В настоящем изобретении частично предусмотрена нуклеиновая кислота, например, вектор, кодирующий полипептид Gene Writer, описанный в данном документе, нуклеиновая

кислота-матрица, описанная в данном документе, или и то, и другое. В некоторых вариантах осуществления вектор содержит селектируемый маркер, например, маркер устойчивости к антибиотикам. В некоторых вариантах осуществления маркер устойчивости к антибиотикам представляет собой маркер устойчивости к канамицину. В некоторых вариантах осуществления маркер устойчивости к антибиотикам не придает устойчивости к бета-лактамным антибиотикам. В некоторых вариантах осуществления вектор не содержит маркер устойчивости к ампициллину. В некоторых вариантах осуществления вектор содержит маркер устойчивости к канамицину и не содержит маркер устойчивости к ампициллину. В некоторых вариантах осуществления вектор, кодирующий полипептид Gene Writer, интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления вектор, кодирующий полипептид Gene Writer, не интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления вектор, кодирующий нуклеиновую кислоту-матрицу (например, РНК-матрицу), не интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то селектируемый маркер не интегрируется в геном. В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то гены или последовательности, участвующие в поддержании вектора (например, гены поддержания плазмиды), не интегрируются в геном. В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то последовательности, регулирующие перенос (например, инвертированные концевые повторы, например, из AAV), не интегрируются в геном. В некоторых вариантах осуществления введение вектора (например, кодирующего полипептид Gene Writer, описанный в данном документе, нуклеиновую кислоту-матрицу, описанную в данном документе, или и то, и другое) в клетку-, ткань-, орган-мишень или субъекту-мишени приводит к интеграции части вектора в один или несколько сайтов-мишеней в геноме(геномах) указанных клетки-, ткани-, органа- или субъекта-мишени. В некоторых вариантах осуществления менее 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2 или 1% сайтов-мишеней (например, ни один из сайтов-мишеней), содержащих интегрированный материал, содержат селектируемый маркер (например, ген устойчивости к антибиотикам), последовательность, регулирующую перенос (например, инвертированный концевой повтор, например, из AAV), или и то, и другое из вектора.

Иллюстративные способы получения терапевтического фармацевтического белка или полипептида, описанного в данном документе, включают экспрессию в клетках млекопитающих, хотя рекомбинантные белки также могут быть получены с использованием клеток насекомых, дрожжей, бактерий или других клеток под контролем соответствующих промоторов. Экспрессионные векторы для клеток млекопитающих могут содержать нетранскрибируемые элементы, такие как точка начала репликации,

подходящий промотор и другие 5'- или 3'- фланкирующие нетранскрибируемые последовательности и 5'- или 3'- нетранслируемые последовательности, такие как необходимые сайты связывания рибосомы, сайт полиаденилирования, донорные и акцепторные сайты сплайсинга и последовательности терминации. Последовательности ДНК, полученные из генома вируса SV40, например, точку начала репликации, ранний промотор, сайты сплайсинга и полиаденилирования SV40, можно использовать для получения других генетических элементов, требуемых для экспрессии гетерологичной последовательности ДНК. Подходящие клонирующие и экспрессионные векторы для применения с клетками-хозяевами, являющимися клетками бактерий, грибов, дрожжей и млекопитающих, описаны в Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

Различные системы культивирования клеток млекопитающих можно использовать для экспрессии и получения рекомбинантного белка. Примеры систем экспрессии на основе клеток млекопитающих включают линии клеток CHO, COS, HEK293, HeLa и BHK. Способы культивирования клеток-хозяев для получения белковых терапевтических средств описаны в Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology)*, Springer (2014). Композиции, описанные в данном документе, могут содержать вектор, такой как вирусный вектор, например, лентивирусный вектор, кодирующий рекомбинантный белок. В некоторых вариантах осуществления вектор, например, вирусный вектор, может содержать нуклеиновую кислоту, кодирующую рекомбинантный белок.

Очистка белковых терапевтических средств описана в Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); и в Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

В некоторых вариантах осуществления стандарты качества включают без ограничения:

(i) длину мРНК, кодирующей полипептид Gene Writer, например, имеет ли мРНК длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК длину более 3000, 4000 или 5000 нуклеотидов;

(ii) присутствие, отсутствие и/или длину поли(A)-хвоста в мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК поли(A)-хвост (например, поли(A)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствие, отсутствие и/или тип 5'-кэпа на мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК 5'-кэп, например, является ли этот кэп 7-метилгуанозиновым кэпом, например, кэпом O-Me-m7G;

(iv) присутствие, отсутствие и/или тип одного или нескольких модифицированных нуклеотидов (например, выбранных из псевдоуридина, дигидроуридина, инозина, 7-метилгуанозина, 1-N-метилпсевдоуридина (1-Me-Ψ), 5-метоксиуридина (5-MO-U), 5-

метилцитидина (5mC) или запертого нуклеотида) в мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК один или несколько модифицированных нуклеотидов;

(v) стабильность мРНК (например, с течением времени и/или в заранее выбранных условиях), например, остается ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% мРНК интактной (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность; или

(vi) активность мРНК в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая мРНК, анализируется на активность.

Наборы, изделия и фармацевтические композиции

В одном из аспектов настоящего изобретения предусмотрен набор, содержащий Gene Writer или систему Gene Writer, например, описанную в данном документе. В некоторых вариантах осуществления набор содержит полипептид Gene Writer (или нуклеиновую кислоту, кодирующую полипептид) и РНК-матрицу (или ДНК, кодирующую РНК-матрицу). В некоторых вариантах осуществления набор дополнительно содержит реагент для введения системы в клетку, например, реагент для трансфекции, LNP и т. п. В некоторых вариантах осуществления набор подходит для любого из описанных в данном документе способов. В некоторых вариантах осуществления набор содержит один или несколько элементов, композиций (например, фармацевтических композиций), Gene Writer и/или систем Gene Writer или их функциональных фрагментов или компонентов, например, размещенных в изделии. В некоторых вариантах осуществления набор содержит инструкции по его применению.

В одном из аспектов настоящего изобретения предусмотрено изделие, например, в котором находится описанный в данном документе набор или его компонент.

В одном из аспектов настоящего изобретения предусмотрена фармацевтическая композиция, содержащая Gene Writer или систему Gene Writer, например, описанную в данном документе. В некоторых вариантах осуществления фармацевтическая композиция дополнительно содержит фармацевтически приемлемый носитель или наполнитель. В некоторых вариантах осуществления фармацевтическая композиция содержит РНК-матрицу и/или РНК, кодирующую полипептид. В вариантах осуществления фармацевтическая композиция обладает одной или несколькими (например, 1, 2, 3 или 4) следующими характеристиками:

(a) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) ДНК-матрицы относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(b) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) некэпированной РНК относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(c) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) РНК частичной длины относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(d) по сути отсутствуют непрореагировавшие динуклеотиды кэпа.

Физико-химические испытания, меры контроля качества и единообразия при производстве (CMC)

Очистка белковых терапевтических средств описана, например, в Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); и в Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

В некоторых вариантах осуществления система Gene Writer™, полипептид и/или нуклеиновая кислота-матрица (например, РНК-матрица) соответствуют определенным стандартам качества. В некоторых вариантах осуществления система Gene Writer™, полипептид и/или нуклеиновая кислота-матрица (например, РНК-матрица), полученные с помощью описанного в данном документе способа, соответствуют определенным стандартам качества. Соответственно, настоящее изобретение в некоторых аспектах относится к способам получения системы Gene Writer™, полипептида и/или нуклеиновой кислоты-матрицы (например, РНК-матрицы), которые соответствуют определенным стандартам качества, например, в которых указанные стандарты качества являются проанализированными. В некоторых аспектах настоящее изобретение также относится к способам анализа указанных стандартов качества в системе Gene Writer™, полипептиде и/или нуклеиновой кислоте-матрице (например, РНК-матрице). В некоторых вариантах осуществления стандарты качества включают без ограничения одно или несколько (например, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 или 12) из следующего:

(i) длину РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы длину более 100, 125, 150, 175 или 200 нуклеотидов;

(ii) присутствие, отсутствие и/или длину поли(A)-хвоста в РНК-матрице, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(A)-хвост (например, поли(A)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствие, отсутствие и/или тип 5'-кэпа в РНК-матрице, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы 5'-кэп, например, является ли этот кэп 7-метилгуанозиновым кэпом, например, кэпом O-Me-m7G;

(iv) присутствие, отсутствие и/или тип одного или нескольких модифицированных нуклеотидов (например, выбранных из псевдоуридина, дигидроуридина, инозина, 7-метилгуанозина, 1-N-метилпсевдоуридина (1-Me-Ψ), 5-метоксиуридина (5-MO-U), 5-метилцитидина (5mC) или запертого нуклеотида) в РНК-матрице, например, содержит ли

по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильность РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остается ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактной (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активность РНК-матрицы в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая РНК-матрицу, анализируется на активность;

(vii) длину полипептида, первого полипептида или второго полипептида, например, имеет ли полипептид, первый полипептид или второй полипептид, длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, длину более 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900 или 2000 аминокислот (и необязательно длину не более 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700 или 600 аминокислот);

(viii) присутствие, отсутствие и/или тип посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% полипептида, первого полипептида или второго полипептида, фосфорилирование, метилирование, ацетилирование, миристоилирование, пальмитоилирование, изопренилирование, гликозилирование или липоилирование или любую их комбинацию;

(ix) присутствие, отсутствие и/или тип одной или нескольких искусственных, синтетических или неканонических аминокислот (например, выбранных из орнитина, β-аланина, GABA, δ-аминолевулиновой кислоты, PABA, D-аминокислоты (например, D-аланина или D-глутамата), аминоизомасляной кислоты, дегидроаланина, цистатионина, лантионина, дженколовой кислоты, диаминопимелиновой кислоты, гомоаланина, норвалина, норлейцина, гомонорлейцина, гомосерина, O-метилгомосерина и O-этилгомосерина, этионина, селеноцистеина, селеногомоцистеина, селенометионина, селеноэтионина, теллуроцистеина или теллурометионина) в полипептиде, первом полипептиде или втором полипептиде, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, одну или несколько искусственных, синтетических или неканонических аминокислот;

(x) стабильность полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остается ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактным (например, длиной более 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450,

1500, 1600, 1700, 1800, 1900 или 2000 аминокислот (и необязательно длиной не более 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700 или 600 аминокислот)) после теста на стабильность;

(xi) активность полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая полипептид, первый полипептид или второй полипептид, анализируется на активность; или

(xii) присутствие, отсутствие и/или уровень одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от пирогенов, вирусов, грибов, бактериальных патогенов или контаминации белками клетки-хозяина.

В некоторых вариантах осуществления система или фармацевтическая композиция, описанные в данном документе, не содержит эндотоксинов.

В некоторых вариантах осуществления определяют присутствие, отсутствие и/или уровень одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов и/или белков клетки-хозяина. В вариантах осуществления определяют, свободна ли или по сути свободна ли система от пирогена, вируса, гриба, бактериального патогена и/или контаминации белками клетки-хозяина.

В некоторых вариантах осуществления фармацевтическая композиция или система, описанная в данном документе, обладает одной или несколькими (например, 1, 2, 3 или 4) следующими характеристиками:

(a) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) ДНК-матрицы относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(b) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) некэпированной РНК относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(c) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) РНК частичной длины относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(d) по сути отсутствуют непрореагировавшие динуклеотиды кэпа.

Пути применения

Благодаря интеграции кодирующих генов в последовательность РНК-матрицы система Gene Writer может удовлетворить терапевтические потребности, например, путем обеспечения экспрессии терапевтического трансгена у индивидуумов с мутациями потери функции, путем замещения мутаций приобретения функции нормальными трансгенами, путем предоставления регуляторных последовательностей для устранения экспрессии мутаций приобретения функции и/или путем контроля экспрессии функционально связанных генов, трансгенов и систем на их основе. В определенных вариантах осуществления последовательность РНК-матрицы кодирует промоторную область,

специфичную для терапевтических потребностей клетки-хозяина, например, тканеспецифический промотор или энхансер. В еще нескольких других вариантах осуществления промотор может быть функционально связан с кодирующей последовательностью.

В вариантах осуществления система редактирования генов Gene Writer™ может предоставлять терапевтические трансгены, экспрессирующие, например, заместительные факторы свертывания крови или заместительные ферменты, например, лизосомальные ферменты. Например, композиции, системы и способы, описанные в данном документе, применимы для экспрессии в геноме-мишени, являющимся человеческим геномом, агалсидазы альфа или бета для лечения болезни Фабри; имиглюцеразы, талиглюцеразы альфа, велаглюцеразы альфа или алглюцеразы для лечения болезни Гоше; себелипазы альфа для лечения дефицита лизосомной кислой липазы (болезни Вольмана/CESD); ларонидазы, идурсульфазы, элосульфазы альфа или галсульфазы для лечения мукополисахаридозов; алглюкозидазы альфа для лечения болезни Помпе. Например, композиции, системы и способы, описанные в данном документе, применимы для экспрессии в геноме-мишени, являющимся человеческим геномом, фактора I, II, V, VII, X, XI, XII или XIII при дефицитах факторов свертывания крови.

В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок, или мембранный белок). В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует мембранный белок, например, мембранный белок, отличный от CAR, и/или эндогенный мембранный белок человека. В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует внеклеточный белок. В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует фермент, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок или запасной белок. Другие иллюстративные белки, которые могут кодироваться последовательностью гетерологичного объекта, включают без ограничения иммунорецепторный белок, например, синтетический иммунорецепторный белок, такой как химерный антигенный рецепторный белок (CAR), Т-клеточный рецептор, В-клеточный рецептор или антитело.

Введение

Композицию и системы, описанные в данном документе, можно применять in vitro или in vivo. В некоторых вариантах осуществления система или компоненты системы доставляются в клетки (например, клетки млекопитающих, например, клетки человека), например, in vitro или in vivo. В некоторых вариантах осуществления клетки представляют собой эукариотические клетки, например, клетки многоклеточного организма, например, животного, например, млекопитающего (например, человека, свиньи, крупного рогатого скота), птицы (например, домашней птицы, такой как курица, индейка или утка) или рыбы. В некоторых вариантах осуществления клетки представляют собой клетки животного,

отличного от человека (например, лабораторного животного, сельскохозяйственного животного или животного-компаньона). В некоторых вариантах осуществления клетка представляет собой стволовую клетку (например, гемопоэтическую стволовую клетку), фибробласт или Т-клетку. В некоторых вариантах осуществления клетка представляет собой неделящуюся клетку, например, неделящийся фибробласт или неделящуюся Т-клетку. В некоторых вариантах осуществления клетка представляет собой HSC, и экспрессия p53 не повышена или повышена на менее чем 10%, 5%, 2% или 1%, например как определено в соответствии со способом, описанным в примере 30 заявки согласно PCT № PCT/US2019/048607, включенной в данный документ посредством ссылки во всей своей полноте. Компоненты системы Gene Writer могут в некоторых случаях быть доставлены в форме полипептида, нуклеиновой кислоты (например, ДНК, РНК) и их комбинаций.

Например, для доставки можно применять любую из следующих комбинаций для доставки ретротранспозазы (например, в виде ДНК, кодирующей белок ретротранспозазу, в виде РНК, кодирующей белок ретротранспозазу, или в виде самого белка) и РНК-матрицы (например, в виде ДНК, кодирующей РНК, или в виде РНК):

1. ДНК, кодирующая ретротранспозазу+ДНК-матрица
2. РНК, кодирующая ретротранспозазу+ДНК-матрица
3. ДНК, кодирующая ретротранспозазу+РНК-матрица
4. РНК, кодирующая ретротранспозазу+РНК-матрица
5. Белок ретротранспозаза+ДНК-матрица
6. Белок ретротранспозаза+РНК-матрица
7. Вирус, несущий ретротранспозазу+вирус, несущий матрицу
8. Вирус, несущий ретротранспозазу+ДНК-матрица
9. Вирус, несущий ретротранспозазу+РНК-матрица
10. ДНК, кодирующая ретротранспозазу+вирус, несущий матрицу
11. РНК, кодирующая ретротранспозазу+вирус, несущий матрицу
12. Белок ретротранспозаза+вирус, несущий матрицу

Как указано выше, в некоторых вариантах осуществления ДНК или РНК, кодирующая белок ретротранспозазу, доставляется с помощью вируса, а в некоторых вариантах осуществления РНК-матрица (или ДНК, кодирующая РНК-матрицу) доставляется с помощью вируса.

В одном из вариантов осуществления система и/или компоненты системы доставляются в виде нуклеиновой кислоты. Например, полипептид Gene Writer может быть доставлен в форме ДНК или РНК, кодирующей полипептид, а РНК-матрица может быть доставлена в форме РНК или комплементарной ей ДНК для транскрипции в РНК. В некоторых вариантах осуществления система или компоненты системы доставляются в 1, 2, 3, 4 или более отдельных молекулах нуклеиновых кислот. В некоторых вариантах осуществления система или компоненты системы доставляются в виде комбинации ДНК и РНК. В некоторых вариантах осуществления система или компоненты системы доставляются в виде комбинации ДНК и белка. В некоторых вариантах осуществления

система или компоненты системы доставляются в виде комбинации РНК и белка. В некоторых вариантах осуществления полипептид редактора генома Gene Writer доставляется в виде белка.

В некоторых вариантах осуществления система или компоненты системы доставляются в клетки, например, клетки млекопитающих или клетки человека, с помощью вектора. Вектор может представлять собой, например, плазмиду или вирус. В некоторых вариантах осуществления доставка осуществляется in vivo, in vitro, ex vivo или in situ. В некоторых вариантах осуществления вирус представляет собой аденоассоциированный вирус (AAV), лентивирус, аденовирус. В некоторых вариантах осуществления система или компоненты системы доставляются в клетки с помощью вирусоподобной частицы или виросомы. В некоторых вариантах осуществления для доставки используют более одного вируса, вирусоподобной частицы или виросомы.

В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид, или ДНК-матрицу, или и то, и другое), доставляемая в клетки, представляет собой ковалентно замкнутую линейную ДНК или так называемую ДНК "Doggybone". В течение своего жизненного цикла бактериофаг N15 использует протеломеразу для преобразования своего генома из кольцевой плазмидной ДНК в линейную плазмидную ДНК (Ravin et al. J Mol Biol 2001). Данный процесс был адаптирован для получения ковалентно замкнутой линейной ДНК in vitro (см., например, WO2010086626A1). В некоторых вариантах осуществления протеломераза контактирует с ДНК, содержащей один или несколько сайтов распознавания протеломеразы, где протеломераза приводит к разрезанию в одном или нескольких сайтах и последующему лигированию комплементарных нитей ДНК, что приводит к образованию ковалентной связи между комплементарными нитями. В некоторых вариантах осуществления нуклеиновую кислоту (например, кодирующую транспозазу, или ДНК-матрицу, или и то, и другое) вначале получают в виде кольцевой плазмидной ДНК, содержащей один сайт распознавания протеломеразы, которую затем приводят в контакт с протеломеразой для получения ковалентно замкнутой линейной ДНК. В некоторых вариантах осуществления нуклеиновую кислоту (например, кодирующую транспозазу, или ДНК-матрицу, или и то, и другое), фланкированную сайтами распознавания протеломеразы в плазмидной или линейной ДНК, приводят в контакт с протеломеразой для получения ковалентно замкнутой линейной ДНК, содержащей только ДНК, заключенную между сайтами распознавания протеломеразы. В некоторых вариантах осуществления подход с фланкированием желаемой последовательности нуклеиновой кислоты сайтами распознавания протеломеразы приводит к образованию ковалентно замкнутой кольцевой ДНК без плазмидных элементов, используемых для клонирования и поддержания в бактериях. В некоторых вариантах осуществления плазмидную или линейную ДНК, содержащую нуклеиновую кислоту и один или несколько сайтов распознавания протеломеразы, необязательно амплифицируют перед реакцией с протеломеразой, например, посредством амплификации по типу катящегося кольца или ПЦР.

В одном варианте осуществления композиции и системы, описанные в данном документе, могут быть составлены в липосомах или других сходных везикулах. Липосомы представляют собой сферические везикулярные структуры, состоящие из уни- или мультиламеллярного липидного бислоя, окружающего внутренние водные компартменты, и относительно непроницаемого внешнего липофильного фосфолипидного бислоя. Липосомы могут быть анионными, нейтральными или катионными. Липосомы являются биосовместимыми, нетоксичными, могут доставлять как гидрофильные, так и липофильные молекулы лекарственных средств, защищать свой "груз" от разрушения под действием ферментов плазмы крови и транспортировать свою нагрузку через биологические мембраны и гематоэнцефалический барьер (BBB) (см., например, для обзора Spuch and Navarro, Journal of Drug Delivery, vol. 2011, ID статьи 469679, 12 страниц, 2011. doi:10,1155/2011/469679).

Везикулы можно получать из нескольких различных типов липидов; тем не менее, для образования липосом в качестве носителей лекарственных средств чаще всего используются фосфолипиды. Способы получения мультиламеллярных липидных везикул известны из уровня техники (см., например, патент США № 6693086, идеи которого, относящиеся к получению мультиламеллярных липидных везикул, включены в данный документ посредством ссылки). Несмотря на то, что образование везикул может быть спонтанным, когда липидная пленка смешивается с водным раствором, его также можно ускорить путем применения силы в форме встряхивания с использованием гомогенизатора, соникатора или экструзионного аппарата (см., например, для обзора Spuch and Navarro, Journal of Drug Delivery, vol. 2011, ID статьи 469679, 12 страниц, 2011. doi:10.1155/2011/469679). Экструдированные липиды можно получать посредством экструзии через фильтры с уменьшающимся размером пор, как описано в Templeton et al., Nature Biotech, 15:647-652, 1997, идеи которой, относящиеся к получению экструдированных липидов, включены в данный документ посредством ссылки.

Липидные наночастицы являются другим примером носителя, который представляет собой биосовместимую и биоразлагаемую систему доставки для фармацевтических композиций, описанных в данном документе. Наноструктурированные липидные носители (NLC) представляют собой модифицированные твердые липидные наночастицы (SLN), которые сохраняют характеристики SLN, улучшают стабильность и емкость загрузки лекарственного средства и предотвращают утечку лекарственного средства. Полимерные наночастицы (PNP) являются важным компонентом для доставки лекарственных средств. Эти наночастицы могут эффективно направлять доставку лекарственного средства к конкретным мишеням и улучшать стабильность лекарственного средства и контролируемое высвобождение лекарственного средства. Также можно использовать липидно-полимерные наночастицы (PLN) - новый тип носителя, который является комбинацией липосом и полимеров. Эти наночастицы обладают взаимодополняющими преимуществами PNP и липосом. PLN состоит из структуры сердцевина-оболочка; полимерная сердцевина обеспечивает стабильную структуру, а

фосфолипидная оболочка обеспечивает хорошую биосовместимость. Таким образом, эти два компонента увеличивают показатель эффективности инкапсуляции лекарственного средства, способствуют модификации поверхности и предотвращают утечку водорастворимых лекарственных средств. Для обзора см., например, Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Экзосомы также можно применять в качестве носителей для доставки лекарственных средств для композиций и систем, описанных в данном документе. Для обзора см. Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; https://doi.org/10.1016/j.apsb.2016.02.001.

Фузосомы взаимодействуют и сливаются с клетками-мишенями и, таким образом, могут использоваться в качестве средств доставки различных молекул. Обычно они состоят из бислоя амфипатических липидов, окружающего просвет или полость, и фузогена, который взаимодействует с бислоем амфипатических липидов. Было показано, что фузогенный компонент можно сконструировать для придания специфичности к клеткам-мишеням для слияния и доставки полезной нагрузки, что позволяет создавать средства доставки с программируемой специфичностью к клеткам (касательно разработки, получения и применения фузосом см., например, публикацию согласно PCT № WO/2020014209, включенную в данный документ посредством ссылки).

Систему Gene Writer можно вводить в клетки, ткани и многоклеточные организмы. В некоторых вариантах осуществления система или компоненты системы доставляются в клетки с помощью механических средств или физических средств.

Составление белковых терапевтических средств описано в Meyer (Ed.), *Therapeutic Protein Drug Products: Practical Approaches to formulation in the Laboratory, Manufacturing, and the Clinic*, Woodhead Publishing Series (2012).

16. Тканеспецифическая активность/введение

В некоторых вариантах осуществления система, РНК-матрица или полипептид, описанные в данном документе, вводятся в ткань-мишень, например, в первую ткань, являются активными (например, более активными) в ней. В некоторых вариантах осуществления система, РНК-матрица или полипептид не вводятся в ткань, не являющуюся мишенью, или являются менее активными (например, не являются активными) в ней. В некоторых вариантах осуществления система, РНК-матрица или полипептид, описанные в данном документе, применимы для модификации ДНК в ткани-мишени, например, в первой ткани (например, без модификации ДНК в ткани, не являющейся мишенью).

В некоторых вариантах осуществления система содержит следующее: (a) полипептид, описанный в данном документе, или нуклеиновую кислоту, кодирующую его, (b) нуклеиновую кислоту-матрицу (например, РНК-матрицу), описанную в данном документе, и (c) одну или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных в отношении ткани-мишени, где одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных в отношении ткани-мишени, находятся в функциональной

ассоциации с (a), (b) или (a) и (b), где при ассоциации с (a) (a) предусматривает нуклеиновую кислоту, кодирующую полипептид.

В некоторых вариантах осуществления нуклеиновая кислота в (b) содержит РНК.

В некоторых вариантах осуществления нуклеиновая кислота в (b) содержит ДНК.

В некоторых вариантах осуществления нуклеиновая кислота в (b): (i) является однонитевой или содержит однонитевой сегмент, например, представляет собой однонитевую ДНК или содержит однонитевой сегмент и один или несколько двухнитевых сегментов; (ii) имеет инвертированные концевые повторы; или (iii) как (i), так и (ii).

В некоторых вариантах осуществления нуклеиновая кислота в (b) является двухнитевой или содержит двухнитевой сегмент.

В некоторых вариантах осуществления (a) содержит нуклеиновую кислоту, кодирующую полипептид.

В некоторых вариантах осуществления нуклеиновая кислота в (a) содержит РНК.

В некоторых вариантах осуществления нуклеиновая кислота в (a) содержит ДНК.

В некоторых вариантах осуществления нуклеиновая кислота в (a): (i) является однонитевой или содержит однонитевой сегмент, например, представляет собой однонитевую ДНК или содержит однонитевой сегмент и один или несколько двухнитевых сегментов; (ii) имеет инвертированные концевые повторы; или (iii) как (i), так и (ii).

В некоторых вариантах осуществления нуклеиновая кислота в (a) является двухнитевой или содержит двухнитевой сегмент.

В некоторых вариантах осуществления нуклеиновая кислота в (a), (b) или (a) и (b) является линейной.

В некоторых вариантах осуществления нуклеиновая кислота в (a), (b) или (a) и (b) является кольцевой, например, представляет собой плазмиду или миникольцо.

В некоторых вариантах осуществления последовательность гетерологичного объекта находится в функциональной ассоциации с первым промотором.

В некоторых вариантах осуществления одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, содержат тканеспецифический промотор.

В некоторых вариантах осуществления тканеспецифический промотор содержит первый промотор в функциональной ассоциации с: i. последовательностью гетерологичного объекта, ii. нуклеиновой кислотой, кодирующей транспозазу, или iii. (i) и (ii).

В некоторых вариантах осуществления одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, содержат тканеспецифическую последовательность распознавания микроРНК в функциональной ассоциации с: i. последовательностью гетерологичного объекта, ii. нуклеиновой кислотой, кодирующей транспозазу, или iii. (i) и (ii).

В некоторых вариантах осуществления система содержит тканеспецифический промотор, и система дополнительно содержит одну или несколько тканеспецифических

последовательностей распознавания микроРНК, где: i. тканеспецифический промотор находится в функциональной ассоциации с: I. последовательностью гетерологичного объекта, II. нуклеиновой кислотой, кодирующей транспозазу, или III. (I) и (II); и/или ii. одна или несколько тканеспецифических последовательностей распознавания микроРНК находятся в функциональной ассоциации с: I. последовательностью гетерологичного объекта, II. нуклеиновой кислотой, кодирующей транспозазу, или III. (I) и (II).

В некоторых вариантах осуществления, где (a) содержит нуклеиновую кислоту, кодирующую полипептид, нуклеиновая кислота содержит промотор в функциональной ассоциации с нуклеиновой кислотой, кодирующей полипептид.

В некоторых вариантах осуществления нуклеиновая кислота, кодирующая полипептид, содержит одну или несколько вторых тканеспецифических последовательностей, контролирующих экспрессию, специфичных для ткани-мишени, в функциональной ассоциации с последовательностью, кодирующей полипептид.

В некоторых вариантах осуществления одна или несколько вторых тканеспецифических последовательностей, контролирующих экспрессию, содержат тканеспецифический промотор.

В некоторых вариантах осуществления тканеспецифический промотор представляет собой промотор в функциональной ассоциации с нуклеиновой кислотой, кодирующей полипептид.

В некоторых вариантах осуществления одна или несколько вторых тканеспецифических последовательностей, контролирующих экспрессию, содержат тканеспецифическую последовательность распознавания микроРНК.

В некоторых вариантах осуществления промотор в функциональной ассоциации с нуклеиновой кислотой, кодирующей полипептид, представляет собой тканеспецифический промотор, при этом система дополнительно содержит одну или несколько тканеспецифических последовательностей распознавания микроРНК.

В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, доставляется в ткань или клетку головного мозга, мозжечка, надпочечников, яичников, поджелудочной железы, паращитовидной железы, гипофиза, яичек, щитовидной железы, молочной железы, селезенки, миндалин, тимуса, лимфатического узла, костного мозга, легкого, сердечной мышцы, пищевода, желудка, тонкой кишки, толстой кишки, печени, слюнной железы, почки, предстательной железы, крови или в другой тип клеток или тканей. В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, применяется для лечения заболевания, такого как рак, воспалительное заболевание, инфекционное заболевание, генетический дефект или другое заболевание. Рак может являться раком головного мозга, мозжечка, надпочечников, яичников, поджелудочной железы, паращитовидной железы, гипофиза, яичек, щитовидной железы, молочной железы, селезенки, миндалин, тимуса, лимфатического узла, костного мозга, легкого, сердечной мышцы, пищевода, желудка, тонкой кишки, толстой кишки, печени,

слюнной железы, почки, предстательной железы, крови или из другого типа клеток или тканей или может включать несколько видов рака.

В некоторых вариантах осуществления описанная в данном документе система Gene Writer™ вводится посредством энтерального введения (например, перорального, ректального, желудочно-кишечного, сублингвального, сублабиального или трансбуккального введения). В некоторых вариантах осуществления систему Gene Writer™, описанную в данном документе, вводят посредством парентерального введения (например, внутривенного, внутримышечного, подкожного, внутрикожного, эпидурального, внутримозгового, интрацеребровентрикулярного, накожного, назального, внутриартериального, внутрисуставного, внутрикавернозного, внутриглазного, внутрикостной инфузии, внутрибрюшинного, интратекального, внутриматочного, интравагинального, внутривезикального, периваскулярного или чресслизистого введения). В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, вводится посредством местного введения (например, трансдермального введения).

В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться для модификации клеток животных, клеток растений или клеток грибов. В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться для модификации клетки млекопитающего (например, клетки человека). В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться для модификации клетки домашнего скота (например, коровы, лошади, овцы, козы, свиньи, ламы, альпаки, верблюда, яка, курицы, утки, гуся или страуса). В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться в качестве лабораторного инструмента или исследовательского инструмента или использоваться в лабораторном методе или методе исследования, например, для модификации клетки животного, например, клетки млекопитающего (например, клетки человека), клетки растения или клетки гриба.

В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться для экспрессии белка, матрицы или последовательности гетерологичного объекта (например, в клетке млекопитающего (например, клетке человека), клетке растения или клетке гриба). В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, может применяться для экспрессии белка, матрицы или последовательности гетерологичного объекта под контролем индуцируемого промотора (например, промотора, индуцируемого малыми молекулами). В некоторых вариантах осуществления система Gene Writing или ее полезная нагрузка предназначены для настраиваемого контроля, например, посредством применения индуцируемого промотора. Например, промотор, например, Tet, управляющий геном, представляющим интерес, может являться "молчащим" при интеграции, но в некоторых случаях может активироваться при воздействии низкомолекулярного индуктора, например,

доксициклина. В некоторых вариантах осуществления настраиваемая экспрессия позволяет контролировать ген (например, терапевтический ген) после обработки, например, обеспечивая эффект дозирования, зависящий от малых молекул. В вариантах осуществления эффект дозирования, зависящий от малых молекул, включает изменение уровней продукта гена во времени и/или в пространстве, например, посредством местного введения. В некоторых вариантах осуществления промотор, используемый в системе, описанной в данном документе, может быть индуцируемым, например, отвечающим на эндогенную молекулу хозяина и/или экзогенную малую молекулу, вводимую ему.

В некоторых вариантах осуществления система Gene Writing применяется для внесения изменений в некодирующие области и/или регуляторные контрольные области, например, для настройки экспрессии эндогенных генов. В некоторых вариантах осуществления система Gene Writing применяется для индукции повышения или понижения экспрессии генов. В некоторых вариантах осуществления регуляторная контрольная область содержит один или несколько из промотора, энхансера, UTR, сайта CTCF и/или области, контролирующей экспрессию гена.

В некоторых вариантах осуществления система Gene Writing применяется для лечения или предупреждения болезни экспансии повторов (например, заболевания из таблицы 26) или для уменьшения его тяжести или симптома. В некоторых вариантах осуществления болезнь экспансии повторов включает экспансию тринуклеотидных повторов. В некоторых вариантах осуществления у субъекта имеется по меньшей мере 10, 20, 30, 40 или 50 копий повтора. В вариантах осуществления болезнь экспансии повторов является наследственным заболеванием. Неограничивающие примеры болезней экспансии повторов включают болезнь Хантингтона (HD) и миотоническую дистрофию. Например, здоровые люди могут иметь от 10 до 35 тандемных копий тринуклеотидного повтора CAG, тогда как пациенты с болезнью Хантингтона часто имеют > 40 копий, что может привести, например, к образованию удлиненного и дисфункционального белка гентингтина. В некоторых вариантах осуществления Gene Writer корректирует экспансию повтора, например, распознавая ДНК на конце области повторов и внося однонитевой разрез в одну нить (фигура 30). В некоторых вариантах осуществления компонент Gene Writer, представляющий собой РНК-матрицу, содержит область с числом повторов, характерным для здорового субъекта, например, приблизительно 20 повторов (например, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35 или 35-40 повторов). В некоторых вариантах осуществления компонент Gene Writer, представляющий собой РНК-матрицу, копируется с помощью TPRT в сайт-мишень. В некоторых вариантах осуществления однонитевой разрыв во второй нити и синтез второй нити затем приводят к интеграции новой скопированной ДНК, содержащей правильное количество повторов (например, как описано в данном документе). В некоторых вариантах осуществления система распознает ДНК на конце области повторов, и при этом матрица несет информацию о новом количестве повторов. В вариантах осуществления таким образом можно применять Gene Writer независимо от количества повторов, присутствующих у индивидуума и/или в отдельной клетке. Из-за

наличия множественных повторов альтернативное терапевтическое средство, отличное от Gene Writer (например, терапевтическое средство для гомологичной рекомбинации на основе CRISPR), в некоторых вариантах осуществления может привести к непредсказуемому поведению репарации. Дополнительные неограничивающие примеры болезней экспансии повторов и вызывающих их повторов можно найти, например, в La Spada and Taylor Nat Rev Genet 11(4):247-258 (2010), которая включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления систему Gene Writing можно применять для лечения здорового человека, например, в качестве профилактической терапии. В некоторых вариантах осуществления системы Gene Writing могут быть нацелены на создание мутаций, например, которые, как было показано, защищают от заболевания, представляющего интерес. Иллюстративный перечень таких заболеваний и мишеней защитных мутаций можно найти в таблице 22.

В некоторых вариантах осуществления компонент, представляющий собой нуклеиновую кислоту, в системе, предусмотренной в настоящем изобретении, представляет собой последовательность (например, кодирующую полипептид или содержащую последовательность гетерологичного объекта), фланкированную нетранслируемыми областями (UTR), которые модифицируют уровни экспрессии белка. Различные 5'- и 3'-UTR могут влиять на экспрессию белка. Например, в некоторых вариантах осуществления кодирующей последовательности может предшествовать 5'-UTR, которая модифицирует стабильность РНК или трансляцию белка. В некоторых вариантах осуществления за последовательностью может следовать 3'-UTR, которая модифицирует стабильность РНК или трансляцию белка. В некоторых вариантах осуществления последовательности может предшествовать 5'-UTR, и за ней может следовать 3'-UTR, которые модифицируют стабильность РНК или трансляцию. В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны из 5'- и 3'-UTR фактора комплемента 3 (C3) (cactcctccccatcctctccctctgtccctctgtccctctgaccctgcactgtcccagcacc) или орозомукоида 1 (ORM1) (caggacacagccttggatcaggacagagacttgggggccatcctgcccctccaacccgacatgtgtacctcagctttttccctcacttgcat caataaagcttctgtgtttggaacagctaa) (Asrani et al. RNA Biology 2018). В определенных вариантах осуществления 5'-UTR представляет собой 5'-UTR из C3, и 3'-UTR представляет собой 3'-UTR из ORM1.

В определенных вариантах осуществления 5'-UTR и 3'-UTR для экспрессии белка, например, в мРНК (или ДНК, кодирующей РНК) для полипептида Gene Writer или последовательности гетерологичного объекта, содержат оптимизированные последовательности для экспрессии. В некоторых вариантах осуществления 5'-UTR содержит GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC и/или 3'-UTR содержит UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA,

например, как описано в Richner et al. *Cell* 168(6): P1114-1125 (2017), последовательности из которого включены в данный документ посредством ссылки.

В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны для усиления экспрессии белка. В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны для модификации экспрессии белка таким образом, чтобы минимизировать ингибирование избыточного продуцирования. В некоторых вариантах осуществления UTR расположены вокруг кодирующей последовательности, например, вне кодирующей последовательности, и в других вариантах осуществления находятся проксимальнее кодирующей последовательности. В некоторых вариантах осуществления в UTR включены дополнительные регуляторные элементы (например, сайты связывания с miRNA, цис-регуляторные сайты).

В некоторых вариантах осуществления открытая рамка считывания (ORF) системы Gene Writer, например, ORF мРНК (или ДНК, кодирующей мРНК), кодирующая полипептид Gene Writer, или одна или несколько ORF мРНК (или ДНК, кодирующей мРНК) в последовательности гетерологичного объекта фланкированы 5'- и/или 3'-нетранслируемой областью (UTR), которая усиливает их экспрессию. В некоторых вариантах осуществления 5'-UTR мРНК-компонента (или транскрипта, полученного из ДНК-компонента) системы содержит последовательность 5'-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3'. В некоторых вариантах осуществления 3'-UTR мРНК-компонента (или транскрипта, полученного из ДНК-компонента) системы содержит последовательность 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3'. Данная комбинация 5'-UTR и 3'-UTR приводит к желаемой экспрессии функционально связанной ORF, как было показано в Richner et al. *Cell* 168(6): P1114-1125 (2017), идеи и последовательности из которого включены в данный документ посредством ссылки. В некоторых вариантах осуществления система, описанная в данном документе, содержит ДНК, кодирующую транскрипт, где ДНК содержит соответствующие последовательности 5'-UTR и 3'-UTR, где T заменяет U в вышеперечисленной последовательности). В некоторых вариантах осуществления ДНК-вектор, применяемый для получения РНК-компонента системы, дополнительно содержит промотор для инициации транскрипции in vitro выше 5'-UTR, например, промотор T7, T3 или SP6. Вышеуказанная 5'-UTR начинается с GGG, что является подходящим сайтом начала для оптимизации транскрипции с применением РНК-полимеразы T7. Для настройки уровней транскрипции и изменения нуклеотидов в сайте начала транскрипции таким образом, чтобы они соответствовали альтернативным 5'-UTR, в идеях Davidson et al. *Pac Symp Biocomput* 433-443 (2010) описаны варианты промотора T7 и способы их обнаружения, которые соответствуют обоим этим признакам.

17. Вирусные векторы и их компоненты

Вирусы являются полезным источником средств доставки для систем, описанных в данном документе, в дополнение к источнику соответствующих ферментов или доменов, описанных в данном документе, например, в качестве источников полимераз и функций полимеразы, используемых в данном документе, например, ДНК-зависимой ДНК-полимеразы, РНК-зависимой РНК-полимеразы, РНК-зависимой ДНК-полимеразы, ДНК-зависимой РНК-полимеразы, обратной транскриптазы. Некоторые ферменты, например, обратные транскриптазы, могут характеризоваться несколькими видами активности, например, быть способными как к РНК-зависимой полимеризации ДНК, так и к ДНК-зависимой полимеризации ДНК, например, к синтезу первой и второй нитей. В некоторых вариантах осуществления вирус, используемый в качестве системы доставки Gene Writer или источника ее компонентов, может быть выбран из группы, описанной в Baltimore *Bacteriol Rev* 35(3):235-241 (1971).

В некоторых вариантах осуществления вирус выбран из вируса группы I, например, представляет собой ДНК-вирус и упаковывает dsDNA в вирионы. В некоторых вариантах осуществления вирус группы I выбран, например, из аденовирусов, герпесвирусов, поксвирусов.

В некоторых вариантах осуществления вирус выбран из вируса группы II, например, представляет собой ДНК-вирус и упаковывает ssDNA в вирионы. В некоторых вариантах осуществления вирус группы II выбран, например, из парвовирусов. В некоторых вариантах осуществления парвовирус представляет собой депендопарвовирус, например, аденоассоциированный вирус (AAV).

В некоторых вариантах осуществления вирус выбран из вируса группы III, например, представляет собой РНК-вирус и упаковывает dsRNA в вирионы. В некоторых вариантах осуществления вирус группы III выбран, например, из реовирусов. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, представляют собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, могут непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты.

В некоторых вариантах осуществления вирус выбран из вируса группы IV, например, представляет собой РНК-вирус и упаковывает ssRNA(+) в вирионы. В некоторых вариантах осуществления вирус группы IV выбран, например, из коронавирусов, пикорнавирусов, тогавирусов. В некоторых вариантах осуществления ssRNA(+), содержащаяся в таких вирионах, представляет собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, может непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты.

В некоторых вариантах осуществления вирус выбран из вируса группы V, например, представляет собой РНК-вирус и упаковывает ssRNA(-) в вирионы. В некоторых вариантах

осуществления вирус группы V выбран, например, из ортомиксовирусов, рабдовирусов. В некоторых вариантах осуществления РНК-вирус с ssRNA(-)-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую РНК-полимеразу, способную копировать ssRNA(-) в ssRNA(+), которая может транслироваться непосредственно хозяином.

В некоторых вариантах осуществления вирус выбран из вируса группы VI, например, представляет собой ретровирус и упаковывает ssRNA(+) в вирионы. В некоторых вариантах осуществления вирус группы VI выбран, например, из ретровирусов. В некоторых вариантах осуществления ретровирус представляет собой лентивирус, например, HIV-1, HIV-2, SIV, BIV. В некоторых вариантах осуществления ретровирус представляет собой спумавирус, например, вирус пенистости, например, HFV, SFV, BFV. В некоторых вариантах осуществления ssRNA(+), содержащаяся в таких вирионах, представляет собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, может непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты. В некоторых вариантах осуществления ssRNA(+) сначала подвергается обратной транскрипции и копированию с образованием промежуточного dsDNA-генома, из которого мРНК может транскрибироваться в клетке-хозяине. В некоторых вариантах осуществления РНК-вирус с ssRNA(-)-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую ДНК-полимеразу, способную копировать ssRNA(+) в dsDNA, которая может транскрибироваться в мРНК и транслироваться хозяином. В некоторых вариантах осуществления обратная транскриптаза из ретровируса группы VI включена в качестве домена обратной транскриптазы полипептида Gene Writer.

В некоторых вариантах осуществления вирус выбран из вируса группы VII, например, представляет собой ретровирус и упаковывает dsRNA в вирионы. В некоторых вариантах осуществления вирус группы VII выбран, например, из гепаднавирусов. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, представляют собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, могут непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, сначала подвергаются обратной транскрипции и копированию с образованием промежуточного dsDNA-генома, из которого мРНК может транскрибироваться в клетке-хозяине. В некоторых вариантах осуществления РНК-вирус с dsRNA-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую ДНК-полимеразу, способную копировать dsRNA в dsDNA, которая может транскрибироваться в мРНК и транслироваться хозяином.

В некоторых вариантах осуществления обратная транскриптаза из ретровируса группы VII включена в качестве домена обратной транскриптазы полипептида Gene Writer.

В некоторых вариантах осуществления вирионы, используемые для доставки нуклеиновой кислоты в настоящем изобретении, могут также нести ферменты, участвующие в процессе Gene Writing. Например, ретровирусный вирион может содержать домен обратной транскриптазы, который доставляется в клетку-хозяина вместе с нуклеиновой кислотой. В некоторых вариантах осуществления РНК-матрица может быть ассоциирована с полипептидом Gene Writer внутри вириона, так что оба они совместно доставляются в клетку-мишень при трансдукции нуклеиновой кислоты из вирусной частицы. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может содержать ДНК, например, линейную ssDNA, линейную dsDNA, кольцевую ssDNA, кольцевую dsDNA, миникольцевую DNA, dbDNA, ceDNA. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может содержать РНК, например, линейную ssRNA, линейную dsRNA, кольцевую ssRNA, кольцевую dsRNA. В некоторых вариантах осуществления вирусный геном может циркуляризоваться при трансдукции клетки-хозяина, например, линейная молекула ssRNA может участвовать в ковалентной связи с образованием кольцевой ssRNA, линейная молекула dsRNA может участвовать в ковалентной связи с образованием кольцевой dsRNA или одной или нескольких кольцевых ssRNA. В некоторых вариантах осуществления вирусный геном может реплицироваться путем репликации по типу катящегося кольца в клетке-хозяине. В некоторых вариантах осуществления вирусный геном может содержать одну молекулу нуклеиновой кислоты, например, содержать несегментированный геном. В некоторых вариантах осуществления вирусный геном может содержать две или более молекулы нуклеиновых кислот, например, содержать сегментированный геном. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может быть ассоциирована с одним или несколькими белками. В некоторых вариантах осуществления один или несколько белков вириона могут быть доставлены в клетку-хозяина при трансдукции. В некоторых вариантах осуществления природный вирус может быть адаптирован для доставки нуклеиновой кислоты посредством добавления сигналов упаковки вириона к нуклеиновой кислоте-мишени, где клетка-хозяин используется для упаковки нуклеиновой кислоты-мишени, содержащей сигналы упаковки.

В некоторых вариантах осуществления вирион, используемый в качестве средства доставки, может содержать комменсальный вирус человека. В некоторых вариантах осуществления вирион, используемый в качестве средства доставки, может содержать анелловирус, применение которого описано в WO2018232017A1, которая включена в данный документ посредством ссылки во всей своей полноте.

Аденоассоциированные вирусы

В некоторых вариантах осуществления вирус представляет собой аденоассоциированный вирус (AAV). В некоторых вариантах осуществления геном AAV содержит два гена, которые кодируют четыре белка репликации и три капсидных белка соответственно. В некоторых вариантах осуществления гены фланкированы с обеих сторон

инвертированными концевыми повторами (ITR) размером 145 п. о. В некоторых вариантах осуществления вирион содержит до трех капсидных белков (Vp1, Vp2 и/или Vp3), например, продуцируемых в соотношении 1:1:10. В некоторых вариантах осуществления капсидные белки получены из одной и той же открытой рамки считывания и/или из сайта дифференциального сплайсинга (Vp1) и альтернативных сайтов начала трансляции (Vp2 и Vp3 соответственно). В целом Vp3 является наиболее распространенной субъединицей вириона и принимает участие в распознавании рецепторов на клеточной поверхности, определяя тропизм вируса. В некоторых вариантах осуществления Vp1 содержит фосфолипазный домен, например, который функционирует для обеспечения инфекционности вируса, на N-конце Vp1.

В некоторых вариантах осуществления упаковочная способность вирусных векторов ограничивает размер редактора оснований, который может быть упакован в вектор. Например, упаковочная способность AAV может составлять приблизительно 4,5 т. о. (например, приблизительно 3,0, 3,5, 4,0, 4,5, 5,0, 5,5 или 6,0 т. о.), например, включая один или два инвертированных концевых повтора (ITR), например, ITR размером 145 оснований.

В некоторых вариантах осуществления рекомбинантный AAV (rAAV) содержит цис-действующие ITR размером 145 п. о., фланкирующие кассеты трансгена в векторе, например, обеспечивая до 4,5 т. о. для упаковки чужеродной ДНК. После инфицирования rAAV в некоторых случаях может экспрессировать слитый белок по настоящему изобретению и персистировать без интеграции в геном хозяина, существуя эписомально в виде кольцевых конкатемеров "голова к хвосту". rAAV можно использовать, например, in vitro и in vivo. В некоторых вариантах осуществления опосредованная AAV доставка гена требует, чтобы длина кодирующей последовательности гена была равна геному AAV дикого типа или превышала его по размеру.

Доставка с помощью AAV генов, которые превышают этот размер, и/или применение больших физиологических регуляторных элементов могут быть осуществлены, например, посредством разделения белка(белков), который(которые) должен(должны) быть доставлен(доставлены), на два или более фрагмента. В некоторых вариантах осуществления N-концевой фрагмент слит с расщепленным интеином-N. В некоторых вариантах осуществления C-концевой фрагмент слит с расщепленным интеином-C. В вариантах осуществления фрагменты упакованы в два или более вектора на основе AAV.

В некоторых вариантах осуществления двойные векторы на основе AAV создаются с помощью разделения крупной кассеты экспрессии трансгена на две отдельные половины (5'- и 3'-концы или голову и хвост), например, где каждая половина кассеты упакована в отдельный вектор на основе AAV (размером < 5 т. о.). Повторная сборка полноразмерной кассеты экспрессии трансгена в некоторых вариантах осуществления может быть затем достигнута при совместном инфицировании одной и той же клетки обоими двойными векторами на основе AAV. В некоторых вариантах осуществления после совместного инфицирования следует одно или несколько из: (1) гомологичной рекомбинации (HR)

между 5'- и 3'-концевыми геномами (двойные перекрывающиеся векторы на основе AAV); (2) ITR-опосредованной конкатемеризации 5'- и 3'-концевых геномов "голова к хвосту" (двойные транс-сплайсинговые векторы на основе AAV) и/или (3) комбинации этих двух механизмов (двойные гибридные векторы на основе AAV). В некоторых вариантах осуществления применение двойных векторов на основе AAV in vivo приводит к экспрессии полноразмерных белков. В некоторых вариантах осуществления применение платформы двойного вектора на основе AAV представляет собой эффективную и жизнеспособную стратегию переноса генов для трансгенов размером более чем приблизительно 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9 или 5,0 т. о. В некоторых вариантах осуществления векторы на основе AAV также можно применять для трансдукции клеток нуклеиновыми кислотами-мишенями, например, при продуцировании нуклеиновых кислот и пептидов in vitro. В некоторых вариантах осуществления векторы на основе AAV можно применять для процедур генной терапии in vivo и ex vivo (см., например, West et al., Virology 160:38-47 (1987); патент США № 4797368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest.94:1351 (1994); каждый из которых включен в данный документ посредством ссылки во всей своей полноте). Конструирование рекомбинантных векторов на основе AAV описано в ряде публикаций, в том числе в патенте США № 5173414; Tratschin et al., Mol. Cell. Biol.5:3251- 3260 (1985); Tratschin, et al., Mol. Cell. Biol.4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); и Samulski et al., J. Virol.63:03822-3828 (1989) (включенных в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления Gene Writer, описанный в данном документе (например, с одной или несколькими направляющими нуклеиновыми кислотами или без них), может быть доставлен с применением AAV, лентивируса, аденовируса или других типов плазмидных или вирусных векторов, в частности, с использованием составов и доз из, например, патента США № 8454972 (составы, дозы для аденовируса), патента США № 8404658 (составы, дозы для AAV) и патента США № 5846946 (составы, дозы для ДНК-плазмид), а также из клинических испытаний и публикаций, касающихся клинических испытаний с участием лентивирусов, AAV и аденовирусов. Например, для AAV путь введения, состав и доза могут быть такими, как описано в патенте США № 8454972 и в клинических испытаниях с использованием AAV. Для аденовируса путь введения, состав и доза могут быть такими, как описано в патенте США № 8404658 и в клинических испытаниях с использованием аденовируса. Для доставки плазмиды путь введения, состав и доза могут быть такими, как описано в патенте США № 5846946 и в клинических испытаниях с использованием плазмид. Дозы могут быть рассчитаны на основе средней массы индивидуума 70 кг (например, взрослого человека мужского пола) или экстраполированы на нее и могут быть скорректированы для пациентов, субъектов, млекопитающих разной массы и разных видов. Частота введения регламентируется практикующим врачом или ветеринаром (например, терапевтом, ветеринаром) в зависимости от обычных факторов, включая возраст, пол, общее состояние здоровья,

другие состояния пациента или субъекта и конкретное состояние или симптомы, на которые направлено лечение. В некоторых вариантах осуществления вирусные векторы можно инъецировать в ткань, представляющую интерес. Для Gene Writing, специфичной для типа клеток, экспрессия Gene Writer и необязательной направляющей нуклеиновой кислоты в некоторых вариантах осуществления может находиться под управлением промотора, специфичного для типа клеток.

В некоторых вариантах осуществления AAV обеспечивает низкую токсичность, например, благодаря способу очистки, не требующему ультрацентрифугирования клеточных частиц, которые могут активировать иммунный ответ. В некоторых вариантах осуществления AAV обеспечивает низкую вероятность приведения к инсерционному мутагенезу, например, потому, что он по сути не интегрируется в геном хозяина.

В некоторых вариантах осуществления AAV имеет предел упаковки, составляющий приблизительно 4,4, 4,5, 4,6, 4,7 или 4,75 т. о. В некоторых вариантах осуществления Gene Writer, промотор и терминатор транскрипции могут входить в один вирусный вектор. SpCas9 (4,1 т. о.) в некоторых случаях может быть сложно упаковать в AAV. Следовательно, в некоторых вариантах осуществления применяется Gene Writer, который короче по длине, чем другие Gene Writer или редакторы оснований. В некоторых вариантах осуществления Gene Writer составляет менее чем приблизительно 4,5 т. о., 4,4 т. о., 4,3 т. о., 4,2 т. о., 4,1 т. о., 4 т. о., 3,9 т. о., 3,8 т. о., 3,7 т. о., 3,6 т. о., 3,5 т. о., 3,4 т. о., 3,3 т. о., 3,2 т. о., 3,1 т. о., 3 т. о., 2,9 т. о., 2,8 т. о., 2,7 т. о., 2,6 т. о., 2,5 т. о., 2 т. о. или 1,5 т. о.

AAV может представлять собой AAV1, AAV2, AAV5 или любую их комбинацию. В некоторых вариантах осуществления тип AAV выбирают с учетом клеток, подлежащих нацеливанию; например, серотипы 1, 2, 5 AAV или гибридный капсид AAV1, AAV2, AAV5, или любая их комбинация могут быть выбраны для нацеливания на клетки головного мозга или нервные клетки; или AAV4 может быть выбран для нацеливания на сердечную ткань. В некоторых вариантах осуществления AAV8 выбирают для доставки в печень. Иллюстративные серотипы AAV, используемые в отношении данных клеток, описаны, например, в Grimm, D. et al, J. Virol. 82: 5887-5911 (2008) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления AAV относится ко всем серотипам, подтипам и встречающимся в природе AAV, а также к рекомбинантным AAV. AAV может применяться для обозначения самого вируса или его производного. В некоторых вариантах осуществления AAV включает AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64R1, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrh10, AAVLK03, AV10, AAV11, AAV 12, rh10 и их гибриды, AAV птиц, AAV крупного рогатого скота, AAV собак, AAV лошадей, AAV приматов, AAV животных, отличных от приматов, и AAV овец. Геномные последовательности различных серотипов AAV, а также последовательности нативных концевых повторов (TR), белков Rep и субъединиц капсида известны из уровня техники. Такие последовательности можно найти в литературе или в общедоступных базах данных,

таких как GenBank. Дополнительные иллюстративные серотипы AAV указаны в таблице 36.

Таблица 36. Иллюстративные серотипы AAV

Ткань-мишень	Средство доставки	Литературный источник
Печень	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2, NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4] Аденовирус (Ad5, HC-AdV[6])	1. Wang *et al.*, *Mol. Ther.* **18**, 118-25 (2010) 2. Ginn *et al.*, *JHEP Reports*, 100065 (2019) 3. Paulk *et al.*, *Mol. Ther.* **26**, 289-303 (2018). 4. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014). 5. L. Wang *et al.*, *Mol. Ther.* **23**, 1877-87 (2015). 6. Hausl *Mol Ther* (2010)
Легкое	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) Аденовирус (Ad5, Ad3, Ad21, Ad14)[3]	1. Duncan *et al.*, *Mol Ther Methods Clin Dev* (2018) 2. Cooney *et al.*, Am J Respir Cell Mol Biol (2019) 3. Li *et al.*, *Mol Ther Methods Clin Dev* (2019)
Кожа	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010) 2. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014).
HSC	Аденовирус (HDAd5/35^{++})	Wang et al. *Blood Adv* (2019)

В некоторых вариантах осуществления фармацевтическая композиция (например, содержащая AAV, описанный в данном документе) содержит менее 10% пустых капсидов, менее 8% пустых капсидов, менее 7% пустых капсидов, менее 5% пустых капсидов, менее 3% пустых капсидов или менее 1% пустых капсидов. В некоторых вариантах осуществления фармацевтическая композиция содержит менее чем приблизительно 5% пустых капсидов. В некоторых вариантах осуществления количество пустых капсидов находится ниже предела выявления. В некоторых вариантах осуществления преимущественно, чтобы фармацевтическая композиция содержала небольшое количество

пустых капсидов, например, поскольку пустые капсиды могут вызывать нежелательный ответ (например, иммунный ответ, воспалительный ответ, печеночный ответ и/или сердечный ответ), например, при незначительном или отсутствующем терапевтическом эффекте.

В некоторых вариантах осуществления содержание остаточного белка клеток-хозяев (rHCP) в фармацевтической композиции меньше или равняется 100 нг/мл rHCP на 1×10^{13} vg/мл, например, меньше или равняется 40 нг/мл rHCP на 1×10^{13} vg/мл или 1-50 нг/мл rHCP на 1×10^{13} vg/мл. В некоторых вариантах осуществления фармацевтическая композиция содержит менее 10 нг rHCP на $1,0\times10^{13}$ vg или менее 5 нг rHCP на $1,0\times10^{13}$ vg, менее 4 нг rHCP на $1,0\times10^{13}$ vg или менее 3 нг rHCP на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации. В некоторых вариантах осуществления содержание остаточной ДНК клеток-хозяев (hcDNA) в фармацевтической композиции меньше или равняется 5×10^{6} пг/мл hcDNA на 1×10^{13} vg/мл, меньше или равняется $1,2\times10^{6}$ пг/мл hcDNA на 1×10^{13} vg/мл или 1×10^{5} пг/мл hcDNA на 1×10^{13} vg/мл. В некоторых вариантах осуществления содержание остаточной ДНК клеток-хозяев в указанной фармацевтической композиции составляет менее $5,0\times10^{5}$ пг на 1×10^{13} vg, менее $2,0\times10^{5}$ пг на $1,0\times10^{13}$ vg, менее $1,1\times10^{5}$ пг на $1,0\times10^{13}$ vg, менее $1,0\times10^{5}$ пг hcDNA на $1,0\times10^{13}$ vg, менее $0,9\times10^{5}$ пг hcDNA на $1,0\times10^{13}$ vg, менее $0,8\times10^{5}$ пг hcDNA на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации.

В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции меньше или равняется $1,7\times10^{5}$ пг/мл на $1,0\times10^{13}$ vg/мл, или 1×10^{5} пг/мл на $1,0\times10^{13}$ vg/мл, или $1,7\times10^{6}$ пг/мл на $1,0\times10^{13}$ vg/мл. В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции составляет менее $10,0\times10^{5}$ пг на $1,0\times10^{13}$ vg, менее $8,0\times10^{5}$ пг на $1,0\times10^{13}$ vg или менее $6,8\times10^{5}$ пг на 1,0 x 10^{13} vg. В вариантах осуществления фармацевтическая композиция содержит менее 0,5 нг на $1,0\times10^{13}$ vg, менее 0,3 нг на $1,0\times10^{13}$ vg, менее 0,22 нг на $1,0\times10^{13}$ vg или менее 0,2 нг на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации бычьего сывороточного альбумина (BSA). В вариантах осуществления содержание бензоназы в фармацевтической композиции составляет менее 0,2 нг на $1,0\times10^{13}$ vg, менее 0,1 нг на $1,0\times10^{13}$ vg, менее 0,09 нг на $1,0\times10^{13}$ vg, менее 0,08 нг на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации. В вариантах осуществления содержание полоксамера 188 в фармацевтической композиции составляет приблизительно 10-150 ppm, приблизительно 15-100 ppm или приблизительно 20-80 ppm. В вариантах осуществления содержание цезия в фармацевтической композиции составляет менее 50 пг/г (ppm), менее 30 пг/г (ppm) или менее 20 пг/г (ppm) или любое промежуточное значение концентрации.

В вариантах осуществления фармацевтическая композиция содержит общее количество примесей, например, как определено с помощью SDS-PAGE, составляющее менее 10%, менее 8%, менее 7%, менее 6%, менее 5%, менее 4%, менее 3%, менее 2% или любую промежуточную процентную долю. В вариантах осуществления общее количество примесей, например, как определено с помощью SDS-PAGE, составляет более 90%, более 92%, более 93%, более 94%, более 95%, более 96%, более 97%, более 98% или любую

промежуточную процентную долю. В вариантах осуществления содержание никакой отдельной неназванной родственной примеси, например, как измерено с помощью SDS-PAGE, не составляет более 5%, более 4%, более 3% или более 2% или любой промежуточной процентной доли. В вариантах осуществления фармацевтическая композиция содержит процентную долю заполненных капсидов по отношению к общему количеству капсидов (например, пик 1+пик 2 согласно измерению с помощью аналитического ультрацентрифугирования), составляющую более 85%, более 86%, более 87%, более 88%, более 89%, более 90%, более 91%, более 91,9%, более 92%, более 93% или любую промежуточную процентную долю. В вариантах осуществления фармацевтической композиции процентная доля заполненных капсидов, измеренная в пике 1 с помощью аналитического ультрацентрифугирования, составляет 20-80%, 25-75%, 30-75%, 35-75% или 37,4-70,3%. В вариантах осуществления фармацевтической композиции процентная доля заполненных капсидов, измеренная в пике 2 с помощью аналитического ультрацентрифугирования, составляет 20-80%, 20-70%, 22-65%, 24-62% или 24,9-60,1%.

В одном варианте осуществления фармацевтическая композиция содержит титр геномов, составляющий от 1,0 до $5,0\times10^{13}$ vg/мл, от 1,2 до $3,0\times10^{13}$ vg/мл или от 1,7 до $2,3\times10^{13}$ vg/мл. В одном варианте осуществления фармацевтическая композиция демонстрирует биологическую нагрузку, составляющую менее 5 КОЕ/мл, менее 4 КОЕ/мл, менее 3 КОЕ/мл, менее 2 КОЕ/мл или менее 1 КОЕ/мл или любое промежуточное значение концентрации. В вариантах осуществления количество эндотоксина в соответствии с USP, например, USP <85> (включенной посредством ссылки во всей своей полноте), составляет менее 1,0 EU/мл, менее 0,8 EU/мл или менее 0,75 EU/мл. В вариантах осуществления осмолярность фармацевтической композиции в соответствии с USP, например, USP <785> (включенной посредством ссылки во всей своей полноте), составляет от 350 до 450 мОсм/кг, от 370 до 440 мОсм/кг или от 390 до 430 мОсм/кг. В вариантах осуществления фармацевтическая композиция содержит менее 1200 частиц размером более 25 мкм на контейнер, менее 1000 частиц размером более 25 мкм на контейнер, менее 500 частиц размером более 25 мкм на контейнер или любое промежуточное значение. В вариантах осуществления фармацевтическая композиция содержит менее 10000 частиц размером более 10 мкм на контейнер, менее 8000 частиц размером более 10 мкм на контейнер или менее 600 частиц размером более 10 пм на контейнер.

В одном варианте осуществления фармацевтическая композиция имеет титр геномов, составляющий от 0,5 до $5,0\times10^{13}$ vg/мл, от 1,0 до $4,0\times10^{13}$ vg/мл, от 1,5 до $3,0\times10^{13}$ vg/мл или от 1,7 до $2,3\times10^{13}$ vg/мл. В одном варианте осуществления фармацевтическая композиция, описанная в данном документе, характеризуется одним или несколькими из следующего: менее чем приблизительно 0,09 нг бензоназы на $1,0\times10^{13}$ vg, менее чем приблизительно 30 пг/г (ppm) цезия, приблизительно 20-80 ppm полоксамера 188, менее чем приблизительно 0,22 нг BSA на $1,0\times10^{13}$ vg, менее чем приблизительно $6,8\times10^{5}$ пг остаточной плазмидной ДНК на $1,0\times10^{13}$ vg, менее чем приблизительно $1,1\times10^{5}$ пг остаточной hcDNA на $1,0\times10^{13}$ vg, менее чем приблизительно 4 нг rHCP на $1,0\times10^{13}$ vg, pH

от 7,7 до 8,3, приблизительно 390-430 мОсм/кг, менее чем приблизительно 600 частиц размером > 25 мкм на контейнер, менее чем приблизительно 6000 частиц размером > 10 мкм на контейнер, титр геномов от приблизительно $1,7\times10^{13}$ до $2,3\times10^{13}$ vg/мл, инфекционный титр от приблизительно $3,9\times10^{8}$ до $8,4\times10^{10}$ МЕ на $1,0\times10^{13}$ vg, содержание общего белка приблизительно 100-300 пг на $1,0\times10^{13}$ vg, средняя выживаемость > 24 дней у мышей A7SMA с дозой вирусного вектора приблизительно $7,5\times10^{13}$ vg/кг, относительная активность приблизительно 70-130% на основании клеточного анализа in vitro и/или менее чем приблизительно 5% пустых капсидов. В различных вариантах осуществления фармацевтические композиции, описанные в данном документе, содержат любую из вирусных частиц, обсуждаемых в данном документе, сохраняют активность в пределах ± 20%, ± 15%, ± 10% или ± 5% от эталонного стандарта. В некоторых вариантах осуществления активность измеряют с использованием подходящего клеточного анализа in vitro или животной модели in vivo.

Дополнительные способы получения, определения характеристик и введения доз частиц AAV описаны в WO 2019094253, которая включена в данный документ посредством ссылки во всей своей полноте.

Дополнительные конструкции rAAV, которые можно использовать согласно с настоящим изобретением, включают конструкции, описанные в Wang et al 2019, доступные по адресу: //doi.org/10.1038/s41573-019-0012-9, включая таблицу 1, которая включена посредством ссылки во всей своей полноте.

18. *Введение AAV*

В некоторых вариантах осуществления аденоассоциированный вирус (AAV) используется в сочетании с системой, нуклеиновой кислотой-матрицей и/или полипептидом, описанными в данном документе. В некоторых вариантах осуществления AAV используется для доставки, введения или упаковки системы, нуклеиновой кислоты-матрицы и/или полипептида, описанных в данном документе. В некоторых вариантах осуществления AAV представляет собой рекомбинантный AAV (rAAV).

В некоторых вариантах осуществления система содержит следующее: (a) полипептид, описанный в данном документе, или нуклеиновую кислоту, кодирующую его, (b) нуклеиновую кислоту-матрицу (например, РНК-матрицу), описанную в данном документе, и (c) одну или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных в отношении ткани-мишени, где одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных в отношении ткани-мишени, находятся в функциональной ассоциации с (a), (b) или (a) и (b), где при ассоциации с (a) (a) предусматривает нуклеиновую кислоту, кодирующую полипептид.

В некоторых вариантах осуществления система, описанная в данном документе, дополнительно содержит первый капсидный белок рекомбинантного аденоассоциированного вируса (rAAV); где по меньшей мере одно из (a) или (b)

ассоциировано с первым капсидным белком rAAV, где по меньшей мере одно из (a) или (b) фланкировано инвертированными концевыми повторами (ITR) AAV.

В некоторых вариантах осуществления (a) и (b) ассоциированы с первым капсидным белком rAAV.

В некоторых вариантах осуществления (a) и (b) находятся в одной нуклеиновой кислоте.

В некоторых вариантах осуществления система дополнительно содержит второй капсидный белок rAAV, где по меньшей мере одно из (a) или (b) ассоциировано со вторым капсидным белком rAAV, и где по меньшей мере одно из (a) или (b), ассоциированного со вторым капсидным белком rAAV, отличается от по меньшей мере одного из (a) или (b), ассоциированного с первым капсидным белком rAAV.

В некоторых вариантах осуществления по меньшей мере одно из (a) или (b) ассоциировано с первым или вторым капсидным белком rAAV, распределено внутри первого или второго капсидного белка rAAV, где первый или второй капсидный белок rAAV находится в форме капсидной частицы AAV.

В некоторых вариантах осуществления система дополнительно содержит наночастицу, где наночастица ассоциирована с по меньшей мере одним из (a) или (b).

В некоторых вариантах осуществления (a) и (b) соответственно ассоциированы с: a) первым капсидным белком rAAV и вторым капсидным белком rAAV; b) наночастицей и первым капсидным белком rAAV; c) первым капсидным белком rAAV; d) первым капсидным белком аденовируса; e) первой наночастицей и второй наночастицей или f) первой наночастицей.

Вирусные векторы применимы для доставки всей системы или ее части, предусмотренной в настоящем изобретении, например, для применения в способах, предусмотренных в настоящем изобретении. Системы, полученные из различных вирусов, использовались для доставки полипептидов, нуклеиновых кислот или транспозонов; например: лентивирус с дефицитом интегразы, аденовирус, аденоассоциированный вирус (AAV), вирус простого герпеса и бакуловирус (описано в работе Hodge et al. Hum Gene Ther 2017; Narayanavari et al. Crit Rev Biochem Mol Biol 2017; Boehme et al. Curr Gene Ther 2015).

Аденовирусы представляют собой распространенные вирусы, которые использовались в качестве средств доставки генов благодаря наличию точных данных по их биологии, генетической стабильности, высокой эффективности трансдукции и простоте крупномасштабного получения (см., например, обзор Lee et al. Genes & Diseases 2017). Они имеют линейные геномы в виде dsDNA и представлены множеством серотипов, различающихся вариантами тканевого и клеточного тропизма. Чтобы предупредить репликацию инфекционного вируса в клетках-реципиентах, в геномах аденовирусов, используемых для упаковки, удалены некоторые или все эндогенные вирусные белки, которые предоставляются в клетках, продуцирующих вирус, в транс-положении. Это делает геномы зависимыми от помощников, что означает, что они могут быть реплицированы и упакованы в вирусные частицы только в присутствии недостающих

компонентов, обеспечиваемых так называемыми вспомогательными функциями. Зависимая от помощников аденовирусная система с удаленными всеми вирусными ORF может быть совместимой с упаковкой чужеродной ДНК размером до ~37 т. о. (Parks et al. J Virol 1997). В некоторых вариантах осуществления аденовирусный вектор используется для доставки ДНК, соответствующей полипептидному или матричному компоненту системы Gene Writing™, или оба они содержатся в отдельных аденовирусных векторах или одном и том же аденовирусном векторе. В некоторых вариантах осуществления аденовирус представляет собой аденовирус, зависимый от помощника (HD-AdV), который не способен к самостоятельной упаковке. В некоторых вариантах осуществления аденовирус представляет собой аденовирус высокой емкости (HC-AdV), в котором удалены все эндогенные вирусные ORF или их значительная часть, при этом сохранены необходимые компоненты последовательности для упаковки в аденовирусные частицы. Для этого типа вектора единственными аденовирусными последовательностями, требующимися для упаковки генома, являются некодирующие последовательности: инвертированные концевые повторы (ITR) на обоих концах и сигнал упаковки на 5'-конце (Jager et al. Nat Protoc 2009). В некоторых вариантах осуществления геном аденовируса также содержит ДНК-заполнитель для обеспечения соответствия минимальному размеру генома для оптимального продуцирования и стабильности (см., например, Hausl et al. Mol Ther 2010). Аденовирусы использовались в данной области техники для доставки транспозонов в различные ткани. В некоторых вариантах осуществления аденовирус используется для доставки системы Gene Writing™ в печень.

В некоторых вариантах осуществления аденовирус используется для доставки системы Gene Writing™ в HSC, например, HDAd5/35++. HDAd5/35++ представляет собой аденовирус с модифицированными волокнами серотипа 35, которые устраняют нацеливание вектора на печень (Wang et al. Blood Adv 2019). В некоторых вариантах осуществления аденовирус, который доставляет систему Gene Writing™ в HSC, использует рецептор, который специфично экспрессируется на примитивных HSC, например, CD46.

Аденоассоциированные вирусы (AAB) принадлежат к семейству Parvoviridae и более конкретно составляют род Dependoparvovirus. Геном AAV состоит из линейной однонитевой молекулы ДНК, которая содержит примерно 4,7 тысяч оснований (т. о.) и состоит из двух основных открытых рамок считывания (ORF), кодирующих неструктурные белки Rep (репликации) и структурные белки Cap (капсидные). Была идентифицирована вторая ORF внутри гена cap, которая кодирует белок, активирующий сборку (AAP). ДНК, фланкирующие кодирующие области AAV, представляют собой две действующие в цис-положении последовательности инвертированных концевых повторов (ITR) длиной примерно 145 нуклеотидов с прерывистыми палиндромными последовательностями, которые могут быть уложены в энергетически стабильные структуры шпилек, которые функционируют в качестве праймеров репликации ДНК. Было показано, что в дополнение к их роли в репликации ДНК последовательности ITR участвуют в интеграции вирусной ДНК в клеточный геном, освобождении из генома хозяина или плазмиды и инкапсуляции

вирусной нуклеиновой кислоты в зрелые вирионы (Muzyczka, (1992) Curr. Top. Micro. Immunol. 158:97-129). В некоторых вариантах осуществления один или несколько компонентов Gene Writing™, представляющих собой нуклеиновую кислоту, фланкированы с помощью ITR, полученных из AAV, для обеспечения вирусной упаковки. См., например, WO2019113310.

В некоторых вариантах осуществления один или несколько компонентов системы Gene Writing™ переносят посредством по меньшей мере одного вектора на основе AAV. В некоторых вариантах осуществления по меньшей мере один вектор на основе AAV выбран, исходя из его тропизма к конкретной клетке, ткани, организму. В некоторых вариантах осуществления вектор на основе AAV является псевдотипированным, например, AAV2/8, где AAV2 описывает дизайн конструкции, но капсидный белок заменен капсидным белком из AAV8. Понятно, что любой из описанных векторов может представлять собой псевдотипированное производное, в котором капсидный белок, используемый для упаковки генома AAV, получен из AAV другого серотипа. В некоторых вариантах осуществления AAV, предназначенный для использования для Gene Writing™, может быть получен методом направленной эволюции для обеспечения тропизма к новым клеткам или тканям, как было продемонстрировано в литературе (например, Davidsson et al. Proc Natl Acad Sci U S A 2019).

В некоторых вариантах осуществления вектор на основе AAV для доставки представляет собой вектор, который содержит два инвертированных концевых повтора (ITR) AAV и нуклеотидную последовательность, представляющую интерес (например, последовательность, кодирующую полипептид Gene Writer™ или ДНК-матрицу или и то, и другое), при этом каждый из указанных ITR содержит прерывистую (или несплошную) палиндромную последовательность, т. е. последовательность, состоящую из трех сегментов: первого сегмента и последнего сегмента, которые идентичны при прочтении 5'→ 3', но гибридизируются при размещении друг против друга, и сегмента, который отличается, который разделяет одинаковые сегменты. Такие последовательности, особенно ITR, образуют шпилечные структуры. См., например, WO2012123430.

Общепринято, что вирионы AAV с капсидами получают путем введения плазмиды или плазмид, кодирующих геном rAAV или scAAV, белки Rep и белки Cap (Grimm et al., 1998). При введении этих плазмид-помощников в транс-положении геном AAV "освобождается" (т. е. высвобождается и затем восстанавливается) из генома хозяина и в дальнейшем инкапсулируется с получением инфекционного AAV. В некоторых вариантах осуществления одну или несколько нуклеиновых кислот Gene Writing™ упаковывают в частицы AAV посредством введения нуклеиновых кислот, фланкированных ITR, в пакующую клетку в сочетании со вспомогательными функциями.

В некоторых вариантах осуществления геном AAV представляет собой так называемый самокомплементарный геном (называемый scAAV), в котором последовательность, расположенная между ITR, содержит как необходимую последовательность нуклеиновой кислоты (например, ДНК, кодирующую полипептид или

матрицу Gene Writer™ или и то, и другое), так и, в дополнение, последовательность, обратно комплементарную необходимой последовательности нуклеиновой кислоты, так что эти два компонента могут располагаться внахлест и гибридизироваться между собой. В некоторых вариантах осуществления самокомплементарные модули разделены промежуточной последовательностью, которая позволяет ДНК сворачиваться складчато, например, образуя структуру стебель-петля. Преимущество scAAV заключается в том, что он готов к транскрипции при попадании в ядро, а не зависит от первоначального праймирования с помощью ITR и синтеза второй нити для образования dsDNA. В некоторых вариантах осуществления один или несколько компонентов Gene Writing™ разработаны в виде scAAV, где последовательность между ITR AAV содержит два обратно комплементарных модуля, которые могут гибридизироваться между собой с образованием dsDNA.

В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид, или матрицу, или и то, и другое), доставляемая в клетки, представляет собой линейный ДНК-дуплекс с замкнутыми концами (CELiD-ДНК или ceDNA). В некоторых вариантах осуществления ceDNA получена из репликативной формы генома AAV (Li et al. PLoS One 2013). В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид, или ДНК-матрицу, или и то, и другое) фланкирована с помощью ITR, например, ITR AAV, где по меньшей мере одна из ITR содержит сайт концевого разрешения и сайт связывания белка репликации (иногда называемый сайтом связывания репликативного белка). В некоторых вариантах осуществления ITR получены из аденоассоциированного вируса, например, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12 или их комбинации. В некоторых вариантах осуществления ITR являются симметричными. В некоторых вариантах осуществления ITR являются асимметричными. В некоторых вариантах осуществления предусмотрен по меньшей мере один белок Rep, обеспечивающий репликацию конструкции. В некоторых вариантах осуществления по меньшей мере один белок Rep получен из аденоассоциированного вируса, например, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12 или их комбинации. В некоторых вариантах осуществления ceDNA получают посредством создания клетки-продуцента с (i) ДНК, фланкированной ITR, например, ITR AAV, и (ii) компонентами, необходимыми для ITR-зависимой репликации, например, белками AAV Rep78 и Rep52 (или нуклеиновой кислотой, кодирующей данные белки). В некоторых вариантах осуществления ceDNA не содержит какого-либо капсидного белка, например, не упакована в инфекционную частицу AAV. В некоторых вариантах осуществления ceDNA составляют в LNP (см., например, WO2019051289A1).

В некоторых вариантах осуществления ceDNA-вектор состоит из двух самокомплементарных последовательностей, например, асимметричных, или симметричных, или по сути симметричных ITR, определенных в данном документе, фланкирующих указанную кассету экспрессии, где ceDNA-вектор не ассоциирован с

капсидным белком. В некоторых вариантах осуществления ceDNA-вектор содержит две самокомплементарные последовательности, обнаруженные в геноме AAV, где по меньшей мере одна ITR содержит действующий Rep-связывающий элемент (RBE) (также иногда называемый в данном документе "RBS") и сайт концевого разрешения (trs) AAV или функциональный вариант RBE. См., например, WO2019113310.

Интеины

В некоторых вариантах осуществления, которые более подробно описаны ниже, интеин-N может быть слит с N-концевой частью первого домена, описанного в данном документе, и интеин-C может быть слит с C-концевой частью второго домена, описанного в данном документе, для присоединения N-концевой части к C-концевой части с соединением таким образом первого и второго доменов. В некоторых вариантах осуществления каждый из первого и второго доменов независимо выбраны из ДНК-связывающего домена, РНК-связывающего домена, домена RT и эндонуклеазного домена.

Используемый в данном документе термин "интеин" относится к самовырезающемуся интрону белка (например, пептиду), например, который лигирует фланкирующие N-концевые и C-концевые экстеины (например, фрагменты, подлежащие соединению). В некоторых случаях интеин может содержать фрагмент белка, который способен вырезать себя и соединять оставшиеся фрагменты (экстеины) пептидной связью в процессе, известном как сплайсинг белка. Интеины также называют "интронами белков". Процесс самовырезания интеина и соединения оставшихся частей белка в данном документе называется "сплайсингом белка" или "интеин-опосредованным сплайсингом белка". В некоторых вариантах осуществления интеин белка-предшественника (белка, содержащего интеин, до интеин-опосредованного сплайсинга белка) происходит из двух генов. Такой интеин упоминается в данном документе как расщепленный интеин (например, расщепленный интеин-N и расщепленный интеин-C). Например, у цианобактерий DnaE, каталитическая a-субъединица ДНК-полимеразы III, кодируется двумя отдельными генами - dnaE-n и dnaE-c. Интеин, кодируемый геном dnaE-n, может называться в данном документе "интеином-N". Интеин, кодируемый геном dnaE-c, может называться в данном документе "интеином-C".

Применение интеинов для соединения гетерологичных фрагментов белка описано, например, в Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (включенном в данный документ посредством ссылки во всей своей полноте). Например, при слиянии с отдельными фрагментами белка интеины IntN и IntC могут распознавать друг друга, самовырезаться и/или одновременно лигировать фланкирующие N- и C-концевые экстеины фрагментов белка, с которыми они были слиты, тем самым восстанавливая полноразмерный белок из двух фрагментов белка.

В некоторых вариантах осуществления применяется синтетический интеин на основе интеина dnaE - пара интеинов Cfa-N (например, расщепленный интеин-N) и Cfa-C (например, расщепленный интеин-C). Примеры таких интеинов были описаны, например, в Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5 (включенном в данный документ

посредством ссылки во всей своей полноте). Неограничивающие примеры пар интеинов, которые можно применять в соответствии с настоящим изобретением, включают интеин Cfa DnaE, интеин Ssp GyrB, интеин Ssp DnaX, интеин Ter DnaE3, интеин Ter ThyX, интеин Rma DnaB и интеин Cne Prp8 (например, описанные в патенте США № 8394604, включенном в данный документ посредством ссылки).

В некоторых вариантах осуществления интеин-N и интеин-C могут быть слиты с N-концевой частью расщепленного Cas9 и C-концевой частью расщепленного Cas9 соответственно для соединения N-концевой части расщепленного Cas9 и C-концевой части расщепленного Cas9. Например, в некоторых вариантах осуществления интеин-N слит с C-концом N-концевой части расщепленного Cas9, т. е. с образованием структуры N-[N-концевая часть расщепленного Cas9]-[интеин-N]~C. В некоторых вариантах осуществления интеин-C слит с N-концом C-концевой части расщепленного Cas9, т. е. с образованием структуры N-[интеин-C]~[C-концевая часть расщепленного Cas9]-C. Механизм интеин-опосредованного сплайсинга белка для соединения белков, с которыми слиты интеины (например, расщепленный Cas9), описан в Shah et al., Chem Sci. 2014; 5(l):446-461, включенном в данный документ посредством ссылки. Способы разработки и применения интеинов известны из уровня техники и описаны, например, в WO2020051561, W02014004336, WO2017132580, US20150344549 и US20180127780, каждый из которых включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления расщепление относится к разделению на два или более фрагмента. В некоторых вариантах осуществления расщепленный белок Cas9 или расщепленный Cas9 содержит белок Cas9, который представлен в виде N-концевого фрагмента и C-концевого фрагмента, кодируемых двумя отдельными нуклеотидными последовательностями. Полипептиды, соответствующие N-концевой части и C-концевой части белка Cas9, могут подвергаться сплайсингу с образованием восстановленного белка Cas9. В вариантах осуществления белок Cas9 разделен на два фрагмента в неупорядоченной области белка, например, как описано в Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014, или как описано в Jiang et al. (2016) Science 351: 867-871 и файле PDB: 5F9R (каждый из которых включен в данный документ посредством ссылки во всей своей полноте). Неупорядоченную область можно определить с помощью одной или нескольких методик определения структуры белка, известных из уровня техники, включая без ограничения рентгеновскую кристаллографию, ЯМР-спектроскопию, электронную микроскопию (например, cryoEM) и/или моделирование белков in silico. В некоторых вариантах осуществления белок разделен на два фрагмента по любому C, T, A или S, например, в области SpCas9 между аминокислотами A292-G364, F445-K483 или E565-T637 или в соответствующих положениях любого другого Cas9, варианта Cas9 (например, nCas9, dCas9) или другого napDNAbp. В некоторых вариантах осуществления белок разделен на два фрагмента по T310, T313, A456, S469 или C574 SpCas9. В некоторых вариантах осуществления процесс разделения белка на два фрагмента называется расщеплением белка.

В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 2-1000 аминокислот (например, 2-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900 или 900-1000 аминокислот). В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 5-500 аминокислот (например, 5-10, 10-50, 50-100, 100-200, 200-300, 300-400 или 400-500 аминокислот). В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 20-200 аминокислот (например, 20-30, 30-40, 40-50, 50-100 или 100-200 аминокислот).

В некоторых вариантах осуществления часть или фрагмент Gene Writer (например, Cas9-R2Tg) слиты с интеином. Нуклеаза может быть слита с N-концом или C-концом интеина. В некоторых вариантах осуществления часть или фрагмент слитого белка слиты с интеином и слиты с капсидным белком AAV. Интеин, нуклеаза и капсидный белок могут быть слиты друг с другом в любой конфигурации (например, нуклеаза-интеин-капсид, интеин-нуклеаза-капсид, капсид-интеин-нуклеаза и т. д.). В некоторых вариантах осуществления N-конец интеина слит с C-концом слитого белка, и C-конец интеина слит с N-концом капсидного белка AAV.

В некоторых вариантах осуществления эндонуклеазный домен (например, никазный домен Cas9) слит с интеином-N, и полипептид, содержащий домен RT, слит с интеином-C.

Иллюстративные нуклеотидные и аминокислотные последовательности интеинов представлены ниже.

ДНК интеина-N DnaE:

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAA
TCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAAT
GGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGT
ATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAAT
TTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGG
ACCTCATGCGAGTTGACAACCTTCCTAAT

Белок интеин-N DnaE:

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEV
FEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

ДНК интеина-C DnaE:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATA
TTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

Интеин-C:

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

ДНК Cfa-N:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTAT
TGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATG
GTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTAT
TTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTC

ATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGA
TCTCAAACAAGTGGATGGATTG CCA

Белок Cfa-N:

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVF
EYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

ДНК Cfa-C:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGG
AAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGG
AGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

Белок Cfa-C:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

Липидные наночастицы

В способах и системах, предусмотренных в настоящем изобретении, может использоваться любой пригодный носитель или способ доставки, в том числе, в некоторых вариантах осуществления, липидные наночастицы (LNP). Липидные наночастицы в некоторых вариантах осуществления содержат один или несколько ионных липидов, таких как некатионные липиды (например, нейтральные, или анионные, или цвиттерионные липиды); один или несколько конъюгированных липидов (таких как липиды, конъюгированные с PEG, или липиды, конъюгированные с полимерами, описанные в таблице 5 документа WO 2019217941; включенного в данный документ посредством ссылки во всей своей полноте); один или несколько стеринов (например, холестерин) и необязательно одну или несколько нацеливающих молекул (например, конъюгированные рецепторы, лиганды рецепторов, антитела) или комбинации вышеперечисленного.

Липиды, которые можно применять для образования наночастиц (например, липидных наночастиц), включают, например, липиды, описанные в таблице 4 документа WO2019217941, который включен посредством ссылки, например, липидосодержащая наночастица может содержать один или несколько липидов из таблицы 4 документа WO2019217941. Липидные наночастицы могут содержать дополнительные элементы, такие как полимеры, такие как полимеры, описанные в таблице 5 документа WO2019217941, включенного посредством ссылки.

В некоторых вариантах осуществления конъюгированные липиды в случае их присутствия могут включать одно или несколько из PEG-диацилглицерина (DAG) (такого как 1-(монометоксиполиэтиленгликоль)-2,3-димиристоилглицерин (PEG-DMG)), PEG-диалкилоксипропила (DAA), PEG-фосфолипида, PEG-церамида (Cer), пегилированного фосфатидилэтаноламина (PEG-PE), PEG-сукцинатдиацилглицерина (PEG-DAG) (такого как 4-0-(2',3'-ди(тетрадеканоилокси)пропил-1-0-(w-метокси(полиэтокси)этил)бутандиоат (PEG-S-DMG)), PEG-диалкоксипропилкарбама, натриевой соли N-(карбонилметоксиполиэтиленгликоль-2000)-1,2-дистеароил-sn-глицеро-3-фосфоэтаноламина и соединений, описанных в таблице 2 документа WO 2019051289 (включенного посредством ссылки), а также комбинаций вышеперечисленного.

В некоторых вариантах осуществления стерины, которые могут быть включены в липидные наночастицы, включают одно или несколько из холестерина или производных холестерина, например, таких как в WO2009/127060 или US2010/0130588, которые включены посредством ссылки. Дополнительные иллюстративные стерины включают фитостерины, в том числе описанные в Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386, включенном в данный документ посредством ссылки.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид, некатионный липид, конъюгированный липид, который ингибирует агрегацию частиц, и стерин. Количества данных компонентов могут варьироваться независимо и для достижения желаемых свойств. Например, в некоторых вариантах осуществления липидная наночастица содержит ионизируемый липид в количестве от приблизительно 20 мол. % до приблизительно 90 мол. % от общего количества липидов (в других вариантах осуществления это может составлять 20-70% (мол.), 30-60% (мол.) или 40-50% (мол.); от приблизительно 50 мол. % до приблизительно 90 мол. % от общего количества липидов, присутствующих в липидной наночастице), некатионный липид в количестве от приблизительно 5 мол. % до приблизительно 30 мол. % от общего количества липидов, конъюгированный липид в количестве от приблизительно 0,5 мол. % до приблизительно 20 мол. % от общего количества липидов и стерин в количестве от приблизительно 20 мол. % до приблизительно 50 мол. % от общего количества липидов. Соотношение общего количества липидов и нуклеиновой кислоты (например, кодирующей Gene Writer или нуклеиновой кислоты-матрицы) можно варьировать по желанию. Например, соотношение общего количества липидов и нуклеиновой кислоты (по массе или весу) может составлять от приблизительно 10: 1 до приблизительно 30: 1.

В некоторых вариантах осуществления ионизируемый липид может представлять собой катионный липид, ионизируемый катионный липид, например, катионный липид, который может существовать в положительно заряженной или нейтральной форме в зависимости от pH, или аминосодержащий липид, который может быть легко протонирован. В некоторых вариантах осуществления катионный липид представляет собой липид, способный быть положительно заряженным, например, в физиологических условиях. Иллюстративные катионные липиды содержат одну или несколько аминогрупп, которые несут положительный заряд. В некоторых вариантах осуществления липидная частица содержит катионный липид в составе с одним или несколькими нейтральными липидами, ионизируемыми аминосодержащими липидами, биоразлагаемыми алкиновыми липидами, стероидами, фосфолипидами, содержащими полиненасыщенные липиды, структурными липидами (например, стеринами), PEG, холестерином и липидами, конъюгированными с полимером. В некоторых вариантах осуществления катионный липид может представлять собой ионизируемый катионный липид. Иллюстративный катионный липид, раскрытый в данном документе, может характеризоваться эффективным значением pKa, составляющим более 6,0. В вариантах осуществления липидная наночастица может

содержать второй катионный липид, характеризующийся другим эффективным значением pKa (например, более высоким, чем первое эффективное значение pKa) по сравнению с первым катионным липидом. Липидная наночастица может содержать от 40 до 60 молярных процентов катионного липида, нейтрального липида, стероида, липида, конъюгированного с полимером, и терапевтического средства, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), инкапсулированной внутри липидной наночастицы или ассоциированной с ней. В некоторых вариантах осуществления нуклеиновая кислота составлена совместно с катионным липидом. Нуклеиновая кислота может быть адсорбирована на поверхности LNP, например, LNP, содержащей катионный липид. В некоторых вариантах осуществления нуклеиновая кислота может быть инкапсулирована в LNP, например, в LNP, содержащую катионный липид. В некоторых вариантах осуществления липидная наночастица может содержать нацеливающий фрагмент, например, покрытый нацеливающим средством. В вариантах осуществления состав на основе LNP является биоразлагаемым. В некоторых вариантах осуществления липидная наночастица, содержащая один или несколько липидов, описанных в данном документе, например, формулы (i), (ii), (ii), (vii) и/или (ix), инкапсулирует по меньшей мере 1%, по меньшей мере 5%, по меньшей мере 10%, по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 92%, по меньшей мере 95%, по меньшей мере 97%, по меньшей мере 98% или 100% молекулы РНК, например, РНК-матрицы и/или мРНК, кодирующей полипептид Gene Writer.

В некоторых вариантах осуществления соотношение липидов и нуклеиновой кислоты (соотношение масса/масса; соотношение вес/вес) может находиться в диапазоне от приблизительно 1: 1 до приблизительно 25: 1, от приблизительно 10: 1 до приблизительно 14: 1, от приблизительно 3: 1 до приблизительно 15: 1, от приблизительно 4: 1 до приблизительно 10: 1, от приблизительно 5: 1 до приблизительно 9: 1 или от приблизительно 6: 1 до приблизительно 9: 1. Количества липидов и нуклеиновой кислоты можно регулировать для обеспечения желаемого соотношения N/P, например, соотношения N/P, составляющего 3, 4, 5, 6, 7, 8, 9, 10 или выше. В целом, общее содержание липидов в составе на основе липидных наночастиц может находиться в диапазоне от приблизительно 5 мг/мл до приблизительно 30 мг/мл.

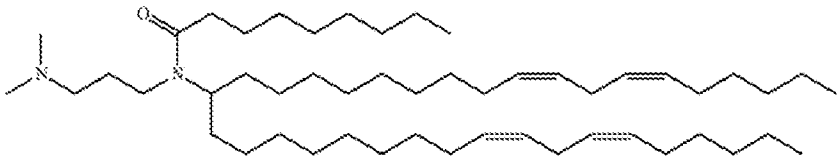
Иллюстративные ионизируемые липиды, которые можно применять в составах на основе липидных наночастиц, включают без ограничения липиды, перечисленные в таблице 1 документа WO2019051289, включенного в данный документ посредством ссылки. Дополнительные иллюстративные липиды включают без ограничения одну или несколько из следующих формул: X из US2016/0311759; I из US20150376115 или из US2016/0376224; I, II или III из US20160151284; I, IA, II или IIA из US20170210967; I-c из US20150140070; A из US2013/0178541; I из US2013/0303587 или US2013/0123338; I из

US2015/0141678; II, III, IV или V из US2015/0239926; I из US2017/0119904; I или II из WO2017/117528; A из US2012/0149894; A из US2015/0057373; A из WO2013/116126; A из US2013/0090372; A из US2013/0274523; A из US2013/0274504; A из US2013/0053572; A из W02013/016058; A из W02012/162210; I из US2008/042973; I, II, III или IV из US2012/01287670; I или II из US2014/0200257; I, II или III из US2015/0203446; I или III из US2015/0005363; I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID или III-XXIV из US2014/0308304; из US2013/0338210; I, II, III или IV из W02009/132131; A из US2012/01011478; I или XXXV из US2012/0027796; XIV или XVII из US2012/0058144; из US2013/0323269; I из US2011/0117125; I, II или III из US2011/0256175; I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII из US2012/0202871; I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV или XVI из US2011/0076335; I или II из US2006/008378; I из US2013/0123338; I или X-A-Y-Z из US2015/0064242; XVI, XVII или XVIII из US2013/0022649; I, II или III из US2013/0116307; I, II или III из US2013/0116307; I или II из US2010/0062967; I-X из US2013/0189351; I из US2014/0039032; V из US2018/0028664; I из US2016/0317458; I из US2013/0195920; 5, 6 или 10 из US10,221,127; III-3 из WO2018/081480; I-5 или I-8 из WO2020/081938; 18 или 25 из US9867888; A из US2019/0136231; II из WO2020/219876; 1 из US2012/0027803; OF-02 из US2019/0240349; 23 из US10,086,013; cKK-E12/A6 из Miao et al (2020); C12-200 из WO2010/053572; 7C1 из Dahlman et al (2017); 304-O13 или 503-O13 из Whitehead et al; TS-P4C2 из US9708628; I из WO2020/106946; I из WO2020/106946.

В некоторых вариантах осуществления ионизируемый липид представляет собой MC3 (6Z,9Z,28Z,31Z)-гептатриаконта-6,9,28,31-тетраен-19-ил-4-(диметиламино)бутаноат (DLin-MC3-DMA или MC3), например, описанный в примере 9 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой липид ATX-002, например, описанный в примере 10 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой (13Z,16Z)-A, A-диметил-3-нонилдокоза-13-16-диен-1-амин (соединение 32), например, описанное в примере 11 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой соединение 6 или соединение 22, например, описанные в примере 12 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой гептадекан-9-ил-8-((2-гидроксиэтил)(6-оксо-6-(ундецилокси)гексил)амино)октаноат (SM-102); например, описанный в примере 1 документа US9867888 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой (9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат (LP01), например, синтезированный в примере 13 документа WO 2015/095340 (включенного в

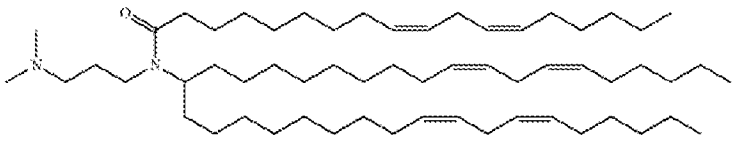
данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой ди((Z)-нон-2-ен-1-ил)-9-(((4-диметиламино)бутаноил)окси)гептадекандиоат (L319), например, синтезированный в примерах 7, 8 или 9 документа US2012/0027803 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой 1,1'-((2-(4-(2-((2-(бис(2-гидроксидодецил)амино)этил)(2-гидроксидодецил)амино)этил)пиперазин-1-ил)этил)азандиил)бис(додекан-2-ол) (C12-200), например, синтезированный в примерах 14 и 16 документа WO2010/053572 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой липид на основе сложного эфира холестерина, содержащего имидазол (ICE)-(3S,10R,13R,17R)-10,13-диметил-17-((R)-6-метилгептан-2-ил)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-тетрадекагидро-1H-циклопента[a]фенантрен-3-ил-3-(1H-имидазол-4-ил)пропаноат, например, структуру (I) из документа WO2020/106946 (включенного в данный документ посредством ссылки во всей своей полноте).

Некоторые неограничивающие примеры липидных соединений, которые можно применять (например, в комбинации с другими липидными компонентами) для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), включают



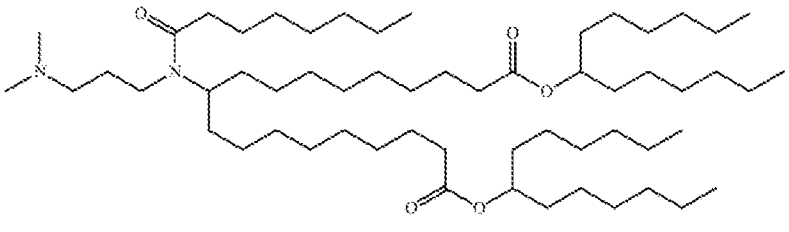
(i)

В некоторых вариантах осуществления LNP, содержащую формулу (i), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



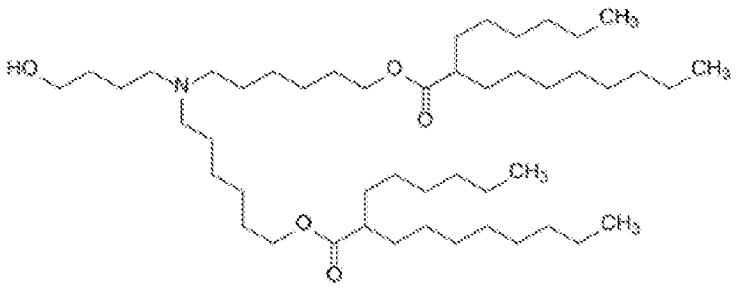
(ii)

В некоторых вариантах осуществления LNP, содержащую формулу (ii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

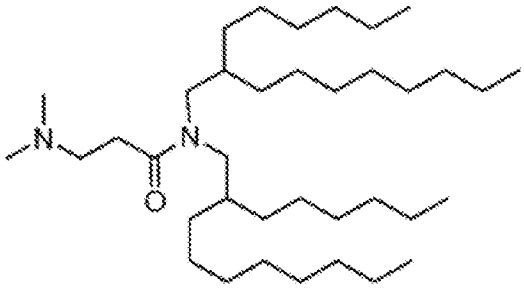


(iii)

В некоторых вариантах осуществления LNP, содержащую формулу (iii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

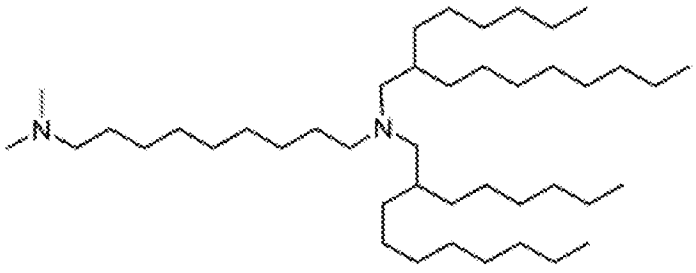


(iv)



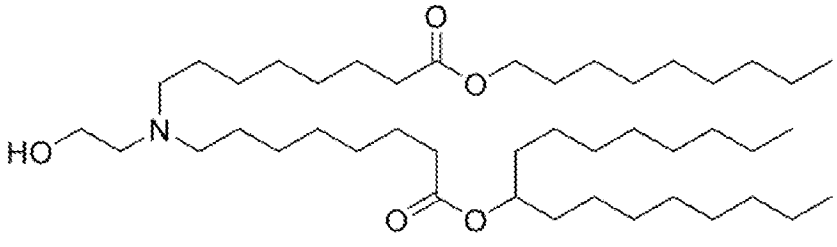
(v)

В некоторых вариантах осуществления LNP, содержащую формулу (v), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

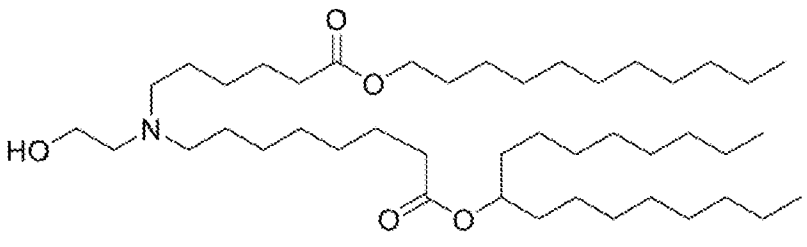


(vi)

В некоторых вариантах осуществления LNP, содержащую формулу (vi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

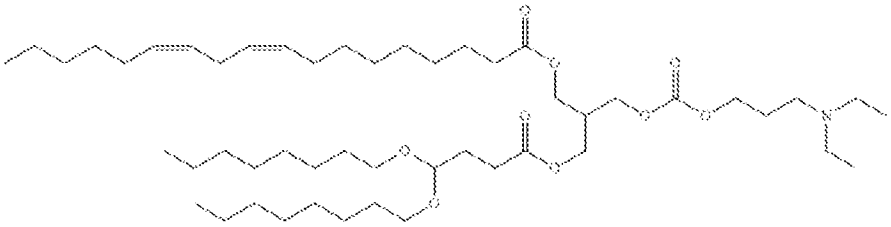


(vii)



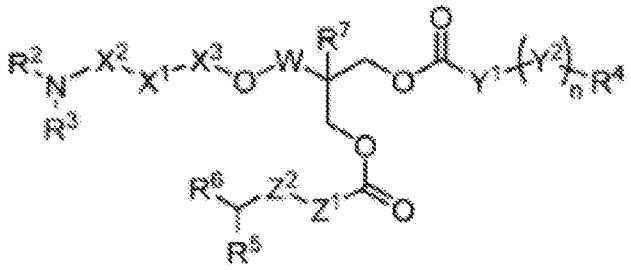
(viii)

В некоторых вариантах осуществления LNP, содержащую формулу (viii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



(ix)

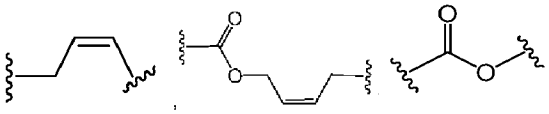
В некоторых вариантах осуществления LNP, содержащую формулу (ix), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



(x),

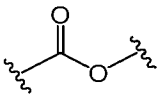
где

X^1 представляет собой O, NR^1 или прямую связь, X^2 представляет собой C2-5-алкилен, X^3 представляет собой C(=0) или прямую связь, R^1 представляет собой H или Me, R^3 представляет собой C-3-алкил, R^2 представляет собой C-3-алкил, или R^2, взятый вместе с атомом азота, к которому он присоединен, и 1-3 атома углерода из X^2 образуют 4-, 5- или 6-членное кольцо, или X^1 представляет собой NR^1, R^1 и R^2, взятые вместе с атомами азота, к которому они присоединены, образуют 5- или 6-членное кольцо, или R^2, взятый вместе с R^3 и атомом азота, к которому они присоединены, образует 5-, 6- или 7-членное кольцо, Y^1 представляет собой C2-12-алкилен, Y^2 выбран из



(в любой ориентации), (в любой ориентации), (в любой ориентации),

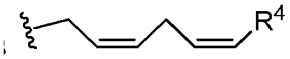
n равняется 0-3, R^4 представляет собой C1-15-алкил, Z^1 представляет собой C1-6-алкилен или прямую связь,

Z^2 is 

(в любой ориентации) или отсутствует, при условии, что если Z^1 представляет собой прямую связь, то Z^2 отсутствует;

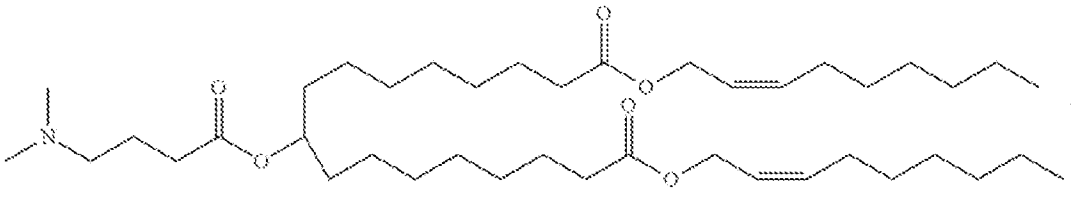
R^5 представляет собой C5-9-алкил или C6-10-алкокси, R^6 представляет собой C5-9-алкил или C6-10-алкокси, W представляет собой метилен или прямую связь, и R^7 представляет собой H или Me или их соль, при условии, что если R^3 и R^2 представляют

собой C2-алкилы, X^1 представляет собой O, X^2 представляет собой линейный C3-алкилен, X^3 представляет собой C(=0), Y^1 представляет собой линейный C1-алкилен, $(Y^2)n-R^4$ представляет собой



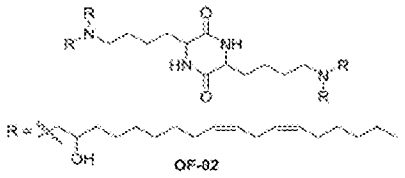
, R^4 представляет собой линейный C5-алкил, Z^1 представляет собой C2-алкилен, Z^2 отсутствует, W представляет собой метилен, и R^7 представляет собой H, то R^5 и R^6 не представляют собой Cx-алкокси.

В некоторых вариантах осуществления LNP, содержащую формулу (xii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

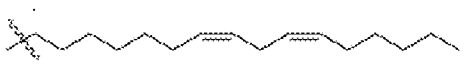


(xi)

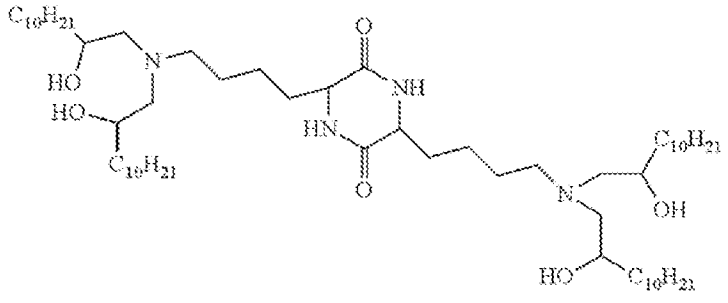
В некоторых вариантах осуществления LNP, содержащую формулу (xi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



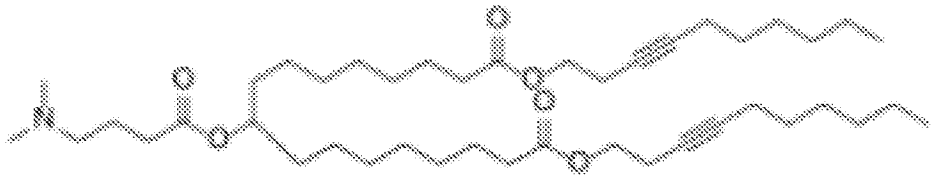
где R=



(xii)

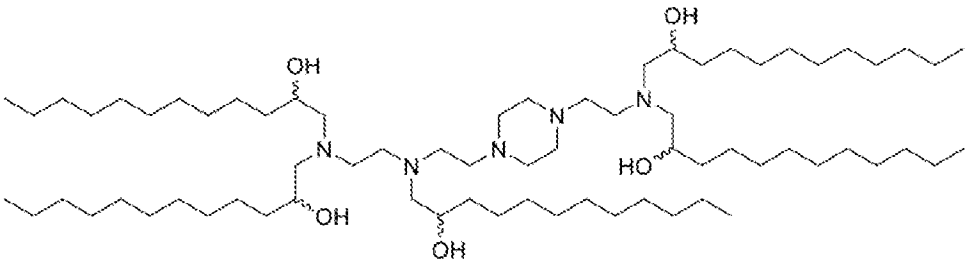


(xiii)



(xiv)

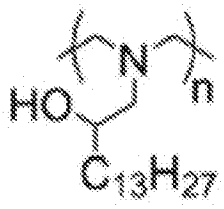
В некоторых вариантах осуществления LNP содержит соединение формулы (xiii) и соединение формулы (xiv).



(xv)

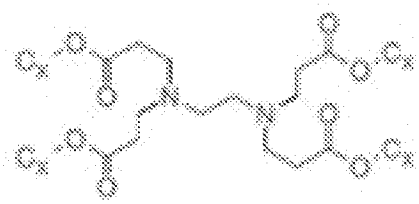
В некоторых вариантах осуществления LNP, содержащую формулу (xv), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

Сердцевинный PEI_{600}

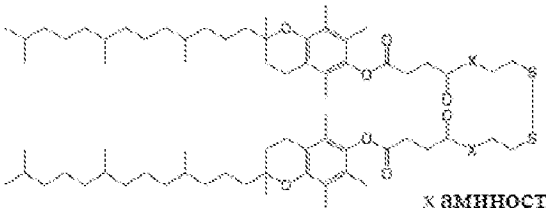


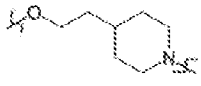
(xvi)

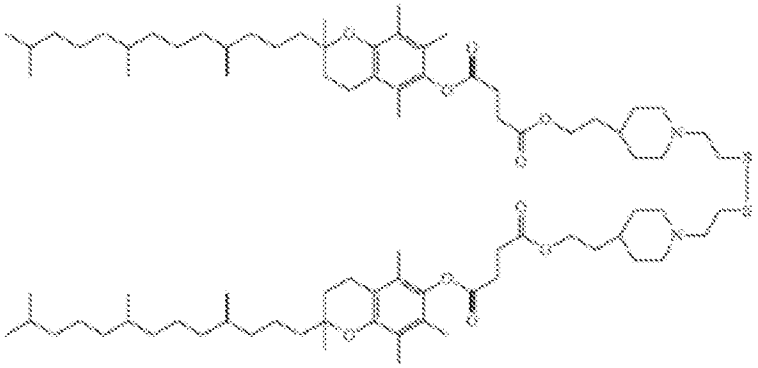
В некоторых вариантах осуществления LNP, содержащую состав формулы (xvi), применяют для доставки композиции Gene Writer, описанной в данном документе, в эндотелиальные клетки легкого.



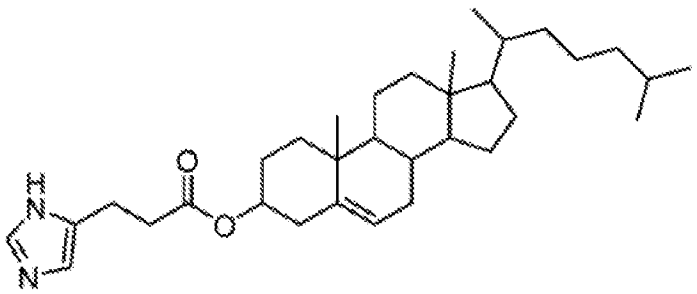
(xvii)



х аминоструктура где X=  (xviii) (a)

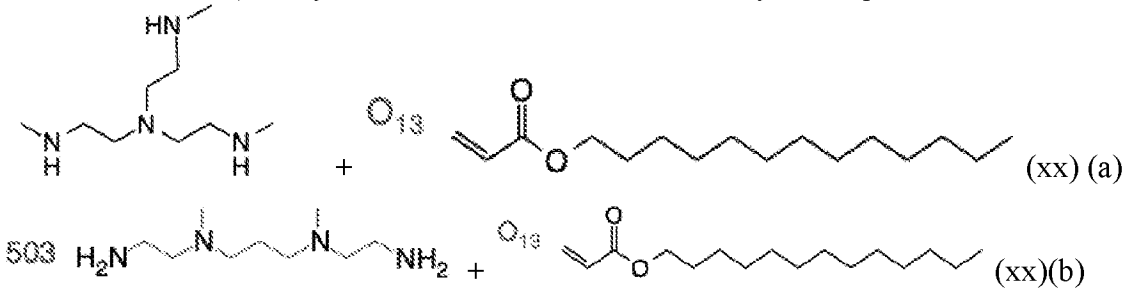


(xviii) (b)



(xix)

В некоторых вариантах осуществления липидное соединение, применяемое для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), получают с помощью одной из следующих реакций:

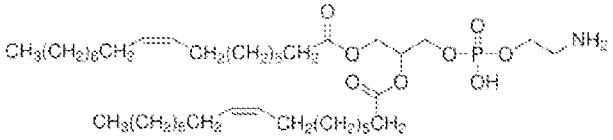


(xx) (a)

(xx)(b)

Иллюстративные некатионные липиды включают без ограничения дистеароил-sn-глицерофосфоэтаноламин, дистеароилфосфатидилхолин (DSPC), диолеоилфосфатидилхолин (DOPC), дипальмитоилфосфатидилхолин (DPPC), диолеоилфосфатидилглицерин (DOPG), 1,2-диолеоил-sn-глицеро-3-фосфоэтаноламин (DOPE), дипальмитоилфосфатидилглицерин (DPPG), диолеоилфосфатидилэтаноламин (DOPE), пальмитоилолеоилфосфатидилхолин (POPC), пальмитоилолеоилфосфатидилэтаноламин (POPE), диолеоилфосфатидилэтаноламин-4-(N-малеимидометил)-циклогексан-1-карбоксилат (DOPE-Mal), дипальмитоилфосфатидилэтаноламин (DPPE), димиристоилфосфоэтаноламин (DMPE), дистеароилфосфатидилэтаноламин (DSPE), монометилфосфатидилэтаноламин (такой как 16-O-монометил-PE), диметилфосфатидилэтаноламин (такой как 16-O-диметил-PE), 18-1-транс-PE, 1-стеароил-2-олеоилфосфатидилэтаноламин (SOPE), гидрогенизированный соевый фосфатидилхолин (HSPC), яичный фосфатидилхолин (EPC), диолеоилфосфатидилсерин (DOPS), сфингомиелин (SM), димиристоилфосфатидилхолин (DMPC), димиристоилфосфатидилглицерин (DMPG), дистеароилфосфатидилглицерин (DSPG), диэрукоилфосфатидилхолин (DEPC), пальмитоилолеоилфосфатидилглицерин (POPG), диэлаидоилфосфатидилэтаноламин (DEPE), лецитин, фосфатидилэтаноламин, лизолецитин, лизофосфатидилэтаноламин, фосфатидилсерин, фосфатидилинозитол, сфингомиелин, яичный сфингомиелин (ESM), цефалин, кардиолипин, фосфатидную кислоту, цереброзиды, дицетилфосфат, лизофосфатидилхолин, дилинолеоилфосфатидилхолин или их смеси. Понятно, что можно также применять другие диацилфосфатидилхолиновые и диацилфосфатидилэтаноламиновые фосфолипиды.

Ацильные группы в данных липидах предпочтительно представляют собой ацильные группы, полученные из жирных кислот, имеющих углеродные цепи C10-C24, например, лауроил, миристоил, пальмитоил, стеароил или олеоил. Дополнительные иллюстративные липиды в некоторых вариантах осуществления включают без ограничения липиды, описанные в Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386, включенном в данный документ посредством ссылки. Такие липиды включают в некоторых вариантах осуществления растительные липиды, которые, как было обнаружено, улучшают трансфекцию мРНК в печени (например, DGTS). В некоторых вариантах осуществления некатионный липид может иметь следующую структуру:



(xxi)

Другие примеры некатионных липидов, пригодных для применения в липидных наночастицах, включают без ограничения липиды, не содержащие фосфор, такие как, например, стеариламин, додециламин, гексадециламин, ацетилпальмитат, глицеринрицинолеат, гексадецилстеарат, изопропилмиристат, амфотерные акриловые полимеры, лаурилсульфат триэтаноламина, алкиларилсульфат, полиэтилоксилированные амиды жирных кислот, бромид диоктадецилдиметиламмония, церамид, сфингомиелин и т. п. Другие некатионные липиды описаны в WO2017/099823 или в публикации заявки на патент США US2018/0028664, содержание которых включено в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления некатионный липид представляет собой олеиновую кислоту или соединение формулы I, II или IV из US2018/0028664, включенного в данный документ посредством ссылки во всей своей полноте. Некатионный липид может составлять, например, 0-30% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления содержание некатионных липидов составляет 5-20% (мол.) или 10-15% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В вариантах осуществления молярное соотношение ионизируемых липидов и нейтральных липидов находится в диапазоне от приблизительно 2:1 до приблизительно 8:1 (например, составляет приблизительно 2:1, 3:1, 4:1, 5:1, 6:1, 7:1 или 8:1).

В некоторых вариантах осуществления липидные наночастицы не содержат каких-либо фосфолипидов.

В некоторых аспектах липидная наночастица может дополнительно содержать компонент, такой как стерин, для обеспечения целостности мембраны. Одним иллюстративным стерином, который можно применять в липидной наночастице, является холестерин и его производные. Неограничивающие примеры производных холестерина включают полярные аналоги, такие как 5α-холестанол, 53-копростанол, холестерил-(2'-гидрокси)-этиловый эфир, холестерил-(4'-гидрокси)-бутиловый эфир и 6-кетохолестанол;

неполярные аналоги, такие как 5α-холестан, холестенон, 5α-холестанон, 5β-холестанон и холестерилдеканоат; и их смеси. В некоторых вариантах осуществления производное холестерина представляет собой полярный аналог, например, холестерил-(4'-гидрокси)бутиловый эфир. Иллюстративные производные холестерина описаны в публикации согласно PCT WO2009/127060 и публикации заявки на патент США US2010/0130588, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

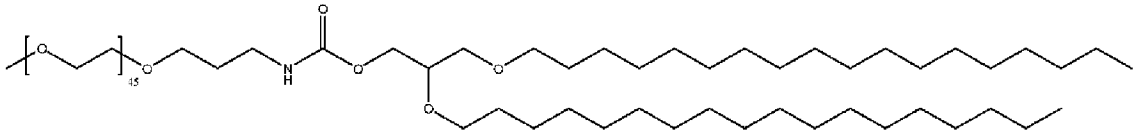
В некоторых вариантах осуществления компонент, обеспечивающий целостность мембраны, такой как стерин, может составлять 0-50% (мол.) (например, 0-10%, 10-20%, 20-30%, 30-40% или 40-50%) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления такой компонент составляет 20-50% (мол.) или 30-40% (мол.) от общего количества липидов в липидной наночастице.

В некоторых вариантах осуществления липидная наночастица может содержать полиэтиленгликоль (PEG) или конъюгированную молекулу липида. Обычно они применяются для ингибирования агрегации липидных наночастиц и/или обеспечения стерической стабилизации. Иллюстративные конъюгированные липиды включают без ограничения конъюгаты PEG-липид, конъюгаты полиоксазолин (POZ)-липид, конъюгаты полиамид-липид (такие как конъюгаты ATTA-липид), конъюгаты катионный полимер-липида (CPL) и их смеси. В некоторых вариантах осуществления молекула конъюгированного липида представляет собой конъюгат PEG-липид, например, липид, конъюгированный с (метоксиполиэтиленгликолем).

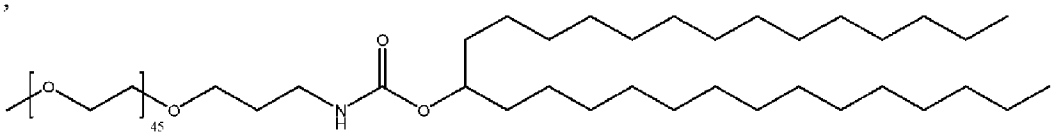
Иллюстративные конъюгаты PEG-липид включают без ограничения PEG-диацилглицерин (DAG) (такой как 1-(монометоксиполиэтиленгликоль)-2,3-димиристоилглицерин (PEG-DMG)), PEG-диалкилоксипропил (DAA), PEG-фосфолипид, PEG-церамид (Cer), пегилированный фосфатидилэтаноламин (PEG-PE), 1,2-димиристоил-sn-глицерин, метоксиполиэтиленгликоль (DMG-PEG-2K), PEG-сукцинатдиацилглицерин (PEGS-DAG) (такой как 4-0-(2',3'-ди(тетрадеканоилокси)пропил-1-0-(w-метокси(полиэтокси)этил)бутандиоат (PEG-S-DMG)), PEG-диалкоксипропилкарбам, натриевую соль N-(карбонил-метоксиполиэтиленгликоль 2000)-1,2-дистеароил-sn-глицеро-3-фосфоэтаноламина или их смесь. Дополнительные иллюстративные конъюгаты PEG-липид описаны, например, в US5885613, US6287591,

US2003/0077829, US2003/0077829, US2005/0175682, US2008/0020058, US2011/0117125, US2010/0130588, US2016/0376224, US2017/0119904 и US/099823, содержание каждого из которых включено в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления PEG-липид представляет собой соединение формулы III, III-a-I, III-a-2, III-b-1, III-b-2 или V из US2018/0028664, содержание которого включено в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления PEG-липид представляет собой соединение формулы II из US20150376115 или US2016/0376224, содержание которых включено в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах

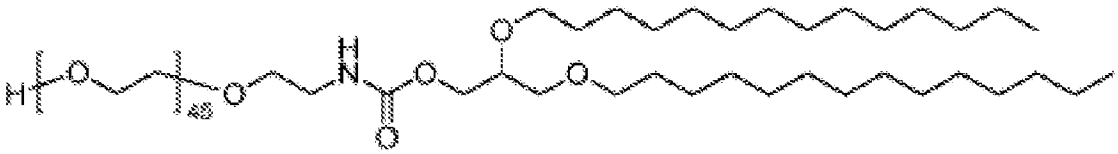
осуществления конъюгат PEG-DAA может представлять собой, например, PEG-дилаурилоксипропил, PEG-димиристилоксипропил, PEG-дипальмитилоксипропил или PEG-дистеарилоксипропил. PEG-липид может представлять собой один или несколько из PEG-DMG, PEG-дилаурилглицерина, PEG-дипальмитоилглицерина, PEG-дистерилглицерина, PEG-дилаурилгликамида, PEG-димиристилгликамида, PEG-дипальмитоилгликамида, PEG-дистерилгликамида, PEG-холестерина (1-[8'-(холест-5-ен-3[бета]-окси)карбоксамидо-3',6'-диоксаоктанил]карбамоил-[омега]-метилполи(этиленгликоля), PEG-DMB (3,4-дитетрадекоксилбензил-[омега]-метилполи(этиленгликолевого) эфира) и 1,2-димиристоил-sn-глицеро-3-фосфоэтаноламин-N-[метокси(полиэтиленгликоля)-2000]. В некоторых вариантах осуществления PEG-липид включает PEG-DMG, 1,2-димиристоил-sn-глицеро-3-фосфоэтаноламин-N-[метокси(полиэтиленгликоль)-2000]. В некоторых вариантах осуществления PEG-липид имеет структуру, выбранную из:



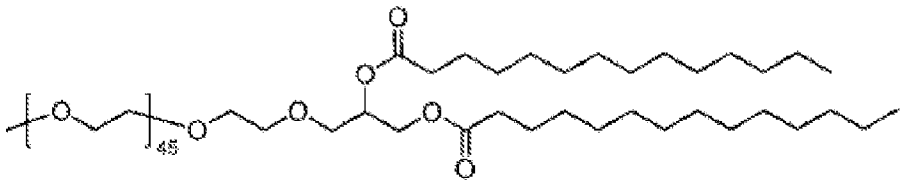
,



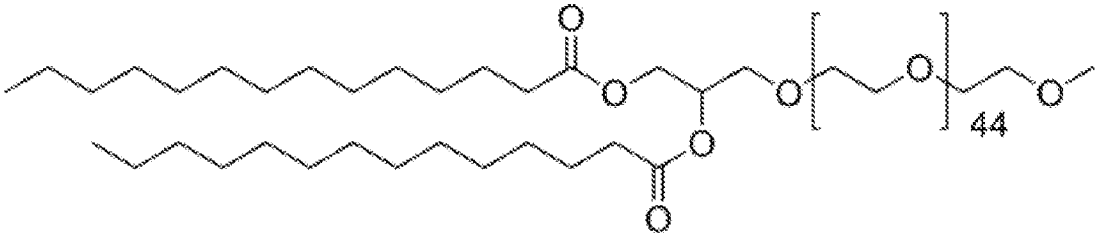
xxii),



(xxiii),



(xxiv) и



(xxv).

В некоторых вариантах осуществления вместо PEG-липида также можно использовать липиды, конъюгированные с молекулой, отличной от PEG. Например, вместо PEG-липида или в дополнение к нему можно использовать конъюгаты полиоксазолин (POZ)-липид, конъюгаты полиамид-липид (такие как конъюгаты ATTA-липид) и конъюгаты катионный полимер-липид (GPL).

Иллюстративные конъюгированные липиды, т. е. конъюгаты PEG-липид, конъюгаты (POZ)-липид, конъюгаты ATTA-липид и катионный полимер-липид, описаны в патентных заявках согласно PCT и LIS, указанных в таблице 2 документа WO2019051289A9 и в документе WO2020106946A1, содержание каждого из которых включено в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления LNP включает соединение формулы (xix), соединение формулы (xxi) и соединение формулы (xxv). В некоторых вариантах осуществления LNP, содержащую состав формулы (xix), формулы (xxi) и формулы (xxv), применяют для доставки композиции Gene Writer, описанной в данном документе, в легкое или клетки легкого.

В некоторых вариантах осуществления PEG или конъюгированный липид могут составлять 0-20% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления содержание PEG или конъюгированного липида составляет 0,5-10% или 2-5% (мол.) от общего количества липидов, присутствующих в липидной наночастице. Молярные соотношения ионизируемого липида, некатионного липида, стерина и PEG-конъюгированного липида можно изменять по мере необходимости. Например, липидная частица может содержать 30-70% ионизируемого липида по молям или от общего веса композиции, 0-60% холестерина по молям или от общего веса композиции, 0-30% некатионного липида по молям или от общего веса композиции и 1-10% конъюгированного липида по молям или от общего веса композиции. Предпочтительно композиция содержит 30-40% ионизируемого липида по молям или от общего веса композиции, 40-50% холестерина по молям или от общего веса композиции и 10-20% некатионного липида по молям или от общего веса композиции. В некоторых других вариантах осуществления композиция содержит 50-75% ионизируемого липида по молям или от общего веса композиции, 20-40% холестерина по молям или от общего веса композиции, 5-10% некатионного липида по молям или от общего веса композиции и 1-10% конъюгированного липида по молям или от общего веса композиции. Композиция может содержать 60-70% ионизируемого липида по молям или от общего веса композиции, 25-35% холестерина по молям или от общего веса композиции и 5-10% некатионного липида по молям или от общего веса композиции. Композиция может также содержать до 90% ионизируемого липида по молям или от общего веса композиции и 2-15% некатионного липида по молям или от общего веса композиции. Состав также может представлять собой состав на основе липидных наночастиц, например, содержащий 8-30% ионизируемого липида по молям или от общего веса композиции, 5-30% некатионного липида по молям или от общего веса композиции и 0-20% холестерина по молям или от общего веса композиции; 4-25% ионизируемого липида по молям или от общего веса композиции, 4-25% некатионного липида по молям или от общего веса композиции, 2-25% холестерина по молям или от общего веса композиции, 10-35% конъюгированного липида по молям или от общего веса композиции и 5% холестерина по молям или от общего веса композиции; или 2-30% ионизируемого липида по молям или от

общего веса композиции, 2-30% некатионного липида по молям или от общего веса композиции, 1-15% холестерина по молям или от общего веса композиции, 2-35% конъюгированного липида по молям или от общего веса композиции и 1-20% холестерина по молям или от общего веса композиции; или даже до 90% ионизируемого липида по молям или от общего веса композиции и 2-10% некатионных липидов по молям или от общего веса композиции, или даже 100% катионного липида по молям или от общего веса композиции. В некоторых вариантах осуществления состав на основе липидных частиц содержит ионизируемый липид, фосфолипид, холестерин и пегилированный липид в молярном соотношении 50: 10:38,5: 1,5. В некоторых других вариантах осуществления состав на основе липидных частиц содержит ионизируемый липид, холестерин и пегилированный липид в молярном соотношении 60:38,5: 1,5.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид, некатионный липид (например, фосфолипид), стерин (например, холестерин) и пегилированный липид, где молярное соотношение липидов находится в диапазоне от 20 до 70 молярных процентов для ионизируемого липида при целевом значении 40-60, молярный процент для некатионного липида находится в диапазоне от 0 до 30 при целевом значении от 0 до 15, молярный процент для стерина находится в диапазоне от 20 до 70 при целевом значении от 30 до 50, и молярный процент для пегилированного липида находится в диапазоне от 1 до 6 при целевом значении от 2 до 5.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид/некатионный липид/стерин/конъюгированный липид в молярном соотношении 50: 10:38,5: 1,5.

В одном аспекте в настоящем изобретении предусмотрен состав на основе липидных наночастиц, содержащий фосфолипиды, лецитин, фосфатидилхолин и фосфатидилэтаноламин.

В некоторых вариантах осуществления также могут быть включены одно или несколько дополнительных соединений. Данные соединения можно вводить отдельно, или дополнительные соединения могут быть включены в липидные наночастицы по настоящему изобретению. Другими словами, липидные наночастицы могут содержать другие соединения в дополнение к нуклеиновой кислоте или по меньшей мере вторую нуклеиновую кислоту, отличающуюся от первой. Другие дополнительные соединения могут без ограничений быть выбраны из группы, состоящей из малых или больших органических или неорганических молекул, моносахаридов, дисахаридов, трисахаридов, олигосахаридов, полисахаридов, пептидов, белков, аналогов пептидов и их производных, пептидомиметиков, нуклеиновых кислот, аналогов нуклеиновых кислот и производных, экстрактов, полученных из биологических материалов, или любых их комбинаций.

В некоторых вариантах осуществления липидная наночастица (или состав, содержащий липидные наночастицы) не содержит реакционноспособных примесей (например, альдегидов или кетонов) или содержит реакционноспособные примеси (например, альдегиды или кетоны) в количестве, меньшем, чем предварительно выбранный

уровень. Без ограничения какой-либо теорией, в некоторых вариантах осуществления липидный реагент применяется для получения состава на основе липидных наночастиц, и липидный реагент может содержать загрязняющую реакционноспособную примесь (например, альдегид или кетон). Липидный реагент может быть выбран для изготовления на основании наличия реакционноспособных примесей (например, альдегидов или кетонов) в количестве, меньшем, чем предварительно выбранный уровень. Без ограничения какой-либо теорией, в некоторых вариантах осуществления альдегиды могут вызывать модификацию и повреждение РНК, например, поперечную сшивку оснований и/или ковалентное конъюгирование липида с РНК (например, образование аддуктов липид-РНК). В некоторых случаях это может привести к неудачной реакции с участием обратной транскриптазы и/или включению неподходящих оснований, например, в месте(местах) повреждения(повреждений), например, мутации во вновь синтезированной ДНК-мишени.

В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента, характеризующегося общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента, содержащего менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента: (i) характеризующегося общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержащего менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием множества липидных реагентов, и каждый липидный реагент из множества независимо соответствует одному или нескольким критериям, описанным в данном абзаце. В некоторых вариантах осуществления каждый липидный реагент из множества соответствует одному и тому же критерию, например, критерию из данного абзаца.

В некоторых вариантах осуществления состав на основе липидных наночастиц характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления состав на основе липидных наночастиц содержит менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц: (i) характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%,

0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержит менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

В некоторых вариантах осуществления один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, содержат менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления один или несколько или необязательно все липидные реагенты, используемые для получения липидной наночастицы, описанной в данном документе, или состава на ее основе: (i) характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержат менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

В некоторых вариантах осуществления общее содержание альдегидов и/или количество какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов) определяется с помощью жидкостной хроматографии (LC), например, в сочетании с тандемной масс-спектрометрией (MS/MS), например, в соответствии со способом, описанным в примере 34. В некоторых вариантах осуществления содержание реакционноспособных примесей (например, альдегидов) и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяют посредством выявления одной или нескольких химических модификаций молекулы нуклеиновой кислоты (например, молекулы РНК, например, описанной в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах. В некоторых вариантах осуществления содержание реакционноспособных примесей (например, альдегидов) и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяют посредством выявления одной или нескольких химических модификаций нуклеотида или нуклеозида (например, рибонуклеотида или рибонуклеозида, например, содержащихся в молекуле нуклеиновой кислоты-матрицы или выделенных из нее, например, как описано в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах, например, как описано в примере 7. В вариантах осуществления химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида выявляют посредством определения присутствия одного или

нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 7.

В некоторых вариантах осуществления нуклеиновая кислота (например, РНК), описанная в данном документе (например, нуклеиновая кислота-матрица или нуклеиновая кислота, кодирующая Gene Writer), не содержит альдегидную модификацию или содержит альдегидные модификации в количестве, меньшем, чем предварительно выбранное. В некоторых вариантах осуществления в среднем нуклеиновая кислота имеет менее 50, 20, 10, 5, 2 или 1 альдегидной модификации на 1000 нуклеотидов, например, где одна поперечная сшивка двух нуклеотидов представляет собой одну альдегидную модификацию. В некоторых вариантах осуществления альдегидная модификация представляет собой аддукт РНК (например, аддукт липид-РНК). В некоторых вариантах осуществления нуклеотид с альдегидной модификацией образует поперечные сшивки между основаниями. В некоторых вариантах осуществления нуклеиновая кислота (например, РНК), описанная в данном документе, содержит менее 50, 20, 10, 5, 2 или 1 поперечных сшивок между нуклеотидами.

В некоторых вариантах осуществления LNP направляются к конкретным тканям посредством добавления нацеливающих доменов. Например, биологические лиганды могут быть представлены на поверхности LNP для усиления взаимодействия с клетками, на которых представлены когнатные рецепторы, что тем самым стимулирует ассоциацию с тканями, где клетки экспрессируют рецептор, и доставку груза к ним. В некоторых вариантах осуществления биологический лиганд может представлять собой лиганд, который управляет доставкой в печень, например, LNP, на которых представлены GalNAc, приводят к доставке груза в виде нуклеиновой кислоты в гепатоциты, которые содержат асиалогликопротеиновый рецептор (ASGPR). В работе Akinc et al. *Mol Ther* 18(7):1357-1364 (2010) изложена идея конъюгации трехвалентного лиганда GalNAc и PEG-липида (GalNAc-PEG-DSG) с получением LNP, зависимых от ASGPR, для наблюдаемого эффекта доставки груза с помощью LNP (см., например, фигуру 6). Другие составы на основе LNP, на которых представлен лиганд, например, в которые включены фолат, трансферрин или антитела, обсуждаются в WO2017223135, который включен в данный документ посредством ссылки во всей своей полноте, в дополнение к литературным источникам, используемым в данном документе, а именно Kolhatkar et al., Curr Drug Discov Technol. 2011 8:197-206; Musacchio and Torchilin, Front Biosci. 2011 16:1388-1412; Yu et al., Mol Membr Biol. 2010 27:286-298; Patil et al., Crit Rev Ther Drug Carrier Syst. 2008 25:1-61; Benoit et al., Biomacromolecules. 2011 12:2708-2714; Zhao et al., Expert Opin Drug Deliv. 2008 5:309-319; Akinc et al., Mol Ther. 2010 18:1357-1364; Srinivasan et al., Methods Mol Biol. 2012 820:105-116; Ben-Arie et al., Methods Mol Biol. 2012 757:497-507; Peer 2010 J Control Release. 20:63-68; Peer et al., Proc Natl Acad Sci U S A. 2007 104:4095-4100; Kim et al., Methods Mol Biol. 2011 721:339-353; Subramanya et al., Mol Ther. 2010 18:2028-2037; Song et al., Nat Biotechnol. 2005 23:709-717; Peer et al., Science. 2008 319:627-630; и Peer and Lieberman, Gene Ther. 2011 18:1127-1133.

В некоторых вариантах осуществления LNP выбирают по тканеспецифической активности посредством добавления молекулы для селективного нацеливания на орган (SORT) в состав, содержащий традиционные компоненты, такие как ионизируемые катионные липиды, амфипатические фосфолипиды, холестерин и конъюгированные с поли(этиленгликолем) (PEG) липиды. Идеи Cheng et al. Nat Nanotechnol 15(4):313-320 (2020) демонстрируют, что добавление дополнительного компонента "SORT" точно изменяет профиль доставки РНК in vivo и опосредует тканеспецифическую (например, в легкие, печень, селезенку) доставку и редактирование генов в зависимости от процентного содержания и биофизических свойств молекулы SORT.

В некоторых вариантах осуществления LNP содержат биоразлагаемые ионизируемые липиды. В некоторых вариантах осуществления LNP содержит (9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат, также называемый 3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропил(9Z,12Z)-октадека-9,12-диеноат, или другой ионизируемый липид. См., например, липиды из WO2019/067992, WO/2017/173054, WO2015/095340 и WO2014/136086, а также приведенные в них литературные источники. В некоторых вариантах осуществления термины "катионный" и "ионизируемый" применительно к липидам LNP являются взаимозаменяемыми, например, где ионизируемые липиды являются катионными в зависимости от значения pH.

В некоторых вариантах осуществления несколько компонентов системы Gene Writer могут быть получены в виде одного состава на основе LNP, например, состав на основе LNP содержит мРНК, кодирующую полипептид Gene Writer, и РНК-матрицу. Соотношения компонентов, представляющих собой нуклеиновые кислоты, можно изменять с целью усиления до максимума свойств терапевтического средства. В некоторых вариантах осуществления соотношение РНК-матрицы и мРНК, кодирующей полипептид Gene Writer, составляет от приблизительно 1:1 до 100:1, например, от приблизительно 1:1 до 20:1, от приблизительно 20:1 до 40:1, от приблизительно 40:1 до 60:1, от приблизительно 60:1 до 80:1 или от приблизительно 80:1 до 100:1 в молярном соотношении. В других вариантах осуществления система из нескольких нуклеиновых кислот может быть получена при использовании отдельных составов, например, одного состава на основе LNP, содержащего РНК-матрицу, и второго состава на основе LNP, содержащего мРНК, кодирующую полипептид Gene Writer. В некоторых вариантах осуществления система может содержать более двух компонентов, представляющих собой нуклеиновые кислоты, составленных в виде LNP. В некоторых вариантах осуществления система может содержать белок, например, полипептид Gene Writer, и РНК-матрицу, составленные в виде по меньшей мере одного состава на основе LNP.

В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от нескольких 10 до нескольких 100 нм, например, при измерении с помощью динамического светорассеяния (DLS). В некоторых вариантах осуществления

средний диаметр LNP в составе на основе LNP может составлять от приблизительно 40 нм до приблизительно 150 нм, например, приблизительно 40 нм, 45 нм, 50 нм, 55 нм, 60 нм, 65 нм, 70 нм, 75 нм, 80 нм, 85 нм, 90 нм, 95 нм, 100 нм, 105 нм, 110 нм, 115 нм, 120 нм, 125 нм, 130 нм, 135 нм, 140 нм, 145 нм или 150 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от приблизительно 50 нм до приблизительно 100 нм, от приблизительно 50 нм до приблизительно 90 нм, от приблизительно 50 нм до приблизительно 80 нм, от приблизительно 50 нм до приблизительно 70 нм, от приблизительно 50 нм до приблизительно 60 нм, от приблизительно 60 нм до приблизительно 100 нм, от приблизительно 60 нм до приблизительно 90 нм, от приблизительно 60 нм до приблизительно 80 нм, от приблизительно 60 нм до приблизительно 70 нм, от приблизительно 70 нм до приблизительно 100 нм, от приблизительно 70 нм до приблизительно 90 нм, от приблизительно 70 нм до приблизительно 80 нм, от приблизительно 80 до приблизительно 100 нм, от приблизительно 80 до приблизительно 90 нм или от приблизительно 90 нм до приблизительно 100 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от приблизительно 70 нм до приблизительно 100 нм. В конкретном варианте осуществления средний диаметр LNP в составе на основе LNP может составлять приблизительно 80 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять приблизительно 100 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP находится в диапазоне от приблизительно 1 мм до приблизительно 500 мм, от приблизительно 5 мм до приблизительно 200 мм, от приблизительно 10 мм до приблизительно 100 мм, от приблизительно 20 мм до приблизительно 80 мм, от приблизительно 25 мм до приблизительно 60 мм, от приблизительно 30 мм до приблизительно 55 мм, от приблизительно 35 мм до приблизительно 50 мм или от приблизительно 38 мм до приблизительно 42 мм.

В некоторых случаях LNP может являться относительно однородной. Индекс полидисперсности может применяться для указания однородности LNP, например, распределения липидных наночастиц по размерам. Небольшой (например, менее 0,3) индекс полидисперсности обычно указывает на узкое распределение частиц по размерам. LNP может характеризоваться индексом полидисперсности, составляющим от приблизительно 0 до приблизительно 0,25, например, 0,01, 0,02, 0,03, 0,04, 0,05, 0,06, 0,07, 0,08, 0,09, 0,10, 0,11, 0,12, 0,13, 0,14, 0,15, 0,16, 0,17, 0,18, 0,19, 0,20, 0,21, 0,22, 0,23, 0,24 или 0,25. В некоторых вариантах осуществления индекс полидисперсности LNP может составлять от приблизительно 0,10 до приблизительно 0,20.

Дзета-потенциал LNP можно использовать для обозначения электрокинетического потенциала композиции. В некоторых вариантах осуществления дзета-потенциал может описывать поверхностный заряд LNP. Липидные наночастицы с относительно низким зарядом, положительным или отрицательным, обычно являются желательными, поскольку более сильно заряженные соединения могут взаимодействовать с клетками, тканями и

другими элементами в организме нежелательным образом. В некоторых вариантах осуществления дзета-потенциал LNP может составлять от приблизительно -10 мВ до приблизительно +20 мВ, от приблизительно -10 мВ до приблизительно +15 мВ, от приблизительно -10 мВ до приблизительно +10 мВ, от приблизительно -10 мВ до приблизительно +5 мВ, от приблизительно -10 мВ до приблизительно 0 мВ, от приблизительно -10 мВ до приблизительно -5 мВ, от приблизительно -5 мВ до приблизительно +20 мВ, от приблизительно -5 мВ до приблизительно +15 мВ, от приблизительно -5 мВ до приблизительно +10 мВ, от приблизительно -5 мВ до приблизительно +5 мВ, от приблизительно -5 мВ до приблизительно 0 мВ, от приблизительно 0 мВ до приблизительно +20 мВ, от приблизительно 0 мВ до приблизительно +15 мВ, от приблизительно 0 мВ до приблизительно +10 мВ, от приблизительно 0 мВ до приблизительно +5 мВ, от приблизительно +5 мВ до приблизительно +20 мВ, от приблизительно +5 мВ до приблизительно +15 мВ или от приблизительно +5 мВ до приблизительно +10 мВ.

Эффективность инкапсуляции белка и/или нуклеиновой кислоты, например, полипептида Gene Writer или мРНК, кодирующей полипептид, описывает количество белка и/или нуклеиновой кислоты, которое инкапсулировано или иным образом ассоциировано с LNP после получения, относительно предоставленного исходного количества. Желательно, чтобы эффективность инкапсуляции была высокой (например, близкой к 100%). Эффективность инкапсуляции можно измерить, например, путем сравнения количества белка или нуклеиновой кислоты в растворе, содержащем липидную наночастицу, до и после разрушения липидной наночастицы с помощью одного или нескольких органических растворителей или детергентов. Анионообменную смолу можно применять для измерения количества свободного белка или нуклеиновой кислоты (например, РНК) в растворе. Флуоресценцию можно применять для измерения количества свободного белка и/или нуклеиновой кислоты (например, РНК) в растворе. Для описанных в данном документе липидных наночастиц эффективность инкапсуляции белка и/или нуклеиновой кислоты может составлять по меньшей мере 50%, например, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 80%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 90%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 95%.

LNP необязательно может содержать одно или несколько покрытий. В некоторых вариантах осуществления LNP может быть составлена в виде капсулы, пленки или таблетки с покрытием. Капсула, пленка или таблетка, содержащие композицию, описанную в данном документе, могут иметь любые пригодные размер, прочность на разрыв, твердость или плотность.

Дополнительные иллюстративные липиды, составы, способы и определение характеристик LNP описаны в WO2020061457, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления липофекцию клеток in vitro или ex vivo осуществляют с использованием Lipofectamine MessengerMax (Thermo Fisher) или реагента для трансфекции TransIT-mRNA (Mirus Bio). В определенных вариантах осуществления LNP составлены с использованием смеси ионизируемых липидов GenVoy-ILM (Precision NanoSystems). В некоторых вариантах осуществления LNP составлены с использованием 2,2-дилинолеил-4-диметиламиноэтил-[1,3]-диоксолана (DLin-KC2-DMA) или дилинолеилметил-4-диметиламинобутирата (DLin-MC3-DMA или MC3), состав и применение которого in vivo описаны в Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012), включенном в данный документ посредством ссылки во всей своей полноте.

Составы на основе LNP, оптимизированные для доставки систем CRISPR-Cas, например, RNP Cas9-gRNA, gRNA, мРНК Cas9, описаны в WO2019067992 и WO2019067910, оба из которых включены посредством ссылки.

Дополнительные конкретные составы на основе LNP, пригодные для доставки нуклеиновых кислот, описаны в US8158601 и US8168775, оба из которых включены посредством ссылки, и они включают составы, применяемые в патисиране, продаваемом под наименованием ONPATTRO.

Иллюстративная дозировка LNP с Gene Writer может составлять приблизительно 0,1, 0,25, 0,3, 0,5, 1, 2, 3, 4, 5, 6, 8, 10 или 100 мг/кг (РНК). Иллюстративная дозировка AAV, содержащего нуклеиновую кислоту, кодирующую один или несколько компонентов системы, может включать MOI, составляющую приблизительно 10^{11}, 10^{12}, 10^{13} и 10^{14} vg/кг.

