

(19)



Евразийское
патентное
ведомство

(21) 202292525 (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 15/10 (2006.01)
C12N 15/113 (2010.01)
C12N 15/11 (2006.01)
C12N 15/90 (2006.01)

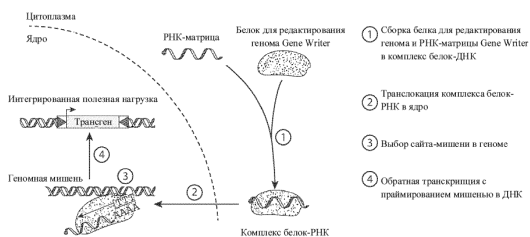
(54) СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА

(31) 62/985,285; 63/035,627; 63/067,828
(32) 2020.03.04; 2020.06.05; 2020.08.19
(33) US
(86) PCT/US2021/020948
(87) WO 2021/178720 2021.09.10
(88) 2021.11.11
(71) Заявитель:
ФЛЭГШИП ПАЙОНИРИНГ
ИННОВЕЙШНЗ VI, ЭлЭлСи (US)

(72) Изобретатель:
Ботмер Энн Хелен, Котта-Рамузино
Чечилия Джованна Сильвия,
Саломон Уильям Эдвард, Рубенс
Якоб Розенблюм, Сайторик Роберт
Джеймс, Ван Цзы Цзюнь, Ким Кюсик,
Котлар Рэнди Мичелл, Рэй Анания,
Альтшулер Роберт Чарльз, Кумар
Сандип, Рокет Натаниэль, Стейнберг
Барретт Итан (US)

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

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



202292525

A1

A1

202292525

СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА РОДСТВЕННЫЕ ЗАЯВКИ

Настоящая заявка заявляет приоритет по заявке на патент США под регистрационным номером 62/985285, поданной 4 марта 2020 года, заявке на патент США под регистрационным номером 63/035627, поданной 5 июня 2020 года, и заявке на патент США под регистрационным номером 63/067828, поданной 19 августа 2020 года, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

1. ПРЕДПОСЫЛКИ ИЗОБРЕТЕНИЯ

2. Интеграция представляющей интерес нуклеиновой кислоты в геном происходит с низкой частотой и небольшой сайт-специфичностью при отсутствии специализированного белка, предназначенного для содействия событию вставки. Некоторые существующие подходы, такие как CRISPR/Cas9, являются более подходящими для редактирования небольших участков, которое зависит от путей репарации хозяина, 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 существует потребность в улучшенных композициях (например, на основе белков и нуклеиновых кислот) и способах вставки, изменения или делеции представляющих интерес последовательностей в геноме.

3. 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, настоящее изобретение относится к композициям, системам и способам для вставки, изменения или делеции представляющих интерес последовательностей в геноме хозяина.

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

4. ПРОНУМЕРОВАННЫЕ ВАРИАНТЫ ОСУЩЕСТВЛЕНИЯ

1. Система для модификации ДНК, содержащая:

(а) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нукlease 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, вторую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

2. 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, нукlease 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.

3. 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, нукlease 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 домен RT содержит последовательность из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

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, нукlease 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 домен RT содержит последовательность из таблицы 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2,

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.

5. 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, нуклеазный 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 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.

6. 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, нуклеазный 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 система способна производить вставку в сайт-мишень размером по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40 или 44 нуклеотида.

7. 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, нуклеазный 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 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.

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, нуклеазный 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 последовательность гетерологичного объекта имеет длину по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70 или 73 нуклеотида.

9. 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.

10. 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, нуклеазный 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 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.

11. 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, нуклеазный 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 система способна производить делецию в сайте-мишени размером по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70 или 80 нуклеотидов.

12. 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, нуклеазный 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 система способна производить замены нуклеотидов, например, транзиции и/или трансверсии, в сайте-мишени размером по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов.

13. 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, нисказный 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, 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 (a)(ii) и/или (a)(iii) содержит домен TAL; домен "цинковый палец" или домен CRISPR/Cas, выбранный из таблицы 4, или их функциональный вариант (например, мутантную форму).

14. 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, нисказный 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, 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 эндонуклеазный домен, например, нисказный домен, разрезает как первую нить, так и вторую нить 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.

15. 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, нисказный 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) последовательность, которая

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

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

17. 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 с мишенью.

18. 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, вторую 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 с мишенью.

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

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

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

гибридизацию в 4x хлориде натрия/цитрате натрия (SSC) при приблизительно 65°C с последующей промывкой в 1x SSC при приблизительно 65°C.

22. Система согласно любому из предыдущих вариантов осуществления, где первый и второй конъюгирующие домены могут быть соединены ковалентно, например, с помощью лигирования с использованием шунтов, например, с помощью способа, описанного в Moore, M. J., & Query, C. C. *Methods in Enzymology*, 317, 109-123, 2000.

23. 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.

24. Система согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы (RT) получен из ретротранспозона или последовательности, характеризующейся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним.

25. Система для модификации ДНК, содержащая:

а) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT) из ретротранспозона или последовательности, характеризующейся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ним, (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, вторую нить сайта в геноме-мишени), (ii) необязательно последовательность, которая связывает полипептид, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

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

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

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

29. Система согласно любому из предыдущих вариантов осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, соответствующую домену обратной транскриптазы из любой из таблицы 30, таблицы 31, таблицы 41, таблицы 44 или таблицы 50, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или их функциональный фрагмент.

30. 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.

31. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая (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 эндонуклеазный и/или ДНК-связывающий домен полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

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

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

34. 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 связывает эндонуклеазный и/или ДНК-связывающий домен полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью,

где (i) содержит последовательность нуклеиновой кислоты, комплементарную последовательности гена из любой из таблиц 9-12 или имеющую не более 1, 2, 3, 4 или 5 отличий от указанной последовательности, характеризующейся указанной комплементарностью.

35. 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.

36. 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).

37. РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен RT содержит последовательность, выбранную из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, которая характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

38. РНК-матрица согласно любому из предыдущих вариантов осуществления, где домен RT содержит последовательность, выбранную из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, 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.

39. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность из (ii) специфично связывается с доменом RT.

40. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, представляет собой последовательность, например, последовательность UTR, из таблицы 1 или из домена из таблицы 2 или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

41. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (ii) последовательность, которая связывает эндонуклеазный 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.

42. 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 эндонуклеазный и/или ДНК-связывающий домен полипептида.

43. 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.

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

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

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

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

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

49. Система для модификации ДНК, содержащая:

(a) первую РНК-матрицу (или ДНК, кодирующую первую РНК-матрицу), содержащую (i) последовательность, которая связывает эндонуклеазный домен, например, нуклеазный домен, и/или ДНК-связывающий домен (DBD) полипептида, и (ii)

последовательность, которая связывает сайт-мишень (например, вторую 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 гетерологичного объекта и (iii) 3'-концевой домен гомологии с мишенью.

50. 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.

51. 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.

52. Система согласно любому из предыдущих вариантов осуществления, 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.

53. Система согласно любому из предыдущих вариантов осуществления, 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.

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

55. Система согласно любому из предыдущих вариантов осуществления, где одна или обе из РНК-матрицы и РНК, кодирующей полипептид (a), содержат химически модифицированную мРНК, например, мРНК, содержащую химически модифицированное основание, например, мРНК, содержащую 5-метоксиуридин.

56. Система согласно любому из предыдущих вариантов осуществления, где одна или обе из РНК-матрицы и РНК, кодирующей полипептид (a), содержат химически модифицированную РНК, например, РНК, содержащую химически модифицированное основание, например, РНК, содержащую 2'-О-метилфосфотиоат.

57. Система согласно любому из предыдущих вариантов осуществления, где одна или обе из РНК-матрицы и РНК, кодирующей полипептид (a), содержат химически модифицированную РНК, например, РНК, содержащую химически модифицированное основание, например, 2'-О-метилфосфотиоат, по одному или обоим из 3, 4 или 5 оснований на 5'- или 3'-конце РНК.

58. 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).

59. 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 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

60. 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 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2, 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.

61. Полипептид согласно любому из предыдущих вариантов осуществления, где полипептид кодируется мРНК, например, химически модифицированной мРНК, например, мРНК, содержащей химически модифицированное основание, например, мРНК, содержащей 5-метоксиуридин.

62. Полипептид согласно любому из предыдущих вариантов осуществления, где полипептид кодируется мРНК, например, химически модифицированной мРНК, например, мРНК, содержащей химически модифицированное основание, например, мРНК, содержащей N1-метилпсевдоуридин.

63. 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 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними; 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.

64. 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 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2, 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; 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, нисказный домен.

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

66. 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.

67. Реакционная смесь, содержащая:

клетку и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, согласно любому из предыдущих вариантов осуществления.

68. Реакционная смесь, содержащая:

ДНК, содержащую сайт-мишень, и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, согласно любому из предыдущих вариантов осуществления.

69. Набор, содержащий:

систему, полипептид, РНК-матрицу или ДНК, кодирующую их, согласно любому из предыдущих вариантов осуществления;

инструкции по применению системы, полипептида, РНК-матрицы или ДНК, кодирующей их; а также

одну или обе из клетки или ДНК, содержащих сайт-мишень.

70. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD содержит домен TAL.

71. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD содержит домен "цинковый палец".

72. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD содержит домен CRISPR/Cas.

73. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен представляет собой нисказный домен.

74. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен содержит домен CRISPR/Cas.

75. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas содержит домен или полипептид из таблицы 4 или его функциональный вариант (например, мутантную форму).

76. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas содержит домен или полипептид из рода/вида из таблицы 4.

77. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен содержит нуклеазу типа II (например, FokI), резольвазу для структур Холлидея или нуклеазу для двухнитевой ДНК, содержащую изменение, которое приводит к устранению ее способности разрезать одну нить (например, превращающее нуклеазу для двухнитевой ДНК в никазу).

78. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT содержит обратную транскриптазу или ее функциональный фрагмент или вариант, выбранный из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2.

79. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT содержит одну или несколько мутаций (например, вставку, делецию или замену) по сравнению со встречающимся в природе доменом RT, или доменом RT или функциональным фрагментом, выбранным из таблицы 1 или 3, или последовательностью домена обратной транскриптазы из таблицы 2, или перечнем последовательностей SEQ ID NO: 1-67 из WO2018089860A1, включенной в данный документ посредством ссылки.

80. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где одна или несколько мутаций выбраны из D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K или D653N в домене RT обратной транскриптазы вируса лейкоза мышей или соответствующей мутации в соответствующем положении другого домена RT.

81. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где одна или несколько мутаций выбраны из WO2018089860A1, включенной в данный документ посредством ссылки (например, мутации C952S, и/или C956S, и/или C952S, C956S (двойная мутация), и/или C969S, и/или H970Y, и/или R979Q, и/или R976Q, и/или R1071S, и/или R328A, и/или R329A, и/или Q336A, и/или R328A, R329A, Q336A (тройная мутация), и/или G426A, и/или D428A, и/или G426A, D428A (двойная мутация), и/или любой их комбинации; где положения указаны относительно SEQ ID NO: 52 из WO2018089860A1), в домене RT ретротранспозазы R2Bm или соответствующей мутации в соответствующем положении другого домена RT.

82. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD и/или эндонуклеазный домен (например, домен CRISPR/Cas) содержит домен или полипептид из таблицы 4 или его функциональный вариант (например, мутантную форму).

83. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD и/или эндонуклеазный домен (например, домен CRISPR/Cas) содержит домен или полипептид из таблицы 4.

84. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT и DBD и/или эндонуклеазный домен (например, домен CRISPR/Cas) слиты посредством пептидного линкера, например, линкера из таблицы 42.

85. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где длина линкера составляет приблизительно 6-18, 8-16, 10-14 или 12 аминокислот.

86. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где линкер содержит глицин и серин, например, при этом линкер содержит только остатки глицина и серина, например, где линкер содержит последовательность GSSGSS.

87. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где линкер содержит последовательность в соответствии с таблицей 42, например, линкер 10, раскрытый в таблице 42, или последовательность, имеющую не более 1, 2 или 3 замен по сравнению с ней.

88. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas содержит Cas9, например, Cas9 дикого типа или никазу Cas9.

89. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT расположен с С-концевой стороны от DBD в полипептиде.

90. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT расположен с С-концевой стороны от никазного домена в полипептиде.

91. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT расположен с N-концевой стороны от DBD в полипептиде.

92. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен RT расположен с N-концевой стороны от никазного домена в полипептиде.

93. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид содержит линкер, например, расположенный между доменом RT и доменом DBD или доменом RT и никазным доменом.

94. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где длина линкера составляет 2-50, например, 2-30, аминокислот.

95. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где линкер представляет собой гибкий линкер, например, содержащий остатки Gly и/или Ser.

96. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где 3'-концевой домен гомологии с мишенью комплементарен последовательности, прилегающей к сайту, подлежащему модификации системой, или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной последовательности, прилегающей к сайту, подлежащему модификации системой.

97. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где 3'-концевой домен гомологии с мишенью имеет длину более 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, 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 (например, имеет длину 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 или 30 нуклеотидов).

98. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где 3'-концевой домен гомологии с мишенью имеет длину не более 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, 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.

99. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением положения или положений, подлежащих модификации.

100. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением положений, кодирующих последовательность, подлежащую вставке в сайт.

101. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением того, что последовательность гетерологичного объекта не содержит нуклеотидов, кодирующих последовательность, подлежащую делеции в сайте.

102. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта имеет длину более 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 или 30 нуклеотидов (например, имеет длину 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 или 30 нуклеотидов).

103. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта

имеет длину не более 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 или 30 нуклеотидов.

104. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта приводит к замене по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидами нуклеотидов не из сайта-мишени.

105. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта приводит к вставке 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 или 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 в сайт-мишень.

106. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта приводит к делеции 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, 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов.

107. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

108. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), имеет длину по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 85 или 80 нуклеотидов).

109. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает полипептид, связывает эндонуклеазный домен и/или домен DBD.

110. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает полипептид, содержит последовательность, соответствующую одной или обоим из предсказанной 5'-UTR и предсказанной 3'-UTR из таблицы 3 или таблицы 41, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или их функциональный фрагмент.

111. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает

полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), содержит gRNA.

112. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), имеет длину по меньшей мере 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21 или 20 нуклеотидов), например, имеет длину 17, 18, 19, 20, 21, 22, 23 или 24 нуклеотида.

113. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт-мишень, комплементарна второй нити сайта-мишени или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной второй нити сайта-мишени.

114. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

115. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, дополнительно содержащие gRNA, нацеленную на вторую нить, которая направляет эндонуклеазный домен (например, никазный) домен для осуществления одонитевого разрыва второй нити (например, в геноме-мишени).

116. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит gRNA, нацеленную на вторую нить.

117. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где gRNA, нацеленная на вторую нить, расположена в нуклеиновой кислоте, отдельной от РНК-матрицы.

118. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где gRNA направляет эндонуклеазный домен (например, никазный домен) для осуществления одонитевого разрыва второй нити (например, в геноме-мишени) в сайте, который находится 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, одонитевого разрыва в первой нити).

119. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где gRNA специфично связывается с редактируемой нитью.

120. Система, способ, набор, РНК-матрица или реакционная смесь 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).

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

122. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид не содержит функционального эндогенного нацеливающего домена (например, где полипептид не содержит эндогенного нацеливающего домена).

123. Система, способ, набор, РНК-матрица или реакционная смесь 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).

124. Система, способ, набор, РНК-матрица или реакционная смесь 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).

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

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

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

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

129. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен содержит домен FokI.

130. Система, способ, набор, РНК-матрица или реакционная смесь 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).

131. Система, способ, набор, РНК-матрица или реакционная смесь 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.

132. Система, способ, набор, РНК-матрица или реакционная смесь 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.

133. Система, способ, набор, РНК-матрица или реакционная смесь 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 сайта одонитевого разрыва (например, создаваемого с помощью никазы, например, эндонуклеазного домена, например, как описано в данном документе) в молекуле ДНК-мишени.

134. Система, способ, набор, РНК-матрица или реакционная смесь 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.

135. Система, способ, набор, РНК-матрица или реакционная смесь 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).

136. Система, способ, набор, РНК-матрица или реакционная смесь 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 одонитевого разрыва молекулы ДНК-мишени (например, сайта, подвергаемого одонитевому разрыву никазой, например, эндонуклеазным доменом, как описано в данном документе).

137. Система, способ, набор, РНК-матрица или реакционная смесь 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).

138. Система, способ, набор, РНК-матрица или реакционная смесь 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.

139. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит направляющую РНК (gRNA), например, как описано в данном документе.

140. Система, способ, набор, РНК-матрица или реакционная смесь 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).

141. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК системы (например, РНК-матрица, РНК, кодирующая полипептид из (а), или РНК, экспрессирующаяся из последовательности гетерологичного объекта, интегрированной в ДНК-мишень) содержит сайт связывания с микроРНК, например, в 3'-UTR.

142. 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.

143. 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.

145. 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.

146. 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.

147. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК, кодирующая полипептид из (а), 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.

148. 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.

149. Система, содержащая:

мРНК, кодирующую полипептид или систему согласно любому из предыдущих вариантов осуществления, и

РНК-матрицу согласно любому из предыдущих вариантов осуществления.

150. Система согласно любому из предыдущих вариантов осуществления, где мРНК, кодирующая полипептид или систему согласно любому из предыдущих вариантов осуществления, и РНК-матрица согласно любому из предыдущих вариантов осуществления расположены в разных молекулах нуклеиновой кислоты.

151. Система, содержащая молекулу РНК, содержащая:

РНК-матрицу (или РНК, кодирующую РНК-матрицу) согласно любому из предыдущих вариантов осуществления и

последовательность, кодирующую систему или полипептид согласно любому из предыдущих вариантов осуществления.

152. Система согласно любому из предыдущих вариантов осуществления, где молекула РНК содержит участок внутренней посадки рибосомы, например, функционально связанный с последовательностью, кодирующей систему или полипептид

153. Система согласно любому из предыдущих вариантов осуществления, где молекула РНК содержит сайт расщепления, например, расположенный между РНК-матрицей (или РНК, кодирующей РНК-матрицу) и последовательностью, кодирующей систему или полипептид.

154. Система или полипептид согласно любому из предыдущих вариантов осуществления, где полипептид содержит расщепленный интеин, например, два или несколько (например, все) из домена RT, DBD, эндонуклеазного (например, никазного) домена или их комбинации транслируются как отдельные белки, которые объединяются в один полипептид путем сплайсинга белков.

155. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система содержит одну или несколько кольцевых молекул РНК (circRNA).

156. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA кодирует полипептид Gene Writer.

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

158. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA доставляется в клетку-хозяина.

159. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA способна линеаризоваться, например, в клетке-хозяине, например, в ядре клетки-хозяина.

160. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA содержит сайт расщепления.

161. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA дополнительно содержит второй сайт расщепления.

162. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где сайт расщепления может быть расщеплен рибозимом, например, рибозимом, содержащимся в circRNA (например, посредством саморасщепления).

163. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где circRNA содержит последовательность рибозима.

164. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность рибозима способна к саморасщеплению, например, в клетке-хозяине, например, в ядре клетки-хозяина.

165. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где рибозим представляет собой индуцируемый рибозим.

166. Система, набор, полипептид или реакционная смесь 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.

167. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где рибозим представляет собой рибозим, отвечающий на нуклеиновые кислоты.

168. Система, набор, полипептид или реакционная смесь 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).

169. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где рибозим отвечает на белок-мишень (например, белок оболочки MS2).

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

171. Система, набор, полипептид или реакционная смесь 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.

172. Система, набор, полипептид или реакционная смесь 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.

173. Система, набор, полипептид или реакционная смесь 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.

174. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где рибозим активируется компонентом, экспрессируемым в клетке-мишени или ткани-мишени.

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

176. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где рибозим содержится в кольцевой РНК или линейной РНК.

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

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

178. Система согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид (а), содержит кодирующую

последовательность, которая является кодон-оптимизированной для экспрессии в клетках человека.

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

180. Липидная наночастица (LNP), содержащая систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, согласно любому из предыдущих вариантов осуществления.

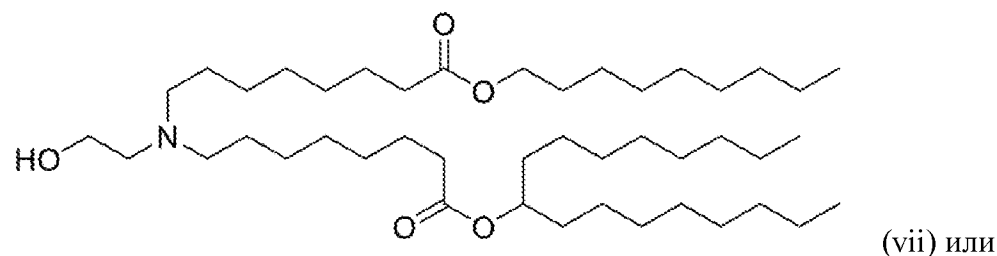
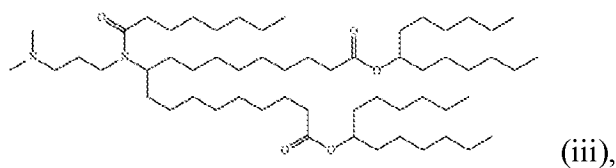
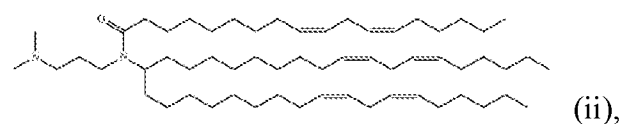
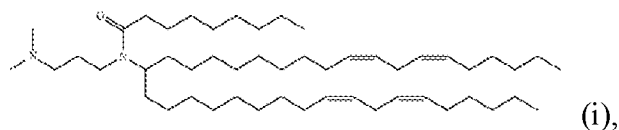
181. Система, содержащая первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы Gene Writing (например, описанной в данном документе); и

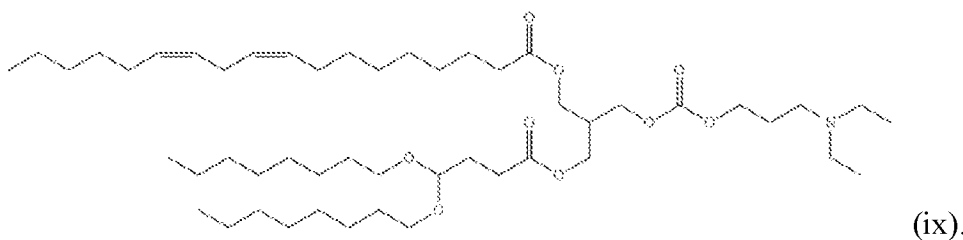
вторую липидную наночастицу, содержащую молекулу нуклеиновой кислоты системы Gene Writing (например, описанной в данном документе).

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

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

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





(ix).

185. 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.

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

187. Система, набор, полипептид или реакционная смесь 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.

188. Система, набор, полипептид или реакционная смесь 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.

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

190. Система, набор, полипептид или реакционная смесь 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%.

191. Система, набор, полипептид или реакционная смесь 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%.

192. Система, набор, полипептид или реакционная смесь 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).

193. Система, набор, полипептид или реакционная смесь 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).

194. Система, набор, полипептид или реакционная смесь 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).

195. Система, набор, полипептид или реакционная смесь 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%.

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

197. Система, набор, полипептид или реакционная смесь 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).

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

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

200. Система, набор, полипептид или реакционная смесь 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%.

201. Система, набор, полипептид или реакционная смесь 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%.

202. Система, набор, полипептид или реакционная смесь 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).

203. Система, набор, полипептид или реакционная смесь 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).

204. Система, набор, полипептид или реакционная смесь 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).

205. Система, набор, полипептид или реакционная смесь 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 26.

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

207. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество разновидностей альдегидов определяется посредством обнаружения 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 41.

208. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида обнаруживаются посредством определения присутствия одного или нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 41.

209. 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.

210. 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).

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

212. Система, содержащая:

первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы или полипептид согласно любому из предыдущих вариантов осуществления; и

вторую липидную наночастицу, содержащую РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу согласно любому из предыдущих вариантов осуществления.

213. Вирус, вирусоподобная частица, фузосома или виросома, содержащие систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, согласно любому из предыдущих вариантов осуществления.

214. Система, содержащая:

первые вирус, вирусоподобную частицу, фузосому или виросому, содержащие полипептид (или кодирующую его ДНК или РНК) системы или полипептид согласно любому из предыдущих вариантов осуществления; и

вторые вирус, вирусоподобную частицу, фузосому или виросому, содержащие РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу согласно любому из предыдущих вариантов осуществления.

215. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы имеет длину более 100, 125, 150, 175 или 200 нуклеотидов или имеет длину по меньшей мере 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 имеет длину менее 15, 10, 5 или 20 тысяч оснований или имеет длину менее 500, 400, 300 или 200 нуклеотидов).

216. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит поли(А)-хвост (например, поли(А)-хвост, который имеет длину по меньшей мере 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов).

217. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы 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)), и/или

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-модифицированных NTP; химической модификации одного нуклеотида (например, окисления 3'-концевой рибозы до реакционноспособного альдегида с последующей конъюгацией альдегид-реактивного модифицированного нуклеотида) или химического лигирования с другой молекулой нуклеиновой кислоты.

218. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где РНК-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, C5-пропинилдезоксцитидина (pdC), C5-пропинилдезоксйридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдезоксйридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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).

219. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит один или несколько модифицированных нуклеотидов.

220. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где после теста на стабильность по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы остаются интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов или длиной по меньшей мере 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).

221. Система, набор или реакционная смесь согласно любому из предыдущих вариантов осуществления, где по меньшей мере 1% сайтов-мишеней являются модифицированными после анализа системы на активность.

222. Система, набор, РНК-матрица, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система, полипептид, РНК-матрица и/или ДНК, кодирующая их, составлены в виде липидной наночастицы (LNP).

223. Система, набор, РНК-матрица, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где ДНК, кодирующая систему, полипептид и/или РНК-матрица упакованы в вирус, вирусоподобную частицу, виросому, липосому, везикулу, экзосому или LNP.

224. Система, набор, РНК-матрица, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где ДНК, кодирующая систему, РНК-матрица или полипептид упакованы в аденоассоциированный вирус (AAV).

225. Система, набор, РНК-матрица, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система, РНК-матрица, полипептид, липидная наночастица (LNP), вирус, вирусоподобная частица или виросома свободны или по сути свободны от контаминации пирогенами, вирусами, грибами, бактериальными патогенами и/или белками клетки-хозяина.

226. Вирус, вирусоподобная частица или виросома, содержащие:
систему, РНК-матрицу или полипептид согласно любому из предыдущих вариантов осуществления или ДНК, кодирующую любое из них, и
капсидный белок аденоассоциированного вируса (AAV).

227. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома согласно любому из предыдущих вариантов осуществления, где система, РНК-матрица и/или полипептид активны в ткани-мишени и менее активны (например, не активны) в ткани, не являющейся мишенью.

228. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома согласно любому из предыдущих вариантов осуществления, дополнительно

содержащие одну или несколько первых тканеспецифических последовательностей, контролирующей экспрессию, специфичных для ткани-мишени, где одна или несколько первых тканеспецифических последовательностей, контролирующей экспрессию, специфичных для ткани-мишени, находятся в функциональной ассоциации с РНК-матрицей, полипептидом или нуклеиновой кислотой, кодирующей его, или ими обоими.

229. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен, например, нисказный домен, осуществляет односторонний разрыв первой нити ДНК сайта-мишени и осуществляет односторонний разрыв второй нити в сайте, находящемся на расстоянии от первого одностороннего разрыва.

230. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где односторонние разрывы производятся в ориентации наружу.

231. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где односторонние разрывы производятся в ориентации наружу.

232. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления,

где последовательность, которая связывает сайт-мишень, определяет местоположение одностороннего разрыва в первой нити,

где система дополнительно содержит дополнительную нуклеиновую кислоту, содержащую последовательность, которая связывает сайт на расстоянии от сайта-мишени, и где последовательность, которая связывает сайт на расстоянии от сайта-мишени, определяет местоположение одностороннего разрыва во второй нити.

233. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительная нуклеиновая кислота дополнительно содержит последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или DBD), например, где дополнительная нуклеиновая кислота содержит gRNA.

234. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт на расстоянии от сайта-мишени (например, связывается с первой нитью сайта в геноме-мишени), имеет длину по меньшей мере 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21 или 20 нуклеотидов), например, имеет длину 17, 18, 19, 20, 21, 22, 23 или 24 нуклеотида.

235. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт на

расстоянии от сайта-мишени, комплементарна первой нити сайта-мишени или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с первой нитью сайта-мишени.

236. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD и/или эндонуклеазный домен содержат домен CRISPR/Cas.

237. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas и РНК-матрица связываются с сайтом-мишенью, и где первая нить сайта-мишени содержит первый сайт РАМ.

238. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas и дополнительная нуклеиновая кислота связывается с сайтом на расстоянии от сайта-мишени, и где вторая нить сайта на расстоянии от сайта-мишени содержит второй сайт РАМ.

239. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый сайт РАМ и второй сайт РАМ расположены между местоположением одонитевого разрыва в первой нити и местоположением одонитевого разрыва во второй нити.

240. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где местоположение одонитевого разрыва в первой нити и местоположение одонитевого разрыва во второй нити расположены между первым сайтом РАМ и вторым сайтом РАМ.

241. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, дополнительно содержащие дополнительный полипептид, содержащий дополнительный ДНК-связывающий домен (DBD) и дополнительный эндонуклеазный домен, например, дополнительный никазный домен.

242. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительный эндонуклеазный домен, например, дополнительный никазный домен, содержит эндонуклеазный или никазный домен, описанный в данном документе, например, домен CRISPR/Cas, нуклеазу типа II (например, FokI), резольвазу для структур Холлидея, мегануклеазу или нуклеазу для двухнитевой ДНК, содержащую изменение, которое приводит к устранению ее способности осуществлять одонитевой разрыв в одной нити (например, превращающее нуклеазу для двухнитевой ДНК в никазу).

243. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительный DBD связывает сайт на расстоянии от сайта-мишени.

244. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен (а) или (b) осуществляет одонитевой разрыв первой нити, а дополнительный эндонуклеазный домен (например, дополнительный никазный домен) осуществляет одонитевой разрыв второй нити.

245. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где одонитевые разрывы производятся в ориентации наружу.

246. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где одонитевые разрывы производятся в ориентации внутрь.

247. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где DBD и необязательно РНК-матрица (например, последовательность, которая связывает полипептид) определяют местоположение одонитевого разрыва в первой нити, а дополнительный DBD определяет местоположение одонитевого разрыва во второй нити.

248. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид (например, DBD) содержит эффекторную молекулу TAL.

249. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид (например, DBD) содержит молекулу с "цинковыми пальцами".

250. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид (например, DBD) содержит домен CRISPR/Cas.

251. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительный полипептид (например, дополнительный DBD) содержит эффекторную молекулу TAL.

252. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительный полипептид (например, дополнительный DBD) содержит молекулу с "цинковыми пальцами".

253. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительный полипептид (например, дополнительный DBD) содержит домен CRISPR/Cas.

254. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где полипептид и дополнительный полипептид связываются с сайтами в ДНК-мишени между местоположением одонитевого разрыва в первой нити и местоположением одонитевого разрыва во второй нити.

255. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где местоположение одонитевого разрыва в первой нити и местоположение одонитевого разрыва во второй нити находятся между сайтами, где полипептид и дополнительный полипептид связываются с ДНК-мишенью.

256. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где в ДНК-мишени местоположение одонитевого разрыва во второй нити находится на противоположной стороне от сайтов связывания

полипептида и дополнительного полипептида относительно местоположения одностороннего разрыва в первой нити.

257. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где в ДНК-мишени местоположение одностороннего разрыва во второй нити находится на той же стороне, что и сайты связывания полипептида и дополнительного полипептида, относительно местоположения одностороннего разрыва в первой нити.

258. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где домен CRISPR/Cas полипептида и РНК-матрица связываются с сайтом-мишенью, и где первая нить сайта-мишени содержит первый сайт РАМ.

259. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где сайт РАМ и сайт на расстоянии от сайта-мишени расположены между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити.

260. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где местоположение одностороннего разрыва в первой нити и местоположение одностороннего разрыва во второй нити расположены между сайтом РАМ и сайтом на расстоянии от сайта-мишени.

261. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, дополнительно содержащая дополнительную нуклеиновую кислоту (например, gRNA), содержащую последовательность, которая связывает сайт на расстоянии от сайта-мишени, и где последовательность, которая связывает сайт на расстоянии от сайта-мишени, определяет местоположение одностороннего разрыва во второй нити.

262. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где дополнительная нуклеиновая кислота дополнительно содержит последовательность, которая связывает дополнительный полипептид (например, домен CRISPR/Cas), например, где дополнительная нуклеиновая кислота содержит gRNA.

263. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывается с сайтом на расстоянии от сайта-мишени (например, с первой нитью сайта в геноме-мишени), имеет длину по меньшей мере 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21 или 20 нуклеотидов), например, имеет длину 17, 18, 19, 20, 21, 22, 23 или 24 нуклеотида.

264. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает сайт на

расстоянии от сайта-мишени, комплементарна первой нити сайта-мишени или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с первой нитью сайта-мишени.

265. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где сайт на расстоянии от сайта-мишени содержит сайт РАМ.

266. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где сайт РАМ и сайт-мишень расположены между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити.

267. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где местоположение одностороннего разрыва во второй нити (например, относительно одностороннего разрыва в первой нити) таково, что полимеризация ДНК под действием домена RT протекает в направлении к местоположению одностороннего разрыва во второй нити.

268. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где местоположение одностороннего разрыва во второй нити (например, относительно одностороннего разрыва в первой нити) является таким, что полимеризация ДНК под действием домена RT протекает в направлении от местоположения одностороннего разрыва во второй нити.

269. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии по меньшей мере 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 друг от друга.

270. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии не более 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, 200 или 250 нуклеотидов друг от друга.

271. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии 20-200, 30-200, 40-200, 50-200, 60-200, 70-200, 80-200, 90-200, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 20-190, 30-190, 40-190, 50-190, 60-190, 70-190, 80-190, 90-190, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 20-180, 30-180, 40-180, 50-180, 60-180, 70-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 20-170, 30-170, 40-170, 50-170, 60-170, 70-170, 80-170, 90-170, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 20-160, 30-160, 40-160, 50-160, 60-160, 70-160, 80-160, 90-160, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 20-150, 30-150, 40-150, 50-150, 60-150, 70-150, 80-150, 90-150, 100-150, 110-150, 120-150, 130-150,

140-150, 20-140, 30-140, 40-140, 50-140, 60-140, 70-140, 80-140, 90-140, 100-140, 110-140, 120-140, 130-140, 20-130, 30-130, 40-130, 50-130, 60-130, 70-130, 80-130, 90-130, 100-130, 110-130, 120-130, 20-120, 30-120, 40-120, 50-120, 60-120, 70-120, 80-120, 90-120, 100-120, 110-120, 20-110, 30-110, 40-110, 50-110, 60-110, 70-110, 80-110, 90-110, 100-110, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 20-90, 30-90, 40-90, 50-90, 60-90, 70-90, 80-90, 20-80, 30-80, 40-80, 50-80, 60-80, 70-80, 20-70, 30-70, 40-70, 50-70, 60-70, 20-60, 30-60, 40-60, 50-60, 20-50, 30-50, 40-50, 20-40, 30-40 или 20-30 нуклеотидов друг от друга.

272. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше двухнитевых разрывов (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше) при модификации ДНК, чем в остальном аналогичная система, где один или несколько из сайта РАМ, сайта-мишени или сайта на расстоянии от сайта-мишени не расположены между местоположением одонитевого разрыва первой нити и местоположением одонитевого разрыва второй нити.

273. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше двухнитевых разрывов (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше) при модификации ДНК, чем в остальном аналогичная система, где полипептид и дополнительный полипептид связываются с сайтами в ДНК-мишени не между местоположением одонитевого разрыва в первой нити и местоположением одонитевого разрыва во второй нити.

274. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше двухнитевых разрывов (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше) при модификации ДНК, чем в остальном аналогичная система, где в ДНК-мишени местоположение одонитевого разрыва во второй нити и местоположение одонитевого разрыва в первой нити расположены между сайтами связывания полипептида и дополнительного полипептида.

275. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше двухнитевых разрывов (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше) при модификации ДНК, чем в остальном аналогичная система, где местоположение одонитевого разрыва во второй нити (например, относительно одонитевого разрыва в первой нити) является таким, что домен РТ инициирует обратную транскрипцию в направлении от местоположения одонитевого разрыва во второй нити.

276. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше делеций, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где один или несколько из сайта

РАМ, сайта-мишени или сайта на расстоянии от сайта-мишени не расположены между местоположением однонитевого разрыва первой нити и местоположением однонитевого разрыва второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

277. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше делеций (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше) при модификации ДНК, чем в остальном аналогичная система, где полипептид и дополнительный полипептид связываются с сайтами в ДНК-мишени не между местоположением однонитевого разрыва в первой нити и местоположением однонитевого разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

278. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше делеций, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где в ДНК-мишени местоположение однонитевого разрыва во второй нити и местоположение однонитевого разрыва в первой нити расположены между сайтами связывания полипептида и дополнительного полипептида, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

279. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше делеций, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где местоположение однонитевого разрыва во второй нити (например, относительно однонитевого разрыва в первой нити) является таким, что домен RT инициирует обратную транскрипцию в направлении от местоположения однонитевого разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

280. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше вставок, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где один или несколько из сайта РАМ, сайта-мишени или сайта на расстоянии от сайта-мишени не расположены между местоположением однонитевого разрыва первой нити и местоположением однонитевого разрыва второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

281. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше вставок, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где полипептид и дополнительный полипептид связываются с сайтами в ДНК-мишени не между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

282. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше вставок, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где в ДНК-мишени местоположение одностороннего разрыва во второй нити и местоположение одностороннего разрыва в первой нити расположены между сайтами связывания полипептида и дополнительного полипептида, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

283. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше вставок, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где местоположение одностороннего разрыва во второй нити (например, относительно одностороннего разрыва в первой нити) является таким, что домен RT инициирует обратную транскрипцию в направлении от местоположения одностороннего разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

284. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит больше желаемых модификаций Gene Writing (например, на по меньшей мере 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% больше) при модификации ДНК, чем в остальном аналогичная система, где один или несколько из сайта РАМ, сайта-мишени или сайта на расстоянии от сайта-мишени не расположены между местоположением одностороннего разрыва первой нити и местоположением одностороннего разрыва второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

285. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит больше желаемых модификаций Gene Writing (например, на по меньшей мере 500, 400, 300, 200, 100, 90, 80,

70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% больше) при модификации ДНК, чем в остальном аналогичная система, где полипептид и дополнительный полипептид связываются с сайтами в ДНК-мишени не между местоположением однонитевого разрыва в первой нити и местоположением однонитевого разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

286. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит больше желаемых модификаций Gene Writing (например, на по меньшей мере 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% больше) при модификации ДНК, чем в остальном аналогичная система, где в ДНК-мишени местоположение однонитевого разрыва во второй нити и местоположение однонитевого разрыва в первой нити расположены между сайтами связывания полипептида и дополнительного полипептида, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

287. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит больше желаемых модификаций Gene Writing (например, на по меньшей мере 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% больше) при модификации ДНК, чем в остальном аналогичная система, где местоположение однонитевого разрыва во второй нити (например, относительно однонитевого разрыва в первой нити) является таким, что домен RT инициирует обратную транскрипцию в направлении от местоположения однонитевого разрыва во второй нити, например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

288. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый однонитевой разрыв и второй однонитевой разрыв находятся на расстоянии по меньшей мере 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450 или 500 нуклеотидов друг от друга, например, по меньшей мере 100 нуклеотидов друг от друга (и необязательно не более 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130 или 120 нуклеотидов друг от друга).

289. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где первый однонитевой разрыв и второй однонитевой разрыв находятся на расстоянии 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 100-150, 110-150, 120-150, 130-150, 140-150, 100-140, 110-140, 120-140, 130-140, 100-130, 110-130, 120-130, 100-120, 110-120 или 100-110 нуклеотидов друг от друга.

290. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше вставок, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где местоположение одонитевого разрыва во второй нити находится менее чем в 100 нуклеотидах от местоположения одонитевого разрыва в первой нити (и необязательно в по меньшей мере 20, 30, 40, 50, 60, 70, 80 или 90 нуклеотидах от него), например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

291. Система, набор, РНК-матрица или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система производит меньше делеций, не кодируемых последовательностью гетерологичного объекта (например, на по меньшей мере 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1% меньше), при модификации ДНК, чем в остальном аналогичная система, где местоположение одонитевого разрыва во второй нити находится менее чем в 100 нуклеотидах от местоположения одонитевого разрыва в первой нити (и необязательно в по меньшей мере 20, 30, 40, 50, 60, 70, 80 или 90 нуклеотидах от него), например, при измерении путем секвенирования длинных ридов PacBio, например, как описано в примере 29.

292. 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.

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

(а) 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.

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

(а) 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.

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

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

297. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (а) 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.

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

299. 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.

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

301. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (а) 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.

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

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

303. Способ изготовления РНК-матрицы, включающий:

(a) получение РНК-матрицы согласно любому из предыдущих вариантов осуществления и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа одного или нескольких модифицированных нуклеотидов (например, выбранных из дигидроуридина, инозина, 7-метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-О-метилриботимидина, 2'-О-этилриботимидина, 2'-фторриботимидина, C5-пропинилдезоксцитидина (pdC), C5-пропинилдеоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдезоксидеоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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 РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу; или

(vii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли РНК-матрица от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

304. Способ изготовления системы для модификации ДНК, включающий:

(a) получение системы для модификации ДНК согласно любому из предыдущих вариантов осуществления и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа 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, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу;

(vii) длины 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, имеют 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) присутствия, отсутствия и/или типа посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержат 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;

(ix) присутствия, отсутствия и/или типа 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), аминокислоты, дегидроаланина, цистатионина, лантионина, дженколовой кислоты, диаминопимелиновой кислоты, гомоаланина, норвалина, норлейцина, гомонорлейцина, гомосерина, O-метилгомосерина и O-этилгомосерина, этионина, селеноцистеина, селеногомоцистеина, селенометионина, селеноэтионина, теллурицистеина или теллуromетионина) в полипептиде, первом полипептиде или втором полипептиде, например, содержат 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) стабильности полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактными (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) активности полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей полипептид, первый полипептид или второй полипептид; или

(xii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

305. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с:

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

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

306. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с системой, полипептидом, РНК-матрицей или ДНК, кодирующей их, согласно любому из предыдущих вариантов осуществления;

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

307. Способ согласно любому из предыдущих вариантов осуществления, где система, полипептид, РНК-матрица или ДНК доставляются в сайт-мишень путем электропорации, например, нуклеофекции.

308. Способ согласно любому из предыдущих вариантов осуществления, который не включает приведение клетки в контакт с ДНК, например, или который включает приведение клетки в контакт с композицией, которая содержит не более 10%, 5%, 4%, 3%, 2% или 1% ДНК по массе или по молярному количеству.

309. Способ согласно любому из предыдущих вариантов осуществления, который не включает приведение клетки в контакт с белком, например, или который включает приведение клетки в контакт с композицией, которая содержит не более 10%, 5%, 4%, 3%, 2% или 1% белка по массе или по молярному количеству.

310. Способ согласно любому из предыдущих вариантов осуществления, который включает приведение клетки-мишени или популяции клеток-мишеней в контакт с по меньшей мере двумя РНК-матрицами и/или по меньшей мере двумя полипептидами GeneWriter таким образом, что по меньшей мере два сайта-мишени (первый сайт-мишень и второй сайт-мишень) модифицируются в клетке-мишени.

311. Способ согласно любому из предыдущих вариантов осуществления, где каждый из первого сайта-мишени и второго сайта независимо редактируются с частотой по меньшей мере 5%, 10% или 15% копий сайта в популяции клеток.

312. Способ согласно любому из предыдущих вариантов осуществления, где каждый из первого сайта-мишени и второго сайта независимо редактируются с частотой по меньшей мере 50%, 60%, 70% или 80% от уровня редактирования, полученного у в остальном аналогичной популяции клеток, приведенной в контакт с в остальном аналогичной системой, нацеленной только на один из сайтов-мишеней.

313. Способ согласно любому из предыдущих вариантов осуществления, где полученная в результате популяция клеток содержит не более 5%, 10% или 20% нежелательных вставок/делеций по сравнению с нежелательными вставками/делециями, полученными у в остальном аналогичной популяции клеток, приведенной в контакт с в остальном аналогичной системой, нацеленной только на один из сайтов-мишеней.

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

315. Способ согласно любому из предыдущих вариантов осуществления, где клетка представляет собой Т-клетку.

316. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт, например, путем нуклеофекции или доставки липидных частиц, с:

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

за счет чего обеспечивается модификация сайта-мишени в геномной ДНК в клетке, где клетка является эуплоидной, не является иммортализованной, является частью ткани, является частью организма, является первичной клеткой, является неделящейся, является гаплоидной (например, клеткой зародышевой линии), является нераковой полиплоидной клеткой или получена от субъекта, у которого имеется генетическое заболевание.

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

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

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

320. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту:

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

за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

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

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

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

324. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту системы, полипептида, РНК-матрицы или ДНК, кодирующей их, согласно любому из предыдущих вариантов осуществления;

за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

325. Способ согласно любому из предыдущих вариантов осуществления, где заболевание или состояние, ассоциированное с генетическим дефектом, представляет собой показание, указанное в любой из таблиц 9-12, и/или где генетический дефект представляет собой дефект гена, указанного в любой из таблиц 9-12.

326. Способ согласно любому из предыдущих вариантов осуществления, где субъект является пациентом-человеком.

5. 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.

Первая/вторая нить. Как используется в данном документе, термины "первая нить" и "вторая нить", используемые для описания отдельных нитей ДНК ДНК-мишени, служат для проведения различий между двумя нитями ДНК на основании того, какая нить домена обратной транскриптазы инициирует полимеризацию, например, на основании того, где инициируется синтез с праймированием мишенью. Первая нить относится к нити ДНК-мишени, на которой домен обратной транскриптазы инициирует полимеризацию, например, где инициируется синтез с праймированием мишенью. Вторая нить относится к другой нити ДНК-мишени. Обозначения первой и второй нитей не описывают нитей ДНК сайта-мишени в других отношениях; например, в некоторых вариантах осуществления первая и вторая нити подвергаются одностороннему разрыву полипептидом, описанным в данном документе, но обозначения "первая" и "вторая" нити не имеют отношения к порядку, в котором происходят такие односторонние разрывы.

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 мышинового 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 (а) 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 транзистентно (например, мРНК), так и полустабильно в течение более чем одного поколения (например, эпизомный вирусный вектор, плазида или другой самореплицирующийся вектор).

Инвертированные концевые повторы. Используемый в данном документе термин "инвертированные концевые повторы" или "ITR" относится к цис-элементам вируса AAV, названным так из-за их симметрии. Эти элементы способствуют эффективному размножению генома AAV. Предполагается, что минимальными элементами для функции ITR являются Rep-связывающий сайт (RBS; 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 1538) для AAV2) и сайт концевой разрешения (TRS; 5'-AGTTGG-3' для AAV2), а также переменная палиндромная последовательность, позволяющая образовывать шпильку. В соответствии с настоящим изобретением ITR содержит по меньшей мере эти три элемента (RBS, TRS и последовательности, позволяющие образовывать шпильку). Кроме того, в настоящем изобретении термин "ITR" относится к ITR известных природных серотипов AAV (например, ITR серотипа 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или 11 AAV), к химерным ITR, образованным путем слияния элементов ITR, полученных из разных серотипов, и к их функциональным вариантам. Под функциональным вариантом ITR понимается последовательность, проявляющая по меньшей мере 80%, 85%, 90%, предпочтительно по меньшей мере 95%, идентичность последовательности с известным ITR, что обеспечивает

возможность размножения последовательности, которая содержит указанный ITR, в присутствии белков Rep.

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.

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. В различных вариантах осуществления нуклеиновые кислоты находятся в функциональной ассоциации с дополнительными генетическими элементами, такими как тканеспецифическая(тканеспецифические) последовательность(последовательности), контролирующая(контролирующие) экспрессию (например, тканеспецифические промоторы и тканеспецифические последовательности распознавания микроРНК), а также дополнительными элементами, такими как инвертированные повторы (например, инвертированные концевые повторы, такие как элементы вирусов или элементы, полученные из вирусов, например, ITR AAV) и тандемные повторы, инвертированные повторы/прямые повторы (например, инвертированные повторы транспозонов, например, инвертированные повторы транспозонов, также содержащие прямые повторы, например, инвертированные повторы, также содержащие прямые повторы), области гомологии (сегменты с различной степенью гомологии с ДНК-мишенью), UTR (5'-, 3'- или как 5'-, так и 3'-UTR) и различные комбинации вышеперечисленного. Элементы систем, предусмотренных в настоящем изобретении, в виде нуклеиновых кислот могут быть представлены в различных топологиях, включая одностороннюю, двустороннюю, кольцевую, линейную, линейную с незамкнутыми концами, линейную с замкнутыми концами и их конкретные варианты, такие как ДНК Doggybone (dbDNA), ДНК с замкнутыми концами (ceDNA).

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.

Функциональная ассоциация. Используемый в данном документе термин "функциональная ассоциация" описывает функциональную взаимосвязь между двумя последовательностями нуклеиновой кислоты, такими как 1) промотор и 2) последовательность гетерологичного объекта, и означает, в таком примере, что промотор и последовательность гетерологичного объекта (например, ген, представляющий интерес) ориентированы таким образом, что в подходящих условиях промотор управляет экспрессией последовательности гетерологичного объекта. Например, нуклеиновая кислота-матрица может быть односторонней, например, с (+) или (-) ориентацией, но функциональная ассоциация между промотором и последовательностью гетерологичного объекта означает, будет ли нуклеиновая кислота-матрица транскрибироваться в определенном состоянии, при этом если она находится в подходящем состоянии (например, в (+) ориентации, в присутствии необходимых каталитических факторов, NTP и т. д.), она точно транскрибируется. Функциональная ассоциация применяется аналогичным образом к другим парам нуклеиновых кислот, в том числе к другим тканеспецифическим последовательностям, контролирующим экспрессию (таким как энхансеры, репрессоры и последовательности распознавания микроРНК), IR/DR, ITR, UTR, или областям гомологии и последовательностям гетерологичных объектов, или последовательностям, кодирующим транспозазу.

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) выпетливания. 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.

Тканеспецифическая(тканеспецифические) последовательность(последовательности), контролирующая экспрессию. Используемый в данном документе термин "тканеспецифическая последовательность, контролирующая экспрессию" означает элементы нуклеиновой кислоты, которые повышают или снижают уровень транскрипта, содержащего гетерологичную последовательность объекта, в ткани-мишени тканеспецифическим образом, например, предпочтительно в целевой(целевых) ткани(тканях) по сравнению с нецелевой(нецелевыми) тканью(тканями). В некоторых вариантах осуществления тканеспецифическая последовательность, контролирующая экспрессию, предпочтительно управляет транскрипцией, активностью или периодом полужизни или подавляет их у транскрипта, содержащего гетерологичную последовательность объекта, в ткани-мишени тканеспецифическим образом, например, предпочтительно в целевой(целевых) ткани(тканях) по сравнению с нецелевой(нецелевыми) тканью(тканями). Примеры тканеспецифических последовательностей, контролирующих экспрессию, включают тканеспецифические промоторы, репрессоры, энхансеры или их комбинации, а также тканеспецифические последовательности распознавания микроРНК. Тканевая специфичность относится к целевой (ткани(тканям), в которых экспрессия или активность нуклеиновой кислоты-матрицы желательны или допустимы) и нецелевой (ткани(тканям), в которых экспрессия или активность нуклеиновой кислоты-матрицы нежелательны или

недопустимы). Например, тканеспецифический промотор (такой как промотор в нуклеиновой кислоте-матрице или контролирующей экспрессию транспозазы) управляет экспрессией преимущественно в целевых тканях по сравнению с нецелевыми тканями. В отличие от этого, микроРНК, которая связывает тканеспецифические последовательности распознавания микроРНК (в нуклеиновой кислоте, кодирующей транспозазу, либо в нуклеиновой кислоте-матрице, либо в них обеих), преимущественно экспрессируется в нецелевых тканях по сравнению с целевыми тканями, благодаря чему снижается экспрессия нуклеиновой кислоты-матрицы (или транспозазы) в нецелевых тканях. Соответственно, промотор и последовательность распознавания микроРНК, специфичные для одной и той же ткани, такой как ткань-мишень, характеризуются противоположными функциями (стимулируют и подавляют соответственно с конкордантными уровнями экспрессии, т. е. высокими уровнями микроРНК в нецелевых тканях и низкими уровнями в целевых тканях, тогда как промоторы управляют экспрессией на высоком уровне в целевых тканях и экспрессией на низком уровне в нецелевых тканях) в отношении транскрипции, активности или периода полужизни ассоциированной последовательности в данной ткани.

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.

6. КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

Фигура 1 представляет собой схематическое изображение системы редактирования генома Gene Writing™.

Фигура 2 представляет собой схематическое изображение структуры полипептида редактора генома GeneWriter™.

Фигура 3 представляет собой схематическое изображение структуры иллюстративных РНК-матриц GeneWriter™.

Фигуры 4А и 4В 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, например, из таблицы в данном документе; "цинковый палец"; домен TAL; Cas9; dCas9; никазу Cas9; фактор транскрипции или мегануклеазу), РНК-связывающие домены (например, включающие РНК-связывающий домен белка В-бокса, белка оболочки MS2, dCas или элемент последовательности из таблицы в данном документе), домены обратной транскриптазы (например, включающие домен обратной транскриптазы элемента последовательности из таблицы в данном документе; другие

ретротранспозазы (например, указанные в таблице в данном документе); пептид, содержащий домен обратной транскриптазы (например, указанный в таблице в данном документе)) и/или эндонуклеазный домен (например, включающий эндонуклеазный домен элемента из таблицы в данном документе; 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.

Фигура 5 представляет собой схему, на которой показаны модули иллюстративной РНК-матрицы GeneWriter. 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.

Фигура 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.

Фигуры 7А и 7В представляют собой схемы, на которых показан иллюстративный процесс осуществления одностороннего разрыва второй нити. (А) Никаза Cas9 слита с белком Gene Writer. Белок Gene Writer вводит односторонний разрыв в нить ДНК посредством своего EN-домена (обозначается как *), и слитая никаза Cas9 вводит односторонние разрывы либо в верхнюю, либо в нижнюю нити ДНК (обозначается как X). (В) Gene Writer нацеливается на ДНК посредством своего ДНК-связывающего домена и вводит односторонний разрыв в ДНК с помощью своего EN-домена (*). Затем никазу Cas9 применяют для создания второго одностороннего разрыва (X) на верхней либо на нижней нитях выше или ниже одностороннего разрыва, введенного с помощью EN.

Фигура 8. Линкерная область на С-конце ДНК-связывающего домена R2Tg может быть усечена и модифицирована. Были сконструированы делеции в природном линкере от домена myb в А или В до положений 1 или 2 наряду с заменой с помощью синтетических линкеров 3GS (SEQ ID NO: 1024) или XTEN (А). Эффективность интеграции измеряли в клетках HEK293Т с помощью ddPCR (В).

Фигура 9. "Посадочные площадки", предназначенные для тестирования мутаций сайтов-мишеней R2Tg Gene Writer.

Фигура 10а. Анализ методом ddPCR, в котором измеряется процентная доля интеграций для всех лентивирусных интегрированных "посадочных площадок" на клетку.

Фигура 10b. Секвенирование ампликонов и анализ методом NGS вставок/делеций, присутствующих в сайтах "посадочных площадок".

Фигура 11. Замена ДНК-связывающего домена ретротранспозазы Gene Writer на ZFP для AAVS1. На этой фигуре раскрыт "линкер 3GS" как SEQ ID NO: 1024.

Фигура 12. Замена ДНК-связывающего домена ретротранспозазы Gene Writer с активным доменом EN или без него на Cas9 или никазу Cas9 (*=мутантный)

Фигура 13. Слияние AAVS1 ZFP с ретротранспозазой Gene Writer с функциональным ДНК-связывающим доменом или без него.

Фигура 14. Схематическое изображение продуктов слияния никаза Cas9-Gene Writer. (A) Схематическое изображение никазы Cas9, слитой с белком Gene Writer. (B) Схематическое изображение gRNA, удлиненной на 3'-конце.

Фигура 15. Схематическое изображение продуктов слияния никаза Cas9-Gene Writer. (A) Схематическое изображение никазы Cas9, слитой с белком Gene Writer. (B) Схематическое изображение донорного трансгена, фланкированного UTR, и гомологии с сайтом разреза.

Фигура 16. Схематическое изображение конструкций. (A) Схематическое изображение белка Gene Writer. (B) Схематическое изображение донорного трансгена, фланкированного UTR, и гомологии с сайтом разреза. (C) Схематическое изображение используемых конструкций Cas9.

Фигура 17. Схематическое изображение мРНК, кодирующей Gene Writer (A). Нативные нетранслируемые области (UTR) были заменены 5'- и 3'-UTR, оптимизированными для экспрессии белка (показаны как 5'-UTR_{exp} и 3'-UTR_{exp}). Экспрессию белка Gene Writer анализировали с помощью анализа HiBit путем исследования экспрессии его HiBit-метки (B). На этой фигуре раскрыт "3GS" как SEQ ID NO: 1024.

Фигура 18. Интеграция в геном, индуцированная белком Gene Writer с его нативными UTR и UTR, оптимизированными для экспрессии белка. Активность Gene Writing с ненативными UTR стимулируется благодаря наличию РНК-матрицы, несущей нативные UTR ретротранспозона.

Фигура 19. Доставка системы Gene Writer с применением мРНК, кодирующей полипептид, и плазмидной ДНК, кодирующей РНК-матрицу для ретротранспозиции.

Фигура 20. Схемы иллюстративных стратегий конструирования 5'-UTR. НА=плечо гомологии; К=последовательность Козак; рА=поли(А)-сигнал; АМа=*A. maritima*; Rх=другие виды ретротранспозонов.

Фигура 21. Возможное расположение интрона (или интронов) в РНК-матрице. Интроны показаны изогнутыми линиями. 5'-НА: 5'-плечо гомологии; 3'-НА: 3'-плечо гомологии; 5'-UTR: 5'-UTR, специфичная для ретротранспозонов; 3'-UTR: 3'-UTR, специфичная для ретротранспозонов; GOI: ген, представляющий интерес. Оранжевые блоки соответствуют последовательности, предназначенной для экспрессии из геномного местоположения, содержащего свой собственный клеточноспецифический промотор, поли(А)-сигнал и UTR для экспрессии белка (5'- и 3'-UTR_{exp}). Последовательность может быть ориентирована в смысловой (показано выше) или в антисмысловой ориентации относительно UTR ретротранспозона и плеч гомологии. Интрон может быть расположен внутри GOI или внутри UTR_{exp}.

Фигура 22. Интеграция в геном в клетках HEK293T по данным анализа методом ddPCR на 3'-конце. мРНК Gene Writer в количестве 0,5 мкг/лунка вводили путем котрансфекции с РНК-матрицами с ферментативно добавленными кэпом 1 и поли(А)-хвостом или без них. Соотношение мРНК Gene Writer и РНК трансгена составляло 1:1.

Фигура 23. Интеграция в геном, обнаруживаемая с помощью 3' ddPCR, индуцированная экспрессией мРНК Gene Writer, продуцируемой с использованием немодифицированных (G0) либо модифицированных нуклеотидов (псевдоуридина (Ψ), 1-N-метилпсевдоуридина (1-Me-Ψ), 5-метоксиуридина (5-MO-U) или 5-метилцитидина (5mC)). Использовали 1 мкг мРНК Gene Writer на лунку. Использовали немодифицированную РНК-матрицу. РНК Gene Writer и РНК-матрицу вводили путем котрансфекции в молярном соотношении, составляющем 1:8.

Фигура 24. Схема конструкции плазмид с управляющим элементом и трансгеном. Плечи гомологии (НА) и спейсерные последовательности являются вариабельными в данном наборе экспериментов.

Фигура 25. (А) Временная шкала эксперимента. (В) Схема конфигураций конструкций R2Tg и трансгена. (С) Вестерн-блоттинг против Rad51 демонстрирует уменьшение уровня экспрессии белка Rad51 в день 3.

Фигура 26. Клетки U2OS обрабатывали ненацеливающейся контрольной siRNA (контроль) или siRNA для Rad51 вместе с R2Tg Wt или контрольными мутантными формами RT и EN. ddPCR в 3'- (А) или 5'- (В) месте соединения применяли для оценки эффективности интеграции в день 3.

Фигура 27. (А) Карта последовательности рибозима элемента R2 из *Taeniopygia guttata* (R2Tg) в окружении модулей трансгенной молекулы РНК Gene Writer. Признаки рибозима обозначаются следующим образом: Р - область спаренных оснований; Р'- нить, комплементарная области спаренных оснований; L - петля в конце Р-области; J - нуклеотиды, соединяющие области спаренных оснований. На фигуре раскрыта SEQ ID NO: 1734. (В) Предсказание вторичной структуры рибозима R2Tg. Заштрихованным прямоугольником указано предположительное каталитическое положение, которое можно использовать для инактивации рибозима. На фигуре раскрыта SEQ ID NO: 1734.

Фигура 28. Карта последовательности рибозима элемента R2 из *Taeniopygia guttata* (R2Tg) в окружении модулей трансгенной молекулы РНК Gene Writer. Признаки рибозима обозначаются следующим образом: Р - область спаренных оснований; Р'- нить, комплементарная области спаренных оснований; L - петля в конце Р-области; J - нуклеотиды, соединяющие области спаренных оснований. На фигуре раскрыта SEQ ID NO: 1734.

Фигура 29. Предсказание вторичной структуры рибозима элемента R2 из *Taeniopygia guttata*. На фигуре раскрыта SEQ ID NO: 1734.

Фигура 30. Система Gene Writing для лечения иллюстративного нарушения, связанного с экспансией повторов. На фигуре раскрыты SEQ ID NO: 1645, 1599, 1645, 1635-1636, 1645 и 1686-1688 соответственно в порядке появления.

Фигура 31. Иллюстрация двух ориентаций однонитевого разрыва второй нити в иллюстративной системе Gene Writing.

Фигура 32. Иллюстрация ориентации и положения внесения однонитевых разрывов второй нити в иллюстративной системе Gene Writing и их эффект в отношении редактирования.

Фигура 33. Демонстрирует образование и экспрессию слитых белков Cas9-RT. Для оценки экспрессии новых полипептидов Gene Writer в клетках человека клетки U2OS трансфицировали экспрессионными плазмидами Cas-RT, содержащими различные домены RT из таблиц 1 и 30, слитые с нисказой дикого типа (WT) или Cas9(N863A). Клеточные лизаты собирали в день 2 после трансфекции и анализировали с помощью вестерн-блоттинга с использованием первичного антитела к Cas9. Первичное антитело к GADPH было включено в качестве контроля нагрузки.

Фигура 34. Демонстрирует улучшение экспрессии продуктов слияния Cas-RT благодаря выбору линкерной последовательности. Чтобы оценить, как линкеры могут изменять экспрессию новых полипептидов Gene Writer в клетках человека, клетки U2OS трансфицировали экспрессионными плазмидами Cas-RT, содержащими различные линкеры из таблицы 42, посредством слияния нисказы Cas9(N863A) с доменом RT ретротранспозазы R2Bm с мутантным РНК-связывающим доменом. Клеточные лизаты собирали и анализировали с помощью вестерн-блоттинга с использованием первичных антител к Cas9. Первичное антитело к винкулину (слева) или GADPH (справа) было включено в качестве контроля нагрузки. Контроли Cas9 слева представляют титрование экспрессионной плазмиды с Cas9. Незакрашенными стрелками указан тестируемый исходный линкер, тогда как закрашенной стрелкой представлен линкер (линкер 10), который, как обнаружено, существенно улучшает экспрессию слитого полипептида. Номера образцов соответствуют идентификаторам линкерных последовательностей в таблице 42.

Фигура 35. Демонстрирует, что активность нацеливания Cas/gRNA на ДНК сохраняется в продуктах слияния Cas-RT. Различные домены RT сливали с Cas9(WT) и вводили путем электропорации в клетки U2OS. Геномную ДНК собирали и анализировали на наличие мутационных сигнатур с помощью секвенирования нового поколения. В соответствующих случаях указаны мутации в РНК- или ДНК-связывающих доменах (RBD или DBD) доменов ретротранспозазы R2. Частота вставок/делеций используется в данном документе как косвенный показатель сохранения активности Cas в качестве составной части продукта слияния с RT.

На фигурах 36А и В раскрыто применение мутаций, улучшающих домены обратной транскриптазы. Консервативные домены обратной транскриптазы из родов ретровирусов Betaretrovirus, Deltaretrovirus, Gammaretrovirus, Epsilonretrovirus и Spumavirus были выровнены и сравнены с мутациями, которые, как ранее было показано, улучшают активность RT (Anzalone et al Nat Biotechnol 38(7):824-844 (2020); Baranauskas et al Protein Eng Des Sel 25(10):657-668 (2012); Arezi and Hogrefe Nucleic Acids Res 37(2):473-481 (2009)).

На фигуре 36А показано, что набор из 3 ключевых мутаций был идентифицирован и применен к RT из этих родов, как указано. На фигуре 36В раскрыто, что дополнительные мутации были применены с первым приоритетом из набора T306K/W313F или в качестве альтернативы из L139P/E607K, если ни одна из первого набора не считалась передаваемой. Выбранные мутации показаны в таблице 45.

Фигура 37. Клетки U2OS подвергали нуклеофекции различными слитыми векторами Cas-RT, в которых домен RT был выбран из базы данных доменов мономерной ретровирусной обратной транскриптазы. Редактирование локуса НЕК3 с использованием матрицы, описанной в таблице 43, оценивали с помощью секвенирования ампликонов и анализа точного редактирования в сравнении с сигнатурами вставок/делеций. Данные представлены здесь в виде коэффициентов активности, которые рассчитываются как соотношение частоты встречаемости ридов с точно намеченным редактированием (вставка СТТ в сайт-мишень одонитевого разрыва) к частоте прочтений с любыми другими мутациями (вставками/делециями). Три проанализированные конфигурации РНК-матрицы давали сходные результаты, поэтому показаны результаты для одной матрицы (матрицы Р2 из таблицы 43).

На фигуре 38 показано, что одновременное нацеливание на несколько локусов приводит к эффективной активности Gene Writing. Клетки НЕК293 подвергали нуклеофекции с помощью систем Gene Writing, содержащих различные композиции плазмид с матрицами, чтобы обеспечить нацеливание на локусы: 1) только НЕК3, 2) только НВВ или 3) как НВВ, так и НЕК3. Процент редактирования указан для каждого локуса при доставке одной или обеих локус-специфических экспрессионных плазмид с РНК-матрицей. Закрашенные столбцы представляют события точной записи, а незакрашенные столбцы представляют частоту вставок/делеций. Редактирование, специфичное для локуса-мишени, наблюдалось при независимой доставке любой матрицы, а высокоэффективные и специфичные редактирования наблюдались в обоих локусах при совместной доставке матриц.

Фигура 39. Демонстрирует эффект длины в отношении активности Gene Writing. Клетки НЕК293Т подвергали нуклеофекции с помощью систем Gene Writing, полностью состоящих из РНК, содержащих различные РНК-матрицы (таблица 48), для тестирования эффективности редактирования в подходе без использования ДНК в локусе НЕК3. Матрица 4, которая кодировала то же редактирование, что и матрица 1, но с добавлением 20 нуклеотидов на 3'-конце матрицы RT, демонстрировала примерно 3,1-кратное падение активности точной записи и примерно 2,4-кратное падение соотношения точных коррекций и вставок/делеций.

Фигура 40. Показан эффект доставки Gene Writer, состоящей только из РНК, с использованием различных композиций мРНК. Нуклеофекция НЕК293Т различными мРНК Cas9-RT(MMLV) (таблица 49) с использованием матрицы 1 (таблица 48А). В данном документе не наблюдалось сильных эффектов при варьировании составов кэпов и UTR.

Фигура 41. Клетки HEK293T подвергали нуклеофекции с помощью системы Gene Writing с использованием набора матриц (матрица 1, таблица 48) для редактирования локуса HEK3 и двух различных конструкций Cas-RT. Анализ последовательности указывал на то, что оба продукта слияния Cas-RT вносили редактирования очень точным и эффективным образом. В обеих системах наблюдалось повышение эффективности в условиях, включающих необязательный вторичный односторонний разрыв. Эти данные показывают успешное клонирование и точную запись доменом PERV RT в качестве составной части этих продуктов слияния Cas-RT.

Фигура 42. Демонстрирует эффект доставки Gene Writer, состоящей только из РНК, с использованием модифицированных нуклеотидов. Молекулы мРНК, кодирующие полипептид Cas-RT(MMLV), варьировали по составу для определения эффектов (таблица 49). Здесь матрица 1 используется для редактирования локуса HEK3 после включения модифицированных нуклеотидов в состав мРНК-компонента. Было обнаружено, что активность Gene Writing с 5mU-модифицированным мРНК-компонентом является как высокой, так и точной.

На фигурах 43 А, В и С показан эффект доставки Gene Writer, состоящей только из РНК, с использованием различных композиций мРНК, доставляемых в клетку с помощью липидных частиц. На фигуре 43А показана липофекция HEK293T различными мРНК Cas9-RT(MMLV) из системы, состоящей только из РНК, выполненная с использованием матрицы 1 (таблица 48) и доставки с помощью Lipofectamine 3000. На фигуре 43В показана липофекция HEK293T различными мРНК Cas9-RT(MMLV) из системы, состоящей только из РНК, выполненная с использованием матрицы 1 (таблица 48) и доставки с помощью реагента MessengerMax. Эти данные указывают на более высокую эффективность точного редактирования при использовании реагента MessengerMax. На фигуре 45С показан анализ двух матриц, различающихся по общей длине, с использованием реагента MessengerMax. В этом эксперименте не было обнаружено никаких серьезных изменений эффективности редактирования, связанных со сменой матрицы. В тех случаях, когда они были включены "голова к голове", добавление gRNA для второго одностороннего разрыва приводило к повышению эффективности системы.

На фигуре 44 показана доставка Cas-RT из системы, состоящей только из РНК, с использованием систем на основе липидов. Cas9-RT(MMLV) и Cas9-RT(PERV) доставляли в клетки HEK293T с матрицей 1 (таблица 48) с использованием липидного реагента MessengerMax. В данном документе активность обоих ферментов составляла около 5% точной записи.

На фигурах 45А и В показана экспрессия системы Gene Writer, состоящей только из РНК, в первичных CD4⁺ Т-клетках человека. На фигуре 45А показана экспрессия белка Gene Writer из мРНК при различных дозах, доставленных в первичные CD4⁺ Т-клетки человека в день 1 после нуклеофекции. Gene Writer обнаруживали с помощью антитела, нацеленного на Cas9-часть полипептида. GAPDH, ген домашнего хозяйства, обнаруживали с помощью антитела к GAPDH. Повышение уровней экспрессии наблюдалось при

увеличении доставляемых доз нуклеофицированной мРНК, кодирующей полипептид, например, 0, 2,5, 5 и 10 мкг мРНК Gene Writer. Показанные данные об обнаружении экспрессии белка включали 2 повторности. На фигуре 45В показана жизнеспособность клеток после нуклеофекции 6 РНК-матриц. Жизнеспособность первичных CD4⁺ Т-клеток после доставки РНК системы Gene Rewriter в день 3 после нуклеофекции. Жизнеспособность клеток оценивали с помощью проточной цитометрии после окрашивания собранных Т-клеток для выявления живых/мертвых клеток (среднее значение $\pm$ стандартное отклонение, n=2 повторности).

[Гейт: живые клетки в синглетной популяции из популяции клеток, отобранной по графику размера FSC/SSC]

На фигурах 46 А и В показана Gene Writing в первичных CD4⁺ Т-клетках человека. На фигуре 46А показано точное редактирование геномного локуса HEK3 с помощью системы Gene Writer в первичных CD4⁺ Т-клетках человека без добавления gRNA для второго одностороннего разрыва. На фигуре 46В показано точное редактирование геномного локуса HEK3 с помощью системы Gene Writer в первичных CD4⁺ Т-клетках человека. Геномную ДНК экстрагировали из клеток в день 3 после нуклеофекции. Геномное редактирование HEK3 исследовали с помощью анализа секвенирования ампликонов на основе ПЦР. ДНК-ампликоны, содержащие ожидаемые геномные изменения, были идентифицированы как события точной записи, тогда как ампликоны с непреднамеренным редактированием (например, вставкой, делецией) были подсчитаны как вставки/делеции. Процентную долю каждого из них рассчитывали, исходя из общего количества ридов для каждого условия (среднее значение $\pm$ стандартное отклонение, n=2 повторности).

На фигурах 47 А и В показано использование gRNA для второго одностороннего разрыва для Gene Writing в первичных CD4⁺ Т-клетках человека. Данные, полученные на фигуре 46, показаны здесь для прямого сравнения потенциальных эффектов gRNA для второго одностороннего разрыва в отношении эффективности. На фигуре 47А показано, что в этом эксперименте добавление gRNA для второго одностороннего разрыва не приводило к усилению сигнала точной записи. На фигуре 47В показано, что использование gRNA для второго одностороннего разрыва вместо этого могло привести к увеличению частоты вставок/делеций. Таким образом, в некоторых вариантах осуществления последовательность gRNA для второго одностороннего разрыва может отсутствовать в системе, описанной в данном документе. Точное редактирование геномного сайта HEK3 с помощью системы Gene Writer в первичных CD4⁺ Т-клетках человека без добавления (фиг. 47 А) или с добавлением gRNA для второго одностороннего разрыва (фиг. 47 В). Геномную ДНК экстрагировали из клеток в день 3 после нуклеофекции. Геномное редактирование HEK3 исследовали с помощью анализа секвенирования ампликонов на основе ПЦР. ДНК-ампликоны, содержащие ожидаемые геномные изменения, обусловленные системой Gene Writer, были идентифицированы как события точной записи, тогда как ампликоны с непреднамеренным редактированием (например, вставкой, делецией) были подсчитаны как вставки/делеции. Процентную долю каждого из них рассчитывали, исходя из общего

количества ридов для каждого условия (среднее значение $\pm$ стандартное отклонение, $n=2$ повторности).

На фигуре 48 показана разработанная конструкция 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(А)-сигнал; SD-SA используется для обозначения донорных сайтов сплайсинга и акцепторных сайтов сплайсинга антисмыслового интрона в последовательности, кодирующей GFP.

Фигура 49. Скрининг ретротранспозонов-кандидатов позволяет идентифицировать 25 кандидатов, функционирующих для интеграции полезной нагрузки в транс-положении в клетки человека. В общей сложности 163 системы ретротранспозонов анализировали на активность в клетках человека, как описано в примере 39. 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.

На фиг. 50 А и В показан анализ люциферазной активности для первичных клеток. LNP, составленные в соответствии с примером 44, анализировали в отношении доставки груза в первичные гепатоциты человека (А) и мышцы (В) в соответствии с примером 45. 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 фигуре 51 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 транзистентной, и уровни ферментов возвращались к уровню, близкому к фоновому уровню при обработке средой-носителем, через 48 часов после введения.

7. 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 последовательности гетерологичного объекта в сайт-мишень в геноме млекопитающего) в одном или нескольких местоположениях в последовательности ДНК в клетке, ткани или у субъекта, например, *in vivo* или *in vitro*. Последовательность ДНК гетерологичного объекта может содержать, например, замену, делецию, вставку, например, кодирующую последовательность, регуляторную последовательность или единицу экспрессии гена.

Более конкретно, в настоящем изобретении предусмотрены системы на основе обратной транскриптазы для изменения последовательности геномной ДНК, представляющей интерес, например, путем вставки, удаления или замены одного или нескольких нуклеотидов в/из последовательности, представляющей интерес. Настоящее изобретение частично основывается на биоинформатическом анализе для идентификации последовательностей обратной транскриптазы, например, в ретротранспозонах различных организмов (см. таблицу 1 или 3).

В настоящем изобретении частично предусмотрены редакторы генома Gene Writer™, содержащие полипептидный компонент и компонент в виде нуклеиновой кислоты-матрицы (например, РНК-матрицы). В некоторых вариантах осуществления редактор генома Gene Writer™ может использоваться для внесения изменений в сайт-мишень в геноме. В некоторых вариантах осуществления полипептидный компонент содержит записывающий домен (например, домен обратной транскриптазы), ДНК-связывающий домен и эндонуклеазный домен (например, нуклеазный домен). В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) содержит последовательность, которая связывает сайт-мишень в геноме (например, которая связывается со второй нитью сайта-мишени), последовательность, которая связывает полипептидный компонент, последовательность гетерологичного объекта и 3'-концевой домен гомологии с мишенью. Не желая ограничиваться какой-либо теорией, полагают, что нуклеиновая кислота-матрица (например, РНК-матрица) связывается со второй нитью сайта-мишени в геноме и связывается с полипептидным компонентом (например, локализуя полипептидный компонент в сайте-мишени в геноме). Считается, что эндонуклеаза (например, нуклеаза) полипептидного компонента разрезает сайт-мишень (например, первую нить сайта-мишени), например, позволяя 3'-концевому домену гомологии связываться с последовательностью, прилегающей к сайту-мишени, который нужно изменить, в первой нити сайта-мишени. Считается, что записывающий домен (например, домен обратной транскриптазы) полипептидного компонента использует 3'-концевой домен гомологии с мишенью в качестве праймера и последовательность гетерологичного объекта в качестве матрицы, например, для полимеризации последовательности, комплементарной последовательности гетерологичного объекта. Не желая ограничиваться какой-либо теорией, полагают, что выбор подходящей

последовательности гетерологичного объекта может привести к замене, делеции или вставке одного или нескольких нуклеотидов в сайте-мишени.

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.

8. Редакторы генома Gene Writer™

Редакторы генома Gene Writer™ представляют собой системы, которые способны модифицировать геном клетки-хозяина и могут применяться для мутации, делеции или другой модификации геномной последовательности-мишени, включая вставку гетерологичных полезных нагрузок. В некоторых вариантах осуществления эти системы происходят из естественно эволюционирующих мобильных генетических элементов, известных как ретротранспозоны. Полипептиды Gene Writer™ также могут содержать домены RT, полученные из источников, отличных от ретротранспозонов, например, из вирусов.

Не содержащие длинных концевых повторов (LTR) ретротранспозоны представляют собой тип мобильных генетических элементов, широко распространенных в геномах эукариот. Они включают два класса: ретротранспозоны типа апуриновой/апиримидиновой эндонуклеазы (APE) и ретротранспозоны типа эндонуклеазы, подобной ферментам рестрикции (RLE). Ретротранспозоны класса APE состоят из двух функциональных доменов: эндонуклеазного/ДНК-связывающего домена и домена обратной транскриптазы. Ретротранспозоны класса RLE состоят из трех функциональных доменов: ДНК-связывающего домена, домена обратной транскрипции и эндонуклеазного домена. 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 транспозазой, и варибельную 5'-область, как правило, имеющую открытую(открытые) рамку(рамки) считывания ("ORF"), кодирующую(кодирующие) белки транспозазы. Последовательность РНК-матрицы может также содержать 5'-нетранслируемую область, которая специфично связывается с ретротранспозазой.

В некоторых вариантах осуществления, как описано в данном документе, элементы таких не содержащих LTR 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 последовательности нуклеиновой кислоты-объекта (например, гетерологичной) 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: (А) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и либо (х) эндонуклеазный домен, который обладает функциональным свойством связывания с ДНК, либо (у) эндонуклеазный домен и отдельный ДНК-связывающий домен; и (В) 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 некоторых вариантах осуществления ДНК-связывающая функция может предусматривать РНК-компонент, который направляет белок к последовательности ДНК, например, gRNA. 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 не содержащих LTR ретротранспозонов, например, ретротранспозонов APE-типа или RLE-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 белок редактора генома Gene Writer способен к синтезу второй нити. В таблице 50 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 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 50.

Семейство	Элемент	Номер доступа	Орган	Предсказанная 5'-UTR	Предсказанная 3'-UTR	Предсказанная аминокислотная последовательность	RT	Стар-кодон	Стоп-кодон
	Crack-3_SP	GCF_000002235.5	Strongylocentrotus purpuratus	AAGGTGGTTCAG TTCCTCCGTTTAG AAACAGGGGCTT AAAATTTGCCCAT TTGAATATTCATA GCTTAGTTGCTAA GATTGACGAACTT AGATTACTTCTGA GAAATGCATCATT TGATGTTATTGGT TTAAATGAAACAT TTTGCGACTCC	TGAATGGCTCAGCCG TACATACACTACGTG CATGTTCTGTAATATT TACATGTCTTATTCAC ATAATTGCGACTTAA TTAATATTTTCTATT GACATTACTGCTTAC TAGGTTAATGATATA TGTTAATTTTGCAAA CGAGTATTATTTTGT AACTGTTCTGTATAA ATGTTATAAACGATG TAAATATGTTCAATTCA GGGCCCCATGGAAG ACCACTACTTGCTGG TGAATGGGCTACCCT	MIDDSLEYIDDYILLRRDRTRHGGGVAV YISKRFNYKRRDDLISSTLECLWFELHYP NKSPILVGEFYRPPNSHVDFEELDGNL DSALSEDMSCIILGDFNCDCMPHQLDL NAEKLMFYTNMYGLSQLIDTATRVNT SSTLIDLIFVSDKELYNEWGIFETSISDHY LVYTVRDFSAKIHNKSEYAEFRSFKNLDE EKFYSELRSTPWQVIKDIDNVDVAWD MWLTLFNNIINKHIPVCKKRIRKKACPW ISDEIVKLMTERDKVLKAAKANKSDDL WDTYKYLKNTVIHLMRKRKREYNADQI SQNHNNQKQMWRLRNLPKSLPISP NSINVDGANVSDNANIAEAFNDFDISN ATKITKDLRRNKSQLRVDEDNIGPHTS QFNLNLITEDFVLKQIDNLSIDKATGED HISCKLLKMTKNVIAESLCDIINKSLSTGV VPREWKKARVVPIFKSGDISSLNNYRPIS	cd01650	452	710

					GTCAAAGATCAGAA ATAAATAAATAAA	ILPIVSKIIERAVHQQLSEFLDVNKLHPN QSGFRPKHSTATALAKLVNQWSLNIDN KEISGVAFIDLRKAFDTVDELLHLKACI GCSNRTIYWFKSYLGDRQQVTQFKGAK SHPSTIKLGVPQGSILGPLLSIYVNSLPE CTPEGIIDMYADDTTLTVSANDATVVEE KLTVALNRVMAWIKENRLVLNTEKTCV MLIGSPANLKKIESFTISLNGNLIKRVQF TKCLGLLIDEELKWSKQIENVSKLTQKNI SMIKRAKSFLPQKSLKLLYNSLVLPRLDY CSVVWSNRFHSHTTKLQKIQKRAARIIL NKTYDTPSADLFTSLKWMTLDKRFELSR VMMIYKCVHNLAPSYLQVGLINPNDV HEHHTRQIDSGCLRVPRFHTDCYKCSPI VSSIFEWNKLDNSLRTASSVNSFKRLFK QSCNL			
	CR1- 2_IS	GCF_00289 2825.2	Ixodes scapul aris	TT	TGATCAGTTGAAGCG CTGATGTAGTATTTG CTTTTGCTCTCACAC CTGGTGATGTCTCTG TTTTGTGCTTTTGATT	MRQHNNILNDQGNAIDLCFSNTDGIH VARSELSMVPEDSYHPALDVVLSIFDPP VLSPSNGPSPHTSYKYSAGDYTGMYHH LSGVDWASILESSDINGQVYLLTTVVQE AMDMYIPLKTSKKSFRPIWFSRELRTLL	cd0 165 0	316	567

					TCCCTTGTTAATCTGA GTGTTTTTCATACTTG TTAATATTTTTATTCT CAGACTGGAGAACTA CGTTCTCCTTCTGCAC TCGTGTGTATGCGTG TTTGTATGTGTGTGT TTTGCTCATTTTTTTTT AGTGTGCCAGCACTA AGACCGATTGGTTGT TCCTGGACACTTTAA TAAACATTGATTGAT TGATTGATTGA	KRKDFLHRRFKKSGLPKWEVEFKICRKR CKKILNRDRQAHSDSVESDLVKNPKKF WRYAKSRLGKNIESEIVLNQHSDSAKG MSELFADHFASVYINNNNCSSSILPHSD VFGDSLGSVVVDEECVRLCIKKMKPTLS AGHDGIPAAIVKAYHELLTPILCTIFNNS LTSGIFPDDWKLAVVVPVLKSGNSGDV LNYRPISLLSSFSKLFEMVIHQFLSKFRSI VIPNQHGFMSTSTNLVSFMSAASE VVCNRGQLDVFYFDLSKAFDVVNHNIL LYKLSSYGVCVSVYNLLKSYLSDRHAYVR VNNTTSSLYKASSGVPQGSVLGPLLFNI FVNDVSRVILNSSFLQYADDIKLYKKISTL EDCIALQNDAYSFGRWCLDNDLRLNHS KTKIMTYSRKTHDIFFPYSYHGELLSRVS ELRDLGVVFDSTLRFDTHTVIKVVQSALR TLGAVSRITKEFKSPSAFFTLYCSLVRSQ EYASVVWNGISKTNSSSVQNKFISI FKYRYLENGDPCLNGVCNSQLVNLLSL QRRVKADLLFLKTLHGSINSPSLLSEV SLRVPRVPTRLQSSFYIARHFSNLNPIQR			
--	--	--	--	--	---	--	--	--	--

						SAECYNRYSDTIDIFNSCGSKFGNAVQIF LLQKEG			
	Jockey -1_DSi	GCA_00438 2185.1	Droso phila simula ns	TCGAGCTTTATTG CAAGACCAACCA AGCAGATATTCTT CTCATATCCGAAA CCCATTTTTCAGA CAGGTCCTACTTC ACCTTAAGGGGG TATGA	TAAAATTAAATTATA TACATACATTATTATA TACATCATACA	MPRVSHPPPEWFSILIRSSIKFDTQPPIQE DWVQCARISVSTTHGDITLAAAYCPPR FRATDDCFSSLLDSLGPKFVIGGDFNAK HTWWGSRVCNPKGRALCRQVLSRNLN CHGTGTPTYWPTDPAKTPDILDFAISGG LDPARIQTREVEDLLSDHSAIAVSVYASA MLRSAGKKLIFKTSISKYHHWLNNSTV TNPPLDTPELIDAAVVNLTSQIQEAATY AAPRFSNQTRDRSRDVHFWNPRVNEL KSEKRRLLRWFLSRNPRDKTALNRAT KELKETLVRLRKEALDSFVGNLAPGDPH HNLWNVTKRIKRPLKRIPPIRRQDNSW CRTETERANAF AEHLEEVFTPANRCSPN EAAETARLLSEPLFDSSESIPQATEEEISNL IAAMGNHKAPGDDAINAIALKLLPPISK SLITTIINKCLQLGHFPSSWKRAEVVMIL KPGKHEANPASYRPISLLPVLSKILERVFL TRVLPVIEEAGLIPDHQFGFRRRHGTPE QGHRVAQYILD AFENKKYCLAVMLDIK	cd0 165 0	438	693

						QAFDRVWHPGLLLKLKNGLPQPYFKFL KSFLEGRRFVRCGEARSDVKEARAGV PQGSVLGPLLYNAYTADLPVIPSQYIMA ATYADDTAFLTTADTPAEASQTMQEQL DLLNTWHTRWNISVNSDKSTATTFATR RGTCPPVSFNGFPIPEVANPNYLGFTLD RRLTWRPHLLKKRKQAENRVREFYWL MGRQSKLPTSTKLLIYKAIIRPIWTYGIQL WGTASKSNINIIQTFENKTLRTVTNAHP FHDNATIHEVLSFPWVKDEIKKSSRQY MQRLQDHPNLAQDLLDTSQRTRRLH RSHPLDLSGIRHAN			
	CR1-4_MU n	GCF_01227 5295.1	Melopsittacus undulatus		TGATGAAGGGAGTG GAACACCTCCCTTAT GAGGAGAGGCTGAG GGAGCTGGGTCTCTT TAGCTTGGAGAAGA GGAGACTGAGGGGT GACCTCATCAGTGTT TACAAATATGTAAAG GGTGGGTGTCAGGA	MAFDTVSHSILIDKLRKGLDDQVVRW IKNWLKGRRQRVVVNGAESSWRSVTS GVPQGSVLGPVLFNIFINDLDEGTECTL SKFADDTKLGGVADTPEGCVAIQRDL RLESWAGRNLMKFNKGKCRVLHLGKN NPMYQYRLGVDLLERVKGKPGGPGG AEPQR	SSF 566 72	2	126

					TGATGGAGCTAGGCT TTTTTCAGTGATATCC AGTGATAGGACAAG GGGCAATGGGTGTA AACTGGAGCATAGG AGGTTCCACGTTAAC ATCAGGAAGAACTTC TTTACTGTAAGAGTG ACAGAGCACTGGAA CAGGTTGCCCAGGG GGGTTGTGGAGTCTC CTACGTTGGAGATAT TCAAGGCCCGCCTGG ACAAGTTCCTGTGTG ATGTACTCTAGGTTA CCCTGCTCTTGCAGG GGGGTTGGACTAGA TGATCTTTTGAGGTC CCTTCCAACCCTTGG GATTCTGTGATTCTG TGA			
--	--	--	--	--	--	--	--	--

	Tad1-65B_B G	GCA_90051 9115.1	Blume ria gramin is	AT	TAAGTCCTGGCGAAA TAGAGGTTTTTCAGG ACAAAGAGTTGTGA GGAAGGAGGGAATG CGCAGGGCGCTTAAT TTTCTTGGAATTATT TCTTTATTTATGTTTT ACGTATAGATTTGAC AAGCAAATATGTAC TTAGGTTGTAGTGTG TCCCGAGGGAGCATA GACCCGTATCACGCC TATTGGGTGTAGGGC GTAAGCCTGATACGA AATGAAATCAAATCA AATCAAA	MKIAECGV PQGSPLSPLLFLLYIAILLQD ENKNTKFGYADDIAILSIGNTASEAVAE AQKEVDKLVQLASENRIDFDSTKSELLVI GGGPKKKLDTSGLSVQIQGHSVPTSPY VRWLGVWLDSQLNFKQHVQEWCCKA QRITQFMRQTNSVQRSAAPGPMIKVV QACVLSTALYGAEAWWPGLTRITTQRG KKVCTGVGWHTDLLDSTILKAVRAALP VWRTTPNMALHRESGITPSAVILQQRN ILAAARIQLDDWHPLVFRSREDPKDTI SRLGLCICQKNRSFKFPEWYLTRLEISAK NVPMAEGPGKLYFPEAESIRKNTGSDF QDEADYAKQWIASLSAETVCAYSDSSS CGPARSAWGYTLYKGGQVIGSDSGPLL GAEVYDAEIFGAQKALERALGKAGNCPI KVLLDNTTAVRALRTGQTKSSQEAVEVF TRLARQARSVEIRWIPAHAGVMGNED TDRLAKEALRNLPDETDTSPEMRDRP KYTFAALTRFVRQRCDEVVESWWRDH RPSRYEELDLQMKRKRPPPELALPRWAH HRLITARTGHGDFIKYHERFGHEFENCE	SSF 566 72	6	148
--	-----------------	---------------------	------------------------------	----	--	--	------------------	---	-----

						CVCGREKRPWHFAECRIALQQWRNLK HEPPPRGREMLAEKGWERFLQFLTGTK CYERQTLHRED			
	RTE- 1_EL	GCF_01100 4845.1	Esox lucius	CTCCTTGAAGTGC CACCTTAACGTGG TGGAGGGGTTTG AGTACCCGAGTG ACCCTAGGAGCTA TGTTGTCTGGGGC TATATGCCCTGG TAGGGTCTCCCAA GGCAAACAGGTC CTAGGCGATGGG TCAGACTAAGAG CGGTTCAAAACCC CTTAATGAGAAAT TGTGCCGTGACGT CGCCCGGCATGG CGCAGCCGGGGC CCCACCCTGGAGC CAGGCCCGAGGA	TAAGCGGAAGATGA TGCAATGCAATGC	MQPDPRTKLALGTWNVTSAGKEPE MVREVERFRLEVVGITSTHGLGSGTTLL ERGWTLHHSOVAHGERRRAGVGLLIAP QLCRHVLEFTPVNERVVSRLRVGDRSL TVVCAYGPNGSAEYPTFLES LGGVLESA PTGDSIVLLGDFNAHVGNDSDTWRGVI GRNGPPDLNPSGVQLLDFCASHSLITN TMFKHKGVHQCTWHQDTLGRRSMID FVVVSSDLRPYVLDTRVKRGAELSTDHH LVVSWIRWRGRKLDRLGRPKRTVRVC WERLAESPVREIFNSHLRQSFDWIPREV GDIESEWTMFSTAIVEAAARSCGRKVS GACRGGNPRTRWWTPEVRDAVKLKKE SYQAWLACGTPEAADGYRQAKRTAAR VVVEAKTRAWEFGEAMEKDYLASK RFWQTIRRLRGKQCPTNAVYSRGGQL LTSTGDVVGRWKEYFEDLLNPAVTSSIE EAEDGESEVDSSITRAEVTEVVKLLGG	cd0 165 0	529	784

				GCGACATGGGGC CGCCCTCCCGTGG GCTCACCACCCAC AGGAGGGACCAT AAGGGGCCGGTG CGAGAGGATCGG GCGGCAGTCGAA GGCGGGGGCCTG A		KAPGVDEIRPEYLKSLDVVGLSWLTRL NIAWQSGTVPLGWQTGVVPLYKKGD RRVCSNYRGITLLSLPGKVYARVLERRIR PIVEPRIQEEQCGFRPGRGTLQLYTLY RVMEGSWEFAQPVHMC FVDLEKAFD CVPRGILWRVLREYGV LGPLLRAVRSLY DRSRSLVRIAGSKSDLFPVHVGLRQGCP LSPVLFVIFMDRISRRSQGPEGVRFGDH TISLLFADDVVVLAPSSQDLQHALGRF AAECEAVGMKISTSKSEAMVLSRKRVA CPLQVGGECLPQVEEFKYLGVLTSEGR MERIDRRIGAASAVMRS MYRSVVVKK ELSRKAKLSIYRSIYVPTLTYGHELWVMT ERTRSRIQAAEMSFLRRVAGRSLRDRVR SSVTREELRVEPLLLHIERGQLRWLGHLF RMPPERLPGKVFRSRPTGKRPRGRPRT RWRDYVSRLAWERLGPPEEELEEVSGE REVWASLLRLLPPRPGPG			
	ORTE-3_AAe	GCF_00220 4515.2	Aedes aegypti	CATACGTTGAGTG ACTACCGTTGAGA GCGGACGTGTCT	TGATGCGATTCAGAA CGACTGGAGTTGAA GAAGAGAGGAGAGA	MMTMTFNGIRRLGWFLSVLSLVKQRL NSSHESNSLSKLILKQICVKLNGVNFTQE DVDELNSIILSKIEGWEQDLNKFIEFFDN	cd0 165 0	1078	1338

				GTGCTTGCCTGCT	AGATCACCATATTGT	IQFEVFDIAEMANHCLLLAVSQIFECEV			
				GTTACAGTGGTTC	GCATTGCCAGAATAA	RVSNDKDDCKIIGEGYKNSWFGVLEIG			
				GAGTGAAAAGAT	AGGTATTGTAAGTAC	MESSERVEKVTRALKGSSKIKLGKCGKD			
				CGTTTTTGCATT	AAATCAATGTTTTAA	RQVNGMAGKGNEPLIVKEECLWYTTIG			
				GTCGTAGGAGGA	ATATGTAAAATGGAA	NWIIIESKGEDDGKLYSGLMMNVIQK			
				ATTGTGCTGGGGT	GTCAAACAAGTATAC	MNEGEIGEQUEMMEMLKQIDLFFEMY			
				TTTTGGGTTTCCG	GTTATACTGGGAAGG	GEILNRLEEEVMINGNALYEIWKIVSKLA			
				GAAGTGCCGTGTT	TTTAATTTAATGGTA	VALNREIIVLMRGGEEIVLGVTERKRKF			
				GGAGCAGTTGGG	AGTGAAATTATTACA	ASFWLVAEFQEVGVRYYSIKSSKLIQEW			
				AGGTTACCGTGCT	AAACATAATCAATAG	KENMSIKPAIDVFEKRAEEVSYNLRFKES			
				GGGTTTCCGACG	CACCAATCGTAATTTT	NQQAVIKRIDGDGSLIRAMLDQMGR			
				GATTGGACTGGA	CATATATTATATATCA	DQNYMSDYNVTKIRNNIASYINNSRE			
				TCCGTGATTTGGC	AACAGGACACCCAAT	RFNVVLYEHIEEVGTIERVIENIREKSYWL			
				CAGCAACTCGGCT	TGAAGCAATATGGAT	GEEAIITVAELFEKQVEVYTNQGLSVRIG			
				GGGTGAAATAGG	GAGGAGAGTGGAAT	VRKDNESALRIFYTGNHYDSVISIAEEEE			
				CTTTTCGGAGAGT	TGCCAACTCCTTATC	GEIEAEMININEAEGNKMQAGGESRAV			
				AGTGAGAAAACG	GCTAACATACTGTTG	KSNKVREVNESSSHSKKITQYGKQSTNT			
				AAAAAGCAGCTG	ATAGAGGGATGGTC	TGDSNNHEQSKKKVEKQLRKDSEIKDK			
				GTAGTGGAGTGC	CTTAAACCCCACTCAT	ENKALKIVSWNVRGCRKVRDEIDDIL			
				GGAGGAGGTAGA	TGAGACATCCTTTTG	VANDVMIAALQEVNTIGDDLSTNNYS			
				AGCCTAGTGGAG	TATGATGAAATGACC	WQVKDNHTNKARGLAIVVRKNQRIRIL			

				GAGTGCGGAGGG	GGTAGATCGGTGACC	QAKECRKGIMWIKLTINGVKLIVINVHA			
				GGTGAAAGTGCG	AGTATCTAGGTTATA	PNSNQKNFLSYLNEIVTRESDRKSLIVLG			
				GTGAGGGCTGTA	ATTTTTGGAATGCCTT	DFNAQIGSEDLQPEDRVWIGTKLGHR			
				GTCGGTAAATTCC	AGGTAAACCTTAAAT	CNDNGELFKYFLHLSKLNSTKIGKSIEI			
				GCATTTTCAGATTG	AATAGCAAAGAACTG	TWRTKDKSSQIDHVLVPQYTSFRIRYIK			
				CTGATTGTTAGTG	AAACAATCAACTGGA	GCWTDLKT DHVMLTIGLVFTNNPVLKK			
				TCCTGAAGAGAA	TAACATCATATAAAA	SNKDTARINVDLLKIPEKQKCFREKLSSIP			
				GACTAAATTCCTT	CGGACAGTCTACTGT	TKTDIQTDTLDKCYEEMASRLRIAAYNT			
				TAATTGGGGCAG	CTAACGGCTGTGCCA	LKVSTVPNTPRRKYALDRLRDALKVSKK			
				CTGGCGTGGTTAA	TGCAACAATGAGTGG	HPDMINYKYRVANRRHEYQI AVKEHRE			
				GGGATTGTTCCG	TTACCCTTACATACAT	KEITDFFKNLNDYDAGVRVRKTYNYLKD			
				GGGAAGCGGAAA	ACATACATACAT	FVKRKKQKKAMISSRVWEDILKTCEGPE			
				GGGAGTGTGAAG		VEFCAEHDYTPLNPPTNDEIKSILFSAC			
				TGTTTGCCAGAG		NGKSPGSDRVVAEYIKA ADEETLERIIEI			
				GAAGTTGGCGGA		KAVYIRNDMPKEWKKSTQIPIPKKSGAS			
				AAAAAAATTGG		EADHFRRISLCNVVYKIYAKWIAKQLRA			
				AACGCCGAAGAA		YAGEPGLHQAAFTANRSTDDHIFVARR			
				AGAAAAAGTTTT		MMEEYWNSGQDLFILALDISKAFDNVK			
				GAGGCGAGAGTG		LQSIGKVLIDLGVPSRLVDRVLSCVKDEV			
				GCGGGGGGGTGT		TRIMWENQYSAEVKRGKG IKGQCPLSP			
				TGCGGTCGAAAG		LLFNYLIQDVLKKVVAEIP ELKLMGINQL			

				CTTCAACGGAAG AAAACACGAGTG GTGTGAGTGAGA AGACCTTGATCGG TTTTGGCGGTAAA GCTCGTTGCAGG AGTTCGGAAAAA AAAAAAAAGAAA ATTGGAAGCGAG AGTGAGTGCCAC GTAGAAGCGAAG TGGAGAAAAGTG GCATCGCACACTG TGAGAGGAGGAA GTGGAGCATCGA AATAGGCAAGCG GCTGAGTAGCGA CAGAGAGAGAGT AAGTGAGAAAGC GCATCTGTAGCG GGGAGGCAGAAA	ILPLILVFADDILIIARTKAELEKILHELELQ LAKVGLEINRDKSQVLIRYPNDKAQHPK EILLNNKIFKVSDSLAYLGICLTATLNRKL TCKQRCLNAVKSSRIIVEFCKTFKPSWEL GKLMYKTVIAPSLLYGTVSVMTKKSRL KIGNYEKLILRNIYQYCKKPKDLKFNARK LLDGKTINRRIRVGRISYFGHVRRREKN HPLKLAYKLKFDKKKEGRPSLTWLKSLD QDFERYQNMNRSSWEELSKDREKIKKK LDEIYKNEESEISDGESSDGDQNSRYKH WKWKRRNEK			
--	--	--	--	---	---	--	--	--

				TAAGAAGCAGGG CTGCTGGTCGATC CGTGGCGAATGC GGTGGAGAGCAC GAAGGAAGATTC CAACGCAGCGAG ACTGAGTGCACC GGTTTCATTGGTG GCAGTATTGGTG AGTCATTCACTTT CTCGTTCCAATAA TTTTTCCAAGTAT CTCGAAGTTTGGT CCTC					
	Proto2 -1_OL	GCA_00434 8135.1	Oryzia s latipes	CA	TAATTTATATCTTTTT ACTTTTTATTATGTAT ACTGTATTTGTATCTT TTTTCTATGGACCAC ATCTGAGTCTGAAAT AAAAGTGTGTT	MVVKWGNVLSEPFYVSNGVQQGGILS PFLFNVYIHDNSTNLNLCKTGCVVGCQT FNHLMYADDLVIFSPYSAGLQTLKVC DYGRDFDIKYNPSKSKVMIVRTKDDRN VNFPVFKLCGTTLEQTSSIKYLGHFITDD WRDDRDIQRQYCKLYAQANMLMRKF SACSLQVKCSLFRAYCTPLYTAPLWWN	SSF 566 72	13	208

						YGRGTFHKFLVAYNDAFRLLLQVPRWY RASQLFVDAHIPTCEALLRKLTYSLMERL DKSKNSIIMSLTDPKRSSCRYTSALRKH WIQRLYSF			
	Penelo pe- 2_Sln	GCA_01036 7695.1	Soleno psis invicta	AAAGATCAATGA CCAGCTAATTAAA GACTGGTTACGA ACCACGGGAAGA TTCGTTAGGGAGC ATCCTAATATTTT GTTTACCAGAGCG GACAAGGGTAAT GTAACAGTTGCCC TTGACAAGAATG ATTATATCTCGAA G	TAGGTTCATCACCCC TTCTCCTTCTATTCTT TTCGTTGCGGCCACC GAGGCTATCTCGTTT CGGTCTATATTATTTG TATCTATTCTCACCGA GATGTAACATCCCAT CTCAAGGCTGTGACA CGAGTTTTGGTGCTT AGAGGTAATTTGTAT TTGATTTTCGCTTTGA CTTATCTCCTCTTTTT CCCTTTTTTGTTGAAT TTACATTACTTTATAG AAGTGTATTTTTTACA ATTTTATATCTAATGA CTTGTACAGCTTCCAT	MEGMLSDESTYVKLSKDPIQKLTKEARS LLSIWKDKGFIEQHIYRSLTTDGVLPRA YGLPKIHKPNNPLRIIISSINSPLHSLAYFL HNIIYNSIPEADSFVKNSFHLVKNLNLGH VESDHVLVSLDVVSLFTNPKNLAVESV TTRWDLISKNTTVPVNEFLKALNLIIDST YFKFNNTCYKQIFGLPMGSPLSPILADLV MQDLECKALKLLSFHLPFYCRYVDDIILA APANSLSGILDTFNSMHDRLQFTMEREI EGKISFLDVMLINVNNRIIFDQYRKPTFS GRYLNFFSHHPLCHKKGVIIFGLVDKLLLL SHPQFHCKNFINLVNILLENNYPLNFIFS MINSRLKHLFYKDDTYVEDKRVRRFFTI PYIESVSEKFSSFCKFDCSVAFTIPSKLN KFITTGKDHTERLSCNDVVYKINCKDCD SSYVGQTKRQLKTRVSEHKADIRKASSP SVISLHRIGENHEFDWENVKILDKEPSY	cd0 030 4	189	269

					TTTTATTCTTATTTATT TATTTATTGAATTTTT TACTTATATTCCGCTT ATTCTTCTTTTTTAAC TATTTAATTTTCATTG TTGTATGCCACAAGT TTTATAGTCATCATTT TAACATCTTTCTCTTA TTTATTTATTTATTTG TTAGCAACCAATCGA TGTTGTTTCACAAAA CTCTTGTCTCGGCAC ACGCTCTGTGA	VKRSISEMLHIKNQMHGLNKQSDTELL SDAYLPILERIPPS			
	PENEL OPE	GCA_00798 9325.2	Грынн a virilis	GGTCGCCAGAGC CATCAATAAATAT CAACGGAAGGCA CGCCGTATGCACA GCAACCAAC	TAAAATAATGTATGT AAAATGTTGAAATA ATGTTTAATTTATTGT ATTATAATTGTTAATT GTTTTTGTATCTTGG TGTTAGTGCCCTGAA GACGGTTTGCCGATG TGCAACCGAAATATA	MSYAKIKTKYKDSKRTINKFQLTLVKLTK LKSSLKFLKCRKSNLIPNFIKNLTQHILTIL TTDNKTHPDITRTLTRHTHFYHTKILNLL IKHKHNLLQEQTKEKAKTNIEQLMT TDDAKAFFESERNIENKITTTLLKKRQETK HDKLRDQRNLALADNNTQREWFVNKT KIEFPPNVVALLAKGPKFALPISKRDFPLL KYIADGEELVQTIKEKETQESARTKFSL	cd0 030 4	470	558

					TCGGAAGAGAATTG AATAAAATTGTTTTTC ATTGTTTGTTTAAACA AACACGGACCTCGAG CCAGCCAACAAATAA ATATTGAAATATGGA AA	VKEHKTNNQNSRDRAILDVEQTRKL LKENINIKILSSDKGNKTVAMDEDEYKN KMTNILDLCAYRTLRLDPTSRLQTKNN TFVAQLFKMGLISKDERNKMTTTTAVP PRIYGLPKIHKEGTPLRPICSSIGSPSYGL CKYIIQLKNLTMDSRYNIKNAVDFKDR VNNSQIREETLVSFVSLFPSIPIELAL DTIRQKWTKLEEHTNIPKQLFMDIVRFC IQENRYFKYEDKIYTQLKGMPMGSPAS PVIADILMEELDKITDKLKIPRLLTKYV DDLFAITNKIDVENILKELNSFHQKQFT MELEKDGKLPFLDSIVSRMDNTLKIKW YRKPIASGRILNFNSNHPKSMIINTALGC MNRMMKISDTIYHKEIEHEIKELLTKND FPPNIIKTLLKRRQIERKKPTEPAKIYKSLI YVPRLSERLTNSDCYNKQDIKVAHKPTN TLQKFFNKKISKIPMIEKSNVVYQIPCGG DNNNKCNSVYIGTTKSKLKTRISQHKSD FKLRHQNNIQKTALMTHCIRSNHTPNF DETTILQQEQHYNKRHTLEMLHIINTPT			
--	--	--	--	--	---	---	--	--	--

						YKRLNYKTDTENCAHLYRHLLNSQTTSV TISTSKSADV			
	I- 3_SPur	GCF_00000 2235.5	Strong ylocen trotus purpur atus	GGTAAAAAGTCG CCTCATATAACTA ATAAGTCCATATC AGACCTATTGATT GTACAATGGAAC GCACAAGGCTTG ACGGCGCATGGC GCAAACTTTTAT CTTCCCTTATTGAT AACAATTTGTTAG GTATTCACCCAGA TATA	TAGCTTTTAAAGCTA TCAGGGATCCCCTTA ACTTATGTCGATCCT GTAATAATATATAAA ATGAAGTATTTGCGG TTTTTCCATGAACAAT TGCAACTAACTCCCA TCTCGCCTTTTATGTG CATTTTATGCAATTTT AATATTGCAGCTTTTT TTTTTTTTTTTTTTT TAAAACCATATCACT ATTAATTCATTTAC CAAGGAAAGTCGCTC GTGGGCTTGATTCC GAAGCCCACGCCAAA GCTATACACGAAATT GACTGGCGCAAATAA TCTTTGAGATTGAAG	MCIQETWFNDFNVL SIPNYVCVCKNRP DGRRGGLVMYLKNDITFQAMDT PRNA EYQKVDIFCKNGVISIFNYYNPCK KIELID LNNMIRNVPGKYLILGDFNSHNPL WGS NKLDTNGRTVESFIENNDLVLLNDG SPT RLDPYYGTYSNLDLSLVSKELARNT EWK VLSNNGSDHNLVCLNFKHVYNRP GGE NMGQRWSFRKVNWEEYREEIEEAL RK EGYENLGAELNIQGRYNVIVKVIK ETSN KCIPRVQIAKRRKVNPNWNNACKK AV QERNRLKNRLRRYYDLNDVKKYLK KKA VAQKVIRRAQYDYWLEYCRRMN RFTDI GKVWEQVKRMRDSTMNKHIFHNL FV DGNILKESEEKANAMATYFKSVDES LSV TKNENENILYNNSSTNDDDCINDDF NM SEMGNVLKEMKNSSSGKDDIYLIM IKN LPNKALVYLLNLYNDIWKISLFP HQWYE VIQIPILKSGKDPTQVTSYRPISLI STFSKI MEKMIKNRLTWFEKNRLLNPIQSG FR	cd0 165 0	462	717

					AGCGCCATCTAAACG AAGAACCAACCAACC	KYRNVKDHLVRLETDIHISLKS KKYTVAV FLDLEKAYDSLWKTGLLRKLSILGIGGK MLLFIKAFLLSRISNVRINNYMSRHYTLE NGIPQGSSLSPLLFTLMLNDLKINVSDV KLSCYADDIAIWCSNKDLELARRKLQEA LHELEIWCWTWKLNVSPKSAVLVFTN KQKKDILLQIKGHNLPICEQYKFLGVFRD SNLTWSVHIDSIIDKCKKRINLLRCVSGT PWGSKYEILMMIYKGLILSILDYGSEMY DSACISLTKKLD SIQYRALKIITGTIHNVSL EALQVLTGELPLRYRRRLNDNFRVYLLS STLNHPTKEVLGDDYDLNNLEWKIGEG PFKLRSFRDEVVVEGQTNLLPKIPEWRL ARPQVSFEHHGKVQKKEMPPFEIKSIVL DSIGNSWNGFLHIYTDGSKLEDGKTGA AFYVSHYNVVKFFRLSHVCIMRAELVAI LMALEWINTFIESSSVIFSDALSALQMI DSYFIRCAIVNEILSIINILQFKDVCVAFE WIPSHCDIRGNEIVDRAAKKGAMKRCI DIVIPNTRSELAASRLIHYNKLWQAEWD NAEHGRFLYALHSDVSKPIHRQGLSRRE			
--	--	--	--	--	------------------------------------	--	--	--	--

						ECLIQQLRVGNCSLNYYLFQINQHDTGL CEVCQEDETIKHFIFYCTKYDKDRRLKE RIKLNNPNLKQLMSNPFTRELLGYIYD TNRFR			
	L2-3_XT	GCF_000004195.4	Xenopus tropicalis	TATCCCTCTGCTG TATACTGCCATAT AACCCACTATATT GCAGCCTCTTTAT TAATGGTAATTCT CATCTGTGTATCC CTCTGCTGTATAC TGCCATATAACCC ACTATATTGCAGC CTCTTTATTAATG GTAATTCTGTCAT CTGTGTATCCCTC TGCTGTATACTGC CATATAACCCACT ATATTGCAGCCTC TTTATTAATGGTA ATTCTCATCTGTG	TAGCAATACCCTGTG CCCCACCTCTCACAA CTCTGGTCATGCCCCA TTCCCACACCTTGTGT CTCAACCCTTTCTCCT TGTAGATTGTAAGCT CTTTTGGGCAGGGCC CTCTTCACCTCTTGTA TCGGTTATTGATTGC TTTATATGTTACTCTG TATGTCCAATGTATG TAACCCACTTATTGTA CAGCGCTGCGGAATA TGTTGGCGCTTTATA AATAAATGTTAATGT AATGTAATGTAAT	MFYPALLLFALSCFLKHLQPHPPSSQKGI TITASLLSSPSLSAHEVFKLLCSITPVNSFS SHKTKTYKSRSHLAFLSFLLLAAGDISPN PGPYPIPVLLGTRSPNPPCTPLKYRTLTPI PKASIQISCALWNSRVCNKLTSIHDLFI AKSLNLFITETWLSPTDTASPAALSFG GLHLTHTPRPGNRAGGGVGLLLSPRYS FKVLP PPPSLSFSSFEVHSIRLFSPVSLRIA VIYRPPGPTSQFFDNFSAWLPYFLHIPTS WDTLTILDSFPPAIHLKSFHLLLLFHFP LRFTLLGSSPLFRFALRSYTDLLDRPHNS LTTFLPGFLTSTFIYRPPGPTSQFFDNFSA WLPYFLSSDSPSIVLGDFNIPVNNPNAP ATSKLLSLTSSFGLSLCSDSPSHSNGNSL DLIFTKCCSTSNFTNSPFPLSDHNLLTFQ LSLTPSTPSAVPQTYTYRDLQSLNTHLS SSFDSLHSNILT VSCPNQATSVYYSTLTT	cd01650	702	964

				TATCCCTCTGCTG TATACTGCCATAT AACCCACTATATT GCAGCCTCTTTAT TAATGGTAATTCT CTCATCTTGTGTA TCTGCTATGCCAA TAACTCATTCTATT AGGCCTACCTTGA TTAGTACTTGCC ATAGCTGTGTCAA ACTCACATCTGTG TCCACTCCTCAGT GCCAATATTCAC AAGTGCTAGGAA AATTGCAGTAAA CAAGGCAGTTCAT GGAACTAACTCAT ATCTCTTCTCACTT CTAATGGTAATTC TCATCTGTGTATC		ALNELAPAKTKRMRPKPLQPWHTAHT KSLQKQSRTLERRWRKSRSESDFCSYKS ALLSYNTTLCQAKQTYFTALINSLSKKPA QLFSTFNSLLSPPPPPPCASVTAQAIAEH FKNKIDTIRSDIAQLNPHNHSPPSLHAP QSLGFSFPVTEEEVSKLLSSSHLTTCPLD PIPTKLLRNANPCLIKALTHLFNLSLSTGI FPSQLKHALVTPILKKPSLDPSNPTNLRP ISLLPFISKLLERLVYNRLTLFLSDNNLLDP LQSGFRQQHSTETALTRLTNDLLSAKAK KHYSLLILLDLAAFDTVDHPLLLQTLRSL GLRDTALSWFSSYLSDRSFSVSYNGESS SPLPLSVGVPQGSVLGPLLFSLYTSSLGK LISSFGFQYHLYADDTQIYLSSPDLNPELL TRVSSCLSAISTWMSQRYLKLNLSTEL VLFPPSNAHIVPEVSITVNNSTITPSSQA RCLGVILDSALSFTPHIQSLIKSCHFHLRN ISKIRSFITQDAAKILIHSLIISRLDYCNLLI GLPLQRLSPLQSIMNTAARLIHLSNRSSS AMPLCQSLHWLPLPSRIKFKLMTLTFKA VHNSAPPYISELISRYHPTRLRSSTDLLL			
--	--	--	--	---	--	---	--	--	--

				CCTCTGCTGTATA CTGCCATATAACC CACTATATTGCAG CCTCTTTATTAAT GGTAATTGTAAC CATCTTGTATCTG CTATGCCAATAAC TTTTCTGTTAGG CCTACCTTGATT AGTACTGTGATA GCTGTGTCAAAC CACATCTGTGTCC ACTCCTCAGTGAG CAAGAGATATTTT TACAGCCCCAAAA CTTGGTTCTGGTG CTTGTAAGTCT CCCCCTCCCCACA CAGCCTGCTTACT TTGATCCCTGATT GAACTGATTAATT	NSSLIPSSHARIQDFARAAPLLWNSLPQ TVRLSPNLSAFKRSLKTHLFREAYPNLV			
--	--	--	--	--	--	--	--	--

				AGTCAATTAGTTT CTGGCTTTCCCCA ATCAGCTGCATAT AAATTCAGATGCA GCCTTGGGGATG GCACTAGTGTCAG GGGCAGGCACTG GTGCTACCATTTC CTAAGTGCCAACA TTTACAAGTGCT AGGAAGAAGGGA ATGGCAGTTAAAC AAGGCAGTTGGA CAAGGCATTGAC AAGGCTACAACT ATCTCCTATCAAC ACTGAACAACGT GGAATCGCAAGA AAACAACTTTGCT TATTCATGGTAGT					
--	--	--	--	---	--	--	--	--	--

				TTGTTATTTCTAT CCCAAA					
	L1- 1_VVi	GCA_01103 7095.1	Vitis vinifera	GGAGTGTTCACT GAATTATTGTGGG ACTAGAAAGAGT GCGGGGTTGCTTC AGATTGCTAACTG AAGTTTGGGTTTG TTGGTCCAGTGAG GGGAGAGGGGCT TATTAAGATGGAT GTT	TAAGGGTGTGGATA GGGTGGCTCTTCGTT CCTTGGCAGGCTGCT AGGGAGCGTGGGGT GTTTTCTTTCTTTTTC TTTTCTCCTTCTTGTA CTTTAGCCACTGCTT CGGTGGTTTTTAATA CAATCTCTTATTTAT CAAAAAAAAAA	MKLRVLSWNVRGANDSEKRKVIKALIK SQKVDVVCLQETKMQEMSRMIVRSLG VGRCLDWKVLNSRGSSGGVLVFWDKR VLHLLEAEVGNFSVSCRFKNCEDDFCW LFSGVYGPSLMKEREDFWAELGAVRGL WSDPWCVAGDFNVVRFVPESSRGGR SASMRRFSEIMEDLELRDLPLQGSFT WKGGLNNQSHSRLDRFLISNEWEDHF SGSVQYVLPKPTSDHFPILLDGGGVRS PMPFRFENMWLKEEGFKEKMQGWW VGLNFGSASYVLVSKLKALKSLLRDWN KLEFGKVEVNKALALSQVDFWDMELS RTLSVQEVDARRGAKEDFKKWALMEEI SWRQKSREVLWREGDRNTKFFHKMA NAHRRGNMLSRVKINGVWLTKEVEK EGVNEFKAMLSSAGGWRPSLRGLSFE RLEAVDAASLEEPFSEQEVMEALKGFC GDKAPGPDGFSMAFWQSSWEFVKEE VLGFFREFHNNHGRFVKSLNATFIVLIPKK	cd0 165 0	496	763

						GGAEDLRDFRPISLVGGLYKWLAKVLA NRLKRVVGKVVSKAQNAFVQGRQILD AVLVANEVLDSVLKNKEEVVMCKLDIEK AYDHVEWSFLFSVMRKMGFGEKWIR WMKWCISTVSFSVLVNGSSSGFFQSSR GLRQGDPLSPYLFVLVMEAFSSLLRKAV AGGFVSACKARSRGGEGVNVSHLLFAD DTLVFCGASKVQLLHLNWILMWFEAM SGLRINLDKSELIPVGCVNNVEELAAAIG CKVGSPTSYPGLPLGAQYRSRAVWDG MEERMRRKLARWKSQYISKGGRLTIRS TLANMPIYFMSMLSMRPRVRLRLRIQ REFLWGGGALERKIHVLVKWELVCLEKD NGGLGVKSLSILNKALLCKWSWRFAME REAFWNQVIRGKYGEEQGGWSSKEAR GETHGVGLWKTLRKEWEVVKSRVLFV VGNGKRIKFWKDIWCGDEPLCVSFPSL FALAVSKDAWVKDVWRCNEGGGSWS PLFSRPFNDWELEEVCSFFVALNRKQIQ QGVDDRVIWRETNCGKFSVKSLYKSLV SGHPISFPSSAIWKVTVQPRVSFFGWE			
--	--	--	--	--	--	---	--	--	--

						ATWGGKALTLDQLQRRGWALANRCYLC QRHEESIDHILLHCDKVRTLWVLLYSMF GVQWVLPATVKETLSGWNGSFVGKKR KGVWKASPLCLFWTVWKTRNKVAFEE EELSIQRLKASFVYFLWSETKRSIKDGPS TLVDFVGWVASR			
	R2- 2_SMe d	GCA_00260 0895.1	Schmi dtea medite rranea	GTCATAGGGTGA ACTGCAATTCTGA CACGATGACCGA GCTGTGTCAGTTT GCAGCTAGTCGCT AAAGACTCGATCA GTCCGCCAAGTG AGGTGGCCGGGT ATCTGCAGCACTA GAGCCACTGGTAT CAAGAGCAGAGA TACGCGAGTGGA AGTTGAGTACGA CTACCTTCACGGG GTCCTCCTGATAA	TAGAAGGGGAAACAA AGGAAAAACGAAAT GACTGGTAACTATGA AGGATATAGCTGAAA GCCGAAAGGAAGGC TAAGTCCTGAAACCG ATCTACATCTTCGATC CCAAGAGGAACTGT GGGTTAAGCTTGAGC CGACGGAAAAAGCG AATGCATGTTAGACG ACGAGGTACAGTCAC CTCCTCGTGGTATTT GGCGGGCAATGCTC ACTAAATTAAGTGTG	MRKELVTIKNLIIESGATAPAPVPLEVA VEVHQSSSVPEITDQSTTTQEGSYSEPPI HRCENCGREFRTRAGVQQHRRKAHTN EFMEEKEKAAPTKKLRWTDEEKEILIESE IKIIKEGSLKEQHEINKILASRMPGRSQD GIAKIRQKQEHKAEIQRRLHGTVTTDET RGSRTSEITEPIRSLPINTKNWSEDELKR MLAEVVKLRNKNEKDINKKLAEIFPNRT MGSIKSKRTKDKEYQDLVKLTMQTISEN PDNETDFNTSNTENNSTD AEKEVKNYL NMLLLSNNEEELTSTLKEAATLALQG KKTEASEKLNEYASKTLFPGLKITNQTRK REKKISKRETRRQEYAEIQKLYKKNISSA AEKAINGKWSIKPEEEYHNNKDLIKAW KPILEAPPFSDCRPIENIKEMDYALMEIS	cd0 165 0	474	730

				CCACAGTGGACT GTGGGAACTAAA TGTGTGCTCAGCG TCCCTACTTTCTC GTAGGGTAAAGG GTAAGATAACCCA GAGAATATCCCAT GGGAGATATCCA TGGAAAAAGCAC CACGTTAGACAAT CCGATGGTCTAAC TCGGCTCCGAGG GGCTAACTATCCC AAAGGGCTTAA	AGTAGCTGAGAACTG TATGTGTATCATGAA AAAAAAA	TAEIFLAIRAMGKTAPGPDGIKYSKLKK NIQSMAILFNTCLLTSFLPLPLKIARTILIP KQENPGILDYRPLTIASVVTRVFHSILAK KLDNNAQLSQRQKGFRKCDGVAENIVI LETILTNSRSEKRPLCMAFVDLRKAFDSV GHESIIRGAKRVGVPPMLLEYISSSYQN ASTNLFGEILNSRRGVRQGDPLSPILFNF VIDEALENLNRNIGYLLKEEKVSCLAFAD DIVLIAETKGLENHIEKLLEKLNAGLE LNASKCATLMVMKNGKEKSTYISTKVIK IKENDIPTMKATETYKYLGLQMFGKARE QNANEVITEGLENITRAPLKPQQRIHILR DFLIPRLIHKLVLGRVAKKSLKRIDQNIRK KVRNWLHLPKDTTAAFIHADAGDGGGL GVPALEHTIPLLRERITNLRKSNDPVTK ECLRMEYTKQVLGKWSRPTKIGETLAT NKSQLKEAFRKQMLITLDGKGLKDHHE TPTIHKWIRRGENMTGKQFITAVKIRG NLVATKSRNSRGRPEQEKLCQAQCGRP DSLGHILQGCWRTHGMRVERHNNICR RIKAIMKGKESEVVEEPRLQTNEGLRKP			
--	--	--	--	--	---	---	--	--	--

						DLLICHKGKIIICDAQVVDSSNCSLESE NQRKIDYYKKDSVVAEARKLIGRDDDEDII IMAVTFNWRGAISKTSIRDLDMLLDIKS KEVIKMSRKIIRDNSIMVEMHRNRTEKR R			
	Utopia - 1_NVit	GCF_00919 3385.2	Nasoni a vitripe nnis	CCATTCCTTCGTA CGGGTTTTCTGTC CGGCATAGCCGG GTGGGAGACTCG CGCGGGGGAGGT CATATCTCACCAC CATCTTCGAGCTT TTTGCTGCACCTG A	TAGTGGGGCCATAAC ACCTAGGCCCCACAG TGTGGCGATGTCCAT GTGTGTCGTCCTTATT TATTTATTTATTTGTC CTGTCGAGCTGTTTA ATTTATTCACATGTAT GTGTTGTGTGTGTCG TCCCTCGTAGCAGTT TTATTCCGTCCAACCA GAGGTCGACCAAGA TTAAAATAAGCATGG CTTGAAGCAGGCCAA GCGCCATGTTCTAAC CCCGTTTTAGGGGAA	MSGLTGREVKCITVNVLMQQPHTKAI REGDFIVILLPQSDDETLCCLCVGRGRY SGKTRVECLNRHVKEVHPDLTTTFRCW GCGFAAPGDKKYPRKIVTQHCATCVPE VSSAPSGRVDGERRVNTRRRLGIAAAT EASPVRRTRRNGLASPPVEQDISQSAAP PEPARVPQHPEIVALGESADDEVFRSPV NSPPRDWRAAAPQQAASSSPPAVPGIT AATPSNTTGTGNGSANSILAEHPIPTPP PTNTEANGRADIPRSGRTPPPGQQAA RRRAPTTEQRRIVGLLEAATGREQLEEA TTQAMLFLARLTGRRPEPRNAIRPGPR QRHPAQEDVQAQAPDRIQEAKKLQRL YRTSKKRAVQKILAGPNMNCQIDKNTIT AHFVELAARRDGGEDWPDVFDREPT AASGEALCTPITREEVFRRLLKGRNNTSP	cd0 165 0	477	740

					GTTACTTAACCTAAA AATACAACCTTTCC	GPDGITYRDLAKADPGAHVLAALYNSI WRIEATPALWGVSNNTTLYKKGDAMDI SNWRPISLGDTVPKLFAAILADRIKRWA VANGRYASQKGFLEFEGCYEHNFVLQ EAIREAKGGRKELVVAWLDLASAFTSVP HSSILQALEGHGLPSKARNIISSLYTGMT TRFHTAEGPTDPILIQSGVRQGCPLSPD VFNLTLLEVVLREIQRTGEGYTIEGRRTSH LAYADDVAILADSPAGMRLLFAAERG ARAVGLTFNPAKCATLHIAGRGEAAVR PTEFSVQGTPVRALASEEAYEHLGIPTG YQVRQTPINTLRDLLADIGSIDRSLLAP WQKLDVGTFLPRLDIFIMQGAHIDKG FLTEADKIIKKAASWLSLPQRASAELVF LPPSQGGGGLLTVASYSKMLYSSDVTV STIAGSTLRRTVSERLKKRASNIDIARFLS GDLDLPRSTSPSTFWTKVRSAAALRIKTKL GLRWSWCQEGEVLLMACGDPRAPGT RVSPQTKHLVTTSLRRCLNRHYAESLLA KKDQGKVFVTRRSGQSNHFLRSGSFT RFCDWRFIHRARLDVLP LNAAKRWQR			
--	--	--	--	--	-----------------------------------	---	--	--	--

						GMDKRCRRCGSDLETLPVLSHCGPHS AARQKRHNNIQDRLVKAASRCPGTISV NQTVVGVRGPDAALRPDIVVRDDVNR RVTIVDVAVPFENRLEAFDGVREAKIAK YTPLARQLTDSGYTVTVEAFVVGALGA WDPRNERVLSLLSISRYAILMRRLMVS DTIRWSRDIYVEHVSGIRQYRE			
	LIN9_S M	GCA_00260 0895.1	Schmi dtea medite rranea	AAACGACATCATG AACGCTTGCCCGC AACAAATCCAGTTA TCCCTGCGGTAAC ATTGTGGAAGTCA TAAGACAAGTACT AAAAGAAGAATT AGAAAAATTAGA AGAAAAAATTGA AAATAATTTATTT ATAAAATTTAAAA ATTTAAATTTAAA TTTAAATGAAGAT AAAAATTTATTTA	TAAAAGGGCAAAAA GATATTTCAAGATGA ATTGTGGACTCATCT AAAAAATGACCACCT CGAGCCCAAATAAGC CTAGCTATCATGGTT GCTGATGGAAACAGT AAGGCACCTGATAGC TAACTCTTCACCGTG GGTGTCTTCAGACAC TCACGGTGACAAGAA AGGACACCACTAGTA AAAACCACTAGTTTT CTGACACCTCTTGCT	MMDSRQLNTPKIRKYQNPKMTNDIM KSYNYAVLSDVTPQETTQQTTHLNVDI DKETTQPKQPLTKSGKPKSKPIAVSYKFK DATFIWDTTTPQTNPPKDCTKLIDKTRPR KTIFKSAFQTYLKKELSNETFVEVKFTL MATHKYRFKDENSRLLAYRIINRYVMET ANEFKETEFDMARFAKFFTIPENWLKH LKPYSTATETSPADRIKVQKLVDLTCRYP FKTQEEQTSVANFLHFFTQRSIIGISRDY KFQKCIPFMARKTNGQETTSTINTTPPT EENRLPMVIITPLEEPKIEHRRPEKRGAS NDTIVLSDEEFLLKRRTLPTRKSKNPTG AGNVPTETECTDEVKFILNNEYQIECKE CGKVVENVRNGMNHRLRQKHDFPNRT	cd0 165 0	793	1057

				ATCCAATAAATAA TCAAGAAAATCAA GAAA	ACAAACTCTGTAAAA ATCAAAAGGATCGAT AGGCCGCGCTTTCAC GGTCTGTATTCGTAC TGAAAATCAAGATCA AGGAAGCTTTTCCCA TTTtagTCAACACCA GGTTTCTGTCCTAGTT GAGCTTCCCTTGGGA CATCTGCGTTACCATT TGACAGATGTACCGC CCCAGTCAAACCCC CACCTGTCCTCAAAA CAGTTCAATTGCATC CAAAGATCGCAATTT TTTCACTAAAAAAA TTAACAAAAGTTAAT TATACTGCTTCATTGA GTAAGTAGAAAAACA ATC	DVMVSCVRCEVPIKGAECVNHIKNHKK DDKEESQAGSLVANTQDIPNESSLSQA AIEVYLRNILKMKENQERNIQYLEPSTA NFLINRNLRAFYNVVKIEKLIGWEQVIW LIHWNKCHWIVYLANCDSTSVILDSD NQMTLQQRCAKAKFDKFLEGTFEKA VLGTLERNVPQQPNNFDCGVYVIQYIS DFLKDPQRMDYHTPDSKRIRKEIGELILE EMKNPASKIKNPNKEIQSLLQKFRLQI NVNDVFHWFAEEYQKSLPKIRTKRDGK LNKLSCSYQIQRLFGLAPKRAVKEIYFQE TSTADLETRVLNEHFKKDESMMKECKIK NGNHYQDWITKAQIDNKEILEALKNST DSAPGEDNIPLRQWIIWNNDGVLFDM FNYIKRTHDIPEMWKNYTTTLLIKPGKS QESNIPANWRPISILPTSIRIFMKVLNKR VLDWANRGELISKWQKAVDKANGCDE HSYVIQALIEKANRSYYKNEQCNLAFLD LADAFGSIPFQVIWHTLKNMGMDEKTI NLLKEIYKDCSTKYKCGKNESEKIKITKGV RQGCPLSMTLFSLCIQYLIQGIAEKKKGA			
--	--	--	--	--	---	--	--	--	--

						TIAGQEVCILAYADDLVIVANTAKDMQ MLLTTIENLAKQADLIFKPAKCGYYRDP RDKKSMMKIYGKEISIVDEKNVYTYLGV RIGDTKKKDLNVRFEVKKKTTAIFKSKL RSDQKLEAYNIFCQSKFVYILQGEDIAKT KIETYDEEIKKMIKEDILKLQDKSPFTDFV IYSPREKGGGITKIIDEQIIQTINRTAKLL NSSHRAIRAIYEELIQVANLRGEKEINSI EEALKWLEGTNKYKKNSNAKTTWITRV REAFQTLKHKHKIKVRFVPKENGIGYKIK CDTQEKVEELDNSKELSKSLHWMIKEA YFKEWKTLCQGYILSLKTSEFMWKM PRGLPDPDWRFCLKVRANMLDINMKQ ANQGGRLGNTKCRKCEDKESASHVINH CASGNWSRVEKHNQVQNELAKELTKR NISFEKDCIPKETKGTLPDLVIRLKDKI MIVDIKCPFDEESAIESARNKNIDKYREL AKEIQAKTVLPTTISTFVVCSLGTWDKR NNELLRQMGIRYEEKEMRINMIHKAI HGSRKTYDQHRNFNG			
--	--	--	--	--	--	--	--	--	--

	R2-1_TG	GCA_012069535.1	Taenio pygia guttata	GTCTAGTTACAAC TGGGCATCGCTGC AGAGATCGCACCT CCTCGTGGTCCCG CTGGTAGCCCTTC GAAGGGTGACTA AGTCGATCTCTGC CCCAGGTACGGA GCCGTTGGGACTC ACCAGTCCAACGT AACTCCTGCCTAA ACTCGGTGAAAC AAATTCCTCGGTA AAAAGCCCC	TAATTCAGGTTATTTA GATGCTTAGTTTATG TACCTTTCTTGATTG TCTAGGATTTTGATA GTGTTATTATTTTAT ATTTTGTACGATTGC ATAATGTTCTTTTTTA TACAGCTCTGTTTTAA TAAAATAGACGATAC CTAGAGACGTTAGG GCAGCCACAAGCCA GTTAGGTAGCGGATA GTAGGTAGGAACAG ACTTTTACTATTTTAT AACGCGTCAATTACC ACCTGATTTGGACCA ATTCACGGGATTTGT CCAAGGTGGACGGG CCACCTTTACTTAACC AGGAAAAGGAACAT	MASCPKPGPPVSAGAMSSSEGLTTHSV LAIERGPNLANSGSDFGGGGLGPVCL LRVSVGTQTSRSDWVDLVSWSHPGPA SKSQRVDLVSLFPKHRVDLLSKNDQVDL VAQFLPSKFPPNLAENDLALLVNLEFYR SDLHVYECVHFAADWEGLSGLPEVYEQ LAPQPCVGETLHSSLPRDSEPFVPEEGS SEKESEDAPKTSPPTPGKHGLEQTGEEK VMVTVDPKNPPCPCCGTRVNSVLNLIE HLKVSHGKRGVCFRCAKCGKENSNYHS VVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIV ASQPKETSNRGAHKRCWTKEEEELLIRL EAQFEGNKNINKLIAEHITTKTAKQISDK TRLLSRKPAEETREEPGVCHRTRRAAAS LRTEPEMSHHAQAEARDNGPGRRLP GRAAAGGRTMDEIRRHDPKGNGQQR PTKQKSEEQLQAYYKKTLEERLSAGALN TFPRAFKQVMEGRDIQLVINQTAQDCF GCLESISQIRTATRDKKDTVTREKRPKKP FQKWMKDRAIKKGNLYLRFQRLFYLD RG	cd0 165 0	690	956
--	---------	-----------------	----------------------------	---	--	--	-----------------	-----	-----

					ATATAATTTATGTGT GTTCGATAAA	KLAKIILDDIECLSCDIPPSEIYSVFKTRWE TTGNFKSLGDFKTYGKADNTAFRELITA KEIEKNVQEMSKGSAPGPDGITLGDVV KMDPEFSRTMEIFNLWLTTGKIPDMVR GCRTVLIPKSSKPDRLKDINNWRPITIGS ILLRLFSRIVTARLSKACPLNPRQRGFIRA AGCSENKLLQTIIWSAKREHRPLGVVF VDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIR VGVKQGDPMSPLLFNLAMDP LLCKLEE SGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTLGLQTQGQKC HGFYIKPTKDSYTINDCAAWTINGTPLN MIDPGESEKYLGLQFDPWIGIARSGLST KLDFWLQRIDQAPLKPLQKTDILKTYTIP RLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSTRDGG LGITKLA GLIPSVQARRLHRIAQSDDTMKCFME KEKMEQLHKKLWIQAGGDRENIPSIWE APPSSEPPNNVSTNSEWEAPTQKDKFP KPCNWRKNEFKKWTKLASQGRGIANF			
--	--	--	--	--	-------------------------------	---	--	--	--

						ERDKISNHWIQYYRSIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCD DIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKP DLIFVKDARALVVDVTVRYEAAKSSLEE AAAEKVRKYKHLETEVRQLTNAKDVTF VGFP LGARGKWHQDNFKLLTELGLSKS RQVKMAETFSTVALFSSVDIVHMFASR ARQSMVM			
	ORTE-1_AAe	GCF_00220 4515.2	Aedes aegypti	ACGGATTAATGGT GTGAATGTGAAC GGGAGGTTTCGA AAGACAAATGCA TGAACGGAGCTA GCGAGTGATAAA TCGGGAGCGGGG AGACGAAGTGAG AGGTTGGTGCGA GAGAGACGAAGT GAAAGGTGGGTG ATAGTAAAGTAG	TAGTAAATAATTGAA GAAGTATCAAAAAA AAAAAAACATACAC AAAAACCTACAAAT TTACACACACGCATA AACATATTGGAAAAT AACACATATGAGTAT GCACAAATACGAATT ATACATATAAAAAGA TACGAAAACAAATAT ACATACACAAAAACA TATGATAACTCACAC	MDLFNFDINDFQTAVVQKIPAGSFITAI MLQINKINSEFKNHQLNEMEFCCLIADF IRCNDHLFSKMWEEDIKEVVHKIFSMES GVEVHMAQAIVELFQTKICFYQKKREV EVIQPNSCSNTNMLTEINILFTFEDGTEK FYSIQSITDKKSSGKKESGYNEQPRNELY KVKEYESGREMMVKKIKGDGNCFIRAI DQMNKSEDIYRCNLSDIILRNKIANH MNKNRERYEMYLIESLEENGFLKFIEDIR KPGVYLGHESMVAVTEIFDKCILLYQTD SETLRIGEEKDREEDVLRLLYTGENGVK NHFDSIRIVKGQTEENTGTEKTSVEEVK	cd0 165 0	1122	1384

				GTGAGATGTGTG	GAAACACTAATACAC	EKESYTSMHGKHNDSEESTIRYRKTQ			
				AGAGGTAAATAT	TCAATCACTCATAAA	DTPITGKQIKETGKLVQEEGKQWQWN			
				ATAGACACTGGA	CACATACAAAATAAT	KERNNQKKSDDKKA VEKNMDERGNET			
				GTGAAGTGACCA	CAATAATGAAAATTA	RKITREKHKPINKDRAITKDFITDNRKEEI			
				TTGGATTATTAGT	CTCTGAAAATACACA	LDLEEKRRRQRSKEREDRLEQLATQIKT			
				AGTAGTGTAGGT	AAACATAGATACTAA	NEERSMLINGSQRQGWEEENDSQRERLD			
				CAATCAAGAAGT	AATAAACACACACAT	REKQAKTDKNTRGRIQMDVQGKTEEP			
				GGTGAAGGTGGT	AAATTACACATAAAC	LSIVNDEQVRKENNTKDVKLGDKKENP			
				AGTGACGATGAC	ACATATACAGTATTG	RISTKQRNDRKRAKNERVDEEGVQENS			
				GGTTAGGATGCC	AACAGAACATTTGCA	DTMDGGTKKHRNHVAEERGETRKGQ			
				ATAGATAATATAT	ACTTTTTCGGTTTTCC	VIESEESRATGVKATGEIGEENEEVKESI			
				AAAGCGACTTGG	ATACGGAATGGCCG	MLIATLNIRGCSKLEKRQEIDNMLESFNI			
				TGGTGTTGAAGA	GTTTTGGTAGGTTTG	DVAALQEVNVKAELITTRNYEWRI GGQ			
				AATGAGAAGGGT	ATCTCGGTTGTTTGT	GQNKSRGLAILIRKDKGIDVMEWKEIGK			
				GTAGAGAAGAGC	GGGCCGAATTGAGT	YGLSISLLSKGRKIILINVHGPNKNGYIFY			
				GGAAAAGAAAGG	GAGATTTTCACGGAG	SSLGKLISQDHLRSNLIMVGDWNAQIG			
				TGAATTGAAACAC	CATTGGATTTTCGGCA	KDSVMPEDAECIGNRLGFSNNNDNGE			
				AGGGGTGTTCAA	CGTGTTTTTGAATGG	EFKMFLTIHKIKNMSTVIGKNTDITWKS			
				CTGAATTGGGCTA	TTTTTCCGATTATGG	GKKESQIDHVLKPREGKIEIRFIKGYWTG			
				CGTGACAGGATT	GTTTGGATGTAGTGG	INTDHKMIVTGIRTREKENKEKGKKKIPI			
				GGAGCGATAAGA	CGGTTTGACTGAAGT	DASVLKYDTIKEKYQEALKKYEIRQKEG			

				GAAAAATAAGTG AAAGGTATTAATT ATTTAAACTAAG GTAAGACGACTG ATGACTGGAATG CATTCTAATACAC TGATGAATTAAG GTAAGACTAGAT ATAAGCAAATATA TATTATGAGTTTT CTTTTCTTTATCA ATTATAGCATAAC ATTTATAACTACA CAAAGAGTACAA A	GAAATTTGAAATTTA ATACCCCTACACAAA AACTGTCTTCAGCAA ACTTGTCATAGCGT CAAAATAGACAACTT TGCTGAAAATATCAT CAAGCTATTCTGTTA ACGTGTTGAGTTATG TCAATAAAAAAAAAA TCGTAATAAGGGTCG CTTTGGTCAAACCAT TAATGCTTTTGAAGTC GTGGTTGGGGGGAT CAGTCGGAGGCATGT GTTGAGGTTCTGGGT GTGCCTGGTGTCTG TGGGGGTTTCATTCG GATCGGTCCGTGGAT AGCTGGGTCTGGCT TGCCGGGGTTGGTCG TTTCGTGTGGGGACC	NTIEQDFKAVAGKMKKAANEVLSARA PLTPIRKAALNKLKTAINLSNKHPDIFPY RWKLKDRRMEFQAAIRSHNERKIKKFY KELNDFDVAVRIKSYQFLKGFLKRKKH KNVYIPMSQWNEALKESEGPEIQLMDE HDTCPVTVPKKEEMMQIVQQQCNG KSAGADGIRMELLKYADKQTQQELADI WKRVWIENHLPEDMEKTIQVPIPKNAK AKGVQDFRRISLCNVAYKPYAKWIKQR LLEFTGEPDINQAAFTTEGRSTDDHMFV TRRILEEYWNAGNPLYVAALDISKAFDN VSLLQLRSILAKLNVP SHLIDRVLQCIRKE KIRVRWQNQYTTECNRGKGVKQGCPL SPFLFNIVMQDVVRKVKEKVPENLMN TGSLKLPLILIFADDILIITKDEEGLKRILIAL EESLREVGLEINNDKSQILIRAPNATKRP PNTINLNGRNYDVKTS MRYLGAWLTD TLNRPATTKQRCVNAAKVSKLVVEFCK KFKPTWEIGKLIYNTVIAPSILYGT KTATL TKRSRVQLGKYEKQILKDIWNQCRKTD ERKFNVRELKKGKTINRRVRVNRISYYG			
--	--	--	--	--	--	--	--	--	--

					GGGGAGGTTGCGAG TGTTTTGTCAAATACT GTAAATACATACAAG ACAAAATAACACAAG CACATGTATATATAT ATGAGAACACATGCA GAAATAGACATTGAC TTATAACACAAATTA AAGCTGTTTTAAAAA GATAATTACTCAATTT TAATAATAAACACAT AGCATTAGTACAAGA ATGACATGCAGTCAT CAGTCGTTTGAGTTG GTAAAAGAAAATAA AATTGTTGATTAA AAAATGCAAGGAATT TGAATATCATATAAT TAACTAGAAATAATA ACTACACTTGTTTACT TTCGAGGAATGATTA	HIQRREETHPIKMAMKLKFNKKKHGRP SFTWKKSLQDFNRYNGITEEEWSQIA QDRDKIKKKAEEIYRNNESEISDGEDSTE QDM			
--	--	--	--	--	--	---	--	--	--

					GAATATCTCGCGTTC GGTATCGTGGCTGTG GTGATCGGTTTCACT TCGTCGATTTTCCTTT TAACTAATTACTTGG CCGGGCGACTGTTTT TGACTGCTATTTAGA TTGAGATTGTCCATT GTCATGCTAAATCGT AAAGTGTTACCAAAA TCTATTGAACTTCGT GTATTAACCCTTACTC GACGAAGAGAGATC GATCACTGCGGGTAC GCCTGTATGATGCTG AACGCAAGTTATGAA TATAGAAAACCTTC TCTTCACTTTTGCATG GGATGTTTTAGATTA GCAGTGTTTGTTTCGT TTTACTGGTGATTTTA			
--	--	--	--	--	--	--	--	--

					GTGATGTCATTGTGT TGTGTGGGTTTTTCTT TTATAGCCAGTAGAT TGTTGCAGAGCGCCC AAAGGGGTTTCAGATT ATGATCATAGTGGAA TTTATTATTTTCTTGT TATAAATAATATAAG AATATAAAGAATAAT TACAATTATCTTCATT GTACAATGCAAATCA TAACTGAACTCCTT AATAATAAATGTCTA GTTATTTTTTATATGT TAACCAATTAGAAGT TTTGAAGTAGTACAA TAAAGTAGATTAGTA AATGTGTATATGCCG AAAAAAAGTCTTAAT GCCAACCAAAGTTAT ACAAGGTTTCAGTCG			
--	--	--	--	--	---	--	--	--

					TAGCATTAAAATTCA GAATAACTTGTCATG CCGACTGAATGTTGT TTATAAAAATCAAAG AAGAATAATCATAAT AATAAAAATAAACAT GGTGGAGGGAAATG ATAGCAACGAGAAG CATCATAATAATTCTA TGAGAACTTATAGC AGGTTATAAATAAAT ATGATGATTATTTCTA CTGTAACTTATTATT CAAATCTTTAAAAAT AGTTATAGAATGAAA CATTGCTATTGCCTG ATATGAAATTAAAAT AATTGAAAAACATT AATAAAATCAATAAT TAGTTTTGAATTCAA AAAAAAAAATAAATA				
--	--	--	--	--	---	--	--	--	--

					AATACAAATATTAAA AATAATAAAGATACA TAATGATTTGGATGA TGTGAGGGAGGATG GGAGGTCGTTGCTAT CATTTGTCTGTATGTA TAAACATAAAATAAA TGTTAAAATAAACAA GTATGTATAATTTTA GTCGCTATTAAGTGA TGGATATACTGGTAA GAATACATATTCGTC TTGTTTGCGAGCAGT GAACAGTAAATTATA CATCTATTGCAGTTTA ACCTGATGAAACTAC AGAGTATAATGACCG TTACATAACCGAATT ATATTAACACTAGTA TGTTAAATTCAAATTT CTAATGAATTAAATT			
--	--	--	--	--	---	--	--	--

					TCATTTCTATTAAATT TATAGAGAACAAAAT AGGAATAATCGATG GATCGATGTTGGAGC TGCAAACAGACTAAA GCTGCATATTGCACA CGGTCCTCTATGGAG TTATATTTCATAATTT CTTCTCTCTATATATA AAACATTTTAGGGGC AGCAGCAGGAAGGT GTGAGCATTTAAATA GTGTTAAAATTGAAG CTCTGAACCGTATTTT TCTTAGGTACGAAAA GCGTTAAAAACAAAA CGTAGGATGTACGG AGTATCATAATGGTT ATATTATATTGTAGTT TATTGTACGAGACCT TTTAGAACAATAAAT			
--	--	--	--	--	---	--	--	--

					CGTCAATTGGGAAAT TCGGATATGAAGTGG TGTTGCACGAAAGTA ACATCAGTAGCAACA GTGTTGTCGTTTCAT ACAAGGGACTAGTGT GTAATGTTATTTTATG TTATATTGTGTAGCG TTGTACTATATCTACT GTTCATAGGTAATAA ATAGGTATACAAATA ATTCCTTTGGCGCAT AAAATAACAGATAAT TCTTATTCCTTATAGC TACTAGTCCAGTGGA GGTGACGAATTACT TATTTATAATTGTTTT GTTATACTAATTCATT GGCAGAAATTCCAGA AGCGTCATGGAGTCA TCGGAGATACTAACA				
--	--	--	--	--	---	--	--	--	--

					TAATCACTGCTATTA ATATTCAGCAGTTCT ATTCCGAAAACGTGA ATTCGTCCTTACTCTG GAGACTACCTAACGA GGTTCAATTCATCAC AATTAAAGAATGGAA ATTCTGAAACATCTA TATTCATATCATATAC AACTATAAATTCCATT CTATATATTTTATTAT AGTTATTTCAATTATT ATATGCTGGGTATTG AGAGGGTGCATTAG GGCATCAGAGAAGT CACAGATGGACGAT GTAGTAGGGTGTCA GCCGAAATCTAGGA GAATGTGCCGTAACA TCCTTACACTACATGC TGCTACTTATAAATG			
--	--	--	--	--	--	--	--	--

					ACTTTAATGTAATTAT CTTTGTTTACTACAGT AGGCTATAGCGAATG AAACGAAGCGTCGG ATGTATTATTATTTG TTTGTACTCCATGAA CATATGGAAGCTGGA ATTTTGTTTTTGATGT ATCCGGAACTTTAT TAACATATCACTGAT TTTTTTATTTTATTCA AATTATTGTTTATTC TTATCATTCTTCTGTC GACCAGCCTAGGAG CATCTGTAGCACAAA TATTCAAAAGAAAAT TATATGTCGGCCGAT ATAATACTGCATGCT GTTTCACTGCGTCAT AATTTATTGTA ACTAC CATTTTATTTTCATAT			
--	--	--	--	--	--	--	--	--

					TTTATTATCATATTAT TATATATATACATT ATAATCTTATCTACTA TATAACATATCAATA TAATTGAATTTTTCAT GATTTGGGAAGGGA GGGAGCATCAGGGC ATCATTGAAATTACA ACTGGACAATGAAGT GGAGTGCCAATCGG AGTCAGGAGGAAAA GTGGTAGTCGTTAC GTCTAGTACTTTTCTC TCCAACAGTTGTGAA AGATTTGACGCGCCC TTCTTCTAACAACCT TTCCGTGGAAAGCTC GACAAGTAATGCAAA TTCTAATACTCAATCT GTTGGATCATATGGC TGGAAGGAGATTCTC			
--	--	--	--	--	--	--	--	--

					TATTTCCCGCGGGTT TCGCGTAAGTGTTTC TACTTACGGGTCCGC CACCCCCCTTTTGGCC CTTCATACAGCGATA TGTTGAATGCAGATT GGTAATAAAATTTGA CCTTACTGTTGAGTG GAGCCGCATGCAGA GCAAGACTCAAGCAA GGATTGGTGGCTACC TACATACATACATAC A				
	CRE- 1B_Fo Ca	GCF_00221 7175.1	Folso mia candid a	TGCATCATTGCCA AA	TAAGTGTATTTCATA CTTTATCAATGTTTAT GTTAATAAATAATT	MPLFNRPPQCRQLCPNPLLSANNATIQ VLDSPVSTHNETSRSTQLDAHRSNFVG GSQIAKVESSDLDRSSPFNGVFYRIETGK IIAQPSSSPLPPSSSPLLPLSSSSS STPLPLSSSSSPPPPSSSSSPLLPLSSSSSP PPPSSSSSPLLPLSSSSSSTPLPLSSSSSP PPSSSSPRPHPGPRKKRNKQKKRTKPRK LPPPTSSQPLPHTTEGLNPVFEDLFCPVS	SSF 566 72	674	871

						GCPGPSSIGTDEWILCNACLIWYHTSCL FLQAIPEGDWFCANCPQIDTQIGPQINS NTSSPANQSYLGTHIQLTQSIPFTNPRA SQPISTLSTPLSNFISSKETCTSLTLEWLN NHIVPGGVIEKAILHDAYRRDSLGFKPI NDAEMGKLIKNSFSNIRTKRIRTELHGTS KYHYVGIQFKSEHSADSVFDFQTIISLR QNCPTLKQCPKACRIPVAQVLKSAVDK CISENTPEAWHLLTFPYITLQCESKANI KNGQSLNSKIRNNLSIFESLDKNTYQDY LITLSSKNSAHDKPGSTIKDIGPKVISKVS EGNIRGAIRLLTSDDIAPNSSETLETLVE KHPPQPLDNFTVLDEPNIPPFTTTVEQV ISAI FSFSGSAGGLDALRPQILKDLVSQ HTGDVGTQLLESMTKFCNFMLAGKVP LDICPFIYGASLTALKKKSGGIRPIAVGCT YRRIVSKICVKVMNHVFQDYFSPVQVG FNVKNGAEAGAHAARRFFSFPKSP FLKIDFKNAFMVDRGKLLSSVHELVP LYPYINQCYRFSSSLVWNSSLIESCRGVQ QGDPLGPPLFCLILHPVVSNLGSEFNM			
--	--	--	--	--	--	---	--	--	--

						WYLLDDGTLGGDPTTVLDDFKHIISCAN EVGLELNFDKCEVSVLGLQTNTQCILESF QSVAPGISCMQTQENAELLGAPLTYESIE HVFRKKINAFSNMAGRLNSIPNHIAYYL LKHSLAIPRLVYLLRAAPIFKSPHLLLEFD SLLHDTLEQITNCQINPLSLEEIALPVKLG GLGIKQSKTLATPCFLASSFATTDLVSLIL PSYVTALSDSLVTEAQDLWKSISCDELP RSTISSIQSSWEIPIFNQVCSRILDSSTPI NTARILASRSKESGAWLQALPSRNLGTL LDDNSFRIAIGLRLGLKLCFPHKVCVCGTQ IDELGLHGLSCKKSAGRHSRHTQCNDII KRALISADIPSIREPLGCCRTDGKRPDGL TLIPYKQGRSLIWDFCTCTDFAPSYLSHT LRHAGTAAKLAETKKHKLYSELEKQYIFV PVAVETSGVWGMGLKFIQDVGRRM KDVTGPMSTTFLIQRISLAVQRGNAASI MGAVPSQRGLEELFYILNDRSFYLQKSC KGPNGDLGRLSGSSKGFPNPPTSGFSST PQAN			
--	--	--	--	--	--	---	--	--	--

	L2- 22_NV	GCF_00020 9225.1	Nemat ostella vecten sis	AAACGGGAGATC TAATCTCATCCTT GGCCTCAGTGAG ACGTTTTTGGACG AATCATGGACGG ATACCGAGTTGGC TATAGATGGTTAC TCGCTGGTTAGAG CTGATCGCACATA CAACTCTGGTGAC GGGCTGCTAGTTT TCATACCCAGCCA TATTTCTTCAAA AGGCGCGTGGAT CTTGAGTTTGAAG GTATTGAAGCCAT TTGGCTTGAGCTA TTATTCCACCGAT CGAGCCCGTGCTT TGCTTGTTTTATT ACCGTCGTCCTTC	TAAAACCGGATATCC ATTCTCATGTTAATTG TTATTGCATGTTAATT AATGTTTGTCTGTTTG CACTTGTTTCTATGTT GACGTCTGCCTGTTT TTGTAGTTAGTTTTGT CGTTAGTTTTTAGTTG CTTTATCAAGATGTA GTATTAGTGATGTCT TGTATAGTTTAATTTT CAAGTTTCCCCTTTGT ATTTCCGTAAATGTCT AACATAGTTTTTTTTTA TTATTATCATAGGTA CTTAGATATAGGGAG GGCCATATATTATAA AACTGTAAAAGGTTA ATATGAAACCCTCGT TAAATAAAGTTTATT GTATTGTATTG	MFSHASSAISQQVMVLGDLNFDLLERP LPAPTKQYLSIFDSLGYQQKITKSTRPISN SLLDHIFVSSSDFVIEADTISNTASDHLPI FVAWKLKAGYTRPTGHKVLRFRRSRKNF CIESFMHDLVLPWSTLDIYDDVNDAL DQWYKLFNDILTVHAPIKEHRIKRPLKL DWFNAEISAAIKARNCLHRIATSLKTD SWSEYRRARNHVVHLIRKAKKTFYNDAI KNNLENPKNLWKKIIRNISPAAKHNLPNR LTVNDTLYEDTFDIANFFNEHFSNISSQV CNDVPEFPDRSVLTNFINFKHQPGSYF SIPPISDFVLNSPSRLDFGKAAGLDELK VFFLKSAAPEIFRSLTKILNLSISSCVFPDP WKSAPVPLFKDGSFLDRTNFRPISVLCI LSKVLERHIHTSFYYFLTSQDLLTDFQFG FRWFRSCELATINLSDSILKNMDSGLPS GMLLIDLKKAALVDHNTLLLKLGLYGC DELTWKWFKSYLSGRTQKTCFKGVLSNA LPINVGVPQGSILGPLFFLLFINDLPLFLS SDLVTDTTMFADDTTFLISGASVPDIELK LNKLATDVHKWATLNRMLVNTKKTKS	cd0 165 0	374	636
--	--------------	---------------------	-----------------------------------	--	--	---	-----------------	-----	-----

				TTCTACCTCTGCA TACATAGCTCTTT TGGACGAA		MLIGSHQKVSNLAESSLSVTLNSCPIEQ VNIARVLGVVFDNHFTCEPHIELFCKYIN SRISLLRSIAPFLTPAGAIHYYNACIHSQF PYCASVWGNCSQTLQLLLLRSQKRAAR TILSADFSVPSVSLFSRLNWLYICDLIKQR KLWTLFSILINPDCPLCLRSKFSFVSHERR TRSSTPNNLVLPQPRTDAGKRTFSFSAIS LFNLLPSLKQLAFSRFLNLNQLPNVTSL AKKKVRDIFHSHLNSVSHLEQLFCHACR FYIDCRCYFS			
	Crack-5_AAe	GCF_00220 4515.2	Aedes aegypti	GGTATTCCAACGG GAATGTCTTCTTA CGTAAGGATGAA AGTTCTAAGGCAA TCCATGTGAGTGA TGTTAGTGTTCTA TCATCGATTGTGT AATATTTCTGAT TAAGTTAA	TGAGATGTCCATATT GAATTACTGAAATGT AATGAGTTAATGTTA GTATTAAGTAGAAAA TAAGAAACATTGTTA GATTATATGCTTGCT CCCTCGAAATATTTTA CTTTTCGAGTGGCTG GACACAGTTAATTAT GTAACTTTTAAAGCT	MITLIKINYKAELINELNINFCLEEIKLDKE NLNILYLNTRSCRNKEEELYQLINEFDGS VHVVFVAETWLYKNEIFNLLGYDAYHG CRDGRGGGVSI FVLSDYRSQMLVINSE DNNFLMVELIDQNVKIMGVYNPGRSF NLFFEQLEQILHNYKRIIVCFDFNINLLDS SGELIRQYTYKIQSLGYIILNKLDARYATR ISNTIATVIAHIITDIFEYDLQFGIKDTERLI SDHRTLTLISMRIENHKKNKPDHKKICVN YENIIRSTSQNTFEICQQFSDIVVVLNEI MTKNRTEILSRKKAMVIRKPYINNALIAE	cd0 165 0	515	769

					GCGTTGTTAAATAAA GCAAAAAAATAAAA	IQNKNELYRQYRTAPGGVLKDELYILYA RKRNMLKNKIKFAKEQYYGRKMHKNIN NYKKTWSTLREIIFQNGSENITPQALTK ENGKIIDSAAQIAKCLNEYFINIGWNIQ QNGTGNDNFLIYLNNMPETRFSSVVE TEKAKNVINSLKSGVAVGLNKISTKCIQK CVTLIAPKLTEIVNNVILSSQFEDCLKIAKI VPLYKSGDKFSKSNYRPVSIIPALSKIPECI MNEQIKHHLLQNELLNKNQFGFLPKKS TETASLELLNFIVKGLDDGEYVGCIFIDLR KAFDCISHELLIAKLSFYGFTPSAVSLLKS YLANRKQQCQVNGVLSEPLLIKSGVPQ GSIIGPILFNIFINDLFNIPLKGGKLQCYAD DAVLKYRAKDLETLSMMQHDLQLLPS WFSANKMQINNKKTNFIIFSLKTSNVSL DLFINNERLSQVTEVDYLGLAIDKNLNW KSHITLVKNKITPFIFALRRVRGCLSTKSC WQLYNAFILPHLSYMNTVWGSAAASH LNVLKVLQNRVIKTIRKLPFLFPSASLYSP EVLPINLLHKYNCLILVYKIKYKLVESNIRL TEFHDLHSHSTRARQNFYVSRPRTLSL			
--	--	--	--	--	-----------------------------------	---	--	--	--

						KHFIYSGITQFNALPNHLKLEIHLSSYKKK LKKYLFENN			
	CR1- 54_AA e	GCF_00220 4515.2	Aedes aegypti	A	TAAGTAAATCAGTCT GTAGGAAATTTATCC AAGACGTTGCAAATA AATAAATAAATAAA	MWLRPDVLNSELCSNYTIFRCRSEQT SSFQRGGGVLIAVNASFRCTSVSLNDCD SLEQVVSVNLPMSTIYLCGIYIRPTSPP EAYSMHATAIQSICNFSSETDAIIVVDY NLPGLSWISDEELNSYLPCNASTEATTF VEAVMATGLYHVNSIRNVNNRILDIAF VSDVCDIEVLEAPSYLLRMDAHHKSFVL RVQIWETRNNSSSDSNDETDFDFSLCD FESLNASLSMINWTNELANKGTDEATA HFYDKLYDVLHNSVPRKRLRPKQKGKLP WWSPELRLRNGVRKARKRYRYSKSH VVRQNLHVIESRYNKCKVFTFRNYINSIE SNLKNNPNSFWSFVKSRSKNRIPEQIR FQNRTSANVDESANLFADFFSSVNSTIS PTLSTATRHSIPTHDITLSFLNFTEQDVK TTLTSLNVKKGAGADRLPPSYWKECAE SMKLPISLIFNRSLADRKFPVWKTASV VPIHKSGSINSVDNYRGISILCCIAKVFES MMHMLVLYNATRHLISDSQHGFVKKLS	cd0 165 0	470	721

						TVSNLMCYTNFLSSEIERRQQVDAVYF DFSKAFDKVPHDLAIKKLEQMGLPDWI TEWLRSYLSGRKAFVRIGDTQSRGFDIT SGVPQGSVLGSLIFVLFINDLSHRLKSKC LFYADDLKIYRAITSVIDCYALPSDINELL WCSENGMQLNNSKCKCITFTRSLSSIYF NYEINGSIIDRVTTINDLGVTIDSKLKFSK HVRVTTAKAFAVLGFIRRNQAFQDIYT LKTLYCSLVRTILEYAACVWSPYCTTQVL HVEKVQRSFLRYALRSLPWNDPVNLPD YESRCMLLNMETLSFRRTKLKQMFLFD LITHNIDCSDLLAEVGFLAPSRYLRSRQL LAPRTHRTAYGQNSPLTSCIRSFIVNDV FDFNLSRVEFKRRISSIR			
	CR1- 1_CBr e	GCA_00014 3925.2	Caeno rhabdi tis brenn eri	TTTCGAAGACAGA CGAGATGCGATC AAGACCCTCACTA CATTCAACTCAAT CAAAAGCCGCGT GACCGGTTGTCAC AACGCCACCATTC	TAACATGTA ACTATTT GTTTGATGTATAAGA GGCCTATTGGTCTTT CACTCAATAAACTCA TCATCATCA	MNCFSIKNKLPQLDLVSSTNNFDVICLS ETKLDDSFPSLLSLNNKFSVVRKDRNK HGGGVAILITKKVKFCPIEIPKHFSSPEIS GVDILLRGTTVRVIVAYHPSHTSSLERLV QSLEFLLCSQKHCVVLGDFNMMPHINWE TLTACDSPCKTFLAFVTNNGLTQHVLSP TRLNPDNILDCLSN TKIVRDVQVGELFS	cd0 165 0	460	710

				GTCAGGACATGTC CAAACCGGAGCT GTCAAAATTTTCGG GACTCGTGGAAG CAGGCCATCAAG AACAACAACGAA GCTGGACAAGGG ATCTGGACTGTGA GAAACCTCGAGCT GGTCAGAATTCCC TACAAGGACAAT GAAACTCCTCACC CTTGGACAGTGC GAAACAGAACCA AGTAAAGTTTTTC TCCCTTGCCCAT		DHKIISVTLKLSNSAKKVVKRVQLFRKAN FQLINYNLSMINWDVEMSCMSLEEKYE FFVSVLQHQIALHVPCKNINVLAKKHGP KILSLQKEKLRIWRAAGNSDNYKKISLILK EELIQEDQMNTEKSLTYGTSKKFFKFINS RYKSNDEIGVIKDSQNIPVNSNILKAELL SDAFSKVFTVDNSSSTLFPSRTNNVLES MTFEPSVIEKVLLKLVPKCNTTPDEIPAIF LKKVGTSIAIPLSIIFSESFRTSQVPRLWK TAIVKPLHKKGPRTDPNNYRPISLTSSVC KVMEKLVRKNLTDYLNSSQQLFSRAQFG FRSNMSTESQLTSYFSTLISNHSRLRKPTY SVYVDFKKAFDTSFEKLTCLMSYGIK GQLLDWIVCFLTNRQTQRVNVNGVLSQE KFVLSGVPQGSVLGPLLFLLFINDIGDKF QSHFLLYADDLKLFTNPSTIQTDLGLLS EWCDKWQMSVAPNKCENIIFNLTKSP RTLNFDFTIKSKIPRTKKIRDGLILSSDL PFSEHIDTILRRSHQRINIFFNILRRASFET FIKCYTIYVRPILEYGSI VFN PVLKEQIKRI ESVQKSFVYRVFKKFHIPYLN YFDAL SVC			
--	--	--	--	--	--	--	--	--	--

						NLTSLELRRRLHADLVFLYKVIVSKETLLYP PFFTLPSPNIRKHKFYFRSNSKNSNKFF SQFISNRSLNCWNDLSPSVFPPTPHSIN FISNISYCNFQKYLTLNPHNF			
	LINE2 H_CE	GCA_00452 6295.1	Caeno rhabdi tis elegan s	AGCTGGTTAATTG CATGTAGGCTAAT GCTCGTTCCATTG CATCGCTTGAAAA ACTTGAATTTGTC CATAAGTTCGTAG CCTCTAATAGTTT AGACATCTTGTTT GTTACCGAAACAT TTCTAAGTAATGC TTTTACTGATTTCT CATGCTCTTCACC TCACTATACCTGC ATAAGATCTGATC GGCTTGAGTGCC ATCCAAAAAATGC GGCGGAGGTACT	TGATCTCATTTCTCA CTCATTTGTACATGAT GTCATATTTTTTTATT TATTTTTTTATTATTTT CTCATCAACTTTTCTT GGAAATCATTTGTAT CATCTGACCTCGGAC ATGTTCCAATCGTGT CACCTTATCATTCACA TTTCTTACCTCATCTC ATTTTTATTCTCATT TTGCTCTTAATTTTCA AGGCAAGAAATAGT GAATAAATAAATTAA	MKFHDLAQMVFNPTRFSKLNSGTFLDL IFTNNSGVQLNLKPGTPFINSDHISIFFSL SLPQPTPAHSAPPKSFMNFRKCDFFNL NRYMVSYNWPQQLSFFAPPDSKLNHFL NIFNELISDFTPQYPTLRVQSPKVTGRKL YRKLKHHPDTESSVLKKRVKSRLKFLKK KNMAFENKLIHNSKNSKSLFSYIRNRIKPS SSLTHFSVDNNNVDPKPIEMAEVLVKEFS KSVVTHDPPFPPTFPSCESTPSPVDTSPLT IASTILKLPKIGYSLDNINFFILKKCCDSIS IPLSLIFKEFITTSFPSCWKTSIVIPVHKK GNSSDPKNYRPIALTHPLSRLFERIVLNP TRKTFSHKISKFQFGFLNNRSCITAHLNA SSNYHKILSVPNQFVDIILFDFAKAFDSV PHELLLCKLRNFGFDDNICLWFKAFISD RVSKVKIGDSLNSSSFKNSSGVLQGTVT GPFLFLVYINDLLDFPPDVHVIAFADDL	cd0 165 0	309	557

				GCTGTGTTTTTC ATAAAAAATGGG CTCTTACGCCTGT CTCAATTGCACCT TCCTTTTATCTCTC GCATTTTTGTCAA ATTCTGGCATTAG ACCATAAGCAATC CAAACTAGATTT ATACTGATATACC GCCCCCATCCACT ACCGCAAGTAGA ACCGCTGAGCTCT TTAAAAATCTAAA CCAGCTTCTCGAC TGCCCCCTCCTATA CTTATGTCCTCCTC GGTGATTTC AATT TTGCAAACACCAA GTGGAATTCAAAT TTTCTTCCCACGG		KLFSNNCQSLRSSIKIENWCKIWQLKLA ENKTKVLHVGGKNPKYKYVNGQNIEC CSKACDLGIWVDDQLAYEKHILVKVNA AMFKCRQILKNFRSLNMMFYFKLFNTYI RPILEYGCEIYHPKSASLTKKLEQPLRYFS RHVFKKCKTQYDSYEHRLNQSNQQSLK HRRVFLILKTFHNIITGKYHFPNLALYLKN SKSPRFPYQMTAIGKTNNNSFLHHTLTI WNLVTKHFPVPVSGYTFSKRISCLPLESY V			
--	--	--	--	---	--	--	--	--	--

				GCTCTGAGTATTT CTATAATTGG					
	RTE- 1_CRe m	GCA_01018 3535.1	Caeno rhabdi tis reman ei	CTTCTTCTGTACT GAGCCTCTTCGCG GCGAAGAAGATC AACAGATGTCACA CTTCTACAAACCC GGCGAGTTCGTC ACGCCGGGAGTG AAACATCTGGCCA AGAAGTACAACG TCTCACGCAGGTC AGAGGTAGCGTC GAATACTCTCTGT GACCGGGCGGCT TTAGCATTGGTTT AGCAGACCATAG TAAAGTAATCAGA GACACCCAAACC GCTCAAAACAGC ACCAAAACCTGCC	TGAGCGAAGAACGG ATCGACTAAGTAACT AAGTAAGTAAGTAA	MDVYHSKRTKLTRPKTRKTDLRICTFNC RSVATDNRLAELLEETNRIKYDIIGLSETK RSAEVHMTNRDGTGVILGKRNDSSVSG GVGFIINKSLMPKIKEIKIVNHRIGYLTQ VNKNQYTIIQGYAPTADYEEEDHSNFYE NLEDVYKSCSRYKMMVIGDFNARIGERK GNEVFIGPHSMENRNDSEGERLATFCEV NRVFMNSQFIKPTHRRWTYISPDQK HRHELDHILANGKYITDVSVPSTNGS DHRMLRANVHINAKQAKFEQVKRRKP PRRVLDPATAALLATENLDSCEDQDIDKE YDTLVHVLKKAQDAAVTIPTNHSRNL KDSTRLLLSKRRFTDRSDPNFKTLSKECR QAVKRDHENFAKDRLLNAANQRKSLR KWARDINEYQSYIPCLLSGETGEKLSRE KMEIEVRRFYSNLFATKQSKATTIIPQQT KALPPFLPEEIQHSLNSFQNGKAAGEDK LSADFLKSCHSSVHKLLANKFSRYLQEG KVPTKWKSSKTTLIFKKGDKENLENYRPI	cd0 165 0	511	779

				TCAGTCCTCCGCA CTCGGCCTCTGCA ACTGTTACGGTAG ATTCAGTAGCACG TCGAGAGAGGCG ATCTAATCCCGGA GTT		CLLPVLYKVFTKCILNRIRKSLEEAQPVE QAGFRRSFSTIDHIHSVQRLLEVGREYQI PITMVFIDFKKAFDTIEPTALWESLKTQ GIDSGYIKLLKECYNDCSTTITPFYNPVNI PITRGVRQGDPISPNLFSALESFRRM SWPHLKEDKDDYDNSPGIRINGRNLTH LRFADDIILISKTPQIAEKMLQELVNSCE TVGLEINASKTKVLRNKFASSHQIHIRKN NTMSAIEDVEEYVYLGRLLNTKNDLEPE IHRRRRSAAALNNIKNTTDALTCPKIR AQLFDSIVLPALTYGSETWTFTKALSERV RITHAALERKLVGITLTEQREKNLHREDI RKMSQVKDPLVFITKRKLSWAGHVMR RNDNRWTTLLQEWIPRNEKRPVGRPP MRWADSLKKEISVRQGTQLIEPWSTIA KDRKKWIAVIRAHTN			
	NeSL- 2_CRe m	GCA_01018 3535.1	Caeno rhabdi tis reman ei	CCAACTCTCATCG TATTAACCTACGG TATTCCTCCTAGT GAGTGTAATAAA GGTTAATTACGTT	TGAATACCGTCAGAT AAGCCCCCAACATAA AAATAAAAGTCGGC GTTAGCTAATCACTA AACCGGCTCCTCATT	MTNVYLKPVNDNQTNKTGDNSRNTM SNSQCEMTWKPVARTYAQAASNPAD DKTVTVLGCKYNLLKLGNTPTTKRSPP KPSRGGARISSVYTLTDELEITHREEGKIT FAIDLPNKNNILCPLCRECTQTRGRGSSF	cd0 165 0	895	1156

				TTCTCTTACCAA GAAAAAGAAATT TCGAATCCTTTTT GTGTAACACAA ACTGACAGAGAC CTATCGAATTTCT TTTGTTCGTATA TAGGAATAGTCAC TCTGGACCACGAA GTGGACAGTTGTT GGCGGACTTCCA AAGTTGAGAGAA AAGGTGTGAAGA GAGGAGGTCTAG AAACACTTCGGCT GTCTAGGACCAGT TCCTGAGTGGA AGAGGAAGGTCT AGAAACACTTCG GCTGTCTAGGACC AGTTCGTGAGATC	GGGGGAGAGTATCA TTCCGGTGCTCTCCG TTTGGGCGGTAGGG AGGAGTTGGGTAGC GACCCGGAAGTATG GATGCCCAACCACCG CAATCTGATCTGGCA ATGTGTTTCGGATGG TCTCTGTCTCTAGATC TGAAATAGAGCTCTG GCCTGAAGAACACAC GCGCGGACCGGTTG GATGCCGACTTGATC TGGAGGGTGCAAAC CTGAAAGGGAAAGT TGAAGGCCGTGAGG CTCCTGGCGGGAAAC TCCGTCATAGTCAGT GGTGTGCCAACACCC GACGACTATGACATA GTTGGAGGAATCCTG	TKHMKLHVKEKHQLKATFIYKCSMCNE YEPEKKCGTKWQTHLQKVHNYKYDES AIVVPVPPNTRQQIANELNNAAPFVDIR KPKANVVEEKKTENGALLKFLTKSNKDE QEHSQSNDSNPVESPEKANQALTIDPK GNNSPSKSSMRSSQSSASSICQEIQEIIT LSEDEEPAARPKPGINVWSLINEKGKD AYIDTDIMMAFLKMRVENYDSVNIIDPL NYQFPARVDLVPLIQRNLEDGKKRVVF PICADEHWTLTISNGIAAFYDPTGSRM SSYIEELVNELGLIIPKEQDEQPRQRDSY NCGVFVMKMAEAFIQDTEWEMEEVE EDVKNFRRNLLEELKPNYEIFAEEKIYYN SPGKSFAQSRPTSRSSQCAVCPTCSRSA TPMMDVGNMEVDPVPQQQETPKSRE PEQDEGWKVVGNTKRGRVVTSPNI SPEAKRQFTGPEIKVSPGKFHPLVGET EEMEVTCDSPPTKEPTTEPKVTPSLPAM RIVSPEVTKKQTSKKKGKYGKKKQQTKK AQPPKGEPTKKAQPKGEPALIEQVRT WFDKQMKSYQEQQSNIQTLTWIADSL			
--	--	--	--	---	--	---	--	--	--

				TCTTGTGAAGAGT TGAAAACAGTTA GCTGAGGCTACT GTATTTCTTGATA GCCCCGCCCCCAA TCCCCCTCCCCC CCCCTCGACAGAA TTTTCTGTTTGAC CTCCTGGAATTTG CGAGGAGTGCGC GAGAATTTTCGAA TTCTTCGCGCGTT TTCTCGAAATTTT CCAGAAGATTCTG AGCGGAGAATCT TCGAGAAAGTGA GCCGAATTTTCGCG CGAATTTTTCGCG ATTTTCAAATTAT CAATTTTGTCTGG AAAATTTATTTTCT	ATCTGATAATAATCA TTGTTTCATATAAGGG AGGGGGGATGGTAAA TACCCAGGGTCCGAA ACCATCAAAGCAGCT ACTGACCAGCATAGT AGTGATGAACACATA GACCCTGGGGTTCCC TGAACCTGACCCATC TGCACAAACCCACTT TGTACAAATGAACCA AACTGATGAAGAGTT TAATGATTTCTTACAT CACAGCTAGCGGACC TACCGTGAGGTAGAC TCCCGCCGCTGTAGC AGGCTCGCCATTG	TAAIFKANNGNKYLVDKITARCPPPLLN EGEMATQTSRRTEAVKPKDRFVKESNE PLRIQYAKNRAKTFNVIIGKHSARCEIDI NVVENHFRQTLKAQPVTEKALNTVCSG IKKAEVDPSIEGPISSGEVKAILAKIKDTS PGTDGVKYSCLKWFDPEGERLALLFDE CRQHGKIPSHWKEAETVLLPKDCTEEER KKPENWRPISLMATVYKLYSSVWNRRI SSVKGVISDCQRGFQAIDGCNESIGILR MCIDTATVLNRNLSCSWLDLTNAFGSV PHELIRRSIAAFGYPESVINIISDMYNGS SMRVKTAEQKTQNMIEAGVKQGDPIIS PTLFNICLEGIIRRHQTRKTGYNCVGNK VRCLAFADDLAILTNNKDEMVDVLNQL DKDCRSVALIFKPKKASLTIKKGSVDQY ARIKIHGTPIRTMSDGDYKYLGVQGTGN GGRASESESLTQIAAELQMVHDTDLAP NQKLDVLKAFILPRLQHMYRNATPKLTE LKEFENTVMKSVKMYHNIPIKGSPLYV QIPVKNGGLGVMSPRFTCLITFLASTLFK LWSDDEYISSIHKKALSRLITAKVMGLKT			
--	--	--	--	---	---	---	--	--	--

				GGCAAAATTTGAT TTAGTTCACGCGG GAGAGAAGGATT TGGTGGAAAAGG GTATTGATTTTG TGGCGGAGGAAA CTCCCACTGAATC AATAACTCTCAA GGAGAACTCATC GAACAACCTCGG GTGACCTGAATCT TGGGCGAAATTT CGCATTGACACAA GATAACACAAATT ACTGTTGAAAACA AATCAGAACAAA CTGTCAAAAAGA AAGACAAAAGT ATTGATTAACAAC ATC		QTATLQEQCDYLNTKKAITKGGYNLFSR MNEASRTLSVNLGAPLKSMQFIPENGE IALEVQASENSQIKVFSKADSMKLVTKL KDLVKSAMLKNFLENKKVKSQVVQVLQ HHPQSNKFVNDGKNISISSQKFVHPARL RLLVCNGNSYGDHPKDCRRCGYECES QAHILQHCTYNFSTGITQRHDRVLRIL HEVIKGRKNNDYYDIMVDTEPGPTRER PDIIMIQLKDGPEVLLADVTPYENGVVA IEAAWEWKIEKYSHFIEYFARLGKRAVIL PLVVGSLGTYPDTSNSLRMLGLSDGQ IRNLIPDISMIALESSKQIYWRHIFGDSYR IVTELYCTKNKQEARFGDEPMENVQVS DRFQPFKTREREKKSEEEKRRRSKSKKG KTWRGSKKQTDSRQSGKSYQNQGFQR SVGQGGSR			
--	--	--	--	--	--	--	--	--	--

	CR1- 2_CRe m	GCA_01018 3535.1	Caeno rhabdi tis reman ei		MSAQSLALIDHLDEYLPLKSCILAETLITH LLLESSIFNHPFSLSSSLKSLPERPFPTRIL PHLQEPPLLNSDHFVSDFSLITTVSNHSS ASLPTSPTFRNEHRLNYKKCDFKSLNADI ANVDWDLVLSSISSPTDKFALFSKMLSD TGATTTTTTTTTTTCTT TCTTTATTCTCCCTCC CCCTTGCCCAACACC TATTTTATAATTCCTT ATTCCCGGTTTCCTTT CTCAGGTTTAGTCTC GTGAGGGACTGTACC TGACCTCCAGTGACA TTTTCTGTATATTTAA ATAGTTAAATAAAAT ATCAA	MSAQSLALIDHLDEYLPLKSCILAETLITH LLLESSIFNHPFSLSSSLKSLPERPFPTRIL PHLQEPPLLNSDHFVSDFSLITTVSNHSS ASLPTSPTFRNEHRLNYKKCDFKSLNADI ANVDWDLVLSSISSPTDKFALFSKMLSD FIISHTPLSTPPKPHRTSFSIRQLKRARQR YSKLLKLPNPPADQIFRLRTLIASSTKRIK SNMVSEERKILSAPNSRSARLLIKKTRRT SLKIPPLYVNNELISSNAAKSSIFSSVFFS NYNCSPSSSLIPVSNNSSNNISPSELFPLP WIIENTLRKVPPRCGFTTHHANFYIIRNC ATTLAIPLSIIYSDSFLKSHVPASWKHAI PIPKKGS�TSPENYRPISLTDPFSSRVCI LCNYIKINLFRNISRFQHGFLSKRSCSSSL VHSISEYRLLSSHKSLDVVYFDFRKA QVDHQFLISKLDSFGIPSNIISWLTDFLS DRTFSVKIDDFVGTPSASIPSGVPQGSV SGPLLFLVFINDLLLKLADISSLCIAAFAD DIKLYSHDPLALQMGIDLVESWASDNS LPLAHSKTALLRLGPRNSSHPYTIGGSLI DSVDSVRDLGLLIEPNLKFTRHINRAVAL	cd0 165 0	350	599
--	--------------------	---------------------	---------------------------------------	--	---	--	-----------------	-----	-----

						ALLRSKQLLKSFKSNSPQFYIFLTKTYVLP LVEYCSVVYSPPPSSKLAHKLETPLRFFS RKILQRCNTPYCSYSDRLSQLDLFSMRH RRLKAQLLLLYNLIIGASYFPSLETHVRLS SSSRRPMSLICINPNCNFFSTVIPIWNA LTINVPHFCLCPSEFNSLLVNNIARF			
	Penelo pe- 6_CoFI	GCF_00064 8655.2	Copido soma florida num	ACTCCTGTAGCCA ACAACAACAATG ATTCAAAGTGGCT GTTTAATCTTTCTA ATGTTACTATACC AGAAGAAGTCTA TAATATTTTGAAA	TAGAGCAACAGATT ACTGTAATTTAAATT GCGAGTTCGTTGACT TGTACTGGGGTGGG GTGGCTGTAGCGTTG TGTGACTTAAATATA CTCATGTACAGTGTC ACACTTACACTGACC GATAGTATATAAGCA CGATCGTAACTTAAG AAGTGCGTATTTTTC GATCTACCATCTATC GTGATGCCTATAGTC GCTATTGTCGTGGCT ATGTGAGTGTTTCATT	MIRETSYFTTNMNTSHKFEFFKLLETSL HFVPTVTDKIRTNLLKMFDKFLKSTIAQ KFDNQKIINDINKTALFLKHNKELLVTKV DKGNPVVLDREHINSVEEQLNNDKY YTELLKNLLNELIKSVTTFFVKRWYDLGIF KDCEFDIRSSEISKNDLSKCYGLVKIHMA GNPIRIIVSAVNSPTYNLSKFFTKFLNDNI PKPNSCIKNSFHIKEKLNSFIIPQHHTIISL DVVSLFTNTSESIILKSIKNNWKYLRNKI NLSSQEFCEGVSLILNSTVFQFNNKFYK QIFGSAMRSPVSPVLCDLALRCLEEDCL KKLSFFVPFYVRYVDDILICVRNNSIETI TVFNNHDRHLQLTVETPVDNEISFLDLR LIFESDGSII TNWYRKPTHSGRYLNYNSH HPLSQKIAIVYTLVDYSIKLSDSKFHKENL	SSF 566 72	203	380

					ATTGTGCGATTTTCTT GTAAAGTATTGATGA CTTGAAATATAATTG TTTCATATTTTATAT TTTTATAATTGTTATT TAGAAAGTTGTACCT GATGAGGGCCTTGTA ACGTGGTTTACTGTA CGGGCCGGAACGTT GTTTACAGCAATAAA CGATTTTGATGGCAT CACGAAGAGAGTCAT TATTTCTTGATAAAA ATTATCTGCAATCAC GAGAGCCACTGACG AACTA	DISKKILFNNQYPLNFINKYIKKRLTILSIN TPLNRVVSNNRGLMILPYFQGFDSEIKR YLYKNNINVIFKCIDKFSKFIRLGKDRTKH TDRCNVYRIVCNDSCYIGQSSRLK TRINEHMNDCKHRREKSALVDHIAKTN HSFNFDEVKILDTESNRFKRLFSEMLCID SYRNNINRMHDTQNLKLCYKNVITEHT			
	RTEX-15_CGi	GCA_902806645.1	Crassostrea gigas	T	TGAGTTTCACTTTGTA TAACTTTTCCTAACTC TATAACCTTTATATTA ATATATGTTTACCTCT GATCAATTTATATGT	MAVTCRNCSTVDYVLASANVFEILEDL DIHDFCNLFSDVHNPVSFSLSTCYLSRNI STGKNHSGVKLWCPDKADQFRSNIDV HAVKSIENYLNEISNSPGDLCSDNINKV VKDICTVFTDAARDTFGPSHFSPCIEKH	cd01650	324	600

					ATATTTTAATGTTATT TTTGTCCTCATGTACC ATATGTTTGAAATGG TTTTGAGCGAATAAA TGAACCTGAACT	NKPWFNRECKSARKQYHLAKSMHNKS KTDDNKNKLKEKSKIYKKAMDNAVKTY NTGITRKLMLQRSTNTKDFWNILNSSK RDSKSPVNIKDMFDFFKNINEGSLSPDS DTLHEPQVDGSSNELLNCEINESEITDA VKKLKNGKAPGYDNIVNEHIASTLNTFL PIYKLFFNAVFDGLVPEEWLVGIIKPIYK NKGDPSPENYRPITLLSCLGKLFTSILSK RLELYSEEELIHNNQAGFRKGFSTLDHIL TLQFLSNTLMKCKKKLFCAFIDFKQAFD TVWRNGLWCKLLNYGINGKCFKYITN MYKGIKSLIKMNGTVSDFFSCNVGVRQ GENLSPFLFSLYINDLENYLVEKNIAGLQ SFTTSIENELCVYIIMCIYIKLLVLLYADDT VILAESGDDLQNALNEFECYCKEWKLN VNIDKTKVLIFSKGRMCKRVFYFNDRL EIVKEFKYLGIIFSRAGSFLKAKKHLCDQA QKAMYGVIRKIRQFNLPVTCQLELFDK MVVPVLLYGCEVWGFEKLDSEIRVHLK FLKHILCLKSSTPNYMVYGETGRFPLSVI VFTRMVSYWCKLISSTNTSITNILYRYLL			
--	--	--	--	--	--	--	--	--	--

						NQTENTTSTNPWIDCIRHVFNSCGLSN VWHSQFLNIVNQKWVSTAVKQRLQD QFIQKWHADINNSSKGVYKIFKTNFGY EKYLDILPLKFRKILGKFRTSNHHLPVELG RWEGIPLNERFCPLCNTKRIAEFHYILE CDAISQVRNKFIDKKFSARPNVLKFSYL MTTCNMKLLRKLCIFITKIYEVVCP			
	CR1- 24_CQ	GCF_00020 9185.1	Culex quinqu efascia tus		TAAATCATGTTTTGTT TAAATTTATGTTTTGA TAACTTGACTATATG TATGTTTGTAAGTT CTTATTAGTTTTAAGC AACCAAACCTTCATTG AGACCACCGGTCA GATGATTTCAAATAA ACAAATC	MEDPRLFVTVVPSHSRNYAFCTSHPSR PGPVFGPRRGVFQNRSVGKYPYAENDL VSLPDARLTSSNVSTNESIRQDSAYRSQ DAASQQSNDNADPGRPPTCYMEDLRL SVTVVPPHSRNYAFCTSHPSRPGPVFG PRRGVFQNRSVGKYPYAENDLVSLPDA RLTSSNVSTNESIRQDSAYRSQDAASQ QSNDNADPGRPPTCYMEDLRLSVTVV PPHSRKYAFCSHPSRPGPVFGPRRGVF QNHVLVGKYPYAENDLVSLPDAHLTSSN VSTNESLRPDSAYRSQDAASQQSTDNA DPGRPPIRIMEDPDPPDTVPPHCCDS SRSGHQSRPGPVFGSREGVFQTPNAGK YQAAFQVGYALDPDPRLLSSDVQTPPV	cd0 165 0	918	1167

						HSDDPTTPVHHDSRRSSAGQPVIFEP FYYQNVRLNTKVDDFFLSCDCDHDV ILLTETWLDGRVTSACLFGNNFSVYHTY RNPANSKRKVGGGVLIAVRNSLDSVEC VQAVAENLEQMFVTIRTPSKRIFVGCW YLPPEKRFETELMEKHLSCIENIRRIAGP HDVIVVAGDYNQGSLVWKHLSGNRSY VDPQLTRTRTEKERLTHGVMLDGMAN CNMHQCNLVKNDRNRILDLIFVSDDDC TVVAAADEPLVPLDAPHPALVFDIDASR TVFVQVDNNDRLVNYKRTNFDELQRTL ADIDWTPVTSAMDVDTAVERFSNLLNF HLAQHTPQFRTPPKPEWSNARLKRLKK ERAAALKTFTLRRCRMTKLAFNAASSRY RRYNRYRYNVYIRKTQSQLRRNPKKFW TFVKSKYKERGLPATMALGDSTATND EKCELFARHFSSVFREPEIISSTHHLDV PRDLVDINTFVISRDMIEKATKKLKSSFS PGPDGIPSAVLKRCANQLIEPLLIVFNLS LRQAIFPTAWKKSFMFIFKKGSKRSIE NYRGITSLCSGSKLLEIIVSDVFLFNCRSY			
--	--	--	--	--	--	--	--	--	--

						VAPEQYGFMPKRSVNTNLEFTSVCIDK IAEGRQVYAIYTDLKAAFDRIDHDIALQK FSKLGFSFMLCRWLESYLKGRIIQVKIGA SVSLELTNGSGVAQGSNLGSIVFSGFFN DANFLFSGNCKLSYADDFKLFIPIDTLAD CYVLQSRDKFSEWCGHNRLELSSAKCY VISFNRKQSRELFTFDYHLGSHVLERLEV VKDLGVLLDLKLDLHQLSSVIDKANRKL GFMFKAAQEFDDPLCLRTLYCALVRSHL ETSAVVWAPYHQNWIDRIERVQRKFV WFALRRLPWNNPAQLPSYETRCNLLGI ETLQNRRTISRAVFAAKVITSEIDSPNLLF QLNARVQPRNLRPTPGFLSRPLVRTVY AENSPIRALATAFNEFYHLFDHHEPVSRF LERLRTELDRSRSSRTAENPQQRASRT RRSSR			
	Crack-10_CQ	GCF_00020 9185.1	Culex quinqu efascia tus	C	TAATTATTTTGTAGT TAATTCTCCGATTCAT TCCCTCGTTTGTATTT GTATCCATCTTGAAC ACATTATCTTGGTA	MLDNRDDVLGLFLDLSKAFDTVDRKILI DKLRYAGVRGIALDLFTSYLSNRAQFVCI DGIHSLTLVNVGVPQGSVLGPLFFIYL NDFSLLPLKGDRLRFADDSSLFYNNKST NMNDLNLRDDLTIVIEYFRLNKLTLNIN	IPR 000 477	1	182

					GTTTTCTTTTATTTAT TTTCAGCCGCTAATC GCCAGAAACCGAACT CCTTAAAAGGTCTAC GCCGATGGAGTTCTG GAGCCAGCCGCATCA ACTTTGTCGTTACATT TTTATTTAATTTCAA TGACATGTAATCTAC TTAATTTGTATTGCAC ATTGAAAAGAAACAG TTTTTTCAGTCAGCAC TTTGCTGGCTTTTCT GTCAAATAAAAACAA TCAATCAATCAATCA ATCAA	KTNIINIKNSSRSIPNSLSLTTSKFPDLEV SDTKYLGIIIDNRLNWSAHINALILKLHKI TGIIFKIKHKLPQKVLFLIYHSLFHSILSYV TAVWGNACGLLINKLQVAQNRILKIILN LPIRSHTVDLVVKNNILPVRGIYVFQVCS FIYLCNNKTHSNTKFVRSNHQYGTRYH DLLNRPNVTTVPGERSINFNGAQLYNY FNNRFGKSSSLNIFKNQLKQCLSRPDIIE KLLKSVNIFA			
	Crack-10_CQ	GCF_00020 9185.1	Culex quinqu efascia tus	C	TAATTATTTTGTAGT TAATTCTCCGATTCAT TCCCTCGTTTGTATTT GTATCCATCTTGAAC ACATTTATCTTGGTA	MLDNRDDVLGLFLDLSKAFDTVDRKILI DKLRYAGVRGIALDLFTSYLSNRAQFVCI DGIHSLTLVNVGVPQGSVLGPLFFIYL NDFSLLPLKGDRLRFADDSSLFYNNKST NMNDLNLRDDLTIVIEYFRLNKLTLNIN	IPR 000 477	7	181

					GTTTTCTTTTATTTAT TTTCAGCCGCTAATC GCCAGAAACCGAACT CCTTAAAAGGTCTAC GCCGATGGAGTTCTG GAGCCAGCCGCATCA ACTTTGTCGTTACATT TTTATTTAATTTCAA TGACATGTAATCTAC TTAATTTGTATTGCAC ATTGAAAAGAAACAG TTTTTTCAGTCAGCAC TTTGCTGGCTTTTCT GTCAAATAAAAACAA TCAATCAATCAATCA ATCAA	KTNIINIKNSSRSIPNSLSLTTSKFPDLEV SDTKYLGII LDNRLNWSAHINALILKLHKI TGIIFKIKHKLPQKVLFLIYHSLFHSILSYV TAVWGNACGLLINKLQVAQNRILKIILN LPIRSH TVDLYVKNNILPV RGIYVFQVCS FIYLC LNNKTHSNTKFVRSNHQYGTRYH DLLNRPNVTTVP GERSINFNGAQLYNY FNNRFGKSSSLNIFKNQLKQCLSRPDIIE KLLKSVNIFA			
	L2B-6_CQ	GCF_00020 9185.1	Culex quinqu efascia tus	AATTGG	TGAAGTGGCATCCCA TAGGTGTGGCTGTGA AGGAGAGCATGTCAT GACGGTCGGCTTCTC TACA	MLAIEIHHQQFNGIFGGVYHSPSSDR DFLDSFENWLQQVMADDKTNVISGDF NICWNNEGYSRDLKNIADAVGLEQKVL EHTRISVRSRTTIDL VFANVEDCTVKVLA DWRITDHETLGINVEGVCKINMPEKKT	cd0 165 0	397	652

						EYTCWRKYSKERLQNTLLCDETLWRNN LPVDEKAQQFSRALATAVGQLTETMR RSADTKRWFTPHLQAVKLKRDEAYKIFK TTNCAEHWERYKQFRNEYVRELQKAK NSSVQQEIQSCSGDSKKLWRCLKDLIKP GGTPEKEIIFDDKTEPCSDEETANRLNH YFIKSVKEIHDSIPTVNQPQQDPLPTVRT LDISEFSQFQPISMEKLRKIVAAMKNCA GVENISKRVLEDAMDVVGDKLLDIINSS LNLGVFPQEWKKSVVVPIPKVPNSTRA EDRRPINMLPIYEKVLETVVREQLVEYV DRAGLLIEEQSGFRKQHSCESALNLLML KWKQSIENNKFILTERVFDLKRAFETIDRR KLEVLRRNRIKGDVWRWFKSYLEYRT QITRYNNSVSEAEAVELGVPQGSLGPL LFILYINDLTALRRACVNLFADDTVIYV AGDKLDECYEIMNNELAVFADWLKWK KLKLVSKTKYMVVTTRRSESNTISVD GELVERVNAIKYLGVMLDEKLSFAEHVD YTIRKAARKLGVLCRINRYLSFDNKIMVY KTLIAPHFDYCASILFLATNQQQRKRMQI			
--	--	--	--	--	--	--	--	--	--

						VQSKAMRMILRCDRLTPRILMLDSLQW MSIRQRVEYSTLVFIFKVVNGLAPQYLT KTVQCGRDVHRHFTRQAGDIRLLNFKK SCTQNSLFYRGYNLFNMLPESTRATTNL REFKKLCKPFVRQRPLE			
	RTE- 3_CQ	GCF_00020 9185.1	Culex quinq efascia tus	TGTGTGTGCGCCG TGCACATCGGTAT TGTTTTTCGTACC GTTCCCGCGAAA GTGCAAAATTTAC ACGAAAATACGC GTTAGTGGGGTG TGAAATTCATTGC ATCCGTGTTCTGA ACCCGTCTCCTGG CCCGGGAAAGTG ATAGTCCGTGTTT TACGACCGCGAG TGAGACAAATCA ATCACCAGATTTG ACCGAATATCGCG	TAGCGCCAATAAAAG TAAAGTAAGTAA	MVLRRDRGLVQATRTRRKTVPQEAHD ASRTNRHGTGHHMRSHDWKLGTWN CRSLKFDGSIRILSDILVRKFSIVALQEV GWIGAEVQAYPRIGCTIYQSRGENKRL GTAFIVLGEMRDRVIGWTPLTDRMCVL RIKGRFFNISIINVHSPHSGSEDDDKDAF YEQLNWTYNSCPKHDKVIGDFNAQV GQEEEFRPVIGKFSAHVRTNENGLRLID FATSKNMAVRSTCFQHNLRDKYTWRS PQGTESQIDHVVIDGRHFSDIIDVRTYR GANVDS DHYLMVMVKMRQRLSLAKSVR YRRPPRLDLERLKLPEVASRYAHSLEAAL PGEGELLEAPLEDCWRSVKAAITNAAES TIGFVERGRRNDWFDEECRAILEEKNA ARRAMLQYNLRDYEAYGQKRRQQH QLFRAKVRHQEELEFEDMEQLHRSNET	cd0 165 0	569	829

				TTCCGCCACGGGC AATTGAAAAAAA GTACAAAGAACT GCTAAGGCATCGT CGTCGTCGTCGCG ATCGATGGTCGTT GTATTGTGTCGCG CGGCGTGTGGAA GTGTACCGCGAA GACCATACTGACG CGTGTCGTCGTCG TCCTCGTCACTGT GTCTGTTGCTACG AGAGTCGTGTATA GAAGAGAAGCCG TTGGTGACCGGT GGGGTGAATTGC TAGGAAAGCTAC AGAAAAAAAAG TGCTGCAAAGTTT TTTGGGTGAAATT		RKFYKKLNGSRNGFTPRVEMCRDKNG AILTNEREVIDRWKQHFDEHLNSAEAE AGVQGGRRREFIGTAGEGEEPVPMTMR EVKDAIKKLKNNKAAGKDGIGAELIKM GPEKLASCLHRLIVRVWESEQLPEEWKE GVICPIYKKGDKLDCENYRAITILNAAYK VFSQILFSRLSPIAEGFVGSYQAGFVMG RSTTDQIFTVRQILQKCREYQVPTHHLFI DFKAAYDSVDREELWKIMDENGFPGKL IRLIKMTMDGARCCVKISGAESDSFTSL GGLRQGDGISCLCFNVVLEGVMRRAGF NMRGTIFSKSNQFICFADDMDIVGRTF KAVADAYTDLKREAEKVGLRVNVAKTK YLLAGGTESLRARIGPSVTIDGDEFEVVE EFVYLGSLVTSDNSCSREIRRRRIAGSRAY FGLHKSLRSRKFSLHTKCSIYKSLIRPVVL YGHETWTMLEEDLRALSVFERRVLRTIF GGVYENDGWRRRMNHELAQLYNPSI RKVAKAGRLQWAGHVARMUPERADHL SQRNQKINPAKLVFVSEPVGTRRRGVQ			
--	--	--	--	---	--	--	--	--	--

				AAAAAGAAGGAA AAATCAAGAAGC GCAACCCTGCTGC ATACACAGCCCCC CCCCCCCCCCTC TAGTTCCGTTTCT GGGCGTCTGCAC CCCATGTTAGGG GCGGCCCAAACG GACTAGGTGGTC GATTCCTTCTATC CGTGAGCGGCTG TTCCCCACGTTAG GGGCGGCTCGAA AGAAGTGGCCGA CACCCCCCCCCC CCCCCCCCACCCT TCTTTGAGCGTCT GTTCCCCAGGTTA GGGGCGGCTCAA AGAAACCGGTGT	RARWVDQVESDLESVGAPRNWRNAA MDRACWRRIVQQAKLMV			
--	--	--	--	--	--	--	--	--

				CCTGCTCTATCGT CGAGGTAAGCGT CTGTTCTCCAGGT TAGGGGCGGCTT ACAGCAGATAGA GTTCCGACCCCC ACCCCCTCCCCC ATCGAGCGTCTGT TCCCCAGGTTAGG GGCGGCTCGAAA CAGCGTCTGTACC CCAGGTTAGGGG CGGCTGAGTAAA AGTCCTTGTCG GCGTGGGACTGT AAACAGTACCGG CACG					
	Kiri-7_CQ	GCF_00020 9185.1	Culex quinqu efascia tus		TAGATCCAGATTAGA TTTTAGTTTTTTTTTC TCACCCGTTTCAGAGA ATTGACAATTTCTTTG	MKNDPAGDVIDADELNAFFSSGHRQLQ AVDRSDNSSEPCHRTTADHGDNGFAF RHTDVNEISRKILEVQTNATGTDDIPISF VKLLCPFVLPMLAHLFNHIIDTSSFPAA	cd0 165 0	111	371

					GCTCCAAATGTACAA TCTCCTTATCCCTTCC TTGCCCTGATACAAA GTGTAAACTAACAGG ATATCGTTACCAAAT AAATTACAATTACAA TTACAATTACA	WKKAIVTPIPKSSSPIQPKDFRPISVLPA VSKVLEKVLLGQISEHLNAAEPPLAQN QSGYRKGYSTTTALTKVVDVYSNLDE NRCTVMVLVDFSLAFNCVNHQILRKKL NTEFKFTRSACDLITSFLSGRSQIVRFGN SMSAALDVPDGPQGSCLSALLFSLYIN SLPRNLKCEWQLYADDLQVYLSGPPIAEV DRIMRDVNEDLAAIADWARVNQLFPN PKKTQAIVFNKTGTVTPQENIVFCDSIIP LSSQVINLGLHMDCNLTWKAQVNDVV RKVYHTLRTFRRFGSVLSLQTRRKLVA VIVPFFTYCDTVYFPLSAAALREQLHRGF KSALRFVHNLRRRDTTVALRNSIMGHD LPDNYQLRVCCFMKKAFEGTLPDYIMQ HLQRGRQERTAGFIIPRHTTSSGKSVLV HGASCWNSLPMAIKREARFNPFKRAVI THMQN			
	Rex1- 20_DR e	GCA_00869 2375.1	Danio rerio	ACGTGGC	ATTGTGCTGAACTTG AATTAGATATTGGAT ATGTACAAGTGTTTA CTCGTTGTTTTTGA	MMAPYLWQPVAALDVFNFVYLLYFL NIFLQWLRTLYLHTFQFESIRKHWIIFIFI LAFLIGSDAVNSPAIADAVIYSREQLLA LRHTSTFSVPTMEVPTELKRRRRGCRA	cd0 165 0	578	828

					ACCAAGGTGTGCAGT ATGAAATGTAATCTT ATTATTGTTTTATAT GTATGTAAC TTATTTA TCTATTTTTAGTATGG AGAGCTCTGCTGTGA CATGTGAATTTCCCTC CGGGGATTAATAAA GTCAA ACTAACTAA ACTAACTAA	GVKRRLLKKRYKPSLPVIIMGNVRS LGN KTEELTALVRTQREYRECSFMCFTETWL NDKTPDSVVALDRFHLVRADRC AKQTG KNKGGGLALFVNERWCRPAHITTKEVV CNKDIELLAVGMRPYYLPREFSYVIAITV YIPPSANASNACEHINTVTASIQLKYPHS LIVISGDFNHVSLQSSLHTFTQYVDCGT RDCKTLDMLYANVKDAYASTPLPPLGR SDHNLIYIRSVYVPAVRKQPPTIRTVQV WSDSVYEALMDCFDSTDWEVLCKSHD EDIDSLTDCITDYITFVLR TQYPQKRLNA TQTISHGLYNFCVENTVPTKKVKCYSNN KPWVTP ELKALLNEKKRAFLAGDREEIK RVQKELKIKIKEGKVIINLKWNINYNKIM LEMYGVIVCTKLEVTIEQVRNGLRRIKKK KSAGPDGISAGVLKACADQICWVVHYI FNLSRLERV PILWKTSCVVPVPKKPNP TEFSHYRPIALTAQLMKVFERLV LNYIKK LLIASEDRLQFAYKTGVGVDDAIYLLHK SLSHLETSGNTVRITFFDFSSAFNTIQPA LLKVKLENAGLQKGMVSWLTDYITNRP			
--	--	--	--	--	---	---	--	--	--

						QYVRMHNCVSQVIMSSTGVPQGTVLA PYLFTFYTSDFQYSSKNCHLQKFSDDSAI VGYIRNHDESEYRLLLKSFADWCDMNC LNLNTSKTKEIIVDFSRHPSQAVPVIIQES EIEIVSSYRYLGVYLNKLDWSENTIQVY KKGQSRIHLLRRLRSFGVCQTLLKTFYDS VVSSVILYAVTCWGGGLEKEKNKLNKL IKRAGSVVGGALPTIEEIAQDRMLDKFV SIMNHDSHPLYDTVQACASSFSTRLIQP RCYRERYRKSFLPTVIRLYNQSHTR			
	L2- 6_DRe	GCA_00869 2375.1	Danio rerio		GAAAAGAGATCATAT CTCCCCTGTCCTTGCC TCATTGCACTGGCTA CCTATAAAATACAGG GTTGATTTTAAAATTC TTCTTTTCACTTATAA AGCTTTGCACAATTC TGCCCCGGCCTATGT CAGCGAGCTTATTAG GCCATATACAGCTTG CAGATTCTTAAGATC	MLTQPVVSFRVHENSALFNSYTNFL SNANASLPSCSDTEQLAETFNSACDFAL NEVAPLKSRSRVSSIPQPWLNSSTRTL RQECRRAERRWKKDGLQVSYQILKDCL NSYQDSVSEAKSKYFSNLIASKASKPKVL FKTINSVLNPTMCSVPVVNLETCIKFSDF FAQKIQDIKSHLFPTDIDHSDSVPPSNSF TNFQPVSSSQLLNTISQMKLTFPCPSDVL PCRLFKEVLDTICPSLLALINSSLINGVFP DVFKHAVVQPLLKRPNLDPAPVDHYRP ISKLSFISKVLEKIVLSQLTSYLDIFYILDPF	cd0 165 0	260	521

					ACATGATCAGTTTCTT TTGTCTGTCCCTCGGT CTCGCTGCAAGTCAA AGGGTGATCGAGCTT TCAGTGTTGCGGCC CAAGGCTCTGGAATA GTCTCCCTCTGAATAT CAGAATCTCTCCTTCT TTGGATGCTTTTAAA TCTAACTTAAAGACA CATTTTTACTCTCTTG CCTTTGACTGATGCA TGCATGTTATTTTTGT ATTATTTATATTTTAG TATTACTTTTAAATAT ATTTTATTGTATTTT AATACCAGTTATTTT AGTTTTATTGTCTAGT TTATTTTATTGTAAAG CACTTTGGATCAACT ACGGTTGTGTAAAT	QSGFRAKHSTESALLKVTNDILLSMDSG SSAILILLDLSAAFDTVHDNILLRRLECVV GVQGLALQWFSSYLKGRSFSVNIGSFTS PPKPIHCGVPQGSILGPLLFSLYMLPLSSI LLKHKISYHCYADDLQLYFPIKLGKDLSE NLFLCLTDIKTWMTNNSLQLNENKTEV LILGPAATNNSIKTQLGALNDNLHNHA KNLGVIFDPLLQFDKHAVNAVVKSSFFQL RLVAKVKRFLSKKRFRNRNPCFNLLSSRL LQFFVYWPTSVNTLSATVGPKCSRQIN NRN			
--	--	--	--	--	---	---	--	--	--

					TGTGCTCTATAAATA AAGTTTGCCTTGCCTT GCCTTGCCT				
	Rex1- 23_DR e	GCA_00869 2375.1	Danio rerio		TGTATGGATGTTGTA TAGTAGTATTTTTTTA TTTTTATTTTTTAATG TGTACATGGAACTTC GCCTGTGTATGGTAC CTTGTTGCAATTTAGT TTCCCTAACGGGATA ATAAAGCTGAGTCTG AGTCTGAGTCTGAGT CT	MLDLCYGSIKGAMKAVAGPPVGFSDH NVVHLLPVYKSVLRREKVRKREIKVWSE DNSLALQGCFDCTDWSVFKETSESIDEL TEVVCDYISFCTDMIIPHKTVIDFPNNKP WLSNDTKSLLQSKHKYFKQGDMSSVK QLKKAVRVQINKDKIRYKDKIENQLSSN NLRSAWQGVKAMAGCKEDRGGSEINL NGFSSAREFAEELNTFYLRFDEDDFKSE NDNLKQHLSCSNSVNISRIQVKGLLQRT NTKKSPGPDNIGGRVLHTCAEQLSEIFH YIFSQSLMLQRPVEIWKKSVIVPIAKTKF PRELNDFRPVALTSLVMKCFERFIKGEIL LQSKDKFDPLQFAFRQSRGVEDANLTIL NYLFNHLENPKTHARLLFIDFSSAFNTM KPHLLVEKLISLFNLDLNICGWILDFLVG RQQCVRVNGALSDFQFCSTGSPQGCCL SPLL FIMYTND CRSIHRDCHIIKYADDSV IVSLLHDQDLGHGPAVQEFTSWCDRFS	cd0 165 0	292	545

						LQLNASKTKDMVIDYRKASPTPNITNIK GLDIEIVDTHKYLGVVIDNMLTFQPNTQ AVYKKVQQRLFFLRKLRSFQVSTSMMT LFYKQFIESVLSFCIVAWYGNLSVSSKNR LSSLVKVAGKIIIGFNQTGLMVIYQNHVI RRAQGILCTPDHPLYSVFQLLPSGRRFR VPICRTNRTRNSFIMVAIRSLNESRGEG DVSIIY			
	REX1- 7_DR	GCA_00869 2375.1	Danio rerio	TGCGCAGGCGCG AGCTTGCTGCAGC ATGTTTTGACTGC TCTGTCTTGTCTT AGTGTTTTAGAGT TTAATTTGTTGTC TTTATTTTTATTTT TTTGTTTTAGTTTA TTAGTGTTAGTAA GCGTGCCCTCCTT AAACATGAGGGT AGAGAGAGCCAT CTCTGTGTTTTTTT	TAATTTATGCAATATC TTGCTGCTCGTGCCA TATATTAATTAATTAT TACGGGTTTTAATGC GTGGAAGTCTGTGCA TTATGTTCTATGATGT GCAATAGTCTATTAG GTGCAATATTTGTAC GTTTTTTTACTGGGTC TTAAATATTTTTAGTG CAATATCTTTGTTTTT TAAATGAGTGCTGAA TGACTTTACTACTTGA	MTWLHECISDSVVDVAGFSSVRADTR RESGKSKGGGLAVFVNERWCHPNHIVI KERICSPDIELLAVGLRPYYLPREFSHVIAI VVYIPPSANATTACDVIHSATARLQTM HQDAFVLISGDFNHVSLNRTLSTFSQYV ECPTRENKTLDLLYANVEEAYSCIAIPPL GRSDHNLVYLQSTYLPVKGSQLLLGGQG SGQKRPVQPKTAWRLRTGMLCACH MVWTLITSQTVSLSISTSVWIVLYHEIHT LLSYCTMKSIRCFPNNKPWVTKDIKALL NEKKVAFRSGDKEAAKEVQRRLSVRLK EGKERYKKKLEQELKQNSRYVWRGM KNITGFKTSSPRTDGDVDRANQFNMFF	cd0 165 0	491	742

				GGTTACTGCTCTT	ATTATTGTATGTGTA	NRFDSRPLSSSDTALTAPSASAPPNSVT			
				TGTTTGAGCTTGC	ATTGTTTTATGTGTAA	SNCSILPYFTPSPSSSFIVPTEASIASPPFI			
				CGCACGCTGCTAA	TATTATTGGTTTTGGT	GVSDSYKPPMIITEDQVRRELCRLHPAK			
				TGCTAACACAGCT	TAATATTATTTTCATT	APGPDGLTTQVLKLCASQLSGVLLRIFN			
				AGCGGTTGGTCC	ACTTGATATATGACT	LSLSIKKVPVMWKTSCLPVPKKARPSV			
				GGTTTTTGTTTAT	TATTAGTATTAGTTAC	PSDYRPVALTSHSMKVFERLVLTLRPL			
				ACCCCGGACCAG	CACTGATTCTTTTAAG	VRSAQDSLQFAYQAKIGVEDAIIYLLHR			
				CTGATCTCGTTGA	TGTAATATCTTATATA	AYSHLDKPSASLRITFFDFTSAFNTIQPA			
				GGCAGGATGCTT	ATAGTTTGTATACTTA	RLGSKLSAMQVHAPLVAWIMDYLTDR			
				CGAGCACAGTGC	TGTATTGTGGCTGTT	PQYVRLQGNKSDTVLCSTGAPQGTVLS			
				GGTCTCTCGGGTA	TTCTTTTATAGATTGG	PFLFTLYTSDFQYNTEGCHLQKFSDDSAI			
				TCCCCGCGGGTTA	TTGGATTTATGTGAC	VGCIKGGDDLEYRMAVDSFVDWCELN			
				CGGAGGGGGAAG	TTATTTCACTTACTTA	QLQLNIQKTKELVVDLRRSRRSPVTPLSI			
				ACTAAGAGAGGC	ATTTATTTATTTTGTG	RGVDVEIVQDYKYLGVYIDNKLDWSKN			
				TGCCGTGGAGGA	TTGTGGTTTTGTGTTT	SLVTYKKGQSRLYFLRKLSFRVCNTML			
				CAAGGTGCGAGG	CTGTGTTCTGTGTGTT	RMFYESVVASAIFYAVVCWGGSVKVA			
				GAAAAGAACGAG	ATGGCAGTCATTACA	DMKRFNKLIRKAGSVVGEELDNMETV			
				GAATGGGAAATT	TAAGCAATTTCTGCT	AEKRTLCLRLSIMHNVDHPLHNVDDEL			
				GTCTGATGAGCA	GGGATAAATAAAGTT	RSTFSHRLTSLKCSTERHRKSFLPTAIRLH			
				GCTGAGAAGACG	TTCTAAGTCTAAGTCT	NVSLSHLPLQ			
				GTATAAACCTAAC	AAGTCTAA				

				CTGCCGTCTGTTG TTATGGGCAACGT GAGGTCACCTGGA AAATAAGATGGA CGAGCTTACGGC ACGGGTCAGATG CACAGAGAATTTT GTGAGTGTAGTTT GATGTGTTTTACT GA						
	Tx1-49_DR	GCA_00869 2375.1	Danio rerio	C	TAAAAGGCTTTTGTA AGACCTATAGCACTC AACTAATTAACAAAA TTTTAATTATTTATTG AGACATATTCGCATA TTAATATAAATGAGT AAATAAATAAAATGA ATAGTATAAATGAAT AAAATGGGTATTA CAAGATATAAT	MYYMQEKEKGGVLIGLDLEKAFDRVEH DFLFAVLKEFGFGENFRKWIQIFLYRYF NLYLVLNVFFFLTECLQPTRSIRQGCPLT ALLYTLEAEPLGLAIKSDRSIKGIQIENND NREAVYQYADDTTSLGILERYCKASGAK VNVGKSLCMRVGRVKEITQHIPPKEEKE HTKMLGIWVGMDSNKADILNWERVL NGIEKRLRFWKARNLTGKYSSKFDVK RKVFILNALMLSKVWYVLSVTALPEVYY EKINDVVLKFLWGNGRAKIAYYTLIGE KEGGLGLLDSFLKMKTLTIKIIQRFLEGK	SSF 566 72	8	159	

						MGVIWKSVLDDYLLKKCGHFNMGCNVL LMKMKKRMIAGLPKFYVEVLESGLRIM RGIFRKCVDYTKSKGADSRTTTFTELNN FKTILSEGNVLYFKEWWDGGLRQVKDV LYEVIKGFVPLKVIVEDGYKRVVKAKLQR QCHELKKAIKPKDCIVIEKKKSWAMKM NYQRPILRRKMERSMSLCTVKSFYAIFR			
	L2- 42_DR e	GCA_00869 2375.1	Danio rerio	GACGTGTGCAAA AG	TAATCTTTATTGTGGT TGTTGCTGTAATACTT TTATTCCTATTATTAC TTTTATTGTTATCTTT TACTCTTTTAGAGTGT GTGCTATGGTTTTGT TCAGCACTTTGGTCA GTGCCAAACTGAAGT AAATGTGCTTTATAA ATAAACTTTACTTACT TACTTACTT	MAPVCLSACRLPRQFLNVCSATILSVLFL LLFMSCVDALISYDRQALLNIRGLLENQ PRDCSITHDYCYSSFPPTAAFVRRLPCC LIYQRKRRRRRGNRGGVRVRMRKEISIS LALHTACSFVPTYGHRGLHPARRSWDV RYTCHLPIFPDKRQLAEFGAPPRVRIRRR GVNLQHIKSLDYVHQSDLQVISHINMA LLNARSVSNKTFILNDFFTSHNLDLFLTLT ETWIGAGESSVFGELCPLNCSFISTPRCI GKGGGVAMVFRNWFRCTISVGDFTT FESQCALIESDTTPFLCVLVYRPPKPD RG FIKEFSEFVSHIITSYDRLLILGDFNIHVCC PGKPLVTEFMHVLDSFGFTQHIVGATH ALGHTLDLIMSYGFSIENIVIEDASFSDH	cd0 165 0	673	932

						KPIVFNTTLSYFSGTTKSEGFHSRCINSLT ANQFSTMFNSNKVSTAILAAAEHSSSP DDLISLFNSSCDILDCVAPLKYKYPKMK SQPWLDDLSRSLRQICRKAERRWRKDH LVVSLEIFRDAQVNFQKAAKKARATYFS DIINKHSHTPKILFDTINAIISPRTSCKIEQ SSEICERFKRFFTEKVADIRSSFAPLCPSS CVDSTVSSATFCNFEPVSLSELRGLVFQT KTSTSPCDPISSNLLKETFNTIGSSIQAIIIN SCLVSGTVPAFLKHAVVQPLLKKPNADI NCLNNFRPISKLPFMSKVLERAVLSQLQ PFLKSSHKLEPLQSGFTMHHSTETALLK VWNDLLLSVDSGSNAVLVLLDLSSAFDT VDHTILLSRLEHWVGIGSALQWFQSYL KSRFSVSVGQFMSTSAPLLCGVPQGSI LGPMLFNLYMLPLGNIIRKHNISFHFYA DDSQLYVPLKSQNSLQSLINCLEDIKCW MAKNFLQLNNDKTEVILFGPSKTRMSYI GGLGSLAPYVKSQVKNLGVIFDSELCLE KQISSVVKNSYYQLRIISRLKSILSFQDLE KVIHAFITSRLDYCNSLYVGLPQSSLARL			
--	--	--	--	--	--	---	--	--	--

						QLVQNAAARLLTGTKKHAHISPVLASLH WLPVNFRIQFKILLIVFKALNGQAPSYIS ELINLQSTPRSLRSSNKILLQVPRSRLKLK GDRAFAVAAPRLWNQLPPDITSASSLSI FKSRLKTHFYIAFPGF			
	RTE-1_DR	GCA_00869 2375.1	Danio rerio	GGTTCGAAGGTTG TCATAGCCGGGG GCGGTAATGGGG ATAAGCTCCCACT ATCTATAAAGTAC CCCTATACGGCGT GCGTCTCAAATAG CCTCTGACAACCA AGTCCAGCTCCTG GCCTTCAAGTGTG GTTTAGCTACCAA ACCCGGCGGAAC TGTTTTCACTGAC AGGAGAAGGGGC GCAGGCGGGTCA CTGGCGCCTTACA	TAGGATACGTCATCC ATGGTCAACACTGAC CGACGGAGGCCTACT A	MLTGLSEDLRGIDDSRKTAVINNELLRL NVDIAALQETRLADSGTLKEKDYTFYW QGRAPDEPRQHGVAFVAVKNNLLSMVE PGRNGTERLLTLRLNTTGTPLTVSVYA PTLNATLETKDKFYGNLTSVINNIPDKE QLVLLGDFNARVGANHESWPSCLGKF GIGKMNENGQRLELCAFHNLCIANSY FQTKPQHKVSWRHPRSKHWHQLDLIL VRRSAINCVLHVRSYHSADCDTDHSLVC CKIRLNPKRFHRLRKQGNPRINVS RMS QPDLTQKFAETLVKELDTAQTGDSALE MWESLRNTMQRTALATFGKRTTKTHD WFEAKASTMIPCIEAKRAALTEYKRSPT QKNLQILRSTRSKAQHIARYCANEWQ ELSNDIQKAAIAGNIRGMYDGIKKALGP TQCKTAPLKSSTGEIISDKQQQMERWV	cd0 165 0	526	803

				ACCAGTTGCTTCG GGGAGATGATAT TCGTTAGCCTGGG AAGGCAGATCAT CTAGGGGAAGGC AACCTGTTTTCA AACCTCCGCTGCC TTGCGGCTATATC CATTGATGGAAAA GGCTTCAGGAGT AAACCTCGAGGA AAAATCCGGAGT CGGAGTCCCTAA GGCAGTTTAACGC TGTTTGAGCCTC ACTCTGGCAACTC CTGCGACGGCGC CGATACCAAACG TAGCAGCCCTGCT GTTCCTTTGGATT TGTCGACAACGTG		EHYSGLYSRQNTVSSAALDVICLPTME ELDEEPTAEELRKAIDKLASGKAPGSDGI PPDLLKQCKDSLHPLHKALCQCWKEG AVPQDMRDAKIITLYKNKGERSDCNMY RGISLLSIVGKVFVKVILARLQKLAERVYP ESQCGFRAERSTVDMIFSLRQLQEKCRE QQMPLYISFIDLTAKFDLVSRDGLFKILP KIGCPPKLQSLIESFHTDMKGTIQFNGS CSEPFSSSGVKQGCVLAPTLFGIFFALLL RHAFGSATEGIYLRTRSDGKLFNLSRLKA KTKVRETLIRDMLFADDAAVTTHTEEL QSLMTRFSMACKDFGLTISLKKTNVLST DTATPPTITVDDYQLDVVHQFTYLGSTI TDNLSLDAELDRRIGKAASLRLTRV WTNHRLTTATKMAVYNACVISTLLYGS ETWTTYARQERRLNTFHLRGLRRILGIT WQDKVSNVKVLTRAGLPSMYTMLRQ RRLRWLGHVCRMEDGRIPKDILYGELSS GKRTTGRPFLRFKDVVKRDMKALDINT KSWEDLAADRLKWRCTLTQKLKSGEKN MMRASEDKRVHRKVQSVSQGATYQC			
--	--	--	--	--	--	---	--	--	--

				GAGAGGGGGGAC CCGCTACATGGGC AACAGCCTGTCCT CCATAATACATTG CCCTGGCTAGTAT CCGATCTCGCACA CCCTGGAGAGGA CACTCCAGCCTCG CTAGCACTCTGGC GTGGATACAACA CGGGGAGCAGTA GTTTACCGGTTAT AAGCCACAGCTC GGTTGGCGTAGA GCAAGGCGCCAG GGA CTGCTCCCGA CGGTGGGAGGGA TCTTCGGGTCCCA CTGGACAGTCACC GCCCCCCTCAAGC TGGGCAGCCCCC	DCCGKKCLSRI GLFSHQRRCLRQSVMP QYKS			
--	--	--	--	---	--------------------------------------	--	--	--

				AGTCAATTAGGTA CTGCCCCGCCACA GTCTGCTTTCCTC ATTGGGTGCATG GGGAATAGGAGC ATTTCGAACAGCA GACTGCAACCATC GCACCAGACAAT AAAATAACACAG AGAAAGAGACCA GCTCTAAACTGG GATGCTGGAATG TCCGTACA					
	Rex1- 26_DR e	GCA_00869 2375.1	Danio rerio		GATGCAGGACTGCCA GACTCAAGAACTCTT TCTTCCCATCAGCCAT CAGACTTCTAAACAG ATAACCAACGCACAT AACAGTCTATTTTCT GCTCAGTACAACACT ACATACTCCATCCCAT	MSWRPSWQTCMLAPHINVVLLFFVC FLSFPRCIHAMHYQHIADKLAGSCNLR QTFRIYQFYSTGLTTHCPLPQPLQRRRR KRHKRGKRGGLHARLKSRANRPPLPSIL LANVRSLDNKLDELRARITSQREVRECC ALIFTETWLSEKVPGTAVQLQTHSLHRG DRTTASGKAKGGGVCFINNSWCGDV QTVHKHCSPDVEFLLKCRPYLPPREFT	cd0 165 0	581	834

					TTTGCAC TATTTTATT	AVFIAAVYIPPRANATAALGKLFNVFNA			
					TTTTATTCTTTTTTTTT	QEMAH PDAVIAAGDFNQC NLRTVLPK			
					GCACATAATCCAATT	YHQHVSIPTRENNTLDNVYSNIRGAYR			
					GCACTAGGCACTTTT	AAPRPHFGQSDHISLFLYPAYRQRLKQT			
					TTGCATAATAAGCAC	NPITKL VKIWN PQTESTLQDCFALTDW			
					AATAAAAAACAACAA	DVFKTAATKEDLSVNVQDYAEYVTGYIS			
					AAACAAAAA CACTG	TCVENI IPTIQVKKFPNQKPWINSKVRH			
					TACACTGTTTACATCT	MLNARSLAFTSGNENEYKSAKYGLRSAI			
					GTTTCATTACTCAAAT	KEAKRQYKEKLDNTYSTASAGQMWRG			
					CAGGTTTACATATTG	LQHITDYKSTTNVTISASESLPDDLNTFY			
					TTTACTTT CATAGATA	ARFESSSINTEQRHTQTQTSQPALPPP V			
					TTTATATACTTATATT	VTPAAVHKALRKINPRKAAGPDNIPGQ			
					TTACATTCAATCTTCG	ALKACASELVDVFTSIFNLSLSQSSVPTC			
					GTCAC TTCTGTGTAT	FKTTTIVPLPKKSPLTCLNDYRPIALTPIIA			
					ATATGTGTATTTATGT	KCFERVVLPLIQSSIPDTLDPLQYAYRTN			
					GTATGTTGTGTATGA	RSTSDAIAAALHTSLSHLEDKDTYIRMLF			
					TGTGTATGTTGTGTG	IDYSSAFNTVIPHNLT HKLSELGLHPTLC			
					TTTGACTTCACTGTG	DWLLDFLTGRLQSVRIGNKSSSTILT NID			
					GACGGCAAAGTAAG	TPQRCVLSPILYTLFTHDCVASHKNNTII			
					AATCTCATTGTACAG	KFADDTAVIGCITGGDEAAYRKEGASLV			
					GGAGACGTGTTTCCT	TWCDNNNLTLNTDKTKEMIVDMRKER			

					TACTGTGCACATGAC AATAAACAGTTGAAT TGAATTGAATTGAAT T	KTHQPLFIRKLEVERVSSFKYLGVHISED LTWTLNNTQLVKKAQQRLYFLRKLRKLG LSSKILSNFYSCVVESILTNCITVWYGNA TVKDRKRLPRVVRTAEKIIRSPLPSLQTIY HHRVHKRTASILKDPTHQPQHGYSHSYP QGGGIGV			
	RTE- 2_DAn	GCF_00328 5975.2	Droso phila ananas sae	GGGGAGAATGGG GGAGGTTATATA AGGTAAAGTGTCT GATTGCACTCAAC TTGGTGAATGCCA GCATGATCCTTAT CACTGCTCTCCCC AAGCCGACGCAC TACCTCCACGTGG TTCTCCCTGGATT GTCACTCAACACC CGTTGTTTGACAA ACTAAAGCGCGC GACGAGAGTCCT AGCTCTTCAAAC	TAGGGTTGTGATGGC CATTATATATTATAT	MRWIGQGCLRKKNCDIYYSCHPERHEF GCGFVVGARLRRRVSHFLPINERLATIRI TAKFFNISLICAHAPTEEKDDVTKDAFYA ELDRAYGRCPSHDIKILLGDFNAKVGRE DIFGATVGQFSLHETTSSNGLRLIGFAA AHNMVVRSTGFRHLDIHKASWLSPDRL TRNQIDHVVIDARHASNVLDVRSCRGP NIDSDHYLVAAKIRTRLCVAKVARRGLL RKLDIGKLQSQRTDAFSTQVTGLLSRA TPTPEDISNMWALISHSLRTSAEEVIGF QRTPKRNPWYDQECRDATAAKDAAYG RTLQAGATRAVVEVYRSRRRDEKRLFR RKKREQUERRECESIESSRDRNEARNFYQ RVKRLTQGFKTGGVACKDDDGNLITDA EIVLRLWRDHFSSLLAGSGHDDYGEYD	cd0 165 0	490	743

				AGGTTGTGGCGT AGGACTTGAAATC CACCCACGGAAA ACAAATCTTCAGT GAATGTAGCAGA TGAAGCCTCGGA ACGGAACCAATCT AAAGTTGACGAC CATGGCACACGTA AAAAGGACAATG ATTTGAGAGTATG CACTTGGAATGTA CGCACATTGTACA GACCAGGTGCTG CTCAACAGCTAGC AGATACCCTCAAT AAATGTAGGGCT GACATCACTGCTA TCCAGGAA		PQTPIYGTDVDVPTPSHAEVKDAIQRLK NNKTAGADGLPAELFKAAGDMLVGSM HKLISKIWLTESMPDDWNHSMICPILKK GDATLRTNYRGISLLPVAYKVLTSVLCEN LKPHAEALIGPYQCGFRPGKSTIDQIFTL RQILEKSYENQIDTYHLFVDYKAAF DSP RRDRLYAAMSELGIPAKLIRLCMMTLSN TISSVRVGNDQSETFYTKCGLRQGDSLS CMLFNILMESIIRKAGVHRTGTILTNR VQLGYADDIDIIGRTKRDVTAAFGAIEK ESAKVGLAVNMDKTKFMVCSSRESRRL DSQLSAGSHSFESVKEFIYLGSAISSTND VSLEIKRRITLANRCYFGLSRHFNSRALS RPTKITLYRTLILPVLLYGAEWCWTVVQSD AAALGVFERKILRKIFGPICVGDAYRTR WNHELYELYGDVDIVSRVKIQRLRWLG HIARMEEDAPARKVFDVIVGKRRRGR PRIRWQDQVMEALSTSGVTNWRARA QDRKAWRQIVQQAVTR			
--	--	--	--	--	--	---	--	--	--

	Jockey - 14_DAN	GCF_00328 5975.2	Drosophila ananasae	AG	TGATACCACACCTGA CCTAGTCTCCATTACT ACCTCACAGATGGCC ATAACTACAGAACCC GACGCCATGACACAG CAACGCCCCGCCAAT AGCCGTGCAGCAACA GCCCACTAATACGAC GCACAGACGCCCCGAT CCAGCCGCCACTGCT CCGATTACAAAAACA AGCGACTGCGTCCAG GCAGAAACCCCGCTA CGCAGCAACGGCAC ATCAACACGACGCTC AGCCACCCGGACCAG CCGTCACCACTCTGG CCACCAGACCAAACC CCCGACGACGCCCAG CAGAACTCCGCCGC GCAGCATCGCCTCAT	MQDIDDAIEIFNRNIYHAAKIATRHIKTS NRNPTGQFLLDRKLLNLQAIKKYTKQIK MARQTPEAAKNYKAANERFLRAIKEAK TKRTNNILNKIDPLDRYRMQKLWKITN HFKRQPEPNWPLKLYTGDGTSAKDWY SWTKSSEEKTEAFTRHLEERFSPKLTNSE SFRAHVISKKEEICTQKLLMTTPCWMH RENYGHNSDKTNLPFLPITPNEVLAEVK LLQNKKAPGPDQIDNKLKLLPHKATLH LTHIYNCLMRYSYFPKQWKRAAITMIKK PGKTPHEVSSYRPISLLPGFSKIFETLLM HRMFKIQDFERAIPSHQFGFRKQHSTE QQLLRVTQHILSAYEQKMYCSAVFIDVS EAFDRVWHDGLLLKLAKLLPHSLYKILQ SYLTDRSFTMNSTNGIPSRTGKISAGVP QGSVLGPILYTVFTSDMPLPEVISSQHT YNPSYFENMMLSTYADDTIILCTKREPQ LAIQTTESFLKYITDWATSWCITLNPNT IHVLFTKRSVANLTGNKSPKMKGILIKNE PKHRYLGLILDRKLLLHHHINQLQNRIR NVAAGLNWIIGTKSKLSPQSKILLYKQII	cd0 165 0	268	540
--	-----------------------	---------------------	------------------------	----	---	--	-----------------	-----	-----

					CAGTATGACGCACAA CCACCTGCATCAGCC GTCACTACTCTGGCC TCCAGACCAAGCTAC AGCAACCAGGCAGC AAACCCGCCACGCAG CAACGGCTTGCCAAC ACGCACACATGATCC TGGAACAGCCGCTGC CGACTTGGCCTCCAA AACCAACAACGGCGT CCAGGCAGCTACATC GCAGCGCTGCAGCA ACGGCCAGCCAACAC GTCGCACAAGAACCT AGAAAAGCAGCCACC ATTCTGACCACCAGA ACCAACAATGGCATC CAGGCAACCACCCCG CACAAAGCACAGGCC CACCAACCACTGAAG	APIWQYALPLWGALTSNTQFQRIEAAQ SKLIRKILKAGRYTRNQVLRDAYNIKSIN EIYSEASNRLATSLASHPNTEARKILMDP VTPNRLEMPRYLDQLNRFFLPLRQQLQ QQNQLNTHQPALPTLISYEQQLLSQQA SIKARQQDEEREQNKRLKIEQDPHYRPS IGWVNAMSYKLRHGITTREWEMADKIK YEPEYIQRILMSPLLSEIPPPTPDQELD ARLSTLKTQIINDLDLNLNYETNLAEGDS RTRFLPNSSSPILQAPQIPCINLVEDHPL GNPAIADIWKSTPEI			
--	--	--	--	--	---	---	--	--	--

					AAAACACCTGAAAGC GCCGCCAATGGATTG ACCACCAAACCTCCCG ACGACGCCCAGCAGT AACCCCGCCGCGCAG CAACGGCTCACCAAT ACGACACACTGATAC CTGGACCAACCGCCA CTGCCCTGGCCAACA GACCAAACGACGGC ATCCAGGCAGCAACT CCGCCACGCAGCAAA GGCTTGTCAA				
	R2_DP e	GCF_00328 6085.1	Droso phila persim ilis	AAGATATGGGTCT GAATAATAGCGT AGAAGGGGAGTC ATTCCGTAATTCG TAAATCGTAAAAA TCAGATCAAGTTG ATTCAAGACCTCC TCGTGGTATCTTC	TAGCCTATACACTAT GTTGGAGAGAAGAC GCTTGCTACCTAGGC AAAATGTGAAATTAG GTATAAACATCGTGG TTGTAAAACTTGAGG TGGGTTTTTAGTACG TATGCGTGATTACTT	MSSFGLIVTNLNSETVLWGCQPLGQFS LIGTNMQNTTPRIINTNSLTNQIPTVSSL GAQSEHSAQVNPNSGYQCTICESSFRS KSLGLGVHMSRRHKDEFDQLRLRTDRKA QWSEEELSMMARKEIELAANGERYLNK KLAEVFTNRSVDAIKKCRQRERYKTKIE QLKGQAVPLPEALESETIQRPSIRERDL LVTPPNTLGTTPELSNSEILAVLQGYPP	cd0 165 0	411	673

				TGGATGCTATTAG ACTGA	CGTAATCATGAATCG TGCATGCTAGTGGGG TTTGGCCTCCACTAG TATCTTTGAAGATTTT CCTTCCTCAGCGATC AAAAAAA	VVCNDQWRVEVLQSIVDGAQASGKEI TLQRLSTYLMEVFPSQNDRIQTRPPRR PRNRRQGRRQYALTQRNWDKHKGR CIKAILDGTGTATMPSQGIMGSYWRQ VMTQTSPTYSGTNTTFRTEHPLEGVWS PITLGDLRVHRVSLTKSPGPDGITPRTVR SIPSGVMLRIMNLILWCGKLPVSIRQAR TIFIPKVGNASRPQDFRPITVQSVMRIL NAILASRLTSSVDWDPRQRGFLPTDGC ADNTTIVDLILRDHHKRCKSLYIATLDISK AFDSVSHAASATLTAYGAPKEFVDYV QNSYEVCGTTLNGDGWRSEEFIPARGV RQGDPLSPIIFNLIDQLLSYPNEIGATI GDHTTNAAAFADDIVLFAETRLGLQTM LDTTVDFLSSVGLTLNSDKCFTVGIKGQ PKQKCTVIPETFRIGSRSCPALKRTDE WKYLGITFTAQGRTRYSPADDLGPKLLR LTRSPLKPQQKLFALRTVLIPQLYHKLTL GSVMIGVLRKCDILVRSTVRKWLGLPLD VSTAFFHAPHTYGGLGIPSVRWWAPML RMKRLSNIKWAHLAQSEAASSFLTDEL			
--	--	--	--	------------------------	---	--	--	--	--

						NKARGRTLGLNELTSRTEIETYWANRL YMSVDGRGLREAGLFRPQHWVCQP TRLLTGQDYRNSIKLRINALPSRSRTTRG RNELERQCRAGCDAPETTNHILQNCYR THGRRVARHNCVVNNLKRILEEKGHTV HVEPSLQLETSVSKPDLVCIRDNHACVI DAQIITDGLFLDDVHHRKVEKYKRPEVIS ALRREFGVSGNVEVLSATLNWRGIWSN QSVRRLIAKGLISSGDSNVISARVVTGGL YCFRQFMYLAGYTRDWT			
	TART-1_DWi	GCF_000005925.1	Drosophila willistoni	AATATTGGCTTAT TTTACTTAAATATT TGGTCTCTTTATG TTTTCCCCACTAA AAGCTAATTGCTC TAAAACTTCTAGC CTATAAAGCTAAC TCAACTAAAAGCT GACTGCTCTAAAA CTTTtagccttta AAGCCAAGCCTTG	TGAGGAAAAAATCA GGGACAACCACTTGT CATTCAAAGGCTATC CTGCCCCCCTTGGGC CAAACGAACAAATTC AATACACGGAAAATG TACTACTCGCTCTAA ACATGGCTAGTCTGG AATCTGAGTTGCTCA AATTAATTGATAACA GCACGAATAGCACTG	MNILNKEISLSKNKHLNKKLCKIDSSDRY RPHKLWQLTNKLKRQPLPNWPLRIPSG VDKKHAWTRSPEGKMEFTDQLEKRFS PNLTNSSLDRAKINEETNAILAQKMLM EYACKTHHSIASSSANFDSSNSPFVKPIT LSELKKEIDSLINKKAPGYDRINNALIKVL PLKAIIFLLKIYNSILRFGHFPLVWKKAI VMVKKPKKTISDASSYRPISLLPAFSKILE SLIAKRLLKFNRLSQSIPNHQFGFRKFHG AEHQLARVTQFLQRAFEDKLYGSPIFIDI TEAFGKVWHCGLRNKLIKLLPINLYKVL	cd01650	195	471

				CTCGCAATTGATG	ATGCTCTAAAAAAA	DSYLKDRTFSVRGNHGIMSRDGKILAGI			
				CGCCAACAGCCA	AGGAAAAGTACT	PQGSVLGPLLYNIYSSDMPKFIIEHETL			
				AGCTAAAAAACTA	GCACCCATGATGGCC	IEYYPTYKHGDTLLSTYADDTIFISTDKNA			
				GTGTTACTCAGAC	CTAACTAACAATCAC	LVAMNKTENYLHKFSTWSNKCFTVN			
				CCAAAATATGGGT	TCTAACACCAATACT	VSKTVHVLFSRRSTSPAVKAATVKFNNK			
				CTCCCTACTTTTTT	AATACTTCTGCAAAC	PIKNTSSCRYLGVLVDSKLLFKTHIAKLCA			
				CTCCAGTGTGGA	ATTCATTACCTAACC	QAKSTYTRYKWLGGAKSHLSTRCKVTIY			
				GGAAAACCAAGTTT	ATAAAGAAATATGAT	KQILVPIWTYAPIWGLISNTLFKKLTGT			
				TTTCCCGACATT	TTTAGGCGGCCCTG	HNKLLRKLTNANIYCPNSEIQEKHNAKP			
				AATCGCCACGTTT	CAAAGCCAGAATCCT	LETIYNDACLRFFNSMSDHETKEIEKLYL			
				TTGTAGCCGCCTT	GCCCACTATTCAGCT	KPFIPKRLPWPRYSKQIDKLVASRTIKKA			
				AGCACCAAATTCT	GCCGTCCTGGCCTGG	QTPPLMPPLNPSVTVTNIKLRIESKLME			
				TATTCGCCAAGTT	CCAGAGGGCCAGTCC	FFSKAPALIADNCHQTYNRITEKQDVLR			
				TATTATTAAAAAA	TCGACAGCTGTGCCT	AKEASIARKNRQKIQSGH			
				TTGCTACCTACTT	CCAGTAGACGCGCCT				
				GACTACTTATGCT	CCTGCACTCTCTGCCT				
				AAAAGTAACATA	CGTTCTCAGCGTCTG				
				ATCTTAAATCTAT	CACCATCTCGTCCATC				
				CATACTACAAACT	GTCCTGTCCAGATTT				
				ATCTTACCTAAAT	CCGCTCGCCTGATGC				
				CGCTAAAAACTAA	ACCAATACCTGAAAA				

				ATGTTAATCTAAT GTCTTATTGTATA TTTTTTATTAGTA TTAATCTAGTTTT AAGATAATTTTAA TCCCCCGCCACT GCTGCTGTAACG GCCTTTTAAATTT AAATTTTGTCTAT ATACCCTTGTAGG TGAAAGGGCAAA AAATAAAAAGCG GCCTCAATCGAAA GACTGAGGCCAG GCGGATCAAACG CCTTACCTTAAAA AGTGGGGTAATT CAAGGTCATGTAA ACGCGGCAGAAC TCTCCTTGAAGTG ATCGGCTGACAA	TAAAACAAATCTAAA AATAAAACCAACTAA AAAGTACAAATACTC ACCTGAAAATACACA ACAGCAAACAGTAGC TCCTTGTCCAGCGTC GTATGCATCGTCTCA TCTGCGTCGTCCAGA ATCTCGTCAAGCTAG TCCGTCTGCAGCTCC TGTATCTGAAAATAT ATTAAAATTATTTTCT AAATCAGCAATAAAA CACAAAAACTCACCT TAAAATCTACGTCAG AGAGCAGCAGCATC GTCTGCGAATCCTCG TCCAGCATCTGGCCA ATCATCTGGTCCATCT CGTCCTTCACTAGCG CCTGCACCTGAAAAA			
--	--	--	--	--	--	--	--	--

				GAAACTGGACAT	AAACACAAATCGTCA			
				CCTGGCGACGGG	GAAAAATCTCTCAAT			
				TGGTCCTACTCAT	AATCACGTAAACATA			
				TTCCACACGACC	CCTGCAAATTCTCTTC			
				GGCGAAAAAATC	AGTCACCTGCAGCAA			
				CCTCAGCCATTGA	TACCAGCTGGTCCTT			
				CTTTGGAGTTTAT	CTGGCCCTCTACCTA			
				TCTGGAATAATCG	AGGCTGCATCTGAAA			
				AGGACAGATTAA	AAAAACAAAATATCA			
				CTTCACAAACTAT	ACAAAGAATCAAGCT			
				ATGCGAACTATCC	AATTTACCTAAATAC			
				TCTGACCACTTAC	ATATCTAAAATCCAG			
				CTGTTTATGTACA	CCGCAGCCGCTAGCA			
				CATCACGATGGAC	GTTTCAACTGGAGCT			
				CAGCACCCAAACT	CCTCCTGCATCTTCTG			
				TACTTCCTGACAA	CTGCACCACGTGGCC			
				TCGAAGACTTTTG	CGTCATCTGGTCCTA			
				CCTGTAGGATCTA	AACGCCATAAAGCTG			
				ACATTAACTGTT	CTCCTACACCTGGAA			
				CCAAAACACCCTC	AAAATATAAAAATGC			
				AGCGACTCCCTCC	TAAACTATGCTACTA			

				ACCTAAATATGGA AATCGAATCTGGA TTAGACATTGAGG ACGCGATTAATAT CTTCACAAACAAT AGTCACCATGCTG CAGCCCTGGCCAC TAAACCTCTACGA AGGGACCTGCCA AATACGAATCGTA CCGAATCGGAGG AGCCATACCTATA AATAAAAAATTGC ACAAATTATTA ACAAAAAATTAA ATAAAATTGATAT AATGCCTCTCACG AAACTTTGAACGT AAAAAACATTAA	AATCAACAAATTACT CACCTGCCATGCAAA TGCCGCCAGTAGCAT CGCGTGCAAATCTTC GTGCTGAATCTGGCC AATCTCCTGGTCCAG CTCGTCCTTCACTAGC GCCTGCACCTGGAAA AGAACACAAATCGTC AGAAAAATCAATCAA TAATCAGGTGAGCTT ACCTGCAAATCCTCTT CAGTCACCTGCAGCA ACACCAGCTGGTCCT TCTGGCCCTCTACCTA AGGCTGCACCTGAAA AAAAACAAAAAATCA ACAAAGAATCAAGTT AAATTAATAAATAC ATATCTAAAATCGAG CCGCAGCCGCTAGCA			
--	--	--	--	---	---	--	--	--

					GTTCCAACCTGGAGCT CCTCCTGCATCTTCTG CTGCACCACGTGGCC CGTCATCTGGTCCTA AACGCCATAAAGCTG CTCCTGCACCTGGAA AAAATATAAAAATGC TAAACTATGCTACTA AATCAACAAATTACT CACCTGCCATGCAAA TGCCGCCAGCCGC				
	CRE- 1_FoC a	GCF_00221 7175.1	Folso mia candid a	GGTAACTTACACC GGAGATACTACTT ATCTATCGTTGGA AGTCCCTCTACTT TCACTCTTGAATT CCGCCCCCGTCTA TATATTTTAAGCG TGGAAGGAAATT GCCAGTATTGTTG	TAAGTGTATTTCATA GTTTATCAATGTTTAT GTTAATAATAAATAA TAC	MPLFKSLDIWKYVNLTLPLPPSSSSPP LPPSSSPLPPSSSPLPPSSSPLPPSSSPLPP SSSPLPPSSSPLPPSSSPLPPSSSPLPPSSS PLPPSSSPLPPSSSPLPPSSSPLPPSSSPLP PPSSSSSRPHPGPRKKRNKQKKRTGPRK CPPPTSSRPLPHTTEGLNPVSDDLFCPV SGCPGPSSSATDEWIECSACLLWYHTSC LLLQAIPEDDWFCANCPQIHTQIGPQV NSNTPSPANQSTLGTHSQIIQSSPCTNP RASQPISTLSTPLSNFISSKATCTSLTLEW	SSF 566 72	619	816

				GATACAAGTCGTC AGACCACG		LNNNIVPGGIIKNNLLHEAYRQDSLGIK PLND AEIGKLIKNSFSNIRTKRLRTEFHG ASKYHYVGIQLKSEQVPETVFDFKSISSL RQNCPTLKQCPKACRIPVAQSIKSAIDK CITENTPEAWHLLTFFPYVTLQCDSRAN IKNGQSLNNKIRNNLSIFESLDKNTYQD YLISLSSKNSVFDKSGSKINDIGPKVISKV SEGNIRGAIRLLTSDDVVAPNSSATLDTL VEKHPQQSLDNFTVLDEPNIPPFTTTTE QVISAI FSFSGSAGGLDALRPQILKDLV SQHTGDVGTQLLES LTKFCNFMLAGKV PLDICPFY GASLTALKKKSGGIRPIAVGC TYRRIVSKICVKVMNHVFQDYFSPVQV GFNVKNGAEAGAHAAARRFFSFPHQSP TAFLKIDFKNAFNMVDRGKLLSSVHEL PTLYPYINQCYRFSSSLVWNSSLIDSCRG VQQGDPLGPPLFCLTLHPVISNLGSEFN MWYLDGTLGGNPTTVLDDFQHIISCA NEVGLELNFDKCEVSVLGPQTNTQCILE SFQSVAPGISCM TQENAELLGAPLTLESI EHVFRKKIHAFTNMAGRLHSIPSHIAYY			
--	--	--	--	---------------------------	--	---	--	--	--

						LLKHSLAIPRLVYLLRAAPIFKSPHLLEEF DSLLHNALEQITNCQINPLSLEEIALPIKL GGLGIRQSKTLATPCFLASSFATTDLVSLI LPSYVTALSDSLVTEAQDLWKSISCEELP RSTITSFQSSWESPIFNYVSSRILESSTTPI NTARILASRSKESGAWLQALPSRNLGTL LDDNSFRIAIGLRLGLNLCPHKVCVGA QVDELGLHGLSCKKSAGRHSRHTQAN DIIKRALISADIPSIREPLGCCRTDGKRPD GLTLIPYKQGRSLIWDFCTDTFAPSYLS HTLRHAGTAAKLAETKKHKLYSELKTNVI FVPVAIETSGVWGTEGLKFIQDVGRRM KDVTGPMSTTFLIQRISLAVQRGNAASI MGAVPSQRGLEELFYILNDRSFYLQKSC KGPNGDLGRLSGSSKGFPNPPTSGFSST PQAN			
	L1- 16_G Ma	GCA_01227 3815.1	Glycin e max	GTTTTGGGAAGTA ATCCCAAGGCCCA GCAGAGTGACAA ATCAAAGGAGAT GATGAGAAGTAA	GC	MIILSYNSRGLGRGVKWA AVRRLIVKHK VDLVCLQETKRENFNKSICQAIWGDST GHWDFVPSVQAAGLLCMWNNSTFV VERKEKGRSFLMLEGTCCSNNQKLIFVN VYAPCDLVGKKALWDELRLKASNPSG	cd0 165 0	497	762

				TACCAAGGCCCA GCAGACTGAGGA ACCTTTCGTGGAC TCCAACCCAGCTT CTAAAGGTCACCC TCTGTCTTCTTTAT ACGAGGACCCTAT ACAGCTGCAATGT GACATCTTGAAG GAAATGGGCTTG TCTTATGAGGAAG AAGATAGCAAGG TTAAGGGGCTGCT GCTGGATATGGA GAATAAAGATGA TTTGATGGCAGCA GAGAAGGGAATT AAATCTCAGCAAT T		MWCFLGDFNSIRSSQERISLSQRRADP YDIAAFNQWIDDMELQEIKCYGNSFT WIRPNGCVKSRLDRFLVSQNWLSLWP ESCQTVLQRNLSDHCP SILQTNMVDW GPKPFRVFDWWLQQKGYQKMOVREA WKKDQQGGWGGIALKNKLKLNKSVIK CWSKAEGNINPKKVLNIQQKLNEVEDL AAHRNLTDQEIKDRTLLQQELWNASN AFESLLRQKSRRWLKEGDYNTAYFHK YMNFRRAYNNIPGIFYGEWIQQPDAV KNEAANFFHNRFTQMHSRPTLDGVH FKSISQRQREQLTVPFSDQEIKQAVWSC GGDKCPGPDGLNFNFIKEFWDLNPEF RRFVDEFYAHESFPRGSNASFVALIPKKS HPQSLNDYRPISLIGCMYKVIKLLSNRL RSVMDELIDERQSAFIKERHILHGILILNE VVEEAMRSKKPAMIFKVDFEKAYDSVS WSFLDYMLLRLGFCQKWRRWISACLH SATISVLINGSPSKEFNPSRGLRQGDPLA PLLFNIIGEGLTGMMREAVHKNLYRSYL VGKKREPVNILQYADDTV FVGEVDWQ			
--	--	--	--	--	--	---	--	--	--

						NVHAIKALLRGFELASGLKINFAKSQFGI IGGGVNWGLAAANILNCSQMNYPFHY LGIPIGANPSSLMVWEPLITKFKAKLSK WAQKDISMGGKVTLINSVLNALPIYLLS FFRIPQRIVQKLISLQRKFLWGGDNTHN KIPWVKWADICLPKNEGGLGIKDISKFN VALMGRWIWAFASGQQQLWVRVLIS KYGGWPEFQHDTDKRGFSHWWRDLR KIFHQSDHNIFKHQLTWKVGCGETIKF WTDKWLGEDYTLQKYNQLFLISRQQ NSLISNMGEFNHDSWEWNLRWRRNL FDHESDL			
	Nimb-3_LG	GCF_000327385.1	Lottia gigantea	GCCTCGAACGGA ACAGGATTAGACT CGTATCTTTAGTA CTGTCCGA	TAAAGTAGCACATCA GAAGAGCGCCCCGCT CCTCTGTAGGTGTGG CGATCATTGGGGACT TAGTTCTCGATAGAT CGCCAAGAAGAAAC ACAGAAGAAGAACG CAGTATCAAGACTAC AGAAGAAGATCAAG	MDRSLKILQHNLNGLRARLTELKFLSTA RYDVLCFQETKLYPGVKDTILPGYRLSRL DRPGAHSKCFGGGLAIYVRTELSYQTLV LHSTEFVNGVSLSLADGSKIAIVNVYLP PGKNVIEPSSYRNLCEQISHRNVLFCD FNSKHNAWGSYKQDERGEGLLQFIED NDMCILNDGRPTLIHNSSGVMSHLDLT IATSNIASNCEWQVSSTETLGSDHLLIDI TYRGQIKREGTVGKPKWRIAKADWIKF	cd01650	506	765

					AAGAAACCATGGGA GCAACCCGTTCCCTA CTCAACGTCTGCGAT ATGGAATTTCTCCGA AGCGGACACCATCCA CCAACAAGAAGAAG A	TENCSQWDVKDIQDADVDVYANNVT MVLMEAAKDAIPRTSGKNRACPKPW WSDACQNAVKLRRKALRRVQRNLNTE TKRTYLEATENARKVISEARTSHWHDFC QTLNQHTQIRDVWAKIKSMNGTKSTR EIPVVTYGDRVGESDRQKADLMAESFA DASSDKHHSPEFMNRKTNFAVDHSNV LTHIDENDPLNLPFTSDELQKALSSKKDS AAGEDEVKYSMLKKLPPEASKILLDFFN ESWTNGHVPTKWKTASIIPILKAGKDP NKTESYRPVSLTSTLCKILETMVTHRLRY YLESNNLLDKNQSGFRKLRSTADHIVRL QHAMSKTLNSSGFLCAVFVDLTKAFDL VWTDGLMHKLRQLGIGGRMHNWIAS FLADRKIRVTIGKSLSEEFSLENGTPQGS VISPTLFSVMVNDIRNPKYLSNKCQIAM FADDSTFWGNGTLYTNLQMRIQTCLD RFAKWANDWGFRISKAKTVGMLVCKD PQLTDMRLVLDGERIQFVSETKLLGVTF DRKLSWEPHIKNITNRCQIVLNLMRNIS GNEWGANVVTLMMIYRALIRPILDYGS			
--	--	--	--	--	---	--	--	--	--

						QAFMTARPMLLHKLDVIQHKALRLATG AGIGIKTESLLQDVGEVPLQIRREELALR YWSKLKHLGESSPTYLETARFRHAKKTIS FGQSVRDSVRKYGLDRLPIKPLVFSQVA PWTVSPPIVDTSLASRLDKKRDDPCIAR MIALTHMDARWSESIRIYTDGSHNPDD GTTGLGIYVPSSSTRISKRLPNNCSVYTS EMLAILYAVKYSAGHYHSNYVVLSDSLS SLTSILNDSSNSQAELVDEIRQACSLVHS QGKSIELAWIPSHVGIQGNEIADELARI GLGLSEIDRIDYSLSEIRSMITKSIKSTWE ATLIDANAFRFRLNDSSISKVFLHHPNR QICRTVARLRMMGEVRNHKYELKLCPC GEAGTAEHLLIHCEQLRELKSLVEIMK RYRLEYTLKNILSHTGEGMSTICETVYNF LSESRMIT			
	CR1- 6_NVi	GCF_00919 3385.2	Nasoni a vitripe nnis	T	TCAGATATATGTATC TCCGTATTATTATTAT TTTTGTACATTCGAAT CTAATGAACTATCTG TCACTTGTTGTATATT	MTYPREGCTGPRDGLSKQEELTAGYSG FSNAWTLTRIIMSHRAWTVELFLNCTI GPIDPSTSLCTTFLSKATTSRCHYWSVPL LVLMLMKSLDAVFPGPQETALFLVRFG SSDDASRVLKNRKLPPSSVEATADHTAY	cd0 165 0	711	962

					TTTCAAACTTGTATT TATTATAGTATCTTTC TAAAGGACGCGTCA GCCGTTAAATTATTA AATCAATAAATATAT AAATAT	QRDLYRKLKLKASLHNTSHPD ^{SL} KKVVY VKGTPTLV ^{DL} RRTSTAPLSEKPSKNLQA SETFKKLAGILKPVN ^{KL} SLYYQNAYSMY GKLHLFKAKLFQ ^{LS} ILPDIVVITETWLQD CVLSSELGMQQFDVFRCD RD LSSHGV QRGGGV ^{LIG} VKKFLQAELVHVDN ^{VIE} Q LIIRISLPSHKLIVAAS ^{YL} RPDSSVQVFSSH VSNLEDANIKFPDHRM ^{VLL} GDFNLHHI TWSSDTSAYSPMAYLSPALRSAADIISG YSTSMDLVQHFKHHPKKGYS ^{LD} LLFAQ KDFISPMDFNEDILPCDDHHVPGMFSC NISVASSASCNLK ^{FK} RKFYSADFEAIRA RLSEEDWATVLYSFSDDPDSGIARFYEIL NGVIDQHVPITKQSLSSYPKWYDSELIN AICKKKIAHSTWKSSNSKRDEIEFKRLRA VCVRLSR ^{SKY} RAYISSVELGLKRNMR ^{TF} WSFIKNLRNDSSIPSKMYHGPTRAASD VEVSNLFSGYFSSVYRTDSLREFESNDEP LVSLSNIVLSIETLRSIVANLDDNINGGP DGIPPYFIKRCWSLLERPIVHIFESMLRS GYFPTFWKYSFVVP ^{IK} NGDRHSVANY			
--	--	--	--	--	---	---	--	--	--

						RPISTLSCIPKILDAYLASELSNSLLSKVAQ QQHGFISGRSTLTNLLVFNDFLSESLCN HEQTDVYTDMSKAFDSVNHSRLISKL FNFGIRGKLLDLLRSYLTNRSQAVRVNG AVSSPVDVPSGVLQGSHLGPLLFCLYVN DLVPKLSAAQVLLYADDIKLFMRVRSAD DSVKLQRDLDTLVAWSVENGLSLNANK CSVISFSRSKSLLLFNYCINGSRLQRVDV VKDLGVIFDSSLTFITANQQTDVVVSRSL RMLGFIKRSTVDFSDPSAIIYLYKTLVLP FLYCPQIWRPHSQYLINKLESVRHSFLRY IAFKMGKPLHRFCHDYSEIAEGCDILTV DSTFYLHDCVMTFKILRGLTNCERLSDLF VSRTVAYPLRNLRDLREGTFSSNLGFYSS INRMKRQWNSIPNNVQSKESVGQFKS SIRSLSLAY			
	Daphn e- 15_NV e	GCF_00020 9225.1	Nemat ostella vecten sis	ACATAAACTCTGA AACTACATCT	TAGATATGATTCTAC TTTATTTTCATCAAATC AGCCACATGTAAATT TAATAATAAGGATTC CTTAGTCAAATCAAG	MESLFIEIINEKSKNTLIGVIYRPPHSNVN DFLNEMELLTTKLNTENKKCYILRDFNL NLINIEYHELTRQFLDIMYSSSMIPTISKP TRITSHSATLIDNIFTNNINMQDSFSGLL CADISDHIPIFHITKIGKTKQNKEKRNM	cd0 165 0	412	668

					AACATTCATATATATT TTAGTAGAGACGCTA CTTAGTCTGCGACAG GGGTCTACACCTCAT AAGCATTGTGCTTTT AGTAGACTCCTCATC ATCACTATTTTTTGCA TTATTATCATTATACT ATTATTTGGTGAAAT AACTTGATTGATTG ATTGA	TYKRIMNDSCIDNFLGEIDQINWDNVV LKENVNDSYNSFLQNITYIYETCFPLKKF RNKKINSKPWITNGLLSIHKHKHKLYKKS QELPSESNFNFKQYRNKLNLSLLRIAKR DHFTDSLEKSKMDAKNTWKILNSLLKK NIKKPNYPDIFKNNNQEISNPKDIADKF ADYFTNIGPNLSKNIENTNIDIKSYLKGN YSNSLLFRLTTPDEIKEITLSLKNKNSCGF DEINIRVIKAAISILSNPLSTLINQSIMAGI VPDKLKIAKVIPLFKSGEKEFSNYRPISIL PCFSKIYERAIYNQLISFINKYNILNNNQY GFRDKHSTVMAVCDFIEKISSIDKGYN TIGIFLDLSKAFDTINHNILLQKLQFYGIR GIPLLWFQNYLSNRKQFVNYKDTKSEY DITCGVPQGSILGPLLFLIYINDVPNATEL FSFTLFADDTNLIFKSKSAEDMDEILNSE LFKLTEWFKTNKLSLNTKKTNYYIFRNSK RKTHTTKIEINNEPIQMVNETKFLGLII DSQLNWKSHINHVRTKITKNVGIMNKL RFYLPKFSLKTLTYFTLVHPYIHYGNINWA SNYDTSCLKRIFLLQKRCVRIITNSHYLSHT			
--	--	--	--	--	---	--	--	--	--

						EPLFKEERLLTVYDINTLETALFMYKFDN NQLPGTFDNFFKRNNTIHSYNMRNSC DYRTTLRKTRRGQFSISYRGPKIWSNLT CQAKQCRSLNMFKKIVKDNILSQKAAID L			
	L2- 8_OL	GCA_00434 8135.1	Oryzia s latipes	T	TAGTGGGACTTGTTT ATTTTATTAATTTAAT TCTTTTATTGTTTTT GCTTTTATTGTTCAAT TGATTTTATTGTTTTA TTGTGATTGTTTCTTT ATTGCATCATATTGTT TGTGATTTTATTGGA TTTTATGTAAAGCACT TTGGTTGCCTTGGCT TTTATAAGTGCTATA CAAATAATAAATTGA TTGATTGA	MRKYNVSFHCYADDIQIYLPKSGGAE GLDRLQNCLADIKSWMELNFLCLNKNK TEMFVFGKSDLLTGCCSAVGPLDPSDA PFIRNLGVNLD RHFKFDKQVSLVVQSSF YHLRLISKVKPYLPPHALQQVIHALIFSRL DYCNSLYSGMDQSLLRRLQLVQNSAAR LLTGTRRCEHITPILRSLHWLPVSFRISFK ILLLVFKCLNGLAPQYLSELLHVYRPCR V LRSEDQMLLVPRTRLVTRGDRAFSVA GPKLWNA LPYSVRSARTVSNFKTLLKTH LFTQAFNSSERI	IPR 000 477	3	89
	Rex1- 5_OL	GCA_00434 8135.1	Oryzia s latipes	G	TAACCTTTGCACATTT TTTTTATTATCTCTTT ATTTAGTTTTTTCTAT	MKLNQHKAAAGPDGVSPTLR TCAHQL SAVLQH LFNLSRLRLERVLLWKT SCLVP VPKKTSPSHPGDFRPVALTSHVMKVLE	cd0 165 0	51	301

					TTTGCACATTTCTTTT GCACATCTCTTTACCT GCACTGTTCTGTAAA TAGCTCTTTAAATTAT AATACCTTGATGTCT GCATTTATCCTACATT TTCTTGTTCTTATTTT ATTTATATTATTCTTA CTCTTTTTTTCCCTCT GCTGTCAGTTGTTAT GGTAACGAACAAATT TCCCACGTGTGGGAT CAATAAAGTATTTCT GATTCTGATTCTGAT	RLVLAQVRPVVSSSLDPLQFAYRPRLGV DDAVIHLLQRTHSHLDGGGGMVRITFF DFSSAFNTIRPSLLGDKLRLMGAGASTV SWITDYLTDRPQFVRLGGVLSDVVLSG VGAPQGTVLSPFLFTLYTSDFQHNSC HLQKFSDDSVVVGCSGGREEEYRRLVD GFVEWAGENHLLLNVDKTREMMVIDFR RKRTVLQPLCILGRDVEAVEEYKYLGVTI DSRLSWRSHSTAVYKKACSRLYFLRKLR SFNVCSKMLGIFYQSVVASTLFSSVVCW GSSIAVSDTNRLNKLIRKAGSVIGCKQET FEAVMERRSLNKLKLSIMDNPEHPLHPVL DKQRSSFRRRLRQLRSKTARYQKSFIPN TITLFNNSQCRR			
	RTE-15_Psi	GCF_000230535.1	Pelodiscus sinensis	CACTACACCTTTTT TTGTCAAGGAAA ACATAGTGATGA GAAACGCCTGCAT GGAGTTGGTTTTG CCATTGCAACTC	TAATGCTGCATCCCA CCGTCTTTTGAGACA TATGGAGGCCATGA	MTPSAESKRIMTLRLFTESGYINIISAYAP TLTTNSADKNHFYESLQKVVDSSISSEQ LFILGDFNARVGIDTSSWPICLGHGIG KMNENGQRLLEFCSINELCITNTFFPGK ECHKVSWHHPRSGHWHQLDFVLVQR KDLNCVKSTRTYHSADCDTDHSLIVSKT KLKAKKFHRTTLNKKPKINCMNTRNSA	cd01650	445	722

				GCTGCTCAAGTCA ATAGAC		KCEDFAKAFAKKWEENAHVESVEETW QTLCSAIYEAAREAFGEKKSSNKDWLEE NADRLMPLIKQKNEAYLEYIKNPSDSA AKLQHAKGLLQREIKHCINQYWIDICSN IQVASDTGNARAMYEGLRKALGPNITKI APLKTKEGISITDKSKQMARWIEHYSEL YSEERNIMSETRDFIPTMELMQELDAV PSKQELIDAIDALSSGKAPGTD CIPAEVL KH NKDCLIPPLYDLLLKCWTKGEIPQDM KDANIVTLYKNKGDKGDCNNYRGISLLS ITGKAFAKVILRRLQLLANSIYPESQCGF CTGRSTIDMIFSLCQLQEKCREQNMPLY ISFVDLT KTFDTVSRQGLYLLERIGCPPK LLQLIAFHEGMRATIOQFDGYTLD SFQV NTGVKQGCVLAPT LFGIFFSVLLSCAFK NSHGS IYLHTRSDGSLFNLARLRASKIK KVLIRELLFAADAAIVAHNETQLQDLMA QLSRTCQQFALTISIKKTVVLAQGNTTSP CIMLNNNPLEVV LKFCYLGSTITSNLSLD AEINIHIGKAATTFSRLTKHAWANSKLT I RTKILIQACVLSILLYGSETWTTYASQEK			
--	--	--	--	-------------------------	--	---	--	--	--

						QLSSFHLHCLHKILNIKWQDKVPNEKVL ERAGMSSMGAILHNRRRLRWLGHMKR MDNNCIPKNILYGELAEGTRPKGRPKLC FKDYCKATLVKFQIDIDSWENLAVDRNT WHQHRLRQGQKHDKTLAENQNKKRM KRKEQMITTNTTTAWSCIYCSRTCLSHI GLLSHTKHCKKRPV			
	Utopia -1_PS	GCA_00984 8525.1	Phytop hthora sojae	AGACGAGCAACG CGCTGGGGCCCA AGACCTGGCATG AACGACACTGCCC AGGCTGTGAACG ACGAGCACGCTG CTAACCCGGCAGC GCACCGGCCTCTG GGCTCCGCTGCAC CGGTCACCGGTG CAACACGGTGGG GTCCACGTACGG CGCGATCGGGGC AGCGGCTGTAGC	TGAAGCTACACAAGC GCGAGGCTCGACAG CTGGACCTTCAGCTG TCGAGCCACTGCATC CGCGCCAGCCACCGC ATCTGGGGCTGGCAC TGCCGGCGCCACCGC GAAGGACAACGGAG CGGCAACGCCTCGCG AACGCCGCGTGGGTC TGGGGGGGCCCCGC GGCGCACATCGCAG GCTCGGGCACAGCG GTAAGCTGGTCATT	MDVDGGPAMPEPWVLRFDGACRRNP GPGGAGAALFKPCGTVAWTCSHYMP NSSETNNTAEYTALLGVQSAVHHGAS HLEIEGDSHLVVAQVKGTAFACRNPLR QLRNRVRHALRAVTSLTCLKHIDRKANA HADRLANRALDLKRS LAECGEHQGAM ESCLHMNPAAQRQREQPAPPARPACA PTRAEPASDHDEDIDAEIAARDGGEAFP TLPIGPGTAPARQPRLRLRLTEDEQEA AASALQAMAEELACKIEDADSWTSGD GYISAIPGRIRQVLQPFTAAQPHRPTL QQQRQRPPRVTRTQREHRLDEALDEM AAVQQRPTSRSAVRRARRRVGRIRAS MRQQQLRHDFARNESKCVEDILRAASA	cd0 165 0	499	763

				TCGCCTGCTCACA	GACCAACAGGGCACT	ETAAEEHPETCPIDSGTLHEHFTAVNSP			
				GGCCTACCCACG	ACCCAGGTAGGGAA	RINFLPDEACGALFREAMADVGTQER			
				GCACCAGCACCA	CCGCCCTTCAAAAAC	RSALTDELTVDEVEDQLMQAATNSSPG			
				GCGACGCGCCGG	CCAGGAAGACACAA	HDGVGYDIYKKFAAQLVPLLHAAFQFC			
				CCTGCTTCGGCTC	ACACCCTCCCTTTAGT	WRHHRVPALWKVGFVRLIHKKGDPDT			
				GGCGCTGCCCAG	GACATGCATATTTTA	PANWRPICLQTAIYKLYSGLLARRLSTYL			
				ACCCGCCCGCTTC	GCCTACATTTTCAGTT	EANGLLPMAQKGFRAVNGCHEHNFVA			
				CCCCGCGGCCAC	ACGGAGAGGTTACTA	TTLTDQTRRMRRRLYQVWYDLRNAFG			
				GACGACAGCCCC	ACTGGTAAACACGAA	SVHQDMLWYVLRLLGVERAFVERCDDI			
				CGATGCG	CACACAT	YEDSYFVVGNAADGATEPVRQEVGVY			
						QGCPLSPLLFI AALVPLLRALEKLDGVGV			
						ALADGVRPCTTAYADDLKVFSDSAAGI			
						ARCHAVVEQFLEWTGLQANPGKCAFL			
						AVTRNARGNPAHDKDMKLSLHDEEVS			
						SIKLHDSYRYLGVGDGFDHVRHRLQLEP			
						KLQQIKREAVALMQSGLATWQVVKAL			
						KTYVYPKVEYALRHLRPLQSQLQSFDREV			
						VAKGLRHLLRLPRSATNEVLYAPTSSGG			
						LGLQPLVEMHRALQIAHAWQMLHSDK			
						PAIREVARAQVCQVARKRHRLREEHW			
						RGRDDELVRCFLNSELAASPHAEALRRH			

						GDIGSLWSDVQRWLRIYHLSLVVQDDR NGLDPLGLRVPHHAKWLDHKSVLRHV KLHLKIRHQTRWKGLADQGKTVRAHG GVGAKFVSTGAGLSDDDYRFGVKARLN QIDTNAVLRKRRLRSHKTCRDPTCSSAE TLAHVLNHCEPNMDAIRQRHDDALEQI GSKIRNALKRGESTAELRLNQTVPEYTG AALRPDIVLRNVAKKMVIADLAVTFEE HAAGARHSSLQLSHDHKTLKYQPIVAEL QLKGWQVQTAAIVYGSLGSVQPSNSTP TRRS			
	Utopia -2_PS	GCA_00984 8525.1	Phytop hthora sojae		TAGGAACCTAGCCGA TTGCAGCAGCACCTG CGCCCCGCCAGCACG CACAAGCGCCCCAGC CAAGACACCAGGGG CAGTCGGGTGACGG AGACCGGGGGGACT CCGTCGCGCTCCGGC CGCCGCTAGACGGCA ACCAGGCCCATGCG	MDVDQGRVRPRQRRPSRRPSERRPD GKRRRLNSEDADAQKPAEHLQLADE DGAENHTPVHTPSAASAHQAPVLAVY VHNAERFTCTLCTFTAASFSSLQRHRTT RHRRAVLQDKFLAGCACGTPIASRPAA AKHAQACAGRSTALPATAPPVAGTSSP TDAAATATVAAATVTPDSPQPATPELA APPPLASPTHPEVVEVQEQPSQGRWG PALPRALVAERVAARLGAIPAPRWEPL PRELVSTRIAARLGEVPAPRWGPPLPRR	cd0 165 0	912	1175

					GCCGACAGGGCCAC ACCCATGTAGGGAAC CGCCCTCGGAAACCC TGCACCGGCTCACAT GGTCCGACCCCCGA TTGGTTCAACCCGTA CACCAACTACACACC AATTAGGGACTCCCA GGTAGTCACTACCAG GACGCAACGGACTA ACCGGCAGGTAATAC ATTAACCCTT	MVASRIAARLLPPEPTADGETKEDTPA DGADLTARDANQENNPAAAPDRTATA AAMDMDSGADGEWLLRFDGACRANP GPGGAGAALFKPDGSVVWTCSHYMPS SSETNNTAEYSALLLSRAAADHGATSL RVEGDSTLVIQQVRGIFAARNAILRRLR DQVKVELARVGSFSLHHIDRQANAHAD RLANRALDLRRTMMECGAHTAGDGCT NTTLDPMPLVPPQPTPRAPPPTPAADT NIEMQEPEDEDLADIDDGEVYAAMRIG PDAVPQRRPRLRLRQLTEAEDDAARE MVERLGATLAAKITDASDWESAEGYIT ALPHLLYDKLQPYSQAQRRPEPRPQRQ RQQHQGSQCPRQQRQQGEQRRQRPL HDQQQDRRSLEGDAHGDNERDGDAS QQDSARRRQCRRRGRSGAAKRRRRRR PPRVTRHHREHRLDEALDDLHAVEQAS PGDRTAVRRARRRVGRVNSAIDQQHL RHRFDTDEKACVEDILAKARAQQETPR TAASSAGTASDPAAPAAGATDDGTCPI PGDELWRYFDGVNTPSQEFAPDAPAG			
--	--	--	--	--	--	---	--	--	--

						AAFRSAMTHLPAASRFRELLTEAPTQD DIESQLQHAHGTSSPGLDGVGYDVYQH FAAQLLPVLATAFQKCWEAKRVPQSW KLGVVRLHKKGAREDPANWRPICLQQ AIYKLYAGVLARRLVRLDANDRHAPG QKGFRAVNGCGEHNFLAATLVDQARR KRRQLYEVWYDFRNAFGSVPFALLWD ALARLGVPEYVTMCKGLYESAAFVVG NAADGTTAPISLRGVFQGCPLSPQLF NAAISPLLALKRLPATGVPLSGDDCPG ASAYADDLKIFSGTEDGIKRQHALVADF LRWTGMAANPSKCSAMSIQRDRRGVL KADDLKLELDGSPIPALSMSASYTYLGV GDGFDHTRRRVELAPTLKQLKDDATAL LQSGLAPWQVVKAVKTYLYPRVEYALR HLRPFQQQLAGFDRHLARGLRHLLRLP ASATAQCFYAPVSRGGLGLLPLTELHAA LQVAHGWQMLNSKDP AIQRIARVQLR QIADARHRLDTQAWAGREEELCELLLN SQLGTSPDAPPKRRNGDIGSLWVDVQ RHLRTLGLQFATAPACADTGTAATLQ			
--	--	--	--	--	--	--	--	--	--

						LRVPHHDKWLDHRTVLRHVKLHIKHRH WTKWAAMRDQGRTARAHGGVGSSFL TRPRGMCEADYRFAVAARLNQLDTHS VLKRRRLRAHDRCRQPGCSRVELAHV LNHCPGTMDAVRGRHDDALKTIERALI ASSSGDRQDRVELRVNQTVPFLAGPAL RPDQLYNHTKKTAVVVDLAVAFEEQV GEDPDSSALARIAAHKRAKYDRIKRHLE RQGWKVHLSALVYGSLGAVAGGNRSV YTEHLGLLKRDAKRLDWQKKEEDREDK CSSWPHEGWVKAPGDRTAIG			
	L1- 2_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	TTATAGTAAGTTG TATACTAAAGAAG ATATTGATGTAA AATGCAGAAGAT TATATTTTTCTCA AAATATTTCTCT CTCTCTGACGATG AAAATAAAATATG TGAAGGTTTATTA ACACACAGAGAA	ATAAATAACTGTGTA TTATTGATGTTCCATT GATTTGCTAAATTGT AAGCAAAATTATGTT ACTGTTGAATATTGA ATAAACACCTCTAAG AGGCAAAGAGAAAA A	MKKNKTPGSDGITAIFYQVFWNEICDI LVAALNECYANEKMSDTQRKGIITLIYKK GDHRELKNWRPITLLNSDYKIIATVLAK RFQKVIGKIIHENQSGYIRGRLAACNVRL TKDIEYMTRNGKNGALMMVDFSKAF DVIDVTFIKKCLSKFNFGPVFQKWVTILY TNISSSVIVNGWLSETFSVLRGIRQGCPL SALLFVMAAELMANKVRENKEIKGLDI DDKGIDLKLLQYADDTLFYVSDINSFENI LIELEVFGIVAGPKLNREKTVLIWIGTAC	cd0 165 0	47	306

				TGTAAACAAGCTG TATTCGCA		QRWSLSCFGLHWTDEPVKYLGHHIAP DLQCALKIDWENKLEKMQKLLDNWRK RNLTFLGRLTIKSLAISQVIHLMIVDTIPS TFLIKLNSLIFKFLWKSIEKVKTFRCLKWL GMFLDNTTGNWKDMCKHWFDCCLGG LHLLNCNYTDYTLNLLKLRNVPHFYIEIL HAWYLINKYTSaecQTQKHVDENILWY NQDIVFEKKMLFYEDWYKSGIYLNDIF KNGTFVSVNELFIRLKTRRSKQFLFDYT KLQKSIPKIWINDKHRVPLVFKDHNFSIP HFVFGKSVKPVSLSSKQWYNMISFEQI KINKCCLYWEDVLQKDVKWSSVFNRNL IGIKDFRLREFNFKLLYNMLPVRNNLFK WGLSYDDLCLKCKKTEDVLHAFISCKAN ETFFAYLKDLVLRafnlRIEIDIVLLLKMN SQTEIDLIMTIAFWsiYKMIIFRNKYGYD NRDLSLIYLFKREMTNRIEINEHLKVQN MKFPANLLSYI			
	RTE- 3B_SP	GCF_00000 2235.5	Strong ylocen trotus	GGGACAACGGCG TCTCTCAGAACTG GTGTTACATGGA	AAGAAGAAGA	MESHNSKSKRNGVKGTSRNRGASLGR PSHVRQVGPDHRSASVRNRNIMRIGT WNVRTLQYQKGLDNVQEMNRMRIIN	cd0 165 0	549	830

			purpur atus	TATATTCATGGTC ACCACACTGCGGT GAGGAAGGCAAC GGGAAACCACCTT CACTATTCTTCCC AAGATATACC		ILGISETRWTGSGEFTVEKHRMIYSGGIR HERGVGVLLDPFTANCVLGFWAVSDR VLLVKLQGKPFNMAIIVVYAPTAESTDE DIDIFMNSQITMTAKSQCKSNEIVIAMG DMNAKVGRGRYGGTVGPYGIGDRNER GNRWVEWCEAKGMVIVNTWFKEHPR RIYTWKSPGDRARNQIDYITINERFKRG VKQAKTYPGADCRSDHIPVVATLQCKL KKMTKAKSIQKVDFEQLQNRDVRLRYT VKVQNRFDSLTDQGEEVTWESMRDIL VETAEETLPKKEKSKQRWMTEEILSK MSKRQKIKDRDGIEYRELDREIHDHCTK AKEEWLNKQCAEIERSFNANDPRVFKK VNDLSGKKLRSSKLGCIKAKDGSVLVEK EEILNRWTEYIGELFDDVRKEPPHCAGSI EGPRILLSEVRS AISMMKR NKAAGPDG VVMEMVEALENYGVMKLTEVINKIYDD GKFPDDLCKSIFITLPKKPGAVECGQHRI ISLM SHITKIILRVLLQRARNHLTPEISNV QFGFVKDAGTRNAVFLRSIAERAVEM QKDVFLCFIDYTKAFDKVRHDQLFEDLT			
--	--	--	----------------	--	--	--	--	--	--

						KIGLDTKDVKLLQNLWEQTACVRVET ELSDYINIKEGGSARMCHVSRSLQLLQV TTSTLKRQVRQGCVMSPDLFNYYSELIL RNLEQERGLKVGGHNITNIRYADDTILI AESEKDLQRLLDVVVRESDKRGLSLNYK KTECMVISKKGGIPCSLKINNNQIKQAT SVNLYGTTITDDARGEQEIKRRIALAKST FSKLENILKNKKLKMKTRLRLVLCYVYP VLMYGCEAWTLTSDLRKRLESCEVWFL RRMLRISWTERVSNDEVFAMAGTRRK LLNNIRGRQLSFLGHVIRKNGLENLALT GRIEGRRSRGRRRILWMSSLRSGISECGI DLGENELLHRIWDRELWHSMVANVSG YGT			
	CR1- 25_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	A	TAGGTGTTGTCTATT GTCTCATGTTATTACA CAACCTCCTCCCTCCA CAAAATAGCAGTAAT CACGGAAGAATCCTG GATCGATGAACAGG CTTAGACCTACAACA	MSGETRYIYDKADYDAMYDKIMEIDW NNEFWEMSMEKSW/SFFKTTLLDIFNRY TPKKTLNVSQRKKHWMNRETLEVRE KHRHFRKWQEMKSEKNREKYSSAEIEQ QRRCYSSNQARWQCRKAVSSYEKYI ASGVKKNARAFWSYVNSKITTRESVCN LTKPDGSKTETDQEKAENVLQDFFSSVFT	cd0 165 0	270	526

					TCGTGTTTTAAAGTA AAGTAAGTAAGTAAA GTAA	QEDLQDIPLPLDVEFDTPLDDIQITSGD VFELLRRLKPNKSPGPDKIHPRLKELAD VLALPISLLFQKSLESKQLPQDWKSANV TPIFKKGNKSDSKNYRPVSLTSIICKMM ENLVRKQLLDHLMTNDLISVHQHGFVV GRSCITNLLVLDQWTQILDDGGTVDAI YLDFMKAFDTVPHHRLLLKLESYGVRG NVLAWLESFLSGRQQRVVVRGKQSKW GKVLSGVPQGSLGPCFLFLFVNDLPKV VNSNISMFADDTKLYVRADTADLRKQL QEDLSRMEEWAEEKWQLCFHPDKCTL RLGKLPAEASTDYCMKKGTDSVTLSTK VEKDLGVLVDSQLSFKDHISQTVLKANR VLGVIRRTFMYLNTDTMPFLFKGLVRPI LEYGQAAWSPYRLGEQRMLESVQRRRA TKIIPGLRNLTQERLTKLKLPTLVHRRKR GDMIDVYKYVNGLYTVQEMPFKMSTE DRTRGHSRKLEKKYSRLDIRKYFFSNRV VELWNGLPENVVAAGSVNSFKNRLDS HWSKSVPVSLLYEPV			
--	--	--	--	--	--	---	--	--	--

	Crack-8_SP	GCF_000002235.5	Strong ylocen trotus purpur atus		TAAATATAACCCAAG TACTCGCAAATAAGT AGCCACTATCACAAA CTTCAATATTTAGTAA ATTACCAATTGCAGA TCACTGTATTTCTCAT AAATATCCTACAAAT TACAAATAATTATGT TTAGACACACTGTCA CCTCAATATAATATTT TTTCTTCTTCAAAATT TGTTTATGTATGATTT CTTGATGACTTGATTT TTTTGGATTTTTTTTG TTTACTTGCTGAGTTT AAATTTACATTTGAG ATTTAATCTGTAATTC TTTCTGGTATGTTTAC TACTCACCATGTGAT GATGTTTTCATGTAT ATGTATGTCTTATATA	MIISKPITRLFNLSLRLGVVPQEWKQAR VTPIHKGGDKDNVNNYRPISILPTIMKV LEKAVHSQLHDHLLLEFDLLNAQQSGFR KGYSTNTLLTYVSDNLHRNIDSGKLTGIV FLDIRKAFDSVDHSKLEKLYTYGVRGN TLSWFESYLRGRYQCTLVQRARSDLQV LQSGVPQGSILGPLLFTLFMNDLPNSVE HSKVMLYADDTAIMFSALNPKEIESTM NSDLDRVKSWLCHNKLSLNTDKTNFLL VGTPQKL RHIEAPIDIKIDGLQIRQVEHV KYLGVWLDQHLNWNEHMKHTSSKIAS RLALLRRVRKYLTIDS AKLLANALVLPYF DYCCTSWSN CNQGLRDTLIK LHKQMA RIVLQMNVLTPSSEMFSKLKWMNMEK RWQFQKCKLMFDIVNGYSPSYLQDMF TQSAATHRYSTRIADTRGLVIPKVKTEN GRRAFSFE GAMTWNNLPSEL RNAQTK QSFCTSYFRNM	cd0 165 0	26	284
--	------------	-----------------	--	--	--	--	-----------------	----	-----

					TCTATTTTATGTTATG TAGGGCCCCCAAGAA TAACAGTATTTTAACT ACTGGTGGGGCTACC CTGTTGAAATAAAAC GAAATAAATAAATAA ATAAATA				
	Daphn e- 19_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	ACCAGAAGGTGT TAGCTGCATGGAC TCGTGATGGTATG ATTATTGCCTCAA TTCAAGCCTCTGG TGGAAAATCGAC CTCTAAAGTTATA CGTTCAGATATCG ACCTCGAAAACCT TTGATAGTTATGT CACGTTAGCGATA TTATGAACGATCG AGCTGACATTTTA ATGACTTATATTT	TAACTCACCAGGCTC TTTACCCCCCCCCCTT GTTTTCCCTCTTTTTG TCATGCTTCCAATCGT ACAACAATTAATTTA CCTTTTATTTACATTA TACTGAGGGATGGG CTCTAGCATAAGGAC CACACTCCTTTTATAGC TCATCCCTTCACTCCC ATTATGTTATGTTATT TTGTTGTTGTTGAAT GTATGATATTTTATGT GTGATTATTCCTCATT	MPGSTVNSIINDQSFVEFLSPRNGENDF YGPSHGSYTCNEFNDLVNDSNDGLFSIL HINCRSLSKNYDNISTFLSSLTVQFTAIAL TEIWIKEYNSYYDIHNYTLVLKGRQNKK GGGIGFYVRNDVKFNRRDLDDINTPYIE SLFIEVVTANISAILGVVYRPPDQPIQSFL ESVENVITTVNTENRIFYMIGDYNINLFK VNCQSHVKDFIEMFMAVGMYPVHVG PTRVVNNSISLIDNIFTNNLENCSSGTILT DISDHFPIFCIYSQSTQTVTETCITMKRQ FSAKNINTFITNIGNTHWILKENDANSS YDCFLKQFQHVYEICFPMKCQSKKKKRF DKPWFNDELRLNRKKSLEYKYLNNR TAYRLDAYKKCRNKLTNRIKYVKYKYFK	cd0 165 0	550	804

				CTCCCTGGAGTGA GAAGTTTTTTACA TTTTATCTTTTGG ATTTTTCTAATACT CGAACTTTTAAAT GTAAAATACCTTT TGTATTTCTAATT GTTTTTGTATAC TGCTTGT	TAATTGTATATGTTCT TGTATTTTAAATGATG AGTGTAATAAACGA ACTTGAACCTGAACT TG	QSFVDAKGDTKQTWKIINGLLGTSKK NEIPSTLQYNNETIDNDTDICESFNEYFI NIGPKLTSTLDIPTGDFKQYLKKMNLFF FYFFFTPVTPREIINIVDKLSNKC SGFDS VDVKMVKRVIHLICYPLCAIFNSSLETGI VPDQLKIAQIPIFKSGTKTEMSNYRPISL LPVFSKILERAVYNRLYSFLNLHNILNNS QFGFRTGHSTSSASVDFTCQVSQVFDK SEVMMGLFLDLSKAFDTLDHNIILAKLF NYGIRGVLLNWFKSYLSNRKQFVKIKGN TSSMQTVVCGVPQGSILGPLLFILSIND MCNISDKLKYILFADDTSVFMSHKNIQS LQQDFGNEINKLLHWLSLNKLILNVNKT NYMVFTKQQKIDFNEVDLQVNNHEFK RVSVNKFLGIEIDSKLTWNEHTGAVCNK ISKNTGILNKLRLPKCVLRMLYNTLILPH LYYCSEVWAMTSQTNLNRILKMQKKAI RIISNSHYRAHSKPIFYKLKLLNVFDICEY HIAVFMYL CRNESLPCALLKYFNLNSTF HDYETRNASNFHLPQIKTTLMKKSISFD			
--	--	--	--	---	---	--	--	--	--

						GPKIWNSIPINIRNCRSLQTFKRKYKMY LIESYVL			
	Daphn e- 26_SP ur	GCF_00000 2235.5	Strong ylocen trotus purpur atus	CCAAACCGTTTTTC AGTCATAGGAGT CACAGAAACGTG GCTTAATGATACA TCTCCATCTACTTT ATGTAATATTGAC AATTATTCGATTG TACGTTCTGATAG AAAAACCGGAAG GGGTGGTGGAGC TGCT	TGACAAATCTCTTAA ATTAACACATGATAT CATATATGTTTGTCA GCATTTCCCTTGATATA GCAGGAAAATCCCCC TCCCATGCACCCTCTA TCCCATTTTTTTTTTTTT TTGGTTGCAAGTCAC TGTTTATTTTTTGTTA TTTTCCCCCAATTATG TTCTGTAATTTACTTT ACTTTACAATCGTTAT GATTTTAAATGTTTCT GGGCTGCAAGCGTAT CAAGCTGTCACGCTT AGGCTGTAGCCCTTC TGTATCATCACCTCTT CGAGCGTATGTATCT ATGATTTACTGAATG	MFISNLLTYKIRSDLSLSDGTDSCIEIN MAQRKNILICTIYRPPHTNVELFLDEFDR MLHKVNLENKDVYLMGDFNIDLLQED NTFHNMLRSNSFNAMIDKPTRLANNSS TLIDNIFTNVTNNYIGSGLLYSEISDHLPI FLICNHVRNYPEDEKNQAMIRKITAEKI AMFNDELLKVTWNDVVACTDTNKAYA LFWNLFSSLLNKHIPMTKKRSARNRIKQ PWITKGIFCSIKKRNRLYKYSLRHPGEES LNNYKSYRNKLTIIQRSRQMYFAKQFE NANGNASSTWNVINQALKKQKSKSKC HITKDNNTMVDSKEIANCFNAYFVGVG PNQANTIASNGVDFSQFLQTPCNKSLFL KPTNLIEILNIARSLKISYANGHDGISTYLL KQVIGNIITPLVHIFNLSLTTGTFPEAFKI AKVVPIYKKDDSSQVSNYRPVSILSSLSK VLERLIYNRLYCFLTENDLLNQDQYGFR KYHSTDALAQLYDKVSNALASRKHVIG VFMDLSKAFDTLDHKVLSSKLHYYGVR	cd0 165 0	424	681

					TGGAATGTAATTTA ACTTATGTTTTATACT CATGTAAAGACATTG TCACATTATTGTTATA ACCTATGATGATTCA GTAACCAATAAATAT TTCAACTATTTCAACT ATTTCAACTA	GVALDWFNDYLSGRKQYVSYESVDSSL LPMKCGVPQGSILGPLLFLLYVNDISNST TSLQYIIFADDTNVFCSNSDFRSLQRILN TELPKLSLWFRSNKLSLNLKKTNYIHFQS RKNATDKPIPEIVLDGEYLEKKQNTKFLG TIINENLKWSDHVHYVSKRVSRTVGLLY KLKHYVPLNILYMLYNSLILPYMSYCNIL WADSKVYTDKMLLLQKKAIRVCTNAGY RDHTDPLFARLKTCLKVTDINFLQKALFM FRLHNDSLPSHFPPMFCLNRNIHTYPTR QSNKYHLCNPKTTMAMKSIRHSGPDI WNSLSVDIRNLKSVYTFKKSIVKSILLSKY			
	CR1-30_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	GTGAGGTGATGG CTACAGGGAGCA GACATGGTCTGGT TTTGGTGATATAA AATCCACTTAATC TTGCACTTTTAAC TGCAGATCGTGTA TTTTATTTACCTGA TTTGCTTCTTGA	TAAGCTTACTAGCTT GGTTTTTACTTGCACT GGACTTAGCATATTT TACTGTAAATTGACA AGTGCTTTGTCCACA CACCATTCCAGCAAT TCATGACACTCCTTGT AGTCACTTGCCTCAC CTACCATCCCAGTAT	MKNYFLPPSSSRSHPLLLAILSIVCALKLL KSGNLYTQIADHRSASLPFFDLHATLCE PRSLHQPTRRRRTYTADIDNCKLLLFISSC NKWNRNTKLCPLDFRSTSLSHLTILLLLS GQVETNPGPSTSAPETFPCAICGDEV DNDPALLCDHCDWCCHISCVGISPD SYDILTCKSRFAWVCCQCSSTNIGSSSSSLQ GLNLSQTNRFSPHHDSSNEEILPSINPL TSTPTSSPGGRRQPRFHPPSRNRKKAFK	cd0 165 0	746	998

				ACATATCTTCAAG CTTTTGATATGAT TATTTTGTTCTTTG ATACCTTCTGGAG GTAGAATATTGTG ATAGTTTTACAGC CATTTTCTGCCAA GGAATTTGAACAC ACACCTGTAGTCC CATATGGAGGAT CCTGAGTTTGCTT ATTTTGCCAGCC TTCTTCATACCTAC ACTACAGGCAAC AATAACTGTAATG TATATTTTCTGACT GAATTATGCTTAT ATTATTCTTATGA TATAGAAGCAATT ACTCACTGGGATT TAAATCATTTACC	AGCTTCCTGAGTGAG CTGCCAAAGGGATTA CCCAACCAAGTACCA AGTAA	VMVVNCNGLKGNACKSSFHAAIAHHT PDFVFGCESKLDDSLSSYSIFPSNYSIYRK DRNIHGGGVFIAIIDTLIVTECEPFDSDC EIQWCNVQLANAKPLYIGSYCPPNNR QQGLEGLHQSLSKIMTRHRRSHPNILIA GDFNHPDINWDDQSTTNPATSTCHQK LLDILLHNSLSQTVREVTRPSSNNILDLV VTSNPALVENVCVKSGISDHNIVTFTLA ANPKISVKPLRKIYQFHKADEQQLRDAA AEFATEFLQSDPERRSVETNWSIIVNFL NRCLSELIPSKMSKGRKHLPWITPPLKR QMRKRDSLFFKKAHSQSCSTAWTAYRQ FRNQVAKNVHRARCEYVNNIIGASLKD NPKSFWSYIKTCRSENIGIPTLRTPSKLC ASALDKAETLNDYFQSVFTPHQLQPLV QKGPSPFPTIGHLHVHRTGVEKRLQQL DPSKASGPDELPPKLLKLVAEIASPLSFL FQQSYNSGIVPAQWKQALVTPIHKSGE KCDPSNYRPISLTCVCKVMEHIILSHVS KHVAVNNILTDAQHGFRRLSTTTQLT TAVHDWSSILQQRSQADIIFLDFKKAFFD			
--	--	--	--	---	--	--	--	--	--

				AGATGATTTAGCT AACCGTCTACTAC TACTAATAGCTAA CCTACATAAAATG GGTGTAGCTGAG TGAGTGAGACTG CTGAGCTTTTCTG ACTTCCCTTTGTTT GTTCTGAATACGT ATTGAATATTGAA TACTGTCCTGCCA AATTTAATCACCG GTATTCATCACCG TTCGACTTGCAAA CGCCATC		RVPHLLLSTKLQYYGISGDSLQWIMSL TDRQQAVVVDGSKSTWRDVTSGVPQ GSVIGPTLFLLYINDIQDNIQSTMRLFAD DSTLYREIRCPEDHQILKDDLQQLSRWS SNWMMDFNVKKCAILTITRKRTPSIREY HLSNGLIPRVKEYKYLGVTVAADLRWN KHCQTIRHKASRTLGLLRRTLSPCTR KSRAYTALVRPQLEYASEAWNPHSVTV TKMLEQVQRTAARFVYRDYRYTTSPSA LVAALGWDTLHTRRILDQCTLLYKIHHR LVSIPAPTIVIPATYFGRHDHNLKYVIPV ATIDSFKFSYYPRAIRIWNHLP TGITNFKEAALPSIRMMQPPVGS SHII			
	L2- 1_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	AG	TACCTGTTCATAAGA GAATAATCTATAAAG TTTCACTATTTGTTTA TAAATCCATGTATAA CCTTATCCCATCCTAT ATTTCTAGTCTGTTTT	MHGHTLDLVITRCADNILHNCMVLDK GISDHYVISCDLNLQKSDAKRTLITSRNY RLINQESFGNELVAVLQSSISQCGDVNEQ VEFINKAICDVLDRHCPVSVRMKRDRP CGPWYDDNIHLLKRKKRKSERKWRKSN LRADRDIFVCHKDELNAAIVKSKSSYFTN	cd0 165 0	306	575

					GTCCCATCCAGCAAT CAAGATATCCATTAC GTTCAAGGTCAGACC CCCTTCTTCTGCAAAT CCCCCGCATCAATTC AAAGAAGGGTCAAC AAAATATAATTTTTTC TGGTCCCACTATTTG GAATAGTCTCCCTTT AAGCATTCTGTTCTTG CACCTCAGTCGCATT ATTTAAGAAACATTT AAAAACCATTGTGTT TTCTTTGTAATTGT	LLEANCKTMYATLHLLLGSSKKVLPYC DDQTDLANKFAGFFDDKISKIRTVLDER EQGLQTSGIVQSMDKHQ CETGFSTFVD VSPSDIKMIISKSANKFCSLDPMPTWL V KDNIEVLVSPFVSLINSSLRSGVFPDKFK DAIISPCCLKKSTLDSLVLANYRPVSNLAFL SKVLETIVSKQLVKHLDRCDLHEVLQSA YKANHSTESALLRVKHDLLAIDKREGVI MVMLDLKSAFDTIDHSVLLRRLHSDFFL SGTVLDWFQSYLEGRSFAVSIGNHLSHK LSLKYGVPQGSVLGPLLFTMYLRPLGDIL RNNGVEFHLYADDVQLYYSFNPKSSSSF QNAILLIEKCVSENVVWMIKNKLLMNE QKTESLLIISRRLRHYLPDHSVHLLLGDAF VEPSESVRTLGVVLDSSLTMSQQVTS LC KSLNFHLYNISRVRRMLTDEACHHAIRS LVLSRLDNANSLININVSIDIRLQRIQN RAARIISNARKYDHSV PFLNNYTG			
	CR1Y_ SP	GCF_00000 2235.5	Strong ylocen trotus	TGTGTGTGTGTCG CGGGAGTCGGAT GTACACTGCAAAT	CAAGACCCTCCAGGA CCTGCCCATGCACAT AGCTACCAAGTCCTC	MAPAQLCGSATSVEYTISAGQIYSFPVP SVATSAAKICHLNHTDIEPCTTNPD SKD EGESDDSLCLGSPKFSSSPRTLKKKPINP	cd0 165 0	621	854

			purpur atus	GCTTCCTGGAAAC TGCCTACATCGAC ACCCTGGGATGTC TTTATCTGGAATT TCTTCATGAATTT AGGAATTTCACCC TTCCTTCTCATGG CAGATTGAATTTT ATGTTCTTTTGG CCAATTCTGTCAA CTGAAATCTAAGC TAGATTTTGCCAC TCATTTCATTAAG AGTGCCATCAAAA TCTGTAGTCCCAC AGCAAGACTCTG ATGTCTACAGCTT TCACATGGGCTAC AGTCATACATGTA CAGACATGTACA GGGAACCACTA	CGTTCAAGAACTGAG CAGCATAAGCAGTCC TTCTTTCCCAGGACT GTGAGAGTGGAACA ACCTCCATCAGAAAC AGTATGGCCTCATCC GTCAATGCCTTCGT CAATGCCTGACGGGA GGCAGCAACTAAGTC CGGGAACCTTTTACT TGCACCATGTAAATA GCATTGCACCATGCA CCTTGCTCTCGTAAG CACACACACACTCAT CAGTCTGCAAGAATC TTCAACTGTGAAGGA GGTAGACTTAACCCA AAGAAGAAGAAGAA	TKPKAKVLTILNVNCQSIRAKKQSFALI HQNHPDVVIETESWLTEDIRSSEVFPTD HYNIAARRDRPNDLHGGVFLAIRKDLLAS REIELETDCELIWCKISMPGAKTFHVAM KLRAHLVQDQHARSQDFPCCDEMSLT ELERSLSRLSPTHSVVIGGDFNLPGWD WRNQQVGNCRYPNLHRRLGSLDDH GLTQVVTEPTRYQNTLDLLLVSNTPTVVN EVTVVPGVSDHDCPLVKLNVRPTRPDE EESFTLQQSSVGEVWDQPGQMRRKVL LYSKAQWEKFGESIEAVRDEIEAKVTYA SVNELWLIFKTGIEKGINDFIPQKLTPHK ANLPWVTNGIRRLIKRDRLSNECRHLR KREGLIPYVLEERLRDLKRTLQFALRKAY WNYVESIFTVEDDDEGPGKYNCMKRF WQFIKNNRNSAGIAELKSDGVLFSEP AEKADALNRQFKKVFTQEIPPHLLPD TSEHESMPDIVITLKGIEMLKCLKPHK APGPDGITPRILKQLSRVIAPILRIIYRLSY DTGEIPDDWHKANVVPVYKKGDRTP SNYRPISLTCICCKMMEHIFASSIMRHG			
--	--	--	----------------	--	---	--	--	--	--

				GCTTCTAGTTCTA CACTGTACAACGTG GCAACCATGGTCA ATACTGTAATGTA TTATTATTATTACA TTCCTTTCTCCACT ATGAGTTGATTGA ATTAACATGCCCA AGAAACCTAAATT TCCATGTGGTTCA TGTGGGAAAAAT GTGGCTTGGAAG TCAAGGGCATTCA GTGTGAAGGTGA ATGTGAACAGTG GTTCCATGCAGAG TGTATTAACCTTA ACTCTACAGACTA CAGTGCCCTAG		DEKDILCPLQHGFQAKKSCELQLTRLCH LGFVTDLLNNMELSQQTDVLITDFSKAF DKVGHQRLKKMEYYGVRGKSIDWISS FLPNRKQQVVLGDGHTSPTVEVQSGVP QGSVLGPCLFLFYINDLPNQLSTKVRLFA DDTVLYHTIQETHEDAELLQQDLQDLC GWSSKWMMALNPKKCVITVSRKKS QVCPSVFPQWRNHWSLCQLPNTLALL SLLTSTGTQHVSATCKKANNTLAFLRRN LRVNSTQLKTTAYQALVRPLLEYGSAV WDPHTIRCVNQLEMVQRRARYVLNR YHHTSSVSDMLSQLGWTSLQKRREHQ RLTLLYKMHYSLVDFDIHHPTSNRQQS			
--	--	--	--	--	--	--	--	--	--

	CR1- 7_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus		TGATGTTACTTTTAAT AATTACTGTCGATGT TCAAACCTGAATGTTT CTTTTAGTTTGCTGTT GTTACCCTGTGAATT GCCATTGCGTACGGG TAGTGAAAATGTTAT ACAAATATTGCTAAG TGTATGAATATTGTTT ATTTTTCTTCATAATA AGTTTCAACTTAATTA AATTACGTTTTTTTAA TATACACAAGACTAT GTCATTGTTGTAACCT TCGGAGCCTATGACC TACAAGCATTGCTTT CTCTTTAGGCTCCATC ATATTTCTTCCCTTAT TCAACTTAGTCTTATG TATCATTCTTTTTTTT TTTTGTATAAACGCA	MDFNLNLLQCDENTSCQDFLDLMLS KSFAPTITKPTRITEYSATLIDNIFVSGTPIV KSGVIVTDISDHFPIYTVISNVISHSNPNT NKCSGIRYLSDPDLRRLKEALTITDWYA VYNCDDVDTSFDNFMEILMLTLNSFPT RQQNCSNYKKHPRLAWISKSLLKSINRK NTLYYKYKKNHTVKSCLKYSRYRNILTS LRLAKKNYFTKQFNVNKGDIKNTWRVI NNVLKSRRVSTHISNIMINENRIQDPLVI AEEFNYYFANIGPNLAQNIPVSDLHFQY FLETKNTNSMFCTPILEHEVIDIVKNMK TKKSAGNDGITNFILKEIIGEIVSPLTHLL NLSIMNGIVPKMKIAKVIPIFKKGNAQ DIGNYRPISLLTSFSKILEKLMYSRTMQF LIDFDILSDTQFGFRQKHSTTHALLMFL DKVAHAIDDMAHTVGIFLDFSKAFDTIN HDILLYKLSHYGIRGKALEWFRNYLTDR KQYVSINGHDSYQRSVSCGVPQGSLLG PLLFVIYINDFRNSSEIMSFILFADDTNIF YSHRDPQYLLRTVNSELKRIQAWINAN KLSLNVSKTHFMLFSNCIKSLPGNILIDN	cd0 165 0	356	610
--	--------------	---------------------	--	--	---	---	-----------------	-----	-----

					TCCTAAAATGAGTTC AAAATGTAACATGCT ATATGTAATGATGTA CTATACAACATGTGA GAAATATGAAAATAA ACTGAAACTGAACT GAAA	TNLNQAESTKFLGIYIDSDLSWKIHVNYL CKLISRNTGILNKLKHEFPTHILLSLYNTLI LPYLNYGILAWGNGTKAQLEKLLIIQKR AIRNINHAGFLSHTSQLFFNNKVLIGD VYVYSVGVFMYQLTINELPAVICALFVR NNTIHSYSTRQRNLFHLPLTRTLFAQKNI TFTGPKLWNDLPIEFTTESTSLYSFKRKLK CHLLSNYALMVEEA			
	Hero- 2_SPur	GCF_00000 2235.5	Strong ylocen trotus purpur atus	ATATATCTTCACT CTCGATATATCGG AACTCTCGATTTT TCCTACCCTACAA CTACCGGCGACG AATCACCTAACAA CCGCA	TTTATATATACATTAA TTTTTCATATTTGTTT TGAACCTGTAACCAT TCTGTACCAAAATGA CAATATTTAACGTAA TACGCAATTTATCTC GATTCAATTATACAC TATTCATTTATGTATC ATGTATGGGGATCTT CTAGCTTCCTCACCAT AATGTTACAATCTTCT TCTTTTACCCTGTTCT GTACTATACCACTTCT	MPPYVLIGSLLLFFTFQANCLRFNVPNL RATEWQPHNAAIKDRLTLLHHDLGAGI VTPKSAADDLTSLTEYFSGQDDFVDKK LSDTYIEHEPKSLIQARKCKNVLRKRAFG KNGTAEDRKAFAAVKTVSYLRKQDKQ KKCDKSRSHMEKRYRNNFYEFSSKAACS GSLNESPTKPLFSEDVANNYFGDKYSTP NVINIGSLNWFYPVNINTPDRFEFDMT PIKPKDVKNVLRKKAMSAPGNDGLM YGMIKHLDSVHHFLATLYNKILDSGDPP ESWSESKVTLIYKSGDTPENFRMISL TSCISKVFHQIAERSMSYLTENNFDKT VQKAFIDGVNGCIEHNIVLHEILSDAKCK	cd0 165 0	280	545

					GTTTCCTCTCCTTTT CTCCTTTTCTCCTTTTC ACATCCTTTTCCCTCT CTCTTTATCCCGGTGC TCCTTGTTATCACATA TATTAT	KRSLHTTFFDLADAFGSVSHQLITSSME RYAIPTPVARYVATLYSRLNGRVRGPG WITKPFEFKRGVFQGDPLSPTIFLMCFN PVLEYLKEQQDQFGYKWGDRKIITTPYA DDFNLITGNKLQHQLMNGINDRLKS MGLTLKPRKCSSLTSGKPNIVDFTIGS DVISSVHGHVPVKFLGSLITFSMKTSEIFE YKKGKIESTLENINKCNIRNEYKLCILTRY ALPSLRMYLTVHVDVHASHLKSLDIILDY IKMWLGIPSRGANLAVVHDPDGLNIPSI SDLYRTCHCLAYSRSRVKGDDVVNHAL DRRLERERGWSRQTSFSLEAHEIHSKIP DSVSWNQTKTVIKKAVQDVCRENWKR VIEPLLQGEFMKVVMHMENTDVTWКСI IYNLPQSVMKFSVNAAIDSLPTNKNLYR WGKRISDKCNLCNCTGTLHHVLNNCER MLDRYTWRHNNILRIMLKHIKNTSNN TKVYADIEGESINGGTIPPNIIRTQHKPDI CHIDNKNKDIIIIELTVPFETNIDKAHQRK QDKYASLTHDITSTGWKSKLICIEIGSRG			
--	--	--	--	--	--	--	--	--	--

						LITPQNKYRLKEILKITNSTRYPESRDEILK AVLFSSYVIFQGRREPAWQFDQYL			
	Crack- 2_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	GGTCTGTTGTATT ACGATATTTCTGA CCATTTACCGATA TTTATTATTTCTAA AGTTCACCTTGTT AAAAAAATTACTA AAAATAATTGTGC A	TTTCAACATGCAGTA GGCAATTGTCAAATT TCGGTCACATTGCAA CTCAGTTTTGCAAATT TTGTTTTCTATTCGT ATTTCTTAATTCCTT TATTGTAAAGGCTGT ATTTGTTTCTACTCGA CTATACCACACAGAG CTACGTACACTTCCA CTATTTGTTACAAAC ACATACAAACACGCA CGCTCGCGTGACGT GCACACACACACACA CACACCCACACACGC ACACACGCACACACA CAATATATATATATAT ATATACTTGACACA ACACACTCACCAACT	MHRKQTKHNIESLNFDLAKEDWVEVLC VSDVNVAYEKFLEIVLHYYHKNIPLVKIK KGKKAKQPWITKGILRSIVARNKLYKEA LTSQCNDKLKKYKTYRNHLTSLIRLSKKL YYSNNIESNKSNNNLLWKTVKELLGKKK TESNKCFLIDEQQITDPIKISKAFNEYIID VGPNLAANIDSGDMDYNKSLNFPRENS LFLTPTNDYEIINVVQHLNSTKSSGHDSI SVHLLKQIIFNMVTPITYIFNLSISTGVFP HLLKIAKVIPIYKKDDSSVISNYRPISLLPSI SKILEKIIYKRLNTFLQKNNSLIPNQFGFR IGHSTDYAILHMYDKIIEAFANSEHIVGV FMDLSKAFDTIDHKILHKLQYYGVRGQ ALSWFTNYLQDRQQYVSFQNHSTKLT IKCGVPQGSILGPLLFIIYINDIETSSLLNF IIFADDTSISYSHKNFDTLITRLNCELMKI SRWFKSNKLSLNLSTNFIHFKQTNVHD MRNCNIEIDGLPLIEKQSVKFLGVSIDSH LNFNEHIRNIHTSISRNIGILYKLKQILSER	cd0 165 0	263	519

					CATACACAAAGATCG CTTGAACACTTTTTCA TTTAAAGTTTACCTT AAGATTTTTTTTTGTT TTGTTCTTGTTTTTT TTAGGCCGCCATCCT TGCAAGCTTTGCTT ATGATTGCGGTCTCC TGCACTTTCCACCA GAATAGTTAGGATAA TACTTTAAATTTATGA TAATATGTTATGGAA ATAATAAGTTTCGTT ATTATAATTATGTAC ATATATTATGTAAAA AAACATTTTGACATG AAGAAATTATTGTTT GTTTATGTTGTATTAT ACCTTACATGTATATT ATGTTATTTGTTATTC AGTTATATTTTATTTG	SMIILYNSLILSHVIYCNIVWGNCSLTRIN SLFLLQKRAIRMITNSHYISHTPLFQRL KTLKIQDIHIFQTCIFMYKYSHNQLPSLF HNIYSLNSNIHTYPTRCSTNYHLENPKIL LAQKSVRHHGPDVWNSLPNSIRHCSSL YSFKASLKKTLSKYLHN			
--	--	--	--	--	--	--	--	--	--

					GAAAAAATGCAGAA ATAAAATCAAATCAA AATCAAA				
	RTE- 11_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	ACAAATGGATAC AAGATGCAAAAA AGTGGAAACCTC ACGTCAGGCTACC AAATCCAAGGCA GTCTACACAACAA TCAAACAAATCAC TAAGAAATCTAGC ATCAGAATGCAG GCAGTGAAAAGC AAAGATGGAGAA ATCCTTACCGAGA TGAAAGATGTAA AGAGAGATGGAA AGAAAGCTTTGA GGAAGTGTACAAT CACCAGAACCAA ACAAATGGAGAA	GTCAAGTAAGTCAAT TA	MPAKPCRESIRIILQRRIMKKTEEVSEA QAGFRPGRSTMDQLFTLRQIMEKRLER NQSLCCFIDYEKAFDSVWQEGLWKA MNFFGYSSKYIHLLQALYQQSSSAVRV NGELTDWFKTTVGVRRQGCALSPQLFNI LLEAVMLYAIHDTKIGVRVQGQLINNLR FADDIVLIADNDRDLQTIVDLVHQSSSN FGLKINITKTEVQAISKQPQLLNINVGGE QLKQVDKFTYLGGTISQTGSCSEEVKHR IGKAIGAFQALKKIWASQDIRCSTKVELY KVLVLSILLYGAETWTLKKVDENRLHTFE MACLRRIMGVTRFDRLRNTHIRTQLN MEETIIDRVATKRLRYFGHINRMNSKRY PHILLNGNIHGKRPRGRPAKRWTDCIK ADCKNRQVDSLTKATRLTEDRKVWQAI VKQKPSYIPVMVWTA	cd0 165 0	12	235

				GCACTACTCAGAT GGCCAGCCTCGA ACCGGTTCCAGAC ATTCTGCGAGATG AAGTAGAATTTGC TGTACGAAAACCTC GCAGAAGAGAAA GGATTTGACAGC ATCACAGGAGAG GAACTCAAAGCTA CTGGGCCAACAG GAATCACCATACT TCACAAAATATGG CGGGAAGAAAGT TTCCGGCTGACT GGGGAAAAGCCA TAATAACACCAAT CTACAAGAAAAA ACAAATCAACTGT GGAAACTACAGG GGTATTAG					
--	--	--	--	---	--	--	--	--	--

	REX1- 2_XT	GCF_00000 4195.4	Xenop us tropica lis	GATTCAAG	TGACATTCTGCCCAA ACCAACACACACTCC CCATAACTACTGGAC TCTTCGCAGTGTCAC TTTAAGCAAAGCCAC TTTAAATCCTCACTGC ACAATTTCAAATATT GCATTGCATTTTATTA TTGTAAATACTTGTA CCTTTATTGCACACTT TTATTATATTTAATAT TTGTATTTTTTTTTTTT TTGTTTTTTGTTTTGT TTGTCTATGTAAAGT TGGGAGGAACACGG GTCAAAAAAATGTCA TTACAGTAATGATGC TTCATTGTTATTTGTA TATGACAATAAACTT GAAACTTGAAACT	MAPVRMAAVAQCPPQLCFFLMFVVLF FSFFATCNALISYDRQTLDIRQNSQSGP SQTEFFMNRGDLPLPWQKKHRRKR GKRAGILVRLNRQHRAPLPSILLANVQ SLENKDELRARITFHRDIKNCNIFVFTET WLVPSIPDSAIVPEGLSIHRRDRTIDSGK CRGGGVCVLTNKRWCTDVQTVSSGCS PDLEYLMIRCRPFYLPREFTAIFCTAVYIP PHADTTRALDELFDAIDRQETLHPEAAF IVAGDFNKANMKKVLPKFYQHVNFPTR GENILDHVYTPYAHAYKARPRPAFGKS DHSSVLLLPAQRQKLKREQPMMRSVQ RWTDQSDSTLRHCFNITDWEVFRTAA DNNINNYAEYVMDFIHKCINDVVPRVN VRTYPNQKPWVNGEVRAALRARTAAF KSGDPSEYKSSRYALRKAIKTAKGQYRE KVESCYSNSNIRNMWSGLKTITDYKGK SCRVGEVTASLPDELNTFYARFDKHL SQ ELLLEEDGRPYQISQADVCKHFRKVNTR KAPGPDNIPGRVFKACAHELADVFTDIF NLSLLQSAVPTCFKRTTIIPVPPKMNVS	cd0 165 0	564	818
--	---------------	---------------------	-------------------------------	----------	---	--	-----------------	-----	-----

						LNDYRPVALTSIAMKCFERIVKAHITSLL PASLDPLQFAYRSNRSTDDAIAIALHTAL SHLDQRNTYVRMLFIDYSSAFNTIVPSR LVMKLRDLNIGSSLCSWILDFLTNRPQV VRIGNITSSTLTLSTGAPQGCVLSPLLYTL FTHDCTATHSSNTIIKFADDTTIIGCISDG DESAYRAEVRALTSWCRDNNLLLNVS TKELIVDYRRLLQGGGHTPIHIEGAEVER VSCFRFLGINISEDLSWSHHVGVITKAA RQRLFFLRRLRRFGMDSRILTNYRCTIE SILSGCITTWYGSCNALDRKALQRVVRS AERITRTELPAMQDLRQRCRRKMQR LSDPSHPSHRLFTLLPSGRRYRSIQTRTS RYRDSFYPQAIRLLNC			
	LINE1-38_ZM	GCA_902373975.1	Zea mays	GATGGCCCTCGTC TCTCCCGCCCCCG CCCTCTCGGCTCC CGACGCCCCTCCC CCTCCCCAGCACC AGCACGCGCCGG TTCCCATTCGGG	TGAGCTCAGGATGTG GTGTAGTGCCGGGG CAAAAGGGTTGAACC TGCTTTCAACTCTTAT GTTTATGCTTGACCC AGGGGGATAGTCTA GTGGCTTGTTGAGAG	MLGVSAAQGNGGPASSRRKRWRKKK RLEREDEPTLEIVAPPPVQMLPASDAT ACVIDWSVQMSIAEDNLKTAAVITVVC SRDISANEIVSVLAPRLEMEEGGLVLRQ LSQSSFLAVLPSRVQLDTLVDSFPIIREEN FMLSSRRWTRFGGAGGNSLKHLELEL RGLPVHAWETSVVQQLINPHACISQVL	cd0 165 0	1115	1378

				GCCGGCGGCCAA	TGGTTGTGTGTGGTT	PDTLEVRSLEAFRCLAWCSDPERIPATK			
				CTCCCTTGCTGCC	GCTGTGTTGCGTCTT	DLWITEPDQASAAEGKKALVYPVQIRW			
				GTCGCCGCAGCA	TTGGTATGTTGTGTC	VPADLSLLTGDPPLLIGGDDQDKSDTPP			
				CCCTCTAGCCCTG	TTGTGATGTGTGTGT	SRRRCSDNPPRTQHDAINDVPGNKHTT			
				GCGCGTCGGCTTC	CGTGTCAGGTGGTAT	QRRPVHDLGLPLRRSDSPRELHPRVQI			
				TATCCCCTCCTTGT	GTTGTGTGCGGCGTG	DFEARSSQPCVLSDAVADSCVGNPPTP			
				CTCCTCCCGTGTC	TGTGTGTGTGTGTGT	VPDGVQCLAAGDVLLRSEEGCEDAEVI			
				CCATGTAGCCAG	TGTACCTTCTTCTATC	PMTMAVSSPCPSSGRRRLLLKTATK			
				GGGGCGCTCCAA	TATCTATTAATGGAA	QAKFRNPPIAFFRRRRRAPRTPEGATLP			
				GTCTCTAAGGTGG	TGATACGCAGATCTC	PCSSGDRGAGSSLGALPPAPAAGVAAS			
				GGTGACCTCAGTC	CTGCGTGTTGAGAA	LLPHGSPPLVESHGSPGQQDFLRRVSK			
				CGGCCTCTTCGGA	AAAAAAA	AVGAVLPLPKAVGGRRRRRQPTSLPDA			
				TGGCGTGACGTC		IRRSRRVAGLGVEQVQHIPKARRTIMKS			
				GCTCGACTCCCGG		ILAEPDSEHFSQQNLDAYAKLFQAPLSR			
				CCTTCCTTCAAGG		AQVEALAALFGWVMPDDEGEVPPGT			
				ACGTGCTGGTGA		ATLVCLQPTVAYYSNCCPMDPSLILFW			
				GACCAGCGAATC		NVRGLNSAARQDVVRSLVDSLVRDVVC			
				CTTTGCCGCCAG		LQETKLVNCDRRTVLSMLGADFDNNFI			
				CCGTGTGGTAGTC		CLPSVGASGGVLITWRACLGRIVASRVD			
				CAAGAGTCAGCC		SFSVSVKFNPADGVAAWWLTCVYGPQG			
				CTGGCAGCTGGA		HDRKIAFINELRAIRAACTGPWLLGGDF			

				ACGTCCTCGGCGC TGGGGAAGGTGA ACCATCGGCCGCC TTGCTCTACCCC CTGTCTACAAAGC GGCGGAGTCCCA GGCAGGCTCCCC ACTCTACACCGGC TTCCGCCCAGCT CCAAGTTTGATG ATCAAGGATGGA CAACGGTGGAAT CCAAACAAGGTA GAAGACGCCGGT TGAGGGCTTCAC GACGCCCAGTC GAGTCCCTGCTGA GCTTGTTGGAAG GTGCTTCAACTGC TTGGGCTGCGATC ACAGGGCTGCGG		NLIFRDEDKNNININRPMMGRFRRRCIN DLSIKDIDLLGRKYTWSGANNVQVKLD RVFCSSDWEEKFPNYLLQSSASEDS DH CPLLLGLKDNHTGKRRHFESFWPRFD GFLEAVSNTWGSVQQSSCPLKTLCLKFK ATTKGLQSWSDRKVGNFKLQLELAREIL LQLEVAKESRILNVLETWLFGRLLKCSLS LASLIRSMARTRSRIAWLKDGDANTRLF HSWSRYRKRKNFIAKIQDGEQLVTS HS DKAEVLRSFYSDLIGTGQQRSTVNLD T LEMQHTDLQGLEQPISEEEVWNTIKQL PPDKAPGPDGFTGRFYKVCWSIIKKDI MAAIMAIWNKDFRNFYLN TAFITLLP KKEDAILAKDFRPISLIHSFAKLITKIMAS RLSGYMDKLVSRNQSAFIKGRFIQDNF MLVQQTARLLHAQKQPRILLKLDISKAF DSVSWSFLLEVLEKVGFGPIWRDVISGL LMTSSTQILLNGVPGEFIQHLRGLRQGD SLSPLLFIIVMDVLNGLIVKAADEGLLQP LSTRSIQHRISLYADDVVIFLRPAASDLHL VTRILELFGDASGLKTNVQKSSVAPIHCS			
--	--	--	--	--	--	--	--	--	--

				GGTGTGACAGTG AGCCTCGGTGCTT CCTCTGCTCGGAG CTTGCCACAAGT CTTCCTCCTCCTGC CCCAAGACCCCG GTGTGGAAGCGG CTCGGGCGCCCCA GGCCTGTGAAGT CCCGGTCTGGGA GCGGCTTTCACCG GCAGTTGAGGGG AGAGTCCAACGA GCGGCAGGCCAC CGTTCGGTTTGGC AGCGTATCGCTAC GCCCCAAGCTGTT CCGATCGCCCCTT CTGTCCTGGTGGG AAGC		PVDMEMVTQIFPCKVETLPLKYLGLPLS LRKLSKEQLQPLIDRLADLLPNWKADL MTRAGRAIQVQHVLTAIVYHAMALD LPKWAIDKIRRGFLWKGRKDVGG HCLVAWGKVTRPKELGGLGISDLQKLN WALRVRWLWLEKTDPSKSWADLPLQS SSILSILSTALITEIGDGRNTLFWRDRW LMGLRIEDLTPHIFAIPTTRITNRRTVAE GIVEMAWIRDLRGITWDVLSDFLTSE VIADYSINVGVPDKHIWKFSPTGQYSSK SAYDMLFLGSTSLGAYERIWESWAPSK CSFFLWLVMMHNRCTADRLARRGLPH PSVCPLCDQEEETIQHLLISCVFARQFW YSLLSRGLPNLSPDNASCSLDDWWSR AVEKIPRDLQSGFNSLVILGAWSLWRH RNDCVFNGEIAKPGECSHPRWR			
--	--	--	--	---	--	---	--	--	--

	LINE1- 24_ZM	GCA_90237 3975.1	Zea mays	CCTGCATGCTTGA GTTCTCCACTGCT ATGGCTCGGGAG GAGGCGGCTCTTC GCCGGGCTTTGTT CGTCTCTATTGTT GGCAGCAGACCG GAGATCAGGGAG CTGAAGTGCGGG ATGAAGTGGCTC GGGCTTTTGGAAT TGACGTAAATTCA ATATCGATTCATC AGTCCTTGCCAGA AGATTTCTATTG TTGTTACCGGATG AAAGTGCTGATA GAGTCTTCAGCG GAGGAAAATTGT TTAGGGGCCCTCT TTCAATCTTCAAT	TAATTCTCCAGGGGG ACTCTTCCAAGCTTTA TCTCCTTAACTCGGC GTATGTGACTCTCTT ACCAAAAAAGCTGAG GCTGTGGAGGTAAA AGACTTCAGACCGAT TAGTTTAATTCACAG TTTTGCCAAGATTCTC ACCAAGATTCTGGCC AATAGGCTTGCTGCA ATCCTCCCTACTTTGG TCTCCCCCAATCAGA GTGCATTCATCAAGG GAAGAAGTATTCATG ACAATTTTATGATGG TTAGAGAAACTGCAA AAGCGATTCACAGGC AGAAAATTCCTAGAA TCTTGCTGAAGTTAG ACATTGGCAAGGCCT	MNLSKILIWNVRGLNRKARRDAVRMMVESTRPDIVCLQETKKE SISRFMVFSLLGNVFDQFTFLPAIGTRGGILLAWKGSVCSALQTR VDEFSISVQFNQNDSLPWWFTGVYGPQSDEKKLQFLQELRMV RSVCVGPWAVGGDFNLIYRAEDKNNNNLDRAMMGRFQRIIN DLQLVELPLLGRKFTWSNERSSPTLVRLDRFFYSRDWEDLFPDCL MQSSSSSISDHCPLVVLHEFTQGKRRFHFESFWTKLEGFMTEV ENSWCQPIAVGCPLQRMADKLKRLSKHLQSWSQKRVGHVKRQ LVMTNEILHRLEVARDSDLSYHEEWLRRKLKHSLGLASLERT MARHRSGLNWLKEGDANTAYFHHSRYRKRKNFIGKLKVEDRI VVDQEQEKKHAVWNFYNNLLGVAASREFTLNLGVFHRPASNLE VLDQIFSEEEVWNTIKSIPSDKAPGPDGFTGRFYKMCLDGCQRG FLGSCQ
--	-----------------	---------------------	-------------	--	---	--

				TCAGACGGTGGT	TCGACTCAGTTTCGT	
				CGCGGTTGGCTCA	GGCCTTTCCTCCTTGA	
				TGCTGAGGCGGC	AGTTCTGGAACATGT	
				TAGGTTGCCAGCA	GGGATTTGGTCTATT	
				CTGGTGGAAGTC	CTGGCGTAATTTAAT	
				GAGGTTTCATGGC	TGCTAAGCTATTGTG	
				ATCCCAGCCCACG	TTCCTCCTCTACTCGA	
				CATGGGATCGTTC	GTGCTAGTTAATGGG	
				TACGGTGGAGTA	GAACCAGGAGATCTC	
				TCTGCTCCGTGAC	ATTTGTCACCAGCAA	
				TCTTGCATTATCA	GGCCTGCGGCAGGG	
				CAGACATTCATCC	AGATCCTCTCTCCCA	
				GAGCACGTCATTG	ATGCTCTTTATATTAG	
				AAGAATGATCTGT	TAATGGATGTGCTGA	
				CTTCCTTCAGGCT	ACTCTCTTTTCAGAA	
				TCGTGCATGGTGT	AGCAAATGAGTTGG	
				TTGACACAGATTT	GCTCTTGGGACCTTT	
				GTTCCCTCGGCCG	ATTAAGAAGGGGAG	
				ATGAAGCTCCTCA	GGGGCCAGCGGGTC	
				TTGTTGAGCCGG	TCTCTTTACGCCGAT	
				GGACGGATGTGC	GATGTTGTATTATTT	

				AGGAGAAAAGAT	GTTCAGCCTAACAGA	
				GTCTTTCCTATGA	AATGAGCTGAGTGTG	
				TATTGAGGTGGCT	GTCAAGGAGCTGTTG	
				GTTACTGCTGTGA	AGGGTGTTTGGCAA	
				TGGTTCCGACGAA	GGCATCAGGGCTGAT	
				TTTGACTCCAGCT	TACAAACATAAATAA	
				CATTCCCCTGGCA	GTGCAGTATGACACC	
				TGGACGGGAGGC	TATTCAGTGCAACGA	
				ATGAGGATGGCG	GCATATCATTGCTGC	
				GCGGGAGGCATT	TGCTCGACAGATGTT	
				CGGAATCCACGTC	ACTCTGCAATGTTGT	
				GGGCATCGCGCC	TCCTTTTCCATGTAAA	
				CAATTTTCCAGCG	TATCTAGGTCTTCCTC	
				TCTGGGCCCAACG	TCTCTGTAAAAAGC	
				TTCGTGGTGCCCA	TCCCTAGAAGATCCT	
				GCGCTGCGGCAG	TTCTGGAGCTTATTG	
				CGGGCCCGGCTG	ATAAGGTGGCAGCA	
				CAGGACGAGGCG	AAGCTTCCTGGCTGG	
				CGCGGTCCTTTGG	AAAGCTGCTAATCTT	
				ACGAGGAGGTGG	AACCTGGCCGGAAG	
				CTGCAGCCCCCCT	GGCAACTTTGGTTAA	

				GTTGCTGCGGGC	AGCTGTGCTCACTGC	
				CTGTGCGGCGGC	TATTCCCATTATCAT	
				TGGCGGTTATGG	CTAATTGCTTCAATGT	
				CGGGAGTCGAGG	CCAAGGTGGGTAATC	
				AGGTGGCCGCAC	AAGGCAATAGACAA	
				AGGAACCGAGGG	GATTCGCAGGGGTTT	
				CCTCGCCGTCCAG	TCTTTGGAAAGGAAG	
				CGCACTGCTGGG	AACTGACATTAGAGG	
				GTGCTGGCGGAG	AGGACATTGTGTCGT	
				GAGGTTGGGCCA	TGGGTGGTCCAAGGT	
				CTGGAATGGGCC	CTGTCGTCCCCTCTCT	
				AGGGAGGTGTTG	CTTGGTGGGTGGGA	
				GCCCAGATGTCAC	ATTCACAATTTGGAA	
				GGCAGCAGATGG	GTTATGGGATGGTCT	
				GTCCCGGGACAG	TTAAATTTGAGATGG	
				TTTCGCCCCGCTC	CTTTGGCTCCAGAAA	
				ACTCCTGTGCAAC	ACTGAACCAGCTAGG	
				AAGATGATGGGC	CCTTGGGCAGATTTT	
				AGGGCTGTGTTG	AAAATAAAGGTGCAC	
				GCTGGGTCGCTG	AGATGCTTCAGCTCT	
				AAATCGCTTGGGT	CTTCGCAGCCTCTGT	

				CGGGCTCTCCCCG	GTGTTCCGGTGGTTGG	
				ACAGAGGTTTGTG	TAATGGGGCAAACAC	
				GGGTTGCAGAGG	TCTCTTCTGGACAGA	
				CTGACCAAATTTTC	TAAATGGCTTCAGGG	
				GCCCCCTTCAATG	CCAATCTTTCTCAGAT	
				TCGGCTTGGTCCC	TGCCCCCGCTCTAAT	
				CCGGCTGACCTCA	AAATCGGGTGCCCGC	
				ACTCAAAATCTTT	GAATATAAGAGGCG	
				GGTTGACTGGCG	CAGATCTGTCCAAGC	
				GACAGAACCTCCC	CGCCTTAACAGCCAA	
				TGAATGTTTATGC	TGGTTGGGTGCATGA	
				TAGGAGGCCAGC	TATCTCTGGAAGCCT	
				TGATTACAGATTG	TACGGCGCAGGTGCT	
				AAAAAGCAGACT	GTTTGAATTTTTATTG	
				TCAGATACCGCAC	GTTTGGGATTTACTC	
				CGGTTCAATCCCT	CAAGAGGTGCAACT	
				TCGGCCAGATGA	GTACCCTAATGTGGA	
				GGTGTCGGCACCT	GGATCAACATCGGTG	
				CTGGTACCTGCTG	GTTGCCATCCTCCTC	
				GTATTGGAGAGG	GGGGCTGTATTCCGC	
				CGACCCCCGTTAT	CAGCTCTGCTTATTTG	

				TGGTGGCTAATAT	CGCTTCATGGTGGT	
				CCAAGAACATTGA	GCAACAACCTTTGAG	
				CTGCTTGCTTCCG	CCTGCTACAAGAATC	
				CAACCTCCAATTC	TGGAAGACTTGGGCC	
				AAAAAAGAAGAA	CCCCCGCGCTGCAAG	
				CTAAACAGCTTCC	TATTTCAATTTGGCTTG	
				GCCAGATTTTGTC	CATCTCTCAACCGAT	
				CCACGTCGCAGTA	GTTGGACGGCTGATC	
				GTCGCTTGTCAAA	GACTGGCACGTCATG	
				GAAAAGGGAAGG	GAATGGATCATCCAG	
				CTTAAATTCAACA	ATAGTTGTCCCTTGT	
				GTCAGACAAGTCC	GTGATCAGCAGGAT	
				AGGCGGAGCTCA	GAGTCAGTGCATCAC	
				TTTGAAGTTGAAC	TTGCTGGTGGATTGT	
				GTCACAAACCAA	GTTTTTACGAGGGAA	
				GGTGGCTGTGAC	GTTTGGTTTACAACA	
				TGATGAAATGCTG	CTAAGCAGGCTAGGT	
				GAGGAATTTGGC	TTGCAACATCTTGCTC	
				AGCTGTTTGACAA	CCACGTCTGATGCTA	
				ACCCCTTTCCAAC	CAGGGTTTCAAGATT	
				TCTCACATCAAGG	GGTGGAGTGAGGCA	

				CCCTGGCAGCTTT	GAAAGACAAGTGTCT	
				GTTTGGATGGTCG	AAGCATAAGAAAAA	
				GTCCCAGATAATG	AGGTTTCAACTCTATT	
				TGCAAGAATGCTT	GTAATTTTAGTAGCT	
				GGTGTCCTTTGC	TGGCAGATTTGGAAG	
				TAGGGAGAAGCA	CATAGAAATGCGTGT	
				TTGAGTGTGCTTG	GTGTTTGACAATGCT	
				CTGCTGTAGCCTC	TCCCCAAACATTAC	
				ACTTAGTGGTTGT	GCAATTTTGCTGTCC	
				ATTTCTGTTATCA	ATCCATGATGAGGCT	
					TTATTATGGGGTCTT	
					GGAGGGGCTGCTGC	
					TCTTAGGCAGGTTTG	
					GTTGTAGTTTTTTTAT	
					TTTTGGTTTTTTTAAT	
					TTTTTGGGGTGCTTC	
					CTGGGCACCTTTGTA	
					CTGGGTCTTTTCTAG	
					ACCCTTTTTCTTATCT	
					TCTTAATATATTTGAT	
					GCGTGGTTCTCCAC	

					GTGTTGAGAAAAAA AAA				
	LINE1- 41_ZM	GCA_90237 3975.1	Zea mays	G	TAATTACGTAACTCT ATGTACCCTTCCTGC GGTGTCTCTTGGGCA CTGCTTCACTCTCTTC TTTATTAATAGAAAC GTGCACAATCTCGTG CCGTTCGTTCAAAAA AAAAA	MADVMAVMNSLHDQRFLNFDLLNRA NIVLLPKHDGAENLSSFRPISLIHSVAKLF MKLLSMRLAPTMRDIISKQCQTAFIKGRS IHDNFLFVRNMARRYHRNGTPMLLVKL DISKAFDSVRWDYLLALLAHLGFPPKW RDWVASLLATCSSQVLLNGIPGQSITHG RGLRQGDPLSPFLVLAIDPLQRLHLAT EAGILSRVAKARTRLRIAMYADDAIIFVK PEKRELDSEFATLLHQFGEATGLRTNIQK SSVVPKICAGLNLDLILAGFPASRSQFPI KYLGIPLSVERLKKVDFQYLLDKVSSKLT CWQGRNFTFAGRLVLVKSVLQAQPVHT LTTINVPKVVLEKIDKFRKFLWAGNEN LTGGKCKVNWPTVARPLDLGGLGVLDI HRFARALRLRWLWREWEVPEAPWAG LQTPCNDTDKLLFAAATRIAIGDGAKTS FWHCAWVNGRRPKDIAPDIFAASSDC RLSVREALTDSTWVNRIIDLHAIHTADHL KQFVELWSFAHSVSLVQNTEDRIDWIY	cd0 165 0	23	286

						NASGEYSAASAYRIQFMGHTCSNLKALI WKPWGPSKCKLFAWLIIRNRVWTSDR LAARGWPHSPSCPLCKRHPETAHHIIA GCRYSRRIWDALASWLSIGPLAPATWP DSTSIKQWWFMVGSVAVPKKGTRSLLL LVVWHIWLERNNRTFQRQERSVADLL AAIKAEARLWGVAGAKHLLALLPML			
	LINE1-35_ZM	GCA_902373975.1	Zea mays	ACAAAACATGGT ACTGCGCTACGAC GGAGCCGACGGG TTGCAGGGGCAG GGGTGGAATTTA ACATGCAGGATCT GGAGAACAGAGC CACAAAACAAGT AATGCGGACTCTG AGAGTCATCACA GAGAATGCAAGC ATCAACCAGCAA GCTGTTAAGGATT ATGAGAATCTTTT	TAGCTCTGGTTTTGTT TTCAGGATCTGTTTG CTCTTGTTTTTGTTTT TTTCCTCTTCTTTTTTTC TTTGGTTTTTTTATCC CCTCCGGGGAGGTCT TTTGTACTATGGTCTA TTTTGGACCCTCTTTT TCCTCTACTTAATATA ATGATGCGCAGTTCT CCTGCGCTTTCGAGA AAAAAAA	MDPSKIICWNVRGLNSADRQNSVRNL VLSSKADVCLQETKMTVISTRILAALG SDFCQFSFLPSCGASGGVLVAWKRLHG VTGLSRIDRYSISIQFCNAHNRCWWLT CVYGPQSNSDKMLFLQELRDVRAACTG PWLIAGDFNLIYKASDKNNSNINRAM MGRFRKLVEDLSLNEVPLFGRKYTWSN QQDSPVLVKLDRVLCSEWEDNFPNVL LQSAASEGSDHCPLLLGIRDNKLKRRF CFEAFWPKLEGFQDVIQEVWDSVAAV NCPFMTLNLKLKATARKLQSWSGKHV GHVRTQLALAKEILHRLEIAQDCRALSLA EIWLSNLLKKHSLFLSSLQRTIARSRSRIS WLKEGDANTKLFHMHSRHRKRRNFVA	cd01650	498	760

				TCAGCATCCTTA ACTCAAAGTCATG TTGAAGCGTTAGC GGCTCTCTTGGA TGGGAGACTCCG GGGAATCTCTCC CGTAGTGTGTTTT TCTTTCTGGGTG TAATTTATAATTG TTGTA		QLVNGDQILTNNHDDKAVLVDNFYARLI GECGDREQTINLEALHLPTFNLHLDEP VTELELWNTIKGLPNDKAPGPDGFTGR FYKSCWNIIKMDLMAVVAAVWNRKFD NFCLLNSAFITLIPKHDGAVHVKDFRPIS LVHSAAKLITKLLANRLSQKLNDMVSPI QSAFIKGRFIQDNFMLVQQTARLFHQQ KRPRLLLKLDITKAFDSVSWPFLIEVMKY LGFGQIWRDIICGLSSSSTQVLLNGIPG GHIEYRRGLRQGDPLSPMLFILVMDVL GLLFSAEHAGLLLDLSPRRPLHRVSMY ADDVALFLHPSAADIAITLDILHLFGEAS GLRNNEQKSNVYPIQCGGVELEMVQD LLPCGISTFPCKYLGLPLSLKNLTKVQVQ PIVDRIADQLPGWKADLMTRAGRKVQ VQHVLTGMLIYLAMAVNIPSWAIAID KIRRSFLWRGRKEARGGHCLVAWGRV CRPLQLGGLGISSLKELIWALRMRWLW LQKTEPNRPWCSLPIKVPDKARAFFSM VLTskVGNGANTLFWTDKWIHGQKIA DLVPRLFDIIPKKITNRRTVQEALDRR			
--	--	--	--	---	--	--	--	--	--

						WIADITGSLSVGVLLDYLHLWNLISDFEL LPDIEDKHTFSIAANGTYSAKAAYDGLFI GAVQFGHYERVWKTWAPSKCHFFLWL SVLNKCWTADRLAKRGLDHPERCPLCD QESESLDHILVSCVFAREFWFRLMCIFRL QNLAPTSTVSFTGWWEMIDNGSGD MIMKGVNSLIALGAWIWNHRNRIVFD GISPSVSAALCQAREEQQLWEMAGAK GLSFLAATIRGT			
LINE1- 2_ZM	GCA_90237 3975.1	Zea mays	GGACCCTGTTGAT GAGCCAGGGACC GTGGAGCAGCTG GCGCCCAGTACTG AGGATTCTATAGC TGTCACCTCTATT CCTTCCGCGAATT TACCTCTGCTTAC TTGGGAAGAGGA TAGACTTGGTGCT GCTGTGGTTCAG ACATTGTGTCTAA	TAGCAGTTTTCTGCT GGTAGTCGTTTTGG TCGGGTTGTTACCTT GTGTGTGTGTGTGGT GCGTTGTGTTATTGC CGGAGGGGTCTCGG CCTAGGGGCCTTTGT AATTAGGGGTTGACA AGTCTTGTCTCTCCTA TGTTCTTTTTTTCTTCT TAATATAATGACACG	MNLSNILVWNARGLNNKARRDYVRDT ILSSKADIICLQETKVASFSSHLLLSVCGS DFDKFLTLPVAVGTRGGILIAWKGAVCQ VISSRIDNFSVSVQFAGMDGMNWWF TGVYGPQEDDRKIQFLQELDFRTLCSG PWLLSGDFNMIYQAADKNNANLNRAL MGRFRRFLDDCLLKEIPLHGRKFTWSN ERSSPTLVRLDRVFCNDWEDFFPNSLL QSASSGVSDHCPLILSLNVQTHKKRRFH FESYWPKLPGFLEAVSHSWNAPVSTAC PVEGIFLKLQRLSKGLQKWSHHKVGNV KLQLATAKEILHRLEIARDNRVLSPGEE	cd0 165 0	496	759	

				CCAGGGTATTCT	CAGCTCTCCTGCGTTT	WLRKKLKLHCLGLASLERTIARLRSRILYL			
				CACCCTGTTGAGA	GTTCAAAAAAAAAAAA	QEGDANTSFFHQQARYRKKKNFIAKFQ			
				GGCATAGGTCTGT		VEERIVTSQEEKQQAALDFYDNLLGTAE			
				TGAGCCGCAGGA		HRDYTLDLHSLGIQQHDLGDLELVFSM			
				TGCTGCCATTAAT		EEVWSVVKDLPLDKAPGPDGFTGRFYK			
				CGCAGAGGTGGC		VCWDIIKEDMMGALLAVQCGHVSKLK			
				CTCCCTTCCAACC		LLNTAFITLLPKKVDALLVKDYRPISLVHS			
				TTCAGGTATATTT		FAKLVTKVLAARLAPHLPSMVSINQSAF			
				TAGGCGAAAGCG		IKGRSIQDNFLLVQQLTRSLHHSKEPHIL			
				TCACTCTAAGAAA		LKLDISKAFDSISWPFLDMLQHLGFGR			
				TGTTCTGCGCCGG		SWCNLISLLLCTSSTRILVNGEPGEYILH			
				ATGCGCCTGCTGT		HRGLRQGDPLSPMLFILVMEALNAMV			
				TTCTTTCAATGAC		SYAFREHMLQPIANQHAKHCISFYADD			
				CTGGAGAAAATC		VVLFMRPSHQDISAIIILDIFGHATGLKT			
				ATCAAGCCAGTG		NISKSSVTPIQCGANELTTISALLPCATK			
				GTACAGCTGCTGC		NFPCTYLGIPLSIRRLANSDLLPLVDKVA			
				CCATGCCGGTTGC		DKLPGWKANLLSRAGRLVMVKTVLSSV			
				CAAGATTAGAAG		PIYLMLALELPKWVFKAIDKRRRGFLWK			
				CAGTAAGACTGTT		GQGDANGGNCLVSWEAVQRPLKYGG			
				GTTGCTGTGCCTC		LGILNLEMFGWALRARWLWLQKT DAS			
				GTCGGAGTAGGA		RPWAGLPIRVHRNAKALVDVAITSVVG			

				GGGTGGCTGGTG TCAAGCCTTGTTT TCCAGGTCCGATC ATCACTGAAGCCC AAAAGAAAGTCA TCAGGAGTTTGG GTCTCGCAGCCAC GGAAGAAGCCAT TGATCAAAAAGCT CAAGACAACTATT TTAAGCTGTTCGA GGGGCGCCTGAC GGACTCTCATCTT GCTGCACTTGCTG CTATTTTTGGCTG GCAGGCTGAGCG ATTTGATGATTCC AGCAGCCCAGTTC TCCTGGCAGGCTG TTAATTCCTGCTG		SGESVKFWSDRWLLGKTVAEHCPTLIQ VISRRALKRRTVAEGLTNRQWVSDIRG GLSVTVLVEYLQLWNLVDGVVLQPDIA DQHIWRLSAHGTYCSKSAYDALFVGSIP FGPWRRVWKTWAPLRCKFFVWLAIKN RVWTADRLAKRGLPHPVACPLCDQAE ETIQHILVSCVFARQIWTSMLHNLGLLAI VPQHGCTRFSNWWCQSIKKVEKSLRK GLNSLIILVAWEIWKHRNACVFEGVVP CTQRVQSAVIEEGNVWCLAGASALQD LLLRQLPRGP			
--	--	--	--	--	--	---	--	--	--

				TCCTTGGTCGTGT TTATTTGTTTTCA					
	LINE1- 18_ZM	GCA_90237 3975.1	Zea mays	CCAAAAAGGAAC TTATTGATTCAGT TTATCTAAAAAAC TTCTGTCCTAATC AGTTTGATTCTTT TGCTTTTGTTCCCTT CGGTGGGGAACT CGGGGGTTCAATT ATCATCTGGAAAA GTTCAAAATAACA GGAAATGTCATCT TTCAAATGACTA TGCTCAGTCAGTG GAATTCATATCCA ATTTATCGCGTGC TCTTGGATTATTA CAAACATTTATGC ACCTTGACTCCA GGGAAAATAGAT	TAACG TTCCTTTTCT ATGGAGGTCATTATT CTCATGACATGGAGC ATTTGGAAAAGTCGA AATGAATGGTTGTTT GAAAACAAAGACCCT TCAGTGCATCACTGT ACGAATGAATTCAGT AAGGAACTACGGTTG GTTATTCACAGAGCG CGAGGGAAATACGA CATCTCCATTCCAAG CTGGTTGGATCAGTG GCAGTCATAAGATTT GTTTTGCTCTGACCA ATCTTTGTAACTTTTC AATTAGATCTTACCA GCTAACCTCTAACTCT TGTAATCAAGTCGAG	MPSDKLWLLVGDFNLIRRPEDRNKAG GDINLMLKFNEAISNLDLVEIPLHGLSYT WSNRQREPLLQRLDWFFISQEWSIVYP DCAKTMPRDISDHVPCLVSFKSKIPKP KIFRFENFWLEFEGFMNIFQTTWLGLP CLPDKAKNLTAKFKIVRKVLKDWQRS LP KIDKTVRDIKLLIEFIDNIEEHRDLSIEEW NFRDIMQQKVAELLNQKIYWKQRASI KWITEGDICSRFFHAHATVKHRRNTIAS LTDENGIHFEHDHKSNULLWEAFKSRLG SSDFLGIDFDLSELLRNEDLQGLDSPFT KQEIDNIIKSLPSDKSPGPDGFNTNFIKK CWHIIAQDFYDLCEKFYQEEVCLRSING SFIVLIPKKENPQKVG DYRPISLLNNNM KILTKLLANRLQSFMTRLVHRNQYGF IK GRTIQDCLAWAYEYIHLCHSSKKEIIVLK LDFEKAFDTLEHELIMQVMTHKGFGSK WINWVRNILHSGTSSVLLNGVPGKSFH CKRGVRQGDPLSPLLFVLAADLLQSIINK	cd0 165 0	365	629

				TTTCTCAACTGGT TCTATAACATTTCT	TTGTACCTTTTCATAT TTAATAAAATTTTACA GTAGGGGTTTCCCCT CCTGATCTTTCAAAA AAAAA	ARHLNLLQLPLPDICEQDFPIVQYADDT LLIMEACPKQLFFLKAILNSFASSTGLRV NYNKSNMYPINVNPEKMEILSRTLNCQI GAMSFTYLGLPLGLAKPRIQQFLPLIQRI EKRLSCSSNFLSQAGRLELVNSVFSALPT FFMCTLKIPKSIKQIDIYRKHCLWRGND TSSKKPPLVAWSTVTQRKVNGGLGVLR LETQNDALLKFFHKFFNSYDLPWVNII WNKYYTPNRLPGHRRIGSFWWKSMK LLHNFKGLAHPIIGNGRSILFWEDQWN KSIPRQQYPELFSFTKSTKLSIAEVKEQD HLFRIFQLPISEQAYEQYLELQVVWEQI VLSNVKDRWRYIWSDSYSTQKAYRHL MGQTQVHSIYKCLWWSKCQPKHKVFY WLWLKNRLNTRNMLRRKNMTLESYSC ENCLWQREETLYHLFLRCNFAKACWNS IGIPPTQNCSSRGGLGKPQTTT			
	LINE1- 28_ZM	GCA_90237 3975.1	Zea mays	GCGGTCATTTTCG GCGGACCGTTTCG GCTCCCCGCCCTC CCGCCTTCCCTGC	TAATGTTTTTCCTATC CGATGTTCTGAGGCG GATCTTGCTTGTGTTC AGGAGCTCCTTCAT	MVVFSMFFQMDPSKIVCWNVRGLNG SARQDVVRCLVDSVKADVCLQETKM SVFSRLFVFRLLGSEFQNFVFTPSVGAS GGILVAWKNNISSCGSRIDHCVSVNFS	cd0 165 0	506	704

				ATGCCCTCCGCC	GTTCCATTTCGGAAT	LETGSSWWLSCVYAPQGNHEKIQFLQ			
				TCCCCGCCCATC	TTCCCTGTCAGTACCT	DLRLIRTHCSGPWLVVGDFNMILCDED			
				GCTCCGTTAACGG	TGGCCTCCCGCTTTCT	KSNSNVDRAMMGRFRRWKDDLALLEL			
				CGCCCGCCCCACC	CTTGGAAGCTTAAG	PLIGRRFTWSNGHACPTLVRLDRMFCT			
				CTCCCCGCACCAC	AGAAGTCAAATTCAA	ADWEDLFPDCLLQSAATQDSDHCPLLL			
				CTAGCCCCTCGCC	GCAATTGTCGATAAA	GLCDINKRKGRFHFQAFWVKLEDFQDV			
				GGCGCTCACCCCG	ACTGCAGCACAGTTA	VANAWSSIPAGQCPLVSLSKKLQATAK			
				ACTCCGGCGACCA	CCTGGGTGGAAAGCT	ALQRWSDNKVGDCLKLQLEMARELLHR			
				CTTCCTCCTCCCC	GATCTGATGACAAAG	LEMAHDSRTLSPSEDWLRRTLKKHILVL			
				TCCCGTGTGGCGA	GCTGGTAGAGCCATT	SSLSRSIARLSRINWLKEGDAHTSLFHA			
				CTTTGTCTCCCGC	TATGTTCAATTTGTG	QARIRKKKNFIGSLSTSEGLILTSBEDKA			
				GGCTTGTCCGTTT	ATGACATCAAGAATG	DELFHFYDALLGTALGRENSLNLDALAM			
				TCCCCGCTGGGC	ATTTATGCTGCTCTG	PSFNLNELDVPFSEEEVWGTVKSMPSD			
				GTTCCAAGGCGC	GCGTTGGACATCCCT	KAPGPDGFTGRFYKVCWVTIKADVMA			
				AGCGGTGGGCGT	TCTTGGGCTTTTAAA	AISCVWARKFRNMGLLSAFITLLPKVE			
				CTTCCTGCCTCAC	GCAATTGATAAGCTT	GATQVKDFRPISLVHSFAKLVTKMMAN			
				CAGCCCGTCCTCA	AGACGAGGCTTTCTG	RLARHLNDMVSPNQTAFIGKRFILEDNF			
				TCGGCCTTGGGC	TGGAGAGGGAGGAA	MLVQHTARFLHRQRQARILFKLDISKAF			
				GTTGGTGCCGTTG	GGAGGTGAATGGTG	DSVSWAFLEVLQKLGGFQIWRDIICGL			
				GAAGCGGTGCCT	GCCATTGTCTGCTGA	LASSSTRILLNGVPGDYIAHQRLRQGD			
				CGCAGAGGACTTT	CTTGGCCCAAAGTCA	PLSPMLFILVMDTLNLLIQRASDEGLLHS			

				CAAGGACGTGCT	CTAGACCCAAGGACC	LSTNSLQHRFSLYADDAVVFLKPAAFRH			
				GCTGCTCCCTCCG	TGGGGGGTCTGGGT	KSDSGNPSDFWGCVRSKDQRSEE			
				CCCCTGCCGTCTG	ATTAGTGATCTTCGG				
				CCCGGCAGTCCCT	ATCATGAGCTGGGCA				
				CCTCGACCTCCGC	CTGCGGGCTAGGTG				
				ACAGACCGAAGT	GCCTTGGCTGCAAAA				
				TTGCGGTGGTGC	ATCTGATAGGACTAA				
				AACGTCCCTGGAA	GCCATGGTCGGACTION				
				CGGTGCGCAGGC	CCAGATTCAAACCTG				
				CTGTTCTGACCG	CGACATTGTGCAAAG				
				GGAGGGCTGGGT	CTTGTTGGCTGCGGC				
				TAGGGTCGAACCT	TGTTATCACTCATTTG				
				CGTCGACGTCAAC	GGAGATGGGTCTGA				
				TCCCCCTTCGCC	TACTCTGTTTTGGAA				
				ACCTCCCCGTCGA	GGATCGTTGGCTGAA				
				TCGATCCCTGAGG	TGGCAAATGTATTAA				
				ATCTCCGGGTAG	GGAGGTGGCACCTG				
				GTGTTTTAACTGT	CTGTCTTCGGTATGG				
				CTGGCCTCCTCTC	TGCCCAAGCGTATTG				
				ATCGGGCGGCGG	CTTCTAGACGCCTGG				
				ATTGCCGACGTTC	TCAAGGATGCCTTGC				

				GGTGCGCTGCTTC	CTGATCTCAATTGGT			
				AAGTGTTGGGGG	TGGCTGATCTGAGAG			
				TTCGGACATCGTG	GGGCTATTTCTGTTA			
				CAGTTCAGTGCAA	CAGCCCTTGCTAATC			
				GGTTCGTGCTGCG	TCATGGACCTTTGTG			
				CAGGTGATCAAA	ACTTACTCGAGGGGA			
				GCTTCAGTTCGGT	TGGTTCTGCAGCCGG			
				GTGGGATCGTCTT	ATGTTCTGATGTGC			
				GGCCCTAAGCTG	ATACCTGGAGGTTCA			
				GTTCAACAAGCCTT	GTGAATCGGGTGAAT			
				TGGACTCTCCTAG	ACTCGGCTAGTTCAG			
				ACGCGCCATGGA	CTTACAGGGCTTTGC			
				GTGGCGCAGGGT	TTCTTGGCTCTGTTAG			
				CTCTTCGCCGATC	CTTTGAGCCTGCTGA			
				CCGGCCATCCAGA	ACGCATTTGGAAATC			
				GGTCCGGTGTGG	CTGGGCTCCAGGAAA			
				CTGATCCTACCTT	GTGTAAATTCTTCCTC			
				GGGTGGCACGAA	TGGCTTGTGGAGCAC			
				GAAACGCCGCCA	GATAGATGTTGGACG			
				TCGTACGCGGGG	GCTGACAGACTTGCT			
				TCGAAGGCCTAA	AAGCGTGGCCTGGAT			

				GACCGCCCCGTT	CACCCAGAGTCCTGC			
				GCACCTTTGGAG	CCTTTATGCGACCAG			
				GAATCCAGTGAA	ATGGAAGAACTATC			
				GTTGCACCTCCAC	AACCATTGCTGGTT			
				CCGGTTCCTCTTC	CAGTGCAGTTTTGCA			
				CCTAGAGCCGGTC	AAACAGTTTTGGTTC			
				AGGGTGCATGCC	AGCCTTCTGCGGTCG			
				CCTCAACCGGTTG	GGTGGTTTCCAGGAG			
				TTCCTGATATGGT	CTCTGCCCTTCCCATG			
				GGAAGCCACACA	ATGACGTCTGTTTTG			
				TCTCTGCATTTTG	AAACCTGGTGGAGG			
				GACCGTGCGGAT	AATAGTGCTGCCAGG			
				GCGCTTGCTAGA	GTGGGCATGCCTTTA			
				GAAGAACTGCGG	TCCAAGGGATTCAAT			
				TTGCGACGGTGT	TCTCTTGTCATTTTGG			
				GCTCTGTTCGTTG	GTGCATGGACTTTGT			
				TTGTCACTGGTTC	GGAAGCATAGAAAT			
				AAGACCGAGGGT	AGGTGTGTTTTTGAT			
				GGCTGCTGAGAC	AACTGTAACCCTAGC			
				TCTTGCTGATGAG	GCTGCGGTGGTGTTG			
				GTGGCTGCTGGTT	GCAGCTGCAAGGGA			

				TCGGGCTGGATG	GGAGTTGATGGGTT			
				CGTCGTCCTTCTC	GGCGTATGGCTGGA			
				CGTTCATCGTTCG	GCCAAAGCGCTCTCT			
				TGCCCAGAAGATT	TTCTTTCATTTAGAAG			
				TTGTCTTAATCCT	AGGTGTCTTAGGTGT			
				GAATTCGGAAGA	CTTAGTTGTTTGGGG			
				GTTGGCCAATCAG	AGTGTTTTTTGGTTTT			
				GTGTACAACAATG	GGTTTGCCAAATGTA			
				GTGCTGTTTTCAA	ACACCTTTGTACTGG			
				TTCTCTTTCTGGG	GGTTTTTTCTCCCCAT			
				TCTTCAAGTTCCT	TTTATTCTTCTTCTTA			
				GAGATGGTCTAG	ATACAAAAATACGCA			
				ACTTGCCCATGCC	GCTCTCCTGCGTATTC			
				GAGGTGGTGGCC	AAGAAAAAAA			
				TTCCATCTCCTGT				
				TCTTGTGGAGTTC				
				AAAGGGATTCCA				
				ATTCATGCTTGGG				
				ATTTGGCTACTGC				
				TCAGAATTTGTTG				
				AGGGATCACTGC				

				ACTGATGTGGAG CTCCATCCTGACA CGGTGAATCGAC ACGACTTCTCTC CTTCAAGGTTTT GGATGGTGCCGA CGTCCACACCTCA TTCCAAATTTGGT GGAGCTCCTTGTT CCGGAACCGTTG ATGATGACGCCTG GCTCTACCTCTCTT GAGAGAAGATTG CTCGCTTACCAAG TGA CTGCCCTTGT GTCTCTGATCTCT CCGCCGGTGCATA GCGCTGCTGCACC TCCGTCGCCGGCG TTTGATGATGATC AGGACAACGGGC					
--	--	--	--	---	--	--	--	--	--

				CTCGTCGGCGCA GGCTCAACAGTTC AGTTCAGATGGG GCATGCTGAGTCC GGGGAAGCACTT GCAGCGCGAAGG CCCGTCCGTGAAA GGCTGGGTGTTG CCCCGGTCTCTGT TCGACATGTGGA GGCTGGATTGTCC GTGAAAGACTCG GCGCCTTCCGTTG CCTTGGTTACTGG GGACACGCATACT GATGTTGTGGAG CCGGTGATTGGCT GCATGGATGCTG GGTTGAGTCTGCC CCTGGAGAATGTT CATGGGCCTGGC					
--	--	--	--	---	--	--	--	--	--

				ACTGAGACTTTGG AGATTCCAGGTGC TTGCGATGGCCCG GGTGGTGAAGTC TTCTTGTCCATGA CTGATGTTGACTC TTCTGTTCCGGCG GTGGAAACTTTGA GTGCATTTTCATC CCTCCCGAGTGGC CCTGCCATCGCTA CCCGTGTGGGGA CTGTGTTGAAGCC GTCCATGAAGGCT GTTGGGGCAGGG CTCAGGAGGTGG CCGTGTTCCCAA TCAAATATTTAG AAGGAAGTGCAG AGGGGCTGCTGC TTCACAATCCCCT					
--	--	--	--	---	--	--	--	--	--

				CTCTCGGCTGCTC AGTCTGCGACCAC TGCTGATTTTCATC GCTAGAGTGTCCA GGCCGGCCGCTC GGGTCCTGTGTG GGCCGCAGAAGC GGCAGCGTCGCC CCAGGGCACCTTG TGAGGGGGCTCT ACCTCGGCGGAG CAGGCGCTTGGCT GGAGCAGATGTT GAAAAGGTAATG GCCCCAGCACAA GTTTCGAAAGGA AAATGGAGGATT GCTAATGAGCTCG GCCTTCTTGGTGG GGATCCATCTGAA TCGCTGTCCCTGG					
--	--	--	--	---	--	--	--	--	--

				AGGTTCTCGATAA CTATGCCAAGGTT TTCGGTGAGCGG CTGACTGATCTGC AAGTTAAAGCTCT TGCGGCTCTCTTC AACTGGACTGTCC CAGAAGACTTGC AGGGTGCAGTGG TGCAGGCCTGATC GCCTCTGTGTGTA TTG						
	L2- 2_SSa	GCF_00023 3375.1	Salmo salar	TACATTACAACGG CTTGATTTGTTTG ATCTGAGCAATTC TTCTTAGCTAGCT ACATAGCCGTCTT TGTATCAAAGATA ATTGTGTAGTTTA GAGTAATTAGAG TAATTATCGAGGC	TAGGATAAAGTAATC CTTCTAACCCCCCCCC CCCTTAAAAGATTTA GATGCACTATTGTAA AGTGGTTGTTCCACT GGATATTATAAGGTG AATGCACCAATTTGT AAGTCGCTCTGGATA AGAGCGTCTGCTAAA	MCLSIPVLSPLHRPYKRSTPRGRCHSNL VVPARTTHVEFQVSGSLWNCRSAANK AEFISAYATLQSLDFLALTETWITTDNTA TPTALSSSAYVFSHTPRASSQRGGGTGI LISPKWTFSLSPLTHLSISSFEFHAVTVTS PFKLNILIIYRPPGSLGEFINELDALISSFP EDGSPLTVLGDFNLPTSTFDSFLSASFFP LLSSFDLTLSPSPPTHKAGNTLDLIFTRCC SSTNLIATPLQVSDHYLVSFSLSLSSNTS	cd0 165 0	519	780	

				TAGCTATCTTCGT CCTCCTAACGTAG TCAACACCGCTGC TAGCTAGCTAGTC ATCACTGCTAGCT AGCCAACTTCTAC CGACTAGCAGCA CTGTAGAACTAT TACATTACAACGT AACGACTTGATTA GTGTGGTGTTAGT GTTAGTTAGCCAG CTACATAGTTGTC TTTGCTGTCTCTG TATCTAAGATAAT TGTGTAGTTTAGA GTAATTATCGAGG TGAGCTAGCCAG CCGCGACGCGGC GCCAGACTTAGTC AACACACCTAGTC	TGACTTAAATGTAAA TGT	HSAPTRMVLRRPNLRSLSPATLSSSILSS LPSAQTFSNLSPDSASSTLLSSLSASFDFL CPLSSRPARSSPPAPWLDDSLRAHRTGL RAAERKWRKTRLPADLASFHSLLSTFSS SVSAAKATFYHSLQASASNPRKLFATF SSSLNPPPPPPSSLSADDFVNHFEKKV DDIRSSFAKSNDTAGPAHTALPCALTSF SPLSPDEISRLVTAGRPTTCPLDPIPSSLL QTISGDLLPYLTSLINSSLTAGYVPSVFKR ARVAPLLKKPTLDPSDVNNYRPVSLLSFL SKTLERAVLGQLSCYLSQNDLLDPNQSG FKTGHSTETALLCVTEALRTAKANSLSSA LILLDLSAAFDTVNHQILLSTLSELGISGA AHAWIASYLTGRSYQVAWRESVSAPRA LTTGVPQGSVLGPLLFSLYTKSLGSVISS HGLSYHCYADDTQLIFSPPSDNQVAN RISACLADISVWMTDHHLKLNLGKTELL FLPGKDCPFHDLAITVDNSLVSSSQSAK SLGVTLDNTLSFSTNIKAVTRSCRFMLY NIRRVRPCLTQEAAQVLIQALVISRLDYC NSLLAGLPACAIKPLQLIQNAAARLVFN			
--	--	--	--	---	------------------------	--	--	--	--

				ATCATTAACCCAC TGCTAGCTAGCCA ACCGTTACCGACT AGCAGCGCTGTA GATACTAATACTT TACAACGGAACG ATTTGACTAGTGC AGTGTTAGCTAGC TAGCTACTAGTTG TCTTTGTCATAGC TTGATAATTGTGT AGTTTAGAGTAAT TATCGAGGTTAGC TAGCCAGCTATTT CCGTCCCCCGCGA CGCCATTTTTCCT AACCCAGCCAACT ATTACCGACGAGC AGCATTGTAGAA ACTAAATACATTA CAAAGGAACGTC	LPKFSHVTPLLRSLHWLPVEARIRYKTM VLAYGAVRGTAAPPYLQALIRPYTQTRAL RSSTSGLLASLPLRKHSSRSAQSKLFAAL APQWWNKLPDARTAESITTFRRHLKP HLFKEYLG			
--	--	--	--	---	--	--	--	--

				TTGATTAGTGTTA TGTTAGCTAGCTA GCTACAAAGTTGC TTTGTATCATGAC AAGGTGTAGTACT GAAACTATCGAG GTTACCTAGCCAG CTACACGTTCAAA CAAAGTCAACAAC GCAGCCACTGCTA GCTAGCCTACTCC ACCAGCCAGCAG TACTGTATCATTTT CGTCATTTTAGTC AATAAGATTTTGG CAACGTAAGCTTA ACTTTCTGAACAT TCGAGACGTGTA GTCCACTTGTCAT TCCAATCTCCTTT GCATTAGCGTAGC					
--	--	--	--	---	--	--	--	--	--

				CTCTTCTGTAGCT TGTCTACT					
	Dong- 1_MM a	GCA_00048 4575.1	Mesob uthus marte nsii	TAACCCCTCCTGA ATCCTGCTAAGCT GCCTCAAAGTGG CGGAGGGGTGTA CGTGGGCGGAAT AGGCGATGCAAG TCAGAAATTATTT TCTAGTAGGGTCA CCCTGGGCTAGA AGGCTATTCCATT ATGCTCAACTAA	TAACTCACAATTA CATCACAAGTAGCTA TGTCTTGACTCTTGTC CGACATAGCACTTTG CCCCGGAGTTATTCC GTGCTCACATTGATG TACCATAATATAATA ATAATAATAAATACA	MLAGVPTLVVFFQYAMAQSRKRKHTG SDVSRVNVKVRVCELGRGSHSESLQSP ATGGESSTNPSLATAIQIPRVGNGSPHQ SPTSQSVSLETQPETTKAGKPRQRMKW TTDMNEFIMRTYYTITKLETDITGYRREL YDKFTSRYPNITVTEQRIADQRRTHHRN LIPHTLESIKTQVAQQISQHETDSHTQI EHDVEQTPEITYIDDNTQPINNTDLDI NKLDTLQETLTKFTGTDPTQRPRIPKL HNSQQLNNTLHCMNTHILPKLINAETT LIEHTLIYAAAVTVTTHLGLKTLPNKIGT NTKHHLKPAWQQRLEKQINEIRTELGR TQYTKNNTSTKLMKSIRKIFRKLKTHTR HENHNTHIKEYIDTLKQKLAASQRLKR YKTS LARKLDNTLFTNNEKTFYRNINTK DPKQTQNREIPTQQSLTSFWSALWAV PVTHNQRASWIAEEETRDLVPEMPYT TITTEELNTIRYTHNWKAPGVDGIQNF WYKHFTCLHDLLAKHLTHLIEPKDTPS	cd0 165 0	534	805

						FLTHGITFMKPKSTNTGNPSQYRPITCL PTLYKILTSCITNKIYAHIEENNILTEQQK GCRRSHMGCKEQLIIDTTILKQVHRQH RHLHAAYIDYKKAFDSIPHSWLIQTLQL YKINTKYITFLQNIQHWRTTLHITINN NHVATEPIHIRRGIFQGDSLSPLWFCLA LNPLSNKLNTSNCGYDIRQNQTTHYIHS HLFYMDDIKLYASTQTQLQQLQITHTF SKDICMEFGLDKCKTLHIEKGVKKGGF TLENGDTIESMEEGEVYKYLGFQQARLT EQKKTQTLISKLYTNRLNNILKTHLNGK NTVKAINTFaipilTYSFGVIQWTRTELE SLQRKTRTTLTkYQYHHPKSAIERTTLPR KEGGRGLIDIPHLHDRQIYNLREFFHKK QHTSQLHKAIVLADRnitPLNLQNNTPC QIDNTAHIQEKkTSWCKKELHGRHPHD LQKHfIDAKASNSWLVRGDLFTETEGFII AIQDQVINTKNYRKYLKDSTLTNDTCR KCHIAPETIQHITSGCKLLAQTDYLHRH NQVANIIHQKLATKHKLINTLTPYYKYIP DTILENSTHKMYDRTILTdkTIHHNRP			
--	--	--	--	--	--	---	--	--	--

						DITLIDKQNKVTYIIDISIPDTHNIQDAFT QKISKYTELSQEIKRLWGMQKVVTVPIL SSTGVIPHTLHKSLKTLDLHPKTYITLQK ASLLNTCNIVRKFLSTDTDS			
	SR1	GCF_00023 7925.1	Schisto soma manso ni		TGATTACCAAATTCT TTTTTCCCTCTATTA TCGTTCATATACCTA GGTTCTGCCTGGAG GTATTGGTGATCCAC TGCTACCAGACACGG AAGCCCGTTAAGCGA AAGCTTCTTCTATTCC GTCCTCAACCATTG AACCATTG	MEFLVDDL SILAPLGKSDHAVLSFSFVSK TELRYPSTNKRWNFKRLNVLALQDYLQ QVDWDVHPQLEVDAHWD FLLHTILCA TEHSVPKMVPKSYKQPPIIKNRTRRLSR KRHCWAEYKRTDNNGAYRQYKHIRNIC TKAIREDR LQFQTKLIDKFVSNPKSLFSY AASLRQGKTGVSQLLGPNGPTNNDSD AANLLAEQYSRTFQLTHINHTDESFTCT CTGLSEVDLSADLVLRKLQHLRKDTSPG PDMVHSAVLREAASILATPLSVMFAHS LSRGKLPEIWKLAHITPIFKGRRSEPSS YRPVALLSIPSKIMESLVYDGILEYLSSSK FFSPQQHGFRKGHSCMTNLLTAVDRW TTILDRKGKVDVIYLD FSKAFDRVNHAC LIKKLGR LGIKPP LIDWLSSYLENRHFKV RVNFTLSQAMECPSGVPQGSILGPLLFL IYINDLPQQVSSDLLLFADDVKLWREIR	cd0 165 0	288	537

						NHNDILVLQEDLTRLQSWADDNGLTF NTSKCKVVHLRHVADYSYNLGNPLEV SQVEKDLGVLVPYDLKSYANCDKNASQ ANLALVTLKRIFGQFDGRTFHIIFNSFIRP HLEYGNIVFPPSLQKDKDTLERIQRRAT KSVRGLKFKPYEERLKSLNLYPLEYRRLR GDLLMTYSILNTSGHPLKHLLKLSHNTN LRGNTQKLETLSRSDCRHNFYSVRVV KCWNSLPTELVQATSQESFKRKLDLFLR TKDNILL			
	CR1_A C_1	GCF_00009 0745.1	Anolis carolin ensis	TTGATCCCAGAAG GGCAGTTGGTTCT CGCAGAAGCAGA GGCAGCAGGAGG ACATTTTTGCTCC CTGGCTTCAGGTA GGATTGGCTAA GATTTGGCTAAGA TTTTAACTGAGT GTGGGCTTGGCTC CTGTGGTCAAAGT	TGAATGCGATTTCCT GCTTCTTAGCAGGGG GTTGGACTAGATGGC CCATGTGGTCTCTTCC AACTCTACTATTCTAT GATTCTATGATTCTAT GATTCTATG GATTCTATG	MQQKQNIPITCTSCGMFRFFTQKLDNY ICSKCKQMTLMEQRIQQLEHRIKTLKDI QALELFLDTAQHAAVDQQPASHQH HQ DHDQEPYGEINSKVDNPQAWKEVTHR RRRRTRQPPQNSSAQLHLHNRFEILTS LYNQETHLVEDHSFLDTTQWTTLDQCA GDSPDGDSASPQSHLCNHANAPSPIID QEQQEHPLGETRSNRNIPWESNGLSD VSQWIVIDECTGDVEEEDNTLHLHDSL QQEHSSGDIHTVLHKRDPVNPQRKQVL VVGDSLLRGTEAIIISRPDGMARETCCLP	cd0 165 0	1008	1259

				CTAACTGTTGTGT GACTGTTGTGTTG AAAGAGAGCCAG GTACTTAAGTTGA TTGACAGCAGGC ATTGTCCCCTGTT AATTGAGTTTCTG GGAAAGCTTTATT TGCCAGTGTGAGT TTCTGCCAGAGCC TGATCCCAGAAG GGCAGTTGGTTCT CGCAGAAGCAGA GGCAGCAGGAGG ACATTTTGTCTC CTGGCTTCAGGTA GGATTGGCTAA GATTTGGCTAAGA TTTAAACTGAGT GTGGGCTTGGCTC CTGTGGTCAAAGT		GAKIHHITQRLSRLLKPHHPPHMLLIHV GTNDTARHTFQKITNDFRALGTLKLYN VHVIFSSLLPVVGHGSTRAGKIVQVNN WLRKWCQEEHFGFLDHGLLFQEDGLL ASDGVHLTQVGKHLFAHRLTNLIRRTLN PWSTLPRGWTTGKRWGASHTSRKTSF CTQTHKPHQAHFKLDPLGEGNNSLANT ILPTTTGNRRKAKRRAAQTQQGPSTEST IIPNKQFEGRSQGLTCLYTNAQSMGNK QDELQLLAQHHTYDVIGITETWWDDS HHWNLTIEGYNLFHRNRTKGRGGGVA LYVKNSYVAEEMQDCNPGNQLESIWIR IKGTGTQKDLVVG VYYRPPSQDEGLDE AFCQQLTKQAQRDIVVMGDFNYPDI CWKTNSAKSTKSNKFLTCLADNFMVQ KVEEATRG SATLDLILT NVEDLINTVEVV GSLGASDHVLLQFAIQRNAETKTSQTRI LDFKRADFQKMKELLSGIPWTPILKNKG VKDGWEEFFKSEILKAQM QTVPTKKKNK TSAKKPEWMSKELLTELKLKSDMHKK WKRGEITKEEFKRIANTCREKVRKAKAQ			
--	--	--	--	--	--	--	--	--	--

				CTAACTGTTGTGT GACTGTTGTGTTG AAAGAGAGCCAG GTACTTAAGTTGA TTGACAGCAGGC ATTGTCCCCTGTT AATTGAGTTTCTG GGAAAGCTTTATT TGCCAGTGTGAGT TTCTGCCAGAGCC TGATCCCAGAAG GGCAGTTGGTTCT CGCAGAAGCAGA GGCAGCAGGAGG ACATTTTGTCTCC CTGGCTTCAGGTA GGATTGGCTAA GATTTGGCTAAGA TTTTAACTGAGT GTGGGCTTGGCTC CTGTGGTCAAAGT		NELRLARDIKNNKKGFFAYVGRKRKKKE AIGPLQGEDGVMATGDREKAELLNAFF ASVFSEKESHLQPQQHGMDEGLGEIQP QIGKQVVQEHSLNEFKSPGPDQLHP RVLKELAEVISEPLAIIFESSWRTGEVPA DWRRANVVPPIFKKGKKNDPNNYRPVS LTSIPGKILEKIIKEVVCEHLETNAVIAN QHGFSTKNKSCQTNLISFFDRVTSWVDT GNAVDVAYLDFSKAFDKVPHDLLANKL VKCGLDKTTVRWICNWLSESTQRVLTN ASSSSWKEVTSGVPQGSVLGPVLFNIFI NDLDEGLEGTIIKFADDTKLGGIANTPE DRSRIQNDLRLERWAKTNKMKFNDR KCKILHFGRKNGNQRYRMGDAWLDSS VCEKDLGVLVDNKLNMSQQCDAAAKK ANGILACINRGIASRSREVMPLYSALVR PHLEYCVQFWAPQLKGDVDKLESVQR RATKMIKGLENKPYEERLKELGMFSLQK RRLRGDMIAMYKYVKGSHREEGASLFS AALQTRTRNNGFKLQERRFHLNIRKNFL			
--	--	--	--	---	--	--	--	--	--

				CTAACTGTTGTGT GACTGTTGTGTTG AAAGAGAGCCAG GTACTTAAGTTGA TTGACAGCAGGC ATTGTCCCCTGTT AATTGAGTTTCTG GGAAAGCTTTATT TGCCAGTGTGAGT TTCTGCCAGAGCC TGATCCCAGAAG GGCAGTTGGTTCT CGCAGAGGCAGC AGGAGGCACCTTT ACTAACTATCCTT TGCAGTACATACA GGAAACAAAGCA ACATAAACAAATA CACAAAGAGGTG GTTTGTTCAGTT TGCACATAAAGTG	TVRAVRQWNSLPGAVVEAPSLEAFKQ RLDGHLSGVL			
--	--	--	--	--	--	--	--	--

				CGTGCATATTACT TAACTGTACAAAG ATCCCACTTGGCC ACCACGCTCACCA AGAG					
	Penelo pe- 2_Psi	GCF_00023 0535.1	Pelodi scus sinensi s	AGAAACTCGCAC AGATACAGACAG ACATCATCTTTCT ATCCAAGTGCAA GAAA	TAACAAGTCTTTGAA CTTTCTCACCCATTAT CAAGACAGTTTCCCC AATTATCACCTGTAA TACCATTAACCTACA AACATCCCACTCTCCC TACCTCTAATATCATC AATTCACAGACACTT ACCTTCCTTCCTCCCC CTTCCCACCCCCCGC CTCCTCCTTCTGTTCT GCAATGTGATTTGTC CTTTTCATATTTGTTT ATTTTTTTTAATTGTA TCCTTTGGTATATATG GTTGTGACTACTTTCT	MDIVPKGLKVNNPLRSTYFTDYSERLCL TLSKKLRNHLISILYKKQKTIKNDISNLDR LISRQTSTPTDTLQDFCSTRQEIIYYTHFY SLQQRKDNKFSNLLHTTGYNCLNPP DNIVNLSSYKLNPAEESVLSRGLSFCPTS PTNWIQFCGDLEAFFRRLRLREYFQTTT EHQSDLPDPYQHQRKNNSIWTPPDG RNYNLDFYIHSFRNNAQTDIIRKQQAT HNLSRAEHHAIQSLKNNLDIVIKPADKG GAVVIMNKADYNQEATRQLSNTTFYKP LPSDPTLDYQKKLHCLLKSPTATREQIL TESPSDPRPGLFYLLPKIHKPGHPGRPIIS GIGTLTTGLSSYVDTLKPFVTNTPSYLR DTTDFLRKLQNIDNLPNNTILATMDVE ALYTNIPHEDGLQAIRNTIPEDTTANLIA DLCNFVLTHNYFRFENNLYLQISGTAM	IPR 000 477	305	558

					TCCACTATTTGATCTG AGGAAGTGGGTCTG GCCCACGAAAGCTCA TCATCTAATAAACCA TCTTGTTAGTCTTTAA AGTGCTACATTGTCC TGCATTTTGCTTCAAC TACCCAGACTAACA CGGCTACATTTCTATC	GTRMAPQYANIFMADLEQRFLNSRPLS PLLYLRYIDDIFMIWTHGQETLEIFHRDF NNLHPTINLSLDHSTREIHFLDTTVQINN GKLDTTLYRKPTDSYSYLHASSSHPEHTT RSIVYSQALRYNRICSNPTDRDQKLQDL YQAFINLNYPPRETKKQIERARRIPRNHL LQDRPKKTNNRTPLVITYNPQLKPVQHI INKLQPILEQDTILKEALGDRPIVSYRQP PNLKMILTNNHHRTYHTNTNPGTFPCNK PRCQLCPHIHSADTIIGPNQVSYKIKNTY SCASRNIIYAIMCRKCPSAMYIGQTSQT LRQRINAHKTDIRQDHKEKTVSCHFNQ KGHSLNDLATCILLQRPFTSALERESSEL SFMLKFDTYQPGT			
	Penelo pe- 2_Psi	GCF_00023 0535.1	Pelodi scus sinensi s	AGAAACTCGCAC AGATACAGACAG ACATCATCTTTCT ATCCAAGTGCAA GAAA	TAACAAGTCTTTGAA CTTTCTCACCCATTAT CAAGACAGTTTCCCC AATTATCACCTGTAA TACCATTAACCTCACA AACATCCCCTCTCCC TACCTCTAATATCATC	MDIVPKGLKVVNNPLRSTYFTDYSERLCL TLSKKLRNHLISILYKKQKTIKNDISNLDLDR LISRQTSTPTDTLQDFCSTRQEIIYYTHFY SLQQRKDNKFSNLLHTTGYN SNCLNPP DNIVNLSSYKLNPAEESVLSRGLSFCPTS PTNWIQFCGDLEAFFRRLRLREYFQTTT EHQSDLPPPYQHQRKNNSIWTPPDG	IPR 000 477	327	504

					AATTCACAGACACTT ACCTTCCTTCCTCCCC CTTCCCACCCCCCGC CTCCTCCTTCTGTTCT GCAATGTGATTTGTC CTTTTCATATTTGTTT ATTTTTTTTAATTGTA TCCTTTGGTATATATG GTTGTGACTACTTTCT TCCACTATTTGATCTG AGGAAGTGGGTCTG GCCCACGAAAGCTCA TCATCTAATAAACCA TCTTGTTAGTCTTTAA AGTGCTACATTGTCC TGCATTTTGCTTCAAC TACCCACAGACTAACA CGGCTACATTTCTATC	RNYNLDFYIHSFRNNAQTDIIRKRQQAT HNLSRAEHHAIQSLKNNLDIVIKPADKG GAVVIMNKADYNQEATRQLSNTTFYKP LPSDPTLDYQKKLHCLLKSPTATREQIL TESPSDPRPGLFYLLPKIHKPGHPGRPIIS GIGTLTTGLSSYVDTLKPFVTNTPSYLR DTTDFLRKLQNIDNLPNNTILATMDVE ALYTNIPHEDGLQAIRNTIPEDTTANLIA DLCNFVLTHNYFRFENNLYLQISGTAM GTRMAPQYANIFMADLEQRFLNSRPLS PLLYLRYIDDIFMIWTHGQETLEIFHRDF NNLHPTINLSLDHSTREIHFLDTTVQINN GKLDTTLYRKPTDSYSYLHASSSHPEHTT RSIVYSQALRYNRRICSNPTDRDQKLQDL YQAFINLNYPPRETKKQIERARRIPRNHL LQDRPKKTNNRTPLVITYNPQLKPVQHI INKLQPILEQDTILKEALGDRPIVSYRQP PNLKMILTNNHRTYHTNTNPGTFPCNK PRCQLCPHIHSADTIIGPNQVSYKIKNTY SCASRNIIYAIMCRKCPSAMYIGQTSQT LRQRINAHKTDIRQDHKEKTVSCHFNQ			
--	--	--	--	--	--	--	--	--	--

						KGHSLNDLATCILLQRPFTSALERESSEL SFMLKFDTYQPGT			
	REX3	GCF_00277 5205.1	Xiphop horus macul atus	C	TAAGCGGAAGAAGA TGGATGGATGGATG GATGGATGGATGGA TGGATGGATGGATG	MTLGRSSMIDFVIVSSDLRPYVLDTRVK RGAELSTDHYLVVSWLRRWGTWLSLG PCSVPPLSRRPIGAVAARLSVPVAAATL EPAGGHLLVRDAVRLKKESYRAFLACGT PEAADGYRRAKRHAARAVAEAKTRAW EEFGEAMEKDFRTASRRFWSTIRRLRG GKRCSTNTVYSGDGVLLTSTRDVVGRW AEYFEDLLNPTNMPSTEEAEPGDSGLG SPISGDEVAEVVKLLGGKAPGVDEIRP EFLKALDVVGLCWLTRLCNIAWTSGAV PLDWQTGVVVPLFKKGDRRVCSNYRG VTLLSLPGKVYSGVLERRVRRIVEPRIQE EQCGFRPGRGTLTDLQLYTLGRVLEGAW FAQPVYMCFVDLEKAFDRVPRGALWG VLREYGVPGPLIRAVRSLYDRCQSLVRIA GSKSGSFPVRVGLRQGCPLSPILFITFM DRISRRSQGVEGIRFGGLRISSLLFADDV VLLASSDRDLQLSLERFAAECEAAGMGI SASKSEAMVLSRKRVECLLRVRGGVLP	cd0 165 0	277	532

						QVEEFKYLGLFTNGGRREREIDRRIGAA SAVKRALYRSVVVKRELSQKAKLSIYRSI YVPTLIYGHELWVMTERTRSRIQAAEM GFLRRVAGLSLRDRVRSSVIREGLRVEPL LLHIERSQLRWLGHLVRMPPGRLPGEV FRARPTGRRPRGRPRTRWRDYVSRLA WERLGIPPEELEE VAGEREVWASLLKLL PPRPDLG			
	CR1-3_IS	GCF_00289 2825.2	Ixodes scapul aris	TCTATTGATTGCT TAGTCATATCTGA AACGAACATAAA GGACTGCGACGA AGCATTATTCCAA AATAAAAAGTTTTG AAACGATTACAAA TAACAGGAAGAA CAAGCGAGGAGG TGGAATC	TGAACAAGTGTA AAA AAAGGGTTTGT TATA AGTGACTAGAGCTCT GTCGCTGCCCTACGC ACAAGCCGCTTAGGC TTATGTGGGGTATGA ATCTCTGTAAAATTG TATGTGGAAAAGATG AAATAAAGATTATTA TTATTAT	MVMVNKKWNINKSKKITFSRAEAVQI DIENEKEQIQILAVYRAPQGNIAEFIDDI DNWIKVENSKNLIITGDINIDITKQSNLV DKYLNCLAEHGLQSMIKDITREEILGLQL TRSCIDHICIRSKKYCKAHVIAEKLADHYF VGIAIPELGIPESERDQVTYINNAKVVDK YVREINWESIANEEDPEIYNKIKNAFTDI YKKSSEQVTVKKRKTGNQWVTKEIDQLI RKKNKAWRLLKSSPESSQRKEEYRSIRN KVTSIIKKEKRN YFWKKFEDTKDPSQCW SVINQVLGRNTRLSIDQTIKKNFGTYCN DEKIANDFNNEFIQSIELLCNHGYQNET ETRARIGNSMYLP EMT EEDLVEIVSNM	cd0 165 0	415	680

						KTNKAVGIDSIRPRDIKENVNIKTVLLAL YNKSLAIGKVPSGLKTAVVRPIYKKGDK KTLLNYRPVCILPVLNYIMEKYVHKALN GFINKYNLLTKTQYGFKTNSSTITLLEDFS DFVNKKLDEGNYVLALFIDFSRAFETVR HETLIDRLFSLGVRGKLLDWFKDYLSGR TQVVKIASSFSEKRLVKQGV PQGTVLGP ILFNIYINDIVNVHLNSVLFQFADDTVILY ADKDYTTAENKIQEDVYKLVKWKYKENQ IIINTKKTELMCFRSPQRRTQMEGNIKIH SDSCKNCNCEPLKLTTSVKYLGLYIDEHL KWNQHVEHICKRLRSVAAQLYHLRSCL PQPILKMIYHSLGKSVLVYGITYGHCAK YLQDKINSILNKVVKNLYVDENRPLHYM YQHLDIQKLSDLLELNIIHQHFSDKYKT KYIPIRNLRPKNLFQEPKVTKYGKAVR ASYVPRVFNKLPPRIEITKISELKRELKQ WFINRKELQLAEKSTRGWNLNCV VLC VLFCRELWSKGVYMSCFHFFFLFARN			
--	--	--	--	--	--	--	--	--	--

	RTE-1_PM	GCF_010993605.1	Petro myzon marinus	GCCTGTGCGTGTG TGTGTGTGTGTG GCAAAACAACCCC CTACCTCCTCGCG GTGCCGGACGGG CAACTCTGTCGTG GACAGTGCTGTG GGGGGAGGTCCA ATTTTGATTGGCT CCTCCCCTCATGG CATATGTGTCCAC CACAGGGGCTAA ACGGGGGCTCTG GACGGGTGCGAG GGGCTGGTAACC CCTCCACCTTCAA TTAATTACCACGA AACCACGAATTAA ACTGAAATATGTA ACTAATAAGCTTG TGCCGGAGGCCC	CG	MTSPGQTPIGGARRLNLYPRGSSRNNE KPLRVATWNVRTLTAERLPLLCRELD YRLDLCGIQEVWRTGSGSCHHEGWTV LYSGQQTAKQQGVALLNKRMGAW ERGGEVWEAVSPRLLWARLPINRRRG MVVVVTYAPTEDEHPWRDAGKADES DAFYDLLSQVLAKVAPRDRVIMLGDFN ARVGHDAGTWSGVIGGHGPDQLNNN GRRLLDFCAAHGLVITNTLFRHRRHTTS WMHAGSKQEHLDDYIVTDQRMRAHV LDTRTYRGADLPDHLVICKMRLPFFH SGGGCTPRTPFQPKVAWSQLSDESKR EFHHRLRLGLASSPTPPDVEGKWARFR TVTHAAAMAALGTRSRPPQRPWLTEE VFRMVEKKRQAHCHRLQNPTSENEEA YRSLRSEVRRARRCKDGWWSRVAEE MESASVARDHKSLFNSIKKVTCSATPITII RGKNGDPISDPQEQVCRFAEHFCEVLG EGKSLSQENIVAQSEDDPAEAAVPQILE EVTPEWLAEVPSVEEVLKAIRSLRPGK AAGRDDLPSSEMLREGGVCAQTALYDII	cd0 165 0	578	856
--	----------	-----------------	---------------------	--	----	--	-----------------	-----	-----

				CGGGGGGCCAGG CGTCT		TTVWTTGRVPQDWKDALVVPLFKKGD RTDCNNYRGISLLSVPGKVLARIIQHRLA RHAEERLAEPQCGFRKGRSCTDAIFSVR LLQQRCREFQQDLHLCFIDLVKAYDTIS RPGLWVVLRAFGVPDPLLTIIKDLHDGE RARVKLRGRESEFPVHLGVRQGCPLA PTLFNIYFDHVVREAFNSCTGGISIWYRY NGRLIDARVRSRDGSLLNHVMYADDL VIAAHSEEELETLLRLERATKRWGLTIS PTKTKHLPGYFVPSDTPPPQLPIGDGAV VERVESFPYLGSVLMTGSSLTAEVSLRIS RAAISFHRLRNAVWKLPGLSLRTKLQVF SATVIPSLLYACETWTPLMEHLRRLETFR LACLSILGLTRLD CVRSSQILERSGQGPI GELLRQARLRWLGHVARMPSHRMPK QLLFGRIEGAKRARHGLEKRWSDVVQE DVELSGLADDWYQRCQDRPLWRRLVK DATGQLEAVALQQRRAEERRSQQRA ERRVAKAQQVGTVGGTGGASASHLSQ STRGTWVCPVTSCQRAFDTS			
--	--	--	--	-----------------------	--	---	--	--	--

	Perere _Smed	GCA_00260 0895.1	Schmi dtea medite rranea		TTTACGTATGTAAAT CATGTTTGT TTTTAA ATATTATTAATAA TTGAAAATTTAAAT GCATTAATTTTAAA AAATTAGAATATAAG TAAAAAATTTAAAT TTAAAATTAATTTAA AGTGAATTGAATATA TTTTAAAAATATTTAT TTATTAAAAAATTAA ATTAATAAAAAATAA TTTGTTTAAATTAAAT GTTAATTCATAATATT TTTAGTTGGTTTTGA GTACCTGAAG	MIAVLSKYIVRKVFYWRYRLIKSLPNLFN YLFINHAAAVPIFKRFMKVYNSLKRKEL DISFLETCLLEKMIPSFYKMKLSDCLSSK DVQLIHRRTLRRMLESEKIRRTSIDKKLT NIGLMLNNATNEFCLQVCVEWCIKNSI VQIKSKENAHIKKLINLRKIFGISKFSVEN TVINLSKIVLNDDEERVLFKFGPNHCINR LDETKLCAKIESLYFHIARSMKGINGNYV KSSLSASTMEYIDKNKKQNIRDFMTLKN LAKKNINICKFDKNGGIVILDHVDVYVVK LNDLLNSPQFVKLPKGRKNGKLPPLKDE EFVQKMLNGFIEKKLIDEEVGKSIRPIGS QPAKLYGLPKIHKDGVMPRPVLSMVGT AQYKIAKFLDGLLKPLMRSEFECKDSFEF VSCIAKLQKRSMNDVMVSFDVCSLFTN VPLVETIDLCCALWNENDSEHHILDRRA FRKLLEFATSNVNFLFNDEWYQQIDGV AMGSPLPTMASIFLASLEKKIASFQFE KPFVYKRYVDDIFLIFENQKHVEPFLQF MNSLHKNIVFTCETERKSSIAFLDLLIQR KDKQYETEIYRKPTDTGLYTSPESFCEFK	cd0 030 4	474	565
--	-----------------	---------------------	-----------------------------------	--	--	---	-----------------	-----	-----

						YKRNMVKGLIYRSWALSSTFANSVKSV DKLVELLIKNGYSKSFLLLMIKETVDKLIK NPKDPCCGKSECNLSDGCELSVDLGDC KNGYKKIEPKYVLVLPYSEGFTNYKRKLG KLIGNLEYKIVSNSCKVRNMFINKSKTPV GLCSDLVYQFTCNGCNATYIGETSRHLC TRVLEHCRLKGLTNISEHNRGCKSDIGM SDFRILLRSFNSYWERVICEALLIRSLDPK INVQSAVTTNVLNVFK			
	L1- 3_MD	GCA_00000 2295.1	Mono delphi s domes tica	GACAGAGTGAAA CAACTTTCAAGAA AAGAAAAACCCA AAAGGTAGGCAG AGAACATCTGGG GCACTGGGGTGA GAGGAGAGCAGA GCCAGCAAGAGG CATAGCAAGAAG TGAAGAGCAAGC CACATACCCCACC ACACCCCACATCT	TAATTATTAAGAACT ATTTTGCCCAATTATA TGGCAATAAATATGA CAATCTAGGTGAAAT GGGTGAATATTTACA AAAATATAAATTGCC TAGATTAACAAAAGA GGAAATAGAATACTT AAATAATCCCATCTC AGAAAAAGAAATT GAACAAGCCATCAAG GAACTTCCTAAGAAA	MRKKLKELKEEIDSKTILVGDLNPLSELDKSNQKINKKEVREVNE MLEKLELIDIWRKLNDRDKKEYTFFSAVHGTYTKIDHVLGHRNIAN KCRKAEIINATFSDHNVIKIIINKGPWKGKFKINWKLNNLILQNW WVKKEIETIMDFIKENDNEETTYQNLWDTAKAVLRGKFISLSAYI NKIEREEINELGLQLKKLEKEQIKNPQIKTKLEILKIKGEINKIESKRT IELINKTRSWYFEKTNKIDKVLVNLIKRKRKRIKLTVSKMKRVISPL MKRKLRL			

				GCTGTCCCAGGTT	AAATCCCCAGGGCCA	
				CATGAACCTAAGC	GATGGATTCACAAGT	
				CCAAGCCAAATCA	GAATTCTATCAAACA	
				GACCCTGAGATCC	TTTAAAGAACAACATA	
				TGCTGGGCAAAT	ATCCCAATACTATAC	
				AGTGGTACAAGC	AAATTATTTGACAAA	
				CAACCCATACCCC	ATAAGCAAAGAAGG	
				TTGAAATTTCAAA	AGTCTTACCAAATTCT	
				CCTGTGGAGAAG	TTTTATGACACAAAT	
				TCCAAGGCAATCT	ATGGTACTGATTCCA	
				GGGCTGAAGTGG	AAGCCAGGCAGACC	
				GCTGATCCATGTG	AAAAACAGAGAAAG	
				GGCCCAGAGCAA	AAAACACAGACCAA	
				AGCCAGGAATCC	TCTCCTTAATGAACAT	
				CAGTCTGGGCTTG	AGATGCAAAAAATCT	
				GGATAAGGACAT	TAAATAAAATACTAG	
				AATTCACAGTTTG	CAAAAAGACTACAGC	
				GGGTGAGGTGAT	AAGTTATCACAAGGA	
				AGTACTCAAGGG	TTATTCACTATGACCA	
				GAGCCTCACTGG	GGTGGGATTATACC	
				GATCTTTTTGCCT	AGGAATGCAAGGAT	

				AAAGCTCAGGGC	GGTTTAATATTAGGA	
				AGAGCTCCAAGA	AAACCATCCACATAA	
				CAGGGCAATCAA	TTGACCATATCAATA	
				GGGTGTGCCTGA	ATCAAACCAACAGAA	
				TTGGATCAAGTAC	ATCACATGATTATCTC	
				ACAGTGAGACCA	AATAGATGCAGAAAA	
				AAAAACTTGAGCC	AGCCTTTGACAAAAT	
				TAAGCCTGGGGC	ACAACACCCATTCTT	
				TAGAATTTGATCA	ATTGAAAACACTAGA	
				GCCCCTGACCTGA	AAGTATAGGAATAG	
				TCAAGATCAGAGT	AAGGGCCATTCTCTCA	
				GAGATCAAAAGC	AAATAATAAACAGTA	
				TTACACCCAAGCC	TATATTTAAAACCATC	
				TGAGGCAAGTCTT	AGCAAGTATCATCTG	
				TAAGCTATCAGCA	CAATGGGGACAAGTT	
				TCTCAAGGGTCAA	AGAAGCCTTCCCAAT	
				TTGAATCAGCAG	AAGATCAGGAGTGA	
				GGGGAGGGGGCT	AGCAAGGATGCCCAT	
				CTCAGAGCTCTCA	TATCACCTCTATTATT	
				GCCCTCAGACCAT	TAATATTGTACTAGA	
				CAGGTAAACTGG	AATGCTAGCAGTAGC	

				TAAAAACCCAGA	AATTAGAGAAGAAA	
				AAATAAGCTGAG	AAGAAATTGAAGGG	
				AATAACATCAAG	ATTAAAGTAGGCAAT	
				GAAGGACTGAAG	GAGGAAACTAAACTA	
				TTAAGAGGGTAC	TCACTCTTGCAGAT	
				CCCAACTTCCCAG	GATATGATGGTATAC	
				AAAACAGAGCCT	TTAAAGAATCCTAGA	
				ACCTATAATTAGA	AAATCAACTAAAAGG	
				TAGGAAAAATGA	CTAGTGGAAATAATT	
				GCAAACAATCAA	AACAACTTTAGCAAA	
				AAAAGCACCTAAC	GTTGCAGGGTACAA	
				TCTAAAGAATTTT	GATAAACCCACATAA	
				TTATGGAGACAA	ATCATCAGCATTTCTA	
				AGAGCAAGGTAC	TATATTCCAACAAA	
				AGACACAGAAGG	ACTCAGCAGCAGGA	
				GGACAGTGAAAG	GTTAGAAAGAGAAA	
				CAAAGGAAACAC	CTCCATTTAAAATCAC	
				ATGCAAAGTCCAA	TCTAGACAATATAAA	
				AAGAAAAATATG	ATATTTGAATCTATCT	
				GATTGGACACAG	GCCAAGACAAACACA	
				ATTCTGGAAGAG	GGAATTATATGAACA	

				CTCAAAAAAGACT	CAACTACAAAACACT	
				TCAAAAACCAATT	TTTCACACAATTAAA	
				AAGAGAGGCAGA	ACTAGATCTAAATAA	
				GGAAAAGTGGGG	TTGGAAAAACATTAA	
				AAGAGAAATGAA	TTGCTCATGGGTAGG	
				AGTGATGCAAGA	ATGAGCTAATATAAT	
				AGAAAAGAAAGT	AAAAATGACAATCCT	
				GATGCAAGAAGA	ACCCAAATTAATCTA	
				AATGAAAAGTGA	CTTATTCAGTGCCAT	
				TGCAAGAAGAAA	ACCTATCAAACCTACC	
				TGGGCCAATTGA	AAAAAACTTTTTTATA	
				AAAAGGAGAACC	GAATTAGAAAAAAAT	
				AAAAACTGACAG	ATAACAAAGTTCATC	
				AGGAAAATCAGG	TGGAAGAACAAAAG	
				TCTTAAAAATCAG	ATCAAGAATATCAAG	
				AATTGACCATCTA		
				GAAACTAATGATT		
				TCATGAGAAATCA		
				AGAAACAATAAA		
				GCAGAATCAAAG		
				GAATGAAAAAAA		

				CAGAGGAAAACA TGAAATATCTTAT TGAAAAACAAC GACCTAGAAAAT AGATCTAGGAGA GACAATTTGAGA ATTATTGGACTAC CTGAAAGCCATG ATCAAGAAAAA ACCTTGGACATCA TATTACAAGAAAT TATCAAAGAAAAC TGCCCAGAGATCC TTGAACAAGAAA ATAAAATTGAAAT TGAAAGAATTCAC AGATCACCTCCAG AAAGAAATCCTCA AATGACAACCTCC AGGAATGTAATA GCTAAATTCAAGA		
--	--	--	--	--	--	--

				GCCACCAAGCCA AGGAGAAAATAC TTCAAACAGCCAG AAAGAAACCATTC AAATACCATGGA GCTACAATCAGG ATCACACAGGACC TAGTAGCTTCTAC ATTAAAGGATCAA AAGGCATGGAAT ATGATATTCAGGA AGGCAAAAGATT TGGGTTTACAACC CAGAGTCACCTAT CCAGCAAAACTG AGTATATTCTTTC AGGGGAAAAAAT GGACATTCAATAA GATAGAAGATTTC CAAGCATTCCTGA GGAATAAGCCAG		
--	--	--	--	--	--	--

				ACCTAAACAAAAA ATTTGAAGTCCAA ACATAAGACACA AGAGAAGCCCAA AAAGGTAAATAA GAAAGAGAAAAT TTAAGGGACTCAT TAAGGTCAAAT GTTTATATGTCTA CATGGAAAGAGG ATTTCTGTAATTCT CAAAAATTACTCT TAGTAGTAGAGA AGATAGAAGGAA TTACTCAGAAAG AGGGTGGAGAAG TAAGCTGATTATG ATGATATGATGTA TATGATGATATAT ATGTAAATGTGAT GATATGGAATGA		
--	--	--	--	--	--	--

				TATGTATGTATGA TATGATATGTATG CATATGAATGATA TGTAATAAATCAA TCAAGGGCTGAA AGAGAGGGTTGC ACTGAGAGAAAA GGGAAGGGAGA GATAGAATGGGG TAAATTATATAAC ATAAAGAGGCAT GAAAAATCATTAT AGAGAGGGGAG GATAAGGGTGGT GATGGGCAATGC TTAACTTTACTCT CATTGTAAGTGGC TCAGAGAGGGAA AAACAAGTATACT CAGTGGGGTATA GAATTCTATCTTA		
--	--	--	--	--	--	--

				CCCTACAGAGAA GTAGAAGGGGAA TAAGACAAGGGA AGGGGGAGGTAA TTGGAGGGAGGG GAAGTGGGGGGG TGACAAAAAGCA AAATACTGGTGA GGAGGAACAGAG TGAAAGGGAATA GAGCAGGATCTA AAGGGGATAATA TGATGGAGGGCA ATACACAGTTAGT AATCATAACACTG AATGTGAATGGG ATGAACTCACCCA TAAAATGGAAGT GGATAGCAGAGT GGATTAAAAACC AGAATCCTACTAT		
--	--	--	--	--	--	--

				ATGTTGTTTACAA GAAACACATTTGA GGCAGGGTGACA CACACAGATTCAA GGTAAAAGGCTG GAGCAGAATTTAT TATGCATCATCTA AAGTAAAAAAA GCAGGAGTAGCA ATCATGATACCAG ACAAAGCTAAAG TAGAAATTGATCT GATTAAAAGAGA TAAGGAAGGAAA TTACATTTTGCTA AAAGGTACTATAA ACAATGAAGTAAT ATCATTACTAAAC ATTTATGCACCAA ATGGTATAGCATC TAGATTTCTAA		
--	--	--	--	---	--	--

	Poseidon-1_PM	GCF_010993605.1	Petro myzon marinus	TGATTGATTGTTTAC GCTTACAGTGGCTTT CAGACAGCATACATA AAAAATTTTCGTCTCC CCCCTCCTCTTTTCAA AAAATAGATGGACTA GTCCTGCTGGGCAAA CTGCCAAATCGTTGC GCGTGACGCAAACGT TTTTTTTAAATAAATA AATATGCCGTTTTCA AATAAATAAATTTAT GCAAACCGACACACC TGATGATGACCGTTT GAGTTATGGTCGAAA CGTCGTGAGAATTTT TATTGTTTTATATATA TTTTGTTATTTTGTG TTTCTCTTCAACGTGA CTTTACCGAAAAAAC TAAGAATATGAATTA	MNRLYGEGFHITLRSFNKLRDKKAQLIE SLTFLKRCRDSGVIPIAFTLTSATVTSSNA KRILFKTSLALISEKIRLIRTELDQLDKTLY KSHLSLTAIMSTSDWEHVDNMTASSAA TMSTKCKENLISKFNLLFNKRPDPHTKK LDVNKVVINRSNSSISDGLKSVLAKGLN FAVTPKHIPVNDIIAQVDVALRGLGHIA ADVARSEIAATLKNKMPASNLSKDER FELYKLKRNNDIVVLSADKGNATVIMDK IEYNNKIKELDDGSYTKIKRDPTNSLVT QTNKIISDSTALIDQSKQFSLKPSAARPP RLYGLPKIHKVDTPLRPIVSTISSPTYKLA RFLSEILRPFTGKTTSAVQNSTHLVSTIK GISLSPDLLVSFDVVSFLTVPVPDTIAI IEDLVHQGLNTEVPPLVRHCLTNTYFT WNGTFYKQTEGTPMGSPSPVIANIFM EKFETDAIQATATLRPTLWKRYVDDTFVI WPHGRPALDQFFDHINSIHPAIRFTME VEIDGKIPFLDVLIKRPDGSMSHSVYRK ATHTDAYLKPDSHHHPSQKNSLIKTLYH RAQCISDQEHISSELKHVHQALRANGYS	cd0304	440	524
--	---------------	-----------------	---------------------	--	---	--------	-----	-----

					CTGCAGTCACGAAAG CTTTAAATCAAAA	DYTIRKAITQTMKKNPLEEPFKTITLPYIA GITKKISKILNKHKIQVRFNTLHKIRDLP SVKDVIPDIQYPGVYKLSCLCGSSYIGET KRHISDRIREHKADLKHGRTNTSAVAD HCSNSVGSHAIDFSKTEVLCKPSGFHER FVHEAIHIYKHRDNFNREDGFYLSNHW KDLINHFSS			
	Poseidon-1E_PM	GCF_010993605.1	Petro myzon marinus	AGCTTAGCTGATT GGAGAATATAGA GTTTATTGATAAC ATGAAGGTAATTT ATATTTTAGTTTTT ATTGGTTCACAT ATAGTACTGTGAA TAGCTCAGATGTG TAGGCCTTGACTT CAACATAGTCATT TCTGTGACCAGTC ACTAGCAAATCAC CCAGTTATTTCAA TGCTAAATTCGCC	TAATTAATTGTATTTT ACGATGACCTTCAGA CCGCAGCTCTTTCGC CAGCATTGCCTCTTA ATACCTCTCACACCA CGGTTAATGGTATGG TTTACAACATGGAGT ATACTAGGGCTTCGC AGCTGTCTCTCTCCCC CTCCTCTGTTCCGACC ATGGACGAACTCGTC CGTCTGAACGGGGCC CGTAGCTATCCGTGG TACGAAATTTGAGGG	MNHCSHESFKSKFNRLYGEGFHTSLRSF NRLRDKKAQLIESLTFLKRCRDSKVIPTA FTLTSAVATSLNAKRILARTSFALISEKIRL TRTELDQLDKTLYKSHLKLATMSTSD WEHVDNMTTSSAATMSNKRDNLSK FNHLSSKKS DPHT EKIDVNKVVINRSNS SISDGLKSVLAKGLNFAVTPRHLPVNDII AHVDVALKGLSHISADVARSEIAATLKN AKLPPSNISKDERFELS QLKKRDDIVVLS ADKGNATVIMDKTEYNNKIQELLNDGS YTKIKRDPTNSLVTKN KIISDSTALIDEN KQFSLKPSAARPPRLYGLPKIHKEDTPLR PIVSTISSPTYKLARFLADILRPFTGKTTSA VLNSTHLVSTIKDISLSPNDLLVSFDVVSL	cd00304	453	537

				TAGTTATTTCAAT GCTAAATTCTCCT GTTTGATTTGTAC TAAGTGCCTCGGT GAATGTGTATGCT CTGACTTCAACGT AACTATTTCTGTG AGCAGTTACAAG CTAATCACTCAGC TATTTCAATGCTA AATTCTATGCTAA TTCCATTGTCAAG GGGTGCCTCTGC GGATGTGTATGCC TTAAGTTAGCTAT TTCAATGCTAAAT TCTATGCTAATTC CTTTAACAATGGG TGCCTCTGCGGAT GTGTATGCTTTAA CTTTAAACATAGC	TAAGAATCCACCCAT TTTTTTCGCCATATAG CATTGTGATATATAA CGACTTATGGTTGGT AGCTTAGCTGATTGG AGAATATAGAGTTTA TTGATAACATGAAGG TAATTTATATTTTAGT TTTTATTGGTTCCT ATATAGTACTGTGAA TAGCTCAGATGTGTA GGCCTTGACTTCAAC ATAGTCATTTCTGTG ACCAGTCACTAGCAA ATCACCCAGTTATTTT AATGCTAAATTCGCC TAGTTATTTCAATGCT AAATTCTCCTGTTTGA TTTGTACTAAGTGCC TCGGTGAATGTGTAT GCTCTGACTTCAACG	FTKVPVPDTIAIHEDLVHQGLHTDVPPLV RHCLTNTYFTWNGTFYKQTEGTPMGS PLSPVVANVFMEKFETDAIQATATLRPTL WKRYVDDTFVIWPHGQPALDQFRDHI NSIHPAIRFTMEVEKDGLPFLDVLIKRR HDGSMSHSVYRKATHTDAYLKPDSHH HPSQKNCLIKTLYHRAQCISDQEHTSEL KHVHQALRANGYSYDTIRKAITPVMKK DPLVEPLKTITLPYIAGVTKKISKILNKHKI QVRFNTVHKIRDLVPSVKDALPDIQYPG VYKLKCLCGSSYIGETKRHISDRIREHKA DLKHGKTNSAVADHCYNSVGS HDIDF SKTEVLCKPTGYHERFVHEAIIHYKHRD NFNREDGFRLSNHWKDVLNHFCS			
--	--	--	--	---	---	---	--	--	--

				TATTTCTGTGACC AGTCACAAGCTAA TCATTCAGCCATT TCAATGCTAAATT TTATGCTAATTCC TTTGTCAAGGCTC TTTGTTCTCTCAC ACAGCTTAAAAGC CGTGATCTGAACG TGGCTCTCTCACT CCAGCGCTGTCCT ACCAGACAGTCTC GCTCCCTTCACCT GATGAAGGCCGC TTGTGTTGTGGCC GAAACGTCGTGA GAATATTATTGTG TTATTTTATTTTAT CAATTTATTATTTT ATTCTCATTCTAC GTGACTTTACCGA	TAACTATTTCTGTGA GCAGTTACAAGCTAA TCACTCAGCTATTTCA ATGCTAAATTCTATG CTAATTCCATTGTCAA GGGGTGCCTCTGCG GATGTGTATGCCTTA ACTTAGCTATTTCAAT GCTAAATTCTATGCT AATTCCTTTAACAAT GGGTGCCTCTGCGGA TGTGTATGCTTTAACT TTTAACATAGCTATTT CTGTGACCAGTCACA AGCTAATCATTCAAGC CATTTCAATGCTAAA TTTTATGCTAATTCCT TTGTCAAGGCTCTTT GTTTCCTCTCACACA GCTTAAAAGCCGTGA TCTGAACGTGGCTCT			
--	--	--	--	---	---	--	--	--

				AAGAACCAAAGA AG	CTCACTCCAGCGCTG TCCTACCAGACAGTC TCGCTCCCTTCACCTG ATGAAGGCCGCTTGT GTTGTGGCCGAAACG TCGTGAGAATATTAT TGTGTTATTTTATTTT ATCAATTTATTATTTT ATTCTCATTCTACGTG ACTTTACCGAAAGAA CCAAAGAAGATGAAT CATTGCAGTCACGAA AGCTTTAAATCGAAA				
	Poseidon-2B_P M	GCF_01099 3605.1	Petro myzon marinus	T	TAAACTAATTTTCAAT GGTACAGTAAATTAT TTTTCAAGTAGTGGT ACCCCAAACCTTCGC AGCTGCTTATAAATC TCCCCTCTTTCACACC ACTCAATAGGTAAAC TCACTCCCCTCCCTCC	MELDTLDRTLLQAHMLTQKMSPDW NHVDNMTASSAESLSVKCKANLISKFN HLSKKNPVPHITKLDV NKIVINQSSTPIS DNLKTVLARGLNFAVTPKHIPVDDIISQ VDVALKGVGHIQADVARAEIATILKNAK KPISNISKNERSELANLKNRSDIVILSADK GNATVIMDKDSYNNKIQNLLNDGSYSK IKNDPTNSLAIKTNKIITSSTSLIDPTKRPS	cd0 030 4	361	445

					CCCCCCCCGCGCCAC ACGCCAACTCCACGC ACGTACTCCACGCAC TCACTCTCTCACTCAG TCAATCAATCCAATC AATCTGCTAATTTATT TATGCCACTTTCTTTT TCACTCCACCCATTCTG TTGCTCTCACTCAGTC CATTACTAGGTGCG CTCGGTGACTGGATT TACTGCTCTGACTGT CTTAACTAGCTGATTT CACTGAGTGCACTCA GTGAATGTAATCACT CACTAACTTCACTAAT TCTACTAATTCCTGTC AGTCCACTCAGTCTA CTAACTCACTCAATTC AATCTAAATTCAATTC TAATTCCTCAAGGGG	LKPSAARPPRFYGLPKIHKEGTPLRPIVS TINSPTYRLARFLSDTLLPFTGKTASAVL NSAQLVSMVKDVLSPDDILVSFDVVSL FTKVPVPDTIEIENLVQHGLDPDTPALV RHCLNNTYFTWNGSFYKQNEGTPMGS PLSPVIANIFMEKFESDAIQTSTLRPSLW KRYVDDTFVIWPHGLPALNLFHHINSR HPAIKFTMEVEKDGLPFLDVLFFERRPD GSLGHSVYRKPTHTDSYLPKPDShHHPA QKNSLIKTLYHRAHCISDQHHLTSELKH VHLALKRNGYSDRTIRQAVSPSIQKQPL DTPVKTITLPYIAGVTKKISKILNKHKIQV RFNTVHKIRDLPVSKDVIPDLQYPGVYK LSCTCGDCYIGETKRHVSDRIREHKSDLK HGKTNSSAVADHCYNAVGDHDIIFSDT KVLCKPTGFYQRFVHEAIQIYKHRHNFN REDGFKLSNHWKDVNLNRTKH			
--	--	--	--	--	---	--	--	--	--

					TACCTCTGCGGATGT GTCATGCTACATACC CGCTTATGGTCCATC ATGGCCCATTACCT CCCAGTTTCATCATG CTAAATTATATGCTA ATTCCTTTGTCAAGG CGCTTTGTTTCTTCTC ACACAGCTTAAAGGT CGTGATCTGAGCGTG GCTTTTTTCACTCCAG CGCTGTCCTCCAGAC AGCTCGGTCCCTTCA CCTGATGAAGGCCGC TTGTGTTGTGGCCGA AACGTCGTGATTTTT GTTATTTTATTTTTTA CCTACGTGACTTTAC CCAATAGAACCAAAG AGTATGGATCCTTGC			
--	--	--	--	--	--	--	--	--

					AGTCACGAAAGCTTT AAATCTAGAA				
	Poseidon-1B_P M	GCF_01099 3605.1	Petro myzon marinus	CA	TAACTGGCTGTGATG TTTCCATCCTACCTAC CGATGTTAGCAGTTA TTTTCGACTCTCTGAG TTACATTCACTAACAC AGCCAACCGTTAGCT GTGTTAGTAAATGGC GGCTTGCCACTGCCT CGTCCATGCCCTGGC CGAGCTTGGCCCCCT CCGATCTTACATTTAC ATGCAAATTTCTCTCC TACTAGCCCCGCCCA CTCGCTCCTACATGA CCCTTGGCTGCTACA CCCATCTCTAAACCTT TATAAACAGCCTGCC TCAGCAGCTTTCCTC TTTTTTCACCTGATGA	MGSPLSPVVANVFMEKFETDAIQASTL RPSLWKRYVDDTFVIWPHGLPALNQFL NHINSIHPAIKFTMEVEKDGLPFLDVLI EKRPDGSMSGHSVYRKATHDAYLKPDS HHHPSQKNSLIKTLHHRARCISDHHHLA SELEHVHQALKMNGYSYDITIRKAITPTV GKEPSEDPIKTITLPYIAGVTKNISKIMSK HKIQVRFNTVHKIRDLPVSKDVIPDMQ YPGVYKLSCVCGSSYIGETKRHVSDRVR EHKADLKHGRTNTSAVADHCFNAVGS HEIDFSKTKVLCKPTGYQQRVHEAIQIY KHRHNFNREDGYKLSNHWKDVINQFK S	cd0 030 4	1	83

					CGATTGCTTGTGTTG CGATCGAAACGTCGT GATTTTAAATTTTAT TATTTTTCTATCTAC GTGACTTTACCGAAA GACCCAAAGAATATG AAT				
	L2- 24_AC ar	GCF_00009 0745.1	Anolis carolin ensis		GATACGGTCGATATA ATCTGATGTATTCGC TGGATTAATGTGTAA TTGAATACAGACTGG CTCTGACTTTTGGCCC AATGCTACTGGCCCT GTTGTAATGGCCACA CAGTGCTTTACTGTTT TAAATGGGGTATAAT ATTGGTTTTTAAGTG TGTTTTTACGATCGTT TGTTTTACTATATTTT GTATTATATGTGGCA TTAATCTTGCCATTAT	MPIKMVRPRRLMDPDGFLRSLGDLPAS SRGDPVDALVTRYNSEMARAITIAPER PLSLRRAKPAPWFTEELAVMKRERRGL ECVWRKLRSSEDRARAKALRKAYSVAL RAAKKSFQTARIAAAKNRPSELFVVG LLNGTPQTGPPDDLAENCAAFSHHFAD KVAQIRSELDATVDAVSEEVARAPVCPI LMDSFQLVQPDDVDKILGEVRATTCAL DPCPSWLKVKEARGGLVEWVRVINASL REGIFPSCLKQAVVKPLLKKPSLDPTNW NNYRPISNLPFLGKVLERVVASQLQGFL DDTDFLDPSQSGFRPGHSTETALVALV DDLRRELDRGSVTLVLLDISAAFDIDH GILLGRLLGMGLGGTVLQWLQSFLEGR	cd0 165 0	257	517

					GTAAGACCGCTCTGG GTCCCCCTGGGGGA GAAGAGCGGTATAT AAATTAAATAAATAA ATAAATAAATAAATA AATAA	SQMVKLGDTCSDPWPLTCGVPQGSILS PMLFNIYMKPLGEVIRSFQVVRCHLYAD DTQLYYSFPPASKEAPQILNQCLAAVSD WMRANKLKLNPDKTEVLLVSRKSDRGI GWQPVLDGVALPLKAQVRSLGVLLDSS LTLEAQVSAVAGRAFAQLKLVRQLRPYL VKSDLATVVHALVTSRLDYCNALYVGLP LKTARKLQLVQRSAARLLTGANYRERTT PLFKELHWLPVIFRAQFKVQVITYKALN GLGPTYLSDRISPYEPARRLRSSGEALLS VPPPSQVRLVGTRERAFSAVAPRLWNS LPREIRQAPTLLSFRKQLKTWMWDQAF GPS			
	Jockey - 8_DPe r	GCF_00328 6085.1	Droso phila persim ilis	GGCTAATGCAGC ACACAAATTTACT TGCCTGTGAGATC GCAACGCACCCAT ACATTCCAAGAAG GCTGGCCAGGCC AAGGTATGCCGC AATGATAGAAGA	C	MSASLPPANAGGDGNWFDMMKRDYQF SSQDEHNRNNIPHISLAPTREDSIQMCT IAAKIKNHQTINIGAIYCPAHEWIDTQF KQVFDSLGSRFIVAGDWNASKQWWG NARSCRGAALLKCIQENDQSILATGSP THFPSNRRHMPSAIDFAVYKGIRHNLIH IETSYDLGSDHNMQIIKLLLRPEYGKKKIL LPKNANIRRFQAHLEQRIHLETELHNGD	cd0 165 0	470	731

				GCACAATCCAAAC AGCAACAGAAGC CTGACAGCCATCC AAACAATACTGCT CCAAGTCAGACG CGAGAACTGGCT GCGCAATCCATTT GCAACACACAATA TGGAAATCAGCC GCAGTAGCGCTA ATGCAGCTGCTAC GCAGGCAAATGA CGGCCCCAGCAG AGAACCTGATGAT GGACGGCGGCAG CTCCAGAATCCAG ATGATGGCGGCG GCTGAAAAGAAA CTAAAAACCTTTT TATTCTTCCACAG ATACGAAGCCACT		DIDDAVAILEKNIYIAAELASQTSSNQHT VYGPQNFHLDSTLKAIIKIRILCKINQK DRCSNPLIARTINQLTNRISKRIKLQKDK QMNRLLDIDPTDRYSMQKLWRLTNT IKRQPIPNLPVRRHVAEGQQAPADPW CKTTEEKAEFAQHLEARFTPKITAALE DREEINAVLQENGPDAAQIIRPVTIQE LKSEIKKLKTKKAPGLDSIDNRVIKSLPTS AIEYLLICNSALQFGHFRKSWKCAKISM LRKPGKPSDRLESYRPISQLTGFSKILERL ILKRLFEEESFAMAIPNHQFGFRSQHSA EQQIGRVTQFILRSYEAQRHCSAVYMDI QEAFDRVWHQGLLCKVKTIIPDNLII TSYLSDRSFHVCCNNITSSLKHITAGVP QGSVLGPVLYNLYTADMPQPPRHANS TQMLATYADDAVYLVSALSPKQANEN VERLLADALSWSKKWGITISLAKTAHVL HTLSPVQQHTPAPRLAGQLLAKKSVMPP YLGIIHMDTKLNFNKHVAAIVTKTRARIK KLDWIIGRNSKLPLPPKVLLYKQLIPIW QYALPVWGPLTSSMQMAKLQTIQSKA			
--	--	--	--	---	--	--	--	--	--

				TCACCTGCACAGA AATTGGTAATCCA CATCCAGAATGAC GATGCACTGGAC CTCGGCCAGCTTC AGCCCTGTTGATG CTGTCCTGCAGTC ACAGTTTAGCGCC GACACTCGCAAG GATTCCACTTCGC AGCCGACGACGA GACCAATAGGCG TAATGATGTTCTT CACGGCTGCATAC TCCACCTGAAATA ATTAAAAACAAAA ACAATGTAAACAA AGTAGCAATAAG TTTAGTGTTACCC TGGCATTTAGCCA AATTATCGATAGC	LRAMTNAPRYVSNNALHADLQVATVE EVYNRLCRNCYT			
--	--	--	--	--	--	--	--	--

				TACTAAAAAAAAA TAAATACCAAATA AGCCAATACCGTA AGAAATGTTACCC TGCCAATTTACAG CCACTTCTGCAAC TACATGTAATGTA GATAAGCAGTTG GCGCGAAACAAT TTGTAAAATATGC TAATTATAAGAGT ATAAACAAATAAA ACGGTTCTCTCTT C					
	R4- 1_PH	GCA_00158 7735.2	Parhya le hawai ensis	CCGACCGCCAGC GGGATAACTGGC AAACCCTGTCTCG ACCACCGGCCCGT GAATCCATCGGG GCGTATGAGTCA GACACAGGGGGG	TGAGCCTTAGGTCGC GGGATGTGACCCGG CGCCAGAGTGTAGA GCTGAACATCGCTCA ACCGATCCAATTTGG GTCGTGAAATCCCCT	MKMSHNRDTPSNGVKGTSVRLGTSLV RSPVGEAGAVRERGTHPSESVSQSDA SVNATGEGSVREQAPLSPPGAEETVP TQRRTRHKWSKEDRVVLWECFVASKR EGPGYLKRLKQLWDERGIPGNFPQASL SGQIRQICSKNLLSEEERLQIAARMEAQ VASPSADEPARQVPTRPVTPPRSPPVEP	cd0 165 0	616	879

				TGTTTAAGGTGAC CCTGTTGCGAGG AAATGCGCAGCA AAAGCCGGATGA GCCTTAGAACATC GAAGGCCAACGA CAACCTGCCGAAA GACTGGCGGCTA ATCCAGTCAACCT CCTGTAGGTCACC GGCTGGTCATGCT GAATCTCAGCTTA ACCAGGCGTACG AACTTACAGTTGG AGGGTCGAGCAC CCCTGATGGCTGA AAAGGACCATCA AAGTCGAAGGTA GCCAACGAGAAG ACCAACGGCTGAT TCAGGCGGAAGA	CGATAATAATAATAA TAA	ARRPSIPSEETPDLGAVPSEIDSADPNRS PSRGPRHLPANHMSQSESEDDVTDPD VSDQQRSDSLEPRDLLRNSSVESTPGHP NQELSDTLLSNYPSEIDSDDPNQSPRR GPRHLPAMDMSLSDSMDEETEPDLS QQRSDLLELRDLLRNSSVETTPKGHPSL RHLPEPKIRAAAFRVNSVIGKIHTNNITE TNALIKAGADLAVRILEVQPRPQRTQRK KDPPWKHRLEKNIAEIRKHLSEWR RGNLHDEEKKTLLESRYRCLEVGLTNLE DTLKQRLSAKRSKVRRFEARVAGFHQN QLFNTNQKRLYQTLRGEETSSDSPNAEE SIRFWSDIWSKEVRHNNAAEWLHDVK EKNVAADPDLTITSQQLKKQLSKTKNW KAPGPD MVQGYWIKTFTSLHSRIAAQL NHCLQRGTVPTWMTTGKTVLIQKDKA KGTEVSNYRPITCLPLMWKVLTGIIYER VYQHLD SKKLLPDEQKGCRRNTRGTD HLLVDKLLTKDARSKKKNLSMAWVDYK KAFDMVPHSWILECLDIYGIAGNIRNLI ATTMPNWKTQLTSANKHLGEVSIRIRGI			
--	--	--	--	--	------------------------	--	--	--	--

				GTCAACTCGTTGA ATGCGTTCGACAG CTTGGGGTAGAT GGAACTCCTAAGC CCTGAAAGGCAG TCCATCTTCGCAG ACGCTAAGGTGC CCCGCCGCCTGAG GGTTATCAGGCCC CGCCGCCTGAGG GTTACCAGGCAAC C		FQGDSLSPLLFVLTMIPLSETLNKAGQG YNYSRTMKLNHLLYMDDLKLYAKSKDQ VEQLLNIVHQYSQDIKMQFGISKCGVL NIERGEVTASEGIIIEEGTIKDIEETGYKYL GVMYNTILHRTMKDSIRKEYLTRRLIL KSHLNGGNTIKAINTWAVPVVRYSAII NWTKKDCTDMDIKTRKLMTIYRALHPR SCVDRLYINRREGGRGLISVEDCVEAEK RALSQHFRESDDPWARCLIEAKLLKETE TADQFKERRRLDRTNWKSMKMSGQ YLEAVQDKIVPDSWNWLLRGELKRETE GTILAAQEQALRTRYIQNKIDKRNVPST CRICRSSDETINHVISECGVLAQKEYKRR HDKVARHLHWTLRLIHNFPVSRWYEH EPAPVVENEAQVIYWDKRMETDRVLH ANRPDIVVKDKQEKSAKLIDVSIPFDSRII DKEAEKKEKYRDLAIELQRLWQMKVDV VPVVIGALGAMSKNLKTALRELKCGHL HPGTLQKSALLGTAHIIRKVL			
--	--	--	--	--	--	---	--	--	--

	Proto2 -4_CTe	GCA_00032 8365.1	Capitel la teleta	GATGAATTGTATA ATTGTGTTGATTA TGATCGTGATTCA	TGATTAAGCTACCTTT ATGTTGCAGTTTTGA TTTGTTTGTTCGTTTG TTTGATTTTCTTTCAC CTACTATAATCCTGTA TATTTATTTATACGAA TAAATTATTATTATTA TTA	MNELIAFFKISVQNVCDQGSCYSSHQF TAQGVRSAVSKLKDKSDCKDNFMSNH VIHGGDTLFLVHLSFLFNCMISHTCDPKQ LRMSVLIPIPKNRRKSLSVSTNYRSIALSS IVGKVFDRMILTEHVNIFSTSNLQYGFK PKHSTVQCTFVLQNVINHYKSRGSPCYV MLLDASKAFDRVQYVRLFYLLRERGLCA RIFMLLALLYTNQSMCVRWKNCISDNF SCSNGIKQGGVLSPTLFCIFYDELLIRLKH CGSGCHLGHFTGALSYADDVALLAPTL TAAKVMLNVCEQFANNFNVLFNADKS ALLVCGRPIRKEVNLRLNGRIIQQNSTA VHLGTTIGDNSSTRNVKKAASDLYCLTN VLMCKFGHC SIDVLSALFFFFFFFYCSSFY GCPLWDLRNI EPILTAWRKCLKRIWKLN IRTRTKYINNLPKGNLISALLIRSLSFYMQ CLSSSNAIVKLASNYCFYGC MYSVISENV HLCSVLSSVPLRFIFDDFENVKSIIQRGM YDEATIAESLALKDLCLYLRDGLMFLNFEI TSEELLCAFNHFTI	cd0 165 0	85	341
--	------------------	---------------------	-------------------------	---	--	---	-----------------	----	-----

	R4-1_AC	GCF_000090745.1	Anolis carolinensis	CCCCCCGCAGGT GCCACCTTGACGT GGTGGGGGGGCT TAGCTGCTCCAAT GATGACTGAGGG CTGTGCTGGCGGT AGTGTAACCTACCG GCAGGTCCAACC AAGCCAGAGAGG CCTCAGCTGAGG AGCACAACCAAG AGCACCTAACCCA TTAGCATT	TAGACACTTGGGAAG TGTTGCGACTTGTGGT TTTGTGAAACGAAAT CCAGCATATCTATCTT GTTTGCTGTGTCATA CAACGTCGTTGTGTT GAT	MENQTQTKSILARSSPRIKRTAVDAADS GHPSTSGLRNHQNPNEQSQRQKYTM PENRTIMKCYKSEPQRRGYQKRMHQ LWKQEYPDSQITEPRLADQRRFIIRNKV FSEVELEEIQKICKVNYHQTTAQTAET PATLGMVEQIEPEESVELLQEFEEPALET SVEPPGTLTARQQELKDKIMAHAAANA IRKRLPTLKTVPKRHLAPLMKDVNAALS TVQITSIEQTNQYAYSAAVIVTEELGLLQ PRQPQRKSTGKPKWKVRLELKIKKLRSD ASNLKNMKEKKLNDKIKQYLIRKYWL NTRKIEEALIVKEQITATARKIERYEARII QYRQNQLFQSDQRRFYQSLNQTTDTV TIKPEKTATTKFWKELWENNKNYNKNA GWIKEFEGKFSQNMELMEITTEMISK RVQKVKNWTSPGSDQLHGFWLKHLISL HGKMAQQFNEMLQKGSISEWLTTGRT YLIQKDKPAKGAAPGNYPITCLPTMFKL LTGIIADRIQDYLEEKNILPDEQKGNKRK SRGTDQLLIDKMILENCKSRKANLHM TWIDYKKAFDLPHSWIICKLDAIGICKN	cd0	1650	465	731
--	---------	-----------------	---------------------	--	---	---	-----	------	-----	-----

						VGTFIENMMEHWKTELFVGNESYGLV NIRRGIFQGDSLSPLLFIAMIPSTILQKT NLGYQTSKNSHKISHLMYMDDLKLYGK TETEIQSLTNTVRIFSTDINMEFGLDKCS TVALKKGKIIIESEGINMPNGQTIKCHQP EAYKYLGILQLDNIKHEHVKT VVSKEYT QRVRKILKSKLNGGNTIKAIN TWAIPVIR YTAGIINWTQVELDNLDRKTRKLM TIH HSLHPRSDVDRLYLPRRSGGRG LLQVK QAVKEEEHALAEYVKQSEEPALIEVKNQ KLLKTQQTKNQYKKTALQTRADSWHN KTLHGKFLDKIEGKADKEKTWLWLTNG TLKKETEG LILAAQEQAIR TNAIKAKIEKS ADDPKCRLCKETDETIDHILSCCKKIAQT DYKQRHNYVAQMIHWNLC LKYHLPAA KNWWDHKPAKVLENEHAKILWDFRIQ TDKVLEHNTPDITVVEKNKVWIIDVAIP GDSRIDEKQQEKLSRYQDLKIELQRLW QKPVQVVPVVMGTLGAVPKDLSRHLE TIDIDKITICQLQKATLLGSARIIRKYITQS			
--	--	--	--	--	--	--	--	--	--

	CR1- 66_AA e	GCF_00220 4515.2	Aedes aegypti	ATCGCTGTTCGTA CCTGA	MYSKRFQPSNETLAISHDVQPGCTPVSL VEASYPLVTVEPILPATSSHPGPVVEYGE EVFQTLFAGKYSCDSSNFSPVTSVASSDI PTRKSDTKSAAPDDFASDNPANRLPNV ITDDLNNNQSCHALRFYYQNVRLRTK VDSFFLAVSECDYDVIVLTETWLDDRFQ NAQLFGSKIIVYRTDRNAANSSKRRGG TAGATTTTTATGTTTT GVLIAVSTHLNSFIDPTLTSLAEQLWIKI TTTTTTCATTAAGACA STPSRTISVGVIYLPDRRDNMDYIRDH AACAGATGTCAGATG LNSMNSVLSNLALNDCVLQFGDYNQP AATAACAACCTCAAT NLLWEASDDGFLRVNSQSILSTAGSTLV AAATAAATAAATAAA DGFHFHGLTQLNGIPNRNSRFLDLVLG TAAATAAATAAATAA NDEARYSCNVTEAIEPLSDLDVDHPALD AT VILNLPDSITFEDCSDNSQMYDFRQADY AGIHAAISRIDWQPIDDANNIDDAVTYF NHVVQDIIAQNVPLCRPKPKPPWSNSH LRQLKRNRKAKSQRHYCRNRSQFTKQEF NRASNAYRTYNRLLYRRYVVRTQTSLRR NPKRFWSFVNSKRKEVGLPRSVHLDGQ TATSEQEKCFLFAQHFQKTFHDSPTA DQIAAAIEYTPRDVDFRSFTIDEQTIQA	cd0 165 0	638	885
--	--------------------	---------------------	------------------	------------------------	--	-----------------	-----	-----

						AFRKLKMSHAPGPDGIPSCILKNCSHSL ATPLAKLFSRSLQEGKFPDEWKVSSMF PVYKKGDKKDVHNYRGITSLCSCSKLFEI LINDALFNSCKSFISTAQHVFFPKRSVTT NLTEFVSLCLLTLDSGGQIDTIYDLKAA FDRVDHGILLAKLDKLGVSLSLVRWFES YLTNRLLYVKIGSSQSESFTNLSGVPQGS NLGPLLFLLFINELSQLLPPGCRLFYADD VKIYMIIDSTADCELLQSRVNIMANWCS SNFLTLSVNKCNCVISFHRRQKPIIFDYRIC DTVLNRVDQVKDLGITLDQEMNFKPHY NDVIAKANRQLGFIFKLSDDFRDPLCLK GLYCSLVRSLIFSVMVWCPYQSCWIAR LESVQRRFLRYAFRHLPWRDPLNLPAYE DRCRLGIETLEQRRRRAQAVFAAKLIV GDIDSPELLRQLNIYAPEQVIRQRNFLQL VPCNRLYGMNEPLRASADALNEAFTAF DFNLP LISFIRRLSLS			
	DongA a	GCF_00220 4515.2	Aedes aegypt i	GGTGGGAGCTCA GGCAACATAACTC CTGGGCGTCCATG	TGAGAAGCTCATCCA GCAGACTCATTAGGA TAAGTTAAACCGACC	MKANPAMETRSMRNRMRSPGEGAP TGAGPGTGDASGRQLEEDNVQERPS SQRAQVPVPRTRNRNGSSRNHQGHAAP	cd0 165 0	530	802

				AAAAACCACCATC TTGAGTAGGTGG GAGCTTGAGGTA AAATATCCTGGGC GTCCATGTAAAC CACCTCGGCGA GCAGCGGCGCTT GAACCAAGCTATT AAGCACTGTAATT GGAGGAAATGCC AATGCAGAACCC GTAGCTTCGGG A	AATGAGATCCACAGA GCCTAATCCCCTTTG GCATACAGATTGCCC GGGGTAGGTGAAAG TTTCCAGCATTTTTTG CTGAGAAGTGCCAAA ACTCTATAATAATAA TAA	TDVAVADRRQSLTAGGRRQRIMWTR EMNMYVIRCYYVCTRMETDMSGRSG MLEMFNERFPRFARQLDLNKLYTRQRA IMSHNMLTAAEVEYIKLEVQREIGEEGT RSSDTSRRSSVGLDAPITSES RDSPAPTA PGPTADMQRQQLRDELALHMGTA VT QFRGTDPM SRHRIPKLRYSYRLTSAVSIL NQDILPQYLEAVENLEDLQFIVYSAAIAV VKTLGLRTRPQGEDGTRPNAQKPVW MRRLESRIATLR SKVGRLTQYKQGNRST RLVRHVAEIVRPAELRDLREVDITEILDT HVQRLSALAKRLRRYGECSKRKEQNRM FNINEREFYNRIRNDKVD FGEGLPEIGD VTQFWANLWENPTEHNGDGMWLAE EERQCDGIGDMTAVVVTAQDIREATRY TRNWAAPGPDFVHNFWYKKLTTIHGR MAECFNTVLGDPTQLPEFITRGVTFLLP KDQHTADPAKYRPITCLSSLYKVLSSVIA RRVQAHCDANNVMTEEQKGCRRTQ GCKDQVIVDAVIVGQATQKQRNLSMA YIDYKKAYDSVPHSYLLKVLQLYKVDGN			
--	--	--	--	---	--	---	--	--	--

						VIRLMQHAMGMWSTSLHITDGTAVLR SRTLIRRGIFQGDTFSPLWFCLAMNPL SKALNQCNYGYQLKSGERSTRVTHTFY MDDLKLF AESVQRLHQLLQLVTTFSNDI RMEFGIDKCRSIHLRRGQVMDASCFRV NEQEEIRNMVEGETYKYLGLQLRGIRH TMIKKELQEFLSRVNCILKSFLSAGNKV KAINTFAVPLLTYSFGVVKWTKTDLEAIE RAVRVAFTKHRMRHPKSSIERVTLPR AGGRGVTDIQALCVSQQQLRAYFVES QNRHEIYRTVCEADHGFSALHLAQEDY QLNCDIKTVDEMIAMWKQKELHGTHP HQLELEHIDKVASNTWLVRGDLFSETE GFMVAIQDRVIATKNYRHYILHEDVED RCRKCNSVGETIEHV VAGCSVLG SAYL DRHNEVAKIVHQQLALKHNLVDRFVPY YKYLDPVLENSCIKLYWDREITDVLIRA NRPDIVVYDKRMKRVTLIDIAVPLDHN V QSTFSNKIAKYHDLAEELKQMWHLEDV RIVPVVLSATGIVPKSLLRSLDELELK KDL HSIQKAVILGTCSIVRRFLNHHN			
--	--	--	--	--	--	---	--	--	--

	RTE-1_Alp	GCF_000816755.2	Arachis ipaensis	ACTCCGCCTATGT TACACTTGCGGTA CATAGCCGGTCCC AAGCCCGGATAA AGGAGGAGGGTT GTGTTAGGTCTTC GGCAACCAACAT AAAAAATATAGCC GAACCCCATGAC	TAGTGGGA	MNQRHYCAKARSLPGSNAPYGSSTVS NEQEPLHRCPDVVLNEQGFSRFRERTR VNKLVHKVKGKGRSDRRLRFGTWNIGT LTGKSMEVVDTMTRRKINIMCLQETK WVGAKARELDTSGFKLWYTGKVKNRN GVGIIVDKQWKKD VVDVKRVGDRIISIK LVVEGGAFHVISAYAPQVGSDEQHKIRF WEDLESLVQGIPLGDKIFLGDLNGHV GREVTGYGSIHGGHGFGVINAEGKTILD FSSTFDLLIANTCFKKRDEHLITYKSGMT SSQIDFFLLRRVDRKFCINCKIIPGESLTT QHRVLVMDFRVEQKLRKRHHTKNPRT RWWRMKGEEQRSFLRRVGEEAKWDG NGSAEEMWREMAEVIRRTAKESFGES KGIGPRDKESWWWNASIQEKIKIKREC FKEWSLCRNADNWEKYKAAKKETKVA VSEARTRAYEGLYQSLGTKEGEKGIYRIA KSRERRTRDLQVKCIKDKDGEVLAQEE KINERWKSIFYELFNEGQKTLPSLGRLC TREEDQNFDDYYRRIRDFEVKEALKQMK NGRAVGPDNIPIEVWKGLGEKGINWLT	cd01650	589	836
--	-----------	-----------------	------------------	---	----------	---	---------	-----	-----

						KLFNEILRSKKMPDEWRKSTLVPIYKNK GDIQSCGNYRGIKLMSHTMKLWERVIE RRLRKETQVTENQFGFMPGRSTTEAIYL LRRMMERYRSNKRDLHMFIDLEKAY DRVPREVLWKVLEKRRVRIAYIRAIKDM YDGATTSVKTQGGVTEEFPIGIGLHQGS SLSPYLFTLVLEVLTEHIQEPVPWCMLF ADDIVLMGESREDLNKKLELWREALEV YGLRISRSKTEYMECKFSLRRENPNIEVK IGENILRKVKSFKYLGCIIQDNGEIEQDV NHRIQAGWSKWSASGFICDKKVPLKL KGKFYRTAIRPAMLYGTECWAAGEHE HKLSVAEMKMLRWMSGHTRLDKIRNE DIRERVGVAPIVEKMVESRLRWFGHVR RRPIEHPVRRVDEMEDGQRAKGRGRP KKTIHEVVKRDLVNGLSVDMIHDRAQ WRRLIHVADPT			
	Ingi- 2_AC	GCF_00000 2075.1	Aplysia califor nica	CCGGAGTGGGGC GACCATCCCCGAC CCGGCGGGTCCC AGGTGGCGGATA	TGATGCGATCGAGAA GAAAAAGAAGAA	MTLEEGNVRYGARPVRTHQAGAVPGP KPRRGPPQRLEGLPSSWSARRRGRKAT GNTNTKQVKKSLNIMHWNNAEGISNKK TELENFLKDNEIHISCIQETHLQEEASF	cd0 165 0	548	804

				GGGAACGGCCAC TGTTATAGAGGTT AGCTGTGAATATA TTGAATAAGCAGC CCCGGACCAACTG GCGATGCTCTTAC CAGGACGGGGTG AGGTGGTAGGTG GCACTACCACGTC ATTAAATGCCCTA GGAGTCCTCTGAC GGGAGCATCATG AGGAAAAGGACC GCATTAAGTCCTG GCTACGGGGGGG ATCCGGTAACAAT TCCTCCCTGTGCT CCCAGGGTAGGT CCTTGGACCTGGA GCTGGAGTGCTG AAATCGGAGGTA		VRGYQCFRNDRQGRKKGGVLTlVRNN LNTVEIKKYTGAEYIHNRITTSRGNLDL VNYTCPNDTNLVLETIEVPDTDfIIAGDF NSQSQSWGYNRLDRRGEEIEAWQDEN HLLL VNDPLDQPTFYSRAWHSTTTPDL ALCTENLHRHLSRKVETQLGGSDHRPV LLTIVESLPEDSPKRPRWNYKKADWRLF SIRTNELTKNIAVFGKNINNAIREFNENIL KSAKETIPRGVRKNYTPYWTPELQKTH NDLIKAREEAETHPGRENNNKLQRCKA HYLKT KLECTRKSWRDKTHKLNLEKDTT KLWRLVKGLSDEGNKNDKISLEVEGEM EMEKRAAHTFVKAYAQESNIKIPRERQK EVRQEERRKRSTNNIPEAMEKDITMEE LKKAINQLKQRKSTGPDEITNEMLQHL GNTALQTLLDIFNLSWKIGQVPQCWKE ATMILILKPGKSKAQPLSYRPISLISCVCK TMERIIACRMQWYLETQHIIVPEQAGF RQFRSTEDQTIYLSQVIEDSFQAQKVML ATFVDLQKAFDKVWKEGLLVKLLRSGIQ GNMYQWTKSYLHNRRARVLVDGYQG			
--	--	--	--	--	--	---	--	--	--

				TTGGGCCGCAAG GCGCACTCTAGA GAACCACAATACC ACGTAATACTGCA ATAT		RKMLLRQGVPPQGGVMSPLLFILFINDL MNELPKGIHAALYADDLVLWCSEEFVT TATYRMQLALDKLTTWTKTWCVTANK EMSSATLFTLSQAEPKHLTLDKTPLKFE NQQKYLGIIFDKRLTWKQHIQNAENKA RKKLNIMRKLGTQWRANEKTLRQVYL GAIRPHLEYGSSAWMTAAQTHKYTLEK VQNQALRIITGAMRSTPIEKMEQKTSIV ALQRRWECKALVQYTKATAMEDHILHT RTERSNTARLKRSNFIKESRKIQRSLQDH LPATVNMLNHTPEAPSWEEKNTNISIH TSIPMFTTKDDHSDLSKSLTSLMLEERY PREAWIRIYTDGSATDAVKRGGAGVYV QFPNRRRQAEAVPTGRHCTNYRAETEA LVQAANIIKNVDPQSQVFLTDAMSV LQAINSGKLPRLRKEIEDIPCLRIMLQWI PSHCGVLGNEQADRMALGAQKEQP DNSVNYTEIKTIINTLFKAPQTPDGYHEL SRQEQVIIIFRLRTQHTRLRQLMYKCFKI VQSPKCQCGEADQTVEHILQHCKTHRA			
--	--	--	--	--	--	---	--	--	--

						LRQETWPTPLTLEEKLYEPPDELTRTAQ FMTDTGLQV			
	JAM1	GCF_00220 4515.2	Aedes aegypti	TTGGCGCGGTAA ATGGCTGGGCAT GGCGTACCATTG GTACCTCGCGTAC CTGCAGGAATAA AATAGACCCCTTT GTGCGGTCCTTAG CCTCTTGCCCAGC AACTCCTATCCCT ACCTCCTCGTGGT ACTGGCCGGGGT ACGAGTAACCTTG GGGAAGATCGGG TAACCAACCCCCG GTGGGAACCTTG GTCGTATGCTGAC AGGGAAGGGGG GGTTTGCTTTTGC AAACCTGGAGCG	TAATTGATGTAAAGC CAAATAAGTAAGTAA GTAAGTAAG	MGDMQRRVIGWWPINERMCKLRIKG RFFNFSIINVHSPHSGSTDDDKDAFYAQ LEREYDRCPSHDVKIIIGDLNAQVGQEE EFRPTIGKFSAHRLTNENGLRLIDFAASK NMAIRSTYFQHSLPYRYTWRSPQQTES QIDHVLIDGRHFSDITDVRTYRGANIDS DHYLMVMVKLRPKLSVINNVRYRRPPRY NLERLKQPDVANAYAQHLEAALPDEGE LDRAPLEDWRTVKAAINDAAESVVG VERSSRNDWFDEECQEVLEEKNAARA AMLQHGTRQNVERYRLKRKQQTRLFR DKKRRLEEEVECQEMELLYRSQETRKFYQ KLNTSRKGFVPRAEMCRDKDGSILTDG REVIERWKQHYDEHLNGAENTGTEGQ DSEGDGYVSTADSGNQPAPTMGEVKD AIQQLKNNKAAGKDGIGAELIKMGPDR LAACLHRLIVRIWETEQLPEEWKQGVIC PIYKKGDKLECENYRAITILNAAYKVLSQL LFRRLSPIANEFVGSYQAGFIDGRSTTD	cd0 165 0	457	716

				TCTGTACTCCACG TTAGGAGCGGCT CACAAACAGCGTCT GTTCCCCATGTCA GGGGCGGCTGAT CATCGTCCGAGTG CCAGAGAAGGAC TCTAAGCTAACT GCGCACTATGGCC CTCCGAACATTTA GGGGGAATGGTC CTCCGGAAATCTA GGGGGTTGGTGT CAGGCCCTGCAA GCCAACCGTAAA AACACATCAGCAC AGGAACGTCAAC GAGAGAATACGG ACCGGAACAATC GGCAAAGACCAC AGCGACGAAAAT		QIFSVRQILQKCREYQVPTHHLFIDFKAA YDSIDRIELWKIMDENSFPGLTRLIRAT MDGVQNCVKISGEHSSSFESRRGLRQG DGLSCLLFNIALEGVMRRAGLNSRGTIF TRSGQFVCFADDMDIIGRKFETVADLFT RLKREATRVGLMVNASKTKYMLVGGT ERDRTRLGSSVTIDGDTFEVVDEFVYLG SLLTADNNVSREIRRRRIISGRAYYGLQK KLRSRKIHPRTKCTMYKTLIRPVVLYGHE AWTMLEEDLQALGVFERRVLRITIFGGV QENGWVRRRMNHELAQLYGEPSIVKV AKAGRIRWAGHVARMPPDNNPVKMF ATNPVGTRRRGAQRARWIDQVHQDLE SVGHSRGWREAAMNRRGNWRNIVGEA LSR			
--	--	--	--	---	--	--	--	--	--

				GGACTAGCGATT GGAAACTCGGTA CGTGGAAGTCA AATCTCTCAACTT CATTGGAAGTACT CGCATACTCTCCG ATGTACTGAAGAC CCGCGGTTTCGAC ATCGTAGCGCTGC AGGAGGTGTGCT GGACAGGAGCAT TGGTGCGAACGTT TAGAGGTAATCAT ACCATCTACCAGA GCTGCGGCAACA CACGCGAGCTGG GAACAGCTTTCAT AGTG					
	RTE_EI e2	GCF_00220 4515.2	Aedes aegypti	TGGGTTGTAAAAT GGTGAGTCGACA ACCAGGAAGGAG	TAGATTCAACGCAAT AGCAAGGAATTGTTG	MGSIPNFSPTTKNGERRNNERYFRQPT WQRNKDYEWKLGTWNVRTLNEPGRA SLLARELQKVGVGVAAIQEVWRWPRSGE	cd0 165 0	561	821

				CGTCCAACATAGC TCTGGTCCTCGCA AGTCCCTACCTCA CGCTTCCTCGGGT CTAACGATGACAA AGACCGCCAGCT AAGGGTTGCGTA CTTAGCTGGTAGT GCAACCTGGGCA CTGTTGTCCTTCT GACATCAGCTAG AGTGAGGGGGTG CGTCCCGAGCGTC TGTTCAACAGGAG GTGCGGCTCAAA CAGCGTCTGTTCT GGTATCCAGCGG CTGAGTATGAAAT GTTGTATCACGTC AACTAACCTAAGG TGGCAGCCCCATC	CCCATCAAGTATCAA GTAAGTAAGT	REFRAVDPIANTTFKYNIYHSGGDKAEH GVGFIVIGKQMKRVMKWKPISERICVL RIRGKFFNYSLINIYAPTNDKPDDVKDAF YEGLDKAYGECPKHDVKIVIGDANAQV GREDFFRPIIGKESLHSVTDNGLRLVTF AAARGMAICSTYFARKDIRKHTWRHPN GDICNQIDHILVDGRHFSDVIDVRSFRG PNIDSDHYLVVSKIRARLSTVANSRPQR TMRFNIRQFSADGVTAEYHQKLDERISE VNVSENLNNLWDAIHGVVSETAREVIG TARRRPRSGWFDEECQRTDEKNVAR RRMLVSGTRLNRERYREARSAEKRIHRR KKKEFDENVIAEAQNCMEQNDMRIFY ETVNGVRRKSAPSPVMCNDREGNLLT DKTMVAARWKEHFETLLNGGNDRASE NRMNIDNDGQAVEPPSLDEVKKAIKEL KNNKAAGKDELPAELLKHGSEQLHRIIH HILLRIWEEEELPASWMDGLICPLFKKG HRLECANYRGITLLNSAYKIMSRILFNRL RPLEESFVGEYQAGFREGRSTSDQMFT LRQILDKFREYNLQTHHLFIDFKAAYDSV			
--	--	--	--	--	-------------------------------	--	--	--	--

				AACGTGATGTAG GTAGCGCAACCCC GGTAA		KRNELWEIMMEHGFPTKLIRLIRATLEG SKSSIRVADEISTSFVTLDGLKQGDALSN LLFNIALEGAIRRACVHRNGTIITRSYML LGFADDIDIIGIDRRAVEEAYVPFKRETT RIGLIINTTKTKYMVAGRHRGPIGDVGS EVVIGGDIFEVVVEFVYLGTLVTCNDNV TREVKRRIAAASRAFYGLRNQLKSRLQ TKTKLALYKTLILPVALYGHESWTLKEVD RRAFGVFERKVLRTILGGKLDNGIWRRR MNHELYQVYNEMDIVKRIKHGRLRWA GHVARMSEERQAKQIFSREPGRGRRLR GRPRTRWLFAVEEDLRSLNVQGDWKR LAQDRVQWRRLLHSA			
	RTE_EI e2B_A Ae	GCF_00220 4515.2	Aedes aegypti	TGGGCTGTGAAA TGGTGAGCCGAC AACTAGGAAGGA GCGTCCAACATAG CTCTGGTCCTCAC AAGTCCCTACCTC ACGCTTCCACGGG TCAAATGATGACA	TAGATTCAACGTAGC GGATTGTAGCCCATC AAGTATCAAGTAAGT A	MGSISNFLPTTKNGERRNNERYFRQPT WQRNKDYDWKLGTWNVRTLNEPGRV SLLARELQKVGVSVAIQEVRWPRTGE REFRAVDPVANTAFKYNIIYHSGSDKAE HGVGFIVIGKQMKRVMRWKPINERICV LRIRGKFFNYSLINIYAPTNDKPDDTKDA FYDCLDKTYGECPKHDVKVVGIDANAQ VGREDDFRPIIGTESLHSATNDNGLRLIT	cd0 165 0	561	821

				AAGACCGCCAGC TAAGGGTTGCGT ACTTAGCTGGTAG TGCAGCCTGGGC ACTGTTGTCCTC TGACATCAGCTAG AGTGAGGAGGTG CGTCTCGAGCGTC TGTCCACCAAGGA GGTGCGGCTCAA ACAGCGTCTGTTC TGGCATCCAGCG GCTGAGTATGAA ATGCTGTATCACG TCAGCTACACCTA AGGTGGCAGCCC CACCATCGTGATG TAGGTATCGCGAC CCTGGTAA		FAAARGMAISSTYFARKDIRKHTWMH PNGELCNQIDHVLVDGRHFSVIDVRT FRGPNIDSDHYLVVSKIRARLSTVANTR TQRPLRFNIQRLSTEGVADEYRRKLDERI RGYNVGENLNNLWESIHGAMSTAARE VIGTAQRRPRNGWFDEECQNVTDEN VARSRMLVSGTRQNRERYKEARAAEKR IHRRKKRQYEEDIITEAQNSIEQNDMRR FYETVNGVRRKTAPSPVMCNDREGNLL TDKILVAARWKEHFEQLNGENTSASE NRLIISNDGQAVEPPTLDEVRRRAIKELKN CKAAGKDELPAELFKHGSEQLYEILHRTL LRIWEEELPASWLEGLICPLYKKGHRLE CANYRGITLLNSAYKIMSRVLFNRLRPLE ESFVGEYQAGFREGRSTTDQMFTLRQI LDKFREYNLQTHHLFIDFKAAYDSVKRN ELWQIMVEHGFPAKLIRLIRATLDGSKS SVRVADETSTSFVTL DGLKQGDALSNNL FNIALEGAIRRAGVQRSGTIVTRSHMLL GFADDIDIIGIDRRAVEEAFVPPFQRETAR IGLTVNTSKTKYMIAGRQRGSNRDVGS			
--	--	--	--	---	--	---	--	--	--

						EVVLGGEKFEVVVEEFVYLGTLVTCNDV TREVKRRLLAAANRAFYGLRNQLKSRL QTKTKFALYKTLILPVALYGHESWTLKEV DRKAFGVFERKVLRTILGGKQENGIWR RRMNHELYQVYKEVDIIKLIKHGRLRW AGHVVRMPPEERQAKIIFSREPGRGRRL RGRPRTRWLFVEEDLRALNVQGDWK RLAQDRAQWRRRIHSA			
	RTE_EI e5	GCF_00220 4515.2	Aedes aegypti	ACCGCTAGCGCAT GACGGCTCACCAT AGTAGCATGTCCG AAAGAACTTGAA ACTATTGAGAACC AGAGCTACAGGT CCAAAGCCCTGAT AAAGGTGGATGG TTGTGATGGCTAT GACCGTGGTCATC ATGTAAAGAACA AGCGGACGATGC TGTATCGTGTCAG	TGATGCTACGCTGTA GCCATCAAGGTATCA AGGTAGGTA	MHDVGSATLVRWPIEDSSPTKKGKSRA NGLILRQQTRQRIKDNDWKVGSWNVR TLNEPARVGLLARELRNVGVNVAIQEI RWPRTGEREFRAVDPIANTSFKYHIYS GGDRAERG VGFIVIGKQMKRIIRWKPV SDRICVLRMKGKFFNYSLISYAPTNDKP DDMKDEFYESLDKAYGECPKQDVKIVI GDANAQIGKEDFFRPVIGTESLHSV TND NGLRLVTFAAARGMAISSTYFARKNIRK HTWRHPSGDACSQIDHVLVDGRHFSD VIDVRTFRGPNIDSDHYLVVAKIRARLSS VTSSATHRTLRFNIQRLSTDGVAAQYR QQLDEKLGRVNVAGDVDSLWDPIHEA	cd0 165 0	570	830

				CATCGAGGAGGC AGCCCCTCTAG		VTTTAREVIGTGQRRRRNDWFDEECQ RVTDMKNVARSRMLVAGTRQNRERYR AARAEKRIHRRKKRQHEESVVADAQE SMNRNDMRRFYATVNGARHNTVPVP AMCNDREGNLLTDKTAVAARWKEHF QQLNGENGNAVARNRMNIDDDQAV DPPTIGEVKKAISELKNCKAAGKDEIPVE LLKHGSEQLYQSIHRVLLKVWEDEELPT GWLDGLIRPIYKKGHRLECANYGITLL NSAYKILSRILFNRLRPLEESFVGEYQAG FREGRSTTDQMFSLRMILDKFREYNLQ THHLFIDFKAAYDSVKRNLWQIMSEN GFPAKLIRLIRATLDGSKSSVRIADEVSTS FVTL DGLKQGDALSNLLFNIALEGAIRRS GVQRNGTIITRSHMLLGFADDIDLIGID RRSVEEAFVPLKRETARIGLTINSTKTKY MVAGRDRVRHGGVSAEVVFDGDVFE VVEEFVYLGTLVTCDNDVSREVKRRVA AANRAFYGLRNQLRSRSLQTETKFALYK TLILPVALYGHEAWTLKESDRKALGVFE RKVLRTILGGKLENGVWRRRMNHELY			
--	--	--	--	-----------------------------	--	---	--	--	--

						QVYKDANIINRVKYGRLQWAGHLVRM SEERIAKIIFSREPGRGRRLRGRPRIRWL YAVEEDLATLNVRGNWRSFAQDRRRW SSTIRPAMA			
	RTE_EI e2C_A Ae	GCF_00220 4515.2	Aedes aegypti	GGGCTGCAAAAT GGTGAGTCGACA ATCAGGAAGGAG CGTCCAACACAGC TCTGGTCCTCACA AGTTCCTACCTCG CGCTTCCACGGGT CAAATGATGACA AAGACCGCCAGC TAAGGGTTGCGT ACTTAGCTGGTAG TGCAGCCTGGGC ACTGTTGTCCTTC TGACATCAGCTAG AGTGAGGAGGTG CGTTCTGAGCGTC TGTACACCAGGA	TAGCCCATCAAGTAT CAAGTAAGTA	MGSIPNSLSTTKNGERRRNERYFRQPT WQRNKDYDWKLGTWNVRTLNEPGRV SLLARELRKVGCVAAIQEVRLRSGER EFRAVDPIANTAFKYNIYHSGGDKAEH GVGFIVIGKQMKRVIRWKPINERICVLRI RGKFFNYSLINVYAPTNDKPDDVKDAFY ECLDKTYGECPKHDVKIVIGDANAQVG KEDFFRPIIGKESLHSVNDNGLRLVNFA ATRGMAISSTYFARKDIRKHTWQHPN GELCNQIDHVLVDGRHFSDVIDVRTFR GPNIDSDHYLVVSKIRARLSTVSSSRNR QSLRFNIQRLSADGVAADYHQKLDERIS EIDESVNLSDLWESVHGAVSTVAREVV GTAQRRPRNGWFDEECQRMTEKNL ARSRMLVSGTRQSRERYKEARAAEKRI HRRKKKEHEEAVIAQAQEQAMEQNNM RRFYESVNGVRRKTAPSPAMCNDREG	cd0 165 0	561	821

				GGTGCGGCTCAA ACAGCGTCTGTTC TGGTATCCAGCG GCTGAGTATGAA ACGCTGTATCACG TCAGCTACACCTA AGATGGCAGTCC CATCAACGTGATG TAGGTAGCGCGA CCCCGGTAA		NLLTDKTMVAARWKEYFESLLNGDNG SGSGSRIQIDDDGQAVEPPTLDEVKKAI NGLKNNKAAGKDELPAELLKHGSEQLH ELLHRIISRIWEEEQMPTSWLEGLISPLY KKGHRLECANYGITLLNSAYKIMSGILF NRLRPYEESFVGEYQAGFREGRSTTDQ MFTLRQILDKFREYNLQTHHLFVDFKAA YDSVKRNELWQIMSEHGFPACLRLIRA TLDGSKSSVRVDEISSSFVTL DGLKQG DALSNLLFNIALEGAIRRAGVQRSGTIIT RSHMLLG FADDIDIIGIDRRAVEEAFVPF KRETARIGLTINTTKTKYMIAGGQRGSG RVSGSEMVLGGDKFEVVVEEFVYLGTLV TCDNDVTREVKRRIAAASRAFYGLRNQ LKSRSLQTKTKLALYKTLILPVVLYGHES WTLKEVDRRAFGVFERKVLRTILGGKLE NGIWRRRMNHELYQVYKEVDIVKRIKH GRLRWAGHVARMPEERQAKIIFNREP GRGRRLRGRPRTRWLFAVEEDLRALNV QGDWKRLAQDRAQWRRLIHSAQIHRS EL			
--	--	--	--	---	--	---	--	--	--

	JAM1C _AAe	GCF_00220 4515.2	Aedes aegypti	GGGTGGTAAATG ACTGGACAATGC GTACCATTGGTAC TTCGCGTACCTGC AGGTATAAAATA GACCCCATTTGTG GTCCTTAGCCTCT TGTCCAGTAACTC CTATCCCTACCTC CCCGTGGTGCCGC CTGGGATACGAG TAACCGTAGGGA AGATCGGGTAAC CAACCCTGGTGG AACCTTGGTCGTA TGCTGACAGGGA AGGGGGGCTCCT CTCTTCTGAGGGT GTAGCTTATCAGA GCGTCTGTTCTCC ATGTTAGGGGCG	TGAAGGACGTAGAG CCAGCAAAAGTAAAG TAAGTA	MKSSVQEAYNVNSDRNNRHRPRHRKR TNDWKLGSWNCKSLNFLGSTRILSELLR VRKFDIVALQEVCKWGSTVHTYRDGYT IYQSCGNRHELGTAFIVMGEMQRRVIG WWPIDNRMCKLRIKGRFFNISIIINVHSP HLASDDDDKDAFYAQLEREYDGCPSHD VKIVIGDLNAQVGQEEEFPIIGKFSAH QLTNENGLRLIDFAASKNMAIRSTYFQH SLPYRYTWRSPQQTESQIDHVLIDGRHF SDIIDVRTYRGANIDSDHYLVTVKVRQR LSVVNNIRYRRPPRYNLRLKQPEVATE YAQSLEAALPEEGELTEAPLEDCWSSLK AAINNAEAGAIGFVEANRRNGWFDEE CQTVLDEKNAARAMMLQQGTRQNV RYKQKRRQQTHLFRDKKRLEEECE MEQLYRSQETRFYKKLNASRKGFVPR AEMCRDKDGGILTDEREVIERWKQHY DEHLNGAEEEDQDSRRNGFISTADEGD VPTPTIGEVKDAIKQLKNNKAAGKDGIG AELIKMGPDLATCLHRLIARIWDTEQL PEEWKEGIIYPIYKKGDKLECENYRAITIL	cd0 165 0	550	810
--	---------------	---------------------	------------------	---	---	--	-----------------	-----	-----

				GCTCAAAACAGC GTCTGTTCTCCAT GTTAGGAGCGGC TGATCATCGTCCT AGTGCCAGCGTG GGACTCTAAACA GTGCTGTGCACG ATGATCCTCCGGC GAGACAGGGGGT TGGTGCAGGCCTT ACAAGCCAGCCG TAA		NAAYKVLSQLIFRRLSPLASRFVGSYQAG FVDGRSTTDQIFMLRQILQKCREYQVPT HHLFIDFKAAYDTIDREELWKIMDENG PGKLTRLIKATMDSVQCCVKISGALSDP FETRKGLRQGDGLSCLLFNIALEGVMKR AGFNMRTGTFNKSSQFICFADDVDIVG RTFQVVAEQYTRLKREADRVGLKVNTS KTKYLLAGGTERDRARIGRRVTIDGDEF EVDDEFVYLGSLITSDNNCSREIRRRRIAG SRAYYGLHKTLRSGKLHFRTKCTMYKTLI RPVVLYGHETWTMLEEDLQALGVFERR VLRTIFGGVCENGWRRRMNHELAQL YGEPSITKVAKAGRVRWAGHVVRMPD NNPAKMVFNSNPAGTRRRGAQRARW FDQVEQDLGSGRSRNWRLAAMDRV SWRNIVAQVMS			
	JAM1B _AAe	GCF_00220 4515.2	Aedes aegypti	GGTAAATGGCTG GGCATGGCGTAC CATTGGTACCTCG CGTACCTGAAGG AATAAAATAGACC	TAGATGTAACTAAA TAAATGAAT	MKNQATNNQRENTNRDNRRRPQRRK GTSDWKLGTWNCKSLNFIGSTRILADV LKDRGFGIVALQEVCWKGSMTVRFRG NHTIYQSCGNTHLGTAFIGMGMQR RVIGWWPINERMCKLRIKGRFFNFSIIN	cd0 165 0	553	795

				CCTTTGTGCGGTC CTTAGCCTCTTGC CCAGCAACTCCTA TCCCTACCTCCTC GCGGTACTGGCC GGGGTACGAGTA ACCTTAGGGAAG ATCGGGTAACCA ACCCCCGGTGGG AACTTTGGTCGTA TGCTGACAGGGA AGGGGGGGTTTG CTTTTGCTTTTGCT TCTGCAAACCTTG AGCGTCTGTACTC CATGTTAGGAGC GGCTCACAACAG CGTCTGTTCCCCA TGTCAGGGGCGG CTGATCATCGTCC GAGTGCCAGAGA		VHSPHSGSTDDDKDAFYAQLEREYDSC PSHDVKIIIIGDLNAQVGQEEEFRTIGKF SAHRLTNENGLRLIDFAASKNMAIRSTY FQHSLPYRYTWRSPQQTESQIDHVLID GRHFSDIIDVRTYRGANIDSDHYLVMVK LRPKLSVVNNVRYRRPPRYDLERLKQPD VAAAYAQQLEAALPEEGELDEAPLEDC WRTVKAAINDAAESNVGYVGRSRRND WFDEECQEVLKEKNAARAVMLQQGTR QNVERYKRKRQQQTRLFREKKRRLEET ECEEMEQLCRSQETRRFYQKLNASRNG FVPRAEMCRDKDGSILTDEREVIERWK QHFDEHLNGAESTGNEGRDNGGNAFV STAEDGNQPAPTLREVKDAIHQLKNNK AAGKDGIGAELIKMGPERLAICLHRLIGT IWETEQLPEEWKEGVICPIYKKGDKLDC ENFRAITILNAAYKVLSQLIFRRLSPVVNE FVGSYQAGFVDGRSTTDQIFTVRQILQK CREYQVPTPLGIMDENSFPGLTRLIRA TMEGVQNCVKVSGEHSSSFGSHRGLR QGDGLSCLLFNIALEGVMRRAGLNSRG			
--	--	--	--	---	--	---	--	--	--

				AGGACTCTAAGCT AAACTGCGCACTA TGGTCCTCCGAAC ATTTAGGGGGAA TGGTCCTCCGGAA ATCTAGGGGGTT GGTGTCAAGGCCCT GCAAGCCAGCCG TAA		TIFTRSSQFVCFADDMDIVGRTFEKVAD LYTRLKREAAKVGLVVNAAKTKYMLAG GAERDRARLGSSVTIDGDTFEVVDEFVY LGSLLTADNNVSREIRRRISGSRAYYGL QKKLRSKKIHARTKCTMYKTLIRPVVLY GHETWTMLEEDLQALGVFERRVLRITF GGVQENGWRRRMNHELAQLYGEPSI QKVAKAGRIRWAGHVARMPDSNPAK MVFASDPVGTRRRGAQRARWADQVR IDLASVGQNRGWRDAATNRVLWREIV DSVLSV			
	Lian- Aa1	GCF_00220 4515.2	Aedes aegypti	TTAGTGGCTGCTT GGGCACGTCAGG AGTTAATCATCAA TATTAACCTCCTTT TCCCGAGCAGCCT AGATAGCCGCGT AGTGTCGGTAGC GGTTGTTTCAACT GGCTAAGAATTA ACACTACGGACTG	TAGGCTCTCTTTACG CTCATGCGTTTTGCA GTGCCCTTTTTAGGG CGCTGTTCGAACCCA TTGTGGTATGGAGCT ACATGCTCTCATTTG CTTATGCGATCTTCCC TCTTCAAGGGACCCC ACTCCTATTTCTCCC ATCTTTCCCTTCCCTT	MQNGAAQVALVQEPYFRRGNFYLGNL VDPVFATFSKLEMANSRSMPRACVLVN KAIVATLISELTTRDVCVAVTIDVSVGDLN RKYVYCSVYLPHDEPSPTDDFKRVVHC VTKGLPLIVGSDANAHHIIWGSSDINLR GSSLMEYLSSTDGLLNIGNRPTFMVSN REEVLDITLCSNRISHELTNWHVSDEESL SDHRYIFFEHSNVTAQTLRFRNPRSTN WELYIELVATKFHGYSPSIENPSDLDDA VDTTTSYIMEAFEEACPLRSVKTTTRGTP	cd0 165 0	468	725

				CCTGTTCCGGTGG TAAAAGTCCACCT CACAGGTGACCCC TAATTTATGGTGT GATGCGTACCGT GCCTAAGAATGA ATGGTTAGGGGG GTCTAAATAAAAC CTAACCGCAAACG GAGCCTGTGGGG TACCAGGGCGCC CTCCACAGTATTG AGTCCTTCCTGTG CTACCCGGAGCA ATGGTGCAGGTG ACCTTGTTTCT CCGAGATAATCG GCTGCCCTTCTTC AGTCTCTATCTTG AGGCTTAATAAG GGTGGGATTATG	TCCTCTCCCATCGGG TAGATGATGAAATAG GCTCAAATATGGCGA TGGCACAAATCTCCC AACTGGTGGGGAAC GTGCCTTTGGAGCCG GCCTTCTGATACCTG ATACCTGA	WWNSDLTRLRKQCRRSWNRRRSAGS ESFKSARKAYKKALRSAERSGWKNLCT NVSSLSEVSRLNKILAKSKDFQVNEIRLP NGDFTSSDEEVLECLFNTHFPGCVDIAS TDEPNVFSCSYESLASARSIVTTESIQWA LNSFAPFKSPGADGIYPVLLQKGFEFIKH VLKKLLVSSFATGYIPKSWRDITVKFIPK GGRASYEEAKSFRPISLTSFLLKCLERIID HHIRDVYLANMPLHVNQHAYQSGKST VTLLHKVVYDIEKAFAQKQSCLGVFLDIE GAFDNVSFDAILEAARNHGLPTMITNW IHQMLKNRHLFSTLRQAAIRKLSVCGCP QGGVLSPLLWNLVADTLRLQLNNCGFP TYGFADDYLALIVGMCISTLFDLMQSAL QVVESWCRQYGLSVNPNKTSIVLFTER RNRDGIRPLRLFGTEINVTDQVKYVGVI LDSKLSWTPHIDFRVKKACMAFGQCRR TFGKTWGLKPKYIKWIYTTVVRPILAYG CLVWWQKGEVRTIQSKLGHQLRMCL MAMSGAFSTTPTAALEALFDVAPLHIYL KQEALSCSYRLWVLDLLEKNPVNRRST			
--	--	--	--	---	--	---	--	--	--

				AATATGTTGACAT TTAATTTAAATTA TCACCTATATGGA TTCGCATTATGCG TTTTACACAGTGT ATTCTGTGCTCTT TCGCCTTTGGCGA CTTTAAATAGATA CCGATCTGGTTTT TTCGTTGCTGGTC TTTTTCGTGCTGA GATTGCAAAAAG CTTAATCCTGCCT AGTTTGGGTAGT GGCTACGGTTAG GTTAGCTCAGATC AATCTTCAGCATA AAAGAACAGCAA CGATCAATCTTTG CAGACTC		HTSLFPLLWNWDKIVLAPSDLTACNFPY RTFTTQFPSREEWTSGYLERSISNNIVCY TDGSLLEGRAGAGVYSRELRLNQFYSLG RNCTVFQAEIFALMCGVQSALQQRVM GKVIYFCSDSQAAIKALASANSRSLVIA CRTQIEELNSVNSVNLVWVPGHSSIAG NELADELARDGASHDFIGPEPAIPISKC WVKLQINSWAATQHKQYWNSLESCR QTKLYITEPSPKVAKYLTNLSKQNCSSLV RALTGHCRNLNYHMANIQRADSFVCDSC DSDYGTSYHLICNCPVFAQMRFQLLGK HLLSETEYRSLNLQDILLFLTRCGNEL			
--	--	--	--	--	--	--	--	--	--

	RTE_EI e4	GCF_00220 4515.2	Aedes aegypti	GGGCTGCAAAAC GGTGAGTCGACA ACTAGGAAGGAG CGTTCAACACAGC TCTGGTCCTCACA AGTTCCTACCTCA CGCTTCCACGGGT CAAACGATGACA AAGACCGCCAGC TAAGGGTTGCGT ACTTAGCTGGTAG TGCAGCCTGGGC ACTGTTGTCCTTC TGACATCAGCTAG ATTGAGGAGGTA CGTCTCGAGCGTC TGTTCAACAGGAG GTGCGGCTCAAA CAGCGTCTGTTCT GGTATCCAGCGG CTGAGCAAGAAA	TGACTGGTAAGGTAA GTA	MPHQDRPSVSWDVKQSWHDGAPA RQECWRRPNKPPVKNLIANNIGDNTTR NNRQRPTRRNKDYDWKLGTWNCKSL GFAGCDRIIYDELHPRNFDIVALQELCW TGQKVWKS GHRAATFYQSCGTTNELG TGFIVLGKMRQRVIGWQPINARMCKM RIKGRFFNYSIINVHCPHEGRPDDEKETF YAQLEQIYDGCSPRDVKIVVGDMNAQ VGREEMYRPPVIGRNSLHAVSNDNGQR CVNFAASRGMVVRSTFFPRKDIHKTTW RSPDNLTENQIDHVLIDGKFFSDITNVR TYRSANIDSDHYLVAVCMRSLSTVTNS RRSRTPRLNIERLRDSEVAQEYAAQQLV ALPTEEQLGAATLEDGWRHIRSAIGSTA TAALGPAAPNQRNDWYDGECEQLKNE KNAAWARMLQHRTRANEARYKQARN RQNSIFRRKKRHQEERDREAMEELFRA KDTRKFYEKLNRSRKGFVPQADMCRDC DGNLLTNEREVVERWRQHYDEHLNGD VASSGGGTEITLGARADDERLPPDLQE VETEIRRLKNNKAAGVDQLPSELLKYGG	cd0 165 0	588	848
--	--------------	---------------------	------------------	--	------------------------	--	-----------------	-----	-----

				TGCTGAATCGCGC ACAGCTAAATCCA AGGTGGTAG		DALARALHWVISKIWEEVLPEEWME GVVCPYKKGDKLDCANYRAITILSAAYK VLSQILCRRLSPIAREFVGQYQAGFMGE RATTDQIFAIRQVLQKCREYNVPTHHLF IDFKSAYDTIDREQLWQIMHEYGFDPKL IRLIKATMERVMCVVRVSGTLSSPFESR RGLRQGDGLSCLLFNIALEGVIRRAGIDT SGTIFRKSVQLLGFADDIDIVARNFETM ADTYIRLRAEARRIGLNINVSKTKYMIAR GSREDTARPPRVHIDGDEIEVVVEFVY LGSLVTADNDTSREIQRRIVAGNRAYFG LRRTLRSSKVRHRTKLTITYKTLIRPVLLYG HETWTLRAEDQRALGVFERKVLRTIYG GVQMEDGTWRRRMNHELHQLLGEPT IVHLAKIGRLRWAGHVTRMSDTPVK MVLENNPTGTRRRGAQRARWVDQVE DDLRTLRRMRGWRRRAAMDRVEWRRL LRTAEATQALA			
	CR1- 10_A Mi	GCF_00028 1125.3	Alligat or mississ	CT	TGACCCCCAGTTGTC TTTCCTGCCTAGTGC AGGGGGCTGGACCC	MGQGFKLRLLDGETLIGACPGTNTQE SPVDRDTRRATLAGAEGIVLPQHDKS RVSVGGLRCLYTNARSMGNKQEELALL	cd0 165 0	539	790

			ip pi en sis		GATGATCTTCCAAGG TCCCTTCCAGCCTTAC AATCTATGAATCTAT GAAT	LSSGHYDLVGITETWWDSSHDWSVAIE GYTLHRRDRVGGKGGGVALYVKEQYTS LEADIGSKEGQLETWVKLWGGGGER DLTVGVYYRPPNQGEELDLEFSRELTEA AHARDMVVMGDFNYPDISWEEHSAK SDQSRSLTCIEELFLTQEVHGPTRGNAL LDLVLAKGDDLVTDLNIEGSLGSDSHET IMFTVHHKADKSDSRIEVLDFKNANFN KLRLVGEVLWDQGPKVKGVHEEWLF LKDTILEAQRKSIPMQRKGSRRAGKPS WLNREIMVLLRKKKEAYKQWKLGTVPK EDYTTMAHTCREQIRKAKAATELRLAT GIKDNKKSFFRYVGNRRKTNGSVGPLL NNSGQLVTDTEKAELLNDYFASVFHQ TKGKIMPGRVQDEHGRNDHLPTIAVEL VYDQLEKLDIYKSAGPDELHPRVLKELA EVIAEPLAKLFQNSWCSGEVPEDWKRA NVVPIFKKGRREDPGNYRPISLTSIPGKIL EKFIKESICDKLAEGKILNDSQHGFIAGR SCLTNLISFYDQVTHHLDKREEVDIVYLD FQKAFDLVSHDVLMRKLEDCGLNCETV			
--	--	--	---------------------------	--	--	---	--	--	--

						RWVGNWLQGRTQRVLVNGSVSSWRE VASGVPQGSVLGPVLFNIFINDLDVGK SSLAKFADDTKLWGSVVTLEDRLQIQA DLDRLASWADQNQMKFNVEKCKVLHL GTNNPQHTYRLGGTKLTSTMNERDLG VIIDHSMNMSHQCDASRANNTLACI NRCISSKTKEVILPLYSALVRPQLEYCIQF WAPHFNKDVEKLERVQRRATHMIRDL QGKPYEERLRDLGLFSLKKRRLRGDLVA AYHYIRGVHQGLGEQLFTRAPLGKTRS NGHKLLEDRLFRLNIRKNFFTVRVSRLW NKLPPPEVVQSPTLEIFKRRLDSHLAGVT			
	CR1- 2_ASi	GCF_00045 5745.1	Alligat or sinensi s	CCTGACCGTATTA TGAGGGACTTCG AGGCTCTGGGGA CCAGGCTGAAGG AGACGGGGGCGC AGGTGGTTTTTC CTCTCTCCTCCG GTGAGGGGGCGG GGACAGAGGCAT	TGAGTCCCGTTTTTGT TTTTAGGGTGAGTGG TTTTGTTTGTGTTTTT GGGTTGTTCTCCCGC CCTGGCGGGGGGTC GGGCTAGAAGATCTA TTAGGTCCCTCCGA CCCTGAAACTATGAA ACTATGAAA	MGNKQEELALRIASSNPDVVGLTETW WDSSHDWAVNIRGYRLYRRDREGRKG GGVALYVKEHYTSSASSMGSGEGRAEV LWVKIQGGQGERDLTVGVYRPPNQE EELDREFSGQLAETLKARDVVVMGDFN YPDICWEEQSAGTDRSRRFLAELQDLHL IQEVHSPTRGDALLDLVLAAGDDLVRGL QVLDHLGNSDHRVLEFTIQRRVSRARS KVVALDFRRANFSEMRRLVGEELGSLR	cd0 165 0	470	720

				GGGACCTGCATT GGTGAGGTCAAC CAGCGCTTACGCA GCTGGTGCCGCC GGGCAGGCCTTG GGTTCCTTGACAA CGACCCATGCTTT CGGGAGGGGGG ATTGCTGGGGCG GGATGGGCTTCAT CTCTCCCCTAATG GCAAGCGTATCTT CTCTTCAGGTTG GCTGATCTCGTAC GGCGAGCTTTAA ACTAGACTCGCCG GGGGAGGGGGCT GCAGGAAGACGA GGAGACGCCATC AGGGCAAGGGGT GCAAAGGTAAGT		AEELSVQEEWSFLKETILGAQGVSIPTRS KAGRSVQKPPWLTKGVRECLLAKRVAY ARWKGGVISKEEYTSSARACRRAVRKA KAEIELGLASRIKDNKKSFFKYIGRMKKA PGNVGPLQDMRGNLVVASEDKADLFN QFFTSVFLCRDRDFPAVIEAGREGIAFRP QVGEDRVRELLEGLDAFKSAGPDALHP RVLRELAGVIAGPLARLYERSWCLGQVP DDWKVANVVPIFKKGKREDPGNYRPV SLTSILGKLFEKIIKEHICDGPASEMMLR GNQHGFIRGRSCQTNLIAFYDQVTKAM DAGVAVDVVFLDFSKAFTVSHPILINK LGDCGIDAYTVRWIADWLKGRTQRRV VDGSYSTWGEVSGVPPQGSVLGPALF NFFISDLVDGVESNLFKFADDTKIWGEV STLAGRERLQKDLRLQGWANRNRMC FNTDKCRVLHLGSRNQHTYTMGSSLL ESTEERDLGVIIDSKMNMGRQCEVM VGRANRTLSCIHRICISSRAREVILPLYATL VRPQLEYCVQFCAPHFRRDVDNMERV QRRATRMIRGQQGRPYSERLRDLNLFS			
--	--	--	--	---	--	---	--	--	--

				AATGGGGGTCTT GCCATCAGGGCA AGGGGTGCAAAG GTAAGTAATGGG GGTCTTGCCATCA GGGCAAGGGGTG CAAAGGTAAGTA ATGGGGGTCTTG CCATCAGGGCAA GGGGCTGCAGGA AGATGGGGGGTC GCCATCAGGGCA AGGGGCTGCAAA GGTAAGTAATGG GGGTCTTGCCATC AGGGCAAGGGGT GCAAAGGTAAGT AACAGGGGTCCT GATGGTCTGGCTA ATCAGGGGGTGG GGGGGCACACAG	LHKRRLRGDLVTYKLRGNQKGLGETL FPLAPAGITRNNGHKLMERRYRLDIRKN YFTVRAARSWNQLPREVVLAPTLGIFKK RLDTYLAGLI			
--	--	--	--	---	--	--	--	--

				TCTTGCAAGGCCT TAAATGCCTGTAC ACCAATGCTCGTA GC					
	L2- 18_AC ar	GCF_00009 0745.1	Anolis carolin ensis	GAGAGGCCTGCT GTTGTGCTGCAGA AGAGGCCAGAAA CATCTCCAGCCAG CCTGCCTGTTGCT GCCTCTGGGACTC CTTGTGAGTAGAC TGCAGCCCCAGG CCCTTTCCATCTG GGTGTGACTCAG CACTTCCTAGGAA GGACCATCTATAA AGCCTGAGTCCAC AAGGCCCTGGCA GGGCTGAGAGGC CTGCTGTTGTGCT GCAGAAGAGGCC	TAAGAATATATGACC TCAACTGACAGACCC AGTCACATGAACATG GAAAGCGCCTTAAAT GTATATTTAAATGTA TAATTATGTATATTT AATGGCTATTCTTAA TTTTATTGTAATTTT AATCTTGATGGATT TTAAATTTGATGAAA ACTGTTTTTAATGTGT ATGGTTATACTGTAA GCCGCCCTGAGTCCC CTGATGGGTGAGAA GGGCGGGGTAGAAG TGATGTAATAAATAA ATAAATAAA	MLLMHPLHFTTAFALCIVGGGWGRLE GYGDCRVGEGNGCERMPIYNRKLVNQ ITSFGPGEEKGKGQCDRGGGCSDBGW APIEVVSGRGRYGKGRPNHLIRPRVFGR TLIVPKRSPDIANLGSQDGGPSRLKVVL NARSVNGKTSFIQDLILDEHVDLACITET WLDEMGGVNLTQLCPPGYSVQHQP SGGRGGGVAVVYKDSIPLTRCPVPQST MFECVHLRVGDRDLGILLVYRPPRCTT VSLPELAEVVSGLALESRRLLVLGDFNVH AETNLSGAAQDFMAAMATMGLSQLV TGPTHSAAGHTLDLVFCQGWEEGGGVE ELTIAPLPWTDHHLIRFRLTAPPNLCRG GGPKMVRPRRLMDPEGFLTTLGEFPAS SAGDPVDALVSLWNGEMTRAIDTIAPE RPLSSNRAKPAPWFTEELAAMKRKKRE LECVWRSGYNESDRTRLGSLRAYAAAI	cd0 165 0	610	870

				AGAAACATCTCCA GCCAGCCTGCCTG TTGCTGCCTCTGG GACTCCTTGTGAG TAGACTGCAGCCC CAGGCCCTTTCCA TCTGGGTGTGACT CAGCACTTCCTAG GAAGGACCATCT ATAAAGCCTGAGT CCACAAGGCCCTG GCAGGGCTGAGA GGCCTGCTGTTGT GCTGCAGAAGAG GCCAGAAACATCT CCAGCCAGCCTGC CTGTTGCTGCCTC TGGGACTCCTTGT GAGTAGACTGCA GCCCCAGGCCCTT TCCATCTGGGTGT		KAARRAHLAANIAASAQNRPAALFRVIR GLLTPTNQGGSPDNSAARCEAFARFFA DKVALIRSDFDILTAVSEDVTSAPACPV LLDSFQLVELEDVDKILGEARPTTCILD CPSWLVRREARGGLAEWVKVVVNASLR EGKFPASLKMAVIKPLLKKPSLDPTQFG NFRPISNLPYLGKVLEHVVASQLQGFLV DTDFLDPAQSGFRPGHGTETALVALVD DLRRELDRGSVSLLVLLDLSAAFDTV DH GILLGRLAGMGLGGTVLQWLHSFLEGR SQMVSLGDACSAPQPLTCGVPQGSVLS PMLFNIYMKPLGEIIRSFGVRCHLYADD VQLCHSFPPVTKEAVQVLNRCLAAVSD WMRANKLKLNPDKTEVLLVSRKAEQG TGLQPVLGDGVTLPKQVRS LGVILDSS LSLEPQVSAVARGAFAQLKLVQRPLPYL GKPDLATVVHALVTSRLDYCNALYVGLP LKTVRKLQLVQREARLITGAAYRERTT PLLRQLHWLPISYRAQFKVLALAYKALN GSGPAYLSERVSRYDPPRSLRSSDEALLA VPSASQVRLAGTRDRAFSV VAPRLWN			
--	--	--	--	---	--	---	--	--	--

				GACTCAGCACTTC CTAGGAAGGACC ATCTATAAAGCCT GAGTCCACAAGG CCCTGGCAGGGC TGAGAGGCCTGC TGTTGTGCTGCAG AAGAGGCCAGAA ACATCTCCAGCCA GTCTGCCTGTTGC TGCCTCTGGACTC CTTGCCTACTGCC CCCCTCCCTTTTTC CCAGCGCTGGTG AGTGGACATAAG AAGGGAGGGAAG ACTTTTAAAACTC CAAAGCCATCCG GAGGAAGAAGGC AGAGGAACTGAC TTTAAATCATTTC	ALPIEVRQAPSLLSFRKRTKTWLWDQA FEHESAAKMR			
--	--	--	--	---	---	--	--	--

				TGTCTCTTGCTGC AGGCCTACTGCCC TCCTCGCTTTCCCC AGCATTGGCTATT GAACTAAGAGTCT GTCTTACTGAGGT CCACTCATTATTT ATATATTGCTGCT ATAACAATCTTTA AATGCTGTTATAA TAACTATTATGCA CTGTGAAGCCTG GTCCCCCTGACTA GTTTACTAAGTCT ATCATCTCTAATT GGTCCTGCCCTGG CACCTTATAATAC CATATTGCACCAT ATTGAATTAGTTC ACCATCATAATCC CCACTGCCTGCCT					
--	--	--	--	---	--	--	--	--	--

				TGCACTTTAGTCA GCTCAGACCCCTA TACCT					
	RTE- 1_AC_ 1	GCF_00009 0745.1	Anolis carolin ensis	CCCCCTCGATGGA TTGTCACCTTGCC GTGGTGAGGGGG CTTGCGTGTTCCG ATGAACCTGTGG GCACAACCACTG GAGTGATGCACTC CCAGGAGTGGCC GCAGGGGAGGTT CCAGACCAAGCA CAATCCGAAGACC CAAAGACCTCAAC GGCGGAGCAGGC GGAGGATAACAT GGTACATGTTACA ACGGCTGTGAAG GCGGAAGAAGGC TGCAACAGACTG	TAAGTAA	MLTKTSRKIFYCQLDTILSEIPKEDKIILLG DFNARVGRDSDLWPGIIGKDGVGNSN SNGILLTKCGEHNLVITNTLFRQKNKLK TSWKHPRSKHWHLLDYVITRARDRRD VLLTRAMTGTDDCWTDHRLIRSTMAIK IVPKRRLQGRKTRRKMNTQALQEPSKR ALLQTTLKDHLPTEHPENVEEHWNKLK TSIITACEESIGYLTKKHQDWFDNDKEI QQLIDNKRKAFQTWQRDTNCAAKKKIY ASAKAEVQRRRTRELKNIWWTKKAEIIQ HLADTHDAQGFFKATKIIYGPRNHGIQP LRSSDGTKILKDKTSIALRWKEHYQNLL NRSSNVAEETLSQIPQQQTRDELAALPS LEEVSNAISQQKNNKASGPDGIPAEIFK EGGPELIQQLHQLIEKVWMTEKIPADFK DATIITLFKKGDRTDCGNYRGISLLTSAG KILARILANRLLPVSEDTLPESQNGFRPS RGTVDMIFTARQLQEKCREQNQPLYM	cd0 165 0	417	698

				AGAAGCCACTGTC GTTGTGTAAACCA CACCACTGCTGGA ACCTCACTCTGTG AAGACTGTGTGTT GACCGGCCGTGC ACCGACCTCCACA CATTAAAAAAAAA TCACGCACAGGC GTCTTCCAACAAA AATAAAAACAAA CCTACAAAAGTCC CATGGCGATCAG CGAGTGGCGACG GGGGCAGGACTG TGAAATCTGGAA GCCCCTAGTCACA GACTGGCACATG GGCGGTGGATAT GGA CTAGTCGTT CTACCTCAAGGAC		AFIDLAKAFDTVNRSALWTILQKIGCPD KFNILRLLHDDMMATVLDSNGSQSD PFKVESGVKQGCVIAPTLFSIFIAMILHL VDGKLPTGVEIYRTDGKLFNLSRLRAKT KVTTTSVIELQYADDNVVCAHSEEDLQ ATLNTFAEAYEKLGLSLNIEKTKVLFQQA PANPSAKPGIQLNGVTLENDHFRYLG SHLSTKVNIDTEIQHRLSSASAAFFRMK QRVFDDRRDIRRETKVLVYKAIVLPTLLYA CETWTVYRRHTKLLERFHRCLRKILQIS WEDRRTNVSVLEEAKTTSIEAMLLRHQ LRWTGHVVRMPDHRPKQLLYSELKN GKRNVGGQEKRFKDKLANKNCGIDT ENWEALALERSNWRSVAVTSSAAEFEEA RTEGLREKRAKRKERQANPDRDRLPPG NRCPHCGRICGSRIGLFSHLRTHAQDTK VGRPSSSNYEGSPK			
--	--	--	--	--	--	--	--	--	--

				CGAGGCAGTTGA GTAGTCCGGCAG CTGTATCCATGAC TGAGCAGCCCTTT TTAGGATCCGCTC TGCTCACCCACA CGGGGAAGGGGC TAGAAAAGGTGC CCTAAACATAGTC TGCCTCTCCTACC CTGACTGGACTGC CGCGTCCAGAGG GGTCACCACTCTG CGGCCAAAAAAG AAAAATGAACTTT GGAACATGGAAC GTACGGACATTGT TAGATAA CACTGA CAGCGAACGCCC CGAACGCAGGAC TGCTCTCATTGCA					
--	--	--	--	--	--	--	--	--	--

				AGGGAGCTGGGA CGCTTTAAGATCG ACATAGCAGCCCT TCAGGAGACCCG GAGAGCAGGAGA GGGACAGCTGAA GGAAGAAAAGGG AGGCTACACCTTC TTCTGGAAGGGA CTGCCTGAAGAA GAGCGAAGAATA CACGGAGTTGGC TTTGCTATCAGAA ACGACCTGGTGA AGCACCTGACTGA AGCACCCACTGGC ATCAACGAACGA CTTTCAACCCTCC GAATTAACCTTGC CAAAAACCAACA GGCAACCATCATA					
--	--	--	--	---	--	--	--	--	--

				TGCGCCTATGCAC CAACTCTAG					
	Vingi- 1_Acar	GCF_00009 0745.1	Anolis carolin ensis	GGGGGACACGGA AAGAGCCTCCCCG AAGATTGAGTGA ATTCAGTCGGGC GTCCCCTGGGCAA CGTTTCTTGTAAG CGGCCGATCTTTC CACCCAAAAGCA TTGGATGA	TAGTTGCTTGTGATTT CTTTCTTTTTTATTTT ATTTCCATTATTTGAA ATGTATTTGTTGTAG CAATGCTTTTGACAC GAAATAAATAAA	MDEYQRSLSRPLLTIMSINIEGLSLAKEE LLAKMSEDISCDILCIQETHRDITMRRPK ILGMQLAVERPHRQYGS AIFVRSGVAIS ATSLTEVN NIEILSVELDSCTVSSLYKPPG ADFYFTPPTSCHNHEAHFVVGDFNSHS CVWGYDEDDRNGEAVLTWADNSRMS LLHDSKLPPSFNSGRWKRGYNPD LIFVK ESISHQCTKRVLNPIPN TQHRPICCVAY AAVRPKSVPFRRRYN FNKANWTKFTET LEAAISDIEPSIENYDLFVEAVKRSSRLSIP RGCRTSYLPGLNEESLNQLQEYLRLFQE NPYSDGTIAAGQKLSTALANAKKDRWI ELLENLDMKSSRKAWQLLRRLSDPL VNPGHANVTPDQIAHQ LIQNGKTNCS RIKMKINRVPELETHQLSSPLNLKELREA IKRCKTGKAPGLDDLMMEQIKHLGPKA ENWLLKFYNQCLAHKQIPRAWRKTKIIA ILKPGKDASNARNYRPISLLCHLYKVYER MLLNRLGPVIEPKLIAQQAGFRPGKNC	cd0 165 0	469	724

						TGQILHLTEHIEEGYEKGCITGTVFVDLT AAYDTVQHRKMLHKVYHITRDFDFTKT VQTLENRSFYVEFQGQKSRWRRQKN GLPQGSVLAPTLFNIFTNDQPQPPLTKS FIYADDLGLTTQAKDFETVEKQLTNALK DLSSYYKENHLKPNPAKTQVCAFHLRN REANRKLKVTWEGQELEHCFHPKYLGV TLDRTLTYRKHCMNTKHKVAARNNILR KLTGSAWGADPQVIRTSALALSFSTAIEY ACPVWHKSAHAKQVDIALNETCRIITG CLKPTPVDKLYKLAGIAPPDVRREVAAN GERKKVEHCESHPLHDYHPPPTRLKSRL GFMRTTTPLDVPPATARVSLWAAKPG NSNWMAPQEGLPPGANQEWATWKS LNRLRSGVGRSKDNLARWHYLEESSTL CDCGAEQTTQHMYACPQCPASCTEEE LFKATDNAVAVARFWSKI			
	RTE- 1_AG	GCA_00154 2645.1	Anopheles gambiae	TGCTCTGTAATGG GATGGGATCCGA AGCCCCATTGAG GGATAATACAGG	TAGCGCCGGATAAGT AAGTAAGTAAG	MGSWNVRTLSEAGALKKLDDALATLS MDLVALQEIRWLGNGVHNRRGKHICY DIYYSCHDRHRVLGTGFAVGPRLPKPAI MDFKAINDRLLCTLRMRGKFFNISLINVH	cd0 165 0	521	783

				CTCTCCCATCCAA CTCCTATTCCGAC ACGTCCTCGTCGT GCAGAGTGGTAA CATGTGTCACCAC ATATCCAAGCTTG GGTACACGGGCT TGACCCAACCCCT TGGGCGGATGGT GGCACATGGCGA ACCAGGAGGGGG GTGGGTATCCCG GGAAACTGGGTG CCTGAACGTCGG GGGGGCAACCCG ACGCTAAACAAA ACGGTCACGTGG GCGCAGGGCCAG TCACCCAGTCCCA TGAATCAACGACT ACGGAAAGCACC		APTEDKEEEEKDLFYGRLARIVDACPRH DLIIILGDFNAKVGREPMYRQYTGCHSL HEHSNDNGSRLVQFAAANNLVVGSTK FARKKIHKITWAHPGGESFNQIDHVLIS RRRQSSLLNVRTYRGANIDSDHYLVGLV IRCRIARPRANGGENTQARLNTDSLRL DIAVQQEFKTALEESLLPEDRYETTSE WNALKTKIINCARNILPPRRGNTKSGW FDDECRQVTERKNTAYRAMQQRHRTR ACAEYSRLRREEKRVHRSKKHALEEQN MRELEQTREAYGPTRKFYQAIAGHRNN VVPKVTCCRNKDGDLVSNQPEVLSRW AQYFDELLNDQFSEQLEAPLADNVMLL PPSIEETRKAIRRLKNNKAPGTDGIAAEL VKNGGARLENEIHQIVTEVWDSESMPC DWNLGIIYPIYKKGDRLCDNNYRGITVL NTAYKIFSLILQDRLVPHVEEIVGNYQRG FRNGKSTTDQIFTMRQILEKMAEYRND TYHLFIDFKAAYDSIARVKLYDAMSSFGI PAKLIRLVRMTMTNVTCQVRVDGKLSG PFATTKGLRQGDGLACLLFNLALEAIR			
--	--	--	--	--	--	---	--	--	--

				AGAAATATTCGAA ACGGACTTAACG ATACCGACCCTAC GCAATGCCCCCG GACTACTCTGAAA		DSRVETTGTIFYKSTQILAYADDIDIIGLR LSYVAEAYQGIEQAAENLGLQINEAKTK LMVATSADLPINNPNNLRRRDVQIGERT FEVVPEFTYLGSKVSNDNSMEVELRAR MLAANRSFYSLKKQFTSKNLSRRTKLGL YSTYIVPVLTYASETWTLSKSDEALLAAF ERKMLRRILGPVCVEGQWRSRYNDELY EMYGDLTVVQRIKLARLRWAGHVVRM ETDDPARKVFLGRPQGQRRRGRPKLR WQDGVESAIAKAGITDWQTKARDRER FRTLRLQAKTAKRL			
	RTE-5_AC	GCF_00000 2075.1	Aplysia califor nica	TGGCTGGAGCTA ATGCTCCTGGTAG GGCCACCCATGCC AGACAGGTCGAC GGGTAGAGGTCA GACATAGCGTGA TCCACCGGCCCTC CAGGTTTCGGAG GTTCAGCTCAGG GCTAGCGACCCA	TAGTAAGTAAATAAG TAA	MNAGESSRQEAVCRKIEDLNTAKTKVII GLWNVKTLYGTGKSAQVTAEMRRYGL HVLGISESQWTGSGWLRTATGETLIYF GREDNQHREVVATILNKGLEKCLMEW KPINSRMMTLRLKGKHTNTSIIQCYAPT NDESESKDYFYEELQRGLTGIPRHDIV VMGDLNVKVGEDNTNYERTMGQHG MGTRNENGERLVDFCALNNLVIGGTLF PHRTIHNLWNPNRGRDKNQIDHLM GGMLRRSLQDVKVRRGADVGSYHHLV	cd0 165 0	557	811

				GACTGGTAAAAA GCATATGTTACGG AAACAGCAACAA AGAGAAACCATA TTTCTGTGTGTGT TGGACCTTCAGAG TCGCTTGTGACTC GT		TSRIKFKLKKFTTREATRTQFNIQKLSDG NVTSAFRLELKNRFQALQDLEETTPNTD EVNTQWKLIVEAYTESSKTCLGVKKKNA SNEWIQQGTWNIIERREIRKRELAACS IRLQEKYAEDYKEVCQKVKRLVREDKRE AMENLATETEEAAKGEQGNVYKITRII CGKYHGGMSGPIKGIGRLLTIEKEQEA RWPEHFREVLNQPPPEDTPDIQEATLN LDISSDPPQRHEVVKAVKSLKEGKSPG MDNLNAELFKVDPELTTTILQPLFKTIRK GEQITEDWRTGVIVKVPKKRALGDLNN WTGITLLSVASKILAKIIINRMSDAVDICL RKEQARLRKERRCTDQIFALRNITEQCT ERQRQLYVNVVDFEKVFDGVHRESLW CILRSYGIPGKIVEIKSFYQSFKHVVG RTTAFEVKTGVRQGCNMSAIMFNLAID WVMRKTTEDTSRGIRWTLFSTFEDLDL VDDLAQLSHTNQHMKEKTNPLHKYER HIGLRISQKKTEMMTLNVSSPTLIMENN AHLPTDKFTHLGSINKTDGGTKEDIMN RLGKERSVHRSVINIWKSEQFSIQTTLK			
--	--	--	--	--	--	--	--	--	--

						QYNSCVLSTLLYGFEWWRKTGHDRKQ SSFHTTSLRQILRIFWPDNISNTDLLQK WRQEDRTTIKRRRWRLIGHVLRREPDS VVKTAHWTPEGHRKRGRPKVTWRRK VESEMKSWNHSWGTCLKGQNRQSW RSFVAARNSRWRNG			
	BovB	GCF_00226 3795.1	Bos taurus	GCGCAGGCGGCT GCGACGGGCGCA CTAAGCGCGGCC GAGAGGAGCTAC CCCACGTCCGAG GTCAGGGGCAGA AGCCGGGAGGAC CCCATGCCCAGG GGCGGCGGCCAA GAGGAGTTACCC CACGTCCGAGGTC AGGGGCAGCGGC CGAGAGTGCCAG ACTGCGACGGCG CAGGAACGGCCG	TGAACTGAACTGA	MKLSHARGATQDGRVMVERSARMW STGEGNGKPLQYSCLNPMNSMKRQN DRILKEELPRSVGAQYATGDQWRNNSR KNEGMEPKQKQYPAVDVTGDRSKVRC CKEQYCIGTWNVRSMNQGKLEVVKQE MARVNVLDILGISELKWTGMGEFNSDD HYIYYCGQESLRRNGVAIMVNKRVRNA VLGCNLKNDRMISVRFQGGKPFNITVIQV YAPTSNAEEAEVERFYEDLQDLLELTPK KDVLFIIGDWNAKVGSQETPGVTGKFG LGIRNEAGQRLIEFCQENALVITNTLFQ QHKRRLYTWTSPDGQHRNQIDYILCSQ RWRSSIQSAKTRPGADCGSDHELLIAKF RLKLKKVGKTTRPFRYDLNQIPYDYTVE VRNRFKGLDLIDRVPDELWNEVRDIVQ	cd0 165 0	610	861

				AGAGGAGCTACC CCACGTCCGAGGT CAGGGGCGGCGG CCGAGAGGAGAT ACCCAGCGTCCGA GGTCAGGGGCGG CGACGAGAGGAG TTACCCCGCGTCC GAGGTCAGGGGC GGCGGCCGGGAG GAGCTACCCACG CCCCTAAGCCCGA GGCCAGGGGCGG CGGCCGGGAGGA GCAACCCACGCC CGAGGCCAGGGG CGGCGGCCGGGA GGACCAACCCAC GTCCAAGGAGCC GTGGCTGCGCGG GCGCAGGAGGGC		ETGIKTIPMEKKCKKAKWLSGEALQIAV KRREVKSKGEKERYKHLNAEFQRIARRD KKAFLSDQCKEIEENRMGKTRDLFKKI RDTKGTFHAKMGSIKDRNGMDLTEAE DIKKRWQEYTEELYKKDLHDPDNHDGV ITDLEPDILECEVKWALESITTNKASGGD GIPVELFQILKDDAVKVLHSICQQIWKT QQWPQDWKRSVFIPIPKKGNAKECSN YRTIALISHASKVMLKILQARLQQYVNR ELPDVQAGFRKGRGTRDQIANIRWIME KAREFQKNIFYCFIDYAKAFDCVDHNKL WKILKEMGIPDHLTCLLRNLYAGQEAT VRTGHGTTDWFQIGKGVRRQGCILSPCL FNLYAEYIMRNAGLEEAQAGIKIAGRNI NNLRYADDTTLMAESEELKSLLMKVK VESEKVGLKLNIQKTKIMASGPITSWEID GETVETVSDFIFLGSKITADGDCSHEIKR RLLLGRKVMTNLDSIFKSRDITLPTKVRL VKAMVFPVVMYGCESWTVKKAERRRI DAFELWCWRRLLRVPWTARRSNQSILK EISPGISLEGMMMLKLKLQYFGHLMRRV			
--	--	--	--	--	--	---	--	--	--

				CTAGAGGAGCTA TCCCACGTTGAAG GTCAGGAAGGGC GGCGGTGAGGAG ATACCCCTCGTCC AAGGTAAGGAGC AGTGGCTGCGCTT TGCTGGAGCAGC CGTGAAGAGATA CCCCACGCCCAAG GTAAGAGAAACC CAAGTAAGACGG TAGGTGTTGCAA GAGGGCATCAGA GGGCAGACACAC TGAAACCATACTC ACAGAAAAC TAG TCAATCTAATCAC ACTAGGACCACA GCCTTGTCTAACT CA		DSLEKTLMLGGIGGRRRRRGRQRMRL DGITDSMDVSLSELRELVM DREAWRA AIHGVAKSRTRLSD			
--	--	--	--	--	--	---	--	--	--

	L1- 1_Cho	GCA_00016 4785.2	Choloe pus hoffm anni	GAGGCGGGGCAA GATGGCGGAGTG GTGAGGTGTATG TTTAGTTACTCCT CCAGGGCAGTAG GTAGAAAGCCAG GAACTGCGTGGA CTGGACACCGCA GAGCAATTTGACT TTGGGCATACTTC ATACAACACTCAT GAACACGTGGAA CAGCTGAGATCA GCGAAATCTGTAA GTTTTGTGGCCA GGGGACCCGCGC CCCTCCCTGCCAG GCTCAGTCCCGTG GGAGGAGGGGCC GTCAGTGCTGGG AACGAGGAGGGA	CAATAGTACAGACAG CCCAATTATAAAATG GGCAAAAGATATGA AAAGACAGTTCTCTG AAGAGGAAATACAA ATGGCCAAGAAACAC ATGAAAAAATGTTCA GCTTCACTAGCTATT AGAGAGATGCAAATT AAGACCACAATGAGA TACCATCTCACACCA ATTAGAATGGCTGCC ATTAAACAAACAGGA AACTACAAATGCTGG AGGGGATGTGGAGA AATTGGAACCTTAT TCATTGTTGGTGGGA CTGTATAATGGTTCA GCCACTCTGGAAGTC AGTCTGGCAGTTCCT TAGAAAACCTAGATAT	MADSRNAFTVITLNVNGLNSPIKRYRFA EWIKKNEPSICCIQETHLRHRDTKKLKVK GWKKIFHASYSQKKAGVAILISDKIDFKC KDVMRDKEGHYILIKGAIQQEITIINVY APNQGATKYMQRQLAKLKEAIDVSTIIV GDFNTSLSPIDRSTRQKTNKEIENLNNLI NELDLTDIYRTLHPKSPGYTFFSSAHGTF SRIDHMLGHKTSLNKFKKKIEIIQSTFS HNGIQLEVNNHQRLRKFTNTWRLNNT LLNNQWVKEEIAREIAKYIETNENENTT YQNLWDAAKAVLRGKFIALNAYIKKEER AKIKELMDQLKKLENEQQTNPKPSRRK EITRIKAEINDIENKKTIERINNTKSWFFE KINKIDKPLARLTskREKTHINKIMNEK GDITADPEEIKKIIRGYEQLYANKLDNV EEMDNFLETYEQPRLTREEIEDLNQPITS KEIQSVIKKLPTNKC PGPDGFTGEFYQTF QKELTPILLKLFQNIENGTL PNSFYEANI NLIPKPGKDATKKENYRPISLMNIDAKIL NKILANRIQRHIKKIIHHDQVGFIPGMQ GWFNIRKSINVLQHINKSKEKNQMIISI	cd0 165 0	512	774
--	--------------	---------------------	--------------------------------	---	--	---	-----------------	-----	-----

				GAAGTGCAGTTGC AGCTCTTATTGGA AACTCATTCTACT GATCCAAACTCCA AACATAGATAGA CTGAGACCAGAC ACCAGAGAATCT GGGAGCAGCCAG CCCAGCAGAGAG GAGATAGGGATA GCAAAAACAGCA AAGAAAAACCCA AAAATAAAAGCA GAGACTTTTTGGA GTTCTGGTGAACA TAGAAAGGGGAA GGGCAGAGCTCA GGCAGAGTCAGG CTCATATGCAAAT CCAGAGGGCGAG TTTCCTTGTCTGA	AGAGTTACCCTTCGA TCCAGCGATTGCACT TCTCGGTATATACCC GGAAGATCTGAAAG CAGTGACACGAACA GATATCTGCACGCCA ATGTTCATAGCAGCA TTATTCACAATTGCCA AGAGATGGAAACAA CCCAAATGTCCTTCA ACAGATGAGTGGAT AAATAAAATGTGGTA TATACACACGATGGA ATACTACGCGGCAGT AAGAAGGAACGATG TCGTGAAACATATGA CAACATGGATGAACC TTGAAGACATAATGC TGAGCGAAATAAGCC AGGCACAAAAAGAG AAATATTATATGCTA	DAEKAFDKIQHPFLIKTLQKVGIEGNFLN MIKSIYEKPTASIVLNGERLKAFLRSGT RQGCPLSPLLFNIVLEVLARAIRQDKEIK GIQIGKEEVKLSLFADDMILYLENPEKSTI QLELELINKFSKVAGYKINAHKSVMFLYA RNEQTEETLKKKIPFSIATKKIKYLGINLT KDVKDLYKENYITLLKEIEGDLKRWNIP CSWIGRLNVIKMSILPKLIYRFNAIPIKIPT TYFADLEKLVIKFIWKGKMPRIAKDTLKK KNEVGGLTLPDFEAYYKATVVKTAWY WHKDRHIDQWNRIENSEIDPQIYGRlif DKAPKVTELGHNSLFNKWGWESWISIS KRMKEDPYLTPYTKINSKWIKDLNIKDS TIKLEDNVGKHLQDLVLGGHFLDFTPK AQATKEKIDKWELLKLSFCTSKEFVKK VKRQPTQWEKIFGNHVSDKRLISCIYKQ RNPTTQ			
--	--	--	--	--	--	--	--	--	--

				AAAGCTGGTTTCT	CCACTAATGTGAACA			
				CTGCTCTCTGGAT	TTGAAAAATGTAAAA			
				CGTTCCTTAATGG	CAAATGGTTTATAAT			
				CCCTAATCTCCTT	GTAGAATGTAGGGG			
				GTCTCTTAGCATT	AACTAGCGATAGAG			
				TCAATAGCCCATT	AGCAATTAAGGAAG			
				AGATCTGCAAGG	GGGGAACGATAATC			
				AGGGCCCTTCTTT	CAATAAGAACAGATA			
				TTCTTCTTTTTTTT	AGCTATCGTGGGTAA			
				TTTTTAATCTTTT	ATTTAACGTTCTGGG			
				TCTCTTTTTCTAAA	AATGCCCAGGAATGA			
				ACAATTACTCTAA	CTATGGTCTGTTAATT			
				GAAGCCCAATAC	TCTGATGGGTATAGT			
				AGAAAGCCTCAA	AGGAACAAGTTCACA			
				AGACTTGCAATTT	GAAATGTTGCTATAT			
				GGGCAGGTCAAG	TAGGTTACTTTCTTG			
				ACAAGAGCAGAA	GGGTAGAGTAGGAA			
				CTAAGAGAGCTCT	CATGTTGGAAGTAAA			
				GAGACACAAGGC	GTAGTTATCTTAGGT			
				AATAAGTCCAGTG	TAGTTGTCTTTTCTT			
				GCTGAGAAAATTC	ACTCCCTTGTTATGA			

				ACTAAACACCACA	GTTTGTGAAATGT			
				ACTTCCCAAGAAA	TTTTTTATTGTATGTT			
				AGGGGGGCGTCC	TTTTTTAAAATTTTTT			
				CCTCACAGCCATC	TTATATAGTTGATTTA			
				TTCCTGGTGGGCA	AAAAAAAAAGTTAATT			
				GGGAACACTCCT	TAAAAAAAAAAAAAAAAA			
				GCCCGGCGCCAG	CAATGAAAAAAAAATAT			
				CCCCACAGCCCAG	GCAGAGCCCCCTTGA			
				AGCTGCCCCAGAC	GGAGCTGGTGGAGA			
				AACCCAGTGTGAC	ATGCAGGGGTATTG			
				GGGAAGTGTTTCC	GGCTTCCCCACCTCG			
				AACAACACACGC	ATGGTTGCTGATGTG			
				ACACACCACAAAA	CTCACAGACATAGGG			
				TTGGGCATGGAC	GA CTGGTGGTTTGAT			
				ATTAGCCTTTCCT	GGGTTGAGCCCTCTA			
				GCACCCTCAGCTG	CCACAGGACTTGCCC			
				ATTGTCCCAGAGT	TTGGGAAGACTGTTG			
				TGGGAAGGTGGA	CTGCAAAGGAGAGG			
				GCAGTGTGAAAA	CTAGGCCTCCCTATA			
				GGGGGAAATTAA	ATTGTGCCTAAGAGC			
				CACGCCCCATTCA	CTCCTCCCGAATGCC			

				GCCATCCTTTCAG CAGACTGGGAAC CTCCCTACACAGC CCTGCAGCCCAGA ACCTCCCTGGGG GGACGGCACTCA CCTGTGACATAGC ACAGTCATCCCTC AACAGAGGACCT GGGGGTGCACGG CCTGGAAGAGGG ACCCACTCGCAAG TCTCAGGGGCCAT ACGCCAATACCAA GGACTTGTGGGT CAGTGGCAGAGA CAAACGTGGCA GGACTGAACTGA AGGATTAGACTAT TGCAGCAGCTTTA AATCTCCAGGAAC	TCTTTGTTGCTCAGAT GTGGCCCTCTCTCTCT AGCTAAGCCAACTTG GCAGGTGAAATCACT GCCCTCCCCCTACG TGGGATCAGACACCC AGGGGAGTGAATCT CCCTGGCAACGTGGA ATATGACTCCCGGGG AGGAATGTAGACCTG GCATCGTGGGATGG AGAACATCTTCTTGA CCAAAAGGGGGATG TGAAAGGAAATGAA ATAAGCTTCAGTGGC AGAGAGATTCCAAAA GGAGCCGAGAGGTC ACTCTGGTGGGCACT CTTACGCACAATATA GACAACCCTTTTATAG GTTCTAATGAATTGG			
--	--	--	--	--	--	--	--	--

				ACCAGGGAGATT	AGTAGCTAGCAGTAA			
				TGATTGTTAGAGC	ATACCTGAAACTATC			
				CGCCCCCCTCCC	AAACTACAACCCAGA			
				TGACCGCCCAGAC	ACCCATGAATCTTGA			
				ACACGCCCCATAT	AGACGATTGTATAAA			
				ACAGGGCGGGCA	AATGTAGCTTATGAG			
				ACACCAACTACAC	GGGTGACAATGTGAT			
				ACGCAAGCTTGGT	TGGGAAAGCCATATG			
				ACACCAATTGGAC	GACCACACTCCCCTTT			
				CCCACAAGACTCA	GTCTAGTTTATGGAT			
				CACCCCCACTCAC	GGATGAGTAGAAAA			
				CACAAAGACAAA	ATGGGGGAAGGAAA			
				GTGGGGGAGAAC	CAAAAAACAAACAA			
				TGGCTTGAGGGG	CAAAGGCACCCAGTG			
				AATAGGTGGCTC	TTCTTTTTTACTTTAA			
				GTGGATGCCACCT	TTGCTCTTTTCACTT			
				GCTGGTTAGTTAG	TAATTATTATTCTTGT			
				AGAAAGTGTA	TATTTTTGTGTGTGTG			
				CACGAAGCTGTA	GTAATGAAGGTGTCA			
				GATCTGACAAATT	GGGATTGATTTTGGT			
				AGAGATAAGGGT	GATGAATGCACAACT			

				TTCAATCGGTCTA CAAATCCTAAAAG AACCTTATCAAGT TAAGCAAATGCCA AGAGGCCAAAAA CAACAGAAAATTT TAAAGCATATGAA AAAACCAGACGA TATGGATAACCCA AGCCCAAGCACCC AAATCAAAAGATC AGAGGAGACACA GTACTTGGAGCA ATTAATCAAAGAA CTAAAGATGAAC AATGAGACCATG GCACAGGATATA AAGGACATGAAG AAGAGCATGGCA CAGGATATAAAG GACATCAAGAAG	ATGTAATGGTACTGT GAACAATTGAATGTA CGCTTTGTTTTGTATG ACTGCATGGTATGTG AATATATCTCAATAA AAATGAATTAAAAAA AAAAA			
--	--	--	--	--	--	--	--	--

				ACCCTAGAAGAG CATAAAGAAGAA ATTGCAAGAGTA AATTAAAAAATA GATGATCTTATGG AAATTAAAGAAA CTGTTGACCAAAT TAAAAAGATTCTG GATACTCATAGTA CAAGACTAGAGG AAGCTGAACAAC GAATCAGTGACCT GGAAGATGACAG AATGGAAAATGA AAGAACAAAAGA AAGAATTGGGGA AAAAAATTGAAA AAATCGAAATGG ACCTCAGGGATAT GATAGATAATATA AAACGTCCAAATA					
--	--	--	--	--	--	--	--	--	--

				TAAGACTCATTGG TGTTCCAGAAGG GGAAGAGAAGGG TAAAGGTCTAGG AAGAGTATTCAAA GAAATTGTTGGG GAAAACCTCCCAA ATCTTCTAAACAC CATAAATACACAA ATCATAAATGCCC AGCGAACTCCAA ATAGAATAAATCC AAATAAACCCACT CCGAGACATATTC TGATCACACTGTC AAATACGGAAGA GAAGGAGCAAGT TCTGAAAGCAGC AAGAGAAAAGCA ATTCACCACATAC AAAGGAAACAGC					
--	--	--	--	---	--	--	--	--	--

				ATAAGACTAAGTA GTGACTACTCAGC AGCCACCATGGA GGCGAGAAGGCA GTGGCATGACAT ATTTAAAATTCTG AGTAGAGAAAAA TTGCCAACCAAGA ATTCTTTATCCAG CAAAGCTCTCCTT CAAATTTGAGGG AGAGCTTAAATTT TTCACAGACAAAC AAATGCTGAGAG AATTTGCTAACAA GAGACCTGCCCTA CTTGAGATACTAA AGGGAGCCCTAC AGACAGAGAAAC AAAGAAAGGAGA GAGAGATACATG					
--	--	--	--	---	--	--	--	--	--

				GAGAAAGGTTCA GTACTAAAGAGA TTCAGTATGGGTA CATTAAAGGACAT TAATAGAGAGAG GGAAAAAATATA TATCTGACAAACA TAAACCAAAGGA TAAG					
	R2NS- 1_CSi	GCA_00360 4175.1	Clonor chis sinensi s	TAGTCGGCGAGC TGA	TGAATGTACATTCTT GCCATTGAATCTCAC ACCAGACCCAGTATG ACGGACACTTTGTGC CTGATGTGCGAGTCT TGACTGTGCTAGCGC TTACCGCGCCTTGAA GAGCATTACAGATTG TTTGTCTTTCTTCGG TTGTTAGACTTTACAC GCATGTTTCCTTACCA AAATTCTTACGTACG	MTTSSWCRLGSLDASLGELRVQQLQTG LTTLFGFNVLVTFDENVHYKTSGASAPVP TSTQERLVGLTCEECKWCKSKAGLVA HHRVHDNDVGTNMVAQLACADCSR LFPTKIGLSQHRRHAHPTQHNADKLSR VKHSGARWSQQESQSLRLANNLYPSC ETQTALFARLEQYFPGRSAISIKTRLRVL NWQAQQDESSGGPDQTIGQIAAYSS EADDYSVWFKQTVDCAVSLLESHADSS LASVDLLAFARGLQSGIMTPEQVLSLLD LHASRTFPHTWKTVSRRRRQLAHRMP VNRKQIRRANYAAIQTLYHQRKDAAS	cd0 165 0	431	689

					TTGGGATTCATCCTA TCTGACTGGAACGT TGGTTGCATGACTTC GAATGAGACATTTCT TTCTTTTATCTCATAT TCTCCTGTAACCCTTT CGCATTGATCGCTTG CATTCACTTTTATGT CTGTGACCATGCTCT TCAAAAATAAACGAT A	AVLDGSWKDLYKGNCGLPPDAEQYWK QVLSAPKHVDSRPSRVVPSDWSLIEPI TGEEVGRTVRSMGNSSPGLDKLTPRML RRFNANVLAGYFNLLLLAGGCPPHLCR ARITLVPKVPNPTSPDQLRPISVSSILVRC FHKVLADRWSRRLQLPSLQFAFLHRDG CLEATSLHALLRHSSATASNLSLAFVDIS KAFDSVSHDTIVRSAEAFGAPSPLVRYIA QSYENAVAVFPSSEVHCHRGVRQGDPL SPLLFIAMDEVGLSMPQLGYQFHDT LVDGFAFADDWVVCAESQARLKEKLEA AAVELGRAGMKINARKTKAMVICGDR KHRATAVSVEPFCFAEELITPLGPTDVT YLGIPFTFKGKGVFNHRQHLLKLLDEV RAPLKPHQRMEITRNYLIPRLTYSVLG QVHRNTLKRLDNYIRQSIRGWLRLPKD TPISYIHAGKQHGGGLIPSLSATIPMQR RVRMEKLLSTQCRVLRNVVNDSAFGKV VRDLSLPIRVHGSCVNTKEELVAAWGD SLHNSVDGRGLRELVASPLSNRWLVFP ERVFPRIFIRGIQLRCNLLRTRVRSARHG			
--	--	--	--	--	--	--	--	--	--

						HGGQTILCRGNCGPESLVHILQSCWI THDARCARHNRVARELAKRLRRLGYTV FEELRAPTSTSFIPDLIAVRERRATVIDV SIVSDGRGVTWVNEKKQKYGADEFSLA IISALRAIGCDVDFLVHQPMIISYRGICFP QSAKAVIGLGLSKVTVSDLCLLAIVGSLR TYDTFMRGTWR			
	L2- 2_CCa	GCA_00401 1595.1	Cyprin us carpio	CAAAAGGTAAAG GGGGAGGTGTTG CTTCAATTTATAA CAATGTTTTTCAGG ATTTCTCAGAGGG CAGGCTTCAAGTA TAACTCGTTTGAA GTATTGGTGCTTA ATATAACATTATC CAGAGAAACAAA TGTTAATGATAAA TCCCCTGTG	TAACATACTAATACA CTTCTAATATCCAAAT CCGTAAAGGATTTT TAGGCTGCATTAATT AGGTAAACCGGAAC CGGGAACACTTCCCA TAACACCCGATGTAC TTGCTACATCGTTAG AAGAATGGCATCTAC GCTCATATTAGTCTG TTTCTCTCTTATTCCG AGGTCACCGTAGCCA CCAGATCCAGTCTGT ATCCAAGTCAGAGG	MFVLATVYRPPGHHTDFIKEFADFTSEL VLAADKVLIVGDFNIHVDNEKDALGSF IDILNSIGVRQHVSGPTHCRNHTLDLILS HGIDVDGVEIMQPSDDISDHVVLCKLH IAKTVNSTSCYKYGRITSTTKDCFVSNL PDVSQFLSISKTSEQLDDVTETMDSLFS ILNTVAPLRLRKVKENSLTPWCNEHTRT LKRAARKMERSWRKTKLEVFRIAWRES NLSYRKALKTAKSDYFSSLLEENKHNP LFNTVAKLTKNKASTSVDISQHHSSNDF MNYFTSKVDTIRDKIVTMQPSATVSHQ TVHYRSPEEQFHSFSTIGEEELYKLVKSS KPTTCMLDPIPSKLLKEVLPEVIDPLLTII NSSLLLGYVPKTFKLAVIKPLIKKPQLDPK	cd0 165 0	385	648

					GTCACTGCAGTCACC CGGATCCAGTATGTA TCCAGCCCAGATGGT GGATCAGCACCTAGA AAGGACCTCTACAGC CCTGAAAGACAGCG GAGACCAGGACAAC TAGAGCCCCAGAGAC AGATCCCCTGTAAAG ACCTTGTCTCAGACG ACCACCGGGACAAG ACCACAGGAAACAG ATGATTCTTCTGCAC AATCTGACTTTGCTG CAGCCTGGAATTGAA CTGCTGGTTTCGTCT GGTCAGAGGAGAAC TGGCCCCCGACTGA GCCTGGTTTCTCCCA AGGTTTTTTTCTCCAT TCTGTCACCGATGGA	ELVNYRPISNLPFLSKILEKVVSSQLYSFL EKNGICEDFQSGFRPYHSTETALLRVTN DLLSSDRGCISLLVLLDLSAAFDTIDHTI LLHRLEHFVGINGSALAWFKSYLYDRH QFVAVNEEVSYRSQVQYGVPPQGSVLG PLLFTLYMLPLGDIIRKHGVSFHCYADDT QLYISSRPGETYQFEKLTECIVDIKNWM TSNFLLNSEKTEVLIIGPKSSACNDLERC LSLDGCSVNSSSSVRNLGVLFDSNLSLES HVSSICKTAFHHLKNISKLRPMLSMSNA EMLIHAFMTSRLDYCNALLGGCSARLV NKLQLVQNAAARVLTRTRKYDHISPVLS TLHWLPIKHRIDFKILLITYKALNGLAPQY LNELLHYSPRPLRSQNSGNLIIPRISKS TAGGRSFSYLAPKLWNNLPNIVREADTL LQFKSRLKTHLFNLAYT			
--	--	--	--	--	--	--	--	--	--

					GTTTGGTTCCTTGCC GCTGTCGCCTCTGGC TTGCTTAGTTGGGGA CACTTCATTTACAGC GATATCGTTGACTTG ATTGCAAATTATTGC ACAGATACTATTTAA ACTGAACTGAGCTGC ATGATGACATCACTG AATTCAATGATGAAC TGCCTTAACTGTCAT TTTGCATTATTGACAC ACTGTTTTCTAATTA ATGTTGTTCACTTGC TTTGACGCAATCTTTT TTGTTTAAAGCGCTA TATAAATAAAGATGA CTTGACTTGACTTGA C			
--	--	--	--	--	---	--	--	--

	L2- 2_DRe	GCA_00869 2375.1	Danio rerio	GCAGGAGAAGCA CTTTAGCAGCATC TAGAACAGCAGC CTGTAAGTACATT TAAGATTTGTTTC AGTTGTTGTGTAT GGTTTGAGGACTT GTTTCCAGCTGTT TGTGTAAAGTTGT AGGACATTTAAAC TTGCTTTTCAGTTG TGTATAATACTAG AAGTTGTTTAGCC ACTGTTTCCTTGG TACTATAAGAGC TTGTGTAGCGAAC GCAGACGCGGTT CGCGTCGTCCGCC TCAGTTTCGGCCC TTGTTTTGACTCG GGAGGCGTGTCC	TAATCTGCAATTGCC TCTTTGAATATCACAC TAATTGTACACAAAA AAAAAAAAAAAAAA AAAAAAAAATAAAAAA ACTACTAACACTTCCC TTCTTAGACTTTACAG ACCTGAAACTTGCCT TTAGTACTTATTCATT GTTGCTCTTAGTTGT GTAAATTGCTTCCTT GTCCTCATTTGTAAG TCGCTTTGGATAAAA GCGTCTGCTAAATGA CTAAATGTAAATGTA AATGT	MCFLIPVVTNTRKTREVRRCRRNPHNLR SIHVSTISQLSLSVGLWNCQSAVNKADF ITSIATYSDYNLMALTETWLRPEDTATH ATLSANFSFSHTPRQTGRGGGTGLLISK EWKFTLIPSLPTISSFEFHAVTIIHPFYINV VVIYRPPGKLGHFLELDVLLSSFSNFAT PLLVLGDFNIYVDKPQAADFQTLASFD LKRAPTSATHKSGNQLDLIYTRHCFTDQ TIVTPLQISDHFLSLNIHITPEPPHTPTL VTFRRNLRLSPNRLSTIVSDSLPPSRKLT ALDSNSATNTLCSTLASCLDRLCPLASRP ARASPPAPWLSDALREHRSLRAAERI WRKTKNPAHLLTYQTLSSFSAEVTSK QTYYRLKINNATNPRLLFKTFSSLLYPPP PPASSTLTDDFATFFCTKTAKISAQFAA PTTNTQDTTPHTLTFSFSQLSESEVSKL VLSSHATTCPLDPIPSHLLQAISPAVIPTL THIINTSLDSGLFPTTFKQARVTPLLKKP NLDHTLLENYRPVSLLPFMAKILEKVVF NQVLDFLTQNNLMDNKSQSGFKKGHST ETALLSVVEDLRLAKADSKSSVLILLDLA	cd0 165 0	508	768
--	--------------	---------------------	----------------	---	--	---	-----------------	-----	-----

				AAACGCCAGGTA ACCACTATATAGT GAGCACGGTAGC ATTAGTCGGCAG GAGAAGCACTTTA GCAGCATCTAGA ACAGCAGCCTGTA AGTACATTTAAGA TTTGTTTCAGTTG TTGTGTATGGTTT GAGGACTTGTTTC CAGCTGTTTGTGT AAAGTTGTAGGA CATTTAAACTTGC TTTCAGTTGTGTA TAATACTAGAAGT TGTTTAGCCACTG TTTCCTTGGTTAC TATAAGAGCTTGT GTAGCGAACGCA GACGCGGTTTCGC		AFDTVNHQILLSTLES LGVAGTVIQWFR SYLSDRSFRVSWRGEVSNLQHLNTGVP QGSVLGPLLSIYTSSLGPVIQRHGF SYH CYADDTQLYLSFHPDDPSVPARISACLL DISHWMKDHHLQLNLAKTEMLVVSAN PTLHHNFSIQMDGATITASKMVKSLGV TIDDQLNFSDHISRTARSCRFALYNIRKI RPFLSEHAAQLLVQALVLSKLDYCNSLL AGLPANSIKPLQLLQNAARVVFNEPK RAHVTPLLVRHLWLPVAARIKFKTLMF AYKVTSGLAPSYLHSL LQIYVPSRNLRSV NERRLVVPSQRGKKSLSRTLNLPSW WNELPNCIRTAESLAIFKKRLKTQLFSLH FTS			
--	--	--	--	---	--	---	--	--	--

				GTCGTCCGCCTCA GTTTCGGCCCTTG TTTTGACTCGGGA GGCGTGTCCAAA CGCCAGGTAACC ACTATATAGTGAG CACGGTAGCATT GTCGGCAGGAGA AGCACTTTAGCAG CATCTAGAACAGC AGCCTGTAAGTAC ATTTAAGATTTGT TTCAGTTGTTGTG TATGGTTTGAGGA CTTGTTTCCAGCT GTTTGTGTAAAGT TGTAGGACATTTA AACTTGCTTTCAG TTGTGTATAATAC TAGAAGTTGTTTA GCCACTGTTTCCT					
--	--	--	--	---	--	--	--	--	--

				TGGTTACTATAAG AGCTTGTGTAGCG AACGCAGACGCG GTTGCGTCGTCC GCCTCAGTTTCGG CCCTTGTTTTGAC TCTCGAGGCGTGT CCAGCTGAATTCA ATCAGCTAGTGCT TTGGGGTTATATA AACAACTAGTTCA CCGCGGCAGCGG TCGCGGCAGCCTC GTGTGAAGACCG ACGAGGGTAAAG ACCATCGACTCTA CCTGCGCGACTCC ACCGAGCAAAGA CACCGACAAAGC ACTTGAGTACTTT ACTGTATTGTTTT					
--	--	--	--	--	--	--	--	--	--

				ACTTTACACTTATT TTTTGTTGTCAGT GCACTTTTATT					
	RTE- 1_DCa	GCF_00160 5985.2	Dendr obium catena tum	GTTGCCAGGTCCC AAGTCCTGGATG AGAAAAGTGGAG GGTTGCGTTAGGT AGTCGACAGCCA GCGTAAACTTGT CGAACAATAATG AC	TAGTGGGATAAAGG CTTGATGTTGTTGTT GTTG	MNPLRNCVNARASPGSDALYGLNTVT SVQGGLHRRPGAVLNRQGFPHQGRM WAKKL VHETKIRIGTWNIGSLTGKLMEL IDIMIHRKLNILCIQETKWVGEKAKEIDS LGFKLWYSGRSRTRNGVGVIIDKSLKDN VIKINRIGDRIIALKIVVDDKIVNIISAYAP QVGLDMSSKLNFWNDLEELIQNIPLEE RVFIGGDLNGHVGREVD DFKSIHGGHG FGTTNEG GKTILDFCLAYDFLIANTLFKK RNEHLITYKNGYSTTQIDYFLIRKLNRLD CKDCKVIPSENLC TQHRLLVLD FKIKGHK IKNKVKQFSRIRWGH LKNINKNIFESKIL DLKILSVHEDANIMWEKVASDIKNI AKD VLGVSKGKAPNSKNSWWWNP NVQEK IKNKKICYKNWFNFKNSDNWEIYKVAR KDSKKAICDAKT LAFSDLYNRLDTKEGE RYMYKLAKQRDRSSKDLVQVKIKDED QKILFLEDDIQQRWMRYFAKLFNEDCE	cd0 165 0	581	829

						NKIDYSDLSTLNENKDFCFYRRISKTEVR EALKKMKGGKAVGPDDIPIEVWKCLGD EGISWLTCLFNTILKTKKMPDEWRMSV LIPIFKNKGDAQDCANYRGIKLMCHTLK LWERVMEHRLRRDTNVSQNQFGFMP GRSTMEAIHLLRGLMEKYRENKEDLHM IFIDLEKAYDKVPREVLWRVLEKKGVNN AYIQIHKDMYAGAITCVQTQGGTLKYFPI SVGLHQGSALSPYLFALVMDVLTRHLQ EDVPWCMLFADDILLVDKTREGVEGKL ELWRSTLESKGFRLSRSTEYMECNFSS NRPSEGIVTLGDQVINKSTRFRYLGSIQ SDGEIDGDIISRIQVGWLKWRNASGLLC DRNVPLKLKGKFYKMVVRPAMLYGAE CWPLKEKHNTKLSVAEMRMLRWMSG FTLRDRIRNEHIREKVGVPVEDKIRESR LRWFGHIKRRPSDDPVRRVEVLDLTYV KKGRRPKKTWLENIRNDLSLLDLNENL TFNRTQWRKRIHVADPT			
--	--	--	--	--	--	---	--	--	--

	Daphne-2_LMi	GCA_000516895.1	Locusta migratoria	TTTCACCAGAATA TTGGGGGACTAA AGAATAAAATAG ATGAGCTTTTGGT TTGTTTAGATGAT ATAGAAACTGATA ATGTAATAGATAT ACTATGCCTGTCT GAGCATCACATTG TCACTGGTATGGA AAATGTGAATGTC AGTGGATATAAAT TAGCTGCAAATGT AAGTAGAGATAA TAAGGTAAGAGG AGGAGTTGCCAT ATATGTCAAAAGC TATCACAGTGTA AAAATTTAGAAAC AAAAAAGTTTTGT GTAGAGCAACAT	TAAAGGAATTTCTGA ATGGCAACTCCTTCT ACACTTTAGATGAAT TTTTAGATAAAAATT AGTAATTTAGGAAAA AAATTAATAAATAA AAAAAATAAATAA ATATTAATAATTGA GTGTCAAGTAATAAT ATGTATAGACACATT TCACTAACTGACAC GTTCCACATCATTAC GAAATGTCGTATTTA TGATCTATGGAACAA GTATTAATCTAATCTA ATCTAA	MCIMSEISTSNNKIKTIWNIITRETGQPR TQEDSITIKVNEKITNKKSEVENTFNMYF LNVVENIGSKCSLEEARQLMEQGVPRQ FDTMEIPPTSQSEIRKLINSLKSKSSHGID GISSRVLKACSQEICRILSHICNSSLEQGI FPDRLKYAIVKPLHKKGDTLDVNNYRPI SLLTALSKILEKVMYSRVASHINKNKVLT KCQFGFQKGFSTENAIYIFTNEILNALDN RKSPVGIFCDLSKAFDCVNHGILLDKLKY CGMSGKVHKWFKSYLTGRVQKVDISN SHNIQKQSPQGTGELSSMGCHRVRYW DLCCSYTKTAESSNWGTIKHGVPPQGSV LGPLLFLIYINDLPFYIHKDAKLVLFAADD SIAITPESQELTDEIVNEVFQRIIRWFVS NGLSLNFEKTQYIQFHTINGITPLINIDFNQ KSVAKVEYSKFLGLCIDEGLNWKKHIED LLKRLSSATYAIRVIANFGDVQLSKLAY YAYFHSLLSYGIIFWGNSSFSKKVFIAQKR VIRIAGAHPRSSCRHLFKDLGIFTIYIYIS LMKFVINNPTEFKCNSSVHSYNTRRKG	cd01650	153	443
--	--------------	-----------------	--------------------	--	---	---	---------	-----	-----

				ATTGAAGCATGTG CCTGTCAGCTTAA ACTCAATGATGAT TCTCTTATAATTGT AACAGTATATAG GTCCCCTTCAGGA AACTTTCATTATTT CTGGAAAACCTG GATGCCTTGTTGT GCTATCTGTCAGA TAGGGGGAAGCA AATTATTATTTGT GGTGATTTC AATG TTGATTCACTGAA AGAGTGTAATAG GAAGAATGACCT GGAAGTCTTACTT GGCTCTTTCAATT TGACATCTGTTAT TGATTTTCCTACTC GGGTAGTAAAGG	DLHYSRLNLALSQKGVNYSATKVFGHLP SSIKILTDSQLAFKKN			
--	--	--	--	---	--	--	--	--

				ACAGCAGTACATT GATAGATAATTTT TTATAGACCAAGA TAAGTTTAACAAC ATAAATGCTTGTC CTGTTGAGAATG GTCTTTCTGATCA TGGTGACACAGCTA GTAACAGTATATG ACATAGCTCCTTT CAGTAATTCGAAA CAGTCCTCCAGAG TTGTGCGCTCAAT TAATGATTTAACA ACTAGAAATTTC AAGAAAGTCTAC AGCAGTTAGACT GGGATGAGGTTT ATAAGGAACCTG ATGCTAATTAAA ATATAACTTATTT					
--	--	--	--	---	--	--	--	--	--

				CATGATACATTTG TAAGGGTATTTGA AAAGTGTTCCTC AAGAAATTAGTTA AATATAATTGTAA GAAACCATGCAA AAAACCATGGCTT ACTAAAGGAATA AAAATATCTTGTA ACCGAAAAAGGG AACTATATCTATC AGCAAGAAAGAG TAATAACCCAGAA ACAATCAAATATT ATAAAACTACTG TGCTACATTAAGA AAAGTTATTAAAA AATCCAGAAGC					
	RTE- 2_LMi	GCA_00051 6895.1	Locust a	CCACCCCCATGAA CCACAGACATTGC CGAGGTGGGGTG	TGAAGACCACAACA	MMASSWVKYSGGKIVPHSDLRAGTTQ EDVVIKKNKTGILRIGAWNVLTLNRVGK LENLKREMDRLKLDMSGISEVRWQDE	cd0 165 0	548	807

			migrat oria	GCTTGC GTGCCTC AACGATA CAGAT AGCCGT ACCGTA GGTGCA ACCACA GCGGAG GGGGTAT CTGTGG AGAGGC CAGACAA ACGTG TGGTTC CTGAAGA GGGGCA GCAGCC TTTTCA GTAGTTA CAGGGG CAACAG TCTGGAT GATTGA CTGAGCT GGCCTT GTAACAT CAGCTA ACATGG CTTTGCC ATGCAG TTA CTGC GAACGG CTGAAA GCAAGG GGAAC TACAGT CGTTACT TTCCCG AGGACA		QDFWSGEYRVINTKSVGGNAGVGLIM NKKIGMRVSYEQHSERIIVVKIDTKPTP TTVVQVYMP TSSADDEDIEEMYEEIKEII QKVKGDENLIVMGDWNSVVGKGREG NVVGEYGLGKRNERGDRLVEFCAEHNL MITNTWFKHHERRLYTWKRP GDTGRY QIDFIMVRNRFRNQVLNCKTYPGADID SDHNLLVMNCRLKLRLQKGRKVR RW NLDKLKEPEVVESFRGSIRQGLAETGES NTVEEEWVALRDRVVKAAEDHIGRRTR PNRNPWITQDILNLIDERRKYKNSASET GKREYRRLKNEIDRKCLAKQEWLEDK CRTVEECITRGKIDTAYRKIKEIFGDKKSS CMNIKNSDGKPVLSKEGKAERWKEYIE GLYKGNKLEGSILEMEDEVEEDEIGDTIL REEFDRALKDLSRNKAPGVDDIPSELLIS LGEPVMKKLFHLVCKMYETGEIPSDFKK NIVIPPKASADKCENYRTISLISHGCKIL TRIIYRRMEKLVESDLGEDQFGFR RNVG TREAILTLRLIEDRLRK GKPTFLAFVDLE KAFDNV DWNTLFQILKVAGVKYRERRV			
--	--	--	----------------	--	--	--	--	--	--

				TGCAGCTCTACTG TATGAATATATG		IFNLYKNQTAVIRVEGHERDAEVGKGV RQGCSLSPMLFNLYIEQAVKESKEKFGE GIKVQGEEIKTLRFADDIVILSETAKDLEE QLNGMDSVLKGRYKMNINKSKTKVME CSVRKSGDAEGIRLGNETIKVVDEF CYL GSKITSDGRSKEDIKCRMAMGRKAFLK KRNLTSNINLNVRKSFLKAFVWSVALY GSETWTIRSSDKKMIEAFEMWCYRRM MKIRWVDHVTNEEVLNRIGEKRSEFWHI LTKRRDRLVGHILRHQGIVNLALEGSV WGKNRRGRPRDEYSKQIQKDVGCRTY SQMKRLAQDRVAWRAASNQSAD			
	RTE- 1B_LM i	GCA_00051 6895.1	Locust a migrat oria	CAAAATAGTGAA GCAAGACCTACA AAAAATCCTTGGA TATCACAGAAATT TAATTAGAAAATA TAAAAATGCAGTA TGAAGGCCAAAAT GAGTACAGACGA CTAAAAATGAGAT	TGAAGACCACAACAA CAACA	MAGVLEAKCKSVEECIKGKSRYRLQENK RTFGTRKQLYEYQELDGKPVLSKEGKAE RWKEYIEQLYKGNELRDDVVGREEEVD QDEIGDTILRDEFDRALKDLGRNKAPGV DDIPSELLKSLGEPAMTRLFELVCRM YE TGEIPSDFKKNV IPIPKSGADRCENYR TISLISHGCKILTRIYYRMEKLVEADLGE DQFGFRNVGTREAILTRLILEDLRLRG KPTFLAFVDLEKAFDNVDWNTLFQILKS	cd0 165 0	148	420

				TGACAGGAAGTC AAAATGGCTAA		AGVKYRERRVIYNLYKSQTAVIRLEGHE REASVGKGVRRQGCSLSPMLFNLYMSRR PFSDVIQFVHEQAVKESKEKFGIGIKVQ GEEIKTLRFADDIVILSETAKDLEEQLNG MDSVLKRAYKMNINKSKTKIMECGLRK SGDAEGIRLGNETMKVVDEFCYLGSKIT CDGRSTEDIKCRMAMGRKAFLKKNLL TSNINLSVRKSFLKAFVWSVALYGSETW TISSADRKRIEAFEMWCYRRMLKIRWV DRITNEEVNLRIGEKRNFWHTLTKRRDR LVGHILRHRGIVNLALEGSVWGKNRRG RPRDAYSKQIQVDVGCRSYEEMKRLAQ DRVGWRAASNQSPD			
	Daphn e- 1_LMi	GCA_00051 6895.1	Locust a migrat oria	TAACTAAAACCTC TAGGACAGCAATT GATAATATTTTAA TAGACAAATCAAA GGAACAAAATCA TATCATAAAACCA ATAATAAATGGCC	TGATGCATGCACTCA TACTATATTAGTATTA TTATATTATCAGCAA AAAAAACCTAAAAAT TATGTATCTGTATTTT ATTATAAGTGTATGT ATCTGTATTTATATTA TTAGTATTATCTATTA	MQLLVLGIDSKQIMKTAKSTYRRRVNP PKIENFRKLLKDMNWKDVYSAHGMNE KYNVFINNVSIMFENCEPLKVTRIKQSII KPWITQGIKISCKTKRKMYSKSSCND DDLTKYKEYCKILKKVIQASKQMHYEKKI ATSGNQIKTIWDIVKEETGRTRKEQEQV ALRVDETLVTDGHSVANLNFNKFISVTD SMGLSGSVNNALEYLKLAFTNNFTYMN	cd0 165 0	288	549

				TATCAGATCATGA T	TTCAGCTTGTAATTT TGTTTCTTGACATGTT CCACATCCACGAGGA TCTCCTCAACATGGA TCTACGGAACGAATA AACTAATCTAATCTA ATCTA	MSLTSPQEITSIIKSLKTKHSSGYDEISTKL IKACSCSELSTILSYLCNQSIITGTFPDWLK YADIKPLFKKGDKEIPSNYRPISLLPAFSKI LEKVMYKQLLNHLTTNDILSRTQFGFLK GSDIEKAIYTYSENVLNSLDNKLQAAGIF CDLSKAFDCVNHNILLSKLEFYGVTGNA AKWFKSYLTNRKQRVSVQVTSELSQQS SSDWEEITCGVPQGSILGPLLFLVYINDL SSVTLPFAEFVLFADDTSIAIKSTSSIVIER SANDIFMNINKWFKANSLSLNFEKTHY MQFRTCKRFPDSICIKYEEEQIEEVDSIK FLGLQLDNKFSWEEHTTELQRLNKS VF SIRVLSDIGDIKITKLAYFAYFHSIMSYGII FWGNSSSQSKVFRVQKRVIRIICGVNSR TSCRSLFKELGILTASQYIYSLMKFVINN ISLFPTNSSVHTYNTRNKNDLHKDLKAL TLVQKGVHYSGTHIFNNLPSSIKNLATN KDQFKRNLKDFLVANSFYISSEFLNCTK			
	CR1- 2B_M Un	GCF_01227 5295.1	Melop sittacu s	C	TAGGTTACCCTGCTC TTGCAGGGGGGTTG GACTAGATGATCTTT	MQIPVAPEGVRTKEEFALVDEDWVRD QLCNLDIYKSMGPDGMHPRVRELAEV IARPLSIIFGKSWETGEVPEDWKMANV	cd0 165 0	77	327

			undulatus		TGAGGTCCCTTCCAA CCCTTGGGATTCTGT GATTCTGTGA	TPVYKKGKKEDPGNYRPVSLTSVPGKV MEQLILDSISRHIKDEGVIKNSQHGFM RKSCLTNLIIFYEEVTRWRDDGRAVDV VFLDFSKAFDTVSHSILIDKLRKCLDDQ VVRWIKNWLGRRQRVVVNGAESSW RSVTSGVPQGSVLGPVLFNIFINDLDEG TECTLSKFADDTKLGGVADTPEGCVAIQ RDLDRLESWAGRNLMKFNGKCRVLH LGKNNPMYQYRLGVDLLERSEGERDLG VLVEGRLRELGLFSLEKRRLRGDLISVYK YVKGGCQDDGARLFSVISSDRTRGNGC KLEHRRFHVNIKRNFFTVRVTEHWNRL PRGVVESPTLEIFKARLDKFLCDVL			
	RTE-1_MD	GCA_000002295.1	Mono delphis domestica	GGGTGTATGGGG TGCTCAGGGAGG GGTAGTATCTCTG GTATGGAGGGCT TGTCGTGCCCTCC TAGGGCAGCTCTC CAGCCTCTGACCC CCACCTGACACCC	TGAAACTGCACAAAG ACAATAGTCATTCTC GATCACCGAGAGACT ACCACT	MDSTAHNPQGRGLEKVSQTLPALQTP GQHTAAGGSSPLSGRNQRKNTKKLLLG AWNIRTLTDRENTPRPERRTALIGKELA RYNIDIAALSETRLPEEGSLSEPTTGYTFF WKGRASNEDRIHGVGLAIKTSLLKQLPD LPVGISERLMKIRLPLSKDRYATIISAYAP TLTSTEETIEQFYSDLSAVLHSVPTNDKLI LLGDFNARVGQDHERWKGVLGKHGV	cd01650	588	870

				AGCTCTCACTTGT GGCTCCCAGTAGC TGCTAGCATGTGG CAGCGGCCACAC CCCGGGCAACGG CTTCGACAGGCCG GCTAAACCTTGTG AGGGTAGCCATC GGGTCATCGACCC CTGGTGAACCAG GGCTTTGCTCACC CAGCATGTGAAG ACTGCTTCGGCTG AACAGACGGAAG AAACCAATAAGA AGGTTCAACGGCT GAGAGGGCGACG CAGCAAAGCACT GTGGAGTGCTTA GGGCGTGTTGGA GCACAAAGGACA		GKMNNNGLLLLSKCSEFELTITNTVFRM ANKYKTTWMHPRSKQWHLIDYIIVRRR DIQDVKITRAMRGAEWCWDHRLVRATL QMRIAPRHPKRAQTVRAFYNVSRLRDP SYLQTFQSCLDDKLSAKGPLTGSSTEKW NQFRDAVKETSKAVLGPKQRNHQDWF DENNTAIEDLLSKKNKAFMEWQNNPN SAPKKDRFKSLQATAQREIRKMQDRW WEKKAEEIQRFADMKNYKQFFSALKTV YGPLKPTTTPLLSSDGDTLIKDKKGISNR WKEHFSQLLNRPSSVDQSALDQIPQNR TIEQLDVPPSIEEVQKAIKQMSAGKAPG KDGIPTEVYKALNGKALQAFHIVLTSIW EEEDMPPELRDASIVALYKNKGSRAAC DNYRGISLLSTAGKILARVILNRLLSSVSE QNLPEQCGRPDSTIDMVFTVRQM QEKCLEQNLSLYIVFIDLTAFDTVNRDA LWVILSKLGCPAKFVKLIQLFHVDMTGE VLSGGETSDRFNISNGVKQGCVLAPVLF NLFFTQVLRHAVMDLDLGVYIKYRLDG SLFDLRRLTAKTKTTERLILEALFADDCAL			
--	--	--	--	---	--	---	--	--	--

				ACACGGCCATCCA ATGCAGCTGAGG AAGTCTCCAGATG TAACAATTTTTCG TGCCACTGGACCC AGGCTTCCAACGC CGAGAGAGTGGG ACTGTCTCTGTGC ATCGGCTTTTCCA CTTAAATCTCTTTC ACGCACAAGTATC TTTGTGCACACTC ATCTATCCTAACC CCGTCCACCCTCT TCAAGACCTGCG GCGATGGGGGAG TGGCGACGCAAC AGGTGGAGGTGA CCACTGGCAGTTG TAGTCACGATCCT GCACGTAGGCGG		MAHQENHLQTIVDRFSTATKLFGLTISL SKTEVLFPAPGRPTNQPCITIDGTQLS NVNTFKYLGSTIANDGSLDHEINARIQK ASQALGRLRCKVLQHRGVSTATKLKVY NAVVLSSLLYGCETWTLYRKHMKQLEQ FHQRSLRSIMRIRWQDRITNQEVLDR NSTSIEVMVLKTQLRWSGHVIRMDPQ RIPRQVFYGELSAGLRKQGRPKKRFKD QLKSNLKWAGITPKQLELAASDRSSWR THINHAATTFEDERRRLAAARERRHQ ATTAPPVTTGVPCPMCHKLCASAFGLQ SHMRVHRR			
--	--	--	--	---	--	--	--	--	--

				CCCACGGACCACT GGTCGCTCGGCCC TGTGGGCAGCAG GGACGTTCGGCA GCATCCTGGGCG ACTGAGCAGCCCT CTCTAG					
	RTE-1_NVi	GCF_009193385.2	Nasoni a vitripennis	ATGGCGGAGTAC TAGCATCGGGTA ACTTCACGTACCT GTATGGGAGTGG GGGATCTCTAGG CACCGGATTTCTC GTAAGCAAAAGC ATCATACATTCGG TTAAAAGTTTCAA ATCCGTCAATGAT AGGCTCTCGTACA TCATTATCGAAGG TGAATGGTATAG ATATGTATTTATT	AAAAAA	MTNRIAYGGTSKKT VKEAANKVLGKKK KPKSKPWFDEEC ELWFERRKKAKLDSL QNRSDRTVEEYSNVRKQTS AIYRNKKRE YQKNLIRRIETNSKENNPREMYRGINAI RKGFRSRAQLMKDENGDLVTNDNELLS LWKNYFDKLLNVHENSEELGDEIHTAEL HVEEPSYQEVEAAIKKLKNNKAAGNDSI PAELLKYGGVELTFKIYKLVCAIWKNETI PENWKESIIPIFKKGDKTDCNNYRGISL LATCYKVLNVIQARLTPFAEDIVGDYQ CGFRRNRSTSDQMFTIRQLLEKKWEFC ETIHQLFIDFKKAYDSIKRSKMYQILVLLG VPKKLVRLIQICLNGSTGKVRVGGNVSE PFMIRDGLKQGDGLSTVLFNLTLEYAVR	cd01650	228	486

				AATGTACACTGTC CTACGGAGGACA AAGAAGAAGAAG CTAAGGATCTCTA TTACGAAACTTTA GAGCAGGTAATC GACCAGTTCGCGT CTTACGACACAAG AATAGTATTAGGC GATTTCAATGCTA AAATAGGTAGGG AGGAAATGTTTA GGCCTACTATAGG GAAGGAAAGCCT GCACGAAGCCAG CAATGATAACGGT ATTAGGGTCATAA ATTTGCGGCGGGC AAAAGATCTTATA ATCAAAACTACGT GTTTTAAGCACAA	KMQVSQLGATLNGTTQILGYADDLDIL GDCRETVARNAEILIKAVEYTGLEVSESK TKYMIVDKLGICRGEEDLRVGNFTFEKV SEFRYLGTTINDRNEINVEINKRLHSGNA CFYAVSNLLKSRLLSKNVKIRIYRTIILPVV LYGCETWALTQADNRFRVFENKVLRLK IYGPKKDEETGEWRR LHNDLHNLyas PNINRIKSRRLGWAGHVARMGDDRT AARVMKGRPMVTRPLGRPRRRWEDN VKADLVEIGRVGVDRRGASWVGLTQD RAAWKACVDEAMNFRVPNAT			
--	--	--	--	--	--	--	--	--

				GGACATACACAA GGCAACGTGGAC ATCGCCGGACGG GGCCACACAGAA CCAAATTGATCAT TTCCTCATTGAAA AAAGACGTAAGG GCTTACAGAGGG GCAGATAGCGAC TCGGACCACTTCC TAGTAGTAGCCAA ATTAAGAGCTAG ACTAGTAGCGAAT CAAAATAGTAAG CGAGCAAACAAG GTAGAAAGCTTC GATATTGAGAAG CTACGAGATAGA ACAGAGCGAATT AGGTACCAGATA GAAATTAATAACA					
--	--	--	--	--	--	--	--	--	--

				GGTTTCAGGCACT TGAAGAAGCAAA CACATCGCCGGA GGGGA					
	RTE- 2_NS	GCF_00039 3655.1	Nicotia na sylvest ris	TACTGTCAGGGCT ACACCAGTGGTA GCGGTGACAGCC CTTTTGGTTTTTG GGTCGTGGTATTT TCTGTGTTTTCAG GGAAATTCTCGAA GTTGTTGGAGAC AAGTTGGGCGCA CGAGCACTAGCA AAGGAGAGCAAC CTGGACGAGCGT CAGCAAAGGGCG GTGGGAAGCGGT GCGTGAGCGTAC AACTACATCGAGT ACAGCAAATCAAC	TAGGGGAAGTTGTG AGATATTTCTACAGC GCACTAGAGTGAGA CTAGCCAGTTAGGAG TTAGTCTTAGGATGC TACTGGTCAGCTACT GATGCAGGGCTTTAT CTGTTGGATATTAGT ATACCTCCATCTTTT TCGTATTTCTATATT TCTTATATTGCTGTTA TTTTGTTTTTATGTAT TTTATGTTATGTTATG TATTTTATGTTATTTT ATTATGAGTCTATTG ATAGTACTAATATAT CGTCTCTTGTTGCTTT	MTIKLVVGECTLNVVSAYAPQVGLDEEI KRRFW EGLDDIVRSIPP SERLFIGGDFN GHIGSSAGGYTEVHGGFGFGERNGGG TSLLDFAKAFDLVIANSSFPKREEHLVTY QSSVAKTQIDYLLLRCDRRLCEDCKVIP GETLATQHRLLVMDIGIMIRRKRSVRG RPRIRWGALTKDKAQELEGRLSAMGA WRSSGDANTMWSTTADCIRKAAREVL GISSGRTGGHKGDWWWNAV VQGKV EAKKAAYLRLVGSTGEEEK RANSERYKV ARKEAKMAVTEAKTAAFARLYEELGNK GGEKKLFRLAKVRERTARDLDQVRCIKD DDDKVLMGDDQIKRRWQTYFHKLLNE EGDQDIVLGELRNADSPHEL SYCRDIEV DEVMEAMRKMRRGRATGPDEIPVEL WRCVGRAGLEWLTGLFN VIFKTNRMP EEWRWSTMVPLYKNKGDIQSCNNYRG	cd0 165 0	438	687

				ACAGGTTTGGTTC	CTTGAGCCGAGGGTC	IKLLSHTMKVWERVVEMRVRRTVSISD			
				TCGGGAGTTGGT	TCCTGGAAACAGCCT	NQFGFMPGRSTTEAIHLIRRMVEQYRD			
				ACCTCCCGTTCTA	CTCTGCCCCTCGGGG	KKKDLHMFIDLEKAYDRVPREVLWSC			
				CTGTTTCCCTTTTG	TAGGGGTAAGGTCT	LEVKGVPVAYIRVIKDMYDGAKTRVRT			
				GTGTTAGCTAAAT	GCGTACATATTACCC	VGGDSEHFPVVTGLHQGSASFSPFLFAL			
				TGATGTTACTTTA	TCCCCAGACCCCACT	VMDAITHHIQGEVPWCMLFADDIVLID			
				GTTACAATAGTT	TGTGGGATTATACTG	ETRGGVNERLEVWRHALESKGFRLSRT			
				CAATCATTCAACT	GGTTGTTGTTGTTGT	KTEYLECKFGAEPTEAGVEVRLDSQVIP			
				AGTGTACTTGTAG		KRGSEFKYLGSVIQGTGEIDEDVTHRIGV			
				TCACTTGTTTCCTT		GWMKWRLATGVLCDDKKVPPLLKGKFY			
				CTAGTTTGATTCTG		RAVVRPAMLYGTECWPVKNSHIQKMK			
				TTGTCCGTTGTGC		VAEMRMLRWMCGHTRMDKIRNEDIR			
				ACTAGCGAGTCTT		EKVGVPAMEDKMRERLRWFGHIQRR			
				CTTTAGATTTGTC		STDAPVRRRCERLAVVGTRRGRGRPKKY			
				GTAGGTGTAGTG		WGEVIRQDMARLRITEDMALDRELWR			
				GCTGCAGTCTGG		SSIKVVG			
				GATGATAGATTAA					
				GGGCATGTCCTCG					
				GGTGGGGCAGAG					
				TGTGGGGCCGGG					
				GTCAGGGTGTGG					

				GACGGGAGGCAG GGGAGGTAGAGG GGGCAAGGGAGC CTATAGGTTGAGA ATCGGGTCATGG AACATAGGTTCGC TAACGAGTAAGTC TATAGAGTTGGC GAAGATCCTTCAG AAGAGGAAGATT AATATAGCGTGTG TCCAGGAGACTA GGTGGGTTGGAT CGAGGGCGAAAA ACGCGGATGGGT ATAAGTTGTGGTA CTCTGGAGTCCTG AGGGGTAAGAAT GGAGTGGGTATC CTGGTAGATAGCC ATCTTAGAGAGTC					
--	--	--	--	---	--	--	--	--	--

				GGTGGTAGAGGT CAGGCGGGTGAA TGATAGACTA					
	RTE- 1_Oaf	GCA_00436 5145.1	Orycte ropus afer	ACTCGTCTGTCAG TTTGTCACTACTGT GGTGGCTTGCGT GTTGCTGTGAGG CTGGAGGCT	TAACAACAACAACAA	MPLVFQMPAGLHMVHRFQWSFQTKT DSEEGRGHPLLKELARENLMNSSETLSD TLLENFMNSPETLSDTVPEDEPLRLEGT QNMTGKELPHQSRTNLNDMDKVCLS GPPFADVARKLRRNSCKHPLIIGTWNV RSMNLGKLEVVKNEMECIKIDILGISELK WTGFGHFESDNHMIYYAGNDKLKRNG VAFIITKNVSRSVLKYNVSDRIISICLQG RPVNTTIIQIYAPTTKAKDEEIEDFYQLLQ SEIDQTCNQDVLITGDWNAKVGNKEE SPVVGKFGLGDRNNAGDHMIEFCKTN DLFIANTFFQQHKWWLYTWTSPDGVH RNQIDYICGKRRWKSSISSAKTRPGADC GTDHQLLICKFKLKLKTIKTSPREPKYDLE YIPPEFRDHLKNRFDALNTNDQKPEEL WDDIKNIIHEESKRSLKRQERKKRPKW MSEETLKLALWVRVAKANGRNDKIKEL NRRFQRAAREDKANYYNEMCKDLELE	cd0 165 0	628	879

						NQKGRTCSAFLKLKELRKKFKPQVAILK DSMGKILNDAGNIKRRWKEYTESLYQK ELVDVHPFQEVAYDQEPMLKEEVEAA LKELAKNKAPGIDGIPIEMFQQMDAPL EALTQLCQEIWKTATWPTDWKRSIFVP IPKKGDPTECGNYQTISLISHASKILLKIIQ KWLQQYIDRELPEMQAGFRRGCGTRD IIADVRWILAQSKYQKDVYLCFIDYAKA FDCVDHKNLWITLRRMGIPEHLIVLMR NLYIDQEAVIQTEQGDTAWFKIKKGVC QGCILSPYLFNLYAEQIIREAGLYEEHGI RIGGRLINNLRYYADDTLLAESEDDLKHL LMKIEDYSLQYGLHLNIKKTKILTTGPISSI TINGEKIEVVKDFILLGSTINTNGSSSQDI RRCIALGKSAAKDLFQVLKSKDVSLKTK VCLTQAMVFSIASYACESWTMKNKEDR RRVDAFELWCWRRMLKVPWTAKRTN KSVLEEVQPECSLEARMARLHLTYFGHV VRRDHSLEKDVMLGKVEGQRKRGRPS TRWIDTVAATMGSSIARIVRLAQDQAV FHSVHRVAMSQRRLDGT			
--	--	--	--	--	--	---	--	--	--

	RTE- 2_OL	GCA_00434 8135.1	Oryzia s latipes	ATCTAGCCACAGG GGTTGCAAGCCA GTTACTTCTGTGC CGGTCCCAAGCCC GGATAAATAGAG AGGGTTGCGTCA GGAAGGGCATCC GGCGTAAAAAAT TGCCAAAATAACC	TGATGGGAAAAGCC GAAAGAGAAAGAAG AA	MRIIHNTLDIPDRSRPGLTTTATDAVNL QGVGGNLTTVGRRKRGGRRVRGQREK GKGRNIGLRIGTLNVGTMTGKGRELAD MMERRKVDVLCVQETRWKGSKARSIG GGYKLFYHGVDKRNGVGVLKGEFVN SVLEVKRVS DRMMSLKLEIEGVMVNVV SGYAPQVGCELEVKERFWSELDEVIESI PRGERVVIGADFNHVGEGNRGDEEV MGRFGVKERNLEGQM VVDFAKRME MAVVNTYFQKREEHRVTYRSGGRSTQ VDYILCRRGHLREVNDCKVVVGESVAR QHRMVVCKMTLEVRKKKRGKTEKTK WWKLQNEETCEEFRQKL RQVLGGQDE LPDDWETTAEIIRETGRKVLGVSSGKRK DGKDTWWWNEEVQECVQRKRLAKRK WDVERTEEGRQEYKEAQR RVKREVAK AKQKAYDELYDRLDTKEGEKDLYRLAR QRDRDGKDVQQIRVIKDRDGKVLTTQE SVQKRWKEYFEELMNEENDRERREED VVVEQEVAEIGKDEV RKALKRMKSGK AVGPDDVPVEVWKCLGETAVEFLTSLF	cd0 165 0	572	818
--	--------------	---------------------	------------------------	--	--	---	-----------------	-----	-----

						NRILESEKMPEEWRRSVLVPIFKNKGDT QNCSNYRGIKLSHTMKLWERVVVEAR LRKKVEICEQQYGFMPRKSTTDAIFALR MLMEKYRDGQKELHCVFVDLEKAYDR VPREELWYCMRSSGVAEKYVRVVQDM YERSMTVVRCAVGQTEEFKVEVGLHQ GSALSPFLFAMLMDRLTDEVRRQESPWT MMFADDIVICESREQVEEQLERWRFA LERRGMKVSRSKTEYMCLNERDQGRS VRLQGAEVKKVQEFKYLGSTVQCDGEC GKEVKRRVQAGWSGWRKVSGLVLCDR RVPARLKGKVYKTVVRPALLYGLETVAV RQRQEAEMEVAEMKMLRFSLGVTRLD RIRNEYIRGTAHVACVSDKVREARLRW FGHVQRRDSGYIGRRMLEMELPGRRRA RGRPKNRYMDVLTEDMKLANVRVEDV HDRVRWKRMIRCGDP			
	RTE-1_OL	GCA_004348135.1	Oryzias latipes	TGTGAACCGCCAC CTTAACGTGGTGG GGGAGTTTGAGA GCTCGGGTGATCC	TAAGCGGAGGAAGA TGGATGGATGGATG GA	MKRNNEDPITSPGLASPGPHSGARPRV GARRRAPGGRGSSHGPRSGTARRSDV GPLSHGPTTTRRPQKGPVQCGLGGSR GRVPRRPKPRSWNLAFGTWNVVTSLGG	cd01650	613	868

				CAGGAGCTAAGC TGTCTGGGGCTTA AGCCCCTGGTAG GGTCTCCCATGGC AGACAGGTCCTG GGTGACGAGCCA GACAAAGAGCGG TTCAGAAACCCCT G		KEPELVREVERYRLDIVGLTSTHSLGSET QSLERGWTLYHSGVAQGERRRAGVGL LVSPQLSRQVLEFTPVDERVMSLRLRVG DRSLTVVSAYGPNSSSEYQAFLVSLEGV LESAPTGDSIVFLGDFNAHVGNDSGTW RGVIGRNAPPDLNPSGVLLLDFCARHSL SITNTMFEHKGVBHQYTWHQDTLGRRS MIDFVVVSGDLRPCVLDTRVKRGAELST DHYLVVSWIRWRGRRPERLGRPVRTVR VCWERLAEPSVREVFNSHIRENFKQVP REVGDISEWTFMSTIVSAAARSCGRK VSGACRGGNPRTRWWTPEVRDAVKLK KESYQAWLACGTPEAADRYRQSKRAA ARTVAEAKTRSWEEFGEAMEKDYRLAS KKFWQTIRRLRRGKQFSTGTVFSAGGE LLTSTGDIVGRWKEYFEDLLNPTDTPSF EEAESEDFGVGLSITRAEVTEVVNELLG GKAPGVDEVPRPEHLKSLDAVGVS WLTR LCSIAWRSGTVPLDWQTGVVVPLFKKG DRRVCSNHRGITLLSLPGKVYARVLERR VRPIVEPRIQEQQCGFRPGRGTVDQLY			
--	--	--	--	--	--	--	--	--	--

						TLSRVLEGLWEFAHPVHMC FVDLEKAF DRVPRGILWGV LWEYGVRGALLRAVR SLYDRSRSLVRIAGSKSDLFPVHVGLRQ GCPLSPVLFIVFMDRISRRSQGPEGIWF GDHRISLLFADDV VLLASSNRDLQRAL GRFAAECEAAGMRVSTAKSEAMVLDLDR RKVVCHLSVGGVSLPQVEEFKYLGVLF SEGRSEREIDRRIGAASVVMRSLYRSVV VKRELSRKAKLSIYRSIYVPVLTYGHELW VMTERTSR IQAAEMSFLRRVAGRSLR DRVRSSVTRGELGVEPLLLHIERSQLRW LGHLVRMPPGRLPGEVFRACPTGRRPR GRPRTWRDYVSRLAWERLGVPPEELE EVAGVRE VWASLLSLLPPRPGPG			
	L2-6_OL	GCA_004348135.1	Oryzias latipes	AATTCACACATCA GTCACAAAACCTA CGACATGCACACA TCAGATTTTACCC ACAACCATGAAG ATGGCTCTTTTAA ATGTCAGATCACT	TGATTTTAACTGCTA ATTTATTCTATTAGTT TTAGTATAGGCTATT TTATGTATTTATTTTA ATTTTATGCATCTTAT TTTCAATTTTATGTTT TTTTTTCTTTTATGT	MFFFSIRGGKKGGGTASIAHKTILSENIL LSKYCSFEHHAFLFGDPPVLTVTVYRPPT PPSCFIEEFSEFLSSLHIRYNRILITGDFNL HVDKNDDIYACEFLNLLNSMDFQQHV TQPTHNRGHTLDLVITYGLSTGVSSVVD LAISDHYCVFFNVTSEIQQEASVRTVRK RYLTPEVAASFIDIFKQTPAQILPAPCDFI	cd01650	421	681

				TTTAAATAAGTCT	TTTAAATTTTAGCATC	VDHFNNRLQSSIDAVAPFKTKLRTTPK			
				TTTATCATTAAATG	TTATGATTTTAGCCCC	MPWKTDDIKGLKQNCRRARERYRKTKL			
				ATCTGATTTAGAC	AGTGTTTCTTGAGGG	IIHKDILKEQLKIYNNNSVKRARTSYFSKIIS			
				CATGATTTAACCT	GATCTTTTCCTCCAG	ENKNNPKFLFKTIDILINKDFNKSSMPSS			
				GTTTGTTTTTAAC	GGATTGCCGGCTTGG	NAACEDFADHFMGKINAIRSNISSMQY			
				TGAAACCTGGCTT	TCAGCGGTTGTTGCT	TDLNRPPALFLPEETLGSFVLVDAEMLG			
				GGCTCTGCAGCTC	GCAGTTGTTGCCCTT	RVVSQVNTTTCLLDPIPTSLFKTFYGFEE			
				TACTGAGTTCTAC	GCTGGGGCGGCAGG	SELLNIVNCSLQTGVFPSACKTAVVRPLL			
				CTTTTTAA	GGTCTAATTTTACAG	KKSQLDAPKNLDNYRPVSNLPFLSKINIEK			
					CTTCTTGGCTGTGAA	SVATQLHEFLNAQKILEKYQSGFRKNHS			
					GGGGACCCTGGTGTT	TETALLKIVNDLRLNYDSQKLSVLVLLDL			
					GTCCCGGTCTGGGTG	SAAFDTVDHQILLDRLSLVGLSGNVLK			
					GGCACTGTGGCGAT	WFYSYLTDRHFYVSMDDTCSSGIHEISCG			
					GTTCCGGTGGCCGG	VPQGSILGPILFNLYMLPLGEVIRRHGV			
					GCTGGCTCCTGGTAT	DFHSYADDTQLYIAVSPDDLEPVNTLLN			
					GAAAGGATTCCCAGA	CILDIKSWMAENFLQLNQDKTEVLIIGS			
					TACTGTTTCCTCGCA	EDKREMILPHLHNFKPSQCVRNLGVLF			
					GCAGAGCTAAGCCAT	DSELNFIPIHKNVVKTFYHLKNIARVRP			
					ACTTGTTTATTTATTT	FLSLASTEVLMMHAFIFSRLDYCNALLSGL			
					ATTTATTTATTTATTT	PKKSISNLQLLQNSAARVLTRTRGREHIT			
					ATTTAAGTTACATAG	PVLKSLHWLPVRFRIIDFKILLMVYKCFY			

					TCATATTTTCATTGTAC GTGGGAGCCATGGG TCATGTGGTTTAATG GAGGGGAGTGGGAA GGGTGAAGGGTGGG TTGGGTTTTTATTGTA ATGTTGTGTGTTTACT GTTTTTAGTGTTAAA GCGCTTCGAGCTGCA CAAAGTGCATGAGA AGCGCTATTTTATAA ATTTGATTTGATTTGA TTTGA	GLGPSYLN DILLKYEPSRTL RSSGAGLLV VPKVRTKTYGEASFCHYGPRLWNLPE NLRAAETVEVFKKRLKTHLFNLAFS			
	L2- 2_OL	GCA_00434 8135.1	Oryzia s latipes		TGAAAGGTCCTATCT GCACTTTTCTTTCCTT TCCTTCCTTTATTTTT AAATCATTTTTCAATT TTTTTCTTTTCTTTTAA AGTAAATTTTACGTT GATTATTTCTATGAT GTATTGTGATTTTAAT	MSSFYVSFKIELKQTPPLLCIVMYKPPQ HSQSFIDDFTEMLSVVCTDFDGLVITGD FNVHVDNVNDRNAKELNAVLETFDLIQ HVSCPTH SRGHTLDLLITKGVNIYNVSV VDVALSDHFCVFFDL SVAPKPPPGPAV VQRRQINESTRAQFVEMINPENASGA NVDEMLNSVTSSILNVLDAIAPMKVKL RKHRQKAPWRNDDL VRAQKQQCRRA	cd0 165 0	387	648