19. Подходящие показания

Иллюстративные подходящие заболевания и нарушения, которые можно лечить с помощью систем или способов, предусмотренных в данном документе, например, включающих в себя Gene Writer, включают без ограничений: Синдромы Барайтсера-Винтера 1 и 2; сахарный и несахарный диабет с атрофией зрительного нерва и глухотой; недостаточность альфа-1-антитрипсина; недостаточность кофактора гепарина II; адренолейкодистрофия; синдром Кеппена-Любинского; синдром Тричера Коллинза 1; недостаточность митохондриального комплекса I, II, III, III (ядерный тип 2, 4 или 8); гипермарганцемия с дистонией, полицитемией и циррозом печени; карциноидная опухоль кишечника; синдром предрасположенности к рабдоидным опухолям 2; болезнь Вильсона; гиперфенилаланинемия, BH4-дефицит в следствие частичного дефицита PTS, BH4-дефицит, D и несвязанный с ФКН; гиперинсулинемическая семейная гипогликемия 3, 4 и 5; фолликулярный кератоз; орально-лицевой-пальцевой синдром; синдром SeSAME; глухота, несиндромальная сенсоневральная, митохондриальная; протеинурия; синдром секреторной диареи при инсулинзависимом сахарном диабете; болезнь моямоя 5; анемия Даймонда-Блэкфана 1, 5, 8 и 10; синдром псевдоахондропластической спондилоэпифизарной дисплазии; синдром ломкой роговицы 2; метилмалоновая ацидемия

с гомоцистинурией; синдром Адамса-Оливера 5 и 6; аутосомно-рецессивная агаммаглобулинемия 2; пороки развития затылочной части коры головного мозга; фебрильные судороги, семейные, 11; мукополисахаридоз VI типа, типа VI (тяжелый) и VII типа; синдром Мардена Уокера; псевдонеонатальная адренолейкодистрофия; миопатия сфероидного тела; ключично-черепной дизостоз; множественные кожные и слизистые венозные мальформации; детская острая печеночная недостаточность; неонатальный внутрипеченочный холестаз, вызванный дефицитом цитрина; дефект межжелудочковой перегородки 1; окуло-денто-дигитальная дисплазия; опухоль Вильмса 1; синдром Вейля-Марчезани; почечная адисплазия; катаракта 1, 4, аутосомно-доминантная, аутосомно-доминантная, множественные, с микрокорнеа, по типу Coppock, ювенильная, с микрокорнеа и глюкозурией, ядерная диффузная непрогрессирующая; одонто-гипофосфатазия; цереброокулофациоскелетный синдром; шизофрения 15; церебральная амилоидная ангиопатия, связанная с APP; гемофагоцитарный лимфогистиоцитоз, семейный, 3; недостаточность порфобилиногенсинтазы; эпизодическая атаксия 2 типа; трихо-рино-фаланговый синдром 3 типа; прогрессирующая семейная сердечная блокада типа IB; восприимчивость к глиоме 1; синдром Лихтенштейна-Кнорра; гипогидротическая X-сцепленная эктодермальная дисплазия; синдром Барттера 3, 3 с гипокальциурией и 4 ; недостаточность карбоангидразы VA, в следствие гипераммониемии; кардиомиопатия; пойкилодермия, наследственная фиброзирующая, с контрактурами сухожилий, миопатией и легочным фиброзом; комбинированная d-2- и 1-2-гидроксиглутаровая ацидурия; недостаточность аргиназы; палочко-колбочковая дистрофия 2 и 6; синдром Смита-Лемли-Опица; муколипидоз III гамма; синдром Блау; синдром Вемера; менингиома; дефект йодтирозильного сочетания; синдром Дубина-Джонсона; недостаточность 3-оксо-5-альфа-стероид-дельта-4-дегидрогеназы; синдром Буше-Нейхаузера; накопление железа в головном мозге; умственная отсталость, сцепленная с X-хромосомой 102 и синдромальная 13; семейная предрасположенность к аденоме гипофиза; гипоплазия мозолистого тела; гиперальфалипопротеинемия 2; недостаточность ферроксидазы; нечувствительность к гормону роста при иммунодефиците; синдром Маринеско- Sj\xc3\xb6gren; синдром Марцольфа; семейный горизонтальный паралич взгляда с прогрессирующим сколиозом; синдром Митчелла-Райли; Гипокальциурическая гиперкальциемия, семейная, 1 и 3 типы; синдром Рубинштейна-Тайби; синдром Эпштейна; ювенильный ретиношизис; мышечная дистрофия Беккера; синдром Лойса-Дитца 1, 2, 3; врожденный мышечно-гипертрофический церебральный синдром; семейная ювенильная подагра; сперматогенная недостаточность 11, 3 и 8; орофациальные расщелины 11 и 7, синдром расщелины губы/неба и эктодермальной дисплазии; умственная отсталость, X-сцепленная, неспецифическая, синдромальная, типа Хедера и синдромальная, типа wu; комбинированные недостаточности окислительного фосфорилирования 1, 3, 4, 12, 15 и 25; лобно-височная деменция; дисплазия Кнеста; семейная кардиомиопатия; доброкачественная семейная гематурия; феохромоцитома; глухота, вызванная аминогликозидами; недостаточность трансаминазы гамма-аминомасляной кислоты;

глазокожный альбинизм типа IB, типа 3 и типа 4; синдром почечной колобомы; гипомиелинизация ЦНС; синдром лимфангиэктазии-лимфедемы Хеннекама 2; базилярная мигрень, семейная; дистальная спинальная мышечная атрофия, сцепленная с X-хромосомой 3; X-сцепленная перивентрикулярная гетеротопия; микроцефалия; мукополисахаридоз, MPS-IH/S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; инфантильный паркинсонизм-дистония; лобно-височная деменция с включениями TDP43, связанная с TARDBP; наследственный диффузный рак желудка; сиалидоз типа I и II; синдром микроцефалии-капиллярной мальформации; синдром наследственного рака молочной железы и яичников; поражение мелких сосудов головного мозга с кровоизлиянием; некетотическая гиперглицинемия; нейрогепатопатия Навахо; аурикулокондилярный синдром 2; спастическая параплегия 15, 2, 3, 35, 39, 4, аутосомно-доминантная, 55, аутосомно-рецессивная и 5A; аутосомно-рецессивный эластолизис типа IA и IB; гемолитическая анемия несфероцитарная, обусловленная недостаточностью глюкозофосфатизомеразы; синдром Хатчинсона-Гилфорда; семейная амилоидная нефропатия с крапивницей и глухотой; надклапанный аортальный стеноз; диффузная ладонно-подошвенная кератодермия Ботнического типа; синдром Холта-Орама; синдром Коффина-Сириса/умственная отсталость; пороки развития лево-правой оси; синдром Рападилино; нанофтальм 2; краниосиностоз и зубочелюстные аномалии; параганглиомы 1; синдром Снайдера-Робинсона; фибрилляция желудочков; синдром активированного PI3K-дельта; синдром Хауэла-Эванса; синдром Ларсена, доминантный тип; синдром Ван Мальдергема 2; MYH-ассоциированный полипоз; недостаточность 6- pymvoyl-тетрагидроптеринсинтазы; синдромы Алажиля 1 и 2; лимфангиомиоматоз; мышечно-глазное заболевание головного мозга; расстройства, связанные с WFSl; первичная гипертрофическая остеоартропатия, аутосомно-рецессивная 2; бесплодие; синдром прогерии Нестора-Гильермо; недостаточность митохондриального трифункционального белка; синдром гипоплазии левых отделов сердца 2; Первичная дилатационная кардиомиопатия; пигментный ретинит; болезнь Гиршпрунга 3; синдром Апшоу-Шульмана; дисплазия Дебукуа 2; диарея 3 (секреторная натриевая, врожденная, синдромальная) и 5 (при тафтинговой энтеропатии, врожденная); врожденная пахионихия 4 и 2 типа; церебральная аутосомно-доминантная и рецессивная артериопатия с подкорковыми инфарктами и лейкоэнцефалопатией; дистрофия формы Vi tel 1i; тип II, тип IV, IV (сочетанный печеночный и миопатический), тип V и тип VI; атипичный синдром Ретта; дефект атриовентрикулярной перегородки 4; синдром Папийона- Lef\xc3\xa8vre; амавроз Лебера; X-сцепленная наследственная моторная и сенсорная невропатия; прогрессирующая склерозирующая полиодистрофия; синдром Гольдмана-Фавра; почечно-печеночно-панкреатическая дисплазия; синдром Паллистера-Холла; амилоидогенный транстиретиновый амилоидоз; синдром Мельника-Нидлса; синдром гипериммуноглобулина E; атаксия задних столбов спинного мозга с пигментным ретинитом; точечная хондродисплазия 1, X-сцепленная рецессивная и 2 X-сцепленная доминантная; эктопия хрусталика, изолированная аутосомно-рецессивная и доминантная;

семейная холодовая крапивница; семейный аденоматозный полипоз 1 и 3; порокератоз 8, диссеминированный поверхностный актинический тип; спектр синдромов избыточного роста, связанных с PIK3CA; церебральные кавернозные мальформации 2; экссудативная витреоретинопатия 6; врожденная мраморная кожа, телеангиэктатическая мегалэнцефалия; TARP синдром; сахарный диабет, перманентный неонатальный, с неврологическими характеристиками; короткореберная грудная дисплазия 11 или 3 с полидактилией или без нее; гипертрихотическая остеохондродисплазия; бета-талассемия; болезнь Ниманна-Пика типа C1, C2, типа A и типа C1, взрослая форма; болезнь Шарко-Мари-Тута типы IB, 2B2, 2C, 2F, 21, 2U (аксональный), 1C (демиелинизирующий), доминантный промежуточный C, рецессивный промежуточный A, 2A2, 4C, 4D, 4H, IF, IVF и X; тирозинемия I типа; пароксизмальная фибрилляция предсердий; УФ-чувствительный синдром; агенезия зубов, селективная, 3 и 4; врожденная мышечная дистрофия с недостаточностью мерозина; недостаточность длинноцепочечной 3-гидроксиацил-КоА-дегидрогеназы; врожденная аниридия; гипертрабекулярность левого желудочка 5; дефицит декарбоксилазы ароматических L-аминокислот; ишемическая болезнь сердца; тотальная лейконихия; дистальный артрогрипоз типа 2B; пигментный ретинит 10, 11, 12, 14, 15, 17 и 19; синдром Робинова-Сорауфа; синдром Тенорио; пролактинома; нейрофиброматоз, тип 2; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями головного мозга и глаз, типы A2, A7, A8, A11 и A14; гетеротаксия, висцеральная, 2, 4 и 6, аутосомная; синдром Янковича Риверы; семейная парциальная липодистрофия 2 и 3 типа; заболевание гемоглобина H, без делеций; мультицентрический остеолиз, нодулез и артропатия; агенезия щитовидной железы; недостаточность семейства ацил-КоА-дегидрогеназ, члена 9; болезнь Александра; болезнь накопления фитановой кислоты; рак молочной железы и яичников, семейный 1, 2 и 4; недостаточность пролиндегидрогеназы; детская гипофосфатазия; агенезия поджелудочной железы и врожденные пороки сердца; витамин D-зависимый рахит, тип 2; иридогониодизогенез доминантный тип и тип 1; синдром аутосомно-рецессивной гипогидротической эктодермальной дисплазии; умственная отсталость, сцепленная с X-хромосомой, 3, 21, 30 и 72; наследственные геморрагические телеангиэктазии 2 типа; блефарофимоз, птоз и обратный эпикантус; недостаточность аденинфосфорибозилтрансферазы; судороги, доброкачественные семейные инфантильные, 2; акродизостоз 2, с гормонорезистентностью или без нее; тетрада Фалло; пигментный ретинит 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; недостаточность лизосомальной кислотной липазы; врожденная мышечная дистрофия типа Эйхсфельда; врожденная мышечная дистрофия Уокера-Варбурга; синдром периодической лихорадки, ассоциированной с рецептором TNF (TRAPS); прогрессирующая миоклонус-эпилепсия с атаксией; эпилепсия, детские абсансы 2, 12 (идиопатическая генерализованная, предрасположенность) 5 (лобная ночная эпилепсия), лобная ночная эпилепсия типа 1, парциальная, с различными очагами, прогрессирующая миоклония 3 и X-сцепленная, с различными нарушениями обучаемости и поведенчискими расстройствами; синдром удлиненного интервала QT; дикарбоновая аминоацидурия; брахидактилия типов A1 и A2;

заболевание, подобное эластической псевдоксантоме, с множественной недостаточностью факторов свертывания крови; синдром мультисистемной дисфункции гладкой мускулатуры; синдактилия Сенани типа Ленца; синдром Жубера 1, 6, 7, 9/15 (дигенный), 14, 16 и 17 и орофациодигитальный синдром xiv; дигитореноцеребральный синдром; ретинобластома; семейная дискинезия с миокимией лица; наследственная сенсорная и вегетативная нейропатия типа IIB и IIA; семейный гиперинсулинизм; мегалэнцефальная лейкоэнцефалопатия с подкорковыми кистами тип 2a; синдром Ааза; синдром Видемана-Штайнера; эксфолиативный ихтиоз; врожденная миотония; гранулематозное заболевание, хронический, сцепленный с X-хромосомой вариант; недостаточность 2-метилбутирил-КоА-дегидрогеназы; саркоидоз с ранним началом; глаукома, врожденная и глаукома врожденная, колобома; рак молочной железы, предрасположенность к нему; цероидный липофусциноз нейронов 2, 6, 7 и 10; врожденная генерализованная липодистрофия 2 типа; недостаточность фруктозо-бифосфатазы; врожденная контрактурная арахнодактилия; синдром Линча I и II; недостаточность фосфоглицератдегидрогеназы; синдром Берна-Маккеуна; инфаркт миокарда 1; ахроматопсия 2 и 7; пигментный ретинит 73; дефект протана; полимикрогирия, асимметричная, двусторонняя лобно-теменная; спинальная мышечная атрофия, дистальная, аутосомно-рецессивная, 5; метилмалоновая ацидурия из-за недостаточности метилмалонил-КоА-мутазы; семейная порэнцефалия; синдром Гурлера; ухо-небно-пальцевый синдром I и II типов; синдром Сотоса 1 или 2; кардиоэнцефаломиопатия, инфантильная с летальным исходом, обусловленная недостаточностью цитохром-с-оксидазы; парастремматическая карликовость; резистентность к тиреотропин-рилизинг-гормону, генерализованная; сахарный диабет 2 типа и инсулинозависимый - 20; аневризмы грудной аорты и расслоение аорты; резистентность к эстрогенам; заболевание мочи “кленового сиропа” типа 1A и типа 3; гипоспадия 1 и 2, сцепленная с X-хромосомой; метахроматическая лейкодистрофия ювенильный, поздний инфантильный и взрослый типы; ранний Т-клеточный предшественник острого лимфобластного лейкоза; нейропатия, наследственная сенсорная, тип IC; умственная отсталость, аутосомно-доминантная 31; пигментный ретинит 39; рак молочной железы с ранним началом; аномалия Мэя-Хегглина; болезнь Гоше 1 типа и подострая нейропатическая; синдром Темтами; спинальная мышечная атрофия с преимущественным поражением нижних конечностей 2 типа, аутосомно-доминантный; анемия Фанкони, группы комплемента E, I, N и O; алкаптонурия; болезнь Гиршпрунга; комбинированная малоновая и метилмалоновая ацидурия; аритмогенная правожелудочковая кардиомиопатия 5, 8 и 10 типов; врожденное липоматозное разрастание, сосудистые мальформации и эпидермальные невусы; синдром Тимоти; недостаточность гуанидиноацетатметилтрансферазы; миоклоническая дистония; болезнь Канзаки; дефект транспорта нейтральной 1 аминокислоты; нейрогипофизарный несахарный диабет; метаболизм гормонов щитовидной железы, патологический; доброкачественная лопаточно-перонеальная мышечная дистрофия с кардиомиопатией; гипогликемия с дефицитом гликогенсинтетазы в печени; гипертрофическая

кардиомиопатия; миастенический синдром, врожденный, 11, связанный с дефицитом ацетилхолиновых рецепторов; синдромная Х-сцепленная умственная отсталость 5; синдром Сторморкена; апластическая анемия; умственная отсталость; нормокалиемический периодический паралич, калий-чувствительный; болезнь Данона; нефронофтиз 13, 15 и 4; тиреотоксический периодический паралич и тиреотоксический периодический паралич 2; бесплодие, связанное с многохвостыми сперматозоидами и дополнительными ДНК; глаукома, первичная открытоугольная, с ранним началом; афибриногенемия и врожденная афибриногенемия; поликистоз почек 2, взрослый и детский типы; поздняя семейная кожная порфирия; церебелло-окуло-ренальный синдром (нефронофтиз, окуломоторная апраксия и мозжечковые нарушения); лобно-височная деменция, сцепленная с хромосомой 3 и убиквитин-позитивная лобно-височная деменция; метатрофическая дисплазия; синдром иммунодефицита, нестабильности центромерного участка и лицевых аномалий 2; анемия несфероцитарная гемолитическая, обусловленная дефицитом G6PD; бронхоэктазы с повышенным содержанием хлоридов в поте или без 3; врожденная миопатия с диспропорцией мышечных волокон; комплекс Карнея, тип 1; крипторхизм односторонний или двусторонний; буллезный ихтиоз Сименса; изолированный дефицит лютропина; DFNA 2 несиндромальная потеря слуха; синдром Кляйна-Ваарденберга; синдром серых тромбоцитов; дефект синтеза желчных кислот, врожденный, 2; 46, нарушение формирования пола XY, типы 1, 3 и 5; острая интермитирующая порфирия; синдромы Корнелии де Фанге 1 и 5; гиперглицинурия; колбочко-палочковая дистрофия 3; дисфибриногенемия; синдром Карака; врожденная мышечная дистрофия-дистрогликанопатия без умственной отсталости, тип B5; инфантильный нистагм, сцепленный с Х-хромосомой; врожденный дискератоз, аутосомно-рецессивный, 1, 3, 4 и 5; микроцефалия с хориоретинопатией, лимфедемой или умственной отсталостью или без; гиперлизинемия; синдромы Барде-Бидля 1, 11, 16 и 19; аутосомно-рецессивная центронуклеарная миопатия; синдром Фрейзера; синдром каудальной регрессии; врожденный фиброз экстраокулярных мышц, 1, 2, 3a (с или без экстраокулярного поражения), 3б; синдром Прадера-Вилли; злокачественная меланома; синдром Блума; заболевание Дарье, сегментарное; многоочаговая остеолизная нефропатия; гемохроматоз 1, 2B и 3 типов; детская мозжечковая атаксия с прогрессирующей наружной офтальмоплегией и мозжечковая атаксия, умственная отсталость и синдром нарушения равновесия 2; синдром гипоплазии левых отделов сердца; эпилепсия, потеря слуха и синдром умственной отсталости; локус 2 количественного признака уровня трансферрина в сыворотке крови; глазной альбинизм I типа; синдром Марфана; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями головного мозга и глаз, тип A14 и B14; гипераммониемия III типа; синдром криптофтальма; врожденная универсальная алопеция; гипофосфатазия взрослых; недостаточность манноз-связывающего белка; макулярная дистрофия «бычий глаз»; аутосомно-доминантная торсионная дистония 4; нефротический синдром, тип 3, тип 5, с глазными аномалиями или без них, тип 7 и тип 9; судороги, ранняя детская эпилептическая энцефалопатия 7;стойкая гиперинсулинемическая гипогликемия

грудного возраста; тромбоцитопения, сцепленная с Х-хромосомой; неонатальная гипотония; синдром Орставика-Линдемана-Сольберга; первичная легочная гипертензия, 1, с наследственными геморрагическими телеангиэктазиями; гипофизарно-зависимый гиперкортицизм; бородавчатая эпидермодисплазия; буллезный эпидермолиз, узловой, локальный вариант; дефицит цитохром с оксидазы i; синдром Киндлера; миосклероз, аутосомно-рецессивный; артериальный ствол; синдром Дуэйна 2 типа; ВЗРОСЛЫЙ синдром; спектр синдрома Цельвегера; лейкоэнцефалопатия с атаксией, с поражением ствола и спинного мозга и повышением лактата, с исчезновением белого вещества и прогрессирующая, с недостаточностью яичников; дефицит антитромбина III; голопрозэнцефалия 7; синдром фокомелии Робертса- SC; синдром истощения митохондриальной ДНК 3 и 7, гепатоцеребральные типы и 13 (энцефаломиопатический тип); порэнцефалия 2; микроцефалия, нормальный интеллект и иммунодефицит; гигантская аксональная нейропатия; синдром Стерджа-Вебера, капиллярные мальформации, врожденные, 1; заболевание Фабри и заболевание Фабри, кардиальный вариант; недостаточность глутаматформиминотрансферазы; синдром Фанкони-Биккеля; акромикрическая дисплазия; эпилепсия, идиопатическая генерализованная, предрасположенность к 12; кальцификация базальных ганглиев, идиопатическая, 4; миопатия с полиглюкозановыми тельцами 1 с иммунодефицитом или без; злокачественная опухоль простаты; врожденная эктодермальная дисплазия лица; врожденный порок сердца; возрастная макулодистрофия 3, 6, 11 и 12; врожденная миотония, аутосомно-доминантная и рецессивная формы; гипомагниемия 1, кишечная; недостаточность сульфитоксидазы, изолированная; заболевание Пика; дефицит плазминогена, тип I; синдактилия 3 типа; колбочко-палочковая дистрофия, несовершенный амелогенез; псевдопервичный гиперальдостеронизм; терминальная костная дисплазия; антенатальный синдром Барттера 2 типа; врожденная мышечная дистрофия-дистрогликанопатия с умственной отсталостью, типы B2, B3, B5 и B15; семейная инфантильная миастения; лимфопролиферативный синдром 1, 1 (сцепленный с Х-хромосомой) и 2; гиперхолестеринемия и гиперхолестеринемия, аутосомно-рецессивная; новообразование яичника; инфантильный ганглиозидоз GM1; синдромная Х-сцепленная умственная отсталость 16; недостаточность рибозо-5-фосфатизомеразы; заболевание Альцгеймера, типы 1, 3 и 4; синдром Андерсена-Тавила; синдром множественных синостозов 3; красная волчанка обморожения 1; гемофагоцитарный лимфогистиоцитоз, семейный, 2; синдром Аксенфельда-Ригера 3 типа; центрально-ядерная миопатия; остеоартроз с легкой хондродисплазией; нарушения биогенеза пероксисом; тяжелая врожденная нейтропения; наследственная невралгическая амиотрофия; ладонно-подошвенная кератодермия, неэпидермолитическая, очаговая или диффузная; дисплазминогенемия; семейный колоректальный рак; спастическая атаксия 5, аутосомно-рецессивная, типа Шарлевуа-Сагене, 1, 10 или 11, аутосомно-рецессивная; лобно-метафизарная дисплазия 3; наследственное заболевание недостаточности факторов II, IX, VIII; спондилохейродисплазия, по типу синдрома Элерса-Данлоса, с иммунодефицитом, Аггреканского типа, с врожденными вывихами суставов, укороченно-

кистевого типа, Седагатского типа, с колбочко-палочковой дистрофией, и Козловского типа; синдром ихтиоза недоношенных; синдром Стиклера 1 типа; очаговый сегментарный гломерулосклероз 5; дефицит 5-оксопролиназы; синдромальная микрофтальмия 5, 7 и 9; ювенильный полипоз/синдром наследственной геморрагической телеангиэктазии; недостаточность бутирил-КоА-дегидрогеназы; сахарный диабет взрослого возраста у молодых, тип 2; уственная отсталость, синдромальная, типа Класа-Йенсена, сцепленная с Х-хромосомой; глухота, кохлеарная, с миопией и умственными нарушениями, без вестибулярного поражения, аутосомно-доминантная, Х-сцепленная 2; синдром спондилокарпотарзального синостоза; STING-ассоциированная васкулопатия, ранее начало; болезнь накопления нейтральных липидов с миопатией; иммунная дисфункция с инактивацией Т-клеток из-за дефекта входа кальция 2; кожно-кардиальный синдром; недостаточность кортикостеронметилоксидазы 2 типа; наследственная миопатия с ранней дыхательной недостаточностью; интерстициальный нефрит, кариомегалический; триметиламинурия; гипериммуноглобулинеимя D с периодической лихорадкой; предрасположенность к злокачественной гипертермии 1 типа; трихомегалия с умственной отсталостью, карликовостью и пигментной дистрофией сетчатки; аденокарцинома молочной железы; недостаточность фактора комплемента В; врожденная мышечная дистрофия Ульриха; гипертрабекулярность левого желудочка; заболевание рыбьего глаза; финский врожденный нефротический синдром; поясно-конечностная мышечная дистрофия, тип IB, 2A, 2B, 2D, Cl, C5, C9, C14; идиопатический фиброзирующий альвеолит, хроническая форма; первичная семейная гипертрофическая кардиомиопатия; доброкачественное увеличение в сыворотке ангиотензин i-превращающего фермента; недостаточность Cd8, семейная; синдром Протея; дефект транспорта глюкозо-6-фосфата; синдром Борджесона-Форсмана-Лемана; синдром Зельвегера; спинальная мышечная атрофия II типа; рак предстательной железы, наследственный, 2; тромбоцитопения, дисфункция тромбоцитов, гемолиз и нарушение синтеза глобина; врожденные нарушения гликозилирования типов IB, ID, 1G, 1H, 1 J, IK, IN, IP, 2C, 2J, 2K, Ilm; пограничный буллезный эпидермолиз Герлица; генерализованная эпилепсия с фебрильными приступами плюс 3, тип 1, тип 2; шизофрения 4; ишемическая болезнь сердца, аутосомно-доминантная 2; врожденный дискератоз, аутосомно-доминантный, 2 и 5; подкорковая ламинарная гетеротопия, сцепленная с Х-хромосомой; недостаточность аденилаткиназы; Х-сцепленный тяжелый комбинированный иммунодефицит; копропорфирия; амилоидная кардиомиопатия, связанная с транстиретином; гипокальциемия, аутосомно-доминантная 1; синдром Бругада; врожденный миастенический синдром, чувствительный к ацетазоламиду; первичная гипомагниемия; склеростеоз; лобно-височная деменция и/или боковой амиотрофический склероз 3 и 4; мевалоновая ацидурия; шванноматоз 2; наследственная моторная и сенсорная нейропатия с атрофией зрительного нерва; поздняя кожная порфирия; рассекающий остеохондрит; судороги, доброкачественные семейные неонатальные, 1 и/или миокимия; синдром удлиненного интервала QT, подтип LQT1; умственная отсталость, передняя протрузия верхней челюсти и косоглазие; идиопатическая

гиперкальциемия грудного возраста; гипогонадотропный гипогонадизм 11 с аносмией или без; поликистозная липомембранозная остеодисплазия со склерозирующей лейкоэнцефалопатией; первичная аутосомно-рецессивная микроцефалия 10, 2, 3 и 5; прерванная дуга аорты; врожденная амегакариоцитарная тромбоцитопения; синдром Германского-Пудлака 1, 3, 4 и 6; синдром удлиненного интервала QT 1, 2, 2/9, 2/5, (дигенный), 3, 5 и 5, приобретенный, предрасположенность к; синдром Андермана; дистрофия колбочек сетчатки 3Б; эритропоэтическая протопорфирия; недостаточность сепиаптеринредуктазы; недостаточность ацил-КоА-дегидрогеназы с очень длинной углеродной цепью; гиперферритинемический катарактальный синдром; синдром спастической параплегии Сильвера; болезнь Шарко-Мари-Тута; дефект межпредсердной перегородки 2; синдром Карнавала; врожденная нечувствительность к боли с ангидрозом; катехоламинергическая полиморфная желудочковая тахикардия; гипокалиемический периодический паралич 1 и 2 степени; синдром внезапной детской смерти; гипохромная микроцитарная анемия с перегрузкой железом; синдром дефицита GLUT1 2; лейкодистрофия, гипомиелинизирующая, 11 и 6; колбочковый монохроматизм; остеопетроз аутосомно-доминантный 1 и 2 типа, рецессивный 4, рецессивный 1, рецессивный 6; тяжелая врожденная нейтропения 3, аутосомно-рецессивная или доминантная; недостаточность метионин-аденозилтрансферазы, аутосомно-доминантная; пароксизмальная семейная фибрилляция желудочков; недосаточность пируваткиназы эритроцитов; дисплазия Шнекенбекена; пируэтная желудочковая тахикардия; дистальная миопатия типа Марксбери-Григгса; недостаточность УДФ-глюкозо-гексозо-1-фосфат уридилилтрансферазы; внезапная сердечная смерть; синдром Ней-Лаксова 1; атрансферринемия; гиперпаратиреоз 1 и 2; злокачественная меланома кожи 1; симфалангизм, проксимальный, lb; прогрессирующая псевдоревматоидная дисплазия; болезнь Верднига-Гоффмана; ахондрогенез 2 типа; голопрозэнцефалия 2, 3, 7 и 9; хаболевание Шиндлера, тип 1; цереброретинальная микроангиопатия с кальцинатами и кистами; гетеротаксия, висцеральная, сцепленная с Х-хромосомой; синдром туберозного склероза; синдром Картагенера; резистентность к гормонам щитовидной железы, генерализованная, аутосомно-доминантная; бестрофинопатия, аутосомно-рецессивная; поражение ногтей, несиндромальное врожденное, 8; синдром Мора-Транебьерга; колбочко-палочковая дистрофия 12; нарушение слуха; овариолейкодистрофия; почечный канальцевый ацидоз, проксимальный, с глазными аномалиями и умственной отсталостью; недостаточность дигидроптеридинредуктазы; фокальная эпилепсия с нарушением речи с задержкой психического развития или без; синдром атаксии-телеангиэктазии; синдром Брауна-Виалетто-Ван Лаэра и синдром Брауна-Виалетто-Ван Лаэра 2; кардиомиопатия; периферическая демиелинизирующая нейропатия, центральная дисмиелинизация; эндотелиальная дистрофия роговицы Фукса 4; синдром Каудена 3; дистония 2 (торсионная, аутосомно-рецессивная), 3 (торсионная, Х-сцепленная), 5 (допа-чувствительный тип), 10, 12, 16, 25, 26 (миоклоническая); эпифизарная дисплазия, множественная, с миопией и кондуктивной глухотой; неспецифическое нарушение проводимости сердца;

бранхиотические синдромы 2 и 3; нарушение биогенеза пероксисом 14B, 2A, 4A, 5B, 6A, 7A и 7B; семейная почечная глюкозурия; кандидоз, семейный, 2, 5, 6 и 8; аутоиммунное заболевание, мультисистемное, с инфантильным началом; ранняя детская эпилептическая энцефалопатия 2, 4, 7, 9, 10, 11, 13 и 14; синдром Сегава, аутосомно-рецессивный; глухота, аутосомно-доминантная 3a, 4, 12, 13, 15, аутосомно-доминантная несиндромальная сенсоневральная 17, 20 и 65; врожденная дизэритропоэтическая анемия I и II типа; синдром усиленного ответа S-колбочек; нейрональный цероидный липофусциноз взрослых; фибрилляция предсердий, семейная, 11, 12, 13 и 16; болезнь Норум; остеосаркома; частичный альбинизм; недостаточность биотинидазы; сочетанные клеточные и гуморальные иммунные дефекты с гранулемами; энцефалопатия Альперса; недостаточность голокарбоксилазсинтетазы; сахарный диабет взрослого возраста у молодых, тип 1, тип 2, тип 11, тип 3 и тип 9; вариегатная порфирия; инфантильный кортикальный гиперостоз; недостаточность тестостерон-17-бета-дегидрогеназы; L-2-гидроксиглутаровая ацидурия; тирозиназо-отрицательный глазо-кожный альбинизм; первичная цилиарная дискинезия 24; понтоцеребеллярная гипоплазия 4 типа; цилиарная дискинезия, первичная, 7, 11, 15, 20 и 22; тдиопатическая кальцификация базальных ганглиев 5; атрофия головного мозга; краниосиностозы 1 и 4; кератоконус 1; РАСопатия; врожденная гиперплазия надпочечников и врожденная гипоплазия надпочечников, сцепленная с Х-хромосомой; синдром истощения митохондриальной ДНК 11, 12 (кардиомиопатический тип), 2, 4B (тип MNGIE), 8B (тип MNGIE); брахидактилия при гипертонии; плоская роговица 2; синдром Аарского; множественная эпифизарная дисплазия 5 или доминантная; эндотелиальная дистрофия роговицы 2 типа; недостаточность аминоацилазы 1; задержка речевого и языкового развития; синдром Николаидеса-Барайцера; недостаточность энтерокиназы; эктродактилия, эктодермальная дисплазия и синдром расщелины губы/неба 3; множественный врожденный артрогрипоз, дистальный, сцепленный с Х-хромосомой; синдром Перро 4; синдром Джервелла и Ланге-Нильсена 2; наследственные неполипозные колоректальные новообразования; синдром Робинова, аутосомно-рецессивный, аутосомно-рецессивный, с брахи-син-полидактилией; нейрофибросаркома; недостаточность цитохром-с-оксидазы; пузырно-мочеточниковый рефлюкс 8; недостаточность дофамин-бета-гидроксилазы; синдром углеводдефицитных гликопротеинов I и II типов; прогрессирующий семейный внутрипеченочный холестаз 3; доброкачественные семейные неонатально-инфантильные судороги; панкреатит, хронический, восприимчивость к; ризомелическая точечная хондродисплазия 2 и 3 типа; нарушение стероидогенеза из-за недостаточности цитохрома p450 оксидоредуктазы; глухота с лабиринтной аплазией, микротией и микродонтией (FAMM); синдром Ротмунда-Томсона; комплексная кортикальная дисплазия с другими пороками развития головного мозга 5 и 6; миастения, семейная инфантильная, 1; трихоринофаланговая дисплазия I типа; заболевание Worth; гипоплазия селезенки; недостаточность кофактора молибдена, группа комплементации A; синдром Себастьяна; прогрессирующий семейный внутрипеченочный холестаз 2 и 3; синдром Вейля-Маркезани 1 и 3; микроцефалическая остеодиспластическая

примордиальная карликовость 2 типа; нарушение обмена сурфактанта, легочное, 2 и 3; тяжелая Х-сцепленная миотубулярная миопатия; рак поджелудочной железы 3; кровотечения тромбоцитарного типа 15 и 8; тирозиназо-положительный глазо-кожный альбинизм; синдром Борроне-ди-Рокко-Кровато; синдром ATR-X; недостаточность сахаразы-изомальтазы; частичная недостаточность компонента комплемента 4 в следствии дисфункционального ингибитора cl; врожденная центральная гиповентиляция; инфантильная гипофосфатазия; недостаточность ингибитора активатора плазминогена 1 типа; злокачественная лимфома, неходжкинская; синдром гиперорнитинемии-гипераммониемии-гомоцитруллинурии; синдром Шварца-Джампеля 1 типа; локус 1 количественного признака фетального гемоглобина; миопатия дистальная с началом в передней большеберцовой кости; синдром Нунана 1 и 4, синдром LEOPARD 1; глаукома 1, открытоугольная, e, F и G; синдром Кенни-Кэффи 2 типа; PTEN-ассоциированный синдром гамартомных опухолей; мышечная дистрофия Дюшенна; инсулинорезистентный сахарный диабет и черный акантоз; микрофтальм изолированный 3, 5, 6, 8 и при колобоме 6; синдром Рейна; преждевременная недостаточность яичников 4, 5, 7 и 9; синдром Аллана-Хемдона-Дадли; цитруллинемия I типа; болезнь Альцгеймера, семейная, 3, со спастическими парапарезами и апраксией; семейная гемиплегическая мигрень 1 и 2 типов; вентрикуломегалия при кистозной болезни почек; эластическая псевдоксантома; гомоцистеинемия в следствии недостаточности MTHFR, недостаточности CBS и гомоцистинурия, реагирующая на пиридоксин; дилатационная кардиомиопатия 1A, 1AA, 1C, 1G, IBB, 1DD, IFF, 1HH, II, IKK, IN, IS, 1Y и 3B; недостаточность гуанинoксидазы АМФ в мышцах; семейный рак молочной железы; наследственная сидеробластная анемия; миоглобинурия, острая рецидивирующая, аутосомно-рецессивная; нейроферритинопатия; аритмия сердца; синдром дефицита переносчика глюкозы 1 типа; последовательность, приводящая к голопрозэнцефалии; ангиопатия наследственная с нефропатией, аневризмами и мышечными спазмами; недостаточность изовалерил-КоА-дегидрогеназы; синдром Каллмана 1, 2 и 6; постоянный неонатальный сахарный диабет; акрокаллезный синдром по типу Шинцеля; синдром Гордона; расстройства, связанные с MYH9; синдром Доннея Барроу; тяжелая врожденная нейтропения и 6, аутосомно-рецессивная; заболевание Шарко-Мари-Тута, типы ID и IVF; синдром Коффина-Лоури; недостаточность митохондриальной 3-гидрокси-3-метилглутарил-КоА-синтазы; гипомагниемия, судороги и умственная отсталость; ишиопателлярная дисплазия; множественные врожденные аномалии - гипотония - судорожный синдром 3; спастическая параплегия 50, аутосомно-рецессивная; низкорослость с неспецифическими аномалиями скелета; тяжелая миоклоническая эпилепсия в младенчестве; пропионовая ацидемия; подростковый нефронофтиз; макроцефалия, макросомия, синдром лицевого дисморфизма; заболевание Штаргардта 4; синдром Элерса-Данлоса 7-го типа (аутосомно-рецессивный), классический тип, 2-й тип (прогероидный), гидроксилизин-дефицитный, 4-й тип, вариант 4-го типа и обусловлен дефицитом тенасцина-X; миопия 6; плоский таз; семейный холодовой аутовоспалительный синдром 2; пороки развития сердца и крупных сосудов; болезнь фон

Виллебранда типа 2M и типа 3; недостаточность галактокиназы; синдром Бругада 1; X-сцепленный ихтиоз с недостаточностью стерилсульфатазы; врожденная колобома глаза; гистиоцитоз-лимфаденопатия плюс синдром; аниридия, мозжечковая атаксия и умственная отсталость; гипертрабекулярность левого желудочка 3; боковой амиотрофический склероз типов 1, 6, 15 (с лобно-височной деменцией или без), 22 (с лобно-височной деменцией или без) и 10; несовершенный остеогенез тип 12, тип 5, тип 7, тип 8, тип I, тип III, с нормальными склерой, доминантная форма, рецессивный перинатальный летальный; гематологическое новообразование; фавизм, подверженность; легочный фиброз и/или недостаточность костного мозга, связанные с теломерами, 1 и 3; доминантная наследственная атрофия зрительного нерва; доминантный дистрофический буллезный эпидермолиз с отсутствием кожи; мышечная дистрофия врожденная, мегакониального типа; множественные атрезии желудочно-кишечного тракта; синдром МакКьюна-Олбрайта; синдром ногтя-надколенника; синдром нейроакантоцитоза Мак-Леода; системный вариабельный иммунодефицит 9; частичная недостаточность гипоксантин-гуанинфосфорибозилтрансферазы; псевдогипоальдостеронизм 1 типа аутосомно-доминантный и рецессивный и 2 тип; недостаточность уроканатгидратазы; гетеротопия; синдром Меккеля 7 типа; Ch\xc3\xa9diak-Higashi синдром, синдром Чедиака-Хигаси, взрослый тип; тяжелый комбинированный иммунодефицит в следствии недостаточности ADA, с микроцефалией, задержкой роста и чувствительностью к ионизирующему излучению, атипичный, аутосомно-рецессивный, Т-клеточный отрицательный, В-клеточный положительный, NK-клеточный отрицательный или NK-положительный; резистентность к инсулину; недостаточность стероид-11-бета-монооксигеназы; синдром подколенного птеригиума; легочная артериальная гипертензия, связанная с наследственными геморрагическими телеангиэктазиями; глухота, аутосомно-рецессивная 1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86 и 89; Первичная гипероксалурия, тип I, тип и тип III; врожденная парамиотония фон Эйленбурга; синдром Дебюкуа; недостаточность карнитинпальмитоилтрансферазы I, II, II (позднее начало) и II (инфантильный); вторичный гипотиреоз; нижнечелюстно-лицевой дизостоз, по типу Treacher Collins, аутосомно-рецессивный; синдром Каудена 1; синдром Ли-Фраумени 1; недостаточность аспарагинсинтетазы; «сотовидная» ретинальная дистрофия Дойна; атрофия зрительного нерва 9; инфантильные судороги и пароксизмальный хореоатетоз, семейные; атаксия при дефиците витамина E; гиперплазия островковых клеток; мышечная дистрофия Миоши 1; тромбофилия, наследственная, обусловленная дефицитом протеина C, аутосомно-доминантная и рецессивная; синдром Фехтнера; недостаточность пропердина, сцепленная с X-хромосомой; умственная отсталость, стереотипные движения, эпилепсия и/или церебральные пороки развития; недостаточность креатина, X-сцепленная; пиломатриксома; цианоз, преходящий неонатальный и атипичный нефропатический; взрослая атаксия с окуломоторной апраксией; гемангиома, капиллярная инфантильная; PC-K6a; генерализованный доминантный дистрофический буллезный эпидермолиз; болезнь Пелизеуса-Мерцбахера; центроядерная миопатия 1, врожденная с избытком мышечных

веретен, дистальная 1, лактоацидоз и сидеробластная анемия 1, митохондриальная прогрессирующая с врожденной катарактой, тугоухостью, задержкой развития и тубулярным агрегатом 2; доброкачественные семейные неонатальные судороги 1 и 2; первичная легочная гипертензия; лимфедема первичная с миелодисплазией; врожденный синдром удлиненного интервала QT; семейная экссудативная витреоретинопатия, сцепленная с Х-хромосомой; аутосомно-доминантная гипогидротическая эктодермальная дисплазия; примордиальная карликовость; семейный легочный капиллярный гемангиоматоз; недостаточность карнитин-ацилкамитин-транслоказы; висцеральная миопатия; семейная средиземноморская лихорадка и семейная средиземноморская лихорадка, наследуемая по аутосомно-доминантному типу; комбинированная частичная и полная недостаточность 17-альфа-гидроксилазы/17, 20-лиазы; ухо-небо-пальцевый синдром I типа; нефролитиаз/остеопороз, гипофосфатемия, 2; Семейная гиперлипопротеинемия 1 и 3 типа; фенотипы; CHARGE-синдром; синдром Фурманна; синдром гипотрихоз-лимфедемы-телеангиэктазии; хондродисплазия типа Бломстранда; акроэритрокератодермия; замедленная скорость нервной проводимости, аутосомно-доминантная; наследственный синдром, предрасполагающий к раку; краниодиафизарная дисплазия, аутосомно-доминантная; спиноцеребеллярная атаксия аутосомно-рецессивная 1 и 16; недостаточность пропротеинконвертазы 1/3; D-2-гидроксиглутаровая ацидурия 2; гиперэкплексия 2 и наследственная гиперэкплексия; миопатия центрального стержня; синдром Опица G/BBB; муковисцидоз; дистрофия боуменовой мембраны II типа Тиля-Бенке; недостаточность бисфосфоглицератмутазы; недостаточность митохондриальной короткоцепочечной эноил-КоА-гидратазы 1; синдром ломкости кожи при эктодермальной дисплазии; вольфрамоподобный синдром, аутосомно-доминантный; микроцитарная анемия; недостаточность пируваткарбоксилазы; недостаточность адгезии лейкоцитов I и III типов; множественная эндокринная неоплазия, тип 4; Транзиторный буллезный дермолиз новорожденных; синдром Примулы; немелкоклеточный рак легкого; врожденная мышечная дистрофия; комбинированная недостаточность липазы; СИНДРОМ КОУЛА-КАРПЕНТЕРА 2; дефект атриовентрикулярной перегородки и общий атриовентрикулярный переход; недостаточность ксантиноксидазы; синдром Ваарденбурга типа 1, 4C и 2E (с неврологическим поражением); синдром Стиклера, типы I (несиндромальный глазной) и 4; синдром ломкой роговицы, голубых склер и гипермобильность суставов; микросферофакия; синдром Чадли-Маккаллоу; эпидермолизная буллезная симплексная и поясно-конечностная мышечная дистрофия, симплексная с пятнистой пигментацией, симплексная с атрезией привратника, симплексная, аутосомно-рецессивная и с атрезией привратника; расстройство Ретта; нарушение миграции нейронов; недостаточность гормона роста с аномалиями гипофиза; болезнь Ли; ладонно-подошвенный полосатый кератоз 1; синдром Вейсенбахера-Цвеймюллера; недостаточность ацил-кофермента А дегидрогеназы со средней длиной цепи; недостаточность УДФ-глюкозо-4-эпимеразы; предрасположенность к аутизму, сцепленному с Х-хромосомой 3; регматогенная отслойка сетчатки, аутосомно-

доминантная; семейные фебрильные судороги 8; отсутствие локтевой и малоберцовой кости при тяжелом дефиците конечностей; гипертрабекулярность левого желудочка 6; центромерная нестабильность хромосом 1,9 и 16 и иммунодефицит; наследственная диффузная лейкоэнцефалопатия со сфероидами; синдром Кушинга; сниженная плотность допаминовых рецепторов d2 в головном мозге; C-подобный синдром; почечная дисплазия, пигментная дистрофия сетчатки, мозжечковая атаксия и скелетная дисплазия; дисгенезия яичников 1; синдром Пирсона; полинейропатия, потеря слуха, атаксия, пигментный ретинит и катаракта; прогрессирующий внутрипеченочный холестаз; аутосомно-доминантный, аутосомно-рецессивный и сцепленный с X-хромосомой рецессивный синдром Альпорта; синдром Ангельмана; синдром детской эпилепсии амишей; аутоиммунный лимфопролиферативный синдром, тип 1a; гидроцефалия; марфаноидный фенотип; синдром голых лимфоцитов 2 типа, группа комплементации E; рецессивный дистрофический буллезный эпидермолиз; фактор H, VII, X, v и фактор viii, комбинированная недостаточность 2, xiii, субъединицы, недостаточность; зонулярная пылевидная катаракта 3; бородавки, гипогаммаглобулинемия, инфекции и миелокатексис; доброкачественная наследственная хорея; недостаточность гиалуронглюкозаминидазы; микроцефалия, грыжа пищеводного отверстия диафрагмы и нефротический синдром; задержка роста и умственного развития, нижнечелюстно-лицевой дизостоз, микроцефалия и расщелина неба; лимфедема, наследственная, id; задержка полового созревания; явный избыток минералокортикоидов; генерализованная артериальная кальцификация младенческого возраста 2; МЕТИЛМАЛОНОВАЯ АЦИДУРИЯ, ТИП mut(0); врожденный порок сердца, несколько типов, 2; семейная гипоплазия, гломерулокистозная почка; цереброокулофациоскелетный синдром 2; болезнь Штаргардта 1; умственная отсталость, аутосомно-рецессивные 15, 44, 46 и 5; недостаточность пролидазы; метилмалоновая ацидурия типа cblB; заболевание Огучи; эндокринно-цереброостеодисплазия; лиссэнцефалия 1, 2 (сцепленная с X-хромосомой), 3, 6 (с микроцефалией), сцепленная с X-хромосомой; соматотрофная аденома; синдром Гамсторпа-Вольфарта; липидный протеиноз; миопатия с тельцами включения 2 и 3; синдром увеличенного вестибулярного водопровода; остеопороз с псевдоглиомой; приобретенный синдром удлиненного интервала QT; фенилкетонурия; CHOPS-синдром; глобальная задержка развития; кристаллическая корнеоретинальная дистрофия Биетти; расстройство, подобное синдрому Нунана, с ювенильным миеломоноцитарным лейкозом или без; врожденная эритропоэтическая порфирия; наследственная бульбовидная атрофия; параганглиомы 3; синдром Ван дер Вуда; недостаточность ароматазы; синдром дисморфизма умственной отсталости Бирка Барела; боковой амиотрофический склероз 5 типа; метгемоглобинемия I 1 и 2 типов; врожденная стационарная куриная слепота типов 1A, IB, 1C, IE, IF и 2A; судороги; рак щитовидной железы, фолликулярный; синдром летальной врожденной контрактуры 6; дистальная наследственная моторная нейронопатия типа 2B; стромальная опухоль полового тяжа; эпилептическая энцефалопатия с началом в детском возрасте, ранняя инфантильная, 1, 19, 23, 25, 30 и 32; миофибриллярная миопатия 1 и связанная с

ZASP; детская мозжечковая атаксия с прогрессирующей наружной офтальмоплегией; недостаточность пуриннуклеозидфосфорилазы; дефекты переднего мозга; эпилептическая энцефалопатия типа Леннокса-Гасто; ожирение; 4, гипертрабекулярность левого желудочка 10; синдром Верхейя; синдром Мовата-Уилсона; одонтотрихомелический синдром; узорчатая дистрофия пигментного эпителия сетчатки; синдром Lig4; синдром Бараката; недостаточность IRAK4; соматотрофная аденома; недостаточность киназы дегидрогеназы кетокислот с разветвленной цепью; цистинурия; семейная аплазия червоподобного отростка; недостаточность сукцинил-КоА-ацетоацетаттрансферазы; лопаточно-перонеальная спинальная мышечная атрофия; пигментная дистрофия сетчатки; тромбастения Гланцмана; первичная открытоугольная глаукома с началом в юношеском возрасте 1; синдромы Айкарди-Гутьера 1, 4 и 5; почечная дисплазия; задержка внутриутробного развития, метафизарная дисплазия, врожденная гипоплазия надпочечников и генитальные аномалии; монилетрикс; низкий рост, ониходисплазия, лицевой дисморфизм и гипотрихоз; метахроматическая лейкодистрофия; недостаточность лизосомной кислой липазы; синдром 3 M 2; врожденный амавроз Лебера 11, 12, 13, 16, 4, 7 и 9; мандибулоакральная дисплазия с липодистрофией типа A или B, атипичная; синдромы Мейера-Горлина 4; гипотрихоз 8 и 12; синдром короткого интервала QT 3; эктодермальная дисплазия 1 ib; анонихия; псевдогипопаратиреоз 1A типа, псевдопсевдогипопаратиреоз; атрофия зрительного нерва Лебера; синдром Бейнбриджа-Роперса; синдром Уивера; низкий рост, атрезия слухового прохода, гипоплазия нижней челюсти, аномалии скелета; недостаточность альфа-маннозидазы; макулярная дистрофия, вителлиформная, с началом во взрослом возрасте; глутаровая ацидурия 1 типа; ганглиозидоз GM1 типа (с поражением сердца) 3; мандибулоакральный дизостоз; наследственная лимфедема I типа; остановка предсердий 2; синдром грима Кабуки; миопатия Бетлема и миопатия Бетлема 2; недостаточность миелопероксидазы; Флековская дистрофия роговицы; наследственный энтеропатический акродерматит; гипобеталипопротеинемия семейная, ассоциированная с apob32; синдром Коккейна типа A; тяжелый неонатальный гиперпаратиреоз; расстройство, подобное атаксии-телеангиэктазии; синдром Пендреда; I система групп крови; семейная доброкачественная пузырчатка; висцеральная гетеротаксия 5, аутосомная; нефрогенный несахарный диабет, нефрогенный несахарный диабет, X-сцепленный; миопатия с мини-ядрами и наружной офтальмоплегией; синдром Перри; гипогидротический/волосистый/зубной тип, аутосомно-рецессивный; наследственный панкреатит; умственная отсталость и микроцефалия с гипоплазией моста и мозжечка; заболевание накопления гликогена 0 (мышечная), II (взрослая форма), IXa2, IXc, тип 1A; полосчатая остеопатия с черепным склерозом; недостаточность глутатионсинтетазы; синдром Бругада и синдром Бругада 4; эндометриальная карцинома; гипогидротическая эктодермальная дисплазия с иммунодефицитом; внутрипеченочный холестаз беременных 3; синдром Бемара-Сулье, типы A1 и A2 (аутосомно-доминантный); болезнь Салла; недостаточность орнитинаминотрансферазы; синдром опухоли гамартомы PTEN; дистихиаз-синдром лимфедемы; недостаточность кортикостероидсвязывающего

глобулина; нейрональный цероидный липофусциноз взрослых; болезнь Дежерина-Сотта; тетраамелия, аутосомно-рецессивная; синдром Сениора-Локена 4 и 5, ; глутаровая ацидемия IIA и IIB; аневризма аорты, семейная грудная 4, 6 и 9; гиперфосфатазия с синдромом умственной отсталости 2, 3 и 4; врожденный дискератоз, сцепленный с X-хромосомой; артрогрипоз, почечная дисфункция и холестаз 2; синдром Баннаяна-Райли-Рувалькабы; 3- метилглутаконовая ацидурия; изолированная недостаточность 17,20-лиазы; синдром Горлина; HFGS; заболевание Тея-Сакса, вариант B1, Gm2-ганглиозидоз (у взрослых), Gm2-ганглиозидоз (начало у взрослых); заболевание Даулинга-Дего 4; заболевание Паркинсона 14, 15, 19 (юношеское начало), 2, 20 (раннее начало), 6, (аутосомно-рецессивное раннее начало и 9; сенсорная, аутосомно-доминантная атаксия; врожденная микроворсинчатая атрофия; миоклонико-атоническая эпилепсия; заболевание Танжера; 2-метил-3-гидроксимасляная ацидурия; семейная почечная гипоурикемия; шизэнцефалия; синдром истощения митохондриальной ДНК 4B, тип MNGIE; синдром Фейнгольда 1; дефект почечного транспорта карнитина; семейная гиперхолестеринемия; Таунса-Брокса-бранхио-ото-реноподобный синдром; синдром Гришелли тип 3; синдром Меккеля-Грубера; буллезная ихтиозиформная эритродермия; синдром нейтрофильного иммунодефицита; миастенический синдром, врожденный, 17, 2A (медленный канал), 4B (быстрый канал) и без канальцевых агрегатов; микрососудистые осложнения сахарного диабета 7; синдром МакКьюсика Кауфмана; хроническое гранулематозное заболевание, аутосомно-рецессивное, цитохром b-положительное, типы 1 и 2; недостаточность аргининосукцинатлиазы; недостаточность переносчиков митохондриального фосфата и пирувата; решетчатая l дистрофия роговицы III типа; синдром эктодермальной дисплазии-синдактилии 1; гипомиелинизирующая лейкодистрофия 7; умственная отсталость, аутосомно-доминантная 12, 13, 15, 24, 3, 30, 4, 5, 6 и 9; генерализованная эпилепсия с фебрильными судорогами плюс 1 и 2 типы; восприимчивость к псориазу 2; синдром Франка Тер Хаара; аневризмы грудной аорты и расслоение аорты; синдром Крузона; гранулезоклеточная опухоль яичника; эпидермолитическая ладонно-подошвенная кератодермия; дисхондростеоз Лери Вейля; недостаточность 3-бета-гидроксистероиддегидрогеназы; семейная рестриктивная кардиомиопатия 1; аутосомно-доминантная прогрессирующая наружная офтальмоплегия с делециями 1 и 3 митохондриальной ДНК; синдром Антли-Бикслера с аномалиями половых органов и нарушением стероидогенеза; синдром Хайду-Чейни; пигментно-узелковое заболевание коры надпочечников, первичное, 1; эпизодический болевой синдром, семейный, 3; синдром Дежерина-Сотта, аутосомно-доминантный; FG синдром и FG синдром 4; недостаточность дендритных клеток, моноцитов, B-лимфоцитов и природных лимфоцитов-киллеров; гипотиреоз врожденный, незобный, 1; синдром Миллера; немалиновая миопатия 3 и 9; олигодонтия - синдром колоректального рака; синдром холодового потоотделения 1; болезнь Ван Бухема 2 типа; глаукома 3, первичная врожденная, d; цитруллинемия I и II типа; миопатия Нонака; врожденная мышечная дистрофия вследствие частичного дефицита LAMA2; синдром мионевральной желудочно-кишечной энцефалопатии; синдром Ли из-за

недостаточности митохондриального комплекса I; медуллобластома; недостаточность пируватдегидрогеназы El-альфа; рак толстой кишки; синдром Нанса-Хорана; заболевание Сандхоффа, взрослый и детский типы; артрогрипоз, почечная дисфункция, синдром холестаза; аутосомно-рецессивная гипофосфатемическая болезнь костей; «Сотовидная» ретинальная дистрофия Дойна; спиноцеребеллярная атаксия 14, 21, 35, 40 и 6; деменция с тельцами Леви; митохондриальное заболевание, связанное с RRM2B; Броуди миопатия; синдром мегалэнцефалии-полимикрогирии-полидактилии-гидроцефалии 2; синдром Ушера, типы 1, IB, ID, 1G, 2A, 2C и 2D; тип гипокальцификации и тип гипоматурации, IIA1 несовершенный амелогенез; недостаточность гормонов гипофиза, сочетание 1, 2, 3 и 4; Симфалангизм Кушинга; почечный тубулярный ацидоз, дистальный, аутосомно-рецессивный, с поздней сенсоневральной тугоухостью или с гемолитической анемией; детский нефронофтиз; синдром ювенильного полипоза; сенсорная атаксическая невропатия, дизартрия и офтальмопарез; недостаточность 3-гидроксиацил-КоА-дегидрогеназы; карцинома паращитовидной железы; Х-сцепленная агаммаглобулинемия; тиамин-зависимая мегалобластная анемия, с сахарным диабетом и сенсоневральной глухотой; множественная недостаточность сульфатазы; нейродегенерация с накоплением железа в головном мозге 4 и 6; недостаточность холестеринмонооксигеназы (расщепляющей боковые цепи); гемолитическая анемия вследствие недостаточности аденилосукцинатлиазы; миоклоническая эпилепсия с разорванными красными волокнами; синдром Питта-Хопкинса; синдром множественных птеригиумов, синдром Эскобара; гомоцистинурия-мегалобластная анемия вследствие дефекта метаболизма кобаламина, комплементарный тип cblE; холецистит; сфероцитоз 4 и 5 типов; множественные врожденные аномалии; пигментная ксеродерма, комплементарная группа b, группа D, группа E и группа G; болезнь Лейнера; комеальная дистрофия Гроеноу I типа; недостаточность коэнзима Q10, первичные 1, 4 и 7; врожденная непрогрессирующая дистальная спинальная мышечная атрофия; микросиндром Варбурга 2 и 4; дефект синтеза желчных кислот, врожденный, 3; АКТГ-независимая макронодулярная гиперплазия надпочечников 2; акро-капито-феморальная дисплазия; заболевание костей Педжета, семейное; тяжелая неонатальная энцефалопатия с микроцефалией; синдром Циммерманна-Лабанда и синдром Циммерманна-Лабанда 2; синдром Райфенштейна; семейная гипокалиемия-гипомагниемия; светочувствительная трихотиодистрофия; узловой буллезный эпидермолиз у взрослых; рак легких; синдром Фримена-Шелдона; синдром гиперинсулинизма-гипераммониемии; задняя полярная катаракта 2 типа; Склерокорнеа, аутосомно-рецессивный; ювенильный GM >1< ганглиозидоз; синдром Коэна; наследственные синдромы параганглиомы-феохромоцитомы; неонатальный инсулинозависимый сахарный диабет; гипохондрогенез; синдром плавучей гавани; эластолизис с остеодистрофией и тяжелыми легочными, желудочно-кишечными и мочевыми нарушениями; врожденные контрактуры конечностей и лица, гипотония и задержка развития; врожденный дискератоз аутосомно-доминантный и аутосомно-доминантный, 3; гистиоцитарный медуллярный ретикулез; синдром Костелло;

иммунодефицит 15, 16, 19, 30, 31C, 38, 40, 8, обусловленный дефектом cd3-дзета, с гипериммуноглобулинемией M 1 и 2 и X-сцепленным наследованием, с дефектом магния, вирусом Эпштейна-Барра и неоплазией ; дефекты межпредсердной перегородки 2, 4 и 7 (с дефектами атриовентрикулярной проводимости или без них); недостаточность GTP-циклогидролазы I; эквиноварусная деформация стопы; недостаточность фосфоглицераткиназы 1; туберозный склероз 1 и 2; аутосомно-рецессивный врожденный ихтиоз 1, 2, 3, 4A и 4B; и семейная гипертрофическая кардиомиопатия 1, 2, 3, 4, 7, 10, 23 и 24.

20. *Показания по ткани*

Дополнительные подходящие заболевания и нарушения, которые можно лечить с помощью систем и способов, представленных в данном документе, включают без ограничения заболевания центральной нервной системы (ЦНС) (см. иллюстративные заболевания и пораженные гены в **таблице 13**), заболевания глаз (см. иллюстративные заболевания и пораженные гены в **таблице 14**), заболевания сердца (см. иллюстративные заболевания и пораженные гены в **таблице 15**), заболевания гемопоэтических стволовых клеток (HSC) (см. иллюстративные заболевания и пораженные гены в **таблице 16**), заболевания почек (см. иллюстративные заболевания и пораженные гены в **таблице 17**), заболевания печени (см. иллюстративные заболевания и пораженные гены в **таблице 18**), заболевания легких (см. иллюстративные заболевания и пораженные гены в **таблице 19**), заболевания скелетных мышц (см. иллюстративные заболевания и пораженные гены в **таблице 20**) и заболевания кожи (см. иллюстративные заболевания и пораженные гены в **таблице 21**). В **таблице 22** представлены иллюстративные защитные мутации, которые снижают риск указанных заболеваний. В некоторых вариантах осуществления система Gene Writer, описанная в данном документе, применяется для лечения показаний, указанных в любой из таблиц 13-21. В некоторых вариантах осуществления система Gene Writer модифицирует сайт-мишень в геномной ДНК в клетке, где сайт-мишень находится в гене, указанном в любой из таблиц 13-21, например, у субъекта, имеющего соответствующие показания, указанные в любой из таблиц 13-21. В некоторых вариантах осуществления Gene Writer исправляет мутацию в гене. В некоторых вариантах осуществления Gene Writer вставляет последовательность, которая была удалена из гена (например, вследствие мутации, вызывающей заболевание). В некоторых вариантах осуществления Gene Writer удаляет последовательность, которая была дуплицирована в гене (например, вследствие мутации, вызывающей заболевание). В некоторых вариантах осуществления Gene Writer заменяет мутацию (например, мутацию, вызывающую заболевание) соответствующей последовательностью дикого типа. В некоторых вариантах осуществления мутация представляет собой замену, вставку, делецию или инверсию.

21. Таблица 13. Заболевания ЦНС и пораженные гены

Заболевание	**Пораженный ген**
Альфа-маннозидоз	MAN2B1

Атаксия-телеангиэктазия	ATM
CADASIL	NOTCH3
Болезнь Канавана	ASPA
Дефицит карбамоилфосфатсинтетазы 1	CPS1
Заболевание CLN1	PPT1
Заболевание CLN2	TPP1
Заболевание CLN3 (ювенильный нейрональный цероидный липофусциноз, болезнь Баттена)	CLN3
Синдром Коффина-Лоури	RPS6KA3
Врожденный миастенический синдром 5	COLQ
Синдром Корнелии де Ланге (NIPBL)	NIPBL
Синдром Корнелии де Ланге (SMC1A)	SMC1A
Синдром Драве (SCN1A)	SCN1A
Глициновая энцефалопатия (GLDC)	GLDC
GM1-ганглиозидоз	GLB1
Болезнь Хантингтона	HTT
Гидроцефалия со стенозом сильвиева водопровода	L1CAM
Синдром Ли	SURF1
Метахроматическая лейкодистрофия (ARSA)	ARSA
MPS 2 типа	IDS
MPS 3 типа	Тип 3a: SGSH Тип 3b: NAGLU
Муколипидоз IV типа	MCOLN1
Нейрофиброматоз 1 типа	NF1
Нейрофиброматоз 2 типа	NF2
Нейродегенерация, ассоциированная с пантотенаткиназой	PANK2
Пиридоксин-зависимая эпилепсия	ALDH7A1
Синдром Ретта (MECP2)	MECP2
Болезнь Сандхоффа	HEXB
Семантическая деменция (лобно-височная деменция)	MAPT

Спиноцеребеллярная атаксия с аксональной нейропатией (атаксия с окуломоторной апраксией)	SETX
Болезнь Тея-Сакса	HEXA
Х-сцепленная адренолейкодистрофия	ABCD1

22. Таблица 14. Заболевания глаз и пораженные гены

Заболевание	**Пораженный ген**
Ахроматопсия	CNGB3
Врожденный амавроз (LCA1)	GUCY2D
Врожденный амавроз (LCA10)	CEP290
Врожденный амавроз (LCA2)	RPE65
Врожденный амавроз (LCA8)	CRB1
Хороидеремия	CHM
Палочко-колбочковая дистрофия (ABCA4)	ABCA4
Палочко-колбочковая дистрофия (CRX)	CRX
Палочко-колбочковая дистрофия (GUCY2D)	GUCY2D
Цистиноз, глазной ненефропатический	CTNS
Решетчатая дистрофия роговицы I типа	TGFBI
Макулярная дистрофия роговицы (MCD)	CHST6
Оптическая атрофия	OPA1
Пигментный ретинит (AR)	USH2A
Пигментный ретинит (AD)	RHO
Болезнь Штаргардта	ABCA4
Вителлиформная макулярная дистрофия	BEST1, PRPH2

23. Таблица 15. Заболевания сердца и пораженные гены

Заболевание	**Пораженный ген**
Аритмогенная кардиомиопатия правого желудочка (ARVC)	PKP2
Синдром Барта	TAZ
Мышечная дистрофия Беккера	DMD
Синдром Бругада	SCN5A
Катехоламинергическая полиморфная желудочковая тахикардия (RYR2)	RYR2
Дилатационная кардиомиопатия (LMNA)	LMNA

Дилатационная кардиомиопатия (TTN)	TTN
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса I типа	EMD
Семейная гипертрофическая кардиомиопатия	MYH7
Семейная гипертрофическая кардиомиопатия	MYBPC3
Синдром Джервелла-Ланге-Нильсена	KCNQ1
Дефицит LCHAD	HADHA
Поясно-конечностная мышечная дистрофия типа 1B (Эмери-Дрейфуса EDMD2)	LMNA
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Синдром удлиненного интервала QT 1 типа (Романо-Уорда)	KCNQ1

24. Таблица 16. Заболевания HSC и пораженные гены

Заболевание	**Пораженный ген**
ADA-SCID	ADA
Адренолейкодистрофия (CALD)	ABCD1
Альфа-маннозидоз	MAN2B1
Хроническая гранулематозная болезнь	CYBB; CYBA; NCF1; NCF2; NCF4
Вариабельный неклассифицируемый иммунодефицит	TNFRSF13B
Анемия Фанкони	FANCA; FANCC; FANCG
Болезнь Гоше	GBA
Глобоидно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Гемофагоцитарный лимфогистиоцитоз	PRF1; STX11; STXBP2; UNC13D
IL-7R SCID	IL7R
JAK-3 SCID	JAK3
Злокачественный детский остеопетроз - аутосомно-рецессивный остеопетроз	TCIRG1; вовлечены многие гены
Метахроматическая лейкодистрофия	ARSA; PSAP
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS

MPS7	GUSB
Муколипидоз II типа	GNPTAB
Болезнь Ниманна-Пика A и B	SMPD1
Болезнь Ниманна-Пика C	NPC1
Пароксизмальная ночная гемоглобинурия	PIGA
Болезнь Помпе	GAA
Дефицит пируваткиназы (PKD)	PKLR
Дефицит RAG 1/2 (SCID с гранулемами)	RAG1/RAG2
Тяжелая наследственная нейтропения	ELANE; HAX1
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB
Синдром Вискотта-Олдрича	WAS
X-сцепленная агаммаглобулинемия	BTK
X-сцепленный SCID	IL2RG

25. Таблица 17. Заболевания почек и пораженные гены

Заболевание	**Пораженный ген**
Синдром Альпорта	COL4A5
Аутосомно-доминантный поликистоз почек (PKD1)	PKD1
Аутосомно-доминантный поликистоз почек (PKD2)	PDK2
Аутосомно-доминантное тубулоинтерстициальное заболевание почек (MUC1)	MUC1
Аутосомно-доминантное тубулоинтерстициальное заболевание почек (UMOD)	UMOD
Аутосомно-рецессивный поликистоз почек	PKHD1
Врожденный нефротический синдром	NPHS2
Цистиноз	CTNS

26. Таблица 18. Заболевания печени и пораженные гены

Заболевание	**Пораженный ген**

Острая интермиттирующая порфирия	HMBS
Синдром Алажилля	JAG1
Дефицит альфа-1-антитрипсина	SERPINA1
Дефицит карбамоилфосфатсинтетазы I	CPS1
Цитруллинемия I типа	ASS1
Синдром Криглера-Найяра	UGT1A1
Болезнь Фабри	LPL
Синдром семейной хиломикронемии	GLA
Болезнь Гоше	GBE1
GSD IV	GBA
Гем A	F8
Гем B	F9
Наследственный амилоидоз (hTTR)	TTR
Наследственный ангионевротический отек	SERPING1 (KLKB1 для CRISPR)
HoFH	LDLRAP1
Гиперхолестеринемия	PCSK9
Метилмалоновая ацидемия	MMUT
MPS II	IDS
MPS III	Тип IIIa: SGSH Тип IIIb: NAGLU Тип IIIc: HGSNAT Тип IIId: GNS
MPS IV	Тип IVA: GALNS Тип IVB: GLB1
MPS VI	ARSB
MSUD	Тип Ia: BCKDHA Тип Ib: BCKDHB Тип II: DBT
Дефицит OTC	OTC
Поликистозная болезнь печени	PRKCSH
Болезнь Помпе	GAA

Первичная гипероксалурия 1 типа	AGXT (HAO1 или LDHA для CRISPR)
Прогрессирующий семейный внутрипеченочный холестаз 1 типа	ATP8B1
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11
Прогрессирующий семейный внутрипеченочный холестаз 3 типа	ABCB4
Пропионовая ацидемия	PCCB; PCCA
Болезнь Вильсона Болезнь накопления гликогена типа 1a Болезнь накопления гликогена типа IIIb Изовалериановая ацидемия Болезнь Вольмана	ATP7B G6PC AGL IVD LIPA

27. Таблица 19. Заболевания легких и пораженные гены

Заболевание	**Пораженный ген**
Дефицит альфа-1-антитрипсина	SERPINA1
Кистозный фиброз	CFTR
Первичная цилиарная дискинезия	DNAI1
Первичная цилиарная дискинезия	DNAH5
Первичная легочная гипертензия I типа	BMPR2
Дефицит сурфактантного белка B (SP-B) (нарушение метаболизма легочного сурфактанта 1)	SFTPB

28. Таблица 20. Заболевания скелетных мышц и пораженные гены

Заболевание	**Пораженный ген**
Мышечная дистрофия Беккера	DMD
Миотония Беккера	CLCN1
Бетлемская миопатия	COL6A2
Центронуклеарная миопатия, X-сцепленная (миотубулярная)	MTM1
Врожденный миастенический синдром	CHRNE
Мышечная дистрофия Дюшенна	DMD

Мышечная дистрофия Эмери-Дрейфуса, AD	LMNA
Лице-лопаточно-плечевая мышечная дистрофия	Хромосомная область DUX4 - D4Z4
Гиперкалиемический периодический паралич	SCN4A
Гипокалиемический периодический паралич	CACNA1S
Поясно-конечностная мышечная дистрофия 2A	CAPN3
Поясно-конечностная мышечная дистрофия 2B	DYSF
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Мышечная дистрофия Миеси 1	DYSF
Врожденная парамиотония	SCN4A
Миотония Томсена	CLCN1
Миопатия VCP (IBMPFD) 1	VCP

29. Таблица 21. Заболевания кожи и пораженные гены

Заболевание	**Пораженный ген**
Доминантная форма дистрофического буллезного эпидермолиза	COL7A1
Рецессивный дистрофический буллезный эпидермолиз (подтип Аллопо-Сименса)	COL7A1
Буллезный узловой эпидермолиз	LAMB3
Простой буллезный эпидермолиз	KRT5; KRT14
Эпидермолитический ихтиоз	KRT1; KRT10
Болезнь Хейли-Хейли	ATP2C1
Ламеллярный ихтиоз/небуллезная врожденная ихтиозиформная эритродермия (ARCI)	TGM1
Синдром Нетертона	SPINK5

30. Таблица 22. Иллюстративные защитные мутации, которые снижают риск заболеваний

Заболевание	**Ген**	**Иллюстративная защитная мутация**
Болезнь Альцгеймера	APP	A673T
Болезнь Паркинсона	SGK1	

Сахарный диабет (II типа)	SLC30A8	p.Arg138X, p.Lys34SerfsX50
Сердечно-сосудистые заболевания	PCSK9	R46L
Сердечно-сосудистые заболевания	ASGR1	NM_001671.4, c.284-36_283+33delCTGGGGCTGGGG; NP_001662.1, p.W158X
Сердечно-сосудистые заболевания	NPC1L1	p.Arg406X
Сердечно-сосудистые заболевания	APOC3	R19X; IVS2+1G→A; A43T
Сердечно-сосудистые заболевания	LPA	
Сердечно-сосудистые заболевания	ANGPTL4	E40K
Сердечно-сосудистые заболевания	ANGPTL3	p.Ser17Ter; p.Asn121fs; p.Asn147fs; c.495+6T→C
HIV-инфекция	CCR5	CCR5-дельта32

31. *Патогенные мутации*

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для коррекции патогенной мутации. Патогенная мутация может быть генетической мутацией, которая повышает восприимчивость или предрасположенность человека к определенному заболеванию или нарушению. В некоторых вариантах осуществления патогенная мутация представляет собой мутацию, вызывающую заболевание, в гене, ассоциированном с заболеванием или нарушением. В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для реверсии патогенной мутации к ее аналогу дикого типа. В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для замены патогенной мутации на последовательность, не вызывающую заболевание или нарушение.

В **таблице 23** приведены иллюстративные показания (столбец 1), лежащие в основе гены (столбец 2) и патогенные мутации, которые можно скорректировать с помощью систем или способов, описанных в данном документе (столбец 3).

32. Таблица 23. Показания, гены и причинные патогенные мутации

Заболевание	**Ген**	**Патогенная мутация#**
Ахроматопсия	CNGB3	1148delC
Дефицит альфа-1-антитрипсина	SERPINAI	E342K
Дефицит альфа-1-антитрипсина	SERPINAI	E342K
Дефицит альфа-1-антитрипсина	SERPINAI	R48C (R79C)
Врожденный амавроз (LCA10)	CEP290	2991+1655A>G
Синдром Андерсена-Тавила	KCNJ2	R218W
Аритмогенная кардиомиопатия правого желудочка (ARVC)	PKP2	c.235C>T
ассоциированное с врожденным дефицитом фактора XI	F11	E117*
ассоциированное с врожденным дефицитом фактора XI	F11	F283L
ATTR-амилоидоз	TTR	V50M/N30M
Аутосомно-доминантная глухота	COCH	G88E
Аутосомно-доминантная глухота	TECTA	Y1870C
Аутосомно-доминантная болезнь Паркинсона	SNCA	A53T
Аутосомно-доминантная болезнь Паркинсона	SNCA	A30P
Аутосомно-доминантный рахит	FGF23	R176Q
Аутосомно-рецессивная глухота	CX30	T5M
Аутосомно-рецессивная глухота	DFNB59	R183W
Аутосомно-рецессивная глухота	TMC1	Y182C
Аутосомно-рецессивная гиперхолестеринемия	ARH	Q136*
Анемия Даймонда-Блекфена	RPS19	R62Q
Монохроматизм синего конуса	OPN1LW	C203R
Синдром Бругада	SCN5A	E1784K
Синдром CADASIL	Ген NOTCH3	R90C
Синдром CADASIL	Ген NOTCH3	R141C
Болезнь Канавана	ASPA	E285A

Болезнь Канавана	ASPA	Y231X
Болезнь Канавана	ASPA	A305E
Дефицит карнитинпальмитоилтрансферазы II	CPT2	S113L
Хороидеремия	CHM	R293*
Хороидеремия	CHM	R270*
Хороидеремия	CHM	A117A
Цитруллинемия I типа	ASS	G390R
Классическая галактоземия	GALT	Q188R
Классическая гомоцистеинурия	CBS	T191M
Классическая гомоцистеинурия	CBS	G307S
Заболевание CLN2	TPP1	c.509-1 G>C
Заболевание CLN2	TPP1	c.622 C>T
Заболевание CLN2	TPP1	c.851 G>T
Палочко-колбочковая дистрофия	GUCY2D	R838C
Врожденный дефицит фактора V	F5	R506Q
Врожденный дефицит фактора V	F5	R534Q
Врожденный дефицит фактора VII	F7	A294V
Врожденный дефицит фактора VII	F7	C310F
Врожденный дефицит фактора VII	F7	R304Q
Врожденный дефицит фактора VII	F7	QI00R
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	E200K
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	M129V
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	P102L
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	D178N
Кистозный фиброз	CFTR	G551D
Кистозный фиброз	CFTR	W1282*
Кистозный фиброз	CFTR	R553*
Кистозный фиброз	CFTR	R117H
Кистозный фиброз	CFTR	дельта-F508
Цистиноз	CTNS	W138*
Болезнь Дарье	ATP2A2	N767S
Болезнь Дарье	ATP2A2	N767S

Болезнь Дарье	ATP2A2	N767S
Буллезный узловой эпидермолиз	LAMB3	R42X
Буллезный узловой эпидермолиз	LAMB3	R635X
Семейный боковой амиотрофический склероз (ALS)	SOD1	A4V
Семейный боковой амиотрофический склероз (ALS)	SOD1	H46R
Семейный боковой амиотрофический склероз (ALS)	SOD1	G37R
Болезнь Гоше	GBA	N370S
Болезнь Гоше	GBA	N370S
Болезнь Гоше	GBA	L444P
Болезнь Гоше	GBA	L444P
Болезнь Гоше	GBA	L483P
Дефицит глутарил-CoA-дегидрогеназы	GCDH	R138G
Дефицит глутарил-CoA-дегидрогеназы	GCDH	M263V
Дефицит глутарил-CoA-дегидрогеназы	GCDH	R402W
Глициновая энцефалопатия	GLDC	A389V
Глициновая энцефалопатия	GLDC	G771R
Глициновая энцефалопатия	GLDC	T269M
Гемофилия A	F8	R2178C
Гемофилия A	F8	R550C
Гемофилия A	F8	R2169H
Гемофилия A	F8	R1985Q
Гемофилия B	F9	T342M
Гемофилия B	F9	R294Q
Гемофилия B	F9	R43Q
Гемофилия B	F9	R191H
Гемофилия B	F9	G106S
Гемофилия B	F9	A279T
Гемофилия B	F9	R75*
Гемофилия B	F9	R294*

Гемофилия B	F9	R379Q
Врожденный дефицит антитромбина I типа	SERPINCI	R48C (R79C)
Врожденный хронический панкреатит	PRSS1	R122H
Синдром Хантера	IDS	R88C
Синдром Хантера	IDS	G374G
Синдром Гурлер (MPS1)	IDUA	Q70*
Синдром Гурлер (MPS1)	IDUA	W402*
Гиперкалиемический периодический паралич	SCN4A	T704M
Гиперкалиемический периодический паралич	SCN4A	M1592V
Гиперкалиемический периодический паралич	CACNA1S	p.Arg528X
Гиперкалиемический периодический паралич	CACNA1S	p.Arg1239
Интермиттирующая порфирия	HMBS	R173W
Изолированная агаммаглобулинемия	E47	E555K
Решетчатая дистрофия роговицы I типа	TGFBI	Arg124Cys
Дефицит LCHAD	HADHA	Glu474Gln
Врожденный амавроз Лебера 2 типа	RPE65	R44*
Врожденный амавроз Лебера 2 типа	RPE65	IVS1
Врожденный амавроз Лебера 2 типа	RPE65	G-A, +5
Синдром Леша-Нихена	HPRTI	R51*
Синдром Леша-Нихена	HPRTI	R170*
Поясно-конечностная мышечная дистрофия типа 2D	SGCA	Arg77Cys
Синдром Марото-Лами (MSPVI)	ARSB	Y210C
Средиземноморский вариант дефицита G6PD	G6PD	S188D
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Эпителиальная дистрофия роговицы Месманна	KRT12	L132P

Метахроматическая лейкодистрофия	ARSA	P426L
Метахроматическая лейкодистрофия	ARSA	c.459+1G>A
Синдром Моркио (MPSIVA)	GALNS	R386C
Муколипидоз IV типа	MCOLN1	406-2A>G
Муколипидоз IV типа	MCOLN1	511_6943del
Болезнь Ниманна-Пика типа A	SMPDI	L302P
Нейрональный цероидный липофусциноз (NCL)	CLN2	R208*
Нейрональный цероидный липофусциноз 1 типа	PPT1	R151*
Болезнь Паркинсона	LRRK2	G2019S
Синдром Пендреда	PDS	T461P
Синдром Пендреда	PDS	L236P
Синдром Пендреда	PDS	c.1001+1G>A
Синдром Пендреда	PDS	IVS8, +1 G>A,
Фенилкетонурия	PAH	R408W
Фенилкетонурия	PAH	I65T
Фенилкетонурия	PAH	R261Q
Фенилкетонурия	PAH	IVS10-11G>A
Фенилкетонурия	PCDH15	R245*
Фенилкетонурия	PCDH15	R245*
Болезнь Помпе	GAA	c.-32-13T>G
Первичная цилиарная дискинезия	DNAI1	IVS1+2_3insT
Первичная цилиарная дискинезия	DNAH5	10815delT
Первичная гипоксалурия	AGXT	G170R
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11	D482G (c.1445A>G)
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11	E297G
Пропионовая ацидемия	PCCB; PCCA	c.1218_1231del14ins12
Эластическая псевдоксантома	ABCC6	R1141*
Дефицит пируваткиназы (PKD)	PKLR	c.1456c->T

Пигментный ретинит	USH2a	C759F
Пигментный ретинит	IMPDHI	D226N
Пигментный ретинит	PDE6A	V685M
Пигментный ретинит	PDE6A	D670G
Пигментный ретинит	PRPF3	T494M
Пигментный ретинит	PRPF8	H2309R
Пигментный ретинит	RHO	P23H
Пигментный ретинит	RHO	P347L
Пигментный ретинит	RHO	P347L
Пигментный ретинит	RHO	D190N
Пигментный ретинит	RPI	R667*
Пигментный ретинит/синдром Ашера типа 1C	USH1C	V72V
Синдром Ретта	MECP2	R106W
Синдром Ретта	MECP2	R133C
Синдром Ретта	MECP2	R306C
Синдром Ретта	MECP2	R168*
Синдром Ретта	MECP2	R255*
Синдром Санфилиппо A (MPSIIIA)	SGSH	R74C
Синдром Санфилиппо A (MPSIIIA)	SGSH	R245H
Синдром Санфилиппо B (MPSIIIB)	NAGLU	R297*
Синдром Санфилиппо B (MPSIIIB)	NAGLU	Y140C
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	Q3*
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E26K
Серповидно-клеточная анемия	HBB	E26K
Серповидно-клеточная анемия	HBB	E7K
Серповидно-клеточная анемия	HBB	c.-138C>T
Серповидно-клеточная анемия	HBB	IVS2

Серповидно-клеточная анемия	HBB	654 C>T
Синдром Слая (MPSVII)	GUSB	L175F
Болезнь Штаргардта	ABCA4	A1038V
Болезнь Штаргардта	ABCA4	A1038V
Болезнь Штаргардта	ABCA4	L541P
Болезнь Штаргардта	ABCA4	G1961E
Болезнь Штаргардта	ABCA4	G1961E
Болезнь Штаргардта	ABCA4	G1961E
Болезнь Штаргардта	ABCA4	G1961E
Болезнь Штаргардта	ABCA4	c.2588G>C
Болезнь Штаргардта	ABCA4	c.5461-10 T>C
Болезнь Штаргардта	ABCA4	c.5714+5G>A
Болезнь Тея-Сакса	HEXA	InsTATC1278
Тирозинемия 1 типа	FAH	P261L
Синдром Ашера типа 1F	PCDH15	R245*
Пестрая порфирия	PPOX	R59W
Миопатия VCP (IBMPFD) 1	VCP	R1555X
Болезнь фон Гирке	G6PC	Q347*
Болезнь фон Гирке	G6PC	Q347*
Болезнь фон Гирке	G6PC	Q347*
Болезнь фон Гирке	G6PC	R83C
Болезнь Вильсона	ATP7B	E297G
X-сцепленная миотубулярная миопатия	MTMI	c.1261-10A>G
X-сцепленный ретиношизис	RS1	R102W
X-сцепленный ретиношизис	RS1	R141C

[#]: См. J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, включенный в данный документ посредством ссылки во всей своей полноте, в отношении подробностей номенклатуры генных мутаций. * означает стоп-кодон.

33. *Компенсаторные редактирования*

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для введения компенсаторного редактирования. В некоторых вариантах осуществления компенсаторное редактирование находится в положении гена, ассоциированного с заболеванием или нарушением, которое отличается от положения мутации, вызывающей заболевание. В некоторых вариантах осуществления

компенсаторная мутация находится не в гене, содержащем причинную мутацию. В некоторых вариантах осуществления компенсаторное редактирование может нивелировать или компенсировать мутацию, вызывающую заболевание. В некоторых вариантах осуществления компенсаторное редактирование может быть введено с помощью систем или способов, предусмотренных в данном документе, для подавления или реверсии мутантного эффекта мутации, вызывающей заболевание.

В **таблице 24** приведены иллюстративные показания (столбец 1), гены (столбец 2) и компенсаторные редактирования, которые можно ввести с помощью систем или способов, описанных в данном документе (столбец 3). В некоторых вариантах осуществления компенсаторные редактирования, представленные в таблице 24, могут быть введены для подавления или реверсии мутантного эффекта мутации, вызывающей заболевание.

34. Таблица 24. Показания, гены, компенсаторные редактирования и иллюстративные признаки конструкций

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Дефицит альфа-1-антитрипсина	SERPINAI	F51L
Дефицит альфа-1-антитрипсина	SERPINAI	M374I
Дефицит альфа-1-антитрипсина	SERPINAI	A348V/A347V
Дефицит альфа-1-антитрипсина	SERPINAI	K387R
Дефицит альфа-1-антитрипсина	SERPINAI	T59A
Дефицит альфа-1-антитрипсина	SERPINAI	T68A
ATTR-амилоидоз	TTR	A108V
ATTR-амилоидоз	TTR	R104H
ATTR-амилоидоз	TTR	T119M
Кистозный фиброз	CFTR	R555K
Кистозный фиброз	CFTR	F409L
Кистозный фиброз	CFTR	F433L
Кистозный фиброз	CFTR	H667R
Кистозный фиброз	CFTR	R1070W
Кистозный фиброз	CFTR	R29K
Кистозный фиброз	CFTR	R553Q
Кистозный фиброз	CFTR	1539T
Кистозный фиброз	CFTR	G550E
Кистозный фиброз	CFTR	F429S
Кистозный фиброз	CFTR	Q637R
Серповидно-клеточная анемия	HBB	A70T

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Серповидно-клеточная анемия	HBB	A70V
Серповидно-клеточная анемия	HBB	L88P
Серповидно-клеточная анемия	HBB	F85L и/или F85P
Серповидно-клеточная анемия	HBB	E22G
Серповидно-клеточная анемия	HBB	G16D и/или G16N

[#]: См. J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, включенный в данный документ посредством ссылки во всей своей полноте, в отношении подробностей номенклатуры генных мутаций.

35. *Регуляторные редактирования*

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для введения регуляторного редактирования. В некоторых вариантах осуществления регуляторное редактирование вводится в регуляторную последовательность гена, например, в промотор гена, энхансер гена, репрессор гена или последовательность, которая регулирует сплайсинг гена. В некоторых вариантах осуществления регуляторное редактирование повышает или снижает уровень экспрессии гена-мишени. В некоторых вариантах осуществления ген-мишень является таким же, как ген, содержащий мутацию, вызывающую заболевание. В некоторых вариантах осуществления ген-мишень отличается от гена, содержащего мутацию, вызывающую заболевание. Например, системы или способы, предусмотренные в данном документе, могут применяться для повышения экспрессии фетального гемоглобина посредством внесения регуляторного редактирования в промотор bcl11a, за счет чего обеспечивается лечение серповидно-клеточной анемии.

В **таблице 25** приведены иллюстративные показания (столбец 1), гены (столбец 2) и регуляторные редактирования, которые можно ввести с помощью систем или способов, описанных в данном документе (столбец 3).

36. Таблица 25. Показания, гены и компенсаторно-регуляторные редактирования

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Гомозиготная семейная гиперхолестеринемия	LDLR	c.81C>T
Формы порфирии	ALAS1	c.3G>A
Формы порфирии	ALAS1	c.2T>C
Формы порфирии	ALAS1	c.46C>T
Формы порфирии	ALAS1	c.91C>T
Формы порфирии	ALAS1	c.91C>T
Формы порфирии	ALAS1	c.226C>T
Формы порфирии	ALAS1	c.226C>T

Заболевание	**Ген**	**Изменение нуклеотида#**
Формы порфирии	ALAS1	c.226C>T
Формы порфирии	ALAS1	c.229C>T
Формы порфирии	ALAS1	c.247C>T
Формы порфирии	ALAS1	c.247C>T
Формы порфирии	ALAS1	c.250C>T
Формы порфирии	ALAS1	c.250C>T
Формы порфирии	ALAS1	c.340C>T
Формы порфирии	ALAS1	c.340C>T
Формы порфирии	ALAS1	c.349C>T
Формы порфирии	ALAS1	c.391C>T
Формы порфирии	ALAS1	c.391C>T
Формы порфирии	ALAS1	c.403C>T
Формы порфирии	ALAS1	c.403C>T
Формы порфирии	ALAS1	c.199+1G>A
Формы порфирии	ALAS1	c.199+1G>A
Формы порфирии	ALAS1	c.199+1G>A
Формы порфирии	ALAS1	c.199+1G>A
Формы порфирии	ALAS1	c.199+2T>C
Формы порфирии	ALAS1	c.199+2T>C
Формы порфирии	ALAS1	c.199+2T>C
Формы порфирии	ALAS1	c.199+2T>C
Формы порфирии	ALAS1	c.200-2A>G
Формы порфирии	ALAS1	c.427+1G>A
Формы порфирии	ALAS1	c.427+2T>C
Формы порфирии	ALAS1	c.1165+1G>A
Формы порфирии	ALAS1	c.1165+2T>C
Формы порфирии	ALAS1	c.1166-1A>G
Формы порфирии	ALAS1	c.1331-2A>G
Серповидно-клеточная анемия	BCL11A	c.386-24278G>A
Серповидно-клеточная анемия	BCL11A	c.386-24983T>C
Серповидно-клеточная анемия	HBG1	c.-167C>T

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Серповидно-клеточная анемия	HBG1	c.-170G>A
Серповидно-клеточная анемия	HBG1	c.-249C>T
Серповидно-клеточная анемия	HBG2	c.-211C>T
Серповидно-клеточная анемия	HBG2	c.-228T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -198 T>C
Серповидно-клеточная анемия	HBG1/2	C. -175 T>C
Серповидно-клеточная анемия	HBG1/2	C. -175 T>C
Серповидно-клеточная анемия	HBG1/2	C. -175 T>C
Серповидно-клеточная анемия	HBG1/2	C. -175 T>C
Серповидно-клеточная анемия	HBG1/2	C. -175 T>C
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	c. -90 связывание с BCLIIA
Серповидно-клеточная анемия	HBG1/2	c. -90 связывание с BCLIIA

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Серповидно-клеточная анемия	HBG1/2	C. -202 C>T, -201 C>T, -198 T>C, -197 C>T, -196 C>T, -195 C>G
Серповидно-клеточная анемия	HBG1/2	C. -197 C>T,-196 C>T, -195 C>G

[#]: См. J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, включенный в данный документ посредством ссылки во всей своей полноте, в отношении подробностей номенклатуры генных мутаций.

37. *Болезни экспансии повторов*

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, могут применяться для лечения болезни экспансии повторов, например, болезни экспансии повторов, представленной в **таблице 26**. В таблице 26 приведены показания (столбец 1), ген (столбец 2), минимальная повторяющаяся последовательность повтора, экспансия которого наблюдается при данном состоянии (столбец 3), и местоположение повтора относительно указанного гена для каждого показания (столбец 4). В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, например, включающие в себя Gene Writer, могут применяться для лечения болезней экспансии повторов посредством сброса количества повторов в локусе в соответствии с индивидуально адаптированной РНК-матрицей (см., например, пример 24).

38. Таблица 26. Иллюстративные болезни экспансии повторов, гены, причинные повторы и местоположения повторов

Заболевание	**Ген**	**Причинный повтор**	**Местоположение повтора**
миотоническая дистрофия 1 типа	DMPK/DM1	CTG	3'-UTR
миотоническая дистрофия 2 типа	ZNF9/CNBP	CCTG	Интрон 1
Дентаторубрально-паллидолуйзийская атрофия	ATN1	CAG	Кодирующая область
Синдром ломкой X-хромосомы с умственной отсталостью	FMR1	CGG	5'-UTR
Синдром ломкой X-хромосомы с умственной отсталостью	FMR2	GCC	5'-UTR
Атаксия Фридрейха	FXN	GAA	Интрон
Синдром ломкой X-хромосомы с тремором/атаксией	FMR1	CGG	5'-UTR

Болезнь Хантингтона	HTT	CAG	Кодирующая область
Болезнь 2, подобная болезни Хантингтона	JPH3	CTG	3'-UTR, кодирующая область
Миоклоническая эпилепсия Унферрихта-Лундборга	CSTB	CCCCGCCCCGCG	Промотор
Окулофарингеальная мышечная дистрофия	PABPN1	GCG	Кодирующая область
Спинальная и бульбарная мышечная атрофия	AR	CAG	Кодирующая область
Спиноцеребеллярная атаксия 1	ATXN1	CAG	Кодирующая область
Спиноцеребеллярная атаксия 2	ATXN2	CAG	Кодирующая область
Спиноцеребеллярная атаксия 3	ATXN3	CAG	Кодирующая область
Спиноцеребеллярная атаксия 6	CACNA1A	CAG	Кодирующая область
Спиноцеребеллярная атаксия 7	ATXN7	CAG	Кодирующая область
Спиноцеребеллярная атаксия 8	ATXN8	CTG/CAG	CTG/CAG (ATXN8)
Спиноцеребеллярная атаксия 10	ATXN10	ATTCT	Интрон
Спиноцеребеллярная атаксия 12	PPP2R2B	CAG	Промотор, 5'-UTR?
Спиноцеребеллярная атаксия 17	TBP	CAG	Кодирующая область
Синдромальная/несиндромальная Х-сцепленная умственная отсталость	ARX	GCG	Кодирующая область

39. *Иллюстративные матрицы*

В некоторых вариантах осуществления в системах или способах, предусмотренных в данном документе, используются последовательности матрицы, указанные в **таблице 27**. В таблице 27 представлены иллюстративные последовательности РНК-матрицы (столбец

5) и необязательные последовательности gRNA для второго однонитевого разрыва (столбец 6), предназначенные для спаривания с полипептидом Gene Writing для коррекции указанных патогенных мутаций (столбец 4). Все матрицы в таблице 27 предназначены для иллюстрации общей последовательности из: (1) gRNA, нацеливающейся на однонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по однонитевому разрыву первой нити.

40. Таблица 27. Иллюстративные заболевания, ткани, гены, патогенные мутации, последовательности РНК-матрицы и последовательности gRNA для второго однонитевого разрыва

Заболевание	**Ткань**	**Ген**	**Мутация**	**РНК-матрица**	**gRNA для второго однонитевого разрыва**
Альфа-1-антитрипсин	Печень	SERPINA1	PiZ	TCCCCTCCAGGCCGTGCATAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTcGTCGATGGTCAGCACAGCCTTATGCACGGCCTGGA	TTTGTTGAACTTGACCTCGG
Кистозный фиброз	Легкое	CFTR	Дельта-F508	ACCATTAAAGAAAATATCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACCAaagATGATATTTTCTTTA	AaagATGATATTTTCTTTAA
Серповидно-клеточная анемия	HSC	HBB	HbS	GTAACGGCAGACTTCTCCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGACTCCTGaGGAGAAGTCTGCC	TGGTGAGGCCCTGGGCAGGT

Болезнь Вильсона	Печень	ATP7B	H1069Q	TTGGTGACTGCCACGCCCAAGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAcCCCTTGGGCGTGGCAGTC	GGCCAGCAGTGAACAcCCCT
ARVC	Сердце	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTTCACtGAACCAGCAGTGTTCC	TTGGTTGAAAATGATTTTGT
Синдром удлиненного интервала QT 1 типа	Сердце	KCNQ1	P343S	CCAGGGAAAACGCACCCACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCcCAGCGGTAGGTGCCCCGTGGGTGCGTTTTC	CTCCTTCTTTGCGCTCcCAG
Муколипидоз IV типа	ЦНС	MCOLN1	406-2A>G	GCCCTCCCCTTCTCTGCCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGGTACtGTGGGCAGAGAAGGGG	TCAGGCAACGCCAGGTACtG

В некоторых вариантах осуществления в системах или способах, предусмотренных в данном документе, используются последовательности матрицы, указанные в **таблице 35**. В таблице 35 представлены иллюстративные последовательности РНК-матрицы (столбец 5) и необязательные последовательности gRNA для второго однонитевого разрыва (столбец 6), предназначенные для спаривания с полипептидом Gene Writing для коррекции указанных патогенных мутаций (столбец 4). Все матрицы в таблице 35 предназначены для иллюстрации общей последовательности из: (1) gRNA, нацеливающейся на однонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по однонитевому разрыву первой нити.

41. Таблица 35. Иллюстративные матрицы для Gene Writing и последовательностей gRNA для второго однонитевого разрыва для коррекции иллюстративных болезней экспансии повторов. Область матрицы, охватывающая повтор(повторы), указана строчными буквами

Заболевание	**Ген**	**Ссылочный номер доступа**	**Повтор**	**Местоположение**	**РНК-матрица**	**gRNA для второго однонитевого разрыва**
миотоническая дистрофия 1 типа	DMPK	NC_000019.10 (45769709..45782490, комплементарная нить)	CTG	3'-UTR	CTCGAAGG GTCCTTGTA GCCGTTTTA GAGCTAGA AATAGCAA GTTAAAAT AAGGCTAG TCCGTTATC AACTTGAA AAAGTGGG ACCGAGTC GGTCCGTG ATCCCCCca gcagcagcagca gcagcagcagca gcagcagcagca gcagcagcagca gcagcagcagC ATTCCCGG CTACAAGG ACCCT	ATCACAGG ACTGGAGC TGGG

миотоническая дистрофия 2 типа	CNBP	NC_000003.12 (129167827..129183896, комплементарная нить)	CCTG	Интрон 1	ACCACTGC ACTCCAGC CTAGGTTT AGAGCTAG AAATAGCA AGTTAAAA TAAGGCTA GTCCGTTAT CAACTTGA AAAAGTGG GACCGAGT CGGTCCGT GTCTGTCTG TCTGTCTGT CTGTCTGTC TGTCTGTCT GTCTGcctgcc tgcctgcctgcctg cctgcctgcctggc tgcctgtctgcctgt ctgcctgcctgcct gcctgcctgcctgc ctgTCTGTCT CACTTTGTC CCCTAGGC TGGAGTGC A	GCCTCAGC CTCCTGAG TAGC

Синдром ломкой X-хромосомы с умственной отсталостью	FMR1	NC_000023.11 (147911919..147951127)	CGG	5'-UTR	GGGGGCGTGCGGCAGCGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTGCGGGCGCTCGAGGCCCAGccgccgccgccgccgccgcctccgccgccgccgccgccgccgccgccgCGCTGCCGCACG	GCTCAGAGGCGGCCCTCCAC
Атаксия Фридрейха	FXN	NC_000009.12 (69035752..69079076)	GAA	Интрон	CAGGCGCGCGACACCACGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAACCCAGTATCTACTAAA	CGCTTGAGCCCGGGAGGCAG

					AAATACAA AAAAAAAA AAAAAAgaa gaagaagaagaa gaaAATAAA GAAAAGTT AGCCGGGC GTGGTGTC GCGC	
Болезнь Хантингтона	HTT	NC_000004.12 (3074681..3243960)	CAG	Кодирующая область	GGCGGCTG AGGAAGCT GAGGGTTT TAGAGCTA GAAATAGC AAGTTAAA ATAAGGCT AGTCCGTT ATCAACTT GAAAAAGT GGGACCGA GTCGGTCC AGTCCCTC AAGTCCTTC cagcagcagcagc agcagcagcagca gcagcagcagca gcagcagcagca gcagcagcaacag ccgccaccgccgc cgccgccgccgcc gcctcctCAGC TTCCTCAG	CGCTGCAC CGACCGTG AGTT

Спиноцеребеллярная атаксия 1	ATXN1	NC_000006.12 (16299112..16761490, комплементарная нить)	CAG	Кодирующая область	TGAGCCCC GGAGCCCT GCTGGTTT AGAGCTAG AAATAGCA AGTTAAAA TAAGGCTA GTCCGTTAT CAACTTGA AAAAGTGG GACCGAGT CGGTCCAC AAGGCTGA Gcagcagcagca gcagcagcagca gcagcagcagca gcatcagcatcag cagcagcagcagc agcagcagcagca gcagcagcagca gCACCTCAG CAGGGCTC CGGG	TCCAGTTC TCCGCAGA ACAC

42. *Иллюстративные последовательности гетерологичного объекта*

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, содержат последовательность гетерологичного объекта, где последовательность гетерологичного объекта или последовательность, обратно комплементарная ей, кодирует белок (например, антитело) или пептид. В некоторых вариантах осуществления лечение одобрено регулирующим органом, таким как FDA.

В некоторых вариантах осуществления белок или пептид представляет собой белок или пептид из базы данных THPdb (Usmani et al. PLoS One 12(7):e0181748 (2017), включенной в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления белок или пептид представляет собой белок или пептид, раскрытый в **таблице 28**. В некоторых вариантах осуществления системы или способы, раскрытые в данном документе, например, предусматривающие Gene Writer, могут применяться для интеграции кассеты экспрессии белка или пептида из таблицы 28 в клетку-

хозяина для обеспечения экспрессии белка или пептида у хозяина. В некоторых вариантах осуществления последовательности белка или пептида в первом столбце таблицы 28 можно найти в патентах или заявках, представленных в третьем столбце таблицы 28, которые включены в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления белок или пептид представляет собой антитело, раскрытое в таблице 1 из Lu et al. J Biomed Sci 27(1):1 (2020), включенного в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления белок или пептид представляет собой антитело, раскрытое в **таблице 29**. В некоторых вариантах осуществления системы или способы, раскрытые в данном документе, например, предусматривающие Gene Writer, могут применяться для интеграции кассеты экспрессии антитела из таблицы 29 в клетку-хозяина для обеспечения экспрессии антитела в хозяине. В некоторых вариантах осуществления система или способ, описанные в данном документе, применяются для экспрессии средства, которое связывает мишень из столбца 2 таблицы 29 (например, моноклональное антитело из столбца 1 таблицы 29), у субъекта, имеющего показания, указанные в столбце 3 таблицы 29.

43. Таблица 28. Иллюстративные белковые и пептидные терапевтические средства

Терапевтический пептид	**Категория**	**Номер патента**
Лепирудин	Антитромбины и фибринолитические средства	CA1339104
Цетуксимаб	Антинеопластические средства	CA1340417
Дорназа альфа	Ферменты	CA2184581
Денилейкин дифтитокс	Антинеопластические средства	
Этанерцепт	Иммунодепрессивные средства	CA2476934
Бивалирудин	Антитромбины	US7582727
Лейпролид	Антинеопластические средства	
Пегилированный интерферон альфа-2a	Иммунодепрессивные средства	CA2203480
Альтеплаза	Тромболитические средства	
Интерферон альфа-n1	Противовирусные средства	
Дарбэпоэтин альфа	Антианемические средства	CA2165694
Ретеплаза	Фибринолитические средства	CA2107476

Эпоэтин альфа	Гематиники	CA1339047
Лососевый кальцитонин	Средства для сохранения плотности костей	US6440392
Интерферон альфа-n3	Иммунодепрессивные средства	
Пегфилграстим	Иммунодепрессивные средства	CA1341537
Сарграмостим	Иммунодепрессивные средства	CA1341150
Секретин	Средства диагностики	
Пегилированный интерферон альфа-2b	Иммунодепрессивные средства	CA1341567
Аспарагиназа	Антинеопластические средства	
Тиротропин альфа	Средства диагностики	US5840566
Антигемофильный фактор	Коагулянты и тромботические средства	CA2124690
Анакинра	Противоревматические средства	CA2141953
Грамицидин D	Антибактериальные средства	
Внутривенный иммуноглобулин	Иммунологические факторы	
Анистреплаза	Фибринолитические средства	
Инсулин регуляр	Противодиабетические средства	
Тенектеплаза	Фибринолитические средства	CA2129660
Менотропины	Средства для улучшения фертильности	
Интерферон гамма-1b	Иммунодепрессивные средства	US6936695
Интерферон альфа-2a, рекомбинантный		CA2172664
Коагуляционный фактор VIIa	Коагулянты	
Опрелвекин	Антинеопластические средства	
Палифермин	Средства против мукозита	

Глюкагон рекомбинантный	Гипогликемические средства	
Альдеслейкин	Антинеопластические средства	
Ботулотоксин типа В	Антидистонические средства	
Омализумаб	Противоаллергические средства	CA2113813
Лютропин альфа	Средства для улучшения фертильности	US5767251
Инсулин лиспро	Гипогликемические средства	US5474978
Инсулин гларгин	Гипогликемические средства	US7476652
Коллагеназа		
Расбуриказа	Супрессоры подагры	CA2175971
Адалимумаб	Противоревматические средства	CA2243459
Имиглюцераза	Ферментативные заместительные средства	US5549892
Абциксимаб	Антикоагулянты	CA1341357
Ингибитор альфа-1-протеиназы	Ингибиторы сериновых протеиназ	
Пегаспаргаза	Антинеопластические средства	
Интерферон бета-1а	Антинеопластические средства	CA1341604
Пегадемаза крупного рогатого скота	Ферментативные заместительные средства	
Сывороточный альбумин человека	Заменители сыворотки	US6723303
Эптифибатид	Ингибиторы агрегации тромбоцитов.	US6706681
Иодированный сывороточный альбумин	Средства диагностики	
инфликсимаб	Антиревматические средства, противовоспалительные средства, нестероидные, дерматологические	CA2106299

	средства, желудочно-кишечные средства и иммунодепрессивные средства	
Фоллитропин бета	Средства для улучшения фертильности	US7741268
Вазопрессин	Антидиуретические средства	
Интерферон бета-1b	Адъюванты, иммунологические и иммуносупрессивные средства	CA1340861
Интерферон альфакон-1	Противовирусные средства и иммунодепрессивные средства	CA1341567
Гиалуронидаза	Адъюванты, анестетики и пермеабилизирующие средства	
Инсулин свиной	Гипогликемические средства	
Трастузумаб	Антинеопластические средства	CA2103059
Ритуксимаб	Антинеопластические средства, иммунологические факторы и антиревматические средства	CA2149329
Базиликсимаб	Иммунодепрессивные средства	CA2038279
Muromo b	Иммунологические факторы и иммунодепрессивные средства	
Иммунный Fab для дигоксина (овечий)	Антидоты	
Ибритумомаб		CA2149329
Даптомицин		US6468967
Тозитумомаб		
Пегвисомант	Гормонозаместительные средства	US5849535
Ботулотоксин типа A	Нервно-мышечные блокирующие средства, средства против морщин и антидистонические средства	CA2280565
Панкрелипаза	Желудочно-кишечные средства и ферментативные заместительные средства	
Стрептокиназа	Фибринолитические средства и тромболитические средства	
Алемтузумаб		CA1339198

Альглюцераза	Ферментативные заместительные средства	
Капромаб	Индикаторы, реагенты и диагностические средства	
Ларонидаза	Ферментативные заместительные средства	
Урофоллитропин	Средства для улучшения фертильности	US5767067
Эфализумаб	Иммунодепрессивные средства	
Сывороточный альбумин	Заменители сыворотки	US6723303
Хориогонадотропин альфа	Средства фертильности и гонадотропины	US6706681
Антитимоцитарный глобулин	Иммунологические факторы и иммунодепрессивные средства	
Филграстим	Иммунодепрессивные средства, антинейтропенические средства и гемопоэтические средства	CA1341537
Фактор свертывания ix	Коагулянты и тромботические средства	
Бекаплермин	Средства, индуцирующие ангиогенез	CA1340846
Агалсидаза бета	Ферментативные заместительные средства	CA2265464
Интерферон альфа-2b	Иммунодепрессивные средства	CA1341567
Окситоцин	Утеротоники, антитоколитические средства и средства для индукции родов	
Энфувиртид	Ингибиторы слияния HIV	US6475491
Паливизумаб	Противовирусные средства	CA2197684
Даклизумаб	Иммунодепрессивные средства	
Бевацизумаб	Ингибиторы ангиогенеза	CA2286330
Арцитумомаб	Средства диагностики	US8420081
Арцитумомаб	Средства диагностики	US7790142
Экулизумаб		CA2189015
Панитумумаб		
Ранибизумаб	Офтальмологические средства	CA2286330
Идурсульфаза	Ферментативные заместительные средства	

Альглюкозидаза альфа	Ферментативные заместительные средства	CA2416492
Эксенатид	Гипогликемические средства	US6872700
Мекасермин		US5681814
Прамлинтид		US5686411
Галсульфаза	Ферментативные заместительные средства	
Абатацепт	Антиревматические средства и иммунодепрессивные средства	CA2110518
Косинтропин	Гормоны и диагностические агенты	
Кортикотропин		
Инсулин Аспарт	Гипогликемические средства и антидиабетические средства	US5866538
Инсулин детемир	Противодиабетические средства	US5750497
Инсулин глулизин	Противодиабетические средства	US6960561
Пегаптаниб	Предназначен для предотвращения респираторного дистресс-синдрома (RDS) у недоношенных детей с высоким риском развития RDS.	
Несиритид		
Тималальфазин		
Дефибротид	Антитромбины	
Натуральный интерферон альфа ИЛИ мультиферон		
Глатирамера ацетат		
Преотакт		
Тейкопланин	Антибактериальные средства	
Каникумаб	Противовоспалительные средства и моноклональные антитела	

Ипилимумаб	Антинеопластические средства и моноклональные антитела	CA2381770
Сулодексид	Антитромбины и фибринолитические средства, и гипогликемические средства, и антикоагулянты, и гиполипидемические средства	
Тоцилизумаб		CA2201781
Терипаратид	Средства для сохранения плотности костей	US6977077
Пертузумаб	Моноклональные антитела	CA2376596
Рилонацепт	Иммунодепрессивные средства	US5844099
Деносумаб	Средства для сохранения плотности костей и моноклональные антитела	CA2257247
Лираглутид		US6268343
Голимумаб	Антипсориатические средства и моноклональные антитела, и ингибитор TNF	
Белатацепт	Антиревматические средства и иммунодепрессивные средства	
Бусерелин		
Велаглюцераза альфа	Ферменты	US7138262
Тесаморелин		US5861379
Брентуксимаб ведотин		
Талиглюцераза альфа	Ферменты	
Белимумаб	Моноклональные антитела	
Афлиберцепт	Антинеопластические средства и офтальмологические средства	US7306799
Аспарагиназа Erwinia chrysanthemi	Ферменты	
Окриплазмин	Офтальмологические средства	
Глюкарпидаза	Ферменты	

Тедуглутид		US5789379
Раксибакумаб	Антиинфекционные средства и моноклональные антитела	
Цертолизумаб пегол	Ингибитор TNF	CA2380298
Инсулин-изофан	Гипогликемические средства и антидиабетические средства	
Эпоэтин дзета		
Обинутузумаб	Антинеопластические средства	
Фибринолизин, он же плазмин		US3234106
Фоллитропин альфа		
Ромиплостим	Колониестимулирующие факторы и тромбопоэтические средства	
Луцинактант	Легочные сурфактанты	US5407914
Тализумаб	Иммунодепрессивные средства	
Алискирен	Ингибитор ренина	
Экстракт пыльцы амброзии		
Секукинумаб	Ингибитор	US20130202610
Соматотропин рекомбинантный	Гормонозаместительные средства	CA1326439
Дротрекогин альфа	Антисептические средства	CA2036894
Алефацепт	Дерматологические и иммунодепрессивные средства	
Липопротеин OspA	Вакцины	
Урокназа		US4258030
Абареликс	Антитестостероновые средства	US5968895
Серморелин	Гормонозаместительные средства	
Апротинин		US5198534

Гемтузумаб озогамицин	Антинеопластические средства и иммунотоксины	US5585089
Сатумомаб Пендетид	Средства диагностики	
Альбиглутид	Лекарственные средства, применяемые при сахарном диабете; пищеварительный тракт и обмен веществ; лекарственные средства, снижающие уровень глюкозы в крови, помимо инсулина.	
Алирокумаб		
Анцестим		
Антитромбин альфа		
Антитромбин III человеческий		
Асфотаза альфа	Ферменты пищеварительного тракта и обмена веществ	
Атезолизумаб		
Культивированные аутологичные хондроциты		
Беректант		
Блинатумомаб	Антинеопластические средства, иммунодепрессивные средства, моноклональные антитела, антинеопластические средства и иммуномодулирующие средства	US20120328618
Ингибитор эстеразы C1 (человека)		
А-субъединица фактора свертывания крови XIII		

(рекомбинантная)		
Конестат альфа		
Даратумумаб	Антинеопластические средства	
Дезирудин		
Дулаглутид	Гипогликемические средства; лекарственные средства, применяемые при сахарном диабете; пищеварительный тракт и обмен веществ; лекарственные средства, снижающие уровень глюкозы в крови, помимо инсулина	
Элосульфаза альфа	Ферменты пищеварительного тракта и обмена веществ	
Элотузумаб		US2014055370
Эволокумаб	Средства, модифицирующие липиды, простые; Сердечно-сосудистая система	
Концентрат фибриногена (человека)		
Филграстим-sndz		
Внутренний фактор желудка		
Иммуноглобулин против гепатита В		
Кальцитонин человека		
Человеческий противостолбнячный иммуноглобулин		
Человеческий иммуноглобулин		

к вирусу бешенства		
Человеческий иммуноглобулин Rho(D)		
Гиалуронидаза (человеческая рекомбинантная)		US7767429
Идаруцизумаб	Антикоагулянт	
Человеческий иммуноглобулин	Иммунологические факторы; иммунодепрессивные средства; Противоинфекционные средства	
Ведолизумаб	Иммунодепрессивное средство; антинеопластическое средство	US2012151248
Устекинумаб	Дерматологическое средство, иммунодепрессивное средство; антинеопластическое средство	
Туроктоког альфа		
Очищенное туберкулиновое производное белка		
Симоктоког альфа	Антигеморрагические средства: фактор свертывания крови VIII	
Силтуксимаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства	US7612182
Себелипаза альфа	Ферменты	
Сакрозидаза	Ферменты	
Рамуцирумаб	Антинеопластические средства и иммуномодулирующие средства	US2013067098

Концентрат протромбинового комплекса		
Поракtан альфа	Легочные сурфактанты	
Пембролизумаб	Антинеопластические средства и иммуномодулирующие средства	US2012135408
Пегинтерферон бета-1а		
Офатумумаб	Антинеопластические средства и иммуномодулирующие средства	US8337847
Обилтоксаксимаб		
Ниволумаб	Антинеопластические средства и иммуномодулирующие средства	US2013173223
Нецитумумаб		
Метрелептин		US20070099836
Метоксиполиэтиленгликоль-эпоэтин бета		
Меполизумаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства, ингибиторы интерлейкина	US2008134721
Иксекизумаб		
Инсулин свиной	Гипогликемические средства, антидиабетические средства	
Инсулин Деглудек		
Инсулин из говядины		
Тиреоглобулин	Гормональная терапия	US5099001
Человеческий противосибиреяз	Производное плазмы крови	

венный иммуноглобулин		
Антиингибиторный коагулянтный комплекс	Факторы свертывания крови, антигемофильное средство	
Антитимоцитарный глобулин (лошадиный)	Антитело	
Антитимоцитарный глобулин (кролика)	Антитело	
Бродалумаб	Антинеопластические средства и иммуномодулирующие средства	
Ингибитор эстеразы С1 (рекомбинантный)	Кровь и органы кроветворения	
Каникумаб	Антинеопластические средства и иммуномодулирующие средства	
Хорионический гонадотропин (человека)	Гормоны	US6706681
Хорионический гонадотропин (Рекомбинантный)	Гормоны	US5767251
Человеческий фактор свертывания крови X	Факторы свертывания крови	
Динутуксимаб	Антитело, иммунодепрессивное средство, антинеопластическое средство	US20140170155

Эфмороктоког альфа	Антигемофильный фактор	
Комплекс фактора IX (человека)	Антигемофильное средство	
Вакцина к гепатиту А	Вакцина	
Человеческий иммуноглобулин против вируса ветряной оспы	Антитело	
Ибритумомаб тиуксетан	Антитело, иммунодепрессивные средства	CA2149329
Ленограстим	Антинеопластические средства и иммуномодулирующие средства	
Пеглотиказа	Ферменты	
Протамина сульфат	Антагонисты гепарина, гематологические средства	
Белок S человека	Антикоагулянтный белок плазмы	
Сипулеуцел-Т	Антинеопластические средства и иммуномодулирующие средства	US8153120
Соматропин рекомбинантный	Гормоны, гормонозаместители и антагонисты гормонов	CA1326439, CA2252535, US5288703, US5849700, US5849704, US5898030, US6004297, US6152897, US6235004, US6899699
Сусоктоког альфа	Факторы свертывания крови, Антигеморрагические средства	

Тромбомодулин альфа	Антикоагулянтное средство, антиагрегантое средство	

44. Таблица 29. Иллюстративные средства терапии на основе моноклональных антител

mAb	**Мишень**	**Показание**
Муромонаб-CD3	CD3	Отторжение почечного трансплантата
Абциксимаб	GPIIb/IIIa	Предупреждение образования тромбов при ангиопластике
Ритуксимаб	CD20	Неходжкинская лимфома
Паливизумаб	RSV	Предупреждение респираторно-синцитиальной вирусной инфекции
Инфликсимаб	TNFα	Болезнь Крона
Трастузумаб	HER2	Рак молочной железы
Алемтузумаб	CD52	Хронический миелоидный лейкоз
Адалимумаб	TNFα	Ревматоидный артрит
Ибритумомаб тиуксетан	CD20	Неходжкинская лимфома
Омализумаб	IgE	Астма
Цетуксимаб	EGFR	Колоректальный рак
Бевацизумаб	VEGF-A	Колоректальный рак
Натализумаб	ITGA4	Рассеянный склероз
Панитумумаб	EGFR	Колоректальный рак
Ранибизумаб	VEGF-A	Макулярная дегенерация
Экулизумаб	C5	Пароксизмальная ночная гемоглобинурия
Цертолизумаб пегол	TNFα	Болезнь Крона
Устекинумаб	IL-12/23	Псориаз
Канакинумаб	IL-1β	Синдром Макла-Уэльса
Голимумаб	TNFα	Ревматоидный и псориатический артрит, анкилозирующий спондилоартрит
Офатумумаб	CD20	Хронический лимфоцитарный лейкоз
Тоцилизумаб	IL-6R	Ревматоидный артрит
Деносумаб	RANKL	Потеря костной массы
Белимумаб	BLyS	Системная красная волчанка

Ипилимумаб	CTLA-4	Метастатическая меланома
Брентуксимаб ведотин	CD30	Лимфома Ходжкина, системная анапластическая крупноклеточная лимфома
Пертузумаб	HER2	Рак молочной железы
Трастузумаб эмтанзин	HER2	Рак молочной железы
Раксибакумаб	PA *B. anthracis*	Сибиреязвенная инфекция
Обинутузумаб	CD20	Хронический лимфоцитарный лейкоз
Силтуксимаб	IL-6	Болезнь Кастлемана
Рамуцирумаб	VEGFR2	Рак желудка
Ведолизумаб	Интегрин α4β7	Неспецифический язвенный колит, болезнь Крона
Блинатумомаб	CD19, CD3	Острый лимфобластный лейкоз
Ниволумаб	PD-1	Меланома, немелкоклеточный рак легкого
Пембролизумаб	PD-1	Меланома
Идаруцизумаб	Дабигатран	Устранение дабигатран-индуцированной антикоагуляции
Нецитумумаб	EGFR	Немелкоклеточный рак легкого
Динутуксимаб	GD2	Нейробластома
Секукинумаб	IL-17α	Псориаз
Меполизумаб	IL-5	Тяжелая эозинофильная астма
Алирокумаб	PCSK9	Высокое содержание холестерина
Эволокумаб	PCSK9	Высокое содержание холестерина
Даратумумаб	CD38	Множественная миелома
Элотузумаб	SLAMF7	Множественная миелома
Иксекизумаб	IL-17α	Псориаз
Реслизумаб	IL-5	Астма
Оларатумаб	PDGFRα	Саркома мягких тканей
Безлотоксумаб	Энтеротоксин B *Clostridium difficile*	Предупреждение рецидива инфекции *Clostridium difficile*
Атезолизумаб	PD-L1	Рак мочевого пузыря

Обилтоксаксимаб	PA *B. anthracis*	Предупреждение ингаляционной формы сибирской язвы
Инотузумаб озогамицин	CD22	Острый лимфобластный лейкоз
Бродалумаб	IL-17R	Бляшечный псориаз
Гуселькумаб	IL-23, p19	Бляшечный псориаз
Дупилумаб	IL-4Rα	Атопический дерматит
Сарилумаб	IL-6R	Ревматоидный артрит
Авелумаб	PD-L1	Карцинома из клеток Меркеля
Окрелизумаб	CD20	Рассеянный склероз
Эмицизумаб	Фактор IXa, X	Гемофилия A
Бенрализумаб	IL-5Rα	Астма
Гемтузумаб озогамицин	CD33	Острый миелоидный лейкоз
Дурвалумаб	PD-L1	Рак мочевого пузыря
Буросумаб	FGF23	X-сцепленная гипофосфатемия
Ланаделумаб	Плазменный калликреин	Приступы наследственного ангионевротического отека
Могамулизумаб	CCR4	Грибовидный микоз или синдром Сезари
Эренумаб	CGRPR	Предупреждение мигрени
Галканезумаб	CGRP	Предупреждение мигрени
Тилдракизумаб	IL-23, p19	Бляшечный псориаз
Цемиплимаб	PD-1	Плоскоклеточный рак кожи
Эмапалумаб	IFNγ	Первичный гемофагоцитарный лимфогистиоцитоз
Фреманезумаб	CGRP	Предупреждение мигрени
Ибализумаб	CD4	HIV-инфекция
Моксетумомаб пасудодокс	CD22	Волосатоклеточный лейкоз
Равулизумаб	C5	Пароксизмальная ночная гемоглобинурия
Каплацизумаб	фактор фон Виллебранда	Приобретенная тромботическая тромбоцитопеническая пурпура
Ромосозумаб	Склеростин	Остеопороз у женщин в постменопаузе с повышенным риском переломов
Рисанкизумаб	IL-23, p19	Бляшечный псориаз

Полатузумаб ведотин	CD79β	Диффузная крупноклеточная В-клеточная лимфома
Бролуцизумаб	VEGF-A	Макулярная дегенерация
Кризанлизумаб	P-селектин	Серповидно-клеточная анемия

45. *Способы модификации растений*

46. Описанные в данном документе системы Gene Writer можно применять для модификации растения или частей растения (например, листьев, корней, цветков, плодов или семян), например, для увеличения приспособленности растения.

47. A. Доставка в растение

48. В данном документе предусмотрены способы доставки системы Gene Writer, описанной в данном документе, в растение. Включены способы доставки системы Gene Writer в растение путем приведения растения или его части в контакт с системой Gene Writer. Способы применимы для модификации растения, например, для увеличения его приспособленности.

Более конкретно, в некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, нуклеиновая кислота, кодирующая Gene Writer), может быть закодирована в векторе, например, встроена рядом с растительным промотором, например, промотором гена убиквитина кукурузы (ZmUBI) в векторе для растения (например, pHUC411). В некоторых вариантах осуществления нуклеиновые кислоты, описанные в данном документе, вводят в растение (например, рис подвида *japonica*) или в часть растения (например, в каллюс растения) посредством агробактерий. В некоторых вариантах осуществления системы и способы, описанные в данном документе, могут применяться в растениях путем замены гена растения (например, гена гигромицинфосфотрансферазы (HPT)) нулевым аллелем (например, содержащим замену основания в старт-кодоне). Системы и способы модификации генома растения описаны в Xu et. al. *Development of plant prime-editing systems for precise genome editing*, 2020, Plant Communications.

49. В одном аспекте в данном документе представлен способ повышения приспособленности растения, где способ включает доставку в растение системы Gene Writer, описанной в данном документе (например, в эффективном количестве и в течение эффективного периода времени), для повышения приспособленности растения по сравнению с необработанным растением (например, растением, в которое не доставили систему Gene Writer).

50. Повышение приспособленности растения как следствие доставки системы Gene Writer может проявляться многими путями, например, приводя тем самым к более высокой производительности растения, например, улучшенной урожайности, улучшенной жизнеспособности растения или качества продукта, собираемого с растения, улучшению признаков, имеющихся до или после сбора урожая, которые считаются необходимыми для сельского хозяйства или садоводства (например, вкуса, внешнего вида, срока хранения),

или улучшению признаков, которые иным образом приносят пользу людям (например, понижение продуцирования аллергенов). Улучшение урожайности растения относится к повышению выхода продукта (например, определяемого по выходу биомассы, зерна, семян или плодов растения, содержанию белка, углеводов или масел или площади листьев) растения на измеримую величину по сравнению с выходом того же продукта растения, полученного в тех же условиях, но без применения композиций по настоящему изобретению, или по сравнению с применением обычных средств для модификации растений. Например, урожайность может быть повышена на по меньшей мере приблизительно 0,5%, приблизительно 1%, приблизительно 2%, приблизительно 3%, приблизительно 4%, приблизительно 5%, приблизительно 10%, приблизительно 20%, приблизительно 30%, приблизительно 40%, приблизительно 50%, приблизительно 60%, приблизительно 70%, приблизительно 80%, приблизительно 90%, приблизительно 100% или более 100%. В некоторых случаях способ является эффективным для повышения урожайности в приблизительно 2 раза, 5 раз, 10 раз, 25 раз, 50 раз, 75 раз, 100 раз или более чем 100 раз по сравнению с необработанным растением. Урожайность может быть представлена в виде количества по весу или объему растения или продукта растения относительно некоторого параметра. Такой параметр может быть представлен в виде времени, площади выращивания, веса полученных растений или количества используемого сырья. Например, такие способы могут обеспечивать повышение урожая тканей растения, включая без ограничения семена, плоды, ядра, семенные коробочки, клубни, корни и листья.

51. Повышение жизнеспособности растения как следствие доставки системы Gene Writer также можно измерять другими способами, как, например, по повышению или улучшению показателя жизнеспособности, густоты (количества растений на единицу площади), высоты растения, окружности стебля, длины стебля, количества листьев, размера листа, листового полога, внешнего вида (например, более зеленый цвет листьев), показателя корней, всхожести, содержания белка, повышенному кущению, более крупным листьям, большему количеству листьев, меньшему количеству мертвых прикорневых листьев, более сильным росткам, потребности в меньшем количестве удобрений, потребности в меньшем количестве семян, более продуктивным росткам, более раннему цветению, более раннему созреванию зерна или семян, меньшему полеганию растений (вылеганию), усиленному росту побегов, более раннему прорастанию или любой комбинации этих факторов на измеримую или заметную величину по сравнению с тем же фактором растения, производимого в тех же условиях, но без введения композиций по настоящему изобретению или с применением общепринятых средств для модификации растений (например, средств для модификации растений, доставляемых без использования PMP).

52. Соответственно, в данном документе представлен способ модификации растения, при этом способ включает доставку в растение эффективного количества любой из систем Gene Writer, предусмотренной в данном документе, где с помощью данного

способа осуществляют модификацию растения и за счет этого осуществляют введение полезного признака в растение или повышение его уровня (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением. В частности, способ может обеспечивать повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

53. В некоторых случаях повышение приспособленности растения представляет собой повышение (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) устойчивости к болезням, засуховыносливости, жаровыносливости, холодовыносливости, солевыносливости, выносливости к металлам, выносливости к гербицидам, выносливости к химическим веществам, эффективности использования воды, использования азота, устойчивости к азотному стрессу, фиксации азота, устойчивости к вредителям, устойчивости к травоядным животным, устойчивости к патогенам, урожайности, урожайности в условиях ограниченного количества воды, жизнеспособности, роста, фотосинтетического потенциала, питания, содержания белка, содержания углеводов, содержания масла, биомассы, длины побегов, длины корней, показателя структуры корневой системы, веса семян или количества собираемого урожая.

54. В некоторых случаях повышение приспособленности представляет собой повышение (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) скорости развития, роста, урожайности, устойчивости к абиотическим стрессовым факторам или устойчивости к биотическим стрессовым факторам. Абиотический стресс относится к стрессовому воздействию окружающей среды, которому подвергается растение или часть растения, которое включает, например, стресс, вызываемый засухой, солевой стресс, тепловой стресс, холодовой стресс и стресс, вызываемый низким содержанием питательных веществ. Биотический стресс относится к стрессовому воздействию окружающей среды, которому подвергается растение или часть растения, которое включает, например, стресс, вызываемый нематодами, стресс, вызываемый травоядными насекомыми, стресс, вызываемый грибными патогенами, стресс, вызываемый бактериальными патогенами, или стресс, вызываемый вирусными патогенами. Стресс может быть временным, например, в течение нескольких часов, нескольких дней, нескольких месяцев, или постоянным, например, в течение жизни растения.

55. В некоторых случаях повышение приспособленности растения представляет собой повышение (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) качества продуктов, собираемых с растения. Например, повышение приспособленности растения может представлять собой улучшение в отношении коммерчески благоприятных признаков (например, вкуса или внешнего вида) продукта, собираемого с растения. В других случаях повышение приспособленности

растения представляет собой повышение срока хранения продукта, собираемого с растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%).

56. В качестве альтернативы повышение приспособленности может представлять собой изменение признака, который полезен для здоровья человека или животного, как, например, снижение продуцирования аллергена. Например, повышение приспособленности может представлять собой снижение (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) продуцирования аллергена (например, пыльцы), который стимулирует иммунный ответ у животного (например, человека).

57. Модификация растения (например, повышение приспособленности) может возникать в результате модификации одной или нескольких частей растения. Например, растение можно модифицировать посредством приведения в контакт с листом, семенем, пыльцой, корнем, плодом, побегом, цветком, клетками, протопластами или тканью (например, меристематической тканью) растения. По этой причине в другом аспекте в данном документе представлен способ повышения приспособленности растения, при этом способ включает приведение пыльцы растения в контакт с эффективным количеством любой из композиций для модификации растений, представленных в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

58. В еще одном аспекте в данном документе представлен способ повышения приспособленности растения, при этом способ включает приведение семени растения в контакт с эффективным количеством любой из систем Gene Writer, раскрытых в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

59. В другом аспекте в данном документе представлен способ, включающий приведение протопласта растения в контакт с эффективным количеством любой из систем Gene Writer, описанных в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

60. В дополнительном аспекте в данном документе представлен способ повышения приспособленности растения, при этом способ включает приведение растительной клетки растения в контакт с эффективным количеством любой из систем Gene Writer, описанных в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

61. В другом аспекте в данном документе представлен способ повышения приспособленности растения, при этом способ включает приведение меристематической ткани растения в контакт с эффективным количеством любой из композиций для модификации растений, представленных в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

62. В другом аспекте в данном документе представлен способ повышения приспособленности растения, при этом способ включает приведение зародыша растения в контакт с эффективным количеством любой из композиций для модификации растений, представленных в данном документе, где способ обеспечивает повышение приспособленности растения (например, на приблизительно 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% или более 100%) по сравнению с необработанным растением.

63. В. Способы применения

64. Растение, описанное в данном документе, можно подвергать воздействию любой из композиций на основе системы Gene Writer, описанных в данном документе, с помощью любого подходящего способа, который обеспечивает доставку или введение композиции в растение. Систему Gene Writer можно доставлять либо отдельно, либо в комбинации с другими активными (например, удобрениями) или неактивными веществами и можно наносить, например, посредством распыления, инъекции (например, микроинъекции), через растения, посредством полива, погружения, в форме концентрированных жидкостей, гелей, растворов, суспензий, спреев, порошков, пеллет, брикетов, плиток и т. п., составленных для доставки эффективной концентрации композиции для модификации растений. Количества и места для применения композиций, описанных в данном документе, как правило, определяются условиями обитания растения, стадией жизненного цикла, на которой на растение можно целенаправленно воздействовать композицией для модификации растений, участком, в котором следует выполняться применение, а также физическими и функциональными характеристиками композиции для модификации растений.

65. В некоторых случаях композицию распыляют непосредственно на растение, например, сельскохозяйственные культуры, например, путем распыления из ранца, распыления с воздуха, опрыскивания/опыления сельскохозяйственных культур и т. д. В тех случаях, когда систему Gene Writer доставляют в растение, растение, получающее систему Gene Writer, может находиться на любой стадии роста растений. Например, составленные композиции для модификации растений можно применять в виде покрытия для семян или средства для обработки корней на ранних стадиях роста растения или в виде средства для полной обработки растения на более поздних стадиях цикла сельскохозяйственного растения. В некоторых случаях композицию для модификации растений можно применять по отношению к растению в виде средства для местного применения.

66. Кроме того, систему Gene Writer можно применять (например, в отношении почвы, в которой растет растение, или в отношении воды, которую используют для полива растения) в качестве системного средства, которое поглощается и распределяется по тканям растения. В некоторых случаях растения или употребляемые в пищу организмы могут быть генетически трансформированы для экспрессии системы Gene Writer.

67. Замедленного или непрерывного высвобождения также можно достигнуть за счет покрытия системы Gene Writer или композиции с композицией(композициями) для модификации растений растворимым или биоразлагаемым покровным слоем, таким так желатин, при этом данное покрытие растворяется или разрушается в среде его использования, что делает тогда композицию на основе системы Gene Writer для модификации растений доступной, или за счет распределения средства в растворимой или разрушаемой матрице. Таки средства/устройства с непрерывным высвобождением и/или распределением могут преимущественно использоваться для устойчивого поддержания эффективной концентрации одной или нескольких композиций для модификации растений, описанных в данном документе.

68. В некоторых случаях систему Gene Writer доставляют в часть растения, например, лист, семя, пыльцу, корень, плод, побег или цветок, или в ткань, клетку или протопласт. В некоторых случаях систему Gene Writer доставляют в клетку растения. В некоторых случаях систему Gene Writer доставляют в протопласт растения. В некоторых случаях систему Gene Writer доставляют в ткань растения. Например, композицию можно доставлять в меристематическую ткань растения (например, апикальную меристему, латеральную меристему или интеркалярную меристему). В некоторых случаях композицию доставляют в постоянную ткань растения (например, простые ткани (например, паренхиму, колленхиму или склеренхиму) или сложную постоянную ткань (например, ксилему или флоэму)). В некоторых случаях систему Gene Writer доставляют в зародыш растения.

69. С. Растения

70. Систему Gene Writer, описанную в данном документе, можно доставлять во множество растений или обрабатывать их ею. Растения, в которые может быть доставлена система Gene Writer (т. е. "подвергаемые обработке") в соответствии со способами по настоящему изобретению, предусматривают целые растения и их части, в том числе без ограничения органы/структуры вегетативных побегов (например, листья, стебли и клубни), корни, цветки и органы/структуры цветков (например, прицветники, чашелистики, лепестки, тычинки, плодолистики, пыльники и семяпочки), семена (в том числе зародыш, эндосперм, семядоли и оболочку семени) и плоды (зрелую завязь), ткань растения (например, сосудистую ткань, основную паренхиму и т. п.) и клетки (например, замыкающие клетки, яйцеклетки и т. п.) и их потомство. Дополнительно части растения могут относиться к частям растения, таким как побег, корень, стебель, семена, прилистники, листья, лепестки, цветки, семяпочки, прицветники, ветви, черешки, междоузлия, кора, опушение, ростки, корневища, вайи, пластинки, пыльца, тычинки и т. п.

71. Класс растений, которые можно обрабатывать способом, раскрытым в данном документе, включает класс высших и низших растений, в том числе покрытосеменные (однодольные и двудольные растения), голосеменные, папоротники, хвощи, псилофиты, ликофиты, мохообразные и водоросли (например, многоклеточные или одноклеточные водоросли). Растения, которые можно обрабатывать в соответствии со способами по настоящему изобретению, дополнительно включают любое сосудистое растение, например, однодольные, или двудольные, или голосеменные, в том числе без ограничения люцерну, яблоню, растение рода Arabidopsis, банан, ячмень, канолу, клещевину, цикорий, хризантему, клевер, какао, кофе, хлопчатник, семена хлопчатника, кукурузу, крамбе, клюкву, огурец, дендробиум, диоскорею, эвкалипт, овсяницу, лен, гладиолус, растения семейства Лилейные, семя льна, просо, дыню, горчицу, овес, масличную пальму, масличный рапс, папайю, арахис, ананас, декоративные растения, фасоль, картофель, рапс, рис, рожь, райграс, сафлор, кунжут, сорго, сою, сахарную свеклу, сахарный тростник, подсолнечник, клубнику, табак, томат, газонную траву, пшеницу и овощные культуры, такие как салат, сельдерей, брокколи, цветная капуста, тыквенные культуры; фруктовые и ореховые деревья, такие как яблоня, груша, персик, апельсин, грейпфрут, лимон, лайм, миндаль, пекан, грецкий орех, лещина; вьющиеся растения, такие как виноград (например, виноградник), киви, хмель; плодовые кустарники и колючие кустарники, такие как малина, ежевика, крыжовник; лесные деревья, такие как ясень, сосна, пихта, клен, дуб, каштан, тополь; такое как люцерна, канола, клещевина, кукуруза, хлопчатник, крамбе, лен, семя льна, горчицу, масличную пальму, масличный рапс, арахис, картофель, рис, сафлор, кунжут, сою, сахарную свеклу, подсолнечник, табак, томат или пшеницу. Растения, которые можно обрабатывать в соответствии со способами по настоящему изобретению, включают любое культурное растение, например, кормовую культуру, масличную культуру, зерновую культуру, плодовую культуру, овощную культуру, волокнистую культуру, пряную культуру, ореховую культуру, газонную культуру, сахарную культуру, культуру для производства напитков и лесную культуру. В определенных случаях культурное растение, которое обрабатывают в этом способе, представляет собой растение сои. В других определенных случаях культурное растение представляет собой пшеницу. В определенных случаях культурное растение представляет собой кукурузу. В определенных случаях культурное растение представляет собой хлопчатник. В определенных случаях культурное растение представляет собой люцерну. В определенных случаях культурное растение представляет собой сахарную свеклу. В определенных случаях культурное растение представляет собой рис. В определенных случаях культурное растение представляет собой томат. В определенных случаях культурное растение представляет собой картофель.

72. В определенных случаях растение представляет собой сельскохозяйственную культуру. Примеры таких культурных растений включают без ограничения однодольные и двудольные растения, в том числе без ограничения кормовые растения или кормовые бобовые растения, декоративные растения, продовольственные культуры, деревья или

кустарники, выбранные из Acer spp., Allium spp., Amaranthus spp., Ananas comosus, Apium graveolens, Arachis spp, Asparagus officinalis, Beta vulgaris, Brassica spp. (например, Brassica napus, Brassica rapa ssp. (канола, масличный рапс, репа масличная), Camellia sinensis, Canna indica, Cannabis saliva, Capsicum spp., Castanea spp., Cichorium endivia, Citrullus lanatus, Citrus spp., Cocos spp., Coffea spp., Coriandrum sativum, Corylus spp., Crataegus spp., Cucurbita spp., Cucumis spp., Daucus carota, Fagus spp., Ficus carica, Fragaria spp., Ginkgo biloba, Glycine spp. (например, Glycine max, Soja hispida или Soja max), Gossypium hirsutum, Helianthus spp. (например, Helianthus annuus), Hibiscus spp., Hordeum spp. (например, Hordeum vulgare), Ipomoea batatas, Juglans spp., Lactuca sativa, Linum usitatissimum, Litchi chinensis, Lotus spp., Luffa acutangula, Lupinus spp., Lycopersicon spp. (например, Lycopersicon esculenturn, Lycopersicon lycopersicum, Lycopersicon pyriforme), Malus spp., Medicago sativa, Mentha spp., Miscanthus sinensis, Morus nigra, Musa spp., Nicotiana spp., Olea spp., Oryza spp. (например, Oryza sativa, Oryza latifolia), Panicum miliaceum, Panicum virgatum, Passiflora edulis, Petroselinum crispum, Phaseolus spp., Pinus spp., Pistacia vera, Pisum spp., Poa spp., Populus spp., Prunus spp., Pyrus communis, Quercus spp., Raphanus sativus, Rheum rhabarbarum, Ribes spp., Ricinus communis, Rubus spp., Saccharum spp., Salix sp., Sambucus spp., Secale cereale, Sesamum spp., Sinapis spp., Solanum spp. (например, Solanum tuberosum, Solanum integrifolium или Solanum lycopersicum), Sorghum bicolor, Sorghum halepense, Spinacia spp., Tamarindus indica, Theobroma cacao, Trifolium spp., Triticosecale rimpaui, Triticum spp. (например, Triticum aestivum, Triticum durum, Triticum turgidum, Triticum hybernum, Triticum macha, Triticum sativum или Triticum vulgare), Vaccinium spp., Vicia spp., Vigna spp., Viola odorata, Vitis spp. и Zea mays. В определенных вариантах осуществления культурное растение представляет собой рис, масличный рапс, канолу, сою, кукурузу (маис), хлопчатник, сахарный тростник, люцерну, сорго или пшеницу.

73. Растение или часть растения для применения в настоящем изобретении включают растения на любой стадии развития растения. В определенных случаях доставка может происходить в ходе стадий прорастания, роста проростков, вегетативного роста и репродуктивного роста. В определенных случаях доставка в растение происходит в ходе стадий вегетативного и репродуктивного роста. В некоторых случаях композицию доставляют в пыльцу растения. В некоторых случаях композицию доставляют в семя растения. В некоторых случаях композицию доставляют в протопласт растения. В некоторых случаях композицию доставляют в ткань растения. Например, композицию можно доставлять в меристематическую ткань растения (например, апикальную меристему, латеральную меристему или интеркалярную меристему). В некоторых случаях композицию доставляют в постоянную ткань растения (например, простые ткани (например, паренхиму, колленхиму или склеренхиму) или сложную постоянную ткань (например, ксилему или флоэму)). В некоторых случаях композицию доставляют в зародыш растения. В некоторых случаях композицию доставляют в растительную клетку. Стадии вегетативного и репродуктивного роста также обозначаются в данном документе как "взрослые" или "зрелые" растения.

74. В тех случаях, когда систему Gene Writer доставляют в часть растения, часть растения может быть модифицирована с помощью средства, модифицирующего растения. В качестве альтернативы система Gene Writer может распределяться в другие части растения (например, по сосудистой системе растения), которые впоследствии модифицируются средством для модификации растений.