					GCATTTTCTGTTTGG TGAAGCACCTTGAAT TACTTTGTGTACGAA TTGTGCTATACAAAT AAACTTGCCTTGCCTT GCCTTGCCTT	ERKWRKSKLQVHYGMYKGELYAYNQIL CRTRERYFSEIIGSCSSNSRVLFTTVNRLT NPPTQLPIELVCTPKCNEFAVFFNDKVQ GIKKAINSTTHITIERPPTHSKLTHFVPVT DQTVQETITRLSSSTCCLDVLPTNFLKSV LNSVLPSITQIVNMSLQSGIFPEALKTAV IKPLLKKSGLDPTVLNNYRPISNLPFLGK VLEKVVYQQLTDFLLNNSFDVFQSGFR PHHSTETALIKVTNDIRLNTDDGKVSVLI LLDLSAAFDTVDHGILLHRLQDWVGISG SALSWLKSYLED RKYFVEIGSCVSDYMA LTCGVPQGSILGPLLFNLYMLPLGQLIRS NNISYHNYADDTQIYVSLTTGEYGPVDS LSHCLQQISAWMQNNFLQLNSNKTEVI IFGPQKQREIVSSYLHSLKPSNQVRNL GVIMDSDLNFNSHIKSVTSAAFYHLKNI AKIKNIVSKPDLEILIHAFVSSRLDYCNGL LTGLSKRAVKQLQYIQNAAARVLTRTRK YDHISPVLRSLHWLPVPQRIDFKAALLV YKSLHGRAPKYISDMLVPYEPSRTLRTS GAGLLIPRVRTKQGESAFQYSAAKIWN			
--	--	--	--	--	--	---	--	--	--

						SLPEVVRQASSVFMFKSRLKTFLESRVY D			
	CR1- 1_PH	GCA_00158 7735.2	Parhyale hawaiiensis	GCGTGGCCTGCG CGTTATCAGGCCG CCCCATTGTCCG GCCACCGGACGC CTCTTTGTTGTAT GAGCTGGCTTTG GGTTTGGGGCTCT GGGTAGCCCCTG AGTTGGAGGAAA TGCCACTCTAAAG GCGGATGAGACC TCTTTTGGTCCAA CGGCATCCCTGAG AGGAGATGAACA GGCTGGTGCCAT AGCTGGGTACTCA GTCTTCCGCAGTC TGCTCCTCTCTCTC TTCCATCAGGTAG	TAGCAGCGTTTTGGG CTTCCCGGTGTTGCA AGATGCTTCATTTGG GTTGTATTACCTGTCC GTGGGAGCTCAAGC GTGAAGTTTAAATAA TAATAATAA	MLYTIFLVYGILCFFVYFCIYLYITVYLCF LFFLFLLLCGDVESNPGPGRARGCRLLY CNIRGLHANLAELDFVSRGVDVCCSET LVAGRRHDAELALAGFQSPFRRLCGSG PGFRGMVAVYVRSGFCAYRQSVHECAC HEIIVVRVCGRLNNYYLFSLYRSPATDDS LYDCLLTAMASIQSTDPKAAFVFGDV NAHHRDWLGSASPTDCHGVAALDFCT LSSCVQLVRGSTHIAGNCLDLVMTDVP DLMTVTVGSPIGSSDHSVLVSLDLNQ VVPVVDTRRVVFKSRANWVAITRAVR TLPWRQIIHSEDPVSELNDLTVSILERFV PKRTLVRSRDKPWFDDQCRLAFEAKQ AAYRAWRRSRDRTLWQTYVDRRAEAK RVYEEAQRRLRQRSRESLLSIDHPHRW WSELKGSVFGAEPSPPLVGPGGGIITD PLARAEALLSAHFDGKQSRDVIALPHGCH PEPRLTSLAFRSGAIKVLLEGLDPYGGVD PVGMFPLFYKQLADVLAPKLAVIFRRLLIR	cd0 165 0	536	795

				TACCACCTGATGC TCTTTTCCTTCTCT TTCTCTTCTTCTTC TTTTCGGCCAAAG CCTGTACTGGTGA TGGCAGGGTCTG GGGCGCGTAGAT AGCTGGCCGCTTT CGTAGCTATCGTT ACTGTCGATTTGT TTTTGCTTTATT		LGNFPRCWRTGNITPIKGPVSPYVAN YRPITLTPILSKVFEKLIAGKLGRFAEVTG LLPAGQFAYRIGLGCCDALLSVSHHLQS ALDCRSEARLVQLDFSAAFDRVNHRL LYKLESGVGVVSGVPQGSVLGPLLFV LYTSDMFSSLENTLINYADDSTLMAVIP APRLRDAVAQSLNRDLSRISAWCSAWS MKLNASKTKSMIISRRTLVPQHPQLEI DDTLLQESSSLEILGVVFDEKLTFFPHIRR LVSRASKIGLLRKVNSVFGDSQVARRC FYAFLLPVLEYCSPVWASAADTHLRLLD RLVSSASRLCADNDLVNLSHRRRVAELC MFYKVYNNERHWLYSSLPALKVFGRET RAACGAHSFTLEAVRCRTNQFCRCFVP WSAKVWNLLPASAFGRIGLQAFKSAV NGFLDL			
	RTE- 1_Psi	GCF_00023 0535.1	Pelodi scus sinensi s	TGGTACGCTAAG AGCACGGCAAGA CATGAAAGATGC CCTGGTCAACCAC	TAGATGTCAAGGTCC TCTTCGAATACGATG GACGAACAACAACAA C	METGGGYTFFWRGRSSDERRESGVGF AVKNHLVRKLASLPKGVNDRLMTLHLP LPRGKQATLISAYPTMTNPDEVKDKFY EDLDALLSSVKRTDRLILLGDFNARVGS	cd0 165 0	488	770

				TGC GCCCAGCCCA TCTCCAACCGTCT CGACTTTAGTCCT GCCATTGGAGTA AGATGGAATCTG GGAAGAGAGAGT GAGGCTGACTAT GCGCACCTCTCCC TCACTTTAATCCA ATTCACGCGCAAG TCTTGACATCTCG ACATCTCCTAAGT CCTGTGGCGACG GGCGAGCGACGA AGCAGCAGGTGT GGATACTACTGGG AGCTGTAGCCGC GGACCTGCACAC AGGCGGCTCAGG TCTAATGGTCGTC TTCTTGCTGGTGG		DHTAWDGVIGKNGIGKCNSNGLLLLKT CAAHDLIITNTIFRLPTRKKTSMHPRS KHWHLIDYVIVRRKDRQDVRVTKAMP SADCWTDHRLIISKLSFRIKPKRRPQGN KAVMKKINVSRLRNPNTATLLAEDLGN QLELEQLEENAEKDWERFRDIVHSSALK VLGPSHRKQQDWFDEENDEEIKAILAEK HRLHRAHQNDPSSTAKKNAFNNIRRIV QNKLRKMQDSWLSAKADEIQKFSDRN DAKRFYEALRSLYGPQPSGSSPLDSDG ATLITEKALILQRWAEHFQSVLNRPSSIN NEAIDRMPQADINHSLDVSPSETETLQ AISQLSSGKAPGSDAIPAEVYKGGGAVL VNKLTQLFQSFWNQGSIPQELKDASIV HLYKRKGNRQVCDNHRGISLLSIAGKIL ARVLLNRLINHLEQGLLPETQCGFRKER GTVDMIFAARQLQEKCEQNRELYTTF VDLTAFDTSREGLWRIMSKFGCPDR FILMVRQFHDGMTARVLDDGEASEAF PVTNGVKQGCVLAPTLFSIMFSAMLSD AFQNSSLGIRLRYRTDGKLFNLRLQAV			
--	--	--	--	--	--	---	--	--	--

				AAGCAACAGCTG GATTCGGCGTCTA CCTGAGTGA CTGA GCAGCCCTCTTTA GGATTGCACTGCT CACCTCCGTAGTA AGAGGAAGGGGC TAGAAAAGGTGC CCTAAACATTGTT TGCTTCACAGAAC CCTGGCCAGCTTA CCGCGGCTGGAG GGGATCCCATCTC ATGCGGT CGAAC AATAAAATTT CGA ACAAAATGCAAG GTGACACCACTCA CTATTGGCACATG GAATGTACGCAC GCTTCAGGACAAC CCTTCAGCAGACC	TKIKETVLRDLLFADDCALNAGSEQEM QASMDKFAAACDNFGLLINTKKTEVM HQPAPKAQHQPVTITVKGQKLQAVDQ FTYLGSTLSRSVTIDIEVNCRVAKASSAF GRLLTKVWDRRGISLATKLKVYRAVVL T TLMYASETWTVYRRHARQLNHFHTICL RRLLRIRWQDKVPDTEVLSRAGLPSVYT LLMKAQTRWAGHVARMPDHRIPKQLF YGELFQGKRSHGGQKKRYKDTLKASLK SLDIDTTSWETLAQDRSTWHTLIYQGC QTSEARRTTIAQEKRALRKARVAKAATV VPTHVCPTCGRTFRARIGLISHLRTHHE QSITC			
--	--	--	--	---	---	--	--	--

				GCCCAGAAAGAA GAACAGCCCTGG TCGCCAGAGAGC TCACAAGATTTAA CATCGATATTGCG GCACTGTGTGAA ACTCGTCTAGCAA ATGAAGGTCAGC TC					
	Poseidon-3_PM	GCF_010993605.1	Petro myzon marinus	ATATTGGTCGCGA AAGCCTACGTGTT AAACTGACTAACC TCTATGGGGAGG ATACTTTTCTTAA AACCAGAAAATT GGAGAAGCTTCG CATTAAGCGG	TAATACACATTACATT TTGATTAAAGAATT ACTAATCTACAAC ATTACTAATCATAAC CAATGACTGATTACT AATCGATTATATAAT TACATTACCCCCCCT CCGCCTCTAATCACT GCACCTGAATGAGTT AAGTGATGACGTCAC CCTGCTGCTCTATATA AAACCGGGAGCAAG	MKLISSLAFLKRCRDSKVIPNAVKITHHI KSSKANQILHKASELLIERIQHNRSELSY ISESLKQLHLDISNSVSMHYWNIIDNITQ QSADNVFKLTKTNQINKFSRLHNKQIPT ENQKCKNKTVLNLNTNTVLTGDMVSILS KGLNFAVTPRTIPVDEIIAEVESAVRLLP HNEAEMTRHEVAKIIENSKVPKVNISN EAKALRELRVNKDIQILAADKGRTTVV MSRADYDNKIRNLLLEESTYIKIKRDPTNT LERKTTAIIANSLIPKNEQYLIKPRASKPP RLYGLPKIHKIGIPLRPVSTIGSATYQLAK YITKQIQPYIGKTSSFIKNSEHLVNVIKDI	cd00304	416	499

					CGAGGGCTTTACCAT TCTGGCGAATTGCCC GATGATGCTGGAAGT ATTTACCAGCGAAAC GTCGTA CTCAA CTG TAAATGTGGATTAAC TCTTCAACACACCGG TGGGCTGAAGTAGT GAATTATTTGTCTTT GTCAACGTGACACTT TTTTTGCTGATTGGCC GTGTCGGGTGGTGG AGGGTTACGACACAT TTATAAATAAACGCG ATCTAACCCAAAAAC CTTTATTATGAATAAT ATTGGTCGCGAAAGC CTACGTGTTAAACT	KLTPDDILVSFDVISLFTKVPVDESLKLLD TLVKNTGLQKDIPKLAGHCLNNTYFLW RGDFFKQKEGAAMGSPLSPAIANLYME NFEIRALASSTLKPSFWFRYVDDTLVVW PHGKDTLNDFLNHLNAQNPAIQFTME MEVNNTLPFLDILLGHGLNGKITHSVYR KQHTDAYLNAESHHPAQKSSLIKTLV HRAKTVSDGESLVREIKHVKDALRQNG YNNAQINRALNAKSRSQSNTTESDQIV KTIKIPYISGVTDKLARVLQKQNITVRFG TVHКИQKLVSCTKDALPKLSTEGVYRINC SCGSCYIGQTGRNISVRLKEHQ RNIKLS QTQLSAVAEHACQQGHSIDFDNVAVV CKTNYWQRLIRESIEIERNKNNFNRS D SYRLSAAWKRIIK			
	Poseidon-	GCF_01099 3605.1	Petro myzon		TAATACACATTACATT TTGATT TAAAGCTTTC GTGACTGCAGTGATT	MVDEIIAEVESAVRLLPHNEAEMTRHE VAKIIENSKVPKVNISSENEAKALREL RVN KDIQILAADKGR TTVVMSRADYDNKIR	cd0 030 4	260	343

	3B_P M		marin us		CATGTTCTTCGCTCTT TCGGTGGAGTCGCGT TGGGGGGAAGCAAT AAAATAATAGAATAT ATATGTTTGTGTTTT GTTGTTTCCCTTCGGC GTGGCTTTGCCGAGG GAGCTAAGAATATGA ATGCACATTACAATT CCTTATAGGGGAAAC ACATACTAATTAACG GATTACTAACGTAGC ATATAATCACTAACA CACATTACAATAAT TACTAATCTACAAC ATTACTAATCATAAC CAATGACTGATTACT AATCGATTATATAAT TACATTACCCCCCCT CCGCCTCTAATCACT GCACCTGATATGAGT	NLLEESTYIKIKRDPTNTLERKTTAIIANSL IPKNEQYLIKPRASKPPRLYGLPKIHKIGI PLRPIVSTIGSATYQLAKYITKQIQPYIGK TSSFIKNSEHLVNVIKDIKLTDDILVSFD VISLFTKVPVDESLKLLDTLVKNTGLQKD IPKLAGHCLNNTYFLWRGDFKQKEGA AMGSPLSPAIANLYMENFEIRALASSTL KPSFWFRYVDDTLVVWPHGKDTLNDF LNHLNAQNPAIQFTMEMEVNNTLPFL DILLGHGLNGKITHSVYRKQTHDAYLN AESHHPAQKSSLIKTLVHRAKTVSDGE SLVREIKHVKDALRQNGYNNAQINRAL NAKSRSQSNTTESDQIVKTIKIPYISGVT DKLARVLQKQNITVRFGTVHGIQKLVSC TKDALPKLSTEGVYRINCSCGSCYIGQT GRNISVRLKEHQRNAIKLSQTQLSAVAEH ACQQGHSIDFDNVAVVCKTNYWQRL IRESIEIERNKNNFNRSYSRLSAWKRII K			
--	-----------	--	-------------	--	--	--	--	--	--

					TAAGTGATGACGTCA CCCTGCTGCTCTATAT AAAACCGGGAGCAA GCGAGGGCTTTACCA TTCTGGCGAATTGCC CGATGATGCTGGAA GTATTACCAGCGAA ACGTCGTACTIONAAC TGTAATGTTGTGGA TTAACTCTCAACACA CCGGTGGGCTGAAG TAGTGAATTATTTGT CTTTGTCAACGTGA CACTTTTTTGCTGAT TGGCCGTGTCGGGTG GTGGAGGGTTACGA CACATTATAAATAA CGCGATCTAACCCAA AAACCTTTATTATGA ATAATATTGGTCGCG			
--	--	--	--	--	--	--	--	--

					AAAGCCTACGTGTTA AA				
	Rex1- 2_PM	GCF_01099 3605.1	Petro myzon marin us	AGATGGCGCCGA TGATGGCAGCCTC GGCCGGTTGGTG CGCCCTGTTTTGT CCAGTTTTGTCGT TTTGTCTTGAGAG GGGCCTGCTGTTT AACACACTCGTGC TGGTGAGAGAGG CCTGCTGTTTCATC ACACCCGTGCTGG TGAGAGAGGCCT GCTGTTGAGGGG CCTGCTGTTTCATC ACACCCGTGCTGG TGAGAGAGGCCT GCTGTTGAGGAG CCCGTGTGAGG GGCCTGCTGTTCA	TGTTTTTAACATTTGA TCTACATAACGACAC CTTGCCGATATTAAC TATGTTACCCTATTTG TATTTTTTATTTTTAA ATTATATCTTAAAGCT ATTTTATATTCGATAT TTATTTAATTTTATAT ACTTCAGCCCCTTGG TGTTAGTTAGACTTTT AACCCCTTATTATAGT AAGACTTGTTTAAAA CTTATTTAAGATTTTA GATTCCTATATGTTG AATGTATGCACCATC CTGCCAAAGTAAATT CCGTGTGCGGTGTGAA CTTACATGGCGAATA	MALMMVASAGRCALFCLVLSVNSVFC QDFRSSLTRDELLNIRETTPVDLIPTFLLP AAELVGILVNAALGFAQATKRRPRGKR SGALVRLRRRGIRTVLPSFFLSNVRSLYN KMDELQLLMVKNKDFLSASALCFTETW LCDLIPDTSLQLAGFHLVRSRDRDIALSGK TKGGGICFYINSGWCVDVTVILQHCSPL LESIFITCKPFYSPREFASLILVGVYLP QVKEAQRMLADQILSVERANPDSLIV LGDFNKGNLSHELPKYIQFIKCP TLDHCYTTVSRAYHAVPRAALGHSDHA MVHLIPAYRQKLKLC EAVEDLRACLDCTDWD ELTEAVTSYISFCEDCCIPTRTRVNFNNN KPWFSAKLRRRLRLEKEEAFRSGDRDRFK ESKNRFSKAVREAKRLYSERLKHQFSTN DAASVWRGLRQITNYKPRAPHTTND LANDLNEFYCRFERQLDCPEPPLPTQEA PSPTNTTETTISFQEGEVNRLFQRQNP	cd0 165 0	574	826

				GCACACCCGTGCT GGTGAGAGAGGC CTGCTGTTCAACA CACCCGTGCTGGT GAGAGGGGCCTG CTGTTCAGCACAC CCGTGCTGGTGA GAGGGGCCTGCT GTTTCAGCACACCC GTGCTGGTGAGA GAGGCCTGCTGTT CAGCACACCCGTG TTGAGGGGCCTG CTGTTCAGCACAC CCGTGCTGGTGA GAGGGGCCTGCT GTTTCATCACACCC GTGTTGGTGAGA GGGGCCTGCTGTT CAGCACACCCGTG CTGGTGAGAGAG	AAACCCTTTCTGATTC TGATTCTGA	KAAGPDSVSPATLRHCADQLSPVFTNIF NTSLETCHVPACLSSTIIPVPPKKLRPTGL NDYRPVALTSVVMKSFERLVLAHLKAIT DPLLDPLQFAYRANRSVDDAVNMALH FTLQHLDAPASYARILFVDFSSAFNTIIPA LLQDKLFQLNVPYTTCKWITDFLSDRKQ CVKLGTVSGSRSISTGTPQGCVLSPLLF SLYTNSTSNHLSVKLLKFADDTTLIGLIS GGDESDYRWEADNLVTWCSQNNLEL NALKTVMVVD FRKNTAPLTPITLLDSP VKIVESFRFLGTILSRDLKWELNISSLTKK AQQRMYFLRQLKKFNLPKTMMVQFYT AIIESILTSSITVWYAAATAKDKSRLQR II HTAEKVIACNLPTLEDLHTTRSMRRARK IMADFSHPGHS LFQLPSGRRLRSIRTKT SRHKNSFFPSAAGLLNKAKGLH			
--	--	--	--	---	-------------------------------	--	--	--	--

				GCCTGCTGTTCAG CACACCCGTGTTG AGGGGCCTGCTG TTCAGCACACCCG TGCTGGTGAGAG GGGCCTGCTGTTG AGGGGCCTGCTG TTCAGCACACCCG TGCTGGTGAGAG AGGCCTGCTGTTC AGCACACCCGTGT TGGTGAGAGGGG CCTGCTGTTCAGC ACACCCGTGCTGG TGAGAGGGGCCT GCTGTTCAACACA CCCGTGCTGGTGA GAGGGGCCTGCT GTTCAGCACACCC GTGCTGGTGAGA GGGGCCTGCTGTT					
--	--	--	--	---	--	--	--	--	--

				CAACACACCCGTG CTGGTGAGAGGG GCCTGCTGTTCAA CACACCCGTGCTG GTGAGAGGGGCC TGCTGTTCAACAC ACCCGTGCTGGTG AGAGGGGCCTGC TGTTCATCACACC CGTGCTGGTGAG AGAGGCCTACTGT TCATCACACCCGT GCTGGTGAGAGG GGCGTGCTGTTCA TCACACCCGTGTT GGTGAGAGGGGC CTGCTGTTCA CTCCCATGCTGGT GAGAGAGGCCTA CTGTTATCACAC CCGTGTTGGTGA					
--	--	--	--	---	--	--	--	--	--

				GAGGGGCCTGCT GTTCACTACTACC CATGCTGGTGAG AGGGGCCTACTG TTCAACACACCAG TGCTGGTGAGAG AGTCCTGCTGTTC AGCACACCCGTGC TGGTGCGAAG					
	RTE1_ Pca	GCA_00402 6925.2	Proca ia capens is	GGTAAGCAGGGT AACACCATCAGGT AAGGAGGGAGTG AGCAGCCCCCACT TGTCTGTCAGTGG AGGCTTGCTTGTT GCTGTGATGCAG GAAGCTATGCCAC TGGTATTTAGGTA CCAGCAGGGTCA ACCATGGTAGAC AAGTTTCAGCTGA	TCTGCTGCAAAGGAC CTCTTCAAAGTATTCA AAAGTAAGGATGTCA CCTTGAAGACTAAAA TGCGCCTGACCCAAG CCATGGTATTTTCAAT CACATCATATGCATG TGAAAGCTGGACCAT GAACAAAGAAGACA GAAGAAGAATTGAT GCCTTTGAATTGTGG TGTTGGAGAAGAAT	MNSSQTLSDMVPESPPRLEGTQKTTG EELPPQSRVNLNDVDGVRPLGSSFADV ARLKMRRNSCKHPLIIGTWNVRSMNL GKLEIVKNEMECINIDILGISELKWTGLG HFESDNHIVYYAGNENLKRNGVAFIIKK NISRSVLKYNNAVSDRIISHLQGRPVTNTI IQIYAPTTKAKDEEVEDFYQLLQSEIDQT CNQDALIITGDWNAKVGNKEGSVVGK YGLGDRNNAGDHMIEFCKTNDFFIANT FFNQHKWRLYTWTSPDGIHRNQIDYIC GKRRWKSSISAVRTRPGADCSTDHQLL ICKFKLKLKIRTSPREPKYDLEYIPPEFKD	cd0 165 0	569	819

				GCTTCCAGACTAA	GTTGAATGTACCATG	HLKNRFDVLITNDRRPDELWNDIKDIIH			
				GACAGACTAGGA	GACCGCCAAAAGAAC	EESKRSLKRQERKKRPNWMSEETLKLAL			
				AGAAGGACCCAG	AAACAGATCTGTTCT	ECRAAKAKGRNGEVKDLNRRFQRAAR			
				CAGTCTACTTCTG	GGAAGAAGTACAAC	EDKRKYNDMCKELEMENQKGRTHSA			
				AAAGGAATTAGC	CAGAGTGCTCCTTGG	FLKLKELKKKFKPRVAIVKDSTGKLLNDS			
				CAGGGAAAACCT	AAGCAAAGATGGCA	ESIKRRWKEYTESLYQKELVDVQPFEEV			
				T	AGACTGCATCTTACA	AYDQEPMVLKEEVEAALKALAKNKAPG			
					TACTTTGGACATGTT	IDGISIEIFQQTDAALEVLTHLCQEIWRT			
					GTCAGGAAAGATCA	ATWPTDWKRSIFMPIPKKGDPTTECGNY			
					GTCCCTGGAGAAGG	RTISLISHASKILLKIIQKRLQQYIDKELPEI			
					ACATCATGCTTGGTA	QAGFRRGHGTRDIIADVRWILAESKEY			
					GAGTGGAGGGTCAG	QQDVYLCFIDYAKAFDCVDHTKLWLTL			
					CGAAAAAGAGGAAG	QRMGIPEHLIVLLRNLYMDQEAVVRTE			
					ACCCTCGATGAGATG	QGDTAWFKVRKGVHQGCILSPYLFNLY			
					GATTGACACAGTGGC	AEQIIREAGLYEEEQGIRIGGRLINNLCY			
					CGCAACAATGGGCTT	ADDTTLLAGNEEDLRHLLMKIKDHSLOQY			
					AAGCATAAAAACGAT	GLHLNIRKTKILTTPVSNIVINGERIEVV			
					TGTGGACATGGCGCA	KDFILLGSTINALGSSSQEIKRHIALGK			
					AGACCGGGCAGTGTT				
					TCGTTCTGTTGTTTCA				
					AGGGTCACTATGAGT				

					CGGAACCAACTCGAT GGCACCTAACAACAA CAA				
	Crack- 1_SSa	GCF_00023 3375.1	Salmo salar	TCCAAGATGGCGT AGCAGTCAGACG TCTTTGTCCTCCTC TTGTCGTGTCCCG TGTATATATTTTAA TATCTTTTTTCTTC GCATATCTTTTAA ATATTTTTCTTAAC CTCAACTTCAAAA CACTCTCCTGCAA CCCGCCTCACCCA ATGTGGAATAGA TATGCTATTTTATT TACCTCGAACTGG AATCCACAACAGA AGCTAGCCAGCTA ACTAGCTACTAGC TAGTAGTCAGCTA	TAGCCCATCTATAATT AGCCCAAACAACACTAC CTCATCCCCATACTGT TATTTATTTATTTTGC TCCTTTGCACCCCAGT ATCTCTACTTGACAT TCATCTTCTGCACATC TACCATTCCAGTGTTT ACTTGCTATTATTTTCG CCACCATGGCCTATT TATTGCCTTACCCCTA TCTTACTCATTTGCTC ACATTGTATATATAA CTTTTTCTACTGTATT ATTGACTGTATGTTT GTTTACTCCATGTGT AACTCTGTGTTGTTG TATGTGTCGAACTGC	MITRLSMPLLNVMPCPLLFWLVFIGLF HCRASSPAHYTLSNLSVPPPTHAMTSP GFNDVSRDNISLIITQCLGLPLYSHPTIP LSVHYTLNLFYRPQKPAPFTLCSGRARR PILIAFSRTLILLLLCSSGDVEVNP GPAVP SSTPTPQALSFD DFCNRKSLGFMHV NIR SLLPKFVLFTALAH SANPDVLAVSESWL RKTTKNSEISIPNYNIFRQDR TAKGGGV AIYCRDSLQSSVLLSRV PKQFELLLLKI H LSKNKSLTVAACYRPPSAPSCALDTICELI APHLSELVLLGDLNWDMLNTPAILQS KLDALNLTQIINEPTRYNP KAVNTGTLID IILTNLPSKYTS AVFNQDLS DHCLIA CIRN GSAVKRPPLITVKRSLKHFSEQAFLIDLA LVSWKDIDLIPSVEDAWLFFKSAFLTILN KHAPFKKFRTRNRYS PWFS PDLTALNQ HKNILWRSALASNSPRDMQLFREVRN QYTQAVRKAKASFFKQKFASCNTNSKK	cd0 165 0	647	905

				GCCACGGTTAGC GGTCATCAGCTAT CCTTTAGCTCGAA AAGCTATCGCCAG TTTGTACAGCGCG ACTCAGACCAGA GCATACCGGACCT ATTTTTCTCCATAT CCCCGGATTCTA CCGCAAGCTCTGG ACATTTACACCTG GATCATTGCAGCT AGCTAGCTGCATC CGAGTGACTACTA TTGGCTAACGTTT CGGTCCCGAAGC AAGCACCAATTAT CCTGGAGCTAGCC CATGCTAGGCCCA TCTCCCGGCTAGC TGAAGAGTTCCAT	TTTGCTTTATCTTGGC CAGGTCGCAGTTGTA AATGAGAACTTGTTT TCAACTTGCCTACCT GGTTAAATAAAGGTG AAATAAAAAATAAAT AAA	FWDTVKSMENKSTSSQLPTALRIGNSV TTDKSTIIENFNKHFSTAGHAFHLATPTP VNSTAPPTATRPSLPHFSFSQIQSADVL KELQNLDPYKSAGLDNLDPPFLKLSAEIV ATPITSLFNLSFVSSEIPKDWKAAAVIPLF KGGDTLDPNCYRPISILPCLSKVFESQVN KQITDHFESHRTFSAMQSGFRAGHGCT SATLKVLNIIITAIDKKQYCAAVFIDLAK AFDSVNHILIGRLNSLGFSDCLAWFT NYFSDRVQCVKSEGLLSGPLAVSMGVP QGSILGPTLFSVYINDVALAAGESLIHLY ADDTILYTSGPSLDTVLTNLQTSFNAILQ SFRGLQLLLNTSKTKMLFNRLPAPAR PSSITLDGSDLEYVDNYKYLGVWLDCK LSFQTHIKHLQSKVKSRIQFLRNKASFT HAAKHTLVKLTILPILDFGDVIYKIASNTL LNKLDVYHSAIRFVTKAPYTTHHCDLY ALVGWPSLHTRRQTHWLQVIYKTLGK VPPYLSSLVTIAAPTCSRSSRYISLVTPK ANSSFGRLSFQFSAANDWNEQKSLKL ETLISLTSFKHQLSEQLTDHCTCT			
--	--	--	--	--	--	--	--	--	--

				CAGCCACTCCTTG GGCTACAATACCT ATTTTGCCAACTG ACCTGGACCCGTT TACTGCCGTGCG GAGCCCCACCGA GCCTTCACGACTG GACTACCGACGTT ATCTGCCCCGAGG GAGTTATCCAACT GGCCCCTCCGTCG CGACGTAACTGA ACGCCCATCTGCT GGCCTGCTAACCG CGGATCGTTAGCT GTCTAGAGCATAT CGGACTGCTATCT GAATAGGTCTATC GGACAATTTCTTG GGCCACTATACCT ATTTTGCCAATTG					
--	--	--	--	--	--	--	--	--	--

				GACTGGTCCCCTC TACCACACGGAAC CCTACTAATCCAT CACGACTGGTCTA CCGACGGAAACG CACGAGGTGGCT AAAACAGACTTCT TCCGTCGCGACGT CCCTCTAAGGCCC TTCTGCCAGCTTG CTACCTATGGCCC GCTAGCTGTCTGA ATCGCCGTGTCCC CAGCCAGCCCAAC TACTCACTGGACC CCT					
	Rex1- 2_SSa	GCF_00023 3375.1	Salmo salar	CTTCCGGCGCCGA CAGAG	TAACATTGAGTGGCT GCTGCCAACATACTG ACTCAACTCCAGCCA CTTTAATAATGGAAA AATTGATGTAATCAA	MAASLRVPRKLC SILFFYVLFLLPQVIL GFITYSREELLD IRATSTHQHYDQEYDFP EADPLFGPPPRTMDRIPAGDPKQRRRR RRGRRSGLLVRLRRRAHRAPLPSILLAN VQSLDNKVDEIRARVAFQRDIRDCNVL	cd0 165 0	579	832

					TTTATCACTTAGCCAC	CFTETWLTRDTLSESVQPPGFFTHRAD			
					TTTATATTATTATAT	RNKHLSGKKKGGGVCLMINETWCDHN			
					TAGATAATGTTTACA	NIQELKSFCSPDLEFLTIKCRPHYLPREFS			
					TACCCTACATTACTCA	SIITVVYIPPQADTSTALKEHWTLCCKLE			
					TCTCATATGTATATAC	TTYPEAAFIVAGDFNKANLKTRLPKFYQ			
					TGTACGCTATACCAT	HIDCATRAGKTLDH CYSNFRDAYKALPR			
					CTACTGCATCTTGCC	PPFGKADHDSILLLPAYRQKLKQEAPAL			
					ATCTTGATGTAATTA	RSVQRWSDQSDSTLQDCFDHVDWDM			
					GATGTATCACTAGCC	FRIAANNIDEYADSVCEFINKCIGDVV			
					ACTTTAAACAATGCC	PTASIKTFPNQKPWIDGSIRAKLNARTT			
					ACTTTATATAATGTTT	AFNQGKVTGNMTEYKQCSYSLRKAIKQ			
					TCATACCCTACATTAC	AKRQYRDKVESQFNGSDTRGMWQGL			
					TCATCTCATATGTATA	QSITDYKRKTSPVADHDALLPDRLNNFF			
					TACTGTACTCTATACC	ARFEDNTVPLTRPATKTCGLSFTAANVS			
					ATCTACTGCATCTTGC	KTFKRVNPRKAAGPDGIPGRVLRACAD			
					CATCCTGATGTAATG	QLAGVFTDIFNQSLSQSAVPTCFKRATI			
					TATCACTAGCCACCTT	VPVPKKAKVTELDYRPVALTSVIMKCF			
					AAACAATGCCGCTTT	ERLVKD HITSTLPDTLDPLQFAYRPNRST			
					ATATGTTTTCATACCC	DDAIAITLHTALHLDKRNTYVRMLFID			
					TACATACTCATCTCAT	YSSAFNTIVPSKLVIKLETGLDPALCNW			
					ATGTATATACTGTAC	VLDFLTGRPQVVRVGNNISTPLILNTGS			

					TCCATACCATCTACTG CATCTTGCCTATGCC GTTCGGCCATCACTC ATTTATATATTTTTAT GTACATATTCGTATTC ATTCCTTTACACTTGT GTGTATAAGGTAGTT GTTGTGGAATTGTTA GGTTATATTACTTGTT AGATATTACTGCATG GTCGGAAGTAGAAG CACAAGCATTTCGCT ACACTTGCAATTAACA CCTGCTAACCATGTG TATGTGACAAATACA TTTGATTGAT	PQGCVLSPLLYSLFTHDCVAMHASNSII KFADDTTVVGLITNNDAYREEVRALG VWCQENNLTLNVNKTKEIMIVDFRKQQ REQPIYIDGTVVERVESFKFLGVHITDK LKWSTHTDSVVKKAQQRLFNLRRLKKF GLSPKLTNFYRCTIESILSGCITAWYGSC SAHNRKALQRVVRSAQRITGCKLPALQ DTYTTRCHRKAKKIIKDNNHPSHCLFTPL SSRRRGQYRCIKAGTERLKNSFYLKAIRL LNSHH			
	RTEX-1_SSa	GCF_000233375.1	Salmo salar	CTATCAGTTAGCA TGTGGAACATTCA GGGCCTAAACTCA TCAACCTTTGGAC TGAAGAGTTTAGC	TGACAGACCAATCAA CCTGCACATGTCCTCT ACTGTATGCTTATTGT TATTGTTGAATGTAT GGTTATTTTGACCCTT	MCSFAQYISPPQNPHITQRRSPPRIPIL LRGDISPPQNPHITQRRYPPLRIPILLRG DIPPESPYSEEIPTLEEETCHFQAQG NVLICGDTNARTGTLPDLTSTRGDSFIT GHTVSNCLHLPHRNNSDSTVNKNGRD	cd01650	499	800

				ACTGGAGTTCAAC AAAAATCTTAAAG ATGTTGACGTCAT CATTCTGCAGGAG ACATGGTGTAAG GCTGACATTGTCA CTCACTGTCCAC AGGCTACAGAGA GGTCATTGTGCCA TCACAGAAACACA GCTCTGTCAATAG AGGCAGAGACTC TGGAGGATTGAT CATTTGGTACAAA TCCGAACTACAAA ATCTAATTGATCC CCTCAAAATTGGC AAATATCACATTT GGTTAAACTGA AAAAAGAACTTGT	GGTTATTGTTGTTACT GTTGTCCCGTTGACA ATATTTGATTCTCATT TTTATTTATTTTATA TTGTAAATATCCAAA ATAAGCTTTGGCAAT ATGTACATTGTTACG TCATGCCAATAAAGC GAATTGAATTGAATT GAATTGAGAGA	LLQLCRSLGLYFVNGRLRGDSLGRFTYC SPLGHSTVDYMITDIDPFSLSSTVKPLT PLSDHSQITLFLKRTDMETTTSHSQPSKL YNIRNSYRWAQNSTEEYQKATWNQNI QTLLDNFLDTTFTHSKEGINLAVKNINYI FRQTAKEAQLKLIKNTKKITDDNWFD ADCKIIRKKLRTLSNQKHRDPNNGELRL HYCETLKLYKRTLRTKKAQYNSKQLTLIE ESINTNNFWQNWKKLKSKQEELAIQN GDIWTTTHFKTLYNTVQIDTNAEQRQIH EKLNGLEKAIKDNQNPPLDSPITDQELYK KLQALKFKKACGPDGILNEMLKLTSAKF QLAILKLFNLILSVGYFPDIWNQGLITPIF KNGDKFDPNNYRGICVNSNLGKVFCSEI NVLSKSQIGFIPKHRTTDHIYTLHTLIDK HVHQNNTKIYACFIDFQKAFDSIWHTG LFYKVIESGVGGKTYDIIKSMTGNTCSI KIGKKRTEFFNQGRGLRQGCNLSPALF NIYINELATILEKSSAPGVSLHNSEVKCLL FADDLCLLSPTAHGLQQSLDLLEQYCQT WALAVNPKNTCPREDPGPVRPVLQDL			
--	--	--	--	---	--	---	--	--	--

				ACTGACAGAAAA AG		CLLSPTAHGLQQSLDLLEQYCQTWALA VNPCKTKIMIFQRRSRSGIRPKFSIGTK YIEYCTHYNLGLKISSTGHLNEAVNELR EKARRAFYAIKKQIQIEPIKIWLKIEYVI EPIALYGSEVWGPLAKQDFIKWDKHPIE TLHAEFCKILLHVQRKTTNNACRAELGQ YPLIIKTQKRAIKFWKHLKYSDPLSYHYQ ALQCQELSKEKSPLIQLVLGLSSQTCSTN TLKPQDQNIQSIRINQITTQSKQNYIAY WETQAQTQSKMQCYLALNRQYTVAN YLTMVTDQNLRKTLTKYRLSEHSLAIEK GRHRKTWLPVEERLCNHCTTAEPETEL HFLTKCQKYKTIRECHFPKFETLIQGFKD LSDEDRLPVLLGEDAESGLAAHYIAAC HKMRDSV			
	Crack- 1B_SS a	GCF_00023 3375.1	Salmo salar	CCAAGATGGCGT AGCAGTCAGACG TCTTTGTCCTGTC GTGTCCCGTGTAT ATATATTTTACAT CTTTTTCTTCGCAT	TAGCCCATCTATAATT AGCCCATCTAACTAA CTACCTCATCCCCATA CTGTATTTATTTATTT ATTTTGCTCCTTTGCA CCCCAGTATCTCTACT	MITRLSMPLLNVMPCPLLFWLVFIGLF HCRASSPAHYTLNHLVPPPTHAMTSP GLNVS RDNISLIITQCLGLPLYSHPTIPL SVHYTLKLFYRPQKPAPFTLCGRHRRPI LIAFSRTLILLLLCSSGDVEVNP GPAVPSS TPTPQALS FVDFCNRKSLGFMHVNIRSL	cd0 165 0	646	904

				ATCTTTTTAAAAT	TGCACATTCATCTTCT	LPKFVLFTALAHSANPDVLAVSESWLRK			
				ATTTTCCTAAAAC	GCACATCTACCATTC	TTKNSEISIPNYNIFRQDRTAKGGGVAIY			
				TCAACTTCAAAAC	CAGTGTTTAATTGCT	CRDSLQSSVLLSRVSPKQFELLLLKIHLK			
				ACTCTCCTGCAAC	ATATTGTAATTACTTT	NKSLTVAACYRPPSAPSCALDTICELIAP			
				CCGCCTCACCCAA	GCCACCATGGCCTAT	HLSELVLLGDLNWDMLNTPAILQSKL			
				TGTGGCGTGGAT	TTATTGCCTTAACTTA	DALNLTQIINEPTRYNPKSVNTGTLIDIIL			
				CTTTTTTTTTACTA	CCTCATTTGCACACA	TNLPSKYTSAVFNQDLSDHCLACIRNG			
				AAGTATTTCTATT	CTGTATATAGACTTTT	SAVKRPPLITVKRSLKHFSEQAFLIDLAG			
				TACCTCAGATCTG	TGTTTTCTTTTGTCT	VSWKDIDLIPSVEDAWLFFKSAFLTILNK			
				GAATCCCACAACA	ACTGTATTATTGACT	HAPFKKFRTRNRYSPWFSPDLTALNQH			
				GAAGCTAGCCAG	GTATGTTTTGTTTACT	KNILWRSALASNSPRDMQLFREVRNQY			
				CTAACTAGCTACT	CCATGTGTAACCTCTG	TQAVRKAKASFFKQKFASCSTNSKKFW			
				AGCTATCAGTCAG	TGTTGTTGTATGTGT	DTVKSMENKSTSSQLPTALRIGNSVTTD			
				CTAACCACGGCTA	CGAACTGCTTTGCTTT	KSTIIENFNKHFSTAGHAFHLATPTPVNS			
				GCGGTCATCAGCT	ATCTTGCCAGGTCG	TAPPTATRPSLPHFSFSQIQSADVLKELQ			
				AACCTTTAGCTCG	CAGTTGTAAATGAGA	NLDPYKSAGLDNLDPFLLKLSAEIVATPI			
				GAAAGCTATCGCC	ACTTGTTCTCAACTTG	TSLFNLSFVSSEIPKDWKAAAVIPLFKGG			
				AGTTTGTACAACG	CCTACCTGGTTAAAT	DTLDPNCYRPISILPCLSKVFESQVKNQI			
				CGACTCAGACCA	AAAGGTGAAATAAA	TDHFESHRTFSAMQSGFRAGHGCTSAT			
				GAGCATACCGGA	AAAA	LKVLNDIITAIDKRQYCAAVFIDLAKAFD			
				CCTATTATTATTTT		SVNHHILIGRLNSLGFSDCLAWFTNYF			

				TATTCTTTTTCTCT CCATATCCCCGGA TTTCTACCGCAAG CTCTGGACATTTT CACCTGGATCTTC GCAGCTAGCTAG CTGCTATCCCGAG TGACTACTCCTGG CTAACGTTTCCGT CCCGGAGCAAGC ACCAATTAGCCTG AAGCTAGCCCCG CCAGGGCTCCTGA GCTACCACCGAA GCCCACTCCTGGG CTACAATATCCGG ACCCCTTTTACTG CCGGTACGGGGC ACGGAGCCCCGC CGATCCTCCACGA CTGGACTACCGAC	SDRVQCVKSEGLLSGPLAVSMGVPQGS ILGPTLFSVYINDVALAAGDSLHLYADD TILYTSGPSLDTVLNLQTSFNAILQSF GLQLLNSTSKTKCMLFNRSPPAPARPSI TTLDGSDLEYVDNYKYLGVWLDCKLSF QTHIKHLQSKVKSRIQFLFRNKASFTHA AKHTLVKLTILPILDFGDVIYKIASNTLLN KLDVYHSAIRFVTKAPYTTHHCDLYAL VGWPSLHTRRQTHWLQVIYKTLLGKVP PYLSSLVTIAAPTCSTRSSRYISLVTPKAN SSFGRSLFQFSAANDWNEQLQSKLKLETL ISLTSFKHQLSEQLTDHCTCT			
--	--	--	--	--	---	--	--	--

				GTTATCTGCCCGA GGATTCCAAGTGG CCCCTCCGTCGCG ACGTAACCTGAAC GCCCATTCTGCTA ACCGCGGCCTACT AGCTATCTAGAGC TACTTGGAACCCT ACTAATTCCACGA CTGGTCTATCGAC GTCACCGCACGA AGAGGCCAAAAC AGACTTACCCCA TCGCGACGTCCCC CAAAGGCTAACTT GCTTGCCCCGGTC TGTTAACTGCTAG CTTGCTTGCCCCG GTCTGCTAACTGC TAGCTTGCCTGCC CCGGTCTACTAAC					
--	--	--	--	--	--	--	--	--	--

				TGCTAGCCCCGGT CTGCCAACTGCTT GCTTGCTAACCCG GTCTGCTAACTGC TAGCTTGCCAGCC CCGGTCTGCTAAC TGCTAGCTTGTTT AGCCCCGGCCTAC TAACTGCTAGCTT GCTAGCATCGGCC TGCTAGCTGTCTG AATCGCCGTGTCC CCAGCCAGCCCAA CCACTCACTGGAC CCAT					
	Rex1-3_SSa	GCF_000233375.1	Salmo salar	AG	TAATCAAATGGCTAC CCAGACTATTTGCAC CCCCCCCCCCCCACG CTGCTGCTACTCTCTG TTATTATCTATGCATA	MAPEGMAAVLSALNQPCYFVCFFALFV TCFVHNVAATVSYDRKELLDIRTAITHLK LDKEFFFNESDGRDILQTPDQAQIPVIR WRRKRRFRGKRSGCLVRIRRRVANLPL PSVLLANVQSLENKWDELKARISYQRDI KNCNILCFTESWLND DIKNIQLAGYTLY	cd01650	581	833

					GTCACTTTAATAACTC TACCTACATGTACAT	RQDRTAASGKTRGGGLCIFVNNSWCTI SKEVSRFCSPEVEYLMISCRPHYLPREFS SVFFVAVYIPPQTDAGTKTALNELYTAIS KQENAHPEAALLVAGDFNAGKLKSVLP NFYQHVKCATRGKKTLDHLYSTHRDAY KALPRPPFGKSDHNSILLIPAYKQKLKQE APVTRSIKKWSDEADAKLQDCFASTDW NMFRDSSDGIEEYTTSVTGFINKCIDDV VPTVTVRTYPNQKPWITGNIRTELKGR AAAFKERDSNPEAYKKSRYALRRTIKQA KRQYRTKIESYYTGSDARRMWQGLQTI TDYKGKHSRELPSDTSLPDELNNFYARF EASNTETCMRASAVPDDCVITLSAADV SKTFKQVNIHKAAGPDGLPGRVLRACA DQLASVFTDIFNLSLSESVIPTCFKQTTIV PVPKNTKVTCLNDYRPVALTSVAMKCF ERLVMAHINTIIPETLDPLQFAYRPNRST DDAISIALHTALSHLDKRNTYVRMLFIDY SSAFNTIVPSKLITKLRTLGLNTSLCNWIL DFLTGRPQVVRVGNNTSATLILNTGAP QGCVLSPLLYSLFTHDCTARHDSNTIIKF			
--	--	--	--	--	-------------------------------------	--	--	--	--

						ADDTTVVGLITDNDLAYREEVRDLTV WCKDNNLSLNVIKTKEMIVDYRKRRT HAPILIDGAVVEQVESFKFLGVHITNKL WSKHTKTVVKRARQNLPLRLKRFG MGPQILKRFYSCTIESILTCITAWYGNC SASDRKALQRVVRTAQYITGAKLPAIQD LYTRRCQRKALKIVKDSSHPSHRLFSLLP HGKRYRSAKRSKRLLNSFYPAIRLLNS			
	L2- 2B_SS a	GCF_00023 3375.1	Salmo salar	GTTCCGGAAGCTCC AATTTGAGCAGCA GCAGCTGAAAAA TGAGCTCCTACAA ATATTGAAATTTA CATGGTTTTTAGA GTGAAGTACATC GGAAAAAAGCAA AAGAAAAGCTGCA AGACTGAATAGG AGAAAAACAAG CAACAAACAAG AAGATAGGCTTTT	TAGGATAAAGTAATC CTTCTAACCCCCCCCC CCCCCTTAAAAGATT TAGATGCACTATTGT AAAGTGGTTGTTCCA CTGGATATCATAAGG TGAATGCACCAATTT GTAAGTCGCTCTGGA TAAGAGCGTCTGCTA AATGACTTAAATGTA AATGTAAATGTAA	MCLSLPVLSSLRPHYKFTPRGRCHSNL VVPARTTHVEFQVSGSLWNCRSAANK AEFISAYATLQSLDFLALTETWITTDNTA TPTALSSSAHVFSHTPRASSQRGGGTGI LISPKWTFSLPLTHLSISSFEFHAVTVTS PFKLNILIIYRPPGSLGEFINELDALISSFP EDGSPLTVLGDFNLPTSTFDSFLSASFFP LLSSFDTLSPSPPTHKAGNTLDLIFTRCC SSTNLIATPLQISDHVLVSFLLSSNTSH SAPTRMVLRRPNLRSLSPATLSSSILSSLP SAQTFSNLSPDSASSTLLSSLSASFDFLCP LSSRPARSSPPAPWLDDSLRAHRTGLR AAERKWRKTRLPADLASFHSLLSTFSSS	cd0 165 0	577	838

				GAAAAATAACTAT TTTTCATAAAATT GTAGTTATCTTAG CTAGCTGAATTGT TTACCAGTTGCTA AGCAGTTGCTAG GGACTCTTTTGGA AGAAGCTAGCTA GCTAACGAAGAA CAACAAAGAATAT CTAGTTAACAGAA GAAAAGAAAGAG AAAATCAGGACA GAAGAAAAAGGA TATCAAAGTGAAA CAGTTCTCTAAAG ACACAGGAGACA ATAATACAAGAA ACAACACTTCTGT AGCTTGTCAACT		VSAAKATFYHSKFQASASNPRKLFATFS SLLNPPPPPPPPSLRMTSSTILKRRLTIS DPRLLSQTTPLVLLTLPYPVLEALCYLLLP PESSSPSPSSLSADDFVNHFEKKVDDIR SSFAKSNDTAGPAHTALPCALTSFSPLSP DEISRLVTAGRPTTCPLDPIPSSLLQTISG DLLPFLTSLINSSLTAGYVPSVFKRARVA PLLKKPTLDPSDVNNYRPVSLLSFLSKTL ERAVLGQLSCYLSQNDLLDPNQSGFKT GHSTETALLCVTEALRTAKANSLSSALILL DLSAAFDTVNHQILLSTLSELGISGAAHA WIASYLTGRSYQVAWRESVSAPCALTT GVPQGSVLGPLLFSLYTKSLGSVISSHGL SYHCYADDTQLIFSPPSDNQAAANRISA CLSDISVWMTDHLKLNLGKTELLFLPG KDCPFHDLAITVDNSLVSSSQSAKNLGV ILDNTLSFSTNIKAVTRSCRFMLYNIRRV RPCLTQEAAQVLIQALVISRLDYCNLLA GLPACAIKPLLIQNAAARLVFNLPKFS HVTPLLRSLHWLPVEARIRYKTMVIAYG AVKGTAPPYLQALIRPYTQTRALRSSTS			
--	--	--	--	---	--	---	--	--	--

						GLLAPLPLRKHSSRSAQSKLFAALAPQW WNKLPHDARTAESITTFRRHLKPHLFKE YLG			
	SR2	GCF_00023 7925.1	Schisto soma manso ni	CCCAAATGCCCTG GTACGGCCGAGA GTGGGGAGAGTC CACTCTCCCTCTC GAAATGCTCTCAC ATGGCCACGCGT ATATAGCCTCTGC CAGGGAAGTCCT ACTCACTGCCTTC TCGTGGCGGGGG TGTTGTTTACGAA AGTGAGAGGACG AAAAGCGAATGT CCGGCGCTTTAAC CGGGTTGGTGGA CACGGAGGGTCC ACCTAGGGGAGT TGGAAAACCTG	AGAGTGGTTCTAACT TTCTTAACTGAAAGA GTCTTCTGGTTGTAC ATTTCACTCCGCCTTA TCTTTTCATCCCTTCT CTTCTACTTTCATTA TTTTGTGTGGCGCAT ATGTATCTGGTGCCC TTTTGTACCAATATAT ATGTGTTTAAATAAA TAAATAAATAAATAA A	MLPPIMNTEAVFLKPRTPFKLASFNVRT LMQIRQQIGLAMSLEGLNVDVCCLSET RIQDSSEVLQIRSPSVASKSLFHVRLSGD PVASSSGLAGVGVALSARAEALIDWIP INSRLCAVRLESSIKVRRNRREKRCLFVIS AYAPDCSPDAIKDEFYHQLTVLLQKAR STDIVVLAGDLNAQVGRLGTEESRLGG RWGLVGRRTDNGDRLLQLCTDHNLF ASTNFRHSHRRCATWRPPSASQAWTQ IDHIAISYRWGCVQDCRSFWSTYLESD HALVCANLTLFSGRRIDHHQRIDVSKL AATSVASKYRAELASRLATIPPKSIDEH WLHLHDAMKMASKAACGFAKRPAYK HWVSSGSLQLMEARRSTPGDREFDHK RRLLRNEIGQSLRKDREAWWSEANE MEAAAASGNCRKLFQLIRATGSKKSGV SETIYEDDGMPITNISRRLGRWAEFFEG QFNWPAAPATSTSLSCPPWPVTTDPP	cd0 165 0	549	808

				ATTCCAAACCAAT GGTGCACATGGG CTCCAGTATCCTG ATGGAACGAATG GCGAATGAACCT GTCATTGGTCACC GGCTACCATGGG ACTGCATCTCCTC ACGATGCTCCACT GCCTTGTGGATCA GACCTTTAGGTCA AAGGCTCGGGGT GTGGCCCCCTAAG AAAACCACCTGCT TCGGTTTGGGCAC CCGGGCAGTATC ACAGCCCACACAC AAATAAATGAAAT GATGAATTCTATT GACGATACTATTC TCGCTCATACTAG		NEAEVRKELQLLKRYKSPGPDDLPPALF KDGGDFLTKELTVLFKVWEQESVPTS WNESIVVPIFKKGSRCSCNNYRGISLLPI ASKLLASIILRRLFKTRERLTREEQAGFRS GRGCIHIFTLRQMLEHRHTYRRPTIVV FLDIRAAFDSLDRTVLWDCLLKKGVPEK FINILKALYTNNTSGRVRAYNHLSPLFHSS SGVRQGCPISPFLNFADDILETALMD VSNGGVDMLPGERLLDLEYADDIVLLC DNAQGMQSALNQLAISVRRYGMCFAP SKCKVLLQDWQDSHPVLTLDGEQIEVV EKFVYLGSIYISAGGGVSDEIDARIMKAR AAYANLGHLWRLRDVSLAVKGRIYNAS VRAVLLYACETWPLRVEDVRRLSVFDH RCLRRIADIQWQHVSNAEVRHRVFG RRDDNSIGVTILKHRLRWLGHVLRMSS QRLPRRALFADSGTGWKKRRGGQCMT WCRGMKESCKGLASVGPSRLPGWGPR DGATQWLETLSDMAQNRSQWRSCCN LLLLST			
--	--	--	--	---	--	--	--	--	--

				AAGCAAGACAAT TTACATTATTATTA TTATTATTACCATT ATTATTATTATTAT TATTATTACTATTA TTTACTACGTCGC TAATTCACCCCCT TTATCCCCAATTT GTGTAACCTTACT TTTCAACTTATGC AGCAGCTCGCCCA TGGTGGAAAATCT GTCCTATCTAAAC GATCTCCAGTCAC TGTCTGGTTGAAA GTGA					
	LIN2_S M	GCA_00260 0895.1	Schmi dtea medite rranea	AACGTT	TACTACCACAAGACG GGTATTTTGGTAAAT CTTGAAAAATTTGCA TTTAATTAATGTTTGA GTAAAAATTGCCATT	MKIEIGEFAALFQRFQEQNLNAKSKQEEE FRNESVPGSRVPSEDMPVFQINERFHA LERSNELLNGKMDSIVNTLNYLISGLNG NFDNNNKLNCICKTNHVKEPLSIVKTNL VSELP SALLVQNKVPCLKLSEEDLIRKIPT	cd0 165 0	1533	1797

					TGACTTGTGTATGTG ATTTATTGTGTTAATT AATGAATTCCACCCT CTTCCAGGGAAAGGT TTAACTTTGCATAAA ATTAAACTCTCTGTCT ACACTTAGTAGCAAA TTAATCTAGAATGAA GAAGCCTGATGGTAT TTTTCCCAAATCTGCT GAAAACCTAGTTCTG AAGGATGGAAATAA AAGCAGATAAAGATT CAATATTCGATAAAG AAATGGACCTCAACA ATCAATGTCAATTAT CAATAATAAATCTGG GCAGATGTCAGAAG GCGATAATCAAATGC TGAATTCCAATGACC GAATCTACTTTGATG	KIVPSKKVSKKTIEHTSRVNFSLFVRVLD SNERQQFFKLFPNLFALLKSKESEKKSEV LNIELTKISKKDNPLNIAGFKNFFFRNNN ICVTIDFYNSSKVSKITKNLKIKTSLINSKI VKNYNTNKFIPASLSIKKFLFVLSDKLAET TNISNVTVREEQINKIITLLKIKNSLPDYL VKSQIKYKDGDFLMLDLLNSCEIEGELG QKITELSRSQIKELFNSFPEIQELFFIEKKA NSGTFFHRKKSIDENPEKLINLINKFKIEK TFPQQISFVGGKFVKIDKLND SQFPQVE SESDIMFNLGLFSLFGLSEYDNILNSLLC QPKSPEKLVDIIFHLNYLNLKGNFDTSTN GTYENYVQNYIVNCLLMHSILDLETCAF LNLDNIELWAPKLIELIPSSSHYSSSCKEK LLKLKQWKLSGEIDRIHNIICPETIIEINNF LGVIDQSYTTKSDQLKFLTSLGEKINNIEI DNDVDDLKAKRMLYFIIRDKKRFLNK DNFNYPFNCRKICTLKMSDFITKLIQS NELFEEREAVVDNNIDVSPNENLNILIVE PVNPTPDTVIVGQMDTFPTDIDETEPV TIDKPLKKITSKQAVKPKAQKSKREVPAS			
--	--	--	--	--	--	---	--	--	--

					TCTTAGATTTGTGATC AGCTTCCACTACAAA GGTGCTGCATCAAAA GCCAATGGAGATTCA ATGCCGGATGAAAAT GCCGATGAAAGAGA ATCAGAGCCGAGGA CAATTTAGAGAAAAT TCGATTGCAACTGCG TCAAGAGTTCCTCAA CAATAAGATGAAATA TCGAAACTCAATTGG ATCTTTTGCTTTGGA GAGTCTCAGAACGGC AAGCTCCATGTAGAG CTGTGTCCGATATGA AAAATTCAAAATTGT TATATTAATAAATTT TACTATAGTACTACT GAGTGATTTTGTAA GTCTACGACTTAGTC	VSISVPISEETDADSTIINYFFPERDPKIW FTDDDIDYYLESHIGNPNFAPVKCFIVEIL SSNCKDEIFPMPDNLFKAQIILCPLNINE SHWILFVYCKLTKISYFVDPILRKRHNLIN KSSALKITMALNNLFDLKA AVTFH PHD NLVYQDNNDYDCGPYICAYAMLISQEW QTFPNNFIETIRKNVHRFQTSKFQNP PK VSGGGITTLPFKRFQNVTRPPKVALPYK ASLQENSCVPSLIKEWFSNKNYITVLS PELTNLIAHNNQFIRENVTRQFTRVK YVIAVIPNNPHWCLYIFSPVFEHSYILNF QGRVLNERLVSIGESMTDYLNKFFISTA KVNFI VNHNQTHFSRSEDPHFFASAFI EFLERFFFGNILSFTKINETLKSFKLQITNE ICKNLNLINFSINDEIITKYFKNLNLPSNFI FLNSAMCTAIMDDCTRYLNDFLNYDNII KAQVIFAIVAPPKFREILFIIDYNTDEYYLL NPTTLLGNDNYLSTCKILRNKINEIRESP NRVLNLGECPHDIRGSGLVSKILFCSYM SNYANDLPLVGLNLRQIGNVVSSFLPIV SEIKKVKSLENNIINQNHKESEN LKERKN			
--	--	--	--	--	---	--	--	--	--

					GTAGACGTAGTTGTA	KIDQLLIKVFKMNVDEIVDVILDQISIKTV			
					GGCGTTGTTGGGCGC	PTHKPYLGNKERVESQINKSNLIKNFNS			
					TGATGGGAGCTGAT	DMKTTVNTIINDVDVDVRPKYENIITNF			
					GGGCGATGATGGGT	EHKIPLDRTWDKVFKFCNRSDKKLKLKY			
					GCGGATGGGCGATG	LSNSEVLFELKKADNTSPGVDGIKYRDL			
					ATGGGTGCTGATGG	VMLDPEGKLLTFLFNKIISKVIPDSWKS			
					GTGCGGATGGGTGC	FKTLLIPKPKGDKYDNVSSWRPIALLSV			
					GGATGGGTGCGGAT	IYKLFASCLANRLTFWINKNNLLHIGQK			
					CGGTGCGGATCGGT	GGSRHDGCVEHNSILSSALEHSKYSKNS			
					GCGGATGGGTGCGG	PLAIAWLDIKDAFGSVPHAYMWSLLRF			
					ATGGGTATGGATGTG	VGVDEEFVTILERMYSdTSTFYSCGPILT			
					TGTGGATGGGTATG	PNIAIRQGVKQGCPI SMILFALAINPVLE			
					GATAGGTACGGATG	AVSRSKCEPFMIGESAVKILAYADDIALV			
					AGTGTGGTTCTAGAA	SKSAYDLQRVTNIAVAIAASEIGFEYRPEK			
					TGTAGGTAGGTGTG	CGYIELPLIETESQILINDIKIKKLASSEFYQ			
					GAAGTTGGTAGGTGT	YLGVPVGEETDQSPYDILKKVVS DTRKI			
					GTGGAAGGTTCTAGA	ADSDLLGWQKLKAYKIFLHSRLVFPFRT			
					AGGTAGGTGTGTGG	REIKTSALAKSNANNNLTVNITSQLRNC			
					AAGGTTCTAGAAGGT	FRKMMSLPNHSEVCYFYNSTENGGAN			
					AGGTGTGTGGAATTT	CVDLLDEYHTQTITHFFRLFTSSCEFARK			
					GGTAGGTGTGTTGGT	VNLDSLNFVAGPRLGIINPTLQERFDWI			

					CCACAACTACGTGAC ATAATAACAGTTAGG TAATGTGTTGGTCCA CGTGACATAATAAAG TTAGGTAATAAGTGT GTCTCCGAGTAGTGA CTGACCAGTAAGTAC GTGACATAACAATGG TGACATAATAAGTGT GTCTCCGTGTAGTGA CTGACCAGTGGTCAA GTGATGTAATCATGA GGTCACTGACCAAAG GTCAATTACTAGGTA GCAACTGAGTGGAC AAAGGTCAATCTCCA GGTAGCAACAGCGC GTGCAGAGGTCAAAT GTTACGTAATAATTA CGTAAGCCAAGGGG GGTTGTATCTGACGG	NGKEPKPRHSGKKTRFQRARIAIAFFKK EHDINIMFIVTKGQPSLFITSEKRGTFILT PDLRKTTSKVLHMALCDSYLSKWEKSCT SNFIASAIKLSPKINKAIYRGDISEFAWHF IHRARTNTLSINAKPHNKGKRLCRLCH TEDETM SHIIQSCKIHLTLGSERHNDCL DLISSHLSKNSNLIVVDHVCSLVPEske RVDLMITDMVRKKIFLVDMKPCDTIN NFALVDLLNLQKYDSLKMQIQDAKPDF KVELDTCIIGSLGSFPPKTPELLLKIGVDR THLKGLLKDCALSNISHSARRWH			
--	--	--	--	--	--	--	--	--	--

					GCTCACATCCTACTA CT				
	Tx1- 2_SPur	GCF_00000 2235.5	Strong ylocen trotus purpur atus	CAATCCCTTGATA TGTGTATTGTGGA AAAGAACTGTTGT GAGAACTTTTGT TTTGCATTAACGA ATTCTCCAAACGT CGAAAATTCAGG AATGTAAAGTGA AGAAAAAAAATT CCCCGCCATATCG TACTTGGAGGTAT AGATTTTAAGACT CAAGAATGTAAG CCTTGTGACATTC CATGAGATGTAAT CTGGTTGGAAGT ACCATGCAATTTA AAGCACCAGTCA AAACTCATGAATG	TAATGGATGTGGGTA TTATTATTATTATTAG TATTATTAATTCTCTC TCGTGTACAGGTGGA CAAACCGGAACAGG TTGCATGCAGTGTGT GTGTCGGCGTGAAG GGGGTATGATGAGA TGGGTGGGCTAGGG GGGGGGGGGGGGT AAGAGTTTGTGTGGT GTGAGTTGAGGGAA CAGGTTGCATGCAGT GTGTGTTTGTGTTT TATGGTTGTTTATGT GTGTGAAAGGGGGT GTGGTGTGGGTGGG CTAGAGGGGAGGGG GGGGGGGTGTGAGA	MSVNLKLLSINVRGLRDRVKRKKIFQW CKMSANMIFLQETFSTEDIEYSWKGD WEGSMFFSHGSNHSRGVIVMFKPGITF VVEKTLIDEGGRYVIKKGKCQGEKLILGS VYFPTRNKENAQCDFLSDLNKIICSVFS QEYSLIIGGDFNVIMNHNLDYMGASNI SKSKYRDSLLDFLDKLDLIDVWRKRHPN KKEFTFRQKQPLVQSRLDYIFISSKVDKLI RKSEILPSITPDHAGVRIEFNNLVDSFVY GKSYWKFNSSLCNEKDFVDGVKKKISEI EETWRNQITSKIKFWDFMCKMKLREYIM KFSRERAKLRRVHIAKLEKEIKDLEYRLTI NIDCHKSIQDEIDMKKMELEKLFDYSRQ GLKVRSRTEWSEEGEKNSKFFEQLLKS KRKSVIKEIYNENKETVRDGGKILKIREF YENLYACKNGEITEDIENMFLNELPKLSK ENKEYCEGKITLAECYQAVKEMKWNKS PGNDGFSTEFYFTFWPILGKTIVEAFNE SFETKMLSNSQRQGVITLIEKEGKDPLY	cd0 165 0	520	774

				TAAAGAGAAAAT AGTACCCGCCAAT TTGTACTAATATA GGCATAGGTTTGC CTTATTGCATTCC ATGAGATGTCTAG CTTGTTACTTAAT CTTTAAAATGTGT CATAGAGGAAT GTGAACACTGTAC AACGCAAATTATA GAACGGGTTCTT TTAAAGATATTTA CAAATGGAGAGG GGGAGTGGTTTG GTGTGCATGTGTG TGTGTGTGTGTAA CTTACATACATAC ATGTACATCGCAT TTCTCTTTCCTTTT CCATTTGGTGAGC	GTGTTGTGATAGGAT GGTAATATGTAATTT TAAACCTTCCTCTACT GTTATTTTGATTATTC ACTGTTACATGCAAG ACTTTTTTTTCCCCTTT TTTTTCTCTATGTTC ATTGATTGATGTAAC TTATGAAATGTCTTG TATCAAAGATC	MKNYRPITLLNVDYKILSKVLATRVKEVL EDIIHYDQVGYYKDRNIGEAVRLIDDTF FQSLYQPVGFIAVDFEKAFDCISHKLLF KALEQFGFGKVLCSWVKTLTYTNVSSCV MNGGHSTGYFGINRGVRQGDPLSPYLF VIVIELLARYIRKDESIKGVGFKNKEIKQV LYADDITLFLKDMDSINKVKKVFKNFEKL SGLKINMDKTNFMWLGEESDRPGLPFF GHRVQHIKILGVYFARHSHKIKDDLNYKEI LSKIKRLIGWWKQRDLTMFGKIKLLKIY VFSKLNIVSSLMTVPQNI FEEIEKISFNFI WGGRDRIKRKILYQDYCAGGLRVTNFKI FIKTQRIMWIKRLLYGEKYSGWKVTFD HFFRSLGGRIIFLCNYDINKMDLKGIPSF YKDMLKAWQELDKCRHFGENKNNPIIF SNKCICLSNKTFDFVELFQKGILQASDIIV RGHLKPILHFLNLGFETKSLLRIQKIFS AIR NDWIQSDFMEVDTENFEIELMIFGR LQ KLCDIKSRRLYDYFIEDLQKEYTLHVTDG NCQFDLGMTTELKETFTRMKCAMLCNK HREFQYKQLQGAVYTKRELFKFGFESD			
--	--	--	--	--	--	---	--	--	--

				AAAGTAATCCAAC TAATCAATAGCCA TGAACACGTGAG AGGCTTCCTCAGC TATGATTGGTCCG CGAGAGCTGGAG GGCGAAAGTGTT TCTTTTGTTCTTT CGTCACCGGCTTC TCGTGGTCACCAA CGCCGGACATGCT GATTGGCTGAAT GGTCTTTGCGGCA GCAGGAGGGCGG GACGTGTTTTGAG ACAAGTCTCGTCT CCAGCTGGTCCTA GACATTTCA GTTG ATCGAGATAAAT GGCATCTGAAATA TACAACATAATTG	HFCSFCRLEVETYQHFLDCLKVKRLWQ EVIQKLHLIEIDDDWKPIFMGISGNTAR IFVINCIIFLVKYIIFISRKEGTLPSVQKTVH MLKNYRDEEYKIAFKKGKLG MHLQKW EQIDAFLL			
--	--	--	--	--	--	--	--	--

				TCGATTTGTTTCT GCTCTCCCCCCCC CCCCCTTTAATAG TAAATATCCATGT AATCTTACAATCT TAGCAGGAGGAT TAACGTTAGTGCC TTGTGAAGCAAG CATAATTAGATCT CGTCACCGGCTGC TTGTGGTCACTGT CTATGACGGCAG GTGGGCGGGACG TTGTTCTTTTATTT AGAAAAGTCTCGT CCCCAGCCAGTTT TAGACACCAGCG AGGTCTTTAAATC AACTATAAATTTT ATATCTAGACTCT GTTTTGAGAGAA					
--	--	--	--	--	--	--	--	--	--

				ATTAAAAGGGGA AAATAAATAATGC CCTTTTACGGTTA CCTTAAGCCGGTC GTTAACAGTGAA AAAAAGAAAGAA AAAAAGAAGAGG AATAGTTCAAAG TTAATTAATTCGT TTTCTTAAATAA AAATAAAATAAA AGATGGGAGGGG AAGAACTCCTTA TATACTGTAAAGT TTATAATGGTATC AAAAGCCATTTGG TGGGGTACAAGC TGTTGAAGCTTAA AGTATGCACGGG GTTTATTTATTAC GTTTCGTTTATTAC					
--	--	--	--	---	--	--	--	--	--

				GTTTTTTTTCTTTT AAAATATTATATA TAGCCGTTTGAGC CTGGGGCTAAAG TATCATTAGTTAC ACATTATCTGAAT CCGTTAATTATGA GGTCAAGTTATAT CGTCTGGAGTTAT GATAAATTGCTCT GTTTTATTTTCGG TGTTGGTTTTTTTT TTTTTTTTTTTTTG TCACAGTAAGAA GACTGTATTAATA CTTTGTGTAAAAG GCTGCCACGCCAT GATTTAGAATAAG TGATGGGAGGGG TTTCGAACATGCC AGCATATATTATG					
--	--	--	--	---	--	--	--	--	--

				CTTATGTGTATGA ATGTGGTTGATAG TGTTGAAGGGTG GGGTGATGGTAT GTGTGGTATGTTT TCCGGTTCCGATG TCTGCCTGTACTT GTTTAATTTGCAA TCTAACTTGAAAG CGTTTGTTTTCT AGAGTTGGAAAT GCATCAAGCTTCC CATCTACTGTGAA ACTGTCACCACTA TGTACATTTCTAA AGGTTCTTTGTTA CAGGTATAGATTC ATAGTTCTTG					
	Rex1- 2_OL	GCA_00434 8135.1	Oryzia s latipes		TGA CTGGAGTCTGGA AAAACAAACTGCACC TTTGCTCATCCTAATT	MQV NKALRRINTRKATGPDNIPGRALR ACANELTDILTSIFNLSLSQRVVPSCFKTT TIVPLPKSPPTCLNDYRPVALTPHIMKC	cd0 165 0	55	309

					GTTGCTACTGTCATT GTCACTTTTGCACATT TGCACTTTTAGAACT GCCACTCGACATCTT GTGTTTTTGTGTGTAT GTT	FERVMLTHIQSSIPDTLDPLQYAYRPNR STSDAIAAALHTSLSHLENKDSYIRMLFI DYSSAFNTVVPKHLTHKLFSLGIHPIICD WLLDFLTGRPQSVRIGNRTSASIITNTG TPQGCVLSPILYTLFTHDCVASHKDNIL KFADDTAVIGRITGGDEAAYRREVASLV TWCEDNNLTLNTDKTKEMIVDMRKER RNPHQPLLIRGLEVERVSSFRYLGVHLS QDLTWLNTSQLVKKAHQRLYFLRRRLR RFGMSPKILSNFYSCMVESILTSCITVWY GSSTAADRKRLQRVVKTASKITRTPLPSL QSIYKRRVCRRAASIIKDPTHQRGLFTL LPSGRRYRSVKCRTRTLKNSFFPSAIRLL NS			
	RTE- 1_DaC a	GCF_00162 5215.1	Daucu s carota	CATGAGGTGATA GTCATGAATCCAA ATATAATAGATAA ATAAGATATTTTT TTAAGGTGCAGTT CTGCTCCCTAACT AATAGTAGAAAT	TGAGGGTTATGGTGT ATGTTTAAAGGGAG GGTTCAGCTCAGTCA TTACCAGATGCATTA GCGCCAATCAAAAAT GTTTGGGTAAGTCCC ATGTGAAGAAGGCC	MANTIESMTFWGCLDDLVRISITLDQFL FVGDFNGHIGARADGYQGVHGGFGY GVRNDNGSTLLEFATAHDLVIVNSCFRK RDDNLITFRSGGHATQIDYLLIRSDFRF CSYCKVFPVEACASQHRLLVMDLSLDG PPMEGLRVVTPRILWRNLHGSKVAEFK DMLEGFFRLEDDVDQMWIRMADLIRG	cd0 165 0	413	662

				AAAGTAATGGTTA	ATGGGCTAAGGTATT	VARRTLGVASGNVKEQRESWWND			
				TAAGGGTTCTCTA	ATATTCTTATTATAGC	DVQAKVQFKKGCLEFMSCEGPDRHI			
				AAACATGACCCAA	ACACACTATTTTATTT	KRELFLAKRMAKQAVAEAKTKAYQD			
				GTTAATAAATGTC	GTTGTGCACTAAGCC	MYRRLGTKEGVNEIFKLAKARNKRSQDI			
				TCTAACCAAAAAA	GGGGGTCTTCACGG	GAVRYIKDEGDRVLLYDGDIIARWGRYF			
				TGTCTACATTTTTA	AAACAGCCTCTCTAC	SELFNAARGREIVFEQENVNVSHNDVG			
				ATGTGTTTTTTAT	TTTAAGGGTAGGGG	MSQDITMAEVRVALGMMGKGKTVGL			
				ACGCAATAAATTA	CAAGGCTGCGTACAT	DEIPIEVWQCLGEQGVRWLTALFNVIF			
				TAATAATTGAAC	CTTACCCTCCTCAGAC	RTSRMPSEWRLSVVVPYKKNKGDAQSC			
				ATCACTAAAATTA	GCTGCTCACTGAGTA	SNYRGIKLLSHTMKLWERVIECRIRRIVT			
				TAGACACAATTGG	TGTTTGGTTTGGTTT	ISVNQFGFMPGRSTIEAIYLIRCLMEKYR			
				GATAAAAAGTTA	GGTT	ECRKDLHMFIDLEKAYDSVPRVVLWR			
				GGCACATTTGAAA		CLAARGVPEVYIRTIQDMYSAVGTCVRT			
				ACTTTACAAATAT		PVGDTQYFPVEVGLHQGSVLSPLLFIIVL			
				AAGGACAAATCC		DVITRDIQAPVPWCMLFADDDIVLIAESR			
				GACAGTGGCTTA		SENVKLEQWHMSLEGYGLRLSRSKTE			
				AAAATCCAAATTT		YLVANFSEEIHEADVAVCIAEGRVPQT			
				GGACCAATTCTAG		NTFKYLGSIQSNGBISADVTHRISTGWL			
				CCTAAATCTTCTC		RWWAATGVLCDKNVPLKLGKFYRVAI			
				AACGGGGGCCTG		RPSLLYGSECWPLRKAQERRLETAEMR			
				AATTACGGTCCAA		MLRRICGNTMTDHIPNDTFRRLLGVESI			

				ATTTAGGCCACTT AG		SKKIREGRLRWYGHVRRKCN SAPVRRV EHISVRGKRKRGRPRRTWADQLSLDM GALNLTGDMTLDKNDWRRRIRVVETR SRGSQRL			
	Penelo pe- 2_LG	GCF_00032 7385.1	Lottia gigant ea	TACGATATATAGA CACAGCTTTATGG AACAACAATCACT GGTTAAAGTTTTA TGCAACGTTTCGA CGTGTTTCCCCCG TCGTCCTCAGGCA CATAATGAGAAG AAGTGATGGGCA AACTATATATACA TTAATGCTTATTA TTAAGAACACACA AGAGATAAATGG TTTATTGATACAT TACATTAAATTTT AACAGGTACAAT AAGAGGTATATA	TAATTATTCATATTCA AACAATTTCCAAC TTT AAGCATAGTTAGTTA TATGAACAATGTAAT TTCTATTAATAAAAG ACTCCAATACCAGCA ATAGTTGTTTATTTAA TCTTCAATTAATCTTT TTTAATTCATTTACTT TATTTCAATTATTTTA CTTAATCAGCCTGTA TGTTATTTTAAGAATT ATATACCTCTTATTGT ACCTGTAAAATTTA ATGTAATGTATCAAT AAACCATTTATCTCTT GTGTGTTCTTAATAA	MRSTMLSNEFLT KSVINLSQKNLTDEQI NLLSRGLKFSPVTKLKNHDDFIVNIERGL HQLAPGGHINYLRHQVTEILAKNKIQPN NLTKVQRDALTDLRQDNKIAILPADKGK ATVIMDKKEYQETVTELLSDAKTYRRLR KNPTSIYKDRLDKKLNELESSNQITNQLK FKLRSSRHPRSPYARATIKIHKNP I KARLL ICSRDSVFYNIAQHLSRILAPLGKTADSF I KDSEEFVTKLKE LQITGKMYSYDVVDLF TNVPCKDVL SIIRDKISKLETPLDTTLTTD SIVDLIQMCMSSTYFTWGDEIYEQIHGL PMGSPLSPLITEIYMTAFEEKTMETSPI S PICWFRKVDDTFVMLRAEDDPSALLLN AQHKRINF TMEAEEDDSIPFLDVRIKRL GGQLKTCVYRKPTH TDQYVNYHSNTP QHVKRAAITALVRRAKRICDPEYFQQEL QHIKKVFCLHNNYPPAFVHQAFRSVLM	cd0 030 4	316	397

				ATTCTTAAAATAA CATACAGGCTGAT TAAGTAAAATAAT TGAAATAAAGTA AATGAATTAAAAA AGATTAATTGAAG ATTAAATAAACAA CTATTGCTGGTAT TGGAGTCTTTTAT TAATAGAAATTAC ATTGTTCATATAA CTAACTATGCTTA AAGTTGGAAATT GTTTATTAAATAC ATTTATGAATATG ATGATAAGTTTAT CATATGCATGCGA TAAACGAACACCT TCATCTCTGTAA GGTCAGGCTGAT ATCTTCTGATGTG	TAAGCATTAATGTAT ATATAGTTTGCCCAT CACTTCTTCTCATTAT GTGCCTGAGGACGA CGGGGGAAACACGT CGAAACGTTGCATAA AACTTTAACCAGTGA TTGTTGTTCCATAAA GCTGTGTCTATATAT CGTA	AKPTNQPPNPSTRLAIPFIGKASYQ IKKILKKSVDIETVFSNSTLKDLLRASGK LNPSNNQKPKGCVYQVECSGSSYIGET KRPLDVRLKDHQDKPGNEGTSAAAEHL AMGRKTGHKVNFDNTNILTINKHIDKT RRLHEAIIHRRYQPDNRDEGVRLSHAY DKLINL			
--	--	--	--	---	--	--	--	--	--

				GATAGCTTCATGA AGTCTCCTGGTTT TATCTATGTGCTT ATTGATTGTCAGT ATGTTGGTGTGT CAAAGTTGACTTT ATGTCCTGTTTTC CTTCCCATAGCAA GATG TTCAGCCAC AGCAGATG TTCCT TCATTGCCGGGCT TGTCTTGATGGTC TTTGAGTCGAACA TCAAGGGGTCTCT TTGTTTCCCCTATA TAAGAACTTCCAC AGGAGCACTCAA CTTGGTAGACACA TCCCTTAGGTTTT TGGTTGTTGGAG GGATTCAATTAC					
--	--	--	--	--	--	--	--	--	--

				CACTGGCTCTGAG CAAATCTTTAAGC GTTGAGTTGCCAG AGAACACTGTTTC AATGCCGACTGAC TTTTTCAGGATCT TCTTGATTTGGTA TGATGCCTTTCCG ATGAATGGGATA GCTAAGCGGGTT GATGGTCTGTTGG GATTAGGCTGTTG GTTGGTTGGTTTT GCCATTAATACGG ATCGAAAAGCTTG ATGTACAAAGGCT GGAGGGTAGTTG TTGTGCAAGCAAA ATACTTTCTTGAT GTGTTGTAATTCT TGTTGAAAGTATT					
--	--	--	--	--	--	--	--	--	--

				CTGGGTCACAAAT CCTTTTGTCTTC TGACCAGCGCTGT GATGGCTGCCCTT TTGACATGTTGAG GGGTGTTGGAAT GGTAGTTCACATA CTGATCTGTATGT GTAGGTTTCCTAT ACACGCATGTCTT CAACTGACCGCCT AACCGCTTAATTC TCACATCCAGAAA GGGTATAGAGTC ATCTTCTTCTGCTT CCATGGTGAAATT GATCCTCTTATGC TGAGCATTAAGAT GCTGAAGTAGAG CTGATGGGTCATC TTCAGCTCTAAGC					
--	--	--	--	--	--	--	--	--	--

				ATGACAAAGGTG TCATCAACTTTCC GGAACCAGCAGA TTGGTGAGATCG GAGAGGTTTCCAT GGTCTTTTCCTCA AAGGCTGTCATAT AGATTTCAGTAAT CAATGGTGACAA GGGGCTTCCCATA GGTAGACCGTGG ATTTGTTCATAGA TCTCATCTCCCAT GTGAAGTAGGTA GAGGACATACAC ATCTGA					
	Crack- 23_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	TCCGTTTCTGCTT AATAAATGTGTTA ATATATTTCCAGA GTTTTGTCTGGGG CTGTCGTATTATC	GTCTGGAACGCTCTT GTTCATTGCCGTGCA ATATGAATGCTGCAA TCTCCTTTAACATTTT ATTTTAATTAGCACA	MIPELCSTGLKICHININSLPHKMDELKIL LGNDPFDIVAISETHCDSTITDLDISLDN FCVTRKDRSRHGGGIALYVKNNISYEVIS QLDNELEVIFVKVKQKQTHPLYLGVVYR PPNSPNEFFDKLSEIIHSISENAGEIIIGD	cd0 165 0	516	775

				AAGCAATATGCTT	CGCCAATAATGAATC	FNCDLLPEICTPFSRALLSLMDTNLLHQL			
				ACAACGACAGCC	TTGCTTACTTATTTTA	ISVPTRVTPHTKSLIDHVFSSEIPEEHQHT			
				CTTCTTCCACATAT	TCAATATGCTAAACA	GVVKTHISDHYLIFTVFGYARAKQKNDD			
				AATTTTATAATGT	ACATTATCATTGATG	NIIEYRCFRHFSNDNFVTMLQNLPFNE			
				GTCATAATGTTTG	TTTTTTTTTTTAAACC	NLNDKDPNHAWQCWVDLLISAVNKH			
				ATAATCCATGAAT	TCAATCAAGGAATCT	APVKSKRIRSKLCPWITGDIEAMNRRD			
				GCAGAAGAAATC	TTGAAATTGTATTTCA	WLYTMATKPSNPDSSEFWSEYKSLRNE			
				AAATCAAAATCAA	TGTACTIONAAGCAATA	ITCMIRNAKNDYINALLHENGCTPNAM			
				ATCAAAATTCATG	CCATGTTATGAATTG	WKTLLKLISSSKESNLKLEVDGEEVIDAE			
				TATATTGCATACT	ATTATGATACTAACA	KIANHFNNYFVDAVDTMFNPLNFEND			
				TTAACTTCTGGCC	TATGCATGTTATTATA	VNNSTIDRASEPSVHSDTVDTVPPVNSF			
				AATCAGCGTAGTT	TTTATGTGCATAATAT	DLPVITDEFVMKEIERIPIKKATGPDDVSI			
				GGCCTTTTATTGT	GTCAACTATTTGTTTT	KLLKFACNSQNVIQSLTYILNSSLDQGIF			
				CAACGATAACATC	TATAATTAATAGTTTA	FDEWKIARVQPIYKAGSKLSVENYRPVS			
				ATCCCTGGACTAC	AAGTATTAATGGGA	LLSIPSKILEKAVNTEFQEYLKVNKILTER			
				ATTTGGCGGCAG	AAGTATTAATAGTAT	QFGFRPNHSCETALLCMVDLWAQNVD			
				ATGGCGTGCATCT	ATATGGTATAACGTA	KGNLNGVAYIDMRKAFDAVNHSTLLM			
				TAACAACCATGGA	ATGGTGACGGCAAA	KLKNVGCTDRAIKWFRSYLGGRSQFVSI			
				ATACGCGTAATG	CTTTTTTAACATTGAA	KGKKSSTRTINYGVPQGSVLAPLLFSIFI			
				GCGCGAAACCTC	ATATATTGCAATAAT	NDLPTSIHSADMFLFADDATMSVQGK			
				ATCACTCGTCTCC	TTGTTTTGATATTTAT	TLSEIQTKLNLALDEVYSWTQKNKLLLN			

				GAGATCAGTCAA	ACCATACTTTGTTTTA	TTKTKVMVIGSKQKLQTLNENNLNVRIK			
				GCGTCACTAGAA	ACTTTAACCTAATACT	TTPIECVTQYKCLGVIVDSNLLFSKHVEK			
				GTAGCCCGCCGTC	ATGTGTTATTGCGTA	VALQMKQKLGLRRLKGTFNKHYSMIY			
				TGACCCACTGCCC	CAAATGTCAACTACA	WGYILPHAMYCCTVWTNRSLHNYETI			
				AATCAGCGCCCGC	TGTATATATATATATA	NKLHKRAAYIISGCSWSTPSHEVLLQLN			
				GATCTCCGAATAC	TTTCAATTGTGGGTTT	WSTLEALYSRAMGIMAFKCVHHLAPAL			
				AATGAGGACCGA	TTGTTGTTGTTTGT	LANKFILHDDVAQRMTRNSHELKLESN			
				TCCATTCTCGAAT	ACAATAGTTTTAATT	RCNTEFFKRSFVNSAITLWNNLPIETRFS			
				CGGCAAGACTCC	ATAAACTGGGAAAGT	PNLNVFKSNLSH			
				GAAAGCTCCTCGC	ATTTAATTGTATACAT				
				ACGCCATGGCTGC	TTTATAATTACTCAAT				
				TGAGAATGATAA	GGTGACGGTAAACTT				
				CTCAGCTTTTCGG	TTTTTAACATTGAAAT				
				AGACGCCAACATC	ATTCGCAACAATTTT				
				GCAAATTCCGCAA	ATAAATATCTATAGC				
				AAAGCCGAAGGG	ATACCTTATCTTAATT				
				GAGAGTTTTCCCC	TGAACTTCGTATACT				
				CGAGACTGGCTC	ATGAAATGTACTATA				
				GACTGTCTCCAGA	TCGCCTTTTTATTGCA				
				CGGCCCGCACGCT	TGTCCGAATTATCTCT				
				CCTGAATCGTCAA	CATATTATAATGTGC				

				GCTCAGCAATAG ATGATAGCGATA ATTTAAGTCTTCA CGCGGAGGATCT TAATGTAAACGCT ACAAATGTAATAA ACATCAATGAGA AACAAACGCAAA TTGAAAAAACTAC AGCTTCCAAGGAT GAAAGTGA ACTA AATGAAGGACAC AATGTGACAAATA CACAAGGATATTA TGA ACTTGTTGAT GGC	ATTTATTGTTGTTTTG ATTGACATTATTATTG AACCATGTATGTTGT AAAGGGCCCTGATG AAAACCAGCTCATGC TGATCGGGCTACCCT TTTTAAAGATGTGAA ATAAATAAATAAATA AATAAA				
	HEROT n	GCA_00018 0735.1	Tetrao don nigrovi ridis	AGATTGGTCTGGC TAAGCCAGTGAC GTCCAGGAACAG ACTGGCTGACGA CCACGAATAGAG	TGATCACTCCCAGTC GGGTGCGCTGGGTG AGGGGGTCTGATGTT GAAAGACCCGAAACC CCCGATGACCCCAGG	MATTQASVKPTAVATCVCCKICKKNPRG LKIHQTKMGCLASVQPEQRARFSLSESR EVPARAEPYGPQQPHSPEALGETQEER GQESPHSAQNLRAQVAQAPDNPQHH RRVKWPPASKVSEWQQQLDEDLEGILES	cd0 165 0	362	627

				TGGTGACAGCTTG GATAGACAGCTG ACAGCAGGGAAA GACGGCAACCGG GGCAGGAAGGGC TAGCAACCCAGCC TGCATCTTCCGTG AGGAAGAACCCA AAACTTGCTACGA AGAGCCCGAAGC AAAGATACCCCA GGGGAGCCCGAG AGGGGGGGGAGA ATGAGCTCCCCAA ACGGACGGATAA C	TACTATCACTGACGA TGTGTCCAAGACATG CATCAATAGGTGTAT TTAGAAATC	TAKGGVDRKLQTM TTLVISFATERYGT MEKRAAPEKYTKNRRAEKISQLRQELR VLKKQFKGASEDQKPGLAELRCTLRKKL LTLRRAEWHRRRAKERAKKRAAFLANP FGFTKQLLGQKRS GHLECAKEEVDSYLH DTFSDAERENSLGECRVLISPPEPACSFN TKAPTWKEIQT VVRAARNNSAPGPNG VPYLVYKRCPKLLARLWKILRVIWRRGK VAHQWRWAEGVWVPKEEKSTLIEQFR TISLLNVEGKIFFSILSHRLSDFLLKNQYID SSVQKGGIPGVPGCLEHCGVVTQLIREA REGRGSLAVLWLDLANAYGSIPHLVE MALARHHVPGPIKTLIMDYYDSFHLRV TSGSVTSEWHRLEKGIITGCTISVIIFALA MNMLAKSAEPECRG PITKSGIRQPPIRA FMDDLTVTTTSVPGCRWILQGLERLMT WARMRFKPGKSRSLVLKAGKVTDRFRF YLGGTQIPSVSEKPVKSLGKMFDGSLKD AASIRETNDQLGHWLTLVDKSGLPGKF KAWVYQH GILPRILWPLLVEFPISTVE GLERRVSSCLRRWLGLPRSLSSNALYGN			
--	--	--	--	--	--	---	--	--	--

						NNKLTLPFSSLAEEFMVTRAREVLQYRE SKDPKVALAGIEVRTGRRWRAQEAVD QAESRLHHKELVGAVATGRAGLGTTPT THLSRLKGKERRDQVQLEVRASIEEQRA SQWVGLRQQGAWTRWEEAMARKIS WPELWRAEPLRIRFLIQSVYDVLPSN LFLWGKVESPCPLCQGRGTLEHILSSC PKALGEGRYRWRHDQVLKAIAESISSA MEYSKRLPLPGRGVRFRAGEQPPPQP RAQPGLLATARDWQLRVDLGKQLKFP ENIVETNLRPDIVLHSQSSKQVILLETVP WEERMEEAYERKAGKYAELVEDCRRRA GWRSRCLPIEVGGRGFAGKSLCKAFSLL GITGMRRRKAICAASEAAERASRWLWI QRDKPWTSASWTQAGN			
	CR1- 65_AA e	GCF_00220 4515.2	Aedes aegypti	C	TGACTGTATGTTTAG CTTAAGTTATTTATTT ATACATTCATTAAGA CAACTTCGTGTCAGA TGGATTAAATCAGAA ATACAAATACAAAT	MNDHLSANMHRTLSHSGRTTSSTMEA PTPPSTVESLQPALISRPDPVVGRGGGV FQPVLSGKYNKITDSSLPDSPSSLRIYY QNIRGLRSKIDSFFLAISELDYDVIVLSET WLDDCILSSQLFGSNYSVFRTDRSALNS DKGRGGGVLIIVSSRISCSVDSAPVSNQ	cd0 165 0	586	838

						LEQLWVKIVLPRNNVSIGVIYLPDCKN DIDIQRHIESIGAVHSHLHENDMAML FGDYNLSDVHWTSDANDNVFVDLTRT RLNPASSTLLDGFCFHGLSQVNTIVNSG GRTLVLVLLNDVVLPNHSIHEAVEALTT VDVHHPPIELVWSCPLPVVYETVSDNTG FDFKKANFESLNVALRSMDWDR LDSIE DVNEAVAFFTNSVMQAVAEHVPPRRP PSKPPWSNSRLRLKKRRRSAALRFYRSH RSQQSKLHFNRSSSEYRSYNAYLFAQYK SRTEQNLRSNPKQFWSFVNTKR NEDGL PTSMYDEL SADSADKCELFAMQFKR AFNDYASPASQIDRALQDTPQDVFSYD MFHVSEEEVIGAIGKLKFSYNPGPDGIP PALLKRCSTLLAPLTKLLNHSLRQQVFP ACWKKSFLFPIHKKGDKRCVSNYRGITS LCACSKVFEIIVNDSL FESCKRYISTDQH GFFPKRSVSSNLVEFSTLCIRAIDAGKQV DAVYLDLKA AFDRVDHRILLQKL R KCGV AINFVDWFRSYLTERSLCVRIGESSESASF TNISGVPQGSNLGPLLFTLFINDVSLLL P			
--	--	--	--	--	--	--	--	--	--

						TGNRIFYADDAKIYMIVESLDDCYRLQH LLNLFEAWCTRNLCLTSIEKCQVISYGRK RSPIRFPYELSGTTLERVKCVRDLGVTLD EKLTFNCHLNDVISRANKQLGFVLKVSS GFKDPLTLKALYCALVRPILEFAAIVWCP FQNYWISRIESVQRKFVRHALKNLPWR DPENLPPYHERCRLLGIDTLENRRHVAQ TMFVAKVLRNEIDSPSLLADVNIYAPER TLRRRQFISLGSRNTRYGQHDPVRFMA LKFNEAYHLDFNVTITTLRNHFQYFRR N			
	I- 4_DAn	GCF_00328 5975.2	Droso phila ananas sae	TT	TAAAATTAATAAAA TATACTGTAACTTA GCTAATTATAAAAAAT AAATAAATAATTA ATACGTAGTCGAAGG CCTCGTAGCTAGCAC TACACCCCAAAGTGT TAGTCATAGTTTTGT AAACTGTACTATCAT	MKINQANWDFATHHTNNNNNLYPPS TNINKETAQINKIIMKSAHLAIPQTLPHR HTHTVPWWNKDLADLQEKKKQTWKA LKRDTTQENIIKYRKTNALFKRAVKIRKK EATSKFTSTIQPNTPTTKIWNNIRRFCGL NPVKQIHCTNNNVTTNDKQKIANLLA QHWSQLSKDSNFSQEFTSHKYNTTTLN HIPSKSAMEIDKPITKLELTIALRSLKGITP GINRISYPMLKNSPLAIQNRILSHFNHIF ENYIPQSYKTSIIIPILKAQKDKTDIASYR	cd0 165 0	260	518

					ATCCAATAAATAAAT AAATAAATAT	PISLNCCVGKTLDKIIANRLWWYLN SNKLLHPHQYGFKKGKSTSDNLLYADHIITKS LNTRKHTSLISLDFARAFDRVGIHSIVDQ LIEWKCSHKIINYIINFTKNRKIAVRIGSSI SNFYPLENGIPQGSPISVVLFLIAFNKLA NILTIIYKEIKFSAYADDFLLLVLNKNQKN TTINLTHIFRKIYDWGSYSGASLSQTKCQ HLHICRKHRCNCTITTQNSNIKNVTS SLKILGLTINNRYNWSSHINSLIPSLSNKHNII KCLASHKFNCNTSTLLNITKSILISKIDYAL HIYGYTPKNFLNKLKTTLNAAALRTVLGA FRTTPTHNLLYEANIPTIEQQRDHLTTKL FTNLTRSKDTPHKKHIQPKTNKKWTST LERTIENCQLHNIPFKPTKLYKTKPPWQ LPQNIINTSLHNYKKTNTSSDTFQKLFFE IKETYKNHTFIYTDGSKINLNTAFAVTTE TDVRLRYGILPPSSSILTGEMIGILEAIKIST TKGKSVICSDSLGAIEAIQNPLNLNKYP TYIRSLQTKQPKIKILWIPGHSNIKGND LADRTAREATKMPLIYTPNENITDINKM LKNILIQEKETLINKCNTWYKNINTNNT			
--	--	--	--	--	-------------------------------	--	--	--	--

						HISALSKIQQEEISRIDQIKFIRLRLGHTSL THKHLLDPNTQRNCYRCQSSTPTNME HILDSCPAISQNKQNLLKNRPATEYLKN PTAENIKVLIKFLKETNLYNEI			
	CR1- 83_H M	GCF_00000 4095.1	Hydra vulgari s	ATGGCGGCGTAA AGATGGCAGCCT AGTTGAGCACGT GTTTTGAAATAAT TTTAATA	ACATTTTAAATACTA ACTTTTTTATGTTTG TATGTTAGTGTGTTA CAGCTGTAAAGTTA CAAACCTACGTAGT TTGGACTCATCACTCT TTAAGTGTACAGCAC TATTATATTATATT	MVVSKDTTQAINAIEVLRKELIERIIKVE KENERLNTENESLKKQLAEQCNQNNNK PLFSSLIKKTSEGAPPISDSETRLLNAISAE QKEKIKKEKNVLVFGLESKRKSKDEIKE DDDESVEVFDAGLDKSCIVKHFRLNS RDKNRPGPLVIELVDPSYQKQVLSVAR DLNKVEAYKNKVFINPDLTYRQRASLKL LLGERNSLNKIEEGQKFDISICYQKRYAA ENNRGRSKVRHFDLLSETICCRKLKSKN TKKINAPSSLISKLSEITCLYMNATSLNNK FDELIEVINSNKANVVMICETWWNNES VTNINGYTLFRRDRGVGRGGGVCIYVQ NNIKAYEVSEKCLNDPNIEQIWCVIELG SEKVLGSMYRTGISNISNCINITKSINFA HSMNLLNKYTGILICGDFNFNAIKWTE DYYGTLTSETDTPAKIFLECLNNCFMHQ NVIKPTFQNNIGNETNVLDLIAENKNR	cd0 165 0	788	1054

						VYKLVHLPPLGGINHGHHIIKFNYSHEEC IDNIVLIKEKFNFNRGKFDKLNNYFKILN WDNEFKGLEINKCYEKWLSLYEEGCKLF IPIIKSGSIQINTKPLWINRELKALSKEKR NVWFECRRQKFRDNCLLAKYKELNKVY KKKSESKNTRFIKKKVNQRIKDFEYDLAL NSKFNPKKVYSYINNKS VKDNIRAIQLK NGSIEDLKVIANTLNGYFASVFNKPQT LSAPIIDIETQKICKNPDFNASVVEEYLNK LDTNKATGMDRVHPKVLKECKSTLAKPI SLLFNKSFETGKLPKLWLCANIVPLFKNG DKLNPCNYRPTCIFNICSLTHATIDLPVSL TSVVCKVMEKIIKDKMMKHLVENKLIN KNQHGFVNNKSCVTNLMESLDFITSSID NGLDVIVLFLDLAKAFDSVDHERLLLKLN ALGFKSKLFDWCQGFLKNRLQRVVIGN NKSEWEEVVS GVPQGSVLGPLLFVYI NDISN KIKSSCKLFADDMKILRAIKNDN DVIELQQDIDLLMEWSNEWLMKFNQ QKCKVMIIGNSQHKVLMNNTALAYTS MEKDLGVYISDDLEWKHHVNKAVNKA			
--	--	--	--	--	--	---	--	--	--

						NQKLQIKHTFKYLDEKTMKLLFIALVR PYLEYAAPWNPYRQYDKDKLECVQQR ASRIKILKGYSYEERLKKLDLLSLENRRRR GDLIQMFKYLKGYDIINFYTKPEMLDNG YKTRGHNSKLRRQYTGNIKRHNFFTNR VVNDWNNLEQETIDAVSINQFKKKIG			
	CR1- 22_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	TATTGTGAATGAA TTAATGACTTTTT TGTAATATTGGA CCGAATTTAGCAG ATGAATTGCCGAA TACAGATGTAGCA TATGGTGATTTTT TGGGGCAGAATA TAGAGCAATCTTT ATTTTGAAACCA ATAACTGCAGTTG AAGTTAAGAGTC AAGTTCTGCTACT TGATACCAAAAA	ATAAATATAATTCAT AATTCATATTTTACAT TATTAGTTTGTATATT GTAAGACATTGTTGA GTAAGACATTGTTGA GTACTGATTCTTGTTT TTTTTAAAGTATCGAT TGTAAGTAAATTC ATTATTTTTTGTGTTG TTACGTTCTCTAAACT GATTTGTACCAAAAT ACTTTGTTTAGCTGTT TAAGGGGTTAGGCTC GACTAGCACATGTGC TATTATCTGGCCCTT TTACCAACCAATAGT	MIPVQLIRDSAPYISHPLAYIFNLSLETGK VPDEIKIAKVTPLHKKGPREDPGNYRPIS VLPVIGKVLEKLINIRLIQFLEKNDIFYH QYGRKKYSTKLSLINLINSTLKSIDQGKV TLGMFIDFKKAFDTINHNILLGKLSSYGI RGLPLEWVKDYLSNRFQFVQCGQTLSE RKKIRCGVPQGSVLGPTLFLIYINDLPTV TDYFNFRLEFADDSNLFHTFSADQGEIDV NEVNNNIQKIHKWCTANKLTINSMTN FMLICSRRKKIGINGKLKIAGEEIKEVDV ATFVGVDHRLSWKSHIRFVNQCIRKK VGVLFRRLHFVPSRTLILLYKAFIQPHISY GIEVWGSTYKTSLSILYSQKMAVRAITF SEWRTPSDPLFKRLRILNVYNLHDLISIST FMFDLVNGNLPHILIEYCDIIEHSYSTRQ	cd0 165 0	37	295

				GGCTTGTTGGTCAC GATAAC	TTCTCTTCCTGTATAT GTAATTATTGTATAA AATGTATGAAATGTA TTGCTGTTGGTAAAT AAAA	KEDGQLRLPKLRTTQGQFSLSFVGSEF WNRLPVDVRLTRSRNFRNSVTSHLLN KS			
	Daphn e- 17_SP	GCF_00000 2235.5	Strong ylocen trotus purpur atus	TCCATTTAAGTTC TTCTTTACATTTTG TACATAGTTGTTG TTGTTGTTTGCTTT TTACTACTTTGTTT TAAGTGTAGAATT CATATTCCTTTGCT TTTTGGTTTTCA A	TGATATTAGCTACAG ATTTTAACGGTTGTT AAATTCATTATATTTT ACATGTAATTGTTGT TTGTTTCACTTTTAT TACAGCTTAACTTTTT ATATTTGATTAATTG GTCAGCCGCATACAA GGACCACCTCCTTTCT GGCTGCTACCAGCAT TAGCAATTCCTTTGT ATTATTCATGTTTGTA AATTGATTCTGTGTA TTTCTGTGTTTATCTA TGTTAGTAATTGTAA TTGTGTGAAACTTTG	MLCRARFVQCLNEKHVSVSTLREFLSPS DPQSTNNFTQSHYTPDEFNTFVTQKSL DKDVNNKLSFININCRSLQHNFDAISVL MMSLRMQVSIIGVSETWLRDENSQYSL CNLDNYVFVGNNRRLGRGGGVGIFIRK DINFKRRTDLINESFLESIFIEIERFNSSIL ICVMYRPPSQPVDDFLMSLVNVFDVV NVEKKLFYLMGDFNIDLLNVGKVQSVN DFLNILMLHSMIPLINEPTRVTDRSSSLL DNIFTNDSCYIYSGILFTDISDHFVYCIC DTDIPKSSCETDFYKRHLNENNIRMYFL QLLYSTEWYLSSDVNSSYNFVFKFMK VYNLCFPVKLVRMKYKRKNKPWITNEII KLSKKKCRLYKLLIKNPVYRKEVYRKIR NELNNKIKIEKKNYLKKSFKDSTGNMCK TWDIINKAMGKKCKMTNIKEINNNDTV	cd0 165 0	557	812

					TATACTGCTTTTGT TACCCTGTATGTTT TATTGCTTGGCAAAA TAAATTGAATTGAAT TGAATTGAA	ITDKNDISNNFNEYFVDVGYNLNCKFDC FDDSECIKYIGMNDKSMFFKLVSSEIE DIVKKIKNNTSAGYDDVDVKIVKKCIYLL SSHLCTIFNQCLEQGVFPDKMKIAKVIP VFKNGVSETMSNYRPISVLTVFSKILEKC MYNRLYAFLT KCNIINDNQYGFRAHGS TSCAIDFIHKVASAIDNNEV MLGLFLDL SKAFDTLDHHILLVKLYQYGIRGIGLDF KSYLLHRHQIVSIANVKS NYQEIKCGVP QGSILGPLLFILYINDLPSVSRILKSILFAD DTSIFLSHQNIQSLQDIFNQEI KLINTWF SVNKL SINISKTHYMIFTNKKCDTDSIIR LCNTEIKQVSSLKFLGITIDNKLKWDHI NAVCNVLSKNIGVMYKLSFPKCILKMI YNSLILPFLYYGILAWGNAHETHIRKVILL QKRAIRTI CHTSYLAHSHPLFKEMKILNF TDMYNYQCAIFMFQCYLQILPKSLLKLFI LNCNIHHYSTRNAENFHLPLVHTTSYKR TVIFNGPCLWSTIPRNIRELK FVNTFKRK YKKYLM DVY			
--	--	--	--	--	---	--	--	--	--

	R2NS-1_SMe d	GCA_00260 0895.1	Schmi dtea medite rranea	CCTTTTACGGTA TAATTATAAACAT TTATCATAACATA CTTTCTGTATCAA CTATATATATGAG TATAAATCTGGAT CTTCTGAGAAGG ACTCATATATATA AGATAATCTATAA AACTGGAAGGTA ATGACGATGATTC AGAGGTCCAGTG ACTATTCCTCGAG AGAGGAACTAGT TACGCAGTTCTGA ATCGGCAATCACT AACTCAACAGGCT GATGTCAGGATAT GGAAAGGGGTCT AATAGTGGTCTTA AAACTCTCTTTTG	TGAGTAGTGGTATAT TTATAAATATAAGTA GCGAATTTATAATAT TTATCTATACTATCTT TATATATTTCAAGTG GATCTTTTATGACTA ATTAAAAATACTTTTT TATATTTTAACTTTT ATTACTTTATTTTCTA TTTTTATATTCTATGA TTCTTATAATGTAATT TTATTGGGAAAGAGA GTAAACGGGGGGCA ATGGGGTTAGTCTCA GTATGGATCCGGACC CTCCGGAATCGGGAT TCAATGACTGGGGCA GTAGATGCGGGATG GAAATTAAGAGTACT TTGGCATTAGGGTTT CTGGGACTGTGAGTG	MAIITGEGDCHLLKKSLLKLKTMKNKTN KLNVNKGNDLDVTFPCDFCGLSFPTKIG RGVHISKAHRVERDAKLSGHASAKKAY YWSTEEETRMAKIHLQEESRGDSNHVL RAMHSACSTGRTLESIRLHMKAKRYVKI YTELRNEKQIQSSGVENILNNISDASPLT LPSETETLTSYLLNLPLKKGVS KWREQDL DFIIHSVSDSEEQKRLLDVYCKTVTEGYR KGKTGPLKKSTKVSPHHATTTKSKKFLK KKLKRELYGKVQQLWIKDRCEAAKLVL GKVSTLICEEDRSARQGEYWKKIFERES EPDIRPVEVYADPMLVRPISEDEIKKV KDIKLTSSGPDNISTEVLREIPVLVLVKL FNLIILLSHLPSEFKSSRTTLIPKVETPKDP GDYRPISVSSVFVRALNKILASRVSN KGAHAQKGFKALDGCLNTVLLNCILG DAKINNKSIALSFIDMKKAFDSVNHDSIL RAATIAGYPKLLLGYIKESYNGADTEVEK IKATFKRGVKQGDPLSPVLFNNIVELAIK KAYESGIGYRMNGTVFPIMAYADDIILL SESKEGLQDLLDIVLTELRHNGLEANAD	cd0 165 0	385	643
--	-----------------	---------------------	-----------------------------------	---	--	---	-----------------	-----	-----

				AAAGTGACTTCTT	GTGAAACCGTAAAAA	KCASIIAKKDGGKQKKWVISSDSVLAAGK			
				ATTGGACACCTTT	GGGAGTAAGAATCC	TMPCMNANSMYKYLGIYFGPLGITTER			
				TTATTGTATAATT	GGGCTATTACCTGTA	LTKIYSNQLYNIGIAPLPQQRLEMINDF			
				ATAAACATTAATT	CGGCTAGAGAAATCT	LLPGILHEAVLGNTNVGELKDLDKMTR			
				ATAAATCATATCA	AGTTCAGCCTTGATG	VAIRKWLHLPKDTCSSFIYTGVKQGGLG			
				TACTTTCTGTATC	CCGTATTTTTTTGGG	VRSILAEVQFGKISRLESLNKSKDTRIQUAL			
				AACTATATATATG	AAGTTTATTCGCACT	TTCEVVKKAQRQWSKPIRVGDSIATSKV			
				AGTATAAATCTGG	GCCGTGGTTTGCAC	QFYEANRELLLSKCDGKGLREAGKIPFA			
				ATCTTCTGAGAAG	CAAGTTTGATGGGAA	HDWLRAKGKLMKGRSFINSIKVRGNLF			
				GACTCATATATAT	GACACAAATAGATTT	GTPVRLNRLNKSGIPLIKCKMGCLMPCS			
				AGGATAATCTATA	AGTTTTGGTTCTTTTG	LNHIAQTCPKVHGLRVERHNKVMNYIA			
				AATCTGGCAGGT	CCATTACTGAAATCT	DVARKGEWKTLVEPIIPYGATIRKPDLV			
				AGTGACGATGATT	ACAACGAAGCCACCG	VIKDKKAMVLDIQVVSDVGNLNERNTL			
				CAAAGGTCCAGT	GCTATAAAGAGTGGT	KSEVYNDKIIADWVKNAGADSVEFGAI			
				GACTATTCCTCGA	ACTCTTGTGGTATAC	TLNWRGLWASDSWETVSKIGVNKSDA			
				GAGAGGAACTAG	CTTTTAAGCACGGTA	MIISLKVLNWTARMYCAWNKG			
				TTACGCAGTGCTG	AGGAAGTGGAATGG				
				AATCGGAAATCAC	ATGGGTTGCTCATTC				
				TAActCAACAGTC	CGGACTGCGCGTTGG				
				TGATGTCAGGATA	AGCAATTATATTGAT				
				TAGAAAGGGGTC	CGGGTGATACTAAAA				

				TAATAGTAGTCTT	AGCCGGGAAATATCA			
				AAACCTCTCTTTT	ATTACTAGTTTTCTGG			
				GAGAGTGACTTCT	AATACTGCTGCTCAC			
				TATTAGACACCTT	TTACAGTTTCTGGATT			
				TTATTGTATAATT	ACAAATTTACCCATA			
				ATAACCATTTATC	GGTAATTGATATATA			
				ATATCATACTTTCT	GCTGGCTGTAAGGG			
				GTATCAACTATAT	GTATGGGACCCTTAC			
				ATATGAGTATAAA	AGGTATCATTATTAA			
				TCTGGATCTTCTG	AGGCTTGGATGATCA			
				AGAAGGACTCAT	GATGGGTATATAAG			
				ATATATAGGATAA	GGAGGGCTATCAAG			
				TCTATAAAACTGG	TTACTGGTTTTCTGAA			
				CAGGTAGTGACG	TGATATTTGGATTAC			
				ATGATTCAAAGGT	ATTTCTATCAAGGTA			
				CCAGTGATTATTC	ACAGCGACGTCTATA			
				CTCGAGAGAGGA	ACTTATATATCCATAC			
				ACTAGTTACGCAG	GAAATGGAACGATC			
				TGCTGAATCGGCA	AGATGGATAAACAA			
				ATCACTAACTCAA	GGGATGGATTATCAA			
				CAGTCTGATGTCA	GTTACTGGGTTTTCT			

				GGATATGGAAAG GGGTTTAATAGTG GTCTTAAACTCT CTTTGAGAGTGA CTTCTATTAAAC ACCTTTTAAATGT ATAATTATAATTA AATATAAGCAG	GGATGATATTCGGAT TACATTTATCTTAAAG GTAACAGCGAAGTCT TGAACCTTAACTTTCCA TACGTCTAGGGAAAT TCGCTAATAGAAATA TTAGTGATTAATAAC CACACGGGCTAACAA CCTCTAAGGTTGGGC AAGTGAGTTGGATTC ATCCAGCTTGCGCGT ATTATATTTTCGATCC GATGTCACCTGAGGT TACGGGGAATTAAAC CACAGCTTTTCTTATA TACTGAATCAGGAT ATAAGGATTGCATAT TAAACGTTGGTGGG TTAATGAACTATAAC TGAGGAGTAATCAAA TAGTAATGATTTGAA			
--	--	--	--	--	--	--	--	--

					TGCAAAAGGTTTCAT TTGTTACTGATTAAC TGCGGGTTTCATTTA AATATCTTATTTTATA CATGGGCAAGAATA ACAATGTGAAGTAAT TCAGCTCTAGTAATC TAATGATAAAGCAAT TACCAAGGTATAACG CCCTATTTTGTAGGG TTGGCCAAATATAAC TTGATAATAAATGCT GATTATTGATCTACT AGGTATCAGGGTGC AATGCCTTACGAAGG TAAGGTTGGACAAAT GTAACCTGGTAACAT GTATAACTATAAGAC CCCCGTTCAAGGATAG TGGTGATTGATAGTG GTGATGTAGTTAATA			
--	--	--	--	--	---	--	--	--

					CAGGCGTGTATGAAT CTGGTGAAGGTGGG AAAAACTTCATAACT CCGGTTATGATTTGA AGGTCAGGTTTGGAT TGGGTCTGATCGGCT AATCAAGGTAGTGAT GAACTCAAAGTAGG GGTCTATGGAGAGG GAGCACTTGATAAGA TAATTCACGATAAGA TGGTTAGTAAAACT CGCGATCCCTGCGTT GGAAAGTTTTGAAGT GCTGAAGGTTGGTTA GCAATAATCAACAAT GCTCGTTAGCGAAGG GAGAATAATTCCGGG CTATCATTCAAACCT GTATTTTATCAATGA ATTATT				
--	--	--	--	--	--	--	--	--	--

	Vingi- 2_SPur	GCF_00000 2235.5	Strong ylocen trotus purpur atus	AGGCTCTGTATAA GCTTGGCTTGTTA GAGCTCTTGTGAG TTGTGTGCCGCGG GGTTTCATATCCT TAAACTGCTGGCG AACAGAGCAACA AGACCAAGCTCTG TAATTGTTAGGTT TCTCCAATGGTAG GTTTGTCCAATAA TGCCGAGTTTGTG CCACGGAACCTA GTCGGACATTTCTG ATTATATTCTTTTT ATGTCATACGTCT TTATCTTTTTTTTT TTAAATATTTTTT ATAGTTTAGTCTA CCTTGCTGATACC TCCTCAAAATATA	TAGTGACTCGATAAG AAGAAGAA	MATNDNDDQFVAEAQGRTHGSLSGP ALSISSNIEGLSGAKQELLADLCITTNCDI LCLQETHRGPTRVRPRISGMTLIAERPH DQYGSAIFARNGLAVDKSYLSDINNIEV LSIDLMGITVSSVYKPPQEPFDLPGRAT DGRMNVVIGDFNSHSISWGYREMNE GELVETWADANQLSLIHDAKLPAFNS ARWRRGYNPDLIFTSTRIANQCEKVILD PIPHSQHRPIGLKVNAVITPRAVPFRRRF NLRKANWEKFAKDFDSYIQDLPATPQN YDTFVELMKKSSRRNIPRGRTSYIQGL TDHSKVMYEDYIEHFQMDPFSEETLES GENLTKEIGESRRRKWQEFIQSTDMTH SSRKAWKTIRILGNDYTKAQPRSEVTAD QVAHQLLLNGRGNPDHHPRRARLPKV PDALSSQDDPGFTRPFTMDDLAAIKA MKNNKAAGLDDILCEQIKQLRPGALN WLLKMFNECLGTNNIPKLWRKSKVIAL LKPGKDAASPKSYRPISLLCHTYKLFERLI LNRIAPFVDEHLIPEQAGFRPGKSCTGQ LLNLTQFIEDGYESKLITGAAFVDLSAAY	cd0 165 0	487	742
--	------------------	---------------------	--	--	-----------------------------	---	-----------------	-----	-----

				TACGATAGACTAT CATTAAATTCTGCT GCATTACCATATA GCTTACTTAGAGG TACTGCAGATCA ATCACCTGTACTG TATCTAGAGTATA GTTAAGTGTAGG AGAACAGTAAAA ACAGTGGGGCCA TTGAAGGAGCGT TCACGTCGCGGTC CGTCAGATGAAA ATCTGCGAATCCT TCAACCACACTAC TCAAGTTTCTTTTA GTA ACTACCACAA CCAACACC		DTVNHRLTRKIFELTKDVRTGLIQTLLS NRRYFVVLNGNRSRWRRQKNGLPQGS VLAPLLFNVTNDQPIHPNTRSFLYADD LCIATQNNQSFVKLEESLSDALAGLIPYYA TNHLRANPDKTQISAFHLKNRDANHQL RISWYGKRLKHTPNPVYLGVTLDRLTY KNHIAN TKAKVGARNSILKKLANTNWG TDARTIRTTALALCFSA AEYASPVWSRS AHAPKIDPALNSSCRAITGCLRPTKVEDI YLLCGIAPPHVRRVLVSQKEKDKQENNP LHPLFEQQPVGKRLKSRHSFLHSVEPLN TSPHNQRMTLWSNHLQTTPHKLSLSPK ESLATGSSETWPTWLSLNRLRTGYGRC RVLMQRWGLGDEQTACECGQEQTME HLLVCPLLPEPCTQDDLEAVNCRAKACV DHWTGLV			
	I- 1_DPe	GCF_00328 6085.1	Droso phila	ACTCTGATATCGA CTCAGAAATAGAC AGTAAACACCCTA	TAACTCTAATATAAA ACCTATGGAATACAT CTCACATATAATAAA	MTLTIIQWNLRGYINNYNELQIIKQYQP HIIALQETHLFNIHNIPIPIFSLYAKNYS NFGGA AVL IHN SIESKEIHLQTDFDATAI	cd0 165 0	491	749

			persim ilis	TCAAATTCCCATA AAAACAGCACAA CCTCTAACATTGT AAAAATTCTACCA CTAACTACCTCTA AGCGCAGACTAG CAGCAATTAAGG CTGCGGACAAATT AAAAAGCTCAAG CAGCAAAAAAAC AAGACAGAAGAA CCCCAATACTGAC GCCTTGAGTAGCA ACAACTGCTCAAT GGAAATAGAATC CGACTAAAATAA CAGATCAAGTAA AAACAACCCAAAT AACCAGCACTCAA GAGATAAAAACT CAAAAAA	CATAATATAATAAGC ATAACATAATAACAA CATAGCCGAAGGTCG TAGTAACCAGTGCTA TCATTAATAAGCTAG GTTAATAGTTTAACT CTAACCTCGCTGTAA ATAAATAAATAAA	EVESKKKFKIISAYISPSQNFNESNLHNV LSNNSSSLILGDFNSWSKLWGSPKNN RGNTVAKFLNNTNFILLNDKSPHFTTH NTYTHIDLSFCSPDLFQYANWKIIDDLH GSDHFPILITLFNNNKQNKQTQNNPIFKI NSANWDAFKDKTDYFNTKNSPSNNIN KESSCLNKILLQSAHLTIPQTSIKNHPRN VPWWNNGQLTLNNERKQAWNKLKK HITTTNIIRFRKAAAKFKKEIKSAKNKSLS SFTSEIQPQTKTSQTSWSNIRRFCGLNN MKNIHTIIHPTTLQPTSDNPSIANSFAEY WSDQSKDSNFSSMFTNNKYSISTYQSS PPSQSAQKIESDITSIELSSALSSLKGHTP GIDRISYPMIANTSPAVKDRILKHNNIL NNYIPQAYKIGLVIPIAKPGYIKTSITSYRP ISLNSCFKVLDKIIAKRLWWFVTTNKHI HTNQVGFKRGKSVTDNLAYLDHLITDA LSHKQHLSIVSLDFNKAFDKIGLHAITKQ LQEWNIGPKILNYIINFMSKRKIKCRVA NNLSSTFPLHNGIPQGSPISVIIFLIAYNK LANLITIQRNLTFTAYADDFHIIVKFKKN			
--	--	--	----------------	---	--	--	--	--	--

						QTQENFNLDLDFQKIDEWCSFSGSSLSF DKCKHLHICRKHNCCTITTRNLNIPSV DHLKILGLTFDSKYRWTRHIEDLASSLSN KINIIKCLSSNKFNCNHTLISITKAIIVSKI DYGLVFYGNSSITWLKKIQTILNCAVRIA LGAFRTTPITNLLMEANINTVADRRDFL QAKLIKNLYFPNNSPLHKISKSLQSRIHF NNKIPSAITYSLQHCKKQNISLLIKPEIKT KLPFWKFQPNTINTELEKLDKNNTHDSI YQQHFLTICKNQFNDHTFLFTDGSKTIST TSFAITTEKNTLKVGLLQNFSSIYSAEIIAI FEAIQIVKSMRGKYMICSDSLSSLQAIG NPDNFNYYESKIRSLLMQLHPKIKLLWI PGHKNIKGNELADQKAKSASTMPLTYT HNNNISDINKHLTKMLKDKLIKLEQTS NWYQFINPNQITADKFTKRLDQNTTSRI KQTKIIRLRLGHTRITHQHYLDPTKINTC THCLNNVPVTIEHLLKDCPSSQLTDIRFK IFANNNPLTFLSNPTKTEIDLIKYLKETNL YPLI			
--	--	--	--	--	--	---	--	--	--

	LOA- 2_DSe r	GCA_00804 2835.1	Droso phila serrata	TGGCTGGCAAAA ATCCTCCAGCTGA ACCCTCCACGAGG GAGCCGGCTAGG CAATGGGCACTGT ATTACCGCGTACA ATGGTAATGCAGT GGCGCATGCTCTG TCCGGGGTAATG CAGTGTCTAGCTA AGGCTTCGATTCA ATTTCTCCCGCCT AAACGCCATGGTT GGAGGTGCCTGG CGGCAGGCATAA AACGCAAGTGGG TGGTTCCCCGGAA ATAAACCAATTTT TCACAGGATGGA GACAGGAGAGGC AACCGGGGCTGG	TAAGATTCTCCAGGA CCATGTGAAAAACAA CATCCGGATAAATAA GGGCCATACTGGCCT ATGATTGGCCCCCTC AGGAGGGCCGTCCG GTTAACCTAACCTA ACCT	MSLRVLQLNLHKS KVASAELLIAMEQG SADVALVQEPWIASGNAVAGLKSSNYN AFYTPSVSRNRTAVLVRKGLQANLMSH YSSDDLTVVLLESKKKRLLVASCYMAHD RPAPPEELRSLVEPGPKDQQLIVGADA NAHHSVWGSPDINDRGESLLDFILSTNL YIANVGEEHTFVGPTSSNVLDITLVAGQ STMVSDWKVLERPSFS DHRYIQFQCEF DHSPKPTVFRNPRNTNWGKFQDIISRK LATSPGTVKSDEIDIEKSLETLSKELLGAY HASCPIRSKKRTRPPWWNKDSLQRD DLREFFKLAKQAKDDIINEEYKILLRNYK KEIRRAQRSSWRNFCSNIETAPETARLR KLLSKQPTIQSQLKRDDGQWTEGSEEA LTALMHGHFPGCTKITNTHPAISFWIG NLLSCRRVQSDWGTASMVKGAHRGTP QGGVLSPLLWNLVVDLIRRFERKAPKI TAYADDISIITGVCPSTLSSIMESTLREIC EWTDKVGLSINADKTDLILFTKRYKVPP WIPPKINTTRLTPKTQVKYLGTVLDSKLA WKPNVLERVKKATIALYASKKMLSSTW	SSF 566 72	436	801
--	--------------------	---------------------	---------------------------	--	---	---	------------------	-----	-----

				GATGTCAGCCCAA ACGTCGGTGGAA GGCGTGACTGCT ACCACCGGCGGG CACGAAGGGGAG CAATCCTCTCTGC GTGTCACCAGCG GTCACACCTCTGA TTTGACGGGAGC AGGTCAAGTCAG TTGTAGCGCTAAT GCCAAATAACGA CGAAAAGAAGGC TAAGCCGCACGG GTGCCGTTGTCTA TACAGACTCGGAC CTGGAGAGCGTG CAGCTTTCGAAAG AAGAGGAAGAAA GCCTTCTTCGCTC GCCAGAACCAGG		GLSPALMHVVYISVVRPTLLYGVLVW WQAMEKRTRYKKLMERTQRQALLCISG ALRSTPTKALETIIGIDPLDIHAQLTAGKA AQLVALGSMSTKGYGHSSIGRTMSSV SDYMTPTPCPDVKTSTFMGPEDWKTV QDHARYLNIYTDGSKMEGGVGAGIYCT DPETRLSYKLPSHCSIFQAEVFAIRKAAE LAQEIARPCAEVNLVDSQAAIRSMQSS AVSSKNVLASRDALDILSTTTSVRIYWVP SHQGIEGNEIADELAKEGVGLAIDRTEN VPVSLRTLQSELERRADVRAKSRWRGN SATCKTSKIMCKERNDKLSHYVLHLSRK DCRLLVGLLTGHCLAAAHAVKLGITNN GNCRKCDESEATETLEHLICKCPALTQA RVKHLGAPVLASLADASKRSPGELLAFA KTTAILKDFNTA			
--	--	--	--	--	--	---	--	--	--

				CACCAGCTCCCCG CGAATAGACGGG CCGAAGCGTTCAA GCGAGGGGAAGA AGGCTAAGGCGT AGTATAAAGCTGC CCTCCGAATCCAG AGCCGGCTCCGG ACAAAAGCTGCCC TAACCGAAGAGG AGAACGTAAAGG TCAACTGACAGG ATCGATGAGAGC AGGCTGCACTTCG CCAAGCTTCCTAG CTTGGCATCAACG GCCGGGTTCGCC AACAAAGTGTAG GTGATGCTTGCCA CCAAGCGGCAAC GCTCGGCCGAGA					
--	--	--	--	---	--	--	--	--	--

				GTGCCATAAAGG ACATTCCGGCGA GCAAGCGGGGAC TCGTGCGCAAGG AAGCAGGCCCCC CCTGGAAGGCGA AAACCAAATCCAG ATTAACAAGAAG GCAGAGGACATC CTGTATCCCAGAC ACGGGAAGATGG CATGGGTAGCGT CTACCTGCGCCTC AAAAAGCGCCAT CCTGGGGATAAG GATGACCACACCC TGAAGGCAGGCG AGATGGAAAAGG ACCTAGGTCTGGA AACCATTTTGGAC GCCGCCCAGGAA					
--	--	--	--	--	--	--	--	--	--

				CTAACGCTGGATG AGTCCGAAGAGG AGGAAGACGGTG ACCTGACGATAGT AGTCAACCCGTCG AGTGGAGTTGAG CCCAGTACCC					
--	--	--	--	---	--	--	--	--	--

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.

Под влиянием успеха ретротранспозонов в природе в данном документе дополнительно обсуждается, что естественная функция ретротранспозона может быть воспроизведена с использованием функциональных частей, полученных из полностью независимых систем. Например, функциональный Gene Writer™ может состоять из неродственных ДНК-связывающих доменов, доменов обратной транскрипции и эндонуклеазных доменов. Эта модульная структура позволяет комбинировать функциональные домены, например, dCas9 (связывание ДНК), обратную транскриптазу MMLV (обратная транскрипция), FokI (эндонуклеаза). В некоторых вариантах осуществления несколько функциональных доменов могут происходить из одного белка, например, никазы Cas9 (связывание ДНК, эндонуклеаза), ретротранспозона R2 (связывание ДНК, обратная транскрипция, эндонуклеаза).

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 размером 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 размером 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 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.

9. Полипептидный компонент системы редактирования генов Gene Writer™

Домены и функции

В некоторых вариантах осуществления полипептид Gene Writer™ обладает функциями связывания ДНК сайта-мишени, связывания нуклеиновой кислоты-матрицы (например, РНК), расщепления ДНК сайта-мишени и записи в нуклеиновой кислоте-матрице (например, РНК), например, обратной транскрипции. В некоторых вариантах осуществления каждая функция содержится в отдельном домене. В некоторых вариантах осуществления функция может быть отнесена к двум или нескольким доменам (например, два или несколько доменов вместе демонстрируют функциональное свойство). В некоторых вариантах осуществления два или несколько доменов могут иметь одинаковую или сходную функцию (например, каждый из двух или нескольких доменов независимо обладает функциональным свойством связывания с ДНК, например, для двух разных последовательностей ДНК). В других вариантах осуществления один или несколько доменов могут быть способны выполнять одну или несколько функций, например, домен Cas9 обеспечивает как связывание ДНК, так и расщепление сайта-мишени. В некоторых вариантах осуществления все домены расположены в пределах одного полипептида. В некоторых вариантах осуществления первый домен находится в одном полипептиде, а второй домен находится во втором полипептиде. Например, в некоторых вариантах осуществления полипептид Gene Writer™ может быть разделен между первым полипептидом и вторым полипептидом, например, где первый полипептид содержит домен обратной транскриптазы (RT), и где второй полипептид содержит ДНК-связывающий домен и эндонуклеазный домен, например, нисказный домен. В качестве дополнительного примера, в некоторых вариантах осуществления изобретения каждый из первого полипептида и второго полипептида содержит ДНК-связывающий домен (например, первый ДНК-связывающий домен и второй ДНК-связывающий домен). В некоторых вариантах осуществления первый и второй полипептиды могут быть объединены посттрансляционно посредством расщепленного интеина.

10. *Записывающий домен*

В определенных аспектах настоящего изобретения записывающий домен системы Gene Writer™ обладает активностью обратной транскриптазы и также называется доменом обратной транскриптазы (доменом RT). В некоторых вариантах осуществления домен RT

содержит каталитическую часть RT и РНК-связывающую область (например, область, которая связывает РНК-матрицу).

В определенных аспектах настоящего изобретения в основе записывающего домена лежит домен обратной транскриптазы не содержащего LTR ретротранспозона APE-типа или RLE-типа. Домен обратной транскриптазы дикого типа не содержащего LTR ретротранспозона APE-типа или RLE-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, ретротранспозона, содержащего LTR, или не содержащего LTR ретротранспозона. В определенных 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 или не содержащего LTR ретротранспозона APE-типа из клады R1 или Tx1. В определенных вариантах осуществления система Gene Writer™ содержит полипептид, который содержит домен обратной транскриптазы ретротранспозона, не содержащего LTR, ретротранспозона, содержащего LTR, интрон группы II, элемент, генерирующий разнообразие, ретрон, теломеразу, ретроплазмиду, ретровирус или сконструированную полимеразу, указанные в таблице 1 или таблице 3. В некоторых вариантах осуществления система Gene Writer™ содержит полипептид, содержащий домен обратной транскриптазы, указанный в таблице 2. В вариантах осуществления аминокислотная последовательность домена обратной транскриптазы системы Gene Writer™ характеризуется по меньшей мере приблизительно 50%, по меньшей мере приблизительно 60%, по меньшей мере приблизительно 70%, по меньшей мере приблизительно 80%, по меньшей мере приблизительно 85%, по меньшей мере приблизительно 90%, по меньшей мере приблизительно 95%, по меньшей мере приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98%, по меньшей мере приблизительно 99% идентичностью с аминокислотной последовательностью домена обратной транскриптазы ретротранспозона, не содержащего LTR, ретротранспозона, содержащего LTR, интрона группы II, элемента, генерирующего разнообразие, ретрона, теломеразы, ретроплазмиды, ретровируса или сконструированной полимеразы, последовательность ДНК которой указана в таблице 1 или таблице 3, или пептида, содержащего домен RT, указанный в таблице 2. В некоторых вариантах осуществления домен RT имеет последовательность, выбранную из таблицы 1 или 3, или последовательность пептида, содержащую домен RT, выбранный из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85% 90%,

95%, 96%, 97%, 98% или 99% идентичностью с ними. В некоторых вариантах осуществления домен RT, содержащийся в полипептиде Gene Writer, был подвергнут мутации по сравнению с его исходной аминокислотной последовательностью, например, имеет по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен. В некоторых вариантах осуществления домен RT получен из RT ретровируса, например, RT HIV-1, RT вируса лейкоза мышей Молони (MMLV), RT вируса миелобластома птиц (AMV), RT вируса саркомы Рауса (RSV). В некоторых вариантах осуществления домен RT получен из RT интрона группы II, например, RT матуразы интрона группы II из *Eubacterium rectale* (MarathonRT) (Zhao et al. PNHK 24:2 2018), домен RT из LtrA, RT TGIRT (или trt). В некоторых вариантах осуществления домен RT получен из RT ретрона, например, обратная транскриптаза из Ec86 (RT86). В некоторых вариантах осуществления домен RT получен из ретроэлемента, генерирующего разнообразие, например, из RT Brt. В некоторых вариантах осуществления домен RT получен из RT ретроплазмиды, например, RT из плазмиды Mauriceville. В некоторых вариантах осуществления домен RT получен из ретротранспозона, не содержащего LTR, например, RT из R2Bm, RT из R2Tg, RT из LINE-1, RT из Penelope или Penelope-подобного элемента (PLE). В некоторых вариантах осуществления домен RT получен из ретротранспозона, содержащего LTR, например, обратная транскриптаза из Ty1. В некоторых вариантах осуществления домен RT получен из теломеразы, например, TERT. Средний специалист в данной области способен идентифицировать домены обратной транскрипции 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 обратная транскриптаза содержит домен InterPro IPR000477. В некоторых вариантах осуществления обратная транскриптаза содержит домен Pfam PF00078. В некоторых вариантах осуществления RT содержит домен InterPro IPR013103. В некоторых вариантах осуществления RT содержит домен Pfam PF07727. В некоторых вариантах осуществления обратная транскриптаза содержит консервативный белковый домен семейства cd00304 RT_подобные, например, cd01644 (RT_perA17), cd01645 (RT_Rtv), cd01646 (RT_бактериальный_ретрон_I), cd01647 (RT_LTR), cd01648 (TERT), cd01650 (RT_nLTR_подобные), cd01651 (RT_интрон_G2), cd01699 (PNHK_зависимые_RNAP), cd01709 (RT_подобные_1), cd03487 (RT_бактериальный_ретрон_II), cd03714 (RT_DIRS1), cd03715 (RT_ZFREV_подобные). Белки, содержащие такие домены, можно дополнительно найти с помощью поиска доменов в базах данных белков, таких как InterPro (Mitchell et al. Nucleic Acids Res 47, D351-360 (2019)), UniProt (The UniProt Consortium Nucleic Acids Res 47, D506-515 (2019)) или в базе данных консервативных доменов (Lu et al. Nucleic Acids Res 48, D265-268 (2020)) или с помощью сканирования открытых рамок считывания на наличие доменов обратной транскриптазы с использованием инструментов предсказания, например, InterProScan. Разнообразие обратных транскриптаз 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.

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 иницирует 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 иницирует 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 иницирует 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 Мейсона-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 пенестости человека (HFV) (например, UniProt P14350), вируса пенестости обезьян (SFV) (например, UniProt P23074) или вируса пенестости/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 миелобластома 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 интегральный домен, например, где интегральный домен может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит интегральный 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 Н, например, где домен РНКазы Н может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы Н, например, эндогенный домен РНКазы Н или гетерологичный домен РНКазы Н. В некоторых вариантах осуществления в домене RT (например, описанном в данном документе) отсутствует домен РНКазы Н. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы Н, который был добавлен, подвергнут делеции, мутации или заменен гетерологичным доменом РНКазы Н. В некоторых вариантах осуществления в результате мутации домена РНКазы Н 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 Н 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 (SEQ ID NO: 1539) или YMDD (SEQ ID NO: 1540) в домене RT (например, в обратной транскриптазе) заменен на YVDD (SEQ ID NO: 1541). В вариантах осуществления замена YADD (SEQ ID NO: 1539), или YMDD (SEQ ID NO: 1540), или YVDD (SEQ ID NO: 1541) 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).

В некоторых вариантах осуществления домен обратной транскриптазы выбран из элемента из таблицы 1 или таблицы 3.

11. Таблица 1. Иллюстративные домены обратной транскриптазы из разных типов источников

Источники включают интрон группы II, ретротранспозон, не содержащий LTR, ретровирус, ретротранспозон, содержащий LTR, 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
-------	-----	------------------	---------	--------------------	-----------------

MarathonRT	Интер группы II	CBK92290.1	D4JMT6	MDTSNLMEQILSSDNLNRAY LQVVRNKGAEGVDGMKYTE LKEHLAKNGETIKGQLRTRK YKPQPARRVEIPKPDGGVRNL GVPTVTDRFIQQAIAQVLTPIY EEQFHDHSYGFRPNRCAQQAI LTALNIMNDGNDWIVDIDLEK FFDTVNHDKLMTLIGRTIKDG DVISIVRKYLVSGIMIDDEYED SIVGTPQGGNLSPLLANIMLN ELDKEMEKRGLNFVRYADDC IIMVGSEMSANRVMRNISRFIE EKLGLKVNMTKSKVDRPSGL KYLGFYFDPRAHQFKAKP HAKSVAKFKKRMKELTCRSW GVSNSYKVEKLNQLIRGWIN YFKIGSMKTLCKELDSRIRYR LRMCIWKQWKTQPNQEKNL VKLGIDRNTARRVAYTGKRIA YVCNKGAVNVAISNKRLASF GLISMLDYYIEKCVTC (SEQ ID NO: 1542)	IPR00047 7, PF00078, cd01651
TGIRT, trt	Интер группы II	AAT72329.1	Q6DKY 2	MALLERILADRNLITALKRVE ANQGAPGIGDVSTDQLRDIYR AHWSTIRAQLLAGTYRPAPV RRVGIPKGPGGTRQLGITPVV DRLIQQIALQELTPIFDPDFSPS SFGFRPGRNAHDAVRQAQGY IQEYGRYVVDMDLKEFFDRV NHDLIMSRVARKVDKKRVLK LIRYALQAGVMIEGVKVQTEE GTQPGGPLSPLLANILLDDLD KELEKRGLKFCYRADDNCNIY VSKLRAGQRVKQSIQRFLEKT LKLKVNEEKSVADRPWKRAF	IPR00047 7, PF00078, cd01651

				GLFSFTPERKARIRLAPRSIQR LKQRIRQLTNPNEWSISMPREIH RVNQYVGMWIGYFRLVTEPS VLQTIEGWIRRLRLCWQLQ WKRVRTRIRELRALGLKETA VMEIANRTKGAWRTTKPQTL HQALGKYTWTAAQGLKTSLQR YFELRQG (SEQ ID NO: 1543)	
LtrA	Интер он групп ы II	AAB06503.1	P0A3U0	MKPTMAILERISKNSQENIDE VFTRLRYLLRPDIYYVAYQN LYSNKGASTKGILDDTADGFS EEKIKKIIQSLKDGTYYPQPVR RMYIAKKNSKKMRPLGIPTFT DKLIQEAVRIILESIYEPVFEDV SHGFRPQRSCHTALKTIKREF GGARWFVEGDIKGCFDNIDH VTLIGLINLKIKDMKMSQLIY KFLKAGYLENWQYHKTYSGT PQGGILSPLLANIYLHELDKFV LQLKMKFDRESPERITPEYRE LHNEIKRISHRLKKLEGEEKA KVLLEYQEKRRLPTLPCTSQ TNKVLKYVRYADDFIISVKGS KEDCQWIKQLKLFHNLK MELSEEKTLITHSSQPARFLG YDIRVRRSGTIKRSKVKKRT LNGSVELLIPLQDKIRQFIFDK KIAIQKKDSSWFPVHRKYLIR STDLEIITIYNSELRGICNYYGL ASNFNQLNYFAYLMEYSCLK TIASKHKGTLSTISMFKDGS GSWGIPYEIKQGKQRRYFANF SECKSPYQFTDEISQAPVLYG YARNTLENRLKAKCCELCGT SDENTSYEIHHVNVKNLKG	IPR00047 7, PF00078, cd01651

				KEKWEMAMIAKQRKTLVVC FHCHRHVIHKHK (SEQ ID NO: 1544)	
R2Bm	Ретро транс позон , не содер жащи й LTR	AAB59214.1	V9H052	MMASTALSLMGRCNPDGCTR GKHVTAAPMDGPRGPSSLAG TFGWGLAIPAGEPCGRVCSPA TVGFFPVAKKSNKENRPEASG LPLESERTGDNPTVRGSAGAD PVGQDAPGWTCQFCERTFST NRGLGVHKRRRAHPVETNTDA APMMVKRRWHGEEIDLLART EARLLAERGQCSSGDLFGALP GFGRTLAIKGQRRREPYRAL VQAHLARFGSQPGPSSGGCSA EPDFRRASGAEEAGEERCAED AAAYDPSAVGQMSPDAARVL SELLEGAGRRRACRAMRPKT AGRRNDLHDDRTASAHKTSR QKRRAEYARVQELYKKCRSR AAAEVIDGACGGVGHSLEEM ETYWRPILERVSDAPGPTPEA LHALGRAEWHGGNRDYTQL WKPISVEEIKASRFDWRTSPG PDGIRSGQWRAVPVHLKAEM FNAWMARGEIPEILRQCRTVF VPKVERPGGPGEYRPISIASIPL RHFHSILARRLLACPPDARQ RGFICADGTLENSAVLDAVLG DSRKKLRECHVAVLDFAKAF DTVSHEALVELLRLRGMPEQF CGYIAHLYDTASTTLAVNNE MSSPVKVGRGVRQGDPLSPIL FNVVMDLILASLPERVGYRLE MELVSALAYADDLVLLAGSK VGMQESISAVDCVGRQMGLR	IPR00047 7, PF00078, cd01650

				LNCRKSAVLSMIPDGHRKKH HYLTERTFNIGGKPLRQVSCV ERWRYLGVD FEASGCVTLEH SISSALNNISRAPLKPQQRLEIL RAHLIPRFQHG FVLGNISDDR LRMLDVQIRKAVGQWLRLPA DVPKAYYHAAVQDGGLAIPS VRATIPDLIVRRFGGLDSSPWS VARAAKSDKIRKKLRWAW KQLRRFSRVDSTTQRPSVRLF WREHLHASVDGRELRESTR PTSTKWIRERCAQITGRDFVQ FVHTHINALPSRIRGSRGRRG GGESLTCRAGCKVRETTAHI LQQCHRTHGGRILRH NKIVSF VAKAMEENKWTVELEPRLRT SVGLRKPDIIASRDGVGVIVD VQVVSQGQRSLDELHREKRNK YGNHGELVELVAGRLGLPKA ECVRATSCTISWRGVWSLTSY KELRSIIGLREPTLQIVPILALR GSHMNWTRFNQMTSVMGGG VG (SEQ ID NO: 1545)	
LINE-1	Ретро транс позон , не содер жащи й LTR	AAC51271.1	O00370	MTGSNSHITILTLNVNGLNSPI KRHRLASWIKSQDPSVCCIQE THLTCRDTHRLKIKGWRKIYQ ANGKQKKAGVAILVSDKTDF KPTKIKRDKEGHYIMVKGSIQ QEELTILNIYAPNTGAPRFIKQ VLSDLQRDLDSHTLIMGDFNT PLSILDRSTRQKVNKDTQELN SALHQTDLDIYRTLHPKSTEY TFFSAPHHTYSKIDHIVGSKAL LSKCKRTEITNYLSDHSAIKL ELRIKNLTQSRSTTWKLNNLL	IPR00047 7, PF00078, cd01650

				LNDYWVHNEMKAEIKMFFET NENKDDTTYQNLWDAFKAVC RGKFIALNAYKRRKQERSKIDT LTSQLELEKQEQTTHSKASRR QEITKIRAEELKEIETQKTLQKI NESRSWFFERINKIDRPLARLI KKKREKNQIDTIKNDKGDITT DPTEIQTIREYYKHL YANKL ENLEEMDTFLDTYTLPRLNQE EVESLNRPI TGSEIVAIINSLPT KKSPGPDGFTA EFYQRYKEEL VPFLLKLFQSIEKEGILPNSFY EASILIPKPRDRTTKKENFRPI SLMNIDAKILNKILANRIQQHI KKLIHHDQVGFIPGMQGWFN RKSINVIQHINRAKDKNHVIISI DAEKAFDKIQQPFMLKTLNK LGIDGMYLKIIRAIYDKPTANI ILNGQKLEAFPLKTGTRQGCP LSPLLFNIVLEV LARAIRQEKE IKGIQLGKEEVKLSLFADDMI VYLENPIVSAQNLLKLISNFSK VSGYKINVQKSQAFLYNNNR QTESQIMGELPFTIASKRIKYL GIQLTRDVKDLFKENYKPLLK EIKEDTNKWKNIPCSWVGRIN IVKMAILPKVIYRFNAIPIKLP MTFFTELEKTTLKFIWNQKRA RIAKSILSQKNKAGGITLPDFK LYYKATVTKTAWYWYQNRD IDQWNRTEPSEIMPHIYNYLIF DKPEKNKQWGKDSLLNKWC WENWLAICRKLKLPFLTPYT KINSRWIKDLNVKPKTIKTLE ENLGITIQDIGVGKDFMSKTP	
--	--	--	--	--	--

				KAMATKDKIDKWDLIKLKSF CTAKETTIRVNRQPTTWEKIF ATYSSDKGLISRIYNELKQIYK KKTNNPIKKWAKDMNRHFSK EDIYAAKKHMKKCSSSLAIRE MQIKTTMRYHLTPVRMAIHKK SGNNRCWRGCGEIGTLVHCW WDCKLVQPLWKSVMRFLRD LELEIPFDPAIPLLGIIYPKDYKS CCYKDTCTRMFIAALFTIAKT WNQPNCPMTIDWIKKMWHIY TMEYYAAIKNDEFISFVGW MKLETILSKLSQEQKTKHRIF SLIGGN (SEQ ID NO: 1546)	
Penelope	Ретро транс позон , не содер жащи й LTR	AAL14979.1	Q95VB5	MERSPEPSININGRHAVCTAT NMSYAKIKTKYKDSKRTINKF QLTLVKLTKLKSSLKFLKCR KSNLIPNFIKNLTQHLTILTTD NKTHPDITRTLTRHTHFYHTK ILNLLIKHKHNLLQEQTKHMQ KAKTNIEQLMTTDDAKAFFES ERNIENKITTTLKKRQETKHD KLRDQRNLALADNNTQREWF VNKTKIEFPPNVVALLAKGPK FALPISKRDFPLLKYIADGEEL VQTIKEKETQESARTKFSLLV KEHKTKNNQNSRDRAILDTV EQTRKLLKENINIKLSSDKGN KTVAMDEDEYKNKMTNILDD LCAYRTLRLDPTSRLQTKNNT FVAQLFKMGLISKDERNKMT TTTAVPPRIYGLPKIHKEGTPL RPICSSIGSPSYGLCKYIIQILK NLTMDSRYNIKNAVDFKDRV NNSQIREETLVSFVVSFLFPS	IPR00047 7, PF00078, cd00304

				IPIELALDTIRQKWTKLEEHTN IPKQLFMDIVRFCIEENRYFKY EDKIYTQLKGMPMGSPASPVI ADILMEELLDKITDKLKIKPRL LTKYVDDLFAITNKIDVENIL KELNSFHKQIKFTMELEKDGK LPFLDSIVSRMDNTLKIKWYR KPIASGRILNFNSNHPKSMIIN TALGCMNRMMKISDTIYHKEI EHEIKELLTKNDFPPNIIKTLL KRRQIERKKPTEPAKIYKSLIY VPRLSERLTNSDCYNKQDIKV AHKPTNTLQKFFNLIKSKIPMI EKSINVYQIPCGGDNNNKCEN SVYIGTTKSKLKTRISQHKSDF KLRHQNNIQKTALMTHCIRSN HTPNFDETTILQQEQHYNKRH TLEMLHIINTPTYKRLNYKTD TENCAHLYRHLLNSQTTSVTI STSKSADV (SEQ ID NO: 1547)	
RT M- MLV	Ретро вирус	ADS42990.1	P03355[660- 1330]	TLNIEDEHRLHETSKEPDVSL GSTWLSDFPQAWAETGGMGL AVRQAPLIPLKATSTPVSIIQ YPMSQEARLGIKPHIQRLDQ GILVPCQSPWNTPLLPVKKPG TNDYRPVQDLREVNKRVEDI HPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPL FAFEWRDPEMGISGQLTWTR LPQGFKNSPTLFDEALHRDLA DFRIQHPDLILLQYVDDLILA ATSELDCQQGTRALLQTLGN LGYRASAKKAQICQKQVKYL GYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAFC	IPR00047 7, PF00078, cd03715

				RLWIPGFAEMAAPLYPLTKTG TLFNWGPDQQKAYQEIKQAL LTAPALGLPDLTKPFELFVDE KQGYAKGVLTQKLGPWRRP VAYLSKKLDPVAAGWPPCLR MVAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRW LSNARMTHYQALLDTRVQ FGPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQ RKAGAAVTTETEVIWAKALP AGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHI HGEIYRRRGLLTSEGKEIKNK DEILALLKALFLPKRLSIHCP GHQKGHSAEARGNRMADQA ARKAAITETPDTSTLL (SEQ ID NO: 1548)	
RT RSV	Ретро вирус	AAC82561.1	P03354[709- 1567]	TVALHLAIPLKWKPDHTPVWI DQWPLPEGKLVALTQLVEKE LQLGHIEPSLSCWNTPVFVIRK ASGSYRLLHDLRAVNAKLVP FGAVQQGAPVLSALPRGWPL MVLDLKDCFFSIPLAEQDREA FAFTLPSVNNQAPARRFQWK VLPQGMTCSPTICQLVVGQVL EPLRLKHPSLCMLHYMDDL LAASSHDGLEAAGEEVISTLE RAGFTISPDKVQREPGVQYLG YKLGSTYVAPVGLVAEPRIAT LWDVQKLVGSLQWLRPALGI PPRLMGPFYEQLRGSDPNEAR EWNLDMKMAWREIVRLSTT AALERWDPALPLEGAVARCE	IPR00047 7, PF00078, cd01645

				QGAIGVLGQGLSTHPRPCLWL FSTQPTKAFTAWLEVLTLLIT KLRASAVRTFGKEVDILLPA CFREDLPLPEGILLALKGFAG KIRSSDTPSIFDIARPLHVSLKV RVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADL GASVQQLEARAVAMALLW PTTPTNVVTDSAFVAKMLLK MGQEGVPSTAAAFILEDALSQ RSAMAAVLHVRSSEVPGFF TEGNDVADSQATFQAYPLRE AKDLHTALHIGPRALSKACNI SMQQAREVVQTCPHCNSAPA LEAGVNPRGLGPLQIWQTDFT LEPRMAPRSWLAVTVDTASS AIVVTQHGRVTSVAVQHWWA TAIAVLGRPKAIKTDNGSCFT SKSTREWLARWGIAHTTGIPG NSQQQAMVERANRLKDRIR VLAEGDGFMKRIPTSKQGELL AKAMYALNHFERGENTKTPI QKHWRPTVLTEGPPVKIRIET GEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPD ITQKDEVTKKDEASPLFAG (SEQ ID NO: 1549)	
RT AMV	Ретро вирус	HW606680.1	-	TVALHLAIPKWKPNHTPVWI DQWPLPEGKLVALTQLVEKE LQLGHIEPSLSCWNTPVFVIRK ASGSYRLLHDLRAVNAKLVP FGAVQQGAPVLSALPRGWPL MVLDLKDCFFSIPLAEQDREA FAFTLPSVNNQAPARRFQWK VLPQGMTCSPTICQLIVGQILE	IPR00047 7, PF00078, cd01645

				PLRLKHPSLRMLHYMDDL AASSHDGLEAAGEEVISTLER AGFTISPDKVQREPGVQYLG YKLGSTYVAPVGLVAEPRIATL WDVQKLVGSLQWLRPALGIP PRLMGPFYEQLRGSDPNEARE WNLDMKMAWREIVQLSTTA ALERWDPALPLEGAVARCEQ GAIGVLGQGLSTHPRPCLWLF STQPTKAFTAWLEVLTLITK LRASAVRTFGKEVDILLPAC FREDLPLPEGILLALRGFAGKI RSSDTPSIFDIARPLHVSLKVR VTDHPVPGPTVFTDASSSTHK GVVVWREGPRWEIKEIADLG ASVQQLEARAVAMALLWPT TPTNVVTDSAFVAKMLLKMG QEGVPSTAAAFILEDALSQRS AMAAVLHVRSHSEVPGFFTE GNDVADSQATFQAY (SEQ ID NO: 1550)	
RT HIV	Ретро вирус	AAB50259.1	P04585[588- 1147]	PISPIETVPVKLKPGMDGPKV KQWPLTEEKIKALVEICTEME KEGKISKIGPENPYNTPVFAIK KKDSTKWRKLVDFRELNKRT QDFWEVQLGIPHPAGLKKKK SVTVLDVGDAYFSVPLDEDFR KYTAFTIPSINNETPGIRYQYN VLPQGWKGSPAIFQSSMTKIL EPFRKQNPDIVIYQYMDDLIV GSDLEIGQHRTKIEELRQHLL RWGLTTPDKKHQKEPPFLWM GYELHPDKWTVQPIVLPEKDS WTVNDIQKLVGKLNWASQIY PGIKVRQLCKLLRGTKALTEV	IPR00047 7, PF00078, cd01645

				IPLTEEALELAENREILKEPV HGVYYDPSKDLIAEIQKQGQG QWTYQIYQEPFKNLKTGKYA RMRGAHTNDVKQLTEAVQKI TTESIVIWGKTPKFKLPIQKET WETWWTEYWQATWIPEWEF VNTPLVLWYQLEKEPIVGA ETFYVDGAANRETKLGKAGY VTNRGRQKVVTLTDTTNQKT ELQAIYLALQDSGLEVNIVTD SQYALGIIQAQPDQSESELVN QIIQLIKKEKVYLAWVPAHK GIGGNEQVDKLVSAGIRKVL (SEQ ID NO: 1551)	
Ty1	Ретро транс позон , содер жащи й LTR	AAA66938.1	Q07163- 1[1218- 1755]	AVKAVKSIKPIRTTLRYDEAIT YNKDIKEKEKYIEAYHKEVN QLLKMKTWDTDEYYDRKEID PKRVINSMFIFNKKRDGTHKA RFVARGDIQHPDTYDSGMQS NTVHHYALMTSLSLALDNNY YITQLDISSAYLYADIKEELYI RPPPHLGMNDKLIRLKKSLYG LKQSGANWYETIKSYLIQCGG MEEVRGWSCVFKNSQVTICL FVDDMVLF SKNLNSNKRIIEK LKMQYDTKIINLGESEDEIQY DILGLEIKYQRGKYMKLGME NSLTEKIPKLVPLNPKGRKL SAPGQPGLYIDQDELEIDEDE YKEKVHEMQKLIGLASVVG YKFRFDLLYYINTLAQHILFPSR QVLDMTYELIQFMWDTRDKQ LIWHKNKPTDPDNKLVAISDA SYGNQPYKYSQIGNIYLLNGK VIGGKSTKASLTCTSTTEAEIH	IPR01310 3, PF07727

				AISESVPLNLSYLIQELNKK PIIKGLLTDSRSTISIIKSTNEEK FRNRFFGTKAMRLRDEVSGN NLYVYYIETKKNIADVMTKPL PIKTFKLLTNKWIH (SEQ ID NO: 1552)	
Brt	Ретро элемент, генерирующий разнообразие	NP_958675.1	Q775D8	MGKRHRNLIDQITTWENLLD AYRKTSHGKRRTWGYLEFKE YDLANLLALQAEKAGNYER GPYREFLVYEPKPRILISALEFK DRLVQHALCNIVAPIFEAGLL PYTYACRPDKGTHAGVCHVQ AELRRTRATHFLKSDFSKEFFPS IDRAALYAMIDKKIHCAATTR LLRVVLPDEGVGPIGSLTSQ FANVYGGAVDRLHDELKQR HWARYMDDIVVLGDDPEELR AVFYRLRDFASERLGLKISHW QVAPVSRGINFLGYRIWPTHK LLRKSSVKRAKRKVANFIKH GEDESLQRFLASWSGHAQWA DTHNLFTWMEEQYGIACH (SEQ ID NO: 1553)	IPR00047 7, PF00078, cd01646
RT86	Ретро н	AAA61471.1	P23070	MKSAEYLNTFRLRNLGLPVM NNLHDMSKATRISVETLRLLI YTADFRYRIYTVKKGPEKR MRTIYQPSRELKALQGWLRL NILDKLSSSPFSIGFEKHQSILN NATPHIGANFILNIDLEDFPS LTANKVFGVFHSLGYNRLISS VLTKICCYKNLLPQGAPSSPK LANLICKLDYRIQGYAGSRG LIYTRYADDLTLAQSMKKV VKARDFLFSIIPSEGLVINSKK TCISGPRSQRKVTGLVISQEKV	IPR00047 7, PF00078, cd03487

				GIGREKYKEIRAKIHHIFCGKS SEIEHVRGWLSFILSVDSKSHR RLITYISKLEKKYGKNPLNKA KT (SEQ ID NO: 1554)	
TERT	Тело мераз а	AAG23289.1	O14746	MPRAPRCRAVRSLLRSHYRE VLPLATFVRRLGPQGWRLVQ RGDPAAFRALVAQCLVCVPW DARPPPAAPSRQVSCCLKELV ARVLQRLCERGAKNVLAFGF ALLDGARGGPPEAFTTSVRSY LPNTVTDALRGSGAWGLLLR RVGDDVLVHLLARCALFVLV APSCAYQVCGPPLYQLGAAT QARPPPHASGPRRRLGCERA WNHSVREAGVPLGLPAPGAR RRGGSASRSLPLPKRPRRGAA PEPERTPVGQGSWAHPGRTR GPSDRGFCVVSPARPAEEATS LEGALSGTRHSHPSVGRQHH AGPPSTSRPPRPWDTPCPPVY AETKHFLYSSGDKEQLRPSFL LSSLRPSLTGARRLVETIFLGS RPWMPGTPRRLPRLPQRYWQ MRPLFLELLGNHAQCPYGVL LKTHCPLRAAVTPAAGVCAR EKPQGSVAAPEEEDTDPRRLV QLLRQHSSPWQVYGFVRACL RRLVPPGLWGSRHNERRFLR NTKKFISLGKHAKLSLQELTW KMSVRDCAWLRRSPGVGCVP AAEHLREEILAKFLHWLMS VYVVELLRSFFYVTETTFQKN RLFFYRKSVWSKLQSIGIRQH LKRVLRELSEAEVRQHREA RPALLTSRLRFIPKPDGLRPIV	IPR00047 7, PF00078, cd01648

				NMDYVVGARTFRREKRAERL TSRVKALFSVLNYERARRPGL LGASVLGLDDIHRAWRTFVL RVRAQDPPPELYFVKVDVTG AYDTIPQDRLTEVIASIIKPQN TYCVRRYAVVQKAAHGHVR KAFKSHVSTLTDLQPYMRQF VAHLQETSPLRDAVVIEQSSS LNEASSGLFDVFLRFMCHHA VRIRGKSYVQCQGIPQGSILST LLCSLCYGDMENKLFAGIRRD GLLLRLVDDFLLVTPHLTHAK TFLRTLVRGVPEYGCVVNLR KTVVNFPEDEALGGTAFVQ MPAHGLFPWCGLLLDTRTLE VQSDYSSYARTSIRASLTFNR GFKAGRNMRRKLFGLRLKC HSLFLDLQVNSLQTVCTNIYK ILLQAYRFHACVLQLPFHQQ VWKNPTFFLRVISDTASLCYSI LKAKNAGMSLGAKGAAGPLP SEAVQWLCHQAFLKLTRHR VTYVPLLGSLRTAQTQLSRKL PGTTLTAAEAAANPALPSDFK TILD (SEQ ID NO: 1555)	
RT Mauriceville	Петро плазма ида	NC_001570. 1	Q36578	MPNHRLPNCVSYLGENHEL WLHGMFGLLKRSNPQTGGIL GWLNTGPNGFVKYMMNLMG HARDKGDKEYWRLGRSLM KNEAFQVQAFNHVCKHWYL DYKPHKIAKLLKEVREMVEIQ PVCIDYKRVYIPKANGKQRPL GVPTVPWRVYLHMWNVLLV WYRIPEQDNQHAYFPKRGVF TAWRALWPKLDSQNIYFDL	cd00304

				KNFFPSVDLAYLKDKLMESGI PQDISEYLTVLNRSLVVLTSE DKIEPHRDVIFNSDGTPNPNL PKDVQGRILKDPDFVEILRRR GFTDIATNGVPQGASTSCGLA TYNVKELFKRYDELIMYADD GILCRQDPSTPDFSVEEAGVV QEPAKSGWIKQNGEFKKSVK FLGLEFIPANIPPLGEDEVKDY PRLRGATRNGSKMELSTELQF LCYLSYKLRIKVLRDLYIQVL GYLPSVPLLRYRSLAEAINELS PKRITIGQFITSSFEFTAWSPL KRMGFFFSSPAGPTILSSIFNN STNLQEPSDSRLLYRKGSWV NIRFAAYLYSKLSEEKHGLVP KFLEKLREINFALDKVDVTEI DSKLSRLMKFSVSAAYDEVG TLALKSLFKFRNSERESIKASF KQLRENGKIAEFSEARRLWFE ILKLIRLDLFNASSLACDDLSS HLQDRRSIKKWGSSDVLULK SQRLMRTNKKQLQLDFEKKK NSLKKKLIKRRAKELRDTFKG KENKEA (SEQ ID NO: 1556)	
RTX	Сконс труир ованн ая поли мераз а	QFN49000.1	-	MILDTDYITEDGKPVIRIFKKE NGEFKIEYDRTFEPYLYALLK DDSAIEEVKKITAERHGTVVT VKRVEKVQKKFLGRPVEVW KLYFTHPQDVPAIMDKIREHP AVIDIYEYDIPFAIRYLIDKGL VPMEGDEELKLLAFDIETLYH EGEEFAEGPILMISYADEEGA RVITWKNVDLPYVDVSTER EMIKRFLRVVKEKDPDVLITY	IPR00613 4, PF00136, cd05536

				NGDNFDFAYLKKRCEKLGINF ALGRDGSEPQIQRMGDRFAV EVKGRIHFDLYPVIRRTINLPT YTLEAVYEAVFGQPKEKVYA EEITTAWETGENLERVARYS MEDAKVTYELGKEFLPMEAQ LSRLIGQSLWDVSRSTGNLV EWFLLRKAYERNELAPNKP EKELARRHQSHEGGYIKEPER GLWENIVYLDFRSLYPSIIITH NVSPDTLNREGCKEYDVAPO VGHFRCKDFPGFIPSLGDLLE ERQKIKKRMKATIDPIERKLL DYRQRAIKILANSLYGYGY ARARWYCKECAESVIAWGRE YLTMTIKEIEEKYGFKVIYS DTFATIPGADAETVKKKAM EFLKYINAKLPGALELEYEGF YKRGLFVTKKKYAVIDEEGKI TTRGLEIVRRDWSEIAKETQA RVLEALLKDGDVEKAVRIVK EVTEKLSKYEVPPEKLVHKQI TRDLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKGS GRIVDRAIPDEFDPTKHKYD AEYYIEKQVLPVERILRAFG YRKEDLRYQKTRQVGLSARL KPKGTLLEGSSHHHHHHH (SEQ ID NO: 1557)	
--	--	--	--	---	--

12. Таблица 2. Описания сигнатур InterPro, присутствующих в обратных транскриптазах, из таблицы 1

Сигна тура	База данных	Краткое название	Описание
cd0030 4	CDD	<i>RT_подоб ные</i>	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]
cd0164 5	CDD	<i>RT_Rtv</i>	RT_Rtv: обратные транскриптазы (RT) ретровирусов (Rtv). RT катализируют превращение однонитевой РНК в двухнитевую вирусную ДНК для интеграции в хромосому 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 дубликации, 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	<i>RT_бактериальный_ретрон_I</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	<i>TERT</i>	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	<i>RT_nLTR_подобные</i>	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 одонитевой РНК в двухнитевую ДНК для интеграции в хромосомы хозяина. RT представляет

			с собой многофункциональный фермент с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы Н). [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	<i>RT_интрон_G2</i>	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	<i>RT_бактериальный_ретрон_II</i>	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 однострессовая (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]
cd0371 5	CDD	<i>RT_ZFRE V_подобн ые</i>	RT_ZFRE 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]
cd0553 6	CDD	<i>POLBc_B 3</i>	Каталитический домен подсемейства В3 ДНК-полимераз В-типа. Белки архей, участвующие в репликации ДНК, являются сходными с белками эукариот. Некоторые представители архей также имеют несколько ДНК-полимераз семейства В (В1,

			В2 и В3). До сих пор не была(были) установлена(установлены) какая-либо(какие-либо) конкретная(конкретные) функция(функции) для различных представителей ДНК-полимераз В-типа архей. Филогенетический анализ ДНК-полимераз семейства В эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства В архей и эукариот. Структурное сравнение термостабильной ДНК-полимеразы В-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	<i>ДНК_полимеразы В_Kod1_подобные_экзо</i>	3'-5'-экзонуклеазный домен ДНК-полимераз семейства В 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 В. Их аминоконцы 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 В содержит бета-шпильчатую 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 В. Филогенетический анализ ДНК-полимераз семейства В эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства В 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	<i>RVT_1</i>	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	<i>ДНК_полимераза_В</i>	Данная область ДНК-полимеразы семейства В, 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	<i>RVT_2</i>	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	<i>RT_домен</i>	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 Н 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_В_с_несколькими	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

		ми_домен ами	(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, А и В, со сходными способами элонгации цепи. Ряд ДНК-полимераз был сгруппирован под обозначением ДНК-полимераз семейства В. Шесть областей сходства (пронумерованных от I до VI) обнаружены во всех или в подмножестве полимераз семейства В. 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 В содержат характерный мотив DTDS (SEQ ID NO: 1558) 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 В, по-видимому, характеризуется более чем одной формой активности, возможно, включающей элонгацию, связывание ДНК и связывание dNTP [PMID: 9757117].
IPR013 103	InterPro	<i>RVT_2</i>	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 3** (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 последовательностями, 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.

13. Таблица 3. Иллюстративные последовательности ретротранспозонов

Семейство	Элемент	Номер доступа	Организм	Предсказанная 5'-UTR	Предсказанная 3'-UTR	Предсказанная аминокислотная последовательность
R2	R2-1_TG	.	Taeniopygia guttata	GTCTAGTTACAAC TGGGCATCGCTGC AGAGATCGCACCT CCTCGTGGTCCCG CTGGTAGCCCTTC GAAGGGTGACTAA GTCGATCTCTGCC CCAGGTACGGAGC CGTTGGGACTCAC CAGTCCAACGTAA CTCCTGCCTAAAT TCGGTGAAACAAA TTCCTCGGTAAAA AGCCCC (SEQ ID NO: 1140)	TTCAGGTTATTTAGATG CTTAGTTTTTTGTACCTT TCTTGTTTTGTAGGA TTTTGATAGTGTTAGTA TTTTTATATTTTGTAC GATTGCATAATGTTCTT TTTTATACAGTTCTGTT TTAATAAAAATAGACGA TAGCTAGAGACGTTAG GGCAGCCACAAGCCAG TTAGGTAGCGGATAGT AGGTAGGAACAGACTT TACTATTTCATAACGC GTCAATTACCACCTGAT TTGGACCAATTCACGG GATTTGTCCAAGGTGG ACGGGCCACCTTTACTT AACCCGGAAAAGGAAC	MASCPKPGPPVSAGAMSLESGLTTHSVLAIE RGPNSLANS GSDFGGGGLGLPLRLLRVSVG TQTSRSDWVDLVSWSHPGPTSKSQQVDLV LFPKHRVDLLSKNDQVDLVAQFLPSKFPPNL AENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRD SELFVPEEGSSEKESEDAPKTSPTPGKHGLE QTGEEKVMVTPDKNPPCPCGTRVNSVLN LIEHLKVSHGKRGVCFRCAKCGKENSNYHS VVCHFPCRG PETEKAPAGEWICEVCNRDF TTKIGLGQHKRLAHPAVRNQERIVASQPKET SNRGAHKRCWTKEEEELLIRLEAQFEGNKNI NKLIAEHITTKTAKQISDKRRLLSRKPAEEPR EETPGTCHHTRRAAASLRTEPEMSHHAQAED RDNGPGRRPLPGRAAAGGRTMDEIRRHDPDK GNGQQRPTKQKSEEQLQAYYKKTLEERLSA GALNTFPRAFKQVMEGRDIKLVINQTAQDC FGCLEISISQIRTATRDKKDVTREKHPKKPF

				ATATATAATTTATGTGT GTTCGATAAA (SEQ ID NO: 1263)	QKWMKDRAIKKGNYLRFQRLFYLDRGKLA KIILDDIECLSCDIPLSEIYSVFKTRWETTGSF KSLGDFKTYGKADNTAFRELITAKEIEKNVQ EMSKGSAPGPDGITLGDVVKMDPEFSRTMEI FNLWLTTGKIPDMVRGCRTVLIPKSSKPDRL KDINNWRPITIGSILLRLFSRIVTARLSKACPL NPRQRGFIRAAGCSENKLLQTIWSAKREH RPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENM NTNISILETFCNLTGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYL LQFDPWIGIARSGLSTKLDFWLQRIDQAPLK PLQKTDILKTYTIPRLIYIADHSEVKTALLET LDQKIRTAVKEWLHLPPCTCDAILYSTRDG GLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIW EAPPSSEPPNNVSTNSEWEAPTQKDKFPKPC NWRKNEFKKWTKLASQGRGIVNFERDKISN HWIQYYRRIPHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPV TQDARIKRHNYICELLLEEAKKKDWVVFKE
--	--	--	--	--	--

						PHIRDSNKELYKPD LIFVKDARALVVDVTVR YEA AKSSLEEAAAEKVRKYKHLETEVRHLT NAKDVT FVG FPLGARGKWHQDNFKLLTEL GLSKSRQVKMAET FSTVALFSSVDIVHMFAS RARKSMVM (SEQ ID NO: 1016)
R2	R2- 1_Gfo		Geospiza fortis	AGACTTAAGTGAG TTTGGTTACAAC T GGGCATAGCTGCA GAGACCGCGCCTC CTCGCGGCCCCGC TGGTAAGCCCTTA ACAGGGTGACTAA (SEQ ID NO: 1141)	GGTAGATAATCTTTGTA TAGTGGGGGGGGATCT CATGTACCGGGTTTCTT TTATTTGATTTTCAATA AAACAGACGGTAGCTA GGTTCGCAAGGCAGCC ACAAGCCAAAGATAGG TAGGGTGCTCATAGTG AGTAGGGACAGTGCCT TTTGATTCACAACGCGT CAATACCATCTGACAC GGATACCCTTACCGGA CTTGTCATGATCTCCCA GACTTGTCCAAGGTGG ACGGGCCACCTTTACTT AACCCGGA AAAAGGAAC ATATATTAATTATATGT	VGLCPSPGVDGTHQPND SFQNFGETNFSVQ VARLVTRNLAPRSVRGNGFGSGMATHPVPA DESGHESDPFLVGRSCGQPARLTRQSVGTQT SRDDILPSKTTKL TENELDLLVNFSLELYRSD LQGFVQEGIHFSVNREVLEGFPEVYEQPAPQ PAVGDDLNTSLPPDNNICVLEKGSSEAVEDG TPEVAHPVPETQ GKESPNNIVMTLPNKNPP CPCCRVRLH SVLALIEHLKGSHGKKRACFR CVKCGRENFN YHSTVCHIAKCKGPKVEKAP VGEWICEVCGRDFTTKIGLGQHKRLAHPLV RNQERIDASQPKETS NRGAHKRCWTKEEEE MLIKLEVQFEGHRNINKLIAEHLTTKTSKQIS DKRRLLP RKQLTDLSKGVAGQKVLPGLSH QPQLGVVDNGLGGGHLPGGPAAEGRTIEPL GHHLDDKDN GHREIADQHKAGRLQAHYRKK IRKRLSEGMISNFPEVF EQLLDCQEAQPLINQ AAQDCFGCLDSASQIRKALRKQNTQKDQGD QPKRPAQKWMKKRAVKRGHFLRFQKLFHL

					GTTCGGAAAA (SEQ ID NO: 1264)	DRGKLAKIILDDVECLSCDIPPSEIYSVFKAR WETPGQFAGLGDFEINRKANNKAFRDLITA KEILKNVREMTKGSAPGPDGIALGDIRKMDP EYTRTAELFNLWLTSGEIPDMVRGCRTVLIP KSSKPERLKDINNWRPITIGSILLRLFSRIITAR LTKACPLNPRQRSFISAAGCSENKLLQTIIR TAKNEHRPLGVVFDIAKAFTVSHQHIIHV LQRRRVDPHIIGLVKNMYKDISTVITTKKNT YTDKIQIQVGVKQGDPLSPLLFNLAMDPLLC KLEEHGKGFHRGQSKITAMAFADDLVLLSD SWEDMNANIKILETFCDLTGLKTQGQKCHG FYIKPTKDSYTVNNCAAWTINGTPLNMINPG ESEKYLGLQFDPWVGIAKTSLPEKLDFWLE RIDRAPLKPFQKLDILKTYTIPRLTYVADHSE MKAGALEALDRTIRSAVKDWLHLPSSSTCDA ILYTSMKDGGLGVTKLVGLIPSVQARRLHRI AQSPEETMKDFLEKAQMEKMYEKLWVQA GGKRKRMPSSIWEALPEVVPSIDTATTSEWEA PNPKSKYPRPCNWRRKEFKKWTCLIAQGW GIRCFKGDKISNNWIRHYRYIPHRKLLTAIQL RASVYPTREFLARGREDNCVKSCRHCEAAE ESCAHIIGMCPVVRDARIKRHNRICERLMEE AGKRDWTVFQEPHIRDVTKELYKPDILIFVK
--	--	--	--	--	---------------------------------	--

						EGLALVVDVTIRFESTKTTLEEAAAEKVNK YKHLETEVRNLTNAKDVIFMGFPLGARGQ WYNKNFELDLTLGLPRSRQDIIAKTLSTDAL ISSVDIIHMFASRGRRQHA (SEQ ID NO: 1386)
R2	R2-1_ZA	.	Zonotric hia albicollis	CGACTTGAGAAGG TCTGGTTACAAC GGGCATAGCTGCA GAGATCGCGCCTC CTCGTGGCCCCGC TGGTAAGCCCTTA ACAGGGTGACTAA GTCGATCTCTGCC CCAGTCCAGGAGC CGCTGGGTTTCAC CAGCCCAGCGATT CCTTCCAAATTCG GTGA (SEQ ID NO: 1142)	GTAGTCACATTGCACTT TCTGTAACTTGCCTGG GTGTGGGATGTGGGCC TGGGGTGTGGGTATG GGGTATATATGTGGGA TATTCTGGTGGGAATGT CCATTCACTGTATGCCT ATCTTTTTAATAAAAAG ACGGTAGCTAGGTTCG CGAAGCAGCCACAAGC CAATAGCCAGTTAGGT AGCTCATAGTGGGTAG GTGACAGGAACCTTTG ACTCAGAACGCGTCCA TTAACATCTAGAACGG ACCAAACCTTCGGACAT GCACCGATTAACCGGA TTTGTCCAAGGTGGAC GGGCCACCTTTACTTAA	NKFLGKSRVAYCLKPGPPVSDRGKEFGSGL TTHPEPESESGHDPTVPNPGPSLGAGEGAQP LPLLRVSVGTQTCEEDFITSRPTKLPGIESEL GPLVKFSLEVYRSDLKGDVQFEGIHFPDNW GVLEGFPVEYEQ LAPQPNGGDELNHS LPGD REGDVLEKDSSEKEKEAAPEALPSVQRARSE QLPDNIVKVTVPDKNPPCPCCGVRLNSVLAL IEHLKGSHGRRRVCFRCAKCGRENFNHHST VCHYAKCKGPQIERPPVGEWICEVCGRDFT TKIGLGQHKRHHMHAMVRNQRIDASQPKET SNRGAHKRCWTKEEEELLMKLEVQFENHK NINKLIAEQLTTKTAKQISDKRRMLLKKGRG TTGNLETEPGMSHQSAKVKDNLGGDHL PGGPVVDKGTIGKPGQHLDTDNSHQITAGK KKGGGLQARYRRRIMKRLAAGTINIFPKVF KELINDQEARPLINQTTEDCFGLLDSACQIRT ALREKGSQEERPRKQYQKWMKKRAIKRG DYLRFQRLFHLDRGKLARIILDNTESLSCDIS PSEIYSVFKARWETPGHFNGLGD FEIKGKAN

				CCCGGAAAGGGAACAT ATATAGTTATATGTGTT CGTAATA (SEQ ID NO: 1265)	NKAFRDFITAKEIEKNVREMSKGSAPGPDGI ALGDIKKMDPGYSRTAELFNLWLTAGDIPD MVRGCRTVLIPKSTTPERLKDINNWRPITIGS ILLRLFSRIITARMTKACPLNPRQRGFISAPGC SENLKLLQSIIRTAKNEHKPLGVIFVDIAKAF DTVSHQHIIHVLQQRVDPHIVGLVNNMYK DISTYVTTKKNTHTDKIQIRVGVKQGDPLSP LLFNLAMDPLLCKLEESGKGFHRGQSSITAM AFADDLVLLSDSWENMKENIKILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINNCPAWTIN GTPLNMINPGESEKYLGLQIDPWTGVAKYD LSTKLKIWLESIDRAPLKPLQKLDILKTYTIP RLTYLADHSEMKAGALEALDQQIRTAVKD WLHLPSCDAILYVSTRDGGGLGVTKLAGLI PSVQARRLHRIAQSPDETMKDFLEKAQMEK MYEKLWVQAGGKKKGMPsiWEALPMTVPP TNTGNLSEWEAPNPKSKYPKPCDWRRKELK KWTKLESQGRGVKNFRNDTISNDWIQYYRR IPHRKLLTAIQLRANVYPTREFLARGRDNY VKFCRHCEADLETCGHIIGFCPVTKDARIKR HNRICDRLCEEAAKREWVVFKEPHLRDATT ELFKPDVIFVKEDRALVVDVTVRYESAKTTL EAAAMEKVDKYKHLEAEVKELTNAKDVF
--	--	--	--	---	---

						MGFPLGARGKFYKGNFNLLETGLPKTRQL SVAKTLSTYALMSSVDIVHMFASRSRKPNV (SEQ ID NO: 1387)
R2	R2Dr	AB09 7126	Danio rerio	AATCCCCCTACC CAATCCCCCGTC GTGACCTCCAGGC CAGGAATCACGAG CGTACGACAGTGG CCATCCGGCAATG ACAATAGCGTGAC TAACGACAATGAG TCAGATCCATGAC CCTTGGAGTGGGT TAACCTCCGCCTC TTTAAAAAC (SEQ ID NO: 1143)	AAATCCCAGCGGGATA CAGCAAGAAGGTATCG GATCTAATAAGGTTGA GCGAGGAGAGGGTGGA GATCCTTTGGGGGGGG TCGGGCTAAGTTCCCCT CTCGGGTCCTCCACGG TGACGCTCTACCCCTCC CTCCTCGCTCGTAGAAC CCAACGGTGAACACGG TTGGCAGGATGAAGTG ACGTGAGGGGTAAGAC ATGCGTACGTGAGCGC GCATTTTTGCTGTTCTC TGGACTGGGTTTCGTCC CCCTCACAACCATCACT TACACTATAGGGGCAC AGCGGCTCCTACCTCCC TCCCTATGACCCCCCT TCCCATACCGATCCATG	MESTAKGKSYWMARRPVEGATEGSLGRVP FVTRDPKRKPEAKRTLTHGLGLRECSVLT RLIEGRRGRDHTPSGWNAQRGMPNDESSVE EPNGPIPSNPIPTGTQALPEPMADGEQGEHPG VVVTLPLRDLNCPLCGGSASTAVKVQRHLA FRHGTVPVRFSCESCCKTSPGCHSVLCHIPK CRGPTGEPPEKVVKCEGCSRTFGTRRACSIH EMHVHSEIRNRKRIAQDRQEKGTSTDGEGR AGVERADAGEGPSGEGIPPKRPRRARTPREP SEPPANPPILSPQPDLP GGRLDLLREVASGW VRAARDGGTVIDSVLAAWLDGNDRLPELV DAATQRTLQGLPAGRLARRPATFVAPNRRR GRWGRRLKLLAKRRAYHDCQIRFRKDPARL AANILDGKSETSCPINEQAIHEHFRNKWANP SPFGGLGRFGTENRANNAHLLGPISKSEVQT SLRNASNASTPGPDGVGKRDISNWDPECETL TQLFNMWWFTGVIPSRLKKSRTVLLPKSSDP GAEMEIGNWRPITIGSMVLRLFTRVINTRLT EACPLHPRQRGFRRSPGCSENLEVLECLLRH SKEKRSQ LAVVFVDFAQAFDTV SHEHMLSV

					GCTGTTCTAGTCTGGAC CGAGGGTCGGACGGGG CATTTGAAGGTAGCTG GAATCCTCCGCTGCTGC GAGCCTGAGGTCGATG GTTAGAGGTGAAATAC TTGGGAGGAGACACAG CCTCCGGAGAGCCCCT CCCGGGTGGTCATCAT GGCAACCGGGTGAAAC CTTACGGTTTCACTTAC GAAACAGCACCATAAC AGCGCCGTAATAGCGC ACCGGTGTGACTACTGT CCAGTGCTGATATTCTC ATCTGGAGAATACAAC ACGGGTAATGGCAGAG TATTCAAAACCCAAAT GTTTACGATCGACCAA CGGAGTCGTTCCCTTGC ATCTAGGCCGGACCCG AAACTGCCGTAATTGC CCGTCCCCAAGGTAGC	LEQMNVDPHMVNLIREIYTNSTSVELGRK EGPDIPVRVGVKQGDPLSPLLNLALDPLIQS LERTGKGCEAEGHKVTALAFADDLALVAGS WEGMAHNALVDEFCLTTGLTVQPKKCHS FMVRPCRGAFTVNDCPPWVLGGKALQLTNI ENSIKYLGVKVN PWAGIEKPDLTVALDRWC KRIGKSLLKPSQKVYILNQFAIPRLFYLADH GGAGDVMLQNL DGTIRKAVKKWLHLPST CNGLLYARN CN GGLGICKLTRHIPSMQARR MFRLANSSDPLMKAMMRGSRVEQKFKA WMRAGGEESALPRVFGANQYQEGEEVAND LVPRCPMPSDWRLEEFQHWMGLPIQGVGIA GFFRNRVANGWLRKPAGFKERHYIAALQLR ACVYPTLEFQQRGRSKAGAACRRCSSRLESS SHILGKCPAVQGARIRRH NKICDLLKAEAET RGWEVRREWAFRTPAGELRRDLVLILGDE ALVIDVTVRYEFAPDTLQ NAGKDKVSYYP HKEAIARELGVRRVDIHGFPLGARGLWLAS NSKVLELMGLSRERVKVFSRLLSRRVLLYSI DIMRTFYATLQ (SEQ ID NO: 1388)
--	--	--	--	--	---	--

					CTCTTAGAAAACCGAA GCCCCGGTCGGGGCGGT GGTTGCGGCGGCGCTG CGGGGGCCTGCTGCTC GGGCGGCGTCGGTGTG CCGCGGTGGTTGCGGT GGTGCGGCGGGGATCT CGGTCCTTGCGGTGCCG CTGTGCCGCCGCGGTC GCGTCGGTGGCGCTGG GGTGGTGGCCCGAGTG GCGTCGGCGTGCCACT GCCCATAGTCGCCCCG GGGGGCGACCGATCTG GAGGGGCGAGGGGGCT CGCGGGACTTTAACGA GAAACGGAACGCAACT TCTCGCATCGCTCCCGG GACTTTCCCCCCTCGTT CAGCCGAGGGATGCCA AAAGGCATGAAAGGTA AGTACCATAACGGTCC GCAAACTCTCTTCTGA	
--	--	--	--	--	--	--

					CTCGGTTCTCTGTTGGT TTTCTAGAGTAACAAC GAGGTGGAGGAGAGGG ACATGGCAGGGACTCC CATTCGTGCCAGCGGG TGGGGACAGATCGAAG GAACGGTTCGAGGGCG TAACAGACGAGAGGGA ATCCGGTCACACATTG ATGCCATGCCTAAATA GGCGAGGTTTGTATTTC TACTTTGTGGGTTTCAGT ATAGTCGGAGCATATG GTCGGTTGTCCCGTTGT TTTCACGGCGGGCAAG CGACTATCATGATAAA GTAGAATGGGAGACGG GCTCCCTGACAAACCC GGAAAGGCGCCCCCCC GTGGTTCGTAGCAGCT GACGGATCACGCTCGA AGAAAAATGAGTGAGA	
--	--	--	--	--	---	--

					GGGGACGCCGCAACCA C (SEQ ID NO: 1266)	
R2	R2- 1_GA	.	Gasterost eus aculeatus	CATATTGGGGTCT CAGGAGGAGACA CAGGGTCTGTTGC GGCTCCGGTAAAC GGTACCGGAGTCG GTAAAGCATCGTT TGGGCCCCGCTCC ACGTGGTGGTCCG CGGTAACACCAAT AGGGTGGCTAAGA GGCCCAGTAATTT CCCCGAATTGTCT TCCCCCCCCGCGCG GGGGGGACCCCCC TTTAGTGTCGGAG CGGTCGCGCCTCC GCGTTTGGGGTGT CGCAGGCGTGAGC CTTCGTCCCCTTA AGTTCAGACGGTC CCGGCTTCTTGCC	GGAGGGGAGTAGGTCT CTACTCTGACCCGAAG GGCCCCCCCCGTTTCAGA CCTGATTCTAGGCTACC TGTGCCTAATTGGGGG GGTCCCAAAGAGATGT TGTCTGTTGTAGAAGG GTTTGCGCCACTGACTG CACGGAAGGGTGGGCC TCGACAGGTAGGGGTT ACATGACTCCGTGCTGC TCAGCAGACCCGCGCC TCTGAGACCGGGTAGG GCTACTTGAACAAGCG ACGCCCTGGTGTATGTC CGTATCCTAACCTGGTT TGGGAAAGCCGATACC GGCAATGCCCGCCACA GGTGTCGCGCACCCCA CGGGATGACGTATGGG CCCCGGGGGACCTCAT	MLRGGVGTTPAGGAGAVGPGMASPGGCSV RFSPGGRRLLGHRTGGLSPSVSWRLKRLSVS LRRWSGPGLLGADGAGGGGA AVASPRGTQV LGSGAGRRWLGHGSRGSSPSAARGLRRLTV RLKRLSGGLLSPKACRDAEEGSSSSPGFRNP KGLGGRGLTPLGSRRFCRLTVSLNRWRGSL VKLNASSRASGRRTPVKPA CDSRAGR GSEH AEGGGVSAAPMVLR SRRKLTF SVDGDSNSG DRARSGSVSAARPGHLLVDGESASSRSGPA GDARLAGPSTRSRRKGCLPPVDFENPKKRT RLMAKMTNGNPTSHVPCPAPCSNGHEGGG RVAVIEGR LPELSGSRSIGIQPALPVETS FVG QSTGRGADGDANANSSPPSPNLGGSVGMVP AVRDGTPPLGRPGEDHSRECAGGNTPLWML EDSFRCDYCPREFGTRAGRSLHMRRAHLAE YDGAGFCWGERLSEFAATRLWSTEETKKLA VFCERGVPSPSECRAIAASLGAGKTHHQVRS KCRLVFEAIRRELLEVA AATERLEKSARRK QPAVPPAPVHGVRGVLRGLLGKRVPREGGT TGSTSARIVRRDDCRQGAVASASLNLIRRLG RKATGRSGRRRVLGRPPRMDVRRSVRMRR

			GGGCCAACCCCCG GTGCAGCGTTCTC CCATGTTGGATCG GCACCCAGCCCCG GGTGCCATGCGAG TTCAGACATTTTG TTTATGTATCGTCT GCGTG GTT GACTT GCTAAGCTCATT CCTCCTCTCACTG CGTCCCCCAGGT GCTGATCGGTTGA AGAGGATTCGTCG TTGACCTCGGCGG TGAATTTGGGATT GTATTATACAGGT AGGTATAGAGGGC GTGCGG (SEQ ID NO: 1144)	GGATACTCCACTGGAC TTGCACAATCCTGGTGT ACTGGATGCAGCGACG TTGGTGACATAAGCAA TCGCTAAGTCGGGGTA GGGGAGGTGGGGACCT CGGCACGGCTGTAGGA ACGGGTGTATGGGCTC CGGCAGCCGTCGTCAC TCCCATACAACACAGG GGCTGCATCCTGGTGG CCGGTGCTAGTTGGTTC TGGAAGCCCGCCCGGG CTGGTTCGCAGAAGCA GGGTGCGCCCAGGGTA GGTTTGGTATATCTGGG TCCGGTGCGATACCTAT CGATGGGCAGCGAGGG CCGCCTCGTGACGCGCT GTGTGGAGCTGGAGCC GGCCTGGGTATGAACA GTTCTTGCGGATGTGGC GTAGCTAGATAGTACC	MRRFLYRLARLGWAKLAMFVLDGQMGAS CPVPLVEVSAVFRERWSIVRAFLGLGQFGGF GTADNAGFGKLIDPAEVRAHLQSIKNRSSPG PDGITKVALSKWDPEGIKLAHMYSTWLVS A GIPKVFKKCR TTLIPKTGDVSLHGDVGQWR PITIASLVRLYSRILTERMTVACPSHPRQRG FIASPGCSENMLLEGCMSLSKAGNGSLAV VFVDFAKAFDTVSHEHLLSVLVQKGLDQH MVELIKDSYENSVTKVHCQEGCSTDIAMKV GVKQGDSMSPLL FNLALDPLIQQLEREGRGF PVNGKSITAMAFADDLAIVSDSWEGMRANL DILVDFCEL TGMRTQPSKCHGFLIEKSGSRS YKVNRC EPWLLNDTALH MVGPKE SIKYLG VQVNPWTGIFAEDTVAKLRQWVVAISKTPL RPLDKVSLLCQFAVPRVIFVADHCMLSAKA LTEMDRSIRQAVKRWLHLARCTTNGLLYSR KSSGGLGIPKLSMIVPAMQARRLLGLSRSKD ETVRWMFLETTDHVAFERAWLRAGGSPDE VPELGPD LVEGSPAEGNADPVSTVRPRKRIV PCDWRQVEFDRWAGQLVQKGK GIRT FEADKI SNCWLYDYPPNKLKPGDFTA AVQLRANVY PTRELAGRGRTDTIDVCCRHCGEAPETCWHI LALCPKVKRCRIQRHHKVCQVLVAEAERHG
--	--	--	---	--	--

				CGTGGTTGTGGGCGTG GTGTCGACCAAATGTT GTCCTGTGTGCACATAG GCCAAGGGTTACGTGG GTGGCAGTCAGAAGCA CCCGCACCTGGAAGTG ATTGCCCCGGGATCCC GGCTCTCTGTGAAGAG CTACCTTGAGGAAAGG TGTTCCGCTGGA ACTCA AGACCCTACAGTAGGG GATATCAACTGGCTTTG AGGTGCTGTGATTCCG GAACCAGGGCGAGGGC GAGTACTTAGAGCATG TCCAAAAGCCCGGGGA ACGTTCCGGGGGCCTG CTTGGGTCGTTGGACCC ACATCCGTAAAACGAT GGATCTCGCGTCGGCG CTCGGGAGAACTTCCC GCATGAACGCTGATTG CATGTGAGAACGCCCC	WEVEREKRWMLPSGECVAPDLICWLDELA LIVDVTVRYEFDEESLERARIEKECKYRPLIP VIRASRVQTKKVTVYGFPLGARGKWPAPNE LLLADLGLSKARTRSFALLSRRVLLHSLDV MRTFMR (SEQ ID NO: 1389)
--	--	--	--	---	---

					CACGGCGGCGGGGCAG GCGCTCCCCCTGGGTGT AAGGCTCGGGGGGGTC ACGGCTCCGCTCTAAA AG (SEQ ID NO: 1267)	
R2	R2_BM	AB07 6841	Bombyx mori	GGGCGATACGCAT AATTTTAATTTCC CGATTGAAATCCA GTCGTCTTAATCT GGTGACCAGTGGC GCGGTCACCAGTA TAGTGACACAGGAC GTGAATGGCTCCG AGGCTGGCGGAGT CACTCACTATAAG TGTGAGAGACGAT GTCCTGTGCCAAG TATACGTCCAACC CTAACGGGTTAAG TGAAATTAGTTGC TCATAACAGGGAC GGTGTACCTGTTT GCTCGTGGCTGGC	GCCTTGCACAGTAGTCC AGCGGTAAGGGTGTAG ATCAGGCCCGTCTGTTT CTTCCCCGGAGCTCGCT CCCTTGGCTTCCCTTAT ATTTAACATCAGAAAC AGACATTAAACATCTA CTGATCCAATTTGCGCG GCGTACGGCCACGATC GGGAGGGTGGGAATCT CGGGGATCTTCCGATCC TAATCCATGATGATTAC GACCTGAGTCACTAAA GACGATGGCATGATGA TCCGGCGATGAAAA (SEQ ID NO: 1268)	MMASTALSLMGRCPDGC TRGKHVTAAPM DGPRGPSSLAGTFGWGLAIPAGEPCGRVCSP ATVGFFPVAKKSNKENRPEASGLPLESERTG DNPTVRGSAGADPVGQDAPGWTCQFCERTF STNRGLGVHKRRRAHPVETNTDAAPMMVKR RWHGEEIDLLARTEARLLAERGQCSGGDLF GALPGFGRITLEAIKGQRRREPYRALVQAH ARFGSQPGPSSGGCSAEPDFRRASGAEEAVE ERCAEDAAA YDPSAVGQMSPDAARVLS EGAGRRRACRAMRPKTAGRRNDLHDDR TA SAHKTSRQKRRAEYARVQELYKKCRS RAA AEVIDGACGGVGHSLEEMETYWRPIL ERVS DAPGPTPEALHALGRAEWHGGNRD YTQLW KPISVEEIKASRFDWRTSPGPDGIR SGQWRA VPVHLKAEMFNAWMARGEIPEIL RQCRTVF VPKVERPGGPGEYRPILIASIPL RHFHSILARR LLACPPDARQRGFICADGT LENSAVLDAV LGDSRKKLRECHVAVLDFAKA FDTVSHEAL

				TATCGAATGGACG GGACCAATACACC CCCCTGTTAGTAA TGGGGTAAGAGA GAGCGGTCTGAAA CTATGGCCGAAAT CACGACGCCCCAC TCCTACCCATAAC CTGCACGTGGTAC CGCCGCACATTGA CCGATACGGGAGG AGGGGCAGCACTT GAATCACGTAGTC TTGGTGTAGCCAT TGCGGGACTACAG CCCTCGTAAGTGC CGCCTTAGAACGC AACGGGGCAATA GGTGGGCCGGGGC GCTAGCGGGGGG GAGTAATCTCCCC TGTTGGCGTGCAC CGCACTGCTCCCA		VELLRLRGMPEQFCGYIAHLYDTASTTLAV NNEMSSPVKVGRGVRQGDPLSPILFNVVMD LILASLPERVGYRLEMELVSALAYADDLVLL AGSKVGMQESISAVDCVGKQMGLRLNCRK SAVLSMIPDGHRKKHHYLTERTFNIGGKPLR QVSCVERWRYLGVDFEASGCVTLEHSISSAL NNISRAPLKPQQRLEILRAHLIPRFQHGFLG NISDDRLRMLDVQIRKAVGQWLRLPADVPK AYYHAAVQDGGLAIPSVRATIPDLIVRRFGG LDSSPWSVARAAAKSDKIRKKLRWAWKQL RRFSRVDSTTQRPSVRLFWREHLHASVDGR ELRESTRPTSTKWIRERCAQITGRDFVQFV HTHINALPSRIRGSRGRRGGGESSLTCRAGC KVRETTAHILQQCHRTHGGRILRHNKIVSFV AKAMEENKWTVELEPRLRTSVGLRKPDIIAS RDGVGVIVDVQVVSGQRSDELHREKRNK YGNHGELVELVAGRLGLPKAECVRATSCTI SWRGVWSLTSYKELRSIIGLREPTLQIVPILA LRGSHMNWTRFNQMTSVMGGGVG (SEQ ID NO: 1390)
--	--	--	--	--	--	--

				CTGGGGGCAGTGT CATCCGGAAACAG GTGGGCCGGGGCG CCACCAGGGGGG AGCAATCCCTCCT G (SEQ ID NO: 1145)		
R2	R8Hm-A	.	Hydra vulgaris	TTCAAGTGGATGA AGCTGGGAAGGTA ATCTGTAGTTGGT TGAGTTGGTTGCA GATTACTGCTGTC GATTTTGCTTTCTA TTGAAAGCCTGTC TCTACGGGTCCTG AAGCTTGAATTTT GGTAGCTATAGTT TTGTGGGAGGAAA GTGGAATTTTGTA CCATCTTTTGTCTC TCGTATCTACTAT AGTAAATCCGGTC ATGCAGCCTCTAC GCGGCGCAACTAG	TAAATGCCAAAAGTTG CTTGGGCTAAATGATA CGTACGCTAGAAAAAG CGACTTGCTGCACGGA TGACGGTTCATCAGAG CCCGATATGTGCATGTC AAGGCGGCAGGGAGAA TCACTAGTGTAGCTGTT CTTTCCATTACGACTTA CGCGGTAAACGTGGCA CGATAGATTTACACCA GGAAATAATACGTGAA GGGTTCCACCATATACT GGAGTTTAGATCTATG AGGGAAACATTTGTAA TAAGTCAGTCTGGTAA CCTGGCGCCGCTGTTGA	MNLLIVTSSIKESDVPSSGKGGVAVNNITAG ASGKDTCVIIHPGTDGIWCCTECVEIHNSGK DLKRHLAKRHPSVTISGYKCNLCPFVSERQL SVGTHLRYCRGVKEVVKREFACASCSFSSD TFSGLQVHMQRKHIAEWNDQLKEKTEFAW TDRELRELAEKELTPSFYRNKIFYAALGTS RTYDAVRKIRYNDRYKSAIAEMRSQIADAA AAAQERDVERGLVSAHSDRGKEMPLPVVET KSDIQVNNDIKKDIELTPNSRQKQTNLALAR PAVIEVEEDLGRQDVKQYLASLRQDDYTSP AERSIFAYCREETNWSATKRQVLKISRTRG LRQPKKVRPFEPPEGFKPNRNMWKWRKYRF LQECYREKRAETVSKILDGTFIDEPEEEIRPE LEEVQRMVIDRLEKRTQLDTTKIVQTDEVFC LQSYGRITIGEVRDALGASKKDSASGPDGLL LQDVRRLGPLLLCNIFNMWYLGIPVEENR CRTILLYKSGDRHLASNYRPVTIGNMLNRLY

				AAACTTGGATCAG TGATCAAGGCTAA TGCATGCCGGGTC TCCTCAGATTAGG AGTATAATACAAA TCTGACTTCATCA CTAAGAGGCTATG GGGCTAACGATCC TATAGTCTCG (SEQ ID NO: 1146)	GTCAAATTA ACTATGTC AATACTCATTAAGTTAT CGACTTTGATATGGCAT GGGGTGATTCCGCGTT ATATCAAAGTCAAACA TGATGATTGCAATGAG AAACTACCACGCTTGG TCACGTTTGTGAGGAG AACATCTCATTCAAGCC TCCCGGATGTCGGCAC CCGCTGACATCTTCTGG CTTATGAAAATTTTCAT TAATTTTGTAAAGTCAT GGGCGGCTTGAAAGC (SEQ ID NO: 1269)	AKIWDKRIRKNVRLHVRQKAFIPVDGCFEN VKTIQCVLQSYRKRKLEHNVFIDLAKAFD TVLHDSIRKALWRKGVPSGVVKVVDSLYA GAVTSISVGKTKTRSICINSGVKQGCPLSPLL FNLILDELAERIEATGCGLDLDGHVLSMAF ADDYVLLAKDSVEMNELIRVCSTFFKEKGL SVNPGKCQSLRVLPVKEKKRSMKVLVRPHR WWRIKDQDQVDIPSMTYDSLGLGYLGVSIDPT GKIALPIE EWKNWMTKLKECKLKPEQKVKI LKEVVCSR VNYVLRMSECGISELRSWTRFV RNWAKNIIHLPTWCSSDWIHSIKGLGIPDVS KGIVIQRM RASEKMSTSEDGIVRVVGARLV QKNRVLWEKAGFEGIELKAARRHCEVERLN NIGNITNGVALKTIAAVSSVNRYWMIEDNL KSGNKILVWKAMAGAIPTKINLSRGVADQT LKKCRRCGLTAETDGHILAGCHTSSDAYSK RHNMLCDKLAKELKLNGGPNRRVWRERTC FTSTGRRYRPDIIVKDDSKITVIDMTCPYEKS EGHLIQCESAKVTKYEPLKLDKYWTRELEG ANGIVA EKVELMGLAIGAIGTIMRSTLRKLC ELKSGRIVRRLQMIACNNSAQIIKGHLSRAT RRNLR (SEQ ID NO: 1391)
--	--	--	--	---	--	---

R2	R8Hm-B	.	Hydra vulgaris	CTTGGGGTCACTG ACACATTTTTCGG TAGCCATAGTTTT TTGAGAGGAAGA GTGGAAGTTTTTC CATGAGTCGTCTC TCGTATAAACTGT GGTAAATCCGGCC ATCCAGCCTCTAC GCGGCGCAACTAG AAACTTGGATCAG TGATCAAGGCTAA TGGATGACGGGAC TCCATGGATAAGG AGATATAAAGATC TTATTTGAACGCA TCTTAAGGGGTTA TGGGGCTAACACC CCCTTAATTCTGG TGCACATTTATTG ACCGTT (SEQ ID NO: 1147)	ATGCCCCGAGGTAGTTG GGATAATGATGCACAA GCTCGTAAGGCGACTT GCTGCACGTATGCCGCT AAACGCTTAGCTCGAT GAGTGCATGTCAAGAC GGTCGGGAGTATGATC AGTGGAGCTGACTTTCC AGACAACTCACGCGGA TTCGCGTGCGGTGGAT ACAACACCTGGTATAA CATATGAAGGGTTCCA TCTAGTACAGGGATAA CGATCCATGGGAGCAA ACTAATTAGTTGGAGG TAATCCAACGCCGCTGT TGAGTCAGTTTTTAACC GCCAGTCAACTCTTGTA GGTTATCGGTCTTCGGC AGACCTTGACCGCCT AGCGCCGGCCAACAGT TTGTCGTCGACTAACAT GATGATTTGCGAGAGA	MSNRITIGDVPSVGKGGGLTVNKQTAGADGA EACVVIHPGAKGIWSSPACLRKFTIGKELRA HLAQIHKLAPSAVRYRCNKCPYEGDVQLSV GTHLRYCKGIAGVVEEKKQFACAICNFSSDT FSGLQVHKQRKHVVEWNEQLKEKTEFAWT DRELRELAVKEVTIPFSVVNTETFAVLDTTR TKDAVRKIRYTDYKXSILAEVRAQVNAVAE EAPQASDESQITLLVNTGRGAELQPAVINITD SIELVTDVNEVEMVTSNSTNEEQPINAPVEP AVIEADLGRQDAKLYLASLRQSDCTNASDR WTLAYCRGEVDWCKTKSRLFKVSRHARGL RQPQRVENWEFPEGFRPNRNLRKWRKYSFL QSCYRTKKKETVSKILDGTFKDTPEEEIRPEL EEVQRVYVDRLEVRTQLDTRTVHIDERFD LVSYGRITIREVQDAISASKKDASGGPDGLL LQDVKKASPRQLCIFNMWYLGIPVVENR CRTILLHKGGEKHLTSNYRPVTIGNMLNRV YAKIWDRRIRKNLQLHVRQKAFVPLDGCFE NVKTIQCILQSYRRSRREHNVFVVDLAKAFD TILHDSIEKALLRKGIPIRSVIKVVDSLYAGAV TSITVGKTKTRPICINSGVKQGCPLSPLLFNL VIDELAERLEATGCGLDLEGHVISSMAFADD YVLLAKDSVEMNVLNMVNCNTFFEEKGLAV
----	--------	---	-------------------	--	---	--

					AACCCACGCTTTGTCAC TTATGTGAGGATAAAA TCTCTTGTCCATATGAT CCTTTGAAGGGAACAG CGCTTTGAGCTTGCTCG GCGTTGGCACCTTTAGT CTGTAATATTTTCTTGA TATTATGGACGAAAAA GGTAGTATGGTTGCA (SEQ ID NO: 1270)	NPAKCQSLRVLPVKGKRSMKVLTRTHRWW KINNQDVEIPSMTYESVGKYLGV MIDPAGKI ALPIEEWKLWLTRLRECKLKPQKVVLKE VVCARANYVLRMSGCGICELRKWSRFVRG WVKSIHFPAWCNSEWMHSSKGLGIPDVVS GIVIQRMRAAEKMAKSTDGVVRVVGARIVQ TNRVLWKRAGLAGIELDAARKFCEVKRVN KIGNQTNGGALKTIAESSVSRHWLLEKNIRP GNKILVWKAMAGVIPTKINLSRGVADQTLK KCRCCGLTAETDCHILAGCPTSRDAYSKRH NLLCDKLAKELRLNGGPSRRVWRERMCLS GNRRYKPDIVVKDDGVITVIDMACPYEKS ERHLSQCEDAKVAKYEPLRLDRSWTQELEG NNGRSANEISVVGIAVGAIGTITRKTQRILSK LKLAKVGRPLQIIACNESAQIIRRHLSGSRLR NLR (SEQ ID NO: 1392)
R2	R9Av	GQ39 8057	Adineta vaga	GAAATAGTTTGCA ATGGTAGGTGTAT GGCGCCTCTGTGT CTCTCTTTCGCTG GATATAGTTTGAC GATTTTGTACCAG GTATCTGTTTCTTG	ACTAGTCTCCTTCTTCT ATTAGTCAGTCTAATTA ATTTTCTTACATTCTA CATCTAGTTCCATTATT AAATTGGTATGATCAG TGCTATCTCTGCTACAC TCAATGCTTAATCGTAT	MNLPIREHAVSVHNINKFNYLCQLCSKSYDT INSVKAHYVACRRQKNASSTTAVPTNVINN NQLAINTNQVISRNPLQCVECLMKQVDFYA KDTKALVTHMRTKHAAAYEESKKVATRRV AWSPDEDQILAELEVKLKKIQKGQLLSRLV VEYNKCADKSKAPSRSKDAIRTRRQQHDYK LLRLSLQSQQPPVGSEDSDDISSNNNPLTT

				TGAGTTCAGCACC AGTTTGAACAGGC TTAGCGATAGACC TTCGAACTTGAAA CACTGTTGTGAAG CTGGCTGGGCCCC TGCAGATTTTCTC GATTAGAACGTGA GTGTTACGTCCAG AATGACCCACCAG TGGTTAGTTCTAC GTTGCCCTGGAAA GGAGAAAAGTTG AGCTAAAATCGCA CGGCCTAGTTGTT TATCAAATAGGCA CGGTGAGGAACTC TTCTATGTACCCT GACTAAAGTACTC ACTTGTGCGCTGG GTTTGCTCCCCCT CGCATTGACTTAT CTGATCGCACTAC	GTTATTGACAGTCTGAC ACTTGATTACTCTTACG ACATATGCACTGTTTGC TTCAGAGAAACCACTG TTCATATAGTGAAGTTC CTCAGTTTTCTGTTGAT ATATTCTTCTTTCATTC TCGCTTCTCCTTTTCTA CTGTGTTCTTTTTATCA GTTTTTTGTGGAAAAAT TGAGAATAAATAAAGT (SEQ ID NO: 1271)	THNVTPTPDSSNVLLIQKIRESVDSIVKITN LKLNTNMLNAASAFINQNNNMDPLELSMR GIEEDVKAIRDKELQKPTRNVPSSSTTSRKPTR NAKRLEKSKKYGYQHLYYNNKKKLVAEI LDGETSGAKPPPMNLVEDYYRNIWSRSTIDD SPVNNIKTVNSDSIFAPISRDEIKLALSNTKK DSAAGPDAVTIKEAKAIIDNLYVAYNIWLG VQGIPEQLKLNKTILIPKGNSDL SLLKNWRPI TISSIILRVYNRLLAYRMNKIFKTNDKQVGF KPVNGCGINISWLHSLK HARLNKNSIYACL VDVSKAFDSVSHQSIVRALTMNGAPSLLVK LIMDQYTNVNTVITCSGSISNKINISSGVKQG DPLSSLLFNLVIDELFDVIKDQYGYTIDNIGT TNARCFADDLTLISSRMGMNKLELTTKFF KERGLNVNPSKCMSIGMSKGYKGKSKIES EPLFSITDAQIPMLGYIDKTTRYLGVNFTSIG AIDAKRIKKDLQDTLDKLEHLKLKAQCKMD LLRTYMIPRFMFQLIHTELYPKLLIKMDILIR KLAKRILHLPISTSSEFFYLPFKEGGLQLTSL KEAVGLAKIKLHKKIMSSNDPMLCYLIESQR SRIVEHFMKDLKLGDSTLNEMNNIKECFM KEKRISFAQKIHGVGFVFSSTPLTNQWINGE IKTMTTKTYINSIKLRTNTLETRVTTSRGLNII
--	--	--	--	--	--	---

				CCACCAAACGAAA CATAAACTTAGCT CGTGGTATCAGTC CACAGCGTGTGCA GTCGGATTGAGGG GAGCGTGTAGTG ACAAGCAGGATA ATATTAACATAGT TAATGTAAAGGCG TTCAACATTCCTT ATCCAATTGGAAG AGTTGACTGTGAA GTTTGTCATGAAG ACATTGGACAA (SEQ ID NO: 1148)		KTCRRCHVADESLMHVLQCCSSTKGLRYSR HHKICAKVANKLVMNGYGVFREKSYPDPN NSGSYLRPDIIVKNGHVIVLDVTVVYEVTG ATFINAYQTKINKYNMVMQIEQMFNCVNG ELHGLVIGSRGSIHHSQHLIWHQMGFSSIELK YVAIGCMEDSLRIMSTFSKAIT (SEQ ID NO: 1393)
R2	R2OI	LC34 9444	Oryzias latipes	CGCACAGGGGAC ACAGAGCCTGCCC AAGTACCGCTCCC GAGGGAGCGGGA AACGGGGGGGTG ACTATCCCCTGGG GTCCGGCGAGAGC GCTGGTCTACGGA	GGGGGACAGCTGGGAG TCTCGGCATGATTACAA ATCTTGCGCTGCACTCG GATGTCGTCCCCGTGAC GGACACATTAATCCGG AAAGCGAGTGGTGACT CGCCTCAAG (SEQ ID NO: 1272)	MGTDTVYVGQDYPGSLSKRVPARLVAGPM LRERSCHAHVFRAGHMWNWRTSLPSGRWD QPALEKSRVLTRSVATATDPEITSYPGKSVS TSTQVQEEDWCSRESGWISPLAPEEPSVVS EITASMVATMRVATEEVVLEPQPEQVVTILP EHGRNVPPGLAEQDTASPIEVSVLLPDLAEN CPLCGVPSGGLRLLGKHFAVRHAGVPVTYE CRKCAWRSPNSHSISCHVPKCRGRARMPSG

				CCAGGGGTGGCTG TGGGCAGGCTGCT CCTCAGGCCAGTT GATTAGTTACGCA TGGGCTGTACCTC CACGTGGTCCCGC TGGTAACGACTTG TCGGCTAAATCAG CCCGCCCACCATC TGGGATATGGTTG ACCGTCTAACCCC AGTACTCAGGTCA CAAACAAA (SEQ ID NO: 1149)		DPGIACDLCEARFATEVGVAQHKRHVHPVE WNKVRLERRGARGGGIKATKLWSVAEVET LIRLIREHGDSGATYQLIADELGRGKTAEQV RSKKRLLRIDTASNSPDDAEVEEERLESLAV RSSSRSPPSLVATRVREAVARGESEGGEIR AIAALIRDVDQNPCLIETSASDIISKLGRDVD GPKRPRPVVREQTQEKGWVRRLARRKREY REAQYLYSRDQARLAAQILDGAASQECALP VDQVYGAFREKWETVGQFHGLGEFRTGAR ADNWEFYSPILAAEVKENLMRMANGTAPG PDRISKKALLDWDPRGEQLARLYTTWLIGG VIPRVFKECRTKLLPKSSDPVELQDIGGWRP VTIGSMVTRLFSRILTMRLTRACPINPRQRGF LASSSGCAENLLIFDEIVRRSRRDGGPLAVVF VDFARAFDSISHEHILCVLEEGGLDRHVIGLI RNSYVDCVTRVGCVEGMTTPPIQMKVGVKQ GDPMSPLLFNLAMDPLIHKLETAGTGLKWG DLSIATLAFADDLVLVSDSEEGMGRSLGILE KFCQLTGLRVQPRKCHGFFMDKGVVNGCG TWEICGSP IHMIPPGESVRYLGVQVGPGRGV MEPDLIPTVHTWIERISEAPLKPSQRMRLVN SFALPRIIYQADLGKVTVTKLAQIDGIVRKA VKKWLHLSPTCNGLLYSRNRDGGGLGLLKL
--	--	--	--	---	--	---

						ERLIPSVRTKRIYRMSRSPDIWTRRMTSHSVS KSDWEMLWVQAGGERGSAPVMGAVEAAP TDVERSPDYPDWRREENLAWSALRVQGVG ADQFRGDRTSSSWIAEPASVGFAQRHWLAA LALRAGVYPTREFLARGKEKSGAACRRCPA RLESCSHILGQCPFVQANRIARHNKVCVLLA TEAERFGWTVIREFRLEDAAGGLKIPDLVCK KADTVLIVDVTVR YEMDGETLKRAASEKV KHYLPVGQQITDKVGGRCFKVMGFPVGAR GKWPASNNTVLAELGVPAGRMRTFARLVS RRTLLYSLDILRDFMREPAGRGTRVALIPAA TGAAN (SEQ ID NO: 1394)
R2	R2_LP	AF01 5814	Limulus polyphe mus	TGGGAGGAGACCC AAACATATCCTAGG ATGGGGCGGAACC GACCATATGAGCC ATATTAACATTGC CCACACTATCCTC TGGAGGTACCTCC TCGTGGTACGGCT GGATATAGGTAAA TCCTGTAACCAAA TCCTCCAACCCGT	ATTTTGTCTCTTTCCCC AATGATGTCTACTAGC ACGCTGCCGAAGCTAG ATAGATTGAGGAATCT GCGTAATCTGTAATGAT TACGCCTCATGGGCATC TATCGGTAGCGTCGAC CCTGACGTAAATTGG GTAATAAGAAATATCG A (SEQ ID NO: 1273)	GIDGYMFGYARASGSTSVSIQSSSMTEGETN ERATPRASDSSSVSIQSSCVTEGECLPPTDNC NPSVENQLPCVTEGRFERVGS LVTVRLPFRK VACDLC SKEFLTYSKFVH QANFHNSETQA CCTYCGKSDGNHHSIACHVPKCPWRRTVTF AANLSNFLCDLCNDSFKTKSGLSQHKRHKH PCSRNAERILSLGVRTPSARPRQVWSEEET RTLREVEVVYSGQKNINVLCAGHLPKGTSK QVSDKRRDLHRIRSSNVHGTPTTQSRGDPVE QVEEYEELDWEGMHPFPDPDSKFCSYLDQL RDQKGLTEPVWQEIEIVAQEWVENLAHVQS

				GAAGGAGAACAC TAAAACCCATATA GTGGCCTCGCCAA CCACTATATGTCC AACGGCAGGAGA AGCTATCTCCCGG ATGGGAAGGAAA ACCCTAAACCGTG ATGGGAACTTACC GGCCCCATCAGCT ATTGGGTACCCGG TAGGGACTTGCAA CCCTACCCTGTAT TTGCATTTTATAG GGAACCGGTCGGC CCTATATCAGAGT AGACCGTTTATTA AATATGGGTGAAA ATATTAACAGTAA AAGCTATGGTTTG GCGTCCGTGTGGT GCCAGGGCGGCG GCCAAACCCGAGC		SWNHERTTKQVPENNTPARRPFKRRLHRVE RYKRFQRMVDLQQRKLAEEILDGREAVTCN LKKEEIKDHYDQVYGVSNDRVSLDDCPRPP GANNTDLLKPFTPTEVMDSLQGMKNGAPGP DKITLPFLQKRLKNGIHVSLANVFNLWQFSG RIPECMKSNRSVLIPKGKSNLRDVRNWRPITI SSIVLRLYTRILARRLERA VQINPRQRGFVPQ AGCRDNIFLLQSAMRRAKRGKTLALGLLDL SKAFDTVGHKHLTSLERFAVHPHFVRIVED MYSGCSTSFRVGSQSTRPIVLMRGVKQGDP MSPILFNIALDPLLRLQLEESRGFMFREGQAP VSSLAYADDMALLAKDHASLQSMGLTVDK FCSGNGLGLNIAKSAGLLIRGANKTFTVNDC PSWL VNGETLPMIGPEQTYRYLGASICPWTG INSGPVKPTLEKWIANITESPLKPHQRVDILC KYALPRLFYQLELGTNLNFKELKELDSMVKQ AVKRWCHLPACTADGLLYSRHRDGGGLAVV KLESLVPCLKIKTNLRLVHSTDPISSLAESD GLVGAIEGIAQKAGLPIPTPDQRSGTYHSNW RDMERRSWERLALHGQGVELFKGSRSANH WLPRPVGMKPHHWVKCLAMRANVYPTKR GLSRGNLSKNKDSAKCRGCTSMRETLCHLS GQCPKLKSMRIRRH NKICEHLIAEASF KGWK
--	--	--	--	---	--	---

				TACTTGGCACCAA CTGGGGATGGTAG CTTCCGAGCGATT CCCTGGCGACGTG GGACCGATCGACG ATGGAGTCCAAAC ATCCGGAATAGAG GAATTGAGAAATA CCTATTCCACCAC CGGCTCACATACC CAAGGTGAACCCG GTGCAACTAGAGT ACAACCTATCTGT GGCGGTAGGTGCC GAACCACTCAGGT GACGGGCTTGTTT ATTGATGTCTCCC TACGAGACACGAA TTGTGACAAATCC ACTCCGGTGGACA ATTACCCGATCTA TGAACCTGTTACC GATATTAGACAAG	VLQEPTLVTDNGERRRPDLIFHRDDKAVVV DVTVRYEISKDTLREAYASKVRRYGCLTEQI KDLTGATSVVFHGFPMGARGAWFPESDV MADLNIRSKYFEEFLCRRTILYTLDLLWKS NEQYLERLAP (SEQ ID NO: 1395)
--	--	--	--	---	---

				AAAATAAAGAACT GACAACGCCTAGA GCTTCAGGCAGCA TGTCTGTAAGTAT CCAGTCATCGAGC GTGACTGAGGGCG AAATTGATAATAA CTCTGAAACTGA (SEQ ID NO: 1150)		
NeS L	NeSL-1	Z8205 8	Caenorh abditis elegans	GCTCACTTTCTAT CGTGTTAACCGTA CGTTTACACTCCC AGTGAGTGTAATA AAGGTTATTCGAT AGAGGGTGTCTCC CTCTTTCTTG GGT AATTCTTCGGCGG TCCGGGGTCTCTC CCTCGTCTTTTTTT TAAACTTTTCTTTC TCATCCACTCTTTT GCTCCTTTTTACTA ACTCTTGTACTCT	CCTCCAGGGCACGCCG CACGCCAAAAGTCCTG GCATAACTCTGCAAAT AACATCAAACGTCAAT CAACTCCACAACTCTC CACTCTCTTCAAGTCTT CTCGGTGCTTCCAACAC CACAATGGTGAAAGCT CCTTCACCTTTTCCCTC CAAAATTCTTCCCATGT GGGGAAGTCCTGTTCTT GTAAGCTCTCCGGAGG CTGCAAGAGCAGAAGA AATTCTTCTTTCTGACA	MLRRKGRHRMVMVNSVKWQPSAHAEAIGT GKSWAPQRSQASEHGWQSNAMFDPPNRILF ARDSWSLNQSTHLQNQRSGSGLGIRPGQVR NNMVGGGPHRAGDPKRRVELVSIQGSEVTV RTIYPSDEIFSCYSKSCDIKTKAGYGPEDLKH LTRHIKNEHGLKARWAYQCGLCNEKSDPSV SEGHWMEAHMVAVHQSSAEKRIKSYQKC TGARVAEQLQAAAPSLTVPGKHKSGSRDAA KDSMTPTKDDDPKTRIYQTRSVVKKSTQKT AEPTDEGSRGPKYASIFQKSVKARKSLALLC ELSSPKPMNPLPTNELTLKEGNSRELAKEEA PSEGIDDIVIIDLDESEESPPRRKRFTWCLD HESSREAWLDDTAIFWYISYLCRGSTKYSAL DPCLWSMYKVKGSRYILDRLESSITYFFPICE

				ATAGTCTTTTCTC ATCCCCATCCGC CGTTGGGCAAAGT TTATTTACTTTGTT AAATCCATATTTT ATCTCTCTCACCC GTACAGAAAGCGT CTCCTTCTCAAAC GCTTTTCTGTACTT TTTCTTATATTTTC ATTAACATATTTT TCCTGTTTATACT AACCTAACCTCCA TTGTCAATTACTA ACTAACTTGTACA ACGGATTTTCG (SEQ ID NO: 1151)	AGGTCAGAAGGAAGTC CTGTTCTTGAGGCGTCC ATCCCGGGCGTCATAG GAGAGATCAGATGCAC CTTCTAGCAGGAGCTA GAAGGGCTGCCCTGTC TTGAGATCCCCACGGG GGTCAATAGACGGGAG GGGCTGCTGGCTTTCTC TTTTAAAGAGGAAGCA CCAATCCGGAGATCCTT AGGGGTCAAAGGATTA AAAGGCAGCAGGTCCA ATTCTCCTCACTGACTT CGGTCAGAGAGGAGTC CCGCCTTGGAGACCTCC CCGGGGAGGTTGCTGA AGAGGCGGAAGCTCCT TCTAGCAAGAGCTAGA GGGAGTTCCCAGTCCT GAAACCCTTGCGGTTG ATGATGGAATGGAAGA GTACTTCGGTACTGCTC	EDHWTLLVLKDNSYYYANSLHQEPRGPVR DFINDSKRARKEFKVQVPLQRDSFNCGVHIC LMTNSIMAGGKWHSEEDVRNFRKRLKCTL QEEGYELYSVNSLGIPFQAPTTEQMDYKETR CKRSYASVLTQISPPAKRPDCKPDNNIFVPT KDCAAEGNPQEKGRNESPEEINTEHIVVAGK PANNISPRCRSTSEMLFEMVKATTSSGRSSL GTMTQDEFIRSTIAEAVPLMSIKLPPMELPR KILPIPPRKPTQTNGGQKKGKQQRVPTGKPD TLNAKVRNWFNNQLESYAMEGRSFQRLEW LTEVLTASIQKAAAGDEGIVDIICKRNPPLEV AKGEMCTQTENKRKTTNNAARIADPIQSSK GAGDVKASYWKERARTYNRIIGSKEELCKIP IDQLEDFFKKSTSRTNVQESIMKEKSSKIPAL KIGNWMEKKFIGKEVAFALRKTKDTAQGA DGLRYHHLQWFDPSGELLAKVYNECQRHR KIPKHWKEAETILLFKNGDQSKPENWRPISL MPVIYKLYSSLWNRRIRAVPNVLSKCQRGF QEREGCNESLAILRTAIDVAKGKRRNLAVA WLDLTNAFGSIPHELIEYALTAYGFPQMVV DVVKDMYQGASMRVKNATEKSDRIPMSG VKQGDPISTLFFNICLETVIRRHLESANGHQC LKTRIKVLAFAADDMAILTDSPDQLQRELSKL
--	--	--	--	---	--	--

				GTTGCTCTCTCTGCGTT TTACTGCCGAGGGCCG GATTTGCTCGAATCGCG AAAGGTCTCAATCGAC CATTCAAGATGACGGC TTATCTAAGGTCCGAA AGCAGTTGGGAGAGTA ACGTGTTCTCCTACCTT TCAAGTTGAATGGTCGT TTTACTGTTTGGGATAG CTGACTTGATGCTAGTA CGCTTCATCTGTGGATG ACGCTCCCCAAGCAGT CAAGTAGACTTGAAAG GTGCCCTCGCCCTAGTT AGCTCTTAGACCTTATG GGTCGCCATGGTTGTG GACGGGTATGCTTGCC GGAGCCGAGTCGTGTT TCTTAGAACCAACCTCG ACGAGGCGAAAGCTTG CACAAGTTAGCACAAAT TGTGGTAGGGCCGACT	DNDCTPLNLIFKPAKASLVIQKGVVRSASI KLKGNAIRCLDENTTYKYLGVQTGSAARIS AMDLLEKVTKELECVVKSDLTPPQKLDCLK TFTLSKLTMYGNSIPLITEIKMFANIVIRGV KVMHRIPVRGSPLEYIHLPVKDGGGLGVACP KTTTCMITFLVSTLKKLWSDDEYIKTLFTSLA EEVVKKESKKSTVTMDDIADYLNVEERINR SEFGYNSITRLRDVMRNLAITGDSPLYRLKM VVKNGKIALLVQATSESMERIYTEEDAKKL QRSLKDQVNKALKHRFNNTTKVVKSKVVRV VQQHPASNRFVTKGGNLSLACHRFVHKARL NLLACNYYNYDKSKSKVCRRRCGKDLETQW HILQNCPFGFSKKITERHDAVLHKVKTliesG GKKNWTMKIDEELPGFSRLRPDICKSPDEK QIILADVACPYEHGVEAMERSWQAKIDKYE TGFAHLRKS GTKLTVLPPIIGSLGSWWKPTG DSLKELGIKGSVINS AIPELCATVLEHSKNTY WNHIFGEAYIPNPMRNGHAKPAGNGWKKE RLQKAPVRPTN (SEQ ID NO: 1396)
--	--	--	--	---	--

					AGAAAATGAGTCCCTT AGGGGGTTACGCCTTG GCGAAAGTGAGGACAA TTGGCATTGACGGGTG CTTCGGCACTAGGCAA AGGCGCCACCACACTG TCCAATCTCTAAAAAGT TCACATTCATCGAAGA ACTACCGGAACCAACC ACACATGTGTTGAAAC CTACACGGTGGAAGGG AAAGGAAAGCTTCGCT GGAACGAAAAGAACGG ATAGGTTCCCCTTCTTG ATGGCTGTGAGGCTTA GGATGGACGGGAAGGC CGTGAGGCCTCAGGCG GGTAACTCGGCCAGAC GCTAGTTGATCTTCGGA TCACGACAGCCCTGGC TAAGAGGAACCCTGGA TGGAGTGTGAAGGATG GGCGGGTAGGGGGTTA	
--	--	--	--	--	---	--

					AGCCTGTTGACAGACC ACCGACTGCAGTCACA AAATCAGTGATTATGC GGGTGGACCAATCTGT TGGCGGGTGTTTCCCTC TACCTGACCCCGCAAT ATGGTATGTACGATCCT CGGATCTAAAATTCAT AATGGCCCACCACAAC CATAAACCTCCCTAGC AGCTGGTGGTCCCGAT AATTCGGGTTCCTTGCCA CTACTGCGACCCAGGC TCGCC (SEQ ID NO: 1274)	
CR E	Cnl1	.	Cryptoco ccus neoform ans	CCCTCTTAATACC CCATAACACATAA CAACCCCCTAATC AACGTTCTCTGCA CCTTAAACACCAC CAAC (SEQ ID NO: 1152)	TGAGGAAGAGGAGGTT GGATTATTTTTTCTTTT CTTTAATAAGTTGTTTA TTTAAGTAGTTTCTTTC ATTCGGGCAACCCACA CGACAACCCAATAAAT TAAACAACGAAAAATG CAACCTCTATAACCC (SEQ ID NO: 1275)	MSLQRAKNARGDPGRCNLCSADYRDLKDH LNKQHSTHFFVPSDLRGSSLVACPRCGTPCS AGTGLSRHQSRYCGLTAPRIRNRVGNSTN TSRCPPSNTAASPIVPSPSPERPSPQPAEVVA SLEPLSEAEVLEVAQVDAETVDTLEGTRRA PESVPRSAEEGSTRVRELNMTAPEEEHRGEE ESSHTNPTAPAGLENAVSSTLGSPGTLPSLL PSQECANERFLYLAHLPVRSKPLPNNLVTDF MDAAERCALAYIAQPSDSTLLAFLALPKVG

						LTQALAPEQPLRPSTFLKQFPHIPWPEQPPAR RPPSNIRPDTTKQVIKLVENGRLGAAERVLE EDASVAELDQGVIDQLITKHPKGPSCPGNA VGPTPGKAPDIDTIQKALDSFKPDTAPGVSG WSVPLLKTAARKREPVKQFLQLLCAAIANNT APGRSMLRTSRLIPLKKDDGSIRPIAVGELIY RLCAKALIISHFQPDFLLPFQLGVKSIGGVEPI VRLTERVLEGSAGAEFSFLASLDASNAFNRV DRAEMAAAVKTHAPTLWRTCKWAYGDSS DLVCGDKILQSSQGVRRQGDPLFFSITLRP TLNALSQSLGPSTQALAYLDDIYLFSDNSQV LSKTTQFLADKQHIIKLNEKKCKLISFDEIRQ EGFKMLGTMVGGKEKRAEFLEGRIRKEMA KVGKLDLPHQHALLLRFCIQQNLRHLQR SLRSDDLVDLWERLDTMLWEEVKRMRMR QREDTAEELALGRSLTKLPARLGGLGLLSFK DVAPLAYRSAAEASDTLLDNLGLLSSPEEPP TPIPQRTRCAELWESQQEAILHNLGDTERKR LTENASRLGRSWLSVIPYLQPLRLSNVEIASG LHDRTLVGSSIPVCRFCGSDSPLGHDELCRA RNPWTQRRHNAINRVIIYQHLKQIQGATVEIE PHTLSGQRRNDLRVRGSSALAFDYLKQVY SLGDRDARSTVTPCAPNGKLADFCLDRCVN
--	--	--	--	--	--	--

						WLDKVGQVVSKNAPKVTGGVFKPIILSTGG LMSRSTADEWKDWRDAMPVGGFEKMEKRI GVELVKARARTLV (SEQ ID NO: 1397)
CR E	CRE- 12_CCri		Chondru s crispus	ACGCCCCCTATCC ATTTCTGCCAGCC TCCCATCGGCTCG CCGTCTCCGCAAC CCCTCTTCCTCGG CTGTACCAGTTCC GCTCCCACAACCT CCCTCGCCACA (SEQ ID NO: 1153)	TAAGTCCTTGACGCCTG CCCCGTGATACAGCAT CGGTACCCCTAGCATT GAATAAAAAA (SEQ ID NO: 1276)	MSQPNISSAETPLSQLPTPVTPPPSPSNPSLSL PTVRDLLLCPIRSSHVYSSIPSSCLHSFTMLLI KTVRAASATMTPTESHRAFIHLHILPIAVLRR SFRGETGWSRRTGQHHALRQRIRRASSGRH WAALWHEALAAHQVDLDYRTRHSRRYQA SATSRHRIGRAMRLAADAQYGRAMSALKA KPLPDLHAAA TRDTLTALHPPASPQPLSP TDLPPVPEITEGQVLRAARALNPTSAAGPDH LSPRILQLLARTTISPEAGVTGLSALTNLVRR LARGDIPDR TAPLLAAATLIPLQPRPHKIRPI AVGQALRRL VTKVLLPPAIQDTRDHLLPEQL ANSVASGMDAIVHDTRMLMHRHGRNPDYI MVSVDARNAFNTFSRQSLLDRLPLQTPSLAR FLNLIYGRTPDLVLPSSPRFLMKSQEGTQQ GDPASMLLFS LAIQPLLRRLTRECRLDLNRW YADDGTLVGPISSEVIKALRILRDDGPQSGFH VNINKCRAYWPTVMPEKLSELLRIFPLHVEC GEGGVALLGAPLGTD AFVRRHLMNKVQSC HASLSLLDEIPDARTRFHLHRVTGSVCKVEH VFRLTPPHLSLPAATKFDEQQIAAYSRLNDV

						AVSTSMATQIGLPFRLGGHGFTPLSPFIHASY AASLIEAAPVRVKGPHNPSESFYRRMARRHI VHVLGALNPEVRTRGILGTHSPLGPFEPEAL LSRPERVHHTLIQAMQGATSRLYWEHTAW DLDPLPRNHSAASVRRRARYNSLRAPGAAS FLCSHPSLTSRVPSAVWSCMLRRHLDTPVY CDSIRPLICSHCCKPMDARGDHAAICRHGFG VVHRHNTVRNLLARHAFRAAGLCCDLEVPS LLPNTANRPADILVQPAPPPSGALPDRPTAY DVTVRSPYCRSTMSLAAKGLAGAAEAADL DKLRVHSRTVRDAFHLQPDSPLPLLDWHFV PLAFDTLGATSSRTMAVLEYLAHRIANRTYS SYGTAKIRLLQRISFAVWSSLASATLSRMPY HGAALSSPAQV (SEQ ID NO: 1398)
CRE	CRE-13_CCri		Chondrus crispus	CNCCAGCCAMCG ATCCCGCCGCCAC TCGCMGCCCCGGCC GTCTCGACCGCCA CCTCCCCGASGCC CCAGCCCATC (SEQ ID NO: 1154)	TAATTCACCTTCATATC TGCTAGTGTCTCTGTAA GCGCACCCCTCATGCAT TGATAAAATTACCCCC A (SEQ ID NO: 1277)	MAXXPXISPPGAPPAPLRYRMLQCPPPLPKX XXXPVPHPMSSPIRXRLPHRXMRGPPSXTTP RDMHRPHGTPGPHSHRXCGRPPXHCTHASX QPRXAXHXLQXPKLRSPPPHPHVSPLILCXG PLPTPMTQPRMKRALXSASAPPTKRPSASQ GPAASSHDXPRTPPPXP RP PPYPFPPTLDQ HXFALSXAYPHXP RP PPSPXRXLRHSPFP FGXQTFSSIPGRLHSTVLLLIRLVRAATAAN TPETTLXSCTFTCSRLPFFERPSXAXLAGGP

						RAVNFMLSACXYGERVXDESGXSYXXKXH CITSHPPRPRRYSRQHRSRNHHTXFLHLHLLP TAXLREAFRGEXGWRSSRGQLHALRLXIRR ACTGREWGLLXXEALDAHSXRTEWQHHA RRPSPVSPSARAARAMRLASQAQYGRAMR TFTNPPLADLNDPATMERLQALHPTPTVPVV PLPPSAQPRPPEVTXEAVXRAVRRLNPNSAA GPDRMSPKLLHLLAHTPISPEAGVTGLSALT NLVSRLARGSLPPCTIPLASAATLLPLQPRPG KIRPIAIGQALRRLVTKXLLPAAIDDCRDHL APEQXANGIPNGIDAIVHDARMLVRRHGND PHYXMVSIDASNAFNNFSRQQVLDQLPTRA PSLSRYLDMVYARAPSPLVLPSXPPTILHSRX GSQQGDPASMLLFSALQPLTRLISRECLX MNRWYADDGTIIGRIDEVXKALDIITKEGPR FQFFLNPSKTRVFWPSRQXDLLSPLMTVGPL RVIDEGGVXLLGAPIGSPSXMAQYIREXLNT CKTALAHLDDHIPEARMRFLHRVSASACRL QHLFRLVPPDFAXPFAQQFDRDQLXAYXRF NSVTMSPRIVPKYGCXFTXATASPHWHL YTPXTLLASSIPLQHGYKVPTFPPSLSISVLHE ARCGSFFEIYLHSXNPHTSRCDYVAKNRAQI RLXFSHGGLTSLASTIHASYAASLIDTAP
--	--	--	--	--	--	---

						ARLQGPFPFPAVSQYQRFARGPLRVVLRNLP SFVQPAHFSTMTXDLGCLEPXALLARPERIH TFLQQAQYSAAASSYWQXPLWESFPNPGDH SAASLRKRVRYNLLAPGATSFLTAPHAATS RVHNATWSTMLRRHLDAPVTNDSISPLRCX HCSKPMDARGDHAXIXSHGFGTLHRHNTVR NVLARQLFRVAGLAYSLEVPFLIPNTAARPA DILVQPPPPAPGLPPDXPTAYDVTICSPFRRG MLYHAARHRGGAADAASVRKXKALERTIR XALLIEDDNXPPPLDWHFQPLSFDALGAPSQ STVHVIEDHAKLMALRNSCTIATAKSRIQQR LSFAIWSSAAAAILSRLPHTHAADISYPIEV (SEQ ID NO: 1399)
CR E	CRE- 1_ACas		Acantha moeba castellani i	TAACCCTAACCT CTCCCTCGGCCCC TCTACCCTAAAGC GCCCTAATCGACC GGCGACGCCCTAA TCGCTACCCTCTA CGCCCTAATCGAC TTTGGCGCCAAAG CGACTTTCCCCGG CCGATTTTCTTCCT	TAAGCCGCGCGACGAG GACGGCCAGGACGACC AGGACGACGGCGACGG CGACCACCTAGCACCG CACGCGCCACGACATA TTGTCGCGCGCTGTACA GGCGGCTAGGTCGAGC CCAGCCGACCGTTCTG AGCCTCAGTCGGCTTG AGCCCCCGGCTTCCCA	MATTTISRSPSSSSSSSSARSRASASTSASVAS IPRLFRDGRFHCPLAHCQTRTSTWQDLSAHL TRMHDGDVPRDVAAACGIVQCLHEGCRKW FRGAAGLASHRGKARHAPPPAPRAALAVAA VPRADSRGRTPAPTPSVAPPXAGPPPRAAPR AAPSP LPCPPALPHPPPSASPPTSSVTSPCSPP TTPPSQPSPDLFSGFANAPTTSPSPSTPXSSPA GSPIPAARRFVLPVATPYAPAPRANRPKLSP VARPFVPKARAGAIPEASSPVT PQDRAVSRR EDAAAAPSSAPGLGLADEHEDDDTYGGDTI

				GCCTTTTTCTTTTC TCTCCAAGCGACG CGCCTTTTACTTTG CCGCCGTTCTGTT TTTTCTTTTCTCTT TGCACTTCGCTTC TACTTCACACCTC CTCCTCCTCCTTCT CGACCCGCGCGGC CTCGAGCGACTTG CTGCAGCGGCTCC CGGCCTCCCCCAC GCGGCCTGCTACT CCCGCTTTCTAGA CGCCCCCGGTCTT GCTCTCAGTCTCC CGCATCGAAGCGG TAGTCGGGGTACG TGCTCAAGTGA CAAGCCTCTTTTC AGCCTCGGCGCTC TCTCAATCCGCCT CAGTCTTAGCCTT	AGGCCTACCGGGGCGG CTCTTTTTTCGCCCTGGT TTTTGCCGGCCTGTTTT TTCTCTCCCCCTTTTCC CCCCTTTTCCATTTGTA CTTAGTTTTTCCTTCGG CCGCGGCAGCTTGTTGC CCGGCATAGTGTTAAT ATGTTTAAAAAACGTG TAAATAAATAACTGTTT AACCCTAACCCTAACC CTAA (SEQ ID NO: 1278)	ALTAPHAPRETRAPFEFEACFLEEEAPATAG DLPPYARAFLACPSARLQEIPRRLKSAWQAA AKTIAEAALDCHTAGDTQGYNAHLRLFIELP ARGLAVPTNCRGAARTKLQRRERLLDIAAGR IPAIPDPPCDAPGADDALRGFPVSGTTAGDV SNDDDSGGVHDRPAATASARQAKRLVEQG LSSRALRALERGEPAVASADTLGRLEALHPP NPTDRGLWPGAPKAAIPRVTAKHLAQVAKE LPRGSAPGPSGWTFELVQAAIDRQPTGTVA AFLIDMAQRALRGTLHWRGLLTASRLVALK KPDGGVRPIAVGEALYRVIGRLVLKADRVM SSADATQYVGRHQYGVAYPPGGVEAPVHAV RELHDSGQLRAVVSLDWRNAFNSLDRVHT ALLIADRAPALARLYEWSYREDSVLVLPRA FEKAGLPASLLSQAGVRQGDVLGPLFFAIGA APVLDEIDAIPYVTPRAYLDDIFVTIPHGVTD AATKAAVAATFATAEREGAAAGLRLNRCK SAVWAADAEALLPPHAAGAREDVESCAPV REGLKILGAPVGSPAFVAKSLDGIIKRAIGTL DLVADAELPLQHKLVLLRQCVAQIPTFWAR AVPDAGPALAVWDTALLRRTGALVGLDVR DGSLQADIARLPVRLGGLGLRSMKDTAPRA FVASILFAAALANTRRSELTCSASTARRLRA
--	--	--	--	---	--	---

			TCAAGTTGCTCGA TTACGCTCTCGAA TCGCTCTCTCTC AGTCTCAGTCTCA GTCTCAATCTCGA TCTTGCCCTTCGCCT TCGTCTCGACGCC TTGCTCTCGWAAT CGCTGCCACTACG TGCCAGCTTTTTTC GTGCCTTGTCTTC GTGTCGACCCGGA CCGTTTGCAAGCC CTCGCCTTCGTAC CCCGCTCTCGTAG CCGTTCTCATCGC TGAAGCGTTCTAC GCGCTGGCAGCAA GCCTCGGCCCTAG CTTGTAGCGCCGC CGGTGGCCGCTCG CCAAC (SEQ ID NO: 1155)		ALPELARTDACNDEAAWRRSIARGVFPD KLGTTQLQRVLQGMADSKSAHRTRRQVPFL FAAVFEDAATPGSGAWLAAIPSDPTLVLPDA ELAEAVRIKLLTTTANAAGVCPACHKTGIDP SHAYTCVSLSHLRTARHDVVRRVELACKT EKPVREHVLAIPPVAPTDNNNNGDEDGSPV TTADDNADGHAVATKRRPETRASARAAAA AATAAAAAAIINDNSLLSDDDDDDDDHDDNC HGEERGEGERNVTCPGHYTATPFAADDTL NSDEDNEDNAHEDDDEDGKDDNDDDVYN NCNSSSSDGDEGGDDLDEYSDQSVTRSD AATGESPNPERPTTPTRALLRADLWLPATST AVDVMVAAACRRSRAKAFDRAVSRKAAK YGPAVADGSIKVVPFVVSFPGVLSRPAKAF LKRAMGDTTAAKQAKARLRLAVALVRGTA RLSYAWGACAALIVGGN (SEQ ID NO: 1400)
--	--	--	---	--	---

CR E	Cre- 1_FCy	.	Fragilari opsis cylindrus	ATCAATCTAATAC TGAAGGCAATACC AAACCTCAACCCGA AATCAAAATCGTT AGAATCAATATAC GACCCCCGCTGCT GTACATGTCCAGC CGGATCTCGTTGT AAAGAAGATTGCA GCTGTAAAAAAGT TGGACTTCTTTGTT CTTCTGTGAAGA AGTTGATTGTGGC TGTTCAAATTCTTT TCATAATAAAGAA TTA (SEQ ID NO: 1156)	TAGCACCACCATCTATT CATATCCACACACTGA CCACCTCCACCTTCACA ACTCCACTCTCAATTCC CCTGACTACTAAGAAA TATTTTCATGGTGGTTAC ATTGGAGGTATCCCAA CACCAAGCACAAAATG AACCCACTAACCCTCTC ATCCTATCCACGGGGA CCACCTTTGAGCAGAA CACCATTTCATATTACAA CCTTTAGCTAGATTAAG ATAATTATTTAGTACAT ATTTTATACTATTAAAA AAAAAAAAAAAAAA (SEQ ID NO: 1279)	MAPLPWNAATSSPPSPVPLTNDKKKDSLPT ATSKNLSKNNNNKNNNTNRINNIKNNNDNTN DGSNKINLKLPPAAVKITNPYKNKKKKNKKK NNAGKSNPKTNQNPNSSPLSDNDDDDTDSS NITINRRLKFGTDDLAPPNPPSNTNTIGTATA ATAATATTTATAATATTATNTTTTTTTTTNN TTGDNLASNINNNNNNNNSGSNNSNTNNIN NTDGNGSNRPPPRVYTVDP RSDLP GAEISA ANKMLDEVYGDHVHDNPGSHLSGLISSSQD QLWQGYFRRLLIPHNQSLYDCPKGKLGKDIT NEYSNLFEAIMNGKCNMEKLLVFPVVVLQR RHGVTKNADVKKRRLLSRLTAWKEGKFKYL VEDTHRDLIAKQSKARGDTTPAHRKAVYSS KLMRGHLQSAVNYITDREGGGILYPYDVDE KSGHTVSRVLQDKHPSMRDPGPAMPAYES VPELPTLEITADTVEIVAGKLSGGAGLSGVD SIQLKHLHLLHHGQASQRLRNVC AKFGRWLA NEHPPWASYRAMLANRLIALDKMPGIRPVG IGD TWRRFFAKLV LAVSMSYATDCCGSDQL CAGLRAGVDGAIHGLSAMWREMESEENTG FVLIDADNAFNEVSRINMLWTIRHEWPAGA RFAFNCYRHHSLLVVRNPGGKPF TFFSKEGV TQGD PFAMIA YGVALLPLIRKLKELNVLLVQ
---------	---------------	---	---------------------------------	---	--	--

						SWYADDASAAGKFDEILRLFQDLLRMGPDF GYFPNASKSILITHPDNVVAAHHFFNETHGL GFKISTGSRFLGGFIGDTTSRDEYVSTKIADW IHGTELAAVARLKYPHAAYTGITKCLQHK WSFTQRVIPGIDDLFQPLEDELNNLLPALFG DPPSTMDDKLRLLTALPVKHAGLALPNPVT SSATNYKNSTLMSSHLLAVVQGKINFSLQDH RDTCQSSLSASREL RQTENDSSLTNLLAALP PAAAGQPSTTRAIKRAGETGLWLTTPNHIN GNILGCDEFIDAIRLRYQKVPHNLPAKCDGC GSAFDVGHALQCKSGGLIIRRHDELNLELAS LAKMALRESAIRAEPEINPSASIMDSPTTITAI DTNGDRGDLLIKGFWDNGMDAIDVRITDT DAKSYRTRDPKKVLQSQEKEKKKKYLDQC LLQRRRAFTPFVVSVDGLIGYEASNVLKQLSK RLADKWNKPYSVTCGIVRSRISACARASNQ CLRGSRIPFKTMSRQIQWEDGAGAGLYRIVR (SEQ ID NO: 1401)
CR E	Cre- 1_HM	.	Hydra vulgaris	TTTCTAATGTTAC GTGATATGATATG GTTAGTTCATGGT TAGTTTATGTTTAT GCTTAGTTTATGG	TAACTTGTATTTTAAA TTGTTTTATTAGTTT (SEQ ID NO: 1280)	MNMVSICKRCDRSFTTLKGLNIHKGQCKIFV SNTNKQINNVNNELTTPNKNKVEINTILNC DEISVEHYSTNTPYLPKINICESIIDPN DY LW GHMPFSFLLNHVNTIYDEIVFYHKNLFKVPS GKGGKMFIEELTFWLKQFNNRTKLNGIAMK

				AAAATCGTTTATT TATGGCACAATAT TGTTTGCTGTTTT AAATTTATGTAAC GTGTGCATTTGAT GTATATTCTTGAA CTTTTAAATCTGA ATTTTACTTGGTT TAATACGTTTATT ATATTCTTCGATT GAGCAATTTATCC TATCAAAGCAATT TATCCTTCGATTC GAGCAATTTATCC TTCGATTCGAGCA ATTTATCCTTCGA TTGAGCAATTTAT CCTATCAAAATTA GCATATATACTGC AATTTTCAAATAA TCTACGAAATAAG TTCACTTACTGAA AATCATTAAGTAA		CFMIVPSLMLQKPSIRSKAKEHAECVRRITL WRNGNFSELMREIRYIQSKINTSKKKRTFEDI SRIFAKLMMEGKVAAALKVLDRESSGILQC SESVLKELKSKHPDETPVQDNCLLYGPLQNT PECLFDSIDEISIFNSALQTKGSAGPSGMDAD LYRRVLC SKCFGPSCKTLREEIATFTKNIATK SYQPDIVQPYIACRLIPLDKNPGIRPIGIGEV RRIVGKTISHHCQKEIKEAAGPLQTCAGHGA GAEEAIHAMQKIFHQEDTDGVLLIDARNAF NCLNRSVALHNIQITCPILAMYL VNTYRKPA KLFIYGGETIFSKEGTTQGDPLAMPWYSLST VTIINTLKL VIPDVKQVWLADDATAAGKLQ SLKKWYKCLEDVGGLYGYVYNQSKCWLIV KSDNQAEEAKLIFGNSINITTQGKRHLGAAL GSEAYKKVYCEDLVSKWSKELNNLCEIATT QPQAAYS AFIKGYRSKFTYFLRTIEAFENFVT PVEKILSEKLLPVLFGTDCSIIKENRDLLALN PSEGGLGICNLITEAKEQHTASKKITNLHIKSI LDQSDVMKEKDDFGKTFSEIKTKTNMDKSK KKKEEVKKIHAGLPENLKL LVEQACDKGAS SWLNTLPIKEQHLDLNKEEFKDALRLRYNV PLANLPSYCACGEKFDELHAMSCKKGGFVC NRHDNIRDLLTVCLNKVCTDVQAEPHLIPT
--	--	--	--	---	--	---

				AAGAAGAAAGGA AGAAAAAATAAA AATAAAAAGTAGT AAATCCTTTCATA ACAATAATCATTC TATTATTAAATTT AAAGGAATATTTT GGTTTTGTACTAA ATCATGCGTTCAT ATTCACCGAAGA AGGGGGCTGCTAT ATTTTTGTTTGAA GTTGTTTATCTTA AACTTTAAACTT GTGTTCAACCAAC CGTAAACATTAGT TCGCTGTTGCTC AAATTATCTACAA TATAAAATTTATC AATCTTTTTTCGTT ACGGTAAACAATA AACATAAAATAA CTATAGTTATTTT		NEKFNFKTANTNDEARLDIAKGFWRKGET AFFDVRVTHVNSKSSKKQPTKHIFRRHEDA KKREYLERVLEVEHGTFTPLIFGTNGGFGDE CKRFTALLAQKLSLKMGERYGAVINWLRTR LSMEITRASLLCLRGSRTPFRRHYNTDDVGLE NVQCGLI (SEQ ID NO: 1402)
--	--	--	--	--	--	---

				ATTGTTTACCGCA TATTGTTTAACTA TAGTTAAACAAAG TATTTGTTTATGG AACATTACCAGTA TCTCTTGTTAAGG TAAACAACAAAAC ATAGACGGCATCT CTTTTAAAGGTAA TTAAGTATACGGC TAATAATAAAAAT ATACAGCTAATAA TAAAATCTTCA (SEQ ID NO: 1157)		
CR E	CRE- 1_LSa		Lactuca sativa	ACATTAAATTAGA GAGGTTGATGTTT CAATGGAAGAAG ATGAAATTCCAAG AAGCTATTTTGT TGCCCAACCAAGTG TTTGATAAAATGT CCAAACTAATTTT TCTCTTGTTGCAG	TGA ACTATATTTTATAT ATTAAAAAAA (SEQ ID NO: 1281)	MASSSTSSSDICLCPFRSFHCCPNGEVGSKGI XRMISHIKRHLLTEDRKCVLREALSSDVGL FMAVEETLKAFGQWMCCKCMTLHALSRYC HHPDGRVXFVTGADGSSRYIVGILKPSTKES VTNALGGLVFDVGLLDVRFKEPITTVKSIPH SCRLAFSQALKTALYKVIAQPGSVDWICLL LLPRCTLQVFRPKNRQECRSGNRKSLQQSSI LKSLDTWGKEDGIRKLVQNMLDNPEVGAM GQGGGILQKESTSSNTNIRQCLRKVADGHFT

				CTTTATTGTTCAA GATAATGTAGTTT GCTTAGTTTGAGC G TTCCTTGTGCAC ACCAACAGTGTGT TGGTGTGCCATTT CCTTTCCTTCCTTT TTAACTATTGCTT CATAGCTTAAGCT TCATCTCGAGGCT TGTTCTCTTGT (SEQ ID NO: 1158)		AAVKVLCSSGVAPYNGDTIKALEDKHPFRP PPSMPSPII SE PPLVADFD CVFGCIK SFPKGTS CGRDGLRAQHXLDA LCGEGSAIATDLIRAIT SVVNLWLAGRCPTILAEFVASAPLTPLIKPD NGIRPIAVGTIWRRLVSKVAMKGVGKEMAK YL NDFQFGVGVSGGAEVVLHSANRVLSEHH ADGSLAMLTVD FSNAFNLVDRSALLHEVKR MCPSISLWVNFLY GQAARLYIGDQHIWSAT GVQQGDPLGPLLFALVLHPLVHKIRDNCKL LLHAWYLDDGTVIGDSEEVARVLNIIRVNGP GLGLELNIKKTEIFWPSCDGRKLRADLFPTDI GRPSLGVKLLGGAVSRDAGFISGLAMKRAV NAVDLMGLLPQLCDPQSELLLRSCMGI AK LFFGLRTCQPVHIEEAALFFDKGLRRSIEDM VVC GGPFFGDIQWRLASLPIRFGGLGLYSAY EVSSYAFVASRAQSWALQDHILRDSGICGM DSDYLCAMTRLRDTIPGFDCSGFTNKDTPPK SQKALACALFSKIVKDMEVDFDMTVRQKA VFECLRAPHAQDFLLTIPIDGLGQHMS PVEY RTILRYRLMIPLFPIDEICPVCRKACLDTFGE HAVHCRELPGFKYRHDVVRDVLFDACRRA GISAKKEAPVNFLTDPQDGRSTLRPADILVF GWVGGKHACVDLTGVSPVLGLRSGGFTAG
--	--	--	--	--	--	--

						HAALKAAACKVAKHENACIENQHVFVPFAF DTFGFLAPEAVELLNRVQRMHNSVISPRST DVVFKRISFAIQKGLAAQLVARLPSIDMY (SEQ ID NO: 1403)
CR E	Cre- 1_MB		Monosig a brevicoll is	CATCTTGGCGTG ACCACGTTGTCAG ACAAAATCTGCAA CCCCGCTCTTTGC GGCCCGCGTTTTG GCGGCGCCCTCGC TCCCACCGTGTCC GCTCGCTTGCTCG CTTGCTTGCCCCG CGGAC (SEQ ID NO: 1159)	TAGGTAGGCACCGTCT CGGGGGTCCCTCTGTG GGGATCCCTGTGTGCA CCTGTGCTCCCTAGGT GGTTCCTCGTTGTGTCT TTTGATGGCTTGACTTG TATTTTTGTTTTAATTTT GCTTTAATTTTTGCTGT ATTTGTGTGGTATTTTT GCTGAATTTTTGTAAGG TCCTTTGTATGATGTCT TTGTCTTCTGTGGTCGG TTGTTTTCTCAATCCG ACGTTGTGTCTCGTTGG ATGTGAGCGTGCCGTG GTGTTCTTTGTGTTTGT GCTGTGATGGCTTGTAG TTGTGATGTGTGACTGC CTTTTGGGTGTCTTGT	MATESGGEDSWTQVRGAKRPSAESPPSNTT TSPSQTHRS AKHTKHGSARHDRNHVFPDPM TTPLRPHARHSVPTARASSHVPSTSPAAGAT ESSARAVVPAEPVTRTSNGGGEQHPIIGNT SNASPRTPRTPSSPRSFAQVAAAMPAAATAT SSAPMTEDLSASVPSEPNNGSGEQQPSPESTG QTHHSIPNTPSDFLTMSDES DSPPRSTALRA PTPIAPPAHDGDGDTNGSATPEPLVQSPTPA QMVLPYPSGTQQTHSDPSPPSASPPATTILPA AISHPVEHSEHANSAPLGEVSESETHNTAGE HSESEQDVLLSDPAPPAAANVLDAQRKVLLK TSGHRQLLACPFGLCKCKGPRLDRKAWVN HVLREHPYDEQATDLVKQVMEAKLVAQCN KCHLFFEAAGISQHRSRCGANLKRATEALFH AAGHDLLEIMRGAWPQQCVGSRISVCELLK LARHPLMQRSRYP SNATETKLMAATLSQLY WSAVHSDYTAEEREMCWALILALPSMLLSA PSTALSTIDLRNMFHDRLRWLV TGQLGRVV DAMRKAVARKQSRRGQLNAGAG AHPNDA

				GTTTGAAATGGCCGTAT CTCTGGTTATACTTGGT CGTTTTGTACGATTTTT GTTTCTATGTGCGTGAT TCTTCGCGCTTGTACTT CTTGGCATGATAGAAG CCAATGAATGTGTCTTG TTCTCTTGTGTTGTTTT GCGTGCCGTCGTGATTT TGATGTCGGGGTTGCA CAGCTTTGCTTTCAGCT CTGAGGTTCAAACACC TAATTTA (SEQ ID NO: 1282)	VDQSLRSLVRDPDLADEAWANHV TNRLNR GQIAKAFDADKARAVIGNSEVQAVRDLLVP PGLTPYIASTPASTSTLAPATAVSSPTVSFTK GELPKALAATKGVTDPYGWSGELLASIYRIK EHFSQVLGPRQGSTSDPTAPSDGDAPQGPTT ATGGPQVALNKIFHHIANNTVPESIRHALCSI NYTILEKANGKFRPVGTD SIFNKVVNRALLE QQQPHIAHLLQASPELAVGVKDGISA AVGM AFGELQACESTPGWTMLSLDFKSAFN YTDR ARLHEIVADKVPGLLRAFERHYARPTTHCIV DKFFKVIDIDVGQGIVQGNELSPFFFALYSCE VLGLLDATTDYRCKVIKYLDDIVLMGPAED VAADVEIVKARAESAGLHLQPSKSRFYMPR HHSASITAIKSVLPDAVRETANTGMTVLGTP IGRREWMKKQLNDKAKHIAGKLN DMLTTG VSLQALLTAMQYVPSLINHLYTLPPSLTSGL SELLNRACKDTFVKAFFAKVNLSAPAGAEG HDVTLEQLLEARLFTRANTGGFGLHDLVER GPVAYVCNMAKLATRYPRVYDRLL EDASR AADFEAHVQRAGFQMATVKDAATQRP AEII ALRSKAALDDLMAKCALDLQQAYLASREW GVSTVL TMRGRDKLRRLSDTTFAIAVV SMM GFGLHELINVKPTDKCPLCSSKTPQPRLTRE
--	--	--	--	---	--

						HLLTCRPIKRHNALRDEMGRLLRYATLSHV WVEKSGYNANGQSCRIDLHCRNPFPGGALG PALPDLGIDVTVRTAQPPPTSQACIKVGAAL RRAEKEKRDYYTGFNHGKTLIVPAAMTTTG GFASSFVDLLGQLARCAEARGVYQPGLDEA FVPRWKGRFAALVHQMNAADHIQRHFGGVC LRSS (SEQ ID NO: 1404)
CR E	CRE- 2_HMa		Hydra vulgaris	AATTTAAAAAAA AAAATCGTTTATT TATGGCATAATAC TGTTTGTAATTTT GAAAATTCGTGCA ACAACTGCAGTTA AATTGAAGAGCTG AAATTTAAGATCT GAGCTTTTCAATC AGAGTTTTTTACC CTAAAACATTAAA TTTTATCATAACA AAAATCGTTCTAA TATTATTAACTT AAAGAAATTCGTT CTTATATCAAATC	TGAGCTCTTATAAATT ATATTATAGCATTTTGT TTTA (SEQ ID NO: 1283)	MSSCKVTIPHVCPYCKVELKTICGINRHILKC KKNPLQIPSLQKTNTSLTLEPNTKVIPSITKQ NDIIASTSSNNLAFNQKKDYTLTPTYSRKTT PVSILSSMKMTPISITSHIVRRKLPELPSQTTN HLFNENFINVPFLPEIMNHLVPNNNVMWG VYSYQQFKLFVDSTYDEIVNYRRNIFNIPSG KAGKEFIEELTFWLRKFNSTSSLNSIALKVT MILPNLLLQKPSAKSKSKEHTLCLTRRIDLW KKGDTSLLLKEVRNIQKKFVNSKXKRSMD ISRIFAKLIMEGKITAALKFLEKEASSGILPLS DNTLKDLKSKHPEPSRVEDYSLLFGPIDLIPK CFFDCIDEQLVMKAAFATKGSAGPSGMDAD IYRRILCSKNFIKEGKELRKEIAKMTQNLLTE TYEPTFLEAFTACRLIPLDKNPGIRPIGVGEV LRRIGKVISWSFNSEIKEAAGPLQTCAGHGA GAEEAVHAMKEIFDNVQTDAILLIDAKNAF

				TTATTTTCAGTGTTT CACAGACGAAGG GTTTTACTAGATT TTTATTTTTTTCAAC TTTTGAATTTGTTT ATTATAAACTGT AAACTAGTGTGCA ACCAACCGTAAAA AWTAGTTAGCTGT TCACCCAAAATAT TATTCAGTATGAA AATATTTAATCTT CTTTATTCGCAGT AAACAATAAAATA TCTAGTTAAACAA AATATTTCTTAAT AATAAAAAACAA AAACTTTTTCTTA ACAAGTACA (SEQ ID NO: 1160)		NCMNRQVALHNIQIICPLISIIYLINTYRNPSRL FVAGGKEISSQEGTTQGDPLAMPWYSCNTTI IIEHLLVNYPQVKQVWLADDAAASGSIANL HSWYQHLIDEGCKHGYVYNQSKCWLIVKSP SLAENAGIVFGKSVNITTEGQRHLGSVGTSQ NFKNKYCTEKVAKWLTELKQLCKVAETQP QAAFIATKGFRRSKFTYFLRTIPKFEQYLAPV DEILSHLLLPTLFGKDTPFEDHIRKLFTLTPR DGGLGIPILVEEAPHQFLSSVKLTKNLVQQII DQDKILKTKNSSGNVLEDLEKILTTDRLKHR KEKIIAVDSMQPDSMLRNIQQTRSECASTWL NALPLENQGFVLNKEEFRDALCLRYNFDLK NIPRICEGEPFNVTHALSCKKGGFISSRHDN IRNLFTTLLKRVCINVQSEPHLIPLDNENFYF HTANKSNQARLDIKANGFWRNGQTAFDVF RVTHVNSMSNKNLDIAAIFRKHEKEKKREY GERVREVEHGSLTPLVFGTNGGMGKECHRF VRRLAEKLAEKQNEKYSVVMTWLRTKLSF EILRSTILCLRGSRTPWTKKNDFEIGVDFKM DALEARI (SEQ ID NO: 1405)
CR E	MoTeR1	JQ747 487	Magnapo rthe oryzae	CCCGAACCCGAAC CCAAACCCAAACC CAAACCCAAACCC	TAATAGGTAACGTCCCT ATTTTTGTCTTTGGTTTT GTTTTTATCTTTGTTTTT	MVCPTCNGVYADYNDHIRKKHPDERYTAL QLQPLGLTPCPICKTACKNDLGVKTHLSKIH KISGASKISTQPRIRTENTDNTNSVPTSSFNVP

				AAACCCAAACCCA AACCCAAACCCGG AGGGTTCCCAAGT CGCCTAAACCCGA AGGGTTTAGGATA TTATTTGTTTATT AGAATTGGATAAT TATTTACCCCTGTT GGACAGGGGGGTT GCAGGGGTAAAT TAAGGTTTTTTATT ATTTATGCGCCGT TTATTTGTTTACCC CCCCAAATATTAT AAAAGCGCGTTCC ATCCTCTTAGGAA AAGCGAAGCTTTT CCTTGTAAGTC GCTAGACTTTTAC TATAAAAGTCGCT AGACTTTTATACC AATCTTTTAACAA AAAGCGTAGCTTT	GTTTTGTTTTCGTTTT GTTTTGTTTTCGTTTT GTTTTTTTTTTTGT GTTTTGTTTTGCCTT GTTTTGTTTTATCTT ATTTTGTTTTGT ACTTGTTTTATTTGTT TATATTTACCTTTGAT TTTTCTATTTTCCCAC CCTTATTATTATAACC CAACCTACTAATATTT TTCTTTTTTCTTTTTCT TTTACGGTTTTATTT CCCGTTGTTTTTCTA TTTTATTTGTACGACAA AACCCTAGCAAATAA GCTTAGAATATAATAA AGCGCGAATTAAAA (SEQ ID NO: 1284)	LPEIQTLPGLNNSRWADNPRKRRADTPSPT RGRNTRPRRFSYTDIDLTNDEPADNPRANNP RVNNPRVNNPPSSPNSLPSISEFHTPGTLPL TNSNISLKDQHDKITGPILQKPLIQKLIEYSKI PIPEHHLHARQAKIFADAANRIAKNFIQSPT KTLFNLLILPRIFGIGLINGKVTIMQNFPSQI PIPKIDFPSEKTDSDPVLNAKKLLEKGYIGR AAKAIDPTPVAPETPESLNILREKHPIGQNN PFNTKSQPISGRQITEKAILLAISSIGREKAPG LSGWTRSLLDAAIKIPTQNDVIPALRLLTDMI RQGTAPGRELLCASRLIGLSKPDGGVRPIAV GDLLYKIAFKAILNTLWSPNCLLPYQLGVNS IGGVEPAIFTLEEAIMGPNINGIKSITSLDLKN AFNSVSRAAIASSVAKYAPTFYRSTCWAYN QPSILITENGSVLASAQGIRQGDPLGPLLFSL AFRPTLETIQKSLPYTYIAAYLDDVYILSKTP VKDKIAKIIKSPFTLNSAKTTETDIDTLKTN GLKTLGSFIGPTELKKEFLQNKIQNFESSINA LKKLPKQYGLLILRKSTQLLLRHLLRTLNSQ DLWELWEKTDKLIADFVINLTVTKRKRKPIT DFVTPLITLPIKDGGFGLLRHNGIAQDIYFAA KDLTTEIRHKIQRISNDFPQNQSPTATEILHLL HNGVLADCKNGLTNAQLNALTENASYLGR
--	--	--	--	---	--	---

			TTGTTGCCAATCT ATTAAAAAAGCG GAGCTTTTTTTAA CTTTTTCTTTTTT TTTTTTTTTCTTT TTTTTTTTTTTTT CTTTTTTTTTTTT TTTTTTTTTATATA TATTATTATTATTA TTATTAGCGGTGG GGCTATTTATGCG CTTTAATTTGTGC GGGGCTATTTATG CGCTTTAATTTGT GCGGGGCTATTAA TGCGCTTTAACTT TACAAATTTTATT TATGCGCTTTAAT TGCTGCGGGCCTG TTAATGCGCTTTA ATTTACAAATTC ATTAATGCGCTTT AACTTTTATATTT	KWLNILPIQKSNRLTDWEMAEAVRLRL LAP VKPLTHPCNHCGNRTNINHEDVCKGAVRKY TARHDQINRSFVNSLKS RPEIDVEIEPDLNNE NNVNNANTTTENPTSPNGQNDTGCLFTTPI RSGTRNGQNGLRADFAVINGVSKYYYDVQI VAINKDSGNTNPLNTLADAANNKRRKYQFL DPFFHPIIISAGGLMEKDTAQAYKQIQKLIGP VAAHWLDTISISLILLRSRTTAAISIAKNRPRA (SEQ ID NO: 1406)
--	--	--	---	--

				ACTAATGCGTTAT TTATATAATTGCT ATTATTATCGTTG CTATTATTATTATT GCTATTATTATCG TTATTATTATTGC AATTTTATTATAT AAACCCTCGTTTG TCCCTCGATTTAT CCCGTTTCTTTTCC ATCCCATCGCGCG TTTTCGTAAGCTTT GGTTTTCGTAGGA TTTGCTTTCGTAG GCTTTGCTTTCGT AGGCTTTCGTCAG CTTTTACCTGCTTT TATTTTTTCTTTTT CTTTTATTCCCCC CCCTTTTTTTTACC TGGTTTATTAGCG GTTTACCTGCTTTT ATTACCTGGTTCC		
--	--	--	--	--	--	--

				CCTTTACCTGTTTT ATTAGCGGTTTAC CTGCTTTTATTACC TGGTTCCCCTTTA CCTACTTTATAAG CGGTTTACCTGCT TTTATTACCTGGTT CCCCTTTACCTGTT TTATTAGCGGTTT ACCTGCTTTTATT ACCTGGTTCCCCT TTACCTGTTTTATT AGCGGTTTACCAG CTTTTATTACCTG GTTCCCCTTTACCT ACTTTATTAGCGG TTTACCCGTTTCTA TTAGTGGGCATTT ATTTCCTGTTTTTA TTAGCAGTTAAAT TTACCCTTTTAAG GTTATTTACCTGC TTTTATTACAGG		
--	--	--	--	--	--	--

				GCACCCCTGTTTT TACTAGCAGTTAA ATTTACCTTTTTTAA GGTTATTTACCTG CTTTTATTCACAG GGCACCCCTGTTT TTACCAGCAGTTA AATTTACCTTTTTTA AGGTTATTTACCT GCTTTTATTAACA ACCCTTTATTTTTT CCTATTAACGGGT ATTTATTTACCTGT TTTATTGGAATTC ACCCGTTGGACGG C (SEQ ID NO: 1161)		
HE RO	HERO- 2_BF	.	Branchio stoma floridae	TTTTCAGTCTGGC TCAGCCAGTGACC GCCGGGAAAGTCC GGCTGACTACCAC GAATAGGGTGGTG ACAGCTGGATAGA CAGACGACAGCTC	TGATTAAAGACCCGAA ACACCCAATGACCCCG GGTTCATCACTGATGAT GTGTCCCTGTTCGCACT ACCAGAGTGTATTCTA GAG (SEQ ID NO: 1285)	MNAVCVCGKVCKNQRGLRIHQTKMACLRR VQAEHRSGAVATTVEPVLSASAPGQTEEDQ GPEAPHSARNLRATPAPPQGRKSDHHRVKW PAANSKEWSQFDEDVDMILESVSRGSTDQK LQSMCTVIMSMGAERFGTIGQRKPTDTMKP NRREVKIRQLRQELKSLRRSFKASTSGEERA ALAEALTHLREKLRTLRRAEWHKKKGKER

				GGAAAGACGGCA TTGGGGCAGTATG GGTTGGCACCCCT AACTGCATCTCCC CTAGGAGAGCATC CCGCAACACGCTA CAAAGAACCACA AAGAGCAATACCC CCAGGGATGCCCCG AGAGGGGGGGGAG GATGAGCATCCCA TTCGGACGGTCCA ATCGGTATTGACC CCAGCAAACGGA GAATCGACA (SEQ ID NO: 1162)		ARKRSAFITNPFGFTKRLLGQKRSGNLTCPV EEINLHLSNTFSDASRDVDLGPCPLL VTSPEP EVHFDISEPTLKEVRETVKAARSSSAPGPSG VVYKVYKHCPRLVVRLWRILKVWVRRGKV AADWRQAEGVWIPKEEESKVDQFRLISLLS VEGKIFFKIVAQRLIKYL LDNQYIDTSVQKG GVPGVPGCLEHTGVVTQLIREAKENRGDLA VLWDLANAYGSIPHKLVETALTRHHVPESI QNLILDYYSNFWLRAGSSTATSAWQRLEKG IITGCTISVPLFALAMNMIVKGAEAGCRGPV SRSGTRQPPIRAFMDDLTVMTATVPVCRWL LQGLERLITWARMSFKPAKSRLVLKKGKV AERFRFTLGGTQIPTVSEKPVKSLGKVFNSSL KDTASVQQTRSDLTTWLEGIDKTGLPGSFK AWMFQHGVLPRVLWPLL VYEVPM TMVEQ LERTISRFLRKWLGLPRSLSNIALYGRSTKLQ LPLSGLTEEFKVTRAREVL MYRDSSDSKVSS AGIHVRTGRKWKAQEAVDQAEARLRHSVL VGSVAVGRAGLGSCP KPRYDKVSGKEKRLL IQDEIRAGEEEDRRCRMVGM RKQGAWTRW EHADSRKV TWPELCRAEPSRIKFLISSVYDV LPSPANLHVWGLAETPSCQLCQRRGTLEHIL SCCPKALGEGRYRWRHDQVLRVLADTVSN
--	--	--	--	--	--	--

						AIQSSRSQPPKKSIVFVRAGEKTRQQPTSA GGLLSTARDWQLLVDLGRQLKFPEHIVATS LRPDMVLVSESTRQVVLELTVPWEERISEA NERKRAKYAELVVQSQSNGWRARCVPEV GCRGFAGQSLAYVLKLLGVRGFRLRKSIRDI LEAAEKASRWLWFRRGEPWKPHGHRSGND QPRLGRPGEVW (SEQ ID NO: 1407)
HE RO	HERO- 2_DR		Danio rerio	TTCAAGCCTGGCG CAGCCAGTGACTC CTAGGAATAGACT AGGTGGCAACCAA GAATAGTTTGGTC GACTACTGGAGAG ACAGTTGACGGCA CGGAAAGACGGC ACTTGGGACAGTA TGGGTTAGCACCC CAGCCTGTGTCTT TCGTGAGAGAGAA CCCAAACAAGCTA CGGAAAGCCCCAC AGAGATATACCCC CAGGAGATCCCGA	TGATCAACCCCGGCTG GGTCACCTGGGTGAGA GTGTATGATGTTGAGA GACCCGAAACACTCAA TGATCCCAGGATACAT CACTGATGATGTGTCCC AAATGCATCCATGAGA TGTTTCTTGCATAA (SEQ ID NO: 1286)	MTHANEQTTNKIYVTCICGKLCKNHWGLKI HQARMKCLEQESKVQRTGPEPGETQEEP EATHRAKSLHVPEPQTPSEVVQQRIKWPPAS KGSEWLQFDEDVSNIIQAIKGDADSRLKT MTTHFSYALERFGCIEKGKTKPTTPYTMNRR ATQIHHLRQELRSLKKLYKKATDEEKQPLA ELKNILRKKLMILRRAEWHRRRGRERARKR AAFITNPGFTKQLLGDKRSGRLECSIEEVNR FIEETVSDPLREQELEPNKALISPTPPAREFSL RGPSLKEVKEIHKASRSASTPGPSGIPYLVYK RCPGLLLHLWKILKVIWQRGRVAEQWRCA EGVWIPKEENSKNINQFRIISLLSVEGKVFFSI VSRRLTEFLLENNYIDPSVQKGGIPGAPGCL EHTGVVTQLIREAHENRGDLVVLWLDLAN AYGSIPHLKVELALHRHHVPSKIKDLILDYY NNFKMRVTSGSETSSWHRIGKGIITGCTISVI

				GAGGGGGGGAGG ATGAGATCTCCAA TCGGACGGATCAA AGGTTA (SEQ ID NO: 1163)		LFALAMNMVVKSAEVECRGPLTKSGVRQPP IRAYMDDLTTTTTPGSRWILQGLERLIAW ARMSFKPSKSRSMVLKKGKVVDKFHFSISG SVIPTITEQPVKSLGKLFDSLSKDSAAIQSK KELGAWLAKVDKSGLPGRFKAWIYQHSILP RVLWPLLIYAVPMSTVESLERKISGFLRKWL GLPRSLTSAALYGTSNTLQLPFSGLTEEFMV VRTREALQYRDSRDGKVSSACIEVRTGRKW NAGKAVEVAESRLQQKALVGT VATGRAGL GYFPKTLVSQVKGKERHLLQGEVRASVEE ERVSRVVGLRQQGAWTRWNTLQRRITWAN ILQADFQRVRFLVQAVYDVLPSPSNLHVWG KNETPSCLLCSGRGSLEHLLSSCPKALADGR YRWRHDQVLKAIAASLASAINTSKNHRAPR KAVHFIKAGEKPRALPQLTTGLLHKASDWQ LEVDLGKQLRFPHHIAATRLRPDI AISEASR QLIILELTVPWEERIEE ANERKRAKYQELVE ECRERGWRTYYEPIEIGCRGFAGRS LCKVLS RLGITGVAKKRAIRSASEAAEKATR WLWIK RADPWTAVGTQVGT (SEQ ID NO: 1408)
HE RO	HERO- 3_BF	.	Branchio stoma floridae	CTGACCAGCAGAC GGGAAGCCCGCG ACCAACTAGTCTC	TAGAAACCCACAAGGC TGAGAAATGTAGAGCA TCTGTATGGACAATATT	MALPAVRSGPASTWTLITLVIVA AKGTDGF MSFKLPLLSTD TWSGYNNDVK TLLGPLHHE LATNEMSPKLAGEGFSDIMCDFMASKPEFS

				CGCAAATATTGCA CACAGGGCGACCC TATGGAGCTGATT CAGTCAAATTTCC TCTGAGATATAACC GATAACTATCTAC AGAAACTGCACAG TTAGTTTGGAAG AGCTTTTCTACTG AAAGACAGCAAA ATCCGCCACTTTA GACGAGCGTCAAG ACTGCCCTCCCCA TAACCAAT (SEQ ID NO: 1164)	GATGATTGAAATGTTGT GATTTTAGATCAAATTT AGAAATATGAAAACCG AACTAAACTAAATATA ATGTTTTTTTTTAAAGTA ATGATAAGCAATACCC ACATTGTGCAATACTAT CTATGTTATGTCCTTTG TCCCCCCTGCATGTTTG GTCAATAATGACCATC GTGTCCTGGGCTCCGTG TACCTTTCTTTACTATG AATAAAGAATGATTTT ACTAC (SEQ ID NO: 1287)	HTTEESHSEGYISHEPQSLAQVKRLKNKLRK KAFRADATPEDRKAFRDAIKTYSFMKRQQK RKETTKSAAHQEKEYHKNFWKFAGKCAKG QLDIPPVKPAFSVYYYANEYYKNKYSHPTRV DFNKLLWPHLPVEEQLPANSFDMSPVRPK DIKAVLSKRCATSAPGPDGIMYGHLKHLPA CHLFLSTLFSKLLESGBPPTSWSNGVSLIHK DGSPEAAENFRMICLTSCVSKIFHQILSERW AKYMTCNLDLIDPETQKAFLTGINGCVEHVQ VMREILAHAKKNRRTVHITWFDLADAFGSV EHELIYYQMERNGFPPHITTYIKNLYSRLKGGK VKGPGWESDPFPFGRGVFQGDNLSPHIFLTVF QPILQHLKGVEQQHGYNLNDKHVYVTLPFAD DFCLITTNRQHQLITQISSNTKSMNLKLK PRKCKSMSIVSGKPSDISFTIDGDPVKTTKDA PEKFLGGYITFLSKTKETYDILAKTIETTVENI NKS AIRNEYKLRVYMEYAFPSWRYMLMVH DLTD TQLQKLDSIHTKAIKTWLRMQPSATN AILYNTRGLNFKSISDLYLEAHALAYSRSVL KADEKVKHALQAKLDRESQWTRKMQKWG IGKCHTIHQQAIVAKDSEWTSVRKHVKQQ VTDMRHDVWTKHQENLLQQGQMLQLLEEE KCDLTWRSAMYNLPRGILSFAVRASIDALPT
--	--	--	--	---	--	--

						LCNLTTWGKRNTDKCKLCGNRETLHHVLN HCGVALQQGRYTFRHNSVLKHITDTIIESIDT SRINATIIYADIQGYTTNGGTIPVHTIPTTQKP DLIIYLPEQKTLHIHELTVPFENIKTSHDRK VNKYSTLAADLETAGISATLTCFEVGSRLV TPENKTRLRTLFKIVKAKPPKTLFTDISRIAM LSSYAIWNSRHEPYWESETLL (SEQ ID NO: 1409)
HE RO	HERODr	.	Danio rerio	AAAGCAGTAGAG (SEQ ID NO: 1165)	TAGCATGCCACTTGGA CACAGGCCGGGGTCTG ATCAGCCTCGGTCGGG TCGCCTGGAGGAGGGT GTCTGTTGCAAGACCC GAAACACCCTGTGAGC CCAGGAAACAACACTG ATGATGTGTCCAAGGTT GTGCATCAGGAGATGT TTCTGTAAAC (SEQ ID NO: 1288)	MTTHRAEVTTS GKTQEPPGPEATHSAQSLL VSPTPAAGRSPATQSCPQVTAAHNSPQSPQS QQVAVTRSDCVPLAQPRIQWPQSSKKAEWL QFDKDVNQILEVTGKGGVDQRLSTMTTLIV NIAAERFGTVTPKPTSTYTPSHRVKEIKRLR KELKLLKRQYKAAGEVERAGLEDLRGILRK QLVNLCRAEYHRKRRRERARKRAAFLANPF KLTKQLLGQKRTGKLTCSKEAINNHLKATY SDPNREQPLGPCGALLTPPEPTSEFNMKEPC RSEVEEVVRRARSSSAPGPSGVYPYKVYKNC PKLLHRLWKALKVIWRRGKIAQPWRYAEG VYIPKEEKSENIDQFRVISLLSVESKIFFSIVA KRLSNFLLSNKYIDTSMQKGGIPGVPGCLEH TGVVTQLIREAREGRGDLAVLWDLTNAYG SIPHKLVEVALEKHHVPQKVKDIIIDYYSKF

						SLRVSSGQLTSDWHQLEVGITGCTISVTLFA LAMNMMVKAAETECRGPLSKSGVRQPPIRA FMDDLTVTTTSVPGARWILQGLERLVAWAR MSFKPAKSRSLLVRKGKVRDEFRLFRLGQHQ IPSVTERPVKSLGKAFNCSLNDRDSIRETSTA MEAWLKAVDKSGLPGRFKAWVYQHGI LPR LLWPLLIYEVPM TVVEGF EQKVSSYLRRWL GLPRSLSNIALYGN TNKLKLPFGSVREEFIVA RTREHLQYSGSRDAKVSGAGIVIRTGRKWR AAEAVEQAETRLKHKAILGAVAQGRAGLGS LAATRYDSASGRERQRLVQEEVRASVEEER TSRAVAMRQQGAWMKWEQAMERNVTWK DIWTWNPLRIRFLIQGVYDVLPSPSNLYIWG RVETPACPLCSKPGTLEHILSSCSKALGEGR YRWRHDQVLKSIAEAISKGIKDSRYRQATA KVIQFIKEGQRPERTAKNCSAGLLSTARDW VMTVDLERQLKIPPHITQSTLRPDILVSEAT KQLILLELTVPWEERMEEAQERKRGKYQEL VEQCRANGWRTRCMPVEVGSRGFASYTLS KAYGTLGITGTNRRRALSNNVEAAEKASRW LWLKRGEQWGQ (SEQ ID NO: 1410)
HE RO	HERO Fr	.	Takifugu rubripes	AGACTAGGTGACA ACCAAGAACAGTT	TGATCACCCCGGCTGG GTCGCCTGGGCGAGGG	MTPAMEMTTT VT CICS KLCKNQRGLKIHQA RMKCLEREVEVQRTGPGPGETQEEPQEAT

				WGGTCGACTACTG GAAAGACAGTTGG CAGCTCGGAAAGA CGGCACCCGGGAC AGTATGGGTTAGC ACCCCAGCCTGTA TCTTTCGCGAGAA GGAACCCAAACA AGCTACGGAAAGC CCTACAGAGAAAC ACCCCAGGAGAT CCCGAGAGGGGG GGAGGATGAGATC TCCAATCGGACGG ACCTAACGTTA (SEQ ID NO: 1166)	TGTATGATGTCGTGAG ACCCGAAACACCCTAT GAACCCAGGATACATC CTGACGATGTGTCCCA GTGCATCCAGGAGATG TAKCTTTAAGT (SEQ ID NO: 1289)	HRSQSLHVPEPPNPNRVVQQQRIKWPPANR RSEWLQFDEDVSNIIQATAKGDVDSRLQAIS TIIVSYGSEFRFGRIEKGNTETTSYTMNRRSFK IHQLRKELRTLKKQFKRAXDGDQKQALKELY NILRKKLKTLLRAEWHRRRGRERARKRAAF IANPFRFSKQLLGDKRSGRLECSREEVNRFL QNTMSDPLRGQDLGPNRALISPAPPSAEFKL AEPSLKEVEEVIKAARSASSPGPSGVPYLVY KRCPEILRHLWKALKVIWRRGRVADQWRC AEGLWIPKEEDSKNINQFRTISLLSVEGKVFF SIVSRRLTEFLLKNNYIDTSVQKGGIPGVPGC LEHNGVVTQLIREAHESKGELAVLWLDLTN AYGSIPHLVELALHLHHVPSKIKDLILDYY NNFRLRVTSGSVTSDWHRLEKGIITGCTISV VLFVLAMNMVVKAAEVECRGPLSRSGVRQ PPIRAYMDDLTVTTTTSVPGCRWILQGLERLI LWARMSFKPTKSRSMLVKKGKVVDKFRFSI SGTVIPSITEQPVKSLGKLFDSLKD TAAIQK STEELGGWLTKVDKSGLPGRFKAWIYQYSI LPRVLWPLL VYAVPVTTVESFERKISSFLRR WLGLPRSLNSAALYGTSNTLQLPFSGLTEEF KVARTREALQYRDSRDCKVSSAGIEVK TGR KWKAEKAVXVAESRLRQKALVGAVATGRT
--	--	--	--	---	--	--

						GLGYFPKTQVSHARGKERNHLLQEEVRAGV EEERVGRAVGLRQQGAWTRWESALQRKVT WSNIMQADFHRVRFLVAAVYDALPSPANLH AWGKSETPTCSLCSGRGSLEHLLSSCPKSLA DGRYRWRHDQVLKAVAESIALAISTXKHHH APKKAISFIKAGERPRAGPQITTGLLHTAXD WQLHVDLGKQLIFPQHIATTSLRPDMIIIEA SKHLIMLELTVPWEERIEEANERKRAKYQEL VEECRGRGWRTFYEPIEVGCRGFAGRSLCK AFGRLGVTGTAKKRAIKXASEAAERATRW XWLKRADPWVATGTQAGS (SEQ ID NO: 1411)
HE RO	HEROTn	.	Tetraodo n nigroviri dis	AGATTGGTCTGGC TAAGCCAGTGACG TCCAGGAACAGAC TGGCTGACGACCA CGAATAGAGTGGT GACAGCTTGGATA GACAGCTGACAGC AGGGAAAGACGG CAACCGGGGCAG GAAGGGCTAGCA ACCCAGCCTGCAT	TGATCACTCCCAGTCGG GTCGCCTGGGTGAGGG GGTCTGATGTTGAAAG ACCCGAAACCCCGAT GACCCCAGGTACTATC ACTGACGATGTGTCCA AGACATGCATCAATAG GTGTATTTAGAAATC (SEQ ID NO: 1290)	MATTQASVKPTAVATCVCCKICKNPRGLKI HQTKMGCLASVQPEQRRFSLSESREVPAR AEPYGPQQPHSPEALGETQEERGQESPHSAQ NLRAQVAQAPDNPQHRRRVKWPPASKVSE WQQLDEDLEGILESTAKGGVDRKLQTM TTL VISFATERYGTMEKRAAPEKYTKNRRRAEKIS QLRQELRVLKKQFKGASEDQKPGLAELRCT LRKKLLTLRRAEWHRRRAKERAKKRAAFL ANPFGFTKQLLGQKRS AHLECAKEEVDSYL HDTFSDAERENSLGECRVLISPPEPACSFNTK APTWKEIQTVVRAARNNSAPGPNGVPYLVY

				CTTCCGTGAGGAA GAACCCAAAACCTT GCTACGAAGAGCC CGAAGCAAAGAT ACCCCCAGGGGAG CCCGAGAGGGGG GGAGAATGAGCTC CCCAAACGGACGG ATAAC (SEQ ID NO: 1167)		KRCPKLLARLWKILRVIWRRGKVAHQWRW AEGVWVPKEEKSTLIEQFRTISLLNVEGKIFF SILSHRLSDFLLKNQYIDSSVQKGGIPGVPGC LEHCGVVTQLIREAREGRGSLAVLWLDLAN AYGSIPHKLVEMALARHHVPGPIKTLIMDY YDSFHLRVTS SVTSEWHRLEKGIITGCTISV IIFALAMNMLAKSAEPECRGPIKSGIRQPPI RAFMDDLTVTTTTSVPGCRWILQGLERLMTW ARMRFKPGKSRSLVLKAGKVTD RFRFYLG TQIPSVSEKPVKSLGKMFDGSLKDAASIRET NDQLGHWLTLVDKSGLP GKFKAWVYQHGI LPRILWPLL VYEFPISTVEGLERRVSSCLRRW LGLPRSLSSNALYGN NNKLTLPFSSLAEEFM VTRAREVLQYRESKDPKVALAGIEVRTGRR WRAQEAVDQAESRLHHKELVGAVATGRAG LGTTP TTHLSRLKGKERRDQVQLEVRASIEE QRASQWVGLRQQGAWTRWEEAMARKISW PELWRAEPLRIRFLIQSVYDVLPSPSNLFLWG KVESPCPLCQGRGTLEHILSSCPKALGEGR YRWRHDQVLKAIAESSAMEYSKRLPLPGR GVRFVRAGEQPPPQPRAQPGLLATARDWQL RVDLGKQLKFPENIVETNLRPDIVLHSQSSK QVILLELTVPWEERMEEAYERKAGKYAELV
--	--	--	--	---	--	--