Таблица X. Иллюстративные последовательности нуклеиновых кислот, содержащие кодирующие последовательности для ретротранспозонов из различных типов источников, и последовательности 5'-UTR и/или 3'-UTR, фланкирующие их ***(следующие страницы)***

Название	Номер доступа	Организм	Клада	NID
5S-Sauria	.	Anolis carolinensis	SINE3/5S	
AARA8_AG	.	Anopheles gambiae	I	
Aca-Nin-DC-SINE1	.	Aplysia californica	SINE2/тРНК	
ACASINE	.	Anolis carolinensis	SINE2/тРНК	
ACASINE2	.	Anolis carolinensis	SINE2/тРНК	
ACASINE2a	.	Anolis carolinensis	SINE	
ACASINE2b	.	Anolis carolinensis	SINE	
ACASINE2c	.	Anolis carolinensis	SINE	
ACASINE2d	.	Anolis carolinensis	SINE	
ACASINE3	.	Anolis carolinensis	SINE	
ACSINE1	.	Anolis carolinensis	SINE	
AFESINE	.	Azemiops feae	SINE	
AFRO_LA	.	Loxodonta africana	SINE2/тРНК	
AFROA1	.	Orycteropus afer	SINE2/тРНК	
AFROA1B	.	Orycteropus afer	SINE2/тРНК	
AFROSINE	.	Proboscidea	SINE2/тРНК	
AFROSINE-1_LA	.	Loxodonta africana	SINE2/тРНК	
AFROSINE-2_LA	.	Loxodonta africana	SINE2/тРНК	
AFROSINE1B	.	Proboscidea	SINE2/тРНК	

AFROSINE2	.	Proboscidea	SINE2/тРНК	
AFROSINE3	.	Proboscidea	SINE2/тРНК	
Ag-CR1-22	.	Anopheles gambiae	CR1	1,3,11,4,2
Ag-CR1-7	.	Anopheles gambiae	CR1	2,11,56,57,68,1,58,4,3,59
Ag-I-2	.	Anopheles gambiae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
Ag-Jen-1	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jen-11	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-1	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-12	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-13	.	Anopheles gambiae	Jockey	16,1,3,82,4,11,2,15
Ag-L1-5	.	Anopheles gambiae	L1	16,1,26,43,3,44,24,11,4,2
Ag-Loner-1	AAAB01008849	Anopheles gambiae	I	
Ag-Outcast-2	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-Outcast-5	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-Outcast-6	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-R1-6	.	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3
ALINE	.	Mammalia	RTEX	16,11,15
ALINE	.	Mammalia	RTEX	16,11,15
ALINE	.	Mammalia	RTEX	16,11,15
ALPINE1	.	Vicugna pacos	SINE2/тРНК	
ALPINE1B	.	Vicugna pacos	SINE2/тРНК	
ALPINE2	.	Vicugna pacos	SINE2/тРНК	
ALU	.	Primates	SINE1/7SL	

Alu1_OG	.	Otolemur garnettii	SINE1/7SL
Alu1_TS	.	Carlito syrichta	SINE1/7SL
Alu2_OG	.	Otolemur garnettii	SINE1/7SL
Alu2_TS	.	Carlito syrichta	SINE1/7SL
Alu3_TS	.	Carlito syrichta	SINE1/7SL
AluJ_Mim	.	Microcebus murinus	SINE1/7SL
AluJr	.	Primates	SINE1/7SL
AluJr4	.	Primates	SINE1/7SL
AluMacYa3	.	Macaca	SINE1/7SL
AluMacYb2	.	Macaca	SINE1/7SL
AluMacYb4	.	Macaca	SINE1/7SL
AluSc5	.	Primates	SINE1/7SL
AluSc8	.	Primates	SINE1/7SL
AluSg1	.	Primates	SINE1/7SL
AluSg4	.	Primates	SINE1/7SL
AluSg7	.	Primates	SINE1/7SL
AluSq10	.	Primates	SINE1/7SL
AluSq2	.	Primates	SINE1/7SL
AluSq4	.	Primates	SINE1/7SL
AluSx1	.	Primates	SINE1/7SL
AluSx3	.	Primates	SINE1/7SL
AluSx4	.	Primates	SINE1/7SL
AluSz6	.	Primates	SINE1/7SL

AluYb10	.	Homo sapiens	SINE1/7SL	
AluYb11	.	Homo sapiens	SINE1/7SL	
AluYb8a1	.	Homo sapiens	SINE1/7SL	
AluYc5	.	Primates	SINE1/7SL	
AluYd3a1_gib	.	Nomascus leucogenys	SINE1/7SL	
AluYf5	.	Primates	SINE1/7SL	
AluYk11	.	Primates	SINE1/7SL	
AluYk12	.	Primates	SINE1/7SL	
AluYk13	.	Homo sapiens	SINE1/7SL	
Ambal-1_FCy	.	Fragilariopsis cylindrus	Ambal	238,10,11,6
Ambal-1_PMul	.	Pseudo-nitzschia multiseries	Ambal	1,17,10,11,6,2
Ambal-1_TP	.	Thalassiosira pseudonana	Ambal	17,11
Ambal-1N1_TP	.	Thalassiosira pseudonana	Ambal	
Ambal-2_FCy	.	Fragilariopsis cylindrus	Ambal	17,10,11,6
Ambal-2_TP	.	Thalassiosira pseudonana	Ambal	17,10,11,247,6,7
Ambal-3_FCy	.	Fragilariopsis cylindrus	Ambal	1,17,10,11,6,2
Ambal-3N1_FCy	.	Fragilariopsis cylindrus	Ambal	
AmnSINE1_AMi	.	Alligator mississippiensis	SINE3/5S	
AmnSINE1_Crp	.	Crocodylus porosus	SINE3/5S	
AmnSINE1_GG	.	Gallus gallus	SINE3/5S	
AmnSINE1_HS	.	Homo sapiens	SINE3/5S	
AMPLICON_AA	AF065437	Aedes albopictus	RTE	
AnolisSINE2	.	Anolis carolinensis	SINE2/тРНК	

AnolisSINE2a	.	Anolis carolinensis	SINE	
AnolisSINE2b	.	Anolis carolinensis	SINE	
AnolisSINE2c	.	Anolis carolinensis	SINE	
AnolisSINE2s	.	Anolis carolinensis	SINE2/тРНК	
ARMER1	.	Hippopotamidae	L1	
ATLINE1_1	.	Arabidopsis thaliana	L1	
ATLINE1_10	AL161518	Arabidopsis thaliana	L1	78,2,24,11,1,81,77,4,16,75,3,76
ATLINE1_10A	AC006300	Arabidopsis thaliana	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
ATLINE1_11	AL161498	Arabidopsis thaliana	L1	43,78,2,44,24,11,12,1,14,81,77,4,16,75,3,76
ATLINE1_12	AC002387	Arabidopsis thaliana	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
ATLINE1_2	.	Arabidopsis thaliana	L1	
ATLINE1_3A	.	Arabidopsis thaliana	L1	
ATLINE1_4	AC006248	Arabidopsis thaliana	L1	
ATLINE1_5	AC007047	Arabidopsis thaliana	L1	
ATLINE1_6	AC007047, AL161499	Arabidopsis thaliana	L1	16,1,75,3,76,77,24,11,4,2
ATLINE1_7	AL138652	Arabidopsis thaliana	L1	43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7
ATLINE1_8	AB073153	Arabidopsis thaliana	L1	1,43,3,44,4,2
ATLINE1_9	AC005897	Arabidopsis thaliana	L1	1,43,3,44,11,4,2
ATLINE1A	.	Arabidopsis thaliana	L1	
ATLINE2	AC005724	Arabidopsis thaliana	L1	
ATLINEIII	.	Arabidopsis thaliana	L1	
AtSB2	.	Arabidopsis thaliana	SINE2/тРНК	
AtSB3	AY033702	Arabidopsis thaliana	SINE2/тРНК	

AtSB4 . Arabidopsis thaliana SINE2/тPHK

AtSB5 . Arabidopsis thaliana SINE2/тPHK

AtSB6 . Arabidopsis thaliana SINE2/тPHK

ATSINE2A AC006528 Arabidopsis thaliana SINE

Au . Aegilops umbellulata SINE2/тPHK

AvaIII . Salmoniformes SINE2/тPHK

AviRTE_AMac . Ara macao RTE 16,1,26,3,4,11,2,15

AviRTE_ApVit . Apaloderma vittatum RTE 16,1,26,3,11,4,2,15

AviRTE_AVi . Amazona vittata RTE

AviRTE_BRh . Buceros rhinoceros RTE 16,1,26,3,11,4,2,15

AviRTE_CAn . Calypte anna RTE 16,1,26,3,11,4,2,15

AviRTE_GRu . Gymnopithys rufigula RTE 16,1,26,3,11,4,2,15

AviRTE_LLo . Loa loa RTE 16,1,26,3,11,4,2,15

AviRTE_MUn . Melopsittacus undulatus RTE 16,1,26,3,11,4,2,15

AviRTE_MUni . Mesitornis unicolor RTE 16,1,26,3,11,4,2,15

AviRTE_MVi . Manacus vitellinus RTE 16,1,26,3,11,4,2,15

AviRTE_OMe . Oreotrochilus melanogaster RTE 16,1,26,3,24,11,4,2

AviRTE_TiGu. Tinamus guttatus RTE 16,1,26,3,11,4,2,15

AviRTE_WBa . Wuchereria bancrofti RTE 65,11

AviRTE_ZCh . Zimmerius chrysops RTE 16,1,26,3,24,11,4,2

B1 . Mus musculus SINE1/7SL

B1_Mio . Microtus ochrogaster SINE1/7SL

B1_Mm . Mus SINE1/7SL

B1_Mur1	.	Muridae	SINE1/7SL
B1_Mur2	.	Muridae	SINE1/7SL
B1_Mur3	.	Muridae	SINE1/7SL
B1_Mur4	.	Muridae	SINE1/7SL
B1_Mus1	.	Mus	SINE1/7SL
B1_Mus2	.	Mus	SINE1/7SL
B1_Rn	.	Rattus	SINE1/7SL
B1-dID	.	Rodentia	SINE1/7SL
B1F	.	Mus musculus	SINE1/7SL
B1F-1_CLan	.	Chinchilla lanigera	SINE1/7SL
B1F1	.	Muridae	SINE1/7SL
B1F1A	.	Cavia porcellus	SINE1/7SL
B1F2	.	Muridae	SINE1/7SL
B2	.	Mus musculus	SINE2/тРНК
B2_Mm1a	.	Mus	SINE2/тРНК
B2_Mm1t	.	Mus	SINE2/тРНК
B2_Mm2	.	Mus	SINE2/тРНК
B2_Rat1	.	Rattus	SINE2/тРНК
B2_Rat2	.	Rattus	SINE2/тРНК
B2_Rat3	.	Rattus	SINE2/тРНК
B2_Rat4	.	Rattus	SINE2/тРНК
B2_Rn	.	Rattus	SINE2/тРНК
B2_Rn1	.	Rattus	SINE2/тРНК

B2_Rn2	.	Rattus	SINE2/тРНК	
B2_RnY	.	Rattus	SINE2/тРНК	
B2L_S	.	Dipodidae	SINE2/тРНК	
B3	.	Rodentia	SINE2/тРНК	
B3A	.	Rodentia	SINE2/тРНК	
B4	.	Rodentia	SINE2/тРНК	
B4A	.	Rodentia	SINE2/тРНК	
Baggins-1_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Baggins-2_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Baggins-3_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,17,3,7
BAGGINS1	.	Drosophila melanogaster	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
Baggins1_Cis	.	Ciona savignyi	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
BALN1_HV	AB016119	Hordeum vulgare	L1	
BALN2_HV	AB016120	Hordeum vulgare	L1	
BALN3_HV	AB016121	Hordeum vulgare	L1	
BC1	.	Rattus norvegicus	SINE2/тРНК	
BC1_Cp	.	Mus musculus	SINE2/тРНК	
BC1_Cpo	.	Cavia porcellus	SINE2/тРНК	
BC1_Dor	.	Dipodomys ordii	SINE2/тРНК	
BC1_Ma	.	Mesocricetus	SINE2/тРНК	
BC1_Mm	.	Mus	SINE2/тРНК	
BC1_Rn	.	Rattus	SINE2/тРНК	
BCS	X51356	Bos taurus	SINE	

BflSINE1	.	Branchiostoma floridae	SINE3/5S	
BflSINE2	.	Branchiostoma floridae	SINE2/tPHK	
BILBO	U73803	Drosophila subobscura	Loa	
BivaCORE-SINE1_RuDe.		Ruditapes decussatus	SINE2/tPHK	
BivaCORE-SINE2_MiYe.		Mizuhopecten yessoensis	SINE2/tPHK	
BivaMD-SINE1_CrVi .		Crassostrea virginica	SINE2/tPHK	
BivaMD-SINE1_TeGr .		Tegillarca granosa	SINE2/tPHK	
BivaMeta-SINE1_HyCu.		Hyriopsis cumingii	SINE2/tPHK	
BivaV-SINE1_BaAz .		Bathymodiolus azoricus	SINE2/tPHK	
BivaV-SINE1_ChFa .		Azumapecten farreri	SINE2/tPHK	
BivaV-SINE1_HyCu .		Hyriopsis cumingii	SINE2/tPHK	
BivaV-SINE1_MiYe .		Mizuhopecten yessoensis	SINE2/tPHK	
BivaV-SINE1_MyCa .		Mytilus californianus	SINE2/tPHK	
BivaV-SINE1_RuDe .		Ruditapes decussatus	SINE2/tPHK	
BivaV-SINE1_SiCo .		Sinonovacula constricta	SINE2/tPHK	
BivaV-SINE2_CrGi .		Crassostrea gigas	SINE2/tPHK	
BivaV-SINE2_MiYe .		Mizuhopecten yessoensis	SINE2/tPHK	
BivaV-SINE2_PiFu .		Pinctada fucata	SINE2/tPHK	
BivaV-SINE3_CoAn .		Coelomactra antiquata	SINE2/tPHK	
BivaV-SINE3_SpSo .		Spisula solidissima	SINE2/tPHK	
BM1	X03547	Bombyx mori	SINE2/tPHK	
BM2A	X70929	Bombyx mori	L1	
BMC1	AB018558	Bombyx mori	Jockey	16,1,87,3,88,11,4,2,15

BmSE	.	Bombyx mori		SINE2/тРНК
BNR1	EU564339	Beta vulgaris	L1	1,3,4,2
BoSB10A	.	Brassica oleracea	SINE	
BoSB10A	.	Brassica oleracea	SINE2/тРНК	
BoSB11	.	Brassica oleracea	SINE	
BoSB11	.	Brassica oleracea	SINE2/тРНК	
BoSB12	.	Brassica oleracea	SINE	
BoSB12	.	Brassica oleracea	SINE2/тРНК	
BoSB13	.	Brassica oleracea	SINE	
BoSB13	.	Brassica oleracea	SINE2/тРНК	
BoSB14A	.	Brassica oleracea	SINE	
BoSB14A	.	Brassica oleracea	SINE2/тРНК	
BoSB14B	.	Brassica oleracea	SINE	
BoSB14B	.	Brassica oleracea	SINE2/тРНК	
BoSB15	.	Brassica oleracea	SINE	
BoSB15	.	Brassica oleracea	SINE2/тРНК	
BoSB1A	.	Brassica oleracea	SINE	
BoSB1A	.	Brassica oleracea	SINE2/тРНК	
BoSB2	.	Brassica oleracea	SINE	
BoSB2	.	Brassica oleracea	SINE2/тРНК	
BoSB3	.	Brassica oleracea	SINE	
BoSB3	.	Brassica oleracea	SINE2/тРНК	
BoSB5A	.	Brassica oleracea	SINE	

BoSB5A	.	Brassica oleracea	SINE2/тРНК
BoSB5B	.	Brassica oleracea	SINE
BoSB5B	.	Brassica oleracea	SINE2/тРНК
BoSB5C	.	Brassica oleracea	SINE
BoSB5C	.	Brassica oleracea	SINE2/тРНК
BoSB5D	.	Brassica oleracea	SINE
BoSB5D	.	Brassica oleracea	SINE2/тРНК
BoSB6A	.	Brassica oleracea	SINE
BoSB6A	.	Brassica oleracea	SINE2/тРНК
BoSB6B	.	Brassica oleracea	SINE
BoSB6B	.	Brassica oleracea	SINE2/тРНК
BoSB6C	.	Brassica oleracea	SINE
BoSB6C	.	Brassica oleracea	SINE2/тРНК
BoSB6D	.	Brassica oleracea	SINE
BoSB6D	.	Brassica oleracea	SINE2/тРНК
BoSB7A	.	Brassica oleracea	SINE
BoSB7A	.	Brassica oleracea	SINE2/тРНК
BoSB8A	.	Brassica oleracea	SINE
BoSB8A	.	Brassica oleracea	SINE2/тРНК
BoSB8B	.	Brassica oleracea	SINE
BoSB8B	.	Brassica oleracea	SINE2/тРНК
BoSB9A	.	Brassica oleracea	SINE
BoSB9A	.	Brassica oleracea	SINE2/тРНК

BoSB9B	.	Brassica oleracea	SINE	
BoSB9B	.	Brassica oleracea	SINE2/тРНК	
Bov-tA1	.	Ruminantia	SINE2/тРНК	
Bov-tA2	.	Ruminantia	SINE2/тРНК	
Bov-tA3	.	Ruminantia	SINE2/тРНК	
BOVA2	.	Bos taurus	SINE	
BovB	.	Bos taurus	RTE	16,1,26,3,24,11,4,2
BovB_Ac	.	Anolis carolinensis	RTE	
BovB_ACo	.	Agkistrodon contortrix	RTE	1,3,4,2
BovB_Af	.	Antechinus flavipes	RTE	
BovB_ALi	.	Amblyomma limbatum	RTE	
BovB_Ano	.	Amphibolurus norrisi	RTE	
BovB_Bh	.	Bothriocroton hydrosauri	RTE	
BovB_BM	.	Bombyx mori	RTE	
BovB_Bp	.	Burramys parvus	RTE	
BovB_Ca	.	Ctenotus atlas	RTE	
BovB_Cg	.	Christinus	RTE	
BovB_Dr	.	Christinus	RTE	
BovB_Ec	.	Equus caballus	RTE	
BovB_Er	.	Eremiascincus richardsonii	RTE	
BovB_Es	.	Egernia stokesii	RTE	
BovB_Et	.	Christinus	RTE	
BovB_Gd	.	Eremiascincus douglasi	RTE	

BovB_Gl	.	Gehyra lazelli	RTE	
BovB_Gv	.	Gehyra variegata	RTE	
BovB_Hs	.	Hydrophis spiralis	RTE	
BovB_Io	.	Isoodon obesulus	RTE	
BovB_La	.	Loxodonta africana	RTE	
BovB_Ma	.	Macropus antilopinus	RTE	
BovB_Me	.	Notamacropus eugenii	RTE	
BovB_Ml	.	Gehyra variegata	RTE	
BovB_Oa	.	Ovis aries	RTE	
BovB_Pb	.	Petaurus breviceps	RTE	
BovB_Pc	.	Procavia capensis	RTE	
BovB_Pg	.	Perameles gunnii	RTE	
BovB_Plat	.	Ornithorhynchus	RTE	
BovB_PMo	.	Python molurus	RTE	65,16,1,26,3,11,4,2,15
BovB_Ta	.	Tachyglossus aculeatus	RTE	1,3,11,4,2
BovB_Tr	.	Tiliqua rugosa	RTE	
BovB_VA	AF332697	Vipera ammodytes	RTE	16,1,26,3,24,11,4,2
BovBa-1_EF	.	Eptesicus fuscus	RTE	
Bovc-tA2	.	Ruminantia	SINE2/тРНК	
BOVTA	.	Bos taurus	SINE2/тРНК	
BRIDGE1_FR	.	Takifugu rubripes	BRIDGE	
BRIDGE1_TN	.	Tetraodon nigroviridis	BRIDGE	74,1,3,71,103,72,2
BRIDGE2_FR	.	Takifugu rubripes	Penelope	

BRIDGE2_FR . Takifugu rubripes Penelope

BS X77571 Drosophila melanogaster Jockey 16,1,87,88,3,11,4,2,15

BS2 . Drosophila melanogaster Jockey 16,1,87,14,3,88,11,4,2,15

BS3_DM . Drosophila melanogaster Jockey 16,1,3,11,4,2,15

BS4_DM . Drosophila melanogaster Jockey 1,3,2

BTALUL1 . Bos taurus SINE2/tPHK

BTALUL2 X05090 Bos taurus RTE

BuceSINE . Buceros rhinoceros SINE

BvL1 FM993986 Beta vulgaris L1 1,3,11,4,2

BvL1 FM993986 Beta vulgaris L1 1,3,11,4,2

BvL1-2 FM993987 Beta vulgaris L1

BvL1-2 FM993987 Beta vulgaris L1

CAM1_GG X70342 Gallus gallus CR1

CAN . Carnivora SINE2/tPHK

CASINE . Oryza sativa SINE

CATS . Bombyx mori R1 6,2,9,10,11,86,13,1,4,15,16,17,3,7

CAVID . Cavia porcellus SINE2/tPHK

CAVID1B . Cavia porcellus SINE2/tPHK

CAVID1C . Cavia porcellus SINE2/tPHK

CAVID2 . Cavia porcellus SINE2/tPHK

CAVID2A . Cavia porcellus SINE2/tPHK

CAVID2B . Cavia porcellus SINE2/tPHK

CAVID2B1 . Cavia porcellus SINE2/tPHK

CAVID2B2	.	Cavia porcellus	SINE2/тPHK
CAVID2C	.	Cavia porcellus	SINE2/тPHK
CAVID2C2	.	Cavia porcellus	SINE2/тPHK
CAVID2D	.	Cavia porcellus	SINE2/тPHK
CAVID2E	.	Cavia porcellus	SINE2/тPHK
CELE45	.	Caenorhabditis elegans	SINE2/тPHK
CERE1	.	Ceratotherium simum simum	SINE2/тPHK
CGT1	L76172	Colletotrichum gloeosporioidesL1	16,11,15
ChaetoSINE1	.	Chaetopterus variopedatus	SINE2/тPHK
CHR-1	AB005033	Physeter catodon	SINE2/тPHK
CHR-2	AB005034	Physeter catodon	SINE2/тPHK
CHR-2_BT	.	Bos taurus	SINE2/тPHK
CHR-2A	.	Ruminantia	SINE2/тPHK
CHR-2B	.	Ruminantia	SINE2/тPHK
CHR2A_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2B_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2C_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2C2_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2D_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2E_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHR2F_Ttr	.	Tursiops truncatus	SINE2/тPHK
CHRL	.	Cetartiodactyla	SINE2/тPHK
CHRL1_BT	.	Bos taurus	SINE2/тPHK

Cics-1	.	Ciona intestinalis	SINE2/τPHK	
CiL2	.	Ciona intestinalis	CR1	
Cili-1	AJ226259	Ciona intestinalis	L1	
Cili-2	AJ227045	Ciona intestinalis	Loa	
CIN4E_ZM	Y00086	Zea mays	L1	1,43,3,44,11,4,2
CIN4L_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
ClySINE1	.	Clytia hemisphaerica	SINE2/τPHK	
Cnl1	.	Cryptococcus neoformans	CRE	1,3,2
CoeSINE1a	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1c	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1d	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1e	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1f	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1g	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1h	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1i	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1j	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE1k	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE2	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE3	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE4	.	Latimeria chalumnae	SINE2/τPHK	
CoeSINE5	.	Latimeria chalumnae	SINE3/5S	
CoeSINE6	.	Latimeria chalumnae	SINE2/τPHK	

Copia3_HV_I	.	Hordeum vulgare	Copia	
Copia3_HV_LTR	.	Hordeum vulgare	Copia	
Coprina_Cc1	.	Coprinopsis cinerea	Penelope	
Coprina_Pc1	.	Phanerochaete chrysosporium	Penelope	3,1
Coprina_Pc2	.	Phanerochaete chrysosporium	Penelope	
Coprina-1_SLL	.	Serpula lacrymans	Penelope	
Coprina-2_SLL	.	Serpula lacrymans	Penelope	
Coprina-3_SLL	.	Serpula lacrymans	Penelope	
CR1_1a_XT	.	Xenopus tropicalis	CR1	
CR1_1b_Xt	.	Xenopus tropicalis	CR1	
CR1_AC_1	.	Anolis carolinensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1_B	.	Gallus gallus	CR1	
CR1_E_PPU	.	Picoides pubescens	CR1	
CR1_E_VSP	.	Veniliornis spilogaster	CR1	
CR1_Ele1	.	Aedes aegypti	CR1	1,3,4,11,2
CR1_Ele10	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele11	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele12	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele13	.	Aedes aegypti	CR1	16,1,3,59,4,11,57,2,15
CR1_Ele14	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele15	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1_Ele16	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele17	.	Aedes aegypti	CR1	16,1,3,59,11,4,2,15

CR1_Ele18	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele19	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1_Ele2	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1_Ele20	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele21	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1_Ele22	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele23	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele24	Aedes aegypti	CR1	53,92,2,54,11,56,57,68,1,58,4,15,16,3,59,60
CR1_Ele25	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele26	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele27	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele28	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1_Ele29	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1_Ele3	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1_Ele30	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele31	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele32	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele33	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele34	Aedes aegypti	CR1	16,1,54,3,58,59,53,11,4,56,2,15
CR1_Ele35	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele36	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele37	Aedes aegypti	CR1	54,1,3,58,59,53,11,56,4,2
CR1_Ele38	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59

CR1_Ele38B_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele39	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59,60
CR1_Ele4	.	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele40	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele41	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele42	.	Aedes aegypti	CR1	53,2,54,11,57,68,1,4,15,16,3,59
CR1_Ele5	.	Aedes aegypti	CR1	42,1,8,14,3,11,4,12,2
CR1_Ele6	.	Aedes aegypti	CR1	53,2,54,8,11,12,1,4,15,16,3,59
CR1_Ele7	.	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele8	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1_Ele9	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_F	.	Gallus gallus	CR1	16,1,3,11,4,48,2,15
CR1_GD	X14729	Gallus gallus	CR1	
CR1_GG	U88211	Gallus gallus	CR1	
CR1_HS	.	Homo sapiens	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1-1_AC	.	Aplysia californica	CR1	16,1,3,11,4,2,15
CR1-1_ACC	.	Aquila chrysaetos canadensis	CR1	16,1,3,11,4,48,2,15
CR1-1_ACo	.	Agkistrodon contortrix	CR1	1,3,4,2
CR1-1_AG	.	Anopheles gambiae	L2B	54,1,8,3,59,53,11,4,57,12,2
CR1-1_ASi	.	Alligator sinensis	CR1	2,63,11,48,1,64,4,15,16,3

CR1-1_BF	.	Branchiostoma floridae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-1_BM	.	Bombyx mori	CR1	1,3,11,4,2
CR1-1_BTa	.	Bemisia tabaci	CR1	53,2,54,11,57,68,1,4,3,59
CR1-1_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-1_CGi	.	Crassostrea gigas	CR1	16,1,3,11,4,2,15
CR1-1_CJap	.	Caenorhabditis japonica	CR1	16,1,3,4,11,2,15
CR1-1_CM	DQ524336	Callorhinchus milii	CR1	
CR1-1_CPB	.	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-1_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-1_Croc	.	Crocodylidae	CR1	16,1,3,4,11,48,2,15
CR1-1_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-1_DAn	.	Drosophila ananassae	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-1_DBi	.	Drosophila biarmipes	CR1	
CR1-1_DBp	.	Drosophila bipectinata	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-1_DEl	.	Drosophila elegans	CR1	1,3,11,4,2
CR1-1_DEu	.	Drosophila eugracilis	CR1	1,16,3,11,4,57,2,15
CR1-1_DF	.	Drosophila ficusphila	CR1	1,3,4,2
CR1-1_DK	.	Drosophila kikkawai	CR1	1,3,4,2
CR1-1_DPer	.	Drosophila persimilis	CR1	16,1,3,11,4,57,2,15
CR1-1_DRh	.	Drosophila rhopaloa	CR1	16,1,3,59,60,4,11,57,2,15
CR1-1_DSer	.	Drosophila serrata	CR1	1,3,11,4,2
CR1-1_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,57,2,15

CR1-1_DVi	.	Drosophila virilis CR1	16,1,3,11,4,2,15
CR1-1_DWil	.	Drosophila willistoni CR1	1,3,11,4,57,2
CR1-1_Gav	.	Gavialis gangeticus CR1	16,1,3,11,4,2,15
CR1-1_HM	.	Hydra vulgaris CR1	16,1,3,11,4,2,15
CR1-1_Hmel	.	Heliconius melpomene CR1	1,3,11,4,2
CR1-1_HMM	.	Heliconius melpomene melpomeneCR1	1,3,4,2
CR1-1_IS	.	Ixodes scapularis CR1	16,1,3,11,4,2,15
CR1-1_LCh	.	Latimeria chalumnae CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_LG	.	Lottia gigantea CR1	21,16,1,3,11,15,4,2
CR1-1_Lme	.	Latimeria menadoensis CR1	16,1,3,11,4,2,15
CR1-1_LMi	.	Locusta migratoria CR1	16,1,3,11,4,2,15
CR1-1_LSal	.	Lepeophtheirus salmonis CR1	16,1,3,11,4,2,15
CR1-1_LVa	.	Litopenaeus vannamei CR1	16,1,3,11,4,2,15
CR1-1_MUn	.	Melopsittacus undulatus CR1	16,1,3,11,4,48,2,15
CR1-1_NV	.	Nematostella vectensis CR1	
CR1-1_NVi	.	Nasonia vitripennis CR1	1,8,3,11,4,12,2
CR1-1_OL	.	Oryzias latipes CR1	16,1,3,4,11,2,15
CR1-1_PH	.	Parhyale hawaiensis CR1	16,1,3,11,4,2,15
CR1-1_PMo	.	Python molurus CR1	
CR1-1_PPac	.	Pristionchus pacificus CR1	1,3,4,2
CR1-1_PPo	.	Papilio polytes CR1	
CR1-1_PSi	.	Pelodiscus sinensis CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_SP	.	Strongylocentrotus purpuratus CR1	

CR1-1_TCa	.	Tribolium castaneum	CR1	1,3,4,2
CR1-1_XL	.	Xenopus laevis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-10_AG	NT_078268.2	Anopheles gambiae	CR1	16,1,3,11,4,2,15
CR1-10_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-10_BF	.	Branchiostoma floridae	CR1	
CR1-10_CGi	.	Crassostrea gigas	CR1	1,3,11,4,2
CR1-10_CJap	.	Caenorhabditis japonica	CR1	1,3,4,2
CR1-10_CM	.	Callorhinchus milii	CR1	16,1,3,11,15,4,2
CR1-10_CPB	.	Chrysemys picta bellii	CR1	82,2,63,11,48,1,64,4,15,16,3
CR1-10_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-10_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-10_Crp	.	Crocodylus porosus	CR1	2,63,11,1,64,4,15,16,3
CR1-10_CTe	.	Capitella teleta	CR1	16,1,3,11,4,2,15
CR1-10_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-10_DTa	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-10_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-10_NV	.	Nematostella vectensis	CR1	
CR1-10_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,11,4,2,15
CR1-100_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-101_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-102_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-103_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-104_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59

CR1-105_AAe	.	Aedes aegypti	CR1	1,3,172,173,59,11,4,2
CR1-106_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-107_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-108_AAe	.	Aedes aegypti	CR1	1,3,4,11,2
CR1-109_AAe	.	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-10B_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-10B_CM	.	Callorhinchus milii	CR1	63,16,1,64,3,11,4,48,2,15
CR1-10B_CPB	.	Chrysemys picta bellii	CR1	82,2,11,48,1,4,15,16,3
CR1-10D_AMi	.	Alligator mississippiensis	CR1	2,63,11,1,64,4,15,16,3
CR1-11_AG	.	Anopheles gambiae	CR1	2,11,56,57,68,1,58,4,127,3,59,60
CR1-11_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-11_BF	.	Branchiostoma floridae	CR1	
CR1-11_CGi	.	Crassostrea gigas	CR1	1,11,4,2
CR1-11_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-11_CM	.	Callorhinchus milii	CR1	
CR1-11_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-11_CRem	.	Caenorhabditis remanei	CR1	1,3,4,2
CR1-11_Crp	.	Crocodylus porosus	CR1	
CR1-11_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-11_DTa	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-11_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-11_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-11_NV	.	Nematostella vectensis	CR1	

CR1-110_AAe	.	Aedes aegypti	CR1	1,3,11,4,57,2
CR1-111_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-112_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-113_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-114_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-115_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-116_AAe	.	Aedes aegypti	CR1	2,11,56,57,68,1,58,4,3,59
CR1-117_AAe	.	Aedes aegypti	CR1	16,1,3,60,4,11,2,15
CR1-118_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,69,3,59
CR1-119_AAe	.	Aedes aegypti	CR1	16,54,1,3,59,53,11,4,57,2,15
CR1-12_AG	.	Anopheles gambiae	CR1	16,54,1,3,53,11,4,244,2,15
CR1-12_AMi	.	Alligator mississippiensis	CR1	
CR1-12_BF	.	Branchiostoma floridae	CR1	
CR1-12_CGi	.	Crassostrea gigas	CR1	1,3,4,2
CR1-12_CJap	.	Caenorhabditis japonica	CR1	1,3,4,2
CR1-12_CM	.	Callorhinchus milii	CR1	
CR1-12_CPB	.	Chrysemys picta bellii	CR1	
CR1-12_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-12_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-12_Crp	.	Crocodylus porosus	CR1	16,1,3,11,4,2,15
CR1-12_CTe	.	Capitella teleta	CR1	2,11,57,1,4,15,16,3,59
CR1-12_DR	.	Danio rerio	CR1	1,3,4,2
CR1-12_DTa	.	Drosophila takahashii	CR1	1,3,11,4,57,2

CR1-12_HM . Hydra vulgaris CR1 1,3,11,4,2

CR1-12_HRo . Helobdella robusta CR1 16,1,3,60,11,4,2,15

CR1-12_NV . Nematostella vectensis CR1

CR1-120_AAe . Aedes aegypti CR1 1,3,11,4,2

CR1-121_AAe . Aedes aegypti CR1 1,3,4,2

CR1-122_AAe . Aedes aegypti CR1 53,2,54,11,57,1,4,15,16,3,59

CR1-123_AAe . Aedes aegypti CR1 1,3,59,11,4,2

CR1-124_AAe . Aedes aegypti CR1 54,1,8,3,59,53,11,4,12,2

CR1-125_AAe . Aedes aegypti CR1 1,3,59,11,4,2

CR1-126_AAe . Aedes aegypti CR1 16,1,3,11,4,2,15

CR1-127_AAe . Aedes aegypti CR1 54,1,3,59,53,11,4,2

CR1-12B_CPB . Chrysemys picta bellii CR1

CR1-13_AMi . Alligator mississippiensis CR1

CR1-13_BF . Branchiostoma floridae CR1

CR1-13_CGi . Crassostrea gigas CR1 16,1,3,24,4,11,2

CR1-13_CJap . Caenorhabditis japonica CR1 16,1,3,11,4,2,15

CR1-13_CM . Callorhinchus milii CR1

CR1-13_CPB . Chrysemys picta bellii CR1

CR1-13_CQ . Culex quinquefasciatus CR1 16,1,3,11,4,2,15

CR1-13_CRem . Caenorhabditis remanei CR1 1,3,11,4,2

CR1-13_Crp . Crocodylus porosus CR1

CR1-13_CTe . Capitella teleta CR1 1,3,4,2

CR1-13_DTa . Drosophila takahashii CR1 1,3,11,4,2

CR1-13_HRo	.	Helobdella robusta CR1	16,1,3,11,4,2,15
CR1-13_NV	.	Nematostella vectensis CR1	
CR1-13_SP	.	Strongylocentrotus purpuratus CR1	
CR1-14_AMi	.	Alligator mississippiensis CR1	
CR1-14_BF	.	Branchiostoma floridae CR1	
CR1-14_CGi	.	Crassostrea gigas CR1	21,1,3,11,4,2
CR1-14_CJap	.	Caenorhabditis japonica CR1	16,1,3,11,4,2,15
CR1-14_CQ	.	Culex quinquefasciatus CR1	1,3,59,11,4,57,2
CR1-14_Croc	.	Crocodylidae CR1	2,11,48,1,4,15,16,3
CR1-14_CTe	.	Capitella teleta CR1	1,3,11,4,2
CR1-14_HRo	.	Helobdella robusta CR1	16,1,3,4,11,2,15
CR1-14B_CQ	.	Culex quinquefasciatus CR1	16,1,3,11,4,2,15
CR1-15_AMi	.	Alligator mississippiensis CR1	
CR1-15_BF	.	Branchiostoma floridae CR1	
CR1-15_CJap	.	Caenorhabditis japonica CR1	1,3,4,2
CR1-15_CQ	.	Culex quinquefasciatus CR1	1,3,4,11,2
CR1-15_Crp	.	Crocodylus porosus CR1	
CR1-15_CTe	.	Capitella teleta CR1	1,3,4,2
CR1-15_HM	.	Hydra vulgaris CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-15_NV	.	Nematostella vectensis CR1	
CR1-16_AMi	.	Alligator mississippiensis CR1	
CR1-16_BF	.	Branchiostoma floridae CR1	
CR1-16_CJap	.	Caenorhabditis japonica CR1	1,3,4,2

CR1-16_CQ .	Culex quinquefasciatus	CR1	54,1,3,59,53,11,4,57,2
CR1-16_Croc .	Crocodylidae	CR1	2,63,11,1,64,4,15,16,3
CR1-16_Crp .	Crocodylus porosus	CR1	
CR1-16_HRo .	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-16B_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-17_AMi .	Alligator mississippiensis	CR1	2,63,24,11,1,64,4,16,3
CR1-17_BF .	Branchiostoma floridae	CR1	
CR1-17_CJap .	Caenorhabditis japonica	CR1	1,3,4,2
CR1-17_CQ .	Culex quinquefasciatus	CR1	1,3,4,2
CR1-17_Croc .	Crocodylidae	CR1	
CR1-17_Crp .	Crocodylus porosus	CR1	
CR1-17_Gav .	Gavialis gangeticus	CR1	1,3,4,2
CR1-17_HM .	Hydra vulgaris	CR1	1,23,3,11,4,2
CR1-18_BF .	Branchiostoma floridae	CR1	
CR1-18_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-18_Crp .	Crocodylus porosus	CR1	2,63,11,1,64,4,15,16,3
CR1-18_DR .	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-18_HRo .	Helobdella robusta	CR1	16,1,3,11,15,4,2
CR1-18_NV .	Nematostella vectensis	CR1	1,3,11,4,2
CR1-19_BF .	Branchiostoma floridae	CR1	
CR1-19_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-19_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-19_HRo .	Helobdella robusta	CR1	1,3,4,2

CR1-19_NV	Nematostella vectensis	CR1	16,1,3,11,4,2,15
CR1-1a_AG	Anopheles gambiae	L2B	53,2,54,8,11,57,12,68,1,4,3,59
CR1-1B_CM	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1B_CPB	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1b_Lch	Latimeria chalumnae	CR1	
CR1-1B_PH	Parhyale hawaiensis	CR1	16,1,3,11,15,4,2
CR1-1B_PSi	Pelodiscus sinensis	CR1	
CR1-1C_CM	Callorhinchus milii	CR1	
CR1-1C_PSi	Pelodiscus sinensis	CR1	
CR1-2_ACC	Aquila chrysaetos canadensis	CR1	16,1,3,4,11,48,2,15
CR1-2_AG	Anopheles gambiae	CR1	54,1,3,59,53,11,4,57,2
CR1-2_AMi	Alligator mississippiensis	CR1	2,11,48,1,4,15,16,3
CR1-2_ASi	Alligator sinensis	CR1	16,1,3,11,4,2,15
CR1-2_BF	Branchiostoma floridae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-2_BTa	Bemisia tabaci	CR1	2,21,23,55,11,57,1,4,3,59,61
CR1-2_CBre	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-2_CGi	Crassostrea gigas	CR1	21,1,23,3,11,4,2
CR1-2_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-2_CPB	Chrysemys picta bellii	CR1	2,63,11,1,64,4,15,16,3
CR1-2_CQ	Culex quinquefasciatus	CR1	1,3,11,4,2,126
CR1-2_CRem	Caenorhabditis remanei	CR1	1,3,4,2
CR1-2_Crp	Crocodylus porosus	CR1	2,63,11,48,1,64,4,15,16,3
CR1-2_DAn	Drosophila ananassae	CR1	1,3,11,4,57,2

CR1-2_DBi	.	Drosophila biarmipes	CR1	16,1,3,11,4,57,2,15
CR1-2_DBp	.	Drosophila bipectinata	CR1	1,3,11,4,2
CR1-2_DEl	.	Drosophila elegans	CR1	
CR1-2_DEu	.	Drosophila eugracilis	CR1	16,1,3,11,4,57,2,15
CR1-2_DF	.	Drosophila ficusphila	CR1	16,1,3,11,4,2,15
CR1-2_DK	.	Drosophila kikkawai	CR1	16,1,3,11,4,2,15
CR1-2_DPer	.	Drosophila persimilis	CR1	16,1,161,162,3,11,4,2,15
CR1-2_DRh	.	Drosophila rhopaloa	CR1	16,54,1,3,53,11,4,2,15
CR1-2_DSer	.	Drosophila serrata	CR1	1,3,59,60,11,4,57,2
CR1-2_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-2_DWi	.	Drosophila willistoni	CR1	16,1,3,11,4,57,2,15
CR1-2_EL	.	Esox lucius	CR1	16,1,3,11,4,2,15
CR1-2_FR	.	Takifugu rubripes	CR1	1,47,40,3,4,41,2
CR1-2_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-2_Hmel	.	Heliconius melpomene	CR1	1,3,4,2
CR1-2_HMM	.	Heliconius melpomene melpomene	CR1	21,1,3,11,4,57,2
CR1-2_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-2_IS	.	Ixodes scapularis	CR1	1,3,4,2
CR1-2_LCh	.	Latimeria chalumnae	CR1	16,1,3,11,4,2,15
CR1-2_LG	.	Lottia gigantea	CR1	16,1,3,11,4,2,15
CR1-2_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,48,62,2,15
CR1-2_LSal	.	Lepeophtheirus salmonis	CR1	16,1,3,11,4,2,15
CR1-2_LVa	.	Litopenaeus vannamei	CR1	1,3,11,4,2

CR1-2_MUn	.	Melopsittacus undulatus	CR1	16,1,3,11,4,48,2,15
CR1-2_NV	.	Nematostella vectensis	CR1	
CR1-2_NVi	.	Nasonia vitripennis	CR1	195,1,8,14,3,11,4,12,2
CR1-2_OL	.	Oryzias latipes	CR1	1,3,2
CR1-2_PH	.	Parhyale hawaiensis	CR1	145,54,1,16,3,53,146,11,4,2,15
CR1-2_PPac	.	Pristionchus pacificus	CR1	16,1,3,11,4,2,15
CR1-2_PPo	.	Papilio polytes	CR1	
CR1-2_PSi	.	Pelodiscus sinensis	CR1	16,1,3,11,4,2,15
CR1-2_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-2_TCa	.	Tribolium castaneum	CR1	1,3,4,2
CR1-2_XL	.	Xenopus laevis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-2_XT	.	Xenopus tropicalis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-20_AMi	.	Alligator mississippiensis	CR1	16,1,11,4,2,15
CR1-20_BF	.	Branchiostoma floridae	CR1	
CR1-20_CQ	.	Culex quinquefasciatus	CR1	1,54,3,59,53,11,4,57,2
CR1-20_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-20_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-20_NV	.	Nematostella vectensis	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-20_SP	081C18	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-21_BF	.	Branchiostoma floridae	CR1	
CR1-21_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-21_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-21_NV	.	Nematostella vectensis	CR1	53,2,54,11,56,57,1,58,4,15,16,69,3,59

CR1-22_BF	.	Branchiostoma floridae	CR1	
CR1-22_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-22_DR	.	Danio rerio	CR1	1,3,4,2
CR1-22_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-22_HRo	.	Helobdella robusta	CR1	1,3,11,4,2
CR1-22_SP	.	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-22-NV	.	Nematostella vectensis	CR1	
CR1-23_BF	.	Branchiostoma floridae	CR1	
CR1-23_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-23_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-23_HRo	.	Helobdella robusta	CR1	16,1,3,11,15,4,2
CR1-24_BF	.	Branchiostoma floridae	CR1	
CR1-24_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-24_DR	.	Danio rerio	CR1	1,3,4,2
CR1-24_HRo	.	Helobdella robusta	CR1	16,1,3,4,11,2,15
CR1-24_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,4,11,2,15
CR1-25_BF	.	Branchiostoma floridae	CR1	
CR1-25_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-25_DR	.	Danio rerio	L2	16,1,3,11,4,2,15
CR1-25_HM	.	Hydra vulgaris	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-25_HRo	.	Helobdella robusta	CR1	16,1,3,4,2,15
CR1-25_SP	.	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-26_BF	.	Branchiostoma floridae	CR1	16,21,1,23,3,4,15,11,2

CR1-26_CQ .	Culex quinquefasciatus	CR1	16,1,3,4,11,2,15
CR1-26_HM .	Hydra vulgaris	CR1	16,21,1,23,3,11,4,2,15
CR1-27_BF .	Branchiostoma floridae	CR1	
CR1-27_CQ .	Culex quinquefasciatus	CR1	1,3,4,2
CR1-27_HM .	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-28_BF .	Branchiostoma floridae	CR1	
CR1-28_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-29_BF .	Branchiostoma floridae	CR1	
CR1-2B_CPB .	Chrysemys picta bellii	CR1	
CR1-2B_Croc .	Crocodylidae	CR1	2,11,48,1,4,15,16,3
CR1-2B_Crp .	Crocodylus porosus	CR1	2,63,24,11,48,1,64,4,16,3
CR1-2b_Lch .	Latimeria chalumnae	CR1	
CR1-2B_MUn .	Melopsittacus undulatus	CR1	1,3,4,2
CR1-2B_PPac .	Pristionchus pacificus	CR1	16,1,3,11,4,2,15
CR1-2C_Croc .	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-3_AG .	Anopheles gambiae	CR1	1,54,3,58,59,53,4,56,57,2
CR1-3_BF .	Branchiostoma floridae	CR1	
CR1-3_CBre .	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-3_CGi .	Crassostrea gigas	CR1	1,3,4,2
CR1-3_CJap .	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-3_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-3_CRem .	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-3_Croc .	Crocodylidae	CR1	16,1,3,11,4,2,15

CR1-3_DAn	.	Drosophila ananassae	CR1	1,3,11,4,57,2
CR1-3_DBi	.	Drosophila biarmipes	CR1	16,1,3,11,4,2,15
CR1-3_DBp	.	Drosophila bipectinata	CR1	1,3,11,4,2
CR1-3_DEl	.	Drosophila elegans	CR1	1,3,4,2
CR1-3_DK	.	Drosophila kikkawai	CR1	1,3,11,4,57,2
CR1-3_DPer	.	Drosophila persimilis	CR1	
CR1-3_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-3_DWi	.	Drosophila willistoni	CR1	1,3,11,4,57,2
CR1-3_FR	.	Takifugu rubripes	CR1	1,2
CR1-3_Gav	.	Gavialis gangeticus	CR1	16,1,3,11,15,4,2
CR1-3_HM	.	Hydra vulgaris	CR1	180,2,11,56,57,68,1,58,4,127,3,59
CR1-3_Hmel	.	Heliconius melpomene	CR1	1,3,11,4,2
CR1-3_IS	ABJB010111403	Ixodes scapularis	CR1	1,3,4,11,2
CR1-3_Lch	.	Latimeria chalumnae	CR1	
CR1-3_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,62,2,15
CR1-3_LVa	.	Litopenaeus vannamei	CR1	1,14,3,11,4,2
CR1-3_NV	.	Nematostella vectensis	CR1	
CR1-3_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
CR1-3_PH	.	Parhyale hawaiensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-3_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-3_PPo	.	Papilio polytes	CR1	
CR1-3_PSi	.	Pelodiscus sinensis	CR1	16,1,3,11,4,48,2,15
CR1-3_SP	.	Strongylocentrotus purpuratus	CR1	

CR1-3_TCa	.	Tribolium castaneum	CR1	1,3,11,4,2
CR1-30_BF	.	Branchiostoma floridae	CR1	
CR1-30_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-30_SP	.	Strongylocentrotus purpuratus	CR1	180,2,11,56,57,68,1,58,4,15,16,3,59
CR1-31_BF	.	Branchiostoma floridae	CR1	
CR1-31_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-31_SP	.	Strongylocentrotus purpuratus	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-32_BF	.	Branchiostoma floridae	CR1	
CR1-32_HM	.	Hydra vulgaris	CR1	1,23,3,11,4,2
CR1-33_BF	.	Branchiostoma floridae	CR1	
CR1-33_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-34_BF	.	Branchiostoma floridae	CR1	
CR1-34_HM	.	Hydra vulgaris	L2A	16,1,3,11,62,2,15
CR1-35_BF	.	Branchiostoma floridae	CR1	
CR1-35_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-36_BF	.	Branchiostoma floridae	CR1	
CR1-36_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-37_BF	.	Branchiostoma floridae	CR1	
CR1-37_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-37_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-38_BF	.	Branchiostoma floridae	CR1	
CR1-39_BF	.	Branchiostoma floridae	CR1	
CR1-39_HM	.	Hydra vulgaris	CR1	1,3,11,4,2

CR1-3b_LCh	.	Latimeria chalumnae	CR1	
CR1-3B_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,48,62,2,15
CR1-3B_PSi	.	Pelodiscus sinensis	CR1	
CR1-3c_LCh	.	Latimeria chalumnae	CR1	63,16,1,64,3,11,4,48,2,15
CR1-3N1_LMi	.	Locusta migratoria	CR1	
CR1-4_AG	.	Anopheles gambiae	CR1	1,3,59,11,4,2
CR1-4_BF	.	Branchiostoma floridae	CR1	
CR1-4_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,4,11,2,15
CR1-4_CGi	.	Crassostrea gigas	CR1	21,1,23,3,4,11,2
CR1-4_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-4_CM	.	Callorhinchus milii	CR1	2,11,48,1,4,15,16,109,3,110
CR1-4_CPB	.	Chrysemys picta bellii	CR1	16,1,3,4,11,48,2,15
CR1-4_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-4_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-4_Croc	.	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-4_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-4_DAn	.	Drosophila ananassae	CR1	1,3,11,4,2
CR1-4_DBi	.	Drosophila biarmipes	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-4_DEl	.	Drosophila elegans	CR1	16,1,3,4,11,2,15
CR1-4_DK	.	Drosophila kikkawai	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-4_DRh	.	Drosophila rhopaloa	CR1	16,1,3,11,4,2,15
CR1-4_DT	.	Drosophila takahashii	CR1	16,1,3,59,11,4,57,2,15
CR1-4_DWi	.	Drosophila willistoni	CR1	16,1,3,11,4,2,15

CR1-4_HM	.	Hydra vulgaris CR1	16,1,3,11,4,2,15
CR1-4_Hmel	.	Heliconius melpomene CR1	1,3,4,2
CR1-4_LCh	.	Latimeria chalumnae CR1	16,1,3,11,4,2,15
CR1-4_LMi	.	Locusta migratoria CR1	16,1,3,11,4,62,2,15
CR1-4_LVa	.	Litopenaeus vannamei CR1	1,14,3,11,4,2
CR1-4_MUn	.	Melopsittacus undulatus CR1	1,3,2
CR1-4_NV	.	Nematostella vectensis CR1	
CR1-4_NVi	.	Nasonia vitripennis CR1	1,8,3,11,4,12,2
CR1-4_PH	.	Parhyale hawaiensis CR1	16,1,3,11,4,2,15
CR1-4_PPac	.	Pristionchus pacificus CR1	1,3,4,2
CR1-4_PSi	.	Pelodiscus sinensis CR1	2,63,11,48,1,64,4,15,16,3
CR1-4_SP	.	Strongylocentrotus purpuratus CR1	
CR1-4_TCa	.	Tribolium castaneum CR1	2,11,57,1,4,15,16,3,59,60
CR1-40_BF	.	Branchiostoma floridae CR1	
CR1-41_BF	.	Branchiostoma floridae CR1	
CR1-42_BF	.	Branchiostoma floridae CR1	
CR1-42_DR	.	Danio rerio CR1	16,1,3,11,4,2,15
CR1-42_HM	.	Hydra vulgaris CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-43_AAe	.	Aedes aegypti CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-43_BF	.	Branchiostoma floridae CR1	
CR1-43_HM	.	Hydra vulgaris CR1	1,3,11,4,2
CR1-44_AAe	.	Aedes aegypti CR1	92,2,11,56,57,1,58,4,15,16,3,59,60
CR1-44_BF	.	Branchiostoma floridae CR1	

CR1-44_DR .	Danio rerio	CR1	1,3,2
CR1-44_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-45_AAe .	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1-45_BF .	Branchiostoma floridae	CR1	
CR1-45_HM .	Hydra vulgaris	CR1	21,16,1,23,3,24,11,4,2
CR1-46_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-46_BF .	Branchiostoma floridae	CR1	
CR1-46_DR .	Danio rerio	CR1	16,1,3,4,11,2,15
CR1-46_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-47_AAe .	Aedes aegypti	CR1	16,1,3,4,11,2,15
CR1-47_BF .	Branchiostoma floridae	CR1	
CR1-47_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-48_AAe .	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-48_BF .	Branchiostoma floridae	CR1	
CR1-48_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-49_AAe .	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-49_BF .	Branchiostoma floridae	CR1	
CR1-49_DR .	Danio rerio	CR1	1,3,4,2
CR1-49_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-4B_AG .	Anopheles gambiae	CR1	1,3,58,59,4,56,57,2
CR1-4B_CPB .	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-4B_LMi .	Locusta migratoria	CR1	
CR1-4C_CPB .	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3

CR1-5_AG	Anopheles gambiae	CR1	54,1,3,59,53,11,4,57,2,68
CR1-5_AMi	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-5_BF	Branchiostoma floridae	CR1	
CR1-5_CBre	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-5_CGi	Crassostrea gigas	CR1	2,21,23,24,11,1,4,16,3
CR1-5_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-5_CM	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-5_CPB	Chrysemys picta bellii	CR1	16,1,3,4,11,2,15
CR1-5_CQ	Culex quinquefasciatus	CR1	54,1,3,59,53,11,4,2
CR1-5_CRem	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-5_Croc	Crocodylidae	CR1	2,63,24,11,1,64,4,16,3
CR1-5_CTe	Capitella teleta	CR1	1,3,11,4,2
CR1-5_DEl	Drosophila elegans	CR1	
CR1-5_DK	Drosophila kikkawai	CR1	
CR1-5_DRh	Drosophila rhopaloa	CR1	16,1,3,11,4,57,2,15
CR1-5_DT	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-5_HM	Hydra vulgaris	CR1	1,3,11,4,2
CR1-5_Hmel	Heliconius melpomene	CR1	
CR1-5_HMM	Heliconius melpomene melpomene	CR1	16,1,3,11,4,2,15
CR1-5_LCh	Latimeria chalumnae	CR1	16,1,3,11,4,2,15
CR1-5_LVa	Litopenaeus vannamei	CR1	2,142,1,14,3,11,4,12
CR1-5_MUn	Melopsittacus undulatus	CR1	1,3,4,2
CR1-5_NV	Nematostella vectensis	CR1	

CR1-5_NVi	Nasonia vitripennis	CR1	1,8,3,11,4,12,2
CR1-5_PH	Parhyale hawaiensis	CR1	16,1,3,11,4,2,15
CR1-5_PPac	Pristionchus pacificus	CR1	1,3,4,2
CR1-5_PSi	Pelodiscus sinensis	CR1	2,63,11,1,64,4,15,16,3
CR1-5_SP	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-50_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-50_BF	Branchiostoma floridae	CR1	
CR1-50_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-51_AAe	Aedes aegypti	CR1	53,2,54,8,11,57,12,42,1,4,3,59
CR1-51_BF	Branchiostoma floridae	CR1	
CR1-51_DR	Danio rerio	CR1	1,3,4,2
CR1-52_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-52_BF	Branchiostoma floridae	CR1	
CR1-53_AAe	Aedes aegypti	CR1	54,1,3,58,59,53,11,4,56,2
CR1-53_BF	Branchiostoma floridae	CR1	16,1,21,3,11,15,4,2
CR1-53_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-54_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-54_BF	Branchiostoma floridae	CR1	
CR1-54_DR	Danio rerio	CR1	
CR1-54_HM	Hydra vulgaris	CR1	21,1,23,3,4,11,2
CR1-55_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-55_BF	Branchiostoma floridae	CR1	
CR1-55_DR	Danio rerio	CR1	

CR1-55_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-56_AAe .	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-56_BF .	Branchiostoma floridae	CR1	
CR1-57_AAe .	Aedes aegypti	CR1	54,1,8,3,59,53,11,4,12,2
CR1-58_AAe .	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-58_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-59_AAe .	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-5B_AMi .	Alligator mississippiensis	CR1	1,3,2
CR1-5B_CJap .	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-5B_CM .	Callorhinchus milii	CR1	16,1,3,11,4,2,15
CR1-5B_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-5B_Crp .	Crocodylus porosus	CR1	16,1,3,4,11,2,15
CR1-6_AG .	Anopheles gambiae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,60
CR1-6_BF .	Branchiostoma floridae	CR1	
CR1-6_CBre .	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-6_CGi .	Crassostrea gigas	CR1	16,1,3,24,11,4,2
CR1-6_CJap .	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-6_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-6_CRem .	Caenorhabditis remanei	CR1	1,3,4,2
CR1-6_Croc .	Crocodylidae	CR1	2,63,24,11,48,1,64,4,16,3
CR1-6_CTe .	Capitella teleta	CR1	16,1,3,11,4,2,15
CR1-6_DT .	Drosophila takahashii	CR1	
CR1-6_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15

CR1-6_Hmel	Heliconius melpomene	CR1	1,3,2
CR1-6_HMM	Heliconius melpomene melpomene	CR1	1,3,4,2
CR1-6_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-6_LVa	Litopenaeus vannamei	CR1	16,1,3,59,11,4,57,2,15
CR1-6_MUn	Melopsittacus undulatus	CR1	
CR1-6_NV	Nematostella vectensis	CR1	
CR1-6_NVi	Nasonia vitripennis	CR1	1,3,11,4,2
CR1-6_PPac	Pristionchus pacificus	CR1	1,3,4,2
CR1-60_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-60_HM	Hydra vulgaris	CR1	16,1,23,3,11,4,2,15
CR1-61_AAe	Aedes aegypti	CR1	1,3,4,2
CR1-61_HM	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-62_AAe	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1-62_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-63_AAe	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1-63_HM	Hydra vulgaris	CR1	21,1,23,3,4,11,2
CR1-64_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1-64_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-64B_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-65_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-66_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-66_HM	Hydra vulgaris	Crack	21,1,23,3,11,4,2
CR1-67_AAe	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59

CR1-67_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-68_AAe .	Aedes aegypti	CR1	53,2,54,11,57,1,4,3,59,60
CR1-69_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-69_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-6B_CM .	Callorhinchus milii	CR1	
CR1-6B_CQ .	Culex quinquefasciatus	CR1	1,11,2
CR1-6C_AMi .	Alligator mississippiensis	CR1	
CR1-6C_CM .	Callorhinchus milii	CR1	
CR1-6D_AMi .	Alligator mississippiensis	CR1	2,63,11,1,64,4,15,16,3
CR1-6F_AMi .	Alligator mississippiensis	CR1	2,63,24,11,48,1,64,4,16,3
CR1-7_AG .	Anopheles gambiae	CR1	1,3,4,11,2
CR1-7_BF .	Branchiostoma floridae	CR1	
CR1-7_CBre .	Caenorhabditis brenneri	CR1	1,3,2
CR1-7_CGi .	Crassostrea gigas	CR1	1,3,4,2
CR1-7_CJap .	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-7_CM .	Callorhinchus milii	CR1	
CR1-7_CPB .	Chrysemys picta bellii	CR1	
CR1-7_CQ .	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-7_CRem .	Caenorhabditis remanei	CR1	16,1,3,4,11,2,15
CR1-7_Croc .	Crocodylidae	CR1	16,1,3,11,4,2,15
CR1-7_Crp .	Crocodylus porosus	CR1	
CR1-7_DT .	Drosophila takahashii	CR1	
CR1-7_Gav .	Gavialis gangeticus	CR1	1,3,4,2

CR1-7_HM	.	Hydra vulgaris	CR1	1,3,4,2
CR1-7_Hmel	.	Heliconius melpomene	CR1	
CR1-7_HMM	.	Heliconius melpomene melpomene	CR1	2,21,11,56,57,1,58,4,15,16,3,59,60
CR1-7_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-7_LVa	.	Litopenaeus vannamei	CR1	16,1,3,11,4,2,15
CR1-7_NV	.	Nematostella vectensis	CR1	
CR1-7_NVi	.	Nasonia vitripennis	CR1	212,1,213,8,3,11,4,12,2
CR1-7_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-7_SP	.	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-70_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-70_HM	.	Hydra vulgaris	CR1	16,1,23,3,24,11,4,2
CR1-71_AAe	.	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-71_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-72_AAe	.	Aedes aegypti	CR1	16,1,3,59,60,4,11,57,2,15
CR1-72_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-73_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-74_AAe	.	Aedes aegypti	CR1	16,1,3,59,11,4,57,2,68,15
CR1-74_HM	.	Hydra vulgaris	CR1	16,1,23,3,11,4,2,15
CR1-75_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-75_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-76_AAe	.	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-76_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-77_AAe	.	Aedes aegypti	CR1	16,1,3,4,11,2,15

CR1-77_HM .	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-78_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-79_AAe .	Aedes aegypti	CR1	54,1,8,3,59,53,11,4,12,2
CR1-79_HM .	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-7B_Gav .	Gavialis gangeticus	CR1	16,1,3,11,4,2,15
CR1-8_AG .	Anopheles gambiae	CR1	1,3,4,2
CR1-8_AMi .	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-8_BF .	Branchiostoma floridae	CR1	
CR1-8_CBre .	Caenorhabditis brenneri	CR1	1,3,4,2
CR1-8_CGi .	Crassostrea gigas	CR1	21,1,23,3,11,4,2
CR1-8_CJap .	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-8_CM .	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-8_CPB .	Chrysemys picta bellii	CR1	
CR1-8_CQ .	Culex quinquefasciatus	CR1	16,1,3,11,4,57,2,15
CR1-8_CRem .	Caenorhabditis remanei	CR1	1,3,4,2
CR1-8_Crp .	Crocodylus porosus	CR1	
CR1-8_CTe .	Capitella teleta	CR1	1,3,4,2
CR1-8_DTa .	Drosophila takahashii	CR1	1,3,11,4,2
CR1-8_Gav .	Gavialis gangeticus	CR1	16,1,3,4,2,15
CR1-8_Hmel .	Heliconius melpomene	CR1	1,3,11,4,2
CR1-8_HMM .	Heliconius melpomene melpomene	CR1	1,3,55,59,11,4,57,61,2
CR1-8_HRo .	Helobdella robusta	CR1	16,1,23,3,4,11,2,15
CR1-8_NV .	Nematostella vectensis	CR1	

CR1-8_PPac .	Pristionchus pacificus	CR1	1,3,4,2
CR1-8_SP .	Strongylocentrotus purpuratus	CR1	
CR1-80_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-80_HM .	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-81_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-81_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-82_AAe .	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-82_HM .	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-83_AAe .	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-83_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-84_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-84_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-85_AAe .	Aedes aegypti	CR1	1,3,11,4,2
CR1-85_HM .	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-86_AAe .	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1-86_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-87_AAe .	Aedes aegypti	CR1	1,8,3,4,11,12,2
CR1-87_HM .	Hydra vulgaris	CR1	1,3,4,2
CR1-88_AAe .	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-88_HM .	Hydra vulgaris	CR1	1,3,11,4,2
CR1-89_AAe .	Aedes aegypti	CR1	1,69,3,59,11,4,2
CR1-89_HM .	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-9_AG .	Anopheles gambiae	CR1	1,54,16,3,53,11,4,57,2,15

CR1-9_AMi	Alligator mississippiensis	CR1	
CR1-9_BF	Branchiostoma floridae	CR1	
CR1-9_CGi	Crassostrea gigas	CR1	1,3,4,2
CR1-9_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-9_CM	Callorhinchus milii	CR1	
CR1-9_CPB	Chrysemys picta bellii	CR1	
CR1-9_CQ	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-9_CRem	Caenorhabditis remanei	CR1	1,3,4,2
CR1-9_Crp	Crocodylus porosus	CR1	
CR1-9_DTa	Drosophila takahashii	CR1	1,3,11,4,2
CR1-9_HM	Hydra vulgaris	Crack	1,3,4,2
CR1-9_Hmel	Heliconius melpomene	CR1	1,3,4,2
CR1-9_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-9_NV	Nematostella vectensis	CR1	
CR1-9_NVi	Nasonia vitripennis	CR1	1,8,3,11,4,12,2
CR1-9_SP	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-90_AAe	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-91_AAe	Aedes aegypti	CR1	53,2,54,11,57,68,1,4,15,16,3,59
CR1-92_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-93_AAe	Aedes aegypti	CR1	54,1,8,3,59,53,4,11,12,2
CR1-93_HMa	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-94_AAe	Aedes aegypti	CR1	1,3,4,11,2
CR1-94_HMa	Hydra vulgaris	CR1	16,1,3,11,4,2,15

CR1-95_AAe	.	Aedes aegypti	CR1	1,3,11,4,57,2
CR1-96_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-96_HMa	.	Hydra vulgaris	CR1	16,1,3,11,15,4,2
CR1-97_AAe	.	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-98_AAe	.	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1-99_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-9B_AMi	.	Alligator mississippiensis	CR1	16,1,3,4,11,2,15
CR1-9B_CM	.	Callorhinchus milii	CR1	
CR1-9C_AMi	.	Alligator mississippiensis	CR1	2,63,11,1,64,4,15,16,3
CR1-9D_AMi	.	Alligator mississippiensis	CR1	16,1,3,4,11,2,15
CR1-B2	.	Gallus	CR1	
CR1-C	.	Gallus	CR1	16,1,3,11,4,2,15
CR1-C2	.	Gallus	CR1	
CR1-C3	.	Gallus	CR1	
CR1-C4	.	Gallus	CR1	
CR1-D	.	Gallus	CR1	
CR1-D2	.	Gallus	CR1	
CR1-E_Pass	.	Passeriformes	CR1	
CR1-F2	.	Gallus	CR1	
CR1-G	.	Gallus	CR1	
CR1-H	.	Gallus	CR1	
CR1-H2	.	Gallus	CR1	
CR1-I_Tgu	.	Passeriformes	CR1	

CR1-J1_Pass	.	Passeriformes	CR1	
CR1-J2_Pass	.	Passeriformes	CR1	
CR1-J3_Pass	.	Passeriformes	CR1	
CR1-K1_Tgu	.	Estrildidae	CR1	
CR1-K2_Tgu	.	Estrildidae	CR1	
CR1-K3_Tgu	.	Estrildidae	CR1	
CR1-K4_Tgu	.	Estrildidae	CR1	
CR1-L1_Tgu	.	Estrildidae	CR1	
CR1-L2_Tgu	.	Estrildidae	CR1	
CR1-L2-1_XT	.	Xenopus tropicalis	L2	1,3,11,4,2
CR1-L3A_Croc	.	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-L3B_Croc	.	Crocodylidae	CR1	16,1,3,4,11,2,15
CR1-X1_3-конец	.	Gallus	CR1	
CR1-X1_Pass	.	Passeriformes	CR1	
CR1-X2_3-конец	.	Gallus	CR1	
CR1-X2_Pass	.	Passeriformes	CR1	
CR1-X3_Pass	.	Passeriformes	CR1	
CR1-Y	.	Gallus	CR1	
CR1-Y_Pass	.	Passeriformes	CR1	
CR1-Y1_Aves	.	Aves	CR1	
CR1-Y2	.	Gallus	CR1	
CR1-Y2_Aves	.	Aves	CR1	
CR1-Y3	.	Gallus	CR1	

CR1-Y4	.	Gallus	CR1	
CR1-YB1_Tgu	.	Passeriformes	CR1	
CR1-YB2_Pass	.	Passeriformes	CR1	
CR1-Z1_Pass	.	Passeriformes	CR1	
CR1-Z2_Pass	.	Passeriformes	CR1	
CR1AVI	.	Neognathae	CR1	
CR1L	.	Vertebrata	CR1	
CR1X_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,11,4,2,15
CR1Y_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,4,11,2,15
Crack-1_AAe	.	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-1_AFC	.	Cichlidae	Crack	1,3,11,4,2
Crack-1_BF	.	Branchiostoma floridae	Crack	21,16,1,23,3,11,4,2,15
Crack-1_CP	.	Culex pipiens	Crack	2,11,57,1,4,15,16,3,59,60
Crack-1_CQ	.	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-1_CS1	.	Capitella sp. 1	Crack	16,21,1,23,3,11,4,2,15
Crack-1_DR	.	Danio rerio	Crack	21,1,23,3,4,2
Crack-1_EL	.	Esox lucius	Crack	
Crack-1_GA	.	Gasterosteus aculeatus	Crack	21,1,23,3,11,4,2
Crack-1_HM	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Crack-1_IS	.	Ixodes scapularis	Crack	2,21,55,11,57,1,4,15,16,3,59,61
Crack-1_LG	.	Lottia gigantea	Crack	1,3,11,4,2
Crack-1_NV	.	Nematostella vectensis	Crack	53,2,21,54,11,56,57,68,1,58,4,15,16,69,3,59
Crack-1_PM	.	Petromyzon marinus	Crack	16,1,3,11,4,2,15

Crack-1_SP	Strongylocentrotus purpuratus	Crack	16,1,21,3,11,4,2,15
Crack-1_SSa	Salmo salar	Crack	1,3,4,11,2
Crack-10_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-10_BF	Branchiostoma floridae	Crack	53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
Crack-10_CQ	Culex quinquefasciatus	Crack	1,3,2
Crack-10_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-11_AAe	Aedes aegypti	Crack	16,1,3,55,60,11,15,57,4,2,61
Crack-11_BF	Branchiostoma floridae	Crack	53,2,54,11,56,57,68,1,58,4,15,16,3,59
Crack-11_SSa	Salmo salar	Crack	1,3,11,4,2
Crack-12_AAe	Aedes aegypti	Crack	2,55,24,11,57,1,4,16,3,59,61
Crack-12_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-12_SSa	Salmo salar	Crack	1,3,11,4,2
Crack-13_AAe	Aedes aegypti	Crack	1,3,4,11,2
Crack-13_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-13_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-14_AAe	Aedes aegypti	Crack	1,3,11,4,57,2
Crack-14_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-14_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-15_AAe	Aedes aegypti	Crack	16,1,3,59,60,11,4,57,2,15
Crack-15_BF	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-15_SSa	Salmo salar	Crack	1,3,4,2
Crack-16_AAe	Aedes aegypti	Crack	16,1,3,4,11,2,15
Crack-16_BF	Branchiostoma floridae	Crack	1,3,11,4,2

Crack-16_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-17_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-17_BF	Branchiostoma floridae	Crack	53,2,54,11,56,57,68,1,58,4,15,16,3,59
Crack-18_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-18_BF	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-19_AAe	Aedes aegypti	Crack	1,3,55,11,4,2,61
Crack-19_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-1B_PM	Petromyzon marinus	Crack	
Crack-1B_SSa	Salmo salar	Crack	1,3,11,4,2
Crack-2_AAe	Aedes aegypti	Crack	16,1,3,24,11,4,2
Crack-2_AFC	Cichlidae	Crack	16,1,3,11,4,2,15
Crack-2_BF	Branchiostoma floridae	Crack	16,21,1,23,3,11,4,2,15
Crack-2_CP	Culex pipiens	Crack	1,3,59,60,11,4,57,2
Crack-2_CQ	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-2_DR	Danio rerio	Crack	1,3,4,11,2
Crack-2_EL	Esox lucius	Crack	
Crack-2_HM	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Crack-2_NV	Nematostella vectensis	Crack	1,3,11,4,2
Crack-2_PM	Petromyzon marinus	Crack	2,1
Crack-2_SP	Strongylocentrotus purpuratus	Crack	1,3,4,2
Crack-2_SSa	Salmo salar	Crack	2,63,11,1,64,4,15,16,3
Crack-20_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-20_BF	Branchiostoma floridae	Crack	1,3,11,4,2

Crack-21_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-21_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-22_AAe	.	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-22_BF	.	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-23_AAe	.	Aedes aegypti	Crack	21,1,3,11,4,2
Crack-23_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-23_SP	.	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-24_AAe	.	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-24_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-25_AAe	.	Aedes aegypti	Crack	2,55,11,57,1,4,15,16,3,59,60,61
Crack-25_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-26_AAe	.	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-26_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-27_AAe	.	Aedes aegypti	Crack	1,3,59,60,11,4,57,2
Crack-27_BF	.	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-28_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-28_BF	.	Branchiostoma floridae	Crack	16,1,3,24,11,4,2
Crack-29_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-29_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-2B_SSa	.	Salmo salar	Crack	1,3,4,11,2
Crack-2C_SSa	.	Salmo salar	Crack	
Crack-3_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-3_AFC	.	Cichlidae	Crack	1,3,4,2

Crack-3_BF	Branchiostoma floridae	Daphne	21,1,23,3,11,4,2
Crack-3_CP	Culex pipiens	Crack	16,1,3,60,24,4,11,2
Crack-3_CQ	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-3_DR	Danio rerio	Crack	16,1,3,11,4,2,15
Crack-3_EL	Esox lucius	Crack	1,3,4,2
Crack-3_HM	Hydra vulgaris	Daphne	21,16,1,23,3,11,4,2,15
Crack-3_NVe	Nematostella vectensis	Crack	1,3,11,4,2
Crack-3_PM	Petromyzon marinus	Crack	
Crack-3_SP	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-3_SSa	Salmo salar	Crack	1,3,11,4,2
Crack-30_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-30_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-31_AAe	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-31_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-32_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-32_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-33_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-34_AAe	Aedes aegypti	Crack	16,1,3,11,4,57,2,15
Crack-35_AAe	Aedes aegypti	Crack	2,11,57,1,4,15,16,3,59,60
Crack-36_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-37_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-4_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-4_CP	Culex pipiens	Crack	1,3,11,4,2

Crack-4_CQ	.	Culex quinquefasciatus	Crack	16,1,3,24,11,4,2
Crack-4_DR	.	Danio rerio	Crack	
Crack-4_EL	.	Esox lucius	Crack	
Crack-4_HM	.	Hydra vulgaris	Daphne	1,3,4,2
Crack-4_PM	.	Petromyzon marinus	Crack	
Crack-4_SP	.	Strongylocentrotus purpuratus	Crack	63,1,64,3,11,4,62,2
Crack-4_SSa	.	Salmo salar	Crack	21,16,1,3,11,4,2,15
Crack-5_AAe	.	Aedes aegypti	Crack	1,3,4,11,2
Crack-5_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-5_CQ	.	Culex quinquefasciatus	Crack	16,1,3,59,11,4,57,2,15
Crack-5_DR	.	Danio rerio	Crack	1,4,2
Crack-5_HM	.	Hydra vulgaris	Daphne	1,3,11,4,2
Crack-5_SP	.	Strongylocentrotus purpuratus	Crack	2,11,12,1,14,4,15,16,3
Crack-5_SSa	.	Salmo salar	Crack	16,1,3,4,11,2,15
Crack-6_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-6_BF	.	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-6_CQ	.	Culex quinquefasciatus	Crack	1,3,4,2
Crack-6_DR	.	Danio rerio	Crack	
Crack-6_SP	.	Strongylocentrotus purpuratus	Crack	2,21,11,56,57,1,58,4,15,16,3,59
Crack-6_SSa	.	Salmo salar	Crack	53,2,21,54,23,24,11,57,1,4,16,3,60
Crack-7_AAe	.	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-7_BF	.	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-7_CQ	.	Culex quinquefasciatus	Crack	53,2,54,55,24,11,56,57,68,1,58,4,16,127,3,59,61

Crack-7_DR	Danio rerio	Crack	
Crack-7_HM	Hydra vulgaris	Daphne	2,11,56,57,68,1,58,4,127,3,59
Crack-7_SP	Strongylocentrotus purpuratus	Crack	2,21,11,56,57,68,1,58,4,127,3,59
Crack-7_SSa	Salmo salar	Crack	63,16,1,64,3,11,4,2,15
Crack-8_AAe	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-8_BF	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-8_CQ	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-8_HMa	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Crack-8_SP	Strongylocentrotus purpuratus	Crack	1,3,4,2
Crack-8_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-9_AAe	Aedes aegypti	Crack	1,3,59,60,11,4,57,2
Crack-9_BF	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-9_CQ	Culex quinquefasciatus	Crack	1,3,4,2
Crack-9_SP	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-9_SSa	Salmo salar	Crack	16,1,3,11,4,2,15
CRE-1_ACas	Acanthamoeba castellanii	CRE	18,5,1,19,3,2
Cre-1_BM	Bombyx mori	CRE	18,5,1,19,3,2
CRE-1_CCri	Chondrus crispus	CRE	2,1
Cre-1_FCy	Fragilariopsis cylindrus	CRE	
CRE-1_FoCa	Folsomia candida	CRE	2,121,56,57,122,1,58,123,124,69,3,59,125
Cre-1_HM	Hydra vulgaris	CRE	1,52,2
CRE-1_HRo	Helobdella robusta	CRE	1,3,2
CRE-1_LSa	Lactuca sativa	CRE	1,3,99,2

| | | |

CRE-4_CCri	.	Chondrus crispus	CRE	1,2
CRE-4_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-5_CCri	.	Chondrus crispus	CRE	2,1
CRE-5_FoCa	.	Folsomia candida	CRE	121,1,124,3,164,2
CRE-5_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-6_CCri	.	Chondrus crispus	CRE	1,3,2
CRE-6_FoCa	.	Folsomia candida	CRE	121,1,124,3,123,2,125
CRE-6_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-7_CCri	.	Chondrus crispus	CRE	1,3,2
CRE-7_FoCa	.	Folsomia candida	CRE	53,2,54,56,57,68,1,58,4,127,3,59,125
CRE-8_CCri	.	Chondrus crispus	CRE	2,1
CRE-9_CCri	.	Chondrus crispus	CRE	2,1
CRE1	M33009	Crithidia fasciculata	CRE	
CRE2	U19151	Crithidia fasciculata	CRE	
CroRTE-1	.	Crocus sativus	RTE	1,26,3,11,4,2
CroSINE-10	.	Crocus sativus	SINE2/tPHK	
CroSINE-11a	.	Crocus sativus	SINE2/tPHK	
CroSINE-11b	.	Crocus sativus	SINE2/tPHK	
CroSINE-12	.	Crocus sativus	SINE2/tPHK	
CroSINE-13	.	Crocus sativus	SINE2/tPHK	
CroSINE-1a	.	Crocus sativus	SINE2/tPHK	
CroSINE-1b	.	Crocus sativus	SINE2/tPHK	
CroSINE-1c	.	Crocus sativus	SINE2/tPHK	

CroSINE-2a	.	Crocus sativus	SINE2/тРНК
CroSINE-2b	.	Crocus sativus	SINE2/тРНК
CroSINE-2c	.	Crocus sativus	SINE2/тРНК
CroSINE-2d	.	Crocus sativus	SINE
CroSINE-3a	.	Crocus sativus	SINE2/тРНК
CroSINE-3b	.	Crocus sativus	SINE2/тРНК
CroSINE-4a	.	Crocus sativus	SINE2/тРНК
CroSINE-4b	.	Crocus sativus	SINE2/тРНК
CroSINE-4c	.	Crocus sativus	SINE2/тРНК
CroSINE-5a	.	Crocus sativus	SINE2/тРНК
CroSINE-5b	.	Crocus sativus	SINE2/тРНК
CroSINE-6	.	Crocus sativus	SINE2/тРНК
CroSINE-7	.	Crocus sativus	SINE2/тРНК
CroSINE-8	.	Crocus sativus	SINE2/тРНК
CroSINE-9	.	Crocus sativus	SINE2/тРНК
CRP1	.	Crocodylus porosus	SINE
CRYI	.	Testudinoidea	SINE2/тРНК
CryIIB	AB125437	Macrochelys temminckii	SINE2/тРНК
CSINE1	.	Oryctolagus cuniculus	SINE
CSINE2	.	Oryctolagus cuniculus	SINE
CSINE2_OP	.	Ochotona princeps	SINE
CSINE2A2_OP	.	Ochotona princeps	SINE
CSINE2B	.	Oryctolagus cuniculus	SINE

CSINE2B_OP	.	Ochotona princeps	SINE	
CSINE2C_OP	.	Ochotona princeps	SINE	
CSINE3_OP	.	Ochotona princeps	SINE	
CSINE3A	.	Oryctolagus cuniculus	SINE	
CSP2096	J01188	unidentified sea urchin	CR1	
Cte-Nin-DC-SINE1	.	Capitella teleta	SINE2/тРНК	
CYN-I	.	Cynocephalus	SINE	
CYN-II1	.	Cynocephalus	SINE2/тРНК	
CYN-II2	.	Cynocephalus	SINE2/тРНК	
CYN-II3	.	Cynocephalus	SINE2/тРНК	
CYN-III1	.	Cynocephalus	SINE2/тРНК	
CYN-III2	.	Cynocephalus	SINE2/тРНК	
CYN-III3	.	Cynocephalus	SINE2/тРНК	
CYN-III4	.	Cynocephalus	SINE2/тРНК	
CYN-III5	.	Cynocephalus	SINE2/тРНК	
Cynin-1_PH	.	Parhyale hawaiensis	RTE	53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
Cynin-2_PH	.	Parhyale hawaiensis	RTE	1,3,59,11,4,2
CZAR	M62862	Trypanosoma cruzi	CRE	2,1
D_CH	M57436	Capra hircus	SINE2/тРНК	
DANA	L42295	Danio rerio	SINE	
Daphe-16_SPu	.	Strongylocentrotus purpuratus	Daphne	21,1,3,11,4,2
Daphne_DS	.	Darwinula stevensoni	Daphne	1,3,11,4,2
Daphne-1_AFC	.	Cichlidae	Daphne	

Daphne-1_BTa	.	Bemisia tabaci	Daphne	53,2,21,54,23,11,56,57,68,1,58,4,127,3,59
Daphne-1_CSav	.	Ciona savignyi	Daphne	1,3,4,2
Daphne-1_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-1_GA	.	Gasterosteus aculeatus	Daphne	1,3,11,4,2
Daphne-1_Hmel	.	Heliconius melpomene	Daphne	16,1,3,11,4,2,15
Daphne-1_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-1_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-1_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-1_LVa	.	Litopenaeus vannamei	Daphne	21,1,23,3,11,4,2
Daphne-1_NV	.	Nematostella vectensis	Daphne	
Daphne-1_OL	.	Oryzias latipes	Daphne	16,1,3,11,4,2,15
Daphne-1_PH	.	Parhyale hawaiensis	Daphne	1,3,11,4,2
Daphne-1_PPo	.	Papilio polytes	Daphne	
Daphne-1_SK	.	Saccoglossus kowalevskii	Daphne	1,3,11,4,2
Daphne-1_TCa	.	Tribolium castaneum	Crack	16,1,3,11,4,2,15
Daphne-10_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-10_HMa	.	Hydra vulgaris	Daphne	16,1,3,11,4,2,15
Daphne-10_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,4,2
Daphne-10_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-10_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-11_CTe	.	Capitella teleta	Daphne	21,16,1,3,24,11,4,2
Daphne-11_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-11_LMi	.	Locusta migratoria	Daphne	1,3,4,2

Daphne-11_SPu	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-12_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-12_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-12_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-12_SPu	.	Strongylocentrotus purpuratus	Daphne	21,1,3,11,4,2
Daphne-13_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-13_HMa	.	Hydra vulgaris	Daphne	1,3,4,11,2
Daphne-13_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-13_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-13B_HRo	.	Helobdella robusta	Daphne	16,1,3,11,15,4,2
Daphne-14_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-14_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2
Daphne-14_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-14_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-14_NV	.	Nematostella vectensis	Daphne	
Daphne-14_SPu	.	Strongylocentrotus purpuratus	Daphne	2,21,23,11,1,4,15,16,3
Daphne-15_CTe	.	Capitella teleta	Daphne	1,3,11,4,2
Daphne-15_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-15_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-15_NVe	.	Nematostella vectensis	Daphne	16,1,3,11,4,2,15
Daphne-15_SP	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-16_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-16_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2

Daphne-16_HRo	.	Helobdella robusta	Daphne 16,1,3,11,4,2,15
Daphne-16_LMi	.	Locusta migratoria	Daphne 1,3,11,4,62,2
Daphne-17_CTe	.	Capitella teleta	Daphne 16,1,3,24,11,4,2
Daphne-17_HRo	.	Helobdella robusta	Daphne 16,1,3,24,11,4,2
Daphne-17_LMi	.	Locusta migratoria	Daphne 16,1,3,11,4,2,15
Daphne-17_SP	.	Strongylocentrotus purpuratus	Daphne 1,3,11,4,2
Daphne-18_CTe	.	Capitella teleta	Daphne 1,3,4,2
Daphne-18_HM	.	Hydra vulgaris	Daphne 21,16,1,23,3,11,4,2,15
Daphne-18_HRo	.	Helobdella robusta	Daphne 1,3,4,11,2
Daphne-18_LMi	.	Locusta migratoria	Daphne 1,3,11,4,2
Daphne-18_SP	.	Strongylocentrotus purpuratus	Daphne 21,1,3,11,4,2
Daphne-19_HRo	.	Helobdella robusta	Daphne 1,3,11,4,2
Daphne-19_LMi	.	Locusta migratoria	Daphne 16,1,3,11,4,2,15
Daphne-19_SP	.	Strongylocentrotus purpuratus	Daphne 1,3,11,4,2
Daphne-1B_LMi	.	Locusta migratoria	Daphne 1,3,11,4,57,48,62,2
Daphne-2_AFC	.	Cichlidae	Daphne
Daphne-2_BTa	.	Bemisia tabaci	Daphne 16,1,3,11,15,4,2
Daphne-2_CTe	.	Capitella teleta	Daphne 1,3,2
Daphne-2_Hmel	.	Heliconius melpomene	Daphne 16,1,3,11,4,2,15
Daphne-2_HMM	.	Heliconius melpomene melpomene	Daphne 1,3,11,4,2
Daphne-2_HRo	.	Helobdella robusta	Daphne 1,3,11,4,2
Daphne-2_LMi	.	Locusta migratoria	Daphne 1,3,4,2
Daphne-2_NV	.	Nematostella vectensis	Daphne 53,2,21,54,55,11,56,57,1,58,4,15,16,3,59,60,61

Daphne-2_OL .	Oryzias latipes	Daphne	1,3,4,2
Daphne-2_PH .	Parhyale hawaiensis	Daphne	
Daphne-2_PPo .	Papilio polytes	Daphne	
Daphne-2_SK .	Saccoglossus kowalevskii	Daphne	21,1,3,82,4,11,2
Daphne-20_HRo .	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-20_LMi .	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-21_HMa .	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-21_HRo .	Helobdella robusta	Daphne	16,1,3,24,4,11,2
Daphne-21_LMi .	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-21_SPu .	Strongylocentrotus purpuratus	Daphne	21,1,23,3,11,4,2
Daphne-22_HRo .	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-22_LMi .	Locusta migratoria	Daphne	1,3,4,62,2
Daphne-22_SPur .	Strongylocentrotus purpuratus	Daphne	2,21,23,11,1,4,15,16,3
Daphne-23_HRo .	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-23_LMi .	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-23_SPur .	Strongylocentrotus purpuratus	Daphne	1,3,4,2
Daphne-24_HMa .	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-24_HRo .	Helobdella robusta	Daphne	16,1,3,24,11,4,2
Daphne-24_LMi .	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-24_SPur .	Strongylocentrotus purpuratus	Daphne	16,21,1,3,11,4,2,15
Daphne-25_HRo .	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-25_LMi .	Locusta migratoria	Daphne	1,3,4,2
Daphne-25_SPur .	Strongylocentrotus purpuratus	Daphne	21,1,23,3,11,4,2

Daphne-26_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-26_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-26_SPur	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-26B_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-27_DRe	.	Danio rerio	Daphne	1,3,2
Daphne-27_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-27_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-28_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-28_LMi	.	Locusta migratoria	Daphne	1,3,4,11,62,2
Daphne-29_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2
Daphne-29_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-29_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-3_AFC	.	Cichlidae	Daphne	
Daphne-3_BTa	.	Bemisia tabaci	Daphne	2,1,3,11,4,62
Daphne-3_CTe	.	Capitella teleta	Daphne	1,3,2
Daphne-3_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,4,11,2
Daphne-3_HRo	.	Helobdella robusta	Daphne	1,3,11,4,183,2
Daphne-3_LMi	.	Locusta migratoria	Daphne	53,62,2,54,11,57,68,1,4,15,16,3,59
Daphne-3_LVa	.	Litopenaeus vannamei	Daphne	2,21,23,11,1,4,15,16,3
Daphne-3_NV	.	Nematostella vectensis	Daphne	
Daphne-3_OL	.	Oryzias latipes	Daphne	21,1,3,11,4,2
Daphne-3_SK	.	Saccoglossus kowalevskii	Daphne	1,3,4,2
Daphne-30_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15

Daphne-30_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-31_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-31_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-32_HRo	.	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-32_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-32B_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-33_HRo	.	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-33_LMi	.	Locusta migratoria	Daphne	54,1,59,53,11,4,57,2,68
Daphne-34_HRo	.	Helobdella robusta	Daphne	16,1,3,24,11,4,2
Daphne-34_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-35_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-35_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-36_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-36_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-37_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-37_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-38_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-38_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-38_LMi	.	Locusta migratoria	Daphne	1,3,4,62,2
Daphne-39_HRo	.	Helobdella robusta	Daphne	16,1,3,59,24,11,4,57,2
Daphne-39_LMi	.	Locusta migratoria	Daphne	1,3,11,62,2
Daphne-4_AFC	.	Cichlidae	Daphne	
Daphne-4_BF	.	Branchiostoma floridae	Daphne	16,1,3,11,4,2,15

Daphne-4_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-4_HRo	.	Helobdella robusta	Daphne	2,11,57,68,1,4,15,16,3,59
Daphne-4_LMi	.	Locusta migratoria	Daphne	53,62,2,54,11,57,48,1,4,15,16,3,59
Daphne-4_LVa	.	Litopenaeus vannamei	Daphne	1,3,4,11,2
Daphne-4_NV.		Nematostella vectensis	Daphne	21,1,23,3,11,4,2
Daphne-4_PH .		Parhyale hawaiensis	Daphne	1,3,11,4,2
Daphne-40_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-40_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-40_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-41_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-41_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-42_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-43_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-44_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-45_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-46_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-47_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-48_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-49_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-5_AFC	.	Cichlidae	Daphne	
Daphne-5_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-5_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-5_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2

Daphne-5_LVa	.	Litopenaeus vannamei	Daphne	2,21,23,11,1,4,15,16,3
Daphne-50_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-51_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,30,2
Daphne-51_LMi	.	Locusta migratoria	Daphne	1,54,3,60,53,11,4,57,62,2
Daphne-52_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-52_LMi	.	Locusta migratoria	Daphne	16,1,3,11,15,4,2
Daphne-53_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-54_LMi	.	Locusta migratoria	Daphne	16,1,11,62,2,15
Daphne-55_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-56_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,30,2
Daphne-56_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-57_HMa	.	Hydra vulgaris	Daphne	21,16,1,23,3,11,4,2,15
Daphne-57_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-58_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-59_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-6_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-6_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-6_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-6_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-6_LVa	.	Litopenaeus vannamei	Daphne	21,16,1,23,3,4,11,2,15
Daphne-6_SPu	.	Strongylocentrotus purpuratus	Daphne	
Daphne-60_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-61_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2

Daphne-62_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-64_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-65_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-65_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-66_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-67_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-68_LMi	.	Locusta migratoria	Daphne	
Daphne-69_LMi	.	Locusta migratoria	Daphne	
Daphne-7_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-7_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-7_HRo	.	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-7_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-7_LVa	.	Litopenaeus vannamei	Daphne	21,16,1,23,3,11,4,2,15
Daphne-70_LMi	.	Locusta migratoria	Daphne	
Daphne-71_LMi	.	Locusta migratoria	Daphne	
Daphne-72_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-73_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-73_LMi	.	Locusta migratoria	Daphne	
Daphne-74_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-75_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-76_LMi	.	Locusta migratoria	Daphne	
Daphne-77_LMi	.	Locusta migratoria	Daphne	
Daphne-78_HMa	.	Hydra vulgaris	Daphne	1,54,127,3,58,59,53,4,56,57,11,2

Daphne-78_LMi	.	Locusta migratoria	Daphne	
Daphne-79_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-8_CTe	.	Capitella teleta	Daphne	1,3,11,4,2
Daphne-8_HMa	.	Hydra vulgaris	Daphne	16,1,3,11,4,2,15
Daphne-8_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-8_LMi	.	Locusta migratoria	Daphne	16,1,3,4,11,2,15
Daphne-80_LMi	.	Locusta migratoria	Daphne	
Daphne-81_LMi	.	Locusta migratoria	Daphne	1,11,2
Daphne-8B_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-9_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-9_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-9_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,48,62,2,15
Daphne-91_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-92_HMa	.	Hydra vulgaris	CR1	21,1,23,3,11,4,30,2
Daphne-95_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-96_HMa	.	Hydra vulgaris	Daphne	2,21,23,24,11,1,4,30,16,3
Daphne-97_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-98_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-99_HMa	.	Hydra vulgaris	Daphne	2,21,23,11,1,4,15,16,3
DASIa	.	Dasypus	SINE2/тРНК	
DASIb	.	Dasypus	SINE2/тРНК	
DASIIa	.	Dasypus	SINE2/тРНК	
DASIIb	.	Dasypus	SINE2/тРНК	

DASIIIb1	.	Dasypus	SINE2/тРНК
DASIIIb1a	.	Dasypus	SINE2/тРНК
DASIIIb2	.	Dasypus	SINE2/тРНК
DASIIIb3	.	Dasypus	SINE2/тРНК
Deceiver-1_CCO	суперконтиг_2.42	Coprinopsis cinerea	Не содержащий LTR ретротранспозон 16,1,3,24,4,11,2
Deceiver-1_MLP	.	Melampsora larici-populina	Не содержащий LTR ретротранспозон 1,3,11,4,2
Deceiver-2_MLP	.	Melampsora larici-populina	Не содержащий LTR ретротранспозон 1,3,11,4,2
Deceiver-3_MLP	.	Melampsora larici-populina	Не содержащий LTR ретротранспозон 1,3,11,4,2
DeuSINE	.	Metazoa Eukaryota.	SINE3/5S
DINOLT1	NW_633168	Dictyostelium discoideum	L1 2,1,3,11,4
DIP	.	Allactaga	SINE2/тРНК
DIP0	.	Dipodidae	SINE2/тРНК
DIPODE1	.	Dipodomys ordii	SINE2/тРНК
DIPODE2	.	Dipodomys ordii	SINE2/тРНК
DIPODE3	.	Dipodomys ordii	SINE1/7SL
DIPODE4	.	Dipodomys ordii	SINE1/7SL
Dl_TRE3-A	LODT01000000	Dictyostelium lacteum	L1 16,1,3,11,4,2,15
Dl_TRE5-A	LODT01000000	Dictyostelium lacteum	Не содержащий LTR ретротранспозон 1,3,11,4,2
DMCR1A	.	Drosophila melanogaster	CR1 16,1,3,11,4,2,15
DMRT1A	.	Drosophila melanogaster	R1 2,8,11,86,12,1,14,4,15,16,3
DMRT1B	.	Drosophila melanogaster	R1 2,8,11,86,12,1,14,4,15,16,3
DMRT1C	.	Drosophila melanogaster	R1 16,1,3,11,4,2,15
DOC	X17551	Drosophila melanogaster	Jockey 87,2,11,1,14,4,15,16,3,88

DOC2_DM	.	Drosophila melanogaster	Jockey	1,16,87,3,88,11,4,2,15
DOC3_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15
DOC4_DM	.	Drosophila melanogaster	Jockey	1,87,88,11,2
DOC5_DM	.	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15
DOC6_DM	.	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15
Dong	.	Bombyx mori	R4	1,3,4,2
DONG_FR2	.	Takifugu rubripes	R4	
Dong-1_ADi	.	Acropora digitifera	R4	1,3,4,2
Dong-1_AFC	.	Cichlidae	R4	
Dong-1_HMM	.	Heliconius melpomene melpomene	R4	1,3,4,2
Dong-1_MMa	.	Mesobuthus martensii	Loa	1,3,4,2
Dong-1_NVe	.	Nematostella vectensis	R4	1,3,4,2
Dong-1_PPo	.	Papilio polytes	R4	
Dong-1_PXu	.	Papilio xuthus	R4	
Dong-1_TCas	.	Tribolium castaneum	R4	1,3,4,2
Dong-2_ADi	.	Acropora digitifera	R4	1,3,2
Dong-2_BM	.	Bombyx mori	R4	1,3,4,2
Dong-2_HMM	.	Heliconius melpomene melpomene	R4	1,3,4,2
Dong-2_Lch	.	Latimeria chalumnae	R4	1,3,4,2
Dong-2_PPo	.	Papilio polytes	R4	
Dong-3_ADi	.	Acropora digitifera	R4	1,3,4,2
DongAa	.	Aedes aegypti	R4	1,3,4,2
DongAG	AB097127	Anopheles gambiae	R4	1,3,4,2

Doro-1_PH	.	Parhyale hawaiensis	CR1	1,16,176,3,11,4,177,2,15
Doro-2_PH	.	Parhyale hawaiensis	CR1	1,3,4,11,2
Doro-3_PH	.	Parhyale hawaiensis	CR1	16,1,3,4,11,2,15
Doro-4_PH	.	Parhyale hawaiensis	CR1	16,1,3,4,11,2,15
Doro-4B_PH	.	Parhyale hawaiensis	CR1	
Doro-5_PH	.	Parhyale hawaiensis	CR1	
Dp_NLTR-A	.	Dictyostelium purpureum	Не содержащий LTR ретротранспозон	6,2,10,11,13,1,4,100,17,3,7
Dp_TRE3-A	.	Dictyostelium purpureum	Не содержащий LTR ретротранспозон	1,3,11,4,2
Dp_TRE3-B	.	Dictyostelium purpureum	Не содержащий LTR ретротранспозон	1,3,11,4,2
DRE	X57034	Dictyostelium discoideum	L1	1,3,11,4,2
DualenCr4	.	Chlamydomonas reinhardtii	RandI	34,35,6,2,33,10,11,13,1,32,4,17,3,7
EbuSINE1	.	Eptatretus burgeri	SINE3/5S	
EbuSINE2	.	Eptatretus burgeri	SINE3/5S	
EdSINE1	.	Entamoeba dispar	SINE	
EGLN1_SM	AB016142	Solanum melongena	L1	
EGLN1_SM	AB016142	Solanum melongena	L1	
EGLN2_SM	AB016143	Solanum melongena	L1	
EGLN2_SM	AB016143	Solanum melongena	L1	
EGLN3_SM	AB016144	Solanum melongena	L1	
EGLN3_SM	AB016144	Solanum melongena	L1	
EGRT1	X86077	Blumeria graminis	Не содержащий LTR ретротранспозон	
EHAPT2	.	Entamoeba histolytica	SINE	
EhRLE2	AB097128	Entamoeba histolytica	R4	1,3,4,2

EhRLE3 AB097129 Entamoeba histolytica R4

EptSINE1 . Eptatretus burgeri SINE2/тРНК

ERE1 D26566 Equus caballus SINE2/тРНК

ERE1B . Equus caballus SINE2/тРНК

ERE1C . Equus caballus SINE2/тРНК

ERE2 . Equus SINE2/тРНК

ERE3 . Equus caballus SINE2/тРНК

ERE3B . Equus caballus SINE2/тРНК

ERE3B2 . Equus caballus SINE2/тРНК

ERE3C . Equus caballus SINE2/тРНК

ERE3C2 . Equus caballus SINE2/тРНК

ERE3C3 . Equus caballus SINE2/тРНК

ERE3C4 . Equus caballus SINE2/тРНК

ERE3D . Equus caballus SINE2/тРНК

ERE4 . Equus caballus SINE2/тРНК

ERE4B . Equus caballus SINE2/тРНК

Eri-1 AF195903AF195904AF19Mesechinus dauuricus SINE2/тРНК

Eri-2 AF195910, AF195911, Mesechinus dauuricus SINE2/тРНК

Eri-2A_EE . Erinaceus europaeus SINE2/тРНК

Eri-2B_EE . Erinaceus europaeus SINE2/тРНК

Eri-2B2_EE . Erinaceus europaeus SINE2/тРНК

Eri-2C_EE . Erinaceus europaeus SINE2/тРНК

ERI1 . Erinaceus SINE2/тРНК

Erine1	.	Erinaceus europaeus	SINE2/tPHK	
Erine1b	.	Erinaceus europaeus	SINE2/tPHK	
Erine1c	.	Erinaceus europaeus	SINE2/tPHK	
Erine2a	.	Erinaceus europaeus	SINE2/tPHK	
Erine2b	.	Erinaceus europaeus	SINE2/tPHK	
Erine2c	.	Erinaceus europaeus	SINE2/tPHK	
Expander	.	Takifugu rubripes	RTE	1,26,16,3,11,4,2,15
Expander1_Cis	.	Ciona savignyi	RTE	
EXPANDER1_DR	.	Danio rerio	RTE	1,26,16,3,11,4,2,15
EXPANDER2	.	Takifugu rubripes	RTE	16,1,26,49,3,11,4,50,2,15
F524	.	Oryza sativa	SINE2/tPHK	
FAM	.	Primates	SINE1/7SL	
FEILAI_AA	.	Aedes aegypti	SINE2/tPHK	
FEILAI_B	.	Aedes aegypti	SINE2/tPHK	
FEILAI-1B_AAe	.	Aedes aegypti	SINE2/tPHK	
FLAM	.	Primates	SINE1/7SL	
FMLN1_SI	AB016110	Setaria italica	L1	
FMLN2_SI	AB016111	Setaria italica	L1	
FMLN3_SI	AB016112	Setaria italica	L1	
FRAM	.	Primates	SINE1/7SL	
FW	.	Drosophila melanogaster	Jockey	16,1,87,3,88,24,11,4,2
FW2_DM	.	Drosophila melanogaster	Jockey	16,1,87,14,3,88,11,4,2,15
FW3_DM	.	Drosophila melanogaster	Jockey	1,87,3,88,11,4,2

G_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15
G2_DM	AC006215	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15
G3_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15
G4_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15
G5_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15
G5A_DM	.	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15
G6_DM	.	Drosophila melanogaster	Jockey	87,88
G7_DM	AC003788	Drosophila melanogaster	Jockey	
GAG_CA	AJ411828	Cicer arietinum	Не содержащий LTR ретротранспозон	
GAMERA	.	Oryzias latipes	Не содержащий LTR ретротранспозон	
GarnAlu1	.	Otolemur garnettii	SINE2/тРНК	
GarnAlu2	.	Otolemur garnettii	SINE2/тРНК	
GarnAlu3	.	Otolemur garnettii	SINE2/тРНК	
GarnAlu3B	.	Otolemur garnettii	SINE2/тРНК	
Garnel1	.	Otolemur garnettii	SINE2/тРНК	
Gecko	.	Aedes aegypti	SINE2/тРНК	
GENIE1_GI	.	Giardia intestinalis	Genie	
GilD	.	Giardia intestinalis	Не содержащий LTR ретротранспозон	
GilM	.	Giardia intestinalis	Не содержащий LTR ретротранспозон	18,5,1,3,4,2
GymnSINE	.	Gymnopithys rufigula	SINE3/5S	
HAL1	.	Eutheria	L1	
HAL1	.	Eutheria	L1	
HAL1_SS	.	Sus scrofa	L1	31,25,21,20,23,27,28,30

HAL1-1_Cja	Callithrix jacchus	L1	31,25,21,20,23,27,28,30
HAL1-1_Eeu	Erinaceus europaeus	L1	
HAL1-1_NS	Nicotiana sylvestris	L1	75,76,77
HAL1-1_Opr	Ochotona princeps	L1	
HAL1-1A_Cho	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1-1A_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1A_ML	Myotis lucifugus	L1	25,21,20,23,28,30
HAL1-1A_Opr	Ochotona princeps	L1	
HAL1-1B_Cho	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1-1B_Cja	Callithrix jacchus	L1	
HAL1-1B_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1B_ML	Myotis lucifugus	L1	
HAL1-1B_Tbel	Tupaia belangeri	L1	
HAL1-1C_Cja	Callithrix jacchus	L1	
HAL1-1C_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1C_Tbel	Tupaia belangeri	L1	
HAL1-1D_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1D_Tbel	Tupaia belangeri	L1	
HAL1-1E_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1E_ML	Myotis lucifugus	L1	
HAL1-1E_Tbel	Tupaia belangeri	L1	
HAL1-1F_Cpo	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1G_Cpo	Cavia porcellus	L1	25,21,20,23,28,30

HAL1-1H_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2_Eeu	.	Erinaceus europaeus	L1	
HAL1-2_ML	.	Myotis lucifugus	L1	
HAL1-2A_Cho	.	Choloepus hoffmanni	L1	
HAL1-2A_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3_Cho	.	Choloepus hoffmanni	L1	
HAL1-3_ML	.	Myotis lucifugus	L1	
HAL1-3_PVa	.	Pteropus vampyrus	L1	25,21,20,23,28,30
HAL1-3A_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3D_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-4_Cho	.	Choloepus hoffmanni	L1	
HAL1-4_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-4A_Cpo	.	Cavia porcellus	L1	
HAL1-4B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30,36
HAL1-4C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30,36
HAL1-5_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1B	.	Homo sapiens	L1	
HAL1B	.	Homo sapiens	L1	
HAL1M8	.	Mammalia	L1	25,37,21,38,23,20,28,30

HAL1M8	.	Mammalia	L1 25,37,21,38,23,20,28,30
HAL1M8	.	Mammalia	L1 25,37,21,38,23,20,28,30
HalFs1	.	Forficula scudderi	R1 2,8,11,132,12,86,42,131,1,14,4,15,16,3
HalSINE1	.	Haliotis discus	SINE2/tPHK
HaSE1	.	Helicoverpa armigera	SINE2/tPHK
HaSE2_DP	.	Danaus plexippus	SINE2/tPHK
HaSE3	.	Helicoverpa armigera	SINE3/5S
HaTad1-1_AB.		Agaricus bisporus	Tad1
HaTad1-1_BG.		Blumeria graminis	Tad1
HaTad1-1_BoBo	.	Botryobasidium botryosum	Tad1
HaTad1-2_AB.		Agaricus bisporus	Tad1
HaTad1-2_BG.		Blumeria graminis	Tad1
HaTad1-2_BoBo	.	Botryobasidium botryosum	Tad1
HaTad1-3_BG.		Blumeria graminis	Tad1
HaTad1-3_BoBo	.	Botryobasidium botryosum	Tad1
HaTad1-4_BG.		Blumeria graminis	Tad1
HaTad1-4_BoBo	.	Botryobasidium botryosum	Tad1
HaTad1-5_BG.		Blumeria graminis	Tad1
HE1_DR1	.	Danio rerio	SINE2/tPHK
HE1_HJ	.	Heterodontus japonicus	SINE2/tPHK
HE1_MM	.	Mustelus manazo	SINE2/tPHK
HE1_SINE	.	Mustelus manazo	SINE2/tPHK
HE2_DR	.	Danio rerio	SINE2/tPHK

Hebe_Av	.	Adineta vaga	Jockey	16,1,3,4,11,2,15
HELENA	.	Drosophila virilis	Jockey	
Helena_Ds	chr3R	Drosophila simulans	Не содержащий LTR ретротранспозон	16,1,87,3,88,11,4,2,15
HELENA_RT	.	Drosophila melanogaster	Jockey	
HER_LINE	.	Scyliorhinus torazame	CR1	1,3,4,2
HER1_HJ	AB027741	Heterodontus japonicus	Не содержащий LTR ретротранспозон	
HER1_ST	.	Scyliorhinus torazame	Не содержащий LTR ретротранспозон	
HERO-1_AC	.	Aplysia californica	Hero	1,3,4,2
HERO-1_AFC	.	Cichlidae	Hero	
HERO-1_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-1_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-1_PP	.	Physarum polycephalum	Hero	1,3,4,2
HERO-1_SP	AAGJ02121261	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-2_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-2_DR	.	Danio rerio	Hero	1,3,4,2
HERO-2_HR	.	Helobdella robusta	Hero	1,3,4,2
Hero-2_SPur	048B05	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-3_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-3_DR	.	Danio rerio	Hero	44,43
HERO-3_HR	.	Helobdella robusta	Hero	1,3,4,2
Hero-3_SPur	.	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-4_DR	.	Danio rerio	Hero	
HERO-4_HR	.	Helobdella robusta	Hero	1,3,4,2

HERO-5_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-6_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-7_HR	.	Helobdella robusta	Hero	1,4,2
HERO-8_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-9_HR	.	Helobdella robusta	Hero	1,3,4,2
HERODr	.	Danio rerio	Hero	1,3,4,2
HEROFr	.	Takifugu rubripes	Hero	1,3,4,2
HEROTn	.	Tetraodon nigroviridis	Hero	1,3,4,2
HET-A_DV	.	Drosophila virilis	Jockey	
HeT-A2_DVi	.	Drosophila virilis	Jockey	87,88
HeT-A3_DVi	.	Drosophila virilis	Jockey	87,88
HETA	U06920	Drosophila melanogaster	Jockey	87,88
HETA_DSi	.	Drosophila simulans	Jockey	87,88
HidaAg1	AB090822	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
HOLN1_EA	AB016148	Equisetum arvense	L1	
HOLN2_EA	AB016149	Equisetum arvense	L1	
HOLN3_EA	AB016150	Equisetum arvense	L1	
HOPEBm2	.	Bombyx mori	R1	2,8,11,86,12,1,14,4,15,16,3
HPAI	.	Oncorhynchus nerka	SINE2/тРНК	
HydSINE1	.	Hydractinia symbiolongicarpus	SINE2/тРНК	
I_Av	.	Adineta vaga	Penelope	1,3,48,62
I_DM	.	Drosophila melanogaster	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I_Ele10	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3

I_Ele11	.	Aedes aegypti	I	6,2,10,11,13,1,4,15,16,17,3,7
I_Ele13	.	Aedes aegypti	I	6,2,7,8,9,10,11,86,12,13,1,186,4,15,16,17,3
I_Ele14	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele15	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele16	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele17	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I_Ele18	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele19	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele2	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
I_Ele20	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I_Ele22	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele24	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I_Ele25	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I_Ele28	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I_Ele33	.	Aedes aegypti	I	6,2,7,8,10,11,12,1,4,15,16,17,3
I_Ele34	.	Aedes aegypti	I	16,1,8,3,11,4,12,2,15
I_Ele35	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele36	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I_Ele37	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele39	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,1,4,15,16,17,3
I_Ele4	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I_Ele42	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele43	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3

I_Ele43B_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele43C_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele4B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4C_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4D_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4E_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4F_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4G_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele7	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I_Ele8	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
I_Ele9	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-1_AA	.	Aedes aegypti	Nimb	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
I-1_AN	.	Aspergillus nidulans	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_AO	.	Aspergillus oryzae	I	
I-1_BM	.	Bombyx mori	Nimb	6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,7
I-1_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,1,4,15,16,17,3
I-1_CH	.	Bipolaris maydis	I	
I-1_CI	.	Ciona intestinalis	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-1_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,17,3,7
I-1_CTe	.	Capitella teleta	I	16,1,8,14,3,11,4,12,2,15
I-1_DAn	.	Drosophila ananassae	I	6,2,9,10,11,86,1,14,4,15,16,17,3,7
I-1_DBi	.	Drosophila biarmipes	I	
I-1_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,13,1,4,15,16,17,3,7

I-1_DEl	.	Drosophila elegans	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DEu	.	Drosophila eugracilis	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DF	.	Drosophila ficusphila	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-1_DK	.	Drosophila kikkawai	I	6,2,9,10,11,86,1,4,15,16,17,3,7
I-1_DP	.	Daphnia pulex	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DPe	.	Drosophila persimilis	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-1_DR	AL672145	Danio rerio	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I-1_DSer	.	Drosophila serrata	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-1_DT	.	Drosophila takahashii	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-1_DVi	.	Drosophila virilis	I	
I-1_DWi	.	Drosophila willistoni	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DYa	.	Drosophila yakuba	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_Hmel	.	Heliconius melpomene	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_HMM	.	Heliconius melpomene melpomene	I	7,1,17,9,3,10,4,6,2,13
I-1_LHu	.	Linepithema humile	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_LMi	.	Locusta migratoria	I	6,2,9,8,10,11,12,1,4,15,16,17,3,7
I-1_NVi	AAZX01003261.1	Nasonia vitripennis	I	6,2,9,10,11,13,1,15,16,17,3,7
I-1_PBa	.	Pogonomyrmex barbatus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-1_PH	.	Parhyale hawaiensis	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-1_PPo	.	Papilio polytes	I	
I-1_RPr	.	Rhodnius prolixus	I	2,8,11,12,1,14,4,15,16,3
I-1_SP	.	Strongylocentrotus purpuratus	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-10_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7

I-10_DBi	.	Drosophila biarmipes	I	17,9,10,6,7,13
I-10_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-10B_LMi	.	Locusta migratoria	I	
I-11_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-11_LMi	.	Locusta migratoria	I	7,16,1,8,3,11,4,12,2,15
I-12_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-13_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-14_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-15_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-2_AO	.	Aspergillus oryzae	I	
I-2_BF	.	Branchiostoma floridae	I	53,2,54,8,11,56,57,12,68,1,14,58,4,15,16,69,59
I-2_BM	.	Bombyx mori	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-2_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_CH	.	Bipolaris maydis	I	
I-2_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DAn	.	Drosophila ananassae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DBi	.	Drosophila biarmipes	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-2_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,86,1,4,15,16,17,3,7
I-2_DEl	.	Drosophila elegans	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DEu	.	Drosophila eugracilis	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-2_DF	.	Drosophila ficusphila	I	
I-2_DK	.	Drosophila kikkawai	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DVi	.	Drosophila virilis	I	6,2,9,10,11,86,1,4,15,16,17,3,7

I-2_DWi	.	Drosophila willistoni	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-2_DYa	.	Drosophila yakuba	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-2_LMi	.	Locusta migratoria	I	6,2,9,10,11,12,13,1,4,15,16,17,3,7
I-2_NVi	.	Nasonia vitripennis	I	2,8,11,12,1,14,4,15,16,3
I-2_PPo	.	Papilio polytes	I	
I-2_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_SPur	.	Strongylocentrotus purpuratus	I	6,2,9,10,11,13,1,14,4,15,16,17,3,7
I-2B_HMM	.	Heliconius melpomene melpomene	I	1,3,4,2
I-3_AC	.	Aplysia californica	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
I-3_AO	.	Aspergillus oryzae	Tad1	
I-3_BTa	.	Bemisia tabaci	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-3_CQ	.	Culex quinquefasciatus	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
I-3_DAn	.	Drosophila ananassae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,6,2,13
I-3_DEu	.	Drosophila eugracilis	I	6,2,7,9,10,11,86,13,1,4,15,16,17,3
I-3_DF	.	Drosophila ficusphila	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DK	.	Drosophila kikkawai	I	7,1,17,9,3,10,4,6,2,13
I-3_DR	.	Danio rerio	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_DYa	.	Drosophila yakuba	I	7,1,17,16,3,11,4,2,15
I-3_HMM	.	Heliconius melpomene melpomene	I	7,1,17,9,3,10,4,6,2,13
I-3_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_PPo	.	Papilio polytes	I	17,10,6,7

I-3_RPr	.	Rhodnius prolixus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_SPur	.	Strongylocentrotus purpuratus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3B_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,12,13,1,4,15,16,17,3,7
I-4_AC	.	Aplysia californica	Nimb	6,2,7,9,10,11,86,12,1,14,4,15,16,17,3
I-4_AO	.	Aspergillus oryzae	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
I-4_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DAn	.	Drosophila ananassae	I	7,1,17,9,3,10,4,6,2,13
I-4_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DEu	.	Drosophila eugracilis	I	
I-4_DF	.	Drosophila ficusphila	I	
I-4_DK	.	Drosophila kikkawai	I	7,1,17,9,3,10,4,6,2,13
I-4_DR	.	Danio rerio	Nimb	17,9,10,6,7,13
I-4_DYa	.	Drosophila yakuba	I	6,2,9,10,24,11,13,1,4,16,17,3,7
I-4_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-4_RPr	.	Rhodnius prolixus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-44_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-45_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-48_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I-49_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-5_AO	.	Aspergillus oryzae	Tad1	2,8,11,12,42,1,14,4,15,16,17,3

I-5_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-5_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,14,4,15,16,17,3,7
I-5_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,4,6,2,13
I-5_DF	.	Drosophila ficusphila	I	
I-5_DK	.	Drosophila kikkawai	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-50_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-51_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-52_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-53_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-54_AAe	.	Aedes aegypti	I	16,1,3,11,2,15
I-55_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,178,13,1,179,4,15,16,17,3
I-56_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-57_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-58_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-59_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-5B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_AO	.	Aspergillus oryzae	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,13,1,15,16,17,3
I-6_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7

I-6_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,4,6,2,13
I-6_DF	.	Drosophila ficusphila	I	
I-6_DY	.	Drosophila yakuba	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_LMi	.	Locusta migratoria	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-6_RPr	.	Rhodnius prolixus	I	7,1,17,9,3,10,4,2,6,13
I-60_AAe	.	Aedes aegypti	I	16,11,15
I-61_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-62_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-63_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-64_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I-65_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-66_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-67_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-68_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,1,14,4,15,16,17,3
I-69_AAe	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,13,1,4,15,16,17,3
I-7_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_DBi	.	Drosophila biarmipes	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-7_DBp	.	Drosophila bipectinata	I	
I-7_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-70_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I-71_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3

I-72_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-73_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-74_AAe	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I-75_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-76_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-77_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-78_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-79_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-7B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7C_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7D_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-8_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-8_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-8_DBi	.	Drosophila biarmipes	I	
I-8_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,1,4,15,16,17,3,7
I-8_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-80_AAe	.	Aedes aegypti	I	6,2,10,11,1,4,15,16,17,3,7
I-80B_AAe	.	Aedes aegypti	I	6,2,10,11,1,4,15,16,17,3,7
I-81_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-9_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-9_LMi	.	Locusta migratoria	I	7,1,17,9,3,10,11,6,2,13
I-9B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-N1_CQ	.	Culex quinquefasciatus	I	12,207,208,209,210,200

I1_As	.	Adineta	Penelope	1,3,62
I2_As	.	Adineta	Penelope	1,3,62
ID_B1	.	Muridae	SINE2/тРНК	
ID_Rn1	.	Rattus	SINE2/тРНК	
ID_Rn2	.	Rattus	SINE2/тРНК	
ID-B1_Cpo	.	Cavia porcellus	SINE2/тРНК	
ID-Spe	.	Spermophilus	SINE2/тРНК	
ID4_	.	Muridae	SINE2/тРНК	
IdioSINE1	.	Idiosepius paradoxus	SINE2/тРНК	
IdioSINE2	.	Idiosepius paradoxus	SINE2/тРНК	
IDL-Geo_Cca	.	Chaetodipus californicus	SINE2/тРНК	
IDL-Geo_Tbo	DQ206804DQ206805DQ20	Thomomys bottae	SINE2/тРНК	
ILRC_GL	AJ294708	Ocyphaps lophotes	CR1	
ILRC2_GL	.	Ocyphaps lophotes	Не содержащий LTR ретротранспозон	
IN25	.	Homo sapiens	L1	
Ingi	.	Trypanosoma brucei	Ingi	6,2,18,7,9,10,11,13,1,4,143,15,5,16,19,17,3
Ingi-1_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,57,13,1,4,15,16,17,3,59
Ingi-1_CCri	.	Chondrus crispus	Ingi	6,2,10,24,11,1,4,16,17,3,7
Ingi-1_HMi	.	Hymenolepis microstoma	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-1_LMi	.	Locusta migratoria	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-1_LVa	.	Litopenaeus vannamei	Ingi	6,2,9,10,11,13,1,4,17,3,7
Ingi-1_Rpro	ACPB01056162.1	Rhodnius prolixus	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-2_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3

Ingi-2_LMi	.	Locusta migratoria	Ingi	6,2,9,10,11,13,1,4,17,3,7
Ingi-2_LVa	.	Litopenaeus vannamei	Ingi	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Ingi-2_RPr	.	Rhodnius prolixus	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-3_AC	.	Aplysia californica	Ingi	6,2,9,10,24,11,13,1,4,16,17,3,7
Ingi-3_LVa	.	Litopenaeus vannamei	Ingi	
Ingi-4_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3
Ingi-4_LVa	.	Litopenaeus vannamei	Ingi	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Ingi-5_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3
Ingi-6_AC	.	Aplysia californica	Ingi	16,1,3,4,11,2,15
Ingi-7_AC	.	Aplysia californica	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-8_AC	.	Aplysia californica	Ingi	53,6,2,7,54,9,10,11,56,57,68,13,1,58,4,15,16,17,3,59
Ingi2	.	Trypanosoma brucei	Ingi	105,6,2,18,7,9,10,11,13,1,4,143,15,5,16,19,17,3
Inkcap-1_CCi	.	Coprinopsis cinerea	RTEX	3,2,1
Inkcap-2_CCi	AACS02000013	Coprinopsis cinerea	RTEX	240,7,1,17,3,10,4,11,6,2,13
ISABELLE_TM	AF326781	Triticum monococcum	L1	
Isc-Nin-DC-SINE	.	Ixodes scapularis	SINE2/tPHK	
ISEAG1	AB097148	Anopheles gambiae	L1	
IVK_DM	.	Drosophila melanogaster	Jockey	6,2,9,10,11,1,4,15,16,17,3,7
JAM1	.	Aedes aegypti	RTE	1,26,3,11,4,2
JAM1B_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2
JAM1C_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2
JN_As	.	Adineta	Penelope	1,3
JN_Av	.	Adineta vaga	Penelope	7,1,17,129,3,130

Jockey	M22874	Drosophila melanogaster	Jockey	16,1,87,88,3,24,4,11,2
Jockey_Ele10	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey_Ele2	.	Aedes aegypti	Jockey	16,1,8,14,3,4,11,12,2,15
Jockey_Ele3	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey_Ele5	.	Aedes aegypti	Jockey	
Jockey_Ele7	.	Aedes aegypti	Jockey	16,1,3,4,11,2,15
Jockey_Ele8	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-1_BTa	.	Bemisia tabaci	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_CoFl	.	Copidosoma floridanum	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_CQ	.	Culex quinquefasciatus	Jockey	16,1,14,3,11,4,2,15
Jockey-1_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DBi	.	Drosophila biarmipes	Jockey	87,14,88
Jockey-1_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-1_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DEr	.	Drosophila erecta	Jockey	87,2,11,12,1,14,4,15,16,3,88
Jockey-1_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DF	.	Drosophila ficusphila	Jockey	16,1,87,88,3,11,4,2,15
Jockey-1_DGri	.	Drosophila grimshawi	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DK	.	Drosophila kikkawai	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DPer	.	Drosophila persimilis	Jockey	87,2,11,1,4,15,16,3,88,174
Jockey-1_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_Dro	.	Drosophila	Jockey	
Jockey-1_DSe	.	Drosophila sechellia	Jockey	16,1,87,3,88,11,4,2,15

Jockey-1_DSer	.	Drosophila serrata	Jockey 16,1,87,3,88,11,4,2,15
Jockey-1_DSi	.	Drosophila simulans	Jockey 16,1,3,11,4,2,15
Jockey-1_DT	.	Drosophila takahashii	Jockey 16,1,87,3,88,11,4,2,15
Jockey-1_DVi	.	Drosophila virilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-1_DWi	.	Drosophila willistoni	Jockey 16,1,87,3,88,11,4,2,15
Jockey-1_DYa	.	Drosophila yakuba	Jockey 16,1,87,3,88,11,4,2,15
Jockey-1_Hmel	.	Heliconius melpomene	Jockey 1,3,4,2
Jockey-1_LMi	.	Locusta migratoria	Jockey 1,3,11,4,2
Jockey-1_LVa	.	Litopenaeus vannamei	Jockey 16,1,14,3,11,4,2,15
Jockey-1_PH	.	Parhyale hawaiensis	Jockey 16,1,8,3,11,4,12,2,15
Jockey-1_PPo	.	Papilio polytes	Jockey 1,3,4,2
Jockey-1_PXu	.	Papilio xuthus	Jockey
Jockey-1_RPr	.	Rhodnius prolixus	Jockey 1,3,4,2
Jockey-1_TCa	.	Tribolium castaneum	Jockey 16,1,3,11,4,2,15
Jockey-10_BTa	.	Bemisia tabaci	Jockey 16,1,3,11,4,2,15
Jockey-10_DAn	.	Drosophila ananassae	Jockey 16,1,87,14,3,88,4,11,2,15
Jockey-10_DBp	.	Drosophila bipectinata	Jockey 16,1,3,11,4,2,15
Jockey-10_DEu	.	Drosophila eugracilis	Jockey
Jockey-10_DF	.	Drosophila ficusphila	Jockey
Jockey-10_DRh	.	Drosophila rhopaloa	Jockey 16,1,87,3,88,11,4,2,15
Jockey-10_DY	.	Drosophila yakuba	Jockey 87,88
Jockey-11_AAe	.	Aedes aegypti	Jockey 16,1,3,11,4,2,15
Jockey-11_BTa	.	Bemisia tabaci	Jockey 16,1,3,11,4,2,15

Jockey-11_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-11_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-11_DF	.	Drosophila ficusphila	Jockey	
Jockey-11B_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-12_AAe	.	Aedes aegypti	Jockey	16,1,8,14,3,11,4,12,2,15
Jockey-12_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-12_DAn	.	Drosophila ananassae	Jockey	16,1,3,11,4,2,15
Jockey-12_DBp	.	Drosophila bipectinata	Jockey	
Jockey-12_DF	.	Drosophila ficusphila	Jockey	
Jockey-13_AAe	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-13_BTa	.	Bemisia tabaci	Jockey	16,1,3,4,11,2,15
Jockey-13_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-13_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-14_AAe	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-14_DAn	.	Drosophila ananassae	Jockey	1,3,4,2
Jockey-14_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-15_AAe	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-15_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-16_AAe	.	Aedes aegypti	Jockey	16,1,14,3,11,4,2,15
Jockey-17_AAe	.	Aedes aegypti	Jockey	16,1,3,4,11,2,15
Jockey-1B_Hmel	.	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15
Jockey-1B_LMi	.	Locusta migratoria	Jockey	
Jockey-2_BTa	.	Bemisia tabaci	Jockey	16,1,131,3,4,11,132,2,15

Jockey-2_CQ	.	Culex quinquefasciatus	Jockey 16,1,8,14,3,11,4,12,2,15
Jockey-2_DAn	.	Drosophila ananassae	Jockey 16,1,3,11,4,2,15
Jockey-2_DBi	.	Drosophila biarmipes	Jockey 1,3,11,4,2
Jockey-2_DBp	.	Drosophila bipectinata	Jockey 16,1,3,11,4,2,15
Jockey-2_DEl	.	Drosophila elegans	Jockey 16,1,87,3,88,11,4,2,15
Jockey-2_DEu	.	Drosophila eugracilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-2_DF	.	Drosophila ficusphila	Jockey 16,1,87,88,3,4,11,2,15
Jockey-2_DGri	.	Drosophila grimshawi	Jockey 16,1,3,24,11,4,2
Jockey-2_DK	.	Drosophila kikkawai	Jockey 16,1,87,88,3,11,4,2,15
Jockey-2_DPer	.	Drosophila persimilis	Jockey 87,2,11,1,4,15,16,3,88
Jockey-2_DRh	.	Drosophila rhopaloa	Jockey 16,1,87,3,88,11,4,2,15
Jockey-2_DSer	.	Drosophila serrata	Jockey 16,1,87,3,88,11,4,2,15
Jockey-2_DSim	.	Drosophila simulans	Jockey 16,11,15
Jockey-2_DT	.	Drosophila takahashii	Jockey 93,16,1,87,3,88,11,4,2,15
Jockey-2_DVi	.	Drosophila virilis	Jockey 87,2,11,1,4,15,16,3,88
Jockey-2_DYa	.	Drosophila yakuba	Jockey 16,1,3,4,11,2,15
Jockey-2_Hmel	.	Heliconius melpomene	Jockey 16,1,87,3,88,11,4,2,15
Jockey-2_HMM	.	Heliconius melpomene melpomene	Jockey 92,16,1,87,3,88,11,4,2,15
Jockey-2_LMi	.	Locusta migratoria	Jockey 1,3,4,2
Jockey-2_LVa	.	Litopenaeus vannamei	Jockey 16,1,8,3,46,11,4,12,2,15
Jockey-2_PPo	.	Papilio polytes	Jockey
Jockey-2_RPr	.	Rhodnius prolixus	Jockey 1,87,3,88,4,2
Jockey-2B_LMi	.	Locusta migratoria	Jockey

Jockey-3_BTa	.	Bemisia tabaci	Jockey 16,1,3,11,4,2,15
Jockey-3_CQ	.	Culex quinquefasciatus	Jockey 16,1,3,11,4,2,15
Jockey-3_DAn	.	Drosophila ananassae	Jockey 16,1,3,11,4,2,15
Jockey-3_DBp	.	Drosophila bipectinata	Jockey 16,1,3,11,4,2,15
Jockey-3_DEl	.	Drosophila elegans	Jockey 16,1,87,3,88,24,11,4,2
Jockey-3_DEu	.	Drosophila eugracilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3_DF	.	Drosophila ficusphila	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3_DGri	.	Drosophila grimshawi	Jockey 87,88
Jockey-3_DK	.	Drosophila kikkawai	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3_DPer	.	Drosophila persimilis	Jockey 87,2,11,1,4,15,16,3,88,174
Jockey-3_DRh	.	Drosophila rhopaloa	Jockey 16,1,87,3,88,24,11,4,2
Jockey-3_DSim	.	Drosophila simulans	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3_DT	.	Drosophila takahashii	Jockey 1,87,88,3,11,4,2
Jockey-3_DVi	.	Drosophila virilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3_DYa	.	Drosophila yakuba	Jockey 16,1,3,11,4,2,15
Jockey-3_Hmel	.	Heliconius melpomene	Jockey 16,1,3,11,4,2,15
Jockey-3_HMM	.	Heliconius melpomene melpomene	Jockey 16,1,3,11,4,2,15
Jockey-3_LMi	.	Locusta migratoria	Jockey 1,3,4,2
Jockey-3_RPr	.	Rhodnius prolixus	Jockey 16,1,87,3,88,11,4,2,15
Jockey-3B_LMi	.	Locusta migratoria	Jockey
Jockey-4_BTa	.	Bemisia tabaci	Jockey 16,1,3,11,4,2,15
Jockey-4_CQ	.	Culex quinquefasciatus	Jockey 16,1,3,11,4,2,15
Jockey-4_DAn	.	Drosophila ananassae	Jockey 16,1,87,3,88,24,11,4,2

Jockey-4_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DEu	.	Drosophila eugracilis	Jockey	16,1,87,14,3,88,11,4,2,15
Jockey-4_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DK	.	Drosophila kikkawai	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DPer	.	Drosophila persimilis	Jockey	16,1,87,88,3,11,4,2,15
Jockey-4_DRh	.	Drosophila rhopaloa	Jockey	16,1,3,11,4,2,15
Jockey-4_DSim	.	Drosophila simulans	Jockey	1,3,4,2
Jockey-4_DT	.	Drosophila takahashii	Jockey	
Jockey-4_DVi	.	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DYa	.	Drosophila yakuba	Jockey	1,3,4,2
Jockey-4_Hmel	.	Heliconius melpomene	Jockey	16,1,3,11,4,2,15
Jockey-4_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_LMi	.	Locusta migratoria	Jockey	1,3,4,2
Jockey-4_RPr	.	Rhodnius prolixus	Jockey	16,1,3,11,4,2,15
Jockey-4B_DVi	.	Drosophila virilis	Jockey	1,87,3,88,11,4,2
Jockey-5_BTa	.	Bemisia tabaci	Jockey	16,131,1,3,11,4,132,2,15
Jockey-5_CQ	.	Culex quinquefasciatus	Jockey	16,1,14,3,11,4,2,15
Jockey-5_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DEu	.	Drosophila eugracilis	Jockey	1,3,4,2
Jockey-5_DF	.	Drosophila ficusphila	Jockey	16,1,3,11,4,2,15

Jockey-5_DK	.	Drosophila kikkawai	Jockey 1,3,4,2
Jockey-5_DPer	.	Drosophila persimilis	Jockey 16,1,87,88,3,11,4,2,15
Jockey-5_DTa	.	Drosophila takahashii	Jockey 16,1,87,3,88,11,4,2,15
Jockey-5_DVi	.	Drosophila virilis	Jockey 87,2,119,11,1,4,15,16,3,88
Jockey-5_DYa	.	Drosophila yakuba	Jockey 16,1,3,4,11,2,15
Jockey-5_Hmel	.	Heliconius melpomene	Jockey 1,3,4,2
Jockey-5_LMi	.	Locusta migratoria	Jockey 1,3,4,2
Jockey-5_RPr	.	Rhodnius prolixus	Jockey 1,3,4,2
Jockey-5B_HMM	.	Heliconius melpomene melpomene	Jockey 1,3,4,2
Jockey-6_BTa	.	Bemisia tabaci	Jockey 16,131,1,3,4,11,132,2,15
Jockey-6_CQ	.	Culex quinquefasciatus	Jockey 16,1,3,11,4,2,15
Jockey-6_DAn	.	Drosophila ananassae	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DBp	.	Drosophila bipectinata	Jockey 16,1,3,11,4,2,15
Jockey-6_DEl	.	Drosophila elegans	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DEu	.	Drosophila eugracilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DF	.	Drosophila ficusphila	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DK	.	Drosophila kikkawai	Jockey 16,1,3,11,4,2,15
Jockey-6_DPer	.	Drosophila persimilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DRh	.	Drosophila rhopaloa	Jockey 16,1,3,11,4,2,15
Jockey-6_DTa	.	Drosophila takahashii	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DVi	.	Drosophila virilis	Jockey 16,1,87,3,88,11,4,2,15
Jockey-6_DYa	.	Drosophila yakuba	Jockey 16,1,3,4,11,2,15
Jockey-6_Hmel	.	Heliconius melpomene	Jockey 1,3,4,2

Jockey-6_LMi	.	Locusta migratoria	Jockey	
Jockey-6_RPr	.	Rhodnius prolixus	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_BTa	.	Bemisia tabaci	Jockey	16,1,3,82,11,4,2,15
Jockey-7_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-7_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DEl	.	Drosophila elegans	Jockey	16,1,3,4,11,2,15
Jockey-7_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DK	.	Drosophila kikkawai	Jockey	1,3,4,2
Jockey-7_DPer	.	Drosophila persimilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DVi	.	Drosophila virilis	Jockey	1,3,4,2
Jockey-7_DYa	.	Drosophila yakuba	Jockey	16,1,3,11,4,2,15
Jockey-7_Hmel	.	Heliconius melpomene	Jockey	16,1,3,11,4,2,15
Jockey-7_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15
Jockey-7_LMi	.	Locusta migratoria	Jockey	1,3,11,4,2
Jockey-7_RPr	.	Rhodnius prolixus	Jockey	16,87,88,11,15
Jockey-8_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-8_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-8_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-8_DEu	.	Drosophila eugracilis	Jockey	
Jockey-8_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-8_DPer	.	Drosophila persimilis	Jockey	16,1,3,11,4,2,15

| | | |

Jockey-N6B_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N6C_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N7_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N8_CQ	.	Culex quinquefasciatus	Jockey		14
JOCKEY2	.	Drosophila melanogaster	Jockey	1,3,4,2	
JVY_As	.	Adineta	Penelope		82
JY_As	.	Adineta	Penelope		
K_As	.	Adineta	Penelope	1,3	
K_Av	.	Adineta vaga	Penelope	1,3	
KARIN_TA	AF325197	Triticum aestivum	L1		
KARIN_TM	AY485644	Triticum monococcum	Не содержащий LTR ретротранспозон	1,43,3,44,11,4,2	
Keno-1_GA	.	Gasterosteus aculeatus	Tx1	43,2,8,44,11,12,1,26,14,4,3	
Keno-1_SSa	.	Salmo salar	Tx1	1,43,3,44,11,4,2	
Keno-5_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3	
Keno-6_XT	.	Xenopus tropicalis	Tx1	1,43,3,44,4,2	
Keno-7_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,1,26,14,4,15,16,3	
Keno-8_XT	.	Xenopus tropicalis	Tx1		
KenoDr1	AB097144	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoDr2	.	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoDr3	.	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoFr1	AB111947	Takifugu rubripes	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46	
Kibi-1_GA	.	Gasterosteus aculeatus	Tx1	1,26,3,11,4,2	
KibiDr1	.	Danio rerio	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3	

KibiDr2	.	Danio rerio	Tx1	1,26,3,11,4,2
KibiFr1	AB097136	Takifugu rubripes	Не содержащий LTR ретротранспозон	1,26,3,11,4,2
KibiTn1	AB097136	Tetraodon nigroviridis	L1	1,26,43,3,44,11,4,2
Kiri-1_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-1_CQ	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-1_DBi	.	Drosophila biarmipes	Kiri	1,3,4,2
Kiri-1_DEl	.	Drosophila elegans	Kiri	1,3,11,4,2
Kiri-1_DF	.	Drosophila ficusphila	Kiri	1,3,4,2
Kiri-1_DK	.	Drosophila kikkawai	Kiri	16,1,3,24,11,4,2
Kiri-1_DPer	.	Drosophila persimilis	Kiri	1,3,11,4,2
Kiri-1_DRh	.	Drosophila rhopaloa	Kiri	1,3,11,4,2
Kiri-1_DTa	.	Drosophila takahashii	Kiri	1,3,4,2
Kiri-1_HMM	.	Heliconius melpomene melpomene	Kiri	21,1,3,4,2
Kiri-1_LMi	.	Locusta migratoria	Kiri	1,3,11,4,2
Kiri-1_PPo	.	Papilio polytes	Kiri	
Kiri-1_PXu	.	Papilio xuthus	Kiri	
Kiri-10_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-10_CQ	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-10_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,11,4,2
Kiri-11_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-11_CQ	AAWU01038353	Culex quinquefasciatus	Kiri	53,2,54,11,56,57,68,1,58,4,127,3,59
Kiri-12_AAe	.	Aedes aegypti	Kiri	1,3,4,11,2
Kiri-12_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2

Kiri-13_AAe	Aedes aegypti	Kiri	16,1,3,24,4,11,2
Kiri-13_HMM	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-14_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-15_AAe	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-16_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-17_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-17_HMM	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-18_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-19_AAe	Aedes aegypti	Kiri	16,1,3,4,11,2,15
Kiri-2_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-2_CQ	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-2_DEl	Drosophila elegans	Kiri	16,1,3,24,11,4,2
Kiri-2_DK	Drosophila kikkawai	Kiri	16,1,3,24,11,4,2
Kiri-2_DPer	Drosophila persimilis	Kiri	1,3,11,4,2
Kiri-2_DTa	Drosophila takahashii	Kiri	16,1,3,24,11,4,2
Kiri-2_HMM	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-2_LMi	Locusta migratoria	Kiri	1,3,4,2
Kiri-2_PPo	Papilio polytes	Kiri	
Kiri-20_AAe	Aedes aegypti	Kiri	1,3,4,11,2
Kiri-20_HMM	Heliconius melpomene melpomene	Kiri	1,3,11,4,2
Kiri-21_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-22_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-23_AAe	Aedes aegypti	Kiri	16,1,3,24,4,11,2

Kiri-24_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-25_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-26_AAe	Aedes aegypti	Kiri	16,1,3,4,11,2,15
Kiri-27_AAe	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-28_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-29_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-3_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-3_CQ	Culex quinquefasciatus	Kiri	16,1,3,24,11,4,2
Kiri-3_DK	Drosophila kikkawai	Kiri	1,3,11,4,2
Kiri-3_DTa	Drosophila takahashii	Kiri	16,1,3,24,11,4,2
Kiri-3_HMM	Heliconius melpomene melpomene	Kiri	21,1,3,11,4,2
Kiri-3_LMi	Locusta migratoria	Kiri	1,3,4,2
Kiri-3_PPo	Papilio polytes	Kiri	
Kiri-30_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-31_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-32_AAe	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-33_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-34_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-35_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-36_AAe	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-37_AAe	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-38_AAe	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-4_AAe	Aedes aegypti	Kiri	1,3,4,11,2

Kiri-4_CQ	.	Culex quinquefasciatus	Kiri	53,2,54,11,56,57,68,1,58,4,127,69,3
Kiri-4_DK	.	Drosophila kikkawai	Kiri	1,3,11,4,2
Kiri-4_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-4_LMi	.	Locusta migratoria	Kiri	
Kiri-5_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-5_CQ	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-5_DK	.	Drosophila kikkawai	Kiri	1,3,4,2
Kiri-5_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-5_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-5_LMi	.	Locusta migratoria	Kiri	1,3,4,2
Kiri-6_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-6_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,2
Kiri-6_DK	.	Drosophila kikkawai	Kiri	16,1,3,4,11,2,15
Kiri-6_DTa	.	Drosophila takahashii	Kiri	1,3,4,11,2
Kiri-7_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-7_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,2
Kiri-7_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-7_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-8_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-8_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,11,2
Kiri-8B_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-9_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-9_CQ	.	Culex quinquefasciatus	Kiri	16,1,3,11,4,2,15

KoshiTn1	AB097135	Tetraodon nigroviridis	Tx1	2,8,11,12,42,1,26,14,4,3
L1	.	Homo sapiens	L1	
L1_AC_1	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_10	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1_AC_11	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_12	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_13	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_14	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1_AC_15	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1_AC_16	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1_AC_17	.	Anolis carolinensis	L1	21,16,1,26,3,24,11,4,2
L1_AC_18	.	Anolis carolinensis	L1	2,21,11,1,26,28,4,3
L1_AC_19	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1_AC_2	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_20	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1_AC_3	.	Anolis carolinensis	L1	1,26,3,4,11,2
L1_AC_3B	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_4	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_5	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1_AC_6	.	Anolis carolinensis	L1	16,1,26,23,3,4,11,2,15
L1_AC_7	.	Anolis carolinensis	L1	1,3,11,4,2
L1_AC_8	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1_AC_9	.	Anolis carolinensis	L1	1,26,23,3,11,4,2

L1_As	.	Adineta	Penelope	1,3,130
L1_Av	.	Adineta vaga	Penelope	1,3
L1_BT	.	Bovidae	L1	
L1_CA	.	Cicer arietinum	L1	
L1_Canid_5-конец	.	Canidae	L1	
L1_Canid2_5-конец	.	Canidae	L1	
L1_Carn1_5-конец	.	Carnivora	L1	
L1_Carn2_5-конец	.	Carnivora	L1	
L1_Carn2_orf2	.	Carnivora	L1	
L1_Carn3_5-конец	.	Carnivora	L1	
L1_Carn5_3-конец	.	Carnivora	L1	
L1_Carn5_5-конец	.	Carnivora	L1	
L1_Carni_orf2	.	Carnivora	L1	
L1_CC	Y00726	Cavia cutleri	L1	
L1_CF	AB012223	Canis lupus familiaris	L1	
L1_CP	M60812	Conilurus penicillatus	L1	
L1_Ele12	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1_Ele13	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele16	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele17	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1_Ele19	.	Aedes aegypti	L1	93,1,26,3,11,4,2
L1_Ele21	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1_Ele26	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2

L1_Ele28	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1_Ele3	.	Aedes aegypti	L1	16,1,26,3,24,11,4,92,2
L1_Ele3B_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,92,2
L1_Ele3C_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele4	.	Aedes aegypti	L1	1,26,3,11,4,2
L1_Ele5	.	Aedes aegypti	L1	43,2,44,24,11,93,1,26,4,16,3
L1_Ele9	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1_Fc	.	Felis	L1	
L1_Fc_5-конец	.	Felis	L1	
L1_Fc_orf2	.	Felis	L1	
L1_Felid	.	Felidae	L1	
L1_Lme	.	Latimeria menadoensis	L1	1,3,4,2
L1_Mars1_orf2	.	Metatheria	L1	
L1_Mars1a_3-конец	.	Metatheria	L1	
L1_Mars1b_3-конец	.	Metatheria	L1	
L1_Mdo3b_3-конец	.	Monodelphis	L1	
L1_MM	M29324	Mus musculus	L1	
L1_Mur1_5-конец	.	Muridae	L1	
L1_Mur1_orf2	.	Muridae	L1	
L1_Mur2_3-конец	.	Muridae	L1	
L1_Mur2_5-конец	.	Muridae	L1	
L1_Mur2_orf2	.	Muridae	L1	
L1_Mur3_5-конец	.	Muridae	L1	

L1_Mur3_orf2	.	Muridae	L1	
L1_Mus1_3-конец	.	Mus	L1	
L1_Mus2_3-конец	.	Mus	L1	
L1_Opos_5-конец	.	Didelphidae	L1	
L1_Opos3_orf2	.	Didelphidae	L1	
L1_Rat3_3-конец	.	Rattus	L1	
L1_Rat4_3-конец	.	Rattus	L1	
L1_RN	.	Rattus norvegicus	L1	
L1_Rn2_3-конец	.	Rattus	L1	
L1_Rod_5-конец	.	Muridae	L1	
L1_Rod_orf2	.	Muridae	L1	
L1_RS1_5-конец	.	Cercopithecinae	L1	
L1_SS	.	Suina	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1_TC	.	Trypanosoma cruzi	Ingi	6,2,18,10,11,1,4,15,5,16,17,3,7
L1_TD	.	Triticum turgidum subsp. durum	L1	
L1-1_AB	.	Agaricus bisporus	L1	2,10,11,13,1,26,4,17,3,7
L1-1_Acar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-1_AFC	.	Cichlidae	L1	16,1,26,24,11,2
L1-1_AIp	.	Arachis ipaensis	L1	43,79,2,44,10,11,80,1,77,4,75,17,3,76,7
L1-1_ALai	.	Albugo laibachii	L1	1,3,11,4,2
L1-1_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-1_AMe	.	Ailuropoda melanoleuca	L1	25,21,20,23,28,30
L1-1_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7

L1-1_BDi	.	Brachypodium distachyon	L1	43,2,44,11,12,1,14,4,3
L1-1_BRa	.	Brassica rapa	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-1_BV	.	Beta vulgaris subsp. vulgaris	L1	43,78,79,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7
L1-1_CCr	.	Condylura cristata	L1	31,25,21,20,23,27,28,30
L1-1_CGr	.	Cricetulus griseus	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-1_Cho	.	Choloepus hoffmanni	L1	16,1,26,3,24,4,11,2
L1-1_Cja	.	Callithrix jacchus	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-1_CPa	.	Carica papaya	L1	43,78,79,6,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
L1-1_CPB	.	Chrysemys picta bellii	L1	
L1-1_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_CQ	.	Culex quinquefasciatus	L1	93,16,1,26,3,24,11,4,2
L1-1_CR	.	Chlamydomonas reinhardtii	L1	16,1,3,24,11,4,2
L1-1_Crp	.	Crocodylus porosus	L1	
L1-1_DN	.	Dasypus novemcinctus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_Dor	.	Dipodomys ordii	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-1_DR	AL645691	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-1_DV	.	Didelphis virginiana	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76
L1-1_EE	.	Erinaceus europaeus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_EL	.	Esox lucius	L1	
L1-1_EPa	.	Schrenkiella parvula	L1	
L1-1_ET	.	Echinops telfairi	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3

L1-1_FV	.	Fragaria vesca	L1	2,43,80,17,1,3,44,10,79,4,7
L1-1_GAr	.	Gossypium arboreum	L1	43,79,2,44,10,24,11,80,1,4,16,17,3,7
L1-1_GM	.	Glycine max	L1	43,78,2,44,11,1,81,77,4,75,3,76
L1-1_GR	.	Gossypium raimondii	L1	43,78,2,44,10,11,80,1,81,77,4,17,75,3,76
L1-1_HAE	.	Hyaloperonospora arabidopsidis	L1	6,2,10,11,80,1,26,230,4,17,3,7
L1-1_HArab	.	Hyaloperonospora arabidopsidis	L1	1,3,4,2
L1-1_HM	.	Hydra vulgaris	L1	16,1,26,3,24,11,4,2
L1-1_JC	.	Jatropha curcas	L1	1,3,11,4,2
L1-1_LA	.	Loxodonta africana	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_LA-5-конец	.	Loxodonta africana	L1	
L1-1_LCh	.	Latimeria chalumnae	L1	
L1-1_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-1_LU	.	Linum usitatissimum	L1	1,43,3,44,4,11,2
L1-1_MAc	.	Musa acuminata	L1	1,43,75,3,76,44,77,11,4,2
L1-1_Mad	.	Malus domestica	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-1_MD	.	Monodelphis domestica	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-1_ME	.	Notamacropus eugenii	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_MN	.	Morus notabilis	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-1_Mpr	.	Mammuthus primigenius	L1	
L1-1_MPu	.	Mustela putorius furo	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-1_NS	.	Nicotiana sylvestris	L1	1,43,3,44,11,4,2
L1-1_NV	.	Nematostella vectensis	L1	

L1-1_OAf	.	Orycteropus afer	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_OES	.	Olea europaea subsp. europaea	L1	43,79,6,2,7,10,44,24,11,80,1,77,4,16,75,17,3,76
L1-1_OL	.	Oryzias latipes	L1	2,21,23,24,11,1,26,4,16,3
L1-1_OP	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_OSp	.	Pecoramyces ruminatium	L1	1,3,4,11,2
L1-1_PAb	.	Picea abies	L1	43,6,2,7,8,10,44,11,12,80,1,14,4,75,17,3,76
L1-1_PAlni	.	Phytophthora alni	L1	1,11,2
L1-1_PApha	.	Pythium aphanidermatum	L1	
L1-1_PArrh	.	Pythium arrhenomanes	L1	1,3,4,2
L1-1_Pca	.	Procavia capensis	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_PCaps	.	Phytophthora capsici	L1	6,2,10,11,80,1,230,4,17,3,7
L1-1_PCinn	.	Phytophthora cinnamomi	L1	7,1,17,3,10,11,4,6,2,80
L1-1_PD	.	Phoenix dactylifera	L1	
L1-1_PI	.	Phytophthora infestans	L1	1,8,3,11,4,12,2
L1-1_PInsi	.	Pythium insidiosum	L1	6,2,10,11,80,1,230,4,17,3,7
L1-1_PIway	.	Pythium iwayamai	L1	2,1
L1-1_PLate	.	Phytophthora lateralis	L1	
L1-1_PPini	.	Phytophthora pinifolia	L1	6,2,10,24,11,80,1,26,4,16,17,3,7
L1-1_PR	.	Phytophthora ramorum	L1	175,1,176,3,120,11,4,177,2
L1-1_PRamo	.	Phytophthora ramorum	L1	6,2,10,11,80,1,26,4,17,3,7
L1-1_PSoja	.	Phytophthora sojae	L1	
L1-1_PTr	.	Populus trichocarpa	L1	43,78,2,44,11,1,81,77,4,75,3,76
L1-1_PTr	.	Populus trichocarpa	L1	1,43,3,44,2

L1-1_PUlti	.	Pythium ultimum L1	6,2,8,10,11,12,80,1,230,4,17,3,7
L1-1_PVa	.	Pteropus vampyrus L1	20,2,21,23,25,1,27,28,4,30,31,3
L1-1_RC	.	Ricinus communis L1	
L1-1_Sar	.	Sorex araneus L1	85,20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_SBi	.	Sorghum bicolor L1	43,2,8,44,24,11,12,1,14,4,16,3
L1-1_SH	.	Sarcophilus harrisii L1	
L1-1_SMo	.	Selaginella moellendorffii L1	16,1,43,3,44,24,11,4,2
L1-1_SP	.	Strongylocentrotus purpuratus L1	1,26,43,44,11,4,2
L1-1_SSa	.	Salmo salar L1	2,21,24,11,1,26,28,4,16,3
L1-1_SSc	.	Sus scrofa L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-1_Str	.	Ictidomys tridecemlineatus L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_STu	.	Solanum tuberosum L1	43,79,6,2,8,44,10,11,12,80,1,77,4,17,75,3,76,7
L1-1_Tbel	.	Tupaia belangeri L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_TC	.	Theobroma cacao L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-1_TS	.	Carlito syrichta L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-1_Ttr	.	Tursiops truncatus L1	16,1,26,3,24,4,11,2
L1-1_VC	.	Volvox carteri L1	16,1,26,3,24,11,4,2
L1-1_Vpa	.	Vicugna pacos L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_VVi	.	Vitis vinifera L1	16,1,43,3,44,24,11,4,2
L1-1_XL	.	Xenopus laevis L1	2,21,24,11,1,26,28,4,16,3
L1-1_XT	.	Xenopus tropicalis L1	2,21,11,1,26,28,4,30,15,16,3
L1-1_ZM	.	Zea mays L1	43,2,8,44,24,11,12,1,14,4,16,3
L1-10_AIp	.	Arachis ipaensis L1	43,78,2,44,11,1,81,77,4,75,3,7

L1-10_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-10_ATr	.	Amborella trichopoda	L1	1,3,4,11,2
L1-10_BDi	.	Brachypodium distachyon	L1	16,1,43,3,44,24,11,4,2
L1-10_CPa	.	Carica papaya	L1	1,43,3,44,11,4,2
L1-10_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-10_DR	AL772154	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-10_ECa	.	Eucalyptus camaldulensis	L1	43,78,79,2,8,44,10,11,12,80,1,81,77,4,75,17,3,76,7
L1-10_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-10_GAr	.	Gossypium arboreum	L1	7,1,43,17,3,44,10,79,4,2,80
L1-10_GM	.	Glycine max	L1	16,1,43,75,3,76,44,24,11,4,2
L1-10_JC	.	Jatropha curcas	L1	1,43,17,44,10,79,4,2,6,7,80
L1-10_LCh	.	Latimeria chalumnae	L1	1,3,11,4,2
L1-10_LSal	.	Lepeophtheirus salmonis	L1	
L1-10_Mad	.	Malus domestica	L1	
L1-10_MD	.	Monodelphis domestica	L1	
L1-10_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-10_NS	.	Nicotiana sylvestris	L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-10_PI	.	Phytophthora infestans	L1	
L1-10_SSa	.	Salmo salar	L1	2,21,23,11,1,26,4,3
L1-10_STu	.	Solanum tuberosum	L1	16,1,75,43,76,3,44,24,11,4,2
L1-10_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-100_CSav	.	Ciona savignyi	L1	1,14,3,11,4,12,2
L1-100_DR	.	Danio rerio	L1	1,2

L1-101_DR	.	Danio rerio	L1	
L1-102_DR	.	Danio rerio	L1	1,3,4,2
L1-103_DR	.	Danio rerio	L1	
L1-104_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-105_DR	.	Danio rerio	L1	
L1-106_DR	.	Danio rerio	L1	
L1-107_DR	.	Danio rerio	L1	20,2,21,23,11,25,1,28,4
L1-108_DR	.	Danio rerio	L1	16,21,1,26,3,24,11,4,2
L1-109_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-10a_MD	.	Monodelphis domestica	L1	
L1-11_AIp	.	Arachis ipaensis	L1	43,91,92,2,44,11,93,1,4,75,3,76
L1-11_ALy	.	Arabidopsis lyrata	L1	1,43,3,44,4,11,2
L1-11_ATr	.	Amborella trichopoda	L1	7,1,17,43,3,44,11,4,2
L1-11_BDi	.	Brachypodium distachyon	L1	14,12
L1-11_CPa	.	Carica papaya	L1	1,3,4,2
L1-11_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-11_ECa	.	Eucalyptus camaldulensis	L1	43,78,79,6,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76
L1-11_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-11_GM	.	Glycine max	L1	43,91,92,2,44,11,93,1,96,97,4,75,3,76,98
L1-11_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2
L1-11_LSal	.	Lepeophtheirus salmonis	L1	
L1-11_Mad	.	Malus domestica	L1	43,78,79,6,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7
L1-11_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7

L1-11_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,10,44,11,80,1,4,75,17,3,76
L1-11_PI	.	Phytophthora infestans	L1	7,1,17,3,10,4,6,2,80
L1-11_SSa	.	Salmo salar	L1	1,3,11,4,2
L1-11_STu	.	Solanum tuberosum	L1	
L1-11_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,15,16,3
L1-110_DR	.	Danio rerio	L1	
L1-111_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-112_DR	.	Danio rerio	L1	1,28,4,2
L1-113_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-114_DR	.	Danio rerio	L1	1,4,2
L1-115_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-116_DR	.	Danio rerio	L1	
L1-117_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-118_DR	.	Danio rerio	L1	1,3,11,4,2
L1-119_DR	.	Danio rerio	L1	21,1,23,11,4,2
L1-12_AIp	.	Arachis ipaensis	L1	1,43,3,44,11,4,2
L1-12_ALy	.	Arabidopsis lyrata	L1	1,3,11,4,2
L1-12_ATr	.	Amborella trichopoda	L1	1,3,11,4,2
L1-12_BDi	.	Brachypodium distachyon	L1	2,8,24,11,12,42,1,14,4,16,3
L1-12_CPa	.	Carica papaya	L1	44,43
L1-12_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-12_ECa	.	Eucalyptus camaldulensis	L1	43,78,79,6,2,7,8,10,44,24,11,12,80,1,14,81,77,4,16,75,17
L1-12_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80

L1-12_GM	.	Glycine max	L1	
L1-12_LCh	.	Latimeria chalumnae	L1	20,2,21,23,24,11,25,1,26,28,4,16,3
L1-12_LSal	.	Lepeophtheirus salmonis	L1	
L1-12_Mad	.	Malus domestica	L1	
L1-12_NN	.	Nelumbo nucifera	L1	43,79,6,2,44,10,11,80,1,4,100,17,3,7
L1-12_NS	.	Nicotiana sylvestris	L1	
L1-12_PI	.	Phytophthora infestans	L1	7,1,17,3,10,11,4,6,2,80
L1-12_SSa	.	Salmo salar	L1	2,21,23,11,1,26,28,30,3
L1-12_STu	.	Solanum tuberosum	L1	16,1,43,75,3,76,44,24,11,4,2
L1-12_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-120_DR	.	Danio rerio	L1	1,3,11,4,2
L1-121_DR	.	Danio rerio	L1	
L1-122_DR	.	Danio rerio	L1	1,26,3,11,4,2
L1-123_DR	.	Danio rerio	L1	44,43
L1-124_DR	.	Danio rerio	L1	1,4,2
L1-125_DR	.	Danio rerio	L1	1,3,4,2
L1-126_DR	.	Danio rerio	L1	16,1,26,3,24,4,11,2
L1-127_DR	.	Danio rerio	L1	
L1-128_DR	.	Danio rerio	L1	
L1-129_DR	.	Danio rerio	L1	
L1-12A_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-12B_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-13_AIp	.	Arachis ipaensis	L1	1,43,75,3,44,76,11,4,2

L1-13_ALy	.	Arabidopsis lyrata	L1	
L1-13_ATr	.	Amborella trichopoda	L1	1,3,4,2
L1-13_BDi	.	Brachypodium distachyon	L1	
L1-13_CPa	.	Carica papaya	L1	1,3,4,2
L1-13_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-13_FV	.	Fragaria vesca	L1	43,80,17,44,10,79,6,7
L1-13_GM	.	Glycine max	L1	
L1-13_LCh	.	Latimeria chalumnae	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-13_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,2
L1-13_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-13_NN	.	Nelumbo nucifera	L1	43,79,6,2,44,10,24,11,80,1,4,100,16,17,3,7
L1-13_NS	.	Nicotiana sylvestris	L1	
L1-13_SSa	.	Salmo salar	L1	16,1,26,3,24,4,11,2
L1-13_STu	.	Solanum tuberosum	L1	16,1,43,75,76,44,24,11,4,2
L1-13_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3
L1-130_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-131_DR	.	Danio rerio	L1	
L1-132_DR	.	Danio rerio	L1	1,3,11,4,2
L1-133_DR	.	Danio rerio	L1	
L1-134_DR	.	Danio rerio	L1	
L1-135_DR	.	Danio rerio	L1	
L1-136_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,3
L1-137_DR	.	Danio rerio	L1	21,16,1,26,23,24,11,2

L1-138_DR	.	Danio rerio	L1	
L1-139_DR	.	Danio rerio	L1	
L1-13B_SSa	.	Salmo salar	L1	
L1-14_AIp	.	Arachis ipaensis	L1	1,75,43,3,76,44,11,4,2
L1-14_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-14_ATr	.	Amborella trichopoda	L1	1,43,3,44,11,4,2
L1-14_BDi	.	Brachypodium distachyon	L1	1,3,11,4,2
L1-14_CPa	.	Carica papaya	L1	16,1,43,3,44,11,4,2,15
L1-14_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,15,16,3
L1-14_FV	.	Fragaria vesca	L1	43,80,17,10,44,79,6,7
L1-14_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-14_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-14_Mad	.	Malus domestica	L1	
L1-14_NN	.	Nelumbo nucifera	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-14_NS	.	Nicotiana sylvestris	L1	
L1-14_SSa	.	Salmo salar	L1	21,1,3,28,11,4,2
L1-14_STu	.	Solanum tuberosum	L1	16,1,43,75,3,76,44,24,11,4,2
L1-14_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,15,16,3
L1-140_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-15_ALy	.	Arabidopsis lyrata	L1	
L1-15_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-15_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-15_CPa	.	Carica papaya	L1	1,43,3,44,4,2

L1-15_DR	.	Danio rerio	L1	1,26,8,14,3,11,4,12,2
L1-15_FV	.	Fragaria vesca	L1	43,80,17,10,44,79,6,7
L1-15_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-15_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-15_Mad	.	Malus domestica	L1	
L1-15_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,15,16,17,3,7
L1-15_NS	.	Nicotiana sylvestris	L1	
L1-15_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-15_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-15_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-15a_Lch	.	Latimeria chalumnae	L1	
L1-15B_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-15b_Lch	.	Latimeria chalumnae	L1	
L1-15c_Lch	.	Latimeria chalumnae	L1	
L1-16_ALy	.	Arabidopsis lyrata	L1	43,80,17,44,10,79,7
L1-16_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,7
L1-16_BDi	.	Brachypodium distachyon	L1	
L1-16_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-16_DR	.	Danio rerio	L1	16,1,26,3,24,11,4,2
L1-16_FV	.	Fragaria vesca	L1	1,43,3,44,4,2
L1-16_GMa	.	Glycine max	L1	16,1,3,24,11,4,2
L1-16_Lch	.	Latimeria chalumnae	L1	
L1-16_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2

L1-16_Mad	.	Malus domestica	L1	1,43,3,44,10,11,4,2,80
L1-16_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-16_NS	.	Nicotiana sylvestris	L1	
L1-16_SSa	.	Salmo salar	L1	1,3,4,11,2
L1-16_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-16_XT	.	Xenopus tropicalis	L1	1,3,11,4,2
L1-17_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-17_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-17_BDi	.	Brachypodium distachyon	L1	16,1,3,11,4,2,15
L1-17_CPa	.	Carica papaya	L1	44,43
L1-17_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-17_FV	.	Fragaria vesca	L1	17,10,79,6,7,80
L1-17_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,2
L1-17_Mad	.	Malus domestica	L1	7,6,1,80,17,43,3,44,10,79,4,2
L1-17_NN	.	Nelumbo nucifera	L1	1,43,3,44,11,4,2
L1-17_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,8,10,44,11,12,80,1,4,75,17,3,76
L1-17_SSa	.	Salmo salar	L1	16,26,24,11
L1-17_STu	.	Solanum tuberosum	L1	
L1-17_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,4,11,2
L1-17a_Lch	.	Latimeria chalumnae	L1	
L1-17b_Lch	.	Latimeria chalumnae	L1	
L1-18_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-18_ATr	.	Amborella trichopoda	L1	7,1,43,17,3,44,10,79,4,6,2,80

L1-18_BDi	.	Brachypodium distachyon	L1	
L1-18_CPa	.	Carica papaya	L1	16,1,3,24,11,4,2
L1-18_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-18_FV	.	Fragaria vesca	L1	17,10,79,6,7,80
L1-18_Lch	.	Latimeria chalumnae	L1	
L1-18_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-18_Mad	.	Malus domestica	L1	
L1-18_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-18_NS	.	Nicotiana sylvestris	L1	
L1-18_SSa	.	Salmo salar	L1	2,21,24,11,1,26,28,4,30,16,3
L1-18_STu	.	Solanum tuberosum	L1	43,78,79,6,2,7,10,44,11,80,1,81,77,4,75,17,3,76
L1-18_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-19_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-19_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-19_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-19_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-19_FV	.	Fragaria vesca	L1	43,80,17,44,10,79,6,7
L1-19_Lch	.	Latimeria chalumnae	L1	
L1-19_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-19_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-19_SSa	.	Salmo salar	L1	21,1,26,23,3,11,4,2
L1-19_STu	.	Solanum tuberosum	L1	16,1,75,43,3,76,44,24,4,11,2
L1-19_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3

L1-1a_Cis	.	Ciona savignyi	L1	1,14,3,11,4,12,2
L1-1A_Cpo	.	Cavia porcellus	L1	
L1-1a_EF	.	Eptesicus fuscus	L1	
L1-1a_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1A_VC	.	Volvox carteri	L1	16,1,26,3,24,4,11,2
L1-1A2_Sar	.	Sorex araneus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1B_Cho	.	Choloepus hoffmanni	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1b_Cis	.	Ciona savignyi	L1	
L1-1B_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1B_LA-5-конец	.	Loxodonta africana	L1	
L1-1B_MAc	.	Musa acuminata	L1	1,75,43,3,76,44,77,11,4,2
L1-1b_MD	.	Monodelphis domestica	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-1B_Opr	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1b_TS	.	Carlito syrichta	L1	
L1-1C_Cho	.	Choloepus hoffmanni	L1	
L1-1C_Cpo	.	Cavia porcellus	L1	
L1-1D_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3,39
L1-1E_Cpo	.	Cavia porcellus	L1	
L1-1F_Cpo	.	Cavia porcellus	L1	
L1-1G_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1H_Cpo	.	Cavia porcellus	L1	
L1-1N_Tbel	.	Tupaia belangeri	L1	25,21,20,23,28,30
L1-2_AIp	.	Arachis ipaensis	L1	43,79,2,7,10,44,11,80,1,77,4,75,17,3,76

L1-2_ALai	.	Albugo laibachii	L1	1,3,11,4,2
L1-2_ALy	.	Arabidopsis lyrata	L1	
L1-2_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-2_BDi	.	Brachypodium distachyon	L1	1,11,4,2
L1-2_BRa	.	Brassica rapa	L1	16,1,43,3,44,24,4,11,2
L1-2_BT	.	Bos taurus	L1	22,29
L1-2_BV	.	Beta vulgaris subsp. vulgaris	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-2_CGr	.	Cricetulus griseus	L1	2,22,24,11,1,26,4,29,16,3
L1-2_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
L1-2_Cis	.	Ciona savignyi	L1	
L1-2_Cja	.	Callithrix jacchus	L1	
L1-2_CPa	.	Carica papaya	L1	16,1,43,3,44,24,11,4,2
L1-2_CPB	.	Chrysemys picta bellii	L1	
L1-2_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-2_CR	.	Chlamydomonas reinhardtii	L1	16,1,26,24,4,11,2
L1-2_Crp	.	Crocodylus porosus	L1	
L1-2_Dor	.	Dipodomys ordii	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_DR	.	Danio rerio	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
L1-2_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_ECa	.	Eucalyptus camaldulensis	L1	1,75,43,3,76,44,77,11,4,2
L1-2_EE	.	Erinaceus europaeus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-2_EPa	AFAN01000035	Schrenkiella parvula	L1	1,43,78,3,44,81,77,11,4,2

L1-2_Fc	.	Felis catus	L1	25,21,20,23,28,30
L1-2_FR	.	Takifugu rubripes	L1	16,1,26,3,24,11,4,2
L1-2_FV	.	Fragaria vesca	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-2_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,24,11,80,1,4,16,17,3
L1-2_GM	.	Glycine max	L1	16,1,43,75,76,44,24,11,2
L1-2_HAE	.	Hyaloperonospora arabidopsidis	L1	6,2,10,11,80,1,4,17,3,7
L1-2_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-2_JC	.	Jatropha curcas	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-2_LA	.	Loxodonta africana	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-2_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2
L1-2_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-2_MAc	.	Musa acuminata	L1	
L1-2_Mad	.	Malus domestica	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-2_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_ME	.	Notamacropus eugenii	L1	2,21,23,24,11,1,26,4,30,16,3
L1-2_MPu	.	Mustela putorius furo	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-2_NS	.	Nicotiana sylvestris	L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-2_NV	.	Nematostella vectensis	L1	
L1-2_Opr	.	Ochotona princeps	L1	84,20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_PAb	.	Picea abies	L1	16,1,3,24,4,11,2
L1-2_Pca	.	Procavia capensis	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_PD	.	Phoenix dactylifera	L1	7,16,1,17,43,3,44,24,11,4,2

L1-2_PI	.	Phytophthora infestans	L1	175,16,1,176,3,24,11,4,2
L1-2_PR	.	Phytophthora ramorum	L1	1,8,3,11,4,12,2
L1-2_PTr	.	Populus trichocarpa	L1	43,91,2,44,24,11,93,1,4,16,75,3,76
L1-2_PTr	.	Populus trichocarpa	L1	43,91,2,44,24,11,93,1,4,16,75,3,76
L1-2_PVa	.	Pteropus vampyrus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_Sar	.	Sorex araneus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_SBi	.	Sorghum bicolor	L1	
L1-2_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-2_SP	.	Strongylocentrotus purpuratus	L1	1,43,3,44,4,2
L1-2_SSa	.	Salmo salar	L1	21,1,26,23,3,11,4,2
L1-2_SSc	.	Sus scrofa	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_STr	.	Ictidomys tridecemlineatus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_STu	.	Solanum tuberosum	L1	
L1-2_Tbel	.	Tupaia belangeri	L1	20,2,21,23,11,25,1,28,4,30,3
L1-2_TC	.	Theobroma cacao	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-2_TML	.	Trichechus manatus latirostris	L1	20,83,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_TS	.	Carlito syrichta	L1	
L1-2_Ttr	.	Tursiops truncatus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-2_VC	.	Volvox carteri	L1	16,1,26,3,24,4,11,2
L1-2_Vpa	.	Vicugna pacos	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_VVi	.	Vitis vinifera	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-2_XL	.	Xenopus laevis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-2_XT	.	Xenopus tropicalis	L1	1,3,11,4,2

L1-2_ZM	Zea mays	L1	43,2,8,44,11,12,1,14,4,75,3,76
L1-20_ALy	Arabidopsis lyrata	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-20_ATr	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,75,3,76,7
L1-20_CPa	Carica papaya	L1	43,44,11
L1-20_DR	Danio rerio	L1	2,21,24,11,1,26,28,4,30,16,3
L1-20_FV	Fragaria vesca	L1	43,17,44,10,79,7,80
L1-20_LSal	Lepeophtheirus salmonis	L1	1,8,3,11,4,12,2
L1-20_NN	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-20_SSa	Salmo salar	L1	
L1-20_XT	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-20a_Lch	Latimeria chalumnae	L1	
L1-20b_Lch	Latimeria chalumnae	L1	
L1-20c_Lch	Latimeria chalumnae	L1	
L1-21_ACar	Anolis carolinensis	L1	2,21,23,11,1,26,4,30,3
L1-21_ALy	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-21_ATr	Amborella trichopoda	L1	
L1-21_CPa	Carica papaya	L1	44,43
L1-21_DR	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-21_FV	Fragaria vesca	L1	43,17,10,44,79,6,7,80
L1-21_Lch	Latimeria chalumnae	L1	
L1-21_LSal	Lepeophtheirus salmonis	L1	16,1,3,24,11,4,2
L1-21_NN	Nelumbo nucifera	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-21_SSa	Salmo salar	L1	2,21,23,24,11,1,26,4,16,3

L1-21_XT	.	Xenopus tropicalis	L1 2,21,24,11,1,26,28,4,30,16,3
L1-21B_ACar	.	Anolis carolinensis	L1 21,1,26,23,3,11,4,30,2
L1-22_ACar	.	Anolis carolinensis	L1 2,21,23,24,11,1,26,4,30,16,3
L1-22_ALy	.	Arabidopsis lyrata	L1 16,1,17,43,3,44,10,79,24,11,4,2,80
L1-22_ATr	.	Amborella trichopoda	L1 1,3,4,2
L1-22_CPa	.	Carica papaya	L1 1,43,3,44,4,2
L1-22_DR	.	Danio rerio	L1 2,21,23,24,11,1,26,4,30,16,3
L1-22_FV	.	Fragaria vesca	L1
L1-22_Lch	.	Latimeria chalumnae	L1
L1-22_LSal	.	Lepeophtheirus salmonis	L1 1,3,11,4,2
L1-22_NN	.	Nelumbo nucifera	L1 43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95
L1-22_SSa	.	Salmo salar	L1 1,3,11,4,2
L1-22_XT	.	Xenopus tropicalis	L1 2,21,24,11,1,26,28,4,30,16,3
L1-23_ACar	.	Anolis carolinensis	L1 1,26,23,3,11,4,2
L1-23_ALy	.	Arabidopsis lyrata	L1 43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-23_ATr	.	Amborella trichopoda	L1 7,1,17,43,3,44,10,79,11,4,2,80
L1-23_CPa	.	Carica papaya	L1 44,43
L1-23_DR	.	Danio rerio	L1 2,21,218,23,24,11,1,26,219,4,30,16,3
L1-23_FV	.	Fragaria vesca	L1 75,8,76,78,81,77,12
L1-23_LSal	.	Lepeophtheirus salmonis	L1 1,3,11,4,2
L1-23_NN	.	Nelumbo nucifera	L1 43,79,6,2,10,44,24,11,80,1,4,94,16,17,75,3,76,95,7
L1-23_SSa	.	Salmo salar	L1 2,21,23,11,1,4,30,3
L1-23_XT	.	Xenopus tropicalis	L1

L1-23B_ACar .	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-24_ACar .	Anolis carolinensis	L1	1,26,3,11,4,2
L1-24_ALy .	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,17,75,3,76,7
L1-24_ATr .	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-24_CPa .	Carica papaya	L1	44,43
L1-24_DR .	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-24_FV .	Fragaria vesca	L1	43,17,10,44,79,7,80
L1-24_LSal .	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-24_NN .	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-24_SSa .	Salmo salar	L1	21,1,26,3,28,11,4,2
L1-24_XT .	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-24B_ACar .	Anolis carolinensis	L1	1,26,3,11,4,2
L1-24C_ACar .	Anolis carolinensis	L1	1,26,3,11,4,2
L1-25_ACar .	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-25_ALy .	Arabidopsis lyrata	L1	43,78,79,6,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-25_ATr .	Amborella trichopoda	L1	16,1,24,11,2
L1-25_CPa .	Carica papaya	L1	44,43
L1-25_DR .	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-25_FV .	Fragaria vesca	L1	17,10,79,6,7,80
L1-25_LSal .	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-25_NN .	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-25_SSa .	Salmo salar	L1	21,1,26,3,28,11,4,2
L1-25_XT .	Xenopus tropicalis	L1	21,1,23,3,11,4,30,2

L1-26_ACar	.	Anolis carolinensis	L1 2,21,23,11,1,26,4,15,16,3
L1-26_ALy	.	Arabidopsis lyrata	L1 43,78,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-26_ATr	.	Amborella trichopoda	L1 7,1,80,17,43,3,44,10,79,4,6,2
L1-26_CPa	.	Carica papaya	L1 44,43
L1-26_DR	.	Danio rerio	L1 2,21,23,11,1,26,4,30,3
L1-26_FV	.	Fragaria vesca	L1 11
L1-26_LSal	.	Lepeophtheirus salmonis	L1
L1-26_NN	.	Nelumbo nucifera	L1 43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,3,7
L1-26_SSa	.	Salmo salar	L1 1,3,11,4,2
L1-26_XT	.	Xenopus tropicalis	L1 2,21,23,24,11,1,26,4,30,16,3
L1-26B_ACar	.	Anolis carolinensis	L1 21,16,26,1,3,24,11,4,2
L1-27_ACar	.	Anolis carolinensis	L1 2,21,23,11,1,26,4,15,16,3
L1-27_ALy	.	Arabidopsis lyrata	L1 43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-27_ATr	.	Amborella trichopoda	L1 43,79,2,44,10,11,80,1,77,4,17,75,3,76,7
L1-27_CPa	.	Carica papaya	L1
L1-27_DR	.	Danio rerio	L1
L1-27_FV	.	Fragaria vesca	L1 43,17,44,10,79,6,7,80
L1-27_LSal	.	Lepeophtheirus salmonis	L1 1,3,4,2
L1-27_NN	.	Nelumbo nucifera	L1 7,1,80,17,3,10,79,4,6,2
L1-27_SSa	.	Salmo salar	L1 1,26,3,11,4,2
L1-27_XT	.	Xenopus tropicalis	L1 2,21,24,11,1,26,28,4,16,3
L1-27B_ACar	.	Anolis carolinensis	L1 21,1,26,23,3,11,4,2
L1-28_ACar	.	Anolis carolinensis	L1 1,26,3,11,4,2

L1-28_ALy	.	Arabidopsis lyrata	L1	43,79,2,44,10,11,80,1,77,4,75,17,3,76
L1-28_ATr	.	Amborella trichopoda	L1	
L1-28_CPa	.	Carica papaya	L1	44,43
L1-28_DR	.	Danio rerio	L1	16,26,24,11
L1-28_FV	.	Fragaria vesca	L1	
L1-28_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-28_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-28_SSa	.	Salmo salar	L1	1,2
L1-28_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-29_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-29_ALy	.	Arabidopsis lyrata	L1	1,17,43,3,44,10,79,4,11,2,80
L1-29_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,75,17,3,76,7
L1-29_CPa	.	Carica papaya	L1	44,43
L1-29_DR	.	Danio rerio	L1	16,26,24,11
L1-29_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-29_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-29_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3
L1-2A_Cpo	.	Cavia porcellus	L1	
L1-2a_EF	.	Eptesicus fuscus	L1	
L1-2a_MD	.	Monodelphis domestica	L1	21,23,30
L1-2B_ALy	.	Arabidopsis lyrata	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-2B_CPB	.	Chrysemys picta bellii	L1	
L1-2B_Cpo	.	Cavia porcellus	L1	

L1-2B_EC	.	Equus caballus	L1 20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2B_LA-5-конец	.	Loxodonta africana	L1
L1-2B_MPu	.	Mustela putorius furo	L1
L1-2B1_Cho	.	Choloepus hoffmanni	L1 20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2C1_Cho	.	Choloepus hoffmanni	L1 20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-3_AIp	.	Arachis ipaensis	L1 43,79,2,44,10,11,80,1,4,17,3,7
L1-3_ALai	.	Albugo laibachii	L1 1,3,11,4,2
L1-3_ALy	.	Arabidopsis lyrata	L1
L1-3_ATr	.	Amborella trichopoda	L1 43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_BDi	.	Brachypodium distachyon	L1 43,2,8,44,11,12,1,14,4,75,3,76
L1-3_BRa	.	Brassica rapa	L1 7,1,43,17,3,44,10,79,4,2,80
L1-3_BT	.	Bos taurus	L1 20,2,21,23,25,1,27,28,4,30,31,3
L1-3_BV	.	Beta vulgaris subsp. vulgaris	L1
L1-3_Cho	.	Choloepus hoffmanni	L1 16,1,26,3,24,4,11,2
L1-3_Cis	.	Ciona savignyi	L1 16,1,14,3,24,11,4,12,2
L1-3_CLan	.	Chinchilla lanigera	L1 20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-3_CPa	.	Carica papaya	L1 1,43,3,44,11,4,2
L1-3_CPB	.	Chrysemys picta bellii	L1 21,1,26,23,3,28,11,4,2
L1-3_Cpo	.	Cavia porcellus	L1 20,2,21,23,11,25,1,28,4,30,3
L1-3_CQ	.	Culex quinquefasciatus	L1 16,1,26,43,3,44,24,11,4,92,2
L1-3_CR	.	Chlamydomonas reinhardtii	L1
L1-3_DR	.	Danio rerio	Tx1 2,8,11,12,42,1,26,14,4,3
L1-3_EC	.	Equus caballus	L1 20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3

L1-3_ECa	.	Eucalyptus camaldulensis	L1	1,43,75,3,76,44,77,11,4,2
L1-3_EE	.	Erinaceus europaeus	L1	20,2,21,23,25,1,28,4,30,3
L1-3_EPa	.	Schrenkiella parvula	L1	
L1-3_FV	.	Fragaria vesca	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_GAr	.	Gossypium arboreum	L1	16,1,43,3,44,24,11,4,2
L1-3_GM	.	Glycine max	L1	16,1,75,43,3,76,44,24,11,4,2
L1-3_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-3_JC	.	Jatropha curcas	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-3_LA	.	Loxodonta africana	L1	1,22,3,4,29,2
L1-3_LCh	.	Latimeria chalumnae	L1	2,21,23,11,1,26,4,30,15,16,3
L1-3_LSal	.	Lepeophtheirus salmonis	L1	1,8,14,3,11,4,12,2
L1-3_Mad	.	Malus domestica	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-3_MD	.	Monodelphis domestica	L1	11
L1-3_ME	.	Notamacropus eugenii	L1	25,21,20,23,28,30
L1-3_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-3_NS	.	Nicotiana sylvestris	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-3_NV	.	Nematostella vectensis	L1	
L1-3_Opr	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-3_PAb	.	Picea abies	L1	43,6,2,10,44,24,11,80,1,14,77,230,4,16,17,75,3,76,7
L1-3_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-3_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-3_PTr	.	Populus trichocarpa	L1	16,1,3,24,11,4,2
L1-3_PTr	.	Populus trichocarpa	L1	16,1,3,24,11,4,2

L1-3_PVa	.	Pteropus vampyrus	L1	
L1-3_SBi	.	Sorghum bicolor	L1	1,43,44,11,4,2
L1-3_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-3_SSa	.	Salmo salar	L1	16,21,1,26,3,24,11,4,2
L1-3_SSc	.	Sus scrofa	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-3_STu	.	Solanum tuberosum	L1	43,2,44,11,1,26,4,94,75,3,76,95
L1-3_Tbel	.	Tupaia belangeri	L1	
L1-3_TC	.	Theobroma cacao	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_TML	.	Trichechus manatus latirostris	L1	16,1,26,3,11,4,2,15
L1-3_TS	.	Carlito syrichta	L1	20,2,21,22,23,11,25,1,26,27,28,4,29,30,15,31,16,3
L1-3_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-3_Vpa	.	Vicugna pacos	L1	16,1,26,3,24,11,4,2
L1-3_VVi	.	Vitis vinifera	L1	1,43,3,44,11,4,2
L1-3_XL	.	Xenopus laevis	L1	2,21,24,11,1,26,28,4,16,3
L1-3_XT	.	Xenopus tropicalis	L1	2,21,11,1,28,4,3
L1-3_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,3
L1-30_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-30_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-30_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-30_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,75,17,3,76,7
L1-30_CPa	.	Carica papaya	L1	44,43
L1-30_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-30_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2

L1-30_NN	.	Nelumbo nucifera subsp. nucifeL1		1,3,4,11,2
L1-30_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-31_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-31_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-31_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-31_ATr	.	Amborella trichopoda	L1	
L1-31_CPa	.	Carica papaya	L1	44,43
L1-31_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-31_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-31_NN	.	Nelumbo nucifera subsp. nucifeL1		7,1,43,17,3,44,10,79,4,6,2,80
L1-31_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-31B_LSal	.	Lepeophtheirus salmonis	L1	
L1-32_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-32_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-32_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-32_ATr	.	Amborella trichopoda	L1	1,75,43,3,76,44,11,4,2
L1-32_CPa	.	Carica papaya	L1	44,43
L1-32_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-32_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-32_NN	.	Nelumbo nucifera subsp. nucifeL1		7,1,43,17,3,44,10,79,4,6,2,80
L1-32_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-33_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1-33_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3

L1-33_ALy	.	Arabidopsis lyrata	L1 43,78,79,2,44,10,11,80,1,81,77,4,75,3,76
L1-33_ATr	.	Amborella trichopoda	L1 1,11,4,2
L1-33_CPa	.	Carica papaya	L1 44,43
L1-33_DR	.	Danio rerio	L1 2,21,11,1,26,28,4,3
L1-33_LSal	.	Lepeophtheirus salmonis	L1 1,43,3,44,11,4,2
L1-33_NN	.	Nelumbo nucifera subsp. nucifeL1	94,16,1,3,95,24,11,4,2
L1-33_XT	.	Xenopus tropicalis	L1 2,21,24,11,1,26,28,4,30,16,3
L1-34_AAe	.	Aedes aegypti	L1 93,16,1,26,3,24,11,4,2
L1-34_ACar	.	Anolis carolinensis	L1 16,1,26,3,24,11,4,2
L1-34_ALy	.	Arabidopsis lyrata	L1 43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-34_ATr	.	Amborella trichopoda	L1
L1-34_DR	.	Danio rerio	L1 2,21,23,11,1,26,4,30,3
L1-34_LSal	.	Lepeophtheirus salmonis	L1 1,26,3,11,4,2
L1-34_NN	.	Nelumbo nucifera subsp. nucifeL1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-34_XT	.	Xenopus tropicalis	L1 1,26,3,11,4,2
L1-35_AAe	.	Aedes aegypti	L1 16,1,26,43,3,44,24,11,4,2
L1-35_ACar	.	Anolis carolinensis	L1 1,26,23,3,11,4,2
L1-35_ALy	.	Arabidopsis lyrata	L1 43,78,2,44,11,1,81,77,4,75,3,76
L1-35_ATr	.	Amborella trichopoda	L1 43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-35_DR	.	Danio rerio	L1 21,1,3,28,11,4,2
L1-35_LSal	.	Lepeophtheirus salmonis	L1 1,26,3,4,11,2
L1-35_NN	.	Nelumbo nucifera subsp. nucifeL1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-35_XT	.	Xenopus tropicalis	L1 2,21,24,11,1,26,28,4,30,16,3

L1-35B_ACar	Anolis carolinensis	L1	
L1-36_AAe	Aedes aegypti	L1	1,26,3,11,4,2
L1-36_ACar	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-36_ALy	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,77,4,16,17,75,3,76,7
L1-36_ATr	Amborella trichopoda	L1	1,43,3,44,4,2
L1-36_DR	Danio rerio	L1	2,21,23,11,1,28,4,3
L1-36_LSal	Lepeophtheirus salmonis	L1	
L1-36_NN	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-36_XT	Xenopus tropicalis	L1	2,21,23,11,1,26,28,4,15,16,3
L1-37_AAe	Aedes aegypti	L1	93,16,1,3,96,24,11,4,98,2
L1-37_ACar	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-37_ALy	Arabidopsis lyrata	L1	43,78,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-37_ATr	Amborella trichopoda	L1	
L1-37_DR	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-37_LSal	Lepeophtheirus salmonis	L1	
L1-37_NN	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-37_XT	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-38_AAe	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-38_ACar	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-38_ALy	Arabidopsis lyrata	L1	43,78,2,44,10,24,11,80,1,81,77,4,16,75,17,3,76
L1-38_ATr	Amborella trichopoda	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-38_DR	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-38_LSal	Lepeophtheirus salmonis	L1	

L1-38_NN	.	Nelumbo nucifera subsp. nucifeL1		1,43,3,44,4,2
L1-38_XT	.	Xenopus tropicalis	L1	21,1,23,3,11,4,2
L1-39_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,92,2
L1-39_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-39_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-39_ATr	.	Amborella trichopoda	L1	
L1-39_DR	.	Danio rerio	L1	1,43,3,44,4,2
L1-39_LSal	.	Lepeophtheirus salmonis	L1	
L1-39_NN	.	Nelumbo nucifera subsp. nucifeL1		43,79,6,2,10,44,11,80,1,4,17,3,7
L1-39_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-3A_Cho	.	Choloepus hoffmanni	L1	
L1-3A_Cpo	.	Cavia porcellus	L1	
L1-3a_EF	.	Eptesicus fuscus	L1	
L1-3A_ME	.	Notamacropus eugenii	L1	25,21,20,23,28,30
L1-3B_ALy	.	Arabidopsis lyrata	L1	75,8,76,77,12
L1-3B_SS_3-конец	.	Sus scrofa	L1	
L1-3C_ALy	.	Arabidopsis lyrata	L1	43,78,79,6,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-4_AIp	.	Arachis ipaensis	L1	43,79,2,7,101,10,44,11,102,80,1,77,4,75,17,3,76
L1-4_ALy	.	Arabidopsis lyrata	L1	
L1-4_ATr	.	Amborella trichopoda	L1	94,16,1,43,3,44,95,24,11,4,2
L1-4_BDi	.	Brachypodium distachyon	L1	16,1,3,24,11,4,2
L1-4_BV	.	Beta vulgaris subsp. vulgaris	L1	43,78,79,2,7,10,44,11,80,1,81,77,4,75,17,3,76
L1-4_Cho	.	Choloepus hoffmanni	L1	

L1-4_Cis	.	Ciona savignyi	L1	
L1-4_CPa	.	Carica papaya	L1	16,1,43,3,44,11,4,2,15
L1-4_CPB	.	Chrysemys picta bellii	L1	2,21,23,11,1,26,28,4,3
L1-4_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-4_CQ	.	Culex quinquefasciatus	L1	93,16,1,26,3,24,11,4,2
L1-4_CR	.	Chlamydomonas reinhardtii	L1	
L1-4_DR	AL807749	Danio rerio	Tx1	2,8,24,11,12,1,26,4,16,3
L1-4_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-4_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76
L1-4_EPa	.	Schrenkiella parvula	L1	7,1,43,17,3,44,10,79,4,11,2,80
L1-4_FV	.	Fragaria vesca	L1	43,78,79,6,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7
L1-4_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-4_GM	.	Glycine max	L1	
L1-4_HM	.	Hydra vulgaris	L1	16,1,26,43,3,44,24,11,4,2
L1-4_JC	.	Jatropha curcas	L1	43,78,79,2,10,44,24,11,80,1,81,4,16,17,3,7
L1-4_LA	.	Loxodonta africana	L1	
L1-4_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2
L1-4_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-4_Mad	.	Malus domestica	L1	43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76
L1-4_MD	.	Monodelphis domestica	L1	
L1-4_ME	.	Notamacropus eugenii	L1	
L1-4_NN	.	Nelumbo nucifera	L1	43,79,6,2,7,10,44,24,11,80,1,4,94,16,75,17,3,76,95
L1-4_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76

L1-4_NV	.	Nematostella vectensis	L1	
L1-4_PAb	.	Picea abies	L1	43,79,6,2,10,44,24,11,80,1,77,4,16,17,75,3,76,7
L1-4_PI	.	Phytophthora infestans	L1	1,8,3,11,4,12,2
L1-4_PR	.	Phytophthora ramorum	L1	
L1-4_PTr	.	Populus trichocarpa	L1	16,1,43,3,44,24,11,4,2
L1-4_PTr	.	Populus trichocarpa	L1	16,1,43,3,44,24,11,4,2
L1-4_PVa	.	Pteropus vampyrus	L1	
L1-4_SBi	.	Sorghum bicolor	L1	
L1-4_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-4_SSa	.	Salmo salar	L1	2,21,23,11,1,26,4,3
L1-4_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-4_TML	.	Trichechus manatus latirostris	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-4_TS	.	Carlito syrichta	L1	
L1-4_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-4_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-4_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,3
L1-40_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-40_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-40_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-40_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-40_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-40_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,11,4,2
L1-40_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3

L1-40B_ACar	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-41_AAe	Aedes aegypti	L1	16,1,3,24,11,4,2
L1-41_ACar	Anolis carolinensis	L1	21,1,26,3,11,4,2
L1-41_ALy	Arabidopsis lyrata	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76
L1-41_LSal	Lepeophtheirus salmonis	L1	1,3,2
L1-41_NN	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,80,1,4,17,3,7
L1-41_XT	Xenopus tropicalis	L1	1,26,3,11,4,2
L1-42_AAe	Aedes aegypti	L1	1,3,11,4,2
L1-42_ACar	Anolis carolinensis	L1	2,21,24,11,1,26,28,4,16,3
L1-42_ALy	Arabidopsis lyrata	L1	43,78,2,44,11,12,1,14,81,77,4,75,3,76
L1-42_DR	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-42_LSal	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-42_NN	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,11,4,2
L1-42_XT	Xenopus tropicalis	L1	21,1,26,23,3,11,4,2
L1-42A_XT	Xenopus tropicalis	L1	21,1,23,3,11,4,2
L1-43_AAe	Aedes aegypti	L1	93,16,1,3,24,11,4,92,2
L1-43_ACar	Anolis carolinensis	L1	1,26,3,11,4,2
L1-43_DR	Danio rerio	L1	21,1,3,28,11,4,2
L1-43_LSal	Lepeophtheirus salmonis	L1	1,14,3,11,4,2
L1-43_NN	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7
L1-43_XT	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-44_AAe	Aedes aegypti	L1	16,1,3,24,11,4,2
L1-44_ACar	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3

L1-44_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-44_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-44_NN	.	Nelumbo nucifera subsp. nucifeL1		1,43,75,3,76,44,11,4,2
L1-44_XT	.	Xenopus tropicalis	L1	21,1,26,23,3,11,4,2
L1-45_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1-45_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-45_DR	.	Danio rerio	L1	26,11
L1-45_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-45_NN	.	Nelumbo nucifera subsp. nucifeL1		43,78,79,2,7,8,10,44,11,12,80,1,81,77,4,17,3
L1-45_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-46_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-46_ACar	.	Anolis carolinensis	L1	21,1,26,3,11,4,2
L1-46_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-46_LSal	.	Lepeophtheirus salmonis	L1	1,11,2
L1-46_NN	.	Nelumbo nucifera subsp. nucifeL1		1,43,3,44,11,4,2
L1-46_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-47_AAe	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1-47_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-47_DR	.	Danio rerio	L1	21,26,23,28,11,30
L1-47_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-47_NN	.	Nelumbo nucifera subsp. nucifeL1		16,1,43,3,44,24,11,4,2
L1-47_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-47B_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3

L1-48_AAe	.	Aedes aegypti L1	1,43,3,44,11,4,2
L1-48_ACar	.	Anolis carolinensis L1	21,1,26,3,11,4,2
L1-48_DR	.	Danio rerio L1	2,21,23,228,11,1,26,4,3,229
L1-48_LSal	.	Lepeophtheirus salmonis L1	11
L1-48_XT	.	Xenopus tropicalis L1	20,2,21,23,11,25,1,26,4,3
L1-49_ACar	.	Anolis carolinensis L1	16,1,26,23,3,24,11,4,2
L1-49_DR	.	Danio rerio L1	2,21,23,24,11,1,26,4,30,16,3
L1-49_LSal	.	Lepeophtheirus salmonis L1	
L1-49_NN	.	Nelumbo nucifera subsp. nucifeL1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-49_XT	.	Xenopus tropicalis L1	
L1-4A_Cpo	.	Cavia porcellus L1	
L1-4a_EF	.	Eptesicus fuscus L1	
L1-4A_MD	.	Monodelphis domestica L1	
L1-4A_ME	.	Notamacropus eugenii L1	
L1-4B_DR	.	Danio rerio Tx1	43,2,8,44,11,12,1,26,4,15,16,3
L1-4B_EC	.	Equus caballus L1	20,2,21,23,25,1,27,28,4,30,31,3
L1-4B_LA-5-конец	.	Loxodonta africana L1	
L1-4B_NS-3-конец	.	Nicotiana sylvestris L1	
L1-4C_NS-5-конец	.	Nicotiana sylvestris L1	
L1-5_A5-конец_EC	.	Equus caballus L1	
L1-5_AIp	.	Arachis ipaensis L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-5_ALy	.	Arabidopsis lyrata L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-5_ATr	.	Amborella trichopoda L1	43,79,6,2,44,10,11,80,1,77,4,17,75,3,76,7

L1-5_B5-конец_EC	.	Equus caballus	L1	
L1-5_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-5_BV	.	Beta vulgaris	L1	
L1-5_Cis	.	Ciona savignyi	L1	
L1-5_CPa	.	Carica papaya	L1	1,3,4,2
L1-5_CPB	.	Chrysemys picta bellii	L1	
L1-5_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-5_CR	.	Chlamydomonas reinhardtii	L1	
L1-5_DR	.	Danio rerio	Tx1	2,8,11,12,42,1,26,14,4,3
L1-5_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-5_ECa	.	Eucalyptus camaldulensis	L1	75,76,78,81,77
L1-5_EPa	.	Schrenkiella parvula	L1	44,43
L1-5_FV	.	Fragaria vesca	L1	1,43,17,44,10,79,2,6,7,80
L1-5_GAr	.	Gossypium arboreum	L1	16,1,43,3,44,10,24,11,4,2,80
L1-5_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-5_HM	.	Hydra vulgaris	L1	16,1,26,43,3,44,24,11,4,2
L1-5_JC	.	Jatropha curcas	L1	
L1-5_LA	.	Loxodonta africana	L1	1,3,4,2
L1-5_LCh	.	Latimeria chalumnae	L1	2,21,23,24,11,1,26,4,30,16,3
L1-5_LSal	.	Lepeophtheirus salmonis	L1	1,8,14,3,11,4,12,2
L1-5_Mad	.	Malus domestica	L1	43,78,79,2,44,10,11,80,1,81,77,4,17,75,76,7
L1-5_MD	.	Monodelphis domestica	L1	
L1-5_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7

L1-5_PI	.	Phytophthora infestans	L1	6,2,10,11,80,1,26,4,17,3,7
L1-5_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-5_PTr	.	Populus trichocarpa	L1	43,79,6,2,8,10,44,11,12,80,1,77,4,17,75,3,76,7
L1-5_SBi	.	Sorghum bicolor	L1	1,43,44,11,4,2
L1-5_SSa	.	Salmo salar	L1	16,1,26,3,24,11,4,2
L1-5_STu	.	Solanum tuberosum	L1	
L1-5_TS	.	Carlito syrichta	L1	
L1-5_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-5_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,4,11,2
L1-5_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,94,16,3,95
L1-50_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-50_DR	.	Danio rerio	L1	21,1,23,3,11,4,2
L1-50_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-50_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95
L1-50_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-50A_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-51_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-51_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-51_LSal	.	Lepeophtheirus salmonis	L1	11
L1-51_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-51_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-52_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-52_DR	.	Danio rerio	L1	21,1,3,28,11,4,2

L1-52_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,15,2
L1-52_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
L1-52_XT	.	Xenopus tropicalis	L1	20,2,21,23,24,11,25,1,26,4,16,3
L1-53_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-53_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-53_LSal	.	Lepeophtheirus salmonis	L1	11
L1-53_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,7,44,10,24,11,80,1,77,4,16,75,17,3,76
L1-53_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
L1-54_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-54_DR	.	Danio rerio	L1	1,11,2
L1-54_LSal	.	Lepeophtheirus salmonis	L1	
L1-54_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,4,2
L1-54_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
L1-55_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-55_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,30,16,3
L1-55_LSal	.	Lepeophtheirus salmonis	L1	
L1-55_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7
L1-55_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
L1-56_ACar	.	Anolis carolinensis	L1	2,21,24,11,1,26,28,4,16,3
L1-56_DR	.	Danio rerio	L1	21,1,23,3,11,4,2
L1-56_LSal	.	Lepeophtheirus salmonis	L1	
L1-56_NN	.	Nelumbo nucifera subsp. nucife	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-56_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,1,26,14,4,3

L1-57_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-57_DR	.	Danio rerio	L1	26,11
L1-57_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-57_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,30,16,3
L1-58_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-58_DR	.	Danio rerio	L1	
L1-58_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-58_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-59_ACar	.	Anolis carolinensis	L1	1,26,23,3,4,11,2
L1-59_DR	.	Danio rerio	L1	21,16,1,26,3,4,15,11,2
L1-59_LSal	.	Lepeophtheirus salmonis	L1	
L1-59_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-5A_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-5A2_EC	.	Equus caballus	L1	21,16,26,23,24,11,30
L1-5B_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-6_AIp	.	Arachis ipaensis	L1	43,79,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-6_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-6_ATr	.	Amborella trichopoda	L1	43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7
L1-6_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-6_BV	.	Beta vulgaris	L1	43,78,79,2,7,8,10,44,11,12,80,1,81,4,75,17,3,76
L1-6_Cis	.	Ciona savignyi	L1	
L1-6_CPa	.	Carica papaya	L1	1,3,11,4,2
L1-6_CPB	.	Chrysemys picta bellii	L1	2,21,23,11,1,26,4,30,3

L1-6_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-6_CR	.	Chlamydomonas reinhardtii	L1	16,1,26,24,11,4,2
L1-6_DR	AL627249	Danio rerio	L1	2,21,23,24,11,1,26,28,4,16,3
L1-6_ECa	.	Eucalyptus camaldulensis	L1	16,1,43,3,44,24,11,4,2
L1-6_FV	.	Fragaria vesca	L1	7,1,80,17,43,3,44,10,79,6,2
L1-6_GAr	.	Gossypium arboreum	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-6_GM	.	Glycine max	L1	1,75,43,3,76,44,11,4,2
L1-6_JC	.	Jatropha curcas	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-6_LA	.	Loxodonta africana	L1	
L1-6_Laf	.	Loxodonta africana	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-6_LCh	.	Latimeria chalumnae	L1	2,21,23,24,11,1,26,4,16,3
L1-6_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-6_Mad	.	Malus domestica	L1	7,1,17,43,3,44,10,79,4,11,2,80
L1-6_MD	.	Monodelphis domestica	L1	
L1-6_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-6_PI	.	Phytophthora infestans	L1	6,2,10,11,80,1,230,4,17,3,7
L1-6_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-6_SBi	.	Sorghum bicolor	L1	1,43,75,3,76,44,11,4,2
L1-6_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-6_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-6_TS	.	Carlito syrichta	L1	
L1-6_VC	.	Volvox carteri	L1	
L1-6_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3

L1-6_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,4,16,3
L1-60_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,11,4,2,15
L1-60_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-60_LSal	.	Lepeophtheirus salmonis	L1	1,8,3,4,11,12,2
L1-61_ACar	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1-61_DR	.	Danio rerio	L1	21,1,26,11,2
L1-61_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-62_DR	.	Danio rerio	L1	
L1-62_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-62_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-63_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-63_DR	.	Danio rerio	L1	1,26,28,11,4,2
L1-63_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-63_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-64_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-64_DR	.	Danio rerio	L1	1,26,11,2
L1-64_LSal	.	Lepeophtheirus salmonis	L1	
L1-64_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-65_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-65_DR	.	Danio rerio	L1	
L1-65_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-65_XT	.	Xenopus tropicalis	L1	
L1-66_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2

L1-66_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-66_LSal	.	Lepeophtheirus salmonis	L1	
L1-66_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-67_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-67_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-67_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-67_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-68_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-68_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-68_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-68_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,118,16,3
L1-69_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-69_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-69_LSal	.	Lepeophtheirus salmonis	L1	1,2
L1-6A_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-6B_JC	.	Jatropha curcas	L1	
L1-7_AIp	.	Arachis ipaensis	L1	1,75,43,3,76,44,11,4,2
L1-7_ALy	.	Arabidopsis lyrata	L1	7,1,43,17,3,44,10,79,4,2,80
L1-7_ATr	.	Amborella trichopoda	L1	1,3,11,4,2
L1-7_BDi	.	Brachypodium distachyon	L1	
L1-7_BV	.	Beta vulgaris	L1	
L1-7_Cis	.	Ciona savignyi	L1	1,14,3,11,4,12,2
L1-7_CPa	.	Carica papaya	L1	1,43,3,44,4,2

L1-7_CPB	.	Chrysemys picta bellii	L1	
L1-7_CQ	.	Culex quinquefasciatus	L1	16,1,3,24,11,4,2
L1-7_CR	.	Chlamydomonas reinhardtii	L1	
L1-7_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-7_ECa	.	Eucalyptus camaldulensis	L1	1,43,3,44,4,2
L1-7_FV	.	Fragaria vesca	L1	43,17,10,44,79,6,7,80
L1-7_GAr	.	Gossypium arboreum	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-7_GM	.	Glycine max	L1	
L1-7_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-7_JC	.	Jatropha curcas	L1	7,1,17,43,3,44,10,79,4,11,2,80
L1-7_LCh	.	Latimeria chalumnae	L1	16,1,3,24,4,11,2
L1-7_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-7_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-7_MD	.	Monodelphis domestica	L1	
L1-7_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-7_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-7_PR	.	Phytophthora ramorum	L1	8,11,12
L1-7_SBi	.	Sorghum bicolor	L1	1,3,11,4,2
L1-7_SSa	.	Salmo salar	L1	2,21,23,24,11,1,26,4,30,16,3
L1-7_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-7_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-7_ZM	.	Zea mays	L1	43,78,79,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7
L1-70_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2

L1-70_DR	.	Danio rerio	L1	44,43
L1-70_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-71_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,3
L1-71_DR	.	Danio rerio	L1	
L1-72_ACar	.	Anolis carolinensis	L1	1,26,3,4,11,2
L1-72_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-73_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-73_DR	.	Danio rerio	L1	
L1-74_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-74_DR	.	Danio rerio	L1	43,2,21,44,24,11,1,26,4,16,3
L1-75_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-75_DR	.	Danio rerio	L1	1,26,3,11,4,2
L1-76_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-77_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-78_ACar	.	Anolis carolinensis	L1	
L1-78_DR	.	Danio rerio	L1	2,21,23,11,1,26,28,4,3
L1-79_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,11,4,2,15
L1-79_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-79B_ACar	.	Anolis carolinensis	L1	
L1-7A_XT	.	Xenopus tropicalis	L1	
L1-8_AIp	.	Arachis ipaensis	L1	16,1,43,3,44,24,11,4,2
L1-8_ALy	.	Arabidopsis lyrata	L1	1,43,3,44,11,4,2
L1-8_BDi	.	Brachypodium distachyon	L1	

L1-8_Cis	.	Ciona savignyi	L1	1,3,11,4,2
L1-8_CPa	.	Carica papaya	L1	1,43,3,44,11,4,2
L1-8_CQ	.	Culex quinquefasciatus	L1	16,1,43,3,44,24,11,4,2
L1-8_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-8_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76
L1-8_FV	.	Fragaria vesca	L1	43,17,44,10,79,6,7,80
L1-8_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-8_GM	.	Glycine max	L1	1,43,75,3,44,76,11,4,2
L1-8_HMa	.	Hydra vulgaris	L1	1,43,3,44,11,4,2
L1-8_JC	.	Jatropha curcas	L1	43,17,10,44,79,7,80
L1-8_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-8_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-8_MD	.	Monodelphis domestica	L1	
L1-8_NN	.	Nelumbo nucifera	L1	43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95
L1-8_PI	.	Phytophthora infestans	L1	16,1,3,24,11,4,2
L1-8_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,1,14,4,15,16,75,3,76
L1-8_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-8_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-8_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-80_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-81_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-81_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-82_ACar	.	Anolis carolinensis	L1	1,23,3,11,4,2

L1-82_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-83_ACar	.	Anolis carolinensis	L1	116,2,21,23,11,1,26,4,3
L1-83_DR	.	Danio rerio	L1	20,2,21,11,25,1,26,28,4,3
L1-84_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-84_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-85_ACar	.	Anolis carolinensis	L1	1,26,3,28,11,4,2
L1-85_DR	.	Danio rerio	L1	1,26,43,44,11,4,2
L1-86_DR	.	Danio rerio	L1	21,1,23,3,11,4,30,2
L1-87_ACar	.	Anolis carolinensis	L1	
L1-87_DR	.	Danio rerio	L1	
L1-88_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-88_DR	.	Danio rerio	L1	21,1,3,28,4,2
L1-89_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-89_DR	.	Danio rerio	L1	2,1
L1-9_AIp	.	Arachis ipaensis	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-9_ALy	.	Arabidopsis lyrata	L1	43,2,44,24,11,1,77,4,16,75,3,76
L1-9_ATr	.	Amborella trichopoda	L1	
L1-9_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-9_CQ	.	Culex quinquefasciatus	L1	92
L1-9_DR	.	Danio rerio	L1	
L1-9_ECa	.	Eucalyptus camaldulensis	L1	1,43,3,44,11,4,2
L1-9_FV	.	Fragaria vesca	L1	1,43,17,44,10,79,2,7,80
L1-9_GAr	.	Gossypium arboreum	L1	43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7

L1-9_GM	.	Glycine max	L1	
L1-9_JC	.	Jatropha curcas	L1	75,8,78,76,81,77,12
L1-9_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,4,11,2
L1-9_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-9_Mad	.	Malus domestica	L1	43,78,79,2,8,44,10,11,12,80,1,81,77,4,17,75,76,7
L1-9_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-9_NS	.	Nicotiana sylvestris	L1	1,75,43,3,76,44,10,79,11,4,2,80
L1-9_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-9_STu	.	Solanum tuberosum	L1	94,16,1,43,3,44,95,24,11,4,2
L1-9_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-90_ACar	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-90_DR	.	Danio rerio	L1	
L1-91_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-91_DR	.	Danio rerio	L1	
L1-92_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-92_Cis	.	Ciona savignyi	L1	1,26,8,3,11,4,12,2
L1-92_DR	.	Danio rerio	L1	21,1,26,23,11,4,30,2
L1-93_ACar	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1-93_Cis	.	Ciona savignyi	L1	16,1,26,3,24,4,11,2
L1-94_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-94_Cis	.	Ciona savignyi	L1	1,26,43,3,44,11,4,2
L1-94_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-95_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3

L1-95_Cis	.	Ciona savignyi	L1	16,1,8,26,3,11,4,12,2,15
L1-95_DR	.	Danio rerio	L1	20,2,21,23,11,25,1,26,28,4,30,3
L1-96_ACar	.	Anolis carolinensis	L1	23
L1-96_Cis	.	Ciona savignyi	L1	187,16,1,26,3,24,11,4,188,2
L1-96_DR	.	Danio rerio	L1	
L1-97_Cis	.	Ciona savignyi	L1	43,2,8,44,11,185,12,1,26,4,15,16,3
L1-97_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,15,16,3
L1-98_Cis	.	Ciona savignyi	L1	43,1,26,3,44,11,4,2
L1-98_DR	.	Danio rerio	L1	
L1-99_Cis	.	Ciona savignyi	L1	
L1-99_DR	.	Danio rerio	L1	
L1-99B_DR	.	Danio rerio	L1	
L1-99ext_Cis	.	Ciona savignyi	L1	
L1-BT	DQ000238	Bos taurus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-N1_CQ	.	Culex quinquefasciatus	L1	
L1-N1_FV	.	Fragaria vesca	L1	75,76,77
L1-N2_CQ	.	Culex quinquefasciatus	L1	
L1-N2_FV	.	Fragaria vesca	L1	75,76,78,81,77
L1-N3_CQ	.	Culex quinquefasciatus	L1	
L1-N3_FV	.	Fragaria vesca	L1	75,76,77
L1-N4_CQ	.	Culex quinquefasciatus	L1	
L1-N4_FV	.	Fragaria vesca	L1	75,76,77
L1-N5_CQ	.	Culex quinquefasciatus	L1	

L1-N6_CQ	.	Culex quinquefasciatus	L1	
L1-Y_CF	.	Canis lupus familiaris	L1	
L1A_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1A_OC	.	Oryctolagus cuniculus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1A-1_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1A-2_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1A2_OC	.	Oryctolagus cuniculus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1B_BM	.	Bombyx mori	Jockey	
L1B_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1B_MM	.	Mus musculus	L1	
L1B_OC	M94361	Oryctolagus cuniculus	L1	
L1B_RN	.	Rattus norvegicus	L1	20,2,21,23,22,24,11,25,1,26,27,28,4,29,30,31,16,3
L1B_SS	.	Sus scrofa	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1B-1_Cho	.	Choloepus hoffmanni	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1BM	AB002279	Bombyx mori	Jockey	16,1,11,2,15
L1C_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1C_OC	M18818, X07786, X077	Oryctolagus cuniculus	L1	
L1C_SS_5-конец	.	Sus scrofa	L1	
L1HS	.	Homo sapiens	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1Lx_I	.	Mus musculus	L1	
L1Lx_II	.	Mus musculus	L1	
L1Lx_III	.	Mus musculus	L1	
L1Lx_IV	.	Mus musculus	L1	

L1M1_5	.	Eutheria	L1	
L1M1B_5	.	Homo sapiens		L1
L1M2_5	.	Homo sapiens		L1
L1M2_5	.	Homo sapiens		L1
L1M2A_5	.	Homo sapiens		L1
L1M2A_5	.	Homo sapiens		L1
L1M2A1_5	.	Homo sapiens		L1
L1M2B_5	.	Homo sapiens		L1
L1M2B_5	.	Homo sapiens		L1
L1M2C_5	.	Homo sapiens		L1
L1M2C_5	.	Homo sapiens		L1
L1M3A_5	.	Eutheria	L1	
L1M3A_5	.	Eutheria	L1	
L1M3B_5	.	Homo sapiens		L1
L1M3C_5	.	Homo sapiens		L1
L1M3C_5	.	Homo sapiens		L1
L1M3D_5	.	Eutheria	L1	
L1M3D_5	.	Eutheria	L1	
L1M3DE_5	.	Eutheria	L1	
L1M3DE_5	.	Eutheria	L1	
L1M4B	.	Homo sapiens		L1
L1M6_5-конец	.	Mammalia		L1
L1M6_5-конец	.	Mammalia		L1

L1M6_5-конец	.	Mammalia	L1
L1M6B_5-конец	.	Homo sapiens	L1
L1M7_5-конец	.	Mammalia	L1
L1M7_5-конец	.	Mammalia	L1
L1M7_5-конец	.	Mammalia	L1
L1MA1	.	Homo sapiens	L1
L1MA10	.	Eutheria	L1
L1MA10	.	Eutheria	L1
L1MA10	.	Eutheria	L1
L1MA2	.	Homo sapiens	L1
L1MA2	.	Homo sapiens	L1
L1MA2	.	Homo sapiens	L1
L1MA3	.	Homo sapiens	L1
L1MA4	.	Homo sapiens	L1
L1MA4A	.	Homo sapiens	L1
L1MA5	.	Homo sapiens	L1
L1MA5	.	Homo sapiens	L1
L1MA5	.	Homo sapiens	L1
L1MA5A	.	Homo sapiens	L1
L1MA6	.	Homo sapiens	L1
L1MA7	.	Homo sapiens	L1
L1MA8	.	Homo sapiens	L1
L1MA9	.	Homo sapiens	L1

L1MA9	.	Homo sapiens	L1	
L1MA9	.	Homo sapiens	L1	
L1MA9_5	.	Eutheria	L1	
L1MAB_5-конец_EC	.	Equus caballus	L1	
L1MAB_EC	.	Equus caballus	L1	
L1MAB_ML	.	Myotis lucifugus	L1	
L1MAB1_5_EC	.	Equus caballus	L1	
L1MAB2_ML	.	Myotis lucifugus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1MB1	.	Homo sapiens	L1	
L1MB2	.	Homo sapiens	L1	
L1MB3	.	Homo sapiens	L1	
L1MB3	.	Homo sapiens	L1	
L1MB3	.	Homo sapiens	L1	
L1MB3_5	.	Homo sapiens	L1	
L1MB3_EC	.	Equus caballus	L1	1,22,3,4,11,29,2
L1MB4	.	Homo sapiens	L1	
L1MB4_5	.	Eutheria	L1	
L1MB4_Laf	.	Loxodonta africana	L1	
L1MB5	.	Homo sapiens	L1	
L1MB6_5	.	Eutheria	L1	
L1MB6_5	.	Eutheria	L1	
L1MB6_5	.	Eutheria	L1	
L1MB7	.	Eutheria	L1	

L1MB7	.	Eutheria	L1		
L1MB7	.	Eutheria	L1		
L1MB8	.	Homo sapiens		L1	
L1MC1	.	Homo sapiens		L1	
L1MC1_EC	.	Equus caballus	L1		21,1,23,3,11,4,30,2
L1MC2	.	Homo sapiens		L1	
L1MC2	.	Homo sapiens		L1	
L1MC2	.	Homo sapiens		L1	
L1MC3	.	Eutheria	L1		
L1MC3	.	Eutheria	L1		
L1MC3	.	Eutheria	L1		
L1MC4	.	Homo sapiens		L1	
L1MC4	.	Homo sapiens		L1	
L1MC4	.	Homo sapiens		L1	
L1MC4_5-конец	.	Eutheria		L1	
L1MC4_5-конец	.	Eutheria		L1	
L1MC4_5-конец	.	Eutheria		L1	
L1MC4B	.	Eutheria	L1		
L1MC4B	.	Eutheria	L1		
L1MC4B	.	Eutheria	L1		
L1MC5	.	Eutheria	L1		
L1MC5	.	Eutheria	L1		
L1MCA_5	.	Homo sapiens		L1	

L1MCA_5	.	Homo sapiens		L1	
L1MCB_5	.	Homo sapiens		L1	
L1MCB_5	.	Homo sapiens		L1	
L1MCC_5	.	Homo sapiens		L1	
L1MCC_5	.	Homo sapiens		L1	
L1Md_F_5-конец	.	Mus			L1
L1Md_Gf_5-конец	.	Mus			L1
L1MD1	.	Homo sapiens		L1	
L1MD1	.	Homo sapiens		L1	
L1MD1_5	.	Eutheria	L1		
L1MD2	.	Homo sapiens		L1	
L1MD2	.	Homo sapiens		L1	
L1MD2	.	Homo sapiens		L1	
L1MD3	.	Homo sapiens		L1	
L1MD3	.	Homo sapiens		L1	
L1MDA_5	.	Eutheria	L1		
L1MDA_5	.	Eutheria	L1		
L1MdA_I	.	Mus musculus		L1	
L1MdA_II	.	Mus musculus		L1	
L1MdA_III	.	Mus musculus		L1	
L1MdA_IV	.	Mus musculus		L1	
L1MdA_V	.	Mus musculus		L1	
L1MdA_VI	.	Mus musculus		L1	

L1MdA_VII	.	Mus musculus	L1
L1MDB_5	.	Homo sapiens	L1
L1MDB_5	.	Homo sapiens	L1
L1MdF_I	.	Mus musculus	L1
L1MdF_II	.	Mus musculus	L1
L1MdF_III	.	Mus musculus	L1
L1MdF_IV	.	Mus musculus	L1
L1MdF_V	.	Mus musculus	L1
L1MdFanc_I	.	Mus musculus	L1
L1MdFanc_II	.	Mus musculus	L1
L1MdGf_I	.	Mus musculus	L1
L1MdGf_II	.	Mus musculus	L1
L1MdMus_I	.	Mus musculus	L1
L1MdMus_II	.	Mus musculus	L1
L1MdN_I	.	Mus musculus	L1
L1MdTf_I	.	Mus musculus	L1
L1MdTf_II	.	Mus musculus	L1
L1MdTf_III	.	Mus musculus	L1
L1MdV_I	.	Mus musculus	L1
L1MdV_II	.	Mus musculus	L1
L1MdV_III	.	Mus musculus	L1
L1ME_ORF2	.	Homo sapiens	L1
L1ME_ORF2	.	Homo sapiens	L1

L1ME_ORF2	.	Homo sapiens	L1
L1ME1	.	Homo sapiens	L1
L1ME2	.	Homo sapiens	L1
L1ME2	.	Homo sapiens	L1
L1ME3	.	Homo sapiens	L1
L1ME3A	.	Homo sapiens	L1
L1ME3A	.	Homo sapiens	L1
L1ME3C_3-конец	.	Eutheria	L1
L1ME3C_3-конец	.	Eutheria	L1
L1ME3C_3-конец	.	Eutheria	L1
L1ME3D_3-конец	.	Eutheria	L1
L1ME3D_3-конец	.	Eutheria	L1
L1ME3D_3-конец	.	Eutheria	L1
L1ME3E_3-конец	.	Eutheria	L1
L1ME3E_3-конец	.	Eutheria	L1
L1ME3E_3-конец	.	Eutheria	L1
L1ME3F_3-конец	.	Eutheria	L1
L1ME3F_3-конец	.	Eutheria	L1
L1ME3F_3-конец	.	Eutheria	L1
L1ME4	.	Eutheria	L1
L1ME4A	.	Homo sapiens	L1
L1ME4A	.	Homo sapiens	L1
L1ME5	.	Eutheria	L1

L1ME5_3-конец	.	Mammalia		L1
L1ME5_3-конец	.	Mammalia		L1
L1ME5_3-конец	.	Mammalia		L1
L1MEA_5	.	Homo sapiens	L1	
L1MEA_5	.	Homo sapiens	L1	
L1MEB_5	.	Homo sapiens	L1	
L1MEB_5	.	Homo sapiens	L1	
L1MEC_5	.	Eutheria	L1	
L1MEC_5	.	Eutheria	L1	
L1MEC_5	.	Eutheria	L1	
L1MED_5	.	Homo sapiens	L1	
L1MED_5	.	Homo sapiens	L1	
L1MEE_5	.	Homo sapiens	L1	
L1MEe_5-конец	.	Eutheria	L1	
L1MEe_5-конец	.	Eutheria	L1	
L1MEe_5-конец	.	Eutheria	L1	
L1MEf_5-конец	.	Mammalia		L1
L1MEf_5-конец	.	Mammalia		L1
L1MEf_5-конец	.	Mammalia		L1
L1MEg_5-конец	.	Mammalia		L1
L1MEg_5-конец	.	Mammalia		L1
L1MEg_5-конец	.	Mammalia		L1
L1MM_F	X57795	Mus musculus		L1

L1N1_MD	.	Monodelphis domestica	L1
L1NH_RN	X05455, Y00164	Rattus norvegicus	L1
L1P_MA2	.	Eutheria	L1
L1P_MA2	.	Eutheria	L1
L1P_MA2	.	Eutheria	L1
L1P4a_5-конец	.	Primates	L1
L1P4a_5-конец	.	Primates	L1
L1P4b_5-конец	.	Primates	L1
L1P4b_5-конец	.	Primates	L1
L1P4c_5-конец	.	Primates	L1
L1P4c_5-конец	.	Primates	L1
L1P4d_5-конец	.	Primates	L1
L1P4d_5-конец	.	Primates	L1
L1P4e_5-конец	.	Primates	L1
L1P4e_5-конец	.	Primates	L1
L1PA10	.	Homo sapiens	L1
L1PA11	.	Homo sapiens	L1
L1PA12	.	Homo sapiens	L1
L1PA12_5	.	Homo sapiens	L1
L1PA13	.	Homo sapiens	L1
L1PA13_5	.	Homo sapiens	L1
L1PA14	.	Homo sapiens	L1
L1PA14_5	.	Homo sapiens	L1

L1PA15	.	Homo sapiens	L1	
L1PA16	.	Homo sapiens	L1	
L1PA16_5	.	Homo sapiens	L1	
L1PA17_5	.	Homo sapiens	L1	
L1PA2	.	Homo sapiens	L1	
L1PA3	.	Homo sapiens	L1	
L1PA4	.	Homo sapiens	L1	
L1PA5	.	Homo sapiens	L1	
L1PA6	.	Homo sapiens	L1	
L1PA7	.	Homo sapiens	L1	
L1PA7_5	.	Homo sapiens	L1	
L1PA8	.	Homo sapiens	L1	
L1PB1	.	Homo sapiens	L1	
L1PB1	.	Homo sapiens	L1	
L1PB2	.	Homo sapiens	L1	
L1PB2c	.	Homo sapiens	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1PB3	.	Homo sapiens	L1	
L1PB4	.	Homo sapiens	L1	
L1PBA_5	.	Homo sapiens	L1	
L1PBA1_5	.	Homo sapiens	L1	
L1PBB_5	.	Homo sapiens	L1	
L1PREC1	.	Homo sapiens	L1	
L1PREC2	.	Homo sapiens	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3

L1Pt	.	Pan troglodytes	L1	
L1Pt_5-конец	.	Pan troglodytes	L1	
L1R2_RN	M36995	Rattus norvegicus	L1	
L1S1_SS	X15410	Sus scrofa	L1	
L1S2_SS	X15411	Sus scrofa	L1	
L1SU_SP	007C10	Strongylocentrotus purpuratus Tx1		1,26,43,3,44,11,4,2
L1VL1_5	X53728	Mus musculus	L1	
L1VL1_5-конец	.	Mus	L1	
L1VL2_5	.	Mus musculus	L1	
L1VL2_5-конец	.	Mus	L1	
L1VL4_5-конец	.	Muridae	L1	
L1VL4a_5-конец	.	Rattus	L1	
L2	.	Eutheria	CR1	
L2	.	Eutheria	CR1	
L2	.	Eutheria	CR1	
L2_AC_1	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_10	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2_AC_11	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_12	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_13	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_14	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_15	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_16	.	Anolis carolinensis	L2	16,1,3,11,4,2,15

L2_AC_17	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2_AC_2	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_3	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_4	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_5	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_6	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_7	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_8	.	Anolis carolinensis	L2	16,1,3,11,4,2,15
L2_AC_9	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_Av	.	Adineta vaga	Penelope	1,3
L2_Ele3	.	Aedes aegypti	L2	16,1,3,11,4,2,15
L2_Plat1a	.	Ornithorhynchus	CR1	
L2_Plat1b	.	Ornithorhynchus	CR1	
L2_Plat1c	.	Ornithorhynchus	CR1	
L2_Plat1d	.	Ornithorhynchus	CR1	
L2_Plat1e	.	Ornithorhynchus	CR1	
L2_Plat1f	.	Ornithorhynchus	CR1	
L2_Plat1g	.	Ornithorhynchus	CR1	
L2_Plat1h	.	Ornithorhynchus	CR1	
L2_Plat1i	.	Ornithorhynchus	CR1	
L2_Plat1m	.	Ornithorhynchus	L2	
L2_Plat1n	.	Ornithorhynchus	CR1	
L2_Plat1o	.	Ornithorhynchus	CR1	

L2_Plat1q	.	Ornithorhynchus	CR1	
L2_Plat1r	.	Ornithorhynchus	CR1	
L2_Plat1s	.	Ornithorhynchus	CR1	
L2_Plat1t	.	Ornithorhynchus	CR1	
L2_Plat1u	.	Ornithorhynchus	CR1	
L2-1_AC	.	Aplysia californica	L2	1,3,4,2
L2-1_AFC	.	Cichlidae	L2	
L2-1_AMi	.	Alligator mississippiensis	L2	
L2-1_BM	.	Bombyx mori	Kiri	1,3,11,4,2
L2-1_CCa	.	Cyprinus carpio	L2	16,1,3,11,4,48,2,15
L2-1_CPB	.	Chrysemys picta bellii	L2	
L2-1_Croc	.	Crocodylidae	L2	1,3,11,4,2
L2-1_Crp	.	Crocodylus porosus	L2	
L2-1_CTe	.	Capitella teleta	L2	1,3,4,2
L2-1_DL	.	Dicentrarchus labrax	L2	2,1
L2-1_DPu	.	Daphnia pulex	L2	16,1,3,4,11,2,15
L2-1_DR	.	Danio rerio	L2	63,1,64,3,11,4,48,2
L2-1_EL	.	Esox lucius	L2	1,3,11,4,2
L2-1_FR	.	Takifugu rubripes	L2	1,3,4,2
L2-1_FR1	.	Takifugu rubripes	L2	16,1,3,11,4,51,2,15
L2-1_GA	.	Gasterosteus aculeatus	L2	16,1,3,4,11,48,2,15
L2-1_HMa	.	Hydra vulgaris	L2	1,4,2
L2-1_Hmel	.	Heliconius melpomene	L2	1,3,11,4,2

L2-1_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-1_HSal	.	Harpegnathos saltator	L2	
L2-1_LCh	.	Latimeria chalumnae	L2	
L2-1_NVi	.	Nasonia vitripennis	CR1	16,1,3,11,4,2,15
L2-1_OL	.	Oryzias latipes	L2	16,1,3,11,4,2,15
L2-1_PM	.	Petromyzon marinus	L2	1,3,11,4,2
L2-1_PSi	.	Pelodiscus sinensis	L2	1,3,11,4,2
L2-1_RPr	.	Rhodnius prolixus	Kiri	66,67,53,144,2,54,55,11,56,57,68,1,58,4,3,59,60,61
L2-1_SIn	.	Solenopsis invicta	L2	1,8,14,3,11,4,12,2
L2-1_SK	.	Saccoglossus kowalevskii	L2	1,3,4,11,2
L2-1_SP	.	Strongylocentrotus purpuratus	L2	1,3,4,2
L2-1_SSa	.	Salmo salar	L2	1,3,11,4,2
L2-1_XL	.	Xenopus laevis	L2	1,3,11,4,2
L2-10_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-10_EL	.	Esox lucius	L2	
L2-10_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-10_HRo	.	Helobdella robusta	L2	53,2,54,24,11,56,57,68,1,58,4,16,3,59,60
L2-10_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-11_CTe	.	Capitella teleta	L2	1,3,4,2
L2-11_DRe	.	Danio rerio	L2	1,3,2
L2-11_EL	.	Esox lucius	L2	1,3,11,4,2
L2-11_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-11_HRo	.	Helobdella robusta	L2	53,2,54,11,56,57,68,1,58,4,3,59

L2-12_CTe	.	Capitella teleta	L2	1,3,4,2	
L2-12_EL	.	Esox lucius	L2		
L2-12_Hmel	.	Heliconius melpomene	L2		16,1,3,11,4,2,15
L2-12_HRo	.	Helobdella robusta	L2		53,2,54,11,57,68,1,4,15,16,3,59,60
L2-12a_Lch	.	Latimeria chalumnae	L2		
L2-12b_Lch	.	Latimeria chalumnae	L2		
L2-13_CTe	.	Capitella teleta	L2	1,3,4,2	
L2-13_DRe	.	Danio rerio	L2		1,3,4,2
L2-13_Hmel	.	Heliconius melpomene	L2		1,3,4,2
L2-13_HRo	.	Helobdella robusta	L2		1,3,11,4,2
L2-13a_Lch	.	Latimeria chalumnae	L2		
L2-13B_DRe	.	Danio rerio	L2		2,1
L2-13b_Lch	.	Latimeria chalumnae	L2		1,3,11,4,2
L2-13c_Lch	.	Latimeria chalumnae	L2		
L2-14_CTe	.	Capitella teleta	L2	1,3,4,2	
L2-14_DRe	.	Danio rerio	L2		
L2-14_EL	.	Esox lucius	L2		16,1,3,11,4,2,15
L2-14_Hmel	.	Heliconius melpomene	L2		1,3,11,4,2
L2-14_HRo	.	Helobdella robusta	L2		53,2,54,11,57,1,4,15,16,3,59
L2-15_CTe	.	Capitella teleta	L2	1,3,4,2	
L2-15_Hmel	.	Heliconius melpomene	L2		1,3,4,2
L2-15_HRo	.	Helobdella robusta	L2		53,2,54,11,57,1,4,15,16,3,60
L2-16_CTe	.	Capitella teleta	L2	1,3,4,2	

L2-16_DRe	.	Danio rerio	L2	63,16,1,64,3,11,4,48,2,15
L2-16_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3,59
L2-16_NVe	.	Nematostella vectensis	L2	
L2-17_CTe	.	Capitella teleta	L2	1,3,4,2
L2-17_DRe	.	Danio rerio	L2	16,1,3,4,11,2,15
L2-17_HRo	.	Helobdella robusta	L2	53,2,54,11,56,57,68,1,58,4,15,16,3,59,60
L2-17_NV	.	Nematostella vectensis	L2	
L2-18_ACar	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2-18_CTe	.	Capitella teleta	L2	1,3,4,2
L2-18_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3
L2-19_ACar	.	Anolis carolinensis	L2	1,3,11,4,2
L2-19_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-19_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,3,59,60
L2-1a_Cis	.	Ciona savignyi	CR1	21,16,1,3,24,11,4,2
L2-1b_Cis	.	Ciona savignyi	CR1	53,2,21,54,24,11,56,57,1,58,4,16,3,59,60
L2-1B_DR	.	Danio rerio	L2	
L2-1c_Cis	.	Ciona savignyi	CR1	2,21,24,11,56,57,1,58,4,16,3,60
L2-1d_3UTR_Cis	.	Ciona savignyi	CR1	
L2-2_CCa	.	Cyprinus carpio	L2	16,1,3,11,4,2,15
L2-2_Cis	.	Ciona savignyi	CR1	2,21,55,11,56,57,1,58,4,3,59,60,61
L2-2_CM	.	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-2_CPB	.	Chrysemys picta bellii	L2	
L2-2_Croc	.	Crocodylidae	L2	1,3,4,2

L2-2_CTe	.	Capitella teleta	L2	1,3,2
L2-2_DL	.	Dicentrarchus labrax	L2	1,3,4,2
L2-2_DRe	.	Danio rerio	L2	16,1,3,24,11,4,2
L2-2_EL	.	Esox lucius	L2	1,3,4,11,2
L2-2_GA	.	Gasterosteus aculeatus	L2	1,3,11,4,2
L2-2_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-2_HRo	.	Helobdella robusta	L2	53,2,54,11,56,57,68,1,4,15,16,127,3,59
L2-2_LMe	.	Latimeria menadoensis	L2	1,3,11,4,2
L2-2_MD	.	Monodelphis domestica	CR1	
L2-2_ME	.	Notamacropus eugenii	CR1	
L2-2_NVi	.	Nasonia vitripennis	CR1	1,3,4,2
L2-2_OL	.	Oryzias latipes	L2	16,1,3,4,11,2,15
L2-2_PM	.	Petromyzon marinus	L2	16,1,3,11,4,2,15
L2-2_SIn	.	Solenopsis invicta	L2	16,1,8,14,3,11,4,12,2,15
L2-2_SP	.	Strongylocentrotus purpuratus	L2	16,1,3,11,4,2,15
L2-2_SSa	.	Salmo salar	L2	16,1,3,11,4,2,15
L2-2_XL	.	Xenopus laevis	L2	
L2-2_XT	.	Xenopus tropicalis	L2	16,1,3,24,4,11,2
L2-20_ACar	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-20_CTe	.	Capitella teleta	L2	16,1,3,11,4,2,15
L2-20_HRo	.	Helobdella robusta	L2	2,11,57,68,1,4,3,59,60
L2-21_ACar	.	Anolis carolinensis	L2	
L2-21_CTe	.	Capitella teleta	L2	1,3,4,2

L2-21_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-22_ACar	.	Anolis carolinensis	L2	1,3,4,2
L2-22_CTe	.	Capitella teleta	L2	1,3,4,2
L2-22_HRo	.	Helobdella robusta	L2	2,11,56,57,68,1,159,58,4,15,16,160,3,59
L2-22_NV	.	Nematostella vectensis	CR1	1,3,4,11,2
L2-23_ACar	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-23_CTe	.	Capitella teleta	L2	1,3,4,2
L2-23_HRo	.	Helobdella robusta	L2	16,1,23,3,4,11,2,15
L2-24_ACar	.	Anolis carolinensis	L2	1,3,4,2
L2-24_CTe	.	Capitella teleta	L2	1,3,2
L2-24_HRo	.	Helobdella robusta	L2	54,1,3,59,53,11,4,57,2
L2-24_NV	.	Nematostella vectensis	L2A	1,3,11,4,2
L2-25_ACar	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-25_CTe	.	Capitella teleta	L2	1,3,4,2
L2-25_HRo	.	Helobdella robusta	L2	2,24,11,56,57,1,58,4,16,3,59
L2-26_ACar	.	Anolis carolinensis	L2	
L2-26_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-26_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-26B_HRo	.	Helobdella robusta	L2	16,11,15
L2-27_ACar	.	Anolis carolinensis	L2	1,3,4,2
L2-27_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-29_ACar	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-29_HRo	.	Helobdella robusta	L2	1,3,4,11,2

L2-29B_ACar .	Anolis carolinensis	L2	1,3,11,4,2
L2-2B_SSa .	Salmo salar	L2	16,1,3,11,4,2,15
L2-2C_SSa .	Salmo salar	L2	16,1,3,11,4,2,15
L2-3_AFC .	Cichlidae	L2	
L2-3_AMi .	Alligator mississippiensis	L2	
L2-3_Cis .	Ciona savignyi	CR1	53,2,21,54,11,56,57,68,1,58,4,3,59
L2-3_CM .	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-3_CPB .	Chrysemys picta bellii	L2	1,3,4,2
L2-3_Crp .	Crocodylus porosus	L2	
L2-3_CTe .	Capitella teleta	L2	1,3,4,2
L2-3_DL .	Dicentrarchus labrax	L2	
L2-3_DRe .	Danio rerio	L2	1,3,11,4,2
L2-3_EL .	Esox lucius	L2	1,3,11,4,2
L2-3_GA .	Gasterosteus aculeatus	L2	16,1,3,11,4,2,15
L2-3_Hmel .	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-3_HRo .	Helobdella robusta	L2	60,57
L2-3_LCh .	Latimeria chalumnae	L2	1,3,4,2
L2-3_LMe .	Latimeria menadoensis	L2	1,3,11,4,2
L2-3_NVi .	Nasonia vitripennis	CR1	16,1,3,11,2,15
L2-3_OL .	Oryzias latipes	L2	1,3,4,2
L2-3_PM .	Petromyzon marinus	L2	
L2-3_SSa .	Salmo salar	L2	16,1,3,11,4,48,2,15
L2-3_XL .	Xenopus laevis	L2	1,3,11,4,2

L2-3_XT	.	Xenopus tropicalis	CR1	1,3,11,4,2
L2-30_ACar	.	Anolis carolinensis	L2	
L2-30_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-31_ACar	.	Anolis carolinensis	L2	
L2-31_HRo	.	Helobdella robusta	L2	2,11,57,1,159,4,160,3,60
L2-32_DRe	.	Danio rerio	L2	1,3,4,2
L2-32_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-33_DRe	.	Danio rerio	L2	2,1
L2-33_HRo	.	Helobdella robusta	L2	1,3,11,4,2
L2-34_DR	.	Danio rerio	L2	16,1,3,11,4,2,15
L2-34_HRo	.	Helobdella robusta	L2	16,1,3,4,11,2,15
L2-35_DRe	.	Danio rerio	L2	1,3,4,2
L2-35_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-36_DRe	.	Danio rerio	L2	1,3,4,2
L2-36_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-36B_DRe	.	Danio rerio	L2	
L2-37_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-38_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-39_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-3b_Cis	.	Ciona savignyi	CR1	
L2-3B_CM-3-конец	.	Callorhinchus milii	L2	
L2-3C_CM-3-конец	.	Callorhinchus milii	L2	
L2-3D_CM	.	Callorhinchus milii	L2	

L2-3E_CM	.	Callorhinchus milii	L2	
L2-4_AFC	.	Cichlidae	L2	1,3,2
L2-4_Cis	.	Ciona savignyi	Crack	53,2,21,54,11,56,57,68,1,159,58,4,15,16,160,3,59,60
L2-4_CM	.	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-4_CPB	.	Chrysemys picta bellii	L2	
L2-4_CTe	.	Capitella teleta	L2	1,3,4,2
L2-4_DL	.	Dicentrarchus labrax	L2	
L2-4_DRe	.	Danio rerio	L2	16,1,3,24,11,4,2
L2-4_EL	.	Esox lucius	L2	2,111,11,1,4,112,15,16,3
L2-4_GA	.	Gasterosteus aculeatus	L2	1,3,11,4,2
L2-4_Hmel	.	Heliconius melpomene	L2	1,3,4,11,2
L2-4_HRo	.	Helobdella robusta	L2	1,3,11,4,2
L2-4_LCh	.	Latimeria chalumnae	L2	
L2-4_NVi	.	Nasonia vitripennis	CR1	54,1,3,53,4,11,2
L2-4_OL	.	Oryzias latipes	L2	1,11,2
L2-4_PM	.	Petromyzon marinus	L2	
L2-4_SSa	.	Salmo salar	L2	1,3,11,4,2
L2-4_XL	.	Xenopus laevis	L2	
L2-4_XT	.	Xenopus tropicalis	CR1	1,3,11,4,2
L2-40_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-41_DRe	.	Danio rerio	L2	16,1,3,24,11,4,2
L2-41_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-42_DRe	.	Danio rerio	L2	1,3,11,4,2

L2-43_DR	.	Danio rerio	L2	1,3,4,2
L2-43_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-44_DR	.	Danio rerio	L2	1,3,2
L2-45_DR	.	Danio rerio	L2	
L2-46_DR	.	Danio rerio	L2	1,3,2
L2-47_DR	.	Danio rerio	L2	
L2-48_DR	.	Danio rerio	L2	1,3,2
L2-49_DR	.	Danio rerio	L2	1,3,4,2
L2-4B_CM	.	Callorhinchus milii	L2	
L2-4B_EL	.	Esox lucius	L2	
L2-4C_CM-3-конец	.	Callorhinchus milii	L2	
L2-4D_CM	.	Callorhinchus milii	L2	
L2-5_AFC	.	Cichlidae	L2	1,3,11,4,2
L2-5_CPB	.	Chrysemys picta bellii	L2	
L2-5_CTe	.	Capitella teleta	L2	1,3,4,2
L2-5_DRe	.	Danio rerio	L2	1,3,4,2
L2-5_EL	.	Esox lucius	L2	1,3,11,4,48,2
L2-5_GA	.	Gasterosteus aculeatus	L2	16,1,3,24,11,4,2
L2-5_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-5_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3
L2-5_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-5_NVi	.	Nasonia vitripennis	CR1	16,1,3,4,11,2,15
L2-5_OL	.	Oryzias latipes	L2	63,16,1,64,3,11,4,2,15

L2-5_PM	.	Petromyzon marinus	L2	1,3,11,4,2
L2-5_SSa	.	Salmo salar	L2	16,1,3,11,4,2,15
L2-5_XT	.	Xenopus tropicalis	CR1	16,1,3,11,4,2,15
L2-50_DR	.	Danio rerio	L2	
L2-51_DR	.	Danio rerio	L2	
L2-52_DR	.	Danio rerio	L2	
L2-53_DR	.	Danio rerio	L2	
L2-54_DR	.	Danio rerio	L2	
L2-55_DR	.	Danio rerio	L2	
L2-56_DR	.	Danio rerio	L2	1,3,2
L2-57_DR	.	Danio rerio	L2	
L2-58_DR	.	Danio rerio	L2	
L2-59_DR	.	Danio rerio	L2	
L2-5B_EL	.	Esox lucius	L2	
L2-5B_HRo	.	Helobdella robusta	L2	16,1,3,11,4,57,2,15
L2-5N1_SSa	.	Salmo salar	L2	
L2-5N2_SSa	.	Salmo salar	L2	
L2-6_AFC	.	Cichlidae	L2	
L2-6_CPB	.	Chrysemys picta bellii	L2	
L2-6_CTe	.	Capitella teleta	L2	1,3,4,2
L2-6_DRe	.	Danio rerio	L2	1,3,4,2
L2-6_EL	.	Esox lucius	L2	16,1,3,11,4,2,15
L2-6_Hmel	.	Heliconius melpomene	L2	1,3,4,2

L2-6_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-6_LCh	.	Latimeria chalumnae	L2	1,3,11,4,2
L2-6_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-6_OL	.	Oryzias latipes	L2	1,3,11,4,2
L2-6_PM	.	Petromyzon marinus	L2	
L2-6_SSa	.	Salmo salar	L2	1,2
L2-6_XT	.	Xenopus tropicalis	CR1	1,3,4,2
L2-68_HMa	.	Hydra vulgaris	L2	16,1,3,11,4,2,15
L2-7_CPB	.	Chrysemys picta bellii	L2	
L2-7_CTe	.	Capitella teleta	L2	1,3,2
L2-7_DRe	.	Danio rerio	L2	
L2-7_EL	.	Esox lucius	L2	1,3,11,2
L2-7_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-7_HRo	.	Helobdella robusta	L2	53,2,54,24,11,56,57,68,1,58,4,16,3,59
L2-7_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-7_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-7_OL	.	Oryzias latipes	L2	1,3,4,2
L2-7_PM	.	Petromyzon marinus	L2	16,1,3,11,4,2,15
L2-7_SSa	.	Salmo salar	L2	
L2-8_AAe	.	Aedes aegypti	L2	53,2,54,11,56,57,68,1,58,4,3,59
L2-8_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-8_DR	.	Danio rerio	L2	1,3,11,4,2
L2-8_EL	.	Esox lucius	L2	1,3,11,4,2

L2-8_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-8_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-8_LCh	.	Latimeria chalumnae	L2	1,3,11,4,2
L2-8_NVi	.	Nasonia vitripennis	CR1	1,3,60,4,11,57,2
L2-8_OL	.	Oryzias latipes	L2	2,1
L2-9_AAe	.	Aedes aegypti	L2	53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
L2-9_CTe	.	Capitella teleta	L2	16,1,3,11,4,2,15
L2-9_DRe	.	Danio rerio	L2	16,1,3,11,4,2,15
L2-9_EL	.	Esox lucius	L2	1,3,2
L2-9_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-9_HRo	.	Helobdella robusta	L2	53,2,54,11,57,68,1,4,15,16,3,59
L2-9_LCh	.	Latimeria chalumnae	L2	1,3,4,11,2
L2-9_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-9_OL	.	Oryzias latipes	L2	1,3,11,4,2
L2-N1_HRo	.	Helobdella robusta	L2	
L2A-1_NVe	.	Nematostella vectensis	L2A	66,67,2,11,56,57,68,1,58,15,16,59,60
L2A-1_PPo	.	Papilio polytes	L2	
L2A-1_SK	.	Saccoglossus kowalevskii	L2A	1,3,11,4,2
L2A-2_SK	.	Saccoglossus kowalevskii	L2A	1,3,4,2
L2B	.	Eutheria	CR1	
L2B	.	Eutheria	CR1	
L2B	.	Eutheria	CR1	
L2B_ME	.	Notamacropus eugenii	CR1	

L2B-1_AAe	.	Aedes aegypti L2B	16,1,3,4,11,2,15
L2B-1_AG	.	Anopheles gambiae L2B	54,1,8,3,59,53,11,4,12,2
L2B-1_BTa	.	Bemisia tabaci L2B	1,3,11,4,2
L2B-1_CP	.	Culex pipiens L2B	2,8,11,57,12,1,4,15,16,3,59
L2B-1_CQ	.	Culex quinquefasciatus L2B	1,3,59,11,4,2
L2B-1_HM	.	Hydra vulgaris L2B	1,3,59,4,11,57,2,68
L2B-1_LMi	.	Locusta migratoria L2B	1,3,11,2
L2B-1_MLe	.	Mnemiopsis leidyi L2B	
L2B-1_PBa	.	Pogonomyrmex barbatus L2B	1,3,11,4,2
L2B-1B_CQ	.	Culex quinquefasciatus L2B	1,3,4,2
L2B-2_AAe	.	Aedes aegypti L2B	54,1,3,59,53,11,4,57,2
L2B-2_BTa	.	Bemisia tabaci L2B	1,54,3,59,53,4,11,57,2
L2B-2_CQ	.	Culex quinquefasciatus L2B	54,1,3,59,53,11,4,57,2
L2B-2_PBa	.	Pogonomyrmex barbatus L2B	16,1,3,11,4,2,15
L2B-3_AAe	.	Aedes aegypti L2B	53,2,54,11,57,1,4,15,16,3,59
L2B-3_CQ	.	Culex quinquefasciatus L2B	16,1,8,14,3,11,4,12,2,15
L2B-3_PBa	.	Pogonomyrmex barbatus L2B	1,3,11,4,2
L2B-3B_AAe	.	Aedes aegypti L2B	1,3,11,4,2
L2B-3C_AAe	.	Aedes aegypti L2B	1,3,11,4,2
L2B-4_AAe	.	Aedes aegypti L2B	53,2,54,8,24,11,57,12,68,1,14,4,16,3,59
L2B-4_CQ	.	Culex quinquefasciatus L2B	1,3,59,11,4,2
L2B-4B_AAe	.	Aedes aegypti L2B	42,1,8,3,59,4,11,12,2
L2B-5_AAe	.	Aedes aegypti L2B	53,2,54,11,57,1,4,15,16,3,59

L2B-5_CQ	.	Culex quinquefasciatus	L2B		1,3,11,4,2
L2B-6_AAe	.	Aedes aegypti	L2B		1,3,11,4,2
L2B-6_CQ	.	Culex quinquefasciatus	L2B		1,3,11,4,2
L2B-7_AAe	.	Aedes aegypti	L2B		1,3,4,11,2
L2C	.	Eutheria	L2		
L2C	.	Eutheria	L2		
L2C	.	Eutheria	L2		
L2D	.	Eutheria	L2		
L2D	.	Eutheria	L2		
L2D	.	Eutheria	L2		
L3	.	Homo sapiens	CR1		16,1,3,4,11,2,15
L3_Av	.	Adineta vaga	Penelope		165,1,166,167,3,168,130,169
L3_ME	.	Notamacropus eugenii	CR1		
L3b_3-конец	.	Mammalia	CR1		
L3b_3-конец	.	Mammalia	CR1		
L3b_3-конец	.	Mammalia	CR1		
L4	.	Mammalia	RTEX		1,3,4,2
L4	.	Mammalia	RTEX		1,3,4,2
L4	.	Mammalia	RTEX		1,3,4,2
L5	.	Mammalia	RTEX		
L5	.	Mammalia	RTEX		
L5	.	Mammalia	RTEX		
Lam1	.	Lethenteron reissneri	SINE2/тРНК		

LDT1	AF081103	Lymantria dispar	Jockey	16,1,87,3,88,11,4,2,15
LFR1_LP	.	Lepidosiren paradoxa	Не содержащий LTR ретротранспозон	
LFSINE_AMi	.	Alligator mississippiensis	SINE2/тРНК	
LFSINE_Vert	.	Latimeria menadoensis	SINE2/тРНК	
Lgi-Nin-DC-SINE1	.	Lottia gigantea	SINE2/тРНК	
Lgi-Nin-DC-SINE3	.	Lottia gigantea	SINE2/тРНК	
Lian-Aa1	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LIN10_SM	.	Schmidtea mediterranea	L1	1,3,120,4,2
LIN10B_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN11_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN13_SM	.	Schmidtea mediterranea	NeSL	1,215,3,4,2
LIN14_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2
LIN15_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN2_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2
LIN21_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2
LIN23_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,120,149,4,2
LIN24_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN24B_SM	.	Schmidtea mediterranea	NeSL	
LIN25_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,120,149,4,2
LIN26_SM	.	Schmidtea mediterranea	NeSL	1,3,120,82,4,2
LIN3_SM	.	Schmidtea mediterranea	NeSL	147,1,148,149,120,3,4,2
LIN4_SM	.	Schmidtea mediterranea	NeSL	2,1,3,120,4
LIN4b_SM	.	Schmidtea mediterranea	NeSL	1,148,3,149,120,4,2

LIN5_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2
LIN6_SM	.	Schmidtea mediterranea	NeSL	1,181,3,4,2
LIN7_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN7B_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
LIN9_SM	.	Schmidtea mediterranea	NeSL	147,18,5,1,148,3,149,120,4,2
LINE_CA	AJ411792, AJ411792.1	Cicer arietinum	L1	
LINE_CH	X72221	Capra hircus	L1	
LINE_LS	Z17425	Lilium speciosum	L1	
LINE-05_OS	.	Oryza sativa	L1	
LINE-1	.	Oryza sativa	L1	
LINE-1_AA	M95171	Aedes aegypti	Jockey	
LINE-4_OS	.	Oryza sativa	L1	43,2,8,44,11,12,42,1,14,4,3
LINE-6_OS	.	Oryza sativa	L1	
LINE-7_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3
LINE-8_CR	.	Chlamydomonas reinhardtii	L1	
LINE1_BM	D26009	Bombyx mori	Jockey	
LINE1_CA	AJ411789	Cicer arietinum	L1	
LINE1-1_SBi	.	Sorghum bicolor	L1	
LINE1-1_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,248,2
LINE1-10_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-10_SBi	.	Sorghum bicolor	L1	
LINE1-10_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-11_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2

LINE1-11_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-11_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-11B_OS	.	Oryza sativa	L1	
LINE1-11B_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-12_OS	.	Oryza sativa	L1	
LINE1-12_SBi	.	Sorghum bicolor	L1	16,1,43,75,3,76,44,24,11,4,2
LINE1-12_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-13_OS	.	Oryza sativa	L1	
LINE1-13_SBi	.	Sorghum bicolor	L1	
LINE1-13_ZM	.	Zea mays	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-14_OS	.	Oryza sativa	L1	
LINE1-14_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-14_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-15_OS	.	Oryza sativa	L1	
LINE1-15_SBi	.	Sorghum bicolor	L1	8,14,12
LINE1-15_ZM	.	Zea mays	L1	250,8,14,251,12
LINE1-16_OS	.	Oryza sativa	L1	
LINE1-16_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,12,1,14,4,16,75,3,76
LINE1-16_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-17_OS	.	Oryza sativa	L1	
LINE1-17_SBi	.	Sorghum bicolor	L1	1,14,3,11,4,2
LINE1-17_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-18_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2

LINE1-18_SBi	.	Sorghum bicolor	L1	1,14,3,4,12,2
LINE1-18_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-19_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-19_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-19_ZM	.	Zea mays	L1	1,4,2
LINE1-19B_OS	.	Oryza sativa	L1	
LINE1-2_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-2_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-20_OS	.	Oryza sativa	L1	1,43,14,3,44,11,4,2
LINE1-20_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-20_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-21_OS	.	Oryza sativa	L1	
LINE1-21_SBi	.	Sorghum bicolor	L1	
LINE1-21_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-22_OS	.	Oryza sativa	L1	
LINE1-22_SBi	.	Sorghum bicolor	L1	
LINE1-22_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-23_OS	.	Oryza sativa	L1	
LINE1-23_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-23_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-24_OS	.	Oryza sativa	L1	
LINE1-24_SBi	.	Sorghum bicolor	L1	
LINE1-24_ZM	.	Zea mays	L1	94,16,95,24,11

LINE1-25_OS.	Oryza sativa	L1	44,43
LINE1-25_SBi .	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-25_ZM .	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-26_OS.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-26_SBi .	Sorghum bicolor	L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-26_ZM .	Zea mays	L1	1,3,4,2
LINE1-27_OS.	Oryza sativa	L1	
LINE1-27_SBi .	Sorghum bicolor	L1	94,1,43,3,44,95,11,4,2
LINE1-27_ZM .	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-28_OS.	Oryza sativa	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-28_SBi .	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-28_ZM .	Zea mays	L1	94,16,1,3,95,24,11,4,2
LINE1-29_OS.	Oryza sativa	L1	
LINE1-29_SBi .	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-29_ZM .	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-3_SBi .	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-3_ZM .	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-30_OS.	Oryza sativa	L1	
LINE1-30_SBi .	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-30_ZM .	Zea mays	L1	1,43,3,44,4,2
LINE1-31_OS.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-31_SBi .	Sorghum bicolor	L1	1,3,11,4,2
LINE1-31_ZM .	Zea mays	L1	1,3,11,4,2

LINE1-32_OS.		Oryza sativa	L1	
LINE1-32_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-32_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-33_OS.		Oryza sativa	L1	
LINE1-33_SBi	.	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-33_ZM	.	Zea mays	L1	94,16,1,3,95,24,11,4,2
LINE1-34_OS.		Oryza sativa	L1	
LINE1-34_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-34_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-35_OS.		Oryza sativa	L1	
LINE1-35_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-35_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-36_OS.		Oryza sativa	L1	
LINE1-36_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,42,1,14,4,3
LINE1-36_ZM	.	Zea mays	L1	1,3,4,2
LINE1-37_OS.		Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-37_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-37_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-38_OS.		Oryza sativa	L1	
LINE1-38_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-38_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-39_OS.		Oryza sativa	L1	43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7
LINE1-39_SBi	ABXC01004903.1	Sorghum bicolor	L1	1,43,3,44,11,2

LINE1-39_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-4_SBi	.	Sorghum bicolor	L1	
LINE1-4_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-40_OS	.	Oryza sativa	L1	
LINE1-40_SBi	ABXC01001255.1	Sorghum bicolor	L1	43,2,44,24,11,12,1,4,16,3
LINE1-40_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,42,1,14,4,16,75,3,76
LINE1-41_OS	.	Oryza sativa	L1	
LINE1-41_SBi	ABXC01007042.1	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-41_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-42_OS	.	Oryza sativa	L1	
LINE1-42_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,42,1,14,4,3
LINE1-42_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,94,16,3,95
LINE1-43_OS	.	Oryza sativa	L1	
LINE1-43_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-43_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-44_OS	.	Oryza sativa	L1	16,1,43,3,44,24,4,11,2
LINE1-44_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,12,1,14,4,16,3
LINE1-44_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-45_OS	.	Oryza sativa	L1	43,2,44,24,11,1,14,4,16,75,3,76
LINE1-45_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-45_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-46_OS	.	Oryza sativa	L1	
LINE1-46_SBi	.	Sorghum bicolor	L1	1,43,75,3,44,76,11,4,2

LINE1-46_ZM	.	Zea mays	L1	43,2,44,24,11,1,4,94,16,3,95
LINE1-47_OS	.	Oryza sativa	L1	
LINE1-47_SBi	.	Sorghum bicolor	L1	44,43
LINE1-47_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-48_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-48_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-48_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,75,3,76
LINE1-49_OS	.	Oryza sativa	L1	
LINE1-49_SBi	.	Sorghum bicolor	L1	1,43,3,44,2
LINE1-49_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,4,16,3
LINE1-5_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-5_ZM	.	Zea mays	L1	43,2,44,24,11,1,14,4,16,75,3,76
LINE1-50_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-50_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-50_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-51_OS	.	Oryza sativa	L1	
LINE1-51_SBi	.	Sorghum bicolor	L1	1,43,14,3,44,11,4,2
LINE1-51_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,94,16,3,95
LINE1-52_OS	.	Oryza sativa	L1	
LINE1-52_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,4,11,2
LINE1-52_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-53_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-53_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3

LINE1-53_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-54_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-54_SBi	.	Sorghum bicolor	L1	7,1,43,17,3,44,10,79,4,2,80
LINE1-54_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-55_OS	.	Oryza sativa	L1	
LINE1-55_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,4,11,2
LINE1-55_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-56_OS	.	Oryza sativa	L1	
LINE1-56_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-56_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-57_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-57_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-57_ZM	.	Zea mays	L1	1,43,75,3,76,44,11,4,2
LINE1-58_OS	.	Oryza sativa	L1	43,78,79,2,7,8,10,44,24,11,12,80,1,81,77,4,16,75,17,3
LINE1-58_SBi	.	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-59_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3
LINE1-59_SBi	.	Sorghum bicolor	L1	1,43,75,76,44,11,4,2
LINE1-59_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-6_OS	AC113336.2	Oryza sativa	L1	43,78,79,2,8,44,10,11,12,80,1,81,77,4,17,75,3,76,7
LINE1-6_SBi	.	Sorghum bicolor	L1	
LINE1-6_ZM	.	Zea mays	L1	43,2,8,44,11,12,1,14,4,75,3,76
LINE1-60_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-60_SBi	.	Sorghum bicolor	L1	43,78,79,2,7,10,44,11,80,1,81,77,4,249,75,17,3,76

LINE1-61_OS.		Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-61_SBi	.	Sorghum bicolor	L1	
LINE1-62_OS.		Oryza sativa	L1	43,2,8,44,11,12,1,14,4,3
LINE1-62_SBi	.	Sorghum bicolor	L1	
LINE1-62B_SBi	.	Sorghum bicolor	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
LINE1-63_OS.		Oryza sativa	L1	1,43,3,44,4,2
LINE1-63_SBi	.	Sorghum bicolor	L1	1,43,3,44
LINE1-64_OS.		Oryza sativa	L1	43,79,2,10,44,11,80,1,4,17,3,7
LINE1-64_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-65_OS.		Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-65_SBi	.	Sorghum bicolor	L1	
LINE1-66_OS.		Oryza sativa	L1	43,2,44,11,12,1,14,4,3
LINE1-66_SBi	.	Sorghum bicolor	L1	43,78,79,2,8,44,10,11,12,80,1,14,81,77,4,17,75,3,76,7
LINE1-67_OS.		Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-67_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-68_OS.		Oryza sativa	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-68_SBi	.	Sorghum bicolor	L1	43,78,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-69_OS.		Oryza sativa	L1	43,78,79,2,7,8,10,44,11,12,80,1,81,77,4,75,17,3,76
LINE1-69_SBi	.	Sorghum bicolor	L1	43,2,44,11,12,1,14,4,3
LINE1-7_OS	AC146581.1	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-7_SBi	.	Sorghum bicolor	L1	43,82,2,8,44,24,11,12,1,14,4,16,3
LINE1-7_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-70_OS.		Oryza sativa	L1	43,2,44,24,11,1,4,16,75,3,76

LINE1-70_SBi	.	Sorghum bicolor	L1		43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7
LINE1-71_OS	.	Oryza sativa		L1	1,43,3,44,4,11,2
LINE1-71_SBi	.	Sorghum bicolor	L1		1,43,3,44,11,4,2
LINE1-72_OS	.	Oryza sativa		L1	
LINE1-72_SBi	.	Sorghum bicolor	L1		7,1,17,43,3,44,10,11,4,2,80
LINE1-73_OS	.	Oryza sativa		L1	16,1,43,3,44,24,11,4,2
LINE1-73_SBi	.	Sorghum bicolor	L1		1,43,3,44,4,2
LINE1-74_OS	.	Oryza sativa		L1	43,2,8,44,11,12,1,14,4,75,3,76
LINE1-74_SBi	.	Sorghum bicolor	L1		1,43,3,44,11,4,2
LINE1-75_OS	.	Oryza sativa		L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,76,7
LINE1-76_OS	.	Oryza sativa		L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-77_OS	.	Oryza sativa		L1	43,78,79,6,2,7,8,10,44,11,12,80,1,81,77,4,75,17,3,76
LINE1-78_OS	.	Oryza sativa		L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-79_OS	.	Oryza sativa		L1	16,1,43,3,44,24,11,4,2
LINE1-8_OS	AP004587.3	Oryza sativa		L1	43,2,44,24,11,12,1,14,4,16,3
LINE1-8_SBi	.	Sorghum bicolor	L1		43,2,8,44,24,11,12,1,14,4,16,3
LINE1-8_ZM	.	Zea mays	L1		1,43,3,44,11
LINE1-80_OS	.	Oryza sativa		L1	1,43,3,44,4,2
LINE1-81_OS	.	Oryza sativa		L1	
LINE1-82_OS	.	Oryza sativa		L1	1,43,3,44,4,2
LINE1-83_OS	.	Oryza sativa		L1	
LINE1-84_OS	.	Oryza sativa		L1	
LINE1-9_OS	.	Oryza sativa		L1	

LINE1-9_SBi .	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-9_ZM .	Zea mays	L1	1,43,3,44,4,2
LINE1-9B_OS .	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-N1_OS .	Oryza sativa	L1	
LINE1-N1B_OS .	Oryza sativa	L1	
LINE1-N1C_OS .	Oryza sativa	L1	
LINE1-N1D_OS .	Oryza sativa	L1	
LINE1-N1E_OS .	Oryza sativa	L1	
LINE1-N1F_OS .	Oryza sativa	L1	
LINE1-N1FB_OS .	Oryza sativa	L1	
LINE1-N1G_OS .	Oryza sativa	L1	
LINE1-N1H_OS .	Oryza sativa	L1	
LINE1-N1I_OS .	Oryza sativa	L1	
LINE1-N1J_OS .	Oryza sativa	L1	
LINE1-N1K_OS .	Oryza sativa	L1	
LINE1-N1L_OS .	Oryza sativa	L1	
LINE1-N1N_OS .	Oryza sativa	L1	
LINE1-N1N2_OS .	Oryza sativa	L1	
LINE1-N1P_OS .	Oryza sativa	L1	
LINE1-N1Q_OS .	Oryza sativa	L1	
LINE1-N1QB_OS .	Oryza sativa	L1	
LINE1-N1R_OS .	Oryza sativa	L1	
LINE1-N1S_OS .	Oryza sativa	L1	

LINE1-N1T_OS	.	Oryza sativa	L1	
LINE1-N1U_OS	.	Oryza sativa	L1	
LINE1-N1UB_OS	.	Oryza sativa	L1	
LINE1-N1V_OS	.	Oryza sativa	L1	
LINE1-N1W_OS	.	Oryza sativa	L1	
LINE1-N1WB_OS	.	Oryza sativa	L1	
LINE1-N1X_OS	.	Oryza sativa	L1	
LINE1-N1X2_OS	.	Oryza sativa	L1	
LINE1-N1Y_OS	.	Oryza sativa	L1	
LINE1-N1Z_OS	.	Oryza sativa	L1	
LINE1-N2_OS	.	Oryza sativa	L1	
LINE1-N3_OS	.	Oryza sativa	L1	
LINE1A_MT	AC148776	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A_MT	AC148776	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A1_MT	AC122166	Medicago truncatula	L1	16,1,24,11,4,2
LINE1A1_MT	AC122166	Medicago truncatula	L1	16,1,24,11,4,2
LINE1A2_MT	.	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A2_MT	.	Medicago truncatula	L1	1,43,3,44,4,2
LINE1B_MT	AC135415	Medicago truncatula	L1	7,1,43,17,3,44,10,79,4,6,2,80
LINE1B_MT	AC135415	Medicago truncatula	L1	7,1,43,17,3,44,10,79,4,6,2,80
LINE1C_MT	AC137081	Medicago truncatula	L1	43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76
LINE1C_MT	AC137081	Medicago truncatula	L1	43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76
LINE1D_MT	AC153004	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7

LINE1D_MT	AC153004	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
LINE1E_MT	AC148239	Medicago truncatula	L1	43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
LINE1E_MT	AC148239	Medicago truncatula	L1	43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
LINE1F_MT	AC140545	Medicago truncatula	L1	43,79,2,8,44,10,11,12,80,1,77,4,75,17,3,76,7
LINE1F_MT	AC140545	Medicago truncatula	L1	43,79,2,8,44,10,11,12,80,1,77,4,75,17,3,76,7
LINE1G_MT	.	Medicago truncatula	L1	43,75,76,44
LINE1G_MT	.	Medicago truncatula	L1	43,75,76,44
LINE1H_MT	.	Medicago truncatula	L1	43,79,2,44,10,11,80,1,4,17,3,7
LINE1H_MT	.	Medicago truncatula	L1	43,79,2,44,10,11,80,1,4,17,3,7
LINE1I_MT	AC165446	Medicago truncatula	L1	43,1,3,44,11,2
LINE1I_MT	AC165446	Medicago truncatula	L1	43,1,3,44,11,2
LINE1J_MT	AC144515	Medicago truncatula	L1	16,1,3,24,11,4,2
LINE1J_MT	AC144515	Medicago truncatula	L1	16,1,3,24,11,4,2
LINE2_CH1	.	Crotalus horridus	CR1	
LINE2_EC1	.	Echis coloratus	CR1	
LINE2_NT1	.	Natrix tessellata	CR1	
LINE2_PM1	.	Python molurus	CR1	
LINE2_VA1	.	Vipera ammodytes	CR1	
LINE2_VP1	.	Daboia palaestinae	CR1	
LINE2_WA1	.	Walterinnesia aegyptia	CR1	
LINE2A_CB	.	Caenorhabditis briggsae	CR1	
LINE2A_CE	.	Caenorhabditis elegans	CR1	
LINE2B_CB	.	Caenorhabditis briggsae	CR1	1,3,4,2

LINE2B_CE	.	Caenorhabditis elegans	CR1	
LINE2C_CB	.	Caenorhabditis briggsae	CR1	
LINE2C_CE	.	Caenorhabditis elegans	CR1	
LINE2C1_CE	U13643	Caenorhabditis elegans	CR1	16,1,3,11,4,2,15
LINE2D_CE	.	Caenorhabditis elegans	CR1	
LINE2E_CE	AF077538	Caenorhabditis elegans	CR1	
LINE2F_CE	.	Caenorhabditis elegans	CR1	
LINE2G_CE	.	Caenorhabditis elegans	CR1	1,3,4,2
LINE2H_CE	.	Caenorhabditis elegans	CR1	1,3,4,2
LINER1	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,15,16,3
LINER1-2_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,4,15,16,3
LINER1-3_NVi	.	Nasonia vitripennis	R1	16,1,3,11,4,86,2,15
LINEX-1_DR	.	Danio rerio	Rex1	
LINEX-2_DR	.	Danio rerio	Rex1	
LinSINE1	.	Lingula anatina	SINE2/tPHK	
LitSINE1	.	Littorina saxatilis	SINE2/tPHK	
Lm1	.	Locusta migratoria	SINE2/tPHK	
Lm1B	.	Locusta migratoria	SINE2/tPHK	
Lm2	.	Locusta migratoria	SINE2/tPHK	
Lm2B	.	Locusta migratoria	SINE2/tPHK	
Lm2B2	.	Locusta migratoria	SINE2/tPHK	
Lm2C	.	Locusta migratoria	SINE2/tPHK	
Lm2C1	.	Locusta migratoria	SINE2/tPHK	

Lm2C2	.	Locusta migratoria	SINE2/тРНК	
Lm2D	.	Locusta migratoria	SINE2/тРНК	
Lm2E	.	Locusta migratoria	SINE2/тРНК	
LmeSINE1a	.	Latimeria menadoensis	SINE3/5S	
LmeSINE1b	.	Metazoa Eukaryota.	SINE3/5S	
LmeSINE1c	.	Metazoa Eukaryota.	SINE3/5S	
LmeSINE1d	.	Latimeria chalumnae	SINE3/5S	
LmeSINE1e	.	Latimeria chalumnae	SINE3/5S	
LMR1_LM	.	Leptosphaeria maculans	Не содержащий LTR ретротранспозон	
LmSIDER2	.	Leishmania	Ingi	
LOA	X60177	Drosophila silvestris	Loa	199,200,201,2,8,202,11,12,203,1,204,14,89,4,90,15,16,17,3
LOA_Ele2	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LOA_Ele2B_AAe	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA_Ele3	.	Aedes aegypti	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA_Ele3B_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA_Ele3C_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA_Ele4	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA_Ele5	.	Aedes aegypti	Loa	6,2,9,10,11,86,13,1,4,15,16,17,3,7
LOA_Ele6	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA_Ele6B_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA_Ele6C_AAe	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA_Ele7	.	Aedes aegypti	Loa	6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
Loa-1_BTa	.	Bemisia tabaci	Loa	7,1,17,9,3,10,4,6,2,13

LOA-1_CQ	Culex quinquefasciatus	Loa	7,1,17,9,3,10,4,6,2,13
Loa-1_DAn	Drosophila ananassae	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-1_DBi	Drosophila biarmipes	Loa	
Loa-1_DBp	Drosophila bipectinata	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1_DEl	Drosophila elegans	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-1_DGri	Drosophila grimshawi	Loa	6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
Loa-1_DK	Drosophila kikkawai	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-1_DPer	Drosophila persimilis	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
Loa-1_DPs	Drosophila pseudoobscura	Loa	16,1,89,11,86,2,90,15
Loa-1_DRh	Drosophila rhopaloa	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-1_DSer	Drosophila serrata	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-1_DT	Drosophila takahashii	Loa	6,2,7,9,10,11,86,13,1,4,15,16,17,3
LOA-1_DYa	Drosophila yakuba	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LOA-1_FoCa	Folsomia candida	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1_LHu	Linepithema humile	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Loa-1_LMi	Locusta migratoria	Loa	6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
Loa-1_LSal	Lepeophtheirus salmonis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LOA-1_MLe	Mnemiopsis leidyi	Loa	7,1,17,3,10,4,6,2,13
Loa-1_MMa	Mesobuthus martensii	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-1_PPo	Papilio polytes	Loa	16,1,3,11,4,2,15
Loa-1_RPr	Rhodnius prolixus	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-1_SIn	Solenopsis invicta	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-10_AAe	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3

LOA-1B_DPer	.	Drosophila persimilis	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1B_DT	.	Drosophila takahashii	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-1B_FoCa	.	Folsomia candida	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1C_DT	.	Drosophila takahashii	Loa	
Loa-2_BTa	.	Bemisia tabaci	Loa	1,3,4,2
LOA-2_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,4,15,16,17,3,7
Loa-2_DEl	.	Drosophila elegans	Loa	
Loa-2_DF	.	Drosophila ficusphila	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DGri	.	Drosophila grimshawi	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-2_DK	.	Drosophila kikkawai	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DPer	.	Drosophila persimilis	Loa	6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
LOA-2_DSer	.	Drosophila serrata	Loa	6,2,9,10,11,86,13,1,15,16,17,3,7
Loa-2_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DYa	.	Drosophila yakuba	Loa	6,2,9,10,11,13,1,15,16,17,7
LOA-2_FoCa	.	Folsomia candida	Loa	
Loa-2_LSal	.	Lepeophtheirus salmonis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LOA-2_MLe	.	Mnemiopsis leidyi	Loa	16,1,3,11,4,2,15
Loa-2_MMa	.	Mesobuthus martensii	Loa	16,1,3,11,4,2,15
Loa-2_RPr	.	Rhodnius prolixus	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-2_SIn	.	Solenopsis invicta	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-3_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,4,15,16,17,3,7
Loa-3_DEl	.	Drosophila elegans	Loa	
LOA-3_DPer	.	Drosophila persimilis	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7

Loa-3_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-3_LSal	.	Lepeophtheirus salmonis	Loa	
Loa-3_RPr	.	Rhodnius prolixus	Loa	16,1,3,89,11,15,4,86,2,90
LOA-4_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Loa-4_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-4_LSal	.	Lepeophtheirus salmonis	Loa	7,1,17,9,3,10,11,4,2,6,13
LOA-5_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-5_LSal	.	Lepeophtheirus salmonis	Loa	7,16,1,17,3,11,4,2,15
LOA-6_CQ	.	Culex quinquefasciatus	Loa	7,1,17,9,3,10,4,6,2,13
Loa-6_LSal	.	Lepeophtheirus salmonis	Loa	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
LOA-7_CQ	.	Culex quinquefasciatus	Loa	7,1,17,9,3,10,4,6,2,13
LOA-8_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-9_AAe	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loner_Ele1	.	Aedes aegypti	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
Loner_Ele2	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
Loner_Ele3	.	Aedes aegypti	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
Loner_Ele4	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
Lun1	.	Lepidosiren paradoxa	SINE2/тРНК	
LX	.	Muridae	L1	
Lx_3-конец	.	Muridae	L1	
LX10	.	Mus musculus	L1	
LX2	.	Sciuromorpha	L1	
Lx2_3-конец	.	Muridae	L1	

Lx2A_3-конец	.	Muridae	L1	
Lx2A1_3-конец	.	Muridae	L1	
LX2B	.	Mus musculus	L1	
Lx2B_3-конец	.	Muridae	L1	
Lx2B2_3-конец	.	Muridae	L1	
LX3	.	Sciuromorpha	L1	
Lx3B_3-конец	.	Muridae	L1	
Lx3C_3-конец	.	Muridae	L1	
LX4	.	Mus musculus	L1	
LX5	.	Mus musculus	L1	
LX5b	.	Mus musculus	L1	
LX5c	.	Mus musculus	L1	
LX6	.	Mus musculus	L1	
LX7	.	Mus musculus	L1	
Lx7_3-конец	.	Rodentia	L1	
LX8	.	Mus musculus	L1	
LX8B	.	Mus musculus	L1	
LX9	.	Mus musculus	L1	
M_Av	.	Adineta vaga	Penelope	1,3,2
M1_As	.	Adineta	Penelope	1,3
M2_As	.	Adineta	Penelope	1,3
M3_As	.	Adineta	Penelope	1,3
MamSINE1	.	Mammalia	SINE2/тРНК	

MamSINE1	.	Mammalia	SINE2/тРНК
MamSINE1	.	Mammalia	SINE2/тРНК
ManaSINE1	.	Manacus vitellinus	SINE2/тРНК
ManaSINE1_ZCh	.	Zimmerius chrysops	SINE2/тРНК
ManaSINE2	.	Manacus vitellinus	SINE
MAR1	.	Metatheria	SINE2/тРНК
MAR1_MD	.	Monodelphis domestica	SINE2/тРНК
MAR1c_Mdo	.	Monodelphis domestica	SINE2/тРНК
MAR4_MD	.	Monodelphis domestica	RTE
MARE3	.	Mammalia	SINE
MARE3	.	Mammalia	SINE
MARE3	.	Mammalia	SINE
MARE6	.	Mammalia	Tx1
MARE6	.	Mammalia	Tx1
MARE6	.	Mammalia	Tx1
MAUI	.	Takifugu rubripes	CR1 16,1,3,24,4,11,48,2
MdoRep1	.	Metatheria	SINE
Meg_PVa	.	Pteropus vampyrus	SINE3/5S
MEG-RL	EU853278.1	Rousettus aegyptiacus	SINE3/5S
MEG-RS	EU853265.1	Rousettus aegyptiacus	SINE3/5S
MEG-T2	.	Megachiroptera	SINE2/тРНК
MEG-TR	.	Megachiroptera	SINE2/тРНК
MeloSINE	.	Melopsittacus undulatus	SINE

MEN	.	Menetes berdmorei	SINE2/тPHK	
MER129	.	Euteleostomi	SINE	
MER131	.	Euteleostomi	SINE	
MER25	.	Eutheria	L1	
Metulj-0_Hmel	.	Heliconius melpomene	SINE	
Metulj-1_Hmel	.	Heliconius melpomene	SINE	
Metulj-2_Hmel	.	Heliconius melpomene	SINE	
Metulj-3_Hmel	.	Heliconius melpomene	SINE	
Metulj-4_Hmel	.	Heliconius melpomene	SINE	
Metulj-5_Hmel	.	Heliconius melpomene	SINE	
Metulj-6_Hmel	.	Heliconius melpomene	SINE	
Metulj-7_Hmel	.	Heliconius melpomene	SINE	
MGR583	AF018033	Magnaporthe grisea	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
MinoAg1	AB090816	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
MIR	.	Mammalia	SINE2/тPHK	
MIR	.	Mammalia	SINE2/тPHK	
MIR	.	Mammalia	SINE2/тPHK	
MIR_Aves1	.	Aves	SINE2/тPHK	
MIR_Aves2	.	Aves	SINE2/тPHK	
MIR_Xt	.	Xenopus tropicalis	SINE2/тPHK	
MIR1_AMi	.	Alligator mississippiensis	SINE2/тPHK	
MIR1_Crp	.	Crocodylus porosus	SINE2/тPHK	
MIR3	.	Mammalia	SINE2/тPHK	

MIR3 . Mammalia SINE2/тРНК

MIR3 . Mammalia SINE2/тРНК

MIR3_MarsA . Metatheria SINE2/тРНК

MIR3_MarsB . Metatheria SINE2/тРНК

MIRb . Mammalia SINE2/тРНК

MIRb . Mammalia SINE2/тРНК

MIRb . Mammalia SINE2/тРНК

MIRc . Mammalia SINE2/тРНК

MIRc . Mammalia SINE2/тРНК

MIRc . Mammalia SINE2/тРНК

MIUSE1_TM AF459639 Triticum monococcum L1

MON1 . Monotremata SINE2/тРНК

MonoRep87A . Ornithorhynchus SINE2/тРНК

MonoRep87B . Ornithorhynchus SINE2/тРНК

MosquI_Aa2 AF134900 Aedes aegypti I

MoTeR1 JQ747487 Magnaporthe oryzae CRE 1,3,2

MoTeR2 JQ747488 Magnaporthe oryzae CRE

MRE1_OL . Oryzias latipes SINE

MRT . Monascus pilosus Tad1 6,2,9,10,11,13,1,4,15,16,17,3,239,7

MSTA1 . Primates ERV3

MusHAL1 . Mus L1

MusHAL1_5-конец . Mus L1

MUSID2 . Rodentia SINE2/тРНК

MUSID3	.	Rodentia	SINE2/тРНК	
MUSID4	.	Rodentia	SINE2/тРНК	
MUSID5	.	Rodentia	SINE2/тРНК	
Mutsu-1_SK	.	Saccoglossus kowalevskii	Tx1	43,2,8,44,11,12,42,1,26,14,4,3
Mutsu-2_SK	.	Saccoglossus kowalevskii	Tx1	2,8,11,12,42,1,26,14,4,3
MVB2	X52381	Neovison vison	SINE2/тРНК	
MyrSINE	.	Myrmecophagidae	SINE2/тРНК	
N_As	.	Adineta	Penelope	1,3
N0_As	.	Adineta	Penelope	
NAVIRTE1	.	Nasonia vitripennis	RTE	1,3,2
Nematis_C4	.	Caenorhabditis brenneri	Penelope	1,3,103,2
Nematis_Cr	.	Caenorhabditis remanei	Penelope	1,3,103,72,2
Nematis_Pp	.	Pristionchus pacificus	Penelope	1,3,103
Neptune_Hyd	.	Hydra vulgaris	Penelope	1,3,2
Neptune_Sp	.	Sphenodon punctatus	Penelope	
Neptune1_Ac	.	Anolis carolinensis	Penelope	73,1,3,103,72,2
Neptune1_Ap	.	Acropora palmata	Penelope	73,1,3,71,72,74
Neptune1_Nv	.	Nematostella vectensis	Penelope	1,3,70,71,72
Neptune1_Ren	.	Amphimedon queenslandica	Penelope	73,1,3,70,71,72,62,74
Neptune2_Ap	.	Acropora palmata	Penelope	
Neptune2_Ren	.	Amphimedon queenslandica	Penelope	205,73,74,1,3,70,206,71,72,2
Neptune3_Ren	.	Amphimedon queenslandica	Penelope	74,73,1,3,70,71,72,2
NeSL-1	Z82058	Caenorhabditis elegans	NeSL	147,1,148,3,149,120,4,2

NeSL-1_C11	.	Caenorhabditis tropicalis	NeSL	147,1,3,149,120,4,2
NeSL-1_CA	.	Caenorhabditis angaria	NeSL	147,1,148,3,149,120,4,2
NeSL-1_CBre	.	Caenorhabditis brenneri	NeSL	147,1,3,149,120,4,2
NeSL-1_CBri	.	Caenorhabditis briggsae	NeSL	1,2
NeSL-1_CJap	.	Caenorhabditis japonica	NeSL	147,1,148,3,149,120,4,2
NeSL-1_CRem	.	Caenorhabditis remanei	NeSL	147,1,3,149,120,4,2
NeSL-1_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2
NeSL-1_TV	.	Trichomonas vaginalis	NeSL	18,5,1,19,3,4,2
NeSL-2_CBre	.	Caenorhabditis brenneri	NeSL	147,1,148,3,149,120,4,2
NeSL-2_CRem	.	Caenorhabditis remanei	NeSL	147,5,19,1,148,3,120,149,4,2
NeSL-2_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2
NeSL-3_CBre	.	Caenorhabditis brenneri	NeSL	147,1,148,3,149,120,4,2
NeSL-3_CRem	.	Caenorhabditis remanei	NeSL	147,1,148,3,120,149,4,2
NeSL-4_CRem	chrUn	Caenorhabditis remanei	NeSL	147,1,148,3,149,120,4,2
NeSL-4_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
Ngaro1_DR	.	Danio rerio	Не содержащий LTR ретротранспозон	220,221,222,2,223,224,1,225,226,17,3
Nimb-1_AAe	.	Aedes aegypti	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_BTa	.	Bemisia tabaci	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-1_CPB	.	Chrysemys picta bellii	Nimb	
Nimb-1_CQ	.	Culex quinquefasciatus	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-1_CTe	.	Capitella teleta	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-1_DPu	.	Daphnia pulex	Nimb	6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,60,7
Nimb-1_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7

Nimb-1_EL	Esox lucius	Nimb	7,1,17,9,3,10,4,11,6,2,13
Nimb-1_GA	Gasterosteus aculeatus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_HRo	Helobdella robusta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_LG	Lottia gigantea	Nimb	6,2,9,10,11,1,4,15,16,17,3,7
Nimb-1_LMi	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_LVa	Litopenaeus vannamei	Nimb	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-1_MMa	Mesobuthus martensii	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-1_PH	Parhyale hawaiensis	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-1_RPr	Rhodnius prolixus	Nimb	12,8
Nimb-1_SSa	Salmo salar	Nimb	6,2,7,9,10,11,13,1,4,15,16,3
Nimb-10_DR	Danio rerio	Nimb	
Nimb-10_LMi	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-11_DR	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-11_LMi	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Nimb-12_DR	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-12_LMi	Locusta migratoria	Nimb	6,2,9,10,11,86,13,1,14,4,15,16,17,3,7
Nimb-13_DR	Danio rerio	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-13_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-14_DR	Danio rerio	Nimb	
Nimb-14_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,86,12,13,1,14,4,15,16,17,3
Nimb-15_DR	Danio rerio	Nimb	2,8,11,12,1,14,4,15,16,17,3,7
Nimb-15_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-16_DR	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7,227

Nimb-16_LMi	Locusta migratoria	Nimb	6,2,9,8,10,12,13,1,14,4,17,3,7
Nimb-17_DRe	Danio rerio	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-17_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-18_LMi	Locusta migratoria	Nimb	6,2,7,9,8,10,11,86,12,13,1,14,4,15,16,17
Nimb-19_LMi	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
Nimb-19B_LMi	Locusta migratoria	Nimb	
Nimb-1N1_LVa	Litopenaeus vannamei	Nimb	
Nimb-2_AAe	Aedes aegypti	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_BTa	Bemisia tabaci	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_CPB	Chrysemys picta bellii	Nimb	
Nimb-2_CQ	Culex quinquefasciatus	Nimb	6,2,8,9,10,11,12,1,4,15,16,17,3,7
Nimb-2_CTe	Capitella teleta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_DR	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_LG	Lottia gigantea	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Nimb-2_LMi	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_LVa	Litopenaeus vannamei	Nimb	6,2,9,10,11,13,214,1,4,15,16,17,3,7
Nimb-2_PH	Parhyale hawaiensis	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_RPr	Rhodnius prolixus	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-2_SSa	Salmo salar	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-20_LMi	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
Nimb-21_LMi	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-22_LMi	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,1,14,4,15,16,17,3
Nimb-23_LMi	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3

Nimb-24_LMi	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-25_LMi	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-26_LMi	Locusta migratoria	Nimb	16,1,14,3,11,4,2,15
Nimb-27_LMi	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,3,7
Nimb-28_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-29_LMi	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-2B_LMi	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2B_SSa	Salmo salar	Nimb	8,14,12
Nimb-2C_LMi	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_BTa	Bemisia tabaci	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_CTe	Capitella teleta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_DR	Danio rerio	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-3_LG	Lottia gigantea	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_LVa	Litopenaeus vannamei	Nimb	6,2,10,11,12,1,14,4,17,3,7
Nimb-3_PH	Parhyale hawaiensis	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-3_RPr	Rhodnius prolixus	Nimb	
Nimb-3_SSa	Salmo salar	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-30_LMi	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-31_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17
Nimb-32_LMi	Locusta migratoria	Nimb	16,1,17,9,10,11,4,2,6,7,13,15
Nimb-33_LMi	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-34_LMi	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-35_LMi	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3

Nimb-36_LMi .	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-37_LMi .	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-38_LMi .	Locusta migratoria	Nimb	7,16,1,17,3,11,4,2,15
Nimb-39_LMi .	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-4_BTa .	Bemisia tabaci	Nimb	17,9,10,6,7,13
Nimb-4_DR .	Danio rerio	Nimb	6,2,21,9,23,10,11,1,26,4,30,17,3,7
Nimb-4_LMi .	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-4_LVa .	Litopenaeus vannamei	Nimb	6,2,9,10,11,12,13,1,14,4,15,16,17,3,7
Nimb-4_RPr .	Rhodnius prolixus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-4_SSa .	Salmo salar	Nimb	6,2,8,9,10,11,12,13,1,4,15,16,17,3,7
Nimb-40_LMi .	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-40N1_LMi .	Locusta migratoria	Nimb	
Nimb-41_LMi .	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-42_LMi .	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,17,3
Nimb-4B_RPr .	Rhodnius prolixus	Nimb	16,1,8,3,11,15,4,12,2
Nimb-5_DR .	Danio rerio	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-5_LMi .	Locusta migratoria	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-6_DR .	Danio rerio	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-6_LMi .	Locusta migratoria	Nimb	6,2,9,10,11,86,13,1,14,4,15,16,17,3,7
Nimb-7_DR .	Danio rerio	Nimb	17,9,10,6,7,13
Nimb-7_LMi .	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-8_DR .	Danio rerio	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-8_LMi .	Locusta migratoria	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7

Nimb-9_DR .	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-9_LMi .	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-N1_SSa.	Salmo salar	Nimb	
Nimb-N2_LVa .	Litopenaeus vannamei	Nimb	
Nimb-N2_SSa.	Salmo salar	Nimb	
Nimb-N2B_LVa .	Litopenaeus vannamei	Nimb	
nimbus EF413180	Biomphalaria glabrata	Nimb	6,2,7,9,10,11,86,12,13,1,14,4,15,16,17,3
NLA M57436	Capra hircus	SINE2/тРНК	
Не содержащий LTR-1_ACar .	Anolis carolinensis	RTEX	
Не содержащий LTR-1_BG .	Blumeria graminis	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_CM.	Callorhinchus milii	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_CPB .	Chrysemys picta bellii	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_CR .	Chlamydomonas reinhardtii	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_DR .	Danio rerio	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_LSal .	Lepeophtheirus salmonis	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_LVa .	Litopenaeus vannamei	Не содержащий LTR ретротранспозон 6,2,9,10,11,13,1,4,15,16,17,	
Не содержащий LTR-1_SBi .	Sorghum bicolor	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1_STu .	Solanum tuberosum	Не содержащий LTR ретротранспозон	
Не содержащий LTR-10_LVa .	Litopenaeus vannamei	Не содержащий LTR ретротранспозон	
Не содержащий LTR-1B_DR .	Danio rerio	Не содержащий LTR ретротранспозон	
Не содержащий LTR-2_BG .	Blumeria graminis	Не содержащий LTR ретротранспозон	
Не содержащий LTR-2_CM.	Callorhinchus milii	Не содержащий LTR ретротранспозон	
Не содержащий LTR-2_CR .	Chlamydomonas reinhardtii	Не содержащий LTR ретротранспозон	

Не содержащий LTR-2_LSal . Lepeophtheirus salmonis Не содержащий LTR ретротранспозон
Не содержащий LTR-2_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 2,16,1,14,3,4,11,12,15
Не содержащий LTR-2_PPo . Papilio polytes Не содержащий LTR ретротранспозон
Не содержащий LTR-2_STu . Solanum tuberosum Не содержащий LTR ретротранспозон
Не содержащий LTR-21_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон
Не содержащий LTR-22_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон
Не содержащий LTR-24_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон
Не содержащий LTR-2B_LSal . Lepeophtheirus salmonis Не содержащий LTR ретротранспозон
Не содержащий LTR-3_BG . Blumeria graminis Не содержащий LTR ретротранспозон
Не содержащий LTR-3_CM . Callorhinchus milii Не содержащий LTR ретротранспозон 1,3,11,4,2
Не содержащий LTR-3_CR . Chlamydomonas reinhardtii Не содержащий LTR ретротранспозон
Не содержащий LTR-3_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 1,3,11,4,2
Не содержащий LTR-3_PPo . Papilio polytes Не содержащий LTR ретротранспозон
Не содержащий LTR-4_BG . Blumeria graminis Tad1
Не содержащий LTR-4_CR . Chlamydomonas reinhardtii Не содержащий LTR ретротранспозон
Не содержащий LTR-4_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
Не содержащий LTR-4_PPo . Papilio polytes Не содержащий LTR ретротранспозон
Не содержащий LTR-4B_BG . Blumeria graminis Tad1
Не содержащий LTR-5_BG . Blumeria graminis Не содержащий LTR ретротранспозон
Не содержащий LTR-5_CR . Chlamydomonas reinhardtii Не содержащий LTR ретротранспозон
Не содержащий LTR-5_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон
43,6,2,7,9,10,44,11,13,1,14,4,15,16,17,3
Не содержащий LTR-6_BG . Blumeria graminis Не содержащий LTR ретротранспозон

Не содержащий LTR-6_CR . Chlamydomonas reinhardtii Не содержащий LTR ретротранспозон

Не содержащий LTR-6_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 16,1,3,24,11,4,2

Не содержащий LTR-6A_CR . Chlamydomonas reinhardtii Не содержащий LTR ретротранспозон

Не содержащий LTR-7_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 7,1,17,9,10,11,6,2,13

Не содержащий LTR-9_LVa . Litopenaeus vannamei Не содержащий LTR ретротранспозон 1,3,11,4,2

Не содержащий LTR-N1_PH . Parhyale hawaiensis Не содержащий LTR ретротранспозон

Не содержащий LTR-N1B_PH . Parhyale hawaiensis Не содержащий LTR ретротранспозон

Не содержащий LTR-N2_PH . Parhyale hawaiensis Не содержащий LTR ретротранспозон

Не содержащий LTR-N3_PH . Parhyale hawaiensis Не содержащий LTR ретротранспозон

Не содержащий LTR-N4_PH . Parhyale hawaiensis Не содержащий LTR ретротранспозон

NotoAg1 AB090823 Anopheles gambiae R1 2,8,11,86,12,42,1,14,4,15,16,3

Nova-1_PH . Parhyale hawaiensis RTE 16,1,3,11,4,48,62,2,15

Nve-Nin-DC-SINE1 . Nematostella vectensis SINE2/тРНК

Nve-Nin-DC-SINE2 . Nematostella vectensis SINE2/тРНК

Nve-Nin-DC-SINE3 . Nematostella vectensis SINE2/тРНК

Nve-Nin-DC-SINE4 . Nematostella vectensis SINE2/тРНК

Nve-Nin-DC-SINE5 . Nematostella vectensis SINE2/тРНК

Nve-Nin-DC-SINE6 . Nematostella vectensis SINE2/тРНК

O_As . Adineta Penelope 1,3

O_Av . Adineta vaga Penelope 1,3

Odin-1 . Oikopleura dioica Odin 11

Odin-10_OD CABV01003716 Oikopleura dioica Odin 1,3,4,11,2

Odin-11_OD CABV01003634 Oikopleura dioica Odin 1,3,11,2

Odin-33_OD	CABV01004174	Oikopleura dioica	Odin	1,3,11,2
Odin-34_OD	CABW01000201	Oikopleura dioica	Odin	1,3,11,2
Odin-35_OD	CABV01005185	Oikopleura dioica	Odin	1,3,11,2
Odin-36_OD	CABV01000234	Oikopleura dioica	Odin	1,3,11,2
Odin-37_OD	CABV01003479	Oikopleura dioica	Odin	1,3,11,2
Odin-38_OD	CABW01000121	Oikopleura dioica	Odin	1,3,11,4,2
Odin-6_OD	CABV01000386	Oikopleura dioica	Odin	1,3,11,2
Odin-7_OD	CABW01000816	Oikopleura dioica	Odin	1,3,11,2
Odin-8_OD	CABV01004396	Oikopleura dioica	Odin	1,3,11,4,2
Odin-9_OD	CABV01000574	Oikopleura dioica	Odin	1,3,11,4,2
OegopSINE	.	Idioteuthis cordiformis	SINE2/тРНК	
OK	D32083	Octopus vulgaris	SINE2/тРНК	
OM2	X59354	Oncorhynchus masou	SINE2/тРНК	
OR1	D32089	Octopus vulgaris	SINE2/тРНК	
OR2	D32092	Octopus vulgaris	SINE2/тРНК	
ORTE-1_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,4,3,177
ORTE-2_AAe	.	Aedes aegypti	RTE	1,175,176,3,120,4,11,177,2
ORTE-3_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,4,3,177
ORTE-4_AAe	.	Aedes aegypti	RTE	175,1,176,3,120,11,4,177,2
ORTE-5_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,26,4,15,16,3,177
ORTE-6_AAe	.	Aedes aegypti	RTE	120,2,175,176,24,11,1,26,4,16,3,177
OrySINE1	.	Oryzias latipes	SINE2/тРНК	
OSLINE1_1	AE017112	Oryza sativa	L1	43,82,79,2,10,44,11,80,1,4,17,3,7

OSLINE1_2	AP008218	Oryza sativa	L1	43,78,79,2,7,10,44,11,80,1,81,77,4,75,17,3,76
OSLINE1-2	AC137696	Oryza sativa	L1	1,43,3,44,11,4,2
OSLINE1-3	AC136524	Oryza sativa	L1	43,2,8,44,11,12,1,14,4,75,3,76
OSLINE1-4	AC104712	Oryza sativa	L1	16,1,3,24,11,4,2
OSLINE1-5	.	Oryza sativa	L1	1,43,3,44,11,4,2
OSSINE1	.	Salmo salar	SINE3/5S	
OsSN3	.	Oryza sativa	SINE2/тPHK	
Outcast_Ele14	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast_Ele3	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-1_BF	.	Branchiostoma floridae	Outcast	53,6,2,54,9,8,10,11,57,12,216,13,68,1,14,4,15,16,17,69,3,59,7
Outcast-1_DAn	.	Drosophila ananassae	Outcast	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Outcast-1_DBi	.	Drosophila biarmipes	Outcast	7,1,17,9,3,10,4,6,2,13
Outcast-1_DBp	.	Drosophila bipectinata	Outcast	16,1,3,11,2,15
Outcast-1_DEl	.	Drosophila elegans	Outcast	7,16,1,17,3,10,11,4,6,2,15
Outcast-1_DEu	.	Drosophila eugracilis	Outcast	7,1,17,9,3,10,4,6,2,13
Outcast-1_DT	.	Drosophila takahashii	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-1_DWi	.	Drosophila willistoni	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-16_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-17_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-18_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-19_AAe	.	Aedes aegypti	Outcast	6,2,8,9,10,11,12,1,4,15,16,17,3,7
Outcast-2_BF	.	Branchiostoma floridae	Outcast	53,6,2,7,54,9,8,10,11,56,57,12,13,68,1,14,58,4,15,16,17,69,3,59
Outcast-2_DAn	.	Drosophila ananassae	Outcast	17,9,10,6,7,13

Outcast-2_DBp	.	Drosophila bipectinata	Outcast	
Outcast-2_DTa	.	Drosophila takahashii	Outcast	6,2,8,10,11,12,1,4,15,16,17,3,7
Outcast-20_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
P_As	.	Adineta	Penelope	1,3
P_Av	.	Adineta vaga	Penelope	1,3
p-SINE1_OS	.	Oryza sativa	SINE2/tPHK	
Pami	.	Patiria miniata	SINE2/tPHK	
PB1	.	Rodentia	SINE1/7SL	
PB1D10	.	Rodentia	SINE1/7SL	
PB1D11	.	Euarchontoglires	SINE1/7SL	
PB1D7	.	Rodentia	SINE1/7SL	
PB1D9	.	Rodentia	SINE1/7SL	
Peci	.	Petrolisthes cinctipes	SINE2/tPHK	
Ped-1	.	Pedetes capensis	SINE3/5S	
Ped-2	.	Pedetes capensis	SINE2/tPHK	
PENEL1_NVi	.	Nasonia vitripennis	Penelope	1,3,72
PENELA_Smed	.	Schmidtea mediterranea	Penelope	73,1,3,103,2
PENELOPE	.	virilis group	Penelope	74,1,3,71,103,72,2
Penelope_Ele2	.	Aedes aegypti	Penelope	73,74,1,3,71,103,72,2
PENELOPE_SM	BK000685	Schistosoma mansoni	Penelope	1,3,103,72,2
Penelope-1_AAe	.	Aedes aegypti	Penelope	1,3,103
Penelope-1_ACar	.	Anolis carolinensis	Penelope	74,73,1,3,71,103,72,2
Penelope-1_ADi	.	Acropora digitifera	Penelope	74,73,1,3,70,71,72,2

Penelope-1_AEc	.	Acromyrmex echinatior	Penelope	1,3,103,72,2
Penelope-1_AFC	.	Cichlidae	Penelope	1,103,72,2
Penelope-1_AMi	.	Alligator mississippiensis	Penelope	1,3,2
Penelope-1_BF	.	Branchiostoma floridae	Penelope	1,3,103,72
Penelope-1_BTa	.	Bemisia tabaci	Penelope	1,3
Penelope-1_CGi	.	Crassostrea gigas	Penelope	73,74,1,3,70,71,72,2
Penelope-1_CM	.	Callorhinchus milii	Penelope	73,71,72,74
Penelope-1_CoFl	.	Copidosoma floridanum	Penelope	1,3,72,2
Penelope-1_CPB	.	Chrysemys picta bellii	Penelope	1,3,72,2
Penelope-1_CQ	.	Culex quinquefasciatus	Penelope	
Penelope-1_Crp	.	Crocodylidae	Penelope	1,3,72,2
Penelope-1_CTe	.	Capitella teleta	Penelope	1,3,2
Penelope-1_DEl	.	Drosophila elegans	Penelope	74,1,3,70,71,103,72,2
Penelope-1_DF	.	Drosophila ficusphila	Penelope	1,3,71,103,72,74
Penelope-1_DK	.	Drosophila kikkawai	Penelope	
Penelope-1_DR	.	Danio rerio	Penelope	1,3,72
Penelope-1_DRh	.	Drosophila rhopaloa	Penelope	74,1,3,71,103,72,2
Penelope-1_FoCa	.	Folsomia candida	Penelope	73,71,72,74
Penelope-1_GA	.	Gasterosteus aculeatus	Penelope	73,1,3,70,71,72,62,74
Penelope-1_Gav	.	Gavialis gangeticus	Penelope	72
Penelope-1_HM	.	Hydra vulgaris	Penelope	
Penelope-1_HMM	.	Heliconius melpomene melpomene	Penelope	
Penelope-1_HRo	.	Helobdella robusta	Penelope	