						EDCRRAGWRSRCLPIEVGGRGFAGKSLCKA FSLLGITGMRRRKAICAASEAAERASRWLWI QRDKPWTSASWTQAGN (SEQ ID NO: 1412)
NeS L	LIN9_S M		Schmidte a mediterr anea	AAACGACATCATG AACGCTTGGCCGC ACAATCCAGTTA TCCCTGCGGTAAC ATTGTGGA ACTCA TAAGACAAGTACT AAAAGAAGAATT AGAAAAATTAGA AGAAAAAATTGA AAATAATTTATTT ATAAAATTTAAAA ATTTAAATAAATT TAAAAATTTAAAT TTAAATTTAAATG AAGATAAAAATTT ATTTAATCCAATA AATAATCAAGAAA ATCAAGAAA (SEQ ID NO: 1168)	TAAAATGGCAAAAAGA TATTTCAAGATGAATTG TGGACTCATCTAAAAA ATGACCACCTTGAGTCC AAATATGCCTAGCTATC ATGGTTGCTGATGGAA ACAGTAAGGCACCTGA TAGCTAACTTTTCTACTG TGAATATCTTCAGATAT TCACAGTGACACGAAA GGACACCACTAGTAAA AACCACCTAGTTTTTCT GACACCTCTTGCTACAA ACTCTGTAAAAATCAA AAGGATCGATAGGCCG CGCTTTCACGGTCTGTA TTCGTA CTGAAAATCA AGATCAAGGAAGCTTT TCCCCTTTTAGTCAACA CCAGGTTTCTGTCCTAG	MMDSRQLNTPKIRKYQNPMTNDIMKSYN YAVLSDVTPQETTQTTTHLNV DIDNETTQPK QPLTKSGKPKSKPIAVSYKFKDATFIWDTTP QTNPPRDCTKLIDKTRPRKTIFKKSAFQSYL KKELSNETFVEVKTFMLATHKYRFDENS LLAYRIINRYVMETANEFKETEFDMARFAKF FTIPENWLKHLKPYSTATETSPADRIKVQKL VDLTCRYPFKTQEEQTSVANFLHFFTQRSIIG ISRDKYKFQKFIPFMARKNTRPETTSTMVTTSP TEQNRLPMVIITPLEEPKSEHRRPEKRGASN DTIVLSDEEFPLLKRRTLPTRKSKNPTGAGN VPTETECTDEVKFILNNEYQIECKEKGKVWE NVRNGLNHLRQKHDFPNRTDVMVSCVRCE VPIKGAECVNHIKNHKKDDKEESEAGSLVA NTQDIPNESSLSQAAIEVYLRNILKMKENQE RNIQYLEPSTANFLINRNLRAFYQNVKIEKLI GWEQVIWLIHWNKCHWIVYLANCD SKTSVI LDSDNQMTLQQR CNIAKFDKFLEGTFE EK TVLGT LERKVPQQPNFDCGIYVIQYISDFL KDPQRIDYHTPDSKRIRKEIGELILEEMKNPA

				TTGAGCTTCCCTTGGA CATCTGCGTTACCATT GACAGATGTACCGCCC CAGTCAAACCTCCCCAC CTGACACTGTCCTCAA ACAGTTCAATTGCATCC GAAGATCGCAATTTTT CACTAAAATAAATTAA CAAAAGTTAATTATACT GCTTCATTGAGTAAGTA GAAAAACAATC (SEQ ID NO: 1291)	SKIKNPNKEIQSLLQKFRLLQINVNDVFHWF AAQYQKSLPKIRTKRDGKLNKLSCSYQIQRL FGLAPKRAVKEIYFQETSTADLETRVLNEHF KKDESTMKECKIKNGNHQDWITKAQIDNK EILEALKNSTDSAPGEDNIPLRQWIIWNNDG VLFDMFNYIKRTHDIPDMWKNYTTTLLIKP GKSQESNIPANWRPISILPTSYRIFMKVLNKR VLEWANRGELISKWQKAVDKANGCDEHSY VIQALIEKANRSYYKNEQCHLAFLDLADAF GSIPFQVIWHTLKNMGMDDEETINLLKEIYKD CSTKYKCGKNESEKIKITKGVRQGCPLSMTL FSLCIQYLIQGIAEKKKGATIAGQEVCLAYA DDLVIVANTAKDMQMLLTTIENLAKQADLI FKPAKCGYYRDPKDKSMMKIYGKEISIVD EKNVYTYLGVRIKDTKKKDLNVRFEVKKK TTAIFKSKLRSDQKLEAYNIFCQSKFVYILQG EDIAKTKIETYDEEIKKMIKEDILKLQDKSPF TDFVIYSPREKGLGITKIIDEQTIQTINRTAK LLNSSHRAIRAIYEELIQVANLRGEKEINTIE EALKWLEGTNKYKKNSNAKTTWITRVREA FQTLEKKHKIKVRFPKENCIGYKIKCDTQE KIVELDNSKELSKSLHWMIKEAAYKEWKAL KCQGYIISLKTSEFMEWKMPRGLPDPDWRF
--	--	--	--	--	---

						LTKVKANMLDVNMKQANQGGRLGSTKCR KCEDKESASHVINHCASGNWSRVEKHNQV QNELAKELTKRNISFEKDSIPKETKESLRPDL VIRLKDKIMIVDIKCPFDEESAIESARNKNID KYRELAKEIQAKTGLQTTVSTFVVC SLGTW DKRNNELLRQMGIRYEESKEMRINMIQKAI HGSRKTYDHRNFNNG (SEQ ID NO: 1413)
NeS L	NeSL- 1_CBre		Caenorh abditis brenneri	AAGGACGCTGGTT TAAGGCCGAATTC GTTCGTTCTTTTTC TGGCGGTCTTGCT TTGAGCTTGTTTT CCGATCCT (SEQ ID NO: 1169)	TAAACCCACACGAGAM CTACGACGCCATAAGA TCAGGCATGTACGGAT GTGAATGAGACTGATG AACGGAATGAGCACGT GCCATAAGATCGGGT ATKAAAGAWCAGAGA CGATCCCTAMCATCGG GAAAACACGAGTTATA CTGCTTCACTGAMCTC GCTAAGCTCTCATAATG ACCGAACTTGTTTCGCA ACTGCCTCCTAACCGG GCGGGTGTGAGAAGGG AGGTCGCCTTGAGGCG GACGCAATGAGGGATG	MPLXISDCVHLVSAEGDTMNGRSTCGPLSR SSSVVSRSRSSPSVPPHPSPSIGPDTGLSAG IIGTSRGC SLWLPEVDNALSQWLRKGLERD HEVLVCGFEAAKPLSLSKARLLRKTTPRNTG VVRHILEFDGRLVHTNCNETECVLSTLXSSX AVEVVRISLKCEPREPCEPKCVLSILCSDKIV XISFECETREPFPPFXDRKFREPIPFVFERMY DPRDPIPSFICWMYDLRQRMTPGTLPXNPLS XENKDSWGRPAVIKNEIRSMRSYLEENVKE NRLNLLRRLRGGGEGKKMIRKLVAEKKSDT EAVCRILYPLDDRYECFVDGCETTSTMGYG SSDLKYMTTHIKKEHGVKVQWTYECSLCN KQAPFMGGAASKWVTAHMA TKHTETVKL KLKPSISTTAKVAAKLDEIAVSLPKPRQVRV LRDPDEVKEKVAKPTLASTREEVKRNALRN MAPLVELSSQNQLTGAERPEETSEAMRLEE

					TGTGCAGGTTCCCCCTC TTGAGATCCGAAAGTC TAAAAGTACTAGACCG AAAGATCGAGGACGGA CGGGATGGCCGCGAGG CACACGGCGGGTAACA CAGCCAGATAACCTAG TAGATCTTCGGATCTCG TCGGCCTGGAGATATG TGGAACCCTGGGAAAG GAGAAAGTTGTTTGTG GGCTGGCAAGAGTGAA GTTTGAATGTGAACCA CCGTCATGCAACCACT AAACCAGTGGCGATGC GGGTGGAGTCATCACA GGAAAATGTTTCTGTTG CTTGACTTATCAGTGTT TGATATCGCCCTCAGGC ACAAGTATGAAGGCC CCACCCACATAAACTC CCTAGCAACTGGTAGT CCAGCAAGCGCTGGTW	CRTPEKIAELEGKIQTRTVTKKLSALKESME KRTREEKVGKPSLAPIHEEVKKTARRSLAPL VEPSTFTHLTGASRLQAVRDAFSKANKDAA AKRRSSLAKPARLSEIMNTTFTKETVNETKE PVNDTDESIATIQPQVRVYRFNTWCLDHETT REAWLTGEVVDWFMGKVTEKKDQYRVFD SLVWSMYKFHGVGYVLDLMRDPLTYFLPIC EHDHWVLLVIDEKGIWYGDSKGAEP CREIA KFIEETKRERRMFPVPVPLQRDGVNCGVHIC LMVKSIVNGEPWYTEEEVKVFRNRNVKRG LK EFGFELYSERIVYVGDDSIKVNDEHDDD VVF LSEETNNTTFTIEQAEDPAEEDAQHLESPVK PVKLMELKIPKIEIKKKEIRRKPKQQIEKKRK VPTGKPDELLVRVRLWLEREVQSYFDSGKR FQRLEWILDVLTAAIHKATAGDEQAIERIEK RSPPLEVEEGEMSTQTEPKKRERKEKESGCE MKASHKEMYFKNRSKAFNVIIGKDSKQCEIP IETLQKFFEGTTAETNVPAEVLKEMGSRLPK LEALDWMEANFIESEVSDAMKKTKDTAPG VDGLRYHHLKWFDPEYKMLTLLYNECKNH RKIPSHWKEAETILLYKGGDETRPDNWRPIS LMPTIYKLYSSLWNRRIRSVGGVMSKCQRG FQEREGCNESIGILRTAIDVAKGKRRNLSVA
--	--	--	--	--	--	---

				CTTGCTACTATTGCGCC CCAGGCTCGCCC (SEQ ID NO: 1292)	WDLTNAFGSVPHELIKSTLESYGFPPEMVTE IVMDMYRGASIRIKSKNEKSEQIVIKSGVKQ GDPISTLTFNMCLENVIRRHLDASGHRCIKT KVKVLAFADDMAILAENRDQLQTELNKLD KECESLNLIFKPVKCASLIERGMVKNKNAEV VLRGKPIRNLDENGSYKYLGVHTGIATRVST MQLLESVTKEMDLVNQSGMAPFQKLDCLK TFVLPKLTMYANAIKPKLTELKVFANLTMR MVKEIHEIPIKGSPLYVQLPPSQGGLGVACP KITALITFLVNVMMKKLWSSDSYIRKLYRDYL DEVAETETGMEEMTKEDIAKYLSGDVPIDK KAFGYNTFTRVRDVCNSLTXIXGAPLHKLKI VERDGDFAILVQATKEGMEKIFTCAQEKKL QQLLKAEVNTALAHRRFFTEKPVKSAVMSV MRQYPQSNAFVKNGKNVSIKVHSHWIHKARL NALHCNFTYGENKSKVCRRCGKDVETQL HILQXCEYGLPKLINERHDAVLHVVRNLIRK GSKKDWKLKIDETVSSCNQLRPDIYMCSPD GKEVIMADVTCPYESGMQAMQESWNRKVT KYEGGFSHFXKMGGKFTVLPVVGSLGTW WKPTTNSLVQLGIEKXTIRRVIPELCSMTME YSKDVYWNHIFGDTFRKPPMRFGVEKPKGN SWKKEGSEPKGAASSD (SEQ ID NO: 1414)
--	--	--	--	--	---

NeS L	NeSL- 1_CJap	.	Caenorh abditis japonica	GCGCCCCGGGTTA CATTGTCGGGGCC ACCTTTCTCTTGG AGTAGAGTACAGT CTACTAATTTTTTG ATAAGCTAGTCGG GTCCGAACCACTA GAGTTTGCTTGAA AATGCGTCAAACC AGCATTTTAGAAC TCGCCCCAAAAGTT CGGCCCCGACCCC CAAACAAATGGG ACCTTCTTGACGA TTTTCCCTGAAAA TCGGAGGATGGAA TGGTCCCCTATTC TTGTAAATAGKAC TGTGCAATACCCC TTCGTCATCTGTG GGGAACAGATGA CACGTGACGTCAT CCGTGTAGACGTC	TAAAAGCCAAAAGCCA CGGAGCATCGGGAAAG AAAAATGGAAAAGGAC TGAAAACGAGACTGAA AAATCCCAAACAAAAC AAATCCAAAACAAACT GAAAAAAAAAAAAA ACAAAACAAAACACTGG ACAGACACTGGAAACA GTGTCAGGCAAAGTCG CCGATTATACTGTTCCA CGCCTTAAAAGTCCCG AAATGGCGCAAAACAA CCTGAATCTATCTGAAA GTGCTCCAAACACGC ACAACCTCGGAGAAAAT CAGGGACAAGTTGCTT CACGCAACGGGCTGGG ACAGGTACCCCCTCCTG AAACCGCGAGGTTGAG GATGGACGGGAAGGCC GCGAGGCTTATGGCGG GTAACCTCGGTTGGTGTG	WRRPAPKQTKNSSLHHLGHEVKRIARLKPGI FEFHAKPKNSSLHHLGHGVKRXARLKPGIFE FHAKXKNSSLHHLGHEVKRIARLKPGIFEFH AKPKNSSLHHLGHEVRRNSRLKPGIFGFYQK SKNSSLHHLGHEVRRRIARLKPGILEFHAKNRI KSGLKVTFLSDLXAHAGALACSRFLASTLK TEHCRQKSFKPVGFLLHFLKNSSINEVASLR NVKKXFLEFFSGKPIGGMASFSRTKITFFKLC LKNFVLSAENPPIIRQKTNQNKASXVQIARG GHLSDCLPSQKMAGVLGRLFLSVQSTLSHR PFDTLRSDDDKRGRKTIKLQFFIKENLVTP XVAROVKILXKQTKNNSGNSDSNSETKNFS KNKVSRQNGPLIGGGNHKKIGENQITRTLEI ESKSDDNKVLVLRILYPTNDWYKCYSQWC QHKSLVGYGAHDLKYLTDHIKSTHSSKKVE WSYQCSICDAKAEGTGTKAARWITAHMPK VHGIEATHRIKQNSEKTTNVKTANSLQEMA LSLQKPKNGPKKVVMATSTTPEKKISELESK IQTREVAKQLSALKESAQKNQQGNKTKNVK SSLKTIAENTNETKKISARKSLINYLPEDVL NHIPKEPKPASAKXGLQELTGAQRLQETRRR FMAGNRRDSIARRESLSLGKISNSFKIELKNA PEKTTLKKPAVTQKQNTSQNVSSSTVVKEN
----------	-----------------	---	--------------------------------	---	--	---

				ACGTTTTCCCGTG CCTGCGGGAGCCC CCAATCGAGCAAT TTTTGCTCTTTTGA GTGTCTGGAACGC TTGAAACCCCAGA CAAATCAGGCCCA GTCGTCGGAAAAT TTCTTTTTTGAAATT TTTTGGCGCCTGC GAAAAAAATTTTT TAACCGCCACAAA CCCCCGGGAGGCG CGGWTAGGGATA TCGATGTCATCGA CTCGTCGGTGATC TTTGATTTTCTCTC TGCGTCTCCTATTT TGGAACAGTCTCG ACCAAAAAACCG GGCCTGGCAACCC ACCGAATCCGGAT GTCGGAGGGATTT	CTAGTAGATGATTTATA TCCGACAGCCCCAACT AAGAGGAATCCTGGGA AAGGAAAACCTTGAAAA AGTTTTTACAGGGCTGG TAATAGTTCAGCACAA TTGTAGTCTACTGTCTT GCAACCACAACAAACC AGTGGTTCTGCGGGTA GATCAAACCTATAATTTG TGTGTTTTCTTTTACTT GACCCGGGCAACACAT TATACCACGTCCACAA GGACGAATTCATAATG GCCCCTCCCTAAATAA ACTCCCTAGCAACTGGT GGTCCGGCGAAGCCGG TTCTTGCCACTATTGCG CCCCAGGCTCGCCC (SEQ ID NO: 1293)	KTGNDVITIDDTETVKRKINTWCLDHESTEN AWMADDIIFWYIQKQIEISLDNKKFKVIDPLI WTTYRIYGVECVQDELVGFEKYFFPICENGH WVLLIIDDKRVWYSDSLADKPIEVIEDLINK LNRTQGKFNQTVPKQKDGFNCGVHVCLVA KSVITENFWYTEKDVNDFRKTVKLWLFSEG FELYSEPYKQIQKNISVNSEKNQISDNEKN WGDKTQTVNESTLKERDEDIFLLRPHISVGV ALKTEDEKNQKAENLKAPQKLKAIRRLKIL KTCLKKLTA VKGKPEETERAAIPNLMAIKLK TPPKVEPVRRNPEKGENYXKSQPNKKRQIPT GKPDDEL VKKVREWF EI QFQAYFEDGKSFQR LEWXTGLLTAAIHKASAGDEQAVGKIIKRCP PLEIEEGEMATQTETKQKPKNQKSTKGANSS SSIREAYAENRARTFNKIIGKDDKCEIPIEKIE KFFENTTSNTNVPTETLARITSDLPKLEIGSW IEEEFREKEVAEALKKTKDTAPGVDGLRYH HLSWFDPKXKLLTKLYNECREHKKIPGHWK EAETVLLYKGGDETQAENWRPISLMPTICKL YSSLWNKRIKSVTGVL SKCQRGFQEREGCN ESIAILRTAIEAAKGTKKSLIAWLDLTNAFG SVPHESIEATLIA YGFPGMVTEVIKDMYNGA SIRVKTNEKSKQILIKSGVKQGDPISTLFNI
--	--	--	--	--	--	---

				GGCAAGAAATGTT GGAAATAACGAA ATTTCGTTATTTTC AGCACAATTGTCA AACCGGCAAGAA AACTGGATGGACA AGACACACAATTT ACCGGAAATTGTG CTTGTTACGTCGA ATTTCCCAATTTT GAAAAAATTCCTC GTTCCACTGGTCG GGACGCGAGGTCA GACGATCTGCACG TCTGAAACCCAAA ATCTTCGGATTTT ATGCAGTAG (SEQ ID NO: 1170)		CLESVIXRHLKSADGHKCIXSNIKLLAFADD MAILSDSKTKLQQELQKMDDDDCTPLNLIFKP AKCASLIIEWGKVQKDQKIKLKGQFIRSLAE QDTYKYLGVQTGIETRVSAMQLMKKTVSEL DKINCSALAXWQKLDVKTFLPKMTYMY ANTVPKLSELKEFANITMRAIKVMQNIPVKG SPLEYVQLPIGKGGLGVACPKTTALITYLVS TMKKLWSTDDYIRKLHTDYLKMVAIKETKT KEVTLEDLASYLSDDKTVCKKAVGYNSFTR VREICKTLSKNKGALLSQLKIIAKDGKLAILV QAXKDGTKIFTHDHVKTLQKXLKKEINEA LLHRFTTEKRVKSEVVRVVQEYPQCNSFVR DGGKVSIGAHRFVHKARLNLLACNYNTWQ DAATKQCRRCGYEKETQWHILSSCPKSMGG KITERHDSVLKTVKEMIQTGSLKNWKLKLD HELPGSTRLRPDIYLRSPNGSEIILGDVTIPYE HGIEAMQTAWQKKIEKYEFGKYLRSTGKK LTIVPIVVGALGSWWKPTTDSLVS LGIDKNT VKRAIPEICSTVLEYSKNIYWNHIFGDSYQK VPMFFGGEKPKGQSWKKVKPPEGKTASNH EPPG (SEQ ID NO: 1415)
--	--	--	--	---	--	---

NeS L	NeSL- 1_CRem	.	Caenorh abditis remanei	CGCGAACCAGTCA T (SEQ ID NO: 1171)	TAGCCGATCGTAAAAG AAACCGAGCCGTAACA ACAAGCAAAGTAAACA AAAGAAAAATCAATAA AAAGGAAGGTTGACCT CAGACCCCGAGGAGGG AAGAGAGACACCSAGA AAAAGAGAGACGCAGA GAAAAGGAGAGACACC TCTCATAAGGAGAGGT AGGTCAATCCAAATGT AAACAGAAAAAACCAG TGGGGAGGAAAGAAAG ACTGATTTCACCCACTA AAATGAATTTGGAAAC AGAATTTGGAAGAGAA AAGAGAAAGGGAAACC TAAAGAAAATAGTTCT CTTGCCAAAATTCTGTA GAGGAATACTTTGTCA AAACATGATAGAAACC AGTAATCTGGTACGAA AGACAAGTAAGACCTG	MTVFIDRGIGERGQMAVCSLHRYFSFSPFSPI PPYVNNGSFGENGCGTDKSLPVIEWVVREV KINWSENILVVECLIMVKSGERVVVKRQNL EKVIQNLARINSTLFSNLGNQIFCVVPRIKDS TNKEQGYRKEKQXKFHVSFRSISQVPPYLR GGGDVMEDTEIRGIRKLEPEAQLDSSKPLIC RVLPTQGYMYKCFYPKCKGHSNGSTDLRS LKKHMVDKHFTNIEFAYKCATCMFLT TGKS ATALKSIKAMASHHKVTMEPGKKSLVQK L NARLEEAAPSLPMPRNRSKVIQLTPEKSISE LEKKKQTRSVAKQLSTLKESAQKKEEEVKI AEVKKREPRLSIIPESNVRRSLAAGLEQCINP EQSVAQRIREKREEYAKASREAAAKRRSSL AMKPARLPDKENEITLQETKKIDDPIVIDLEK ECILT TVLQVPRNQFNSWCLEHETTIDAWLT DEVIHMYMCTITENRKYFMAIDPVLWPVYV R NGAEDLLRRTSCPGTFFFPICESNHWLLVI EHDVYWYLDPKGEEP KGNVEILLES MKRKR QYYEFPPPSQRDNVNCGVHVCLMAKSIVDE CGYNWYSEEDVRSFRTNMKDILKSKGYELC PEPYNRQNLLKTEKQKEVILEEMIDSFVVED DMTFTVHRDS DHGDDEVEHLKTIEQEPENEI SEIENVEGSVDSVIPKLMEMRVQTPPVINEK
----------	-----------------	---	-------------------------------	--------------------------------------	--	--

					AACTGACAAGAAGGAA GTCAGAAAGAAATACC GCTCACAAAGCCTGTG ATCGATTCTCTTACCTA CTGAACTTGTTCTCTTG GCCTCGTAACCGGCTA AAGGGAGAAGGAATGT TAATTGGAGATAGACA TAAAGATAGGTGGAGT GAAGGTCCTGTTCTTGA AACTAGGAGGAATGTG GAAAGAGCAGAAGGCC GCGAGGCTTTAGACGG GTAACTCAGTCAGTTGC TAGTGGTCTTCGGATCC AACGGCTTCGGACATA GTGAGGAACCCTGGGT ACGGAGAAGAAATGGA AAAGAGATAGGGCGGG CAAAGGCTAAGTTCAT ACACTGTCATGCAACC ACTAAACCAGTGGGAT CTGCGGGTGAATCACTT	RGKKRVSAKEKPRKQKEKEQKVPTGKPDEL VKRVRVWFEKEFKSYVEDGKSFQRLEWXT DVLTAAIQKASAGDEKAVELIEKRCPPEXE EGEMCTQTEKKKKPKSGKGNGGQESMKS MASYSENRAKTYNRIIGKHSKQCEIPIAKVQ KFFEGTTAETNVPKETLKEMCSRLPKVEVG TWIEGEFSESEVTEALKKTKDTAPGVDGLR YHHLKWFDPELKMLSQIYNECREHRKIPKH WKEAETILLYKGGDESKXDNWRPISLMPTIY KLYSSLWNRRIRAVKGVMSKCQRGFQEREG CNESIGILRTAIDVAKGKKRNIAVAWLDLTN AFGSVPHELIKETLESYGFPEIVVDVVEDMY RDASIRVTTRTEKSDQIMIKSGVKQGDPIST LFNMCLESVIRRHLDERSVGHRCCLKTKIKVLA FADDMAVLAESSEQLQKELTAMDADCSAL NLLFKPAKCASLILEKGIVNRLNEVVLRGKPI RNLMENETYKYLGVQTGTETRVSIMDHTE VSREIDLVNMSQLAMHQKLDILKAFILPKMT YMYQNTPKLSELKVVFANLVMRSVKEFHNI PLKGSPLYVQLPVGKGGGLGVACPKNTALL TFLVTIMKKLWSSDSYIRKLYTDYLEEVAK VEIGKFEVNLNDLAEFLSDERAVIDSKLFGFN AFTRVREVVRSLCKNKDSPLHSLKIIEREGK
--	--	--	--	--	--	---

					TCGAAAAGAAGTGAAT GGACGTGCTGATGTCT GACTTTAAAGAAGTCT GAAATTAAAAAACAG ATATAAAGGCCCTCA CTATAAACTCCACAGC AACAGGTGGTCCGGCG AGGCCGGTTCTTGCCAC CATTGCACCCCAGGCTC GTC (SEQ ID NO: 1294)	LAISVQATEESIEKIFTEDQEKKLMYLLKGEL NTALQHRFFTQKVFKSEVMRVVQQHPQSNS FVRNGGKMSFSAQRFVHPGRLNQLPCNYNT WAKGRTKLCRRCAKNENETQSHILQVCDYS IGNIIKERHDAVLYKFRELKRGSKGHWLER TDRTVPNTGSQPKPDLYLESPDGKHVILADV TVPYERGIEGMQKAWNEKINKYTDGYKEIF RRQGKSLVVLPLVVGSLGTWWKPTEESLIK LGVEKTTVRRRIIPETCGMVAEYSKNCYWRHI YGEKYVQTPMINGGKKPEGNDWKKCEKGI EVPKVTN (SEQ ID NO: 1416)
NeS L	NeSL- 1_TV	.	Trichom onas vaginalis	GGGTGAGTAGTCT AGTGGT (SEQ ID NO: 1172)	TAAGAAGAGATAAGAC GAGTGAGAAGAACAGA AGCATAGTAGGATTGG CAGAGCTTAAGCGATG TCACTCGGTACGAAAC GTGTACCAAACACCGG ATTCCGTGCTAGGAATC ACAAGCCAAAATAAAA GAGACACCACGAAAAT TACTCACCTCCCTCAA ACAGATAATAATATTA ACCTCCCATCCATCAGT	MIPVLGTGGPEKLPLQSYVYCGNTAITDSFT PTAKTILKPEEQNLDIVLKNIAALNPENYS DL IRSLSKMEFRLDYPKEIENYWISEKLFSQSIA SLPISLLVASMFSPEDRDLSTEPFHCNADGC NFHCDNCERMVEHIREHHNTDPMINTFETTE DTFRRITAIKIDKTGIEELNPLKYRCSYCDL FTEAEDHAIHMISHLTEKLSPDISFFFNDILRL YKTIDKPTVQNLFPETQVAIFDTLEETNRFL IVGREAIETIEEAFPPSPPGTDRKPSIIITDTCQ LRFVPCMDEPPKGD LGILTLLLRDFS AHNPI KSLNNKELIADKDIDYSPDFVEGALANAEEH DTTNSQNNNGRYINSAEKLTEFLIQCEDYLT

				CCGTATGGTCTGATAAC AGACTAGCACCACATC CATGATACTCATTGG AGTGAAAACCACCAAC AACAAATCCACCTAGA CCAAATCCTGCCCCACC TCCACCCAAGTAGCTC GCTTCGCTCGCTCACCT AAAACTTTGCTCGCTCG CTTGCTCGCTCGTCTT AACCTTTCCGAATAA AACTTACAATTCCCGG CTCGCCCCATTTTTT (SEQ ID NO: 1295)	NIKTLEDLERFYTTIKDYRVNKEVIAEDTPIF VYFLVEEGKLPKPGLRCPLESYEGHEDKAFF SLRKLCDHFKGEIAKTSFDPKVHTIDIWVEF LAQAYGTGTFVYKDENGNIIDLDTHVFKCPY ADCSYTNNDRSKLMHMKTKKHAKNVYIE RYGFFWGIVIEGVNRPKGIVYPTLKDIEHA CRKCPEAGCNTYVTELSDIKEHLKKKHST TAGVDGEIAHTDATYCWITKEELDALHAER ARERAEQVDNTPVQQIINADNNEENNENQE DNGNNEEADALDPPNNTTETEDAEVHAVII NPPATEEEEEVAIIAEARRNIPELQQAEEERGCV TPKMTSLVRLKLLKGGGELFNKKLTPLATR YAATGNTEADKIKVDYLTLCNAALREMIY TNNHSESKFMTAENGEDTAPPPRISEDTRDR IQKAANEIKGTLIKVVKHISHARCLKDSTRD DEHNKFVEMIAKIKNDLRDNKFEQYNIEEIF QGPISDQSILNIVNTEDNNEFIKKMDYINRIL GTPQDASPYARKKLQACFADNPTKTLRNIL ADKVPQQSLKPSEYLDYYGPQWANEAEY ENFLHHDYALPERYGQVFANDFLDFMTNES KIIEVIRNKNHLSAHGLDGIPNSVYMLFPVSA AKFLSILFRSIIISGHIPDCWKLSKTVMLFKK DDPSLAKNWRPIGITSCTYRIFMTLVNKALQ
--	--	--	--	--	---

						MIPMFHAMQKGFVRGATLSEHIAVANEVLC QSTRTQSEMFQTAIDFTNAFGTVPHQLIFDSL EAKKVPDSIINLLKDLYKGARTAIYTRHAHS EIVPVRRGVIQGCPLSPILFNCCLDPLLYAVQ RRHFEDGYRFQDKAGQYSIAIQAYADDVLV ISPTHEGMQRILNTVDEFQKIAKLKVAPQKC VTLAKTSTAIQPFRIGPDEIPIKTSMDNITYLG IPISGTKTSRFAAATGILEKVKAQIRVVFASH LALSQKIIALRVFILPQLDFYMFHNVFRVND LKATDQMIRGLIDKEAPTSNIPVSFFYMPKN KGGFGLVKLELRQPQLVLTKFARLWLSQQA ETKAFFHTMAQEEKSFRKVVEDQENGFLGI KMENGKIVQKNERSKRTNCFITQAAKAADK LEVRFKEWDKGGIQVRGVGENATDWYRSK HIGQISPLIGRVIQQRQYEEFKKDETHSHTFC EPAALAESHDIMKRPQAVPNNLYSAAIALRT NTAPTPANMHFHNPEVLANCPLCGCQSCCTL FHTLNMCRNRFSLYKWRHNIICDDIYQFIHD HYPGVTIKCSARITSDGYQTTGPELDDTVKD LLPDLVVYDEANKMIKIIEVTCPYGTDNNVG NSLDAAYDKKVNKYKSLAEQTERLFNWTT TLSIIVVSSLGVIPLRTKLDALRISPADHIQLL KRLSMHAIAASACIVFEKVPEFFGMRCRPLP
--	--	--	--	--	--	--

						GRVTAPNAAIPPNNNNENNNDTDHGGQENQQ ATSEEQPTNNGNAQEDNGQGGEQINNSTEQTI SVDQIIIEEDAENNAIEQALDQPDEDEFLN (SEQ ID NO: 1417)
NeS L	NeSL- 2_CBre		Caenorh abditis brenneri	GACTCGCCTTGGG GAAGGTWTTTCAG GGGKSAATTGCCG MAGGCAAGGCAG CCCCSMMTAGCT TACAAAGTAAGTA CMCATTTTCATTT CTTGTGAATTCTTT AAACATATTTTTC TTGTTTTTTTGATTT CTTTTTTCTCTACC TTCCCCCAATTCTT CCCCTCATCTTGT GTATACATCCCCC TCCTCCAACCAAT CAATACATTGACC TCTCTCTTCTGTCA AAAAATCAATACT AGTATATTGTCCC	TAAACCGGCTCCTCTGG GAGGAGGTATGTCAGA GGACATTCTCCGTGGG CGGATGGGAGGAGTAG GGTAACGACCCGTCAT TCTGGATGCCTAAACC ACCACAATCTGTCAAG GCAAAGTGCCCCAAAA GCACACGCGTGGATCG GTTTGGATGCCGACTG AGCCAGAGGGCAAAGT CGAAGGCCGGTAGGCT CCCGGCGGGTTGTCCGT CATAGTCAGTGGTGCG CCTACACCCAAGTCTA TGACACACAAGGACAA CCCAAATAAATAAGC CAAGGCGGCGTTAGCT TCGAGCTAACAAGCTC	MRYHXSXPAXRTSDNXWRSIXKDVRPPD PSTIEEKSRYNRSIGIPDSLKXRSSAVRSXSSX PPSGPQDVRLXNSPSLDDRRRLVDCETTLGS YREWTDKPMMGKMTYAAVTKRAPPRPQT GGARLSTNLLADEMEIKYRDTNDIRLVIDLP NPHLIKCPCKSCISARGRGANALKYMKRHI ADAHHLNADVFYKCSRCQEHEPENVC GAK WIVNHLKRVHGYTLEDAVSTAKPSTRQQIA NAFNDSAPFIDARKTSDVPEKKSREAGLEKF LAPTKSEDTREKTPPSTRKSSESSEASIQSTIQ ETLSESSDTLTVQEIIINISSEDEMDEEPPKRRV NVWALIHENGKDAWIDSDLMVIFLESRARG YESCSIIDPLNFICTDMSYLTIVRRRMEEGY KKIIFPLCANDHWTLVTITGSTATFYDPMGN EPTETVKKMIDELDLEMQLAPSNSPRQRDS WNCGVFVMKMAEAYIKDTQWDLTDVDTD VKTFRRSLLTELKAKFNIFAEDIQTYRPPSRK ALTRNSQSPVVVCHKCSR PATPIQDVSRME VEEAPVLVPTPEEPPQEWTFVGKNRKRGV T

				TTGTATAGTATTA TTTGACGTCGTCT TTGTATTAGGAGT AGGTAACAAWCT GTGTATGGCTTCA AAAAGCATGCACA AACWCCTGTCAA AAAGTAWTTCCCA TCMTGTGAATAGC TCAACGACWKGA AGMCCAATGAT (SEQ ID NO: 1173)	CCCGAGAGGATGGTTG CCACAGGGCACCATCC TGGGGAACGACCCGAT CTTTCGGATGCCCAACC ACCGCCAATCTGTCAG GCAACGTGCCCCAAAA GCACACGTGCGGAGCG GTTGGATGCCGACTGA GCCAGAGGGCAAAGTC GTAGGCCGGTAGGCTC CCGGCGGGTTCTCCGTC GTAGTCAGTGGTGTGC CTACACCTAACTGCTAT GACAAGCGTATAGGAG GCCCCGAAAAACAAGC CAAGGCGGCGTTAGCT GAGAGCTAACAAGCTT CTCGTGGATGGGTGCC AGAGGGCACCATCCTG GTGGGTGGATGGGGGG AGCTTGGAACGTCCC GATCCTTCGGATGCCCA GACCACCGCAATCTGT	S RTPNTSPEAKRPAFPPVPLKPSANRWHFPE EETEKMEVSSADEVKNSTPPKPPKIPNLLAM KIASPVPLKRGNPSKKHKGKGHMMNTARKG PTKKEMPKGEPANLIVKIRSWFDEQLKMYK DEGSNLQRLTWLSDSLTAAIGKAFNGNKYI VDQIIKRNPPLVEKGAMSTQTSRKRDEFKP RERMAQEPNEPLRIQYAKNRQKTFFKIIIGKQ SEQCTINIETVEQHFRKTLKAPVVSENAIVT CGSIKKVLMPKTIEDPISSVEVKSILTKVKDT SPGTDGVKYSNLRWFDPEGERLAKLFEECR KHREIPSHWKEAETILLPKDCSDEEKKKPEN WRPIALMATIYKLYSAVWSRRISGVQGVISP CQRGFQSLDGCNESIGILRMCIDTASVLNRN LSCSWLDLTNAFGSVPHELIRRSLESFGYPQS VIQIVTDMYKGATMKVK TADQKTQSIKIEA GVKQGDPISTL FNICLEG IIRMHQMREKGY DCVGHKVRCLAFADDLAILTNNKDEM QEVI DKLDADCRSVSLIFKPRK CASLTIVRGAVDK YAKIRINGDAIRTMADRDTYRYLG VKTGVG GRASETEALIQVVKELQKVHETDLAPHQKL DILKTFLLPRLQHL YRNATPKLSELREFENV VMKSVKRYHNIPIKGSPVEYVQIPVKKGGL GVLSPLRLTCLITFLTSTLCKLWSDDPFISSIHK
--	--	--	--	---	---	--

				CAAGGCACCGTGCTCC AAAAGCACACGCGCGG GTTGGTTTGGATGCCGA CTGAGCCAGAGGGCAA AGTCGTAGGCCGGTAG GCTCCCGGCGGGCTCTC CGTCATAGTCAGTGGT GTGCCTTCACCCAACCTG CTATGACATGCGTACA GGAGGCCCGGAAAAAT AAGCCAAGGCGGCGTT AGCATAGGGCTAACAA GCTTCTCGTGGATGGGT GCCAGAGGGCACCATC CTGGTGGGTGGATGGG GGGAGCTTGGAACGT CCCGATCGTTCGGATGC CCAACCACCGCAATCT GCCAGGCAACGTGCTT CGGAWGGTCATTGGTT CTAGACTTGTAATAGA CCATTGGCCGGAAGAG CACACGCGCGGTTGGT	DALSRITVKAMGLTTQSATIKETCEYLNTRK AVTKGGYSLFCRMNESLRTL SVIQGAPLKS MEFIPVNNEIGIAVQATKDSEIKVFTKADSLK LMSKCLKDLVRSAMLKRFLEEKSVKSRVTQV LQHHPQSNRFVRDGRNCSIAAQRFVHPARL NLLSCNANTYDVNHPKGCRRRCQADFESQQ HILQNCHYSLAGGITQRHDRV MNRLQEIGN GRKAHYKIMVDMETGATRERPDII MEERDG PEVLLADVTPVPYENG VQAVERA WDKKIEK YKHFLDY YRKIGKKATILPLVVGSLGTYWP DTSHSLKMLGLSDGQIRNVIPEICQIALESSK NIYWKHILGDSYKTVEGLFCQRNNKEVRFE GKGEKHHVSQRFQPLKCEKVRTMKSTKEEG RSRSNAKKGPNWRRSKSESDGRSVSKGRY WRDPSNKPPHSKMTQSALAKR (SEQ ID NO: 1418)
--	--	--	--	---	---

					TGGATGCCGACCGAGC CTAGAGGGTGCAAACC TGAAGGGCGAGGTCGA AGGCCGTGAGGCTCCC GGCGGGAAACTCCGTC ATAGTTAGTGGTGTGCC TACACCCGACGACTAT GACACATAGGAGGAAT CCTGATCTGATATGATC ATGTATATAGGGAGGG CGAAGGTAAATAGTCA GKGTCAAAGTCCACGT GGCAGCTACTCCCCAG CATAGTAGTGATGCGA GTGGAWCCAACCTTGA CACTGATGTTCCCTGAG CCTGACCCATCTGCACA AATCCAACAGTGTATG ATGGCCCACACACTGA GGACGAGTATCACTTG TGATACTCAGAGGTGT CCCCCATGATCAACCA ATATCACAGCTAGCGG	
--	--	--	--	--	--	--

					ACCTACCGTGAGGTAG ACCCCCGCCGCTGTAG CAGGCTCGCCTC (SEQ ID NO: 1296)	
NeS L	NeSL- 2_CRem	.	Caenorh abditis remanei	CCAACTCTCATCG TATTAACCTACGG TATTCACCTCCTAG TGAGTGTAATAAA GGTTAATTACGTT TTCTCTTGCMAGA GAAAAAGAAAAT TCGAATCCTTTTT GTGTAACCTCACAA ACTGACAGAGACC TATCGAATTTCTT TTGTTTCGTATAT AGGAATAGTCACT CTGGACCACGAAG TGGACAGTTGTCG GCGGACTTCCAGA GTGGAGAGAAAA GGTGTGAAGAGA GGAGGTCTAGAAA	TGAATACCGTCAGATA AGCCCCCAACATAAAA ATAAAAGTCGGCGTTA GCTAACCACTAAACCG GCTCCTCATTGGGGGA GAGTATCATTCCGGTGC TCTCCGTTTGGGCGGTA GGGAGGAGTTGGGTAG CGACCCGGAAGTATGG ATGCCCAACCACCGCA ATCTGATCTGGCATTGT GTTTCGGATGGTCTCTG TCTCTAGATCTGAAATA GAGCTCTGGCCTGAAG AACACACGCGCGGACC GGTTGGATGCCGACTC GATCTGGAGGGTGCAA ACCTGAAAGGGAAAGT TGAAGGCCGTGAGGCT	MTNVYLKPVNDNQTNKTGDNSRNTMSNSQ CEMTWKPVARTYAQAASTNPADDKTVTVL GCKYNLLKLGNTQPQTSKRSPKPSRGGARIS SVYTLTDELEITHREEGKITFAIDLPNKNNIL CPLCRECTQTRGRGSSFTKHMKLHVKEKHQ LDATFIYKCSMCNEYEPEKKCGTKWIQTHL QKVHNYKYDESAIVVPVPPNTRQQIANELN NAAPFVDIRKPKAAAVEEKKTENGALLKFL TKSNKDDQVKSPSXDIPDAESPEKETQALTI DPKGNNSPSKSSIRSSQSSASSVCQEIQEITL SEDEDPKGARKPKGINVWSLINETGKDAYID TDIMMAFLKMRVENCDSVNIIDPLNYQFPA RVDLVPLIQRNLEDGKKRVVFPICADEHWT LLTISNGIAAFYDPTGSRMSSYIEELVNELGL IIPKEQDEQPRQRDSYNCGVFVMKMAEAFIQ DTEWEMEEVEEDVKNFRRNLLEELKPNYEI FAEKIKYYNSPGKSFAQSRPTSRSSQCAVCP TCSRSATPMMDVGNMEVDPVPQQQETPKS REPEQDEGWKVVGKARKRGVVTERS PNISP

				CACTTCGGCTGTC TAGGACCAGTTCC TGAGTGGAAGA GGAAGGTCTAGAA ACACTTCGGCTGT CTAGGACCAGTTC GTGAGATCTCTCG TGGAGAGTTGAAA ACAGTCAGCTGAG GCTACTGTATTTT TTGATAGCCCCGC CCCCAATCCCCCT CCCCCCCCCCTCG ACAGATTTTTCTG TTTGACCTCCTGG AATTTGCGAGGAG TGC GCGAGAATT TCGAATTCTTCGC GCGTTTTCTCGAA ATTTTCCAGAAGA TTCGAGCGGAGAA TCTTCGAGAAAGT GAGCTGAATTCG	CCTGGCGGGAACTCC GTCATAGTCAGTGGTGT GCCAACACCCGACGAC TATGACATAGTTGGAG GAATCCTGATCTGATA ATAATCATTGTTCATAT AAGGGAGGGGGATGGT AAATACCCAGGGTCCG AAACCATCAAAGCAGC TACTGACCAGCATAGT AGTGATGAACACATAG ACCCTGGGGTTCCCTGA ACTCGACCCATCTGCAC AAACCCACTTTGTACA AATGAACCAAAGTAT GAAGAGTTTAATGATTT CTTACATCACAGCTAGC GGACCTACCGTGAGGT AGACTCCCGCCGCTGT AGCAGGCTCGCCATTG (SEQ ID NO: 1297)	EAKRQFTGPEIKVVSPGKFHPLVGETEEMEV TCDSPPTKEPTTEPKVTPSLPAMKIASPEVTK KQTSKKKGKYGKKKQXTKKAQPPKGEPTK KAQPKGEPKALIEQVRTWFDKQMKSYQEQ GSNIQTLTWIADSLTAAIFKANSNGNKYLVDK ITARCPPPLLNEGEMATQTSRRTEAVKPKDR FVKESNEPLRIQYAKNRAKTFNVIIGKHSAR CEIDINVVENHFRQTLKAQPVTEEALNTVCS GIKKAKVDPSIEGPISSEGEVKAILAKIKDTSP GTDGVKYSDLKWFDPEGERLALLFDECRQH GKIPSHWKEAETVLLPKDCTEEERKKPENW RPISLMATVYKLYSSVWNRRISSVKGVISDC QRGFQAIDGCNESIGILRMCIDTATVLNRNL SCSWLDLTNAFGSVPHELIRSLAAGYPES VINIISDMYNGSSMRVKTAEQKTQNIMIEAG VKQGDPISTLNFNICLEGIIRRHQTRKTGYNC VGNDVRCLAFADDLAILTNNQDEM QDVLN QLDKDCRSVALIFKPKKASLTIKKGSVDQY ARIKIHGMPIRTMSDGDYKYLVQVTGNGG RASESESLTQIAAELQMVHDTDLAPNQKLD VLKAFILPRLQHMYRNATPKLTELKEFENTV MKS VKMYHNIPIKGS PLEYVQIPVKNGLG VMSPRFTCLITFLASTLFLKLSDDDEYISSIHK
--	--	--	--	--	---	---

				CGCGAATTTTCCG CGATTTTCAAATT ATCGATTTTTGTC GGAAAATTTATTT TCTGGCAAAATTT GATTGAGTTCACG CGGGAGAGAAGG AATTGTTGGAAAA GGGTATTGATTTT TGTGGCGGAGGAA ACTCCCACTGAAT CAATAACTCTCAA AGGAGAACTCATC GAACAACCTCGGG TGACCTGAATCTT GGGCGAAATTTTC GCATTGACACAAG ATAAMACAAATT ACTGTKGAAAATA AATCAGAACAAAC TGTCAAAAAGAGA GACAAAAAGTATT		KALSRITAKVMGLKTQKATLQEQCEYLNTK KAITKGGYSLFSRMNEAIRTL SVN LGAPLKS MQFIPENGELALEVQASENSQIKVFSKADSM KLVTCLKDLVKSAM LKNFLENKKVKSKVV QVLQHHPQSNK FVNDGKNXSIS SQKFVHPA RLSQLVCNGNSYSKDLPKNCRWCGYECESQ AHILQHCTYSLSSGITQRHDRV LNRILXEVIK GRKNNDYYDIMVDTEPGPTRERPDII MIQKD GPEVLLADVTPYENG VVAIEAAWDWKME KYSHFIDYFARLGKRAVILPLVVGSLGTYWP DTSNSLRMLGLSDGQIRNLIPDISMIALESSK QIYWRHIFGDSYRIVSDLYCRKDQQEIRFGD EPMENVQVSDRFQPFKTREREKKSEEEKKR RSKSKKGK TWRGSKKQTDSRQSGKSNQNNQ GFQRSVGQGVSR (SEQ ID NO: 1419)
--	--	--	--	--	--	---

				GATTAACAACATC (SEQ ID NO: 1174)		
NeS L	NeSL- 4_CRem	chrUn	Caenorh abditis remanei	CCCTTTTCTATCGT ATTAACACTACGATA ACCGCTCATTTGA GTGTAAAAAAGGT TCCCCCCTCCTCG CCTGCCTTACCCA CGCATCTCTGCCT CTGGGAAGGCGG AGGGTCAACTTGC GGGTCTGTGGATT TCCTTTCCTATCCA CCGCCCATATTCT CTGTCGAAAGCCT ACCTAGATCAGCC GGGAGTTTTTCCT ATCCCATTCAGGC GATCGCTCAAGGC TGTTTTATCGACA CTCCTTCTTGACA AGTATTTATTTCTT GACAAATTCTATT	TAACATGCCTTGGAAG GCACCACGCCAAAAGT CCTGGCAACTGATTTGA ATAATGTATAAAAAGTA ACTGGAACCAAATGCC CGATAGGTAGGGCGGG AGAAAATGACCTAGAA AACACAAAGTCCCAAG CCCCCGGATTCGAAAG ACCTATAGGAAGTCAG TGAATAGAGAGAAATA TCAAACAAATCTCACC CATTCACAAGGACTTA CTGGTCGAGTAGAAAA CAAGCCAAAACATCAA GCACGACGCAAAAAGG GGTAACTTTGGGCAAC TAATTAACGGATACCTC CGTGTATCAGGCAAAG CCGCCACCAACAGCAA ATTACTGCCCGATAGGT	MTKTEWSWRHRSRSGVIVVKIDTSDYAN VRVHVAADLSNEDGHTSHNNGIILPIPMKPS VDRFCQIQYPPRGYYVPHPQSQKGHDAKPS RHWNEEAQPPYYHNNNHGRRGRSAKPSGR RPPRKPIQEESLAAHPQIPGDTASAVPLYSD VVNNENKSQGKPPQGSRRSGRPGTKPSVP VGEAEQETNSRPIAPEPIVKFKHDKHGWTTV QGSHTSGRPVPKPSVPVSEANRFQLQEGD FPPLTTSESSQEEIKVPNYQRIVSPIPLPSEEDS KLPTKSNYRAPKGRKSRNYKKPQQQNPKK YQQLPYQPKVNNAPTDRMAPEQLKGGGG KTAHNDIEEMEIEEDTDEKIIQVKRIKIVNKL TPHHFVCMPTYPTDNIYRCFVKGCTATSQG GWGAEDLKYLTVHIRQECHKIKVEWTYECGI CGDLSSGAGKHISKWIKPHMRKKHNRDAPT NFKMGSRSSGKPKITELLEESAPSCSNPRRKT LNQKKTAITQVTPEKLKTGYQTRSVTKALS VLKESRQKELEVLREEEKANAKQKSKLHPF FTKAPHIDGVKPTVRRELSKMITPGGEHKGT KIPMVHTKRGLIQKINRKAKKAKPMHLDES THIEASQLDVITIDDDDEDDNMTPMRRRFNT

			TTCCTTTTATCGA TTTTCTCTTATTTA TCGATTCTTGTGA AAAT (SEQ ID NO: 1175)	AGGGCGTGAGAAAATG ACCTACAACCTCCAAG ACCCGAGCCCACGGAA TCGAAAGACCTATAGG AAGTCAGTGAATTGAT GGAAATACAAAACCAA ATTTCTTCCATTCACAA GGACTTACTGGTCGAG TAGAGCACAAGCCAAA ATATCAAGTATGACGC AAAAATGGGTAACCTT GGGCATCCAATCAACG GATACCTCTGCGTATCA GGCAAAGTCGCCACCA AACTGTACTACTCCGA AAAAACCAAGAAACAT GATTTTCCCACTCCGTT AAAGCATCTCAACCAA GCTAAAGCGGTAAGGT TATCATGTCAAAAGGT GTAGCTACAGCAACCT AAAGCCCGAAAGGTAG GGCCGTATAAAAAGAC	WCLDHETTQEAWLTDDVINWYWKDLFCGN EQYMLVDPLVWLIYKMGGMAGVEQRFKSK KTCLFPICEADHWILLVFDETNLCYANSLGS QPNGQVKNFIQQLNRKLCSEFEKEVPLQKDS VNCGVHVCLIAKSIVNGQFWYDDSDVRTFR TNAKAALKAQGYELFSEAPKQIENPDSSHRE DIKENSMEMCSESLMIVATPQRSEAPMELV DTEPSDLESPKSDRVVYEDCITALSDVSEPR MTPEKSETPEVPVVEERDLDPKLESPKSDR VVYEDCITDLSDVSEQRMTPPEKCEPEAPLV VECVELERLPKDLPTDRSTVVAIPEAVKLE EKSEVVIPRLMELSYTPPEPSPVVEYTQPYT HTHTKPKVKATCQMGKKRKVPTGKPDLELIQ IVRQWFEKEFNDYVTEGRNFQRLEWLTNLL TAAIQKASAGDEETIEKIRKRCPPPEVRENE MSTQTSQRQKPTTTNQQKRSRNTTQSDTQA NTYWRNRAKTYNQIIGQDFKQCDIPAIILEEF YKKTTSVTNVPQETLVKVTSLRPLRDIGKWI EDPFTEQEVFGALKKTKDTAPGTDGLRYYH LQWFDPDCKMLSSIYNECQHHLKIPAQWKE AETILLFKSGDESKPDNWRPISLMPTIYKLYS SLWNRRTVKGIMSKCQRGFQEREGCNESI GILRSAIDVAKGKRSHLSVAWLDLTNAFGS
--	--	--	---	---	--

					CTACACCCTCCAAGAC CTAAACCCACGAACTC GAACGACCTACAGGAA GTCCGTGAATGGAGAG AAATATCTCACCAAAT CTCTTCCATTACAAAG GCTAACTGGTCAAGTA GAGCACAAGCTAAGCC TCCAAGCACGAAGTGA TATGGGTAATTTAGGC AACCAATCAACGGATA CCTCCGTGTATCAGGCA AAGTCGCCACAAACAC TGTACTIONTCCGTTACT CCCAAACACATGGATC TCCTTCTCTCACCAAAA AGCTTTATAACCAAGCT AACGGTGGAAAGGACA TCATGTCACGAGGAGT AGCTACAGTAACCTCTC TCTTGAGACTGCAAAG TCGAGGATGGATTGGG AAGGCCGCGAGGCAAA	VPHELIESTLSAYGFPEMVVHIVKDMYKDA SIRVKNRTEKSEQIMIKSGVKQGDPISTLNF MCLETVIRRHLKESSGHKCIDTRIKLLAFAD DMAVLAESKEQLQKELTEMDDEDCTPLNLIF KPAKCASLIIEFGKVRTHEQIMLKREPIRNLN DDGTYKYLGVHTGADARTSEEELIISVTKEV DLVNRSALTPPQKLDCLKTFTLPKMTYMYA NAIPKLTELSAFANMVMRGVKIIHYIPVRGS PLEYIQIPTGKGGLGVPCPRITALITFLVSTM KKLWSDDEYIRKL YNSYLKKVVEAETGIVE VSTKDLAEYLSNKVPSRKHEFGYNCYSRIRE VCNGLALNQAAPLYKLEFIEQDNELAVVVQ PTEESKERIFTKDHVKKLQSLKASVNDALL HRFLTTPKPVKSEVVQVLQQHPQSNSFVRMG GKVSISVHVWIHRSRLNQLTCNYNIFDPKQP KNCRRCGYKNETQWHILQDCTYGWAKLIR ERHDAVHHKVVTMICAGAKKNWGRKIDQE LPGFTSLRPDICLTSPDGKEVIFADVCPVYSR TRNIEFAWKEKIRKYTEGYSHLVAQGIVTV LPIAIGSLGTWWTPTNESLYQLGISKSDIRSAI PLLCSTVMEYSKNAYWNHIYGNSTSVPLR YGHQKPDGDDWKKELSCEPVLALQQ (SEQ ID NO: 1420)
--	--	--	--	--	--	---

					AGGCGGGTAACTCGGC CAGACGCTAGTGATCTT CGGATCCGACAGCCCT GGCCTTAGAGGAACCC TGGGATAAGGAGCACG ACGGGAAGGATGTTCC GCAAGGATTTCCCTTCC CATTAGTCAGGGCTGG CAGTTGGTAATATAGC CTTTCTACACACCACCG TCTTGCACCCACTAAAC CAGTGGGATATGCGGG TGGACTCAATGTAGAA AGGTGTTCCCACTGCCT GACTCGCCAACTTTATA TGTCTTGTCAACATAAT GGCCCCTCACTATAAA CTCCCTAGCAACTGGTG GTCCGGCGAAGCCGGT TCTTGCCACTATTGCGC CCCAGGCTCGCC (SEQ ID NO: 1298)	
--	--	--	--	--	--	--

NeS L	NeSL- 4_SM		Schmidte a mediterr anea	TTAAATCATTTTT AAATGTGTTTGAA TATCTTAAATTAT CAAATCATATTAA TATCAATGCTAAA AAAAAATCGTGCK CATCAGGCGCACG AAAATAATGGACA CAACTCGTCGACC TGCTGTCGACTCA CAGAGAACCTCAA TTTGGAAGAATGG GAAGCCTATAATG CTACAATTCCGCC AACCCCTATTTGA ATGACAGATAGTC AAATATCAAAAAA TATACAAACTGCT GTCAAGCGTGACT CACTTCCTTCCAA TCGAAAAATAGGA AKATGTAAGAAAC ATGAAAGTCAAGC	TGAGTGTGCTACGAGG CAGCGCTGGTAATTGC ATCGGCGTTGCAGATTT GTGTACGATAGATAAA AACCAATAGTAATAAA TGCTGAGCCTAGCTCGC ATATCTAAGCCGAAAG GCAGCATATATATGAG ACAATTTAAAAAAA (SEQ ID NO: 1299)	MNVDLDATIKSIGMNTKETTPNSQLRVET TPCTSTTIMHASCNTTSTISYSPLPSAVSLPES PASSITITTTDDNCDIIETPYPLPQTNGDLSEIL KDIEANKDTTMSNKVLDCSDSGDDRDMII ENDRESMDLFSQSLNNTNQSDERREKNLT ENAPTEITTEKSYFDIISKASDNTTSKKLLNV KNELTAGLPPMPPVTNTAKFIRNVRPEDIAD PTLYRLDSRGKLGCRRTQYKKPGCGDIAVYD YEAIVEHAAFIHTIPFNEQNNVDCQPCHPKK GKDVHTIVLIKYADIFNHIEAHSHVVQTAIT DNMKTYLRLTKENXFYCSYRNNKKKNKCK KAFNLESNMMDITEHMKTHTGYSFDXNLNI LCYCGIWKPFTELIAHIKTEHLQEYINSIPNK ENIHNTTIVSPLNFAGILASGETQNIPDEEII KPRDLPENLAFNRNIENELSWSQLHVKAYIF SYAVKTSTIFINPYTCNALIQCNYKTFETFP FKDFAKWNEIVLPIHNNTSSWSFFFLNKKKR VAMIIDPTADDSHTLHFELATDILRTLNVQN IFEDLNFPLTEVEYPVCHEANLSAFXVCHFL KCLMSDLPIDIPDIDHMKETMRPIIRKYNCA KFPESDVRNYRVLIEDLIYQLNLDITITCEEIL CEIERINGRLNPKRYFKESKPKTDIIHLQKKK SAELLCVKRLKFQISQKTEIGKIWENDDVVDH
----------	---------------	--	-----------------------------------	---	---	--

				TGAAAAACCAATA ATATGTCCTAAAA TAAAACAATTTGA AAATATGCAAAAA ATACCTATAAAAT CACAGCCGAATAA ATTCCCATCCGTT CTAAGCAGAAACC GCTACGAACTACT GCAAGAATCGGAT CAAGTATATTAAT TTCCCCCMGGGG GAAATTAATATAC TTGTTAKAAAATT AATTTTTTAATAA AAATAAATAAATC GAATAAATATAAA ATAAAAATAAATC AAATTAACTTTT ATTAACAATAAAA TCGCAGTAAGTAA ATTTCCACTGTTA TTAAATTTAAAAC		RPPMARFLKTFASQDCPVSNTSSINLPYYMD TDTDXTDCENLSHIMKNLDSSAPGMDLIT GGDWKKISPKHELITAICNCILRNKVCPEKW KLFRTVLILKPGKMSESWRPLAIMDT AYRIFTTLLNNRLLQWIRNGNLISPNQKAIGI PDGCAEHNATLHFAIDRAKRCKTELHIVWL DIADXFGLPHDLIWYTLANMGLKNETLTLI KELYKDVKTFIDCQGTLEPVPITKGVKQGC PLSMTLFCLSIDYILKSILTNYPFLHDLNISIL AYADDLVLLSDSYLEIKKSLESTVELAAFAN LKFKPSKSGYLSINNVNSDILKLHLYNEEIPTI SENNKYRYLGVDFSYKRNDVDGRLGSAL ALTRSLFKSYLHPAQKLNAYKTFIHSLKIFSL RNCVIGHRILDCDRNRVTQGREKQLGFDQEI KALLKTMIGDKFQAXNNYFPYTHCKLGGLG ITSAIDEYLIQSITGITRLFHSSNLSFRKMLITE LAHSRGGKNFEAGLKWLNCEVNKAFPNTSF FVKFQKSALALKRKFCCVNLKFVEDNFSLE MTYKKRTSYVNHQNLSTLSKELHDFVGLYY AEQXCQMRVQGHIAITAIGDSITAKYLIASDI LNDAYYYFLVRARNLLNLNAYRLKYN IGTKCRLCHLDEETQAHXFNHCRAKPNARR VKHENVLVSIVAFLEKIGFEIDVEKSPKYISIP
--	--	--	--	---	--	--

				AAAATTCCTTTAA AAATGCCTCTCTT TTTCAGTAATAAC ACCTTTTCTTGCTT TTATTACTATTTCT TGTGTACTGTACA AATCGAGCACAGT TATTGCAAATAGG ACATAGAAATTCC TTTTTAAGTAAAT TTAAATCCATGAG AAATAAAATAAA ATCCTTTTGATTC AAAGTTTCTATGT TGCTTTCTAATAG AATGGTGTAAGCA TTAATGGGTCTTG ATTTTATAAATT AAATATATTTAAT CTATTAAATTAAT ATGTTTTTATTAAT TATTAATTTTATA GTGGGGGGAAATT	TKLKPDMVIRSKRNKDIHVLDLKVYPYDSGE GFEKAREDNYVKYKDLAIEIGKAFNQKATIS AVVIGCLGTWDDKNNNAALSKIGLTKTEIISL ARIACPNAVIACYHIYREHVSFTKSAMALPF SLA (SEQ ID NO: 1421)
--	--	--	--	--	--

				AATATACTTGATC CCAAGAATCAACT GATGATGAAGAAT ATGTTATTTCAAA ATACATACAAGAA GCTGGAAAAAAC AAATCAATCGCTA CA (SEQ ID NO: 1176)		
NeS L	R5	AY21 6701	Girardia tigrina	GTAGGTAACATG ACTGCAAAATAAT AATTCTACACCTA TTGTTGATAACTC ATCTCGTGCGCAA ACGGAGCATGTTA TTTCTAATCATTTT GTCACACAGGATT CTTCTAATTCTGA TAGTAATATTATA GATAGAGATAGG AACCTTGTTGATT TAGATGCGTCAAT AACTTCTCCTACT	TGATCCGTGTGTTTGTG TCGTATGATTGTTTCCG TGTGTGTCTATATTTT CTTTTTTATACTTTCAA TTACCTCGTTGTAATGT TATAACTTCATATGGAA TATATGTAATTTAGTTT AGTTTAGTTAGTTTAGT TTAGTTTAGTTTAGTTT AGTTAGTTTAGTTAGTT TAGTTAGT (SEQ ID NO: 1300)	TTGRNLGQWSCYSRSIQQSNYSFKLSSTEVG ELVEQSPAPLQSPQFSNNYNNLNINNNLYYS LNTFNQSNLCCLVNIEFFPTQHLLGDIVNS GCINYMNNYNNFDNINLYINSNVLSYNNYN HSFLASPYTTNITEHADINMHVQEVNMQQD NNTQHAIQQVSLQATSLQHTLDEMIVQFN TAVRLKKKHKVAKIFRGHNHRKDLPTLPAR EQYKTKPKLAIREVLHRKTTATSSPSENAIK AFFSSYSRPAELFTGQELLESWFPVHPEDDF EFRIPGRDQIAKYIKFASKSAAGLDWITYEDI KLGDPGSEILQPIFEYIVQNNICPSEGKASRTI MIPKPGKSDYSDPSSWRPITITSVYRLLMK YLTWELYNWILLNQMLSRSQKSLGKFEGCH DHNAMLNMLIQDVERRQTNPSNPINKNKRLY

				ATTATACAGCCAG AGGATAGTAAGAT ATCTGAGGATGAG GACTTCATCTTAG TCAATAGGAAAAA GAGCAAAAATAA GAAAAAATCTAAG AAAACAACCTGAA AATAAAAATGAA ATTCCTATTCAAA AGAGTAAAGATA AGAAAAAGAAGT CTAAAATTAATAC CGAAAAACTAACT GAAAATATTACTA CTTCTGAAATACC ACTTGAAATTGCT CCTTCCATACCTTT ACCTTCAGCAAGT ACCTCGGGTTCTC AACAACCGGCCAA TCCTCCAGAAGAC GCTACTCTAAGTG		IVFLDFTNAFGSVPLDTLMYVPQRFGLGTSA LTLIKNLYLDNYTNVTCGESKIENVKLNKG VKQGCPLSMLLFNIFINIIIRAIEAMPDVHGY PLGDMDIRILAYADDIALISDSHKDLQEMVY KAEYIGRILGLLFNPSKCALMDIPHDKKRTP PILVNGEMIKCVGKADPYKYLGTFRSWFRK LDIKELLQMMMDETCLITESNLHPHQKIHA YETFIHSQLPFHLRHSRIPFSDFITNRKTNKTT NNSNDSEKSIQKAYDPESGQLFLNTFALPSG CAKDDFFYITKDAGGPQLTSGLDEYLIQSIMYI FRLLGSEDPTLNSAIKHDLISHLNLKGFVNIN FSQAISIFNSNFTDRDTHFSHLSRTEWARLQL ARKKLKSTLAIQTNVCLINGHLVLTLSLENN VLLIDSKEKGDVKKIHASLMGFLRLAHLIRL QKHGWSKLLFSATTHHEILNKRILNGHVPY KIWYFIHRARLGLLPTKLFSVSNLCRKCGGK KETMSHALVNCPMMQTLINERHDALEISLV QILSSKFQGTVIRQKTYVNELRPDITMESDT QYYLVEVKCPFDTKMSFELRTQQTTDKYNII IEILEDVHPGKEVRLVTFIVGTLGSWGPQNS DFLRDLGFSKDEIDQVKTRLMLQNINSSCEQ WKRFBVQYAPTITPGPIPAESEDQGTSDNG PTAATVQGPVIGDEEEELQIYDSGLDESSDD
--	--	--	--	--	--	--

				ATACGGATCTCTT CCTTACACAGGAT GATCCCGATAGTC TTATTCTTTCTGGA AGTACTCAACCAA CCTTTGTTGACCT CAACCCTTCACAG CAATCGGAACTTC CTTCAAATACTGA CAGCCAAAGATTT GAGGCGGGTGAA ACACCCAAAATCA TAACTTCTTACAG GGATGACCTTTTC TACTCTACAGTCC TTCACTACAACCTC AGATACAGGTTAC GGTATAAGTGTTG ACAATGGGGAGC AGAGGTTTCGAAT TCTTGCTAGGAAT CTTGTCAGGAAAA CCAAGGATAAGTT	EPDPDDAELLFTIDIEQYLNSVITD (SEQ ID NO: 1422)
--	--	--	--	---	--

				CCCCTCTTTATAT GCTGGACAAGTAA TTAGACACACAGT CTTCTTCAATCAC TTCAACCAGGCAT ACTACGCCAATAA TATAACTGATAGT AAAGGTAATCTAA TTGAGTTTTCTGA TGATAAGCCTTTT CAAAGTATACCGA CTGACCCAAAAAC TGA ACTAGAGCAA ATTAGGAGAGAG AGACAACATCTAG TTGATAGAGCTCT TAGACATAATCAG TTACGGGAAACTT ATATTTTAAATAA ACTTAATAATAAT AATGGGGGGGGT GGCGAACATTTGA AAAGGAAAAAGA		
--	--	--	--	---	--	--

				TCAAAGTCAATAC GGATGATGTCTCC AGCAATGATGGAG ACAGAAAACATA G (SEQ ID NO: 1177)		
NeS L	Utopia- 1B_CPB		Chrysem ys picta bellii	GTTTAATTCCTTCT GATGGACATCTGC AACACCCGTCCTG AAG (SEQ ID NO: 1178)	TAACCGAGACCGCCGA CCAGGGAAATAACCCA CTTCCTTCCCTGACGAA CCAAGGGACGCACCCC ACCCATGTACTTATTCG CTACACCGATACTGACT TGGACTCCTTATACATT CCATGGGTGGCGTACC CGAGCCCACTTATCCAC TGACACTTTAAAACTC TTGCACCCCAATCTGGG TCTATGCCGGTTATGCG ATATGTATGTATCTCTT CATCCTTGCAACCGATA CCTGTAATCCCTCATAA CCCAAGCCTGACCCCA GATGTACAGTACCTTCC CTCTTAACCTCGTGTATA	MESPAXIFEKIDAALXIYSAAAXLXXNSLSL SPXXAXXSXXAAPASSTPQKTQXKPIPXTTL GASRKXRTTXKDEXIXXWXXKAPVDTSXG RXSTRRTALRDLTSRSXNIXXALQEEDPRRT PPXSRDQDAERRPAAPEKAATRGA PPTIQDQ DADRCPAGRDATGGAPRRPRTRMLXAAPL GRMPPEEPPTTRDQDADRRPAAPERDAPE GTTSSTPDPETTYHPPVRRRAAPRGTHSXAX DLDAARCPSPGQRDIVASESSTPPGATSPPQA SLPDXEESPAESAGTTEVRPTEGEAGEDDCI YLQYPXPTGLLLCPFCLPXHGVTLGALSK HVRKAHNKRIAFRCSRCDAPFETQKKCKXH XATCKGPLTTAKVNPTDTRLRVPTPTPDGPA SAPQPASPEPQXVRGDQPPTEGSATPASRTD DATKRTSPASRIPTLDPVVRGITATSQVSDLT RCLSDLIKTIRHNTDTRRXSAPPQVTSCRPA VGATSTAPQAARRDPANGGASRSPQIPRPDP APGRPNTSSKVTQRDSDRQKPHAPPRTPQPD

				TTTAATTTTAAACATTA ACTTTAATAAAATTTT AAA (SEQ ID NO: 1301)	TTRRRTRTIPSASKHDRAPTKPNTGVSRTPLP PGRSSAASETPRAAPPPPHQDPRLKTHLNTA PQSEGQQGHRLSPQHLPNRRQRSRRNDGRE QRVATPWQSAWMEELAKAEDFETFDLMD RLTAELSAEITARRREPQEASRATRRFPAPTR NNTAREGRRGDVGRRYDPAAASRIQKLYRT NRTKAMREILDGTSSYCAIQPERLYSYFKDV FDHEAQTNLQRPECLLPLPRINLTEDLERDFS PQEVQARLMRTKNTAPGKDGIRYHLLKKRD PGCLVLAAIFTKCKQFHRVPRSWKKSMTVLI HKKGERDDPGNWRPISLCSTIYKLYASCLAA RITDWSVCGGAVSSVQKGFMSCEGCYEHNF LLQTAIQEARRSKRQCAVAWLDLTNAFGSIP HHHIFATLGEFGMPETFIQILRDLYKDCTTTI RATDGETDAIPIRRGVKQGCPLSPIIFNLAME PLIRAISSGPTGFDLHGKKISILAYADDLALV ADSSESLQQMLDVTSQAAEWMGLRFNPKK CASLHVDGGARALVRPSRFLIQGEPMASLEE GEVYQHLGTPTGVRVRQTPEDTIAEILRDAA QIDSSLAPWQKINALNTFLIPRISFVLRGSA VAKVPLNKADSTIRQLVKKWLYLPQRASTD IYISHRQGGANVPRMGDLCDVAVMTHAFR LLTCPDPTVRSIAQEAVRDVVRKRIARAPSE
--	--	--	--	--	---

						QDIATYLSGSLEAEFGREGGDLSSLWSRARN ASRRLGKRIGCCWKWCEERRELGILVPRIKT PDHTIVTPTARAMLERTLKDAIRCHYAENLK RKPDQGKVFVSSKWDASNHFLPGGSFTRF ADWRFVHRARLNCVPLNGAIRHGNRDKRC RKCGYANETLPHVLCGCKQHSGAWRHRHN AIQNRLVKAIPPSLGKITLDSAIPGTDSRLRP DIVVTDAEKKKVLMDVTVPFENRSPAFHE ARARKALKYTPLAETLRAQGYEVQIHALIV GALGSWDPHNEPVLACGVGRRYARLMRQ LMVSDTIRWSRDIYTEHITGHRQYHTE (SEQ ID NO: 1423)
NeS L	Utopia- 1_ACa		Acantha moeba castellani i	CCCGTCAAGGGTG CTCCACGAGATCC CTGTCGCTAGCCG ACCGGTTTTACCA CCCCACCCCGCCC GGACAACCACGG ACCCTGCTCCGCA GCAGGACCCACG CACG (SEQ ID NO: 1179)	TAACAACCATGTATGG TGAACCACACCTCTCTC GATCTTGTATTCTGTGA TTGGACATCAGAGTTCC TGCGAAGGGATACACT CTGCCAATCTCGTGGGT TGTAATAAATCCACAC CTTCAACA (SEQ ID NO: 1302)	MAAKSVACPHDGCANKYASEASLRRHIKN KHATDEEGDETSHSCPHCHRPFSTARGLSVH IGKSHRQAPPEPTRPPPAPAPADPGLDPDPGP TVTPPSRDEEDREEDDDPVEIADLSCPHCA QALPSAHGLANHLRACKDHRVPAPGAPRSG PPSSRYWTAVEHHRYVEAMARFADHPDLL ARAAAHIGTRTYKQVDSHRTKVIAAEREGR PVRTLDPMDWRMRPYCASTARWLAEQG RSPVAPRSPCPEPHAPPPAAALLYIPATPPAP TPRAPVAPPKLAPPAESTVPATPDGNPEAPA PPFSAPGPPTPKALPPPPPSRRNLRPHLVPKD

						AWQGVADAVAPAASRLLRTPLAHLSTEQW ATFEAALAGLEATLHHAARSAEAVPTRCAS RAREDAERQLREARKTREIFGKAAALYAAG KDPTATIERIPPEVRLHLPTPGSAEWPARAA AARRVIRRAVARADRLRKRMGILDSRDRLQ RLFNANQKKAVRQILAPSTKAPRCQLDPAA VEEAYIQTLAKPPPIDPSPPWKNSVQWPRPP TAADDGGSPFSVAEVRAQLRRLPNGSAPGID GIPYEAYKRTKLDATLAHVFEVVRLNARLP ARWDVARTVLLYKKGDPNDTGNWRPISLQ VTIYKIFTAALSKRLISWAGKHNTFSASQKG FLPAEGCHEHAFVLRSVLDDARRHKQNVYL AWYDLRNAFGSVSHDLIAWCAAMLGLPRY LRDAIGAIYRHSALFVQVG DQETTGVIPMRC GVKQGCPLSPLLFNLCVEPALRCLRRTTGYK FYGTSITVEGQAYADDLLTAAPSAYHAARQ VATIEEWANWAGVSFVVQALS LDAPAGKC AALAINFEGGLMHSIDPALKVQGAAIPAMSR NNVYRYLGVHVGLTDALGQANELLEKASR DARTICASGLEPWQKVVAIKTFILSRLPFFFH NGKIQRGRCQQFDRELREN LRAALRLPVCT TNAFFHSRVASGGLGILPIAEEQQVYLAHV FKLLTSPDLSIRAIARHQLAEVTHARHTTPV
--	--	--	--	--	--	--

						QDGEASPFPGWLMRGQEVASTTPSGDVSSI WFAAAGAYSRMGWSVRDALHPTLTVGPGV QFEGRFQRANVIPALRASAFSRHAVEWSAL RTQGRAAAAYQHAVHPATHHHVHNSAGLTT KEYRFAIKCRLGLLPTRAAPHHRNGPTACR ACSYARETANHVLGHCPATKAEVIARHNRI CRALAQAAEASWTSVLEDVPIPGVDSPLRPD IYCSRPGQCAIIEVAVSYEDAFNASMEGRAK QKTDKYAGLAATVEEQLRLQTRHAAFVVG FSGVVLPA SVTATATSLDLPPKTWNVLLKR CVAASIKGSYTAWRRFRRSTP (SEQ ID NO: 1424)
NeS L	Utopia- 1_AEc		Acromyr mex echinatio r	GGTGCACAACGGA TGCATCATACGTG TACCGGAGCATAC GGGCTGTCACGGC GGCTGCATGCGCG ATCTAGCTCGGAG ATTTTATTTATTTA TTTATTAATTTATT TATTTATTCATCG AGTGTGAGTGTTT GCGTTTTGCCGAG	TAAATTATTTTGTCTTT GTCTTGGCCCCCCTTT TTAAACCAAGCAGGAG AGAGTGGCCCAATGCC CAACTATTATATATTAA CTATTTACTGTGATATT TATTATTTGACTGTTGG GCGGGCCCCCTCTCTGCT GGTTTTATTTATATATA TTTTTTACTCGCGTACT TTTTGTACTACTCTATT	VCSVRGCRREDSRRFYKFKFPLNFVKVPKTI VIGSAFQKSSVSARSQNHSTRVPKTRQPR TSNTIGRYTAASANNYLTVIITGNYTVFAQW ICYRECTWLLSKFVNFFLTIIGYFFQLRLVVI YEGPVILDTFSNCGSSLFMRGQXSKALLVRL NRSALAMADPQVHYIDYPLPPRVKCVKCFG AEGAGKVKGEYSDPPHLAKHLKKCHPGDT LNYKCSICDLRGTKYPLRDVKAHYAECHV SPAVDAAGPSTRGSLGEC SGAGQPTASRAA KATTRLAETVGGTDKRRATSGSRQLTLPF AATPSPSTAAGEARAPRSXSTTPTSRSPSYA

				AAGCGATTTTCGT TAAGTGATACGCG CCGCGTTCATAGG TTAG (SEQ ID NO: 1180)	TTTCTTTTTATTTTAGCT ATGCTATTTTTATCTCT TTCTTTGTCTCTATTTTC TTTCTTTTTTTCTTTTCCT TTTCTTTTCTTTCTTTTA TTCTTCTTTTATTTATCT TTTTTCTTTTCTGTTGT GGGGCCCTGACCGTCC GAGTGTGAATGCCGCG AAAAACAATATTATGT TTTATACGAGTGTGCAT GTGCGTGATATATTTAT CTATTTTATTTTATTTAT TTATTATAATTTATTGC CGCGCGCGCTCCTCCG GGACTTTTATTCGTTGA CAATACTGTGATATTTT TCTGCKCAGGCTGGGG GGGCTTGCCCCCAGC CCCTTAGTTTTAATTGC CTATGCGGGGGGGGCT TTTGTCCCCCGCAAATG TATATATATATATATT	AVTAGPPSMRSTTTSTTARSKTVAKGAAPN TTTTTTARRSGEAAATRKPTTATVSKPRVL SVETVRLPVDDIQRAGVQNAAKPARAPSRP PQRTSPEAGGPRTTGAKEKCGEGAYKKLPA NSGNPISTRTRRATSVPVEKSEGTARRERVS PHPPPKGIDIILSSTSEEEGTPYQPGGVGRLRL RRKKVTGPPPKMTPREGVVTRARRSTSAPV EKSALDARLTALDRTSSRATGNPTSQIAGGL YTSRGQPERTPPARLPSLSPTTRGSPSGSLGEI RTPISPATSLPATLTTCTVTTTTCGSPITSTGF TGGVGRLITPPSLPQTNILPTIGEEGTSPCVA VVTTHPRPTGEDAPCEAPQPVSDHQRQSIGE PRRDTDTHLACDVATGNAHHLHGDDDLW KPHNIHGLHRWRGEADNTXEPPNEHPPDH RGGRNVTVRGGRHHPSLGGRSTSPLILPRPT TPEPERGQEERRLEGAAQPPTTPVVEGDNQ WDGQWTVSVRRRARRQQLNDTSPSNSSESPP TAGPSRSPRIAPLSALIAASTSRHETSLNLNC TNGNICMDRTPPRNILPVXAERRRETSPQDR VEGDIGYGAGKVSAEHPSAPVNVRGVMSR GRATASSIVPPRANRGEGRQHHSRRRPDAP VGQPSRDHPAPATVARQRRRERVAARDALL DRAKDVATIADLEAFAASVAFFGEDASAT
--	--	--	--	--	---	--

				AGCGCGCGGCTTAGCC GCTTTTGTGTATTAC CCCAGAGGGGAATTGT CCCTCTGGGGAAAAAA AATGATTGAAAAATA AAGTGAGCTAA (SEQ ID NO: 1303)	GAAARARDSVRSREAGARRGVKGGERPE REGAGRPGSAPADPGASGEARGDWVREAK RLQALYRANRRKAVREVLQGPADQCQVPK RQVQEYFERLYSGGEDLAGAGVEAERPDP SPREVS AVLGPLAEREVDRRLRRMNNSAPG PDGVSYRDLRGADR GARLLTALYNICLRLE AVPASWKTSNTVLIHKKGDRGMLNWRPL ALGDTVPKLFAALLADRLTDWAVTRGKLCS AQKGFLRDEGCYEHNFLQEVLT HAKRSKR QAVVAWLDSLNAFGSIPHATIRRALIRSAVP RGLIAIWDSMYDGCTTRV RTAEGHTAPIPIR SGVRQGCPLSPIIFNL AIDSVVRVAAEXNDG YSLHGNTWSALAYADDIALLAQTPEGMER MLASVEAEAASVGLRFNPAKCATLHV GAG NGGRVLPTS FQIQGETINPLAQGESYTHLGV PTGFSVDQTPYAAVGDIVSDLRAVDRSLLAP WQKIEMLGTFILSR LD FLLRGARVFKGPLTA VDLNIRRHVKSWLNLPQRASAEGVYMPPR WGGCGLLPLSDLADVLTV AHAYRMLTVRD GAVRELAWESLRGVVGRRIGHAPSCEDIASF LSGSLDGRMRGGGEASLWSSARNAALRQSE RLSLRWRWVEATEEMTLECRGPRGAAIKIPP EARGQVVNRLRS AVAEHYASRLLSKPDQ GK
--	--	--	--	--	--

						VFEVSSRSRVSNHFIRGGSFTRFADWRFIHK ARLDVLPNGARRWEANDKRCRRRCGEVSE TLPHVLCHCGIHSAAIQLRHDAVLHRLWKA TRLPGVVRVNQRVEGVSDDELGALRPDLVVR HEPSKSVVICDVTVPFENRWTAFEDARARKI AKYSPLAEELQRRGYRVVVTAFVVGALGS WDPRNEAVLRLLRVGNQYAAMMRRLIVSD TIRWSRDIYVEHVSQTRQYLAPSRPSGDLAT PPRAVRRRWLAEERSAQDAARRGSDSVSVA (SEQ ID NO: 1425)
NeS L	Utopia- 1_AMi	.	Alligator mississip piensis	TGCTGGAAAGACG GAGAACCGCTTCC TTTTTCCCTGCGCC TGGCCTGGTATTG CAGTACCTCCAGG ATTAGCGCCAACT AGTCCGGCAGACT GTCGGAATACAGC AATAGAAAGWGA GCTGACTAGCAGC TTGCTTTCCTTCCT CCGGTGCAGCATG GGTTCTCGTCAGT	TGAACCCCCCTCTGCA CCAGATGGACCTTCACT TCGAGAGGATTCTTCA GCAATGGACGACCCCG CTCCACCCGAAGAGGA CCCCCGCGATGAGACT CTATATGGACTGAGAC ACTTTTTCTTCGAACCA CTTCCTCCACCATTGCG GACCATTGTAACGGGT TTGTGTGTATCTATCTT CTTTCTCTCTCAGCGTC GCGAACCCCCCTCCCTCC	CHHAGLRPGTPNRRTRPDQTAPLPDPRGHP MPPNRRGSRSRPEEPSRREPPXPACQGLRV WSPPQQRMPWPQTLWLEELSRATTFKAFF ASVARLTEELSAAARPGQPRGGNNRPATRR DHRLQPQRRPRRQRYDPAAASRIQKLYRAN RPKAVREILEGPSAFCQVPRETLENYFSRVF NPPAEAAAPRPATVEALTPVPPAEGFEDAFT PQEVEARLKRTRDTAPGRDGIRYSLLKKRDP GCLVLSVLFNRCREFRRTPTTWKRAMTVLI HKKGDPTDPGNWRPIALCSTVAKLYASCLA ARITDWAVTGGAVSRSQKGFMSSTEGCYEH NFTLQMALDNARRTRKQCAVAWLDISNAF GSVPHRHIFGTLRELGLPDGVIDLVRELYHG

				CMTGACGGGCTA GGGAAGGCGGTG CTGCCAGTACGTC CGAAAGAGTGCCG GTTGCGCAAGCGA CCGCGCCACTCAG GTGAGTAGCCAAG GGTCTTACAGTTC ACCGGACCCGAW AACGCGAAAACCC CAACTCGGGCTAG TAGCCGAAGACCT GGGTCCCCCCCMG GTCAGAGTAGGCG AACGCCWGKGCT CAGAGGACGGAA CGCGGAAAACACC CCCAGGTCCCAAG GACGCCCTGATCC ACTGACAAGAACG CTCGAGGCACGCC AGGAGACCCCCAG CTAGGGTGGACCG	CCTTCCCCTCCCCCTCC CCCCACCCCCGGGCTT AGTTGGCTAACATTGTA TCTCCTGTAACTAGTT GCGTTCCCCTCCTCACC CCCATCCCTCTATTGTT AGTCCCTCGCTCGGGC GCTCTGTATTTCCCTAC CGGCTTTGTCATCTTTT TTGGATTCACAATCCTA AACATCTACTAATAAA AGTCAATC (SEQ ID NO: 1304)	CTTTVRATDGETAEIPIRSGVRQGCPLSPIIFN LAMEPLLRAVAGGPGGLDLYGQKLSVLAY ADDLVLLAPDATQLQQMLDVTSEARWMG LRFNVAKCASLHIDGRQKSRVLDSTLTIQGG AMRHLRDGEAYCHLGTPTGHRAKQTPEETI NGIVQDAHKLDSSLLAPWQKIDAVNTFLIPR VAFVLRGSAPKTPKKADAEIRRLKKWL HLPLRASNEVLHIPYRQGGANVPRMGDLCD IAVVTHAFRLTCPDXTVSIIAASALEETARK RIGRQPTRRDLATFLSGSLEGEFSRDGGDFA SLWSRARNATRRLGKRIGCAWTWTEERREL GVSLQPAPHADRVTVTPRTRTFLEFLKDA VRNKYAGDLRAKPDQGKVFDVTSKWSSN HFMPSGSFTRFADWRFLHRARLNCLPLNGA VRFGHRDKRCRRCGYVAETLPHVLCSCCKPH ARAWQLCHNAVQDRLVRAIPAAAGEISVNR TVPGCESQMRPDIVITNEEAKKVVIVDVTIPF ENRRQAFTDARARKREKYAPLADILRGRGY DVTVDALIVGTLGAWDPSNESVLHACRVSR RYAKLMRCLMVSDTIRWSRDIYVEHITGHR QYTDPTRRTAAGPDPEGTA (SEQ ID NO: 1426)
--	--	--	--	--	--	---

				CCGACTGCAGGTC CGGAGGACCCTCC CAGGAGGGTGGA CCAGCGAACCCTAA GTTGGCGACGAAC CCTGACGCACCCC CCACGATGTCAGG ACCCCGACAGGCG GCGGTGGACCACT GACCATCGACCGA CCCCCAGAGGCAG AGAGACTCTCAGA GCCCCGAACCCCG GCTGACGAGAGCC GCCTCCCGGCGGA GGACCCCGGAGCC TGAGGATGCCCCC CGGATGACGGCGG AGCGCCCCGAGCG ACAGCGGACCCCT CCGGACCCCCACG GCCCCTCGGTGAC GATGGCGGGCCCC		
--	--	--	--	---	--	--

				GAACGACGACGA CCCCCGGACCCCG GCGGTCCCGAGGA CGCCCCCCCCGAG GGTCTCCCCACGC TGGTGGAGGAGCC CCGGACCCCCCCG ACACCGGACCCCC CCACGGACGACCC AGGCGAAGGCGT AGACATGACAGCA CTCACGTTCTCC CCTTCCCCCTCCC GGCGAAGCTGTTC TGCCCGACCTGCC ACCCGCCAAGACA GTACAGGTCGCAC GGCGACATGAACA AGCACCTACGGCG CTTCCACCAGCTG CGCCTAGCCTTCT ACTGCGCCCTCTG CGGCACCGAGTAC		
--	--	--	--	--	--	--

				GAGGCCCTGAAGC TCCTGAAGAACCA CCAGAAGGGATGC GAGGGCCACGGA GCCGAGAGGAGA CCCGGCACGCTGG TGAGGTCCGCTGC CCCGGCCCCGCCGG ACCCAGGCCGCGG TGCGAAGGCCCGC CAGACTGGCCACC CCGCCGACAACCC CACCGGACCAGAC CTCCAGGGACCAC CCGACGGAGAGA CCTGCCCCAGTGA (SEQ ID NO: 1181)		
NeS L	Utopia- 1_CMy		Chelonia mydas	CTCTTCTTATGAA TACTTGCAACACC TGCACTGAAGATG GATTCTCCGGCTG CTATTTTGTAAAA ACTGATGCTGCTT	TGAGCCGGTACGACAT CGTGCATCAACTATGA GAAAGGGACTGAGAGA CTTTTCCATTGGACCA TATGAACTGGAACCAT AAACTCACTGAACATT	MTTKKVLGASTTLQTSSTKGKNSGCSKDPL RDAVPGRSWILRPACRDITTRRNIPPAPQQQ QPPMESPPTLQLQDALRRPSPTPAAAQVAD AGGALAALHTIKRGISVDWTSISPKXXQRXT SASPDACPASETTQRDXRXLLDARPAGPLDP TRPHQDEPASDTADAAGTPLLQGNEDTIYL

				TGAAGGTGTATTC TGCTGCTGCTACC TTGGAAGGAAATT CTCTCTCTGCTCCT GAGACATCCCCAG CTGCACCGTGTAC CACCACCACCACT GCTGCTGCTCCAC AGAAGGTTTCTCG GACA (SEQ ID NO: 1182)	AAATCTCACCAAATGA GGGTAAATCCATCCTC ATCATCGTATCCACTCA TTATACTCCACACCTGA ACATAGCCATTATATG AACAAACATACCCCAT ATCTCAATGTCTGTACT TTGACCCGTTAACCTTT TACCCCAATCGGGGA TATTGCAGATTATGTAT TCCTTACGCCACCCGAT CCTAAACCGAATTTTCGC ACCCCTTGATAATCTGT ACCTTATTCCTGATAA CCAGAACTTCTATGCT TAAACTCTGTACCGTTT TTTTTTATTTCAACATC ATCTTAATAAAATTATT AAA (SEQ ID NO: 1305)	QYPLAADMLICPICSPQSFHLLGVVTRHLK RCHSKRVAFSCALCSLPFETQKQCKMHQVA CRKCLKGTTQSPAPAPSPPAARRPAAPEPQR RKXTSQAQVKKPAPVARPAERDAAIEKVPA ASGNITQVLASRRPVSPSHVAKXISMLRRLS AASPPVQHVPVPRRISAPPRIAARDPVAGRA SAAPQ TALRTPAAGGASTTPQTALRTPTAG GASAMPQTTLXPXRRPDWRNQPRSHSKAPG LHRQTDQHGPQVHSAGHCLREISRSSSNRLG SSHSAAATHRRTGGVPATPEPDRVSPPTSNA XIPPEIPPQHPTGPNPDPRDRRQADHTAGSEP APDEVEDXEGQRPVMVRAATPWQTAWTEEL QAAASFDDFDLLVDRLTRELSAEIAPRRSSN QENAPPAHRTAPNHNTTTRGARSRDASRR YDPAAASRIQKLYRANRSKAMREILDGSPSY CTIPSERLYSYFKDVFDRIRNDAQRPECLRP LPRVDEAGVLETDXTPKEVMARLSKTKNTA PGKDGIPYSLLKKRDPGCLVLATLFNQCKRF CRTPSSWKKAMTVLVYKKGERDDPSNWRPI SLCSTMYKLYASCLASRITEWSVSGGAISSIQ KGFMSCEGCEYHNFLVLTQTTIETARRARRQC AVAWLDLANAFGSMPPHHIFATLQEFGMPE NFLRVIREVYEGCSTTIRSVEGETAEIPISGV
--	--	--	--	---	--	---

						KQGCPLSPIIFNLAMEPLLRAISNGTDGFNLH GERVSVLAYADDL VLTADDPESLQGMLDA TSRAADWMGLRFNAKKCATLHIDGSKRDS VQTTGFQIQGEPVIPLAEGQAYQHLGTPTGF RVRQTPEDTIQEILQDAAKIDASLLAPWQKI NALNTFLIPRISFVLRGSAVAKVPLNKADKI VRQLVKKWLFLPQRASNELVYIAHRHGA NVPRMGDLCDIAVITHAFRLLTCPDAMVRNI AANALHDATKKRIGRAPSNQDIATFLSGSLD GEFGRDGRDIASLWSRARNATRRLGKRIGC RWEWCEERQELGVLVPQIRSNDNTIVTPSAR GMLERTLKA AIHSLYVETLKRKPDQGKA FE LTSKWDASQPLPRRGRLHPFRRLAVHPPCPA QLRPAQRSRPPPREPRQALQEVRLLQRDPAPR PVQLQAPLQSLAAAPQCHPEPPGESHRTAPG GGRRELRHPRYPASGTPANHFLAGGGFTRF ADWRFIHRARLNCVPLNGAVRHGNRDKRC RKCYSNETLPHVLC SCKPHSRAWQLRHNA IQNRLVKAIAPRLGEVAVNCAIPGTDSQLRP DVVVTD EAQKKIILVDVTVSFENRTPAFREA RARKLEKYAPLADTLRAKGYEVQMDALIV GALGAWDPCNERVLRTCGIGRRYARLMRR
--	--	--	--	--	--	---

						LMVSDTIRWSRDIYIEHITGHRQYQEV (SEQ ID NO: 1427)
NeS L	Utopia- 1_CPB	.	Chrysem ys picta bellii	TTTTTCTGATGCT TGACTGCAAACAC CCATCCAGAAGAT GGAATCTCCTGCA GCCATTTTGGAAA AAATTGATGCTGC TTTAAAGATATAC TCCATTCTCCTWK TTTGKAAGAAAAC TCTTTTTTCAGCTTC AGCTATTCTGTCA TCGGCTGCTGCTG TTCCTGCTTCCCA GAAAGCTCAGCM AAAACCTATCCTG AAGACCWCCCTTG GTGCCTCACGGAA GACCCGGASCACC TGCAAGAACCAAA ACATTAGGAGCTG GCTGAAGAAACCC	TGAGCCAGAGTGACAT CGTTCTCCCACTACGAG AAAGGGACCAAGTGAC CTTCTCCGTTGGATCAT ATGAACTGGAACCATA AACTCCCTGAACATTA AATCTCACCAAATGAG GGTCAATCCATCCTCAT CATCATATCCACTCATT ATAMTCCACACCCGAA CACAGCCACTCTATGA ACTTCATACCCTCATAT CTCAATGTCTGTACTTT GACCCATCAACCTTTTA CCCCCAATCGGGGATA TTGCAGATTATGTATTC CTCATGCCACCTGATCT TAAACCAAACCTTTGCA CCCTCGATAATCTGTAT GTTATTCCCTGATAACC AGAAACTTCTATGCTCA	MTQDQDADCCPAGKDATRGAAPPMTQDQD ADRCPAAPERDAPEGTTSSTPDPKTTYHPAV RRRAARRGMHLRAQDLDAARCPSGQRDNV ASESSAPPRATSPPPQASLPDPEESPGESAGTT EIRPTEGEAGEEDRIYLQYPLPTGLLLCPFCL PVHGVQTLAALSKHVRKTYNKRIAFRCSRC DLPFETQKKCKFHQATCRGPPTTAKVNPTDI LRVPTLTPTDDLASAPQPASPESQQIRGDQPP TEGSVTPASRTDDATKRTSPVSRIPTLDPAV RGTTATSQVNNLTRRLSDLIKTIRHNTDTRR CSAPPQVTSCRPAVGATSIVPQAARRDPANG GASRSPQIPQDPAPGRPNTSSKVTQRASDR QKPHAPPRTHQPDAARRRTRTIPSASKHDRA PTKPSTGASRTPLPPGRSSAASET PRAALPTT PGPPPQDPPEHRSTVRGTTTRPQTVPAAPEPA ETTQQEERRPRARVATPWQSAWMEELAKA EDFENFDLMDRLTAELSAEITARRREPQEA ARATRRFPAPSRNNTAREGRRGDVGRRYDP AAASRIQKLYRMNRTKAMREILDGTSSYCAI QPERLYSYFKDVFDHEAQTNLRRPECLSPLP RIDLTEDLERDFSPQEVQARLSRTKNTAPGK

				CCCGTGGATACCT CWGCAGGGAGAC CTGGSTCCAGMAG GACAKCTCTTCGG GACCTCMCATCSA GGAGCAAGAATAT CTCAACAGCTCTT CAGGAGGGGGAC CCCC GGAGAACCC TGCCCGCTTCCCA GAACCAGGATGCT GATCGCCGCCCA CCGGGAAGGATGC CACCGCAGGAGCC CCCCCA (SEQ ID NO: 1183)	AACTCTGTTCACTATTT TTTTAAACATCATCTTA ATAAAATTTTAAATCT GTT (SEQ ID NO: 1306)	DGIRYPLLKKRDPGCLVLA AIFNKCKQFHRV PRSWKKSMTVLIHKKGXRDDPGNWRPISLC STIYKLYASCLAARITDWSVCGGAVSSVQK GFMSCEGCYEHNFLQTAIQEARRSKRQCA VAWLDLTNAFGSIPHHHIFATLGEFGMPETF IQILRDLYKDCTTTIRATDGETDAIPIRRGVK QGCPLSPIIFNLAMEPLIRAISSGPTGFDLHGK KLSILAYADDLVLTADDPESLQGMLDATSR ATDWMGLRFNAKKCATLHIDGSKRDSVQT TGFQIQGEPVIPLAEGQAYQHLGTPTGFRVR QTPEDTIQEILQDAAKIDASLLAPWQKINAL NTFLIPRISFTLRGSAVAKVPLNKADKIIRKL VKKWFLFPQRASNELVYIAHRHGGANVPR MGDLCDVAVITHAFRLLTCPDATVRNIAAN ALRDATEKRIGRAPSNQDIATFLSGSLDGEF GRDGRDIASLWSRTRNATRRLGKRIGCRWE WCEERQELGIRVPQIRSDDNTIVTPTARGLL ERTLKAAIRSLYVETLKRKPDQGKAFELTSK WDASNHFLDGGGFTRFADWRFIHRARLNC VPLNGAVRHGNRDKRCRCKGYPNETLPHV LCCKPHSRAWQLRHNAIQNRLVKAIAPRL GEISVNCTIAGTDSQLRPDVVVVTDEAQKKIIL VDVTVSFENRTPAFREARARKLEKYAPLAD
--	--	--	--	---	---	--

						TLRAKGYEVQMDALIVGALGAWDPCNERV LRTCIGIGRRYARLMRRLMVSDAIRWSRDIYI EHITGHRQYQEA (SEQ ID NO: 1428)
NeS L	Utopia- 1_DYak		Drosophi la yakuba	AAAGTGTAGTTCT TTTCTGTTTTAGTG TAGTGGGAAGTCT GTTTCTTTTTATTA TGTTTTTTACGAA AAAGTCCTGGTCT TTGAAATTCATTG TCTAAATTTTAAA TAAAATTATAAAA TTTAAAAAGAAAA TTAATTAAAGAAG CGATGAAATATCT CTGAAATTCAATC AATCAATTAATCA TGGCGTCTCAGCG AGTGCACGTATTT GCCTACCCCTTCG TGGGACCATTCCG GTGCTCCGTATGC ATGGATGCGTCCG	TAAAAAATTTAAATGC CTTAAAAATAAATAAA TATATCAAAATTTAAA AAAAAAAACGAGGAAC AAATAAACACAAATTC TGAAAGATTTATATAAT TTAAAAWATAAATCGA AAATAAATGTTGAAAA CAAAAAAAAAATAATA ATAATAATAAAACAC AATAACACTCACCCGG CCTGCCCCAGAGGCAG GTAAACATTTACTGGCC ATATGGCTTTTTTTTTTA A (SEQ ID NO: 1307)	YAPGYEAAQSPCGREPPRDHRRRPRDACGS SHSPEPCLTTPRLLPETVSAEPCDDESQRTRY ASPHKQARTLHDAEPRDASREHAPSCAEPR CHRCQWTHWKDCCPHSTNTTDGPEGTDRC ADTITSPATAACPQRSPCLGSSNGCDETAPE KRQPAADLVHTAPFAVLVRAGPFADLVRAG PFADHHQDDDPLPHRSGSLGPLCSKQKDPR KTHQHRHSGQAGNQTHTDIPRAAPSRRAAI CLMANAAATREDLLRAATSLSEMAAANQP TRSPTGGGEPTSQGRRGPQALADAAKRIQQI YRTNIPRAMRKVLRTLLTAVFSACLRTGHV PDLCKKSRTVLIHKKGDRTDLSNWRPLSMG DTIPKLFAAVMADRLTAFLTNGGRLSEEQK GFLQHEGCHEHNFVLGQVLEESRRQGKDLV MGWLDLSNAFGSIPHATIMDAVAGMGIPSRI RTIIHQLATGAATAKTIDGMSEEIPIEAGVR QGCPASPILFNIAIERVLRKIKTVNAGYLLYG SRISPLAYADDLVLIASSPEEMRSLRAADD AAIEAGLHFNPKKCATLHLTGKKSSRAVQ TGFLVRGTPIPAMTEGDAYEYLGIPGLKKN

				GGATGCATCCCAC TAGGTCGCTGGGC GAATACGGCACAT ACGCTGCGGCATA TCAGCACATAACC CGGCGCCACCCAC AAGTGGTTATTAC ATACCGTTGCCGG GTCTGTGGCGCTG A (SEQ ID NO: 1184)		QTPRAAMEAIVGDIAKIDDSLLAPWQKIDAA RTFVAPKLDFVLRSGATLRAPLRHLDTVIKK HIKKWLYLPQRASAEVVYTPLKKGGAGILP SSILADVL TIAQAHMVS CPGEVVSRIASEG LREAVKRKINREPSGDEMAHFLSGSTLSGET ASFGDAGFWSRVRMATKRQAVHLGVRWA WRGGELLVESRGQRNRPVATDSNSRSQLIQ RLRCAAQDEFLTILINKPDQGKVAKLSTLTP VSNAFIRDGSFTRFADWRFIHRARLGVLPLN GAIRWGS GDKRCRVCGYQLESVPHVLCHC MHHSNAMQQRHNAVMDRLAKAGSRLGTP RVNCRVEGVAEDMAALRPDLVWRDERSRK IVIVDVTVPFENGAEAFDNARGEKEEKYRPL AEALRAMGYQVKLEAFIVGALGSWDPKNE RVLKTLGVSRFYAGLMRRLMVADTIRWSR DIYVEHVSGIRQFTLPSGAPSN (SEQ ID NO: 1429)
NeS L	Utopia- 1_Gav	.	Gavialis gangetic us	CGCTGGAAAGACG GAGAACCGCTTCT TTTTCTGCGCCC GGCCTGGTATTGC ACTTCTCCAGGA CCAGCGCCAACCT	TGAACCGCCCCCTCC GCGCCAGACGGACCTT CACTTCACTCCGAGAG GATTCTTCGACCACGG ACGACCCCGCTCCACC CGAAGAGGACCCCCGC	MSGPRQAAADPRPSTDPRRQRDSQSPEPRLT RAASRRRTDPEDAPRTTAEHPERRRTPPDP RGPSATTAGPERRRPPDPGGPEDDPPEGLPT LVEEPRTPTPDPPDGRPRRGCRGSAHVPP LPPPCEAAVPDLPPAKAVQVAQRHEQTPTA LPPAAPSVLLLPLRHRVRGPEAPEEPPQGMP

				AGTCCGGCAGACT GCCGGAATAATAG CCTCAGAAAGAGA GCTGGCTAGCAGC CCTCTTTTCTTTCC TCCGGTGCAGCGT GGGTTCTTGTCAG TCCTGATGGGCTA GGGAAGGCGGTG CCGCCAGTACGTC CGAAAGAGCGCC GGTTGCGCGAGCG ACCGCGCCGCTCA GGCGAGTAGCCCA AGGGTCTTACGGT TCGCCGGACCCGA TAACGCGAAAGCC CCGACTCGGGCCA GTAGCCGAAGACC NTGGGCCTCCCTC CCCAGGTCGGAGT AGGCGAACGCCCCG TGCTCGGAGGACG	GATGAGACTCTATACG GACTGAGGCACTTCCTT CGAACCACTTCCTCCAC CATTGCGGACCATTGTA ACGGGTTTGTGTGTATC TATCTCCTTTCTCTCTC AGCGTCGCGAACCCCC TCCCCACCCCCCACCC CCGGGCTTAGTTGGCTA ACATTGTATCTCCTGTA ACCTAGTCGCGTTCCCC TCCTCACCCCCATCCCT CTATTGTTAGTCCCTCG CTCGGGCGATCTGTATT TCCCTATCGGCTTTGTC ATCTTTTTTCTGGATTC CCGATCCTAAACATTTA CTAATAAAAGTCAATC TGTTCTTT (SEQ ID NO: 1308)	GPRGREETRHAGEVRRPTTTRAAARRPARPA APPATPPDQTSGDRPTERPAPATPPRRSAPR DPRPDVTPRPDGGPPGPPGPPDAPDPPRIPEPP GEPEPPGALQLPSVPGSPGAETSAQQRMPTP RQALWLEELSRATAFEAFEASVARLTEELSA AARPGQPRRGADNGPTTRRDHRPQPQRRPR RQRYDPAAASRIQKLYRANRPKAAREILEGP SAFCQVPRETLFNYFSRVFNPPAEAAAPRPA TVEALTPVPPAEGFEEAFTPREVEARLKRTR DTAPGRDGIRYGLLKKRDPGCLVLSVLFNR CREFRRTPAAWKRAMTVLIHKKGDPTDPGN WRPIALCSTVAKLYASCLAARITDWAVTGG AVSRSQKGFMSTEGCYEHNFTLQMALDNA RRTRKQCAVAWLDISNAFGSVPHRRIFGTLR ELGLPDGVIDLVRELYHGCTTTVRATDGET AEIPIRSGVRQGCPLSPIIFNLAMEPLLRVA GGPGGLDLYGQKLSVLAYADDLVLLAPDA TQLQQMLDVTSEAARWMGLRFNVAKCASL HIDGRQKSRVLDSTLTIQQQAMRHRLRDGEA YCHLGTPTGHRAKQTPEETINGIVQDAHKL DSSLLAPWQKIDAANTFLIPRVAFVLRGSAV PKTPLKKADAEIRLLKKWLHLPLRASNEV LHIPYRQGGANVPRMGDLCDIAVVTHAFRL
--	--	--	--	---	---	--

				GAACGTGGACAA AACACCCCCAGGT CCCAATGACGCCC TGATCCACTGACA AGAACGCTCGAGG CACNCCAGGAGAC CCCCAGCTAGGGC AGACCGCCGACCA CGGGTCGCGGAGG ACCCTCCCAGGAG GGTGGACCAGCGA ACCCGAGTCGGCG ACGAACCCCGACG CACCCCCCCC (SEQ ID NO: 1185)		LTCPDATVSIIAASALEETARKRIARQPTGRD LATFLSGSLEGEFGRDGGDFASLWSRARNAL TRRLGKRIGCAWTWTEECRELGVSLQPAPH ADRVTVTPRTRTFLEFLKDAVRNKYAGDL RAKPDQGKVFVDVTSKWDASNHFMPSGSFTR FADWRFLHRARLNCLPLNGAVRFGHRDKR CRRCGYAAETLPHVLCSCCKPHARAWQLRH NAVQDRLVRAIPAAAGEISVNRTVPGCESQ MRPDIVITNEEAKKVVIVDVTIPFENRRQAF DARARKREKYAPLADTLRGRGYDVTVDALI VGTLGAWDPSNESVLRACRVSRRYAKLMR CLMVSDTIRWSRDIYVEHITGHRQYSDPTRR AAAGPDPEGTA (SEQ ID NO: 1430)
NeS L	Utopia- 1_LV	AGC V013 58106	Lytechinus variegatus	ATCTACTATC (SEQ ID NO: 1186)	TGAATAGCATTATATT GTGTTCCAAACAACAT ACTCATTATTATATCTA AACATTTTTTTTTCTGT TCCTGACAATCTACGTA AAGTCTGCTAACCAAC TGGCATGATGAAATAA GATAAAATCCCCTTAC	MSCPREGSDHLGPDPETPALHQQGSDIRVTSS RLRGSRGKSSRQPSSRHQVPASEASATAQQT AANECQVCGSSFATSSGLRRHMARLHRAAS ADPEGAAPASITEIFDYPLPSRWKCSACSENF FNQQLKRHQTRHHPATTTFAYAFRCSSCRSE FDSARRAANHWQVHKKERSQLSGTEPQASS QARVSMASHPPPLPNTSWAELASNP AEIPSF VWESPPKNRPSVEEFGSSLPTDVTMMSQSPP

				ACATTAATTTCTTGTCA CATCATAATGCTTTGTC AAAGCAATGTCCTACA TAATATCTCGATGTCAC CCCAATTAATTTTACAT CCTTCGGTAACCTTTAT ACCGTTGGATCAACAT ATATGATTTGTAAACT GTTATTTCTGAGTTTTT TCTATGCTAATAAA (SEQ ID NO: 1309)	PQVQSSPVPALTPLSPAATASSSPPGAARQLT PPTQTNTPTVTQRRARLQPEADVPELPPSVTE HPVSDAQHWVDAVSSASDWSEFEAVCDQF VIHAVAVSRPNLARPPQQQDRQRSGDHPPRQ QRGQHRPTFDVREASRIQKLYRTSKKRAIRH ILKEKSPSFSGSESDVLDFFREVSASKEVDEE AVGKLASSLFDVPQGDDSATSLSLPTSAKEI GARLSRMTNSAPGKDRLEYRHIRRADGSFSI SEAIENKCLAEGRIPAPWKTAstillHKAGPT DDPANFRPIALQSCLYKLFMAVLADRLTKW ACENQYLSPEQKSARPCGCFEHSFLLSAAL KDCRRNQKTICIGWLDLRNAFGSIPHPVIKIV LSSLGVPDSLVTLLMDAYNGASTSFTLTGG QTDTPVIRSGVKQGCPMSPIFNLAIELIIRAV KKNASDNHLGVTVQGKNLSILAYADDLVLL SRDTEGLQSLLQVAGSSASTLQMQFKPQKC ATLTLDCKRGTNVRQSAHHIQGAAIPSLTEE ERYRYLGVPIGLPRLTSLQESSRKLSSDIETIS SSLLAPWQKLDAIKTFVIPILQYTLRATEYLK SDLKPLRAAIIKHVKKICHLPVRSSNAFVFAS RPSGGLAFVDPGVDADILVVTQAVRTLASD DDTVRAVALGQLTSVVHRTVHSAPSDDCID KFLSGSSEGPLANSGNSGQASSLWSRTRAAS
--	--	--	--	---	--

						RRLKIRIVGASSGDIKVESGGRAIPSKKVTAG LRSDHHNEMSEKLRLSLPDQGKVARALSLDS FANATSWLTSGSFIRFCDWRFIHRARLNCLP TNAAVRRWKQNANTKCRRCDHQLETLPHII NNCRPNMVPPIRRRHNSIQERLVKAIHYGDIY QDQHVPGDPNPRERPDIIVVEGNKVTIVDITI PFDNGPDALSTAANAKVMKYDTLRQELASR GMDVEVHAFVIGSLGSHWDNERVLGRLGI SRRYRTLMMRLCCIDAIKGSRDYIEHVTGH RQY (SEQ ID NO: 1431)
NeS L	Utopia- 1_NVit	.	Nasonia vitripenn is	CCATTCCTTCGTA CGGGTTTTTCGTGC CGGCATAGCCGGG TGGGAGACTCGCG CGGGGGAGGTCAT ATCTCACCACCAT CCTCGAGCTTTTT GCTGCACCTGA (SEQ ID NO: 1187)	TAGTGGGGCCATAACA CCTAGGCCCCACAGTG TGGCGATGTCCATGTGT GTCGTCCTTACTTATTT ATTTATTTGTCCTGTCG AGCTGTTTAATCTATTC ATATGTATGTGTTGTGT GTGTCGTCCTCGTAGC AGTTTTATTCCGTCCAA CCAGAGGTCGACCAAT ATTAAAATAAGCATGG CTTGAAGCAGGCCAAG CGCCGTGTTCTAACCCC	SGXTGREVKCITVNVLMQQPHTKAIREGD FIVILLPQSDDETLCCLCVGRGRYSKTRV ECLNRHVKEVHPDLTTTFRCWGCGFAAPGD KKYPRKIVTQHCATCVPEVSSAPSGRVDGE RRVNTRRRLGIAAATEASPVRRTRRNLASP PVEQNISQSAAPPEPARVPQHPEIVALGESAD DEVFRSPVNSPPRDWRAAAPQQAASSSPXA VPGITAATPSNTTTRTGNGSAXSILAEHPIAP PPTNTTEANGRADIPRSGRAPPPGXQAARRR APTTEQRRIVGLLEAATGREQLEEATTQAM LFLARLTGRRPEPRNAIRPGXRQRHPAQGDV QAQAPDRIXEAKKLQRLYRTSKKRAVQKIL AGPXMNCQIDKNTITAHFVELAARRDGGED

					GTTTTAGGGGAAGTTA CTTAACCTAAAAATAC AACTTTTCC (SEQ ID NO: 1310)	WPDVFDREEPTAASGEALCTPITREEVFRRL KGRNNTSPGPDGITYRDLAKAXPGAHLVLA LYNXIWRIEATPALWGVSNNTLIYKKGDAM DISNWRPISLGDTVPKLFAAILADRIKRWAV ANGRYSASQKGFLEFEGCYEHNFVLQEAI RE AKGGRKELVVAWLDLASAFTSVPHSSILQA LEGHGLPSKARNIISSLYTGMTTRFHTAEGP TDPILIQSGVRQGCPLSPDVFNLTLEVVLREI QRTGEGYTIEGRRISHLAYADDVAILADSPA GMRLLFAAERGARA VGLTFNPAKCATLHI AGRGEEAVRPTEFSVQGTPVRALASGEAYE HLGIPTGYQVRQTPINTLRDLLADIGSIDRSL LAXWQKLDAVGTFLLPRLDFTMQGAHIDK GFLTEADKIIKKA A KSWLSLPQRASAELVFL PPSQGGGGLLTV AHSYKMLYSSDVTVSTIA GSTLRRTVSERLKKRASNIDIARFLSGDLDP RSTSPSTFWTKVRS AALRIKTKLGLRWSWC QGGEVLLMACGDPRAPGTRVSPQTKHLVTT SLRRCLNRHYAESLLAKKDQGKVFEVTRRS GQSNHFLRSGSFTRFCDWRFIHRARLDVLP L NAAKRWQRGMDKRCRRCGSDLETLP HVLS HCGPHSAARQKRHNNIQDRLVKAASRCPGT ISVNQTVVGVRGPDAALRPDIVVRDDVNRR
--	--	--	--	--	--	---

						VTIVDVAVPFENRLEAFDGVREAKIAKYTPL ARQLTDSGYTVTVEAFVVGALGAWDPRNE RVLSLLSISRYYAILMRRLMVSDTIRWSRDI YVEHVSGIRQYRE (SEQ ID NO: 1432)
NeS L	Utopia- 1_PCa	.	Phytopht hora capsici	GCCCGCCGGTGGAGTAGCCATGTTGG CCACCACCGCCCAAGTCTCCGCCGCA GCTGCGACTGCTGCTGCTCATCGAGT CGCAGTAGTCGCA GCTGCACAAGCAC CACCTCCTGCCGGACGCGCCGCCGTG GAGCACCACGCGCGCGCTGAGCCGTA CCAAGACCAAGGNTTCCAGGCTCGC GCGCGCGTGGMGCTCCCAGCCAGCG ACGCGTTCAGCAGCAGGGTCGGCTGC GGCAGCTCGACGC	TAGACGGCACAGTTCTGGCCCACGTAGGCCGA AAGGGCCCCACCCATGTAGGGAACCGCCCTCG GGAAATCCATTCCGGTGTTGCGACTGAAGAGAT GCTCCTTCGCCTTGACGGAGGTACATCTCGACA GTCGAACTTCAACTCGCAACATATCCGATACAG TTACAAACCACAGTTAGATATCAGATAGGAAC CTTCCTTTAGGAAGCTAACGGGTACACTGGATG GTAAATACACATACATTTT (SEQ ID NO: 1311)	MVVSRTITARLEATPAPRWDPPLPRRVIASRI ADRLVPATAPCRSALNAAFPSPSRDTVTESF TQEDRQLEPLTRHVDEETKDSELPGRAPT VLDDEESKDNDATAGEWLLRFDGACRANPGPG GAGAVLFNPSGAASWTCSHFMPGATETNNT AEYTAFLLGARAAADHGATLLRVQGDSQL VLRQVKGIYGAKSTRLLRLRDAVRAELARV GQFSLHHIDRQDNAHADRLANRALDMKSTL VECATHPGRNACTTTLTTSAAAESPASPPPV GARDTPMADAGEERLADVDDGEVYAAMR LGPGEVPERRPRLRLRQLSDEELEAASEMVE RLGAALSAKITDAEDWASAEGYITALPYML YDKLQSYSQAPRGPQQPVLTRSPRGDDRPA SSEPNASSTTGGVASEHQPRRRRRRGRKRG RRQRRNPRRSGREGATGGHQHKKHKPRPP RETQHHREHRLDEALDELHALERTDPHNRP AIAKARRRVGRIRSAINQQLLRHKFDTDEKA CVDGILSTARAERAARAATPSPPASGAPTTT VSAPGAIVTNDGTCPIPSDKLWRHFDAVN

				TGCACGGGGACGG CAGCGACGGCAG GCACCAAGACCAC GACGACCAAGCCG TCGCTGCCGTCCC CATGGACGTCGAC CAAGGTGCTCGGT GGCCGACAGCGG ATCAMCACCCGCT GTCGCCGCCACCG GAGCTCCGAGTCG GCGGCAAGCGCCG CCGCCTGAACGAC GGCGACGACGAA GACATCCGCGAGC TGGCCGAGCTTCT GCTGCCCCGACGAG GAGGAGGCCGAC GACCACAAACCAG CGCCCAGGTTACC CGCGACCAGCGCT CATCCGGCCTCCG TCCTCGCTGTGTA		TPRLDFDAEAPGSAAFRAAMDHLPAATRLL DLLKEAPSTDEIETQLQHVKASSSPGLDGVG YDVYKRFTIQLLPVLRAAFRC CWLYKKVPQ SWKLGVVRL LHKKGPREDPANWRPICLQQ AIYKLYTGILARRLTRWMDANDRHAPGQK GFRAVNGCGEHNFLAATLIDNARRKHRPLY EVWYDFRNAFGSVPFALLWDSLQRLGVPPD YVDMCKGLYNQASFVVGNAVDGSTAPVEQ RVGVFQGCPLSPQLFNAAISPLLYALRRLPD TGVQLSSVDRPGASAYADDLKIFSGTKAGIT QQHEL VATFLRWTGMQANPAKCRSMGVRR NTNGAVEADNVHLELDDTPIPSMTHMQSYT YLGIGDGFHDVRRRVELAPKLKTLKHDVTA LVESGLAPWQVVKAVKVYLYPRVEYALRH LRPDDQHLESFDLHLRRGLRHLLRLPKNAT NEFFYAPVSRGGLGLLPLVELHAALQIAHG WQTLHSPDPAIRRVAREQLYQIADARHRLD KDHWP HRREELCELLNGELG TSAHAPPKR RNGDIGSLWVDVRKNLKT FGLKVATAPAN QETGVPAQPLQLRVPHHAEWLDHGNVLRH VKLHIKNLHWQ TWCALSDQGKTARVHGGV GSAFLTRPRGMWESDYRFAVAARLN VVDT VNTLSRRRLRAHDCRYPACRWKETLAHV
--	--	--	--	--	--	---

				CGCGCACAAACGCG CAGCGCTTCAACT GCACGTTGTGCGT GTACACGGCTGCC AGCTTCGCTGCTC TTACGCGACACAG GGACTCTCGGCAC CGGCGCGTGACCT TCCTGGACAGGT CTCGGCGGGTTGC GCGTGCGGCAAAC CTTTTGCCTCGAG GCTGGCCGCAGCA AGACACGCACAA ACGTGCGCCAGCC TCAGCACCACT GGTCGCGGTTTCG ACGACAGGAGGA GCATCAAGCCACA CTGTCGTCGGAGC CAACACCACCGTC GCCACGGCGGTCA CCGCCGAACCCCC	LNHCPGTMDAVRGRHDDALKEIEHTLRASS GDRRELRVNQTVPLPGPPLRPDIQVYNHD KRTVAVVDLAVAFDEQPSDPESSGLAKAV QIKKAKYAGIKEHLENQGWKVHLSAIVYGS LGSVAASNHKVYTEHLGLLKRDAKRLDRQ LSSACIQSSRRIWNFHC AKHRARQHEHQAPP SQATRGRRTTETGGNPSRTDRR (SEQ ID NO: 1433)
--	--	--	--	---	--

				CCTGCTCCACCAT CAAGCCTCGGAAC TCACTGTGCCCCC CCCCACGTGTGAG TTCCCCCGACGTC GATGTGCAGCTGC ACAGTCCGCCACA GGAAGATCAGCAC GAGGACGCCACCC ANCACCCGGAGA GCACGCAACACCA ACCTCCTGAGGCA ACCCGCTGGGGTT CGCCGCTCGCGCC CACGCTCGTTGCC TCCAGGATTGCTC AGCGACTCGGCGA GCTGGAACCTCCA CGCTGGGGCCCAC CATTACCCCGCGC G (SEQ ID NO: 1188)		
--	--	--	--	--	--	--

NeS L	Utopia- 1_PI	.	Phytopht hora infestans	AGCTCGGCCTCGC GGCTGCCTTCCCA GGCGCCGCCGACT TCGCGCTCTGGCG CGGCCCACACGCC GCCGCCGAGCCTC CAAGCGCGCCCGT TGGCTTTCGCAGA CGCAGGGGCTCGCG GCGACGGCCCTCG CCAGCCCCCAAGA CCCCCCTACGAT GTGGCACCACCCG GCAGGGCGGCCG GCAGGCTGCCCCGA CTCGGTATCGCCG GGTGCTACACTCT CAGCCGCCACAGC TCGGGCCTTGGCG GTCCGCCATTGGC CCTTGAGCTCGA CAGCGACAGCAGC GACGACGAGGAC	TAAGCTGGTCATCATG ACCGACAGGGCACTAC CCAGGTAGGGAACCGC CCTTAAAAAACCCAGG AAGACACAAACACCCT CCACTTAGTGACATAC ATATTTTAGCCTAGATT TCAGTTACGGAGAGGT TACTAACTGGTAAATA CGAACACATATTCTGTT CTAATCAGTGTGAAAA CTGGTTTTTCGCCTTTTG GCGGACTTTTTCACTCG CATTTTTGGGCAATCGT CTGCGGCTAGCTTGCTA GCGGCGGACGAGCGGT CTCCGGGGGCGTTCAC CTTTCCCCCGCGAGGCC AACTACACCGATCTTCT CTACACTTTTCTAATTC GCCTCCGTCTTCGGTCT TCGGTTGTGGGCTTTT TTCTTTTTGACCAATCA	MLADPAALAAGLARAPPPPSAPQDPSPAFPA GPAGQNPRAAAPARVEVHTVVAPPGRAGG MLPDPGLVEEPIQATYAHDAAQFECALCPY VAESMAVLVQHRRSAHRGTRFKDIFTSGCQ CSLVFYARIVAASHAVACARRNQRAVPPAP TPVAPTRPEATPQPTGYLAAAMTAAAAAAS SDTVVAAATNMQSAVPAAAKTTGLQLVPPE LEPALPQRASCHAGKRRRLNADEAVTPCTP TARVSPQTEVAMAPHDAPQDDTVLQREAA EPQPDPAATPGAQVQRVEDTTAAQDDTVQ QDHDADTAQVSPPRRTPTRWGPRPSSTQEPS PMTGEPAATLAARRPLTPAATGTRATRWGP CHRAIGAAAIARLVTGLSTEPAQPQRRQPPP PQEPPSQPEPLAAAATAAADIAATVAADIAA AAANAAMDVDGGPAADETWLLRFDGACR RNPGPGGAGAALFAPSGAVVWTCSHYMPS RSETNNTAEYTALLLGVSQSAVHHGASHLEV EGDSSLVIAQVKGTFACRNARLRQLRNRVR HALRSVDTHKLRHIDRQANAHADRLANRA LDQRRTSSECGTHGSCMDSCLAVPTALAAQ ETPPAAPPSTSATPAEGNAMDDIAAEIAARD EGETFPVLPIGPSAPERQPRLRLRQLSDEER DAAADALQELADTMASKIEDADSWTSSEG
----------	-----------------	---	-------------------------------	---	--	---

				GCTCAAGACCCCC ACGCCGCCGCCCC AGAACCCCCAGAA GACGTCGCGAGTG TGCTTGCCCCACC CGGCAGGGCAGG CAGC (SEQ ID NO: 1189)	GAGCGCGCCATGCGCC TCTTCTGGCCAATCAGA GACCGGGCCCTGTCCTC GGACAGCGAGGCCTCC ACGGCCAGCCAATCGA GTCTCGGCAGCGACGC GTCTTTCTATAGCGCAG CTGACGAGGCCGATCT GGCGGCCCCCGATTGG TCCGACTTTCGGCCAAT CAGCGACGACGAGGGG GCAGGGGTTTACACTTT TGCCCCCGTTTCGACTT CAACTTCAGGCCAAAA TGGCGATTTGGACCCTC CACGCGCCGTGCCACT GCTCGGCACCGGCGGC GATTCAGCGGGTGCAA CTTCGGGCACGTGTGC AACACATGCAGCGCCC ATTGCACGCCAAGCGG CATCGCGGGACGACGC CTCGGCCGCCCAAGCG	YISSIPERIREVLQPYATAPPQPGRSRRQQR RPPRVTRNQREHRLDEALDDMAATQQATPR DQRAVRRARRRVGRVRASMAQQELRHEFA KDESKCVAKILKTASTETAAEDEHPETCPID AATLHAHFTGVNAPRTDFDYDATSGREFRA AMSDLPPATVEIDAFDAELTIDEVEDQLTRA AKTSSPGHDGIDYGIYSRFAAQLVPLLHAVF QFCWRHRRVPRLWKVGIVRLIHKKGDP RPQ TNWRPICLQPTIYKLYSGLLAHRLSRWLEGN DRLPMAQKGFRANGCHEHNFMATTL LDQ TRRQHRKLYQVWYDLRNAFGSLPQQLMWR VLRHLGVDSGFIDRCRDIYRDSAFVVANAA DGATDPVRQEVGVYQGCPLSPLLFVAALVP LVRRLKLDGVGVPLADGVRPCTTAYADDL KVFSDSAAGIRKCHDTVAGFLAWTGLRANP GKCASLAVTTNARGNPTRDSSMRLEVHDA AITTL SLHESYRYLGVDGDYDHVRHRLQLE PKLKQLKREAVALLTSG LAPWQVVRA LKV YVYPKVEYALRHLRPLQSQLQAFDRV VSKG LRHLLSLPRSATSEVLYAPTSSGGLGLQPLV ELHRA LQLAHAWQMLH SKDPAIQA VARAQ ACQVVRKRYRLQEDHWRGRDDEL VRSFLN SELAASPHAEVLRNRNGDIASLWSDVQRWLR
--	--	--	--	--	---	--

					CAGCCCCGCCCTTCCAG CACGACCTCGCGCCGTT TGGCGGATCGCCATCA AGACGTGCGAGAGCCA GGCGGGGTCGGGCAA ATATACTTACTCTAAGT ATGCCC GAATCCCTGCC CTCTCAGGCTGAACGC GGCCCCATACTTGATCT AAGTATGGGAGGATCC CTGGCCTCTCAGGCTGT ACGCGAGACCC (SEQ ID NO: 1312)	IYHLRFEHCDETEAHGPLSFRVPHHNKWL HKTVLRHVKLHLKIRHQTRWKG MVDQGKT VRVHGGVGAKFMTTGAGLSDDDYRFGVK RLNQVD TNSVLKRKRLRAHTTCRDPTCSSA ETLAHV LNHCESNMDAIRQRHDDALEQIGS KIRGALDRAKSPT ELRLNQTVPEYTGAALRP DIVLRNVAAKTMVIADLAVTFEDQAARARH SSLQLSHDHKTLKYQPIVAELQHKGWRVQT AAIVYGT LGSVQPSNFKAYTEKFKLHKREA RQLDLQLSSH CIRASHRIWGW HCRQHRDRQ RSGTASRASRGSGGAPRRTSQAPARR (SEQ ID NO: 1434)
NeS L	Utopia- 1_PMi		Patiria miniata	CTGATGTGGATAC CTTGGAATTACTC AACCGTGTCGGAG TCTTTTGTCTTTTG CGCCCAACACCTC ATGGATAACCATGC TTGTCGCTGGAGC GACGTTACAAGCG TGAGGGCGCCCTC CATGCCGGACAGC	GATTAGCGAACACTAA TATCCTGCCMTAGACG TGATTGCTAATCCGCAA ACCAACCGGATCTACA GCCTGAACACTGAACT TTAATCTTCACCCATGT CACATCTGGACACTAG GTTTTTGCTCTGTTTGT GTTTTCCTGCCTTTWCM TTGGAWCTTTCGCMCT	MCLKSFSSTSGLRRHMARLHRQSPDASTPS TMTEVFPYPLPKVWPCVVC RENFYHNQTLK RHQKNFHPNVDLTTVYQCSVCGQEFVTGRK ASFHFKVHRRMSASAIPTPSAMPSSPMDLIR GLVGEPLPPSPARTPPPLARYISPAPRSSFSPP WNPSPPPRSPTPLPRPLTPPPRSPPPPRSPTPP PPVTLTAPVTEPAVPVALTTAQVTEPSAPA VHTAAPVTL SNAPVTEPATPATDPATPVTRL HSPVTHISCSISFTASHAPYSCAAPTSPSVYA CSPRRRQCSSTIAAVCNSEASSGNPCLLALP

				TGGTCTGTGCC (SEQ ID NO: 1190)	GGAATTTATTTGTCGCT TGGATTATTTTTTTTCT CACAATTTGGATCTATT TTCGTTTGTTCACTTCG AACTCTAGCTGCCCTTT CTTCGGACACTGAACTT TAATCTTCGCCATGGCT GTCAGTCGCCGGTTCAC TTGCTGCGGTGGGATCC TGTTGTGATAATCCCCG TGCATTGCCCATGGATT TATTTCCGCCTTAGTTG TTCCTAACCTTGGATTT ATTGCTGTGGGTGATGC CCGGGTTTTGTTTACAT CGGGATCCCGCTGCGG CTCGGTGTTCCATGCGA CTGGCAGCCCCTTTGTT TACTCTSGACTCTATTC ATTGTTGTATTTCTTCC ACACTGGCCAGTGATC ACACTGCTGTGTTTCCC GGGAAGATATCCTCTG	VHRHHL PDTSPQRPGLLFHHPGIPPHRPGRPP HCHGHS LHRHDHRARRPGRQHHLPRSPSP PRSPSPSRSPSPPPRSPSPPPRSTPPPRSPSPT PRSPSPPPRPILPPRSPDSTHRS LTSHARSPSR PATPIPVQPRHLRPSTPAHPGVDNAVPPSQ QSAIDVWLAELSR SADFESFEDVCDRFVEFA AAEGRNNGRPARPAHQPPRDRGNQGPRPQR PPRPHRPGLGPEFDAQEASRLQKLYRTSKKR AIR TIL TGSDVRYSGYRGPHPALMSDTAATE VLTSGFLSLEVFSAREVD TDTIATDTSLLFPN SAQARESGQDLLRPVTQREVSLRLGRMSNS APGKDRLEYRHIRQVDGA FRVTLEIFNRCLR ESRVPSSWKTATTVLHKKGDATDPANFRPI ALQSCLYKLLMAILSDRVTTWALDNDLISS QKSARPGEGCYEHTFLLSTVVKDARRNQKN MYAAWLDLQNAFGSIPHDAMFTVLTSIGAP EGLVSLVRDVYTDASTDFVTPTGR TAAVPIH SGVKQGCPISPVLFNL TLELIIRAVNASATRD RSAPVVHGQAVPILAYADDLVILSRSSDGLQ SLLTTASIMATKIQLKFKPAKCASLSLECR GTKVRPLEFNVQDKIIPALTEE QHYRYLGVP IGLYRTDDSLETLVAKMTDDIQRIDSSLLAP WQKLDAIRTFVQPCLAYTLRAGDCAKKHL
--	--	--	--	----------------------------------	--	---

					CGGTTTTACGCTCTGG GTGTCTCCCCGGGCAA CGCACTGGTTGCTTGCT GCGCCATCACCTTTTC GTTTATATTCATTTTCA GTCTGCCGTTATCTTGG CCAGCGCTCATTCTTTT GTGATGGCCGTGGACT GACCCTCTGCGGTTTTC TGCGGTCASCACTCTCG GTGAATGCTGTGCCAC CATTTTTTCATTTGTTTA CTTTTTCAGCTACAATT ATCCTGGCCAGCTTTCA CTCTTTTGTGATGGCCG TGGACGGACCCTCTGCT GGTTTTACGCTCCGGG TTTGTCTGCGGTCAGCA CCCTTGGTGAGTGCTGG TCGTTTGTGCTCATT TTGCTTAGTTCACCATT ATCTTTMTTCTCTTTTG TWTGGTTTTCTAGCG	KRLRGQLVKTARKVCNLPTRATTNYIFADR RAGGLGFIDPNVDADIQIITQAVRMLSSPDDI TRAIATGQLSSVVHRTIHRAPTQEETDEFLSA SMEGDFANSGNSGQASSLWSRARAAARRL KVTISGSLSGSVITKSTENREMAAKSITTALR AQSR AHYTHRLLSLPDQGKVGQSLNQDQY MNSSSWMSSGSYILFCDWRFIHRARLNTLPT NATAQRWKPNTSPACRRCQHPQETLPHILN HCPPNMVPIRRRHNLVQQRIVSAVRHGRVF VDQHVPEDPNPRERPDITVVEGDKVTIIDVC CPFDNGRDALMTAAAKETKYADLKQALV AAGKDVEVFGFAVGS LGSWLPSNERALRRL GIAKRFR TLMRKLLRIDAIKGS RDVYIEHMC GHRQYT (SEQ ID NO: 1435)
--	--	--	--	--	--	---

					GTTGTCTGGGAGTTGA GCTGCAGTTGTCTGGTC TTGGTTTTACCCCCATT TGTTTTCTTTTAACGCG GGGCGTATTGCCTTGAC CGGCCGTCTCAGCTTTT CTCCTAGAGCAACCTTC CGTTCATCCAACTTTTA GTTTTCTCAGTTCTTGG CCATTCCGGTTGGTTAA TTTTTATTTATACTTAA TTTTATGTTTACATTTT CTGGTTGGAGACCATT TAGCTTGTTTTAATAGC TTTTCTTCTTTAATTAA TACCCTCTGCCATTGAG GGTTTTTATTACTATTA ATTTTGTTTACTCTTTG TAACTTGTTTGATTGAA TATTTTAATAAACCAC (SEQ ID NO: 1313)	
--	--	--	--	--	---	--

NeS L	Utopia- 1_PS	.	Phytopht hora sojae	AGACGAGCAACG CGCTGGGGCCCAA GACCTGGCACGAA CGACACTGCCCAG GCTGTGAACGACG AGCACGCTGCTAA CCCGGCAGCGCAC CGGCCTCTGGGCT CCGCTGCACCGGT TACCGGTGCAACA CGGTGGGGTCCAC GTCACGGCGCGAT CGGGGCAGCGGCT GTAGCTCGCCTGC TCACAGGCCTACC CACGGCACCAAGCA CCAGCTACGCGCC GGCCTGCTTCGGC TCGGCGCTGCCCA GACCCGCCCCGCTC CCCCCGCGGCCAC GACGACAGCCCCC	TGAAGCTGCACAAGCG CGAGGCTCGACAGCTG GACCTTCAGCTGTCGA GCCACTGCATCCGCGC CAGCCACCGCACCTGG GGCTGGTACTGCCGGC GCCACCGCGAAGGACA ACGGAGCGGCAACGCC TCGCGAGCGCCGCGTG GGTCTGGGGGGGGCCC GCGGCGCACATCGCAG GCTCGGGCACGGCGGT AAGCTGGTCATTTGACC AACAGGGCACTACCCA GGTAGGGAACCGCCCT TCAAAAACCCAGGAAG ACACAAACACCCTCCC TTAGTGACATGCATAT TTAGCCTACATTTAG TTACGGAGAGGTTACT AACTGGTAAACACGAA CACACAT (SEQ ID NO: 1314)	MDVDGGPAMPEPWVLRFDGACRRNPGPGG AGAALFKPCGTVAWTC SHYMPNSSETNNT AEY TALLLVQSAVHHGASHLEIEGDSHLV VAQVKGT FACRNPRLRQLRNRVRHALRAV TSLTLKHIDRKANAHADRLANRALDLKRSL AECGEHQGAMESCLHMNPAAQRQREQPAP PARPACAPTRAESASDHDEDIDAEIAARDGG EAFPTLPIPGTAPARQPRLRLRQLTEDEQE AAASALQAMAEELACKIEDADSWTSGDGYI SAIPSRIRQLLQPFTAAQPHPRPPLQQQRQRP PRVTRTQREHRLDEALDEMAAVQQRERPTS SAVRRARRRVGRIRASMRQQQLRHDFARNE SKCVEDILRAASAETAEEHPETCPIDSGTLH EHFTAVNSPRINFLPDEACGALFREAMADV GTPQERRSALTDELTMD EVEDQLMQAATNS SPGHDGVGYDIYKKFAAQLVPLLHAAFQSC WRHHRVPALWKVGFVRLIHKKGDPNDPAN WRPICLQTAIYKLYSGLLARRLSAYLEANGL LLMAQKGFRA YNGCHEHNFVATLLDQTR RMRRRLYQVWYDLRNAFGSVHQDMLWYV LRLLGVERAFVERCDDIYEDSYFVVGNAAD GATEPVRQEVGVYQGCPLSPLL FIAALVPLL RALEKLDGVGVALADGVRPCTTAYADDLK
----------	-----------------	---	---------------------------	---	--	---

				GATGCG (SEQ ID NO: 1191)		VFSDSAAGITRCHAVVEKFLEWTVLQANPG KCAFLAVTRNARGNPAHDKDMKLSLHDEE VSSIKLHDSYRYLGVGDGFDHVRHRLQLEP KLQQIKREAVALMQSGLAPWQVVKALKTY VYPKVEYALRHLRPLQSQLQGFDREVVAKGL RHLLRLPRSATNEVLYAPTSSGGLGLQPLVE MHRALQIAHAWQMLHSDKPAIREVARAQV WQVARKRHRLREEHWRERDDELVRCFLNS ELAASPHAEALRRHGDIGSLWSDVQRWLRI YHLSLVVQDDRNGLDPLGLRVPHHAKWLD HKSVLRHVKLHLKIRHQTRWKGLADQGKT VRAHGGVGAKFMSTWAGLSDDDYRFGVK ARLNQIDTNAVLRKRRLRSHKTCRDPTCSS AETLAHVLNHCESNMDAIRQRHDDALEQIG SKIRNALKRGKSTAE LRLNQTVP EYTG AAL RPDIVLRIVA AKKMVIADLAVTFEEHAAGA RHSSLQLSHDHKTLKYQPIVAELQLKGWQV QTAAIVYGSLGSVQPSNSTPTRKS (SEQ ID NO: 1436)
NeS L	Utopia- 1_PU	ADO S0100 1321	Pythium ultimum	GCGGTGTACGCGC ACAACGCCGCGCT CTTCGAGTGACG TTGTGCGCGCACA	TGATGCGGGTCATATTG ACCGAAAGGGCACCAT CCACGTAGGACACCGC CCTCAAAAACCCAGTT	MGTQSARERGAPSAPHSHTLGPRTPPRPPAC SKHGELESAAGGRDGGCSDGAERERDAERD IRANERDCNGDGDGDDADSDSDDRNDARR RSRRPRATATTTTSAPTTTTTTTTTSATTSATT

				CCGCGCGGGATCT CGCCGCGCTCCAG CAGCATCGGCGCT CCGCGCACCGCAG CGTCCGCTTTGTG GATCACTTCCACA GCGGATGCGCGTG CGGCGTGAGCTTC CACTCGCGTGCGG CGGCAACCAAGCA CGCGCGCGAATGT CCAGAGAGCGCGT TCTCGGTCGCCGC CGCCGCGCGCACT GCAGCGGCCAACA CCGCAGGTATGTC TCTCGGCGCCGAC GAACGCGACCACC TCGCGTCCGTCGG CGCCTTGCAATGAT GTTGCATCCCCCG CTTTTGGCAACAT TTTGCCGGTTGCG	CAGTTTATTGACACCCT CCACTTAGTGACATGC ATATTCAAACCGATAC ATATTCGTTAGGAGAG GTTACTAACTGGTAATA TATCACCATTTC (SEQ ID NO: 1315)	PATDSSPWVLRFDGACRRNPGPGGAGAALF EPGGAVVWTVSHYLPGETNNTAEYSAMLL GVRSAlHHGATRLRVEGDShLALSQVRGTF ACTNRRLRKLNRVQAALRELGDYRLVHID RQANAHADRLANRALDLRKTkVDCGPHAT TTDACVQPAEILAPTARLSSSSSSSSSSSSSDEP MPGLEEPAADDETDADAEADIAMRDGGEIF PTLQIGPGSAPAQQPRLRLRQLSDDESEAAA RTLEHFANDMASKIADADDWRSGEgyISAIP VRLRELLAPYAVPIRSPPRNASSRPPRPQSRP PRPPRVTRHQREHRLDEALDDLAAAQRSTS TDQRSIRNARRRVGRIRTAQAQSDLRSQFAT NERACVESILRAAKPDGTEPQASAGTCPIDR ATLHAHFAGVNTPRERFDfDDALGADFRAA LDVLPppDQAADAFADELSLGEVEDQLDRV VASSSPGLDGVGYDVfKRfRLQLLPLLHAA YQCCWRHRRVPATWKVGLVRLlHKKGDR AEPNNWRPICLQQAiYKVYSGLLARRLSRW LEANERFTTAQKGfREFNGCHEHNFVASSLL DQTRRLHRKLYAVWYDLRNAFGSMPQPLM WRVLARLGVDtAFLQRCEdiYADsFFVVGn AADGATDPVRQEVGVYQGCPLSPLLfISALI PLLRALQRLPGVGvPLADGVRPCTTAYADD
--	--	--	--	--	---	--

			TTTGCGACCGCGG CAGACGCATCAAG CGCCACCGTGATC GCAGACGCAGCCA TGCAGCACAGTGC TGTGCCCTCTGCT GCCGCCCAATCCC CTCGGCGTGCGCA CGTCCCCCCCCGTG CCGCGCGCCACCA CGACACCATCCGC GCTGCGGATTGGT GGCAAACGCCGCC GCCTGAACGACGA CGGCGACAACGA AAACAGCGACGG CCGCGACGCCGAC ATCGAGATGCGCG CCGACGACACCGA CGCACCAGCGCCG ACCAACCCCGCGA CCAGTGCGGCTGC AACGCCCCGCGCGC		LKVFSDSAAGIQQCHGTVARFLRWTGLRAN ASKCALLPVTTTARGNPAIDDTLQLELHGD AIARLTQSSYAYLGVGDGFDHVQHRVQLA PKLAELKRDAVALLRSLAPWQVLKAIKVY LYPRIEYALRHLRPLQSQLEGFDRAVAKGFR HLLRLPANATNELLYAPVSSGGLGLLPLVEL HKALQIAHWQMLHSKDAAVQAIARAQVR QVVQKRYTLADHWQGRDDEL VQLFLNSE LAASPHATIKRRNGDIGSLWSDVQRHLKTL QLRLETREPTADAPDSPNGLLHLRVPHHRK WLSHKTVLRHMKLHIRLCHKHKWQSMSDQ GRTVRAHGQAGSHFVSRGVGLWDADYRFA LQARLNQLDTNSTLKRRRQRTNATCRAPNC SRTETLAHVLNHCETNMDVIRQRHDALEQ IGAAINAAIKGRRTDTEVRLNQTVPEFNGPA WRPDIQVRDARSKTMVIADLAITFEDQPND QSASSSLQHSREHKIAKYQPIAAALERQGWR VHTSAIVYGSLRSVHPSNFTVYTELLGLLKR DARRLNTTLSCHCIRSSRRVWNWHCGQHR ARQHQRCEGRAHGSGGNQRAEGGTATT (SEQ ID NO: 1437)
--	--	--	--	--	---

				ACAGCACCAGCAC CAACGGATGCCGC GACCGCGCGCCGC CGCCACACGCG (SEQ ID NO: 1192)		
NeS L	Utopia- 1_SM		Strigami a maritima	GGAGTGTTCTTTT CGGAGACGCCGCC TACTTTAGAGGAG AGAATCCCCACGG GCATCCTCATTTG ATCTGATCCATCG AGTATCTGCGAAT AGTCGGCGCACTC CTTTTGCCATGAT CCCGGGGGTCTCA TGGTAAAAAGGTT TGTGGCACGGCTT AGTTGACGCCCT CTCCACGTCAC CGGCCTGCATCGA TCGACTGCTCTCT CCTCTTCCTCCCTC CCTCCTCTTCACG	TGATGGGAGAGTGAGG AATCTTCTCCACTGTGC AAAACCATACAGTCAG AAGATGCTAACTACTA GTTTGATAACCCTGTGCC CCCTGCAATGTCCCGCG TGTCGTACCCAAGCCC GGCTGGCATTGAGACA CATTAGGCTCTCGCTCC CCCGTATACTCTCATAA TTTCGTGTACGCTAATC CTACCCTACCCCTCCCT TTGACCACTACCCAAC CATGTGTATAGCTGTGC TGGTGATCCCGGGGCG GTTATTCACTGGTTATC ATATCATTCTAAAATGA TCTTTGATCTCTCAATT	MATVRLKYPYPPEGILCGPCAANTNAPQTR PYSDKSGLAKHLKLYHKATLVVECRHCGHE ESDLRKMKKHISTNHPVAAAAAPTVPRLG PTAPPPPRVILRPRFIPRPTSPSSSSSSASSP ASSRRSVSLPPASPPVSSASSPAARSGRNSPD SQGTAPVTPIGTVRNSPAGSPALSYSTASPIA STITTPRHLSPASPALSAGPGSLGASPPVSPT AATVPPAPPATVPAVMAATVPFVAATTVP VGSSTVPQRPAGPRRPPFPIDDWIGRIARVS SLPELDAVSRLLEDEVVKRRPPDPNARPASL HPTRRPPPPSTRPRPCHGTGGTSVSLAALTSS CIREDHRGSLPLLCTSGWDSPWPLSPSPSSHC PSSNPCPSSSTPPSSLLGPPRHSHLTWRGSGS TTPSHRHCSRARYHHLLWRLPYPSSPRLPYP PSALARYPSVPPALDDPLPSLSTIGSEGSPEC HLCRNWTPSQGCSRMKWSRDAPLTPTPDPP HSIRHAALPLHLLALVPAMGPAELQSLWRR SRPRAFAKITEGASPCALPVGTVHGHFLQV

			CTCTCTTCACGTG ACCCCTTCCCATC CCGCCCCTCGGCT TTTGGCAAAGATC TGTGTGTCCTCCA AAGCACCCATCTA CCATTTGCTCGAG TTGCGATTGGTCG AAGCTGCCACGCC ACTCGTCTACTCT GCCTCTCTGACCC CTTCCCCCTCTCT CCCTCTGCCGCCC TCGCCTCTCGGGT CCCTTCCCATCTC CCTCCTCCCATCC ACGTGCTCCCTCC TCTCGTCACGTGA TCGTTGCGGCTCG AC (SEQ ID NO: 1193)	AACAACTAACTTATTC TGTTTGTTTCATTGTTT CACCTCGTAAGAGGAA GTTCAATTGTGCGATAAA TCAA (SEQ ID NO: 1316)	HQATAHLPTVPVLPPLRPPRSSDPLVTPISPG EVLDRLLRRATDTAPGPDTHIYSEWRAIDPTG RLLSSLFQKVQTFGAPTRWKESTTTLIHKGG DHTAMSSWRPIALLSTVAKIYGSILSHRLTT WAVQNGRLSLSQKGFLPFRGCLDQNYLVQS CLQDARRNKKTLSLAFLDLKNAFGSIPHLTI RHSLEWLGLAPSSIDILEASFLGSSTRVRTET GLTPPISLDTGVVQGAPLSPILFNLAIEPLLRT VPSAHSGFSLHGHVVSVVAYADDLAILAPST PALQSQLDAISGMADWAGLSFNPAKCATVT LTGKDNSRDTLSSLQGSPVPSISDGDAYKHLG VPTGTTTFPSGTD AIKKMTTDLQAIDHSDLA PAQKLDALRTFIMPQLSFHLSHGSPKAPLT QLDKKIKRAAKHWLFLPQRASNEILYMSHL HGGQSLLPLSVLADIGQVTHAVALLQSRDP AVADLALRTCREVASKRAKKTVNGPELAQ YLSGSTDGIYCTPTSDIPSLWTTARAATRRLS STLPLTWTSPGPSVPFLSINGSPLSPFRVQST LTNAIRHNHLSTLIAKRDQGNSYRTSHDPDP SNYWVKGGDFLRFCDWRFIHRARLNLLPVN GARRWDANSIKTCRRCGAPNETLAHVNLVC PVGLPEMKKRHDAIHARIKKALRPSPTVV HHDRTVPGCGPLRPDILRISERDKSVAIVDIH
--	--	--	--	---	--

						VPFDNGTDAVERAHETKRAKYELIRRHYE QGYRVTFDSL VVTALGRLWRGSEAAALQAL QISSQYSKLLRKLLVADAIHGSRNVYAHM TGMVM (SEQ ID NO: 1438)
NeS L	Utopia- 1_SP	AAGJ 02140 537	Strongyl ocentrotu s purpurat us	AAGGCTCAAACCA GGCTGCCAACCAA GCTGCCAGCTTAG GCACCAACCAAGC TGCCAGTCCAGGC TCCAACCAGGCTG CCCACCAATCTGC CAGCTTAGGCTCC AACCAAGCTGCCC ACCTAGCTGCCAG CTTAGGCACCAAC CAAGCTGCCAGCC GAGGATCCAACCA GGCTGCCACCTA GCTGCCAGCTTAG GCTCCAACCAAGC CGCCAGCCCAGGC TCCAACCAGGTTG CCACCCGAAAGTC	TGACGAAATGTTCAAT ATATGACATTCTAATGT TCATGTATTTTTTGT GCTTGACAAAGCTAGA TGAATATATTCCTGTC CTACTACATACATCTCG ATGTGCCCTTGCTAGAC GATCCATTGGTAACCAT GATATCAATGGATTAA CACATGATCTGTAAAC AATATATGATTTACACA ATGTATTTATTGTTTCA ATAAATCTGTTCTTTTC ACTTTAATACATGAAA CATGACTGTGCCTTCTC CAACTGGAGACCTACA TAATTTGTTAAAATGAT ATAAATATTTCGAAGA	MSHSITEVFDYPLPSRWKCTVCLENFFNQQT LKRHQARHHQTTSFLYVFRCSACQAEFDSA RKASNHWWQSHKRKPILSQPAVNEIPSSGLDP SPPRSRPPVEVIGSSFPDDVSMLEPSTPSTSL QMDPEVVHPPSRISFSPMHLSTQPASPIQI GVEVSFNSSSSLQMDPEVVQPPSPSISFSPMH LSPTQPASPIQIGVEVSFNSSGSLQMDPEVVQ PPSPSISFSPMHLSTQPVSPIQIGIEVSFTSSSP LQMDPEVVQPPSPSMSYSPMHLSTQPDSP PVDIDVIPAAEVPLPDIEIPPSPDRHPAAEVPL PDIEIPPSPDRHPVAEVPLPDIEIPPSPDRHPQS PPRPVMMEQPVHTPPPADTQQANGPQHWV TVLANATNWEDFGRVCVEFANHAVEAARS RQDAPQVRPAAQRQPRRPTTRPRQPTFDVRE ASRLQKLYKRSKKRAVRHILRDDAPSFSGSN EQLLDYFKEIYAPPEIDENRAQQLAESLFTDL EEAKESAAALMSPISQQEISTRLSRMSNSAP GKDRLEYRHIRQADGACRVTHIMFNRCLQE HRIPSAWKEATTILIHKSGTTDDPANFRPIAL

				TGCCAGTTTAGGC ACCAACCAAGCTG CCAGCCGAGGCTC CAACCAGGCTGCC ACTCGAGGCTCCA ACCAGGCTTTCAC CCGAAGCATTACC CAGGCTGCCCCGTT TAGGCACAAACCA GGCTGCCAGCCGA GGCTCCAACCAAG CAACCAGCCCAGG TTCCAATCAAGCT GCCAGCCGAGGCT CCAAGTAAGCTGC CAGCCGAGGCTCC AACCAGGCTGCCA CTCGAGGCTCCAA CCAGGCTTTCACC CGAAGCATTACCC AGGCTGCCCCGTTT AGGCACAAACCA AGCTGCCAGCCGA	TGAAATTATTATTAATA A (SEQ ID NO: 1317)	QSCLYKLFMGILSDRMTQWACNHNLLSPEQ KSARPCEGCHEHTFLLSSVIKDTKRNQKTAN IAWLDLRNAFGSIPHQAIHAVLTTIGAPVSLV MLLKDTYTGASTSFLSTSGETDPIQIQSGVK QGCPMSAILFNLIELIIRAVKKKATDDGLGL VVHGQRLSIMAYADDLVLMSTPEGLDAIL SVASEQAETLRLAFKPTKCASSLSCRHGTS VLPREYTVQGHLM PALDEEEQYRYLGVPFG LPRFTNLKDLIGKLKGNIE TIASSLLAPWQKL DAIKTFVQPGLSFVLRAADYLKSDLRSLKSA ITTNVKKICQLPLRAANAYIFAAKESGGLAFI DPNVDADIQVITQAVRVLSSDDEVVQTIATS QLKSVVHRTIHAVPTEEDIDNYLSGSNEGLL ANSGNSGQASSLWSRTRSAARRLHLTLRAT TSGTVVVNQADIDHTRDILPASITRGLRLIQ RTTNAEKLKSLPDQGKVARSLSDNPFANGS SWHATGKFIRFCDWRFIHRARLNCLPTNVA TKRWKANANGKNGHQQETLPHVLNHCLPN MVPPIRRRHDNIQQRLVTAIRHGDVFNQHV PGDPNPRERPDITVIEGNKVTVIDISVPFDNG PNACTTAAQAKVEKYSALRQALRDMGRDV EVHGFIVGALGTWHQGNERALGRLGVSRW
--	--	--	--	--	--	---

				GGCTCCAACCAAG CAACCAGCCCAGG TTCCAATCAAGCT GCCAGCCGAGGCT CCAACCAAGCTGC CAGCCGAGGCTCC AACCAGGCTGCCC ACCAAGCTGCCAA CTTAGGCTCCAAC CAAGCTGCCAGCC CAGGCTCCAACCA GGTTGCCACCCGA ATCTCTGCCAGTT TAGGCACCAACCA AGCTGCCAGCCTA GGCTCCAACCAGG CTGCCACTCGAGG CTCCAACCAGGCT TTCACCCGAAGCA TCACCCAGGCTGC CCTTTTAGGCACA AATCAAGCTGCCA GCCGAGGCTCCAA	YRTLMRKLCCIDAIQASRDIWVEHVTGHRQ YE (SEQ ID NO: 1439)
--	--	--	--	---	--

				CCAAGCAACCAGC CCAGGTTCCAATG AAGCTACCAGCCG AGGCTCCAACCAA GCTGTCAGCCGAG GCTCCAACCAGGC TGCCCACCAAGCT GCCAGCTTAGGTA CCAACCAAGCTGC CAGCCCAGGCTCC AACCAGGTTGCCA CCCGAAACTCTGC CAGTTTAGGCACC AACCAAGCTGCCA GCCGAGGCTCCAA CCAGGCTGCCACT CGAGGCTCCAACC AGGCTTTCACCCC AAGCATCAACCAG GCTGCCC GTTTAG GCACAAACCAAGC TGCCAGCCGAGGC TCCAACCAGGCCT		
--	--	--	--	--	--	--

				CCAACCAAGCTAC CAGCCGAAGCTCT GCCAGATTAGGCA CCAACCAAGCTGC CACCGTTAGAGGC CCAGAGCCCACCA ATCCCATAACGTG TGAGATATGTGAA GCTTCCTTCCACA CCTCCGCCGGTCT CCGTCGCCACACG GCCAGGCTTCATC GCACCACTGCTGA ACACACTGACGAC AGC (SEQ ID NO: 1194)		
NeS L	Utopia- 1_TSP	.	Trichinel la spiralis	TTTCTGGTATGAA TCCCAAGCGGATT CGTTACGAAATTT GCATAAGTTTTTG AAAAAATAGGCAT TTGGTCGAGTGCT CGCACCACCATT	TGATCCGTCCCGAACC AACGGAACCACATTGC CGCATGACTTCGATTTC GCTTTTGCTCTTTTTGT ATTTAATTTTGCTATTA ACAATTCAGTTTGTTAA CTGTTTTGTATTCATT	MCSAKTPALKSGRRRGKEVNYEGQIVRVER RRGSRSTSATDLGTRMVTRGRKKLMEASV REAGHHGGESASTVDVDVVESKKITGKTAR RNRRAPSGDGKRRESCGAECGQAVCGNAV ADRSEASSPRTPNVSKSGRDKCGQPTIKAST PSPPKRKPTTSSSPRTPCLSKRGARSKIPSTPD TPSTSGGSGKQRVLVSPLLRTEKLPDLEVLQ

				GTCGCGGGTCGTC CTGATATTGCACT ACATTCAGGAACG GCCTATTCCCTTC GGGGAATTGTGTT TTAGGAATTGGAA TCGGTTTGGTTAC GATCGGTCGAGTG GTTCGTGAGATCG AGTGACAGCCGGG TGGCAGCGACA (SEQ ID NO: 1195)	GAAGATCCAAATAAAA C (SEQ ID NO: 1318)	RTEEQVTVRATFPIAQAVVCPLGCEKPYTAV RPDGQFAHQTLTRHFMRVHNCHSVQWHYR CRNCNTDFLPADHRYPLRVVNTHVRSCVSR WEITRKLGESEDLHGVRCDLCDYVGVSKRA VGLHRRRHANENIMQNTGTAAQIEALSKQV GEIRVAGDYSQFKFGKRVRQYVAPTQRRDG LDEEEVHEAEEEEVPAESRTILGEPSTATAIG AEEISATGPVRADTAAQQMICRIGQWCWVP QDYHSIPAPQCWTDTLMDLMIEQIVLQRYP DGAGVSVMSCSAVSAAIHHEISAEFAAQVM SSHDASLYCIIPVNVRNHWQMIVLDVAERV VHYCYSLREHNTVVLSSLLSLVELSGKHTG CTSWKIETHDGAPVQTNAFDCGPFSCFLKH LLHGIDMNF GDRESAALRTDLKFMIDAVSTP VVPATDKLKKKPDGSATQLTQFQQKFLSAS DNWQSPDVDLQAVYDEVVESIVSGHNEPSS RNTSRLQKKSPGKGGKGQVRRRSAVTRDPA WLKSASAVQKAFNSAPARTVNAILRRPNAC PSFTATQVADHYFNLRPAVTS LAPEVIDILPP PATDHSM LVAELSESEVWEKMQKAPNSAP GADRITIRMVRMADPGAMILTRFYRACLLR KWVPLQWKQSVCKLLYKDGDKERLANWR PIALEPVLQRVLSAVVASRVTNWARANGLIS
--	--	--	--	--	---	--

					LEAQKGFQPADGTSEHNFVMEVAIQEARRT NAQLAISWLDISNAFGTVSHQLLFSLLERYG LDPTFTSFIQNLYKDATIVVKGANGTHVTAR WSVGVRQGDPCSGILFCLFVEPLLRSVLPSL PCEAETTAVNVLGQPITALAYADDIALFAPSI GVMQQQLCKIQGMASAMGFRFNPKKCASL YLNRAVVNAATFTISGEEIPALVHGDTRYL GVAAGLGKPQTPFSLLRENLEAELIFRSKL APWQKMDAYRTYVLPRLTFQLMIAKFNNIK QSAGQYDRAILRLVKRCFQLPVETSTDFIRA PRQCGGLGVPSLRELYATAKVSRALKMLWS PCRVVSSLAASQLQRVASAYFAKRLRDVEA ADLSTFMNAARSTPLDRSGYPTCLWMDAR KQMSYLTKVAGVDCYFLVGEAGTSFFIRNG LGQTVSVLSPLRKNKVMSVLGGAIQTRHLD AWLQCKRQGKTASCIVLDRSSSRFITTGRYT SFAAMRFALPARLDLLPCRARSSMRSYQNC RRCGYDRETLPHILQHCRQFSAPAYQARHD AVQGRLETVMRRRFPNLRVNRALPEIGSNK RPDLVVVDEEKRLVILLDIAIVFENTAAAFV DARTRKWAHYEKEILAYRLRGYSVTYDAIV VGALGTWDPKNDAILKRIGVVSQRYLRMLK VLVVSEMLEHSSRIYRKHLGLRDLLPDTGTK
--	--	--	--	--	---

						RRPVGTTETDPPGGDLRQKKRNTISARASGG KCLERRFTSPVGTPSQRGELQCQPCPGPRRP ALAGIAPNPPSLQPRKPPPRQHQKPVTKSTA H (SEQ ID NO: 1440)
NeS L	Utopia- 2_CMy		Chelonia mydas	GGCAGAAACTGCA CSTTCTAGAAGAC TCACTGCCTATCC TGAGGAAGACTAC CGCTTTGGAGATG GATTCTACTGCCG CTGCTTCAAAGA AAATCTTCATGCT GCTTCGGAGGCTC CAGGACAGASTGA GAAGATGATCGCT CCCTCGCCGATTC CACAGAAACCTTC AACTGCCGCTCGG ACCGCTGCTGCTC CACGGAGCGCCCA TCGGGGAACAGCC TCTAAGAGACCCT CAGCGCTTCCAGG	TGAGCTGGAGTGCCGA TGAGAAGCGCAGTCGG GAAAGTAACTGAAATA CTTTCCTCATGGATTGT ATTTTCTAAATGGACAA CCTACCTAATTCTCAAT TACTGAGGGACAATCT CCACTCATTGATATATT TTGCTTTCCACAACCAA ATCTCTGTACAACCTTT CATGAGTGATGTACCC GAGTACTTGGATTCTAA TATCTAAACTGTATTGT TAAATCTATTCACCTAA ATTTGGGTTATTGCTGA TTATGTACTCTATGTAT CATATGACTTTTAAAAA CAAACCTTTGTATTTGTG GATAATCTAAGCACTA	MLQLRLPTPQTLRLLHPSQLPQSHSTKRWSN IYEERARAPKDTTIERVSAASKIPKLDPAKRR IGAPLQLMQGNSISRQLSASSQYVQHNAV RRVSAPPHTGNFTGSCRRKLALPHRPPSETQ PAESYLHCWTTQRDPTLADQHGLQDHPTGL PPEGSHCIKDPRQNSSTEGQRRSGNSQARRIP KRASTAASKTVLPKR TSAALKSVREDTALV LEDPAKWSSQHREGXRQQANPTAVFQPEPA EIEQQPXVRAATPWQAAWMEELARTASFX DFDLLVDRLTKDLSAEIVSGRKGQTQENTPTA HRQNQNNMREARRRNISRCYDPAAASRIQK LYRSNRPKAMREILDGPSSYCAIPSERLFLYF KGVFDRVAQNDMQRPECLXPXPRVDYAED XEQDFTSWEVEARLTKTKNTAPGKDGIRYN FLKKRDPGCLVLTAIFNKCKQFR RTPSSWKK SMMVLVYKKGKQDNPNTRRPISLCSTMYKL YASCLAARITDWSVNGGAISSIQKGFMSCKG CYEHNFLVQTAIHMARRAWRQCAIAWLDL ANAFGSMPHQHIFDMLREFGMPENFLQLVR

			ACATCGCAGATGA GCAGCATCGATTG GAAAAGCAGCGC CTCCCTGAAGAAA ACCCGTGGAGATG CTGCTGCAGGGA WATCTTGCACCTC GAGGACGGCCTCC CAGGACATCGCAG CCAGGACAAACAT CATCTCGCCTGCT CTTCAGGASAAGG ATGCCMSAAGAAC TTMTCCCACCTCC TCMACTGCCCAGG ATCCTSATGCTGG TCGTCGTCCTGCT GTGCTGGAAGMK ACCCCAACTGGCG AGACCTCAGAAAC CACCCAGMWGGA CCKCAACATCTCA CTAGATGCCTGCC	TACCCAGATGTACAGA CACTCTTTTCCCAACCT ATGTATTATATTTTTT AACATTAGCTTTAATAA AATTTTAA (SEQ ID NO: 1319)	ELYEGCTTTICSMGETPEIPIRSGVKQGCP SPIVFNLAMEPLIRAISGLGGFDLYDNRVNI LAYADDLVLIADNPESLQQMLDITSQAANW MGLRFNARKCASLHIDGSRRDSVQATSFQIQ GEPMIFLEDGQAYQHLGTPTGFRVQQTPED TIAEILRDVARIDSSLLAPWQKINALNTFLIPR ISFVLRGSAMVKVPLNKADNTIRQLVKKWM FLPQRASNELVYISHRQGGANVPRMGDLCD VAVITHAFRLLTCPDAMVRNIAESALQDAV KKRIARTPSNQDVATYLSGSLEGEFGRDGG DFASLWTRARNATTRLEKRIGCHWTWCEE RQELGVLVPQVKNTDHTIITPRARTMLERTL KDAIRCQYVENLKRKPDQGKA FEVTCKWD ASNHFLPGGSFTRFADWRFIHRARLNCVPLN GAVRHGNRDKRCRKCGYANETLPHVLCSC KPHSRAWQLRHNAIQDRLARAIPPPVGKVA VNSAIPGTDSQLRPDIVITNEDRKKIIMVDVT VPFENRTPAFHDARARKVEKYAPLAETLRA KGYQVQTHALIVGALGAWDPSNERVLREC GIGQRYARLMRQLMVSDAIRWSRDIYIEHIT GHRQYQEG (SEQ ID NO: 1441)
--	--	--	---	---	---

				CAGCCGAACCTCT CCATGCWACTCTC CCAGAGCAACSAG AACCATCCAGMG AATCCGCTGATAT GACTGAAGCCMA TCCAACAGAGGGA GAAGGAAAGGAG AATGACTGCATCT ATCTCCAGTATCC CCTCCCTACGGAC ACGCTCCTCTKCC CCTTCCGCTATCC GAGGGTTCCAGTA CATTGGCAGTCTC AGCAMACACCTC AAGAGAATCCATA SCAAGCGGATCAC CTTCCGGTGTGCC CTCTSCGACCTGC CTTTCGAGACGCA GATGAAATGTAAG TCTCATCAAGTCA		
--	--	--	--	---	--	--

				CCTGCAAAGGACA TCTCGAACTGGAA GAGTCCAACCTTA CCAGTCTATGTTG CCGCCACCCCATC TCTGCTCCGAAAG CAGAAACACCAC (SEQ ID NO: 1196)		
NeS L	Utopia- 2_PCa	.	Phytopht hora capsici	ACCGCCCAAGTCT CCACCGCAGCTGC GACTGCTGCTGCT CATCGAGCCGCAG TAGTCGCAGCTGT ACAAGCACCACT CCTACCGGACGCG CCGTCGTGGAGCA CCACGCGCGCGCT GAGCCGCNTCAAG ACCAAGAATTCCA GGCTCGCGCGCGC GTGGCGCTTCAAG ATGGCGACGCGAC CAGCAGGAGTGGC	TAAGCTGGTCATTTGAC CGACAGGGCACTACCC AGGTAGGGAACCGCCC TTCAGAAACCCAGGAA GACACAAACACCCTCC CTTTAGTGACATACATA TTTTAGGCTACATTTCA GTTACGGAGAGGTTAC TAACTGGTAAATAAAA AGCACTTT (SEQ ID NO: 1320)	MQDMEEELLLDVEMETETTEPQTSTAXDAT TTTDRPTRWGPHPRAVAAAAIAQLVTGEXA XPALPSRQDRRPAPRSHAPTRSRWGPRHQA VGAAAIASLATGLPASAAPVSRATKHGEGR RRLQTRWGPRVSIPRAARRPGSRWGPPRAA GASGQLPASASGATGQLPEHVEAITTTPRVA SDADEGPTPPDPWILRFDGACRRNPGPGGA GAALFKPSGAVVWTC SHYMPSSNETNNTAE YTALLLVQSAVHHGATRLDIEGDSSLVIAQ VKGTFACRNAKLRLRNRVRHALRSVEKY TLRHIDRKANAHADRLANRALDRRSSSEC EPHGSCMERCCGTDTPPAVQGPTPQAAAAV PVQVWPQWQRQTMVAWTTSHGGRCRDCS TRCGRSLPSLAHRPRLSPRRQPRLRLRQLSD EERDXAADALQELSDVMASKIVDADSWDT

				CGCTCTACTGACG GCGGAGACGAGG CCAAGACCGAGG ACAGCGACGCGG ACGCAGCCAAGCG CCAAGCCGCTCAA GCAGCACCGAGAC CCCATACCCMGGGA CACCGCGCCGTAC GACGGCCACAGG GACTCGGTGCTGG CTGTGTACGCACA CAACGCACCTGCA TTCACCTGCGCGT TGTGTGTGTACAC AGCACGCAACTTC GCCGAGCTGACCA AGCATCGCCATGC AGCGCATCGTCAC ACCCGCTTCGTGG ATCACTTCCACAG CGGGTGACGTGT GGCATCGGCTTCC		GEGYISSIPERIREVLQPYTTTRPPRPGHQQQQ RRRPPRVTRNQREHRLDEALDDMQATQQA APRDQRAIHRARRRVGRVRASMAKQELRQ AFAKDESKCVSKILAGASAETAEEHVDEC PIDAATLHAHFTGTNAPRTDFDYDAACGQE FRGALDSMQPPTVATDAFEEELTIDEVEDQL TRAAKTSSPGHDGIGYDIYSRFAAQLVPLLH AAYQFCWLHRRVPALWKLGIVRLIHKKGDP MQPTNWRPICLQPAIYKIYSGLLARRLSRW MEQNQRLPMAQKGFRAFNGCHEHNFVATT LLDQTRRSHRRLYQVWYDLRNAFGSLPQQQL MWSVLRHLGVDASFIARCKNIYQDSAFVVA NAVDGATDPVRQEVGVYQGCPLSPLLFI SALVPLIRRLEKLDGVGVPLAEGVRPCATAYAD DIKVFSDSAAGIRKCHDAVTRFLEWTGLRA NPGKCASLAVTTNARGNPVRDDGVHLELQ GEVIAPLSLHDSYRYLGVGDFDHVRHRLQ LEPKLQQIKREAVALMQSGLAGWQVVKAL KTFVYPKVEYALRHLRPLQSQLQGFDRAVV RGLRHLLRLPQSATTEFFYTPTSGGGLGLQS LVEMHQALQVAHAWQMLHSKDAAVVAVA KEQVCQVARKRYRLQEEHWRGRGDELVRL FLNSELAASPFADCLRRNGDIGSLWTDVQRT
--	--	--	--	---	--	---

				AGTCGCGCGCGGC AGCTACGCGACAC GCTCAAGCCTGTG CAGACAGCACACA CGCCACCGTAGCT GCCTCGCGCGACC CGGCMNCCASTGC SSCCSGNGCCGGM AGCGASGAGGAG GACCNCGCACCCC CCGGTCCTCTSSTC GCGGCAGCATCTG CAGCTGCAGCAGC CGCATCAAGCGCC ACTCCAGCTGCAG ATACTGCCACCAC GCAGAGCGCCGTG CCCATCGCTACTC CTGGGCCCCAGTA CGCCCCCACGTG CTGGAGCCACCTC CAGAGCTCCGAGT TTCCGGCAAACGC	LRLHHLSLTAQDDRDGQDPLALRVPHHTK WLDHKTVLRHVKLHMKIRHQTRWKGLVD QGKTVRVHGGGLGAKFVSTGAGLSDDAYRF GVKARLNQVDटनाVLKRKRLRSSKTCRDPT CSSAETLAHALNHCASNMDAIRQRHDDALE QIGSKIRGALERAKSTTELRLNQTVPEYTGA ALRPDIVLRNVAAKKMVIADLAVTFEDHAA GARHSSLQLSHDHKTLKYQPIVAELRVQGW QVQTAAIVYGSLGSVQPSNFKTYTEKLKLH KREARQLDLQLSSHCIQASHRIWGWHCRRH REGQRSGNTSRASRGSGGTPRRTSQVRARR (SEQ ID NO: 1442)
--	--	--	--	--	---

				CGMCGCCTCAACA CGCCGATCGACCT GCAGCCGCTGGAC GTGGACGCGCTG (SEQ ID NO: 1197)		
NeS L	Utopia- 2_PI		Phytopht hora infestans	GCTCGGCCTCGCG GCTGCCTTCCCAG GCGCCGCCGACTT CGCGCTCTGGCGC GGCCCACACGCCG CCGCCGAGCCTCC AAGCGCGCCCGTT GGCTTCCGCAGAC GCAGGGCTCGCGG CGACGGCCCTCGC CAGCCCCCAAGAC CCCCCTACGATG TGGCACCACCCGG CAGGGCGGCCGGC AGGCTGCCCCGACT CGGTATCGCCGGG TGCTACACTCTCA GCCGCCACAGCTC	TAGGCGGAAACCAGGC CCAAGACGGCCGACAG GGCCCCACCCAGGTAG GGAACCGCCCTAGAAA CCCATTTCGGTGGTCGA CTCGAAGGCCTTACCTA TTTTTTCCTTAGACATT CAATTAGGTAGCGACC AAATTACAAATTTGGT AACGAGTAAGCCAAAT GGTAATACACAAAAC TTTCTGTTCTAATCAGT GTGAAAACCTGGTTTTCG CCTTTTGGCGGACTTTT TCACTCGCATTTTTGGG CAATCGTCTGCGGCTA GCTTGCTAGCGGCGGA CGAGCGGTCTCCGGGG	MLADPAALAAGLARAPPPPSAPQDPSPAFPA GPAGQNPRAAAPARVEVHTVVAPPGRAGG MLPDPGLVDSSPAAATAATPAPVAATATTA RAAARVAVEHHAHAEPNQEHLPMARVLVE PMQVDECSSCDRSTLTADDGSGDDVAAPSS MLSNDVAAPMDVDSGTSCPPTLQQPLQRPR ALHVGSKRRRLDADDGEEAHQLQEEEEAGI HAPALRLSAASAQPASVLAVYTHNASRFDC TLCAYTAGSFASLLTHRNSRHRRTAFLDRFS AGCACGVPFASRLAAARHAQACASLSSAPS AEASSAAGTSSPTADGADSTVSAVAHAEPG LPHHNDTELTAAPPLVSSSDVEVQATKTEAT DNRWGAPLPRVLVASRIAGRLAQVPPPRWG PPLPRTTIAARIATRLAATPAPRWDPPLPRSL VVSRIAARLLPALPDAPACEEEAKDSDTMD WAPTWTNEETKESEPHDEAPGQVDEETIDD ADGEWLLRFDGACRANPGPGGAGAAALFKP SGPVVWTCSHYDPSTTATNNTAEYTALLG

				GGGCCTTGCGGT CCGCCATTGGCCC TTGGAGCTCGACA GCGACAGCAGCG ACGACGAGGACG CTCAAGACCCCCA CGCCGCCGCCCA GGACCCCCGCAG ACGTCGCGAGTGT GCTTGCCCCACCC GGCAAGGCAGGC AGC (SEQ ID NO: 1198)	GCGTTCACCTTTCCCCC GCGAGGCCAACTACAC CGATCTTCTCTACACTT TTCTAATTCGCCTCCGT CTTCGGTCTTCGGCTGT CGGATTTTTTTCTTTTT GACCAATCAGAGCGCG CCATGCGACTCTTCTGG CCAATCAGAGACCGGG CCCTGTCCTCGGACAGC GAGGCCTCCACGGCCA GCCAATCAAGTCTCGG CAGCGACGCGTCTTTCT ATAGCGCAGCTGACGA GGCCGATCTGGCGGCC CCCGATTGGTCCGACTT TCGGCCAATCAGCGAC GACGAGGGGGCAGGGG TTTACACTTTTGCCCCC GTTTCGACTTCAACTTC AGGCCAAAATGGCGAT TTCGACCCTCCACGCGC CGTGCCACTGCTCGGC	ARAAADHGVTKLRIEGDSTLVIQQVRGIFAT RSTRLRALRNKVKLELARVGSFSLHHIDRQA NGHADRLANAGLDRRRTKLECSVHPDGRG CTNTSVATAAPTAPAAPLPSARPPASTAAPS PDDDHSDQGDIDDGEVYAAMCISPDVPHR RPRLRLRRLTDEESEEAGNVVERLAASLAA KIADAPDWETAEGYITALPYALYDKLQPYS QSQHQPQRQQQQQRQRPRQQQQQTRQRRQ RRCKRGGGSQHRQRKTRRRRPPRVTRHHRE HRIDEALDDLHALESRRPQDRTAISKARRRV GRIRSALDQHQLRHRFDTDEKACVDGILAA ARDKDRAASVTTTAQTAAPPHSAPASAPSS AVDDGICPIPGDLLHAFFTDVNTPRTEFDAD SPIGARFREALAQLPAAIAATELLMEPPSPDE VEDQLQVRVGTSSPGLDGVGYDVYKTFTTQQ LLPALHAAFSRCWTDQRVPSWKLGVVRL LFKKGDRQDPANWRPICLQQAVYKLYAGIL AHRFTRWLDANTRHADAQKGFRAVNGCGE HNFLAATLTDNARRRRRELHV VVYDIKNAF GSVPHELLWEVLRRMGVPAQFIACCQGIYD AAAF TVGNAADGTTAPIQLRLGVFQGCPLSP HLFTAVISPLLHALKRLPGTGVQLSAVDRPG ASAYADDLKVFSDTKDGITRQHQLVTDFLR
--	--	--	--	---	--	--

				ACCGGCGGCGATTGAG CGGGTGCAACTTCGGG CACGTGTGCAACACAT GCAGCGCCCATGTCAC GCCAAGCGGCATCGCG GGACGACGCCTCGGCC GCTCAAGCGCAGCCCC GCCCTTCCAGCACGAC CTCGCGCCGTTTGGCGG ATCGCCATCAAGACGT GCGAGAGCCAGGCGGG GTCGGGCAAAATATAC TTACTCTAAGTATGCCC GAATCCCTGCCCTCTCA GGCTGAACGCGGCCCC ATACTTGATCTAAGTAT GGGAGGGATCCCTGGC CTCTCAGGCTGTACGCG AGACCCGTACGGCCGA ATCCCCTGGCCTCTCAG CCTGTACGCGGGGC (SEQ ID NO: 1321)	WTGMVANPSKCSTMSVQRDNRGVLKTANL TLQLDGAQIPALGMTEAYAYLGIGDGFHDV RRRVELAPKLRELKADTTALMQSGLAPWQ VVKALKVYIYPRVEYALRHLPFQQQLQGF DRHLARGLRHLLRLPTSATTEFLYAPTSRGG LGLLPLTEVHGALQIAHAWQTLHSPDPAIRR IARVQLRQVADARHRLDAEHWKERGEELC ERLLNSQLGTSAHAPPKRRNCDIGSLWVDV QRHLRSLGLQLQTAPADTHTGAPAQPLQLR VPHHDKWLTHKDVLRHVKLHIKNNHWHR WTSMRDQGKTARAHGGECSGFLTQPRGM WEADYRFAVAGRLNQVDITYSVLKRRRLRS HDCRCRQPGCHRAETLAHVLNHCPGTMDAV RGRHDGALKRIERELHASATDRRDRVELRV NQTVPFLAGPALRPDLQLYNHTKKTAVVD LAVAFEEQASDDASSALSILASHKRAKYDR IKRHLEQGWKVHLSALVYGS LGAVASGN YQVYTTHLGLLKRDAKRLDRQLSVECIQSS RRIWNLHCSQHRTRQHQARPSQGPRGSRAT ETGGTPSQTSRR (SEQ ID NO: 1443)
--	--	--	--	---	--

NeS L	Utopia- 2_PR	.	Phytopht hora ramorum	TCAAGCCCCGCCG CCAAGCCCAGCTG CGGCTGTTGCCGC CCCTCCAGCAGCA GCAGTCGCAGCTG AACCTACAGCCCC TCCTGCTGATCGC GCCGCCGTGGAGC CCCGCGCGCGCGC CGAGCCGCCCAA GAACAAGCACCCC CAGCTAGCGCGCG CGTGGAGCCC (SEQ ID NO: 1199)	TGAGCACCTTGGGTTC TCAAGCGTGATGCGAA GCGGCTGGACCGGCAG CTCTCGGTGGCGTGCAT CCAGTCCAGCCGCCGC ATCTGGAACCTGCACT GCAGCCAGCACCGCGC GCGCCAGCACCAAGCA CCAGGGGGAAGTCGGG CGGCGGAGACCGGGG GACTCCGCCGCGCACC GGCCGCCGCTAGACGG CACACAGGCCACAGC GGCCGACAGGGCCACA CCCAGGTAGGGAACCG CCCTCAAACCCCGCCG GTACATTATGGTCCGAC ACCTATGAGGTGCAAC CTGTACACAAGTTACA CACCACATAGCGACTA CCAGGTATTTACTACCT GGAAGCCAAGGATTAA CCGGTCGGTAATACAC	MSGDVVSSDGSSRTTDASGDGDDGAGSSDA AGDVGVVAMDVDQGARRQQPPWQRVGGK RRRLNDVDDDETRELAELLLEEEDEAGDHA PAPRLSAASARPASVLSVYAHNAQRFQCTL CTYTAASFASLKRHRDSRHRRTAFLDRFSA GCACGAPFASRLAAANHAHACASLNRTL SV AATPAAGELSPTAGAANATVKAATVTPDSP RQDPPELAASPPPLASSPDVAVQAADMQAPT RWDPPPLPRTLVA TRVASRLTDLTPPRWGPPL PRATVVSRIAARLEAAPT PRWGPPLPRVVVA SRIAERLAPPELAADDETKDGEEDQSFTPEV AAARSXGGEDANGEWLLRFDGACRANPGP GGAGAALFKPSGPVVWTC SHYMPSSSETNN TAEYTALLLGMRAAADHGATR VHVEGDST LVIQQVRGIFATRSTR LRGLRKS VKAEMAR MEHVTLHHIDRQANGHADRLANAALDRRK TKLECGLHPDGQGCSSAATTAVPSVVPDR PPSSTAAPTSAEPDETEQGDIDDGEVYAA MCIGPDSIPERRPRLRLRQLSETEEEEAGAIV ERLAATLAGKITDASDWATAEGYITALPYTL YDKLQPFAQHRHQPRPQHRQQPQRDPPLGT HDGDHGPSTSRSSRRRRRRRAKDRLRRRPPRI TRHHREHRLDEALDDLRAVEHASPHDRPAV
----------	-----------------	---	-----------------------------	--	--	---

					ATAACTTT (SEQ ID NO: 1322)	ARARRRVGRVNSAIAQQQLRHKFDKDEKA CVDGILAAARASRGLATPSASASRHPPPVPS TAADDGSCPIPSDELHAFPTAVNTPAGTFEP MAPVGAPFRSAVAHLPAATSQPELLSDAPTT DDIEDQLQRARGSSSPGLDGVGYDIYKAFA AQLLPALHAAFACCWRHKQVPQSWKVGVV RLLFKKGERTEPANWRPICLQQAIIYKLYAG VLARRLTRWLDANGRHADTQKGFRAMNG CGEHNFLAATLVDQARRKRRELHVWYDF ANAFGSVPHDLLWEALERQGVPSPFIACCR GLYADAVFTVGNAADGTTAPIALRVGVFQG CPLSPHLFTAIIAPLLHALKRLPDTGVQLSR VDCLGASAYADDLKIFSGTEGGTKRQHALV ADFLRWTGMRANPAKCCTMSVQRDTRGVL KACNLGLQLDGAPIPALTMSASYAYLGIGD GFDHVRRIELAPKLQELKHDATAALLQSGL APWQVVKAVKVYLYPRVEYALRHRLRPFHQ QLEGFDRHLVRGLRHLLRLPANATTAFFYA PVSRGGLGLLPLTELHAALQVAHGWQMLN SKDPAIRRIARVQLRQIADARHRIDAQAWQ DREEELAQLLLNSQLGASTGAPPKRRNGDIG SLWVDVQRHLRHLCLKLETAPACAETGTAA AMLQLRVPHHDKWLDHKTVLRHVKLHYK
--	--	--	--	--	-------------------------------	--

						NKHWARWAAMXDQGKTARTHGGAGSGFL TRPRGMWEADYRFAVAARLNQLDTHSVLK RRRLRXHDRCRQPGCTQGGDAGARAQPLR RHHGRGPRPPRRRPQAHRARAARVVAGRP GPRRAPGQPDGAVARRPRATARPPAVQPHQ EDGGGGRPGRGVRGAGERRPGELGAGTHR RTQAREVCRRQATPRAPRVEGPPLGARVRL ARRGAGRQPQGAY (SEQ ID NO: 1444)
NeS L	Utopia- 2_PU	.	Pythium ultimum	TGACTGGTGTTTG ATCACGATCAATG AGGTGATTAACAT GAGCCGGAGCAG GCCCCTTACACGC TGGTGCTGTAATG GTTTCAGGAATGCT CTTATGAGTAACT CCACAGTATAATT TTTGTTAGCGGGG GTGAGCGCGTGCG CCGCCCCCACCTT TCTCTTTTCGTTTG TATTTGGTTCCTCT GGAGCGACTTCCG	TAAGCGGGGGGTCCAG ACCCACAAAGAGAGAA GCAGGAATCATGGTCC GCATGGACCAGTAGGG CACGCTCCACAAAGGT TATCGCCCTCAAACCCA TCACACGAAGGATCTA AAAAGAAAGCAACAAT CGAAATAGTAAATAGT AAATAGCTTAGAAAGG TCAACATGCGAAATGC ATGAGGAACGTAAGAT GGTAATACATTTTATA (SEQ ID NO: 1323)	MDYDDSEFFDAICIPDEDADVLDGDEGDE GGNDDESSEPLPLAITNAPSAPLHATMLCGT VTQPWLLRFDGACRRNPGPAGAGATLTRPN GIILWTHYRYIPDKTATNNVAEYEALLDGLR CAAHHGVKHLRIEGDSNLVIEQVKGIFACST SLRPRRDQVREILRHFETYSFRHIDRALNRQ ADRLANQALDLLKTVSVCALSQTRVQDDT GAAHGCWHWTPPDASPTDDASTSILTQDVP VPMIDIDDYDPDEPMNAADDPVSINAEREGG TVYPVLRLLGPNVVPERQKRLQIPWLPPREM QKLEKKIEVLGETFASRIRDAPDWFS AEGYI TALTSELATLIRQSTAATTGPNAARPCERTIS KEKRRARRTTPLQRALAEAKHELQIIQPDAS RKSVRKAARRVKRISQAQQRHDLRRLFSTN ERRCVEKILRDPPVGPSSTSSSLPATDDDRCT

				TGGCTTTTGTGCGC GCGCTATCGCGCC CGGCATGAGGATC GTCGCTCCTGACG CCATGTGCGCCAG TATATCGGAGCCC CGTGCTCCGCTTG GGTGTGTGTTCCG CCACGACGCCACG CGATTTGCGTGCA CGGCATGCGCTGG CGACTTCCGCTCA GTGCGCGAACTCG CGGTGCATCGTCG ACGTGCGCACCGG TCTCTGGCCTTCC AGGACTGGTATGA GTCGTCGTGCGCG TGCAGAGCGCTCT TCACCGCCCGTCT CGACGCGTTCGTT CATTCGAGCCGTT GCGACCACAATGC		IDPADLFAYFQTQATAPTDFDDEGGELFR SVLDELPRADQEVHLLDEITRDEIEDQLSRI SKSTAPGLDGITNAVYVRFLQLLDALQAA FNACWRYNRVPSMWKAAFVRLIYKKGNRA VPSNWRPICLQQTIVYKLYTAILASRLQRWM DANARFTMSQKGFRAFNGCHEHNFVATCL HDQTRRLRKKLAIVWYDLRNAFGSLPHEYL WRVLARLGMPPQFVARVRQLYADASFTVE SRDGTTPVQLERGVYQGCPLSPYLFI AALIP LVRALHKLKDQHGI V LAPGVTDCVSAYAD DIKIFARSGTGAKALHEIVVRFLSWTNMAA NPAKCALMVTGDARGGDDTDASMTLSIEGE TIPRLTGKEGYVYLGVEDGLAHERRATCLR DSLKAASADVRLLRSDLAPWQIVRAIKSH VLSRFDYVLRHLRPFLSLFDGFDKMLVRGIK RLCQLPQTATSEFLFSPTSAGGLGFLPLKELF AALQIVHALQMLHSKDANVRAIARHQALQ VVRKRYALQSDHWSDREEELLEEFFNGTLE RSPFALAKKVSGDIASLWTDVRVNLTKYGL KFGEAHGRRLQPLVSHTDKQLAPQQWASAI KTHMRLRHLKRWTTLVDQGKTARMHERIG SAFLTRPSGVYDASYAFAVRARLNQVDTRS ALKRK RIVNNSRCRVSGCSELETLAHVLNH
--	--	--	--	--	--	---

				GAACCAATCAGCG AACGCCGTCCCCT CGTCCGCCGCCTC GACAACGAGAGC GGACGCGGATGCG ACGGCCCCCCTC CCGCTATCTCTCTT CCGCGGACGGGCT TCCCCGCCCCTG TCCTAGTTGCGCC GCTCTCCTTCTGG ATGCGGCAACTTG GACGACCCACGTG CACGCGTGCGCTG CCGCCAGTGCGAC GGCGTTGATCCAG CGTCATCTACACG CCGTTCCTCCCGT ATCGACAGCAACC GATCTCAACGGCG ATGATAGCCAACC TCCAGACCCGCGT GCCAACATCCAAC		CRFGSDSIRARHAETLLLIKTTMERELTRPGR QHQRLLVDATVPEARDPVPSNDAAESNIGA MAPSISHLRPDIQLYDNKTMEAVIIDLAVAF EDQSTDDAASSSSFARVKGVKTKKYEVIKQ FLEYKGYTVHVAALVYGSLGSVDTGNFAV YTERLGLRKGAVRRLECSLSARHINFAHRM WRRHAIAHTTGLRLIGTNSVQQQGVQRAPA EKQQHQRPVQRPSRAQAPRDQPSQQQSQSQ SFQQSQSQSQSQSQSQSQSQSRQRHAPTPTP VPVPVPVPVSTPTLTPTPTPTRRPKPAPSSTQ PQQGAPAQRRQQREQKKQPACRRRHATAV RETPPAAPTAARTATPTARPTATTRSTSTTRS TATATSTTRSSAPTSRPSAPRPRSSAPRPRSS APTTRSAAPTTRSAAPTATTSVYKSRRTT NAISGATTRRSASSKRAPMQPRTALTPTQQQ QRQ (SEQ ID NO: 1445)
--	--	--	--	---	--	---

				AGCTTCCAGCCAT CCATTGACAACG ACAACGACAACG ACAACGACGACG ACCCGCTGCTCGA CGACGTGCTC (SEQ ID NO: 1200)		
NeS L	Utopia- 3_PI		Phytopht hora infestans	GTACGGCCGAATC CCTGGCCTCTCAG CCTGTACGCGGGG CTATACTTGGTCT AAGT (SEQ ID NO: 1201)	TAGACGGCAAAGTTCT GGCCCGCGTAGGCCGA AAGGGCCCCGCCCAGG TAGGGTAACGCCCTCG GGAAAACCATTCTGGT GTTTGGCTGTTTTTCAA CAGTCGAACTTCAACTC GGAGCATATCAGATAC ACTTACTCACACATTTA GATATCAGATAGGGAA CCTTTATTAGGGAGATA ACGGGTACACCGGATG GTAAATATACAAAACC TTCTCTGTTCTAATCAG TGTGAAAACCTGGTTTTC GCCTTTTGGCGGACTTT	MPRSLASEPVHSSASRLPSQAPPTSRSGAAH TPPPSLQARPLASADAGLAATALASPQDPPY DVAPPGRAAGRLPDSVSPGATLSAATARAL AVRHWPLELDSOSSDDEDAQDPHAAAPGPP ADVASVLAPPGRAGSMLADPAALAAGLAR APPPPSAPQDPSPALPAGPAGQNPRAAAPAR VEVHTVAPPGRAGGMLPDPGLVDSSPAAA TAATPAPVAATATTARVAVEHHHAHAEPNQE HLPMARVLVEPMQVDECSSCDRSTLTADDG SGDDVAAPSSMLSNDVAAPMDVDSGTSCPP TLQQPLQRPRALHVGSKRRRLNADDGEEAH QLQEEEEAGIHAPALRLSAASAQPASVLAVY THNASRFDCTLCAYTAGSFASLKTHRNSRH RRTAFLDRFSAGCACGVPFASRLAAARHAQ ACASLSSAPLAEASSAAGASSHTVDGADST VSAAGHAEPDLPRHNATELTASPPLVSSTDV

					TTCACTCGCATTTTTGG GCAATCGTCTGCGGCT AGCTTGCTAGCGGCGG ACGAGCGGTCTCCGGG GGCGTTCACCTTTCCCC CGCGAGGCCAACTACA CCGATCTTCTCTACACT TTTCTAATTCGCCTCCG TCTTCGGTCTTCGGTTG TCGGGCTTTTTTCTTTTT GACCAATCAGAGCGCG CCATGCGACTCTTCTGG CCAATCAGAGACCGGG CCCTGTCCTCGGACAGC GAGGCCTCCACGGCCA GCCAATCAAGTCTCGG CAGCGACGCGTCTTTCT ATAGCGCAGCTGACGA GGCCGATCTGGCGGCC CCCGATTGGTCCGACTT TCGGCCAATCAGCGAC GACGAGGGGGCAGGGG TTTACACTTTTGCCCCC	EVQATETEATENRWGTPLPRVLVASRIAGR LAQVPPPRWGPPLPRTTIAGRIATRLAATPAP RWSPPLPRSLVASRIAGRLLPALPDAPACED EAKDSDEMDWEASEPHVEAPGPVDEETIDD ADGEWLLRFDGACRANPGPGGAGAALFKP SGPVVWTCSHYDPSTTATNNTAEYTALLG ARAAADHGVTKLRVEGDSTLVIQQVRGIFA TRSTRLRALRNKVKLELARVGSFSLHHIDRQ ANGHADRLANAGLDRRRTQLECSVHPDGR GCTNTSVATAAPTASAAPSTPTRPPATTAAP FHSDQGHIDEDDERRADIDDGEIYAPMTLGP DEVPARRPRLRLRQLSDEELEAAGAIVERLS ASLSAKITDAEDWGTAEGYITALPHLLYDKL LPYSRTAPRHQRPPRPSRNQQDHPQPRRDQP QRNVDEQQHAESQQGEDQRQQQPPTRRRR RRGKRRGRRQRRHPRQPGQSTRASQQSRQR RPRPPRVTRHHREHRIDEALDELHTLERARP QDRSAIDKARRRVGRVRGAINQHLLRHRFD TDEKACVADILEKAHAARAARTQAAGAA TSTGGAATAPTQQAATSALGDADDGTCPIL ADELWQYFTGTNTPRWEFNPATPVGEAFRT AMARLPPATRLRELLTEAPTADEIETQLQHV RGSSSPGLDGIGYDVYQRFAQQLLPVLTASF
--	--	--	--	--	--	--

					GTTTCGGCTTCAACTTC AGGCCAAAATGGCGAT TTGGACCCTCCACGCGC CGTGCCACTGCTCGGC ACCGGCGGCGATTAG CGGGTGCAACTTCGGG CACGTGTGCAACACAT GCAGCGCCCATTCAC GCCAAGCGGCATCGCG GGACGACGCCTCGGCC GCTCAAGCGCAGCCCC GCCCTTCCAGCACGAC CTCGCGCCGTTTGGCGG ATCGCCATCAAGACGT GCGAGAGCCAGGCGGG GTCGGGCAAAATATAC TTACTCTAAGTATGCCC GAATCCCTGCCCTCTCA GGCTGTACGCGGCCCC ATACTTGACCTAAGTAT GGGAGGATCCCTGGCC TCTCAGGCTGTACGCGR GGACCAAGTACAGCCG	KRCWTAKMVPQSWQVGVVRLLYKKGAHD DPANWRPICLQQAIYKLYTGVLARRLVRWL DVNDRHAPGQKGFRVNGCGEHNFLAATLI DQARRKRRSLYEYVWYDFRNAFGSVPFQLL WDSLQRLGAPADFIDMCKGLYHQAAFVIGN AADGPTAAIRQQVGVFQGCPLSPQLFNVAIS PLLFALRRLPETGVQLSGDDRVGVSAYADD LKTFSSTKAGATKQHELVA AFLAWTGMKA NAAKCSSMGVRRNSNGATEADNLDLALDG TPIPSMTHMQSYTYLGIGDGFHDVHRRIELA PKLKTLLKQD TTALLESGLAPWQVVKAVKV YLYPRVEYALRHLPEDQLLESFDLHLRAG LRHLLRLPKNANNDFFYSPVSRGGLGLLPLV ELHAALQIAHGWQMLNSTDPATRRIAREQL HQIADARHRLDKAHWKERGDEL CQLFLNL DLG TSAHAPPKRRNCDIGSLWVDVRKNLQA FGLKLETAPADAESGTPALPLQLRVPHHEK WLTHRDVLRHV KQHLKNKHWRAWCAFDQD QGR TARAHGGVGSSFITRPRGMWESDYRFA VAARLNMVDTSATLARRRLRAHDRCRYPG CRWKESLEHVLNHCPGTMDAVRGRHDGVL REIEHALRAPSGARREL RVNQTVPGLP GPAL RPDIQVYNHDQRTVAVVDLAVAFDRQDRD
--	--	--	--	--	---	---

					AATCCCTGGCCTCTCAG CCTGTACGCGGGGCTA TACTTGGTCTAAGTATG CCCCGGTCGCTGGCCTC TGAGCCCGTACGC (SEQ ID NO: 1324)	DPETSLAKAAAEKKAKYTGIQRHLERQG WKVHLSALVYGSLSVAPNNYKVYTEHLG LLKRDAKRLDRTL SVACIQSSRRIWNLHCA KHRARQHQTPTSQSRGRRVTETGGAPSRTDR R (SEQ ID NO: 1446)
NeS L	Utopia- 3_PR	.	Phytopht hora ramorum	TGCGCGGCAGACC AAGACGCGCAGCC AACAAACAGACCGT GCAGCGTGCGGGT GGAAAGCGCCGCC GCCTGAACGCTGG TGACGATGAAGAC CAGCGAGAGCTGG CCGAGCTCCTGCT CGTGGACGAGGAC AAGGCTGGCGCCG AACACCCCGCGCT CAGGCTGCCCACG GCCAGCGCTCATC CGGCCTCCGTCCT CTCCGTGTACGCG CACGCTGCAACTC	TAGACGGCAACACTCT GGCCCGCGTCGGCCGA AAGGGCCCCACCCACG TAGGGAACCGCCCTCG GGAAACCCAGTCTGGT GTTCGGCCAAGAAATG CACCACCACCACGGCG GAGGTGCATTTGACA GTCGAACTTCAACCCG CCACATATCGGATATA GTTACAGCTCTAGTTAG ACATCGGATAGGAACT TCTTAGAAAATTAACG GGTATAACGGATGGTA AATAAAATAAAAACTT C (SEQ ID NO: 1325)	MQDQVDAEQQARNRWGPPLPRPLVASRVA ARLGEVPPPRWGPPLPRGVVVSRIAARLEAV PVPRGGPPLPRSFVATRIADRLAPPSDLSLL DEEMKESEPPDPTHHSADEDSTDAETADAV MEPAFVSDPPTATPREWRLQFDGACRGGPN PGGAGALLYNPEGAVVWTGSHYMPGAKET NNSAEYTALLIGARAAADHGARQLRIEGDS LLVIRQVKGLYATKSTRLRQLRNAVRHELA RVGQHSLHHIDRQGNFADRLANRALDLKS DKVECKEHPVAGACTTCMGSPSAGPPATPP PTTADIEMADAGSDDEL RADIDDGEVYAPM RLEPGVIPTRRSRLRLRQLTDDEMEAAGEVV ERLSAGLSAKIADADDWETAEGYITALPYM LYDKLQQYTQVRHGTARSPAPHPQRRDVQ GQVETHREPRHETIGQPDQPGEPSPTRRRRR GKRKGRRQRRHPRRTNCGGGGRQQRKQRH PRPPRGTRHHREHRIDEAIDELHALERARPQ

				GCTTCGACTGCAC GCTGTGCACGTAC ACGGCTGCCAGCC TCGCTTCGCTCAA GCGCCACCGCTCG TCTCGGCACCGAC GCACGGCCTTCCT CGACAAGTTCTTG GCGGGCTGCGCGT GCGGCACGCCCTT CGCATCGAGGTTG GCCGCAGCCAGAC ACGCGCAAGCGTG CGCCAACCTCTGC ACCACCTCGGCGA CGACTTCGACGGC AGCAAAGGCATCA AGCCCCACTGCTG CCGGAGGCAGACC CACCGTCCGTGCA GTGGTCACCGCCG CGCCCGACCTGCC CCGCCAGTATCCC		ARPAIAKARRRVGRIRSAIDQQLLRHRFDTA EKECVDGILAAARTARDARTTVRAAAATGT TATPETAVTSGTEQQDDNGTCPIPSEVLWRH FDSVNTPQRDFDPEAPEGAAFRSAMARLPA ATRFMELLKEEPSTDGIEVQLQHASSTSSPG LDGVGYDVYKRFA SQLLPVLKAAFKCCWT HKQVPQSWKLG VVRLLYKKGDREDPANW RPICLQQAIYKIYTGVLARRLTRWQDANDR HAPGQKGFRPVNGCGEHNFLAAMLIDHARR KHRPLYEVWYDFRNAFGSVPLGLLWDALE RTGVPAEYIAAVQGLYDHA AFMVGN AVDG STAPILQRVGVFQGCPLSPPLFSA AISPLLHA LQLLPSSGVQLSGDDRPGVSAYADDLKTFS GTKAGVTEQHEL VAMFLRWTGMAANPAK CRSMGVRRNGNGAIEADHLELALDDTPIPTL THLQSYTYLGIGDGF DHVRRRVALAPKLKL LKQDATA LMESGLAPWQVVKAVKGYLYPR VEYALRHLPDDQLLESFDLHLRRGLRHLL RLPKSANND FVYAPVSRGGLGFLPLVELHA ALQIAHGWQMINS PDPAIRRIAREQLHQVAD ARHRLDKDHWKQRGDELCELLN GELGTS AHAPPKRRNGDIGSLWVDVRKNLKA FGLKL ATAPADPESGAPAKPLQLCVPHHA EWLDHR
--	--	--	--	---	--	---

				TCGGAGCTCGTTG CGTCCCCCCCCGCA GCCGAGCTCCACC AACGTTGCA (SEQ ID NO: 1202)		NVLRHVKQHMKNKRWRRAWCSHVDQGRTA RAHGGVGSGLTRPRGMWESDYRFAVAAR LNMLDTVNVLARRRRLRAHDCRHPGCRWK ETLAHVLNHCPTGTMDSIRGRHDDALKEIERT LHASSGDRQGRTELRTNQTVPGLAGPALRP DLQVYNHDQRTVAVVDLAIADFDEQPRDDPE SSGLAKAAAEKKAKYAGIKRHLERQGWKV HLSALVYGSLGSVAPSNYKVYTEHLGLLKR DAKRLDRQLSVACIQSSRRIWNLHCAQHRA RQHQQDPAPRGRRVTETGGTPSRTDRR (SEQ ID NO: 1447)
NeS L	Utopia- 4_PI	AAT U010 01281 .1	Phytopht hora infestans	AGCTCGGCCTCGC GGCTGCCTTCCCA GGCGCCGCCGACT TCGCGCTCTGGCG CGGCCCACACGCC GCCGCCGAGCCTC CAAGCGCGCCCGT TGGCTTTCGCAGA CGCAGGGCTCGCG GCGACGGCCCTCG CCAGCCCCCAAGA CCCCCCTACGAT	TAAACGGGTCACTTGA CCGACAGGGCACCACC CAGGTAGGGAACCGCC CTTTAAAACCCAGGAA GACACAAACACCCTCC ACATAGTGACATACAT ATTTTAGCCTAGATTTC AGTTACGGAGAGGTTA CTAACTGGTACATAAA ATTACACATTCTGTTCT AATCAGTGTGAAAAC GGTTTTCGCCTTTTGGC	MLADPAALAAGLARAPPPPSAPQDPSPAFPA GPAGQNPRAAAPARVEVHTVVAPPGRAGG MLPDPGLVEEPIQATYAHDAAQFECALCPY VAESMAVLVQHRRSAHRGTRFKDIFTSGCQ CSLVFYARIVAASHAVACARRNQRAVPPAP TPVAPTRPEATPQPTGYLAAAMTAAAAAAS SDTVVAAATNMQSAVPAAAKTTGLQLVPPE LEPALPQRASCHAGKRRRLNADEAVTPCTP TARVSPQTEVAMAPHDAPQDDTVLQREAA EPQPDPAATQGAQVQRVEDTTAAQDDTVQ QDHDADTAQVSPRRTPTRWGPRPSSTQEPS PMTGEPAAATLAARRPLTPAATGTRATRWGP

				GTGGCACCACCCG GCAGGGCGGCCG GCAGGCTGCCCCGA CTCGGTATCGCCG GGTGCTACACTCT CAGCCGCTACAGC TCGGGCCTTGGCG GTCCGCTATTGGC CCTTGAGCTCGA CGGCGACAGCAGC GACGACGAGGAC GCTCAAGACCCCC ACGCCGCCGCCCC AGAACCCCCAGAA GACGTCGCGAGTG TGCTTGCCCCACC CGGCAGGGCAGG CAGC (SEQ ID NO: 1203)	GGACTTTTTCACTCGCA TTTTTGGGCAATCGTCT GCGGCTAGCTTGCTAG CGGCGGACGAGCGGTC TCCGGGGGCGTACACC TTTCCCCCGCGAGGCCA ACTACACCGATCTTCTC TACACTTTTCTAATTCG CCTTCGTCTTCGGTCTT CGGCTGTCGGATTTTTT TCTTTTGGACCAATCAG AGTGCGCCATGCGACT CTTCTGGCCAATCAGA GACCGGGCCCTGTCCTC GGACAGCGAGGCCTCC ACGGCCAGCCAATCAA GTCTCGGCAGCGACGC GTCTTTCTATAGCGCAG CTGACGAGGCCGATCT GGCGGCCCCCGATTGG TCCGACTTTCGGCCAAT CAGCGACGACGAGGGG GCAGGAGTTTACACTTT	CHRAIGAAAIARLVTGLPTEPAQPQRRQPPP PQEPPLQPEPQAAAATVAADIAATVAADIAA AAANAAMDVDGGPAADETWLLRFDGACR RNPGGGAGAALFAPSGAVVWTCSHFMPSR SETNNTAEYTALLLGAQSAVHHGAKRLNIE GDShLILSQVRGAFACNNKRLRSLRNRVQA SLRQLDWYRLQHIDRKANQHADRLANRAL DLRRTVTECGPHAETRNRCFQTPQPLVEPGE THCVPGSDEVLAANTAMEDATAVPTEDDE AEVAARDGGEVFPTIAIGPDSAPAKQPRRLRL KKLDEDDFDAAAAAVTRVSEELASKIVDAG DWTSGEGYISAIPERLRAALRPFALPTQPARP QPREPRMQPPRRPPRVTRDHLEHRLDEAL DTMENVQRSTPQNQKAVRRARRRVGRLRS AMDRTLRLKKFATHERECVAEILRRASTEE AANPSQEKCPIDRATLHEYFTATSTQRTPFD YDSAKGTEFRTFLEVMSTPSHETSALTAEPT LDEIEDQLAHVKAGSSPGHDGVGYDVYRRF QVQLLPLLHAAFRFCWRHRRVPALWKVGF VRLHKKGDPQQPNNWRPICLQTAIYKLYS GLLARRLSKFLEANELLPMAQKGFRAFNGC HEHNFVATTLLDQTRRMHRRLYQVWYDLR NAFGSLPQQLMWGVLRQLGVTEEFVARCS
--	--	--	--	---	--	--

					TGCCCCCGTTTCGACAT CAACTTCAGGCCAAAA TGGCGATTTTCGACCCTC CACGCGCCGTGCCACT GCTCGGCACCGGCGGC GATTCAGCGGGTGCAA CTTCGGGCACGTGTGC AACACATGCAGCGCCC ATTGCACGCCAAGCGG CATCGCGGGACGACGC CTCGGCCGCTCAAGCG CAGCCCCGCCCTTCCAG CACGACCTCGCGCCGTT TGGCGGATCGCCATCA AGACGTGCGAGAGCCA GGCGGGGTCGGGCAA ATATACTTACTCTAAGT ATGCCCGAATCCCTGCC CTCTCAGGCTGTACGCG GCCCCATACTTGATCTA AGTATGGGAGGATCCC TGGCCTCTCAGGCTGTA CGCGAGACCCGTACGG	GIYEDSYFVVGNASDGATEPVRQEVG VYQG CPLSPLL FITALVPLLRALENQDGVGVPLAD GVRPCATAYADDIKVFCDSATGIQRCHALV TRFLEWTGLQANPAKCAFLPVTRSQHSNPT RDRDIELRIHGEAIALTLGLQESYRYLGVGDG FDHVRHRLQLEPKLKQIKREAVALLHSELVP WQILKALKVYIYPKVEYALRHLRPLKSQLQ GFDSAIVRGLRHLLRLPENSHDGLFFSPTSA GGLGLLSLVELHEALQVAHAWQMLHSDP AIRAIARTQVGQVARKRFLVEEHWRGRED DLAQRFLNTELAASPHATETRRNGDIGSLW NDVRDTLQTLGLKFAAGDEEEAPGLLQLRV PHHTKWL SHSTVLRHVKLHMKLRRMDTW KSKVSQGTTVREHGGVGSRFITAGAGLSDA EYRFAIAPRAHLIDTNSTLKRRRLRANDTCR APGCSYTEPPAHILNKCSPNMDAIRKRHDDA LERIADALRRKVEKSGGRLEVAINKTVP EYD GAALRPDIVLRNTETKRAIADLAITHENQPT DATTSSALQQSRDNKITKYQTVAAAMMRA GWRVRVTGIVYGSLSVLP SNFKVYTELLA LLKRDARRLNRQLSSH CIRASARIWSAHCRR HRERQRSGNASRASRGSGGAPRRTSQASAR R (SEQ ID NO: 1448)
--	--	--	--	--	---	--

					CCGAATCCCTGGCTTCT CAGCCTGTACGCGGGG CTATACTTGGTCTAAGT ATGCCCCGGTCGCTGG CCTCTGAGCCCGTACAC A (SEQ ID NO: 1326)	
NeS L	Utopia- 4_PR	.	Phytopht hora ramorum	CCAGGAATCACCC CCGCCGCCCCCAA GCGCTCCGCGCC GAGCCCAGCTGCG GCTGCTGCCGCCC CACTGGTGACGGC AGTCGCAGCTGCC CCTGCCACCTCTC CTGTTGGCCGCGC CGCCGTGGAGCCC CGCGCGCGCGCCG AGCCGCCCCAAGA ACAAGCACCCCCA GCTAGCGCGCGCG TGGAGCCC (SEQ ID NO: 1204)	TAGACGGCACACAGGC CCACAGCGGCCGACAG GGCCACACCCAGGTAG GGAACCGCCTTCAAAC CCGGCCGGTACATTAT GGTCCGACACCTATGA GGTGCAACCGGTACAC AAGTTACACACCACAT AGCGACTACCAGGTAT TTACTACCTGGAAGCC AAGGATTAACCGGTCG GTAATACACATACTTT (SEQ ID NO: 1327)	MSGDVVSSDGSSRTTDASGDGDDGAGSSDA AGDVGVVAMDVDQGARRQPPWQRVGGK RRRLNDVDDDETRELAELLLEEDEAGDHA PAPRLSAASARPASVLSVYAHNAQRFQCTL CTYTAASFASLKRHRDSRHRRTAFLDRFSA GCACGVPFASRLAAANHAHACDSLNRTFSV AAAPAAGELSPTAGAANATVKAATVTPDSP RQDPPKLAATPPLASSALVDPDHAEQQAR ERWGPPLPRTL VAGRVAARLSEVPAPRWGP PLPRGVVAFRIGHRVLPPEMTSDEETKDDSS VQDGDQRQDYPVAAMDVDMSGMSGEWLLRF DGACRANPGPGGAGAALSQPDGSVVTCS HYMPSSSETNNTAEYTALLLGTRAAADHGT TTLRVEGDSTLVIQQVRGIFATRSVTLRHLR DQVKLELARVGRFSLHHIDRQANAHADRLA NRALDLRRTVSECGVHPDGNGCTPTAIDDR PLAPTQQPPDAPPPPPAADIEMEDPDDEDLA

						DIDDGEVYAAMRVGPNATPQRRRRGRSGT AKKHRRQRPPRVTRHHREHRLDEALDDLH AVERSTPSDRTTVRRARRRVGRVNSAIEQQ RLRHRFDTDEKACVTDILAKACATREAART TASGGDPPAGPATPAAGSADDGTCPILGEEL WRFSDSVNTPRQEFAPDAPVGAAFRSALAR LPAATSCKELLTAAPSAGEVEDQLQHVRGA SSPGLDGVGYDVYQHFAAQLLPALTAAFKA CWTAKRVPQSWKLGVVRLHKKGAREDPA NWRPICLQQAIYKLYTGLLARRLVRWLDAN DRHAPGQKGFRAVNGCGEHNFLAATLVDQ ARRKRRTLFEVWYDFRNAFGSVPFALLWD ALARLGVPDDYVTMCKGLYESAAFVVGNA IDGTTDPIALRVGVFQGCPLSPQLFNAAISPL LFALQRLPATGVQLSGDDCPGASAYADDLK IFSGTEDGIKRQHALVADFLRWTGMAANPN KCCTMSVQRDGRGVLKTDDLQLDLAGTPIP ALSMSASYTYLGIGDGFHDVRRRVELAPAL KQLKDDATTLLQSGLAPWQVVKAVKTYLY PRVEYALRHLRPFQQQLEGFDRHLARGLRH LLRLPGNATAECFYAPVSRGGLGLLPLTELH AALQVAHGWQLLNSKDPAIRRIARVQLRQI ADARHRIDSRAWEGRDEELCELLNSQLGT
--	--	--	--	--	--	---

						SPDAPPKRRNGDIGSLWVDVQRHLRTLGLK FATAPACADAGSAATTLQLRVPHHDKWLD HRTVLRHVKLHVKHRHWSKWAAMRDQGK TARAHHGAGSGFLTRPRGMWEADYRFAVA ARLNQLDTHSVLKRRRLRAHDHCRQPGCSR AETLAHVLNHCAGTMDAVRGRHDDALKHI ERALHASSPGGQDRVELRVNQTVPSLAGPA LRPDLQLYNHTKKMVAVVDLAVAFEEQAS DDPESSALARIAAHKRAKYAGVKRHLERQG WKVHLSALVYGSLGAVPAGNHKVLTEHLG LLKRDAKRLDRQLSVACIQSSRRIWNLHCSQ HRARQHQAPEGGSRAAETGGTPPRTGRR (SEQ ID NO: 1449)
NeS L	Utopia- 5_PI	.	Phytopht hora infestans	CTCAAGCCTAGCA GCGGCTACCGCAG CTACTCCAGCTCC GGTAGCTGCTACT GCTACAAGCTGCTC GCGCTGCTGCTCG CGTCGCCGTGGAG CACCACGCGCACG CTGAACCGAACCA	TGACGCACCGTGACAT AGTGCGGCACGTGAAG ATGCACATGAAGCTCC GACACTGGGCCAAGTG GGCGGCCATGCGCGAC CAAGGCAAGACAGCTC GTGCACATGGTGGGGT TGGTAGTGGCTTCCTCA CACGGCCGCGAGGCCT GTGGGAAGCCGACTAC	MARVLVEPMQVDECSSCDRSTLTADDGSG DDVAAPSSLNSNDVAAPMDVDSGTDCPPAL QPPQRPRALHVGSKRRRLDADDEEEARQL QEEEEAGIHAPALRLSAASAQPASVLAVYTH NASRFDCTLYAYTAGSFASLKTHRNSRHRR TAFLDRFSAGCACGVPFASRLAAARHAQAC ASLSSAPLAEASSAAGASSHTVDEADSTVSA AGHTEPDLPRHNATELTASPPLVSSPDVEVQ APETEATENRWGTPLPRVLVASRIAGRLAQ VPPPRWGPPLPRTTIAGRIATRLAATPAPRW

				AGAACATCTACCG (SEQ ID NO: 1205)	CGGTTCGCGGTGGCCG GCCGCTTAAACCAGGT AGACACGCACAGTGTC CTCAAGCGCCGGCGCC TCCGAGCACATGACAG GTGCAGACACCCAGGA TGCACGCGCTCCGAGA CGCTGGCGCATGTGCTT AACCACTGCGACGGAA CCATGGACGCAGTCCG TGGCCGGCCATGACGC CGCACTCAAGATTATTG AGCGTGCGCTCCTCGC ATCGTCGGCCGACCAG CAGGACCGTGCTGAGC TCCGCGTGAACCAGAC CGTGCCGTCCTCGCCG GCCCCGCGCTACGGCC CGACCTTCAGCTCTACA ACCACACCAAGAAGAC GGTGGCGGTGGTCGAC CTGGCCGTGGCGTTGA GGAGCAGGCGAGTGAC	DPPLPRSLVVSRIAARLLPALPDAPACEEEA KSDTMDWAPTWTNEETKDSEPHDEAPGQ VDEETIDDADGEWLLRFDGACRANPGPGGA GAALFKPSGPVWWTCSHYDPSTTATNNTAE YTALLGARAAADHGVTKLRVEGDSTLVIQ QVRGIFATRSTRLRALRNKVKLELARVGSFS LHHIDRQANGHADRLANAGLDRRRTKLECS VHPDGRGCTNTSVATAAPTAPAAPLPPARPP ATTAAPSHDDHHSVQGDIDDGEVYAAMCIG PDAVPHRRPRLRLRHLTDEESEEEAGDVVER LAASLAAKIADAPDWETAEGYITALPYALY DKLQPYSQAQPQPPSQQQQQQQQRPRQQQ QTRQRRQRRGKRGGGSQRRQRKTRRRRPPR VTRHHREHRIDEALDDLHAIESRRPQDRTAI SKARRRVGRIRSALDQHQLRHRFDTDEKVC VDAILAGARASQGATTAPPSATTDPPAPMD DSRCPPIPGDDLWRFFDSVNTPRRSFDAEAPD GAAFREAMACLPAAATRAQELLTEAPTVDEV EDQIQHARASSSPGLDGVGYDIYKQFAAHL LPALHTAFVCCWNHKRVPQSWKLGVVRL HKKGDRQDPANWRPICLQQTIYKLYAGILS RRFVRWLDANARHAEAQKGFRAMNGCGE HNFLAATLVDHARRKRKELHVWVWYDLANA
--	--	--	--	------------------------------------	---	--

					GACGCGAGTAGCTCGG CACTGTCCCGGATCGCC AACCACAAGCGAGCCA AGTACGACCACATCAA GCTACACCTCGAGCGC CAAGGATGGAAGGTAC ACCTCTCGGCACTCGTG TACGGGTCGCTTGGGG CGGTCGCTAGTGGCAA CTACCAGGTGTACACC ACACACCTGGGGCTAC TCAAGCGCGATGCAAA GCGGCTGGACCGGCAG CTGTCTGCCTAATGCAT CCAGTCCAGCCGCCGC ATCTAGAATCTACTG CAGCCAGCACCGGACT CGCCAACACCAGGCGA GGCCCAGCCAAGGACC AAGAGGCAGCCGGGCG ACGGAGACCGGGGGGA CTCCGTCCCAGACAAG CCGCCGCTAGGCGGAA	FGSVPHDLLWETLARQGVPTFVDCCRGIYS DAAFTIGNAADGTTAPIRLRVGVFQGCPLSP HLFTAAIAPLLHALKRLPVTGVQLTGVDRPG AAAYADDLKTFSSSVDGIKRQHELVATFLR WTGMAANLSKCSAMSVQRDSRGVLKTGDL CLKLNDAEIPALSMTASYAYLGIGDGFHDV RRRLELAPMMKQLKHDATALMQSGLASWQ VVKAVKVLYLPRIEYALRHLPFKQQLEAF DEHLRRGLRHLLRLPTNATSAFFSAPTSRGG LGLLPLTELHAALQIAHGWQILNSPDGATQR IAREQLREIPDARHRLDTAHWRNRDAELCE LLLNSQLGQSSHAPPKRRNCDIGSLWIDIRR QLGTLGLKFETAPGRRSHQPARPAIAAFACR TTTSG (SEQ ID NO: 1450)
--	--	--	--	--	---	---

					ACCAGGCCCAAGACGG CCGACAGGGCCCCACC CAGGTAGGGAACCGCC CTAGAAACCCATTTTCG GTGGTCGACTCGCAAG CCTTACCTATATTTTAG ACGTAGCGACCAAATT ACAAATTTGGTAACGA GTAAGCCAAATGGTAA TACACAAAACCTTT (SEQ ID NO: 1328)	
NeS L	Utopia- 5_PR		Phytopht hora ramorum	CCAGGCATCACCC CCGCCGCCCCCAA GCGCTCCGCCGCC GAGCGCAGCTGCG GCTGCTGCCGCC CACTGGTGACGGC AGTCGCAGCTGCC CCTGCCACCTCTC CTGTTGGCCGCGC CGCCGTGGAGCCC CGCGCGCGCGCCG AGCCGCCCCAAGA	TGATGCCCCGCAACTA CAAGGTGCTTACTGAG CACCTTGGGCTGCTCAA GCGTGATGCGAAGCGG CTGGACCGGCAGCTGT CGGTGGCGTG CATCCA GTCCAGCCGCCGCATCT GGAACCTGCACTGCGC GCAGCATCGAGCACGG CAGCACCAGGGCCAAG CGCCAAGGGGCAGTCG GGCGGCGGAGACCGGG	MSGDVVSSDGSSRTTDVSGDGDDGADGAG SSDAAGDVG VVAMDVDQGARRQRPPWQR VGGKRRRLNDVDDDETRELAE LLEEDEV GAQAPALRLFAAS AHPASVLSVYAHNAQRF VCTL CAYTAASFASLKRHRDSRHRRVSFVD KFSAGCACGTPFGSRLAAARHAQACASLSIP RTVTAPAAAGDLSPTATGANATASAAATSP DLPRPASPELAASPPQTSPFDVAIQADAAEQ TAWTRWD PPLTRA AVAARVASRLAVVPAP RWGPPLSRTL VASRIAARLDAQTSRWGPPLP PAMVASRIASRLAAMPAPRWGPPLLRTVIAS RIADRLLPELA ADEETKDDDVHMDNAASV

				ACAAGCACCCCA GCTAGCGCGCGCG CGGAGCCC (SEQ ID NO: 1206)	GGGACTCCGCCACAGA GCGGCCGCGCTAAGC GGACATCGGGCCCGTA GCGGCCGACAGGGCCA CACCCATGTAGGGAAC CGCCCTCTAAACCCGCC CGGTACATTATGGTCCG ACACCTATGAGGTGCA ACCGGTACACAAGTTA CACACCACATAGCGAC TACCAGGTATTTACTAC CTGGAAGCCAAGGATT AACCGGTCGGTAATAC ACATAACTTT (SEQ ID NO: 1329)	DVDEESEVADVVMTDHDGEWLLRFDGACR ANPGPGGAGAALFKPSGPVWACSRYPSS SATNNTAEYTAALLGARAAADHGATHLRV EGDSTLVIQQVRGIFAARSTRMRALRNQVQ SELARVGSFSLHHIDRQDNAHADRLANRAL DLRRTVIECGIHCDGVGCTATTTEVQSSSAP EIPTRPVADHDHEHEVVDVVDVCGVCGVC GDRGTCGVCDVSGDIDDGEVYAAMRTGPD AVPARRPRLRLRKLTDDEEQEEAGTLAERLG ATLAAKIADARDWESAEGYITALPYLLYDK LLPYSQGPARSPLVRQHQRQQQPDGQFQR PTQSRSAARRQRRQRHRARRRPPRVTRHHR EHRLDEALDDLHAVERATPSDRRSIRRARRR VGRVNSAVEQQRLRHHFDTNEKGCVEILLA KARAQRSTTVARTAVGEPNSGAAEDDGTCP IPSERLHRHFTEVNTPGSSFDAMAPVGAPFR AALAHLPAAATEASELLTEAPTPDEIEDQLQR AKGTTSPGLDGVGYDVYKAFSTQLLPVLHA AFQCCWQHHRVPQSRKQGIVRLLYKKGPRE DPANWRPICLQQVIYKTYAGVLARRFTRWL AANGRHADAQKGFRTVNGCGEHNFLASTLI DHARRSRRELHMVWYDLKNAFGSVPQELL WEVLQRMGVPPAFVEVCQGLYQDAAFTVG
--	--	--	--	--	---	---

						NAADGPTDLVRQLVGVFQGCPLSPHLFTA AI SPLLHALDRLKDTGVRLSADDRPGASAYAD DLKIFSGTADGVKRQHALVADFLRWTGMV ANPNKCCTMSVQRDGRGVLKACDLELQLD GARIPSLIMNASYAYLGTGDGFDHVRRIEL VPALMQLKDDATALLQSGLAPCQVVKAVK TYLFPRVEYALRHLRPFQQQLEGFNRHLVC GLRHLLRLPVSATTSFFFVPVSRGGLGLLPLT ELHAAPADRRPAPPPRPRPLEGAGGENMRA ADQLAARDVGPRPTQAPQRRHRLVVGRRP APPPRTRPQARDRAGVRGDRHRGGDAAAS RAAPREVAGPPHGPAAGEAAHEEQALAAV GRDEGPRQDRPHPWWCRERLPHAASRPMG DRLPLCSGGSAPAGHAQRAEAPAPPRARP VSTAGLLPCRDGTGTRAESLRRHHGRGPRPPR RCAQDHRARAHVVRARTAPSSGSTRPCP RSPAPRCGPTSSSSTTPRRRWRWSTWPWRS RSRRATTPRALRWASPRTSEPSTPASSGTSS AKGGRSTSRRSCTARWAR (SEQ ID NO: 1451)
NeS L	YURECi	.	Ciona intestinal is	ATCAACCCACTAC TACACCCTCTACA AAGAACCCACTAC AGCAAGATCTTGG	TAGCCAAAAGATTTGG TGTTCTTGCGAGCTGC AGTCCCGGCAGAGTAG GCGAGATCGTGAGCTC	MATSSSSVSSGNVQTEVRCVYHGKGDLLLE CPVAHCPSIHPTVATITKHLKKHHTPQFEQIT TKNLTITYTCSQCSFSTTGLTQHHISKHYKTC KGVGAVQEGNKGRFCCPACGTRWALLCKA

				CGACCAACTACAC	TACACCGCTTCACGATT	RHHFNNVHFEYDTPPIAAFSGTPYKLKKRKF
				CTGCAGCCTCCCG	TGACGAGTTCTTCTTGA	TIINKALTYSCPLPLNQLLCPLWSCSLTILNK
				TCGACTCTTCACC	GTCGATCCCCAACGGC	PLSSVQQETAHGDGSQGQSYVPTQLRQVLR
				ACCTGCGCACCTC	AACGCACTCCAAATTT	ARCHCGNPPIGKGHWASQCQGRPLSSPKGG
				CCTCCGACTCCAA	AACGACCGCATGGCGA	RSSPTPPANLTLHFLNYLPFQLPSQSSSPQSST
				GCGGCAGACAGTG	CGCCACTGATCTCCCAG	LDPTACKARVPIPSFLRGDCEVTFFIIPSVNFY
				GCACCACCAGCAC	TCATCCGTGCTCCGAGG	RPYLSYPLRMFWRNRTSSGHCSLHRVVRGF
				TCCAAGGGTCCAC	GCACTTAAGGGGAACC	VRERLVPHPRSKSPARTPLEFLCEFRLAGVFP
				AGCCAGAAGAGT	GATTCACTGATGCTGCA	DPGKVASLRPVPAPLTLCLSPVAGPMISCE
				GCACCCCCACTG	AGTTGGCGGGGGCGCA	DHSAPPSVRSSSPIPNPASVSSVEAHLSDLL
				CTAGGAAGAGCTA	CTCACCAATGGAAGAT	DKVSSGELRPLSPTLPSSGFFGPLLPTPPPRP
				GGCCATTGCCCGC	GAGGGAGCCCAACGAT	TPSAEKASPSGLSYLPCREVKIASIARPSPAS
				TGTGGGCCTTGTG	GCCCCGAAC TACCGCA	QRVGCDADRTGPSLNP NYQQTSPSTPSFSPI
				CTGGCTGTCTGCC	CCGTCCCCAGCACCGT	VRPPKFPRSGAKVNSKSKPPGVRPRRAKPIE
				GCTGATTGGCACG	GTCCTTGTGAGTAATTA	PGTESASPVDVDTISSSVQEPCTPENRTPEFF
				TTAGGGAGTGTGC	TTTTCACTAATCCAGGG	YERKWLVSILNIHEREGSNFFQFN RDLEYWT
				GCTGAGCACGATA	ACGGGGCCATGAACTG	QLLSGSQKGGRAKRASYNRGAANQAMKNR
				GGGAGGCTGCTGA	AACTATGTCTGCACGCT	DSGRKDFDPRPVAGHSSGGGTELGSRPRYP
				AGCCGAAGAACCC	GCCCCGCCTAGCCTCG	KGARLRADFWRDMKGTVRKLLDGSNGERR
				GCGCCGGGGACTT	GCCACAAATAAATAGT	CGIPLDIIERKFRQVSM PGWIDHRRYAAGAS
				TAGGACTTTAACT	CAAGCCGGGGCGTACT	PSLVTQAETDVAITSEEVEAVLSGLNVQSAP
				AAAATCTGTAACA	ACCAATTGATTTCGGATT	GSDGLSYRFWKGLDPSGRLLSCLFEIVRRHG
				CTAAAATTGGAAT	TGGCGGGGCCCATCAC	RIPGAWPTCSVILLCKDAQGDVQDVGNWRP

				ACTGGATTACTAT TGGATTACTACTG GACAATAACCTGA TCTACAACAACCA ATTGGATTACAGA ACCAAATTACAAT CTCAACTACAAAC AAAAAGTGAGTCC CCGGCCGGGTTTC TATATTACCCAAT TACTGTTTTTACA ATTTTAAAATTTT ATAGTATTTTACT AATTTTTCCGCCC CGCTAGCACTTAA ATTGGCACCCCCC CCCCCSTCCAAA AAAAAATAGATA ACCCTCCACCACC TAAACCCCGGTCA CCCCCTCAGAACT AATTGGAACCCGG ATCTTTTCCCTCTA	GCGCCTCCGCCCCAC ACACAAAAACACTTTT AAACTCTTCGGTTCCCC AACCCTACAACAAAG CGAGCGGCCCTCTTA GATCCAATTTTAAAATT TTAAATCAGTGCCTCA ACTTTTTTACGTGTTGT GTTTTTTGTATTTTTTCC CACAGATTTTGTATTTT TATATTATATTTTATAT ACACAACACTATTTTAA TACACTACCTTGCACTG TCCCCTTTTTGTAATT ATCACCTTTTACCTTTT ATGCCGCTCGCTGAGCT CCTTGACACGGATGACC AGACAAACTTTTATAA AATTATAACATTGTTTT AATTGTCGCGGGGTAT CAGTGGCGCCCCCTTGC GGCCGTGAAAGCCCTT TTCACCTTAATTCGCCC	ITICRTLYKLYAAVIARRIQTWAKQGGVLSR LQKGFMPVEGVFEHVFMLDTVLSDAKLRR KNLLAVFLDVRNAFGSVRHECLLKVLRHFD APHYLVELVRDIYTGATCRVRSSVGETGDIP CDRGVRQGYPLSGILFNLVTEVLIPGLSAGN DGYRMACLGGKLTQVLAYADDLVVVTENR DQMLRQLGVCEEFGRWAGLAFNQRKCGLI GWRTLRGRRVALEDPLLLNGVEIPLLRPGE HYKYLGAMTGVMSPRTGSQLIKDFRARLQ RLFTSFLTPHQKLIALKRFLPSLSFHRLVRPI ARSELIALDRRVRECLRVAFRLTKPSCQAVF HTPTDMDGLGVPSVCSSESSILTIAQGFKVLTS PDGTVSATASARVKLYAAKFGGLTEAGPSD WARYLSGDDVNGNSTRKPGANLPSGLWTR VRCASRQLGAVWRVCPENGITVRVRNSVIT SRDRRKLIRSFHDCSNQQWKEQWMQHPNQ EKTAAAHMAYADANRWVKQPSVMEPHTY FFALRARLNLLPTRVSRAIYSRDQHPDILCRR CGASVESLPHVLNHCPPNMSIILGRHNLVLQ EVLNAVDKTQFKEISVDRTVPEHMSETGEA LRPDIVARRNDGSVVVVDVACPFQKANFD EAAKRKLLKYDKLCCNIAASTGKPVECHSIV
--	--	--	--	--	--	--

				TCAAATTTCCGGT TCAGTTTGGCTTA ATTACCTAAAACC CGTCTTTATTTTAT CCCTTTATTGCCC CTTAAACCAAATT ATTCAAATTGAAA TCCGGCCAAATTG CCTTGCTGAACCA CCATTTTTGTTTTT CTTATAAAAGTAA ATTTTTCAACAAC CCAAAATCATAAA ATTAAAACTGTTA TCTGATTAGTCTA AATTCTGTTATTG ACAAAAACCTCCA TACCAACCTTAAT CTTATCATTATCTT CACCCAGACTGCT CAATCTAGTTCTT TTTTCAATCAATA AGGAGGTGTTTTTC	CTTAGTCCAAATTTTTC CCCATTGCCTGTAAAA GTGCGTTGCCGTCGACT CGAACTCGTCCTTTTTA CCGCCTCTCTGTTACTG AACTATGACCAGCTTG GCTGTTGAAGTCGGCTT TAATTCGCCGGCTTCAC ATTATATTTTTTTGTTG GGTAATCTGTTTTTTTT CATATATAAAACCATTC GGCTTAAAATCACCAA TCCCCATCATCCACCCT GGTCACCCATTGAAAC ATCTCTTTTTAGTCAAT TATTTTTTTCAGATTACC CCTCCCAAATTGTCATA TAGTTTAACACCCCGTT TGCCAAGTTGGTCTTTT CCCCATCCCCTTCTGTC CTTCCGGTAATCCCATA ACATTTTATTCTTTAAT TCGGCCCCATAAATAG	VGS LGSLAEGLSTSLRALGITDFARSKLVAC HQQ (SEQ ID NO: 1452)
--	--	--	--	---	---	---

				TTTGGTCTGTTGTT TTATAGTTTACCG TTATAATAGTCTG TTGTCCTATAGTTT TCCGTGGTGTATT TCTGGTCTACTAT AGTTGGTGCCTCT AAAACATTATTAT ACCGAAGAGTGCT AGCGATTTCTATT ATTAAC TCACCCA CCAGTACTGCTGT GCATCCACGCAGC ACCACGCCTGCCA TCACGTACAGTGC GCTTGCTGCCTTG TGTTTTGTTTTGCA GACACCACTGGAC GATAG (SEQ ID NO: 1207)	TGTAATCTTAAGACGA AGTCCCCAGAGGCCCG GTCCCCCCTTCTTTGT CATGGACCGGGCCAGC CCCCCTGTTACCACAAC AACCACCATTTTATTC TTTCTTTCTTTTATTAG TATTTATTTATATTGCC GCCCAACTGTCGTGTCG TGGCGCAGGGGGGTTC CCTCTGTCGGCCCAGTG GACGACCGTCTAAAAA ACAGGCACAGGCAGCA GGCACTTATCACCCGGT ACCTCCGGGTACCGGA GGCTGGTTTTTCAGCGCA CTGTACGTGATGGCCA ATTTTATTCATTGCATT TTATCCGCGTCGTGGTG TTTGCGTGGATGCATTA ATAAAAGATGAAATTC C (SEQ ID NO: 1330)	
--	--	--	--	--	--	--

R2	PERERE -9	BN00 0800	Schistos oma mansoni	ATCTCACGTTTTA ATTTATTTTTGAA CTACTGCAGTCTG AGTGCTTCTAACG ACCCGAAGGCTCA GAAACTACCCACT TCTTGAAGTCTA CTTTTGTCTGTTA TCCACAACAACAG TTGTGATTCTATTC TCCANATATTCCT TGTGCTTTTGTCA ACATTATTCTATA CCAACTGTACCAC CTACTTCTTCATCT CACGTTTTAATTC TGGTCTTATTTTCT CATCATTAGTCAC GGAGAGGGCCTAT GAACGGTCCGTGA CGCGAAATTCTAT CCGCGATTTTCGAC CTCTCCTGCTAGT	TAACGGCTGAACGAAT AGCCCCCTTCACTCTTA GACATTCCCCCACTGTT GTTGCTTATCTTCATGC TCTTGTGTTAATTGACT GCTCTCTTCTGGGTTGA CGTCTGATTGTCTCTCT CTCTTTCCATATTGCTT GCTCTGCCCCTTACTT CCAATAGTTGTCATATT ATGTCTTTGTTTACTTG CCATGTCTAACGACAA TTACTTTATCTACCTTA GTTTGTCTCTTGGTTT CGATTGCCTTCATATGT TCATGGCGGAATCTGA TGTTTATAATGACTATT CCTATTACCACCACTAC AACTACTATTATTATT TCATTACTATTAACATT ATTATAAACATTATTAC TATTATTATTACTA TTATTACTTCTACAATT	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSS STCLICFAIFPTHNILLSHATAIHHISCPPTPVQ DGSQQMSCVLCAAFFSSNRGLTQHHRHRHIS EYNELIRQRIAVQPTSRIWSPFDDASLLSIAN HEAHRFPTKNDLCQHISTILTRRTAEAVKRR LLHLQWSRSPTAITTSSNNHTITDIPNTEARY IFPVDLDEHPPLSDATTPNASTHPLPELLVILT PLPSPTRLQNISESQTSHESNKNSMHTPPTYA CDPDETLGATPSSTIPSCFHSYQDPLAEQRG KLLRASASLLQSSCTRIRSSSLLAFLQNESTL MDEEHVSTFLNSHAEFVFPRTWTPSRPKHPS HAPANVSRKKRRKIEYAHIQRLFHHRPKDA SNTVLDGRWRNPYVANHSMIPDFDCFWTT VFTKTNSPDSREITPIIPMTPSLIDPILPSDVTW ALKEMHGTAGGIDRLTSYDLMRFGKNGLA GYLNMLLALAYLPTNLSTARVTFVPKSSSPV SPEDFRPISVAPVATRCLHKILAKRWMP LFP QERLQFAFLNRDGC FEAVNLLHSVIRHVHT RHTGASFALLDISRAFDTVSHDSIIRA AKRY GAPELLCRYLNNYYRRSTSCVNRT ELHPTC GVKQGDPLSPLLFIMVLDEVLEGLDPMTHL TVDGESLNYIAYADDLVVFAPNAELLQRKL DRISILLHEAGWSVNPEKSRTL DLISGGHSKI
----	--------------	--------------	----------------------------	--	--	--

				GGTCCCCGAAGTA CGGTTCTCTGGC CTGTCAGTTGTGT TAAAACTATATAA TAACG (SEQ ID NO: 1208)	AATATTATGGCTACTCC TCTCAGCACACCAATA AAATATCAATCAAACA TCTCAATTATATCCACC TATTAAACTCTCTCTAT TTCCCCTGAGTTATAAA CTTACAATTCAGTCTAA CCGAATATCTCTCTTTT ACAAATCTTAAGTATGT AATTTTGTGCCAAACCC ATTTGGGTCTGTACAAT TTGATACTTAAAAATA AATGTTATTAGCC (SEQ ID NO: 1331)	TALSQTEFTIAGMRIPPLSAADTFDYLGIKFN FKGRCPVAHIDLLNNYLTEISCAPLKPQORM KILKDNLLPRLLYPLTLGIVHLKTLKSMDRN IHTAIRKWLRLPSDTPLAYFHSPPVAAGGLGI LHLSSSVPFHRRKRLETLLSSPNRLLHKLPTS PTLASYSHLSQLPVRIGHETVTSREEASNSW VRLHSSCDGKGLLLAPLSTESHAWLRYPQ SIFPSVYINAVKLRGGLLSTKVRRSRGGRVT NGLNCRGGCAHHETIHHILQHCALTHDIRCK RHNELCNLVAKKLRRQKIHFLQEPCIPLEKT YCKPDFIIIRDSIAYVLDVTVSDDGNTHASRL LKISKYGNERTVASIKRFLTSSGYIITSVRQTP VVLTFRGILDRASSQSLRRLCFSSRDLGDLC LSAIQGSIKIYNTYMRGT (SEQ ID NO: 1453)
R2	R2- 1_BTe	.	Bombus terrestris	TCAATAGTTACTC GGGGGCAGGCGG GATATTGGTCTTG CCTTGCCCAAGTC ACACTCCTACCTC CTCGTGGTACCGC CGGTAACACGCGC ACGTCCACATCAG CGAGGGGCGTACT	TGAAAACACGATAACG ATTATGAATCAAATA AGAAAGTAAACATCCC AGAAATTGTCTACGTCT TATTTGTTATCTATTTA TTTGTTTA (SEQ ID NO: 1332)	IAKFDNNTNSASDAAPLSPGGAVADLSASEG TTDNDQAMSPAMSLXTVPLVGNRVACPXC EKREANLFFLNLSDLDRHLTQHHPDAPIXWS CIDCAKCFPKLHGARCHIPKCGGASSQARTG EFQCEACPMSFGSRRGLSTHERHAHPAVRNI KRRGADPPEENTKSWKVEXVARLKLWEIF KDHKYPNKEISKFLTTKTVDXKXQYQRKKLN LIGXESPQEATSLATEGGCDLVSSGNASFGS PVGRNENEEELIHEWKLKSLKNEINKPTEVPPI

				CCCCCGGATGTGG CGGCGCGTGGCTA AACGGAGTGTGGC GACGAAGGAGCG AAAGACTAACAAC TATAACGGTCTTC CGTAACGGCTACT TGGAGCCGTGAAT AATGGAGCCTATA TTAAACCCTGGAA CTCGTTCCTTCGTT CTGTTGACGACTG GAACGGCAAAGG ACATGATTTGGAT AACAATTGGAAAC TTAATATCGAATT CACAATTAA (SEQ ID NO: 1209)		LKEVYNRLMLIWEEHQDDRDSLTESLDHFIR TALYELINKINKNQTDLKTAKRAAKTKSPKN NRNSRKRF SYARCQELFHECPRRLADAVVN NDQAYLEPARQPPGSEEVRGLYEKLWGQV GSTYVPAPVTRVPKLSLSEIFPPIAAEDVGER IGKIRKKAAGPDGLQRDHLTIPGLPIIMAKI YNILVYCSYFPSAWKENRTTLIPKINKPCSLV ENWRPITISPILGRIFSSIIDGRIRRGTVLNM QKGFTSENGCKINIELLSALNYSKRNSGGIF TIVDISKVFDTPHAALKPCLAKKGVPA DLIDEMYKNVKTITKTDGGVEIMIRRGVK QGDPLSPLLFNLCLEPLLEEIEEQASGINVSE HRKVSVLAFADDIVLLGADAREAQHQINVL TDYLQSLMMNLSIEKCQTFEVVAKKDTWFI KEPGLKIGNQIMPTVDPDEAFKYLGA KGVHCGVIVPELLSVVKRVRKLSLKPGQKV ELLTKYIFPRYIYHLLVSPPSDTVLKLLDSEV RQEVKTILHLVPSTATGFFYTPKACGGIGIPR FEHIIKLGTLKSAIKIANSIDPAVAGLIDDAI KKLKQTANSLRINWPASLEDIEKARKRLRKE HISQWADLKCQGGQGVPDFIKNKTGNLWLED HSLKPSRLIDALRLRTNTFGTRSVLARADK NIDVTCRRCRAQPETLGHILGLCQHTKGLRI
--	--	--	--	---	--	---

						KRHDEVKSLLEGRLKSKKNNEVFVEPTIKA GGSLFKPDLVIKNGERVLVVDVTVRYENKN YLALAEKEKIEKYRPCLRALKEIFNAKGGEI LPVVLGSRGTITPNTEKVLKRLGIANNDIKTI LLNVLRSSIELCNIFIDD (SEQ ID NO: 1454)
R2	R2- 1_Crp		Crocodyl us porosus	AGGCGTCTCCTTT AAGGGCAACGGTC TGGTTACGCGGTC GCAGCAGKCTTGC KCCAGGTACCTCC WCGTGGTTCCCGC CGGGTGCCGMAG MCCCAGGSCTGTC GGTAGCTCGATCC TGGCACAGTAMG GCCAGGGGAGTKC TTCCTTGCTGCWG GTGCCCCACAAGC GTKTGGCAGSMC MCCTGCTTCWTCG CAAMAATMASAG KGTSMT CAGTAGT CGGCCCCGCCGCT	TGAATCCCCTCTGGG GACCCCCAAAAATTAG AAAAACCCAAAACAGT TGTGTTTAAGTGTGTTT TTGTTTGTCCCTTTGGC TTCACCTCCAAGTTGCG ATCCCCCATCTCCCCTG CGCTGTCTTTCTGAATG ACCAGTGGTGTGAGG CTGGTGTGACCTCGGTC ACCTCCAAGGCCAAGT GCCCTGGCCCCGAGTA GGACCAGGTGGCCCAG CTTGCTGGGCACCCGTC ACCGCCCAGGGCAAAT GGAAGGGATCTATCCT GACCACTACCAGGCTA AGTGTGGTGCTGCCTA	VPPGAEARGRYHHPRXEXARQGEPPSXRVF LVXLPDSNPPCPICGDHVXXXSVLALHCVE GHXWAXVQYQCTHCGILCHIPRCQGRVXE XTGKDXXCPECPASFDEKAGLSQHKRHTVT XSERVAGXLLRAXLRHGCWSVEEEETLTR LDAMFXGARNINQLIAAEXVSKMXKQISDK WRXLXLXPEQTTXGGXAESASVXXESMT PEMEAQSPAXPPGKIRKIFTGQDGHAGGXA WENQEDFWTRWARRWLKRGQXLSDKVQ EVLGXWVEGQPRIXAWVDXVSLDVLTLFL GVPPGPQRAPSKKGPXEGGKPTSWMKNRAI KWGTFLRYQHLLFGANRKLAAAXILDGAXR NQXTLLLEEVSXQXYXGKWEAEPPFEGLGRF GXXRDVDSFAFEALITXEEAVKHMMXMAX XSAPGPKLTLRDLRRADPEGDALAEFLSL WXITGVVPDRLKEXQXVLIPKAVDFEKLRQ LGNWRPITIXSIVLQLXSRVLTARLTAAPIX PHQQGFISAPXCAENLKXPELIXRKVKXDRR

				AGCCAAAACGTGT CGCCACSCAGTTA CAGATGGTGCCTG TGCTGACCKGKKG CCCCGTGGGCTCG GWGGTGCAGGGG WGGCCGGCCCAT GGCTGGGCCAGAC STGGGCKTGGAG CCCGCTCCCAMCC CAGAGTTCCCCCT CAAMGCCGCCAG GKCAGMAMCAKC CGGGGAGKGMCA CGGCCCCGGCCGT GA (SEQ ID NO: 1210)	GCCTGCCGTAAGGTCA AGCGCCCTGCTGCCGCT CGGGTAKCAGTCCTCG TTCACCTCGTCCCTCCTA GTACCCTCCGCMCTCTG CGCTTCTGCTATCCACT GTGGCCGGWGATGCCG AGGWTGSWMCAWCCT CGACACCTCCAGGGCC AAGCGCCTTGGCCMCA GGTAGGACCTGGCACC TGCCCAGGGGGCCAGA CACTGCCTGCGGCAAG GGAAAGGGAGCCGTCC CTGACCGTTACCGGGCT TGATGGTGCTGGCTAG CCCGCCGTATGTCAAG CACTCCACAGCTGCTCG AGTTCGCTGGCTTCACC TTCATCCCWCCTAGTG CCTTCCGCCTCTGCGCT ATTTTCGTCCCGACTCG TACCTCCCCACCTCTGC	PLGVAFVDXARAFDSVSHDXISWVLKAKGV DQHIVNLIEDSYQKVTMRVQVFSGSTPPISIK XGVKQGDPMSPLLFNIAMDPLIXKLKTVRQ GVKVGASLTTLAFADGLXLLXDSWEGMQ HNITTSXTPGRACNTTSRHRPGLLQPHGPTS ATXKMXGVLLESXMRLLYGEQLRGLEDXR PXXHDAXARRADTISGLEXRSLGWDXQTRF GYATXLLAREDGDCXAQTNAEALCWXSXP FPGCAXRPXYANXGWVASEALDSMSRRXV KEWFHLPACTDXLLXSRHRDGGGLGLRLAR XXLAAXVRRPIRVATSSDEVTRKVSYACGIS DEVERLXLAXGGDXSNVPRFEDPXAPKSXX VQGPHEAAQETPRVVRTQAIPWPSNWRAEE HSKWAQLSCQGERVELFCNDPVSNGWINSR GQLAERLWIMALKLRSNIYPTREFLGRGQA GTNIGCRHCTHPRET LGHILGICPAMQEARIL RHNKLCKILAAEGKNCEWTVYYEPHLHNA AGELRKPD LIFVRDGTALVVDITVWYEGGP ATLLSTTAEKATKYLDLNTQIQELTGAEQVT YFGFPIGARGKWHADNWRVLSELGLSNSRK ERVTRLLSWRALLGSVDMVNIFVSKHRQEN LLDEHCTPAEQVVSSYAS (SEQ ID NO: 1455)
--	--	--	--	--	--	---

					GCTACMGCTATCCCTG MAAGGACCAAGTGGGC AGGGGSGTTCGCCCCC GCCCCGAGGAGGCTCG GCGTATCCGTGGCKTC WTGCCTCCACCGTCTT GTGCCGCTAGAGGGGT ACCTCMGAGACCGGCG CAACACGACCTTGACG STTAGACAGTAGGGTG MAACAKCCCTGCTGCA GGCSTGAAGGGCCAAA CGGCTGTGCCATGAGA GGGGAACCTTGAAGAC CGGGSCAGTCAGCCAG TTAGTCAGTTGGGCGA AACAATCCCAGCTGCA GGCCCAMMAGGGCTGT CAGGTGAGGGGGTATC CCCAWCCACCCCCCGC CGCCGACTACGGAGGC AKGAAGTCCCTAGTGA CTTCKGACCCCCACGTC	
--	--	--	--	--	---	--

					TTGTGCCGGGAGAGGG GAACCTTGAAGATCGG GGCAAGCCGCACTTGA TAGTTAGCCAGTCKGG TGAAACAATCCCAGCT GCGGGTCCGAAAGGGC CGACTWCCAGGCGAGG GGGGCCTGCGGAAAMC CCCCTCCATGGTACGG 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 AGTCATTTTGTGTTGCA TGGGGCATSCTGAGTCA CTTAACCGAAAGACTC WAAATAACTCAAAAGA	
--	--	--	--	--	---	--

					GGKAMCCTCTGSGGTT CGGTAAA (SEQ ID NO: 1333)	
R2	R2- 1_DWi	.	Drosophi la willistoni	GAAGCTGGGTCGG ATGAGCGCAGAA GGGGTGTTCTTTG GAACACTGTAATT CATAAGTCGTAAG TCTGATCAAGTCG ACTCGAAACCTCC TCGTGGTGTTTCC TGGGTGCTGTTGA GTTCTAGTCTCT AGGTTCTTTTCAG TAGCTAA (SEQ ID NO: 1211)	TAGATGTACTAACCTCT AGCTTTTCTTATACTTT TGCCTGCTACCTTGGCA TTACATCTAAAAAGGT ACAAACATCGCATTGT CATAAAGAGGTGGTTT TAGTACGTAGGCGCTG TGGGACTTCATTGTCCC GGTGATGCAGTGAATC GTGCATACGAGATTGT CCAGTAGTTGGTTGCTC GTATCTTTAGAAGATTT CCTTCCTCGGCGATCAA AAAAAAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 1334)	FERRSNSWGYRPLEPRSVGTESNNNSPRSNI TITSATSRPGDQPREAIAVVNLAGEIPCAVCG RLFNTRRGLGVHMSHQHKDELDTQRQRED VKLRWSEEEAWMMARKEVELEASGNLRF NKKLAEVFTHRSSEAIKCFRKRGEYKAKLE QIRGQSTPTPEALDSITSQPRPSLLERNHQVS SSEAQPINPSEEQSNWEIMRILQGYRPVECSP RWRAQVLQTIVDRAQAVGKETTLQCLSNYL LEVFPNPNEPHTIGRSNLRPRTRRQLRQQE YAQVQRRWDKNTGRCIKSLLDGTDESVM NQEIMEPYWKQVMTNPSTCSCDNTRFRME HSLETVWSAITPRDLRENKLKLSSAPGPDGI TPRTARSVPLGIMLRIMNLILWCGKIPFSTR LARTIFIPKTVTANRPQDFRPITVPSVLVRQLN AVLASRLASKVNWDPRQRGFLPTDGCADN ATLVDLILREHHKRWKSCYLATVDVSKAFD SVSHQAIKTLQAYGAPT NFVSFIEEQYKGG GTS LNGAGWSSEVFIPARGVKQGDPLSPLLF NLIIDRLLRSYPREIGAKVGNTMTSAAAFAD DLVLFAETPMGLQTLDDTTVGFLASVGLSL

						NADKCFTVSIKGQAKQKCTVVERRSFCVGE RECPSLKRTEEWKYLGRFTADGRARYSPA DDLGPKLLRLTRAPLKPQQKLFALRTVLIPQ LYHQLTLGSVMIGVLRKCDRLVRQFVRRWL DLPLDVPVAYFHAPHTCGGLGIPSIRWIAPM LRLKRLSNIKWPHLEQSEVASSFIDDELQRA RDRLKAENVQLCSRPEIDSYFANRLYMSVD GCGLREAGHYGPQHGWVSQPTRLTGTKEY LHGVKLRINALPSKSRTTRGRHELERRCRAG CDAPETTNHILQKCYRTHGRRVARHNSVVN AVKRGLERKGCVVHVEPSLQCDSGLNKPDL VGIRQNHIIYVIDVQVVTGDHSLDQAHQRKV ERYDRADIRSQMRRFFGATGEIEFHSVTLNW RGIWSGQSVKRLLIAEDTKLISVRVAV NGGVTSFKYFMYCAGYTRS (SEQ ID NO: 1456)
R2	R2- 1_Gav	.	Gavialis gangetic us	AGGCATCTCCTTK AAGGGTAATGGTC TGGTTACATGGTC ATAGCAGGTTTGT GTCAGGTACCTCC CAGTGGTTCCCGC CGGGTGSCAMAGC	TGAGAAGTTGCGAGTT CTTATGCAAGTTGAATA CCACTCTKGKGACCCC AAAAAAWWAAACCCC AAAACAGTTGTGTTTA AGTGTGTTCTTGTTTCGT CCCTTTGGCTTCACCTC	PAAPRAWGAVEAGPWPGRTRAVEPAPSPES SPSEAARAAPAGEGHGPGHESPSVQRPEADT TAPGVSAPTREGEPPSTRVFLVRLPDSNPPCP ICRDHVGKPSALALHCVESHAWADVQYQC THCKKVSANKHSILCHIPCCQGRVPEWTGK DWACPECPASFNNKVGLSQHKRHVHPVTR NVERVAGSLSRAGLRPQTRRGCVSVEEEET

				CCCAGGGCTGTCG GTAGCTCGATCCT GGTACAGTACGGC CAAGGGAGTTCTT CCTTGCTGTCGGG TGCCTCGCAAGCA CKTGGCAGCCCCA ATCGCTTCATTGC GAAAAACACAAA CGTCCTAAGGGGA TGATCAGCTAGTC AGTTCTGCCGCTA GCCAAAACGTGTT GCCACCCAGTTAC AGATAGCGTCTGT GCTGA (SEQ ID NO: 1212)	SAAGTTGCGATCCCCC ATCTCCCCTGCGCTGCC TTTCAGAACGGCCGGT GGTGTCGAGGCTGGCG CGACCTCGGTCACCTCC AAGGCCAAGTGCCCTG GCCCCGAGTAGGACTG AGTGGCCCAGCTCGCT GGGCACCCGTCACCAT CTGGGGCAAATGGAAG GGATCTGTCCTGACCAC TACCAGGCTAAGTGTG GTGCGGCCTAGCCTGC CGTAAGGTCAAGCGCC CTGCTGCCACTCAGGTA TCAGTCCTCGTTCACTT GTCCCTCCTAGTACCCT CTGCCTCTGCTCTTTTG CTATCCACTATGGCCAG TGATGTTGAGGTTGGTG CATCCTTGGTCACCTCC AGGGCCAAGCGCCTTG GCCACAGGTAGGACCT	LTCLDAMFRGARNINQLIAAEMVTKMPKQI SDKRRQLGLCPEQTTLGGDAESTSVVEEES MTPEMETQSPINPPGKIRKILAQRARRWLKK GQGLSDKVREVLGAWVEGQPRIHAWVDSV SLDVLTFLFLGVPSGPQRAPNKKRPKEGGKPT SWMNKCAVKWGTFLRYQHFLFGANRKLLV AIVLDGADRNOCTLLLEEVFQAYREKWGLE EVLRAYRGKWEVESSFEGLGRFGVRRDAD NFAFKALITPEEVVKHMMAMASKSAPGPKD LTLRDLRRADPEGDALAEFLSLWLITGTVPD GLKECRSVLIPKTVDREKLGQLGNWRPITIG SIVLRLFSRVL TARLAAACPINPRQRGFIAAP GCAENLKVLELLLRKRKRDRQPLGVVFVDL ARAFDSVSHDHISWVLKAKGVDEHIVNLIE DSYQKVTTTRVQVFNGVTPPISIKTGVKQGDP MSPLLFNIAMDPLIAKLETGQGVKVGASL TTLAFADDLVLLSDSWEGMLKNISILEDFCN LTGLRVQPKKCQGFFLNPTCDSFTVNNCEA WKIAGREITMLGPGESTRYLGLNVGPWVGI DKPDLGTQLSSWLERIGTAPLKPMQKLSLLV QYAIPRLNYQADYAGIGRVALEALDSMNRR KVKEWFHLPACTSDGLLHSRHRDGGGLGLPR LAKAIPAEQVRRRLIRVATSSDEVTRKVSAC
--	--	--	--	--	--	--

					GGCACCTGCCCAGGGG GCCAGACACTGCCTGT GGCAAGGGAAAGGGAG CCGTCCCTGACCGTTAC CAGGCTTGAGATGGTG CTGGCTAGCCCACCAT ATGTCAAGCACTCCAC AGCTGCTTGAGTTTGT GGCTTCACCTTCATCCC ACCTAGTGTCTTCTGCC TCTGCACTATTTTCATC CCAACCTCGTACCTCCCC ATCTCTGCGCTCCTGCT ATCCCTGCAAGGACCA AGTAGGCAGGGGGGTT CATCCCCCTACCTGCAG GAGACTCAGCATATCC ATGACTTCTTGCCTCCA CCGTCTTGTGGCGCTAG AGGGGTACCTCAGAGA CCGGCACAACATGACC TTGACGGTTAGACAGT AGGGTCAAACAACCCT	GISDEVERLWLARGGDMSSVPRFEDPEAPRS PGVQGPCEAAQEIPSVVRKLAIPRPSNWRSK KHSKWAQLSCQGEGMELFCNDPVSNGWNN SRGQLAEHLQIVALKLRSNIYPTREFLGRSQ ASTNVGCRHCTHPHETLGHILGICPAVQEAR IIRHNKLCKILAAEGKKCEWTVYYELQLLN AAGELCKPDLIFVRDGTXLVVNVTVGYEGG PAXLLSTAAEKATKYLDXNAQIQELTGAEQ VTYFGFPIGARGKWHADNXRVLSLGLSNS RKERVARLLXWRALLGSVDMVNIFASKHR QENLSDXALSPS (SEQ ID NO: 1457)
--	--	--	--	--	--	---

					GCTGCAGGCCCAAAGG GCCAACAGCTGTGCCA CGAGAGGGGAACCTTG AAGACTGGGGCAGTCT GACCATGCTGGTTAGTC AGTTGGGTGAAATAAT CCCAGCTGCAGGCCCA AAAGGGCTGACAGTCA GGTGAGGGGGTATCTC CATCTGCTCCCCACTGC CAACTACGGAGGCATG AAGTCCGTAGTGA CTTC TGACCCCCACGTCTTGT GCCATGAGAAGGGAAC CTTGAAGATTGGGACA AACCGCACTTGAAAGT TACTCAGCCGGGTGAA AATAAGTCCCAGTTGC GGGCCCCTCGGGGCTG ACAGTCAGGTGAGGAG GGCTGCAAAGCCCATC TCCTGACTCCAGAGGC CTGGCGTCCTAACCGA	
--	--	--	--	--	---	--

					CTTCTTGCCACCAATGT CTTGCGCCAGGAGAGG GCAACCTTGAAGATCG GGGCAAGCCGCACTTG ATAGTTAGCCAGTCGA GTGAAACAATCTCAGC TGCGGGTCCGAAAGGA CTGACTTCCAGGCGAG GGGGGGGCCTGCGGAA AACCCCTCCATGGTAC GGAGGTCTGGCATCCT AACCGACACCTTGCCA CCAATGTCTTGTGCCAG GAGAGGGGAACCTTGA AGACTGGGGCAAGCCG CAGTTGATGGTTAGTCA GTCGGGTGAAATAATC CCGGCTGCACCCTGCTG TGACTGCTAAGCCCGG TCCCAAGGGGCATGA GGCATGTGCGCTGAGA CGGGAGGGGTGACATC TGGCGATCAGCACAGC	
--	--	--	--	--	---	--

					ACAGACTGAAGGGAGG CACTTGCCGAGAATGC TTCTGAGGCCCCAGACT TGGGGTGGTGCAGCTTT GTCTCGTGTATAGTACA GCACCCTACTGCTCCCT TTGGGCAGCAGAATTT GTCCTGACCTCTTACCC ACCCGAGTCTGCGCTTT TGTTCCACCTCGCTGTC TCCCTGCTGTGCTGTTT TTCTCTCAAGTGGGTTA AATCTCAACATGATTAT CTCCACGTTTCCGCTC AAGGGCAATGCCCAAC ATGACGGAGATCGTTG GTGCATGGTAGTCACG AGACCATCCGGACCCT CCAGTGGTCGCTATAGT CATTTTGTGTTGCATGG GGCATGCTGAGTCACTT AACCGAAAGACTGTAA ATAACTCAAAAGAGGT	
--	--	--	--	--	---	--

					ACCCTCCGGGGTTCGGT AAA (SEQ ID NO: 1335)	
R2	R2-1_IS	.	Ixodes scapulari s	GTTCCAAAGGAAG GCACTCCTTTGGT TCGTGATGAGATG TTCATGGTGCTTG CCTAGCTGGAGAA ATCCGACTCACAC CTGCACGTGGTCC CTGCCGCCTGCCA GTATGCCGAGGAA ACGGGTGCAACTT AATCCGTGGATAC TGGTAGCAACGTG AGCAACGGTACGG TCCTTCGCGGACC ACCCTGGGCGTTC GGGTTGCCAGCCC GTTCCGCCGAAAT ATCTTGGCCCTGA AACTAAAAGAAA A (SEQ ID NO: 1213)	TAGTGTGACGGAGTCC TCAAGCCCCCACAAGT GCCTGCCAGGTGGCAG GAAAGGGCAACTACTG GTGAGCGACCCAAGCA AGGCGGAGCCAAGACC AAGCTGGAGCCAAGAG CAACTCCAGGAGGCAG GGGTGGATATCAAGAG CAACCCCAAGGGACAC AGACCACGGGCAACTA CTGGTGAGCGCCCAAG ACAGGGGTGGATATTA AGAACAGCCCCACAAA GTGTTACCTATATTAAC AATAAAGTTGAAGCCT CAACCACGCATTGCGG GTTAGATGGCGTGGCTT GGCCCGCCGCCATGAT GAGCTGGAACCCTCCA CCTGGTGGGCCGCACG	MQCTSRLADAPRFARVGVEGEGVGASGNG TDAQLWYGCTGCDEAFSSLRGLRIHAAQKK HGNQDGLLRLPAGRPRKRRVGKSTTAGASD RVTTDPVPAPVPESPGLLPGLPGPSLPGCSDL PPGVLPGGWSASPGPLSWPPSLDAGPLPGPS RVSPGPSRSPGKPTGPPSLDAGPLPGPSRVS PGPSRSPGKPPGTPEPLPGSPGGRRGVSPGQ PGSRTPSSSAGAGHFVCPQCSRAFSKIGM SQHQKHAHLEEYNAGINITRTKARWDPEET YLLARLEATLNPDKNINQTLHAALPRGSC RTLESIAHRKQAAAYRDLVTSLSARESSEA QHVPDRPLETPEPQTPANPQRDSKQAVIEAL QSLIGRAPPGSFQGARLWDIARQATRGTNIL PLLNSYL RDVFTLPTKPTRKKPAVRPARSRR KQKKQEYARTQDLFRKKQSDCARAVLDGP TSSSVPGTGAFLQTWREIMTGPSPALEAPPLP TRGEVDLFFPATAQEIQSAEIAVNSAAGPDG FSARLLKSVPALLLRVMVNLLLLVRRVPAA LRDARTTFIPKVPDAVDPSQFRPITVASVLQR LLHRILAKRALEAIPLNFRQRAFQPVGDCAE NIWLLSTALNEARTRRRPLHMASVDLTKAF

					AGACCACCGGCTCTTTC TACTAAGGCCGGTCTCC GTGACTGCGGTTGGGA TAAACTCCAAGCACTG AGCGGTAAAAAAAAAA AAAAAAAAAAAAAAAA AA (SEQ ID NO: 1336)	DRVTTDAILRGARRAGLSGEFIGYLKELYTT SRTLLQFQGESLLVEPTTGVRQGDPLSPILFN LVLDEYLSSLDPDISFVSGDLRLDAMAFADD LIVFASTPAGLQDRLDALVEFFDPRGLRVNV KKSFTLSLQPGRDKKVKVVCQIFTIGGTPL PASKVATPWRYLGMTFTPQGSINKGTSEQL DLLLTRTSKAPLKPQQRLVVLARNYLLPRLY HRLVLGPWSAALLLKMDTTIRGAIRRWMDL PHDTPLGFFHAPVTEGGLGINS LRASIPAMV LQRLDGLHFSTHPGAEVAIQLPFLTGLHRR EAAAQYQGQRLLSKADVHRMWSARLHGSC DGRPLRESKRVPAAHRWAAEGTRLLSGRDF ISITKLKINALPTLERTSRGQHKDIQCRAQC AVESLGHVLQACHRGHRGRIRRH DN IARYV CGRLTQIGWAVKWEPHYSVAGRTLKPDIVA HRGAETVVLD AQVVG TSMRLGFHHAQKKE KYSLPDLLHQVCEGRRDAARVSTITLNFRG VWAPESAQDLKSLGLTDNDLKLLTVRCLQG GAQCFRLHRRMTTVVKATGDEANALPAHS GLPPTQLGGRTLGPSAHNQSARTT (SEQ ID NO: 1458)
--	--	--	--	--	--	---

R2	R2- 1_MLe	.	Mnemio psis leidy	TGGGGGCCCCTTG GACTTGCTCCCTG GGGCAGGACACC AGTGAAAGGAGA TCCTCAAGACAGG ACAGAGAGACAG GCACACCAACCCT TCGAACCTGAGGA GACACCCAGATCT GGTTACCCCATTC CTGTAACCATGGT TGCTCTCCGTGCG CTGTAAAGGAAAC CCAGCCTAGTACC CTCGGGGAAAGGT TGCGTAGTACTTA GCAGTGTGCGGAT CAAACCTCTACCG GTCTCTCTAGCGA TGAAAGTTTCTCC GACTGGAACCTGA GGGACTGGCTAGC CCAGCTAGCCTGT	TAGTGACGAGAAAAGT CGCTTTATCCTTACATA ACAGTGTGATAGTCAT CCTTATCACAACTGTCC TGGCGAACAAAATGGC AGAGGGATAGTACTCG GTCCAACCAGAAGGAA GCCACACGATGCCAA ACTTGCTGTAACCCACG TGAGCGGAAAACATCC TCCGAGTATAGTATGAT GGAAGGATACAGGATG TGGACCGCTCTAGGCG GCGGACGGTGTAGACG GCGACCTTACCATAGCT GCGACTGCTGTGCGAT CCGGAACGGGCCTTTCT CACTTACCTCAAGCACC GTGTCGGATCGCGGGG TGGGCGCGGGATTCTT GCTGGGAAGTTGTCTG CACCCAAGTCGTCAGC TAACTGTCGGTACTTAT	MSNTSHSKLNLKMDNKLKTSLETSGVRAD SIITRVRTSSNRGEHSNGVTYPRCEQGVAPL DTHGGICDAPPQVTVPATETDKQKKCEYCE FTYLKPRQIGTHMRKRHPQEWNDIKRTKFL SEKRQKRWLDEDFELLCIGQEEYLVLSIGK QGKGINQYIQTKYFPTLSTDAIKSQRKSRRFS EYSEKRSRELQPCNTSSDPEELPNEAVTENS LSFDPLDRDVVKKISSKDHGDQILLVQEHLI NGRYQEANTLAKAIFEKLSGKFPNLKTGDH RPGKQQTARKVGKKRVRGSGKKLSPSKQN RRELYAIVQKQWRTKKRSKVINQILTGNLN KEQSYTHTPDQLAQFWSTLFGRVSPRDDRPI NHRRSVIPELDKPLSVEEVEAALKGAKDAA TGIDGVPISHLKHLGSAALTILYNGLYVTGSI PDPWKRARTILIPKSNPPASPGDYRPISISSYF YRIYTSASKRLASAVSLDDRQKGFIKEDGIR DNLSLIDTLINETKAGSKSLFMTFMDVKKAF DSVSHYAIARSLEWAGVPDGMRSVIADLYQ DCTTDICGRSVKVTRGVKQGDPLSSTLFLNV IEMVMSNVPERLGIQFQGHRLFYLAFADDL VLLTRGPTANQKLVSLVHEQLARVGLELHP GKCKSIAIMADPKRKTTTFVDQGSSVLIGGEP VSSLGPQEWYKYLGIKLGSGGMPQGIYRDQ
----	--------------	---	-------------------------	--	---	--

				AGAGCAATGCGTA TACGATGCCTTGG CGACAATGGCGAC CGCTGCTTAGCAG ACGGAGGTTAGTG AAAGGGCGACTTG CTGTTCATAGTCA CGTGAGTCGTCTA GAAACTGCACCGA TGCTGCCCTCTGT TCCAGGAGAAGG ACAGTGGAGGACT AAATCGTAGCGCG AGCGGTGTT (SEQ ID NO: 1214)	ACGCACCCAGAGCCAT GCTCTTACCAGGCCACT TGACGGCGCTACACCG GTCTGTAGGAGGGTAT CTCAGCGACTTGGACA AAATGGATCTTGACTG CCTGAGGGCGCTTATC GAGTGACCCAGAGTAA GCTGGTAGGAAGAATC TTGCAGTTGAGAGGGC TAGTAGGGCCAGCCTC TGCTCGTCAACTGTCCT AGGCGTTAACGTGTGC CTCCCTATGAGGTAACC GTGAAGTATTACTTGTT TCCCTGTAGAGACCAT ATAACTAGCTGAGAAA CGCCCTGCGGGTTAAT ATACAGTTACCGCCGTC CTGCATCCCCGAGTTGT GACCAGCGTACAGCCA TGTACCACGGCACTTCC AGCATTCTCTGGTCTAA	LADLLAKTDSAPLKPQQRLYLRSILPKFN HRLMFERVTCQTLEGLDKLIRTHVRKWLKL PKDTPGPAFYADKGSGLGLITLRYRVPLLK LRRHKKMADSPDPVIRLIPNAEPTISLLARW TKMCSLYGKQYQHKSELSKIIRDKYWTMCD GKGLRTEVPPDTAKKTLSSLFEDRTPLKPGQ LIGAIGVRLNTLGTARNNRAGYSPEANIC DKCPGNRQATLGHISQTCPATHGRRVKRHD KIVNRIAKALKERGSVKNILTEPHLRHDKLP LRKPDLIVHTEKSVEIIDVQVVADQGISRHE DEDQQKKIVKYDVDGYKRAAYKMLGIDYG SIPCNVSAFTITWRGNLAPHSLKLASRLQFSP VLKYIVADSLVDTWGAFLIWGKTS (SEQ ID NO: 1459)
--	--	--	--	--	--	--

					GAATGTATCAGCTGGC CCGCCGAAAGATAGAG CGACCCCGCCTCTATCA CTAGAGTAAGCAGGCA CGCGAAATATATGCAG GACGCCTTGGCTCAGT AGCTCTGGTACTGAGT ATAGCATGGCTGAACG CCCCTTTAAGTTCGGTG GCCCAGAGACTCTCCG AGACTTCCCAGTCCTGG AGAGGGTACACCTTTA CCAGAACTTTCTGTGGT CGGTTG (SEQ ID NO: 1337)	
R2	R2- 1_PM	.	Petromy zon marinus	CTATTAATGGGAT GAAGAAGGGGGA CACGAGTTTGTGT GTGCATCCAGTTT CCATGGTGCATGC AGGAGTGGTGGTT TAAATGGCGAGAC TCTACAGGGCTTC	TAATTTAAGGTAAAAT CGTGGGATTGTTTTGAT GGCAATCTGCCTAGTC GCGGCCTTCCATTTTGG GTAGGCAGCAGACCCA TCTATATAACAACTAC TTTGCCTTTCATAGGGG TACCCGACCCTACCAA	MNERLTDELTTEFILSDMFLWDYPCTDQNK CYPCNLVFLDHRTWSSHMARVHPHANKTY KCRICNRTADSIHKIASHYGRACKSLIGKTN AITTTIDETLFSLHCSRGTFTKTGLGVHTRR THPTEHEAILQQNTPGRKVVRWGEEVEIMA HKEAQQKDEDINMNQLIQNSVMPHRTLEAI KGKRRNIKYKELVRTLKETTYKVENQCLVN LVLPTTSEITTTTPSEGDQPAIRAEKEQSPTAA

				CATGGCTACACGG GATGCAAGGCATC AGACATTTTGGCA CAGGCAATCCTTT TGGTCTCTACCGC AATCATGTCTTAG ACCTCAGTAGCGA CCACTACAACCAC AGTGGTGACTGCT GTTGAGTGAAGGA CGACTGAGCGCTG GATAACAACCTTC TTGCGTGGCCCAA CATCGAAGCAACC ACTTCGGAGCTGG CACAAGGCAAGA GGGCAGCCCAAG GTGTGAATCATCT CAACTTCACTGCA GGAAGAAATGCTG TGCAAGGATGAGT GTGAACGACACCA ACGGGATTGTTGC	CTTTCGGGGAAGTAAA AGAAA (SEQ ID NO: 1338)	EDLQVIINDLKSQNFSHNQALLLNSHVEKF LNRSKPIKRKDHVNQQEIDENRHRRQSKQT KYRRYQYLYHTNKKALLDEITSDRSGPSIYP TEESIRGTFVTLFESNSPPDNIPSKLKNQSCI DIVKAITLDELIKTLAIMKDKSPGQDNITLSD LRTLPIKYLLDILNIILYIQDIPQIWKQHRTRLI PKTKEELEKPSNWRPITISSIVIRLLHKILSYR LGQQLKLNRYRQKAFLPVDGCFENSALLHFII HNARQKHENTQIVSIDLSKAFDSVSHESIIRA LNRFNLSKESITYLTNIYKCNLTDIVFGSTIM RNINLKRGVKQGDPLSPLLFNMIMDELNDN LPTYIGVNVGNQKVNSMMFADDLILFAETE CGMNKLLDITTKFLDDRHLKININKCNLSRFI KYGKQKTFSVATTSSYFINNEPINPVSYVKG FKYLGIEFDPRGKRSISCNLLAAMLNKLTRA PLKPEKKVYLINNNLIPRIIHQLVLGKVTKGL LMSLDSEIRKTVKLLLRLPHDTPDSFFYTSVS NGGMGIRNLCDSVALSIINRHNKLITSDDLVI RALSQQSYTIATLKQAHIIAGSKFPSKSLNQN KWSNKLYQTTDGRGLVYCQSQTENNSWIT GNHRTIKSYNYIDMVKLRINALPTKSRNCRG TLETQCRFKCRSINNQISEETLAHILQKCDR SHYSRIARHDSLQFLATAAQKLNWEVIKEP
--	--	--	--	---	--	---

				TGACCAGGAGGTG CCAACCAAATTTG AATGGATTGACTT TGGGCCTGGTTTC TCCTGCGTGTATT GCACGGAAAAAC AAGTGGCTACACG TGTGGCCGTCGTG TCCTGGGGTTTCG CAACACAACCTCCA CAAGATCGACAAC TATGAGGATGACA ATGTACTTAAAGA ACAAAGAGACTG ACGCCAAAGGGG ATTTAAACCGCCA AATCGTACATTGG GTCTCACTACAAT TTTTTTACGTGTAT TTATTTTCCTAAGT GTCTGTACTTGCC ATTCTTCGCTGCTT TTTCTGCATTAATT		TLPSDTNKAKPDLILVRDSHVLIVDVAVPWE SRSLAHAYDFKVKKYATDKKMQAYLKTIY PEKEIRTEALIISARGGWCALNNMVTKKVGL SSAWVKLALIKVMEGSVKIWRWSWSKG (SEQ ID NO: 1460)
--	--	--	--	---	--	---

				GCATATCGTATGC AAATAAGCGAATT AACCACCACCGTG CAACTATATGCAG ATGTTACAGCTGA GCCCTCTATCATA CCGGTGTACTAAT CTGGTATGGTGTT GGCATGCTATGCT TGCGTAACGACCT TTGCTGATTGGTT CAGTCGGCTGATG GTGGGTTCAGGCG AAACATTTGTATA TTGGTTTAATCAA ACCGAAACACTAA AATTTTGAACACA GTTTTCCATTACA CCAGTTGTATTGC TAGAAGTGCAAAT CGAAGGAGTCAAT TTTGACCGACGAT TAGCTGCCGATGT		
--	--	--	--	---	--	--

				GCGGTGAAAAAG CTGATCACAATAG CATACTTGGGC CGACAACCCCGTG TGCTATAAACGTA AGTCGCGAATTAT AAAGAAAACAAA CCGGACGGACTAC TCGGTGACGAACT AACATCGCTC (SEQ ID NO: 1215)		
R2	R2-1_SM		Schmidtea mediterranea	CAGTGCTATTCGA ATGTCAATGTGAA GAAATTCAACTAA GCTCTGGTTAACG GCGGGAGTAACTA TGACTCTCTTAAG GAATTAAGAATTT ACCTGCCKTAKTA AAARTGAAATCM GTTGTTCATWGCA AGTGGTATTGTAC ACCTTCCCGCGGT	GGCCACGCGCGTCGTC CTTGTTTGATCACTAGT GGATCAACCTTCGACTC CCCGGAAGTGTGGGAG TGGCGGAAGAAAGGCC AGAGGATGTCCTGAAA CCATATATTTATTTATA GAAGTTTTACTTCATCC TATTTACGTATTTCACT ATGAAAATGAGTAAAG TTCTCGACTCGATGAGT TGGGGGCAACCATTGG	MKKVLNNETEKLPGSNLTFMCGFCDREFDT ARGRGVHESRGHLVERDAAVQSRVKAVVS KKYYSNEEDVALAKMQLXHADLAKSEXL EAMYLALGKGRTREAIEQHIRKSLRYKGVL EEQRKLEETARGNVRQNNVGVPASNATKN LQRFLESLPLGTNRREERLDRIIRSNSIESQRL ELIHYCNDMCQDFVQLDCQXNPINAIRRRNP KRLSKKQLKRAKFSALQRLWIRDRKAAAQL VLKDKLDSLLSNKEDSKDLGSYWQQVFERE SELDRRPIPQVVENEELNSPVLEKEVEWAVK NIKKSTAAGPDGLTALALKKIPYSELVKLFNI ILLVGFLPDVLKNSRTILPEVDNPQGGGDYR

				GCTAGTCGTTTAA AACTAAGTTACAA ACCACGAGGGGC GTCCTGACGGACT GSAAAAGCATTGA GRGTCMTGAAGA GAGGCTCTTATTG TACGAATCTCTTC AACGATCGAAGTC TGGACCGATATGA GAACTAATACATT AGTTGACAGGTGA AAAATACTGTTGA TACTTAGTTCTC AGTCATGTGGTAT ATTGCCAGTCAAT TACTACATWAATA TTAGTGTGGCTCT CAAAGGAACACG ATTGRTCAGCAGT CCAATGCGCGACT GGCGGGCTTGTTG TTTGCAATTTGTTAC	GGGTCCTGAAGAGAGG CTCTCACTGTAAAAAAT CTCTTCGTGTCTGTTTA TTCCTAGGCACNTGCTG CATTATGAAGCGGWGA AAGTAAAGTTAGAGCT GAGAGATAGGTACTTG CTGCATTATGAAGCGG AGAAAGGCCTCGAATA AATAGGGTGTTAGAGT TATTGATGGAGAGTAT ACTAGTAAGCTTAAGC TGCGCCTCGCGCGGTG CCCAAAAATATACTTA ATGAGAGCAATAACTC AAGGNGAGTTTAATTC ATATGCGCATGCGGCA CCAAGGTGCTGAATGG CATCGATTAAACCTCTC CTGTTGTAGAAGCAGG TCATAAATGGAGGRGG GCAACCACTGAAACTT ATGAGCCAGAAGAAGC	PISINSVLTRLNKLAKRVSEGDFGINGQKG FKSVDGCLENLATVESILADARMKKNKLAV VFLDMSKAFDSVNHESIVRAGEIKGYPKLL MTYVKECYNDATTNVAGVTAKFNRGVQKQ GDPLSPALFNNVIDLAIERVSGTGIGYNMGG KKYSVVAYADDLVLFGESREGLQIALTALL EELKLNLTPNPAKSASLTFERSGPHWFAST DTVTALGDQIPAMGNIETYKYLGIKFNSCGV VKGSLPGIYTKKLELISKAPLKPQQRLAMLT DFLIPGVLHQAVFGQTNAGDLRSLDKRTRR AVRSWCHLPSTDTSTAFIHAKAKDGGMGIPSI RAEVQFGKLDLDRFGKLPNVKDERSKVLADN AHIKKMLEKLGVGPIKGVRCNKNLEFYN KMREELIKSNDGIGLKEASLVPSANTWLKLS DLHMSGRTFVGCLKTRGNLMATVTRTSRG GQNPGLNCKKGCQYQGSLNHIVQKCPVV KGLRIKRHDEVVKYVEEITKKAGWSATMEP IIPFEGSHRKPDVLVRGDLGKVVDIQIVSDH CGLDEKNSCKIGKYDNDIIRNYVRGLGPSRV EVAAITLNRGVWSRDSFNLIKRLGMTMD AKIISMRLASTAKMFKTCKKVLEPVCRTK TADCDGYGPEETSARPCHELNLKESST* (SEQ ID NO: 1461)
--	--	--	--	---	--	---

				CGGCTACTTGAAA AGGTTATATATAG CAGACGCTTAAAG CGCGACTGTAATT TACATYTCATTGC CCAGTATTTGTCT TTTGTCAGATTTA GCAAAATTTTCATA TTTTGTTAATTACC TTAAGTGGTTAAA CGATCCCATAATT GCTTGCAATTATT ATAAAGTAATTCA GGTA AAAATTACA TATCTGGCTGATC CTGCCAGTAGTCA TTTTACTTCCGCC GCGCTATAAAACA GTTTAAAAACTGA ATAGGAATCAAAA AGAACATGGCAA GCGACTATATGTA ACTGGGCATTCAA	TTAAGTACAWAAGTTT TAGGCAATTACTGAAC GGAGTTAACTGTAGTT AACACTACCATGTAGTT GTTTATAAAGCAAATA TCAGGTTTCAGTCTATA TACTAAAAGTATTTTTT GATACCGTGGTATATA GGCAACTAGTTAGGAA ATAGTAAGGGATGACG CATTGTCTCTCTTTATG GTAGTGAGGAACTTA TCGACTCGCGAGGGAT GAAACCCGTAAACCGA TCGATYTAGCCTATAA GTACCAGCGACAGTTA AACCATCTTACGCGAG GGGTAAAACCTGAGGA CCGATTATGGTATAACT TCTCAAGATTAGCACA AAATGCGAGTGCAACT TGAGGAGGAGGATTTG AGTGTTAATTCATAATG	
--	--	--	--	---	---	--

				CATTCCCTATTAC ATATGGTGGTGCC TGGGGTCTGTTTT ATATAATGGGTAC CCGGGAAGTGGAT CTGTATCACCAGT CATGGTGCCATAT CTTTKGATAAAGA TACAGTTTAA AAC TGCGATGATACTA ATAGAGATCCTCT TAGACCTTCGTAA AGAAGTGGGGATT GATGACATTAGCA TTGGAAGAATTAA ATCTCCAAGGAAA TGGAGTAACTTCA ATGAAGTCCCACA ACCCCGTTGAAGG GCTGGGTTCGAGT ATCGAGAGAAAA CTCTAAATTCTCTT CGGTTMTGTCCAA	TACTAATCTAATTAAAC TGTGACGGGAATTGCA GCTTCGGCTGTAATTAC TTTGAGGCCTATCACGG ATTGTAAGGAACATAT TGACACCGTAAGTCTA ACGTGTTCCCGATTTCC AACCAGGTCATATGAA GGGCTGCCCTTGATAA GGCGGATTTGACCCAA TTCTTCATATGAGAGGC TTATTCCAGCCTTCCCG TAGTACCGTGAGGTTTT CCCGCCTCGAACGGAA CAATGTTGCAGGGTAA TTAAGTACATCGGGCT ATATMGCGATATTTAA CGTTTTA (SEQ ID NO: 1339)	
--	--	--	--	--	---	--

				CGGAGGGGACATT ACTGTAAAATATC CTCTAAAAACAAC T (SEQ ID NO: 1216)		
R2	R2-1_SP	.	Strongyl ocentrotu s purpurat us	TCTCGCGACGCGT TCTTCTGCCTGAT GAAGTCACGTCAG GTAATAGACTTAG AAGGTTGATGAGC GTTCTCTCCTGG ACCGGGGGTGAG GATGTGTTCTACT GAATCATCGTTCC GGTTGTGAGGTCC GCTGCAAATAGGC CTTGGGGTGGTCT ACCTCCGGGTCGC TACTCCTTTTGAG CTAGTCGATCCAG GTGAGAGTCGGGG AAGCCCACTTAGG TGGGCCAGCTAAG CAGATCACCCCCC	TAAACCTTGCCTCCCCG GGCCCCCCTCAGTGA AAGACAACCTTTCACCG TAATAATCATATATTTG TATACCATGTATTAATC TAGGTAACAATTGAAA GTAACATTGAACCGTA TTCATACTTTAATGAGT ACAATAGTGAAGGGAA CTGTATATTACATACCT CGAGATAGAGGTTTTT GTACCTTAAGGGTTCGT GAGAATCCATACTGCA TAAAGGGGTAACTTG TAACTCAACCCCAGGG AGACAAGCCGACAAAT TCGGCATTACTATGTGA GGGTCATAAGTGTTAA AGGACCCATTGTATAT	MENSFAWEGTSSAEGRTTVEDSPSSSDDFVS NVGFKVAKADPTVWEEANMSEDNTIIEDPP SSSDDFANNVGFKVTRADPTAWEEASTSTE TEDLPSSSNFIDNVETQIDMAGPTAWEDADS NEDNIDEGTPNNINNNLAIVRGRADAYACSC CERNFISLKAIGTHLKETHNKKVVFECACQ HTFVKAHGLACHVPKCKEDADTPMLNRLL HGCGEGLAFNTRRGLSQHERHRHPSACLS TRRRSRLDGIARKKSLRNRRDIWTNDEIRLL KQLMIQYEHAKKINIKIAEHFNHKNQVM HKRRSLREKDMALGAPHDAPPPLAEPIIEV VEGAREELEQAPVDVILPDLEALTVNDGRG GGSPVLTEGGESTRDMEENGTDTRSPSPRE ERAGSTPWERGWQPRVDRGRGEYKGYGGE RGDRHTVSLSQRGESGVDPLTPGGVVEDYD DSYLEDYFPGWDEDEHMHIIGRLDLSDESEQ GEAAVSPRLGSFGDLLEEVDNAMEGKDNLSE ALAETLGLVLHEGHRVEYIKEKMNINVKQM ATEILAHGANKGNPKRRKEAVAANKRNPGR

				CAGCACGACAGTG	AACTTGTAATGTACA	LDRAQRDNQAKAKAKEKKRIFSETQTQYKK
				CGCTGTAATATAG	GAGAATTCGAGTTGAG	NPHRLVEKLLDGKGDERCSVSLEVIQRTYM
				CCTGTTGAGAGTG	TTCGAAGAATAAACAA	NRFSRESKEVDIGAYVDPETVEDNQGIVDPI
				CACCCATTTATAC	ACAAAAAAGAAAAACAA	SKAEIERAISTTKKGSAPGPDGVTYDALKAY
				TGTTATAAAAGTA	AGAGAATTCGTTTCGTCC	GNCQLYLLIMYNTWLAMGKVPSEAKTYRSI
				TAAATGGTTCTTA	ATAGAAGCGGAAAAAAG	LIPKGQGDPMDINNFRPLTLANVISRLYSKIL
				GCTAACCTATCCA	CGCAAAAAGGCGTCCT	TRRLDGAVSVCPRQRGFTHKASIEDNTLILR
				ATGATTTTTATTGT	AACCTCACGGTCTAAC	ELIMKSKRNKECLAVVLLDLAKAFDTVSHD
				TATGTATAATACT	CCAGAAACATTTAAGG	LIKALRRHRVHEHLISVIMDLYEGGTTSTTT
				GACCACTTTAGTC	GGGGAAGGAGCTCATT	DEGTTCPAIRSEVKQGDSLSPVLFNLALDPL
				TAGTGTCATGTAA	ACTCCAATCCAAAACC	LATLEQRGKGVEIGGHTFVSLAYADDTALV
				GGATCCGAGACAA	ACGTTCCCTAGCCCAA	SSSHLDMTANLDITVEYLNATGLSLNVRKC
				TAGCCTATGTCTC	AAGCGCGCTTCAATGA	QGFLLPINKSFLVNEAESWVVEREAIPWVE
				TAGTCTAGCAATA	GGGAGCCGATATTGGA	PGDTAKYLG VQVGPWSRPWPSIQPVIKRLT
				ACACAAGAGGGA	GGACAGAGTTCATCAA	AYCESIDKAALKPRQRIHILTTYIAPRIAFEIA
				TAATCCACCTCCC	AAGCCACCTGTAAGTA	EGGYSTLVDCRGGIQYTRIREDMTIRNYVR
				TACTCCAGCCACC	GGCCCATTCTGCCAAG	KWLFLPACLSNSFLYTRRGEGGLGLVSFYD
				ACCTTTACCTTTCT	GACATGCGCAAAAGGA	YVPTERMRLKVRVCDSEDPVIAGAAASLGL
				TATCTTCCATATA	AGCAGATTATCAAACC	RERAAKISAQTGLPVPVKPKGAHNAWRKV
				GATGCCAACCAGA	AGTCAAACAAGCACAA	QKKKWKAQPTQGKGVSCYQHRLGNKWLG
				CGGGTCGATGGGC	ACATTGGGGGGATGGG	APSFLTENDYIWAIKLRTNLVPTREAMGRGII
				AGTGAGAAACAA	ATAACCCCGGAAAGAG	GRNQVECRHCHTTIETMGHISGYCQMVLDI
				GAATGGGGATAAT	AGGGATCTTTAGATAG	RLIRHNRICKALIKAATATGLRVTEEPRIVGT