Penelope-1_LCh	.	Latimeria chalumnae	Penelope	
Penelope-1_LG	.	Lottia gigantea	Penelope	73,1,3,70,71,72,74
Penelope-1_LHu	.	Linepithema humile	Penelope	74,73,1,3,103,71,72,2
Penelope-1_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-1_LSal	.	Lepeophtheirus salmonis	Penelope	73,72
Penelope-1_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,2
Penelope-1_MLe	.	Mnemiopsis leidyi	Penelope	
Penelope-1_NV	.	Nematostella vectensis	Penelope	
Penelope-1_NVi	.	Nasonia vitripennis	Penelope	74,73,1,3,70,71,72,2
Penelope-1_OD	.	Oikopleura dioica	Penelope	1,3,71,2,74
Penelope-1_OL	.	Oryzias latipes	Penelope	18,74,5,1,71,103,72,2
Penelope-1_PBa	.	Pogonomyrmex barbatus	Penelope	73,1,3,71,103,72,74
Penelope-1_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-1_PSi	.	Pelodiscus sinensis	Penelope	74,1,3,71,72,2
Penelope-1_SIn	.	Solenopsis invicta	Penelope	74,73,1,3,71,103,72,2
Penelope-1_SK	.	Saccoglossus kowalevskii	Penelope	
Penelope-1_SM	.	Schmidtea mediterranea	Penelope	73,74,1,3,70,71,2
Penelope-1_SP	.	Strongylocentrotus purpuratus	Penelope	
Penelope-1_SSa	.	Salmo salar	Penelope	
Penelope-1_TCas	.	Tribolium castaneum	Penelope	73,74,1,3,70,71,103,72,2
Penelope-1_XL	.	Xenopus laevis	Penelope	72
Penelope-10_ACar	.	Anolis carolinensis	Penelope	74,1,3,71,72,2
Penelope-10_CoFl	.	Copidosoma floridanum	Penelope	71,103,72,74

Penelope-10_FoCa	.	Folsomia candida	Penelope	1,3,70,71,103,72,74
Penelope-10_HM	.	Hydra vulgaris	Penelope	3,1
Penelope-10_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-10_LVa	.	Litopenaeus vannamei	Penelope	1,3,71,74
Penelope-10_NV	.	Nematostella vectensis	Penelope	
Penelope-10_PH	.	Parhyale hawaiensis	Penelope	1,3
Penelope-10_XL	.	Xenopus laevis	Penelope	74,70,71,72
Penelope-10_XT	.	Xenopus tropicalis	Penelope	73,71,74
Penelope-100_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-101_LMi	.	Locusta migratoria	Penelope	
Penelope-102_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-103_LMi	.	Locusta migratoria	Penelope	1,3,103,72
Penelope-104_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-105_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-106_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-107_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-108_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-109_LMi	.	Locusta migratoria	Penelope	
Penelope-11_ACar	.	Anolis carolinensis	Penelope	73,1,3,71,72,74
Penelope-11_CoFl	.	Copidosoma floridanum	Penelope	1,3,2
Penelope-11_FoCa	.	Folsomia candida	Penelope	73,1,3,103,72
Penelope-11_HM	.	Hydra vulgaris	Penelope	1,3,2
Penelope-11_LMi	.	Locusta migratoria	Penelope	1,3,103,72

Penelope-11_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,103,72,2
Penelope-11_NV	.	Nematostella vectensis	Penelope	
Penelope-11_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-11_XL	.	Xenopus laevis	Penelope	
Penelope-11_XT	.	Xenopus tropicalis	Penelope	1,3,72,2
Penelope-110_LMi	.	Locusta migratoria	Penelope	
Penelope-111_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-112_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,2
Penelope-113_LMi	.	Locusta migratoria	Penelope	
Penelope-114_LMi	.	Locusta migratoria	Penelope	
Penelope-115_LMi	.	Locusta migratoria	Penelope	
Penelope-116_LMi	.	Locusta migratoria	Penelope	
Penelope-117_LMi	.	Locusta migratoria	Penelope	
Penelope-118_LMi	.	Locusta migratoria	Penelope	
Penelope-119_LMi	.	Locusta migratoria	Penelope	3,2,1
Penelope-12_HM	.	Hydra vulgaris	Penelope	3,1
Penelope-12_LMi	.	Locusta migratoria	Penelope	1,72,2
Penelope-12_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,103,72,2
Penelope-12_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-12_XL	.	Xenopus laevis	Penelope	
Penelope-12_XT	.	Xenopus tropicalis	Penelope	3,1
Penelope-120_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-120N1_LMi	.	Locusta migratoria	Penelope	

Penelope-121_LMi	Locusta migratoria	Penelope	1,3,70,71,74
Penelope-12B_LMi	Locusta migratoria	Penelope	
Penelope-13_HM	Hydra vulgaris	Penelope	
Penelope-13_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-13_LVa	Litopenaeus vannamei	Penelope	73,1,3,71,74
Penelope-13_XL	Xenopus laevis	Penelope	1,3,72,2
Penelope-13_XT	Xenopus tropicalis	Penelope	1,3,72
Penelope-14_HM	Hydra vulgaris	Penelope	
Penelope-14_LMi	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-14_XL	Xenopus laevis	Penelope	
Penelope-15_HM	Hydra vulgaris	Penelope	
Penelope-15_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-15_LVa	Litopenaeus vannamei	Penelope	1,3,71,74
Penelope-15_XL	Xenopus laevis	Penelope	73,70,71,72,74
Penelope-15N1_LMi	Locusta migratoria	Penelope	
Penelope-15N2_LMi	Locusta migratoria	Penelope	
Penelope-16_HM	Hydra vulgaris	Penelope	
Penelope-16_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-16_XL	Xenopus laevis	Penelope	74,1,3,71,2
Penelope-17_HM	Hydra vulgaris	Penelope	
Penelope-17_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-17_XL	Xenopus laevis	Penelope	1,3,72
Penelope-18_HMa	Hydra vulgaris	Penelope	

Penelope-18_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-19_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-1A_NV	.	Nematostella vectensis	Penelope	
Penelope-1B_CM	.	Callorhinchus milii	Penelope	
Penelope-1B_CPB	.	Chrysemys picta bellii	Penelope	1,3,103,72,2
Penelope-1B_XL	.	Xenopus laevis	Penelope	
Penelope-1C_CPB	.	Chrysemys picta bellii	Penelope	1,3,71,72,2,74
Penelope-1N1_OD	.	Oikopleura dioica	Penelope	
Penelope-1N1_PSi	.	Pelodiscus sinensis	Penelope	
Penelope-1N2_CPB	.	Chrysemys picta bellii	Penelope	
Penelope-2_AAe	.	Aedes aegypti	Penelope	1,3,103,72,2
Penelope-2_ACar	.	Anolis carolinensis	Penelope	1,3,71,72,74
Penelope-2_ADi	.	Acropora digitifera	Penelope	73,1,3,70,71,72,74
Penelope-2_AEc	.	Acromyrmex echinatior	Penelope	73,74,1,3,71,72,2
Penelope-2_BF	.	Branchiostoma floridae	Penelope	1,3
Penelope-2_BTa	.	Bemisia tabaci	Penelope	73,1,3,72
Penelope-2_CGi	.	Crassostrea gigas	Penelope	3,1
Penelope-2_CoFl	.	Copidosoma floridanum	Penelope	1,3,72
Penelope-2_CPB	.	Chrysemys picta bellii	Penelope	1,3,70,71,2
Penelope-2_CQ	.	Culex quinquefasciatus	Penelope	
Penelope-2_Crp	.	Crocodylus porosus	Penelope	
Penelope-2_DR	.	Danio rerio	Penelope	
Penelope-2_FoCa	.	Folsomia candida	Penelope	1,3,71,103,72,74

Penelope-2_GA	.	Gasterosteus aculeatus	Penelope	73,1,3,70,71,48,62,74
Penelope-2_HM	.	Hydra vulgaris	Penelope	
Penelope-2_HMM	.	Heliconius melpomene melpomene	Penelope	
Penelope-2_HRo	.	Helobdella robusta	Penelope	103,3,1
Penelope-2_Lch	.	Latimeria chalumnae	Penelope	
Penelope-2_LG	.	Lottia gigantea	Penelope	1,3,103
Penelope-2_LHu	.	Linepithema humile	Penelope	74,73,1,3,70,71,72,2
Penelope-2_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-2_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,72,2
Penelope-2_MLe	.	Mnemiopsis leidyi	Penelope	1,3,2
Penelope-2_NV	.	Nematostella vectensis	Penelope	
Penelope-2_OD	.	Oikopleura dioica	Penelope	73,70,71
Penelope-2_PBa	.	Pogonomyrmex barbatus	Penelope	
Penelope-2_PH	.	Parhyale hawaiensis	Penelope	1,3,71,103,74
Penelope-2_PSi	.	Pelodiscus sinensis	Penelope	1,3,2
Penelope-2_SIn	.	Solenopsis invicta	Penelope	73,74,1,3,71,103,72,2
Penelope-2_SK	.	Saccoglossus kowalevskii	Penelope	
Penelope-2_SSa	.	Salmo salar	Penelope	1,3,70,71,72,74
Penelope-2_XL	.	Xenopus laevis	Penelope	1,3,72
Penelope-20_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-21_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-22_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-23_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2

Penelope-24_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-26_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,72,2
Penelope-27_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-28_LMi	.	Locusta migratoria	Penelope	73,1,3,103,2
Penelope-29_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-2B_Crp	.	Crocodylus porosus	Penelope	1,3,71,72,2,74
Penelope-3_AAe	.	Aedes aegypti	Penelope	73,1,70,3,71,103,72,2
Penelope-3_ACar	.	Anolis carolinensis	Penelope	73,74,1,3,70,103,71,72,2
Penelope-3_ADi	.	Acropora digitifera	Penelope	
Penelope-3_BF	.	Branchiostoma floridae	Penelope	1,3,71,103,72,74
Penelope-3_BTa	.	Bemisia tabaci	Penelope	73,1,3,71,72,74
Penelope-3_CGi	.	Crassostrea gigas	Penelope	73,1,3,70,71,72,74
Penelope-3_CoFl	.	Copidosoma floridanum	Penelope	73,1,3,70,71,72,74
Penelope-3_CQ	.	Culex quinquefasciatus	Penelope	1,3,72
Penelope-3_Crp	.	Crocodylus porosus	Penelope	
Penelope-3_DR	.	Danio rerio	Penelope	
Penelope-3_FoCa	.	Folsomia candida	Penelope	
Penelope-3_HM	.	Hydra vulgaris	Penelope	
Penelope-3_HRo	.	Helobdella robusta	Penelope	1,3,103
Penelope-3_LCh	.	Latimeria chalumnae	Penelope	
Penelope-3_LG	.	Lottia gigantea	Penelope	73,70,71,72,74
Penelope-3_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-3_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,103,2

Penelope-3_MLe	.	Mnemiopsis leidyi	Penelope	3,1		
Penelope-3_NV	.	Nematostella vectensis	Penelope			
Penelope-3_OD	.	Oikopleura dioica	Penelope		73	
Penelope-3_PH	.	Parhyale hawaiensis	Penelope			
Penelope-3_PSi	.	Pelodiscus sinensis	Penelope	74,1,3,70,71,103,72,2		
Penelope-3_SIn	.	Solenopsis invicta	Penelope	1,3,103,72,2		
Penelope-3_SSa	.	Salmo salar	Penelope	1,3,72,2		
Penelope-3_XL	.	Xenopus laevis	Penelope	71,74		
Penelope-30_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2		
Penelope-31_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2		
Penelope-32_LMi	.	Locusta migratoria	Penelope	1,3,72,2		
Penelope-33_LMi	.	Locusta migratoria	Penelope			
Penelope-34_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2		
Penelope-35_LMi	.	Locusta migratoria	Penelope			103
Penelope-36_LMi	.	Locusta migratoria	Penelope	1,3,72,2		
Penelope-37_LMi	.	Locusta migratoria	Penelope	1,3,103,2		
Penelope-37B_LMi	.	Locusta migratoria	Penelope			
Penelope-38_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2		
Penelope-39_LMi	.	Locusta migratoria	Penelope	1,3,103,72		
Penelope-3B_DR	.	Danio rerio	Penelope			
Penelope-3B_PH	.	Parhyale hawaiensis	Penelope	74,1,3,71,2		
Penelope-3B_PSi	.	Pelodiscus sinensis	Penelope	1,3,70,71,103,72,2		
Penelope-3N1_LMi	.	Locusta migratoria	Penelope			

Penelope-4_AAe	.	Aedes aegypti	Penelope	74,73,1,3,70,103,71,72,2
Penelope-4_ACar	.	Anolis carolinensis	Penelope	1,3,71,74
Penelope-4_ADi	.	Acropora digitifera	Penelope	1,3,71,103,72,74
Penelope-4_BTa	.	Bemisia tabaci	Penelope	73,1,3,71,72,74
Penelope-4_CoFl	.	Copidosoma floridanum	Penelope	73,72
Penelope-4_CQ	.	Culex quinquefasciatus	Penelope	1,3,103
Penelope-4_Crp	.	Crocodylus porosus	Penelope	
Penelope-4_DR	.	Danio rerio	Penelope	
Penelope-4_FoCa	.	Folsomia candida	Penelope	73,70,71,72,74
Penelope-4_HM	.	Hydra vulgaris	Penelope	
Penelope-4_HRo	.	Helobdella robusta	Penelope	103,3,1
Penelope-4_LG	.	Lottia gigantea	Penelope	73,70,71,72
Penelope-4_LMi	.	Locusta migratoria	Penelope	1,3,103,72
Penelope-4_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,72,2
Penelope-4_NV	.	Nematostella vectensis	Penelope	73,1,3,70,71,74
Penelope-4_OD	.	Oikopleura dioica	Penelope	1,3,103,2
Penelope-4_PH	.	Parhyale hawaiensis	Penelope	1,3,71,2,74
Penelope-4_SIn	.	Solenopsis invicta	Penelope	74,73,1,3,71,72,2
Penelope-4_SSa	.	Salmo salar	Penelope	1,3,72
Penelope-4_XL	.	Xenopus laevis	Penelope	72
Penelope-40_LMi	.	Locusta migratoria	Penelope	73,74,1,3,70,103,71,2
Penelope-41_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-42_LMi	.	Locusta migratoria	Penelope	73,1,3,2

Penelope-43_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-44_LMi	.	Locusta migratoria	Penelope	1,3,71,74
Penelope-44B_LMi	.	Locusta migratoria	Penelope	
Penelope-45_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-46_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-47_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-48_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-49_LMi	.	Locusta migratoria	Penelope	1,3,72,2
Penelope-4N1_LVa	.	Litopenaeus vannamei	Penelope	
Penelope-5_AAe	.	Aedes aegypti	Penelope	1,3,103,72,2
Penelope-5_ACar	.	Anolis carolinensis	Penelope	73,74,1,3,71,72,2
Penelope-5_BTa	.	Bemisia tabaci	Penelope	103,72
Penelope-5_CoFl	.	Copidosoma floridanum	Penelope	73,71,72,74
Penelope-5_DR	.	Danio rerio	Penelope	1,3,71,72,74
Penelope-5_FoCa	.	Folsomia candida	Penelope	1,3,103,2
Penelope-5_HM	.	Hydra vulgaris	Penelope	
Penelope-5_HRo	.	Helobdella robusta	Penelope	1,3,2
Penelope-5_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-5_LVa	.	Litopenaeus vannamei	Penelope	1,3,70,71,103,72,2
Penelope-5_NV	.	Nematostella vectensis	Penelope	1,3,70,71,72
Penelope-5_OD	.	Oikopleura dioica	Penelope	
Penelope-5_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-5_SIn	.	Solenopsis invicta	Penelope	74,73,1,3,71,72,2

Penelope-5_SSa	.	Salmo salar	Penelope	73,1,3,70,71,72,74
Penelope-5_XL	.	Xenopus laevis	Penelope	74,1,3,71,103,72
Penelope-5_XT	.	Xenopus tropicalis	Penelope	74,1,71,103,72,2
Penelope-50_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-51_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,2
Penelope-52_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-52B_LMi	.	Locusta migratoria	Penelope	
Penelope-53_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-54_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-55_LMi	.	Locusta migratoria	Penelope	73,74,1,3,71,2
Penelope-56_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-57_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-58_LMi	.	Locusta migratoria	Penelope	74,1,3,71,2
Penelope-59_LMi	.	Locusta migratoria	Penelope	73,1,3,103,2
Penelope-5N1_BTa	.	Bemisia tabaci	Penelope	
Penelope-6_ACar	.	Anolis carolinensis	Penelope	1,3,71,72,74
Penelope-6_BTa	.	Bemisia tabaci	Penelope	1,3,2
Penelope-6_CoFl	.	Copidosoma floridanum	Penelope	73,70,71,72,74
Penelope-6_FoCa	.	Folsomia candida	Penelope	1,3,70,71,103,72,2
Penelope-6_HM	.	Hydra vulgaris	Penelope	
Penelope-6_HRo	.	Helobdella robusta	Penelope	1,3,2
Penelope-6_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-6_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,2

Penelope-6_NV	.	Nematostella vectensis	Penelope			
Penelope-6_OD	.	Oikopleura dioica	Penelope		103	
Penelope-6_PH	.	Parhyale hawaiensis	Penelope	1,3,2		
Penelope-6_SIn	.	Solenopsis invicta	Penelope	73,74,1,3,70,71,103,72,2		
Penelope-6_SSa	.	Salmo salar	Penelope	73,62		
Penelope-6_XL	.	Xenopus laevis	Penelope	73,72		
Penelope-6_XT	.	Xenopus tropicalis	Penelope	1,3,103,72,2		
Penelope-60_LMi	.	Locusta migratoria	Penelope	73,1,3,103,2		
Penelope-62_LMi	.	Locusta migratoria	Penelope	1,3,2		
Penelope-63_LMi	.	Locusta migratoria	Penelope	1,3,2		
Penelope-64_LMi	.	Locusta migratoria	Penelope	1,3,103,2		
Penelope-65_LMi	.	Locusta migratoria	Penelope	1,3,2		
Penelope-66_LMi	.	Locusta migratoria	Penelope			72
Penelope-67_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2		
Penelope-68_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2		
Penelope-69_LMi	.	Locusta migratoria	Penelope			73
Penelope-69B_LMi	.	Locusta migratoria	Penelope			
Penelope-6N1_NV	.	Nematostella vectensis	Penelope			
Penelope-7_ACar	.	Anolis carolinensis	Penelope	73,1,3,103,72,2		
Penelope-7_BTa	.	Bemisia tabaci	Penelope	1,3,71,74		
Penelope-7_CoFl	.	Copidosoma floridanum	Penelope	73,1,3,71,72,74		
Penelope-7_FoCa	.	Folsomia candida	Penelope	1,3,103,72,2		
Penelope-7_HM	.	Hydra vulgaris	Penelope			

Penelope-7_HRo	.	Helobdella robusta	Penelope	
Penelope-7_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-7_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,2
Penelope-7_NV	.	Nematostella vectensis	Penelope	
Penelope-7_PH	.	Parhyale hawaiensis	Penelope	74,1,3,71,103,2
Penelope-7_SSa	.	Salmo salar	Penelope	1,3,71,72,74
Penelope-7_XL	.	Xenopus laevis	Penelope	
Penelope-7_XT	.	Xenopus tropicalis	Penelope	18,74,5,1,71,103,72,2
Penelope-70_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-71_LMi	.	Locusta migratoria	Penelope	1,3,70,71
Penelope-72_LMi	.	Locusta migratoria	Penelope	73,1,3,70,71,74
Penelope-72N1_LMi	.	Locusta migratoria	Penelope	
Penelope-73_LMi	.	Locusta migratoria	Penelope	73,74,1,3,71,2
Penelope-73N1_LMi	.	Locusta migratoria	Penelope	
Penelope-74_LMi	.	Locusta migratoria	Penelope	
Penelope-75_LMi	.	Locusta migratoria	Penelope	3,1
Penelope-76_LMi	.	Locusta migratoria	Penelope	74,1,3,71,2
Penelope-77_LMi	.	Locusta migratoria	Penelope	74,1,3,71,103,72,2
Penelope-78_LMi	.	Locusta migratoria	Penelope	
Penelope-79_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-8_ACar	.	Anolis carolinensis	Penelope	73,70,71,72,74
Penelope-8_BTa	.	Bemisia tabaci	Penelope	1,103,72,2
Penelope-8_CoFl	.	Copidosoma floridanum	Penelope	1,3,72

Penelope-8_FoCa	.	Folsomia candida	Penelope	1,3,103,72
Penelope-8_HM	.	Hydra vulgaris	Penelope	
Penelope-8_HRo	.	Helobdella robusta	Penelope	1,3,103,2
Penelope-8_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-8_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,2
Penelope-8_NV	.	Nematostella vectensis	Penelope	
Penelope-8_PH	.	Parhyale hawaiensis	Penelope	1,3,103,2
Penelope-8_SSa	.	Salmo salar	Penelope	
Penelope-8_XL	.	Xenopus laevis	Penelope	
Penelope-8_XT	.	Xenopus tropicalis	Penelope	
Penelope-80_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-81_LMi	.	Locusta migratoria	Penelope	
Penelope-82_LMi	.	Locusta migratoria	Penelope	3,1
Penelope-83_LMi	.	Locusta migratoria	Penelope	74,1,3,71,103,2
Penelope-84_LMi	.	Locusta migratoria	Penelope	
Penelope-85_LMi	.	Locusta migratoria	Penelope	
Penelope-86_LMi	.	Locusta migratoria	Penelope	73,1,3,103,72
Penelope-87_LMi	.	Locusta migratoria	Penelope	
Penelope-88_LMi	.	Locusta migratoria	Penelope	
Penelope-89_LMi	.	Locusta migratoria	Penelope	
Penelope-9_ACar	.	Anolis carolinensis	Penelope	73,1,3,72
Penelope-9_CoFl	.	Copidosoma floridanum	Penelope	1,3,72
Penelope-9_FoCa	.	Folsomia candida	Penelope	1,3,103,72

Penelope-9_HM	.	Hydra vulgaris	Penelope		
Penelope-9_HRo	.	Helobdella robusta	Penelope	1,3,103,2	
Penelope-9_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2	
Penelope-9_LVa	.	Litopenaeus vannamei	Penelope	18,73,5,1,3,71,74	
Penelope-9_NV	.	Nematostella vectensis	Penelope		
Penelope-9_PH	.	Parhyale hawaiensis	Penelope		
Penelope-9_XL	.	Xenopus laevis	Penelope	73,1,3,71,74	
Penelope-9_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,103,72,2	
Penelope-90_LMi	.	Locusta migratoria	Penelope		
Penelope-91_LMi	.	Locusta migratoria	Penelope		
Penelope-92_LMi	.	Locusta migratoria	Penelope		
Penelope-93_LMi	.	Locusta migratoria	Penelope		
Penelope-94_LMi	.	Locusta migratoria	Penelope		
Penelope-94N1_LMi	.	Locusta migratoria	Penelope		
Penelope-95_LMi	.	Locusta migratoria	Penelope		
Penelope-96_LMi	.	Locusta migratoria	Penelope	1,3,70,71,103,74	
Penelope-97_LMi	.	Locusta migratoria	Penelope	71,73,74	
Penelope-98_LMi	.	Locusta migratoria	Penelope		
Penelope-99_LMi	.	Locusta migratoria	Penelope		73
Penelope-N1_PH	.	Parhyale hawaiensis	Penelope		
Penelope-N1_XL	.	Xenopus laevis	Penelope		
Penelope-N1_XT	.	Xenopus tropicalis	Penelope		
Penelope-N10_XL	.	Xenopus laevis	Penelope		

Penelope-N12_XL	.	Xenopus laevis	Penelope
Penelope-N13_XL	.	Xenopus laevis	Penelope
Penelope-N14_XL	.	Xenopus laevis	Penelope
Penelope-N15_XL	.	Xenopus laevis	Penelope
Penelope-N15B_XL	.	Xenopus laevis	Penelope
Penelope-N16_XL	.	Xenopus laevis	Penelope
Penelope-N17_XL	.	Xenopus laevis	Penelope
Penelope-N18_XL	.	Xenopus laevis	Penelope
Penelope-N19_XL	.	Xenopus laevis	Penelope
Penelope-N2_XL	.	Xenopus laevis	Penelope
Penelope-N20_XL	.	Xenopus laevis	Penelope
Penelope-N21_XL	.	Xenopus laevis	Penelope
Penelope-N22_XL	.	Xenopus laevis	Penelope
Penelope-N23_XL	.	Xenopus laevis	Penelope
Penelope-N3_XL	.	Xenopus laevis	Penelope
Penelope-N4_XL	.	Xenopus laevis	Penelope
Penelope-N447_AMi	.	Alligator mississippiensis	Penelope
Penelope-N5_XL	.	Xenopus laevis	Penelope
Penelope-N6_XL	.	Xenopus laevis	Penelope
Penelope-N7_XL	.	Xenopus laevis	Penelope
Penelope-N8_XL	.	Xenopus laevis	Penelope
Penelope-N9_XL	.	Xenopus laevis	Penelope
Penelope1_CI	.	Ciona intestinalis	Penelope

Penelope1_Dw	.	Drosophila willistoni	Penelope	1,3,71,103,72,74
Penelope1_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2
Penelope1A_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,2,74
Penelope1B_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2
Penelope1C_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,2,74
Penelope1D_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2
Penelope1N_CPB	.	Chrysemys picta bellii	Penelope	
Penelope2_Dw	.	Drosophila willistoni	Penelope	74,1,3,71,103,72,2
Penelope2_XT	.	Xenopus tropicalis	Penelope	1,3,72
Penelope3_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2
Penelope4_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,74
Perere_INT	BK004067	Schistosoma mansoni	CR1	16,1,3,11,4,2,15
Perere_Smed	.	Schmidtea mediterranea	Penelope	1,3,103,2
PERERE-10	BN000801	Schistosoma mansoni	Penelope	1,3,103,2
PERERE-2	BN000793	Schistosoma mansoni	CR1	1,3,59,4,57,2
Perere-3	BN000794	Schistosoma mansoni	RTE	16,1,26,3,11,4,2,15
PERERE-4	BN000795	Schistosoma mansoni	CR1	16,1,3,4,11,2,15
PERERE-5	BN000796	Schistosoma mansoni	CR1	16,1,3,59,60,11,4,57,2,15
PERERE-6	BN000797	Schistosoma mansoni	CR1	1,3,59,4,57,2
PERERE-7	BN000798	Schistosoma mansoni	CR1	16,1,3,11,4,2,15
PERERE-8	BN000799	Schistosoma mansoni	Rex1	1,3,4,2
PERERE-9	BN000800	Schistosoma mansoni	R2	18,5,19,1,3,4,2
Perere2_Smed	.	Schmidtea mediterranea	Penelope	73,1,3,103,2

PILN1_PT	AB016145	Pinus thunbergii	L1	
PILN2_PT	AB016146	Pinus thunbergii	L1	
PILN3_PT	AB016147	Pinus thunbergii	L1	
Plat_L3	.	Ornithorhynchus	CR1	16,1,3,11,4,2,15
Plat_L3b	.	Ornithorhynchus	CR1	
Plat_R4	.	Ornithorhynchus	R4	
Plat_RTE1	.	Ornithorhynchus	RTE	
Plat_RTE1_SINE	.	Ornithorhynchus	SINE	
PlatCR1	.	Ornithorhynchus	CR1	
PlatCR1_old1	.	Ornithorhynchus	CR1	
PlatCR1_old2	.	Ornithorhynchus	CR1	
PlatSINE1	.	Ornithorhynchus	SINE	
Platy-1-1	.	Callithrix jacchus	SINE1/7SL	
PNL_SM	.	Schmidtea mediterranea	Penelope	1,3,2
PNL2_SM	.	Schmidtea mediterranea	Penelope	73,1,3,70,71,103,2
POLN2_ST	AB016141, AB016141.1	Solanum tuberosum	L1	
POMSINE	.	Podarcis muralis	SINE2/тРНК	
Poseidon_Ac	.	Anolis carolinensis	Penelope	1,3,103,72,2
Poseidon_Ap	.	Acropora palmata	Penelope	1,3,71,103,72,74
Poseidon_Hyd	.	Hydra vulgaris	Penelope	71,72,74
Poseidon-1_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-10_HM	.	Hydra vulgaris	Penelope	71,74
Poseidon-1B_PM	.	Petromyzon marinus	Penelope	1,3,71,103,74

Poseidon-1C_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-1D_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-1E_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-1F_PM	.	Petromyzon marinus	Penelope	1,3,71,103,72,74
Poseidon-1G_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-1H_PM	.	Petromyzon marinus	Penelope	
Poseidon-1I_PM	.	Petromyzon marinus	Penelope	
Poseidon-1J_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-1K_PM	.	Petromyzon marinus	Penelope	
Poseidon-1N1_PM	.	Petromyzon marinus	Penelope	
Poseidon-1N2_PM	.	Petromyzon marinus	Penelope	
Poseidon-2_HM	.	Hydra vulgaris	Penelope	71,72,74
Poseidon-2_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2
Poseidon-2B_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2
Poseidon-2C_PM	.	Petromyzon marinus	Penelope	74,71
Poseidon-2D_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2
Poseidon-3_HM	.	Hydra vulgaris	Penelope	
Poseidon-3_PM	.	Petromyzon marinus	Penelope	74,73,1,3,71,103,72,2
Poseidon-3B_PM	.	Petromyzon marinus	Penelope	74,73,1,3,71,103,72,2
Poseidon-4_HM	.	Hydra vulgaris	Penelope	
Poseidon-4_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2
Poseidon-5_HM	.	Hydra vulgaris	Penelope	71,74
Poseidon-6_HM	.	Hydra vulgaris	Penelope	1,3

Poseidon-N1_PM	.	Petromyzon marinus	Penelope	
Poseidon-N2_PM	.	Petromyzon marinus	Penelope	
Poseidon-N3_PM	.	Petromyzon marinus	Penelope	
Poseidon-N3B_PM	.	Petromyzon marinus	Penelope	
Poseidon-N4_PM	.	Petromyzon marinus	Penelope	
Poseidon-N5_PM	.	Petromyzon marinus	Penelope	
Poseidon-N6_PM	.	Petromyzon marinus	Penelope	
Pp_NLTR-B	.	Polysphondylium pallidum	Не содержащий LTR ретротранспозон	6,2,10,11,13,1,4,17,3,7
Pp_NLTR-C	.	Polysphondylium pallidum	Не содержащий LTR ретротранспозон	1,3,2
Pp_TRE3-A	.	Polysphondylium pallidum	Не содержащий LTR ретротранспозон	1,3,11,4,2
Pp_TRE3-B	.	Polysphondylium pallidum	Не содержащий LTR ретротранспозон	1,3,11,4,2
Pre0_SS	.	Sus scrofa	SINE2/тРНК	
PRE1_SS	Y00104	Sus scrofa	SINE2/тРНК	
PRE1f2	.	Suina	SINE2/тРНК	
PRE1h	.	Suina	SINE2/тРНК	
PRE1h2	.	Suina	SINE2/тРНК	
PRE1i	.	Suina	SINE2/тРНК	
PRE1j	.	Suina	SINE2/тРНК	
PRE1k	.	Suina	SINE2/тРНК	
PRLN1_LP	AB016125	Lolium perenne	L1	
Proto1-1_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,2,6,13
Proto1-2_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,6,2,13
Proto1-3_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,2,6,13

Proto1-4_NG	.	Naegleria gruberi	Proto1	6,2,9,10,11,13,1,4,17,3,7
Proto1-5_NG	.	Naegleria gruberi	Proto1	6,2,7,9,10,11,13,1,4,17,3
Proto1-6_NG	.	Naegleria gruberi	Proto1	7,1,17,9,3,10,4,6,2,13
Proto2-1_BF	.	Branchiostoma floridae	Proto2	16,1,3,24,11,4,2
Proto2-1_BM	.	Bombyx mori	Proto2	1,3,2
Proto2-1_CS1	.	Capitella sp. 1	Proto2	16,1,3,24,11,4,2
Proto2-1_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-1_DR	.	Danio rerio	Proto2	2,1
Proto2-1_Hmel	.	Heliconius melpomene	Proto2	1,3,4,2
Proto2-1_LVa	.	Litopenaeus vannamei	Proto2	1,3,11,4,2
Proto2-1_OL	.	Oryzias latipes	Proto2	1,3,2
Proto2-1_PH	.	Parhyale hawaiensis	Proto2	16,1,3,24,11,4,2
Proto2-1_PPo	.	Papilio polytes	Proto2	16,1,3,11,4,2,15
Proto2-1_PXu	.	Papilio xuthus	Proto2	
Proto2-1_SK	.	Saccoglossus kowalevskii	Proto2	16,1,3,11,4,2,15
Proto2-1_SPur	.	Strongylocentrotus purpuratus	Proto2	2,11,56,57,68,1,58,4,69,3,59
Proto2-10_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-11_CTe	.	Capitella teleta	Proto2	1,3,11,4,2
Proto2-12_CTe	.	Capitella teleta	Proto2	1,3,11,4,2
Proto2-13_CTe	.	Capitella teleta	Proto2	16,1,3,24,4,11,2
Proto2-14_CTe	.	Capitella teleta	Proto2	1,3,11,4,2
Proto2-15_CTe	.	Capitella teleta	Proto2	16,1,3,24,11,4,2
Proto2-16_CTe	.	Capitella teleta	Proto2	1,3,4,2

Proto2-2_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15
Proto2-2_CTe	.	Capitella teleta	Proto2	
Proto2-2_Hmel	.	Heliconius melpomene	Proto2	
Proto2-2_HMM	.	Heliconius melpomene melpomene	Proto2	1,3,11,4,2
Proto2-2_LVa	.	Litopenaeus vannamei	Proto2	
Proto2-2_PH	.	Parhyale hawaiensis	Proto2	1,3,11,4,2
Proto2-2_PPo	.	Papilio polytes	Proto2	
Proto2-2_SK	.	Saccoglossus kowalevskii	Proto2	1,3,4,2
Proto2-3_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15
Proto2-3_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-3_Hmel	.	Heliconius melpomene	Proto2	1,3,4,2
Proto2-3_PPo	.	Papilio polytes	Proto2	
Proto2-3_SK	.	Saccoglossus kowalevskii	Proto2	1,3,82,11,4,2
Proto2-3B_PPo	.	Papilio polytes	Proto2	
Proto2-4_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2
Proto2-4_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-4_PPo	.	Papilio polytes	Proto2	
Proto2-5_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15
Proto2-5_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-5_PPo	.	Papilio polytes	Proto2	
Proto2-6_CS1	.	Capitella sp. 1	Proto2	16,1,3,24,11,4,2
Proto2-6_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-7_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2

Proto2-7_CTe	.	Capitella teleta	Proto2	1,3,4,2
Proto2-8_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2
Proto2-8_CTe	.	Capitella teleta	Proto2	3,2,1
Proto2-9_CTe	.	Capitella teleta	Proto2	3,2,1
PSINE1	.	Procavia capensis	SINE2/tPHK	
PSLINE	AB005891	Acanthochelys spixii	CR1	2,63,11,48,1,64,4,15,16,3
Pt1	.	Phaeodactylum tricornutum	Penelope	82
PTr-3	.	Populus trichocarpa	SINE2/tPHK	
Q	U03849	Anopheles gambiae	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,60
Q_As	.	Adineta	Penelope	1,3
Q_Av	.	Adineta vaga	Penelope	1,3
R_Av	.	Adineta vaga	Penelope	1,129,3,130
R0_As	.	Adineta	Penelope	1,3
R1_DEr	.	Drosophila erecta	R1	2,8,11,86,12,1,14,4,15,16,3
R1_DM	X51968	Drosophila melanogaster	R1	2,8,11,86,12,1,14,4,15,16,3
R1_DMo	.	Drosophila mojavensis	R1	2,8,11,12,157,1,14,4,15,16,3,158
R1_DPe	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1_DPs	.	Drosophila pseudoobscura	R1	128,2,8,11,86,12,1,14,4,15,16,3
R1_DSe	.	Drosophila sechellia	R1	2,8,11,86,12,1,14,4,15,16,3
R1_DSi	.	Drosophila simulans	R1	2,8,11,86,12,1,14,4,15,16,3
R1_DVi	.	Drosophila virilis	R1	2,8,11,12,42,1,14,4,15,16,3
R1_DWi	.	Drosophila willistoni	R1	2,8,11,12,1,14,4,15,16,3
R1_DYa	.	Drosophila yakuba	R1	2,8,11,86,12,1,14,4,15,16,3

R1_Ele1	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
R1_Ele4	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1_Ele7	.	Aedes aegypti	R1	1,3,4,2
R1_Ele8	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1_Ele9	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1_SC	L00945	Bradysia coprophila	R1	
R1-1_AEc	.	Acromyrmex echinatior	R1	16,1,3,11,4,86,2,15
R1-1_AP	Контиг61185	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_BM	.	Bombyx mori	R1	16,1,3,11,4,86,2,15
R1-1_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-1_CFl	.	Camponotus floridanus	R1	
R1-1_CoFl	.	Copidosoma floridanum	R1	16,1,3,11,4,86,2,15
R1-1_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-1_DBi	.	Drosophila biarmipes	R1	2,11,86,12,1,14,4,15,16,3
R1-1_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-1_DEl	.	Drosophila elegans	R1	16,1,3,11,4,86,2,15
R1-1_DEr	.	Drosophila erecta	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DEu	.	Drosophila eugracilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DF	.	Drosophila ficusphila	R1	2,11,86,12,1,14,4,15,16,3
R1-1_DGr	.	Drosophila grimshawi	R1	2,8,11,12,42,1,14,4,15,16,3
R1-1_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DRh	.	Drosophila rhopaloa	R1	2,8,11,86,12,1,14,4,15,16,3

R1-1_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DSi	.	Drosophila simulans	R1	16,1,3,11,4,86,2,15
R1-1_DT	.	Drosophila takahashii	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DVi	.	Drosophila virilis	R1	2,8,11,86,12,42,1,14,4,15,16,3,46
R1-1_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-1_Hmel	.	Heliconius melpomene	R1	1,3,11,4,2
R1-1_HMM	.	Heliconius melpomene melpomene	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7
R1-1_HSal	.	Harpegnathos saltator	R1	16,1,3,11,4,86,2,15
R1-1_LHu	.	Linepithema humile	R1	16,1,3,11,4,86,2,15
R1-1_LMi	.	Locusta migratoria	R1	6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,7
R1-1_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_PBa	.	Pogonomyrmex barbatus	R1	1,3,4,2
R1-1_PPo	.	Papilio polytes	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-1_PXu	.	Papilio xuthus	R1	
R1-1_SIn	.	Solenopsis invicta	R1	16,1,3,11,4,86,2,15
R1-1_TCas	.	Tribolium castaneum	R1	2,8,11,12,42,1,14,4,15,16,3
R1-10_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,12,2,15
R1-10_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-11_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-11_PPo	.	Papilio polytes	R1	
R1-12_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-12_PPo	.	Papilio polytes	R1	
R1-13_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3

R1-13_PPo	.	Papilio polytes	R1	
R1-14_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3
R1-14_PPo	.	Papilio polytes	R1	
R1-15_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-15_PPo	.	Papilio polytes	R1	
R1-15B_PPo	.	Papilio polytes	R1	
R1-16_DWi	.	Drosophila willistoni	R1	16,1,3,11,4,2,15
R1-16_PPo	.	Papilio polytes	R1	
R1-16B_PPo	.	Papilio polytes	R1	
R1-17_DWi	.	Drosophila willistoni	R1	16,1,3,11,4,86,2,15
R1-17_PPo	.	Papilio polytes	R1	
R1-18_DWi	.	Drosophila willistoni	R1	
R1-1B_CoFl	.	Copidosoma floridanum	R1	42,8,14,12
R1-1B_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1B_DVi	.	Drosophila virilis	R1	
R1-1C_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2_AP	Контиг3122	Acyrthosiphon pisum	R1	16,1,8,3,11,4,86,12,2,15
R1-2_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-2_CoFl	.	Copidosoma floridanum	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-2_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DBi	.	Drosophila biarmipes	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DEl	.	Drosophila elegans	R1	2,8,11,86,12,1,14,4,15,16,3

R1-2_DEu	.	Drosophila eugracilis	R1	16,1,8,3,11,4,86,12,2,15
R1-2_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DGr	.	Drosophila grimshawi	R1	2,8,11,12,42,1,14,4,15,16,3
R1-2_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2_DM	.	Drosophila melanogaster	R1	2,8,11,12,1,14,4,15,16,3
R1-2_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DTa	.	Drosophila takahashii	R1	2,11,86,12,42,1,14,4,15,16,3
R1-2_DVi	.	Drosophila virilis	R1	2,11,86,12,1,14,4,15,16,3
R1-2_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-2_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-2_Hmel	.	Heliconius melpomene	R1	16,1,3,11,4,86,2,15
R1-2_LMi	.	Locusta migratoria	R1	1,8,14,3,4,12,2
R1-2_PBa	.	Pogonomyrmex barbatus	R1	16,1,3,4,11,86,2,15
R1-2_PPo	.	Papilio polytes	R1	7,1,17,9,3,10,4,6,2,13
R1-2_SIn	.	Solenopsis invicta	R1	1,3,4,2
R1-2_TCas	.	Tribolium castaneum	R1	16,1,3,4,11,86,2,15
R1-2B_AP	Контиг17211	Acyrthosiphon pisum	R1	2,8,11,12,1,14,4,15,16,3
R1-2B_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2B_PPo	.	Papilio polytes	R1	
R1-2C_PPo	.	Papilio polytes	R1	
R1-3_AP	Контиг971	Acyrthosiphon pisum	R1	16,1,14,3,11,4,86,2,15
R1-3_BM	.	Bombyx mori	R1	

R1-3_BTa	.	Bemisia tabaci	R1	2,16,1,14,3,11,4,15
R1-3_CoFl	.	Copidosoma floridanum	R1	
R1-3_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DEl	.	Drosophila elegans	R1	16,1,8,11,12,2,15
R1-3_DEu	.	Drosophila eugracilis	R1	
R1-3_DF	.	Drosophila ficusphila	R1	16,1,3,11,4,86,2,15
R1-3_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-3_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DTa	.	Drosophila takahashii	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DVi	.	Drosophila virilis	R1	1,3,4,2
R1-3_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DYa	.	Drosophila yakuba	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_LMi	.	Locusta migratoria	R1	16,1,3,11,4,2,15
R1-3_PBa	.	Pogonomyrmex barbatus	R1	190,2,8,11,86,12,42,1,14,4,15,16,3
R1-3_PPo	.	Papilio polytes	R1	
R1-3_SIn	.	Solenopsis invicta	R1	1,3,11,4,2
R1-3B_DEl	.	Drosophila elegans	R1	1,3,4,2
R1-4_AP	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_BM	.	Bombyx mori	R1	1,3,4,2
R1-4_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-4_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DBp	.	Drosophila bipectinata	R1	16,1,3,11,4,86,2,15

R1-4_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DVi	.	Drosophila virilis	R1	
R1-4_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-4_PPo	.	Papilio polytes	R1	
R1-4B_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-5_BM	.	Bombyx mori	R1	2,8,24,11,86,12,42,1,14,16
R1-5_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-5_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DK	.	Drosophila kikkawai	R1	1,3,11,4,2
R1-5_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-5_PPo	.	Papilio polytes	R1	
R1-5B_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-6_BM	.	Bombyx mori	R1	
R1-6_BTa	.	Bemisia tabaci	R1	2,8,11,86,12,93,42,1,14,4,15,16,3
R1-6_DK	.	Drosophila kikkawai	R1	43,2,8,44,11,86,12,42,1,14,4,15,16,3
R1-6_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-6_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-6_PPo	.	Papilio polytes	R1	

R1-7_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-7_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,2,15
R1-7_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-7_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-7_PPo	.	Papilio polytes	R1	
R1-8_BTa	.	Bemisia tabaci	R1	2,8,11,12,1,14,4,15,16,3
R1-8_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3
R1-8_PPo	.	Papilio polytes	R1	
R1-9_BTa	.	Bemisia tabaci	R1	16,1,8,3,11,4,12,2,15
R1-9_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-9_PPo	.	Papilio polytes	R1	
R1-N1_CFl	.	Camponotus floridanus	R1	8,14,12
R1-N1_SIn	.	Solenopsis invicta	R1	8,14,12
R1A_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,157,1,14,4,15,16,3,158
R1A_DS	AF015489	Aphonopelma sp. WDB-1998	R1	7,1,17,9,3,10,4,6,2,13
R1A_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1A_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,1,14,4,15,16,3
R1B_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1B_DSer	.	Drosophila serrata	R1	2,8,11,12,1,14,4,15,16,3
R1B_DVi	.	Drosophila virilis	R1	2,8,11,12,42,1,14,4,15,16,3
R1B_NVi	.	Nasonia vitripennis	R1	2,8,11,12,1,14,4,15,16,3
R1B2_DAn	.	Drosophila ananassae	R1	
R1C_NGi	.	Nasonia giraulti	R1	16,1,8,3,11,4,12,2,15

R1C_NLo	.	Nasonia longicornis	R1	16,1,8,3,11,4,86,12,2,15
R1C_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1D_NGi	.	Nasonia giraulti	R1	16,1,3,4,11,2,15
R1D_NLo	.	Nasonia longicornis	R1	16,1,3,11,4,2,15
R1D_NVi	.	Nasonia vitripennis	R1	16,1,3,4,11,2,15
R1E_NGi	.	Nasonia giraulti	R1	2,8,11,86,12,1,14,4,15,16,3
R1N_NLo	.	Nasonia longicornis	R1	
R1NS-1_CQ	.	Culex quinquefasciatus	R1	2,8,11,86,12,1,14,4,15,16,3
R2_AM	AF015815	Anurida maritima	R2	
R2_BM	M16558	Bombyx mori	R2	18,5,19,1,3,4,2
R2_CI	AB097121	Ciona intestinalis	R2	
R2_CPB	.	Chrysemys picta bellii	R2	
R2_DA	U13640	Drosophila ambigua	R2	1,3,2
R2_DAn	.	Drosophila ananassae	R2	18,5,1,3,4,2
R2_DM	X51967	Drosophila melanogaster	R2	18,5,19,1,3,4,2
R2_DPe	.	Drosophila persimilis	R2	18,5,1,19,3,4,2
R2_DPs	.	Drosophila pseudoobscura	R2	18,5,1,19,3,4,2
R2_DSe	.	Drosophila sechellia	R2	18,5,19,1,3,4,2
R2_DSer	.	Drosophila serrata	R2	1,3,4,2
R2_DSi	.	Drosophila simulans	R2	18,5,19,1,3,4,2
R2_DYa	.	Drosophila yakuba	R2	18,5,19,1,3,4,2
R2_FA	AF015819	Forficula auricularia	R2	
R2_HC	AF015816	Hippodamia convergens	R2	

R2_KF	GU949558	Kalotermes flavicollis	R2	18,5,19,1,3,4,2
R2_LP	AF015814	Limulus polyphemus	R2	
R2_PS	AF015818	Porcellio scaber	R2	
R2_RL	GU949555	Reticulitermes lucifugus	R2	106,18,104,5,19,1,3,4,2
R2_RU	GU949554	Reticulitermes urbis	R2	106,18,104,5,19,1,3,4,2
R2-1_AAm	.	Amblyomma americanum	R2	1,3,2
R2-1_ACC	.	Aquila chrysaetos canadensis	R2	
R2-1_ACh	.	Acanthisitta chloris	R2	
R2-1_AFo	.	Aptenodytes forsteri	R2	
R2-1_AMi	.	Alligator mississippiensis	R2	106,18,104,19,1,5,3,4,2
R2-1_ApA	.	Apteryx australis mantelli	R2	
R2-1_APi	.	Acyrthosiphon pisum	R2	18,5,1,3,4,2
R2-1_BRG	.	Balearica regulorum gibbericep	R2	
R2-1_BTe	.	Bombus terrestris	R2	18,5,1,19,3,4,2
R2-1_CAnn	.	Calypte anna	R2	
R2-1_CAu	.	Cathartes aura	R2	
R2-1_CBr	.	Corvus brachyrhynchos	R2	
R2-1_CCa	.	Antrostomus carolinensis	R2	
R2-1_CCan	.	Cuculus canorus	R2	
R2-1_CPu	.	Calidris pugnax	R2	
R2-1_Crp	.	Crocodylus porosus	R2	5,1,19,3,4,2
R2-1_CSt	.	Colius striatus	R2	
R2-1_CU	.	Chlamydotis macqueenii	R2	

R2-1_CVo	.	Charadrius vociferus	R2	
R2-1_DWi	.	Drosophila willistoni	R2	18,5,19,1,3,4,2
R2-1_EGa	.	Egretta garzetta	R2	
R2-1_FAl	.	Ficedula albicollis	R2	106,18,104,19,1,5,3,4,2
R2-1_FCh	.	Falco cherrug	R2	
R2-1_FPe	.	Falco peregrinus	R2	
R2-1_GA	.	Gasterosteus aculeatus	R2	18,5,1,19,3,4,2
R2-1_Gav	.	Gavialis gangeticus	R2	18,5,1,19,3,4,2
R2-1_GFo	.	Geospiza fortis	R2	106,18,104,1,19,5,108,3,4,2
R2-1_GSt	.	Gavia stellata	R2	
R2-1_HAl	.	Haliaeetus albicilla	R2	106,18,5,104,19,1,3,4,2
R2-1_IS	.	Ixodes scapularis	R2	18,5,19,1,3,4,143,2
R2-1_LCh	.	Latimeria chalumnae	R2	18,5,19,1,3,4,2
R2-1_LDi	.	Leptosomus discolor	R2	
R2-1_LSal	.	Lepeophtheirus salmonis	R2	
R2-1_LV	.	Lytechinus variegatus	R2	106,18,104,19,1,5,3,4,2
R2-1_MDe	.	Mayetiola destructor	R2	18,5,1,3,4,2
R2-1_MLe	.	Mnemiopsis leidyi	R2	18,5,1,3,4,2
R2-1_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-1_MUn	.	Melopsittacus undulatus	R2	
R2-1_MUni	.	Mesitornis unicolor	R2	18,5,104,19,107
R2-1_MVi	.	Manacus vitellinus	R2	
R2-1_NNi	.	Nipponia nippon	R2	18,5,1,19,3,2

R2-1_NV	.	Nematostella vectensis	R2	
R2-1_OHo	.	Opisthocomus hoazin	R2	106,18,5,104,19,1,3,4,2
R2-1_PAd	.	Pygoscelis adeliae	R2	
R2-1_PBa	.	Pogonomyrmex barbatus	R2	1,3,4,2
R2-1_PCar	.	Phalacrocorax carbo	R2	
R2-1_PCau	.	Priapulus caudatus	R2	1,3,4,2
R2-1_PCr	.	Podiceps cristatus	R2	
R2-1_PCri	.	Pelecanus crispus	R2	
R2-1_PGu	.	Pterocles gutturalis	R2	
R2-1_PLe	.	Phaethon lepturus	R2	
R2-1_PM	.	Petromyzon marinus	R2	18,5,1,19,3,4,2
R2-1_PPap	.	Phlebotomus papatasi	R2	18,5,19,1,3,4,2
R2-1_PPu	.	Picoides pubescens	R2	
R2-1_PRR	.	Phoenicopterus ruber ruber	R2	
R2-1_PSi	.	Pelodiscus sinensis	R2	1,3,4,2
R2-1_RMi	.	Rhipicephalus microplus	R2	18,5,1,19,3,4,2
R2-1_RPr	.	Rhodnius prolixus	R2	1,3,4,2
R2-1_RPu	.	Rhipicephalus pulchellus	R2	1,3,4,2
R2-1_SCa	.	Serinus canaria	R2	104,105,2,18,106,1,4,5,19,107,108,3
R2-1_SK	.	Saccoglossus kowalevskii	R2	1,3,4,2
R2-1_SM	.	Schmidtea mediterranea	R2	18,5,1,3,4,2
R2-1_SP	.	Strongylocentrotus purpuratus	R2	106,18,5,104,19,1,3,4,2
R2-1_SSa	AGKD01072455	Salmo salar	R2	18,5,19,1,3,4,2

R2-1_StC	.	Struthio camelus australis	R2	
R2-1_TAl	.	Tyto alba	R2	
R2-1_TCas	.	Tribolium castaneum	R2	18,5,1,3,4,2
R2-1_TG	.	Taeniopygia guttata	R2	18,5,1,19,3,4,2
R2-1_TGut	.	Tinamus guttatus	R2	106,18,104,1,19,5,3,4,2
R2-1_TSP	.	Trichinella spiralis	R2	18,5,19,1,3,4,2
R2-1_TUr	каркас_6	Tetranychus urticae	R2	1,3,4,2
R2-1_XM	.	Xiphophorus maculatus	R2	18,5,1,19,3,4,2
R2-1_ZA	.	Zonotrichia albicollis	R2	18,19,1,5,108,3,4,2
R2-1_ZLM	.	Zosterops lateralis melanops	R2	18,5,19
R2-2_APi	.	Acyrthosiphon pisum	R2	1,3,4,2
R2-2_CCan	.	Cuculus canorus	R2	
R2-2_CMa	.	Chlamydotis macqueenii	R2	
R2-2_DWi	.	Drosophila willistoni	R2	18,5,1,19,3,4,2
R2-2_HAl	.	Haliaeetus albicilla	R2	
R2-2_IS	.	Ixodes scapularis	R2	1,4,2
R2-2_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-2_MUn	.	Melopsittacus undulatus	R2	
R2-2_MUni	.	Mesitornis unicolor	R2	
R2-2_NNi	.	Nipponia nippon	R2	
R2-2_NV	.	Nematostella vectensis	R2	1,3,4,2
R2-2_PBa	.	Pogonomyrmex barbatus	R2	18,5,1,19,3,4,2
R2-2_PM	.	Petromyzon marinus	R2	18,5,19,1,3,4,2

R2-2_RPr	.	Rhodnius prolixus	R2	1,3,4,2
R2-2_SMed	.	Schmidtea mediterranea	R2	18,5,1,19,3,4,2
R2-2_TCas	.	Tribolium castaneum	R2	18,5,1,19,3,4,2
R2-2_TUr	каркас_37	Tetranychus urticae	R2	18,5,1,3,4,2
R2-3_IS	ABJB010555169	Ixodes scapularis	R2	
R2-3_MR	.	Megachile rotundata	R2	1,3,4,2
R2-4_MR	.	Megachile rotundata	R2	18,5,19,1,3,4,2
R2-5_MR	.	Megachile rotundata	R2	1,3,4,2
R2-6_MR	.	Megachile rotundata	R2	
R2-7_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-8_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-N1_Gav	.	Gavialis gangeticus	R2	
R2-N2_Gav	.	Gavialis gangeticus	R2	
R2-N2B_Gav	.	Gavialis gangeticus	R2	
R2A_NVi	.	Nasonia vitripennis	R2	18,5,1,19,3,4,2
R2A_TM	AF015817	Tenebrio molitor	R2	
R2Amel	.	Apis mellifera	R2	18,5,19,1,3,4,2
R2B_DM	AF015685	Drosophila mercatorum	R2	
R2B_NVi	.	Nasonia vitripennis	R2	106,18,104,19,1,5,3,4,2
R2B_TM	AF015822	Tenebrio molitor	R2	
R2C_NGi	.	Nasonia giraulti	R2	18,5,1,3,4,2
R2Ci-B	AB097122	Ciona intestinalis	R2	18,5,1,19,3,4,2
R2Ci-D	.	Ciona intestinalis	R2	18,5,19,1,3,4,2

R2CIA_CI	AB097121	Ciona intestinalis	R2	
R2Cs-D	AB097125	Ciona savignyi	R2	
R2D_NGi	.	Nasonia giraulti	R2	
R2Dr	NM_001030097	Danio rerio	R2	18,5,19,1,3,4,2
R2E_NLo	.	Nasonia longicornis	R2	
R2Eb	AB201408	Eptatretus burgeri	R2	1,3,2
R2Ha	AB201415	Hasarius adansoni	R2	1,3,2
R2La	JN937617	Lepidurus arcticus	R2	1,3,4,2
R2LcA	.	Lepidurus couesii	R2	18,5,19,1,3,4,2
R2LcB	JN937619	Lepidurus couesii	R2	18,5,1,19,3,4,2
R2LcC	.	Lepidurus couesii	R2	1,3,2
R2Ll	JN937616	Lepidurus apus lubbocki	R2	18,5,1,19,3,4,2
R2Mr	AB201414	Metacrinus rotundus	R2	
R2NS-1_CGi	.	Crassostrea gigas	R2	18,5,1,19,3,4,2
R2NS-1_CSi	.	Clonorchis sinensis	R2	18,5,19,1,3,4,143,2
R2NS-1_PMi	.	Patiria miniata	R2	18,5,19,1,3,4,2
R2NS-1_SMed	.	Schmidtea mediterranea	R2	18,5,19,1,3,4,2
R2Nvec-A	.	Nematostella vectensis	R2	18,5,1,19,3,4,2
R2Ol-A	.	Oryzias latipes	R2	106,18,104,5,19,1,107,3,117,4,2
R2Pc	AB201416	Procambarus clarkii	R2	1,3,2
R2Sm-A	.	Schistosoma mansoni	R2	18,5,1,19,3,4,2
R2Ta	AB201409	Tanichthys albonubes	R2	3,1
R2Tc	EU854578	Triops cancriformis	R2	18,5,1,3,4,2

R2Tc_it	JN937621	Triops cancriformis	R2	1,3,2	
R2Tl	AB201417	Triops longicaudatus	R2		
R4_AL	U29445	Ascaris lumbricoides	R4	106,104,1,3,4,2	
R4_HC	U29590	Haemonchus contortus	R4		
R4_Hmel	.	Heliconius melpomene	R4	1,3,4,2	
R4-1_AC	.	Anolis carolinensis	R4	1,3,4,2	
R4-1_ADi	.	Acropora digitifera	R4	1,3,4,2	
R4-1_BM	.	Bombyx mori	R4	1,3,4,2	
R4-1_BX	CADV01008175	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-1_CJap	ABLE03011482	Caenorhabditis japonica	R4	1,3,4,2	
R4-1_CM	.	Callorhinchus milii	R4	2,1	
R4-1_CPB	.	Chrysemys picta bellii	R4		
R4-1_ED	.	Entamoeba dispar	R4	1,3,4,2	
R4-1_HG	.	Heterodera glycines	R4	1,3,4,2	
R4-1_HMe	.	Heliconius melpomene melpomene	R4	1,3,4,2	
R4-1_LVa	.	Litopenaeus vannamei	R4	1,3,4,2	
R4-1_MI	CABB01003843	Meloidogyne incognita	R4	1,3,4,2	
R4-1_PH	.	Parhyale hawaiensis	R4	1,3,4,2	
R4-1_SRa	CACX01002001	Strongyloides ratti	R4	1,3,2	
R4-1_TCa	.	Tribolium castaneum	R4		196,197
R4-1B_AC	.	Anolis carolinensis	R4	1,3,4,2	
R4-2_AS	.	Ascaris suum	R4	1,3,4,2	
R4-2_BX	CADV01009048	Bursaphelenchus xylophilus	R4	1,3,4,2	

R4-2_HG	ABLA01000389	Heterodera glycines	R4	1,3,4,2
R4-2_LVa	.	Litopenaeus vannamei	R4	1,3,4,2
R4-2_SRa	CACX01002006	Strongyloides ratti	R4	1,3,2
R4-3_BX	CADV01008832	Bursaphelenchus xylophilus	R4	1,3,4,2
R4-3_SRa	.	Strongyloides ratti	R4	1,3,2
R4-4_BX	.	Bursaphelenchus xylophilus	R4	1,3,4,2
R4-4_SRa	.	Strongyloides ratti	R4	1,3,2
R4-5_BX	.	Bursaphelenchus xylophilus	R4	1,3,4,2
R4-N1_ED	.	Entamoeba dispar	SINE	
R5	AY216701	Girardia tigrina	NeSL	1,3,4,2
R5-1_SM	.	Schmidtea mediterranea	NeSL	1,148,3,149,120,4,2
R5-2_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2
R6Ag1	AB090817	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3
R6Ag3	AB090819	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
R7Ag1	AB090820	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,242,15,16,3
R7Ag2	AB090821	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3
R8Hm-A	.	Hydra vulgaris	R2	5,1,3,4,2
R8Hm-B	.	Hydra vulgaris	R2	1,43,3,44,4,2
R9Av	.	Adineta vaga	R2	1,3,4,2
Racla	.	Rana clamitans	SINE2/tPHK	
RandI-1	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,11,4,6,2
RandI-1_ACas	AEYA01000779	Acanthamoeba castellanii	RandI	7,1,3,10,11,4,6,2,13
RandI-1_VC	.	Volvox carteri	RandI	6,2,7,33,10,11,13,1,32,4,237,17,3

RandI-2	.	Chlamydomonas reinhardtii	RandI	6,2,10,24,11,13,1,4,16,17,3,7
RandI-2_ACas	AEYA01000457	Acanthamoeba castellanii	RandI	7,1,3,10,11,4,6,2
RandI-3	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,11,4,6,2,13
RandI-3_ACas	AEYA01001211	Acanthamoeba castellanii	RandI	7,1,17,3,10,4,6,2,13
RandI-4	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,4,11,6,2
RandI-4_ACas	AEYA01002376	Acanthamoeba castellanii	RandI	7,1,3,10,4,6,2,13
RandI-5	.	Chlamydomonas reinhardtii	RandI	
RandI-5_ACas	AEYA01000341	Acanthamoeba castellanii	RandI	7,1,17,3,10,4,2,13
RandI-6	.	Chlamydomonas reinhardtii	RandI	7,1,32,17,3,10,33,4,11,6,2,13
RaR2	FJ461304	Rhynchosciara americana	R2	18,5,1,19,3,4,2
Ras1	.	Rasbora pauciperforata	SINE2/тРНК	
RAVLIN_MT	.	Medicago truncatula	L1	1,3,4,2
RAVLIN_MT	.	Medicago truncatula	L1	1,3,4,2
RAVLIN2_MT	.	Medicago truncatula	L1	43,78,79,44,10,80,81,77,75,17,76,7
RAVLIN2_MT	.	Medicago truncatula	L1	43,78,79,44,10,80,81,77,75,17,76,7
RAVLIN5_MT	AC152406	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
RAVLIN5_MT	AC152406	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
RDRE1_RN	K03501	Rattus norvegicus	SINE1/7SL	
REP-1_ED	.	Entamoeba dispar	SINE	
REP2-2_TT	AY371729	Tetrahymena thermophila	Не содержащий LTR ретротранспозон	
REP6_TT	AY371731	Tetrahymena thermophila	Не содержащий LTR ретротранспозон	
REP6_XT	.	Xenopus tropicalis	Penelope	
REX1_CY	.	Poeciliidae	Rex1	

REX1_FS .	Fundulus sp. Laguna de LabradoRex1		
REX1_FURC .	Takifugu rubripes CR1		
Rex1-1_ACar .	Anolis carolinensis	Rex1	1,40,3,4,41,2
REX1-1_AFC .	Cichlidae	Rex1	1,40,3,11,41,4,2
REX1-1_BF .	Branchiostoma floridae	Rex1	1,3,11,4,2
Rex1-1_CCa .	Cyprinus carpio	Rex1	1,40,3,4,41,11,2
Rex1-1_CTe .	Capitella teleta	Rex1	1,40,3,4,41,2
REX1-1_DR .	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-1_EL .	Esox lucius	Rex1	1,40,3,4,41,11,2
Rex1-1_FoCa .	Folsomia candida	Rex1	1,3,11,2
Rex1-1_FR .	Takifugu rubripes	Rex1	1,40,3,4,41,2
REX1-1_GA .	Gasterosteus aculeatus	Rex1	1,40,3,11,41,4,2
Rex1-1_HRo .	Helobdella robusta	Rex1	1,3,11,4,2
Rex1-1_OL .	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-1_PM .	Petromyzon marinus	Rex1	1,40,3,11,41,4,2
Rex1-1_SP .	Strongylocentrotus purpuratus	Rex1	16,1,3,24,11,4,2
Rex1-1_SSa .	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-1_XL .	Xenopus laevis	Rex1	1,40,3,11,41,4,2
REX1-1_XT .	Xenopus tropicalis	Rex1	1,40,3,4,41,2
Rex1-10_SSa .	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-11_SSa .	Salmo salar	Rex1	1,40,3,4,41,2
Rex1-15_DRe .	Danio rerio	Rex1	
Rex1-19_DRe .	Danio rerio	Rex1	1,3,4,2

Rex1-1B_EL	Esox lucius	Rex1	
Rex1-1B_FR	Takifugu rubripes	Rex1	1,40,3,4,41,2
REX1-2_AFC	Cichlidae	Rex1	1,40,3,11,41,4,2
Rex1-2_CTe	Capitella teleta	Rex1	1,40,3,4,41,2
REX1-2_DR	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-2_FoCa	Folsomia candida	Rex1	1,3,11,2
Rex1-2_FR	Takifugu rubripes	Rex1	1,40,3,4,41,2
Rex1-2_GA	Gasterosteus aculeatus	Rex1	1,40,3,11,41,4,2
Rex1-2_OL	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-2_PM	Petromyzon marinus	Rex1	1,40,3,4,41,11,2
Rex1-2_SSa	Salmo salar	Rex1	1,40,3,4,41,11,2
REX1-2_XT	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-20_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-21_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-21_NV	Nematostella vectensis	CR1	41,40
Rex1-23_DRe	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-23_NV	Nematostella vectensis	CR1	1,40,3,11,41,4,2
Rex1-24_NV	Nematostella vectensis	Rex1	1,40,3,4,41,2
Rex1-25_NVe	Nematostella vectensis	Rex1	1,3,11,4,2
Rex1-26_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-28_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-29_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
REX1-3_AFC	Cichlidae	Rex1	1,40,3,11,41,2

Rex1-3_CTe	Capitella teleta	Rex1	1,40,3,11,41,4,2
REX1-3_DR	Danio rerio	Rex1	1,3,4,2
Rex1-3_EL	Esox lucius	Rex1	1,40,3,4,41,2
Rex1-3_FR	Takifugu rubripes	Rex1	1,40,3,11,41,4,2
Rex1-3_GA	Gasterosteus aculeatus	Rex1	1,40,3,4,41,2
Rex1-3_OL	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-3_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-3_XT	Xenopus tropicalis	Rex1	1,40,3,4,41,11,2
Rex1-30_DRe	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-31_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-38_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-3B_SSa	Salmo salar	Rex1	1,40,3,4,41,11,2
Rex1-3C_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-3D_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-3E_SSa	Salmo salar	Rex1	
REX1-4_AFC	Cichlidae	Rex1	40,41
REX1-4_DR	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-4_EL	Esox lucius	Rex1	1,40,3,11,41,4,2
Rex1-4_OL	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-4_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-4_XT	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-40_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-40B3_DR	Danio rerio	Rex1	

Rex1-43_DRe	.	Danio rerio	Rex1	1,40,3,41,2
Rex1-5_AFC	.	Cichlidae	Rex1	40,41
REX1-5_DR	.	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-5_EL	.	Esox lucius	Rex1	1,40,3,11,41,4,2
Rex1-5_OL	.	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-5_SSa	.	Salmo salar	Rex1	1,3,11,4,2
REX1-5_XT	.	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-50_DR	.	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-52_DR	.	Danio rerio	Rex1	
Rex1-53_DR	.	Danio rerio	Rex1	
Rex1-54_DR	.	Danio rerio	Rex1	
Rex1-55_DR	.	Danio rerio	Rex1	
Rex1-56_DR	.	Danio rerio	Rex1	1,40,3,41,2
Rex1-57_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-58_DR	.	Danio rerio	Rex1	
Rex1-59_DR	.	Danio rerio	Rex1	
Rex1-6_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-6_OL	.	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-6_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-6_XT	.	Xenopus tropicalis	Rex1	
Rex1-60_DR	.	Danio rerio	Rex1	1,40,3,41,2
Rex1-61_DR	.	Danio rerio	Rex1	
REX1-7_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2

Rex1-7_OL	.	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-7_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-7_XT	.	Xenopus tropicalis	Rex1	
REX1-8_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-8_SSa	.	Salmo salar	Rex1	1,40,3,4,41,2
REX1-8_XT	.	Xenopus tropicalis	Rex1	
REX1-9_DR	.	Danio rerio	Rex1	
Rex1-9_SSa	.	Salmo salar	Rex1	1,40,113,3,4,41,2
REX1-9_XT	.	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
REX3	AF125983	Xiphophorus maculatus	RTE	1,3,4,2
REX5	AY228504	Xiphophorus maculatus	L2	1,3,11,4,2
Rex6	.	Takifugu rubripes	R4	1,3,4,2
Rex6-1_OL	.	Oryzias latipes	R4	1,3,4,2
Rhin-1	AC150242.3, AC152492	Rhinolophus ferrumequinum	SINE2/тРНК	
RILN1_OA	AB016096	Oryza australiensis	L1	
RILN3_OA	AB016098	Oryza australiensis	L1	
RILN4_OS	AB016099	Oryza sativa	L1	
RILN6_OS	AB016101	Oryza sativa	L1	
RILN7_OS	AB016102	Oryza sativa	L1	
RILN8_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3
RMER1B	.	Murinae	Не содержащий LTR ретротранспозон	
RNALUIII	M13501	Rattus norvegicus	SINE1/7SL	
RSG1	M37214	Oncorhynchus mykiss	SINE	

RSINE1	.	Rodentia	SINE2/тРНК
RSINE2	.	Rodentia	SINE2/тРНК
RSINE2A	.	Rodentia	SINE2/тРНК
RT_BC	AJ430549	Brassica carinata	Не содержащий LTR ретротранспозон
RT1	M93690	Anopheles gambiae	R1 2,245,8,11,86,12,246,1,14,4,15,16,3
RT2	M93691	Anopheles gambiae	R1 2,8,11,12,1,14,4,243,3
RTAg3	AB090812	Anopheles gambiae	R1 2,8,11,86,12,42,1,14,4,15,16,3
RTAg4	AB090813	Anopheles gambiae	R1 43,2,8,44,11,86,12,42,1,14,4,15,16,3
RTE_BOV_B_AC_1	.	Anolis carolinensis	RTE 16,1,26,3,4,11,2,15
RTE_Ele2	.	Aedes aegypti	RTE 1,26,3,4,11,2
RTE_Ele2B_AAe	.	Aedes aegypti	RTE 1,26,3,11,4,2
RTE_Ele2C_AAe	.	Aedes aegypti	RTE 1,26,3,11,4,2
RTE_Ele3	.	Aedes aegypti	RTE 16,1,26,3,11,4,2,15
RTE_Ele4	.	Aedes aegypti	RTE 1,26,3,4,11,2
RTE_Ele5	.	Aedes aegypti	RTE 1,26,3,11,4,2
RTE_SJ	AY027869	Schistosoma japonicum	RTE
RTE-1	AF054983	Caenorhabditis elegans	RTE 16,1,26,3,24,11,4,2
RTE-1_AC	.	Aplysia californica	RTE 16,1,26,3,24,11,4,2
RTE-1_AC_1	.	Anolis carolinensis	RTE 18,5,1,19,3,11,4,2
RTE-1_AFC	.	Cichlidae	RTE 16,1,26,3,11,4,2,15
RTE-1_AG	.	Anopheles gambiae	RTE 16,1,26,3,24,11,4,2
RTE-1_AIp	.	Arachis ipaensis	RTE 65,1,26,3,11,4,2
RTE-1_ATr	.	Amborella trichopoda	RTE 1,3,11,4,2

RTE-1_BDi	Brachypodium distachyon	RTE	1,3,4,2
RTE-1_BF	Branchiostoma floridae	RTE	1,127,3,59,60,4,56,57,2,68
RTE-1_BM	Bombyx mori	RTE	1,3,4,2
RTE-1_BTa	Bemisia tabaci	RTE	1,26,3,4,11,2
RTE-1_CBre	Caenorhabditis brenneri	RTE	1,26,3,11,4,2
RTE-1_CJap	Caenorhabditis japonica	RTE	16,1,26,3,11,4,2,15
RTE-1_CPB	Chrysemys picta bellii	RTE	18,5,16,1,26,3,11,4,2,15
RTE-1_CQ	Culex quinquefasciatus	RTE	1,26,3,4,11,2
RTE-1_CRem	Caenorhabditis remanei	RTE	1,26,3,11,4,2
RTE-1_Crp	Crocodylus porosus	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-1_CTe	Capitella teleta	RTE	1,3,11,4,2
RTE-1_DaCa	Daucus carota	RTE	1,3,11,4,2
RTE-1_DAn	Drosophila ananassae	RTE	1,26,3,4,11,2
RTE-1_DBi	Drosophila biarmipes	RTE	1,3,4,2
RTE-1_DBp	Drosophila bipectinata	RTE	1,26,3,11,4,2
RTE-1_DCa	Dendrobium catenatum	RTE	65,1,26,3,11,4,2
RTE-1_DF	Drosophila ficusphila	RTE	1,3,11,4,2
RTE-1_DPl	Danaus plexippus	RTE	1,26,3,11,4,2
RTE-1_DR	Danio rerio	RTE	16,1,26,3,11,4,2,15
RTE-1_DSa	Drosophila santomea	RTE	1,26,3,11,4,2
RTE-1_DSer	Drosophila serrata	RTE	1,26,3,4,11,2
RTE-1_DT	Drosophila takahashii	RTE	1,26,3,11,4,2
RTE-1_DYa	Drosophila yakuba	RTE	1,3,4,2

RTE-1_EC	.	Equus caballus	RTE	1,3,11,4,2
RTE-1_EL	.	Esox lucius	RTE	16,1,26,49,3,11,4,50,2,15
RTE-1_GA	.	Gasterosteus aculeatus	RTE	16,1,114,26,3,4,115,11,2,15
RTE-1_Gav	.	Gavialis gangeticus	RTE	18,5,1,19,3,11,4,2
RTE-1_GM	.	Glycine max	RTE	65,1,26,3,11,4,2
RTE-1_HMe	.	Heliconius melpomene melpomene	RTE	16,1,26,3,4,11,2,15
RTE-1_HRo	.	Helobdella robusta	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-1_JC	.	Jatropha curcas	RTE	
RTE-1_LVa	.	Litopenaeus vannamei	RTE	1,26,3,11,4,2
RTE-1_MAc	.	Musa acuminata	RTE	65,1,26,3,11,4,2
RTE-1_Mad	.	Malus domestica	RTE	65,1,26,3,11,4,2
RTE-1_MD	.	Monodelphis domestica	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-1_MLe	.	Mnemiopsis leidyi	RTE	1,3,11,4,2
RTE-1_MMa	.	Mesobuthus martensii	RTE	1,3,4,2
RTE-1_MN	.	Morus notabilis	RTE	65,1,26,3,11,4,2
RTE-1_NN	.	Nelumbo nucifera	RTE	1,3,4,2
RTE-1_NS	.	Nicotiana sylvestris	RTE	1,26,3,11,4,2
RTE-1_NV	.	Nematostella vectensis	RTE	
RTE-1_NVi	.	Nasonia vitripennis	RTE	1,3,4,2
RTE-1_OAf	.	Orycteropus afer	RTE	16,1,26,3,4,11,2,15
RTE-1_OL	.	Oryzias latipes	RTE	1,26,16,49,3,11,4,50,2,15
RTE-1_PBa	.	Pogonomyrmex barbatus	RTE	
RTE-1_PD	.	Phoenix dactylifera	RTE	1,3,4,2

RTE-1_PH	.	Parhyale hawaiensis	RTE	16,1,26,3,11,4,2,15
RTE-1_PM	.	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-1_PPac	.	Pristionchus pacificus	RTE	16,1,26,3,24,4,11,2
RTE-1_PPo	.	Papilio polytes	RTE	1,3,2
RTE-1_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,4,11,2,15
RTE-1_PXu	.	Papilio xuthus	RTE	
RTE-1_RC	.	Ricinus communis	RTE	
RTE-1_SBi	.	Sorghum bicolor	RTE	1,3,4,2
RTE-1_SIn	.	Solenopsis invicta	RTE	16,1,3,4,11,2,15
RTE-1_SIt	.	Setaria italica	RTE	65,16,1,26,3,24,11,4,2
RTE-1_SK	.	Saccoglossus kowalevskii	RTE	1,3,59,60,4,2
RTE-1_SSa	.	Salmo salar	RTE	2,18,24,11,1,26,4,65,5,16,19,3
RTE-1_STu	.	Solanum tuberosum	RTE	1,26,3,11,4,2
RTE-1_TD	TREP1820	Triticum turgidum subsp. durum	RTE	65,16,1,26,3,24,11,4,2
RTE-1_TML	.	Trichechus manatus latirostris	RTE	16,1,26,3,24,11,4,2
RTE-1_TP	.	Thalassiosira pseudonana	RTE	5,1,3,11,2
RTE-10_AMi	.	Alligator mississippiensis	RTE	1,26,3,11,4,2
RTE-10_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-10_CPB	.	Chrysemys picta bellii	RTE	
RTE-10_Crp	.	Crocodylus porosus	RTE	
RTE-10_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-10_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-10_PPac	.	Pristionchus pacificus	RTE	16,1,3,11,4,2,15

RTE-10_PPo	Papilio polytes	RTE	1,3,4,2
RTE-10_PSi	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-10_PXu	Papilio xuthus	RTE	
RTE-10_SP	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-10B_CPB	Chrysemys picta bellii	RTE	16,1,26,3,4,11,2,15
RTE-11_AMi	Alligator mississippiensis	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-11_BF	Branchiostoma floridae	RTE	1,26,16,3,11,4,2,15
RTE-11_CPB	Chrysemys picta bellii	RTE	1,26,3,11,4,2
RTE-11_Crp	Crocodylus porosus	RTE	
RTE-11_Hmel	Heliconius melpomene	RTE	1,3,4,2
RTE-11_HMM	Heliconius melpomene melpomene	RTE	16,1,26,3,11,4,2,15
RTE-11_LCh	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-11_LMi	Locusta migratoria	RTE	1,3,4,2
RTE-11_PPac	Pristionchus pacificus	RTE	1,3,4,2
RTE-11_PPo	Papilio polytes	RTE	
RTE-11_PSi	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,24,4,11,2
RTE-11_PXu	Papilio xuthus	RTE	
RTE-11_SP	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-12_AMi	Alligator mississippiensis	RTE	
RTE-12_BF	Branchiostoma floridae	RTE	16,1,26,3,24,11,4,2
RTE-12_CPB	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-12_Crp	Crocodylus porosus	RTE	16,1,26,3,24,11,4,2
RTE-12_Hmel	Heliconius melpomene	RTE	1,3,2

RTE-12_HMM	.	Heliconius melpomene melpomene	RTE 1,3,4,2
RTE-12_LCh	.	Latimeria chalumnae	RTE 18,5,16,1,19,3,11,4,2,15
RTE-12_LMi	.	Locusta migratoria	RTE 1,26,3,11,4,2
RTE-12_PPac	.	Pristionchus pacificus	RTE 16,1,26,3,24,11,4,2
RTE-12_PPo	.	Papilio polytes	RTE
RTE-12_PSi	.	Pelodiscus sinensis	RTE 65,16,1,26,3,24,11,4,2
RTE-12_PXu	.	Papilio xuthus	RTE
RTE-12_SP	.	Strongylocentrotus purpuratus	RTE 1,3,4,2
RTE-13_AMi	.	Alligator mississippiensis	RTE
RTE-13_BF	.	Branchiostoma floridae	RTE 16,1,26,3,4,11,2,15
RTE-13_CPB	.	Chrysemys picta bellii	RTE 65,16,1,26,3,11,4,2,15
RTE-13_Hmel	.	Heliconius melpomene	RTE 1,3,4,2
RTE-13_Lch	.	Latimeria chalumnae	RTE 65,1,16,26,3,4,11,2,15
RTE-13_LMi	.	Locusta migratoria	RTE 1,3,172,173,4,2
RTE-13_PPac	.	Pristionchus pacificus	RTE 170,1,3,4,171,2
RTE-13_PPo	.	Papilio polytes	RTE
RTE-13_PSi	.	Pelodiscus sinensis	RTE 65,16,1,26,3,11,4,2,15
RTE-13_PXu	.	Papilio xuthus	RTE
RTE-13_SP	.	Strongylocentrotus purpuratus	RTE 65,16,26,11,15
RTE-14_BF	.	Branchiostoma floridae	RTE 1,26,3,4,11,2
RTE-14_CPB	.	Chrysemys picta bellii	RTE 16,1,26,3,4,11,2,15
RTE-14_Hmel	.	Heliconius melpomene	RTE 1,3,2
RTE-14_Lch	.	Latimeria chalumnae	RTE 18,5,16,19,1,26,3,11,4,2,15

RTE-14_LMi .	Locusta migratoria	RTE	16,1,26,3,4,11,2,15
RTE-14_PPac .	Pristionchus pacificus	RTE	1,3,11,4,2
RTE-14_PPo .	Papilio polytes	RTE	1,3,4,2
RTE-14_PSi .	Pelodiscus sinensis	RTE	5,16,1,19,26,3,4,11,2,15
RTE-14_SP .	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-15_BF .	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-15_CPB .	Chrysemys picta bellii	RTE	
RTE-15_Hmel.	Heliconius melpomene	RTE	1,3,4,2
RTE-15_LMi .	Locusta migratoria	RTE	1,3,4,2
RTE-15_PPo .	Papilio polytes	RTE	1,3,4,2
RTE-15_PSi .	Pelodiscus sinensis	RTE	1,3,11,4,2
RTE-15_PXu .	Papilio xuthus	RTE	
RTE-15_SP .	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-16_BF .	Branchiostoma floridae	RTE	1,26,3,11,4,2
RTE-16_CPB .	Chrysemys picta bellii	RTE	
RTE-16_Hmel.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-16_LMi .	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-16_PPo .	Papilio polytes	RTE	
RTE-16_PSi .	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-16_SP .	Strongylocentrotus purpuratus	RTE	1,3,2
RTE-17_BF .	Branchiostoma floridae	RTE	5,1,19,4,2
RTE-17_CPB .	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-17_Hmel.	Heliconius melpomene	RTE	18,5,19

RTE-17_LMi .	Locusta migratoria	RTE	16,1,26,3,24,11,4,2
RTE-17_PPo .	Papilio polytes	RTE	
RTE-17_PSi .	Pelodiscus sinensis	RTE	1,26,16,3,11,4,2,15
RTE-17B_PPo .	Papilio polytes	RTE	
RTE-18_BF .	Branchiostoma floridae	RTE	1,3,11,4,2
RTE-18_CPB .	Chrysemys picta bellii	RTE	
RTE-18_Hmel.	Heliconius melpomene	RTE	65,16,1,26,3,4,11,2,15
RTE-18_HMM .	Heliconius melpomene melpomene	RTE	1,3,2
RTE-18_LMi .	Locusta migratoria	RTE	1,3,11,4,2
RTE-18_PPo .	Papilio polytes	RTE	
RTE-18_SP .	Strongylocentrotus purpuratus	RTE	1,26,3,11,4,2
RTE-18B_PPo .	Papilio polytes	RTE	
RTE-19_BF .	Branchiostoma floridae	RTE	1,3,4,2
RTE-19_CPB .	Chrysemys picta bellii	RTE	
RTE-19_Hmel.	Heliconius melpomene	RTE	1,3,4,2
RTE-19_LMi .	Locusta migratoria	RTE	1,3,4,2
RTE-19_PPo .	Papilio polytes	RTE	16,1,26,3,11,4,2,15
RTE-1B_DAn .	Drosophila ananassae	RTE	1,26,3,11,4,2
RTE-1B_DCa .	Dendrobium catenatum	RTE	1,3,4,2
RTE-1B_DSa .	Drosophila santomea	RTE	26,11
RTE-1B_LMi .	Locusta migratoria	RTE	1,3,4,2
RTE-1B_Mad .	Malus domestica	RTE	65,1,26,3,11,4,2
RTE-1B_PPo .	Papilio polytes	RTE	

RTE-1BN1_DAn	.	Drosophila ananassae	RTE
RTE-1C_DAn	.	Drosophila ananassae	RTE 1,26,16,3,11,4,2,15
RTE-1C_LMi	.	Locusta migratoria	RTE
RTE-2_AC	.	Aplysia californica	RTE 1,26,16,3,24,11,4,2
RTE-2_AFC	.	Cichlidae	RTE 1,26,3,11,4,2
RTE-2_AG	.	Anopheles gambiae	RTE 1,26,3,4,11,2
RTE-2_BF	.	Branchiostoma floridae	RTE 16,1,26,3,4,11,2,15
RTE-2_BM	.	Bombyx mori	RTE
RTE-2_BTa	.	Bemisia tabaci	RTE 1,26,3,11,4,2
RTE-2_CBre	.	Caenorhabditis brenneri	RTE 16,1,26,3,4,11,2,15
RTE-2_CJap	.	Caenorhabditis japonica	RTE 1,26,3,11,4,2
RTE-2_CPB	.	Chrysemys picta bellii	RTE 18,5,16,19,1,26,3,24,4,11,2
RTE-2_CQ	.	Culex quinquefasciatus	RTE 16,1,26,3,24,11,4,2
RTE-2_Croc	.	Crocodylidae	RTE 18,5,1,19,3,11,4,2
RTE-2_CTe	.	Capitella teleta	RTE 18,5,16,19,1,26,3,11,4,2,15
RTE-2_DAn	.	Drosophila ananassae	RTE 1,26,3,11,4,2
RTE-2_DBp	.	Drosophila bipectinata	RTE 1,26,3,11,4,2
RTE-2_DF	.	Drosophila ficusphila	RTE 1,26,3,11,4,2
RTE-2_DPl	.	Danaus plexippus	RTE 1,26,3,11,4,2
RTE-2_DR	.	Danio rerio	RTE 1,2
RTE-2_EL	.	Esox lucius	RTE 1,49,3,4,50,2
RTE-2_Gav	.	Gavialis gangeticus	RTE 1,3,4,2
RTE-2_Hmel	.	Heliconius melpomene	RTE 16,1,26,3,11,4,2,15

RTE-2_HRo	Helobdella robusta	RTE	2,11,57,68,1,26,4,3,59,60
RTE-2_LCh	Latimeria chalumnae	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-2_LMi	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-2_LVa	Litopenaeus vannamei	RTE	16,1,26,3,11,4,2,15
RTE-2_MD	Monodelphis domestica	RTE	1,3,4,2
RTE-2_ME	Notamacropus eugenii	RTE	1,26,3,11,4,2
RTE-2_MLe	Mnemiopsis leidyi	RTE	1,3,4,2
RTE-2_MMa	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-2_NS	Nicotiana sylvestris	RTE	1,3,11,4,2
RTE-2_NV	Nematostella vectensis	RTE	
RTE-2_NVi	Nasonia vitripennis	RTE	16,1,3,11,15,4,2
RTE-2_OL	Oryzias latipes	RTE	1,26,3,11,4,2
RTE-2_PD	Phoenix dactylifera	RTE	
RTE-2_PH	Parhyale hawaiensis	RTE	65,16,1,26,3,24,11,4,2
RTE-2_PM	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-2_PPac	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-2_PPo	Papilio polytes	RTE	
RTE-2_PSi	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-2_PXu	Papilio xuthus	RTE	
RTE-2_SIn	Solenopsis invicta	RTE	1,3,4,11,2
RTE-2_SK	Saccoglossus kowalevskii	RTE	65,16,1,26,3,11,4,2,15
RTE-2_SP	Strongylocentrotus purpuratus	RTE	65,16,1,26,3,11,4,2,15
RTE-2_STu	Solanum tuberosum	RTE	

RTE-20_CPB . Chrysemys picta bellii RTE

RTE-20_Hmel. Heliconius melpomene RTE 16,1,26,3,11,4,2,15

RTE-20_HMM . Heliconius melpomene melpomeneRTE 1,3,2

RTE-20_LMi . Locusta migratoria RTE 1,3,11,4,2

RTE-20_PPo . Papilio polytes RTE

RTE-21_Croc . Crocodylidae RTE 5,16,1,19,26,3,11,4,2,15

RTE-21_LMi . Locusta migratoria RTE 1,26,3,11,4,2

RTE-21_PPo . Papilio polytes RTE 1,3,11,4,2

RTE-22_LMi . Locusta migratoria RTE 16,1,26,3,11,4,2,15

RTE-22_PPo . Papilio polytes RTE

RTE-22B_PPo . Papilio polytes RTE

RTE-23_CPB . Chrysemys picta bellii RTE 18,66,16,5,19,1,26,67,3,11,4,2,15

RTE-23_LMi . Locusta migratoria RTE 16,1,3,11,4,2,15

RTE-24_CPB . Chrysemys picta bellii RTE

RTE-24_LMi . Locusta migratoria RTE 16,1,26,3,11,4,2,15

RTE-24_PPo . Papilio polytes RTE

RTE-25_CPB . Chrysemys picta bellii RTE

RTE-25_LMi . Locusta migratoria RTE 16,1,26,3,11,4,2,15

RTE-25_PPo . Papilio polytes RTE

RTE-25B_PPo . Papilio polytes RTE

RTE-26_CPB . Chrysemys picta bellii RTE

RTE-26_LMi . Locusta migratoria RTE 16,1,26,3,11,4,2,15

RTE-26_PPo . Papilio polytes RTE

RTE-27_CPB .	Chrysemys picta bellii	RTE	1,3,4,2
RTE-27_LMi .	Locusta migratoria	RTE	1,3,11,4,2
RTE-27_PPo .	Papilio polytes	RTE	
RTE-27B_CPB .	Chrysemys picta bellii	RTE	
RTE-28_LMi .	Locusta migratoria	RTE	1,3,11,4,2
RTE-28_PPo .	Papilio polytes	RTE	
RTE-29_CPB .	Chrysemys picta bellii	RTE	1,3,4,2
RTE-29_LMi .	Locusta migratoria	RTE	1,3,11,4,2
RTE-29_PPo .	Papilio polytes	RTE	
RTE-2B_PPo .	Papilio polytes	RTE	
RTE-3_AC .	Aplysia californica	RTE	16,1,26,3,4,11,2,15
RTE-3_AG .	Anopheles gambiae	RTE	1,3,4,2
RTE-3_AMi .	Alligator mississippiensis	RTE	18,5,16,1,19,3,11,4,2,15
RTE-3_BF .	Branchiostoma floridae	RTE	16,1,26,3,4,11,2,15
RTE-3_BM .	Bombyx mori	RTE	1,3,4,2
RTE-3_BTa .	Bemisia tabaci	RTE	16,1,26,3,24,11,4,2
RTE-3_CBre .	Caenorhabditis brenneri	RTE	
RTE-3_CPB .	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-3_CQ .	Culex quinquefasciatus	RTE	1,26,3,11,4,2
RTE-3_Crp .	Crocodylus porosus	RTE	18,5,16,19,1,26,3,24,4,11,2
RTE-3_CTe .	Capitella teleta	RTE	16,1,26,3,24,11,4,2
RTE-3_DF .	Drosophila ficusphila	RTE	1,26,3,11,4,2
RTE-3_DPl .	Danaus plexippus	RTE	16,1,26,3,11,4,2,15

RTE-3_DR	.	Danio rerio	RTE	
RTE-3_Hmel	.	Heliconius melpomene	RTE	1,26,3,11,4,2
RTE-3_HRo	.	Helobdella robusta	RTE	2,11,57,68,1,26,4,3,59,60
RTE-3_LCh	.	Latimeria chalumnae	RTE	18,5,16,1,26,3,24,11,4,2
RTE-3_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-3_LVa	.	Litopenaeus vannamei	RTE	1,26,3,11,4,2
RTE-3_MD	.	Monodelphis domestica	RTE	
RTE-3_ME	.	Notamacropus eugenii	RTE	
RTE-3_MMa	.	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-3_NS	.	Nicotiana sylvestris	RTE	
RTE-3_NV	.	Nematostella vectensis	RTE	
RTE-3_OL	.	Oryzias latipes	RTE	1,26,3,11,4,2
RTE-3_PM	.	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-3_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-3_PSi	.	Pelodiscus sinensis	RTE	
RTE-3_PXu	.	Papilio xuthus	RTE	
RTE-3_SIn	.	Solenopsis invicta	RTE	16,1,3,11,4,2,15
RTE-3_SP	048B05	Strongylocentrotus purpuratus	RTE	1,3,11,4,2
RTE-30_CPB	.	Chrysemys picta bellii	RTE	
RTE-30_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-31_CPB	.	Chrysemys picta bellii	RTE	
RTE-31_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-31_PPo	.	Papilio polytes	RTE	

RTE-32_CPB	.	Chrysemys picta bellii	RTE	
RTE-32_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-32_PPo	.	Papilio polytes	RTE	
RTE-33_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-33_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-33_PPo	.	Papilio polytes	RTE	
RTE-34_CPB	.	Chrysemys picta bellii	RTE	1,3,11,4,2
RTE-34_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-34_PPo	.	Papilio polytes	RTE	
RTE-35_CPB	.	Chrysemys picta bellii	RTE	
RTE-35_PPo	.	Papilio polytes	RTE	
RTE-36_CPB	.	Chrysemys picta bellii	RTE	
RTE-36_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-36_PPo	.	Papilio polytes	RTE	
RTE-37_CPB	.	Chrysemys picta bellii	RTE	
RTE-37_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-37_PPo	.	Papilio polytes	RTE	
RTE-38_LMi	.	Locusta migratoria	RTE	1,3,2
RTE-38_PPo	.	Papilio polytes	RTE	
RTE-39_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-3B_SP	048B05	Strongylocentrotus purpuratus	RTE	1,26,3,11,4,2
RTE-4_AC	.	Aplysia californica	RTE	1,26,3,11,4,2
RTE-4_AMi	.	Crocodylidae	RTE	16,1,26,3,11,4,2,15

RTE-4_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-4_BM	.	Bombyx mori	RTE	1,3,11,4,2
RTE-4_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-4_Croc	.	Crocodylidae	RTE	
RTE-4_CTe	.	Capitella teleta	RTE	1,3,4,2
RTE-4_DF	.	Drosophila ficusphila	RTE	1,26,3,11,4,2
RTE-4_DPl	.	Danaus plexippus	RTE	65,16,1,26,3,11,4,2,15
RTE-4_DR	.	Danio rerio	RTE	
RTE-4_Hmel	.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-4_HRo	.	Helobdella robusta	RTE	53,2,54,11,57,68,1,4,3,59,60
RTE-4_LCh	.	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-4_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-4_LVa	.	Litopenaeus vannamei	RTE	1,3,4,2
RTE-4_MMa	.	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-4_NV	.	Nematostella vectensis	RTE	
RTE-4_PM	.	Petromyzon marinus	RTE	1,16,26,3,11,4,2,15
RTE-4_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-4_PPo	.	Papilio polytes	RTE	16,1,26,3,11,4,2,15
RTE-4_PSi	.	Pelodiscus sinensis	RTE	65,1,26,3,11,4,2
RTE-4_PXu	.	Papilio xuthus	RTE	
RTE-4_SP	048B05	Strongylocentrotus purpuratus	RTE	1,3,11,4,2
RTE-40_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-41_LMi	.	Locusta migratoria	RTE	16,1,3,11,4,2,15

RTE-42_LMi .	Locusta migratoria	RTE	65,16,1,26,3,4,11,2,15
RTE-43_LMi .	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-44_LMi .	Locusta migratoria	RTE	1,3,11,4,2
RTE-45_LMi .	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-4562_AMi .	Crocodylidae	RTE	
RTE-46_LMi .	Locusta migratoria	RTE	16,1,26,3,4,11,2,15
RTE-47_LMi .	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-48_LMi .	Locusta migratoria	RTE	
RTE-49_LMi .	Locusta migratoria	RTE	
RTE-4B_PPo .	Papilio polytes	RTE	
RTE-5_AC .	Aplysia californica	RTE	65,16,1,26,3,24,11,4,2
RTE-5_AMi .	Alligator mississippiensis	RTE	16,1,3,11,4,2,15
RTE-5_BF .	Branchiostoma floridae	RTE	1,3,4,2
RTE-5_BM .	Bombyx mori	RTE	65,1,26,11,2
RTE-5_CPB .	Chrysemys picta bellii	RTE	1,26,3,11,4,2
RTE-5_CTe .	Capitella teleta	RTE	1,3,4,2
RTE-5_DPl .	Danaus plexippus	RTE	16,1,3,11,4,2,15
RTE-5_Hmel .	Heliconius melpomene	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-5_HRo .	Helobdella robusta	RTE	16,1,26,3,24,11,4,2
RTE-5_LCh .	Latimeria chalumnae	RTE	18,5,1,3,11,4,2
RTE-5_LMi .	Locusta migratoria	RTE	65,16,1,26,3,11,4,2,15
RTE-5_LVa .	Litopenaeus vannamei	RTE	
RTE-5_MMa .	Mesobuthus martensii	RTE	65,16,1,26,3,4,11,2,15

RTE-5_NV	.	Nematostella vectensis	RTE	
RTE-5_PM	.	Petromyzon marinus	RTE	5,1,19,26,11,2
RTE-5_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-5_PPo	.	Papilio polytes	RTE	
RTE-5_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-5_PXu	.	Papilio xuthus	RTE	
RTE-5_SP	.	Strongylocentrotus purpuratus	RTEX	21,1,3,11,4,2
RTE-50_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-51_LMi	.	Locusta migratoria	RTE	65,1,3,4,11,2
RTE-52_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-53_LMi	.	Locusta migratoria	RTE	65,16,1,26,3,11,4,2,15
RTE-54_AMi	.	Crocodylidae	RTE	
RTE-54_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-55_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-56_LMi	.	Locusta migratoria	RTE	1,26,3,4,11,2
RTE-57_LMi	.	Locusta migratoria	RTE	
RTE-58_LMi	.	Locusta migratoria	RTE	
RTE-59_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-6_AC	.	Aplysia californica	RTE	16,1,26,3,11,4,2,15
RTE-6_AMi	.	Alligator mississippiensis	RTE	1,26,3,4,11,2
RTE-6_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-6_BM	.	Bombyx mori	RTE	1,3,4,2
RTE-6_CPB	.	Chrysemys picta bellii	RTE	18,5,16,19,1,26,3,11,4,2,15

RTE-6_Crp	.	Crocodylus porosus	RTE	
RTE-6_DPl	.	Danaus plexippus	RTE	1,11,2
RTE-6_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-6_HMM	.	Heliconius melpomene melpomene	RTE	16,1,26,3,11,4,2,15
RTE-6_HRo	.	Helobdella robusta	RTE	65,16,1,26,3,4,11,2,15
RTE-6_LCh	.	Latimeria chalumnae	RTE	1,26,3,11,4,2
RTE-6_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-6_LVa	.	Litopenaeus vannamei	RTE	
RTE-6_NV	.	Nematostella vectensis	RTE	
RTE-6_PM	.	Petromyzon marinus	RTE	1,3,2
RTE-6_PPac	.	Pristionchus pacificus	RTE	16,1,26,3,11,4,2,15
RTE-6_PPo	.	Papilio polytes	RTE	1,3,2
RTE-6_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-6_PXu	.	Papilio xuthus	RTE	
RTE-6_SP	.	Strongylocentrotus purpuratus	RTE	16,1,26,3,11,15,4,2
RTE-60_LMi	.	Locusta migratoria	RTE	100,16,1,3,24,4,11,2
RTE-61_LMi	.	Locusta migratoria	RTE	
RTE-62_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-63_LMi	.	Locusta migratoria	RTE	
RTE-64_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-65_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-66_LMi	.	Locusta migratoria	RTE	
RTE-67_LMi	.	Locusta migratoria	RTE	

RTE-68_LMi .	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-69_LMi .	Locusta migratoria	RTE	
RTE-6B_HRo .	Helobdella robusta	RTE	65,16,1,26,3,11,4,2,15
RTE-7_AC .	Aplysia californica	RTE	1,3,11,4,2
RTE-7_AMi .	Alligator mississippiensis	RTE	16,1,26,3,11,4,2,15
RTE-7_BF .	Branchiostoma floridae	RTE	
RTE-7_CPB .	Chrysemys picta bellii	RTE	18,5,1,19,3,4,2
RTE-7_Crp .	Crocodylus porosus	RTE	
RTE-7_DPl .	Danaus plexippus	RTE	1,3,4,2
RTE-7_Hmel .	Heliconius melpomene	RTE	16,1,3,11,4,2,15
RTE-7_HRo .	Helobdella robusta	RTE	1,3,4,2
RTE-7_LCh .	Latimeria chalumnae	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-7_LMi .	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-7_LVa .	Litopenaeus vannamei	RTE	1,3,4,11,2
RTE-7_NV .	Nematostella vectensis	RTE	65,16,1,26,3,11,4,2,15
RTE-7_PM .	Petromyzon marinus	RTE	16,1,3,11,4,2,15
RTE-7_PPac .	Pristionchus pacificus	RTE	1,3,4,2
RTE-7_PPo .	Papilio polytes	RTE	
RTE-7_PSi .	Pelodiscus sinensis	RTE	1,3,4,2
RTE-7_PXu .	Papilio xuthus	RTE	
RTE-7_SP .	Strongylocentrotus purpuratus	RTE	65,1,16,26,3,4,11,2,15
RTE-70_LMi .	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-71_LMi .	Locusta migratoria	RTE	1,3,11,4,2

RTE-8_AMi	Alligator mississippiensis	RTE	16,1,26,3,11,4,2,15
RTE-8_BF	Branchiostoma floridae	RTE	53,2,54,11,57,68,1,26,4,65,3,59,60
RTE-8_CPB	Chrysemys picta bellii	RTE	16,1,26,3,24,11,4,2
RTE-8_Crp	Crocodylus porosus	RTE	
RTE-8_Hmel	Heliconius melpomene	RTE	1,3,4,2
RTE-8_HRo	Helobdella robusta	RTE	18,5,1,19,3,11,4,2
RTE-8_LCh	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-8_LMi	Locusta migratoria	RTE	1,3,11,4,2
RTE-8_LVa	Litopenaeus vannamei	RTE	16,1,26,3,11,4,2,15
RTE-8_NV	Nematostella vectensis	RTE	
RTE-8_PM	Petromyzon marinus	RTE	18,5,16,1,19,3,11,4,2,15
RTE-8_PPac	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-8_PPo	Papilio polytes	RTE	
RTE-8_PSi	Pelodiscus sinensis	RTE	1,3,2
RTE-8_PXu	Papilio xuthus	RTE	
RTE-8_SP	Strongylocentrotus purpuratus	RTE	16,1,26,3,11,4,2,15
RTE-9_BF	Branchiostoma floridae	RTE	1,26,3,11,4,2
RTE-9_CPB	Chrysemys picta bellii	RTE	16,1,26,3,4,11,2,15
RTE-9_Crp	Crocodylus porosus	RTE	1,3,11,4,2
RTE-9_Hmel	Heliconius melpomene	RTE	1,3,4,2
RTE-9_LCh	Latimeria chalumnae	RTE	2,18,11,1,26,4,15,5,16,19,3
RTE-9_LMi	Locusta migratoria	RTE	16,1,26,3,24,11,4,2
RTE-9_LVa	Litopenaeus vannamei	RTE	1,3,11,4,2

RTE-9_NVe	Nematostella vectensis	RTE	1,3,4,2
RTE-9_PM	Petromyzon marinus	RTE	1,3,11,4,2
RTE-9_PPac	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-9_PPo	Papilio polytes	RTE	16,1,26,3,4,11,2,15
RTE-9_PSi	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-9_PXu	Papilio xuthus	RTE	
RTE-9_SP	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-N1_ATr	Amborella trichopoda	RTE	
RTE-N1_Crp	Crocodylus porosus	RTE	
RTE-N1_Lch	Latimeria chalumnae	RTE	
RTE-N1B_ATr	Amborella trichopoda	RTE	
RTE-N2_Lch	Latimeria chalumnae	RTE	
RTE0_Mars	Metatheria	RTE	
RTE1_ET	Echinops telfairi	RTE	1,3,11,4,2
RTE1_LA	Loxodonta africana	RTE	65,16,1,26,3,24,11,2
RTE1_MT	Medicago truncatula	RTE	1,3,4,2
RTE1_MT	Medicago truncatula	RTE	1,3,4,2
RTE1_Pca	Procavia capensis	RTE	16,1,26,3,24,11,4,2
RTE1_Sar	Sorex araneus	RTE	1,3,11,4,2
RTE1_ZM	Zea mays	RTE	1,3,11,4,2
RTE1-N1_LA	Loxodonta africana	RTE	
RTE1-N1b_LA	Loxodonta africana	RTE	
RTE1N_Sar	Sorex araneus	RTE	

RTE1X_SP	048B05	Strongylocentrotus purpuratus	RTE	16,1,26,3,24,11,4,2
RTE2_ZM	.	Zea mays	RTE	65,1,26,3,11,4,2
RTESINE1	.	Monodelphis	SINE	
RTESINE2	.	Metatheria	SINE	
RTETP-1_PT	chr_7	Phaeodactylum tricornutum	RTETP	
RTEX-1_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-1_AMi	.	Alligator mississippiensis	RTEX	1,3,11,4,2
RTEX-1_BF	.	Branchiostoma floridae	RTEX	62,2,63,24,11,1,64,4,16,3
RTEX-1_BTa	.	Bemisia tabaci	RTEX	21,16,1,3,4,11,2,15
RTEX-1_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-1_CoFl	.	Copidosoma floridanum	RTEX	1,3,11,4,2
RTEX-1_CPB	.	Chrysemys picta bellii	RTEX	
RTEX-1_CR	.	Chlamydomonas reinhardtii	RTEX	1,3,11,4,2
RTEX-1_Crp	.	Crocodylus porosus	RTEX	1,3,4,2
RTEX-1_DR	.	Danio rerio	RTEX	16,1,3,24,11,4,62,2
RTEX-1_DRh	.	Drosophila rhopaloa	RTEX	
RTEX-1_DT	.	Drosophila takahashii	RTEX	1,3,11,4,2
RTEX-1_EL	.	Esox lucius	RTEX	1,3,11,4,2
RTEX-1_FoCa	.	Folsomia candida	RTEX	16,1,3,11,4,2,15
RTEX-1_LG	.	Lottia gigantea	RTEX	1,3,11,4,2
RTEX-1_NV	.	Nematostella vectensis	RTEX	62,2,63,24,11,1,64,4,16,3
RTEX-1_PBa	.	Pogonomyrmex barbatus	RTEX	16,1,3,11,4,2,15
RTEX-1_RPr	.	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15

RTEX-1_SK	Saccoglossus kowalevskii	RTEX	2,63,11,1,64,4,15,16,3
RTEX-1_SP	Strongylocentrotus purpuratus	RTEX	1,3,11,4,2
RTEX-1_SSa	Salmo salar	RTEX	1,3,11,4,2
RTEX-1_VC	Volvox carteri	RTEX	1,3,82,11,4,2
RTEX-10_ACar	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-10_BF	Branchiostoma floridae	RTEX	53,62,2,63,54,11,56,57,48,68,1,64,58,4,127,3,59,60
RTEX-10_CGi	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-10_RPr	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15
RTEX-10_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-11_ACar	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-11_BF	Branchiostoma floridae	RTEX	1,109,3,110,11,4,62,2
RTEX-11_CGi	Crassostrea gigas	RTEX	
RTEX-11_RPr	Rhodnius prolixus	RTEX	21,11
RTEX-11_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-11A_BF	Branchiostoma floridae	RTEX	
RTEX-11B_ACar	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-12_ACar	Anolis carolinensis	RTEX	16,1,3,24,4,11,2
RTEX-12_BF	Branchiostoma floridae	RTEX	16,1,3,24,11,4,2
RTEX-12_CGi	Crassostrea gigas	RTEX	1,3,4,2
RTEX-12_RPr	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15
RTEX-12_SK	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-13_ACar	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-13_BF	Branchiostoma floridae	RTEX	16,1,3,11,4,2,15

RTEX-13_CGi	.	Crassostrea gigas RTEX	1,3,11,4,2
RTEX-13_RPr	.	Rhodnius prolixus RTEX	1,3,11,4,2
RTEX-13_SK	.	Saccoglossus kowalevskii RTEX	1,3,4,2
RTEX-14_ACar	.	Anolis carolinensis RTEX	16,1,3,24,11,4,2
RTEX-14_BF	.	Branchiostoma floridae RTEX	2,1,109,3,110,11,4,62
RTEX-14_CGi	.	Crassostrea gigas RTEX	21,1,3,11,4,2
RTEX-14_SK	.	Saccoglossus kowalevskii RTEX	1,3,11,4,62,2
RTEX-15_CGi	.	Crassostrea gigas RTEX	1,3,4,2
RTEX-15_SK	.	Saccoglossus kowalevskii RTEX	1,3,11,4,2
RTEX-16_BF	.	Branchiostoma floridae RTEX	16,1,3,24,11,4,2
RTEX-16_CGi	.	Crassostrea gigas RTEX	21,16,1,23,3,24,11,4,2
RTEX-16_SK	.	Saccoglossus kowalevskii RTEX	1,3,4,2
RTEX-17_BF	.	Branchiostoma floridae RTEX	16,1,3,11,4,2,15
RTEX-17_CGi	.	Crassostrea gigas RTEX	
RTEX-17_SK	.	Saccoglossus kowalevskii RTEX	16,1,3,24,11,4,2
RTEX-18_CGi	.	Crassostrea gigas RTEX	21,1,23,3,11,4,2
RTEX-18_SK	.	Saccoglossus kowalevskii RTEX	1,3,4,2
RTEX-19_CGi	.	Crassostrea gigas RTEX	21,1,23,3,11,4,2
RTEX-19_SK	.	Saccoglossus kowalevskii RTEX	16,1,3,24,11,4,2
RTEX-2_ACar	.	Anolis carolinensis RTEX	16,1,3,24,11,4,2
RTEX-2_BF	.	Branchiostoma floridae RTEX	189,190,62,2,191,192,193,11,194,1,4,15,16,3
RTEX-2_BTa	.	Bemisia tabaci RTEX	16,1,3,11,4,2,15
RTEX-2_CGi	.	Crassostrea gigas RTEX	1,3,4,2

RTEX-2_CoFl.	Copidosoma floridanum	RTEX	16,1,3,4,11,2,15
RTEX-2_DR .	Danio rerio	RTEX	1,3,4,2
RTEX-2_DT .	Drosophila takahashii	RTEX	1,3,11,4,2
RTEX-2_FoCa .	Folsomia candida	RTEX	16,1,3,11,4,2,15
RTEX-2_NV .	Nematostella vectensis	RTEX	62,16,1,3,24,11,4,2
RTEX-2_PBa .	Pogonomyrmex barbatus	RTEX	
RTEX-2_RPr .	Rhodnius prolixus	RTEX	1,3,4,2
RTEX-2_SK .	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-2_SSa .	Salmo salar	RTEX	2,16,1,3,24,11,4,62
RTEX-2_VC .	Volvox carteri	RTEX	1,3,11,4,2
RTEX-20_CGi .	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-20_SK .	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-21_CGi .	Crassostrea gigas	RTEX	1,23,3,4,11,2
RTEX-21_SK .	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-22_CGi .	Crassostrea gigas	RTEX	1,3,4,2
RTEX-22_SK .	Saccoglossus kowalevskii	RTEX	16,1,3,11,4,2,15
RTEX-23_CGi .	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-23_SK .	Saccoglossus kowalevskii	RTEX	16,21,1,3,24,11,4,2
RTEX-24_CGi .	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-24_SK .	Saccoglossus kowalevskii	RTEX	16,1,3,4,11,2,15
RTEX-25_CGi .	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-25_SK .	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-26_CGi .	Crassostrea gigas	RTEX	16,1,3,11,4,2,15

RTEX-26_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-27_CGi	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-27_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-28_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-3_ACar	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-3_BF	Branchiostoma floridae	RTEX	2,1,3,11,4,62
RTEX-3_CGi	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-3_DR	Danio rerio	RTEX	
RTEX-3_NV	Nematostella vectensis	RTEX	2,16,1,3,11,4,62,15
RTEX-3_RPr	Rhodnius prolixus	RTEX	1,3,4,2
RTEX-3_SK	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-3_SSa	Salmo salar	RTEX	62,1,3,11,4,2
RTEX-3_VC	Volvox carteri	RTEX	16,1,3,24,11,4,2
RTEX-4_ACar	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-4_BF	Branchiostoma floridae	RTEX	53,62,2,63,54,11,56,57,68,1,64,58,4,15,16,3,59,60
RTEX-4_CGi	Crassostrea gigas	RTEX	21,1,3,11,4,2
RTEX-4_DR	Danio rerio	RTEX	2,1,3,11,4,48,62
RTEX-4_NV	Nematostella vectensis	RTEX	16,1,3,4,11,48,62,2,15
RTEX-4_RPr	Rhodnius prolixus	RTEX	1,3,2
RTEX-4_SK	Saccoglossus kowalevskii	RTEX	1,3,11,4,62,2
RTEX-4_VC	Volvox carteri	RTEX	1,3,11,4,2
RTEX-5_ACar	Anolis carolinensis	RTEX	16,1,3,11,4,2,15
RTEX-5_BF	Branchiostoma floridae	RTEX	16,1,3,24,11,4,62,2

RTEX-5_CGi	Crassostrea gigas	RTEX	1,3,2
RTEX-5_DR	Danio rerio	RTEX	1,3,11,4,2
RTEX-5_NV	Nematostella vectensis	RTEX	
RTEX-5_RPr	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-5_SK	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-5_VC	Volvox carteri	RTEX	1,3,11,4,2
RTEX-6_ACar	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-6_BF	Branchiostoma floridae	RTEX	16,1,3,24,11,4,2
RTEX-6_CGi	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-6_NV	Nematostella vectensis	RTEX	
RTEX-6_RPr	Rhodnius prolixus	RTEX	2,21,23,11,1,4,15,16,3
RTEX-6_SK	Saccoglossus kowalevskii	RTEX	1,3,4,11,2
RTEX-7_ACar	Anolis carolinensis	RTEX	16,1,3,4,11,2,15
RTEX-7_BF	Branchiostoma floridae	RTEX	1,3,4,2
RTEX-7_CGi	Crassostrea gigas	RTEX	16,1,3,24,11,4,2
RTEX-7_RPr	Rhodnius prolixus	RTEX	21,1,3,11,4,2
RTEX-7_SK	Saccoglossus kowalevskii	RTEX	63,2,16,1,64,3,24,11,4,62
RTEX-8_ACar	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-8_BF	Branchiostoma floridae	RTEX	1,3,4,2
RTEX-8_CGi	Crassostrea gigas	RTEX	
RTEX-8_NV	Nematostella vectensis	RTEX	
RTEX-8_RPr	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-8_SK	Saccoglossus kowalevskii	RTEX	1,3,11,4,2

RTEX-9_ACar	.	Anolis carolinensis	RTEX 16,1,3,24,11,4,2
RTEX-9_BF	.	Branchiostoma floridae	RTEX 1,3,11,4,2
RTEX-9_CGi	.	Crassostrea gigas	RTEX 21,1,23,3,11,4,2
RTEX-9_RPr	.	Rhodnius prolixus	RTEX 1,3,11,4,2
RTEX-9_SK	.	Saccoglossus kowalevskii	RTEX 1,3,11,4,2
RTEY-1_CTe	.	Capitella teleta	RTEY 16,1,3,11,4,2,15
RTEY-2_CTe	.	Capitella teleta	RTEY 1,3,4,2
RTEY-3_CTe	.	Capitella teleta	RTEY 1,3,4,2
Ruka	.	Ixodida	SINE2/тРНК
S_Av	.	Adineta vaga	Penelope 1,3
S0_Av	.	Adineta vaga	Penelope 1,3
S1_BrN	L24035	Brassica napus	SINE2/тРНК
SacSINE1	.	Squalus acanthias	SINE2/тРНК
Sadhu-1_ALy	.	Arabidopsis lyrata	SINE
Sadhu-2_ALy	.	Arabidopsis lyrata	SINE
Sadhu-3_ALy	.	Arabidopsis lyrata	Не содержащий LTR ретротранспозон
Sadhu-4_ALy	.	Arabidopsis lyrata	Не содержащий LTR ретротранспозон
Sadhu1-1	NC_003071.3	Arabidopsis thaliana	SINE
Sadhu1-2	NC_003070.5	Arabidopsis thaliana	SINE
Sadhu1-3	.	Arabidopsis thaliana	SINE
Sadhu10-1	NC_003074.4	Arabidopsis thaliana	SINE
Sadhu2-1	NC_003070.5	Arabidopsis thaliana	SINE
Sadhu3-1	NC_003074.4	Arabidopsis thaliana	SINE

Sadhu3-2	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu4-1	NC_003076.4	Arabidopsis thaliana	SINE	
Sadhu4-2	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu5-1	NC_003075.3	Arabidopsis thaliana	SINE	
Sadhu5-2	NC_003076.4	Arabidopsis thaliana	SINE	
Sadhu6-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu7-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu7-2	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu8-1	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu9-1	NC_003070.5	Arabidopsis thaliana	SINE	
Sake_BM	.	Bombyx mori	Daphne	16,1,3,11,4,2,15
Sake2_BM	.	Bombyx mori	Daphne	1,3,4,2
SalSINE1	.	Salmo salar	SINE2/tPHK	
SART-1_APi	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_CF	.	Camponotus floridanus	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_HMM	.	Heliconius melpomene melpomene	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_LHu	.	Linepithema humile	R1	2,11,86,12,1,14,217,4,15,16,3
SART-1_MOc	AFFJ01003038	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-1_PHu	.	Pediculus humanus	R1	
SART-1_PXu	.	Papilio xuthus	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-1B_HMM	.	Heliconius melpomene melpomene	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-2_APi	.	Acyrthosiphon pisum	R1	8,14,12
SART-2_MOc	AFFJ01001302	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3

SART-2_PXu	.	Papilio xuthus	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-3_APi	.	Acyrthosiphon pisum	R1	8,14,12
SART-3_MOc	AFFJ01002771	Galendromus occidentalis	R1	1,3,4,2
SART-4_APi	.	Acyrthosiphon pisum	R1	2,16,1,14,3,11,4,86,15
SART-4_MOc	.	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-5_APi	.	Acyrthosiphon pisum	R1	16,1,3,4,11,86,2,15
SART-5_MOc	.	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-6_APi	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
SART-7_APi	.	Acyrthosiphon pisum	R1	1,3,4,2
SART-8_APi	.	Acyrthosiphon pisum	R1	
SART-9_APi	.	Acyrthosiphon pisum	R1	16,1,3,4,11,86,2,15
SART1	D85594	Bombyx mori	R1	2,8,24,11,86,12,1,14,4,16,3
SART1_R1_Hmel	.	Heliconius melpomene	R1	8,14,12
SARTBm3	.	Bombyx mori	R1	2,8,11,86,12,1,14,4,15,16,3
SARTPx1	.	Papilio xuthus	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc1	AB593321	Tribolium castaneum	R1	2,8,11,86,12,42,1,14,4,15,16,3
SARTTc2	AB593322	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc3	AB593323	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc4	AB593324	Tribolium castaneum	R1	2,11,86,12,1,14,4,15,16,3
SARTTc5	AB593325	Tribolium castaneum	R1	2,11,86,12,1,14,4,15,16,3
SARTTc6	AB593326	Tribolium castaneum	R1	2,8,11,86,12,42,1,14,4,15,16,3
SARTTc7	AB593327	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc8	.	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3

SepiaSINE	.	Sepia officinalis	SINE2/tPHK	
SepiaSINE2	.	Sepia officinalis	SINE2/tPHK	
Sepioth-SINE1	.	Sepioteuthis lessoniana	SINE2/tPHK	
Sepioth-SINE2A	.	Sepioteuthis lessoniana	SINE2/tPHK	
Sepioth-SINE2B	.	Sepioteuthis lessoniana	SINE2/tPHK	
SHALINE10_MT	.	Medicago truncatula	L1	78,2,8,11,12,1,14,81,77,4,75,3,76
SHALINE10_MT	.	Medicago truncatula	L1	78,2,8,11,12,1,14,81,77,4,75,3,76
SHALINE11_MT	.	Medicago truncatula	L1	100,1,3,11,4,2
SHALINE11_MT	.	Medicago truncatula	L1	100,1,3,11,4,2
SHALINE12_MT	.	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
SHALINE12_MT	.	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
SHALINE14_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
SHALINE14_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
SHALINE15_MT	AC147498	Medicago truncatula	L1	2,93,1,75,3,76,92
SHALINE15_MT	AC147498	Medicago truncatula	L1	2,93,1,75,3,76,92
SHALINE16_MT	.	Medicago truncatula	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
SHALINE16_MT	.	Medicago truncatula	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
SHALINE17_MT	.	Medicago truncatula	L1	1,3,11,4,2
SHALINE17_MT	.	Medicago truncatula	L1	1,3,11,4,2
SHALINE18_MT	.	Medicago truncatula	L1	1,3,4,2
SHALINE18_MT	.	Medicago truncatula	L1	1,3,4,2
SHALINE2_MT	.	Medicago truncatula	L1	44,43
SHALINE2_MT	.	Medicago truncatula	L1	44,43

SHALINE20_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,80,1,4,17,3,7
SHALINE20_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,80,1,4,17,3,7
SHALINE21_MT	.	Medicago truncatula	L1	75,76,78,81,77
SHALINE21_MT	.	Medicago truncatula	L1	75,76,78,81,77
SHALINE4_MT	.	Medicago truncatula	L1	75,76,77
SHALINE4_MT	.	Medicago truncatula	L1	75,76,77
SHALINE5_MT	.	Medicago truncatula	L1	43,79,2,8,10,44,24,11,12,80,1,77,4,16,17,75,3,76,7
SHALINE5_MT	.	Medicago truncatula	L1	43,79,2,8,10,44,24,11,12,80,1,77,4,16,17,75,3,76,7
SHALINE6_MT	.	Medicago truncatula	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
SHALINE6_MT	.	Medicago truncatula	L1	43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
SHALINE7_MT	.	Medicago truncatula	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
SHALINE7_MT	.	Medicago truncatula	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
SHALINE8_MT	.	Medicago truncatula	L1	78,2,24,11,1,81,77,4,16,75,76
SHALINE8_MT	.	Medicago truncatula	L1	78,2,24,11,1,81,77,4,16,75,76
SHALINE9_MT	.	Medicago truncatula	L1	2,6,1,17,3,10,79,7,80
SHALINE9_MT	.	Medicago truncatula	L1	2,6,1,17,3,10,79,7,80
SHANSINE_MT	.	Medicago truncatula	SINE	
SINE_AFC	.	Lepidiolamprologus elongatus	SINE2/τPHK	
SINE_CP	X79504	Chironomus pallidivittatus	SINE	
SINE_CP1	X79507	Chironomus pallidivittatus	SINE	
SINE_DR1	.	Danio rerio	SINE	
SINE_FR1A	.	Takifugu rubripes	SINE2/τPHK	
SINE_FR1B	.	Takifugu rubripes	SINE2/τPHK	

SINE_FR1C	.	Takifugu rubripes	SINE2/тРНК
SINE_FR1D	.	Takifugu rubripes	SINE2/тРНК
SINE_FR2	.	Takifugu rubripes	SINE2/тРНК
SINE_RR	.	Pelophylax ridibundus	SINE
SINE_SM	.	Prosopium spilonotus	SINE2/тРНК
SINE_TE	.	Teleostei	SINE2/тРНК
SINE_UT	AB108791, AB108792,	Urotrichus talpoides	SINE
SINE_VV	.	Vicugna vicugna	SINE2/тРНК
SINE-1_AIp	.	Arachis ipaensis	SINE2/тРНК
SINE-1_ALy	.	Arabidopsis lyrata	SINE
SINE-1_ATr	.	Amborella trichopoda	SINE
SINE-1_BM	.	Bombyx mori	SINE
SINE-1_CQ	.	Culex quinquefasciatus	SINE
SINE-1_ECa	.	Eucalyptus camaldulensis	SINE
SINE-1_EPa	.	Schrenkiella parvula	SINE2/тРНК
SINE-1_HM	.	Hydra vulgaris	SINE
SINE-1_LMi	.	Locusta migratoria	SINE
SINE-1_Mad	.	Malus domestica	SINE2/тРНК
SINE-1_MD	.	Monodelphis domestica	SINE2/тРНК
SINE-1_Pmo	.	Python molurus	SINE3/5S
SINE-12_OS	.	Oryza sativa	SINE2/тРНК
SINE-2_CQ	.	Culex quinquefasciatus	SINE
SINE-2_EL	.	Esox lucius	SINE3/5S

SINE-2_LMi	.	Locusta migratoria	SINE
SINE-2_Mad	.	Malus domestica	SINE2/тРНК
SINE-2_MD	.	Monodelphis domestica	SINE2/тРНК
SINE-2019_Crp	.	Crocodylus porosus	SINE
SINE-3_CQ	.	Culex quinquefasciatus	SINE2/тРНК
SINE-3_LMi	.	Locusta migratoria	SINE
SINE-3_Mad	.	Malus domestica	SINE2/тРНК
SINE-3B_CQ	.	Culex quinquefasciatus	SINE2/тРНК
SINE-4_CQ	.	Culex quinquefasciatus	SINE2/тРНК
SINE-4_LMi	.	Locusta migratoria	SINE
SINE-4_Mad	.	Malus domestica	SINE2/тРНК
SINE-5_CQ	.	Culex quinquefasciatus	SINE2/тРНК
SINE-5_FV	.	Fragaria vesca	SINE
SINE-5_LMi	.	Locusta migratoria	SINE
SINE-5_Mad	.	Malus domestica	SINE2/тРНК
SINE-6_CQ	.	Culex quinquefasciatus	SINE2/тРНК
SINE-6_CR	.	Chlamydomonas reinhardtii	SINE2/тРНК
SINE1_BB	AF327249	Bubalus bubalis	SINE
SINE1_CE	.	Caenorhabditis elegans	SINE
SINE1_MT	.	Medicago truncatula	SINE
SINE1_MT	.	Medicago truncatula	SINE2/тРНК
SINE1_SO	.	Solanum demissum	SINE
SINE1_SS	AC129961, AF018149,	Sus scrofa	SINE2/тРНК

SINE16_OS	D85060	Oryza sativa	SINE2/тРНК
SINE1A_SS	.	Sus scrofa	SINE2/тРНК
SINE1a1_Cis	.	Ciona savignyi	SINE
SINE1a2_Cis	.	Ciona savignyi	SINE
SINE1B_SS	.	Sus scrofa	SINE2/тРНК
SINE1b1_Cis	.	Ciona savignyi	SINE
SINE1b2_Cis	.	Ciona savignyi	SINE
SINE1C_SS	.	Sus scrofa	SINE2/тРНК
SINE1D_SS	.	Sus scrofa	SINE2/тРНК
SINE1OS	D10673	Oryza sativa	SINE2/тРНК
SINE1R5_OS	D10674	Oryza sativa	SINE2/тРНК
SINE2_Gav	.	Crocodylidae	SINE2/тРНК
SINE2-1_AC	.	Aplysia californica	SINE2/тРНК
SINE2-1_ACar	.	Anolis carolinensis	SINE2/тРНК
SINE2-1_ADi	.	Acropora digitifera	SINE2/тРНК
SINE2-1_AFC	.	Cichlidae	SINE2/тРНК
SINE2-1_ALy	.	Arabidopsis lyrata	SINE2/тРНК
SINE2-1_AMi	.	Alligator mississippiensis	SINE2/тРНК
SINE2-1_AP	.	Acyrthosiphon pisum	SINE2/тРНК
SINE2-1_BDi	.	Brachypodium distachyon	SINE2/тРНК
SINE2-1_BG	.	Blumeria graminis	SINE2/тРНК
SINE2-1_BT	.	Bos taurus	SINE2/тРНК
SINE2-1_BTa	.	Bemisia tabaci	SINE2/тРНК

SINE2-1_BV	Beta vulgaris	SINE2/тРНК
SINE2-1_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-1_CCr	Condylura cristata	SINE2/тРНК
SINE2-1_CM	Callorhinchus milii	SINE2/тРНК
SINE2-1_CoFl	Copidosoma floridanum	SINE2/тРНК
SINE2-1_CQ	Culex quinquefasciatus	SINE2/тРНК
SINE2-1_Croc	Crocodylidae	SINE2/тРНК
SINE2-1_DaCa	Daucus carota	SINE2/тРНК
SINE2-1_DL	Dicentrarchus labrax	SINE2/тРНК
SINE2-1_DR	Danio rerio	SINE2/тРНК
SINE2-1_EC	Equus caballus	SINE2/тРНК
SINE2-1_EL	Esox lucius	SINE2/тРНК
SINE2-1_FV	Fragaria vesca	SINE2/тРНК
SINE2-1_GA	Gasterosteus aculeatus	SINE2/тРНК
SINE2-1_HM	Hydra vulgaris	SINE2/тРНК
SINE2-1_JC	Jatropha curcas	SINE2/тРНК
SINE2-1_Laf	Loxodonta africana	SINE2/тРНК
SINE2-1_LMi	Locusta migratoria	SINE2/тРНК
SINE2-1_LVa	Litopenaeus vannamei	SINE2/тРНК
SINE2-1_MAc	Musa acuminata	SINE2/тРНК
SINE2-1_NS	Nicotiana sylvestris	SINE2/тРНК
SINE2-1_NV	Nematostella vectensis	SINE2/тРНК
SINE2-1_OL	Oryzias latipes	SINE2/тРНК

SINE2-1_OS	Oryza sativa	SINE2/тРНК
SINE2-1_Pca	Procavia capensis	SINE2/тРНК
SINE2-1_PD	Phoenix dactylifera	SINE2/тРНК
SINE2-1_PH	Parhyale hawaiensis	SINE2/тРНК
SINE2-1_PL	Paracentrotus lividus	SINE2/тРНК
SINE2-1_PM	Petromyzon marinus	SINE2/тРНК
SINE2-1_PMaj	Parus major	SINE2/тРНК
SINE2-1_PPo	Papilio polytes	SINE2/тРНК
SINE2-1_PSi	Pelodiscus sinensis	SINE2/тРНК
SINE2-1_PTr	Populus trichocarpa	SINE2/тРНК
SINE2-1_PTr	Populus trichocarpa	SINE2/тРНК
SINE2-1_PXu	Papilio xuthus	SINE2/тРНК
SINE2-1_RPr	Rhodnius prolixus	SINE2/тРНК
SINE2-1_SBi	Sorghum bicolor	SINE2/тРНК
SINE2-1_SLy	Solanum lycopersicum	SINE2/тРНК
SINE2-1_SP	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-1_SSc	Sus scrofa	SINE2/тРНК
SINE2-1_STu	Solanum tuberosum	SINE2/тРНК
SINE2-1_TAe	Triticum aestivum	SINE2/тРНК
SINE2-1_VC	Volvox carteri	SINE2/тРНК
SINE2-1_VMa	Vaccinium macrocarpon	SINE2/тРНК
SINE2-1_XT	Xenopus tropicalis	SINE2/тРНК
SINE2-1_ZM	Zea mays	SINE2/тРНК

SINE2-11_OS	Oryza sativa	SINE2/тРНК
SINE2-11_SBi	Sorghum bicolor	SINE2/тРНК
SINE2-12_OS	Oryza sativa	SINE2/тРНК
SINE2-12_SBi	Sorghum bicolor	SINE2/тРНК
SINE2-1a_SBi	Sorghum bicolor	SINE2/тРНК
SINE2-1B_ACar	Anolis carolinensis	SINE2/тРНК
SINE2-1B_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-1B_CCr	Condylura cristata	SINE2/тРНК
SINE2-1B_DR	Danio rerio	SINE2/тРНК
SINE2-1B_VMa	Vaccinium macrocarpon	SINE2/тРНК
SINE2-1C_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-1c_SBi	Sorghum bicolor	SINE2/тРНК
SINE2-1D_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-2_AC	Aplysia californica	SINE2/тРНК
SINE2-2_ACar	Anolis carolinensis	SINE2/тРНК
SINE2-2_ADi	Acropora digitifera	SINE2/тРНК
SINE2-2_AP	Acyrthosiphon pisum	SINE2/тРНК
SINE2-2_BT	Bos taurus	SINE2/тРНК
SINE2-2_BV	Beta vulgaris	SINE2/тРНК
SINE2-2_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-2_CM	Callorhinchus milii	SINE2/тРНК
SINE2-2_CoFl	Copidosoma floridanum	SINE2/тРНК
SINE2-2_CQ	Culex quinquefasciatus	SINE2/тРНК

SINE2-2_DaCa	Daucus carota	SINE2/тРНК
SINE2-2_DR	Danio rerio	SINE2/тРНК
SINE2-2_FV	Fragaria vesca	SINE2/тРНК
SINE2-2_HM	Hydra vulgaris	SINE2/тРНК
SINE2-2_LMi	Locusta migratoria	SINE2/тРНК
SINE2-2_LVa	Litopenaeus vannamei	SINE2/тРНК
SINE2-2_MAc	Musa acuminata	SINE2/тРНК
SINE2-2_NV	Nematostella vectensis	SINE2/тРНК
SINE2-2_OL	Oryzias latipes	SINE2/тРНК
SINE2-2_OS	Oryza sativa	SINE2/тРНК
SINE2-2_PH	Parhyale hawaiensis	SINE2/тРНК
SINE2-2_PM	Petromyzon marinus	SINE2/тРНК
SINE2-2_PPo	Papilio polytes	SINE2/тРНК
SINE2-2_PTr	Populus trichocarpa	SINE2/тРНК
SINE2-2_PTr	Populus trichocarpa	SINE2/тРНК
SINE2-2_PXu	Papilio xuthus	SINE2/тРНК
SINE2-2_RPr	Rhodnius prolixus	SINE2/тРНК
SINE2-2_SP	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-2_STu	Solanum tuberosum	SINE2/тРНК
SINE2-2_VMa	Vaccinium macrocarpon	SINE2/тРНК
SINE2-2B_HM	Hydra vulgaris	SINE2/тРНК
SINE2-2B_PH	Parhyale hawaiensis	SINE2/тРНК
SINE2-2B_SSa	Salmo salar	SINE2/тРНК

SINE2-2C_SSa	Salmo salar	SINE2/тРНК
SINE2-2D_SSa	Salmo salar	SINE2/тРНК
SINE2-3_ALy	Arabidopsis lyrata	SINE2/тРНК
SINE2-3_BT	Bos taurus	SINE2/тРНК
SINE2-3_CCa	Cyprinus carpio	SINE2/тРНК
SINE2-3_CM	Callorhinchus milii	SINE2/тРНК
SINE2-3_CoFl	Copidosoma floridanum	SINE2/тРНК
SINE2-3_DaCa	Daucus carota	SINE2/тРНК
SINE2-3_DR	Danio rerio	SINE2/тРНК
SINE2-3_EL	Esox lucius	SINE2/тРНК
SINE2-3_FV	Fragaria vesca	SINE2/тРНК
SINE2-3_LMi	Locusta migratoria	SINE2/тРНК
SINE2-3_LVa	Litopenaeus vannamei	SINE2/тРНК
SINE2-3_MAc	Musa acuminata	SINE2/тРНК
SINE2-3_NV	Nematostella vectensis	SINE2/тРНК
SINE2-3_OL	Oryzias latipes	SINE2/тРНК
SINE2-3_OS	Oryza sativa	SINE2/тРНК
SINE2-3_PH	Parhyale hawaiensis	SINE2/тРНК
SINE2-3_PM	Petromyzon marinus	SINE2/тРНК
SINE2-3_SP	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-3_VMa	Vaccinium macrocarpon	SINE2/тРНК
SINE2-3B_DR	Danio rerio	SINE2/тРНК
SINE2-3B_OL	Oryzias latipes	SINE2/тРНК

SINE2-4_ALy.	Arabidopsis lyrata	SINE2/тРНК
SINE2-4_CCa.	Cyprinus carpio	SINE2/тРНК
SINE2-4_DR .	Danio rerio	SINE2/тРНК
SINE2-4_FV .	Fragaria vesca	SINE2/тРНК
SINE2-4_NV .	Nematostella vectensis	SINE2/тРНК
SINE2-4_OS .	Oryza sativa	SINE2/тРНК
SINE2-4_PH .	Parhyale hawaiensis	SINE2/тРНК
SINE2-4_PM .	Petromyzon marinus	SINE2/тРНК
SINE2-4_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-4_VMa .	Vaccinium macrocarpon	SINE2/тРНК
SINE2-4a_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-4a1_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-4B_OS .	Oryza sativa	SINE2/тРНК
SINE2-4b_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-4c_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-5_ALy.	Arabidopsis lyrata	SINE2/тРНК
SINE2-5_DR .	Danio rerio	SINE2/тРНК
SINE2-5_NV .	Nematostella vectensis	SINE2/тРНК
SINE2-5_OL .	Oryzias latipes	SINE2/тРНК
SINE2-5_SP .	Strongylocentrotus purpuratus	SINE2/тРНК
SINE2-5_SSa .	Salmo salar	SINE2/тРНК
SINE2-5_VMa .	Vaccinium macrocarpon	SINE2/тРНК
SINE2-6_ALy.	Arabidopsis lyrata	SINE2/тРНК

SINE2-6_SP	.	Strongylocentrotus purpuratus	SINE2/тPHK
SINE2-6_SSa	.	Salmo salar	SINE2/тPHK
SINE2-6_VMa	.	Vaccinium macrocarpon	SINE2/тPHK
SINE2-7_SP	.	Strongylocentrotus purpuratus	SINE2/тPHK
SINE2-8_SP	.	Strongylocentrotus purpuratus	SINE2/тPHK
SINE2B-1_CM	.	Callorhinchus milii	SINE2/тPHK
SINE3_AFC	.	Cichlidae	SINE2/тPHK
SINE3_IP	.	Ictalurus punctatus	SINE3/5S
SINE3-1	.	Danio rerio	SINE3/5S
SINE3-1_AO	.	Aspergillus oryzae	SINE3/5S
SINE3-1_CCa	.	Cyprinus carpio	SINE3/5S
SINE3-1_TC	.	Tribolium castaneum	SINE3/5S
SINE3-1a	.	Danio rerio	SINE3/5S
SINE3-2_AO	.	Aspergillus oryzae	SINE3/5S
SINE3-2_DR	.	Danio rerio	SINE3/5S
SINE6_OS	D85050	Oryza sativa	SINE2/тPHK
SINE8_OS	D85052	Oryza sativa	SINE2/тPHK
SINE9_OS	.	Oryza sativa	SINE2/тPHK
SINEB1_MU	.	Meriones unguiculatus	SINE1/7SL
SINEB2_MU	.	Meriones unguiculatus	SINE2/тPHK
SINEC_c1	.	Canidae	SINE2/тPHK
SINEC_Fc	.	Felis	SINE2/тPHK
SINEC_Fc2	.	Felis	SINE2/тPHK

SINEC_Fc3	.	Felis catus	SINE2/тРНК
SINEC_PV	Z33497	Phoca vitulina	SINE2/тРНК
SINEC-1_MPu	.	Mustela putorius furo	SINE2/тРНК
SINEC-1B_MPu	.	Mustela putorius furo	SINE2/тРНК
SINEC-1C_MPu	.	Mustela putorius furo	SINE2/тРНК
SINEC-1D_MPu	.	Mustela putorius furo	SINE2/тРНК
SINEC1_AMe	.	Ailuropoda melanoleuca	SINE2/тРНК
SINEC1_AMe2	.	Ailuropoda melanoleuca	SINE2/тРНК
SINEC1_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1A_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1B_AMe	.	Ailuropoda melanoleuca	SINE2/тРНК
SINEC1B1_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1B2_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1C1_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1C2_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC1D_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC2_AMe	.	Ailuropoda melanoleuca	SINE2/тРНК
SINEC2A1_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEC2A2_CF	.	Canis lupus familiaris	SINE2/тРНК
SINEL_SM	.	Schmidtea mediterranea	NeSL
SINEU-1_CM	.	Callorhinchus milii	SINEU/snRNA
SINEU-1A_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1B_CM	.	Callorhinchus milii	SINEU/snRNA

SINEU-1B_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1C_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1D_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1E_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1F_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1G_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1G2_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1H_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1I_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-1J_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-2_Crp	.	Crocodylus porosus	SINEU/snRNA
SINEU-2_Gav	.	Gavialis gangeticus	SINEU/snRNA
SINEU-3A_Croc	.	Crocodylidae	SINEU/snRNA
SINEU-3B_Croc	.	Crocodylidae	SINEU/snRNA
SINEX-1_AG	.	Anopheles gambiae	SINE
SINEX-1_CM	DQ524330	Callorhinchus milii	SINE2/tPHK
SINEX-1_CR	.	Chlamydomonas reinhardtii	SINE
SINEX-2_AG	.	Anopheles gambiae	SINE2/tPHK
SINEX-2_CR	.	Chlamydomonas reinhardtii	SINE
SINEX-3_CR	.	Chlamydomonas reinhardtii	L1
SINEX-4_CR	.	Chlamydomonas reinhardtii	SINE
SINEX-5_CR	.	Chlamydomonas reinhardtii	SINE2/tPHK
Sismon-1_PH	.	Parhyale hawaiensis	RTE 18,5,1,4,11,2

SK	D14862	Heterololigo bleekeri	SINE2/тРНК	
SLACS	X17078	Trypanosoma brucei	CRE	1,3,4,2
SlmI	.	Salmoniformes	SINE2/тРНК	
SlmII	.	Salmoniformes	SINE2/тРНК	
Sm1	.	Selaginella moellendorffii	Penelope	
Sm2	.	Selaginella moellendorffii	Penelope	235,236
SMAIDIV	.	Oncorhynchus keta	SINE2/тРНК	
SOD_SINE	.	Solanum demissum	SINE	
SOD_SINE	.	Solanum demissum	SINE	
SOR1	.	Sorex	SINE2/тРНК	
SP-D-Geo	.	Dipodomys	SINE1/7SL	
SPEC1	J01207	Strongylocentrotus purpuratus	SINE2/тРНК	
SPLN3_SO	AB016137, AB016137.1	Spinacia oleracea	Не содержащий LTR ретротранспозон	
SPOCK_DMi	AY144571	Drosophila miranda	Outcast	16,1,3,11,4,2,15
Squam2	.	Squamata	SINE2/тРНК	
SR1	U66331	Schistosoma mansoni	CR1	1,3,4,2
SR2	AF025672	Schistosoma mansoni	RTE	1,26,3,11,4,2
SR2A	AF025680	Schistosoma mansoni	RTE	
SR2B	AF025681	Schistosoma mansoni	RTE	
STR1	.	Ictidomys tridecemlineatus	SINE1/7SL	
STRID1	.	Ictidomys tridecemlineatus	SINE2/тРНК	
STRID2	.	Ictidomys tridecemlineatus	SINE2/тРНК	
STRID3	.	Ictidomys tridecemlineatus	SINE2/тРНК	

STRID4	.	Ictidomys tridecemlineatus	SINE2/тРНК
STRIDE_Cpo	.	Cavia porcellus	SINE1/7SL
STRIDE1	.	Ictidomys tridecemlineatus	SINE1/7SL
STRIDE2	.	Ictidomys tridecemlineatus	SINE1/7SL
STRIDM	.	Ictidomys tridecemlineatus	SINE2/тРНК
SURF1	X07085	Strongylocentrotus purpuratus	SINE2/тРНК
SUSINE2	.	Sus scrofa	SINE2/тРНК
SVA_A	.	Homo sapiens	SINE
SVA_A	.	Homo sapiens	SINE
SVA_B	.	Homo sapiens	SINE
SVA_C	.	Homo sapiens	SINE
SVA_D	.	Homo sapiens	SINE
SVA_E	.	Homo sapiens	SINE
SVA_F	.	Homo sapiens	SINE
SVA2	.	Eutheria	MSAT
Swimmer	AF055641	Oryzias latipes	L1
Syrinx_DS	.	Darwinula stevensoni	Jockey 16,1,3,11,4,2,15
T_Av	.	Adineta vaga	Penelope 1,3
T1	M93689	Anopheles gambiae	CR1 54,1,3,59,53,11,4,2
T1_As	.	Adineta	Penelope 1,3,184
T2	X15618	Schistosoma mansoni	SINE2/тРНК
T2_As	.	Adineta	Penelope 1,3
T3_As	.	Adineta	Penelope 1,3

TA11	L47193	Arabidopsis thaliana	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
TA12	.	Arabidopsis thaliana	L1	
Tad1	L25662	Neurospora crassa	Tad1	16,1,14,3,11,4,2,15
Tad1-1_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_AF	.	Aspergillus fumigatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-1_CCi	AACS02000007	Coprinopsis cinerea	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_CGlo	.	Chaetomium globosum	Tad1	16,1,3,11,4,2,15
Tad1-1_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_GLuc	.	Ganoderma lucidum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_NC	.	Neurospora crassa	Tad1	
Tad1-1_PGr	.	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Tad1-1_PPl	.	Postia placenta	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-1_PSt	.	Puccinia striiformis	Tad1	
Tad1-1_PTrit	.	Puccinia triticina	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-1_RGl	AEVR01000393	Rhodotorula toruloides	Tad1	6,2,9,10,11,1,15,16,17,3,7
Tad1-1_TMe	.	Tuber melanosporum	Tad1	16,1,3,11,4,2,15
Tad1-1_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Tad1-10_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-10_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-10_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-10_PGr	.	Puccinia graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-10_TSt .	Talaromyces stipitatus	Tad1	
Tad1-100_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-101_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-102_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_BoBo .	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-11_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_TSt .	Talaromyces stipitatus	Tad1	
Tad1-12_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-12_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13B_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-14_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-14_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15B_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15C_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-16_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-16_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-17_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-17_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-17B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-18_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-18_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-19_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-19_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-1B_PGr	AAWC01000331	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-2_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,4,11,2,15
Tad1-2_CGlo	.	Chaetomium globosum	Tad1	16,1,3,11,4,2,15
Tad1-2_EP	.	Erysiphe pisi	Tad1	7,16,1,17,9,3,10,11,4,6,2,15
Tad1-2_PGr	AAWC01004263	Puccinia graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_PTrit	.	Puccinia triticina	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-2_TMe	.	Tuber melanosporum	Tad1	16,1,3,11,4,2,15
Tad1-2_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-20_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-20_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-21_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-21_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-22_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-22_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-23_BG	.	Blumeria graminis	Tad1	1,3,11,4,2
Tad1-23_EP	.	Erysiphe pisi	Tad1	7,1,17,16,3,11,4,2,15

Tad1-24_BG .	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15
Tad1-24_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25B_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25B_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25C_BG.	Blumeria graminis	Tad1	17,9,10,6,7,13
Tad1-26_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-26_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-26B_BG.	Blumeria graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-26C_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-27_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-27_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28B_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-29_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-29_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-3_ACap .	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-3_BG .	Blumeria graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-3_BoBo .	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-3_CGlo .	Chaetomium globosum	Tad1	6,2,8,10,11,12,42,1,14,4,15,16,17,3
Tad1-3_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-3_PGr	AAWC01001353	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-3_PTrit	.	Puccinia triticina	Tad1	16,1,3,11,4,2,15
Tad1-3_TMe	.	Tuber melanosporum	Tad1	16,1,3,4,11,2,15
Tad1-3_TSt	.	Talaromyces stipitatus	Tad1	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Tad1-30_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-30_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-30B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-32_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-32_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-32B_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-33_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-33_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-34_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-34_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-36_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-36_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-37_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-37_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-38_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-38_EP	.	Erysiphe pisi	Tad1	16,42,1,8,3,4,11,12,2,15
Tad1-38B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-39_BG	.	Blumeria graminis	Tad1	16,1,8,14,3,11,4,12,2,15
Tad1-39_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-4_ACap	.	Histoplasma capsulatum	Tad1	2,8,11,12,1,14,4,15,16,3
Tad1-4_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-4_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-4_CGlo	.	Chaetomium globosum	Tad1	6,2,10,11,1,14,4,15,16,17,3,7
Tad1-4_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-4_PGr	AAWC01002745	Puccinia graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-4_PTrit	.	Puccinia triticina	Tad1	
Tad1-4_TMe	.	Tuber melanosporum	Tad1	6,2,18,7,9,10,11,13,1,4,5,17,3
Tad1-4_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-40_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-40_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-42_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-42_EP	.	Erysiphe pisi	Tad1	1,17,9,3,10,6,7,13
Tad1-43_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-43_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-43B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-44_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-44_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-45_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-45B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46C_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46D_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-47_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-48_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-48_EP	.	Erysiphe pisi	Tad1	6,2,7,9,10,11,1,4,15,16,17,3
Tad1-49_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-49_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Tad1-5_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-5_CGlo	.	Chaetomium globosum	Tad1	2,8,11,12,42,1,14,4,15,16,17,3
Tad1-5_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_PGr	AAWC01000958	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-5_TMe	.	Tuber melanosporum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_TSt	.	Talaromyces stipitatus	Tad1	

Tad1-50_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-50_EP .	Erysiphe pisi	Tad1	
Tad1-51_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-52_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-52_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53B_BG.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-53C_BG.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-54_EP .	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-55_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-55_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-56_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-56_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-57_BG .	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15
Tad1-57_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-58_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-58_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-59_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-59_EP .	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5B_TMe.	Tuber melanosporum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-6_ACap .	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-6_BG .	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-6_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-6_CGlo	.	Chaetomium globosum	Tad1	1,3,11,4,2
Tad1-6_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-6_PGr	AAWC01003552	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-6_TSt	.	Talaromyces stipitatus	Tad1	7,1,17,9,16,3,10,11,4,6,2,15
Tad1-60_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-60_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-61_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-61_EP	.	Erysiphe pisi	Tad1	16,1,8,3,11,4,12,2,15
Tad1-62_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-62_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-63_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-63_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-63B_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-63C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-64_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-64_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-65_BG	.	Blumeria graminis	Tad1	
Tad1-65_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-65B_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,6,2,13
Tad1-66_BG	.	Blumeria graminis	Tad1	6,2,241,9,10,11,13,1,4,15,16,17,3,7
Tad1-66_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-66B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-67_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-67_EP	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-68_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-68_EP	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-69_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-69_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-69B_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-69C_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,239,7
Tad1-7_BoBo	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-7_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7_PGr	Puccinia graminis	Tad1	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Tad1-7_TSt	Talaromyces stipitatus	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-70_BG	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15
Tad1-70_EP	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-70B_BG	Blumeria graminis	Tad1	16,1,3,11,4,2,15
Tad1-71_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-71_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-72_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-72_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-73_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-73_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-74_BG	Blumeria graminis	Tad1	

Tad1-74_EP	Erysiphe pisi	Tad1	6,2,7,9,10,11,1,4,15,16,17,3
Tad1-75_BG	Blumeria graminis	Tad1	16,1,3,11,4,2,15
Tad1-75_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-76_BG	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-76_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-77_BG	Blumeria graminis	Tad1	17,9,10,6,7,13
Tad1-77_EP	Erysiphe pisi	Tad1	16,1,3,4,11,2,15
Tad1-78_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-79_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7B_BG	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7B_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-8_BG	Blumeria graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-8_BoBo	Botryobasidium botryosum	Tad1	
Tad1-8_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-8_PGr	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Tad1-8_TSt	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-80_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-81_EP	Erysiphe pisi	Tad1	7,16,1,17,3,10,11,4,6,2,15
Tad1-82_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-83_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-84_EP	Erysiphe pisi	Tad1	16,11,15
Tad1-85_EP	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-86_EP	Erysiphe pisi	Tad1	16,1,8,3,11,4,12,2,15

Tad1-87_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2
Tad1-88_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-89_EP	.	Erysiphe pisi	Tad1	16,11,15
Tad1-9_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-9_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9_PGr	.	Puccinia graminis	Tad1	7,1,17,9,3,10,6,2,13
Tad1-9_TSt	.	Talaromyces stipitatus	Tad1	16,1,17,3,11,4,2,15
Tad1-90_EP	.	Erysiphe pisi	Tad1	2,8,11,12,42,1,14,4,15,16,3
Tad1-91_EP	.	Erysiphe pisi	Tad1	7,1,17,16,3,11,4,2,15
Tad1-92_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-93_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-94_EP	.	Erysiphe pisi	Tad1	
Tad1-95_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-96_EP	.	Erysiphe pisi	Tad1	2,8,11,12,42,1,14,4,15,16,3
Tad1-97_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-98_EP	.	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-99_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
TAHRE	AJ5425818	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15
TAL	.	Talpidae	SINE2/тРНК	
TAL-1	AB108790, AB108813,	Mogera robusta	SINE2/тРНК	

Talua	.	Reticulitermes lucifugus	SINE2/tPHK	
Talub	.	Termitoidae	SINE2/tPHK	
Taluc	.	Termitoidae	SINE2/tPHK	
Talud	.	Termitoidae	SINE2/tPHK	
TAR2_TA	AB012296	Triticum aestivum	L1	
TART_B1	U14101	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15
TART_DV	AY219709	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
TART-1_DWi	.	Drosophila willistoni	Jockey	1,3,4,2
TART-A	AJ566116	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15
TDD3	AF002669	Dictyostelium discoideum	L1	1,4,2
TE-18_LVa	.	Litopenaeus vannamei	Penelope	
TE-3_XT	.	Xenopus tropicalis	Penelope	
TE-4_LVa	.	Litopenaeus vannamei	Jockey	
TE-5_LVa	.	Litopenaeus vannamei	Penelope	
TE-6_LVa	.	Litopenaeus vannamei	Penelope	
TguSINE1	.	Estrildidae	SINE2/tPHK	
TguSINE1_Ucy	.	Uraeginthus cyanocephalus	SINE2/tPHK	
THER1	.	Mammalia	SINE2/tPHK	
THER1	.	Mammalia	SINE2/tPHK	
THER1	.	Mammalia	SINE2/tPHK	
THER1_MD	.	Monodelphis domestica	SINE2/tPHK	
THER2	.	Mammalia	SINE2/tPHK	
THER2	.	Mammalia	SINE2/tPHK	

THER2	.	Mammalia	SINE2/тРНК		
THER2_MD	.	Monodelphis domestica	SINE2/тРНК		
THER2_ME	.	Notamacropus eugenii	SINE2/тРНК		
Togen-1_CCa	.	Cyprinus carpio	Не содержащий LTR ретротранспозон 1,3,2		
Togen-1_DR	.	Danio rerio	Не содержащий LTR ретротранспозон 1,3,4,2		
Togen-1_SSa	.	Salmo salar	Не содержащий LTR ретротранспозон		
TP2	X52770	Physarum polycephalum	Не содержащий LTR ретротранспозон		
TRAS-1_HMM	.	Heliconius melpomene melpomene	R1	7,1,17,9,10,4,6,2,13	
TRAS-1_MSe	.	Manduca sexta	R1	6,2,9,10,11,86,1,4,15,16,17,3,7	
TRAS-1_PXu	.	Papilio xuthus	R1		
TRAS-1_TUr	каркас_45	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-2_PXu	.	Papilio xuthus	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7	
TRAS-2_TUr	.	Tetranychus urticae	R1	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3	
TRAS-3_TUr	.	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-4_TUr	.	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-5_TUr	.	Tetranychus urticae	R1	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3	
TRAS-6_TUr	.	Tetranychus urticae	R1		
TRAS-7_TUr	.	Tetranychus urticae	R1	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7	
TRAS-8_TUr	.	Tetranychus urticae	R1		
TRAS1	D38414	Bombyx mori	R1	6,2,7,9,10,11,86,13,1,4,15,16,17,3	
TRAS1_R1_Hmel	.	Heliconius melpomene	R1	133,130,134,135,136,000	
TRAS3_BM	AB046668	Bombyx mori	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7	
TRAS3_SC	AB046673	Samia cynthia	R1		

TRAS4_SC	AB046674	Samia cynthia	R1	
TRAS6_BM	AB046671	Bombyx mori	R1	
TRAS9_SC	AB046675	Samia cynthia	R1	1,3,11,4,2
TRE3B	AF134170	Dictyostelium discoideum	L1	16,1,3,24,4,11,2
TRE3C	AF134171	Dictyostelium discoideum	L1	1,3,11,4,2
TRE3D	NW_633189	Dictyostelium discoideum	L1	1,3,11,4,2
TREP20	AY013246	Hordeum vulgare	L1	
TREP51	AJ270053	Hordeum vulgare	L1	
TREP52	AJ270056	Hordeum vulgare	L1	
TREP53	AJ270054	Hordeum vulgare	L1	
TREP58	AJ270055	Hordeum marinum	L1	
TREP80	Y13390	Secale cereale	L1	
TrepLINE1	.	Trifolium repens	L1	43,2,8,44,24,11,12,1,77,4,16,75,3,76
TRIM	X59239	Drosophila miranda	Loa	
TS	D17453	Nicotiana tabacum	SINE	
TS2	.	Solanum demissum	SINE	
TS2	.	Solanum demissum	SINE	
TSCL	U65470	Arabidopsis thaliana	Не содержащий LTR ретротранспозон	
tSINE_Fc	.	Felis catus	SINE2/тРНК	
Tu-III	AB090261AB090262AB09	Tupaia glis	SINE2/тРНК	
TUB1	.	Tupaia belangeri	SINE2/тРНК	
TUB2	.	Tupaia belangeri	SINE2/тРНК	
TUB2B	.	Tupaia belangeri	SINE2/тРНК	

TUB3	.	Tupaia belangeri	SINE2/τPHK	
TUB4	.	Tupaia belangeri	SINE2/τPHK	
TUBE1	.	Tupaia belangeri	SINE2/τPHK	
TUBE2	.	Tupaia belangeri	SINE2/τPHK	
TUBE3	.	Tupaia belangeri	SINE2/τPHK	
TUBE4	.	Tupaia belangeri	SINE2/τPHK	
TUBE5	.	Tupaia belangeri	SINE2/τPHK	
TUS	.	Tupaia belangeri	SINE2/τPHK	
TWIN	.	Culex pipiens	SINE2/τPHK	
TWIN2_CP	AF282725	Culex pipiens	SINE2/τPHK	
twinID-Spe	.	Spermophilus	SINE2/τPHK	
TX1_FR1	.	Takifugu rubripes	L1	
Tx1_FR1B	.	Takifugu rubripes	Tx1	1,26,45,3,11,4,2
TX1_FR2	.	Takifugu rubripes	L1	
Tx1_XT	.	Xenopus tropicalis	Tx1	1,26,8,14,3,11,4,12,2
Tx1-1_AAe	.	Aedes aegypti	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-1_AC	.	Aplysia californica	Tx1	16,1,26,3,24,4,11,2
Tx1-1_ACar	.	Anolis carolinensis	Tx1	16,1,3,46,24,4,11,12,2
Tx1-1_ADi	.	Acropora digitifera	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-1_ASi	.	Alligator sinensis	Tx1	16,1,26,8,3,24,11,4,12,2
Tx1-1_BF	.	Branchiostoma floridae	Tx1	43,2,8,44,11,12,1,26,14,4,3
Tx1-1_CGi	.	Crassostrea gigas	Tx1	150,43,2,44,11,1,26,151,152,4,153,15,16,3,154
Tx1-1_CM	.	Callorhinchus milii	Tx1	

Tx1-1_CPB	.	Chrysemys picta bellii	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-1_CQ	.	Culex quinquefasciatus	Tx1	1,8,14,3,11,4,12,2
Tx1-1_Crp	.	Crocodylus porosus	Tx1	16,1,26,8,3,24,4,11,12,2
Tx1-1_DPu	.	Daphnia pulex	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-1_DR	.	Danio rerio	Tx1	1,8,26,3,11,4,12,2
Tx1-1_FoCa	.	Folsomia candida	Tx1	16,1,26,8,3,24,11,4,12,2
Tx1-1_Gav	.	Gavialis gangeticus	Tx1	1,8,11,12,2
Tx1-1_HM	.	Hydra vulgaris	Tx1	1,26,3,11,4,2
Tx1-1_LCh	.	Latimeria chalumnae	Tx1	2,8,24,11,12,1,26,14,4,16,3,46
Tx1-1_LMi	.	Locusta migratoria	Tx1	16,1,3,24,11,4,2
Tx1-1_NV	.	Nematostella vectensis	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-1_PM	.	Petromyzon marinus	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-1_PSi	.	Pelodiscus sinensis	Tx1	
Tx1-1_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,24,11,1,26,4,30,16,3
Tx1-1_SP	.	Strongylocentrotus purpuratus	Tx1	1,43,3,44,4,2
Tx1-1_SSa	.	Salmo salar	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-10_AMi	.	Alligator mississippiensis	Tx1	
Tx1-10_BF	.	Branchiostoma floridae	Tx1	1,26,43,3,44,11,4,2
Tx1-10_CGi	.	Crassostrea gigas	Tx1	138,2,11,1,26,139,140,4,15,16,3,141
Tx1-10_CQ	.	Culex quinquefasciatus	Tx1	92,16,1,8,3,24,11,4,12,2
Tx1-10_Crp	.	Crocodylus porosus	Tx1	
Tx1-10_DR	.	Danio rerio	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-10_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46

Tx1-10_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-10_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,3,24,11,4,2
Tx1-10_SSa	.	Salmo salar	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-10B_Crp	.	Crocodylus porosus	Tx1	
Tx1-11_AMi	.	Alligator mississippiensis	Tx1	
Tx1-11_BF	.	Branchiostoma floridae	Tx1	43,2,21,44,24,11,1,26,4,16,3
Tx1-11_CGi	.	Crassostrea gigas	Tx1	150,43,2,44,11,1,26,151,152,4,153,15,16,3,154
Tx1-11_CQ	.	Culex quinquefasciatus	Tx1	2,1
Tx1-11_Crp	.	Crocodylus porosus	Tx1	
Tx1-11_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-11_LCh	.	Latimeria chalumnae	Tx1	16,1,26,3,24,11,4,2
Tx1-11_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-11_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-11_SSa	.	Salmo salar	Tx1	1,43,44,4,2
Tx1-12_AMi	.	Alligator mississippiensis	Tx1	
Tx1-12_BF	.	Branchiostoma floridae	Tx1	43,2,44,11,12,1,26,14,4,15,16,3
Tx1-12_CGi	.	Crassostrea gigas	Tx1	16,1,43,26,3,44,4,15,11,2
Tx1-12_CQ	.	Culex quinquefasciatus	Tx1	16,1,26,3,24,11,4,2
Tx1-12_Crp	.	Crocodylus porosus	Tx1	16,1,3,24,11,4,2
Tx1-12_DR	.	Danio rerio	Tx1	21,1,26,23,43,3,44,11,4,2
Tx1-12_Lch	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,14,4,3,46
Tx1-12_LMi	.	Locusta migratoria	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-12_SK	.	Saccoglossus kowalevskii	Tx1	1,8,26,14,3,11,4,12,2

Tx1-12_SSa	.	Salmo salar	Tx1	2,8,11,12,42,1,26,14,4,15,16,3
Tx1-13_BF	.	Branchiostoma floridae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-13_CGi	.	Crassostrea gigas	Tx1	16,1,26,3,24,11,4,2
Tx1-13_CQ	.	Culex quinquefasciatus	Tx1	1,3,11,4,2
Tx1-13_DR	.	Danio rerio	Tx1	1,43,3,44,11,4,2
Tx1-13_Lch	.	Latimeria chalumnae	Tx1	
Tx1-13_LMi	.	Locusta migratoria	Tx1	16,1,26,43,44,24,11,4,2
Tx1-13_SK	.	Saccoglossus kowalevskii	Tx1	1,26,3,11,4,2
Tx1-13_SSa	.	Salmo salar	Tx1	
Tx1-14_BF	.	Branchiostoma floridae	Tx1	
Tx1-14_CGi	.	Crassostrea gigas	Tx1	1,3,11,4,2
Tx1-14_CQ	.	Culex quinquefasciatus	Tx1	93,16,1,26,3,24,11,4,92,2
Tx1-14_DR	.	Danio rerio	Tx1	1,26,43,3,44,11,4,2
Tx1-14_LMi	.	Locusta migratoria	Tx1	16,1,43,3,44,24,11,4,2
Tx1-14_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-14_SSa	.	Salmo salar	Tx1	1,26,11,2
Tx1-14a_Lch	.	Latimeria chalumnae	Tx1	
Tx1-14b_Lch	.	Latimeria chalumnae	Tx1	
Tx1-15_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-15_CQ	.	Culex quinquefasciatus	Tx1	16,8,24,11,12
Tx1-15_DR	.	Danio rerio	Tx1	1,43,3,44,4,2
Tx1-15_Lch	.	Latimeria chalumnae	Tx1	
Tx1-15_LMi	.	Locusta migratoria	Tx1	1,8,3,11,4,12,2

Tx1-15_SK	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-16_BF	Branchiostoma floridae	Tx1	1,43,23,3,44,4,2
Tx1-16_DR	Danio rerio	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-16_SK	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-16a_Lch	Latimeria chalumnae	Tx1	
Tx1-16b_Lch	Latimeria chalumnae	Tx1	
Tx1-17_BF	Branchiostoma floridae	Tx1	16,1,26,43,3,44,11,4,2,15
Tx1-17_DR	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-17_Lch	Latimeria chalumnae	Tx1	
Tx1-17_SK	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-18_BF	Branchiostoma floridae	Tx1	2,1,8,26,14,3,11,4,12
Tx1-18_DR	Danio rerio	Tx1	21,1,26,23,43,3,44,11,4,2
Tx1-18_Lch	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-18_SK	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-19_DR	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,15,16,3
Tx1-19_SK	Saccoglossus kowalevskii	Tx1	26,11
Tx1-19a_Lch	Latimeria chalumnae	Tx1	43,2,8,44,11,12,1,14,4,3,46
Tx1-19b_Lch	Latimeria chalumnae	Tx1	43,2,8,44,11,12,1,14,4,3,46
Tx1-1B_CM	Callorhinchus milii	Tx1	
Tx1-1B_CPB	Chrysemys picta bellii	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-1B_PM	Petromyzon marinus	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-1B_PSi	Pelodiscus sinensis	Tx1	16,1,26,3,24,11,4,2
Tx1-1C_CPB	Chrysemys picta bellii	Tx1	

Tx1-1N1_ACar	Anolis carolinensis	Tx1	46,12
Tx1-2_AAe	Aedes aegypti	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-2_AMi	Alligator mississippiensis	Tx1	1,26,3,11,4,2
Tx1-2_ASi	Alligator sinensis	Tx1	1,26,8,3,11,4,12,2
Tx1-2_BF	Branchiostoma floridae	Tx1	21,1,26,23,3,11,4,2
Tx1-2_CGi	Crassostrea gigas	Tx1	2,16,1,8,26,3,11,4,12,15
Tx1-2_CPB	Chrysemys picta bellii	Tx1	
Tx1-2_CQ	Culex quinquefasciatus	Tx1	2,8,24,11,12,1,14,4,16,3
Tx1-2_Crp	Crocodylus porosus	Tx1	42,8,14,46,12
Tx1-2_DR	Danio rerio	Tx1	21,1,26,23,3,11,4,2
Tx1-2_FoCa	Folsomia candida	Tx1	42,1,26,8,14,3,11,4,12,2
Tx1-2_LCh	Latimeria chalumnae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-2_LMi	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-2_NV	Nematostella vectensis	Tx1	
Tx1-2_PM	Petromyzon marinus	Tx1	1,26,3,11,4,2
Tx1-2_SK	Saccoglossus kowalevskii	Tx1	1,26,3,11,4,2
Tx1-2_SPur	Strongylocentrotus purpuratus	Tx1	1,26,43,3,44,11,4,2
Tx1-2_SSa	Salmo salar	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-2_XT	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
Tx1-20_DR	Danio rerio	Tx1	21,1,26,43,3,44,11,4,2
Tx1-20_SK	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-21_DR	Danio rerio	Tx1	1,26,11,2
Tx1-21_SK	Saccoglossus kowalevskii	Tx1	1,8,26,14,43,3,44,11,4,12,2

Tx1-22_DR	.	Danio rerio	Tx1	1,43,3,44,11,4,2
Tx1-23_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-24_DR	.	Danio rerio	Tx1	1,3,4,2
Tx1-25_DR	.	Danio rerio	Tx1	1,26,8,14,3,11,4,12,2
Tx1-26_DR	.	Danio rerio	Tx1	1,26,8,14,3,4,11,12,2
Tx1-27_DR	.	Danio rerio	Tx1	1,43,3,44,2
Tx1-28_DR	.	Danio rerio	Tx1	44,43
Tx1-29_DR	.	Danio rerio	Tx1	21,1,43,23,3,44,11,4,2
Tx1-2B_SSa	.	Salmo salar	Tx1	1,43,44,11,4,2
Tx1-3_AAe	.	Aedes aegypti	Tx1	2,1,8,3,11,4,12
Tx1-3_AMi	.	Alligator mississippiensis	Tx1	16,1,26,8,3,24,4,11,12,2
Tx1-3_ASi	.	Alligator sinensis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-3_BF	.	Branchiostoma floridae	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-3_CGi	.	Crassostrea gigas	Tx1	21,1,26,3,11,4,2
Tx1-3_CPB	.	Chrysemys picta bellii	Tx1	43,2,8,44,24,11,12,1,26,14,4,16
Tx1-3_CQ	.	Culex quinquefasciatus	Tx1	2,8,11,12,93,1,14,4,3
Tx1-3_DR	.	Danio rerio	Tx1	21,1,26,3,11,4,2
Tx1-3_FoCa	.	Folsomia candida	Tx1	16,1,26,3,24,11,4,2
Tx1-3_FR	.	Takifugu rubripes	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-3_LCh	.	Latimeria chalumnae	Tx1	16,1,26,43,3,44,46,24,11,4,2
Tx1-3_LMi	.	Locusta migratoria	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-3_PM	.	Petromyzon marinus	Tx1	2,8,24,11,12,1,26,4,16,3
Tx1-3_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,11,1,26,4,3

Tx1-3_SPur	.	Strongylocentrotus purpuratus	Tx1	1,43,3,44,11,4,2
Tx1-3_SSa	.	Salmo salar	Tx1	93,1,8,14,26,3,11,4,12,2
Tx1-3_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-30_DR	.	Danio rerio	Tx1	1,3,2
Tx1-31_DR	.	Danio rerio	Tx1	1,3,4,2
Tx1-32_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-33_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-34_DR	.	Danio rerio	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-35_DR	.	Danio rerio	Tx1	
Tx1-36_DR	.	Danio rerio	Tx1	44,43
Tx1-37_DR	.	Danio rerio	Tx1	44,43
Tx1-38_DR	.	Danio rerio	Tx1	21,1,43,23,3,44,11,4,30,2
Tx1-39_DR	.	Danio rerio	Tx1	93,1,26,3,11,4,2
Tx1-4_AAe	.	Aedes aegypti	Tx1	92,2,8,11,12,42,1,14,4,3
Tx1-4_AMi	.	Alligator mississippiensis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-4_BF	.	Branchiostoma floridae	Tx1	43,2,44,24,11,12,1,26,14,4,16,3,46
Tx1-4_CGi	.	Crassostrea gigas	Tx1	138,2,24,11,1,26,14,139,140,4,16,141
Tx1-4_CPB	.	Chrysemys picta bellii	Tx1	1,8,3,46,12,2
Tx1-4_CQ	.	Culex quinquefasciatus	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-4_DR	.	Danio rerio	Tx1	21,1,26,43,3,44,11,4,2
Tx1-4_FoCa	.	Folsomia candida	Tx1	16,1,8,3,24,11,4,12,2
Tx1-4_FR	.	Takifugu rubripes	Tx1	1,26,3,11,4,2
Tx1-4_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46

Tx1-4_LMi	.	Locusta migratoria	Tx1 16,1,43,44,24,11,2
Tx1-4_PM	.	Petromyzon marinus	Tx1 1,26,3,11,4,2
Tx1-4_SK	.	Saccoglossus kowalevskii	Tx1 43,2,21,44,24,11,1,26,4,16,3
Tx1-4_SPur	.	Strongylocentrotus purpuratus	Tx1 16,1,26,43,3,44,82,24,11,4,2
Tx1-4_SSa	.	Salmo salar	Tx1 2,21,23,11,1,26,4,15,16,3
Tx1-40_DR	.	Danio rerio	Tx1
Tx1-41_DR	.	Danio rerio	Tx1 1,26,8,14,3,11,4,12,2
Tx1-42_DR	.	Danio rerio	Tx1
Tx1-43_DR	.	Danio rerio	Tx1 1,8,26,3,11,4,12,2
Tx1-44_DR	.	Danio rerio	Tx1 1,26,3,11,4,2
Tx1-45_DR	.	Danio rerio	Tx1 16,1,26,3,24,11,4,2
Tx1-46_DR	.	Danio rerio	Tx1 44,43
Tx1-47_DR	.	Danio rerio	Tx1 2,8,11,12,42,1,26,14,4,3
Tx1-48_DR	.	Danio rerio	Tx1 2,8,11,12,42,1,26,14,4,3
Tx1-49_DR	.	Danio rerio	Tx1 1,3,2
Tx1-4B_Crp	.	Crocodylus porosus	Tx1
Tx1-5_AAe	.	Aedes aegypti	Tx1 2,8,24,11,12,42,1,26,14,4,16,3
Tx1-5_AMi	.	Alligator mississippiensis	Tx1 16,1,26,3,24,11,4,2
Tx1-5_BF	.	Branchiostoma floridae	Tx1 2,8,24,11,12,1,26,14,4,16,3,46
Tx1-5_CGi	.	Crassostrea gigas	Tx1 16,1,43,26,3,44,11,4,2,15
Tx1-5_CQ	.	Culex quinquefasciatus	Tx1 16,1,3,24,11,4,12,92,2
Tx1-5_Crp	.	Crocodylus porosus	Tx1 1,8,26,3,11,4,12,2
Tx1-5_DR	.	Danio rerio	Tx1 21,1,26,23,3,11,4,2

Tx1-5_FR	.	Takifugu rubripes Tx1	1,26,43,3,44,11,4,2
Tx1-5_LCh	.	Latimeria chalumnae Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-5_LMi	.	Locusta migratoria Tx1	92,2,8,11,12,93,42,1,26,14,4,3
Tx1-5_PM	.	Petromyzon marinus Tx1	16,1,26,3,24,11,4,2
Tx1-5_SK	.	Saccoglossus kowalevskii Tx1	43,155,2,23,44,11,1,26,4,15,16,156,3
Tx1-5_SPur	.	Strongylocentrotus purpuratus Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-5_SSa	.	Salmo salar Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-50_DR	.	Danio rerio Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-51_DR	.	Danio rerio Tx1	
Tx1-52_DR	.	Danio rerio Tx1	21,1,26,43,3,44,11,4,2
Tx1-53_DR	.	Danio rerio Tx1	
Tx1-54_DR	.	Danio rerio Tx1	1,26,43,44,11,2
Tx1-55_DR	.	Danio rerio Tx1	16,1,26,43,8,3,44,24,11,4,12,2
Tx1-56_DR	.	Danio rerio Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-57_DR	.	Danio rerio Tx1	1,43,3,44,11,2
Tx1-58_DR	.	Danio rerio Tx1	1,8,26,14,3,11,4,12,2
Tx1-6_AAe	.	Aedes aegypti Tx1	16,1,26,3,24,11,4,2
Tx1-6_AMi	.	Crocodylidae Tx1	42,8,14,46,12
Tx1-6_BF	.	Branchiostoma floridae Tx1	2,8,24,11,12,42,1,26,14,4,16,3,46
Tx1-6_CGi	.	Crassostrea gigas Tx1	21,1,198,26,43,3,44,11,4,2
Tx1-6_CQ	.	Culex quinquefasciatus Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-6_Crp	.	Crocodylus porosus Tx1	
Tx1-6_DR	.	Danio rerio Tx1	1,26,43,3,44,11,4,2

Tx1-6_LCh	.	Latimeria chalumnae	Tx1 16,1,26,43,3,44,24,11,4,2
Tx1-6_LMi	.	Locusta migratoria	Tx1 16,1,26,3,24,11,4,2
Tx1-6_SK	.	Saccoglossus kowalevskii	Tx1 43,2,21,23,44,24,11,1,26,4,16,3
Tx1-6_SPur	.	Strongylocentrotus purpuratus	Tx1 1,26,43,3,44,11,4,2
Tx1-6_SSa	.	Salmo salar	Tx1 2,8,24,11,12,1,26,14,4,16,3
Tx1-6B_AMi	.	Alligator mississippiensis	Tx1
Tx1-7_AMi	.	Crocodylidae	Tx1
Tx1-7_BF	.	Branchiostoma floridae	Tx1 2,8,24,11,12,42,1,26,14,4,16,182,3,46
Tx1-7_CGi	.	Crassostrea gigas	Tx1 16,1,26,3,24,11,4,2
Tx1-7_CQ	.	Culex quinquefasciatus	Tx1 2,1,8,3,11,4,12,92
Tx1-7_Crp	.	Crocodylus porosus	Tx1
Tx1-7_DR	.	Danio rerio	Tx1 21,1,26,23,3,11,4,30,2
Tx1-7_LCh	.	Latimeria chalumnae	Tx1 2,1,43,3,44,46,11,4,12
Tx1-7_LMi	.	Locusta migratoria	Tx1 16,1,26,3,24,11,4,2
Tx1-7_SK	.	Saccoglossus kowalevskii	Tx1 1,26,3,11,4,2
Tx1-7_SSa	.	Salmo salar	Tx1 43,2,21,23,44,11,1,26,4,3
Tx1-8_AMi	.	Alligator mississippiensis	Tx1 16,1,8,26,3,24,11,4,12,2
Tx1-8_BF	.	Branchiostoma floridae	Tx1 16,1,26,43,3,44,24,11,4,2
Tx1-8_CGi	.	Crassostrea gigas	Tx1 138,2,11,1,26,139,140,4,15,16,3,141
Tx1-8_CQ	.	Culex quinquefasciatus	Tx1 42,1,8,14,3,11,4,12,2
Tx1-8_Crp	.	Crocodylus porosus	Tx1
Tx1-8_DR	.	Danio rerio	Tx1 1,26,3,11,4,2
Tx1-8_LCh	.	Latimeria chalumnae	Tx1 43,2,8,44,11,12,42,1,14,4,3,46

Tx1-8_LMi	.	Locusta migratoria	Tx1	142,16,1,26,3,24,11,4,2
Tx1-8_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-8_SSa	.	Salmo salar	Tx1	21,1,26,43,3,44,4,11,2
Tx1-9_AMi	.	Crocodylidae	Tx1	1,26,8,3,4,11,12,2
Tx1-9_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-9_CGi	.	Crassostrea gigas	Tx1	1,3,4,2
Tx1-9_CQ	.	Culex quinquefasciatus	Tx1	1,3,11,4,2
Tx1-9_DR	.	Danio rerio	Tx1	43,2,8,44,11,12,1,26,14,4,3
Tx1-9_LMi	.	Locusta migratoria	Tx1	16,1,8,26,3,24,11,4,12,2
Tx1-9_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-9_SSa	.	Salmo salar	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-N1_CQ	.	Culex quinquefasciatus	Tx1	8,14,12,92
Tx1-Sat-1_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,8,3,4,11,12,2,15
Tx1-TGTA-1_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,3,11,4,2,15
Tx1-U5-1_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-U5-2_SK	.	Saccoglossus kowalevskii	Tx1	43,2,44,11,12,1,26,14,4,3,46
UCON13	.	Euteleostomi	Penelope	
UCON3	.	Euteleostomi	SINE2/τPHK	
UCON3_CM	.	Callorhinchus milii	SINE2/τPHK	
UCON49	.	Mammalia	L2	
UCON49	.	Mammalia	L2	
UCON49	.	Mammalia	L2	
UCON55	.	Mammalia	SINE2/τPHK	

UCON55	.	Mammalia	SINE2/тРНК	
UCON55	.	Mammalia	SINE2/тРНК	
UCON82	.	Mammalia	RTE	
UCON82	.	Mammalia	RTE	
UCON82	.	Mammalia	RTE	
UCON86	.	Mammalia	L2	
UCON86	.	Mammalia	L2	
UCON86	.	Mammalia	L2	
UnaL2	.	Anguilla japonica	L2	16,1,3,24,4,11,2
UnaSINE1	.	Anguilla japonica	SINE2/тРНК	
UnaSINE2	.	Anguilla japonica	SINE2/тРНК	
Utopia-1_ACa	.	Acanthamoeba castellanii	NeSL	18,5,1,19,3,4,2
Utopia-1_ACar	каркас_474	Anolis carolinensis	NeSL	1,3,4,2
Utopia-1_AEc	.	Acromyrmex echinatior	NeSL	1,3,4,2
Utopia-1_AMi	.	Alligator mississippiensis	NeSL	1,3,4,2
Utopia-1_APi	.	Acyrthosiphon pisum	NeSL	1,3,4,2
Utopia-1_APl	.	Agrilus planipennis	NeSL	1,3,4,2
Utopia-1_CFl	.	Camponotus floridanus	NeSL	1,3,4,2
Utopia-1_CMy	.	Chelonia mydas	NeSL	1,3,4,2
Utopia-1_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-1_Crp	.	Crocodylus porosus	NeSL	1,3,4,2
Utopia-1_DPo	.	Dendroctonus ponderosae	NeSL	1,3,4,2
Utopia-1_DPu	.	Daphnia pulex	NeSL	1,3,4,2

Utopia-1_DYak	.	Drosophila yakuba	NeSL	1,3,4,2
Utopia-1_EBr	.	Eimeria brunetti	NeSL	
Utopia-1_EMi	.	Eimeria mitis	NeSL	1,3,4,2
Utopia-1_ENe	.	Eimeria necatrix	NeSL	5,1,19,3,4,2
Utopia-1_Gav	.	Gavialis gangeticus	NeSL	1,3,4,2
Utopia-1_GG1	.	Ganaspis	NeSL	1,3,4,2
Utopia-1_HAra	.	Hyaloperonospora arabidopsidis	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_HG	.	Heterodera glycines	NeSL	1,3,4,2
Utopia-1_HMM	.	Heliconius melpomene melpomene	NeSL	1,3,4,2
Utopia-1_HSal	.	Harpegnathos saltator	NeSL	1,3,4,2
Utopia-1_IS	.	Ixodes scapularis	NeSL	1,3,4,2
Utopia-1_LAl	.	Lasioglossum albipes	NeSL	1,3,4,2
Utopia-1_LFu	.	Ladona fulva	NeSL	1,3,4,2
Utopia-1_LV	AGCV01358106	Lytechinus variegatus	NeSL	18,5,1,19,3,4,2
Utopia-1_MRo	.	Megachile rotundata	NeSL	
Utopia-1_NVit	.	Nasonia vitripennis	NeSL	18,5,1,3,4,2
Utopia-1_PAlni	.	Phytophthora alni	NeSL	7,1,17,10,230,6,2,80
Utopia-1_PArrh	.	Pythium arrhenomanes	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PBa	.	Pogonomyrmex barbatus	NeSL	1,3,4,2
Utopia-1_PCa	.	Phytophthora capsici	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PCinn	.	Phytophthora cinnamomi	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PCu	AHJF01004292	Pseudoperonospora cubensis	NeSL	1,3,4,2
Utopia-1_PI	.	Phytophthora infestans	NeSL	6,2,10,80,1,230,4,5,19,17,3,7

Utopia-1_PInsi	.	Pythium insidiosum	NeSL	7,1,17,3,10,230,4,2,80
Utopia-1_PKern	.	Phytophthora kernoviae	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-1_PLate	.	Phytophthora lateralis	NeSL	7,1,17,3,10,4,2,80
Utopia-1_PMi	.	Patiria miniata	NeSL	18,5,1,19,3,4,2
Utopia-1_PPac	.	Pristionchus pacificus	NeSL	1,3,4,2
Utopia-1_PPini	.	Phytophthora pinifolia	NeSL	80,17,10,6,7
Utopia-1_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,2,6
Utopia-1_PRe	.	Panagrellus redivivus	NeSL	1,3,4,2
Utopia-1_PS	.	Phytophthora sojae	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PSi	.	Pelodiscus sinensis	NeSL	1,3,4,2
Utopia-1_PT	.	Parasteatoda tepidariorum	NeSL	18,5,19,1,3,4,2
Utopia-1_PU	ADOS01001321	Pythium ultimum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-1_PVexa	.	Phytopythium aff. vexans	NeSL	7,1,17,3,10,4,2,80
Utopia-1_SaPa	.	Saprolegnia parasitica	NeSL	1,3,4,2
Utopia-1_SDicl	.	Saprolegnia diclina	NeSL	1,3,4,2
Utopia-1_SM	.	Strigamia maritima	NeSL	1,3,4,2
Utopia-1_SP	AAGJ02140537	Strongylocentrotus purpuratus	NeSL	18,5,1,19,3,4,2
Utopia-1_TSP	.	Trichinella spiralis	NeSL	147,1,148,3,149,120,4,2
Utopia-1B_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-2_APi	.	Acyrthosiphon pisum	NeSL	1,3,4,2
Utopia-2_CMy	.	Chelonia mydas	NeSL	1,3,4,2
Utopia-2_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-2_DPu	.	Daphnia pulex	NeSL	1,3,4,2

Utopia-2_LFu	.	Ladona fulva	NeSL	
Utopia-2_PCa	.	Phytophthora capsici	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PI	.	Phytophthora infestans	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-2_PR	.	Phytophthora ramorum	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PS	.	Phytophthora sojae	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PU	.	Pythium ultimum	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-3_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-3_DPu	.	Daphnia pulex	NeSL	1,3,4,2
Utopia-3_LFu	.	Ladona fulva	NeSL	
Utopia-3_PCa	.	Phytophthora capsici	NeSL	
Utopia-3_PI	.	Phytophthora infestans	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-3_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-4_LFu	.	Ladona fulva	NeSL	
Utopia-4_PI	AATU01001281.1	Phytophthora infestans	NeSL	6,2,10,80,1,230,4,5,19,17,3,7
Utopia-4_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-5_LFu	.	Ladona fulva	NeSL	1,3,4,2
Utopia-5_PI	.	Phytophthora infestans	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-5_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-6_LFu	.	Ladona fulva	NeSL	1,3,4,2
UVIR_DV	.	Drosophila virilis	Jockey	16,1,3,11,4,2,15
Ves	Y19045Y19046Y19047Y1	Myotis daubentonii	SINE2/τPHK	
Ves2_ML	.	Myotis lucifugus	SINE2/τPHK	
Ves2B_ML	.	Myotis lucifugus	SINE2/τPHK	

Ves3_ML	Myotis lucifugus	SINE2/tPHK	
Ves4_ML	Myotis lucifugus	SINE2/tPHK	
Vingi-1_AAl	Atelerix albiventris	Vingi	16,1,3,11,4,2,15
Vingi-1_Acar	Anolis carolinensis	Vingi	16,1,3,11,4,2,15
Vingi-1_AMi	Crocodylidae	Vingi	
Vingi-1_APis	Acyrthosiphon pisum	Vingi	16,1,3,11,4,2,15
Vingi-1_BF	Branchiostoma floridae	Vingi	16,1,3,11,4,2,15
Vingi-1_BM	Bombyx mori	Vingi	16,1,3,11,4,2,15
Vingi-1_BTa	Bemisia tabaci	Vingi	16,1,3,11,4,2,15
Vingi-1_CE	Caenorhabditis elegans	Vingi	16,1,3,11,4,2,15
Vingi-1_CJ	Caenorhabditis japonica	Vingi	94,16,1,3,95,24,11,4,2
Vingi-1_CPB	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-1_Croc	Crocodylidae	Vingi	16,1,3,11,4,2,15
Vingi-1_EE	Erinaceus europaeus	Vingi	16,1,3,11,4,2,15
Vingi-1_Gav	Gavialis gangeticus	Vingi	1,3,4,2
Vingi-1_Hmel	Heliconius melpomene	Vingi	16,1,3,11,4,2,15
Vingi-1_HR	Helobdella robusta	Vingi	16,1,3,11,4,2,15
Vingi-1_Lch	Latimeria chalumnae	Vingi	1,3,11,4,2
Vingi-1_Lme	Latimeria menadoensis	Vingi	1,3,11,4,2
Vingi-1_LVa	Litopenaeus vannamei	Vingi	1,3,11,4,2
Vingi-1_PH	Parhyale hawaiensis	Vingi	16,1,3,11,4,2,15
Vingi-1_PMa	Petromyzon marinus	Vingi	16,1,3,11,4,2,15
Vingi-1_Pp	Phlebotomus papatasi	Vingi	16,1,3,11,4,2,15

Vingi-1_PSi	.	Pelodiscus sinensis	Vingi 1,3,4,2
Vingi-1_SPur	.	Strongylocentrotus purpuratus	Vingi 16,1,3,4,11,2,15
Vingi-1_Tcas	.	Tribolium castaneum	Vingi 16,1,3,11,4,2,15
Vingi-10_AMi	.	Alligator mississippiensis	Vingi
Vingi-10_CPB	.	Chrysemys picta bellii	Vingi
Vingi-10_LMi	.	Locusta migratoria	Vingi 1,3,4,2
Vingi-11_AMi	.	Alligator mississippiensis	Vingi 1,3,4,2
Vingi-11_CPB	.	Chrysemys picta bellii	Vingi
Vingi-11_LMi	.	Locusta migratoria	Vingi 16,1,3,4,11,2,15
Vingi-12_CPB	.	Chrysemys picta bellii	Vingi 16,1,3,11,4,2,15
Vingi-12_LMi	.	Locusta migratoria	Vingi 2,1
Vingi-13_CPB	.	Chrysemys picta bellii	Vingi
Vingi-13_LMi	.	Locusta migratoria	Vingi
Vingi-14_CPB	.	Chrysemys picta bellii	Vingi 16,1,3,11,4,2,15
Vingi-14_LMi	.	Locusta migratoria	Vingi 16,1,3,11,4,2,15
Vingi-15_LMi	.	Locusta migratoria	Vingi 1,3,11,4,2
Vingi-16_LMi	.	Locusta migratoria	Vingi 1,2
Vingi-17_LMi	.	Locusta migratoria	Vingi 1,3,4,2
Vingi-18_LMi	.	Locusta migratoria	Vingi 16,1,3,11,4,2,15
Vingi-19_LMi	.	Locusta migratoria	Vingi 16,1,3,11,4,2,15
Vingi-1B_PM	.	Petromyzon marinus	Vingi
Vingi-1N1_BM	.	Bombyx mori	Vingi
Vingi-1N1_EE	.	Erinaceus europaeus	Vingi

Vingi-1N10_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N11_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N2_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N3_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N4_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N5_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N6_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N7_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N8_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N9_EE	.	Erinaceus europaeus	Vingi	
Vingi-2_Acar	.	Anolis carolinensis	Vingi	16,1,3,11,4,2,15
Vingi-2_AMi	.	Alligator mississippiensis	Vingi	
Vingi-2_BF	.	Branchiostoma floridae	Vingi	16,1,3,4,11,2,15
Vingi-2_CE	.	Caenorhabditis elegans	Vingi	1,3,2
Vingi-2_CJap	.	Caenorhabditis japonica	Vingi	16,1,3,11,4,2,15
Vingi-2_CPB	.	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-2_Crp	.	Crocodylus porosus	Vingi	
Vingi-2_Gav	.	Gavialis gangeticus	Vingi	16,1,3,11,4,2,15
Vingi-2_HR	.	Helobdella robusta	Vingi	16,1,3,11,4,2,15
Vingi-2_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-2_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-2_LVa	.	Litopenaeus vannamei	Vingi	16,1,3,4,11,2,15
Vingi-2_PH	.	Parhyale hawaiensis	Vingi	16,1,11,2,15

Vingi-2_PM	.	Petromyzon marinus	Vingi	16,1,3,11,4,2,15
Vingi-2_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,11,4,2,15
Vingi-20_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-21_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-22_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-23_LMi	.	Locusta migratoria	Vingi	16,1,3,11,2,15
Vingi-24_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-25_LMi	.	Locusta migratoria	Vingi	16,1,3,11,2,15
Vingi-26_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-27_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-28_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-29_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-3_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-3_BF	.	Branchiostoma floridae	Vingi	1,3,4,2
Vingi-3_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-3_Crp	.	Crocodylus porosus	Vingi	
Vingi-3_Gav	.	Gavialis gangeticus	Vingi	1,3,4,2
Vingi-3_HR	.	Helobdella robusta	Vingi	16,1,3,11,4,2,15
Vingi-3_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-3_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-3_PH	.	Parhyale hawaiensis	Vingi	
Vingi-3_PM	.	Petromyzon marinus	Vingi	1,3,4,2
Vingi-3_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,11,4,2,15

Vingi-30_LMi	.	Locusta migratoria	Vingi	16,1,3,24,11,4,2
Vingi-31_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-32_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-33_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-35_LMi	.	Locusta migratoria	Vingi	
Vingi-36_LMi	.	Locusta migratoria	Vingi	
Vingi-37_LMi	.	Locusta migratoria	Vingi	16,1,3,24,4,11,2
Vingi-38_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-39_LMi	.	Locusta migratoria	Vingi	1,4,2
Vingi-4_AMi	.	Alligator mississippiensis	Vingi	16,1,11,4,2,15
Vingi-4_CPB	.	Chrysemys picta bellii	Vingi	1,3,4,2
Vingi-4_Crp	.	Crocodylus porosus	Vingi	
Vingi-4_HR	.	Helobdella robusta	Vingi	16,1,3,60,11,4,2,15
Vingi-4_LCh	.	Latimeria chalumnae	Vingi	16,1,3,24,11,4,2
Vingi-4_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-4_PM	.	Petromyzon marinus	Vingi	
Vingi-4_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,24,11,4,2
Vingi-40_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-41_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-42_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-43_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-44_LMi	.	Locusta migratoria	Vingi	
Vingi-45_LMi	.	Locusta migratoria	Vingi	1,3,4,2

Vingi-46_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-47_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-48_LMi	.	Locusta migratoria	Vingi	
Vingi-49_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-5_AMi	.	Alligator mississippiensis	Vingi	16,1,3,11,4,2,15
Vingi-5_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-5_Crp	.	Crocodylus porosus	Vingi	
Vingi-5_LMi	.	Locusta migratoria	Vingi	16,1,3,24,11,4,2
Vingi-5_PM	.	Petromyzon marinus	Vingi	1,3,4,2
Vingi-6_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-6_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-6_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-6_LMi	.	Locusta migratoria	Vingi	1,3,11,4,2
Vingi-6_PM	.	Petromyzon marinus	Vingi	
Vingi-7_AMi	.	Alligator mississippiensis	Vingi	
Vingi-7_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-7_Lch	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-7_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-7_PM	.	Petromyzon marinus	Vingi	
Vingi-8_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-8_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-8_Lch	.	Latimeria chalumnae	Vingi	
Vingi-8_LMi	.	Locusta migratoria	Vingi	16,1,24,11,4,2

Vingi-8_PM	.	Petromyzon marinus	Vingi	
Vingi-9_AMi	.	Alligator mississippiensis	Vingi	
Vingi-9_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-9_Lch	.	Latimeria chalumnae	Vingi	
Vingi-9_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-9_PM	.	Petromyzon marinus	Vingi	1,3,4,2
VINSINE	.	Varanus indicus	SINE2/тРНК	
VLINE1_VV	.	Vitis vinifera	L1	1,43,3,44,4,2
VLINE1_VV	.	Vitis vinifera	L1	1,43,3,44,4,2
VLINE10_VV	.	Vitis vinifera	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
VLINE11_VV	.	Vitis vinifera	L1	16,1,75,43,3,76,44,24,11,4,2
VLINE2_VV	.	Vitis vinifera	L1	16,1,3,24,11,4,2
VLINE2_VV	.	Vitis vinifera	L1	16,1,3,24,11,4,2
VLINE3_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,2
VLINE3_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,2
VLINE4_VV	.	Vitis vinifera	L1	1,3,4,2
VLINE4_VV	.	Vitis vinifera	L1	1,3,4,2
VLINE5_VV	.	Vitis vinifera	L1	94,1,3,95,11,4,2
VLINE5_VV	.	Vitis vinifera	L1	94,1,3,95,11,4,2
VLINE6_VV	.	Vitis vinifera	L1	1,43,75,3,44,76,11,4,2
VLINE6_VV	.	Vitis vinifera	L1	1,43,75,3,44,76,11,4,2
VLINE7_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,4,11,2
VLINE8_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,4,2

VLINE9_VV	.	Vitis vinifera	L1	16,1,75,43,3,76,44,24,11,4,2
W_As	.	Adineta	Penelope	3,1
W_Av	.	Adineta vaga	Penelope	1,3,130
W1_As	.	Adineta	Penelope	1,3
W1_Av	.	Adineta vaga	Penelope	1,3,167,130,169
W2_As	.	Adineta	Penelope	1,3
W2_Av	.	Adineta vaga	Penelope	1,3
Waldo-1_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-1_CQ	.	Culex quinquefasciatus	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-2_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo-3_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo-4_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-5_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-6_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo1_AG	AB090814	Anopheles gambiae	R1	2,8,11,86,12,1,4,15,16,3
WaldoAg2	AB090815	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
WaldoFs1	AB078939	Forficula scudderi	R1	16,1,3,4,11,2,15
WALLSI1	.	Notamacropus eugenii	SINE	
WALLSI1A	.	Notamacropus eugenii	SINE	
WALLSI2	.	Notamacropus eugenii	SINE2/тРНК	
WALLSI3	.	Notamacropus eugenii	SINE	
WALLSI4	.	Notamacropus eugenii	SINE2/тРНК	
WHLN1_TA	AB016116	Triticum aestivum	L1	

WHLN2_TA	AB016117	Triticum aestivum	L1	
WHLN3_TA	AB016118	Triticum aestivum	L1	
WORF_DMi	AY144572	Drosophila miranda	CR1	1,3,11,4,2
WRETRO	.	Notamacropus eugenii	Не содержащий LTR ретротранспозон	
WSINE1	.	Notamacropus eugenii	SINE2/тРНК	
WSINE1a	.	Notamacropus eugenii	SINE2/тРНК	
X1_LINE	.	Euteleostomi	CR1	
X11_LINE	.	Mammalia	RTE	
X11_LINE	.	Mammalia	RTE	
X11_LINE	.	Mammalia	RTE	
X12_LINE	.	Mammalia	Vingi	
X12_LINE	.	Mammalia	Vingi	
X12_LINE	.	Mammalia	Vingi	
X13_LINE	.	Mammalia	Crack	
X15_LINE	.	Mammalia	L2	
X15_LINE	.	Mammalia	L2	
X15_LINE	.	Mammalia	L2	
X17_LINE	.	Mammalia	CR1	
X17_LINE	.	Mammalia	CR1	
X17_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	

X19_LINE	.	Mammalia		CR1
X19_LINE	.	Mammalia		CR1
X19_LINE	.	Mammalia		CR1
X2_LINE	.	Euteleostomi		CR1
X20_LINE	.	Mammalia		CR1
X20_LINE	.	Mammalia		CR1
X20_LINE	.	Mammalia		CR1
X21_LINE	.	Mammalia		CR1
X21_LINE	.	Mammalia		CR1
X21_LINE	.	Mammalia		CR1
X24_LINE	.	Mammalia		L2
X24_LINE	.	Mammalia		L2
X24_LINE	.	Mammalia		L2
X3_LINE	.	Mammalia		RTE
X4_LINE	.	Vertebrata	R4	
X5A_LINE	.	Euteleostomi		CR1
X5B_LINE	.	Euteleostomi		CR1
X6A_LINE	.	Euteleostomi		CR1
X6B_LINE	.	Euteleostomi		CR1
X7A_LINE	.	Euteleostomi		CR1
X7B_LINE	.	Euteleostomi		CR1
X7C_LINE	.	Euteleostomi		CR1
X7D_LINE	.	Euteleostomi		CR1

X8_LINE	.	Euteleostomi	CR1	
X9_LINE	.	Euteleostomi	L1	
YJ_As	.	Adineta	Penelope	3,1
Ylli	YLI319752	Yarrowia lipolytica	L1	16,1,26,3,24,11,4,2
Yoyo-1_PH	.	Parhyale hawaiensis	CR1	1,3,82,4,11,2
YURE_CSa	.	Ciona savignyi	NeSL	5,1,3,4,2
YURE-2_Cis	.	Ciona savignyi	R2	5,1,19,3,4,2
YURECi	.	Ciona intestinalis	NeSL	5,1,3,4,2
YW_As	.	Adineta	Penelope	1,3
ZENON_BM	.	Bombyx mori	CR1	16,1,3,11,4,2,15
Zenon-1_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-1_PPo	.	Papilio polytes	CR1	1,3,4,2
ZENON-1_PXu	.	Papilio xuthus	CR1	
ZENON-1B_PPo	.	Papilio polytes	CR1	
Zenon-2_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-2_PPo	.	Papilio polytes	CR1	
ZENON-2_PXu	.	Papilio xuthus	CR1	
ZENON-2B_PPo	.	Papilio polytes	CR1	
ZENON-2C_PPo	.	Papilio polytes	CR1	
ZENON-2D_PPo	.	Papilio polytes	CR1	
ZENON-2E_PPo	.	Papilio polytes	CR1	
Zenon-3_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-3_PPo	.	Papilio polytes	CR1	

ZENON-3_PXu	.	Papilio xuthus	CR1	
ZENON-4_PPo	.	Papilio polytes	CR1	
ZENON-4_PXu	.	Papilio xuthus	CR1	
ZENON-5_PPo	.	Papilio polytes	CR1	
ZENON-5_PXu	.	Papilio xuthus	CR1	
Zepp	AB008896, D83919	Chlorella vulgaris	L1	16,1,26,24,11,4,2
Zorro	AF254443	Candida albicans	L1	1,3,11,4,2

Таблица Y. Иллюстративные домены InterPro из таблицы X *(следующие страницы)*

NID	Домен	Название	Описание
1	IPR000477	Домен обратной транскриптазы	"Применение РНК-матрицы для получения ДНК с целью интеграции в геном хозяина и использования клетки-хозяина является стратегией, применяемой при репликации ретроидных элементов, таких как ретровирусы и бактериальные ретроны. Фермент, катализирующий полимеризацию, представляет собой РНК-направленную ДНК-полимеразу или обратную транскриптазу (RT) (). Обратная транскриптаза встречается в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Ретровирусная обратная транскриптаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Обнаружение ретроэлементов у прокариот поднимает интригующие вопросы, касающиеся их роли у бактерий, происхождения и эволюции обратных транскриптаз, а также того, являются ли бактериальные обратные транскриптазы старшими, чем эукариотические обратные транскриптазы. Было определено несколько кристаллических структур домена обратной транскриптазы (RT)."
2	PF00078	Обратная транскриптаза (РНК-зависимая ДНК-полимераза)	"Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы."

3	PS50878	Профиль каталитического домена обратной транскриптазы (RT)	"Ретровирусные обратные транскриптазы (EC 2.7.7.49) (прототипы RT) проявляют три вида ферментативной активности: РНК-зависимой ДНК-полимеразы, рибонуклеазы H и ДНК-зависимой ДНК-полимеразы. Они используются для копирования генома плюс-нити РНК для получения минус-нити ДНК, удаления РНК-матрицы и синтеза плюс-нити ДНК с использованием минус-нити ДНК в качестве матрицы соответственно [PMID:7526778]. Обе полимеразные активности осуществляются посредством домена RT, тогда как рибонуклеазная активность обеспечивается доменом РНКазы H. Было определено несколько кристаллических структур домена RT (см., например, {PDB:3HVT}) [PMID:1377403]. Структурно домен RT можно разделить на четыре субдомена. Его анатомическое сходство с правой рукой привело к тому, что субдомены были названы "пальцами", "ладонью" и "большим пальцем". Четвертый субдомен находится между остальной частью RT и доменом РНКазы H, что привело к тому, что называют соединяющим субдоменом. Домен "ладонь" содержит каталитическую триаду, ключевую для полимеразной активности, и его складчатая структура напоминает структуру соответствующего каталитического домена в других ДНК- и РНК-полимеразах. Положение "пальцев" и "большого пальца" определяло глубокую расщелину с активными остатками полимеразы в ее основании. Белки, которые, как известно, содержат домен RT, перечислены ниже. - Ретровирусная обратная транскриптаза. - Каталитическая субъединица теломеразы. - Эукариотические ретротранспозоны. - Ретроэлемент Escherichia coli - ретрон, отвечающий за синтез многокопийной однонитевой ДНК

			(msDNA). - Бактериальный мобильный интрон группы II, который может обеспечивать обратный сплайсинг в субстрат-мишень ДНК. Субдомены "палец", "большой палец" и соединяющий субдомен менее консервативны среди всех подсемейств доменов RT [PMID: 9703504] и не могут быть идентифицированы посредством способа профилирования. Таким образом, профиль, который разработали авторы настоящего изобретения, направлен на каталитическую область (субдомен "ладонь")".
4	cd01650	RT_nLTR: ретротранспозон, не содержащий LTR (длинных концевых повторов), и ретровирусная обратная транскриптаза (RT), не содержащая LTR.	"RT_nLTR: Ретротранспозон, не содержащий LTR (длинных концевых повторов), и ретровирусная обратная транскриптаза (RT), не содержащая LTR. Это подсемейство содержит как ретротранспозоны, не содержащие LTR, так и ретровирусные RT, не содержащие LTR. RT катализируют превращение однонитевой РНК в двухнитевую ДНК для интеграции в хромосомы хозяина. RT представляет собой многофункциональный фермент с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). [PMID: 1698615, PMID: 10605110, PMID: 10628860, PMID: 11734649, PMID: 12117499, PMID: 12777502, PMID: 14871946, PMID: 15939396, PMID: 16271150, PMID: 16356661, PMID: 2463954, PMID: 3040362, PMID: 3656436, PMID: 7512193, PMID: 7534829, PMID: 7659515, PMID: 8524653, PMID: 9190061, PMID: 9218812, PMID: 9332379, PMID: 9364772, PMID: 8828137]"
5	IPR013087	"Цинковый палец" типа C2H2	"Цинковые пальцы" (Znf) типа C2H2 (классические) были первым классом, который был охарактеризован. Они содержат короткую бета-шпильку и альфа-спираль

			(структура бета/бета/альфа), где один атом цинка удерживается на месте остатками Cys(2)His(2) (C2H2) в тетраэдрическом массиве. Znf C2H2 можно разделить на три группы в зависимости от количества и паттерна "пальцев": тройной C2H2 (связывает один лиганд), несколько смежных C2H2 (связывает несколько лигандов) и разделенные пары C2H2. Znf C2H2 являются наиболее распространенными ДНК-связывающими мотивами, обнаруживаемыми в эукариотических факторах транскрипции, и они также были идентифицированы у прокариот. Факторы транскрипции обычно содержат несколько Znf (каждый с консервативной структурой бета/бета/альфа), способных образовывать множество контактов вдоль ДНК, где мотивы Znf C2H2 распознают последовательности ДНК, связываясь с большой бороздкой ДНК посредством короткой альфа-спирали в Znf, причем Znf охватывает 3-4 основания ДНК. Znf C2H2 также могут связываться с РНК и белковыми мишенями. Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают

			ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет классический домен "цинковый палец" C2H2".
6	PS50879	Профиль домена РНКазы H	"Рибонуклеаза H (РНКаза H) (EC 3.1.26.4) распознает и расщепляет нить РНК гетеродуплексов РНК-ДНК. У прокариот РНКаза H участвует в удалении РНК гибридов РНК-ДНК, особенно праймеров фрагментов Оказаки при синтезе ДНК с отстающей нитью [PMID:2177653]. Дополнительная функция бактериального домена РНКазы H

			заключается в создании праймеров для инициации репликации ДНК [PMID:2446774]. Домен РНКазы H также присутствует в ретровирусных белках обратных транскриптазах, где он выполняет три функции: (a) образование полипуринового тракта, который выступает в качестве праймера для синтеза плюс-нити ДНК, (b) удаление РНК-праймеров из формирующихся минус- и плюс-нитей ДНК и (c) разрушение исходной РНК-матрицы. В случае с HIV инактивация РНКазы H делает вирус неинфекционным, что делает эту ферментативную активность мишенью для терапии AIDS [PMID: 2177653]. У эукариот белок РНКаза H1 присутствует в ядрах, но он также был выделен из митохондрий многих организмов. У мышей инактивация гена РНКазы H1 приводит к остановке развития приблизительно в день E8.5 из-за неспособности образовывать mtDNA, что указывает на важную роль РНКазы H1 во время эмбриогенеза млекопитающих [PMID: 12667461]. Была расшифрована 3D-структура домена РНКазы H из различных бактерий и ретровирусов (см., например, {PDB:1RNH}) [PMID:2169648,PMID:8108376,PMID:1707186]. Все они имеют четыре бета-нити и от четырех до пяти альфа-спиралей. Белок РНКаза H1 Escherichia coli связывает один кофактор, представляющий собой ион Mg(2+), в активном центре фермента. Двухвалентный катион связывается карбоксильными группами четырех кислотных остатков - Asp-10, Glu-48, Asp-70 и Asp-134 [PMID:8108376]. Первые три кислотных остатка являются высококонсервативными во всех последовательностях

			бактериальных и ретровирусных РНКаз H. Разработанный авторами настоящего изобретения профиль охватывает весь домен РНКазы H".
7	IPR012337	Суперсемейство белков, подобных рибонуклеазе H	"Каталитический домен нескольких полинуклеотидилтрансфераз имеет сходную структуру, состоящую из 3-слойной укладки типа альфа/бета/альфа, которая содержит смешанные бета-листы, что позволяет предположить, что они имеют сходный механизм катализа. Полинуклеотидилтрансферазы, содержащие этот домен, включают рибонуклеазу H класса I (РНКазу HI) и класса II (РНКазу HII), РНКазу HIV (домен обратной транскриптазы), ретровирусную интегразу (каталитический домен), транспозазу Mu (сердцевинный домен), ингибитор транспозазы Tn5 (содержащий дополнительные субдомены, полностью состоящие из альфа-спиралей), DnaQ-подобные 3'-5'-экзонуклеазы (экзонуклеазные домены), резольвазу RuvC и митохондриальную резольвазу ydc2 (каталитический домен). Это суперсемейство также включает домен YqgF, описанный как подобный РНКазе H и типичным примером которого служит белок YqgF Escherichia coli. Белки, содержащие домен YqgF, согласно предсказаниям представляют собой рибонуклеазы или резольвазы, исходя из гомологии с резольвазами для структур Холлидея RuvC."
8	PS50158	Профиль "цинкового пальца" типа CCHC	"Домен "цинковый палец" CCHC из 18 остатков в основном обнаруживается в нуклеокапсидном белке ретровируса. Он необходим для упаковки вирусного генома и для процесса раннего инфицирования [PMID:2695083,PMID:10074407]. Он был также обнаружен в эукариотических белках, участвующих в связывании РНК или

			связывании однонитевой ДНК. Некоторые белки, которые, как известно, содержат "цинковый палец" CCHC, перечислены ниже. - Ретровирусный белок GAG. - Нуклеокапсидный белок. - Белок CNBP человека. CNBP участвует в стерол-опосредованной репрессии. Также было показано, что он связывает 5'-нетранслируемую область рибосомных белков мРНК (область, которая, как известно, важна для регуляции трансляции). - BYR3 дрожжей; гомолог CNBP. - Белок glh-1 Caenorhabditis elegans; вероятная ATP-связывающая РНК-хеликаза. - Главный гексамерсвязывающий белок Leishmania (HEXP). Он связывает 5'-нетранслируемую область гена GP63. - Белок PolX транспозона TNT1 табака. Связанный с ретровирусами полипротеин pol. Разработанный авторами настоящего изобретения профиль охватывает весь домен".
9	cd09276	Не содержащий LTR домен РНКазы HI обратных транскриптаз	"Рибонуклеаза H (РНКаза H) подразделяется на два семейства: тип 1 (прокариотическая РНКаза HI, эукариотическая РНКаза H1 и вирусная РНКаза H) и тип 2 (прокариотическая РНКаза HII и HIII и эукариотическая РНКаза H2). Рибонуклеаза HI (РНКаза HI) представляет собой эндонуклеазу, которая расщепляет нить РНК гибрида РНК/ДНК неспецифичным в отношении последовательности образом. РНКаза H широко представлена в различных организмах, включая бактерии, археи и эукариот. РНКаза HI также наблюдалась в качестве домена, прилегающего к гену обратной транскриптазы в ретровирусах, содержащих длинные концевые повторы (LTR) ретротранспозонах и не содержащих LTR ретротранспозонах. РНКаза HI в содержащих LTR ретротранспозонах осуществляет разрушение исходной РНК-

			матрицы, образование полипуринового тракта (праймера для синтеза плюс-нити ДНК) и окончательное удаление РНК-праймеров из вновь синтезированных минус- и плюс-нитей. Каталитические остатки для ферментативной активности РНКазы H - три остатка аспарагиновой кислоты и один остаток глутаминовой кислоты (DEDD) - неизменны во всех доменах РНКазы H. Положение домена РНКазы в не содержащих LTR и содержащих LTR транспозонах находится на карбоксильном конце домена обратной транскриптазы (RT), и их РНКазные домены группируются вместе, что указывает на общее эволюционное происхождение. Многие не содержащие LTR транспозоны утратили домен РНКазы, поскольку их активность сосредоточена в ядре, и может быть достаточно клеточной РНКазы; однако содержащие LTR ретротранспозоны всегда кодируют свой собственный домен РНКазы, поскольку в цитоплазме необходима РНКазная активность в частицах РНК-белковых комплексов. Ингибиторы РНКазы H исследовались в качестве мишени для лекарственных средств для лечения HIV, поскольку инактивация РНКазы H ингибирует обратную транскрипцию. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198, PMID: 11435400]"
10	IPR002156	Домен рибонуклеазы H	"Домен РНКазы H отвечает за гидролиз РНК-части гибридов РНК x ДНК, и эта активность требует присутствия двухвалентных катионов (Mg2+ или Mn2+), которые связывают его активный центр. Этот домен является частью большого семейства гомологичных ферментов РНКаз H, из которых лучше всего охарактеризован белок РНКаза HI из Escherichia coli. Предсказания вторичной структуры ферментов из E. coli,

			дрожжей, печени человека и различных ретровирусов (таких как вирус саркомы Рауса и спумавирусы) подтвердили в каждом случае пять бета-нитей (от 1 до 5) и четыре или пять альфа-спиралей (A, B/C, D, E), которые были идентифицированы с помощью кристаллографии в домене РНКазы H обратной транскриптазы вируса иммунодефицита человека 1 (HIV-1) и в РНКазе H E. coli. Обратная транскриптаза (RT) представляет собой модульный фермент, несущий активность полимеразы и рибонуклеазы H (РНКазы H) в отдельных доменах. Обратная транскриптаза (RT) превращает однонитевой РНК-геном ретровируса в двухнитевую ДНК-копию для интеграции в геном хозяина. Этот процесс требует активности рибонуклеазы H, а также активности РНК- и ДНК-направленной ДНК-полимеразы. Ретровирусная РНКаза H синтезируется как часть полипротеина POL, который содержит аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Бактериальная РНКаза H катализирует эндонуклеолитическое расщепление до сложного 5'-фосфомоноэфира, воздействуя на гибриды РНК-ДНК. Была расшифрована 3D-структура домена РНКазы H из различных бактерий и ретровирусов. Все они имеют четыре бета-нити и от четырех до пяти альфа-спиралей. Белок РНКаза H1 E. coli связывает один кофактор, представляющий собой ион Mg2+, в активном центре фермента. Двухвалентный катион связывается карбоксильными группами четырех кислотных остатков - Asp-10, Glu-48, Asp-70 и

			Asp-134. Первые три кислотных остатка являются высококонсервативными во всех последовательностях бактериальных и ретровирусных РНКаз H."
11	IPR036691	Суперсемейство эндонуклеаз/экзонуклеаз/фосфатаз	"Это суперсемейство доменов обнаруживается в большом количестве белков, включая магнийзависимые эндонуклеазы и фосфатазы, участвующие во внутриклеточной передаче сигналов. Этот домен обнаруживается во многих белках, включая белки AP-эндонуклеазы (лиазу (апуринового или апиримидинового сайта) ДНК), дезоксирибонуклеазу-1 (ДНКазу I;), синаптоянин инозитол-1,4,5-трифосфатфосфатазу () и сфингомиелиназу ().Структура домена ДНКазы I содержит укладку типа "бета-сэндвич" и дупликацию мотива альфа+бета".
12	IPR001878	"Цинковый палец" типа CCHC"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев".

			Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет домены "цинковые пальцы" типа CysCysHisCys (CCHC) и имеет последовательность C-X2-C-X4-H-X4-C, где X может представлять собой любую аминокислоту, а число указывает количество остатков. Эти домены "цинковые пальцы" CCHC из 18 остатков в основном обнаруживаются в нуклеокапсидном белке ретровирусов. Они необходимы для упаковки вирусного генома и для процесса раннего инфицирования. Они были также обнаружены в эукариотических белках, участвующих в связывании РНК или связывании однонитевой ДНК.

13	PF00075	РНКаза Н	"РНКаза Н расщепляет нить РНК гибрида РНК/ДНК. Важный фермент в цикле репликации ретровирусов, часто обнаруживаемый как домен, ассоциированный с обратными транскриптазами. Структура представляет собой смешанную укладку альфа+бета с тремя слоями a/b/a."
14	IPR036875	"Суперсемейство доменов "цинковые пальцы" типа CCHC"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают

			белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет суперсемейство доменов "цинковые пальцы" типа CysCysHisCys (CCHC) и имеет последовательность C-X2-C-X4-H-X4-C, где X может представлять собой любую аминокислоту, а число указывает количество остатков. Эти домены "цинковые пальцы" CCHC из 18 остатков в основном обнаруживаются в нуклеокапсидном белке ретровирусов. Они необходимы для упаковки вирусного генома и для процесса раннего инфицирования. Они были также обнаружены в эукариотических белках, участвующих в связывании РНК или связывании однонитевой ДНК.
15	PF14529	Эндонуклеаза-обратная транскриптаза	"Этот домен представляет эндонуклеазную область ретротранспозонов ряда бактерий, архей и эукариот. Эти ферменты в основном относятся к классу EC:2.7.7.49".
16	IPR0051 35	Эндонуклеаза/экзонуклеаза/фосфатаза	"Этот домен обнаруживается в большом количестве белков, включая магнийзависимые эндонуклеазы и фосфатазы, участвующие во внутриклеточной передаче сигналов. Белки, в которых обнаруживается этот домен, включают: белки

			AP-эндонуклеазы (), белки ДНКазы I (), синаптоянин инозитол-1,4,5-трифосфатфосфатазу () и сфингомиелиназу ()."
17	IPR036397	Суперсемейство рибонуклеазы H	"Каталитический домен нескольких полинуклеотидилтрансфераз имеет сходную структуру, состоящую из 3-слойной укладки типа альфа/бета/альфа, которая содержит смешанные бета-листы, что позволяет предположить, что они имеют сходный механизм катализа. Полинуклеотидилтрансферазы, содержащие этот домен, включают рибонуклеазу H класса I (РНКазу HI) и класса II (РНКазу HII), РНКазу HIV (домен обратной транскриптазы), ретровирусную интегразу (каталитический домен), транспозазу Mu (сердцевинный домен), ингибитор транспозазы Tn5 (содержащий дополнительные субдомены, полностью состоящие из альфа-спиралей), DnaQ-подобные 3'-5'-экзонуклеазы (экзонуклеазные домены), резольвазу RuvC и митохондриальную резольвазу ydc2 (каталитический домен)."
18	PS00028	Предполагаемая нуклеаза HARBI1	NA
19	PS50157	Профиль домена "цинковый палец" типа C2H2	"Домены "цинковые пальцы" [PMID:3125980,PMID:3292287,PMID:4040853,PMID:3124104] представляют собой белковые структуры, связывающие нуклеиновые кислоты, впервые идентифицированные в факторе транскрипции TFIIIA у Xenopus. С тех пор эти домены были обнаружены во многих белках, связывающих нуклеиновые кислоты. Домен "цинковый палец" состоит из 25-30 аминокислотных остатков. На обоих концах домена расположены два остатка цистеина или гистидина, которые

			участвуют в тетраэдрической координации атома цинка. Было высказано предположение, что такой домен взаимодействует приблизительно с пятью нуклеотидами. Схематическое изображение домена "цинковый палец" показано ниже: x x x x x x x x x x x x C H x \ / x x Zn x x / \ x C H x x x x x x x x x x Многие классы "цинковых пальцев" характеризуются по количеству и положению остатков гистидина и цистеина, участвующих в координации атома цинка." В первом охарактеризованном классе, называемом C2H2, первая пара координирующих цинк остатков представляет собой остатки цистеина, а вторая пара представляет собой остатки гистидина. Ряд экспериментальных отчетов продемонстрировал цинк-зависимое свойство связывания ДНК или РНК у некоторых представителей этого класса. Некоторые белки, которые, как известно, содержат "цинковые пальцы" C2H2-типа, перечислены ниже. Авторы настоящего изобретения указали в скобках количество областей "цинковых пальцев", обнаруживаемых в каждом из этих белков; символ "+" указывает, что доступны только частичные данные о последовательности и что могут присутствовать дополнительные домены "пальцы". - Saccharomyces cerevisiae: ACE2 (3), ADR1 (2), AZF1 (4), FZF1 (5), MIG1 (2), MSN2 (2), MSN4 (2), RGM1 (2), RIM1 (3), RME1 (3), SFP1 (2), SSL1 (1), STP1 (3), SWI5 (3), VAC1 (1) и ZMS1 (2). - Emericella nidulans: brlA (2), creA (2). - Drosophila: AEF-1 (4), Cf2 (7), ci-D (5), Disconnected (2), Escargot (5), Glass (5), Hunchback (6), Kruppel (5), Kruppel-H (4+), Odd-skipped (4), Odd-paired (4), Pep (3), Snail (5), Spalt-major (7), Serependity, локус бета (6), дельта (7), h-1 (8), Suppressor of hairy wing su(Hw) (12), Suppressor of variegation

			suvar(3)7 (5), Teashirt (3) и Tramtrack (2). - Xenopus: фактор транскрипции TFIIIA (9), p43 из частицы RNP (9), Xfin (37 !!), Xsna (5), гаструла от XlcGF5.1 до XlcGF71.1 (от 4+ до 11+), ооцит от XlcOF2 до XlcOF22 (от 7 до 12). - Млекопитающие: базонуклин (6), BCL-6/LAZ-3 (6), эритроидный krueppel-подобный фактор транскрипции (3), факторы транскрипции Sp1 (3), Sp2 (3), Sp3 (3) и Sp(4) 3, репрессор транскрипции YY1 (4), белок опухоли Вильмса (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), ZfX (13), ZfY (13), Zfp-35 (18), ZNF7 (15), ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3). В дополнение к консервативным остаткам лиганда цинка было показано [PMID: 8129873], что ряд других положений также важен для структурной целостности "цинковых пальцев" C2H2. Наилучшее консервативное положение находится через четыре остатка после второго цистеина; обычно это ароматический или алифатический остаток. Также был разработан профиль, который охватывает весь домен".
20	PF17490	dsRBD-подобный домен подвижного элемента L1	Эта запись представляет домен, подобный домену связывания двухнитевой РНК.
21	IPR004244	"Транспозаза L1"	"Многие элементы L1 человека способны к ретротранспозиции. Было показано, что некоторые из них проявляют активность обратной транскриптазы (RT), хотя функция многих пока неизвестна."

22	PF08333	Белок с неизвестной функцией (DUF1725)	Это семейство включает множество эукариотических последовательностей и одну бактериальную последовательность. Многие из его представителей аннотированы как предполагаемые ретротранспозоны L1 или гомологи обратной транскриптазы LINE-1. Рассматриваемая область обнаруживается повторяющейся у некоторых представителей семейства.
23	IPR042497	"Суперсемейство транспозаз L1"	"Это суперсемейство представляет консервативный компонент в белке ORF1 ретротранспозона, представляющего собой длинный диспергированный ядерный элемент 1 (LINE-1 или L1). Этот домен обозначается как RRM, что означает мотив распознавания РНК, и он участвует в связывании РНК. Для этого требуется кооперация между этим доменом RRM и С-концевым доменом, несмотря на то, что эти два домена не контактируют друг с другом."
24	PF03372	Семейство эндонуклеаз/экзонуклеаз/фосфатаз	"Это большое семейство белков включает магнийзависимые эндонуклеазы и большое количество фосфатаз, участвующих во внутриклеточной передаче сигналов [PMID:10838565]. Это семейство включает: белки АР-эндонуклеазы EC:4.2.99.18, например, Swiss:P27695, белки ДНКазы I, EC:3.1.21.1, например, Swiss:P24855, синаптоянин инозитол-1,4,5-трифосфатфосфатазу EC:3.1.3.56 Swiss:O43426, сфингомиелиназу EC:3.1.4.12 Swiss:P11889 и ноктурнин Swiss:O35710."
25	IPR035300	"dsRBD-подобный домен подвижного элемента L1"	Эта запись представляет домен, подобный домену связывания двухнитевой РНК подвижных элементов LINE-1 (L1).
26	cd09076	"Эндонуклеазный домен (L1-EN) не содержащего LTR	"Это семейство содержит эндонуклеазный домен (L1-EN) не содержащего LTR ретротранспозона LINE-1 (L1) и родственные домены, включая эндонуклеазу

		ретротранспозона LINE-1 (L1) и родственные домены"	Xenopus laevis Tx1. Эти ретротранспозоны принадлежат к подтипу 2 клады L1. LINE можно разделить на два подтипа. Подтип 2 имеет две ORF: вторая (ORF2) кодирует модульный белок, состоящий из N-концевого домена апуриновой/апиримидиновой эндонуклеазы (EN), центральной обратной транскриптазы и домена, подобного "цинковому пальцу", на С-конце. Элементы LINE-1/L1 (полноразмерные и усеченные) составляют приблизительно 17% человеческого генома. Эта эндонуклеаза вносит однонитевые разрывы в геномную ДНК в консенсусной последовательности-мишени 5'TTTT-AA3', образуя 3'-гидроксильный конец рибозы в качестве праймера для обратной транскрипции ассоциированной РНК-матрицы. Эта подгруппа также включает эндонуклеазу Tx1 Xenopus laevis - другой представитель клады L1. Это семейство принадлежит к большому суперсемейству EEP (экзонуклеаз/эндонуклеаз/фосфатаз), которое содержит функционально разнообразные ферменты, имеющие общий каталитический механизм расщепления фосфодиэфирных связей. [PMID: 15274918, PMID: 17626046, PMID: 9922177, PMID: 12832649, PMID: 11810275, PMID: 8945517, PMID: 11700292, PMID: 15983781, PMID: 16336743, PMID: 18767160, PMID: 12870715]"
27	PF17489	Домен тримеризации подвижного элемента L1	Эта запись представляет домен тримеризации.
28	IPR042566	"С-концевой домен подвижного элемента L1"	"Это суперсемейство представляет консервативный компонент в белке ORF1 ретротранспозона, представляющего собой длинный диспергированный ядерный

			элемент 1 (LINE-1 или L1). Этот домен обозначен как С-концевой, и его единственная функция заключается в содействии связыванию РНК с центральным доменом RRM.
29	IPR013544	Домен с неизвестной функцией DUF1725	Этот домен находится на С-конце многих эукариотических последовательностей и некоторых бактериальных последовательностей. Многие из его представителей аннотированы как предполагаемые ретротранспозоны L1 или гомологи обратной транскриптазы LINE-1. Рассматриваемая область обнаруживается повторяющейся у некоторых представителей семейства.
30	PF02994	RBD-подобный домен подвижного элемента L1	Эта запись представляет RBD-подобный домен.
31	IPR035301	"Домен тримеризации подвижного элемента L1"	Эта запись представляет домен тримеризации, обнаруживаемый в белке ORF1 длинного диспергированного ядерного элемента 1 (LINE-1 или L1).
32	PF02099	Josephin	NA
33	IPR006155	Домен Josephin	"Домен Josephin представляет собой эукариотический белковый модуль из приблизительно 180 остатков, который встречается самостоятельно в Josephin-подобных белках и в виде аминоконцевого домена, ассоциированного с двумя или тремя копиями убиквитин-взаимодействующего мотива (UIM), в белках, подобных атаксину-3. Белки, содержащие домен Josephin, функционируют как ферменты деубиквитинирования. Хотя первоначально предполагалось, что домен Josephin может представлять собой домен, полностью состоящий из альфа-спиралей, отдаленно родственный доменам ENTH и VHS, участвующим в функции мембранного транспорта и регуляторного адаптера, теперь считается, что этот

			домен представляет собой домен цистеиновой протеазы, в основном состоящий из альфа-спиралей, который согласно прогнозам является активным в отношении цепей убиквитина или родственных субстратов. Домен Josephin содержит два консервативных остатка гистидина и один остаток цистеина, которые необходимы для активности убиквитин-протеазы, и два сайта связывания убиквитина. Некоторые белки, которые, как известно, содержат домен Josephin, представляют собой следующие. * Белок 1 болезни Мачадо-Джозефа животных (атаксин 3). Он взаимодействует с ключевыми регуляторами (CBP, p300 и PCAF) транскрипции и подавляет транскрипцию. * Белок, подобный белку болезни Мачадо-Джозефа растений (MJD1a-подобный) (гомолог атаксина 3). * Josephin млекопитающих 1 и 2. * Josephin-подобный белок Drosophila melanogaster. * Josephin-подобный белок Arabidopsis thaliana.
34	IPR022035	"Домен WW PCIF1"	"Этот домен обнаруживается у бактерий и эукариот и имеет длину примерно 180 аминокислот. Он представляет собой домен WW PCIF1. PCIF1 взаимодействует с фосфорилированным карбоксиконцевым доменом РНК-полимеразы II (CTD). Домен WW PCIF1 может непосредственно и предпочтительно связываться с фосфорилированным CTD по сравнению с нефосфорилированным CTD. PCIF1 связывается с гиперфосфорилированной RNAP II (RNAP IIO) in vitro и in vivo. Двойное иммунофлуоресцентное мечение в клетках HeLa показало, что PCIF1 и эндогенная RNAP IIO совместно локализованы в ядре клетки. PCIF1 представляет собой кэп-специфическую аденозинметилтрансферазу, которая катализирует образование

			N6,2'-O-диметиладенозинового кэпа (m6A(m)) путем метилирования аденозина во втором транскрибируемом положении кэпированных мРНК."
35	PF12237	Домен WW фактора 1, взаимодействующего с фосфорилированным CTD	"Это семейство доменов обнаруживается у бактерий и эукариот и имеет длину примерно 180 аминокислот. Этот домен представляет собой домен WW PCIF1. PCIF1 взаимодействует с фосфорилированным карбоксиконцевым доменом РНК-полимеразы II (CTD). Домен WW PCIF1 может непосредственно и предпочтительно связываться с фосфорилированным CTD по сравнению с нефосфорилированным CTD. PCIF1 связывается с гиперфосфорилированной RNAP II (RNAP IIO) in vitro и in vivo. Двойное иммунофлуоресцентное мечение в клетках HeLa показало, что PCIF1 и эндогенная RNAP IIO совместно локализованы в ядре клетки. Таким образом, PCIF1 может играть роль в синтезе мРНК, модулируя активность RNAP IIO."
36	cd14845	"Пептидаза L-Ala-D-Glu, также известная как L-аланил-D-глутаматэндопептидаза"	"Это семейство пептидаз L-Ala-D-Glu включает L-аланил-D-глутаматпептидазу (бактериофага T5) (также известную как L-аланоил-D-глутаматэндопептидаза) и пептидазы L-Ala-D-Glu Ply118 и Ply500. Эндолизин бактериофага разрушает пептидогликан бактериального хозяина изнутри, что приводит к лизису клеток и высвобождению вирионов-потомков. Эндолизин бактериофага Ply118 расщепляет связи между остатками L-Ala и D-Glu пептидогликана клеточной стенки Listeria. Это семейство принадлежит к подсемейству C пептидаз M15 MEROPS. [PMID: 15044722, PMID: 18560152, PMID: 16789876, PMID: 9702193, PMID: 9614968, PMID: 8852545, PMID: 8631706, PMID: 7810996, PMID: 7121588, PMID: 10500118, PMID: 9265630]"

37	IPR015898	Домен, подобный гамма-субъединице G-белка	"Эта запись представляет гамма-субъединицу G-белка и домен GGL (подобный гамма-субъединице G-белка), которые родственны в отношении своих последовательностей и состоят из удлиненного альфа-спирального полипептида. Гамма-субъединица G-белка образует стабильный димер с бета-субъединицей, но не вступает в какой-либо контакт с альфа-субъединицей, которая контактирует с противоположной стороной бета-субъединицы. Домен GGL обнаруживается в нескольких белках RGS (регуляторах передачи сигналов посредством G-белка). Домены GGL могут взаимодействовать с бета-субъединицами с образованием новых димеров, которые предотвращают связывание гамма-субъединицы и могут предотвращать образование гетеротримера путем ингибирования связывания альфа-субъединицы. Взаимодействие между нейронспецифическими изоформами бета-5 G-белка и доменами GGL RGS может представлять собой общий способ связывания между бета-пропеллерными белками и их партнерами."
38	PS50058	Профиль домена гамма-субъединицы G-белка	"Связывающие гуаниновые нуклеотиды белки (G-белки) [1] выступают в качестве посредников при передаче сигналов, генерируемых трансмембранными рецепторами. G-белки состоят из трех субъединиц (альфа, бета и гамма). Альфа-субъединица связывается с GTP и гидролизует его; функции бета- и гамма-субъединиц менее ясны, но они, по-видимому, необходимы для замены GDP на GTP, а также для заякоривания на мембране и распознавания рецепторов. Гамма-субъединицы представляют собой небольшие белки (длиной от 70 до 110 остатков), которые связываются с мембраной посредством изопренильной группы (либо

			фарнезила, либо геранилгеранила), ковалентно связанной с их С-концом. У млекопитающих существует по меньшей мере 12 различных изоформ гамма-субъединиц. Белок egl-10 Caenorhabditis elegans, который является регулятором передачи сигналов посредством G-белка, содержит домен, подобный гамма-субъединице G-белка. Авторы настоящего изобретения разработали профиль, который охватывает всю длину гамма-субъединицы."
39	cd16857	"Домен ингибитора роста (ING) из белка ингибитора роста ING1, ING2 и сходных белков"	"ING1 представляет собой эпигенетический регулятор и супрессор опухолей типа II, который влияет на рост, старение, апоптоз клеток и репарацию ДНК, воздействуя на конформацию хроматина и экспрессию генов. Он выступает в качестве считывателя метки активного хроматина - триметилирования лизина 4 гистона H3 (H3K4me3). Он связывает и направляет белок 45a (Gadd45a), индуцируемый остановкой роста и повреждением ДНК, к сайтам-мишеням, таким образом связывая гистоновый код с деметилированием ДНК. Он взаимодействует с ядерным антигеном пролиферирующих клеток (PCNA) через домен PCNA-взаимодействующего белка (PIP) УФ-индуцируемым образом. Он также взаимодействует с PCNA-взаимодействующим белком p15 (PAF). Кроме того, ING1 ассоциирует с представителями семейства 14-3-3, что необходимо для цитоплазматической релокализации. Эндогенный белок ING1 специфично взаимодействует с проапоптотическим представителем семейства BCL2 BAX и совместно локализуется с BAX УФ-индуцируемым образом. Он стабилизирует супрессор опухолей p53, ингибируя полиубиквитинирование мультимоноубиквитинированных форм

			посредством взаимодействия и совместной локализации герпесвирус-ассоциированной убиквитинспецифической протеазы (HAUSP)-деубиквитиназы с p53. Он также участвует в индуцированном трихостатином A апоптозе и передаче сигналов посредством каспазы 3 в дефицитных по p53 клетках глиобластомы. Кроме того, тирозинкиназа Src может связывать и фосфорилировать ING1 и дополнительно регулировать его активность. ING2, также называемый белком, подобным ингибитору роста 1 (ING1Lp), или p32, или p33ING2, является основным компонентом многофакторного модифицирующего хроматин комплекса, содержащего транскрипционный корепрессор SIN3A и гистондеацетилазу 1 (HDAC1). Он участвует в контроле клеточного цикла, стабильности генома и дифференцировке клеток мышц. ING2 независимо взаимодействует с H3K4me3 (гистон H3, триметилированный по лизину 4) и PtdIns(5)P и модулирует взаимное влияние метилирования лизина и ацетилирования лизина на гистоновых белках посредством ассоциации с хроматином в присутствии повреждения ДНК. Он действует совместно с SnoN, опосредуя индуцированную трансформирующим фактором роста (TGF)-бета Smad-зависимую транскрипцию и клеточные ответы. Его экспрессия повышается при раке толстой кишки, и он увеличивает инвазию за счет повышенной экспрессии MMP13. Он также выступает в качестве кофактора p300 для ацетилирования p53 и играет положительную регуляторную роль во время p53-опосредованного репликативного старения. Как ING1, так и ING2 содержат N-концевой домен ING, содержащий мотив, подобный лейциновой "молнии" (LZL), и хорошо изученный C-

			концевой домен "цинковый палец" типа гомеодомена растений (PHD). [PMID: 18450387, PMID: 18836436, PMID: 18655775, PMID: 19085961, PMID: 19442115, PMID: 20066899, PMID: 20179197, PMID: 20681406, PMID: 21078114, PMID: 21227930, PMID: 21286670, PMID: 21707358, PMID: 21731648, PMID: 21896275, PMID: 22163049, PMID: 23388825, PMID: 23459830, PMID: 23585863, PMID: 23803967, PMID: 19766113, PMID: 19730436, PMID: 19442116, PMID: 19437536, PMID: 19184981, PMID: 18334480, PMID: 18160212, PMID: 17208481, PMID: 16024799, PMID: 18533182, PMID: 8944021, PMID: 10626813, PMID: 10866301, PMID: 10807544, PMID: 9440695, PMID: 24008732, PMID: 19962781, PMID: 20676127, PMID: 20621832, PMID: 20890119, PMID: 21124965, PMID: 20948283, PMID: 21767350, PMID: 22808232, PMID: 23823870, PMID: 23864195, PMID: 16728974, PMID: 16387653, PMID: 16893883, PMID: 12859901, PMID: 11784859, PMID: 11481424, PMID: 10072587, PMID: 16728977, PMID: 23412501]"
40	PF09004	Домен с неизвестной функцией (DUF1891)	Этот домен обнаруживается в наборе гипотетических эукариотических белков.
41	IPR015095	"N-концевой гомолог 8 белка репарации алкилированной ДНК AlkB"	Этот домен находится в крайнем N-концевом положении гомологов эукариотического белка репарации алкилированной ДНК.
42	PF00098	"Цинковый сустав"	"Цинковый сустав представляет собой цинк-связывающий мотив, состоящий из следующего: CX2CX4HX4C, где X может представлять собой любую аминокислоту. Мотивы в основном происходят из ретровирусных белков gag (нуклеокапсидных). Структура прототипа получена из HIV. Он также включает представителей,

			участвующих в регуляции эукариотических генов, таких как GLH-1 C. elegans. Структура представляет собой цинковый палец с 18 остатками".
43	IPR026960	Цинк-связывающий домен обратной транскриптазы	Этот домен, по-видимому, представляет собой цинк-связывающую область предполагаемой обратной транскриптазы.
44	PF13966	Цинк-связывающий домен в обратной транскриптазе	Этот домен, по-видимому, представляет собой цинк-связывающую область предполагаемой обратной транскриптазы.
45	cd05920	"2,3-дигидроксибензоат-AMP-лигаза"	"2,3-дигидроксибензоат-AMP-лигаза активирует 2,3-дигидроксибензоат (DHB) путем лигирования AMP из ATP с высвобождением пирофосфата. Однако она также может катализировать обмен ATP-PPi для аналогов 2,3-DHB, таких как салициловая кислота (о-гидробензоат), а также 2,4-DHB и 2,5-DHB, но с меньшей эффективностью. Белки этого семейства представляют собой самостоятельные аденилирующие компоненты нерибосомных пептидсинтаз (NRPS), участвующих в биосинтезе сидерофоров, которые представляют собой низкомолекулярные соединения, хелатирующие железо, синтезируемые многими бактериями для облегчения получения этого жизненно важного микроэлемента. В Escherichia coli 2,3-дигидроксибензоат-AMP-лигаза называется EntE и представляет собой аденилирующий компонент системы энтеробактина NRPS. [PMID: 12221282, PMID: 21166461, PMID: 16678412, PMID: 2531000, PMID: 11018148, PMID: 16632253, PMID: 9373621, PMID: 17762044, PMID: 11255012]"
46	IPR042509	Белок 3, содержащий домен "цинковый палец" CCHC	"Белок 3, содержащий домен "цинковый палец" CCHC (ZCCHC3) представляет собой белок "цинковый палец", который выступает в качестве косенсора cGAS во

			врожденном иммунном ответе на цитозольную dsDNA и ДНК-вирус. Было показано, что он связывает ДНК и РНК в цитоплазме и способствует распознаванию вирусных нуклеиновых кислот сенсорами вирусов, такими как DDX58/RIG-I, IFIH1/MDA5 и CGAS."
47	cd05844	Белки семейства 4 гликозилтрансфераз	"Гликозилтрансферазы катализируют перенос сахарных фрагментов от активированных донорных молекул к специфическим акцепторным молекулам, образуя гликозидные связи. Акцепторная молекула может представлять собой липид, белок, гетероциклическое соединение или другой углеводный остаток. Эта группа гликозилтрансфераз наиболее близкородственна семейству 4 гликозилтрансфераз (GT4). Представители этого семейства могут переносить сахарные фрагменты, связанные с UDP, ADP, GDP или CMP. Разнообразная ферментативная активность представителей этого семейства отражает широкий спектр биологических функций. Белковая структура, доступная для этого семейства, имеет топологию GTB - одну из двух белковых топологий, наблюдаемых для зависимых от нуклеотидного сахара гликозилтрансфераз. Белки GTB имеют разные N- и C-концевые домены, каждый из которых содержит типичную укладку Россманна. Два домена характеризуются высокой структурной гомологией, несмотря на минимальную гомологию последовательностей. Большая расщелина, разделяющая два домена, содержит каталитический центр и обеспечивает высокую степень гибкости. [PMID: 12624205, PMID: 7596293, PMID: 16564135, PMID: 16037492, PMID: 16263268, PMID: 15653326, PMID: 12691742]"

48	cd00229	"Гидролаза_SGNH или гидролаза_GDSL представляет собой разнообразное семейство липаз и эстераз"	"Гидролаза_SGNH или гидролаза_GDSL представляет собой разнообразное семейство липаз и эстераз. Третичная укладка фермента существенно отличается от укладки семейства альфа/бета-гидролаз и уникальна среди всех известных гидролаз; его активный центр очень похож на типичную триаду Ser-His-Asp(Glu) других сериновых гидролаз, но может не содержать карбоновой кислоты. [PMID: 7610479, PMID: 15522763, PMID: 12842470, PMID: 10801485]"
49	PS00354	Предполагаемая нуклеаза HARBI1	NA
50	IPR000637	"Консервативный сайт ДНК-связывающих HMG-I/HMG-Y"	"Группа белков с высокой подвижностью (HMG) представляет собой семейство относительно низкомолекулярных негистоновых компонентов хроматина. HMG-I и HMG-Y (HMGA) представляют собой белки, состоящие из приблизительно 100 аминокислотных остатков, которые образуются в результате альтернативного сплайсинга одного гена. Белки HMG-I/Y предпочтительно связываются с малой бороздкой AT-богатых областей двухнитевой ДНК неспецифичным в отношении последовательности образом. Предполагается, что эти белки могут функционировать в фазировании нуклеосом и в процессинге 3'-конца транскриптов мРНК. Они также участвуют в регуляции транскрипции генов, содержащих AT-богатые области или находящихся в непосредственной близости них. Связывание ДНК этими и несколькими родственными белками осуществляется доменом из 11 остатков, известным как AT-крючок. В известных белках HMG-I/Y обнаруживаются три высококонсервативные области, близкородственные консенсусной

			последовательности TPKRPRGRPKK. Синтетический олигопептид с этой последовательностью специфично связывается с ДНК-субстратом способом, напоминающим таковой у интактных белков HMG-I/Y. Предсказания структуры позволяют предположить, что пептид имеет вторичную структуру, сходную с противоопухолевыми и противовирусными лекарственными средствами нетропсином и дистамицином, а также красителю Hoechst 33258. Эти лиганды, которые также предпочтительно связываются с AT-богатой ДНК, эффективно конкурируют как с синтетическим пептидом, так и с белками HMG-I/Y за связывание с ДНК. Пептид также содержит новые структурные особенности, такие как предсказанный изгиб Asx или "крючок" на его N-конце и выступающие латерально катионные "щетинки" Arg/Lys, которые могут играть роль в связывании белков HMG-I/Y. Предсказанная структура пептида - AT-крючок - представляет собой ранее не описанный ДНК-связывающий мотив."
51	IPR001952	Щелочная фосфатаза	"Эта запись представляет щелочные фосфатазы () (ALP), которые выступают в качестве неспецифических фосфомоноэстераз для гидролиза сложных фосфатных эфиров, в оптимальном случае при высоких значениях pH. Механизм реакции включает атаку алкоксида серина на атом фосфора субстрата с образованием переходного ковалентного комплекса фермент-фосфат с последующим гидролизом фосфата серина. Щелочные фосфатазы встречаются во всех царствах живых организмов, за исключением некоторых растений. Щелочные фосфатазы представляют собой металлоферменты, существующие в виде димера, каждый

			мономер которого связывает ионы металлов. Ионы металлов, которые они несут, могут различаться, хотя наиболее распространены цинк и магний. Например, щелочной фосфатазе Escherichia coli (кодируемой phoA) требуется присутствие двух ионов цинка, связанных в сайтах для металлов M1 и M2, и одного иона магния, связанного в сайте M3. Однако щелочным фосфатазам Thermotoga maritima и Bacillus subtilis для максимальной активности требуется кобальт. У млекопитающих существует четыре изофермента щелочной фосфатазы: плацентарная, подобна плацентарной (зародышевых клеток), кишечная и неспецифичная в отношении ткани (печени/кости/почки). Все четыре изофермента заякорены на наружной поверхности плазматической мембраны ковалентно присоединенным гликозилфосфатидилинозитоловым (GPI) якорем. Щелочные фосфатазы человека имеют четыре сайта связывания металлов: два для цинка, один для магния и один для ионов кальция. Плацентарная щелочная фосфатаза (ALPP или PLAP) является высокополиморфной и имеет по меньшей мере три распространенных аллеля. Ее активность подавляется рядом эффекторов, таких как 1-фенилаланин, 5'-AMP и п-нитрофенилфосфонат (PNPPate). Уровень изофермента, подобного плацентарному (ALPPL или PLAP-подобного), повышен в опухолях из зародышевых клеток. Кишечный изофермент (ALPI или IAP) обладает способностью к детоксификации липополисахарида и предотвращению инвазии бактерий через барьер слизистой оболочки кишечника. Неспецифичный в отношении ткани изофермент (ALPL) может играть роль в минерализации скелета. Дефекты в ALPL являются причиной

			гипофосфатазии, в том числе инфантильного типа (OMIM:241500), детского типа (OMIM:241510) и взрослого типа (OMIM:146300). Гипофосфатазия представляет собой наследственное метаболическое заболевание костей, характеризующееся нарушением минерализации скелета. Эта запись также содержит родственный фермент стрептомицин-6-фосфатфосфатазу () (кодируемый strK) из видов Streptomyces. Этот фермент участвует в синтезе антибиотика стрептомицина, специфически расщепляя как стрептомицин-6-фосфат, так и, более медленно, стрептомицин-3-фосфат."
52	IPR040260	Белок, подобный фактору репликации A	"Rfa2 (также известный как RPA32) является компонентом комплекса белка репликации A (RPA), который связывается с ssDNA и удаляет ее вторичную структуру. Комплекс RPA участвует в репликации, репарации и рекомбинации ДНК. Stn1 является компонентом комплекса CST - комплекса, который связывается с однонитевой ДНК и необходим для защиты теломер от разрушения ДНК. Комплекс CST связывает однонитевую ДНК с высокой аффинностью независимым от последовательности образом, тогда как выделенные субъединицы сами по себе связывают ДНК с низкой аффинностью. В дополнение к защите теломер, комплекс CST, вероятно, играет более общую роль в метаболизме ДНК в нетеломерных сайтах. Эта запись содержит белки с ДНК-связывающим доменом "крылатая спираль", включая белок 2 фактора репликации A (RFA2) и субъединицу STN1 комплекса CST."
53	PS01359	Профиль домена CUB	NA

54	IPR01978 6	"Консервативный сайт "цинковый палец" типа PHD"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии,

			сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет домен "цинковый палец" типа PHD (гомеодомена), который представляет собой мотив, подобный "цинковому пальцу" C4HC3, обнаруживаемый в ядерных белках, которые, как считается, участвуют в опосредованной хроматином регуляции транскрипции. Мотив "палец" типа PHD напоминает "палец" типа RING, содержащий мотив C3HC4, но отличается от него. Функция этого домена еще не известна, но по аналогии с доменом LIM он может быть вовлечен в белок-белковое взаимодействие и может быть важным для сборки или активности многокомпонентных комплексов, участвующих в активации или репрессии транскрипции. В качестве альтернативы, взаимодействия могут быть внутримолекулярными и играть важную роль в поддержании структурной целостности белка. По аналогии с "пальцем" типа RING и доменом LIM считается, что "палец" типа PHD связывает два иона цинка. Сигнатура этой записи начинается с первого остатка цистеина области "цинкового пальца" и заканчивается последним. Расстояние между остатками цистеина в "пальце" типа PHD близко к таковому в "пальце" типа RING. Таким образом, проведение различий между этими двумя доменами либо по паттерну, либо по профилю затруднено, и некоторые редкие домены распознаются паттернами и профилями как RING, так и PHD".

55	PF03258	Белок FP бакуловируса	Белок FP отсутствует у мутантных форм бакуловируса (с малым количеством полиэдров) [PMID:8760443].
56	IPR019787	"Цинковый палец" типа PHD"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например,

			Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет домен "цинковый палец" типа PHD (гомеодомена), который представляет собой мотив, подобный "цинковому пальцу" C4HC3, обнаруживаемый в ядерных белках, которые, как считается, участвуют в опосредованной хроматином регуляции транскрипции. Мотив "палец" типа PHD напоминает "палец" типа RING, содержащий мотив C3HC4, но отличается от него. Функция этого домена еще не известна, но по аналогии с доменом LIM он может быть вовлечен в белок-белковое взаимодействие и может быть важным для сборки или активности многокомпонентных комплексов, участвующих в активации или репрессии транскрипции. В качестве альтернативы, взаимодействия могут быть внутримолекулярными и играть важную роль в поддержании структурной целостности белка. По аналогии с пальцем типа RING и доменом LIM считается, что палец типа PHD связывает два иона цинка."
57	IPR011011	"Цинковый палец" типа FYVE/PHD"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают

			другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Домен "цинковый палец" типа FYVE является

			консервативным от дрожжей до человека и назван в честь четырех белков, в которых он был обнаружен: Fab1, YOTB/ZK632.12, Vac1 и EEA1. Он участвует в привлечении цитозольных белков к мембране путем связывания с фосфатидилинозитол-3-фосфатом (PI3P), который обнаруживается в основном на эндосомах. Домен "цинковый палец" типа гомеодомена растений (PHD) содержит мотив типа C4HC3 и широко распространен у эукариот, при этом он обнаруживается во многих регуляторных факторах хроматина. Мотивы "цинковые пальцы" как типа FYVE, так и типа PHD демонстрируют на удивление похожие укладки альфа+бета, связанные с диметаллом (цинком)."
58	PS50016	Профиль "цинкового пальца" типа PHD	"Цинковый палец типа гомеодомена растений (PHD) (C4HC3) [PMID: 7701562] обнаруживается в ядерных белках, которые, как считается, участвуют в опосредованной хроматином регуляции транскрипции. Мотив "палец" типа PHD напоминает "палец" типа RING, содержащий мотив C3HC4, но отличается от него. Общая функция "пальца" типа PHD еще не известна, но в CREB-связывающем белке (CBP) он является неотъемлемой частью ферментативной сердцевины ацетилтрансферазного домена, а также вносит ключевой вклад в транскрипционную активность CBP. 3D-структура транскрипционного фактора синдрома Вильямса-Бойрена показывает, что "палец" PHD координирует два атома цинка крестообразным образом [PMID:11124022]. Эта атипичная конформация также характерна для FYVE. Два сайта для цинка расположены на противоположных концах короткого бета-листа. Этот признак также является общим для "пальца" типа

			RING и домена типа FYVE, но структура отличается окружающими петлями, которые значительно различаются по длине и последовательности, что может объяснить большое разнообразие функций между этими тремя доменами. Один или несколько "пальцев" типа PHD обнаруживаются в следующих белках. - Белок Drosophila trithorax и гомологичный белок млекопитающих HRX (также известный как ALL1, MLL и Htrx). Ген trithorax необходим для правильной регуляции гомеозисных генов у мух [PMID:8098166]. Ген HRX участвует в транслокациях при остром лимфобластном лейкозе. Trithorax содержит 4 "пальца" типа PHD. Наиболее близкий к С-концу домен в HRX является вырожденным. - Polycomb-подобный белок из Drosophila - белок, кодируемый геном группы Polycomb. Ген Polycomb-подобного белка и гены группы Polycomb необходимы для правильной регуляции гомеозисных генов у мух. Эти гены противодействуют эффектам генов группы trithorax, по-видимому, путем создания неактивной структуры хроматина [PMID:7907493]. Polycomb-подобный белок содержит 2 "пальца" типа PHD. - Xe169 человека и Smcx мыши - белки, кодируемые генами, сцепленными с X-хромосомой, которые избегают инактивации X-хромосомы. Smcy мыши представляет собой родственный ген на Y-хромосоме. Xe169 содержит 2 "пальца" типа PHD. - RBP2 человека - белок, родственный Xe169, который, как полагают, связывается с белком Rb. - Br140 человека (также известный как перегрин) - белок, который также содержит бромодомен. - Альфин-1 Medicago sativa - предполагаемый фактор транскрипции. - Группа гомеобокс-содержащих белков растений, таких как HAT3 и PRHA растения рода Arabidopsis, ZMHOX1a

			кукурузы, PRHP петрушки. - Аутоантиген человека Mi-2. - Белок Shuttle Craft Drosophila. - BS69 человека - белок, связывающий E1A аденовируса. - CBP мыши (CREB-связывающий белок), близкородственный E1A-связывающий белок p300 человека и YNJ1, родственный белок из C. elegans - белок с "цинковыми пальцами" млекопитающих семейства "requiem". - TIF1 человека - предполагаемый медиатор транскрипции ядерных рецепторов. Консенсусный паттерн, который авторы настоящего изобретения разработали, начинается с первого остатка цистеина и заканчивается последним, при этом разработанный авторами настоящего изобретения профиль охватывает весь домен."
59	IPR013083	"Цинковый палец" типа RING/FYVE/PHD"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев".

			Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Это суперсемейство представляет домены "цинковые пальцы" типа RING, PHD и FYVE, которые имеют общую структурную укладку альфа/бета, связанную с диметаллом (цинком), а также не содержащий цинк домен U-бокса, который сходен с "цинковым пальцем" типа RING за исключением того, что в нем отсутствуют остатки, связывающие ионы металла (U-бокс, ассоциированный с множественным убиквитинированием)."
60	cd15489	Суперсемейство "пальцев" типа PHD	"Суперсемейство "пальцев" типа PHD включает канонический "палец" типа гомеодомена растений (PHD), обычно характеризуемый как Cys4HisCys3, и неканонический удлиненный "палец" типа PHD, характеризуемый как

			Cys2HisCys5HisCys2His. Вариации включают "палец" типа PHD RAG2, характеризуемый как Cys3His2Cys2His, и "палец" типа PHD 5, обнаруживаемый в белках, содержащих домен SET, связывающих ядерный рецептор, и характеризуемый как Cys4HisCys2His. "Палец" типа PHD также называют мотивом LAP (белок, ассоциированный с лейкемией) или доменом TTC (консенсусная последовательность trithorax). Единичные или множественные копии "пальцев" типа PHD были обнаружены в различных эукариотических белках, участвующих в контроле транскрипции генов и динамики хроматина. "Пальцы" типа PHD могут распознавать немодифицированный и модифицированный хвост гистона H3, и было обнаружено, что некоторые из них взаимодействуют с негистоновыми белками. Они также функционируют как эпигеномные считыватели, контролирующие экспрессию генов посредством молекулярного привлечения мультибелковых комплексов регуляторов хроматина и факторов транскрипции. Домен SF "палец" типа PHD структурно сходен с суперсемействами RING и FYVE_подобные. [PMID: 12480335, PMID: 16297627, PMID: 17949986, PMID: 18682256, PMID: 19011376, PMID: 19442115, PMID: 19925420, PMID: 20048137, PMID: 21813457, PMID: 24291127, PMID: 11884585, PMID: 7568208, PMID: 7701562, PMID: 7583761, PMID: 21514168, PMID: 11124022, PMID: 17142463]"
61	IPR004941	Белок FP бакуловируса	Функция белка FP неизвестна. Белок отсутствует у мутантных форм бакуловируса FP (с малым количеством полиэдров).

62	IPR036514	Суперсемейство гидролаз SGNH	"Гидролаза SGNH имеет укладку, аналогичную таковой у флавопротеинов, а именно трехслойную структуру альфа/бета/альфа, где бета-листы состоят из пяти параллельных нитей. Ферменты, содержащие этот домен, выступают в качестве эстераз и липаз, но характеризуются низкой гомологией последовательностей с истинными липазами. Белки, содержащие этот тип эстеразного домена, были обнаружены во множестве гидролаз; те, в отношении которых есть структурная информация, включают эстеразу из Streptomyces scabies; эстеразный домен поверхностных гликопротеинов вирусной гемагглютининэстеразы (вирус гриппа C, коронавирусы и торовирусы); ацетилгидролазы млекопитающих; грибную рамногалактуронанацетилэстеразу и многофункциональный фермент тиоэстеразу I (TAP) из Escherichia coli. Эстеразные домены типа гидролазы SGNH содержат уникальную сеть водородных связей, которая стабилизирует их каталитические центры; они обычно содержат каталитическую триаду Ser/кислота/His."
63	PF13472	Семейство GDSL-подобных липаз/ацилгидролаз	Это семейство предполагаемых липаз и родственных ферментов сходно с Pfam:PF00657.
64	IPR013830	Эстеразный домен типа гидролазы SGNH	"Эта запись представляет эстеразный домен типа гидролазы SGNH, который имеет укладку, аналогичную таковой у флавопротеинов, а именно трехслойную структуру альфа/бета/альфа, где бета-листы состоят из пяти параллельных нитей. Ферменты, содержащие этот домен, выступают в качестве эстераз и липаз, но характеризуются низкой гомологией последовательностей с истинными липазами. Белки, содержащие этот тип эстеразного домена, были обнаружены во множестве

			гидролаз; те, в отношении которых есть структурная информация, включают эстеразу из Streptomyces scabies; эстеразный домен поверхностных гликопротеинов вирусной гемагглютининэстеразы (вирус гриппа C, коронавирусы и торовирусы); ацетилгидролазы млекопитающих; грибную рамногалактурононацетилэстеразу и многофункциональный фермент тиоэстеразу I (TAP) из Escherichia coli. Эстеразные домены типа гидролазы SGNH содержат уникальную сеть водородных связей, которая стабилизирует их каталитические центры; они обычно содержат каталитическую триаду Ser/кислота/His."
65	IPR027124	Белок 5 комплекса SWR1/белок 2 развития черепно-лицевой области	"Это семейство включает белок 5 комплекса SWR1 (Swc5) и белок 2 развития черепно-лицевой области (CFDP2). Swc5 является компонентом комплекса SWR1, который опосредует ATP-зависимый обмен гистона H2A на вариант H2A - HZT1, что приводит к регуляции транскрипции отобранных генов путем ремоделирования хроматина. Он участвует в обеспечении стабильности хромосом. Дупликация предкового гена bnct (bucentaur) приводит к гену bnct (cfdp1) типа h и гену p97bcnt (cfdp2). Cfdp2 специфичен для жвачных животных и содержит область, полученную из эндонуклеазного домена ретротранспозонного элемента RTE-1, не обнаруживаемого у человека и мыши. По этой причине bcnt использовали в качестве модели молекулярной эволюции."
66	IPR018527	"Рубредоксин, сайт связывания железа"	"Рубредоксин представляет собой низкомолекулярный железосодержащий белок, участвующий в переносе электронов, иногда заменяющий ферредоксин в качестве переносчика электронов. Он обнаруживается у бактерий, архей и растений. Были

			расшифрованы 3D-структуры ряда рубредоксинов. Укладка относится к классу альфа+бета, с 2 альфа-спиралями и 2-3 бета-нитями. Его активный центр содержит ион железа, который координируется ионами серы четырех консервативных остатков цистеина, образующих почти правильный тетраэдр. Консервативные остатки цистеина находятся в двух петлях, которые являются наиболее консервативными областями белка. Кроме того, кольцо кислотных остатков вблизи центра [Fe(Cys)4] также является высококонсервативным. Эта запись представляет консервативную область, которая содержит два цистеиновых остатка, связывающих атом железа."
67	PS00202	Предполагаемая нуклеаза HARBI1	NA
68	IPR001965	"Цинковый палец" типа PHD"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от

			аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет домен "цинковый палец" типа PHD (гомеодомена), который представляет собой мотив, подобный "цинковому пальцу" C4HC3, обнаруживаемый в ядерных белках, которые, как считается, участвуют в опосредованной хроматином регуляции транскрипции. Мотив "палец" типа PHD напоминает "палец" типа RING, содержащий мотив C3HC4, но отличается от него. Функция этого домена еще не известна, но по аналогии с доменом LIM он может быть вовлечен в белок-белковое взаимодействие и может

			быть важным для сборки или активности многокомпонентных комплексов, участвующих в активации или репрессии транскрипции. В качестве альтернативы, взаимодействия могут быть внутримолекулярными и играть важную роль в поддержании структурной целостности белка. По аналогии с пальцем типа RING и доменом LIM считается, что палец типа PHD связывает два иона цинка."
69	cd15517	"Палец" типа PHD, обнаруживаемый в факторе транскрипции 19 (TCF-19), лизин-специфической деметилазе KDM5A и KDM5B и других сходных белках"	"TCF-19 был идентифицирован как предполагаемый транс-активирующий фактор, экспрессия которого начинается на позднем этапе перехода G1-S в делящихся клетках. Он функционирует как новый островковый фактор, необходимый для пролиферации и выживания линии бета-клеток INS-1. Он играет важную роль в восприимчивости как к сахарному диабету 1 типа (T1DM), так и к сахарному диабету 2 типа (T2DM); было высказано предположение, что он может положительно влиять на массу бета-клеток в условиях стресса бета-клеток и повышенной потребности в инсулине. KDM5A первоначально был идентифицирован как партнер по связыванию белка ретинобластомы (Rb), и его инактивация может быть важна для того, чтобы Rb способствовал дифференцировке. Он участвует в транскрипции посредством взаимодействия с TBP, p107, ядерными рецепторами, Myc, Sin3/HDAC, Mad1, RBP-J, CLOCK и BMAL1. KDM5B имеет ограниченный паттерн экспрессии в яичках, яичниках и транзиентно в молочной железе беременных женщин, и было показано, что уровень его экспрессии повышается при раке молочной железы, раке предстательной железы и раке легких, что позволяет предположить потенциальную роль в онкогенезе. Как KDM5A, так и KDM5B функционируют как деметилазы гистона

			H3, триметилированного по лизину 4 (H3K4me3). Это семейство также включает гомолог лизин-специфической деметилазы 7 Caenorhabditis elegans (ceKDM7A). ceKDM7A (также называемый белком 1.2, содержащим домен JmjC, гомологом белка 8 "пальцев" типа PHD или гомологом PHF8) представляет собой белок, содержащий гомеодомен растений (PHD) и домен JmjC, который функционирует как деметилаза гистонов, специфичная в отношении H3K9me2 и H3K27me2. Связывание "пальца" типа PHD с H3K4me3 направляет H3K9me2- и H3K27me2-специфическое деметилирование с помощью его каталитического домена JmjC в механизме транс-регуляции гистонов. Кроме того, это семейство включает белки растений OBERON 1 и OBERON 2, белки, подобные альфину 1 (AL), гистонацетилтрансферазы (HAT) HAC и белок 4, содержащий AT-богатый взаимодействующий домен (ARID4). [PMID: 1868030, PMID: 10777095, PMID: 23860123, PMID: 10336460, PMID: 10616211, PMID: 10878660, PMID: 12237901, PMID: 12657635, PMID: 15803180, PMID: 16645588, PMID: 17320161, PMID: 17311883, PMID: 17373667, PMID: 17363312, PMID: 23907094, PMID: 23777767, PMID: 23637629, PMID: 23408432, PMID: 22020125, PMID: 22420752, PMID: 22615382, PMID: 21980403, PMID: 21937684, PMID: 21821892, PMID: 21369698, PMID: 21105039, PMID: 20863814, PMID: 20403335, PMID: 19336002, PMID: 18591252, PMID: 18048344, PMID: 17973255, PMID: 17709396, PMID: 20406991, PMID: 21348942, PMID: 21562575, PMID: 21960634, PMID: 20231316, PMID: 22937203, PMID: 23093672, PMID: 23112189, PMID: 17573780, PMID: 18483221, PMID: 18270511, PMID: 18722178, PMID: 19762557, PMID: 19850045, PMID: 21788502, PMID: 23722541, PMID: 17320160, PMID:

			17320163, PMID: 15949438, PMID: 11358960, PMID: 9129143, PMID: 7935440, PMID: 1857421, PMID: 8414517, PMID: 17351631, PMID: 17351630, PMID: 20567261, PMID: 20567262, PMID: 20346720, PMID: 18403411, PMID: 19392692, PMID: 14963126, PMID: 19154204, PMID: 12466527, PMID: 11160878]"
70	PF01541	Каталитический домен GIY-YIG	"Этот домен, называемый GIY-YIG, обнаруживается в аминоконцевой области субъединицы "c" эксцинуклеазы abc (uvrC), эндонуклеазах segA, segB, segC, segD и segE бактериофага T4; он также обнаруживается в предполагаемых эндонуклеазах, кодируемых интронами группы I грибов и фагов. Структура I-TevI, эндонуклеазы GIY-YIG, показывает новую укладку альфа/бета с центральным трехнитевым антипараллельным бета-листом, фланкированным тремя спиралями [PMID:12379841]. Наиболее консервативные и предполагаемые каталитические остатки расположены на неглубокой вогнутой поверхности и включают сайт координации металлов."
71	IPR00303 05	Эндонуклеаза GIY-YIG	"Нуклеазы семейства GIY-YIG участвуют во многих клеточных процессах, включая репарацию и рекомбинацию ДНК, перенос мобильных генетических элементов и ограничение поступления чужеродной ДНК. Суперсемейство GIY-YIG объединяет нуклеазы, характеризующиеся наличием домена, состоящего обычно из ~100 аминокислот, с двумя короткими мотивами "GIY" и "YIG" в N-концевой части, за которыми следует остаток Arg в центре и остаток Glu в C-концевой части. Домен GIY-YIG образует компактный структурный домен, который служит каркасом для координации иона двухвалентного металла, необходимого для катализа разрыва

			фосфодиэфирной связи. Домен GIY-YIG характеризуется архитектурой типа альфа/бета-"сэндвича" с центральным трехнитевым антипараллельным бета-листом, фланкированным тремя спиралями. Трехнитевой антипараллельный бета-лист содержит элементы последовательности GIY-YIG. Наиболее консервативные и предполагаемые каталитические остатки расположены на неглубокой вогнутой поверхности и включают сайт координации металлов. Домен GIY-YIG участвует во множестве клеточных процессов, включающих расщепление ДНК, от саморазмножения с интронами или без них до ограничения поступления чужеродной ДНК, для восстановления ДНК и поддержания стабильности генома. Некоторые белки, которые, как известно, содержат домен GIY-YIG, включают следующие. * Эукариотические белки Slx-1, участвующие в поддержании числа копий рДНК. У них имеется С-концевой Zn-связывающий домен "палец" типа RING. * Анкириновый повтор млекопитающих и белок 1, содержащий домен LEM (ANKLE1). * Субъединицы UvrC эксцинуклеаз (A)BC бактерий и архей, которые удаляют поврежденные нуклеотиды, делая разрез поврежденной нити с обеих сторон от места поражения. Вирус Paramecium bursaria Chlorella virus 1 (pbvc1). * Эндонуклеазы SegA-SegE фага T4, вероятно, участвующие в перемещении ДНК, кодирующей эндонуклеазу. * Интрон-ассоциированная эндонуклеаза 1 фага T4 (I-TevI), специфичная в отношении границы сплайсинга гена тимидилатсинтазы (td) и участвующая в хоминге интронов."

72	cd10442	Каталитический эндонуклеазный домен GIY-YIG Penelope-подобных элементов и сходных белков	"Эта модель соответствует домену EN PLE, который содержит каталитический модуль эндонуклеаз GIY-YIG бактериальных интронов/интронов органелл группы I, а также бактериальных белков репарации ДНК UvrC. Он может расщеплять ДНК с низкой специфичностью в отношении нуклеотидной последовательности. Однако домен EN PLE отличается от других эндонуклеаз GIY-YIG наличием высококонсервативного мотива CCHH (CX(2-7)CX(33-39)HX(3-5)H, при этом X может представлять собой любой остаток. Роль мотива CCHH еще не была идентифицирована. Penelope-подобные элементы (PLE) представляют новый класс эукариотических ретроэлементов, которые не принадлежат ни к ретротранспозонам, содержащим длинные концевые повторы (LTR), ни к ретротранспозонам, не содержащим LTR (часто называемым LINE), а вместо этого образуют сестринскую по отношению к теломеразным обратным транскриптазам (TERT) кладу узкоспециализированных немобильных обратных транскриптаз (RT), ответственных за добавление теломерных повторов к концам эукариотических хромосом. Одна открытая рамка считывания (ORF), кодируемая PLE, состоит из двух основных доменов - домена RT и эндонуклеазного домена (EN), соединенных линкерной областью переменной длины. Оба эти домена функционально активны. [PMID: 17345670, PMID: 16093704, PMID: 15465912, PMID: 18261821, PMID: 12743378, PMID: 16646971]"
73	IPR035901	Суперсемейство эндонуклеаз GIY-YIG	"Нуклеазы семейства GIY-YIG участвуют во многих клеточных процессах, включая репарацию и рекомбинацию ДНК, перенос мобильных генетических элементов и

			ограничение поступления чужеродной ДНК. Суперсемейство GIY-YIG объединяет нуклеазы, характеризующиеся наличием домена, состоящего обычно из ~100 аминокислот, с двумя короткими мотивами "GIY" и "YIG" в N-концевой части, за которыми следует остаток Arg в центре и остаток Glu в С-концевой части. Домен GIY-YIG образует компактный структурный домен, который служит каркасом для координации иона двухвалентного металла, необходимого для катализа разрыва фосфодиэфирной связи. Домен GIY-YIG характеризуется архитектурой типа альфа/бета-"сэндвича" с центральным трехнитевым антипараллельным бета-листом, фланкированным тремя спиралями. Трехнитевой антипараллельный бета-лист содержит элементы последовательности GIY-YIG. Наиболее консервативные и предполагаемые каталитические остатки расположены на неглубокой вогнутой поверхности и включают сайт координации металлов. Домен GIY-YIG участвует во множестве клеточных процессов, включающих расщепление ДНК, от саморазмножения с интронами или без них до ограничения поступления чужеродной ДНК, для восстановления ДНК и поддержания стабильности генома. Некоторые белки, которые, как известно, содержат домен GIY-YIG, включают следующие. * Эукариотические белки Slx-1, участвующие в поддержании числа копий рДНК. У них имеется С-концевой Zn-связывающий домен "палец" типа RING. * Анкириновый повтор млекопитающих и белок 1, содержащий домен LEM (ANKLE1). * Субъединицы UvrC эксцинуклеаз (A)BC бактерий и архей, которые удаляют поврежденные нуклеотиды, делая разрез поврежденной нити с обеих сторон от

			места поражения. Вирус Paramecium bursaria Chlorella virus 1 (pbvc1). * Эндонуклеазы SegA-SegE фага T4, вероятно, участвующие в перемещении ДНК, кодирующей эндонуклеазу. * Интрон-ассоциированная эндонуклеаза 1 фага T4 (I-TevI), специфичная в отношении границы сплайсинга гена тимидилатсинтазы (td) и участвующая в хоминге интронов. Эта запись представляет суперсемейство нуклеаз GIY-YIG и идентифицирует домен GIY-YIG в различных и отдаленно родственных представителях суперсемейства, таких как Slx1."
74	PS50164	Профиль домена GIY-YIG	"Нуклеазы семейства GIY-YIG участвуют во многих клеточных процессах, включая репарацию и рекомбинацию ДНК, перенос мобильных генетических элементов и ограничение поступления чужеродной ДНК. Суперсемейство GIY-YIG объединяет нуклеазы, характеризующиеся наличием домена, состоящего обычно из ~100 аминокислот, с двумя короткими мотивами "GIY" и "YIG" в N-концевой части, за которыми следует остаток Arg в центре и остаток Glu в C-концевой части [PMID:10219084,PMID:12379841,PMID:15692561,PMID:16646971,PMID:19361436]. Домен GIY-YIG образует компактный структурный домен, который служит каркасом для координации иона двухвалентного металла, необходимого для катализа разрыва фосфодиэфирной связи. Домен GIY-YIG характеризуется архитектурой типа альфа/бета-"сэндвича" с центральным трехнитевым антипараллельным бета-листом, фланкированным тремя спиралями. Трехнитевой антипараллельный бета-лист содержит элементы последовательности GIY-YIG. Наиболее консервативные и предполагаемые каталитические остатки расположены на неглубокой вогнутой

			поверхности и включают сайт координации металлов [PMID:12379841,PMID:15692561,PMID:16646971,PMID:19361436]. Некоторые белки, которые, как известно, содержат домен GIY-YIG, перечислены ниже. - Эукариотические белки Slx-1, участвующие в поддержании числа копий рДНК. У них имеется С-концевой Zn-связывающий домен "палец" типа RING. - Анкириновый повтор млекопитающих и белок 1, содержащий домен LEM (ANKLE1). - Субъединицы UvrC эксцинуклеаз (A)BC бактерий и архей, которые удаляют поврежденные нуклеотиды, делая разрез поврежденной нити с обеих сторон от места поражения. - Вирус Paramecium bursaria Chlorella virus 1 (pbvc1). - Эндонуклеазы SegA-SegE фага T4, вероятно, участвующие в перемещении ДНК, кодирующей эндонуклеазу. - Интрон-ассоциированная эндонуклеаза 1 фага T4 (I-TevI), специфичная в отношении границы сплайсинга гена тимидилатсинтазы (td) и участвующая в хоминге интронов." Разработанный авторами настоящего изобретения профиль охватывает весь домен GIY-YIG".
75	PF14111	Домен с неизвестной функцией (DUF4283)	"Это семейство доменов обнаруживается у растений и имеет длину примерно 100 аминокислот. С учетом очень широкого разнообразия других доменов, с которыми он ассоциирован, возможно, что этот домен является связывающей/направляющей областью. В нем имеются два высококонсервативных остатка триптофана."
76	IPR025558	Домен с неизвестной функцией DUF4283	"Этот домен обнаруживается в белках растений. С учетом очень широкого разнообразия других доменов, с которыми он ассоциирован, возможно, что этот

			домен является связывающей/направляющей областью. Он содержит два высококонсервативных остатка триптофана."
77	IPR040256	Неохарактеризованный белок, подобный At4g02000	"Эта запись представляет группу белков растений, включая At4g02000. Их функция неясна."
78	PF14392	"Цинковый сустав"	"Цинковый сустав" представляет собой цинк-связывающий мотив, состоящий из следующего: CX2CX4HX4C, где X может представлять собой любую аминокислоту. Это конкретное семейство обнаруживается в белках растений.
79	cd06222	"Суперсемейство белков, подобных рибонуклеазе H, включая РНКазу H, HI, HII, HIII и РНКазоподобный домен IV сплайсосомного белка Prp8"	"Ферменты рибонуклеазы H (РНКазы H) делятся на два основных семейства - тип 1 и тип 2 - на основании сходства аминокислотных последовательностей и биохимических свойств. РНКаза H представляет собой эндонуклеазу, которая расщепляет нить РНК гибрида РНК/ДНК неспецифичным в отношении последовательности образом в присутствии двухвалентных катионов. Она широко представлена в различных организмах, включая бактерии, археи и эукариот. Большинство прокариотических и эукариотических геномов содержат несколько генов РНКазы H. Несмотря на отсутствие гомологии аминокислотных последовательностей, РНКазы H типа 1 и типа 2 характеризуются общей укладкой основной цепи и стерической конфигурацией четырех кислотных остатков активного центра и характеризуются одинаковыми каталитическим механизмом и функциями в клетках. РНКаза H участвует в репликации, репарации и транскрипции ДНК. Важной функцией РНКазы H является удаление фрагментов Оказаки во время репликации ДНК. Ингибиторы РНКазы H исследовались в качестве мишени для

			лекарственных средств для лечения HIV, поскольку инактивация РНКазы H ингибирует обратную транскрипцию. Эта модель также включает домен IV Prp8, который принимает укладку РНКазы, но демонстрирует низкую гомологию последовательностей; домен IV участвует в ключевых сплайсосомных взаимодействиях. [PMID: 19228198, PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 14734815]"
80	PF13456	Домен, подобный обратной транскриптазе	Этот домен обнаруживается у растений и, по-видимому, является частью ретротранспозона.
81	IPR025836	"Цинковый сустав" CX2CX4HX4C	Этот "цинковый сустав" представляет собой цинк-связывающий мотив, состоящий из последовательности CX2CX4HX4C, где X может представлять собой любую аминокислоту.
82	PS51257	Профиль сайта прикрепления липидов липопротеинов прокариотической мембраны	"У прокариот мембранные липопротеины синтезируются с сигнальным пептидом-предшественником, который отщепляется специфической сигнальной пептидазой липопротеинов (сигнальной пептидазой II). Пептидаза распознает консервативную последовательность и делает разрез выше остатка цистеина, к которому присоединен липид в виде комплекса глицерид-жирная кислота [PMID:2202727]. Некоторые из белков, которые, как известно, подвергаются такому процессингу, в настоящее время включают следующие (свежие перечни см. в [PMID:2202727, PMID:3253732]). - Основные липопротеины наружной мембраны (муреин-липопротеины) (ген lpp). - Липопротеин 28 Escherichia coli (ген nlpA). - Липопротеин 34 Escherichia coli (ген nlpB). - Липопротеин nlpC Escherichia coli. - Липопротеин nlpD

			Escherichia coli. - Осмотически индуцируемый липопротеин B Escherichia coli (ген osmB). - Осмотически индуцируемый липопротеин E Escherichia coli (ген osmE). - Ассоциированный с пептидогликаном липопротеин Escherichia coli (ген pal). - Редкие липопротеины A и B Escherichia coli (гены rplA и rplB). - Белок гомеостаза меди cutF (или nlpE) Escherichia coli. - Белки traT плазмид Escherichia coli. - Белки лизиса плазмид Col Escherichia coli. - Ряд бета-лактамаз Bacillus. - Белок, связывающий периплазматический олигопептид Bacillus subtilis (ген oppA). - Белки наружной поверхности A и B Borrelia burgdorferi (гены ospA и ospB). - Вариабельный основной белок 21 (ген vmp21) и 7 (ген vmp7) Borrelia hermsii. - Белок наружной мембраны 3 Chlamydia trachomatis (ген omp3). - Эндоглюканаза cel-3 Fibrobacter succinogenes. - Белки Pal и Pcp Haemophilus influenzae. - Пуллуланаза Klebsiella (ген pulA). - Белок секреции пуллуланазы pulS Klebsiella. - Белок p37 Mycoplasma hyorhinis. - Варианты поверхностных антигенов A, B и C Mycoplasma hyorhinis (гены vlpABC). - Белок наружной мембраны H.8 Neisseria. - Липопептид Pseudomonas aeruginosa (ген lppL). - Эндоглюканаза egl Pseudomonas solanacearum. - Субъединица цитохрома реакционного центра Rhodopseudomonas viridis (ген cytC). - Антиген 17 Kd Rickettsia. - Плазмидные белки инвазии mxiJ и mxiM Shigella flexneri. - Белок транспорта олигопептидов A Streptococcus pneumoniae (ген amiA). - Антиген 34 Kd Treponema pallidium. - Мембранный белок A Treponema pallidium (ген tmpA). - Хитобиаза Vibrio harveyi (ген chb). - Плазмидный белок вирулентности yscJ Yersinia. - Галоцианин из Natrobacterium pharaonis [PMID:8195126] - мембраноассоциированный белок,

			связывающий медь. Это первый известный архебактериальный белок, модифицированный таким образом. Из последовательностей-предшественников всех этих белков авторы настоящего изобретения получили профиль, который начинается в начале последовательности и заканчивается после посттрансляционно модифицированного остатка цистеина."
83	cd05654	"Белок утилизации аргинина пептидазы M20, RocB"	"Семейство пептидаз M20, подсемейство RocB ArgE (белок утилизации аргинина RocB; белок разрушения аргинина RocB). Эта группа белков, возможно, родственна ацетилорнитиндеацетилазе (ArgE) и может участвовать в пути разрушения аргинина и/или орнитина. У Bacillus subtilis RocB является одним из трех генов, обнаруживаемых в опероне rocABC, который зависит от сигма-фактора L и индуцируется аргинином. Функция представителей этого семейства пока неизвестна, хотя согласно предсказаниям они представляют собой деацетилазы. [PMID: 7674922, PMID: 8113162]"
84	cd17374	Транспортеры органических анионов суперсемейства мембранных транспортеров	"Транспортеры органических анионов (OAT) обычно демонстрируют широкую субстратную специфичность и способствуют обмену внеклеточных органических анионов (OA) на внутриклеточные. Было охарактеризовано несколько OAT, включая OAT1-10 и уратный анионообменник 1 (URAT1, также называемый SLC22A12). Многие OAT встречаются в проксимальных канальцах почек - месте, где происходит активная секреция лекарственных средств. OAT опосредуют всасывание, распределение и экскрецию широкого круга токсинов окружающей среды и клинически важных лекарственных средств, включая терапевтические средства

			против HIV, противоопухолевые лекарственные средства, антибиотики, гипотензивные и противовоспалительные средства, и поэтому имеют решающее значение для выживания видов млекопитающих. OAT относится к семейству SLC22 (переносчиков растворенных веществ 22) из суперсемейства мембранных транспортеров (MFS) белков мембранного транспорта, которые, как считается, функционируют с использованием одного сайта связывания субстрата посредством механизма чередующегося доступа, включающего перемещение по типу кулисного переключателя. [PMID: 24745985, PMID: 23789957, PMID: 23506881, PMID: 23054972, PMID: 22841915, PMID: 16189627, PMID: 17245646, PMID: 17103332, PMID: 25540139, PMID: 27614971, PMID: 17502342, PMID: 17038320, PMID: 23920220, PMID: 26358290, PMID: 25904762, PMID: 25481222, PMID: 21446918, PMID: 23196129, PMID: 24646860, PMID: 19854166, PMID: 26758938, PMID: 26234418, PMID: 26098515, PMID: 23530251, PMID: 23403214, PMID: 18537473]"
85	cd00548	Цитохром-с-нитритредуктаза и сходные белки	"Это семейство содержит цитохром-с-нитритредуктазу (также известную как цитохром c552 или NrfA) и сходные белки. Пентагемный фермент NrfA катализирует восстановление электронов нитрита до аммиака в азотном цикле. Этот фермент также может превращать монооксид азота и гидроксиламин, два потенциально связанных промежуточных продукта реакции, в аммиак. Он представляет собой гомодимер, каждый мономер которого содержит четыре классических сайта связывания гема типа CXXCH вместе с альтернативным мотивом связывания гема типа CXXCK, который важен для катализа. Это семейство также включает октагемную

			нитритредуктазу (TvNiR) из галоалкалофильной бактерии Thioalkalivibrio paradoxus, которая катализирует восстановление нитрита и гидроксиламина до аммиака, а также восстановление сульфита до сульфида. [PMID: 25658043, PMID: 10440380, PMID: 11863430, PMID: 12618432, PMID: 18201106, PMID: 18597779, PMID: 22382353, PMID: 12165429, PMID: 12196156]"
86	cd09077	Эндонуклеазный домен, кодируемый различными ретротранспозонами, не содержащими длинных концевых повторов, принадлежащими к кладе R1 и I	"Это семейство содержит эндонуклеазный домен (EN) различных ретротранспозонов, не содержащих длинных концевых повторов (не содержащих LTR), длинных диспергированных ядерных элементов (LINE), которые принадлежат к подтипу 2, кладам R1 и I. LINE можно разделить на два подтипа. Подтип 2 имеет две ORF: вторая (ORF2) кодирует модульный белок, состоящий из N-концевого домена апуриновой/апиримидиновой эндонуклеазы (EN), центральной обратной транскриптазы и домена, подобного "цинковому пальцу", на С-конце. Большинство не содержащих LTR ретротранспозонов встроены по всему геному хозяина; однако многие ретротранспозоны клады R1 проявляют специфичную в отношении мишени ретротранспозицию. Это семейство включает эндонуклеазы SART1 и R1bm тутового шелкопряда Bombyx mori, которые принадлежат к кладе R1. Оно также включает эндонуклеазу улитки (Biomphalaria glabrata) Nimbus/Bgl и комара Aedes aegypti (MosquI), обе из которых принадлежат к кладе I. Это семейство принадлежит к большому суперсемейству EEP (экзонуклеаз/эндонуклеаз/фосфатаз), которое содержит функционально разнообразные ферменты, имеющие общий каталитический механизм расщепления фосфодиэфирных связей. [PMID: 12644555,

			PMID: 12949131, PMID: 15247245, PMID: 17537809, PMID: 10605110, PMID: 9364772, PMID: 9092665, PMID: 2554335, PMID: 2447482, PMID: 1690812, PMID: 17521654]"
87	PF07530	Домен, ассоциированный с "цинковыми пальцами"	Эта функция этого домена неизвестна [PMID:11779830] и часто обнаруживается ассоциированной с Pfam:PF00096.
88	IPR006579	Домен pre-C2HC	"Этот домен (PRE_C2HC) присутствует в белках, ассоциированных с ядерными функциями, и обнаруживается исключительно у членистоногих, включая ряд видов дрозофил, тутового шелкопряда и непарного шелкопряда. Эти белки, возможно, участвуют в связывании РНК или связывании однонитевой ДНК."
89	IPR031961	Домен с неизвестной функцией DUF4780	"Этот предполагаемый домен функционально не охарактеризован. Этот домен обнаруживается в белках насекомых и обычно имеет длину от 132 до 144 аминокислот. Существует единственный полностью консервативный остаток W, который может быть важен с функциональной точки зрения."
90	PF16012	Домен с неизвестной функцией (DUF4780)	"Этот предполагаемый домен функционально не охарактеризован. Это семейство доменов обнаруживается у эукариот и обычно имеет длину от 132 до 144 аминокислот. Существует единственный полностью консервативный остаток W, который может быть важен с функциональной точки зрения."
91	IPR012677	Суперсемейство связывающих нуклеотиды доменов со структурой типа альфа-бета-складки	"Это суперсемейство представляет связывающие нуклеотиды домены со структурой типа альфа-бета-складки, которая состоит либо из ферредоксиноподобной укладки (бета-альфа-бета)2, такой как обнаруживаемая в РНК-связывающих доменах различных рибонуклеопротеинов или в вирусных ДНК-связывающих доменах; либо

			из укладки бета(альфа)-бета-альфа-бета(2), такой как обнаруживаемая в рибосомном белке L23."
92	cd00590	Суперсемейство мотивов распознавания РНК (RRM)	"RRM, также известный как RBD (РНК-связывающий домен) или RNP (рибонуклеопротеиновый домен), представляет собой широко распространенный у эукариот домен, обнаруживаемый в белках, участвующих в посттранскрипционных процессах экспрессии генов, включая процессинг мРНК и рРНК, экспорт РНК и регуляцию стабильности РНК. Этот домен имеет длину 90 аминокислот и состоит из четырехнитевого бета-листа, упакованного между двумя альфа-спиралями. RRM обычно взаимодействует с ssRNA, но также известно, что он взаимодействует с ssDNA, а также с белками. RRM связывает различное количество нуклеотидов в диапазоне от двух до восьми. Активный центр содержит три ароматические боковые цепи, расположенные в пределах консервативных мотивов RNP1 и RNP2 домена. Домен RRM обнаруживается в различных гетерогенных ядерных рибонуклеопротеинах (hnRNP), белках, участвующих в регуляции альтернативного сплайсинга, и белковых компонентах малых ядерных рибонуклеопротеинов (snRNP). [PMID: 15853797, PMID: 22516180, PMID: 22278943, PMID: 18515081, PMID: 20458169, PMID: 7509757, PMID: 17473849, PMID: 18059478, PMID: 17589525, PMID: 16936729, PMID: 16432215, PMID: 16179478, PMID: 17036044, PMID: 16362037, PMID: 16885237, PMID: 10323862, PMID: 15769465, PMID: 8807823, PMID: 7535921, PMID: 10896472]"
93	IPR035979	Суперсемейство РНК-связывающих доменов	"Известно, что многие эукариотические белки, содержащие одну или несколько копий предполагаемого РНК-связывающего домена, состоящего из приблизительно

			90 аминокислот, связывают однонитевые РНК. Самой большой группой белков, связывающих однонитевые РНК, является семейство эукариотических мотивов распознавания РНК (RRM), которое содержит консенсусную последовательность RNP-1 из восьми аминокислот. Белки RRM характеризуются разнообразными предпочтениями в отношении связывания с РНК и функциями и включают гетерогенные ядерные рибонуклеопротеины (hnRNP), белки, участвующие в регуляции альтернативного сплайсинга (SR, U2AF, Sxl), белковые компоненты малых ядерных рибонуклеопротеинов (U1 и U2 snRNP) и белки, регулирующие стабильность и трансляцию РНК (PABP, La, Hu). RRM в гетеродимерном факторе сплайсинга, представляющем собой вспомогательный фактор U2 snRNP (U2AF), по-видимому, имеет два RRM-подобных домена со специализированными признаками для распознавания белков. Этот мотив также присутствует в нескольких белках, связывающих однонитевую ДНК. Типичный RRM состоит из четырех антипараллельных бета-нитей и двух альфа-спиралей, расположенных в укладке типа бета-альфа-бета-бета-альфа-бета с боковыми цепями, которые укладываются с основаниями РНК. Специфичность связывания РНК определяется множественными контактами с окружающими аминокислотами. В некоторых случаях во время связывания РНК присутствует третья спираль. RRM рассматривается в ряде публикаций."
94	IPR020847	"Сайт связывания АР-эндонуклеаз 1"	"Средства, повреждающие ДНК, такие как противоопухолевые лекарственные средства блеомицин и неокарциностатин, или такие, которые генерируют

			кислородные радикалы, вызывают различные повреждения ДНК. Среди них потеря оснований, при которой образуются апуриновые/апиримидиновые (AP) сайты или разрывы нити с атипичными 3'-концами. Репарация ДНК в AP-сайтах инициируется специфическим эндонуклеазным расщеплением фосфодиэфирного остова. Такие эндонуклеазы также обычно способны удалять блокирующие группы с 3'-конца разрывов нити ДНК. AP-эндонуклеазы можно разделить на два семейства на основании сходства последовательностей. Это семейство содержит представителей семейства AP-эндонуклеаз 1. За исключением Rrp1 и arp, эти ферменты представляют собой белки, состоящие из приблизительно 300 аминокислотных остатков. Оба из Rrp1 и arp содержат дополнительные и неродственные последовательности в своих N-концевых участках (приблизительно 400 остатков для Rrp1 и 270 остатков для arp). Эта запись представляет мотив, содержащий остаток глутамата, который, как было показано в ферменте Escherichia coli, связывает двухвалентный ион металла, такого как магний или марганец."
95	PS00726	Предполагаемая нуклеаза HARBI1	NA
96	PS50102	Профиль эукариотического мотива распознавания РНК (RRM)	"Многие эукариотические белки, которые, как известно или предполагается, связывают однонитевую РНК, содержат одну или несколько копий предполагаемого РНК-связывающего домена, состоящего из приблизительно 90 аминокислот [PMID:2470643,PMID:3072706]. Этот домен известен как мотив распознавания РНК (RRM). Эта область была обнаружена в следующих белках. ** Гетерогенные ядерные

			рибонуклеопротеины ** - hnRNP A1 (белок, дестабилизирующий спираль) (два раза). - hnRNP A2/B1 (два раза). - hnRNP C (C1/C2) (один раз). - hnRNP E (UP2) (по меньшей мере один раз). - hnRNP G (один раз). ** Малые ядерные рибонуклеопротеины ** - U1 snRNP 70 кДа (один раз). - U1 snRNP A (один раз). - U2 snRNP B'' (один раз). ** Белки, ассоциированные с пре-РНК и мРНК ** - Фактор инициации синтеза белка 4B (eIF-4B) [PMID:2390971], белок, необходимый для связывания мРНК с рибосомами (один раз). - Нуклеолин (4 раза). - Дрожжевой белок, связывающий однонитевую нуклеиновую кислоту (ген SSB1) (один раз). - Дрожжевой белок NSR1 (два раза). NSR1 участвует в процессинге пре-рРНК; он специфично связывает последовательности ядерной локализации. - Белок, связывающий поли(A)-хвост (PABP) (4 раза). ** Другие ** - Белок определения пола дрозофилы Sex-lethal (Sxl) (два раза). - Белок определения пола дрозофилы Transformer-2 (Tra-2) (один раз). - Белок 'elav' дрозофилы (3 раза), вероятно, участвующий в метаболизме РНК нейронов. - Антиген HuD паранеопластического энцефаломиелита человека (3 раза) [PMID:1655278], который весьма сходен с 'elav' и может играть роль в процессинге нейронспецифической РНК. - Белок "bicoid" дрозофилы (один раз) [PMID:2752425], гомеобоксный белок группы segment-polarity, который также может связываться со специфическими мРНК. - Антиген La (один раз), белок, который может играть роль в транскрипции РНК-полимеразы III. - Белок Ro 60 кДа (один раз), предполагаемый белок комплекса RNP. - Белок кукурузы, индуцируемый абсцизовой кислотой в ответ на водный стресс, который, по-видимому, представляет собой РНК-связывающий

			белок. - Три белка табака, расположенные в хлоропластах[PMID:1698606], которые могут участвовать в сплайсинге и/или процессинге хлоропластных РНК (два раза). - X16 [PMID:2030943], белок млекопитающих, который может участвовать в процессинге РНК в связи с клеточными пролиферацией и/или созреванием. - Инсулин-индуцированный белок ростового ответа CL-4 крысы (два раза). - Нуклеолизины TIA-1 и TIAR (3 раза)[PMID:1326761], обладающие нуклеолитической активностью в отношении клеток-мишеней, представляющих собой цитотоксические лимфоциты. Может участвовать в апоптозе. - Дрожжевой белок RNA15, играющий роль в обеспечении стабильности мРНК и/или длины поли(A)-хвоста [9]. Внутри RRM содержатся две высококонсервативные области. Первая представляет собой гидрофобный сегмент из шести остатков (называемый мотивом RNP-2), вторая представляет собой октапептидный мотив (называемый RNP-1 или RNP-CS). Положение обоих мотивов в домене показано в следующем схематическом представлении: xxxxxxx######xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx########xxxxxxxxxxxxxxxxxxxxxxxxx RNP-2 RNP-1. Авторы настоящего изобретения разработали профиль, который охватывает домен RRM."
97	PF00076	"Мотив распознавания РНК (также именуемый доменом RRM, RBD или RNP)"	"Мотив RRM, вероятно, является диагностическим признаком РНК-связывающего белка. RRM обнаруживаются в различных РНК-связывающих белках, включая различные белки hnRNP, белки, участвующие в регуляции альтернативного сплайсинга, и белковые компоненты snRNP. Этот мотив также присутствует в

			нескольких белках, связывающих однонитевую ДНК. Структура RRM состоит из четырех нитей и двух спиралей, расположенных в виде альфа/бета-"сэндвича", и в некоторых случаях во время связывания РНК присутствует третья спираль. С-концевую бета-нить (4 нить) и конечную спираль трудно выровнять, и они были опущены в выравнивании SEED. Белки LA (Swiss:P05455) имеют N-концевой rrm, который включен в SEED. Существует вторая область ближе к С-концу, которая имеет некоторые признаки, характерные для rrm, но, по-видимому, не имеет важной структурной сердцевины rrm. Белки LA (Swiss:P05455) являются одними из основных аутоантигенов при системной красной волчанке (SLE), аутоиммунном заболевании."
98	IPR000504	Домен мотива распознавания РНК	"Известно, что многие эукариотические белки, содержащие одну или несколько копий предполагаемого РНК-связывающего домена, состоящего из приблизительно 90 аминокислот, связывают однонитевые РНК. Самой большой группой белков, связывающих однонитевые РНК, является семейство эукариотических мотивов распознавания РНК (RRM), которое содержит консенсусную последовательность RNP-1 из восьми аминокислот. Белки RRM характеризуются разнообразными предпочтениями в отношении связывания с РНК и функциями и включают гетерогенные ядерные рибонуклеопротеины (hnRNP), белки, участвующие в регуляции альтернативного сплайсинга (SR, U2AF, Sxl), белковые компоненты малых ядерных рибонуклеопротеинов (U1 и U2 snRNP) и белки, регулирующие стабильность и трансляцию РНК (PABP, La, Hu). RRM в гетеродимерном факторе сплайсинга, представляющем собой вспомогательный фактор U2 snRNP (U2AF), по-

			видимому, имеет два RRM-подобных домена со специализированными признаками для распознавания белков. Этот мотив также присутствует в нескольких белках, связывающих однонитевую ДНК. Типичный RRM состоит из четырех антипараллельных бета-нитей и двух альфа-спиралей, расположенных в укладке типа бета-альфа-бета-бета-альфа-бета с боковыми цепями, которые укладываются с основаниями РНК. Специфичность связывания РНК определяется множественными контактами с окружающими аминокислотами. В некоторых случаях во время связывания РНК присутствует третья спираль. RRM рассматривается в ряде публикаций."
99	IPR004263	Экзостозин-подобный белок	"Существует пять идентифицированных белков семейства EXT человека (EXT1, EXT2, EXTL1, EXTL2 и EXTL3), которые являются представителями семейства супрессоров опухолей при наследственных множественных экзостозах. Они представляют собой гликозилтрансферазы, необходимые для биосинтеза гепарансульфата. Наследственные множественные экзостозы (EXT) представляют собой аутосомно-доминантное заболевание, характеризующееся появлением множественных выростов длинных костей (экзостозов) на их эпифизах. Мутации в двух гомологичных генах, EXT1 и EXT2, ответственны за синдром EXT. Гены EXT человека и мыши имеют по меньшей мере два гомолога у беспозвоночных Caenorhabditis elegans, что указывает на то, что они не функционируют исключительно как регуляторы роста костей. Было показано, что EXT1 и EXT2 кодируют гликозилтрансферазы, участвующие в стадии удлинения цепи биосинтеза

			гепарансульфата. Эта запись также включает гомологи ксилоглюкангалактозилтрансферазы Arabidopsis KATAMARI1 и EXT Drosophila melanogaster."
100	IPR004808	АР-эндонуклеаза 1	АР-эндонуклеазы можно разделить на два семейства на основании сходства последовательностей. Это семейство содержит представителей семейства АР-эндонуклеаз 1. Они представляют собой эндонуклеазы, которые удаляют поврежденную ДНК в положениях цитозина и гуанина путем расщепления с 3'-стороны от АР-сайта с помощью реакции бета-элиминирования. Белки содержат глутамат, который, как было показано в ферменте Escherichia coli, связывает ион двухвалентного металла, такого как магний или марганец. Средства, повреждающие ДНК, такие как противоопухолевые лекарственные средства блеомицин и неокарциностатин, или такие, которые генерируют кислородные радикалы, вызывают различные повреждения ДНК. Среди них потеря оснований, при которой образуются апуриновые/апиримидиновые (АР) сайты или разрывы нити с атипичными 3'-концами. Репарация ДНК в АР-сайтах инициируется специфическим эндонуклеазным расщеплением фосфодиэфирного остова. Такие эндонуклеазы также обычно способны удалять блокирующие группы с 3'-конца разрывов нити ДНК.
101	IPR000595	Домен, связывающий циклические нуклеотиды	"Белки, которые связывают циклические нуклеотиды (cAMP или cGMP), имеют общий структурный домен из приблизительно 120 остатков [PUB00013967], [PUB00013968], [PUB00005543]. Наиболее изученным из этих белков является

			прокариотический белок-активатор катаболитных генов (также известный как рецепторный белок cAMP) (ген crp), где известно, что такой домен состоит из трех альфа-спиралей и характерной восьминитевой антипараллельной структуры бета-"бочки". В этом домене содержатся шесть инвариантных аминокислот, три из которых представляют собой остатки глицина, которые, как считается, необходимы для поддержания структурной целостности бета-"бочки". cAMP- и cGMP-зависимые протеинкиназы (cAPK и cGPK) содержат две тандемные копии домена, связывающего циклические нуклеотиды. cAPK состоят из двух разных субъединиц - каталитической цепи и регуляторной цепи, которая содержит обе копии домена. cGPK представляют собой однонитевые ферменты, которые содержат две копии домена в своем N-концевом участке. Ионные каналы позвоночных, управляемые циклическими нуклеотидами, также содержат этот домен. Два таких катионных канала были полностью охарактеризованы, и один обнаружен в палочковидных клетках, где он играет роль в передаче зрительного сигнала."
102	PS50042	Профиль мотива связывания cAMP/cGMP	"Белки, которые связывают циклические нуклеотиды (cAMP или cGMP), имеют общий структурный домен из приблизительно 120 остатков [PMID:2550070,PMID:1710853,PMID:1313416]. Наиболее изученным из этих белков является прокариотический белок-активатор катаболитных генов (также известный как рецепторный белок cAMP) (ген crp), где известно, что такой домен состоит из трех альфа-спиралей и характерной восьминитевой антипараллельной структуры бета-"бочки". Известно, что такой домен существует в следующих белках. -

			Прокариотический белок-активатор катаболитных генов (CAP). - cAMP- и cGMP-зависимые протеинкиназы (cAPK и cGPK). Оба типа киназ содержат две тандемные копии домена, связывающего циклические нуклеотиды. cAPK состоят из двух разных субъединиц - каталитической цепи и регуляторной цепи, которая содержит обе копии домена. cGPK представляют собой однонитевые ферменты, которые содержат две копии домена в своем N-концевом участке. Нуклеотидная специфичность cAPK и cGPK обусловлена наличием аминокислоты в консервативной области бета-"бочки" 7: инвариантный в cGPK треонин в большинстве cAPK представляет собой аланин. - Ионные каналы позвоночных, управляемые циклическими нуклеотидами. Два таких катионных канала были полностью охарактеризованы. Один обнаружен в палочковидных клетках, где он играет роль в передаче зрительного сигнала. Он специфично связывается с cGMP, что приводит к открытию канала, и тем самым вызывает деполяризацию фоторецепторов палочек. В обонятельном эпителии аналогичный cAMP-связывающий канал играет роль в передаче обонятельного сигнала. В этом домене содержатся шесть инвариантных аминокислот, три из которых представляют собой остатки глицина, которые, как считается, необходимы для поддержания структурной целостности бета-"бочки". Авторы настоящего изобретения разработали два паттерна сигнатуры для этого домена. Первый паттерн расположен внутри бета-"бочек" 2 и 3 и содержит первые два консервативных Gly. Второй паттерн расположен внутри бета-"бочек" 6 и 7 и содержит третий консервативный Gly, а также три других инвариантных остатка."