				TTGGAATTTGACA TTCTTTCCTAAAG CAGACCTGAAGGG TTAAGAGTCACAA GCGGAGTCCGGAC TGTCTTTATAAAC AGACAATATTTTC TTATTTTACCACCT ATATGGGGATATT TCCCAACTCGTAA TATAGGGCCCACT GTCAAGTGGCTGG TATAAGTTACCCT GTGGGTAACAAAT AAATTCAAACAAA GA (SEQ ID NO: 1217)	TGGATGGAAGGGGTGG GACGTTGAATCCAACC ATGCCGTTTTTATGTTC CCGATAAAGAAGGATA AGGTCACTCCAGCCTG ACACACAAAGTGGGGT AAAAGAACTCCGCTCG TACGGACTCCAAATAG AA (SEQ ID NO: 1340)	DGKNYLPDLIFSAGAGEPCYVVDPTVVWDD DPKNLREAWRGKVRKYTPPIPAVEAMLHPS SVQIFGFVCGARGTWCPMNDDIAKIVGLKN SGISRTLQIVLCDTIRMVKAFMAR (SEQ ID NO: 1462)
R2	R2- 1_SSa	AGK D010 72455	Salmo salar	AATCTTTAACCCC GGACTCTTGGGGT TCTTACGACTCTG TATGAGGAACAGT CGAAGAGAGGGC GCTACCAATCCAA	TAGATCCGTTTGTTATG ATTGGAGGGAGCCTGC CGAGTGGTATGAGCGC TCCAAC TATTGAACCCA TATGATTCCCGAGGCCT GGCCAGACGCCTAGAT	MSGKRIVEMSGCDEKICQNKHCLKRRWAW ISGPKGETSPPRKRGTCENVSFQDKSHASDP DPLKAPEAREDAGSVAPQWVGEIKTPSLTSR DGVSEVVLPPQPVHAEGVSPASDSKDKATKI TLLISLPVCDLRCGR CERPLETVGKAVRHFA VAHPTVSVVFKCQKCEKSSKNSHSISCHIPK

				GTATATGTCCCAA	GCCTGCCACAATTGAA	CKGMTETRTDVEGDHGCDCHEKFTTAMG
				GAGGGCTGGGAC	CGCAGCCCTAGCTTGCT	LTQHKRHRHIVQYCKEKEGEMTARRKGEV
				AGGGTGGGAAGAG	AGGAGATCCATAGGAA	EAVKWSEWEESEVARLSGLAGLKMNRRI
				TGCACCTCGCGAT	CTGGCCTATGGGGCGT	ADSLGTGKTAEQVRQKRRRMRPEKVRCDK
				CTGGGGCAGGGA	CATGACGGTTGAAGTT	PKEAKDKSNLIKMLSIPSATPTPQTGLKGFL
				AGGAATGGAGAG	CCTCCATAGCGTGCTTG	GELNGVATKGEVQIGGVTLSLRGVEQDSAL
				AAGTCGAAGAAG	GGAGGGGACGACAATG	LNTSALELQRLGGRAGSANPLSLQRERETT
				GCTTGTAGAGAAG	ACGAGTCATGACGTAC	LPSEERRKTKQGEYRRVQKMFRSNEKKIAKY
				GGGCTCTCCTAGA	CGAGAGAACCCCAACC	ILDGNGDGEAASPPLEIALAFKSRWEEVETF
				TCCTAACCTGTAT	CAGGTTGGGGGAGAGA	HGLGQFYSRGEADGVVFRSLISMSEVCENL
				GACGCCCCGTAAAA	GCCAGCAAGAGCGGAG	GAIKNNTAAGPDGITKPALLEWDPTGAKLA
				CGGTGACCCCACT	ATGCTTGGTATACCAA	AIFSIWLTSGTLPGPFFKKCRTTLIPKTDDPILL
				AGCGAATAAAGG	GCTAGCAGAGAGAGGG	TQVAGWRPLTIGSVVLRLYSRILTHRLEAC
				AGGCAGGTGACA	TTGAAGAGGATGACTA	PINPRQRGFISSPGCSENLMILGGLIKRSWAK
				ATAGAGGGCAGG	CTGGGCTCAGAGTCAT	GERLAVVLVDFARAFDSVSHSHILEILRQRG
				GCCGACTTCCCAG	CTCACCCCTAAAAGGCG	LDEHIIIGIVGDSYTDVTTTITVSGEQSPPIDM
				GTTTACATTGTTG	GTGGGGCATCGGTTGA	RVGVKQGDPMSPLLFNALDPMIDTLERYG
				TACTTGTCAACAT	ACACCTACCCATACCG	LGYRMGEQQITALAFADDLVLVSDSWEGM
				AAAGAGGTGTCTC	GGATGGGAGGTGGTAG	ACNIRILEEFCRLTGLRIQPRKCHGFLIQKIQR
				AATAGTTTGAATC	GCCGAAAAAGAACAGG	ARSVNLCKPWIVCGEELHMGVGPESVSYL
				AACAAGGGAGAG	AAGATGGTGGAGTAAG	MKVSPWHGIMEPDPVERLCNWISSIGRSPLK
				GAATACCGACCTG	TTGAGAGCGGTTGCTC	PSQKVRMLNVYAAPRMTYQADHGGLGPIV
				CTCCCTTGGGGCG	GGGAAGTTATGTTGTG	LNVLDMIRKAVKVWLHLPLCTCDGLLYSR

				GGGGTACTGGTCT TAGCCCGGTTCCC CGCAAGTTTCCTT TGCCTGGGATGTG CCTGACTGGCTCC ATCCCCCTTTCCCC ATTAGGCACGGCT AGATGACGCACCG ATGGGCGGGTGTG TAGGTCGCTACCG AAGGGGACTGGG GGTGTCCGGTGAA CCAGGACTTCCCA AAATGGTCTCACA TTTTTAAGCGGCT TGAGTATCGCCA GTATCCTCGCGCG GCACTGGGAACCC AGTCAACCGCTCT GTGCCCCGGCGCA GGCGGGGGTTTAA TGTCTCCCCGGCT TCACCGGCGCTTC	ATAACTCCATTAAGGC CGGTGGGCATGGTGCG GATAATGGAACTATA AAAACAATAAAAAGAA AGACCAAAAAAATGTT CTGTTATGATGCCTTAC ACATGTCTGGGAGACC CCATAAGGGTCTCCCCT TATACTTCACTGGGAA ACCCCATAAAGGGTATC CCCCTATATTTACTGGG AGACCCCATAAAGGGTC TCCCCCTATAGATGTAG AGCGTAAGGGGTCTCC AAAGTACCGGCCGATA TGGCCTTATGGCAAAC CTGGTGGTAGGGACAA GGAGGTAAGGGCAGTG CCAACCCCTACTTGATC GGGACCATCCAGGGAA TGCCATCCTCCCGCGAA GGTGATGTGGTGAGGT AAGGGGGGAGCCCGTC	CQDGGLGIVKLACQIPSIQARRVYRLWHSKE AITRVVTRRTVEAEEYRGMWLRAGGSEAGL PPLEDREEGAVQCTDTAGSVKPKNPVIPDW RRAEFLKWQNLTAQGVGVQVFGGDKNSNH WMANPETLGSKERHYIAGLQLRANVYPTRE ALSRGRPDLPKVCRQCLAGTESCAHILGQCP AVKDSRIRRHKLCDLLASEAESAGWTVIR EMCCRTRAGALRRPDLVFVKTGFALVVDVT VRYEMAYDTLMGAAAEEKVARYTPITPYVA MTLKARRVKVFGFPLGARGKWPGSNDRL KAMGVGGGRRKQLAKLFSRRALLYSLDVL RDFYRAEGETGDLDDDESVDHDL (SEQ ID NO: 1463)
--	--	--	--	--	---	---

				GGCGACGACGCA GAGGAGCACCCG GAGGCCCCCATGA ACTTAAACCAACC TATCTTGAAATAT GGCCTCTCGTTCG GGTGAAGGGCAG GTGGGAAGAGAG GGCTGCCTCACGA TAAACACCTAGTC AATAGCCAGTCGG GAAAAATGTGGA ATGTTAGGACAGG GAGGTAAGGGAG GCGGCTTTTTCGT AAGGCTCCTTCAA CCCCTACCTGTAG TCCACCTATTGCA GGTGTTGACAACA TGCAAGATGACCT GCCTCGTTACGGG TCGCGTATCATTG CTACAGGTCGTGT	TTCGAGTTTCCCCAACC CCTACCCACAGGTGAG AGGAGGAGAAGAGGA ATCTGTCCCCAACGGG AGGAGGGTGAGGTGTA AGGGGGAGACCTTCTA GTAGGGTCTTCTCAGTC GCCTGACGTCCTGACTG TGGGGTGGATCAGTAC CCTACAGGTGAGACCG GTGAGGTAAGGGTGTG GCCCTCTTGAGGGCTGC GCCAACCCCTACTCGA GGTAACCTGAGGGAGT GGTGGAATGGCGGCAT GTTAGTGCTGGGACTTG ATTGCGAGGGTTTAAT GAGAGTGGCCTGCTGA GAGCAACACTTGTGGT GCTTAAAGCGGGGCGG CCCATGACCACCGTGA GATAGGACACTGCACA GTGCAGCCATGAGGTT	
--	--	--	--	---	--	--

				GCCGCTTCTAAGA GGATAGTAAGGA GAGGTTATAGGGA GGTCCTGTTAGGG CTTCCTCAACCCC TCTCTATGCGATT CCTTACAGGAGTG GATCGAGAAGTCC CGGACGTAATACA CCCTGGAGGTAAG GGAGTGGCCTTCT AGTAGGGCTGCTT CAACCCCTCTGAT GGGAGTGTACCGG GAACCTCGACTTG TAAGCACAGGTTA GTATGGGAGCAGA AGGGGGAGCCGT AATGGGCTTCTCT TTCACCTGCTTAC ATAATACCTGTGG TGCATGTATCTAG GTCTTGGCGGGAG	CCTGGAGGATGATGCG ATGAGGTGGGGGCCTC ATCAGCCCCTCCTGGCA GGGCGTCGGCCAGGGA AACTAAATGTCTCTAGC ATGTCAGTGCAGTGAG GTAAGGGGAGAGCACT CTAGTAGGGCTCTTCCA ACCCCTACCTGTAGGTC ACCTGGTCCAGGTGTC GATGATGTGAAAACAA GAGCTACTTTGGTACCG GTCTGTTGCAAAAAGG GTTCTGCAGAGGACGA CGGCTATCCCTATCGGG AGGGAATAGTCGGTCC CAGGTAGTGGAAAATG GGGCTTTCCACTGAGC ATGAAAATGTGGTAGA GGTTGCGTCCAACCCA ATGATTTGCAGCAGAG CTCTTGGACACGAAGT CTGTATAGTCCCATGCA	
--	--	--	--	---	---	--

				AGTACTGAGAGAC AAGGTTGAGACCC CAAGATTGGGTCT CCCTAGCCTCTAT AGCTGCGACTCTT AGCGGGGATATGG AGTAACATGTACC AAGGGAGTGAAT AATAAAGGAATTG ACGGGGTACAAGT GACTGTTGGCCCG AATCCTAGCCACT TGATGACCAGGGA TATATTACAGACT GAGAGCGAATCTA GAGACGTGAGAAT AAAGTGAGAATTG GTTGAGCGAATAC AGAGGAAG (SEQ ID NO: 1218)	GGCAGCCAACCAGAGA ATGGTGGCAAGACCCC AGCTCCGTATGGGAGG GGAGGGCCAAGATATA CGGAACGGCTGCTAAA GCGTTCTGCCGGTGTCA GTCTAATCACAGACAG CTGTGACGAAACAAAG TATGGGTTCCGACATGC TTGGTCAGCTCTTAGCC GCAAGGCTTAAATCGA ACGCAGCCCGCCGAGA GTGAACATTAAACGGG GATGGAATGTGTCTAG CGGTTACGTACTACCA GGGCTCAGGTTTCGCCT GAGCCGAGGCTCTACA CGTCATGGTGGGAGTT CTCCCCACGCTCGTGAG GGCATGTAGTGGGATG GCATGTGGCGGACCAT CAGCTGGCACTACCAG GCCTCGGGCTTGCCCG	
--	--	--	--	--	---	--

					AGTGCGGGACCTCACA CATTGTAGGTGTGCTTG TCCCCCTACGTTCGAA GACTTGAGGCGGAGAA TACTCATAGGCCCCAC GGCAAAGGGACACAAC ACGGAGGCTTGTGTCC GACGAGCCGTGGACTC CTATAGACAGCCCGGG ATATCACTGGGCACGC TCATACTGAAGAAATT CGATGAACCGGGCCTA CCGGAGCAAATGCACT CTAATCGCCTTTGTGGG CGACTGTGGCCCCCTCA TGCGAGTGAGGAATAT CATAAACTGCAATGGT TCAAAAAGTGATTCCT ATGGCTCGTCGGGGAG GGCTGACTGGGGCAAG CAAATGATTGAAAGGG GAAGAACCTTTTCAAC TGTTTCTTGCCAAGCCC	
--	--	--	--	--	--	--

					GGTTGATGGTGGCGCT AGTAATTGCGACGGGA AAATGCGGTTTAAGTCT CCGAAGTAGTGCGTAG CACCGGATGTCGACCG GGTGTAAGGCCCTTC GTAAAGTCCCTGGGGA GGTCAGTCCTGGGGCT ACTGATGCGCAGTATG TAATTCGCAGAATAGG GCCATCGATACCGCCT GCGTGACTCGACTGGG TTTCCACTTGAGGATAT CCGACCGTAGCGTGCA CCCTCTTGTAGTTGCGC CGGAAACGGCTGTGTT CCCTCACGTATGTGAG GAAACTCAACAATGTG AGTGGGTAAACGGCGG GACGAACTATGGCTCT CGT (SEQ ID NO: 1341)	
R2	R2- 1_TCas	.	Tribolium	AGTCATAGAGCCA GAACCTCCTCGTG	TGAAAAGAGTGGCAGT TGTGGAGACTTCCTTCG	MSRRPGKSNEPPVRSRAMGLTTLSGKTSNS GAQGPSTSAPMQNMAGGFVDCGRSYALK

			castaneu m	GTCCCGCTGGGCA CAGGGATTAATTT TTCTGTGGCAAAT TTGACTGGCTTCA GAGAGCGTTTTTC GAAGTGGWCTGT GTGACTGCGTTCC CCCCTTAGTTGCT ATWTCCGCTKMG ATTAACATCTCAC CTCGACGTWTAAG ATCATT (SEQ ID NO: 1219)	GACGTCGGGTCGGATT TTCGGACGCCAGGGTA CCTCCACCGCTGGGTTC ACAAACTAGGCGAACA TCTGCCGATACCCTCTT TAGGTCATAGGACCAC ATGTCTCTGCACGAGAT TAACCCA (SEQ ID NO: 1342)	TSLARHKKECGKNNAECRWCGTRFNTLAG TRQHERKAHFVQYQSDLAKALPQPESELME KIAIVEARSXNGIFYKEMMASTGLTHQQVRS RREKPEYKGFLEARRSLAQTNIRAGSISPAS TXAGSLESASPKAGCSSSASPGPTTRSAPT KGVPRSSNSARIVVEAQVHTRAPPNTGETE VALRESRRTVPRLGXNPSRPCGISPLMAIAID EDSVLGGLRVQAGPSPTAVHSVEAFPGTSS MTPMETDRVHNKSGIDPILEHNGTRQVRRE ESSTREDPVEQWSPNYPKTPVTMPNITTTAD AXXTSYNRTPQTLPGNRRRRSRSLPPVQRKS ASDXLESVDSLGPWAVFLQDQVDAGSLSGN DSLADLVRVALTKSDRGVLNDVNRILAQ RAESLRIRKRGSKGKRKSKTGRHYGQTTSG SGQRAALFKKHQDLFLKNRRGLAETILSGK EDFGPRPEPPVTSVEEFYGGIFESPSPDNEPF EVRATGVEDPPPLTSPWTKRTLRSPLYWRR RPPPTYITMDEIKAARAGWQISAPGSDQIPV AAVKTMSELELAILFNILFRNVQPSAWGVL RTTLVPKDGDLRNPANWRPITISSALQRLH RVLAARLSKLVSLSSSQRGFTEIDGTLANALI LHEYLQYRRQTGRITYXVVS LDVRKAFTVS HCSVSRALGRFGIPSVIREYILATFGAQTTIK
--	--	--	---------------	--	---	---

						CGSVTTRPIRMLRGVRQGDPLSPVLFNLVM DELLEKVNEKYEGGSLQSGERCAIMAFADD LILIADRDQDVPAMFDDVSTFLERRGMSVNP AKCRALIAGAVSGRSVVRTGSSYKIHNTPIP NVDALDAFKYLGLEFGHKGVERPTIHNLVS WLNLRRAPLKPDKCLFIRQYVIPRLLYG MQNPQVTSRVLREADRLIRRHCLKTYHNLV HTPDSLIIHASVSDGGLGIMELRKAIPRIFLGR LVKLLNKNKDSVLSSVLQSNRVRTLGMKLS TMAGEVPESTFWRNQIASGPLSKGLEQAAE DSASRLWISEKPSGWSGRDHVRAVQLRTGN LPTKAIPSPVVGQRRCRHGCACDESISHVLQ MCPLTHADRIRRHDEVVKKVARHCTSRGW TVEVEPHIRSRCGRLEFKPD LAVHQPGGAIVI ADVQVSWDSESLTVPYERKRAKYDVPQFH QAAQHAWPGKALTFAPVIVGARGIWPRINN DRSAALQIPPVVRACVNSVVKWGSSIHATF MRSVWANRLNPRPLRA (SEQ ID NO: 1464)
R2	R2- 1_TGut	.	Tinamus guttatus	CTGGGGACCGTGG TTACAACCCGGGC TTAGCTGCAGAGA CAGTACCTCCCCG TGGTTCCTCCCGG	TAGGGGGCTTGGCATT CTCATTGCCTGCTCCTG AAAGGATATGGGTCCT GCGTCGCGTGGTAGGC AGACCCATTCGTCCGA	MGSWIVNFVSVATQTGEFPVDTARRAPVPV TSYPESECHXPLPLTFCNSDVTIWGGVRPEP VDCLGDLPEXYDALPGVAGPREXVGGSPPG EGVRSPGIASPSGTAVQHDFGSPILVPGAEE AEVSTPVVKVPQDHPACPCCGTRVVKVTAL

			ACCCCGTAACATC GGGTGACTGAATC TGTCTCTGCCCCG GGAGTAGTTCCTC CTTGCCCTATTGA CCAGCGGTCGCCG GCTGCTCAATAGT ATTCTAGGCGTGA AATATAGCGATAG TCCTAGTGGTTGT CTTACTGGGCCAT AGCCCCTTGCTTC AGGGGTCATTCGC GAAGTCTCTCAGG AGAACTGGGGGTG GTGTTCTTCTGGG TATAGCTAAACCC CCTAGACTGTGTC CGATCC (SEQ ID NO: 1220)	GTAGGGGGCTTGGCAG TNTCCATTGCCTGTGCC CGAAAGGACGTGGGTC ATCTGGTCTGTCTGCCT ACACCTCTCTAGACTTG TAACATCTAGTCTGTCA ACAAGATCAAATTCT TCACACAGACGACCGA GCTTGCTCAGTCTTCCT GTACCCGCAGAATTTTG CTCTTGCTCTCCTTTGG CTGTGTCCTGGACGTGG GACTATTCCATCTCGTC CCAAATGCCGCGTCCA ATTATACCGGATTTGAC AAAGCGGACGGCCCGC TTTATAAGCCGAAAA GGTGCCTTGTAATAATTG CAAGGTTCATTAAATA G (SEQ ID NO: 1343)	SEHLRRAHGRKRVLFQCSRCGRMNEKHHSI ACHFPKCRGPPVEEGPLGAPEWCCEECGQK FNTKSGLSQHKRSVHPLTRNVERIEAARPKG KGKRGAGHKGCWTEAEVAQLIELEGRFKNQ RFINKLIAEHLPSKSAKQISDKRRQLAATK TSSPEKRVTSSTSGESSPEVEKVEGIKREYRR RVGEWLCAGSLXDQTSFQKILEDVESGSEIV TGPLEELASFARGKLAAARVRHHRKHAEAE VPAREEQRWMKRRVGRRGLYLRFQRLFAL DRRKLAGIILDDVESIKCPLMEEVADVFR RWEEVAPFTGSGSFRSLGKADNGAFKPMIS AKEVMKNVXEMSRRSAXGPDGLSLRDLMK IDPQGSRMAELFNLWLLAGRVPDQVKAGRT VLIPKSADPGKIGNIDNWRPITIGSVXLRMFS RILSARLRRACPINRRQRGFIAAPGCSENKL LQALIKSAKRDHRTLGVVFVDLAKAFDSVN HQHIFQVLVQKGVGDGHIIDILRDLYTNAGTY LESGSQRSGFIKILRGVKQGDPLSPILFNLAL DPLLCRLEDRLGYKYGDQQIXSLAFADDL ALLSDSWEGMQQSIRVVEEFCQRTGLRVQA PKCHGFLIRPTKESYTINDCDPWTIADMQLD MIDPGSSEKYLGLGIDPWIGLSRPELSEVLTR WVKNIGGAPLKPLQKVDILRSYALPRLLFIA
--	--	--	---	---	---

						DHAGLSATCLHSLDLIRS AVKGWLHLPST CDAIYVSYKDGGGLPLRLASLIPNVQARRL VRIAQSEDDVIRSVVLQEGIQEEIRK VWISAG GRPEKVPSVTGEFPVMEAQA ADEALSEWER RAPRTIYPIPCWK RKRE MENWTNLKSQGHG IRNFENDRISNDWLLHYGRIPHRKLITAIQLR ANVYPTREFLARGLGEGAPRGCRHCPAEWE SCSHIIGYCPAVQEARIKRHNDICGVLAE EA RKLGWVIFIEPHLRDNTNELFKPDLVL VKGS CAKVVDVTIRYESGLTTLSDAAAEKARKYQ HLAGEVRALTSATTVD FLGFPIGARGKWYV GNNGLLSDLGFSTSRVVRIARALSKKALLSS VDIIHIFASRARQAQTSE (SEQ ID NO: 1465)
R2	R2- 1_TSP		Trichinel la spiralis	CTCCTGACTAACC TGATTTTCGTCCGT GCGGCGGCGTTTT CTTTTCGCTCTCCG CTCGTCGAAATTT GCTGTAGTTGATT CGCTTTTCTTTGCG TTTTCTTCTACTTT CGCAGTTTTTTCT	TGAGGTTTTTGT TTTCT TTTTTCCTTTTACCATT C TTGTTCCATTGTTGTTA TTTGCTTTAATCCTGTA TTTTACCGCCGGCAATT CCATTGTTATTATTACT GTTACTGTTATTATTGT TACTATTGTTTTTACTT TTACTTACTACTGTTAT TATACTTTAATTCGTTA	MSNRLANTAAAGGVPEKTS GTLDIPGQPSSS GEKRAISYPGPFGCNSCSFTSTTWLSLELHFK SVHNIRDFVFLCSKCKKSWPSINSVASHYPR CKGSVKA AVVPTSLANTCTTCGSSFGTFSGL QLHRKRAHPDVFAASCSKKT KARWSNDEFT LLARLEAGLDPACKNINQVLAERLMEYNIT RGVEMIKGQRRKDQYKALVRQLRSNSETQ QCVGLAGSMDSNVPANDTSSSVASEVSITYP EYGAVMSCDLIKEATGMAIVDINELQSNLR KAFLSGRKLPMKFH GARETAQKKMANPRV

				GCATTGCCACG (SEQ ID NO: 1221)	ACTTACGTTATTGTTAC CACTACTTACTTTGCTC TCTCGCAAACGTTTCGTT GTTGTTTCTTTTGGACC AGGTTTAGAGAAATCG CACGCACAGCGGAACT GGACCGCTTAAGCCAG AAATAGTAAAGTAACA A (SEQ ID NO: 1344)	AKFKRFQRLFRSNRRKLASHIFDKASLEQFG GSIDEASDHLEKFLSRPRLESDSYSVISGDKSI GVAHPILAEVELELKASRPTAVGPDGIALE DIKKLNTYDIASLFLNLWLKAGDLPASVKAS RTIFLPKSDGTTDISNCRPITIASAMYRLFSRII TRRLAARLELNVRQKAFRPEMNGVFENSAI LYALIKDAKVRSREICVTTLDLAKAFDTVPH SRILRALRKNNVDPEVDLISKMLTGTTYAE IKGLQGKLIPIRNGVRQGDPLSPLLFSLFIDEII GRLQACGPAYDFHGEKICILAFADDLTLVA DSAAGMKILLKAACDFLEESGMSLNAEKCR TLCITRSPRSRKTFVNPAAKFIISDWKTGISSE IPSLCATDTFRFLGHTFDGEGKIHIDTEEIRS MLKSVKSAPLKPEQKVALIRSHLLPRLQFLF STAEADSRKAWLIDSIIRGCVKEILHSVKAG MCTDIFYIPSRDGGMGFTSLGEFSLFSRQKA LAKMAGSSDPLSKRVAEFFIERWNIARDPKV IEAARRVYQKKRYQRFFQTYQSGGWNEFSG NTIGNAWLTNGRARGRNFIMAVKFRSNTAA TRAENLRGRPGTKECRFCKSATETLAHICQR CPANHGLVIQRHDAVVTFLEGEVARKEGYQV MIEPKVSTPVGALKPDLLLIKADTAFIGVDVGI AWEGGRPLKLVNKMKCDKYKTAIPAILETF
--	--	--	--	----------------------------------	--	--

						HVGHAETYGVILGSRGCWLKSNDKALASIG LNITRKMKEHLSWLTFEIIFITQISRIYNSFMK K (SEQ ID NO: 1466)
R2	R2- 1_TUr	Карка с_6	Tetranychus urticae	CTCTCTTATTTTAA CATATTCGATGTA CTCGTACATTGAA TATGCTTTTATTTT TTTTCAAAGTTTTT TGGGTGCATACCC CTGGAAAATTCTG AGATGTATAAATC TCCCATCAGCTTT GGCTGAAACGTTG GCTAAGTTTTGTA GGTTGTTTGCCCC CTACTACTTAGTC GCAAATGGTATTT GCTAACAGTTGTT AAATTGTTACATT TACAAGTCCTATC CAGTGCCTCCTCG TGGCGCTACCCGG TAACACTTAGAGT	TGATGTATCCCTTCAAT ATATTGTAATCCTCATT CGTCCCTATCCTTTCAT TTGATAAAAAGAACTT TGTTGCTCCTTTTAATA GTTGGTCCCTCCTGTCC CTTTTCTGGAACCTGGT TGTATCGATTATTGAAA GTTGCAATAAACGGAT TTAA (SEQ ID NO: 1345)	MCILGGLTSHSREGGLSRGSSQLKTVKPQNE EDNGTTQLKAGSADSFPRPSGDLNPEEPLSI DICPVCFRQMKSYLGVVRVHMQKMHLEEYN ASIPDPVVSHTRWSDDEEAQALAFTEAKIEVD KLLPRGKGINKFLELLPGR TLESIKSHRKRQ SHKDLVRKYVKEFVDTLAADNDDDTIICQD NGDIFNDPIVGATDSQSETETVADPAEFKTFI ELADDPTKPKVVAKLRNLIKDKPKSEILGSD ILVRILRRTLHGLPVEDEL DQYLEVYFTGKIK QRRSKTQTALSKKQIKQRDYGRLQELYSR RKRCANEILNPTSMSSGGFGHQELSEFWTKTF GPDEQPTLGEVEIIPKENCWW DIFSPISSDEIK ASYPSIGKAAGPDNFSAYQLRKVPVWHEC LYNIFAFYKDIPSRLKDAKTILIPKKDNAESP GDFRPITLSSIITRHFHKILATRVNNFVR FHP MQRGFIQSDGCLENTALIQTVIREAKVRRKQ VHITFCDVRKAFDSVRYDSIIAAIAKKGAPG SFIMYLSNLYRGNKTLLTAGGETRITPTRG VRQGDPLSPILFNCVMDQILTALPSRTGFTLS AGDESVNVNCLAFADDIILISKTKNGHQELL

				AATCTGAGTGGCT AAACTGGAAGGG CGGAAAATGCAA ACAGGCGGTTGGT AGATGCTTCGGCA TTTTGCCAAAAAT CCACGGCTTTTTA GCCCAACAACATC AGGGTGATGGACC CGCCAGCTTGTGG TCAGGATCCCATC CATGAATAAAGCA TGGCTCTGCTTCT GGTGCATCCTCAA CGGGATCGGCTTC GGCTGGATGTAAG TCTTGCGGAGGC (SEQ ID NO: 1222)		DVTQRILKENGLDLNPDKCCSLSLIPHSKTK KIKVVRADFFVNGVKVRSMSIGDSTCYLGV SINVTGQVAPVKMYQALCEKLDCAAIPHQ RLYILKHFVITKMFHPLILSTIAAHKIKNLDLI SRRYVRKWLHLPDCSGSMIHAKVSDGGL GVPLLFRTIADLKVRRKEKLQVHENPIFRILA KLSTVSKELNCKKIASKTTDIQEKTFKEML ATYDGLSLKEARAVPEVHKWVDSYDKRYK FAGRDFVQVIQARFNALPTRSRVWRGRGAD EKSLRCRAGCNARETLNHVSQSCFRTHRVR TARHDKILDFICERLDVVGVKYVREKPISFP GKKLIPDLIVENTDQALVLDLQIVGDNSELP LDERGKNKVIKYNCSQMELYKRKKKTALA VKALTLHYKGLMAPETSNILRSFGFKSKDLE KMAYMALFGTVAAWGIFNRSTETMRSVAN WPRPEEL (SEQ ID NO: 1467)
R2	R2- 2_DWi	.	Drosophi la willistoni	GAAGCTGGGAAG CTGGGTCGGATGA GCGCAGAAGGGG TGTTCTTCGGAGC ACTGTAATTCATA	TAGATGTACTAACCTCT AGTTTCTCTATACTTTT GCCTGCTACCTTGGCAT TACATCTAAAAAGGTA CAAACATCGCATTGGC	FERRSNSWGYQNLEPSNVGQDMNTVPRINN TTTTPATSRPGDQPREAIAVVNLAGEIPCAV CGRLFNTRRGFGVHMSHQHKDELDTQRQR EDVKLRWSEEEAWMMARKEVELEASGNLR FPNKKLAEVFTHRSSEAIKCFRKRGEYKAKL

				AGTCGTAAGTCTG ATCAAGTCGACTC GGAACCTCTTCGT GGTGTTTCCTGGG TGCTGTTGAGTTC CTAGTCTCTAGGT TCTCTCCAGTAGC TAA (SEQ ID NO: 1223)	AAAAAGAGGTGGTTTT AGTACATAGGCGCTGT GGGACTTCATTGTCCCG ATGATGCAGCGAATCG TGCATACGAGATTGTCC AGTAGTTGGTTGCTCGT ATCTTTAGAAGATTTCC TTCCTCGGCGATCAAA ANAAAAAAAAAAAAAA AAAAA (SEQ ID NO: 1346)	EQIRGQSTPTPEALDSITSQPRPSLLERNHQV SSSEAQPINPSEEQSNWEIMRILQGYRPVECS PRWRAQVLQTIVDRAQAVGKETTLQCLSNY LLEVFPPLNEPHTIGRSNLRRPRTRRQLRQQE YAQVQRRWDKNTGRCIKSLLDGTDESVMPP NQEIMEPYWKQVMTNPSTCSCENTRFRMEH SLETVWSAITPRDLRENKLKLSSAPGPDGITP RTARSVPLGIMLRIMNLILWCGKIPFSTRLAR TIFIPKTVTANRPQDFRPITVPSVLVRQLNAV LASRLASKVNWDPRQRGFLPTDGCADNATL VDLILREHHKRWKSCYLATVDVSKAFDLVS HQAIKTLQAYGAPT NFVSFIEEQYKGGGTS LNGAGWSSEVFIPARGVKQGDPLSPLLFNLII DRLLRSYPREIGAKVGNTMTSAAAFADDLV LFAETPMGQTLLDTTLGFLASVGLSLNADK CFTVSIKGQAKQKCTVVERRSFCVGERECPS LKRTEEWKYLGRFTADGRAQYSPADDLGP KLLRLTRAPLKPQQKLF AHRTVLIPQLYHQL TLGSVMIGVLGKCDRLVRQFVRRWDLPLD VPVAYFHAPHTCGGLGIPSIRWIAPMLRLKR LSNIKWPHLEQSEVASSFIDDELQRARDRLK AENVQRCSRPEIDSYFANRLYMSVDGCGLR EAGHYGPQHG WVSQPTRLLTGKEYLHG VK
--	--	--	--	---	--	---

						LRINALPSKSRTTRGRHELERRCRAGCDAPE TTNHILQKCYRTHGRRVARHNSVVNAVKR GLERKGCVAHVEPSLQCDSGLNKPDLVGIR QNHIYVIDVQVVTGDGHS LDQAHQRKVERY DRADIRSQMRRFFGVTGEIEFHSVTLNWRGI WSGQSVKRLLIAKDLLIAEDTKLISVRAVNGG VTSFKYFMYCAGYTRS (SEQ ID NO: 1468)
R2	R2- 2_PM	.	Petromy zon marinus	CGGTGCGTTCCCT TGGGTAAGGAACA CGAGTCTTAGTGG CCTTGACCTCCAC GTGGTCCCGCTGG TAACATCATCTCT TGATGATGGCTAA CAAGGCTAATGCA CCCATTCCATCTC CTATCTCGCATGG AGGCCGCTATGCG TGATTACTAG (SEQ ID NO: 1224)	TGAGAATAATGAGGTG CTAACCTCCCTGGGCCT GACCAAACCCAGAACA CATCACTGGCCAAGAT GATTTCCCGCAGCACGT TGCTTTTCTCTCTCGAT ACCCGAGAATGTTCTG CGGA ACTACGCAGTCT ATGAACAGTCACAGAC AACCTCTGATCCAAG (SEQ ID NO: 1347)	RPQTKLMTDKLKFSSQLARGLAKQRAMDG ARVGDPPITVRPTETDLCNTEGSWGRRPMK LLFVSVSTQTQNE DALWASDVAKPMASRSA LKMTSIPSMTFHNS SLEKEEEMNYDFYEQIK SLVESDDSSDDFTEDDEDVEESFLDISAEEPV LGKFPIDTKGTITVVLPSLEYICVICKQHMGK ASELVAHFNIKHRDIPLVFKCAKCDKTNSNH RSIACHAPKCGGIKLTEESLPMVCECCQARF ATLSGLSQHKRHAHPVTRNEERIKDGIGKTS QRGVHRSCWSLKEVEQLALLELQFQGKKNI NKIIAEALGKTKNKQVSDKRRDLSKKTGAP MSDSLHFSSRPLETLPSPPNVTTGTSSILAQA AERLTNENSGTLEKPAMEAIKAWLNGEGQH DALVETATALMLCPMRLVKNKGKRSKPEN DIIKPRILPTRSWMKKRAEKRGFSFMKHQKLF FKNRSL LASLVLDGTERHECRIPNADVYRFY

						CEKWEKVL PFNGLGQFKSSGVANNEYFEPLI SVEEVQTAIRAIKPTSAAGPDGLTRAAICAA DPEGRTL TALFNAWMITGIIPKELKKNRTILI PKVMDDEKLKELGNWRPITIGSMILRLFSRI MTARLARACPLNPRQRGFIAASGCSENKLV LQDLMRHAKKLHRPLAVMFIDIAKAFDSVS HAHILWVLRHKKVDEHVVGIIQNAYDRCTT SFKSNGESTREISIRVGVKQGDPMSPLLFNLA MDPLICTLESHGVGYSIDTDHVTALAFADDL VLVSESWVGMAANLAILESFCGLSGLEVQA RKCQGFMI SPTKDSYTVNNCDPWTIKNKDV HMIQPDESTKYLGLKICPWTGIIRSDLHVQL KTRISKIDEAPLKPTQKV ELLNAYALPRLLY PADHSDCKQSTLRVLDQEIIKAVKGWLHLP ASTCDGLLYARARDGGLAILKLENAIPSVQV RRLQRIANSSDAIARNIASSQGV EEEYRSLW VRAGGDSEAIPTFFLRGSESKEPVYPRPCDW RKRESRRRCEKPVQGRGIVNFAQDRISNAW LGPRCGFKQCFFIAALQLRANIYPTRESINRG RDGASRSCRKCSARLESLSHILGQCPAVQKF KDCATQTEAEMVPHGPAGNALPGWSLSRTF LVNVPQYKNSRIARRNKISDILADEAARLG WWVYKEPRFTSEAGELRKPALVFAKGEEAL
--	--	--	--	--	--	---

						VIDVTVRFELSRKTSSEAASHQVAYYTPPCD QVKVLTKASNVTFFGFQVGARGKVAP (SEQ ID NO: 1469)
R2	R2- 2_SMed		Schmidte a mediterr anea	AGTCATAGGGTGA ACTGCAATTCTGA CACGATGACCGAG CTGTGTCAGTTTG CAGCTAGTCGCTA AAGACTCGATCAG TCCGCCAAGTGAG GTGGCCGGGTATC TGCAGCACTAGAG CCACTGGTATCAA GAGCAGAGATAC GCGAGTGGAAGTT GAGTACGACTACC TTCACGGGGTCCT CCTGATAACCACA GTGGACTGTGGGA ACTAAATGTGTGC TCAGCGTTCCTA CTTTCTCGTAGGG TAAAGGGTATGAT	TAGAAGGGAAACAAAG GAAAAACGAAATGACT GGAACTATGAAGGAT ATAGCTGAAAGCCGCA AGGAAGGCTAAGTCCT GAAACCGATCTACATC TTCGATCCCAAGAGGA ACTGTGGGTAAAGCTTG AGCCGACGGAAAAAGC GAATGCATGTTAGACG ACGAGGTACAGTCACC TCCTCGTGGTATTTGGC GGGCAATGCTCACTAA ATTAAGTGTGAGTAGCT GAGAACTGTATGTGTA TCATGAAAAAAAAA (SEQ ID NO: 1348)	RKELVTIKNLFEESEGATAPAPVPLEVAVEVH QSSSVPEITDESTTTQEGSYSEPPIHRCENCG REFRTRAGVQQHRRKAHTNEFMEEKEKAA PTKKLRWTDEEKEILIESEIKIIEGSLKEQHE INKILASRMPGRSQDGIKIRQKQEHKAEIQ RRLHGTVTTNETRGNRTSEITEPIRSLPINTK TWSEDEMKRMLAEEVKLRKNEKDINKKL AEIFPNRTMGSIKSKRTKDKDYQDLVKLTM QTISENPDNETDFNTSNTENNSTD AEKEVK N YLNMLLLTINEEEWLTSTLKEAATLALQGK KTEASEKLNEYASKTLFPGLKITNQTRKREK KISKRETRRQEYAEIQKLYKKNISSAAEKAIN GKWSIKPEEEYHNNKDLIAWKPILEAPPFS DCRPIENIKEMDYALMEISTAEIFLAIRAMGK TAPGPDGIKYSKLKKNIQSMAILFNTCLLTSF LPLPLKIARTILIPKQENPGILDYRPLTIASVV TRVFHSILAKKLDNNAQLSQRQKGFRKCDG VAENIVILETILTNSRSEKRPLCMAFVDLRKA FDSVGHESIIRGAKRVGVPPMLLEYISSSYQN ASTNLFGEILNSRRGVRQGDPLSPILFNVID

				AACCCAGAGAATA TCCCATGGGAGAT ATCCATGGAAAAA GCACCACGTTAGA CAATCCGATGGTC TAACTCGGCTCCG AGGGGCTAACTAT CCCAAAGGGCTTA A (SEQ ID NO: 1225)		EALENLNRNIGYLLKEEKVSCLAFADDIVLI AETKGGLENHIEKLLEKLNGAGLELNASKC ATLMVMKNGKEKSTYISTKAIIKIKENDIPTM KATETYKYLGLQMGFKAREQNANEVITEGL ENITRAPLKPQQRIHILRDFLIPRLIHKLVLGR VAKKSLKRIDQNIRKKVRNWLHLPKDTTAA FIHADAGDGGGLGVPALHTIPLLRERITNL RKSNDPVTKECLRMEYTKQVLGKWSRPTKI GETLATNKSQLKEAFRKQMLITLDGKGLKD HHETPTIHKWIRRGENTMTGKQFITAVKIRGN LVATKSRNSRGRPEQEKLEAQCGRPDSLGL HILQGCWRTHGMRVERHNNICRRIKAIMKG KESEVVEEPRLQTNEGLRKPDLICHKGKIII CDAQVVADSSNCSLESENQRKIDYYKKDSV VSEARKLIGRVDEDIIMAVTFNWRGAISKTS IRDLDMLLDIKSKEVIKMSRKIIRDNSIMVEM HRNRTEKRR (SEQ ID NO: 1470)
R2	R2- 2_TCas		Tribolium castaneum	TGGAAGACCCCGC CCATGAGGCTTGG AGAGTGTGATCCT GATCACACTTGAA AAGTTATGCTGAG TACGTCGTGAGAG	TGATGCTCCTTTGGTTT TACCATCTGTGGGGGC ATCGGTCCTCACGGTTT CCTCGGGTTTCTATTG TTTTTCCTAAACCCGAC AAGGAGCCCTTTGGCC	MKSRSFRRIGDCAAGSSRRGVRLTGKAGRE GRFAASPHLSPRYLAGSVSGNVPSVPPGPGL GAGAPAFAGRNADGGPAQNPCPYCARSFT TANGRGLHIRRAHPDEANNAIDIERIHARWS HEETAMMARLEAGAIQRGGVRFMNQFLVP RMPGRTLEAVKSKRRDATYKALVQRFLQAP

			TCGGTAACTGTCC CAGGATGGTCTGG GATAGGCTAAACC TCAGCAGGGGAA AGTTGTAGGGGCC TGCCACCCCTACA CTTTTATAGATAT GGCATTTCGATACC TCAAATAGAGCCT CGGACTTGGAGGA GCATGGTTCCCCT CCTCCTCGTACTA GACCTGGAACCAA CGGTCTTGACAAC CCCATTGGACCTA CGGGAGCGGACC ATGCTCATGGACA TGGATTCCGAAGA CGAAGCGGGGGA ACACGGACCCCCC GCCGATAATGCTC ACTTAACGTCAGG CGAACCCATCGAA	CTCCTCCTTAAACACCT CTCCTTCATCCTSTTAG TCCATTCCGGCTAAAAT GATGAAGACCGAGGAG TGTCACTCTCTTGGCGG GGTTAACCCGTCCAAG TGTAATGTGACCTCGC CATTCGGGCTCTGATA (SEQ ID NO: 1349)	QINLPELRDGDAPRQPDPPQENPPEPPSFDG AIRGAVADLVGGVDWQRLGFQGDRLCNIA RRACDGGDVSGQLLGWLRDVFPVKRVSTR GDQSDLDVDGALVSRRTARTREYARVQEL YRKEPKACLARILGDRREGANRAPNRDPAFI DFWRGVFSEASAEVEGWAEVSDHGELAR RVWDPISVEEVGRSRVRNGAAPGPDGIAVS VWNKLPPEAAALLFNVLLLGRCLPAELTRT RTVFIPKTDAPRTPADYRPISIASVVARHFHR VLSARVQRIPDLFTKYQRGFLSGVDGIADNL SVLDTMLTMSRRCKHLHLAALDVSKAFDT VSHFAIVRACRSIFGSAETVLEEGRRHFVQ VRXGVRQXDPLSPLLFNLVLDRAKRLSTD VGFRLTATKVTALAFADDVVLCATTARGL QTNLDVLEAELRLAGLLLNPKNKCQALSLVA SGRDHKVKLVTKPTFKVGQNTIHQVDASSI WKYLGIIQFRGSGMCGCGSEGVAAGLKRITC APLKPQQRMHLLRVFFLPKFYHAWTFGRNL AGVLRRLDVVVRTSVRTWLRRLPHDIPVGYF HAPTKSGGLGIPQLSRFIPFLRLKRFDRLLGRS AVDYVRECAFTDIADRKIRWCRERLSGIVD QVAGGRDALDAYWTAQLHQSV DGRALRES ASVASSTQWLRCSTRAIPASDWLHYTAVHI
--	--	--	---	---	---

				ATCATCTTGATGT TACCCTTTCAAAG CAGGTCATGCGGC ATATGTCTCAATG CCGGAAAGGGTG GCTCCCGGGCGTA ACGGACATCTATC GCTGATGGAASCT AGCTCGCTCTTGA AGAACGGCGGAC GGGGACTTGGAAT CGTGGTGTGGTTC TGATGTAAGTCCT GAAATTATGGCGT GATGGCCCGCCCG CCCGACCGGAGGG ACTTAGAACCCCC TTCCGCGAGGGTC CTGTCTGTAGGTC CWCCATCTCCGTA AAACGAGTTGGAG GAAACCGCAGAC		GALPSRVRTSRGRRGGQDVSCRGGCLLDET PAHCIQVCHRTHGGRVLRHDAIAKRISADL MELGWIVTREVSFRTTAGVFRPDMVAVKEG VTVILDVQIVSPAPTLDEAHRRKVAKYRDR ADLARYLAEAAVARGRAPPANIRFASATIS WRGVWSAESVGSRELGLSARHFDRYTTM ALCGSWRNWVRFNASTASRMGRGRGDASP RRHENQQ (SEQ ID NO: 1471)
--	--	--	--	---	--	--

				GGGG (SEQ ID NO: 1226)		
R2	R2-7_MR	.	Megachile rotundata	TCTAGTTAGCAAGCGGCCCCCTCTAA (SEQ ID NO: 1227)	TAAATGTCGAACCGAA TTTTGGGTAACGTGCAC CCCACCATCCTTAATCG GCAGCACGCAATAAAG CCGTGGGCAGTGGTTTT AGTGGGTAGTCATTAG GAGTCCCACAGTACCC AGCGAACATCTTAGTG GGTCTGCGTAAACGCA TTTCCACTGCCTATCCT CCGGGAAAAAAAAAAAA AAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAA AAAA (SEQ ID NO: 1350)	SGPATSTFGETKSRLCEPTSALGCRPGAVVI QWAQIHKEKRKRIVGWPLGHLGSPTSLKLR HPRLQAKRIVPVLAELMQCLCARHVSGRSP QKSAWAFTLEERTSACGVDAMFVCSTCQRS FATKIGLGVHVRRAHVEVANAAISVERVKD RWSEEERRIMAAVEVRGVLSGARFINEYIMS HLQTSRTLESVKGTRKNPKYKELVATLLEE ARTSVREESPRSAVNDSATQPSGSPDTRSLR TEHLFTESTEPFEHRIRELIGDLEGVTDFAE LLVSIAEQQLQGDEVAESLTRWLGEVFKPE NQQQQVQRKRRRQRKAPVSGQLPKWRERR RDYAAMQTLFHRNPSLAAGRVLGDGKNESRP PDLPEMTAFWEPILTEQSAEHRAVGPASEKS ELCSVWGPVEKEELLSSVPPLDTAVGPDGV TARQWRVLPVAVRALLYNIILKRGSPASM LESRTVFLPKKQHSVNPADFRPISIASVVVRQ LHKILAMRLRRTNLVDERQRCMDDGCAENI TVLASLLDDARHGLKELHLVSLDCAKAFDS VSHHAIDATLKECGLPAGFVQYISRTYSDSS TRLEVGRNRSEPIKTNRGVRQGDPLSTLIFCL CFDRVARTLSPHIGYDLNNTRISTLLYADDA

						FLVSTTAPGMNILLRSVEESAGEVGLSFNTS KCSALSLIPSGKEKKMKVGTTPTFKTSQGFIT QITPSQEWRYLGVDFQYSGPKKASRSLKIEL ERISKAPLKPQQRLLILRVYLLPRYYHHLVL SRTTLGHLRGLDLQVRAAVRRWLSLPRDIPI AYFHTTAKEGGLGLPAFETSIPCLMLARLRS METSTCKAARAAVQGFVWQKRIHWATAAL TKNGEALTCKADVDRWWASRLHKSVDGRE LRECSGVGSSSTWVNSALNITGRDYVQYHH VRINSLPTRIRTSRGVRREGMEVTCRAGCQV TETAHVIIQSCHRTHGGRILRHNAVCKVLA SGLRDKGWEVREEPKLRTRQGLRKPDIVAI KDGVARVIDAQVVS GSGPLDEAHETKRKY SDNGDVTAAIARECNIAPSNVAYSSCTISWR GVWSPRSAADLLQVGLSKKLLGFITLRVLR GSHLNWTRWNKMTTMRVHHQRTGIG (SEQ ID NO: 1472)
R2	R2Amel	.	Apis mellifera	TGGTAATCAAATG CCTCGCTATTTTA GTAGCGGTAGCGC TCCGCCC GCGCAG GAACCATTGACGC CGCCGTAGTGTGG	TGATCGTTAAAAGTAA AAATCTATTTATT TTTATTCCTATATTATA ACACATTATTTATTAT TTTACTTATTGTTTTAAA GATGACGAAGCCGCAA	MSSNEEGASDTGAPGPGVPVADVSAADGR ATYDDHGMSTDYEEKQTIELPLNGQIQCLWC HIEGRNQRF LQESQYLKHKDTQHPKGEIWR CAACQKEFEKLHGCRCHLPKCKGRKEAKG VAKFKCDSCEESFLTQRGLSMHELHRHPAIR NLKRTQGT SRGNTRPINRASVWSKEETDLLI

				GTGATTTTATATC CAACCAATCACGT CAACTACGATCAT TTGTAATCACCGA CGGTACTTGGTAG GGGTACCACATGG GCATTCTTGCTCA TTCCACAACGCCG CCTCCATCATGGC AACAAATTTAAAT ATATATAAATTCT TAAGGTTTGACCG TATTCATATATAT ATATATATATTTA ATATTACAACCAT AAATCTTATATCG AGCCTTCTATTTG GTCTCAAAGCAA TACGTTGTCAGAT CTTGTAGAACATC AGGAGTGAGCGGT GCGCTGTGGTATC CGTGCTTTGTGCC	GGCCAATCCAAATTTA ACAAAAGAACGAGACT ACTGGTCGACATTA AAGACGAAGCAGCTGC CAGCTGATAAACAACA GAGCCCGTCTCGGCCTT TACACCGAGCGGTGCA AGTCCTGACGTACTATT GTACGTCTAGGGCGCG GGGCAGATTCTACCGT GTAGAATCTGGGGCGA CGCCTCCGCGAGGCAC TCCCTGGACAACGTAC GCTAAAGCGTACGGCT AAGTGCGCCTCCCGAA AGGGTCCCCGTTCTAA TTTTTCCGAGCCCGCGG GCAGATCTCGTGGCAG TGACGCTAGAAAGTTA AGTCCGCGGACATATA AAATTACAGCCTTAAA TAATGAACCCACGAA GGAGGTATCCTCGAAA	KLNERYKHLKQPNVALKEYFPDKTLKQISD KRRLLPVQEPEDVATTTDETGPPPSDSSEESIY ESATEDEGGGDMQQTAPNDSWKEPFIQSIRT NHLEEDSLRKVEEAIERMAMNEGVTEQEV GTLLEQFVDSLTSPTTERKGSRRKSQKTTK RKTTHNNRKKFLYAKHQELYKKSPPRRLLEL ALSGESSSGREVVNLPEADSVGPLYKSLWG QIGPEKTHRNQPMCNNIDMSEIWTPIALES VEKFKKIKSDTAAGADQIKKFHLRKKGALH VFAKLCNLLMLHRIYPAQWKTNR TTLPKP GKSAEEVENWRPITIGSLLGRIYSAMIDRKL RSKIKQHIRQKGFTQEDGCKNNIAILSSALTK MKEDSGGIITIIDISKAFTVPHGEISQSLMN KGVPSPICEYIQKMYIGCKTHIYCRDKKTLPV DILRGVKQGDPLSPLLFNLIIDPIIGTLDETTE GIKLENENISVLAFAADDLVLLAKDKETADK QNRLINEYLDDLKMKVSAEKCTTFEIKRQN KTWFLGDPQLTLGQQRIPYADPEAAIKYLG NFPNWRGLCKTSIKEIIDAARTVKQLKLKPH QKINLIRTYLLPRYIHLKVANPPPLGTLDLID KELKTHKEILHLHPSTTDGLIYTDKSHGGLGI QRVANIVKLAKLKHSILMTRSEDNAVKIAL NGQEGMVKRYATSIGLQWPCGIEEIEETRKK
--	--	--	--	--	--	--

				GCGGCGACAAACC AATACGCTGCTGC CTGTCGCAAAGCA ATACGCTGCTGAT TCTCGGATGCGGG TGTCGACGGTCAC GCAAAGCGATACG CTGGTGGGGTTTC AAAACAATACGGC GCTGGTGCTAAAA AGCATTATGCCGC TAACGGCTGGATT GTCGATCGCCGCT GCGGGGGCTAGTG GCGCACCCAGAGA GGTGCGACGCGCA AGCATTGGTTCTG TGCGAAGCGGAGT TCTTGAGAGTAAT GGTTGCTGGGGGC ACAAAGCGCAAC ATATAGCCTCTTA TGCCTCAAGTCGT	TTCCGCCACGATCCTTC TGATCGTAGGCGCAAA ACA (SEQ ID NO: 1351)	LKRADTNKWKTLISQGQGIKEFFGDKTGNA WLYNPEMLRPSRYLDALKLRTNTYGTKAA LHRAKRDIDINCRRCGVQVETLGHILGLCTH TKNKRIKRHDEICDLIAKNVSKEYVIFREPEV EVNGDRRKPDMMVIKDHDKVYVVDVTVRYE NNDLNLKAYKEKENKYKETAEIMRRDLKA KESRVLPVVIGSRGAVPRATIENLKVGLQT KHALTASLIALRSSIEMANEFLDYDHTT (SEQ ID NO: 1473)
--	--	--	--	--	--	--

				AGTTCGTACCTCC ACGTGGTCCCGCT GGAATGCCTATCG ACTCCTCCCCGGA GGATCATAGAGTT CGAAACCGGCTAC GGCGAGGCAAGG GCGGTGAGGTGCA CACCGATGGGGAG CAGCGACCCCACC TACCCTTAGCTAA GAGAGCAGGCGA TCCGCCAACTGTC AGCACGAAATAA ACTAATCATATGT ATACGAGGGAGA ATTTACAACGGGT ACCTTGTGCCCGA ACCGCCTGTAGGT ATCACCTACAGGT GTTAAAATGAATC TGATAGCTGGCGG ATCGTCGACCCTC		
--	--	--	--	---	--	--

				TTTGATGGCTCTG CGCCAACGACTGG AAAGAATAGGAA CGGAAGTCTAATG GAAGGAAAGTGTC GGGAGCACTATAA ATTCCCAAAGAAG AAAAGAAAAGAA AAAAAATAAAAA ACCCAAATTAA (SEQ ID NO: 1228)		
R2	R2B_NV i		Nasonia vitripenn is	GACTAGACTATGG GTTTCAGTCAGTCC CAAATAGCCGATC CTGGCGCGTCCGG CAGTAATGCCACG TATGAGTCGGTTA CCCATCTCTAAAC GCGTAGAGGTGGG GAGCTAAAGGCCA GGCGGTTTACCCG ACGTCGAATTTCT CCAGGTCTGTGTC	TGACCTGAACAAAACG TGTTGTCTTGTCTTGTC TAAAACTATTTATTCGA AATAAGGGGAGGCTAA CTGCCTGCAAGTTGAA CGCGAAAGTTAGACCT TCCCACCTAAAGCCCA AAAGTGATCGGGGAAT GAATCCGCGGGTGACC CCAGAGTTGGGTAAAC CCTTGAAACGTTGGAG AAGCGGAAGAGAGTCC	TFAPTHPMVRSGPCRKTKRPGSDYRESLIMD SGNNVASEPRGAVDVTSAAPIGAELNAEPCE GRNQRREAALSAQTRRRNXARRARNAQQA DEPGDDEEIETHGPLTIRTXEPMEIVAIKNP QACPKCLQGGTQLLCMGSWELSRHINKEHP SVDVTWVCGACQRRCTTLRSWSCHVLHCK GRQEPKDLPFKCEHCSLSFDSQIGLSQHERH VHPEVRNDKRAAEANKPKGKSGRRPSIWSD EDLLLIRELESEYHGARNINEKIAEHFPDRTG RQVSDARRRKDYAALRGRGGPQGPAGVE AIEEVDEGEIPEGEELVATDGAALSGPPEN GGSAPAEQVNAPALESSSQDRECSPAVGS

				AGTCGACGGAATA	CGCCACCGAGCATCGA	DEQIEDSSDDDEFSDALGEISLPEPLSVERTTI
				AAGGTACTACAAC	GTGCTGCGGCGCCCGA	SPPPRDDWKGPMRWEICNASEEAGSYANW
				ATCTACTATCTAT	ATGAAACCGATCGCGG	VTGLQELVRNNALSEIGLDSLYDQLIQIMRH
				CGGGATCGGAAG	ATGGTGCAAGTCGTAG	PSDDNEQDRLQLNARGPPRRGHRKNRRRRR
				ACGCCTTACAGCG	GACGGGGCACGACCTA	LTAADRKRFAFARCQDLWNNNPKKLAELVI
				TTTTCCGATTTTTG	AGCCTCTGTACGGCG	ANDLSILQRRQAPGR TETQTLYNELWGRVG
				CTCTTTGAGCATT	GCGAAGCCAGGAATCA	PNIEAPRRTEDPIPVSRIFTPTPQEIMGRIRRI
				TTTCTTCAAATTG	CCATGCAAAGGTGTGA	KNDSAAGPDGVTKDDLGRGRVSIASLSKLFN
				CGATAACCGACCC	ACTGGGGCGGATACCT	SILLAGYYPKAWRENRTTLLPKPEKDPADV
				GATCACGCGGGGC	CCACGGGGTTTCCCTGG	KNWRPITISSMVSRVYSGLLDQRVRAVIKQC
				TTTGACAAAGCAA	GCATCGCGCGAGCGAT	DRQKGFTEENGCF SNIQLLDDAVSNAKKAG
				TGCGTGGTCGGTA	GGCCAAAGTCCGCTTTC	GVITILDVSKAFDTPH AVIQGCLEKKGIPET
				AGATGGTTGCAAT	TCAGCTACAAAACAAA	VAAYISSMYRDCSTAIRTRSGDVKIGMKRG
				CTTTTCCACCTCGT	AATGGTATGAGACTTC	VKQGDPLSPLIFNLVLEPLLERLQETSGVEIE
				TTCTTTTACGGAA	GTTAACACTAATTTTTC	GMNLSCAAFADDIVCFANTAPEAGRQLRM
				CGAAAGCAATGCG	CGAGCCTAGCAGGCTC	VADYLGRLDMSLSVSKCIAVEYVPHRKTW
				TGTGGGGAACGTT	CCTTGACAACGCTTATG	YTKNPGLEVNGNAVPSISPSETFKYLGAKVS
				AAAAACTCCCTTC	AATCTGGAAAAGGACA	PWKGLLEGFESDAFREVISRVQRLPLKPMQ
				ATGCATCCCAGGA	CAAAGTGGA AAAAGCG	KVDLLQMYIFPRYTYGLITSPPAKAVLKTID
				TTTATCCTGCTTAC	CTGATGGTGGACAAAA	RIIRTRIKEILHLPESVSSSFLYTPRKQGGLGL
				TGCAAAGCAATGC	GTCAGTTGAGACTTGAT	LEVEKMVLIAALRNGLRARQSHDPVTRAA
				GTGTGGAGCGACT	ATCAGTTGTTTTGACTA	MNSNAADDRLKSYADALRLHWPLTTKELD
				TTACCACGAGTCG	AGAATTTTATTATCGTT	TYKYQLRLSYAQKWAEQKWQGGQGVEEFA

				CTCCACCGCAAAG CAATGCGTATCGC GCAAAAGCAATGC GTGTGGGGGACTT GTCAAAGATCCCC CGCCGCAAAGCAA TGCGTGTCGGCAC CACGTAGAGCAAA GCGTGTAGGCAGA CTTTGTCAAAGT AGTTCTGCCGCAA AGCAATGCGTGTG GAGATCTTCGCCG GTGAAAGCAATAC GTGTGGGCGAACT TAG (SEQ ID NO: 1229)	GACTTTTAAATATTTTA TTATTGACTGTTAATAT ACTGACTTGGGACCAA GTCATCTCTGTTACCCG GTACCGGTTCTGTGCAT CAAACCGGAAAGTCCG TCCCACGTAATGTGGTA GACGCAGGAG (SEQ ID NO: 1352)	QDPVGNSWLQRYDLLPASRYIDAIKLRTNT YPTRALMKIIDGRVDSSCRKCQGSSETLGHI LGRCRYTKDKRISRHNEIKDLLKARLAKNH QVMDEPQITVRGQRFKPDLLVKTNEGRVHV IDVTVRYEHRTYLDEGRTEKIGKYRQILSTL RRDLHSNAEEVIPIVIGSRGAIPRETRKALSK LGIGKSDWLTISLIALRSSLEIVNAFMDD (SEQ ID NO: 1474)
R2	R2Ci-B	AB09 7122	Ciona intestinal is	CGACGGTGAACCA CCTTGTCGCGGTG TAAGAGCTTTAGT GTCTCGAACAAGA AATAGCTTGTGTG CTGTCCTTCTGGG	TGACAGTAATATGAAA ACATCACATCTGACCG GCACAGAATCACCATG CCGTAATGCACCCAAC TAAGGATTCCAATGGG TAAAAAAAAAAAAAAAAA	MGEWPVWSWSLTVLVEKWRPFTILQPYPM PGQLRVDVYLPRKTSYLMDKNIYENTTSPG GGPLCGEKTHRSDVIPPPGFAPSTDASNTL GENVDASATTSSANPLSQEPGWCESSKLFK SQRGLRVHQRSKHPELYHSQNQPLPRSKAR WSDEEMVIFAREEIANRKIRFINQHLHKVFP

			CGGTGCACATACT TCTTAACCTCCCG AGGCCATGCCGGC GGGGGCTTTAGCC CCCGGCAGGTTTT ACCATGCCGGACG GGTTCGAGAGGTA GAGGCCAACTAA GAGTTCACCAGCA GACTTCGCACGCG GCTGGCCACTGGC CGAAGTTTAAACA ACAGGGCCGCATC TTCCCAAACCTCAA TATATGGTGTTAA GTGAACCGTGCCG (SEQ ID NO: 1230)	AAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAA (SEQ ID NO: 1353)	HRTLESIKGLRGKNVRYARIMADLEAEMTS QPEAATSLCTETSENLASSNVLPQTRGWAE NLVENIDTAHLANLGPLSQFEPGKPPSSSTKE AINTEYNDWISKWLPSGAAHRERRANPPST KLNARATRRLQYSRIQNLYKLNRSACAQEV LSGAWKVQSGELNLKEVQPFWEKMFRRKES AKDRRKPKPTGEVLWGLMEPLTIAEVGSTL KSTTPSAPGPDKLTDGVKRIPIAELVSHYNL WLYAGYQPEGLREGITTLPKIKGTRDPAKL RPITVSSFICRIFHRCLAQRMETSLPLGERQK AFRKVDGICHNIWSLRSLIHNSKDNLKELNI TFLDVRKAFDSISHKSLGIAAARLGLPPPLIT YISNLYPNCSTKLKVNGKISKPIEVRRGVRQ GDPLSPLLFNAVMDWALSELDPVGVQIGE QRINHLAFADDIILVSSTKIGMVSSINTLSRH LAKSGLEISAGKEGKSASMAIVVDGKKKM WTVDPDLPFRKVN SQKIPALSITQQYKYLGINI DAQGARNDAARILTEGLAELSRAPLKPQQR LYLLRVHLLPKLQHGLVLSSCAKRALTYLD KSVRSAIRRWLTLPKDTPTAFYHAKACDGG LGITRLEHTIPILKRNRMMKLTLSDEPVIMEL VKLTYFTNLLHKYSNVKLLNSWPVTDKDSL ARAEASMLHTSVDGRGLSNCSDVPRQSDW
--	--	--	---	---	--

						VTNGASLLSGRDFIGAIKVRGNLLPTKVSAA RGRQREITCDCCRRPESLGHILQTCPRTWGP RISRHDSLLKRVNRQACLKNWTPHIEPSIPTNI GLRRPDLVLAKGNIAFLVDATVVADNANM QLQHEAKVEKYNNSDIKEWIKVHCPGVDEV RVTSLTANWRGCLYGGASFLTEDLGLPKA ELSLLSAKINEKGYLWCAHYRG TARLWN RPLRS (SEQ ID NO: 1475)
R2	R2C_NG i	.	Nasonia giraulti	CGGGTCCCCCGA CTTCGGCTTGCCG TGGTCTGGGGCTC ACTGCTTTTTGTG GAGTCATGGTTAC ATGGTGACCCTGG TTCCTCGCACCCC CGCTGGAAACTAT CTGGGGAGGCCAT GATTGGGTAACGA TAAAGGTCCTGGT CGTGTCTCCTGA GATAGGCTGAATG GGTCACTAAGTGG	TAGCGGACTGGACTGT CTGGAGGAGTGTTAA CTCGGGTTCTCATGGGA ACCCGACAACGTTGTT ATCTTGTATGACAATTC ATAAAAAAAAAAAAAA AAAAAAAAAAAAAAAA (SEQ ID NO: 1354)	WVTSPRRPRYVGPQKKKASDGNDGRAAAR AEPTNPGGPDRADDDEGDVKFWCEFP GCDR FFMTRSGRGLHHKKGHPDWNDQRNLAGKQ HRKEIWSEEERLLLAKKEAELAISGARFINV ELRDFTARSLDAIKGQRKRPDYKILVEKFVR ELRVRGIRQGVASRSQQARAMAVAGAPAA TSSGAPPVATQPPPSGRVLR SQVVEAPAMEI PVAESEGDSSGDELFEDVEPVRLSDLPPDRF TIYFAGLEIPGTEDIYAHRLHTICLMTTWRTK EEVRLELGLFLKDLFPSKGSQERPRTNLPD PRNRIERRRGEYKKCQDLWRRNKSTCVQRI LKEDLSQGECLPRELMEPFWNATFTQNP GT APVLPPTTEVYSSVWEPIR PENIKGNYP PQT AAGIDGLTVGDLKGV SREMLARIFNLFMWC GKLPEHLCASRTILLPKKPGAKVPGEFRPITV

				CACCTAA (SEQ ID NO: 1231)		TSVLIRTFHKVLAERLKVVPLDPRQGRGFRES DGCAENVMLLDMTIRYHHERRRKMFLALL DMAKAFDSVSFESMREVLTTKGIPTPFIEYF MTHLEDSTVLQHGNWQSGKIHPTCGVKQ GDPLSPPIFNFIMDEMLKRLPKEIGVNLDGLF VNAMAFADDLSLVANTEQGLQILIDEATSFL GLCGLRANPNKCVTLAIKTIPKEKKT AIDPSS HFRIGNAVIPSLKRTDEWVYLGIFNSNGRLI SDAKPKLIKDLELLTKAPLKPQQLWALKVI VIPGILYRGTLGSSTAGYLRSLDCVIRAYVR RWLRLPGDCPNGYFHAAVADGGLGVHPIR YKAMVDRLARLRKLEKSAYITGPEAARYLQ RQVSIAENRLRDGANRIMSDASMLREFLREL LYKSFDGRPLENSSKVPGQHRWVEEPTRFLS GADYMNCIRARIAALPTAARCARGRLKDKH CRAGCGNVETLNHVLQFCHRTHGTRIGRHD AVVKYVVGGLKKRGYAVKEEPKIVLQDVV YKPDMVATKEGKTLILDAQVLGDQRDMRL AHEDKLRKYGAPEFKRKIRSETGSATIKSLS VTLSWRGLWGPDSVKGLLEEGVILKKDLKI LSTRVLIGALAGWRRFNERTSMATSGRREE VTTRMVRRWKRRERVGVG (SEQ ID NO: 1476)
--	--	--	--	---------------------------	--	---

R2	R2La	JN937 617	Lepidurus s arcticus	GGGGTAGCAATTG ATCGATTCCCGCT TCCTCGTGGCGCT ACCCTGGGTAATA CTATGAGGAATTG ATCACACCGTAGC GAACGTCATCAGT CACAGCTGCACGA ATCCAGATAGAAA TATAACAGACGAG TAATTCTTTTAAA AGCTCGTCAGTAA TCTTCCCAG (SEQ ID NO: 1232)	TGATAATCGCTCCATCC TGCAACTAATTATGAAT GCAAATCTGTAAAGTG ACATTAGTGATACTTAC CTGATACTTACCCTGGT ATTTATTTGACCTATAC TTACCCTGGTATCTACC TGACATATATTTATCTA ACCACCTACCTATGATG ACTCCCGCGGAACTC TCACTTACCTTATTACC CACTTGGTCTTTTATTT TCTCGTTCCTTATTACT TTGTTCTTTGGTGTAG GGTTCTCTGGTTTTTG AACGGCTTCCTTAGCCG GAATTTTGTCTGATGTA TCTTGCTTGTGTCCTTG AAATATACGACCCAGG CTTGCGTCATTTAGGCT CTGGGAAA (SEQ ID NO: 1355)	MSGKSSKPRTVSSGSSSQETPPSGSNACDICG KCFMKPVGLSRVHPSQYHARLEKNQPKAK KFRWTDLEDLYFLAKKEAELLHLGSIK FVNK ELAEFFPEKSV DQIRGQRRSETYKQQVLSIHS ELLKLQTVADSPPPSRIPAKEVSAWLDFFLA LPKTKNKFSEDKLDQLIRTAQDGT LILDDLD LYLREVLVQPTSQGEKQAKLLPPP KSSREKR DREYARAQNL YRKNKTACVNAILDGNKKC ENKIPDIDDFWK TIFESHSPDAEPVCYV VDE EPTNIWSWISFFEMNHNYPDSSTSPGPDGVT ARMLRSIPARVLNKL NLLLFIEDLPAVFKC HRTVLIPKIDNPTSPGEFRPITISSIVVRQLNKI IAARVSEGV PINPRQKA FRQIDGCAENVFLL DFILRDAKTKIKSLSLATVDIKKAFDSVSHHS IFRAIRGARCPENLVNYIQNSYSGCTTQISVG GSISASKIPMNRGVKQGDPLSPVLFNLVINEII RKLPASIGYPINSELSINCIAYADDLILVTNTR EGLKLLLGLLNEELPKRGLELNASKCFGLSL TALGKLK KTHLCTSDQLDLHGT LIKNLTAE SWVYLGVPF SHIGRSKSFSPDLEALLNKLQK SPLKLQKQLFALRVYLIPRL LHGLVLSRVAI GELKIMDKLILKHLRVWLRLPKDTPLGFFYS PVKLGGLGIKNLRTNVLKCRKQRIERMLVSP
----	------	--------------	-------------------------	--	---	--

						DDVVRLVAESEIFLKETDKLKDLLTINGMCL DXRNVPRTGKNNKFWSERLYTSFDGKTLAY SEYFTQGGGWIREDKILQPAHVFAEICKLRI NALPTKSRVAHGRPTKDRSCRAGCLDVQKV PTIETINHIAQVCPRTHGARIKRHDRLVQFLS LNLKRNPKRNVLVEYNFRTVAGIRKPDIIVIE DTRAVILDVQVVGDSNLEMEYLEKSRKYS NDANFINALQKLYPTVTNLTFHAVTFNNRG LIAKSTVAALRMLGVPPRCIMILCVISLEKTL EVWRMFNQSTASARK (SEQ ID NO: 1477)
R2	R2LcA	.	Lepiduru s couesii	TTTGGGGTAGCAA TTGATCGATTCCC GCCTCCTCGTGGC GCTACCCCGGGAT AGCCTCAAAGAAA TTTGACGGTAAAG CAAAGAGGAATTG ATCACCCAAGGCA GTACATCGGCCTT CCTGCAGGAGCTC TGATAAAGATATT AGTGAGTTATTCT GTTGAAGCTCGCT	TAGTGCTTGAGTGATGC CTATCCTTTCTTTGATT AACTCTTACCATATACT TACCAGTTCTTACCCGT ACTTACCCTGTATACTT ACCTGTGTGCGTACCTG TGTACTIONGTCCTTTAGC CGCCTTGTGTTTTTACC ATTGGTACTTACCTTGT GTGGTTGCCCGATACTT ACCTTGATTTGCCTTG TAATTCTGCATGATATT TATTGTGTAGGTTCTG	MSEESRPKQTASKRGAAVEKTMMSGTYVC TLCGRSFEKSVGLSLHTNRMHPEAYNKLKE AKKPVLKKARWSEEEVFLLAQKEAELFIG GIKFMNIELHKIFPERELEGIKGQRKNPTYKA QVVSLLAEIRESKANDSSSSSSSSSSCDSASL GISNWLEFLALPKTSNQFQEGRLDRLISDA LRGVDVLENLDAYLLEVFAPMAQNPCPKP PPPAKNSRERRDREYSRVQNFYKKNRSACIN SILDGNTRSQNVIPGLTKFWTETFEKNSPPD DEAPDQFVADEPRDMYKWITFYEMSQDYL DSSTAPGVDGFSKQLRSMSPRVLNKILNLL LLENLPNSFKMHKTVLIPKIDDPKSPGDFRP ITISPVLARLLNKILAAARLSKLVPIQRQKAFL

				ATTCATTCCCCT G (SEQ ID NO: 1233)	ATGCTTACCTGATTTGT CCCCCTCATCATCTTTA GTTTCGTTCTATTTAC TCCATTATGGAGTTCCG TTTGTTTTTTGGTGGAG GTACAGCACCCTTTAA GCTGGAATTGAGTGAG TTTATGTACTTTGGATG GTTGTAATAAACTACCC GGAGGCAT (SEQ ID NO: 1356)	PVDGCGENIFLLDYILRSSKKSSKSVAMAVL DVKKAFDSVSHHSILRALNEAKCPINFINV RNSYDGCTTKLTCGGTSFPDSVRMNRGVKQ GDPLSPVLFNLIIDSAIRKLPDSIGYVIRDGLK INCLAYADDLILVASSRAGLKTLLNIVAHL SLRGLDLNAAKCHGLSIIASGKAKTTYVSAA DSLDLDGQPIKNLGVLDTWTYLGIPFSLGR AEKVSPDLTNLLNKLQKAPLKLQQKLYAVR NFVIPRALHGLILSKTNLKECNTLDRAIRVFL RTLLYLPKDTPLGFFHSPIKSGGLGITCFRTS VLKCRQLQRIARMRSSCDGVIQAVAESDIFAD EYAKLRDLIRINGNVLDTTESIKRYWAQRLH SSVDGKTLAYMDYFPQGNLWMSEDKVSQR SYVFADCVKLRINAIPTVRVSRGRPNKEMC CRAKCFDSQRMPAFESLNHITQVCPRTHGSR IQRHDKIAKFLFKNLNNCPSRSVLIEPHFVT VDGLRKPDIHYYDDSHMVVLDVQVVSDSAN LEKEFECKAKKYANDVALRSAMLIKYPFIKS FSFVAATYNNRGLIAKSSVQVLRQLGLSPRS IMVSILICLEGTLETWRIFNQSTMNAH (SEQ ID NO: 1478)
R2	R2LcB	JN937 619	Lepiduru s couesii	TTTTGGGGTAGCA ATTGATCGATTCC	TGATAATCGCTCCATCC TGCAACTAATTATGAAT	MSGKSSKPRTVSSGSSSQETPPSGSNACDICG KCFMKPVGLSLHMSKVHPTQYHARLEKNQ

				CGCCTCCTCGTGG CGCTACCCTGGGA TAACCTCAAAGAA ATTTGACGGTAAA GCTAAGAGGAATT GATCACACCGTGA CGAATATCATCAG TCACAGCTGCACG AATCCAGATAGAT ATATAACAGGCGA GTAATTCTTTTCG AAGCTCGTCAGTA ATCTTCCCAG (SEQ ID NO: 1234)	GCAAATCTGTAAAGTG ACATTAGTGATACTTAC CTGATACTTACCCTGGT ATTTATTTGACCTATAC TTACCCTGGTATCTACC TGACATATATTTATCTA ACCACCTACCTATGATG ACTCCCGCGGAAACTC TCACTTACCTTATTACC CACTTGGTCTTTTATTT TCTCGTTCCTTATTACT TTGTTCCCTTTGGTGTAG GGTTCTCTGGTTTTTGG AACGGCTTCCTTAGCCG GAATTTTGTCTGATGTA TCTTGCTTGTGTCCTTG AAATATACGACCCAGG CTTGCGTCATTTAGGCT CTGGGAAA (SEQ ID NO: 1357)	PKAKKFRWTDLEDLYFLAKKEAELLLLGGIK FMNKELAEFFPEKSVDQIKGQRRSETYKQQ VVSIIHSELLKLQAVADSPPPSRIPAKEVSAW LDLLLALPKTKNKFSEDKLDQLIRTAQEGTP VLNDLDLYLREVLVQPTRQGERQAKPLPPP KSSREKRDREYARVQNFYRKNKTACVNAIL DGNKKCENKIPDIDEFWKAIFESQSPPDAEP VSYVVDDEEPKNIWSWISFFEMNRNYPDTSTS PGPDGVTARMLRSIPARVLNKLNNLLLFIED LPAVFKCHRTVLIPKVDNPALPGEFRPITISSI IVRQLNKIIAARVSEGV PINPRQKA FRQIDGC AENVFLLD FILRDAKTKIKSLSLATVDIKKAF DSVSHHSIFRAIRGARCPENLVNYIQNSYSG CTTQISVGG SISTTKILMNRGVKQGDPLSPVL FNLVINEIIRKLPASIGYPINSELSINCIAYADD LILVANTREGLKLLLNNLLNEELPKRGLELNA SKCFGLSLTALGKLKKTHLCTSDQLDLHGT LIKNLTAESWVYLGVPF SHIGRSKSFSPDLE ALLNKLQKSPLKLQQKLFALRVYLIPRLLHG LVLSRVAIGELKIMDKLILKHLRVWLRLPKD TPLGFFYSPVKLGGLGIKNLRTNVLKCRKQR IERMLVSPDDVVR LVAESEIFLKETDKLKDL LTINGMCLDXRNVPR TGKNNKFWSERLYTS
--	--	--	--	---	---	--

						FDGKTLAYSEYFTQGGGWIREDKILQPAHV FAECIKLRINALPTKSRVAHGRPTKDRSCRA GCLDVQKVPAIETINHIAQVCPRTHGARIKR HDRLVQFLSLNLRKNPKRNVLVEYNFRTVA GIRKPDIIVIEDTRAAILDVQVVGDSSNLEME YLEKSRKYSNDATLSMRINALQKLYPTVTSL TFHAVTFNNRGLIAKSTVAALRMLGVPPRCI MILCVISLEKTLEVWRMFNQSTASARK (SEQ ID NO: 1479)
R2	R2Nvec- A	.	Nematos tella vectensis	GGTTGGGGCCTTC TCGTGGCGGAGTC GTGAGTAAGGGGT ATAGGGGTAAGG GACACCACGGACC GAGAACGGTTACC GCTCAAGGCGAGT GGTGGAAGGCATA AAATCGTAACGCC GCCCTCCGACCTG CTCCTGAAACTAA TGCCAACCAACTG ACTGTGGGGCTAA CCTCCCCAGAGTC	TGATGGTGGGTACTCG CCTCTGTGTAACAGGC AAATGAAAGCTGCGCA AGCAGTCGATGAGCCA AAGCCGCACAGCCCCC GACTGGGGTACAGGGC AGCCCTGGGCTATGCC CGAAGTCTTTTGACAG GTCCAAACTTCACCCTT GCCGCCAGTAGGGCAC CAGGGCCAGGTGCAGG TGGGCGCTTTGTTCTAT TTGGTTTCGCTTAATTT TTGTAAATTTTCCCG	MLRGTGNMNDKRDGSATADPTSALLGAVG DGS LVCNLCGLACKSRGGLSIHRRSKHATV YHAERQPAPRAKARWTNDEMILVARKQIAS EKSRC SAVVEGMREAVPHRTF DAVKSLKTK NRNYTRILEQIRAECSEEEVIESGVLKDRTEN VCVQTTSNVPGSAGRAASVELEGNIQVGHQ LAQKTMAGNNSRKQPANHTNWAEFNIEEG NITLRKSKRKANGMPDATHRPGPPTVDSLK HPVCLLQGAADKRDEPHTVEQLYYNIEEGM PLAEEQQWSEKLFDAIDSSLLSVEVELGRIVP GCPDEETRQLIDREFLDFIHSYSREKPPQRGL AKSKPPPKGPKSLRRQQYRQLQRLWDKNRS AAAEQALTGWQEVRTAAGVPLSLMEVPW REIFETPSTTDVREPAPAGPVLWQLLRPVTIA

				AGG (SEQ ID NO: 1235) ACAGGTGAACAC (SEQ ID NO: 1358)	TCGCCCCACCCTTTTAA ACCCTTTTAACACCAAT TTTTTTTACAACCCTCT ACTCAATAATCCAAAT GAATAAAAGCGGCATA ACAGGTGAACAC (SEQ ID NO: 1358)	EVEDAISSKKSASGPDGVPCAALQTMGAAS LAAHFNLWLLAGTQPKRLTECRTIFVPKEV NTHLPLHHRPITIGSVVVRLFHQILGKPMEA VLPLGSGQRGFRKGDGICQNIWLLHTLIRRS TDLLRPLKLVFLDVKKAFDSVSHESLLIAAK RLGVPGPLLTYNELYSRSETVFEVGGESSGS VKVSQGVKQGDPLSSTLFNCVIDWAVSDLD PHIGVLLGESRVSLAYADDLVLLSETEAAL TSQLNSIEKSLAHCGLKLSTGDSGKSASLNIV IDGKAKRWVFNPTPFLRASGGEIRSLVANET YKYLGINIGAQGVKAAEYNFKEALDNLRS APLKPQQLFLLKTYLLPQLHHSVLRSRTTG KLLNSLDALVRKAVRGWLKLPHDTHRAFF YAHQADGGLNVPSLYHLIPLLRSSRYERLTR VEDPEIREVSRTDYFKRVLGAAAAATTVAG HRIDSKVTLRLAWREALYASADGRGLSQCP LVPEVHSWVTDVSGLQTGSQYISAVRLRGA LLPTAVRKSRGRGGVNSNCDCCGRGQPEFL GHVLQTCPRTWGSRISRHNHVLSLIAKACRS RRWQVLEEPHQTTPAQLLKPDVWLNHQA YVVDVSVPGDNTPLSTCHNRKVAYYSGESV REWVRSKTGHNPVSSVVINWRGAMAKES
--	--	--	--	---	---	---

						YRLTKDLRLAKTLPRLVLRVLEGGHGIW LNFHRSTFAVGVT (SEQ ID NO: 1480)
R2	R2Sm-A	.	Schistosoma mansoni	ATGTTTTAATTTAT TTTTGAAC TACTA CTGTCTGAGTGCT TCTTACAACCTGA AGGCTCAGAACT ACCCACTTTTTGC TGTTTATCCACAA CAACAGTTGTGAA TCTATTCTCCAAA TATTCCTTGTGCTT TTGTCAACATTAT TCTATACCAACTG TACCACCTACTTC TTCATCTCACGTTT TAATTCTGGTCTA ATTTTCTCATCATT AGTCACGGAGAG GGCCTATGAACGG TCCGTGACGCGAA ATTCAATCCACGA ATTCGTCCTCTTCT	TAGCCCCCTTCACTCTT AGACATTCCCCCACTGT TGTTGCTTATCTTCATG TTTTGTGTTAATTGAC TGCTCTCTTCTGGGTTG ATGTCTGATTGTCTCTC TCTCTTTCCATATTGCT TGCTCTCCCCGCTTACT TCCAATAGTTGTCATAT TATGTCTTTGTTTACTT GCCATGTCTAACGACA ATTACTTTATCTACCTT AGTTGGTCCTCTTGGTT TGGTTGCCTTCATGTGT TCATGGCGGAATCTGA TGTTTATAATGACTATT CCTACTACCACCATTAC AACTATTATTATTATCA CTATTATTAACATTATT ATTACTTCTACAATTAG TATTATGGCTACTCCTT	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSS STCLICFAIFPTHNILLSHATAIHHISCPPTPVQ DGSQQMSCVLCAAFFSSNRGLTQHHRHRHIS EYNELIRQRIAVQPTSRIWSPFDDASLLSIAN HEAHRFPTKNDLYQHISTVLTRRTAEAVKR RLLHLQWSRSPTAITTSSNNHTTTDIPNTEAR YIFPVDLDEHPPLSDATTPDASTHPLPELLVI LTPLPSPTRLQNISESQTSHESNRNSMHTPPT YACDSDESLGVTPSSSTIPSCFHSYRDPLAEQR SKLLRASASLLQSSCTRIRSSSLLAFLQNAST LMDEEHVSTFLNSHGEFVFPRTWTPSRPKHP SHAPANVSRKKRRKIEYAHIQTLFHHRPKDA ANTVLDGRWRNPYVANHSMIPDFDCFWTT VFTKTNSPDSREITPIIPMTPSLIDPILPSDVTW ALKEMHGTAGGIDRLTSYDLMRFGKNGLA GYLNMLLALAYLPTNLSTARVTFVPKSSSPV SPEDFRPISVAPVATRCLHKILAKRWMPLFP QERLQFAFLNRDGCFEAVNLLHSVIRHVHT RHAGASFALLDISRAFDTVSHDSIIRAARKRY GAPELLCRYLNNYYRRSTSCVNRTELHPTC GVKQGDPLSPLLFIMVLDELLEGLDPMTHLT

				GCTAGTGGTCCCC GAAATACGGTTCC TCTGGCCTGTCAG TTGTGTTAAAACT ATATAATAACG (SEQ ID NO: 1236)	TCAGCACACCAATAAA ATCTCAATCAAACATCT CACTTATTAAACTCTCT ATTTCCCCTTCGTTATA AACTTACAATTCAGTTT AACCGAATATCTCTCTT TTACAAATCTTAAGTAT GTAATTTTGTGCCAAGC CCATTTGGGTCTGTACA ATTTGATACTTAAAAAT AAATGTTAT (SEQ ID NO: 1359)	VDGESLNYIAYADDLVVFAPNAELLQRKLD RISLLLHEAGWSINPEKSRTLDDLISGGHSKIT ALSQTEFTIAGMRIPPLSAADTFDYLGIKSNF KGRCPVAHIDLLNNYLTEISCAPLKPQQRMK ILKDNLLPRLLYPLTLGIVHLKTLKSMDRNI HTAIRKWLRLPSDTPLAYFHSPVAAGGLGIL HLSSSVPFHRRKRLETLLSSPNRLLHKLPTSP TLASYSHLSQLPVRIGHETVTSREEASNSWV RRLHSSCDGKGLLLAPLSTESHAWLRYPQSI FPSVYINAVKLRGGLSTKVRRSRGGRVTN GLNCRGGCAHHETIHHILQHCALTHDIRCKR HNELCNLVAKKLRRQKIHFLEPCIPLEKTY CKPDFIIIRDSIAYVLDVTVSDDGNTHASRL KISKYGNERTVASIKRFLTSSGYIITSVRQTP VLTRFGILERASSQSLRRLCFSSRDLGDLCLS AIQGSIKIYNTYMRGTQRLNE (SEQ ID NO: 1481)
R2	R2Tc	EU85 4578	Triops cancrifo mis	TTTTTGGTCTGGC ATTTGATCGTTTC CGCCTCCTCGTGG CGCCAGACTGGGT AAGCTGATTTATC AGTGAGCTAAGAG	TAGATGACTGCCCTACC CCTTTGCTGCCGAAGA ACTACTGAAGACTATT GAACTAACCAACTGTG TTAAGTAAGAACTAAT GCCTCTTTTTCCCTGCA	MSQKRRPEKAVPDEGATAHDVAQPDKSKC SVCGETFKGPASVTMHMVKKHPVEFNELK MAKKPVPKKVRWSEEEIFQLARTEAELTLQ GVRFINVELQKIFPAREIEGIKGQRKLAKYKE LVKDQLDEIGRAPNPPEQEIGEDVPSPFKAW LELLLALPKTPNDFLEHKLDNIIVQALKEDV

				AAACGATCACCGC AGGAGTCCATCTA CCTGCGCTGCACG TGATTTTCATTGTG CACTTGGCGAGTT ATCCCTTGTGGAG CTCGCTCGTTAGC TTTTGAG (SEQ ID NO: 1237)	TGTATCCCCTGATCAGT GACTTATTTTCTTTTCC TGTTGCGCCCTTTGTTT AGTTATTTTCCTTTAATT ACTAGAATTATCTTTTC GTTCTCCGTCTAATTGC TTTTTCGGTGTAGGAAC GGCTACCTAAAGCTGG AAGTGGGAAGTGTTTT CAATGTACTTTGTGATT ATAGAAATATATGACC CGAGGTGCATTGTTTGG CATTCCTCGAGAAA (SEQ ID NO: 1360)	NSDQVFNDLNSYLKLILEPSGRAKSVPGEIIH GDPSGSAKTSVTKAPKPATVSSSRKKRRDA EFARIQRLYRKNRTSCINTILDGNTREHEAPK NMEGFWREIFERESPDDPDDPDIFLEEEASDI WKYISFYEMCNLYPPPSTAPGPDGFSSKDLR RMTPRVLNKILNLLLHLRDLQPILKSHRTVLI PKTDLPTKPGDFRPITISNILVRHLNKKILANR VSHLIPINERQKAFLPIDGCAENIFTLDFILHH ARTKIKSLMAILDISKAFDSVSHHSIFRALR EARCPIGFIKFIENCYGGCFTKLFCGGVKYPS EVSMNRGVKQGDPLSPVLNVLVIDGLIRQIP SALGFNVSDQVKVSCIAAYADDLILIATTRAG LKTLLDLTNSYLAKRGLSLNPDKCSALSIVA SGKQKL VYIASSEHFDLAGQKMRNLNVGDS WRYLGIQFSHLGRAEKVTPDLTCLINRLQKA PLKLQQKLYALRIYLIPRLIHGLTSLKTNLGE LKTLDKLRKYIRAWLHLPDDTPMGYFYTPL KAGGLGLPSRLVLNNRLERILRMKASQDII VRTIAESETLGVEIRKLHDLLSIDGTILDTSV KIHSFWAERLYSSYDGKCLCNSANFPPGNK WIGEDSLNQRSHIFADCLKLRINALPTRSRT ARGRPLKDKPCRAGCRNSDGVKVIETLNHIT QVCERTHGARVKRHDRLVDFAVKGLQRPH
--	--	--	--	--	---	---

						RVVLKEPHYKTVNGVRKPDIVIKIPDHTYIC DFQVVSDTSCLELEFRKKALKYAEDKGLCD QLTRDHPGELSFTAITFNTRGLIAKSSVTALR KLGMPPRSIMTLQKICMEGSLEIWRFNQTT AMARN (SEQ ID NO: 1482)
R2	R2_DAn	.	Drosophila ananassa	AGAATATGGATTT GATTGTGCAGAGG GGGTGCTATACCG TAACTCGTAAGCC ATGCAATCAGATC AAGTCGACTCAAA ACCTCCTCGTGGT ATTCTCTGGGTGC CAGTATTTACTGG TAGCTGA (SEQ ID NO: 1238)	TAGCCAATGCACGGGT TCCAGATTAAGCTTGCT GCCGAAGCATACCATC AAAATCGGCATAAAAT TCGCTTAATAAAGGAG GTGGTTTTAGTACGTAG GCGTCCCGGGACTTGTC TCGGGATGAATCGTGC ATGCGTATAATTGGGA TCGATAACAAATACCA ACTAAGTTATTACTAAT ATATCGAAATACATAA ATATCCCGTCCTTACGT ATCTTTGAAGATTTCCA TCCTCAGCGAACAAAA AAAAAAAAAAAA (SEQ ID NO: 1361)	FERRKDPWGYRPPGTLKQIGATENNEPRNL NRFVRGESTASSLESTQFGTSAEVNLAGRVP CTICEMTFSSKRGLGVHMSHRHKDDLDAQR LRVDKKARWSEEETLMMARKEVELAASGV RFLNKKLAEIFTHRSADAISSYRKRSEYKAK LEQIRGQSVPTPEAAEEINTTQRRPSNSEQNRR VPRSEGGPIAPTEQTNNEILRVLQGLAPVVC LPRWRAEVLQNIVDNAQVSGQETTLQSLSS YLMEIFPPRNEPHILTRPRTEPRNMRQRRRQ QYARVQRNWDKHPGRCIKSLLEEDDESVMF NQEVMPEPYWRRVMTQPSSSSIKRDMFNME HSLERVWSAVNQDRLRATKVKLSSSPGPDG ITPKTARSVPEGIMLRIMNLILWCGNLPYSIR LARTIFIPKKATANQPQDYRPISVPSVIVRQL NAILASRLSAAINWDTRQRGFLPTDGCADN TTIVDLVLREHHKRFKSCYIGTLDVSKAFDA VAHEAVYNTLASYGAPKGFINYLRKAYEGG GTMLAGNGWVSEAFIPARGVKQGDPLSPILF

						NLVIDRLLRSLPSEIGAKVGNAMTNAAFA DDIVLFAETPMGLQKLLDTTVCFLLSSVGLTL NTDKCFTVSIKGQAKQKCTVVERRSFLIGGR ECPSLKRTDEWKYLGIKFTAEGRRARYDPAE DLGPKLLRLTRAPLKPQQKLFALRTVLIPQL YHKLTLGSVTIGVLKKFDKLVRYTARKWLG LPVDVPVSFFHAPHKSGGLGLPSLRWTAPM LRLKRLSNIKWPHLERSEVASSFVEEEMRRA RDRLQAGSEELLTRSQVDSYLANRLHMSVD GCGLREAERFAPQHGWSQPTRLTGKEYT DGIKLRINALPSRSRTTRGRHELERRCRAGC DAPETTNHILQQCYRTHGRRIARHNGVVNF LKRGLERRGCVVHVEPSLQGETGLNKPDLV AIRQNRIYVIDTQIVTDGHSLDQAHQRKVGK YDTPDIRTNLRRSFGAFDIEFHSA TVNWRGI WSGQSVKRLIASDLLSSGDSNIISVRVISGGL WSWRQFMYLSGYTRDWT (SEQ ID NO: 1483)
R2	R2_DM	X519 67	Drosophi la melanog aster	TTGGGGATCATGG GGTATTTGAGAGC AGAGGGGGAGTA TTCTTCTGTAATTC GTAAGTCATATCA TATGATGTGCGGA	TAGCTAAATCGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC AATAAAGGCATAAACA TCGCAAAATAATGGTT ATAATTAAATGGCTAT	MTTRPSVDIFPEDQYEPNAAATLSRPCTVC GRSFNSKRGLGVHMRSRHPDELDEERRRVD IKARWSDEEKWMMARKEVELTANGCKHIN KQLAVYFANRSVEAIKKLRQRGDYKEKIEQI RGQSALAPEVANLTIRRRPSRSEQDHQVTTS ETTPITPFEQSNREILRTLRYSPVECHSKWR

				AGGGGAATTTTAC TCTGTAAC TCACA AGTCTCTCCTTTA CTCAAGTCGACTC AAAACCTCCTCGT GGTGGTCCCGGTA ATGCTAAACTCGT TTAGCAGCTAATT TGAGCGGAAAAA CTTTTCCGATGGG CTGGTTCCCCAGA GGAAATTTATTCA TATTGGA ACTACA AGCACAAATAACG AGCCTCGGATACC TTTACACAATCTG (SEQ ID NO: 1239)	GAGGATGGTTTTAGTA CGTAGGCGTTGCGGAA CTTCGGTTCATATAGAG CAATGAATCGTGCATG CTAGGAAA ACTGACCA CACACAGTGTTGGCAG ACCTAGTATCTTTCGAA GATTTCCATACCTCCGC GATCAAAAAAAAAAAAA AAAAAAAAAAAA (SEQ ID NO: 1362)	AQELQTIIDRAHLEGKETTLQCLSLYLLGIFP AQGVRHTLTRPPRRPRNRRESRRQQYAVVQ RNWDKHKGRCIKSLLNGTDESVMPSQEIMV PYWREVM TQSPSSCSGEVIQMDHSLERVW SAITEQDLRASRVSLSSSPGPDGITPKSAREV PSGIMLRIMNLILWCGNLPHSIRLARTVFIPK TVTAKRPQDFRPISVPSVLVRQLNAILATRL NSSINWDPRQRGFLPTDGCADNATIVDLVLR HSHKHFRSCYIANLDVSKAFDSL SHASIYDT LRAYGAPKGFVDYVQNTYEGGGTSLNGDG WSSEEFVPARGVKQGDPLSPILFNLVMDRLL RTL PSEIGAKVGNAITNAAAFADDLVLFAET RMGLQVLLDKTLDFLSIVGLKLNADKCFTV GIKGQPKQKCTVLEAQSFYVGSSEIPSLKRT DEWKYLGINF TATGRVRCNPAEDIGPKLQR LTKAPLKPQQLFALRTVLIPQLYHKLALGS VAIGVLRKTDKLRYYVRRWLNPLDVPIAF VHAPPKSGGLGIPSLRWVAPMLRLRRLSNIK WPHLTQNEVASSFLEAEKQRARDRL LAEQN ELLSRPAIEKYWANKLYLSVDGSGLREGGH YGPQHGWVSQPTRLLTGKEYMDGIRLRINA LPTKSRTTRGRHELERQCRAGCDAPETTNHI MQKCYRSHGRRVARHNCVVNRIKRGLEER
--	--	--	--	--	--	---

						GCVVIVEPSLQCESGLNKPDLVALRQNHIDV IDTQIVTDGHSMDDAHQRKINRYDRPDIRTE LRRRFEAAGDIEFHSATLNWRGIWSGQSVK RLIAKGLLSKYDSHIISVQVMRSLGCFKQF MYLSGFSRDWT (SEQ ID NO: 1484)
R2	R2_DPe	.	Drosophila persimilis	AAGATATGGATCT GAATAATAGCGTA GAAGGGGAGTCAT TCCGTAATTCGTA AATCGTAAAAATC AGATCAAGTTGAT TCAAGACCTCCTC GTGGTATCTTCTG GATGCTATTAGAC TGA (SEQ ID NO: 1240)	TAGCCTATACACTATGT TGGAGAGAAGACGCTT GCTACCTAGGCAAAAT GTGAAATTAGGTATAA ACATCGTGGTTGTAAA ACTTGAGGTGGGTTTTT AGTACGTATGCGTGATT ACTTCGTAATCATGAAT CGTGCATGCTAGTGGG GTTTGGCCTCCACTAGT ATCTTTGAAGATTTTCC TTCCTCAGCGATCAAA AAAAA (SEQ ID NO: 1363)	SSFGLIVTNLNSETVLWGCQPLGQFSLIGTN MQNTTPRIINTNSLTNQIPTVSSLGAQSEHSA QVNPNSGYQCTICESSFRSKSGLGVHMSRRH KDEFDQLRLRTDRKAQWSEEELSMMARKEI ELAANGERYLNKKLAEVFTNRSVDAIKKCR QRERYKTKIEQLKGQAVPLPEALESETIQRR PSIRERDLLVTPPNTLGTTPTELSNSEILAVLQ GYPPVVCNDQWRVEVLQSIVDGAQASGKEI TLQRLSTYLMEVFPSQNDRPIQTRPPRRPRN RRQGRRQQYALTQRNWDKHKGRCAKAILD GTEGTATMPSQGIMGSYWRQVMTQTSPTY GTNTTFRTEHPLEGVWSPITLGLDLRVHRVSL TKSPGPDGITPRTVRSIPSGVMLRIMNLILWC GKL PVSIRQARTIFIPKVGNASRPQDFRPITV QSV MVRILNAILASRLTSSVDWDPRQRGFLP TDGCADNTTIVDLILRDHHKRCKSLYIATLD ISKAFDSVSHA AVSATLTAYGAPKEFVDYV QNSYEVCGTTLNGDGWRSEEFIPARGVRQG

						DPLSPIIFNLIIDQLLRSYPNEIGATIGDHTTN AAAFADDIVLFAETRLGLQTMLDTTVDFLSS VGLTLNSDKCFTVGIKGQPKQKCTVVIPETF RIGSRSCPALKRTDEWKYLGITFTAQGRTRY SPADDLGPKLLRLTRSPLKPQQKLFALRTVL IPQLYHKLTLGSVMIGVLRKCDILVRSTVRK WLGLPLDVSTAFFHAPHTYGGLGIPSVRWV APMLRMKRLSNIKWAHLAQSEAASSFLTDE LNKARGRTLAGLNELTSRTEIETYWANRLY MSVDGRGLREAGLFRPQHWVCQPTRLT GQDYRNSIKLRINALPSRSRTTRGRNELERQ CRAGCDAPETTNHILQNCYRTHGRRVARHN CVVNNLKRILEEKGHTVHVEPSLQLETSSVK PDLVCIRDNHACVIDAQIITDGLFLDDVHHR KVEKYKRPEVISALRREFGVSGNVEVLSATL NWRGIWSNQSVRRLIAKGLISSGDSNVISAR VVTGGLYCFRQFMYLAGYTRDWT (SEQ ID NO: 1485)
R2	R2_DPs	.	Drosophila pseudobscura	CAATTGGAAAGAT ATGGGTCTGAATA ATAGCGTAGAAGG GGAGTCATTCCGT AATTCGTAAATCG	TAGCCTATACACTATGT TGGAGAGAAGACGCTT GCTACCTAGGCATAAT GTGAAATTAGGTATAA ACATCGTGGTTGTAAA	SSFGLIVTNLNSETVLWGCQPLGQFSLIGTN MQNTTPRIINTNSLTNQIPTVSSLGAQSEHSA QVNPNSGYQCTICESSFRSKSGLGVHMSRRH KDEFDQLRLRTDRKAQWSEEELSMMARKEI ELAANGERYLNKKLAEVFTNRSVDAIKKCR

				TAAAAATCAGATC AAGTTGATTCAAG ACCTCCTCGTGGT ATCTTCTGGATGC TATTAGACTGA (SEQ ID NO: 1241)	ACTTGAGGTGGGTTTT AGTACGTATGCGTGATT ACTTCGTAATCATGAAT CGTGCATGCTAGTGGG GTTTGGCCTCCACTAGT ATCTTTGAAGATTTTCC TTCCTCAGCGATCAAA AAAAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 1364)	QRERYKTKIEQLKGQAVPLPEALESETIQRR PSIRERDLLVTPPNTLGTTPTELSNREILAVL QGYPPVVCNDQWRVEVLQSIVDGAQASGK EITLQRLSTYLMEVFPSQNDRIQTRPPRRPR NRRQGRRQQYALTQRNWDKHKGRCAIKAIL DGTEGTATMPSQGIMGSYWRQVMTQTSPT YSGTNTTFRTEHPLEGVWSPITLGDLRVHRV SLTKSPGPDGITPRTVRSIPSGVMLRIMNLIL WCGKLPVSIRQARTIFIPKVGNASRPQDFRPI TVQSVMVRILNAILASRLTSSVDWDPRQRGF LPTDGCADNTTIVDLILRDHHKRCKSLYIAT LDISKAFDSVSHAAVSATLTAYGAPKEFVD YVQNSYEVCGTTLNGDGWRSEEFIPARGVR QGDPLSPIIFNLIIDQLLSYPNEIGATIGDHT TNAAAFADDIVLFAETRLGLQTMLDTTVDF LSSVGLTLNSDKCFTVGIKGQPKQKCTVVIP ETFRIGSRSCPALKRTDEWKYLGITFTAQGR TRYSPADDLGPKLLRLTRSPLKPQQKLFALR TVLIPQLYHKLTLGSVMIGVLRKCDILVRST VRKWLGLPLDVSTAFFHAPHIYGGLGIPSVR WVAPMLRMKRLSNIKWAHLAQSEAASSFL TDELNKARGRTLAGLNEELTSRSEIETYWAN RLYMSVDGRGLREAGLFRPQHGWCQPTR
--	--	--	--	--	---	--

						LLTGQDYRNGIKLRINALPSRSRTTRGRNEL ERQCRAGCDAPETTNHILQNCYRTHGRRVA RHNCVVNNLKRILEEKGHTVHVEPSLQLETS VSKPDLVCIRDNHACVIDAQIITDGLFLDDV HHRKVEKYKRPEVISALRREFGVSGNVEVL SATLNWRGIWSNQSVRRLIAKGLISSGDSNV ISARVVTGGLYCFRQFMYLAGYTRDWT (SEQ ID NO: 1486)
R2	R2_DSe	.	Drosophila sechellia	GGGATCAGGGGTA ATTGCGAGCAGAG GGGGAGTATTTTT CTGTAATTCGTAA GTCATATCATATG GTGTGCGGAAGGG GAATTTTACTCTG TAACTCACAAGTC TCTCCTTTACTCA AGTCGACTCAAAA CCTCCTCGTGGTG GTCCCCGGTAATG CTAAACTTGTTTA GCAGCTAA (SEQ ID NO: 1242)	TAGCTAAAACGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC AATAAAGGCATAAACA TCGCAAATAATGGTAA TATATAAATTGGCTATG AGGATGGTTTTAGTAC GTAGGCGTTGCGGAAC TTCGGTTCAGATAGAG CAATGAATCGTGCATG CTAGGAAACTGAAGTG TTGACAGACCTAGTATC TTTGATAGATTTCAT ACCTCCGCGATCAAAA	FERQNFSDGLVPQRKFIHIGTTNRNNEPRSN LRNLMTTRPSVDIFPEDQYEPNAAATLSRVP CTVCGRSFNSKRGLGVHMRSRHPDELDEER RRVDIKARWSEEEKWMMARKEVELTANGH KHINKQLAVYFANRSVEAIKKLRQRGDYKE KIEQIRRQSALVPEVANLTIRRRPSRSEQNHQ VTTSETTPITPFEQSNREILRTLRYSPVECHS KWRAQELQTHIDRAELEGKETTLCCLSLYLL GIFPAQGVRRHTLTRPPRRPRNRRESRRQQYA VVQRNWDKHKGRCIKSLNGTDESVMPSQ EVMVPYWREVMTPQSPSSCSREVIQMDHSL ERVWSAITEHDLRASRISLSSSPGPDGITPKT AREVPSGIMLRIMNLILWCGNLPHSIRLART VFIPKTVTAKRPQDFRPISVPSVLVRQLNAIL ATRLNSSINWDPRQRGFLPTDGCADNATVD

					AAAAAAAAAAAAAAAAAA AA (SEQ ID NO: 1365)	LVLRHSHKHFRSCYIANLDVSKAFDSLHAS IYDTLRAYGAPKGFVDYVQNTYEGGGTSLN GDGWSSEEFVPARGVKQGDPLSPILFNLVM DRLLRNLPSEIGARVGNAITNAAAFADDLVL FAETRMGLQVLLDRTLDFLSLVGLKLNADK CFTVGIKGQPKQKCTVLEAQSFYVGSREIPS LKRTDEWKYLGINFATGRVRCNPAEDIGP KLQRLTKAPLKPQQRMFALRTVLIPQLYHK LALGSVAIGILRKTDKLRYYVRRWLNPLD IPIAFIHAPPKSGGLGIPSLRWVAPMLRLRRL SNIKWPHLTQNEVASSFLEAEKQRARDRL AEQNELLSRPAIEKYWANKLYLSVDGSGLR EAGHWGPQHGWVNQPTRLTGKEYIDGIRL RINALPTKSRTTRGRHELERQCRAGCDAPET TNHIMQKCYRSHGRRIARHNCVVNRIKRGL EERGCVVIVEPSLQCESGLNKPDLVALRQNH IDVIDIQIVTDGHSMDDAHQRKINRYDRPDI RTELRRRFEAAGDIEFHSATLNWRGIWSGQS VKRLIAKGLLSKYDSHIISVQVMRGS LGCFK QFMYLSGFSRDWT (SEQ ID NO: 1487)
R2	R2_DSi	.	Drosophila simulans	GGGATCTGGGGTA ATTGCGAGCAGAG GGGGAGTATTTTT	TAGCTAAAACGTTTGGT TCAAAACATTTGCTTGC TGTCTTGGCATAACATC	TCLAANLSGKNFSDGLVTQRKFTHIGTTNTN NEPRISLHNLMTTRPSVDIFPEDQYEPNAAA TLSRVPCTVCGRSFNSKRGLGVHMRSRHPD

				CTGTAATTCGTAA GTCATATCATATG GTGTGCGGAAGGG GAATTTTACTCTG TAACTCACAAGTC TCTCCTTTACTCA AGTCGACTCAAAA CCTCCTCGTG GTG GTCCCCGGTAATG CTAA (SEQ ID NO: 1243)	AATAAAGGCATAAACA TCGCAAAATAATGGTT ATATATAAATGGCTAT GAGGATGGTTTTAGTA CGTAGGCGTTGCGGAA CTTCGGTTCAGATAGA GCAATGAATCGTGCAT GCTAGGAAA ACTGACC ACACGCAGTGTTGGCA GCCCTAGTATCTTTCGA TAGATTTCCATACCTCC GCGATCAAAAAAAAAA AAAAAAAAAAAAA (SEQ ID NO: 1366)	ELDEERRRVDIKARWSEEEKWMMARKEVE LTANGHKHMNKQLAVYFANRSVEAIKKLR QRGDYKEKIEQIRGQSALVPEVANLTIRRRP SRSEQNHQVTTSETTPITPFEQSNREILRTL GYSPVECHSKWRAQELQTIIDRAELEGKETT LQCLSLYLLGIFPAQGV RHTLTRPPRRPRNR RESRRQQYAVVQRNWDKHKGRCIKSLLNG TDESVMPSQEVMPYWREVM TQSPSSCSG EVIQMDHSLERVWSAITEHDLRASRISLSSSP GPDGITPKSAREVPSGIMLRIMNLILWCGNL PHSIRLARTVFIPKTVTAKRPQDFRPISVPSVL VRQLNAILATRLNSSINWDPRQRGFLPTDGC ADNATIVDLVLRHSHKHFRSCYIANLDVSK AFDSLSHASIYDTL RAYGAPKGFVDYVQNT YEGGGTSLNGDGWSSEEFVPARGVKQGDPL SPILFNLVMDRLLRNLPSEIGAKVGNAITNA AAFADDLVLFAETRMGLQVLLDKTLDFLSL VGLKLNADKCFTVGIKGQPKQKCTVLEAQS FYVGSREIPSLKRTDEWKYLGINF TATGRVR CNPAEDIGPKLQRLTKAPLKPQQRMFALRT VLIPQLYHKLALGSVAIGVLRKTDKLIRYYV RRWLNPLDVPIAFIHAPPKSGGLGIPSLRW VAPMLRLRRLSNIKWPHLTQNEVASSFLEAE
--	--	--	--	---	---	--

						KQRARDRLLAEQNELLSRPAIEKYWANKLY LSVDGSGGLREAGHWGPQHGWNQPTRLLT GKEYIDGIRLRINALPTKSRTTRGRHELERQC RAGCDAPETTNHIMQKCYRSHGRRVARHN CVVNRIKRGLEERGCVVIVEPSLQCESGLNK PDLVALRQDHIDVIDIQIVTDGHSMDDAHQR KINRYDRPDIRTELRRRFEAAGDIEFHSATLN WRGIWSGQSVKRLIAKGLLSKYDSHIISVQV MRGSLGCFKQFMYLSGFSRDWT (SEQ ID NO: 1488)
R2	R2_DYa	.	Drosophi la yakuba	GGGAACATGGGGT AAAGGTGAGTAG AGGGGGAGTATTT TTTATACTCTGCA ACTCATAAGTCTT GCCTTTACTCAAG TCGACTCAAACC TCCTCGTGGTGTT TCCCGGTAATGTT AAAGTTGTTTAGC AGCTAA (SEQ ID NO: 1244)	TAGCTTAAAACGTTTGG TTCACATACATCTGCCT GCTGCCTTGGCACAAT ATCAAAAAGGCATAAA CATCGCACATATTGGTT ATTTACGGCTATGAGG ATGGTTTTAGTACGTAG GCGTTGCGGAACTTCG GTTCGGATAGAGCAAT GAATCGTGCTAGCTAG GAACTGACCAAATAAC AGCAGCCCTAGTATCTT TCGAAGATTTCATACC	FERRIFPKGLVPLTKDNHIGTTNLQNEPRIFT NDLLTTRPSVDHVPEDQYEPNAAATLSRVP CTVCDRSFNSKRGLGVHMRSRHPDELDEER RRVDIKARWSEEKWMMARKEVELMANG FKHINKQLAVYFANRSVEAIKKLRQRGDYK EKIEQIRGQSALAPEVANLTIRRRPSRSEQDH QVPTSEASPITPLEQSNREILRTLRYSPVVC PSKWRAQELQTIIDRAEFEGKETTLCCLSLY LQGIFPVQGVRLTLTRPPRRPRNRRESRRQQ YAVIQRNWDKHKGRCIKSLNGTDESVMPS REFMEPYWREVMTPQSPSSCNGEVIRTDHSL ETVWSAITEQDLRASRVSLSSSPGPDGVTPK TAREVPSGIMLRIMNLILWCGNLPHSIRLAR

				TTTGCGATCAAAAAA AAAAAAAAAAAAAA (SEQ ID NO: 1367)	TIFIPKTVTAKRPQDFRPISVPSVLVRQLNAIL ATRLTSSIDWDPRQRGFSPTDGCADNATIVD LVLRHSHKYFKSCYIANLDVSKAFDSL SHA AIYGTLRAYGAPKGFVDYVQKTYEGGGISL NGEWCSEEFVPARGVKQGDPLSPILFNLVI DRLLRALPSEIGTKVGNAMINAAAFADDLV LFAETRMGLQTLDDKTVDLSTVGLKLNAD KCFTVGIKGQPKQKCTVLEAQSFCVGSREIP TLKRTDEWKYLGIFHTASGRVRCNPAEDIGP KLQRLSEAPLKPQQLFALRTVLIPQLYHKL SLGSVTIGVLRKTDKLRIFYVRRWLNLPDVS PIAFVHAPPKCGGLGIPSLRWVAPMLRLRRL SNIKWPHLVQSEEASSFIEAEKQRARGRLIAE QNELLSRPAIEKYWANRLYLSVDGGGLREA GHYGPQHGWVSQPTRLTGKEYLDGIRLRI NALPTKSRTTRGRHELERQCRAGCDAPETT NHIMQKCYRSHGRRVARHNCVVNRIKRGLE ERGCVVIAEPSLQCESGLNKPDLVVLRQNHI DVIDVQVVTDGHSMD EAHQRKINRYDRPDI RTELRRRFEAAGDIEFHSATLNWRGIWWSQS VKRLIAKGLLSKYDSHIISVQVMRGS LGCFR QFMYLSGFSRDWT (SEQ ID NO: 1489)
--	--	--	--	--	---

R2	R2_KF	GU94 9558	Kaloterm es flavicollis	GAGAGGTCACTGT TGCTGATCAGCTG GACACCTAGCTGA CTGCGTGCATGCG TGCACGCCGCGCG CCGCCTCCTCGTG GTACCGCTGGTAA CGCAGGGTCACTG AGACTACCTGCGG CTAAAGGCGCGCG CGAGGGATTGTAA GTCCACCACCTCC ACGTGGTTCCCCG CGGGCAACGGCAT AGTTCATCTGCCG GCCAAGCGTGCCG TTCCCTCGATCCT CCTCTGATTAGGG ATAGGAGGGGGG CGGTGGCTCCCGC CGACGACCTGGCA AACACCTTGGACA CGCTAAGATAATA	TAATGTCCCTTTTGGCT TGCCCCCACCTGCTTAA AGGAACTGGCAGGAAA GAGAGTGATCCGTGCC ATAGAAATATGGTTAT CCGGGGCAAGTCACTA GCAATATGGGACTTCTC CGGGTCCGTGCGGTCCT TCCAACATGAGCTGGA CGTAGTCCACTCTATGA CTTGAACGATACGGGG GCGTATCTCCCCCGGA AGAGGTCGCCTAGGCG ACTTAGAA (SEQ ID NO: 1368)	MEVPLTSSLGSGATQAPGTPELLGEHTVERP GLDQGHSYGLLMDDVELPVRLPFFGPLICPG CRTLLTSEETISSHHRRVHPDARTRWVCYGC DSPFMTYRAIKCHLPKCSGRKVVTGDHICN GCTKRFESQRGLSLHKRRRAHPGLRNEEMLE PPVRAERRPNAHKSSIWSIDEIRILEQYEAAY VGDLHINMKIAAHLPFKTNKQVSNYRNDRR KKSRTATDASQQGLGPNDGNRGIVPSGQSSP LFLEGSDAEGDEDVFNVLPPTLGGLEPAG QVHSLSEGETSPLVGEADPCFMGGTPSAGE ASGSTLLGPDPTPADGYSLVRKDLQLSVQTS PLLAVGSVGTESVQFERGVLSCGTPPEFLHP EQFAHCANNDPVLNASEEQVHAPLGEEAND LPDNNHPSELGVPEDPTCSPATEQVQPSSE EEADDPFAQFKAWRRRVASYALKIETGVLP AQVDDLRLRLRDGDTQSKVTCEEVEEVLS LTRTILGGTAPKKRVEGR TKW TYKSRTNHE ARKRIMYARCQDLYRRRPQRPVERAVGYQ AEESLLDNQDERPSHGAFETFYTGLWGKSG QCNITMPPGVPRHTGHVLRVTPKDIYSRLR KLKKDYAPGPDGVTKLKVQSMGAYPSLLA KVYNLVMLTGYFSSCWKEHKTS LIPKDRGS PMDVSNWRPITIGSLLSRIYTGLIERRLRTVS
----	-------	--------------	-------------------------------	--	--	--

				GGCCGTCCTCCGG GCCGGCCGAATCA TAAGCACACA (SEQ ID NO: 1245)		DIHQRQVGFMPVNGCAANLFIFDECIRQAK KEGTIVGSLIDVAKAFDTPHEAILRALSSQ GVDEHTMAHIRDMYSGIRTRINGKGSIDIPLV RGVKQGDPLSPMLFNMVMDPLIRDLQRKGF RIGGHEIGALAFADDDIVLLADSIDGAQDHVD QVGRYMNKLGM TLNPRKSSSFLITAMRKT WICRDPGLSIGETKVPGARPSALKYLG VNY TLSEGLSGALIDKLMQAVNRARGLALKPL QKVNILIERIIPKFLYG IILGGPSLTRLHAADK CVRMAVKEILHLHPSTTDHVL YARKKDGG MGIPRLAHLVRLASLRSGLALLASGDVAVQ AAGMAGDLEGRCKKVANDLRLNWPVTLR DVVRASNKFKSQESKDWERLASQGHGVKD FRNDRLGNCWLYDPTVLSSSRYTDALRLRT NTFGVNVALRRADKDLEVNCRRCHGKPETL GHVLGECVAGKGMRIQRHDKMAAFVATKC EEKGYQTTREQLFSIEQGKLPDLVVIDGER ALIVDVTVRFESGNALSRGASEKIEKYQPLA DYFVSQGAVREANVLPVVGSRGAIQTATL KSLATLGLDVERVGKYLAICAVASSVEIAC MHLDYT (SEQ ID NO: 1490)
--	--	--	--	---	--	--

R2	R2_RL	GU94 9555	Reticulit ermes lucifugus	CTCACTGCTGTCA TGCCATTGTTAAT GCGTTGGGTGATG GCGGGTGATGGGA GACGAGTTACAGC AGAGCTGGCTTCA CGGGGGGGGCTGT AGTTGCCCCGAACC AGTCTGCTTCTGA GTGCCTCCACGTG GCCTCGCTGGAAA CGTCGGAGCTGCT TACGGCTAACCCG TAGACAACTCCGG CGGCCAAACTCAG AAAACGGCCTTAC TACCAGGGTCATC CCTCGCGGCGGCG CTAAGACCTTACT GTATGTACTACGG ACTACAGTTGAGT AGCGCGGGAGAG CACTTGATGTGAG	TGAGTTAGATATATATG ACGACTGATTCTTCCCT AACTAAAATATAATTT GGAATGTGCATTACTTA TACTTTATTGTATTTAT TGACATCCGTAGTAGTC TGTTTAGTGGATTTAGA TCGTAAACCATTCTGTG ACGTCGCTGATGTCTAT ATGTACCAGTGGCAGA ATGCTGACTAAGAACA AAATAATTTTAAACAA TAGACAGCACTAAGAA ATTCTGCAAACCTGGAA CGTGGCCACGGTAGG CCAATATACCGGAAAG GGAAAATGACACATCC CCCCTTAAA (SEQ ID NO: 1369)	MMADYNNSVDHALEDNTRLIFARDAVLAR VCGPFDNLECGLCGVLLTSLQGVREHCHRA HHNLDLTFQCTKCDKGFSSYRGICCHFSKCI GARISVSEGPLSCSECEREFDSKRALSTHERH MHPGIRNAKRLKDFNPRGGGGKTIHGNTK WTEEEVQLLVSLSKRFEGYKSINKEISLILTS KTCKQISDKRRYLNHNGNGGLAAAEAVLE FCEDSHPEVTESDGAVLSEIMDEECHQSSVT MRSSIVHGDIGREVQGKELVRIPPDNSVMGN CVVLLRKLATDKRSDLDSLKDKELRIDIEKA TKANRESADGRVIQSVAGNEIDPDTFQWKE LLLQVVRGFPRVDENSELFDLDDKLTKESS DSPVWNDNCELIVSDLCQVLCKKKYELGRQ HHVRKGKRHRGIHHKREKFRECQKIFRKSP RKLAEYL YRDKDL SHISKDASTPQGIEQYYS QLWGEPELLESNTIEEKLPSSSLFDCLPPITPE EVEGRIHKIRPSSAPGLDGVRKIHLVGKGITL VLVKLYNLLFLTGGYPECWKRNRNRTVFIPKIG KDLSEVGGWRPLTIGSLLARMYSAFLERRIR RVTSLSLSQRGFTNIQGCHVNLTILKEGIRQA KVKNGGVIVSVDIEKAFTIPHVSFISRLASQ GVPPLLRKIISNMYKDVYTVIEGQCIPIKRGV KQGDPLSPLLFNIAIDPVLRSLEEFQGGPLG
----	-------	--------------	---------------------------------	---	--	---

			GGTAGCAACTTGT TGCTACTTCGACT GTCTCCCTGAAGA TTTCCGAAGGGGT GCCGCAAGTCCAT GGGGGCGGTCCTG CTGTGGGAGTTAT CTTGACCCGGGA AGCCCTCAGTGGA TAAATACTCCAAA ATGGCTTAGCCAC CTCACGTGGAGAG AGGGGAATTAGG GGTTTACTCCTAA GGGCCGATGCCTC CAATCGGATTTCC CCGGGCAACGGCT ATTATCAGCCGGC TAAGTTCGGACCC TCATCTACTGATG GGATACTCTCTCT CCCCCGCCTGCG TGATTAGTTGGGA		NSAIKILAFADDIILGASSAGQAQQMVDMLG IGLTSCGLGVSHRKCFGFQIVNKNKTWTIVD PMITLNGSSLPFSGPEDRLPYLGVDINPWDR KSRVDAGQRLISAARKGSQLSLKPQQKINLI TAFLLPKFLYILIEDPPSPAYLKSIDHDLRQIY KNILHLPNCVSTAFMYSPKRDGGLGLPRLSC LVPLAHLKAGIKLGS LQDSL VREITTSDRFV RTMGSLVAHSLWASWSLTLQDIYKLKSALKR REAKAWESCVSQGQGAQFRGDSIGNNWL HNP GTYRPGQYIEALKLRANLTGVRVNLKR SGYNVPITCRFCKDIPETQAHVLGLCPKTKG MRIQRHDSIVNRVRDKLKTSPVALMHEQN FTVEEGQVFKPDIVTILGEVGYVIDVTVRYD DRDYIKDASVEKIRKYEALKGYLKDLYPQL NKVEVLPLVFGSRGAVPGSTVHNMGLLGFT KREMVHISRKVIADSLIISNFLEVY (SEQ ID NO: 1491)
--	--	--	--	--	--

				ATGCACGCGGGCG GGTGATTCTGACA AGCCCAGCCCAA AGTTCTTACCACG ACGTACCCGGGGG TGTCACCCGAACT AATATATTTCCG CCGATCGGTTCGC TGTGGGGGGATGT CGGGGCCTGGTGG CCGGGAACCGTCA GCCGCTCTTTGAG TGTCCTGCGTAAC CCGGAGGCGGTGA CGTCGAAGGGCTA GGCTAGCACGGTC CAGACCCCTGAAC GCCTGGGGAGACG GGCCCACCACCTG AGTAGGAGTGGGT CCTTTACCTCATTT GAGGTGTCTCCTC GTTCTTGTAGGGG		
--	--	--	--	--	--	--

				GCGAAGGACGAG GAATGCATCCGTC CCTCCAGGATGCT GGTGGTTTCCGTC TGGTGGGCTCATG CTGAAACGCGCAG GTGCCGATTTGTC GAAGAGGACATAT GGGGTAACCCATA GAACCTAGGCAGG AGTAATCCCTTAG CTGGGGGGGTCAG CTGGTGGGTCCTG TCAATTTATCCTC CCCTCCATGCCAG AGCCGGTCCGAGG TTAGAGACGGACT CATTTTCTCCTTTT TATATGTC (SEQ ID NO: 1246)		
R2	R2_RU	GU94 9554	Reticulit ermes urbis	CACACTGCTGTCA TGCCATTGTTAAT TCGTTGGGTGATG	TGAGTTAGATATATATA ACGACTGATTCTTCCCT AACTAAAATATAATT	MMADYNNNSVDHALEDDTRFIFARDSVLAR VCGHFDNLKCELCGVLLTSLQGVREHCHRS HHNLDLTFQCTKCDKGFSSYRGICCHFSKCR

				GCGGGTGATGGGA GACGAGTTACAGC AGAGCTGGCTTCA CGGGGGGGCTGTA GTGGCCCGAACCA GTCTGCTTCTGAG TGCCCCCAAGTGG CCTCGCTGGAAAC GTCGGAGCTGCTT ACGGTTAACCCGT AGACAACTCCGGC GGCCAAACTCAGA AAACGGCCTTACT ACCAGGGCCATCC CTTGCGGTTGCGC TAAGACCTTACTG TATGTACTACGGA CTACAGTTGAGTA GCGCGGGAGAGC ACTCGATGTGAGG GTAGCAACTTGTT GCTTCTTCGACTG TCTCCCTGAAGAG	GGAATGTGCATTACTTA TACTTTATTGTATTTAT TGACATCCGTAGTAGTC TGTTTAGTGGATTTAGA TCGTTAACCATTCTGTG ACGTCGCTGATGTCTAC ATGTACCAGTGGCAGA ATGCTGACTAAGAACA AAATAATTTTAAACAA TAGACAGCACTAAGAA ATTCTGCAAACTGGAA CGTGGCCACGGTAGG CCAATATACCGGAAAG GGAAAATGACAAATCC CCCCTTAAA (SEQ ID NO: 1370)	GARISVSEGPLSCSECERKFDSKRALSTHER HMHPGIRNAKRLKDFNPRGGGKTIHGNTK WTEEEVQLLVSLSKRFEGYKSINKEISLILTS KTCKQISDKRRYLNLNHNGNGGLAAAEAVL VFCDDSHLEVTDSDGAVLSEIMDEEYYQSSL TMRSSIVHGDIGREVQGKDLVRIPPDNSVMG NCVLLRKLATEKRSDLDLSKDKELRIDIEK ATKANRESADGRVIQSVADNEIDPDTFQWK ELLGQVRGFPRVDENSELFDLDDKLTSELS SDSPVWNDNCELIVSDLCHVLCKNKYELGR QHHVRKGKRHRGIHHKREKFRECQKIFRKS PRKLAEYL YRDKDL SHISKDVSTPQGIEQYY SQLWGKPELLESNTTEEMPLSSSLFDCLPPIT PEEVEGRIHKIRPSSAPGLDGVGKIHLVGKGI TLVLAKLYNLLFLTGGYPECWKRNRVTFIP KIGKDLSEVGGWRPLTIGSLLARMYSAFLER RIRRVTSLSSSQRGFTNIQGCHVNL TILKEGI RQAKVKNGGVIVSVDIEKAFTIPHSVIFSRL ASQGVPPLLRKIISNMYKD VYTVIEGQCIPIK RGVKQGDPLFPLLFNIAIDPVLRSLEEFQGG PLGNSAIKILAFDDD IILGASSAGQAQQMVD MLGIGLTSCGLGVSHRKCFGFQIVNKNKTW AIVDPMITLNGSSLPFSGPEDRLPYLGVD TNP
--	--	--	--	--	---	---

				TTCCGAAGAGGTG CCGCAAGTCCATG GGGGCGACCCGGC TGTGGGAGTTATC CTGCACCCGGGAA GCCCTCAGTGGAT AAATACTCCAAAA TGGCTTAGCCACC TCACGTGGAGAGA GGGGAATTAGGG GTTTACTCCTAAG GGCCGATGCCTCC AATCGGATTTCCTC AGGGCAACGGCTA TTATCAGCCGGCT AAGTTCGGACCCT CATCTACCGATGG GATACTCTCTCTC CCCCCGCCTGCGT GATTAATTGGGAA TGCACGAGGGCGG GTGATTCTGACAA GCCCAGCCCCAAAA		WDRKSRDAGQRLISAARKGSQLSLKPQQK INLITTFLLPKFLYILIEDPPSPAYLKSIDHDLR QIYKNILHLPNCVSTAFMYSPKRDGGLGLPR LSCLVPLAHIKAGIKLGSQSLVREITTSR FVRTMGSVSHSIGASWPLTLQDIYKLKSALK RREAKAWESCVSQQQGAAQFRGDSIGNNW LHNPGTFRPGQYIEALKLRANSTGVRVNLK RSGYNVPITCRFCKDIPETQAHVLGLCPKTK GMRILRHDSIVNRVRDKLTKSPVALMHEQ NFTVEEGQVFKPDIVTILGEVGYVIDVTVRY EDRDYIKDASVEKIRKYEALKGYLKDLYPQ LNKVEVLPLVFGSRGAVPGSTVHNMGLLGF TKREMVHISRKVITDSLIIISNFLEVY (SEQ ID NO: 1492)
--	--	--	--	--	--	--

				GTTCTTACCACGA CGTACCCGGGGGT GTCACCCGA ACTA ATATATTTTCCAC CGATCGGTTCGCT GTGGGGGGATGTC GGGGCCTGGTGGC CAGGAACCGTCAG CCGCTCTTTGAGT GTCCTGCGTAACC CGGAGGCGGTGAC GTCGAAGGGCTAG GCTAGCACGGTCC AGACCCCTGAACG CCTGGGGAGACGG GCCCACCACCTGT GTAGGAGTGGGTC CTTTACCTCATTG AGGTGTCTCCTCG GTCTAGTAGGGGG CGAAGGACGAGG GATGCATCCGTCC CTCCAGGATGCTG		
--	--	--	--	---	--	--

				GTGGTTTCCGTCT GGTGGGCTCATGC TGAAACCCGCAGG TGCCGATTTGTCTG AAGAGGACATATG GGGTAACCCATAG AACCTAGGCAGGA GTAATCCCTTAGC TGGGGGGGTCAGC TGGTGGGTCCTGT CAATTTATCCTCC CCTCCATGCCAGA GCCGGTCCGAGGT TAGAGACGGACTC ATTTTCTCCTTTT ATATGTC (SEQ ID NO: 1247)		
R2	RaR2	FJ461 304	Rhyncho sciara american a	CAGAACGTGGAG AAACGGAATAACT ACCCAGATCCGTT GGTTAACCGGTGG CAAAGTTAATCAA GGTTGCCATAGGC	TAGAAATGTGTGCGAT AAGGTGTGAATAGAAG GGTTACCAAGAGGGA GACCTAGTTTGGACCTC AGAAATGGGGTCATAG GAGTGATAGGTTGTAA	MSNYNETNTSGGDNPRMATQTTGSLSSGPI NQHTCELCCRTFGTRAGLGQHVRKTHPIES NQSINVERKKRRWSPEEIRRMANMEAQATI NNIKHLTQYLATYLPQRTLNAIKGRRRDAE YKELVTGIIANLRSNSSTQQTNQVCNESEMS QRSKILQSIRESVRDLRSRRNKYAKALQELG

				TTAATAACCCTAT GGAAATGTTTCCA CACACCTCCACGT GGTGCCTGCCGGA AATTGTTCTAGGG TGAACAAGCTAAG TTGTGAGAAACGG GCTCCACCACAAT ATGGAGCCTGCCA GGGCGCGAGACTC AGGACTCTCCATG TACAAAGTGGTTA GTTGCAAAAAGAG TGCGCCTAGCATG ACTGATAATTTTT CACTGAAATAACG TTGAACTTTATCT GTGTCATGTGCAC AACACTATGGTGT CTGATCAAGCACC ATCAGTGGTGGAC CTGCTAATGTATT AGTAGAACGTGTC	AGCCGTTGGGGAATCC GGCTACACATGGTATCT CAGGAGCCATTCATGC GCTGATCTCATTAAGGC GTAATAAACTGTGAAA CAGATCCTGATAATGC CGTGCTACCAAATGAT GTAACGAGGCGGAAAT AAAATTAATCTGGGGC GTTCTGCGGAATGACT ACTAAATATAGCGATG CTATATATACAAACGA CTGATGGTAACACCGG CCTTA (SEQ ID NO: 1371)	EAALCGKMLNEEQLIHCISMFNTAKCPKG PRFRKTATHSGTNKQQRQQRYPARVQKLYK MNRKVAAMVLEETDKIQIKLPDHDPMFKF WESEFKEGEGMPERMPKDLKESPDLKAIWD PVTEEEVRKAKVANNTAAGPDGIQPKSWNR ISLKYKTLIYNLLLYYEKVPKHLKVSRTVFIP KKKDGSSDPGEFRPLTICSVVLRGFNKILVQ RLVSLYKYDERQTAYLPIDGVGTNIHVLA LNDSENTKLSELHVALLDITKAFNRLHHTSIK SLVGKGFYPYGFITFIRRMYPGLQTMQFEG HCKMTQVNRGVYQGDPLSGPIFLLAIEKGL QALDKEVGYDIGDVRVNAGAYADDTDLVA GTRLGLQDNINRFSSTIKQVGLEVNPRKSMT LSLVPSGKEKKMKVETGKPFRAVDVPLKEL SINDFWRYLGISYTNEGPERLSLTIEQDLERL TKAPLKPQQRIHMLNAYVIPKYQDKLVLSK TTAKGLKRTDRQIRQYVRRWLKLPHDVPIA YLHAPVKSGGLNIPCLQYWIPLLRVNRVNKI TESQRSVLA AVGKTALLTSTVYKCNQSLAT LGGNPTMLAYRTYWEKELYAKVDGKDLQN ARDDKASTRWNGMLHSDISGEDYLNHYHKL RTNSVPTKVRTARGRPQKETS CRGGCKSTE TLQHV VQQCHRT HGGRTL RHDRIVGLLQHE
--	--	--	--	---	---	---

				CAGGCGATAATGC ACACACGGCCTCC GGGCCCATCGCTT TTTTTGAGATTCC CTAGAAACTTCAG TGTGTGCGACAAC TGTATAACCCATA AGGGATGGACAA AGGTTATACTAGG GGGTAAAAACCCT AATCGGCTAATGG CAAATGGGATGTA GAAATGCCAAAG ATTACTCGCACCG AATAATGGTGGCC GAAAAGCGGGTA ATCGAATGAAATG GTAATGCTGTAGC GGAAACATGATCA CATTCTGTGACAG TAAACCATTAGAC CTAGGGGGAACTA TGATTAACAAGAT	LRRDYNVLAKQELKTGIGLRKPDLVLIKDD TAHIVDVQVARCSKLNESHVRKRSKYDKKE IEVEVKSRYRVSKVMYEACTISYKGIWDKQ SVMSMRRLGVSEYCLFKIVTSTLRGTWLCW KRFRNMITSVRS (SEQ ID NO: 1493)
--	--	--	--	--	--

				ACCAGCTTACATG GAAGCAATGAAAT GTAAGTCACAGTA GTGATAAGTGGTG AAGAGTCTTGTA TCACCGTAACTAG GCCAGGTTCTGGG GATGCCATGAACT TAGGGGGAGTATG GTTAGCAGATCTA CCAGCTAACACTA TTACTGATAATAT GTAAGCCGCAGTA GCGCTAAGTGGTG TACAGATTTGCAA CCACCGTAACTAA GTTCTGTTTCGAT GGACTAGGGGGA ATCATGATTAACA AGATACCAACTTA CATGGAAGTTATG AAATGTAAGTCAC AGTAGTGATAAGT		
--	--	--	--	---	--	--

				GGTGAAGAGTCTT GTAATCACCGTAA CTAGGCCAGATGT AGTCAAAGCACAT GTTTAGGGGGAAC AAGGTTAACATTG GTAAAAGACCAAT GCAACCTCCGTAA CTAAACGTGAAGG CACAAAACATAAAA GTCCAGTGGATGA CAGGTGAGGTCAT CACTGGACCCAAA TGTTTTAAGCTCA TCATAACAACACG GTGAAAAATCCAG CATTTATTTGCCT GATTGAGTAGCTT CCACACTATTCCA AAGCCGAACCTAT CTGGGTTTTTCTTG AAAGGCCGTATAG GGCATATGTCGAG		
--	--	--	--	---	--	--

				AAATAAGTCCAAG GTGAGGTAGTGTG GCCCTGTACCCAG GGGTAAGGTACTA TACGGTCGAGTGG CTCAGTAGGCCTA ACTAGCCACTGAG TCACTATAATGAC TAGTT (SEQ ID NO: 1248)		
R2	YURE- 2_Cis	.	Ciona savignyi	CCAAAATTACTTC CAGCACCTCCACA GCAGACGAACGA AGAAAGAAGACTT ACGAAATAAGAAT AAGCAGTTTAAAG ACGAAGCAGACG AACACCACTCCAC CAACGACGCTCTG CAGCTACACCACC ACCATCTCGCCGC AACAAGAAGAATT TTCCGCTGCTCTC	TAAGATCCGCGGCTGT GGCGCCGAACGAGCAC CTGCCATTCTTCTTGT AGGGACTTTTTACCCCT CACTCCCCCAATAGTT TTTTTTCGTTTTTTCGTT TTTTACCCCCACCCCA CTCGCCTCTGGGCTGCA CATCCCACACGTAGGG ACCTGTTTATATTATTT GCCTTTTATATGTACCA CTTTTAAATATATTTT TGTACCCCAAGATG	MAGHKITMSEGKLEAVRYGGVRNVSYE CPVPDCTKTFSQANNLIRHLNCFGNTKHRA HNFTYFFTCEKCKIQIHSNTKHNSNHYKQC CATGGGPSCETGQYFCPACEQAGLGNLESA LRHFQSSHPEFNLPPRSQFSKSHPNSTLSLK PKDHLMKILYSGPLTPGQLVCPIKICLRSSAA RLFHDVSKLRKHMLVDHNRTL VYETTCGK CLRPVDTSKNMRKTTSHFEKCSGESFISSPSP IPQKTYKLDLPSTSTPPPRKSPKLQPYKPIRTF KNPLTKSSQSKSDNPPKPTPFFSPRTLERSAS WPALSEVVDPLPKLKEKHPSLPCALDKCPPS PRIKPSTLVPPCHTANNSPKPTSPESPSTLKPL PRPIRPSKPLEDWLTVRSVGPDPREIVLNIGPR

				CACTCTGCTCCAA	CTTTTCGCCAAAAAAA	PRPGPAAGSRTTSPSTAPAKRVAANPIAAP
				CACCACCTCTCCT	AAAAATTTTTGTATCAC	LSGEPGATLDCGQTGRKVQPPKKRPTESAG
				GCTCTGTGGACTG	ATTTTTATATTTTGTA	SLPPPAEPATDLLTGREGLARLVEEYHLSGD
				CTGCCTTGCTGCT	AACACAGATTTTTATAA	FGAFCRDLERWTALSSNRRPKPRRGRYNR
				GGACCAACCTCTA	ACTTTGCACTATTTTTA	GAAARATRNRGRDDRQDPQDRDDQGGPGP
				CCCGAAGGAACCC	TATAAACTTCGCACTTA	VTCGRPQRYKRAAALRSAFGRDMKATVRRI
				TTCGAACCCAGCA	TTTAAAATGAATCGCAT	IDGERGDARCEIDPKTIEGRFRDELSPPVREG
				GCAAGGTACGTGT	CTTTTTTATATACACCA	PECSLPPWMAEAQAGEHAPSNDSPGDAYD
				CACCACCTCTCCA	ACACAAACAGGATGTG	GPITALEVEMVLSTLNVGSAPGSDGLSYGF
				CAGAGCCAAGGCC	CAGCTCAGGGGAACCA	WRALDPKGLVLSELFVCRIERRVPGPWKSS
				AGAGTGGATAGA	ATCCTGCGTCCCTCCTA	RVTCLICKDAEGDLDDLGNWRPISICQTVYKI
				GCAGCGGCCTCTA	GCGGCGGGAGGGGCGC	YAAVLARRLQSWALDGGVISRSQKGFMPFE
				CAACCAAGTTCTC	CCACCTACCCCCACGCT	GVYEHVFLLDVSVADARATRRSLAVCWLD
				CACTCCGACGACA	CCTCTTGAGCACCAAC	LRNAFGSVDHTTIVEALSFRGAPAGLVEMIS
				AAACACCTGCCTC	AGGGACTCCCTTCCGG	DIYTGGSCRIRTRAGFTPDIPVGRGVRQGCP
				GTGGCCAGAGTCC	AGCCCCTGCACCCTCA	LSGIIFNLVMEVLLRGVEANNACGYRLSCA
				TGCCGAAGAAGA	ACTTTTCTTATTTTTAA	GGASVRVLAYADDVALVGSSRAEXKIQLGV
				ATCTTCGACCCCA	AAAAAAAAAATCATAT	CERFAAWAGFSFNNKKCAAMVLKHQRGGR
				ACACCGTCTCGGC	ATATTGATCTTGACGAC	RLLDSAPLRLCGEEVAILGPDSFYKYLGAHT
				AGGGCTCCAAGCA	GGGGGCTACATTCAGC	GYGRQTGGQLVDRVERQVVRLFTSFLTPTQ
				ACGACATCAGCCA	CCCCAAAACCCACCC	KLSALKRIVLPAMSFHLRVRPCAEGHLRRLD
				CGAGTGCCACCG	ACCATCCCCAACGAGT	NTVRRCVKTALRLPKGSCRAFFHTSPDAGG
				GAGTGGACGCGGC	GCCGGGGCATTGAAGA	LGITSVVAECDILTVTQAFKMLSSPDHLVSL

				GAACCCGAGGACC GCTGTCGCCAAGA ACAATAACATCAC CGCCTGCAGCAGC TACCAAACAAAGG TTAGTCTCCTACC TCACCTACAACT CGTCTGAATAGAC GCCCCGCGTGGG AGCTAACCTTGGT AGCTAGCTGCAGT GCCCTGCGACAGC GGCCTCGAAGAGC TGCAGAGCGTCAG CCTGCTTCGACCT CGCTTGTTCTCATT TCAACCTACTCCG TCCTGTGGAATCA AAAGAGCCCCACT AACAATTACATTC ATAAAATCTAGCA AGAACGAAGAGA AGCGACCAACTTT	GCTCCGGCACAATTAG CACTTAGCTTATTTATT ATTTTTTGTCAACATTT TTGTTTTTTCAAATTTTT TCACCCCTCACCCCCAC CCTAATAGGTCCCTCGG GCTTGGGCCCCTTTTTC GTGCTCGAGAAGCGTC ACATCGCCCCACTGAC CACGACCTTCCCCGAC ATTGGAGTCCTTGGCGT CTCCCAGGTCGAAACA GTCCCAAGTGATAGCA CCTAATGCTCGACTTGT TTCGGCCTGGGCCGCC GAGGATTCCCAGAACG ACCATTCCTTCTAAATAA TATTTATATTTTCAAGAT AAAATATATATATATC GTTGGCGGGACTTGTCC CGCCTCGATACCGAGT GCTGCAGAGCGGCAA ATAAAGAAGAAACCGA	VAKGRLGMHAARMGRSETASACAMADYL SGDSVMGHXSWKTGYRMPADLWTATRAA SRRLSLRFSPQPQGEFGLESGTFKIAPRERRS LTRRLHHRQNLWWRNQWAALPNQGKTVA AHSAYAASNWVKGPSSLAPQALFFGLKAR LNQMPTRSVKACYSRAPNYDKSCRRCGAE VETLPHVLNHCPSMKSLERHDSVLAEVLA AIPRGTFASVDVDRTSREHFRRVGEALRPDI VARRHDGSVVVADVTCPFESCASALDTAAA RKIEKYDQLCANLRQLYRKPVESHALVVG LGSWGRTNNTALAALGIRGAVRSRLAKQLV NLSVEGSHNIWLRWSGGIPKDLVR (SEQ ID NO: 1494)
--	--	--	--	---	---	--

				TAATCCATAACTT TTAGATCTTTTTAT TTATTACTGTTTTT AAGCCCTAAGCAT ATTGCCTTTTTTTA GATCTTATAATTA TAAAAATAGATTC AAAGTTAACACCA CCAGGCCGCTACA GAGCATTTTATTT AATCAATTTGTTA CCGACCTCCTGCT GCTTCTTTTTACTT TCTCCAGACTACT ACCCGGATACAAC CCTTGGAACGAG AGGA (SEQ ID NO: 1249)	CGTCGCTCTGCAGCCA AGGACCACCCAAACTC AAGCCAGCACCGTCGA CAACCAACATCCTCAA GTCGGCGGTTGCTGGA ACAACTCATAACATCTT CAANATAAATTATCAC CCTGTGCAGCAGGAGG CCGTGCTTTTAAACTA CTCTGTAGTGGCTCATG ATAATATTTGCTCCTT TTTTGCCCCGTGTAAAC TTAGTNGATGCGAATA AAATCAGTTGAATCA (SEQ ID NO: 1372)	
R4	Dong	.	Bombyx mori	GCTAGCTCCCTAA AATCCTACCTTAC GTCCGAGGCGAAC ATCTGTCCACGTG GGGAGCGGAAAC	TAAAAACTAGCATAAT TATTAATCATAACTAA TGTATATTACTTGGCCA AAAGCCCGTATATACA GTTCCACCGGCTCTGTC	MLRRGRIFLPASTKAGKTRGRMKWSREVN L FIMRTYYYVTKLETDLTIYRKKLHEHFS LKY PNVIISQQRISDQKRAIERNKLLSQETL DRLK EEVRKQLEDEQTNNVENEKLNSETYSHE YT TLTPQTILTKKTQQHTNIISSTQTS HSSTQTES

				GCGTACTATCGAA ACTTACGCGGCTA ACAAGGTAAAGGT AACCCATTAATAT GGAGACAAGACT AAAAAGAAAATT GAGAGGGCCGCTT CCCGGGGGCGATC GCCGGGGCACACC TGGAGCTGGCGCT GGGTGTTCCAGCA TAGGATCCGTCAG CGGCGGAGAGTTG AGCAGGCGGGCTC TCGCCGGGGATTT ACAACCCGAAAAT GCTACAGCCGCCA ACAGAAATCACGA TGTAGAAAGTAGG AGCAATAGCCCAT GTGAACCTTACAG CCCGAGTACCGGT TCATACAACCCCT	GACAGACTGAACTGAG AAAGGGGAAACATATG GAAATAATAATAA (SEQ ID NO: 1373)	ITLLENEVDILNTNPTEGATQTQEVDKDFE TNLTMYSGMDPKARPLPKLYSSKLNELIR LFNNDILVDYISPDQLSDVHTLTYCTAVTIS EQLKYKIIAIEGNARHKKNFKPPWQQRLEK DIAKL RADIGKLTQYINNNRSKKVVQSVEQI FKNTKIHTSHENG NKKSQEFLDTLKQKLAL KAHRLKRYNNSQKRKNENTIFLTNEKLFYR NLIKPKTDRDNSNIDIPTAEQLEMYWARLW ENSAKHNDKANWITEEKERWDTIEEMQFDD VTEEEIT TITARLHNWKSPGIDKIHNFWFKK LICLHKTI AKNLTDIISGNQSIPEFIATGITYMI PKGDFSIEASQYRPITCLPTIYKILTTVITKKIN SHIEHNNILAEQKGCRRGHMGCKEQLIDS TIMKHATTKNRNLHCTYIDYKKA FDSIPHS WLIQVLEIYKINPIISFLRNIMTHWQTTLKLK NPPNFVTTRQIAIKKGIYQGDSL SPLWFCLAL NPLSHQLHNDRAGYRIKQQDNTETIISHLIY MDDIKLYAKNDKEMKKLIDTTTIFSNDISMQ FGLDKCKTVHIIKGKVQPGDYTIDDTNTITA MEPSDLYKYLGFQQLKGLDHITIKQSLTSEY KKRINAICKTKLSGKHLIKALNTYAIPILTYS FGIIKWSKTDIEQIERITRTTLTKHNNLHPKS AIERLTIKRQDGGGRGMIDIWHLWRKQIHSLK
--	--	--	--	---	--	---

				CGGTACAATCATC ACCATCATCCTCG GGTCATAGAGGCT CACCAACGTCGAC TATGG (SEQ ID NO: 1250)		TFFYIKSDLSEIHRAIAQNDNNYTPLNLKQK ELIDNTENLRNRNPQKDMEENWKKKALHG RHPHDLSSQSHIDSKASNMWLKTGSLFPETEG FLIAIQDQVINTKNYRKYIIKDPTIRDDKCRK CNTQPETIQHITGACSTLTQTDYTHRHNQLA NIIHQQLALKHKLIQNTNTPYYNYKPQTVLE NDSCKLYYDRAILTDRTIHYNRPDITLQDKN NKVTYIIDIAVPNTHNIQKTFTEKMTKYTEL KEEIVRIWKQKKAYIVPIIISTTGVPNHHNS LKLLDLKDNIFISLQKAAILNTCRIVRKFMQL EENQTYTYTQ (SEQ ID NO: 1495)
R4	DongAG	AB09 7127	Anopheles gambiae	GAAGGCTAACCAC AATA (SEQ ID NO: 1251)	TAACATCCGGTGCAAA CTCATTAACATTAAGA AAAGAGAGAGGAGAA ATGAGAATGAGATTCA TTCACCTTTGGCATTG AATAGCCCGGGGTAGG TGAAAAGTTCCCAGCA TATTGCTGAGAAGTGA CAAAATTCGGATAATA ATAATAATAATAATA TAATAATAATAATAAT AATAATAATAATAATA	METRSMRKRTTRLPEEGAPTGAGPGTGDR SIQRLEDEM VQERSFSQRALPVPRTQNRNGS PINHQGNAASANVAVADRQQSLILAGGRRQ RIMWTREMNHYVIRCYYVYTRMETDMPGR VKMLGMFNDRFPRFAHQDL SKLYIRQRAII LPEELEFIKLEVRREFGEEEAGWRESSRISAR LNTIDQNTSRASEDRDLDEPTAPGLSVDIQH QMATAVTQFHGTDPLSRHRLPKLHYSYRLK TAVSIINQDVLQPQYLD SVGSIEDLQLIVYSAA VAVVRTLWLR TYPQGDSEGRPCSKAEKPA WMRRLENRINATRTKIGRMQEYQRGNSSM KVVRQIAEMVKPKELRDLTDANITEVLDIHL

					TGCATAATAATA (SEQ ID NO: 1374) QRLSALAKRLRRYAEC SKRKEQNRMFNINE REFYNWIRNDKPNFREGLPDIGDFTQFWAN LCEKPVQHNSEGMRLAEDERFSDGIEDMPV LVVNAQDIREATQYTRNGAAPGPDFVYNF WYKKLITIHEQIAACFNTVLEDSRKLPKFITG GVTYFLPKDQNTKNPAKYRPLTCLSNLNKV LSSVITQKVKDHC DTNNVMTEEQTGRRKNT QGCKDQVIIDAVIVGQA AKKQRNLDMAYID YKKAYDSVPHSYLLKVLQLYKVDGNVIKL MQHAMGMWSTSLHVT DGKVVLRSRSLNIR RGIFQGDTFSTLWFCLAMNPLSRTL NQQCNF GYLLKSEEISTRITHTFFMDDLK LFAETVQK MHLLKNVQGFSNDIKMEFGIGKCRSIHLH RGQVLDADSFRANEQEEIRH MVQGETYKFL GFLQLRGIHYAVIKKELQDKFLHRVSCILKS FLSVGNKVKAINTF AVALLTYSFGVMKWSN TDLEALERTIRVVSTKHQMRHPKASVERVIL PRKIGGVGIIDIQALCISQIHLRSYFVESQNR HELYRTVYKADHGLSALHLAQQDYQLNCNI KTVDGKGATWKQKELHGTHTHQLNLEHID KVSSSTWLVRCDLFCETEGFMVAIQDRVIAT WNYRRCILREDVEDRCRKCNSGGESIEHVIA GCPVLAGSAYLDRHNDVAKIVHQQLALRH
--	--	--	--	--	---

						KLVERFLPCYRYLPDPVQENDCIKLYWDREI ITDILIRANRPDILVYEKRKKRATIDIDIAVTL DHNVQTTTFSTKVMKYHDLAEELKQTWYLE DIRIVPVIISATGIVPMALLRSLDELELQRELP RIQKAVILRTCSTLRRFLNPYN (SEQ ID NO: 1496)
R4	R4-1_BX	CAD V010 08175	Bursaphe lenchus xylophil us	GGGATCCTGGGTT CCTACTACCTCGC TCCACCTCCTCGC GATGGATCCTGGG GAAGTCTCCGGAC TGAGCTAAGAGAG CGTTAAAGTAGAG GGTGACGGCGTAA GTACCTCCAAGTT GCGGTGGAGCGG AACATCTACTCTT CGGAGAGAGGGG AAGCTCTATGGCG GCGTTAGAAAGGT TGGACTACGGCAA CGCCAGGGAGATG GGGAAGGTTTCATC	TAAGAAAAGCATGAAA TAATAAGAAATCAGAT AAGAATAACAAGAATA CTAATAAGTATATCATG TAACTATGACAAAAAG AACGCACCAATAAGAA CATGCTTGAGTGGCCA GCTCTGCAGGCAAAAG TCGAATTTGGAACAGC CGGTAATGGAAGACCT GCAACAAACGTGGGGT AGCAGGCAATATGTAA CTATGACAGACCAAAA CTCCGAAACTCTGGTA ATGAGCCCGTGCCCCC CAAGCATGTGGTCTCGT TCGATGTAGTTAGGAA	MTCNNAVVFPPADGNPAGTADRNF AIRFPS SEPPGPSGIRPSEPLDGRTGIGDVEHAQAGN GGFLVDVLEYKEAHRYGSKCEFCYVQTKGT VCSKPRTDAWLKCEILFLLHHAYTANQNKSI ELAESAFRRAGITRRSKATIAKRWSLIQRGK GTDYKEYWDEYFEKFRYECNPTPIVRRKRN RLAAGLQSPSSVPNGYEFERKRTCETPLDTK ASSLPLICNLLTGIVGVENVEENMSVECTEP KELSGTANSSVPGLAEGVYERRHNNVNEPA AGCPQDVPVANNLIDSPTTNDRLAEFKAQ LDRAERSYMRRRLPRLKNLSPDERMWIGTT VERLRLETVSEPVCEQWRLANAGLYAAIRSI AVMRPLDAAREAHKTWLLNMKMTERKLR QQIGWVETTRRTKNEARTERQEIVYRKVAK LRRERFPMDLDSVSVHLKRKLELLKGRIQV RTAERLRRDTREAAGPYGKTALRGQGFAPN VKDATQYW SGLAQPSGQKCSSENSAILSDWK

			AGGTGATACTAGT TCGCTACTGTCAT TCGATGTATCCGG AACCATACTCGCC AAGTTGTGAACTA TGTGAAAGTCTGG ATCCAATCCAAGA CCACGGGGCGCAA TTAAAAGGTGTGA AGCAGCTTGCTGG TGATCACCACGGT GGTACCTACACCG GCGGGGGAACATC TTTAATGCCGAGA TGACCGCGCGAAT AATAGGGAGCCG GCAAAAGCCGAG CGTAAGTGGAAG GATACAGAAATT GCTCAAAAGTATA CCAACCCGCCACT TATAAAAGACCGA GAAAAGGGTACC	CAGTTCTCTTTAACCCG TGATGATTACGCCCTGT CTTAAATGGCAGGTGC CACCAAATACCGAACA CTCGTTGAGGTATGGTG GTCCGAATGTGAAGCT GGGAGTACAATTTGGT ACGAGAGCACCAGCGC CCCCGATCTAAGTGAT GACGCATGCGTCGGAA CAATGAAGACGGCTGG CAAACATTGAGGAGTC GCAAA (SEQ ID NO: 1375)	ELVECNLSSLPDQMEPLVVQGISRASPWKSP GPDGIFNYYWRQDFIVDWLKQLMLDSLRTG HYPWKLSSGRTVLLYKDGDPKAENYRPIT CLNGCFKMINSVVSEVILKRVENTIALPIEQ MALRRKVVACVESQIWDQIKQRKLSDRTO KCKVAWVDFSKAYDSLNHDAIKFVIGVLKL PTGINNYLLDSMQNWSTRHLELKSSGKVVRG PSYPIKRGVLQGDSLSPTLFVVVTSIIVRHIKT IESSDIQMYMDDIKLYGKDQETLTRLIKELQ TVSNKLGLCMNLKKCAILGDDLPEEINGIEH LKESYKYLGVPPQREITQVRATMAALEKKILT EVDTSLGAAELSYRQRISRVNSKIAPLVRFV VQSM LVTPRDVLKVYNRLGGIDVEIRRLV KYEIRYKKS NVARLYLDRKVGGIGFVNLCRI MVEA VAARAVYCRLAPSFNEFQDFLAEQNT SPITAAQTILDKCGINIELSTSTLGDVKKIVRN HYHELWLTAWKNTGLYKRWENDHVDIKRS SLWINRGNLSANNARIGIGIQDNSIFCRGFVG NKCDTKYCRLCGDGIESVSHIVTGCPTHRTN LYIERHDCVARNVYAYLAIRYGIPVPHYTQR VKTIEKNGDQSVELYWNYKFPCTRALEACR PDIVLIDKVSKRTHIEVAVSWRGRLQEMVD RKVYKYTVNGEYEADGSSRGWNIVRELND
--	--	--	--	---	---

				ACTAGAGCTCATT ATAAAACATCT (SEQ ID NO: 1252)		QYGFPEVYTLVIGAGGEILPCTVKDVERLT GGAATDNLIERMERSAVLGSCRIIKRHLAL (SEQ ID NO: 1497)
R4	R4- 1_HG	.	Heterode ra glycines	TGGCGATACTCGG AACCTCCGGGGAG CCTGGTAGGAGTT GGCCTACAGGTCG CGAAAGTCCCTAG GTGCTGCACGGGT TGCGCTAATCCGA GGGCGCTGGGTTA CCTTCCCATCGGC CAAAAACGTCTGG GCCTTCTTAGCCG CGGGTCCAGTATT TCTGTTGAGCCTG ACAGTTCTTCCCG GATATGGCGAAAG ATTACAGGGCGGT ATTTCGTGAAACC TAAAAWGGTCG GGCCGAATGGCAC GGACAGACTCACT	TGAGGACTCANAATTG ACAATACACCTCAGA (SEQ ID NO: 1376)	MISCDLERETLTQMALFRARSDKTPTHAGIP APDEVREGGCGQNRTNPAAPRGKAAAIQRQ NGITIPXACAQSGLVRTQRVQQWSAVEESA LKDVVVRNTDDRGLINWAKGVLPEWQRLC QLNPTMYMARSSPSLSNKWASLRRTHVGP CPSKEGSGPSQDLSDVKIQPARLAHDTVAEL PQRTVPCGTDGHGVIDSDETETALAEVSRSS PFGEREPLDLGATERITRKRLRNAVRDVVPP RKRRVPSTPSRKEQDLVPEVDGPAPTDLTH PPTSEPEPMLDPLSLVQLVRPQLGRAMGW AAEEMELGNVVMDELKREFNREVRRVGR TPPDQMYKRGAGPPLPQKREPERVALLEQLI AARVERGINRGLDWFLELNVAVFAAARVLS RRERVETLADRLHINDSATLSEVSRRRAKAE RKLRCAREQPWMSRRIRXLGVRVERLKQLA DLVRQRIAGRGNRSSYEGPRRRFRLRPSLRS VTEAPVNPPLNGNEVYTFWHSLSWAQSLRA NTDDCQLREFKNQLSARHTDLTSVGTSSL VQMFSALRKMKKGKAPGPDGIRAAWWG VFRRIAPYVATWVVRVIRGAEPVANWICNG

				TCGGAGTGAGCTC GGGGGCATCCGTG TGTTACCCCGCTG CACCACGCCGAAG CTGTCATAGCGAG CCCGAAGGGGAAT GGCCATGGAGACT CCAGCCTCACCT GTAACTCGAACCT AAGTCCAGGCCCC TTTCTGGTGTGG CTCGCACTGGTTA GGAACACGACAGT CTCGTGTAATCCC ACACGCACCGAAA GCCCTAGCCCTCG TAGGCGAAGGCTG CGTTGCTGGTTTC AGAACTGTGAGC WAGTGGGGTTCG GATGGCCGAGTGT ACCAACCTGCTTG CTAACAGGTAGCA		LTVLLPKSSDNADPSNYRPITCLNTCYKLFT AVIAQITASVVDVLGGLPRQQVALRKGVWG TSVSLMIDALTVADARRAKRPLGVCWFDK KAFDSVPHNLIRWILRVIGLPPVILSVIVSVM DQWATRLKIGGKVMPKTIPVRTGVFQGDTL SPLLFCLSVWPISFALDQFPQYQFRCANHLQ QGFSVGHVIFYMDDLKCYCPDREVLTAVIQQ VQKSASALGLTIHYKKSAWLDQDGGKSGK AVLGVPXLVGTYKYLGMHERFMIVSKDSLE SVRGKFMGRLKTLWTSKLTFGQAMLGTKS XCMPVVRYVLQNLFLPKSEFNQTRLVLREW DRQIRDLLDECNIRQVFRSKTELYVSREEGG WGLPSMEDALEEEVVTKLAILVARQETEPLF RVCEALERKRCPTPLSLGLQILKDWGVGVE LQGRLLLLNGNTVGPSQATRKL TGELVLRR EAERLSRWRSKVKPGCGMTGGAWRDVPGI DVHLSNRWL VKGALSPTVVSNSLAIRANTVI LRGSGGGYTKGTLLRCRGCNTGETRRHIV SACSLGRQKGAASRRHDNVCRILVRAICHK LNIEPPNSANFPHVVVLEGS GAKMWIDFPFV VPHKIRHTRPDIVVLFEWNGVRRLSVIEVAV SDVANMQTQHIRKSHRYGTNSTEPFVAGVT PTYRNDCLAAQLRAKFKAQQVDVIPHVGTT
--	--	--	--	---	--	--

				TAGAGTAATATGC TAGTAAGCAGGGC ACGAGAAAGGGC CGTAAAGGCTCCG GCTATGAAGGACC TTGCGACCACGCG TGTGTCTCCCACG TGCGGATTCTTGA AGCCAGAGTCTTG CACTGCGCGCAGG ATGGAGCCTGTGC AACTCCTCCCTCG CTGATCGCAGGAG TGGAGGATCACCA CTCTTTTACCTTG CTAGCTTGGGGTA CCACCTTGAGCTG GGGCCGGCCTTGC TAGCTTGGGGTAC GACCCTTTGAGCT GGGGCTGGCCTTA CTAGCTTGGCGCG CCACCCTTGGAGC	GETLDGEFGRIRKGLPMLTKLQMPRLWSEIQ RAVILGSYRILVEHLALPKGGA (SEQ ID NO: 1498)
--	--	--	--	--	--

				TGGGGTGGCGCAA GATCACTTGTATA CGGTCTAACCAAT ACATTTGAAAAGC GATCAAAGCGAA (SEQ ID NO: 1253)		
R4	R4-1_PH	.	Parhyale hawaiensis	CCGACCGCCAGCG GGATAACTGGCAA ACCCTGTCTCGAC CACCGGCCCGTGA ATCCATCGGGGCG TATGAGTCTGACA CAGGGGGGTGTTT AAGGTGACCCTGT TGCGAGGAAATGC GCAGCAAAAGCC GGATGAGCCTTAG AACATCGAAGGCC AACGACAACCTGC CGAAAGACTGGCG ACTAATCCAGTCA ACCTCCTGTAGGT CACCGGCTGGTCA	TGAGCCTTAGGTCGCG GGATGTGACCCGGCGC CAGAGTGTAGAGCTGA ACATCGCTCAACCGAT CCAATTTGGGTCGTGA AATCCCCTCGATAATA ATAATAATAA (SEQ ID NO: 1377)	MKMSHNRDTPSNGVKGTSVRLGTSLVRSPV GEAGAVRERGTHPSESVSQSDASVNATGE GSVREQAPLSPPGAEEATVPTQRRTRHKWS REDRVVLWECFVASKREGPGYLRKQLW DERGIPGNFPQASLSGQIRQICSKNLLSEEER LQIAARMEAQVASPSADEPARQVPTRPVTPP RSPPVEPARRPSIPSEETPDLGAVPSEIDSADP NRSPSRGPRHLPAHNMSQSESEDDVTDPDV SDQQRSDSLEPRDLLRNSSVESTPGHPNQEL SDTLLSNYPSEIDSDDPNQSPRRGPRHLPA HDMSLSDSMDEETEPDLSDQQRSDLLELRD LLRNSSVETTPKGHPSLRHLPEPKIRAAAFR VNSVIGKIHTNNITETNALIKAGADLAVRILE VQPRPQRTQRKKDPPWKHRLEKNIAEIRKH LSWISEWRRGNLHDEEEKTLLESRYRCLEV GLTNLEDTLKQRLSAKRKSVRRFEARVAGF HQNQLFNTNQKRLYQTLRGEETSSDSPNAE

			TGCTGAATCTCAG CTTAACCAGGCGT ACGAACTTACAGT TGGAGGGTCGAGC ACCCCTGATGGCT GAAAAGGACCATC AAAGTCGAAGGTA GCCAACGAGAAG ACCAACGGCTGAT TCAGGCGGAAGA GTCAACTCGTTGA ATGCGTTCGACAG CTTGGGGTAGATG GAACTCCTAAGCC CTGAAAGGCAGTC CATCTTCGCAGAC GCTAAGGTGCCCC GCCGCCTGAGGGT TATCAGGCCCCGC CGCCTGAGGGTTA CCAGGCAACC (SEQ ID NO: 1254)		ESIRFWSDIWSKEVRHNNTAEWLHDVKEKN VAADPDLTITSQQLKKQLSKTKNWKAPGPD MVQGYWIKTFTSLHSRIAAQLNHCLQRGTV PTWMTTGKTVLIQKDKAKGTEVSNYRPITC LPLMWKVLTGHIYERVYQHLDISKLLPDEQ KGCRRNTRGTDHLLVDKLLTKDARSKKK NLSMAWVDYKKAFTDMVPHSWILECLDIYGI AGNIRNLIATTMPNWKTQLTSANKHLGEVSI KRGIFQGDSLSPLLFVLTMIPLSETLNKAGQ GYNYSRTMKLNHLLYMDDLKLYAKSKDQV EQLLNIVHQYSQDIKMQFGVSKCGVLNIERG EVTASEGITIEEGTIKDIEEAGYKYLGVMEY NTILHRTMKDSIRKEYLTRLRLILKSHLNGG NTIKAINTWAVPVVRYASAGIINWTKKDCTD MDIKTRKLMTIYRALHPRSCVDRLYINRREG GRGLISVEDCVEAEKRALSQHFRESDDPWA RCLVEAKLLKETETADQFKERRRLDRTNKK KSMKMSGQYLEAVQDKIVPDSWNWLLRGE LKRETEGTILAAQEQALRTRYIQNKIDKRN PSTCRICRSSDETINHVISECGVLAQKEYKRR HDKVARHLHWTLRLRIHNFPVSERWYEHEPA PVVENEAVQIYWDKRMETDRVLHANRPDIV VKDKQEKSAKLIDISIPFDSRIVDKAEKKEK
--	--	--	---	--	--

						YRDLAIELQRLWQMKVDVVPVIGALGAM SKNLKTALRELKCGHLHPGTLQKSALLGTA HIIRKVL (SEQ ID NO: 1499)
R4	R4-2_BX	CAD V010 09048	Bursaphe lenchus xylophil us	GAGGATCCTGGGT TCCTACTACCCTG CTCCATCTCCTCG CGATGGATCCTTG GGGAAGTCTCCGG ACTGAGCTAAGAG AGCGTTAAAGTAG AGGGTGGCGGCGT AGTGACTTCCAAG TTGCGATGGGGCG GAACATCTACTCT TCTGAGAGAGGGA AAGCCCTATGGCG GCGGTAGAAAGGT TGGGCTACGGCAA CACTTGCCATGAT CAGATTCGATCAA AATTAGCCTCTGG GGCTGGCAACCCT ACAACGGATTGTA	TAACAAGTGTAATAAA AACCACCCATGCGTAG TAAACCGATCAATTATC TAGCAAAATCGCAGGT CAGAAGACCAAAGAAC CGACCCAGAGGAATAG GACCAGAGCTGAACT CTCAGATACGCCAACG GTCCTAATAAAACGTC GTTAAGTAAAGCATCG TTAAGTACAAAACAAA AGCACTGTAAACGCGA GGCCCCCTCTTTGCCAA ACTCCGGTAATCCTCGT AGTACGGTGCTTTTTCC CGCTTCACGAATCAGA ACGCTGCCAGATCTTGT CCGACATGGGCCGTAG GGTTTGGGAGTACCAG CGTGGGGCGGAGAGCG	MVFNNCKPKHLCPAIRPTGQQETNGGSEGT AEPTAGPSRPAVSEDAAQPVPPLFEEGEYIRA HRDKTCPYCEVLWIGARSSKARSDSWPLCQ ILYLMKRNDLDRGQRTRYPLLESSLRAAGIA RTKFAIIKCIRNVLRDRYVPNGPYSEHWKIY RANSGEVPQGATITKGKRSARVAGLPSPSQS GHHTKRIQAGTG IETETT VTETNTTPEVSHE HRDPCGEPETSAANVDKVTELTEDGSETRG TANVANGGVSVSDPGRKRQSSSQNRGNIET TNPELVGMWEDMFGVQLDGAMRTTERPRL PKLKHLSEPERLWIRAKLEQAWLQCVSYDV EQQWLNANAVLYAAIRSVAA SRPCKEAREA QKTWLDNKKKDEAKLRRLLIGRISSVHSMMPK GDRTPREKKLVKNITKLKNTHYPDMDWGG LLNHFKVKLSQLKEKISVRVAEHKRKVNRN AAGQYGKSVAGSAGLAPDVVSATAYWSGL AQPGPKKFKASSPIFQTWKDDVAKNLNTEP VLLYPIIKECIRKPSPWKAPGPDGIYNYWQ QEFVAQWIQTLVKRTL DIGRFPTALMCGRT VLLFKSGDKSMPQNYRPITCLNGCFKITNAV

				AACTGAACTATGC TATGCAAAATGAA AATAAAAAATGG GGGCTTTACAATC TAAGGTGTTGGCA GATCACGAAAAC GCCCCGTCGATGAG GGTGAGATAACAT CCTACGGAACAGC CCCTGCTGGACCA AACCAAATCATCC ACAAATTGGAGGG TTTTCTTGGTAGTT TCCCTTGGCACGT CTTGTTTCATAAG CCAGAATAAAAAC GATACCATACAGA CATGAGGCTGGGT ACCTGCCAGCCGC GACACGGAAAAC CGGTAGGTGCAAC CGGAGACAGCTAG GGGAAAGAAAAT	TACCTGGGTACACTGC ATAATCGGGTCTCAGA AAACCACCTATGGTTA TTATTCTGTCTCCCATC TGCAGGGTAGCTTTTCG TTGGGCCATAGGAGCC TAGGGGCAAGAGTGCA TGTAGTCTTCAACGGCC ATGCCAGGGAACTTG TGAGAGGTGAGGGATA ACTAGCATCAGATAAT ATCAGTCATGAAATTA GTAACAACCAACGTTT ACCGTCGTTGGCAAAA CACCGACTAACGATGC TAGTTAGAAAGAGTCG GGTCTTCCCAAAGTTAG GTGCTTGCACCGAAGC CGATCCGCTCTACCCAC AGCTCTGCCAGCGTT (SEQ ID NO: 1378)	LTKVILQRVQDTCALPREQMALKPKVWSC MEAQLRDQALQSEIGDDCKTAWIDFSKAYD SLDHDALRFVIQTIALPAGMEEYLLKSLDSW RTQLVLSDAGKVVSQKPYPIKRGVLQGDSL SPALFVLTTSPIVAHLQRTCPTGRIQLYMDDI KLYGKTESDLCMLIKETQRVANKLGLNINL KKCALFGKSIKQSIAGFDPLGDRTYKYLGP QRDVADIKQAYDELKAKTVQTIGETMACD YLTTTRQVINRLNSKIPPVVRFTQSALCSAP MTRGLYNKITELDNVSRAELRKVLIYKATN VSRFYLATKEGGFGYASLQQVFVEAVVSRA IYCLRAPSLCDIREFILSKFDPVKVARIALARS KIDMDIERMDMASATRTRQHYQAKWKTFL QQSKLYQKWVQHKIDIPNSSRWLQRGEISPR NCRIAVAVQDNTLLCRGFVGSKDPNKQCRL CNAGIETASHIVTECSTHRVHMYIERHDSVA RNIYAVLAKNCGFWIPHYSQKIPTVKITKSY ELYWNYKFPCTQALEACRPDIVLIDRAKKRI LVVEVAVSYVTRLEQMTQRKLYKYGVNGE YQADGETRGWNICREL VQKYNMRIDLCIVV IGACGEILPCMVKEIEKISKVSGRQLLERCQR SAVLGTVRTVRRHLAN (SEQ ID NO: 1500)
--	--	--	--	---	--	--

				AGTAAAGTGTCGA AAACAAAACAGG TAGCCCCTGGCTA GAGGGAATGGGA CATTGTCCGATTA GGTTGCCTTGACC CAATGAAAGCCAA CCGGGTTTACATT ATCGGTCTACCAA GGGCAATGACCAG AGAAACTGCGACT ATGCCGTAACCCT TCGCAGATTGCCG ATGAGAACCATCG ATCGTAAGTCGAA GCCAAGCGGATTG ACCAGTGGAGGGT TATCCCGACAACA GC (SEQ ID NO: 1255)		
R4	R4- 2_HG	ABL A010 00389	Heterode ra glycines	TGTGGCGATACTC GGAACCTCCGGGG AGCCTGGTAGGAG	TAGTCGTAGCCCAGAT GTCGACAGCCCCTACA CCGGTAGATGGAAGGA	MRKSFLQHIPELSSHIA MSVPARNYPKMCSL AQGSSGTL SHNGKGVAMHRCPSDDCAGKD PPQRGSQKGNLRSVRWTPSEEKAVFEYWSR

				TTGGCCTACAGGT	GGTATGTATCTAAGCCC	LEQHAMLNGSEARGTCAITRSQFLIHW DGE
				CGCGAAAGTCCCT	ACGGCAAGCCACCAGT	RESRSLSDGVPEYPMRTERAYYERVLLRQ
				AGGTGCTGCACGG	GGAAACGGTGACCACC	RGWQWDCANECLVIGQCAEPCRKPNVVAI
				GTTGCGCTAATCC	TCTGTTCGCGGAAATGC	KADKGMKRSLVKGKLLSLPHVMGEINQVS
				GAGCCCCTTCCGG	CCCGTAGAGTATATGA	VQVEVPLPSVPASVPQVEGVESKGFTEPES
				TGTTGGCTAGCAC	AGGTCGGAGCATTGAA	NKPSLEG NPAEEGLRKPERVNV PVHGIISDS
				TGGAGGGGGGGG	GCACAAGTGAGAACCC	ERKDLKDRFWSAYKTAKRSVGFRPALKIEP
				GGTCCAAAGAATA	TGGAAGTATGGTGGTG	NRVNRAQWEVLDSCVVEVLKKRETSNGYR
				CGTGAAGCGGACC	AGCTGGTAACCTCGAA	GCVLRHLNVAVYAAGYVLAEGNKERRQVI
				CTGTGTTTTCTCTG	TTCTCCTATGGGTGCT	RRQSAEWLLRQKSEINNIRRHIGWITDELTR
				ACCTCATACGGGT	TGCGCCCGTAGGTCATT	RRTGKNPTSRQLSNFAWLQRRYQVLGKPVR
				ATATCCCGATGAT	TGTGTATGTAGGGATG	ETRDLEVQRERLVSRLRLAQDRINSSMDREE
				ATCTATACAGTGT	GAAGGAATCTCGATAG	RVRKRMLPLRRKLEEPLGDSKLDTKQARTF
				TCATTTTCATTTCT	CGCAAATCGGGATACC	WASLIGERKEFGKIPLENWAE EVRSKVTD
				TAACCTGCTTTTTC	GCACTGGCGATCATGC	GQGFASDHVDQTVWKKILGKARPLKAPGP
				CTCAAGAGAAATA	TGCCAGTGGGCCCTGCT	DGIPNLLWKRLPSANQALFKWLMGIKRKQL
				TGACCACATTCGT	GAGTTGGGCGGGTTAG	SVPSWLTKGRVLLPKGGDPVDPANYRPIA
				TTGCCATTTTCGAA	CAACCCGTCCAGGGAC	CLNTQYKLV TGMVTAWVSEHLTTYSILPIEQ
				GGTGCGACTTTTG	GCAGATGCCAAAAGCC	RAMVSGTWGCTHAMVIDRAITSYAEATGLP
				AGGGAGTTCTCAG	TGGTGACACTCTAAAG	LYVGFVDFAKAFDSVSQPWIRYALKVAGVH
				GGACATCAAGCTG	GGAGTCTACGGCGACG	KRIRCLIGILMKCWSVRYEVFKSGRVLRSAP
				TTCATGGACGGCT	AGAACATCCCCTAAGT	LAVKNGVLQGD TLSPLL FCLSVAVVSSAVG
				TGACGGCTCGTGG	C (SEQ ID NO: 1379)	SLFDFEVTIPGRGVMQQQNHLFYMDDFKGF

				CGGCACGCAGCCG CGCGGGCCAGTTA GGGCTGGTGCCAA TGAGGCAGCGAA GAAGTCTCTGAAA CGCCAAAAGCAGC GGGAGAAGAGAG AGAAAGCGGGGA CGGGTTAGAGATA ATAGCATGATTGA CGGCGAAGCAATA CTTTGACTTCCAG GTACGGAGGACG GCGGCAAAAAGC AAAAAGGGGACG AAACCGATGGCAT TGCCAAAGGGGCC GAAACCATGAGC TGGACGCTGATGG GCCGTCTGCCGCG CAAAAGGAGGGG GGAGCCCCTCCGT CTGGGAAGAAGGT		APSEASLTRMLVTLERTASALGLKINKRKCA LVHPRERENEETGSDIPVLGLRDTYKYLIE ERFGIVFEDAWDRVRTKMFERMRTLCTEH TFGELRAAFASIAPVARYLFLNVIVGGPSW SETLTKAKDMDLRIRLLWERRDNEPGWRF KHCSADRLYLRVQYGGGLGFVSVEDTLSEII YCWAYVQCRPELELARELFGTLNRSARSGI KQSIAGARKVFRSYALLSKNSAQRVSDLD GDASPGFRVGEMIFMEPTRGARAIVKILRKE NDSRRLAAWKGRPMGGRVVSLPELDQVHS YHWLIRARIGRRSFRDCIAAQEGQLKAREL MCPHINAKAKWCRRCGDGRVETEQHILSGC AWSRTGTMLDRHNGVVRQVHTALCRKYGL PVSSHVVPLHAVIENEHAKILYDVALHTSPA GVLPRDGTSTYTGLRSTRPDMVIFDKKART ILIVEISVPWRENLVKQELIKWRKYAINSMIE PLELAEAEIPGPNLKHALGLAYGTSFPTVKV VPIVVGSCGEVLPNITKRLSELGIPKRGIPSL ESIQRAAIIGSGHVIRAHLSVPRSESET (SEQ ID NO: 1501)
--	--	--	--	---	--	---

				GCGGCGCCATTCG TCCAACAAACCGG AATAGATCTTGAT GATCTCAGCAATT GATCATATTGATC TCTTTTCCAGTTG TTGCGATAAATTA TCGTGCATTATTT CTTCGATTTCTCC AAAGCTTAACTCC TTTCCTAATACCT ACTCATGTATACG TTAGTAGGCATGT TTTATGCAGGTAA AGATAGACCTGGT GCCCCCGGGCGAC TTGGGATGTGGAT GATGGTTGGCGGA GAGTTCTGATGAC GCCGTAGTTCCGG AGGAACAGTTCCT ACTAGTGCCAGCC TGGGCTAGTGGAG		
--	--	--	--	---	--	--

				CTGGTCTGGCGAG TGCTTTGCTCGTC AGATAAGGGGTG GTTGGGGTGGTAC CTCTGGAATGATG TGTCGGATAAGCA CTCTGCCTTATAA AGCTGTCGTTCTG CCCAGGTTTCTCC TAACCAGGTTGAA CTTGTAATGAGCC TTTGGGTATCTGG TCGGGGTCTGCGC GGAGATAGCTCGT GAGGAGCTGTTAT TACATTATTGGTT GAACGCTTTTGGG TCGTACTGCACAG ATAAAT (SEQ ID NO: 1256)		
R4	R4- 2_SRa	CAC X010 02006	Strongyl oides ratti	CGTCAGAAGAGCA GGTGTTTTTCAAA GCAAAGACTTATT	TAATTAGATTTATATGT GCCACTTTACCTCAAAG ATATATTAATACTAACT	MQKFSVPKDSSQIFLVDSILNKHICSTKNKV KDVIKRRSIIKTLICAAGLTLRKLVCCKLGN NKYNSKINQLWKKERKIINCIEDLKHLIETN

				CTACGAAGGGGA AAGATGATCAAAC ATGCAGATTTGGC TGCAATGAAATAG AATCAAACCTACCA TGTGGTGACATCT TGCAAATACCATT CATACACAAGAAG ACACGATATGGTT GTGTACAATATAT TAAAAAGACTAAA AGAAAAGTATAA GTTAGATGGTGAT TTACAATATGGTA GATCAGTATTTAA TGGGAAGAAGGA AAAAATTAGGATC AGATCTGGCCATA AGTTTTTGA CTGA ACAACCGTTAGTA ACAATAAACCGG ATATTGTTATAAC GATGAAAGAAGG	TAATTATAATTTATTAT GAATTATAATATAATA AGTTTTAAAATAAAA (SEQ ID NO: 1380)	KRRHNFGKRLRNAKVSPSEMLKDYYNKLK YIKNEITACLDEHKKAILRTKFKLTPSIKIISNI QNHNDDEEAELPKEEEFVKYYKELFTNKDGD DKETPHLDNWLKKFSKTLIVDWTINDKEILE ALKYCGNFKAPGSDMVMKVCYKWFKSAQ NYLIRWIKSTWYGEYTINKKDTNAVTFMIW KRDGKPKNDVKSYRPISCLNCDFKLLNKLIA NKIYESIEKILPINQMAVIKNKHGTCEALLY KSLVQSMKFRRTKDVKEIWCSWIDFSKCYD SISHKCLKKMIQSIKAPPIHKLILDGIDSWNI SICNGKNISKTKIPVKSGILQGKVASSLYFVL LTGEISYALNKEEQVPIETITPSNTLKINHISFI DDYQLYATSQKKVEKLTIKLREIAEEMNLK LNPQKCGIYGTDDL GKRLMLKESSLNFPYTS EYKYLGLVENSLLDKDINIQLFKDKILSKYS TIFESRLTTHQKRKVFNSTISPCAAYYLGNLI TNKCSIQELLNECKKFDQMVRNQLVNQNIK KLQVSNSRIYLPKEYNSLGLNEIEIEVAANIIR KACYIKKRETLRGVDKLYIAMSKNGHRNTL SDALYITKKYSN FQINWNIMGMVKDQNNIL LDAKKIENIKEKRRNLWLEHWKKGNMTYA NEAIKKEFHLPDLNIDSKYLMLCYAGSEEQII YNGHVSLVNQSSPSSRLCRKCNKLEETSYH
--	--	--	--	--	---	---

				GAAAAAAAAGAT AACGTACATACTA GAGATGTCGATAC CACATATCTAGAA CTTGAAGATGCAA GAAAGAATAAAG TATAAAAAATATT GTGTGAGCTCCAT GGTAAAAATAACG AATGATAATGTTG ATTCGATAGCTCG TGATTTTAATCTA CTAAATTTAATGG AAAGAAAAGAAA GGTGTAAGATAAA GTTTGGATCATTT GTATTTGGCTGTT ATGGAGAATATGT CTTGACTGAATCT GCTCTAAGAACAC AGAAAATCCTGAT TGAATTGGGATTC TCCAAAAAAGAG		VASVCEFHKKNLHLMRHNSAVYHIITELCRI MKVKCTLRYPEASGIIKSGNMKIAAGVKYT FGTAKIYHNKPDLVWYTPEVIYVIEVSISSLK NAKSQMKMKTARYAVNSTKKLENFAALNN LKKGENFVEILSHKANFKRVHFMPLVFCTFG EIPKETMKYLEKLNFSNEKIKTIASPIARYTG RTLKAHFTN (SEQ ID NO: 1502)
--	--	--	--	---	--	--

				ATGGATAGCTTGA TCAAAGAGTG TTC TTACAGTTAAATG AATGAGACAGCA AGAATTATTATGA AACATCTAGAAGG CGAAAAGAATCA AGATGAAATCAGC CATTAATAACTTT TATTTACCCAATT ATTTATTGCGTTA ACTTGATTATATG ATATTTATTGTAT ATTTATTGTTTAAT TTAGTTTCATAGT ACAAATAAAAAGA ATATACTACGATT CTTTACTTTTAAGT TCTACGAGAACGT TGTTTTAGAAATAT TTTAATAATATTT CTACTACAATTAA GTAATTAGAAGAA		
--	--	--	--	---	--	--

				ATCAACGAAAGCA GCTAAACTTACTC GCAAAATTCGTTG ATCGAGGCTGGAA TGGCACCAACTAA TATACTAACCAAC AATAGAAAAAAA AAGAAGAACTTCA GTGGACTTCAAAT AAAATAAAATTAT TGGTAGAATTGTA TGATAAAACTGAG AAAATTTGACTAA AACAAAAAGATTA GAGCAAATATGTC ATCATTTTCCTAA TCATACAATTAAA GCGATGATGACGA AATTGAGAGAATT TGAAAGAGAAAA AAGGGAAAATTGT GAAATTAAAATGA GGGAAAATGAGG		
--	--	--	--	--	--	--

				AAAAACCTGAGG AAAAATTAAATTT TGACAACCTATGAA GAAGCAAAATTGA AAAGAGGAATTA AGTGTAAAAAAG AAGTAAAACCAAT TGTTATTGAAAAC AAGGATTAGGAAT TTTTGAAAACGGA AAAATTATTTCTC AAGTTTGTCAACA CAATCAATTACCA AAGAAAGAAGGG TAAATCAACCAGA (SEQ ID NO: 1257)		
R4	R4-3_BX	CAD V010 08832	Bursaphe lenchus xylophil us	GGGTCCTCGGTTT TTACTACCGTGCT CCACCTCCTCGCG ATGGACCCTGGGG TAGGCCTCCGGGC TGAGCTAAGCAGA GCATTAAAGTAAA	TGAGACCACCCATGCG CAGAGTATCCGAATCA GTGAAAGTCCAAGTTT CAGGACAGAAACGTCA GATAAGTCCAAGAGAA ACGAGAAAACAAGTTC AAGTATGCAAGAGTTA	MISRSQADRPVEGQPVTAMSFHNLEPNNLY PENLRPTGSQDANRGVADIAEEVTGPSGLVT NEEAARAPPLFVEGEYKRAHCGGGKCHYC RVLWIGARSSKARTDSWNLCEILFLINKCME LGNVRRRIYSPLESSLKEAGINRTRHAIVKRL AVMRDRFVDNAPYSEHWRLYNACAENRAV VVPMDSATTVKKRTARQAGLESPSQIGVAG

				GAGTGACGGCGCA	ATCAATAAGAGAGTAC	KRVHEAETGTDRINAVIETNTTPLEDIDLSPE
				GTTGCTTCCAAGT	CGTAAATGTATGACCC	TPEGLAELPSTVEIMELTEDGSRSRGTANDA
				TGCGGTGGGGCGG	CCCCCTTTGCCAAGTCG	DGGVSISDPLRNRPSSSQESRNVPEQVDPDG
				AACATCTACTCTT	ACAACGTGCATGCAGG	ELVWESLYGAQLRGAMRTTDRPRLPKLTKF
				CTGAGAGAGGGG	TGTCTCTCTTTTCACCC	SAAEQLWIKSKVEKARLECVSYGIEQQWLR
				AAGCCCTATGGCG	GCCATATGGACCAAAC	ASAVLYATIKTVAACRPYNKAREAHKVWL
				GCGTTAGAAAGGT	GCTATCCAGCCTCGCTC	ENKRAEEKRVRRRIIGRIETVRTMPKGKRTDK
				TGGACTGCGGCAA	AGAAGAGCCTTAGGGC	QIRLARKINRLKRVSFPEMDWHGFLNHFA
				CACTAGCCATGAT	TGGGGAGTACCACATG	KLDLLKKLISVRVAEHERKISRKIAGTYGKS
				CAGATTCGATCAA	TGGCGGAAACTGAATC	VSGQSGFTPDVVAATTFWSGLAQPGPKKFK
				AATAGCCTCTGGG	TGGATGCGATGCATAC	KNSLIFQTWKDSVVENMNTEPVLLHPLIEC
				GCTGGCGACCCTA	CGGGAGCGCAGCGAAA	MNKPSPFKATGPDGIFNSYWRQGFIANWVK
				TAACGGATTGTAA	TCACTTAACGCTGGTGC	SLIQRITQTFEFPASLMCGRTVLLYKNGDTA
				ACTGAACTATGCT	ACCTCCTGCTATCGTAG	KPENYRPITCLNGCFKMTNAVITKVIVQRVQ
				AACCTGTCAGTAA	TACTTCCTAGATAGATG	DTCALPGEQMAKPKVWACMEAQLRDQAL
				AGACAGAACGGG	AGTAGGGTGGGCTAAA	QSEIGNDCKTAWIDFSKAYDSLHDHAIRFVI
				GGCTTTGCAATCT	GGTAGTCGTCTTCAATG	ETLALPDGMEKYLLKSLESWKTKLVLSNRG
				AAGGTGTTGGCAG	GCAATACCGAGGGGTC	KVATGRPYKIKRGVLQGDSLSPALFVIATSPI
				ACCACTAAAACCTG	TCGTGACAGGTGAGGG	VSHLKRVCPSGRIQLYMDDIKLYGKSETELR
				CCCTTTGATGAGG	ATAACTAGTATCAGAT	MLIKEVQKVANKLGLQMNLKKCSTYGAGL
				GTGAGATAACATC	AATATCAGTCATGAAA	TESIAGFDPLGDRAKYKLGVPQRSVADTNL
				CTACGGAACAGCC	TTAGCAACAACCAACG	AFGELEGKVIRSIEETMACEYLTMRQVVTRL
				CCTGCTGATCCAA	TCACCGTCGTTGGCAA	NSVIGPLVRFVAQSVLTSQAKVSWIYNKISD

			ACAAAATCATCCA CAAATCGGAGGGT TTTCTTGGTAGTTT CCCTTGGCACGTC TTGTTTCAAAAGC CAGAATATAAAAC AATACCGAAGTAC AGAATGGCTGGAA CCTAACAGCCATG GCACAGAAAACC GGTGGGTGCACAC CGGAGACAGCAG GGGAAAGATTGCA AAGTGCCGTAAAT AAAGATGGTAGCT CTCTGACCTGAGG GAATGGGACACTG TCCGATTAATGCC TTATCCCGAAGAC GGTCAAGGTTCTA TCTTATCGGTCTC CAAAGGGCTCTGG CCATAGAAGCTGC	AACACCGACTAACGAT GCTAGTTAGAAAGAGT CGGGTCTTCCCAAAGTT AGGTGCTTGCACCGAA GCCGATCCGCTCTACCC ACAGCTCTGCCCAGCG TT (SEQ ID NO: 1381)	LDSKIRAKLAQTGLRYKKSNNVARLYLSKSK NGIGLVNVQQVLVEALVSRAIYCLRAPSLVE IREHILTAEFDPVGAARTVLRRSRIQLEIERV EMASAISAIKTNYQARWMTKFTQSKLYQK WVHHIDIDLANSNLWLERGEISPQNARIAVA AQDNTLLCRGFVGNRESEKQCRMCMNGIET CSHILTECSYHRAHMYIERHDSVARNIYAVL AKDHGLWIPHYSQPVSSTTKTPTCELYWNY KFPCTRALEACRPDIVLIDRAKRTILIVEVAV SYVTRLKQMVSRKVYKYGVNGEKGADGES RGWNMIRELSEVYNMKVNLCVVIGASGE VLPCTVKAIQSISSTKSSRQLLERCQRSAVLG STRVVKRHLAEFH (SEQ ID NO: 1503)
--	--	--	--	--	---

				GAGGAAGCCGTA ACCCTACGCAGAT TGCCGCCGTGC (SEQ ID NO: 1258)		
R4	R4-4_BX	.	Bursaphelenchus xylophilus	TGCCAGCGGTGTT GATTAGGTCCAAG TTCTTTGGCCAAA GATCCGCCCTCGG TTTAGCAGTACCG AACGAGTATACCT TCAAGTGGTGGCA CTGAATTAGACTG AATACTCTGAACT GTAGACTTTTGTG CAACTGTGTATGG TGTGGAAGACTTG TTTGTACCACTAT CAGCTTTATTGGG GCTCGTTACTGTT TCATACAGGTAGA TGTCCCCTTTAGA GATTTCCCTGCAG TTTGCGCTCCAAG	TAATTAGGCAGTGCTCC TAGCGGAGTGCCGTGA AGTGGTGTCAGTACCC GTCGTGTGGAAAGCCC AGGAGGGTTAGTACCG ACAGTGGGAAACCCGC TGCACGCAACCTAAAG ACATTTGCCCTTCGGGG GAGAGGTATGAGACCA CACTAGTCATGGTTGCT TGCGCAAGCATGACTC ATTCATTGTAAGTTCGT ATTATGAGGCCTCGCG GAATGCGAAAGTCCAA GAGCGCCAACTTGTCTC TTTGGGAGGTTTCCTCCG GGTTCCTTAGGGTTGGC GTACCATGCTTCGTGTG AGACCTAGGCCGCTTG	MEILWEDLRLKIEDRYGVTL PQRSASSLKNQ YPKVILRGLPDSGLPWAGVQVNDTGQVVV VDHAEAATLRGSSPAAVDGEAEPPVPPLP AAEVVESAADA AVDPDQSEIVADQGVETRP VENPPANSRETETEPVEVEPYLEGQYKFFVS KILGKSMWRKPIKYPRRVPETLWRQANELIE RSIRQGEVSIQSLNCMVYAAGCAVKSSLDK KDQEAKRRESEWYACRKA EIKALERYLNFI DLELKRRSASRPLTSRQRQNLGVLITKYGRA RVRSGVRLSELQAMLRDALVGIRKCMARKS ADKKRKQGKFVPIQRYLEPSSAEPRLSPDTV RAYWNDIVGSSQQSTSDSTIQDWSSNLSVPS QELNASKIMGWWR AAVSKSKPNKAAGPDG IPGVLWKRFRSASEWVCTWLYRLLQKRRIT PRWLSVGRVVLLPKKGPLEDPANYRPIACL NTVYKLITSVVEMAVREQIQACPLVPYEQI ANRKG VWGCTHASIVDRMITGASREGKGG GFPDLRVLFYDCKKAFDSVNRDHMF AVL R VANVNVKVVHLLHTLSQQWCVR YELRRNN

				TCGCTAGCCTCTT GCGTGTAGTCAAA GGAATACATTTCGC CGTCGGTGACAGG GCTATACCCGGCG ACTACGGACTTGT TTATTACGTAGTG CAGCCTCGTTTAA GACGAATGTGAAA AGAAGGTGTGATT ACTAAGCGTTATG AGTCGGGTATCT GGAAACTCCGCCC CGCCGCAATGGCT TCAAGGGTCAAAC TGCTAACATTTTA AACCAATATTTGG CAGCAGCAGACTA TGATGTCAGAGTG CCGGGGGTCTATC GATATAAGACTGA AAGCGATAGACGT GGAGTGAAAGGA	GACCTAGAGTTAAGCT GTTGCATGTTTGAGGAT GCCTTAGGCGGACGCC ATG (SEQ ID NO: 1382)	RVERSSPLRVKRGLLQGD TLSPTWFCLCMA PISASIKTLNPGPTLRPNMGNGRNRGQVAIQ VSHVFYMD DLKVYCPRVADQRRMEQNIPQ LFGEIGLSINASKSAAAAAVGRYVESELPVL GTKDEYKYL GIESGFVVNEVAALDRMQAV LLNRVEAILS VKEHTVGQRRDAIRAKAIPGG AYILGHIILSDLDPRGAAERMRRLDIEIRRLV KSAGILHDKCSTARIHLSCEQGG LAWPSME RAYYVAVAYSASYLLTSQDETISRARDYFV SGRLSNKFTVYKHLTSIVDSLGLSVELPDPN GLPTGQPSVLARTIARAIDAKLEAQWKETLL TYQRAGRVERADPTVVDHANSYHWLRKA WINEKAYQH AVSVMEGTLLEGVNPHGVLT MCRACKAPSASIAHIITGCAELRKSHMKVRH DGVTRWLYNALTEVDGSLPKFH YTQQIPAE MRGERLTVRYDS DIVTPNKPRHNRPD LVVF DSTRKVIYIVEVSVTWLSVLQKQYDNKLN R YAVNSNHEFSESIPYPPGVNLANEIRVLYPQF TGGVKVFPMIISPTGEVHMQFVPHLAELLEN PNIPRILEKIQRSVVLGTDYIIRSYFAM (SEQ ID NO: 1504)
--	--	--	--	--	--	---

				TTCCCTTTGTTAA GAGAATCGGTAGA ATTCACTTTTTACT TATTCAAGAACTT AACAGCAACAAG CACTCGCGAGGAT TACCGCCCCAGAT TCGGTCGGCGGTA CTTCACCTGCTTTC TTCCACTTTCGGA ATCTGGCATCCTG GCTTTCAGTGTGG TGATGGCCGGCTT GAGTTTTCTTGAG TCGTGCGAGTGCC TCATCTGGGACGT CCGGACCGATTGA TGGAGTCTGCAGT GGACGAGGACTTG ATGGACCGTAACC ATAGTATATCCCT CATGCGTCTTCTC GACTCGAGGGGGT		
--	--	--	--	--	--	--

				AGCTTGCACTACC CACCCCTTCTCTTCT CCGATTGGGATTT AGACCTAGCCCTC TGGTGTGTCTCGA CCCGCGATATCAG ATTCCTGAATCGA CTGTGAGAAATGT CTACGCGCAAAGA TCGACCCATTGCC ACCCGGCTATGTG GATCGGGCTCTTG ACTGCTTATCTCC GGCTTTAATCGCT TGAGGAAAGGGG GGTGTCGCCCGAA AGGGTTACGCGAT CATCCGATTCCTA CCGTAAAAACGTA GTTGAAGTAGGAT TGAACCTGAGTAC CAAGTGAAAAGA GTGGCTATAATGC		
--	--	--	--	---	--	--

				TCCATGGTATACC CTAGGGATAACCG TGTCGAGCGCACT GCTCAATACCCTG ATTTGTTAGTGTA GGTATGTCTAGTG GCTGCCTAGGCAG ATTCGCCTATTCA CTACGTAAAATCT GGGTGAGTACCAT AAGAACCTCCTGT AGGGCCAGGAGTC AAACTAGTCCAGG TTCGGTATGCTTC GGTATACTCTCTC ACGGGCTAGTCAC CTAAGAGTTAAGA ACCGCCTTTTCT GCACTGTGAATAG AAAAAGAAGGGC GGGAGAATATACG GGCGAGGTAAGC ACGTCGGAACGGG		
--	--	--	--	--	--	--

				GTGTGCCCAACTC GTACCGTACGCCC GGACAAGGACCAT CTTTCGTACCCCG TTCACCCGGCACA AGTCCGATTGTCT CTCCCGAGAGGCG TCGGCGGAGGTTG GCTAACGCCGTCT CTCCAACCATCGG TTTGGGTTTAAGT AACGCCCCAGTGG CCGGTAGCCAAAC TGATGGAGGCAGA AATGCCGGTCCGT TGCTAAATGCGGG AACGGACCAAAA CGCCGGTGTGGTG GATACAGGTGGAA GAGTCTTTTGGTC TACGCAAGAAAA GACCAGGCTCAAG TACGAATACGACC		
--	--	--	--	---	--	--

				GTCTTCGTGAACG ACGCCGTCGAGAC AAAGTTCGAAATC CTACGGTCGATAA GGCC (SEQ ID NO: 1259)		
R4	R4_AL	U294 45	Ascaris lumbrico ides	GGGGCCGGTGGGT TTACTCACTTCTG ACCCACCACCAAC GGAACGAGGGAA AGCAGAGCTGGG GCCCTCTTCCGAT TGGCATGGAACCG ACCTCCACGTGGT GGCCCTGGGCAAC GGAATTCAAGAGA GGATTTAATCCTC TCTATCATTTGCA AGATGGATGAGAT CGAGGTATCCGGC AAACAGGTTCCAA GTGAGCACCTTTC CCATAGCTGGGAA	TAGTCGCTAAGGGGTC CGGAAATGGTCCGGTC CTGCGCTACCCGGTTCT GGTAGCACGTTCAAGC GCTCAATCGCCTGCCTT GTAGGCAGTCCATCTGT GGAAGTCGCGCTCTTG ATACAGATGTGGACGG ATGGAAGCAGATGATA GAGCCGGTGACGGCCC TACTAGCCAAACGC (SEQ ID NO: 1383)	MPCSTNSFFERGTPEPHREPISGTDSSSESLGM GTHRSPLNDDEVINGPKGHESDPVHVVRA PRTLHPRRLELPiGVNNLGEASQLRQDSAIA EEAQLESTENHDGRRPPLRGGRKLWSEKEIA TLRRLCEAYGNRQVCWKEVQRKFADFHEE RTVAALATKWGALKRPRAPMVGAPPTPDH DPERGPAGEGDGGTTSQENVPTDDPIPANGP TEGKESDVRLPAVACRCTEPEEQLMESDVRL PAVVRLADPEQHTMKSGVKPVALDGSADL EERPKEKDIEQMGVDFEGEPRFRAFRKAFY GYFRWAVNSFDREPVKRVRRDCPKVIFYAY ADYLIATGSSKALGPNQSRIGRLNGLVYAA ARTIHQFWREEVGHRQQGEKGWYTKTKAT REDLQMLISMMESELARRKEKRKPGAKELE NIHKLVARLGTRSTSGIVRRLEMTRQRLKLL EDRISLHEQEKRRLRLKQFAETPSLKLLTK GAKDRGDTMVTMKSVMDFWRPIIGRRVTS

				TATGGGTTAGGCG TCCTCTGACATAT AAGAGGAATCAG ACTCGTTCGCGCC CGGTCATTAACAT CGATCAGCGGGAG GGCCGGACTGAAG TAAATTTCTGTT GGCCCGAGTGCAG GTGGAGCTCGGAC CCGAAAAACGATC CCTAAGAGGACCA CAACCCGAAGGG ATGGACGCAGTCG CCCCGGCACGTTT GGTGTTGGTTATC CTGGAGTGTTGTG GGACGAATAGCT (SEQ ID NO: 1260)		NPDQLQVLRDWRDEQKKAYPADLDLEKAD LEEKYEGAIRRIQPWKAPGPDGLHAHWWK ALPSAKRLLGELVVDWLTTGKVTTGWMCR GRTILIPKKGDRGDPSNYRPITCLNTCYKVL SVMNSVILSHLSRGEALPMNQRAMRKREW GCTHAMVLDRAVMMDAMAQKKHSLSVA WLDYRKAYDSVSHEYIRWAINSVNIPRSVQ LTLKRLMSDWETRFESTQCRPKLRSDKMKV LNGIFQGDSLPTLFLCIAPISYALNKGVGQ CQSSSGWSAGYGFEIGHQFYMDDLKLYART PAMLDSDQVQVSEVSEAMGLHLNLSKCAKA HYAPHGAGGAQEAVEGAEGSRKGEIPILGL RSTYKYLGVQRLLPMEVALKEFEDKFMDR AETIFASELTWGQMATAYNIAIAGLRYVYS NTNGASPKLLEALKRAATLDTRIRDLLRRH KCRFRNSFVERLYIPRECGGYGLKSVEDTLR ESILATWSYIATNPHLAGQQYFFERLAARGK RTPMADGVKILLDLGVEPQVDLKRRTVTVD GIVFEDPTKLHRYLVGKLLKARTEARIRRW KEASLAGRLVNDTSIDMRLSCLWMKKGFVS ARNLRDALAVQEGSLLTRACPALKGKGGQE VCRCCHAAPETAEHITSACRYWLPSLYVER HDSVARNLYYVICCRYGITPVHYSNRVSPLS
--	--	--	--	--	--	---

						ENSQCRVLWNMDMQTRTPMKHRKPDIVVF DLKREKILMFEVSIAHASGLLKQREIKINRYT VNSEELPDETITPYPPGPNLAADLAATYGWQ VEFAPVVVGTCGEHVPVAVKEDLQRTLCLKP HQVEALLERISRSVIGTARVVRAHLACS (SEQ ID NO: 1505)
R4	R4_Hmel	.	Heliconi us melpome ne	ATAAATAATAATA ATAATAATAATAA TAAGCCCCCTAAA ATCCAACCATACG TCCGAGTCGAACA TCTGATTCTCGTG GGGGGCGGACAC GTGAAAATAA (SEQ ID NO: 1261)	TAACCTATTGTCAGAAT TCCTTACTAGTAATAAT AATTATCGCTGAAAAT CTCCACCCAAATATTGC TTGGCTATATGCTCGCA ATTTTGGTTAACGTAC CCCAATGATTTGGGAG AACAAAAATGGTAAAA CTATAATAATAATAATT ATATTAATAT (SEQ ID NO: 1384)	ITYTANMALVTLFMENMENKRYNLRPLPGG RRGASGANAGCHSMRTVGDGGLSRRVPLE KNVTAEQSSSPLTSSSSHSPVSSIPSPSSTRTL LNSPNSPTSSHSSLVIRSADVQVEALANY APTAGSIRARKKWDIMNRYIWRTYLIITKC ETLLNNYLEPLHQEFSSKFPEMQVTRQRIG DQRRAIIRNKLLSDDTLAQILIEVKELLQIGD QPLTQNNIHSTQLSHSNTRIKWSNELNEEIV KCYFEVTLLEVNKTSYRKNLYSLFISRNPHL SHLTEQRIADQRRLLIFMNKSVHNDRIIELKRE VEIKLANSNSLTKNITESNSPSSQTNEINDSA YVQSNLQPVEPLDQHCINRHNLIKHYVEQE FNNALIQFNNTNPETRPYIPRQKSSRKFSQIV SFLNSEVLPKHLNNELDFNALHNIIYTAXYT ASLCNGTKFSFIDNYRPRNSKPSWQRRLES IDKYRLQIGRLTQYISGNRNRKILKTVEEIKT QYKIHSHHEEPNTELPFLDTLQKLNATSN

					RLRRYLTCTKRKQQNNTFVNNEKHFYRTLS STNQNTTTQLXEHPTENNLQQYWANIWETS IEHNADAEWLNKIPDXEINXMKFKDISIETFN QIIQORTHNWKAPGTDNIHNYWYKKLTCTHS LLLKHINQFIQSPCTLPLFITNGITYMLPKGL DPTNPANYRPITCLQTIYKIITACITDIIYKHID QNNILAEQQKGCRKNSQGCKEQLTIDAIVM KQAHNKNXNTMYIDYRKAFDSVPHSWLLYI LKKYKIHPILITFLSSVMLSWKTRLKLINNNE TLITDWIKIQRGIFQGDALSPLWFCLALNPLS ELLNNTNTGFKLKHNNTYHIIISHLMYMDDI KLYASNKELKILADLTQSFSTDIRMEFGIEK CKVHSIKRGKSQQNTYILNTGEQIESMDENS TYKYLGFQQAQKQIQKQTKIELTNKFKFRL NQILRSQNLNSRNIKAINTYAIPILTYSFAINW SQTDLNLQRIINTHMTTHRKHHPKSCIQRL TISRLDGGRLIDIRNLHNNLVTKFRNYFYA KAEISELHKFIVNIDNKYTPLNLNDRNIQLNQ TLITKQQKIEAWSLSLHGRHLADLSQTHV DKVASNEWLRRGDLFPETEAFMMAIQDQVI DTRNYQKHIIKRPNMVNDLCRRCYSSPETIQ HITGACKTIVQTDYKHRHDQVAIIHQHLAF KHSLITQAQKTPYYKYSPQAILESTNFKLYW
--	--	--	--	--	--

						DRTIITDKTVHYNRPDILLHDKVKXSVYLIDI AIPNTHNLASTFSNKIDKYTDLTIELKSQWK VQSVTTVPIVLSTTGVVPHTLHTSLETGLIHR LSYILLQKAAILNTCRIVRKFLSSNN (SEQ ID NO: 1506)
R4	Rex6	.	Takifugu rubripes	TTCTATGCGCCTT ATGCGACTGGATA GGCCAGTGGTTTA CGCCGCTGACTTT GGTGCGGAAGGTT GTCGGTTCGAATC CAGGCGAGCCCTT AGGCAAGGCTCCT TACGCATATATGC CTACACCTCGGT (SEQ ID NO: 1262)	TAGAGGACCCGAGTCT GAAGGAAGGAGGCACC GCCAGGAGGGCGAGG AAGAGATTTTTTTTAT ATATATATATATATATA TA (SEQ ID NO: 1385)	MSGTXTDRVIPARTSPGSTRSASGVGEPGPP DVKLATGTRHSWSRAENVVLMACYYSNP SERGYMQRMWEKWVLRNPTSSLTKKQLLA QCSNIRNKKLLSQLEIDEARRCASPTVQICY GKGEPGRQVSXGVISSPPNIEIGYKAPMTD GLGTRAADLRERIMKSWGNSSTSLPRLTHK VPDQSLLEDMTALSTIPTTTITETNQLMYA AATVILQMLGYKMKSMNSQKEQMAPWRR RLEAKIMATRREVSLTTEL SRGVNLRTEXPK KYNKLSTTEALETAKQRLTALATRLKRYTR EVEARRINKVFSTNPAKVYSQWQGNKMTT DPPRAETEYQYWKSIWEKEATHNTXAQWLQ DLQTEHSQLEQDPVVITLADIQTRVSKMKS WTAPGPDKIHAYWLKKLTALHERLAAQMN QLLTSGNHPEWLTQGRTVLIMKDPQKG TIPS NYRPITCLSTTWKLLSGIIAAKISRHMDQYM SRAQKGIGNNTRGAKHQLLDRAIAQDCRT RHTNLCTAWIDYKKA YDSMPHTWILECLKL

						YNINRTLREFIQNSMKLWNTTLEANSKPIAR VSIRCGIYQGDALSPLLFCIGLNPLSQIITKSG YGYQFRSGTTVSHLLYMDDIKLYAKNERDI DSL IHL TRIYSKDIGMSFGLDKCGRMISRRG KVIATDGVELPEGNITDVQDSYKYL GIPQAN GNHEEAARRSATAKYLQRLRQVLKSQLNG KNKIQAINTYALPVIRYPAGIIPWPLEEIQAT DIKTRKLNGKHKIQAINTYALPVIRYPAGIIP WPLEEIQATDIKTRKLLTMHGGFHPKSSVLR LYTKRKEGGRGLVSVRTTVQEETSLREYIK KMAPTDRL LSECLRQQKPTKEEEPEGLSWK DKPLHGM YHRQIEEVADIEKTYQWLEKAGL KDSTEALLMAAQEQALSTRAIEARVYHTRQ DPRCRLCGDAPETVQHITAGCKMLAGKAY MERHNQVAGIVYRNICTEYGLEVP GSRWET PPKVLENKQAKILWDFQIQTDKMVVANQPD IVVVDKHQKTVVVIDVAIPSDSNIRKKEHEK LEKYQGLKEEMERMWGMKATVVPV VIGTL GAVTPKLSRWLQQIPGTTSEISVQKSAVLGT AKILRRTLRLPGLW (SEQ ID NO: 1507)
--	--	--	--	--	--	---

Таблица 30. Иллюстративные мономерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	Номер доступа	Организм	Последовательность	Сигнатуры RT
Q4VFZ2_9 GAMR - только остатки	Q4VFZ2	Эндогенный ретровирус свиной	MGATGQQQYPWTTTRRTVDLGVGRV THSFLVIPECAPLLGRDLLTKMGAQI SFEQGKPEVSANNKPITVLTQLDDE YRLYSPLVKPDQNIQFWLEQFPQAWA ETAGMGLAKQVPPQVIQLKASATPVS VRQYPLSKEAQEGIRPHVQRLIQQGIL VPVQSPWNTPLLVRKPGTNDYRPVQ DLREVNKRVQDIHPTVPNPYNLLCAL PPQRSWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPGTGRTGQLTWTRLPQG FKNSPTIFDEALHRDLANFRIQHPQVT LLQYVDDLLLAGATKQDCLEGTKAL LLELSDLGYRASAKKAQICRREVTYL GYSLRDGQRWLTEARKKTVVQIPAPT TAKQVREFLGTAGFCRLWIPGFATLA APLYPLTKEKGEFSWAPEHQKAFDAI KKALLSAPALALPDVTKPFTLYVDER KGVARGVLTQTLGPWRRPVAYLSKK LDPVASGWPVCLKAIAAVAILVKDAD KLTGQNITVIAPHALENIVRQPPDRW MTNARMTHYQSLLLTERVTFAPPAAL NPATLLPEETDEPVTHDCHQLLIEETG VRKDLTDIPLTGEVLTWFTDGSSYVV EGKRMAGAAVVDGTRTIWASSLPEG TSAQKAELMALTQALRLAEGKSINIY TDSRYAFATAHVHGAIYKQRGLLTSA GREIKNKEEILSLLEALHLPKRLAIHC PGHQKAKDPISRGNMADRVAKQAA QGVNLLPMIETPKAPEPGRQYTLEDW QEIKKIDQFSETPEGTCYTSKGKEILPH KEGLEYVQQIHRLTHLGTKHLQQLVR	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			TSPYHVLRLPGVADSVVKHCVPCQLV NANPSRIPPGKRLRGSHPGAHWEDF TEVKPAKYGNKYLLVFVDTFSGWVE AYPTKKETSTVVAKKILEEIFPRFGIPK VIGSDNGPAFVAQVSQGLAKILGIDW KLHCAYRPQSSGQVERMNRTIKETLT KLTAETGVNDWIALLPFVLFRVRNTP GQFGLTPYELLYGGPPPLVEIASVHSA DVLLSQPLFSRLKALEWVRQRAWRQ LREAYSGGGDLQIPHRFQVGDSVYVR RHRAGNLETRWKGPYHVLLTTPTAV KVEGISTWIIHASHVKPAPPPDSGWKA EKTENPLKLRLHRVVPYSVNNFSS (SEQ ID NO: 1559)	
--	--	--	---	--

POL_ SFV1 - только о остатк и	P23074	Вирус пенистост и обезьян 1 типа	MDPLQLLQPLEAEIKGTKLKAHWDS GATITCVPEAFLEDERPIQTMLIKTIHG EKQQDVYYLTFKVQGRKVEAEVLAS PYDYILLNPSDVPWLMKKPLQLTVLV PLHEYQERLLQQTALPKEQKELLQKL FLKYDALWQHWENQVGHRRIKPHNI ATGTLAPRPQKQYPINPKAKPSIQIVID DLLKQGVLIQQNSTMNTPVYPVPKPD GKWRMVLDYREVNKTIPLIAAQNH SAGILSSIYRGKYKTTLDLTNGFWAH PITPESYWLTAFTWQGKQYCWTRLPO GFLNSPALFTADVVDLLKEIPNVQAY VDDIYISHDDPQEHLEQLEKIFSILLNA GYVVSLLKKSEIAQREVEFLGFNITKEG RGLTDTFKQKLLNITPPKDLKQLQSIL GLLNFAFNIPNYSELVKPLYTIVANA NGKFISWTEDNSNQLQHIISVLNQAD NLEERNPETRLIKVNSSPSAGYIRYY NEGSKRPIMYVNYIFSKAEAKFTQTE KLLTTMHKGLIKAMDLAGQEIIVY SPIVSMTKIQRTPPERKALPVRWITW MTYLEDPRIQFHYDKSLPELQQIPNVT EDVIAKTKHPSEFAMVFYTDGSAIKH PDVNKSHSAGMGIAQVQFIPEYKIVH QWSIPLGDHTAQLAEIAAVEFACKKA LKISGPVLIVTDSFYVAESANKELPYW KSNGFLNNKKKPLRHVSKWKSIAECL QLKPDIIIMHEKGHQQPMITLHTEGN NLADKLATQGSYVHCNTTPSLDAEL DQLLQGHYPPGYPKQYKYTLEENKLI VERPNGIRIVPPKADREKIISTAHNIAH TGRDATFLKVSSKYWWPNLRKDVVK SIRQCKQCLVTNATNLTSPPILRPVKP LKPFDKFYIDYIGPLPPSNGYLHVLVV VDSMTGFVWLYPTKAPSTSATVKAL	IPR043502, SSF56672, IPR000477, PF00078
---	--------	---	---	--

			NMLTSIAIPKVLHSDQGAAFTSSTFAD WAKEKGIQLEFSTPYHPQSSGKVERK NSDIKRLLTKLLIGRPAKWYDLLPVV QLALNNSYSPSSKYTPHQLLFGVDSN TPFANSDTLDLSREEELSLQEISSLH QPTSPPASSRSWSPSVGQLVQERVAR PASLRPRWHKPTAILEVVPNPRTVIILD HLGNRRTVSVDNLKLTAYQDNGTSN DSGTMALMEEDESSTSST (SEQ ID NO: 1560)	
--	--	--	---	--

POL_ MPM V - только о остатк и	P07572	Вирус обезьян Мейзона- Пфайзера	MGQELSQHRYVEQLKQALKTRGVK VKYADLLKFFDFVKDTCWFPQEGTI DIKRWRRVGDCFQDYNTFGPEKVP VTAFSYWNLIKELIDKKEVNPQVMAA VAQTEEILKSNSQTDLTKTSSQNPDLDL ISLDSDEGAKSSSLQDKGLSSTKKPK RFPVLLTAQTSKDPEDPNPSEVDWDG LEDEAAKYHNPDWPPFLTRPPPYNKA TPSAPTVMVAVNPKEELKEKIAQLEE QIKLEELHQALISKLQKLKTGNETVTH PDTAGGLSRTPHWPGQHIPKGKCCAS REKEEQIPKDIFVTETVDGQGQAWR HHNGFDFAVIKELKTAASQYGATAPY TLAIVESVADNWLTPTDWNTLVRAV LSGGDHLLWKSEFFENCRDTAKRNQ QAGNGWDFDMLTGSGNYSSTDAQM QYDPGLFAQIQAAATKAWRKLPVKG DPGASLTGVKQGPDEPFADFVHRLIT TAGRIFGSAEAGVDYVKQLAYENAN PACQAAIRPYRKKTDLTGYIRLCSDIG PSYQQGLAMAAAFSGQTVKDFLNNK NKEKGGCCFKCGKKGHFAKNCHEHA HNNAEPKVPGLCPRCKRGKHWANEC KSKTDNQGNIPIPHQGNRVEGPAPGP ETSLWGSQLCSSQQKQPISKLTRATPG SAGLDLCSTSHTVLTPEMGPQALSTGI YGPLPPNTFGLILGRSSITMKGLQVYP GVIDNDYTGEIKIMAKAVNNIVTVSQ GNRIAQLILLPLIETDNKVQQPYRGQG SFGSSDIYWVQPITCQKPSLTLWLDD KMFTGLIDTGADVTHKLEDWPPNWPI TDTLTNLRGIGQSNNPKQSSKYL TWR DKENNSGLIKPFVIPNLPVNLWGRDL LSQMKIMMCSPNDIVTAQMLAQGYS PGKGLGKKENGILHPIPNQGQSNKKG	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
--	--------	--	---	--

FGNFLTAAIDILAPQQCAEPITWKSDE
PVWVDQWPLTNDKLAAAQQLVQEQ
LEAGHITESSSPWNTPIFVIKKKSGKW
RLLQDLRAVNATMVLMGALQPGLPS
PVAIPQGYLKIIIDLKDCFFSIPLHPSDQ
KRFAFSLPSTNFKPEPMQRFQWKVLPQ
GMANSPTLCQKYVATAIHKVRHAWK
QMYIIHYMDDILIAGKDGQQVLQCFD
QLKQELTAAGLHIAPEKVQLQDPYTY
LGFELNGPKITNQKAVIRKDKLQTLN
DFQKLLGDINWLRPYLKLTTGDLKPL
FDTLKGSDPNSHRSLSKEALASLEK
VETAIAEQFVTHINYSPLIFLIFNTAL
TPTGLFWQDNPMWIHLPA SPKKVLL
PYYDAIADLIILGRDH SKKYFGIEPSTII
QPYSKSQIDWLMQNTEMWPIACASF
VGILDNHYPNKLIIQFCKLHTFVFPQII
SKTPLNNALLVFTDGSSTGMAAYTLT
DTTIKFQTNLNSAQLVELQALIAVLSA
FPNQPLNIYTDSAYLAHSIPLLETVAQI
KHISE TAKLFLQCQQLIYNRSIPFYIGH
VRAHSGLPGPIAQGNQRADLATKIVA
SNINTNLESAQNAHTLHHLNAQTLRL
MFNIPREQARQIVKQCPICVTYLPVPH
LGVNPRGLFPNMIWQMDVTHYSEFG
NLKYIHVSIDTFSGFLLATLQTGETTK
HVITHLLHCFSSIIGLPKQIKTDNGPGYT
SKNFQEFCSTLQIKHITGIPYNPQGQGI
VERAHL SLKTTIEKIKKGEWYPRKGT
PRNILNHALFILNFLNLDDQNKSAAD
RFWHNNPKKQFAMVKWKDPLDNTW
HGPDPVLIWGRGSVCVYSQTYDAAR
WLPERLVRQVSNNNQSRE (SEQ ID
NO: 1561)

POL_ MMT VB - только о остатк и	P03365	Вирус опухоли молочной железы мышей	MGVSGSKGQKLFVSVLQRLLSERGLH VKESSAIEFYQFLIKVSPWFPEEGGLN LQDWKRVGREMKRYAAEHGTD SIPK QAYPIWLQLREILTEQSDLVLLSAEAK SVTEEELEEGLTGLLSTSSQEKTYGTR GTAYAEIDTEVDKLSEHIYDEPYEEKE KADKNEEKDHVRKIKKVVQRKENSE GKRKEKDSKAFLATDWNDDDLSPED WDDLEEQA AHYHDDDELILPVKRKV VKKKPQALRRKPLPPVGFAGAMAEA REKGDLTFTFPVVMGESDEDDTPVW EPLPLKTLKELQSAVRTMGPSAPYTL QVVDMMVASQWLTPSDWHQTARATL SPGDYVLWRTEYEEKSKEMVQKAAG KRKGKVS LDMLLGTGQFLSPSSQIKL SKDVLKDVT TNAVLA WRAIPPPGVK KTVLAGLKQGNEESYETFISRLEEAV YRMMPRGEGSDILIKQLAWENANSLC QDLIRPIRKTGTIQDYIRACLDASPAV VQGMAYAAAMRGQKYSTFVKQTYG GGKGGQGAEGPVCFCGKTGHIRKD CKDEKGSKRAPPGLCPRCKKGYHWK SECKSKFDK DGNPLP PLETNAENSKN LVKGQSPSPAQKGDGVKGSGLNPEAP PFTIHDLP RGT PGSAGLDLSSQKDLILS LEDGVSLVPTLVKGTLPEGTTGLIIGR SSNYKKGLEVLPGVIDSDFQGEIKVM VKA AKN AVIIHKGERIAQLLLL PYLKL PNPVIKEERGSEFGSTSHVHWVQEIS DSRPMLHIYLNRRFLGLLDTGADKT CIAGRDPANWPIHQTESSLQGLGM ACGVARSSQPLRWQHEDKSGIHPFVI PTLPFTLWGRDIMKDIK VRLMTDSPD DSQDLMIGAIESNLFADQISWKSDQP VWLNQWPLKQEK LQALQQLVTEQLQ	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
---	--------	---	---	--

LGHLEESNSPWNTPVFVIKKKSGKWR
 LLQDLRAVNATMHDMGALQPGLPSP
 VAVPKGWEIIIDLQDCFFNIKLHPEDC
 KRFAFSVPSPNFKRPYQRFQWKVLPQ
 GMKNSPTLCQKFVDKAILTVRDKYQ
 DSYIVHYMDDILLAHPSRSIVDEILTS
 MIQALNKHGLVVSTEKIQKYDNLKYL
 GTHIQGDSVSYQKLQIRTDKLRRTLND
 FQKLLGNINWIRPFLKLTTGELKPLFEI
 LNGDSNPISTRKLTPEACKALQLMNE
 RLSTARVKRLDLSQPWSLCILKTEYTP
 TACLWQDGVVEWIHLPHISPKVITPY
 DIFCTQLIIGRHRSEKELFSKDPDYIVV
 PYTKVQFDLLLQEKEDWPISLLGFLG
 EVHFHLPKDPLLTFTLQTAIIFPHMTST
 TPLEKGIVIFTDGSANGRSVTYIQGRE
 PIIKENTQNTAQQAIEIVAVITAFEEVS
 QPFNL YTDSKYVTGLFPEIETATLSR
 TKIYTELKHLQRLIHKRQEKFYIGHIR
 GHTGLPGPLAQGNAYADSLTRILTAL
 ESAQESHALHHQNAAALRFQFHITRE
 QAREIVKLCPCPDWGHAPQLGVNP
 RGLKPRVLWQMDVTHVSEFGKLKYV
 HVTVD TYSHFTFATARTGEATKDVLQ
 HLAQSFA YMGIPQKIKTDNAPAYVSR
 SIQEFLARWKISHVTGIPYNPQGQAIV
 ERTHQNIKAQLNKLQKAGKYYTPHH
 LLAHALFVLNVNMDNQGHTAAERH
 WGPISADPKPMVMWKDLLTGSWKGP
 DVLITAGRGYACVFPQDAETPIWVPD
 RFIRPFTERKEATPTPGTAEKTPPRDE
 KDQQESPKNESPHQREDGLATSAGV
 DLRSGGGP (SEQ ID NO: 1562)

POL_ MLV MS - только о остатк и	P03355	Вирус лейкоза мышей Молони	MGQTVTTPLSLTLGHWKDVERIAHN QSVDVKKRRWVTFCSAEWPTFNVG WPRDGTfNRDLITQVKIKVFSPGPHG HPDQVPYIVTWEALAFDPPPWVKPFV HPKPPPPLPPSAPSLPLEPPRSTPPRSSL YPALTPSLGAKPKPQVLSDSGGPLIDL LTEDPPPYRDPRPPPSDRDGNGGEATP AGEAPDPSPMASRLRGRREPPVADST TSQAFPLRAGGNGQLQYWPFSDDL NWKNNNPSFSEDPGKLTALIESVLITH QPTWDDCQQLGTLLTGEEKQRVLL EARKAVRGDDGRPTQLPNEVDAAFP LERPDWDYTTQAGRNHLVHYRQLLL AGLQNAGRSPTNLAKVKGITQGPNES PSAFLERLKEAYRRYTPYDPEDPGQE TNVSMFSIWQSAPDIGRKLERLEDLK NKTLGDLVREAEIFNKRETPEEREER IRRETEEKEERRRTEDEQKEKERDRR RHREMSKLLATVVSGQKQDRQGGER RRSQLDRDQCAYPEKEKGHWAKDCP KKPRGPRGPRPQTSLTLDDQGGQGQ EPPPEPRITLKVGGQPVTFLVDTGAQH SVLTQNPGLSDKSAWVQGATGGKR YRWTTDRKVHLATGKVTHSFLHVPD CPYPLLGRDLLTKLKAQIHFEQSGAQ VMGPMGQPLQVLTNIEDEHRLHETS KEPDVSLGSTWLSDFPQAWAETGGM GLAVRQAPLIPLKATSTPVSIKQYPM SQEARLGIKPHIQRLDQGILVPCQSP WNTPLLVPKKPGTNDYRPVQDLREV NKRVEDIHPTVPNPYNLLSGLPPSHQ WYTVLDLKDAFFCLRLHPTSQPLFAF EWRDPEMGIQGLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYV DDLILLAATSELDCQQGTRALLQTLGN	IPR043502, SSF56672, IPR000477, PF00078, cd03715
---	--------	-------------------------------------	--	--

LGYRASAKKAQICQKQVKYLGYYLK
EGQRWLTEARKETVMGQPTPKTPRQ
LREFLGTAGFCRLWIPGFAEMAAPLY
PLTKTGTLFNWGPDQQKAYQEIKQAL
LTAPALGLPDLTKPFELFVDEKQGYA
KGVLTQKLGPWRRPVAYLSKKLDPV
AAGWPPCLRMVAAIAVLTKDAGKLT
MGQPLVILAPHAVEALVKQPPDRWLS
NARMTHYQALLDTRVQFGPVVAL
NPATLLPLPEEGLQHNCLDILAEAHGT
RPDLTDQPLPDADHTWYTDGSSLLQE
GQRKAGAAVTTETEVIWAKALPAGT
SAQRAELIALTQALKMAEGKKLNVY
TDSRYAFATAHIHGEIYRRRGLLTSEG
KEIKNKDEILALLKALFLPKRLSIHCP
GHQKGHSAEARGNRMADQAARKAA
ITETPDTSTLLIENSSPYTSEHFHYTVT
DIKDLTKLGAIYDKTKKYWVYQGKP
VMPDQFTFELDLHLHLTHLSFSKMK
ALLERSHSPYYMLNRDRTLKNITETC
KACAQVNASKSAVKQGTRVRGHRPG
THWEIDFTEIKPGLYGYKYLLVFIDTF
SGWIEAFPTKKETAKVVTKKLEEIFP
RFGMPQVLGTDNGPAFVSKVSQTVA
DLLGIDWKLHCAYRPQSSGQVERMN
RTIKETLTKLTLATGSRDWVLLLPLAL
YRARNTPGPHGLTPYEILYGAPPPLVN
FPDPMTRVTNSPSLQAHQALYLVQ
HEVWRPLAAAYQEQLDRPVVPHPYR
VGDTVWVRRHQTKNLEPRWKGPYT
VLLTTPTALKVDGIAAWIHAAHVKA
ADPGGGPSSRLTWRVQRSQNPLKIRL
TREAP (SEQ ID NO: 1563)

POL_ HTL1 А - только о остатк и	P03362	Вирус Т- клеточног о лейкоза человека 1 типа	MGQIFSRASPIRPPRGLAAHHWLN LQAAYRLEPGPSSYDFHQLKKFLKIA LETPARICPINYSLLASLLPKGYPGRV NEILHILIQTQAQIPSRPAPPPSSPTH PPDSDPQIPPPYVEPTAPQVLPVMHPH GAPPNHRPWQMKDLQAIKQEVSQAA PGSPQFMQTIRLAVQQFDPTAKDLQD LLQYLCSSLVASLHHQQLDLISEAET RGITGYNPLAGPLRVQANNPQQQGLR REYQQLWLAFAALPGSAKDPSWASI LQGLEEYPYHAFVERLNIALDNGLPEG TPKDPILRSLAYSANKECQKLLQAR GHTNSPLGDMLRACQTWTPKDKTKV LVVQPKKPPPNQPCFRCGKAGHWSR DCTQPRPPPGPCPLCQDPHTWKRD CPRLKPTIPEPEPEEDALLDLPADIPHPK NLHRGGGLTSPPTLQQVLPNQDPASIL PVIPLDPARRPVIKAQVDTQTSHPKTI EALLDTGADMTVLPIALFSSNTPLKNT SVLGAGGQTQDHFKLTSPLVLRPFR TTPIVLTSCLVDTKNNWAIIGRDALQQ CQGVLYLPEAKRPPVILPIQAPAVLGL EHLPRPPQISQFPLNPERLQALQHLVR KALEAGHIEPYTGPGNNPVFPVKKAN GTWRFIHDLRATNSLTIDLSSSSPGPP DLSSLPTTLAHLQTIDLRDAFFQIPLPK QFQPYFAFTVPQQCNYPGTRYAWK VLPQGFKNSPTLFEMQLAHILQPIRQA FPQCTILQYMDDILLASPSHEDLLLLS EATMASLISHGLPVSENKTQQTPGTIK FLGQIISPNHLTYDAVPTVPIRSRWAL PELQALLGEIQWVSKGTPTLRQPLHSL YCALQRHTDPRDQIYLNPSQVQSLVQ LRQALSQNCRSRLVQTLPLLGAIMLT LTGTTTVVFQSKEQWPLVWLHAPLP	IPR043502, SSF56672, IPR000477, PF00078
---	--------	--	--	--

		HTSQCPWGQLLASAVLLLDKYTLQS YGLLCQTIHHNISTQTFNQFIQTSDHPS VPILLHSHRFKNLGAQTGELWNTFL KTAAPLAPVKALMPVFTLSPVIINTAP CLFSDGSTSRAAYILWDKQILSQRSFP LPPPHKSAQRAELLGLLHGLSSARSW RCLNIFLDSKYLYHYLRTLALGTFQG RSSQAPFQALLPRLLSRKVVYLHHVR SHTNLPDPISRLNALTDALLITPVLQLS PAELHSFTHCGQTALTLQGATTTEAS NILRSCHACRGGNPQHQMPRGHIRRG LLPNHIWQGDITHFKYKNTLYRLHV WVDTFSGAISATQKRKETSSEAISSLL QAIAHLGKPSYINTDNGPAYISQDFLN MCTSLAIRHTTHVPYNPTSSGLVERSN GILKTLLYKYFTDKPDLPMDNAL SIAL WTINHLNVLTNCHKTRWQLHHSPRL QPIPETRSLSNKQTHWYYFKLPGLNS RQWKGPQEALQEAAGAALIPVSASSA QWIPWRLKRAACPRPVGGPADPKE KDLQHHG (SEQ ID NO: 1564)	
--	--	---	--

POL_ FOA MV - только о остатк и	P14350	Спумарет ровирус человека	MNPLQLLQPLPAEIKGTKLLAHWDSG ATITCIPESFLEDEQPIKKTLIKTIHGEK QQNVYYYVTFKVKGRKVEAEVIASPY EYILLSPDVPWLTQQPLQLTILVPLQ EYQEKILSKTALPEDQKQQLKTLFVK YDNLWQHWENQVGHRKIRPHNIATG DYPPRPQKQYPINPKAKPSIQIVDDL KQGVLTQNSTMNTPVYPVPKPDGR WRMVDYREVNKTIPLTAAQNQHSA GILATIVRQKYKTTLDLANGFWAHPI TPESYWLTAFTWQKGQYCWTRLPPQG FLNSPALFTADVVDLLKEIPNVQVYV DDIYLSHDDPKEHVQQLEKVFQILLQ AGYVVSLLKSEIGQKTVEFLGFNITKE GRGLTDTFKTKLLNITPPKDLKQLQSI LGLLNFAFNFIPLFAELVQPLYNLIAS AKGKYIEWSEENTKQLNMVIEALNTA SNLEERLPEQRLVIKVNTPSPSAGYVRY YNETGKKPIMYLNIVFSKAEKFSML EKLLTTMHKALIKAMDLMGQEILV YSPIVSMTKIQKTPLPERKALPIRWIT WMTYLEDPRIQFHYDKTLPELKHIPD VYTSSQSPVKHPSQYEGVFYTDGSAI KSPDPTKSNNAGMGIVHATYKPEYQ VLNQWSIPLGNHTAQMAEIAAVEFAC KKALKIPGPVLVITDSFYVAESANKEL PYWKSNGFVNNKKKPLKHISKWKSIA ECLSMKPDITIQHEKGISLQIPVFILKG NALADKLATQGSYVVCNTKKPNLD AELDQLLQGHYIKGYPKQYTYFLEDG KVKVSRPEGVKIIPPQSDRQKIVLQAH NLAHTGREATLLKIANLYWWPNMRK DVVKQLGRCQQCLITNASNKASGPIL RPDRPQKPFDKFFIDYIGPLPPSQGYL YVLVVVDGMTGFTWLYPTKAPSTSA	IPR043502, SSF56672, IPR000477, PF00078
---	--------	---------------------------------	---	--

			TVKSLNVLTSAIPKVIHSDQGAAFTSS TFAEWAKERGIHLEFSTPYHPQSGSK VERKNSDIKRLLTCLLVGRPTKWYDL LPVVQLALNNTYSPVLKYTPHQLLFG IDSNTPFANQDTLCLTREEELSLLQEIR TSLYHPSTPPASSRSWSPVVGQLVQE RVARPASLRPRWHKPSTVLKVLNPRT VVILDHLGNNRTVSIDNLKPTSHQNG TTNDTATMDHLEKNE (SEQ ID NO: 1565)	
--	--	--	---	--

POL_ BLVJ - только о остатк и	P03361	Вирус лейкоза крупного рогатого скота	MGNPSYNPPAGISPSDWLNLLQSAQ RLNPRPSPSDFTDLKNYIHWFKTQK KPWTFTSGGPTSCPPGRFGRVPLVLA TLNEVLSNEGGAPGASAPEEQPPPYD PPAILPIISEGNRNRHRAWALRELQDI KKEIENKAPGSQVWIQTLRLAILQADP TPADLEQLCQYIASPVDQTAHMTSLT AAIAAAEAANTLQGFNPKTGTLTQQS AQPNAAGDLRSQYQNLWLQAGKNLPT RPSAPWSTIVQGPAESSVEFVNRLQIS LADNLPDGVPKPIIDSLSYANANREC QQILQGRGPVAAVGQKLQACAQWAP KKNQPALLVHTPGPKMPGPRQPAPKR PPPGPCYRCLKEGHWARDCPTKATGP PPGPCPICKDPSHWKRDCPTLKSKNK LIEGGLSAPQTITPITDSLSEAELECLLS IPLARSRPSVAVYLSGPWLQPSQNQA LMLVDTGAENTVLPQNWLV RDYPRI PAAVLGAGGVSRNRYNWLQGPLTLA LKPEGPFITIPKILVDTSDKWQILGRDV PSRLQASISIPEEVRPPVVGVLDTPPSH IGLEHLPPPPEVPQFPLNLERLQALQD LVHRSLEAGYISPWDGPGNNPVFPVR KPNGAWRFVHDLRATNALTKPIPALS PGPPDLTAIPTHPPHIICLDLKDAFFQIP VEDRFRFYLSFTLPSPGGLQPHRRFA WRVLPQGFINSFALFERALQEPLRQVS AAFSQSLLVSYMDDILYASPTEEQRS QCYQALAAARLRDLGFQVASEKTSQTP SPVPFLGQMVHEQIVTYQSLPTLQISS PISLHQLQAVLGDLQWVSRGTPPTRR PLQLLYSSLKRHHDPRAIIQLSPEQLQ GIAELRQALSHNARSRYNEQEPLLAY VHLTRAGSTLVLFQKGAQFPLAYFQT PLTDNQASPWGLLLLLLGCQYLQTQAL	IPR043502, SSF56672, IPR000477, PF00078
---	--------	---	---	--

			SSYAKPILKYYHNLPKTSLDNWIQSSE DPRVQELLQLWPQISSQGIQPPGPWK TLITRAEVFLTPQFSPDPIPAALCLFSD GATGRGAYCLWKDHLLDFQAVPAPE SAQKGELAGLLAGLAAAPPEPVNIWV DSKYLYSLLRTLVLGAWLQPDVPVPSY ALLYKSLLRHPAIVVGHVRSRSSASH PIASLNNYVDQLLPLETPEQWHKLTH CNSRALSRWPNPRISAWDPRSPATLC ETCQKLNPTGGGKMRTIQRGWAPNHI WQADITHYKYKQFTYALHVFVDTYS GATHASAKRGLTTQTIEGLLEAIVHL GRPCKLNTDQGANYTSKTFVRFCQQF GVSLSHHVPYNPTSSGLDERTNGLLK LLSKYHLDEPHLPMTQALSRAWTH NQINLLPILKTRWELHHSPLAVISEG GETPKGSDKLFLYLLPGQNNRRWLGP LPALVEASGGALLATDPPVWVPWRL LKAFKCLKNDGPEDAHNRSSDG (SEQ ID NO: 1566)	
--	--	--	--	--

O4189 4_9RE TR - только о остатк и	O41894	Вирус пенистост и крупного рогатого скота	MPALRPLQVEIKGNHLKGYWDSGAEI TCVPAIYHIEEQPVGKKLITTIHNEKEH DVYYVEMKIEKRKVQCEVIATALDY VLVAPVDIPWYKPGPLELTIKIDVESQ KHTLITESTLSPQGQMRLKKLLDQYQ ALWQCWENQVGHRRIEPHKIATGAL KPRPQKQYHINPRAKADIQIVDDLRL QGVLRQQNSEMNTPVYPVPKADGRW RMVLDYREVNKVTPLVATQNCBSASI LNTLYRGPYKSTLDLANGFWAHPKP EDYWITAFTWGGKTYCWTVLPQGFL NSPALFTADVVDILKDIPNVQVYVDD VYVSSATEQEHLDILETIFNRLSTAGYI VSLKSKLAKETVEFLGFSISQNGRGL TDSYKQKLMDLQPPTTLRQLQSILGLI NFARNFLPNFAELVAPLYQLIPKAKG QCIPWTMDHTTQLKTIIQALNSTENLE ERRPDVDLIMKVHISNTAGYIRFYNH GGQKPIAYNNALFTSTELKFTPTEKIM ATIHKGLLKALDLSLGKEIHVYSAIAS MTKLQKTPLSERKALSIRWLKWQTYF EDPRIKFHHDATLPDLQNLVPVQQDT GKEMTILPLLHYEAIFYTDGSAIRSPK PNKTHSAGMGIIQAKFEPDFRIVHLWS FPLGDHTAQYAEIAAFEF AIRRATGIR GPVLIVTDSNYVAKSYNEELPYWESN GFVNNKKKTLKHISKWKAIAECKNL KADIVHIEPGHQPAEASPHAQGNAL ADKQAVSGSYKVFSNELKPSLDAELE QVLSTGRPNPQGYPNKYEYKLVNGL CYVDRRGEEGLKIIPPKADRVKLCQL AHDGPGSAHLGRSALLKLQQKYW WPRMHIDASRIVLNCTVCAQTNSTNQ KPRPPLVIPHDTKPFQVWYMDYIGPL PPSNGYQHALVIVDAGTGFTWIYPTK	IPR043502, SSF56672, IPR000477, PF00078
--	--------	--	--	--

AQTANATVKALTHLTGTAVPKVLHS
DQGPFTSSILADWAKDRGIQLEHSA
PYHPQSSGKVERKNSEIKRLLTKLLA
GRPTKWYPLIPVQLALNNTPNTRQK
YTPHQLMYGADCNLPFENLDTLTLTR
EEQLAVLKEVRDGLLDLYSPSPQTTA
RSWTPSPGLLVQERVARPAQLRPKW
RKPTPIKKVLNERTVIIDHLGQDKVVS
IDNLKPAAHQKLAQTPDSEICPSATP
CPPNTSLWYDLDTGTWTCQRCGYQC
PDKYHQPQCTWSCEDRCGHRWKECG
NCIPQDGSDDASAVAAVEI (SEQ ID
NO: 1567)

POL_ MLV BM - только о остатк и	Q7SVK7	Вирус лейкоза мышей	MGQTVTTPLSLTLEHWGDVQRIASNQ SVGVKKRRWVTFCSAEWPTFGVGWP QDGTFNLDIILQVKSKVFSPGPHGHPD QVPYIVTWEAIAIAYEPPPWVKPFVSPK LSLSPTAPILPSGPSTQPPPRSALYPAF TPSIKPRPSKPQVLSDDGGPLIDLLTED PPPYGEQGPSSPDGDGDREEATSTSEI PAPSPMVSR LRGKRDP PAADSTTSRA FPLRLGGNGQLQYWPFSSSDLYNWK NNNPSFSEDPGKLTALIESVLTTHQPT WDDCQQLLGTLTGEEKQRVLLERAR KAVRGNDGRPTQLPNEVNSAFPLERP DWDYTTPEGRNHLVLYRQLLLAGLQ NAGRSPTNLAKVKGITQGPNESPSAFL ERLKEAYRRYTPYDPEDPGQETNVSM SFIWQSAPAIGRKLERLEDLKSCTLGD LVREA EKIFNKRETPEEREERIRRETEE KEERRRAGDEQREKERDRRRQREMS KLLATVVTGQRQDRQGGERRRPQLD KDQCA YCKEKGHWAKDCPKKPRGP RGPRPQTSLTLDDQGGQGQEPPPEP RITLTVGGQPVTFLVDTGAQH SVLTQ NPGPLSDRSAWVQGATGGKRYRWTT DRKVHLATGKVTHSFLHVPDCPYPLL GRDLLTKLKAQIHFE GSGAQVVGPKG QPLQVLT LGIEDEYRLHETSTEPDVSL GSTWLSDFPQAWAETGGMGLAVRQ APLIPLKATSTPVSIIQQYPM SHEARLG IKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDI HPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPGM GISGQLTWTRLPQGFKNSPTLFDEAL HRDLADFRIQHPDLILLQYVDDILLAA TSELDCQQGTRALLQTLGDLGYRASA	IPR043502, SSF56672, IPR000477, PF00078, cd03715
---	--------	---------------------------	---	--

		KKAQICQKQVKYLGYPVLLREGQRWLT EARKETVMGQPVPKTPRQLREFLGTA GFCRLWIPGFAEMAAPLYPLTKTGTL FSWGPDQQKAYQEIKQALLTAPALGL PDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCL RMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHY QAMLLDTRVQFGPVVALNPATLLPL PEEGAPHDCLEILAEHGTTRPDLTDQP IPDADHTWYTDGSSFLQEGQRKAGA AVTTETEVIWAGALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFA TAHIHGEIYRRRGLLTSEGREIKNKSEI LALLKALFLPKRLSIHCLGHQKGDSA EARGNRLADQAAREAAIKTPPDTSTL LIEDSTPYTPAYFHYTETDLKKLRDLG ATYNQSKGYWVFQGKPVMPDQFVFE LLDSLHRLTHLGYQKMKALLDRGESP YYMLNRDKTLQYVADSCTVCAQVN ASKAKIGAGVRVRGHRPGTHWEIDFT EVKPGLYGYKYLLVFVDTFSGWVEA FPTKRETARVVSCKLLEEIFPRFGMPQ VLGSDNGPAFTSQVSQSVADLLGIDW KLHCAYRPQSSGQVERINRTIKETLTK LTLAAGTRDWVLLLPLALYRARNT GPHGLTPYEILYGAPPPLVNFHDPDM SELTNSPSLQAHQALQTVQREIWKP LAEAYRDRLDQPVIPHPFRIGDSVWV RRHQTKNLEPRWKGPYTVLLTTP TAL KVDGISAWIHAAHVKAATTPPIKPSW RVQRSQNPLKIRLTRGAP	
--	--	---	--

Таблица 31. Иллюстративные димерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	Номер доступа	Организм	Последовательность	Сигнатуры RT
Q8313 3_AVI MA	Q83133	Вирус, ассоциированный с миелобластозом птиц, 1 типа	RATVLTVALHLAIPKWKPNHTPVWID QWPLPEGKLVALTQLVEKELQLGHIEPS LSCWNTPVFVIRKASGSYRLHDLRAVN AKLVPGFVAVQQGAPVLSALPRGWPLMV LDLKDCFFSIPLAEQDREAFATLPSVNN QAPARRFQWKVLPQGMTCSPTICQLIVG QILEPLRLKHPSLRMLHYMDDLALLAASS HDGLEAAGEEVISTLERAGFTISPDKVQR EPGVQYLYGYKLGSTYVAPVGLVAEPRIA TLWDVQKLVGSLQSVRPALGIPPRLMGP FYEQLRGSDPNEAREWNLDMKMAWRE IVQLSTTAALERWDPALPLEGAVARCEQ GAIGVLGQGLSTHPRPCLWLFSTQPTKA FTAWLEVLTLITKLRSASVRTFGKEVD ILLLPACFREDLPLPEGILLALRGFAGKIR SSDTPSIFDIARPLHVSLKVRVTDHPVPG PTVFTDASSSTHKGVVVWREGPRWEIKE IADLGASVQQLEARAVAMALLWPTTP TNVVTDSAFVAKMLLKMGQEGVPSTAA AFILEDALSQRSAMA AVLHVRSHSEVPG FFTEGNDVADSQATFQAYPLREAKDLH TALHIGPRALSKACNISMQQAREVVQTC PHCNSAPALEAGVNPRGLGPLQIWQTD TLEPRMAPRSWLAVTVDTASSAIVVTQH GRVTSVAAQHWWATAIAVLGRP KAIKT DNGSCFTSKSTREWLARWGIAHTTGIPG NSQGQAMVERANRLLKDKIRVLAEGDG FMKRIPTSKQGELLAKAMYALNHFERG ENTKTPIQKHWRPTVLTEGPPVKIRIETG EWEKGWNVLVWGRGYAAVKNRDTDK VIWVPSRKVKPDITQKDEVTKKDEASPL FAGISDWAPWEGEQEGLQEETASNKQE RPGEDTPAANES (SEQ ID NO: 1568)	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

POL_ SIVM 1	P05896	Вирус иммуноде фицита обезьян	MGARNSVLSGKKADELEKIRLRPGGKK KYMLKHVVWAANELDRFGLAESLLEN KEGCQKILSVLAPLVPTGSENLSLYNT VCVIWCIIHAEKVKHTEEAKQIVQRHLV METGTAETMPKTSRPTAPFSGRGGNYPV QQIGGNYTHLPLSPRTLNAWVKLIEEKK FGAEVVSQFQALSEGCLPYDINQMLNCV GDHQAAMQIIRDIINEEAADWDLQHPQQ APQQGQLREPSGSDIAGTTSTVEEQIQW MYRQQNPIPVGNIYRRWIQLGLQKCVR MYNPTNILDVKQGPKEPFQSYVDRFYKS LRAEQTDPAVKNWMTQTLLIQNANPDC KLVKGLGTNPTLEEMLTACQGVGGPG QKARLMAEALKEALAPAPIPFAAAQQK GPRKPIKCWNCGKEGHSARQCRAARRQ GCWKCGKMDHVMACPNRQAGFFRP WPLGKEAPQFPHGSSASGADANCSRRRT SCGSAKELHALGQAAERKQREALQGGD RGFAAPQFSLWRRPVVTAHIEGQPVEVL LDTGADDSIVTGIELGPHYTPKIVGGIGG FINTKEYKNVEIEVLGKRIKGTIMTGDTP INIFGRNLLTALGMSLNLPIAKVEPVKSP LKPGKDGPCLKQWPLSKEKIVALREICE KMEKDGQLEEAPPTNPYNTPTFAIKKKD KNKWRMLIDFRELNRVTQDFTEVQLGIP HPAGLAKRKRTVLDIGDAYFSIPLDEEF RQYTAFTLPSVNNAEPGKRYIYKVLPPQ WKGSPAIFQYTMRHVLEPFRKANPDVT LVQYMDDILIASDRTDLEHDRVVLQLKE LLNSIGFSSPEEKFKDPPFQWMGYELW PTKWKLQKIELPQRETWTVNDIQKLVG VLNWAAQIYPGIKTKHLCRLIRGKMTLT EEVQWTEMAEAEYEENKIILSQEQEGCY YQESKPLEATVIKSQDNQWSYKIHQEDK ILKVGKFAKIKNTHTNVRLLAHVQKI	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	--	---	--

			GKEAIVIWGQVPKFHLPVEKDVWEQW WTDYWQVTWIPEWDFISTPPLVRLVFNL VKDPIEGEETYYVDGSCSKQSKEGKAG YITDRGKDKVKVLEQTTNQQAELEAFL MALTDSGPKANIIVDSQYVMGIITGCPTE SESRLVNQIIEEMIKKTEIYVAWVPAHK GIGGNQEIDHLVSQGIRQVLFLEKIEPAQ EEHSKYHSNIKELVFKFGLPRLVAKQIV DTC DKCHQKGEA IHGQVNSDLGTWQM DCTHLEGKIVIVAVHVASGFIEAEVIPQE TGRQTALFLLKLASRWPITHLHTDNGAN FASQEVKMVAWWAGIEHTFGVPYNPQS QGVVEAMNHHLKNQIDRIREQANSVETI VLMAVHCMNFKRRGGIGDMTPAERLIN MITTEQEIQFQQSKNSKFKNFRVYYREG RDQLWKGP GELLWK GEGAVILKVGTDI KVVPRRKAKIIKDYG GKGEMDSSSHME DTGEAREVA (SEQ ID NO: 1569)	
--	--	--	---	--

POL_ RSVP	P03354	Вирус саркомы Райса	MEAVIKVISSACKTYCGKTSPSKKEIGA MLSLLQKEGLLMSPSDLYSPGSWDPITA ALSQRAMILGKSGELKTWGLVLGALKA AREEQVTSEQAKFWLGLGGGRVSPGP ECIEKPATERRIDKGEEVGETTVQRDAK MAPEETATPKTVGTSCYHCGTAIGCNCA TASAPPPPYVGSGLYPSLAGVGEQQGQG GDTPPGAEQSRAEPGHAGQAPGPALTD WARVREELASTGPPVVAMPVVIKTEGP AWTPLEPKLITRLADTVRTKGLRSPITM AEVEALMSSPLLPHDVTNLMRVILGPAP YALWMDAWGVQLQTVIAAATRDPRHP ANGQGRGERTNLNRLKGLADGMVGNP QGQAALLRPGELVAITASALQAFREVAR LAEPAGPWADIMQGPSESFVDFANRLIK AVEGSDLPPSARAPVIIDCFRQKSQPDIQ QLIRTAPSTLTTPGEIHKYVLDRQKTAPLT DQGIAAAMSSAIQPLIMAVVNRERDQG TGSQGRARGLCYTCGSPGHYQAQCPKKR KSGNSRERCQLCNGMGHNAKQCRKRD GNQGQRPGKGLSSGPWPGPEPPAVSLA MTMEHKDRPLVRVILTNTGSHPVKQRS VYITALLD SGADITIISEEDWPTDWPVME AANPQIHGIGGGIPMRKSRDMIELGVINR DGSLEPPLLLFPAVAMVRGSILGRDCLQ GLGLRLTNLIGRATVLTVALHLAIPKWK KPDHTPVWIDQWPLPEGKLVALTQLVE KELQLGHIEPSLSCWNTVPFVIRKASGSY RLLHDLRAVNAKLVPFGAVQQGAPVLS ALPRGWPLMVLDLKDCCFSIPLAEQDRE AFAFTLPSVNNQAPARRFQWKVLPQGM TCSPTICQLVVGQVLEPLRLKHPSLCML HYMDDL LLAASSHDGLEAAGEEVISTLE RAGFTISPDKVQREPGVQYLYGYKLGSTY VAPVGLVAEPRIATLWDVQKLVGSLQW	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
--------------	--------	---------------------------	---	--

			LRPALGIPRLMGPFYEQLRGSDPNear EWNLDMKMAWREIVRLSTTAALERWD PALPLEGAVARCEQGAIGVLGQGLSTHP RPCLWLFSTQPTKAFTAWLEVLTLITK LRASAVRTFGKEVDILLPACFREDLPLP EGILLALKGFAGKIRSSDTPSIFDIARPLH VSLKVRVTDHPVPGPTVFTDASSSTHKG VVVWREGPRWEIKEIADLGASVQQLEA RAVAMALLLWPTTPTNVVTDSAFVAK MLLKMGQEGVPSTAAAFILEDALSQRSA MAAVLHVRSHSEVPGFFTEGNDVADSQ ATFQAYPLREAKDLHTALHIGPRALSKA CNISMQQAREVVQTCPHCNSAPALEAG VNPRGLGPLQIWQTDFTLEPRMAPRSWL AVTVDTASSAIVVTQHGRVTSVAVQHH WATAIAVLGRPKAIKTDNGSCFTSKSTR EWLARWGIAHTTGIPGNSQGQAMVERA NRLLKDRIRVLAEGDGFMKRIPTSKQGE LLAKAMYALNHFERGENTKTPIQKHWR PTVLTEGPPVKIRIETGEWEKGWNVLV WGRGYAAVKNRDTDKVIWVPSRKVKP DITQKDEVTKKDEASPLFAGISDWIPWE DEQEGLOGETASNKQERPGEDTLAANE S (SEQ ID NO: 1570)	
--	--	--	---	--

POL_ HV2D 2	P15833	Вирус иммуноде фицита человека 2 типа	MGARGSVLSGKKKTDELEKVRLRPGGKK KYMLKHVVWAVNELDRFGLAESLLESK EGCQKILKVLAPLVPTGSENKSLFNIVC VIFCLHAEKVKDTEEAKKIAQRHLAAD TEKMPATNKPTAPPSGGNYPVQQLAGN YVHLPLSPRTLNAWVKLVEEKKFGEV VPGFQALSEGCTPYDINQMLNCVGEHQ AAMQIIREIINEEAADWDQQHPSPGPMP AGQLRDPRGSDIAGTTSTVEEQIQWMYR AQNPVPVGNIYRRWIQLGLQKCVRMYN PTNILDIKQGPKEPFQSYVDRFYKSLRAE QTDPAVKNWMTQTLLIQNANPDCKLVL KGLGMNPTLEEMLTACQGIGGPGQKAR LMAEALKEALTPAPIFAAVQQKAGKR GTVTCWNCGKQGHTARQCRAPRRQGC WKCGKTGHIMSKCPERQAGFLRVRTL KEASQLPHDPSASGSDTICTPDEPSRGHD TSGGDTICAPCRSSSGDAEKLHADGETT EREPRETLQGGDRGFAAPQFSLWRRPVV KACIEGQSVEVLLDTGVDDSIAGIELGS NYTPKIVGGIGGFINTKEYKDVEIEVVGK RVRATIMTGDTPINIFGRNINLNTLGMTLN FPVAKVEPVKVELKPGKDGPQIRQWPLS REKILALKEICEKMEKEGQLEEAPPTNPY NTPTFAIKKKDKNKWRMLIDFRELNKV TQDFTEVNWVFPTRQVAEKRRITVIDVG DAYFSIPLDPNFRQYTAFTLPSVNNAEPG KRYIYKVLPPQGWKGSQSICQYSMRKVL DPFRKANSDVIIIQYMDDILIASDRSDLE HDRVVSQKELLNDMGFSTPEEKFKQD PPFKWMGYELWPKKWKLQKIQLPEKEV WTVNAIQKLVGVLNWAAQLFPGIKTRH ICKLIRGKMTLTEEVQWTELAELQEN KIILEQEQUEGSYYKERVPLEATVQKNLA NQWTKYKIHQGNKVLKVGKYAKVKNTH	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	---	--	--

			TNGVRLLAHV VQKIGKEALVIWGEIPVF HLPVERETWDQWWTDYWQVTWIPEW DFVSTPPLIRLAYNLVKDPLEGRETYYT DGSCNRTSKEGKAGYVTD RGKDKVKV LEQTTNQQAELEAFALALTDSEPQVNIIV DSQYVMGIIAAQPTETESPIVAKIIEEMIK KEAVYVGWVPAHKGLGGNQEVDHLVS QGIRQVLFLEKIEPAQEEHEKYHGNVKE LVHKFGIPQLVAKQIVNSCDKCQQKGEA IHGQVNADLGTWQMDCTHLEGKIIIVAV HVASGFIEAEVIPQETGRQTALFLLKLAS RWPITHLHTDNGANFTSPSVKMVAWW VGIEQTFGVPYNPQSQGVVEAMNHHLK NQIDRLRDQAVSIETVVL MATHCMNFK RRGGIGDMTPAERLVNMITTEQEIQFFQ AKNLKFQNFQVYYREGRDQLWKG PGE LLWKGE GAVIIKVGTEIKVVP RRKAKIIR HYGGGKGLDCSADMEDTRQAREMAQS D (SEQ ID NO: 1571)	
--	--	--	---	--

POL_ HV1A 2	P03369	Вирус иммуноде фицита человека 1 типа	MGARASVLSGGELDKWEKIRLRPGGKK KYKCLKHIVWASRELERFAVNPGLLETSE GCRQILGQLQPSLQTGSEELRSLYNTVA TL YCVHQRIDVKDTKEALEKIEEEQNKS KKKAQQAAAAAGTGNSSQVSQNYPIVQ NLQGQMVHQAISPRTLNAWVKVVEEK AFSPEVIPMFSA LSEGATPQDLNTMLNT VGGHQAAMQMLKETINEEAAEWDRVH PVHAGPIAPGQMREPRGSDIAGTTSTLQ EQIGWMTNPNPIPVG E IYKRWIILGLNKI VRMYSPTSILDIRQGPK EPFRDYVDRFY KTLRAEQASQDVKNWMTETLLVQNAN PDCKTILKALGPAATLEEMMTACQGVG GPGHKARVLAEAMSQVTNPANIMMQR GNFRNQ RKT VKCFNCGKEGHI AKNCRA PRKKGCWRCGREGHQMKDCTERQANF LREDLAFLQGKAREFSSEQTRANSPTRR ELQVWGGENNSLSEAGADRQGTVSFNF PQITLWQRPLVTIRIGGQLKEALLDTGA DDTVLEEMNLPGKWKPKMIGGIGGFIK VRQYDQIPVEICGHKAIGTVLVGPTPVNI IGRNLLTQIGCTLNFPISPIETVPVKLKPG MDGPKVKQWPLTEEKIKALVEICTEME KEGKISKIGPENPYNTPVFAIKKKDSTK WRKLVDFREL NKRTQDFWEVQLGIPHP AGLKKKKS VTVLDVGDAYFSVPLDKDF RKYTAFTIPSINNETPGIRYQYNVLPQGW KGSPAIFQSSMTKILEPFRKQNPDIYIYQ YMDDL YVGSDLEIGQHRTKIEELRQHLL RWGFTTPDKKHQKEPPFLWMGYELHPD KWTVQPIMLPEKDSWTVNDIQKLVGKL NWASQIYAGIKVKQLCKLLRGTKALTE VIPLTEEA ELELAENREILKEPVHEVYYD PSKDLVAEIQQGQGW TYQIYQEPFK NLKTGKYARMRGAHTNDVKQLTEAVQ	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	---	---	--

			KVSTESIVIWGKIPKFKLPIQKETWEAW WMEYWQATWIPEWEFVNTPLVKLWY QLEKEPIVGAETFYVDGAANRETKLGK AGYVTDGRQKVVSADTTNQTTELQAI HLALQDSGLEVNIVTDSQYALGIIQAQP DKSESELVSQIEQLIKKEKVYLAWVPA HKGIGGNEQVDKLVSAGIRKVLFLNGID KAQEEHEKYHSNWRAMASDFNLPPVV AKEIVASCDKCQLKGEAMHGQVDCSPG IWQLDCTHLEGKIILVAVHVASGYIEAE VIPAETGQETAYFLLKLAGRWPVKTIHT DNGSNFTSTTVKAACWWAGIKQEFVIP YNPQSQGVVESMNNELKKIIGQVRDQA EHLKTAVQMAVFIHNFKRKGGIGGYSA GERIVDIIATDIQTKELQKQITKIQNFRVY YRDNKDPLWKGPAKLLWKGEGAVVIQ DNSDIKVVP RRKAKIIRDY GKQMAGDD CVASRQDED (SEQ ID NO: 1572)	
--	--	--	---	--

POL_ FIVPE	P16088	Вирус иммуноде фицита кошек	KEFGKLEGGASCSPSESNAASSNAICTSN GGETIGFVNYNKVGTTTTLEKRPEILIFV NGYPIKFLLDTGADITILNRRDFQVKNSI ENGRQNMIGVGGGKRGNTNYINVHLEIR DENYKTQCIFGNVCVLEDNSLIQPLLGR DNMIKFNIRLVMAQISDKIPVVKVKMKD PNKGPQIKQWPLTNEKIEALTEIVERLEK EGKVKRADSNNPWNTPVFAIKKKS ^G KW RMLIDFRELNKLTEKGAEVQLGLPHPAG LQIKKQVTVLDIGDAYFTIPLDPDYAPYT AFTLPRKNNAGPGRRFVWCSLPQGWILS PLIYQSTLDNIIQPFIRQNPQLDIYQYMD DIYIGSNLSKKEHKEKVEELRKLLLWWG FETPEDKLQEEPPYTWMGYELHPLTWTI QQKQLDIPEQPTLNELQKLAGKINWASQ AIPDLSIKALTNMMRGNQNLNSTRQWT KEARLEVQKAKKAIEEQVQLGYDPSK ELYAKLSLVGPHQISYQVYQKDPEKILW YGKMSRQKKKAENTCDIALRACYKIRE ESIIRIGKEPRYEIPTSREAWESNLINSPYL KAPPPEVEYIHAALNIKRALSMIKDAPIP GAETWYIDGGRKLGKA ^A AAAYWTDTG KWRVMDLEGSNQKAEIQALLALKAGS EEMNIITDSQYVINIILQQPDMMEGIWQE VLEELEKKTAFIDWVPGHKGIPGNEEV DKLCQTMMIIEGDGILDKRSEDAGYDLL AAKEIHLLPGEVKVIPTGVKLMLPKGY WGLIIGKSSIGSKGLDVLGGVIDEGYRGE IGVIMINVS ^R KSITLMERQKIAQLIILPCK HEVLEQGKVVMDSERGDNGYGSTGVFS SWVDRIEEAEINHEKFHSDPQYLRTEFN LPKMVAEEIRRKCPVCRIIGE ^Q VGGQLKI GPGIWQMDCTHFDGKIILVGIHVESGYI WAQIISQETADCTVKAVLQLLSAHNVTE LQTDNGPNFKNQKMEGV ^L NYMGVKHK	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
---------------	--------	--------------------------------------	--	--

			FGIPGNPQSQALVENVNHTLKVWVIQKFL PETTSLDNALSLAVHSLNFKRRGRIGGM APYELLAQQESLRIQDYFSAIPQKLQAQ WIYYKDQKDKKWKGPMRVEYWGQGS VLLKDEEKGYFLIPRRHIRRVPEPCALPE GDE (SEQ ID NO: 1573)	
--	--	--	--	--

POL_ EIAV Y	P03371	Вирус инфекцио нной анемии лошадей	TAWTFLKAMQKCSKKREARGSREAPET NFPDTTEESAQQICCTRDSSDSKSVPRSE RNKKGIQCQGEGSSRGSGPGQFVGVTY NLEKRPTTIVLINDTPLNVLLDTGADTSV LTTAHYNRLKYRGRKYQGTGIIGVGGN VETFSTPVTIKKKGRHIKTRMLVADIPVT ILGRDILQDLGAKLVLAQLSKEIKFRKIE LKEGTMGPKIPQWPLTKEKLEGAKETV QRLLESGKISEASDNNPYN SPIFVIKKRS GKWRLQLDLRELNKT VQVGTEISRGLP HPGGLIKCKHMTVLDIGDAYFTIPLDPEF RPYTAFTIPSINHQEPDKRYVWKCLPQG FVLSPYIYQKTLQEILQPFRETYPEVQLY QYMDDL FVG SNGSKKQHKELIIE LRAIL QKGFETPDDKLQEVPPYSWLGYQLCPE NWKVQKMQLD MVKNPTLNDVQKLMG NITWMSSGVPGLTVKHIAATTKGCLELN QKVIWTEEAQKELEENNEKIKNAQGLQ YYNPEEEMLC EVEITKNYEATYVIKQSQ GILWAGKKIMKANKGWSTVKNLMLLL QHVATESITRVGKCPTFKVPFTKEQVM WEMQKGWYYSWLPEIVYTHQVVHDD WRMKLVEEPTSGITIYTDGGKQNGEGIA AYVTSNGRTKQKRLGPVTHQVAERMAI QMALEDTRDKQVNIVTDSYYCWKNITE GLGLEGPQNPWWPIIQNIREKEIVYFAW VPGHKGIYGNQLADEAAKIKEEIMLAYQ GTQIKEKRDEDAGFDLCVPYDIMIPVSD TKIPTDVKIQVPPNSFGWVTGKSSMAK QGLLINGGIIDEGYTGEIQVICTNIGKSNI KLIEGQKFAQLIILQHHSNSRQPWDENKI SQRGDKGFGSTGVFWVENIQEAQDEHE NWH TSPKIL ARNYKIPLTVAKQITQEC P HCTKQGS GPAGCVMRSPNHWQADCTH LDNKIILHFVESNSGYIHATLLSKENALC	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	--	--	--

			TSLAILEWARLFSPKSLHTDNGTNFVAE PVVNLLKFLKIAHTTGIPYHPESQGIVER ANRTLKEKIQSHRDNTQTLEAALQLALI TCNKGRESMGGQTPWEVFITNQAQVIH EKLLLQQAQSSKKFCFYKIPGEHDWKGP TRVLWKGDGAVVVNDEGKGIIAVPLTR TKLLIKPN (SEQ ID NO: 1574)	
--	--	--	--	--

POL_ BIV29	P19560	Вирус иммуноде фицита крупного рогатого скота	MKRRELEKKLRKVRVTPQQDKYYTIGN LQW AIRMINLMGIKCVCDEEC SAAEVA LIITQFSALDLENSPIRGKEEVAIKNTLKV FWSLLAGYKPESTETALGYWEAFTYRE REARADKEGEIKSIYPSLTQNTQNKKQT SNQTNTQSLPAITTQDGTPrFDPDLMKQ LKIWSDATERNGVDLHAVNILGVITANL VQEEIKLLLNSTPKWRLDVQLIESKVRE KENAHRTWKQHHPEAPKTDEIIGKGLSS AEQATLISVECRETFRQWVLQAAMEVA QAKHATPGPINIHQGPKEPYTDFINRLVA ALEGMAAPETTKEYLLQHL SIDHANEDC QSILRPLGPNTPMEKKLEACRVVGSQKS KMQFLVAAMKEMGIQSPIAVLPHTPEA YASQTSGPEDGRRRCYGC GKTGHLKRNC KQQKCYHCGKPGHQARNCRSKNREVLL CPLWAEPTTEQFSPEQHEFCDP ICTPSYI RLDKQPFIKVFIGGRWVKGLVDTGADE VVLKNIHWDRIKGYPGTPIKQIGVNGVN VAKRKTHVEWRFKDKTGIIDVLFSDTPV NLFGRSLLRSIVTCFTLLVHTEKIEPLPVK VRGPGPKVPQWPLTKEYQALKEIVKD LLAEGKISEAAWDNPYNTPVFVIKKKGT GRWRMLMDFRELNKITVKGQEFSTGLP YPPGIKECEHLTAIDIKDAYFTIPLHEDFR PFTAFSVVPVNREGPIERFQWNVLPQGW VCSPA IYQTTTQKIIENIKKSHPDVMLYQ YMDDLIGSNRDDHKQIVQEIRDKLGSY GFKTPDEKVQEERVKWIGFELTPKKWR FQPRQLKIKNPLTVNELQQLVGNCVWV QPEVKIPLYPLTDLLRDKTNLQEKIQLTP EAIKCV EEFNLKLKDPEWKDRIREGAEL VIKIQMVPRGIVFDLLQDGNPIWGGVKG LNYDHSNKIKKILRTMNELNRTVVIMTG REASFL LPGSSEDWEAALQKEESLTQIFP	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661
---------------	--------	--	--	--

			VKFYRHSCRWTSICGPVRENLTYYTDG GKKGKTAAAVYWCEGRTKSKVFPGTN QQAELKAICMALLDGPPKMNIITDSRYA YEGMREEPETWAREGIWLEIAKILPFKQ YVGVGWVPAHKGIGGNTEADEGVKKA LEQMAPCSPPEAILLKPGEEKQNLETGIY MQGLRPQSFLPRADLPVAITGTMVDSEL QLQLLNIGTEHIRIQKDEVFMTCFLENIPS ATEDHERWHTSPDILVRQFHLPKRIAKEI VARCQECKRTTTSVVRGTNPRGRFLWQ MDNTHWNKTIIWVAVETNSGLVEAQVI PEETALQVALCILQLIQRYTVLHLHSDN GPCFTAHRIENLCKYLGITKTTGIPYNPQ SQGVVERAHRDLKDRLAAYQGDCETVE AALSLALVSLNKKRGGIGGHTPYEIYLE SEHTKYQDQLEQQFSKQKIEKWCVVRN RRKEWKGPYKVLWDGDGAAVIEEEGK TALYPHRHMRFIPPPDSIDIQDGSS (SEQ ID NO: 1575)	
--	--	--	--	--

A0A1 42BK H1_A LV	A0A142 BKH1	Вирус лейкоза и саркомы птиц	TVALHLAIPLKWKPDHTPVWIDQWPLPE GKLVALTQLVEKELQLGHIEPSLSCWNT PVFVIRKASGSYRLHDLRAVNAKLVPF GAVQQGAPVLSALPRGWPLMVLDLKD CFFSIPLAEQDREAFATLPSVNNQAPAR RFQWKVLPQGMTCSPTICQLVVGQVLE PLRLKHPSLRMLHYMDDLALLAASSHDG LEAAGEEVISTLERAGFTISPDKIQREPG VQYLGYKLGSTYVAPVGLVAEPRIATL WDVQKLVGSLQWLRPALGIPPRLMGPF YEQLRGSDPNEAREWNLDMKMAWREI VQLSTTAALERWDPALPLEGAVARCEQ GAIGVLGQGLSTHPRPCLWLFSTQPTKA FTAWLEVLTLLITKLRAVVRTFGKEVD VLLLPACFREDLPLPEGILLALRGFAGKI RSSDTPSIFDIARPLHVSLKVRVTDHPVP GPTVFTDASSSTHKGVVVWREGPRWEI KEIADLGASVQQLEARAVAMALLLWPT TPTNVVTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHVRSHSEV PGFFTEGNDVADSQATFQAYPLREAKDL HTALHIGPRALSKACNISMQQAREVVQT CPHCNSAPALEAGVNPRGLGPLQIWQTD FTLEPRMAPRSWLAVTVATASSAIVVTQ HGRVTSVAARHHWATAIAVLGRPKAIK TDNGSCFTSKSTREWLARWGIAHTTGIP GNSQQQAMVERANRLLKDKIRVLAEGD GFMKRIPTGKQGELLAKAMYALNHFER GENTKTPIQKHWRPTVLTEGPPVKIRIET GEWEKGWNVLVWGRGYAAVKNRDTD KIIWVPSRKVKPDITQKDELTKKDEASPL FAGISDWAPWKGEQEG (SEQ ID NO: 1576)	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	<i>RT_Rtv</i>	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 дубликации, 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	<i>RT_ZFREV_подобные</i>	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	<i>RVT_1</i>	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	<i>RT_домен</i>	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 Н 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	<i>Суперсемейство ДНК/РНК-полимераз</i>	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	<i>ДНК/РНК-полимеразы</i>	Данное суперсемейство включает ДНК-полимеразы и РНК-полимеразы.
PF06817	Pfam	<i>RVT_большой палец</i>	Данный домен известен в качестве домена "большой палец". Он состоит из четырехспирального пучка [PMID: 1377403].
IPR010661	InterPro	<i>RVT_большой палец</i>	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 интегразой, 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	<i>RVT_соединительный</i>	Данный домен известен в качестве соединительного домена. Данный

			домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
IPR010659	InterPro	<i>RVT_соединительный</i>	Данный домен известен в качестве соединительного домена. Данный домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
cd03715	CDD	<i>RT_ZFREV_подобные</i>	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]
--	--	--	---

В таблице 41 представлен перечень 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 41 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 41, 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 обоим из предсказанной 5'-UTR и предсказанной 3'-UTR из таблицы 41, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или ее функциональный фрагмент.

Таблица 41. Системы ретротранспозаз, не содержащих LTR, с аннотацией сигнатур доменов обратной транскриптазы.