103	cd00304	"RT_подобные: семейство, подобное обратной транскриптазе (RT, РНК-зависимой ДНК-полимеразе)."	"RT_подобные: семейство, подобное обратной транскриптазе (RT, РНК-зависимой ДНК-полимеразе)." Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]"
104	IPR001005	Домен SANT/Myb	"Ретровирусный онкоген v-myb и его клеточный аналог c-myb кодируют белки, связывающие ядерную ДНК. Было показано, что в myb одна из наиболее консервативных областей, состоящая из трех тандемных повторов, участвует в связывании ДНК. Домен SANT присутствует в корепрессорах ядерных рецепторов и в субъединицах многих комплексов, ремоделирующих хроматин. Он имеет сильное структурное сходство с ДНК-связывающим доменом Myb-родственных белков. Оба из них состоят из тандемных повторов трех альфа-спиралей, которые расположены в виде мотива спираль-поворот-спираль, причем каждая альфа-спираль содержит объемный ароматический остаток. Несмотря на общее сходство, существуют

			различия, указывающие на то, что домен SANT функционально отличается от канонического ДНК-связывающего домена Myb."
105	PF00096	"Цинковый палец", тип C2H2"	"Цинковый палец" C2H2 представляет собой классический домен "цинковый палец". Два консервативных остатка цистеина и гистидина координируют ион цинка. Следующий паттерн описывает "цинковый палец". #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C], где X может представлять собой любую аминокислоту, а числа в скобках указывают на количество остатков. Положения, отмеченные #, важны для стабильной укладки "цинкового пальца". Конечное положение может представлять собой либо his, либо cys. "Цинковый палец" C2H2 состоит из двух коротких бета-нитей, за которыми следует альфа-спираль. Аминоконцевая часть спирали связывается с большой бороздкой ДНК, связывающей "цинковые пальцы". Принятая консенсусная связывающая последовательность для Sp1 обычно определяется асимметричной гексануклеотидной сердцевиной GGGCGG, но эта последовательность не содержит, среди прочего, повтор GAG (=CTC), который представляет собой высокоаффинный сайт для связывания Sp1 с промотором wt1[PMID:12683994]."
106	cd00167	"ДНК-связывающие домены 'SWI3, ADA2, N-CoR и TFIIIB'. Тандемные копии домена связывают тандемные повторы	"ДНК-связывающие домены 'SWI3, ADA2, N-CoR и TFIIIB'. Тандемные копии домена связывают тандемные повторы теломерной ДНК как часть кэпирующего комплекса. Связывание является зависимым от последовательности в случае повторов, которые содержат G/C-богатый мотив [C2-3A(CA)1-6]. Этот домен также обнаруживается в

		теломерной ДНК как часть кэпирующего комплекса."	регуляторных комплексах репрессоров транскрипции, где он также связывается с ДНК. [PMID: 8882580, PMID: 9048479, PMID: 10652526, PMID: 11572773]"
107	PF00249	Myb-подобный ДНК-связывающий домен	"Это семейство содержит ДНК-связывающие домены белков Myb, а также семейство доменов SANT [PMID:8882580]."
108	IPR009057	Суперсемейство гомеобокс-подобных доменов	"Белки с гомеобоксным доменом (также известным как гомеодомен) представляют собой факторы транскрипции, которые имеют общий родственный ДНК-связывающий гомеодомен. Гомеодомен был впервые идентифицирован в ряде гомеозисных и сегментационных белков дрозофилы, но теперь известно, что он является высококонсервативным у многих других животных, включая позвоночных. Домен связывает ДНК посредством структуры спираль-поворот-спираль (HTH). Мотив HTH характеризуется двумя альфа-спиралями, тесно контактирующими с ДНК и соединенными коротким поворотом. Вторая спираль связывается с ДНК с помощью ряда водородных связей и гидрофобных взаимодействий, которые происходят между определенными боковыми цепями и доступными основаниями и метильными группами тимина в большой бороздке ДНК. Первая спираль помогает стабилизировать структуру. Многие белки содержат гомеодомены, в том числе Engrailed Drosophila, белки, определяющие тип спаривания дрожжей, ядерный фактор гепатоцитов 1a и белки HOX. Мотив гомеодомена весьма сходен по последовательности и структуре с доменами в широком диапазоне ДНК-связывающих белков, включая рекомбиназы, белки Myb, регуляторы ответа GARP, теломерные белки человека (hTRF1), белки с парными доменами (PAX), RAP1

			дрожжей, белки, связывающие центромеру, CENP-B и ABP-1, регуляторы транскрипции (TyrR), активаторы транскрипции типа AraC и белки, подобные репрессору тетрациклина (TetR, QacR, YcdC)."
109	IPR001087	Липаза/эстераза GDSL	"Эстеразы и липазы GDSL представляют собой гидролитические ферменты с многофункциональными свойствами. Этот новый подкласс липолитических ферментов содержит отдельный мотив последовательности GDSL, отличный от мотива GxSxG, обнаруживаемого во многих липазах. Представители включают: липазу Aeromonas hydrophila, арилэстеразу Vibrio mimicus, термолабильный гемолизин Vibrio parahaemolyticus, фосфолипазу кролика (AdRab-B) и белок, богатый пролином, специфичный для пыльников Brassica napus."
110	PF00657	GDSL-подобная липаза/ацилгидролаза	NA
111	IPR033447	Нектин-3	NA
112	PF17182	Домен OSK	Эта запись представляет домен OSK, определенный Jeske и соавторами [PMID:26190108]. Домен родственен гидролазам SGNH, но в нем отсутствуют остатки активного центра. Домен связывается с РНК [PMID:26190108].
113	cd09881	"VapC-подобный домен PIN Mycobacterium tuberculosis VapC4 и VapC5, а также FitB	"Это семейство включает подобный ассоциированному с вирулентностью белку C (VapC) домен PIN (N-концевой PilT) рибонуклеазных токсинов VapC4 и VapC5 системы токсин/антитоксин (TA) VapBC Mycobacterium tuberculosis и токсина FitB системы TA FitAB Neisseria gonorrhoeae. Это семейство принадлежит к семейству

		Neisseria gonorrhoeae и родственные белки"	VapC-подобных белков суперсемейства нуклеаз с доменом PIN. VapC представляет собой рибонуклеазный токсин с доменом PIN из прокариотических систем токсин-антитоксин (TA) VapBC. VapB представляет собой белковый антитоксин, подобный фактору транскрипции, выступающий в качестве ингибитора. Другие представители семейства VapC-подобных нуклеаз включают эукариотические рибонуклеазы, такие как Smg6, фактор сборки рибосом NOB1, эндорибонуклеазу субъединицы экзосомы Rrp44 и белок процессинга рРНК Fcf1. Структурные свойства домена PIN (N-концевого PilT) указывают на его активный центр, состоящий из трех высококонсервативных каталитических остатков, которые координируют ионы металлов, и у некоторых представителей могут быть обнаружены дополнительные остатки, координирующие металлы. У некоторых представителей суперсемейства отсутствуют некоторые из этих ключевых каталитических остатков. Активный центр PIN геометрически сходен в активном центре со структурноспецифическими 5'-нуклеазами, рибонуклеазами с доменом PIN эукариотических белков, редактирующих рРНК, и бактериальными токсинами оперонов системы токсин-антитоксин (TA). [PMID: 21036780, PMID: 22140099, PMID: 26833558, PMID: 27998932, PMID: 19445953, PMID: 18952600, PMID: 17496075, PMID: 16982615, PMID: 15993073, PMID: 14734548, PMID: 11137022, PMID: 28575517, PMID: 28508407]"
114	IPR019793	Сайт связывания гемового лиганда пероксидаз	"Пероксидазы представляют собой гемсодержащие ферменты, которые используют пероксид водорода в качестве акцептора электронов для катализа ряда окислительных реакций. Большинство гемсодержащих пероксидаз следуют схеме

			реакции: Fe3+ + H2O2 --> [Fe4+=O]R' (соединение I) + H2O [Fe4+=O]R' + субстрат --> [Fe4+=O]R (соединение II) + окисленный субстрат [Fe4+=O]R+субстрат --> Fe3+ + H2O+окисленный субстрат. В этом механизме фермент вступает в реакцию с одним эквивалентом H2O2 с образованием [Fe4+=O]R' (соединение I). Это двухэлектронная реакция окисления/восстановления, при которой H2O2 восстанавливается до воды, а фермент окисляется. Один окислительный эквивалент приходится на железо, образуя промежуточный продукт оксиферрил, тогда как во многих пероксидазах порфирин (R) окисляется до пи-катион-радикала порфирина (R'). Затем соединение I окисляет органический субстрат с образованием радикала субстрата. Гемсодержащие пероксидазы включают два суперсемейства: одно обнаруживается у бактерий, грибов, растений, а второе обнаруживается у животных. Первое можно рассматривать как состоящее из 3 основных классов. Класс I, внутриклеточные пероксидазы, включает: дрожжевую цитохром-с-пероксидазу (CCP) - растворимый белок, обнаруживаемый в митохондриальной цепи переноса электронов, где он, вероятно, обеспечивает защиту от токсичных пероксидов; аскорбатпероксидазу (AP) - основной фермент, ответственный за удаление пероксида водорода в хлоропластах и цитозоле высших растений; и бактериальные каталазы-пероксидазы, обладающие как пероксидазной, так и каталазной активностью. Считается, что каталаза-пероксидаза обеспечивает защиту клеток при окислительном стрессе. Класс II состоит из секреторных грибных пероксидаз: лигниназ или лигнинпероксидаз (LiP) и марганец-зависимых пероксидаз (MnP). Они представляют

			собой мономерные гликопротеины, участвующие в разрушении лигнина. В MnP восстанавливающим субстратом служит Mn2+. Белки класса II содержат четыре консервативных дисульфидных мостика и два консервативных сайта связывания кальция. Класс III состоит из секреторных пероксидаз растений, которые выполняют множество тканеспецифических функций: например, удаление пероксида водорода из хлоропластов и цитозоля; окисление токсичных соединений; биосинтез клеточной стенки; защитные ответы на повреждения; катаболизм индол-3-уксусной кислоты (IAA); биосинтез этилена и т. д. Белки класса III также представляют собой мономерные гликопротеины, содержащие четыре консервативных дисульфидных мостика и два иона кальция, хотя расположение дисульфидов отличается от такового у ферментов класса II. Кристаллические структуры ряда этих белков показывают, что они имеют одинаковую архитектуру - два домена, полностью состоящих из альфа-спиралей, между которыми встроена группа гема. Эта запись представляет сайт связывания гема в ряде пероксидаз."
115	PS00435	Предполагаемая нуклеаза HARBI1	NA
116	cd16965	Альфа-киназный домен канальных киназ	"Эта группа состоит из канальных киназ 1 (ChaK1) и 2 (ChaK2) и сходных белков. ChaK1 и ChaK2 также называют представителями 7 (TRMP7) и 6 (TRMP6) подсемейства M катионных каналов с транзиторным рецепторным потенциалом соответственно. Они представляют собой слитые белки, содержащие трансмембранную ионную пору или канал и C-концевой альфа-киназный домен, оба

			из которых являются функциональными. Оба они являются катионоселективными каналами, которые предпочтительно обеспечивают проникание ионов Zn2+, Mg2+ и Ca2+. Они являются центральными регуляторами гомеостаза Mg2+ и Ca2+. TRMP7 экспрессируется повсеместно, тогда как TRMP6 экспрессируется на высоком уровне в определенных тканях, таких как почки и кишечник. Альфа-киназа представляет собой атипичный каталитический протеинкиназный домен, не имеющий заметного сходства с традиционными серин/треониновыми протеинкиназами. Семейство альфа-киназ было названо в честь уникального способа распознавания субстрата его первоначальными представителями, киназами тяжелой цепи Dictyostelium, которые были нацелены на последовательности белков, принимающие альфа-спиральную конформацию. Недавно было обнаружено, что альфа-киназы также нацеливаются на остатки в неспиральных областях. [PMID: 11161216, PMID: 10021370, PMID: 11389851, PMID: 26218331, PMID: 26179995, PMID: 25437439, PMID: 25204892, PMID: 24756720, PMID: 24752033, PMID: 24634592, PMID: 24906182, PMID: 24756719, PMID: 20932259, PMID: 19780399, PMID: 18660673, PMID: 20012461]"
117	cd00603	Домен IPT плексинов, и рецепторов клеточной поверхности (PCSR), и родственных белков.	"Домен IPT плексинов, и рецепторов клеточной поверхности (PCSR), и родственных белков. Эта подгруппа содержит домены IPT плексинов, рецепторов, таких как рецепторы факторов роста, родственных плазминогену, факторов роста гепатоцитов - рассеивающих факторов, а также рецепторов, стимулирующих макрофаги, и фиброцистина. Плексины участвуют в регуляции клеточной пролиферации и рецепторов клеточной адгезии и отталкивания. Как правило, присутствуют три копии

			домена IPT_PCSR, которым предшествуют домены SEMA (семафорин) и PSI (плексин, семафорин, интегрин). [PMID: 11919560, PMID: 12367632, PMID: 10027283, PMID: 10520995]"
118	cd17772	Два тандемных повтора цистатионин-бета-синтазных доменов (пара CBS), ассоциированные с доменами DHH и нуклеотидилтрансферазы (NT)	"Этот cd содержит два тандемных повтора цистатионин-бета-синтазных доменов (пару CBS), ассоциированных с вышерасположенным доменом DHH, который выполняет функцию фосфоэстеразы, и нижерасположенным доменом нуклеотидилтрансферазы (NT) ДНК-полимераз семейства X. Домен CBS, названный в честь CBS человека, представляет собой небольшой домен, первоначально идентифицированный в цистатионин-бета-синтазе и впоследствии обнаруженный в широком диапазоне различных белков. Домены CBS обычно встречаются в тандемных повторах. Они связываются с образованием так называемого домена Бейтмана или пары CBS, исходя из кристаллографических исследований бактерий. Пара CBS использовалась в качестве основы для этой иерархии cd, поскольку белки CBS человека могут принимать типичную сердцевинную структуру и образовывать внутримолекулярную пару CBS. Область контакта между двумя доменами CBS образует расщелину, которая является потенциальным сайтом связывания лиганда. Пара CBS сосуществует с множеством других функциональных доменов, и это было использовано в данном документе для содействия в ее классификации. Было высказано предположение, что домен CBS может играть регулирующую роль, хотя его точная функция неизвестна. Мутации консервативных остатков в этом домене ассоциированы с различными наследственными заболеваниями человека, включая

			врожденную миотонию, идиопатическую генерализованную эпилепсию, гиперкальциурический нефролитиаз и классический синдром Барттера (представители семейства хлоридных каналов CLC), синдром Вольфа-Паркинсона-Уайта (субъединица гамма 2 AMP-активируемой протеинкиназы), пигментный ретинит (IMP-дегидрогеназа-1) и гомоцистинурию (цистатионин-бета-синтаза). [PMID: 16275737, PMID: 9020585, PMID: 9106071, PMID: 9590298, PMID: 10200156, PMID: 11230183, PMID: 11524006, PMID: 14722619]"
119	IPR006939	SNF5/SMARCB1/INI1	"SNF5 представляет собой компонент дрожжевого комплекса SWI/SNF, который представляет собой ATP-зависимый комплекс ремоделирования нуклеосом, который регулирует транскрипцию подмножества дрожжевых генов. SNF5 представляет собой ключевой компонент всех охарактеризованных до сих пор комплексов класса SWI/SNF. Это семейство состоит из консервативной области SNF5, включая мотив с прямыми повторами. SNF5 необходим для нацеливания на промотор сборки и обеспечения активности ремоделирования хроматина комплекса SWI-SNF. SNF5 также известен как SMARCB1, что означает родственный SWI/SNF, матрикс-ассоциированный, актинзависимый регулятор хроматина, подсемейство b, представитель 1, а также как INI1, что означает интерактор интегразы 1. Считается, что мутации потери функции в SNF5 способствуют онкогенезу злокачественных рабдоидных опухолей (MRT)."

120	IPR038765	Суперсемейство папаиноподобных цистеинпептидаз	Эта запись представляет суперсемейство папаиноподобных цистеинпептидаз. Его представители обычно состоят из одной альфа-спирали и 4 нитей антипараллельных бета-листов и содержат каталитическую триаду Cys-His-Asn.
121	PF02257	ДНК-связывающий домен RFX	"RFX представляет собой регуляторный фактор, который связывается с X-боксом генов MHC класса II и необходим для их экспрессии. ДНК-связывающий домен RFX является центральным доменом белка и связывает ssDNA либо в виде мономера, либо в виде гомодимера [PMID:2253877]. Он распознает X-боксы (ДНК с последовательностью 5'-GTNRCC(0-3N)RGYAAC-3', где N представляет собой любой нуклеотид, R представляет собой пурин, а Y представляет собой пиримидин), используя высококонсервативный ДНК-связывающий домен (DBD) из 76 остатков [PMID:10706293]."
122	IPR039779	RFX-подобный ДНК-связывающий белок	"Эта запись включает Rfx1 почкующихся дрожжей, Sak1 делящихся дрожжей и RFX1-8 человека. Они представляют собой факторы транскрипции, которые содержат ДНК-связывающий домен RFX и консервативную С-концевую область."
123	IPR036390	Суперсемейство ДНК-связывающих доменов с "крылатой спиралью"	"ДНК-связывающие белки с "крылатой спиралью" имеют родственный ДНК-связывающий мотив типа "крылатая спираль"-поворот-спираль, где "крылья" или петли представляют собой небольшие бета-листы. Мотив типа "крылатой спирали" состоит из двух крыльев (W1, W2), трех альфа-спиралей (H1, H2, H3) и трех бета-листов (S1, S2, S3), расположенных в порядке H1-S1-H2-H3-S2-W1-S3-W2. Спираль распознавания ДНК образует специфичные в отношении последовательности контакты с ДНК в большой бороздке ДНК, тогда как "крылья" образуют другие

			контакты с ДНК, часто с малой бороздкой или остовом ДНК. Некоторые белки с "крылатыми спиралями" демонстрируют доступные участки из гидрофобных остатков, которые, как полагают, опосредуют белок-белковые взаимодействия."
124	IPR003150	ДНК-связывающий домен с "крылатой спиралью" типа RFX	"Семейство факторов транскрипции RFX характеризуется уникальным ДНК-связывающим доменом, состоящим из примерно 75 остатков." Гены RFX были выделены у дрожжей, нематод и позвоночных. Характерный ДНК-связывающий домен типа RFX со структурой HTH был привлечен в очень разные в иных отношениях регуляторные факторы, функционирующие в разнообразном спектре неродственных систем, включая регуляцию митотического клеточного цикла у делящихся дрожжей, контроль иммунного ответа у млекопитающих и инфицирование вирусом гепатита B человека. ДНК-связывающий домен типа RFX является необычным представителем подсемейства доменов со структурой спираль-поворот-спираль с "крылатой спиралью", поскольку он использует бета-шпильку (или "крыло") для распознавания ДНК вместо распознающей спирали, типичной для доменов со структурой спираль-поворот-спираль. Он состоит из трех альфа-спиралей (H), трех бета-нитей (S) и трех соединяющих петель (L), расположенных в порядке H1-S1-H2-L1-H3-L2-S2-W1-S3. Третья петля, соединяющая бета-нити S2 и S3, образует "крыло" W1 мотива типа "крылатой спирали". В отличие от ранее описанных DBD с "крылатой спиралью" (обычно длиной 110 остатков с двумя "крыльями", W1 и W2), более короткий ДНК-связывающий домен с "крылатой спиралью" типа RFX имеет только одно "крыло"."

125	IPR036388	Суперсемейство доменов, подобных ДНК-связывающим доменам с "крылатой спиралью"	"ДНК-связывающие белки с "крылатой спиралью" имеют родственный ДНК-связывающий мотив типа "крылатая спираль"-поворот-спираль, где "крылья" или петли представляют собой небольшие бета-листы. Мотив типа "крылатой спирали" состоит из двух крыльев (W1, W2), трех альфа-спиралей (H1, H2, H3) и трех бета-листов (S1, S2, S3), расположенных в порядке H1-S1-H2-H3-S2-W1-S3-W2. Спираль распознавания ДНК образует специфичные в отношении последовательности контакты с ДНК в большой бороздке ДНК, тогда как "крылья" образуют другие контакты с ДНК, часто с малой бороздкой или остовом ДНК. Некоторые белки с "крылатыми спиралями" демонстрируют доступные участки с гидрофобными остатками, которые, как считается, опосредуют белок-белковые взаимодействия. Многие различные белки с разнообразными биологическими функциями содержат ДНК-связывающий домен с "крылатой спиралью", включая репрессоры транскрипции, такие как репрессор биотина, репрессор LexA и репрессор аргинина; факторы транскрипции, такие как белки ядерного фактора гепатоцитов 3, участвующие в клеточной дифференцировке, фактор транскрипции белков теплового шока и общие факторы транскрипции TFIIE и TFIIF; хеликазы, такие как RuvB, которая способствует миграции ветвей в структуре Холлидея, и CDC6 в пререпликационном комплексе; эндонуклеазы, такие как FokI и TnsA; гистоны и транспозазу Mu, где гибкое "крыло" домена, связывающего энхансер, необходимо для эффективной транспозиции."

126	IPR037800	Гистонацетилтрансфераза GCN5	"Комплекс ацетилирования гистонов, регулирующий транскрипцию, Spt-Ada-Gcn5-ацетилтрансфераза (SAGA) участвует в зависимой от РНК-полимеразы II регуляции транскрипции примерно 10% генов дрожжей. SAGA предпочтительно ацетилирует гистоны H3 и H2B и деубиквитинирует гистон H2B. SAGA известен как PCAF у позвоночных, и PCAF ацетилирует нуклеосомный гистон H3. Комплекс SAGA состоит из по меньшей мере TRA1, CHD1, SPT7, TAF5, ADA3, SGF73, SPT20/ADA5, SPT8, TAF12, TAF6, HFI1/ADA1, UBP8, GCN5, ADA2, SPT3, SGF29, TAF10, TAF9, SGF11 и SUS1, и некоторые из этих компонентов присутствуют в двух копиях. Комплекс образован отдельными модулями, каждый из которых имеет отдельную функцию и сшивается либо с другими белками, либо с другими модулями в комплексе. SLIK (SAGA-подобный комплекс) представляет собой многосубъединичный комплекс гистонацетилтрансфераз, который предпочтительно ацетилирует гистоны H3 и H2B и деубиквитинирует гистон H2B. Он представляет собой усовершенствование комплекса SAGA. Комплекс SLIK дрожжей состоит из по меньшей мере TRA1, CHD1, SPT7, CC TAF5, ADA3, SPT20, RTG2, TAF12, TAF6, HFI1, UBP8 (деубиквитиназы), GCN5, ADA2, SPT3, SGF29, TAF10 и TAF9. Комплекс SALSA дрожжей представляет собой измененную форму комплекса SAGA и состоит из по меньшей мере TRA1, SPT7 (формы, усеченной на C-конце), TAF5, ADA3, SPT20, TAF12, TAF6, HFI1, GCN5, ADA2 и SPT3. Комплекс ADA представляет собой комплекс ацетилирования гистонов (HAT), регулирующий транскрипцию. ADA предпочтительно ацетилирует нуклеосомные гистоны H3 (в положениях 'Lys-14' и 'Lys-18') и H2B. Комплекс состоит из по меньшей

			мере ADA2, ADA3, AHC1 и GCN5. AHC1 необходим для общей структурной целостности комплекса ADA. Эта запись включает гистонацетилтрансферазы GCN5, KAT2A и KAT2B (все они включены в EC:2.3.1.48). GCN5 ацетилирует гистоны H2B, H3 и H4, обеспечивая образование специфической метки для эпигенетической активации транскрипции. GCN5 представляет собой компонент комплексов ацетилирования гистонов (HAT), регулирующих транскрипцию, SAGA, SLIK, SALSA и ADA. У млекопитающих имеется два паралога: KAT2A (также известный как GCN5) и KAT2B. KAT2A ацетилирует коровые гистоны, обеспечивая образование специфической метки для эпигенетической активации транскрипции, но не коровые частицы нуклеосомы. Он также ацетилирует белки, такие как CEBPB. KAT2A представляет собой компонент комплекса ATAC, обладающего ацетилтрансферазной активностью в отношении гистонов H3 и H4. KAT2B (также известный как фактор, ассоциированный с P300/кальцийсвязывающим белком (CBP), или PCAF) может ацетилировать коровые гистоны H3 и H4, а также коровые частицы нуклеосомы и негистоновые белки, такие как ACLY."
127	PF00628	"Палец" типа PHD	PHD укладывается в Zn-"палец" чередующегося типа, хелатирующий 2 иона Zn, аналогично доменам RING и FYVE [PMID:11124022]. Несколько "пальцев" типа PHD были идентифицированы как модули связывания метилированного гистона H3 [PMID:17142463].
128	IPR000104	"Антифризный белок, тип I"	"Морские костистые рыбы из полярных океанов могут быть защищены от замерзания в ледяной морской воде с помощью сывороточных антифризных белков

			(AFP) или антифризных гликопротеинов (AFGP): они функционируют путем связывания с кристаллами льда внутри рыбы и предотвращения их роста, а также снижения неравновесной точки замерзания до уровня ниже точки плавления. Несмотря на функциональное сходство, антифризные белки являются структурно разнообразными и включают гликозилированные и по меньшей мере 3 негликозилированные формы: AFGP нототениевых и тресковых представляют собой полимеры с трипептидным повтором Ala-Ala-Thr с дисахаридом, присоединенным к остатку треонина; AFP типа I обнаруживаются у камбал и скорпен; AFP типа II у волосатой рогатки, корюшки и сельди представляют собой Cys-богатые белки; и AFP типа III, обнаруживаемые у бельдюг, богаты бета-структурой. У насекомых и растений также были идентифицированы негомологичные антифризные белки. AFP типа I представляют собой Ala-богатые амфифильные альфа-спиральные белки. Сайты связывания льда всех AFP относительно плоские и гидрофобные и имеют непрерывный участок из остатков аланина, проходящий по длине повтора спирали, составляющей примерно 16,5A. На основе структуры с минимизированной энергией была предложена модель для описания связывания белка с кристаллами льда, при этом белок связывается со структурой, отвечающей за образование центров кристаллизации льда, по типу "молнии" посредством водородных связей метильной группы треониновых боковых цепей (с промежутком из 11 остатков) с атомами кислорода в кристаллической решетке льда. Таким образом, рост кристаллов льда останавливается или замедляется, а точка замерзания снижается. Высокое

			содержание лизина в этих пептидах может способствовать растворимости этих белков."
129	cd00009	Суперсемейство AAA+ (ATPазы, ассоциированные с широким спектром клеточной активности) представляет собой древнюю группу ATPаз, принадлежащих к части ASCE (что означает "дополнительная нить, каталитический E") укладки NTPазы с P-петлей.	"Суперсемейство AAA+ (ATPазы, ассоциированные с широким спектром клеточной активности) представляет собой древнюю группу ATPаз, принадлежащих к категории ASCE (что означает "дополнительная нить, каталитический E") укладки NTPазы с P-петлей. Категория ASCE также включает ABC, RecA-подобные, VirD4-подобные, PilT-подобные и хеликазы SF1/2. Представители ATPаз AAA+ функционируют как молекулярные шапероны, ATPазные субъединицы протеаз, хеликазы или ATPазы, стимулируемые нуклеиновыми кислотами. Белки AAA+ содержат несколько отличительных признаков в дополнение к консервативной структуре сердцевинного домена типа альфа-бета-альфа и мотивам Уолкера A и B NTPаз с P-петлей. [PMID: 18208389, PMID: 15037234, PMID: 18466635, PMID: 9927482, PMID: 11473577, PMID: 11916378, PMID: 15381417]"
130	IPR027417	Нуклеозидтрифосфатгидролаза, содержащая P-петлю	"Укладка NTPазы с P-петлей является наиболее распространенным доменом нескольких различных укладок белков, связывающих нуклеотиды. Наиболее распространенной реакцией, катализируемой ферментами NTPазами с P-петлей, является гидролиз бета-/гамма-фосфатной связи связанного нуклеозидтрифосфата (NTP) Энергия гидролиза NTP обычно используется для индуцирования конформационных изменений в других молекулах, что составляет основу биологических функций большинства NTPаз с P-петлей. NTPазы с P-петлей демонстрируют существенное предпочтение субстрата в отношении либо ATP, либо

			GTP. NTPазы с P-петлей характеризуются двумя консервативными сигнатурами последовательности: мотивом Уолкера A (собственно P-петля) и мотивами Уолкера B, которые связывают соответственно бета- и гамма-фосфатные фрагменты связанного NTP и катион Mg2+. Домены P-петли ATPаз принадлежат к одной из двух основных категорий. Категория киназ-GTPаз (KG) включает киназы и GTPазы, а категория ASCE характеризуется наличием дополнительной нити в сердцевинном листе, которая расположена между нитью P-петли и нитью мотива Уолкера B. Большинство представителей категории ASCE используют ATP, и представители этой группы включают AAA+, ABC, PilT, HerA-FtsK, хеликазы суперсемейства 1/2 (SF1/2) и суперсемейства ATPаз RecA/ATP-синтазы и т. д."
131	PF15288	"Цинковый сустав"	Этот "цинковый сустав" обнаруживается в белках млекопитающих FAM90A.
132	IPR041670	"Цинковый сустав"	Этот "цинковый сустав" обнаруживается в белках FAM90A и фактора инициации транскрипции TFIID.
133	PF01443	Вирусная РНК-хеликаза (суперсемейства 1)	"Для этого семейства была продемонстрирована хеликазная активность [PMID:8269709] и NTPазная активность [PMID:10217401]. Эта хеликаза играет множество ролей на разных стадиях репликации вирусной РНК, как представлено с помощью мутационного анализа [PMID:8057461]."
134	IPR027351	Сердцевинный домен хеликазы РНК(+)-вируса	"Хеликазы были разделены на 6 суперсемейств (SF1-SF6). Все белки связывают ATP и, следовательно, все они несут классические мотивы Уолкера A (фосфатсвязывающая петля или P-петля) и Уолкера B (Mg2+-связывающая аспарагиновая кислота). Два крупнейших суперсемейства, обычно обозначаемые

			как SF1 и SF2, имеют сходные паттерны из семи консервативных мотивов последовательностей, некоторые из которых разделены длинными низкоконсервативными спейсерами. Мотивы хеликаз, по-видимому, организованы в виде сердцевинного домена, который обеспечивает каталитическую функцию, тогда как необязательные вставки и амино- и карбоксиконцевые последовательности могут содержать отдельные домены с различными вспомогательными ролями. Сердцевинный домен хеликазы содержит два структурных домена: N-концевой ATP-связывающий домен и C-концевой домен. Предполагаемые хеликазы SF1 чрезвычайно широко распространены среди вирусов с положительной нитью РНК(+). Они были идентифицированы в различных семействах вирусов растений, а также в альфа-, руби-, артеривирусах, вирусе гепатита E и коронавирусах. Ряд этих вирусных ферментов участвует в различных аспектах транскрипции и репликации, а также в обеспечении стабильности РНК и межклеточном движении. Сердцевинный домен хеликазы РНК(+)-вируса содержит два RecA-подобных домена типа альфа/бета. N-концевой ATP-связывающий домен содержит параллельный шестинитевой бета-лист, окруженный четырьмя спиралями с одной стороны и двумя спиралями с другой. C-концевой домен содержит параллельный четырехнитевой бета-лист, расположенный в виде "сэндвича" между двумя спиралями с каждой из его сторон. Сердцевинный домен хеликазы РНК(+)-вируса, вероятно, связывает NTP в расщелине между N-концом ATP-связывающего домена и началом C-концевого домена. Эта запись представляет сердцевинный домен хеликазы РНК(+)-вируса."

135	PS51657	Профиль сердцевинного домена хеликазы РНК(+)-вируса	"Хеликазы были разделены на 6 суперсемейств (SF1-SF6) [PMID:22573863]. Все белки связывают АТР и, следовательно, все они несут классические мотивы Уолкера А (фосфатсвязывающая петля или Р-петля) и Уолкера В (Mg2+-связывающая аспарагиновая кислота) [1]. Два крупнейших суперсемейства, обычно обозначаемые как SF1 и SF2, имеют сходные паттерны из семи консервативных мотивов последовательностей, некоторые из которых разделены длинными низкоконсервативными спейсерами [1]. Мотивы хеликаз, по-видимому, организованы в виде сердцевинного домена, который обеспечивает каталитическую функцию, тогда как необязательные вставки и амино- и карбоксиконцевые последовательности могут содержать отдельные домены с различными вспомогательными ролями. Сердцевинный домен хеликазы содержит два структурных домена: N-концевой АТР-связывающий домен и С-концевой домен. Предполагаемые хеликазы SF1 чрезвычайно широко распространены среди вирусов с положительной нитью РНК(+). Они были идентифицированы в различных семействах вирусов растений, а также в альфа-, руби-, артеривирусах, вирусе гепатита Е и коронавирусах. Ряд этих вирусных ферментов участвует в различных аспектах транскрипции и репликации, а также в обеспечении стабильности РНК и межклеточном движении [PMID:11000230]. Сердцевинный домен хеликазы РНК(+)-вируса содержит два RecA-подобных домена типа альфа/бета. N-концевой АТР-связывающий домен содержит параллельный шестинитевой бета-лист, окруженный четырьмя спиралями с одной стороны и двумя спиралями с другой. С-концевой

			домен содержит параллельный четырехнитевой бета-лист, расположенный в виде "сэндвича" между двумя спиралями с каждой из его сторон. Сердцевинный домен хеликазы РНК(+)-вируса, вероятно, связывает NTP в расщелине между N-концом АТР-связывающего домена и началом С-концевого домена [PMID:22573863]. Разработанный авторами настоящего изобретения профиль охватывает весь сердцевинный домен хеликазы РНК(+)-вируса."
136	cd18809	С-концевой домен хеликазы в хеликазах семейства RecD	"RecD является представителем комплекса RecBCD (EC 3.1.11.5, экзонуклеаза V). Он представляет собой альфа-цепь комплекса и функционирует как 3'-5'-хеликаза. Фермент RecBCD представляет собой как хеликазу, которая раскручивает или разделяет нити ДНК, так и нуклеазу, которая вносит однонитевые разрывы в ДНК. Хеликазы семейства RecD представляют собой DEAD-подобные хеликазы, принадлежащие к суперсемейству (SF)1 - разнообразному семейству белков, участвующих в АТР-зависимом раскручивании РНК или ДНК. Подобно хеликазам SF2, хеликазы SF1 не образуют тороидальных структур, как хеликазы SF3-6. Их сердцевинный домен хеликазы состоит из двух сходных белковых доменов, напоминающих укладку рекомбинационного белка RecA. Эта модель описывает С-концевой домен, также называемый HelicC. [PMID: 3526335, PMID: 7603978, PMID: 1552844, PMID: 2563148, PMID: 10322435, PMID: 11545728, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"

137	cd17933	Домен хеликазы DEXS-бокса белка RecD и сходных белков	"RecD является представителем комплекса RecBCD (EC 3.1.11.5, экзонуклеаза V). Он представляет собой альфа-цепь комплекса и функционирует как 3'-5'-хеликаза. Фермент RecBCD представляет собой как хеликазу, которая раскручивает или разделяет нити ДНК, так и нуклеазу, которая вносит однонитевые разрывы в ДНК. RecD является представителем суперсемейства DEAD-подобных хеликаз - разнообразного семейства белков, участвующих в АТР-зависимом раскручивании РНК или ДНК. Этот домен содержит АТР-связывающую область. [PMID: 3526335, PMID: 7603978, PMID: 1552844, PMID: 2563148, PMID: 10322435, PMID: 11545728, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"
138	PF08719	Домен с неизвестной функцией (DUF1768)	Это домен с неизвестной функцией. Он имеет структуру альфа-спирали. Аннотация GO для этого белка позволяет предположить, что он участвует в развитии личинок нематод и положительно регулирует скорость роста.
139	IPR012816	NADAR	"Это семейство содержит белок роящейся подвижности YbiA E. coli. Мутации в YbiA вызывают дефекты роения Escherichia coli, но не обязательно подвижности. Было предсказано, что это семейство участвует в путях использования NAD, вероятно, воздействуя на производные ADP-рибозы, и оно было названо NADAR (NAD и ADP-рибоза). Недавно было показано, что YbiA участвует в утилизации промежуточных соединений рибофлавина. Он катализирует гидролиз N-гликозидной связи в первых двух промежуточных соединениях в биосинтезе рибофлавина, которые

			представляют собой высокореактивные метаболиты, с образованием относительно безвредных продуктов."
140	IPR037238	Суперсемейство YbiA-подобных белков	"Это суперсемейство содержит белок роящейся подвижности YbiA E. coli. Мутации в YbiA вызывают дефекты роения Escherichia coli, но не обязательно подвижности. Было предсказано, что это семейство участвует в путях использования NAD, вероятно, воздействуя на производные ADP-рибозы, и оно было названо NADAR (NAD и ADP-рибоза). Недавно было показано, что YbiA участвует в утилизации промежуточных соединений рибофлавина. Он катализирует гидролиз N-гликозидной связи в первых двух промежуточных соединениях в биосинтезе рибофлавина, которые представляют собой высокореактивные метаболиты, с образованием относительно безвредных продуктов."
141	cd15457	Белок роящейся подвижности YbiA Escherichia coli и родственные белки	"Это семейство неохарактеризованных доменов изначально было классифицировано как домен с неизвестной функцией (DUF1768). Оно содержит таких представителей, как белок роящейся подвижности YbiA E. coli. Мутации в YbiA вызывают дефекты роения Escherichia coli, но не обязательно подвижности. Относительно недавно было предсказано, что это семейство участвует в путях использования NAD, вероятно, воздействуя на производные ADP-рибозы, и оно было названо суперсемейством NADAR (NAD и ADP-рибоза). [PMID: 22399070, PMID: 17122336]"
142	PF16588	"Цинковый палец" типа C2H2	NA

143	IPR036236	Суперсемейство доменов "цинковые пальцы" типа C2H2	"Цинковые пальцы" (Znf) типа C2H2 (классические) были первым классом, который был охарактеризован. Они содержат короткую бета-шпильку и альфа-спираль (структура бета/бета/альфа), где один атом цинка удерживается на месте остатками Cys(2)His(2) (C2H2) в тетраэдрическом массиве. Znf C2H2 можно разделить на три группы в зависимости от количества и паттерна "пальцев": тройной C2H2 (связывает один лиганд), несколько смежных C2H2 (связывает несколько лигандов) и разделенные пары C2H2. Znf C2H2 являются наиболее распространенными ДНК-связывающими мотивами, обнаруживаемыми в эукариотических факторах транскрипции, и они также были идентифицированы у прокариот. Факторы транскрипции обычно содержат несколько Znf (каждый с консервативной структурой бета/бета/альфа), способных образовывать множество контактов вдоль ДНК, где мотивы Znf C2H2 распознают последовательности ДНК, связываясь с большой бороздкой ДНК посредством короткой альфа-спирали в Znf, причем Znf охватывает 3-4 основания ДНК. Znf C2H2 также могут связываться с РНК и белковыми мишенями. Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы

			как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью."
144	cd17335	Белок 6, содержащий домен суперсемейства мембранных транспортеров	Белок 6 человека, содержащий домен суперсемейства мембранных транспортеров (MFSD6), также называют гомологом рецептора 2 MHC класса I макрофагов (MMR2). Предполагается, что он является возможным рецептором лейкоцитарного антигена

			человека (HLA) B62. MFSD6 сохранил консервативность в ходе эволюции и появился раньше билатеральных животных. Он принадлежит к суперсемейству мембранных транспортеров (MFS) белков мембранного транспорта, которые, как считается, функционируют с использованием одного сайта связывания субстрата посредством механизма чередующегося доступа, включающего перемещение по типу кулисного переключателя. [PMID: 20123006, PMID: 21044875, PMID: 20648054, PMID: 17010536, PMID: 26758938, PMID: 26234418, PMID: 26098515, PMID: 23530251, PMID: 23403214, PMID: 18537473, PMID: 16842212]"
145	IPR000571	"Цинковый палец" типа CCCH"	Домены "цинковые пальцы" (Znf) представляют собой относительно небольшие белковые мотивы, которые содержат несколько пальцеобразных выступов, образующих тандемные контакты со своей молекулой-мишенью. Некоторые из этих доменов связывают цинк, но многие не делают этого и вместо этого связывают другие металлы, такие как железо, или вообще не связывают какой-либо металл. Например, некоторые представители семейства образуют солевые мостики, чтобы стабилизировать пальцеобразные складки. Сначала они были идентифицированы как ДНК-связывающий мотив в факторе транскрипции TFIIIA из Xenopus laevis (африканской шпорцевой лягушки), однако теперь установлено, что они связывают ДНК, РНК, белковые и/или липидные субстраты. Их свойства связывания зависят от аминокислотной последовательности доменов "пальцев" и линкера между "пальцами", а также от структур более высокого порядка и количества "пальцев". Домены Znf часто встречаются в кластерах, где "пальцы" могут иметь различную

			специфичность связывания. Существует много суперсемейств мотивов Znf, различающихся как последовательностью, так и структурой. Они проявляют значительную изменчивость в способах связывания, даже между представителями одного и того же класса (например, некоторые связывают ДНК, другие связывают белки), что позволяет предположить, что мотивы Znf представляют собой стабильные каркасы, у которых развились специализированные функции. Например, Znf-содержащие белки функционируют при транскрипции генов, трансляции, транспорте мРНК, организации цитоскелета, развитии эпителия, клеточной адгезии, сворачивании белка, ремоделировании хроматина и восприятии цинка, и это лишь некоторые из этих функций. Цинк-связывающие мотивы представляют собой стабильные структуры, и они редко претерпевают конформационные изменения при связывании со своей мишенью. Эта запись представляет домены "цинковые пальцы" (Znf) типа C-x8-C-x5-C-x3-H (CCCH)". Белки, содержащие домены Znf типа CCCH, включают белки Znf из эукариот, участвующие в регуляции, связанной с клеточным циклом или фазой роста, например, TIS11B человека (фактор 1 ответа на бутират) - вероятный регуляторный белок, участвующий в регуляции ответа на факторы роста, и ядерный белок TTP мыши, индуцируемый факторами роста, выполняющий ту же функцию. Белок TTP мыши индуцируется факторами роста. Другой белок, содержащий этот домен, представляет собой субъединицу фактора сплайсинга человека U2AF размером 35 кДа, которая играет критическую роль как в конститутивном, так и в зависимом от энхансеров сплайсинге, опосредуя важные

			белок-белковые взаимодействия и взаимодействия белок-РНК, необходимые для выбора 3'-сайта сплайсинга. Было показано, что разные белки Znf типа CCCH взаимодействуют с 3'-нетранслируемой областью различных мРНК. Этот тип Znf очень часто присутствует в двух копиях."
146	PS50103	Профиль "цинкового пальца" типа C3H1	"Высококонсервативный порядок расположения цистеин-гистидин Cx(8)-Cx(5)-Cx(3)-H "цинкового пальца" типа C3H1 присутствует во многих эукариотических белках от дрожжей до млекопитающих. Было показано, что в различных белках "цинковый палец" типа C3H1 взаимодействует с 3'-нетранслируемой областью мРНК (см. ниже). Он очень часто присутствует в двух копиях [PMID:15963892,PMID:15718139]. Структура двух "цинковых пальцев" типа C3H1 белка TIS11d, связанного с его мРНК-мишенью, была расшифрована [PMID:14981510]. Два "цинковых пальца" сворачиваются независимо друг от друга, образуя небольшие компактные домены, соединенные линкерной последовательностью. Полипептидный остов принимает малорегулярную вторичную структуру. В каждом "цинковом пальце" существует короткая альфа-спираль сразу после первого цистеинового лиганда и поворот спирали 3/10 между вторым и третьим цистеиновыми лигандами. Все аминокислотные остатки, демонстрирующие межмолекулярные контакты с РНК, располагаются на одной поверхности каждого "цинкового пальца". Некоторые белки, которые, как известно, содержат "цинковый палец" типа C3H1, перечислены ниже. - Белок тристетрапролин (TTP) млекопитающих, прототип белков с "цинковыми пальцами" типа C3H1, ингибируют продуцирование TNF-альфа

			макрофагами путем дестабилизации их матричной РНК [PMID:15718139]. TTP представляет собой специфический РНК-связывающий белок; область, охватывающая два "цинковых пальца" типа C3H1, распознает нонамерные последовательности UUAUUUAUU, присутствующие в 3'-нетранслируемой области (UTR) TNF-альфа [PMID:10330172,PMID:17030620]. - Эукариотическое семейство TIS/CTH. Регуляторный белок, участвующий в регуляции ответа на факторы роста подобно TTP. Тандемный домен "цинковый палец" белка TIS11d связывается с AU-богатым элементом класса II (ARE) в 3'-нетранслируемой области (3'UTR) мРНК-мишеней и способствует их деаденилированию и разрушению [PMID:14981510]. - Эукариотические белки U2AG, U2R1, U2R2, малые ядерные рибонуклеопротеины (1 или 2 копии). - Белок PIE1 Caenorhabditis elegans. Он необходим для определения линии дифференцировки эмбриональных зародышевых клеток и для отсутствия транскрипции мРНК в этих клетках (2 копии). - Белок POS-1 Caenorhabditis elegans. Цитоплазматический белок, сходный с TIS11, необходимый для определения зародышевой линии у C. elegans (2 копии). - Белок субъединицы 4 фактора специфичности расщепления и полиаденилирования эукариот (Cpsf4). Он играет ключевую роль в образовании 3'-конца пре-мРНК, распознавая сигнальную последовательность AAUAAA и взаимодействуя с поли(A)-полимеразой и другими факторами, вызывая расщепление и добавление поли(A)-хвоста. Cpsf4 связывает РНК-полимеры с предпочтением в отношении поли(U)-хвоста (5 копий). - Белок ZFS1 Schizosaccharomyces pombe. Он был выделен как супрессор стерильности,

			вызванной сверхэкспрессией двухнитевой РНКазы (2 копии). - Белок Z183 человека, содержит один "палец" типа RING и один "цинковый палец" типа C3H1. - Нуклеопорин-подобный белок 2 млекопитающих, необходимый для экспорта мРНК, содержащих поли(A)-хвосты, из ядра в цитоплазму (1 копия). Разработанный авторами настоящего изобретения профиль охватывает весь "цинковый палец" типа C3H1."
147	PS50600	Профиль семейства убиквитин-подобных протеаз	"Деубиквитинирующие ферменты (DUB) образуют большое семейство цистеиновых протеаз, которые могут деконъюгировать убиквитин или убиквитин-подобные белки из конъюгированных с убиквитином белков. Их можно разделить на 3 семейства в соответствии с гомологией последовательностей [PMID:10603300,PMID:8982460]: убиквитин-карбоксиконцевые гидролазы (UCH), убиквитин-специфические процессинговые протеазы (UBP) и убиквитин-подобные протеазы (ULP) (EC 3.4.22.68), специфичные для деконъюгации убиквитин-подобных белков. В отличие от пути UBP, который является крайне избыточным (16 ферментов UBP в дрожжах), убиквитин-подобных протеаз мало (только одна в дрожжах, ULP1). ULP1 катализирует две критические функции в пути SUMO/Smt3 благодаря своей активности цистеиновой протеазы. ULP1 подвергает процессингу C-концевую последовательность Smt3 (-GGATY) с превращением ее в зрелую форму (-GG) и деконъюгирует Smt3 от эпсилон-аминогруппы лизина белка-мишени [PMID:10094048]. Кристаллическая структура дрожжевого ULP1, связанного с Smt3 [PMID:10882122], показала, что каталитическая и взаимодействующая области

			контакта расположены в неглубокой и узкой расщелине, где консервативные остатки распознают мотив Gly-Gly на С-конце белка Smt3. ULP1 принимает новую архитектуру, несмотря на некоторое структурное сходство с другими цистеиновыми протеазами. Вторичная структура состоит из семи альфа-спиралей и семи бета-нитей. Каталитический домен содержит центральную альфа-спираль, бета-нити 4-6 и каталитическую триаду (Cys-His-Asp). Авторы настоящего изобретения разработали профиль, направленный на С-концевую часть белков ULP, который проявляет полную протеолитическую активность [PMID:10882122]."
148	PF02902	"Семейство протеаз Ulp1, С-концевой каталитический домен"	Этот домен содержит каталитическую триаду Cys-His-Asn.
149	IPR003653	"Семейство протеаз Ulp1, С-концевой каталитический домен"	"Эта запись представляет С-концевую часть убиквитин-подобных протеаз, которая проявляет полную протеолитическую активность. Деубиквитинирующие ферменты (DUB) образуют большое семейство цистеиновых протеаз, которые могут деконъюгировать убиквитин или убиквитин-подобные белки из конъюгированных с убиквитином белков. Их можно разделить на 3 семейства в соответствии с гомологией последовательностей: убиквитин-карбоксиконцевые гидролазы (UCH), убиквитин-специфические процессинговые протеазы (UBP) и убиквитин-подобные протеазы (ULP), специфичные для деконъюгации убиквитин-подобных белков. В отличие от пути UBP, который является крайне избыточным (16 ферментов UBP в дрожжах), убиквитин-подобных протеаз мало (только одна в дрожжах, ULP1). ULP1

			катализирует две критические функции в пути SUMO/Smt3 благодаря своей активности цистеиновой протеазы. ULP1 подвергает процессингу С-концевую последовательность Smt3 (-GGATY) с превращением ее в зрелую форму (-GG) и деконъюгирует Smt3 от эпсилон-аминогруппы лизина белка-мишени. Кристаллическая структура дрожжевого ULP1, связанного с Smt3, показала, что каталитическая и взаимодействующая области контакта расположены в неглубокой и узкой расщелине, где консервативные остатки распознают мотив Gly-Gly на С-конце белка Smt3. ULP1 принимает новую архитектуру, несмотря на некоторое структурное сходство с другими цистеиновыми протеазами. Вторичная структура состоит из семи альфа-спиралей и семи бета-нитей. Каталитический домен содержит центральную альфа-спираль, бета-нити 4-6 и каталитическую триаду (Cys-His-Asp). Цистеиновые пептидазы с химотрипсиноподобной укладкой включены в клан PA, который также включает сериновые пептидазы. Цистеиновые пептидазы, представляющие собой N-концевые нуклеофильные гидролазы, входят в клан PB. Цистеиновые пептидазы с третичной структурой, сходной с таковой у аспартилдипептидазы серинового типа, входят в клан PC. Цистеиновые пептидазы с интеиноподобной укладкой входят в клан PD, в который также входят аспарагинлиазы. Цистеиновая пептидаза представляет собой протеолитический фермент, гидролизующий пептидную связь, используя в качестве нуклеофила тиольную группу цистеинового остатка. Гидролиз обычно предусматривает каталитическую триаду, состоящую из тиольной группы цистеина, имидазолиевого

			кольца гистидина и третьего остатка, обычно аспарагина или аспарагиновой кислоты, для обеспечения ориентации и активации имидазолиевого кольца. Только в одном семействе цистеиновых пептидаз роль общего основания отводится остатку, отличному от гистидина: в пептидазах семейства C89 (кислые церамидазы) общее основание представляет собой аргинин. Цистеиновые пептидазы можно сгруппировать в четырнадцать различных кланов, при этом представители каждого клана характеризуются третичной укладкой, уникальной для клана. Четыре клана цистеиновых пептидаз имеют структурное сходство с сериновыми и треониновыми пептидазами и аспарагинлиазами. По сходству последовательностей цистеиновые пептидазы можно объединить в более чем 80 различных семейств. Кланы CF, CM, CN, CO, CP и PD содержат только одно семейство. Цистеиновые пептидазы часто активны при кислых pH и поэтому ограничены кислыми средами, такими как лизосомы животных или растительные вакуоли. Цистеиновые пептидазы могут представлять собой эндопептидазы, аминопептидазы, карбоксипептидазы, дипептидилпептидазы или омега-пептидазы. Они ингибируются тиол-реактивными хелаторами, такими как йодацетат, йодуксусная кислота, N-этилмалеимид или п-хлормеркурибензоат. Клан CA включает белки с папаиноподобной укладкой. Существует каталитическая триада, которая представлена в следующем порядке: Cys/His/Asn (или Asp). Четвертый остаток, обычно Gln, важен для стабилизации ацильного промежуточного соединения, образующегося в ходе катализа, и он предшествует Cys в активном центре. Укладка состоит из двух субдоменов, между

			которыми находится активный центр. Один субдомен состоит из пучка спиралей с каталитическим Cys на конце одной из них, а другой субдомен представляет собой бета-"бочку" с His и Asn (или Asp) в активном центре. К этому клану принадлежат более тридцати семейств, и для представителей большинства из них были расшифрованы третичные структуры. Пептидазы клана CA обычно чувствительны к низкомолекулярному ингибитору E64, который неэффективен в отношении пептидаз из других кланов цистеиновых пептидаз. К клану CD относятся белки с каспазоподобной укладкой. Белки в клане имеют структуру альфа/бета/альфа-"сэндвича". Существует каталитическая диада, которая представлена в порядке His/Cys. His активного центра встречается в мотиве His-Gly, а Cys активного центра встречается в мотиве Ala-Cys; обоим мотивам предшествует блок гидрофобных остатков. Специфичность в основном направлена на остатки, которые занимают карман связывания S1, так что каспазы расщепляют аспартильные связи, легумаины расщепляют аспарагинильные связи, а гингипаины расщепляют лизильные или аргинильные связи. К клану CE относятся белки с аденаиноподобной укладкой. Укладка состоит из двух субдоменов, между которыми находится активный центр. Один домен представляет собой пучок спиралей, а другой представляет собой бета-"бочку". Субдомены расположены в порядке, противоположном обнаруживаемому в пептидазах клана CA, и это отражено в порядке остатков активного центра: His/Asn/Gln/Cys. Это вызвало предположение, что белки в кланах CA и CE родственны, и что представители одного клана произошли в результате

			циклической пермутации структуры представителей другого. К клану CL относятся белки с укладкой, подобной сортазе B. Пептидазы в клане гидролизуют и переносят пептиды клеточной стенки бактерий. На укладке показана замкнутая бета-"бочка", вынизанная спиралями, с активным центром на одном конце "бочки". Активный центр состоит из каталитической диады His/Cys."
150	IPR001498	"Белок Impact, N-концевой домен"	"Белок Impact представляет собой регулятор трансляции, который обеспечивает постоянный высокий уровень трансляции при дефиците аминокислот. Он действует путем взаимодействия с Gcn1/Gcn1L1, тем самым предотвращая активацию протеинкиназ Gcn2 (EIF2AK1-4) и последующее подавление синтеза белка. Он является эволюционно консервативным от эукариот до архей. Эта запись представляет N-концевой домен белков Impact."
151	IPR023582	Семейство белков Impact	"Белок Impact представляет собой регулятор трансляции, который обеспечивает постоянный высокий уровень трансляции при дефиците аминокислот. Он действует путем взаимодействия с Gcn1/Gcn1L1, тем самым предотвращая активацию протеинкиназ Gcn2 (EIF2AK1-4) и последующее подавление синтеза белка. Он является эволюционно консервативным от эукариот до архей."
152	IPR020568	Укладка типа домена 2 рибосомного белка S5	"Домен 2 рибосомного белка S5 имеет левостороннюю 2-слойную укладку альфа/бета с сердцевинной структурой, состоящей из бета(3)-альфа-бета-альфа. Домены с такой укладкой обнаружены во многих РНК/ДНК-связывающих белках, а также в киназах из семейства киназ GHMP. Белки, содержащие этот домен с укладкой альфа/бета, включают следующие. * Компоненты трансляционного

			аппарата (рибосомные белки S5 и S9 и домен IV факторов элонгации EF-G и eEF-2). * Белок рибонуклеазу P (РНКазу P). * Рибонуклеазу PH (домен 1), а также различные экзонуклеазы комплексов экзосом (RRP41, RRP42, RRP43, RRP45, RRP46, MTR3, ECX1, ECX2). * Белки модификации ДНК (белки репарации ошибочно спаренных оснований ДНК MutL и PMS2, ДНК-гиразу B, ДНК-топоизомеразу II, IV-B и VI-B). * Киназы GHMP, которые переносят фосфорильную группу от ATP к акцептору (галактокиназу (), гомосеринкиназу () и мевалонаткиназу ()). * Ранний белок-переключатель Xol-1 Caenorhabditis elegans (дивергентный представитель семейства киназ GHMP, утративший сайт связывания ATP). * Шаперон Hsp90 (средний домен), родственный семейству ДНК-гираз/MutL; этот домен содержит дополнительный C-концевой субдомен альфа/бета. * Имидазолглицеролфосфатдегидратазу, которая содержит дупликацию, состоящую из двух структурных повторов этой укладки. * Каталитический домен ATP-зависимой протеазы Lon (La), содержащий дополнительный C-концевой субдомен альфа/бета. * Фермент FAE, активирующий формальдегид, который содержит модификацию этой укладки, состоящую из дополнительной единицы альфа/бета после нити 2."
153	PF01205	Неохарактеризованное семейство белков UPF0029	NA
154	IPR036956	"Суперсемейство N-концевых доменов белков Impact"	"Белок Impact представляет собой регулятор трансляции, который обеспечивает постоянный высокий уровень трансляции при дефиците аминокислот. Он действует путем взаимодействия с Gcn1/Gcn1L1, тем самым предотвращая активацию

			протеинкиназ Gcn2 (EIF2AK1-4) и последующее подавление синтеза белка. Он является эволюционно консервативным от эукариот до архей. Эта запись представляет суперсемейство N-концевых доменов белков Impact."
155	IPR033134	"Рацемаза Asp/Glu, активный центр 2"	"Аспартатрацемаза () и глутаматрацемаза () представляют собой два эволюционно родственных бактериальных фермента, которые, по-видимому, не нуждаются в кофакторе для обеспечения своей активности. Глутаматрацемаза, которая обеспечивает взаимопревращение L-глутамата в D-глутамат, необходима для биосинтеза пептидогликана и некоторых антибиотиков на основе пептидов, таких как грамицидин S. В дополнение к охарактеризованным аспартат- и глутаматрацемазам эта запись также включает один гипотетический белок из Erwinia carotovora и один из Escherichia coli (ygeA). Два консервативных остатка цистеина присутствуют в последовательности этих ферментов. Ожидается, что они будут играть роль в каталитической активности, выступая в качестве оснований при отщеплении протона от субстрата. Эта запись представляет консервативную область, содержащую второй остаток цистеина."
156	PS00924	Предполагаемая нуклеаза HARBI1	NA
157	cd05931	Лигаза жирного ацил-AMP (FAAL)	"FAAL принадлежит к семейству ферментов, образующих аденилат, класса I и гомологична лигазам жирного ацилкофермента A (CoA) (FACL). Однако FAAL продуцируют только ациладенилат и неспособны осуществлять реакцию образования сложного тиоэфира, тогда как FACL осуществляют двухстадийную

			каталитическую реакцию: лигирование AMP с последующим лигированием CoA с использованием ATP и CoA в качестве кофакторов. FAAL имеют вставочные мотивы между N-концевым и C-концевым субдоменами, которые отличают их от FACL. Этот вставочный мотив препятствует связыванию CoA, тем самым предотвращая лигирование CoA. Было высказано предположение, что ациладенилаты служат субстратами для многофункциональных поликетидсинтаз для обеспечения синтеза сложных липидов, таких как фтиоцеролдимикоцерозат, сульфолипиды, миколовые кислоты и микобактин. [PMID: 21185305, PMID: 21601529, PMID: 12406578, PMID: 15112993, PMID: 12598647, PMID: 9373621, PMID: 17762044, PMID: 11255012]"
158	IPR040097	Лигаза жирного ацил-AMP	"Лигаза жирного ацил-AMP (FAAL) принадлежит к семейству ферментов, образующих аденилат, класса I и гомологична лигазам жирного ацилкофермента A (CoA) (FACL). Однако FAAL продуцируют только ациладенилат и неспособны осуществлять реакцию образования сложного тиоэфира, тогда как FACL осуществляют двухстадийную каталитическую реакцию: лигирование AMP с последующим лигированием CoA с использованием ATP и CoA в качестве кофакторов. FAAL имеют вставочные мотивы между N-концевым и C-концевым субдоменами, которые отличают их от FACL. Этот вставочный мотив препятствует связыванию CoA, тем самым предотвращая лигирование CoA. Было высказано предположение, что ациладенилаты служат субстратами для многофункциональных поликетидсинтаз для обеспечения синтеза сложных липидов, таких как фтиоцеролдимикоцерозат, сульфолипиды, миколовые кислоты и микобактин."

159	IPR002219	"Протеинкиназа С-подобный домен, связывающий сложные эфиры форбола/диацилглицерин"	"Диацилглицерин (DAG) представляет собой важный вторичный мессенджер. Сложные эфиры форбола (PE) представляют собой аналоги DAG и мощные стимуляторы опухолей, вызывающие разнообразные физиологические изменения при введении как в клетки, так и в ткани. DAG активирует семейство сериновых/треониновых протеинкиназ, известных под общим названием протеинкиназа С (PKC). Сложные эфиры форбола могут напрямую стимулировать PKC. Было показано, что N-концевая область PKC, известная как C1, связывает PE и DAG зависимым от фосфолипидов и цинка образом. Область C1 содержит одну или две копии (в зависимости от изофермента PKC) богатого цистеином домена длиной приблизительно 50 аминокислотных остатков, который необходим для связывания DAG/PE. Домен, связывающий DAG/PE, связывает два иона цинка; лигандами этих ионов металлов, вероятно, являются шесть остатков цистеина и два остатка гистидина, которые являются консервативными в этом домене."
160	PS50081	Профиль домена типа "цинкового пальца" для сложных эфиров форбола/DAG	"Диацилглицерин (DAG) представляет собой важный вторичный мессенджер. Сложные эфиры форбола (PE) представляют собой аналоги DAG и мощные стимуляторы опухолей, вызывающие разнообразные физиологические изменения при введении как в клетки, так и в ткани. DAG активирует семейство сериновых/треониновых протеинкиназ, известных под общим названием протеинкиназа С (PKC) [PMID:1396661]. Сложные эфиры форбола могут напрямую стимулировать PKC. Было показано, что N-концевая область PKC, известная как C1 [PMID:2500657], связывает PE и DAG зависимым от фосфолипидов и цинка образом.

			Область C1 содержит одну или две копии (в зависимости от изофермента PKC) богатого цистеином домена длиной приблизительно 50 аминокислотных остатков, необходимого для связывания DAG/PE. Такой домен также был обнаружен в следующих белках. - Диацилглицеролкиназа (EC 2.7.1.107) (DGK) [PMID:2156169] - фермент, который превращает DAG в фосфатидат. Он содержит две копии домена, связывающего DAG/PE, в своем N-концевом участке. У млекопитающих известно по меньшей мере пять различных форм DGK. - N-химерин. Специфичный для головного мозга белок, который демонстрирует сходство последовательностей с белком BCR в его C-концевой части и содержит одну копию домена, связывающего DAG/PE, в его N-концевой части. Было показано [PMID:2268301,PMID:1660266], что он способен связывать сложные эфиры форбола. - Семейство сериновых/треониновых протеинкиназ raf/mil. Эти протеинкиназы содержат одну N-концевую копию домена, связывающего DAG/PE. - Белок unc-13 из Caenorhabditis elegans. Его функция неизвестна, но он содержит копию домена, связывающего DAG/PE, в своем центральном участке и, как было показано, специфично связывается со сложным эфиром форбола в присутствии кальция [PMID:1445255]. - Онкоген vav. Vav был получен в результате генетической перестройки во время анализов переноса генов. Его экспрессия, по-видимому, ограничена клетками гемопоэтического происхождения. Vav, по-видимому [PMID:1660266,PMID:1614545], содержит домен, связывающий DAG/PE, в центральной части белка. - Белок rotund дрозофилы, активирующий GTPазу. Домен, связывающий DAG/PE, связывает два иона цинка;

			лигандами этих ионов металлов, вероятно, являются шесть остатков цистеина и два остатка гистидина, которые являются консервативными в этом домене. Авторы настоящего изобретения разработали как паттерн сигнатуры, так и профиль, которые полностью охватывают домен DAG/PE."
161	PS50045	Профиль домена, взаимодействующего с сигма-фактором 54	"Некоторые бактериальные регуляторные белки активируют экспрессию генов из промоторов, распознаваемых ядерной РНК-полимеразой, связанной с альтернативным сигма-фактором 54. Они имеют консервативный домен из приблизительно 230 остатков, участвующий в АТР-зависимом [PMID:2041769] взаимодействии с сигма-фактором 54. Этот домен был обнаружен в белках, перечисленных ниже. - acoR из Alcaligenes eutrophus - активатор оперона катаболизма ацетоина acoXABC. - algB из Pseudomonas aeruginosa - активатор гена биосинтеза альгината algD. - dctD из Rhizobium - активатор dctA, белка транспорта С4-дикарбоксилата. - dhaR из Citrobacter freundii - регулятор оперона dha для утилизации глицерина. - fhlA из Escherichia coli - активатор структурных генов формиатдегидрогеназы H и гидрогеназы III. - flbD из Caulobacter crescentus - активатор генов жгутиков. - hoxA из Alcaligenes eutrophus - активатор оперона гидрогеназы. - hrpS из Pseudomonas syringae - активатор hprD, а также других локусов hrp, вовлеченных в патогенность для растений. - hupR1 из Rhodobacter capsulatus - активатор генов [NiFe]-гидрогеназы hupSL. - hydG из Escherichia coli и Salmonella typhimurium - активатор гидрогеназной активности. - levR из Bacillus subtilis, который регулирует экспрессию оперона леваназы (levDEFG и sacC). - nifA (a

			также anfA и vnfA) из различных бактерий - активатор оперона nif, обеспечивающего азотфиксирующую способность. - ntrC из различных бактерий - активатор генов ассимиляции азота, таких как гены глутаминсинтетазы (glnA), или оперона nif. - pgtA из Salmonella typhimurium - активатор индуцируемой системы транспорта фосфоглицерата. - pilR из Pseudomonas aeruginosa - активатор транскрипции гена пилина. - rocR из Bacillus subtilis - активатор генов утилизации аргинина - tyrR из Escherichia coli, участвующий в транскрипционной регуляции биосинтеза и транспорта ароматических аминокислот. - wtsA из Erwinia stewartii - активатор гена патогенности для растений wtsB. - xylR из Pseudomonas putida - активатор Tol-плазмидного оперона катаболизма ксилолов xylCAB и xylS. - гипотетический белок yfhA Escherichia coli. - гипотетический белок yhgB Escherichia coli. Приблизительно половина этих белков (algB, dcdT, flbD, hoxA, hupR1, hydG, ntrC, pgtA и pilR) относится к двухкомпонентным системам передачи сигнала [PMID:2694934] и содержит домен, который может фосфорилироваться сенсорной протеинкиназой в их N-концевом участке. Почти все эти белки содержат ДНК-связывающий домен со структурой спираль-поворот-спираль в их С-концевом участке. Домен, взаимодействующий с сигма-фактором 54, обладает АТРазной активностью. Это может быть необходимо для содействия конформационному изменению, необходимому для взаимодействия [PMID:1534752]. Домен содержит атипичный АТР-связывающий мотив А (Р-петлю), а также форму мотива В. Два АТР-связывающих мотива расположены в N-концевом участке домена; авторы

			настоящего изобретения разработали паттерны сигнатуры для обоих мотивов. Другие области домена также являются консервативными. Авторы настоящего изобретения выбрали одну из них, расположенную в С-концевом участке, в качестве третьего паттерна сигнатуры."
162	IPR002078	Домен, взаимодействующий с сигма-фактором 54 РНК-полимеразы	"Некоторые бактериальные регуляторные белки активируют экспрессию генов из промоторов, распознаваемых ядерной РНК-полимеразой, связанной с альтернативным сигма-фактором 54. Они имеют консервативный домен из приблизительно 230 остатков, участвующий в АТР-зависимом взаимодействии с сигма-фактором 54. Приблизительно половина этих белков, в которых обнаруживается этот домен (algB, dcdT, flbD, hoxA, hupR1, hydG, ntrC, pgtA и pilR), относится к двухкомпонентным системам передачи сигнала и содержит домен, который может фосфорилироваться сенсорной протеинкиназой в их N-концевом участке. Почти все эти белки содержат ДНК-связывающий домен со структурой спираль-поворот-спираль в их С-концевом участке. Домен, взаимодействующий с сигма-фактором 54, обладает АТРазной активностью. Это может быть необходимо для содействия конформационному изменению, необходимому для взаимодействия. Домен содержит атипичный АТР-связывающий мотив А (Р-петлю), а также форму мотива В. Два АТР-связывающих мотива расположены в N-концевом участке домена."
163	IPR000851	Рибосомный белок S5	"Рибосомы представляют собой частицы, которые катализируют мРНК-направленный синтез белка во всех организмах. Кодоны мРНК являются доступными

			на рибосоме для обеспечения связывания тРНК. Это приводит к включению аминокислот в растущую полипептидную цепь в соответствии с генетической информацией. Входящие аминокислотные мономеры попадают в А-сайт рибосомы в виде аминоацил-тРНК в комплексе с фактором элонгации Tu (EF-Tu) и GTP. Растущая полипептидная цепь, расположенная в Р-сайте в виде пептидил-тРНК, затем переносится на аминоацил-тРНК, а новая пептидил-тРНК, удлиненная на один остаток, транслоцируется в сайт Р с помощью фактора элонгации G (EF-G) и GTP, тогда как деацилированная тРНК высвобождается из рибосомы через один или несколько сайтов выхода. Приблизительно 2/3 массы рибосомы состоит из РНК, и 1/3 - из белка. Белки называются в соответствии с субъединицей рибосомы, к которой они принадлежат, - малой (S1-S31) и большой (L1-L44). Обычно они вынизывают рРНК-сердцевины субъединиц. Многие рибосомные белки, особенно белки большой субъединицы, состоят из глобулярного доступного на поверхности домена с длинными пальцеобразными отростками, которые простираются в рРНК-сердцевину для стабилизации ее структуры. Большинство белков взаимодействуют с несколькими элементами РНК, часто из разных доменов. В большой субъединице приблизительно 1/3 нуклеотидов 23S рРНК находятся по меньшей мере в обеспечиваемом ван-дер-ваальсовыми силами контакте с белком, а L22 взаимодействует со всеми шестью доменами 23S рРНК. Белки S4 и S7, которые инициируют сборку 16S рРНК, расположены в местах соединения пяти и четырех спиралей РНК соответственно. Таким образом, белки служат для организации и

			стабилизации третичной структуры рРНК. При том, что важнейшие функции декодирования и переноса пептидов зависят от РНК, белки играют активную роль в функциях, которые, возможно, появились в ходе эволюции для усовершенствования процесса синтеза белка. В дополнение к их функции в рибосоме многие рибосомные белки выполняют некоторую функцию "вне" рибосомы. Рибосомный белок S5 является одним из белков малой субъединицы рибосомы и представляет собой белок, содержащий 166-254 аминокислотных остатков. Известно, что у Escherichia coli S5 играет важную роль в сборке и функционировании 30S субъединицы рибосомы. Было показано, что мутации в S5 увеличивают частоту ошибок трансляции. Он принадлежит к семейству рибосомных белков, которые на основании сходства последовательностей относят к одной группе S5 хлоропластов бактерий, цианелл, красных водорослей, митохондрий архей и грибов; S2 млекопитающих, Caenorhabditis elegans, дрозофилы и растений; а также S4 дрожжей (SUP44)."
164	PS51526	Профиль ДНК-связывающего домена с "крылатой спиралью" типа RFX	"Семейство факторов транскрипции RFX характеризуется уникальным ДНК-связывающим доменом, состоящим из ~ 75 остатков." Гены RFX были выделены у дрожжей, нематод и позвоночных. Характерный ДНК-связывающий домен типа RFX со структурой HTH был привлечен в очень разные в иных отношениях регуляторные факторы, функционирующие в разнообразном спектре неродственных систем, включая регуляцию митотического клеточного цикла у делящихся дрожжей, контроль иммунного ответа у млекопитающих и инфицирование вирусом гепатита B

			человека [PMID:8600444,PMID:10767550]. ДНК-связывающий домен типа RFX является необычным представителем подсемейства доменов со структурой спираль-поворот-спираль с "крылатой спиралью", поскольку он использует бета-шпильку (или "крыло") для распознавания ДНК вместо распознающей спирали, типичной для доменов со структурой спираль-поворот-спираль. Он состоит из трех альфа-спиралей (H), трех бета-нитей (S) и трех соединяющих петель (L), расположенных в порядке H1-S1-H2-L1-H3-L2-S2-W1-S3. Третья петля, соединяющая бета-нити S2 и S3, образует "крыло" W1 мотива типа "крылатой спирали". В отличие от ранее описанных DBD с "крылатой спиралью" (обычно длиной 110 остатков с двумя "крыльями", W1 и W2), более короткий ДНК-связывающий домен с "крылатой спиралью" типа RFX имеет только одно "крыло" [PMID:10706293]. Некоторые белки, которые, как известно, содержат ДНК-связывающий домен с "крылатой спиралью" типа RFX, перечислены ниже. - RFX1 человека - клеточный трансактиватор, используемый высокопатогенным вирусом гепатита B - RFX2 человека. - RFX3 человека - фактор транскрипции, необходимый для цилиогенеза и дифференцировки островковых клеток во время развития эндокринной части поджелудочной железы. - RFX5 человека - регулятор транскрипции генов MHC класса II. - ARID2 человека - компонент комплекса ремоделирования хроматина SWI/SNF-B (PBAF) [PMID:15640446]. - Crt1 дрожжей - эффектор контрольной точки повреждения и репликации ДНК. - sak1 Schizosaccharomyces pombe - важный регуляторный белок в жизненном цикле. Он позволяет клеткам выйти из

			митотического цикла и войти либо в стационарную фазу, либо в путь, ведущий к половой дифференцировке. Разработанный авторами настоящего изобретения профиль охватывает весь ДНК-связывающий домен с "крылатой спиралью" типа RFX."
165	IPR001487	Бромодомен	"Бромодомены обнаруживаются в различных ДНК-связывающих белках млекопитающих, беспозвоночных и дрожжей. Бромодомены могут взаимодействовать с ацетилированным лизином. В некоторых белках классический бромодомен подвергся дивергенции до такой степени, что части области либо отсутствуют, либо содержат вставку (например, белок HRX млекопитающих, гипотетический белок ZK783.4 Caenorhabditis elegans, белок YTA7 дрожжей). Бромодомен может встречаться в одной копии или в двух копиях. Точная функция домена неясна, но он может участвовать в белок-белковых взаимодействиях и может играть роль в сборке или обеспечении активности многокомпонентных комплексов, участвующих в активации транскрипции."
166	cd08197	2-дезоксисциллоинозозосинтаза (DOIS) катализирует образование углеродного цикла от D-глюкозо-6-фосфата до 2-дезоксисциллоинозозы.	"2-дезоксисциллоинозозосинтаза (DOIS) катализирует образование углеродного цикла от D-глюкозо-6-фосфата до 2-дезоксисциллоинозозы посредством многостадийной реакции в биосинтезе аминогликозидных антибиотиков. Аминогликозидные антибиотики, содержащие 2-дезоксистрептамин (DOS), включают неомицин, канамицин, гентамицин и рибостамицин. Они являются важными антибактериальными средствами. DOIS является гомологом дегидрохинатсинтазы, которая катализирует циклизацию 3-дезокси-D-

			арабиногептулозонат-7-фосфата в дегидрохинат (DHQ) в шикиматном пути. [PMID: 17879343, PMID: 16258246, PMID: 15313224, PMID: 17035031, PMID: 16915821, PMID: 17368605, PMID: 15928851, PMID: 15788164, PMID: 16861807, PMID: 14757238, PMID: 14750780, PMID: 16506694]"
167	IPR022690	"Белок, ассоциированный с репликацией геминивируса AL1, каталитический домен"	"Геминивирусы характеризуются геномом в виде кольцевой однонитевой ДНК, инкапсулированной в сдвоенные (парные) квазиизометрические частицы, от которых эта группа вирусов получила свое название. Большинство геминивирусов можно разделить на две подгруппы в зависимости от круга хозяев и/или насекомых-переносчиков: т. е. на те, которые инфицируют двудольные растения и передаются одними и теми же видами белокрылки, и на те, которые инфицируют однодольные растения и передаются различными цикадками-переносчиками. Геномы передаваемых белокрылкой вируса мозаики африканского маниока, вируса золотой мозаики томата (TGMV) и вируса золотой мозаики фасоли (BGMV) имеют двухкомпонентный геном. В отличие от этого, для передаваемых цикадками вируса полосатости кукурузы (MSV) и вируса карликовости пшеницы (WDV) был идентифицирован только один ДНК-компонент. Вирус курчавости верхушки свеклы (BCTV) и вирус желтой карликовости табака относятся к третьей возможной подгруппе. Подобно MSV и WDV, BCTV передается определенными видами цикадок, но, как и передаваемые белокрылками геминивирусы, его круг хозяев ограничивается двудольными растениями. Сравнение последовательностей передаваемых белокрылками вируса курчавости листьев кабачка (SqLCV) и вируса

			желтой курчавости листьев томата (TYLCV) с геномными компонентами TGMV и BGMV показывает близкое эволюционное родство. Выравнивание аминокислотных последовательностей белков вируса желтой мозаики картофеля (PYMV) с таковыми последовательностями, кодируемыми другими геминивирусами, показывает, что PYMV является близкородственным геминивирусам, выделенным из Нового Света, особенно в областях гена предполагаемого белка оболочки. Сравнение белков, кодируемых ДНК MSV, с белками других геминивирусов, инфицирующих однодольные растения, включая вирус полосатости проса и вирус полосатости мискантуса (MiSV), выявило высокий уровень сходства. Белки AL1 кодируют белковый инициатор репликации (Rep) геминивирусов, который представляет собой репликон-специфический фермент-инициатор и является важным компонентом реплисомы. Для белка Rep геминивируса эта N-концевая область имеет решающее значение для распознавания точки начала репликации и расщепления ДНК, а также переноса нуклеотидов. Он встречается в ассоциации с"
168	PS50014	Профиль бромодомена	"Бромодомен [PMID:1346755,PMID:7580139] представляет собой консервативную область, состоящую из приблизительно 70 аминокислот, обнаруживаемую в следующих белках. - Фактор инициации транскрипции высших эукариот TFIID, субъединица 250 кДа (TBP-ассоциированный фактор p250) (ген CCG1). P250 ассоциирован с белком, связывающим TATA-бокс TFIID, и, по-видимому, необходим для прохождения фазы G1 клеточного цикла. - RING3 человека - белок с неизвестной функцией, закодированный в локусе MHC класса II. - CREB-связывающий белок (CBP)

			млекопитающих, который опосредует регуляцию генов с помощью cAMP путем специфичного связывания с фосфорилированным белком CREB. - Гомеозисный белок стерильности самок дрозофилы (ген fsh), необходимый по материнской линии для правильной экспрессии других гомеозисных генов, участвующих в формировании паттерна, таких как Ubx. - Белок brahma дрозофилы (ген brm) - белок, необходимый для активации нескольких гомеозисных генов. - Гомологи белка brahma млекопитающих. У человека известны три brahma-подобных белка: SNF2a(hBRM), SNF2b и BRG1. - BS69 человека - белок, который связывается с аденовирусом E1A и ингибирует трансактивацию E1A. - Перегрин человека (или Br140). - BDF1 дрожжей [PMID:7580139] - фактор транскрипции, участвующий в экспрессии широкого класса генов, включая snRNA. - GCN5 дрожжей - общий активатор транскрипции, действующий совместно с некоторыми другими ДНК-связывающими активаторами транскрипции, такими как GCN4, HAP2/3/4 или ADA2. - NPS1/STH1 дрожжей, участвующие в контроле фазы G(2) при митозе. - SNF2/SWI2 дрожжей, являющиеся частью комплекса с белками SNF5, SNF6, SWI3 и ADR6/SWI1. Этот комплекс SWI участвует в активации транскрипции. - SPT7 дрожжей - активатор транскрипции Ty-элементов и, возможно, других генов. - Белок cbp-1 Caenorhabditis elegans. - Гипотетический белок YGR056w дрожжей. - Гипотетический белок YKR008w дрожжей. - Гипотетический белок L9638.1 дрожжей. Некоторые белки содержат область, которая хотя и сходна в какой-то степени с классическим бромодоменом, но отличается от него либо отсутствием части домена, либо наличием вставки.

			Этими белками являются следующие. - Белок млекопитающих HRX (также известный как ALL-1 или MLL) - белок, участвующий в транслокациях, приводящих к формам острого лейкоза, который, возможно, выступает в качестве фактора регуляции транскрипции. HRX содержит область, сходную с С-концевой половиной бромодомена. - Гипотетический белок ZK783.4 Caenorhabditis elegans. Бромодомен этого белка содержит вставку из 23 аминокислот. - Белок YTA7 дрожжей. Этот белок содержит область со значительным сходством с С-концевой половиной бромодомена. Поскольку он является представителем семейства AAA, он также находится в другом окружении с функциональной точки зрения. Вышеуказанные белки обычно содержат один бромодомен, но некоторые из них содержат две копии, например, в случае с BDF1, CCG1, fsh, RING3, YKR008w и L9638.1. Точная функция этого домена еще не известна, но считается, что он участвует в белок-белковых взаимодействиях и может быть важен для сборки или обеспечения активности многокомпонентных комплексов, участвующих в активации транскрипции. Консенсусный паттерн, который разработали авторы настоящего изобретения, охватывает большую часть бромодомена; более чувствительное выявление доступно благодаря использованию профиля, охватывающего весь домен."
169	PF00799	Каталитический домен Rep геминивируса	"Белки AL1 кодируют белковый инициатор репликации (Rep) геминивирусов, который представляет собой репликон-специфический фермент-инициатор и является важным компонентом реплисомы [PMID:12130667]. Для белка Rep

			геминивируса эта N-концевая область имеет решающее значение для распознавания точки начала репликации и расщепления ДНК, а также переноса нуклеотидов [PMID:12130667]."
170	IPR008162	Неорганическая пирофосфатаза	"Неорганическая пирофосфатаза () (PPаза) представляет собой фермент, ответственный за гидролиз пирофосфата (PPi), который образуется в основном как продукт многих биосинтетических реакций, в которых используется ATP. Все известные PPазы требуют присутствия катионов двухвалентных металлов, при этом магний придает наибольшую активность. Предполагается, что среди других остатков лизин является частью активного центра или находится близко от него. PPазы были секвенированы из бактерий, таких как Escherichia coli (гомогексамер), Bacillus PS3 (термофильная бактерия PS-3) и Thermus thermophilus, из архебактерий Thermoplasma acidophilum, из грибов (гомодимер), из растений и из бычьей сетчатки. У дрожжей была охарактеризована митохондриальная изоформа PPазы, которая, по-видимому, участвует в выработке энергии и активность которой стимулируется средствами, разобщающими синтез ATP. Последовательности PPаз имеют некоторые области сходства, среди которых есть область, содержащая три консервативных остатка аспартата, которые участвуют в связывании катионов."
171	PS00387	Предполагаемая нуклеаза HARBI1	NA
172	PS00290	Предполагаемая нуклеаза HARBI1	NA

173	IPR003006	"Иммуноглобулин/главный комплекс гистосовместимости, консервативный сайт"	"Основная структура молекул иммуноглобулина (Ig) представляет собой тетрамер из двух легких цепей и двух тяжелых цепей, соединенных дисульфидными связями. Существует два типа легких цепей: каппа и лямбда, каждая из которых состоит из константного домена (CL) и вариабельного домена (VL). Существует пять типов тяжелых цепей: альфа, дельта, эпсилон, гамма и мю, все из которых состоят из вариабельного домена (VH) и трех (в альфа, дельта и гамма) или четырех (в эпсилон и мю) константных доменов (от CH1-CH4). Молекулы Ig представляют собой высокомодульные белки, в которых вариабельные и константные домены имеют четкие консервативные паттерны последовательностей. Домены в молекулах Ig и Ig-подобных молекулах разделены на четыре группы: V-группа (вариабельные), C1-группа (константные-1), C2-группа (константные-2) и I-группа (промежуточные). Структурные исследования показали, что эти домены имеют общую сердцевинную структуру типа бета-"сэндвича" с топологией типа "греческого ключа", причем типы различаются количеством нитей в бета-листах, а также их паттернами последовательностей. Иммуноглобулиноподобные домены, которые родственны как по последовательности, так и по структуре, могут быть обнаружены в нескольких различных семействах белков. Ig-подобные домены участвуют в различных функциях, включая межклеточное распознавание, деятельность рецепторов клеточной поверхности, обеспечение мышечной структуры и деятельность иммунной системы. Гликопротеины главного комплекса гистосовместимости (MHC) представляют собой гетеродимерные рецепторы клеточной поверхности, функция

			которых заключается в представлении антигенных пептидных фрагментов Т-клеткам, отвечающим за клеточноопосредованные иммунные ответы. Молекулы MHC можно разделить на две группы на основании структуры и функции: молекулы класса I представляют внутриклеточные антигенные пептидные фрагменты (~ 10 аминокислот) на поверхности клеток-хозяев цитотоксическим Т-клеткам; молекулы класса II представляют антигенные пептиды экзогенного происхождения (~ 15 аминокислот) Т-клеткам-хелперам. Молекулы MHC класса I и II собираются и нагружаются своими пептидными лигандами с помощью различных механизмов. Однако в обоих случаях они представляют пептидные фрагменты, а не целые белки Т-клеткам, и они необходимы для формирования иммунного ответа. Некоторые белки этой группы отвечают за молекулярную основу антигенов групп крови - поверхностных маркеров на наружной стороне мембраны эритроцитов. Большинство этих маркеров представляют собой белки, но некоторые представляют собой углеводы, присоединенные к липидам или белкам [Reid M.E., Lomas-Francis C. The Blood Group Antigen FactsBook Academic Press, London/San Diego, (1997)]. Гликопротеин системы групп крови Лютерана (гликопротеин клеточной поверхности B-CAM) (антиген системы Аубергера B) (антиген F8/G253) принадлежит к системе групп крови Лютерана и ассоциирован с антигенами Lu(a/b), Au(a/b), LU6-LU20."
174	IPR039091	Рецептор ароматических углеводородов/репрессор	"Рецептор ароматических углеводородов (AHR) представляет собой активируемый лигандом фактор транскрипции, который опосредует биологические и токсические эффекты диоксиноподобных соединений. После связывания лиганда рецептор

		рецепторов ароматических углеводородов	перемещается в ядро, где он гетеродимеризуется с ARNT (ядерным транслокатором AHR). Репрессор рецепторов ароматических углеводородов (AHRR) конкурирует с AHR за связывание ARNT и тем самым подавляет транскрипционную активность AHR."
175	PF02338	OTU-подобная цистеиновая протеаза	"Это семейство состоит из группы предсказанных цистеиновых протеаз, гомологичных гену Ovarian Tumor (OTU) у дрозофилы. Представители включают белки эукариот, вирусов и патогенных бактерий. Консервативные цистеин и гистидин и, возможно, аспартат представляют собой каталитические остатки в этой предполагаемой группе протеаз."
176	IPR003323	Домен OTU	"Область гомологии, содержащая четыре консервативных мотива, была идентифицирована в белках эукариот, нескольких групп вирусов и патогенных бактерий Chlamydia pneumoniae. Ни один из этих белков не имеет известной биохимической функции, но низкое сходство последовательности с полипротеиновыми областями артеривирусов позволяет предположить, что он может обладать активностью цистеиновой протеазы. В этом случае каталитическими остатками могут быть консервативные остатки цистеина и аспартата в мотиве I и остатки гистидина в мотиве IV. Мотивы II и III имеют более ограниченную консервативность последовательности и могут участвовать в распознавании субстрата. Было высказано предположение, что эукариотические белки, содержащие домен OTU, могут опосредовать протеолитические события, вовлеченные в передачу сигналов, ассоциированных с модификацией структуры

			хроматина и контролем клеточной пролиферации. В вирусах белки, содержащие этот домен, аннотированы как репликаза или РНК-зависимая РНК-полимераза. Эукариотические последовательности родственны гену Ovarian Tumor (OTU) у дрозофилы, деубиквитинирующей пептидазе Cezanne и индуцированному фактором некроза опухоли-альфа белку 3 (семейство пептидаз C64 по MEROPS), а также отубаину 1 и отубаину 2 (семейство пептидаз C65 по MEROPS). Цистеиновые пептидазы с химотрипсиноподобной укладкой принадлежат к клану PA, к которому также принадлежат сериновые пептидазы. Цистеиновые пептидазы, представляющие собой N-концевые нуклеофильные гидролазы, входят в клан PB. Цистеиновые пептидазы с третичной структурой, сходной с таковой у аспартилдипептидазы серинового типа, входят в клан PC. Цистеиновые пептидазы с интеиноподобной укладкой входят в клан PD, в который также входят аспарагинлиазы. Цистеиновая пептидаза представляет собой протеолитический фермент, гидролизующий пептидную связь, используя в качестве нуклеофила тиольную группу цистеинового остатка. Гидролиз обычно предусматривает каталитическую триаду, состоящую из тиольной группы цистеина, имидазолиевого кольца гистидина и третьего остатка, обычно аспарагина или аспарагиновой кислоты, для обеспечения ориентации и активации имидазолиевого кольца. Только в одном семействе цистеиновых пептидаз роль общего основания отводится остатку, отличному от гистидина: в пептидазах семейства C89 (кислые церамидазы) общее основание представляет собой аргинин. Цистеиновые пептидазы можно

			сгруппировать в четырнадцать различных кланов, при этом представители каждого клана характеризуются третичной укладкой, уникальной для клана. Четыре клана цистеиновых пептидаз имеют структурное сходство с сериновыми и треониновыми пептидазами и аспарагинлиазами. По сходству последовательностей цистеиновые пептидазы можно объединить в более чем 80 различных семейств. Кланы CF, CM, CN, CO, CP и PD содержат только одно семейство. Цистеиновые пептидазы часто активны при кислых pH и поэтому ограничены кислыми средами, такими как лизосомы животных или растительные вакуоли. Цистеиновые пептидазы могут представлять собой эндопептидазы, аминопептидазы, карбоксипептидазы, дипептидилпептидазы или омега-пептидазы. Они ингибируются тиол-реактивными хелаторами, такими как йодацетат, йодуксусная кислота, N-этилмалеимид или п-хлормеркурибензоат. Клан CA включает белки с папаиноподобной укладкой. Существует каталитическая триада, которая представлена в следующем порядке: Cys/His/Asn (или Asp). Четвертый остаток, обычно Gln, важен для стабилизации ацильного промежуточного соединения, образующегося в ходе катализа, и он предшествует Cys в активном центре. Укладка состоит из двух субдоменов, между которыми находится активный центр. Один субдомен состоит из пучка спиралей с каталитическим Cys на конце одной из них, а другой субдомен представляет собой бета-"бочку" с His и Asn (или Asp) в активном центре. К этому клану принадлежат более тридцати семейств, и для представителей большинства из них были расшифрованы третичные структуры. Пептидазы клана CA обычно чувствительны к

			низкомолекулярному ингибитору E64, который неэффективен в отношении пептидаз из других кланов цистеиновых пептидаз. К клану CD относятся белки с каспазоподобной укладкой. Белки в клане имеют структуру альфа/бета/альфа-"сэндвича". Существует каталитическая диада, которая представлена в порядке His/Cys. His активного центра встречается в мотиве His-Gly, а Cys активного центра встречается в мотиве Ala-Cys; обоим мотивам предшествует блок гидрофобных остатков. Специфичность в основном направлена на остатки, которые занимают карман связывания S1, так что каспазы расщепляют аспартильные связи, легумаины расщепляют аспарагинильные связи, а гингипаины расщепляют лизильные или аргинильные связи. К клану CE относятся белки с аденаиноподобной укладкой. Укладка состоит из двух субдоменов, между которыми находится активный центр. Один домен представляет собой пучок спиралей, а другой представляет собой бета-"бочку". Субдомены расположены в порядке, противоположном обнаруживаемому в пептидазах клана CA, и это отражено в порядке остатков активного центра: His/Asn/Gln/Cys. Это вызвало предположение, что белки в кланах CA и CE родственны, и что представители одного клана произошли в результате циклической пермутации структуры представителей другого. К клану CL относятся белки с укладкой, подобной сортазе B. Пептидазы в клане гидролизуют и переносят пептиды клеточной стенки бактерий. На укладке показана замкнутая бета-"бочка", вынизанная спиралями, с активным центром на одном конце "бочки". Активный центр состоит из каталитической диады His/Cys."

177	PS50802	Профиль домена OTU	"Область гомологии, содержащая четыре консервативных мотива, была идентифицирована в белках эукариот, нескольких групп вирусов и патогенных бактерий Chlamydia pneumoniae [PMID:10664582]. Ни один из этих белков не имеет известной биохимической функции, но низкое сходство последовательности с полипротеиновыми областями артеривирусов позволяет предположить, что он может обладать активностью цистеиновой протеазы [PMID:10664582]. В этом случае каталитическими остатками могут быть консервативные остатки цистеина и аспартата в мотиве I и остатки гистидина в мотиве IV. Мотивы II и III имеют более ограниченную консервативность последовательности и могут участвовать в распознавании субстрата [PMID:10664582]. Было высказано предположение, что эукариотические белки, содержащие домен OTU, могут опосредовать протеолитические события, вовлеченные в передачу сигналов, ассоциированных с модификацией структуры хроматина и контролем клеточной пролиферации [PMID:10664582]. Некоторые белки, содержащие домен OTU, перечислены ниже [PMID:10664582]. - Ovarian Tumor (OTU) - белок дрозофилы, участвующий в морфогенезе ооцитов. Поскольку предполагаемый каталитический цистеин заменен серином, неясно, является ли OTU гомологом активной протеазы или инактивированной протеазы. - Предполагаемый белок, индуцированный HIV-1 (HIN1) - человеческий гомолог OTU. - Белок B0546.2 - гомолог OTU Caenorhabditis elegans. - Белок M2 вируса болезни "Ла-Франс". - Индуцированный фактором некроза опухоли-альфа белок A20 (TNP3) млекопитающих. Он защищает клетки от

			апоптоза и блокирует активацию лимфоидных клеток. В дополнение к домену OTU, TNP3 содержит несколько копий мотива, подобного "цинковому пальцу", который также присутствует в других белках, участвующих в убиквитин-зависимом протеолизе. - Гипотетический белок YFL044C из дрожжей - белок, который также содержит "цинковый палец" типа C2H2. - Гипотетический белок F12K2.7 - белок Arabidopsis thaliana. - Гипотетический белок C25D7.8 - белок Caenorhabditis elegans. - Гипотетический белок CPN0483 Chlamydia pneumoniae. Разработанный авторами настоящего изобретения профиль охватывает весь домен OTU."
178	IPR001849	Домен, гомологичный плекстрину	"Домены, гомологичные плекстрину (PH), представляют собой небольшие модульные домены, встречающиеся в большом количестве белков. Эти домены могут связывать фосфатидилинозитол в биологических мембранах и белках, таких как бета/гамма-субъединицы гетеротримерных G-белков и протеинкиназа C. Благодаря этим взаимодействиям домены PH играют роль в привлечении белков к различным мембранам, тем самым направляя их в соответствующие клеточные компартменты или позволяя им взаимодействовать с другими компонентами путей передачи сигнала. Было обнаружено, что домены PH содержат вставленные домены (как, например, в PLC-гамма, синтрофинах) и подвергаются вставке в другие домены. Мутации в тирозинкиназе Брутона (Btk) в пределах ее домена PH вызывают X-сцепленную агаммаглобулинемию (XLA) у пациентов. Точечные мутации группируются на положительно заряженном конце молекулы вокруг предполагаемого сайта связывания фосфатидилинозитоловых липидов. Была

			определена 3D-структура нескольких доменов PH. Все известные случаи имеют общую структуру, состоящую из двух перпендикулярных антипараллельных бета-листов, за которыми следует С-концевая амфипатическая спираль. Петли, соединяющие бета-нити, сильно различаются по длине, что делает относительно сложным выявление домена PH. В домене PH нет полностью инвариантных остатков. Белки, которые, как сообщается, содержат один или несколько доменов PH, принадлежат к следующим семействам. * Плекстрин, белок, в котором этот домен был впервые выявлен, является основным субстратом протеинкиназы С в тромбоцитах. Плекстрин представляет собой один из редких белков, который содержит два домена PH. * Протеинкиназы Ser/Thr, такие как принадлежащие к семействам Akt/Rac, киназы бета-адренергических рецепторов, мю-изоформа PKC и семейство трипаносомальных NrkA. * Тирозиновые протеинкиназы, принадлежащие к подсемейству Btk/Itk/Tec. * Субстрат 1 рецептора инсулина (IRS-1). * Регуляторы малых G-белков, такие как фактор высвобождения гуаниновых нуклеотидов GNRP (Ras-GRF) (который содержит 2 домена PH), белки обмена гуаниновых нуклеотидов, такие как vav, dbl, SoS и CDC24 Saccharomyces cerevisiae, белки, активирующие GTPазу, такие как rasGAP и BEM2/IPL2, и белок области кластеризации точек разрыва человека bcr. * Белки цитоскелета, такие как динамин (см.), кинезиноподобный белок unc-104 Caenorhabditis elegans (см.), бета-цепь спектрина, синтрофин (2 домена PH) и белок ядерной миграции NUM1 S. cerevisiae. * Фосфатидилинозитол-специфическая фосфолипаза С млекопитающих (PI-PLC) (см.), изоформы гамма и

			дельта. Изоформа гамма содержит два домена PH, при этом второй расщеплен на две части, разделенные приблизительно 400 остатками. * Оксистерол-связывающие белки OSBP, OSH1 и YHR073w S. cerevisiae. * Белок цитрон мыши - предполагаемый эффектор rho/rac, который связывается с GTP-связанными формами rho и rac. * Некоторые белки S. cerevisiae, участвующие в регуляции клеточного цикла и формировании почек, такие как BEM2-, BEM3-, BUD4- и BEM1-связывающие белки BOI2 (BEB1) и BOI1 (BOB1). * Белок MIG-10 C. elegans. * Гипотетические белки C04D8.1, K06H7.4 и ZK632.12 C. elegans. * Гипотетические белки YBR129c и YHR155w S. cerevisiae."
179	PS50003	Профиль домена PH	"Домен, "гомологичный плекстрину" (PH) представляет собой домен, состоящий из приблизительно 100 остатков, который встречается в широком спектре белков, участвующих во внутриклеточной передаче сигналов или представленных в качестве составляющих цитоскелета [PMID:8500161,PMID:8497315,PMID:8236453,PMID:7985225,PMID:7531822,PMID:7890802,PMID:7583640]. Функция этого домена не ясна, было предложено несколько предполагаемых функций: - связывание с бета/гамма-субъединицей гетеротримерных G-белков, - связывание с липидами, например, фосфатидилинозитол-4,5-бисфосфатом, - связывание с фосфорилированными остатками Ser/Thr, - присоединение к мембранам посредством неизвестного механизма. Возможно, что разные домены PH имеют совершенно разные требования в отношении лигандов. Была определена 3D-структура нескольких

			доменов PH [PMID:7634082]. Все известные случаи имеют общую структуру, состоящую из двух перпендикулярных антипараллельных бета-листов, за которыми следует С-концевая амфипатическая спираль. Петли, соединяющие бета-нити, сильно различаются по длине, что делает относительно сложным выявление домена PH. В домене PH нет полностью инвариантных остатков. Белки, которые, как сообщается, содержат один или несколько доменов PH, принадлежат к следующим семействам. - Плекстрин, белок, в котором этот домен был впервые выявлен, является основным субстратом протеинкиназы С в тромбоцитах. Плекстрин представляет собой один из редких белков, который содержит два домена PH. - Протеинкиназы Ser/Thr, такие как принадлежащие к семействам Act/Rac, киназы бета-адренергических рецепторов, мю-изоформа PKC и семейство трипаносомальных NrkA. - Тирозиновые протеинкиназы, принадлежащие к подсемейству Btk/Itk/Tec. - Субстрат 1 рецептора инсулина (IRS-1). - Регуляторы малых G-белков, такие как фактор высвобождения гуаниновых нуклеотидов GNRP (Ras-GRF) (который содержит 2 домена PH), белки обмена гуаниновых нуклеотидов, такие как vav, dbl, SoS и CDC24 дрожжей, белки, активирующие GTPазу, такие как rasGAP и BEM2/IPL2, и белок области кластеризации точек разрыва человека bcr. - Белки цитоскелета, такие как динамин, кинезиноподобный белок unc-104 Caenorhabditis elegans, бета-цепь спектрина, синтрофин (2 домена PH) и белок ядерной миграции NUM1 дрожжей. - Фосфатидилинозитол-специфическая фосфолипаза С млекопитающих (PI-PLC), изоформы гамма и дельта. Изоформа гамма

			содержит два домена PH, при этом второй расщеплен на две части, разделенные приблизительно 400 остатками. - Оксистерол-связывающие белки (OSBP). - Белок цитрон мыши - предполагаемый эффектор rho/rac, который связывается с GTP-связанными формами rho и rac. - Некоторые белки дрожжей, участвующие в регуляции клеточного цикла и формировании почек, такие как BEM2-, BEM3-, BUD4- и BEM1-связывающие белки BOI2 (BEB1) и BOI1 (BOB1). - Белок mig-10 Caenorhabditis elegans. - Гипотетические белки C04D8.1, K06H7.4 и ZK632.12 Caenorhabditis elegans. - Гипотетические белки YBR129c и YHR155w дрожжей. Профиль домена PH, разработанный Toby Gibson в EMBL, охватывает всю длину домена. Некоторые белки содержат вставки крупных участков в домене PH, и поэтому их трудно выявить с помощью этого профиля. В некоторых из этих случаев профиль будет соответствовать только половине домена PH."
180	cd15551	"Палец" типа PHD, обнаруживаемый в белках PYGO	"Семейство включает белок pygopus (dPYGO) Drosophila melanogaster и два его гомолога - PYGO1 и PYGO2. dPYGO представляет собой важнейший транскрипционный компонент сигнального пути Wnt у дрозофилы. PYGO1 необходим для ассоциации с Legless (Lgs)/Bcl9, который выступает в качестве адаптера между Pygopus (Pygo) и Arm/бета-катенином. dPYGO и PYGO2 функционируют в качестве зависимых от окружения коактиваторов бета-катенина, и они связывают ди- и триметилированный лизин 4 гистона H3 (H3K4me2/3). Более того, PYGO2 выступает в качестве считывателя метилирования гистонов и средства ремоделирования хроматина специфичным в отношении семенников и не

			связанным с Wnt образом. Он также опосредует регуляцию хроматина и связывает сигнальный путь Wnt и сигнальный путь Notch для подавления компетентности в отношении люминальной/альвеолярной дифференцировки стволовых и базальных клеток молочной железы. PYGO2 также играет новую роль в транскрипции рРНК во время роста раковых клеток. Он регулирует появление и гетерогенность опухолей молочной железы у мышей MMTV-Wnt1. Все представители семейства содержат "палец" типа гомеодомена растений (PHD). [PMID: 11955446, PMID: 19555349, PMID: 17499269, PMID: 17113272, PMID: 18498752, PMID: 23637336, PMID: 11988739, PMID: 20637214, PMID: 23684539, PMID: 23517060, PMID: 23334328, PMID: 20361361, PMID: 24183574, PMID: 12015286, PMID: 18404694, PMID: 24265259, PMID: 24046329]"
181	cd17065	Убиквитин-подобный (Ubl) домен, обнаруживаемый в убиквитин-карбоксиконцевой гидролазе 24 (UBP24) и сходных белках	"UBP24 (EC 3.4.19.12), также называемый деубиквитинирующим ферментом 24, или убиквитинтиоэстеразой 24, или убиквитин-специфической процессинговой протеазой 24 (USP24), представляет собой деубиквитинирующий белок, который взаимодействует со специфичным в отношении повреждений ДНК-связывающим белком 2 (DDB2) и регулирует стабильность DDB2. Он также может играть роль в патогенезе болезни Паркинсона (PD). Белки UBP24 содержат N-концевой убиквитин-ассоциированный (UBA) домен, убиквитин-подобный (Ubl) домен и C-концевой пептидазный домен C19. [PMID: 16917932, PMID: 20302855, PMID: 23159851, PMID: 22923019, PMID: 18854859]"

182	cd18013	Домен хеликазы DEXQ-бокса бактериальных белков семейства SNF2	"Белки, принадлежащие к семейству SNF2 ДНК-зависимых АТРаз, являются важными представителями комплексов ремоделирования хроматина, которые участвуют в эпигенетическом контроле экспрессии генов. Семейство Snf2 включает большую группу белков, гидролизующих АТР, которые широко распространены у эукариот, но также присутствуют у эубактерий и архей. Бактериальные SNF2, присутствующие в этом семействе, являются представителями суперсемейства DEAD-подобных хеликаз - разнообразного семейства белков, участвующих в АТР-зависимом раскручивании РНК или ДНК. Этот домен содержит АТР-связывающую область. [PMID: 15627498, PMID: 16738128, PMID: 18508648, PMID: 19234488, PMID: 18786849, PMID: 15207721, PMID: 16935875, PMID: 16568948, PMID: 1552844, PMID: 2563148, PMID: 11545728, PMID: 10322435, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"
183	IPR037931	Субъединица BAP18 комплексов ремоделирования хроматина	BAP18 представляет собой компонент комплексов ремоделирования хроматина, таких как MLL1/MLL и NURF. Было показано, что этот белок способствует привлечению субкомплекса MLL1 и андрогенового рецептора (AR) к элементу ответа на андрогены (ARE) генов-мишеней AR. BAP18 содержит домен SANT.
184	cd04643	Два тандемных повтора цистатионин-бета-синтазных доменов (пара CBS), присутствующих в бактериях	"Домен CBS, названный в честь CBS человека, представляет собой небольшой домен, первоначально идентифицированный в цистатионин-бета-синтазе и впоследствии обнаруженный в широком диапазоне различных белков. Домены CBS обычно встречаются в тандемных повторах. Они связываются с образованием так называемого домена Бейтмана или пары CBS, исходя из кристаллографических

			исследований бактерий. Пара CBS использовалась в качестве основы для этой иерархии cd, поскольку белки CBS человека могут принимать типичную сердцевинную структуру и образовывать внутримолекулярную пару CBS. Область контакта между двумя доменами CBS образует расщелину, которая является потенциальным сайтом связывания лиганда. Пара CBS сосуществует с множеством других функциональных доменов, и это было использовано в данном документе для содействия в ее классификации. Было высказано предположение, что домен CBS может играть регулирующую роль, хотя его точная функция неизвестна. Мутации консервативных остатков в этом домене ассоциированы с различными наследственными заболеваниями человека, включая врожденную миотонию, идиопатическую генерализованную эпилепсию, гиперкальциурический нефролитиаз и классический синдром Барттера (представители семейства хлоридных каналов CLC), синдром Вольфа-Паркинсона-Уайта (субъединица гамма 2 AMP-активируемой протеинкиназы), пигментный ретинит (IMP-дегидрогеназа-1) и гомоцистинурию (цистатионин-бета-синтаза). [PMID: 16275737, PMID: 9020585, PMID: 9106071, PMID: 9590298, PMID: 10200156, PMID: 11230183, PMID: 11524006, PMID: 14722619]"
185	cd18357	"Первый домен BTB (Broad-Complex, Tramtrack и Bric a brac)/POZ (поксвирус и "цинковый палец"), обнаруживаемый в Rho-	"RhoBTB3 представляет собой атипичную небольшую гуанозинтрифосфатазу (GTPазу) семейства Rho и является представителем подсемейства RhoBTB, которое характеризуется наличием GTPазного домена (в большинстве случаев нефункционального), за которым следуют богатая пролином область, тандемные домены BTB и консервативная С-концевая область. Карбоксиконцевое удлинение,

		родственном белке 3, содержащем домен BTB (RhoBTB3)"	которое содержит два домена BTB, способно к сборке куллин-3-зависимых убиквитинлигазных комплексов. RhoBTB3 представляет собой ассоциированную с комплексом Гольджи Rho-родственную АТРазу, которая регулирует переход S/G2 клеточного цикла путем нацеливания на циклин E для его убиквитинирования. Он участвует в везикулярном транспорте и в направлении белков на их разрушение в протеасоме. Он напрямую связывается с GTPазой Rab9 и функционирует с Rab9 в транспорте белка из эндосом в транс-сеть Гольджи. Он также содействует разрушению в протеасомах фактора альфа, индуцируемого гипоксией (HIF-альфа), способствуя гидроксилированию, и подавляет эффект Варбурга. Эта модель соответствует первому домену BTB. Домен BTB/POZ представляет собой распространенный мотив белок-белкового взаимодействия, состоящий из приблизительно 100 аминокислот. [PMID: 27314390, PMID: 18298893, PMID: 12426103, PMID: 26215701, PMID: 25670879, PMID: 24923387, PMID: 24145166, PMID: 22245496, PMID: 19531346, PMID: 19490898, PMID: 18835386]"
186	cd06503	"АТР-синтаза F-типа, мембранная субъединица b"	"Мембранная субъединица b является компонентом комплекса Fo в FoF1-ATP-синтазе. АТР-синтазы F-типа (FoF1-АТРазы) состоят из двух структурных доменов: комплекса F1 (фактора сборки 1), содержащего растворимую каталитическую сердцевину, и комплекса Fo (фактора, чувствительного к олигомицину), содержащего мембранный протонный канал, связанных друг с другом центральным "стеблем" и периферическим "стеблем". F1 состоит из субъединиц альфа (или A), бета (B), гамма (C), дельта (D) и эпсилон (E) в стехиометрическом соотношении

			3:3:1:1:1, тогда как Fo состоит из трех субъединиц: a, b и c (1:2:10-14). Олигомерное кольцо из 10-14 c-субъединиц (c-кольцо) образует ротор Fo. Поток протонов через АТРазный канал (Fo) управляет вращением c-кольца, которое, в свою очередь, сопряжено с вращением гамма-субъединичного ротора комплекса F1 вследствие постоянного связывания между гамма- и эпсилон-субъединицами F1 и c-кольцом Fo. F-АТР-синтазы главным образом обнаруживаются во внутренних мембранах эукариотических митохондрий, в тилакоидных мембранах хлоропластов или в плазматических мембранах бактерий. F-АТР-синтазы являются основными продуцентами АТР, использующими протонный градиент, генерируемый в ходе окислительного фосфорилирования (митохондрии) или фотосинтеза (хлоропласты). В качестве альтернативы, в условиях низкой движущей силы АТР-синтазы функционируют как АТРазы, генерируя таким образом трансмембранный градиент протонов или Na(+) за счет энергии, полученной в результате гидролиза АТР. В эту группу также входит F-АТР-синтаза, которая также была обнаружена у архей Candidatus methanoperedens. [PMID: 28001127, PMID: 24650630, PMID: 11533110, PMID: 12370007, PMID: 10580496, PMID: 26460036, PMID: 27791192, PMID: 26439008]"
187	IPR038738	Nedd8-подобный убиквитин	"Nedd8 (также известный как Rub1) имеет единственный консервативный убиквитин-подобный домен, который является частью пути модификации белка, подобного таковому убиквитина. Nedd8 модифицирует семейство молекулярных каркасных белков, называемых куллинами, которые отвечают за сборку ROC1/Rbx1 - убиквитинлигаз E3 на основе RING, некоторые из которых играют непосредственную

			роль в онкогенезе. Эта запись также включает Nedd8-подобные белки грибов и растений."
188	cd01806	Убиквитин-подобный (Ubl) домен, обнаруживаемый в подавляемом в ходе развития белке 8, экспрессируемом в клетках-предшественниках нейронов (NEDD8), и в сходных белках	"NEDD8, также называемый неддилином или белком, родственным убиквитину (RUB/Rub1p), в растениях и дрожжах, представляет собой убиквитин-подобный белок, который конъюгирует с ядерными белками способом, аналогичным убиквитинированию и сентринизации. Он модифицирует семейство молекулярных каркасных белков, называемых куллинами, которые отвечают за сборку ROC1/Rbx1 - убиквитинлигаз E3 на основе RING, некоторые из которых играют непосредственную роль в онкогенезе. Дезамидирование NEDD8 и ингибирование им активности убиквитинлигаз куллин-RING (CRL) отвечают за цитопатический эффект, индуцируемый ингибирующим клеточный цикл фактором (Cif)/гомологом Cif в Burkholderia pseudomallei (CHBP). NEDD8 содержит один консервативный убиквитин-подобный (Ubl) домен, характеризующийся укладкой типа бета-"захвата" Ubl - общей структурой, участвующей в белок-белковых взаимодействиях. Полиубиквитинирование обеспечивает сигналы для разнообразного набора клеточных событий посредством различных изопептидных связей, образующихся между C-концом одного убиквитина (Ub) и эпсилон-аминогруппой K6, K11, K27, K29, K33, K48 или K63 второго Ub. Ubl NEDD8 содержит многие из тех же остатков лизина (K6, K11, K27, K33, K48), что и Ub, где K27 играет роль (отличную от конъюгации) в механизме неддилирования белка. [PMID: 24773014, PMID: 3041007, PMID: 9857029, PMID: 15775960, PMID: 17220875, PMID: 18652489, PMID: 18805092, PMID: 20129059,

			PMID: 23175788, PMID: 15567417, PMID: 14690597, PMID: 9857030, PMID: 15242646, PMID: 12759363, PMID: 12730221, PMID: 12215427, PMID: 11953428, PMID: 11259415, PMID: 10597293, PMID: 9790970, PMID: 9353319, PMID: 10318914, PMID: 9694792, PMID: 22708563, PMID: 25833379, PMID: 18793140]"
189	IPR001623	Домен DnaJ	"Аппарат шаперонов hsp70 играет множество разнообразных ролей в клетке, включая сворачивание формирующихся белков, транслокацию полипептидов через мембраны органелл, координацию ответов на стресс и направление выбранных белков на разрушение. DnaJ является представителем семейства молекулярных шаперонов hsp40, которое также называют семейством J-белков, представители которого регулируют активность hsp70. DnaJ (hsp40) связывается с dnaK (hsp70) и стимулирует его АТРазную активность, создавая ADP-связанное состояние dnaK, который стабильно взаимодействует с полипептидным субстратом. Структурно белок DnaJ состоит из N-концевого консервативного домена (называемого "J-доменом"), состоящего из приблизительно 70 аминокислот, богатой глицином области ("G-домена"), состоящей из приблизительно 30 остатков, центрального домена, содержащего четыре повтора мотива CXXCXGXG ("CRR-домена"), и С-концевой области, состоящей из 120-170 остатков. Такая структура показана в следующем схематическом представлении: +----------+-+-------+-----+----------+---------- ----------+ | J-домен | | Gly-R | | CXXCXGXG | С-конец | +----------+-+-------+----- +----------+----------+ Структура J-домена была расшифрована. J-домен состоит из четырех спиралей, вторая из которых имеет заряженную поверхность,

			содержащую основные остатки, необходимые для взаимодействия с АТРазным доменом hsp70. J-домены обнаружены во многих прокариотических и эукариотических белках. У дрожжей было идентифицировано три J-подобных белка, содержащих области, очень похожие на J-домен, но не содержащие консервативный мотив HPD; эти белки, по-видимому, не действуют как молекулярные шапероны."
190	cd14686	Домен лейциновой "молнии" с основной областью (bZIP) факторов транскрипции bZIP: домен связывания ДНК и димеризации	"Факторы типа лейциновой "молнии" с основной областью (bZIP) составляют один из наиболее важных классов факторов транскрипции энхансерного типа. Они действуют в сетях гомо- и гетеродимеров при регуляции разнообразного набора клеточных процессов, включая выживание, обучение и память клеток, метаболизм липидов и прогрессирование рака, помимо прочего. Они также играют важные роли в ответах на стимулы или сигналы стресса, такие как цитокины, генотоксические средства или виды физиологического стресса. Структурный мотив bZIP содержит основную область и лейциновую "молнию", состоящую из альфа-спиралей с лейциновыми остатками, отстоящими друг от друга на 7 аминокислот, которые стабилизируют димеризацию с параллельным доменом типа лейциновой "молнии". При димеризации лейциновых "молний" образуется пара прилегающих основных областей, которые связываются с ДНК и претерпевают конформационные изменения. Димеризация происходит специфическим и предсказуемым образом, в результате чего сотни димеров оказывают уникальные эффекты в отношении транскрипции. [PMID: 16731568, PMID: 16580748, PMID: 19519454, PMID: 19150388,

			PMID: 18424222, PMID: 21454302, PMID: 21438114, PMID: 21257317, PMID: 21061914, PMID: 23661758, PMID: 19880309, PMID: 8504253, PMID: 7780801, PMID: 16410410]"
191	cd06257	Домен DnaJ или J-домен. Белки DnaJ/Hsp40 (белок теплового шока 40) являются высококонсервативными	"Домен DnaJ или J-домен. Белки DnaJ/Hsp40 (белок теплового шока 40) являются высококонсервативными и играют решающую роль в трансляции, сворачивании, разворачивании, транслокации и разрушении белков. Они действуют в первую очередь путем стимуляции АТРазной активности Hsp70s - важного семейства шаперонинов. Белки Hsp40 характеризуются наличием J-домена, который опосредует взаимодействие с Hsp70. Они могут содержать и другие домены, а их архитектуры предоставляют возможности их классификации. [PMID: 16952052, PMID: 15170475, PMID: 10830586, PMID: 10441399, PMID: 10221986, PMID: 8016869, PMID: 9585179, PMID: 17919282, PMID: 17428722, PMID: 8764403, PMID: 9271376, PMID: 9644977]"
192	PF00226	Домен DnaJ	"Домены DnaJ (J-домены) ассоциированы с системой белков теплового шока hsp70, и считается, что этот домен опосредует взаимодействие. Таким образом, домен DnaJ является частью системы шаперонов (сворачивания белков). Т-антигены, хотя и не в базе Prosite, подтверждены как домены, содержащие DnaJ, из литературы [PMID:9271376]."
193	PS50076	Профиль домена dnaJ	"Аппарат шаперонов hsp70 играет множество разнообразных ролей в клетке, включая сворачивание формирующихся белков, транслокацию полипептидов через мембраны органелл, координацию ответов на стресс и направление выбранных белков на разрушение. DnaJ является представителем семейства молекулярных

			шаперонов hsp40, которое также называют семейством J-белков, представители которого регулируют активность hsp70. DnaJ (hsp40) связывается с dnaK (hsp70) и стимулирует его АТРазную активность, создавая ADP-связанное состояние dnaK, который стабильно взаимодействует с полипептидным субстратом [1,2]. DnaJ состоит из N-концевого консервативного домена (называемого "J-доменом"), состоящего из приблизительно 70 аминокислот, богатого глицином и фенилаланином домена ("G/F-домена"), центрального домена, богатого цистеином ("цинкового пальца" типа CR), содержащего четыре повтора мотива CXXCXGXG, который может координировать два атома цинка, и C-концевого домена (CTD) [2]. Такая структура показана в следующем схематическом представлении: +----------+-+-----------+-----+-----------+-------------+ | J-домен | | Gly/Phe-R | | CXXCXGXG | CTD | +----------+-+-----------+-----+-----------+-------------+ Структуры "J-домена" и "CR-домена" были расшифрованы [PMID:8764403,PMID:10891270]. J-домен состоит из четырех спиралей, вторая из которых имеет заряженную поверхность, содержащую основные остатки, необходимые для взаимодействия с АТРазным доменом hsp70 [PMID:12454054]. "Цинковый палец" типа CR имеет общую топологию V-образной удлиненной бета-шпильки и два симметричных сайта связывания цинка, обозначенных как Zn1 и Zn2: Zn1 образован двумя цистеиновыми мотивами, наиболее удаленными друг от друга в первичной последовательности, тогда как Zn2 образован двумя центральными смежными цистеиновыми мотивами [PMID:10891270]. Было показано, что Zn1 важен для автономной независимой от

			dnaK шаперонной активности, тогда как Zn2 является необходимым сайтом взаимодействия с dnaK, который, по-видимому, имеет решающее значение для функционирования in vivo в системе dnaJ/dnaK [PMID:12941935]. Семейство J-белков подразделяют на три класса [2]. Тип 1 содержит белки с J-доменом, G/F-доменом, "цинковым пальцем" типа CR и доменом CTD (истинные гомологи dnaJ): - Белок MAS5/YDJ1 дрожжей, участвующий в сворачивании белков и митохондриальном импорте белков. - Белок MDJ1 дрожжей, участвующий в сворачивании белков и митохондриальном импорте белков. - Белок SCJ1 дрожжей, участвующий в сворачивании белков в просвете эндоплазматического ретикулума. - Белок APJ1/YNL077w дрожжей, который может модулировать реакцию сворачивания. - Гомолог dnaJ огурца, заякоренный на мембране глиоксисом. - Белок XDJ1 дрожжей. - Белок HDJ2 человека. Тип 2 содержит белки с J-доменом, G/F-доменом и доменом CTD: - Белок nolC Rhizobium fredii - белок, участвующий в сортоспецифическом образовании клубеньков сои. - Белок cbpA Escherichia coli - белок, связывающий искривленную ДНК. - Белок SIS1 дрожжей, необходимый для ядерной миграции во время митоза. - Белок CAJ1 дрожжей. - Гипотетический белок YFR041c дрожжей. - Гипотетический белок YIR004w дрожжей. - Гипотетический белок YJL162c дрожжей. - Поверхностный антиген эритроцитов, инфицированных кольцевой стадией развития Plasmodium falciparum (RESA). RESA, функция которого неизвестна, ассоциирован с мембранным скелетом недавно подвергнувшихся инвазии эритроцитов. - HDJ1 человека. - HSJ1 человека, нейрональный белок. - Белок с цистеиновой нитью (csp)

			дрозофилы. Подгруппа типа 3 содержит белки, имеющие только J-домен: - Белок SEC63/NPL1 дрожжей. Он важен для сборки белков в эндоплазматическом ретикулуме и ядре. - Эукариотический белок Tim14. Важный компонент комплекса PAM - комплекса, необходимого для транслокации белков, содержащих транзитные пептиды, из внутренней мембраны в митохондриальный матрикс ATP-зависимым образом. В комплексе требуется стимуляция активности mtHSP70 (SSC1). - Белок Jac1 дрожжей и HscB эукариот"
194	IPR036869	Суперсемейство J-доменов шаперонов	"Аппарат шаперонов hsp70 играет множество разнообразных ролей в клетке, включая сворачивание формирующихся белков, транслокацию полипептидов через мембраны органелл, координацию ответов на стресс и направление выбранных белков на разрушение. DnaJ является представителем семейства молекулярных шаперонов hsp40, которое также называют семейством J-белков, представители которого регулируют активность hsp70. DnaJ (hsp40) связывается с dnaK (hsp70) и стимулирует его ATPазную активность, создавая ADP-связанное состояние dnaK, который стабильно взаимодействует с полипептидным субстратом. Структурно белок DnaJ состоит из N-концевого консервативного домена (называемого "J-доменом"), состоящего из приблизительно 70 аминокислот, богатой глицином области ("G-домена"), состоящей из приблизительно 30 остатков, центрального домена, содержащего четыре повтора мотива CXXCXGXG ("CRR-домена"), и C-концевой области, состоящей из 120-170 остатков. Такая структура показана в следующем схематическом представлении: +----------+-+-------+-----+----------+----------

			----------+ | J-домен | | Gly-R | | CXXCXGXG | C-конец | +----------+-+-------+-----+----------+----------+ Структура J-домена была расшифрована. J-домен состоит из четырех спиралей, вторая из которых имеет заряженную поверхность, содержащую основные остатки, необходимые для взаимодействия с АТРазным доменом hsp70. J-домены обнаружены во многих прокариотических и эукариотических белках. У дрожжей было идентифицировано три J-подобных белка, содержащих области, очень похожие на J-домен, но не содержащие консервативный мотив HPD; эти белки, по-видимому, не действуют как молекулярные шапероны."
195	cd17282	"Домен распознавания мишени и консервативный домен (TRD-CR) субъединицы специфичности (S) системы рестрикции-модификации типа I, сходные с TRD1-CR1 S-субъединицы Escherichia coli G4/9 (S.Eco16444ORF1681P) и TRD2-CR2 S-субъединицы Zobellia galactanivorans DsiJT (S.ZgaJTORF2697P)"	"Распознающие последовательности S-субъединицы Escherichia coli G4/9 (S.Eco16444ORF1681P) и S-субъединицы Zobellia galactanivorans DsiJT (S.ZgaJTORF2697P) не определены. S-субъединица системы рестрикции-модификации (RM) состоит из двух вариабельных доменов распознавания мишени (TRD1 и 2) и двух консервативных областей (CR1 и CR2), которые разделяют TRD. Каждый TRD связывается с разными специфическими последовательностями ДНК. Системы RM защищают бактериальную клетку от инвазии чужеродной ДНК путем эндонуклеолитического расщепления ДНК, в которой отсутствует сайт-специфическая модификация. Геном хозяина защищен от расщепления посредством метилирования определенных нуклеотидов в сайтах-мишенях. В системах типа I активность как рестрикции, так и модификации присутствует в одном гетеромерном ферментативном комплексе, состоящем из одной субъединицы специфичности (S)

			ДНК (это семейство), двух субъединиц модификации (M) и двух субъединиц рестрикции (R). Эта модель содержит как TRD1-CR1, так и TRD2-CR2. Она также содержит TRD-CR-подобные домены распознавания последовательности различных предполагаемых ферментов рестрикции и метилтрансфераз типа II и может также содержать ДНК-метилтрансферазы типа I. [PMID: 8336674, PMID: 9837717, PMID: 25378308]"
196	IPR001611	Богатый лейцином повтор	"Богатые лейцином повторы (LRR) состоят из 2-45 мотивов длиной 20-30 аминокислот, которые обычно сворачиваются в дугообразную или подковообразную форму. LRR встречаются в белках от вирусов до эукариот и, по-видимому, обеспечивают структурный каркас для формирования белок-белковых взаимодействий. Белки, содержащие LRR, включают тирозинкиназные рецепторы, молекулы клеточной адгезии, факторы вирулентности и гликопротеины, связывающие внеклеточный матрикс, и участвуют во множестве биологических процессов, включая передачу сигнала, клеточную адгезию, репарацию ДНК, рекомбинацию, транскрипцию, процессинг РНК, устойчивость к болезням, апоптоз и иммунный ответ. Анализ последовательности содержащих LRR белков позволил предположить существование нескольких различных подсемейств LRR. Значение этой классификации состоит в том, что повторы из разных подсемейств никогда не встречаются одновременно и, наиболее вероятно, возникли в ходе эволюции независимо. Однако теперь ясно, что все основные классы LRR имеют изогнутые подковообразные структуры с параллельным бета-листом на вогнутой стороне и

			главным образом спиральными элементами на выпуклой стороне. Идентифицировано по меньшей мере шесть семейств содержащих LRR белков, характеризующихся различной длиной и консенсусной последовательностью повторов. Сегменты LRR с одиннадцатью остатками (LxxLxLxxN/CxL), соответствующие бета-нити и прилегающим петлевым областям, являются консервативными в содержащих LRR белках, тогда как остальные части повторов (называемые в данном документе вариабельными) могут сильно отличаться. Несмотря на различия, каждая из вариабельных частей содержит два полуповорота на обоих концах и "линейный" сегмент (поскольку цепь в целом следует линейной траектории), обычно образованный спиралью, в середине. Вогнутая поверхность и прилегающие к ней петли являются наиболее распространенными поверхностями взаимодействия с белками у содержащих LRR белков. 3D-структура некоторых комплексов содержащие LRR белки-лиганд показывает, что вогнутая поверхность домена с LRR идеальна для взаимодействия с альфа-спиралью, что тем самым подтверждает более ранние выводы о том, что удлиненная и изогнутая структура LRR обеспечивает превосходный каркас для достижения разнообразных белок-белковых взаимодействий. Молекулярное моделирование позволяет предположить, что консервативный паттерн LxxLxL, который короче предложенного ранее LxxLxLxxN/CxL, достаточен для придания характерной подковообразной кривизны белкам с повторами, содержащими 20-30 остатков."