Семейство	Элемент	Номер доступа	Организм	Предсказанная 5'-UTR	Предсказанная 3'-UTR	Предсказанная аминокислотная последовательность	RT	Старт-кодон	Стоп-кодон
Penelope	Penelope-2_A Di		Acropora digitifera	TTTCAT CTCTC CCACC GACGC AGCAC CACAG TTTCTT	TAATTTA TTCACAA CTTCATG TGACCAT ATCTTCA CCAATGG CAAAGCT	MPNLKTLTKELRQEIVSYLYPLPSISSNLAFLRA IHIFTKTRFSSHINFISRCLQHKVIPKGFRSNFHA SIFSNVSNFNLKYLHEIQRTQNTFSHNIMRSTIR AMCSKRNELNKQIIECRSQLSNVCPSVLLRSIH AKIQALNSKLFHHLEQTKKKHKLENLTNSPKNR TEPLENQNHHTVVTIPENLPLSNAEKS SVLSKGL NFVPISKKSDEFTTRQDVEKFLRRVQLKAFFHN	IPR000 477	371	610

				TAGAA	CCTCTAC	KEDKSDNTEKDAFETLTAKKSKWTPPEGQFAS			
				ACTAG	ACTCTCA	IDYFIKKCRHDVHKLKSNCNTKLSNLSKEEWT			
				AAATT	TATAAAC	ALINLKNRKDLVIKAADKGGATVWVRTDLYQ			
				CATTT	CACAACA	QEAIRQLSDPTFYTKVNKDLTPANQKIVKDTIQ			
				ACTAC	ACCCACA	ELIRKQELPVTAQNLIITTPRTSCIYFKPKIHKPN			
				TTAAC	ATTCCTC	NPGRPIVSACSCPTELISSYLDKVMTPIVKSLPS			
				CGCTG	TATTCGC	YIKDSNHALETFGNFNFSGENKIIFTMDITSLYT			
				CGACC	TCTGACG	VIPNNEGLQALKYFFNQRPIKKPSSETLLRLAE			
				TCATT	AAGGGCT	LVLTLNCFSGDNYYKQINGVAMGTKMGPSY			
				CATTC	AACGCTC	ANLFVGFIENTKFFSNYHGPKPDLYKRYIDDCV			
				CTCGC	GAAACGT	GATSSTKEELNLFINSVNSFHPALKYTWEISEN			
				TACCG	CAGCTTT	SLAFLDIKLSINGNGLSTSVHYKPTDSHNYLLH			
				CGACG	CTAAATC	SSSHPQHVKNAIPFSQFLRLRRLCSDDTDFNNK			
				CTGAA	TTTCACG	CEEMCQFFKKRGYPDSAVTTGKHRAQEIDRET			
				ACTCA	GTGGTAA	ALQTSQNEETDRIPFTLTYPQNLAIKNVILKN			
				CG	TTCAACC	FKILSNDPETKHIFSLPPLISFKRDKNLGNFLVR			
					TTTATCA	SAFKFNNKPGTFTCKRTRCKTCPFISNTVNISGP			
					ACTCGTT	NRSVKVTDHFTCISSNVIYCITCTLCKKIYIGET			
					TGATAAA	GRRLADRFREHLRDAEQNNTDASKPVARHFN			
					ACCAAAT	LPNHSHHNMTICGLYLHHGNTESRKNLEQKLI			
					TTTTGT	FQLGTLSPHGINERLSFH			
Pene lope	Pene lope-		Acropora digitifera	GCCTC TTCGC	TAATATT GTTACCT	MRLAEYFFDENSNNRTANEHDTPFHINKSTWN PPNDRERALNTFLDAVKLDITTTKPKPTQDNLT	IPR000 477	143	383

1_A Di		TTGAA	CTTATCA	VTERQAIGQLKQRQDIIKPADKSGSTVVMDK			
		TCTCT	TGTTATTT	TWYIDECNRQLTDTKFYQRLDEDITADIQKRV			
		GCTGG	CATTATT	TFYVNRMHKDKLINDKTKQYLIQSDVKPGRFY			
		CCCAC	TTTTCAC	ILPKVHKPGNPGRPIVSSNSHPTERLSHFVDYH			
		AGACG	CTTATAC	LQPLVHKLPSFVKDTNDFLNKLLTIGNLPANSL			
		TGCAA	TTATCTTT	LVTLDVSSLYTNIPHNEGINACERFLRTSSHKTI			
		AAAAA	AGTTTAC	PTSTLCDLIRMILTMNNSFNDNHYLQIHGTA			
		GCCGC	ACTGTCA	MGTKMAPSYANLFLGFFEANALKNAPFQPHT			
		TCGTT	ATTATTTT	WLRYIDDIFMIWTEGLDNLKIFIDYLNNIHSTIK			
		CCCGC	TACTTAT	FTSSHSTNIPFLDVNVSLTNDGNISTDLYTKPT			
		GGTAT	TTCCGGT	DKHQHLLYSSCHPLHTKKAIPFSLALRLRRICS			
		TATTA	TATTACA	TDATFHTRTAQLATYLLKRGYNRNRFVNKQIRR			
		ACCAC	AAATATG	AADIPRQLTLQTKDINKPKRIPFITTFNPSLPHIS			
		AATCT	CAACATA	HIKKHFNLLLSSNRCKSVFQHPPVVAFRSPN			
		CACCG	AAAATCA	LRDLLVTAKLSPNSTNPQLPCGSFRCGKNCAT			
		NCGAT	CGTACCA	CPYISHGLTTYTFFSTGETRPIKFNFCTETKNLI			
		AACCT	ATGTAAA	YMIQCNRCNLQYIGETKRRLKDRFNEHRRTID			
		CAATC	ATTCTCA	NPNNKSKPTTAAEHFLSSPNHTANDMILPIEKI			
		AACAC	TTGTACC	FSNRDSIRKAREAFLIQKGRITIDPDGLNIREETY			
		ATCAA	TGAAGAA				
		TAACA	GGCTGGT				
		CTAAT	TTGGCCA				
		GCCAG	GCCGAAA				

				TACTA	TATAGTA				
				ATGAC	CACCTAC				
				ATACT	AAAAAAG				
				ATCTA	ATCCTTC				
				ATGTT	TACGTTG				
				ACACA	TGTCGAC				
				TAACG	TTCTGCTT				
				CCTCT					
				ACCTC					
				AACTG					
				TCACA					
				CGCAA					
				ACGCT					
				CTCGC					
				CGCGA					
				AACCA					
				AACAT					
				GCCAA					
				CATTC					
				ATTTG					
				GATCA					
				TTCTTC					
				TGTTA					

				TCAAC					
				CTATC					
				CAGTT					
				GCTCT					
				TTATCT					
				ACCGA					
				CGAGA					
				TATCT					
				ATCCT					
				TTCAC					
				GTGGA					
				CTCAC					
				TTTCTG					
				CCCTA					
				CTCCA					
				CGACA					
				TATCA					
				ACTGG					
				CCGGA					
				AGTTT					
				CAGCC					
				GACAT					
				TTACG					

				ACTTTT CTCGG CGC					
Pene lope	Pene lope- 4_A Di		Acropora digitifera	TTTAT GAT	TGATTTA CTCTCAC AGTACTG CCCAACG TATTCTA CGATACA CGTACTG ACCTTAG ACCTGTA GACGGAT CGAAACG CACCAAT CACTGTT TAGTCTT TCCAGCC AATTACA TCTAGGC TCAACTG ACAGTCG ACCAATA	MTPGFKPFTILRSLGRQSERLANNYISYASNVR RIRQQLWFNHRCKDLRLVPSGLRLKSPLNTQE AIQIVKATCRRLRARINDCHRRLNHYNRLQ QRLDKLKQLIPTNLLDVIRDIADQRAKKTTEQ YRIKTELKLTRLQRTKNKKRHKTDENWVRNIS SRPLDKTETQVLSYGLKHSVTPKRIPTETIVSSV EAVLSRQRDLSEPTKDNIRSRIASTIQSASLTDN NLTKDERQALKRLRNDNDILILPADKGRVTVV MDKTDYHDKMDEL VNDKQTYEVLKRDPTPA LQRKLNSKLLQLKKADAIDIRRYNRLRCPVPQ PAKLYGLPKLHKPNVPMRPVVSFCGSPTYELSK YLTTILKPLTDESRLKLQSTETFIDAIKTAQVPD DHKLVSFDVKSLFTSIPLQLALDCTETAINNST LQLPLL TNDLMDLLNLCLTSTYFQYNGQHYK QLHGTAMGSPVSVVVAEIVMQNIEEQALASY KRTIPLWLRVDDTFTTLHKDEIDDFHEHLNR QNAHIQFTKEIEDNGKIPFLDCLVIRDNNRLQT TVYRKPTHDRLLDESSYNPSSHKATTIRTLTR RALLVCNSHDSLADERKYLDNVFSKNNYNCD FVTRNTYRTEPNETNTNLTPTTTTVTIPYIKGTSE	cd0030 4	459	541

				ACATCAC	IIARILQPYNIRVAHRPITTLRKLLTNVKDKDQP			
				GAATAAG	RDRQGAVYKIKCCDCQATYIGETGRNLNTRLT			
				TTGACCA	EHRRATRNGDINNIAEHLQTNHRIDWDSAT			
				ATTACAT	CVTYNTNYYQRIVLESWFTNLEQTPINRCLQLP			
				CTACGAC	APYKRLIDDINRQTTH			
				CGGAGTT				
				TTTATAG				
				TATCTAC				
				TGACGTT				
				ACACAAA				
				TCACTTG				
				ACTCTGA				
				AGATGAC				
				TTCCGCT				
				CAGGTTG				
				TCGAAAC				
				GTCAGTC				
				AATGTCA				
				TCTCAAA				
				CAGTCCT				
				TCTCAGG				
				ACTACAC				
				TCACCCG				

					GACGATC GTACTTT AC				
I	I_El e14		Aedes aegypti	CAGTT GACAG CCGAC CGCCG TCGCG TACAG TCGCC TTTCG CATTG TTCCG TACTG TTAAT ATCAA TTAAC AGCGC ATCAG CCAAC TCGAT CGTTG TCTGA	TGATCCT GCCCAAT ATGACGA CGACATT GATGCCA GCCCACG GATACGA AACTGTT TTACTAA ATTGTAC TATGTAT AGATGTA AGCTTGA AATGTTA GATTGAC AGTTATC TCCTTTTC ATCGGTT CAGTCTT CTGCTGA	MPGPALGGAPLESATRTLCPNGFSQTYFSTA APSPLPVVGSTPGPTLGGALSGSVSSTFLGSTN SKETSLSSFPVPALPTSHNFILQWNLNGYRSRL CDLELIQNNQPWILALQETNNISADEMGRTL GQYAWTVKRLANQRHSA AIGILKGIPHKILDL PSDLPIVGIQLLGSPSVSVACAYLPGGNIPNLHG QIQNCLQALPEPRIFLGDLNSHHPVWGSPKAD KRGSTLINAFEEEDMLILNDGSPTFFNGHFSSEI DVTAVSLSFARNIQWSVSSDLHGSDHFPIRIAL TTGASIPKNTRRPRWKYDKADWAEYDLLIRSA LRNSPPDNMADFISTIVESAKATIPRTSSTPGRK ALRWWSDETKTAVKARRKALRQLRRLPLDHP DRQSTLENYRKLHLECRNIIGDAKKATWESFL ESMNASQTTSELWNRVNALNGKRKTTPLTLQ VQGNAISDPEAVASELGKYFASLSAIESYETHF XNSVKPSTSAINNFIIPPNTADPEMEQPFSAREL SFALSRSNGRSAGPDEVGYPLLKNLPSVGKLA LLDLYNKIWANNDFFPAWKESLVVPIPKANIPT RDPTKFRPISLTCCAGKVMERMVNRRLRYHLE ANGLLDFRQHAFRQGYGTSTYFAGLSDVLKE	cd0165 0	575	833

			CGGGG	CCGTATT	AYDKGNHAEVISLDISKAFNRTWTPLVLEQLA			
			TGCTC	TCTCCAG	AWGIGGRTLHFVRNFLADRTFQVVIGNTKSST			
			CACAG	GGTCTCA	FPEETGVPQGSVLAVTLFLIAMNGVFSRLPKNI			
			CGGGG	GCCTTCT	YVFVYADDIVIVVCGSTPTMTRIRAQTAVKSV			
			CGAGT	GCTGAGC	AKWASDNGFQLSASKSIRCHICPSGHRITGPD			
			GTGCA	CTTCTCC	TIDGQPIPLRKTVRILGVIVDRALSFRQHFDTVK			
			TTGTTT	CGTGTTG	SACRSRLNLIKSI SRPHRSNNREIRFRVARAIVD			
			TATTC	AACTAGC	SRLVYGLELTSIAMDRLVEVLSPIYNSYIRIISG			
			CGGAT	ATAATGT	LLPSTPSDSACVEAGLLPFRIFIRSSICCKTAAFL			
			CAGTG	TAAAAAA	SKTAGEDRVFLLDEGNRALSTAANLSLPPVTR			
			CAGCA	CACGTTA	VHWLGDRSWRSVPPKIDNKKIKNSFSAGSNSAA			
			CCCAG	ATAAAGA	LRRSVAELLQTSYSDFALRYSDGSLTSAGVGIG			
			TGGTG	TGAAAAA	VAGDIPDVMSLPSQCSVFSAEAAAAFIAATTP			
			TGGTG	AAAAA	SDRSILVLTDSASVISALQSDSPHPWIQAILKY			
			AGAAG		ALPDTVFTWIPGHCGVPGNEMADHLARSGLS			
			TGAGG		GQRYTSEVPFMDLKRWIKSTFRQHWEDSWYR			
			TTAAA		TRTLFLRKIKNSTTTWSDLPILKDQRILSRLRTG			
			TAGAC		HTLISHNMGGGPFHKECESCHIPASVDHVLCA			
			TCGGT		CPMYEHLRQIHGLSDNIGEVLRDDATTIAALLS			
			AGCGT		FLRDANLYSSI			
			GGTTC					
			GAGTA					
			CTGTA					

				TAKTA					
				TTATTT					
				CGGGC					
				GCGAG					
				TGTAT					
				GTGTT					
				AGTAA					
				CTTGT					
				GAGTG					
				GTATA					
				AATAA					
				AATTA					
				TCTTC					
				AAATT					
				AAAAA					
				GCGGT					
				TCAGT					
				GGTTT					
				TCAAC					
				TGTAG					
				TGTGA					
				CTGAC					
				CAACC					

				AAAAC					
				AGGTG					
				AGTCT					
				TCCAC					
				GAGTG					
				CCTCC					
				TTCCG					
				GCCCG					
				CCGGC					
				GCCAT					
				GCTGG					
				CGAAC					
				GGCCA					
				TCCCC					
				TCCAC					
				CCAGG					
				GGACC					
				CAGGG					
				GGCCC					
				CGGAG					
				GAGGA					
				TTTGG					
				CAAAG					

				TAGGC					
				AACAG					
				AGTTC					
				TCGGC					
				GACTA					
				CACCG					
				GTGCA					
				TTGCT					
				ACCGG					
				AGTTT					
				ATGGA					
				CCGTG					
				AAGGT					
				ACTGC					
				AGGAC					
				AACTG					
				CAGTA					
				TCTAA					
				GAATG					
				GAAGC					
				GACTG					
				CTGGA					
				ACCAT					

				GCCCC					
				AGGAT					
				CCGTT					
				TCTTCT					
				TCGGC					
				TATCC					
				GTCGA					
				AAAGG					
				CCATT					
				GGAGG					
				TCCCA					
				TTTTCG					
				GGGCC					
				TTCAA					
				GGAAA					
				ATAAA					
				GGCAT					
				GTCCT					
				ACGTG					
				TTGAA					
				GGTGC					
				GCAGT					
				CAAGA					

				CCAGG					
				TMAAT					
				AAGCT					
				CCTCC					
				GCATG					
				AGCCA					
				ACTGC					
				AAGAT					
				AAMAC					
				CCCAA					
				TCCGG					
				ATCGC					
				TGAGC					
				ACCCG					
				CGGCT					
				MAATC					
				AAAGG					
				GAATG					
				TGTGG					
				TGACC					
				AACCA					
				TGACA					
				CAATC					

				GGACT					
				ATCGG					
				ATGAC					
				TACCT					
				GCTTC					
				AACAG					
				CTCTC					
				CTCCC					
				AGGGA					
				ATCAA					
				GGGMS					
				TGCGA					
				AGGAT					
				CACGC					
				GACGC					
				GGACC					
				GGCAT					
				GGCCA					
				AAGCC					
				GTGGT					
				AAACA					
				CCGCC					
				ACCAT					

				CGTAC					
				TCACC					
				ATCAA					
				GGGCA					
				CCGTT					
				ATTCC					
				GGAGC					
				ATGTC					
				GTTTTC					
				GGTTG					
				GTCAC					
				GGTGC					
				AAGAC					
				GACGT					
				CTTTTC					
				TACCC					
				ATCTC					
				CCATG					
				CTATG					
				CTTCC					
				GCTGC					
				TGGGA					
				GTATG					

				GGCAT					
				ACCGG					
				TAAGC					
				GTTGC					
				ACAAG					
				CCCCC					
				AGCAG					
				CATCT					
				GCGGA					
				CGCTG					
				CTGCG					
				AGGTA					
				CACCC					
				GGAAG					
				AACAA					
				ACCCC					
				CGATC					
				CTACA					
				AACAC					
				CCAGG					
				AAGGA					
				GCCGC					
				CAGTG					

				CCCAC					
				GTTGC					
				GTCGA					
				AAACG					
				CGGTA					
				CCAGT					
				GCACC					
				GCAGA					
				GCCGT					
				TGCTG					
				CAAGA					
				CCAGT					
				GCAAA					
				ACAGG					
				AGACA					
				GCACT					
				CAGTC					
				TCAAG					
				CCGCA					
				AGTGT					
				CCGGT					
				GTACW					
				WGAAG					

				GAACA					
				GGCCA					
				TACAA					
				CACCT					
				TCGCG					
				TCGAC					
				GTGGA					
				CATTC					
				CGTAT					
				CCGCA					
				GGCTC					
				GACGT					
				GAATA					
				CGAAG					
				CCCAA					
				CAAGC					
				GAGCA					
				AGTCG					
				AAGTG					
				GTGCC					
				AACTT					
				CTGCC					
				GGTGT					

				GGTCA					
				GTGCC					
				AGCAA					
				GGACG					
				CTGAA					
				ATCGA					
				CAGTC					
				TGAAG					
				GCCAT					
				GGTGG					
				CTGCA					
				AAACT					
				CCGCG					
				GACGG					
				AGAAG					
				AAAAG					
				AGGAT					
				GGCTG					
				AAATG					
				GAACG					
				GGCTC					
				TGCAG					
				AGCCG					

				CACCG					
				TCAGC					
				GACCG					
				CCTCG					
				AGGCG					
				GTTAA					
				AGAAC					
				ACGGC					
				ACCAT					
				CGGAG					
				AGCTG					
				GTCAA					
				GCAAG					
				TGGCA					
				ACCCT					
				AACTG					
				AAACG					
				GTTCA					
				GAACC					
				TCCAA					
				AATGC					
				TCTAG					
				CAGTA					

				AGGAC					
				AAACA					
				AATCC					
				CGGAT					
				CAAAG					
				TCGCA					
				CCATC					
				GCTGA					
				GCTTC					
				TAGGC					
				AGGCA					
				ATCAG					
				CTGCC					
				GTCCA					
				ACGAC					
				AAACC					
				CGCGA					
				CAAGC					
				TTTGG					
				CTTAC					
				CGCGC					
				AAGAS					
				GATCA					

				AACGC					
				AAACC					
				GCCCC					
				AAGTC					
				GCCAG					
				AGGTT					
				CCCGT					
				CATGA					
				AACCG					
				CCCAG					
				TCTGA					
				ACCCC					
				AAAGT					
				GGACG					
				TACAG					
				GTGTC					
				CAAAT					
				GGATT					
				TCAAA					
				TATAA					
				CACCG					
				GACCT					
				CAAGA					

				AACTC					
				TCACA					
				GGCGA					
				TGAAA					
				CCAAC					
				GATGG					
				AAATG					
				CGTCT					
				GATCA					
				TAGCA					
				TGATG					
				TCAGC					
				CACAT					
				CCGAA					
				GCCAC					
				TGGAA					
				TCTCC					
				AACAC					
				GTCTC					
				ACGTT					
				TCGAA					
				CCCTA					
				CGAAA					

				AGAAT					
				GCACT					
				CAACC					
				TCGGA					
				TCACA					
				GTTGC					
				CCATC					
				CGGCG					
				ATTCA					
				AATGC					
				CTCAT					
				CCACC					
				AACTC					
				GAAGC					
				GGAAA					
				AACAA					
				GCACC					
				GCAAA					
				AGAGG					
				TGGTA					
				GAACC					
				AAACA					
				GTAAA					

				ACAGT					
				AATTT					
				ACGCT					
				CAGTC					
				CTCTTC					
				AAGCA					
				CATGG					
				TCTAA					
				ACAAT					
				TCGAC					
				AATAT					
				TTTCTT					
				CCCGT					
				CATCT					
				TCACT					
				CCGCA					
				AATTA					
				GCGTT					
				GAGAC					
				AGTCA					
				AATCA					
				CAAGA					
				GTCGA					

				CTGAG					
				CAAAT					
				ACCGA					
				TGAAC					
				ACACT					
				ACAGC					
				CTCCG					
				AACGA					
				AGTAC					
				TCACG					
				GACAA					
				CTCAG					
				CGACA					
				ACAGG					
				GGTCA					
				CGGAC					
				AGCCG					
				GGGCC					
				CCGTC					
				AGTGC					
				GGACG					
				CTACA					
				CCCCA					

				ACCGG					
				AACTG					
				GCGGA					
				CAACC					
				CCCGG					
				CATCC					
				GGGCG					
				TAACT					
				GTTGT					
				TATCG					
				GGACG					
				CAAAA					
				GGGAC					
				ACAGT					
				GCCTA					
				TCCCC					
				CGAAA					
				GTCCC					
				AACTA					
				GCTGT					
				ATGGG					
				TACGC					
				CAGGG					

				CCCCG					
				TCGGT					
				GCGGA					
				AGTTT					
				TGGAA					
				GAACC					
				GGAAC					
				CGGCG					
				GACAA					
				CCTCT					
				GGCGC					
				CCATA					
				TACGG					
				TCTGG					
				AAGAG					
				GGACG					
				CAAAA					
				GGGAT					
				CAACA					
				TCCTA					
				TTCCC					
				CTKAG					
				GACAG					

				CGTGG					
				GAAGC					
				AAGCC					
				ACAGG					
				AAACC					
				TTTTTT					
				CTGCC					
				CCTGC					
				CAACA					
				CACAG					
				AGGGT					
				GAAGG					
				TCGGG					
				ACGCC					
				CAATA					
				CCCCG					
				AAGGA					
				CAACG					
				AGGGA					
				AGCAA					
				GCCTC					
				CGTCG					
				AGTCG					

				AAACC					
				CCCCC					
				GGAAS					
				AGAAA					
				ACCGC					
				CGGAC					
				CACTA					
				CACCT					
				CCAAC					
				TGGCA					
				CAACC					
				ACCCC					
				GGTCA					
				GCGAC					
				AAGCT					
				GAATC					
				CCCCG					
				ACTTC					
				CCGGG					
				CAGCC					
				TTGGG					
				ACCCG					
				GAAGC					

				AAGCG					
				ACGCC					
				AGGTT					
				GTATG					
				ACCTG					
				TTACA					
				CAACG					
				CTCGT					
				TACAA					
				TCGCT					
				CAGCC					
				CGCCG					
				ACTTC					
				AGCCC					
				GGGTA					
				TTCCG					
				GAACC					
				GTTTC					
				CAGAC					
				TTCAA					
				CTCCA					
				AACGC					
				GGAAG					

				CATAC					
				GAAAA					
				CCGAT					
				CGCCA					
				TCCGC					
				GATAA					
				TKATA					
				GCTCA					
				ACCCG					
				CTGCT					
				GGCCC					
				CTCAA					
				GCCAA					
				ACAGC					
				CTCCC					
				CAGAT					
				CCCTA					
				TCAGG					
				ATCGC					
				CCTGG					
				GCACC					
				TCGGA					
				CTCAA					

				CAATC					
				CGAGC					
				GGTAA					
				GGCCG					
				ACCCT					
				TTCCC					
				GCTGG					
				GCCTC					
				CACCC					
				GGTGA					
				ATCTT					
				CGTCC					
				GCCAC					
				AACAA					
				CCTAT					
				GTCAA					
				CCCAT					
				TTTTCG					
				ACGCG					
				GCTCC					
				CAGTC					
				TTCTTC					
				GCTCT					

				CAAAC TCACA GCAGC ACTCC TTGAT AAGGT TTCCT AGCAG CCCTG TTGGA AACGA CTGTC TTGCC GCAGC TCCCG AAGTT CCGGC AGTAG TAGGT CCC					
RTE	RTE _Ele 3	.	Aedes aegypti	GGGCT GCAAA ATGGT	TAAGTAA GTAAGTA AGTA	MKLEILGLSEVRWPNFGEHRIPSGQILLYSGLR GEHAPRHRGVGFLLSAQAYAALMKWEPINERI IVARFRTRVRNLTIVQCYAPTDAAELQDKENF	cd0165 0	503	762

			GAGTC	YSQLNAVVDKIPKGDIKIYMGDFNAKIGSDNS		
			GACGG	SYEPVMGRHGLGEMSENGELLAEF CGNNDMV		
			AGTCC	IGGSLFPHRPVHKVTWVSRDGFTENQIDHICIS		
			TCACA	RKWKRSLLDVRNKR SADIASDHLLIGE VRLR		
			AGTTC	IARIRRQEERVGRRFNTRRLEDATVKRSFVEEL		
			CTACC	ETXAANISEGGSVEEQWTAIKNAFITTSENN LG		
			TCATG	ELRTRRKQWITDETWRKIEERREAKAAIERAK		
			CTTCC	TRGAKALSQRYSALEREVKRSCRRDKRAWA		
			ACGGG	DSLADEGEKA AATGDIRLLYDISRRLSGAKMN		
			TCGAC	TTMPVKDTTGQLLTDPTDQLKRWFEHFEQLFQ		
			GGATG	VSTRPPTPQHDP PRVRRISRVNTEAPSLHEIETA		
			ACAAA	IQSMKSNKAPGVDRISA EMLKADPAISAQLLH		
			TGACC	QLFRNIWDTATFPADWMQGV LVKVPKKGDLT		
			GCCAG	VCDNWRGIMLLCIVLKVLC KVILNRIQE KIDAT		
			CTAAG	LRRQQAGFRAGRSCVDHIVTLRIILEQVNEFQE		
			AGTTG	SLYMFIDYEKA FDRLNHENMWGALRRKGVP		
			CGTGC	EKIIGLIEAQYEAFTCRVLHNGVLSEPIRVVAG		
			TTAGC	VRQGCILSPLLFLIVIDEILIGAIDREQNRGLLW		
			TGGTA	QPITMEHLNDL DLADDIAILAQR RS DMQSKLD		
			GTGCA	DLAERSSAAGLNINVNKTKSLDVNTAGPSSFT		
			GCCTG	VAGQTVENVENFQYLGSQ LASDGGTKIDIGAR		
			GGTCC	IKKARAAFA SLRNIWKNSQISQRTKIRIFNSNV		
			TGTAG	KSVLLYASETWCVSAENTQRLQVF INRCLRYII		

				TCCTTC	RAWWPHNWISNADLHRRCHQQPIQTEIRQRK			
				TGACT	WGWVGHTLRRGGNEICKQALEWNPAGHRSR			
				TCAGC	GRPRGSWRRSLNKEIKEVDGNLTWPQVKAIA			
				TAGAT	GNRPGWKSFTSALCTRMGAQDP			
				TGAGG				
				AGGTA				
				CGTCC				
				ACCGA				
				GCGTC				
				TGTTC				
				ACCAA				
				GGAGG				
				TGCGG				
				CTCAA				
				ACAGC				
				GTCTG				
				TCTGG				
				CATCC				
				AGCGG				
				CTGAA				
				TAAGA				
				AATGC				
				TCTTCC				

				CCGGA					
				AGCTA					
				CACCT					
				AAGAT					
				GGCAG					
				CCCCA					
				TCACG					
				GCGGA					
				TAGGG					
				AACCT					
				TAGGC					
				TAACA					
				ACCTA					
				CTGTC					
				CCCGA					
				AACAT					
				CAAAA					
				AATGT					
				TCAAG					
				AAACC					
				GATAG					
				AGAAA					
				GATTA					

				CGAAC					
				TGATT					
				TAACG					
				GCAAC					
				GACTT					
				TTAGC					
				GCGAA					
				ACAAC					
				GGACA					
				CGAAT					
				TGGAA					
				CATGG					
				AACGT					
				ACTAA					
				CCCTA					
				GCCCA					
				GCAGG					
				GTAAA					
				CTGGC					
				ACAAC					
				TCGCC					
				AACGA					

				GGCGT GTCGC					
Pene lope	Pene lope _Ele 2		Aedes aegypti	AGTTA GGGAT TTGTA TAATT ATTAT TATCC CCCTTT TCAAT TTTTG ACCAA GACAT GACAG GTATA CTCCT GTCGA AATTT TTTCCT AGGTA GTTTG ATTGT GTAGG	TGAATAC ATAGGTA GATAGTT AGGGATT TGTATAA TTATTATT ATCCCCC TTTTCAA TTTTTGA CCAAGAC ATGACAG GTATACT CCTGTCG AAAATTT TTTCCTA GGTAGTT TGATTGT GTAGGTA CTTATCT GGTGTTA GTATTTA	MASTLSESNTFIKLLKSSIAGLKKDIAFLSECRKL NLTPKSHKIKVRGQVPQDIKKMESEYLRSSIK KLHSKLNSQTLKCYSLHLKLAKQYPWEFPLFL TKVIQAECEARRKKDIHKKLAILKNKQHNK VVTALPKVHSIEGFVVNRSTQQFSEEQLTLLNK GLAYAVATKPDIEQTIIDVETAITHTIPIEFQESA RNDTEESLKGNGFQLRGQNSDEDKVLSELRDK SVYYVKADKGNVAVVILDKDDYDNQMNEKITT GPYRHFRVDPLPTMVRHVEKTLKECKPLLGD NTSRLKVSNPVLPRIKGLPKIHKPGNEMREIISA EGSPTHKLAKWLVEFQSMPKPFESRSVNNTQ EFASQLQASGYIQDDEIMVSFDVKALFPSVPVK EAINLLEEWLLNQHNESNWKNKVRYYIKLTRL CMEENYFSFRGNFYKQTKGAPMGNPLSPFLSE LFMANFESILKKKDLLPQRWWRYVDDVFSIIK KDYLPKILEAINSIHKDIKFTHEEEKDNKLPFLD LLITRESSHFEFEIYRKPTNTMRVIPSTSNHTYQ HKMAAFHHMIHRMQTLPLSETGKSKELDYIFE TARLNGYRNSTIQAVVDKRARDKYRKSM TTL TAEKEDLKRVAVNFDTKITQPLAKKFRKYGV DLVFSSRSSQLKSRLGSTKDKINKLNRAGVYKI	cd0030 4	445	525

			TACTT	AGGTAGA	SCPHCNKIYIGQTKRTLEIRFKEHIAEVKKADK			
			ATCTG	GAGATGC	ELEKGLAYDFKSKVAEHVFQDGHQMTSDNIDI			
			GTGTT	GACGATT	VRSVSSSWKLDVAESLEIYKQKPSLLLNRDQG			
			AGTAT	TCTTCAG	NGQSRLFKFVPPKKYKRA			
			TTAAG	TGGATAC				
			GTAGA	GAGTAAC				
			GAGAT	TGACACT				
			GCGAC	GAAGAAG				
			GATTT	ACTGCAA				
			CTTCA	GTGGTAG				
			GTGGA	TCGAAAT				
			TACGA	ACGCGTA				
			GTAAC	TCTGTCA				
			TGACA	AAGATAA				
			CTGAA	GCATTTA				
			GAAGA	GGAGCGG				
			CTGCA	AATTAAA				
			AGTGG	AGGTACA				
			TAGTC	CGGTACT				
			GAAAT	CCATCTA				
			ACGCG	CTTTAAA				
			TATCT	GTTTCTCT				
			GTCAA	ATTCGAG				

				AGATA	ATTCTGC				
				AGCAT	TCAGAGG				
				TTAGG	ATTCGAA				
				AGCGG	CATTCAT				
				AATTA	TAAAGAG				
				AAAGG					
				TACAC					
				GGTAC					
				TCCAT					
				CTACT					
				TTAAA					
				GTTTCT					
				CTATTT					
				GAGAT					
				TCTGC					
				TCAGA					
				GGATT					
				CGAAC					
				ATTCA					
				TTAAA					
				GAGTT					
				GTA					

CR1	CR1 - 61_ AAe	Aedes aegypti	TT	TAATGTA TCCTGTT AAGTGTT GTCCAAG AGTTGAT GCTTAGT TTAGTTT AGTTTTA ATTTTATT CATTAAG ACCATGA TGTCGGA TGAATCT TCAATAC AAATACA AATACAA ATACAAA TAC	LDNFAFRGLKQRNLLRNHLDRTLIDLIFTNDAIT DCSSVDEACQPIVDIDPHHPAFVLTVRSPITVPF VEVSDDEVRDFGKVDYETLHVLLSRVDWSTV YQCNDVDNAVSIENGKLHECFTHCVPLAMPR RKPAWSNARLLYLRRNQSKMQRRFSKRKNDC NKLAFNRASRSYRYNRYLYGKYVRRVQDNL RRHPKQFWSFVNSKRNETGLPSFSLNGHSAS SAEGKCELFASWFASVFDTSQSQADPDAASQ SLPTDIIDVDIFEVSHMMVEEALKKLKPSISPGP DGIPACILKHCKAVLIAPLKAIFNLSLRCQQFPK AWKMSYMRPIFKKGSRTVEVSNYRGITSLPAGS KCFEIMNDVLLQACKTYISQNHGFFPQRSVI SNLCEYTSFCIKEMDKGGQVDSIYTDIKA AFDT VDHSILLAKLNRLGASSRICSWLYSYLTDRQLL VKIGKAISAPFTPHCGVPQGSNLGPLLFVLFFN DVSHVLSNCYLLFYADDLKIYLIVRSINDCRQL QEQLDLFVSWCTLNHLSSISVSKCCIISFRRMKS PVLNYAINGTILNRVDNVKDLGVLLDHQLTF KMHYSETVDKANRQLGFIFKIASEFDDPLCLKS LYCSLVRSILEFASVIWSPYDAVWIARIESIQRR FVRCALRSLPWSDPLRLPPYEHRCALLGIETLE ERRKMNQAVFGAKVLRSEIDSPALLRQLNIYA	cd0165 0	332	579
-----	------------------------	------------------	----	---	---	-------------	-----	-----

						PQRVLRSRQLLYQPPRRRTMYGSNNPIDTISRRF QDNYNLFDNFISTNTFKNRLRNRQV			
CR1	CR1 - 81_ AAe	.	Aedes aegypti	T	TAGTTTG GGCCCCG TATCACT CTGTACA CATCAAT CGACTTG AACGAGT TCAAAAA AGTTTTA TTAGGTA TGCACTC CAGAGAC TACCTTG GCGAAAT AGGATGC TGCTGCC GCCATAC GAGCATC GCTGTGC ACTCGTC GACTTAC	SQHQRGGGVLIADVKKALQSNVILINDANQLEQ ISVRITYGTMSLVVCTVYLPNTEPALYQQHA ASVNQLCKLLGTQDEIVVLGDYNLPYLRWRF DDDIGAFLPINASSEQEISLVEDIVATGLMQMN HLLNSNERLLDLAFVSNASVCELLEPPLPLMK VDRHHTPFVLTLDVIQTPLENRKNETFFDFRQC NYDELNAEIAQINWRDILAFGSLDDAVTRFYS ELNVIFERHVPRRSRACRPDNKQPWWNAQLR NLRNRLRKARTRYFRVRNRLNRDAVRDLESQ FNELNVSCFQSYIDRLESNMKDDPKQFWSYLR SRTTSRGFPQLINYDGKTATTPDDAVKLFSSFF QSVLSITPPPSSASYLSALKTFDLNLQPTSFTQH EVYKKLCSIDGSKGPGSDGLQPSFIKACSASLA EPAALLFNRS LAESVFPAKWKEALITPIHKAGN VHDVKNYRGISILSCLPKVFESMLLDFLYPAVR HIITEDQHGFVRKRSTTTNLMSYVSTLIDKIEK RQQIDAVYVDFTKAFDRVPHQLAVEKFQRMG FPDWLTKWVLAYLINRSASVRLGTTLSDPFEIT SGVPQGSHLGPLLFILFVNDLCSELKSSKVMYA DDLKIFRVINSPVDCCALQMDIDKILDWCSRN GMDVNIACNVITFTRKRSPTVFEYSMHGATI	cd0165 0	441	690

					CTACACT	ERVRLVKDLGISLDCKLRFVEQISSVIAKAYAV			
					ACGCAAC	LGIIRRNTKDFRDVYCLKALYIYHLFAAFWNM			
					AGGAGAA	EL			
					CTCTTCT				
					AAAGCAG				
					CTTTTCAT				
					TTTGTGAT				
					CTGCTCG				
					AAAATAA				
					CATCGAC				
					TGCTCTG				
					GTCTGTT				
					AGGCAAA				
					TTAAGTT				
					TCAACAC				
					CCCTGTG				
					AGGCTCC				
					TACGGAG				
					GACTGCA				
					CTATTCC				
					GACACGC				
					TGCTCAT				
					CGAACGT				

					TATATGG				
					GCAAAAT				
					AATCCGC				
					TTGATGT				
					ATGTTGT				
					GAAATGT				
					TTAATAA				
					TGTGTAT				
					CACTTGT				
					ACGATTT				
					TAACATA				
					GCTAAGG				
					AAACATT				
					TAGATTA				
					AGGATTC				
					TACGGAA				
					TTGAAGT				
					CTGTGCG				
					ATGTATT				
					TTTATTTA				
					AATCGAA				
					GACCAAA				

					TAAATCA AATCA				
Crack	Crack-13_Ae		Aedes aegypti	CA	TAATATG CATTTTTT CTAATGC CACTTCC GTATCTG TTCTTTGT CATTTCA CGCCACC ACCACCC TTCGCCA CCACCCA CCGCCAG CCACCAA TCACCAC CACCCAC TGCTAGC CACCCGC CGCCAGC CACCCGC CGCCATC GCCACCA	HGGLAVFVRNELQMDVCRNIVIEGFHHIHC RF KFEANAINLHAVYRPPSFDVRRFLHEIEGMISL SKNNEKCVLVGDLNIPLNITTNNIVGEYIRLLE TYNMLPTNTVVSRQSSNNILDHVVCSTSMIED VFNETIFTDVSDHCLICSSINMKCHTSESTLHKR IIDHSRLNELFERSILNIPQNLSVNERLNYVITH YNEILNDCSRLVTQRVKLKGSCPWMTLDLWK LIKIKDKLLKKKRNNPNDNHTTELF AHVSKILQ DKKAKTKRDYYEGLLSGNNQKAAWKIKNV MGNQKSNSKPTAIRNGAQLITDPQQVCSSFND HFCAVGPCLSSTINTNRNINRFGTLGSLQSTIFL KPSTRNEILLINRLDSKKSAGPDGIPVSFIKHH HSFFASLLMETFNEIIVSGEFPDCLKIARVIPVF KSGDPKDMNNYRPISLTIIDKLLLEKLLVSRILE FSTKHHLIYGQQYGFRQGSSTLTACHELVDEIY DALDNKQIVGALFIDLKKAFTVDHDLRLRKL ESYGIRGIAKDLLKSYLTDRHQYVSIGEHKSQL RLVTGVPQGSNLGPILFLFVNDVARLNLHG KLRLFADDTAVFYRGNTCDTILNQIKMDLEVL LEYFGENVLSLNLNKTKYMLIHTPRRKIPAHGP ISIGGHNLDKVYEHEFLGLTIDSVMSWAHIN	cd01650	420	676

					TTCATCA	KLKSKVSSLCGILRKISSFMPQTCLKQIYFALV			
					CTTCACC	HSRLQYVTAVWGSASKSHLRELQVLQNRCLKI			
					CATCGCT	VLRKPLLYPTVNLYSNRNDSSLPIKALYEQV			
					TGCCACC	LVQMQKIVKGTSTHHNTVLHMSQQSRSSRQA			
					AACTTCA	NHIFLFRPNSEFGKKKITYFGSKLYNALPENCK			
					GTAAAGA	SFQNMFHFKRFVRNALKQKVEQFVL			
					TCAATGC				
					TCATAAA				
					TATTTAT				
					AATGTGA				
					ATTGAAA				
					GTTTGTG				
					AAAATTC				
					TAAACTA				
					TAGCGCT				
					TCCTTCA				
					AAGAGCT				
					ACGCTCA				
					TTGGAAG				
					TGCAGTT				
					ATCGCTC				
					AATCGTT				
					GAATTAT				

					ACTTATT TTTTGA AAAGATG AGAAGGT TTTATGC CTATGGG AGAAGTG GCTTAAA GTAGCTA CACTCTC ACGGGCT TTTCCCTT CTCCAAG AGAAAAA AAAGCAG TAAATAA ATAAATA AA				
CR1	CR1 - 121_ AAe	.	Aedes aegypti	AA	TGATTTTT ATTGTTGT ACTTGTT TTTATAT ATCTTGA	VNDVNVNENGRCLDLCFVSTQDVAPSIAPSA LVKNVSHHPPLVLLADSEPCEPLTISPVSDF NRANQQGIVEFLASVNWLEVLDNDVENAAL TFTHILGHVIERHVPKKVHGTAGPPWQTQELR RLKTSKRAALRLYSKRGTQSHRNHYVSVNNK	cd0165 0	316	566

					CCTATTTT GTCGTAA ATTTAAG TAACCAC ACATCAT TGGGACT AATTGTC TGTTGGT GTAAACA AATAAAC AAAAAAA A	YKSTSRRRCFRRYQQHIQQNLKTKPKSFWRYVN QQRKESGLPSTVKLDDEVASDRNQACRLFAK KFSSVFCDEMLSEEYVALAASNVP LVMETLNA FSVDDGMIARAATQLKSSINPGPDGIPSSFVKS YIEQLLLPIRHVFALSLTSGTFPSVWKNAIIFPV HKKGDRKDVSNYRGISSLCAISKLFELVMEP LIAHCKQFISNEQHGFIPGRSTTTNLLSLTSHVS DSFTERVQTDVIYTDLTAAFDKINHLITIAKLER LGISGLMLRWFKSYLTERQLTVGLEGHFSETFT ATSGIPQGS HLGPVIFLLYFNDVNYVIKGPRLS YADDLKMYLRVRSICDAITLQNDLDAFAGWC SLNRMVVNPEKCNVISFSRRKEPVYFNYQLFG KILQRVDHV KDLGVILDAQLTFKHHMSYVIAK ASRTLGFIFRIAKEFTDVYCLKSLYCSLVRATL EYCSTVWHPHYQNGVQRIESVQRRFIRFALRR LPWRDPHRLPSYHSRCQLIHLETLSRRDINRA MIVADLLQGRIDCPAILGQVDISARHRNLRANS MLRPPLRRTNFGQNVAINGLQRVFN RVAPAFD FHLPRHTVRRNFAEILTRRIV			
Kiri	Kiri- 9_A Ae	.	Aedes aegypti	ATGAA CTGTT CCACT CCTAA	TAGAACA TTCAGGA AAATCGA ATTCAGG	NICHINVQSLIARNFTKFHELKLT FIDSKVDIICF TETWLNHTISNTMIGIEGYKLIRNDRNRHGGGI CIYLRNSLLYKVVSMSQSSSNDFYGTEYLNIEI KVGFDKVLLGVIYNPPNIDCIDVLHDILENCTI	cd0165 0	467	726

			AAGTT	AACATAA	NYDCSFFVGDFNTDILKFSSRSRRFGEMLDAM			
			ATATT	TTCGAGT	SYECMNSEPTYFHQSGCSLLDLFITDSPDFVAK			
			GGGAA	TAGTTAA	LDQISMPGVSNHDLVFCSLRLSLNRVEDIVQY			
			GCTAC	ATAGTTT	RDYVNFDDSSALLDGFYNIEWNNLFCMEDPNEI			
			TGTTTC	TAATGGA	LNFINYHILKLHDDFIPLRRKKKNRTPWFNSDI			
			CCCTT	ATCACCA	SHAFINRNLAYDQWKRSRTEIDRNTYKRLRNF			
			CTGCT	CAGTGTA	ANDLVRQAKINHDKQNFNVDLPSKQLWKNIK			
			GCTGT	ACATTTC	KLGISNDSSFLDQDCLNANQINDYFISNYINDSS			
			TGTTG	AAGAAGG	SVPIPFSSDGFKFNFVHEEDIVNSIFSISNAVGL			
			CTGTG	AGTTTCC	DNIPIRFIKILLPIALPVYKFLFDSIIKTSVFPRAW			
			TGCTG	TTACGTT	KNSKVIPIKKKGNSASLSNLRPISILSSLSKVFEK			
			TTGAT	ACATAAA	LLKIQITKFITDNDLLHPLQSGFRQNHSTNSALI			
			GCTAC	TTCAAAT	KVHDDIARVIDRRGIAVLLLDFAKAFDRVSHS			
			TGCTG	AAAGAAA	KLVKKLSSKFSFSHNAVKLIESYLNERNKQAVF			
			TTGCT	TAAATAA	YNGVLSDFSFIESGVPQGSVLGPLLFSFLNDLP			
			GCTGG	AATAAAT	ASLEFCSIHMFADDVQVYLCASDDVDMNDFA			
			GAGGC	AAATA	RKINHDLNNILRWSQNNLLAVNPSKTKAMLIC			
			GTTGC		KLKNRPLPPSIYFDGKRVQFYDKLENLGVIFTS			
			TGTCG		NLCWDAFINGQCGKIYGTLLKKLNLTTKHLDTA			
			AATAA		MKIKLFKSLILPHFIYCDFLFSNATASALNRLRV			
			GATTT		ALNACVRYVFNLTRYSHVSHLQDVLLGCQFS			
			TTCTG		KFYSFRAQLQVYIIKRRKTPRYLYDKLQFMRNT			
			CTGCT					

			CTCGA	RTMSLVIPNHSSAYYGQSLFVRGINNWNRLSP		
			TTGTG	TLKSTANIYGFKGGLLRELERLQ		
			GTGAA			
			TTAGA			
			TCAAA			
			TTAGA			
			CTTTTT			
			CGTTTT			
			CGGCA			
			TATTA			
			GCAGA			
			CATTA			
			TTTAG			
			CTGTTT			
			TTTCTT			
			TGTAA			
			TGTGC			
			TTTTCT			
			CTAAT			
			GTCAA			
			ATGAA			
			CTTCA			
			GACTA			

				AAAGT					
				TAATC					
				ACAAT					
				GAACA					
				TATTA					
				GTAGT					
				TAGTA					
				GTTTT					
				GCTTTT					
				CCTGA					
				TTTCTA					
				CACTG					
				CATGT					
				TGCTT					
				GGGTT					
				CGATG					
				GGTAG					
				ACTGC					
				TTTTG					
				AAGCC					
				TATAC					
				TGTTC					
				AACAT					

				TTAGA TATGG ATGAC TTTTCC CTTGA TAATG CTTCTC CGAAC AATGT AATTC CACGT GTTGT TCTTA ATGCT GCATT AAAAC ATGAC TGTTT GA					
CR1	CR1 _Ele 32	.	Aedes aegypti	GG	TAGGATT TAGTTTC AATTCAT GTAGACA	LRTKIDALYCNAIDCNYDVII LIETGLDAAINSV QLFGDNFNVFRCDRNAKNSTKSSF GGVLIAVH RRFTSSSISITTGESLEQVCVSVLVGNQRILLCAI YIPPDKSKDVNVMD EHI ESVREL RNLSSGNDY	cd0165 0	497	748

					AAAGTCA GATGGAT AATATAC AAATATA CAAATAT ACAAATA T	LLVFGDYNQPELCWKRCGDEIRPTNSGSISAAS SSLIDGMDFCNLHQMNIERNHLNRTLDLVFTS LELSPEITCSVAPLLPTDQHHPPELDDLMLPNG VNNVPTVSNATGELDFTKIDFEALSLFLSETVW HVLLDGLTADNMAETFCSVIANWLNTNVPTR RPAVSPVWGTPLLRNLKRRRNALQRRLRRHRS LENIRSFHRACDEYRELNSSLYKSYVLRVQSNL QRNPKQFWNYVNSKRKSSTIPMSVFLNDSVSS SPLEKCELFAAHFSSVFDKASPSEVDSAVSD VPRDLLDLDFQITPAMVSQAACKLKCTFSPGP DGIPSAVYCRCMSALAEPLCRIFNQSFVDSTFP AIWKQSFLVPVHKKGEKRVINYRGITSLSAA SKLFEMIVSDVILAQAQGYISIDQHGFMPGRSV TTNLLDFSTTCFEQLENGAQVDVIYTDLKAFF DTINHDILLAKISKLGASNRLSSWLLSYLTGRT LRVKVDSCMSSSFTSMMSGVPQGSNLGPLLFL FFNDVMLLLGHGCRLAYADDLKLYFVVRTVE DCVRLQALLDMFVDWCRRNRMTLSVPKCIVV SFYRITRPIMHNYVVNDVVLKRADSFCDLGVL LDHKLTFNLHRSNVIAKANRQLGFIKISRDF DPHCLKALYCSLIRPIESA VVWTPYQLSWSL RLESVQRKFMRLALRNLPWRDPLNLPYPDRC RLDLDTLDRRRKTQQVAFIAKLMNGEVDCS			
--	--	--	--	--	---	--	--	--	--

						RLLSMLNFRVAPRVLRGGALLQQRFHRTAFGS NEPFTSMIRLFSMFEDLYDFDGSSNQFVSRVRS SDLI			
CR1	CR1 _Ele 41		Aedes aegypti	CTCCC AGCCT GCACT TCCAG GCAAG TATGC GTGTT ATACC AGACG TTCTTT SCGTG ATGAA TTGCC AAATT TCAGC TCTCT ATCCG CCGCT GATGT	TAATGTT GTTTTAT GTTGTTG TATTTTAA TCGTTGA TTGTAAT TTTGTTA GTAACGT CTTCATA TTTGTTA AACTAGA ATAAGAT ATACATA AGTTATA CATCATT GGGACCT GTTTTGT CTGTTGA TGTCATC CAATAAA	MTPKCNPVMSSNGQQPGRSVESLMEAPGPSDP VERNAVTVLHSRPGPAVGCGTGVPQSATSGEY SSIIDNAPSPSIVLNSRTCNARKMPISVNAKKSD DLIIYYQNAGGMNCDVDKYLLATSDDCYDIIV LTETWLDTRTISTQVFTPEYEVFRCDRNPRNSR KSTGGGVLIAVKKKLKANIENDLWCSTEQLW VHVQFFGHSLFLCGIYIPDRARDETLIETHLQS VTSVIEIAKPMDEIVVLGDFNLPGISWISSGNGF LHPDMNHSHVHSAASRLLDGYSTATLRQINHV TNENLRSLDLCFVSAQDVAPPISNAPSSLVKQV NHHPLVLKLIVQHYSTSESIAQVTYDFTKADH EGILEFLAAVEWTEVLDTRDVDNAVSTLAHIL GHVIERHVPKKTASKGLPWQTRELRTMKTA KRTALRRYTKNPTPRMRDQYVRLNAAYKRTS RHCFNRYQQQVQRDLTAKPKSFWRYVNEQRK VSGLPSSLFLNGDVASEPQSVCLFAQKFSSTF CDEVIPEVLVTQAASNVPVLVNDSWATLTVDD DMISRAASQLKSTMKPGPDGIPSAFLKRHIDSL LVPIRHVFNLISGAFPSLWKNASVFPVHKKG ERKDVNNYRGISLCAISKLFELVMDPLMTH	cd0165 0	607	857

				AACAT CA	TAAATAA A	CKNHLSDDQHGFISGRSTATNLIRLTSHIADSF AEGAQTDAIYTDLTAAFDKINHAIATAKLERF GISGNMLRWLHSYLIGRKLTVTIEGFQSNEFLA TSGIPQGSHLGPLIFLIYFNDVHYTLKGPRLSFA DDLKIYRRIHSDYDAMCLQQDLHAFASWCTL NRMTVNP GKCSVISFVRIRRPVVFRYEL YNTVI QRVDSVKDLGVILDPQLTFKHHVAYVVAKAS RTLGFIFRVAKDFTSVYCLKSLYCSLVRSTLEY CSVVWNPYYQNGSDRIESIQHRFIRFALRQLP WRDPLRLPSYRSRCQLIDLETQLRRDIARAM VVADTLQGRIDCPVLLQSLDLNVRPRVLRNNT MLRLPFRRTNFGQQSAYIGIKRVFN RVASV FDF HLSRQTIRRKFKTVFLSFLDE			
CR1	CR1 - 109_ AAe		Aedes aegypti		TAATTTTT GTAATCT TGGTTTG AATATTT TAGTGTT TGTATTTT GTTTAAG AAACAAG GGTCATT GGGGTAA	AAPASSNILMYQNVGGINSSLAEYNRAVSDS CYDVYAFSETWLNANTLSNQLFDSSYIVYRQD RSASNSNKRTGGGVLFVRSNFKSRVLNSPNG SSIEQLWVAITTVDATLYLCVIYIPPDRVNDESL IEKHTDSL DWVVSQ LKPRDSIMIIGDFNLSGIS WQYDXSGSLFPVASRSTISLTSRKLLDAYSTAC LRQTVGIMNENSRVLDLCFVSEELRDASTVMQ APSPLVKTCRHHPILLKMEICPQKRFDVSES VFYDFRNADFN GMNDF FENV DWDDVLRDSD ANLAASTISAILLYGIDQYVPVKVKREPSKPAW	cd0165 0	518	768

					TGTTTTA CCTGTTG ACTAATA AATAAAT AAATAAA	SNSELKNLKRVKKAALRRHSKFRTDSTRARY MQANTEYKQLNDCLYNAHQDHLQSILKANPR SFWKHVNEQRKESGLPPTMSDGLHEADSKEDI ADMFRTQFSNVFTNEHVDPDIVNDAIRDVPRL STFVQQIIITDDMVASASKELKSSTGSGPDGIPS LVLKRC AETIAAPLARVFNL SLASGVFPNCWK QSYVFPVFKKGSKQLVSNYRGIAALSATSKLF EVIILQKL VQSYAHYISPDQHGFMPKRSTTTNL ACFTSFILRLMEAGQQVDAIYTDLSAAFDKMN HQIAVAKFDKLGMGNSMLLWLQSYLTGRNM SVKIGDHVSSPFRVW SGVPQGSHLGPFLFLLY MNDVNFKLKCLKLSYADDLKL YWTIKQPEDA VFLQHELEAFAEWCQINRMSLVSKCSVISFSR KHAPFHFGYFLAGTQLERVSTVKDLGILLDSK MTFKDHVAYAVSKASSQLGFLFRYAKKFRDV YCLKALYCSIVRPTLEYSSIIWSPYYQNGIKRV EAIQRKFVRFALRHLRWRDPLNLPSYGSRCLI KLESLVSRRNVAKACFVGDLLQGNIDCSTLLS CLDINVRRLRLRTHSFINIPSARTNYGLQEPMR SMSRVFNM CYHVDFDNVSRETNKSSYRRVLC			
CR1	CR1 -	.	Aedes aegypti	AATTC TGGCA CCACT	TAACAGT CTGGCGG ATTAAGG	MRTKTNEFRTKMETCDYDIIVLTETWLRPDVV NAELASNYCIFRCDRNDNTSTLQRGGGVLIIV KSTLRCSAVELTDYIHLEQVVAVRLPDTTIYV	cd0165 0	495	745

48_ AAe		GCGTG	TATTGAT	CGIYIRPNSHFDVYDSHASAVQEICNLSSTRDSI		
		TTTTCG	CCAAGAC	VVVG DynLPHTWHFDDDCNSYLPRNASSES		
		TTTGT	G TAGCAA	EIMFTQSMIGSGLHQINSLRNMNGRILD LAFVS		
		GTTGT	TTAAATA	EXSFVELIDPPSSLLKVDSHHKPFVLR IEIETRT		
		TTATA	AATAAAT	RAPVEEADVDDDFDKRCDF AALNDMLVAVD		
		ACGTC	AAATAAA	WNETFGDAPLDRITVLFYDKVFDIIRHHVPRRR		
		GCCCT		RCSAVH SKLPWWTPELRHLRNIVRKARKRYF		
		AAAAT		RNKSHDN RKNLKSLETRY SVCQENAFRSYTNR		
		CCTGT		MESNLKRDPK SFWTYIKKLKNINRIPEELSYNG		
		GAATA		TTAETA EASANLFADFFSSVHNNHSPTFSSSAR		
		TTCGT		QNIQSFNVSLPMITISQHEVSSALKKLDVSKGA		
		GCAAT		GTDCLPPTFLKECADSLKTPISVIFNRSLREQVF		
		CAACT		PEIWKVASITPIHKSGSVHSVENYRGISILCCIA		
		CTATA		KVFEEI IHRSLYSAVRPIISDVQHGFVKKRSTVS		
		AATCA		NLMSFTNVLSNSVEKRRQVDAIYFDFAKAFDK		
		GGGTT		VPHELVISKLKHIGLPDWITEWLRSYLTDRKA		
		TATTG		HVKVGRARSHSYDISSGVPQGSVLGPLIFILFIN		
		CTAAA		DLAVRLKSGKLMYADDLKIYREISSTLDCCAL		
		CTGCC		QSDVNELISWCTENGME LNIDKCKTITFTRRQT		
		TCGTT		YVTHDYAIGNNIIERVFSIRDLGV IIDSKLKFTE		
		CCGCT		HISTVTAKGFAVLGFIRN SQSFRDVYTLKALY		
		TCTGTT		CSLVRSVMEYAVCVWSPYHATHMLRIEKVQR		
		AGCCG		CFIHYCLRQLPWN DPTNLPDYVSRCLIDLETL		

			CAAGC	TSRRTKIQRLFVYDLLSNNIDCSELTNELRFYA		
			ATCAG	PTRQLRERQLLVIRPHRTTYGQNCPMSSCLRAF		
			GTTTTC	NDVGCYFDFNVSKCTFKHRIKYLN		
			GCATC			
			ACTTTT			
			GGCAA			
			GGTAG			
			GACAG			
			TGATT			
			TGTGA			
			TCTAA			
			ATCCG			
			AGCAG			
			GCTAG			
			AGAGT			
			GGAAG			
			CATCC			
			GCCAT			
			CGCCC			
			GATAA			
			CATAT			
			TTCCC			
			ACCTC			

				GCCGC					
				TCTCC					
				ACCTT					
				CCAAC					
				CGATC					
				GAAAG					
				CTGCA					
				ACAAG					
				ATCAA					
				AAATC					
				AACCC					
				GCCCC					
				CATTT					
				GCCCC					
				ANCGC					
				CCCAC					
				CAACG					
				CCCTA					
				CTAYT					
				GTTTG					
				CTGCA					
				CCGAA					
				CGAAA					

				TCAAA					
				ACAAG					
				AAATC					
				GTCCT					
				GTAAT					
				ACTCC					
				GCTTC					
				ATATT					
				GGAAA					
				TAAAC					
				TACTT					
				ACGCT					
				GCCTC					
				GTCTTT					
				CATTC					
				CAACG					
				CTCCG					
				AAAGG					
				GACCA					
				ACGAA					
				CGATC					
				ATCAT					
				ACAGC					

				CTCAA					
				TCGCC					
				GTCCA					
				AAGCC					
				AACCC					
				GTTGC					
				TACCA					
				AGGAT					
				TAATC					
				GCCAA					
				ACTCT					
				GCATC					
				CGTAG					
				TATCA					
				TCCAT					
				TCACG					
				TAAGG					
				AAATT					
				CTGTTT					
				TGGTT					
				CTGCA					
				CTTTGT					
				TTATG					

				TATTG					
				AACTG					
				TGAAG					
				CTCAC					
				CGTTG					
				ACAAG					
				CCACC					
				GTTTG					
				TGACA					
				GAGAA					
				CACAC					
				GCGCA					
				GCTCG					
				TGGTC					
				ATTCC					
				GAGAA					
				TCAAT					
				AGTTC					
				ACAGT					
				ATATC					
				ACCCC					
				TTCGT					
				AACTG					

				GGAGC					
				TTGTTT					
				CTCGC					
				AGCTA					
				AAATG					
				CAGGG					
				TGTTT					
				GTGAA					
				AAATG					
				CGCTT					
				CCGAA					
				CTATC					
				CGCTG					
				ATTTTT					
				GGAGC					
				TGTGG					
				TGGAT					
				TTTGTT					
				CTGGT					
				CGCTT					
				TTGCTT					
				CAGGT					
				GTACC					

				GGAAT					
				GAGTC					
				CTGAT					
				ATCCA					
				AGCTG					
				CCGTT					
				GCGAA					
				TAACG					
				GGTGC					
				CTTCA					
				CTGGA					
				TGTGT					
				AACGC					
				CTGTA					
				GCGGA					
				ATAAT					
				GATTA					
				AAGCT					
				CGCTT					
				TACTA					
				AGGCA					
				ATCAC					
				ATCAG					

				TGAAT					
				GCTGC					
				CTATG					
				AAGGC					
				ATAAT					
				CGAAT					
				CAATG					
				AAAAT					
				CCGAA					
				ATTCG					
				AGACA					
				ACATT					
				TTGTC					
				CGAAA					
				TCCGC					
				TCGGA					
				ATTGC					
				AGAAT					
				AACCT					
				TAAAC					
				GAATT					
				CCGGA					
				GATTG					

				TTGCA					
				AACAC					
				ACCGA					
				TTCGA					
				AATAT					
				TCAAC					
				GTCGC					
				ACATT					
				TGATT					
				CGGCT					
				GTCCG					
				TAATC					
				CTGCA					
				AAACG					
				TCTTC					
				GTGCA					
				AATGA					
				TGATG					
				TGCCT					
				CTCGA					
				CCCGC					
				CTAAA					
				AAACT					

				GGTCT					
				GTGGC					
				ACTGA					
				TGGCC					
				ACAGT					
				ACTAG					
				TTCTTC					
				AGACG					
				TCATC					
				GCAGT					
				TCCTG					
				AACGA					
				ACGGA					
				TGAAA					
				ACACG					
				TTTTG					
				GCTTT					
				ACTTA					
				TCAGG					
				AGTGT					
				CCCCG					
				AAAGC					
				GCCTG					

				ATGAG					
				AAGGT					
				TATCG					
				AAATG					
				GTCAA					
				GAATA					
				GGCTC					
				GMAAC					
				AGACG					
				ACTTA					
				ACTGT					
				CGTTA					
				AACTT					
				GTTCC					
				TCGTG					
				GAAAG					
				GATAC					
				AAACA					
				ATCTC					
				ACGTT					
				TGTAT					
				CCTTC					
				AAAAT					

				AGGAA					
				TGAAG					
				CTTGA					
				TCTTA					
				GGGCA					
				AAAGC					
				AATGA					
				CTTCA					
				TCAAC					
				CTGGC					
				CAGTT					
				GGAAT					
				TCGTTT					
				TCGCG					
				AATTC					
				GAGAA					
				CCACA					
				GTTCC					
				TCAAG					
				GGCGG					
				GGTTC					
				TGGAA					
				TCCCT					

				CGGAG					
				AACCA					
				GCCCG					
				AACCA					
				CAGAT					
				CGTAA					
				CAGTA					
				TTGAA					
				CTAGT					
				ATCCC					
				CATCG					
				TGTCC					
				CATAG					
				ATCTT					
				CATTC					
				TCCTC					
				GTGAT					
				GAGAC					
				CGTTTT					
				AAGCC					
				ATTTG					
				ACAAT					
				CTACT					

				ACCAA AACGT GCGGG GT					
CR1	CR1 - 64B _AA e		Aedes aegypti		TAGCTTA AGAAACA GTCTGTG GAATGTA ACAATTC GAGACGG TGACAAT AAATAAA TAAATAA ATAAATA AA	EQSIVCVKLQKSSIIYICGIYLRPNSQPVLVYSSHS AAIQQLSERISRFDSSIVVVG DYNLPQLIWQTDD DINGLLPSNASSEQEVTLLETMVASGFQQINNI VNSNGRLLDLAFVNDNNDVELLEPPTPLL RMD NHHKPFVLRIDVND SCTQVGDSNHADDFDFR YCNFDELKSAIATVEWATLFIGKDTDET VCTF YEV LHRILDEHVPRKRRRRTHPFKHPWWTSEL QHLRNIVRK SRRRYFQSRTVENRDKLRLLEAS YIDCQTASFRRYVARMETTAKEDPSAFWTFIR NRKRVNKFP AEMAFRDTVANSPEGIANLFA DF LENVHCTNPPTFSPDSL RNCPTFDLNLEPF DFA QQDVLSALQKLDTSKGPGTDNL PPLL FKECAD SFQIPLTIIFNKSLQSR TFPMLWKTASVIPIFKAG STRGVENYRGISILCCLAKIFEELVHNVLYTAS QPLISQFQHGFVKKRSTTTNLMTFTSFLSTEIEN KHQVD AIYFDFSKAFDKVPHDLAISKLRHLGF PNWIADWLRSYLTNRKAFVNINGTYSRVISVT SGVPQGSVLGPLIFVLFINDLCFRLKSGKLFYA DDLKIYRTIASHLDCCALQADVNELELWCQEN	cd0165 0	413	664

						GMELNIKKCKSIAFSRRQSRTEFDYKIGSEPLE RVESIRDLGVIIDTKVRFNDHISVINAKAFAAL GFVRRSTNDFNDIYALKSLYCSLVRSILEYAAAC VWAPHHTTQIVRLEKVQRSFIRYALRQLPWS PVNLPDYPARCMLINLEMLDARRNNLQRLFVF DLIMGNIDCPALLEDVQFYAPLRQLRERDLLQI RRHRTSYGFNNPLSRCFRLFNCASALFDFNVS KYVFKNRLKDF			
CR1	CR1 - 8_A Mi		Alligator mississippi ensis	TA	TGATCTC AGCAGTA TTCCTGC CCAGAGC AGGGGGG TTGGACT CGATGAT CTTTCGA GGTCCCT TCCAGCC CTTAATT TCTATGA TTCTATG ATT	GIYYRPPHQEEELDLEFSRELMEAACSRDMVV MGDLNYPDICWEEHLARSDRSCRFLTCVEELY LVQEVNGPTRGKALLDLVLAKGDDLVGDLRI EGNLGDSHEMIEFTIRRRAGKSASRAEVLDF RKADFNELRKIVGETLWEEGLEGKGVQEGWS FLKGTILRAQKRSIPVCRKGGKGAGKPSWLTR GVKECLKKKKEAYGQWKQGVVTKEEYTSIAR ICRERIRKARAVTEIRLVTEIKDNKKSFFRYIGS RRKTKGNVGPLQNGTGQLITDSQEKAELFNEY FASVFLHANGGRPPGRILDGCGRGDCQPMVSV ELVREHLERLNVFKSAGPDGLHPRVLRELAGV IAGPLAKVFEDSWCSGEVPEDWKRANVVPIFK KGRREDPGNYRPVSLTSIPGKILEKIIKKSICERL AEGMMLKDSQHGFVAGRSCLTNLVSFYDQVT NRLDVGDAVDVIYLDFFKKAFFDMVSHDALIKK	cd0165 0	375	626

						LGDCGLAESTVRWVGNWLRGRTQRVLVDGS VSSWREVASGVPQGSVLGPVLFNIFINDLED VESTLAKFADDTKLWGKAVTLGDRMRIQADL DKLERWAERNRMQFNKDKCRVLHLGRSNQR HEYRLGGDVLASTMAERDLGVMVDHRMNMS RQCDAVVNRANSILGCISRCVASRAKEVILPLY SALVRPQLEYCVQFWAPHFKKDVENLERVQR RATRMIKLEDKPYEERLRDLGLFSLGKRRLR GDLVAVYKYVRGEYRELGEKLFTRAPQGRTR SNGHKLLEDGFRLDVRKNFFTVRVTRIWNRLP KEVVQSPTLEVFKRRLDRHLAGII			
CR1	CR1 - 9D_ AMi	.	Alligator mississippi ensis	TCAGA ACAAG ACACA TGCTC GGGTG GGATG GGCTT CACCT CTCCC CCAAA GGTAA GCATG	TGACCCC AGGGTTG GTGACTC TTTTCTTG CCTGCCA GGGGCAG GGGGTCA GACTAGA TGGCCCA CTGAGGT CCCTTCC AACTCTA	SPPAGFKLGSSGEDEYGGSHGTPNQETPTRS AQVCSKDPIRQQFKGGPGTATRGIRCIYTNARS MGNKQEDLELLASNNFDLVGLTETWWDSSH DWVVGIEGYRLYRRDRVGKKGGGVALYVKE QYTSSLIKLGSEGGQAETLWVRIQGARGERDL VVGVYYRPPHQDDELDQEFSKQLEEAVRSRD MVVMGDLNYPDICWEEQSARSNRSRRFLTICIQ DLHLTQEVRCPTRGNALLDLVLATGDDLVDG LRIQGNLGSDHQMIKFTIQRAGIIPSRARVL NFKKADFNRLRGLVSRVLGAKNLDALRVQDG WTFLKETILWAQSESIPWQRKGGKGARKPLW LTRGIQDTLRAKKEAYRKWKQGVASKEEYTS	cd0165 0	536	787

				TGTTCT CTTTC AGGTT GGCTG A	TGAATCT ATGA	LARTCRESVRKAKATLELRLASGIKDNKKSFF RYVGSKKRVRGSIGPIQDGQGGLVTDSRDKAE LFNEFFASVFLNGNLDEPSTYITDTNKGNISRPT VSADTVKRHLEGLDVFKSPGPEGLHPWVMKE LAGVIADPLARIYERSWCSGQVPEDWKRANV VPIFKKGRKEDPGNYRPVSLTSIPGKILEKIIKE HICGSPLANVMRRGNQHGFFLAGRSCLTNLVSF YDKATNCLDAGVEVDVIYLDFFKAFFDKVSHTI LINKLRGFALDDYTVDWVANWLKGRTRQVV VDGSVSTWKDVGSGVPQGSVLGPVLFNIFISD LDEGVGSTLSKFADDTKLWGAANTLERRNQI QADLDRLGKWAESNRMQFNKDKCKVLHLGR KNCHHTYRLGGDLLSSTEAEKDLGVTVDSRM NMSRQCDEAVGRANRTLSCIRRCFTNRSREVI LPLYAALVRPQLEYCVQFWMPQFKRDADNLE RVQRRATRLIRGLQGRPYEERLRDLNLFSLSKR RLRGDLVAAYKFIRGEQHEVGDDLFTRAPPG MTRNNGHKLLERRFRDIRRKFFTVRVARIWN GLPREVVLAPSLGVFKRRLDRHLAGIT			
RTE	RTE - 1_A Tr		Amborella trichopoda	TAATT AGTGC TTATG CACCA	TAATTGG GAAAAGG CTATAAT GATGAT	MGETILDFAMAYDLAIANTYFKKRDEHLITYK SGSHKSQIDLFLVKKVDRPSCDKCKVIPGESLT TQHRLVMDVCIKTWKRRGESNGCPRTRWW NLKNEKIDLFRDKLVEEGKWNSEHDVNVIWN	cd0165 0	353	603

			CAAGT	EMADCIKKVAKEVLGESRGKRNFDKETWWW			
			AGGAT	SDEVQKAISEKRSCYKTWQCNRNDQNLEEYK			
			TAGAT	NAKKNAKQIVSKAKLKAYDNLYNKLGTREGE			
			GATAG	KDVFKLAKIREKKCRDLNHIKCIKSEDQRVLV			
			AATTA	KDDEVNERWRVYFQKLLNDEHNGNIVVEGTV			
			AGAGA	GTSEGGGHRYIRRIRMSEVKEALKRMKIGKAL			
			CAATT	GPDGIPIEVWKCMGEIGIDWLTKLFNKVLRTK			
			TTGGG	KMPDEWRKSIVVPIYKNKGDIQNCANYRGIKL			
			AGGAC	MSHTMKLWERVIEQRLRHETSVSENQFGFMP			
			ATAGA	GRSTTEAIFLLRQLIEKFRAKGKDLHMFIDLE			
			TGGAA	KAYDRVPRDLIWRVLEKKGVSRHYIDLVRDM			
			TAGTC	YDGALTSVRTTSETSEFSVTIGLHQGSALSPY			
			CAAGA	LFVLVMDDLTRHLQDEVPWCMLFADDIVLIDE			
			CATAC	TRNGVNTKLELWRDALESKGFKISRKTTEYME			
			CAAAT	CKFSNNRSRDEEVVKIDGQEVQKRPHFRYLGS			
			GGAGA	IIQENGEIEEDVAHRIRAGWAKWRCATGILCD			
			GAGGA	RRIPMKLKGGFYRTAVRPALLYGAECWADKK			
			CCTTTT	QYTHRMSVAEMRMLRWMSGKTRRDRIRNESI			
			TAGGA	RENLGVAPIGDKMRESRLRWAGHVWRRPSTT			
			GGGGA	PVRRCELVQVEGLKRARGRPKRTWLEVVRKD			
			CTTAA	MGTYGLTEGMAFNRAEWRTRIHVADPN			
			ATGGT				
			CATGT				

				AGGGA AGGAT AGTAG AGGAT ATGAA GGGGT ACATG GGGGA TATGG GTTTG GGGAA AGAAA TGAG					
L2	Una L2		Anguilla japonica	CAGTG TGCAT CTGAT TGTGT CGTCG CTTCT GCCGT CCCCC GCGAT TCAGA	TAACCAC CACCATT CTGTATA TTTCACT CTTTAAT TTAAAAA AAAAACA AAAAAAA AAAAAAA AACCCCC	MFINPIPLCALPRPKWSIHRRRNLSNMIYPPLST PQEFTVTGGLWNCQSAVKKADFITAFKTKSL DFLALTETWITPDNTATPAALSAGYSFSHTPRA SGRGGGTGLLISLAWKFCVLPSPNPTSSSFESH AITVTHPSKFHIIIVLYRPPGPLGSFLDELDTLSS LPEDGTPAILLGDFNLSPETSQLSRVASLLQSFG LSLSSSPPTHKAGNQLDLVFTRSCSIPQLTVTPL HVSDHFFLSFPLSYLHPLNSSNSSPTVTFRRNLS TLSQSSLISTALSTLPPPASFSNLPVDLATSTFNS CLSQSLDSLCPVSKPARSSPPCPWLTESLRSSR	cd0165 0	513	773

			TAAGC	CCCCCCT	TSLRAAERKWRKSRSSDDLTA YHTLLAAFTSA			
			GTATC	TTTTCAT	LTSAKASFFQSKISACVSNPRKLFSTFSSLLTPP			
			TTAAC	GTCACCT	PPPPSSLSADDFLAAFEKGK VATIRNSFAHPVSP			
			TTGAT	GTATTTG	PPPRDRPTATLTSFEMLESDVLQLITSHRATT			
			TTGTCT	TCTTTGTC	CPLDPIPSKVLQSSISSEFLPYLSSIVNSSLSSGHV			
			CTGCT	CTAATAC	PSDFKMARVTPLLKKPTSDPSDVTNYRPVSLLP			
			GTTGC	TGTAGCT	FLSKTLERAVLKQLTSFLHQNNLLDPHQSGFRS			
			TAGTT	TACTCTT	GHSTETALLGVTEALATARASRLSSVLILLDLS			
			AGAGA	CTGCCTA	AAFDTVNHKILLSTLAGMGIAGTALSWFASYL			
			ACATA	GTTGGCT	ADRTYQVTWKGSASASHPLVTGVPQGSVLGP			
			GTTGT	TTGCACA	LLFSLYTRSLGSVINSHGFSYHCYADDTQLFFS			
			GCGTA	GGTTAGG	FPPSVTQVNDKISSCLADISTWMASHHLKLN			
			ATTTA	TTAGAAT	NKTELLFFPCKTSLREL SITVDGTTVTASHSA			
			GATAA	AGTGTC	KSLGVVLDDQLDLKEHIKATSRSCRFLLYNIRR			
			TCTTTT	ACTGTGT	IRPYLTTHSTQLLVQATVTSRLDYCNLLASLP			
			TTAAA	GAACTGT	ACAIQPLQMIQNAAARLIYNLPKF SHVTPLLRS			
			CGTGT	GTTCTTA	LHWLPVAARIRFKTLT LAYTAANRTAPIYLQD			
			CTTTA	GCTAGAA	MIRFYV PARPLRSAEAGRLVTPPTRPKGSQSFS			
			CTGTT	ATAGCTG	TLAPQWWNELPVPLRTSPSLPIFRRGLKTHLFR			
			GCTAG	TACAAAA	LYLD			
			TCAGC	TAAGTAT				
			GAACT	TATACCT				
			TAGTT	TTCTGAA				

				GTGCG	CTTGTGT				
				TTAGC	TCAGCAG				
				TGAGA	ATGCCTA				
				ATCTC	CGACCAT				
				TGTAG	GATATGC				
				TTGAC	ACTTTTG				
				TCACT	TACGTCG				
				GTTGT	CTTTGGA				
				TAGTT	TAAAAGC				
				AGTTA	GTCTGCG				
				AACGC	AAATAAA				
				GTTAG	TGTAATG				
				TGAAA	TAATGTA				
				CTGTG	ATGTAAT				
				TGTGG	G				
				GGGTT					
				GGTGT					
				TTAAC					
				TGCCC					
				GGTAT					
				TGTTG					
				AGCTA					
				ATTTC					

				AAGTA					
				GCTTC					
				ACCTG					
				GTGCT					
				TATCT					
				GCCTT					
				AATGA					
				AGGTG					
				ATGCA					
				AGCAC					
				GTAAT					
				TGTCA					
				CCCGG					
				TATTT					
				ATAGC					
				TCCAG					
				CGGAG					
				GCTGC					
				CATAG					
				GCAGC					
				CTCGT					
				CGTCA					

				GTTTG TG					
R4	R4- 1B_ AC	.	Anolis carolinensi s	ATGMT GTTGA GAGCT GTGCT GGCGG TAGTG TAAAC TACCA GCAGG TTCAA CCAAG CCAGA GTGGT CTCAA CTGAG AAGTA GAACT AAGTG CACCC AACCC	TAGACAC TTGGGAA GTGTCCG ACGTGTG ATCCAAT ACAACAG CCAGCAG AGTGTCT GCTGTGG ACTCATC TTGTTGT GTTTCAA	MVKQTLKKSILARSSPRIRTTASATPGHLSTSG LQEPGPIAEQPGHGPVGQLEAKQLQKWQKFS MPQNRVTVMNCYYRSEPQKRGYQKRMYYQLWK LEYPDSTVTEQRLADQRRYITKKKVLSEVELE EIKSNCETEEQTTVETAEDLMTSEPTHEMEPE ENAGIFQDDEHEEPVAIAYTTPGTLTMKQQEL KENFLTHAALKEGRQRLPTLKTVPKHLAPLM NDVNAALATVEITTMEETNQLAYSTAVIVTQE LGIQQLKPARKMIGKPKWKIRLELKIRSLRSDA SNLKNMKEQKLKNVKTKNYLIKKYWLNTRKI EEALEIVKQQITATARKIERYEARVIQYKQNQL FQSDQRRFYQSLNQQVAVKNEKPEKTATNAF WKDLWETEKGYNKKAEWIKDFENKFSENKM RTLEITTEMITERVKKVKNWTSPGNDQLHGFW LKYLSSLHSKMAEQFNEMLQTGNISKWITTGK TYLIQKDKPAKGAIPGNYRPITCLPTMFKLLTGII ADQIQDYLEENNILPVEQKGNKRRSRGTDQL LIDKMILENCKSRKTNLYTTWIDYKKAFFDSLP HCWIIKCLDVIGVCENIRKFTENAMKQWKTEL FVGNESYGTVNIRRGIFQGDSLPLLFIIAMVPL SVILQKTNLGYQTARNSPKISHLLYMDDLKLY	cd0165 0	476	742

				ATTTA AAT		GKSETEIQSLTNTVRIFSTDISMEFGLDKCATV ALKKGKITHSEGIEMPNGQTIKCNQPEAYKYL GILQLDNIKHGHVKNVVSKEYIQRVRKVLKSK LNGGNTIKAINTWAIPVIRYTAGIVNWTQAEL DDLDRKTRKLMTIHYSLHPRSDVDRLYLPRKS GGRGLLQVKQTVVEEKHALADYVKGSQEPTL REVNSSKLLQVQKTKSEYHKNTIQSRRENWRK KALHGQFLGKIESQIDKEKTWLWLTNGTLKKE TEGLILAAQEQAIRTNAIKAKIEKSTTDPKCRL CKEADETIDHILSCCKKIAQTDYKQRHNTVAQ MIHWNLCHKYHLPATKNWWDHKPEKVTENE HVKLLWDFRIQTDRLVLEHNTPLTIVLKDKV WIVDVAIPGDSRIEEKQLEKLTRYEDLKIELQR LWHKPVKVVPVVGITLGAVPKDLGLHLNTIGA DKITICQLQKATLLGSARIIRRYITQS			
L2	L2- 19_ ACa r	.	Anolis carolinensi s	ATGCA TTACA GAAAC CTGGA TACGA GAAGG AGATA ATATT	TAAATCA GATCTAA TTTGCCA AATAGTC ACTGTTG AATGTTT TTAGTTG AATGTTT	MGIECIGLTLAIRESVTILLVYRAPGAPNENIPK LLDLVADWALKFPSLIVLGDFNVHADRLSSSI ARDLVQSMETIGFSLMVTGP THRAGHTLDLIFS IGIQVTLSSVLEIPWSDHYILNTQIKSSPPQHPET EQIFARPRKLMDPIRLSSVLKELDYTADSLHER VENWNKNLSTALDIVAPKRAFRPRRNRSPWFT DALRQMKLERKRLEKWRKKPDETSKLALET HQKIYEAATREAKKEYNSSLIASARSRPAQIFKI	cd0165 0	388	648

			CCCTT	TTATGTT	VRNIINPTHPAKDMASQDITAETFQSYFINKINL			
			GCATG	GAATGTT	LREGLPPTEDTLRELEAFWPSSDKLLELFVPLE			
			AACTA	TTTATATT	MEDLEKTLAAARPTTCSLDPCPSRLVKECWEG			
			ACACC	GTGGAAT	LRDPLRDIINSSLEQGIFPECLKEAVVHPLLKKP			
			GTCAG	GATATTT	SLDAAIPSNYRPVSNLPFLSKVIERTVAGQLQQ			
			GTTTC	TAAATTG	FLDETAGLDPFQSGFRSGHGTETVLVAITDEIR			
			CAGAT	TAAGTCG	RQLDRGGSALLVLLDLTAAFDTVHDLLIHRL			
			GTTTC	CTCGGAG	AMSGVRGLALXWFNSFLRDRRQRVEWSGQSS			
			AAAGA	CACTCTG	ESSCLCCGVPQGAILSPLLFIYVRPLAGLVRSF			
			TCTCG	GTGGAGA	GIECYQYADDTQLLLRLEPGTTTIPENFKLCLE			
			AAAGG	GCGACTA	AVMSWLRASRLKVNPAKTEILWPGQPPELIQS			
			GGGGT	ATTAAGA	LPDFDGEVLVPSATVKSLGVVLDSSLTMEAQI			
			CGAGG	AATAAAG	TAISKQAFFHLRQARKLASYLSTEALATVIHAT			
			GGGGT	TGAAGTG	VTTRLDYCNALYAGLPKTKTQKIRMVQNAAA			
			GGGAA	AAGTGAA	RLLAGSSIRSHITPILKQLHWLPIEHRISYKILIT			
			TAGTG	GTGAAGT	FRARNGQGPLYLRDRLIPFFHQWSPRSSQENLL			
			GCTCG	GAA	YVPGPREVHLEATRRRAFSTYAPILWNSLPPY			
			CGATC		VRAMSELRPFVKALKTWLFGCAFK			
			ATTTCT					
			CCATT					
			AGAAC					
			TATTC					
			CTACC					

				CCTGA TATT					
Pene lope	Pose idon _Ac	.	Anolis carolinensi s		GCTTTTC AATGCTA ATTAAAG TGATTAA CTACACA ACATTCA CACTGAC CTCTCTC ACCCTAG ACTTTCC ACAGATA TATATTA ACCTCTT TGCTTAG TTTTCTCC ATACCTC ACAACCT CTGAGG	MPAIDVGETSGENASGTWPHSPKDIQQPCDPG HESLRQHIEHLYGEKLFQDTRRLEKLRTKRAH LLCSLTFLLRCDTDTIPQFLAAKRTFKTPQAH RIYNRLERNLLRERIHTIRKELANTDKELLHLHI NISQKMNTQDWDKIDSLTYKKMEKDMVLHT KRQKQKFDKCRKQQPKPELDSRTIINLTDRQ LTEDQVSILEKGGNFAVTTTRIPVENIIANVESA IYQLPEEEAEEVRMETARILRNAKLPPSNITRK ERQAIKDLNSDPEIILPADKGNATVIMETKQY KEKIRQLLDPTIYKKLKQDPTNKITRKTNTLIK NSSINFDIRQQLCKSEALPPRLYGLPKIHKDSIP LRPIVSAIGSPTYNLAKFLATQLQTHIGLTAHYI KDSTHFIEKISNLNLSTKDILISFDVVSFLTKVP VADTLTLIKQNFEDITALFHHCLTTSYFQWDT GFYEQKDGVMAGSPLSPVVANFYMEYFEKQA LETAPKKPTVWFRYVDDTFTIWSHGEEELSKF LDHLNSIHPNIQFTMEKEKEGKL PFLDVL VIRK PNQQLGHTVYRKPTHTDRYLHKNSNHHPSQK RSTIKALTDRAQRICEPHLLQGELNHLN WALQ ANGYSTTDIRRAARPRTSHESQDKDPPRGKVF LPYIKGTTDRIGKLMKKHNLQTIYRPTKKIQQ	cd0030 4	470	554

						MLRSAKDKRDPLSSAGVYRIPCSCGQVYIGTT KRSAQTRVKEHERHCRLIQPEKSAIAEHLMNQ PGHRILFENTKMLDHSNNYHVRLHREAIEIHK HVDNFNRRKEETMKMNIWLPVLQNSKIKTVN KKQYSENRGFPMNQPRAVNDSKQRMPQRQE EDSR			
Ving i	Ving i- 2_A car		Anolis carolinensi s	GGGGG ACATG AGAGA AGCCT CCACG CAGGA TCGTG ACACA TCCGG GCGTC CCCTG GGCAA CGTCT TTGTA GACGG CAGAT TCTCTC	TAGTATT AATGCCA AGTTTTT GTTTTTTT TTCTTAA AAAAAAT CTCTATG TTTGCAA ATCCATT ACAACCT GTACCCT CGGTTTC GCTTCTG ACACGAG AAATAAA TAA	MTIISINIEGMSAAKEQLLYELCRDKKCDVLCV QETHRDKQQKRPKVPGMILVAERPHAQYGSA VFVKPGLQVSSVHNTEENNIEVITVELNNITITS VYKPPNEEFCFSSPENFDSHQRTVVIGDFNSHS HVWGYAQEDKNGEAVLSWAELHKLSLIHDSK LPPSYNSGRWHRGYNPDLLFVSDSIVQQCSKS VGPPIPNTQHRPIVCQLFPTIRPLEVRFRKRRNF RKADWSKFSDMLDDKIISVAPIPEEYERFIKIVK TCSRASIPRGCRTHYLQGITPDTAALLEDYYRL HNEDPFSEQTLQTAQSILSSISEAKKEQWIKLIT EVDMMGRSSQKAWRLLRRLSNDPSQTNTHANI KADQIAHQLLKNGKPKLSTKVGRKPIVRQPDN ENNHLHEPFTSTELDMALNKCKNGKAAGLDD LRMEQIKNFGPKARCWLLKMMNNCIASCQIPK IWRKARVIAILKPGKDRNDPKSYRPISLLCHLY KVLRLILNRIMEKIDPCLIQAGFRKGKSCTS QVLNLTQHIEDGFERQQITGAVFIDLSAAYDTV	cd0165 0	459	714

				ACAAC CAGAA GCGAC GTGCA ACTTG CAACA TGCAA GCAAA CAAAC CAACC AACGG ACCTT TTCAG GACCA ACT		NHRLLLRKTYNITKDYHLTRFIGNLLQNRSFFV EFQGQRSRWKQKNGLPQGSVLAPSMFNIYT NDQPLPEGTESFIYADDRAITAQAGSFEMVEQ KLSEALGALTAYYRENQLIPNPSKTQTCAFHL KNRQASRALRITWEGIPLEHCSTPKYLGVTLD RALTYKKHCFTIKQKVGARNNIVRKLTGTTW GSQPDTVKTSALALCYSAAEYACPVWNTSHH VKTVDVALNETCRIITGCLRPTPLEKLYCLAGI APPDICREVAANNERTKALTSPAHLFGYQPA RQRLKSRNSFLRSTEILAGTPQQARVQKWQAK TRNLNQWLTPDERLPPGHTEDWATWKALNRL RSGTTRCRANLKKWGYKVESATCDCGEEQTT DHLLQCILSPATCTMEDLLTATPEALQVASYW AKDI			
CR1	CR1 - 9_A G		Anopheles gambiae	TTGAC AGTTG ACAGC TGACG ACWGA GTGTG TTTTCT CACTG	TAAATGT TCTGTTC GTTATCT AATGTAA TGTATTG TAAACAA ATTTTGA CTCGAGA	MEGVCCDCSVPLESLECGIMCSFCDAVYHSAC TKSPFTVIEEVKRTASLHWSCVGC SNALGNPH SKAVKTTGMQVGFQAALTA AVEVMKAALVP PVMHEIREGFASFAAAHQAPSEHFNVPPPDALP SGKRRRLFRDVVASTEAI VVDNAPLYPVNTDT DNTHQRNRPSLPPIITGTDTTTTSIPTVSQLPRT NYMWLHLSRLAKTVTV DQVVSMVKSQLDTT DVIAFSLLKAGTTISSVSSLTFKVRIIAALRDKA	cd0165 0	909	1158

			TGCTC	GGGCTTC	LTAGSWPVGLGVREFISLPQRPSHHRPDLTNTN		
			CGCGT	ATAGTCC	TPIPVRHSPMHPVQLSPDMNCTLISPAPQKQLQ		
			AAATC	ATCGATT	PHQTPLKCTKLLFNTFFLSNNHTTVSSKGTVP		
			CGTGA	AATAAAC	NLTSHLFTRNNNTSTTRSLTDDLAPTYTRIPLN		
			ATTAT	TAAACTA	VRSTSITVYYYQNVRLRSKAEDEFRLTVLEADY		
			TTTGTC	AACTAAA	DVVVLTTETWLDPSLPTALLFGDAHCVYRCDR		
			GTTTC	CTAAA	NAANSSLSRGGGVLIIVSSALRSHALPVL AQT		
			GGCCT		LEFACICIQLPTYRLYIAAAYLPPNHSTDEGKIN		
			WATTT		ALLDTVNSICDTLGPNDRFVLVGDFNQPALSW		
			TCGTC		SPAQSNHETPFVFFEPHTSSVRS AQFVDGLHQN		
			AAACT		ALYQLNNNTNSMGRILDLAYGNWSSAATSSLI		
			GCCGC		RVSEHPLLPIDCYHPPLEFDLEITASSMNHATT		
			TTAGC		VTEIKLNYARADLTKLKR LICSFNQSFECSNYT		
			TGTGA		SIDHATNDFSEFMRAALRECAPIKKPHRGPPW		
			CTTAA		GDRTLHALKKAKRAAYDYLRANRSTPCRLYY		
			GCTGT		NGAHALYRRYNRV CYRRYIRQTERSLRKYPR		
			CCCTG		RFWSFADNMRKSTGLPGTIRYKDNCAHSPLDI		
			GTAGT		CELFATRFEDSFISNITPPEAVASALVNTPADAI		
			CCTGC		HVTLPIVTETLVTQKIERLKSSYSEGPDGIPATI		
			CTTCA		LKRC AVSVAPVLTKIFNESLRTGTFFPSLWKVS		
			ATCGT		WLTPIHKKGSKNEAVNYRGITSLAACAKVFEL		
			TCGTC		IIEPLMASAGKYISTNQHGFM PKRSTTTNLM		
			ACAAC		QFVSGCYKSIDDGMQVDAIYTDIKA AFDSVSH		

				CATTG CTGAC TTTGCT GTATT TCCTG CGTTT GTTCC TGAGT CCTGT GGCAT CTATC TGCAA CGGCA ACTAA A	NILLAKLDRLGLSPPLVKWMKSYLCGRSYAVR MGSHQSRSIGASSGVPQGSNLGPLLFLLYISDL CTALPDNSCLLYADDAKIYREIREPEDHLQLQ ATLTEFVSWCKRNALSVCIDKCAIISFSRSRAP ALFNYTLDGQNLERVNCVKDLGVLLDAKLSF EEQIDQVVASGNRLLGLVINMTRELDPMPCK TLYCALIRPLLEYASIVWWPASTRALARLESIQ RKATRFALRDWPRRLDYRTRCLLLGIPPLAER VEHTRLAFITGILNGTIDCPELLSRIHLYVPARIL RRRPMLAVAETRTTFGSRNPLLCMCRLLSAN DIYEPGMTITELTSLLSVRNAFNNNNI			
CR1	CR1 - 10_ AG	NT_0 7826 8.2	Anopheles gambiae	TCCTTTAT AATCCTT TTCCCTG ATACTCT TCCCTAT CCTCTTTT TTCTCAT CCTCTTTT	SPRIPRIFGLQCSPGFSSTTAAIPSQAINRCDPR TRITSVYPYRFSAANFPPVDEPAATTAQSNSTS TPITPLHHAERFETLPSVASPITVSAPLPLHAS FPHPIDDLRFYYQNTRGLRTKLDDLRLASADT EYHVLILTETWLSNIPSSLAFDNNFVIYRCDR NSRNSSRSRGGGTIAVSRKLAACEISPASDSL EQCWIRLKLPSCSLFIGSIYLPPTLSADASTLNT LDASIDEIISRMKKNDRLLLFGDFNIPSITWHAS	cd0165 0	634	882

				TTCCGAC	PDPAPPFLPLLVPASSAHSSSLFIDIVRYHSLSQ			
				TATCATA	LSSVSNFRQVQLDLIFTFPKNQSNATSTCTYVS			
				TCTAATA	QAPVPILQIDPHHPPELSLKITGSLTPSNRSQH			
				TTATTCA	SSSRLTTAPSLDFRKTDFNLLNTSLSNIDWSQL			
				CTCTTCTT	YSCKDVNLAVSTFTSLFLSTLPQCCPKRNPPRS			
				TTCTTAC	PPWSDATLKSLKRDKSRYQRAYWSTRSVNAL			
				GTTTGTA	RTYKYAAAAYRLYNRTRYNGYILILQTRLRW			
				TTTTATCC	YPRAFWKFMNARRKGSSLPSTMHLESVYASSP			
				TAATATC	EDICSLFRNHFSNIFSPPSQIPNTASQCLSYTPQD			
				TGTCTTTC	VIDVSCFGFDEHSVSEALSKLKPSFSAGPDGIPP			
				ATGTGCT	SLLKKTAEVIAPLLSFIFNLSLRLCCFPSSWKSG			
				CAAAAGT	WLCPIFKKGNSADVSNYRGIVLLPACSKVFEIA			
				TCGCTTT	VHSVISFQVKSWISVNQHGFMPSRSTVTNLMQ			
				AGTATTA	FASHVIPSLDSGLQVDTIYTDFKAAFDTVPHHL			
				GGAATTT	LLLKLKKLGF SNMLIDWLYSFLNSRSIRVKTD			
				AGATAGG	TFSSSFCPSSGVPQGSVLSPLLFLVLFVSDVHAIL			
				TTAAGTT	PSEGHLLYADDLKIYLTISTHSDCLRLQEFLNSF			
				AGGAATT	SLWCSYNQLSLCPEKCSVISFCHNRNPIHYEYS			
				AGTTAGA	LSNSVLSRTSLIRDLGVLWDDKLSLNEHLESV			
				TTTAGGA	VASASRTLGLICRLTRDFRDPLCLKSLYCSLVR			
				ATTGTGA	STLEYASVVVGVPVRSGWILRFEKVQKRFRFA			
				AACCATC	FRRMAGPNAAMPSEYESRCNLLGLESLEVRRSQ			
				AGGCTGA	ARVLFIGRMLNGLIDAPTLLEKIGLYAPSRNLR			

					CAATTTG AATAAAC TTGAAAT GAAATGA AA	ARSALNMVSRQTTFGSNEPLLLCVRTFNTSTD PPHPRL			
CR1	CR1 - 8_A G		Anopheles gambiae	TATAT ATCTT CAATC GAAAG GCATC CCCGA AAATT TTGGC GCTTC ATGGA CAGCA GACGA CGCTC CAAGC ACAAT CATGG GCACG TCACC	TAATTTT ATTTTCA AAATACA TTTTCTAT TCTTGTG CACCTGT TCTCCCC TATCGAT TTATGTA CCTTACA TAGTCTA TAAGCAC TAATTTA AGTACAT TCATGTA GGATCTT CAAGATC CCGATGG	MPSPIVDQTVISRATSYTPLNMIDYTIPTFDIRT VEKCLKSLKPSTAPGPDGIPSYIISKCYQSLAPV LTNIYNTSLSSGIFPRLWKTSWMVPIFKKGDORS CASNYRGITSLCAVSKAFELLYGPMLSATSN YIMDNQHG FV PKRSTLTNLTEFVSFCKRNLDS GGQVD AIYTDLKAAFDLISHDILIAKLQKLGF S EQIVKWLHSYLT ERSYKIKVNDHTSREVLGTS GVPQGSNLGPLLFI LYINDVGQLLAEHRFLLFA DDVKLFAPINDTND CINLQHAINLFSIWCRDNA MVL CIEKCRVLSFYRSR SCTR FNYQIDNISVER TDTFRDLGIILDTRLTFNDHLENIVSRGNQLLG WIIRSTQGFRNPMTIKTIYCA YVRSVLEYGSIVS DPCTEQWSNR IEAIQRKATRYAVRLLPWRQGD VLPSYHSRCLLLGIQSLKKRREIAKCLFISGLLN HHIDAPTLLATIEFNAPS RNLRTRQFISVPRYRT RFGQSDPMSAMC RTFINNFDFDNIPQNVFR DRLRTQLT	cd0165 0	92	340

				CGAAG ATATT TGCAA CATTTT CGCTG AACGA TTTTCC G	ATTAATC AATAAAC ATATTAA ACATATT AAACA				
RTE	RTE - 1_A C		Aplysia californica	TTAGC CGTGC AGTCT CGGCA TTGAG AAGAA TAGGA CTGAC GTGGT GCTAG CGTGT CGTAC GAGGC GACTA AATGG	TGATGAT GATGAT	MMQGGESRKEAAGQKPPSSSKSKSLRIGSWN VRTLYEAGKTAQVLQEMHRYQLHILGISETHL NQSGQKRMKEGELFVYSGNDSGDPHREGVG MIISKTAQKTFRGWEAHGPRIIMASFGTKKELK VRDGRKKVINMNIIQVYAPTNDAEVENKDDF YNRLQNVIDKLPGKDINIVMGDLNAKVGSDNS HNEEIMGRHGLGEANDNGERFQNFCTFNNLVI GGTIFPHKRIHKATWVSPDGTTENQIDHFCISR KFRRSLEDVRVLRGADVASDHHLILGKMKLK LRRYAQVAAGNRLKYEVSLLKDPVKKNEFKL ELKNRFQALQDLDECEVNEHWDKLREAVNAT CQSTLGPKDNTRNEWISQDSLDKVKRRRELKA QVNNSRTRTEKLAAQQEYTEVNREVKNSIKK DKNTYLNGLAEKAEKAATDGHRLRVHQTTRT LSGRFSKPTLPVKDKKEGKTIFNPEGQLNRWRE	cd0165 0	562	819

				CGAAC	HFDQLLNRPPEQPEILPARNDLPIDTEPPSRE			
				CAGGT	EIAAAIKTLKSHKAAGPDHIPPEALKAEKDTFV			
				AGTGA	EILYALFEKVWQKEEIPSDWKEGHLIKLPKKG			
				CTCCT	DLSNCNNYRGITLLSIPGKVFNRIILLERIKLATD			
				GAAAC	EKL RGNQAGFRKNRSCVDQIATLR IIVEQSLE			
				AGAGT	WNSPLVVNFIDYEKAFDSVDRETLWKL MRHF			
				CATCC	GIPEKIVKLVKATYDGTSCQVFHEGQLSEPFRV			
				TGCCT	QTGVRQGCLLSPFLFIL AIDWVMKEATR GKRN			
				CCAGG	GIQWIPWQQ LDD LDFADDLALLSHTHAQM QE			
				TCGGG	KTKELNSLSRSVGLRIHPGKSKVVKL GK VSNL			
				GGTTG	AVTVEDTPLEEVD SFTYLGSVIDGKGGTEADV			
				AGTCA	KARTAKARSAFKQLSKVWKASNISLRTKIRLF			
				CGAGG	NSNVKSVLLYGCETWKTTKTVLNKVQTFINRC			
				CTAAC	MRSILRIKWQDRIRNEELWERTGQKDIEVEIGL			
				AACCT	RKWRWMGHTLRKPSGNITRQSLQWNPQGAR			
				CTCTC	GRGRPRETWRRRCVERDMREMKL TWNELGRM			
				GTAAA	AQDRDKWKELVCGLYPARGERH			
				AAGAA				
				CCTGT				
				TACAG				
				AAACA				
				CATAC				
				AAGCA				

				AAAAC CATT TCTCT GCTTG GCTGA TCGTG GATCT TCTCC AAATG GAGAG TTT					
CR1	CR1 - 1_A C		Aplysia californica	AATTG AACAA ATTAT GGAGG AACTG GGAGT GACTC GCAAA CCCTTT GATAT CAGGA GATTA	TGATTTG TCTAGCC TGCACGC AATGCAT TATTGTC GTTCATT ACCAGAT AGTTGAA CAGGCTC TACTTAG AGCCTCA AACAACT	MSLKSMVEIYQPDIIAVTEVKPKYVRFKIQESE LSLDGYDYELFHNLEEEGRGLVLYVRKALKPG LCDDLDNDFCEKRFVKCVLKDGVSLIGLAIIY RSQHSPTAEEDARKLNKLFELVSLRKETHKIIM GDFNFRQIDWETETSNAGEKHIATKFLKATKD AFFIQHQKTPTRYRKGETPNILDLVFTNTEQLV TDIKVEAGIGKSDHLCLLMTLSLSAGTAPKVT RRNFRKTDVEKLREDLGAYDWESELANKNVE DTWSVIKSKIEGAIGASTPLMNISGRKNKRYM DKDTLEAVRKKHRAIRKWHRTGKTTDEIAYN KANNKARKACRKANRNYEKKLAIEAKKNPKV FYSYANSKISKGTGIADLNKDDGKTKSDTEK	cd0165 0	479	737

			TTTGG	AAATCGT	AELLNKFFQSIFTNEDTGPLPDFEYDYNTELQ			
			TAGGA	CAATGTC	DFNIETVTVKELLVSLDVNKATGPDGIPPRVLS			
			ACAGA	AATGTCA	QAADILALPLRILFRRTLNDGVLPVEWKTAFVS			
			ACAGA	A	PIFKKGNKAAVNNYRPVSLTCVICKVMEKLVR			
			GGATA		GHIMDHIVNNMLISKDQHGFVTGRSCITQLLE			
			GCAGA		ALDDWTSVMDEGGSVDVIYMDYQKAFDSVP			
			GACAG		HRRLIQKLSACGVKGRVLKWVRDFLTSRKQR			
			CGGTA		VVINGSRSEEA EVTSGIPQGSVLGPLL FVIYIND			
			ATGAC		LPRGLKTTAKMFADDTKLYVRSDTVNGPQDL			
			CGGGG		QDDLDTLQEWSQRWLLRFHPQKCCVLKVGK			
			AAGCG		DNDNEYFMDQAVDGSTSRHKLNVTKREKDLG			
			AGAAA		VVVDNKLTFKEHVAQMTSKANRIVGLIRRSFD			
			AGGCC		YLSSEDMFVLLYKSLVRPLLEYAHC VWQPFEK			
			AAGAC		GLCSEIEQVQRRATKMLSHLKEKPYSERLKKL			
			CTACA		KLPSLEHRRLRGDMIETYKYIKGY YITERPAFT			
			CTACT		KATTTQLRGNSEKLQKQRSRLCVRANFLENRV			
			GAAAT		NTIWNNLPESVVSAPNVDTFKRRLDEYWKDLP			
			TTGAA		TIYNPTCQN			
			ACTGA					
			AGAAG					
			CAAGT					
			GCTAA					
			TACGA					

				TGAGG					
				AAATT					
				TATTG					
				AGGCA					
				AGAAA					
				AGACG					
				CAGCA					
				AAGGA					
				CAAAA					
				ATGAA					
				GGAGA					
				AGAGG					
				ACAAG					
				TTATA					
				CTTGA					
				CTATT					
				GGGAT					
				GCGAA					
				GAGAT					
				ATKAC					
				AAAGC					
				AGGAG					
				AGAGA					

				GGAGG					
				ACTAT					
				GCCCT					
				CTACC					
				AGGAG					
				TTGAA					
				GGCTA					
				AGAAA					
				GAGAG					
				GTAAA					
				AAAAC					
				TTGGG					
				GGACA					
				ATTTC					
				GCTCA					
				CTGGG					
				TGAGG					
				AGAGA					
				CGGCA					
				AAATT					
				CGGAA					
				CATTG					
				GGAAA					

				TACCC					
				AACGG					
				AACGT					
				CAGAA					
				AGAGA					
				GGGAG					
				ACAAT					
				ACAAG					
				AATAA					
				AGCCA					
				AAGAC					
				AGCAA					
				AGAAT					
				GTAAA					
				TAGTT					
				TTGGC					
				CATTG					
				TTCGC					
				CTTTTT					
				CATAT					
				TTTAA					
				TTATT					
				GCTCA					

				TCTCG					
				TTCCT					
				GACGT					
				WCAGA					
				ATAGA					
				AGTTT					
				CACTT					
				CATTTT					
				CTTCTT					
				TTTCA					
				CCTAC					
				ATTTC					
				ATTCC					
				AATCG					
				GAATG					
				CACAT					
				ACTTT					
				AGAGA					
				ATTGT					
				ACTTG					
				CTTGT					
				ATACA					
				AATGC					

				AGATC AATTG TTGAA CAAAA GAA					
Tx1	Tx1- 1_A C		Aplysia californica	TCAAG CGACA ATCCG TGAGG CTAAC GCTGA CACAC AGGCG GAGAC ACAGG CAGAG ACACA CACCC TGCGT GCCGT AAGGC AAACA AATAT	G	MSIRIASLNCRGLNKKLKRRAIFRKCTQYDIICL QETYVTNTTAREWSLEWSGDDFFYSEG TNKSK GQIILVNKHFKYKTVELCIKTERILGLSFSXDD VSYVAINVYGPVVKNDKLD FIDKLYSVLKDEK ADHTIICGDFNVVHNNTLDIVTGEKHDDMSTQ KFRDWIANCQIVDTWRGVHGEEKDFTWSRPT PFTARRLDYIFCNTELPFVIDSKHAFVYGTDH KAVITTVRTEKFKRGPSYWKFNN SLLKDKEYV NSMEKCIDEYLERTLDETDKTIRWELLKVMIK SHTIDYCIKRCRALKLREN LITEELLKLNDQVV KNPQDLTLLKQFESKKKELEIYQLHKARGAQI RSRQKWIEEGEKNTKYFLRMEQMRGGSNVITS LQDDKNNNLYGHVDIVDSIKNHFERLYKKDH NIETNRVDDFMQGCSPVLDDVDKEICDSELSI VELDAIVNKLNIDSAPGIDGITTPFYKFFWSKL RKPLFTNFQQCLETGELSPTQRRGVISLLHKGK NLRRDNIKNWRPITLTNTDYKIFSKILAIRLQRV MPVLINENQTGFLKGRSISEHIRTLD DIIYVASE	cd0165 0	507	772

			GGACT	KNLPGMVVSLDFQRAFDTVKEGTILATLKKFN			
			CTTTTT	FGEGFINMVGVLMANTECCVQNGGWLSEWL			
			TCCCG	NTERGIKQGCCTSPLLFVLVAELMSIKLRNNS			
			CAGCC	ISGMSTILSGEIINFCKVLQYADDTTLILKDQED			
			TGACC	LASALLDIDSFGSISGLKLNKDKSQGMWIGSDR			
			GCTCC	TRTEKLGGISWVPNGNTIKILGIHFSATEEASRI			
			CGCCC	ELNWIDKIENMKNIMKMWMKRNPSLYGKVII			
			CCCAG	CKTFLLSQISYIIQSLSLPERVLNTIDTLFFKFLW			
			TACGC	QKRHSNKKAREKVKRTVMCRSIEEGGLGVISV			
			TCACC	KDQQKVFLLNWLKRMATVTEGRPNNIPNRISD			
			CCGAA	LAFTEVGGCDYFFSATANCKKTGKLCHASKF			
			GCCAG	WESVMDTWIEMASIKKEPSSTIEILMEPLFNNR			
			GACAC	NIKYKGKSLYLKRWMNSGIKYVYHLYDRRIL			
			AAGGT	KSKEELSHITGTAAFDYALMNALPTQW			
			CTCCA	KQDVSDNTEELLVNTREVINTNGKKTVLEIL			
			AGTGT	SKSNREIRQLLKYRRHAEICGRNFWMRKMNI			
			CAAGA	ITCKYNIAHRSTKESRLRLHFKILHNIFPSNILL			
			CGAGG	NRLGIKNTELCEHCGVKDVIEHMFHCTLKKNF			
			AGACA	WKYVFHTICLKTSSRFPMTDCNILFGFSERGQG			
			TAAAA				
			GACAG				
			GTGAG				
			CGACG				

				AGCCA					
				GCTCG					
				GTCAC					
				CGAGC					
				GTGAA					
				ANCTC					
				GAAAA					
				CAGGA					
				GGTGA					
				GTGGC					
				ACGAG					
				AAGAA					
				GGCAA					
				GACAC					
				GGTAG					
				ATTAT					
				TTTGA					
				TTATG					
				AGGAG					
				CCAGT					
				GACTA					
				CCGAA					
				ACGTG					

				TGAGC					
				ACGAA					
				GAGGT					
				CGGTG					
				CGTGA					
				GTGGT					
				GTGTA					
				CACGT					
				TTCGG					
				TAGCC					
				ATTGG					
				TAACG					
				ACTTG					
				GACTG					
				GTAAA					
				AGGAC					
				TTTGA					
				GTTTG					
				GTTTT					
				GCGTC					
				TGACT					
				GTTTT					
				GG					

Hero	HER O- 1_A C	Aplysia californica	GTAGA	TAAAGTT	MDNDLSVKRTMQKENQENSFMCEQCYRKFD	cd0165 0	355	614
			GAGGA	GGCAGCA	TWRGLRVHQGKVCLKKRRQCRSSDCKTRSKS			
			AGAGC	GAGGAAA	SQDENHSGAISASVDSPTLSQQEMEQQTEPGL			
			CCGAA	ACAGTCG	RKPRILWPAANEKAKYKLEEEVQKKLKEVK			
			ACAGC	AACTAAC	KSSNRGDAKEILSAFATIIYDSAIKEFGICEGKK			
			GACGC	ATGGAGA	KTEKKSGLSRRQRVLAQLRKQKRDLDQKQKKY			
			ATTCT	CCTCACA	APPEEIEGLRILFEDLKKKSRDLQRHERRCTRR			
			CTCCT	TCGGTGA	KERKREREQFLKNPYEAAKKLFTEARSGKLEC			
			GATGA	GGGGGCT	TKEELDAYVKQTYSDSRREEPLPYMRGLKRPA			
			TCTCC	CGGTGCG	APVVKYYLGPIREKEVDEFVKKARAKSAPGG			
			ATGCT	TCAGAAA	DGVSYSKVFKYCDTLRHTLFEMLRGLWKDKEI			
			AGAGT	AGCTGTA	VKDWCRATEGVYLPKEQNAKEIGQFRPISIINVT			
			CCGGG	GAGAGGA	CKIFMGILAKRTVAYLQSNQYVDETVQKAGV			
			ATACG	AGAGCCC	PGIPGCIEHAFTIWDTIQEAKKNKESLNVVWLD			
			GAGCG	GAAACAG	LANAYGSVPHKLLMKAMDFFYVPQEVQNIMR			
			TTCGG	CGACGCA	EYYSNFMRFSTANFTTEWHRLEIGIAAGCSIS			
			CGGTG	TTCTCTCC	VIWFILVMEMLLSAECCEEEAKVRSPKKAFM			
			AGCAG	TGATGAT	DDVTLLTRDGDTMQSVLNRLDKLITWARMKF			
			TGCAT	CTCCATG	KAKKSRLTINKGKQKQLKFTIAGEQMPTVKE			
			ACGGC	CTAGAGT	QPVKSLGRWYSGTSDRSQGVEIMKQAENGL			
			TGTTT	CCGGGAT	KAIDQTKLPKGHKIWCLQFALYPRLAWPLTM			
			AGACT	ACGGAGC	YEVALSRVEMIERKCNSYIRKWLGLPRTINTSS			
			ACGTA	GTTCGAC	LYRHRGALQLPLTSIVEIYKAGKIRTVLMLRES			

			CCTAC	GGTGAGC	RDKEISGKPPDVKTARKWKA EVAANEIISALE			
			CAGTA	AGTGCAT	HGDIVGPAQPDRLGLGNGDFRPFRKMSQRDR			
			GCAGG	ACGGCTG	RKAATRHMQKMEAERREIHLIQCSQQGQMTR			
			TGTAT	TTTAGAC	WEEHVVERKVGWSDIWKWNTSRLSFLLRSTY			
			GCTAC	TACGTAC	DVLPSPVNLVRWKVREDNKCRCGKLGTMKHI			
			GGACA		LSNCNLALNRYTWRHNEVLKVLTKMATQQV			
			TCCTG		EEGKYAPKIQKHGPTKIAFVPQGSKVAKRKA			
			GAAAG		NEAMKDQSGGTWEVAADLKGFERFFPIPTTKK			
			ACAGC		PDIVLWEEKDKEVHLVELTVPHE DNIYAAHER			
			ACGAC		KENRYEALVEQCEEAGWKATHFPVEVGCRGF			
			AACCC		VATSARKWMRVAGLTAKKSNELTNSLQETA E			
			GGAGT		KASHWIWLKREDSSWNE S			
			GACGG					
			GTGGA					
			CATCG					
			GGAAA					
			GACCG					
			AGGAC					
			GTCGG					
			GAAAG					
			ACCGA					
			CGGTG					
			GCAGG					

				GGCTA					
				ACTAC					
				CCAGG					
				TCACT					
				GCAAA					
				TNTAA					
				GTTAC					
				GTCAA					
				CCANT					
				AGAAA					
				AGACA					
				GCCCC					
				GGGAT					
				CCCCC					
				GAGAG					
				GGGGG					
				AGTAG					
				AGGAT					
				GGATC					
				CACAA					
				GGACG					
				GATTC					
				A					

Ingi	Ingi-6_A C	Aplysia californica	CCCC TAGAGAC MLFTNDRIPYELRRRLGNQGLPQVLVGNQAGL CCCAA ACTTGAG VDGYWKKKKIRKQRKWKNKKKTPPTTGNGK ATAGA AGTGGGA EKLIDFNILQANVCGLDKKKTQLGKILYERKV CCCAC CTGATGT HIALLQETLHSSCDINITGYTPYPCKCEGCQGII CTCTTC GAACAAA TYIRNDIQGEVELLNWHPTDAHKVTLWYGNS GTGGT GATATAT KLILYNIYCPPSSTLNFLEQSTIFKKTVIAGDLN GGGGG TGCACTT GHSPLWGYADQNNTGKRIEELCXSSNLILLQD GGGGG GATTTGG EDSTKTLLHRRHGTLHRPDLTLVSADLHNYCS ATTAG AGTCTAC EEVLDCISSDHRPVLTKISILKHRQRKRRTRWN TACTC ATAATTT FKKANWDLFRKTAEDTVKLDDLDNKSIDALN TTCGA GATAAGT SELTSTILDAAKKSIPQGCRTNYKPFWNERLEK ATAGG AGGCACC ATKTKNDARHEYEKDLKNMDKKIRYKKA AAAAG ACGTCAG EAKLVTKESKKEAWTQKCEGLNLHQGGREA AAAAA AGTAGTC WSLNNLTGEKRKENPKPLHTETEVLTSELKK GAGGA CTACAAT AEHFNKHFASINKSSKRNDLDRGLLSVLKGKE AGCTA GGATTCC RDLMTSPSIFQDFLT ACGGG TGCCCAC DKVHNEMQLNLGEKAKLLLLKLYNRTWTEG TCAAG ACTGGGA NLPKAWKLATINPILKKGKKANDPKNYRPISL GGGTG TACCGGG TSCIGKLCERILNSRLYWWLESSGLISQCQAGF ACCCC AAATGAA RSKGRTEQDLFRLTQKIIDGFQRGEQTTAVFVD GAAAA AGAGCTG LQQAYDRVWRKGLLLKLQNMGIKGNMYSWI GAGAC ATGCTCT KSFLTDRLIQTQMDETLSSKAVLEEGLPQGSSI ATATT GGCAAAG SCTLFLVFLNDVSDVLKTEKALFADDLVLWHT	cd0165 0	550	806
------	---------------	------------------------	---	-------------	-----	-----

			ACTTT	AAAGGAG	GNSTTISRRLQDDLNSLGTYCRLWKMKVNC			
			ACTAT	CTAGCCT	EKTVYSVFTRSYNLANAIEPLKIDGTNLKKEN			
			GTTTCT	CCCTCAG	NPVYLGVLQDSKLNKQHTNNLRKKAMKRL			
			TACTC	CCTGAAA	NLIKRLASTNWGSDKNTLRGLYLG YTRAVFD			
			TTTCTA	TTCCTAC	YNLVLQNICSKTTKNSMDTIQNQALRLISGGM			
			TTTATT	GAATTAC	RSSPTAAACEIHTNIEPLETRRKRAALEVYERSK			
			GTTTTT	ACCACCT	RLENNHPNRVLVDKWKPNQRLKNTKSVLDTV			
			ATTGT	GCTGCCA	LNLQEKHHLPERREPLERVPSTLPPHLTIKPEI			
			TTTACT	AATGATC	RKTLLDKSNKNTYPVALKHSALETIDSYPTWI			
			TTATT	AAGTCAA	HSYTDGSAFKATINAGYGSVINLP SGIKKKKSL			
			ACATT	ACCTCAA	TPAGPSAPISLPNNRLSPTQQTSTTCLTLTHLL			
			TATTA	GGAAGAG	PQTSLSLQTAFAQH			
			ACTTTT	TGGATGA				
			CTTTTC	ATGATTG				
			TCCTCT	GGCCACA				
			CTCCA	GGTACGA				
			TCGAC	CAGGAAG				
			AACCT	GAGGATG				
			TTCTTT	TATGGCC				
			GACCT	ACATGAA				
			GGAGT	TAAACCA				
			TGGGA	CTGGTCA				
			CGGGT	AAGATCC				

				ATTGG	AATCAAT				
				TGGGG	GAGTTGA				
				TGGAA	GGAGGGG				
				AGCAG	AGATCAA				
				AGGTT	TCCATAA				
				CATTG	TCTTCCA				
				TCAGC	ACTTAGA				
				ACTGT	ACACAAC				
				ACTGG	ATGTACC				
				ATGTG	CCTCAAC				
				ACTTT	CAACATC				
				ACGTA	TCAACCG				
				GGGAA	AATTGGA				
				TACCC	GTAAAGC				
				AGCTA	AGTCTGC				
				CATAC	TGCCTGC				
				ATGTA	CCACTAT				
				TTAAG	GTGACTA				
				CATGC	CCCCTCA				
				ACTTA	GAGACAG				
				TCGGA	TGGAACA				
				GAGCC	TCATCTG				
				CTATG	TTCTTCTG				

				ACTGC	TAGAAAA				
				GGAGA	CTAACTG				
				AGGTG	ATTTAAG				
				CAGAG	GGGATGT				
				ACCAT	TTTCTTCC				
				TTGCG	CAAGCTA				
				TCTTG	CCCCACA				
				CTTGTT	CATCTAA				
				GACAA	TTGCCTC				
				CATTA	TACTCCT				
				ACCCT	CAAGACA				
				CTGGG	ACAATTA				
				ATCTG	GAACAAA				
				CCATG	CATGCAA				
				CCTTC	CTATTAC				
				CCCTG	CGAATGG				
				CCCCC	CATCATG				
				ATCTC	CAGAAGA				
				TCGGA	GTTAACG				
				TGAGA	CTCAGAT				
				GGACT	GCTACTG				
				TGGTT	GATCGCT				
				GCTCT					

				GGGAG	AACAACA				
				GGGGG	ACAACAA				
				GGGGG					
				GGAAT					
				CTAGC					
				TTCCA					
				ATTGG					
				TATTA					
				TTATT					
				ACTAT					
				TAGTA					
				TTATT					
				ATTAT					
				TATAT					
				TATAA					
				TTTTTG					
				TATTTC					
				AGCAA					
				CACAC					
				AGCAA					
				CTATG					
				CTCTT					
				AATAG					

				CGTCT					
				TTCTCT					
				TCTTTA					
				TCTCCT					
				CAAAC					
				AACTT					
				CTCAA					
				TATAA					
				TACAA					
				CAGAC					
				ACCAA					
				CATAC					
				CCATT					
				CATTTT					
				TCTTA					
				CAATC					
				TCTCT					
				GTGAT					
				GGAGT					
				TGGGT					
				TTTTGT					
				TTGTG					
				GCATT					

				GCACC TACAG CCACT GCTTT ATCTC CTACC CCGGG ATGGA AGGAA AATCA CAGAA CCACC AGTTT TAG					
RTE	RTE - 6_A C		Aplysia californica	GTAAA GCTGG GGCGA CCATC CCCGA CCCAC CGCGT CCACA GGTGG	TGATGAT GAT	MDHITSINKDXDVKGSSERTASSGRVSHAQQV TPDRHQDTAHKRKNQTRNIATWNVRTXLEKG KLDNIKKEMKRMNVNIMGLCEVRWTGAGRIL SDSYNIIYSGGDKHERGVAVIMDMPTSKTLKG YLTISDRVMLVRLKGSPFDTTIIQAYAPTSESTE EEIEKFYDEINQAKTHCKSQDIVIVMGDFNAK VGATRHEDIVGPYGLGXRNERGEKLIWAILN DMIIGNTWYEQPPRRLWTWKSPGDSARNQID YIMIKKRYKNGMLNVKTRPGADCYS DHVPVV	cd0165 0	549	806

			CGGAT	GKFRIRLKKTHMSKNKNMMLNALLQSNKE			
			AGTGG	MQDKFRITVQNKFGPLEEIQDIEKQWEQLKTSI			
			AACGG	NEAAKDVIPERTTKAKKKWMTEEILNLMEER			
			CCTAC	RTNKNNXEKYVNLHKEIRHKCNLAKETWLKE			
			ATTAT	ECEEMENHSNANSKDLHRKIKEVCHTRTRGST			
			GTAGG	GCLKSKDGNIMDKDKILERWTEYIGDLYNDK			
			GTAGC	RDEQRRVENNNEGPPIMKEEVIKALRKMKNNG			
			TGCGA	KAAGPDNITTEMIKALEDFGITTTLKLLNNIYA			
			ATAAC	TGHIPSDLSKSIFIALPKKPNANECETHRTISLM			
			CAATA	SHITKILLRIIMMRIRNKILPEIAQEQC GFIEGKG			
			AGCAG	TVNAIYVIRTTIERALEVQKDLHLCFIDYTKAF			
			TCGCG	DKVRHDKIIDMLEVLNIDGKDLRIIKNMYWDQ			
			GACCA	TAAMRIENDISEYIQIKRGVRQGCVLSPDLFSL			
			ACTGG	YSENIMRHIEEMPGISIGGHNNLNRYADDTVLI			
			CGACA	AEDDEHLQNLLNVIVQESDNMGLSLNIKKTET			
			GTCTT	MVISRXTPIPTCNIKINDTTLKQVQHFKYLGTHI			
			GGAGA	SSDGRCHTEIKSRICQAKATFHKMKQILTNNL			
			GCGAG	SLETRKRVLKS YIEPILLYGCEAWTINKQLERTI			
			GTGGG	VATELWFIRMLRIPWTARKTNEEVLVEANYR			
			ACAGG	KQLITNIRKRQANFFGHTMRKEALENIVTTGKF			
			GTGGT	DGKRGRGRQREKILDKMTTWLGAARTTDTIA			
			GGCAA	AXKDRXKWRSMAHASRQGTG			
			CCCTG				

				TCTCT					
				ATAAA					
				ATACC					
				TCATG					
				TGCTC					
				AGTAA					
				AGAGA					
				CTGTC					
				AGGCG					
				AGAGG					
				GAGGC					
				GCCGA					
				AACCT					
				GCCAC					
				CTGCC					
				CCCAC					
				TGCCT					
				TGTGG					
				GTTAG					
				TCCTTT					
				TGGAC					
				AAAGG					
				CTAGC					

				GGAGT AAACC GTACA AAAAT CCAGC AGAAG GCAAG GGGAA ACCAC TGCTG TATTCT TTCCCT AGAAC ATT					
Ingi	Ingi- 7_A C	.	Aplysia californica	GGGCT GTAAG GGTTG AGGGT GAAAC CCCAG ACGCT CTCTG GCAAG	TAAAGGG ATTTGAA ACAACA ACAA	MHTSLHNPXTNKMPLRQMKLYELGCRLGYQ GPRSLPSPGGGSDKSATGHTAGIGATSNTQTL NILQLNISSIQNKKMELAKTLSDLRIHVALLQE TQLPKRIELHLTGYTMYHCKCTKCQGIATLIRN DITAEVSHPETDDETDMQEVXVQQSDGTTYTI YNVYCPPRXTLQLNFPNTDCHKHTILAGDFNGH SPLWGYSDDXTGTVLEELLRSSNLISLQNDKT PPTLLHRASGTLSRPDLTLVSADLHPHCSMTVL DDIGSDHRPILLQISSNRKREPTNPIRRSWNYK	cd0165 0	545	802

			CCAGC	KANWELYQQKSESRLLDVCTTNPIDECYKDIT		
			CTCAA	TALKTSAKEAIPRGNPKKFRPYWDKELEQLTK		
			CCAAG	ARKKARRRAEKNPTAENRKHYNLYTAKVKLK		
			TCAAC	TKSAKRQTWREHCGELDLRKDGRKAWNLLT		
			TCCCC	NLSGTKRRKNPAPMKPTEKGETCRAEVMNKY		
			GTCCT	FASINKKGNQNKLSHAIKQLTKRAERTPKANL		
			TGTCG	EVFEKEFTLTTELKAALKCKKRRKAPGPDHVTN		
			AGCAG	EMILNLTDSSLSILLGFFNRTWNEGQLPTEWKK		
			CCTGG	ANITPILKKDKPASQPSSYRPISLTSCLGKVVER		
			GGACC	MVNDRLYWWMESTGTISTSQAGFRKGKRTM		
			CCTGC	DHLIRFAQKTSDAFQDKDNVVAVFIDLRQAYD		
			TCTGT	RVWRQGLFYKMQKLGIAGRMYSWIKNFLSCR		
			GCATA	TIQTSCDSKISSQRTLEEGLPQGSALSCTLFLLF		
			GGAGC	MNDLPDVISCQKALYADDLLXWSKGSDMRAL		
			AGGGG	NQAINADLTTISSYCQLWKMEINTDKTVFGLFS		
			TTGCA	LKNKVTASQVNLKINGTTIRHEDYPKYLGVQL		
			GACGG	DRKLSLKEHLGELTKKATSRLNLLKHLSSLNW		
			AGCTC	GADKATLRQLYIGYVRSVIDYSLPLQSIASQSS		
			TGGCC	LAQLDRVQNQALRFVSGAMKTTPTSACEIHTN		
			AGGAG	IEPLDLRRKSATLETYERFVRLEDTNENRKMA		
			GTACC	DEWEPRTRIQKTSFMREATKLLEEFDPPEENREP		
			ATAGC	ITNISSEPPYWQTPEITIRTRLLDPSCDKTTPEEI		
			CCTGT	LRTSALETVDSYAETPIHAYTDGSATNATRNA		

				GAGGT	GYGSTIKYRNNLGEKKQTQLFGPCGTYSNYD			
				GCGAN	AERIAIVKTLEQIAQDFFEGAVTPGDVVVFS			
				GTGAG	MSVLQAVEGGGDYRGDGNILSLVKQMGTSYG			
				CAACG	VRVTLQWVPGHSGVPGNETADHLSKTGASLP			
				ACCCT	QTDTTASLQTCRGIIQAKSRDEWLNRWANAET			
				TCGGC	GRSLFGYMDKPNPKDPSNLLCRADQTLIFRAR			
				TCGTA	TGHLHTNKHINRINPMWEPHCRHCDYHEETVE			
				GAAGG	HLLHCPNLAVLRRRLHTPTDLQSLLYSGVD			
				GTCCG	QMTRTCGYLREALRCQERNARS			
				TACAC				
				ATCCA				
				CCATT				
				ATGGT				
				AGGTC				
				TGGCG				
				TCACA				
				GCAGG				
				CCGAC				
				CTGTT				
				CATTC				
				AGGAA				
				CTCAC				
				TGTCG				

				ACATC					
				TTCTTC					
				CCTCC					
				CTCTC					
				GTGGG					
				AGCTC					
				TACAG					
				ACGGG					
				GCATG					
				GTACC					
				CACAG					
				ACTGC					
				CAGAT					
				GACTC					
				ATCGT					
				GAGAT					
				GTCTC					
				CGCGG					
				ACCCA					
				GAAGG					
				CCTTG					
				TAGCT					
				GTAAG					

				GAATA					
				GGCTG					
				CCTGG					
				AAACA					
				GATGA					
				CGTCC					
				CTACA					
				GATAT					
				CCTGG					
				CAAAT					
				GCTCG					
				CTGAC					
				CCATG					
				ACGTG					
				ACGAT					
				CCTCT					
				GTCTG					
				AGTAG					
				ATTTG					
				GCCAC					
				CCCTC					
				GGATT					
				ATTCT					

				GCCCC					
				ATGAA					
				CGGAA					
				ATGTT					
				AATAG					
				CAGAT					
				AGCCA					
				GGAGA					
				CAGCA					
				GGTGG					
				CGGGG					
				ATTGG					
				GCAAA					
				CGCAA					
				GCTAC					
				GCTTG					
				GTGAT					
				GTAGA					
				CGACT					
				CACTG					
				TGGAA					
				GGAAC					
				GCGTT					

				ACCCC					
				TTACT					
				CAATG					
				GGACC					
				CCTTA					
				CTGGA					
				CTGCT					
				CAGAC					
				GATCA					
				CTTGT					
				CATTG					
				GTTGC					
				ACTTT					
				GTGGC					
				TCCCA					
				TCTCTC					
				AGAAC					
				TAGAG					
				TGTGC					
				CTACA					
				CCCCC					
				TATGG					
				TTGTCT					

				ACTGT					
				CTAAG					
				AAATC					
				CCCGT					
				GAAGG					
				TTGAT					
				GGAAA					
				ACAGG					
				ACAGC					
				GAACA					
				TTGCA					
				CCTTT					
				GTTCT					
				CTTGT					
				GGCAA					
				GATGT					
				TTGGG					
				CACAA					
				GCCAG					
				TGTCA					
				GGACA					
				AGACT					
				ACCAC					

				CTACC TGAGA TCCTA CAGGG					
RTE	RTE - 4_A C		Aplysia californica	CTCCC CTGCT GGTGG GGGGC CCCGT AGTTT TACTA TGGGG GAGCA AAGAA TACAC CTGCC GTGAT ACTGC GTGTC GTAAG AGGCG ACTAA AAGTG	TGAGAAA GGCAATG ATGAT	MKDSGQSRKDSSDPKPSLSNKSLRIASWNVRT LYEAGKSSQVVQEMKRYRLHILGVSETHWIS GQKQLSSGEQVLYSGREDKYHAEGVGILSKY AQKSLRGWEPHGERIIMASFTTRSKSINMNIVQ IYAPTNDASDEKDAFYSQLQGVIDKLPRKDV NIIMGDANAKIGQDNVGYERIMGQHGLGQMN ENGERFANICAFNSFVIGGSVFPKHRIHKATWV SPDGV TENQIDHFCISRKFRRSLEDVKVQRGA DVGSDHLLAKLKLRLKKHGTSTGIKRTKFQ VSLLDGHKKEEFQLQLRNRFQPLADLGGEDTE VDTHWTKVKEVFTSTCQEV LGGKKWEHKEWI SQHSLDLIRIRRLKAEANASRTRGSRADALEQ YRTTAKDVKK SIRKDKEEFMNRLAEKAEAAA VGGHMRILYQTTRTLAGKYGKPEVPVKDQEG KSIFGKEAQSKRWVEHFNSLLNRPPPSNPPNIL PARNELPICCDAPSKQEILDAIKKL NKGKAAGP DQIPPEALKADANVTADILLPLFTRIWNESKFP KDWKEGHLVKLPKKGDLSNCSNYRGITLLSIP GKVFN RIVLERIKTGTDPRLRDEQAGFRKNRST	cd0165 0	551	808

				CATAG GTCTG GTCGC CCGAA AAGTG CCAGT TTGCG GAGGG CCAAC TACCC AACCC AGTAA AAAAC AGTAA GTTAC AGAAA CACGG ACAAA CACAG CTTTTT ATTTTT CGATC TTGGG	TDQIATLRIIVEQSLEWNSQLIVNFLDYEKAFD SIDRATLWKIMRHHGIPQKLVNLIKSMYEGTS CRVVHDGQLTDSFDIKTGVRQGCLLSPFLFILA IDWLMKETTRGHRNGIQWTPWTQLDDLDADFAD DLALLSHTQEQQMAKTSVLNSLSESVGLRIHP GKSKVLRTDAQRKEQVTIEGKALENVESFCYL GSIIDREGGTGAIEKSRIGKAQA AFISLDRIWKT REISLRTKLRIFNSNVKSVLLYGCEAWNASQTC IKRIQVFLNRCLRKLLRIRWTDRTNVEVLRKT RQIPVRDEVGRRRWRWIGHTLRKSGSCITKKV LDWNPQGKRSRGRPRGTWRRITEIDVQKSGKT WNNIKRLAQDREKWGVFVRGLYPGAG			
--	--	--	--	---	---	--	--	--

				AGACG GAGGA GCTTC TGCAT CAGCA GGAAC A					
RTE	RTE - 3_A C		Aplysia californica	GTCAG GTCAG CCTTA GTGCT GCTAC CATAA TGACA GTAGA TGGTA CTACC TGCAG CTGTT CCTAC TGCCT TGTAG GTGCA	TAAGCTA AATTGAC CGTCTGC ATAGACG AAAAGAG CCATATA TATA	MSPRVTFDDGSDYSVALISYRPSQTGQPPVSK VLIRHVCLLLWVYGARTKTERQTLNPAPGKK RVKLTPWKLGTWNVRTMTTGLDTTNLEKTED IRKTAIIDRELTRLNVDITALQETRLANS GTIRE RNYTFFWKGLGEDERRIHGVGFVHNRLLNRI STPQGISERITVMEMETKTGKAHFISAYAPTLT TENEKKDQFYEELQDAINNVPKADQLFLMGD FNARVG TENK VWPTCLGPFGLGKINENGQRLL EFCTNNNLCITNTFYSGKDRHKVSWCHPRSGH WHQIDFIIVRKKDANMVKQTRSFYSADCNTDH LLVIAKAKIVTKKTPRAEKQKKPKINIRNTKDP ALKAKFQEHVKEELQHLDSSSPDNMWNSMRS TIYNSASITFGIERPCKEDWVEASASTLAPLEK K RKILIEHNTNPTESTKQQLRAAKSNLQRETRK CANQYWDKLCSEIQQASDTGNIRKMYENIRV AIGPTANKVAPLKSATGEQLTDK K KQMERWV	cd0165 0	605	868

				TACGG	EHYSDLYSEERHVDRQLHEAIPQLSEMSELDT			
				GAGTA	TPTEELSSAIDELACRKAAGNDNIPAEVLKEN			
				TGAAA	KETLLPYLYKLLTTCWLKAEIPHDMRDAKITT			
				GGTTA	LYKNKGDKGDCNNYRGISLLSITGKAFARIILK			
				AGGGA	RLQKLAERVLPE SQCGFRSNRSTMDMIFSIRQL			
				GTAAA	QEKCREQRLPLHTAFVDLTKAFDTVSRSGLYM			
				CCCTA	VLGKIGCPPILLQLISFHEDMKATI QFDEGCVL			
				AACAA	APTLFSTYFAVLLQHAFKESPGDVFLHWRTDG			
				ATCCG	SLFNLARLKAKTKINKTTVRDLLFADDAALVA			
				GAGGG	HSESTLQTMMDRLSTACKVFSLVISVKKT VILT			
				AGGGC	QEGAEQSHIWLNDKELETVDKFCYLGSTISST			
				CCCTA	SLDYEINIRIGKAATTFGRLTKRAWENKRLTIK			
				AGGCA	TKIKIYEACVLSTLLYGSETWTLYSRQEKRLDI			
				GTCGA	FHLRCCRKILGISWRDKVTNVEVLARANLQSM			
				CAGGG	TTILCLKRLRWLGHVKRMDDNRIPKQLLYGEL			
				AGGAA	AQGKRQRGRPKLRYKDTCKNSLSKLGVDVSI			
				CGATC	WEEIAEDXKTWRKXIKERTKMLENSLRNNQA			
				CGAAC	EKRQRRKENSGNTELRSTTLTCKHCDRICASNI			
				TATCTT	GRISHERSCMKTTLRPS			
				CTGGC				
				AGCCC				
				CTGCA				
				ACTTT				

				ACTGG					
				TTCCA					
				AATGT					
				ATTGA					
				CCTCG					
				TCTCTC					
				CTTTG					
				GACTA					
				AATCA					
				GCTAA					
				GTCGA					
				GAGGG					
				GGTTC					
				GTGCA					
				AATTG					
				GGCAA					
				GGCAC					
				GTTCT					
				CCATA					
				TAGTC					
				AGGCC					
				CAGGC					
				ATTTA					

				GCTTA					
				GTCAA					
				GATTT					
				CAAGA					
				TTTCTG					
				GGAGG					
				ACGCT					
				GCAAA					
				AGAGC					
				AAATA					
				CAGCG					
				GATCA					
				GACCC					
				ACAGC					
				TCAAC					
				CACAG					
				CTACA					
				ACAGC					
				ATTTA					
				TATGA					
				TTATG					
				GAGAG					
				GACAC					

				TTCAT CCTTG TTGTG ACGAG CAAGG GAAAA AACAC GGGAG GTGGC AGTTT ACGGG TTATA AGACA TCCTT ATTTG GCGAA ACAGG					
RTE	RTE - 7_A C		Aplysia californica	GGTGT ATGCA CCGTG CAACG ACTCC TACTC	TAAAGTA A	MGDFNGRVGER RTPWTKHLGPHSDHRTPCNY NGNHVLELCAEHDLFITNTFFQHRASQIYTWY RWNNIMVSSQIDFILTRTRMRITITDSRAIPNAG LDTDHRPIITTLTTKKERKPKKHKRQQERLNM HKLGDDQVQAQIRSTLNEKLGSINSTILTVEEA WDTFKTTLLDTMKEVCGTKKTRGKHRKATA	cd0165 0	392	640

				AGACG AAGAA AAGGA CTCTTT CTTCG AACAA CTCTC GGACT TAATT AGCAA CATCC CAGAC GAGGA AGACC TTTAT GTA		WWNEEVKDAIKEKKRLYNVWVKSKNEEDYI KYRLARRHSSKKVVITSKEKSWTQYGEKLSETC KTSPREFYKKVKTMRTRDEPYVPTTVINDTNG EALHEEKEIIRWEEYFQELLNPSGVQAQNTQ CPFTANYPEHSEPTILESEVREALKTSPKNKAA GDDGITTEAILACGETGIHWLTIIFQKAWKERK VPADWQNAIVVPIWKKKGSKKDCNTYRGISLL SHAGKMYAKILERRTRAKTEHLLSDAQFGFRK GRGCTDAIFALRQLCERALEYDKDLHLVFVDQ EKAHDRVNRNKLWKILEQYDIKGQLLDNIRAI YANSRSAVRTTSGTSDWFPVTSGVRQGCNLSP LLFVIYMDQITKEANPDPEALNELMFADDLAM INNKTQLQEHINQLNASCEKYDMKISISKTEVM TISRRPGKVDININGSQLKQSREFKYLGSIFTEN GKLDREIETRCQRANAVSYQLGPLLKHPNIPM STKAKLINAIPLTLTYQCQTWTLTKALERKLV TCMKCLRRAVNKTRRDKIKNEVIRDTVGATP VLPHIEKQKVWFGHLTRMPSNHPALRAYNIK YSGWRARGRRRRWSESVADTLKDHGMSLLQ ATRLAADRHLCLPATPQGTSGRKK			
Tad1	Tad1 -	.	Blumeria graminis	ACATC CACAT TCCGA	TGAAGGT TGACAGA CTTCTCC	MLGAHDLNLILPKGTITWKSRRGSQSTLDLAFV SNALESTVMACKPAFDLEASSDHIPICTQLHIEP PNREEQQPRPQWKKADWEEVNRRLALKLSRL	cd0165 0	332	595

23_ BG		AAAGT	CAGTAGC	STDHLELNSPESIDQRVSAITQAIKETIPTARPST		
		CGACC	AAGGAAG	FAKPYWTKECSETVKEARKARRLWTKLGSEE		
		CCATA	TGGACGG	SWVEYQKATNRKKSQIKKAKMIGWRAIVSEA		
		GATTT	TATCCGC	TNDPAKLWKLAKWAKKSPEERSRLPQLPDIKD		
		TTATC	ACTGGAC	KDGNMFTEDTEKADILAQHFFPQPVQANLEDI		
		CATAA	ATGACAG	DGTRYPEELTTVSSIITASEIQDTMANLASEKA		
		AGTTG	TGATAGT	AGPDGISNWMLKNCRITLGKVLEKLFNACISIG		
		ACAGC	GCACTTC	YHPKGFKESVTIVLRKPQKESYDNPKAYRPIAL		
		AAATC	TTGAAGG	LNTLGKVLEKLVANRLSKAAEAHHLLPDEQM		
		GTAAT	TAACACA	GARQLRSTISAVELITEQVHSIWGKNKRRVASL		
		CGTGA	ACATCAA	LSLDISGAFDNVSHKRLIHNMRAKGIPRWTTQ		
		TGACG	A	YIESFLEHRTTSMNLGSFKGEQFPTSTGIPQGSS		
		AGTAT		LSPILFLFFVSTLLPELQSPSSTAVGFVDDTNILT		
		TCCTA		WSNSTEDNCRNLERLHGICEKWADRHGVKFA		
		CGAGA		PEKYQLMHFSRARKRHNMQATVQIQGHLAKP		
		GGAAT		QPSLRVLGIHLDPKLNWGAHVKSIQLKVEPQI		
		TAATG		QALGRLAQSTWGATFHKAKLVYNAVVRPAV		
		GTGGA		TYGSPIWAEPGMAGKIPERVIKPLRSIQRRCLKI		
		ACGCC		VTGAYNSTSSKVLEHETSTLPMEIYLKQRRVQ		
		TTGCG		HAGLSKSLPVQDTISSACRKIKQPRAGRQDAR		
		CGGGC		AVTRIMDIADWGKIGANIXTKERQKAAGKIAA		
		TCTAG		FKEWERSWTRLAARQPTNHHATADPKIWNA		
		AGGTG		NIFTDDKTGRTRMDQRGTPSSLHKHLSRAQSSI		

				GTAGC	AMQIRSEHIGFNSYLHRRKVPGVDSPRCQCGY			
				GTCGC	PSQNVKHMVMACTSWAKGRGEILRKAENRSF			
				GTCAA	EAMMNNPKDLARITRWIQQEGRLEQFRLTLEV			
				AGGCG	EATMRAREEG			
				CCACA				
				AGACA				
				CAGAC				
				GCGTC				
				CCTGA				
				AAAGA				
				TTCAA				
				GGACG				
				CAACT				
				AAAGG				
				CGAAG				
				AGCTC				
				TCCAG				
				TAGAG				
				CGTAC				
				AGGAG				
				AATGA				
				TTGAA				
				GAGGC				

				CATTC					
				AGGTT					
				ATCGA					
				AGGCC					
				TTCAG					
				AAAGC					
				GAAAG					
				TCACC					
				CCCCC					
				CATCC					
				CCGCC					
				CAGAC					
				AGGCC					
				ATAGT					
				GATGA					
				TAATT					
				CTGCA					
				ATGGC					
				TAAGA					
				TCCTA					
				CAGGA					
				GATAA					
				AGGAC					

				ATCAA					
				AGCCT					
				CCATA					
				GCCCA					
				AAATA					
				GCCCA					
				CCAAT					
				GCAAG					
				CCCCA					
				GGCCG					
				CTCCT					
				GGGCA					
				AAAGT					
				CGCAA					
				GCAAG					
				GGAGA					
				GTCGT					
				ACGGA					
				ACTAC					
				TATCC					
				GAATA					
				CAGGA					
				TGAAC					

				AGGAA					
				AAGAG					
				AGAAT					
				AGCCA					
				AGCTC					
				TCGAG					
				CGAAG					
				AGTTA					
				GTCAA					
				AAAAA					
				TCGGA					
				GTACA					
				GCAAA					
				TAGTA					
				GGTGC					
				TAGAC					
				AAATG					
				ACAAA					
				TGGTC					
				AAGTG					
				AAGGT					
				CTACT					
				TCGCC					

				GGTGA					
				AGGAA					
				CGAAA					
				GAATT					
				ATGGA					
				GCGTC					
				ACAAG					
				GACTG					
				GACAA					
				CGAAG					
				CTGGC					
				TGCTT					
				CTGCG					
				AAAGT					
				TGCCG					
				GCCCT					
				GTTTA					
				CCAAG					
				TTCTA					
				GTGCA					
				TGACA					
				TGCCG					
				CTCTC					

				CTTTG					
				AGCCA					
				CAAAA					
				TGCAG					
				AACAT					
				CTGAA					
				GCTGC					
				TACAA					
				AAAGC					
				CAACC					
				ATGTT					
				ATATA					
				CAAGG					
				CATGG					
				TAATM					
				CAAAG					
				AGCCG					
				CATGG					
				CTTAA					
				GAAAT					
				CTCAG					
				CAGCC					
				AGGGA					

				AAAAG					
				TCTGG					
				ATCTT					
				AATCC					
				TGTGG					
				ATTGA					
				TCAGG					
				CCGTA					
				CACGC					
				AGATA					
				TGGCG					
				ATAAA					
				AAAGG					
				GCTTG					
				ATGTG					
				GCAAT					
				ATGAG					
				CTCAA					
				GTCAA					
				CAGAA					
				ATATT					
				CAGGA					
				GTGGC					

				TTCCG					
				CATGA					
				TGCAA					
				TGCTT					
				CAATT					
				GCCAA					
				AAGTA					
				CGGAC					
				ATATT					
				GCGAA					
				AATCT					
				GCTCC					
				GCAGA					
				CTCGA					
				AGTGT					
				GGCCA					
				CTGTG					
				CAGGA					
				GAGCA					
				CAACA					
				CTCGA					
				AACTG					
				TACAG					

				GAAAA					
				CAAGA					
				CGTCA					
				GATGC					
				AGCAA					
				CTGTG					
				GAAAA					
				GGCCA					
				CAAAA					
				CATGG					
				GACCA					
				GAGCT					
				GTCCA					
				GTTAG					
				GGCAG					
				CTGCA					
				AAAAA					
				TAGGG					
				CTGTC					
				CAAAA					
				CAGGA					
				CCCAA					
				GACCC					

				AGGAA					
				TGTTCT					
				TGACC					
				GACGA					
				AAAAT					
				CGACA					
				AGAAA					
				TCAAG					
				ATGGC					
				GAATG					
				GCAGA					
				TTGCA					
				GGAAG					
				CAGGA					
				AAAGA					
				AGAGC					
				AGTAA					
				TGACT					
				GCACC					
				ACTTG					
				TTGGG					
				GCCGA					
				TGGTA					

				GTATA					
				GATGG					
				TCGTA					
				GGCGT					
				CCAGG					
				CAGAC					
				CACCA					
				AAGAT					
				CCCAT					
				CAGCT					
				GTCCC					
				CACAG					
				TCGCA					
				GGGCT					
				CTTCG					
				CAAAT					
				CTCAG					
				CGAGG					
				TGCCA					
				TCAAA					
				AATGA					
				CACCC					
				CGGCC					

				TGTGA					
				CACGG					
				GGAAA					
				TCAAG					
				CAGAT					
				GAGAC					
				TGAGA					
				GCGAA					
				ATGGA					
				TTCAT					
				GAAGG					
				ACAAA					
				TCAAT					
				CCTTTC					
				AATCA					
				TACAG					
				TACAA					
				CGTCA					
				ACAGG					
				AGTAA					
				GAATA					
				AAGTG					
				CAGCA					

				CGATT					
				TCCTC					
				CAAGC					
				TTTGA					
				ATCCA					
				CTGAA					
				GCACC					
				ACATA					
				GTGGC					
				CCTGC					
				AAGAA					
				CCATG					
				GCGTC					
				ACCCC					
				ACMGA					
				ACAAC					
				ACGAC					
				CGTCA					
				CCCAC					
				CCTGC					
				CTACC					
				ACCTG					
				GTCTT					

				CCCAA					
				GTGGC					
				CACAA					
				AGGCA					
				GGACA					
				TGCAT					
				CTATA					
				TCAGC					
				AAAAG					
				TCTGG					
				CTATT					
				AGCAA					
				ATGGA					
				GAAGA					
				GAAAG					
				AGTGC					
				CAGAG					
				GAATC					
				GCAAG					
				GAGAC					
				ATCAC					
				ATCKG					
				TCAGC					

				CTCGA					
				TACCG					
				ACCTT					
				GGTAA					
				AATTT					
				GGATA					
				CATAG					
				CGTGT					
				ACAAT					
				CCACC					
				CCCAT					
				TACAC					
				CACAG					
				CAGCC					
				TAGAA					
				GAAGG					
				AACAC					
				TCAAG					
				TGGCT					
				CCCGC					
				AAATC					
				CTATC					
				TCAAG					

				AGGGG					
				GAACA					
				CATTC					
				TAGTC					
				GGAGA					
				CTTTA					
				ATGTG					
				CACCA					
				TCCGC					
				ATTGG					
				GGGGG					
				ACAGT					
				CAGTA					
				CTTAG					
				GCACC					
				ATAAG					
				ATCGC					
				AGACA					
				GCTTC					
				ATCGA					
				A					

CR1	CR1 - 7_C Bre	Caenorhab ditis brenneri	TGATTTC TCTTGTCT TTCCCTTT CTTCACT TGTCGCT TTTGTTTG ACCTCCC ATTCGCT CCATTTT AGGCATA TCCCCCG TTTCTCTT TCTCCCC CGGTGAG GCCTCGT GAGGGGC ACTTGCC AGGTTCT TCTGATG TATCATA TTGATTTT TCCCTCTT TTGGAAA	LLQKLINFGIPSSFVSWFAVFLADRTFSVKINDS IGDTSPTIPSGVPQGSVSGPLLFLIFINDLLL TIPLHIAAFADDIKJFSNDPFVVQTGIDTVVG WAKSNSLPLAHSKTALLHLGSLNPFFPYTIAGN PIIPTSSVRDLGLLTEPNLKFSAHINRTVALALL RSKQLLKSFKSSSPQFYVYLFKCYVLPLIEYCS VVYCPPSTSKLSLKLEYPLRFFTKKVFHRCNIS YTSYPDRLAHLDLFSLRHRRLKAQLQLLYNFL SGSSYFPNLDSHVRFSSSPRRPMNLVNVNPRV KDFFFITVPIWNAITCHCPTFLRPSEFNSSLQCSI QRF	SSF56 672	45	118
-----	------------------------	--------------------------------	---	---	--------------	----	-----

					TAAATAA ATAAATA AA				
RTE	RTE -1	AF05 4983	Caenorhab ditis elegans	CTCTTT CTGTA CTAAG CCTTTT CGTGG CGAAA GGGAT CAACA GATAA CCACT TCAAC AAACT CGATG ATTTC GTCAC ATCGA GAGTG GCTTG GTTTG GAACC	TGAAGAA CGGATCG TCTAAGT ATCTAAG TATCTAA G	IGLAGHYRVSRDTTFALNSNPSCLSPPHPASAK GKTVPGARRKRRSNLGVQTICQSDLEIKKPRD KTRRTDYRVCTFNCRSLSSDDRLAELLEETRI QFDVIGLCETKRAAEAHLIHHDGTGVFLGKRN ESSVSGGVGFIVRSTLLPKIVEVRFLSHRIGLLT FKVSRKFNCTVIQVYAPTADSDLEEICDFYDSV EDVFRECRSKYKL VIGDFNARMGCRQNNERYI GPHAMEPRNDTGELLATFCETNRLWHTNSMF KKPMHKRWTFVSPDGNHRHEIDHILANGKFV TDTTVLPSTNGSDHRLRLCNLHFNNCLAKLE QVRRRKPPKRVLDPAIAYAISATVTVQSDPDL DTDYDNLIQSLKELQDQAIVRPANHSSNRLSEE TRKLLNKRRFMDRNDPQFKSISDKCREAVQKD HEAFASRLLSAANQKSLKRVARDINEYKSV IPCLKSTSTGERITSRVKMEQEIEKFYTELFKSA VSNSQTSSIPATATPPPFLPEEIRHVLRSPNGK AAGQDKISADFLKSCHDNVIDLITDRFNRYLHS RNVPKPWKTSKTTLIFKKGDRENLENYRPICLL PVLYKVFTKCLLNMRRSLDEAQPVEQAGFR RSFSTIDHIHSLQRLLEVGREYQIPLTLVFIDFK	cd0165 0	562	831

				AAGTG CAAAT TATCT GGCCA AGAAG TACAC CGTCT CACGC AGGCC AGAGG TAGTG TCGAA TACTC TCTGT CCGGG CGGCT CTAG		KAFDSVEHQAIWKSLEQGADGAYIDLLKEC YKNCTTNFTPFHRPVA VPVTKGVRQGDPISPN LFSACLEHVFRKLSWIELKGEAEDYDTIPGMR VNGRNLTNLRFADDIVLIANHPNIA SKMLQEL VQKCSEVGLEINTGKTKVLRNRFADPSKVYFG SPSPTTQLDDVDEYIYLGRQINAQNNLMPEIHR RRRAAWAAFNGIKNTTDSITDKKIRANLFDSIV LPALTYGSEAWTFTKALSERVRITHASLERRLV GITLTQQRERDLHREDIRTMSLVRDPLNFVKK RKLGWAGHVARRKDGRWTTLMTEWRPYGW KRPVGRPPMRWTD SLRKEITTRDADGEVITPW STIAKDRKQWLAVIRRNTTNS			
Ving i	Ving i- 2_C E	.	Caenorhab ditis elegans	C	TGAAAAC GGCGCTG CGCCGCC TGTTGCA CAACAAC AACAA	NMKQPAKRTVMTCLDQRAAFDKVPIKKLCVD LHESLNEAKISKWLTNYLSKRIIKANFGNHSSK WCTLLGGVPQGSVLSPNLFSFYMRNMPIHPKT EITVYADDTTILAQDTNPFKAGELIQEHLDLLY DFFNERKMSISAEKSTTTLFTCDTKEFKTKIPV NWNGAEIPTTNQVRLLGVHLNTMCGSKDHVA	SSF56 672	8	150

						ETRKKMMVSTNILKALSGTKWGCSKELLISTA KALVKPIAIYGAPAWSQLLSETNLNKLEAAYR NTLRTCCGLTSDTPIEHIYAEAQMMPLKEEFEL AKQQTYIATTKSGTHPNRNLESRRNTERRSNR PREPPLNLQEATKTHLSRLKGTKAEAQKANHT RFVREFLNSAPPNRILGATPPPLNSSEESSLPKA TRTELARLRCGHSLLIPRYKARIEGTPIGTCCG GDNEGGIDHYLSCGRSHPISEKELWTSPMEAA RALGLAYKPYDPGGGATTSV			
CR1	LIN E2G _CE	.	Caenorhab ditis elegans	AGTGA CGGGT CTCAG TAATC AATAC TCTGT CTTCC GTACG GATCG TGCTA AAGGT AAACG AGGTG GAGGT	TGATCTA ATGTTAC CTTATCC TCACTTC TATCCAT CATTTAT TCTTGCC TTTTATA CATCGAT ATTTGTG TTTTTTGC TTCTTCTC CAATGTA GCCATTA	MPTRKSKILDVALSNSVISNLNIMAPIGSSDHN TIKFTVQLLQNISVPKKMIPDFSKLDWPCCATY FENIEWGKVFDTTHSHVNDIYECFLGHIRTMFSSQ AVPTMPMPETSKPVPQYLRSEFQMVKNQFSKT IMSKTMNDFIEYIYLARKYRKKLRKYLSNLEV KLFKKKGNSKFSKHAKTLLKPRNCSVPALIDN DENEIVSELDKANLLASYFEAQYTNQNNTNFA LDDDFVSDQPIPWVTD AEVLKVIMASKKSCSV TSDQVPFTFIKILSPLISSVLAQIFNLSMMRGQV PHAWTESIVIPLDKITNANNPTDFRPISLTSHLC RVYERCILAKIELFLIDNKFWSQNQHGFPRPKKS TVTNMLESNDWSKSIDLGEQIDIYLD FSKAF DRLPHDLLLDKLVSLRMNKNLLIWLNEFLSNR SFRVRIGQTLSSRKLATCGVPQGAVLSPLLFGI	cd0165 0	299	552

			GTTTG	TTTTCTTC	YVNDISSILPEKVACKQFADDTKLYVSTPKTES			
			TAATG	TCATCTC	ENNLQSALDAVVDWTRGSKLTLNQSETVHVT			
			TGGTA	TTTTTTAT	VGKRRTDFKYHLDGYPIERKAITRDLGFLISEK			
			AAAGT	TTTTGTTC	LDFSEHWRKCINLAKFQLANMFNQYSTSNKK			
			AGAAA	CGTTTTA	LMILLYKTFIRPRLEYGTVVSSPTKKSDEKAIES			
			TCCCA	TTTATGT	VQNAFTRRLYSRIKKRYINPTDKDYKTAMERN			
			TGTGA	AATTTAT	ELFDLQTLSTRRKIIDRIQVIRMNNGKVDLKTS			
			AACAG	TTATGTA	EFFEQQETHTRSKKKYIWTGKSKLRRHFFVN			
			TAGTT	ACCACTT	RTLATMKQH			
			TCAGT	GTGATGG				
			TGAAA	TTTGTAT				
			ACGCT	CTATTTA				
			TAACG	TGTGTGT				
			CTTAT	TTTTTTTC				
			GAAAT	TACCTCA				
			TCTCG	CGATATC				
			TAAAT	TTACTAT				
			GACCT	CGTAATA				
			GAACA	AATTAAA				
			TTGAC	TTAAATT				
			GTAGT	AAATTAA				
			TCCCTT	AACTT				
			TAGAA					

				TCATA					
				AACAT					
				ATATC					
				GTGCT					
				CCTTCT					
				TGCAC					
				TTATC					
				AAAAT					
				ACTGT					
				TAGAT					
				TAATT					
				AGTAA					
				AATTG					
				AAGAC					
				CTTGT					
				GGAAC					
				ATCAA					
				GGCAA					
				AGTTG					
				TAGTT					
				GTGGG					
				AGATC					
				TTAAT					

				ATTGG AGGTA TCGAC TGGGG CCTCA ACGAC TCTAA AACTA AATTA GGTTC TTTATT AGTAG ATCTG TGCAA AAACT ATAAT CTGGA ACAAT ATGTC AAA					
CR1	CR1 -	.	Caenorhabditis japonica		TAACTTA ATGTAAA AATTGAC	STKLDLXXXEDJXFNEXHLNKSSPKAFFNFINS RXKXXQEIXXXRDXXXXQVTSDYMKAELLSN CFSNAFTEDKNTNTDFPYRTNHFINSLCFEPFIV	cd0165 0	151	401

	15_ CJap				TTGTATT AGTGGCT TTGCCCT TCAATCA ATAAATA AAA	ESILSKLTPKCNTTPDGIPAIFLKRTCTSIALPLSI IFRNSYNQXVLPSEFWKTALVKPLHKKGSRSDP NNYRPISLTS AVGK VMEKMVRKSIILFLDDFKL LSNSQYGFRSKMGTESQLLSYQNKLV TNYLSK QTTFSVYIDFRKAFD TVGTKKLLIKLERYGIRN HLLRWLENFLSNRTQRVIINGVMSEEKSVLSG VPQGSVLGPLLFLLFVNDIGDNFNSDYLLYAD DLKIFSTDKSSIQRDLDSLSTWCENWQMSVAP NKCETIAFYHSKRGMDKDTKFTIGGATVPITSR IRDLGMFFTS DLSFSNHIDTVLRKAHQRVNIFF NVL RHAGFDTFIKCFIIVRPLIEYGSIIFSPTLK DLVKSVESVQKSFLYRCFKKFRIEYISYFDSLS YCKLESLEHRRLIIDLTFLYKLIVSKEIFIHNFTC ITRSPLPSLRRHPYIVRSTLRNSSKLSHQFFSNRI ITCWNSLPSYLFPKLICSETFKSALKCLDLSKFL TLNMHNY			
RTE	RTE - 1_CJ ap		Caenorhab ditis japonica	TGCTT GCTGG ATACC TCCTC GCGGT GCAAA CTGGG	TGAGGTC GAACCTC ACAAGTA TCACAAG TAAGTAA GTAA	MFGRSTTDTEITEASYGPKNRRKGNDFHIGSLN VRSISDQSKAEAFDLLAVNSGCKVIALQETKR KEGEWVLSSGAELICTRREAREGG LAFWIHKD WLDSL VETKRESDRISTASFDFHGSHLVIN VY AP TSAANETQSDKFYKDLKKVYTRL SKWDSSI IICGDFNAVVGCSDDL PFVGGHGTGQRSDNG DRLVNFAFSSRVSILNTIYPKRPNRKWTWISPN	cd0165 0	531	773

			CAATG	GKTKNEIDFILTNRREIAVDCEVYCDYLNCS		
			CTGTG	DKWREVKSL		
			CTCCT	DPVRFKNTITEMLTENPPTCYEELTESIRSA		
			ACGAC	ATHATHQLSRSEATKKMMLSRFQMIHAGQTK		
			AGGGG	NADLAKINNEVRRSLENDLDKWREVKSL		
			AAATT	VKGRSLKKTKRALTLKSSPFTNLERPDGSTTY		
			GGACA	NRPEIRKLVLTHFSNLYAATTREPRSRYNHPDE		
			CCACT	PEILQHEVEHAIVTSKTNISPGPDQITMLQLKLG		
			GTTAG	IESIAPHLTAALNQVLNTRTPEDWKTVKISLL		
			GCACA	PKTTKPKKVVDYRPVALSSIVSKLFTKILTRRIT		
			GTAGC	AKSELYEESQAGFRRGRGCADNIQVVAQMW		
			CAAAA	EKCTEFKIPLVAVFLDFTCAFDNVNWT		
			TCCAC	NNLQIGRNVIRALNNSNSAIGELNVLNKKMR		
			TCTTTG	FKIKRGVRQGDSSSPLLALQAILDELD		
			CCACG	CEDDKTGIDINGESIHRLEFADDVVL		
			GTAGT	AEDRANRIAMGCPKYGLNINPSKTQILMNKYV		
			GTCCG	TRSDIDILNTRIEISQKVLYLGR		
			ATACT	VARRIRAGWAAFNAIRRELLQSPQKIRIRLLTT		
			CGTGA	TVIPAMTYGCETWTSKEADTKRLQRAVSNMF		
			AGAAG	EIAGCDPPDMEKQVLRRLKLVAGHVS		
			GGGCT	DRWAKIASEWDPRGSARPRGRPPKRWADDIT		
			TCACG	DSIKIFIHNEALRGNLGHGRRRIGILEARRAWS		
			CCGGC	TVGRDRVTWRS		

				CTAGT					
				AAGCC					
				TGCTG					
				AAGTA					
				CAAAC					
				CTAAA					
				CTTAC					
				TAAAA					
				AGCTG					
				AATCA					
				AAAGA					
				ATGTA					
				AAGGA					
				CCCGT					
				CCAGG					
				AATTC					
				ACCAG					
				CTCGG					
				CCACG					
				GTGCT					
				CGCCG					
				AGTTG					
				TCTTG					

				GTCGG AAGAA A					
CR1	CR1 - 6_CJ ap		Caenorhab ditis japonica	GTCGG AAGAA A	TGACCAC TGATGAC GGCGG TCGGC CTATTT TTCTCC GACCG TCCAT GTCGC GTGAA GAATT GGTGC AATAC CACGA CAACC GCCGC AAGAA CT	MPTNHSQRTLMPRRCKYERTRQRIFSRIQKTNF AQHRXFNF FXHSPDLILSILSANXRSLMANKLE FQYIVNTTSPDILCIQETWLSSKFGNASFLDPNS NYHVFRADRCATSRGGGVCTLVKKEFPXTIIT AETKNSAYELITVDIDVXVTPLRINIYRPPSCSF ENTITLLNKVSDLSNHVGQJVILGDMNINGINW NSAVPTSQTRLGSLLCDFCNQNNLTQYIRKAT RKENLLDVVFSNSRLSDVRILSPIGSSDHNSINF KLDILQRPPEKKILFYDYKNIKWPKCLDYLGSI DWGRLLHPGNSISTMYQNFIDHMREAFELSVP KIYVEPTIRNTPDYLKSFYKLVQLKLERVKASR GRXDFLDYIKTAKKFRKKLLKFRNNEEVKLE SSGVSFRFSKKAKYLLKPKSLRIPTLTNPLGVLS VTNAEKAEILADSFESQYVNRPLLNFELPSNFK SPQEIPWVTDAEMLSMMLKCKNSCSQTPDNV PYKFLKLIAHLISSPLAQLCNLSMMRXECPSLW KESIVIPLNKKEKPMAPTDFRPISLTSQLCRIYE RCLLKKILQHLDTIKFWNNDQHGFRPKKSTVT CMLEAVNDWTESIDEGAQVDVIYLDYAKAFD RVPHDILLGKLVDANLNPNIKWIRSYLSNRXF	cd0165 0	528	785

					ACCTCTT CTGTATT TATTACT TTTCACG TACCGCA ATGGGCG TATACAA ATAAATA AATAAAT AAATAAA	SVKVNNESSTPKHAQCGVPQGAVLSPILFGIFV NEIPSILPSGIRCKQFADDIKIYTAIPKSDTIEDH ETLMQSAIDAVCNWSENSKLALNDSKTLVISL GKNPRTTKYQINNNPIRRETSIRDLGFIITPMLD FTGHWKKAICKAKFSVFQIFNHYCSNNPRLMT LLYKIFIRPILEYGTEISNPTRKLDCKAVESVQN AFTRRLYSRIKGRYLTPNDPDYLTAPQRNSLFN LSSLESRRQTSRKLILKMISGLVDIDTSPFFTY EKNTRTRAKNKFKNHCKTKIRRNFFVNRTLS SFKL			
CR1	CR1 - 12_ CJap	.	Caenorhab ditis japonica	TT	TAAATTA ATTCATT TCAATTG TAGGATT GAAACAT ATTTAAT CTTTTTTT TTCTATTT CTATTTG TTTTTCTT CATACTT ATTGTAC ATTTTTAT	LINILNELLTELCPRIPERPXSKVCPLLRKLRL RRYKSSSLSHADIANRLKRYRALASRYRKWVL KHETKLINESNCTSIRKFVSKRLKSKQSIPALLD NGKAVTSDFVKCDLFSXSFMKNFACPPEEYST PTPSHYGSDSLSSFSSPSSSRCLFLPYIVEKVLQ HLPPKCGFSPHLANYLVLKKCATPLALPLSIIFS QSYCESEVPAAWKLAVVIPVHKKGNPACTDN YRPISLTDPIARVMERVICNYIRSELAYKFSPYQ HGFLHFRSCTTALASSTAKYRQILKKSKTLDIIF FDFSKAFDQVSHSILLDKLAKFGVDPLTCSWFE NFLCDRSFTLKVNNTFGDKLYPVYSGVPQGSV SGPLLFIIFINDLLFSLPPPIHYSCFADDIKIFSDSP EILQSGIDTVVQWSIRNCLPLAPTKTSLLHLGD	cd0165 0	215	460

					TGAATTT	KNRMLKYFIHGIEIQHSSCVRDLGLITDSSLTFK			
					TTCCCCC	IHISRIAALANLRANQLLKCFSQSPQFYAHLF			
					CTTCATTT	NTYVRPLLEYCSIVYSPSPSSSLAGKLENPLRT			
					ACAAAAT	YSRRVLQRCNISFSSYENRLEILNLTSTRHNRL			
					TGTACTG	KSQLLFLYRIVTNNSHFPNLSDFVSFVKSNRRP			
					TTTTATCT	MLLVRNDTCTLHFFARTVPIWNSLLRNCSTFLS			
					CTTTCTCT	PIEFSNLLNVNISRF			
					TTCACTA				
					TAGCGCA				
					ATTGTAA				
					TCTAGCT				
					TAAATTA				
					TCTCCAA				
					CATGACT				
					TTGTTTG				
					ATTATTA				
					TCGAAGG				
					ATCCTCT				
					TGCGGGG				
					TGATGTC				
					GAGTAAT				
					TTTATTTA				
					TTCATTG				

					TGCCATT CCCACCT CTTTCTCT GATCTCG AAATGTT TTTTTTTC TCAACCC GTGAATA AATAAAT AAATAAA TA				
CR1	CR1 - 16_ CJap		Caenorhab ditis japonica	TGCCT CTCAG ACTTT AGAAC TTTTCA AATCC CTTAC CGCAC TTCTA G	TAAATTT TGTGCTT GTGATAT CATCCCA CCCGTCT TTCTCTC ATCTATT TGACCGC GGATTTT TCCAATT GGTCATC TTACTTA	IFXTIILLYW AISIFPQSTGDLLPDSPEFLVDXMT SFHLSQLVTFPTRTSRLGKKNFLDLLFSSNTSII KNLKAGSPFILSDHCSVLFSLSLATSTIPKPTE DPQHLHFRKCDFNKLNCHLFSYNWPKQLSYF ATADSKLYHFLKIFNELIEQYTPISTKRTVHSKI TGKKLYRKLRKRHPNSLSLSLKKRVKSRLKNI RKSFESQEKKVIASKNPRSLYSFIKSRIKTKSTL THFEVDKKVIQKPSDMADLLVNIFSKNFNTPPP PFPALPSPKPLNTSIDTSPFGIYQVIKIKPKIGY SKDNINNFIKNCADSISIPLSLIFGEFLLSRTFPD CWKISTVIPVFKKGNIHDPLNYRPIALTHPLAR LFEKVVGPIRRTFSPKLSKKQFGFLNNRCTL	cd0165 0	343	590

					TTTATTTT CTCTCCC CTTATTTT TATACCG TTGTCCT CTTTTCC ATGTGCC ATATACA TATTTTAA TGCGAGA AATGAAT AAACATA TTATTATT ATTA	HLHASSIFHSVLSYPKSYLDVILIDFRKAFDSVP HELLLCKLSNFGFDTKLCEWFREFLTGRSSIVK IDNNYSSIKFKNSSGVLQGTVTGPFLFLVYIND LLESFPDDVHVVAFAADDIKIFGNNRSSLQKSIKI IEEWCSIWQLKLAENKTQVLHLGVHNPRVEY HIGGITIDKCTSVRDLGIWVDEKLQFRKHINIIV NSCMYKVRQILNSFRSSNLNFYFKLFNTYIRPV LEYGCEVFNPCSVSMSTSLEKPLRFFSRAVFKR CRVNYLSYADRLLQANQQTLKHRRVLLILKTY HNIISGWYHFPNISQYIKLAHTPRFPYLLQPRG KWTPNNFLHKYLSLWNKIAKKFPKTVSGATF HLHISALPLDSYL			
Ving i	Ving i- 1_CJ	.	Caenorhab ditis japonica	TA	TGACCGA AGATCGT CAAGGAA TGAGTAG AGTTGCA CAACAAC AACAACA A	RPAPPPPSQHTTKCRAKPAKTRILSWNVNGWS KRRSELATIIAKEKPDIVCLQETKLDKTNLGYT PTGYKVALASRNKFGGGVCTMVRDDWIFTISK TTSKKTAEILSITIDKLMVTNVYFPPRSSDSPTA EVREDVMEILDMASEKHIITGDWNAHAKTWH SHKTDERGAMIEDELDTRGVHVVLNEDSFTRA ASGVLSSPDITISTSDLAACSSWRTCGLNTDH LPVIIDTMAEIPKLVRPXRIMANFKKANWTDFS DSIETKVACYAGPRDAYSLNFLSATITRAASK	cd0165 0	502	758

						HIPHGKFTPTREHTGDTGEIAELTRERERIQKR NPLDSNLKALNDKIEAARRELRR LXWHQKIED MEDKRRKIDCNALWRVIKGLKPKTKSTGSIRLP NGTTTTNPKTIANAMASSLVAVGKGRDCREER IAWRRTKRRLKKLRPQLPASSISHAEVKLAILK GKPSKALGPDGICHLHLKHVPDNCINLMAELF NASIDENKVPDSWKKANVFMLPKPGKDPSEIK SFRPVSLSPVAKVLERIILERIRREIESPKDQHA FKKNHSTTTAITEATSCIVGGLNMEKPACRTIM TCLDMKAAFDTISIAKLLCGLEDAIADKKVCG WLGNYLHKRRINVQYDDVKSSWLT LRGGVPQ GAVLSPNLFNFFIRDVPAIPDTKIITYADDMTLL VQDPLLSKAAEKSQDALDSLQKWFKDKHLDIS AGKSTVTIFTADTKEFKYDPGLKWEGSVIPMQ NRTRLLGVELDTMLRMTGHVEQVSARLAKAN NILRALAGAKWGSSKETMLRTFKAIKPLATY AGPAWHQLMSETQKTKLERQYVGGLKACCG LTKDTPRQLVYHETRMLPLSEELTLGSEQLAIA AIKTPGHPTSDL SRRRPKERTCGNRPRTPPLGC LEDRGDEWRRMRGDTRAVQKRNHTSFVKSYL SNHGIHPILGRQPPALSEEESTLPRNTRVELARL RAERSLLLEKYKAKVENRPVVCCIKCNDDVG			
--	--	--	--	--	--	--	--	--	--

						DLKHFLKCYPVKPLPMSKLWKDPVAAATALG LAVTPFDPGGDADS			
Ving i	Ving i- 2_CJ ap		Caenorhab ditis japonica	CACTA AAATA AAGGA CCTGA ACGTG TACAT TAAAA CGAAA CAAAC ACACA CTTGC ACGAA ATGCG ACCAA GTCAC TAACC GCCAC CGCCA CACCW CTAGC AAGAC	GAA	MPPGNEDRYRTRGIHSNWLQDRYGPEEQARR RSVXNGQARLGLHDHSIHDEESGLLITNVYFPP RSSDSPXAEVREEVLQILDMEDRKHILTDGWN AHSSAWHSHKTDERGSMIEDELDSRGNHVILN ENTFTRAATGILSSPDITMATADIASRATWHTC GVVNTDHLPIIIAMDSEIPKREKPKRVIANRYK ADWKGFKTDVETRIRSYVGPRDAYSLLEKLISK AILEAAGKHIPSGRFVPTRECQGDTKEMDDL KKRDEIQKLNPLDGKIAEINDVIEKSRRELKKE RWLNKVEDMEIKRKTDNCLWSLLKGLKPKA RSSGPVNLPNGTTTTNAKAIANAMAKNLVTV GEGRENKDEKMAWRRAKRNLVKSKTTLATTI THAEVKLAITNGKSSRAMGPDGLCHHLKNLP ENGIHALAELFNATISENKIPDSWKKANVFMLP KPGKDLTEIKSYRPVSLLSPVAKVLERVILERIK PEIQSPTSQHAFKKAHSTTTAIEVTNRIVGGLN MPKPANRTIMTCLDLKAAFDTVSIARLLCELD GAIQDKAITRWLGNYLYKRKIRTQFGEEKSKW RTLRRGGVPQGAVLSPALFNFYIREVPTIPGAIEI GYADDTTILVQDPILQKAADTSQNALDXMKT WFDERHLDISAGKSSVTIFTADTKEFNFDPLT	cd0165 0	442	698

			CAAAG	WDGTPIPMQNRTRLLGVEFDTMMKMNGQVA		
			ACCAG	KVSERLAKSNNILRALSGASWGSSKETMLKTF		
			CAAWG	KAIKPLATYAGPAWHQLMSDTQMAKLERQYI		
			GGACG	GGLKACCGLTKDTPKELVYYETKMMPLKEEL		
			WGCAG	ELGSEQFAISAMRTPGHPAKDLHTRRPA		
			ACTCA			
			GGCCA			
			ATCCC			
			ACCAC			
			GWCGC			
			GCTCA			
			ATACA			
			CAATC			
			AAACG			
			TGTGA			
			AGATT			
			CTGAC			
			ATGGA			
			ACGTA			
			AATGG			
			CTGGG			
			CCAGA			
			AGAAA			

				CAGTG WGCTC GGTGC TATCA TTAGC AAGGA GAAGC CGGAC GTTKT					
RTE	RTE - 2_CJ ap	.	Caenorhab ditis japonica	CTTCT ACACG TAATG AGCCT CTCCG TGGCG TAGAA GATCA ACACC TGGAT CTTTTC TTCAA CTCCG GCGGA	TGATCGA AGAACGG ATCGACT AAGTAAC TAAGTAA GTAAGTA A	MMHRDGTGVFLGKRNTTSVSGGVGFIVNSTL LPKIQEIKFLNHRIAYLTFQVDRRLSCTVFQVY APVADSDPEELADFYESLEEAYTACRSRYKL VGDFNARIGPRQHTERFIGPHSMEPRNESGELL ATLCESNRLWHMNSQFGKPSHRRWTHISPNKL RGNLHFNTKIARFEQVKRRKPPKRVLDPAVAH AITATATITPSADIDLIDYINLIAMKELQDQAV TQPSNHSTNRLSGTTRQLLTKRRFMDRXDPNF KSLSNECREAVKKDHEEFAKDRLQAAQNKK SMKKVARDIQEYKTFIPCLKSSTSGTRITSRTE MEKEIQQFYSKLFRSNRGPPVIASPIGMEIPPPF LPEEVRAAFRSFPNGKAAGDDKITADFLKSCH DNVILXLDRFTKYLQSGKVPEKWKTSNTTLIF	cd0165 0	447	720

			TTTCGT	KKGDKEDLENYRPICLLPVLYKAFTKCVLNRI			
			CACCC	RTTLEEAQPVEQAGFRRSFSTIDHIHSIQRILEV			
			CGGCA	GREYQQPLTLVFIDFHKAFDSEVPQAIWESLKS			
			AAGAT	QGVQPAYIALLLQQCYTNCTTTITPFHKKIEIPIT			
			TCAAG	RGVRQGDPISPNLFSACLESVFNKMSWQHRLRG			
			TGGCT	DEDDENDKANLLPGIRINGQNLTHLRFADDIV			
			AATCA	LVASNPKAASLMIQELVEMCGKVGLRMNTNK			
			TTACA	TKVMRNKFASQSPVNVTLNNVTTCIEEVNEYV			
			CACTG	YLGRLLNHNNELEPELHRRRRRAAWAAFNNIK			
			CCTCG	NTTDALSCPRIRAQLFDSIVLPALTYGSEVWTF			
			CTCAG	TQALSERVRVTHAALERRLVGISLSEQRQRNL			
			GTCGG	HREDIRAQSEVRDPLLVIKKKKLKWAGHIMRR			
			ATGTA	NDGRWTRLVQEWYPIGEKRPVGRPRMRWSDS			
			GTGTC	LKEQISLFDGNHLKTHWSTIAKDRMAWKAVIR			
			GAATA	DHIR			
			CTTCCT				
			CCCTG				
			CGCAC				
			ATTCC				
			TCGGT				
			TTAGC				
			AGGCC				
			GTTGA				

				AGAGT					
				AGGCA					
				TACCC					
				GCTAC					
				GAACA					
				GCACC					
				AACAA					
				CTGCC					
				TCAGT					
				CCTCC					
				GCATC					
				CGGCC					
				TCTAC					
				TCTTGT					
				TACGG					
				CAGAT					
				TCAAG					
				AGTAT					
				GTCGG					
				AAGAG					
				GCGAT					
				CTAAT					
				CACGG					

				AGTAT					
				CGAAC					
				GTCTG					
				CCACA					
				CTGGA					
				CGGAA					
				CAAGA					
				ACACC					
				CGGCC					
				AAAAA					
				CAAGA					
				AGAAT					
				CGACT					
				ACAGA					
				GTGTG					
				TACGT					
				TCAAC					
				TGCAG					
				ATCGG					
				TTGCC					
				ACCGA					
				TGACC					
				GACTT					

				GCAGA GCTGC TAGAG GAGAC AAAAA CGATC AAGTA CGACA TTATC GGACT GTCGG AAACG AAGAG AGCAG CTGAA ACACA C					
CR1	CR1 - 13_ CJap	.	Caenorhab ditis japonica	G	TAACTAC TGCTGAA CGGACGT CACCAAG ATCCTTT ATCACTT	VLVNRHEIQFLIDRAKPDILCLQETWLSEKTPS PTLVNPPSRYQIFRTDRAHEKRGGGVCTIVKSE IPAKVITSETKNMAYEILALDINISTTPIRLINVY RPPTCSYANTLTLLTKISDMTNXNGQIIILGDFN IKNIKWIPAGPMTKDRLGTILSDFCKNNNLKQF VNXPTRKNNILGIVFSNTKLENVEISXPIGSSDH	cd0165 0	470	724

				CTACCCC	RKITFTLDILEKPTKHLIEIWDFKGINWKVCAA			
				ACCTTTTT	ELNLIDWPTIIDPVKTISEIYNEFLTIMRRICELA			
				TTTTCTGC	VPKKRINKELYNVPEYLVTFHKTIENKFDKLL			
				TCTCATT	KSRSWADYALYIELSRKFRKKLLKFRRNQELL			
				CTTGATC	TLKKSGNSHFTHIAKNILKPKFSMVPTFVDQSG			
				TGTTTCT	HHATSDHEKAEVLA VNFESQYVNKSNTSFTLP			
				ACCCAAC	PDYKSDQEIPWVTDAEMLKFIQKCKNSCSDTP			
				TCCTTATT	DKVPFKFLKLIAPFIASPLSQICNLSMMRGEVPE			
				GCTCTCA	KWQQSIITPINKIAKPQVPTDFRPISITSQICRIYE			
				TTATTTCC	RCILSKLLPFLDKIRFWNNNQHGFRPKRSTVSC			
				CTGCACA	MLESNDWTNSIERGSQVDVIYFDYAKAFDRV			
				TACCTTT	PHDLLLDKL VSIKLNGNLLKWLASYLSNRTFT			
				AATGTAT	VKVTKNQSSTKMAQC GVPQGAVLSPILFGIYV			
				CTGTGTA	NEIPSKLPMGVHCKQFADDIKLYSEIPKDSDPQ			
				CACAATA	NNKLQQAIDAICQWSETKKLALNNSKITILISLG			
				TCAACAC	KTTKNCDYVINGSPIGRKQLTRDLGFMISPKLD			
				GTCACAT	FSEHWHKATNQAKYALSLIFKKYCSDNSRLLT			
				TGGCGTA	LLYKIFIRPILEYGTVISNPHKKSDIKSVESVQN			
				AATAAAT	AFTRRLYSRQKGRYIRADDPEYLSAPERNSLY			
				AAATAAA	GLSSLENRRKYFDRKFLTKMIAGNVDIDTRPFF			
				TAAATAA	TFFQNTRTRAKTKFIWARSKTKIRQEFFTNRTL			
				ATAAATA	TAITL			
				AA				

CR1	CR1 - 17_ CJap	Caenorhab ditis japonica	AT	TAATTTT TGTCTAC ATAGATT TTCCTAA CATCTCT TTTAAAC TTCTTTCG TTTGTTTC TAAATCT TCCGTCA TATCGTT TCTTATA CGTGTGT AATAGTT TGTTATTT ATATATG TGCTGAT GAATAAG TGGCAGC TGCCCTT CAATCAA TAAACAA	KILVSGSSKNFFKFINQRMKESSDVGMIMDGK TXVYDNYRKAELFAESFGKVYSTDDGQYTTC DLRTSAHLSEIIFEPYAVELELSKLASKVNTSPD EIPAIFLKKICTSIALPLSIIFNTSMRTGFLPDLW KTAIVKPLFKKGSRSDANNYRPISLTSTICKTM ERMIRKEMMKHIYGKNLLSNAQYGFRPHRGT EAQLLHYISCIQSNMEKKKDMSSVYVDFRKAF DTVSLPKLQIKMESYGICGRLLNWLSSFLTDRH QSVVLNEVFSPRMVVKSGVPQGSVLGPVLFL YVNDIGDSLECDTLLYADDLKIFSTDAAALQR DLDRISNWCRTWQMDVAPNKCEFIFNYSKK NSALSFPLMIGNFNLQRTTEIRDLGVIFSNNLSF KSHISVIVRKAHQRIINILFNIIRLSTIDVFIKCYVI YVRPILEYGSVVFSPISELVKTIESLQKSFIYRC YKKFGIVYNSYFEALEQNNLETERRRLMADL VFIYKWLVSREINIGHDLFISSANLLPLRRHNY YIRCQLTGATRIGSQFLSNRIIRCWNSLPVKTFP VKFSSTAFKNRLQNLSLNNFLLLHFHNF	cd0165 0	134	384
-----	-------------------------	--------------------------------	----	---	---	-------------	-----	-----

					ATACAAC AAATA				
CR1	CR1 - 8_C Rem		Caenorhab ditis remanei		TAGAGCT TATTCTT ATTCTGT ACTCCCA AACCACC TTTCCCC GCATTAC CACATTG CTCAATT CTTAATG CCTTCTTT TATTCCC CTCATAC CGGTCAT TCTTTTTT TTTGTTG CCCCACA TCTCTCA CTTTAAC CCACCCT TTGAATG	VIPIPKKGSLSSENXPISLTDPFARLFERVICE YIKLHFAHKFSQNHGFLAYRSCTSSLVHSISC YKSSLSSNNSLDVIFDFKKAFTDKVNHKLLQ KLALFGIPHLFIEWFSNFLSGRTFSIKIEDFTDTS ITQIPSGVPQGSVSGPLLFLIFINDLLLDLALIPSL QVSAFADDIKIYSSNPVAVQKGIDLIETWASSN SLPLAHTKTSLLRLGSKNISFPYFIAGQPIETSKS VRDLGLITDSTLKFKSHINKTASALLRTKQLL KSFKSTSPQFYIFLFNCYVLPPIEYCSVVYSPPPA SKLSLSLETPLRFFTRKIFQRCNITYSSYSDRLA QLNLFSLRHRRLKAQLLLLYKFLSGTSYFPHLD SYIRFSSSTRRPMNLICIKPKCSDFFSHTIPIWNA ITSQSSYFLSPSEFNTLISSITRY	cd0165 0	1	246

					GTCTCGT GAGGGAC TCATTCT CAGCCTC CCGCCCC ATATATT GATAATT TTTTATA AGAATAA AATACAA ATACAAA ATACA				
CR1	CR1 - 5_C Rem		Caenorhab ditis remanei		TAAATTA TAACTGT TTCTTTTT TTCTTATT TGTATAA GTGGTGA AAGTCCT TCAAATA AATAAAT AAATC	VQKGPNSLAVDNSTKQVKFLSVGYANCFSVK NKLAQLELVTITNNFDIVCLTETKLDNTFPDSL LSLSNNFSVVRKDRNKHGGGVAILISK SIRFLPI EIPSSLLSSEIAGVDIMAGGVSIRIVVGYHP SHH SKLDGMISCLEFLLSTRKNCVILGDFNMPHISW PSLTASDSHCKKFLSFVTKNGLTQHIASPTRLK PDNILDLCFTNTDILRDVRVGDLFSDHKLIHVT LSVKNRTKKLTKEVKLFRKADYASINYLLSNT DWVLRFSNLNADGMYENLLSILHELIASYVPV KAINTLSKRHSAEILKLQKAKLNVWRKEGNST NYKNISADLKVALIKEERRVNDEKLTNGSVKD	cd0165 0	484	735

						FFKFINSRYKDNQEIGTLKND SGAPINCDSEKV ELFSDFSFSKVFTEDNNVQPHFDKRTEELVSSPD FEPYIIHTLSKLTPKLNTTPDGIPALFLKNVCT AIALPLSIIFRESFRTSIVPTAWKTAIVKPLHKK GSRANPNNYRPISLTSSVGKVMKLVVRKQLTN YLNSNRLLSNCQYGFRSSMSTEAQLLSYQADI LTNYICKKTTHSVYIDFKKAFDKVSLSKLKIKL RSYGIHDDFFNWLC AFLTNRTQRCINNVFSS DRSVLSGVPQGSVLGPLLFLLFINDIGDAFESN YLLYADDLKL FSTNADCIKKDLVRLSVWCDN WQMGVAPEKCEVISFNHSSKKHSNSASLNF SIN DAIPQTKIIRDLGIIFNSDLNFSNHL DVT LRKA HQRVNIFFNVL RHADFEIFIKCFKIYVRPLLEYG STVFSPTSKEQVRLIESVQKTFIFRVFRKFNIPYI SYFESIKHCDILSLEHRRIIDLVFMYKILVSKEV RIYNDVHISLPRFSNNLRRHPYYIKSKLSNSTK NTCQFFTNRIISCWNKLPPHIFPSFPNSDVFR CN ISFNFVEPFLILKHSNF			
RTE	Avi RTE _CA n		Calyp te anna	CCCTC CACGG GCACC CCCTT GTCGT	G	MLDTADSGRPERRSALIAHEL SRLNIDIAALSE VHLHEEGSLKEHGASYTL YWSGKPKTERHLS GVGFMIKNSIASKLENLXTGHSDRIISLRLPLHN KQHVVLF SIYAPTLQADHAEKVKFYTDLRRLT QKVPAGDKIIILGDFNARVGKNYEAWKGVLG	cd0165 0	531	809

			GGGGG	KHGVGNCNDNGRLLLEFCAEQQLTITNTIFQQ		
			AGGAG	KDSLKTTWMHPRSKHWHLXDYVLRQRNVG		
			CTTGC	DVRHTRAMPSAECHTDHRLVRCKLTLYFKPK		
			GTGCC	PKRGGIPRRRLKVNDLQKDTVNRNSFQVNLQAR		
			CTTGT	LLDHPIDPSPEALWQHIKNSILQSSEESLGFSK		
			GAGGT	KNKDWFDENNQEIQELLTKKRTAHQAHLAQP		
			TGGGA	SCHVRKAAFHLACSKLQQKLRDIQNKWWINL		
			GCTCT	AEKTELCAVTGDHRGLYEVLKAVYGPTHQVQ		
			GCTGG	SPLLSADGQMLLTDKTSILNRWSEHFQTLFSAS		
			CGGTG	CVVQGSIAQHITQQPVKTELDIAPTLGETLKAI		
			GCATA	QQLKTGKAAGVDGIPSEVWKHGGQALHAKFH		
			TGCCT	ELLVSCWEQGKLPDLRDAVIITLYKKKGEKS		
			CCAGT	NCSNYQGITLLSIAGKILARILLNRLVPAIAEEV		
			AGGGT	LPESQCGFRANRSTTDMVFLRQLQEKCREQN		
			CTCCC	KGLYVTFIDLTKAFTVSRKGLWQILERLGCPP		
			ATGCC	KFLKMIILLHEDQQGQVRYGDALSESFPINGV		
			AGACA	KQGCVLTPTLFTIFFSTMLQRATADLDEENGIY		
			GGTCA	IRYXTDGSLFNLRLKAHTKTLNHHVPELLFA		
			CAACT	DDAALVAHTEAALQRITSCFAEAADLFGLEVX		
			GAAGG	LKKTEVLYQPAPQEDFHHPHITIGESSELKSVQQ		
			GTCAG	FSYLGCISSDAKINKEIDNRLAKAYRAFGRLH		
			ACAAA	KRVWSNKHLKKSTKISVYRAIVLSTLLYGSES		
			GTGTG	WVIYRHHLRLLERFHQCCLHTILNIHWTDCVT		

				TCCAT	NVPVLEQAGVTSIEAMILRTQLCWAGHVSRM			
				GGGCA	EDHHLPKIVLXGELATGCRNRGAPKRRYKDSL			
				GGATG	KQSLSFGHTDCHQXSTLASSRETWRHTIHNA			
				GGCTC	ASFXNTHRISLEEKRQHRKNRALAVPSQETFC			
				AGTAG	CAFCXHTACPTLALSAISVLAASVGRALLKSSF			
				CCTTC	VKPSH			
				AGGCA				
				ACCAT				
				CCTAG				
				GAGAA				
				GGACA				
				ACTCT				
				AACTC				
				CAAAC				
				CCGGG				
				CAGAT				
				GGAGC				
				TCGTTT				
				AGCCC				
				TGTAA				
				GGCCA				
				TCCAT				
				CTAAG				

				AGAAG					
				GATAC					
				TCTAA					
				CTAAA					
				CCTAC					
				GTCCT					
				GAGGA					
				CTTCG					
				CTGTC					
				ACTGT					
				CCAAG					
				CTTGC					
				TGGGC					
				CCTGG					
				CAGAT					
				GAACC					
				TTAGG					
				AGTAA					
				AGGGT					
				GGGGC					
				CAGTT					
				CCGTG					
				CACGC					

				TGCGC					
				CTCAC					
				CTAAA					
				AAAGC					
				CACAG					
				CGCAG					
				GCTCG					
				AAGGG					
				TGAAC					
				CCACA					
				CTTGT					
				AAAGC					
				CCCTG					
				CAGCG					
				ACAGG					
				CGAGG					
				GWCGA					
				AACGG					
				CAGGG					
				GAAAG					
				GTGCA					
				CTGGA					
				AGCCG					

				CAGAC					
				CCAAC					
				CCTGC					
				ATGCA					
				GGCGG					
				CTCAG					
				GACAT					
				TGGTC					
				GCTCG					
				AGACT					
				GACCC					
				AGGAG					
				ATGAC					
				AGTCT					
				TGTTC					
				GGCAG					
				CATCC					
				TGAGT					
				GACCA					
				AGCAG					
				CCTTTT					
				TCAAG					
				GAGAG					

				CACTG					
				CTTGC					
				TCCAC					
				ATGGA					
				GAGGG					
				GCCTA					
				GAAAA					
				GGTGG					
				CCTAA					
				CCAAA					
				GCTCA					
				TCTCC					
				ACCCC					
				CCACC					
				CGGTT					
				GGCTA					
				SCCAC					
				GGTCA					
				ACGGG					
				CATCT					
				TCTGA					
				TGCAG					
				TCGAA					

				CAAAG ACAAA GAGAC AAAGG CATAC ACCTG CCTCT AAAGG TGTGT TCAGA TTGAC ACTCG CATGT TGGAA CATCA GAACC					
Dap hne	Dap hne- 17_ CTe		Capitella teleta	CACGC GGCGG TGCAA CCGCA CCAGC GAAAG	TAGGGCA ACTACAA GATCTTA TTCCTAT GTAAAAC ACCCTAG CTACTCT	MTSLTTMAAIATTHARTDWMTNQPNMAERT DNDSL VIGTYTATVHLDSTDNDALLRSHYNLN QVLQTNQQEQDNTPINFSPYYDHEHLIQHLAQ PTNNVKFKILSLNCQSLNAKFDQLMILIAQLNH NXVIFDAICLQETWIKESLPPVFNIKGYQAIHQ ARHCSGHGGLLIFLKDSYKFKIIPFEKSAVWES QFINIDFENGKDSLLVGNIYRPPHKVKDDLDL	cd0165 0	615	877

				ACGAC	TACACTT	FSNQFTRNLLQC	SKYKNIAIAGDF	NINLLDVNS			
				G	TTTCTTCT	TTSPA	AHDFPSSII	SNDFLPKITL	PTRISESS	TLI	
					CTTTCCT	DNFFCKL	KNDSTSSG	VVIHKISDH	QAYFTCL	PI	
					ATTTGAT	RSNKI	HNIEKIKM	RRVVSQES	IIAIGNDIR	NATGR	
					TAGTCCT	LDLST	ETDPNVN	YEALMETL	HDSMSQHI	PIKE	
					CCTAGTT	VKFNR	HKHKGSS	WITAGIVR	SIRERDKL	YCKLI	
					CTTCTAG	KAKPNS	NKRESIKL	TLQLYTRIL	QNIIRKAK	SSY	
					CTCTATC	YSMHFA	QLKNDM	KKTWKA	INDILHRN	KNLSN	
					CCCGCTC	RNYPEQ	FKINDQL	ISDKNIIA	EEFN	SFFT	SVGKIL
					TTTATTCT	ASRIAH	NSTNTAT	DYLSGN	VTSRFL	QKVSES	
					TTAACCG	DIERIIA	QLKSVSS	CGYDMV	TPKLLKA	LSPFLIM	
					TTTTGTGT	PLTHVN	QAISTGI	FPDSLKRA	KVVPIYK	KRCPI	
					CTCACTT	KKDDET	VLTNYP	ISLLPAIS	KIFEKVIH	NQLQH	
					ATATGTT	YFTSND	LLYHSQH	GFRPKHST	EALQLLD	QIT	
					TACCCTG	TDLDAN	KIPIAVY	LDLSKAFD	TINHNILL	KKLN	
					TTTTCGG	YYGITG	TAYEMMK	SYLTNRK	QYVSFAD	CNSR	
					TACTTTCT	DIMIDT	GVPQGS	ILGPLLF	IIYINDL	HKSCSN	FLP
					TTAACAA	IVYADD	TTLYKT	INSPNS	INSINQ	ELEHVNL	WL
					TTTCTATT	KIXKLS	LNASKTRY	MVFHKPR	KRLPTLN	ISIDN	
					TCTTTTGT	ENVERC	ISFN	YLGHLN	QNL	TWNDHI	KRTSLKI
					ACTTTTA	ARATGR	INCLKH	FLPKILL	TLYN	SLILPH	LT
					TAACCTT	GILAWG	AKSNDIA	KLQKKS	SVRIIT	ASKYNA	HT
					ACTGTAC	EPLFKA	HLLKIEDI	YKVYQL	NFYFKW	KNGNL	

					TATGTTT GTATTAT TTTTGTGA TTGTATT ACCATCA TGATAAT CATTCAA TTCTCAC TCATTTG GTGGCAT ACATATG TTACGTT CAGTTAT GAACTTC GTATGTC TCCTGAC CCTTGTC GAATAAA TAATCTA TCTATCT ATCT	PKYFNSFSCNTNRDTHSYHTRTTSTVSARVNH EFAKHSLRHSLIILLNQTPNIITDKITTHSYHGFS SYVKNYYINTYTNVCYNSTCYICRNQQQL			
CR1	CR1 -	.	Capitella teleta	GAACA GGAGA	TAAAATG TATCAAA	MTMYIKDHIDYTMLDIGEDTCEVQAARLLME QGQATIFASVYRSPNSSKENSNDNINAFIRNFSNE	cd0165 0	423	675

6_C Te		AAGAT	ATAAAAT	RAYTLIAGDFNYPTIGWKTMTGRTEADIEFIEA		
		GCCAT	AAAATAA	LRDGFLIQHVDRPTRGRGSNNPSLLDLVLTKE		
		GAGAA	AAGGAGC	HTATPTVEYHQPLGKSDHSILDIRIEMKQSETFI		
		AGAGA	TGATCGT	KTIPNYNRGNFDRMRRVLSDVDWEEGLGREE		
		AATGT	AAGAGGC	TTQKKWELFHETITNAVNRCVPKIKIKRKVLK		
		CATCA	TATTTAA	NIPVDNKTLTKIRRKKRLWKQYLETGNQTIYH		
		TACAC	GCCTGAA	DYCRTRNQVRQLTRKNMRELEKNIAKEAKTN		
		GGACT	GATCAGA	PKKFWAYANQKTKYRETIPQLLKTGNPNDGM		
		GCCAG	AGAGGAC	LTENNQEKAQTLATFFNSVFSEKPNDHWNLS		
		AAAAG	CTGTAAG	SSYPTIEQDLITEASVLHHL SKLKKTKSPGPDG		
		AAGAC	TACCTGT	IHPRIEIKEIIAKPLSIIYQHSLDQGGQLPDTWK		
		CAACA	AAGT	TANITAIHKKGDKRIAGNYRPVSLTSVPCKIME		
		ACAGA		KIIRKSIMDHLIRNNILSNKQYGFITGRSTLLQL		
		GAGGA		LTVLDQWTEILDRGSEIDVIFLDFKKA FDRVPH		
		GAGAA		QRLIDKLEY YGINGKILTWISAF LINRKQRVTIH		
		AGAAA		EATSAWAEVKSGIPQGSVLGPTLFII FINSLPDV		
		GAAGA		TSSPTFLFADDAKIFREITKPTDSL LLQEDINNM		
		TGAGC		YHWTLESLLTFNPDKCSSMTISRSGTNHRQYQ		
		AACAC		MNNSPLNKTIERDLGVILDNRLTFDHHIQAKI		
		AATCG		KKSNSIMALIRRTYTYLDNHTFLLLYKSFVRPH		
		AGAGG		LEYANQVWKPHLQKHITAIENVQRRATRILPG		
		ATTAT		MKDLNYESQLRKLKLPTLQYRRIRGDLIETYKI		
		GGACG		ATEYYDTQITQNFQFNSRSTRGHPLKLYHQP		

				CATTG	ARTELRKNSFKIRIAQLWNQLPSEIVASPSIATF			
				GAACT	ERRLDHHLKNHPKHNPFRPP			
				GAACC				
				TAGAA				
				ATTGA				
				AAACT				
				TCTAC				
				CGACT				
				GGGAA				
				AANTA				
				CCCAA				
				TGAAG				
				ACGCC				
				GACAG				
				GGAAC				
				ATACT				
				ACTAG				
				CAGAC				
				CCCTG				
				AAGTT				
				GTGCT				
				TCAAT				
				AGTGA				

				AGACA				
				AGGTG				
				AAGAT				
				GTTCA				
				TGGGG				
				AGGCT				
				GAACA				
				AATTA				
				AGAGA				
				TGCAG				
				AGGAA				
				GATAT				
				CAAAC				
				GAATT				
				AGAGT				
				TCAAC				
				ACGAC				
				ATGAA				
				TTTAC				
				AAGAA				
				AGAGC				
				AGAGA				
				TTAGA				

				GAACT					
				GGTAC					
				AGGAA					
				GCAAA					
				AAACT					
				TAACA					
				TTAAC					
				AGAGA					
				GTGGA					
				GACTT					
				TGTAC					
				ACGTC					
				GTGAG					
				AGGAA					
				AAAAG					
				ATAAT					
				CAGAG					
				TGAAA					
				GCCAG					
				AAAGA					
				ACGGT					
				CAGAG					
				CAGGA					

				GAGAG					
				ACCCA					
				AGAGG					
				ACTCA					
				ACATC					
				TGGAT					
				CACGA					
				ACATA					
				GACAC					
				CCTTA					
				CAATA					
				GCAAA					
				ATTTA					
				ACGAA					
				TTAAA					
				AACCA					
				TAATA					
				AAGGA					
				ATCGC					
				CACCG					
				GATAT					
				CATAG					
				CCCTA					

				CAAGA					
				AGTTA					
				AACCG					
				AAAAA					
				CGCAC					
				GTACA					
				AACTG					
				GGAGG					
				CAAAG					
				AATTA					
				TATAA					
				TACCA					
				GGATA					
				CAATA					
				TAGAA					
				GAAAG					
				GATAA					
				GAGAA					
				GATGG					
				AAGAG					
				GA					

L2	L2-6_C Te	Capitella teleta	T	TAAAAAG AAGTTAA AGACTTT TTTTATTC AATAATG CTTTTAA TAACGCA GAACAGT TTTAAGT GGAAATG AACTTGG CTTGTTTT TATGTTT CTATGAT CTTTTATC CAAATAC ATTTTCA TATGTTTT ATTTTGT GCTTTGA AACTTTT ATGTATT TTAAGCA	QLIKDKKAAYYKKLFSEANHKETFQLLKGTTK PPAKVLRQEQEDVTLAFEFQQFFNTKIRKIHDT LQTDENNSPHSHDKRTSASFTAFRTHQTQSQVE KIIKSAPSKTCELDPMPTWLLKEQNVLDQVLP WVTGITNSSLELSHVPKDHKTANVKPRAKKD GAAYFDLSSYRPVSNLQFLSKVLERAVASQLT EYLEKETLMDPLQSA YRRGHSTETA VLKVK N DIDLAIDEGDG IILVLLDLSAAFD TLDHAIL LKR LEKCLGIKGSALDWIESYLRNRPQRVKVGESY SSDSILQTGVPQGSVLGPLLFLLYVLPLQEVISR FRIDRHHYADDTQLYGRIRLNSIKTIDEDVREM EASIQDVRIWMKKKNKLKLNKKTEALIITSKE MKKHAEGIKLKIGGDTV TATSKVKNLGGWFD EHMKMQEQVNRTTKAGYFHLRTINRIRKNLD TPTCAKAINATVTSRLDYHNALISGARQCIMKP MQRLQNHSARVITMTDKREHIRPVLEDLHWLP VKERSDYKVLTFVHNALHDELAPAYMKEMY RVYHPNRALRSGEDRWTLTVGTAARHEGHRS PXIHGAIMWNTLPEALRSEASHPRF	cd0165 0	150	418
----	--------------	---------------------	---	---	---	-------------	-----	-----

					CTTGAAT ATAATAA CATTTAT TTATTTAT TTATTTAT TTATTTAT TTATTTGT TTATTTA				
Nim b	Nim b- 2_C Te		Capitella teleta	GCGGC GGAGA TTAAA TAGAT ATGGT GTCTT AGATC CCGAT GCCAT TGATC CTTCTC TTGAG GAGGA CATGG CCGTG	CAA	MTILQWNCRGFRANFQELKTLVAQKDATVLC LQEIKVKDSDSCSLNGFNTLKKCTLGIDGRAV GGAAVLIKKDTPHQHIPLTTNLQAVAVRVTLH RLLTVCSIYLPGRSHGDLVELDRLVQQLPPPL LLLGDFNAHSELWGSEEPRPSGRVVEDFIAGN DLSILNTGSQTYLHPASGSFTVIDLSLCSPSAHI DFTWEVDTDQHGSDHFPIFISSECRPSLSHPKW VFKRADWTAFGSLCEIEITEDIFEDVDDPLDIFT TLLIKIAHETIPKSCSSTKCRKPWYTDECKKAIR LRRALQKFRHHPNDDNLLSYRKAQAVVRRV IKASKRASWKNYISKLTIRSSVKKTWDMVRKI SGKYKGHQIHHLTTRSGDTATDKESIAEALAD EFALNSSTSHYTDKFQKYKDKAERKVLNFNSN NTESYNLPFTLLELLMSLKSSHDTSGPDDIHY QLLKHLPINCKMILLNLFNRIWVGEPFPSAWR	cd0165 0	489	746

				GATGT	NAIVIAIPKPGKDISLPTNYRPIALTSCCLKTME			
				GCCTC	RMVNGRLVYFLEYNNLLSPSQSGFRSQRSTAD			
				CTTCC	HLVSLETYIRDGFAHGDHVLVSFFDLEKAYDT			
				ACTTC	TWKCGIISDLFNMGLRGHLPVFIRNFLSHRQFS			
				CCTGA	VRLGSVDSSPHPQEMGVPQGCVLSTLFSIKIN			
				CGAGC	SITNAVGAIDHKCLYVDDFTISYRSRYLPTAER			
				CCAGG	KIQLALNGLQKWADESGFRFSPTKTVSVCFSG			
				GAAGA	RRGVFPDPLLTLNGQKIPIVEKTKFLGVIFDKK			
				TGCAA	LNFKHEHIQFLRGKCQKVLNLLRVISKMNWGA			
				TCCAC	DREVLLRLFRSLLRSRLDYGAIVYGSARISYLE			
				AAAGG	RLKPIQNQALRLCLGAFRTSPVTSLHTEANEMP			
				AGGGA	MAIRQKLLSVQFALRVCSDTTNPAPHQCIFNYN			
				CCCAG	NDRFYLSKPNAIRPLALRIKEDLQTICPDIAIT			
				CCCTG	PNQLFNTPYWLLRPPELDSLIFGKKTANPNY			
				AAAAG	TLRNEFYNLMDKYPDHKVIFTDGSKSDSAVAC			
				GGCCC	SATADNLRIQIRLPDSASIFSAELLAIIYQVLTL			
				CTCCT	ECSANDQQQFLIATDSLSSLQAIGNFNIKHPYV			
				AATAG	FKILEKCTLLHKKGIDLVMAWCPSHVGVMGN			
				ACAAG	ERADLLAKEALSFTTCTIRIPSSDFKPITHEFYK			
					EKWQEQWSSEQENKLYCIQPTLGKWAKSSREI			
					RREEIVLARARIGHSHLTHGYLLRREMPVVCIP			
					CQSILSVKHIFIECVDFDILRRRFFNVSSMK			

CR1	CR1 - 13_ CTe	Capitella teleta	TGATTTT ACCCTAA TTTCCTTG ACACATA TTCTTTGT TGCCATT GTATTTT TTAAGCT GGCAGTT CTTATAG GTTTTCCT GTTCTGC CATGCTG TTGCATG CATTGAT TTTATTGT ACATATT ATTTTGT CATGCAT GAAGCTT AATAAAT AAAAAAA TA	LDCPQIPDPHFTVENVLSVLASLDTNTAFSPQD IPPIILKKCRSLLAPSLCIIFNVSIRCKTVPLEWK KARVVPVFKKGDKSAVSNYRPISILCIVSKVME RCVFNHMYPLLSSFHPLQHGFCKGRSCTSQLL DVYNRIGSILDKGGQIDVVFLDFSKAFCDCVSH KLLLYKLKHFYGVDGNLLEWIRSYLSERSQSV IVEGKSSTWLGVPSGVPQGSILGPLLFLLFIND MPYKTLHCTTALFADDSKCFKSINTTHDCELL QRDLNAMHEWSLIWQMSFNPSKCKVLSITRSH APIHFNYAMNGVILDHVN VF KDLGVFIDQTLS FSTHIESTILKCNKMCGIVKRTIGFNAPVNVKSS LFMSLCRSHLDFASPVWSPHLKSKVKQIESVQ RSMTRFILNNADLSYIERCQFLEILPLSFRREITD LLFLFKYFNGAMNVDFTSIVAKRCFNTRLRSG PFIELLTPTRTDSFINFYFN RVAHMWNILPLEI RSCLSVSTFKSKLLSFYSNKMCFYDVNNSTW TSTCRCNGFYHT	cd0165 0	69	322
-----	------------------------	---------------------	--	---	-------------	----	-----

L2	L2- 9_C Te	Capitella teleta	AGATC AGGTA GCACC CTCCG GTAGC ACCTA AAATT TGCTC CGACC ATTTT GCTGT TTTCC GATAT CTTCC ACTCC GCGGA TTGCA GCGAA TCTTG GTCAA GAC	TGATGTT TACATGT TCCCCCT TTTTATTT TCAAATT GTATTAG CGTTCTG AGCATTT TTATGGA AGCGACG CTTTATA AAATAAA TTTTCATT TTCATTTT CATTTTC ATTTTC	MWKVIFVICVLSQIKAKTGSGETRPGFTICLPQ PRHASPTSMPSLCLEFIRALAVKAPARLPTDLY KRLNGLRISRRRLSKRGTKGYGRGVIPVQVSA LRSADRYVNTARRDASLPTSLVTIEPMTIASSS SCVLLNCRSICNKASEIRDFISDRRMDFLCLTET WLTGTDEDKVVSAALLPDDYAMVHRPRDAR GGGVAFIHRRSITLSQIRLG DYSSFEVLDALVC TARRLRMCVVYRPPPSRRNNLTVAQFMSEFA DFLSDIVSLPGEPLIVGDFNFHLDDCGDADAST FSDLLTSFGLKQHVRGP THEKGHTLDLVITKT DTALIWSSVSRDLCISDHYPVFTTLAIPKVLNPT REISYRRIKAIDKEQLDRLIEASSLSPPENASLD ENV TNYQRTL TGILDTIAPLSRTVRIRVNADW YNDEIRAEKQRRRQAERRWRKSKLPADRELF KVAKNRTQSMIDEAKSLHFREKVSACSDDPKQ LFRVVD TLLGNSTESPLPVHDSRVELSNRFSDF FVEKIATIQSAIPSVDNPPRPQAPLPQSQLATFQ PLDVADTCALISGAAPKSCELDPLPSPLVKDLP SLRPIITRIINQSLATGVFPMEYKSAVVRPLLKK PSLDRDSLKNYRPVSNLSFVSKLIERAVASQLA SHLRNNDLMEPRQSA YRPGHSTETALLCIHNEI VCAIGEGKAVLLVLLDLSAAFDTVHHPTMLSS LERLGVTGTPLKWFESYLSERSQAVQIGDAKS	cd0165 0	620	878
----	------------------	---------------------	---	--	---	-------------	-----	-----

						DPAAINCGVPQGSVLGPILFTIYTSSLGALLREQ GVSYHLYADDSQLLLTCDPRDPAPAVARMEA CIRSVQAWMSCHSLKMNGDKTEILLITSKPMR KVLPDVQLCIDGHQLQPSTTARNLGVTFNHRA SMESHITSVCRLSYMQLRRLRRIKKLLPRDVLE SLVHAFVTSRLDYCNLSYLGIPQTQLDRLQRIQ NSAARLVTD SGAREHITPVLQDLHWLPVKERV IFKVLLIVFKSVHNLTPQYLSEMITVHIPTRHLR SSNQNFLTVPFTRSSFIKNSAFSVAGPTIFNSLPI NIRQAESVPIFKRLLKTHLFHNYFLT			
Rex 1	Rex 1- 1_C Te		Capitella teleta	GGTGT ACAAA CTTCG TCATC AAACA CACTT CGTGC ACAGC TGATG TTGAG ATCCT ATGTG TGCAG	TAATGTT ATTATCA TCATTAA TGGCTTA ATTGTCA TATTCAT CATTCAT TTTACTC ATTATTA TCTGTTT GGCAAAA TCATTCC TTCGGGC	MIYCLPTYIRKLERQKPMIKTIRQWTEDAAEQL GGCFAC TDWNIFEDTTTNINEYTNVITDYIKFC EETIIP TKEIKIYPNNRPWVTNDLIHAVRKRRRL WMSGHHEECKEAQKEVNTIIAQCKDNYKKKI EESFRSGNPKKTWSGLNTITGYKKKPMTRTSIS NDEGWANELNVFYTRFEKEDEDLVFDQPAIAP IRFEEEDVRRVFMGIKQDKAQGPDGLSPRLK ACAQDLAPVFTGLFNRSLEESCVPTLWKCSTIV PVPKRPTAKELNDFRPVALTSIVMKSMERLVL RRL LHQVQLSMDSNQFAYKARRCVEDASALM VHLVSEHLETPGSYARVLFVDFSSAFNTMRPSI LISKLEAQGVAPPLCQWILSYLKNRPQQVRVN QSLSTIRHTNIGAPQGCVLSPVLFTIYTDEHRAT	cd0165 0	257	510

			TGTCG	GAATAAA	SQETRVIKYADDTALVALLSPGKEHLYTDAVS			
			ACCAT	ATCATCT	HFIDQCDNDDLKLNSTKTKEMCIDFRTHNTIIT			
			TTTATC	TATCTTA	PIIIKGQPIEQVTTYKYLGITIQSDLRWTS HISTQ			
			TACCG	TCTCATC	CKKASARLYHLRKL RQFHVDRHLQELFFTATI			
			CGTGA	TCATCTC	DSLLLFGITVWGG RCTARDRNNISRVQRIAGKI			
			GATCA	A	MNTTVQPWGVSHRQRTIMRAKKIIREPLHPLH			
			GCTGT		YTFQRLPSDRRYRAL KAKTNRLRQTFVPSAIRL			
			GTCGT		INSQPI			
			CTTCA					
			TCGTG					
			ATCTA					
			CATCC					
			CTCCC					
			AGTGG					
			CGATG					
			CCAAC					
			AGAGC					
			CACGG					
			AAGCG					
			ATAGC					
			GTCTG					
			TTGCTT					
			TGGAC					

				ATACA					
				ACAAA					
				CAAAG					
				CCAGA					
				GGCAG					
				CTATC					
				GTCAT					
				CACAG					
				GAGAC					
				TTTAA					
				TGGAG					
				CTTCG					
				TTACA					
				TGACG					
				CCCTC					
				CCGAC					
				CTACG					
				TACAA					
				TACGT					
				AAATT					
				TCAAC					
				ACAAG					
				AGGTC					

				GCAGC TGTCT CGATC TACTG TACAG CAACA TAAAA GATGC CTACA AGACC ACAAG CCTGC CACCC ATCGG CAGAT CTGAC CACAC G					
L2	L2- 3_C Te	.	Capitella teleta	GACTT CAATT TTCAA CAACA TGTGA	TAATGAT TTTCCTTT TTACCTC TGTTTGT ACAGCGC	MSLTHEHLSIKVIDVDRSVTSDHFGVVFTINCP KPAREKQKLKVRKWKNVNISALDETIASHIPAI ESAQDPFAAYNNCLTQVVDEFAPETTIEVTKR TDTPWYNAELRHLKTECRRLERKWRRTQLKI DQEIYRKKCRCLRRLRDRTKMTYISSKLQNAT	cd0165 0	306	560

				CTGAT GCAAC ACATG TCAAT GGTCA CACAA TCGAC CTAAT C	ATTTGAA TGTTCTTT GATATTG CTGAAAA TGC GTTT TAAATGT AATAAAT GAAATGA AATGAAA T	STKDTYNILNQLLHKQSSSPLPRHDDEIELAET FSSFFSEKIERVRSAGHPDPDVTIEHHLSTEIT PLTSFRPVSRDWISKTLSSSTPKSCNLDPIPTAL LRKCPSAVPLLCTINWSLANSTMPDSHKTAH VRPRLKKTSLDHDQLKNYRPVSNLSFISKMTE KAIVAQLKDHC FVNNIDVQM QSA YRAGHSTE TAMVRVQNDLLDAVDRDGGAILVLLDLSAAF DTLDHPLILHALKYNIGVTDSALKWFESYLTN RKQLVTIGSQTS DTKNL VYGVPQGSVLGPILFT MYTRPLSNIISNCGMNHHL YADDTQLYIAVNV RSPTSVSSVKESIQECSNQVHRWMTANFLKVN EEKTEVLVITKPSL TEIAPSSFNIIGETITSSASVR NLGVMLDNKLSMADQVRNICQKSFFHLHMIR KIRNFITEDAARQLVHANVMSYIDYCNALLAG SPSYLIDRLQHIQNCAARL VKCADRHSPSLPLL KELHWLPIKFRIMFKINVITFKALRGLAPPYICS LINVYTPARSLRSSNFTLLSVPKFHLQSYGYRT FAHQAPILWNRLPSNLQNQQSLDAFKAQLKTY YFNIAFSSI			
L2	L2- 13_ CTe	.	Capitella teleta	C	TGACATG GCTTTGT GAAGCGC ATAGAAA	DDLNDLVSSFNRRCLDVTNAHAPLRQITLKGD GMKPWYSDAIHSARVRRKYERQFRKSPLEV HRQIYVDQCKEVVQLIQNTKKEYFHQKFASAS AKEVFRLVDNLLHKEPNHTLPTNVPLGDLPQT	cd0165 0	233	497

					TGCATCA FNKFFYDKVHQIRAELDASPTLPSLTTPQPLVA TGCATCA ERGGAPALETFEICSAEDLRKIIQTSGSKTCMQ TGTTGCG DSIPTPLLKEGAVLEAALPAIVNIINRSITEGCV CTATATA AELKKAVIAPILKKTGLDPNNLAHYRPVSNIGF AAATAAT IAKVMEKVIARQLTDHLTRHNLHDQLQSAYR AAATTCA KGCSVETSLLYIKSDIELILDAGDAVLMILLDL ATTCAAT AAFDTIDHALLTRLSEIGLEGSALAWVKSYL TCAATTC MERTQVVRIGDKISESSDLDIGVPQGSVLGPLL AATTC FLIYMIPLGSIIDNRGVLRHSYADDTQLYCRVN IKDPTKLQKQIRDMEECITSVRSWMKSNKLKL NDAKTEVLLISRHEPRQPVEIRVGNAVIVKAPL VKNLGAYLDCRLSMKDQVNRTTSAVYYHLQ RIAKVRPFLSDSACANAINATVTSRLDFQNALL MGLPDCQLRKLQVAQNNAARLLTRTKKREHI SPVLTALHWLPIKHRIEFKVLSMIHKSFHSL SAPDYLKKLCRPYVPSRNLRSRDDPFKVVQLRTRN LYGARTLGGLGARLWALPMSLRKPMFTSTF KKNLKTHFFRISQ			
Dap hne	Dap hne- 10_ CTe		Capitella teleta	CT	TAGTATC KMFLDLMHSHAFFXLIDKPTRITATTSTLIDNIF ATCCTCG SNVINXTHLSGVLYTDISDHLPLFTVNQFICRN ATCTACA NSIVTKTIGYRSFNETNKLAFQNAPLAIDDXL ACTAACA FXCHNCGAAYTLFTSNFVDCYNESFPIKSHKIN TATGAAA PKRNQPWLSPELKRSINRKNRLYRRYHARPSV	cd0165 0	337	594

				CGAACTA	FNEINYKSYKSMITKEIRVAKRTYYHDLLEQN			
				TTTCACT	RNKIXKTWDIIKEVIGIDNTKKPIASITINDSTVS			
				CTGAATA	EPIEIAQHFNKYFSSIGNNLASSLPINHLDPMARY			
				TGCGCTG	MNPNTPRSLFLSPVTSNELYTHIFNLKDASAGID			
				CACCACC	DIKPIIIEFSQLIIPLLHIFNLSFTQGSVPCAMK			
				CACTATT	KANVTPIFKGGDPTSLGNYRPISVLPVFAKILE			
				TTTTTCATT	KLMYNRVYNYLTENRILFNKQFGFRRGYSTE			
				TTTTTTTC	MALLTALDELSSALDMKRVIGVYLDLRKAF			
				TCTCCGC	DTVNIDILLAKLFKYGIRGIAHQWFTSYLNDRS			
				ACCAAAG	QRVKIGCSFSEMKSISCGVPQGSTLGPLLFLLYI			
				CTGAATA	NDLPNVLTNIFPIIFADDTSLFMTGVDINDMTE			
				AAATTTT	AFNHQLSKINEWLIANKLSLNIQKTYSMFLTLS			
				AAATTGC	PHLYSSPIITINNTEIQRVHSTRFLGVIIDRLT			
				ACAATT	WANHITHVSNKIAKSIGIINKVKHSLKQDTLMI			
				GAACAGG	LYYSLIYPYLYCYFLLWGKASHSHINKLTILQK			
				CGACTCA	RVIRIITNSGFRDHTRPLFNTLRRIHMDDLFTYL			
				CAAGTAG	CSVYIFKCIKIIFPRSFLDILNYFTDYNSENAPRSS			
				CCATACT	TRASLTRSIIVPLCRTTLRQKQLRFQSIKLYNNF			
				TAGTCGC	LMPLEILDNSNSISHFRRFLKNVLS			
				CTATTGA				
				CATGACC				
				CTATTCTT				
				TTTCGCTT				

					CTGTATT				
					TCCATTT				
					CTCTAGT				
					ATGTATT				
					TCCATAT				
					TTCATTG				
					TCGCTTC				
					CTGCTTT				
					CTTTACC				
					AACCTCC				
					CTTTTTCT				
					ATTTTGG				
					CATCTTA				
					TATTTGC				
					TAATATT				
					TCATTCTT				
					TCTGTCG				
					TGTCAAT				
					AAATGCA				
					ATTCAAT				
					TCAATTC				
					AATTCAA				

					TTCAATT CAA				
L2	L2- 11_ CTe	.	Capitella teleta		TAAATAA ATAAATA TAAATTA ATTAA	KRNQLTRATRRAKIKYYSDKINESASPKEIFQV TSFLLQSKREVLPKATTMSSGSLADEFAEFFE NKITAIRNGISAAPVQSSPLSDVVALHTFALTN AEEVSTLIRKSTSKSSAVDPIPTWLPKKCILTFS PILAAIVSASLVSGCVPDQLKAAVVTPLLKRPT LDPEVYKNYRPVSNLSYVSKLIERVVAERLKE HCALNNIQCKFQSAYKQNHSTETALIRVHNDI LRAVDRDGGAILVLLDLSSAFDTIDHTLLHHL EHQMAIGDTALKWFSSYLSERTQRVKVKESLS SPRRMLYGV PQGSVLGPVLF TLYTSPISAIIRAQ QLQYHLYADDTQLYMSFTPTSAASTTSSGASM ESAVSAIKRWMDHLYKL NEDKTEMLVIQRP NPRSSQPTHRPTLTLDGNNINPSTSIRNLGVLFN SAFNLEDHVNALCRTMFFHVNRIWRIREYLT SATRALMQAHVISRLDYCNALFYGLPVVVRN RLQKAQNAAARVVVRAKRSDHITPILLQLHW LPVEFRVQYKICLMTYKALHGLAPAYIDELISV HRPTRRLRSSDIRISPVQYKLKSVGGRCFSSA APSLWNELPQSIRLSSHLL EYKSKLKTHFFRKA FNDIYLNKSTSYHFIYILLYIAFEQFYFFLNFAIH KFEVINK	cd0165 0	155	425

Prot o2	Prot o2- 1_C Te	Capitella teleta	GCATC AAGGA AGGTG TGCAA GGGAG GACGA ATATT GTTGA TGGTC GAGTA GGGGA ATCTG AGATC TGTGA TGTGT TCCAT GATGG CTATG AGCAT CTGTA TCAGT CTGTG CCTTCT	TA	MSDHFHIGCELYVHLSLFLNAMLHLDVSSV MSRSVLVPIPKNRKKS MN SASNYRSIAISSIVG KCLDNVILVKHADVLSSCDLQFGFKRTHSTTQ CTFVLNEVVD FYMRGGSSVYVTLLDATKA FD RVHYTKLFRLLVDRSMCNATAKLLLCMYVNQ SLVVRWQSVTSDSFPVTNGVKQGGVLS PMLF CVYIETLLERLRRSGIGCYIGSTFAGAIGYADD VALLAPSLIT AQMLNVCEDFSREFFVQFNSSK SKAIVFNAPPNAVRS MRKLT L NGSPIDYVADA LHLGHVVGDKSASLNMENACRDLVVRVNML CSNFNFVHLD TILL LFNSYCTSYYGCSFWDFSP RTINGFVVCWRKCVRRLDL P YRTRSKXLHLI VLKPEPRHQLFFRFQSFYVGCTTSTNHIVSLCT RLLPFSSSSVACNLRQLMSFVNINYDQLNNLID NHSIRKRLSEFCEATTDQESLCCASSVIELMKIR EGSLISPINPTEATDLLKHMCCSND	cd0165 0	35	295
------------	--------------------------	---------------------	--	----	---	-------------	----	-----

				GATCA					
				GAGTG					
				AACTC					
				GAGGA					
				ACTCC					
				AGCGC					
				GAACT					
				TTCTG					
				CGGGA					
				GTAGA					
				TCGCA					
				CATGT					
				AAAGT					
				AGGCT					
				TATGT					
				TACTG					
				TGATC					
				ACGTG					
				ATCAG					
				CCCAC					
				GTGAT					
				ATTGA					
				GTTTG					

				CCGTT CGTAA ACTGA AACCA GGTAA ATCTG ATGCC TGTCC GCAGT TA					
L2	L2- 15_ CTe	.	Capitella teleta	AGCAC ATATC TGTTG CGACC CATCA CTCCG GACAC ACTCT TGACC TGGTC ATTAC GCGAG AGAAT	TGATTTT ATTAATA CGTCCTC ATACTGT TTTTATC ACTTTTG TGCTCTT AAAAGTC GCTCATT TTGTTTTT AGTCACT TTTAATTT TAGTTAC	RTGSSRHSMLSLLLPA SPMPWKSSETLTLLTS THTTKYLPQHLTDMHLRKPESSH YALTLLGTM MNSEHQNMSEKEGTQVENKWSRDRKET YKE QRDIYNRLLNQTKTQYYSKMVTDKARNPKDL FQLVDLLIHGKETNPVLPEHTSLDELSDRFATF FSEKIETIRLNLSNRQTDSQLLVHQEAEP ELVSS WMVFSRVTXEEVTRVIKKSPSKTCSLDPIPTWL LKACLPALCETITAIINDSLSSGYVPDSMKVAA VSPILKKTNLDQNELKHYRPISNLTFVSKTLERI VAHQLNIYMAENHLHEQFQSAYRQHHSVETA LVRVQNDILMNIDRGDAVLLILLDL SAAFDTV HHDTLIRRLEGIGIGGIVLNWFRSYLKDRHQSV TIDKHSSPRRQLGQGVPQGSVLGPILFSIYMIPL	cd0165 0	258	520

				GACCT	TTTG TTC	AQIERHNMSGHYYADDTQIYISFRPSSPLADIE			
				TGTAT	ACTTTTT	AKANACCSEIQQWMARNMLQLNEGKTDVILF			
				TGAAC	ATATCCA	GTKNILRTL DQKASISIGGSVVLSSASVRNLGC			
				TCAAC	TTCAGTG	VQDSELKLDLHV NQICRSAYYHLRNINKIRRY			
				TGTCA	TTCATCTT	LPLDVTEQLVHSFVTSRIDFCNSLLIGIKKANQ			
				AGGAT	CTTTTACT	QKLQ RVQNAAARCILYIKSREHITPYLHALHW			
				TCCCT	CAGCGCT	LPVEYRIQFKVLLLTFKSLNNLAPAYLTSLLHQ			
				GATCT	ATTGAGC	NAV LASSRRNSSRLLSIPRSRLVGAGDRAFQH			
				CAGAC	ATTTTTG	AAP TLWNASPPAMRTLSDVNVFKKGLKTFLR			
				CATAA	GATTTTA	RAYSDYC			
				GGTGA	GCGCTAT				
				TCCAG	AGAAATA				
				CTCAG	AATTATT				
				CATTA	CTTATTCT				
				ACATT	TATTCTT				
				GCTCG	A				
				TCCCC					
				GAGCA					
				GAACA					
				GAAGA					
				GAATG					
				A					

Dap hne	Dap hne- 5_C Te	Capitella teleta	TG	TAATTAT GTTCTTTT CCATTTA TATATAT ATGCATA TATATAT TTGTTTAT TTGTTTGT TTGTTTTT TTTTTTCT CGTTTTT AAATTTT TTTGTCG CTCTCTCT TTGTGCA AGACTCG TTAAAAC TATTATTT TCTCTCTT CCTTTTCT ATGTACT TTTCATG TGGCGCA	ISSNDPCDAYPRFLNKITTHYESSFPIIKKKPKK RDNNKWITNGLIISIKYKNKLYLDYKRHETIAN ELAYKKYKNRLSGIISQAKKNYYQGLIDENRH NMKNLWRTLKELIGSKKRSSLNEFLINNKP KDKDMISNTFNQYFVNVGPKLASIIPNVDSNPE DYLSGSYPNSLFLTPVTINEIKLALNKLKTTAS AGQDKLKPNIKLISDEISLPLTHIINLIFEHNKV PDEFKFANITPIYKAGNSMEVQNYRPISVLP SKLMERLLHDRIYVFLQQNEIISDYQYGRKXF NSEMALAVTTDNISSLDSQKHVMGLFLDLKK AFDTVDFNILMKKL VHYGIRGNALDLLKNYLS NRKQSVKIENSVSSELPVTCGVPQGSILGPLLFL IYINDLPNAXTSCNPIMYADDTNLFLSGKNIDE MENKFNTDLVALTKYLCNRLSLNVSKTHSM LFTLNNSLRDSSLSLSIDGSIIDTVKCINFLGVKI DNKLNSEHITHTCNKVSKSIGIIRKISKTVNRP TLLMLYNLILPYLTYCNMIWGKSANIHINRLF ILQKKALRLINKAPLLTHTAPLFHECSXLNIHD LYIYRSTIFIFKLSHHMLPTPVSKFFNLTPLVHG LNTRRQLNYKVIVPFCRTSLGQKSLNFNIAKLY NNILQPLSLLNLDSEKFKKSIHCILVAKYL	cd0165 0	239	495
------------	--------------------------	---------------------	----	--	---	-------------	-----	-----

					AAAGCGA TCACGCT TCTGCCA CTACGTT ATTTTCG CACCTTT TATGTTG TGATTGA ATAAAAT TATTATT ATTATAA AAAA				
L2	L2- 17_ CTe	.	Capitella teleta	T	TAAATTT CACATAA TTTTATAT TTCAATA TGTTTTTC GGACTAT TCTATGT GAAGCGC ATAGCAG TGTTTTA CCTTTTG	TLKIAQVKPRPKKPGVNTMDPKLFRPISNLPTI EKLLEKCVIRDLHLYIQQHNIIDPLQSAYRPGH STETAMLRIQSDIKTSLDEGNGVLLVLLDLSAA FDTLDHNILLSRLQNSAGITGVAIDWMKSYLSS RQQFVQIEEKRSVTHDLDIGVPQGSCLGPLLFV MYTTPLGAIFRKHGLSYHQYADDTQVYCSLPS DDQNSLKNVLTRVEKCLDEVKCWMTQNKLK LNEKTECIIFSSKRNKQLFHQTTIRFGGSVITP VTSVRNLGAHLDCEMSMDKQINNTIKSAYYH LRRIAKIRRHLDRSSCTQAVVAFVSSRLDYHN GLLAGITQKNINRLQKVQNHAARLVSQTAHH	cd0165 0	12	272

				CGCTATA TAAATAA ATAAATT GATTGAT TGATTGA TTATAA	CHITPVLADLHWLPIRARIDFKLMLNAHKAVH CNDAPKYMKELCMKYTPSRTLRSSSDPHLLAV KKSKRNIGASTFDVYSCKTWNELPRDVRDIEE TICFKKTLKTHLYKRCF			
CR1	CR1 - 10_ CTe	Capitella teleta	GCAAC CTTTTT GTGGA AGCTA AGACC ACGGA AGGCC TCGAC GAGCT TCGAA ACAAG ATAGC CCCTC TCACA AGCGA TCTCTT TGGGG	TAAATAT TCTGCAA ATTTATG ACCTGTC AAGAGGA CCTCCTC ATGACAT GTACTGA ATAAAGT TTTCTATT GAAAAAA AAAA	MACEWDGKNSLLKLGSENQPSVSDVSISAMQ APPVAQSPVVSTARLSTLALPYAQVILNVPTAK ASTRHPASPVHPTSKPKKENMDSHNIWITNIRG LVSNFDSFRLEIEHAHPYL VIVTETFLSPDVPDQ ALLIPGYSFLRRDRRTHGGGVALYYRSSLRLR RLTSLETQDHEVIWFQASTREHDIIGCAAYWPP GSDL DITQHLQHALDEV SIEFNTIVLLAGDLNA HHPDLSESISTNTAGRLLQRFATVNALTQVIQA PTRVTAHSSSCLDILLTNIPEEINFLGVYPGIGTS DHGVACVDLKTTSVITHTEDSATIRPLHFNTH STNWKT VNDHFSNINWHHRFAACNVEETWTL FKDAYEEVLGRLSKKSRPTSRRTNHERHQHNT LNQELIRLRATQQA AWR LHSTARTRETHQSLV TARQEFTLACANHIRTDEATVKRMTESNAK QWHKYCKQLYRGSAVKEAIPIMTVGTESYITA KDKATILNQTFIAKSRASSRPRFPSLKKRTDSTL ADILFNEYRVKKILQDLNINKASGPDGIQPSTL	cd0165 0	587	851

				AAGCG GCTTC ATTGC GCCTG ATGGA CGACA GCAAG ACCAG GCCAA AACAC GTCAT CATCA GAAAT GTTCC TGCCC ACTAT GACGC GGACA CCATC CTGGA GGAGG CCCGA GTCGA	KNCSDSLHVPLAILFTKSFTEGVVPTEWKLAA VVPVYKNKGQRTDPNQYRPISLTSCVGKVME KIINDALLQHLMGNELIVDNQFGFLPHHSTTDQ IALLLHELLAATTTKKSJVACFLDLAAAFDSV PHPAILQKLPAYGVRGRLFRWIESYLSFRTQYV VVDQSSSPIAQVSSGVPQGSVIAPTLFLLFINDL CHSLSANTSLSHPDCHPEALVYADDTMLYRTG KDPALICPAIEEDLRIAAKWADIWGMRFNASK TVAMYISRTTATPPPITFAGATLEYSHEHRHLG FILDSALSFHAHVNALTRKGATEVFLLRRLSYK VKDRDLLLLKIYKMYVRPHLEYASPAWAALTV TQTGLLES LQRRAMRIILALPRNHITDADYAA LNITSLRHRRNFALACFMFKLNSNRLPRKLSK YRPIPHTNPYNTRNCVLRAPYVPYPCLRLLDRS PFLFGLKILNFISHTFHSLPDISSFKSSLRESSYY MPLAQIYY			
--	--	--	--	---	--	--	--	--

				ATACA					
				GCAGC					
				GCCTC					
				ATCCG					
				TCATT					
				AACTT					
				GACCA					
				AGCCA					
				GGGAG					
				CACCG					
				CCAGA					
				AGGAG					
				GCCCA					
				TAAAG					
				CTGAG					
				CCTTG					
				ACGAC					
				AGCAA					
				TCAAT					
				TGGAG					
				CGACT					
				TCTCA					
				ACC					

Dap hne	Dap hne- 4_C Te	Capitella teleta	CGATG TAAAATC TACTC CTTCACA TTTGC ATCCCCC GAATT CTGCTAA CTTTTC AATTATC TGCCA TATCCCT ACAAT AACAGCT CAACA CATCTCC AGCCA TAACGCT ACTAG TTTGTGA AATCA ATATCCT CAGGG GTTTGTT TCGTC AGTTTGT TGCGA TTATTCG CTATT TTGTTTGT ATTGA TTGTAAA TAACA ATTGTAA TTTTCA ATATCCA TCAAT TTGCGGG AATCT CACATAC AACAA CCATGTT GAGCA CTGGGTC TCAAC AGCCCCT	MKNSIKMKNKLYCKYLKTKNPNDQKQYKSY KITLRSILYKAERTHYDQLFCRNKNNLRKSWE VIKTVIHKQKNNNNKQQYNMEVNGGKKTSDLDII TKHFNEYFINVGPNLSKSIPYSNIDPISYITQTNR YSMFTKPTNETEIGNIIRALKNSAPGPDGIPSSIL KENTAHFIQPLTHILNLSLSQGIVPNEMKIAEIK PLFKSGNKHLVNNYRPISLLPAFSKIFEKCMAS RLIEFITKHNLLEYQYQFGFREKHSTNMALHLLI DKITNAFDQNKVLVGVALDFRKAFTDIDPSILI SKLEKYGIRGCTLDWFRDYLHNRKQYVHLEGI KSSPLSIKCGVPQGSILGPLLFLLYINDLPQSSQ LLPIIFADDTNVFYSGDDSEECIGIINQEMINLIK WIRSNKLSLNVEKTQYIVFSKRKIAPDNNDVNI NGERLNRVDSFKFLGVIIDDRLSWKQHIAYIKT KVSKSIGILRQARPVLNKTLLTVYYSFIYPYL NYCIDVWGHCSKQDLQSLFIVQKRAIRIINYAE RNAHTDQLFKSLEILPLHHL YIFSVIIFMHKFHH KRLPPVIDELFQTNASIHSINTRQRHQLHTPLLH TNISMKSIRYRGAFIWNQTFNEIDTNLTINRFKT CLKNKLFTDFPIRTSPIQ	cd0165 0	195	449
------------	--------------------------	---------------------	--	---	-------------	-----	-----

				TACAT	CTCACTC				
				CCGGA	CACTTGT				
				ATCCT	GTTTTTTT				
				TGCCA	CTTTTCTT				
				ACTCA	TTTGAAT				
				ATCTC	AAACACA				
				TGATC	TTTTTGA				
				ACTGT	ATTTGAA				
				CCTGT	TTGA				
				ATTCT					
				GCTTC					
				ACACC					
				ATTCC					
				ACAAT					
				AATAA					
				GAGCA					
				ACCAA					
				AATAA					
				CTTCA					
				TCACA					
				AAAGG					
				CTATT					
				CACTG					

				ACCGA				
				AATAA				
				AAATA				
				AATTC				
				CACGA				
				CCTCA				
				TAAGC				
				TCCAC				
				GCAAT				
				GGGAT				
				CAAGT				
				ACTCA				
				ATGAA				
				GCAAA				
				CTGCC				
				GAACA				
				GCATT				
				CACAT				
				CATTC				
				TACAC				
				ACAAT				
				TCACC				
				ACCAT				

				TTTTG ATCAA AGTTT TCCAT TAACC AAAAT AAAAC TTGGA TATAA AAATC GAAAA CCATG GCTAA CAGAT GCA					
Dap hne	Dap hne- 15_ CTe		Capitella teleta	TCCTT GACAC CAGCA ACAAG GTCGT GTCTA TCACC AGTGA	TAACGAG CAGACGA CGCCACA GCACCAG TAGCGGC CAACTGA GTGTTAC ACAATGG	MVPDVNPSFSIRAEDCRYIAPXEIPNLCEHYTH PLRILHVNIRSLRANYDELLNLIASYEESNRAL HVLLLCETALSDANTVLFGIPGFDVYANRKS HQRGGVAIYVRKGLNFTQRPDLSIFSEGKFETI FIELSLCQNSQPIIIGEIYRVPSLNAPETIAFFDSF TSLHRSNASIIFGTDQNLDLGKIDEHALTRNV YNLLESEGFLPTVNVPTRIEGVTKTIIDNIYIRGS LPSARTSVVEVKIADHLPIVLTVD FHMHRKKN	cd0165 0	543	800

			GAACG	TATGACG	SPLVIMHRKMDETIAQRIVDELGQCDWQSNEN			
			ACCTA	CTAATTA	RDIDSAFEFFLSEMTRVLDRHAPAKQIKIPSKSF			
			AGGGA	TAACGCC	RREPWFTKGLQRSSRRLSKLYSTAAKCAKNSH			
			CTACT	CACACCC	QHTAYVKYRNVYNSAKRAAKQLYYDNLFSES			
			GACAC	CACCTCT	KNDIKRTWKILNDLTKKSGKDNSPVEKLQVD			
			AGATG	CTTCCAA	DVTITEPKAIADAFARFFADVGPQAGLIASHV			
			ACTGC	AAAAGCA	QTDDQQNRAHLGVPPSHHSVFLHPVTQQEIIRI			
			CTAAT	ATGTTTT	VDSLKNKKSLGLDLVSTHFVKKLKHALSAPLS			
			ATCAC	ACCTTTG	ALVNKSFCEGAFPNLLKRAKTIPIYKKGERFN			
			GGACT	CACAGCC	MDNYRPISLLPIFGKIFEKAYCNRLMPFLED RG			
			CCGAT	TGATCAA	FLCDSQYGFRKRNTTYAVSELYLNSINALLK			
			CTCTC	TATCGAA	NEITLATFIDFSKAFDTISHEILLNKMP SYGIRG			
			ACAAT	ATTGCCA	RALDWIRSYLSNRSMKVSIQGCLSEEVQLSQY			
			CATTC	AATTTAT	GVPQGSILGPILFIIYANDLNACLTHTKAIFYAD			
			AACCT	GCACCTG	DTTLFLSGKSANVLAERMNDDLGRAKRWCEA			
			CTTTCT	CACTCAA	NSLAVNIKKTKYMLIGPNALTRCTIDL SMANE			
			GTTCTT	AAAATAG	KIERVTTFKFLGITIDERL TWKAHIESLSKKLSS			
			GTTCTT	TTTACCA	GLYALRNTKRLTSRHGLLSLYHSLIHSHLTYG			
			GCTAT	CCTCGAC	CLLWGSASDNALKPLKIAQKKAIRLVYDAIV			
			GTTAC	TCTCTGC				
			AACCC	TATCCTA				
			CCATC	GACTCAC				
			CCTAC	ATGCTCA				

				CTTTA	ACCTCAT				
				CTTGC	CAGTACG				
				ACAAA	CCACACG				
				GCACC	CTCCATC				
				TATCT	CTACCAT				
				GCACG	ACCGTCC				
				TCTTCC	TCCCAGA				
				ACTCT	ATAACCT				
				GGTGA	CTCTAGC				
				GTTCC	CACTGCA				
				TATCC	AAAAGTT				
				AACTT	TTTTATCT				
				TCTTC	GAAGGTA				
				ATTCC	TCAGAAT				
				TACCC	CTGGTCC				
				ATCAT	TCCCTCA				
				GCGAG	ACCCTGA				
				TATCT	AACTCGC				
				CATAC	GCACTTC				
				CTTCG	CCCCTCG				
				CCATA	CCGCTTC				
				CCTTA	TCCAAGC				
				CCTAC	AACTAAA				

				CTGTA	ACGAAAA				
				TCTGA	TATCTGA				
				CCCTA	ATGAACT				
				GCTGC	CACCACC				
				GCGAA	TCTCTCT				
				CAGAA	AACTTAC				
				ACCTC	ACATGTC				
				ACTAC	CATCACC				
				GAA	ATTTTTTT				
					TCTCTCTC				
					TCTCTCTC				
					TCTCTCTC				
					TCTCTCTC				
					TCTCTCCT				
					CCTCCTC				
					CTCCTCC				
					TCCTCCT				
					CCTACTC				
					CTCCATC				
					TACATTG				
					TTTTTGTT				
					TTTTTTTC				
					TTTTTGCT				

					TGACAAA				
					TTCCCCA				
					AAAGTTT				
					CAATCTG				
					CTTTATTT				
					GGCCCCC				
					TGGCCCC				
					GTACAGA				
					CGTTTGT				
					GTCTCGC				
					CGGGGGG				
					CCTCAGG				
					TCACGTT				
					TAATGTC				
					CAAATTT				
					ATAATAC				
					CCTGTAA				
					TATTCTG				
					CCCTGAA				
					ATAAACA				
					TTCATTC				
					ATTCATT				
					CATTCA				

Ran dI	Ran dI-4	Chlamydo monas reinhardtii	TAACATC CTGGCGG GGTGCAG GCCTGGC CTGCACG CCCGCCT TCACATG CTCGGTA GTGGGGG CGTACAG GAGGCGG GCCTGGC CCACCTC CTCATCT CCCCCA CTCCCTT AATCCAC ACCACTT GCCTACG GCAGGGT GTCTCTA TTTCTATT TCTATTTC	KLGPTPAALAQDPPTNPTGGPPAGPNTTNSPSP PTSANPGPAARSNTATRIVPKPAGPTKRPARRT PNPTPKLRKLATGTPSAPPPAKRLHQTTIATSH KRRHTATLQNTPTPTENGQAQASPSPPPDGRIS ILTLNTRELLSHHESILHTVTTHTPDILFLTETKI PTRLHYSMGSAKKGLSSHYQTYTSNAPYNPRE GTRPHAGVMVGVTHRFSGTSDMNAHNTD SRLAGHVCHIRIAPPHSTPVHAIGIYAPSDDATR AAIYKYCTDQAAKTAAKEHLLIGGDWNAVL FPSDRSTGQLNATDRAHIAFTTSAGLHPLTPPT HEHPRPHTYLQDQATAPHMSRIDDLLVSHHTH ATITTTGGGQEA VHECGDALDHRALLSSVPLA ALRLYPFPTAPAAPSDPDWVNYHPPTKDQRE KARIKMEETIGQEIADLLNTLHPAHLAITNKLY PAGLTPAQAERNTPSPAEFRALPEVSAADTTSI SNTLESILQKCLDVLKSYCDKKPPRASTTSRPH HTSKTVTRLLTATSKLKERLTAALVTNAAGA TPPTTSPHPAPTPPPNTTSPTPTPPTPAEADTPT AIRSQLLNARTLLRAAKKEARKLNTKRAARAF RRLVATQPKRAHQKIFDETDQPRGTPALTDPT TKEVHTAPEKILSSIHQHFTAQLLPINGPRTGH YDIPHSEKQYPWSRPDAPDRMPALSKPTDAPA TLLPGIADSTTFFGCVNHL SRRKAPGPDGIPNE	cd0165 0	786	1074
-----------	-------------	----------------------------------	--	---	-------------	-----	------

					LLQTIPTSLQTAIHQLMTLMWVKAHIPHTWTQ SDTVLLPKKGDPLLIQNKRPIALANTLYKLYTS LITVNAALFSEQAHIFTESQEGFLRGRNTERQV QNLLHAIEDAATTGRDLFLLYIDFTSAFNTIDH DKLLIVMHDLDGYPTDLIEVVRDLYGKAATSVR TEHGTTPIPIQRGTVQGDVLSPLLFIIFIEPLLR WLHVGGRGYQYGCLTKAENDKHNLSGAF DDLVTATTQISNLRIQVEKIEKYAAWGGLKVN VSKCQLQGILHGRARADPTLGGPRSKPHAAL ARMLEGQIRIDGTAIPYKPPHESYKYLGIHITPT LDWSAQVQYIREKIKKRGQRIATSLATPDQKL RMIQLTIHPAVAYSLPFMPYSTHDVTSLDSLIA STAKMCYGLKRGFPTRAILQPTNALGLGVGSI RELYVRRAGKALVQALNDGGRLGAVTRAML HLQHSVAQKIPADQLRTQCRFYTSRLQSLITD AGLELTQQGSAYHTTLHGLSALLARTVPPGAD AEQALPFRLIQPLLELTSDFRELLTPDRTCVIDA NTLKNTWPRAQRRHRLALNRLTLALHAETRT RAKTHNSTAPLTTAQRTPDGHTFTGLHTAHT APAAVDWGTNLNRFLRQRAPATGQTANAGETA AATTAAATTAAAEPEEDHHAGENHAPAQTRR PGTQQRKRACIKMVREPEGRGSPPADAPAAFD WDPSLDPLTEAHCADLIHTAPSLLGKSKPQNP		
--	--	--	--	--	---	--	--

					AAHLRGLLLGPTPPATFTHRHKLSQEEQDAYL AATSDTWKHFSETYRTKALTTNTLYGQQR VQQVIAHTTSKGAPFYRVQWEPTVITPGSLTL YARNRYRPKPGGLTWLLTPGNEPRVLVEWED LEEPAAILDAQDNKQELLDALNARLAVPLGPR PPPPATDTHLSHDQRQGRWTDASNDTARGQR TLRHYVDINVEPCNPSRDIHPTQAYTVQLGTV LDPGSPACGTHLAHVYDPRGRCIGTISAEQLSL LWQRYQHAQTTWPKVHAKHDQGRGFAAAIA RLLLRYKPGGGSKSHPVTEKGECMLCEGLTQT LHTAIGSTTELFASPLNVSDAATYYTYLYEED MLFGANYDAYSQPWDGVAVACPEPNTEDGA HAVRWALAWAEHTETLTNTRNPPTTATLLILP TQRNAPYTRYLAHPRVTLLGSISATTTHALLG QIHNDTGITLKHDLVLVSNTPGAQLAQPIRDAL GTSLYLSGMPVDPQPQLTWAHTRWVTKQVRR MTRTRPAAPRVTPPPVPATPATLAPFAATQPLR LNPYEGFYTDGSVQKDALGQPSAWGAAIDA RTLRCYQIDVNPQGQHCNTITRCELSAIHQVLK LDEVQALPEVTVYTDSAASLFLIRKTINQPQLL TFSKHKALLDNILALLKRRSHAGRRTNLLKVK AHSGIAHNELADQAAVKTAKEQIPGAEWLV EESDNDPRSKYYWIRKIIKPTVEGVPREKNEGH		
--	--	--	--	--	--	--	--

						YAHDLNRGIATLAAPHIAQGHTRETLYTGFVA NAIPDLDIKSSTRTMNLQSGRHRKAKRCMDYL YGQLWNNKLAARFGRTVTGRRAPRAGPDTPA PCPLCRGPDSGGHILGGCHNPQLRAMYIKRHN IAVQTIARAISRGDCGGCYMVMDATSKADLPD YAADNRLPPWLLPDQEGKPAGRHLHYRPDILL IPSISLAAALNPDFVVLPSHDTIHIEAGYTADT NHAAKQHEKAQQHQALAADLREAGWKVQYT PQSAISLGFAGTIRKDLHPLLTSLGVTSHKARQ CCDTLHDHAVATLNLHIVLTRRRLERGLPPGDP GGT			
L1	L1- 6_C R	.	Chlamydo monas reinhardtii	GGCGG CGGCG GAGGC GGCGG CGGCG GCGGC TGCCC TGGCG GCGGC GGGTG CGGCG GCGCC	TAATTTG CTCGTGG GTGGTGG GAGCGGT GTGCCTC TGGTGTG TGCGCTG TGGGTGT GGGGCGT CGACGTG GGCGTCG CCGGACT	QWGWWRRRWRRHAAGGGGRPCRCAPRPRPH RGD LGCPAGGS AVVRRPRQQAAGPRPLRVGS HNVRGLRSRVQHLGCPRLHALLRCWERQQLD VVCLQETHVRAEDIDLVTADLRAAALRLHVR EWTAWWAPAPTPHQRVSGGTAILIRTCLLEQG AVTFPGGVAAPSVADGAWVGRGVSLPVQWG GHDFTLAAAYYPLASPEQRAFTQQWVASAVG GEGNHLLAADFNFVSDVALDTASGVPRPGDP AAATALAAACPGLIDVLRVRRPGQRVCTYIHS HGASRLDRILCSGSLVAQVLACAVEAGVPSDH RLVTTLAATPEALPPSPGLPKARLGFLDFKDL TKDYRTWLTAALGARPGAPAE LLAWWPQLK	cd0165 0	604	863

			TGGTG	ACGTGCC	RAAAASAARLSRVAAVRRIAASEEEARALAA			
			GTGCG	CGGTGTA	AAAAAAAVEVGAALPEAAAAVRARCAAAT			
			GCACC	CCTGCGT	AAVQAAAGAARRTRHEWLRGGEKPCPVLTR			
			CGGTG	TGACGCC	MLRPSGGPRTIARLKRPDGSATSDPTVMGQVM			
			CCGCG	TCGCCGG	ATYWQDVCAAPPPAPAAQAQVLAALQQHGR			
			GAGGC	TTTGATT	RCTTEEAQQLGQLAVTAAEVRAALAAAPAGR			
			CCCCG	GCGCCTC	AAGPDGIPVELYQYYESDFSPVLADVFTAVGE			
			CTGCC	GTGCGCT	GAGVPGGFLDGVWTFYKAGDPLEPSNYRPIT			
			CCGGC	AGTGCGG	LTDTDYRTLARVLCRRLQPVFGRVIDPEQTAF			
			CGCCG	TTTGCTG	LSDRRIADNVLLLQLTPGLLKAAGQNSGVAAF			
			GGCCG	TACGTTG	LDFYKAYDTVDRSFLACLERMGVGAGFLTW			
			GCGAT	GTGGGGC	VSRLLDTRGAALVNGRVSGWVPLAAGVRQG			
			GGCAC	TTTGCGC	CPLAPLLYLAIQAALLSWLRSGHGVLAGLAS			
			CCGGT	GGCTTGG	ALASQYADDCTPFLAGMQAVPGFLADMVFR			
			GGCGC	GCTTAGC	CASGQRLNMSKVELLVLGTSAGLGVAPAAVA			
			CGCCG	TTCGACT	GPPLPPGWRVVNTAKSLGVHFGDWKTATPPV			
			CGGCC	CGCTGTG	AFDGVVGVGLLIVRMPVSMFGRAAAASAYAL			
			ACCGC	CAGCTGC	SKVLYHLEFAGFSPWAPVGKLLGQVAAVVD			
			CCCGG	CGAGACA	RRLSPAKYAVAPHARPVGVSLEIMQVPPALGG			
			CAGTG	CGCTTTA	VGLLPLQLHVQARQAGLAVRCVLGACGLLAF			
			GCGAC	CGCGAAG	VPPWTKVAEAWLRARHPAATPLRLLVRCDPS			
			CGACG	ACCTGTT	RVLDKLVTPCECKPFLRLLGAVAHLPITLVAPP			
			CGGCG	GCACGGT	AAGDWVYHVPLWANPALCDGAQTFDVAFAD			

			CCTGC	TGTGGAA	LFALPGLVYVGQLVAAHDQLEAVRAAHTRPG			
			GGCCG	GGTCGGT	GWGTRFGEAMLDLYLTSVWRPVLGWGPRRH			
			CTGCA	AGCCAAC	LPSLLGPASPLAAAERIAAAVALLPGGWEAA			
			GCCGC	TGGGGGG	ARAARLARGRAWVMPLPADAVRQAAAGALT			
			GGTGA	AGGGTGG	QVVRGLGWVLDGGPPLLLLNYSVKAGTAMQ			
			CGCTG	TTTCCTTT	LGPQVAALRSKHLLYVRDAGVSADGALAAAG			
			CCGCC	CGGGCCG	AFAGTLARLWALKWENYHKEALWRVAANAC			
			CGCGT	GAGATGG	WSFPKNAAQRAAGVAMPLCRACGRDMCDGD			
			CCGCC	TCCCTCTT	RRHYFWDCVVARALRENMAMAMLGSPPDTA			
			CCTCT	GACCTGG	DDCFSRADLWL VQPPAGLAPPVWDVVCLAAL			
			GGCTC	GTGCACG	SALDSGRQQVIMTGLRVRRALTSAEVMRISVS			
			TGGCG	GGCACCC	VIADFWGRLQSYVSLGLRPKAWATVPAHHPF			
			CTGGC	TTGCTTG	LSRGAGDCVLLVLPDDVGSPPPSP			
			GCCCG	GTCGGAT				
			GCGCT	TGTCCCT				
			GGTGC	TCTTTGT				
			CGGTG	AATATCG				
			CCGCC	CAACCAT				
			GCATC	AA				
			TGCGG					
			CGGCC					
			GCCAC					
			GGATG					

				GCAGC				
				GGCAG				
				CGGAA				
				CTGGC				
				AGCGG				
				CAGCG				
				GCAGC				
				GGCCA				
				CGGCG				
				GCGGC				
				GGCGG				
				AGTGG				
				CGGCG				
				GCGGC				
				GACGG				
				CAGCT				
				GCGGC				
				GGCCG				
				CGGCC				
				ATGAC				
				AATGG				
				GGATG				
				GTGGG				

				CGGCG GCGGC AGTGG CGGCG GCGGC GACGG CAGCT GCGGC GGCCG CGGCC ATGA					
Ran dI	Ran dI-2		Chlamydo monas reinhardtii		TGAGCCA CGGCTCC CCCACAC CTCCTCG GAGTGCG TGGTGAG GTGGCAG CGCGAGG GGGATGG GCCTCCT CGCCCTA GCCACCT	MSSNPMTLQIGTQPGQQQPDLTSPSHPLRGGN NTERTTTPAAENGHIPVPLSWPRQIPSTYFREQ QHLNFCTIHAINNMLGAPWLRAHAVLAFCQV VHTALADTPELQATWGHFHKHNGLFHEAAIN HYLRHRPAPIDTTGTRLLLRLAAPNLDPGHSQ DTILGLLPECSRNMGMVHDGRHAYAIRKEED SWFLVESFVHIGRKPMTPDLWQAVKGDIFCLH PYHLDMPDLPRTYPTIPNLPPPQPPSPPPPNQT KPKRGPTQRALSTM FARQTTQLPTNHSAGDN AAPTPHPGPAETTVTRPLTQATLNITRREATEP ARTQNERPATPNSAARPTRRPRHDNQLDPEPT SPTRANTTTTRPRANANMHAPNTNPTQPTANNT	cd0165 0	1154	1442

TACCCCG	TQPTWAQVVATHAEAAGNALRTVGRITTQAP
ATAATCC	HQRPETPPSAARPTQRPRADPPDSPTTHQTVNP
ACGCCAG	NTPRRTPPGPTPDPGAGAQDNTETRPDNTTWL
GCCCCAT	HIATHNVRGLGTNLLSTICHLREWNADIVILTE
GGGCCGG	TKLGPRTRWLKDTLRQEKLSYRTFTSTKPGTE
CGTTTGA	QYKRRSAGVLMASKRYHSGSNVHTATIPPHL
ATTATTA	HGYIVHCTIRTPHSTPLHILGVYAPEDMPTRRRI
TTATTATT	YKYCEETINTANKTDAHVIIGGDFNAVLENGD
	RTGPPDEADKHHRFVSKNKLKRLPEFGGTIS
	MSFYQTRDGTTPQPTSRIIDDILICQNIHSALPTAS
	RTPATLQVQPCGGLFDHYSLHARVPTFPIRLW
	PSHTATQADTQPPETPGPWRTVVLPPIPKSTLEA
	ARTKIEHTLANPIATLASAIQPATTAIQQHLAQ
	HTAGATNASELQHALQTDPAIQTPDLDALANQ
	LLEILQDGLSILVDMCPKKPPTRGVFAPRKINK
	RIRESHNEIKQLRTQCDALLQQQHPYANITTTT
	APNTGIPLPDPTTEQDAENPPPQPPPAHDATT
	TAEPQRRPRPPPATPAIANLIDKIKRAGDTLRQI
	LKKQTQHDIQKARDRFQHKLATRAKQAHKDI
	FSTQEQSRGIPALIHPQNKAHCTDKASILDAIH
	AYFQQGASITGPRSGKYKPEDRTQVYPFDPS
	LKNSAPDTFTLHSATTPPTHALLPSIADSTNFDS
	CIDHLSRNKSPGPDGIPNELLRALPQTLR TTLH

NLLTLMWVRAHTPASWTASETVLLPKPGDPL
LLKNYRPIALANTTYKLWTSKITVAISDISHDL
DLFSETQEGFLRYRNTERTQLNLVHALEDAGL
TKQDIYVMYADFSSAFNTISHDRLLLEIMWDLG
MPHDLIRVVQNLYSNARTYIRSEHGLTAPIYIE
RGTVQGDTLSPVLFLVFIEPLIRWLHVGGRGYS
YGCLPPHLNARYNCSSAAYADDLAVLTNTLS
DLRIQCDKIHRYSEWAGLQVNHAKCRVTGILH
RRAMQDKGLNGPTCNRTLRSMLENKIHIGTTS
IPYLPATEPFKYLGVTMTLNLWGPQFAYLCK
NITDKSLSLQTCQASPEQCLRIKSCIQSVAAYS
FAVMPYTENDIRTLDNMIARLARTCYRLSRGF
PTRAVLTPAEMSGLVGSGLLPIYVQKATRALT
LALNDPGRLGVMVTHALLELQTRIAGTTDMHK
LRHESHYYTTLKQLSIMRAYNLTLWDHGAPY
KGPPNTLLTAISRSPHPISHHLVLTLTQAGLDLR
NLVGKERTPEGRACLIPSDILFQHLPRNQCHK
HRVALNRLSLCLSAECDGAREAPFNQAESYKC
HLPLPQRYRALPNTPAFAALLPRNATPGELTTQ
STMDAHLQHMTALHPNTEAQTQHPNGLGPR
QRQTTLAHLRRTTGAPPTAPTQSPTATLTQTP
TGTLPTPPLAQPMPPPLTQTPPLTTPAGPDAP
NTQTPPTVASLTTAHQQPKSKKQRPNAIKRLQ

					QKIDADPRHDPLPPTDMRSHTTEKSGFILNNDPD ENDSGPPNFVMWAPHYFANVQKDMQGTNRM PRLGIPRHHSRFPQNRTRPNPTRTIPAHLADPYF LGSQVDDTNPLNPVTAYISANPPTTPPTCAYDS RPRWADTGDFMIAPEPPNWPNITGATPGDSDH PRQADHDPQGMRAEYMQERERLRAHLDDH PPQPRPFKRWMGIRLQLLTGKLPWPTQAYHV AQAAKQLVYSLYHDIEEISHIQAGPITKAVTAP EKGASLHTERYYEVAWKDTTIVKGAAEAYTR MKYTGTA-AAHMGYGGQACHKYFLRMKWDPT SEPVDKCPTLQGWTPAISEHLRKLAEPTPTPQ PKPKDARLTPAQQQGRWTQADVAASHVRRTT KQNTRIDTNPCNPYLDVHPTGRYTIQIPDTYTD ERTLAHVYNPTGRHIASVTLGTLLTLKARYEH ALSTHHDATSPHRSQGGFPEDLAKLLLRTRVD SKGKTCKSAPARERSLPESLKLALYNCLASTHS DPDRPVGLVERFASPLNCSGRATKYYSQHAAD SLFGASHDAYTAPFDGLSLAHPDPDPDACTKA LAWATASIEHLEHQATPAITFMIPQGKGAASR CLASPHVTILALIPPGVSASSMLDGLGADEAA SSKQGLQLLLIANTTGVQALNSNNNLTRLQTS LAELGAVPPPTSQPPGRRQPLEWPARVTLPRD RTAPPGRPYSHRAKARAKQLTRPQAPAARGL		
--	--	--	--	--	--	--	--

						DTCPTPMP TLSALYTTTAAFPTRSSYLYDAATI VWTDGSCIKVAAGEDGNEMNKL GACAWSQQ TEHIYVEPGGLNSTNTIQRAELSAILVALEYIRP RLCLLGMTQVTIASDSLVS LYLRALTAPQRL LLSKHKNLLENILHELQMYADAGVPVQLLKV RSHTGLHGNEEADKGAAKVAANDMEAMGQA IRLAPDEPWSNIWWPKRNSDNFYLS DLNRGVL NSLPPEAQQGFTNNTPVLEQWAEASANMCPTL SNRTLASSTRHPWLVKQILYARYGYLFNAQLK RRYRLGGNGNCPLCHTPDSGGHILGGCTHPK MKGAYINRHNWAVQRVAKCISKGTYGNGFM VMDAGPLADLPHYASAQRPPAWLAAERTSSP LITRLRPDILFIPNLPAHETHREGYRPPLNKARL PIYILEIGYSSDVRHLDKREEKSQQHQELKRL EAAGWTVHYSTNECLALGVTGAIPCNLTKLLT SQLGVTHTAARVCGERLHTHAVNSAGGIKMR RVTEQELKPRGG			
Dap hne	Dap hne- 1_C Sav		Ciona savignyi	TG	TAAATAT TATACCG TAACTGT CCAATGC CTTAGCA ATATTGT	RFFEAEXFLSDLYXNFCNLLSHKPPEYSGDCLL QNFFSVFSEIVDKHAPLRSISRKKRRLLSKPWIT KGLLQSIKNKNSMFSSIVKNKPNLQKYSVFKV YRNKLNHLLEISKQMYKHEFSRSKNDIGRSW KLIRELIQIKSPTMTISK SVGITASDVTSDPINVA NKFNDFFVDIGKNLASKIKPPTNHVPLKPRKS	cd0165 0	266	523

					AACTTCA	SSFFISPSTPDEVFSLINCLDVKKGTRHEDIPTKF			
					TTAATTT	IKLANSIISPILSEIFNTCVKQGIYPDFLKNAEVIP			
					AGATTTT	IFKNGSDNLCTNYRPISLLSQFSKIFEKLLYNRL			
					CCATACC	IKYLDKHKLLSNHQYGFRENCSTTLAVADITN			
					TCGATCT	MLLSNTDKGLYSCGIFIDLSKAFD TVDHSKLLA			
					CTTTAAC	KLEHCYGIRGVALDLFKNYLTNRSQYTFVHNV			
					ACATTGT	RSSIKPVL CGVPQGSTLGPLLFLMYINDLPLVS			
					AAATCAC	NFVTKLFADDTFLCMSSKCPYELQAKVNLELI			
					TGTTATA	AIDDWLRSNKLSLNYSKTSYVLFTGNLKPIGIN			
					TATTCAT	DFHLNINNNKIIRSQSAKYLGVLLDEKLNWNC			
					TATGCAT	HILNLSQKLARYCGIMFKLRKYVDIKTLLNIYY			
					TTTGTTTG	SLVYSQIQYAIVCWGASNKKTLYPLNVIHNKII			
					TTTTCTTT	RAMSYSPARCHVTPLFSDLKILKLCDIYKLEM			
					TTTTAGG	GKL MYKFHNKTL PDSFN YIFQPLYCAHNHITR			
					GTTTTGTT	QVAHHDYFVPYFHTLLGQNSIGYRGPVAWTAI			
					AAAGTTT	DVSLKKLGWFSFKKQLKNQYLLTYIT			
					TTTTGTC				
					ATTGTAA				
					CGTTTTT				
					GCACAAA				
					TTTTTCAC				
					ACTGCTA				
					TTCTGTTT				

					ATTTCTT				
					ATTTATTT				
					ATCTTTA				
					TATTTTA				
					ATTTACA				
					TTTATTAT				
					AAGGTGG				
					CACCTGC				
					TTCGATA				
					ACCGCAT				
					GGTTTCT				
					CGGTGTC				
					TCCTTGTT				
					AAACTTC				
					AGTTTAA				
					TTTATTTT				
					CGTATTG				
					TTTATTAT				
					TTTGTTTT				
					CTGTGCG				
					TTTAACA				
					AATAAAA				
					TATGTTT				

					GATTGAT TGATTGA TTG				
R2	YU RE- 2_Ci s		Ciona savignyi	CCAA ATTAC TTCCA GCACC TCCAC AGCAG ACGAA CGAAG AAAGA AGACT TACGA AATAA GAATA AGCAG TTTAA AGACG AAGCA GACGA ACACC ACTCC	TAAGATC CGCGGCT GTGGCGC CGAACGA GCACCTG CCCATTC TTCTTGT AGGGACT TTTTCAC CCTCACT CCCCCA ATAGTTT TTTTTCGT TTTTTCGT TTTTTCAC CCCCACC CCACTCG CCTCTGG GCTGCAC ATCCCAC	MAGHKITMSEGKLLLEVAVRYGGVRNVSYECP VPDCTKTFSQANNLIRHLNCFGNTKHRAHNFT YFFTCEKCKIQIHSNTKHNIHNHYKQCCATGG GPSCETGQYFCPACEQAGLGNLESALRHFQSS HPEFNLPPRSQFSKSHPNSTLSLKPKDHLMKI LYSGPLTPGQLVCPKICLRSSAARLFHDVSKL RKHMLVDHNRTLVEYETTCGKCLRPVDTSKNM RKTTSHFEEKCSGESFISSPSPIPQKTYKLDLPSTS TPPPRKSPKLQPYKPIRTFKNPLTKSSQSKSDNP PKPTPFFSPRTLERSASWPALSEVVDPLPKLKE KHPSLPCALDKCPPSPRIKPSTLVPPCHTANNSP KPTSPESPSTLKPLPRPIRPSKPLEDWLTVRSVG PDREIVLNIGPRPRPGPAAGSRTTSPSTAPAKR VAANPIAAPLSGEPGATLDCGQTGRKVQPPKK RPTESAGSLPPPAEPATDLLTGREGRLARLVEEY HLSGDFGAFCRDLERWTALSSNRRPKPRRGR YNRGAAARATRNRRGRDDRQDPQDRDDQGGP GPVTCGRPQRYKRAAALRSAFGRDMKATVRR IIDGERGDARCEIDPKTIEGRFRDELSPPVREGP ECSLPPWMAEAQAGEHAPSNDSSQPGDAYDGPI	cd0165 0	710	981

			ACCAA	ACGTAGG	TALEVEMVLSTLNVGSAPGSDGLSYGFWRAL			
			CGACG	GACCTGT	DPKGLVLSELFVCRIERRVPGPWKSSRVTLIC			
			CTCTG	TTATATT	KDAEGDLDDLGNWRPISICQTVYKIYAAVLAR			
			CAGCT	ATTTGCC	RLQSWALDGGVISRSQKGFMPFEGVYEHVFL			
			ACACC	TTTTATAT	DSVVADARATRRSLAVCWLDLRNAFGSVDHT			
			ACCAC	GTACCAC	TIVEALSFRGAPAGLVEMISDIYTGGSCRIRTRA			
			CATCT	TTTTTAA	GFTPDIPVGRGVRQGCPLSGIIFNLVMEVLLRG			
			CGCCG	ATATATT	VEANNACGYRLSCAGGASVRVLAYADDVAL			
			CAACA	TTTGTAC	VGSSRAEXKIQLGVCERFAAWAGFSFNNKKC			
			AGAAG	CCCACAA	AAMVLKHQRGGRRLLDSAPLRLCGEEVAILGP			
			AATTT	GATGCTT	DSFYKYLGAHTGYGRQTGGQLVDRVERQVVR			
			TCCGC	TTCGCCA	LFTSFLTPTQKLSALKRIVLPAMSFHLRVRPCA			
			TGCTC	AAAAAAA	EGHLRRLDNTVRRCVKTALRLPKGSCRAFFHT			
			TCCAC	AAAATTT	SPDAGGLGITSVVAECDILTVTQAFKMLSSPDH			
			TCTGC	TTGTATC	LVSLVAKGRLGMHAARMGRSETASACAMAD			
			TCCAA	ACATTTT	YLSGDSVMGHXSWKTGYRMPADLWTATRAA			
			CACCA	TATATTTT	SRRLSLRFSPQPQGEFGLESGTFKIAPRERRSLT			
			CCTCT	GTAAAAC	RRLHHRQNLWWRNQWAALPNQGKTVA AHS			
			CCTGC	ACAGATT	AYAASNWVKGPSSLAPQALFFGLKARLNQM			
			TCTGT	TTTATAA	PTRSVKACYSRAPNYDKSCRRCGAEVETLPHV			
			GGACT	ACTTTGC	LNHCPKSMKSILERHDSVLAEVLA AIPRGTFAS			
			GCTGC	ACTATTT	VDVDRTSREHFRRVGEALRPDIVARRHDGSVV			
			CTTGC	TTATATA	VADVTCPFESCASALDTAAARKIEKYDQLCAN			

			TGCTG	AACTTCG	LRQLYRKPVESHALVVGSLGSWGRTNNTALA			
			GACCA	CACTTAT	ALGIRGAVRSRLAKQLVNLSVEGSHNIWLRWS			
			ACCTC	TTAAAAT	GGIPKDLVR			
			TACCC	GAATCGC				
			GAAGG	ATCTTTTT				
			AACCC	TATATAC				
			TTCGA	ACCAACA				
			ACCCA	CAAACAG				
			GCAGC	GATGTGC				
			AAGGT	AGCTCAG				
			ACGTG	GGAACC				
			TCACC	AATCCTG				
			ACCTC	CGTCCCT				
			TCCAC	CCTAGCG				
			AGAGC	GCGGGAG				
			CAAGG	GGGCGCC				
			CCAGA	CACCTAC				
			GTGGA	CCCCACG				
			TAGAG	CTCCTCTT				
			CAGCG	GAGCACC				
			GCCTC	AACAGGG				
			TACAA	ACTCCCT				
			CCAAG	TCCGGAG				

				TTCTCC	CCCCTGC				
				ACTCC	ACCCTCA				
				GACGA	ACTTTTCT				
				CAAAA	TATTTTTA				
				CACCT	AAAAAAA				
				GCCTC	AAAATCA				
				GTGGC	TATATAT				
				CAGAG	TGATCTT				
				TCCTG	GACGACG				
				CCGAA	GGGGCTA				
				GAAGA	CATTCAG				
				ATCTT	CCCCCAA				
				CGACC	AAACCCA				
				CCAAC	CCCACCA				
				ACCGT	TCCCCAA				
				CTCGG	CGAGTGC				
				CAGGG	CGGGGCA				
				CTCCA	TTGAAGA				
				AGCAA	GCTCCGG				
				CGACA	CACAATT				
				TCAGC	AGCACTT				
				CACGA	AGCTTAT				
				GTGCC	TTATTATT				

				CACCG	TTTTGTC				
				GAGTG	AACATTT				
				GACGC	TTGTTTTT				
				GGCGA	TCAAATT				
				ACCCG	TTTTCAC				
				AGGAC	CCCTCAC				
				CGCTG	CCCCACC				
				TCGCC	CTAATAG				
				AAGAA	GTCCCTC				
				CAATA	GGGCTTG				
				ACATC	GGCCCCT				
				ACCGC	TTTTCGT				
				CTGCA	GCTCGAG				
				GCAGC	AAGCGTC				
				TACCA	ACATCGC				
				AACAA	CCCCTG				
				AGGTT	ACCACGA				
				AGTCT	CCTTCCC				
				CCTAC	CGACATT				
				CTCAC	GGAGTCC				
				CTACA	TTGGCGT				
				AACTC	CTCCCAG				
				GTCTG	GTCGAAA				

				AATAG	CAGTCCC				
				ACGCC	AAGTGAT				
				CCCGC	AGCACCT				
				GTGGG	AATGCTC				
				AGCTA	GACTTGT				
				ACCTT	TTCGGCC				
				GGTAG	TGGGCCG				
				CTAGC	CCGAGGA				
				TGCAG	TTCCCAG				
				TGCCC	AACGACC				
				TGCGA	ATTCTTCT				
				CAGCG	AAATAAT				
				GCCTC	ATTTATA				
				GAAGA	TTTCAGA				
				GCTGC	ATAAAAC				
				AGAGC	TATATAT				
				GTCAG	ATATCGT				
				CCTGC	TGGCGGG				
				TTCGA	ACTTGTC				
				CCTCG	CCGCCTC				
				CTTGTT	GATACCG				
				CTCAT	AGTGCTG				
				TTCAA	CAGAGCG				

				CCTAC	GCAAAAT				
				TCCGT	AAAGAAG				
				CCTGT	AAACCGA				
				GGAAT	CGTCGCT				
				CAAAA	CTGCAGC				
				GAGCC	CAAGGAC				
				CCACT	CACCCAA				
				AACAA	ACTCAAG				
				TTACA	CCAGCAC				
				TTCAT	CGTCGAC				
				AAAAT	AACCAAC				
				CTAGC	ATCCTCA				
				AAGAA	AGTCGGC				
				CGAAG	GGTTGCT				
				AGAAG	GGAACAA				
				CGACC	CTCATAA				
				AACTT	CATCTTC				
				TTAAT	AANATAA				
				CCATA	ATTATCA				
				ACTTTT	CCCTGTG				
				AGATC	CAGCAGG				
				TTTTTA	AGGCCGT				
				TTTATT	GCTTTTA				

				ACTGT	AAACTAC				
				TTTTA	TCTGTAG				
				AGCCC	TGGCTCA				
				TAAGC	TGATAAT				
				ATATT	ATTTCGC				
				GCCTT	TCCTTTTT				
				TTTTTA	TGCCCCG				
				GATCT	TGTAAAC				
				TATAA	TTAGTNG				
				TTATA	ATGCGAA				
				AAAAT	TAAAATC				
				AGATT	AGTTGAA				
				CAAAG	TCA				
				TTAAC					
				ACCAC					
				CAGGC					
				CGCTA					
				CAGAG					
				CATTTT					
				ATTTA					
				ATCAA					
				TTTGTT					
				ACCGA					

				CCTCC TGCTG CTTCTT TTTACT TTCTCC AGACT ACTAC CCGGA TACAA CCCTT GGAAA CGAGA GGA					
Pene lope	Pene lope- 1_C oFl		Copidoso ma floridanum	TCCAA TGCTTT AAAAC GCTGG AAGGA CCTGC ACATA CAACT ATCAA CCAGT	TAAGCGC TGGCTTG TTGTTGT AATTTAT TGTCAGT CGATAGC ACGACAG GTGACGA GTAAGTC AACCAAA	MAQRRFLLKCRYARVFPNHIKHLYNKCSSVNF RCKPIHADFNKSTLHFCQKLLNFEISDINTHIRF LQNSISRLKDHIASLKLDYNIRERHYKFVNIKQ NTIYRRSNHTLHKKFKKILPKTNTTTSNND SIH NLTNIAIPKDVAEIASLGPNFNLSTFRERDFLC IICKNLEVGLSHLDTNDRVKTEIRDKVSQYLVN ACKQKPAISRKDRQYAKNLKSTKNFLNKNPDI FFTSADKGNVTVCCLKKSDYTERMHNLLSDKST YKIISKNPNSLQKRTSNILKNLNSNGFLSRTY HPNALS LTDTTLAKCYGLPKIHKVDAPLRPIVS	SSF56 672	363	534

				CTTACGA	LINSPTYFLAKVIYDELKKSVNPISHINNSFEL			
				ATTCTCG	KNKLKNIKPTTNYRLMSLDVSSLFTNIPLELVL			
				CGTACCT	RSIDRRIVTIMNKCRIPVDEIRQCTEFLFNNTFF			
				GCGGTAA	TFDNIFYQQIFGTPMGSPISPLFADLVMDDLLEN			
				CAATTTA	DCLSILKNNYNIIPSFYFRYVDDVLLFINVDHID			
				CATCAAA	QILNIFNSYHPKLQFTYEMEIDKKLNFLDLTIIR			
				ATTTATT	NSNYFNFDWFEKPTSSGRILNYFSGHPVSYKR			
				GTATTTT	NIVFNLVDRAILLASTRFHGVNIARVKNLLIKN			
				GTGTTTG	NYPFNFINKNIKVRLKRHKYSNHNIRSTTNTYP			
				AGCCTTC	LCKIGIPYIDNNFEKISTMLKPYNIGHTYPLINNK			
				CAGATCA	LQHIVRLGKDDTTNKHKQTNVVYKFYCKNCPA			
				CTCATTG	SYVGETKRAAEVRISEHMKSKNSDSVVCAHQI			
				TGTGTTTT	NFDHEFDWDNFKILDKESNYKKRIISEMLYIKL			
				TAATTTT	DKHNINRKEDVLSFNRAYLPLLRFKFCPDRVH			
				AATTTTG	RPLPSTFAAPPI			
				TAGCCTG				
				ATGAGGC				
				TGACGCA				
				TATTGTC				
				AGCGAAA				
				CGTTGCG				
				CATTAAT				
				AAATCAA				

					CGCAAGA AAATTTA AATACAC TTTACTT GTATTTA TTTATTTT CTCACAT TTCGCCC CCAACGG ATCACTG ATCACTT TGAATTT CCAAATT CA				
Pene lope	Pene lope- 7_C oFl	.	Copidoso ma floridanum	T	TAAATTA ATATTTG TGTTTTTA TATCCTTT ATTGTAA TTTGAAT TTATCGC GCGTTTT CACTGTA	LFNNICFDAVKMAIEKFWHHIHNKINLSLTDFL NVIEFIMNNNYFQFNNNFYLQRRGSAMGSPISP VLADIVMRDLETACLNLCPFNIIPVYFRYVDDIL LCVPVDKVEYVLNVFNSYSHFLNFTIEKSENN TINFLDMTIVANNDSSISTNWFRKDTHTSGSYLN FYSHCPISQKKAIVSHLVDKNYKLSSPEYFSNN FKYIYTVLLANNYPLDFIESCIKKRLYFLKSINN SQNNISPVNNAIGSLVIPYCRELDYKLRVFCSK SFINVIFKSSNKLPFNRMYGKDILKKLNNPNVV	SSF56 672	34	146

					TAGATTG	YKIDCSDCSNCYIGQTLRKLSIRLNEHKSDIKH			
					AATATAT	KREKSALYQHHLDTGHSFNFS DNEINIIDIEQN			
					ATGTAAC	YYKRLLSEMIFIEKTNNFNRLDCDNLNKVFK			
					ACATGTT	RYIHLLLK			
					GGTTTCG				
					TCTATAT				
					ATGTATG				
					TAAAATT				
					TGTATTT				
					GATAATT				
					TGAATAT				
					GTAAGGT				
					CAGTCGA				
					CTATGTG				
					CCCTCGT				
					TGGGTAC				
					TGACGTT				
					GTTTGAA				
					TCTTCGA				
					ACAAAAT				
					AATCTTT				
					TTGTAAG				
					TCTCTAA				

					TTTGTA A				
					CTGTATT				
					TAAAATT				
					GTTTATT				
					ATTGTTC				
					TTAGTTTT				
					TAAAATG				
					TTATTAG				
					ATGTGAT				
					AACGTGG				
					ATGTATT				
					GTTACCT				
					CAAGTTA				
					ATTTATTT				
					CGTGTAT				
					ATTTTAT				
					AAAATTA				
					GGACTTG				
					AAAAGGC				
					TCCTACC				
					CAAAGAG				
					GAGAGAA				
					ACGTCGT				

					CAGGGGA TATAATA AAAAATT ATATACA ACAGCAA TAGCCAT TGGTTTT ATTATTT GATGAAT TTCGTCTT GCTCGCT GTTGTTG GAGTCAT AAACATC ATAA				
Pene lope	Pene lope- 5_C oFl		Copidoso ma floridanum	TC	TAACAAT TTTCATA ATTTGAT AATTCAT AACAATT TATATCC ATCAATC TCTTTTAA	SDINIYKNYLQKELDKIEKSLIFTLGIRFFELFER FFHYKSKIGIKTFDDILTNKFKILYFRKYKHTYE KKININNNSIRVSKTNSNIKSKWLHNLTINLPK EVQFCINLGKNFNFFDFENQNETKRVYLELFKS LNKPLHKNQISISNNIKFKILDVVKTHIKNNNKS SLDILISNMLNNCNKFLKNNKNIMITKSDKGG AFVIIDKQDYINKVEESLNDKKLYEVVEHNPLP SLINKTSNLLNELNGSDLFGDSGNKKVFNTIRD	SSF56 672	321	494

				TCGGTTA	KCILAKCYGLVKVHKPNQPMRLIISNIGSPTYF			
				GTGGCCA	LSKWVSEFLKNNLPPSPHNVKNSFELKNDIKN			
				CTTTCCTT	FLIPDGFCIVSFDVVAlFNNIHFESVKVALESV			
				GATATTG	WHILNLKINLEFFSFLNIIRFIMENNYFSFNNKN			
				AGTTTGA	YRQKFGSAMGSPCSPSLADIVMRHIIKMLLFLG			
				TTCTTTA	PFEIKLIKVYVDDLLACIKISESDKVLNYLNSID			
				AAGTAAA	KHIQFTMETPNNNSINYLDMTLTWDNNGSIKT			
				AATAAAC	NWFQKPTHANLHLNYWSHSPLSQKIGVIYSLV			
				TTAGTTT	DRAIKLSDTCYHNNNLKKVYEILKENNYPPDFI			
				CATAAAT	KKHAERRKKIYYNNKNDNNNNNSNKKLNDDN			
				TACCGCT	KNYIVLPYIKNCDSNFKSLLNKVSYKTIFKSVN			
				GATTTTG	KQKFLNQLGKDSIKNKDKVCSVYCIKCKNCD			
				TACTACA	KCYVGQSSRVVSKRINEHKRDVKSKKNESAIF			
				CCAGGTT	NHVQDNGHSFDFDNFKVIDNEPNFYKRLISEM			
				CACCACT	VFIKKYNTVNKVTDTQSLSSVFEFYINKYL			
				GTTGTAG				
				TACTACA				
				CACACAT				
				GTTCTCA				
				CGTAGCT				
				TTAAAAA				
				GAATTCC				
				CATATAT				

					TGTACTA				
					TATCGCA				
					GTTGAGT				
					AGAATAA				
					ACATCTC				
					ACATTGA				
					CGACACT				
					CTGCGAT				
					TATTGTTT				
					AATATTG				
					TAAGTAA				
					AGATAAA				
					ACTTTTA				
					TACATTG				
					TTAGTGC				
					GTTTTCT				
					GTTGTGT				
					ACATCTT				
					TATATAC				
					ATATTTA				
					TATTACC				
					ATTGCTG				
					TATATTA				

					ATTTTTA GTGACTT GATAAGG ACTGCCT TGGAGTA GTCGAAA CGTCGTT TTTTAAA ATAAAGT CAATTTG AACAGCA TAGCCCT TAGTACT TGTCATT ATTTATT GTATTAT TAGCTGT TCTTGGA ATCAAAC AATATC				
Pene lope	Pene lope-		Copidoso ma floridanum		TGACGAG TGTTGCT TGATAGG	LLKPLEAALAVNKKIEVDTSKKIRQTYLQVIKK NIFNSKKISYFDKTINNHLNETKMFLKEHREIL VTRADKGGATVVINKTDYINKVERALGDSMY	cd0030 4	283	365

10_				TGTGCCA	YHRLDSQKRANHLTRIINSTKKLVDVWDNRG			
CoFl				CTTGATG	VFDTVEIDRPDVNQLSGCNLAKAYGLIKVHKP			
				GCCACGT	GLPVRIIVSAIGSPTYTLDKWFSKFFNKFCRSP			
				TTCTGTT	SSLKNSYDLKSTLRDLIVPHNYLLVSFDVVSLY			
				GGTCCGT	SNLPFSLIKKSVEEKIGRTLDSMHLDRKEFLDG			
				GTTTCCTT	LDFLINSNYFTFNNHTYKQTCGTPMGGCTSPW			
				GTACATA	IAEIVLESLEKSVLPRCRDSCLMYRRYVDDCLL			
				CATTGAC	IIQSDSVDTVLGIFNDFNPFIQFTVEREDNGVIPF			
				CGTGGGT	LDVTITRIADSVCEIDLYSKPSSSGRLLNFFSHN			
				GCACCCT	PVSQKIAIVFSLIDKCFKLADPKFRSRNIDRARR			
				TATGTGG	ALTMNGYPTYFIEKYIKARISIINRFNKNKNHIN			
				TGACAAG	SLTNCEMSPPVMEGSSRLSFILPFYPRLEKTLR			
				ACCTTAA	SFFSKHKLKTIFQNSNKFTNIIRLGKDQTPKLKR			
				CACCATA	SGIVYRIECECGLFYVGQSKRRLATRIDEHKRM			
				TTACTAT	VHKCDKNSVIAAHCSEFSHTFKFDEVKILDQE			
				TTTAAAT	XNKFKREFSEMLYISYFNRRNVNRKLDTNKLHD			
				ATCTTCC	HYKSFLYELRDRMHGWRC			
				AGTTACA				
				TGTGCTG				
				AGTCAGG				
				ATGTGAC				
				TTATATT				
				GTAGATA				

					ACTTAAC				
					TGTACAT				
					ACATTTT				
					GACTGAT				
					AAAGTTT				
					ATAATAT				
					TAAGATT				
					ATTTAAT				
					TATTAGT				
					ATTCCAT				
					TGCCGTG				
					GTTATGA				
					CTGACTA				
					GTCTTTA				
					TCAGTCA				
					TGCTGCG				
					TTTGGCA				
					TTTTCAT				
					GTTCGCA				
					TCTGCAT				
					AGGTGAC				
					TTATATT				
					TCATTTA				

					ATGTAAT				
					TTGTCAT				
					AATGTTT				
					TTTATATT				
					ATAAATA				
					TTGTAAA				
					TTGTACA				
					GTTCCCT				
					GAAGAGG				
					GAGTCAG				
					GATGACT				
					CCGAAAT				
					ATTGGAT				
					CACATTG				
					TCCAAC				
					AATATAT				
					GAAATAA				
					ATAGCCA				
					TGCTCCA				
					AGATTCC				
					AAGTTTT				
					TTCATTT				
					AAATTTC				

					CTTAATT ATGGCAA AGGAGCA AGAAGAC AGTAGTT TTATTAT CAAA				
RTE X	RTE X- 22_ CGi		Crassostre a gigas	GA	TAATTTA TTCATTCS AGGATAT ACACTTT GATGTAA CATTGTG ATTTAAT ATGTACC ATAACCC AAATATA AGTTAAT AATTGAT ACACCTT GTATAAC GCGTGAT GAACTAT	MDNPTYSPNETLNRNFLFDEINRLVFKTKSKSA SGYDEIPYGVLSQKVETLTRLFQLILDTSIVP SVWRKAIICPILKDASTDPRIPMNYRGVSLLSCI SKIYSAFINQRITSHLDAHEYLMSTVSEKTGR VKTRIHLNSVIRIKITIRSFHDLRKCFDFIDREM MLYKLLVHDIDGKVYNSVKSIYQNSESCVRW QSVQLCEKYLPKLRIVCALNGQLTNSCNTGVK QGDNRSPTLFSIFINDLVQDINELGLGVDVGEK KLSLLLYADDIAMIAKSKAELQRMLDKLHQW CKRWRVLINTNKYKCVHFRKGKIKRSNCNFTI GSNALETVDHYKYLGVTFNYNSCFNSNAELL AKLLAKSGGRALGQIIAKIHNNKYFGFTSYEKL FSSCVAPILDYGAGVWGYRKFAIDNIQNRAI RYFLGVHRFTPILFLYGDVGWLP SQFRRWFCM LRYWNKLCIFEDDRITKMAFNLDYERCRDNW CSEVKHIMNQIGIGQYYDTKCAIDLGLVESHLT	cd0165 0	72	345

					AATTGTT	NFYSEIWKDSLENVPKLRITYKTFKTDGLENYI			
					AATATTA	SFNLSKAERSHMAQLRXGILHXRDPPPLRIESGR			
					CCAAGGT	YVGEPPNERLCTLCTQQVIENELHFVLDCPFYN			
					GACTTAT	QYRTEILGEQLLSPALLNRDRQYVFTYLMCHK			
					AGGCCCA	FRQLSKFVVKAFRIRRRQFLYR			
					ATGGGCC				
					GGGTGCC				
					CAATATT				
					CATGTAC				
					AAGTAGT				
					TGTTATA				
					AAATTTG				
					TATCATA				
					CGGCACA				
					ATTGTAC				
					ATAATAT				
					GCACATG				
					TCACTAT				
					AATAAAT				
					AATTTAC				
					TTACTTA				
					CTTA				

CR1	CR1 - 1_C Gi	Crassostrea a gigas	GAGGA	TAAATTC	MGNSLEQYRAAIGNFYFSCRCVCRSVLVNFFG	cd0165 0	577	830
			GCAAT	CTTTTTGC	FSFFSKIFMFPSIGLFALAMKHLSRHIVVYNFYL			
			CAATC	TTTTGCTT	QFFMFLLLLCCGDIEVNP GPIMTNVLDILHLNIRS			
			AATCC	TCTATAC	IRNKVANLNTIVHDFDILCFTETHLDRNVTNES			
			AGTCT	TGTACAT	LMLDGFNTIYRKDRNCFGGGIMIYLSDLLRAIR			
			TTTGT	TAACGTT	RPEFETTCNESIWIEIESHTCSYFLCCVYRPPNS			
			ACATT	ATCTACT	DGNFWTSLSWSIEKVGEITDNIIIGDLNVDLLSI			
			TATTTT	ATATATT	PQTHVIHDIMSNNGNFVNNINEPTRVTNTSSTLI			
			CGGTG	TGCAATT	DLVLSTPAVKVHESGVLNIDSSVSDHKAIYISF			
			TTATT	ACTTTGT	SIDIELNRAYERNIWVYKDADFELNSLISNCD			
			GATTT	AAATGTA	WDITINSAPNIDAAVAQFTSTFMAYVRECVPE			
			TTCATT	TGTCTAT	KTITIRPRDKPWYDSMLRKTMRIRDRLRNKAL			
			ACACC	GAATTTA	RTNNDADWTKFRHVRNKVNNMKKHAITNYY			
			ACCTG	ATTATTA	DNIESALHDSSRNGKIYWKLMDLKFVKATTE			
			TTTTAT	TAAAGTC	IPPLQFFENGQPTLAFSDIEKTETLNKYFSSISNI			
			ATAAC	AATAACT	ESANRSLPNLYYMCNNFFSDIHIEVQEVFDIISI			
			TCATA	ATAAGTA	LPVNKAIGPDNISHRMLKYTKFSISKPLCMLFN			
			TTTTTT	ACATGTA	KSLVENSFPPNYWKIAHVLPFLKKDDPSVASNY			
			TCATTT	ATATTAC	RPVSLSCVSKIMERIIFKHVYNFFHENNLFYK			
			TCTTTC	ATGTACT	YQAGFLPGHSTVFQLLETYHSIVKSIEDGKMC			
			ATATT	ATGACAA	CMIFCDLSKAFDRVWHEALIFKLQTYGISGHLL			
			TACTA	TTGTATA	QWFKSYLSDRVTQRVVMYKNKLSSTLCINAGVP			
			CATTC	TATTGTA	QGSVLGPLLFLIYINDVATNMLSFCRLYADDN			

			TAACG	TGTATTA	SLQHCDLYTKNIEIVLNHDLKILSDWSDKWLL			
			CAATA	TCATTAG	KFNPSKTKAVFFTKKHVEEFPKLFFQGCQLDY			
			CGTTT	GAGAGGA	VSTHRHLGLLFSNDLSWSQYINNIVNSAYKKL			
			ATGCG	ACTTGTA	GLLKKLKFTLGRTNLSKMYITFIRPVLEYASVV			
			TTATTT	AGTTGTA	WDGCNMFDMERLEKVQLSAARIVTGLPILAST			
			AACTA	AGAACTT	DSLYTETGWESLVSRRNKAKLITMFKIHTNQV			
			CCTTT	GTTTCCG	PEYLSNIIPPVTKNVSIIYHTRNSENYSVPISRLEL			
			GTATA	ATCCTTTT	YKKSFPVPTIRKWNSLKLEVREATSLNIFKKNI			
			TTGTT	GTGTTAC	TESIPQPPKYFSFGKRTINILHTKLRHNCILNYD			
			ATTTA	TTACACA	LYRRNIIESPNCICGQKEDVYHFFFVCKNYVNA			
			CCTTTT	ATAAAAT	RNNLFELLFNRLDETILINSHLLLWGSDNLSYN			
			GTTTT	ATGTTCA	TNCFIFSAVQKFINESGRFNN			
			GTCAT	AATCAAA				
			TGTGC	A				
			ACATA					
			CATTTT					
			TTATG					
			CTATT					
			ACTTT					
			CACCC					
			CTACG					
			TTTTGT					
			TTATCT					

				CTTTTG					
				TTGAA					
				CAATA					
				TTTTGT					
				CAATT					
				GTCAT					
				ATTGA					
				GGTAT					
				CGTCA					
				TTTTCA					
				TTTGA					
				ATACC					
				TTCATT					
				CAGCT					
				GTAAG					
				TTTACT					
				ACTAG					
				TGGTG					
				TATCTT					
				CCGGT					
				TTTTG					
				GTCTT					
				ATTCA					

				TTACC CATAT ACGAC ACATT ATACA GATTG CAGCC ATCAT TTCCC CCCTA TTGTT ATCAA AAAGT TAATC ATTGA CA					
RTE X	RTE X- 23_ CGi		Crassostre a gigas	GAGAG AGAGT TGTTA GACGA GATGA AGAGA AACAT	TGATTAT GTACACT TTGTTTTT CTTACTG TTTATTTA ATATGTA TATGTTC	TYFCSRMLYTPESXVFYKKGKVDIFDCLEEXY CSFHRKGHIYXVGDFXARCGTLDDFISDDKLG KSVEXXLQSFXYDNNANVYCXHRNADRGVN NFGKKLINLCRTTGLKIVNGRNENDYNGNITF YNANGTSTIDYLLTDVYNFDNIAYFSTGIFNCF SDHAPVSFCIKTNYIDGGSKALCCANTSQGNG SINQCTSVRWKEGADLEATWSISEHYDLFLSY	cd0165 0	477	736

			CAAAT	ATATGTA	VENQLYTQEDVDNCFVANFSKTLCKIMKPYCII			
			CTTTG	TAACCTA	TGGLANTVIDAERCIDKPWFNDICKSKYRDYR			
			CAAAA	TGAGACA	GALHNFNVCKSDVNRIVLCAKKKVYKKNSPE			
			AGAAA	TTTGTCTC	VIELYFVQRKRFIKKTQKFKHNYERYQGDR			
			AAAAG	TTGGAAT	MEYLRKCNPKQFFRLFKNNKKISCDVSAKEFF			
			ACATT	AAAGAAC	EHFKNLFENVKTTTYDNTSDDYEILPCYEELD			
			AGGCA	TTGAACT	VCISELEVEKAILKLKSNKSNAEDGVINXFFSK			
			ACTAA	TGAACTT	CKEVMIPILCRLFNNIFDSGFYPDNNWSKGCIVPI			
			AAKAT	GA	FKKGEMQDLNNYRGITLVSCLAKLFTSIINTRL			
			TCAAT		LEWDQRYNIITDAQFGFKPGVGTIDAIFVLQNL			
			CGTTG		ISRTLKNKKRLYCCFVDYTKAFDLIDRSKLWY			
			ATCTG		KLQMEGFGGKLLRIKSLYKNVRACVKHKGTF			
			CAATG		TETFRISAGVLQGEILSPLLFSMYLNDFFERYFIN			
			TCGCT		SNCTSIELQMINMFLIMYADDTVLLAESPEGLQ			
			CTATG		DMLNTLHRYTDEWNLSVNVQKTKIVIFRNGK			
			AAAWA		CVKDNEKWEYNSKYVEVVNQFNLYGMLFNY			
			CAATC		NGKFLSTQKHCAAQGNKAMFSSSKMKDLNF			
			TAGTG		NVETQLNVFDITYVKSVL SYGAEIWGFHKGQD			
			TTTAC		VEKVHINFCKKVLGVRKNTCNSAVYYEVGRF			
			GGGGC		PLEIFRKQKIFKYWFKLRNSKNCILKACYEDM			
			TGGGT		VKNNDIWISNIRKEL SILGLADLYQSTMKESVI			
			GGCGA		LDIIFNRMCDVFKQKVNDISNASKCILYKHV			
			AACAA		VDHFCLQVYLKPIHV KYRKLITRLRLSSHDL			

				AGGAA	KIETGRYDNLTRDCRKCESCNLNVIEDEFHLL			
				GAGGA	ICPSLCSLRDKYIKKYYSRKPSVYKVIQLLSTSN			
				CACAG	TKELCNLGKYLYYANKQRTLHVNYPN			
				GGGCA				
				AGCTG				
				CGGGA				
				CTCTA				
				TATAT				
				TTTGA				
				ACTGG				
				GAATT				
				GACTG				
				GAAAG				
				TGGAG				
				TTCTGT				
				AGGGT				
				GCATC				
				GGTTC				
				GGTCG				
				ATTTC				
				AGCAS				
				AACAG				
				GAGCA				

				SACCT					
				ACTGT					
				AGCGA					
				AATTC					
				CTTTAT					
				GAAAA					
				AGACA					
				GCTGC					
				AGGTC					
				TGTAG					
				CTCCC					
				GGTCT					
				GAAAA					
				AAGTT					
				GAATG					
				GATCC					
				CAGGT					
				CGTCC					
				ATCCG					
				AATTT					
				KTTCA					
				GACCG					
				GKAGC					

				TACAG					
				ACCTK					
				CAGCT					
				ATGAA					
				AAAGA					
				CAGAT					
				TAASTS					
				TCCTT					
				GAAAS					
				AGCCT					
				ATCAA					
				TTAAA					
				WKGA					
				MGTGG					
				ATTCG					
				GTATC					
				AATGA					
				ACAAT					
				TCCCG					
				CACGC					
				TTTTG					
				AGGAA					
				AGACG					

				GAAAC					
				AGCTT					
				TATCC					
				TTAAC					
				AAAAA					
				TATGT					
				TAGAC					
				ATGAA					
				CATAT					
				CACTT					
				GTAGA					
				CCAAT					
				TTTGT					
				AGTAC					
				TGTTG					
				TCTGG					
				ATACG					
				TTTTG					
				AACAT					
				AAAAG					
				CAAAC					
				AGGGT					
				AA					

CR1	CR1 - 3_C Gi	Crassostrea a gigas	GATGA TAATCAC TAACT AAATCTG CTGTA AGTATTT TTACC TTGTTGA TCAAG TAACATT AAATA TTGCTTA TTTAT TTTTATT GTGTC TATTTAT ACAGT CTCGTTC ATGTT ATCCTTG TGASA CATACAT GTATG GTAACAT TGTAT TAAGTAG TAATG ATAACAT AACAA TAACTTT GAAAT CTTATTA AATTG CTACATT ATATA TATCAAA CTATC TATCTAC TATTCT GTACTGG TCCTG TGTAATT TAAAC TATTACT AAGGC GTATAAG	SHPERIVFKHLYNYFHSNNMFYKYQAGFLPGH STVFQLLETYDDIVRRIDEKGSCCMVFCDLSK AFDRVWHKGLLYKLKMYGIDGKMYXWLKSY LANRTQRVMYKNLSTKAQIHAGVPQGSVLG PLLFLIYVNDVAHNMSSFCRLFADDNCLQYSS KCVKTIENCLNHDLNVLENWSKKWLLKFNPS KTKVIFFNTRKNLEPPKLIFQDSQLEYVQTHKH LGVTFSNLSWSSHIEILLSNAYKKIGLLKKLKF KANRKTLSTMYTSFIRPQLEYASVWVGCHAF DSEKLEKLQLXAAARIVTGLTIFASRESLYFETG WEPLNARRKIQRLNIMYKIHNNLVPEYLSDIVP DMRCNASAYYTRNSQNYNIPVCRLQIYKSSFV PTVVNEWNSLPLDIRDSSSLSLFKNKISKSLAV APPYFSYGKRRTNILHTRLRQNCALNIDLYRC NISSPLCSCGKIEDAHHFFFSCPKYATVREEFF NKFLRFGKIRILDSRSLWGGDSLSVDENKYLE SIVQIYIQGTGRFE	cd0165 0	4	226
-----	-----------------------	------------------------	--	---	-------------	---	-----

				CATAG	TACAAAC				
				GTGCA	TGAGACG				
				GACCT	TATTTGC				
				TATTA	AATAATA				
				GTCAC	TTGTATTT				
				AAAAT	GTTGTTT				
				GTTGA	AACATGT				
				AAGCT	AGGGAGA				
				ACTAA	GGGCGAA				
				ACTTA	TTAAGTT				
				CCATT	TTTATAA				
				GTCAA	CTTGTGC				
				ACCAT	CCAATCC				
				TATTTC	CATTGCA				
				TATTG	ATATATA				
				TTTAA	ATGCAAT				
				CCGCT	AAAATAT				
				CACTT	GTTCAAA				
				AGAGA	TCAA				
				ATCTG					
				TTTTTC					
				CTATT					
				ATGTG					

				GAAAC AAGCT AACGT TATTC CACTC TTCAA AAAAG ATGAC CCTTCT GTTTT GTCAA ACTAT AGACC AGTTT CGTTA TTAAG TTGTG TTGGT AA					
Pene lope	Pene lope- 2_C Gi		Crassostre a gigas	TTATT	TAGACCA TAATTAA TCACCTA ATTGTCT	MKGPKFREPRSFTWRRNFIHIMNSVEDYAKR WAKSEKEELDTLSEWVKSIRCILKSRIKRLRGQ MKTIYPSVFNKAIEVRKELDRLHDQYVLVPAD KASNNIVFVCKAHYINCIFKELGFDSTHGNPTY	IPR000 477	147	467

				ACGGACT	THSSLKQEILQNHKSVMDIFNIPNTQNEFDLP			
				TTTCCGG	YLYWIPKLHKSPYKQRYIAGSSKSTKPLSLLL			
				TTTTTTTC	TKILTVVKEKLQEYCTTIYSRSGVNQMWILKN			
				CCGATTA	SKELLETLRSQNLTKINSIKTYDFSTLYTTIPHD			
				CGACAAA	KLKSRLFDIIDSCFFNKNNGSRKNAYLVIGNLKN			
				GAGCACA	YFVKNHSDCTHKYSEVDIKKMLEFLIDNIYVV			
				CGGCGGG	FGNQVFQQSVGIPMGTCNAPLLADLFLYSYEA			
				TGTGACC	EFIQKLVREKNKSLAVAFNSTFRYIDDVLSINN			
				GGTCAGC	CYFHTYVDSIYPSELEIKDTTESASSVSYLDILL			
				AGAGGAT	EMDIDGNLTTKLYDKRDDFNFSIVNFPYLC SNI			
				GCTCACT	PSSPAYGVFVSQ LIRYARACSSYEQFLRRGKLL			
				CCTCCTA	TNKLKQDYQQSRLKSSFRKFYGRYNDLVSKY			
				GGCACCT	NLPLGRMLTDVFHTNC			
				GATCCTA				
				CCTCTAT				
				CTATTTG				
				GAGGTCC				
				GTGTTGC				
				TCTGCTTT				
				GAATTTG				
				TATTTG				
				TTTTATG				
				GATTTTT				

					GAGATGG TTGACGG TTTGTTAT TGTCATT TTTTCATT				
RTE X	RTE X- 26_ CGi		Crassostrea gigas	TGCGC ATGCG TGAAA GTATC GACAT GGCGT AAAAT TCAGC TGCGA GAATT TTCTGT TTATTC CAAAA TTTWG GTWGT TGTTA TGTC A ACAAA	TAATATG TGTATAT TTATGTT ATAATTG TATTTGC AGTGACA TATAGGC CCAATGG GCCGGGT GTTTCAT TATCTAT TTTGTAT AACATTT TGTAATT GTTATAA CACATGT CACTATA ATAAATT	RAILTLSCPFIKIKSTHCHRPLTLLWNDLXWLR MGIARFMKXVSVGLYXIPWKTQISQLLQQIKF EPSEDILSVAKELVAHLDDSAQVVAAARLKNR KIDKPGLVKISFARVDQKIRVLRRAKRKLGEISD YRSVFLRSSKSHTERLIELNARTILNEMPNGKD FRITANGRIVKKTDNEPNTRDESSSELPRPFETI RIGHLNVCGWTNKNHDLRVEIHKYTDCDIVSIN ETHLLNKENLDLDGYQWMGLNRSNVHVKAP KGSGGVGLFIKKHLCKNYKVDIIDNSFDGILGV KFTNKFSDYVFVVFSCYLPPESPWGRDASSFF SHLLTQIYIQYDSDSIFICGDFNARISNIPDFDKD LDTIPPRLTIDNVHNQHGSSFLEFLNDCKFCVL NGRFGNKSNDYTFQSTRGHSVVDYICVPHDV YNSVLRLSNNFLMEDLAINQMTIHFNRQEVIP WLTIIYVFLMMFTIQCSDFQIISCSDIVENAGLVS MQGENCRIPDHGFLIFDFMIRDFENAHPSNTV TSSTRPIKKYKLRCIPPSFMSSVNVRGSLNLMIR IETCRESRENIDRTYTQLCNIITKEMECSIPKFD	cd0165 0	831	1088

			AACTC	ATTACT	WSKRSRKKFRTHKPYWNNELTELWRVMREK			
			CGACA	TACTTAC	EKTFLKHKKGNSVLKQKSFSEYKRAQQELDK			
			AAAGT	TTAC	RMRFFRRKYNLDLAINLDNFSSNNPKKFWHEEL			
			CCCGA		RSLGPKRKNKVPTECYTDDNDGSISSDPTYVQ			
			CTAAC		NVWKSEFCNLYNPSTDVAQQPSNDAFLQRVK			
			GTGCC		ESVNFMENNMLDPLREHNPYLNRPIELDEVKN			
			AAATA		VIMKSKNGKSSGVDEIPYEVLYDCVIEVIHRL			
			AATCT		FMLCFESHIIPSVWRQAIICPILKDSNSDPRVPL			
			GGACC		NYRGISLLSTIYKMYSSILNHRLVSYLEDNNLL			
			TGGAG		VDEQNGFRRERSCADHVFTLNSIIQNRKETFVA			
			CTGAT		FIDLQKAFDTVDRDLLKFSLNNGIDGDFYNSI			
			GTTGG		KSIYGNTVSCVRVSELCTDWFSTSAGVRQGDN			
			GAAGC		LSPTLFALFINDLAKEIKELRKGITVGEINLGILL			
			GAAGG		YADDIAILSDSKENLQKILDTVSYWCEKWRLH			
			CTCTC		INMKKSGIIHFRKKGPRSDFQFQINSNILQYNTD			
			GTCGG		YKYLGVIFHEKMNFQNNADTLISKAGGRALGG			
			ACGGT		MISKIRSIFPKFGPYKEIGISTFEKLFLNCVAPVL			
			CAAAC		DYCSEVWGLQNYHSIEMVQNRALRFYLGVHR			
			CCCAA		FTPLPALHGEFGWKLCYHRHCINAVRLWNRL			
			AGCCA		LTMEDDRLTKRVFLWDFNQTGLSRWTSRIKSL			
			ACACC		FESMNLLHIYDNLSICDVNEFENKLFLTYENE			
			CAAGC		WLGKINQLPKLRLYKSIKTNFKTENYLEMNIS			
			AAGTC		KSQRSILAQFRCGILPIRIETGRYKGEPLDERIC			

			AAAAT	NFCVLNEIEDESHFLLNCTIYSDFRRELFNNIGY		
			GGCGA	NPTLVMSNLDLCIFFLLSNFPRQTSKYLJESYLY		
			CGAAC	RQSLLYKKRNA		
			AGCTC			
			ACCGA			
			ACACA			
			CCCGG			
			CAAGT			
			CCGGT			
			AGGTC			
			AAACA			
			TTTCTC			
			ATGCC			
			GATTA			
			TGAAA			
			CATTT			
			GTGGA			
			AAAAA			
			TGACT			
			ATTAT			
			TGCTG			
			ATGCT			
			GTTGA			

				TAAGT					
				TACAG					
				AAAGG					
				GCAAT					
				CGACA					
				CTACA					
				ATCAA					
				TTGTT					
				GAAAG					
				CAAAT					
				TAGAC					
				AAGTT					
				CAAAA					
				ATGAG					
				TTCGTT					
				CAGGG					
				CATCA					
				ATGAA					
				AAATT					
				CAAAG					
				CCATG					
				A					

RTE X	RTE X- 21_ CGi	Crassostrea gigas	ATCAC GTGAT ATGTA TGGAG GTGGA CGTGT TTTTAC CGGAG TACGA CCTAC GGGAA TTTCAT ACCAT ATATA GCGAA GTTTA GTAA AGGTT GCGGT AAATG TGATA CACCA TTGAT	TAACTTT AGATGTT GTTCTTTT TTGATTC TTTTCGC CTTTACTT ATGTCAA TTTTTCAT ATTTATG TATGTAC AATGTAA TAGGGTC TCCTGTA AAATACC CACTTCT GTACTTT ACTATGT TTAATTT GTAAAAT CTGATGA GCTGCAA AGCTTAA AGAAATA	MNLSSPSPSKPPKKKKQRTKSSTQGKMATTQQ GACTNSVYPNDNLSQSVQPIQQVGNLTYHQLQ PMQYIQSAHAPMQTHMQHQPQNVQSNTTPQVT PQWVSDIVTQLKNMNQKLDITDITLSSMKREI KGLSNRVVEVEKSQEFISKSLEDNEKNTSKALT ELSDLKMALDQSLQENVritesVLDMQCRSM RDNLLFFGIEEAQSGEQENCEDIITSLCASKLEI RSEILIERAHLGSKKNNKPRPIVVKFNRFPHR EMVRKCSYKLKDTHISIGE QFPKEIQERRKLL PIYKRAKENNQKAVLARDKLFIDGKLYSDDGP VQMKATPTQPKASGRRGQKHGRGKRGQENQ TGLRIIAWNIHGLENKWSERNFKSFLCTYDII CLSECWIGKDFNISDQNFQSFVFPKSGKSQGG GTVVLIRNVYTPFISFVESLHDTIVWLKIDKSFC NHTKDIYIAFTYFPPSNSTFLKERDIDLFMELEN QLSYYSIGDIFLFGDFNSRTKTLPDYIVNDQV HESVLDNNLQYSSDNQLTERVNPDLGHNEYG QRLLSLCRSTGVRIVNGRHSGGHSNDYTFNGA RGLSTIDYLLTTYDMFDFISKFIVCNFTTFSDHA PLHIEFKCTLLDPCDTSSHYRTRNVFRWDSSY VDSCNDFNSIEGTCNFVDFEKVTSQSDLDEC MLKITDMLSKDIGNFFYVSSKNSNQSKPANSR SRYSTYLSTNDKPWFNENLKSLEYRDIYITALRT	cd0165 0	920	1181
----------	-------------------------	----------------------	--	--	---	-------------	-----	------

				ACGTG	AAGAATG	FNVCKNAQNHLNLVSKKRTYKKLEAKLKRSY			
				AGTAT	AATG	KQIEGNRLNILKKNNPRQFFAKFRKKPKSVGN			
				TATAC		VPIEIFFDHFKKITNVDGSFDTNCEQANMEVLIP			
				TCTGTT		IVNKPTDECVFYELDKEIDVDEIRKAISNLKRG			
				CCT		KSHGEDGILNEYFINFQQHLLPLLYSLFNCILDT			
						GFFPKSWSSSIIVPVYKKGDCTNPDNYRGISLV			
						SNMAKLFTSILNNRLLDWSKTNDVITDAQFGF			
						KPNCGTRDAIFALHSIVTNTLCKRKRLYCAFID			
						FKKAFDSIDRSKLWLKLSKVGIQGKLLNVIKA			
						LYNDVKSCVVVNGHLSDYFCNNVGLMQGEV			
						LSPILFNLYVNDFEMNFLKSGCIPHELLSLNLFL			
						LMYADDMVLFSESVEGLQLLLNELSHYCKTW			
						NLCINVEKSKIVVFRNKGKIKCNEKWCIGNDV			
						LEIVEQFTYLGIIIFQYNTKFTKAQKQLSDQGRK			
						AMFAFRKNIRGMILNHETLLSLFDCYVGGIINY			
						ASEIWGSHKGNNVEKLHLDFCKQTLGVKRSSF			
						NAAVYAELGRNTSYYKKVLFNFKILEGCVMV			
						GIPLITKRFCILKFWKDVLCNNCHIQKCYEEM			
						YINCELHGHKNWASDVKKILNEIGFNEVWNK			
						QTIDSLTLKVINQRIADLAKQDIFGKIENSQKC			
						HFYKYLVDGFYLQYYLRKSPVKFQQCISKLR			
						LSSHRLAIETGRYNNTHRAQRNCFSCKNYVED			

						EFHFILVCPVYIEYRQKYIKKYFYEKPSMFKLL QLFNNENFKDMCNLGKYIYYSMKLRNDICN			
Tx1	Tx1- 5_C Gi	.	Crassostrea gigas	TATAA TTGTA TTTCAT GAACA TTAAT ATATC TCACA TAATT TTGGT ATGAT TTTTAT AATAT GCTGT CATAT TAACA AGCTC AAATC ATCAG AAAAT CAAGG TACTTT	TAAAATG TGTTTAT AATAGCT GCAAGGA AATTGTA AGCAAAT AATGACA AAAAAAT ATAAGTC TACTTTT ATGTATC CTCTTGC ATATGTA TGAGTTT TTGATAG ATAGTGT TCCCCTC TCTTTCTC CACCTC TTTCACC CCCCCCC	MDDKVQQSYLSILTLNCQGLGDINKRRDIFKV LKSKNYNIYFLQDTHFVEKEESLIKGLWGYKA YFNSYKSNSRGVAILINNNCDITILNEHKDSGD NFLLLDVTIESLPFLLINIYGPNSDSPAFYINVLN KIQDIYSNQHIVIGGDFNLVFDKDLD SMNYKN LNNPKARLEV LKIMETLTLKD VYREHNPTLKK FTWRKKNPIKQARLDFFLISESLNLSLYVKYE NSYRSDHCPVVL YVRTNEFLKGKGFWKFNNS LLKDKEYVYLIKEKIKNVKQQYACPVYNLENI ETIDDNIQFIISDQSFLDTLLMEIRGKSISYSTYK KKKTTEREKQLEQEILLLEQNFTVETKDELSNK QTELELIRKNRLQGQCIRSKVKWIEEGEKPSKY FISLES RNFSSKIIPKVEKEDGSTVFNQFEILNEL RMFYDNL YQKRETIDSENIYEKLKLYDIPKLN KQESDLLEGPLNKLEVYNFLKNMKNDKSPGP DGFTSEFFKFFWNDLGSFITRAINNTYELGHFT ESNKLGVITCLPKAGKPKQFLKNWRPISLLNVL YKIASGCIANRIKTVLYKLINEDQTGFLKGRFIG ENIRMIYDIMQYTEQNQIPGLMLIDFEKAFDTI SWKFIYQTLDDFFNFGISIKNWIQTFYSDIKSCVI QNGIASDYFFPQRGCRQGDPI SPYLFLLCAEVL	cd0165 0	530	791

			TTTTTT	ACCTCAT	GILIRNNKEIRGIEIDGVEYKLSQYADDTSIIFDG			
			AAACA	AACTAAA	SPQSMGILRELDYFACISGLKINFSTKMVWI			
			CATAG	AATACTC	GSKKISKEVFHHSRWKLDWNNPTFDLLGVKFS			
			TATGC	NAACTTT	VNLNEMVELNYPSTFLEVTSLIKQWKLRLNLT			
			GCAGC	CATGGGT	VGRLTVIKTLIIPKLNHLILTIKPSVEYLQQFEK			
			ATAGG	GATAAAA	EIYIFLWGGKVHKKVKNIIQDYRYGGLKMID			
			GTTTA	GCAAGTT	YMEFVCALKSTWIRRLHSNTKWVKLFESELG			
			AGGTT	TTCGAGA	LKIETLWEKGSDFITNISNKNKFWKEVLLDW			
			TTTAA	TTGTTTGT	VKIVMSYSQNDIKVLNEHIWFNPNIKINNSSIFI			
			AAAAA	GCATATA	KNYFNCGYKFIKDLMTTEGNFVSLETLENSNV			
			CAGAC	TATGTCT	KTNFLEYRSLKGAVFERLKALNNNKVEYGPFI			
			AATAA	GTCTGTC	PNSLSIFYKQSKGCKDMYYVLIDNKKESIRSVK			
			GTTAT	GGTGTAT	KWTESGFAYTEIEWNNIFELPFKTSKESKFQWL			
			ACATA	ACACCTG	QFQILHRILPTNEYLFKLNIDSPACSFQNDIET			
			TAAAA	TCTGTAC	IEHLFVDCFHVKEIWL YFEWLRSKFNFMTF			
			CTCCT	GTGTGCT	DKKTILFGKFKNRNMYRLQNILILTVKQFIFAS			
			CACGA	TTTTTATG	KYKNVPNLSTLMLKRTVTDRIHIEKLLLLKNC			
			CCGAT	TGACTGT	KFAEYKEYWQTICDLL			
			TCTAA	GTTTATA				
			TTATTT	TTATACT				
			ATACT	TATATTC				
			ATTGT	TAAATCT				
			ACGAT	CAAATGA				

			TTTTTT	AGTGAAT				
			TTTTA	ATGTGAT				
			AGTTG	GTGTATG				
			ACAAA	CACTGTA				
			AAAAA	AACTAAA				
			GGTTG	CATATAA				
			TTTTTG	GAAGATG				
			GTCAA	TAAATAA				
			CTTCA	CTTGTAC				
			TGCAG	ATGAAAG				
			TCAAC	TTTGAAT				
			TATTTT	AAAAAAA				
			TTAGT	AAAAAA				
			GTTTG					
			TAAAA					
			TTAGC					
			AAGCG					
			TGATG					
			TGCTTT					
			CGCAA					
			TTCCA					
			AACTT					
			CTGCC					

				TTTTA ACTGT TTTTTC TCCCTT TTATTT GCCCC CTTTCT ATTCTT TTCGG GAACC GTAAA CCTTA TAAAG CATTT ATATA GAC					
RTE X	RTE X- 5_C Gi		Crassostre a gigas	AAAAA TCGAG CGGCW ATGAC AACAT CATCA ACGAG	TAAGCGT AAATCAA GTTTATA TATAACT ACTGCTT TGTTGAT CTTTTTG	KQSRVLQGDILSPMLFSLFLEDLETHLHENXSA GIEIDQINIFLLL FADDLAI ISETPNGLQKHLNTL HTYCERWGLTVNIDKTKVMIFQKGRTRKGYK FCYNNKNLEIVDSFAYLGTVLSPNGLFNKHVK ITA EKTLQLYNTFMYKTQKLSLDCKTMMELF DVACSSKMNYGAEVWGYTNAKDLERVHLKF CKRLLGVKTSACTPAVYCELGRKPLLLTRTLRI	SSF56 672	4	137

			ATGCT	ATATTTT	IKYWFKVLRSQNCLMKSIYDYDLKNMLSSWII			
			CAGTG	TCATCCA	FVKNTLLSNGFGDVWYYPESVNEKMFYVYFE			
			CAGAT	CTTTATA	QRLSDCFYQNLRVEIENAHPLRLYKYLNLPH			
			GTGTC	CCTTTTAT	QAEYLNHILYDKHRKALSRLRLSSHRLRIESGR			
			GACCC	ATATATT	YDKKCPINERKCIYCKSNDVEDEYHFVLVCF			
			TTGAA	TATATAA	YTELRKLYIKKYFYVRPSMAKFIDLLNSDSIFM			
			CCAGT	CCTGCCT	LKKLGKYVYESFNKRDNFTIRRE			
			GTAA	CCCACCT				
			ATGTG	GTACCCA				
			TTATT	CAATTAT				
			AATTC	TGTAAAG				
			GATAT	TACTGTT				
			TTGAT	AAATTGT				
			AGTGG	ATGCTAT				
			TTATT	GGCAATG				
			ATCCT	TATGTAA				
			TTAAA	TGTATAC				
			TTGGA	GCCAATA				
			AATTG	AACTACT				
			GGTAT	ACTACTA				
			AATTG	C				
			TGCCG					
			ATTTTC					

				AAAAA					
				AGGAG					
				ATAAA					
				ATGGA					
				TCCAC					
				AGAAC					
				TACAG					
				GGGCA					
				TAACC					
				CTACA					
				CTCTA					
				AGCTG					
				CCTGG					
				CGAAG					
				CTTTTC					
				TCCCG					
				TATCC					
				TAAAC					
				AATAG					
				ATTAC					
				AAAGA					
				TGGGC					
				AGAAG					

				AAAAT					
				AGTGT					
				GATAT					
				CTGAT					
				GCGCA					
				ATATG					
				GGTTA					
				AAGAA					
				AGGTT					
				GTAGT					
				ACTAT					
				TGATG					
				CATTA					
				TATGT					
				TTTAA					
				AAAGT					
				ATCGT					
				TGACT					
				TTAAA					
				ATTGC					
				AAATG					
				GCAGT					
				AAGCT					

				CTTTTG					
				TACAT					
				TTGTA					
				GATTT					
				TCGGA					
				GAGCG					
				TTTGA					
				CACGG					
				TTGAT					
				AGAAA					
				TGCAT					
				TGTAT					
				TATAA					
				AATGT					
				TTAAG					
				TTAGG					
				TATAG					
				AAGGT					
				AAAAT					
				GTTAC					
				GCATA					
				ATCAA					
				ATCTA					

				TGTAT GATGA TGTGA AAAGT TGTGT ACGAG TCAAT GGTAA TCTTA CAGAT TTTTTT TTGA					
RTE X	RTE X- 2_C Gi		Crassostre a gigas	GAATC ATGGA AAAGG ATCTG GTGTC AATAA AAGAA GCTAA TGAAA GTCTA CAAAA	TAGTTGT TTTCCTAT TATGATC CTTTATA CTAAAAT CTAACAT AAGTAGT GATTCTA TAGTCTG TTCTTTTT TATTTAT	MKNRHKLSNRRSGGLILAFKDTIVKYIKVLET DCKFIFLFKLSKTLLNISEDAVCGLVYIPPEGSR YSSPDCFSEIEQEIINVESNCKYLCLFGDFNARV GKMQDYYPDDFLAHVTHCEIQSEVATRSIIFD NYQIPLDRIVRDHTINNFGYKLVEFCKNNDIYI MNSRVGNDKNIGGTTCKDSSTVDYVLASANF FAIVESFNIKNFCDLFSDVHNPISFSFSICINDAK NNIPGKNQPSVKLWCADKADDFRNNIQMSAV CEIENELNVLSSCIDNICKDNISKIAKDICTVFV NTARETFGMSKHVDNFSNEKSSKPWFNRVCR TARRKYHLAKRLYNRNNSTDNKENLKVLSKQ	cd0165 0	506	776

			CTCTG	ATTTTTTC	YKRTMDGCIKDYKKTITKKLVNLRSTNPKDY			
			TTACA	CCTGTAT	WKILHGSRRDEKCAVSIDDMFQFLKDMNEGC			
			GACCT	GAAAATG	ENGESNIDNPFIGTCMYDDILNGEISDEEIVAA			
			AAAGG	TACCTGT	VKKLKNSKAAGYDLIVNEHISSTLSIFLPVYKK			
			CTCGT	TTGTCTG	FFNIIFCSGFVPDEWLIGIHKPIYKNKGDPTQPEN			
			TCAAT	TACCTCT	YRPISLLSCLGKLFTSILANRLEAYATETNLIYE			
			GCGCG	GACGATT	NQAGFRKNYSTLDHILTLQFLCDIFMKRKKKL			
			ACAAT	TGTATTT	FCAFIDFKQAFDTVWRTGLWSKLILNGINGKC			
			CTGAT	GTATGTA	FRYIKNMYMGIKSLVKLNGIASNFFECNVGVR			
			ATTTTC	TTATATA	QGENLSPFLFSLYINDLENFLLEKNIVGLQISISN			
			TGGAA	TATGTTC	EIENELLHLKLLILFYADDTVIMAESTEDLQN			
			TCCCG	CTCTTGT	ALNEFYSYCTQWKLNVNVNKTIMVFSKGPT			
			GAGGA	ACCAAAT	SKKLFYYKGEVIENVKEFKYLGIVLSRTGSFIK			
			GAGAC	TATTTGG	ARKHLCEQAQKAMYGVIRKIRNFNLPVNCQF			
			ATGAA	TCTGAGT	DLFDKVVLPLVLLYGCEIWGFENLDIHERVHLKF			
			GACAC	GAATAAA	LKTVLNLIRSTTPNFMIYGETGRYPLYINVYTRII			
			CGAAG	CTGAACT	TYWGKMLISPENKIVYTVYRYLFSQHFNSDCK			
			CTGTC	TGAACTT	SPWIDCIQRILNMCGLQYVWQNHNFVNIEWLS			
			CTACA	GAA	ATVNQRLKDQFMQIWSNDINCSSRGQVYKILK			
			AGAAT		PYFGFEKYLDILPVKLRKKFIKFRTSNHRFPVE			
			TTATC		TGRWYNIPLNERLCLLCNKGLIGDEFHYILECS			
			CAAAA		ALEEIRKKYINTKYWKRPNFFKFSELMTNCNC			
			AAAGT		KSLRKLCVFISKIFDAVCSS			

				ATAAA					
				CTGGA					
				TTACG					
				AAGTC					
				CCATT					
				CGAAA					
				GGGTC					
				CATAG					
				AATGG					
				GTAAA					
				TGGAA					
				GGAGT					
				ACAGT					
				TCCCA					
				TCCCA					
				GAAAT					
				ATTGT					
				CGCCA					
				AATTT					
				ACCTA					
				TTTTA					
				AGGAT					
				AGAGA					

				ATTAA					
				TCCGT					
				ACACG					
				TGCCC					
				CCCAG					
				AAGCT					
				CCACA					
				AATCC					
				AACGT					
				TTGGG					
				TAAAC					
				GAACA					
				ATTTC					
				CCCCT					
				GAAAT					
				CGAGG					
				AAAAG					
				AGAAG					
				GAAGC					
				TATAT					
				CCCGT					
				CCTCC					
				GCCAG					

				GCAAA					
				GAAAG					
				ACCAC					
				AAGAA					
				AGTAA					
				AACTG					
				GTCCG					
				GGACG					
				TGCTTT					
				ACATT					
				GACGG					
				ACAGG					
				AATAT					
				ATCCC					
				TCCTA					
				TAGGA					
				TCCGC					
				TCCGA					
				GAGCA					
				TTTCC					
				AACGC					
				CCACG					
				CCAGC					

				TCAGA					
				CGCGA					
				TCCGC					
				CAGCA					
				ATCAA					
				ACAGC					
				CGAGC					
				GACAA					
				TACAC					
				CCCGC					
				AAAAC					
				CCGAC					
				CAAAG					
				ACGAC					
				CGACA					
				GGCTT					
				TTAAA					
				CGACA					
				ACGTC					
				AAGGA					
				TCAAC					
				CCCCG					
				ACCGA					

				ATTGG					
				ATAGG					
				GAGAC					
				GCAGG					
				AGAAC					
				AACTC					
				TCGCG					
				AACAA					
				ACAGC					
				TACAC					
				GTCTC					
				TGAAA					
				TTTCTC					
				TCGCT					
				TAACG					
				TATGC					
				GGTGT					
				CAAAC					
				AAAAA					
				CTAAA					
				TTACC					
				CCGAA					
				TTTTGT					

				GATTT TATTG CAAAC TATGA TATCA TTGGA CTATC TGAAA CAAAA ACTGA TGATT CGGAC AATAT TCAAA TACCC GGTTA CGTCA CTTAC					
CR1	CR1 - 6_C Gi		Crassostrea gigas	CCAAA TTCCA ACAAC AGCAC AGACA	TGATCTT ATAACTA CTAGTAC ACCGTTG TTTACTA	MAGDTIEQYRAAIGLFYCCVRCVKKCFYLNLF QLRXLFKCLLQYLLWFGNLYKTLSMCLAQHS LVIDYFFQVIFMLLFLSGDIETNPGPCTGTQYTL DIFHLNIRSIRNKLDSLLSLVSDFDVLCTESH NSSIPDRNLFIDGFNTIFRKDRNSFGGGHIVYVS	cd0165 0	595	851

			GTATT	TATACAT	NSLRVKRRTDLEPTNIECIMGNQQTLNALWVE			
			ATTAA	GTCAATT	ISDPTCNILLCCTYRPPNCDNSFWRNLSWSIEK			
			AGATA	GCATATA	AYDESDKIIIVVGDLNVDFLNIPHThLVKDLLSS			
			TTTGC	TGTATAC	HfLENTITEPTRVTQNTSTLIDPILVTRDVSFFD			
			AGAGA	GTCCATT	SGVLDVHHTISDHKATYIYVKSNLNLNSAYKR			
			AAAAC	CTATCCA	KVWLYKNADFDKLNFLITHTDWETLIMGADD			
			TGGGG	CTGACCT	VDQAAXNFSETFLSHIRKSVPEKTVTIRPKDKP			
			ATCCA	TTATTTAT	WFDSALRKEIRQRNRLRKIAMNSRNALDMQK			
			AATAA	GTTGTGT	FKKSRNKVNNMKKYAIDNYNNLEINLLDSS			
			CCGAA	ACACATA	KNNSKLYWRLLKNVFDIKMSTEIPPLQFALKN			
			AACGA	TTTTGTAT	GNQSIAYTNTEKAEVLNDYFSSISCLDDNDAR			
			TATTTT	TATATCT	LPEFNSECNNFICDVCISEQDVIDIISIIQV NKAV			
			TCGTT	TTGTATT	GPDNISHKMLKNTMYTIAKPLCSLFNRSLRDC			
			CCCAT	GGGAGAG	TFPSSWKISNVLPYKKGDSSQMSNYRPVSLLS			
			CCCAT	GGCGGTC	CVSKIMERIIFRHVYNFFHDNNLFYKYQAGFLP			
			TGGAA	TAAGTTA	GHSTVYQLIETYDHILRAIDQGQSCCMIFCDLS			
			ACAGA	TGGAAct	KAFDRVWHKGLLFLKHTYGITGNLFKWFSY			
			AACAA	TGTGCCC	LVNRKQRVMYREVLSSIKSINAGVPQGSVLGP			
			CCATG	AATCCCA	LLFLIYVNDIADNMLTFCRLFADDNSLQHAAIN			
			GGGGC	ATTGTAT	VHDMESYLN RDLLTLEVWSKKWLLNFNPSKT			
			ATCCA	TTTTGTA	KAVYFSNKLTQTNTVPTLKFQGCDFEVPTH			
			AATTA	CAATAAA	KHLGVIFSEDLSWSSYINSIIANAQKKLGLMKK			
			TTTGTC	AATATGT	LKFKLNRKSLSLLYTSFIRPQLEYASDVWGGCS			

				GTTTC	TCAAATC	INDSDKLEKIQLIAARIVTGLPIFASRESLYFETG			
				CAAAA	AA	WETLYSRRKISRLKTMYKIDRNLLPSYVMDIFP			
				TTGGA		NKRANVSNYSTRNAHNYXIPKCRLELYKSSFV			
				AAACA		PTVVDEWNALSTDVRGSDSVRIFKEKIHVSVK			
				AAGAA		NENXLTCRTRPNFFTYGDRFLNIIHTKLRHNCA			
				CGCAG		LNNDLFRNCNIINSPLCSCGKJEDTYHYFFTCNK			
				TGTAC		YKDARNVLFNDVFQIVNLNIVNTHVLLWGDSS			
				AAAGC		ISDNENKHLFTLVYNFIKNTERFN			
				AAAGA					
				GTAAA					
				CTAAA					
				AAATT					
				CCGAT					
				ATCTTT					
				GTGAC					
				CGAGG					
				ACCTC					
				ACCAG					
				ATACC					
				GACAA					
				GGAAT					
				CATAC					
				ATGAA					

				CTGTC					
				AGCTG					
				CTAAA					
				CGCTT					
				GCGTA					
				AAGTT					
				CACTC					
				TTTCTG					
				GACCA					
				ACGAC					
				GGGAG					
				AATTT					
				TCACT					
				AAACT					
				GTCCG					
				AAAAT					
				GGTCA					
				AAAAT					
				ATTTG					
				ATCAA					
				ATCCA					
				TCAGT					
				GAACT					

				GCACG					
				AAATC					
				GCGCC					
				GCTAC					
				CGAGT					
				GACGA					
				ACAAC					
				AAGGC					
				TAAAA					
				CGCAA					
				ATCCC					
				GTACT					
				CTCTTC					
				TGAAC					
				CACAA					
				ACCCA					
				TGTGA					
				GCAAT					
				GTATA					
				GTATA					
				AATTA					
				GATAC					
				AATGT					

				TCAAT					
				CAGTT					
				ATGTG					
				TGATA					
				TTCTA					
				AGTAT					
				CGTGT					
				CTTCA					
				AAAGT					
				AGATA					
				CATGT					
				ATTTG					
				CAGAT					
				GTATG					
				AAAAT					
				AATTT					
				TTAGT					
				TATAT					
				GTACA					
				CCAAT					
				CTAAT					
				AGTTG					
				TGGTT					

				AAATT					
				ACATA					
				GCTAC					
				ATTTT					
				GTTTG					
				TTATA					
				TAGTA					
				GTATT					
				ATAAC					
				TGTGT					
				AAACC					
				TATGA					
				ATTTA					
				TATTTC					
				CCATT					
				TTTTCT					
				CAAAT					
				GCAAT					
				CAATT					
				TGGTT					
				TTTAC					
				CAATA					
				TTGTA					

				CTTGA					
				TGCCA					
				TAATG					
				TTTTTT					
				TTTTCT					
				ACATT					
				TTCTTG					
				TTTTTG					
				TTCATT					
				TGTTC					
				CTTTCT					
				TGTTTT					
				TCACT					
				TTGCA					
				TTCTTT					
				ATTTG					
				TTTGC					
				ATACA					
				CTTTTT					
				CAAGT					
				GTTTG					
				CTTAC					
				TGTGT					

				TTTTAC					
				CTTCA					
				ACTTTT					
				GTCTT					
				GATGT					
				TACGC					
				AACGA					
				AGGTC					
				ATGGT					
				CGTTC					
				AGTAC					
				ATGTA					
				TGAAT					
				TCACA					
				CCTGT					
				CTTTTC					
				AATCG					
				CAGTG					
				CTCTCT					
				CATAG					
				TTTCA					
				ATAAC					
				GTTAA					

				TCAAC ATGTC TTTCA					
L2B	L2B - 1B_ CQ	.	Culex quinquefas ciatus	G	TAGAAGT GACCCAC GATGGTA CTGTGAC GAAGAGC ACGTTAT GACGGTC GGCCATC TTCATTA TCGGTAC ACATCGC ATGGGAT CACTTTG GGCCGCA TATGATA AATTTAT TCAAAAG TAACGCG AATCTGG GCGCGGT	CESALNLLLLKWKQCIEKGNIVLSVFVDLKRA FETIDRTKLIRVLKKSGIRGTVLKWFSSYLENR KQVTRYND AVSAETAVDLGVPQGSVLGPLFI LYINDIKQALKRVQVNLFADDTVLFVVGDSFD ECFDIMNEELGGFSEWLKWKKLQLNITKTKC MLVTTRQTKDCRRSVQMDGGIVERVETMKYL GVMLDEKLNFNHIDYTIRKAARKFGVLCRIN RYLTAETKVQLYNSLIAPHFDYCSSILFLATQR QMKRMQVLQSKIMRLILRCDRLTPRRSMLECL QWMSVKQRIEYNTLVFVFRVKERMAPQYLTE AMVRGRDIHEHDTRGADDLRLQIWRKACTQN SLFYKGFKLYNQLPEAAKMTSNINEFKRSCKE FVRQRPLE	cd0165 0	1	195

					TTAACCC				
					TATGGGC				
					TCATATG				
					TGCGTGG				
					AATAGCA				
					AATGGTT				
					CCAATAC				
					CCTAAAT				
					GGATAAA				
					AATACTG				
					CAATCTG				
					ATTGATA				
					AAATGAC				
					TCGACAC				
					TGGATTT				
					GTTAGAG				
					CGCCTTG				
					AGATAGG				
					ACACGAG				
					AGAAATG				
					GATGGGC				
					ATACACG				
					GAAGTAG				

					AGTCAAG ATTACAG GACACTC TCAAAAC ACGAACG ACGAATT ATCGAAA GATATCT GCTCGTA AACCTTC CATGCTA CAAAAAC TGTGTAT GGGTAAG AGGTGGG CCATCCA AGGAAAA AAAAAA				
CR1	CR1 - 10_ CQ	.	Culex quinquefas ciatus	TC	TAAATTT AGTTTTA AGTKCAT CATTGGG GCAAAAT	RRADFESLSRALRGIEWSSVLDPVDIDLAVDTY TXILGRLIEQFVPKIRKVSSHLPWQTQELRXL KTQKRAAFKVSNGSVHLRDYYLRINHSYQRLS RRCFANYQRKMQKKLKTDPKKFWKFVNESRR ESGLPTSMRLADXEGDDADKICQLFAKKFAGV	cd0165 0	250	500

					AAGTGCC TGTTGAG AAGAAAT AAACAAC	FCAEPITDEEVRAAAENVPRRDELLASIEIDEGT IQAAAANKLKHSCSPGPDGIPSTLLKRCSSSLSAP LLHLFRLSLSSGKFPVAWKQAYMFPVHKKGD RTNVENYRGISALCATSKLLELVVIDPVFVSHCR QDLADEQHGFFPKRSTATNLLCFTEIINSFDA HSQTDAIYTDLSAAFDKINHRIA VAKLERLGFR GSLLAWFKSYLTGRSLRVRIGDALSDIIDATSGI AQGSHLGPVVFLFYFNDVNYTVTGPRLVYAD DLKLFARIDNSNDAEALQRDLDRFAGWCXVN RMILNPGKCQXITFSRRHFPVTFDYRINDTSVE RVTHVKDLGVILDTQLTFKQHLSYIIAKASRQL GLVIRMTRKFTDIYCLKTLYCSLVRSTLEYCST VWSPHYNNAVCRLESIQRRFVRYALRQLPWT NPLXLPPYEQRCRLIDLDTLQLRRDLARAMTS ADVLLGRIDCPDLRNQIQLRVPTRQLRHTPALE VPFRRTNYSANGAINGLKRAFNVSSVFSLDV SRTTLRFKFVSLLRSLLRM			
CR1	CR1 - 25_ CQ	.	Culex quinquefas ciatus	CACGT GGCCT CGCGG TCTTGT GATCC	TAATGTT AGTTTTTT AAGATTA TACAGCC TGAGCAT TTTAGAT	RALRRIFLGADKRXQPDSAXSASSSCPTCTNKT VNAESHAKVNAETNTKVNVEVNPEIYAGHGA DPKGYGTNFACESQWFTYGSRDRLTNTLRSTP SLPQQSSVQHLPASSPGRAVESIMRAPRLPDAT APSPSYRTRSPASRSHQRRPGAGVGVGGGVVSQ TAPLGKYPSQLQVSLPDACPTYSSSLITSSSTRT	cd0165 0	933	1184

				GAGAA TTTGA	GGAGGCG GTGAAAT AAATAAA TA	ANASPPAQAIQSSPGRTVESIMRAPRLPDATTS SFSYRTRSPAGCSHQSRPGAGVGVGGGVSQPV LFGKYPFQSQRVLPDASPTHRS CPSTSTSTRTA NASPLPAQAIQSSPGRTVESIMRAPRLPDATTSS FSYRTRSPAGCSHQSRPGAGVGVGGGVSQPVL FGKYQFSTPHTSPDARSDFSATVPPSTFVTLAA PGCVYHGTMEDLGSANSGMPAALTASDPASA SSERGFLNVYYQNVRGLRKTNSVLLALSSCD YDVIIFTETWLVPGIKNSEFAAGYDIFRCDRSA ETSSLQRGGGVLVAVKTGRNCKDVELVCGGS LEQIAVRIVLPHQALFICCIYIRPASNIDVYQLH TGAVQQILGMAGVRDLVLVVG DYNLPALTWF FDD DTHSFLPINASAEQEVLLLESMLSCG LLQI NNLASAHGNLLDLAFVSDATAAELIDPPCPLL KPD AHHKPILLKLDYDHGGNDDLDDADEIFDF NQCDYALLNERISAVDWD TLLDTSSVDESTAN FYERLNSILHEVVPKKIRRM SHRYKLPWWNDE LRTL RNRLRKATSRYMESKTAWDKVTVQNIE DEYNALNQESFRSYINNTQ RSLKSDPSKFSY VNSRK NANRTPTDVSYRDLNSTSPAIAANFFA DFFKSVHSSDAIPPRDHEFEYVPTHYLP LPLPV LNDEETLKALSGVDGAKGPGPDGIPPSLIKACA DSLTVPVKRIFNDSL RKG VYPAMWKVASITPI			
--	--	--	--	----------------	-------------------------------------	--	--	--	--

						HKSGSTKKVENYRPISILNCLAKVLEKIVYDRL YAVVRPIISDDQHGFMQKRSTTTNLLTFVHST VNALESGEQVDAIYIDFEKAFDKVPHALTVRK LRKLGLPAWITNWLLSYLTNRKAFVNLRIARS STFDIPSGVPQGSHLGPLIFILFVNEICALTNSNT LMYADDLKLFRSVKNNVDCCSLQMDVNALL HWCESNGMKVHVKKCKVISFSRSRNPLRFDY TMGDGCPLDRVNSIRDLGVTVDKMKFNEHI ALTTGKAFAMLGFLKRNTAHFDDPFALKVLY ASLVRSVLEYAAVWWAPYHVTLAARIERVQR AFVRFALRKLPWNNPLILPSYEGRCLLFRLQSL AERRTLMQRLFVFDLLRGNIDCRSIRNNVRFN EPARQGLRGERPLLWIPRHIFDYGYNNPLDVC FRQFNEICSCFDYVVSKTVFKTRIS			
CR1	CR1 - 9_C Q	.	Culex quinquefas ciatus	C	TAATCTG TTTTTGTG TTTCATG TTGTTAG CATTTGT CTTTAGT TAAGTTA AGTTTTA TCATTAA	GESESRGREHGRVTRPPLAASPGRAVESHVRV PRLPDTTTSSPSHRLIRPACCSLQSRPGAGVGA GGGVSQTATLGKYXSXSQXPLPDARPTSSSCAI SSTSPPVRTPNAVRAFWSSPGRAVESHVRVPH LPDTTTSSPSHRLISPACCSLQSRPGAGVGAGG GVSQTATLGKYHSSLQNPLPDARPTSSSCATSP PVRTPNAVRAFRSSPGRAVESHVRVPRLPDTT TSSPSHRLISSACCSLQSRPGAGVGAGGGVSQT ATLGKYHSSLQNPLPDARPTSSSCATSPPV RTP	cd0165 0	910	1157

					GACCACA AGTTGTC GGATGAT TCAAATA AACAAAT AAACAAA TAAACAA AAAAAAA A	NAVRAFWSSPGRAVESHVVRVRLPDTTTSSPS HRPISPXCCSLQSRPGAGVGAGGGVSQTATLG KYPSGLQNPLSDARPTSSSAAFPSTINAKGTSTP CKLRIYYQNVRLRTKIDAFFLAVTECDYDVI VLTETWLDIEILSPQLFGSTYRVFRNDRNERN DKKSGGGVLIAISSRLDCVLDPTPVCDTLEQL WVRLTVGASQRCIGAFYLPPGLNRSIDLIKDH LESIGNVLSHLNLNDSFIQFGDYNQPGLKWIPS EEGGLEIDPVQSRVPLVSSYLIDGFNLHGLSQL NPCVNVYNNCLDLVLVNEVALPNCAVLAAVD PLLPQDIAHPALEMFVDVPAPSATAEPPLNSQQ SFNFRKANYAAIQEELSRIDWSFLETAPSIDAA VDAFQSILNNLFTEHVPLRRPPQKPPWSNPTLR KLNRLKTAALQEFSNHRSPFFKSKFNLASKRY RGYNRTL YRRYVIRTQRDLRRNPKNFWSFVNS KRKADGLPPTLYLDDVAAATSEEKCALFAAHF KDAFN SRFASEAEIIEATRNTPIN AIECSL TEVA NCQVSAAINKLKLSYSPGPDGIPSSVIKRCSGA LLSPLTCLFNLSLRQSTFPDKWKASFMPVHK KGDKSSVKNYRGITSLSACSKVFEILVNNVLF SCKNYISTDQHGFYPKRSVVTNLAQFVSHCLQ EMDAGAQVDAVYTDLKAAFDRVDHGVLLKR LERLGVSASLVRWLKSYLTNRILRVKIGNSESD			
--	--	--	--	--	---	--	--	--	--

						MFISLSGVPQGSNLGPLLFILFANELSMLLSEGC RLFYADDVKIFTVVRSVLNCSALQRQLNHFNE WCVRNFLTISIEKCSSTISFHRKLKPVVFDYAIAG NTLARVTQVRDLGVTLDSELCFTIHLHDIVSRA NRQLGFIFRIADDFRDVACLRSLYCALVRSILE FCSVWCPYQSTWTAKIESVQRKFTRLALRRRA RNPAGWNTYEERCRLVLQLETLERRRHVSQAT FVAKVLAGEVDTPWIRAQIQIYTPARPLRQRQ YLQLAHRDTNYGQHEPVRFMCSRNFVFSHLY DPNISTPQFQQRARLWYSHQL			
CR1	CR1 - 5B_ CQ	.	Culex quinquefas ciatus	ATACT ACCAG AACGT CGGAG GC	TAATTTA TGTATTA GTTTGTA AGGTAAA ACACCAT TGGGGCC CGAGCCT GTTGGTG TCCAAAT AAATAAA TA	MNTRALEYRLACSAISYDVIAVTETWLKNNTL SSQVFGNGYEVFRCDREQFINSQKKDGGGVAI AVRQGLNARIVSSDEWKCVQVWVAIDLADH TVYICVVYIPPDQTNNASVFEINSSSVSTVSSM ARPYDELVILGDFNLPRLIWRPSRDGFLYADPE LSTQTLHVTEFLDNYSSALLQQINQFPNENNVL LDLCFVSCPDIAPIAIAPAPLVKDVRHHPPLHL SLPVHLASEFTPIVCEVRYDYRKSDVPSMLELL RSIDWENTLDKDNLDDAVQTFTNIIGYVIDRH VPKKPVTNARVPWQTSELRRVKA AKRSALKK YSKHRTLPLHNYVRLNSCYKRMSKQCHADH EVRTQKKLKS NPKGFWKYINEQRKEIGLPSSM FLGDKTASNTEDICQLFAAKFSSMFTQEDLSQ	cd0165 0	499	750

						QQITAAANNVLPNGLALNRITIDRSQILEAATK LKSSNTAGPDGVPSIVLKKCIDGLLDPLAHLFS LSLSTGSFPKLWKHAYMFPVHKKGDKRNADN YRGISALCATSKLFELAVMDPVFFFCKNLISDD QHGFMPSRSTTTNLLTFTTFVTDSTAKSQTDA VYTDLSAAFDKLNHSIAIAKLERMGIGGVLLR WFQSYLTGRKISVKIGDWLSAVFEAFSGIAQGS HLGPLVFLLYYNDCIHSISVPRLSYADDMKIYC RIDNASDAQVLQQQLINFAEWCKINRMVVNPT KCSVISFSRKRNPFIKFEYQICGTSIPRENCVKDL GVLLDTELTYKQHISYIVSKASRQLGFVFRTAK NFTDIYCLKALYCALVRSTLEYCCSVWSPYYE NSVARIESVQRRFIRYALRSLPWRDPFRLPSYE SRCELINLDTLAVRRNVCRSLLVADVLTSRVQ CPAILRGLNIFAPIRTFRNSPFLRVPIRQTNYAM YSALTGLHRAFNFVASLFDNFNLSRDVIKRFLL FFRR			
RTE	RTE - 1_C Q	.	Culex quinquefas ciatus	TAGGG CTCCTT AGGCT AGCCA CTT	TAAGTAA GTAAGT	MEPESGLRATETARRKTDWRNGTTLGMKTRT RIGTWNVLTLAQAGKLETLGREAVRLKLEILG LSEVRWPDSAHEHKMPTGQVLLYSGLRSENAQ RGFLSPNARSALITWKPINERIIVARFRTRVRN LTFVQVYAPTDAAADLQEKEEFYSQLSXVVNEI PKGDIRIYAGDFNAKIGSDNTDLERIMGPHGX	cd0165 0	554	813

					GEMSENGELFTEFCGNQDMVIGGSLFPHRSVH KVTWVSRDGKTENQIDHICISRKWRRSLLDVR NMRSADIASDHHLVVGEIRLRVARVVRQEEK VGCRFNVQRLSNPEVXRAFVGELQARALEIPD STVEEEWQFIKDAFAVTGENTLGMLRTQRKE WISDATWDKIEERKQAKAAIVSARTRATKART RQTYAELEKAVKRSCRRDKRAWTDSLADEGE KAANNGDMRLLYDISRRLCGARTNTRIPVKDR SGQLITDPADQLKRWFHFEQLLQISHPRPSPQ YHPQNVRRINRVNSSPPTLREIEEAVKGMKSG KAPGIDRISAEMLKADAHLSARMLHRLFSKIW DTATFPVDWMQGILVKVPKKGDLTNCDNWR GITLLCIALKVLCKVILCRIKSKIDESLRRQQAG FRSGRSCVDHIVTLRVILEQVNEFQDSLHLVFV DYEKAxFDRLDHENLWEALGRKGVPERIINLIK AQYEAFCRVVHNGVLSDPPIRVVAGVRQGCIL SPLLFLIVIDEVLGALDRVQNRGLLWQPIRME HLNDLDFADDIVLFSQRGNDMQNKLDDLAAC SSAAGLKVNVSCTKSMAVNTERAQGFTVAGE NVEKVDTFQYLGSQIAPDGGTKLDISTRIKKAR SAFAGLRTLWRSNQISLRTKLRIFNSNVKSVLL YGSETWCVSSQNTSKLQVFVNRCLRNILRAW WPRNWVSNAELHRRCCQRPPIHVEIRERKWRW		
--	--	--	--	--	--	--	--

						IGHTLRKDAHEICREALDWNPPQGQRGRGRPRG SWRRSLHQEISAVDGTLSWRQVKAKAENRQQ WKSFNAAALCSTGEQXA			
Pene lope	Pene lope- 4_C Q		Culex quinquefas ciatus	GAATT GTGGT TTGGT TGAGC GTTTT AGCAG AATGC ATTAC TGTGA ATACA AAGAG ACGAG TTGTTC CTTTTA ATTCC GCTAT ATATT TTAAT ATCTT GACAG	TGATGAA GTCTGCA AGTAGTA GACGAAA TACGTAT CTGTCAA GATATTA AAAATAT ATAGCGG AATTAAA AGGAACA ACTCGTC TCTTTGT ATTCACA	MHSAKTLNQTTITMATKPSEHNKFIGLKKIAG IHKDIAFLKKCVEHRVTPTSHKIRTKGHTDSKII RRVELELVKESIKKLYGRLSIKTLECYSLHLKL AKEFPIEFEAFLTKVKVAEKCESERKRKILDNK FRQLISHIAPTMPVVKPLEEFVVNRSSETFSEE QLTLLNKGLKYAVSSKPNLENIVIDIETVINAN SIDEKQTKPPPLPAQINSVRMTVSKLIREATTK HQDNAEVRVVKELKEKPVYYLKADKGNRVVI MDRDEYDKELLEKLGNREKYSLQRGYTLIDM VKKVENTIKECAGLLKSVGKTAKSLRVPNPTL PRIKALPKVHKPGNEMREIISAVDAPTSRIAKW LVEEFKSMPKPFPSRSIISQVFTKELLES GPIAE DEIMVSFDVTALFPSIPVKNAIGTLRDWLQSQY EGEPWRQKTIQYITFVKLCMEQSYFQFRDCIY QORMGASMGNPLSPFMSDVYMAALEHELQT KNLLPERWWRYVDDVFCIIKRDSLTTVLDTIN SARRSIKFTYEMEVEGKLPFLDILILREPGCPSS FSFEIYRKPTNTRRTIPATSNHSFQHKMAAYHY FIHRMTTLPLSEAGKQKELEYIFETAEINGYSRT TIQAIIDKKERKLHRTSLTTLTPSAEPLRRAAVE	cd0030 4	469	548

			ATACG	FDYRITRPLRQKLAKFGIDLVFSSRNSQLESILG		
			TATTTT	STKDPIPTLGKPGIYEVSCGHCDMSYIGQTKRL		
			GTCTA	LQTRLDEHIHTYVKIAMEEKRKGFVPHFKSAV		
			CTACT	TEHIIEEKHNITASDAVILRHITNPSKLAVAESL		
			TGCAG	EIFKAGNKRLLNKDSGNGSTWLFKLLPIQAQK		
			ACTTC	KNEEVVPTVTTNTDTPAMNLQSRTEH		
			ATCAG			
			TGTTCT			
			GTTCT			
			CGACT			
			GAAGG			
			TTCAT			
			CGCTG			
			GTGTA			
			TCCGT			
			ATTTG			
			TAGTT			
			ACTGT			
			AGGTA			
			CTACC			
			TCCTC			
			GTTCTT			
			CTTTTG			

				TGCCT					
				GTATA					
				GGTAA					
				CAGCT					
				TAAAC					
				AGCCA					
				CGTTG					
				AGCCG					
				TTACC					
				TGAAT					
				CCTTG					
				TTGAG					
				TAAAC					
				GTTTG					
				TTGCC					
				TGCCT					
				TGAAG					
				ATTTC					
				CAAAC					
				TTTCG					
				GCGAC					
				AGCCA					
				GCTTC					

				GATGG					
				GTTAG					
				TGATG					
				TGTCTT					
				AAGAT					
				GACCG					
				CGTCG					
				CTAGC					
				TGTGA					
				TGTTA					
				TGTTTT					
				TCCTC					
				GATGA					
				TGTGT					
				TCGGT					
				TACTG					
				CTGAC					
				TTGAA					
				ATGGG					
				GGACA					
				AATCC					
				TTTCC					
				GTTTTT					

				CCTCC					
				ATGGC					
				AATTT					
				TGACA					
				TATGT					
				GTGTA					
				TATGC					
				TCATC					
				CAACC					
				TGGTT					
				TGCAG					
				CAATC					
				TCTTA					
				GTTTG					
				TCCAA					
				TGTAG					
				CTCAT					
				ATCAC					
				AGTGG					
				CCGCA					
				GGATA					
				CCTCG					
				TAGAT					

				GCCGG					
				GTTTG					
				CCTAA					
				AGTCG					
				GTATG					
				GGGTC					
				TTTCGT					
				TGAAC					
				CTAAG					
				ATGGA					
				TTCCA					
				ACTGG					
				CTGTTT					
				CTACT					
				GCTGA					
				ACACT					
				AAATC					
				GATGC					
				CAAAT					
				TTAGC					
				CAGTT					
				TTTGTC					
				GTAAT					

				GGGCG					
				TGTTA					
				TCCTG					
				TAGTC					
				GAATT					
				CTACT					
				GCTGC					
				TCTTC					
				GCAGT					
				GGCTC					
				TGCTG					
				ATGGC					
				GTAAG					
				GGTTG					
				TGAGT					
				GAAGT					
				TCTGT					
				GCAGT					
				TTTCTT					
				TCCTTC					
				TTATC					
				GATGA					
				TTGCTT					

				GGATA					
				GTTGT					
				TCTGC					
				TGTAA					
				CCGTT					
				GATTT					
				CCGCT					
				GTTTC					
				AAATA					
				TGTAC					
				TCCAA					
				TTCTTT					
				CTGTTT					
				TCCTG					
				CTTCG					
				CTGAG					
				GGGTA					
				AAGTT					
				GTCAT					
				CCTAT					
				GTATA					
				AAATA					
				ATGAT					

				ATGCT					
				GCCAT					
				CTTAT					
				GTTGG					
				AATGA					
				ATGGT					
				TTGAC					
				GTGGC					
				TGGGA					
				TTGTTC					
				TTCTG					
				GTGTT					
				TGTTG					
				GCTTC					
				CTGTA					
				GATCT					
				CGAAA					
				GAAAA					
				CGAAG					
				ATGGG					
				CAACC					
				TGGTT					
				CTCTT					

				AGGAT					
				CAGGA					
				TGTCT					
				AAGAA					
				AGGTA					
				ACTTT					
				CCTTC					
				GACTT					
				CCATT					
				TCGTA					
				AGTGA					
				ACTTG					
				ATGCT					
				CCTGC					
				GTGCA					
				CTGTTT					
				ATGGT					
				GTCCA					
				GTACC					
				GTTGT					
				TAGTG					
				AGTCC					
				CTTTTG					

				ATGAT					
				GCAGA					
				AGACG					
				TCATC					
				TACGT					
				ATCTC					
				CACCA					
				ACGTT					
				CTGGT					
				AGTAG					
				GTTTTT					
				GGTTT					
				GTAGT					
				TCGTG					
				TTCCA					
				AAGCT					
				GCCAT					
				ATAAA					
				CATCG					
				CTCAT					
				GAACG					
				GTGAC					
				AGTGG					

				GTTAC					
				CCATC					
				GATGC					
				TCCCA					
				TCCTCT					
				GTTGG					
				TAGAT					
				GCAAT					
				CCCTG					
				AACTG					
				GAAGT					
				AGCTC					
				TGTTC					
				CATGC					
				ATAAT					
				TTGAC					
				GAATG					
				TTATA					
				TACTG					
				GATCG					
				TTTTTT					
				GTCTC					
				CATGG					

				TTCTCC					
				CTCGT					
				ATTGT					
				GATTG					
				CAGCC					
				AATCC					
				CGTAG					
				AGTTC					
				CTATT					
				GCGTT					
				TTTTAC					
				TGGTA					
				TGCTT					
				GGGAA					
				CAGTG					
				CTGTT					
				ACGTC					
				GAAGG					
				AAACC					
				ATTAT					
				TTCAT					
				CCTCA					
				GCTAT					

				CGGTC				
				CCGAC				
				TCTAA				
				GAGCT				
				CCTTT				
				GTGAA				
				CACCT				
				GCGAA				
				CTGAT				
				GATCG				
				AACGG				
				CTCGG				
				GATCA				
				CTAAC				
				CCATC				
				GAAGC				
				TGGCT				
				GTCGC				
				CGAAA				
				GTTTG				
				GAAAT				
				CTTCA				
				AGGCA				

				GGCAA					
				CAAAC					
				GTTTA					
				CTCAA					
				CAAGG					
				ATTCA					
				GGTAA					
				CGGCT					
				CAACG					
				TGGCT					
				GTTTA					
				AGCTG					
				TTACC					
				TATAC					
				AGGCA					
				CAAAA					
				GAAGA					
				ACGAG					
				GAGGT					
				AGTAC					
				CTACA					
				GTAAC					
				TACAA					

				ATACG					
				GATAC					
				ACCAG					
				CGATG					
				AACCT					
				TCAGT					
				CGAGA					
				ACAGA					
				ACACT					
				GATGA					
				AGTCT					
				GCAAG					
				TAGTA					
				GACGA					
				AATAC					
				GTATC					
				TGTCA					
				AGATA					
				TTAAA					
				ATATA					
				TAGCG					
				GAATT					
				AAAAG					

				GAACA ACTCG TCTCTT TGTAT TCACA GTA					
CR1	CR1 - 27_ CQ		Culex quinquefas ciatus	GGATA CCATC GTAGT TGTAG GCGAT TACAA TCTGC CTCAC CTCAA CTGGT CTTTC GATTC GGACA TCAAC GGATA CCTGC CAACA	TAAGTTG TAGTCTG TGCGGCG AGTGCCG AAGATGA AATAAAT AAATAAA TAAAA	ARNCFRRVYACNXDTANLQPAELQPKTPGPA VTPVPESTESLTYDFTRCDFEAVNJALSQINWT GILAHCSLDEAVSRFYDALXDVIQLHTPLRRLH PPSNRKPPWWNATLRNLRNRLRKARTRLHKRR CQPNIAEVRELEELYESQQNAAFSEYIHRIETD VKQNPRSFWTFLKSRKKGSGIPTEISNGCVEAN SVEEQASLFADFFSGVYSTSLPAPCNEVVDHIP THDLHLPVVHFSIEDVSRALDDVDVTKGPGPD SIHPSFLKNCAASLALPTTIIFNRS LAESTFPTA WKLASITPIHKAGNVHKVENYRGISILSGFAKL FEHLVHGALYPVLKPIIANEQHGFVKRRSTTSN LLLFTSTLLTNIEKGQQIDAVYVDLSKAFDKVP HALLLEKLRRYGLPEWIVRWIHSYLADRKAFFV KVRCTSSHLLAIPSGVPQGSVIGPLLFI LFVNDL CSVLESDKLMYADDLKVFRSVSSPLDSCALQQ DIDRLLRWCTANGMEVNVQKCKIISFTRKRTP LLSSYRMGQNELERVNTIKDLGVTIDSKACFN	cd0165 0	297	547

				AACGC ATCTG CTGA		EHIALTTAKAFATLGFLRRNAADFVDLFALKTI YCSLVRSQLEYAVQVWAPYYAVQAERIERIQR AFTRFAVRRLPWVRPQGLSSYDDNCERLKLPA LASRRVLLQRIYVFDLLSCNIDCPTLRDKITLHI PARNLRNPAFLEVPGHRTNYGYYSPLSACCR VFNDVADVFFVGMSKCVFKTRIKNRT			
CR1	CR1 - 23_ CQ	.	Culex quinquefas ciatus	A	TGACAAA GGTTTAG ACCGGTT TATTTTAA GTCAGTT TTATTTAT TTATTTAT TTATTTTT TACATTG TGCTAGT TTAGTTT GWTAGCT TAAGTTT GATTGTA TATATCC TCTATGT CATTTGG	ALKKLGSSKRAALKKFSKCPISVNRLNLNQISN DYRRLNKSFLAHQSRVQNRLKQNPKGFWKF VNEQRKESGLPSTMIRDGVEVSTSSEICDAFRV QFSSVFADNNLSDPDVAAAADSVPDSGLVCSD LQVSREDFNAACSKLKSSTSPGPDGVPAlVIKK CANSLSEPLLRIFNLSLSLGVFPNNWKESYVFP VFKKGCKQDVKNYRGIAALCAVSKLFEVIVLE FLRHNLSHVVSDDQHGFPPKRSTNTNLVTYIS KIQRAISDGYQVDAIYTDLSAAFDKINHRIAIA KLRRVGLHGSLLTWVNSYLTGRTMLVKISDN LSLPFSVTSGVPQGSHLGPFLFLLYINDVNFLLR CPKLCYADDFKLYAIIKRPTDMEDLQAQVDLF ADWCRLNQMVNLNPSKCSTITFTRKRCPTFFSY KILGNPLTRDSFVKDLGVLLDSKLSFKNHISYL CSKASKQLGMIFRMTKYFRDVNCLKILYLSLV RSTLEYGSVVWAPYYHNDIGRIEAIQRKFTRFA LRHLGECPNYETRCMSMLHLDPLSVRREAAKAF	cd0165 0	190	441

					ATTGTAA AATCTGT TGATAAT AAAAAAG AGGTTTT GTGCCTA TTTGAGA AGGACCC ATCGGTA GGCGTTC CACTCAA ACGGGCT TTTCCCTC ATATAAA ACATAAA ACA	FVADLFRSNINCPFLTSELNVNIRRRVLRSHSFL VIPGARTNYAFNEPVSSMCRVFERCYPFFDFHL SRPRLKHDIXNFLREAYVF			
Kiri	Kiri-5_CQ		Culex quinquefasciatus	TTGTTT GGATT TAGAT GATAG TAGAT TTGCC CAGCA	TGAGTTT TGTTTTG GTTTTTG ATTAAGC AATGCCT TGTATTA CACACTC	MTNTVNNVQELRFEPINATLKNLNSKENGFKV GHLNTCSIICHDDVKNIFTRTGFDICIGFTETW LNSDIKDALIKFDNFEILRHDRNSTTRVXGGGV ALFVRPEYKTKIIKTSPRDSIEYMFCELKVVS KLWL VLSINPQPIKSSQLQVVVGVIYKPPTLXD FSFFFEDLGNISSTYNNIILMGDFNINILVSNPMS NNFTRLINSFSLSIVNKSFPPTHFHNTGSSLLDYFI	cd01650	519	778

			GTACA	TTTTATA	CSEN LXSTLFSQVSVPAVSHHDLIYTVFNLPVP			
			GGATT	ACTACAT	VASLKTISFRNFNRIDKEAVRNDFPNFDQLN			
			TTTTGT	ACACAGC	GIVDCETLASNFNHKLIQLVNHAPLTKTILP			
			TGTTA	TGATAAT	TSCPWYSREAYNACVDRDLAYRFWKNNRTLE			
			TTGTT	CAGTCTT	NKLIYNRFRNRATNLIRIAKKQYSAKLFDPTLP			
			ATTTT	AAATTTA	SKTIWKNFKSLGFRDDPDNLDDKFKPEDVNNC			
			GTTTG	TTACTGA	FIQKIENYPTGNRQPTIIPDIVEEDQFKFRAVTE			
			TTTAA	ATTCTAC	DEVHKAILRIKSDAIGYDEIPLKFVKLFLDLLVP			
			GSAGT	TTAAACA	YITLLYNKIISTNTFPTIYKLAIVKPIPKIKNPSNP			
			TTATCT	AGTAATT	NDLRPISILPSLSKGCEILLKDQIVQYVTFKKLIT			
			GAAAA	TGTCGTA	PFQSGFREHHSTSTIITKVCQDINASFDSNYVTV			
			CCAAA	TTATCGA	LILLDFTKAFDLIDHRIFIDKLKEQFQFSSDACD			
			ACAGC	CATTTTT	LMKSYLSDRSQLVQINGVRSQFRSTKRGVPQG			
			ACTAT	ATATCTT	SVFGPLAYTLSVNDLPSFITHCNKQSFADDTQI			
			TATCTT	TAGCTTA	YLSCPVDQLPNKISLINSDLAISRYCKDNGLV			
			TTGTTT	TTTCGTTT	LNTSKTKAIXFSRQNTVLPPLPIQIDNENVEIV			
			TGCCC	TAGTTTTT	DECLNLGIKMKNHLLTWXDHVSLIKKKVYSTL			
			TGATT	AATTTAG	RTLNCHRHTLSIDTKIKLVHTLLMPHFYGNVI			
			GCTCT	TTGTTAG	FSKMNV DTERHLQVCFN SCIRFIYNLRRFDHVS			
			TTTCCT	TTTTTAGT	QFQRSIFDLSLSKFYQFQLLLFLFKLINFKEPQY			
			ACGCG	TTTAATG	LFESLVFSRSNRNTNNIIPVHHTNFMGNSFAVR			
			TGTTTT	AAACATA	GARLWNDLPHNLKSITLFEFKRRCKEHLKNSI			
			TCTCT	AGGTCCT				

				ACATA GATTT ATATA TTTTT	TATCATT TAAGTCA ATGTATA CAAAAAA GGTTGTA CCTTCGG TATATTG TTGAAAA ATAAAAT ACAAATA CAAATAC AAATACA AT				
Tx1	Tx1- 11_ CQ	.	Culex quinquefas ciatus	TG	TAGAAAC ATCCAAA GTTAAAT TATCTCT ACCGAGC TTATTTGT CGTCTTA TTAGCAA GTGGTGA AATGTAA	LYENIFGCLIANNFVKVVAYADDLNILVRNNH EFDTVLQLVSYFSIYSKIKLNFAKSQYLRFGC LSGPHQIKEVESLNVLGVLICQNFNKLVEINYD SYLNKLKYTLSLHQKRILNLFQKAWLLNTYVL SKIWYICQIFPPNNKHLAAIKSICGQFIWQGFIF KVPRNELYLPVIKGGGLALTDVESKAKSLFIKNI LYSNADGNAPLDNFMLGQINNSTLTRNMREW IVKADELKLESHLNSSKKIYDFFIHSLNIKCKT MEDMPNLMWSNLLNNLDSNFLNSKCKTVLFC IFKDIVLNKKKLHLYGIRGTDSKFCENCGLVES	IPR000 477	11	83

					ATTCAAG AAAAATG TTGTAAA ATAAAGT GAAAAAA AAA	VSHLVKYCQTSEEIWRWLKNILNKRLKILVKD PDELLSLSIGDRDFKYKAALWLTVQTIVYNLE KRGFGDLNEFKQGIRIVRFNNKAIFERHFHFL NIF			
CR1	CR1 - 19_ CQ	.	Culex quinquefas ciatus	A	TAGGAAA CAGTCTG TTCAAGA CGAAGAT AAATAAA TAAATAA ATA	IXRTQRRXFLGXFKCCREHSGKSATPPVPTLND ADDAEFAARVCTAIDSKDTGACFGVDWRSGR FTHGSSDRITCGTIVPTIPASFRSSSQHLPVRVL QSSPGRAVESLMRAPRLPDATTSSPRQRLRTPA SRSHQSRPGAGVGVGGGVSQPAPFGKYPSRSQ HFLPDVRPTSSSCASTSATQTAIVSPPSFRAIQF SPGRAVESPMGAPRLPDATTSSPRQRLRTPASR SHQSRPGAGVGVGGGVSQPAPFGKYPSRSQHFL LPDVRPTSSSCASTSATQTAIVSPPSFRAIQFSP GRAVESPMGAPRLPDATTSSPRQRLRTPASRS HQSRPGAGVGVGGGVSQPAPFGKYPSRSQHFL PDVRPTSSSCASTSATQTAIVSPPSFRAIQFSPG RAVESPMGAPRLPDATTSSPRQRLRTPASRSH QSRPGAGVGVGGGVSQPAPFGKYPPRSQHFLP DVRPTSSSCAGPSATQIANASTLSATVPSQLAN DVQPQRPGSVDRRDSSLLTIYYQNVRGLRKT NQLHLSLSSCDYDVIAFSETWLNKIVDSELSS	cd0165 0	1011	1261

					DYTFYRADRSRETSVLERGGGVLLIGVKKRFHP EPIHLRGGEKLEQIAVRIPLPGKTLFVCCVYIPP NSDATIYTQHSSCVQQLCDLAQGPDEVIVVGD YNVPNLTWHFDEDINGYLPSGQSTEQEVALIE DVLGTGLQQMFDLPNANGRLDLALVSDSSRF QLIEPPRAILKVDAHHRPFILVLETQSTGTAAE QSGTYYDFANCDTMALNNSIASLDWANLLL GSVDDIVERFYDKLNEIIREVVPLKHRRPHPNS HQPWWNRETRNLRNRLRKARKRFLKHPTEML GRELADRELHYEMSLEAAFRGYISGIESNLKRE PKSFWAFVKKRKQDKGIPQDMVYRDSSASTP DDSVQLFADFFKTVYSANQPTTSNESLDKILPH ELHMPLVCLTEEDVRKVLASTDASKGPGDCI PPSIVKQCAATLANPLAIIFNRSLESGVFPAKW KVASITPIHKSGNVHNVENYRSISILSCLPKVFE KIVHEFLSRAVQPIISACQHGFMAKRSTTTNLM TFVTDVLRSMESRCQVDAVYVDFAKAFDRVP HQLAVEKLRRLGLPEWITTWLHSYLTSR SAYV KLGGSKSGLFMIPSGVPQGSHLGPLIFLLFIDDL CHWIKNDKVL YADDLKFYRTITSPLDCLALQS DIDSLMRWCDQNGMEANAKKCQVITFDRKLR PITHAYMMGISLLDRVCSVKDLGVTLDRKMTF NEHITSVTAKAFATIGFIRRNTTAFKDIHALKA			
--	--	--	--	--	--	--	--	--

						LYSALVRSQLEYAAQVWAPYHNVHIARLERV QRAFVRLALRTL PWRNPPEQTSIEDRRRLGL DTLKRRQRMQQLFIFDLLTKRLDCQKLSQLLI VYVPPRNLRNPSMFRLSVHHTVYGHNQPF VCCRLFNVVSDRFVLNMTKDSFKKLISC			
RTE	RTE - 2_C Q		Culex quinquefas ciatus	GTGTG TGTGC GTCGC GCGTT ATTAT CGCGC CTCGG TGCGG CTCCG CGAAA GTGCC CTCAG AACCG CAGAA TTACC GTTCC GTCGT GGCCG	TGATTGG TATGGTA TGGTATG	MARRLEIRNKRQKPIEKQTTDWRLGSWNCRL NFLGFEFALAXELQPRNFDVVALQEVCLLEE VLDWPGRQTNSRFXLSGGKDKKLGTGFIVRG KMQDRVFGFTAXSERMCKLRIRGRFFNYSIIN VHCPHEEKTDDEKEAFYATLEEVDGCPRQD VKIIIGDMNARFGREEMFRPTIGPESLHAVTND NGQRCIDFAASRGMVVRSTYFPRKDIHKATWT SPDQRTKTQIDHVLIDGRFFSDVTHVRTFRGAN IDSDHYLVGVDMRSKLSTVFNQRRSRRAPPFN TACLQNGEVAHSYAQQLEANLPGEEELGAAS LEDGWSRIRSAIGSAAEATLGSAIRVSRNDWY DDECQRITAEKKAAYDRKLHKATRGNVERYR QARNRQVAVFKLKKRQQEDRDCAEMEQLFR ANETRKFYEKVNRSRKGFVPRADVCRDNGGN LIVNKSEVLDRWKQYFNEHLNGDEADGDGVG VNLGAPAADEQFPAPDLETVKREIRKLKNNRA AGKDRLPGELFKYGGEKLARALHCVISKIWE EKLPEEWMVGVCPIYKKGDKLDCGNYRGIT	cd0165 0	547	807

			AAAGT	LINAAYKILSQILCRQLSPHARRFVGPHYQAGFT		
			GTGAT	GARATTDQIFCLRQILEKCREYNVPTHHIFIDFK		
			CAATC	AAYDQTVREQLWQIMHENGFPDKLTRLIKATL		
			AACGC	DGVMCHVRVSGXLAEPFGSRRGLRQGDGLSC		
			CGAAG	XLFNIVLEGIIRRAGIETSGTIMNKSYQLLAFAD		
			TGCGT	DIDIVARNLETVMDEVYTRLKVQARRVGLGMN		
			CGCGA	TTKTKYMRGRGSKDVGPPSLTPLTVDGDELEE		
			ACCGG	VSEFVYLGSLVTADNDTSKEIRTRIFAGNRAYF		
			CGTCG	GLRKTLTSDRVQRRTKLTMYRTLIRPVVLYGH		
			TGTGT	ETWTMRQEDERALGVFERKVLRTIYGGVQVS		
			TATAA	DGVWRRRMNHELHAFLEEPTIATQAKIGRLR		
			TTGAC	WAGHVARMDSQPVRLLEDRAEPAGGTGRRRA		
			CCGCA	GKPRARWGDQVNGDIRKICNLGNWRVAAQD		
			ATCGA	RERWKRLATARVPGALC		
			CGGCT			
			GGGCA			
			ATAAA			
			AGTGA			
			AGTTA			
			TTTTTG			
			CCGTC			
			CATCG			
			TCGTC			

				GTCGC					
				GGGAA					
				AACAG					
				GCAGA					
				ACAAA					
				GCAAG					
				AAGTG					
				TGGTG					
				TAGCA					
				ACAAG					
				ACCAG					
				TGGCA					
				GTGCC					
				GTTGT					
				GTCGG					
				TTGCG					
				TATAC					
				GCACA					
				CTAAT					
				CGAAA					
				CCGCG					
				CGCGG					
				TGTGC					

				CGAGA					
				GTGCG					
				AATTT					
				TTGAA					
				AAAAG					
				AAAAC					
				GGGCA					
				AAAAG					
				TTTTG					
				AACTT					
				TTTTGC					
				AAAAA					
				ATAAA					
				ATAAA					
				TTTTAT					
				TATTG					
				CGAAA					
				TTTGG					
				CGCTA					
				GTAGT					
				GAAGC					
				GCATT					
				TACGC					

				ACGCA					
				GCGGA					
				GGACG					
				TCCTCT					
				TCCGG					
				ATATG					
				AAGAC					
				CGTTA					
				GGCTT					
				CGAGA					
				TAAAA					
				GCCCC					
				GTTAC					
				GGTAK					
				CGTGA					
				GAAAA					
				CCGGG					
				GCTAG					
				GTCTC					
				GGTAG					
				CCCGT					
				AGGTG					
				GTTCT					

				CCACC CCGAG ACTAG GTAGC GCAGC CCTGA TAAGG CTGCC TACCG AAAAA GAGGA TCGCG A					
Rex 1	Rex 1- 1_C Ca		Cyprinus carpio	CAAG	TGAACCC CCCCACC CCCCCTC CCTCTAA CATACGA TGTCATG CACCAGT CACTTTG TGCAGCA GTGGTCT	MAARPHAAASLCPDRTVFSCFLSCESQCFCCTSS SRLSVFKRAYSREFLLDVGRAAFLELNSAHAE QLRDLGLLRRPTSSPTTTASRPQRRRHKRER KQKRGKRGGIRARLTANPHKPSIPTIVLANVRS LDNKLDYIRLLRSTQRTVRDCCVFVFTETWLS DSVPDCAIQLDQLTCYRADRARVAGGKTRGG GLCVYINDAWCRDAVVICKHCSPLVEFMIKC RPFYLPREYTAILLVSVYIPPNNNNNSNRNDALN ELYQHISEQQT AHPDAFLIAGDFNHADLKSVF PKIHQHINFPTRGNNILDFVYTTQRGAYKALPL	cd0165 0	597	850

				GCTCACT	PHLGASDHITVMLMPAYRPLVKVAKPVQKQI			
				ACCTCAT	QVWPEGSSDALQDCFDTTDWNMFKQAATYN			
				TCACCAT	NXTDLQEYSETVTAYITKCIEDVTVTKTITVRA			
				GGAACTG	NQKPWMTGEVYRLLLETRNAAFRAGDEGGLRT			
				ACGTCAT	ARANLSRGIREAKRQYSRRIAHQFSDSRDTRSL			
				TCTACCA	WQGIQTITDYKPPQRTCDNISLLNELNTFFAR			
				CCTCATC	FEAQNSTTAQKTPPPSDQVMMLTPDSVRRSF			
				AGTCAGT	SRINARKAPGPDNIPGRVLRDCAAELTDVFTDI			
				TTAAATA	FNISLSQAVVPTCFKATTIIPVPKKPSPSCFNDY			
				AAATAAC	RPVALTPILMKCFERLVMHHIKSALPPSLDPFQ			
				TGCTCTT	FAYRSNRSTDDAISTTLHSALTHLEKKDSYVR			
				GAGCCCT	MLFIDFSSAFNTIIPQQLIHKLVGLNTSLCNW			
				TTGTCAC	LLDFLTGRPQAVRVGSNTSSTITLNTGAPQGCV			
				TTTGTC	LSPLLFTLLTHDCTPSHNSNLFIKFADDTTVVG			
				CTTTTAA	LISNRDETNYRSEVSRLAGWCSDNNLSLNVEK			
				TCAGACT	TKEIVVDFRRAHTQHVPLTINGATVERVSSTKF			
				GAATAAG	LGVHITEDLSWTDNTAALAKKSQQRLYFLRKL			
				CTATTTT	RRARAPPPIMYTFYRGTIESILTSCITVWYGAC			
				TTATGCA	NASCRKTLQRIVRAAEKIIGVSLPSLQDIYGTRL			
				CTAAAAA	TRKALCIAGDPHPHSHSFFSLLPSGRRLRSLQA			
				TCTTTTTA	RTSRLKDSFIHQAVRKLNSLPNLP RPSSASGT			
				TCTGCAC	TEL			
				TGTTTGTT				

					CACTGGT				
					TTGCACT				
					CTATCTG				
					CCATGTG				
					CCTTGCG				
					CTGCTTTT				
					ATTTAAC				
					CTTATTTT				
					TATTTTAT				
					TTTTTCTA				
					TTATGTC				
					TTTTTTAC				
					ATCCCT				
					TATTGTA				
					TAGTTT				
					ATCTTAT				
					ATTTTAT				
					ATTTGAT				
					CTAGATT				
					TTAGGCT				
					CTACTGT				
					TAGTGTT				
					ATCTGTA				

					TGCACCG GGGGTCT GAGAGTA ACGCAAT TTCGATT CTCTGTA TGTATGT ACTGTAC ATGTGGA AGAATTG ACAATAA AGCAGAC TTGACTT GACTTGA CTTGACT TG				
RTE	RTE - 2_D Pl		Danaus plexippus	ATCCA GAGGG TGTCC TACTA AGTAT TTTTCC GATAT	TAATGCT GGGATAA TTGCCAG GAAGAAG AAGA	MVAKVYGLGTARASPINDAQHRPTPAISGQG LPCQVRARLKTLPQRKLR FATWNIGSLTGRC RELADVLMRRRVQCAFLQETRWKGNKSRNIG QGYRLIYTGSPSGKAGVAVVLSEELRNGLLEV DRRCDRLMRVRVLIEGVITNLISAYTPQAGCSG SEKESFWEQFEEVLRAIPAAEVIIVGGDLNGHV GRAAETFDRVHGGFGYGRRNAEGENILRTCIA	cd0165 0	561	805

				TTAGC	SDLAVVNTFFQKTPQHLITYKSGSHSTQIDYLL			
				CTTGG	TRRCHISKVTNCKVIPGESLTAQHRLLVMDYV			
				GCATA	VTPKKKVAEKRKPRIRWWLLNGTMQTSFRAE			
				CGGGT	VESQNLSTNTETAQEVWDRAQSAIITAGKRVL			
				ACGCG	GLSKGGRVIDKETWWWNDEVQEVIREKKTAF			
				ACCCC	KKWQQSNSPEDRLEYIAAKRASKRAVARARS			
				GGATC	DRLSPLYDTLETAEGQKLIYKLARARDKATQD			
				CTCGA	IAKCLSVKDSQGTLLCNHASVKERWRCYFKEL			
				GCCAC	LNTQHPCSLPTEIPPNLGLIAPITPDETRNCLRR			
				CACTT	MKNRKAVGPDDIPIEAWKSLGSLGVLILTDLF			
				TCATA	NRVLNTGTMPHQWRYSYITPIYKGRGSVQDC			
				CCTCT	GSYRGVKIMSHTMKLFERMIDLRLRRECTVSE			
				ATCCT	CQYGFQPGSGTLDAIFAIRTLMEAYREKRRAL			
				CAGTC	HVAFLDLQKAFDCVPRQCIWWALRFKGIPEAY			
				CATGA	IDIIRDMYRDSVSMVRTAVGDTKPFPISVGLHQ			
				AGGTG	GSALSPFLFNVLDTVSANIQDQPPWLMMYA			
				GTGTC	DDIALIDESRLTLERRVNLWKGTLENGGLKLN			
				CTGTA	VTKTEYMACGSPDSCTIHIGPEPAVKSEKFRYL			
				CCTCC	GSILHESGGIDHDVQARISAAWAKWRDVTGV			
				CCTCA	VCDRRIPTKLKGIIYKSIIRPVLLYGSECWPTLS			
				TTCCC	RHTQELHVTEMKMLRWMCGVTRADRIRNTFI			
				CTTCCT	RGSLGVRDVADKLQESRLRWYGHVARRPENY			
				AACCT				

				CTTTCC CCCCT TCTCT ATCTTT TCCTTT CCTGG TTTTAC CCGTA AAACG GGGTT TTGGA GGTCA GACGG CAGTC GCTCC GTTAA AACCA CTCCC GCCAC TCTT		VGKICLDMSVPGARPPGRPRKRWLDTVKQDM RANGLTTADAKDRAKWRSLSRKADPG			
Rex 1	Rex 1-	.	Danio rerio	CCAAG ATGGC GGCGC	TAAATTG AGGGTGT TTCTGGC	GHSENTGSPWFVIPGNNGRRREQKAETRKKSG CVDDSGKPHTSHRSPASSYECQIIANKMDELGF SWTMQGLLPLMLTESWLHPGIPDAAVELTDFT	cd0165 0	546	799

6_D R		GCATT	CAAAATT	LHRHDRNHL	SQKQRGGGLCIYISEKWAKDTK			
		GACGC	ATTTTAA	AVESLLPRFGV	SDSNHCSPDLEYLTVKCRPFFL			
		AGCGG	TCTTTTAA	PREFSSAVFIT	AVYIAPDANANSALNLLHNAISS			
		CTCAG	ACTTGCT	QLNKYPEAAHII	AGDFNHVDLKIVFPKLHQHV			
		AGCTC	GCTAGGA	NCATRGGNLLD	KVYSNIKRGYKTIPAPHIGQS			
		TCCTTC	TATGCAC	DHISLFMLPVY	SSLRKRDPPTLKT VTVWPEGA			
		TCAGT	TTGCCAC	SLQLQDCFAHT	DWSVFESQDLEEYTTSVLSYI			
		GCTTT	TTTATAC	KHCTDTVTTEK	HIRVYANQKPWMTGGLGFC			
		GTTTG	ACTTGGA	LYRHCHHRETH	QSVCPKTLDDRGVRILLKER			
		TTTGTT	CATTAAA	NIAFKSGDKERY	SAARANLKTGIRGAKLKYKH			
		TTTAT	TTGTATG	KIEGYLHSNNS	RQVWQGIQNITNYRPNPRSVG			
		ATGTC	CCTTCCT	VDNTLAEELNC	FFARFEQQSAEEATLHSLSNPS			
		TGTTT	GGCTCAC	HVFTVEEDDVR	KVLRTVNPRKAAGPDGITGR			
		AATTT	ATTTTATT	VLKDCADELTG	VFTKIFNLSLSQAIVPLCLKAS			
		TTTTA	AACTTTT	TIIPVPKKTINS	LNDYRPVALTPVVMKCFEKL			
		ACCTC	AAATTTG	VRSHILQRLPP	AFDAHQFAYRENRESTEDAIAIA			
		AGCCA	AAAAGCT	LHTTLTHLEHK	DSYARILFIDFSSAFNTIIPKRL			
		GTCAG	GCTTGGA	VSKLNHLGFPF	SICLWIQDFLSNRSQRVKVEHH			
		GCGAA	AATGCAC	VSSALSLSTGSP	QGCVLSPLLYTLYTHDCTPFH			
		AATAT	TTGCCAC	SSNTILKFADDI	AVVGLITDGDETAYRDEVNQL			
		TTTAT	TTTATAT	ALWCTDNNLAL	NLTKTKEIIVDFRRNRPDTLPI			
		AGTCG	GCACATC	HIHGDCVEIVS	NFKYLGLHLDQNIWKTNTIAL			
		ACAAG	ATACGCA	VKKAQQRLHFL	RVLKNNNLAQELLVSFYRCTI			

			ACCTC	CTTTTAA	ESVLSYCI	PVWYASCTAQEKAALQ	RVVNTAH			
			CTCTC	TATTGTA	KIIGCTLP	SLEELFKARCYSKAC	SILKDSSHPAY			
			AGTGG	TTGTAAC	HLFQMLP	SGRRYRTIKAKTNRL	KNSTYPRAVA			
			GATTA	TCTTCATT	IMNGH					
			CAGAG	TGATTTT						
			CAAAG	ATATTTTT						
			AGACA	ATCTCTTT						
			GTTAC	TTAATTA						
			AACAA	TTGTTGTT						
			TGTTT	GTTTTAG						
			ATTCA	ATAAATA						
			GAACC	GGACTGC						
			ACAGC	ACTGAAG						
			ATACC	GGAAGGG						
			TGA	CAAATGT						
				AAATTTC						
				GTTGTAT						
				TTTGTAC						
				AATGACA						
				ATAAAGA						
				TTCTGAT						
				TCTGATA						
				TT						

Rex 1	REX 1- 1_D R	Danio rerio	TCGAT CCAAG	TGACTCT TTTGCCC CCCCAAT ACACCCC CCACTCT CCTCTAA CTTAAAC TCCTCGC AATAACT GCACTGT TTAACAT TTGCACA TTTAAAA TTTGCAC ATTCACT GCACCAC ATTGAAC TGTTTAC TTATTGA ACTGTAC ATACCCA CTGCATA TGGGCAT	MAARAQREAQGLSSFCIFAVLFLISGLFVQNS GAFTSYTRQELLDIGLCIPDSFISNLRLIHEIART PRAGRPARPGGSARRPRRERKQRRGKRGGLR AKLKLTPHRLSLPSIFLANVRSLVNKMDEIRLR INHSKRLWNCNVMIFTETWLNSGIPDNAVSLT EHHTFRADRTADDSGKTRGGGLCIYINKAWCT NSVVVGRHCSANLEFLMVKCRPFYLPREFTSTI ISAAYIPPDADAKLAMNELHAAISKQQTAAHPE AAFIVAGDFNHSNLKTVLPKFHQNILCHTRGN KTLDHVYTNIAEAYAVTPLPHLGQSDHLSLFL TPKYSPLIDRVKPSVRTIKVWPAGVDSTLQDRF QHTDWGMFASQATHGSHTDIDSYTSSVLEYIN TTIDSVTTQKQITTPNQKPWMNKEVRLLLKA RNTAFRSGDAQAYSTSRANLKRGIKKAKHCY KLKLEEHFSNSDPRRMWQGIQAITNYKPSQST PTATDVSFLNELNDFYARFDRDNKEPYTRITSS TDHSPITLTSSEVHTALSRINGRKAAGPDGIPGR VLKVCAEQLAGVFTDIFNLSLNLAAPVTCFKT TSIVPVPKHSSPTCLNDYRPVALTPIIMKCFERL VLAHLKDSLPLSTLDPHQFAYRGNRSTDDAVSI ALHSVLTSLDNKNTYARMLFVDFSSAFNTVIP SKLLIKLRDLIDITSLCNWIMDFLTNRPNQNVRS GHICSTTVTLNTGVPQGCVLSPFLYSLFTVDCR	cd0165 0	584	838
----------	-----------------------	-------------	----------------	---	---	-------------	-----	-----

				TTGTAAT	PVNRSNTIIKFADDTTVIGLISNNDETAYREEIQ			
				TATGTAC	HLATWCTDNNLLLNTNKTKEIVDFRKGRGTS			
				ACACCCA	HDPIHINGMAVEPVSSFKFLGTHISKDLSWTTN			
				CTATACA	TSSLIKKAHQRLFFLRQLKKNQLSSAVLVNFI			
				TATACAC	RCTIESILTNCVTWVWYGSCSVAERKALQRVVK			
				TTGTAAT	TAQRITGTTLPAIEDIQKKHCLRRARSILKDTFH			
				CATGCTT	PAHRQFSLPSGRRFRLPRTKTSRLRNSFFPRA			
				ATTTATC	VSLNLSAPH			
				TGCATAC				
				TACTGAT				
				TATTAAT				
				AGCAACC				
				TGTACAT				
				ATATTCA				
				TATATTG				
				TAACATA				
				TACACTT				
				GTAATCA				
				TGTTTAT				
				CTATCCC				
				TGTACAT				
				ATATTCA				
				TATATTG				

					TAACATA				
					TACACTT				
					GTAATCA				
					TGTTTAT				
					CTATCTG				
					CACACTA				
					CTGATTA				
					TTAATAG				
					CAACCTG				
					CACATAT				
					ATTCATA				
					TATTGTA				
					AATCTGT				
					TCATAGC				
					TTATCCA				
					ACCTGTA				
					TATAATG				
					TTCATAG				
					TACATCC				
					ATCTGTA				
					AATATCA				
					CCATAGT				
					TTTCTAT				

					AACTGCA CTTTATA ACTTATT CCTGTAT CCTGCAC TTGCTGC TATTGCA CTGCTGG TTAGACC TAAACTG CATTTCTG TTGCATT GTAATTG TACATGT GTAATGA CAATAAA GTTGAAT CTAATCT AATCTAA TCT				
CR1	CR1 -	.	Danio rerio	ATTGT ATAGC TTTGA	TAATATA TTCAAAA AACATTT	MATGGTGVVITLLVYCYFTSLTAAIRVDPVRM FCSIYPLSYHNDNPHIYKDLGYKRKGNELAKT STGYGTYALAVTMVYPSLYNKITKKKKAMTT	cd0165 0	800	1056

10_ DR		GGCGT	AAAAATC	TQSSRTGLIIAICMLSGDIHPCPGPHHFKTPRE		
		AATGA	TGGTTTA	ANPQTEDKRATLTLQVCSSCTLLCSSAQQANP		
		ATAGT	TAGAGAA	LLSLCPTPNTLKL SVQGVCEVKGAAAVPGWS		
		GCGAC	TCAGACG	DGERQGEKTHPRLCVGMSDRSFGREFAGRVC		
		AAATA	TGTCATC	AFRGNKRPDPETQQQQQTFLNDSASTMATNS		
		GAGTT	ATTCCGA	VNRKRLNPGVVKNRKHAFFNTVNNSKIVWDP		
		AAACA	TTAAAAT	KMKPKGLLGRHLNIRSILSKSEQITHLLTNSNL		
		ACCTC	GAACCTA	DYLCSETWLYKYAPSAAFNISGYTMFRKDRE		
		TTTGCT	GCATGTT	GAKGGGVMIYVKDNLQCEEINWLCDYEFEC		
		TGCCA	TCCAATT	GLNVRLSPQMSFKIIVIYRPPSSDGTIFYDRFEKL		
		TTTTA	GTGCCYA	LKESNLNKETVILGDFNINWEDTSCRNKLKRIT		
		AGACT	CCATCTT	DKFDLVQMINGPTKITTSSSTQIDLIFTNRPERII		
		TGCTG	AATGATC	KTYNMLAGISDHNLT LVARKLSKKRKFKSARE		
		TCTTTT	TGTTGTTT	HEFIGIPRNR RDGFKN AVEQIDWDRLLL GTDLE		
		TTTAG	AAATGTG	KDRGSFVG TIEKTVKNSALNLNANTIKIYYPGL		
		TCTCT	TTTCAAC	TLTYHTKTNERERFCVKKVIKSKSASDRNMFT		
		GTGTG	TATTTTTA	TLRNRVVKNIRKAKADYFLT LIEKAKGDSKTI		
		CAAAA	TCTCTATT	WNQLKKVVRHDTKNKKMLEVMIDGHITNNA		
		ACGTC	ACAAAAT	CKIAEAFNNFLLTQFQIWSHCVHQKFLTCALFF		
		TTGCC	GTTGGTG	INSVSNMVPLCPSKISNMCSLNSSQPIFSISNISE		
		TCCTA	TTTTAGY	SDALQAIKSLKSSRTKETFGIDSLMLKELSSTLL		
		TTCAG	TTAATTT	TPITKLLNRSISEAIFPCVWKSSAVIPIFKNGDH		
		CGTTG	GTAAATA	YSLSNYRPISIIPTVSKVAEKLIANQLIKHLNTTT		

				GCGCT	TTCTCTTT	FSLHPMQFGFRSQHSTETANCFLIENIKHLLDK			
				GTGGC	TAACATC	GGVVGAVILDLKKAFTVNHKILLSKLTKFNM			
				TTAGT	TAYATRA	SPESLKWFESYLSNRSQSVRINNYHSSPLSLPT			
				TGGTC	CCATGTC	GVPQGSILGPLLFSLYINDLPSVCQNLIQMYAD			
				AAAGT	TATTTTA	ETVIXTHGKNNTTQVAEKLTQAMAHVSEWLNQ			
				GCCTG	GTACTTT	SCLKLNVSCTVAMYFSKTNIVNNLTDIYVDGD			
				TCTGG	TTAACTA	KLQIVSDFKYLGVQIDSKLTFKSQIKKVCNRXK			
				TAAGA	TGTTTGA	FNIRSHLSLQAAXIFMHSMILSHITYCLTTWSQ			
				AACTT	TWGATTG	ANKTTLKPIESLYKQSIKILXKKPLRHHHCILQ			
				GGCTG	TATAATT	KHKLLSWENVIKYSNICLLYKIIHGLSSPPLSQF			
				AGGCC	CTTGTAT	VNIRNSTHRATRGAARGELYRPTPEKYLWSGC			
				AAG	TGTTTTTT	LFSYSCTGVEFCPWQHQRYNLL			
					TTAGGTT				
					G TTCAGG				
					GACAACA				
					GATAAAA				
					AATAGCC				
					TTCKGGC				
					TAATTCT				
					GGTGCAT				
					TTGCAGA				
					AATGCTA				
					ATTAATG				

					TACACTG TCCCTGC CAAATAA ACAAATA AATAAAT AAATAAA T				
L2	L2- 1_D L	.	Dicentrarc hus labrax	T	TAGGGCT GGATCAG GTGAGTC CTGAACC ATCCCTT AGTTATG CTGCTAT AGGCCTA GACTATC GGGGGAT TTCCCAT GGTGCAC TGAGCTC CTCTCTTC CTCTCTCT TTCTGCA	VRNHSINFHCYADDTQLYLSIKPDEXNQLSKL QTCLKDIKSWMSCNFLMLNSEKNEILVLGPKN LRDSLKDIVTLDGITLASSTTVRNLGVIFDQD LSFNHSHIKQTSRTAFFHLRNITKIRHILSQTDAE KLIHAFVTSRLDYCNSSLSGCPNKSIIKTLQLIQ NAAARVLTGTRKRDHISPVLASLHWLPVKCRI EFKILLTYKALHGQAPSYLKELIVPYYPXRTL RSLNAGLLVVPIVSKSRTGARAFSYQAPLLWN QLPVVVREADTLXTFKSRLKTFLFDKAYS	IPR000 477	5	93

					CTCATTC ATGTCCC ATTAATG CATGTTA CTAACCT CACTTCT TCCCCGG AGTTCTT GTGCTTT CTCGTCT CGCAGGT TCTCATG GATCCTG GTGGAGC CTCCTGC CATGGTC CTGCTTG ACTGCCA TT				
CR1	CR1 - 3_D Bi		Drosophila biarmipes	AGGCA ATCGA TTTGA GGGGA	TGACCAA AAAGACG ATTTAAG ATTGTTG	MKAFMRQTRKALTEFRKQLTSSFEVFSIMETQ FLDCKLLNESPKRKIANNNKNSLQIPSPCIAAV GIPSPISACLNSILAVNNSAGPGTSGLQSSFAFN NTSPGSEPGASGLQPNQNTGTTAEGSSTPQQL	cd0165 0	775	1027

			AAAAA	TGATGGG	APGPGSLHINPDHATTIAPAALPSSQPAANQPL			
			TCACT	TTTCCGC	SAIPLSGVAPRSNAADLPLRGIAPNRSVSVFVSRL			
			CTACC	ATTAGGA	NPEATADCVKSYLCQKLGAQPNSISVFKFNFK			
			TCAAA	TTATTAG	QPRNISSFKITLPESMFSNALDPTIWPEHTIMHE			
			GCCAC	GATTCCC	YIRKGASQDPNSVDAKNCKKLNRSISVAYQNV			
			AGAAA	ACAATGT	RSLLGKLKRVYSNSTLLDFDIIAITETWLNTTV			
			AATAA	TTTGGAG	YDQEIVNKFAYRLDRPTRGGGVLI AVLNSHS			
			AACCA	CAATTTC	SELIPYDNMHSIEFIAVKVTTDDKCNVFLTCSYIP			
			TACAT	TTGATCT	PGSDPAIYLHHLAISAVSLAVGDADQLIVLGD			
			GGAAC	GCAGCGA	FNLPHLSWLPNSDSNILFPENQHEFIDGLTNLSL			
			TGCCA	AGTGTTT	GQVNSIKNSRERILDLIFVSDVSVATVSRCDPV			
			AAATT	GCCAAAT	STPEDSYHPALYLSVHINFETTNQSASQRGNTI			
			CGACG	GCTGTAT	RDNYESICLTTREHYTGLNSHFDRLNWSTRLYS			
			GTAMC	CAGCAAT	FTDLDSAVLWFNESVCSALDTFVPSYRWKLIK			
			TACGT	TTCAGCA	ASLVLQAFVLHIPLSQVTVGNSLKPPWFSRHLS			
			CGAGT	TACGTTA	YIKNTKSLLKDIKNPRFFYAEYIAECRQNIRSDP			
			TCAAA	ACTGAGA	KKFYSFVNKRKSNVFPPCLHLGNLSESSDINIS			
			ATTAT	GACATAA	NLFATFFQSTYSLPHINNIHIPPFKISDVLNGML			
			TACTA	TTAGCGT	TIDESFSAGPDKIPSCIIKQCAASLCGPITFLFNL			
			CTATT	CAAATGT	SLQLSFFPSVWKTSFLIPLHKKGSKTDVGNYRG			
			CAATG	ATTGACT	IAKLSAIPKLFERLVTCQIQHMCSSIISTSQHGF			
			TCAAA	TATGTAT	VKKRSPSTNLLEFSRLIHCSFLQKKQADVIFTD			
			AAAAT	TCTTATG	FSKAFDSVNHSLLIYKLSLIGFPPYLLDWISSYL			

			CAATA	TTACCTT	SNRSQRVLFKNVSSLVNVTSGVPQATSHLGP			
			ATTTT	GCAAACC	LLFILFINDLPQVMSYCTTFMYSDDVKICMSFS			
			ATATC	ATTTACT	NLSSHLLIQSDLNKFQAWASTKLLSLNYSKCK			
			CGCTT	TAATCTC	TMSFSRTTFHTTTSYTLGNSLLERVVKINDLGVL			
			TGAAA	GTTTAAA	FDQKLCFNHISVMVNKATGVLAFIKRWAKEF			
			AGGAA	AATTATC	NDPYTTKSLYVSLVRPIMEYCSEVWNPQYVY			
			SAGTA	TGATTTT	QQNLIESVQKQFLLFALRNLNWDPRHLPPHL			
			AAGGC	GACAAAA	SRLKLIDLNTLKHREFNCKIF			
			TTACC	TTAATGA				
			AAATC	TAATTGT				
			AMGTC	TATAAGC				
			GACCA	CGAGGTT				
			TGAAT	TGATAAT				
			AGGGG	TAACCAA				
			GTATG	ATGCAAT				
			TAATC	CCAGTTG				
			CGTAA	GGCTTTC				
			TGCAA	CAGTTTC				
			GCGCC	GAATCCG				
			ATCCG	AATGAAT				
			CCGCA	TTAGTTG				
			GTGTT	TGTTCAG				
			ACGAT	AAGAATT				

				CTCTCT	ATTGCAC				
				TCGAC	AACTAAA				
				GCCCT	GATGTTA				
				TACTA	TGAAAAA				
				GTTTTT	ATAGATT				
				ACGGA	CAT				
				TTTCTT					
				CTCTC					
				GGCAA					
				CGAAA					
				AGTTG					
				GAAAA					
				AGTCA					
				TGTCC					
				AGGTT					
				AGATC					
				GAAAA					
				ACGCG					
				GTCGC					
				GATGG					
				GAAGA					
				ATAAA					
				CCAGT					

				TAAAA					
				TTCTA					
				GTCAG					
				ATATG					
				GAAAT					
				ATTCTT					
				GGGAA					
				GCGAT					
				TAATA					
				AACGT					
				TATCA					
				ACACT					
				TAGTG					
				GCTGA					
				GGATT					
				CCACG					
				ACAGG					
				AAGGA					
				GAGCT					
				TCCAA					
				GCGGG					
				AGCAG					
				CCAAA					

				CAAGC					
				CATGC					
				TTTAG					
				AAAGA					
				AGGCA					
				TTCGT					
				AGTTT					
				CGTGA					
				CCTGM					
				CGAAG					
				ACTCC					
				CAATT					
				GAAAT					
				KTCAA					
				TTTTGT					
				AAGTC					
				AAAGT					
				TCTAT					
				CAATT					
				GCCAA					
				ACTGT					
				TTACA					
				TTTCG					

				CCCAA					
				AGTCT					
				TAAGA					
				AAGGA					
				CGAAA					
				TTTCA					
				ACAGA					
				TGTAA					
				TATAT					
				GCAAA					
				ACATA					
				GACCG					
				CATTA					
				AACAT					
				ATATT					
				GTTTC					
				GGTTT					
				TCCCTT					
				GCCCG					
				GTTTC					
				GCCTT					
				TTGTTT					
				CTAAG					

				AATAG					
				ATCGG					
				GAGTT					
				ACAAA					
				ATTAC					
				GATAG					
				TATTG					
				AAATA					
				TTGCTT					
				ACATA					
				GTCAT					
				TATTG					
				CTTTTC					
				GGCGA					
				GCCGA					
				GCAGC					
				TCGAT					
				TGCTC					
				AAGAA					
				ATATA					
				TAAAC					
				TTAAT					
				TATTT					

				ATTTA					
				TAATG					
				ACAAT					
				TATTA					
				TTTAT					
				ATAAA					
				AATTA					
				TACAG					
				TGTAT					
				TGGCC					
				AGCCA					
				TTCCA					
				TTGAA					
				CAAAC					
				GTATT					
				GCCGT					
				TGTGC					
				TCTAT					
				CCTTA					
				AATTA					
				CGAAC					
				TATAG					
				GCAAG					

				TCAAA					
				CAGTC					
				ATTTC					
				GCAAC					
				GATAC					
				TGACC					
				GGTTG					
				TGTTTT					
				ATTTG					
				AGCTC					
				TTGAA					
				ATTAC					
				CTTTC					
				AATCG					
				CGGTT					
				ACACC					
				GAACG					
				TATAT					
				AATCT					
				TAAAA					
				TTGGA					
				GAATA					
				ACTTT					

				CTCAA					
				CCCCC					
				GCTTA					
				AACT					
				TCACG					
				CACCG					
				TATAT					
				AAACA					
				TTACA					
				TCGCG					
				CCAAT					
				TAACA					
				TATTT					
				AATTG					
				AATCG					
				CAATT					
				CATCT					
				GCTGT					
				TGCCA					
				CAAGK					
				GCTTC					
				TGTTC					
				CCTAT					

				CGCCG					
				TCGCG					
				TCTGTT					
				CTTTA					
				CACCC					
				GCTCG					
				CCAAT					
				TGCTC					
				TCTCA					
				AGGCC					
				ACGGT					
				CCTCT					
				CTCCG					
				TGCAC					
				TTCAC					
				TTTGA					
				GCGCT					
				TCGCC					
				TCTCTC					
				ATCGC					
				CTGGA					
				TATTC					
				GAATT					

				GGCTC					
				TCCGA					
				ATCGA					
				CGACT					
				CTCCC					
				ATCAA					
				GTATT					
				CTAAC					
				GGTCC					
				TCTCTC					
				CGTTT					
				ACTTC					
				GAGTG					
				GGTCG					
				CCTCT					
				CTCAT					
				CGCCT					
				TGCTT					
				GCTGC					
				GGATC					
				TTCGA					
				ATTGG					
				CTCTC					

				CGAAT					
				CGTCG					
				CCTCT					
				CCCAT					
				CGCCG					
				TATCT					
				GCCAC					
				TACCT					
				GGTGT					
				TCTGC					
				GTGTC					
				TCTTTA					
				ATCTC					
				CCACG					
				GCCTG					
				CTCTTC					
				TGCAA					
				GATTC					
				TCTTA					
				ACTGC					
				CTCTT					
				AACTG					
				CATAG					

				TGTCC					
				AAACA					
				ATCAC					
				TAAAA					
				TGAGT					
				GTCGT					
				CATCT					
				GTTCC					
				GCAAA					
				GAAAT					
				GCGAG					
				TTCAA					
				TTGTG					
				TTATC					
				ATTGG					
				AGATA					
				ATTAT					
				TTGTCT					
				TGTTG					
				GCTTT					
				GCGAC					
				AATTG					
				CGCTC					

				ATGTA					
				AAATG					
				CGCAG					
				GAATG					
				GGTCA					
				AAACG					
				GCCGT					
				GTAGT					
				TGATC					
				TGATC					
				TCGAA					
				GAGAG					
				GCGGC					
				TTAAC					
				TTGGT					
				TCTGC					
				CCTAC					
				CTGCC					
				TCGTA					
				CTTAC					
				TCGGA					
				G					

RTE	RTE - 1_D F	Drosophila fusciphila	ATGTA GGGCC AGACC AATAA CAATC GTGCA GTGTT ACGCC CCAAC CGAAG ACGCA AACGA CGACA TCAAA GACGA CTTCT ATACC GCCCT TACTA CCACT CTCAC TGACA TACCT	TAAAGGA AAAAAAA AA	MGDFNAKIGPNFTRMRSATGGHGIGTRNNNG DRLVEVCQMFQLIIGGSMFPHRDIHKYTWTS SGNTRNQIDHICISRKWRKSLLDVRNKR GASIDSDHELIIGELSIKLDNRNRPRNADDSR NHRPSPLNIHLLGETDISARLASTLRTHLLPQ QGQTWEQTCDTLRRTAEGIIGTQQRNRNNW ISSATWDTINKRNSLKLLADRDSNRRDEHRA LCKAVKRAARNDKRALLDRLADNAERAANE NNMRSLYQTIEQISGGYRRKTQPVKDANGIL LTSDDDQIHRWRQHFMESISNDMPSRSDDYR SVAPANGNRRISSAPPSVREIGNAIKKLKNNK AAGMDGIPAEILQVDTQLMAEILHPHFSSA WESEHIPDSWKCGVIVKIPKKGDLSDCNNWR GITLLNTSYKVLATILNERLTEKFEPTLRDE QAGFRPHRSCVDQANTLRIITEQAVEWRAP LYLLFIDFKKAFDSINRAAIWGALARKGVPH KVIALIKALYDDANLAVLHNGKTSAPFQTNT GVKQGCPLSPLLFNIVVDEVMREVCKHKRG ITWSFTRHLEDLDYADDICLISHKLSDIQSK VDDLVLNASRVGLQINTAKTKAMRLNHA SQGNIIVDGNPIEYVESFPYLGTVISSNGG VTEDEVTSRLAKARSAFGRLYRIWRSRQIR RKTKIRIFNSCVKSVLLYGSETWLTTKTII QQLQAFTNKCLRIICGIFWPNRISNIDLWN MTGETAIRLQIRKKK	cd0165 0	381	637
-----	----------------------	--------------------------	---	--------------------------	---	-------------	-----	-----

				CGAGG CAACA TCACC GTTCT C		WRWIGHALRKDQXSIVRTALEWNPQGSRRVG RPKATWRR TALQELSQYNITWDTAKQTAXNR VRWRNLVEALCSQEEL			
Pene lope	Pene lope- 1_D F		Drosophila ficusphila	AC	TAGGTTT AGTTTAT AACTCCA CGTCACT TGGCGCA GACGTGT CAAAATG TATGTCA ATATTTA AATTGTA ACAAAAA AGTGTAT TTTGATG TTGATTG TTTTTTGT TTTCGAT TAGTTGC CCTGAAG	IVNNKWTILKEHTXTPSKVFLDIVKFCIQDNRY LKYETKLXIXLKGMMPMGSPASPVVADIVXETL LQNTTEKLRRPPRL LTKYVDDLFGIIKEDEIENT LTTLNAFNXNIKFXMELEKDNCIAYLDTV VIR RGNELKLKWKYKKPIASGRIINFNSRHPKMMIR NTASSCINRMLRISDAIFHEEIINEIKTILTN NDF PKHMIKSLLDKYLKSDNNQRPPKKERTYISVT YVPGLSERLAQSDCYDKENILIAHKPDNTLKQI FNKTKSKIAPTDKKDVVYRIPCNGDGKNNCNS IYIGTTKSKLKTRLSQHKSDFKLRHQTNTQKTA LMSHCAKQGHSPNFEETTILAQEQHYRKRYM MEMLHIINTPTNIRMNYKTDTENCAHIYRHLL GK	cd0030 4	47	129

					ACGGTTG CCGAGGA GCAACCG AAATATA TTGGTCA AAAAACA TAAATTG TGTTTTTA ATATCAC CCAGACC TTGAGCC GGCAAAA CACCAAA TAT				
RTE	RTE - 2_D F		Drosophila ficusphila	GGGGG GATGG GGGAG AGCAT ATAAG GCAAA GTGTC TGACT GCATT	TGATGGC CAATTAT TATTATT AA	MKEAEEASERDRFNVDH GKRRKKDNDLRVCT WNVRTL YRPGAAQQLADTLKKCRADIAAIQE MRWIGQGCIKKKDCDIYYSCHPERHEFGCGFV VGARLRRRVSHFLPINERLATIRITAKFFNISLIC AHAPTEEKDDVT KDTFYAELDRA YGRCP SHDI KILLGDFNAKVGREDIFGATVGRFSLHETTSSN GLRLIGFAAAHNMVVRSTGFRHLDIHKATWLS PDRLTRNQIDHVVIDARHASSVLDVRSCRGPNI DSDHYLVA AKIRTRLCVAKVARRG LLRRLDIG	cd0165 0	551	807

			CAACT	KLQSQQTRDAFSTQVSGLLSRTTPTPEDISDQW		
			TGGTG	ALISHSLRTSAEEVIGFQRPPTRNPNWYDQECRD		
			AATGC	ATAAKDAAYRRTLQAGATRAVVEVYRLRRR		
			CAGCA	DEKRLFRKKREXEKRECESIESSRCRNEARNF		
			TGATC	YQRVKRLTQGCKTGGVACKDNDGNLVTDAEI		
			CTTAT	VLKLWRDHFSSLLAGSGTGYEDYDPQTPIYGT		
			CACTG	DIDVPTPSHAEVKDAIQRLKNNKSAGADGLPA		
			ATCTC	ELFKAAGDMLVGSMHQLISKIWRTESMPEDW		
			CTCAA	NLSMICPVLKKGDATLRTNYRGISLLPVAYKV		
			GCCGA	LSSVLCENLKPHAEALIGPYQCGFRPGKSTIDQI		
			CGCAC	FTLRQILEKSYENQIDTHHLFVDYKAAFDSPRR		
			TACCT	DRLYAAMSELGIPAKLIRLCRMTLSNTISSVKV		
			CCACG	GNDQSETFSTKCGLRQGDSLSCMLFNILMESIV		
			TGGTT	RKAGVHRTGTILSNRCVQLLGYADDIDIIGRTK		
			CTCCC	RDVTAAFGAIEKESAKVGLAVNMDKTKFMVC		
			TGGAT	SSRESRRLDSQLTADSYSFESVKEFIYLGTAITS		
			TGCCA	TNDVSLEIKRRVTLANRCYFGLSRHFNSRALSR		
			TTCAA	QAKITLYKTLILPVLLYGAECWTVVQSDAAAL		
			CATCC	GVFERKILRKIFGPICVGDAYRIRWNHEL YELY		
			GTTGA	GDVDVVSrvKIHRLRWLGHIARMEEDAPARK		
			GTGGC	VFDAVIVGKRRRGRPRMRWQDQVMEALSTSG		
			AAACT	AANWRRRAQDRDAWRQIVQQA VTRQGL		
			AAAGC			

				GCGCG ACGAG AGTCC TAGCT CTTCA AACTA GGTTG TGGCG TAGGA CTTGA AATCC ACCCA TGGAA AAACC AAAAT TCA					
R1	R1- 1C_ DK		Drosophila kikkawai	GGAAC AAACC AACCA A	TAGGGGT CACAGCG CCCGCGA AGCAATG CCAGGCG GCGGTAC CGCTGGG	MQLIQLNLNHCRAAQDLLTQTVRERGSEVAIL SEPYRIGNSRDWATDRTGKAALWLCGVGAQQ MCDIKAADGFVRANVGGTWLYSCYLAPSLSL ESFGRILDELSIDLRGHSNVVVGDFNAWAIE WGSSRTNARGRAVLET FASLDV VLLNEGSRQT FSRAGAGSVIDLTFVSSALARRARWTIGEVYT ASDHEAIECSLGMAGRPSGPPPLPRRAFRQDTF	cd0165 0	455	717

				CGTCCGA	RPRAFASALEGFAVDEADGANVMANKLADA			
				TTCGTAT	VEDACDRSMLQRRPFRGNHRPVFWWSEEIAE			
				TCTTTCTG	LRRICHRSRLLQRARGSPLLVECSERYKAAR			
				TTAATTG	KALKIAIRDSKRESFLKLCDAAEEDPWGGAYK			
				TTAAGCC	MTVKKLNAGGNAPSDPEALEGIVRALFPDGQP			
				ATCCATG	VGNFPRAEPGEFRSVTEAEVLQAGRNIKPKKA			
				TTTATGT	PGPDGIPSRATKLALALLPEQVAGLFNKCLQE			
				GTAGTGT	GTFPVRWKRQRLLLLAKPGKPPGEASSYRPICL			
				ATGTATA	LDIIGKVFERVIATRMTEAVEAAGGLSTEQYGF			
				TATTGTA	RKGRSTMDAIMKVVKTAREAIAGSRWKGGTK			
				GAGTAAT	SYCLVVTLDIRNAFNTADWGRTLESLRMFNIP			
				AAATATA	SYLLNVALSYFSNRVLTDTCLGSRAYKVSAG			
				ACGAATA	VPQGSVLGPLLWNAMYDGVLRRLPMPEGTSLV			
				TAAAAAA	GFADDVAIVVVAKELATVEAHADAADVQAVES			
				AAAA	WLANAGLQLAAHKTEAVLVSSRKA VETASVR			
					VGGTTIRSQRAIRYLGVLLDTRL CFKEHLEFVH			
					KKASGTAGALSRMLLNTRGPKQATRKL LTTV			
					VTSQMLYAAPVWAEAAKVKS YVRGVEATYR			
					LCAVRIACAFRTISED AALVIAGQVPLTEL VRE			
					RAEIYAAAQXREATQQTRAESKITARRRSLEN			
					WQNRWDSSQKGRWTH TLIPNLTSWVERSHGQ			
					VDFYLTQVISGHGCFRSYLKRFGHETEDWCPV			
					CGSGVVEDARHVLFECHRF DHERQELEAVAG			

						SSIXAETLVPMMLADQRVWEATATFAADVMK SLREQERRRKEQTE			
Kiri	Kiri- 5_D K	.	Drosophila kikkawai		TAGAGCC CTTTCATT TTTCCAC AAAATAG TGACAAC CCATGAG CCTTCTT ACTTGTT TAATAAA TTAATAC CGGCAGC CTCGCAA AGGACTA AAAATTT TATATGC CCTCGAG TCCAAAT GCTCAAC TCCGAGA GGCAGTT CTTTGTC	SVAPCDNIFNFVCVSQCDVLDKILKVKNAEG LDNINPKFLKLLLPCLLPYVTYIFNSILTQSSFPT AWKLAKILPIPKQNNEYRPISILPFLSKVFENVL AGQIQGFLDRNLLHKNKQSGFRRKRSCLTAIT DITEDIRQQLDKNHVTILTLLDHSAFDSVNHN ILCTKLKNLNFSDSSVRLIASYLTDRQLQAVVA GAVTSVAVKKLTRGIPQGSVLGPLLFAIYVNDLP NVLSDCEVHLYADDVQIYVSSPVREINVCISTC NRMLSLVQTWAEHNGLSINPKKSKSILISRRPI DTNDLTKPIINNIPISYVDHAVNLGLTFNNTLS WGNHISKATGRVYGILRQLANTRNFLPEHIRLL ISKTILIPTLFYGIEIFGFCDTRDTRKLVLTALNSI VRYVFNLNRFDHVSHTTKLLDLNLNSWINL	cd0165 0	72	326

					AACGCTA				
					TCCGTCT				
					ATGGAAC				
					CAATTAC				
					CCCATGA				
					ACTTAAA				
					ACTATTA				
					GGAATGC				
					AACTCAA				
					TTTAATT				
					CCAAAGT				
					AAAACTA				
					TTTTTAA				
					GAAACAA				
					CGCATAA				
					CAAACAA				
					ATAAACA				
					AACAGCA				
					AACTTCA				
					TATTAAT				
					ACGTAAT				
					TAAAGAT				
					ATAAAAA				

					TAAATAA				
					AATAAAA				
					TAAAATA				
					AAATAAA				
					AATAACT				
					AAAAATA				
					ATAATAT				
					AAAATAA				
					AATAAAT				
					AAATAAT				
					TAGTTCA				
					TATAATA				
					ATTAATG				
					TACTAGG				
					TATTAAG				
					TAGAAAT				
					CCTGTCT				
					AATTTTA				
					ACTCAAG				
					TGACCAG				
					GGCATTA				
					ATAATAA				
					GTATTTT				

					ATACTTG TAGCACT GGGCTAA TTTTGTC AAAATAA ATAAATA AA				
I	I- 3_D K	.	Drosophila kikkawai	A	TAATTAT AACATAT TCCACAC ACATATC TTTGTTA ATACACT GTAGAGC GGAAGGC CTCGGTA GCCATTG CTCATTA TCCTTTGT TTAACCTT TAGTTTA TAACTAT TAGTTAG	TIDPAEIANLGLDLSKQAKDSNFSPEFRTNK MLITPHNTYAPTPTAIEIEKNINLIELNSAINSLK GHTPGLDRISYSMIKNTPISEKFRILQHFNDILD SHIPQSYKLSLVLPILKPNKPKTNPNSYRPISLN PCLAKVLDKVVANRLWWFVLSNKLHNNQFG FKRGRSVLDSLLYVDHLATRSLSLKKHISIISLD FAKAFDKIGIHSLIHQLEKWKLGPRIIKYVWNF MTNRKIIVKNGTSLSISFPLDNGIPQGSPISVILF LIAYNSLTNIISIPNXLXFTAYADDFHLILQHNR NNNPQVNLSTLFHQIGEWCSYSGSSLSIDKCQL LHICRKRSCRANIITGSINIPTTDHTKILGLTINK RYSWNHHITNLKTSLSRPLNIIKHLSSRKYN CN TTSLIQIVNSILISKIDYCLPIYGKAPTSILNCLKS TQNSAIRAALGAFRTTPINNLLIEAKINTLENRC KLLTSKLGTKLILSHNSPIGKLSKSIKTKRNYR YPSTLDRIIDSCKSFSLPFSPAIFYTSKKYNPGII	cd0165 0	111	368

					CATTAAA TAAATAA ATAAATA AA	LKATNTSLSLYNKNTTSPITYHQLFHSLKNSLH DHKFIYTDGSKTNDTVGYSVTNNTDIIKSGLLP QYSSVISSELIHAIHEAIKICSGSPGKYAICSDSLSS IKSISNVNNYSYYPNTIRSLITKLFKIRLIWIPS HIGIQGNERADTAAKETHKYPLIFTDNFN TKDI TQLLKKHYTDKLNLSLQAQTTQWYKQILLNTP ANFSFKTTNLNRQQQIIINRLRLGHTQLTNAHY MNRSLTSICPFCNTSAIDIQHILIDCTALSHVRTI CFSNDNPIHLLANPTAANSNKIIYFLHKMNLIH KI			
Kiri	Kiri-1_DK	.	Drosophila kikkawai	GGCCA TGCGA TGCAA AAAGC ACAAC CAGCT CACAG GTGTT TTTAC ACGAC GTGGT GTGGT GTACG	AG	MHNICATGNFSYILEMEGSVDVSHSHTLLRILA KQRSLKICHINAQSLPKKIEEFYMFVNSGLD VICVSETWFVQDFRDSLISCDGYYVYRGDRVS HGGGVAIFVSKKLKTKFVCKHPDNSQIEYIFLE ISNKFASSKILIGCIYRPRCDITYNEIISLLGDISL NYADIIVAGDLNSNMLRETHLKNEFESIGVHL VNSVFPTHFSRTCSSLDDVFFVSDVKKVYLF QLIVPSFSKHDLIYLTYNFDLQYNDESIVFRDF KNLDYGLLMTEFDKCNWNNIFVSPTVDQQLA QFQNNISYLYENFVPVKT KVIKHSVKSWFNND IKLAIKSRNDAYQKWRYRTPEHYGRFKLLRK TVTKLIDSAKTNFYSSRFQTTASTSRTWKEIKS LGIGRKPAAPEQXVDIEELNTSLSKCNVPEICE	cd01650	505	758

				TGAAG	NNMYLRSVPRNLCDSRFRVCVNQCEVLVNIL			
				CTGAG	KIKSDAVGLDDXNPKEIKVLLPKILKYITYIFNT			
				CGACG	AFTTSTFPDRWKIAKIIPVPKANKEYRPISILPYL			
				GAGGG	SKAFENIMASQINSFIHDNALLNNNQSGFRKHR			
				GAACC	SCTSAILKIANDIREYADNSNVTLLVLLDYSKA			
				TCGAT	FDTVNHNVLC AKLVNQFKFSKTAVKLMTSYL			
				ATGTG	QNRQSVVVS NVSSSFATKRGVPQG SVLGPL			
				CAAAG	LFTLYVNDLPSMLSSCDVHMYADDVQMYVSR			
				CTCAC	PLNRINECVDICNNELVKISNWAKENGLGLNPS			
				ACGAG	KSKCIIYKKKLSLNIGPLLLNDSPIEYVERATN			
				CTTTCT	LGFI FNNTLNWSDHINTAVGKTYGLLRSLYLV			
				GAGCT	RNFLPLNVRMLLAKSCLPTLFYGIELFANCDS			
				GATCA	NSQHKLKMAFNSITRFVFNLRFDHVTRLSYNI			
				ACGAT	FELHLADWIKMRVLVYLQKIIFSKLPYLYETL			
				GCTGG	RFSTSRRSNIIIQIRH			
				CGGCG				
				CCGAT				
				AAGAG				
				CACAT				
				TTCGA				
				GATTA				
				AGAAG				
				CTTCA				

				CTCAA					
				TTTTGT					
				ATCTC					
				CTATG					
				TTTAC					
				ACATA					
				ACTAT					
				AATAT					
				TTATT					
				AGATT					
				TAAAA					
				TTAAT					
				GTAGG					
				TCGCT					
				ATGTT					
				TTCTTT					
				GTGTC					
				CTTTTC					
				AGTAT					
				GTTTC					
				CAATG					
				CCTTA					
				ATTT					

CR1	CR1 - 4_D Rh	Drosophila rhopaloea	GCTCA TTTTTG TCTCTC GCCTT AATCC TGGTG CTTCT GAAGA TGACG TCAAA AACTA CCTGA GCAGC AAACT GAATG TTTCTC GAGCC GAACT AACCG TCTTG AAGTT CAAGT TTAAG	TGATGTG AGTCGGG TACTATG TTTATGTT GGCAGAT ATTTGTT AACTTTA ATGTAAT AAATAAA TAAA	SSGPRASDSFNQRSKKLNNFSLHYQNVRSL GKLSKLHVNSVSFDANVIAFTETWLNSSVSDC EIFAGSYVIYRCDRSMRLGGGVLIAVKSDLVS QRIPHIGTDSILEFVAVRIKLHDYFVFISCSYIPPR SDISLYLQHLAQIQSIHSTLGYNDHLIVMGDFN LPDISWMTSTDSIALVPTTAHEFVHGLIDL SLG QLNSILNCSKKVLDLIFVSEFANAFVTRISPLSN PEDAYHPTLEVTLEASAPLQSCSEIEPASFRCFR RAKFAELNRHLSELDWSFMDSNDDPLHIVEQF YGILYSAFDTFVPSFPSANAPQKPPWFSRRLSF LKNSMSRLFKKFKRTGLQVDHSMYLIARSNFL VHNSVCYNGYLNRCRIEFQDNPKQFYAFVNSK RKSQQFPSCLHLDTTSSASSDADIANLFAKFFQS TYSSFSYDNTSSYPYPFLTPTVIFTPLISEDDVL QNLKTIKMSFSSGPDLPSCVIKKCADALCKPL SKLFQISLLHSFFPEVWKESFIPLHKKGSKSDIK NYRGIAKLSAIPKFFEQLITAQLQHQCSSVMSP TQHGFMRKRRSTTTNLEFTSVIRSAFKNNIQTD VIFTDFSKAFDSVNHQLLLKKLTLLGFPIMLTK WISTYLSNRTQRVAFKSTISNSVHVNSGVPQGS HLGPLLFGIFINDLPQVVLSSYVMMYADDVKL CLPLSNPNSHDLLQADLNNLHSCIQNLLFLN NSKCKLMSFYRKRLLLTSYSIGSNDLERINSVT	cd0165 0	514	765
-----	-----------------------	-------------------------	--	--	---	-------------	-----	-----

				CAGAG	DLGVLFNHNLFSDHIGTIVSKAKGVFAFVKR			
				ACGTG	WSKEFEDPYTTKHLYISLVRPILEYCSCVWSPQ			
				AAATA	YTNHQDLIESVQKQFLLFALRGLNWETHRHLP			
				TCTTC	SYVDRLRLNLPSLKDRRTCHGVMFLHKLITG			
				GTTTA	LVNSSYLLGQIRFSIPVRSTRYFRPIALDICKTNF			
				AAATT	EVHDPFRVLCSQYNELYFLIYSECSVNLIVTNIL			
				GGACT	THLSLPPPT			
				TCCTG				
				AGGAA				
				CATCT				
				TAATA				
				AGGCT				
				CTCGA				
				TGTTTC				
				TATAT				
				GGCCT				
				AAACA				
				TGTTA				
				TTGTTC				
				ACGAG				
				TATGT				
				AAACA				

				TGCCT AA					
R1	R1- 1_D Si		Drosophila simulans	CACCA CGGAA GACGA GTTGA AGAGT GCCTT GGAGG CCCAG GCAGA GATCC CTGCA GCAGT CGTGG CTATC AAGAG CCTTC GGCAA TCGCA GTACG GAGGG AAGAC	TGAGCAA ATCCTGA TCGTCCT CGCGAAG CAATGCT TCACGGC CATAACG CGGGGAG TCCTGGA TTTGCCC TCNTAAA TGTTATT ATGTTGT ATAGTAT ATAGTTT AGAATAA GGGTTCA ATAAACG AATTAAA AAAAA	MMRLIQLNLNHCTAAQDLLVQTVRERRVELA LLSEPYRTPASPDWAFDRTKKAAIWRCsRETQ QITDVLSDIGFVRARVGRWWVYSVYLAPSLTL IAFSRALDRLAADARGRTQVLIAGDFNAWSES WGSSTTNARGRMVLEAFATLDLALLNHGNRH TFRRAGLGsVVDLTFTSGSSYRLARWRLSEDY TGSDHLAIICDLGSPSSQAQTPAQARIKYKTD SLDTQVFREQFLPTASRQGAELTAVELMRRLK AACDASMQSSRPHSRQRAPVYWWNQEIETAR RECLSARRRYQRARGAESFAERQSEYRARRKT LKLAIRESKRKCFLDLCDsADSDPWGSAYKVV VKKAYTRTPKLLDPAMLRsVAEHLFPVMDGL RPADPATGDHVDSDATVSSDEILELAKLLKDG KAPGPDGIPNRALRLALHLQPDSFAKAFTKCL EEGVFPSCWKIQKLLLLPKPGKPPEEPSSFRPIC LIDGTGKLLLEKLCIRLERAIADAGDLsRSQFG FRKARSTVDAVN RVVEVAAQAIDGTRWKGGS KEYCLMVTLDIRNAFNTARWDRILEALIGFSVP AYLVRIIRS YFSDRVLLCESSEG VFRHQVTGGV PQGSVLGPLLWNTMYDGILRLPLVGRSEIVGF ADDVALLVVDKYPGKAEEsCNQNIRAIEHWLS	cd0165 0	473	720

				TGCGG TGATA GCAGT CCCAG CCAAT CTGAC GGACC CGCTG ATCAA GCGTG GCAAG CTGAG GGTAG GATGG TCCCA ATGCT TGATC AAGGA GCTGG AGCCA CGCCA AAGAT GCTTC	SMGLELAPEKTEAVLISSRKAVETATVEVGCT KVSSSRAIKYLGVMIDTRLSFREHLAYANSKA AGVNRALSNIMLNSRGPKQASRLLTSVTRAT MLYAAPVWAKALGTESYAQGLNATHRLSALR VCCAFRTVSNAALVIAGIPPLDLLAKENEVVF NQTHGRRLSPGVKKAIRAEARQHTLETWQHR WTNEPKGRWTQQLIPNLKPWLERKHGDITFHL TQLISGHGCFRSLMRFGHEDADDCPWCGGG RSETAEHVLFSCVKYARERSSLETVLGGRLNA DNLVPFMLQGEAEWQAVNSFAAAITTELRRRA ERARRVHE			
--	--	--	--	---	---	--	--	--

				AGATG					
				CCTGG					
				AAGAA					
				GGCCA					
				CATAG					
				CGGCC					
				CACTG					
				CAGAA					
				GCACC					
				GTGGA					
				CAGAA					
				GCCAG					
				TGCTG					
				CTTCA					
				GATGC					
				GGGAC					
				CGCGG					
				GACAC					
				AAAGC					
				CGCCG					
				AGTGC					
				CCTAA					
				CGAAG					

				CTAAG TGCTT CTTGT GCGCA AGCAG AGGTA GCCAA GCGAC CAACC ACCAA ATGGG CACCC GGAAG TGCCC ATTGG CGGGT AAAGG AGCAC CAAAG GCACT GCA					
Jockey	Jockey-		Drosophila yakuba	GAAAC GTCAA	TAACAAA MGTCGAC	LRRLPHLPRPKPQN NYQPNCPTHLSKSHRNRL YTKCSHSPNPLKWHPC KYDNCCPIYCPPVYK	cd0165 0	479	749

3_D Ya		AAGGA	ACAAACC	WTHENFHKLF SHFETLLDGKSNFILSGDCNSK		
		TAGAT	GCTCCCC	HRCWGNYS SCARGRSLMQAIMANKLLDVVA		
		TTTCA	AAAAATG	TGQATHFPFD RRKQPSALDFAICKGLRRRITRT		
		CAAAA	CAGCTCA	FSSDDLSSDHLPIQIILDLEEANLQLKKVSNSIL		
		TTAGG	TTCTACA	KRNTNMEKF KRILEDKMLLNTEIRTGQDIDDCI		
		ACAAT	AAATAGG	DIFLNNIKTAANEATPKNRPFTNPREN SRRREN		
		WWTG	AATAAAC	NLSLDEATTRLLDEKKELNRILRAAGTEEAKK		
		GACGC	AGGCTAC	WFKSAQNKLAK EIRKLRTSQINNELQLIDTTDR		
		AAAAA	ATGGAAG	YRMQKLWKWTNNVKKQPQPIWPLKKDDDCI		
		GCAAC	ATTCTTA	SMRTASPWTRTLDERAEAFASHLES RFKANLR		
		AATAA	AAATCTA	NSIEDREI IKKELDKFKSLKPRIETGTSGLK PVT		
		CGAAC	CACTAAA	LAELDLLIKSLESKKAPGADNISNKVISNL PKK		
		GGCAG	GCCATGG	ARIYLIL IYNKILNTGHFPPKKWKHACITMILKP		
		CCCCC	ATGCGGG	GKSPSLLNSYRPISLLSGLSKIFERIILNRLYNIES		
		CCCTTT	TTATTTC	FSKAIPNHQFGFRKGHGAEHQLARVTQFILKA		
		CAACA	ACTGAAA	YEDKEVCSATFLDITEAFDRVWHDG LLLKLTK		
		AGACC	CAATAAT	LIPKQLFNVLESYLTGRTFQVRINNETCSRIGTI		
		CCCCA	ACTTAAG	NAGVPQGSILGPVLYSIYTS DMPFPIIKGYIRNIS		
		CCTAC	CGCCAGC	LPDYHPKNIILATYADDTHIISRSNNTRSAININQ		
		CATAT	AAAGTGT	TYINKFCRWAKKWDIAINGKKTGHVLFSLKNI		
		AGTTC	CTTCATC	SENVHTTPIVDGQRAIKLCKHG YLGVM LDRRL		
		CGAAA	TCTATCC	TFGAHLILLKGKIIAAYKRLEWLLGKN SHLPKR		
		TGGCC	TGATGAC	TKLLLWKQIISPIWHY AIAIWGSLTSATQTKKI		

			AAATC	CTGCAAA	QTLENKLIRRIKANRYTKQSAIREKYKIKSFEE			
			TCGGA	ACCCGGA	VFDQASKRYANSIADHENPLIKELLTKSFKPIR			
			AGGAA	CGAAGTA	LAASRTRYVNQYTKYLLPLLQNIXKPPDEPTY			
			GCCGC	TCCCTCC	KSLYSYTDKEELEIVSKIRYDYRQTLPTLLQLE			
			CAGGA	GACGACG	DEEMEARDCKAEEERKEREKEERKKALEKGP			
			AAACT	CATCCCT	PDRWCELLINKFSKMYKKGMITRERVIELMLG			
			ACATC	GCTGACC	QPPTVIKLVIPDYEP			
			AAACC	CCGACGA				
			CGTTC	AGCGCAA				
			CGGCA	GTGAAGG				
			AAATG	ATGTCAT				
			CGAAG	CCACTAG				
			GACGA	GACTGGC				
			TAGGC	AAAGAAG				
			CAAAC	TATCTAG				
			CAAAC	AAGCAGA				
			GGGAT	GGGACAC				
			TCTCTT	GTGAAGG				
			CACAT	ATCAACG				
			GAAGC	TGCGGCA				
			TGCCA	AAAACAT				
			TCCAA	GATGAAT				
			AAACG	CTAAGTA				

				GCTGA	GACTCAT				
				ATAAA	TGCAGCA				
				TGGCG	GCGCATG				
				CCGAC	CATAATG				
				ACGCG	TCACTTA				
				GACAA	CCTGAAC				
				AGCGA	CTTTTGCT				
				CCACA	GCAAATC				
				AAAAA	CCTTGAA				
				CAGGA	ATGTCCT				
				CCAAC	GAAATAA				
				TCGAA	CAGCTGC				
				GGACA	AATCTCC				
				ACCCC	CCGAAAA				
				AAGCC	CCGTACT				
				GCGAC	CAATGAC				
				CCACA	AAAAGAC				
				AAAAC	GACCCGA				
				GACAA	CTTGTCG				
				AAAAC	CAATAGA				
				CCGGC	TAAACTA				
				ACAAA	ACTTTTT				
				GACAC	GGCACTT				

				CTTGA	AGAACAC				
				GGCCT	AAAAAAC				
				TCAAC	ATACAAG				
				GACAA	GTTTTGA				
				ACTCC	TCCAGTC				
				GCAAG	TCTTTATT				
				GAAAG	GGGATTC				
				GAGGG	CCCCACG				
				ATGCT	TCCTCAC				
				ATAGA	AGACCCT				
				ACACA	AGGCGTG				
				CCCGC	TTGCTGC				
				GACGA	CGACCCG				
				AGCTA	GCAAAAG				
				TGACA	ATCGAAA				
				AGCCC	TCGTAAC				
				CAGGA	CCCTGAA				
				GGGGT	ATTGACC				
				GATGC	TTTTTTTT				
				TGGTA	CCCTTAG				
				GCCCG	AATTAAT				
				CCGAC	ATCAACG				
				AACAA	AAATCTT				

				GTAGA	TGTTAAA				
				GCCGC	GATCGAA				
				CAAAA	ATAATTT				
				CTAAT	AATCAAA				
				TTCAA	ACATAAA				
				GACAA	ACAGAGT				
				GGACC	ATTATAT				
				TAAAA	AATAATA				
				ATTAA	TATAAWA				
				CCGCC	AAAAATT				
				GATGG	CACAATG				
				GACCG	TTTTGAA				
				AGGAC	GCCGTCT				
				GTGCA	CAACGTC				
				TAAGC	AAGAAGC				
				ACTTG	CCACTCG				
				CAAAG	ATTTTATT				
				AGGAC	GACCCTT				
				CCCCT	GGCGCGG				
				GACGA	CMCTGCC				
				CCACA	GACCCGA				
				GCCTA	CAAAAGG				
				AACCT	TCAAAAT				

				GGCGC	CGTGACC				
				AACGA	CTTGAAT				
				GTACT	TTCGACT				
				GGAAA	TTTTCCTG				
				TGGAA	GTAGCAA				
				AAGAA	CGATTAT				
				AATTG	ACTGACA				
				ACACA	ATATCCC				
				CTTAT	TGAAAAA				
				GGAAC	GAATAAC				
				TCTTCT	AGATCAA				
				TCAAA	TAAAAGC				
				TGCTT	TTAAAAC				
				AGAAA	TACGCAA				
				ATAAC	CCTACTA				
				AAAAA	CTCCCCT				
				CCTGA	ACTAGGC				
				CCACA	GCAACAT				
				GCACA	AAAACGA				
				CCCCA	AGATTTG				
				TGTCC	GCACCTA				
				TCCTA	AAACAGT				
				AACTT	AAAAACC				

				CTAAG	TAAAATG				
				TTCCTC	TTTGTAT				
				GCAAA	GTCGTTT				
				AATAA	CCATGWT				
				GTAGC	GGGATGC				
				ACCAT	CCCCTCG				
				GACAT	TCCTTCT				
				TAAAC	GAACCCT				
				ATAAA	AGGTGCG				
				CAACA	CCGCTGC				
				CACAT	TGACCCG				
				ACTAT	GCAAAGG				
				CAATA	GTCAATC				
				CCCTT	TCGCGAC				
				AAAAT	CCTTGAA				
				AGGGT	TTCGACT				
				ATTGG	TTTTCCTG				
				AACTC	GCAGCAA				
				ATGTG	CGGTAAA				
				GAGTT	CAGACAA				
				ACCAA	AATCCCT				
				TAAAA	GAAAAAT				
				GAGAC	TATAAAA				

				GAACT	CAAATTA				
				AGAAG	TAAAACA				
				CTTAC	AATTATA				
				ATAAA	AAAGCAT				
				AAAAG	AAAACCA				
				AGGGT	CGACTCG				
				GTCCA	GCGAAAT				
				GATCA	GAAWAA				
				TGCTG	CGACGAT				
				GTAAS	TTTGGCA				
				AGAAA	CCTAAAA				
				CCAAG	CATTAAA				
				CTAGA	AACCTAA				
				ACGAA	AATGTTT				
				ATTCT	TKATGCC				
				ACTAT	GTTTCCA				
				GCTTA	TGCTGGG				
				A	ATGCCCC				
					CCCGCCC				
					TTMTGAA				
					TTCCAAG				
					CGCGGCG				
					CTGCTGA				

					CCCGGCA AAAGGCC AAAATCG TGACCCT TGAATTC GACTTTT CCTGGTA GCAACGG CATACTG ACAATAT CCCTGAA AAAGAAC AATAAAT TAA				
R4	EhR LE2	AB09 7128	Entamoeba histolytica	AGTAA TACAA AGGAA GATAA ATAAA CTAAA AGACA TGCTA ATAAG	TGAACAA TAAGTTG AAAGGTA TTAAACC TTAATTT AATATTT ATTCTTT ATTCTT	MSTTGKKEIRKEIGREEQPSEEDTKEFLNKIYN RGKRIKKENERIKEIIEESNRRNLKRIITKENIE NVITRIPSWKAPGIENMHGYYWKRLESVRDRL VIVFNEWLDPENIPIEMLTGRTILHKGGDSKK AENYRPITCLNVIMKIFTSIINRKIEEQLANNEE KYQINKSQYGNKKRSLAAKEVNINSLLENER EVQKKRYLESFYDIRKAYDTVNDEWLENVLR YFKIPFQITELIKSMSSRWIRIGYRYKEEIREV KIKNGILQGDSISPLLFILQMNIISDVIDRTNEKR	cd0165 0	120	376

			AGGAA	LKVKHVLYMDDIKVTSETKKGMEMVHKNIE			
			ACTAT	TIEVIGMEVNVNKS GVMKKGNIGIQDNMKDIP			
			AAAAA	IVDSEHQYKYLGVYQFTRNDDVGCIERISKSIE			
			ATTAA	EGIKENNKGNDSMNYINRINTEIIPKFTYSASV			
			GAAGG	VKLPIMKLIIDKMIRIEVKEGKIIGAGMCTSRL			
			ATATT	YVNRNKMGDGLKSIRDEYAIKLIRSILYYTWK			
			AAGAA	RGNRIGDMIGINENKNRMLNRLYKSMLKKRIN			
			TAAGA	KEEFRKIVEEERGKRDKNRIIKYVREKFQEDYI			
			AGAAT	NKWRNKKVNSEILKGFEDKTMEQKKTIVWK			
			GATAG	GINVIRKAYIQIKKMEEKARNVGVKALMKH			
			AAGAA	DMRYAICPICNKIATINHILLNCIVTKKTQMNK			
			AAGAT	HDRIGEYIWRSLKIKYQLYENKEKMKMREEEN			
			AAGGG	TKKLIIRKEELGKNERRIEKDKILIVWNHEVVN			
			GAGAA	RSEDKFHKRIDIYVRDLNKKEAMIIDMTIVRDK			
			GTAAA	NISKAFTDKINMYSKLHEHIRKIERLVRVIVIPV			
			TGAAG	VITVSGLINKETVILLNAEGIEIPREIMTREIVISN			
			AAATA	MKDLMRYCGDHSGEYSNVEIPEEEDSQESEGR			
			AAACA	EEILT			
			ATTAG				
			AATAT				
			TTAAA				
			AGAAA				
			TGAGG				

				AAAAA GAAGT AGACT TAAAC AAAAG TGAAA GAACA AGAAT CTACA GAAAC AAAAT C					
RTE	RTE - 2_E L		Esox lucius	CCTCT CCGGG AGCCA TCACC TTATC GTGGT GGAGA GGTTT GTGTG TTCCT ATGAA	GC	MVQNNPMYKRGRGGVTLPGGSPGPPSGARPR GRARRRAPGGRVCHGARSGIARKSNVAPPSPS YGPTTHGRNRWGRVRCHMGGSEGQGPRRTRP GRQGLALGTWNVTSLWGKEPELVREVERYQL DLVGLTSTHSLGSGTVLLDRGWTLFFSGVAQG VRRRAGVGILTSPRLSAAMLEFTPVDERVASL RLRVVGGGISDCCLCICPEQQFGVFGLLDPE WSPVWGSSRGLHNSAGGLQRTRGQFCDRVIG SETACFGHSGKERGGAVNRSPSGGELGQGVGE DSGQTGKTQTGSAGELGTSGGGPRPKDLHSHL RRSFSGIPVEVGGIEPEWSMFKASIAEAAAGSC	cd0165 0	558	813

				CCTGA GGGCT GTGTT GTCCG GGCCT TAGTG CTCCT GGTAG GGTTT CCCAT GGCAA AGTGG TCTCA GGGGA GGAGC CAGAC TAAGA		GLKVLGASRGGNPRTPWWTPVVREAVRLKKE AFRDMLSRRTPETVAVYRQTRRAATSAVKEA KQRVWEEFGEAMEKDFRSAPRCFWKTVRHLR RGKRGTIQAVYSKDGTLTSTEEVIGRWKEHF EELLNPTNTPSIVEAELEADGEASSISMAEVTE VVKQLHSGKAPGIDEIRPEMLKALGVEGMSW MTRLFNIAWESGKVPKEWRTGVVVPLFKKGD QRVCANYRGITLLSLPGKVYSKVLERRVRQIV EPQIEEEQCGFRPGRGTTDQLFTLARILEGAW YAHVPVYMC FVDLEKAYDRVPREILWEVLREY GVRRSLLRAIQSLYVQSESCVRVLGSKSDSFQV GVGLRQGCALSPILFVTFMDRISRRSRGGEGLO FGGLGISSLLFADDVVLMASSVCDLQHSLDRF AAECEAVGMRISTSKSEAMVLSRKPMCELLRV GNEALPQVKEFKYLGVLFASEGTMEIREGRRI GAAGAVLHSLYRTVVTKRELSRKAKLSIYRSIF LPTLTYGHEGWVMTERTRSRVQAAEMGFLRR VAGFSLRDRVRSSAIREELGVEPLLLCIERSQLR WFGHLVRMPPGRLPREVFQARPAGRRPRGRP RTRWRDYISTLAWERLGIPQSELANV			
CRE	CRE -	.	Folsomia candida	AATTT TTCAA AAAAA	TAAAGAA TTTCATT ATATATC	MMTPTSTTAANLSTHDDQHCPSTCPGASNTD QWILCDHCQKWYHTACVAVVIIPEEEWLCPTC VISVTSSQSQVTPSSLNTTTTTKKKHIIATTQY	IPR000 477	404	656

3_Fo Ca	AAAAT	TATTGTA	WLKSNLVEDPLGLIPKSEIVARYAHHIQTSGEG		
	AAAAC	GTTATGC	RVVDATVGKYILQLFPSTKFSRPPTRGAANAYI		
	GCCCA	TAATATT	GIRWKDQPPLLPHAPRNWGAELSQLKNTVPTL		
	AAAAT	ATGGATT	RRVPKGARMEVAKALAKIIEEVTTTNSEESWF		
	ATAGA	ACAATTA	KLFSFPYAVVPVPTKSDKVKNLTTWVKSRVLQ		
	GAATG	TGTATTA	WSITQIVTPSRPRPQFAKPSRPQETIAKKVEAKL		
	AA	TGATAAA	ADGDVRGAIRLLTSDDTIAPNNQETLDALITKH		
		GTA	PPHPEPTYFPNAPDKAEVLAPVDPKEVLIAISSF		
			APGSAGGLDSLRPQILKDLVISENGEPATNLLN		
			AMLKFLDLVLRGDVPKSICPIFFGASMTALRK		
			KTGGIRPVAVGNTWRRLLISKIAVARMSSTLVEI		
			FSPHQLGVGKGGAEAAHAARIYWNYSHDSP		
			KAFCLKLDFRNAFNEIRRDTLQIIREKFPNFYRF		
			VSQAYSSPSEIFFAGTPISSLCGVQQGDPLGPAL		
			YALVIQPIVDLIDTELNMWYLDDSTIADSPEKV		
			LMAMQTITRMGTEVGLHLNTTKCEVGILGVV		
			DDQTRDEIFELFRSTAPGIQKISPESAMLLGVPL		
			TDEAIPTILGAKTEEMEKISTRLLKLSSHSAFFL		
			LRASISTPRLIYFLRCAPTWRRFDALTLYDAKL		
			KNAMEGIINCSLSPQSWLQSSLPVNKGGLGVR		
			HAIDVAIPCFLASIYSSLPLAERLLPPHLHGIDT		
			AIQEGEERWKS LGSSPLPPLEIRGIQAIWEIPQQ		
			EGVINHLTTNATTLEDKARYHAMSQPNVGAW		

						LNAIPSPQLGTHLPNETFRIAAALRLGCDICQPH KCSCGEQVSTKGYHGLSCKNSAGRHSRHAAA NDVIARALRSAGVPIIKEPAGCAREDGKRPDGL TLIPWARGKSVVWDFTCSDTFAPSYLPQTSIRP GVAAEKAEDRKKKRYEQLMDSFLFVPIAIETT GIWGKDGLSFIKEIGSRITNITEEKRATSFLLQR MSITVQRGNVASILGSLPKGHKLEEIFML			
Rex 1	REX 1- 1_G A		Gasteroste us aculeatus	CTTCC GTGGA ATTAC TGGAC ATTCT AGTCA AAGGT GCGCT CACCT TTCGT CACGC GGTGA GACGC CGTAG GAGAG GAAAA	TGACTAT AACACTC CACCGGT CACTCCC CCTCATA CTGCACA TGCCACT TTAACTG CAATTCA TCACTTT GTCGTCA CTCGTCA CTTTGTCT CTTGTCT GTTACTT GTTCGTT	MRRDRDFSTSSVLCFTETWLCGSIPDSALQLA GFQLLRADRDTELSGKTKGGGICFYFNNSWCN DVTVILQH CSPDLETFFINCKPFYSPREFASFIL VGVYMPPAGNVHEAQRTLADQIRRVERTNPD SLVIVLGDFNKGNLTHELPKYRQFIKCPTREEN TLDHCYT TVSRAYHAVPRAALGHSDHVMVHL IPAYRQKLKLCKPAVREFKKWTSEALEDLRAC FDCTDWDVFRTATDSLDEFTEAVTSYIGFCED SCVPSCTRVSYNNDKPWFTAELRTLRLQKDQA FRSGDKDLYTEAKYKFSKAVRDAKRLYSEKL QQQFSASDSASVWRGLKHLTNYKPKTPHSMN DLRLADELNEFYCRFERQCPDPIPHSSTNLLQS PSPSPSPSGAHASSSISPSPPPPATTL SILERDV NRLFRRLNPRKAAGPDSVSPHSLKHCADQLSP VFTDIFNTSLETCHVPACFKASTIIPVPPKKPRITG LNDYRPVALTSVVMKSFERLVLTHLKSLTDPL	cd0165 0	472	725

			CGGGC	AGTGCAC	LDPLQFAYRANRSVDDAVNMALHYILQHLDS			
			CGGCT	TTTATGC	PGTYARILFVDFSSAFNTIIPSLQLDKLSQLHVP			
			CGCTG	TTAATAT	DSTCKWITDFLSDRKQHVKLKGKHSVASRTIST			
			GTGAG	TTTTCTTT	GSPQGCVLSPLLFSLYTNSCTSSHQSVKLLKFA			
			ACTCC	TTAACTT	DDTTLIGLISGGDESAYRWESDHLVSWCSQNN			
			GCAGG	TTTAATA	LELNALKTVEMVVDFFRRNRAPPSPITLCDSPVT			
			CGTGG	TTTAAAA	IVDSFRFLGSIITQDLKWELNISSITKKAQQRLF			
			TTCTC	TTTTTTAA	FLRQLKKFNLPKTMMVHFYTAIIESILCSSITV			
			GCACT	CTTTATTC	WYAAATAKDKGRLQRVIRSAERVIGCNLPSLQ			
			CCGCT	CCTTGTTT	DLFASRSLKRAKKIAADPSHPGQKLFVPLPSGR			
			GCCAG	TATACTA	RLRSIRTKTSRHTNSFFPSAVGLINRARSPTA			
			GCATC	ACCCATA				
			TTTCTC	GCCTTAG				
			TCCAA	CCTTATT				
			CGTGC	CTACTAA				
			GCTCA	CCCATTG				
			CTGTG	CATTAGC				
			CAACA	ATTTCAT				
			AACTG	TCCACTT				
			GACGA	TATTTTAT				
			AATTC	TACTTGT				
			AGCTG	GCACTGT				
			CTG	TGTCTTG				

					TCTGTCT ACTGTCG CGCACTA ACCGCCA AGACAAA TTCCTTGT ATGTTTG ACATATT TTGGCTA ATAAATG TTCCTG ATTCCTG ATTCCTG A				
RTE	RTE - 1_G A	.	Gasteroste us aculeatus	CCTCT CTGGG AACCG TCACC TTATC GTGGT GGAGA GGTTT GTGTG	TAAGCGG TTGAAGA TGGATGG ATGGATG GATGGAT GGAT	MKKRKREGVTRPGGSPGPPFGARPRWRARQR ASGGRVYRGARPGTARRSDAAQHSPTHGLT TCGGNRWGRVRCYKGGSDSGGSRRNRPGRQ MLALGTWNVTSLGGKEPELVREVERYQLDLV GLTSTHSLGSGTKLLDRGWTLFYSGVAHGVR RQAGVGILISPRLSAATLEFTPVDKRVASLRLR VVGKTLTVVCAYAPNSSEYSAFLETNGVL HGAPVGDSIVLLGDFNAHVGDGDGNWRGVIG RNGLPDLNPSGCLLLDFCASHGLSITNTMFEHK	cd0165 0	616	870

			TCCCT	DVHKCTWYQSTLGQRSMIDFVIVSSDLRPHVL		
			ATGAA	DTRVKRGAELSTDHHLVVSWSVRGWGKTRDR		
			CCTGA	PGKPKRVVRVNWERLEEAPVQKIFNSHLRRSF		
			GGGCT	SHIPVEVGDIPEWSMFKTSIAEAAAVSCGLKV		
			GTGTT	LGASRGGNPRTTPWWTPVVREAVRLKKEAFRV		
			GTCGG	MISGGTPEAVAVYRQARRAAASAVMEAKQRV		
			GAGCC	WEEFGVTMEKDFRSAPKCFWKTIRHLRRGKR		
			TTGTG	GTIQAVYSKDGTLLTATEEVIGRWKEHFEELL		
			CTCCT	NPTTTPSLVEAELEAEEGSSSISLVEVTEVVKQ		
			GGTAG	LRSGKAPGIDEIRPEMLKAMGVGGLSWMTRLF		
			GGTTA	NIWKSQTVPKQWQIGVVVPLFKKGDQRVCA		
			CCCTT	NYRGITLLSLPGKVYSKVLERRVRPIVEPRIEEE		
			GGCAA	QCGFRPGRGTTDQLFTLSRIIEGAWEYAQPVY		
			AGTGG	MCFVDLEKAYDRVPREILWEVLREYGVRGSL		
			TCTCA	LGAIQSLYAQSESCVRVLGSKSKAFPVGVLRL		
			GGCGA	QGCALSPILFVVFMDRISRRSRGEEGLQFGGLR		
			GGGGC	ISSLLFADDVVLMASSVCDLQLSLERFAVECEA		
			CAGAC	VGMRISTSKSEAMVLSRKPMDCLLQVGNVSLP		
			TAAGA	QVKEFKYLGVLFTSEGKMECEFGRRIGAAGAV		
			ATGGT	LHSLYRTVVTKRELSRKAKLSIYRSIFVPTLTY		
			TCAAG	GHEGWVMTERTRSRVQAAEMGFLRRVAGVS		
			ACCCC	LRDRVRSSAIREELGVEPLLLCVERSQLRWFG		
			C	HLVRMPPGRLPREVFQARPAGKRPRGRPRTR		

						WRDYISALAWERLGIPQSELVDVAREREVWG PLLELLPPRPDPG			
CR1	CR1 - 1_G av		Gavialis gangeticus	AGCTT AGTTG CCTGT ACACT AACKC CAKAA GCATG SGAAA CAAAC GTACT GACCA ACAAG CAACA CGACC AGCTC CACCT MTGA	TGATCCC AGCAGTC TTTCCTG CCCACGT GCAGGGG GGCTGGA CCTGATG ATCTGTA AGGTCCC TTCCAGC CCCTAAC AACTATG AAA	LDXRHXGIXXVQEGSYWKXGGVALYVKEHY TSLXVKIGCKEGXLETLWVKIRGERREEXLIVG VYYRPLTKNSLENWQRLHGIMGXFNYPDISW DEYSARSDRSQSFLTCIEELYLTQEVYEPTRGK ALLDLVLAKGDDLVSNLRIXGKLGNSDHEMIT FSIHCKPGKSASRVEIPDFRKADFHKLRLVGX GLRSNEEGNTRRVVFPQGSNPKSSKKVYSLSQ ERQQKSWETPLAQQGSHGPPEIKARSLQGMEV RNNYQRGVLSSCPHLKSLILGKLSTNSGDWL VRDYGPMRKGIQEEWSFLKEAILKAQRKSIPS HKKGSKRAGRPPWLNREVMDLLKLKREAYK GWKSGTTTKEEYSAAARTCREQIRTAKALTEH KLATRIKDNKKSFFRYVGSCKRNKGSIGPLINE MGRLLTDTQDKTELLNEYFMSVFHQSKGISLL DRIQDDHGGGDHPPIVDVDLVGEHLQRLNIYK SAGPDELHPRVLKELADVIAQPMAKLFEDSWH SGEVPEDWKKANVVPIFKKGRREDPGNYRPIS LTSIPGKILEKIIKETITHRLTEGSILKGSQHGF MGRSCQTNLISFYDQVTCHLDRGNEVDVIYLD FKKAFDLVSHDVLMMVKLGKCGLDYFMVWWL GNWLCGRTQQRVLIDGTESSWQRTSGIPQGSV	cd0165 0	521	770

						LGPVLFNMFINDLDSGITSNLAKFADDTKLWG SVVTLKDRLAIQADLDRLMKWADLNLMSENT EKCKVLHLGRNNPQHTYRLGSALLASTTTERD LGVMTDNKMNMESHQCDVAVAGRANQTLACIH RGILSKAKDVILPLYSALVRPQLEYCVQFWVL HFKKDVEKLKRVQRRATCMIRGLETKTYEERL RDLGLFSLEKRRLRGDLVAAYKYIRGVHQGL GEHLFTKAPQGRTRNNGHKLIEDRFRLDIRKN FFTSPVPRVWNRLPPEVVQSSTLAIFKKRLDAH LAGII			
CR1	CR1 - 7_G av		Gavialis gangeticus	CC	TGACCCC AGTGCTC TTCCTGC CATGGCA GGGGGTC GGACTAG ATGATCT GCTCAGG TCCCTTC CGACCCT AGAAACT ATGAAAC TATGAAA	LVRGLRVLDPLGDSHCLLEFTIQRKVAKASS QVVALDFRRANFNELRRIVGEALKTQRGGEPG VQGEWSLLKETILQAQRKVIPRRVKGGRRAYK PPWLTKNIRECLLAKKEAYSQWKGGVITKEDY SAVARGCRGMVRKAKAEMEQLASRIKDNK KSFFKYIGSRKKAPGNVGPLQDKTGNLVTPD DKANIFNFFASVFLSKDQNCPLPAGPPLGPG GGSPGPRVSEDLVRELLEGLDVFKSAGPDDLH PRVLKELAGVIASPLARLYEHSWCSGEVPEDW KRANVVPIFKKGRKEDPGNYRPVSLTSILGKLF EKIILAHIQEGPAGNGMLRGNQHGFISGRSCQT NLVAFYDQVTKSLDAGVAVDIVFLDFRKAFTD VSHPIIRKLGDCGVDTHTVQWVANWLESRT	cd0165 0	288	539

						QRVVVDGSFSTWKDVGSGVPQGSVLGPALFNI FISDLDEGVQSTLFKFADDTKMWGDVGTIEGS NRLQLDLRLQGWAEENRMGFNTDKCKVMH LGRNNRHHTYRLGNSPLVSTEVEKDLGVIIDS KMNMSRQCGEAARKANCTLSCIHRCISSRSKE VILPLYAALVRPQLEYCVQFWAPHFKRDVDSI ERVQRRATRMIGQQGRSYEERLRDLNLFSLQ KRRLRGDLVACYKLVRGDQQALGESLFPRAP QGVTRNNGHKLAEGRFKLDIRRRYFTVRAAR VWNQLPSEVVQAPTLGVFKRRLDDHLTGVI			
CR1	CR1 - 17_ Gav		Gavialis gangeticus	ATACA GGACC TCCAC CTAAC TCAGG AGGTG CACAG TCCCA CCAGG GGAAA TGCCC TGCTG GACCT	TGACCCC AGCATCC ATTCCTG CCCAAGG CAGGGGG TCGGACT TGATGAT CTGTTTA GGTCCCT TCCGACC CTAGCTA CTATGAT ACTATGA	MGVQDGWSFLKETILRAQRESVPMHTKGGKS AKKPPWLSKGIQECLRAKKEAYIRWKQGAIK EDHTSVARGCRAAVRKAKAEMELGLVTRIKD NKKSFYKYIGSKKAAPGNVGPLQDALGNLVIA ANEKADLFNKFFASIFLSRDRDISPTGSTGLR GGTARPKVRDDVVRELLEGLDVFKSAGPDAL HPRVLRELAVIAGPLARLYEHSWSSGQVPDD WKKANVVPIYKKGRKEDPGNYRPVSLTSVLG KIFEKIIKEHICEGPAGEIMLRGNQHGFIKGRSC QTNLISFYDQVTKSLDEGVGVVDVIFLDFKKAF DTVSHPIIKKLGDGDAYTVRWVANWLDG RTQRVVVDGSFSTWRDVGSGVPQGSVLGPVL FNIFINDLDEGVSTLSKFADDTKMWGEVGTP	cd0165 0	225	476

			GGTCC	EGRDRIQLDLRLQRWADENRMEFNADKCRV		
			TTGCC	LHLGRRNQHTYRLGNTLLISTEAERDLGVIV		
			ACAGG	DSRLNMSRQCEEAVSKANRTLSCIHRCITSRSR		
			CGATG	EVILPLYAALVRPQLEYCVQFWAPHFKRDVAS		
			ACCTG	IERVQRRATRMIRGQQGRPYEERLEGLNLFSL		
			GTGAG	HKRRLRGDLVAVYKFTRGDQQGIGDALFPRA		
			GGAAT	PHGVTRNNGHKLIESRFRLDIRRHYFMVRVAR		
			TGCGG	TWNQLPREVVLAPTLGVFKRRLDRHLTGVI		
			GTTCT			
			CGGCC			
			ACCTG			
			GGCGA			
			TAGCG			
			ATCAT			
			CGCCT			
			GCTGG			
			AATTC			
			ACTAT			
			CCAGC			
			GCAGG			
			GTGTC			
			AAGGG			
			CTAGC			

				AGTAA GGCGG AAGTC CTTGA CTTCA GGAGG GCTGA CTTCA ATGAG TTAAG AAGAC TAGTA TGAGA GGCAT TGGA TCCCG AAGTT TTGGG GAG					
L1	L1- 5_G Ar	.	Gossypiu m arboreum	GACTT CGTTT GTGTG TCTGC	TAGACCG AACGGTC TTGTAAT AATGTAG	MKFMCWNCRGLGNPAKLRELKQLIAANNPDL IFLSETKMSANDFRRVQNKCRMQNGHAVNSE GRSGGLALMWRDGMNVSIQNYSKHHIDSMV KLENNDNIRVTGFYGHANPNLRSSSWDILRRV	cd0165 0	498	764

			CTATT	CCCATT	GESVREEWIVGGDFNAILNDAEKEGGRRGVR			
			ATTCTT	AATTTGT	AQMNEFKEVMDDLALVDIKPDSGWFTWVNN			
			TGAAT	AAGACTA	RRGGVTWGGLIKERIDRFISSVSVAEKFPFISTK			
			TTTAT	TGTGTAG	VVRQTQSDHDAIILDLWGRKPKAYPTDKRLCF			
			GGCAG	TGAGGCT	KFDVCWAGDVEAKNVIERAWNRAATDFGEKI			
			TCTTGT	AAAAATT	ERVRLALGPWQRKKYGNLKRDMQKLEKHID			
			GGTGT	GTTTTTTT	WIIDSVSSEESGRTLKETRRRLDFLYAREESYW			
			TGAGT	GATATGG	AQRSRSIWLREGDRNTRYFHAKATGRLKKNVI			
			GTTTCT	ATTATCC	EKLQNAEGQWVTNSEDKVAKDYFVKLFRS			
			TTGTT	TTCGGAA	NGQNANTQEMGFIKDCVTRETNEWLCMDYTE			
			GTTTTT	ATCCATG	KEVLQAIKQMDPNKAPGIDGLSGSFFKHHWDI			
			GGTAT	ATGCTGT	VGKDTISYCLEVLNGNKNISSVNETMIILPKIK			
			TATTTT	TATTAGT	NPCELTNFRPISLCRFVYKIISKVLANRLKAALP			
			ATTTTT	GATGTTA	SCISQNQSAFVPGRMIHDNVLIAHELLHYLQSS			
			TTTTGT	CTTAATA	KNGPNKGLVVKLDMSKAYDRVEWNFIENVM			
			GTCTTT	ATATGTG	KKMGFVDKWIAKIMECVRSVTYTVKCNNTLS			
			CTTCG	TTGACCC	DIFVPERGLRQGDPLSPYLFLFCMEAFSRMLIH			
			GTGGG	TTGTGGT	AQENNTLRGIRASRNGPRINHLFFADDALLFVR			
			TTTGC	CGTTTTT	NKRSDIECLVTMLDIFSNISGQEINFDKSMILFS			
			GGTGC	GCTCAAA	PNTPRAQRILFSDLLGMTVVEKLNNYLGLPIPI			
			CGGGG	AAAAA	GKKKTEAFKEITNRMSCRINSWTKRLLSFEGK			
			GACCG		EVFIKAVLQSIPTYAMSIFLAPKGIIDDIQAMLS			
			ACGTG		RTWWSGKDKGGFWTMIPWKTLCCKPKGMGGL			

			ACATG	GFRDVRLFNLALLGRQVWRLLNNTDSLCKV		
			TTGCT	LSSKYFPDGNIFHAKKVDKASFTWSSIAKAVE		
			GTTGA	AFQDGFQWQVGNGEKINIWADNWGMEGLNG		
			AGTCT	AAVISNTLNHSEKFVKDLWHLDRRSWDENKV		
			GGCGT	KQAYGQDWGDKICDIPIGDEVQEDKIIWFHNT		
			TAGGT	HGCFTSKSAYSWLLLKEMGCGPHRFFWRALW		
			GGGGA	KLDTLPKVRVFTWRVGHEILPTNGKIASIRPGF		
			TTTTCA	NKGCPRCGAEVETLLHALRDCPTSRAVLSIGG		
			GAGGG	WSRSFISKSYDHCIDWLEDLMRVLDKRAMAD		
			TTTCTT	LMTILWNCWNNRNNFIFRGKEEEAQLIWERAS		
			GGCTC	NLSKEFRICNMLHAPLISQNIAEKRWVKPPKGF		
			TGTGA	LKINFDTVGETGIGYGTVIRDEEGFVLGGGG		
			CGTTC	GFKEGRRSVEEAECVAFEQSIDVACKLNIKDH		
			GTTGC	VLFETDHAGLVNRINNLANVTHIGARLKDCT		
			TTTGC	VAFTFFKSASLI		
			ATGGC			
			CAAGG			
			AGTTC			
			TTGCG			
			ACTTT			
			ATGCG			
			CTTTA			
			GCGGT			

				GGCGG					
				GCCAT					
				TCGAG					
				GTGGG					
				ACGGG					
				CGATA					
				CATAC					
				GTGGC					
				TCGTC					
				ACGAT					
				CTTTTC					
				TAGCA					
				TCGTC					
				AGACC					
				CTGTTT					
				TGATG					
				GGATT					
				GTTTC					
				ATTGG					
				CAGCA					
				ATAAG					
				GTTTT					
				GTCGT					

				TTCTTT					
				TGTCTT					
				GTTCA					
				GCTAT					
				TTTATT					
				GATTT					
				TCTTTA					
				CGTTC					
				AGTAG					
				GTTGC					
				ATGGG					
				TCTTG					
				CATGT					
				ATCTT					
				GCATG					
				GATTG					
				TTTCGT					
				CTTGC					
				ACGTA					
				TTCCA					
				CTGGT					
				GGTAT					
				TTCTTA					

				TTTTCT					
				TTGAC					
				TTTTCA					
				ACGTT					
				ATTCA					
				ATCTG					
				AGGGT					
				TTTTTT					
				GGTTT					
				AACAG					
				GCATT					
				ATATG					
				AATGC					
				GGAGG					
				ATAAG					
				TTTTA					
				ATGAT					
				TTTCCC					
				TATTT					
				GTTTCG					
				ATTTA					
				GTTGT					
				TTTTGT					

				TACAG					
				TAGCT					
				ATTAG					
				GTTTCT					
				GGGTT					
				TTCTTG					
				GACTG					
				CTGGT					
				GTTAG					
				GTTGC					
				AGTTT					
				TTTGG					
				TACTTT					
				CTTGC					
				ATGTT					
				TTCCTT					
				TCCTT					
				GATGT					
				TCGGG					
				GATGG					
				GGGTT					
				TGATG					
				GGGGG					

				TGTTT					
				GTTTCG					
				TAGCT					
				TGAGT					
				CGGGG					
				TGTTC					
				CTGGC					
				TGTTG					
				TTGTTT					
				TGCTG					
				TGTTTT					
				TGCAG					
				GTTCTT					
				GGACT					
				GTTTCT					
				TTGCT					
				CTGCT					
				CCCGG					
				TGGCG					
				GTTTTC					
				GGGGT					
				TCTGTT					
				TTTCTC					

				TATTCT					
				TGGGG					
				TGTTTC					
				TTGTC					
				GAGGT					
				TGGGG					
				GAGGC					
				GACAA					
				GAGAA					
				AAGTG					
				TTGAT					
				CAGTC					
				TCCCT					
				GGCAC					
				TCGTTT					
				AAGGT					
				AGCAG					
				TCTCG					
				TGAGT					
				TGTCG					
				TTTGA					
				AA					

R1	R1- 1_H Sal	Harpegnat hos saltator	CGCCG	TAGCAGG	MQRMSELGAGLGVVAEPYWVPVRSVNWFGS	cd0165 0	474	707
			CGCTC	GAGGGGG	PDGSVAMVANPAAGTPPCVFRENGSSYVAVD			
			KGCGG	GTAAGGA	WGPITVVG VYLP HHRNKAGLGELAPVKRVLD			
			CCGTC	GGGAGGG	RIGELVRRCHPRPVVIAGDFNAHSEGWGCSPR			
			ACCAT	TGGATGT	QRDPRGEAVIGWAAGLG LLLMNRGSTSTCVR			
			AACGT	CCCCTCC	PGGQSVIDLTWATPSAAGLFQEWRVEAEGESL			
			GCCCC	CTCCTTG	SDHRHIAWALRLPKPARVLARCPGEGNFPLPK			
			GACGG	ACAGGAG	RWALNRLDTRDFRAAVLAATWHLDVGPGEW			
			AAAAT	ACACCAG	DPSPAEGAGEIVGIVASACDAAMPQSHPRRRR			
			ACCAT	AAGGCCA	ATYWWTEELENLRDSSVQARRRYRARRRGG			
			GAGTT	ACCGGCG	DVAGTKEGYLRAREALKAAIVEAKRAAWEQ			
			CATGG	GGGCAAC	MVSSLDDDPWGRPYKRVMGKIRPWAPPPTES			
			TCGAG	CCGCCGG	LDRQALEGLLDALFPRAEGGPPEIPTPPLGED			
			GCCAG	GGACCGC	WDEGLAVSPEELTAARRRLGAKGKAPGPDGIP			
			GCGCC	GCCCTGG	GRAWALALGEGNLSAATCRVLSGCLREGVFP			
			GAATC	ATTCCCT	PEWRRAKLVLLPKEGKAPGSASAYRPICLLDE			
			AGACC	TCGTGGA	AGKMLERIVADRLVQHLSSKGPNLHQRQYGF			
			GGCGG	ACAGGGC	RPGRSTLDAIQCFRDLTRAVVEEEGGVLLAVS			
			ACCTS	GCGGTAC	LDITNAFNTLPWPEIGRALEHHGVPVYLRRILS			
			GGGAT	ACAGGTG	AYFGGRDLAFPVEGGVQGRRAMERGVPPQGSV			
			AAACT	GCGAGGG	LGPLLWDIAFDRVLRTPLPPGCHVVCYADDTL			
			CTTCC	GCTCCCG	VVAGGRAGGKPASERRRRHWHRWWEPSSETWV			
			CTGAG	GAAGGGA	SRWRPKRPRPCSFTAGVGGRRRRRPGSSPGGEL			

			GGTCC	GCCCTGG	GGSPPPSGDGIGIGGGNHQRPGSQGGAPKDPG			
			GCCGG	GGCCGGG	RVLSRQGSGBAAGDPGPRRWRPRRRADDEIS			
			GCCCT	GGGTTGG	RPDAGQPMELRTPLQGTGPQGAQGRTGCSQP			
			GAACG	GGAGGGA	DAHPRGPRLEGSPVRHGGALDRPVRGAHMG			
			GGGCA	ATACGGG	ATAGGLPRRKARAAAGPEARANQGDSGVPIR			
			ATGCT	AGTAGTT	VVHGGDHPGRRPPCGAPRRGEAPPLLEGEGAP			
			GATAG	CCCCTGA	RGGEPVGQGPEGSEAPGRGPDGPGEVEGLAVRP			
			AGGTG	CTCCCCT	PGARASDGRGCPPLPVRVAGRRGRGLSFRLAQ			
			CCTGG	CCTCACG	VLTGHGCFGRYLRRIGKELTTRCHHCPALEDT			
			CCTCA	GTCCGAC	AHHTLAACPAWAGERRVLAGAVGCREGEJSL			
			ACGTC	AGGTAGA	PRMVKAMCGSEEAWAAVVSFCEAVMSQKET			
			GGGGC	CATCCCG	AEWGRRPHRARGAAPGGRGTRQRARGGAGR			
			GTCCG	CAGGGGT	GGLPDPRGPGAPSAARARHRSQGGSRGRRAA			
			CCGAC	TTTAGTC	FPAFPPPSDRTSRDRGERCPGWGWARSPTNPP			
			CGGCT	GGTAAGA	LPD			
			TGCGG	GTCCGAC				
			AGGAG	ATAACCC				
			CTCCG	CCTCCCC				
			CAAGC	CCCAAGG				
			TGGCG	GGGGGGG				
			GCGGA	TATCCAT				
			AAAAG	GATGGAT				
			GCCCC	TTCCCCT				

				GGTTA	GCGTGAC				
				TCGCA	CAAAAAA				
				TCCAG	AA				
				CGGCC					
				GGCAA					
				AGTTC					
				GCAGC					
				ACTGC					
				GAATT					
				GCTGG					
				GCTGG					
				ATGCG					
				ACGAC					
				CGGGG					
				CCGAA					
				GAGGT					
				GGCGG					
				CCGCG					
				GTGGC					
				CCTCG					
				CTGGG					
				GGATG					
				CTCGC					

				CTGAG					
				GAAGT					
				TTCCTC					
				GGGCG					
				TGATG					
				AGCGT					
				CTCCC					
				AGCGG					
				GGAGT					
				CGGCG					
				CCATG					
				GTGGT					
				GCGGT					
				GTCCG					
				CTGGC					
				GGCGG					
				CCGCT					
				AAGGT					
				CGCCG					
				CCGCG					
				GAGCG					
				GGTTC					
				CAGTG					

				GGCTG					
				GACCC					
				GCGCC					
				GGGGT					
				GACGG					
				CGCTG					
				CCCGC					
				TCGCG					
				CAACG					
				CTGTG					
				CTTCC					
				GCTGC					
				TTAGA					
				GCGGG					
				GGCAC					
				GTGCG					
				CGAGC					
				GGTGC					
				ACCAG					
				CGCCG					
				TCGAC					
				CGCAG					
				CGGTA					

				TGTGC					
				TACCG					
				CTGCG					
				GTAAC					
				CCCGG					
				CCACC					
				GCGCC					
				AAGGG					
				GTGCC					
				AGGAG					
				GCGGT					
				GCGCT					
				GCCCG					
				GTGTG					
				CGCCG					
				AGGCG					
				GGCAG					
				GCCCT					
				CCGGT					
				CACAA					
				GATCG					
				GAGGA					
				CCTGC					

				CTGCA					
				AACCC					
				CCTCC					
				CGGAA					
				GGAGA					
				AAGGG					
				GGGAA					
				AGAAG					
				AAGGA					
				AGGAG					
				GAGAG					
				AGGCA					
				GAAGG					
				GGAAG					
				GAGGA					
				CACCG					
				GCGGG					
				GGGAG					
				GCCCG					
				AACCA					
				CCACC					
				GCCGG					
				GGGAG					

				GACCG				
				GACCC				
				GACGA				
				AGCCT				
				GGGGA				
				GGAGA				
				SCGGA				
				CCAGT				
				CTCCT				
				CCCCG				
				CCGAG				
				GGACT				
				GAGGG				
				GGAGC				
				CGGGG				
				CGGAT				
				CCGGC				
				CCCCC				
				CCTGG				
				CAAGA				
				ACCCC				
				AGAGG				
				AAAAG				

				GGGAG					
				GAGGG					
				GGACT					
				CCTCC					
				CCCGA					
				CCTAA					
				TAGAC					
				TGGGA					
				CTGGG					
				AGCCG					
				GAGAG					
				CCCCG					
				ATCCA					
				GGAGC					
				CGGAG					
				GAACC					
				TCTGC					
				CCCGC					
				AGGAA					
				GAGGA					
				GAGGG					
				GAAGA					
				AGAGG					

				CCAAC					
				AGGGA					
				CGAGG					
				ACGAC					
				CCGGA					
				GGAGG					
				ASGCC					
				CAGGC					
				CGCAA					
				AGGTG					
				AGGAG					
				GGTGG					
				AGCCC					
				TCGGG					
				CACAC					
				CCCCA					
				ACAGG					
				TGTAG					
				AGGAA					
				GGGGA					
				GTCGG					
				TGGGA					
				GCGCT					

				GCCCC					
				CGACT					
				CCCCT					
				GCAGG					
				TTGAC					
				TCCCC					
				CTCGT					
				ACCCC					
				CGAGA					
				GGAGG					
				AGGTC					
				ACGCA					
				GGTAG					
				GGACC					
				CCCCG					
				TGTCT					
				CCGCG					
				GAGAC					
				ACCCG					
				CGTTG					
				GAGGC					
				CACCA					
				TGGCC					

				GAAAC					
				CCCCT					
				CCGGG					
				GAGAC					
				CGAGA					
				CGGCT					
				TGCCG					
				TCGGA					
				AGTCT					
				CCCAG					
				GAGGA					
				GGGGG					
				TGGAC					
				GTAGG					
				CAGCC					
				CGGAG					
				GAGGC					
				GTGTA					
				AGGCT					
				TAGCC					
				TTCAC					
				GACAC					
				TCCGG					

				GCCCT ACAGG CCAAC CTCGG CCACG GCGCC CGGGC TCAGG ACCTC TGC					
Dap hne	Dap hne- 14_ HMa		Hydra vulgaris	GTGTA TATAA AAACC TTCAA ATATT TTCAA AATAA CAAAA ACATA CAAA	TATTTCA AGTTATT TGATATT TTTGTCG ACTGTGT GATGAAT ATTATTT AAACAAT ATATGCT GAATACT TTATATT ATAATAT TGTTCAA	MDSTYRNDFEKISNLFQTNTFNLDNESDPDINY FNEINGSKFEC SYFYDPDKIKSIFKNNDNRDYFNI IHINIRSLKKNFDFNFLNMICQTENFFNVICLTET WCTEEDFKSNSNLHLTGfENTVFLERNVNKRG GGVLFYIKENLAYNIRSDMSVSDDNIEILTVEII NKKSKNILLSCCYRPPTGRTEYLSNYLVHNIIE KTSQEKKKNYILGDFNLDCFQYNEKKYITKFY NHLFETGTIPIINKPTRITKSSSSLIDNVLTNDYF NITLKKGIIKTDVSDHFPIFFCLGADHKTNNG KITLKKRIYSNNNLNLFKEQLLHQDWSYINFNE DINIIYKSFFNIFYKIYEANFPIREIVLNAKDISCP WITKGIKKSSKIKQKLYIKYLKNKSDKHKNNY TVYKNLFEKLRKKAKQIYYSDLLIKHKHNSRR	cd0165 0	570	829

					TATTGAG	VWQIMKEITSKHKTNTNTMPNAIKFENKLITDS			
					AATGTTA	NEIAAEFNNFFSSIGHNLSNKIPYVDNNSVDEFI			
					TATATCC	SSPNSTINFSNLTYNEFEIAFKSLKRNKAIGPDD			
					ATATTAT	INGNIVIDSFNEIKDILFKVFKASITQGSFPNSLK			
					AAAATAT	IAKVTPIFKTGDHTNITNYRPISVLPVFSKILERI			
					TATGTTA	MYNRIYTYLIDNKLLYKNQFGFQRNSSTEHAIL			
					AAACTTG	QVTRSIADSFKNSQFTLGIFIDLSKAFDTVNHHI			
					CTTAACA	LIKKLESYGIKDKTLNWFKSYLSNRKQFVYNK			
					TCTGAAG	ESSSQIKNITCGVPQGSILGPLLFLLYINDLYKA			
					AATATTC	STLTTVNFADDTNLFQSHENIEILFNNMNNELK			
					TAACGCT	RISYWFRKNKLSLNTDKTKFILFHPTSKKKRIQ			
					TATGTAA	NILPDLFIDNTIIKREKFTKFLGVIIDENLSWSQQ			
					TTTTTAAT	IDSISMKVAKNIGVLYKARYILNKKQLTQLYY			
					ATTTGTA	SFIHCHINYANIIWGSAHKTKLKSLYRHQKHA			
					AATTTTC	ARIINFGNRFSSSTPLLIEMKALNIYELNIFNILC			
					TACCTTTT	FMFKCKENLSPSVFINLYNLKPKNKYELRDSK			
					AATTCGT	NIQEPFCKTKIDQFCISYRGPFLLWNKIVLPNFDF			
					TATATAG	SFKWSYSSFKKKTKKEIIFSTQNIFIYF			
					ATTGTAA				
					AGGGCTT				
					CATGATA				
					AGATCAT				
					GTGATCT				

					TCTGGAA GTCCTAC CAGTAAT ATATTCA ATGTAAA TTTACCTT G TTCATA TTTTTCTG GAAAAAA AAA				
CR1	CR1 - 53_ HM		Hydra vulgaris	CGCGC GCCGT TAAAG ATGGC TGTTAAT AGACG TGACT TTGTG AAAAA ATAAG AATAA CAAAA CATAA ACATT	TCTCTAT GACTATT ATGAAAC TGTTAAT AGCAATG GCTCATA ATAGATT CAACATG GCGTTGA ATCTCAA GGAGCAT AAATAAA	MVRKEDFEALVSRVLKLEAELQIKDKEINDLKI VVNDMKNQSKLVSNSNDVSNTWSDVVKGRK KNTSVQLDLLNVVVKESHGREIRQKNLVVFGF KTSTNKDESVAKEDDKQELQNLFKQINVKVTI QRFVKLNTKTDKPPPVIIVLTNKEERNAVLKT AKNLRNSPETNNIFINADMTEAERFKMKALRE ERAKLNEANQDLKYYYGIRNERVVKLLKKLY ESTQPYQQISPKKVLXNLNNITKNPNDRHLL NSNKRLIVYYTNSTSLNNKKNDFRALIDIEKPD VFAVTETWYTPLSDVNIEGYVIYRNDRLNGKK GGGVCIYVNSSIKSYEISNIELKSDLIEQIWTWI EYGQEKVLLGCIYRPPDSSCVTNKNIINSIIAAK RAQDKKQFTSIITGDFNYNTLKWKDSVEESH	cd0165 0	757	1007

			TTGTA	TAAATAA	QQAEDFVACLNDCCFEQCVTEATFQSNENELT			
			ATAAT	ATTATTT	NILDLVIVDDSNRIFGLSHLPPLGLSSKGHHILK			
			AACTA		WFYKMKDRIDDFSSSTKKFAYRRGNYEKINV			
			AATAA		YIKSINWADEFQNKCIDKCYDKFLYHYNQSCK			
			TAAAA		NYIPIVKLKNKASEVWMSNSLRKEIKYKNNY			
			GATAA		WTTNKKMNWSIAGSKEKYNKMKRNILKQINN			
			ACTTA		SVKNYERNLVNQVKNEPKLLYSYIRKKQKVK			
			TCTTGT		NHIRALNDNIGSTMYDKHSIVDIFNKYFHSVYI			
			CAAAT		KEENINDCNFEVQTNSTRFGKIFVDQALVKSYS			
			TTGAC		KLDINKSSGYDEVNPYVLKVCSEFSIPLSIYQ			
			A		QSLTNSDIPDLWKMANVSPIFKNGSKLQAINY			
					RPVSLTSIPCKVLEKIISDNLKNYLESNNLFCQ			
					QHGFLKGKNCITNLLESDFLTQMLMIGNSIDI			
					LYTDFQKAFDRVPHKRLISKLCGYGIEKEMLN			
					WIETYLSNRKQQRVIIGDTKSDWLEVLSGVPQG			
					SVLGPLLFLLYINDLPISKIKNKCELYADDNKIIA			
					VVNNQVDSKSLQDDINNLTEWSDKWMINFNI			
					KKCKIMHFGKHKNKFEYFIKDSKLDSTTNEKD			
					IGLLISNTMKWNQHVNMMVVKANSRLGLLSK			
					SFTYKDKNTMKILYCTFVRPILEYASPIWNPYN			
					KNDIHKLEIVQQRATKLIPELRHLPYEERLKKL			
					QLTSLKIRRLRYDLIQYYKIFHNLDKVCWYQQ			
					PKTAISISANGPASNIRGHKARVDIELIKRNRPR			

						ENFFTNRVAKWWNKLPEEIIISAKSVNIFKDKL DKHVMIMKL			
CR1	CR1 - 7_H M		Hydra vulgaris	AAACG TTATC ATCTA TGGAC TAGAA GCAAC CCAAT GTTCTT CTCAG ACGGA AAAAC TAGAA AATGA TAAAT CGCAA GTAAA TAATT TACTT AGTGC TATCG AATGC	AACATTA AACCGAT ACAACAC GCGATCT ACCGGAA GATTTAA AAAACCT TTCCGTT CATCAAA ACTTACG AGCTTTG CCATTTC CTATCGT GGCCCAA TGCTATG GAACCAT TTTACAA GTCTGAG TGGTAAA CTTAAAG AAGCCAA	MMPIINKSTRITQHSCSAIDNIFTNTLFHSTFKS GIIRSDITDHPVFAIFKKPKDLQKKDWYFKKR IFTTRAKNMFKLELSTVNWNFVFLETCAANSY DLFLDKFFNLYDKHFPNKIIRIKTKGLLNPWFT KSLLKSSKRKQKLYIRFINSKKERDKHNYLEH KRLYERTKKIARTTYFKNKLLELKGNCCKKTW DAIKEILGKQKIHNKALPSVITVDGTQISDVKS VATKFNEFFVNIGPNLASKIKHTPAHFLSYLSK SDKSLAFKKLTSKELDDSIKKLNKNKATGYDA ISASVIKDCYELIKDVLLDIFNKSLKTGVFPDK MKVAIITPIFKSGDESLVTNYRPISVLPVFSKILE RIMYNRIFAFLYDNNLLYYNQYGFRSNYSTEH ALMQLVEVISEASSHNKFTLGIFLDLSKAFTDV NHQILLEKLKYYGISHEYHKWFTSYLTDRKQFI SLNDGFQTNKLKISCGVPQGSILGPLLFLIYVN DLYKASNTMSTIMFADDTTLFYSHSNIISLFQV ANNELEQISNWLANSLSLNADKTKFSLFHSKR KFQSIPKSLPMLYINNINIERSTNKLLGILIDEN LSWRDHINHISNKISRSIGILYKTRQILDKSTLT QLYFAFINSYLSYGNIVWGSTNKSCLDKLLKQ	cd0165 0	330	590

			GATAG	TACCTCT	QNHAARIHFQNRFTNAKPLLKSMRALDIYEL			
			AGCAA	CAAAAAT	NIFQNYTIYVSTSVKKNAEGFRKYFP			
			AAGTG	TTAAAAA				
			AGATC	GAACATA				
			TATTC	ATCGAAC				
			GCCGT	TTATTGC				
			CTTAA	CAATCAC				
			ATCTA	GATTCTT				
			AAGGA	TACATGA				
			ACTAT	TTTTTTCT				
			CGATC	AACAATA				
			CTTTG	CACGATT				
			CTTAT	ATTCCA				
			CGAGC	ATTTTAT				
			TTGAA	TATCAAA				
			GACGT	GGTTACT				
			TAATA	GATGAAA				
			ACAAA	AGACTTT				
			AATAT	ATCTTCTT				
			GGTTC	CAGGTTC				
			TTAAA	TCCGGTC				
			GCATC	TTGTAAC				
			AATTA	TATTATA				

				AATTA	AATAATG				
				AGAAA	AAATGTA				
				AATCA	CTTGTAT				
				AAAAA	TTAATAT				
				TATGA	TTCTAAA				
				AAATG	TATTTTA				
				TGTAT	CGGAAAT				
				ATAAA	TATCTTT				
				TCGCG	AATTGTA				
				ATCTA	AATGGTT				
				ACAGC	GCTTGAC				
				AAGTC	AAAAATT				
				AAAGA	GAAATAA				
				TTCATT	AAAAAAA				
				CTAAA	AA				
				AAATA					
				AATTA					
				ATAGA					
				ACGAA					
				ACGAA					
				AAAAA					
				TGCGT					
				TAAGA					

				AGTGA					
				ACAAG					
				AAAAA					
				ATTAA					
				ATTTC					
				ATTAC					
				GGTAT					
				GACAA					
				TCTTA					
				CTCGA					
				ATTTT					
				CACAA					
				ATAGC					
				CACTA					
				GCAAC					
				CACTA					
				ATTTA					
				GCAAC					
				GTCAT					
				CAAAA					
				GTATC					
				TCCCC					
				CAAAA					

				AAGAC					
				AAATA					
				TTGAC					
				TATAC					
				TCGAA					
				TAAAA					
				TCTCA					
				ATTTA					
				TTCAA					
				GATCC					
				ACTTC					
				TCAAC					
				ATAAA					
				CAAAC					
				TGAAC					
				ACTTC					
				TTACTT					
				TGATT					
				CGGAT					
				AAAGC					
				CGCTA					
				TTTTCC					
				TTAAA					

				AGATG					
				TGAAT					
				TATAT					
				TTTTCT					
				AATCA					
				ATTCT					
				CAGTA					
				TTTCAT					
				TTAAA					
				TATTC					
				GTAGT					
				ATTAC					
				TAAAA					
				ACTTC					
				GACGA					
				ACTAA					
				AAGTC					
				CTACT					
				AGATA					
				ATCTT					
				AACTT					
				TAACT					
				TTAGC					

				TTAAT					
				TTGTTT					
				AACGG					
				AAAGC					
				TGGAT					
				TTTAG					
				ATACA					
				ACATT					
				TAATC					
				AAAAT					
				TCTAA					
				ATTTA					
				TATTG					
				CCCAA					
				CTACA					
				AAGCC					
				ATTAG					
				TTTTG					
				AACGT					
				AAGTC					
				TTCAA					
				AAACA					
				GGCGG					

				TGGTA					
				TATGT					
				GTCTA					
				TGTGC					
				TTGAT					
				GAATT					
				TACTT					
				ATCAA					
				ATTAG					
				ATATG					
				ACCTG					
				TCGTT					
				CTCTG					
				AGCCG					
				GATTT					
				TGAAT					
				CTTTC					
				ACTTT					
				AGAAA					
				TCTAC					
				AATAA					
				GTCAT					
				CTAAA					

				AATAT					
				TATTG					
				TGACA					
				ACTAC					
				ATACC					
				GACCA					
				CCAAA					
				CGGAT					
				GTTTA					
				AATAA					
				CTTTC					
				AAAAT					
				CACTT					
				AAGTG					
				ATTAT					
				TTTGA					
				AACAG					
				CATGC					
				AAAAA					
				AAATA					
				AAGAC					
				CACTC					
				ATTTT					

				ATTTCT TGGAG ATTAT AACGT CGACT ATTTA AAATA TGGTG AAAAT AATAA AGTAC AACAA TTTTAC GATAG CTTTA ATGAG TTTGG A					
CR1	CR1 - 5_H M		Hydra vulgaris	AGAAG CTGGA ATGTT TGC GT ATATT	AAACAAT TTAAAGG TTAAAC TAAATCA CTAAAA	MENNLSYNINAKNIVFSDNNSDNHTETLLQSE YYSIPESTQYLCENKNKFSILHINIRSMNKNFEN FKCLLGELNHDFKIICFTETWLKSNETNSNFEL RNYTSVHQMRDICVGGGISIYVHNSIDFIQRKD LNVNNTDCEALCVEIINKLAKNFIINAIYRTPA	cd0165 0	551	810

			TCATA	ATAAATT	GSLKTFKTYLRTFLTTKNILQKHVYVVG DINID		
			CGACA	AAATCTT	LLNHALNSEAKTFIDILLEYNLIPTINKATRVTK		
			AGTTG	TAACACT	KSSTLLDNVITNNFHNSRFKTGIIKTDLTDHFPI		
			GTGGT	TTCTTTA	FFITESVTLNNAATHKSTVFMQRQINESSICQFKNL		
			TCGCG	GCTTTTA	LNNYVDWNLVLQSHDVNNAYDLFLNQFSKM		
			ATGGT	TTAAAAT	YDKAFPLKVKLINSKSVXSPWMTKGLLKSSRK		
			CTCGA	GGTTTCA	KQKLYDKYLKNKTYKNETNYKNYKNLFEKT		
			AAACC	AACCATT	KKQSKVNYYYAKLLEKNKGNPQKTWSVIRDLI		
			CACGT	AGTATAA	GKNKIEKNNLPQKLIIEGKMIYHKEVIIKLN NF		
			AAATA	AAGATAA	FLDVGPNLAAKIPIGQKKFDSYLATTDLIMEEPI		
			TTTCTT	CGAAAAT	LTKSELHTAFNSLKKNK SAGIDQIN VNVIKSVF		
			TTAAC	AAAATTG	DIIEPSLFHIFNLSLKS GHIPDKLKI AKITPIFKSG		
			GTTAC	TCTTTTTC	DETNISNYRPISVLPCFSK LIERIMYNRLY KYLS		
			TTTATT	G TTCATA	ENKILYNNQYGFKKNHSTDHAIXELVKHISNG		
			TTTATT	TTTTTGAT	FNSDCYTIGVFIDLSKAFDTVNHEILLKKLENY		
			TTTTAT	GATATTA	GVKNQSLWFKSYLTNRKQFILKESKDSKNNL		
			TTATC	AATAATC	ITCGVPQGSILGPLLFLLYINDLYLSSKVLNTILF		
			AATAG	AGCCCAA	ADDTNLFYSHKDITVLFKIFNEELDKINEWFIS		
			TTTTTC	ATACTTA	NRLSLNVEKTKFILFHKPSKAENIPLKLPNLLIN		
			AAATT	TGTTTAC	NKIIKRESTTNFLGVLLDEKLSWK FHIKYIEGKI		
			TTAAT	ATATATT	SKNIAMLYRTKPFLNNE SLKNLYFSFIHSYLSY		
			TTAAC	GCTTTTTT	CNISWGSTNHAKLKKIYSKQKHACRIVFGANR		
				TTTTTTTT	KTQCEPLL RQLGALNIYKLN IHQVMIFMFKAK		

				TTTTA	AATTAAA	LGFSPVTFQSYFNEISHKYPTKFSVNNFVVPINF			
				GAA	CAACACA	LKLSSYQIQYRGPLVWKEFLPIFSKKQNFVLNC			
					TTTTTTTT	TLQSFKDDSKRHLLNMEFDILDFKYL			
					GTGTGTT				
					TTTATAA				
					CTTTCAT				
					GACTTTT				
					TGTCGTT				
					TTTACTTT				
					TTTATTTT				
					TAAAATT				
					ATTCTGC				
					TTGAAGA				
					TTGTTAG				
					CAACAAA				
					TTATTTTA				
					TTGCTGA				
					AGTTTTT				
					GATTTGT				
					TTTTGTTT				
					TTGTATTT				
					ATACTGT				
					CTTTTTGT				

					CTTTTATC				
					ATTTATT				
					GTGCATG				
					TAAATTT				
					TTGTGTG				
					ACGGGGC				
					TCGGTGA				
					TAAGACA				
					ACCTGTC				
					TTCTTCTT				
					GCTCCGG				
					CCAGGAT				
					TTTCTTTT				
					ATTGTAA				
					TTTCTTTT				
					TTGTAAT				
					ATTTATC				
					AACGGCA				
					AATTAAA				
					TAAATAA				
					TTAAAAA				
					AAAA				

CR1	CR1 - 42_ HM	Hydra vulgaris	TGCGC	ATATTTTC	MKNKTLCIFCFKTVTDNHKAIYCDCCSSWVH	cd0165 0	640	894
			GATGA	TACCAAC	AKCNQIDLLTYNLLSEDSSEWYCFNCISKMP			
			AGGAA	TAATATT	FGLLTDNELLTTLSCKNPQTASSNFLVSPHSR			
			AGATG	GACATTC	NILKELNTISGSAINCKFYDVVDENKAIKPNSDI			
			GCGGT	TTTACTA	YLHLNIASLPYHIDDLCSLISTLNIPMAIGITES			
			CTWAA	TATATGT	NLYLNDTNITDITINGFNIEHCPTEAKKGGALL			
			TTCTG	CAATTCG	YLQSNLNYVVRHDLKIYLPKLLESVFVEIHKPH			
			AGTTG	TAATCCT	KTNVLIGCIYRHPSMNQNEFNTTFLTPLLEKLS			
			AAGTT	TCAGTTT	KEQKLIFLMGDFNVDLLSYSESNPSTSNYLDTL			
			TGGTA	TTAATTT	SYSFNPSIILPTRITTKSQTLIDNIFMNFYSSDHL			
			TACTTT	GATATTT	SGNLTISISDHTAQFVCVPGTSNQPKNVKISRC			
			TATTTT	TTAACAT	LKRCLKKFDPNKFHIEISKINWELIHKDDNINES			
			AWCGA	TTTGAAA	INNFLKSFNRLVDQCAPYKTLTKNQIKLKSKP			
			GTCGA	TTCTCTA	WISKGLLKSSISVKNKLYKKFVRCKNINTKNELF			
			AATTT	CTTATCT	TKFKFYRNKISNLLKSSKKSYYITYFNNNINNI			
			ATATT	ATCTACA	KNTWKGIDINIKGSLHKTPTQLNLNENIITDH			
			TATCC	TCTAATT	ATVSNIFNNYFVSIREKLLNNIMPSEYTYSDFL			
			ATCTA	TCTCTGG	NFPNANSFFIDPVSEDEVSSLISNTLKNDKSYGP			
			TTATT	TTTTCTAT	NSIPTFFLKLTSIIISKPLSIMINNSFKNGKFPDV			
			AAATC	CTGCCAT	FKVAQVIPIYKAGSLDLSNYRPISSLNLKSLF			
			AATGA	TTTTAAA	EKAMHNRLYKFLDKFKCLYQHGYGFRSKHST			
			AAAAG	CTTTATT	TLALIEITEKIRRALDNKHFACGVFVDLQKAFD			
			ATAAA	ATTAAAC	TVDHSILLTKLEYYGIRGIPLKWFTSYLKNRTQ			

			AAATT	TATCTAT	FVSINSTKSDLQNCSNGVPQGSVLGPLLFLFIN			
			AATAA	ACTAATA	DLNTSFKFATAYHFADDTNMLLVDKSLKKINK			
			AGTAA	TTATTAA	HINHDLGNLVQWLRSNKLSLNSNKTELIIFKSN			
			TTAGT	TTTTAGT	LAPIKKQLNFRLSGQKIYPLDSIKYLGKIDSNL			
			ATCAA	ACCAATG	SFQGHNLNDLAMKLSRSNGMLAKVRHYVNLET			
			AATAA	GGTTGGC	LLNIYHAIFGSHLRYACQVWGQSHKQIILRLSH			
			TTTATT	ATTCTTT	LQNKALKIIYFQYNNNSKSNVLYLISKILKISDLV			
			CGTGA	ATTTATC	VFLNCLFVWNYHHKNLPLSFNNFFSITNCHYP			
			TCAAT	TATCTAT	LRSISSLNLVIPKHQSAKYGNKSIKYQCISSWN			
			ATTAA	TTATCTCT	KLPPDLKSLNSISEFKIKLFNYLLKKYN			
			TACAA	TTTCATTC				
			AGTAA	AAAAATT				
			CCCTA	CTCTATTT				
			AAGTA	CTCAACT				
			GCAAA	ATTCTAC				
			CAAAA	TTCGTAC				
			AAACG	TAAATCT				
			AGCCA	TTGTTGTT				
			TATTTC	GCCATCA				
			ATTAT	CTCTTTA				
			CCATC	CTAAATT				
			AAACA	TTTGTTG				
			AGAAG	CTGCTGT				

				AAAAG	TATTGTC				
				AAGCA	TTTATTTA				
				AGAAG	TGAAATT				
				AAAAG	CTTCTCTT				
				TATAA	ATTATAA				
				CATAA	TTGTTATT				
				ATAAA	ATTTCGT				
				ATGAG	TACTATT				
				GAGAA	ATTATTA				
				GAACC	TTGTTAT				
				GAAAA	AATTATT				
				AAAAA	ATTTTGT				
				ATAGC	TGTTGTT				
				ACGAA	GTTGCTA				
				TCTATT	TTATTATT				
				GGAAA	CTTATCA				
				GCATA	CAATCAT				
				AGATT	TGTTATT				
				TAAAG	ATTGTTA				
				ATACT	TTATTGTT				
				TAAAC	ATTATTA				
				CTCCTT	TTTTAYT				
				ATTTT	GTTATTA				

				TTCTTT	TTGTTATT				
				TGATG	ACTATTA				
				ATTTTT	TTATTGTT				
				GGTTA	ATTATTG				
				GCCAC	TTATTAC				
				ATTTG	TATTATT				
				TTTATT	GTTATTA				
				ATTTTT	TTGCTAT				
				CTGGA	TATTAGT				
				CTATTT	ATTATTA				
				ATCAA	ATAACAG				
				TATTT	TTATTATT				
				GTAGG	ATTA				
				TTATTT					
				ATTTTC					
				ATTAC					
				TTTAT					
				AGAAT					
				TATAA					
				TTATA					
				TCTAA					
				ATAGC					
				TAATA					

				TATCT					
				GTAAT					
				ATAAT					
				TTATTC					
				ATATA					
				TATAA					
				AATAT					
				TATTA					
				TATAC					
				CTATA					
				TAACT					
				ATCTA					
				TATAT					
				CTATA					
				TAGCT					
				TTTATT					
				AGTGA					
				TTATA					
				TAAAT					
				GTCAT					
				ATAAC					
				TTAAT					
				CTTTC					

				AGTTT					
				ATGTT					
				AATTA					
				TATAT					
				TTACA					
				TAACT					
				ATCAT					
				TATTT					
				ACATA					
				ACTAT					
				CACTA					
				TTTAC					
				ATAAC					
				TATCA					
				TTATTT					
				ATCAT					
				TTGAA					
				CTTTAT					
				TTATT					
				ATTTG					
				TACAT					
				TTCTCT					
				TCATTT					

				ATTCTT					
				TTCTTT					
				GTGCT					
				TGTTTT					
				AACAT					
				TTTATT					
				TGTAT					
				ATTTA					
				TTTTTA					
				TTTATT					
				CTTTTC					
				TTCAT					
				GATTA					
				TTTTA					
				ACATT					
				CTTCC					
				CAAGT					
				ATATT					
				ATTTT					
				ATTAT					
				TATAT					
				TGGCT					
				TGTTA					

				GTGTT					
				TTTGTT					
				AGTTG					
				TCAAC					
				ATCTA					
				TTTAA					
				TCTCTT					
				CAACA					
				TCATC					
				TACTT					
				GATTT					
				TGTAT					
				AACCT					
				ACCTG					
				TACTG					
				TGTTA					
				TATAT					
				ACTGT					
				TCTTAT					
				GCAAA					
				TCCAT					
				TTTGC					
				CTCAC					

				AATTA					
				CTCTCT					
				TATTA					
				TATTA					
				TTATA					
				TTATT					
				ATAAG					
				TTTATC					
				TTATTT					
				ATTAT					
				ATCCT					
				TAGTC					
				ACTCC					
				TTTTTT					
				TTTCAT					
				CACCT					
				TCTGT					
				GTTTA					
				TCAAT					
				AATAA					
				AACAT					
				AATAA					
				CATTC					

				TGCCT ATCAA GTT					
CR1	CR1 - 20_ HM		Hydra vulgaris	AACAT ACAAA ATGGA TAACG ATAAA GTAAT TACAA TGGA ATGCT TCGAG AACTC ATGTC AATAC AAGGA GATAC AATTT TCAAA TGCTTT AAGGA AAGTT	CATATAT TTTATAG TTTACAG TATTTGTT TTCAGTA TTTTTTTA TTTTTTTA TTTCTTCT TCTTCTTA TTTTTCTT GTTTTTA ATATTTTT TGTTTTTG AATTTGT TACACTA AAAAAAT ATGAGCT TTGTTTTT ATTGAAT TTTTATTT	MSENSLYNLVSYKMISQPRVNSKGGGVAVYIL DKFEFKKNNTFSKTNKHIETLCIEVLNTSGKSF LISSLYRPPCGKQVNFLNILKXFMSQMSKQNK HVFFAGDTNMDALSYEFANIKNFFNLLFEFNI IPTIFKPTRITNQSSSVIDNILTNC SLNTKFESGLF IADFS DHFPVFHIAREVLSNIDNKKCFVTYRNL SIKKLKKNLNNFKNKLQVEQWDKVYSSTDPNI AFDHFSSTFQNI FDNVCPKITTEIKNKEYKNPW MTITLIKASKRKQKLYVKFLKSKNKEDEKIYK NYKKFFQHSMKEEKIRYFSNQLDKHKFDIKRT WAVINELTGREKKKHTCTCLPQKLISDKKNYN KYNKNMSRIFPIKKIITSTTKICQEFNKYFVNVG PSLASNIVQPNKTFDNFLGEKPTTSINNEKITKL EFNKAIHELKRNKSCGYDGISSNVAIYIMDSISK PLFYLISSSFENSIFPNKLLAKINPILKKGDCN NVSNYRPISLLPVLSKVFERIINRVNNYFTLN NLLYKNQFGFQSKCSTEHAIIKLADKIFKSFDS DELVLGVFVDLSKAFD TVDHSILLSKLKYYGII GSTFKWIQSYLSNRKQFV LLEKSGALDITCGVP QGSILGPLLFLIYINDMHKASQKISTIMYADDT	cd0165 0	484	751

			CGGTC	TATTTT	NLFFNHPBAKTLFETMNIELEKFNLWFIANKLS			
			AGTGT	AAAATTA	LNCQKTSFTLFHKKKNNQLISLKKKQSINLPLK			
			CCTGG	AATTGTT	LPKLLINKNEIKREKCTKFLGVLFDENLSWRN			
			AAAAA	TTTATAA	HIGLLETKMSSVIGVLYKSRPF LDSKSRNMLYF			
			GTTGA	TGAGTGA	SLVHSQ LLYANIVWASTHKTKLLKLQSLQNHA			
			GAACT	CTATAGG	CKVINYLNR LVNPAPVMKSMMVLDIEKLNTL			
			TAGCC	GGCTCTA	QILIFMYKYKNNLLPSVFS DIFLISFSEKYNLRS			
			AAAGR	TGAAAAG	NTRYNYHLGKIKSHQSNFLITNRGPSIWNHFQ			
			CATTG	GTTGTGA	NKNIKDLKSISSF KRET KMHLLNS			
			AAGAA	TGTTAAA				
			ATAAA	CACATCA				
			ACGTG	TCCTTAC				
			GGCTT	CTTCTTTG				
			ACTTTT	AGTCCCT				
			GTTGG	GACTGAT				
			TGATC	TAGAACA				
			AAATT	GAATATA				
			GAGAC	CATAATA				
			AAAGT	TAAAATA				
			AAGCA	AAAAATT				
			AGGTT	ATGAAAA				
			GATAG	AACGTTT				
			TAAAA	TTTATCA				

				TGGCA	TGTTTAT				
				GAAAT	CTTTACG				
				TAAAG	ACGTGTA				
				AAACT	TTTTATGT				
				ATTGT	ATTTTTTA				
				GTATT	CATGATT				
				AAATC	ATATGAA				
				ATTCC	AATTGTT				
				ACGGA	ATTGTAA				
				AAATT	CTGTTTA				
				AAGCA	ATCAGCA				
				TTGAA	AAATAAA				
				AATAA	AGTWTAT				
				AAKAA	ATAATAA				
				AAACT	TAATAA				
				GTCGA					
				TCTTG					
				AAACC					
				GACGT					
				AACAA					
				TTTAA					
				GAATT					
				GTTGG					

				AATCA					
				AAGAA					
				GAAWA					
				TAACG					
				AATCT					
				TTTAA					
				TGATG					
				TTGTA					
				RAAAA					
				AGTAA					
				ACCTA					
				CTAAT					
				TAAAG					
				AAGAT					
				CTGGA					
				AATTG					
				CGGGT					
				CYAGT					
				TATTA					
				TAGAA					
				AGAGC					
				RCATC					
				GTGTC					

				GGTAA					
				ATTTA					
				GTGGA					
				AAGCC					
				TAGAA					
				CTATT					
				GTCTG					
				CAAAA					
				TATTA					
				AACTG					
				GCAGG					
				ACAAG					
				GAAAA					
				CATCT					
				TATCC					
				AACGC					
				TCGTA					
				AACTC					
				AAGGG					
				GAAAA					
				ATATA					
				TATAT					
				TAACG					

				AAGAC					
				TATAG					
				TGATG					
				AAACA					
				ATGCG					
				AATTC					
				GTAAA					
				GAACT					
				TTTTGT					
				ACAAG					
				CYAAA					
				ATACA					
				TAGTG					
				GTAAC					
				GGAAA					
				ATACG					
				CGAAA					
				GTCAT					
				TTACA					
				GTCGA					
				TTAGT					
				GGTAA					
				GAGAA					

				ATGAT					
				AAACG					
				AGAAC					
				GAAGG					
				TGTTG					
				AAAGY					
				AAGTA					
				ATTTA					
				TTTATT					
				TACGT					
				TTTAA					
				TAGTT					
				TTATT					
				AGATA					
				TTAAA					
				TTTTA					
				WGGTT					
				TTTATT					
				TAAGG					
				ATAAT					
				ATTAA					
				CAATG					
				AATAT					

				RTNTA					
				AAACA					
				ATTCA					
				TAAYA					
				ATGAA					
				GCAAA					
				TTTAA					
				CAAAA					
				TTAAT					
				ATCTC					
				CAATT					
				TTTGA					
				AATTT					
				TGAAA					
				TATAA					
				CAAAA					
				ATGAT					
				TAYAA					
				TTTTTT					
				GACTG					
				AAATA					
				GATAT					
				TAAWT					

				GCTTA					
				AATAA					
				TGAAT					
				ACTAT					
				AAACC					
				GGAAG					
				AATTT					
				AAGAG					
				AAAAA					
				AAGCT					
				TAACA					
				AAGAT					
				TTAAA					
				CATTTT					
				AAGTA					
				TAAAT					
				ATACG					
				GTCAA					
				TGAAC					
				ATTTTT					
				ATTAT					
				TCTTTA					
				AATTA					

				TCTTTT TGACG TAATA TGTAT ATCAG AAACT TGGGA AGATG CAAAC TTGCC C					
CR1	CR1 - 3_H M		Hydra vulgaris	AGGTC ATCGC AAATG GCGGA CGTAT TTTTCT AACGG TAAAA AATAA ATAAA AGAAA AAGGA	ATCTTAT ACCTTTTT ACAACCTT ATTATTT ATAACCC TAACGCA GTGCTTC TATTCAT CTTTTAA TTTTTCTC ATTCTA TCGTGAT	MKICHHCKFCNKKIKDNHRAIQDCCSFWIHA KCIPVDKHAYNMLSNDSSSEWFCPNCISNNMPF GVLSDELKSTLSCKNPSNKKIKFHEPPDYLNK FKTMNNVFNPTTKCKYLDVNELNKSINPGTEI YLHLNISSLPFHINDLSSFVGTINTPPMVGITES NLYANDSNITDITIQQYINIEHCPTESKKGGALL YLNSNLNYIVRSDLQIYATKFLESIFVEVIYPLK SNTIFGCIYRHPSLNITEFLSIHFNPLLEKLSHEA KNIVLMGDFNIDLLNYRESQVISNYFESLCSHS LFPTIILPTRVTAKTKTLIDNIFMNSFPTDIVSGN LTISISDHMAQFVCIPNNPPIKKKVKMFKRSFK KFDSDFSFIQEISDINWELLIKDDDNINESINFFLK	cd0165 0	634	890

			TAAGA	ATATTAC	TFEKILDRYAPYKELTKKQIKLQSKPWITSGIL			
			TAAAC	TAATTTT	KSISNKNILYKKFTRSKNVNTKNILFTKFKLYR			
			AGTTA	AATATCA	NKISNLLRYSKKLYYASFFNNNINNVKNTWKG			
			ATTTT	ATTATTA	IKEIINIKPSTSTKSFNLKINENVISDHTSVANIFN			
			CATTTT	AATTTAT	NYFTTIQDELLKNTVPSKCAFGNFLKIPSINSFFI			
			CAATT	TATTATA	EPVTENEVANLIKDTLKNNKSLGPNSLPTFLLK			
			CGCTA	AACTGAT	LVSHIISKPLCTMMNNSFKNGIFPEAFKVAKVI			
			AAAAG	ATGTTTC	PIHKMGSYLDYSNYRPISLLSNLSKLFEKAMFQ			
			TTTTT	TATAGTA	RLEHFLEKHKFIYKHQYGFRNKHSTTHSLIEIT			
			TAAAT	TTCTCTTA	EKIRQAIDNKHFA CGVFIDIRKAFDTVEHTILLE			
			TGTTA	TCACTGT	KLKHYGIRGIPFLWFSSYLCNRTQFVSINGINSG			
			AAAAT	TGTTGTT	LAKSFNGVPQGSVLGPLLFLIFINDLNVSLKFST			
			CTGAT	ATTATTA	AYHFADDTNLLLINKSLKKLNKNMNHDLANV			
			TTAAA	CTATATT	VQWLRSNKLSLNSKKTEIIFKSAKTKINKQLN			
			TCAAA	TTATATC	FRLSGQKINPVNSIKYLGKIDSNLSFASHLQDL			
			CGCTT	GTAATGA	ALKLSRSNGILAKIRHFVNHETLLNLYHAIFHS			
			AAAAA	ATATTAT	HLRYACQVWGQSKQLAFLRLTYLQNKALKLI			
			TAAAA	TGTAATT	YFQHTNSNCSILYFLSKVLKLCDLIQLSNCLFV			
			ATCAA	GTTATTA	WNQNHNNLPLTFINFFSYRENCKYILRSALNFK			
			ATAAA	TTGTTGTT	LSVPKYRTVHYGYESIQHKSQITWNNLPSQLK			
			AAAAA	GTTATTA	SLKSFSKFKTALFNHFLEKYSL			
			GAACT	TTTTTGTT				
			TCGAA	GTTATTA				

				TTATT	TTGCTAT				
				AAAAA	TATTATT				
				AAAAG	GYTATTA				
				AACAC	TTATTATT				
				TTTAA	ATTATTA				
				ACAAA	TTATTATT				
				GAAAA	ATTATTA				
				GAACG	TTATTATT				
				AAGAA	ATTATTA				
				AAAAA	TTATTAT				
				AAAAA	AGTATAT				
				AAAAG	TCTTTTTA				
				AAAGA	TAATCGC				
				AAGAA	TACTATT				
				AGAGA	ATTTTTTT				
				AAGTA	TATCACT				
				AACAA	ATAATCA				
				AGAAG	CTACTAT				
				AAGAA	TACTTTT				
				AAAAA	ATTATTC				
				GAAGA	CTACTTG				
				AATAA	TATTATA				
				GTAAA	ATTATTA				

			CAAAC	TTATTAC			
			AAGAA	TATCTCC			
			AGAGA	ATCTTAT			
			AGAAC	TTTTATTT			
			AAAAC	ATTTTGA			
			CAATT	CAGGTGT			
			TTAAA	TTACTTT			
			GAGAA	ATATTAG			
			TAAAG	TACAAAT			
			AAGAT	TACTATT			
			TGGTT	TGTAAAT			
			ATTCA	ATCCTTG			
			TTTATT	ATGTAAC			
			TTCGC	AACTTTA			
			GATTT	TTCAGAA			
			AATTT	TATATWT			
			GTAAT	GATTTGA			
			TATTA	TTT			
			TTTTTT				
			AATAT				
			AATTA				
			TTTAA				
			CTTATT				

				TGTTA					
				TTTTAT					
				TATTA					
				TTATT					
				ATTAT					
				TTTTTT					
				TTTTTT					
				TCACT					
				TTTCA					
				CATAT					
				TTTTTA					
				ATTAT					
				TTATTT					
				ATCAT					
				TAATT					
				ATATT					
				CATTA					
				GTTCT					
				AGTTA					
				ATAAG					
				GTTTTC					
				CCGTT					
				ATCTA					

				TCTAA					
				GTTGT					
				ATTAT					
				CACAT					
				ATTAT					
				TTATTT					
				ATTTA					
				TCATT					
				AGATT					
				ATCTA					
				TTTAG					
				GAAGT					
				TTTGTT					
				ATATA					
				TTTTAT					
				TATAT					
				TTAAA					
				TTATTT					
				TACTA					
				TATAT					
				AATTT					
				TATTA					
				TATTTT					

				TTGAT					
				CATTA					
				TTATA					
				ACTTT					
				ATCAT					
				TATTA					
				TATAT					
				TCTTAT					
				ACATT					
				TATTT					
				ATTAT					
				GGATT					
				TATAT					
				TATTTT					
				AATTT					
				TATCT					
				ATTGT					
				TTATT					
				ATTTT					
				ATCTT					
				AGTTA					
				TTAAT					
				TAATA					

				TTATA					
				AATCT					
				ATACT					
				CTTTAT					
				TTATA					
				ATTTC					
				GATAA					
				TAATT					
				AACTA					
				CGTAC					
				TACTA					
				CTGAT					
				TACTG					
				TTGCT					
				ATTTA					
				TCCAT					
				TTACT					
				GTTTA					
				TTGTTT					
				GATTA					
				TTTTCT					
				CTTTTA					
				ATTTTC					

				TATAT TTTCTG TTCTAT AAACG GCTTT ACCAT CTATA TC					
CR1	CR1 - 12_ HM		Hydra vulgaris	AAAGG TCTAT AATGA AGCCT ATTTTC GGAAA AAAAG TTGCA TTTAA CAGAA ACTCT ATTGC AGGCG AACGT ACGGT	TCAAATA CAGGAAA CACAGAA TAAAAAA TCCTTTA CAAACGA AGCGTTT GATGCAT GCAACAC ACTTGAT TTAAATT TCATACA GTGCATG TTAAATG GCAAAAA	MEGNIESKTFENISFCAYNCQSFKSNSIYISELV KSFDIIFLSEHWLLNIELFVLKNIALPTHKVFFH AAEKKA FGRPF GGN AFIKKDVLLSPHIIYEDD NILAIKGKNKCHNLVFIGIYLTSCRNNEESLSK YEQQLNLITSIVKNYEDVSECIIAGDFQSY PETL YDSEIRNNKKRNKFSNHLTNFIKSNKLSFIDIIS GAGPTYTYQHKTL SNSSYIDHIAALKESSLSFIN TSVIPSPFNFSDHLPIATNIVVTHTNLSTVINI MHKYNYPKYAWNNEKFINIYNHNLSVAFNY YTFNEEEIEQELKQTCEKIQNAALLAVKQCFK QQICKYKSWWTTDLKKCKENLSFNYKQWQ KSNYNKTQECIEYKRYIFSRKNFRKAVKAAHN KKIHEKTINIENLR TTNPQR FWDNIRKIKTKAT TRLFTINKSQNIKDIVQEFRQHFQTLLNTPLITN NNHNPSQIPPLSKEPNSLIITSADIINCISQLNSN	cd0165 0	542	799

			GCGAG	AATTAAA	KSPDSCGLSAEHLKNSHNNNLLLWLAKFYNSI			
			AATCT	TTGACGC	LSNGKVPNDLSTSTIIPLVKSYKKSLNNPDNYR			
			TATAT	AACCCGT	GISILPIFTKLLEYLILLICPSITDSHSHQFGFKKN			
			CTTTGT	CGTAGAG	SSTLHAEFLLSETVKHYKNNNTPLYMCSLDAN			
			AGGAG	CTACAAA	KAFDTCNWDLLFEKLYFQKKLPLSVVHTISSL			
			GGGTC	CTAATGT	YHSGTANVSYQGVTSNQFSLSQGVRQGSILSP			
			TGTAA	TATGCTA	HLYNIYTESILNEIVSECKVGSTINGIYTGVIAY			
			CTCAA	ATATTGA	ADDVILLSPTISGLQELINRFQTYGVANFIKLNT			
			TGTCG	AAGAAAC	EKTEFLISGKSYIXNNVIKINGVSIYLNKHL			
			GAAGA	GTTCCAG	GFQWDNDTSKNKIATLQKTNLNERITQFQSV			
			AGATC	TTTTGGA	KALIDSGIRYCQPSTIVHLYNTLTVPKLIYGLEL			
			TATCT	ATTTGAA	CKYDTNFLSKIDAVGRAVLKSFFCISKYSKNYL			
			AACCA	AGAAAAG	HSFFKIEDISTILLRNKVNLFIRLLKNQTCLXTD			
			CATCA	AGAGAGG	N			
			AAAAT	CGTTTAA				
			GAAAT	AATCTTG				
			AGGTA	ATGGAGG				
			TTGAG	AGAATGT				
			CCTAT	TCCTAGA				
			TAAGG	ACTCTGT				
			TAGAG	CCCAATA				
			CTTAA	ATATGTG				
			CCGGG	GAAAACA				

				AAAAT	AACCGGC				
				AAATA	TGTTTAT				
				CAATC	AAGCATT				
				GTTCA	TATTGCT				
				TTCAA	AACAAC				
				AGTTA	ACAAGAG				
				TCATA	AGCCCGT				
				AAAAG	GTTGTTTT				
				TACTG	AGCTGAT				
				AAAAA	TTTAAATT				
				GACAT	TATTGTTT				
				AATCC	GACTATT				
				TAAAC	TAAATAT				
				CCTGA	TATTTTA				
				TATTT	CCTGTAA				
				GGGAT	TTTATTA				
				ACTAA	CCGGGTG				
				TATTA	ATAAAGA				
				TCGTT	GAAAATA				
				AAACC	TAATATA				
				ATTCA	ATAATAA				
				GGCTT					
				CATCG					

				TGTTA					
				CCTCA					
				AGAAA					
				CGAAA					
				CACCA					
				CAGCA					
				GAGTG					
				CAGTT					
				AATAC					
				GCTAC					
				ACTCT					
				GTACA					
				ACCAA					
				CTCAA					
				CAAAT					
				ATTTTC					
				TACAT					
				CTAAA					
				GAACA					
				ACCAT					
				CCTTT					
				GCACA					
				GCAAC					

				ATTCTT ACTCA ACTCG AATAC TAAAT TCTAA TTCAA A					
Dap hne	Dap hne- 10_ HMa		Hydra vulgaris	GGGAA ATTTTT CAAGA AATGT TYCAA AAACA ACAAA ANGAT ATTTT AAMAT TAATA AGTGG AAATT TAAAA CTGAC	AAACTCA AAAATTA TGGTATA AAAAATA CAAATTT AGCTTGG TTTAAAA GTTACTT GACTAAT AGAAAAC AATATAT TTCATAT GATGGMG GAAAAAC TGAATAT	MIFLRFYIIFLNFNIALSYFLFKMADETVLTNLN FNSFESKKSILLNNYS DPDINFYNDEV IKNINP LYYNPN SKNIANGMDDNSFSMLHINIRSIQKNF ESLKQFLYKIKINFQIICLSETWCHDKNIENNSN FQLPNYKVIHQVRASNKEGGGLCIFIKN SLLYK VKPNLSSTTNDYESLCIEIINETS KNIIHALYRP PSGSIKVFENHIKKIIKNKSSINKTVYFVGDINL NILDYDKNKNIKNFFNVIFQNSYIPLINKPTRIT KESATSIDQIITNEFINKKIKTGIFKTDISDHFPIFI JSQKCIYNAQNKRKKIITTRLINDNSIKHFHNLL SCVNWDTLKLNLNABKAYDIFLTEFLKLYNK AFPEITKFIKTKTLLNPWITKGILKSSKKKQRLY NKFLKKKTLKNETDYKNYKRLFESILKRSKKH YYSEQLIKHKND SQKTWQIIKEVIGKKNLYRN LLPKNLKFNNKYIVNKS LVAETLNQFFVNIGPT	cd0165 0	592	694

			TAATG	ATGACAA	LASKIATTQLNFKSYLTSNNINVMXNYKLTEK			
			ATCGA	TTACTTG	ELLDVAVSLLKPNESLGFDISSNVIKSIKYITIP			
			ATCGA	CGGTGTT	LLHIFNLSLKQGVFPEKLKIAKVIPIFKSGDPSN			
			TGGAT	CCCCAAG	VANYRPISILPCFSKILERIMYNRLYSFLNINNIL			
			TGCTA	GATCAAT	YNKQFGFKSGHSTDHAIHVLVHDIFKAFDEKKF			
			AAAGA	TCTAGGA	TLGVFIDLSKAFDITVDHKILL			
			AATTA	CCACTTT				
			CTACA	TATTTCT				
			TTAAA	AATTTAT				
			AGAAW	ATAAACG				
			CTTGT	ATTTAAA				
			GAAWC	TAAATTT				
			TTTAA	TCTAACA				
			AAAAA	TTTTAAA				
			GAAAA	CTCAATT				
			CGAAA	TTATTTG				
			ACMGA	CWGATGA				
			AAWGC	TACTAAC				
			CGATC	TTATTTTA				
			AAAAT	TTCTAAT				
			AAAAA	AAAGATA				
			CAAAC	TTAATAT				
			AAGCA	TTTATTTA				

				GTAA	AAACAGT				
				TAAAG	AAACAAA				
				TATGA	GAACCTA				
				AAAAA	TAAAAAC				
				TACAA	TAACCGA				
				AAAGA	ATGGTTT				
				CATTG	AAATCAA				
				AAGAA	ATAAATT				
				AGTCT	GTCCTTA				
				AACGT	AATATAA				
				TCTAT	ATAAAAC				
				CAAGA	AAAATAT				
				TAGCC	ACTTTAT				
				AWGAT	TTCATCG				
				AAAAA	TCTTCAT				
				AGTTA	AAAAAAG				
				GCGAY	AAAATAT				
				TTAGA	TCCATTA				
				AAAAA	AAACTTC				
				AAAAG	CAAATCT				
				TCATC	TTTTATTG				
				GAATG	ATAACTC				
				CAAAT	ATTAATA				

				ATAGG	AAAAGAG				
				TGAGA	AGCAATC				
				ATGAA	AATAAAA				
				AAAAA	TTTTTAG				
				TAAAT	GCGTKAT				
				TTAGA	TCTTGAT				
				CARCT	GAAAAT				
				GGAAG	WTAACAT				
				ATCGA	GGAGGGA				
				CAAAG	MCACATA				
				AAGAA	AGTATAA				
				ATAAT	TTGAAAA				
				CTTCG	TAAAATT				
				AATTG	TCAAAAA				
				AAGGA	ATATWGG				
				GTTAA	TATATTG				
				AGAAA	TATAAAG				
				ATGAT	CAAAACA				
				AACGA	GTTTTTA				
				AAGTT	AATCAAA				
				GGGAT	CTTGTTT				
				GATAC	AAAAAAC				
				GGAAA	TTATATTT				

				TAAAA	TTCTTTTA				
				ATAAT	TTCATTG				
				AAACA	YTATCTA				
				TTTTTG	AATTATG				
				AAAAT	CAAACAT				
				AATCT	TGCTTGG				
				CAATG	TGCAGCA				
				TAAAT	CTAATGC				
				GGAGT	AACAAAA				
				AATCA	ATAAAAA				
				TAGAA	AACTGTT				
				AGAGC	TAGTAAA				
				GCATA	CAAAAAC				
				GAACT	ATGCTAT				
				GGAAT	AAGGAT				
				TATTG	WATTACA				
				AAAAA	AATGTAG				
				AAAAA	ATCGTTT				
				ACCAA	CTCTCAT				
				GAACA	ACTAAAA				
				ATTGT	CACTATT				
				AATTA	TAATAAA				
				AGTTA	CTTAATA				

				CTAAA	TTCTAAA				
				TTATA	TGTATAT				
				AAGAC	CAACTAA				
				AAAGT	ACCTTTA				
				AAAAA	TCATATT				
				TTYTA	CTTATTTT				
				AAAAA	TATGTTT				
				TGCCA	AAACTTG				
				ACAAA	ATAATAA				
				CTAAA	AATATCA				
				AGGWT	CCAATTT				
				CTGGA	TATTTAA				
				ATTTA	TTCATTTT				
				TATCG	TTAAAAA				
				ACGAG	AATAAAT				
				GATTT	CATATAT				
				TTCGC	ACCCTAC				
				TAGAA	AAGATTT				
				ACTAC	TCAAAAA				
				AATAA	ACAATTA				
				TAAGA	TATTCAA				
				AAAAA	TCCAAAA				
				ACTTC	CACATTA				

				TTGAA	TTCAGCA				
				GAAAG	ACTAAAT				
				CAAAA	TCTCAAT				
				TGCAT	TGCAAAT				
				AGGAA	CGAGGTC				
				AAATG	CTAAATT				
				GTAAG	ATGGAAT				
				TATTCT	ACAATAT				
				GTTGT	TAAATAA				
				TATAT	TGAGCTT				
				ACGAT	AAAAC TA				
				AARCT	ATTCTTCT				
				TATTG	TTAAACC				
				TTAAA	AATTTAA				
				GAATT	AACAAAA				
				TATAA	CTTAARC				
				AAAAA	AAAAACT				
				AAATA	TTTGATA				
				ATGAT	AATGAAA				
				AA	ATGAATT				
					AAATTTC				
					TTCTAAA				
					AAGCTTT				

					TTAAATT				
					AGCAAAC				
					AAAAATA				
					AAAATTA				
					ATCTATT				
					AATTACT				
					CTTTAAT				
					ATTAATT				
					TAATATT				
					TTAAACA				
					ATTGTCT				
					TATTGAA				
					ATCTATT				
					TATTATC				
					TTATAAG				
					TTTTATTT				
					ACGACGA				
					TTTTTTAA				
					GTTTTATT				
					ATTTTTTT				
					TKAATTT				
					CTTTTTAT				
					TATTTTCT				

					TTTTTATT				
					ARTACTT				
					ACTTATT				
					TTACTAA				
					TCAATTA				
					TTATTAT				
					AAATTTT				
					TCGAATT				
					ACWAAAT				
					AGTTATA				
					TATTTTAT				
					TTTTTATT				
					TTGTTAA				
					CGTTTTA				
					CTTATGA				
					ATGCATT				
					TTGGGGC				
					TTAGTGA				
					TAAGACG				
					AATTTTG				
					TCTTCTTC				
					TAGCCCC				
					AGTCATG				

					TAAATAT TTGTTTT ATAAATT ACATTGT AAATTAT TTTCAAC GGCAAAT AAATAAT AAAAAAA AA				
Dap hne	Dap hne- 16_ HMa		Hydra vulgaris	GAAAG TCTTA ACTTC CAAGA AACAA CTATT ATGGA AAAAA TAAAT AAGAT CACGA AGCGT TACGA	AAATTTA TGTA AAA TGTATTTT TCAAATC TTCAGCA TATTGAG TATTA AA ATGTTAG TATATTG TATATTT GCATTAT GTTATAT TGTATTA	MALISDFETLVFNFFETKFFSLHENS DPDINYFH DTKIECSYYYPNELKGILFKNVNNDQIRILHVN IRSINCNEKFQNLLEETKYFFNIICLTETWATS NDIKKNFNFYLSHFEMISLERQINKRGGGVLIYI HKDIKHCCRNDLSVSDGDGEILTIEIINERTKNI LVSCCYRPPDGVSENLSMFLQQNVIAKGNNEK KKNFLVGDFNMNCFVYNDDNKVKNFYDTIFE SGAIPLINRPTRITKNSATLLDNIITD LFNNDIK LGILKTDISDHFPIFLTINTDQLPKQKLNKIVKK RFYTSSNIEAFKNQLSLLHWKHLNFNENANNL YEKFYETFISVYESNFPIVTITLKT KHFDNPWIT KGFKKSSKTKQRLYIKYLKTKTPDSEKIYKDY KNLF EKIRKKLKKNYYSKLINKAKNDSKRTW	cd0165 0	569	828

			AAAGG	GTATATT	QILKEITGKQKTCSSSLPQMLKVDNNSLHEPQII			
			AGATA	GCATATT	AHEFNKYFTEIGSILSSKIQKTQTSFYDFLLPFD			
			AACAA	TGCATTA	KNISSEELSAELSFDEFEKAFKSLKKNKAPGAD			
			CTTAA	TGTTAAA	EINGNIVIDCYEQLKNILFKIFRASIHQGIFPERL			
			ATAAG	TCTTTGC	KLARVTPLHKEGDRSNINNYRPISVLSIFSKILE			
			AAGAC	ATTATGT	RIIFNRVYNYFNYNLLYNNQYGFRKGSSTEH			
			ACTAG	CTTTTTTA	AIIQFIHNISKSFEKSQYTLGVFIDLSKAFDTVD			
			ATTTG	AGGTTCC	HHILIQKIKYYGLNNKILKWFESYLTNRKQLV			
			GAGAA	GGTGATA	YSNDGYQSEPLSITCGVPQGSILGPLLFLIYVND			
			TCGAT	AGATCCT	LNKASKLKSIMFADDTNLFLSHTDIYELFSTTN			
			CTCGA	TCTGATC	KELKHISNWFKANKLTLNINKTKWIIFHSCAK			
			AGAAA	TTCTTTTCG	KRFLTNNMPQIYIDETIIKRDTVIFLGVYLDEN			
			CAACC	GAAACCT	ITWRKHIDHISTKVSKNIGILYKARNYLNKKNL			
			TAAGA	AGTCTTA	TQLYYSFIHNYINYAIIAWGSTESKLQHL YRR			
			TTAGA	TATTGTT	QKHAIRVVNYADRF SRSKYFFDYMNVLDVYK			
			TGGAA	AGTACGC	LNIFNVLCFTFMWKNDLSLFVFNDLFS LKPINK			
			TATTT	TGAATTG	YTLRSNSFLNEPFCKTKFNQFCIAYRAPHLWN			
			GAAAA	TATATAC	KIVLPNFD PSTTLPVFKTKLKNFILCMDNILKFY			
			GCCAA	ATATATA				
			ATGAA	TATATAT				
			AATTG	ATATATA				
			GAATG	TATATAT				
			AGTGC	ATATATA				

				GAAAA	TATATAT				
				TGCAG	ATATATA				
				TAAAG	TATACTC				
				GAAAT	ATTTTTTT				
				GTTTA	TTTCTTTT				
				AAAAA	CTTGTAG				
				CAACT	AAGTCTT				
				AAAAA	AAATATT				
				TTAGT	TTTATGT				
				AATGA	ATCAGAC				
				AATTA	TTATATT				
				TCATA	ATTATAT				
				GAAAG	ATCACGA				
				AGCTC	CTTATAT				
				ATAGA	TATTATA				
				ATAGG	TATGTAT				
				CCAAC	CAGACTT				
				CTAAA	ATATTAT				
				GAAGA	TATATAT				
				TAAGA	CACGACT				
				AACCA	TATATTA				
				AGAAC	TTTTGTAT				
				GATTG	CACGACT				

				TCCTA	AACATTG				
				AAGTT	TAAACTA				
				ATTAA	AAAAAAA				
				ATTTTC	AAAA				
				AAGAC					
				AAAAC					
				AAAAA					
				TACTT					
				AACGC					
				TACAA					
				AAAAC					
				TTGCG					
				AGGAA					
				CTGGG					
				ATATA					
				TGTTA					
				ATGAG					
				GATTT					
				CGCTA					
				AAGAA					
				ACGAT					
				GGAAA					
				GCCGA					

				AGGAT					
				GTTGT					
				GGGAA					
				GAGGT					
				CAAAA					
				AGTTG					
				CGACT					
				TGAAG					
				GTAAG					
				TATGC					
				AGTTA					
				TAAAG					
				TATGA					
				TAAAA					
				TTGTTT					
				CTAGA					
				GAATT					
				TCGAA					
				AGTAA					
				ATGCG					
				CTCAT					
				TCTTA					
				ATAAG					

				AACTC CTTTT AATTT ATTTG TGATA ATTAA TC					
CR1	CR1 - 11_ HM		Hydra vulgaris	GGCTA AAGGA TTATC ATTTA CTCAA ATTAA AGAAT TACTG GAAAT TCATG AAAAC ACAAT ATTAA AAATM TTTAR CGATA	TTTTTTGT TTCTTTTT TTAACTT TATATTT AACACTG CATTAAT AAATTTT ATAATTT AACAATT TATAAAC AGTTGTT TAACACT TTTGTTA AAGGTAT CTTAAAA CTGTATA	MDSNEPVNFESTLNFNFQTNDFLLDSDSDPDLN YFSEAGALQNKCSYFYFNHEIKEFLDRDNFNVI HINIRSLKKNFENFRNXIEETFNIFSIICLTETWC DSDDVNFNSNFHLPGFNLISLARKTNKRGGGV JIYVKEXLXFFTRLDMISISDGDKEVL TIEILTKK TKNIIISCCYRPPSGVIESFNSFLRNDIIKKS NHE KKFNYLIGDLNLNCFEYHVNNNIKR FYNDIFE NGAIPLINKPTRITQSTSSLIDNIITTDVFNESLK KGIKSDISDHFPIFFSINTNNEXVTKTHSLCLKT IKFNDKFLYDKKEIANELNKYFVSVGP NLAKKI PKAINPINNLFIPNTSQFNSFELSFE EFESA FKML KPKDAVGPDDINGNIVINSYDIIKDILFKIFSCSI KQGIFPDQLKIAKVPIFKGGDLTNVS NYRPISIL SVFSKVLERISYNKIYNHLSVNNLLYSNQYGFK KNNSTEHAILQFTRSITDSFENSQFTLG VFIDLA KAFTIDHKILFKKLKYYGITGNVLNWLKSYL	cd0165 0	418	677

			AATTT	ACAAAAC	NNRKQFVYVDVVSPVNLFDITCGVPQGSILGP			
			GATAA	CTTTAAC	LLFLIYVNDLHKASNLMTIMFADDTNLFLSHN			
			AATGG	AAAAAGT	NITTLFQNMNIELVKISDWFKLNKLSLNIDKTK			
			AAAAC	TTAACAA	WIFFHPNRKKHLLPSEMPLLFIDNILIKRVTVTN			
			AAAAT	TTGTGCT	FLGIYIDENLSWKNHIENLCSKISKSIGILYKAR			
			GTAAG	AAAGATA	SVLNKHTLIQLYYSFHCHINYANIAWGSTNKS			
			TACTT	TCTTAAA	KLXPLYRQQKHVARLINFKDRFTHAKPLL FEM			
			AAAAG	ACTGTAT	KVFNIYELNIFNVLCFMFKCKTNISPVSFHNLY			
			MGAAA	AACGCAT	XKKEKNKYILRNDNFIRQPVFQTNFGKFCISFR			
			ATAAA	CTATGTA	GAFLWNKIVLKNFKDFSQECNYRSFKQKLKK			
			ATTTT	TTTATAA	VIFSIDDISIYF			
			AAAAA	CGGATTT				
			ATGAA	CTTTTTCT				
			ATAGC	TTTTTTCT				
			AGATA	TTAAATT				
			TTAAA	TGTTTAT				
			AGTGC	GTTTTCA				
			AATGA	ATTCTTA				
			CTTTTC	ATTCTTA				
			ATAAA	ATWATTT				
			TGAAA	GTTATCT				
			AGTAT	TCATGTA				
			GAAAA	TGTTAAG				

				AGAGA	TATATAT				
				CTAAA	TCCTATT				
				GAATT	GAAAGTT				
				ATACG	TATTTATT				
				ATTTT	TTGAGAA				
				AAAGA	TATTTTAT				
				ACATC	TAATACG				
				TTAAA	AAATATT				
				CACAA	AACACGA				
				CATGA	AAATTTT				
				AAGTA	TTAGCGG				
				AATGA	TTCTCGA				
				AAAAG	CGACAAG				
				CTTTA	ACCTTAC				
				AACAA	GGTCTTC				
				TGATG	TACGAGT				
				AATTT	WCCCGCG				
				AAAGA	TTCTTTTA				
				CGAAC	TTTCTTA				
				TTGCK	ATATCAT				
				GAACT	TTATATT				
				TGAAG	ATTATAT				
				ACAGA	CATTGTT				

				AGTCG	AACGAAT				
				AAGAA	ATTTATT				
				ATAAT	AAATACA				
				TTAAG	AAGTTGT				
				ATTTA	AATAAAG				
				ATGGA	AACAAAA				
				ATTGA	AAAA				
				AGAAA					
				AGGAA					
				AATGA					
				AACGT					
				GGGAA					
				GAAAG					
				CGAAA					
				AAAAA					
				ATTAA					
				AGAAN					
				TCTTA					
				AAAKA					
				AAAAC					
				TTGGT					
				ATTAC					
				GAACG					

				AAATA					
				TAAAT					
				TGAAA					
				GRGCT					
				CATAG					
				AACAG					
				GAAAA					
				AAGAA					
				TGAAG					
				TTGAA					
				ATGAA					
				AAAAC					
				GAAGA					
				ACTAT					
				TATTG					
				TAAAA					
				TTTTTA					
				AACTA					
				TAAAG					
				ACAAA					
				GAAAC					
				CGTTTT					
				ARAAA					

				AATAC					
				CGATT					
				ACTCA					
				AATTA					
				TGGAA					
				CGAAA					
				AACTA					
				TACAT					
				AAATG					
				AAGAT					
				TTTAG					
				CGAAA					
				GAACA					
				ATGGA					
				AATAC					
				GAAAA					
				AATYT					
				WTTTA					
				AAGAA					
				GCAAA					
				AGATT					
				TAAGA					
				GCCAA					

				AGGTA AGTAT GCTAA AGTTG TATAT AATAA ATTAG TTACA CGCGA TTTTTA AAAGA ATTAA TTTTTT TATTTT TGGAT TTCGA ATAAC TAAA					
CR1	CR1 - 72_ HM		Hydra vulgaris	AGTCC CTGTG TAATA AGCAA AACTA	CGTTTTTT ATTATTT AGTGACT GAACTAA CTTG TTC	MVNLQLGDFNTYKCEQEAIIVDLKKLIETLRT DNANLTKRVLDLENSKSKSSSAVTGDWASKV KKTPDQMQUIINTIANETKERDKRENNLVIFGLE ESTLSDKVKAIEEDKQNIETIMKKLKVNVIKIN VIKLKSKNGYKAPYIVVLNDKLERN SVLKN AK	cd0165 0	791	1046

			ATAAT	ATTAATT	ELKKDKTYESVFINPDLTESERYKSKLLREECK		
			AATAA	AATTTAT	KLNNDNINKANFYFGIRNDKVVKIKKQLNAK		
			CATTA	ACCGGCT	KDLNHIRCFNYETNNKINLFSGGGQNLNKLLS		
			CAATT	GTCATAG	NKRQLDIEIVNFHERKKCDLSINKEHDTQKCN		
			TCACT	ATATTGC	ELMIWYTNATSLNNKIDELRLQCEIYKPDVICV		
			GACAG	CTGGCGC	TETWFKAESDVNIISYNIFRSDRKTHGGGVCIY		
			GAGGT	TTTATCTC	VKQCINQFETNNLQLNNNFIEQIWTCICFKNNK		
			AATCC	AACAGCA	MLIGCIYRPPDISHEYNNYVLLSINAAMMSIYQ		
			ACTGT	TTAATTA	KEYSDLIITGDFNYNTIDWSDNNCPFSATDNES		
			TATTT	ATAA	QTNKFIECLEDCFLFQCVCEPTFQIANYSKSTNIL		
			AGGGT		DLIITNDKNKIVSITHSAPLGNVKQGHLLKCC		
			GGTAG		YLMNDYKFSTEHHKSIRLYQKGNIELFSCHFNI		
			CCCTC		NVDWENGFKSKSINEMYDLFLYHYNATNLFI		
			CTATC		PRINYQNRNRNRNKWITPELKHKIRTKNNIWLA		
			CAGTA		NKRNGWNEESTKIQYNKLKIRIKNDIKNNIKQF		
			ATCCA		ESTLASKFKSNPKLLYKYINEKKSVKTHIRALE		
			A		IDNGEINDQFAIATELNKYFHSVYVNDNCTNII		
					EFHNIPSNLLLNTVLITQNDVQTRLQALDKSKS		
					VGFDVHPYVLKECSSSISYPLYLIFKKSMETG		
					TLPDMWKKANVMPLFKKGSKLKASNYRPVSL		
					TSIPCKVLERIADCIMQHLLIKNNLLSKKQHGFI		
					MKSKSCTTNLLEYLDILTDAFHNRTVPDVLYT		
					DFKKAFDSVSHKKLLSKLLTFGISGKLLTWIVC		

						FLANRKQRVVLGKSVTDWVDVLSGVPQGSVL GPLLFLVFINDLPENFINECRLYADDNKIIAPISS QNDSQLFQNDINKLDEWSTKWKLGLNFEKCKI MHFGSNNNEYSYFMNNNNTLMEIEKSKLEKDI GVYISSDLKWAKQVKYAASKANSILSLLNNTF KYKDKDLIKTLYCTYVRPNLEFSIQAWSPYYT KDINELEKVQRRATKMIPELRHLEYKKRLEILG LTTLETRRLRCDLIQQFKIFKGIEIIDIKQNQTM NSICSSGPAANIRGAKHRMKPELVRNCLKRQY FFSNRVVDGWNSLPAEVVQSTTVKSFKLNLQS VDLEAITNKKIKTKKAYY			
CR1	CR1 - 55_ HM		Hydra vulgaris	TTAA	TACAATT TTGTTGT AACTTGT CCAGCAC TTGATAC AATTTAA CATTTCA GTAAAAA AAAAAAA ATGTTTTT GTTAATG TATACTC	RWQPKFCIKYYNFNFLPKLKIFIQLTLIKMVRK EDFEALVSRVLKLEAELLTKDKEITQLKILVND LKNQPSNVLNSSACAWSDVVKGKKSNAQAQL DLVNVVVKESQDRDLRQKNLIVFGFNTSKNK NDDVAREEDILGIQNIFNQINAKATIQRVFKLSI KSDKPPPVIIVLTNKEERNAVLKTAKTLRNLP ETNNIFINADLTVAERLKMALREERAKLNET NKDLQYYYGIRNERVVKLIKRFISSQQDQLLF SKNLLFKNSNKFIEKNNKRQLINLKNQELFNHP KNNYCGKWALTNIRSLGNKIADFQHFVHLHSL DICCVIETWLNKNLPDSSLCPRFYTVLRCDRSS KGGGIAIFFRSFIQFTKVDIPPEFHELEIVCADLN	cd0165 0	787	1043

				AGTGTTA	LNGQSFRVIGYYRSGGFGKVAIDYMQSSVKCF			
				GTGGCCT	SILCSTKKNVVILGDFNLPDIDWHYYHGPDSVI			
				TCGGGTC	HNSFLHFINRYSLTQHVDQPTRENSILDLILSSS			
				CTTCTGTT	SSFISNLCIEPPIGTSDHNVIIFTPNLSCLYNKTE			
				GTTTTCA	VPVQQFFCWKNADYQVINKYLKSVDWSIIFQT			
				AATAAAT	CFNIEACWHTFSTILNNIITQHXPKVRVCEEHK			
				AAATAAA	LSHKIHYPYYVKRLINNKAMAWKRSKVTGKS			
				TAAATAA	NDKAAYKSIAAQCSDSIRKYHAAKELKMIRKD			
				ATAAATA	NLGSFFNFVNGKLKTC SKLNDIRRS DNTLCHD			
				TAT	KEEISCIFNNFFASVFTVDNGLSPKFTDRVNVH			
					STCLSSINFSPSTVEKALRAKPTTSTGLDGIPN			
					VFLKNCANSLSIPLSHLFDTSFKDGAIPLDWKV			
					ASVIPIHKKGPTVDPGNYRPISLTSTCCRVMERI			
					INNKLINYLLLNNLITKEQHGFIKRSTCTNILE			
					SLHDWTLNLQRRIPTDIIYFDFKKA FDSV SHPK			
					LLMKLPAYGIVGNLLNWL TNFLHMRSQV VNL			
					NNVTSHAISVTSGVPQGSVLGPTL FLLFINDVS			
					HINKQLDVKLKLFADDIKLYSVYDVRGLQSDL			
					HTAVNXLYEWSCTWQLQIATEKS FVCTIANQS			
					QNIAQRVYTINNHEL AHVNCIRDLGVTIDRNL			
					KFNQHINLIVHKAMSRAFLILKSFYSRDRLLMV			
					KAYCTYVRPMLEYCSSVWSPYTICLINKVEKV			
					QRFFTKRIAGLWSVSYDKRLAILKLNTLEYRR			

						VFKDLVLCYKILNGKLDTDLSNILILNSYLNNR GHNFKLYKHHCSLNSTKCFFSNRIVNVWNSLS EDVVSASSVLMFKKRLAVLLCSS			
L2	L2- 1_H Ma		Hydra vulgaris	AGGAG CATGT TTATG AAAAT GTAA GGATT GGTGG GAGGC TGGTA AGGAG AGAGT AAAAA TAATC TCGAA AAAAT TTAGC AAAAG ATTAA AACAA GAAAA	TAAAGGA ATGTCAA AGTTATG ATTTTAA AAAGTCT TCATTTT ATTTAGC CAAAAGG AAAACAA AATACGA ATATAAA AATCTAA CTGTCTT AACTCTT TACAAAT TATTAAG CAAACCA ACAAAAG TTTAGAC TATAAAT	MLKKEISESEIKQALGQMKNKSPGSDGLTIEF YKHFFXLIKKDLIESLNNVLFTGEMTNSQKEAJ ITCIFKKGNKNDIGNWRPISLTNVDYKILTKVL ANRLKDVLPNIHNSQTACIKNRTIYYNLSYTR DIISIAKKSKLDACILSIDQVKAFDRVDRAFLFD SLKHFGFGKIFISFIKTIYNNINAKVKINGLLSEK IDILRGVRQGCPLSMILYIIQAEIFSCYVRQKN IKGIHVKESETKJQQYADDTNFFLIGDDSIREV KALKLNQKATGAKINIAKCQGIWLGKNIFKDK SEYLNFNWDXNTFKSLGIFFSNNANFQYETHW KERFNKMENSMRKWEKYKLSIKGRKTIINSLII PQALHLAFIAPLPSNGILNKIEAKINNFYWGGKI CRLNKTQTQQKINEGGAGVNSLTNKLKAIQLQ WISKLYNESICGPWKDAMVYILNLYRNANQS KHTFETIPTQMKLPPFYSQLLVNWHSLKQKH VLPITKIEEILNQPLFFNSMIKTMPNKFLTPSET HIKNDIILVADITKTFQPGFLHHKSIGLTEKELE FLTKALPRKWYKKF	cd0165 0	64	321

				GAACA	TTTATAA				
				AAAAA	ATGGAAA				
				GAATT	GATTGTT				
				TAAAT	TTGAAAG				
				TAATC	CATGAAA				
				AAATT	AATATAA				
				ATTAA	ATGAAAA				
				AAAAC	AGATTGG				
				GCATC	AAACGTT				
				GAAAA	TCTTTAT				
				AAATG	AAATCTT				
				CATTT	CAAAAAA				
				AAATT	GTAATAC				
				GCAAT	TAACAAC				
				ATGAA	AAAGCGT				
				AAAAC	GCGAAAT				
				TTTAC	TCTATTT				
				GACGA	AAAACAG				
				TTTAA	CACACTT				
				AAGTT	TTTGTTG				
				AAACT	CCAACCA				
				TGAAA	ACGATAA				
				ATATA	AATGTTT				

				GAAAA	GATTTTA				
				AGAAA	AAATTAG				
				AAGCC	AAAAGAC				
				GTAGG	AAACTTT				
				CGCGG	GCCCACA				
				CAATA	GTGCAAA				
				AGAGC	CTTAAAA				
				AAAAA	ATGAAAA				
				TTCAG	TACTGCC				
				TGGAG	CACATGA				
				GTTAG	TGTATGA				
				AAGGG	ATGTGAA				
				GAAAA	AAAGTAC				
				ATGCT	AGCCACT				
				CCCAT	AATATTT				
				TTTTTT	TATGTTC				
				TTCGG	TAGATAT				
				ATTGG	TTTAGAC				
				AAAAA	CAGCTCT				
				CAAAA	ACCCTAC				
				ATTAG	TGAATAC				
				CTAAA	CCTCATA				
				CAATA	ATAACAA				

				TGTCT	TTTTAAA				
				CAGAA	TTTATTTT				
				ATAAA	ATTTGGA				
				CGGCA	TACGAAA				
				GAGAC	AAAACGC				
				GGGCA	AGCCCAG				
				GGTTA	AGGCTTT				
				AAATC	TTGGCAG				
				GACTC	TTTATTAT				
				AAAAA	TAGAAAG				
				AATTT	TTTTATTT				
				TATTA	GTGAAAT				
				GAATT	ATATTTT				
				TGAAC	AACAGAA				
				ATTTT	TGAAAAG				
				ATAGA	TCAATTC				
				AACTT	GAAAAGA				
				ATACA	AAACATT				
				AATTC	CACTAAG				
				AAAGA	AAATATT				
				AAACG	TGCTAGA				
				ACAAA	AAAGTTT				
				CCTTCT	GATAGAA				

				CAAGA	AAGTCAA				
				ATTTT	AAGATTA				
				GTTCTT	TTTTTAA				
				AAATG	ATGAGTT				
				CTAAA	ACTTATG				
				GTTAA	CTGAGGC				
				WAAAG	AACAAAA				
				TTACT	TAAACAA				
				AAGGA	GAAAAAC				
				AGAAA	ACATAAA				
				ATAAA	CGAAATA				
					AGAAACA				
					TTTTTAA				
					AAATGGA				
					AGATTAA				
					ATTTGAT				
					AGCAGAT				
					GTTCGTG				
					ACTATTT				
					CGTATAA				
					TTTTCGT				
					TTTTCTTT				
					TTCTTTTT				

					TCTTTTGT				
					TTTTTCG				
					ATTTTTAT				
					TTTTTATT				
					TTTTATTT				
					ATTTTTTA				
					GTTTTATT				
					TTATACG				
					GGTCCAA				
					TGGACAT				
					CTTTGTA				
					AACGTTT				
					TTTTTCGT				
					TTTTTCGT				
					TTTTTTTT				
					TAAACCT				
					TGTATTT				
					ATTATGA				
					GGACGTC				
					AGTATGG				
					CAATTAG				
					CCGGATG				
					ACGGAAT				

					AAATTAC AATACCA				
CR1	CR1 - 6_H M		Hydra vulgaris		AGTAAAC TAMAATA AAAAAAA GGGAA ATAAAAA TTGGA AAAACAG AGATT ATTCAAT AACAA TATATCA TTTAA ATGATAT GAGTA ATGATAA GATGG TATGAAT AATTG AAGAAAA ATGAA GAAAGAA AATGA AAAGGAA AGGTG AAAAAAA AAAGC AAAAAAA TGGA AAAAAAA AGAAA AAAAAGA GCGAA ATAAGAA ATCCAAG TGATGAT GTCACAA	MKVKKIFEDILSVKNVQIERAHRTGNKKLKMP RTIVIKLLDFRDKVAILSKSSKLKGKNIYINEDF CFETTQIRKNLREKMKIERAAGKFAFISYDKLII RDWVAKKSKNLSSTVLKMEENVIFKSSNENDI IGDNDTDSNYFNHKMLESQYYSVSESSQYLGQ TKNTFSILNINIRSLNKNFENLKILLDEIKHDFKI ICLTETWCKCGETNYEFELVNYTSIHQPREVN AGGGVSIFIHNSVNYILRNDLCVNETDYESLCV ELVNNKEKNIIINTTYRPPSGNLKKFKTHLKT LNKITNSRKHIYLVGDFNINLINYASNNNAKNF INTLLQYNFIPTINKSTRITNRSSTLLDNIITNNF HNNPLYTGIIQTDLSDHFPTFLATNNILJNNIST KSTIFRRQINENSLKQFKNCLKNCVDWNLILQS NXANNAYDLFLXQFCEQYEIAFPEIKKVINTKS LXNPWMSKGLLKSSKKKEKLYXKYLKKKTY KNEITYKNYKNIFEKTKKRSKKLFYDQLLXKT NGNTKKTWNLIKDIIGKNKRKKNNLPKNLLID GELVYDKSIIAEELNNXFLNIGPNLAAKIPTNSX PFDSYLKTYDKLMDETNLKPSELRTAFSSLKS KKSEGFDKISVDVVKSVFDIIEPSLFHVFDLSLK SGIVPDKLKIARITPIFKSGDKSNILNYRPISVLP	cd0165 0	670	948

				GTGATGA	CFSKLLERIMYNRLYNYLTENNMLYCKQFGFK			
				TGTCATA	KKQFNLRRTICCIVNNLVKKNNSTXHAVVELF			
				AGTGATG	NQISNAFNND CFTLG V FVDLSKAFDTVNHDILI			
				ATGTTAC	KKLEXYGVKNKNLLWFEDYPTNRKQFLYYNQ			
				ATATTCA	NNTIKHSITCGVPQGSILGPLLFILYINDLYLAS			
				TTTAGTG	NLLBLILFADDSNLFYSHRDIKALFKIVNEELC			
				TTAATTA	KVNEWFISNKL SLNVEKTKFMLFHKPNKSENI			
				TGAGATT	PLKLPDLIINKTYIKRESNINFLGVILDENLSWN			
				TTATTTTT	SHINSIEKKISKNIXMMYKAKPFLNIXSLKKLY			
				TCAATTT	FSFIHSYLSYCNIAWASTNYTKLKKLYSKQKH			
				CTTTTAT	ACRIVFGANRNVPCPELLSELGALNVYKLNIH			
				ATTCAAT	QVLLFMFKTKNELSPKIFQSYFNEIXHKYPTKF			
				TTTTTTTT	SDNNFVVPKYNLKLTSYSIQYRGPYLWKKFPX			
				TTTTTTTT	IVNKKKTLRDISLEQFKNESKRLLLFMDLNKFD			
				TTTTATAT	FKCFF			
				TAACTTA				
				TTTCTTA				
				AACATTT				
				TAGCAAC				
				TGTTAAG				
				CATCTTA				
				ACGCTTC				
				TTAATTA				

					ATTCTTA				
					TTTTTTAT				
					TTTTTTTA				
					TTATATT				
					AGTTATT				
					TTCTATTT				
					CATTTCT				
					AATTGTT				
					TCTTTTAT				
					ATTTATA				
					TTTGTAT				
					GTTACGG				
					TTATGTA				
					GAATTTT				
					TAATGGG				
					GCTCGGC				
					GATAAGG				
					CGAAATG				
					TGCCTTC				
					TTCTTGCT				
					CCAGCCA				
					TCAATTT				
					ATATAAA				

					TGTATAT GCAATGT AAACATT TTTTAAC GGCGAAA TAAATTA CACTAAA AAAAAAA AAA				
CR1	CR1 - 19_ HM		Hydra vulgaris	CTAA GCCTT TGAGC CGATA AGACG TGTTTT TGTTT GAGCG TAAGT TTTCA GGAAT TTAA ATTTTT TTCTTT	TGGGGTG TTTCCCCT GCTCGCT GTATGTG TACATTC TATGTAA ATTTACG GCTACAA CTATAAT AATAATA ATAAA	MTNLTPAQLKQVEDAIKKLIHAETAKWSAKIS ELEQVIKQQQLIIDKLIQNPNELTNTTSSIESS WVTVASKNIKKKQPIEQLKIINTFAIEQQQKKK REKNVIIYGIKESAKLSVADKHIDDKETVQNIL VTISDAKPKSVKRIKSTSSKPGPIIVELSDVSER NPLLTAACKLKNSDNYKNIFISPDLTEAERYLN YQLRMKRNELNSNINLKTSPFRYGIRSSGVVK LKNIQTTTKHRPLDTVYEA AESSNSKIYKPLHK HYYPSSTLLPANLKYKSKNLHNKKTNYSVNC YYTNATSLDNKMADLISISRTFDYHVIAISETW FRNDSVPNIDGYSLYRKDRQDGRRGGGVAIYI TDVFNSHEVNQFYTNLLEQVWVSVKFNNKHI LIGCIYRPNDASLIK YVAESFKLAHSLIKKPYD DCLIVGDFNFPNIKWSNDGIESFNPNNPTEYDF	cd0165 0	791	1046

			TTTTTT	WNIFQNSLFTQHITFPTFQTDYGSSNNMLDLLF		
			TTTTTT	TYNNSNSILNLS PSTILGNTTKGHFSITWRFLVSE		
			TTTTTT	LFTESSESQSFNYKKGDYDSISNFIASIDWEFLFL		
			CCGTA	HKTADQMFNILISNYRTACTYFIPKINKTKKKY		
			TTCTTT	DPWVTPHIKNLIREKKSAWHINCSTKWENSEL		
			ATCTTT	KIKYNRLNRSVKTEIFKAKKAHEKEIISKSQNN		
			CTTTTT	PKIVYKYINKQLKIKDRIRSLKDTSGRIIEALPDI		
			TACTT	VDLLNSQFKSMFTLESSENFPLLNKSGILSECIY		
			ATTTTT	NENLDFSISDIIQELADLKPYKSFGSDGLHPLIL		
			CTAAC	KICAKSFAIPLTKIFHKSFNSGVVPTQWKLANIS		
			TTTTTC	PIYKKKGDKLLAENYRPISLTSVACKIFEKLVK		
			TAACC	KKILKHINDNKMLSANQHGFLYHKSCVTNLIE		
			ACTAC	TFDIISSSIASHQFIDVIYLDFAKAFDSVPHNRL		
			TTCTTT	LKLRSFGINGKLLQWIKSFLVGRKQRTINNTS		
			TTCTA	SAWVDVLSGVPQGSVLGPLLFIYTNDLAKLT		
			CTTATT	ETSEIFADDTKLFSTLSQSNFVDLQDRLRLVY		
			GCTTT	EWSNTWLLKLNTSKCCVIHYGRNNPHYNYFL		
			ATCTA	QNNQGINQSLRSSTKERDLGVIFTSNLKWSAHI		
			ATTAG	QTVVSKSLQILGMLKKTFTFIDEKMALKLYKV		
			CTTATT	FIRPHLEYGVSIWAPFLKADMALVESVQRKAT		
			AATTA	KWAYSRLRHLQYCERLKRLNLPSLEFRRLRCDL		
			CTTTTC	IQIFKLIHRFDIVNDKYLPKYINSVSRGHHKIR		
			TACTT			

				AAAAT	REYVLNCLPRYNHLYNRS AKYWNNLPEQITL			
				TATAG	YSTVLSFKNRLDSWLWKKFGTNIYKTGCL			
				TAATA				
				TATTA				
				ACTAT				
				TCGTG				
				TCTAT				
				AATAC				
				AGTTT				
				TTTGTT				
				GTA CT				
				TCGCA				
				TACCT				
				AAATT				
				AAATA				
				TAATT				
				CTTAT				
				ATTTT				
				ATACC				
				TATAC				
				TTATA				
				TTTTAT				
				TATAT				

				ATATC					
				TTTATT					
				TATAT					
				ATTAC					
				TATTA					
				TAAAA					
				ATAGT					
				TATAA					
				CTGAG					
				ATTTTT					
				TTTTAT					
				GCTTC					
				AGTTC					
				GCTTA					
				ATTTTT					
				TTTTTT					
				TTTTTT					
				TTTAC					
				AGAGT					
				TATAA					
				AAAGT					
				TTTTTT					
				TTTTTT					

				TTTTTT					
				TTTTTT					
				TTTCTT					
				TTTCTT					
				CGCCA					
				AAAGT					
				TGCGT					
				TTGCA					
				CCTGT					
				TCTTTG					
				TTTTTA					
				GTTTTC					
				TGAAC					
				AATAA					
				GCTAA					
				ATTTTT					
				GCGCT					
				CTGAC					
				ACAAA					
				CTATA					
				TTAGT					
				AAATT					
				ATATA					

				TCCAT					
				ATATT					
				ATATA					
				TAAAA					
				TACAT					
				ATAAA					
				AAAGT					
				TATAT					
				ATAAA					
				CATAT					
				ACAAA					
				TAATA					
				TAAAG					
				AAATT					
				AAATA					
				TATAC					
				AAATA					
				AATAT					
				GATAA					
				CACGG					
				AAAAA					
				AACAC					
				ATACT					

				GAATC					
				TTTTTA					
				GTCCC					
				TGTGT					
				AATAA					
				GTTAT					
				TTGTC					
				ATAAT					
				TGTAA					
				CAAAT					
				AAATT					
				TGCAA					
				TAATT					
				GTAAC					
				AAATA					
				AATAA					
				ACATA					
				AAAAC					
				ATCAC					
				TGATA					
				TAAAA					
				AAAAA					
				AAAAA					

				TTCAT CTGTG					
CR1	CR1 - 69_ HM		Hydra vulgaris	TACAA GTTAC AATTT AACTAGC GTGTT TTGTGT TTACA AGTTC GGCACGT GAGTA AAATA A	CTGGTTT TTGTTTTA TAACTAT GGTCTTA TATATAA TTGTAAT ATATATT TTAAATT AACTAGC ATTTGTA TTATGTG ATTGTAG GGCACGT TGTCAGA GTCCTTC ATATTTT ATGGACC TGCGTGT CCTTTAG AACATGT ATCACTG	NLIKILIANMVNVQSTEFNVYKADQDKLIKSLI KTIDELKNCNLLLTERLNVLEKAKVVPSTSISW ASVVSGNMPPKRSTEQLHLMNAVVAEAKERE KRESNVVVFGEASKKADLLSAREEEKNSILH MMNSINAKIEIKQIIKLKSKSDKESPFVIVLNNR NDRNSFLKAAKSLKNSTEFNKVFINPDLTVAE RYKAKLLREECKQKNLENSEPSLFYYGIRNEK VIKISKQSFYNCKNQKYFSHGSLTSLVIPKSLQ KNHASSISKQRITAVNNKSIKKVICSKRSNFINS YNHLTTNSFNNFTVYLFNARSLANKLNDFFLY VESNKPALICVCETFFNDKLPNSMVC PKNYSIY RKDRVRIGGGVAIYCRNDVKS AIVSITQTDKDI DIICVDLIFGSSKLRLITCYRPPSYSLDDYIYLES MISIISNLCESVPQFVIVGDFNLPKIDWVNYVSP IEKCHNLFLT LVNNLGM YQYVLEPTRENNILD LVLSNIISLLSNISVECPFSTSDHNSVFFYINTFN YYQPETNSESFYDFKNTDIIAFNSYLSKISWDF DFS FVFTTEEYWNIFSNHLRKGIDRFVPIRTIYH NKKSRSPKYIYKMLIRKSFLWKRWSITKTTL HKKAFIQYASKCKKAISTHHKNVELKLVKENN LSKFFSYVNKNLNNPKKIYPLKTNSNFTTTNVE	cd0165 0	789	1046

					TTCTAAT AAATTTT AATCTAA TCTAATC TAATCTA ATCTAA	IAEVFNKHFGSVFTIDDGILPKINTFVSEHIKMN FVDFSAKIVYQKLKNLKPSTSFGPDGIPNILLK QLAHIICIPLSYIFESSFTSSSLPKQWLQATVSPIF KKGSTSDANNYRPISLTCTSCRVMESIVSSSIAD YLNINNLITPNQHGFLSKRSTCTNLLLESTNDWN KALDSNLITDVVYIDFQKAFDSVPHPKLLKKL AMYGIIDNLLHWISAFLSNRYQRVLVGNSLSN PTCILSGVPQGSVLGPTLFLLFINDLPSIVKDLN CSLMLYADDIKLYSSFKDADYSHDLAVAIYKV YLWSKKWQLKIANNKCFTHRIAPTSFCKDINY NYQLGDHNLTWSTNPKDLGVTMDSYLNYSN HISNIVRASNIRGYLILKCFKSRDPRVIVKAYTT YIRPILEYCSPVWSPNRTKMNNTVERVQRRFT KRISNLNNVSYSSRLSILGLELLEIRRLKQDMIT CYKILNNLVCLNKSSFFLFNNYIYTRGHNFKLR KLKCKLDVRKHSFSLRVVNIWNNLPSDVVNA KNIKSFKIKINSIDFEKYCSG			
CR1	CR1 - 81_ HM		Hydra vulgaris	TTTTTT GGACT TTTAT AGCGA TCGGA CATGT	CTTTTCTT ACCGAGT TGTTGAC ATGTGGA ACAATAT GACCTCG	MVNVQSTEFNVFKAEQEKLIALNRTIEELKN LNLLSERVKMLENAESVKSSPLLSWASVVSG KTPVKKTVEQLNLINVVTAEKKEREKRENNVII FGINPSKAIDKSSAIKEDKNSIQEVLNSINLKVE VKQIIKLKSKIDNQPPFVIVLNNKTDRNSILKAA KLLKNSSEFKNVYINPDLTEAERYKAKLLRDE	cd0165 0	881	1138

			TATAA	GAGGCTG	CKKNNLENSESLLYYYGIRNDRVVKIAKQLLF			
			GAGCT	TAAATGC	NNRKLSHMSHCATELANSSRISCEYLISKSSKP			
			TTTCAT	AAATAAC	DKIICTKKCRNKPINLNLINGNSNNGCNSNNGC			
			TAAAA	ATTTTAA	SSNPYSKNSVYLLNARSLANKLHEFFMFTEAN			
			AAAAA	TGTTCAA	KPAIICVCETFFTSKISDSIVCPKDYSIYRKDRIK			
			ATTTA	AAATAAA	IGGGVAIYCRNDIKSEQVQITQTDTDIDIVCIDLI			
			CGGAA	ATTAAAT	FSSHNLRLITCYRPPFYSIADVAYIELMTAFIYN			
			AAATT	CTTTCGA	LCHSVSQFVIVGDFNFPKMDWLNYSVPNEKC			
			CAAAT	CTTTAAC	HILFLNLVNNLGLHQFVQEPTRENNILDLVFSN			
			AAAGT	AAATATC	NISFLSLISVECPFSTSDHNTVHFSININDLDQSN			
			TAACT	TATTGTA	KNIESFYDFVNADIDGFQSYLLNINWDFEFSFL			
			ATACT	AAAGGAA	FLQLKIIGTFSPHIFAMVFINLFQLERVILIKKLGL			
			TTGTA	CAACTAT	RVFLFVFTVEDYWNVFSNHFCNGIHQFVPIRKS			
			ACTTTT	TTGTTTCT	NLNKKIKHYPSTFIGKMLNRKAFLWKRWSNTK			
			TGTGA	TTATAGT	NITHKKAYNKYASKCKKSLITYQMRVELKL			
			AAAAA	TTTGTTAT	KENNISKFFNYVNRKLNNKSIFPLKDKNNNN			
			GTATA	TTTATTTT	KITCSNEEIADVFNKHFSGSVFTNDNNYSPFINS			
			TTTGC	TTTTTTGT	YVDDSIGIEDCTISTRSSVLVVTKKSLTCLTNL			
			ATAGC	AGCGATA	AVSLQMITITLHLLIVMLMIQLVLKTVLFPPEA			
			GTACC	TTTTTAA	VYKSLKEIKASTSYGPDGIPNILLKLAHLLCIP			
			AACT	ATTTCTTT	LSFIFDASFKSSSLPKQWRQAFVSPVFKKGATA			
				GGGCACG	DPNNYRPISLTCTCCRMERMINVYVIKYLGH			
				TTGTTAG	HNLISLNQHGFLNKRSTSTNLLESANDWNTAL			

					TGTCCAT CTTTGTA TGGGCCT TCGTGTC CTTTAGA ACATGTA TTTATAC TTTGTTCT AATAAAT ATCAATC AATCAAT CATCAT	DNHQITDVVYIDFQKAFDSVSHTKLLVKLAS NIKGNLLNWISAFLTDRYQQVKVGNSISNSIRI VSGVPQGSVLGPTLFLLYVNDLTAIXNNLDCC VKLYADDIKLYSSYCDMDLSQDLVVALNRLL QWAETWQLTIACSKCFALRIAPPSLCKGIKVN YKLGEYSLAWSSNPKDLGVTMDCNLNYKKHI FNVIHQANSRAYLILKCFKTRDPCILVRAFTTYI RPILEYCSPVWSPHQIGLIKSVESVQQKFTKKIR SLRNLSYHNRLQLLGIESLEVRRLKNDLVTCFK IFRKLVCTDQSMFFAIDNNSHTRGHNYKISKQS CRLDIKKI			
Dap hne	Dap hne- 78_ HMa		Hydra vulgaris	TGGTA AAGAT GGCGG AGTTA TAGAT AGAAT TGTA ATTTA TCAAA TAAAG TAATA	AGACTCT GAAATAT GTTGTTA TGCTTTT ATTATTA CTACTYT CTTGATA CCTACTA GTTTTATT ATTAAC TTTTGTTT	MLIRNLCKFCNKTVVDSHKGILCNYCSSWLHP KCNYINKHTYSLIEKDSTDWYCSDCINKNLPCF SLSDFELSLTLSSQTIPTNNLKSLLKSPPHFKKLF NNINKISNKSINCKYYDILELNKVLNNNCNL HLNIASLPYHIDGLRSLISSLTTLPIVIGISESRLN INDSSKTDVNINGFTIEHSPTEAKKGGALLYLR SNLNYFIRSDLIVYSPKYLEIFVEIVNPLKQNI VGSIRHPSMNSNEFITDHFNPLLERLSLENKQ VVLMGDFNMDLLNYNESKVISSYLDTLCSYSF FPSIIQPTRITATSKTLIDNIFLNFQTPDLISGNLT ISISDHMAQFICIPCLSPKKTKEVIFRRTFQNF	cd0165 0	635	891

			TATCTT	TCATATT	NKKFIKDISNADWQIQITNNDNVNESLHYFFKT			
			TACAT	TTTATTGT	FNKILDQHAPYKKLTAKQIKLKSKPWITNGILK			
			GACAG	TATTATT	SISIKNKLYKQYIKAKNHNIKIHIFNKFKWYRN			
			CAATT	AGTATTG	NICNLLKLSKKSYYINFFNNNLXNVKNTWKGI			
			AAGAG	ATATTAA	KEIINIKPSSFSSKSINLKFNGNTISDSKAVSNIF			
			GAAAA	TAATATT	NNYFVNTQQTLLTKIVSPKCNFSDYLNPNINS			
			ATATT	CTTCTAT	FFIDPVTENEVSNLKSLQSKKSLGINSIPTFLLK			
			TATAG	CTTTATT	LVSHIISKPLCTMINSSFENGLFPDVFKIAKVIPL			
			AAAAA	AGTAAAT	YKKGSRDHTNYRPISSLNISKIFEKAMHNRL			
			AAGAC	TGTTATTT	YNFLDKYQCLYKHQYGFRSKHSTTHALIEITE			
			TAAAA	TTATTTAT	VIRKALDDKLFACGVFVDLQKAFDTVDSILL			
			AGTAA	AACATAA	SKLEYYGIRGISLQWFTSYLYNRFQFVSINGNT			
			AAAAA	TAAGTAT	SSLVKCTTGVPQGSVLGPLLFLVFINDLNLSLK			
			AATAT	TATTATT	YTTSYHFADDTNLLLINSKSLKKINKYINHDLN			
			ATCGG	GCTTTTTT	LIQWLRSNKLSLNTSKTEIIIFKSKQSKISKHLNF			
			AAAAT	AATTTTTT	RLSGQKINPVMYLKYLGIKIDSNLTFSQHLKDL			
			TAATT	CTTAAGT	SIKLCRSNGMLAKIRHYVNLETLINIYHAIFGSH			
			AGAAA	TTATTTA	LRYACQVWGQIRTQEILRLSNLQNKALKIIFYQ			
			ACCAA	CTACTTA	HYQSNSDHL YFLSKILKIDDLVQLLNCHFVFN			
			AGTAC	TACTATG	YKQKCLPSTFTNFFT VRESCRYHLRSANSYNL			
			TTAAA	TAATTGT	HVPNHSSVKYGLKSIKYQSVKSWNSLPSPKSI			
			GTTTA	TATTACC	DSITNFKNCLFN YFLARYYT			
			GAGAT	ACTATTA				

				GTGTA	TTGCTAT				
				AGAAC	TATCTAA				
				TTCAT	TATTGAT				
				AGACT	TGTATAA				
				CAAAT	TATCTTT				
				AAATA	ATTATTA				
				TAATT	TTGTTAG				
				AAAGA	TATTATT				
				ACAAG	ACTAATA				
				ATACA	ATATTCC				
				TTACA	TTTTCATC				
				AAATT	GTCATCT				
				AATAA	TATTATT				
				ACTAA	ATTGTTA				
				GAGAA	ATATTAT				
				AGTAA	TTCTATA				
				AGAAA	ATTATTA				
				GAATC	TCAATAC				
				AAGTT	TATTATT				
				AATGA	ATTGTTA				
				AAGRA	TTATAAT				
				AAAGA	CACTTTT				
				TCAAG	ATTGTAA				

				AGTTT	TGTTGCA				
				TTTAC	GGTTTTT				
				ATTTG	ACTTTAG				
				TTTTG	ATTAGTA				
				ATTAT	CTTGTAC				
				TTTTTG	TATTTGT				
				CTTTA	AAAAGTC				
				AATAT	CTTGAAA				
				TATTT	TTTTACA				
				GTAAA	ACTTATT				
				TTTAA	TCAGAAT				
				GTATT	ATATATT				
				TTCCA	GAATTGA				
				GGATT	ATTGAAT				
				CGTTTT	ATATATA				
				TATTA					
				TTGAC					
				TTTTTG					
				ATTAT					
				TATTA					
				TTAGC					
				AGTTA					
				TCATT					

				GTTAT					
				TATTA					
				TTATA					
				ATTAA					
				ATATT					
				TATCA					
				TTATT					
				ATTAA					
				ATAAT					
				ATAGC					
				ATTAT					
				TTATC					
				CTATA					
				TTTATC					
				CATTA					
				TTTAC					
				CTCTT					
				ATTAC					
				AATTT					
				ACATC					
				ATAAT					
				TTATCT					
				TATAT					

				AAATG					
				TTATTT					
				ACCTA					
				ATATT					
				ACTAC					
				TTATCT					
				CCTTCT					
				TTACTT					
				ATTAT					
				TACTT					
				ACCTT					
				ATTAT					
				CATCA					
				AATTA					
				TTTAA					
				ATAAC					
				ATATA					
				TTATTT					
				ATATT					
				ATACT					
				ATGTC					
				TAAAG					
				CTACT					

				AGAGG					
				TATTA					
				ATATA					
				TCTAA					
				ATTTA					
				TATTCT					
				AAATC					
				AAGTT					
				AATAA					
				TATTCT					
				AATTT					
				TATCT					
				AGCTG					
				TAATA					
				CTTATT					
				CTCCT					
				GTGTA					
				AAGTA					
				TTTAA					
				ATCTC					
				TTATA					
				TCTTA					
				CTATC					

				ATTTCT					
				TTATA					
				ACTGA					
				TCATT					
				CCTCA					
				ATATT					
				TAAAA					
				TAGCA					
				ACCAT					
				TTCTA					
				GATTA					
				ACATT					
				TCTTA					
				GTTTT					
				AATCT					
				CTCAG					
				TTTCTT					
				TATAA					
				ATATA					
				TAGTA					
				GATAT					
				CTTTAT					
				TTAAT					

				TATTTT TTCTCT ATACT TTGTTT AAGCA TAATA AGTC					
Ingi	Ingi- 1_H Mi		Hymenole pis microstom a	GGGGT GTGCA ATAGA CTTAC CTCCA CGTGG TGCAC CCTGG GAAAC GTATA CTCAT ACGGC AAAAA GTCTA TAAAT CGGCA	TAAAGCC CGAACAA AGGAGAA GCAGTGG ACAGCGG CACTCTC CAACATA GCCGACT GGCCTAG ACTCGAA GCTGTTG CAGAATT TAGACTG CGTACCG GACACGA TTGCCTG	MTSLKRNSNRRNRPCQGEPQLLQILQWNAGG MSQDKKIQIKILDTYNIDIFTIMESNLTDKLT YYQFPGYTLHLLPKDRQVASGILTGVKEGLTS HCEIIKSMGSMQDICEIIRLNVWKSQNHFKMY AVYNPPQNRPNFDLLNISHKTIVLGDFNAHSTR WGYKNRNTAGKEIEDILNSSPLELIYSDEDPAT YLHYNGTRTTPDLLLVSSDISELTQRKIIDDPGS GHKPVIASITINSKSMTPKMPTKVLWKFKKAD WPKFTNLLETENASPINYNQHPDKLCXNITNI IIKCAKKTIPRGKVKHRYVFWSKNLEELKRKR EVL RNTAEQTGKTEDVQAWRRQSAVLRQAIL QAKRTTFNSFISNINYQNDSQRTFKFLGNLLNN KKKPKKEPIRLNNKLLTTDSEIANSFARFYSSXK QKKNPFVRNSSRDIRKQVTKFDEIKEDTAIPHK VLTEPFSSHELNAAIKQLKCKKSPGEDCIHPEF LIRMGPKAKEILLTLFNKIWETSLVPTQWKVAI	cd0165 0	521	782

				GTTGA	GCGAAAC	VIPILKKGKDPSNFDNYRPISLTSMLAKLMERM			
				CTTCA	ATCTCCA	VNTRLTWFLETNNILGNEQAGFRPQRSTNQQV			
				AACTA	CAGAATA	ATLSQHIKDALDARNTLTAVFIDFKSAYDLVW			
				CT	GGAGTAT	KENLILKLAKIGIRHNMLNWIKAFIQRTCKVR			
					ACGCCCA	YGNDLSKSNLLQTGLPQGAVTSCSLFNVFINDI			
					ACCAACA	AELVQTVTGVKCLLYADDLVWYSAPKKNA			
					TGCCCCC	QKRTEIALNYALKLLANWCDNNGMVINTSKT			
					TATGTGA	AYQTFSLAHHSINPHLMYKDTPLEQTNESTYL			
					CCTTCAG	GVTFDTKLTWKNHIAKIAERASNRLNVLKRLA			
					GAGGAAA	GSLWGSARSTLNTTYKMFVQPIMLYCCETLVT			
					TGGAGAA	ASEAILKPLEKTHNQALRLITGGIKSTPIDAMLL			
					GACCCAT	VTGNSTIRSSVEEKALTLYEKLLRIPSDTFWSS			
					CTGATAC	YENRPRHLKTQVGLIQKAIELKKRLQTTDKPES			
					GGTGTCC	LSLPQNPLADNEVQFCTQLLGHSMTADTPPDQ			
					AGCCCTA	MRLLALETINANYPADQWLQVFTDGSYVENQ			
					CAGACAG	ANVGAGVYSELFTHFAAAGQNRSAFEGEIEAI			
					CGACAGA	KIALGQLCCRDTKFTNAVILSDSQSAIQSIGNRE			
					AACCCAG	PPKTAEIHECKKLYQLLKEKNKTIVLQWIPGHC			
					AGATACT	GIIGNERADTLAKKGTTILQANIFLHNEDLNQE			
					GGGAAGC	RIQDSKKQRT			
					CAGAAGT				
					CAACTAA				
					AGGGCTA				

					ATGATAA CTATTAT AAAGCCT GCCATT GATATAA TAATAAT AA				
CR1	CR1 - 1_IS		Ixodes scapularis	CATTT GGCCG GCAGG GTGCG TTTTTC GCCCT TTTTTC GGCAT TCTTA AAGAA CTCTC GGACA CGGAC ACCTC GGTCG ATGTA	TAGTCAC GACCTTG TTTTGTTG CCTGCAT TTATCCTT GATGACG AGCGGTG TACCACT ATTTAGG CCTTGTG CTGTTTTT GGACACC CAACAAA AATAAAG ATTATTA TTA	MECTLYYQNVRLRTKSSEFLSGVIGNSYTIIS LTETWLEDTPSSQYFPPCYSVFRKDRDYAATR QRYGGGVLTAVDTSLCARRRYDLELYSECVW IEVTCRDGLNFLIGNYYFPPHLSSDV FARHFRD LEGKLDVTKFRVHIYGDFNLPIDWLANYNLI TSDVTSKASSLFDIYFNGFEQRNTIQNSAGN VLDLVLVSTPINYISPITDPLTRVDHFHPPFAVS FCFPLRKQALCSYEYLLYSSGDYLNMYKYFQC YDWKPILNNKCADSAAEALTEVVSSAINDFVP KRVSRLNKYPAWFSKELRSCLRRKLHYHRLFK RSGLDKWYAKFSECRALAKWLFKKDESAHLH GLES DMYAKPQNFWKYVPTKKKRNSSVAPM RNSSGYTCDSIEIANMFAEHFKSCFSSASVPDN NATCHGTSDFLSFIPIDEPEIRESIRNMKSKLSA GSDGIPSFIVKGCAELLTPVLLHIFNLSIHSGIFP SIWKRSVVVPIHKAGDPSTVN NYRPVSLLCFS	cd0165 0	493	745

				AACGC ATAAT CAGGT ACC		KVFEMIMHKRLYRHFRRHKMNPEQHGFLKGKS VETNLC SFLDYAVPLVCNREQVDVIYFDMTKA FDKVNHSTLLTKLDSYGLCLTYCQWFKSYLTC RTNYVRVSGSMSDYFSSPSGVPQGSNLGPLL LMFVNDISLCAKNSKVLLYADDIKLFRRVQSH SDCVLLQEDALQISQWCAKNYLPLNEQKTRT MSLTRKRDSLAFSYTIGGSVITRVS VIRDLGIFI DSSLNFNHHVLTICNASFRNLGAISRFTRKFSSP VCLVRLFCCLVRTKLEYASSVWNCINLTASCCI ERVQKRFIGIHYDRYFDRTCYYSYEQLLKKLGL CTLLDRRRLRDL MFLHKVVNGAVDSPLLLSTV CFHAPLR SNRFQSPFYPSALNVASPLTRMQLEF NKLDQSAVDIFSDLIVFRHRLESHICT			
Ving i	Ving i- 35_ LMi		Locusta migratoria	AGGAT GTCGG ATGTG CCTTT AATAG CCAGG AGATC GTTCT GATTG TCTGG	TGACTGT GTTAATT TATTCTTT GTGTGCT GTATGCA TCTTATA TACCTTA TTTTATGT CTCTTTTG TATTCTT	RRGIFFISYAAPLGGSSADTLRTTALGLVYSTA EYCAPVWMNSAHTRKIDVQLNGTMRIVSGTIR STPTSWLPILSHIPPADLRRSNALLREYRKITNN PRLPIHHDTPDLERNRLRSRHPTMKTAKELHN TDFNLIESWKERWRQACPEQCQALPCITERPPG YELPRKAWSSLNRIRTNHGRCADSLYRWGKIP SPNCDCGADRQTIRHIIEDCEMRAYDGNVSDF LTAKDAAIDYIRNLDVCI	NA	NA	NA