197	PS51450	Профиль богатых лейцином повторов	"Домен с богатыми лейцином повторами (LRR) образован из тандемных массивов (2-52 повтора), содержащих консенсусную последовательность, богатую лейцином. Содержащие LRR белки были идентифицированы у вирусов, бактерий, архей и эукариот. Домены с LRR опосредуют макромолекулярные взаимодействия в таких разнообразных процессах, как инвазия бактерий в клетки-хозяева, иммунный ответ растений и ингибирование связывания РНК. Все повторы LRR можно разделить на высококонсервативный сегмент и вариабельный сегмент. Высококонсервативный сегмент состоит либо из отрезка из 11 остатков LxxLxLxxNxL, либо из отрезка из 12 остатков LxxLxLxxCxxL. Вариабельные последовательности, объединяющие консенсусную последовательность, обеспечивают функциональное разнообразие для связывания [PMID:16231091,PMID:11751054]. Различные кристаллографические исследования богатых лейцином повторов выявили, что каждый LRR содержит бета-нить и альфа-спираль, соединенные петлями (см., например, {PDB:2BNH})[PMID:8264799]. Несколько LRR расположены таким образом, что образуют неглобулярную подковообразную структуру, в которой параллельные бета-листы выстилают внутреннюю окружность подковы (на вогнутой стороне), а альфа-спирали вынизывают внешнюю окружность (на выпуклой стороне). Хотя последовательности и количество LRR могут различаться, разные LRR имеют эту общую подковообразную форму. Исследования мутагенеза и структурный анализ комплексов LRR-лиганд выявили, что вогнутая поверхность LRR участвует главным образом в связывании лиганда [PMID:11751054]. Было показано, что различные

			типы повторов сворачиваются высококооперативным образом. В целом сворачивание отдельного повтора энергетически невыгодно, тогда как образование областей контакта между повторами весьма благоприятно. Таким образом, сворачивание инициируется только тогда, когда два смежных повтора сворачиваются, образуя благоприятную область контакта. После преодоления этого энергетического барьера активации сворачивание происходит быстро посредством механизма нуклеации-роста. В некоторых повторах LRR было показано, что нуклеация инициируется в N-концевом кэпирующем домене [PMID:18462667,PMID:18462675]. Разработанный авторами настоящего изобретения профиль охватывает 22 аминокислоты повтора LRR."
198	cd05676	Цитозольные неспецифические дипептидазы M20, включая ансериназу и сывороточную карнозиназу	"Семейство пептидаз M20, подсемейство CNDP (цитозольные неспецифические дипептидазы), включая ансериназу (Xaa-метил-His-дипептидазу, EC 3.4.13.5), сывороточную карнозиназу (бета-аланил-L-гистидиндипептидазу; EC 3.4.13.20) и некоторые неохарактеризованные белки. Были идентифицированы два гена, CN1 и CN2, кодирующие белки, разрушающие карнозин (бета-аланил-L-гистидин) и гомокарнозин (гамма-аминомасляная кислота-L-гистидин) - два встречающихся в природе дипептида с потенциальными нейропротекторными и нейромедиаторными функциями. CN1 кодирует сывороточную карнозиназу и обладает узкой субстратной специфичностью в отношении дипептидов Xaa-His, где Xaa может представлять собой бета-аланин (карнозин), N-метил-бета-аланин, аланин, глицин и гамма-аминомасляную кислоту (гомокарнозин). CN2 соответствует

			цитозольной неспецифической дипептидазе (CNDP; EC 3.4.13.18) и не ограничивается дипептидами Xaa-His. CNDP требует Mn(2+) для полной активности и не гидролизует гомокарнозин. Ансериназа представляет собой дипептидазу, которая главным образом катализирует гидролиз N-альфа-ацетилгистидина. [PMID: 18550540, PMID: 21262300, PMID: 7674922, PMID: 22525515, PMID: 17012794, PMID: 16181134, PMID: 16302965, PMID: 16722645, PMID: 10493853]"
199	IPR005162	Домен gag ретротранспозонов	"Подвижные элементы (TE) содействуют различным хромосомным перестройкам более эффективно и часто более специфично, чем другие клеточные процессы. Ретротранспозоны структурно сходны с ретровирусами и ограничены длинными концевыми повторами. Эта запись представляет эукариотические белки Gag или связанные с капсидом белки, родственные ретротранспозонам. Существует центральный мотив QGXXEXXXXXFXXLXXH, который является общим для белков gag Retroviridae, но является низкоконсервативным."
200	PF03732	Белок gag ретротранспозонов	"Белки Gag или капсидоподобные белки из содержащих LTR ретротранспозонов. Существует центральный мотив QGXXEXXXXXFXXLXXH, который является общим для белков gag Retroviridae, но является низкоконсервативным [PMID:11600699]."
201	PS00141	Предполагаемая нуклеаза HARBI1	NA
202	IPR021109	Суперсемейство доменов аспарагиновых пептидаз	"Аспарагиновые пептидазы, также известные как аспартилпротеазы ([IntEnz:3.4.23.-]), представляют собой широко распространенные протеолитические ферменты, которые, как известно, существуют у позвоночных, грибов, растений, простейших,

			бактерий, архей, ретровирусов и некоторых вирусов растений. Все известные аспарагиновые пептидазы являются эндопептидазами. Молекула воды, активированная двумя остатками аспарагиновой кислоты, выступает в качестве нуклеофила в катализе. Аспарагиновые пептидазы можно разделить на пять кланов, каждый из которых демонстрирует уникальную структурную укладку. * Пептидазы в клане AA являются либо двухдольными (семейство A1 или семейство пепсина), либо гомодимерными (все остальные семейства в клане, включая ретропепсин из HIV-1/AIDS). Каждая доля состоит из одного домена с замкнутой бета-"бочкой", и каждая доля привносит один Asp для формирования активного центра. Большинство пептидаз в клане ингибируются встречающимся в природе низкомолекулярным ингибитором пепстатином. * Клан AC содержит единственное семейство A8: семейство сигнальной пептидазы 2. Представители семейства обнаруживаются у всех бактерий. Сигнальная пептидаза 2 подвергает процессингу предшественник премуреина, удаляя сигнальный пептид. Пептидаза имеет четыре трансмембранных домена, а активный центр находится на периплазматической стороне клеточной мембраны. Расщепление происходит со стороны аминогруппы цистеина, где тиольная группа была замещена диацилглицерильной группой. В ходе сайт-направленного мутагенеза были идентифицированы два незаменимых остатка аспарагиновой кислоты, которые встречаются в мотивах GNXXDRX и FNXAD (где X представляет собой гидрофобный остаток). Третичные структуры не были расшифрованы ни для одного представителя семейства, но из-за

			внутримембранного расположения предполагается, что структура не является пепсиноподобной. * Клан AD содержит два семейства трансмембранных эндопептидаз: A22 и A24. Они также известны как "пептидазы GXGD" ввиду наличия общего мотива GXGD, который содержит один из пары каталитических остатков аспарагиновой кислоты. Структуры известны для представителей обоих семейств, и они демонстрируют уникальную общую укладку, содержащую до девяти трансмембранных областей. Остатки аспарагиновой кислоты активного центра расположены внутри большой полости в мембране, в которую может попасть вода. * Клан AE содержит два семейства - A25 и A31. Третичные структуры были расшифрованы для представителей обоих семейств и демонстрируют общую укладку, состоящую из альфа-бета-альфа-"сэндвича", в котором бета-лист состоит из пяти нитей. * Клан AF содержит одно семейство A26. Представители клана представляют собой мембранные белки с уникальной укладкой. Гомологи известны только у бактерий. Структура омптина (также известного как OmpT) представляет собой цилиндрическую бочку, содержащую десять бета-нитей, встроенных в мембрану, с остатками активного центра на наружной поверхности. * Существует два семейства аспарагиновых пептидаз, для которых не известны ни структура, ни остатки в активном центре, и они не отнесены к какому-либо клану. Семейство A5 включает термопсин - эндопептидазу, обнаруживаемую только у термофильных архей. Семейство A36 включает фактор спорообразования SpoIIGA, который, как известно, подвергает процессингу и активирует сигма-фактор E - один из факторов

			транскрипции, который контролирует спорообразование у бактерий. Это суперсемейство доменов обнаруживается в аспарагиновых пептидазах, включая пепсин A и другие представители семейства пептидаз A1, индуцируемый повреждением ДНК белок 1 дрожжей (Ddi1) (подсемейство пептидаз A28A по MEROPS), аспарагиновую протеазу животных, подобную ретровирусной (SASPазу; подсемейство пептидаз A28B по MEROPS), и пептидазу PerP Caulobacter (семейство пептидаз A32 по MEROPS). Третичная структура представляет собой ретропепсиноподобную укладку, а пептидаза активна в качестве гомодимера."
203	IPR001969	"Аспарагиновая пептидаза, активный центр"	"Аспарагиновые пептидазы, также известные как аспартилпротеазы ([IntEnz:3.4.23.-]), представляют собой широко распространенные протеолитические ферменты, которые, как известно, существуют у позвоночных, грибов, растений, простейших, бактерий, архей, ретровирусов и некоторых вирусов растений. Все известные аспарагиновые пептидазы являются эндопептидазами. Молекула воды, активированная двумя остатками аспарагиновой кислоты, выступает в качестве нуклеофила в катализе. Аспарагиновые пептидазы можно разделить на пять кланов, каждый из которых демонстрирует уникальную структурную укладку. * Пептидазы в клане AA являются либо двухдольными (семейство A1 или семейство пепсина), либо гомодимерными (все остальные семейства в клане, включая ретропепсин из HIV-1/AIDS). Каждая доля состоит из одного домена с замкнутой бета-"бочкой", и каждая доля привносит один Asp для формирования активного центра. Большинство пептидаз в клане ингибируются встречающимся в природе низкомолекулярным

			ингибитором пепстатином. * Клан AC содержит единственное семейство A8: семейство сигнальной пептидазы 2. Представители семейства обнаруживаются у всех бактерий. Сигнальная пептидаза 2 подвергает процессингу предшественник премуреина, удаляя сигнальный пептид. Пептидаза имеет четыре трансмембранных домена, а активный центр находится на периплазматической стороне клеточной мембраны. Расщепление происходит со стороны аминогруппы цистеина, где тиольная группа была замещена диацилглицерильной группой. В ходе сайт-направленного мутагенеза были идентифицированы два незаменимых остатка аспарагиновой кислоты, которые встречаются в мотивах GNXXDRX и FNXAD (где X представляет собой гидрофобный остаток). Третичные структуры не были расшифрованы ни для одного представителя семейства, но из-за внутримембранного расположения предполагается, что структура не является пепсиноподобной. * Клан AD содержит два семейства трансмембранных эндопептидаз: A22 и A24. Они также известны как "пептидазы GXGD" ввиду наличия общего мотива GXGD, который содержит один из пары каталитических остатков аспарагиновой кислоты. Структуры известны для представителей обоих семейств, и они демонстрируют уникальную общую укладку, содержащую до девяти трансмембранных областей. Остатки аспарагиновой кислоты активного центра расположены внутри большой полости в мембране, в которую может попасть вода. * Клан AE содержит два семейства - A25 и A31. Третичные структуры были расшифрованы для представителей обоих семейств и демонстрируют общую

			укладку, состоящую из альфа-бета-альфа-"сэндвича", в котором бета-лист состоит из пяти нитей. * Клан AF содержит одно семейство A26. Представители клана представляют собой мембранные белки с уникальной укладкой. Гомологи известны только у бактерий. Структура омптина (также известного как OmpT) представляет собой цилиндрическую бочку, содержащую десять бета-нитей, встроенных в мембрану, с остатками активного центра на наружной поверхности. * Существует два семейства аспарагиновых пептидаз, для которых не известны ни структура, ни остатки в активном центре, и они не отнесены к какому-либо клану. Семейство A5 включает термопсин - эндопептидазу, обнаруживаемую только у термофильных архей. Семейство A36 включает фактор спорообразования SpoIIGA, который, как известно, подвергает процессингу и активирует сигма-фактор E - один из факторов транскрипции, который контролирует спорообразование у бактерий. Эта сигнатура содержит остатки активного центра, которые являются консервативными в эукариотических и вирусных аспартилпротеазах. Аспарагиновая пептидаза, также известная как аспартилпротеазы ([IntEnz:3.4.23.-]), представляет собой широко распространенное семейство протеолитических ферментов, которые, как известно, существуют у позвоночных, грибов, растений, ретровирусов и некоторых вирусов растений. Аспартатпротеазы эукариот представляют собой мономерные ферменты, состоящие из двух доменов. Каждый домен содержит активный центр, в котором каталитический аспартильный остаток является центральным. Эти два домена наиболее вероятно произошли в ходе эволюции вследствие дупликации предкового

			гена, кодирующего примордиальный домен. В настоящее время известны следующие эукариотические аспартилпротеазы. * Желудочные пепсины A и C позвоночных (также известные как гастриксин). Химозин позвоночных (реннин), участвующий в пищеварении и используемый для производства сыра. * Лизосомальные катепсины D (EC 3.4.23.5) и E (EC 3.4.23.34) позвоночных. * Ренин млекопитающих (EC 3.4.23.15), функция которого заключается в образовании ангиотензина I из ангиотензиногена в плазме крови. * Протеазы грибов, такие как аспергиллопепсин A (EC 3.4.23.18), кандидапепсин (EC 3.4.23.24), мукоропепсин (EC 3.4.23.23) (реннин мукора), эндотиапепсин (EC 3.4.23.22), полипоропепсин (EC 3.4.23.29) и ризопуспепсин (EC 3.4.23.21). * Сахаропепсин дрожжей (EC 3.4.23.25) (протеиназа A) (ген PEP4). PEP4 участвует в посттрансляционной регуляции вакуолярных гидролаз. * Барьерный пепсин дрожжей (EC 3.4.23.35) (ген BAR1); протеаза, которая расщепляет альфа-фактор и, таким образом, выступает в качестве антагониста феромона спаривания. * sxa1 делящихся дрожжей, который участвует в разрушении или процессинге феромонов спаривания."
204	cd00303	Ретропепсины; пепсиноподобные аспартатпротеазы	"Семейство включает пепсиноподобные аспартатпротеазы из ретровирусов, ретротранспозонов и ретроэлементов, а также эукариотические белки, индуцируемые повреждением ДНК (DDI), и бактериальные аспартатпептидазы. Тогда как пепсины грибов и млекопитающих представляют собой двухдольные белки со структурно родственными N- и C-концами, ретропепсины вдвое короче, чем их аналоги у грибов и млекопитающих. Мономеры структурно связаны с одной

			долей молекулы пепсина, а ретропепсины функционируют как гомодимеры. Аспартат активного центра встречается в мотиве (Asp-Thr/Ser-Gly), как и в пепсине. Ретровирусная аспартилпротеаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. В аспартатпептидазах остатки Asp являются лигандами активированной молекулы воды во всех примерах, где были идентифицированы каталитические остатки. Эта группа аспартатпептидаз классифицируется по MEROPS как семейство пептидаз A2 (семейство ретропепсинов, клан AA), подсемейство A2A. [PMID: 1851433, PMID: 17199106, PMID: 2194475, PMID: 7674916, PMID: 1455179, PMID: 8841139, PMID: 16970402]"
205	PS50035	Профиль активного центра фосфодиэстеразы фосфолипазы D	"Фосфолипаза D (PLD) представляет собой повсеместно распространенный фермент, обнаруживаемый в бактериях, грибах, растениях и млекопитающих. PLD участвует в формировании везикул, транспорте белков, передаче сигналов и митозе. Он катализирует распад фосфатидилхолина на холин и предполагаемый вторичный мессенджер фосфатидную кислоту, которая, в свою очередь, распадается на два вторичных мессенджера (диацилглицерин и лизофосфатидную кислоту). PLD также катализирует реакцию переноса фосфатидила с использованием первичных спиртов в качестве нуклеофильных акцепторов с образованием фосфатидиловых спиртов. Анализ последовательности выявил, что PLD принадлежит к суперсемейству, которое включает кардиолипинсинтазы, фосфатидилсеринсинтазу, белки оболочки

			поксвирусов, мышиный токсин Yersinia и несколько эндонуклеаз. Все представители этого суперсемейства фосфодиэстераз содержат мотив последовательности HxKx(4)-Dx(6)-GSxN, обозначенный как "HKD". Несмотря на различную субстратную специфичность представителей суперсемейства, консенсусный мотив HKD, по-видимому, важен для их ферментативной активности. Большинство ферментов суперсемейства содержат две копии этой консенсусной последовательности, но бактериальная эндонуклеаза содержит только одну копию [PMID:8669123,PMID:8732763,PMID:8755242]. Предполагается, что два мотива HKD ассоциируют с образованием одного активного центра. Было высказано предположение, что при связывании субстрата один из остатков гистидина функционирует как нуклеофил, атакуя фосфодиэфирную связь, а другой гистидин может выступать в качестве общей кислоты, протонируя атом кислорода уходящей группы [PMID:10074947,PMID:10825182]. В профиле, который разработали авторы настоящего изобретения, основное внимание уделяется активному центру фосфодиэстеразы фосфолипазы D."
206	IPR001736	Фосфолипаза D/трансфосфатидилаза	"Изоформы фосфолипазы D, гидролизующей фосфатидилхолин (PLD), активируются факторами ADP-рибозилирования (ARF). PLD обеспечивает продуцирование фосфатидной кислоты из фосфатидилхолина, которая может быть необходима для образования определенных типов транспортных везикул или может обеспечивать конститутивный везикулярный транспорт к путям передачи сигнала. PLD, гидролизующая PC, является гомологом кардиолипинсинтазы,

			фосфатидилсеринсинтазы, бактериальных PLD и вирусных белков. Каждый из этих ферментов, по-видимому, содержит дупликацию домена, что видно по наличию двух мотивов, содержащих высококонсервативные остатки гистидина, лизина и/или аспарагина, которые могут привносить аспарагиновую кислоту в активный центр. Эндонуклеаза Escherichia coli (nuc) и сходные белки, по-видимому, являются гомологами PLD, но содержат только один из этих мотивов."
207	PS50020	Профиль домена WW/rsp5/WWP	"Домен WW [[PMID:7846762,PMID:7802651,PMID:7828727,PMID:7641887] (также известный как rsp5 или WWP) был первоначально обнаружен как короткая консервативная область в ряде неродственных белков, среди которых был дистрофин, кодируемый геном, отвечающим за мышечную дистрофию Дюшенна. Домен, который охватывает приблизительно 35 остатков, в некоторых белках повторяется до 4 раз. Было показано [PMID:7644498], что он связывает белки с определенными пролиновыми мотивами [AP]-PP-[AP]-Y и, таким образом, несколько напоминает домены SH3. По-видимому, он содержит бета-нити, сгруппированные вокруг четырех положений консервативных ароматических аминокислот; как правило, Trp. Название WW или WWP происходит от присутствия этих Trp, а также от консервативного Pro. Он часто ассоциирован с другими доменами, типичными для белков, участвующих в процессах передачи сигналов. Белки, содержащие домен WW, перечислены ниже. - Дистрофин - многодоменный белок цитоскелета. Его самая длинная полученная в результате альтернативного сплайсинга форма состоит из N-концевого актинсвязывающего домена, за которым следуют 24

			спектриноподобных повтора, богатый цистеином кальцийсвязывающий домен и С-концевой глобулярный домен. Дистрофины образуют тетрамеры и, как полагают, выполняют множество функций, включая участие в обеспечении стабильности мембран, передачу сократительных сил во внеклеточную среду и организацию специализации мембран. Мутации в гене дистрофина приводят к мышечной дистрофии типа Дюшенна или Беккера. Дистрофин содержит один домен WW с С-концевой стороны от спектриновых повторов. - Утрофин - дистрофиноподобный белок с неизвестной функцией. - Белок YAP позвоночных является субстратом неизвестной сериновой киназы. Он связывается с доменом SH3 онкопротеина Yes с помощью богатой пролином области. Этот белок появляется в полученных в результате альтернативного сплайсинга изоформах, содержащих один либо два домена WW [PMID:7782338]. - NEDD-4 мыши играет роль в эмбриональном развитии и дифференцировке центральной нервной системы. Он содержит 3 модуля WW, за которыми следует домен HECT. Ортолог человека содержит 4 домена WW, но третий домен WW, вероятно, подвергается сплайсингу, что приводит в результате к альтернативному белку NEDD-4 только с 3 модулями WW [PMID:7828727]. - RSP5 дрожжей сходен с NEDD-4 по своей молекулярной организации. Он содержит N-концевой домен C2. Родственный белок dodo (ген dod) существует у дрозофилы и млекопитающих (ген PIN1). - Белок DB10 табака. Домен WW расположен с N-концевой стороны от области, сходной с АТР-зависимыми РНК-хеликазами. - IQGAP - белок, активирующий GTPазу человека, действующий на ras. Он содержит N-

			концевой домен, сходный с белком mp20 мышц мухи, и С-концевой домен, активирующий GTPазу ras. - Белок процессинга пре-мРНК PRP40 дрожжей, ZK1098.1 Caenorhabditis elegans и SpAC13C5.02 делящихся дрожжей являются родственными белками со сходством с миозином типа MYO2, каждый из которых содержит два домена WW на N-конце. - Гипотетический белок C38D4.5 Caenorhabditis elegans, который содержит один модуль WW, домен PH и С-концевой фосфатидилинозитол-3-киназный домен. - Гипотетический белок YFL010c дрожжей. Для чувствительного выявления доменов WW авторы настоящего изобретения разработали профиль, который охватывает всю область гомологии, а также паттерн."
208	IPR001202	Домен WW	"Синоним(синонимы): домен Rsp5 или WWP. Домен WW представляет собой короткую консервативную область в ряде неродственных белков, которая сворачивается в виде стабильного трехнитевого бета-листа. Этот короткий домен, состоящий из примерно 40 аминокислот, может повторяться до четырех раз в некоторых белках [PUB00005431],[PUB00000239],[PUB00001680],[PUB00001690]. Название WW или WWP происходит от присутствия двух сигнатурных остатков триптофана, которые расположены на расстоянии 20-23 аминокислот друг от друга и присутствуют в большинстве известных на сегодняшний день доменов WW, а также от присутствия консервативного Pro. Домен WW связывается с белками с определенными пролиновыми мотивами [AP]-PP-[AP]-Y и/или мотивами, содержащими фосфосерин/фосфотреонин. Он часто ассоциирован с другими доменами, типичными для белков в процессах передачи сигналов. Известно

			большое разнообразие белков, содержащих домен WW. Они включают следующие: дистрофин - многодоменный белок цитоскелета; утрофин - дистрофиноподобный белок с неизвестной функцией; белок YAP позвоночных - субстрат неизвестной сериновой киназы; NEDD-4 Mus musculus (мыши), участвующий в эмбриональном развитии и дифференцировке центральной нервной системы; RSP5 Saccharomyces cerevisiae (пекарских дрожжей), сходный с NEDD-4 по своей молекулярной организации; FE65 Rattus norvegicus (крысы) - активатор фактора транскрипции, преимущественно экспрессирующийся в печени; белок DB10 Nicotiana tabacum (табака обыкновенного), помимо прочего."
209	IPR036020	Суперсемейство доменов WW	"Синоним(синонимы): домен Rsp5 или WWP. Домен WW представляет собой короткую консервативную область в ряде неродственных белков, которая сворачивается в виде стабильного трехнитевого бета-листа. Этот короткий домен, состоящий из примерно 40 аминокислот, может повторяться до четырех раз в некоторых белках [PUB00005431],[PUB00000239],[PUB00001680],[PUB00001690]. Название WW или WWP происходит от присутствия двух сигнатурных остатков триптофана, которые расположены на расстоянии 20-23 аминокислот друг от друга и присутствуют в большинстве известных на сегодняшний день доменов WW, а также от присутствия консервативного Pro. Домен WW связывается с белками с определенными пролиновыми мотивами [AP]-PP-[AP]-Y и/или мотивами, содержащими фосфосерин/фосфотреонин. Он часто ассоциирован с другими доменами, типичными для белков в процессах передачи сигналов. Известно

			большое разнообразие белков, содержащих домен WW. Они включают следующие: дистрофин - многодоменный белок цитоскелета; утрофин - дистрофиноподобный белок с неизвестной функцией; белок YAP позвоночных - субстрат неизвестной сериновой киназы; NEDD-4 Mus musculus (мыши), участвующий в эмбриональном развитии и дифференцировке центральной нервной системы; RSP5 Saccharomyces cerevisiae (пекарских дрожжей), сходный с NEDD-4 по своей молекулярной организации; FE65 Rattus norvegicus (крысы) - активатор фактора транскрипции, преимущественно экспрессирующийся в печени; белок DB10 Nicotiana tabacum (табака обыкновенного), помимо прочего."
210	cd00201	Домен с двумя консервативными остатками триптофана, также известный как домен WWP или rsp5	"Домен с двумя консервативными остатками триптофана, также известный как домен WWP или rsp5; состоящий из примерно 40 аминокислот; функционирует в качестве взаимодействующего модуля в разнообразном наборе сигнальных белков; связывает определенные богатые пролином последовательности, но с низкой аффинностью по сравнению с другими белками, распознающими пептиды, такими как антитела и рецепторы; домены WW имеют единственную бороздку, образованную консервативными Trp и Tyr, которая распознает пару остатков последовательности X-Pro; вариабельные петли и соседние домены придают специфичность этому домену; существует пять различных групп, выделяемых на основании связывания: 1) мотивы PPXY; 2) мотив PPLP; 3) мотивы PGM; 4) мотивы PSP или PTP; 5) мотивы PR. [PMID: 9401204, PMID: 9407065, PMID: 9305852, PMID: 8757138, PMID: 7828727, PMID: 7782338, PMID: 7802651, PMID: 7846762]"

211	PF00397	Домен WW	Домен WW представляет собой белковый модуль с двумя высококонсервативными остатками триптофана, который связывает богатые пролином пептидные мотивы in vitro.
212	IPR001368	Богатая цистеином область TNFR/NGFR	"Было обнаружено, что ряд белков, некоторые из которых, как известно, являются рецепторами факторов роста, содержит в своей N-концевой части богатый цистеином домен, состоящий из приблизительно 110-160 аминокислот, который можно разделить на четыре (или в некоторых случаях три) модуля, состоящих из приблизительно 40 остатков, содержащих 6 консервативных остатков цистеина. Некоторые из белков, содержащих этот домен, перечислены ниже. * Рецепторы фактора некроза опухоли типа I и типа II (TNFR). Оба рецептора связывают TNF-альфа и TNF-бета, но сходны только в богатой цистеином области. TNFR содержит четыре модуля, представляющих собой богатые цистеином домены (CRD), обозначенных как CRD1-CRD4. CRD2 и CRD3 известны как TNF-связывающие домены. * Растворимый рецептор TNF вируса фибромы Шоупа (белок T2) * Рецептор лимфотоксина-альфа/бета * Низкоаффинный рецептор фактора роста нервов (LA-NGFR) (p75) * CD40 (Bp50) - рецептор цитокина CD40L (или TRAP) * CD27 - рецептор цитокина CD27L * CD30 - рецептор цитокина CD30L * T-клеточный белок 4-1BB - рецептор предполагаемого цитокина 4-1BBL * Антиген FAS (или APO-1) - рецептор FASL, белка, участвующего в апоптозе (запрограммированной гибели клеток) * T-клеточный антиген OX40 - рецептор цитокина OX40L * Wsl-1 - рецептор (еще не определенного лиганда), который опосредует апоптоз * Белок A53 вируса

			осповакцины (SalF19R). Было показано, что все шесть остатков цистеина участвуют в образовании внутрицепочечных дисульфидных связей. Схематическое изображение структуры модуля этих рецепторов, состоящего из 40 остатков, показано ниже: +---------+ +--------------+ | | | xCxxxxxxxxxxxxxCxCxxCxxxxxxxxxCxxxxCxx | | +------------+'C': консервативный остаток цистеина, участвующий в образовании дисульфидной связи."
213	PS00652	Предполагаемая нуклеаза HARBI1	NA
214	cd00890	"Префолдин представляет собой гексамерный молекулярный шапероновый комплекс, обнаруживаемый как у эукариот, так и у архей"	"Префолдин представляет собой гексамерный молекулярный шапероновый комплекс, обнаруживаемый как у эукариот, так и у архей, который связывает и стабилизирует вновь синтезированные полипептиды, обеспечивая возможность их правильного сворачивания. Комплекс содержит две альфа- и четыре бета-субъединицы, и эти две субъединицы являются эволюционно родственными. У архей обычно имеется только один ген для каждой субъединицы, тогда как у эукариот имеется два или более паралогичных гена, кодирующих каждую субъединицу, что добавляет гетерогенность в структуру гексамера. Структура комплекса состоит из сборки в виде двойной бета-"бочки" с шестью выступающими суперспиралями. [PMID: 11884745, PMID: 9630229, PMID: 10550210, PMID: 11171552, PMID: 12456645]"
215	cd17756	Фактор лицензирования репликации ДНК Mcm5	"Mcm5 представляет собой хеликазу, играющую важную роль в репликации. Он является частью гетерогексамерного кольцеобразного комплекса Mcm2-7, который

			является частью репликативной хеликазы, которая раскручивает исходную двухнитевую ДНК на репликативной вилке с получением однонитевых ДНК-матриц для репликативных полимераз. [PMID: 8332451, PMID: 28383499, PMID: 22918583, PMID: 26222030]"
216	cd17059	Убиквитин-подобный (Ubl) домен, обнаруживаемый в убиквитинтиоэстеразе OTU1 и сходных белках	"OTU1 (EC 3.4.19.12), также называемый YOD1, или DUBA-8, или HIV-1-индуцируемой протеазой 7 (HIN-7), или белком 2, содержащим домен OTU (OTUD2), представляет собой p97-ассоциированную деубиквитинилазу, которая функционирует в качестве ключевого участника разрушения, ассоциированного с эндоплазматическим ретикулумом (ERAD). Ее деубиквитинилазная активность также необходима для отрицательной регуляции ретротранслокации холерного токсина A1 (CTA1). OTU1 содержит консервативный убиквитин-подобный (Ubl) домен с укладкой типа бета-"захвата" Ubl, "цинковым пальцем" типа C2H2 и доменом OTU. [PMID: 19818707, PMID: 23827681, PMID: 26463207, PMID: 25747718, PMID: 24068323, PMID: 24610782]"
217	cd17218	Ras-ассоциирующий (RA) домен, обнаруживаемый в белке 1, содержащем Ras-ассоциирующий домен (RASSF1)	"RASSF1 является представителем семейства из шести родственных белков RASSF1-6 (классические белки RASSF). RASSF1 имеет восемь транскриптов (A-H), возникающих в результате альтернативного сплайсинга и использования различных промоторов. За исключением некоторых минорных сплайс-вариантов (RASSF1F и RASSF1G), RASSF1 содержит RA-домен и C-концевой мотив белок-белкового взаимодействия SARAH. Домен RA классических белков RASSF имеет убиквитин-подобную укладку типа бета-"захвата" с низким сходством последовательности с убиквитином. Домен RA опосредует взаимодействия с Ras и другими небольшими GTPазами, а домен

			SARAH опосредует белок-белковые взаимодействия, имеющие решающее значение в путях, индуцирующих остановку клеточного цикла и апоптоз. RASSF1A и 1C являются наиболее широко изученными RASSF1, оба из которых локализованы в микротрубочках и участвуют в регуляции роста и миграции. [PMID: 10888881, PMID: 11333291, PMID: 12024041, PMID: 11857081, PMID: 15205320, PMID: 14729613, PMID: 14743218, PMID: 15109305, PMID: 15753381, PMID: 21134643, PMID: 21199877, PMID: 18566590, PMID: 16510573, PMID: 15949439, PMID: 17692468]"
218	IPR025913	Домен центросомной локализации Cep57	"CLD или домен центросомной локализации белка Cep57 находится на N-конце и расположен примерно между остатками 58 и 239. Эта область расположена внутри первого альфа-спирального суперспирального сегмента Cep57 и обеспечивает локализацию в центросоме с внутренним по отношению к гамма-тубулину расположением, что позволяет предположить, что она находится либо в обеих центриолях, либо в центроматриксном компоненте. Эта N-концевая область может также мультимеризоваться с N-концом других молекул Cep57."
219	PF14073	Домен центросомной локализации Cep57	"CLD или домен центросомной локализации белка Cep57 находится на N-конце и расположен примерно между остатками 58 и 239. Эта область расположена внутри первого альфа-спирального суперспирального сегмента Cep57 и обеспечивает локализацию в центросоме с внутренним по отношению к гамма-тубулину расположением, что позволяет предположить, что она находится либо в обеих центриолях, либо в центроматриксном компоненте. Эта N-концевая область может также мультимеризоваться с N-концом других молекул Cep57. C-концевая часть,

			семейство Cep57_MT_bd, Pfam:PF06657, представляет собой область связывания микротрубочек Cep57."
220	IPR011010	"Фермент, разрывающий и вновь соединяющий ДНК, каталитическая сердцевина"	"Фаговые интегразы представляют собой ферменты, которые опосредуют однонаправленную сайт-специфическую рекомбинацию между двумя последовательностями распознавания ДНК - фаговым сайтом прикрепления attP и бактериальным сайтом прикрепления attB. Интегразы можно разделить на два основных семейства - тирозиновые рекомбиназы и сериновые рекомбиназы - в зависимости от их способа катализа. Интегразы тирозинового семейства, такие как интеграза бактериофага лямбда, используют каталитический тирозин для опосредования расщепления нити, склонны распознавать более длинные последовательности attP и требуют других белков, кодируемых фагом- или бактерией-хозяином. Интеграза лямбда, состоящая из 356 аминокислот, состоит из двух доменов: N-концевого домена, который содержит остатки 1-64 и отвечает за связывание сайтов attP плечевого типа, и C-концевого домена (CTD), который связывает сайты сердцевинного типа с более низкой аффинностью и содержит каталитический центр. CTD можно дополнительно разделить на связывающий домен сердцевинного типа (остатки 65-169) и каталитический сердцевинный домен (170-356), последний из которых представляет эту запись. Каталитический сердцевинный домен принимает укладку альфа3-бета3-альфа4, где одна сторона бета-листа является доступной. Рекомбиназы Cre из фага P1, XerD из Escherichia coli и Flp из дрожжей являются представителями семейства тирозиновых рекомбиназ и

			имеют двухдоменный мотив, напоминающий мотив интегразы лямбда, а также имеют общий консервативный механизм связывания. Структурная укладка их каталитических сердцевинных доменов напоминает структуру интегразы лямбда. Каталитическая сердцевина эукариотической ДНК-топоизомеразы I обладает значительным структурным сходством с семейством ДНК-интеграз бактериофагов. Топоизомеразы I способствуют ослаблению напряжения суперспирали ДНК, внося временный однонитевой разрыв в ДНК-дуплекс, и являются жизненно важными для процессов репликации, транскрипции и рекомбинации."
221	cd09275	Семейство DIRS1 РНКазы HI в ретроэлементах с длинными концевыми повторами	"Ферменты рибонуклеазы H (РНКазы H) делятся на два основных семейства - тип 1 и тип 2 - на основании сходства аминокислотных последовательностей и биохимических свойств. РНКаза H представляет собой эндонуклеазу, которая расщепляет нить РНК гибрида РНК/ДНК неспецифичным в отношении последовательности образом в присутствии двухвалентных катионов. РНКаза H широко представлена в различных организмах, включая бактерии, археи и эукариот. РНКаза HI также наблюдалась в качестве доменов, прилегающих к гену обратной транскриптазы в ретровирусах, содержащих длинные концевые повторы (LTR) ретротранспозонах и не содержащих LTR ретротранспозонах. РНКаза HI в содержащих LTR ретротранспозонах осуществляет разрушение исходной РНК-матрицы, образование полипуринового тракта (праймера для синтеза плюс-нити ДНК) и окончательное удаление РНК-праймеров из вновь синтезированных минус- и плюс-нитей. Каталитические остатки для ферментативной активности РНКазы H - три

			остатка аспарагиновой кислоты и один остаток глутаминовой кислоты (DEDD) - неизменны во всех доменах РНКазы H. Филогенетические модели РНКазы HI содержащих LTR ретроэлементов подразделяются на пять основных семейств: Ty3/Gypsy, Ty1/Copia, Bel/Pao, DIRS1 и ретровирусы позвоночных. Структурные признаки элементов группы DIRS1 отличаются от типичных содержащих LTR элементов. Ингибиторы РНКазы H исследовались в качестве мишени для лекарственных средств для лечения HIV, поскольку инактивация РНКазы H ингибирует обратную транскрипцию. [PMID: 19228198, PMID: 11435400, PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265]"
222	IPR013762	"Суперсемейство интеразоподобных каталитических доменов"	"Фаговые интегразы представляют собой ферменты, которые опосредуют однонаправленную сайт-специфическую рекомбинацию между двумя последовательностями распознавания ДНК - фаговым сайтом прикрепления attP и бактериальным сайтом прикрепления attB. Интегразы можно разделить на два основных семейства - тирозиновые рекомбиназы и сериновые рекомбиназы - в зависимости от их способа катализа. Интегразы тирозинового семейства, такие как интеграза лямбда, используют каталитический тирозин для опосредования расщепления нити, склонны распознавать более длинные последовательности attP и требуют других белков, кодируемых фагом- или бактерией-хозяином. Интеграза лямбда, состоящая из 356 аминокислот, состоит из двух доменов: N-концевого домена, который содержит остатки 1-64 и отвечает за связывание сайтов attP плечевого типа, и C-концевого домена (CTD), который связывает сайты

			сердцевинного типа с более низкой аффинностью и содержит каталитический центр. CTD можно дополнительно разделить на связывающий домен сердцевинного типа (остатки 65-169) и каталитический сердцевинный домен (170-356), последний из которых представляет эту запись. Каталитический сердцевинный домен принимает укладку альфа3-бета3-альфа4, где одна сторона бета-листа является доступной. Рекомбиназы Cre из фага P1, XerD из Escherichia coli и Flp из дрожжей являются представителями семейства тирозиновых рекомбиназ и имеют двухдоменный мотив, напоминающий мотив интегразы лямбда, а также имеют общий консервативный механизм связывания. Структурная укладка их каталитических сердцевинных доменов напоминает структуру интегразы лямбда."
223	PF13495	"Фаговая интеграза, N-концевой SAM-подобный домен"	NA
224	IPR004107	"Интеграза, SAM-подобная, N-концевая"	"Белки, содержащие этот домен, расщепляют ДНК-субстраты посредством внесения серии ступенчатых разрезов, во время которого белок ковалентно связывается с ДНК через каталитический остаток тирозина на карбокси-конце выравнивания. N-концевой SAM-подобный домен фаговой интегразы почти всегда обнаруживается с сигнатурой, определяющей семейство фаговых интеграз (см.)."
225	cd03714	RT_DIRS1: Обратные транскриптазы (RT),	"RT_DIRS1: Обратные транскриптазы (RT), встречающиеся в группе ретротранспозонов DIRS1. Представители подсемейства включают элементы DIRS-1 Dictyostelium, kangaroo Volvox carteri и PAT Panagrellus redivivus. Эти элементы

		встречающиеся в группе ретротранспозонов DIRS1.	отличаются от содержащих LTR и традиционных не содержащих LTR ретротранспозонов. Они содержат концы с разделенными прямыми повторами (SDR), и было предложено интегрировать их через промежуточные соединения двухнитевой замкнутой кольцевой ДНК с помощью кодируемой рекомбиназы, которая сходна с сайт-специфической интегразой гамма. [PMID: 11606703, PMID: 12524337, PMID: 2175908, PMID: 1318387, PMID: 1315955, PMID: 1698615, PMID: 8828137]"
226	IPR010998	"Интеграза/рекомбиназа, N-концевая"	"Фаговые интегразы представляют собой ферменты, которые опосредуют однонаправленную сайт-специфическую рекомбинацию между двумя последовательностями распознавания ДНК - фаговым сайтом прикрепления attP и бактериальным сайтом прикрепления attB. Интегразы можно разделить на два основных семейства - тирозиновые рекомбиназы и сериновые рекомбиназы - в зависимости от их способа катализа. Интегразы тирозинового семейства, такие как интеграза бактериофага лямбда, используют каталитический тирозин для опосредования расщепления нити, склонны распознавать более длинные последовательности attP и требуют других белков, кодируемых фагом- или бактерией-хозяином. Интеграза лямбда, состоящая из 356 аминокислот, состоит из двух доменов: N-концевого домена (NTD), который содержит остатки 1-64 и отвечает за связывание сайтов attP плечевого типа, и С-концевого домена, который связывает сайты сердцевинного типа с более низкой аффинностью и содержит каталитический центр. NTD принимает укладку пучка из 4-5 спиралей с двумя ортогонально

			упакованными альфа-шпильками. Рекомбиназы Cre из фага P1, XerD из Escherichia coli и Flp из дрожжей являются представителями семейства тирозиновых рекомбиназ и имеют двухдоменный мотив, напоминающий мотив интегразы лямбда, а также имеют общий консервативный механизм связывания. NTD фаговой интегразы почти всегда обнаруживается с сигнатурой, определяющей семейство фаговых интеграз (см.)."
227	IPR024064	Суперсемейство FdhE-подобных белков	"Это суперсемейство представляет домен, обнаруживаемый во вспомогательном белке формиатдегидрогеназы FdhE и его гомологах. Он состоит из N-концевого пятиспирального пучка, тандемного повтора из двух рубредоксиноподобных доменов и С-концевой спирали, упакованных вместе."
228	IPR000727	Суперспиральный домен гомологии SNARE мишени	"Процесс слияния везикул с мембранами-мишенями зависит от набора SNARE (SNAP-рецепторов), которые ассоциированы со сливающимися мембранами. SNARE мишени (t-SNARE) локализованы на мембране-мишени и принадлежат к двум разным семействам: семейству синтаксиноподобных белков и семейству SNAP-25-подобных белков. Для слияния требуется один представитель каждого семейства вместе с v-SNARE, локализованным на везикулярной мембране. N- и С-концевые суперспиральные домены представителей семейства SNAP-25 и наиболее близкий к С-концу суперспиральный домен семейства синтаксинов родственны друг другу и образуют новый домен гомологии, состоящий из примерно 60 аминокислот. Этот домен также обнаруживается в других известных белках, участвующих в

			везикулярном мембранном транспорте, некоторые из которых принадлежат к разным семействам белков."
229	PS50192	Профиль суперспирального домена гомологии t-SNARE	"Процесс слияния везикулярных мембран в эукариотических клетках зависит от консервативного аппарата слияния, называемого SNARE (рецепторы белка прикрепления растворимого фактора, чувствительного к N-этилмалеимиду (NSF) (SNAP)). В процессе стыковки везикул белки, присутствующие на везикулах (v-SNARE), должны связываться со своими аналогами на мембране-мишени (t-SNARE), чтобы образовать коровый комплекс, который затем может привлекать растворимые белки NSF и SNAP. Этот так называемый комплекс слияния может затем распадаться после гидролиза ATP, опосредованного ATPазой NSF, в процессе, который приводит к слиянию мембран и высвобождению содержимого везикул. t-SNARE состоят из двух разных семейств белков: интегральных мембранных белков типа II синтаксинов и SNAP-25 (ассоциированного с синаптосомами белка размером 25 кДа), который заякорен в плазматической мембране присоединенными липидами и не пересекает мембрану [PMID:9096343]. N- и C-концевые суперспиральные домены представителей семейства SNAP-25 и наиболее близкий к C-концу суперспиральный домен семейства синтаксинов родственны друг другу и образуют новый домен гомологии, состоящий из примерно 60 аминокислот. Этот домен также обнаруживается в других известных белках, участвующих в везикулярном мембранном транспорте, некоторые из которых принадлежат к разным семействам белков [PMID:9096343]: - Белки SFT1, UFE1, BET1, VAM7 и

			YDR468c дрожжей. - Белки TSL-4 млекопитающих. Разработанный авторами настоящего изобретения профиль охватывает весь суперспиральный домен гомологии t-SNARE."
230	cd09279	"Семейство РНКаз HI, которое включает архейные, некоторые бактериальные, а также растительные РНКазы HI"	"Рибонуклеаза H (РНКаза H) подразделяется на два эволюционно неродственных семейства: тип 1 (прокариотическая РНКаза HI, эукариотическая РНКаза H1 и вирусная РНКаза H) и тип 2 (прокариотическая РНКаза HII и HIII и эукариотическая РНКаза H2). РНКаза H представляет собой эндонуклеазу, которая расщепляет нить РНК гибрида РНК/ДНК неспецифичным в отношении последовательности образом. РНКаза H участвует в репликации, репарации и транскрипции ДНК. РНКаза H широко представлена в различных организмах, включая бактерии, археи и эукариот, и большинство прокариотических и эукариотических геномов содержат несколько генов РНКазы H. Несмотря на отсутствие гомологии аминокислотных последовательностей, РНКазы H типа 1 и типа 2 характеризуются общей укладкой основной цепи и стерической конфигурацией четырех кислотных остатков активного центра (DEDD) и характеризуются одинаковыми каталитическим механизмом и функциями в клетках. Одной из важных функций РНКазы H является удаление фрагментов Оказаки во время репликации ДНК. Геномы большинства архей содержат только РНКазу H типа 2 (РНКазу HII); однако некоторые из них также содержат РНКазу HI. Хотя в последовательностях архейной РНКазы HI сохраняется мотив активного центра DEDD, в них отсутствуют другие общие признаки, важные для каталитической функции, таких как основная область выступа. Гомологи

			архейной РНКазы HI являются более близкородственными ретровирусной РНКазе HI, чем бактериальной и эукариотической РНКазе H типа I, по ферментативным свойствам. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198]"
231	IPR001509	NAD-зависимая эпимераза/дегидратаза	Этот домен обнаруживается в белках, которые используют NAD в качестве кофактора и используют субстраты, представляющие собой нуклеотидные сахара, для различных химических реакций. Одним из наиболее изученных из этих белков является UDP-галактозо-4-эпимераза, которая катализирует превращение UDP-галактозы в UDP-глюкозу в ходе метаболизма галактозы.
232	IPR005886	UDP-глюкозо-4-эпимераза	"Синоним: UDP-галактозо-4-эпимераза (UDP-глюкозо-4-эпимераза) обеспечивает взаимопревращение UDP-глюкозы и UDP-галактозы, которые являются предшественниками глюкозо- и галактозосодержащих экзополисахаридов (EPS). Arabidopsis thaliana имеет пять генов, кодирующих функциональную UDP-D-глюкозо-/UDP-D-галактозо-4-эпимеразу. Набор родственных белков, некоторые из которых предварительно идентифицированы как UDP-глюкозо-4-эпимераза, у Thermotoga maritima, Bacillus halodurans и нескольких архей, но которые сильно ответвлены от этого набора и для которых отсутствуют экспериментальные данные, не включен в это семейство."
233	PF01370	Семейство NAD-зависимых эпимераз/дегидратаз	Это семейство белков использует NAD в качестве кофактора. Белки этого семейства используют субстраты, представляющие собой нуклеотидные сахара, для различных химических реакций.

234	IPR036291	Суперсемейство NAD(P)-связывающих доменов	"Эта запись представляет суперсемейство NAD- и NADP-связывающих доменов с укладкой сердцевины типа Россманна, которая состоит из трех слоев (альфа/бета/альфа) с параллельным бета-листом из 6 нитей. Многие различные ферменты содержат NAD/NADP-связывающий домен, в том числе следующие. * С-концевой домен алкогольдегидрогеназ * Тирозинзависимые оксидоредуктазы (также известные как дегидрогеназы короткоцепочечных жирных кислот) * N-концевой домен глицеральдегид-3-фосфатдегидрогеназы * NAD-связывающий домен формиат-/глицератдегидрогеназ * N-концевой домен сирогемсинтазы * N-концевой домен лактатдегидрогеназы * N-концевой домен 6-фосфоглюконатдегидрогеназы (бета-лист удлинен до 8 нитей) * С-концевой домен дегидрогеназ аминокислот (дополнительная N-концевая спираль замещает С-концевую спираль * NAD-связывающий домен некоторых калиевых каналов * С-концевой домен репрессора транскрипции Rex * Орнитинциклодеаминаза * CoA-связывающий N-концевой домен альфа-цепи сукцинил-CoA-синтетазы"
235	PF05879	GTP-связывающий белок дефекта корневых волосков 3 (RHD3)	Это семейство состоит из нескольких эукариотических GTP-связывающих белков, подобных белку дефекта корневых волосков 3. Было высказано предположение, что белок RHD3 является представителем нового класса GTP-связывающих белков, который широко распространен у эукариот и необходим для регулируемого увеличения клеток [PMID:9087433]. Семейство также содержит гомологичный белок усиления синтеза YOP1 (SEY1) дрожжей, который участвует в мембранном транспорте [PMID:12427979].

236	IPR008803	RHD3/Sey1	"Это семейство состоит из белка дефекта корневых волосков 3 (RHD3) растений и его гомологов из других эукариот. RHD3 представляет собой консервативный белок с предполагаемыми GTP-связывающими мотивами, который участвует в контроле везикулярного транспорта между компартментами эндоплазматического ретикулума и комплекса Гольджи. Он необходим для соответствующего увеличения клеток корней и корневых волосков. Он может играть роль в биосинтезе клеточной стенки и организации актина. Его гомолог из Saccharomyces cerevisiae, Sey1, представляет собой динаминоподобную GTPазу, которая опосредует гомотипическое слияние в ER."
237	PS50957	Профиль домена Josephin	"Домен Josephin представляет собой эукариотический белковый модуль из приблизительно 180 остатков, который встречается самостоятельно в Josephin-подобных белках и в виде аминоконцевого домена, ассоциированного с двумя или тремя копиями убиквитин-взаимодействующего мотива (UIM), в белках, подобных атаксину-3. Хотя первоначально предполагалось, что домен Josephin может представлять собой домен, полностью состоящий из альфа-спиралей, отдаленно родственный доменам ENTH и VHS, участвующим в функции мембранного транспорта и регуляторного адаптера [PMID:12486728], теперь считается, что этот домен представляет собой домен цистеиновой протеазы, в основном состоящий из альфа-спиралей, который согласно прогнозам является активным в отношении цепей убиквитина или родственных субстратов [PMID:12944423,PMID:14559776]. Домен Josephin содержит два консервативных остатка гистидина и один остаток

			цистеина, которые необходимы для активности убиквитин-протеазы [PMID:12944423,PMID:14559776]. Некоторые белки, которые, как известно, содержат домен Josephin, перечислены ниже. - Белок 1 болезни Мачадо-Джозефа животных (атаксин 3). Он взаимодействует с ключевыми регуляторами (CBP, p300 и PCAF) транскрипции и подавляет транскрипцию. - Белок, подобный белку 1 болезни Мачадо-Джозефа растений (MJD1a-подобный) (гомолог атаксина 3). - Josephin млекопитающих 1 и 2. - Josephin-подобный белок Drosophila melanogaster. - Josephin-подобный белок Arabidopsis thaliana. Разработанный авторами настоящего изобретения профиль охватывает весь домен Josephin."
238	IPR010301	"Ядрышковый Nop52"	Считается, что Nop52 участвует в образовании 28S рРНК.
239	IPR036280	Суперсемейство мультигемовых цитохромов	"В белках, принадлежащих к семейству цитохромов типа c, группа гема ковалентно присоединена тиоэфирными связями к двум консервативным остаткам цистеина, расположенным в центре цитохрома c. Цитохромы c обычно функционируют в переносе электронов, но центры цитохромов типа c также обнаруживаются в активных центрах многих ферментов. Этот домен содержит несколько мотивов CxxCH. В мультигемовых цитохромах присутствует различное количество спиралей, а также небольшая бета-структура, но они не образуют истинной структурной укладки. Цитохром (cyt) c3-подобные белки представляют собой мультигемовые цитохромы, включая cyt c3 с четырьмя группами гема, cyt c7 (cyt c551.5) с тремя группами гема (делеция одного сайта связывания гема в cyt c3), cyt c с девятью группами гема

			(тандемный повтор из двух cyt c3-подобных доменов с дополнительным сайтом связывания гема) и cyt c HmcA с 16 группами гема (тандемный повтор из четырех cyt c3-подобных доменов). Фотосинтетический реакционный центр, состоящий из субъединицы цитохрома, также представляет собой мультигемовый цитохром. Кроме того, мотив "дигемовый локоть" демонстрирует аналогичную структуру, и основным характерным признаком этого мотива является упаковка его двух гемов; многие представители этой группы белков содержат один или несколько полных мотивов, фланкированных неполными мотивами и/или другими доменами. Например, мотив "дигемовый локоть" присутствует в домене мультигемового цитохрома, обнаруживаемом в субъединице NapB периплазматической нитратредуктазы, в гидроксиламиноксиредуктазе HAO (содержит 3 полных мотива), в cyt c554 (содержит один полный мотив), в cyt c-нитритредуктазе и в N-концевом домене флавоцитохрома c3 (дыхательной фумаратредуктазы)."
240	cd09280	Эукариотическая РНКаза H является необходимой и более длинной и сложной, чем ее прокариотические аналоги.	"Рибонуклеаза H (РНКаза H) подразделяется на два семейства: тип 1 (прокариотическая РНКаза HI, эукариотическая РНКаза H1 и вирусная РНКаза H) и тип 2 (прокариотическая РНКаза HII и HIII и эукариотическая РНКаза H2). РНКаза H представляет собой эндонуклеазу, которая расщепляет нить РНК гибрида РНК/ДНК неспецифичным в отношении последовательности образом. РНКаза H участвует в репликации, репарации и транскрипции ДНК. Одной из важных функций РНКазы H является удаление фрагментов Оказаки во время репликации ДНК. РНКаза H широко представлена в различных организмах, включая бактерии, археи и эукариот, и

			большинство прокариотических и эукариотических геномов содержат несколько генов РНКазы H. Несмотря на отсутствие гомологии аминокислотных последовательностей, РНКазы H типа 1 и типа 2 характеризуются общей укладкой основной цепи и стерической конфигурацией четырех кислотных остатков активного центра (DEDD) и характеризуются одинаковыми каталитическим механизмом и функциями в клетках. Эукариотическая РНКаза H является более длинной и сложной, чем у прокариот. Почти все эукариотические РНКазы HI имеют высококонсервативные области на своих N-концах, называемые связывающими гибриды доменами (HBD). Предполагается, что HBD способствует связыванию гибрида РНК/ДНК. Прокариоты и некоторые одноклеточные эукариоты не нуждаются в РНКазе H для обеспечения жизнеспособности, но РНКаза H необходима высшим эукариотам. У мышей с нокаутом РНКазы H отсутствует репликация митохондриальной ДНК, и они умирают на стадии эмбрионов. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198]"
241	IPR036864	Суперсемейство ДНК-связывающих доменов Zn(2)-C6 грибного типа	"N-концевая область ряда грибных белков, регулирующих транскрипцию, содержит Cys-богатый мотив, который участвует в зависимом от цинка связывании ДНК. Область образует двухъядерный кластер Zn, в котором два атома Zn связаны шестью остатками Cys. Известно, что этот домен содержится в широком диапазоне белков. К ним относятся белки, участвующие в метаболизме аргинина, пролина, пиримидина, хината, мальтозы и галактозы; катаболизме амидов и GABA; биосинтезе лейцина, помимо прочего."

242	IPR016067	"S-аденозилметиониндекарбоксилаза, сердцевинный домен"	"S-аденозилметиониндекарбоксилаза (AdoMetDC) катализирует удаление карбоксилатной группы S-аденозилметионина с образованием S-аденозил-5'-3-метилпропиламина, который затем выступает в качестве донор н-пропиламиновой группы в синтезе полиаминов спермидина и спермина из путресцина. В каталитическом механизме AdoMetDC участвует ковалентно связанная пирувоильная группа. Эта группа образуется посттрансляционно вследствие самокатализируемой реакции внутримолекулярного протеолитического расщепления между глутаматом и серином. В результате этого расщепления образуются две цепи: бета (N-концевая) и альфа (C-концевая). N-концевой сериновый остаток альфа-цепи затем преобразуется путем негидролитического серинолиза в пирувоильную группу."
243	IPR037794	Субъединица 12 фактора инициации транскрипции TFIID	"Комплекс ацетилирования гистонов, регулирующий транскрипцию, Spt-Ada-Gcn5-ацетилтрансфераза (SAGA) участвует в зависимой от РНК-полимеразы II регуляции транскрипции примерно 10% генов дрожжей. SAGA предпочтительно ацетилирует гистоны H3 и H2B и деубиквитинирует гистон H2B. SAGA известен как PCAF у позвоночных, и PCAF ацетилирует нуклеосомный гистон H3. Комплекс SAGA состоит из по меньшей мере TRA1, CHD1, SPT7, TAF5, ADA3, SGF73, SPT20/ADA5, SPT8, TAF12, TAF6, HFI1/ADA1, UBP8, GCN5, ADA2, SPT3, SGF29, TAF10, TAF9, SGF11 и SUS1, и некоторые из этих компонентов присутствуют в двух копиях. Комплекс образован отдельными модулями, каждый из которых имеет отдельную функцию и сшивается либо с другими белками, либо с другими модулями в комплексе. SLIK (SAGA-

			подобный комплекс) представляет собой многосубъединичный комплекс гистонацетилтрансфераз, который предпочтительно ацетилирует гистоны H3 и H2B и деубиквитинирует гистон H2B. Он представляет собой усовершенствование комплекса SAGA. Комплекс SLIK дрожжей состоит из по меньшей мере TRA1, CHD1, SPT7, CC TAF5, ADA3, SPT20, RTG2, TAF12, TAF6, HFI1, UBP8 (деубиквитиназы), GCN5, ADA2, SPT3, SGF29, TAF10 и TAF9. Комплекс SALSA дрожжей представляет собой измененную форму комплекса SAGA и состоит из по меньшей мере TRA1, SPT7 (формы, усеченной на C-конце), TAF5, ADA3, SPT20, TAF12, TAF6, HFI1, GCN5, ADA2 и SPT3. Комплекс ДНК-связывающего общего фактора транскрипции TFIID играет центральную роль в инициации транскрипции ДНК-зависимой РНК-полимеразы II. TFIID является единственным общим фактором инициации транскрипции, который связывается с TATA-боксом. Связывание TFIID с TATA-боксом является первой стадией образования комплекса, способного инициировать транскрипцию. TFIID состоит из TATA-связывающего белка (TBP) и 14 TBP-ассоциированных факторов (TAF). В нем представлены одна копия каждого из TAF1, TAF2, TAF3, TAF7, TAF8, TAF11, TAF13, две копии каждого из TAF4, TAF5, TAF6, TAF9, TAF10, TAF12 и три копии TAF14. Субъединица 12 фактора инициации транскрипции TFIID (TAF12) представляет собой компонент комплекса ДНК-связывающего общего фактора транскрипции TFIID и комплексов ацетилирования гистонов (HAT), регулирующих транскрипцию, SAGA, SALSA и SLIK".

244	cd04949	Гликозилтрансфераза GtfA вспомогательной системы Sec и сходные белки	"Это семейство является наиболее близкородственным семейству гликозилтрансфераз GT4 и названо в честь gtfA в Streptococcus gordonii, где он играет роль в O-связанном гликозилировании GspB - гликопротеина клеточной поверхности, участвующего в связывании тромбоцитов. В целом гликозилтрансферазы катализируют перенос сахарных фрагментов от активированных донорных молекул к специфическим акцепторным молекулам, образуя гликозидные связи. Акцепторная молекула может представлять собой липид, белок, гетероциклическое соединение или другой углеводный остаток. Эта группа гликозилтрансфераз наиболее близкородственна ранее определенному семейству 1 гликозилтрансфераз (GT1). Представители этого семейства могут переносить сахарные фрагменты, связанные с UDP, ADP, GDP или CMP. Разнообразная ферментативная активность представителей этого семейства отражает широкий спектр биологических функций. Белковая структура, доступная для этого семейства, имеет топологию GTB - одну из двух белковых топологий, наблюдаемых для зависимых от нуклеотидного сахара гликозилтрансфераз. Белки GTB имеют разные N- и C-концевые домены, каждый из которых содержит типичную укладку Россманна. Два домена характеризуются высокой структурной гомологией, несмотря на минимальную гомологию последовательностей. Большая расщелина, разделяющая два домена, содержит каталитический центр и обеспечивает высокую степень гибкости. Представители этого семейства обнаруживаются у бактерий.

			[PMID: 15489421, PMID: 12691742, PMID: 9004408, PMID: 16564135, PMID: 16037492, PMID: 16263268, PMID: 15653326]"
245	IPR014012	Домен, ассоциированный с хеликазой/SANT	"Домен, ассоциированный с хеликазой/SANT (HSA), представляет собой предсказанный ДНК-связывающий домен из ~ 75 аминокислот, обнаруживаемый в эукариотических семействах SRCAP/p400/DOM и SNF2/brahma. Хотя каждое семейство имеет сердцевинные последовательности, определяющие домен HSA, каждое из них также имеет дополнительные последовательности, которые отличают эти семейства друг от друга. Например, последовательность HWDY(L/C)EEEM(Q/V) обнаруживается в семействе SRCAP/p400/DOM, тогда как последовательность HQE(Y/F)LNSILQ обнаруживается в семействе SNF2/brahma. Помимо доменов SANT и хеликазы, домен HSA также обнаруживается в ассоциации с бромодоменом."
246	PS51204	Профиль домена HSA	"Домен, ассоциированный с хеликазой/SANT (HSA), представляет собой предсказанный ДНК-связывающий домен из ~ 75 аминокислот [PMID:11779830], обнаруживаемый в эукариотических семействах SRCAP/p400/DOM и SNF2/brahma [PMID:16024792]. Хотя каждое семейство имеет сердцевинные последовательности, определяющие домен HSA, каждое из них также имеет дополнительные последовательности, которые отличают эти семейства друг от друга. Например, последовательность HWDY(L/C)EEEM(Q/V) обнаруживается в семействе SRCAP/p400/DOM, тогда как последовательность HQE(Y/F)LNSILQ обнаруживается в семействе SNF2/brahma [PMID:16024792]. Помимо доменов SANT и хеликазы, домен HSA также обнаруживается в ассоциации с бромодоменом [PMID:11779830].

			Некоторые белки, которые, как известно, содержат домен HSA, перечислены ниже [PMID:16024792]. - Белок, регулирующий транскрипцию, SWI2/SNF2 дрожжей. - Белок brahma дрозофилы. - SNF2-альфа млекопитающих (BRM). - SNF2-бета млекопитающих (BRG). - SNF2-родственный CBP-активирующий белок (SRCAP) дрожжей - коактиватор активации транскрипции, опосредованной CREB и ядерным гормональным рецептором. - Белок domino дрозофилы (DOM) - гомолог SCRAP. Он необходим для обеспечения жизнеспособности организма и участвует в нескольких аспектах развития мух, включая гемопоэз, развитие крыльев и плодовитость самок. - SRCAP млекопитающих. - SWI2/SNF2-подобный содержащий АТРазу p400/Domino млекопитающих. Разработанный авторами настоящего изобретения профиль охватывает весь домен HSA."
247	cd17248	"Домен распознавания мишени и консервативный домен (TRD-CR) субъединицы специфичности (S) системы рестрикции-модификации типа I, сходные с TRD2-CR2 субъединицы S Actinosynnema mirum DSM 43827 (S. AmiI)"	"S-субъединица S. AmiI распознает 5' CAGNNNNNNNTCGA 3'. S-субъединица системы рестрикции-модификации (RM) состоит из двух вариабельных доменов распознавания мишени (TRD1 и 2) и двух консервативных областей (CR1 и CR2), которые разделяют TRD. Каждый TRD связывается с разными специфическими последовательностями ДНК. Например, TRD1 S. AmiI распознает CAG/CTG, а TRD2 S. AmiI распознает TCGA. Системы RM защищают бактериальную клетку от инвазии чужеродной ДНК путем эндонуклеолитического расщепления ДНК, в которой отсутствует сайт-специфическая модификация. Геном хозяина защищен от расщепления посредством метилирования определенных нуклеотидов в сайтах-мишенях. В системах типа I активность как рестрикции, так и модификации

			присутствует в одном ферментативном комплексе, состоящем из одной субъединицы специфичности (S) ДНК (это семейство), двух субъединиц модификации (M) и двух субъединиц рестрикции (R). Эта модель содержит как TRD1-CR1, так и TRD2-CR2. Она может также включать TRD-CR-подобные домены распознавания последовательности различных ферментов рестрикции и метилтрансфераз типа II и ДНК-метилтрансфераз типа I. [PMID: 8336674, PMID: 9837717, PMID: 25378308]"
248	cd01776	Ras-ассоциирующий (RA) домен семейства белков-интеракторов Ras и Rab (Rin)	"Семейство белков, взаимодействующих/интерферирующих с Ras (Rin), также известных как интеракторы Ras и Rab, состоит из Rin1, Rin2 и Rin3, которые имеют многофункциональные домены, включая SH2 и богатые пролином домены в N-концевой области и домены RH, VPS9 и RA в C-концевой области. Белки, содержащие домен RA, функционируют, взаимодействуя с белками Ras прямо или опосредованно, и участвуют в нескольких различных функциях, начиная от подавления опухоли и заканчивая деятельностью в качестве онкопротеинов. Белки Ras представляют собой небольшие GTPазы, участвующие в передаче клеточных сигналов. Домен RA имеет убиквитин-подобную (Ubl) укладку типа бета-"захвата" с низким сходством последовательностей с убиквитином; убиквитин является модификатором белков у эукариот, который участвует в различных клеточных процессах, включая регуляцию транскрипции, контроль клеточного цикла и репарацию ДНК. Домены RA Rin1, Rin2 и Rin3 являются высококонсервативными, и все они обладают характеристиками связывания с Ras. [PMID: 12574403, PMID:

			23765536, PMID: 21209314, PMID: 22249975, PMID: 11784866, PMID: 11703925, PMID: 11305092, PMID: 18796434, PMID: 22976291, PMID: 16757336, PMID: 15886098, PMID: 9208849, PMID: 9144171, PMID: 7862125, PMID: 11733506, PMID: 19631308, PMID: 24449201, PMID: 22890385, PMID: 22825554, PMID: 20424861, PMID: 12972505, PMID: 21586568, PMID: 26296892, PMID: 18486601, PMID: 23185384]"
249	cd17305	Каталитический фосфатидилинозитолфосфатказный (PIPK) домен, обнаруживаемый в фосфатидилинозитол-5-фосфат-4-киназе типа II (PIP5KII) и сходных белках	"PIP5KII, также известные как PIPKII или PI4P5KII, отвечают за синтез фосфатидилинозитол-4,5-бисфосфата (PtdIns4,5P2), важной липидной молекулы в различных клеточных процессах, из фосфатидилинозитол-5-фосфата (PtdIns5P). В эритроцитах были охарактеризованы три различных PIP5KI: изоформы PIP5K2A, PIP5K2B и PIP5K2C. [PMID: 8869744, PMID: 11294622, PMID: 16801950, PMID: 18314871, PMID: 18545987, PMID: 19644675, PMID: 24389605, PMID: 25548108, PMID: 7852364, PMID: 7639683, PMID: 9367159, PMID: 18364242, PMID: 20583997, PMID: 19896968, PMID: 9038203, PMID: 9753329, PMID: 27132569, PMID: 25866244, PMID: 24127122, PMID: 26774281, PMID: 27313209, PMID: 25495341, PMID: 26710750, PMID: 22496355, PMID: 23758345]"
250	PS00018	Предполагаемая нуклеаза HARBI1	NA
251	IPR018247	"EF-"рука" 1, сайт связывания кальция"	"Многие кальцийсвязывающие белки принадлежат к одному и тому же эволюционному семейству и содержат общий тип кальцийсвязывающего домена, известный как EF-"рука". Этот тип домена состоит из петли из двенадцати остатков, фланкированной с обеих сторон альфа-спиральным доменом из двенадцати

			остатков. В петле EF-"руки" ион кальция координируется в пентагонально-бипирамидальной конфигурации. Шесть остатков, участвующих в связывании, находятся в положениях 1, 3, 5, 7, 9 и 12; эти остатки обозначаются как X, Y, Z, -Y, -X и -Z. Инвариантный Glu или Asp в положении 12 предоставляет два атома кислорода для лигандирования Ca (бидентатный лиганд). Связывание Ca2 индуцирует конформационные изменения в мотиве EF-"рука", что приводит к активации или инактивации белков-мишеней. EF-"руки", как правило, встречаются парами или в большем числе копий. Этот паттерн сигнатуры включает полную петлю EF-"руки", а также первый остаток, который следует за петлей и, по-видимому, всегда является гидрофобным. Примечание: положения 1 (X), 3 (Y) и 12 (-Z) являются наиболее консервативными. 6-й остаток в петле EF-"руки" в большинстве случаев представляет собой Gly, но количество исключений из этого "правила" постепенно увеличивается, поэтому этот паттерн сигнатуры включает все различные остатки, которые, как было показано, существуют в этом положении в функциональных сайтах связывания Ca. Об этом паттерне известно, что в некоторых случаях в нем отсутствует одна из областей EF-"рука" в некоторых белках с несколькими доменами EF-"рука"."

Все публикации, заявки на патенты, патенты и другие публикации и литературные источники (например, регистрационные номера в базах данных последовательностей), процитированные в данном документе, включены посредством ссылки во всей своей полноте. Например, все последовательности из GenBank, UniGene и Entrez, упомянутые в данном документе, например, в любой таблице в данном документе, включены посредством ссылки. Если не указано иное, последовательности, указанные в данном документе (например, по названию гена в Repbase или по номеру доступа), в том числе в любой таблице в данном документе, относятся к записям в базах данных по состоянию на 4 марта 2020 г. Если один ген или белок относится к множеству номеров доступа последовательности, то охватываются все варианты последовательности.

ПРИМЕРЫ

Настоящее изобретение дополнительно проиллюстрировано с помощью следующих примеров. Примеры представлены только для иллюстративных целей, и их не следует истолковывать как ограничивающие объем или содержание настоящего изобретения каким-либо образом.

Пример 1. Доставка системы Gene Writer в клетки млекопитающего

В данном примере описывается система редактирования генома Gene Writer, доставляемая в клетку млекопитающего для вставки экзогенной ДНК в геном клетки млекопитающего. Составные части Gene Writer получены из ретротранспозона, например, ретротранспозона, указанного в таблице X.

В этом примере полипептидный компонент системы Gene Writer™ представляет собой ретротранспозазу Vingi_1_EE из *Erinaceus europaeus*, закодированную в плазмиде, экспрессия которой управляется промотором CMV. Матричный компонент содержит плазмиду с последовательностью гетерологичного объекта между 5'-UTR и 3'-UTR из ретротранспозона Vingi_1_EE, все из которых транскрибируются под управлением промотора CMV. Последовательность гетерологичного объекта представляет собой репортерную кассету GFP, управляемую EF1a, которая представлена в антисмысловой ориентации по отношению к транскрипту матрицы, управляемой CMV. Кроме того, ген GFP прерывается прямым интроном, так что сплайсинг РНК происходит только при транскрипции с промотора CMV, тогда как при транскрипции в противоположном направлении с промотора EF1a интрон сохраняется. В результате ORF GFP остается интактной только тогда, когда сначала происходит транскрипция с промотора CMV, а затем подвергнутый сплайсингу транскрипт подвергается обратной транскрипции и интегрируется в геном, обеспечивая возможность последующей транскрипции, управляемой EF1a, с образованием мРНК, кодирующей функциональный GFP.

Клетки U2OS подвергают нуклеофекции следующими тестируемыми средствами.

1. ДНК, кодирующая полипептид, описанный выше

2. ДНК, кодирующая полипептид, описанный выше, с мутацией, обеспечивающей инактивацию эндонуклеазы

3. ДНК, кодирующая РНК-матрицу, описанную выше

4. Комбинация из 1 и 3

5. Комбинация из 2 и 3

После трансфекции клетки U2OS культивируют в течение по меньшей мере 3 дней и затем анализируют в отношении геномных вставок. Для выявления экспрессии клетки анализируют с помощью проточной цитометрии для измерения уровней GFP, который должен продуцироваться только после успешных событий интеграции. Для молекулярного считывания образцы геномной ДНК, выделенные из каждой группы клеток U2OS, подвергают однонаправленному секвенированию для определения количества и принадлежности сайтов вставки. Чтобы выполнить однонаправленное секвенирование, геномную ДНК сначала экстрагируют, а затем фрагментируют. С фрагментами лигируют адаптеры, и проводят ПЦР с использованием праймера для адаптера и праймера, специфичного для матрицы Gene Writer. Праймер, предназначенный для охвата места соединения интрона, который, таким образом, связывается только с интегрированным GFP, используется для уменьшения количества фоновых ридов от несплайсированного GFP, присутствующего в плазмиде с матрицей. Ампликоны первого цикла затем подвергают второму циклу ПЦР для амплификации сигнала и последующему анализу на MiSeq для определения геномных местоположений успешных событий интеграции. Включение уникальных штрих-кодов в праймеры позволяет оценить количество событий интеграции. На опосредованный ретротранспозазой Gene Writing указывает присутствие событий интеграции, которыми обогащено условие (4) выше.

Пример 2. Направление ретротранспозона в определенный сайт в клетках человека

В данном примере описывается система редактирования генома Gene Writer, доставляемая в клетку млекопитающего для вставки экзогенной ДНК в геном клетки млекопитающего в заранее определенный сайт. Составные части Gene Writer получены из ретротранспозона, например, ретротранспозона, указанного в таблице X, и дополнительно содержат гетерологичный ДНК-связывающий домен.

В этом примере полипептидный компонент системы Gene Writer™ представляет собой ретротранспозазу Vingi_1_EE из *Erinaceus europaeus*, закодированную в плазмиде, экспрессия которой управляется промотором CMV. Ретротранспозаза Vingi_1_EE дополнительно содержит N-концевой домен dCas9. Полипептид Gene Writer направляется путем добавления U6-управляемой gRNA с гомологией с сайтом AAVS1, составляющей 20 нуклеотидов, таким образом, домен dCas9 в комбинации с gRNA обеспечивает активность связывания гетерологичной ДНК, направляя полипептид к сайту AAVS1 для запрограммированной ретротранспозиции матрицы. Матричный компонент содержит плазмиду с последовательностью гетерологичного объекта между 5'-UTR и 3'-UTR из ретротранспозона Vingi_1_EE, все из которых транскрибируются под управлением промотора CMV. Последовательность гетерологичного объекта представляет собой репортерную кассету GFP, управляемую EF1a, которая представлена в антисмысловой ориентации по отношению к транскрипту матрицы, управляемой CMV. Кроме того, ген

GFP прерывается прямым интроном, так что сплайсинг РНК происходит только при транскрипции с промотора CMV, тогда как при транскрипции в противоположном направлении с промотора EF1a интрон сохраняется. В результате ORF GFP остается интактной только тогда, когда сначала происходит транскрипция с промотора CMV, а затем подвергнутый сплайсингу транскрипт подвергается обратной транскрипции и интегрируется в геном, обеспечивая возможность последующей транскрипции, управляемой EF1a, с образованием мРНК, кодирующей функциональный GFP.

Клетки U2OS подвергают нуклеофекции следующими тестируемыми средствами.

1. ДНК, кодирующая полипептид и gRNA, описанные выше

2. ДНК, кодирующая полипептид и gRNA, описанные выше, с мутацией, обеспечивающей инактивацию эндонуклеазы

3. ДНК, кодирующая полипептид, описанный выше, но без gRNA

4. ДНК, кодирующая РНК-матрицу, описанную выше

5. Комбинация из 1 и 4

6. Комбинация из 2 и 4

7. Комбинация из 3 и 4

После трансфекции клетки U2OS культивируют в течение по меньшей мере 3 дней и затем анализируют в отношении геномных вставок. Для выявления экспрессии клетки анализируют с помощью проточной цитометрии для измерения уровней GFP, который должен продуцироваться только после успешных событий интеграции. Для молекулярного считывания образцы геномной ДНК, выделенные из каждой группы клеток U2OS, подвергают ddPCR для определения частоты вставок в сайте AAVS1. Праймер, который отжигается с 3'-концом матрицы, работает в паре с праймером, который отжигается ниже сайта вставки AAVS1 в геноме, и зонд располагается между парой праймеров. Таким образом, события амплификации происходят в случае наличия интеграция в сайте AAVS1, и праймеры могут отжигаться с каждой стороны от места соединения интегрированного элемента. Согласно этой схеме сигнал в анализе ddPCR, нормализованный по стандартному эталону RPP30, указывает на число копий событий интеграции в сайте AAVS1. На опосредованный ретротранспозазой Gene Writing указывает обогащение сигналом интеграции в условии (5) выше. По сравнению с условием (7), где в системе отсутствует gRNA для управления специфичностью нацеливания полипептида, условие (5) демонстрирует по меньшей мере пятикратное увеличение эффективности интеграции в сайте AAVS1 согласно измерению с помощью ddPCR.

75. **Пример 3. Gene Writer™, обеспечивающий замену нуклеотидов в геномной ДНК для коррекции дефицита альфа-1-антитрипсина**

Систему Gene Writer можно применять для изменения ДНК-мишени, например, геномной ДНК, путем вставки, делеции или замены нуклеотидов. Без желания ограничиваться единственным подходом в данном примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности в одном нуклеотиде.

В данном примере полипептид Gene Writer™ и матрица для записи представлены в виде ДНК, трансфицированной в клетки HEK293T, которые обладают генотипом PiZ (E342L) - распространенным аллелем, ассоциированным с дефицитом альфа-1-антитрипсина. Полипептид Gene Writer™ использует никазу Cas9 в качестве как ДНК-связывающей, так и эндонуклеазной функциональной единицы. Матрица для записи разработана таким образом, что она характеризуется гомологией с последовательностью-мишенью и в то же время содержит дополнительные нуклеотиды в необходимом положении таким образом, что обратная транскрипция РНК-матрицы приводит в результате к образованию новой нити ДНК, содержащей замену.

Для создания трансверсии в пораженном локусе гена *SERPINA1* человека, которая восстанавливает триплет GAG, кодирующий глутамат у здоровых пациентов, используется полипептид Gene Writer™ со специальной матрицей, которая кодирует спейсер для хоминга полипептида, домен гомологии с мишенью для обеспечения TPRT и последовательность матрицы для обратной транскрипции, которая содержит замену на G.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени. После периода инкубирования из клеток экстрагируют геномную ДНК. Затем геномную ДНК подвергают амплификации на основе ПЦР с использованием сайт-специфических праймеров, а ампликоны секвенируют на Illumina MiSeq в соответствии с протоколами производителя. Затем выполняют анализ последовательности с определением частоты ридов, содержащих необходимое редактирование.

76. **Пример 4. Выявление интеграции, опосредованной ретротранспозазой, в клетках человека**

В этом примере описывается идентификация ретротранспозонов, демонстрирующих функциональные свойства в клетках человека. Путем анализа нативных или модифицированных ретротранспозонов в отношении активности интеграции в этом примере демонстрируется способ отбора ретротранспозаз, содержащих белковые домены, которые можно использовать для воссоздания ретротранспозаз с их нативным составом доменов или в качестве компонентов химерных или синтетических элементов Gene Writer для модификации генома в клетках человека. Например, ожидается, что ретротранспозон, успешно продуцирующий сигнал интеграции, будет характеризоваться функциональными формами активности связывания ДНК, эндонуклеазы, обратной транскриптазы и необязательно синтеза второй нити. В некоторых вариантах осуществления домен обратной транскриптазы из ретротранспозона, который, как было показано, демонстрирует активность, описанную в этом примере, используется для обеспечения активности обратной транскриптазы в полипептиде Gene Writer, например, в качестве RT слитого полипептида Cas-RT, например, как раскрыто в таблице 9. В описанном в данном документе скрининге используется нуклеофекция двухплазмидной системы, содержащей полипептид ретротранспозона и инактивированную репортерную матрицу, в клетки человека для

определения характеристик эффективности зависимой от RT ретротранспозиции ретротранспозонов, выбранных с помощью вычислений.

В этом примере использовали двухплазмидную систему, содержащую: 1) плазмиду с управляющим элементом для экспрессии белка, кодируемую ретротранспозоном, например, плазмиду, кодирующую полипептид ретротранспозазу из таблицы 3B, содержащую кодон-оптимизированную последовательность, кодирующую ретротранспозазу человека, слитую с меткой HiBit для выявления экспрессии белка и находящуюся под управлением промотора CMV млекопитающих, и 2) плазмиду с матрицей, например, плазмиду, содержащую (i) промотор для экспрессии в клетках млекопитающих для управления транскрипцией молекулы РНК-матрицы, например, промотор CMV, при этом молекула-матрица дополнительно содержит (ii) репортерную кассету, которая неактивна в ситуации экспрессии, происходящей из плазмиды, например, кассету экспрессии EGFP с кодирующей последовательностью, нарушенной интроном, кодируемым в противоположной ориентации (GFPai), фланкированную (iii) нетранслируемыми областями (UTR) нативного ретротранспозона, который в естественных условиях содержит ретротранспозазу (1) (см. фигуру 8). В данном документе репортер GFP кодируется в отсутствие промотора, управляющего его экспрессией, чтобы избежать какой-либо потери сигнала из-за токсичности GFP (см. фигуру 8).

Чтобы доставить двухплазмидную систему в клетки U2OS, ~ 400000 клеток подвергли нуклеофекции 88,3 нг плазмиды с управляющим элементом (1) и 161,7 нг плазмиды с матрицей (2) с использованием 96-луночного набора Nucleofector™ Lonza SE для линии клеток в соответствии с инструкциями производителя. Через три дня после нуклеофекции эффективность интеграции измеряли с помощью ddPCR для определения числа копий интегрированных элементов на геном. Активность ретротранспозиции, зависимую от обратной транскрипции, измеряли с использованием подхода ddPCR, в котором использовался антисмысловой интрон, как описано ниже. Экспрессию белка, являющегося управляющим элементом, измеряли с помощью биолюминесцентного анализа на основе HiBit.

При использовании репортера с антисмысловым интроном, содержащего интронную последовательность в репортерной кассете плазмиды с матрицей, например, описанной в данном документе системы GFPai, интрон присутствует в плазмиде, но вырезается путем сплайсинга во время транскрипции, таким образом, только репортерная ДНК, полученная из транскрипта путем обратной транскрипции, не будет иметь последовательность интрона (фигура 8). Для ограничения выявления только событиями, происходящими в результате обратной транскрипции, зонд ddPCR Taqman был разработан так, чтобы охватывать границу сплайсинга для гибридизации с ДНК, в которой отсутствует интрон, но не с плазмидной ДНК, все еще содержащей интактный интрон. Прямой и обратный праймеры были разработаны выше и ниже зонда и внутри последовательности GFP. Эта конструкция позволяет избежать возможного фона от плазмиды с матрицей,

непосредственно рекомбинированной в геном без первого этапа транскрипции, или от интактной плазмиды с матрицей, контаминирующей экстрагированные образцы gDNA.

Системы Gene Writing, полученные из ретротранспозаз в таблице 3B, анализировали согласно этому примеру для определения активности в клетках человека. Анализ эффективности интеграции 163 систем ретротранспозонов-кандидатов с помощью ddPCR показан на фигуре 9. Из анализа, описанного в этом примере, 25 ретротранспозаз-кандидатов продемонстрировали успешную интеграцию в транс-положении последовательности-матрицы ретротранспозона, фланкированной UTR, в среднем при более 0,01 копии/геном (см. фиг. 10).

В некоторых вариантах осуществления системы-кандидаты дополнительно оптимизируют, как описано в другом месте данной заявки. В некоторых вариантах осуществления две плазмиды с управляющим элементом разрабатывают для одной системы ретротранспозонов на основе различных консенсусных подходов (см., например, пример 8). В некоторых вариантах осуществления плазмиды с управляющим элементом являются кодон-оптимизированными в соответствии с предпочтительными кодонами человека. В некоторых вариантах осуществления молекула РНК-матрицы разработана так, чтобы дополнительно содержать последовательности, фланкирующие UTR ретротранспозона, для способствования интеграции, например, плечи гомологии, характеризующиеся идентичностью последовательности с областями выше и ниже ретротранспозона в его эндогенном геномном окружении или с последовательностью ДНК-мишени в человеческом геноме (см., например, пример 9).

77. **Пример 5. Консенсусное картирование генома для аминокислотной последовательности ретротранспозазы с улучшенной активностью в клетках человека**

В этом примере описывается использование консенсусных подходов к интеллектуальному анализу генома для разработки улучшенной активности ретротранспозонов. С учетом того, что ретротранспозон может существовать во многих копиях в геноме, сравнение всех копий может быть использовано для реконструкции предковой последовательности или последовательности, которая, как ожидается, будет обладать наивысшей активностью в организме хозяина.

В этом примере ретротранспозоны из таблицы 3B идентифицировали с помощью вычислений в геноме их хозяина. Идентифицировали геномные нуклеотидные последовательности, которые характеризовались высокими уровнями совпадений с концами каждого ретротранспозона и по меньшей мере 80% общей нуклеотидной идентичностью с полным элементом. Предсказанную кодирующую последовательность в каждом попарном выравнивании с ретротранспозоном-мишенью подвергали трансляции и использовали для получения консенсусной аминокислотной последовательности для соответствующей ретротранспозазы на основе всех копий в геноме хозяина.

Для создания конструкций для повторно выведенных на основе консенсуса ретротранспозаз каждую аминокислотную последовательность подвергали обратной

трансляции, оптимизировали для человека и клонировали в плазмиду с управляющим элементом, как в примере 7. Полученные с помощью консенсусного подхода ретротранспозазы затем сравнивали с последовательностями ретротранспозаз первого поколения в отношении активности интеграции. Для каждой системы ретротранспозонов сравнивали эффективность интеграции, обусловленную полученными с помощью консенсусного подхода и исходными ретротранспозазами, с использованием двухплазмидной системы, описанной в примере 7. Одни и те же плазмиды с матрицей подвергали совместной нуклеофекции с одним из двух управляющих элементов в клетки U2OS. ~ 400000 клеток подвергли нуклеофекции 88,3 нг плазмиды с управляющим элементом и 161,7 нг плазмиды с матрицей с использованием 96-луночного набора Nucleofector Lonza SE для линии клеток в соответствии с протоколами производителя. Геномную ДНК экстрагировали через 3 дня после нуклеофекции для количественной оценки эффективности интеграции с помощью ddPCR, как описано в примере 7.

Некоторые ретротранспозоны демонстрировали восстановленную или улучшенную активность, исходя из считываемого показателя интеграции в ddPCR. В частности, полученные с помощью консенсусного подхода последовательности для CR1-3_IS, CR-10_Ami, CR1-54_AAe и TART-1_DWi интегрировались в клетки человека только с управляющими элементами, полученными с помощью консенсусного подхода (фигура 11). Кроме того, L2-18_ACar и RTE-2_LMi демонстрировали заметно улучшенную активность при использовании полученной с помощью консенсусного подхода последовательности. Улучшенные последовательности показаны в таблице 11.

78. **Пример 6. Добавление плеч гомологии, полученных из эндогенных сайтов интеграции у нативного хозяина, улучшает интеграцию матрицы Gene Writer в клетки человека**

В этом примере описывается использование подходов к интеллектуальному анализу генома на основе мотивов для разработки улучшенной активности Gene Writing.

Более конкретно, ретротранспозоны из таблицы 3B идентифицировали с помощью вычислений в геноме их хозяина. Идентифицировали геномные нуклеотидные последовательности, которые характеризовались высокими уровнями совпадений с концами каждого нативного ретротранспозона и по меньшей мере 80% общей нуклеотидной идентичностью с полным элементом. Соседние 3'- и 5'-последовательности из генома хозяина экстрагировали из каждого идентифицированного случая появления данного элемента, и получали мотивы, соответствующие 3'- и 5'-концам. Консенсусные мотивы затем выравнивали с человеческим геномом (сборка hg38), и плечи гомологии конструировали с идентичностью последовательностям человека на основе наилучших совпадений с каждым мотивом. Используемые в данном документе плечи гомологии, специфичные для человека, указаны в таблице 11.

Для получения новых плазмид с матрицей 3'- и 5'-плечи гомологии, описанные выше, вставляли непосредственно ниже и выше 3'- и 5'-последовательностей UTR ретротранспозона, как описано в примере 4. Плазмиды с матрицей, содержащие или не

содержащие плечи гомологии, подвергали трансфекции с соответствующей плазмидой с управляющим элементом, как описано в примере 4, и сравнивали характеристики двух версий матрицы, чтобы охарактеризовать любые изменения в активности. В частности, ~ 400000 клеток U2OS подвергли нуклеофекции 88,3 нг плазмиды с управляющим элементом и 161,7 нг плазмиды с матрицей с использованием 96-луночного набора Nucleofector Lonza SE для линии клеток в соответствии с протоколами производителя. gDNA экстрагировали через 3 дня после нуклеофекции для количественной оценки эффективности интеграции каждой системы с помощью ddPCR, как описано в примере 4.

Некоторые матрицы Gene Writing с добавленными плечами гомологии демонстрировали улучшенную или восстановленную активность, исходя из считываемого показателя интеграции в ddPCR. В частности, разработка и добавление плеч гомологии приводили к улучшению интеграции L2-18_ACar, Dong-1_MMa, L2-2_Dre, CR1_AC_1 и RTE-1_Aip (фигура 12). Добавление плеч гомологии приводило к восстановлению интеграции для Tad1-65B_BG по сравнению с конструкцией без плеч гомологии, в которой сигнал ранее не поддавался выявлению.

Пример 7. Выбор липидных реагентов со сниженным содержанием альдегидов

В этом примере липиды выбирают для дальнейшего использования в составах на основе липидных наночастиц, содержащих компонент Gene Writing, представляющий собой нуклеиновую(нуклеиновые) кислоту(кислоты), и липиды выбирают, по меньшей мере частично, на основании отсутствия или низкого уровня контаминирующих альдегидов. Реакционноспособные альдегидные группы в липидных реагентах могут обуславливать химические модификации компонента, представляющего собой нуклеиновую(нуклеиновые) кислоту(кислоты), например, РНК, например, РНК-матрицу, во время составления LNP. Следовательно, в некоторых вариантах осуществления содержание альдегидов в липидных реагентах сводят к минимуму.

Жидкостную хроматографию (LC) в сочетании с тандемной масс-спектрометрией (MC/MC) можно использовать для разделения, определения характеристик и количественной оценки содержания альдегидов в реагентах, например, как описано в работе Zurek et al. The Analyst 124(9):1291-1295 (1999), включенной в данный документ посредством ссылки. В данном документе каждый липидный реагент подвергают анализу методом LC-MS/MS. Посредством метода LC/MS-MS сначала отделяют липид и одну или несколько примесей на колонке HPLC C8, а затем следует выявление и определение структуры этих молекул с помощью масс-спектрометра. Если альдегид присутствует в липидном реагенте, его количественно оценивают с использованием стандарта, меченного основным изотопом (SIL), который структурно идентичен альдегиду, но тяжелее вследствие мечения с помощью C13 и N15. Соответствующее количество SIL-стандарта добавляют в липидный реагент. Затем смесь подвергают анализу методом LC-MS/MS. Количество контаминирующего альдегида определяют путем умножения количества SIL-стандарта на соотношение пиков (неизвестный/SIL). Любой(любые) идентифицированный(идентифицированные) альдегид(альдегиды) в липидных реагентах

оценивают количественно, как описано. В некоторых вариантах осуществления было обнаружено, что исходные липидные материалы, выбранные для состава на основе LNP, не содержат каких-либо контаминирующих альдегидов на уровне, выше выбранного. В некоторых вариантах осуществления один или несколько и необязательно все липидные реагенты, используемые для составления, составляют менее 3% общего содержания альдегидов. В некоторых вариантах осуществления один или несколько и необязательно все липидные реагенты, используемые для составления, составляют менее 0,3% какой-либо отдельной разновидности альдегидов. В некоторых вариантах осуществления один или несколько и необязательно все липидные реагенты, используемые в составе, составляют менее 0,3% какой-либо отдельной разновидности альдегидов и менее 3% общего содержания альдегидов.

Пример 8. Количественная оценка модификации РНК, обусловленной альдегидами, во время составления

В этом примере молекулы РНК анализируют после составления для определения степени любых модификаций, которые могли произойти в процессе составления, например, для выявления химических модификаций, обусловленных контаминацией альдегидами липидных реагентов (см., например, пример 7).

Модификации РНК можно выявить с помощью анализа рибонуклеозидов, например, согласно способам из работы Su et al. Nature Protocols 9:828-841 (2014), включенной в данный документ посредством ссылки во всей своей полноте. В этом способе РНК расщепляется до смеси нуклеозидов, а затем подвергается анализу методом LC-MS/MS. РНК после составления содержится в LNP и должна быть сначала отделена от липидов посредством соосаждения с GlycoBlue в 80% растворе изопропанола. После центрифугирования осадки, содержащие РНК, осторожно переносят в новую пробирку Eppendorf, в которую добавляют коктейль ферментов (бензоназы, фосфодиэстеразы типа 1, фосфатазы) для расщепления РНК до нуклеозидов. Пробирку Eppendorf помещают в предварительно нагретый до 37°C прибор Thermomixer на 1 час. Полученную смесь нуклеозидов анализируют непосредственно методом LC-MS/MS, с помощью которого сначала отделяют нуклеозиды и модифицированные нуклеозиды с помощью колонки C18, а затем выявляют их с помощью масс-спектрометрии.

Если альдегид(альдегиды) в липидных реагентах обусловил(обусловили) возникновение химической модификации, анализа данных позволит установить наличие ассоциации между модифицированным(модифицированными) нуклеозид(нуклеозидами) и альдегидом(альдегидами). Модифицированный нуклеозид можно подвергнуть количественной оценке с использованием SIL-стандарта, который структурно идентичен нативному нуклеозиду, за исключением того, что более тяжелый вследствие мечения с помощью C13 и N15. Соответствующее количество стандарта SIL добавляют в гидролизат нуклеозидов, который затем подвергают анализу методом LC-MS/MS. Количество модифицированного нуклеозида получают путем умножения количества SIL-стандарта на

соотношение пиков (неизвестный/SIL). LC-MS/MS позволяет осуществлять одновременную количественную оценку всех молекул-мишеней.

В некоторых вариантах осуществления использование липидных реагентов с более высоким содержанием контаминирующих альдегидов приводит в результате к более высоким уровням модификации РНК по сравнению с использованием липидных реагентов более высокой чистоты в качестве материалов в процессе составления липидных наночастиц. Таким образом, в предпочтительных вариантах осуществления используются липидные реагенты более высокой чистоты, в результате чего уровень модификации РНК находится ниже приемлемого уровня.

Пример 9. Gene Writer™ позволяет осуществлять вставку крупного участка в геномную ДНК

В этом примере описано использование системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки крупной последовательности нуклеотидов.

В этом примере полипептид Gene Writer™, gRNA и матрица для записи представлены в виде ДНК, трансфицированной в клетки HEK293T. Полипептид Gene Writer™ использует никазу Cas9 в качестве как ДНК-связывающей, так и эндонуклеазной функциональной единицы. Функциональная единица обратной транскриптазы получена из высокопроцессивного домена RT ретротранспозазы R2. Матрица для записи сконструирована так, что она характеризуется гомологией с последовательностью-мишенью и в то же время содержит генетическую полезную нагрузку в необходимом положении таким образом, что обратная транскрипция РНК-матрицы приводит в результате к образованию новой нити ДНК, содержащей необходимую вставку.

Для создания вставки крупного участка в ДНК клетки HEK293T человека полипептид Gene Writer™ используют в сочетании со специфической gRNA, которая направляет Cas9-содержащий Gene Writer™ к локусу-мишени, и РНК-матрицей для обратной транскрипции, которая содержит RT-связывающий мотив (3'-UTR из элемента R2) для ассоциации с обратной транскриптазой, область гомологии с сайтом-мишенью для праймирования обратной транскрипции и генетическую полезную нагрузку (экспрессионную единицу GFP). Этот комплекс осуществляет однонитевой разрыв в сайте-мишени, а затем выполняет TPRT на матрице, инициируя реакцию с использованием праймирующих областей на матрице, которые комплементарны последовательности, непосредственно прилегающей к сайту однонитевого разрыва, и копируя полезную нагрузку GFP в геномную ДНК.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени. После периода инкубирования из клеток экстрагируют геномную ДНК. Затем геномную ДНК подвергают амплификации на основе ПЦР с использованием сайт-специфических праймеров, а ампликоны секвенируют на Illumina MiSeq в соответствии с протоколами производителя.

Затем выполняют анализ последовательности с определением частоты ридов, содержащих необходимое редактирование.

Пример 10. Gene Writer могут интегрировать генетический груз независимо от пути репарации посредством однонитевой матрицы

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации посредством однонитевой матрицы (SSTR) является ингибированным.

В таком примере путь SSTR будут ингибировать с помощью siRNA к основным компонентам пути: FANCA, FANCD2, FANCE, USP1. Также будут включены контрольные siРНК ненацеленного контроля. Нуклеофекцию 200 тысяч клеток U2OS будут осуществлять с использованием 30 пмоль (1,5 мкМ) siRNA, а также с использованием плазмиды с управляющим элементом R2Tg и плазмиды с трансгеном (транс-конфигурация). В частности, 250 нг плазмид, экспрессирующих R2Tg, контрольный R2Tg с мутацией в домене RT или контрольный R2Tg с мутацией, обеспечивающей инактивацию эндонуклеазы, применяют в сочетании с трансгеном в молярном соотношении 1:4 (управляющий элемент к трансгену). Трансфекции клеток U2OS проводят в буфере SE с применением программы DN100. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят ddPCR для оценки интеграции в сайте рДНК. Интеграцию трансгена в рДНК выявляют по отсутствию основных компонентов пути SSTR.

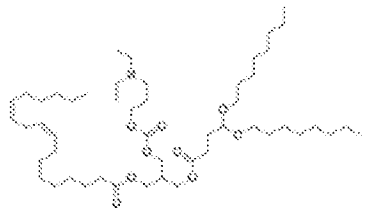
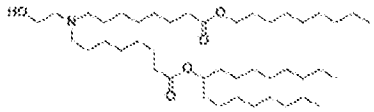
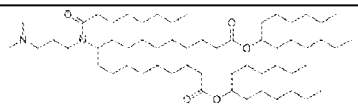
Пример 11. Состав на основе липидных наночастиц, инкапсулирующих мРНК люциферазы светлячка

В этом примере мРНК репортерного гена, кодирующую люциферазу светлячка, составляли с включением в липидные наночастицы, содержащие различные ионизируемые липиды. Компоненты липидных наночастиц (LNP) (ионизируемый липид, вспомогательный липид, стерин, PEG) растворяли в 100% растворе этанола с липидным компонентом. Затем их получали при молярных соотношениях 50:10:38,5:1,5 с использованием ионизируемого липида LIPIDV004 или LIPIDV005 (таблица A1), DSPC, холестерина и DMG-PEG 2000 соответственно. мРНК люциферазы светлячка-LNP, содержащие ионизируемый липид LIPIDV003 (таблица A1), получали при молярном соотношении 45:9:44:2 с использованием LIPIDV003, DSPC, холестерина и DMG-PEG 2000 соответственно. мРНК люциферазы светлячка, используемую в этих составах, получали путем транскрипции in vitro, и она кодировала белок люциферазы светлячка, дополнительно содержащий 5′-кэп, 5′- и 3′-UTR и поли(A)-хвост. мРНК синтезировали в стандартных условиях для транскрипции РНК-полимеразой T7 in vitro с кэпированием во время транскрипции, однако с 100% заменой нуклеотидтрифосфата UTP на N1-метилпсевдоуридинтрифосфат в реакции. Очищенную мРНК растворяли в 25 мМ цитрате натрия, pH 4, до концентрации 0,1 мг/мл.

мРНК люциферазы светлячка составляли с включением в LNP с молярным соотношением амина липида и фосфата РНК (N:P), составляющим 6. LNP составляли посредством микрофлюидного смешивания растворов липидов и РНК с использованием

настольного прибора Precision Nanosystems NanoAssemblr™ с использованием параметров, рекомендованных производителем. Соотношение водного и органического растворителей 3:1 поддерживали во время смешивания с использованием различных скоростей потока. После смешивания LNP собирали и подвергали диализу в 15 мМ Tris, 5% сахарозном буфере при 4°C в течение ночи. Состав на основе мРНК люциферазы светлячка-LNP концентрировали с помощью центрифугирования с использованием центрифужных фильтров Amicon с размерами пор 10 кДа (Millipore). Полученную смесь затем фильтровали с использованием стерильного фильтра с размером пор 0,2 мкм. Конечный LNP хранили при -80°C до дальнейшего использования.

Таблица A1. Ионизируемые липиды, используемые в примере 11 (формула (ix), (vii) и (iii))

ID ЛИПИДА	**Химическое название**	**Молекулярная масса**	**Структура**
LIPIDV003	(9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат	852,29	
LIPIDV004	Гептадекан-9-ил-8-((2-гидроксиэтил)(8-(нонилокси)-8-оксооктил)амино)октаноат	710,18	
LIPIDV005		919,56	

Полученные LNP анализировали в отношении размера, однородности и % инкапсуляции РНК. Измерения размера и однородности проводили с помощью динамического светорассеяния с использованием прибора для DLS Malvern Zetasizer (Malvern Panalytical). LNP разводили в PBS перед измерением с помощью DLS с определением среднего размера частиц (нанометры, нм) и индекса полидисперсности (pdi). Размеры частиц мРНК люциферазы светлячка-LNP показаны в **таблице A2**.

Таблица A2. Размер и однородность частиц LNP

ID LNP	**Ионизируемый липид**	**Размер частиц (нм)**	**pdi**
LNPV019-002	LIPIDV005	77	0,04
LNPV006-006	LIPIDV004	71	0,08

LNPV011-003	LIPIDV003	87	0,08

Процент инкапсуляции мРНК люциферазы измеряли с помощью анализа для количественной оценки РНК на основе флуоресценции Ribogreen (ThermoFisher Scientific). Образцы LNP разводили в буфере 1× TE и смешивали с реагентом Ribogreen в соответствии с рекомендациями производителя и измеряли на спектрофотометре i3 SpectraMax (Molecular Devices) с использованием длин волн возбуждения 644 нм и испускания 673 нм. Для определения процента инкапсуляции LNP измеряли с помощью анализа Ribogreen с интактными LNP и разрушенными LNP, где частицы инкубировали с буфером 1× TE, содержащим 0,2% (вес/вес) Triton-X100, с обеспечением разрушения частиц для обеспечения взаимодействия инкапсулированной РНК с реагентом Ribogreen. Образцы измеряли повторно на спектрофотометре i3 SpectraMax для определения общего количества присутствующей РНК. Суммарную РНК вычитали из количества РНК, выявленного, когда LNP были интактными, с определением доли, которая была инкапсулирована. Значения умножали на 100 с определением процента инкапсуляции. Значения мРНК люциферазы светлячка-LNP, которые измеряли с помощью Ribogreen, и процент инкапсуляции РНК приведены в таблице А3.

Таблица А3. Инкапсуляция РНК после составления LNP

ID LNP	Ионизируемый липид	% инкапсуляции мРНК
LNPV019-002	LIPIDV005	98
LNPV006-006	LIPIDV004	92
LNPV011-003	LIPIDV003	97

Пример 12. Тестирование активности мРНК-LNP in vitro в первичных гепатоцитах

В этом примере LNP, содержащие мРНК репортерного гена люциферазы, использовали для доставки РНК-груза в клетки в культуре. Первичные гепатоциты мыши или первичные гепатоциты человека размораживали и высевали в покрытые коллагеном 96-луночные планшеты для тканевых культур с плотностью 30000 или 50000 клеток на лунку соответственно. Клетки высевали в 1x среду E Уильяма без фенолового красного и инкубировали при 37°C с 5% CO_2. Через 4 часа среду заменяли поддерживающей средой (1x среда E Уильяма без фенола, содержащая Hepatocyte Maintenance Supplement Pack (ThermoFisher Scientific)) и клетки выращивали в течение ночи при 37°C с 5% CO_2. мРНК люциферазы светлячка-LNP размораживали при 4°C и осторожно перемешивали. LNP разводили до соответствующей концентрации в поддерживающей среде, содержащей 7,5% эмбриональной бычьей сыворотки. LNP инкубировали при 37°C в течение 5 минут перед добавлением к высеянным первичным гепатоцитам. Для оценки доставки РНК-груза в клетки LNP инкубировали с первичными гепатоцитами в течение 24 часов, и затем клетки собирали и лизировали для анализа люциферазной активности. Вкратце, среду аспирировали из каждой лунки с последующей промывкой с помощью 1x PBS. PBS аспирировали из каждой лунки и обратно добавляли 200 мкл буфера для пассивного лизиса

(PLB) (Promega) в каждую лунку, а затем помещали на шейкер для планшетов на 10 минут. Лизированные клетки в PLB замораживали и хранили при -80°C до проведения анализа люциферазной активности.

Для проведения анализа люциферазной активности клеточные лизаты в буфере для пассивного лизиса размораживали, переносили в круглодонный 96-луночный титрационный микропланшет и центрифугировали при 15000*g* при 4°C в течение 3 минут с удалением клеточного дебриса. Концентрацию белка измеряли для каждого образца с использованием набора для анализа белка Pierce™ с BCA (ThermoFisher Scientific) в соответствии с инструкциями производителя. Значения концентрации белка использовали для нормализации по числу клеток и определения подходящих разбавлений лизатов для анализа с использованием люциферазы. Анализ люциферазной активности проводили в 96-луночных титрационных микропланшетах с белыми стенками с использованием реагента для анализа с использованием люциферазы (Promega) в соответствии с инструкциями производителя, и люминесценцию измеряли с помощью планшет-ридера i3X SpectraMax (Molecular Devices). Результаты в виде зависимости "доза-ответ" для активности люциферазы светлячка, опосредованной мРНК светлячка-LNP, показаны на фиг. 13 и указывают на успешную LNP-опосредованную доставку РНК в первичные клетки в культуре. Как показано на фиг. 13A, LNP, составленные в соответствии с примером 11, анализировали в отношении доставки груза в первичные гепатоциты человека (A) и мыши (B) в соответствии с примером 12. В результате анализа люциферазы была выявлена дозозависимая люциферазная активность в лизатах клеток, что указывает на успешную доставку РНК в клетки и экспрессию люциферазы светлячка из мРНК-груза.

Пример 13. LNP-опосредованная доставка РНК в печень мыши

Для измерения эффективности LNP-опосредованной доставки частиц, содержащих люциферазу светлячка, в печень LNP составляли и определяли их характеристики, как описано в примере 11, и тестировали in vitro (пример 12) до введения мышам. Самцам мышей C57BL/6 (Charles River Labs) в возрасте примерно 8 недель вводили LNP внутривенным (i.v.) путем в дозе 1 мг/кг. Животным с введением контроля, представляющего собой среду-носитель, вводили i.v. 300 мкл фосфатно-солевого буферного раствора. Мышам инъецировали внутрибрюшинно дексаметазон в дозе 5 мг/кг за 30 минут до инъекции LNP. Ткани собирали при вскрытии через 6, 24 или 48 часов после введения LNP при размере группы 5 мышей на момент времени. Образцы печени и других тканей собирали, быстро замораживали в жидком азоте и хранили при температуре −80°C до проведения анализа.

Образцы замороженной печени измельчали на сухом льду и переносили в пробирки для гомогенизации, содержащие гранулы лизирующей матрицы D (MP Biomedical). В каждую пробирку добавляли ледяной 1x реагент для лизиса культуры клеток с люциферазой (CCLR) (Promega) и образцы гомогенизировали в гомогенизаторе Fast Prep-24 5G (MP Biomedical) при скорости 6 м/с в течение 40 секунд. Образцы переносили в чистую микроцентрифужную пробирку и очищали с помощью центрифугирования. Перед

анализом люциферазной активности для каждого образца определяли концентрацию белка в гомогенатах печени с использованием набора для анализа белка Pierce™ с BCA (ThermoFisher Scientific) в соответствии с инструкциями производителя. Люциферазную активность измеряли с использованием 200 мкг (общий белок) гомогената печени с использованием реагента для анализа люциферазы (Promega) в соответствии с инструкциями производителя с использованием планшет-ридера i3X SpectraMax (Molecular Devices). Для образцов печени была выявлена успешная доставка мРНК со всеми липидными составами, при этом активность репортерного гена соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003 (фиг. 14). Как показано на фиг. 14, LNP, содержащие мРНК люциферазы светлячка, составляли и доставляли мышам iv, а образцы печени собирали и анализировали в отношении люциферазной активности через 6, 24 и 48 часов после введения. Активность репортерного гена в различных составах соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003. Экспрессия РНК была транзиентной, и уровни ферментов возвращались к уровню, близкому к фоновому уровню при обработке средой-носителем, через 48 часов после введения. В этом анализе подтверждали использование данных ионизируемых липидов и их соответствующих составов для систем на основе РНК для доставки в печень.

Без ограничения примером, липиды и составы, описанные в этом примере, свидетельствуют об эффективности доставки in vivo и других молекул РНК помимо мРНК репортерного гена. Системы Gene Writing, состоящие полностью из РНК, могут быть доставлены с помощью составов, описанных в данном документе. Например, системы, состоящие полностью из РНК, в которых используется мРНК полипептида Gene Writer, РНК-матрица и необязательная gRNA для второго однонитевого разрыва, описаны для редактирования генома in vitro путем нуклеофекции, путем использования модифицированных нуклеотидов, путем липофекции и редактирования клеток, например, первичных Т-клеток. Как описано в настоящей заявке, эти системы, состоящие полностью из РНК, характеризуются многими уникальными преимуществами в отношении клеточной иммуногенности и токсичности, что важно при работе с более чувствительными первичными клетками, особенно иммунными клетками, например, Т-клетками, в отличие от линий клеток на основе иммортализованной культуры клеток. Кроме того, предполагается, что все эти системы на основе РНК могут быть нацелены на альтернативные ткани и типы клеток с использованием новых систем доставки липидов, как указано в данном документе, например, для доставки в печень, легкие, мышцы, иммунные клетки и другие, учитывая, что функцию систем Gene Writing проверяли в данном документе in vitro на нескольких типах клеток, и функция других систем на основе РНК, доставляемых с помощью нацеленных LNP, известна в данной области техники. Доставка in vivo систем Gene Writing может оказать большое влияние на многие области медицины, например, исправление патогенных мутаций, внедрение защитных вариантов и улучшение клеток, эндогенных для организма, например, Т-клеток. Предполагается, что

при надлежащем составлении Gene Writing, состоящая полностью из РНК, сделает возможным изготовление средств клеточной терапии in situ в организме пациента.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, 10 или 11, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

2. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, 10 или 11 или последовательностью, имеющей не более 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидных отличий от нее; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

3. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту (например, ДНК или мРНК), кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) домен связывания ДНК-мишени, где один или оба из (i) или (ii) имеют аминокислотную последовательность, кодируемую последовательностью нуклеиновой кислоты элемента из таблицы 3B, 10 или 11, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта.

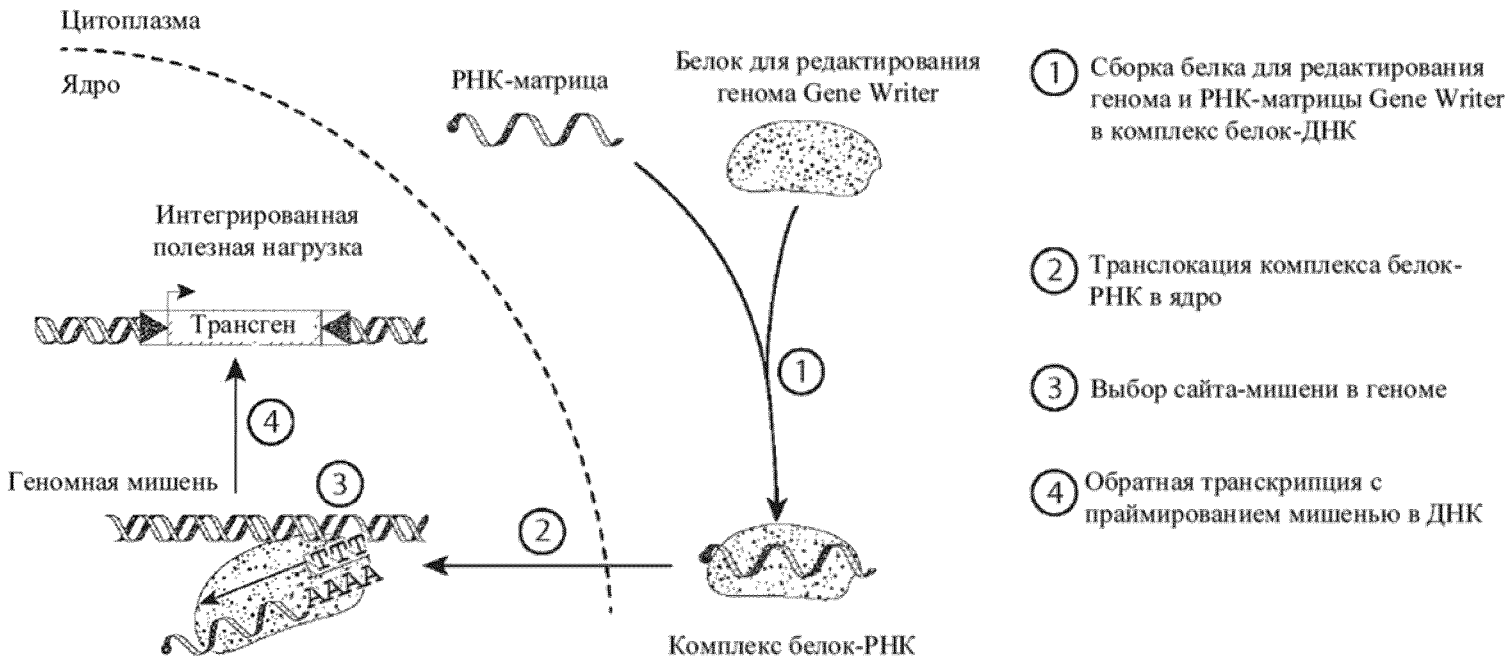
4. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, никазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (например, в направлении от 5' к 3') (i) необязательно последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии,

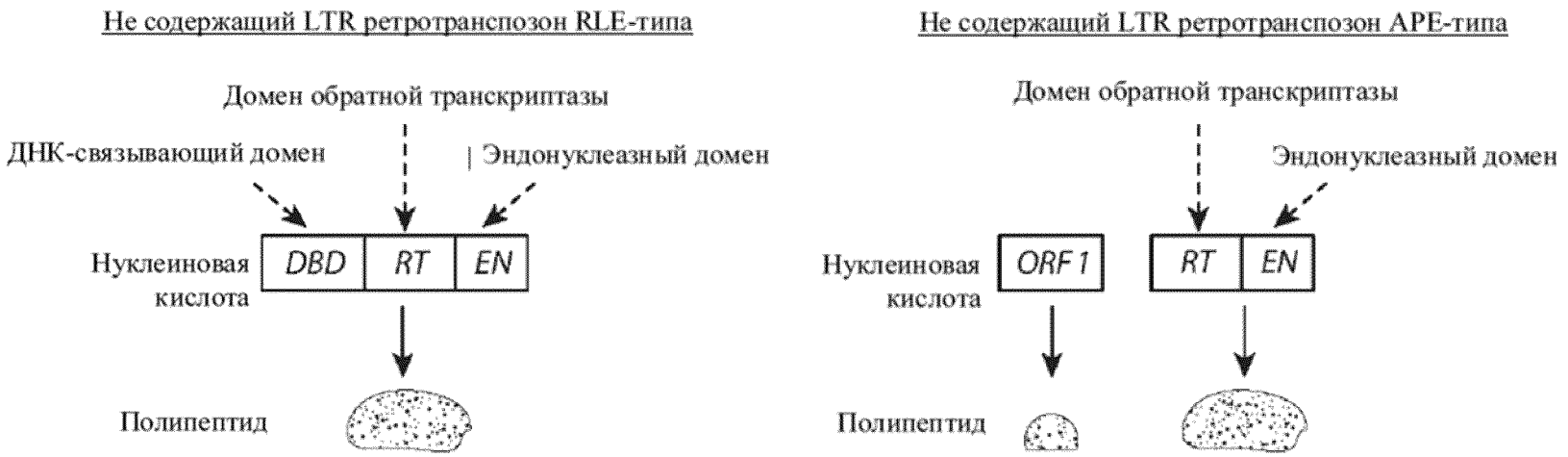
где домен RT имеет последовательность из таблицы 3B, 10 или 11 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

По доверенности



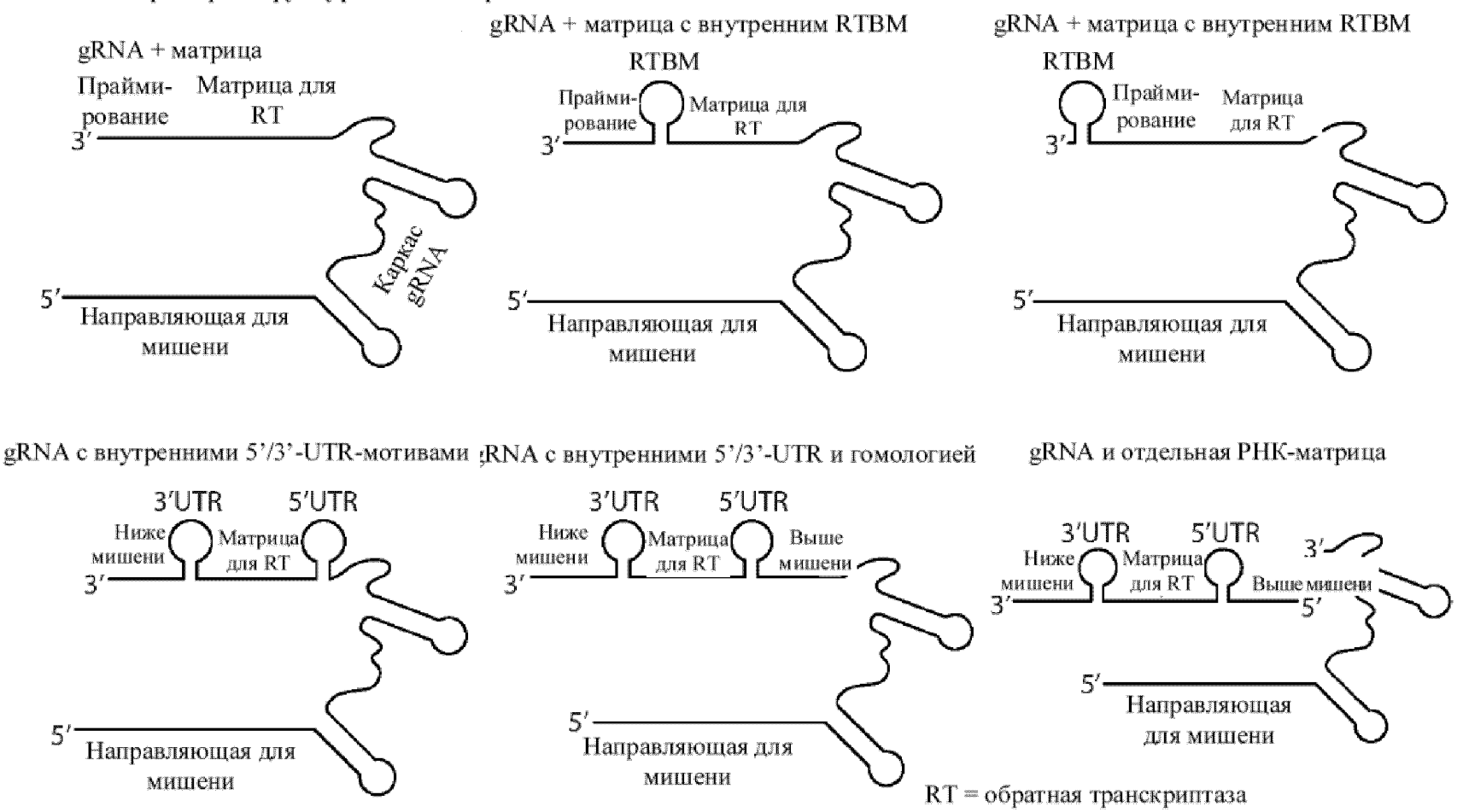
ФИГ. 1

5755561


Не содержащий LTR ретротранспозон RLE-типа
Домен обратной транскриптазы
ДНК-связывающий домен
Эндонуклеазный домен
Нуклеиновая кислота
DBD
RT
EN
Полипептид
Не содержащий LTR ретротранспозон APE-типа
Домен обратной транскриптазы
Эндонуклеазный домен
Нуклеиновая кислота
ORF 1
RT
EN
Полипептид

ФИГ. 2

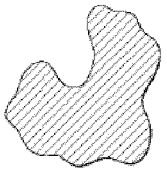
Примеры структуры РНК-матрицы Gene Writer



RT = обратная транскриптаза
RTBM = мотив связывания обратной транскриптазы
UTR = нетранслируемая область

ФИГ. 3

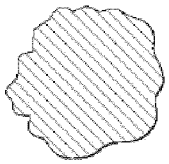
Модули Gene Writer



ДНК-связывающий
RTn
"Цинковый палец"
Домен TAL
Cas9, dCas9, никаза Cas9
Фактор транскрипции
Мегануклеаза



РНК-связывающий
Белок B-бокса
Белок оболочки MS2
Белок dCas
RTn



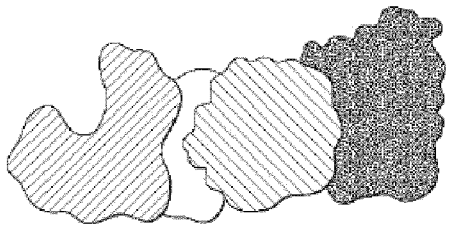
Обратная транскриптаза
RTn
Другая RT
Пептид, содержащий домен RT



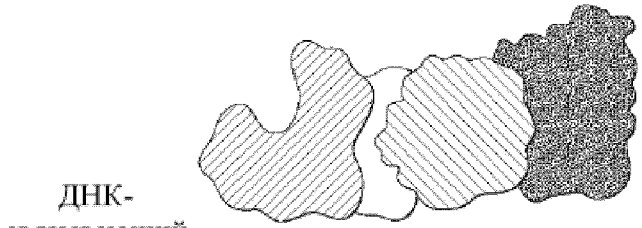
Эндонуклеаза
RTn
Cas9, никаза Cas 9
Фермент рестрикции
(например, типа II, например,
FokI)
Мегануклеаза
Резольваза для структур Холлидея
RLE
APE
GIY-YIG

ФИГ. 4A

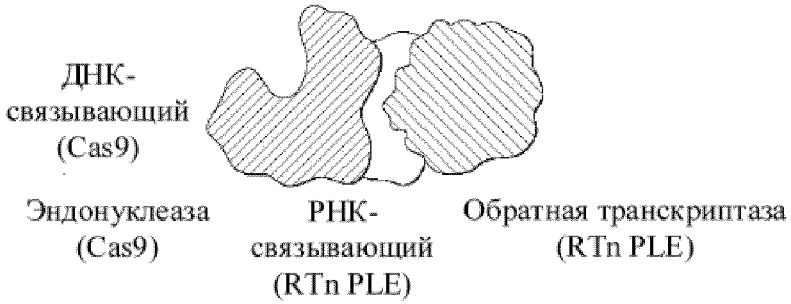
Gene Writer с ретротранспозазой дикого типа
(например, R2)



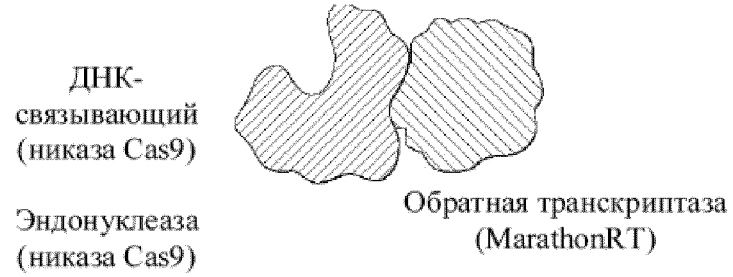
Gene Writer, все домены которого являются
гетерологичными
(например, ZF-R2-MarathonRT-FokI)



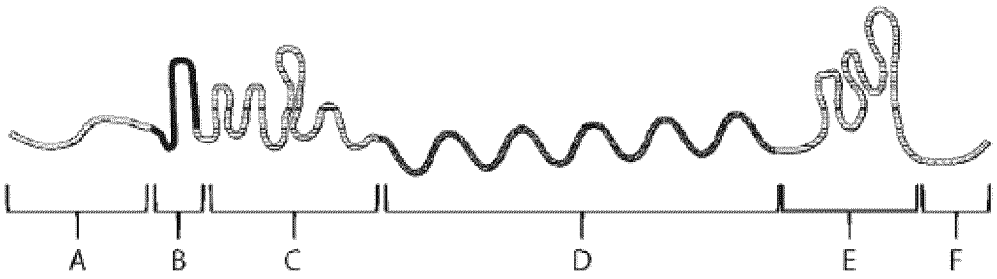
Gene Writer, перенацеленный с
помощью Cas9
(например, Cas9-PLE)



Gene Writer, перенацеленный с
помощью
Cas9 (например, Cas9-MarathonRT)



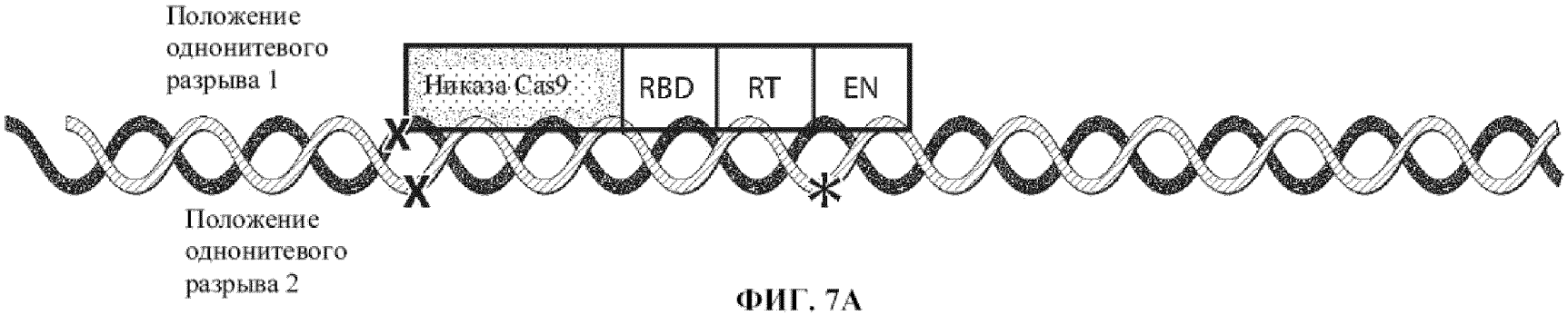
ФИГ. 4В



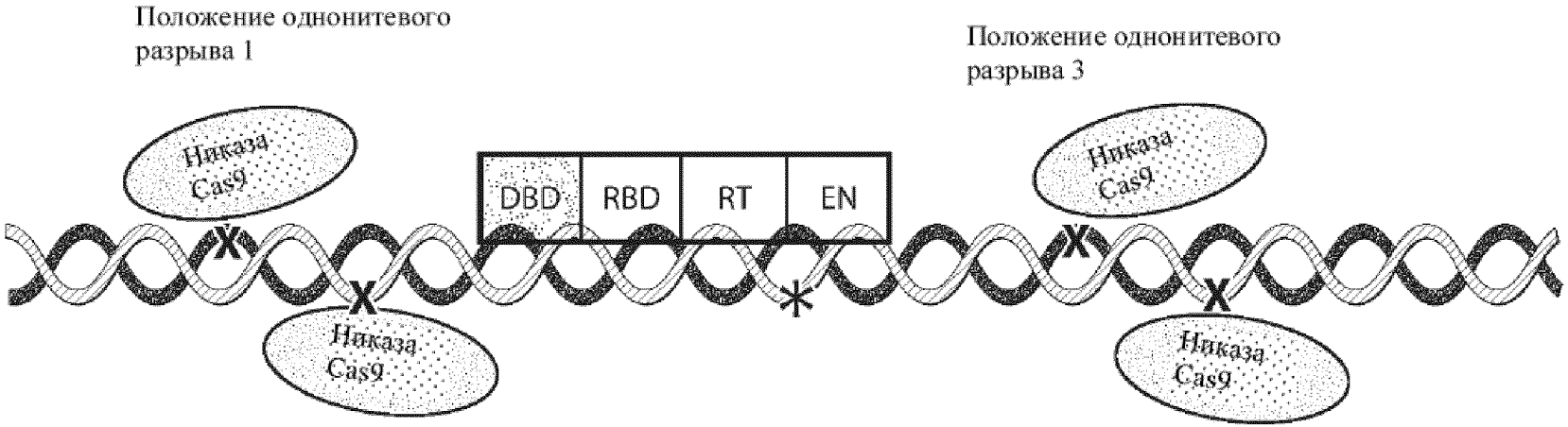
ФИГ. 5

Модуль	Функция
A: 5'-плечо гомологии	Модуль 5'-плечо гомологии является комплементарным последовательности ДНК с 5'-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени
B: Рибозим	Модуль рибозим является подчастью последовательности 5'-UTR ретротранспозазы
C: 5′-UTR	Модуль 5'-UTR представляет собой последовательность РНК или взаимодействующий компонент, который использует управляющий элемент для ретротранспозазы в процессе ретротранспозиции
D: Последовательность гетерологичного объекта	Последовательность гетерологичного объекта представляет собой последовательность РНК, которая служит в качестве матрицы для полипептида Gene Writer для встраивания желаемой полезной нагрузки в целевое местоположение генома
E: 3′-UTR	Модуль 3'-UTR представляет собой последовательность РНК или взаимодействующий компонент, с которым взаимодействует полипептид Gene Writer для связывания с молекулой РНК-матрицы, которая используется полностью и/или частично в качестве матрицы для ретротранспозиции
F: 3'-плечо гомологии	Модуль 3'-плечо гомологии является комплементарным последовательности ДНК с 3'-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени

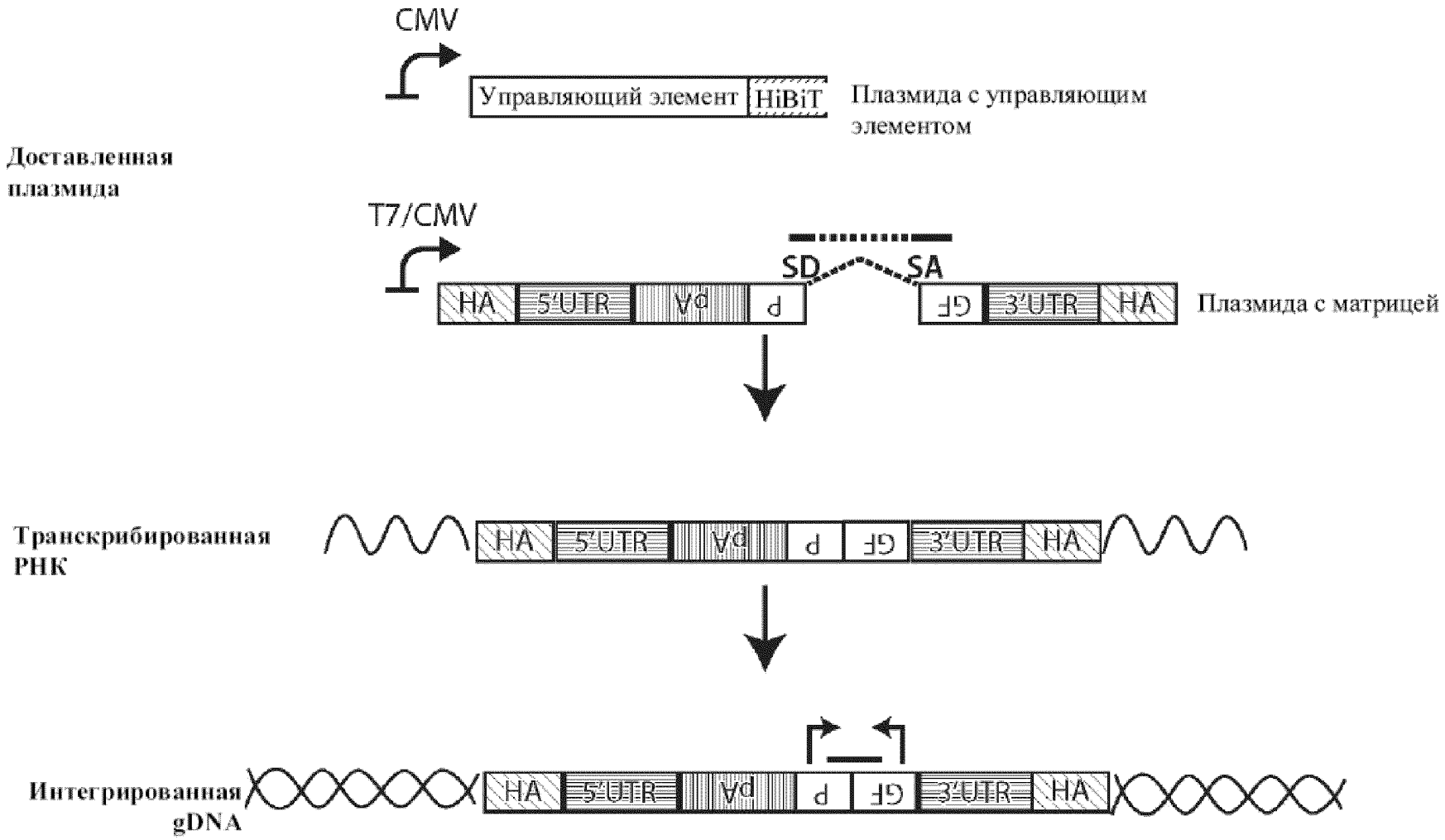
ФИГ. 6


Положение однонитевого разрыва 1
Никаза Cas9
RBD
RT
EN
Положение однонитевого разрыва 2

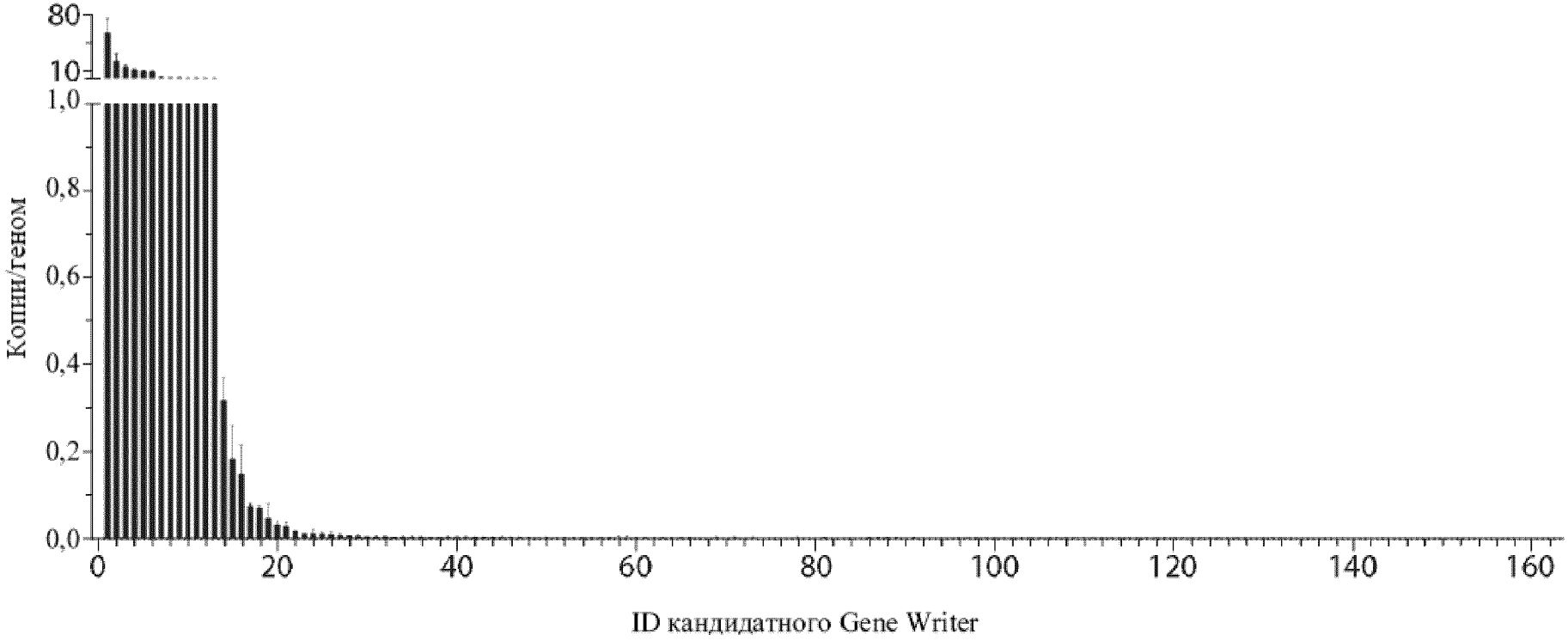
ФИГ. 7A


Положение однонитевого разрыва 1
Положение однонитевого разрыва 3
Никаза Cas9
DBD
RBD
RT
EN
Никаза Cas9
Никаза Cas9
Никаза Cas9
Положение однонитевого разрыва 2
Положение однонитевого разрыва 4

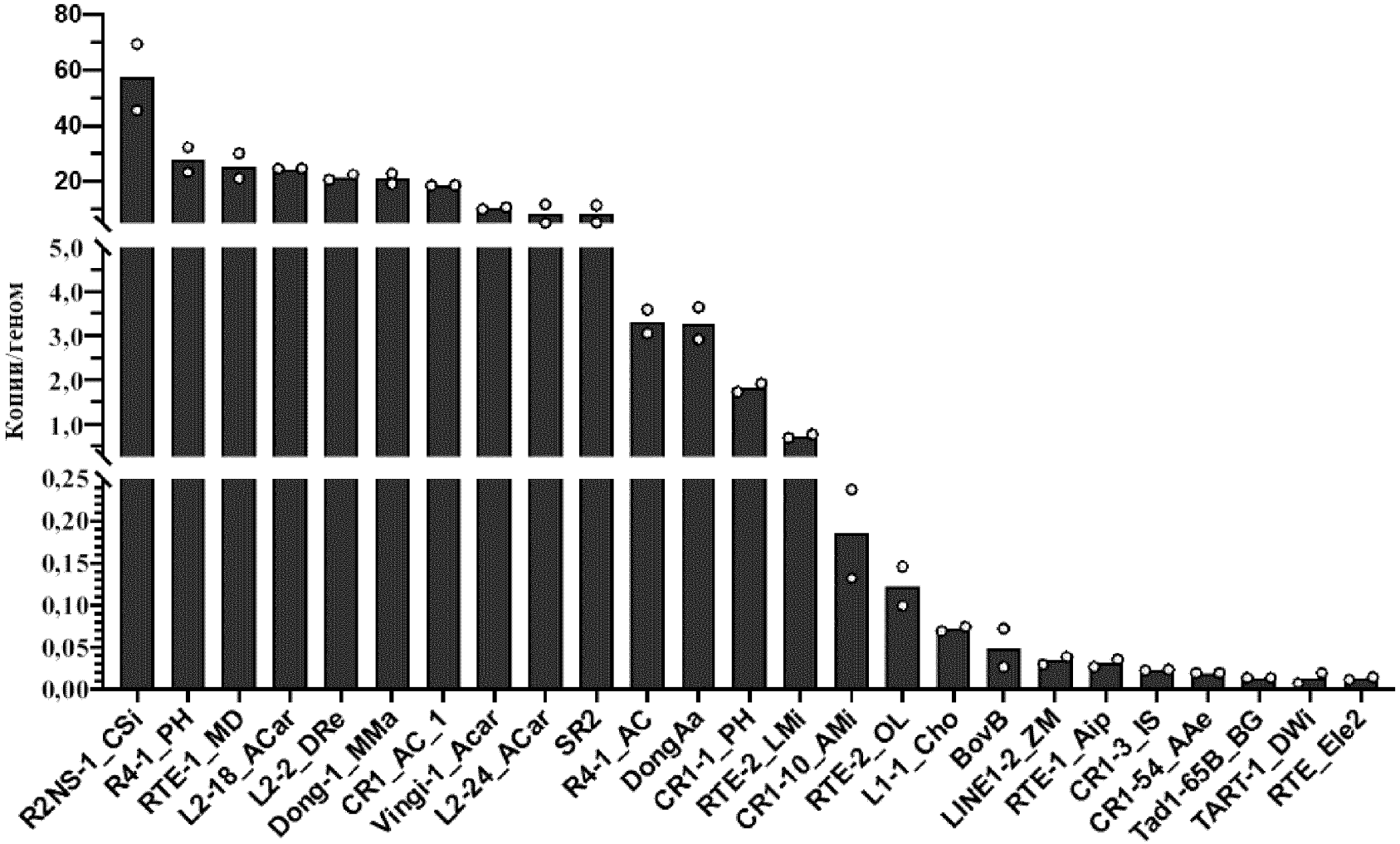
ФИГ. 7B


CMV
Управляющий элемент
HiBiT
Плазмида с управляющим элементом
Доставленная плазмида
T7/CMV
SD
SA
HA
5'UTR
pA
P
GF
3'UTR
HA
Плазмида с матрицей
Транскрибированная РНК
HA
5'UTR
pA
P
GF
3'UTR
HA
Интегрированная gDNA
HA
5'UTR
pA
P
GF
3'UTR
HA

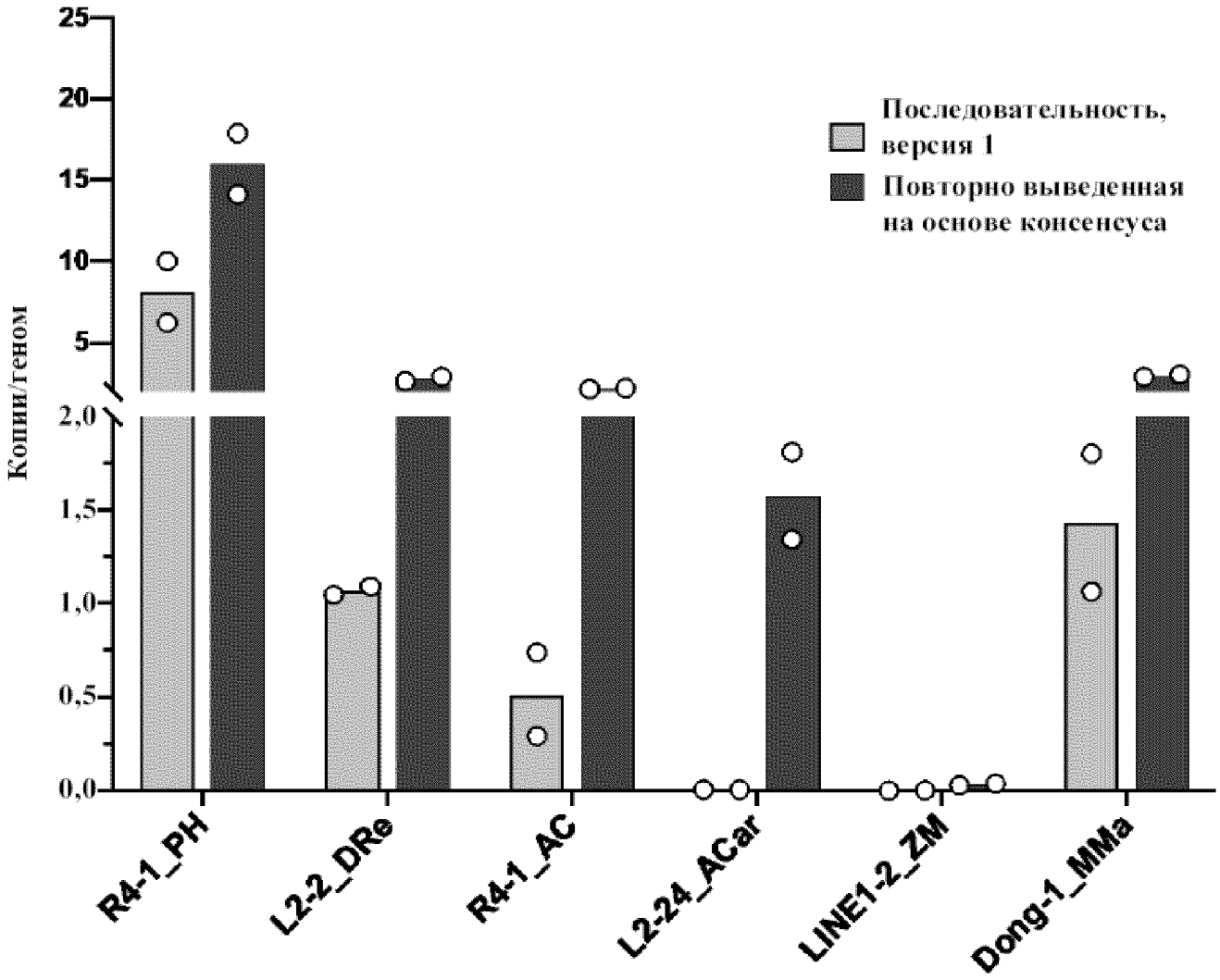
ФИГ. 8


80
10
1,0
0,8
0,6
0,4
0,2
0,0
Копии/геном
0
20
40
60
80
100
120
140
160
ID кандидатного Gene Writer

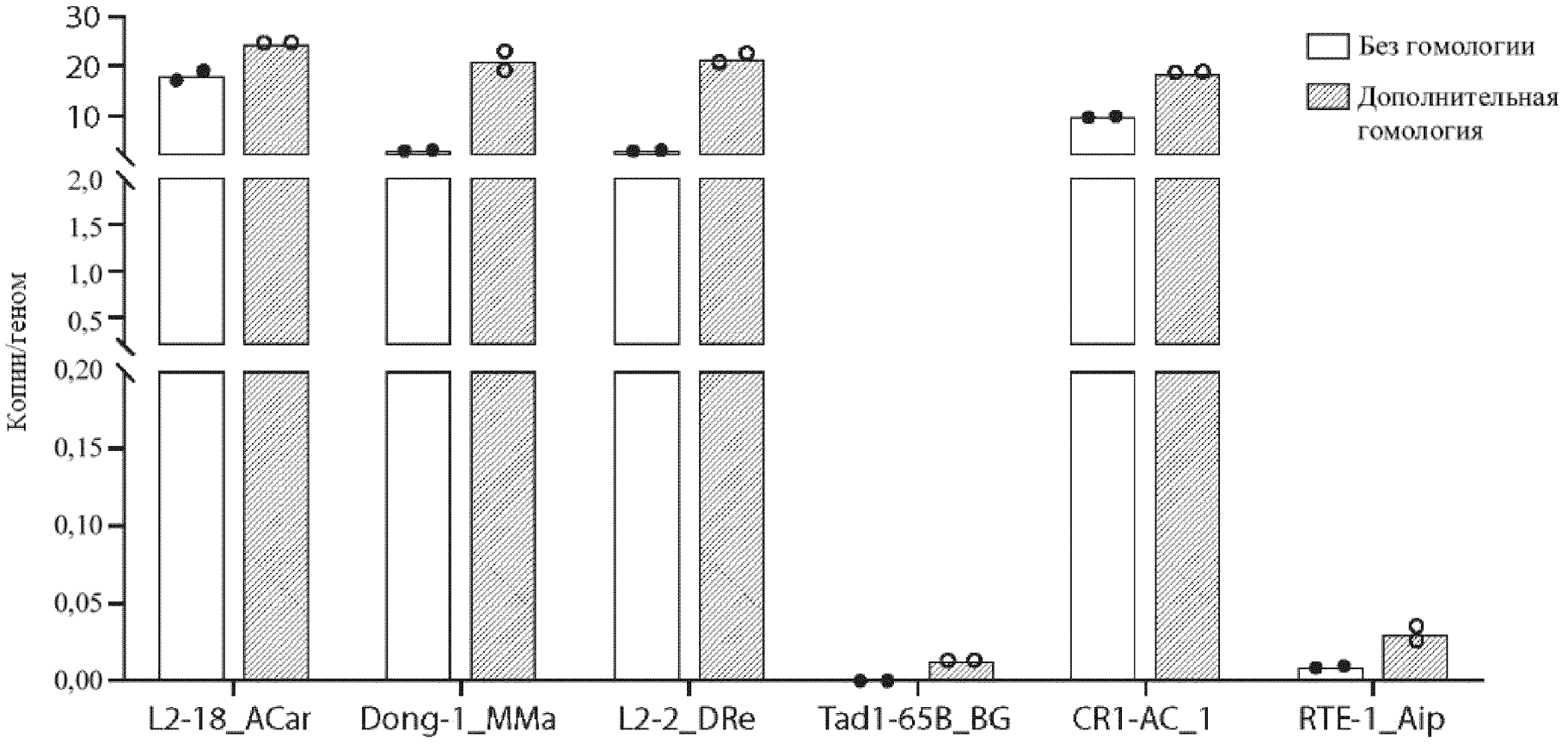
ФИГ. 9


Копии/геном
80
60
40
20
5,0
4,0
3,0
2,0
1,0
0,25
0,20
0,15
0,10
0,05
0,00
R2NS-1_CSi
R4-1_PH
RTE-1_MD
L2-18_ACar
L2-2_DRe
Dong-1_MMa
CR1_AC_1
Vingi-1_Acar
L2-24_ACar
SR2
R4-1_AC
DongAa
CR1-1_PH
RTE-2_LMi
CR1-10_AMi
RTE-2_OL
L1-1_Cho
BovB
LINE1-2_ZM
RTE-1_Aip
CR1-3_IS
CR1-54_AAe
Tad1-65B_BG
TART-1_DWi
RTE_Ele2

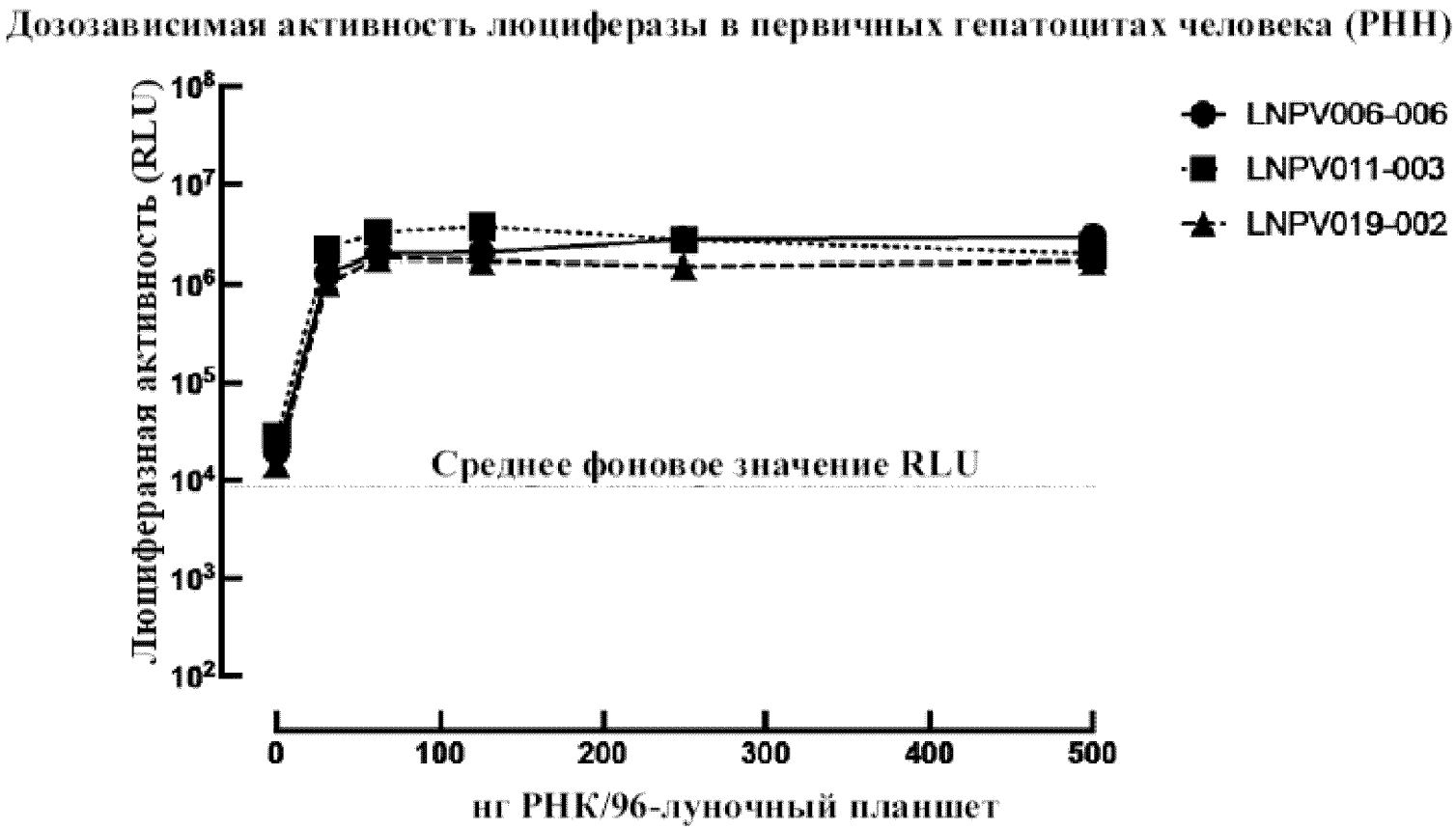
ФИГ. 10


Последовательность, версия 1
Повторно выведенная на основе консенсуса
Копии/геном
25
20
15
10
5
2,0
1,5
1,0
0,5
0,0
R4-1_PH
L2-2_DRe
R4-1_AC
L2-24_ACar
LINE1-2_ZM
Dong-1_MMa

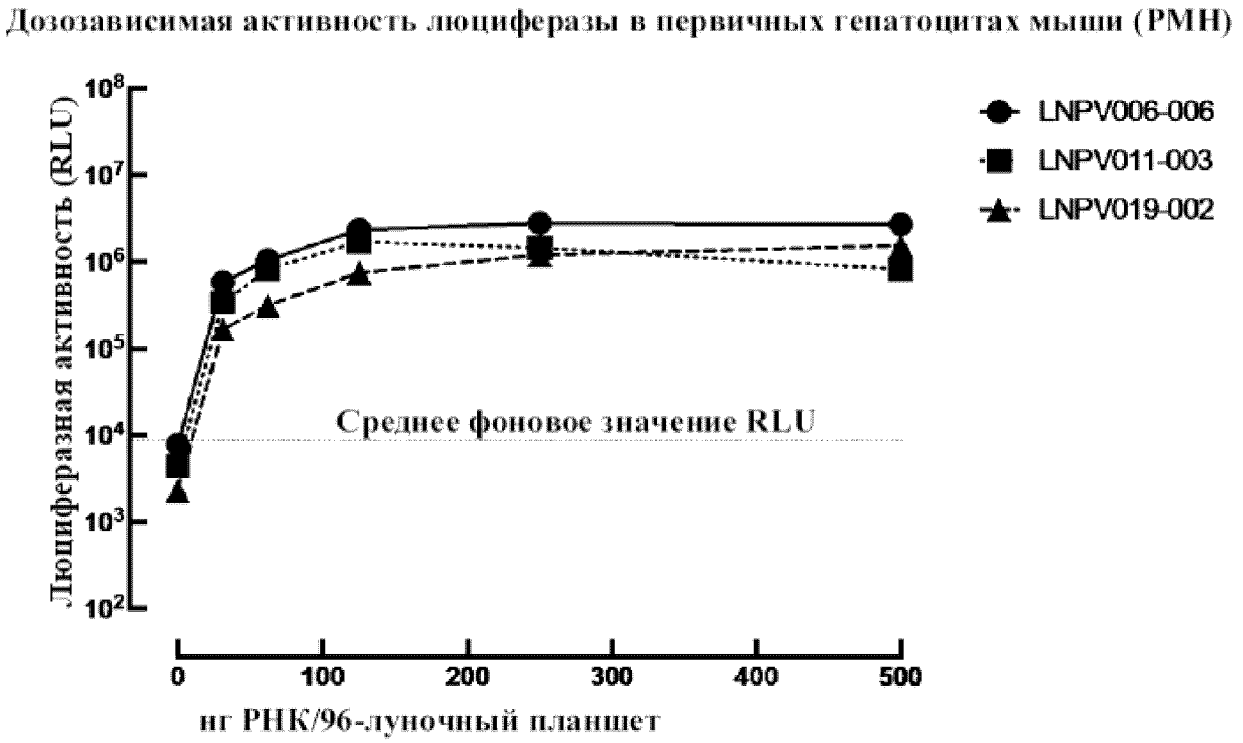
ФИГ. 11


Копии/геном
30
20
10
2,0
1,5
1,0
0,5
0,20
0,15
0,10
0,05
0,00
Без гомологии
Дополнительная гомология
L2-18_ACar
Dong-1_MMa
L2-2_DRe
Tad1-65B_BG
CR1-AC_1
RTE-1_Aip

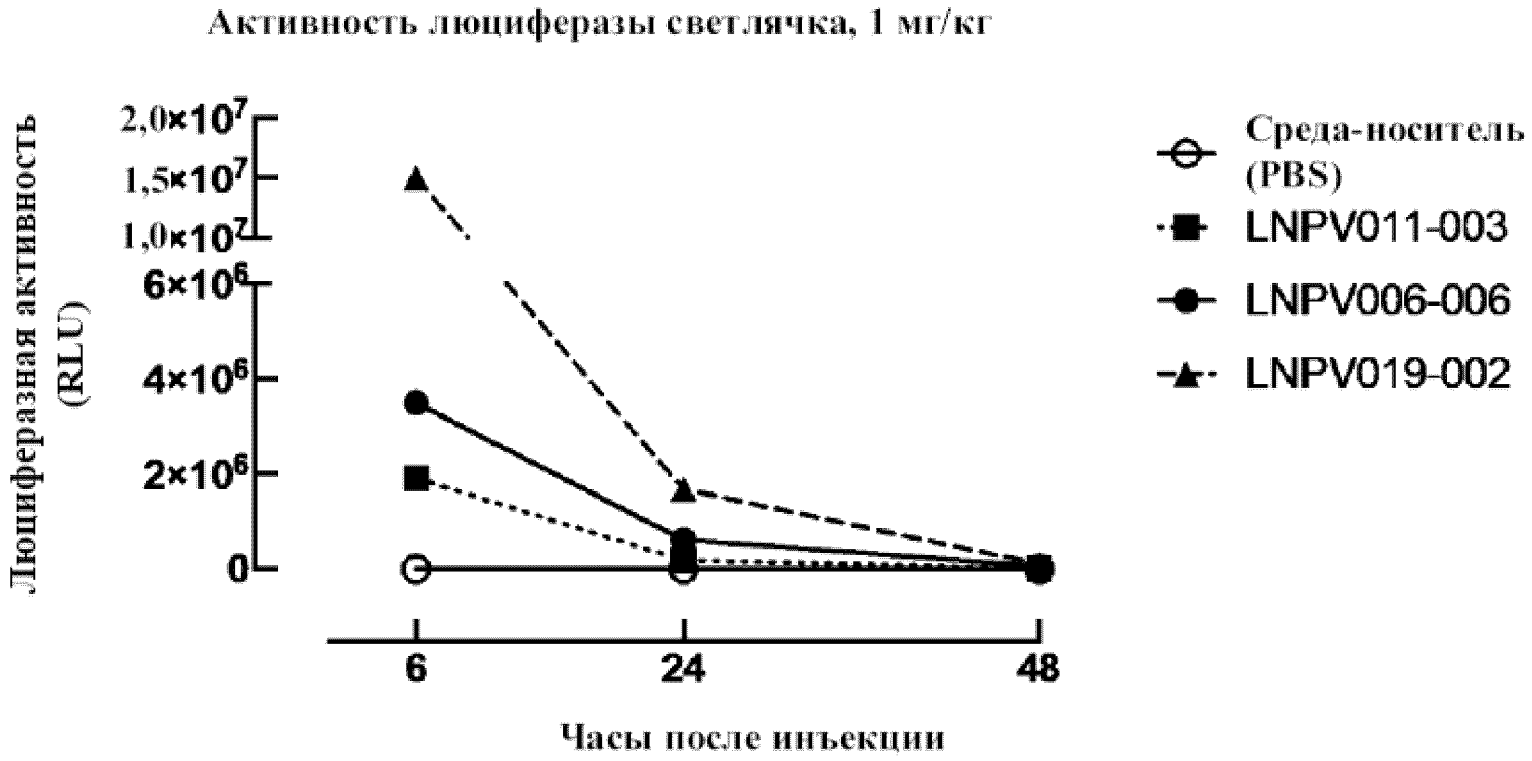
ФИГ. 12


Дозозависимая активность люциферазы в первичных гепатоцитах человека (PHH)
Люциферазная активность (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Среднее фоновое значение RLU
0
100
200
300
400
500
нг РНК/96-луночный планшет
LNPV006-006
LNPV011-003
LNPV019-002

ФИГ. 13A


Дозозависимая активность люциферазы в первичных гепатоцитах мыши (PMH)
Люциферазная активность (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Среднее фоновое значение RLU
0
100
200
300
400
500
нг РНК/96-луночный планшет
LNPV006-006
LNPV011-003
LNPV019-002

ФИГ. 13B


Активность люциферазы светлячка, 1 мг/кг
Люциферазная активность (RLU)
2,0×10^7
1,5×10^7
1,0×10^7
6×10^6
4×10^6
2×10^6
0
6
24
48
Часы после инъекции
Среда-носитель (PBS)
LNPV011-003
LNPV006-006
LNPV019-002

ФИГ. 14