				CTGCC	GTATGTT				
				AGTTA	TTACACA				
				GCGCT	TTTTTGTA				
				AGTTC	TTTTTTGT				
				ATTTA	ATGTATA				
				GTTCA	GTTGTAT				
				CGTTTT	AAGACCT				
				ACAAC	GCTAAAT				
				GCTAT	GATGTGC				
				CCGGG	ACTGCCA				
				GTGCA	TACGATA				
				CCACG	AATAAAT				
				GGGAG	AAATAAT				
				GCGAA	A				
				CTAGC					
				ATTAC					
				ACCCT					
				TACAA					
				ACGCC					
				CGACT					
				TATAG					
				ATAGC					
				ACGCT					

				TGCCG					
				CAGAA					
				CATTC					
				CAGGT					
				GACAG					
				TGGGG					
				GACCA					
				AATAT					
				GCCCA					
				AAGAA					
				CCCTA					
				AACAA					
				CGGAT					
				TACCA					
				CAGGG					
				CTCCG					
				TCCTA					
				GCCCC					
				ATTGC					
				TATTC					
				AGCAT					
				GTACA					
				TCGCT					

				GACAT					
				GCCCC					
				AAACA					
				ACTCC					
				AGAAA					
				ATTTG					
				AATAT					
				GCAGA					
				CGACT					
				GGGCT					
				TAGCG					
				TCAAC					
				ATAAG					
				AGCTT					
				TGAGA					
				CAATG					
				GAGTG					
				ACCTC					
				GTCGT					
				CCAAC					
				TTGGC					
				GACTG					
				GCGCA					

				TATTTC					
				CGGGG					
				CCGAG					
				CCTAC					
				AAAAC					
				GGGAC					
				GTGCT					
				TCCAT					
				ATCTG					
				TAATA					
				AATCA					
				GCAGG					
				CAGAC					
				GACGA					
				CGTCT					
				ACTTT					
				GAGGG					
				AACTC					
				GCTCC					
				GGCAT					
				AACGG					
				TATCC					
				AAAGT					

				ACCTC GGAGT CACAT TAGAC AGAAC CCTCA CCTAT AAGCG ACATC TGGAG AACAC TGCAG CCAAA CTAG					
Ving i	Ving i- 24_ LMi		Locusta migratoria	CGACT GTTCG CGTCC GTAGC AATCC TCAAG CTACA CGTGT CCCCT	TGACTGC AACATCA GGTCTAC CCAGGCC AAAAACA GACCTGT ATAAAGT GACAGAT GATGCAA	MIMKKKQQTPESVLTRELGSTTRICQLNVEGL SSSKSQFLSKFLMENDVDVAVLQETHIENXEQ QRSRGKIHGYDLLGAVYHRHYGLATYVRSSIE NAALVSLSADCFQQIVTKIGELTVVNIYKPPQT EWPQNVIEHPHTIYAGDFNSHHTDWKYRTT DINGEKLSNWAEENHLHLVFDKDKNTFKSAI HNTETNPDLSTTRNSQSHPIPTRRKIXESFPHS QHRPILLEIGINIPLIRSLRPRWNLQKANWDKF SNDIDQTLRWIPARWYNYDRFVNIIIAAKRYI	cd0165 0	481	707

			GGGTG	TTTCATG	PRGYRKEYIPGWNEETEELHQRFLQENDAEIA			
			GTGCT	GCTGGAT	DKLLRNLDVARRKKWEETTEALNFTKSSRKA			
			AAAGA	GGTTTAG	WNLLRKLGDASAPEREPGVSPNQIASRIVSLS			
			ACAGT	ACATTAA	RAPRDKKRTTFVKRHLKSFRQSVEAPNKISRPF			
			CACTT	TATATAA	TVEEVNTAISKIKKGKAPGFDGVLPEFLIHSGR			
			CGAGG	GGACCAT	RTREWLSQFFSHIMQTGLVPPXFKKAKIISLNNI			
			CACGA	GTTCGCT	FVKPGKPKDKPESYRPISLLSTTYKLLERLILGR			
			AGTCA	GCTTATC	IEPILEEHLPEVQXGFRHNRGCEDQVLSLTSHIE			
			GAGCC	TTTCTTA	AGFQQXLKTTVAFVDLTVAYDTVWKDGLLY			
			AGAAG	ATTCAGT	KLYRTVPCQKLLRLFNMLRDRTFQIIMGDKA			
			TCACG	TAACATA	SRVRKINNGLPQGSVLAPALFNLYISDMPETKG			
			GCAGC	AATTGTG	RKFGYADDLAIIVQAKXMEDADTLLSQDLEV			
			CAGAC	TAAATTT	VGKFYADWRLIVSTSKTETTSFHLTIGKQQNL			
			CACTT	ACTGCTC	XKYFSTIKCSPIISSQSILVSHWIAPSPSGNTWK			
			GTGGA	ATTTTCAT	QRQPKSTLAITSCTNCVAAGGEHMLPLFSLNN			
			CTGTC	TGTTGTTT	REATKSPKVLLNNKVLPHNQFPKYLGVTLDRDRT			
			CTGGC	TTCTATT	LSFRKHLEATAAKVNTRNNILHKLCGSRWGA			
			TATCA	ACTTATT	HASTLRGSALALVYSAAEYCSSSWLNSSHVSK			
			AACAC	TTAGTTT	LDARLHDTMRTISGCIKSTPVQWLPVLSHIAPP			
			TTGGG	CCTAACT	PLRRKGCLVRQLKRIQSDPDLPVHKDIPHLSLR			
			CTGAT	TAATGTT	RLKSRKPPLVSAQQLQDADFNITEQWRTDWST			
			TAATG	CCTTTATT	VNFTALWPDPTERTGSGNFPADNGKRLTGSEH			
			CA	AGTGTTT				

					ATTCAAT	LTESVRILCTSGAYRKSSVRLWCFASDNRPYC			
					TGTTTCAT	V			
					CTTACTG				
					TACGATG				
					TATATAT				
					CTTTCAT				
					AAATTTG				
					TATTAAT				
					ATTTTAT				
					AAATATT				
					TTTTTCCA				
					AATTCCA				
					CTGCATC				
					ACACATG				
					TAATGGA				
					CTTTTTAC				
					TGCATAC				
					CTGTATA				
					CTTCTCT				
					GTAATTA				
					ATGCACT				
					TAAATTT				
					TGATATA				

					AGCTGTT TGTTGTC TCTGCTT CCCACCT GGAGAAG TGTCATT GTGTAAT GTATCTT CATGTCC AAGTGTT TATGTCA TACGATA AATAAAT AAAT				
Nim b	Nim b- 38_ LMi		Locusta migratoria	CCCCT CAGGG GCTCA GCATT AATTT GCTGA GTACG GGCTT GGCGA	TAGAAAT TGGTCTT CAGGAGA GGGACCT CCGATTA AGTTACA CCAGAAG GTACTTG GGCTTTG	MRCRSSHPSILQWNCGGFFHHLAELQQLMCHS LLCIAYRKRGFLFRGPLPYVATGAISKIAPSVT GSLEVCASMSLPLSPVRCAPHHESQGLWRCVP LCPCHCPQCAVPLTTNLEAVAVRIRVRQELTIC SLYLPPDEVVCLADLVPWSLNFPRRFSFWGIL MPTTLCGVGRRPRLVVGMPRALVAQLPPPFL LGDFNAHHPMWGGTQTSARGRDVEMLLTQR NLCLLNTGAPTHFSVARGTYSIDLSICSPSLLP HLNWNVHDDLCGSDHFPIFLSLPETQAANRLP	cd0165 0	568	826

			ACCCG	GAACACT	HWSLHRADWPAYASAAVTTPLPAGIEVAVQ			
			GGGTC	GAGTGGG	QVTKAIVSAAEKTVP TASASLRRKSVPWWTQE			
			CTGAG	TTGACTC	IAEAIRERRRALQRTGDTLLGHPGTSARASHR			
			CCGTG	GCCTGCC	RHPSVENLIAFKRLRARARRLRQSKKDSWERF			
			GGCTG	CCAAACA	VASIGSRASPSRVWTQVRRLFGYRPPTGIPSLTI			
			TGTCG	AGCTCCG	QGALCTDPQAIAEYMAEHYARCSATETYSADF			
			CGCCA	TGCAGTG	RTFKERVERKQLSFRTRRQELYNAPFSEWEFE			
			ACAGT	AAGCTTT	TALAACPDTPGPDRIHSQFLRHLSAEFQHQLL			
			CCGCT	CTACAGC	SLFNRIWADGVFPSQWRESTVIPVLKPGADPL			
			ACACC	TCCATGG	VADSYRPISLTNVLCKLLERMVSRRLMWILED			
			GTAAG	AAGTCCT	RGLLAPQQGGFRQGRSAIDNLVSLESAIRTAFT			
			CTCTG	CCCTGAG	RREHLVAVFFDLRKAYDTTWRHHILATLHW			
			TGCGT	GGCCTCT	GLRGSLPIFIQNFLSPRSFQVRVGTCLSSPHIQE			
			GCTTC	CGGAAGG	NGVPQGSVLSVPLFLIAINGLAAAVGSSVSSSL			
			AACGA	AGTCAGT	FADDFCIYFSSATIGVAERRLQGAIRKAQSWAS			
			CCATC	TGTCCTTT	AHGFHFSGSKTRVMHFCRRRALHLDPELYLG			
			GTTTCG	GTCGCCT	NELLRVEETHRFLGMVFDARLTWQPHLRQLR			
			GCGCG	GCGTATT	QKCWKHLNILRCLGHTAWGASRPTLLQLYKA			
			ACGGT	GGTCACA	LIQSRLDYGSIVYDSAAPSALQVLDPVLHSGVR			
			GGAAC	CGTGGAT	LATGAFRTSPVTSLLAEAGVPPLRFRRQRLAS			
			GTTGT	GACACAT	YAAHVHALPHHPNYRLLFPDVALPSRRRPRS			
			ACGTT	GGCCATC	GLPIAVRVRSLSSLESPIPPTFHYPSPRAPPWS			
			TCAGA	TACTTCG	VPVRSFGIPSYSPHFSLPFPCSSLVRTRPQLRLD			

			GCCCG	ACGGGAG	LSRDSKASVAPLNLRQFQAVLGSFHGAEIVY			
			GGGAT	GATCCTC	TDGSLVDGRCGYAYAHDEHIEQHSLPCGSSIFT			
			CTTGG	CACAGTG	AELIAIRRALEHLRACPGESYVICSDSLSGLQAL			
			CTTAA	TCGCTGT	DQCFPKHPLVRSIQESVTFLFSLAALWSSFGR			
			TCGCC	GGCACTT	DTLEFRGTNAWIGWLSRQLLQAP			
			CGGGT	TATTGAC				
			CGCGA	AGTGGTC				
			GGACG	CACATTC				
			GTACA	TGTTGGA				
			AACCT	CTGTCCC				
			CTATA	CTTTTAG				
			AAAAA	CTGCCTT				
			AAAAC	ACGACAG				
			CTCAA	TCTTTTG				
			TCTCA	ACCTGCC				
			AGGTG	AGCTGCG				
			TGCTA	CTTCCTTT				
			CTGCC	TATGCTT				
			GAGGA	GGTGACG				
			GATGC	ACGCTGC				
			ATGGC	TACAGCT				
			TGTTG	GATCGTG				
			AGGTT	TTTTACG				

				GAACA	TTTTATCC				
				GTCGC	GTGCAGC				
				TAGCG	TGGGTTT				
				AGCAA	TATCATT				
				CCTCT	TGATCTA				
				GAGGA	AAGGTGG				
				ACCTG	GTACCTC				
				CCGCA	AGCCTGA				
				CCTCA	TCTCTCA				
				GTTGT	GGCACCC				
				ATAAG	ATCTCTC				
				GCTTA	ACAGATT				
				CCCAG	TTTACTCT				
				GCAAG	TCTTATTT				
				CGGGG	TGTGTTG				
				CTCTG	TCTGACC				
				CCTGG	TGCTTTT				
				GTGGA	ACTCGTA				
				CTTAG	TGTTTTA				
				TTCCCT	GAATGTG				
				AGCTG	CAGAGTG				
				CTCGT	GCTGGCT				
				GGGAC	CATCCTT				

				GAAAT	TATTATC				
				GAAAC	CTCGTGA				
				GAGCC	TCAGCCA				
				AAATA	GCCACAA				
				GCGTG	TATTCTG				
				GTGGA	CTGTAAC				
				CGTAA	TTTTTAGT				
				ACCAC	CCAGTTA				
				CGTCT	TGTTTATT				
				CCAAG	TCCTCTTT				
				CGTGG	GTGTTTT				
				AAGAG	CTTTTTG				
				TTGCTT	GTGTTCC				
				TAATG	GAGCTAG				
				CAGCG	TG TTCCTT				
				ACGTA	ATATTGT				
				TGATC	GTACCCA				
				CCCGA	TTGTGTG				
				ACGTT	TTTCTCTT				
				CCCTT	CTGTGTG				
				CCCTG	ATTCAGC				
				GCCAC	GCTCGTT				
				CCCTG	TTTTCAG				

				GGAGG	CTTGCAT				
				AACGG	TACACTG				
				CGGGC	TATTGTC				
				TAAAG	TTTGCTC				
				ACAGG	GCAACAG				
				GATGA	GGGACTG				
				GCTTA	ATGACCA				
				CTCCC	GCAGTCT				
				CAAGT	AGTCCCC				
				ATCTT	TCTAAAC				
				GTTTG	AACAACA				
				CAGTC	ACAAC				
				GAACA					
				GATGG					
				GGAAA					
				CTTTC					
				ATGTC					
				GACGA					
				AGCCA					
				CTTTTC					
				TTCGT					
				GAAAA					
				CCTTG					

				AAGAT					
				AAGTT					
				TGGTG					
				AAGTC					
				TCGGA					
				TGTCA					
				AGATG					
				AGGAG					
				CGGTT					
				CGGTA					
				CTGAT					
				CGAAA					
				CAGCG					
				TCATC					
				CGCCC					
				AATCG					
				AAAGC					
				ATTGC					
				TTGCTT					
				GCGAT					
				AAGCT					
				GGGTG					
				ACGTT					

				CCGGT					
				CAGTA					
				TAACA					
				CCCCA					
				CAAAT					
				CGCTA					
				ACATG					
				ATCCA					
				GGGTC					
				TGATC					
				TATCA					
				CAGGG					
				ACCTT					
				CTTTTA					
				CAGTC					
				GATGA					
				GAGCT					
				ACGCG					
				CCAAC					
				CTAGA					
				GCGAC					
				GGGGT					
				GTCCA					

				CTTCG					
				TCCGG					
				CGCGT					
				GCATA					
				GGGGT					
				CCGAA					
				GGATA					
				ATAGG					
				GTAGC					
				CACCG					
				GTGCT					
				TTCAT					
				CTTGG					
				CCTTT					
				GAGGC					
				GATAC					
				ATTGC					
				CCGAA					
				AAGGT					
				CAAGG					
				TCATG					
				ATCTA					
				CCGGT					

				GTGAC					
				GTAAA					
				GCCTT					
				ATATC					
				CCTCC					
				CCCTA					
				TGCGG					
				TGTTTC					
				AAATG					
				CTGGA					
				AATTT					
				GGACA					
				TATGT					
				CCTCC					
				AGATG					
				TAGCT					
				CCAGC					
				CCCAC					
				ATGTA					
				GGGAA					
				TGCGG					
				GCGTC					
				CATCA					

				CACCC					
				GGACG					
				CTCCG					
				TGTAC					
				TCCTC					
				CACCT					
				ATCTG					
				TGTAA					
				ACTGT					
				GGGCA					
				GCAAC					
				ATCCA					
				CCCTG					
				TTCTCC					
				AGACT					
				GTACC					
				GTGCT					
				CCGCA					
				AAGAG					
				CGAAA					
				GATTA					
				TGGAA					
				TCAAG					

				ACCCT					
				GGATC					
				GACTG					
				ACATA					
				TACGG					
				AGGCA					
				CGTAA					
				AAAAT					
				ATGAA					
				CGGCT					
				TCACC					
				CAGTT					
				TCCAT					
				GACGA					
				CGTCT					
				TTCGC					
				TGCAG					
				TTACA					
				TCGTC					
				GCCAC					
				CCTTCT					
				GATAA					
				CACCT					

				ACCGT					
				GCAAC					
				AACCT					
				TCTGT					
				GGGCC					
				CAAAG					
				GGCCG					
				TCAAA					
				GCTCA					
				TCTGC					
				CTCCT					
				CAAGG					
				TCGGG					
				GGCAC					
				TCCTCT					
				CACTG					
				CCCCC					
				CACTA					
				CCACC					
				TGGGG					
				GGCAA					
				CTCCA					
				CCCCC					

				ACCAT					
				CGGGG					
				ACACC					
				AGTCC					
				CCACC					
				TCCCA					
				GCCGG					
				AGAAG					
				GAGCA					
				GCCTC					
				CTCCG					
				GCTCC					
				CCTCG					
				CTCGG					
				AAGGG					
				GTCCC					
				TCGGG					
				GCACT					
				CCCTT					
				CTGAG					
				GCTTC					
				TCCTA					
				GTGCT					

				CCAGC					
				GATCC					
				AGCCA					
				ATGGC					
				TGAAG					
				GATCC					
				AAAAC					
				CTGCT					
				GGTAG					
				AAGGG					
				CTTCC					
				CGCTC					
				TTCTTC					
				TACCC					
				CTGGC					
				GCCGC					
				TACAG					
				AGAAG					
				CCCTC					
				CCGGC					
				CCTGC					
				AAACC					
				TAAGG					

				GCCAG CGAGA GGGAC GGAAG AAGAA GACAA AAGAC TCTTC GGTGG CTCTA AGCCA CCGCT ACTCA TCTCT ACATC GGACG					
Ving i	Ving i- 46_ LMi		Locusta migratoria	GAGCG AAAAT TATAG CCTTA CTMAA CCAGT AGCAA	TAGTTTG TTACACG TGTAATA TTATATG TATCTTT GTTATTA AAACATC	TKTWCMMGGFPLTLLFICSQTSNRTVIXXEQA GFRPNRNCIDQVLSLTTFIEAGFQNKEKTSVVFI DLTAAYDTVWRPGVILKLLQAIPSIKINHLNE MLSNRCFQVIMGNNISSPMKLNNGLPQGSVLA PLLFSLYISDMPETKSRKFGYADDWALATRHK DLETTESILTNDLSTIGEYFRKWRLQPSATKTE VSTFHLSNKLANKELTVHFEGKVLRRHNRNPKY	cd0165 0	26	233

				ACSAA GTGAT AA	AAATAAC TGTAAT TTTGTGA TGTAGCC ATACGCT AAATAAA TAAGAAA TA	LGVTLDRTLSYKQHLAKTAAKIKTRNNILQKL CGTSWGSSASTLRTSALGLVYPVAEYCSPA WLNSSHTKLVD AQLNHTMRLITGTIRPTPTFWLPI LSHIQPPSLRRKNALVSEYKKLQKNPHLPVHE DIPSIRKARLRSRHPPLETARTLV TNSFNTDDC WRQTWQINATPEQLKMP CISTKPPGFDLPRKD WSALNRIRTNTGRCADLLYKWK KIASPSCDCG AEKQTVYHV VQECPLRAFAGDQTEFLMVTEG SLDYIHGLDISL			
Kiri	Kiri- 5_L Mi	.	Locusta migratoria	ATGTA CATAG TCAGA AGTGC TCAAC TGAAC AGCAC ATGTC CTCTTC TCAAT CGGAA TTACA GTTAC TCACT	TAAAATA TG TAGAA CGCCTGG TTAGGTG TAAGAGA GGGCCTG AAGGCCC TAATCTT GCCAGGT TAAATAA ATGAGTA AATAAAT AAA	MMNSRDAAHRRFKASTKPEHYEEYRKL RNKV KQCIRNAKIRYARSFVCSNTNPATLWKNLRSL GVGKATPESVFQISADELNEFFSAPINCQAQIN NRPEESPSHITNDDTFHLKHVTINTTRKAIMRIS SEAIGNDGISITMIKNVADSLLPVLTDIYLISLSF SANQHIACVSKALEYIVHDQITEHLCKFSLFDK YQSGFRKNHSTNTALIKVTDDLKFAMNNREA TILTLLDFSKAFDTVNF DILLNKMRQLKFSNSA LKWFE SYLKNRQQSVVCGNEKSSWRNVVSGV PQGSVLGPLLFSLYVNDISSVLTFCKYHFYADD LQLYQSTRPEDINTAIAQMNE DLSSVVTWAKN LGLKLNAKKSQVILVAHQKLISSEL RDQLPPIQ LDGIPIPYQKTVKNLGVIIDEHLNWADNTVTV CRKTSACLYALKKFRNIFPQDVKCKLVQSLVL	cd0165 0	173	410

			ATCAA	PNLHYCDMIQHGTSSENTRRLELTMNACVRYI			
			TATGA	CNIRRYDHISASYTQLGWLRPDKLRDYHTLCL			
			CACGT	LHRLLEQHHSTSPQNLTFRAIILTQDPTYLVP			
			CATCG	LCPFIKRKHLQTPSLSLLFASGTSCPALFARFSL			
			TGATG	LLRLRKSNTSLCHPLKTLAKTHVYSQVSLCVIV			
			GGAGA	KPMLFSPLFLFLLFLSPNQCSSLRYFSSSSFFHH			
			CCTTA	NLLFRYISLCEYYYTHFSPPFPPLPPCPLLLAAII			
			ACATG	KTFPYCNTCHVSYLSNTNINCVYVLLLLVLIEN			
			ACTTG	FNFCRFSITSVILEM			
			TTAAG				
			AGACA				
			CTCCA				
			TCCAC				
			AATAA				
			ACTAA				
			GAGAT				
			TGCTC				
			CAACA				
			ACATG				
			AACAT				
			TCTGC				
			CATTA				
			GAACC				

				TACTC					
				ACCAT					
				ACGGA					
				GCACA					
				CTCTT					
				ATAGA					
				TGTGA					
				TCGCA					
				ACGAA					
				AAAGA					
				TCGAC					
				AAAGT					
				ACAGA					
				TCTAA					
				GTCAA					
				ACGTC					
				AGCCC					
				GCCTC					
				TCAGC					
				ACACG					
				ATGCA					
				GCGCC					
				AAAAG					

				TTAAA					
				TCACG					
				TTATA					
				TTACT					
				CGTAG					
				AAACG					
				TGAAA					
				CAAAT					
				AGACC					
				GTGAG					
				GCTCT					
				AACAG					
				CCAAC					
				TGCTC					
				AGAAA					
				TCTCA					
				TGGCA					
				TCAAG					
				TAACC					
				AGAGA					
				ACCCA					
				CGATT					
				GATGG					

				AAAAA TTTTTG AACTT AGAAT AAACT TATTTC CTATA TGACA AGCAT GCCCC CTACA ATCCG TGTA GAAAA CCCCT GCTCC ATGGC TGACT AAAGA ATTAC GTCAA					
Ving i	Ving i-	.	Locusta migratoria	TCGTTT CAGAA	TAAATAA TTGTCTT	PHQNLLRPFTQAEVDLALSKCKNGKAAGLDDI RVEQIKNFGPITKAWLMELMNNCVQSCKIPKL	SSF56 672	19	256

16_ LMi		CCTGG	GTAAATA	WRKAKVIAIRKPGKDQNDPKSYRPISLLCHLY			
		TCTAA	TGATGTA	KLLERLILHRLMEKLEEYTYHTGWLTGEELHIS			
		CAGTT	TTCACAT	SDETYTAHRRWVSKERDHRSCIYRSDSSLRYC			
		ACTTC	AACATGT	KQKIFLSNFRDKEADGDRKRMDYPRESVLAPA			
		AGTAA	AAATATT	LFNIYTNDQPLPEGTQSFYIYADDCAITAQAHCF			
		ATATA	GTACATT	ETVEENLSKALREMSAYYRENQLRPNPGKTQT			
		GAAGG	AGTTTAT	CAFHLKNRQAARTLKITWENTTLEHCTYPKYL			
		CATGT	ACATTGT	GVTLDRALTYKRHCLNTKQKVAARNNVIRKL			
		CACAA	AAGTGCT	TGTTWGAHPNTVRTSSLALSYSAGEYASPVWF			
		TCCAA	TCTGACA	KSSHAKIVDVALNDTCRIITGCLKPTPLNKLYS			
		GCAAC	CGAATAA	LAGIAPPDIRREVAASKSERGKSETSEAHPLYGH			
		AGCTG	ATAAATA	QPANQRLKSRKSFLRSTEALVGTSQQCRTELW			
		CTTCA	AATAA	RARYTGTAGWFTPCEKLPPGYTEKWSTWRSL			
		AGAAC		NRLRSGVTKCKSNLVKWGFLVESPLCECGAT			
		TGTGC		QTTSHLLQCTQCPATCTPEDLLHATPNAIQVA			
		CATGA		NFWASKV			
		ACAGA					
		AGTTT					
		GATGT					
		ACTGT					
		GCATA					
		CAAGA					
		AACTC					

				ACAGA					
				GATAT					
				CCATC					
				AAATG					
				CGACC					
				GACAA					
				TACCT					
				GGTTT					
				GCAGC					
				TAGCT					
				GCAGA					
				AGTAA					
				CCCAT					
				GCTAA					
				ACGTG					
				GAGGT					
				GCTGT					
				TTTTGT					
				CAAAC					
				CAGGA					
				ATACG					
				AGTCT					
				TGAGT					

				GCATA					
				TAACT					
				AAATG					
				ACAAT					
				AACAT					
				CGAGA					
				TCATC					
				ACCTT					
				GGAAC					
				TTAGT					
				AACAG					
				TGTGA					
				TCACT					
				TCGGT					
				GATAA					
				GCGGT					
				CCTCC					
				TTCCG					
				AAAAT					
				TCTCCT					
				TGCTC					
				CTCTA					
				AAAAC					

				TTCATT					
				CAGGA					
				AAAAT					
				ATCTT					
				GGTAA					
				TATGG					
				GATAT					
				GAAAA					
				TGTGA					
				ACCCG					
				TTATG					
				TCTCG					
				AGCTG					
				AAACC					
				AACAC					
				CCWCA					
				CTCTC					
				ACTCA					
				CGACA					
				GCAAA					
				CTACC					
				TCATC					
				CTTCA					

				AAGTG					
				AAGGT					
				GGCAA					
				CGTGA					
				CCTCT					
				ATTTG					
				TGAGT					
				GAAAC					
				TAGCC					
				AACAG					
				TGTTA					
				CAATC					
				AATGT					
				GTAAT					
				TTCAT					
				ACCCA					
				CCACA					
				CAACA					
				CGGGT					
				ACCAA					
				TAATG					
				TGTCG					
				AATTC					

				TTCCC					
				ATCAT					
				AAGAC					
				CCCAA					
				GAAAC					
				TGGTT					
				TTAGA					
				CGAAG					
				GGATA					
				GTTTC					
				AAAAA					
				GGCAG					
				ATTGG					
				CCGAA					
				ATTCT					
				CCAAA					
				ATGCT					
				GGATG					
				AAAAA					
				GTTTA					
				TACAC					
				TTAGC					
				CTGTG					

				CCTGA					
				AAGTA					
				TGAAC					
				ACTTTT					
				TGAGC					
				TCTTA					
				ATTCC					
				TCGCG					
				AGCAT					
				TATCC					
				CAAGA					
				GCTGC					
				AAACA					
				AACTA					
				CCTGA					
				GGGCA					
				TTACTT					
				CAGAA					
				ACAGC					
				TACCT					
				TGCTC					
				GATAT					
				TATAC					

				CCAAT					
				ATGAT					
				AATCC					
				CATTC					
				AGTGA					
				TGAAA					
				CCCAA					
				AACAG					
				CCAGA					
				TCCGT					
				CCTCT					
				CCTCA					
				GTCTC					
				CGAGG					
				CAAAG					
				AGAGT					
				ACAAT					
				GCGAG					
				AAACT					
				CATGG					
				AAGAC					
				TGTGA					
				TATGA					

				GTAGA					
				AGTAG					
				TCACA					
				AGGCT					
				GGAGG					
				CTTCT					
				AAAGG					
				CTTAA					
				TAATG					
				ACCCT					
				TCACA					
				GGAAT					
				ATGCA					
				GGAAT					
				ATTCG					
				ATCTG					
				ATAAA					
				ATTCA					
				TACAT					
				GAAAA					
				CACCT					
				CATTC					
				AACCA					

				GGTCT GGTAA GAAGA CATGT GAGAG ATAA					
Pene lope	Pene lope- 2_L Mi		Locusta migratoria	AAGTT GTTTG AACAT TCATA TGCGG GAAAA GTTAT TCACA TTGTTT TTACC TCTAC GGGAG GAAAT GTATC GGTGT TTACG GAAGT	TAACAGT CACTATC GATATTT TTGATGT TTGGTCA TCTTTGG TTGTGTT GACTCTG CAGTGAC TTGTACT GTGTGTG TTATCTTC CTTGTTT ACCCTTT GCAAACC GAGGTAT TAAATAC	MIAANVSEFSWDWLDGASWTQSDWEFQSVT ARHSAKFERFLGSESDSNLISRRSVINLTEKTY DDAALSVLGKGLNFAPTPKRLPVVDFISSIEQA VSKLPSDSAEEARREACRVLTGARPPKCNITA AEASALRSLRMDPDIVILPADKGNATVLLNKD DYIQKMQLLSDSAYRRIGADPTKSVERKTNN LLKKSPFSQETIKSLNSYCAVPPRLYGLPKVHK ENIPLRPIVSNIGAPTYRVAKHLASLLSPMVGR CEHHIKNSADFLRRLQGLRLNDSILVSFDVVS LFTRVPLSDSLQLIEVKFGVELTNLFRHVLST YFLFNGQYYEQTDGVAMGSPLSPIVANLFMED FEERALVSATLKPACFFRYVDDTFVWPHGSE NLNGFLEHLNSIHPNIQFTMEVEKNGCLPFLDV LVRRKVDGTLGHSVYRKPTHDLYLRAADSCH HPAQREGVLRRTL VHRAHVISDPESLSAELSHLE VTFRQNGYSERQIKRALRYQPSVQQRVSDDSNE VAPKSMAFLPYAGSISNKIGRILRKYDVKCVFR	cd0030 4	339	423

			TGGAT	GCTTGCA	PPSKIKALLGSVKDDLGLRKAGVYRIPCSCGM			
			AAACT	CATTGCT	SYIGQTIRTVEDRCIEHKRYTRLQQPNKSAVAE			
			CCGCA	TCCTTGA	HCLGTGHPMEYNNTEILACTSTYWDSVIKEAV			
			GCAGT	GTGCAGT	EIKLASDLINRDGGFCLNSAWNPALSLVKKQR			
			AGAGT	TGTGCCT	DKINAASPVD			
			AAGGT	TGAGAAT				
			TATGT	GGCAGGG				
			GCTTG	TGTCCTC				
			TCGTG	CTGCCGA				
			TTGAA	AATATCG				
			GTGTC	GCGGTTG				
			CATAT	TCGACGA				
			TGTTC	CGTCACC				
			CACGT	CGGCAGC				
			TTGCT	ATACCCG				
			AAGGT	TAAGTTG				
			TGTGC	TTTGAAC				
			ACCAT	ATTCATT				
			ATTAA	AC				
			TTCTCC					
			TGCCG					
			CCACG					
			TATCA					

				AGCAA CGTGC TGCTT AGCTC TTGTTC GTGAA AGGAT CCACT TTATC GTCGT CAATT AGATT TTGTTT CCAAG GAGTT ATTTC GGCTT CATTT A					
Ving i	Ving i- 26_ LMi		Locusta migratoria	ACGAG TACTA CTAAA AGGAT	TAGATGT TTGTTGT GAATACA TGTAATG	TYENNSLSKSDKLIGNMIGNRGFQVIMGNISSQ KFLQNGLPQGSVLAPLLFSLYIADMPETACTKF GYADDWXIATRHTNFDTMZETLTSDLSALAA YFRKWRLRPNATKTEATCFHLNNRGADRTLE	SSF56 672	34	112

				TGCAA	TAATATA	VYFDDTRLNYNKEPSIQHTSVNGGLGLMPRRQ			
				AGAAT	TGTAATG	KRLASISITEERTELWKYTLTTHASTTTRSQVS			
				ACTCA	AATGTAC	RVTLDRTLSFKTHVTKTAAKVRTRNNLIHKLC			
				CGGCC	CATAATG	GTTWGSTATTLRSTSALSLVYSAAEYCAPVWL			
				ATCCC	TAATCAG	NSSHVKKIDVELNNTMRIISGTIRSTPNIWLPVL			
				AGTAG	ACACTGA	SHILPPKMRRETALLREYNKIEANPELPIHSDIP			
				AAAAG	AATGTAC	DLRIKRLRSRHPPLSLDSL TGADINSWRREWLQ			
				CCGGC	TTAAATG	NAPQLATHQLLEKRMAPKRPTTCHLMPCITEA			
				TTTCGT	CCATACG	VPGFDQPRKIWSTLNRLRTGHGRCADSLFKW			
				SCCAG	CTAAATA	GLLSSPACDCGAPKQTIQHIREECSARAFSGDF			
				CCGAA	AATAA	EEFLMATQKLLNIYKD			
				GCGTG					
				CAGAC					
				CAAGT					
				CCTGC					
				TCTTA					
				CTACA					
				CATAT					
				TGAAG					
				CAGCT					
				TCCAA					
				AAGAA					
				ATTAA					

				AAACT CTGTG GCCTT TTTGA CTTGT CAGCA GCCTT CGATA CCGCC TGGGA CAAGG CCTGC TCTAT AA					
Pene lope	Pene lope- 69_ LMi		Locusta migratoria	GTATT CTGAA CTCAA TTGAT TTAGT GAACA AGTTA AAAGA CTTCC	TAAATTG TAATGCC TTATTTTA CTAGTTA TAAAATC CTTGTGT AATCACA ACTGTTT TGGCTGC	HKVSHTYLPITKLFSFQWQGIQTVRWSSYGFP SPILAEIFMDNFENQFLESNPLSSKVGWFRYV DDVLCLWTGTTRQLQTFLKQLNSAHKSINFTM EFGGNCINYLDLTIDISNHCHTFKIYRKPSTTDT VIPAQSQHPVTHKHAAFHSMIHRLLSIPMSKES FQIELNTIRQIASNNGYPLPWWLPPTLVDSIMH KKQKQHVSSLLYPVINTPTSQDHSKWCTIPYL GKVSVNVAKKLKS RNFKTSFYTHNTIGHFLPN GKDNPPALEKSGVYKITCNCGKFYIGQSGRAIC	SSF56 672	12	120

			CTATC	TTATGTG	TRLKEHHRSWKLRQGNSTFADHLLKEGHSYN			
			CCCC	ATATAAA	VTHEVLHHVQKGRKMNFFEIMEINRHMYTSP			
			TAGTG	CTTAATG	NLVLNDQTQFPYSPLLSAEPNHNVPQTMPCAD			
			CCAAG	TCTTATTT	NVNYHLLPTCLIPLLLHSHVIKALCNHNCCLAA			
			CTAAT	TACTAGT	YQI			
			TTCTTT	AACCATA				
			TGATG	GTTTAAT				
			CCAGC	TTAACCA				
			AGCTT	GTAAAAA				
			ATTCA	TTTTAAC				
			CTAAT	ATATATT				
			ATACC	TACAGAT				
			AGTTG	GTTTCAT				
			GTGAG	TCAACTT				
			TGCCT	TGTCACA				
			AAACA	CGCTTCC				
			TCCTC	ATGTTCC				
			ACAGA	AGACATT				
			CTTAC	GGACCTG				
			TGGAT	TTTAAGA				
			AAATC	TAATTTTT				
			AGGAG	TGGTCAA				
			TTAAT	GTTTAAT				

				CCACT	TTAACCA				
				TGTCA	GTAAAA				
				CAGAT	TTTTAAC				
				GA	ATATATT				
					CACAACG				
					CTTCCAT				
					GTTTCTA				
					GACATTA				
					GACCTGT				
					TTAAGAT				
					AAGATAC				
					GTTTTGA				
					TCAATAT				
					TTTTCAG				
					CCAATTT				
					CTATTGT				
					TTCACTTT				
					ATATTTC				
					TTAGGAT				
					AACGTCA				
					TAATCTA				
					TTATATT				
					GTCACGT				

				ATTTTAT			
				CAATGTC			
				AGGAATT			
				TACACCA			
				CAACTTG			
				AATTGTA			
				ATATTTC			
				CCGCAGG			
				TTAGTTA			
				TAAACCA			
				AAACACT			
				AGTCTCC			
				TTTGAAA			
				GTTATAT			
				ATCTAGC			
				AAGTTAA			
				GATCTTT			
				GAAACTT			
				AAACTTG			
				TAATGTT			
				TACCTGA			
				TGATGGC			
				GTAACAG			

					CCGAAAA CCGGTTT GTAAACC AATAAAT TTGCAGA ATATCTC ACAGTTA TCTTGTG TATTTCA TTCATAA TGTATCA AATATTC TGTCAAG AACCGAC GGAACAG TCAGTGA ATAT				
Daphne	Daphne-60_LMi		Locusta migratoria	TCAAA ATAAA AATAG TGTA AGATA ACAGA	TGATATA GACGTAC ACAGAAT TTAAAT TGTAACA TTACCAT	MDSTAPCKKVLAKEILQAIKTASVDTFEKLAL PAIGLLPMANEKENVPTVHKEETTMATPPEET CPQPTPTSASEALSPVETPEPSVQEESGGARAS QAIPGVSQETVTASTPKGKGRSTADEAVSSVV SKRKTVPPLRLNDFLVECRKIKKPTKICIPLLHK KKYVEMTSRISYPDESEADSLVPPCKEKKLIFL	cd01650	693	926

			CAATA	TGTGATA	HTNVQSLKNKLFELQYWLEKLNCSILMINEHW		
			AGGAA	TATTGAG	LSCGELELSVPNGYELATGYCRXNKAHGGVSI		
			CTCAA	TCACTGT	YINKMLQLNYETLDVNEFCIESHFEEVAAVLLT		
			GGCGA	TCTATAT	ELEIIAVSVYRPPNHSEDLFIENLQKLIVSLLKK		
			AAACT	TGTACTG	KKSQSSNFWRHLSHSSRKKNHRVLIFGDINIDI		
			GAAGC	TACAACA	RYRSSTVTNLLDTRLRSLNFYCINDKPTRKNACL		
			AGTGA	ATGAGAT	DNII CNFHIGRDQFNTIKLAISDHDGLWLRLPLS		
			TGACG	AGAGGAC	HKTDTSQQRIIRPINEQKLNTLKS KLTLVNWNE		
			AAAGA	TTTGACC	LLTRDMGANESTKRFLEVVT KIFEETCPAYYK		
			AGAAA	ATTSATG	KH SKNRSQSKPGTLKQWYTAELNNLRILMLM		
			TTTTCA	TAGATGT	YYDRSQHSNIDRLRYKNIKKYYRSAIKEAKCK		
			TATTG	ATACAGA	SNDDYILNSVNKCKAAWEVIKTEINHQKKDQ		
			TCAAC	ATTTTAA	NIPISPD TLNKYFANPLMENNTEEDVNKAEELL		
			AAGTC	TTGTATTT	PQLHNKAMGHFEWNHVTCRQVSVAISKLSNS		
			ATGGT	TTTCTAA	RTEDHYGLSNFIIKQISQEILTPITNIINKILTEGV		
			AAGGG	CAAATTT	YPDSLKISVIKPVHKKGSTLSPDNYRPISLVPIIS		
			TTAGC	ACTATTG	KITEYCMHQQLNNYFEHNNLICSEQFGFRPTLS		
			AAAAA	TTACTAC	TVKAVDNIITKVYDCFEKKTLS CATLLDLRKA		
			TAATG	TTTTCAT T	FDMVSHN ILIRKLQYYGIGENALRLLQLVNVN		
			TCGAT	GTACTGT	NKKSDPLPVKKGV PQGSVLSPFLFIVYINDLPSI		
			ATCCA	ATTGTAT	LPCEAVLYADD TTLIAQNSDMDELLLNVAVAI		
			ATCAG	TAGGAAA	ENCNNWFQANXLHINSSKTENIIFDLNAHKHE		
			AATAT	ATTTAAA	NKSVKLLGFHIDQSLSWETHTDKLC SKLARVI		

			AGTCA	ACTAAGT	YLLHRLRGCVSKTLLLYAYFAFFSSHLTYGILL			
			GTGGA	ATATGAT	WGNSPGSKAVFKWQKKAIRCVAGLAPQESCR			
			ATCTC	AATTGTA	HFFKEMKIMTVPSLYIYSCLVYVKENLHKYDL			
			AAAAC	CCTATGA	RMSVHNHNTRHRHAIDKPHVRLAKVHSSYKY			
			CTGGA	TAAACAC	LGPTFFNKLPNEAHSVPLNKFKNIISKWLKSKA			
			GCTCG	TGCAAAT	FYTIKEFEESTIDDLEF			
			TTGCG	TACTAGC				
			AGAGG	TGTTTAG				
			TACTG	ATATGTA				
			AAAAA	TTTAAAT				
			CAGTG	GTATATG				
			AAACT	ACGACAC				
			ACTGT	CAATTAC				
			AAAAA	ATGTAAA				
			GTGGA	CATGTTT				
			GTGAC	AAAGGTG				
			TGTGT	AATAAAA				
			AGTAA	TATTCTA				
			TTGGT	TTCTATTC				
			GGTGG					
			TAATG					
			ATGTG					
			TATAG					

				AAATG					
				AGAGC					
				AGCCT					
				TGCCG					
				TGAAA					
				GCACG					
				AGAGA					
				ATCAC					
				TAGGA					
				AAATT					
				GTCAC					
				AAGCG					
				AAGGT					
				ATTTTT					
				GGTAA					
				GCATT					
				CCTCA					
				TAGAC					
				ATGAC					
				CGGAG					
				ACGAT					
				TCATG					
				TGTAA					

				ATAAA GAAAT CGTCA CAACA AACAG GAAAT TCAGG AAGAT CTGCA GCAGT TTTGT GCACA CTACC TTTATT GAGGC TACTA GCTCT ATCAG AGAAC ATTTT AGAGA					
Nim b	Nim b-	.	Locusta migratoria	CCCCC CAGGG	TGAGTCC TCTGGGT	MSKLRDGSVLIQAETPAQSRSLTCDHLGDIPV SITPHKSLNRVQGVYHRDLLLQTDEELRDNLQ	cd0165 0	901	1159

24_ LMi		GGCCC	GGTCTTT	RRGVHFVRRVFRGPKDNRVATGAFVLA FEGD		
		ACAGC	TAGATCA	NLPEKVKVMMYRCDVQPYIPPPMRCFKCWRF		
		TCTTTT	TTATCTA	GHMSSRCNAAATCRDCGRPSHPDTACAPPPY		
		GTGGG	TGTTC CA	CVNCGGHHSPCSQDCPALHKERKIVEYKTLDR		
		TACGT	CTTGACG	LTYSEARRKYEKLHPVSL LSYAAATTMAPPP		
		GTGTG	CGTCTTT	VATSRPALPPSVGPMGRPSTSASQTSGGTPSSA		
		GCGCG	CGACTAC	GPVLPPTSGVVTPAPQPEKGSPSLASSSGKGTC		
		CACGG	AGCTTTT	SPLPTMSPQKSDTSQRLKDPRPESRLAPQSSHT		
		GTCCC	CTTGTGT	PQAPPEKPPSSGRPKERREDKKEKKERRKAQG		
		CGAGC	ACGCTAT	DAGGLGAPSPCTGHPSRDARRFGLCRSRGCHG		
		CATTG	GTGGTTG	HGGSDFTARDECSCHPSSIFIVFMHRQSGDSLIL		
		CGGCC	GCTTTTC	QWNCGGFFCHLAELRQLLSSFPTLCVALQETW		
		TTCCTT	CCTTTTTA	FPGDRTPTLRGYRGYFNNRTVYNRVSGGVCL		
		CCTGT	TACCCAT	YVHTSVYSSPVPLQTALEAVAVRVRASRDITIC		
		TTCCC	GGTCGGT	SLYLPPDADVAPADLAALIAQLPSPFLLL GDFN		
		AGGCC	CAACCAC	AHHPMWGSNVTTGRGRDFELLFNQLDLCILNT		
		GCTTTT	ATTCATT	GAPTHFSVAHGTFSAIDLSVCSPGLRPLL NWW		
		CCTCA	GGCTCTG	VHDDL CGSDHFPIFLSLPHHQRHARLPRWSFH		
		CCCCG	TCATTTT	RADWAA YTSAAMLDTPLPADATMAVDQIISV		
		GTTCC	ATCTCGT	VVSAAESAI PSSSALPRRRPVPWWSPEIAAAIR		
		CTCCC	ATTTGTT	NRRRALQQYRRHPTVDHLVAFKRLRAQTRRLI		
		CTCCC	ACTGTGT	RQRKQACWEEYVSTIGPRVSSTQIWTKIRRISG		
		TAATC	GTCTGGC	YHPPAGVAGLSIRGTIHTDPTIIAEYFAAHFALN		

			CTTCTT	GTCTTTT	SASTAYSQGFCTLKERAERKQLVFRTNRQEPY			
			CCCAC	GTATCGT	NAPFTEWEFQSILASCPNTSPGPDRIHCQMLKH			
			CCTAC	CTCTTGTT	LSVNSQRHLLDFFNRLWIEGAYPSQWQESIVIP			
			CCTTCT	GACTCTG	LLKPGADPLNVDSYRPISLTNVLCKLFERLVSR			
			CCAAC	TCAATAT	RLLWVLEYRGLLAPQQGAFRQGRSAIDNLLSL			
			CCCGT	TTCATTG	ESAIRTAFAARRHHLVAIFFDLRKAYDTAWRHH			
			CTCTA	ACGGTGC	ILATLHGWGLRGSLPLFIQNFLLSRRFRVRVGT			
			GGTGT	GTTCTTTT	SFSSPHIQENGVPQGSVLSVPLFLVAINGLVAA			
			TCTCG	ATTCTGC	VGPSVSCSLYADDFCLYYSTASTSVAERRLQA			
			CGATG	GCTATTT	AIHRAQNWASAHGFTFSPAKTRVMHFQRQT			
			ACGGC	TAAATCT	VHPHPELYLDGELLQVVDTHRFLGLLFDSRLT			
			CCGGC	AGCTCAG	WPHLRQLKQRCRHLVDLRCLSHTVWGAAR			
			TATTTT	GCGCATT	PILLRLYNALVQSRLDYGSQVYGSASPSALRV			
			CACGG	TATCGAC	LDPVHHTGIRLATGAFRTSPVTSLLAEAGVAPL			
			TACTG	GATTATG	RYRRNTYSPSMLPESTVPLGIQITDSFSRGRSFV			
			CTGGT	TTTTTCTG	CRRGDPGRDSRRSGAVTSGLREPGLWLSLPIC			
			TTCGC	CGTGTCG	ESSRPSPPYRHPSGHWLSYEPCHESSRRGWSR			
			AATTG	CCTAGCT	PIAIPAQHLLAQYAARVHCTPRHPNNRLLFPGE			
			TGTCTT	ATGAAGT	VIRLPQRRPRSGLPAAARVQSLLSTLDSFPLNRP			
			TGTTT	GGGGGAC	FRSTFHAPPWTIPRPRIRLDLSREPKTSTSPLVL			
			GCCGT	TGATGGC	RHQFYALLREFRNSEVIYTDGSSDGDHTGYAY			
			TTTCA	CATAGTA	MHGAYVEQRSLLPPGSSVFTAELVAIYRALKHI			
			CCGAT	GTCTAGT	RTCPGTSFLICSDSLSGLQALDQCYPRHTLVRA			

			TTCGG	CCCCCTA	IQESILALDDSGRSVTFIWTPGHVGIPGNESVD			
			CCCAG	AAACCCC	QLAKQATTCTCRPLAIGLPEADFRPLLNRNKIRELW			
			TTAGG	CCATACC	NAEWSALDVPNKLRTIKDTSVWKSSQLGSRR			
			TCCCC	AACCAAC	DSVALCRLRIGHTRLTHGHLLRREDPPVCQCG			
			CTCGG	CAACCAA	APLTVVHLLLHCPILAPLRHSFNLPNTLSLVLG			
			GGTTT	C	DDASVANMVLRLRAGGIYQLL			
			GACCT					
			CCATC					
			AAAAT					
			TTCAG					
			AGTAA					
			GTGTG					
			AGCCA					
			TTTGG					
			GGAAG					
			AACTC					
			CCAAC					
			CTAGC					
			CCCTA					
			CGGCG					
			CGGAT					
			TCCTA					
			CCCTTT					

				CCTTC					
				CAAAT					
				CCCTC					
				TTCCTC					
				GCCGT					
				AGCTC					
				CCTCC					
				CCCGC					
				TAGGT					
				CACCA					
				GCACG					
				GTAGC					
				CAGTC					
				CGTGT					
				GGTGG					
				GGCTG					
				TTATG					
				TACCC					
				TCATG					
				GTTGA					
				GCCCC					
				CTGAA					
				TTGCA					

				CAGGG					
				ATCAC					
				ACTGC					
				CGATG					
				CCTGA					
				GCTGT					
				GACCT					
				CCTCA					
				TGAAT					
				GCCAA					
				GGAGT					
				CGATG					
				CTTGT					
				CTCTG					
				CGGGG					
				CATCA					
				GGACT					
				CCCAG					
				CAACG					
				GTCAT					
				CCTGC					
				CAGGT					
				GGCGT					

				TGCTG					
				TGGCT					
				GGGTG					
				GCGCC					
				CTTGA					
				GGAGA					
				GCCCC					
				TGTTTC					
				GGAGT					
				GGGTG					
				GCATC					
				GGGGC					
				GGACA					
				TTCCG					
				CAATG					
				AAGAG					
				GAAAC					
				AAACC					
				TCTGA					
				AAATC					
				CTGGC					
				TCTAA					
				GTCAG					

				GTGTC					
				TCAAG					
				TGTGG					
				TAAGT					
				TTAAT					
				TTCAA					
				TGCTG					
				CAAGA					
				TACGA					
				CCCCA					
				AATCG					
				TTTCCT					
				TCCCT					
				AACTC					
				CATCA					
				TGGGG					
				GGAAC					
				GTGGG					
				CCCAC					
				AAAGC					
				AGACT					
				GAACC					
				ATACT					

				CACCA					
				CGATA					
				CCTAG					
				TTGTT					
				ACCAG					
				ACAGG					
				ATGGG					
				GAATC					
				TTTCC					
				AGGCA					
				ACTAA					
				GCCTT					
				CCTTCT					
				TCGTC					
				GAACA					
				CCTCG					
				AAGAA					
				AGGTT					
				TGGAG					
				AAGTT					
				GCCGG					
				T					

Dap hne	Dap hne- 23_ LMi	Locusta migratoria	C	TAAAATA GAATTTT GTAATGA AGAACTG TGACAAA CATATAT ATGTAAA TTTGTATT ACTGTGT TAAAGTG TTCTAAT AATATAT GTGAAAT TGCATTA ATCCTAT AACATTT AATACTT ATCTGTA AAATTAA TTACCTG TGTTTCA TGGTATT CATTTAT	LYLKQPQQLHKVGLPIIHSLVKCLKPQEQHQKI ALSLQCKTVSTPTTSPGRHPRRSYASALNTAAT TPSSPVNTEETATVAVPSEEGGGKSASFSSKFF SRKQSKEXHNITRXKTLKEVSGSSEVTKSISCN KQNKSFHIMYHNVQSLTNKLQELEIILSSDLNA DVICITEHWLEKHQIEALKIPGFHLATSFCRTTI NMVGQLFLQNRITWWDSYFCKTEFKAMQSYS DLIKEKVFEMSAIELTDKHLVIVTVYRSPDGI EEFLCKFEMMLNIVEPTKKKIICGDFNVDLKS SNYKEELLYITKNFSLIPTVNTPTTRITKHSKTP DQIFIQGTFSYNSKVVKMGYSDEAVLLTLE LYSEKTPPIKMIECRDFSDGNIANFCNQLSQET WYEIFYETDVNKMFNTHQTYKYYFDIAFPLQ QRKLKPAKNKCWITNGIKISSTMKRYLHSYKK QNHLSLEMQHYIKEYNKIYRKVIKEAKKMGN DEYIRKSKNKIQATWNIVRYETDTKPVAKSISL KLENRQVSEATTVANIFNDDFTKIADDLIQSNF QTPNLHHTVNYKSPKGSMFLKQVTGSELMLA MSSLKKLSCGFDDIPDYIHKQTGYHIVEPLMHII NHSFSTGTFPNLLKIAKVTPHKKGATDDVNN YRPVAQLSGFSKISLKSSFLIDCQLQACSSAQW LLKNIFEKLFLNRLLEFVNKNSLLSKHQHGFRK SRSTTTAIYEYLSILKAVDEKEMATGIFLDLS	cd0165 0	637	930
------------	---------------------------	-----------------------	---	--	---	-------------	-----	-----

				TAAGTGT	KAFDVIDHGILLHKLERKGIRGISNEWIKSYLT			
				GTTTCAC	NRQQRVVLKQDTTIGNISRRSTYTSEKTTIKYG			
				GGTGTAT	VPQGSTLGPIFLLYIDDLNDTVLAKQKTLFAD			
				ATTCATT	DTSILITGIDQKHLQQSTDTSLLELTHWFKINKL			
				TACTAAG	IINTEKTVAMNFHLSHQKQQNNHTLKINNRS			
				TGTACCA	KLVDVVKFLGIYLQSDLKWNTHIKALTSKLSSI			
				GTACTTA	CYMLRILSTSCNKEIVIQAYHTHFQSAIRYGVIF			
				ATTTTGA	WGNAPCSHSVFKMQKRALRIICNMKKQESCR			
				GTCAATC	PLFPSLKILNLPCLYMFETIVFVKEYLGSTSAYQ			
				ATATTGT	LNSSVHSYRTRNDSNIHVSYARTSAYQKSVIHI			
				GTAAAAC	GTQLYNHLPGNIKSIKEDSIFKNKLKSFLMDHS			
				CTTTCAC	FYSVDEYLS			
				CATTATA				
				TTCATTT				
				ATTACAT				
				ATACCAG				
				AACATAA				
				GTCATAA				
				ATTGAAC				
				CATATTT				
				TATGTAG				
				GCCTAAC				
				TACGCTT				

					TTTTGTTT TGATTTA ATGTCAA TCATATT GTATAAA ACCTTTT GTAAATA CCTCATG ACAACAT CGATTGC ATGGACA TGCTTAA AGATGAA TAAATAA ATA				
Ving i	Ving i- 19_ LMi		Locusta migratoria	GGGTA CGACG TGTAG ACAAC CTCCC CGATG GTGTA CCCTG	TAGTACT GTAAGAG CCTTGTA TTGTTAG AATAGCT GTATAAG AACAATG TAATTAA	MATLVKNEILRNVDTITPPTEHVIGIKLGELTL YNVYKPPSQKWPVDVLPKANPQQFMLAISTPN IPDGVTRDPQLKALDPPAVYAGDFNSQHTRW GYPRSSAEGIGLNDWADNQDLHLLFDAKDKA TFKSKAWGTTSTPDLTFVTRGATGLPVQGTRR VVEAFPRSQHHPVIVEIGLSIPVIKSPPIPRWNLR KADWNKYRCFVEKTINRIPPKPENYKRFTNLL HKAALNAIPRGSRKEYTPCWTKECETLMEEFE	cd0165 0	421	675

			GGCGC	CTAAAGA	KCGDTETADRLINAMNEERRSRWKEAMENLD			
			TGTAA	ATATTTG	FTHSSRKSWGLLRKLGGATIPQRPSLGTTASEI			
			CATGG	CCTTTGT	SAVLKHTSKIQLKTAEKKEVSRDLSNLLVHNP			
			GACAG	ATATGCC	TNMDIVKEVETEEITQALKLMKKGKAAGPDEI			
			CGGCG	GAACGAG	LPEFLKELGTLGRRWLAKLCTECISTGALPKIW			
			AAAAG	TAAATAA	KEAKVIAVLKPGKTGDNPKNYRPISLLCTSYK			
			TCTAC	AATAAAT	LFERILLTRLTPYIEANLPTYQAGFRAGRNCCD			
			ACAAC	A	QVLSLTTYTEKGFQNMKTGLVLIDLTSAYDT			
			TATGG		VWIEGLLLKLTRIIKCKQTYTLLKNMLTGRKIK			
			CGGCC		VSLHGTNSKSVVLNNGLPQGSALAPALFNLYI			
			GGACG		ADLPETKSRKFAYADDLAIAIYQSKNFNDLNET			
			CACAA		LNADLKVLNTYYSKWHLHPNPKNKTTSTIMHL			
			ATGCC		NNREAHTKLQLTMNGQQIRHENKPKYLGVKL			
			TCCCG		DRTLSTYNAHLEDTKQKLKSRNAILSKLAGTTW			
			GTCCA		GCGASTLRTSGLALVYSVAEYCSPVWMRSAH			
			CTCAA		VAKVDVQLRETMRIISGTLRATPCAWLPVLAN			
			AGGCA		IEPPEIRRSQLAIRTHQKILGNLDLPIHEDLRAT			
			AATGG		DRLKSRTAFWKTSEELQEFNSKTAWKELWAT			
			AGCAA		SAPKNSNLVDDPNKKIEGFSLPRKLWTTLNRV			
			GCAAG		RTGVGRCKQLMKKWGLTDSSECECGETQTM			
			GAACT		DHLLACEVYGYNGDLKDIHRLSDSALKWLEKI			
			TTAAC		CNVV			
			ATGAA					

				GAGAC					
				TGAAT					
				GTATG					
				TCATTT					
				CAACA					
				TAGAA					
				GACTA					
				CACAA					
				GGACG					
				AAGGG					
				AGAAA					
				TCCTT					
				GAAAA					
				AATTT					
				TAACA					
				GAGAA					
				CAAGT					
				AAGTG					
				TCGTA					
				CTTCT					
				ACAAG					
				AGACC					
				CACCT					

				CACAG ATGAA ACTAT CGACC GACTC AAAAT ACCGG GATTC ACTAA TGTGA GCTAC ATTGG CCACG ACAAG CGTGG T					
Ving i	Ving i- 22_ LMi		Locusta migratoria	GGACT TCCCT CATAG CCAAC ACCGC CCAGT	TAATTAA ACAATGT ATAAAGT GTGTAAA GATAGCG ATGTATG CCATACG	NWNHIPLITTTQRPRWNFRKAKWDIHPEKLDK CLGWIHPTANNDRFMGAVKSAAKAAVPRGFR KQYVRGWDKECENKYQRFLESGDPEIADDLL HSLDNARREKWIETVESLDFKTSSRKAWSLLR KLGGANKLHHNSTDISANRISAHIQATSRAPRD KKHTRTIRRELSFLKKNSTPLDQYSSKFTTTEV REALKSVKAGKAPGLDGVHPEFLLTVGNMPK	cd0165 0	287	563

				CTTCA TTGA	ATAAATA AATAAA	NGLLNFIQMSQRS AE ECQGRKGTWIGWSTPRI PSYCGEYAKKWLA EFYS DVMATGNIPQKMKN SKIIALLKPGKPSNQPKSYRPIALLSVLYKLLER LLYNRLSPKILESIPDEQSGFRPNRSCTDQVLSL TTHIEAGYQKSKKHQLHLSSTFPDNAYRGRLS KKQKTSIAFIDLSAAYDTVW IQGLLYKLIK IIPC SSTIRLINNMLTGRAFQVITGERTSSTKVLNNG LPQGSVLAPLFFNLYISDIPNTQSRKFGYADDW AIAAQHESFEATEDILCSDLKLLCGYFRKWRL QPNPTKTEVSSFHLCNHSARRELVIFEGKRLS HNNYPKYLGVTLDRTLSYKEHLTNTANKMRT RNNIIHKLCGTTWGSTANTLR TSAVGLVYSVA EYCAPVWINSPHTNRIDVVLNDTMRTISGTIKS TPTVWLPTLSHIPPEMRRKNALLREYNKITNN PQLPIHSDVADLERRRLKSRRPALLSAKEMMN TNFEVEDLWKQHWQEHCPPKCQSFMTGEKPP GFNQPRKVWTTLNRIRTGHGRSADAFYKWGK TPSPNCDCGA EKQTVRHIVEECPFRAYKGT FN DFLAATEEA INYIKCLDVNL			
Pene lope	Pene lope- 47_ LMi		Locusta migratoria	AACCA TTAAA TTACC CATTT	TAGTTCC TGCCTCT ACCACCC TTCCCTTC	PPPVQISSIHSTHQPNPTKPSKPGPNSLAAPHLH RKILLLCNPKYLHPIPQTDVLALQVLEQHSHTT SKKLSNLFTSYSRLGLPLSTPSTTSRHTPQPLIA DKPCLADLLHLPHPQKLPHTTTQNPQPKQTVM	IPR000 477	326	566

				CAGGC	CTCTCCC	NLSSKSLSPIEVSVLSKGLTFCPTPKFNHAGLV			
				TGTAA	TGCCACT	KDLLSFSQSLQWKHHFATNPTIQHTNTNIEPC			
				TCCCT	CTCTATC	LTQFTPPSSQDPPTLPPNHPLLTFQNFLTSLNLAS			
				CCCTC	CTTACAT	PSFPKSLKMDTNLTPSERNAIYHLKTDPDLMIL			
				CACAA	TCGTCTC	PADKGSTTVVLDRKDYVAEGLRQLSDSSTYTP			
				TGA	CCATCCC	CHNDPIPKTQQDLQSLKSLGPSQDLSPESISLL			
					GCCACAC	TPPTPRTPTFYMLPKVHKPNHPGRPIVAGYGAP			
					ATTTCAT	TERISALVDQHLQPITHNLPSYIKDTNHFLHRLS			
					TCCCTTC	TVPVPLPPGTMLVTVDATSLYTNIPNAHGLAA			
					CCTTCCTT	VEHYLSQCPTDCKPTTSFLVTLTNYILTRNYFS			
					TTCCGCA	FEGITYKQICGTAMGTRMAPSYANLFMGHLEE			
					CTGTCCG	SFLTTONPKPLIWFRFIDDIFIWTEGEDTLSTFL			
					TCAGACA	QNLNTFSSIHFTWSYSTQRATFLDVDLFLKDGL			
					GCACTGG	ISTSVHIKPTNHQQYLHFDSCHPFHTKKSLPYS			
					CGTATTC	LATRGRRICSDKQSLSKYAKRLTEAFTDRNYP			
					CCCCACC	PNLVQKQISRALSVQSPSSSPTPTVRPQRSTPLV			
					CATGCCA	TQYHPGLEQLNQVLRQGF DYLS SCP E M R N T L P			
					CACATCC	NILPTPSYSGIPPPPTKPT E I P Y Q I S S P P P P T V V F R R			
					CCCCCCC	PPNLRN I L V H P Y S I P A P N P L P H G S Y P C N R P R C K T			
					CTCCTCT	CPIHHP T T T Y S S P V T G I S Y P I K G R A T C E S S H V I Y			
					CCTGCTT	KLSCNHCAA FY V G M T S T Q L S I R M N G H R Q T V A			
					CCCTCTC	RRQLDHPVAEHAALHSVTHFNDCYTACAIWIL			
					CTTGCCA	PTNTSFSELRRWELSLQYILHSRNPPGLNLR			

					CCTCCTC				
					ACCCCCA				
					CCACTGG				
					TTGCTCT				
					CGCCTAG				
					GATGCCA				
					TGCTGGG				
					TAGCGCA				
					GATCGGC				
					CACAGCG				
					GCCGGAG				
					ACGGCGG				
					TCGTGTC				
					GTGTGTG				
					TGTGTGT				
					GTGTGTG				
					TGTGTGT				
					GTGTGTC				
					TCTAACG				
					ACGAAGG				
					CCTAAAG				
					GCCGAAA				
					GCTATAA				

					TGTTGTT GTCTGTT ATTTGTG CCTTTCT GCGGCTC ATCGGAC ATCTCTG TGGTGAG TAAGCAA CCTATCC TTCTCAA AAATATT GTTGA				
RTE	RTE - 47_ LMi		Locusta migratoria	GCCAT AGCCG GTCCC AAGCC CGGTT GAGAA AGGAG GAGGG GGAGG AGTAA	TAATAAT AATAATA A	MKIIGTAELAE EATLQLNVQVVEDDAPLMAA NMVRWYRAAAGTSIELMHRANDASDGSRSKT HGRNQSCFRWIKIEDSWAKPIIVWFFGHDTAM WKGNGKTTTLLYFLNVGRRRIMMAFLGVKYS GDKLLPRSDLRVGSISGGVIKRDEHFNLRIGHTW NVRTLQRAGKLENLKREMDKCEVNIMGLSEV RWTDSGEIVSGNYTMFYSGGKSFERGV AIVM GNELVQSLTKVVCHSDRLMFVKISAKPVDILIV QVYMPTSDHEDEEVEKMYDEIEDIIQSEGTGK VNTIIMGDWNSVVGQGNEANIVGQYGSKHSW	cd0165 0	655	900

				CACCT CGTAA AAAAC CTTGG TTCAC CTGGC CCGGG TGATA GTACC TGTGA GGGCC CCTTG CCAGG GATAC GGTGA AGACC TCAAC GGCAG TGAAG GCGGA G	TIWLGKRNDRGGMLVEFCQRNNLVVMNTWF KKRKSKLYTWKSPGDMNRYQIDYILVKQFR GSVKDAKTMPAADIDSDHNLLVADISTRMKKI QQHGRRKQRWDLEKLRENNVAVRESLEKKFN EDEIRASAEQWQNVKMILLNTLKSEVGRET RARKPWITSEMLEKMEERRKCKNVNERNTED RNRGTTKKRKVRPHVPENEGTGEKENKSVRTF GIEDSRGQVVTEQVEVLKTWEEYIRELYDTEH RPHDIDIEPEEAIDEDEKGPSILTSEVEKAIDM KSRKATGDDIPVDLLKDIGNDSLKALTQVIN KIYETGEWPKDFLDVTMTLEKKKQAKKCSD YRTISLISHVAKIARLLNRRLEHKIEEVMGEDQ FGFRKGKGTRDAIGMMRILSERVLAVNEEICA CFIDWQKAFDRVNWNLNMILKEIGINWRDRR LIRNLYLGQRVKVRLKCGETNSVEIGRGVRQG CCLSPSLFNLYGEMLAREALKGCGNFKVGGRL INTIKYADDLVVLAKNQRQLQHMMNRLVEM GRKYGMEINIEKSKVMRISKREHLKIIVDNRE LENVNQFKYLGCFITKDAHCTNEIRARIAMAK AAFNKKRTLTSKLSLALRKKLVKCYIWSIAL YGAETWTMRKKEKKYLESFEMWCWRRIEKIK WTDRVTNDDVLRRVGEKKSILRTLQRKANWI GHILRHEGLIHDVIEGKINGTAGRGRRIQHLD		
--	--	--	--	---	---	--	--

						DLKDGRSYIRLKEEAEDREHWHHKFSVRRHG PATRQNTT			
Pene lope	Pene lope- 38_ LMi		Locusta migratoria	CACTC CTAGG GAGCT ACCAG TTTCC AGTTT CATCA GCTCT GTGGA GCAAG TAGTC AACAC TCTAC CATCA GTGTT GCAGA AGAGA TCGGA GAGAG ACGCA GGGCG	TGATGCG ACGGGTA GCATACC TCGCAGG CGCACAC CGAATTC GAAGCCA GGAGAGT AGCGTGG CCACGAT AACATTC TGTAGTG GTCACCA AMAGCA GAGGAGA TCGCCAT GAAAGTG CTAGGAA CGCGCAT GGCGGGC GCGGACC	EFAGRRKLGVLPQTRGTPLSFLKTDYDQKICQ LLEDPAYXLLENDPTDKVDRKTRSLLKETGWP EKVVKQLRAKAPVPPRLYGLPKVHKQDVPLR PIVSNIGAATYATASHLKKMLAPYVGKCVHHI RNSTDFVERLKNLRISNTDIMVSFDVVSFLFTRV PLQDTLDLISERFDGTVDLDFKHILTSTYFLCG GQFYEQTEGVAMGSPLSPVVANLFMEKFEEEA LTTAAYQPTCFFRYVDDTFVIWPHGQNRLEF LDHLNSRHPNIKFTMELENEGQLPFLDVLVKR KPDGTLGHSVYRKPTHDTLYLQADSCHHPAQ KNGVLRTL VHRAKTLSDSDNLAAELDHLQRV FTNNGYSSRDIRKALHPTRRPDTTGEERRRDRE GSLPTICRAYICQDQQNLEETQNQKCFLPTKQD QQSWIISSECSQTTATRAETSARRYTPPDAQIQ QEKKEEETGKVAFLPYAGPISAKISRILRKHKIK SVFCPPNKIRALIGGVKDDLGLRKPGIYQIPCEC GMSYIGQTTRTVXIRCKEHQRHTRLRQATKSA IAEHCLELDHTMDYDKTRILAQTPRFWDSSIITE SIEIKMGEDLINREAGYQLSTAWNPALELLQK QRKKGHPPVELQTTTRDSSSGWPPCAPTTA	cd0030 4	203	287

				CTACC	ACCTATG				
				AGAGC	GAACGCC				
				AGGCC	TGGGACG				
				CCAAA	GCCGCCA				
				ACAAA	CTAGAGA				
				CATCT	GCCGCAG				
				CAAGA	GGCGCGG				
				GACGA	ACCGACG				
				GAGGT	AGGGATC				
				GGCGC	ACCAGCG				
				TAG	TCAGATA				
					CCACCGC				
					CAGACTG				
					GAGCACA				
					TTGGGCG				
					CGGACCG				
					CATAGAA				
					AGCACCA				
					GACGCCG				
					CCTAGGC				
					ATAAATA				
					TGGGACG				
					CGGTCGG				

					CTTGTCG ATCAGTC TCAGGGA ACACCTG ATGAAGA CGCCGAG TTACGGT GTCGAAA TATCGTG TTTGGAA GACGCGG ATCACCG GCAGTAC ACCCGAT TCCACGA GA				
Pene lope	Pene lope- 32_ LMi		Locusta migratoria	TG	TGACGTC ACAGCCC GCAGTGC AGCTATA TAAGGAC GACAGCA GCAATCA	WNKHFARCLVNYRQIRRETSHSIQATKTEHDK SRTQWPSELRNDPLIVVLPADKGNAAVIMERA TMTRRFERCWRRRHMQKLSRDPTSKILRKTSE LLKRSSLEDSVIKNLLPQAPRPPTLYGLPKVHK MNTPLRPIVSTIGAPTYRLAKHLANLLTPHVGH CEHHIKNTADFISRIKCLRLGEEDMMVSFDVVS LFTNVPVKDCIDLLSQLFDDTVVDLFKHTLTSS	IPR000 477	105	351

					CTCGACA GTCAACC ACTTGAC AATGGCC GAGGAGA TCTCGGC CGAAACG TCGTGGA CAGTAAA CCACATG ACGCGGC TGGAAGC CCGAGAA AATTTTA TCAACAT AACTCGC GCGAAAC AATGCAT TCA	YFLYGGGEYFNQTDGVAMGSPLAPVMANLFME HFEDIALDTAPVRPSCFYRYVDDTFVWPMD VKNSVNSSTTSIASMTRSSSRWRWRKRVPCRS LTSSSAVKQMDVSAMVFTENPPTQTGTCTPQA IIHRRNSLFHVRCLAHGREKLGEFFDHINSIHD KIKFTMEVEKEGALPFLDVLVSRKANGCLSHG VYRKPTHTRYLHATSHHPSQKQSVLRTL HRAEVVSDTENLQKELSHLRKVFRDNGYSNS QITKAIFPKTQKQKTPEEDSKKLAFLPFCGSVT GKISRLLKRHKIDSVFRAPSKIRQLMRPVKDDV GLRTPGVYKIPCGCGQFYIGQTVRTIEQRCVEH ERCYRLRYPEKSAIVEHALENRHRIAFEETSIT RTNSFWDSVIKEAIEIKIADNTLNRDGGGLQLST AWNPAIGRLKRARQARDQNIIFYGHVAEHQ			
RTE	RTE - 33_ LMi		Locusta migratoria	GCCCT AGCTA AGTCT GGCAT	TGATGAT GA	MRIGTWNVRTLLKAGKLEEVKMLMEKSQLDI LGICETRWGGNGDFMSGDFRIIHSGSARSGRN GVAILRGKWKNVNLNTYHCSDRIIMIKIKAQP MDLCVIAVYLPPTNSKDEEVEVVYEQIDDLMK	cd0165 0	533	722

				TGCTT	LVKEKDNLIVIGDFNASIGEDKKGLWNGKFGL			
				CCACT	GKMNSRGERLLDFCEQYEMIVTNTCFEMPKR			
				ACTTG	RRYTWKSPVDGERYQIYILVKKKFINQIKSSHS			
				TGTCA	YPGYDIDSDHVLVIADYNSKFKKYKRTNQHR			
				GACTT	WCVEKLKEKQVAADMKELTDNLIGNREGWE			
				CTGCT	GFKETVEVAAKELVGLRKLEPRKPWMTKEILD			
				TTCTTC	LIEERNRLRKKDFDKYKRVKNAITQKCRLEKE			
				GTTCC	KWMHDKHKEIQQDLGGQTDKAYSIIKTMQT			
				TATCG	KTRTKTNMVKNKNGDLLFELEDVCERWKEYI			
				GCCTC	EELYEGEDIANEGERILEEENSVDIDMIGPPISKE			
				TCTTG	EFNSALNDLSNGKATGTDNIPAEILKEIGDQQK			
				GTCAA	STYINFTFKEYYEDGKEPPTDLLNSKTVIIPKKE			
				CTCTG	MLWTVKQEKTNLVSDLEKAKAILALRMVLER			
				CTGCT	RIGVNRKTYLTFIDLQKAFDMVNLEGQRLIWN			
				TCGAC	LYKRQEAIEVNGVKKKAKIRRGVRQGCPLSP			
				CCCGA	YLFNLFIEKAIEKLRITKGIKINGERIHCIRFAD			
				CGGTA	DIVVLADSVRKWSLCCLSKEMEFMLLKFAKIL			
				TTAGG	REFNLKINKKKTCTMVVGKTKEKGKVNIKLE			
				TTGCG	DDVLEQVENFCYLGSTITDDNKCITDIKRRIAW			
				AGGCC	QSKLSKIKGIYMYHRYKEKNSLAKQAFQNKR			
				GAGGG	NLLTDNHISLESRKKFAKTFVWSVLTYGCEAW			
				TGCTTT	TLGKAEEKRLEAMEMWIWRRMTKTSWVDRK			
				CATTTT	SNEQVLREVNENRTLLNYIGRRKSKMIGHIMR			

			CTTTTC	HNQFLKNVFEGKVLGKKPRGRPRATYFNNIKE		
			TAACA	DMGIKSYEELKRMTMERGTWLNRRQGIASFSL		
			CATGG			
			CTGTT			
			AACCG			
			AATGG			
			CTCCC			
			GGTAT			
			AAGGA			
			GACCA			
			ACCTT			
			ACCAA			
			GTGCC			
			AACGG			
			CCAGT			
			AAGGG			
			GGGAT			
			GAGCG			
			GTAAA			
			GAAGG			
			GGCAC			
			CTCCTT			
			GTCCA			

				GGGAT					
				GAGAA					
				ATTGT					
				CCCTA					
				AAGGC					
				GGAGG					
				AACCA					
				TTAAC					
				ATGGT					
				CAACG					
				GCAGG					
				AGGAT					
				GCAGA					
				AGGCA					
				ACGGG					
				AACCA					
				CTGCA					
				TTAAC					
				GATCC					
				CATAC					
				AGGAT					
				TCCCA					
				AATAG					

				AAATA					
				TGGCT					
				ACGTC					
				CAATT					
				CAAAA					
				TTTTCC					
				GGAGT					
				TTCAA					
				TTCCC					
				CCATT					
				CGGAT					
				CTCCG					
				AGGGA					
				AATGT					
				CTTGA					
				GCAAG					
				GCTGT					
				AAGGA					
				GTAAA					
				GGAAT					
				AAAAG					
				AACAG					

Pene lope	Pene lope- 49_ LMi	Locusta migratoria	TTTAA CAGAT GCCGG AATGA AGGTA TTGTTC CTAAG TTTGCT AGAAT AGTTC ACTAA TTGAT TCTAA TGCAC AAGAT TAAGA AGCTG CCAGT CTTGC GCTAG TACGT GAACG CTTCG	TAGCCTT GTTTCGA CCCATCG TTTTGCC GATTTTC TCCACCG GCGCGCT ACTAGTT CTATTGC GCCTAGA GGGCCTC TACTTTG GTTTCGC GCACGCG CGGTGGG TCGTCAG CGGTGTA TAAAGGG GCCGCTG TTTCTCGT TCTGCCT CAGTTTC CGGCGAG	ASVSRIPEDQADEIRQEVA YKLRRAPPPRSNIPS SERA AIRSLREDKDL VILPADKGNATVLLSRGD YVSKMNLLDDSA YRRLRNDPTERIKRKTFFH SHGTNKTEDLSLLKKSSLS EDVVKKLHPGAAV PPRLYGLPKVHKHGIPLRP IVSNLGAPTYNLAK YLASVLAPHVGKCEHHILNSSHFIRRLKSVCLG PSDLLVSFDVVS LFTRVPLEESLLLISDKLDD ET VKLCRHVLTSTYFLFNDAFFEQTDGVAMGSPL SPIVANLFMEDFEDMALRTSFLKPTCFWRYVD DIFMVWPHGMDALNHFEDGKL PFLDVLVQKN SDGTLGHGVYRKPTHTDLY LHATSCHPPHQR MGVLR TLVKRAYAISDAEHLGEELEHLKVVF KQNGYTAQQICRAFQFGPKGEPLDNYKSIAF LPFVGNISSKISRILRRFDIKSVFMPPPKTRALLG SVKDDLGLRKPGVYRIPC YCGKTYIGQTIRTV QDRCVEHHRHTRLLQPEKSAVAEHCLNEG HK MLYDETKLIAPASRYWECVLKESMEIRMADNI INRDKGFP LSKTWNPILSDIKQQRSSFRPAVSFN	IPR000 477	119	362
--------------	-----------------------------	-----------------------	---	--	--	---------------	-----	-----

				TTTTAC	TAAGTGT				
				TCGTA	ATCAGCG				
				GACAA	CACTCAC				
				CTGGA	CTGAAGA				
				TTTTGT	TGGCAGC				
				TTCCA	CAGTTGG				
				AAGAC	TCTGCCG				
				TTATA	AAATATC				
				TTTTTT	GTGTCTA				
				GCATC	GTTGACG				
				TTGAG	ACGGTAT				
				ATTTC	CCGGCTG				
				ATCAC	CACACCC				
				GGCTT	GAGATGA				
				TCGGC	ATATTAT				
				CGCGT	CA				
				CATGG					
				GAGTG					
				GGTTG					
				ATAGC					
				GCTTC					
				TTGGG					
				CCTTTT					

				CCGAC					
				TCGGC					
				ACAGA					
				GGGGT					
				GCTGC					
				CTGCA					
				AACAG					
				TTATC					
				AAATT					
				TGAGC					
				GACTT					
				TGCAC					
				TCAAC					
				AGACA					
				ATTCT					
				GGTGA					
				CGATT					
				TCCGT					
				CGGCG					
				GGTTG					
				TGTCA					
				ATCTT					
				ACAGA					

				TAAGG AACTG GACAA CGCTG CATTA TCGGT TTTGG AAAAA GGTTT AAATT TTGCA CCCAC ACCTA AGGTT GTCCC TATTA CGGAA TTTATT TGTGC TGTTG A					
Ving i	Ving i-	.	Locusta migratoria	AGAGG CTGCC	TAGTGGT TTTGTA	MNKELTSKLHEYLRRLFDEDPFSTDTIRSGMTLS QELAKAKQKRWSEMVESMDMSKSSRKAWQL	cd0165 0	176	431

	2_L Mi			GTACC TCATA CATTC CGGGC	TTTGTTTA TATATAA TCTCTTTT GTTTTGT ATTATAG TTGTACT CTTGTCTT TGACCTT TTGTATG CAACGCT TTTGACA CGAATAA ATAAAT	LKRLSNDPTTPKERFNVTPNEVAHQVLLNGKT EGRKEKIRITRHNLETNDLSAKFTLEELKAAIK MCKSGKAPGLDDILVEQIKHFGPVTMGWLLRF YNNCLTRRHIPKIWRRTRIAILKPGKPASDAK NFRPISLLCHLYKVLERMMLLNRLSPVIEPQLIPQ QAGFRPKKNCTAQILNLTQNIEDGFERGLITGA VFVDLSAAYDTVQHKIMLHKLYQLTRDYDFT SFIGTLLQNRKFYVELDGKKSRWRIQKSGLPQ GSVLAPTLFNIYSNDQPQPPNTKSFIYADDLAL VSQAKDFETVERHLYGLKTL SAYYEKNHLKP NPAKTQVCAFHLRNKDANKKLKVIWNGTELE HCHNP KYLGVTLDRAL TFRKHCLNIKQKVSAR NNIIRKLTNSGWGAQPTVLRRTTAIALCYSAAE YACPVWYTSVHARHVDVALNETCRITTGCLR ATPTHKLYNL AGIAPPDIRREVAANIERXKVEQ NAAHPLHGHQTPPSRLKSRKSFMKTTVNCSTS PSTARQNLW TTRNADIATWIPPLEALPPGENLP WKTWKS LNRLRSGVGRTKDNMARWNYLGDE STLCECGSEQTTKHL YSCPQCPV SCTLEDLLQA KENAVELANFWSDCI			
Nim b	Nim b-		Locusta migratoria	CCCTC AGGGG CTCAG	TGATGCT TGTGGCG TGCAAAT	MKKRVDGELVKSDSFILTFNSPRLRTMLRQAS CVSQTPDHVKAGFLRLNVRPYVPNPMRCFKC QRFGHTTLGCKGRLPVANVVRQPMMLTVLH	cd0165 0	865	1123

37_ LMi		CATTA	CACTATA	LRXVSTVQGTQSGVGIVQCTXRRRKYRKLRP		
		ATTTG	CGCCACA	QGVSPMQRPKRSSRPCNLLYFPPPSLQYCKNQ		
		CTGAG	TTTTAATT	LLKLMPPPRLRLLVWAPTPALASARVRQQKLQ		
		TACGG	GAGTGTA	LRSPDNADKATGTAFAPWFHNCRACHHVT		
		GCTTG	TTTTATAT	ATATTPPTPPPHHSCNHNLCCLHHQATLQSLPSR		
		GCGAA	CGTGCTG	HRHRHNPTNATTSPQLQPQLSPPPSHLNPRRL		
		CCCGG	AGAGGGC	PANTVTGKTVIXLSDFVEVIPPEPSSDPAPELME		
		GGTCC	AGCGGCT	LDLPQGQXSRPNTEPPPSVSPSRKKERVKIQP		
		TGAGC	AATCTTG	PRSNLRDRWPQYCSGTCRVSGRMWRNSVILT		
		TGGGG	GTCGTGC	QGRPICICLQETNFKSTDNPGLRGYVLHKKDDL		
		ACTGG	CCTACCC	SGERARGGVAIFTNTAYHSEPLSLTTHLQAVA		
		TAAGC	ACTGTTT	VSVMCVTVAPLSHHTFTSRCSICYVRHRLTVC		
		GCCGC	TAGCTGA	SLYLPPHDTFDVAALEDLLTQLPQPFIKCGDN		
		CAGTC	CGACGAG	AHNVLWGSASTCPRGRATERLLLSSSAHLLNT		
		CTCTA	GCGAGTG	GQPTHFCTATGSFSAIDLKSLCSPALAHHTQWEV		
		GACAC	TGACGAA	VDDLHGSDHFPIWLHLPDRVEPECRPPQWVIS		
		CGTAA	CGTTTTA	KADWTLFSQLAELETRDNVQEWVDCITQTIHS		
		CCTCT	CAATTTT	AAEASMPKSLGQPKRQPVPPWTAECRRAIRA		
		GGGCA	GTGAACG	RRAALRRFKSRPTADNLAAFRMARAKCRQIIR		
		CGCTT	TTCCCGA	ESRKRSWQQFLSTVTRSTKSSTVWETIRRISGK		
		CAGCG	CTCTGGC	GGRCPVAAVMGNGVLHTSPRDIAQTMAIYFA		
		ACCAC	CTAAGCT	NVTATAHQDPAFQRHRVVAERGNLDFRSTGH		
		CGTGT	CTTAGGC	QDYNCPFTLWELDSALSVARDTSPGNDRIHYS		

			GGCGC	TGGAGGT	MLRHLTQRNKEILLNLFNAIRPGHFPSWREAI			
			GGCGG	TTTAGTG	LIPLLKPGKDRTCPSSYRSVALTSCVGKTLERM			
			TGGAA	TGTTACA	VNRRLVWMLESRQLLSRFQCGFRRYRSTFDNL			
			CGTTG	GGAGGGC	ASLEAAIQQAFLRRHLLGVFFDIEKAYDTTW			
			TGTGT	TCGCTGA	RHLILRQLHEWGFRCGLPIFIRSFLSPRYFRYRV			
			ACCGG	TGCCTCT	GGVLSDRFEQENGVPQGSVLSVTLFAIAINSITS			
			GGAAT	AACATCA	VVKSPVQCSLFVDDFSVFCSSSSLATTTRQLQL			
			GGGGA	GCGATCA	TLRRLDEWAQKTGFKFSTEKSVCVLFNRSRSIF			
			TCTTG	GCGTCCC	NLPELRMRDTVLNFKDTVRFLGLTFDCKLTWL			
			GCTTG	TCTATAC	PHLKDLKRRCLKALSILKCLSHTSWGADRTRL			
			ACCGC	TCCGACA	LQFYRAFVRSRLDYGCTVYGSARPSYKMLD			
			CCGGG	TTTTCTTT	VVHHEGIRLATGAFRTSPIPSLCAEAAEPPLSIR			
			TCGCG	TAAGGAC	RQLLMVRQAYAALSTPYTPAYNTVARPPMER			
			AGGAT	TTGTAAG	LFRNRLRATRPF GIRAKDCFLDMGVSDFTVLR			
			GGTAC	GTTGGAT	RGWSRYPPWLLQRPSLLLDLTSFKKDCTPDITF			
			AAACC	ACTTCAG	QSEFFNILNKHRDCTVVYTDGSKQGDSVGC SA			
			TCTAT	TCTGCCA	VFPGHIVGIRLPAAYSVFSAELHAILQALEHIRR			
			AAAAA	CCTTACA	ARGTKFLLCSDSLSALQSLLVMYPAETITQLIH			
			ACCCT	TCGCTGA	DQLYLLQRRGKEVSFLWVPGHVG I WGNEAAD			
			TCAAT	TTTGCAA	VAAKAACSDPNVTECAIPLRDVYSVLRQRS LQ			
			CCCAA	TATGTGT	LWEETWLAVTG NKLRLVKPTVRPWRTSCQPF			
			GGTGT	TCGTGTT	RRFEVTLTRLRIGH S PLTHSFFIPSGGSTDL			
			GCTAC	GTTGCCA				

				GTGCT	AGACCTA				
				GATGG	CATTCTA				
				GATGC	TTGCCTG				
				ATGGC	TTCTTATT				
				TGTTG	TATTTAC				
				AAGGG	TGTTGTC				
				GAACA	TAATTCT				
				GTCGC	TTTGTCA				
				TAGCG	GGGCGCT				
				AGCAA	AAAAACC				
				CCTCT	TTGCTGT				
				GGGGA	TGTGCGC				
				TACTG	CCAAAAC				
				CCGCA	ATCCTAT				
				CCTCA	CCTATCC				
				GTTGT	TA				
				ATAAG					
				GCTTA					
				CCTAG					
				GCATG					
				CGGGG					
				CTCTG					
				TCTAG					

				GTGGA				
				CACTT				
				ACTTC				
				CCTAG				
				CTACT				
				CGTGG				
				GACCG				
				AGATG				
				GAAAA				
				ATTCT				
				AATTT				
				ACCCA				
				TGAAC				
				CTCCT				
				AGTGG				
				GATGG				
				GTGTA				
				CCACT				
				GGGGG				
				GTACC				
				AACAC				
				CCAAC				
				CAAMC				

				AAGAG					
				GGCTM					
				GGGTA					
				GCCAG					
				TCCTC					
				CGGAT					
				TCAGG					
				TTCTC					
				AAAAA					
				AAAGT					
				TTAGT					
				TTATA					
				GTAAC					
				AGAAC					
				ACATG					
				CTGCT					
				MMAC					
				AGAAT					
				GTGTT					
				TATTG					
				TCGTG					
				AAGAG					
				AAAAG					

				AGGGC					
				TCGTTT					
				GAGAA					
				GGTGT					
				CGCCC					
				TTCTAT					
				ATCCA					
				GAAAG					
				GTCTC					
				GAGGG					
				AATTG					
				CTGGT					
				GMTT					
				GAAAT					
				CAGTA					
				AAACG					
				GTTGC					
				GTAAT					
				GGSAC					
				ATTGT					
				TGGTG					
				GAAAC					
				ATCAA					

				CTTCC					
				CAACA					
				GGCGG					
				CCAAC					
				CTACA					
				GACGG					
				CTAAG					
				TGCCT					
				TGGGG					
				AATAT					
				CAAAT					
				CGAGA					
				CGGAG					
				CTGCA					
				CAACA					
				CCCTT					
				AACTA					
				TAGCA					
				AAGGC					
				GTTGT					
				CACCT					
				GTAGG					
				GATCT					

				GGTGG ACATT CCAGC GGATG AACTG AAGGA TGAGT GGGCT TCAGA AGGCA TTGTT GATGT GCAAA AC					
Pene lope	Pene lope- 46_ LMi		Locusta migratoria	GTTCG TAAGT TTCCC AGGAA CAGAA TGGA ACGCC CATTC TTGTA	TGACGTC ACTGGCG GCGGCGC GGCTATA TAAGCAC GGCAGCG GCCGCGA AATGACA GTCATCC	RNTNSELRKSELKDPGRYNYLGIVTIGLLQDLP NTCTLLSMHVGHCEHHIKNTADFIDRIKGRRL REEDIMVSFDVVSFLTCTPITDSIDLLAQLFDDT IVDLFKHTLTSSYFLYGGDYFNQTDGVAMGSP LAPVVANLFMEHFEEIALDTAPAKPSCFYRYV DDTFVIWPHGQKLLLLSLRRRYLRHLASWTEK IEEFLEHINSIHGHIFTMEVEKDGSPLFDVLV RRKPDGSLGHSVYRKPTHTDRYLHATSYHHP AQKQSVLKTLVHRAETVSDAXNLSNELSHLR	cd0030 4	126	229

			AAATC	ACTTGAC	KVFKDNGYSKRQISQAVSLKARNQKTSEEEVK			
			GAACC	AATGGCA	KTAFLPFCGSTTGKISRLLKRHKIDTIFRPPAKI			
			TCCAG	GAGGAGT	RQLMKPVKDDVGLRTPGIYKIPCECGQFYVGQ			
			CCAAA	TCTCTGT	TVRTIEQRRIEHERCFRLRYPEKS AVVEHALEN			
			CTGAC	CGAAAGC	GHRIAFDETSIIKRTNSFWDSVINEAVEIKISENT			
			AGCTG	TCGTGGG	LNKDGGQLQLSTAWDPVIAELKRARRTRDQNI			
			AACTC	CAGTAAA	LYGQGTGHQ			
			CGTGG	CCACATG				
			CTTCTC	ACGCGGC				
			TCCGA	TTGTAAC				
			TCCTCT	CCGAGAA				
			CACAG	GATTTCA				
			TGTTCT	TTAATGG				
			GCCAG	ATATCGC				
			CGATA	CGCGAAA				
			AGGGA	ACATGCA				
			TGCTA	CTCA				
			CCTGA					
			CGCCG					
			ATTAT					
			GATGG					
			AAAAC					
			CGACA					

				GATGT TGA					
Ving i	Ving i- 45_ LMi		Locusta migratoria	GACGA ACAAC ATGTT ACGGA ATAGC AATGT ATGTC AGGAT GAACA TAGCT GATTA TAGAG TGGTT TCTAA GGACT CTACA GACAA CATTT ATATC TTAGC AGTGA	TAGTATT TTCTCTG GTTTGTTT GTTTGCT TGCTTTT ATTATGC ATATTTG TACATAG TTTTTGTA TTTGTA TTCTTCTA TATATGA AAATTTT CAAATAA GACTGTA GCATGAC CATGTAA GCCATAC GAATAAA TAAATAA ATAA	SPIQSHGFKPRAGGSPGKCAYKQSGCFDGIYPE FLKNAGPKTKRWLLQLYNIILT TGKLPTTFRAA KVIAVRKPGKEGSEASHYRPISLLSVVYKLLER MLLNRIQPLIDKATPLHQAGFRKNRSCCDQVL ALTSFIEAGYQRKLKTAVAFVDLTAAFDTVWR HGLLLKFAEAIPCQKLVNLMDSMLSNNRRFQVF LGDKGSRRRLNDGLPQGSVLSPTLFNLYTSD MPNTSAKKIQFADDLALAYQARNLEECEGTLT TDLHIMGDYFKSWRLKPNPTKTEVNAYHLNN REANRKLNVVFENVPLEHNPYPKYLGVILDRA LTFKNHLEYVAKKVNTRIDIVRKL AGSGWGA NADTLRTATLALVYSTAEYCAPVWKN SAHVR KVDVKLHQAMRIISGTTKSTPLPWLPVLCNIPP PSIRREAAAIREFHKSSTHKNSLLYNQLRNIPTS RLKSRKPPWPMQDSPLVYDVCQEWKDVWKK FPPANSRLITDPTKRVNGFGLPRKLWTS LNRF TGHGRCSACLYEWGYKDAPACDCGAPVQTM QHLIEECPPRKFPGGLSELHLVTTRAVEWLRDL DVQL	cd0165 0	65	319

				CGATA					
				TGTGG					
				AACTA					
				CTATT					
				GTCAA					
				CATCT					
				ATAAA					
				CCCCC					
				ATTAG					
				TCATC					
				TGGCC					
				TACGC					
				CCCCA					
				TTCAA					
				ACTTTT					
				TACTG					
				ACCCT					
				GCCAT					
				CTATG					
				TTGGA					
				GACTT					
				TAATA					
				GCCAT					

				CACCA					
				TGCTT					
				GGGGA					
				TACAA					
				GAATA					
				ATGAT					
				ATAAA					
				TGGTA					
				ACATT					
				TTACA					
				TGACT					
				GGCTG					
				TCCTT					
				GAATA					
				ATTTG					
				TATCT					
				GGTCT					
				TTAGT					
				GCAAA					
				GGACA					
				AAGGA					
				ACGTT					
				TCGAT					

				CAGCA					
				AGATG					
				GCAGA					
				AAGAC					
				ACATC					
				ACCAG					
				ACCTA					
				TGTTTT					
				AACT					
				ACCAG					
				TCAAG					
				CTAAT					
				GGAAA					
				TCAAC					
				TGGCA					
				AACAG					
				GAAAG					
				TTATA					
				GATTC					
				TTTCC					
				ACATA					
				GTCAA					
				CATAC					

				GCCCA					
				TTATC					
				ATTGA					
				CTATG					
				GCATC					
				CAGAT					
				ACCAT					
				TAGTA					
				AATTC					
				AACTG					
				CCAAA					
				GCTCG					
				TTGGA					
				ATTTT					
				AAAAA					
				AGTGA					
				ACTGG					
				CAACA					
				GTATG					
				CTGAC					
				AAACT					
				TGATG					
				TGACT					

				GTTGA					
				CTCAT					
				CCCTG					
				TGCAA					
				TTTAA					
				TAACT					
				ATGAC					
				AGATT					
				TGCTA					
				AAGCT					
				GTCAG					
				AGAGG					
				TAGCA					
				AAAAG					
				AACGA					
				TACCT					
				AGAGG					
				GAGGA					
				AACAA					
				ATATT					
				CCAGG					
				ATGGT					
				CCAAT					

				GATGT					
				GATGA					
				GCTAT					
				ATGAG					
				GAATA					
				CAACT					
				TACAC					
				ATTCA					
				AATAA					
				AACAG					
				CAGAC					
				GAACT					
				TCTTC					
				AAGCT					
				TTGAA					
				TGAGC					
				ACAGA					
				AGAAG					
				GAAAT					
				GGCAA					
				GAAGC					
				TACAG					
				AGAGC					

				CTCGA					
				CTTCA					
				CCCAT					
				TCGAG					
				TCGTA					
				AAGCG					
				TGGCA					
				GCTAT					
				TGAAG					
				AAACT					
				GGGTG					
				CAGAC					
				TAAGC					
				CTGAG					
				AGATG					
				CACCA					
				TAACG					
				TCACA					
				CAAAT					
				GAAGT					
				CGCGA					
				CACGC					
				TTATT					

				GCAAA CTCTA AAGCA TCTGT GAACA AGAAC CGCAC TAGGC TGGTG AAGAA CAGTT ACGAT CCAAA CGCTC TCAGC TAG					
Pene lope	Pene lope- 120_ LMi		Locusta migratoria	AAAAC GGCCT TGACT ATGCA AAAAT GAAGT TCGCA	TGATGTT GTACTTC GTGCAGT GTCCGAA ATATTTC TTCCGAC TTGATCC	MNPQPPKLR SQFKVHKLNHPIRPISNSLNSAYH DLARFLHETLKKSFVFQNNYSVPNSHVLVHKV KDLDCGPDTKLFSLDIKNLYTNVPVVETLKIVE KNLCQFKKLSSEQITDFMNLINVVKYNYFEF NGKLYQQDDGLAMGNPLAGILADIFINSLEEK FFHKFSTTTLGIVSYSRYVDDILVIYEGPTTGID HIFTLFNGLHENITFTCELENEARQLNFLDLTLT	SSF56 672	34	231

			AACCA	AAAGCAT	VEENKISFNIYRKETYTDQIVPASSMHPQCHKN			
			GAATT	TTGTATT	AFFHSAIHRAISIPLSKEKFEEEINLIKTIAVNNE			
			GCTTT	CCAAAAG	YDPAIVDNILKKKTAKRVTTLGTPIESDQKKK			
			GACGT	CCAATTT	YFSIPFLGPISYQIQRLLRNKYNCNVAFSTNNN			
			AGCTA	TTAACTG	LKKNFIHNLKSTCSPLDNSGVYKIVCDTCSSYY			
			ACATA	TTTTGCTT	IGQTGRAFKIRFKEHLLRKNGTNPQNSSFADHL			
			GTTAG	ATTAAAA	LITGHAPKAIHDLNILHTEKKGHRLDVLEELQI			
			TAATA	CTTTGTT	FKHFSRNDGLILNEQLQLRNRSFFNGIKPLLGL			
			ATTCA	ACTCGTT				
			TCAAA	TATTCTCT				
			TGCTG	CTACTAA				
			TGTAC	TATTGTA				
			ATGAT	TTGCCAA				
			AGTTA	AATGACG				
			ATCTT	CCTCAGC				
			AAATA	CATCTCT				
			ATAAG	CATGTTA				
			CTAAA	AACAAAAG				
			GAACC	CATTTTG				
			GATGC	TAGCAAT				
			TTTAA	TGTTTTTT				
			TAACC	CAATACT				
			AAATC	GTTTCTCT				

				TGACA	CTGTACA				
				AAGGT	GTAGCCT				
				TGTTC	AGTTGTA				
				CTTGG	ATTGCGG				
				TTATT	CTACTAG				
				GCTAG	ATGGCAC				
				CAATT	TACTGAG				
				CTGAA	GTAACGT				
				TATGT	CTTCCTG				
				CTCTA	TTCACAA				
				AAACC	CTCCACT				
				TTAGA	TAGCACA				
				ATTTT	CGCCAGT				
				TAATG	CTGATGC				
				AAAAT	CTCCGTG				
				AATAT	CAGGCGT				
				TTCGG	GCAGCCC				
				AACTG	ATAGTGG				
				CACGA	CTAGCTT				
				GGACC	TTTCTAA				
				CGACC	AAATTTC				
				GCAGA	TTTTTAA				
				ATTAC	ATTTTGT				

				AAAAA	GACGTAA				
				GAAGT	GCTTTAC				
				CAGGT	AGCTGAG				
				CGGCC	TTGTATA				
				GTCAG	TTTTATTA				
				TGATG	TGTGACT				
				CAAAA	TGTTTTAT				
				TATTT	TTTATAA				
				GCTTA	TTTCTGA				
				AGCCA	TGAAGGC				
				TTTCAT	AATCATA				
				AAAAA	GTCATTG				
				AACTC	CCGAAAC				
				TCGTT	CAGGTTA				
				AAT	AATAATT				
					TTTAATTT				
					GCGACCG				
					AGGGCTG				
					CWTTTTT				
					ATCCTTT				
					GACGTTG				
					TAACGGT				
					CGCTGAC				

					AATAGCA GCTATGT TCAAAAT TTTA				
RTE	RTE - 34_ LMi		Locusta migratoria	TAGAG AAACA GTAGG AAATG AAAAA GCCTG CATGA GGTGT CAAAG ACTAT AACTT TGCC	TAGTGAC TGAAATT AAGTGTA ACAGATT GAGATGG CTGGGCC ATGTAGA GCGAATG TCGGAGG ATCGAGC TGTTAAA AAGGTCT ACAAGGA AATCCAG GTGAAAG AAGCAAA GGGAGGC CTCGGAA ACGATGG	MGNAV VVKVHGSRETQESTRCSPGTTFNQTDI GNHRGMHRLIAAGRLQGGDRSLYGKMEYTR DPEEGKALKTQCFQIEGKGAKPEASGGSTIKEA LQETAEEVLGLEERVKRADGTMKIVKRLWKK EMRQGRKGKEENDKQDNKIYYRKFMKKRKE ADRICRRKKRELERDGLWSRGMDYGAREEYN DEEVRKFYKKVRELKKGFQPRTLQRRKASNH VLCRDKEGNFIGGEEQVLDQWAKYFSELLNV TTEREDKMSLQKEKTNETEGDQPWQNVSQDE RPVPIPTIEEVKRAIKSLRNNKAPGNDNISAEMI KGRGEHLTRLVHQLIVEIWEKEEMPDQWNMA IICPIHKKKDKANCENYRGIALVSVVYKVLAK VIAMRFAPITEEILRRLPVRLSAGQINNRPNFHY KANNGKVLDLHQLQKKYSGDYQCGFRQGRST TDQIFIIRQIMEKCYEYNVDVHQLFVDFKQAY DSVKRKKLYQTLMEMEFPSKLIRLVQMTLTST KCQVKIEGKLSKEFEVNQGLRQGDVLTLLFNI VLERVMRRVEIDNPGGTLFNRMNQNLAYVDD VCMLSRRLKELEEGFDRLEKEAEQLGVNNTKT	cd0165 0	345	632

					CTGGAGG ATGTGGA AAACGAT TTGAGAA AGATGGG AGTAGGA GATGGAG GAAACGA GCAGAAA ATCGTCA GGATTGG GCTGCGA TCATCAG AGAGGCC AAGGCCC TACATGG GCTGTAG CGCCAAG GAGTAAG TAAGTA	QYLFASRKNAVVKKEYIKLNRGN YERCEKFQS WGAVTEDSRQIREEEKSRIAAGN RAYWAMIKI LQSRSLSRKSKVTVYRTVIRPVVMYGSETWTM TVAAEEELLRRWERKILRKIFGAVIEVGQWRIRR NKELEELYRNLT			
RTE	RTE -		Locusta migratoria	ACCTG TTACG CTGTC	CAC	MTIKCKNKMYTLVNAHAPT NIDNKKNPTKVQ TFWEELED TIQKIPKENI ILLGDFNAQLGKERK YKNIIGEYPAHKRTNNNGIRLVELCQVFNLKIM	cd0165 0	419	678

44_ LMi			GTAAC AGGCG GAGAG GTGAT ACTCC GACGT GGCGC GTCCC AGGTG GCGGA TAKGG GGGTC CTAAC TGGCT TGCCA GCGGA CTTGA GCGAA ATAAA ATAGC TCTCG CGGAC CAAAC	TTHFKKHPKKMKTWKAPNQLLGEFQIDHIAIS RREFKTIQNIKVQKGAKIDSDHYLVKSKIKFIP NSVKNNKKPTRKNFETNNIDADTKNIFKSKLQ GETWEEMKSQIMKAAEESLPTKIKRKNIWWN DKCNEALDNRQTLWQKFKSNPTQINLERFREG RKIASYTIRNEKRLFEKLSLQQIQVHFTQHKSR DYYQSFKRHLNGYEPPSLNFRDKSGKLGTDNDQ ENCNILAHYFKDLLNCKAPPEKPEFKTTEEIPA ESHPPSLDEVQRQAIKSLKNNKSSGEDGITAELL KLGEKVTLHLGTIIEIWQTEKIPEDWRTALI HPLHKKGDRSDVNNYRGISLLSVPYKVLKILL NRIEVLDAQIGEYQGGFRKGRSCSEIINLKNI IRTRAGQPHVTTFIDFRKAYDSIDRESLIRLEE FGIDRKTREIVKQTLTDTKSKVKFRGKISDPFKI QTGLRQGDGLSPILFNITLEKIIRETLAEYTQEK VKGISFGGSVKGLTIHALAFADDIALSTNIXD AIKQIEILAKHAAKFGLQISQEKTKFMSNKFKE VNKLHTIIGCIERVKHFKYLGETITDTGTEKEA VKLRVAKLRRRAQILTMKLYNKKCLSRNAKLL			
------------	--	--	---	---	--	--	--

				ACAAA					
				CCCTC					
				TGGAT					
				TAACA					
				CTCCA					
				GTTGT					
				ATCTC					
				GAGCA					
				ATCCT					
				CGGAA					
				AGCTC					
				CACCA					
				CTCCA					
				ACCTT					
				ACAAA					
				TKTTA					
				TGGAT					
				TATCA					
				GTGGA					
				AAAGA					
				AAACA					
				AATTA					
				CCCCA					

				GGGAG					
				TAAAA					
				TCTCTC					
				GACCT					
				CCGGT					
				CGCAA					
				GCAGG					
				CATTT					
				CGGAT					
				TCTGG					
				GGAGC					
				CTGCA					
				CTTTCT					
				GCTCT					
				TGTTTT					
				ACCAA					
				GACAA					
				CCTGA					
				AAAAC					
				TTAAA					
				TTCAA					
				GAMGA					
				AATAC					

				AGTAT					
				CGGAA					
				CACTG					
				AACAT					
				TAACT					
				CGCTC					
				ATCAA					
				AACAG					
				GAAAA					
				CATAA					
				AGAAC					
				TGACT					
				CAAGC					
				ACTTG					
				AGCAT					
				CATGA					
				TATCC					
				AAATT					
				CTGGC					
				ACTAC					
				AAGAA					
				ACCCG					
				CTACA					

				CTGAC					
				GAAAA					
				CTCTTT					
				TGAAA					
				CAGGC					
				AACTA					
				CAGAA					
				TTTAT					
				AAAGG					
				AAAAC					
				CAGCA					
				GAGTA					
				TATCTT					
				TGGAA					
				AATGC					
				CCCTT					
				ATTAG					
				GTACA					
				GCTTTT					
				ATTGT					
				TAGAA					
				AAAAT					
				ATAAC					

				GGAGT CCGTA ATAAA CTTCA ATTCT ACGTC CAACA GGCTC TCAAC A					
CR1	CR1 - 2_L Mi		Locusta migratoria	AGTCG CCGAG TGGAC GTCGT CCAGT GCACA CCTGT CCAGC GCCAG CACAT TTTCA GTTGT AAGTG	TAAATAG TATAAGG TGGTTCG ATGAACC CTCTGCC AGGCACT TAAATGT GATTTGC AGAGTAT CCCATGT AGATGTA GATGTAG A	MNRDCDCCVRMQSELVTLRSQLQAVLTSSVAQ LEAAIEGRRCCWPVVGTERTPVVSTDRPAPAT CPDTASSEADTPPVVAWEVAPVRGRRRKTFRG TDREASPVRLANRFEALSVQDNVPESDGVSHP EVTQTQPARSGLSQRWVCLRDGCQGGEQGQA LRVHTGGVIPDVERVLPDAMKGTGCSQLQVV AHVGTNDVCRFGSEEILSGFERLSEVVKTASLA CKMKAELTICSIVDRDTCGPLVQSRVEGLNQR LRRFCDRVGCRFLDLRHRVVGFRVPLNRSQVH YTQEAATR VAGAVWRGLGGFLGVRSPLYAGS GYTGSGGCVAWTGRFFRIKSFGIEIENGQSQRV QVDHRRNVDPETIGITVVNCRSCVGVKVPQLQA LIESTETHIIIGSESWLKPEISSEEVFSKDLTVFR	cd0165 0	852	1140

			TAACC	KDRLNKIGGGVFVAVKSSISCSEVEVDSSCELL		
			GGCTT	WVEVVPENRKKLIIGSFYRPPDSDDTIAERFKE		
			CTCTCT	NLSLITKTYPTHHTIIVGGDFNLPSICWQNYIFKS		
			AGAAG	GGRHKTSSENALNAFSENYFEQLVHEPTRIVN		
			CTACA	GCENTLDLLATNNPELIMSIKTDTGISEHRAVV		
			TATTTT	AKLNIESPKVTKNKRKVYLFRKADRNKIHCFL		
			GACTT	KDSFHSFLTNLVGVDMWLKFKEVVSTAIESF		
			TTCAG	IPNKLIKEGGDPPWYTKQVRTLLQKQRKRHTK		
			TGCGC	FKRTRNPQDWRSYTEARNLARTSMRDAYNTY		
			TTTCCC	HNETLSRNLTENPKRFWSYVKYTRGKTQTIPS		
			TTCAG	LRDSNGDPADDSATKAELLNIAFRNSFTKEDE		
			TTCATT	TNIPDFELRTEARMSNVEVDIIGVVKQLKSLNK		
			AAAGT	NKSSGPDICIPRFLSDYAEAIAPYLTAIYNRSLK		
			TATTG	EGSVPKDWKVAQVTPIFKKGSRSNPLNYRPISL		
			TGTGC	TSICSRVLEHILYSNIMNYLEQNDLLTQGQHGF		
			TCGCG	RKYRSCETQLALYSHEVMSAIDKGSQIDSVFL		
			TTTTTA	DFQKAFDTPVPHKRLLIKLRGYGISPQICDWIGD		
			TTTGA	FLSERTQFVVIDGKASTKTEVISGVPQGSVIGPL		
			TTCAT	LFLIYINDLGEICYWSSAVPYLHKRFRGNLSSS		
			ATTCA	LRLFADDAVIYRLVKSSDQNLQNDLNKISL		
			CGTTA	WCKNWQLTLNNEKCEVVHMSTKRNPLNFVY		
			AATAT	KINQSKSRPHEYEEKSVKLCLQDKSIKVKAVN		
			CTGGT	STKYLGITITSDLNWKEHVDNVVGKANQRLRF		

				TAGTG TATTA TTCCG CCGTG TGAAT CGCGT CGCGA CAGTC GTCTG TCTTTG TTTTCT GATTA GAAGC GACAT ATAGT ATTTT GATTT TCGTC TGCGC GTTCT CGTTT CAGTT TGTA	IGRTLKRCNRTTKETAYTTLVRPFLEYCCPVW DPYQVGLTEDIERVQRRARFVLSRYRGDSVT EMIQDLGWESLQQRRLSLRRGLLTKLQSPTFSS ECENILLTPYIGRNDHQNKIREIRARTERFRCS FFPRAIREWNGR			
--	--	--	--	--	---	--	--	--

				ATTTA					
				TACTG					
				TGCTC					
				TCGTG					
				CATAT					
				TAGAT					
				TCAAA					
				TCCGC					
				GTTTT					
				AACAA					
				ACAGT					
				GTAGT					
				CTTCCT					
				TTGAG					
				CGCAT					
				CGCGA					
				CTGTA					
				GTTTTT					
				GTTTTC					
				TCGGC					
				ACTGT					
				AATC					

RTE	RTE - 14_ LMi	Locusta migratoria	CTCTG AAAGT ACTGG AGATT CTAGG AAACG CCAAT ACGGT CCTAG AATCA ATGAA AGGAA AATCC TGACA TTCAC CTCCA CGGGG TAGAG AGGGG GCAAC GTTTG CATAG CAAGC	TGAGAAC GAGAGAA GAAGA	MMTGfVTKQSGGKIVPNSDLRQGTvSKCvIKD DRRQVRIGTWNIrSLsICgKLEnvKKEmEEHDI DILGISEIRWPQKGDFWSGNfRIIySGKEGSyA GVGIILRKHVgTRVKSyVQySDRIIMLRIENDPI DTVIIQVYMPTSSHADEEVEQVYEQLLEEIMEM VKDKDNLVILGDWNASVGEGEDGNTVGKYG LGRRNDRGQQLVDFCkEKELIIANTWFQqPKR RRYTWIAPGDRARYQIDYIMVRRRYRNQAKQ CKSYPGADVATDHNlVMLKSNLRLRRKGQNT KGYCKYDVKKLRQDSVKERYQKQAQqKVKE IPNGTVNEKWIGLKHSILQAADREIGNIKKQAT QPWISDHTLELMEERKKFKNSKSEEGIKSYKR LRNMVNREAKKDKEEYIMRRCKEVEEKLEKG NIETAYELIKKQFKKVNGRSNNILDENGKRLFE NKEVGERWKRYIEKLYEDPDEISNLENEETVT KDQLGPSLTKEEYLKAVRELKNNKAPGCDAL TGEMIQGLDNETHNYIFKLIQDMYETGEIAEDF KKTHIIPKKPGTIKCEEHRTLsLISHMSKILTKII YNRIISKIEQNlDEDQFGfRRNRGTREAILMLR QIMDKMLDLKRPLFIaFVDLEKAfDnVnWNL LMEIVTNVGLDFRDRRIVHELYKNQRGVINIH GEIQEAHIKKGVRQGCNLSPVLFNLYIEEGIKE CKESMEGIKINGVQIKMIRfADDIAIIEESEENL	cd0165 0	545	801
-----	------------------------	-----------------------	---	-----------------------------	--	-------------	-----	-----

				GGCTA	QKALSRMNEILPAKYNMKINTKKTEIMVCSHE			
				ACGAA	QITTDIKINDVPLKQVNVFTYLGSKITPDGKSS			
				TGCAG	ADIKSRIAQAKLAFNNKRMLLTSRNIKEETKK			
				GGACC	AFIKCYVWSVALYGSETWTLGVCDKRRLEAF			
				GAATA	ELWCWRRMQKIPWTAKMTNEEVLRRIGEKRA			
				TCCCC	LVKTIMKRRDKLIGHLYRHNEFIINVMEGMVE			
				CGGGA	GKRGRGRPRLQYLKQIQRDIGVESYQEMKRLS			
				TAAGG	ENREVVRRRIAANQSSD			
				GGATC				
				CCCTA				
				TACCA				
				AGTGC				
				CAAAG				
				GCCTC				
				CATGA				
				GGGCG				
				GATGA				
				GCGGG				
				TATAG				
				GATAG				
				GGCAC				
				TCTCTT				
				GTCCT				

				TGGAG					
				TGTGA					
				AATCG					
				CTCCT					
				AAAGG					
				CGGAA					
				GAACC					
				AGCAA					
				GGTCA					
				ACGGC					
				ATAGA					
				ATGGA					
				GAAGG					
				CAAGA					
				GGGAA					
				ACCAC					
				TCCAT					
				TAAAG					
				ATCCT					
				TTTGG					
				ATATC					
				CCATG					

				TCAAG CAGCA					
Dap hne	Dap hne- 8_L Mi		Locusta migratoria	GATCT TTGGC GTGAC GTCAC AGTGC TGGAT GCTGT GCACA TCGCT CTCGA TTTTTT ATAAT TTTAG GATTA ATATA GTTAT TTTAA TTGTT ATTTC GTGAT AATAT	TAAGCTA ATTCATT TTTTTCTT TATTGAT TCTTTTCA CTTTGTCT TAAGATT GTCGTGT GCATAGG ACTACTT AAAAATT TTCTATT ACCATTC TTGTGTA AAAATTA ATGTATT GACTTGT TCCATGA CCATGGA GTCTCAC TCCTCAA	MIEEQERLICALQVQLKNAEEELGRLRRGRGA GEWELAVGKKVARRRYSDSFTLQVANRFDQL SELSGDEPPVAVGVGSTQQTSAFGKPKSVERS LRKKKVLLLGS SHGRGVGQQLQEVLGSEYQV TSIVKPSAGLDQVTCNIGDL CRNFSKEDQVVIV GGAGNSLDRDGEYDIGGDLER IATQTGGTNVH FVQLFQRHDRPCLNTAVSRVNMGLEKALMTE GMDHIAVVPVETIQRSDFTRHGLHLNRYGKGR LAKLIGDSVVGGGGXTHGKIPVVVGVR AAPFL DGCTFFRLKTADRY SCLKEVPLRXEPPLTKSK NISNKGITVFHQNIQGIRDKNEXLIDTDS DIIGI SEHFLSGVTIQRLPLPGYRLAAFYSRSSLQCGG VAMYVKKRFFIGVNRFFKNDFSSESIDFSKHCT EQIFECCAGVVNVNGTKLVIVVIYRSPDSDFK MFLPKLELVLGSLYRKYKNLIVCGDFNINCVG DSSRKSQLLDLFNSYNMMQTVFFPTRVQGNSS STIDNIFVHSSLLEGHSVSRIVNGLSDHDAQLLI LKSICSQTNVKHKYQLLRKGKVFVLKQMLNIN INYSEKVN PVAIETFLSLIKEQEWQDVYSADNA DDKYNAFLNTFIMLFESCFPLERSKQGTSSNRQ PGWLTKGIRVSCRTKRELYQNVRSSQNEDAV	cd0165 0	851	1112

			TCAAT	TTGGGTC	AHYKQYCKVLKNVIKNAKSMWYAKRIANSR			
			ATAAA	CTATGGA	DKIKTIWSVVKEVSGQQQQVEDITSVRSRNV			
			GGTGT	ACAGAAA	VTDKLDICAFNNHFLNIVGELNKTLVSTGNHI			
			TCTAG	AAACAAT	TFLENAPRLLPEMVLHDTSIGEIESIIKSLKTK			
			AAAAG	AAATAAA	DSCGYDGVSSRILKYCSPYVSPVLSHICNFSFQ			
			GATCG	TAAATA	NGQFPDRLKYSVVKPLYKKGERDNVDNYRPIS			
			TCAGT		MPSVFSKVIEKAVYSRLIDHFSSHNLLSNVQFG			
			AATTT		FRSGLTTENAIFSFLCEALDGLNKRLRTLGVFF			
			TAGCG		DLTKAFDCVDHKILLQKLEHYGIRGVAHNWF			
			TTTGT		TSYFNNRQQKVILRSVENGYDVQSQWGTIKCG			
			AGTGC		VPQGSVLGPLLFLIYINDMPSSITGDCKIFLFAD			
			CGTAT		DTSLVVKDVARNIDTVSNNNAVQEISSWLVENK			
			TGCTTT		LMLNHSKTQFLQFLTHNLTKADILITHNGHMIS			
			AAGTG		ETEQFKFLGVHIDSKLSWKAHVQDLVPKLNSA			
			GTGTT		IFTIRTISEISDFSTRKVVYYAYFHSLMTYGIVF			
			ATTTG		WGNSPDAKRIFLAQKRAVRAISGVSSRTSCRPL			
			TGTTC		FSSLGILTLASQYVFSLMSFVVNNISLFPRVSSF			
			GCGTG		HSVNTRQKSNLHLHHTTLTLVQKGVQYSAASI			
			TGCGG		FNKLPQELKNLSTNPHAFKSKLKSYLMAHSFY			
			CACCG		SVEELLEKLNN			
			ACGTC					
			TTACT					
			GCTTT					

				GTTTA					
				CATCT					
				GGTGG					
				CGAGT					
				GGTTC					
				CAGCC					
				GTAGC					
				GGTCT					
				CCGAA					
				AGCTA					
				AGTTC					
				CAATA					
				AGTGT					
				TTGTG					
				TTATT					
				GTAGT					
				TAAAT					
				TATTA					
				AAGTG					
				ACTAG					
				ATTCA					
				GGAGG					
				TGTTA					

				TCACG					
				CTAGT					
				TGTTT					
				ATTAA					
				ATTTA					
				GTACT					
				TTTCA					
				ACAGG					
				TGACT					
				TGCTG					
				TTTTTA					
				ACGTA					
				GTTGA					
				GGTTG					
				GGAAA					
				CTTTA					
				AGTTA					
				CGGTA					
				TTTGA					
				TAACT					
				CTTKT					
				ACTCT					
				CACTG					

				TGTTT					
				AGTTA					
				GTTTA					
				TTAGT					
				GACGC					
				TTGAA					
				TTTGA					
				ACTTG					
				TTTAT					
				ACAAG					
				CAATA					
				GTGTA					
				AAGTT					
				TCATG					
				TGGTA					
				GTGTT					
				ATTTG					
				TTCTTT					
				TCTGA					
				TACTA					
				TTAGT					
				TGTTG					
				TTACTT					

				GATTT					
				TGTAG					
				AGACT					
				GAGGT					
				TTGTT					
				GATTG					
				CTGTA					
				GTTAA					
				GAGGT					
				ACTTC					
				CTAGT					
				TCTTTT					
				GTAGA					
				AACGT					
				AACTT					
				AATAA					
				CTTAG					
				ACGTC					
				TAGTT					
				TTTTA					
				AATAT					
				TGTGA					
				AATTT					

				TACCA					
				TGAGT					
				GAAAA					
				ATGTG					
				GGCTC					
				TGTCG					
				TAGGG					
				TCGTG					
				CGTAG					
				TGGGC					
				TGCAG					
				TGTGG					
				GATTT					
				GTAGT					
				AAGTA					
				TTTTCA					
				TGGGG					
				GGAGT					
				GCAGT					
				GGGGA					
				AGACA					
				GTGGG					
				CATTG					

				TGGTG AGGTT CTCTC CTGGA ACTGT AGGAT TTGCA GCAGG AATAA G					
Jockey	Jockey-7_LMi		Locusta migratoria		TGACCTC TCTCTGT AGATTGC CGCCGCC TCCTCCC TCCCCCT CCCGTCC CCCGCTT GGCTCAC CCCCTCC TTTCTCC ACCTCCC CGGGAGC	PPSVVPIMAGSPHRILYHNIRSLPANKPLFMHT LSQHCVDAFLLNETFLQPRHSVSTTPYTLHRTD GPPPLARGGVAIGHLRHHPVRPQPPVNNPAEH LILTLFLPSLTVTLATIIYIRARVPIPYDFISHIDNS FTTYVIAADLNLHSRSAAELQRWHQFLPTLRG DLVPLPSPTRPESNTTPDAVLVSPNLWDACRR KSLTSLAVTTPPYSSPSTPAAVPAAPPVLCPLL GPLGRLQAEVLDLIGSDHAPVLLTFNPRGRPG RATRPVPPPRLVHDYSRADWDAYRRSINTSVA SHPLSLHTPDDIVHASSFLHQTAVADAVATHIPT KAIHPHRPPLPPQAVLLLRESRRLYRSFLRTRD RDTLLRHRQLQRHIRNLLNAKKRRDWRQTCT RLNATLPINSSKYWKAFHRLTGSHSVPAYPLL	cd01650	551	810

				TCCTTCCT	HNNLPLPDNASKAQLFADHLSNIFDLPDNPRF			
				TCCGGTT	DYSRFPDVHERTDTSVPPLAPGFQYLDHIPPSD			
				TCCCCCA	LNTPITAQDIVHTLRTKRNTAPGHDRITYRHLR			
				TTCTCCC	ACPPSFHSALASLYNVILFTGFYPDLWKTSRIL			
				CTCCCCC	MFLKPDKPPSEASSYRPICLTSVFSKVLESILSR			
				ACCTCCC	RIHEHLRLHQLLPLSQCGFRPSFSVDDQLLRLV			
				TTCCCCTT	HLLSVQVNSRRAAILVSLDIEKAYDRVWHSGL			
				CCCCCGC	LFKLQTYALPINYVRLIASFLSKRPSYVTVNNQ			
				GCCTTCT	NSRLFYPSAGVPQGSVLSPLLYLLYSADMPRP			
				CCATCCT	QPPVHLLQYADDTAFLALAPSLARSQHLLQSH			
				CCCCCCT	LSEFSSWCNQWLLTVNPTKSQAIIGRTTPSFRH			
				CCCCCTG	RDLYLTLHGRPIQTKTLRYLGLTLDRDLSWN			
				CAGGGCC	PHLKTVRQKALSRLNLLKCLSGRTWGLNPSTI			
				AGTACAG	LYTYRSLMRPVLCCFHPAWISAPPSYYESLKIL			
				TGCTAGT	ERHALRLAYRIRLPSPTRILYDLIPFPHLVQFLQ			
				GTTTTTA	HIRLKYIHRKIESHNPLVSPLLLDPGPLPRLYRH			
				AGTG TTC	IPPSLHLHTLYVLSEGGFRHLPLPDDVLHNRPD			
				AGTGCTT	IYPSFQL			
				CGTGTGT				
				ATAGTGT				
				TTCGT CG				
				CCCAGTG				
				TTTTTTAA				

					CGTTGCG				
					TACACGC				
					CTGTGGG				
					AAACCAG				
					TG TTCAG				
					TG TTTTT				
					ACCTCAA				
					TGTCGCC				
					GTCATGT				
					TATGTCT				
					CGTCGTC				
					TGCAGTG				
					AGTGTTT				
					TTTAGCC				
					ACCGTGG				
					CAATGTC				
					TTGTTTTA				
					ACAACCA				
					TGTATGT				
					TCTGTCC				
					ACTGTCA				
					TGTTGTC				
					TTTTAAC				

					ATTCTCC				
					CATATTT				
					TCATCTT				
					GTAATTG				
					TAAAAGC				
					GCCATTT				
					TATCACC				
					TTTAATTT				
					TATATTC				
					TGTTTTGT				
					CCATTGT				
					TTACCAT				
					TCGCTCG				
					GCTGTAG				
					AGCAGCA				
					TCATCGG				
					CTGCTGC				
					CAGCCTG				
					CCTGTAT				
					TAGGGCA				
					TTAAAAT				
					TCAATAA				

					AGGAAAA AAA				
Pene lope	Pene lope- 9_L Mi		Locusta migratoria	CACAC ATGTA GTGAT GACTG ACACT GTTAT CAATA AGTCA AAAAG AGTAT TGCTC AAGAC GAGAT TTCCG TCCTT ACAAA AGGTG GAAAC TATGC TATAA CGCCT	TAGTGAT GTTTTAT CTCTGAC AAGGACT ATCTCTG CTGATAA CGTGACA GCTTGGC CACGCCC ACTTTTC ACAGTAT ATATGGC GCTCTCA GACACCC GACTCAT CAGTCGG CAAGACT CAACGCA AGTAAGA AGACCTC CGAAGAT	MIGERKALKELNADNDIVVLPADKGNATIVLN KDDYESKILDLLTSGQYRKLKKDPTASVLKKT NTLIKSSSTSIPKEDKKALLKSEAMCPRLYGLPK VHKINIPLRPivsAINSPTYEIARYLATCLQPYIG KTDSYIKDSSHFIGKLEGLTLRPEDMLVSFDVV SLFTMIPVNEVMTFITDIFPEDLTTLFKHCLTSS QFQWGGDFYEQVDGVAMGNPLGPAIANFYM EKFEQSALGIAKKKPVLYWRYVDDTFVWPH GKEALDEFFSHLNNINKKIQFTMEMEKDSAISF LDVLVMREVDGKLKHKVFRKATHTDRLHK NSNHHPQQKRGVikSLVDRAKKICTPEHLDAE LKHLKQAFKNGYSDKEINRVLRPNNSLRTR DQTKRWKNTVSLPFIGKVTDQIGKILRKHNIRP VFRPTKKICQVLRsvKDKRPTLSACGVYKIPCT CGKVYIGTTKRSIETRLKEHKSCLRLGKTEKSA VAEHALQSGSHIVRFSETSVLSTMTNYYPRQY REAIEIYKHKENFNRKEEAMKLSDIWTVALKD RQVLSLTDCNR	cd0030 4	206	298

				GTGGA	GTCCGGC				
				AGTAC	GCAGCTC				
				AGTTG	TGGACGA				
				AGGAT	AACGTTA				
				ATTAT	GGAGAAA				
				TGCGA	CTGAAAA				
				ATGTG	GTTTTAT				
				AAGCA	GGACCAC				
				GGAAT	GGCCTTA				
				AAGAA	TAACCCG				
				CTCTC	GAAGTTC				
				CCGAA	TACCAGC				
				AGAAA	AACTATG				
				CTGCA	ATCCGGT				
				AATGC	CGTGAAA				
				AATTC	GCCTTCA				
				GGATT	TGCA				
				GAAAC					
				CACCA					
				GAGTA					
				CTATG					
				TCGCA					
				GTAAC					

				CTCTG AAAAG TAATT TG					
RTE	RTE - 25_ LMi		Locusta migratoria	CCCGT GTGGA GTTTG CTGGT CTTCC ATCGG GCACC TCCCC AGGTG GCGGA TAGGG GAATG CTCAC CAGAT ATGGT GGGTA CCGGG GAAAT AAAAT	TAGTGTA AAACCTT ATGTACT AGGTGTA TTCATTTT TGGGCGT ATTAGTA TGTTGGA GGTAAAC CTCTGTA ATGAGGA CAATCTC AATAATA AAATAAA ATAAA	MPRKNWNCGRRTGDEKRRKMMFGCWNVQGIS TKIDLLPAELDMFNIDVVVLSETKRKGKGEEE LDNYVHIWSGVSKAVRAKAGVSIMIQKKWKK RITNWTFINERIITVEMTLFAREVVIIGVYAPT DTKDKEKDAFWDTLRETIEKIPRRKELIIMGD MNGRVGIRESCKIVGKHGEAEYNDNGERLIDI CAQFDLKITNTFFKHKDIHKYTWQQNTKELRS IIDYIIIRQTSSFKAADVRSYRGAQCGSDHYLV KMKSFWPWKNATNDTSNINKMNCSEKVQNV HFNIDSLQDESIRTTFFKARMERTLDESFEGSTEE IYEYIKTKVKNVASEVLGIKENNPKRAAEWWS EEIETSVKEKRNAFVQWLNDKSEGTRSKYKEK KNEVEKKIRLAKNEAWERTCANVNSKLGFR AKEAWSVLKALRQDTKGKSNLQLVTQKEWE EYFKKLLNEDRDEYLEEGTVEENEHHDDEIXIS ESEVLQVLRTGKNGKSPGPGNINMEFLKYGGD KIVKLILQLFNKMJHGDSVPKEMKLGYSTIFK KGDRKICSNYRGICVTNTLMRIFGKIKNKLEK NFRTQQEQCGFTAGRSCVDHIFTLRQILEKHRE	cd0165 0	536	788

				ACCCG GGGTG GACCA AAACC AGCAA CTGCT GCCTT GTAGT ATGAT ATTGG CTTAT CAAAG GCTAA AGGAA GAAAA CCTTG AATAC AAATT ACCTG GTCCT CCAGG TTGGG GGTTG	KSKNVGLIFIDLEKAYDTVPRKLLWRALHRAN INTSLIKIIEQMYKDNICQVKIGNTLSQKFRTSK GLLQGCPMSPTLFKIYIDICLRTWSQKCNSMGL EIRDGVYLHHLLFADDQVVIAQDGEDANYMC NQLAIAYKNWGLKINYQKTEYLTNDPHELRIE GKKIKKVNTFCYLGSILETEGKSDSEINKRISSG RKVIGMLNSVLWSRNVNMNRTKKIYKSIFESA VLYGAETWTINQKHTKKLQALEMDFWRRSAR ISRKEKRRNTEVIKRMEIHERIDEVMDRKKLRW YGHVRRMEDTRIPKL VLEWQPEGRRRRGRPV TTWIKNVQLTMNRLGAEEDTQDRHTWRNIV NN			
--	--	--	--	---	--	--	--	--

				TGCAG TGGGC CAGCT CCTCA CTCAC ATAAA AATAT AAAAT GCTAA AAAAC CTAAT AAT					
Pene lope	Pene lope- 55_ LMi		Locusta migratoria	AAATC ACAGT AAAAG TAGCT AATGC TCTAG AAAAT AAATC TAACT ACATG TATCC	TAAACAA TCCCCT GCCAATC ATAGATA ATAGCGC AACAAAA TTATCAA AGATAAA ATATCCC AACATAT TTCGTTC	RRHLVLINKQDYIQKVFEFFDKNGIIKIPCDPTN KYIAQTRKIIKSCSSILSDKDKRKYIPMNPKTPQ LRGHPKTHKELSPIRPVVDSSINCPTYLISKKLNT LLQKYYTYKKKYTIKNSRTLAANINKIDIPING KFVSFDIINLFTNIPVKETINIIELNLLSXNKLPI STYEIINLLHLVLEQNYFKFEDNFYTQPDGVA MGCNLASTIADIFLNTLEERFFLHQETNKIIYYY RYVDDIIVLIDGNETDINNIHGLINKLHPKIKFT MELETNKTINHLDLTIQNVQNKHRFEIYRKPTT TTNIIHAHSCHPSSHKLAFHSMINRLINVPLDD AAKQKEMTILQYIAQQNEYDPNIILKLYNNKT	IPR000 477	124	263

			AAGAC	AATTTTA	SPRPPTTRTEITLQRETSKLNKPIYISTPYLGKIS			
			ACACA	AATAAAA	NKVQQIFRKTNTTFAFQNRHTLKHKFVTHHTK			
			GAATT	GCTTTGC	PSKFLNSGIYKISCNSCNKYYIGQTGRNFTIRYK			
			TTTGT	CTCATAG	EHTRESEKNPSTFYQHLQSENHTVDTIDNALEI			
			AATAC	ATATATA	LHTVQKGYEMNILEEMEVLHAIKNTNLLN			
			AGAAC	TTCTCTGT	DQTDfKHKNYFDNFIHILVHDNG			
			GATCA	GCAAATA				
			ACTCT	ATGTAAG				
			AAAAA	CATATTA				
			ATAAA	CATCTCT				
			CCACA	GTAGAGC				
			AGTTA	TGTCAAG				
			GGCAA	TAAATGT				
			ATAAT	CAAAATC				
			GTTAT	CGTAACT				
			AGAAA	TGATAAA				
			TAAAG	GATCTTG				
			CTGAT	CATGACA				
			AA	GCACAAT				
				ATTATAA				
				GTTACGT				
				GATATAT				
				AATACAA				

					CATAACA				
					GAATGTG				
					GTATGTA				
					GCATTTC				
					TTTTACA				
					TCTGTAT				
					AAAACAA				
					GTAGGCA				
					GCATAGT				
					ATAAATA				
					TTATGAA				
					AACTTAT				
					GTATTGC				
					AGGCATA				
					ACACCTG				
					AAGATGA				
					CTATTAA				
					GTCGAAA				
					TAAGCTT				
					ATMGTGT				
					AACTAAT				
					ACGAATT				
					AAAATAT				

					ACAGTGT TGAAATT ATATAAT TGTTATT ATTAAAA				
RTE	RTE - 5_L Mi		Locusta migratoria	CCCTC GGAAT ACCAC CTTAA CGTTG GTGGG GAGGT TTGTG TGTGT CAGTG ATCCA CTGAG CTATG CTGGC GGGAG CTGTG CTCCC GGTAG	TAGGAGA TACAGGA AATAAGT CAAGTCA AG	MNRSRMDKKDLTFSMKSDTRIGMWNVQTMW QAGRLAEISREMKRYNIQGDEKILGLSETRWT GFGEIKTQEGDTLLYSGREEEHRGGVGLLITQK TKKSLIEWYPVSERIVVARFASRVRPITVVQCY APTEAGSEEEKEQFYGQLHRTLLKAKKKDIMV VMGDLNAKVGNDNQGIVESIMGKHGIGNRND NGERLIEFCSNHRLVIGGTIFPHKNCHKVTWVS PDHRTENQIDHFTISKTWRTSLDVRNRRGAD VGSDHHLVIATLRLKIASTRNTQQKRRIKYNW DKLKNKEVQQCFSLLELSNRFQQLREDEEQDIE RNWKSIESTYIDTCEKVLGKRERGREEWITSET WEAVNQRKECKKRLNTARTQEEKVQARTEFN ELNKMARRFRDRKRAFINNLAQAEMAAQK NDSKTLYDITKRLTGKCFRSEKPVKAKDGKLL TSQKEQLERWREHFLEILNRQTVEEEVTEAAT EEEQEDVDISIEPPSRIEIQRAIQQLKNGKAAGM DNITPEALKVDVSTSVEILFPLFKKIWEEEEKIPT DWKKGLLIKLPKKGDLSECDNWRGITLLSVPS	cd0165 0	550	782

			GGTCA	KVLSRVILERIKKRVNSRLRKVQAGFREGYSC			
			CCCAT	TDQINTLRILEQSTELQSSL YLLFVDFEKA FDS			
			GCCAG	VMRKKMWKAMQEFGIPKKIIKITQEMYENYT			
			ACAGG	CQVLHQGMLSEPIPVD TG VKQGCLLSPILFLM			
			TCAGT	VLDLVMQKAMNRRRG IQWGLTGRLEDL DFA			
			TGGTG	DDLCLLASSYKDMEAMLNDLKTEASEVGLKI			
			AGATT	NGKKTKE MRCNNRNEALFFFKPRDRASQAVL			
			CCAGA	LPGKF SHPGGRTIEDIGNRINKAKGAFAMLRPI			
			CAAAG	WKSRELNSYTKLRIFNTNVKSVLLYGCE SWKT			
			AGTAT	NKQLISRLQIFVNRSLRNILRIWWPQVISNEEL			
			TCCCC	WRRTHQTPIDIEIRRRKWKWIGHILRRGKDNV			
			CTAGG	ARQALEWNPQGSRRRGRPRNTWRRSVVAEGT			
			GGGCT	EKG GKAWTEIKTLAQNRVRWRKFVDALCST			
			CCTGC				
			CTTGG				
			GAGTA				
			TGTGG				
			GTGGC				
			AACCA				
			CATGG				
			GCCCT				
			TAGCT				
			AAGTC				

				TGGCA					
				TTACTT					
				CCACT					
				TACTT					
				GTGTC					
				AGGCT					
				TCCTCT					
				GCATT					
				CTTTCC					
				TATCG					
				GCCTC					
				CCTTG					
				GCCAA					
				CTCTT					
				GATTT					
				TTCCG					
				ACCCC					
				GACGG					
				TATTA					
				GGTTT					
				CGAGG					
				CCCGA					
				GGGAG					

				TCTTTC					
				ATTTTC					
				CTTCTT					
				CTACT					
				CTTATT					
				GTCAA					
				CTGCT					
				GCCTT					
				GTAGT					
				TAGAA					
				GTGGG					
				AACTT					
				CAAAG					
				GCTAA					
				GGGAA					
				GAAAC					
				CCCGA					
				CAGAA					
				AATCA					
				GTAGG					
				CCTGG					
				CTACT					
				TAGAA					

				GGCAG CCCCT TCACC AGGCT ATGGG TGCTG AAGGC TAATA ACCTA CACAC CCGAG ATTAA ACTAT TACAG AAACA TCA					
Ving i	Ving i- 20_ LMi		Locusta migratoria	GGTGG TGCTA AATGA GTAAT AGCAA CAAGA TGAGA	TAAACCA GTGGTGA ATGCCCG AAACGAG CCTTTGG GGGATCA ATAAATG	MRLSIQIIIEKPAVDVIALQETHTADDQQIKSRG KIPGYKLIASNNHSCGLAIYVKHTINSLKQPF KMWSSNLRQAYIKNYQVVSSENIFTISAKI DSATIVNVYKPPRVNXPTPLPSQQHPTLFLGD FNSHHNLWGYCENDSNGEALCEWMETENLK VIFDAKDPPTFHSGCWNRGYTPDLCFTTQLHS SHTQIRRSALRISLTVSTDLSFSRDWNMCTPCA	cd0165 0	461	715

			CGGAA	ACATCCA	VTPETKVEFQESRKS NVDKATDLLQSLDENRK			
			CTGCC	CCATGCA	KIWHETVESFDHTHSSRKAWSIIRKLDATNQTP			
			CGAAA	TCACCTG	VRPKTAIKLNDFAKHLISNSKSPKDKQTRREIK			
			AGGCC	AAGGGCT	TQLNKKRAAPNSSKFSNPFTEEEVRDAIRNM			
			TCCCG	CAACTGG	KVGKALDSMGSILNFSHTVDQQLSNGFLTSSQ			
			GTCAA	ATCACCT	TFSPLIFHHYSRKPSRGSKRCHQKHESWKGA			
			AGAGT	GTTGGAC	GFDGIYPEFLTHCGPATIKWLSNFFSDILSTGNL			
			GCAGA	TTAGAAT	PPLFKKTKLIAILKPNKDPTKVESYRPIALLSVT			
			GCAAA	CTAAGAA	YKLLERLIYNRIHXTINKXIPIEQAGFRPGRSCN			
			CAACA	CTTGTGT	DQVFALTTHIESGYEKKFVSIAAFLDLSAAYDT			
			AGAAG	AATGTGT	VWREGLLLKFVSIIQCRKLTNLLSNMLSNRLFQ			
			TCCTTT	ACATAAT	IYSDNDRSKVFKLNDGLPQGSVLAPLLFNLYIY			
			TCAGG	TTAATAT	DLPATSSRKFIYADDICLVTSQSPSFTTAESTLTA			
			ATGCA	AACACTG	DMEILDYFRKWCLRLNLQKTVTSAFHLRNR			
			CAAGT	TACATAT	HSNYKPEVKLRGQTLQYCSTPKYLGVTLDRLSL			
			TCGTA	GATCCTG	TFRQHLSNVAACKVKTRNNIIQKLAGTSWGCSA			
			GAATT	TACATAT	QTLRSTALALSYSVAEYCCPTWMNSVHSRKID			
			GGACC	ATTGTCC	VQLNQTMRLITGCIHSTQTAGYQFHPTSSSTRS			
			ATCTA	TTGTCAT	DALLKIWRNIQKNPQLPIHSDTSTAETHQRLKSR			
			TTCGC	ACATATC	KPSWKTAHHLDVDTNFDINEAWKSHWEESSVS			
			ATACT	TTACAAT	TVIWLKTLQCEFQVSHSPDASGESSTASEQDK			
			CCAAC	GTATACA	GTADIYSTNGAGQHLPSTVELPYKQ			
			TKAAC	ATTAGAA				

				ATAGA AGGAA TAAGC AGGTC GAA CTATGTA CTTGAAC CAAGCTC TGTATCA TCATTCG ATAAATA ATAA					
Nim b	Nim b- 22_ LMi		Locusta migratoria	CCCCT CAGGG GCTCA GCATT AATTT GCTGA GTACG GGCTT GGCGA ACCCG GGGTC CTGAG CCGTG GGCTG TGTCG CGCCA	TGAAGAT TGGTCTC CAAGACA GGGATTT CAGACTG CAACTGA CCCGTAA GGTTCTG GGATTTG GACCTCG GAATGGG CTGCCTT GAACGAA CCGAATA AGCTGCG CGTAGTC	GQVWRSGIFFENASGAVLIQTASPAQSRALLA CDKLGDVPVSITPHKSLNMVQGLIYHRDLLSQ TDEELRANLERRGVHFVRRVYRGPKDQRVAT GAFILAFEGDTLPEKVKVMIYRCDVKPYVPPP MRCFRCWKFGHMSSRCTSSPNCRECGRPSHPD APCTPPPVVCVNCGQHHPCCSPNCGVLHRERKI MEIKTLDRDLTIQRHVKSMSGSIRAAPSGKDDNI FCRSHSGTSFTAINGQCSLLCGPYGPTTSTAPS MGGGTSSSVAPTTPPVGAITPSPSGTSPAPSR RRNSLLRLLSLGRGPWDTPFRGFSQCTGDPSQ RLKGPRPAGRRASRSSSNPDATTEKPSQASKP KGQREGRQKKSKKKETLVATKPQAPSSSASED EVEILASPEDLAIPGSPEsRIADTDGSEQLAADV MDEPVLLYGFSARRPVLLQWNCGGFFNRLAE LQQLLCRFPVLCLALQETWFPAQRTPALRGYR GYYKNRAAYDRASGGVCIYVLATVPSAAVPL	cd0165 0	993	1246

			ACAGT	AAGCCCC	HTTLEAVAVRIRAMQEFTICSIYLPPEAVSIAD		
			CCGCT	CACAGCC	IAALIAQLPPPFLLLGDFNAHHPLWGGTQTTGR		
			ACACC	GTGTGGA	GRDVEALLSQRDVCLLNTGAPTHFCVARGTYS		
			ATAAG	GATCCTC	AIDLSVCSPSLLPHFRWKVHDDLCGSDHFPILV		
			CTCTG	CCTGCGA	SLLHFQPAQHQPRLRRADWQRYASA AVVP		
			TGCGT	GCCTCTC	VPLPADVDM AVHQVTSSIVSAAEATVPTTSGS		
			GCTTC	GTAGGGA	PRRKSVPWWTPEIAEAIRDRRRALQRHRRHPT		
			AACGA	GTCTGTA	MENLIAFKRLRARARQLIRQSKRDCWEQYVA		
			CCATC	GTCCTCT	NIGSRASSSQVWSQVRRLFGHRPPAAIPSINIQG		
			GTTTCG	GCAGACT	TVCTDPTAIAEHMADHYACNSATAAYSAAFR		
			GCGCG	ACGTATT	ALKVRVERRHLSFHTRRDETYNAPFSQWEFQT		
			ACGGT	GGMCATA	ALAACPDTSPPGPDRIHSQFLRHLSAESQHQLLA		
			GGAAC	CGTGGTT	LFNHLWAEGAFPSQWRESIVIPRNVQCAIQSV		
			ATTGT	GACACAC	GIPNCPCCMPRYLPGTRPDQPVPSPPF SRVPTP		
			ACGTT	GGCCACC	ASRPVQPSMGRGGFPITMAGKHRYPVMKPGA		
			TCAGA	TTCTACG	DPQVADSYRPISLTNVLCALLERMVSRRLMWI		
			GCCCG	CGAGAGG	LEHRGLLSPRQGAFRQGRSTVDNLVALES AIR		
			GGGAT	ACCCACC	TAFSRREHLVAVFFDLRKAYDTTWRHHILATL		
			CTTGG	AAATTGC	HGWGLRGPLPLFIQNFLSLRSFRVRVGTSLSSP		
			CTTAA	CGTTGTG	RIQENGVPQGSVLSVPLFLIAIDGLAAAVGSSV		
			TCGCC	GTGCACT	SSSLYADDFCLYFSSTGMGVAERRLQG AITKA		
			CGGGT	ATTGACG	QSWAMTHGFQFSASKTRVMHFCRRRAVHPDP		
			CGCGA	TGGTCCA	ELYLGXDLLRVEQTHRFLGMQFDARLTWVPH		

			GGACG	CATCCTG	LQQLRKKCWRHLNVLRLSHTAWGAARTSLL			
			GTACA	GTGGACT	QLYKALIQSRLDYGSLVYDSAAPSALRVLPIT			
			AACTC	GCCCACT	IVASVSWGSPPESESAGSSYWLLVARIHASPDHP			
			TATAA	TTTAGCC	NHRLLFPAAVLPSPRRRPRSGLPPIAVRVRSLLS			
			AAAAA	GACTTGA	SLASLPIPTFHLYLSRTPPWFPRLQLRLDLSSD			
			ACCTC	GGCAGTC	SKASVSPMVLCRQFQALIDSFHGAEVVYTDGS			
			AATCT	TTTTACC	LVDGRAGYAYVHGDHVEQHSLPCGTSVFTAE			
			CAAGG	CTGCCTT	LVAISRALTHFRDCPGESIVICSDLSGLLALDQ			
			TGTGC	CTGAGCT	FMPKHHLVRSIQDAVFTLGESGRSVVFWWTPG			
			TACGT	TCCATTG	HMAFRETNESITWPSRPPHGP			
			GCCGA	ATTTTAT				
			GGAGA	CTGACGA				
			TGCAT	TGCCTCT				
			GGCTG	ATGGTTG				
			TTGAG	ACCGTGT				
			GTTGA	TTTAGGT				
			ACAGC	TTTATTC				
			CGCTA	GCGCAGC				
			GCGAG	TGGCTTT				
			CAACC	TATTGGT				
			TCTGA	CTATTTA				
			GGAAC	GTGGTGG				
			CTGCC	GGCCTAT				

				GCACC	TTGCCCG				
				TCAGT	CCTTCTC				
				TGTAT	GGGCTCC				
				AAGGC	CACCCCT				
				TTACC	CCAGAAC				
				CAGGC	AATAGGC				
				AAGCG	CGTTGTC				
				GGGCT	CTTTTAC				
				CTGCC	GGTTTAT				
				TGGGT	GACTTGT				
				GGACC	TTTTCAA				
				TTTAG	TTAAGGA				
				TTCCCT	TTTTAGT				
				AGCTG	GTGTGCA				
				CTCGT	GAGTGGC				
				GGGAC	TGGCTCA				
				CAATG	TCCTGTA				
				AAACG	TTATCCTT				
				AGCAA	GTGATCA				
				ATAAA	GCCAGCC				
				GCGTG	ACACTCT				
				GCCGG	TCTGCTG				
				CGCAA	TCCTATTT				

				AACGG	TAGTTTT				
				CGTCT	CTGTCTA				
				CTAAG	TTTGGTC				
				CGTGG	AACTCCT				
				CAAAA	GTCTGTC				
				CGAGT	CTCTGTT				
				TCAAT	ATATGTC				
				GCGTC	GTCATTT				
				ACTAT	TACCATT				
				GACCC	GACTGTC				
				TAAGA	TCTCTGT				
				CGTTT	ACCCGGA				
				CCTTC	GTATACT				
				CCTGC	GGACGTG				
				CACCC	GAGTACT				
				CGTGG	ATGCGAC				
				GAGGA	AAAAGGG				
				ACGAA	AAACTTT				
				GGGCC	TAAAAGT				
				CTTAA	GAACTCT				
				TAGAT	TAGCTGG				
				CAGAG	GAACAGG				
				CCATA	GGACTGA				

				CTCGC CACAA TACCT TGTAT GTGCG CGTTC AGATG GTGAA TCCTTC ATGAC CACTA AGCCC AGTTC TTTGTT GAACA GCTGA	TAACCCA GCAGTTT AGTCCCC TTTAAAC ACCCAAC AACAACA ACAAC				
Dap hne	Dap hne- 12_ LMi		Locusta migratoria	CAAGA TGGCG CCAGG TGGAA AGTAC TCACG TAACG	TAAGATG CAAATGT ATGTTGT CAATTAT CATCAGC AGTGATA TGAACAT	MLIKLLQKDDHTTRQPIAIQSVSTKDTQKCQN NPKRRKLCQLSLKIRLHHGWKKNLHQYQPQQI PLKKTRPNRRNERRQGPTEETSSEVSDKIMTPF LKILKKKKRKFKAFLFSVPSKHPVHKNKVQQLE VELQTLNSSLVCITEHWCKGHELTHVALKSYV LASHYSRTTLQCGGSCIYVKEGIVFKVRDDIIA LSEDKHIEISAIEISKLNVLKRLTVLCIYHAPSG	cd0165 0	634	915

			ACGAG	GTAAATT	DMDTFIAKLNTSIGHGLMYSKDSL YYVFTMLP			
			ATTTC	TACTAAA	VVTWILLQSLTQALDMVSKPKGRILICGDLNI			
			ATAGT	TATAATA	NTLNADGTCITFLNTLHSYNISPLINTATRITKN			
			GACGT	TTTTATTT	SASCIDHMLTDIEKDMCNVVVDHLGISDHSSQ			
			GTGTA	GTCTGTA	FLKVNINVETNTKLLKYIRVFSESKRSVFASQL			
			AAATC	AGTGTAG	ASESWAEVYRESTVNGKFSKFLSIFKFKSEEVF			
			TGCGA	ATGATGT	PKVLRVIKAQNKNWVTNGIRQSSKTLRHLST			
			AAAAA	ACTTAAG	IKNYNNTIQNFYNTYKKVYRKVLLAAKRYQN			
			AGTTA	ATTTTAT	DKLIEKANNKIKAAWDVTNRVTNRKKAGQINI			
			AAAGA	GTTTGTA	QFQLNGTSVNDPKKLANFANKYFSNVATKLQ			
			GGTAT	TTGTATA	QKFVNNTQHPHTLNNVPHTMFLQPSTAEVSK			
			CGTAT	CAAATAT	VVQNLKSKSSAGIDEIPVCLLKDCTQSISHPLA			
			GTGTT	GATGACA	HIINESFTKGCFPETLKHAKVLPIHKKGDPEIE			
			GACTG	ACATCCA	SIENYRPISLLSSF SKVMETIMKNRLMAYLTKY			
			CAAGT	TACAATT	NLLCKEQSGFRSGKSTSSAIAQFTNAVLEALDK			
			CTTGG	TTTATTGT	GDHVTGIFLDLTAKAFD TVDHTILLNKLEALGV			
			TATCA	ATACGGA	RGVANMWFHSYLQNRVQQVEITKTNSNHTIK			
			TGATA	TGGATAA	HLSEPEYINIGVPQGSVLGPILFLIYINDFPQSVK			
			AATGT	ATAAATA	HGKSVLFADDSNILITDNSQIKLNQKAEDTLVH			
			GTAAA	AATAAA	VSRWVENNKLT LNVKKT TASTSTQINPQCKKN			
			AGTAA		NSINFHIKRKPVTNFLKIKDENIECVSTTKFLGM			
			ATCTA		HIDSQLKWTEHITVLGKKIASACYALRIINSVC			
			AATAC		SSSCTR TAYFANIHSLISYGIIFWGTNEQNRQAI			

			ATAAA	FKLQKRAVRIMTNSSSRAHCTELFKYMGILTV		
			AGAGA	PCEYILQQSSTLEGPLYRIIQIYGNTNCSMRIHF		
			TCATG	TAVIYIKENINQYPPNSSIHEHETRSSRHLHLNR		
			ACTGG	KNTTKTQNSILYNGIKLYNRLPQDIKNKDKIQD		
			CATGC	FRNSLKKFLEKCYITVQDYLN		
			CGGCA			
			GTGCA			
			TTGTG			
			AATGC			
			CACGA			
			AAAGT			
			GAACT			
			AGTGA			
			ATACA			
			CTAAA			
			GTCAA			
			AAGAA			
			GAAAT			
			AATAC			
			GTGTG			
			CTGAC			
			AGACG			
			ATTTA			

				GTGGC					
				TATAA					
				ATGCA					
				AAACA					
				CCAAG					
				AACTT					
				CTGAA					
				TAAAT					
				ATCAA					
				GAACT					
				AGTAA					
				CGAAA					
				CACAA					
				AGAAG					
				TGGAG					
				CAAAA					
				TCATT					
				GCCAA					
				GAACA					
				AGAAA					
				ATATT					
				ATTGT					
				TAATG					

				GACGT					
				TCTGT					
				GCGAC					
				TAAGG					
				TGGCA					
				TGTAC					
				CAAAT					
				AAAAA					
				MTGCA					
				AACAA					
				ATGAT					
				AAGTG					
				CAGCT					
				CCGGC					
				GTTAA					
				TGCTTT					
				CAAAT					
				AAATT					
				GCCGG					
				TATAG					
				AAGAT					
				TTTGG					
				CTCTG					

				AAACA					
				GAAGT					
				CGCGA					
				ACGAT					
				CTCCA					
				GTGAC					
				AAATT					
				ACCAT					
				CAGAA					
				TGCAG					
				TACAA					
				GAAAT					
				CGAAT					
				TCTATT					
				GCAAA					
				ACGTG					
				AAAAA					
				ACGGG					
				TTCTG					
				ATACT					
				GGTGA					
				CAGCC					
				ACACC					

				AGGAA					
				AATTT					
				CCGAG					
				AAAAT					
				AAATG					
				AATAT					
				AACAC					
				ACTTT					
				CTATG					
				GCCAC					
				AGGAA					
				TAACA					
				AAGCC					
				GAGTG					
				CGCCT					
				CTATG					
				TGAGA					
				TACTG					
				AGAAA					
				CAGTA					
				ACTCA					
				GTCTG					
				TGACT					

				CACTA					
				ATAAT					
				GATTT					
				TGTCA					
				TTATC					
				ATGGG					
				AGGCA					
				GAATG					
				ATGTC					
				TATCG					
				AAATG					
				AGAGT					
				AAAAA					
				TKCCA					
				CACGG					
				TGTTT					
				AAAAT					
				CTACA					
				CTGCG					
				TGATC					
				TTAAA					
				GTGCC					
				AAAAT					

				CATAG					
				TTACT					
				AGCAT					
				GCCTC					
				ACAGA					
				TATGA					
				CCTGC					
				AAGAA					
				AACTC					
				AATTG					
				TAAAT					
				ATGGA					
				ACTTG					
				CAAAA					
				ATAAA					
				CAGAC					
				AATTC					
				CACAA					
				AATGT					
				GCAAG					
				TCCTTC					
				CAAAA					
				CKTAA					

				CGTTC TGTGA GCTTC CAAAT TCAAG AGATC TATTC ACAAA ACATG GACAA CACTA SAACA CGCTG GGAAG TCATA CATTG GTAAA					
Nim b	Nim b- 26_ LMi		Locusta migratoria	CCCCT CAGGG GCTCA GCATT AATTT GCTGA	TGAAGTG ACCTGAC GCGGCTC CGAATCG GGCACTG CCCATTA	MGVPLGGTNTHSKKRARLGSPPDGGLCNMSM LKS NRAHSAVQNVFLVIKRKEGSFEKISPFYIQ KGLEGIAGDLKSVKRLRNGTLLVETSSSQQAV NLRNAKCLGEYEIETEMHSTLNYSKGVVSCRD LVDIPNEELKEEWASEGIVDVQNITKRVDGGL VKSESFILTFNSPKLPEYVKAGFLRLKVRPYVP	cd0165 0	1150	1451

GTACG	ACGCACA	NPMRCFKCQRFGHTTLACKGEATCGKCGKAA		
GGCTT	AGTTTTT	HEGAECTPPAKCVNCSGEHPVWSRDCTVYLE		
GGCGA	ATTGTCA	EKKVQEIKTTKRISYGEAKKIFKAMQPPSFVTS		
ACCCG	AGAAGAT	FAAVVQKVPKVDVSTQTDVASVGTNTCTCE		
GGGTC	CCTCCTTT	CTCKAAVVXEPELPHAHQKKLLXSTLQSFCHP		
CTGAG	TTGTGAT	PRQRPLARPWHLQSQRLFRRSSLTFPLLSXTT		
CTGGG	GCCTGTG	ATRTSEKATGVNITELLSPSPAIAIGQALAPSKP		
GACTG	GCGTGCA	EAIPKVITDLSTPAVTNIKVAKPKKAPGKSSEQ		
GTAAG	ACTTACA	AKAGNKPSDYIEVIIPDLPSDSAPELMELDVPQ		
CGCCA	TAGCCAC	GQTSRPKTEPPARSVSPTKRKERVKIQPPRWLP		
CCAGT	ATTTTAA	YCSGTAGFRTHVEELRLLTRDRPMCVCCLQETC		
CCTCT	TGAGTGC	FKETDSPELRGBYFLYRKDDLGTGVRAGGVAIF		
AGACA	ATTTATA	VNATYHSSPVSLTTGLQAVAVSVYLRHRLTIC		
CCGTA	TCGTGCT	SLYLPPHDALDEDALRDLLTQLPQPFIICGDFN		
ACCTC	GCGAGGG	AHNVLWGSANTCPRGRAIERILLSAHLNNT		
TGGGC	CAGAAGC	GQSTHFCAATGSFSIDLSLSSPALAHSAQWEV		
ACGCT	CAATCTT	AMTYMAVTTSPSGFICQTQWYPPEPHHDGCFV		
TCAGC	AGTGGTG	RQTGPCSANWLNLNNVIASRNGWTILHXSLAL		
GACCA	ATCTGCC	FSSPCPLCSMGSGDDLHGSDHFPLWIHLPTDV		
CCGTG	CTCCATT	VPARTPPRWVLRQADWTLFSQLAEFEQRDSV		
TGGCG	TTAGCTG	QEWVDHITQXIHRAAEASIPQSTGQLQRHPVP		
CGGCG	ATGACGA	WWTEECRSAIRARRAALRKFKCRPTAENLA AF		
GTGGA	GGCGAGT	RLARAKCRKVIRESKKRSWQHFLNSVNRSTPS		

			ACGTT	GTGAAAA	TKVWETIRRISGRGGRCPVAAVMANGTLHTDP		
			GTGTG	GGTTTTA	RDIAQTMAEYFATVTSTAQQDPAFQRHRMGA		
			TACCA	AGATTTT	EKGS�DFRSTTAEEYNCPFSMWELDSACLQLV		
			GGGAA	GTGAAGC	IRALDMTGYITVCYSTLQNGTEKSSSHFLMSFG		
			TGGGG	GTCTGCA	RQVIFPTHGVKQFEGQFGLQVHYCRRIQLPFLH		
			ATCTT	CTCTGGC	VGARFCLSAARDTCPGHDRIHYSMLQHLLTERN		
			GGCTT	CTAAGCT	REILLSLFNVVWASGHFPDSWREAVLVPILKPG		
			GACCA	TTTAGGT	KDRTSPSSYSRIALTSCVGKTLERMVNRRLVW		
			CCCGG	TGGAGAT	MLESRQLLSRFQCGFRKYRSTFDNLAFLEAAIQ		
			GTCGC	TTTAGTTT	QAFLRRQHLLGIFFDIERAYDTTWRHLILRQLH		
			GAGGA	GATGCGA	EWGFRGAFFLLCPFFLHAILGIESVTCSRIVLS		
			TGGAT	GTGGCTG	KRTVFLRVAFMGFSWSLPLFIVSFLSPRYFRYR		
			AAACC	GCTGAGT	VGDVLSDRFEQENGVPQGSVLSVTLFAIAINSL		
			TCTAT	TCCTCCTT	TSVVKSPVQCSLFVDDFSIFCSSSSLATTTRQLQ		
			AAAAA	GTGTGTC	LTLLRLDDWAKKSGFKFSTEKSVCVLFNRSRSI		
			AAACC	AGCCAGC	LNLPELRMRDRTLHFKETVRFLGLIFDSKLTW		
			TTCAA	CGTGCTA	LPHLKDLKKRCLKALSILKCLSHNSWGADRTC		
			TCCTA	ATCTACC	LLRFYRAFVRSRLDYGCTVYGSARPSYLKMLD		
			AGGTG	TAGTCCT	VVHHEGIRLATGAFRTSPIPSLCAEAAEPPLSIR		
			TGCTA	CTCTTGG	QLLLGRQGYKMLSTPHAPAYHSVARPPMAGC		
			CGTGC	TAACTGT	FILGRGQHDPLGFARIVSHRWVYLILMFTSGLE		
			TGATG	GATATGA	QISSLAPSEAQHYFRLGSFKKGCTPDFTYRLNF		
			GGATG	ATGTTTTT	SKFIGITILRWSTPMVQAGDSLGCSTVFPTMLPG		

			CATGG	ATGGCCT	FASHSLFCLHSRASRDLSIRNAACSEHTVSYLL			
			CTGTT	CATTTAA	RFTRTAITDSLYPDEKIHSSYRTNCTPSSGGVSR			
			GAGGT	CTAGCTT	CHSAGSWHVGIRGNEVADGLQRRPAATTTYTI			
			GGAAC	GTCTCTT	CHSPAGCPLCGSAEVL SAVGRHVAAVTGNKL			
			AGTCG	ATCTTGT	RSVKPTVRRGNLLPVTDA			
			CTAGC	TTGACAC				
			GAGCA	TCGTCTC				
			ACCTC	ACTCTGT				
			TGGGG	TTTCGTC				
			ATACT	CGGGTGC				
			GCCGC	TAAAGAC				
			ACCTC	CTCGACG				
			AGTTG	TCTGCGC				
			TATAA	CCTGATC				
			GGCTT	CATAATC				
			ACCTA	ACCACCA				
			GGCAC	CACCACC				
			GTGGG	A				
			GCTCT					
			GTCTA					
			GGTGG					
			ACATT					
			TACTT					

				CCCTA GCTGC TCGTG GGACC AGCAT GAATA TTTCTT TGTTTT TCCCC TCCCA GTGGG					
Ving i	Ving i- 14_ LMi		Locusta migratoria	GAGAA GCCTC CCCGC GGAAC ATATT AATAC GTTTT GGCGT CTCCC GGGAC GGCGG CAAAT	TAACATT ACAATTA CCTATTG TACATGA CATTTAT CCTACAT TTTAGCT TATAAAT TTTATTA CTCTCTG GACATGG CCAGCTT	MLTIISVNIEGLSSSKENLLHEMCQSNKCDVLC LQETHRDNNHRRPNIPGMKLVAERPHAQYGS AVFTKPDQLQVKSTCHTEENNIETITVELKDYTV TSVYKPPSEIFAFTPPANFVKEKISFVIGDFNSH SHVWGYASEDKSGEAVLAWAESSQLSLIHDN KLPPSFYSRRWHRGYNPDLLFVSENIAIQSCKS VCPPIPNQHRPIMCQLFPAIRPQNVPFRRRYN FKKANWDRFSGLLDEKIQTINPAPEEYDLFVD TIRTCSKASIPRGYRSNYVKGLDSEAATNLDRY YTLFERDPFGEETLAQANLVTSSISEAKKKEW VKQMEELDMGRSSNKAWRLKRLATDQLNL KHANVKSNIQIAHQLLKNGKPVRRNRNVQNSR	cd0165 0	461	717

				TCCTC AAACC AGAAC ATCAA CACAG TCATT AAAAT GCCTC ATAAA GAACC CAAAA GCTCT TTTCA GAACC A	TAGCCGA AAAGTCT TAGCCCT ACAAAAT ACATGTT CATGTAA TATCTAT GAAGAGC CTAACAA ACCTATC TATCCGA TGTCTAA TGTACGT ATAAAGT TTGATGC TTCTGAC ACGATTA AATAAAA TAAA	QIERDISQETSEISKPF ₅ SMTELDKAISKSRNGKA AGLDDICVEQIKNFGPAACKWLLAXMNNCMS ACKIPKIWRKTRVIAIPSETWEEPDDXXSYRPIS LLCHLYKTLERMILNRLTEVVEPLLIPEQAGFR TSKSCTAQILKL ₅ VQHIEDGFEERKITGAVFIDLT AAYDTVNHRLLLQKFYTSCKDYGMTQMIGNL LQNRFFVDFQGGQSRWRTQKNGLPQGSVLA PLLFNIYTNDQPLPEGTQSFYADDRAITAQGN TFETVESKLSEALEQLAIYYKENQLTPNPRKTQ TCAFHLRNRQATKKLVSWDGIPLVHCATPK YLGVTLDRTLTYRDHCMNTKQKVAARNNIVR KLAGTTWGAEPSTLRTTTLALCYSAAEYACPV WYNSSHAKQVDIPLNESCRIITGCLKPTPTDKL HCLAGIAPPSIRRKVAASKEKLKSVTSAAHPLF GQQPARRRLISRRSFLHTTECLTDTPQNTRTEM WRNKNVHLHDWLVPTEELPPGHGEKRVWR SLNRLRSGVARTKSNMHRWGLSDESVLCDCG QEQTTSHLMQCELC ₅ PTCTVEELMKATPNALD VASFWSKIS			
Ving i	Ving i- 4_L Mi		Locusta migratoria	GGGGG TCACA GCAGA AGCGT	TGATTCA ATTGTAC TATATGT TGTATTT	MRLVNEIAHEKHGSAIFVRQGLQVSGSRTIQH NNIELIGVELDNTEIISIYKPPAQTFNLNIPLLSH TRDNMHTVVLGDFNCHSTQWGYSETNEDGHL LEQWAENNQMCLIHDPKQPPSFNSGRWRKGY	cd0165 0	412	667

			TCCCG	GTCTATC	NPDLIFVSKDLSALSYKQVLNAIPRSQHRPIKLE		
			TTCGC	TCTGTTTT	IRPVVQPVEIPFRRRFNFKRAKWEDFGNKLDES		
			GACCC	TATAGTT	LKLIEPKPEHYDKFVQAIKDASFKTIPRGCRTE		
			CTGTC	TTAAATC	YIPGFTDKNKDAYERYCDLYDKSPLAEETIEC		
			GAGTC	TGTACTT	GEELLEALSDQRRQTWMKTVVGMDMTHSSR		
			TGATA	TATATTTT	KAWRLIKKLSNDPTIRTPHVNITPNQIASQLLL		
			CCAGT	ATCTCTG	NGKVNGKNNAAARSSLKVKKSKESNVSDVDFS		
			CTGGT	TATTTGA	VEELEEAISSLKVGKAAGLDDIRNEIKNLSEN		
			ATCAG	CTACTGC	XKKWLLKFFNCIWRTAKIPRIWRHAKVIAVLK		
			TCCGC	TTTCTGT	PGKDPNDPKSYRPISLLCSLLKLFERLIMNRIGD		
			GAATT	AAAAGCA	YIDSCLIDEQAGFRPGKSCTAQALNLTQFIEDG		
			CTGCT	CTTACTTT	FERGLPTGVVFIDLTAAYDTVNLRTLISKVHR		
			ACAAG	GTAGTTT	MTNSDRLAEVLRALLFNRRFFVDFQQKRSRW		
			ACGAG	AAATTAT	RNQKNGLPQGSVLSPLLFNIYTNDQPIGKKTCS		
			CTACA	GTGCTGG	FIYADDRAIACQAKTSNEIEAILTEDLTTLTEYY		
			TCGGA	CTTCTTTC	RRNHLKPNPSKTQVCLFHLNNREAGKVLDIHW		
			TCAAT	AGTTACA	NDVRLNNCSHPTYLGITLDRTLTYKKHIEKTK		
			ATGCC	TTTGTGA	AKVQTRNGILNKLTSAKWGANAKVLRSTALA		
			GAAAC	TGTAGTT	TCFSVAEYGCPVWERSHHTKKLDVVLNNTCR		
			AAAAA	CGTACAC	KITGCLKPTRVDYLYSLSGIAPPNIRRSFLSSIER		
			GATAG	GAATAA	NKQTNDRHLLYKNQVPKRLQSRRSFMYSV		
			CCAAA		EPTVASIKSLWQEYYPEALITEEALPQGSHLDW		
			AGAGT		QHWKTLNRLRVGVGRSKQLMKRWGYTDSDE		

				GGACT	CECGAIQTMEHLLKCPLVDASRECTQEDLLKY			
				CGTTTT	TGTAAAYTDFWTSKI			
				AGAGC				
				CCACC				
				TGACA				
				ATTAT				
				TTCGC				
				TAACA				
				TCGAA				
				GGAAT				
				TTCAG				
				TATCC				
				AAACA				
				AGATT				
				TAATT				
				GCAGA				
				AATCT				
				GCAGA				
				ACCAA				
				CAGCT				
				GCGAC				
				GTCCT				
				GTGCC				

				TACAG GAGAC CCACA GGGGC CTCTC AAAAT AAGCC AAACA TTGTA GGT					
CR1	CR1 - 3B_ LMi		Locusta migratoria	TTCTC GGTAT TGTA TCATG AATAG GGACT GTGAC TGCTG TGTGC GGATG CAGTC TGAGC TGGTG	TAAAAGT GGTTCGA TGAACCC TCTGCCA TGCACTT CAATGTG ACTTGCA GAGTATC TATGTAG ATGTAGA TGTAG	MEPLREMAAKEGKKANVHVCIPGGVIPDVER VLPDAMKGTGCSQLQVVAHVGTNDVCRFGSE EILSGFERLSEVVKTASLACKMKAELTICSIVD RTDCGPLVQSRVEGLNQRLRRFCDRVGCRLD LRHRVVGFGVPLNRSGVHYTQEAATRVAGAV WRGLGGFFFFREL CGVDWAVFFFFRLEGLGNT QKGFQSQRVQVKDRTNTDLKTIGLTVVNCRSC VGKVPQLALIESTESQIIIGTESWLKPEISSAEI FTKGQMVFRKDRLNKIGGGVFIAVTGGLSCSEI EVDCSCEL VWVEVIPDNQNKLIIGSFYRPPDS DSIAEMFKESLDQISK TYPKHTILIGGDFNLPSI CWQKYMFKTGGKHKTSSEIVLNAFSENYLEQ LVHEPTRIVNGCENILDLLATNNPELIKSIKTD	cd0165 0	772	1022

			ACACT	GISEHRVVVARLNIGLSKSTKIKRKYIYSKKQI		
			TCGCT	KIRLTLSKKIYLFKKADKNSINAFLSENLQSFET		
			CCCAG	NNVSVDQMWLTFKEIISTTIDRFIPNKLTKRGL		
			CTTCA	TRRGQTQNRLDHCYKSKEKSMPNLNAQKIPKT		
			AGCTG	GNLLQKLEIQIHTKQINKEGADPPWYTKQVRS		
			GAGGC	LLQKQRKKHAKFKRTKNPQDWQSFTETRNL		
			TGCGG	RTSMREAYNSYHNETLSRNLVENPKKFWSYV		
			TCGAG	KYTSGKTQSIPSLHDRNGDATDDSSAKAELLN		
			GGGCG	TAFRNSFTREDEANTPEFVSRTAASMRNLEVD		
			TCGTT	TPGVLKQLKSLNKS SPSPDCIPRFLSEYADAI		
			GTTGT	APYLTIINYNRSLSERSVPKDWKVAQVTPIFKKG		
			TGGCC	NRSDPQNYRPISLTSVCSRILEHILSSNIMNYLE		
			GGTCG	ENDLLSHSQHGFRKYRSCETQLALYSHDVMS		
			TGGGA	AIDKGCQIDSVFLDFRKAFTVPHKRLLIKLRA		
			CAACG	YGISPQLCDWIRDFLSERTQYVVIDGKSSSRTE		
			GATCC	VISGVPQGSVIGPLLFLIFINDLGDNLSSSLRLFA		
			GTTCC	DDAVIYRLVKSPEDQKQLQNDLDKISAWCKN		
			ACCGA	WQLTLNNDKCEVIHMTTKRNPLNFSYTINQSN		
			TCGGC	LQAVNSTKYLGVTTITNNLNWKEHIENIVGKAN		
			CTGCG	QRLRFIGRTLRRCNSSSTKETAYTTLVRPILEYC		
			GCAGC	CALWDPHQVGLTDDLEKIQRRARFVLSRYR		
			GACAT	GESVTEMMHELGWKSLKERRYLLRRDLFAKF		
			GCCCT			

				GATAC	LSPTFSSECEHILLAPTYIGRHDHKNKIREIRSK			
				TGCCA	TERFRCSFFPRAIREWNYREIS			
				GCAAT				
				GAGGC				
				AGACA				
				CCCCA				
				CCGTG				
				GTTGC				
				GTGGG				
				AGGTG				
				TCCAG				
				TGCGT				
				GGCAG				
				GCGGC				
				GAAGA				
				CATTC				
				CGGGG				
				GACCG				
				ATCGT				
				GAGGC				
				CTCCC				
				CGGTT				
				CGTCT				

				TGCGA					
				ACAGG					
				TTTGA					
				GGCGC					
				TATCT					
				GTTCG					
				GACAA					
				TGTTC					
				CTGAG					
				CCGGA					
				TGGTG					
				TCGCT					
				CATCC					
				TGAAG					
				TCACC					
				ACGCA					
				GCCTG					
				CAGAT					
				CCGGG					
				CTTTC					
				ACAGA					
				GGGTG					
				GGTTT					

				GCTTG TAGTT GGGAG CTCCA ACGTT AGGCG GGTA					
Pene lope	Pene lope- 59_ LMi		Locusta migratoria	CTTGA TGAAG AAAGA CGGAT ATATA AATAA AACGC TTGAC TTCTTT AAMTG CAAAT AACAT TACAC GTATA GACAA AGACC	TAAAGTA CAATCCA TCCTACC GTAATAT ACTTCCA AATATGC AACAAAA CGCGATA AATTATA GTAGCAC CGCAATA TTATCCC TATATTA CAATATA GACACAA ACAATTT	NKCSDSTNCESDKQPSNLIKKNKIICENYTY SRQYTVKSSTDIAKIQDIKIPDNSRFLSLDVTN LYTNVPIDETIAIIKENLLKYRKIPKQSLNNSTIT IILSHNYFSFNHGVYTQNDGLAMGSCLSGTLA NIFLDHLENKFFEEQNQWTKVVYYFRYIDDT IILFNGCIEEIEDFHQRFNNMHKKITFTIEHEINN SINFLDLTISKTNVHTFNIFRKPTTTDIIHNSSC HPQSHKEAYFRSAIDRITKYPLCNKNINSELNT LKDIAISNGYNPELVNNLYKSKLNSKQKSIKDE ITLTKTETRKKDKFVILPYWGQISDRIKYALKK TNLKIGFYTNNTLGTNIIHKIEKIDRYSQSGVYK IECSKCDKAYIGQTGRPFGKIRFREHINTSENNVS AFGTHLRESKHSTGSIEENLKILHIVCKGQLMD VLEELQIYTHMTKYPEQMLNEQTEMNRNKKFIE NFHNTLASI	cd0030 4	123	210

				CTACA	ATTACAC				
				AACCG	TTAAAAT				
				ATTTC	AGTTCGG				
				AAGCT	AATGAAA				
				TTGAT	TTAAAAT				
				AAAAA	TTATTGT				
				GAGCT	ACGATAT				
				ATAAA	TTGTATG				
				AATCT	GTACTGT				
				GAGTT	TTATGTC				
				TCTTTT	ACACTGC				
				AACAC	ATATAAC				
				AAAAA	CAACGAT				
				GAACA	ATTGTTA				
				AAACA	CAAACAA				
				GATAT	AATTAAG				
				ATAGT	CAAAGAA				
				TGTGA	TATTTAA				
				ATCCA	TAAAAAC				
				CAATG	ATCTTTG				
				CCCCA	GATCAAC				
				AGCTC	AGCTCTG				
				AGAGC	TAAATTA				

				ACAAC	CACCAAA				
				CAAAG	GTAATTA				
				GTCCA	CAACAAA				
				TAA	CAGCGTG				
					ACCTAAA				
					CAGTAAC				
					CAAATAT				
					AAATGTA				
					TATATGA				
					ATGTAAG				
					TAATTAC				
					AACTTGT				
					ACAGTCA				
					ATATAAC				
					ATTGCCA				
					TATTAAT				
					GTATTTT				
					CATGTAA				
					AACACAC				
					TTTGTTCT				
					CTGTAAT				
					AGGTGGC				
					TTAGAAA				

					ATGGGCA AAGCCCG AAAC TAG TCGTCAG CATTTCA AATAAAA GCAGCTT TGGTGCA TATATAA AAATGTC TTTTCTAC AATATTT AAGAGCT GCAAAAC GCGACAT ATCAAAA ACATTT				
Pene lope	Pene lope- 6_L Mi		Locusta migratoria		TGAGAAT GTCCTCC GCAGTCG GGGACGA AACGTTG GGTTTCA	GKQNCKLLSVHITISNGVCSSDWDRIDRLLHD TINKRVLTTSSRQRKKFDSL RPQPTISQLDSSRT VINLSGHLLSDDERSVLARGGNFAVSPRVVPT EEIIANIEAGIRNLDKVTAE EIRIEAVRVLARAK PPKSNLTTGEKNAIKSLNEDDSIIILPADKGNCT VVLNVTDYHCKIKELLD PQT YRKLSKDPTDKV	cd0030 4	354	438

					ATAATAA ATCCACA AGACCAC GGCATAA TAGCCCG GAAATT TTAGTAA TTTTGAC CATTCG GCCGTGA AAGTTA CATTTA CA	LRTVTQLIKKSSIEQCVQKSLFNSVALPPRLYG LPKVHKEKENVPPLRPIVSAIASPTYLIAKFLTT LLRPHVGRSDSYIKNSRHFISKLNILVNPEDIL VSFDVVSFLT MVPVNEVLQQLNLLFPPDIADLF KYCLTTTYFKWNDDFYEQTDGVAMGSPLSPVI ANFYMEKFEEQALKTANKKPAIWFYVDDTF VVWRHGREELDRFLQHLNGINSRIQFTMEKEA DGKLNFLDVSFVKKGNGRLGHTVYRKPTHDT RYLHRDSNHHPQQKRGIIKTLADRARNICEPEL LDAELKHL SNALLKNGYSSTEIRRALRPTRRVE NDERVPVKSKVFLPFIKDVTDRIGKILTKRNVA VIYKPTRKIQDHLKSAKDARKPLEKAGIYRIPC SCGDVYVGTTKRTVKKRLDEHKGNCRRGEID RSVAEHVFQPGDHLIQFDKTQVLAATNGYH ERLYREAIEIVKHPNNFNRMEEGVKLNEIWIPV LKKKCNSLPPWDDINGRRQSATTNCARMSKTS DVTPLRASPRKRS LASVSSEPGCASDIQKCHDP P			
Ving i	Ving i- 27_ LMi		Locusta migratoria	GGGGG CGAAG TTGTTC GCGTC CGGAA	TGACAGA GATATGT TTTATCTT ATTAATT AGACACT	MSATTTIPNGSCQGQQNCQQNDLNRAIGPVLN VCHLNIEGISRSKCDFLSKNLLSNNTDVLALQX THAEDLKN CARGNIHGYRLVGEIGHRTYGVA TYARQDLDDVSIHESCD DDIFVLAVKVS GTTI VNVYKPPNIQWQD TVLPVFEHPSLYVGDFNSH	cd0165 0	490	744

			GCAAT	AGTATAT	SSHWGYSSDNENGIALAEWAENNNLSLVFDA			
			CCTCA	CTTACTA	KDSSTFFSARWKRGYSPDLFCFCSCDQTLRSLS			
			AAGCT	AAGTATT	VTRKVLGTFPRSQHRPVLISVGIKIPLIRSFPRPR			
			CCACG	ATATGTA	WNFNKADWEAFTSKLEDNIRFIPPTTENYDRF			
			TGCCC	GCTGACA	AGLLKATAKACIPRGYRKTYIPGWSQECEDLY			
			CCTGG	TTATACC	NEFISTEDSNTADLLLNSLDNARREKWVKKVQ			
			GTGGT	ATA	SLDFTHSSRKAWSLLRKLSTSVSEKKLPPTHDP			
			TCTAA		PSDIAKRIVALSKSKRQQSDRLISYKLNEVKKC			
			ACGGA		QQAPLEVSKPFCLDELHVAISDVKTGKAAGID			
			CAACA		DIFPEFIKHAGLKTRLWLLNLFNNILKDSSLPK			
			ATCTC		AFKRSKVI AVLKPGKPADRADSYRPIALLSVC			
			TGGTA		YKLLERLIYNRIVPILTASIPVEQAGFKQGRNCC			
			GCTGG		DQVLALTSFIETGFEEKCTTSVAFLDLSAAYDT			
			TCCTG		VWKKGLLVKLYNLIPCILLRLLNSMLSGRLIK			
			TCACC		IYLGNTESGFRQLHDGLPQGSVLAPILFNVYTA			
			ACGGT		DIPATQSRKFIYADDIAIASQHRDIHVTETTLTS			
			GATCT		DLTLLRNYFAMWKLKPNPNKTEVACFHLNNR			
			CCCAG		MADYKLNISFDGVQLEHSEHPKYLGVILDRSL			
			CT		TFKYHLHKTADKVKTRNNILQHLAGSSWGAS			
					AKVLRTTALS LVYSAAEYCSPVWLNSAHVRDI			
					DVQLNRAMRLISGTLMATPLQWLPILCNIVPP			
					DLRRKEALMREYYKILSTPELPILQDLPPQVSR			
					LKSRKPPLKTAVQLHESGFSPKSNWARTWDSF			

						NGXNSIWIDDPSKELQGMDLPRREWSLTNRFR TGVGRCNYWKFKWGQADDQSCDCGADIQTM THILTNCPLRSFSGGMTELHJVSQEAVQWLKE LDLAL			
Dap hne	Dap hne- 6_L Mi		Locusta migratoria	AATAT GATAA ACAAA TTAGA AAACA CAAAT GTAAC AATAG TAAAT CAACC TCGCG ACATG ACGCG TAAAT AAATT AATAG AAAAA GTAAA CAAAG	TAGACAT GCTAAAA ACTACTT AATAAAT GTAATAT AGTAATG ACTCATT TTTGTTGT GATGTGC CATATAT ACTATGT GTATATA GAAATCT GTATACT TATGTAA AAAACCT GTAAGCA ATATCCT ATATCAA	KEQKISKGAQWCPSAINRPIQDQDRSSGTRSKK SNATPHPARPQQKGERKTSKPQILEPSADLVVF HQNVQSLTNKIDEFQIMLENELKKVSVLCLTE HWLQDDTLNMVHFPNFILGSHFCRKKRYASP QHRKQSFVVADLHKSIHLHPSFPEAFQRKIPW KYQDLSYCMKTYLNXECVLHKLSKKNRHIIIC GDFNIDFSVGLETGPKTDILNLLLTFTNMTPTIQT PTCVTKTSATTIDQIFASVPMYKARTLCTGFG DHTAQEITIKMDLPLHCKTSSLTITRSFNLQNT NTFTQYL TNESWEEVYSCRDTNEKFEKFLSIFQ HYFELCFPLKKRRIKKSNIKSSWITPGILTSRE KRVLHTISKQTKTTPEFDTYFKRYKQILNKVIK KSKKLTNFSYIEKPQIKCNVDYCERNRQNTGH THSNINLLENGEEITDPHRVADLFNVYFTEIAS NLIDKQQKPASENKNQKLFIRKNNFPIPNRRK GSTEYYKRIKTKIFIRHKCATVVTKPLTHLCNS SLSNGVFPEKLKLA VKPLFKKGDKRNEKNY RPISLLPGFSKILEKVFYRRLISFIEENNFLSNTQ HGFRKQKSTETAIFDFVNEALSTIDQKEHVSAL	cd0165 0	538	802

			AAATT	TAATTTG	FLDLTKAFDMIDHNLLLKKLEYVGIRGLAHK			
			GATGA	ATGAAGT	WVTSYLKNRAQIVEIAHQEGSNIYTYQSERKDI			
			CTATG	GGATGCT	KYGVPPQGSILGPVLFLIFINDLETSSSSQKYIKF			
			CAAAA	TAAATAA	ADDTNILIKGKTAAELQHEIQNCTSHIEEWFTT			
			AATTT	ATAAATA	NNLVINIEKTVTMHLHTIQNKHPLTPTIKLFNK			
			CGGTA	AATAAAA	KLETVNSTKFLGISIQDDMKWDTHIKYISKKLN			
			TGTAA		TICYGVRILARCTTKETLKNVYYAQAESLLRY			
			ACTTG		GLIFXGNSPESRYCFVAQKRIIRSIENATPRKSC			
			ATTGA		KPLFKALQILPLPCLYILEIVTYIKKISVAQNNL			
			TGTAA		VYTNKSVHTHDTRQTLDLHSEHLRTTLRQNGP			
			GCAAG		LQMGEKIVQQTSSPYQNNQRTPEVQIGCKKLF			
			TTAGA		TRKIIYKINVITLSVNICQLNLMSCFLTNYCNY			
			GAGAT		CHLLYDVDCIII			
			CTATG					
			CATAC					
			CAGTC					
			ATGGT					
			CTTCA					
			TCTCA					
			ATTAC					
			AGTGG					
			GAAAC					
			AGTTC					

				TTAGC					
				TGAAG					
				AAACT					
				GTAAT					
				TCAGT					
				AAAAT					
				TTAAC					
				CACAA					
				GAGGT					
				TTAAT					
				CTGGA					
				AGCTG					
				TACAA					
				AGTAC					
				TTTGG					
				ACAAA					
				TGGGT					
				AGTAA					
				GGAAG					
				GACCA					
				GGTCT					
				CCCAG					
				TTGTC					

				AAATA CGTGG CCCCA GTGAC TGAAG AAACT AATGC TGCTG CAAGA CCTGA AGTCC TGAAG ACGCC AGCCC AACAC CAAAA TGTT TTCTA GACAA AATAA					
Nim b	Nim b-		Locusta migratoria	GGGGC CCACA GCTCT	TAGCTCT TCCTTTTG TTGTATTT	MGDNFAMTSSSSSSTLPFTQKRKLDESKSKKV LPNLPKFLVVSRTDGHDFSSVNPFIQRGVDAI AGPVKTCRQRNGTLLLETEGAFQVQKLLRTT	cd0165 0	987	1244

36_ LMi		TTTGT	TACCCTA	LLHTFPVKVEPHRTLNSSRGVVFSRSLDGLSDE		
		GGGTA	TGTTTTTA	ELQEYLSDQGVTAVHRVMKRVNNNLTPQTQTN		
		CGTGT	GCTGTTT	CVLDLHRPKQTVFLTFDKVDLPSKIKAGYEIIS		
		GTGGC	TCATATA	VRPYVPNPTRCYQCQRFNHTRQSCPNAAKCV		
		GCACG	GGCTTAT	TCGRDAHEGDCPPSPRCINCMGDHAANSREC		
		GGGCC	TATCCTT	PVYKDERAIQEIRVKEKVSTSAARKLLASRKP		
		CCGAG	CGGTATA	AVLPAGTYSNVLSSPRPTKETGTQTCDFTFSDT		
		CCGAA	TCTCAGT	VVRSASPkiARSTSPMPSAHSKSQSTSVSQSR		
		GCAGC	TTTGATT	ATKSEARTSKKDPVREDFLKTSSSQPSAAHSSN		
		CTTCCT	CGGGCGC	RTASKKAHKKRSSSPPRRDSSVTPSSSRPR		
		TCCCG	TAATGAC	PSSVSPGRSTACRSAGISPAEEAAPSDQSRKEA		
		CCCCA	CTCAGCA	EQMSEPMEHDCPSVDCSSSAASQGPQRPSRS		
		GGCTG	GTTGTGC	PLAFSLFFFSAFFMALLQWNIRGIRSNRDELRL		
		CTTCA	GCCCTAA	LLRLHCPLIVGLQETKL RPTDHINFQYTA VRF		
		GTCCC	AACCACA	DLPPVEGIPA HGGAMLLIREDVYYHPIPLQTPL		
		CCCCC	AAAAAAA	QAVAVRVTISTFTFSLCNIYTPSSSAVLRADLL		
		CCGGT	A	HLVAQLPTPFL LIGDFNAHHPLWGSPASCQRG		
		TTATG		SLLADLFNHLNLVCLNTGEPTFLSDTTHTYSHL		
		CCTTC		DLAICTTPLAHRFEWYVLS DRYSSDHFPCIISLL		
		GCCCG		HHTTSPCSTSWNISKADWGLYSSMAVFHDQN		
		GAGAC		FPSCDSQVNYLTDVILTAAEHSIPRTSTSPHRVP		
		GGGTG		VPWWNAACSDAIRARRRALRLFQRRPTLENCI		
		TCCCG		NYKRLRAKCRRVIKESKKACWTAFTSSFNIRST		

			TGTGA	PSSVVWGNLRRLSGTKVHSPIPGLNIANQVLV		
			GTTTA	EPVDICNALGRSFAEVSSSANYDPAFLPRKQAE		
			GGTGG	EARTPSFTSPNRESYNATFTMQELERALALSRS		
			TCTCA	SAPGPDAlHIQMLKNLSPAGKCFLLRTYNRVW		
			TGTAC	TEGEVPTRWREAIVVPIPKPGKDKHLPSSYRPIS		
			CAGCT	LTSCVCKVMERMVNARLVWILESRHLLTNVQ		
			TTTTTC	CGFRRHRSSIDHLVTLSTFVMNNFLRKRQTV		
			TTTTCT	VFFDLEKAYDTCWRAGILRTMHTWGLRGLP		
			CGGTT	LFICAFLKDRMFRVRVGSVLSDTFRQENGVPQ		
			TCTCTT	GSVLSVALFAIAINPIMDCLPADVSGSLFVDDF		
			CTCCT	TIYCSAQRTCLLERRLQHCLDSLYSWSVANGF		
			AACCT	RFSHEKTVCINFWRYREFLPPSLHLGPVALPFV		
			TTGCTT	ETTTFLGLTLDRKLSWSPHVSSLAARCTSLN		
			CCGCC	VLRVLSGTSWGADRTVLLHLYWSTVRSKLDY		
			TTTCCC	GSFVYSSARPSILRRLNSIHRGLRLAIGAFYTS		
			TCCTT	PVDSL YAEAAELPLTYRRDILLCRYACRLLSM		
			ACTGT	PHHPSYPSFFNDSLDSHYGLSAIALPPGVRFR		
			ACTCT	RLLRQLDFTLPSTFQIGESPVPWLQAQVRVHL		
			TCTAG	ELTSLPKESTSDAVYRSRFLELRAWLENSVFIY		
			CCTTC	TDGSKTNEGVGCAFVIGDDGFQYRLNDHCSNF		
			CTCTC	TAELFALYQAVLYIRRHRSPSYVICSDSLSAIQS		
			CTCCTT	LRDPFSAHPVVHRINRSLQLLAADGPRVHLLW		
			TAGTC	VPGHAGVPGNEAADAATAAATLLPRTATYCV		

				TCCGC	PSNDIRGALSRRLISSWHNVSSLTNNKLRTIK			
				CTCGG	PLPVAWRTTSRRSRREEVILARLRIGHCRFSHR			
				CGTTT	HLLTGDPAPLCSCGHPLTVRHVLLSCPNFPTVR			
				GAGAC	IALNLPDTLEEILADDARAAARVLNFINLTNLS			
				TCAAC	KDI			
				CTTTCC				
				CACTT				
				CCTTCT				
				TCTTTC				
				CTTCTT				
				TCCTCT				
				TTCTTC				
				CTCCC				
				TGTGC				
				GTTCC				
				TGAAG				
				GCCGC				
				CCACG				
				CACTC				
				GCTGC				
				GTAAC				
				AGGTG				
				ACTGG				

				GTAAT					
				GCGTA					
				AGTCC					
				CAGCC					
				CCGGG					
				TTGGC					
				ATGTA					
				GGGCC					
				CGCAC					
				GTACC					
				CCCTG					
				GTAAA					
				GGCCA					
				GGCCC					
				AGGGA					
				GGGGT					
				GATTA					
				CCCGA					
				GCTGA					
				TACCT					
				TCCGA					
				CCGTG					
				CCGAT					

				TGGTC					
				CCTCC					
				GTCCG					
				TTCTC					
				GGGAG					
				GTGTG					
				ACCTG					
				AGGTG					
				TGGAC					
				AATCA					
				CCTAA					
				GGCGG					
				GTGAG					
				CCCTC					
				AGTGA					
				GGGTT					
				CCCCA					
				TTAGG					
				AAGGA					
				GCGCG					
				CCATC					
				GGAGA					
				CGCCG					

				GTAAT C					
Dap hne	Dap hne- 57_ LMi		Locusta migratoria	CCAAA CCATC ACTGG AACAA CAACC CCACT AGAAG CAACT CCATA ATTGA AGTAC ACAGT CAACA AAAAC TGAAC TACCA GCATC AGAGG CAACT GCTAT GCCTC	TGATGTT CCTTATA TAGCTAA TCAAAAT TGAGAAT ATAAGTT TTGTTGC AAACATT GTACTAT ATACCTC ATAAGTA TTCCCCC TATAAAT ATTGTGC TATGTAT CCCTTGT TTTGTTGT AAACATT GTACTAT ATACCAT ACCTACC	KTTYIVTTAEHRRAKDQEDLLCVRIFYGFSTRR KAVTNQPGDSSNTQVGAKSSVTVQESQCSFRII HQNFGLGKHNHNDLDVLAFFVNKPDVLVITEH WYPKDKISVMHPLMMNLCSSYSRMNKSGGG VAIFAKSGSNFIVSFYCVIDVSSLSVEGVIELSAI NYRWERQHIIIVGVYRPPSANIDIFLTNLNQLL VEIDNKYSKNGCINIITGDINIDTLSNNSTTKR LQLLLQEFNLNLLIKEPTREVGGRTRTCLDNIITN MDKFTYKTAVLKLAI SDHHA VEFQCVPNINLH LVKHKPPP GKMKRYETKRTINDKNIRNLTHVL QSATWAETNCYSYENFASDFFYYFDMTCPTKS VIVNNNEKCRKVPGRKMQKSTWIDAEVTQAR NMLTLFQCKMLENRENTKLKEAFNKYHEYK NLIGDKRSKHVAHILNNGNNTAACIWNVINSF RAGKDNVINDLVISYNGSVIECQQDIANIFNEY FSSVGKSINDRLHNNKYAFKGKASSYSIYFTPT NSNEIRNIVKTLRNSKSAGYDGLNINIVKSCIDS ISGPLSTAINNILTSGSFPDSLKIAQVTPIFKKGD RLKIENYRPISILPVL SKIIEKVIYNRIVTFLNKH NIIFENQNGFIKNKSTVTAAAQYIQLLHSIFDEV ITAIENKEYVTGIFLDLEKAFDCVNIPILLDKLW	cd0165 0	581	845

				AGCAC CACTG ACAGC AGCAA CAGGA CCAAC TGCAA CACCA ACCCA ACATT ATCAC CAGTA ACAGC AGCTA A	TCATAGT TCATAAG TATTCCC CTTATAA ATATTAT GCTATGT ATCCCTT GTAACCA ACTGACG TATTCAG AAAAATG ATACTTG TATCATT TGATGTT GAATAAA GTTTATTT ATTTATTT A	NIGVRGIAHDLIKSYLTNRKQFVVIRTSSGTYR STVTDMKYGVPQGSVMGPLLFLIYVNDIKYCT QNSKVILYADDTSIVCKHKELDQLEILSNCLTN EIVQYFNESHLSVSTAKTAVMDFNPSRQTDFEI QLCIGKDIASKENCVKFLGITIQDTLKWDVHV DTLASKLAKNIFAINMLSKHCNDTQVLR TAYH ALMSSNISYGIEVWGATTQANLNKILKLQKKA VRIICGAKPKDSCRSLFPKAGILTVVNMYIYKSI MLIRKLNLNLSNDTHNYNTRNKDKFHLLSHRT ALYERNPLYAGQRLANTIPSCITSLPTEKFKVT LKRILITNPHYSLDEYFCFAQNKENFIPDSLTSV			
Jockey	Jockey-9_LMi		Locusta migratoria	CAAAA CGGCG GACGA CTATA GGGCC	TAGGAGC CGTACGC ACAAC TA GGATCAC CAAGACG	MPPLRQMQR LTRDQNQQMRNPRNHTARRQR NGGENSTTQPLTR SERTQEQRDRTHQAREEDV ATAALSSSSTVPATPRENAPRFEDMVTLLQKL RDLIKEIPRQITAAV TAAIQAISTGPVMARSQQ GLTIAAWNANGIRTQEGEFRQFLREEAVDVCL	cd01650	621	939

			ATCAC	ACGCAAA	VNETFLKPGVHVRAANYVCYRTDRTTHGGGT		
			AGCGA	CCGCAAG	AVYVRASLAHHRVQLPPLQAIQESHSGSGRHI		
			AGTTT	GACAATC	HRQDHIRRSLQTTKDCGTTCRRRNFTVTPGKV		
			GGGGC	AACAAGC	FAAGDFNAKHHIWNRSVINNHGWHILREAQH		
			GCTTG	CAGGACA	NDAMAIGPNDPTHFPENGGQPDVIMLRSDPFP		
			ATAAG	CAGAATA	REWGTAGCYYVALIKGVPIYVISASTRNVLSSD		
			CCCTT	CCAGCAC	HVPVVLDDVDVDAGPAKRRGLPTNSTNWDRYR		
			CTATA	CCAGTTA	EHLGNSLQDAPMVEDAGVEDNVRTFTKNAIQ		
			CATTT	TTTTGGA	AVKRATPRLQRDPPDRSRQLPPDILAMITAKN		
			CCTTCT	GGCGTAG	RMFREWQLTRNPATKRAVNRMQREIKAAVK		
			GCCGC	CTCACCA	AHRHQEWSNKLATLRVDDGTGWKATQMFLR		
			CGACA	GCTTGGA	RRLKIPPLQVGPTIVSAPEDKANAIADVFEANF		
			AAACA	CTCCACC	TPVQDPIDPEHVRLVQERLPTFLNAQKSVQKR		
			TTGAA	GACCATC	CPGFYAASINEKLEAPTSSLTKCFDRCRQKPQR		
			GATTG	GGAAACT	TLQQRHCIMRTGHXPQAWKHAEVVPVPKPGK		
			TTTTCC	CATATGT	DHRLPSNYRPISLLPAJSKVFERVYLKRLMDHI		
			GTGGC	TTCCGTC	RRENILPDEQCGRPGHATTHQLLRLVEEALD		
			CTGCC	ATTGCAC	XTERREYFGAVFLDVSRAFXSVWHDGLLFKLL		
			CAGTA	GAGCGAG	DXRPAPAARPINSVVSQRQDVPGAHWKRPSXR		
			ATTAC	CGAGCGA	PNVGNTSGQYSLTFRERSXPSGMTVYSSSCWT		
			AGTCT		SGLPLPHVRLIRSYLQDRTFQVRTEELSTRRPIR		
			GGAAC		AGVPQGSVLGPILYILYTADMPRIDRVHLALY		
			ACATA		ADDTALFSRSMNARTLGTRLQRASDELSRWA		

			AAACA	TKWRLSFNGAKTQALIIARKRVPVDLPRLSVR		
			ACAAT	GTAVPWGRTAKYLGVTLDRLTWQQHVDDV		
			TGTCG	IGRARGRLRDLYPFLNPTSTLPPQLGVAIYTTLI		
			GACAT	RPIFEYAAVVWGNAAQTHIKRLQGLQNRILRL		
			GGGGT	ALHLPRDFPTEDLLRVTEIPPIQARFRQAAKLF		
			ATGGT	YQKAARSPNALIRALGHRVQRLNATRWPHELL		
			TTTCG	RSD		
			ACTCC			
			GCTCC			
			GTGAC			
			TCGGG			
			TGAAT			
			TCTGG			
			ACGAA			
			CGAGG			
			AAAGA			
			GACAC			
			CACTG			
			GTCTT			
			GGTTG			
			TCACC			
			TGTGA			
			CATAC			

				CTCAT					
				AACAG					
				GGGTC					
				TCTGG					
				CAAGT					
				TACCA					
				GGATC					
				GCTGC					
				GACGC					
				CTGTC					
				ACCAC					
				GGAAA					
				AACTC					
				AGATC					
				AAAGC					
				GGGGT					
				GAACC					
				ACAGT					
				GCTAC					
				AACTG					
				CCAAG					
				CTGTG					
				GGACA					

				CGTCG					
				CCCGA					
				TTTTGC					
				TAATT					
				ACCTC					
				CGCGC					
				TGTGT					
				GAAGT					
				GCGCG					
				GGAGA					
				ACACG					
				CCAGC					
				CGAGA					
				TTGCA					
				CACGG					
				AAGAA					
				AGAAG					
				ACGGC					
				CCACC					
				CAAGT					
				GCGTT					
				GGTTG					
				TGGTG					

				GGGAC					
				CACGT					
				CGCCT					
				CCTAC					
				CGTGG					
				ATGTT					
				CTGTA					
				TTCCTC					
				CGGGC					
				TGCTG					
				TAAAG					
				AGCCG					
				CGGAC					
				AACGT					
				CCTGG					
				GGCCA					
				CTAGG					
				AAACC					
				AGCGA					
				CAGTA					
				CGCCC					
				AGGAG					

				TAAGT T					
Ving i	Ving i- 21_ LMi		Locusta migratoria	GGGGG ATGGT GTTAT TCGCG TCCGC AGCAT TCCTC AAGCT GCACG TATCC CCCGG ATGGT GCCAA ACGAA TAACA ATTCT GGTAG CTGGA CTGTC ACACC AGTGA	TAAATGA AGCAGGA CTTAAAT GGCTTAA ATCACTG GACCTGG CTCTGTG ACGGGAG ATCCTAT TCACACA TAATTTTT AATCCCC TGAACGC AAGTACT ATGTACT GTAGTAA CAATAGT TTAATAT GGTGGCT GACAAGA ACCATAC	GLHQXNVDILALQETHIEEPLNNARGNIPGYRL VAAIGHRAYGVSTYVRQGLDDVTVKHNSND DTFILAVKVSEITIVNVYKPPNTQWTDTVLPVF EHPALYVGDFNSHHSQWGYPADDENGRRLTS WAENNNFSLIFDAKDNSTFY SARWRRGYSPDL CFCSCERSSKPM TARRRVLGAFPHSQHRPVIIS VGTQIPIIRSFP RPRWNFSKADWETFTSKVEQSI RFIPPIITTTVSRIFSWLLLNSVSPGASGKNTCQ DGVENATIFTRSFKTPKALTQPTICCSLLTLP GK RSGQSRLGDLYFKGRAEHQVYP SHNNYDRFA NLLLAI AKQCIPRGFRKEYVPGWSRECDDL YK EFQDTESPD TADHLLQSLD TARKEKWIKTVEN LNFTHSSQRAW SLLKKLSSGKPADNCTVTHPN PCDI AKRIVSSSKSARLP SDKRVAYKLNKILRS QFTPLEVSRPFTEEELEAAITTLQTGKAAGLDK IFPEFLKHIGPKTKFWLLTVYNAILATGNLP KA FKRSKIVAILKPGKSPEQAESYRPIALLSVCYKL LERLIYNRISQILDASIPVEQAGFRPGRNCCDQV LSLTSFIEKGYEMKKKTSAAFIDL SGAYDTVW RKGLLLKFYNIIPCPLLLRFLNSTLSNR LIKVYL GHKESA FRQLNDGLPQGSVLAPLLFN VYIADIP	cd0165 0	524	778

			TCTCC	GATAATA	ATQSKKFIYADDIALATQHTDLPTTEATLTNDL			
			CAGCT	A	ATLQEYFRTWKLKPNPNKTEVSCFHLNNKLA			
			ATGTC		NQKLNIYLDGVKLTHNNHPKYLGVILDRSLTF			
			AGCTA		KSHLQKTA AKVKTRTNILQQIAGSSWGANAQ			
			CCCAC		VLRTTALSLVYSSAEYCAPVWLNSAHVREVD			
			CAAAA		VQLNRAMRIISGTLLATPLPWLPVLCHIPPEIR			
			TCCAA		RQEALIREYKKILSTPQLPITQYLPQNLSRLKSR			
			ATGGA		KPPLRLASQLIAEDFSPNTTWAASWDXFEGRN			
			CCTTCT		KHLIPSPSTKVGGMNLPRREWTLNRFRTGVG			
			CAGGA		RCNYWKCKWGMTLDQSCDCGADTQTMDHIA			
			CCAAC		IDCPCVLSLRGWL VFTI			
			ATAAT					
			TTTAA					
			GAATA					
			TAGTC					
			CAGTT					
			CTCAG					
			TGTTT					
			GTCAT					
			TTTAA					
			CGCCG					
			AGGGT					
			TTGAG					

				CAAAT CAAAA TGTGA GTGTC TAAGT AA					
Pene lope	Pene lope- 80_ LMi		Locusta migratoria	GGGCG AAACG TCAGG GGAAA GTTTC ATGTA TGGAC CACGG CATCT CAGCC CGGAA GCTTT AAGTA	TGACTGA CAGCAGC GCCCTCT GTTTTGC CTATATA TACGACG CTTACTC GCGCTCT CGTTGCA GTACGCC GCTTTAC CTCCGAG GATGCCT TCCGCAG ACGAGGG CGAAACG TCAGGGG	MTTPAVKAFIYDQTPLRRGYVNIGSAVREHAK KSQQLCSLTFLRCRDTSTVPKFLKVKRTFRSA QAHRIYSRMESAMLRERIHQTRRALAETNSKL FQLHLLISNTMQCDDWNKIDNLTFRTMEASAE IAGNRQKRKFDLQKNVLVGTMEDNSRTVINL SDKVLTADDEVLVLAAGGNFAVTPRKVPTEDIV ANIEAGIRQLSQQSADVIRMETARILRTCKPPTS NLSQAQRKALRDINADENIIVLSADKGNATVV MKSEDYHEKIMDILDPSNYKQLRKDPTSKILK MTNQLVKASSFSSDDKKHLCKTEAYSPRLYGL PKVHKQLVPLRPIVSAIGSPTHEVAGHLASLLQ PYVGGTDSHIKNSCHFIDIIKELSVDPDDILVSF DVVSLFTMIPVNEAITCIADIFPSDIVALFKHCL TTTYFQYKNVFYEQIDGVAMGSPLSPAVANLF MENFEQRALQTARKRPAKWYRYVDDTFVIWT HGEEELDAFLHLNSINPRIQFTMEKETNGQLN FLDVTVIKREDGSLGHKVYRKDTHTDRLQK	cd0030 4	435	526

					AAAGTTT CATGTAT GGACCAC GGCATCT CAGCCCG GAAGCTT TAAGTAA TGACGAC ACCTGCC GTGAAAG CCTTCAT TCTATGA	DSNHHPRQKKRRDRLSNERTDHWVIRSIGKTH TLIVIFKKILTTTHDKKRGVIKTLVDRAHKICEP AHLDDIEIHLRVTFKKNGYSDREIDRALRPRR TTRTDEERQQPKGKVFLPFINKVTD RIGKILSK YDVETVFRPTTKIKEFLRSAKDKRHPLATPGV YKIPCSCGKVYIGTTKRSVNTRLGEHKRNCRL GHTDKSAVAEHVYQEGDHEIKFNETS VIAKTS HYYARLYREAIEIFKHRNNFNKKEEGLKLDKI WMSTLQQQHDNRLISIEEVGNSRDHSTADTT			
Ving i	Ving i- 15_ LMi		Locusta migratoria	ACCAA GGTGC TGCTG CACAG CAACA AGTAT GCAAG TAAAC CTACA ATGAA TAGTC	TAATTGT ATGTATA TATTCTT GTATACT TGTGAAA ACCATAC GATAAAT AAATAAA TAA	MLVILSHSHVWGYQEEDRNGAILNEWMDMY NLQLVHNAKDPGTFRSGRWQKDYSPLCIISK SGNNNTTQRKVIKSFPHSQHRPVIIQFGTEIPLV KSVPLPRWNFQLANWEQFRKSLDNSIQWIPPL PSNYERFLGAIKVNAKRYIPRGFRKEYIPGWSE ECDELYAEYQDSHSDSTADDLLRALNKNRRE KWHKTTAGLNFTHSSRKAWSLMRKLGTDPRL SNRVAHVNANSVAARLMNISKSKVEKAYTRT VKKELHIKRSQLEPHSDYSREFSIEEVETALKM VKVNKAPGFDGVYPEFLKAAGPKTKLWL TDF FNAILRTGKLPKLFKQAKIIAILKPGKDGT DAS	cd0165 0	335	588

			CTTTTC	HYRPVSLLSMTYKLLERLILNRIQETIDQIIPAE		
			AGGAC	QGGFRKHRSCCDQVLTTLTTLVETGFQRGLKSA		
			CAACC	AVFVDLTAAYGTVWREGLMLKLIQIIPCQKLA		
			AAGCT	FLVNNMLSDRRFKVFLNGQSSRWRTLNNGLP		
			CAGAA	QGSVLAPILFNL YIHDMPDTNSTKIQYADDLAL		
			CACCC	VFQSNDLKHCEQVLTSDLIKLSDYFHKWRLQP		
			AGCTC	NPNKTEVTAFHLNNKEAQRKLHVTFDGIEVPH		
			ATGGT	NYNPVYLGITLDRSLTFKEQCLKLSKKLASRV		
			AACAA	NLLRKIAGSNWGADANTLRS AALALVYSAGE		
			CTTAA	YCAPVWENSAHTTKIDVQLNEAMRIITGTVRS		
			GAATA	TPIQWLPVLANIAPQDLRRKAAKVKLIK NVKA		
			TGTCA	HKNSPLYDALQDPPACRLKSRKPPWCDTESTT		
			CTTGA	SFDVTSKWQERWENENPPHNGHLVQDPSKKVE		
			ACATG	GFHLPRQTWSRLNRIRTGQGRCAYNMKKWGF		
			AGTCG	LPSSDCDCGAEEQTVGHIVKQCPLRAFNGDLE		
			ATTTC	SLHQVTESALDWLSNLDISI		
			ATCTG			
			CTAAG			
			ACTCA			
			GATCT			
			TTCAC			
			GCCTG			
			ATGCT			

				GAATC					
				GGTGT					
				GCAAT					
				TCAGG					
				AACCA					
				ACCAA					
				TCTAC					
				AGAAA					
				AGAAA					
				TCTTTT					
				GGATA					
				TGTTCT					
				GGCTG					
				GGCTA					
				TAAAT					
				GTGAC					
				AGTAG					
				GCATT					
				GCAAC					
				ATATG					
				TCAGA					
				TCCAT					
				TTCTCT					

				ACAAA					
				ATCAT					
				CATAA					
				GACAA					
				ACATC					
				AGAGT					
				CTGTC					
				TTGGC					
				TATAG					
				AGATG					
				CAGGC					
				ACACT					
				ATTGC					
				GAATA					
				TCTAC					
				AAACC					
				TCCGC					
				TTATT					
				GGCCT					
				AATCC					
				ACTGT					
				TAGGT					
				CCTTC					

				ATCCC CCAGT TGTGT					
RTE	RTE - 56_ LMi		Locusta migratoria	CTTCC CAGGG GACAC AGCCT TG TAG TAGGC ACCTT CAGTG CCGGG CGGGA GGGCT TGCGT GCTTC CGTGA AGTTG AGAGC TGTGC CGGCA GTAGC GTAGC	TAGCGCC ACAAAAG AAGAAGA AG	MFYSCDDNRHEFGTGFLVHKYLVHKLHVMNFTPI NPRLSVIRIKGRFFNYSIINGHAPTEESIEDLKD DFYDALEKAYDKCPRNDIKIVLGDFNAQLGRE DMFQPVIGAHSKHGSTNDNGIRLILFAASRNMI VGSTSFPHKEIHKGTWKS PDGRTINQIDHVLID QRHKS NLLDVRTFRGPNVDS DHFLVMAKIRA RISNAQKERNHRTKMAAIKRAIDISAKETLGIE RESIRPTWFDEECKRVTEEKNTAYQKTLQANC TRSLLEDYREKRRIEKMTHRRKKREWTKAEIE CMELSKKKMKQGILKTICSITSLIRDEEGNIISE VNRLWKDGEDTLMTCESRDEQAEDISNDCEQ GLDDEQPLDGPGLI KKA KSLRKLKNNKAPGT DSIPAELLKYGGATLEQSLFELITLIWEQERIPN EWN EGILCPIHKKGDRMECRNHRGITLLNLGY KVLSNIIFDRLLPQVDKEIGMYQSGFRRGRSTI DQIFNLRQILEKTVEYRIGTHHLFIDFKAAYDSI NRNQLFHALQEMGIHGKLLRMVKMIVVQGN Y SDPLEIKNGLRQGDALACLLFNVALEKVIREA GLQTRGTIFYKSVQILAYADDIDIIARTVPALKE AFTALSQASKKMGLMINEDKTKYMACGKAY	cd0165 0	427	677

				TACTG	QPNMPSTFSVNGYTFERVDTFKYLGSTVTCFN			
				GCAGG	DTSTEIKIRLISANKAYFALSNLFQSRLLSRPTK			
				GTCAC	FTIYKTLVRPVLTYASETWPLTETDARLLDGFE			
				CCAAG	RKVLRRRIIGPVYDGEQWRKGTMQNCTLYMPM			
				CCGTA	LLWRTMEERYNAELYTIYADAPVTNIIRSGRIR			
				TTGGT	WAGHVARMEESEVTKKILFGNPGGQRGRGRP			
				CTCAA	RARWLDGVEEDLRKLGCRNWRRVAYDRDG			
				CTGAG	WRKLVGKAKVHPGL			
				AAGCC				
				AGACG				
				AAGAG				
				TGTCC				
				CACAA				
				CATAG				
				GGCAG				
				GGAGG				
				GCTCT				
				AGAGG				
				CGTAG				
				TGAAG				
				CCAGC				
				TTCCCT				
				AGGGG				

				AGGAA				
				ACCTT				
				ATACA				
				AATCC				
				CGGTC				
				CTCCA				
				GGTTG				
				GGGGT				
				TGGGC				
				TTGAG				
				GCTAA				
				CAACC				
				TCTTCC				
				TGTAA				
				AAAAA				
				GTCAG				
				GTTAT				
				GAAAA				
				CCCAA				
				AGAGC				
				AGAAG				
				CGGAC				
				AGATC				

				AATAG					
				ACGAC					
				GACAT					
				GGCAC					
				GGATA					
				TTGGA					
				CAACG					
				AATGG					
				TTAAA					
				AAGCT					
				GAAAA					
				CTGAT					
				CTTGA					
				ATTGG					
				TACCT					
				GGAAT					
				GTTCG					
				ATCTTT					
				ATATA					
				GCCCA					
				GGCGC					
				TCTAA					
				AGAAG					

				TTATT AGACC AGCAG GGGAA GCTCA AGGAG AGCAT TATTG CCATC CAAGA AATTC GATGG ACAGG GACGG GGTAT TGGAA AAGAA GGCAC T					
Ving i	Ving i- 31_ LMi		Locusta migratoria	GGGTG CAATG AAGGA GCCTC	TAGTTCC GGACACG GAAAGTA	MKTNKTTPPKRSFPRPLVTISINIEGISRDKEVLL SDLCKEYQCDFLLIQETHRGITRPSPKIDGMKM ILERPHEQYGSAIFARPGINVTASLTDRDDVEI LTVRTQQYSVTSVYKPPDSTFVFTEPANFSYQ	cd0165 0	473	728

			CCCAA	AGTAAGT	DINFILGDFNCQSTAWGYKETNSSGAELEKWA			
			AGATT	A	EVHDLIIIHDPKLPSSFNSSGRWKKGYNPDNIF			
			GCGAC		VSKKLAAQSEKLIGEPHPTQHRPIICTVSAVVK			
			AGCAA		PEQVPFKRRFNFKKANWKKFREIIDLEVTKILH			
			GATAA		NLRCMTLLSVFRMCLGRQYARMQNTVYRWT			
			AAAAA		SEDAKPLLERYQSSSTGNELMSVIAENRKAKW			
			AAACC		RDLVENLDMKVNSRGLGNCRTSTVMPQSQTI			
			GCAAT		NLLMSLQTSRHHTLLNGKPRQKIPQPLTRVLA			
			CGGGC		EKPHLKAPFTMSELTAAISSLKTNRAPGIDELR			
			GCACC		VSKLKTLAVIPYSGCTESQQIKNLGSNTLQWLL			
			CCGGG		DLMNACVERLHIPKMWRKARVVALLKPGKDP			
			CAACG		NDPKNFRPVSLCHCFKILERLTLSRIENVLEQ			
			TTGAT		QIIPQQAGFRQKGSCCGQVLNLTQHIEDGFERK			
			CATCA		EITGVAFIDLTAAAYDTVNHRLLRKVYDVTKD			
			GCGGC		YHLTSLIATFLQNRFFVCLQGKKSRRWRNQRN			
			TAATC		GLPQGSVLAPALYNIYTNDQPVPNTTRQFMYA			
			CTTCA		DDTAVAAQGTNFLDVEEKLSSSTLESLSQYYEA			
			ACAAC		NHLRPNPSKTQVCAFHLKNAEANRELNVTWR			
			CGGAA		GERLAHCPTPKYLGVTLDRTLTYKHHCNATRE			
			CGGAC		KVSARNNIIRKLTNGSWGANKPKVLRTSALALS			
			TTCTG		VSAAEYAAPVWSSSTHARKVDIAVNEMARIV			
			AGAA		TGCLKPTPVHKMYLIMGIAPPNIRRDIAAEAEK			
					TKQEKDMRHPLYGHQPADRRLRSRKSFLSRTR			

						ALEGSAAENRIARWEMDLRSPHKAAKEELAP GGDLPYTVWRTLNRMRVEVPKCKTNMQRWG FLKDNENTLCLCGAVQDPQHLLICPYLEKPCT MEDLWEANDKAISAANFWKLEV			
Dap hne	Dap hne- 7_L Mi		Locusta migratoria	GGCGG CTAGC GTGAA AATAC GGAAT GTGAC TCCGC GAAAA AACGT GAATT TTAGT GTAA AAATA AACAG TGTGT AAATT GTAAA GCCGA AATTC	TAATTTA AAGGTGA ACTGATG TAGTGCT GGAAACA GAGAAGT GGACTGT GTCTTTTC ACTAAAT TTCTAAT ACTTAAA TACCACA TGTTATA TTTATAA ATGATAA ATTTGTA AAATTTT CTTTCTTT CTTATGT	MESKIDVFGDSHAKGISQELQVISKSSSTVTGMV KPGAGFKEVTKNIVKYKYGKQDSVVICAGAK DIYENNAMKMYKQLLSLILVLSNTNIILVNFXL RHDLPEWSCVNKEIHRINEKLKSCCRLFSNVK LIDSSTLHRECFTKHGLHLNNNGKARLGNLIR DAIKCNSKWTVTELGWYKCPTAETPAQTATA PRSTIQKTVATPTEQPTTEPKLTATSEPTSEVTV KTSKFNQPESSAFTTTTVAEQTSVAVAPKQITSQ PESLAYTATTAVAVVVAHVVTPEEGSHLQSS HQKKAATFEKPGFLLGPGSEQQQIIGKHLNDC NQKNLQQGQLVQEADSNQGREGNINRTAVS KNSNLKIMHQNIQYASKKSEELEVLLAAEVPNI FIMTEHGLKDSELKVFNLEGYKNVAYYCRNE HKSGGVAIFVDSKVGTTPLDISQIAQISTLKQQ QYKSSWGGKCLIIIGIYRSPSSSIENFLLELEVLL QKVMKKYTFVILAGDLNIDMLRDNDCTRFRFQ DILGSYNLQLHVKSPTRCTETSSTLIDHILTNIPT ERQKVQVINTVIADHSGQTIELKDYAKAEDKK MIKYRQTNIANLNSLEILVKAEEWDDVYNEHS	cd0165 0	847	1111

			AAAGC	TATATAA	VEKSWESFCDILRHHINVACPVICKTVSAKQN			
			TAAAT	AATTAAC	KKWITKEIKKAREELIKLYEDYQKDKEQASQE			
			GAACG	AACATGA	AYRSKKKLYTRMIREAKASQYKHEIKTSTNIS			
			TGTGT	CTTGTGT	KTVWNLINDNRKESQTSKDTNIILEMEGVKVT			
			CCAGT	CTAATAT	DPLQVANSLNQHYVAIVEKLLNQXTPVTKSIK			
			TTACA	CATTGTA	SPKWVAESLYLLPTNELEVSKIIGALKKNKNTG			
			GCTTA	AATATGA	LDGISAKIVKHCAYLATPLTFLINLTMKEGVF			
			TTATC	TCTACAG	PNCLKRSKVVPVFKGGDKEDPGNYRPITITTVL			
			AGCGC	ACAAATA	SKIVERVINDRIVNFIDKHNILNEAQHGFRKGR			
			ACTGA	AATTATT	STKTAVANFITYITEALEQRRKVSGIYLDLSKA			
			AGATG	ATTATTA	FDCVCHKILLKKLNCYGIRGNAADIKTF AENR			
			GATAT	TT	QQCVELKHKCRNKEEKVYSEFQQVKYGV AQ			
			TACGA		GSILGPLLFILYVNDMPTEINEKVIMYADD TTL			
			AACTT		LVCGDTVKDLEKRATENFQMAERYFTENNL LI			
			AAAGA		NRKKTCTKFETKITKSSEVFTLKIGSADLAEV			
			TGAAA		DCNRFLGIEVDKHL SWENHIDKVC SKISTNIFL			
			TCAGT		MRKMCYTWTQKPCGLFENKHKYILNEKNVLY			
			GAACT		VDTETLLKMYYG LIFPHLSYGIPLWGGAADV H			
			AAACA		IQRLLKLQKKAIRCC THTETFKTSKESHQMLA			
			CATCA		KLSPRESCKNKFKEFKILTVPSLYIYMRQDTNC			
			AGATC		TKSVYIYETIVFLMVNCTAPLNSDIHNNH NTRTK			
			GCGCC		EQYHFFSKRLNMTDRDPINAGCRFFNKLPLSIR			
			ATGAT		CIKERSAFKRKLKEHLISGCWYKVKEYLTENL			

				ATTTC					
				GACTC					
				TGTCG					
				GAAAA					
				GTTGG					
				CGGAA					
				GTAAA					
				GTTAA					
				AGGAT					
				GCAAA					
				CAAAA					
				AGTGA					
				AAAAG					
				CAAGA					
				AAACA					
				GTGAA					
				ACAAA					
				GAAAA					
				KTGTA					
				GTGCA					
				CGAAA					
				ATTAC					
				TTCCA					

				GGTTC					
				TTACA					
				GATAC					
				GGTTT					
				GTGAT					
				ATCGC					
				TGTGA					
				CAAAT					
				GCAAG					
				TAGGC					
				CAAAA					
				CAACA					
				TAAAA					
				AACAA					
				AACGA					
				ATCCG					
				TAAAA					
				AACAA					
				TCAAA					
				GTGCA					
				AATAA					
				ACAAA					
				ATSTA					

				ATACA AGTGA CAGAA AAG					
RTE	RTE - 7_L Mi		Locusta migratoria	CAGGG ACAAG CATCT TGCCT GACGT GGTGG CGAGC ATGGG GACCC AAGCT GGCCT GGCAT TGCTT CCCCT TACCT GTGCC AGGTC CCTCC TGTTA	TGACAAG AAGAGGA TCAACCC ATGACAA GACCTGC CGCAAGG CAGAATA CAATTGA TGATGAT GAT	MDRNLKLDVLMCEVRWGGCGELES GSDSDCT TXVKKMECMELEYLRLYYSDKNGMYGV GILIGKRLKDKVMKVEYINGRIMMIRLKGKKE DLVIIQVYMPTSNHSEEEVEECYEKIE NLAERE KRNACIVVMGDWNAVVGEGSEGKALGK FGL GERNERGQRLLEFCNENLMVVGNTLFD NHKR RRYTWISRFDNKRHQIDYIMIQQR YRNCLKSA KSYPGADIYSDHILVVAEVLVKLKRTW KGQK REFINIEKLKDKGICEKLENAFSEIIMP GQEMEC IERRWDHYKNGLRQLRKRKWVKKA AKETL GVKERKRVKKEWITPEMISKMDDRRK WKTVN TEEGKRNYRRLNNELRRETRKARVKW LTEKC DKIENLZKEGKYDLMYKEAMD LVFREKNSNS GIKEIEAADGKMVNETQEV MRRWEEYIEWLY DGDTRPSEDELGLEENTVAEDDKG NAILTSE VEASVKEMKNGKAEGIDEIPVELL KSLQREGK KELIGLYNQIYMTGNWPKDFCESIL PIEKKKN SKKCEEHRTVSLISHTAKVLLRTL NRRLYGKL NAVMGEEQFGFRRGVGTRDAIGLL RVIGERY	cd0165 0	523	779

				TCTTTC CTATC GACCA CCCTC GGTCA ACTCT TGTTT ATTTC CGACC CCGAC GGTAT TAAGC TATGG AGTCT TCATTT TCCTCT CTCAG TGCAA GGTAA CAGGA AACCA CCACT GACAT	MEKKRKVYAIFIDLEKAFDRVRWDKLMAILR KHGVDWRDRKLITNLYLNQTARVRIGGVISEE IELGRGVRQGCCLSPTLFNIYLEDIIKESFNISR GVCIGGRRVECFIRFADDMVVMAESAKEMEQM LDDLNTKLEEYGMKINKKKTKSMVIGGRNTK CRLKIGGEEIEQVHSFKYLGSMIMDDMYCTAE IKKRIAMAKEGFKRKQRLLCGPLKKDLRKRLA KCYIWSVVL YGAETWTLRKEDERRLEALEMW IWRMEQVKWEDRIRNEEVLRRVGEERNMLE VIRKRKRNWIGHCLRRDCLLKEGIEGMVKGK RGRGKPRYQMLDNIQGNKYEEMKRLALDREK WKRINP			
--	--	--	--	--	--	--	--	--

				TATTTT					
				CTCTA					
				GACGC					
				TCATG					
				GGGTG					
				ACCTC					
				TTTGT					
				AAAAG					
				ATTTG					
				GAGGT					
				AAAGC					
				TTCCC					
				CTCAA					
				TCGGA					
				TCTCC					
				GGGAG					
				GGGAA					
				TACAA					
				CGAGG					
				TTTCC					
				AGACA					
				TAAGG					
				AGAGA					

				TGGAA CAGGA AGGAA GGAGA AGGAA AAAGA AGAAT TATTG AGGAT AGGAA CGTGG AATGT TTTGA CCCTTT TGCAG GAAGG GAAAT TAGAA AACGC TAAAC ATGAA					
Ving i	Ving i-	.	Locusta migratoria	GGGGA CACAG	TAAATTT TAAAATT	MRMSV NKRSFSEPALKMTSVNIEGLSANKELL LSDICSKCKCDVLILQETHRGPTSHRPNIKGMIL	cd0165 0	524	819

37_ LMi		AAGGA	TATTACA	TLERPHEQYGSAIFTKPGLMVRSAAMTHVDDI		
		GCCTC	TGGACTC	EILTVELKACTVTSLYKPPGADFTFREPSNFQT		
		CTAGA	AAACTAG	QPLKFITGDFNSHSTAWGYQETDSNGEKVER		
		AGATC	ATGTAAA	WAESSGLKL VHD PKLPASFNSGRWRRGYNPD		
		GAAAG	GGATGGT	NIFVTEQIAQQSFKSVEDALPRTQHRPISLTVK		
		ATCGG	GTTCTGG	AVVCPVTVPFKRRFNFKKARWRLFTEELDSKI		
		GCGTC	ACACGAA	DEILPAACSYDKFIQLIRRVSRKYIPRGCRNEYI		
		CCTGG	AAATAAA	AGLDDNIEILLKKEYEELYNTDPFSEDLETGED		
		AAACG	ATAAAT	LLRALTTTRREKWCDLVTSLDFKQNSRQAWS		
		GCACA		LLRNLSNDPTSSKTLHTSVTPNQIAHQLLLNGK		
		CAGCC		TTHRSRPSIVRDMENEHHSLEAPFTLEELQAA		
		GGMTA		IDKMKNNAAGEDDLRVEQIKHFGPKTLRWL		
		ATCCT		THLMKQLRQHHADSQSVEKSKSGSPIKTRKGA		
		TCTAA		NRCQKLSPSILAVPSVQGPRENGANNCVNTMQ		
		CACCA		IPRVWRKARVVALLKPGKEPTDVKNFRPVSL		
		GAACA		CHLFKALERMVLNRISNHVDEVLSPEQAGFRP		
		CCAG		GRSCTGQILNLTQHIEDGYERKQVTGVAFVDL		
				SAAYDTINHSRLVHKVYQATKDFGLSRFVQCI		
				LQNRIFYVTLQGKSSRWRIQKMVLLRAVYGL		
				WTFKICAVHTTEQAFLCHTARKKQQMENPKN		
				GLAQGSVLAPVLFNIYTNDQPMHPDVRQFIYA		
				DDTAVAAQGSTFEEVEEKLTRALELLGNYYH		
				QNHLKPNPAKTQVCSFHLRNKEAKTKLHVSW		

						KGQELEHCYAPKYLGVKLDRTLSFKQHCLDT KKKVSTRNSIIRKLTNTNWGADPQVLRTTAMA LSLSAAEYAAPVWKN SAHVKHIDVAVNESARI ISGCLRPTPTTEKIYPLIGIAPPSIRRQVAAEVERK KQLTDNRHPLHGHQQQPRRLRSRRSFLWTSEP IDVTVEERRLQLWKEATPNPILALKEQLAAGG NKSFTTWKALNRLRTGVSNCKFNLSRWGYLD GEPLCDCGNVQTDDHLLTCPNLSTCCTPLDLQ LANDKAICVADYWARNI			
RTE	RTE - 21_ LMi		Locusta migratoria	CCTGG TCAAC GGTCT CAAAG GCGGA AGAGG AATCT TAGTT CCCAA GGAGA AGAGC GGAAG AACTA ATGGA	TGAACAG CAACAGG TTAACCA AACAGAT TTTTAAA TATCTCT GGGACAA GAAGTCA ACAATAG CTGGATT CAAGAAG TTAGGAA AGATTTG GAAAGAA	MYNFNCRMESPLKVSTTPGGNSNEESTIALCN ASGPRVLGSPNICIKGESEHPWYPSKCLKFNRF FAGTLNVNSLLKIGKQKELEILLDKHEIQILAV QETRFLDENITETKNYRIFKGKPAIKIMKGMPI L GTAFCVHKNLVDKITEFKSSSERLSLLTIKCKN KQYTFVNCHAPINDDNKKNL SKVEAFWELLE DEISKIPKSNVVILLGDFNAQIGREKTRKIVGE YPAHKRTNKNNGTRLIGLCKTYXLKLMSTFFKK LPRKAKTWISPNPMLGEFQLDHVAISKHNIKEI MNVKVTRNREFDS DHYLT KIKMKLLPXKIQK RQORLIKYNTDRLTITKDAIKHFQEEIKISKSGK WEELSKEINRVAKNVFGATKIMKKAWWNETC EEAVAERTKKWKKWKCSGKTEHLDEFKIQRK ETARIKSTKNFYQMFKNHLKGYQAPSICFRDE	cd0165 0	561	821

				GAAAA	ACAATAT	NGKVGLSNXHNCEILRKYFEQLLNCEEPNCKL			
				CCATG	AAGTGAA	EFPDLHENTETDPPPTAEEIKKAINALKNNKAS			
				AAAAA	AAAGATG	GEDSITAELIKWSQPKIISNLQLMFEEIWETEKI			
				TAATC	CAACAGA	PEDWKVALIHPLHKKGDKQDVNNYRGISLLPV			
				ATTTC	AAGAGAA	GYKIFSTIILNKVEATLDSHLGEYQSGFRKGRS			
				GGACT	ACTTTTA	CSEQIFNLKSIIRHRLTNSKDIVVTFIDFKKAFDS			
				AACAA	AGAAGAA	VDRETLDKVIREFGIKSKLANLIRETLTDTKSK			
				TCCGA	AGTGTTA	VKLQGEISKAFDIKTGVRQGDGLSPLLFCNVLE			
				TCTTAT	AAAATGG	KIVREWKQKLKDHKLSPIRLGTKHKGIEIHCLA			
				CTTGG	AAGGATT	FADDFAILSENLTNAEIQINLLEEIANKAGLRIS			
				AATTT	CCAAGGC	VEKTKFMTNIKNAPEILVTDIGXIKRVNKFYKL			
				ACTTC	AGGAGGG	GEMIQENGLEKFAVEERVHKMERAYGITRNIY			
				CTCCC	AGAAGAA	NKRSLSKNLKIQHYNTVVKPECLYASECLVLN			
				TGTAC	GACAGGA	YKLHKLEVLERKIIRKILGPLRNTEVWKLRSNA			
				A	GTAAAGT	EVYRNIENTETLRKRRLFFGHL YE			
					GGTCTGA				
					GGAGAGA				
					AAAAGGA				
					TGCATAG				
					TGAAAAG				
					ATGAAAG				
					AATACTG				
					GAGGAAA				

					AAGAAAG AAGAACA AAGGAGG AAACATT GAAATTG GTGCGTG GTCCTTA GATGGCC GAAACGA AGAAAGA AGAAA				
Ving i	Ving i- 42_ LMi		Locusta migratoria	GCTAA TAAGC ATAAA CATTG AGCGC TTGGC CAACG ATAAG CAGGA GCTGC TCTCG GAATG	TGACATG GATTCCG ACACGAA AAAGTAA TGT	RNGQMLIILYLFMTQNC HHFTVKYGNKALIR ISASAVLIWRQXASRQF ITHSQRHNIDQLATSW PECLRRHKTADDTSRR RFNFRKANWTDFTSTL DNCIDELEPIPGNYER FVEKVRYSISRITIPR GCRQQYIPGLSGESKGL LDQYNEMYEKDPFSNE TEEVGNLLKELSRTRQ EKWVETIERLDMTXSS SKMAWNMMKKLSADPI APKQASNTTADQIAHQ LLISGKTGTKPRSRKI HRNLESEVSLIDTPFG IKDLDNSIQSLKTGKA AGVDDICTEQIKHFGK NKTVDPPTNKVLLNKQ IPKSWRKSKVIALLKPS KDPNLASSYRPISLLC HHSKFMSGNSELCGRK TNSRTRRFPPRKIMYR AGIGADRTHRERLRGE NITG	IPR000 477	299	551

				TGCAG AAATA CAAAT GTCAC ATTCT GTGCA TCAGA GACCC AGACA CCAAT TAACC TCGTC CAACA TACCA GGAAT GAAGC TTGTA ATTGA GGTAT CACAC AAAAA AATAT GGTAG	VVFVDLSAAYDTINQRTLLAKVYNITKDYELT KVIESLLQNRRFYVTLEGKKSRWRTQKNGLPK AVFCPQCCSTFTQTTNLLLVIQSTSFMLMTWPL GHSSSGQKLPAVEERLQNALSQMSVYYKENS KPNPAKTQACAFHLKSKEARRKLQIKWDGFSI EHTAQPKYLGVILDRSLTYKSHCQKTRQKTAT RNQLLKKLIGSKWGPASPHVIRTTAEAVCFSTA EYACPVWNRSAHAKEVNIALNETCRLMTGCL KATPLGQLYKAAGCEAPATRLAVEHTERFR QAFDDRHPLFEAPEPLRRLKSRRSFLRSVPMEP PDKYPLTQKPPPGYHLDWRSWKSLNRLRTGV APVKTNLVKWGIIDESDVRCNCGERQNMEHL LTCKNCPRTCSLDDLWNARDNALDVARYWA DVL			
--	--	--	--	---	--	--	--	--

				TACAA					
				TCTAT					
				GTAGA					
				ACCGG					
				ACCAA					
				ATTAT					
				AACAT					
				CAGTG					
				GGCTC					
				AACAA					
				ACATG					
				ACAAT					
				TTAGA					
				AATTC					
				TAACA					
				GTGGA					
				AACTG					
				ACAAT					
				GCACA					
				ATAAC					
				AAGTG					
				CGTAC					
				AAGCC					

				CCTAG					
				CCAAA					
				AATTC					
				ATATT					
				AATGA					
				ACACA					
				TAATT					
				TTAGA					
				ACAAA					
				ATAAC					
				AACTG					
				CTTCA					
				TCCTG					
				GGTGA					
				TTTTA					
				ACAGC					
				CATAG					
				CACTA					
				CATGG					
				GGATA					
				TGAAG					
				AACTA					
				ACACC					

				AATGG AAGTC TTGTA G					
Pene lope	Pene lope- 30_ LMi		Locusta migratoria	TCTTA AACGA AAATG GCTAA AAACT ACGGC CAACA TAAAT TTCAA CAAGG CTTGT GTAAA GAATA AATGT ATCCC GAACT ATATC AAGTT TAAGA	TAAATGA ACAAGTT CAAGCAA TACATAC AGCTTTT TTAAAAA CTTTGTG GACCTAT TATAACA AAAGTCG TTTATTTT AAATTTC AACCTCT GTCTTAC TGTGCAA ATGATTT CGTGACA CCTGAAT GTACCCA	MIWDTVKSELDEQLQVEFLNLKSRQDQKLKK LMRQIDCRVNNAQNVDNFQFYPRVVNETDISF NDEELRLLNKGLKFNVQPSLNDKTVKELIVAT KMAVDVAKLNENEKGHVGYKLANAIAETN KSCNSMHNNSDHKTVTSINDKLISNKAIVTKA DKGNTTVIVSEEQYINKTLQFFSENGVTEIAKD PTKKVQAEKQLLKRTSFLLEKEKFQCYVMN PTAPRLRSQKLVHKNHGHPIRAIVNSMKGPLCKI TKKLDKILKEFYVFGDNFSVSNNTTLNIEIKDK PIADTSHLMSLDIKNLYANVPVDEA VELIKANL LKHKRLSQQEIVEFIDLLQFILKYNYFAFNGKF YIQKDGLAMGSSISGTIADIFINNLENRLFTNM QHPVDKILYYKRYVDDTLLLIDGGQTEITEIVN KFNSLHEKIEFTVEQEKGDTLQFLDLNITKQDG RHLFKVYRKPTTTDILIHNSCHPNGQKLSALR CLVDRAFKVPMNEKDRTEFNIIRQLAKVNGY QENVVNNLIRKKQRAKSNKNGTXNIKRVMVP YFGNLSQKVANILKPFKVEAAFKTNNTLHHKL RNNLNQETDKYDKAGVYKIQCSSCNANYIGQ	cd0030 4	353	448

			TATGT	CTAGCTT	TGRSFKVRYKEHMDAFRLNNFDKSAVAMHIR			
			AATCC	GTGGATG	ETGHPLKRIEDDLKILHKQDKGRKLDLLEQLEI			
			AAAGT	ATCTTTG	TIHNAKATFH			
			CCAAT	GAATTTA				
			GTAGG	GAGGTCT				
			TAGGG	TTGTTTTTC				
			CAGTG	ATGTGCA				
			GTTGA	TACGTGC				
			TAAAG	GCCCCAC				
			CGCGG	TCGTGTT				
			AAATT	GTTGACT				
			TTGGT	CCGTGCA				
			TGAAT	TTTCGCT				
			TCCGA	AGTGTTA				
			AATTA	TGCGCCG				
			AATTC	CCAGTGC				
			CTCCA	TGGGGTC				
			TAAGA	TGTTTGA				
			AAAAA	ATCCTGA				
			GATTG	CGAAGCA				
			TATTA	CAGTATG				
			GTGGG	CTGAGGA				
			CAATT	CCTTTCT				

				ACTTA	GCGTAGA				
				AGGCG	CACGGTT				
				CATTT	ACAAGGG				
				AGAAG	AAGTAAC				
				CAGCT	CACGTTT				
				AGTGC	TCTTTTAT				
				GGTAC	TACGTTT				
				ACA	TAGATTT				
					TTATTGT				
					CTCAGGT				
					TTTAAGA				
					TTCTTGT				
					ATAGATT				
					ATCTGAT				
					GATGAGG				
					GCTTGAC				
					CCTTGAA				
					ACGCGTC				
					ATATTTT				
					AGTGTTT				
					TTACACG				
					ATCTCTG				
					TGGTTGA				

					GCCTACT TGACAAA ATAAT				
Nim b	Nim b- 19_ LMi		Locusta migratoria	CCCCT CAGGG GCTCA GCATT AATTT GCTGA GTACG GGCTT GGCGA ACCCG GGGTC CTGAG CCGTG GGCTG TGTCG CGCCA ACAGT CCGCT ACACC GTAAG	TAATGGG AAGCCTG ATATTTC AGGCACC CACCATT CAAACGG TTTTAAT CTTTTCGT CCTGTCA GTCCACT TTGACTG GCAGTTT AGTGTGT GCAGAAT GGCTGGC TCATCCT ATATTAT CTTTGCG ATCAGCC AGCCACA	MALPGAKPASLSVXGPNLMLRHMTPRRSLPW PRHGRNAGREIEVRLMPPXTWYVPGLTANRLP PRSPCFLWKTLRISSGRWMACQDEIWXCXHX DVVPSPTGTACLVRGGGWHVKMRSGXVLIT TSSPAQSRALLACDKLGDVPVSITPHKSLNMV QGIIFFHKDLLLQXDDELKKNLERRGVHFIRRVH RGPKDNRVATGAFILAFEGDSLPEKVKVMIYR CDVKPYVPPPMRCFKCWRFGHMSSRCXGNPT CRECGRAXHPDAPCAPPPVCVNCGQHHSPCSP XCTVLHKERKIMEYKTLDRLSYSEAKRKYXK LHPVSMKTSMPQLHHTASAIATALHCLVFHS GPPGXTAHJCPHGLWGHNLNLRPQGPYFGSSN HHSTCGXISPHPPPLHLWGHQSPPLSQRHNSXL WLLSRGRGPSDPFLLRLLLLKDTSRWLKEPPT AGRRASRAPQFLGTEPAEPSQHEKPKEQRETK EKDKKPPKRRIPWWPPCHHPPPLLLRTRETTQ EKDSL VATVPPPPSTSSSEDEIEILASPKDLAIAE SSEAMDVEALPQKSATAEDTVNPTAPSKSACP SDQEALGSASKIGDSRRYCEPYCAVEISMSQRP GGSLILQWNCGGFFRHLPELQQLLCRFPAFCV	cd0165 0	1172	1430

			CTCTG	TTCTTCTG	ALQETWFPAQRTPALRGYRGYYKNRTIYDRV			
			TGCGT	CTGTGAT	SGGVCIYVLATVPSAAVPLHTTLEAVAVRIRA			
			GCTTC	ATTTTAT	MQELTICSLYLPPDEMVS LTDLAALVAQLPPP			
			AACGA	GTTGTTG	LLLGDFNAHHPLWGGAQTTGRGRDIETLLSQQ			
			CCATC	TCTGTTT	DICLLNTGAPTHFCVARGTYS AIDLSVCTPSLL			
			GTTTCG	GGTCAAC	PHFQWRVHDDLCNSDHFVLM SLLDFQPAGC			
			GCGCG	TACTTTG	PSRWSLRRADWQRYTSAATVPVPLPADVDMA			
			ACGGT	TGTTTTCC	VHQVTSTIVSAAEATIPTTSGSPRRKSVPWWTP			
			GGAAC	TTTAAAT	EIAEAIRD RRRALQRHRRHPTVENLIAFKRLRA			
			GTTGT	GTTGTCT	RARQLIRQSKRDCWETYVASIGCRASRLGCGP			
			ACGTT	CTTAGCA	RLGAFLGTVHQQRSPAYTSRVTYAPTRLPSLN			
			TCAGA	CGCATT A	TWQSIMLAPLQHQNTVQTSVHTGLLGDVCRQ			
			GCCCG	TCACTCT	HRLSRLASRVWSQVRRLFGYRPPAAIPSLHIQG			
			GGGAT	GACCGTC	NLCTDPTSIAEHMAEHYARTSATSEYSADFRA			
			CTTGG	TCTGCAT	LKERVEQRRLSFHTRRAESYN S AFSEWEFETA			
			CTTAA	TAGACGT	LAACPDTSPPGP DRIHSQFIRHLSVEFQHQLLAL			
			TCGCC	GGAGTAC	FNRLWAEGAFPSQWRESIIIPVIKPGADPQVAD			
			CGGGT	TTTTTACC	SYRPISLTNVLCKL FERMVSRRLMWILEHRGL			
			CGCGA	GCTAATT	LSPHQGGFRQGRSTVDNLVALES AIRTAFSRRE			
			GGACG	AAAAGAT	HLVAVFFDLRKAYDTTWRHHILATLHGWGLR			
			GTACA	GTGCTCT	GPLPLFIQNFLSFRSFRVRVGTSLSSPRVQENG			
			AACCT	CTTGGCT	VPQGSVLSVPLFLIAINGLA EAVGSSVSSSLFAD			
			CTATA	AGGAACA	DFCLYFSSASIGVAERRLQGAIRKAQSWASTH			

			AAAAA	AGGGACT	GFQFSPSKTRVMHFCRRRAVHLDPELYLYDEL			
			ACCTC	GATAACC	LRVEETHRFLGMLFDSRLTWVPHLRQLKQKC			
			AATCT	GTTGCTG	VRYNILRCLSHTSWGAARTSLLQLYRALIQSR			
			CAAGG	TTTAGCC	LDYGSFVYGSAAPSALQVLDPV LHSGVRLATG			
			TGTGC	CCCTTTC	AFRTSPVASLLAEAGVLPLRFRRQQLLSSYAA			
			TACGT	AAAAACC	HVHATPHHPNRRLLFPAAALPSRQRPRSGLP			
			GCCGA	CAACAAC	AVRVRSLSSLES LPIPPTFHCPYRTAPWSLPRP			
			GGAGA	AACAACA	QLRLDL SHASKASVAPLSLRRQFQAVLESFCG			
			TGCAT	AC	AEVVYTDGSVVDGRTGYAYVHGEHIEQHSLP			
			GGCTG		CGSSVFTAELVAIYRALEHLRACPGESLVICSD			
			TTGAG		SLSGLQALAKCFPKHPLVQSIQESVSRLVRSGR			
			GTTGA		SVVFVWTPGHIGIPGNERVDQLAKQATTSSPL			
			ACAGT		AIGLPESDLRLEFHRKILGLWNSEWANSAPN			
			CGCTA		KLRAVKLSTVAWKSSLRSSRRESVVLCLRLRIG			
			GCGAG		HTWLTHGHLLRREDAPQCDCGAPLTVVHILV			
			CAACC		DCPLLVALRQSFNLPDALSFMLGDDAAMVER			
			TCTGA		VLDFIRAAGFYHVI			
			GGAAC					
			CTGCC					
			GCACC					
			TCAGT					
			TGTAT					
			AAGGC					

				TTACC CAGGC AAGCG GGGCT CTGCC TGGGT GGACC TTTAG TTCCCT AGCTG CTCGT GGGAC GAGAA TGAAA CGAGC AAATA					
Pene lope	Pene lope- 1_L Mi		Locusta migratoria	TCTCCT GGCGA ATAAC AACTT CGAAA AACTA CGGGT	TGATGTT TATTATC GATATTC TGACCGG TTTCATCT TTGGTCG TGTCGGC	MDPDIVILPADKGNATVVLNKQDYLQKMQCL LSDSAYRRISADPTKSVERKTNNLLKKSSLSQE TIKSLNSYCAVPPRLYGLPKVHKDGCPLRPIVS NIGAPTYRVAKHLAGLLSPLVGRCEHHIKNSV DFLRRLEGLHLADSDILVSFDVVSFLFTRVPLSD SLQLIETKFGTDLTNLFRHVLSTYFLFNDEYY EQTDGVAMGSPLSPIIANLFMEDFEDRALESAA	cd0030 4	201	285

			ATGCT	GCTGTGT	LKPACFFRYVDDTFVWPHGSENLYFLEHLN			
			GCCAG	GCTTGCT	SIHPNIRFTMEVEKNGCLPFLDVLVRRKADGTL			
			GTATC	TGCTCGG	GHAVYRKPTHDLQADSCHHPAQREGVLR			
			AAAAT	TGTGTGT	TLVHRAHVISDPESLPTELSHLEVTFRQNGYGE			
			CGACG	TGTCTTTC	GQIKRALRYRPHVQRASDDSTESAPTSVAFLP			
			TCAAA	CTGTTTA	YVGSTSNKIGRILRKNNVKCVFRPPSKIKALLG			
			TTTTTC	CTCATTG	SVKDDLGLRKAGVYRIPCSCGMSYVGQTIRTV			
			GAAGC	AAAACCG	EDRCIEHKRHLRLQQPNKSAVAEHCLDTGHP			
			TGTTA	AGGTTTT	MEYNKTEILACTSNYWDCVIKEAIEIKLGDNL			
			TTCGC	AAAGATC	MNRDGGFCLNSAWDPALSLVKKQRDKTNPTS			
			CAGGA	GTTGCAC	SVEN			
			GAAAT	ATTGCTT				
			TAAGG	CCTCACT				
			AATCC	GCAGTTT				
			ATTAC	CGCCTTG				
			TCTAC	AAAATGG				
			GGGGA	CAGGGTG				
			GGAAA	GTCTCCT				
			TGTAT	GCCGAAA				
			CGTTTT	TATCGGC				
			TGCGG	CGTCGTC				
			AAGTC	GATTTTG				
			AGATA	GTACCTG				

				AGCTT	GCAGCAT				
				CGTAG	ACCCGTA				
				CAGCA	AGTTTTT				
				GGGTC	CGAAGTT				
				AGGTT	GTT				
				GCAAT					
				GTGCA					
				TGTCG					
				TCTTTG					
				TTGGA					
				ATGTC					
				GTGAT					
				AATAG					
				TGTTG					
				TTCCA					
				ACTTTT					
				GCCAA					
				AGTTG					
				TGCAC					
				ATATC					
				AATTC					
				TCCTG					
				CCGCT					

				GGTCG					
				CATTA					
				AACAA					
				CGTGC					
				AGGCT					
				TAGCT					
				CTTGT					
				CCGCG					
				AGAGG					
				ATTCA					
				TTACA					
				CCCGC					
				CGAGA					
				ATTGG					
				ATTTT					
				GTGTC					
				CAAGG					
				AATTA					
				TTTCG					
				GCTTC					
				ATTTG					
				TACAT					
				TGCCT					

				CTGAA					
				ATTTC					
				GGAAT					
				TTTCTT					
				GGGAC					
				TGGTT					
				GGACG					
				GTGCT					
				TCCTG					
				GTCAC					
				AATCC					
				GATTG					
				GGAAT					
				TTCAA					
				ACTGT					
				AACGG					
				CCCGG					
				CATTC					
				GGCGA					
				AGTTT					
				GAACG					
				GTTCC					
				GTAGC					

				TCAGG					
				GTCTG					
				CTTTTA					
				CAACT					
				CGACG					
				TTCTGT					
				GATTA					
				ATCTC					
				ACAGA					
				GAAAA					
				CCTTT					
				GACGA					
				CGCAG					
				CTCTA					
				TCCGT					
				TCTGG					
				GGAAA					
				GGTTT					
				GAATT					
				TTGCA					
				CCGAC					
				GCCTA					
				AACGT					

				TTACC					
				TGTTG					
				TTGAT					
				TTTATT					
				AGTTC					
				TATCG					
				AACAG					
				GCTGT					
				TTTA					
				AACTA					
				CCTTCT					
				GACAT					
				TAAAG					
				AGGAG					
				GTTAG					
				GAGGG					
				AAGCT					
				TGTCG					
				TGTGT					
				TGACT					
				AGGGC					
				ACGTC					
				CGCCT					

				GTGAG TAATA TAACA GCAGC TGAGG CAGCT GCCTT ACGTT CACTT AGG					
Dap hne	Dap hne- 26B _LM i		Locusta migratoria	TGCAT GAACA GTGTA ATACT TTAAG AGCAA TATTTT TGGAC CTAAA TTCATT TACAG TAAGT GTCTA	TAGAAAT GTAGGTC TATGATA AAGTGTG TAGAGAA AATGTAA ATTTACT ATTAGTT ATTACGA TGTATTT ATTTATA TAAAAAA TGTAATC	MLFHQNVQSLNNKLSEIKIMLSNELIDVSVLC MSEHWLRDDALQTVSILNYELAAAYTCRQQILG GGICIFVKKGLKKSFFFFLFFFFFFEISAIELPSL RDNFNMHLQIPNGDIKTFYQNLNTLTQTVKN KQTHICGDFNIDFKGSAKEKTDLMNILVSHTLK ITIHSPTREIVTSLQVYNRSNNNKPTKEQYKA EAVKTGLGDHHGQTISFHIKGIEVNKLTSTYKT RTYNDQEIEIFKNLLKSESWEQIYDTSTENISEK FDIFIEKIKYYFDVVPVKTRRLKDTLPKKIPSQ NKWITTGIRISCANKRFLYRLTKCDTTPAFDR YFKNYKTVLRKVIKSAKKLENDRIYISNSSNKM KAVWNIVKKETGNNKTKNKNIQLKENQQTTS KPKEVANMFNTYFINIAEKLIEKNVIKNPSQLA	cd0165 0	503	794

			GCACA	ATGCTGT	TKIKHIDKTIFLTPTNTKEVQQEIKLLKSKHSSG			
			AGATG	GTAAAAG	IDNIPDFIVKKCADFLSEPLTYLINLSFSTGSFPK			
			GACTG	TGGCAAA	CLKIAKVKPLHKKGSQEEVSNYRPVSQLSAFS			
			CATCT	TATCCAT	KJLEKLFIKDCPISQNPNMVLEKNHSTETAVFS			
			AAATA	TTTATTC	FLYETLQALDRREHTTSIFLDLSKAFDTINHNK			
			GGCTA	AATATGT	LLSKIQNWGLGAAHXXIKSNPHDRKQFQKAV			
			GGCAC	AGAGTAA	LIEITYESNGTVSTYHSDSLVVKHGVPPQGSILG			
			AAATA	CAAAAAT	PILFLLYVNDLYINNTNGKLYQFADDTSILITGS			
			CTTGT	TCTAGCA	TKEIFESNIAAITKNIRDYFDANGLIINVEKPLQ			
			AGCAA	GTGAAAA	WIYLRHRIDHQLTQILICKWPDNKCCKTVAMD			
			AGTAT	ACATTAT	IFTTQNRSPANPNIDINGKTIEKTECTKFLGIWI			
			TGGAG	GTGGTAA	QQNIKWDSDYINRKLSTCYIMRTLKNSIS			
			ATGTA	ATCGTAC	YNNLKSIFAYAHSQMKYGIVFWGNSGASMK			
			TGTGT	CAAGTAA	TFRIQKKIIRIMEGVDQRTSCKPLLKRNTPTSKH			
			GTGAT	AATGTTA	VHNGKPNYYKRRIKKEILPLPSMYIMENLIIKE			
			AAAAA	ATCTACT	KLNSKSPCITRNETVHCHNTRQKNNVHVKQT			
			CAAGG	TTTAATA	NTALFQKGVLPGLRYNALPKAIKSKTDTTKF			
			GAAAC	TTACAAC	KTAVKDYLVMHCFYSVDEYLEFQNYNIS			
			TAATG	ACAAGAA				
			TATAG	AAGAAAT				
			GCAGG	ATGTTAG				
			GGGTG	GCCAGAA				
			AAAAT	ATGGTCC				

				CCAGA	AATATCT				
				CAAGA	ATTGTAC				
				ATACA	CCTGTAC				
				AAAAG	AGTTCTG				
				ACTAC	GATTCAC				
				AACCT	AGGACCA				
				CTGTG	ATAAATA				
				AACTA	AATAAAT				
				GAGCA	AAATAA				
				TGTAC					
				TAGAG					
				ATATT					
				TAAGG					
				AAAAC					
				AAAGT					
				AAGTG					
				CCAAC					
				CAAGT					
				AATCA					
				GAAAA					
				AAGTG					
				TATTA					

				TTATTT T					
Pene lope	Pene lope- 8_L Mi		Locusta migratoria	TCACA TCAGA AACCT CTACG GGGAG GAGAT GGTTG AATCG TGGA AGACT GGATA AACTT CGCAA GAAGA AGGGT AAGCT GCTGA GCGCT CTTGC TTTTTT ATTGA	TGAGGAT GGCGGGC AGTTGCG CCGCTGA AATATTG TGTC AAG TTGACTT TAGTATC CGGCTGC AAACCCG AGAAGAT TACA	MSDWAHRS AVDKQSSKFQRLSRPSISHNTNRR TVINFTDKELDDAMLSVLEKGLNFAPTPKNVP LTEFICSVEQAVSKLPGDQAE EIRQEVAHTLRR APPPRSNISPLERAAIRSLREDQDLIILPADKGN ATVLLSRDDYLRKMDCLLDDSA YRRLQKDPT ERIKRMTLSLLKKSSLSDVLKKLHPGAAVPP RLYGLPKIHKQEVPLRP IVSNLGAPTYNLAKHL ASVLSPHVGKCEHHIPNSMDFIQRLKSLRLSPS DLLVSFDVVSLFTRVPLEESLQLIGEKVDKETA ELCRHVL TSTYFLFNDKFYEQTDGVAMGSPLS PIVANLFMEDFEDMALKTAALKPTCFWRYVD DTFMVWPHGEEALNRLLDHFNCMHPNIRFTM EVEKEGKL PFLDVLVYRRDDGKLGH SVYRKP THTDRYLHASSCHPPVQRTGVLRTL TRRAYAI SDAESLTQELDHLKVVFKQNGYTDKQIRCAFQ FGPSPEPLNETHKSVAFLPFAGSISSKIGRILKK HDIKSVFRPPVKTRALLGSVKDDLGLRKPGVY KIPCQCGKTYIGQTSVPFRTDVLNINATHVCSS PRSRDLSL MWENVYWSDIRTIQDRCVEHQRH TRLLQPEKSAVAEHCLDEGHRMLYNETHMVA SASRYWDCVLKESIEIRLS DN LINRDKGYLLSK	cd0030 4	318	402

				GATGT	SGTLCYQMLKNNGWYFRNLNNNFLIIDYRFRR			
				CGTAG	ILPPERCMSCAPEGSYCHFLRTRYGSSAMYKG			
				TGAAA	AADCTGNSVPARLCTSTLT			
				GTATC				
				ATACC				
				GAAGT				
				TCGCC				
				AAGTT				
				TGTGC				
				ATTAC				
				ATTGA				
				TCGAA				
				AGTGG				
				CCAAT				
				AACAT				
				CAAAA				
				AACGT				
				GCTAG				
				TCTTG				
				CCCTA				
				ATAAG				
				GGAGC				
				GCATT				

				CGTTTT					
				ACTAG					
				AAGAC					
				AATTG					
				GATCT					
				CGTAT					
				CCAAG					
				GACTT					
				ATTTT					
				ATTTA					
				CATCT					
				AGAGG					
				TATCTT					
				CACGG					
				CTGTC					
				GGCTG					
				CATCT					
				TGGGA					
				ATGGG					
				TTGAT					
				AGCTC					
				TTCGT					
				GGGCT					

RTE	RTE - 62_ LMi	Locusta migratoria	TTGCTT	TAATTCA	MKRPFREKFPPSNPIGKEKATSDSVGSQLEDSN	cd0165 0	596	831
			GTCTC	TAATATT	ESEHLKSPNYHQARNAKIKHKHINYIATHNINS			
			CTACC	TATGTTG	LLQTGKLKNFTDELNRQNIHAGLQEMRNTIED			
			GAGCG	GCCATAA	PFEXQGYRIYNGKPGQRMKQCPQFGTGFIVN			
			CCTCC	AAGTAAT	LKVIDSITDFKAISPRIATLTLKATNKIYTIINAH			
			CCAGG	AAATAAA	APTNQKNFLTRTMVEVEKFWDLLDQTINQVN			
			TGGCG	AAAAATA	KNHVKILIGDFNAQLGKERKYRDIIGKWAPHK			
			GATAG	AATAACC	HTNRNGQRLVEICRTYGLISKSTYFKRKPSKL			
			GGGGA	TCTGTAT	MTWKHPDWTKEWQLDHVCMDKFFHKEIQN			
			TGCTC	TATTAAA	VKVLRGMDTGSDHYMVNIKIKFTPLSKKRQR			
			ACCAG	TTGTCAT	ALEIKESMIHINKKTESIGDKRKYPHKLKND			
			ATATG	GTTTATTT	KYQKNTQNTVITDNLQELIPILKQHAEESAPVN			
			GTGGG	TTGAACT	QTSKKHAWWTTECDEIHDERHQAUIKFQTHR			
			TACCG	AAATTGT	TEENAYNLKNIRKKFSQIIRKTKRLYQKDVLS			
			GGGAA	ATTAATG	IEENYHETNSRDYFKAFDKQLRKFNPPPTLLRN			
			ATAAA	CATAACA	KDGRIAHNNKENAEIMAEAFNKLLNLNLKSF			
			ATACC	ACAAATT	KAFNNRHKHPCNNQTRKYSPTVQEIETALKE			
			CGGGG	GTAATAA	MRNYKACGENQVFAEMWKYAGTTTTRTSLHTI			
			TGGAC	CCTCTGT	LQKIWTTEKFPEEWTTAITPLHKKGDKSDPDN			
			CAAAA	ATTATTA	YRGISLLDCTYKIFSKILHNRIREQLDKELGEYQ			
			CTAGC	AATTGTC	GGFRPWRSACAEQIITLKLIMAYYRKRNKPLSIT			
			ACAAG	ATGTTAT	FIDFKKAYDSIHRPSLLKILRHFGLHLKLIRLTE			
			GGCAC	TATTTTG	LTLTNTNSKIKFEGNFLDHSIHMREWYKANPK			

				GACGG	AACTAAA	NIKVGTQKDNIKLNCLGFADDLALLANNIQET			
				ATCCA	TTTGTTA	KQQIRSLQEIAQCIGLKISFEKTEIMVTEPPLAN			
				TATGT	ACAACAA	KIRIGVHNIKIVEKFKYLGEIITHNLNEKPSWQS			
				CAGTG	ATTGTAA	RINKLIKMHVTKNTYNKKCLSIATKLKHYKT			
				GCGGT	TAACCTT	VTQPQITYASETIFKTTNTVAIDRLLKIERRIVR			
				ACCCC	TGTAATA	TCINKKYQKDGWRIASNETVYKEVEPVTSTM			
				AACAG	TTGACTA	KKKRISFLGHLMRTPENRISRKIIEKLWNSKSNI			
				CGTCT	AAGTGGT	RWITEIKEDMMKLQITMEDLKDKTDTLKILQD			
				GTTCC	CCAATGA	KHVTLPINKQRLGRVISDEERRARSERMKK			
				ACATG	AGGCCAT	YWANKKKKNKNKNRRRT			
				TCAGT	AAAATAA				
				GGCGG	ATAAATA				
				CTGAT					
				CAGTA					
				TCCTCT					
				GTGGC					
				TAACT					
				ACCAT					
				AGTTG					
				TAACA					
				CGGAG					
				CTCGC					
				TCCAC					

				TCAAG GTTGA TAGCC TCTGA CAGTC AATC					
Dap hne	Dap hne- 13_ LMi		Locusta migratoria	CAAGA TGGCG TCAAA AGTTC AGTGC AGTGA CACTT CGATC TCCTC AAAGA AGCAG TTATA CAACT CTTCT GCTTG TAAAA AGTGC	TGATTTT ATAACTA CATTGTA TGTATAT ATTATTA TTTTAAA CTGACGA GTCTCCT GTACACT AAATCAA TGATTTA ACTTGTA TGTTATG AGACAAT AAAATTC TGATTCT GA	MIKKNVWFQTVKSMCVQEKKRVVPDCEIDVR PGIRTNQLNNHFKNIQQAkhNSKNVTNSENyK GVFIHVGtNSLRSCSEEDLVNETRNVIRSAKNI YPYSRIIISSILRRRSVSDKYIDKINCNIrDQCNR LGAIFVDSNKFIRDDCIGQDGLHLNRLGSATFS KMFIDLCKIIRNRGNWQSVRGVNPVQNYTKQ GKLAVSEGGEKLVVDKDTTKEDVLKTVNRY NSSFLMHWNIHGLRATYLNKIAKIDELQIILS EGRNIKVVCLNEHWITEDSITVLNKLNFNVAS SFCRKNKTHGGSCILINSSLEYKERDDFNFLNE ELIFESCCVELEQQNIVIIISIYRIPGNDVTKLFFD KFQCLLQKLTkEMRKKIIIAADFNIDTASETSIS TKFNDLIQQYGfRLNFTEFTRVSDKSATCIDNI LTNYTLNENAKFTLNlGISDHYALLMELLKTN KENGRKCFLKRQFNPDNIHLfCCRLNDIDWSL QHESSCSENFDRFLNSFLHVFNETfPPMSAKER QQVSNISTHVCQRKTTskLKWITTGIKISSKRK	cd0165 0	767	1046

			GATAA	RELHAELKSNKSPEFIKYVKSYKKLFRKVVKA			
			ACGCG	AKEMANTKFIRKHKNKSKAVWTVVKSELGIK			
			TAATA	PSNQEISKLKIQDKIIVNPAQISETLMNFSTKNT			
			AAAGG	RQNNCKSCSDIRNFNEFFITVAKSDVNVPSYQD			
			TATTCT	NVHLFGSSLSKNECLIKFDTVTVEDVEKTILNL			
			GTGCG	KNKNSVGWDEIPSKVIKAASNIIAYPLCKIINQS			
			TCACA	LEEGYFPDSLKYAEVKPLFKKGSKEDMGNYRP			
			TGCAA	ISSFPCFPKYLKYQKQFGFQKGKSTIYAINEFV			
			ATTTT	NKVTSTLDKSQKAAGIFCDLSKAFDSVNHSL			
			GGATG	LYKMEKYGIMGTALQWCTSYLTERKQRVVIT			
			CATGA	SSKGKYSSKWRLSHRGYHKAQFVETIPQGVPO			
			AAGAT	GSILGPILFLFYINDLPININYHSVLFADDTSVLI			
			GTGCT	ENKKADEIPDNVLKTLENLDVWFHLNGLKLN			
			AAAGT	KKLRCLKDPRKPRCLVPSKWSKIKYKKTQVM			
			GAACC	QFKTKQSKLNDIQIKHNSQELHEADHVKFLGM			
			TTAAG	KLDRIYPGPHTILGNETGQNLSWSSHIDYLNK			
			TTCGT	LNSLAFAMSILANVTSLDTRKIVYRSYFESIVR			
			AAATG	YGIIFWGNASTMIRILTLQKKIETCGAKPRESC			
			ACGAG	RPLLRLNLNLSVPSLYLFEIILSIIQNYFFIITSST			
			TTKCC	VMKQEQRQFHAPTHRLNLYAQAPQYMGMKI			
			ATGGT	CNKLKNTHFRKVETESMKRKLSTLVQKCY			
			CGTGT	SIDEFLNDDLSIEQDKTYNYVL			
			GGCGA				

				ATGTT					
				TATGT					
				GATGA					
				ACGGA					
				ATGAA					
				CTGTT					
				AAACA					
				CAATA					
				AACTC					
				GAAAA					
				ATGAA					
				ATCAT					
				CAAGA					
				TACTA					
				AATAA					
				TGACG					
				TTGAA					
				GAACT					
				TAAAA					
				AGGTA					
				ATAGT					
				TGCGT					
				TGAAA					

				GAAGA					
				AAATG					
				CGAAA					
				CTTGC					
				AAGTG					
				ATGGA					
				AATAC					
				ACAAA					
				AGTGC					
				GAAAA					
				GAATG					
				TATCA					
				CTGTG					
				TGGAA					
				AAAAC					
				CAAGA					
				AAATC					
				CAAGT					
				GTTTCT					
				AATTC					
				TGGAA					
				ATGAT					
				GCAAC					

				AGTAA				
				ACAAA				
				TCGGG				
				TAGGC				
				CTAGC				
				GTGGA				
				AAAGT				
				GACAA				
				ACAAT				
				GCCAC				
				TTATG				
				CGGAT				
				ACAAC				
				AAACC				
				AATAC				
				AAATG				
				GACTG				
				TCGTA				
				CAGCA				
				CAAAA				
				AGAGG				
				GACAA				
				GCAGG				

				AAAAG AAAGA AAAAG AGAAC ACTAA AAACG ACTTC TTAAT AATCG GTGAT TCC					
Ving i	Ving i- 7_L Mi		Locusta migratoria	GGGGT ACATG AAAGG AACCT CCAMG AAGAT TGCAA TAATA GAAGA ACAAG AAATG AACAG	TGACCGA CCGGACA CGGAAAT GAATGA	MLRSALTADPFTTGLHSNGQHSDRSFSRPFVFI SINVEGFSREKEVLVANLCQKYSCDIXLVQET HGGSTKNRPRIAGMELILERPHDKYGSAIFARP DVNITSASLTEDNNIELLTVQTPFFCVTSVYKPP VPILASKHQPTSTIRTRQQHRTPHSANSVFLCYI SLQTTCANFSIKTPANFNDQDVNFVIGDFNCRS TSWGYPDSDTNGTSLETWAEAHDLHLIHPSS HLHLTVAGGKEGTTQTTFCCFKGSCTMYEDCR RAHPSYATQWQVEKRVQPRQHFVSSKVAAQC TKTVGEPIPRTQHRPVICTVNPVVVP EEPFKR RFNFKKANWQKFSEELDSKVINIPPQPEMYETF IKCVQQVSRKTIPRGCRTQYVPGMSSDDKLLL	cd0165 0	561	816

				CAATC	ETYQTLYEEDPFAKETIQAGEILLETISATRKAK			
				GGGTG	WSNLVENLDMXLNSRQAWKLLKNLSGDPVK			
				TACCT	SQDNFTSMTPDKVATQLLQNGKIKGRTPKEAL			
				GGGCA	QRPETEHHFLRDPFSMAELEAAISSLKMHKA			
				ACGTC	PGVDDLRLVEQIKNFXCKTMEWLLSLMNKCVA			
				AGAAG	KLHIPKMWRKSRVIAILKPGKEPTDPKNFRPIS			
					LLCHTFKILERLVNLRISATLEXKFIPQQAGFRA			
					GKSCSSQILNLTQHIEDGFERGEISGVAFVDLT			
					AAAYDTVNRKLLRKVYDLTKDYQLTSVISTFL			
					QNRRFQVSLQGKRSRWRTXKNGLPQGSVLAP			
					ALYNIYTNDQPIPANTRLFYADDTAVAVQGPT			
					FKEVEKKLSLAMDELSKYYDANHLKPNPSKT			
					QVSAFHRLRNREAEHKLDITWRGEQLKHCPTPV			
					YLGVTLDRALTYRQQCLNVKQKVSARNGILK			
					KLVGSSWGANPHLLRTSAFALCISAAEYAAPV			
					WDASAHAKQVDIAVNNTARIISGCLKPTPVQK			
					LYPIVGIAPPNIRRAIAADYERTKQETDHRHPL			
					YGHQAARHRLKSRKSFLARTAPLEGAPEANRL			
					ATWKSSLPADVSPKEEIAPGANLQYPVWRS LN			
					RLRVGVPRCKSNLKKWGILPASADTSCDCGA			
					DEDASHLLVCPLLDNTCTAQDVHEGNDKAIA			
					AAIFWKC			

Ving i	Ving i- 5_L Mi	Locusta migratoria	GGGGG ACTTG AAAGG AGCCT CCCAG CAGAA TGTGC TGTTA AGCTA GCATC ACATT CGGGC GTCCC CTGGG CAACG GTTTTT TCTTG CCGGC CAATC CTTTC AAATC CAGAG ACATC	TAGAAAC CTTTCTTT ACATCTG CATAATT TTTCTTTT TAAATTA ATATGTA GGTATTT TATGATG TCTCTGA CACGAAT AAGAA	MTSQHRTFSGPIMTIISVNIEGLSLCKEELIKDL CKKNKCDVLCIQETHRDSILNRPKIQGMRLVA ERPHRQYGS AVFKSEIQVLSTSVSDINDVEIV TVELKNCTITSVYKPPLSNFNVNATHKHNSTK VNFVIGDFNSHSTRWGY PQDDENGMAVISWA ESNKLALIHDKLPPSFNSGRWQRGYNPDLIFA TESVSQQCIKTVCPPIPKTQHRPIMCQVMTAVR PVCVPFRRRYNFRKANWLKYQEILDREVAFLP PNPDSYDKFTDAVKRSSRLSIPRGCKTEYVQG LTPETVADLHTYTELFKDPFDESTVAAGQQL LASISEAKRNSWVNLMEELNMSINSHKAWRLL KRLSNDPTSLNNHANITADNIATQLLKNGKPE NRSRRQTKNMSNTTEDETNSLSQPISIDELNEAI AKCKNGKAAGLDDILVEQIKHFGPQTKQWML NMFNKCMA DCKIPSIWRKSKVAILKPGKDIN DPKSYRPISLLCHMYKIFERILQTRLTYVIEPKLI PQQGGFRPERNCTSQILGLTQFIEDGFQKKRVT GAVFIDLTAAYDTVNTRQLIEKLYEITRDRKFT DIIRNLLQNRFFVEFQGKRSRWRNL TNGLPQ GSVLAPSLFNVYTNDQPLPVNTSSYIYADDRVI AAQQGSFEDVEKSLNEALVELDSYYCRNQLKP NPMKTQTCAFHLRNKQASRQLQVRWRGVDL EHCKNPKYLGVTLDRALTFKQHCLNTKLKVC	cd0165 0	471	726
-----------	-------------------------	-----------------------	--	--	---	-------------	-----	-----

				AAGGC ATAGC TAGTG TA		GRNNILRRLTGSTWGARPVVRTSALSLAYAA GEYASPVWCERSAHAKQVDVALNESARIVTGC LQPTPVNKIYHLAGIAPPAIRREQAAYKERQKS IVSKWHPLHGMPEAAQRLKSRNSFLTSTEKES DYRTTVERWKDAERHPWMEPSEALAAGHNES WTTWKSLNRLRSGIARSKVNLSKWGYQNDST LCDCGAIQTTTHMYNCPLSPSHCTFEELMSAQ PNAIDVALFWAKNV			
Dap hne	Dap hne- 37_ LMi		Locusta migratoria	TGGCG GCCTG TATAG ATGTA CTAAT AAGTG GCTCT GTGAA AAACG ATTAT GTTGT AAATA TAAAG TGTAT TTTGT	TAATAAT AAATTAC AAATAAA AAGTGTT GTTTATTT TATAACT ATGTACA CAGTATG GTACTAT AACTTTG TTACTGT CTTGAAC ACATTGT AATACAT CTGTATG	MYIFSDSHGRGIASIMQTMQMFKVTGITKPG APLCIITNSKLLALPNLVPLYVLLHKTDNLKQIS HVSNMWQSNHQGGFQNVMFVEATSSVRTISP NMVCTGIIIMAKQFCARKYDNINSQRQEIHSPIP LSLVKEKEVTSHTQNEEKEQETSLSSPKKESPL LTLKRKDNWKRVSQKEKKNHHPVLRKRTRNI TVFPQKRITIAHLEKKGQLEKKRAEREEESSSC SIEEVLSLSAPEEETFSSAEKLPKLRRGERTKK DIPLKNQDFFIPAAGEKTTRYKLDLRKTKVVPR DQLHKFQGISCTNSIGQNVMLHINVQSLKYKI NELQFWLENLNC SILSVNEHWLNCEELKLYVP FGFDLASSYYRTNVEHGGVSMMLKIMTWWSVY VKNNVNLIYEVTDVSEL CIEAKFEAAAIFPKQ EIIVGSVYRPPKHNESLFVENLEKLILLLLKHKN YRIVIFGDINIDIRFKNRNVNHLLNMLRSLNFY	cd0165 0	818	1089

			GCAGT	TTTATTGT	CINTKPTRQAACLDNIICNFHLNKTDFTISLCI			
			GAAAA	CTGATAT	SDHDGLWLRQTLNYKILETSTEHRIRPINDSKL			
			TGTTCT	CTATTTA	NQLISKLNIVKWDEIITNVDANEATNRFLLLLA			
			CAGTG	TGGTAGT	ELFETCCPKHYRKS VKQRQCSPQNLKKTGIL			
			GCATT	GTACATA	HITTKSKKNWYTPHLNRIRILMLLYHDKAKHN			
			AGGCT	TCTGATA	NADKEMYMKIKKLYRSEIKAAKCKANDAYIL			
			GTGTT	ACTATAA	NSTNKCKAAWEIIKTEMNYD HKVYQIPISPNTL			
			CAGTG	TGTAATT	NEHFVGPHSDSSTMQDVRKAEALLPETSSNLT			
			ACTCA	GTTTATT	NNFCWSRVTEKEVYESVAKLSNSRAEDYYGL			
			AAAGT	ATTCTTG	SNFVIKSVRKQIVKPLTKIINKILEGGHYPECLK			
			GGGTA	ACATTCA	ISVVKPVYKKGEVTSPENYRPISLIPISKIIEYC			
			CATAC	TTTCATA	MHKQMYQYFETNNILSASQYGFRSRLSTTKAV			
			GAAGT	TTATTTTC	ESIVSKIHSSFEKKEIVCATLLDLRKA FDMVSH			
			GCATA	TTCTATCT	NILIKKTPALWSERECYMPDGVLLGEQETVSQ			
			GCATA	TATGGGA	DKKKLQHYGVRGNAICLMESYLENRKQLVKI			
			TTTTG	TTTTCTA	NNESSECLPVLKGVPQGSVLGPFLFIVYINDLP			
			ACACT	AGTCATT	TVVPCDAVLYADDTTLITCDTNLE NVQHNMA			
			ATAAA	TATTGTT	VAMEKCNTWFEANVLQKNDSKTEEIVFNLT A			
			CACAA	GCTATTT	PSQDNKTVKLLGFHIDQKLNWDTHLDKLC SRL			
			AATGT	TGTTTTAT	ARVLHLLYKLRSSVSKELLYAYS AFFHSHLD			
			AATTG	CATTATT	YGII LWGNSVGSKTVFKWQKKAIRCIAGLS PR			
			TGGCG	ACTATTT	DSCRGYKELNIMTVPSLYIYNCLVYVKENLN			
			TAATT	TATCTAA	NYDLRMTIHDHSTRSGHTIDVPYARLAKTHNC			

				CATTG	CCTTGAC	YKYLGLLYFNKLPKEAYTVPVDQFKNTIVKW			
				TGATT	GACACCG	FKSKVFYSAQEFLECGTKDLNFL			
				TGCAA	ATTGCAT				
				TTGGT	GTAAAAA				
				AGCTT	TGCTCAA				
				TAGCA	AGGTGAA				
				AAATA	TAAAGCA				
				TGATG	TCTAATC				
				CTGAG	TAATCTT				
				CGGGA	A				
				AGAAA					
				ATGTG					
				CTATTT					
				ACGAT					
				AGAAC					
				ATGAC					
				ATATT					
				GATCG					
				TTAAA					
				AAGTA					
				CTCAA					
				ACAAC					
				TTGGA					

				AACAT					
				TGGCA					
				ATAAC					
				TCTTA					
				ACGAA					
				GAACA					
				AAACA					
				CGAAA					
				GAGCA					
				TCCAT					
				TTATT					
				AGGCC					
				TAAAT					
				GGAGC					
				TGGTG					
				TTCGA					
				AACAG					
				AACGA					
				GAAAG					
				AAAAA					
				CATCT					
				TGGAA					
				AAACA					

				AATGC					
				TTTAG					
				CAGTA					
				GAAGC					
				AGTTG					
				CGCCT					
				AAACA					
				AGCAA					
				AAACA					
				AGAAG					
				TATTTC					
				GAAGA					
				AACAA					
				GAAAA					
				AGCAA					
				AGCTC					
				CTACA					
				AAATG					
				AATCC					
				ATCAA					
				CACAG					
				AAAAA					
				GAGAA					

				GGAAG AG					
RTE	RTE - 26_ LMi		Locusta migratoria	GTGGA GTTGC TAGTC TCCCA TCGGG CACCT CCCCA GGTGG CGGAT AGGGG AATGC TCACC AGATA TGGTG GGTAC CGGGG AAATA AAATA CCCGG GGTGG ACCAA	TAATCTA ATATAGG TGCTGGA AAAATGT ACTGACA TTGTAAA ACCAGAA TAATAAA TAA	MGQLPIHWKKTLPVKKPIRLQQDGYNRRRRR LRKRTVLIGVWNVQGLQNKVGEVIKEMEARN QDIVVLPETKKKGHGIERIGPYTHFYSGVPKNQ RAKRGISILMKNKFNRMVTNWEPIDENLIKINI NIQEHKITLIGVYGISEDEPIRKKDEFYNRLNGV IEDIGKTREIIIAGDFNSRTGSKIGNKIIGPFGEN VVNDNGNRLIDICEQNSLKILNGFYQHKDIHK YTWHKDTLNQKSIIDYIIKQQTRIKIHDIRAYR GVTLSDHYLVNSKFLFPFRGKNSEENQNCTTE DLQTEIKITKYNLNSFENESTVFLYQKRLDEKL IETEFDNTEQYQYIISKLHESAKEALGETDEK HTXHSLYYWDEEIKKEVKLKKEKYLKWLNTR NVQDKIELNAQQAQAKIRKKVAEAKNKAWEQSC QKIESYIGGKRNTTEAWKTIKNLRRNSKEXTHIK YITHDSWEKYFQNLLSENRRILGKPEGILKXQH ILNILLMTLGKNIFKIYYQKTEEFLGNPEYNEIT NELPDNYIHLEIETVKSIAKSLKNGRATGIGGIP AELLKFGTEKLIELLRNLFERCLNGEDVPNDLK VGYISVIHKKGRKDDCKNYRGITVTNTFSRLY GRIIKYFLEQEYKEKEAEEQAGFRAGRSTIDHIF CLQQIIEKNKKLGRAQPIHLVFVDIEKAYDSVP	cd0165 0	592	851

				AACTA GCACC TACTG CCTTG TAGTT AGGTA TTGGA TTACC AAAGG CTAAG GGAAG AAAAC CTTGA TCAAA AATTA CCTGG CCCTC CAGGT TGGGG GTTAA TGCA		LPNLWKALLSIGINPTIHKAIQNL YKNSISKIKIG KFLSSGFQVTKGLRQGCSMSPTLYKIYTEVTL QNWKKKCQTMGIPINDRTIYSLQFADDQLIAQ DYEDIEYMTRKLIQEYKKAGLNVNMNKT KYM VIGGINEDLILEEGMGTITTTTEEYKYLGVKITN DGKQEKEIRSRINAGKTTISLLNGILWDKHITT DNKIRIYKTIVRSIITYGSEVWTTKWQLKSKLL ATEMDFWRRSARISRRERIRNEVIREKMKCKN SIIEFIEHKHLKWYGHVRRMGQERLPKCILDW TPNGRRKKGRPPTTWMQEVQSAMRKNNIPED LWEDREEWRTIIXNLM			
RTE	RTE -	.	Locusta migratoria	GCCAT AGCCG	TAATTAC TCAAGTT	MSTQFLKPVNKLTTWKSPNSTLGEFQIDHVAIS NKNTKEILNVRTRKGFFESDHLLQIKIKFQPN	SSF56 672	393	550

38_ LMi		GTCCC	CAATCGC	RKKPQMNKTFRIDPEYLKLNKKEISQEIRQSNA		
		AAGCC	TCTCTTTA	TNWKELAEVLKSTMHLGKPPRKRKHRWWNS		
		CGGTT	AAGGGAA	SCDQAIEERTQAWKTFSSNKTTEKWQEFLRIQ		
		GAGAA	TAATCGA	KNTSKAIRREKRRYDINRLKEIEQEFLKNNTRN		
		AGGAG	AATAATA	FYTIFKENLTGYQPPNLCFRRPDGSLETNTRNN		
		GAGGG	ATAATAA	CQILANIFTHSSTASRRQSNISKTPSNPDSKPPSL		
		GGAGG	TA	KEVKEIINSLNNKSPGEDGIIAEMWKLNDSTLT		
		AGTAA		QATHRIISKVWETEQIPAEWKCALIHPLHKKV		
		CACCT		QKLIQITTGEFHCSQLQGNFTAPSNIQNFFKGP		
		CGTTA		VKQTRRTGGPVNRRVSGRVSRILRRTNMEPK		
		AAAAA		DILQMRRCRNTVVTFVDFKKAYDSIDRQTLFH		
		CCTTG		TLEEFRIDRKTRSIIQQALTGTTSKVKFMGETSE		
		GTTCA		AFHIHTGVRQGDGLSPLLFNLVLEKIVRTWEK		
		CCTGG		EIKGIQLGIKKEQRIRVKCLAFADDLAILTNNR		
		CCCTG		EEARLALLEKLHEISQKTGLQISYEKTQYMENK		
		GTGAT		LQIVSPSQPSMEKWHKFLTSDIKDSVHGKQAA		
		AGCAC		DSLPIITTQYGKVAQVSHFRYLGEMIQPSGLNSI		
		CTCGT		ANQERSRLQRAYKLTWNHYNKKSISVNAKL		
		GAGGG		RHYNTVVLPEALYAAETTVIHGHTKIREIEKQE		
		CCCCT		RKILRKIYGSVFREGIWIKRPTREIYKDMETITD		
		GCCAG		TIRKRRMKFYGHISRMNETRLTRQIFEIANKSK		
		GGTAC		NKTKWITETEEDIKELGITQDNINNREQFRNIK		
		GGTGA				

			AGACC	KHKLKQEHTENRTGKKWSERKREHSERMKRI			
			TCAAC	WEERKKKKKS			
			GGCAG				
			TGAAG				
			GCGGA				
			GATGA				
			AGATC				
			ATCGG				
			TACGG				
			CGGAA				
			CTGGC				
			GGAAG				
			AGGCA				
			GCCCT				
			AATCC				
			ACTAG				
			CGTCT				
			GCCCT				
			GGAAT				
			CCAGC				
			GGTTA				
			GTGGT				
			CAGCG				

				TCTGTT					
				CACCT					
				AGTTG					
				GTGCG					
				GCTGC					
				TCCTA					
				GTTCTT					
				CTTGA					
				TCTCA					
				ACAAT					
				CAAGT					
				CGTAA					
				ACATG					
				TGACC					
				TTCGT					
				AGAAG					
				ATCAC					
				CAAAA					
				CAAGT					
				TTCAA					
				CTTGA					
				ATAAG					
				CATGA					

				TGGAA					
				CAAAA					
				GGAAA					
				GAATT					
				CCACT					
				ACCCC					
				AGGCA					
				TCAGT					
				CTCTA					
				GACTG					
				AATTC					
				TCCTG					
				GTCAT					
				CGGAT					
				CTGGG					
				GGACC					
				AGGAA					
				CTTCA					
				TGCGC					
				AGAAG					
				AGTCG					
				GAGCT					
				TCTCA					

				AAATC					
				TCTAA					
				GACCA					
				AAACA					
				GAAAA					
				ACTTC					
				ATAGG					
				AACTC					
				TCAAC					
				ACAAA					
				CACAC					
				TCTCA					
				AAAGT					
				AGGTA					
				AACTT					
				AAACA					
				ACTCA					
				CAAAC					
				ACTGT					
				GGAAA					
				AATTC					
				CACAT					
				AAAAA					

				TCCTT					
				GCGAT					
				TCAAG					
				AAACA					
				CGATT					
				TACCG					
				ACGAA					
				AATCA					
				TTTTGT					
				ACAGG					
				AAACT					
				TCAGA					
				ATCTA					
				CAAAG					
				GAAAA					
				CCTGM					
				AATCC					
				AGTTG					
				AACAA					
				AGGAC					
				CCAAA					
				CAATT					
				TGGCA					

				CAGCT					
				TTTGTT					
				GTGCA					
				TAAAT					
				CAATT					
				ACAGA					
				CTCAA					
				TAATA					
				AACTT					
				CCATT					
				CAACA					
				TCAGA					
				AAGAA					
				TATCC					
				ACAAT					
				CTCAT					
				TAAAA					
				TCTGG					
				AAACA					
				AGGCA					
				TACAC					
				ACTGA					
				TAAAT					

				GCACA					
				CGCTC					
				CCACA					
				AATGA					
				TTACA					
				ATAGA					
				AAAGA					
				CCCAC					
				AGGTA					
				ATAGA					
				GGACT					
				TCTGG					
				GAACT					
				ACTGG					
				AGGAA					
				ACCTC					
				AAATA					
				AAATT					
				CCTGC					
				ACATC					
				ATGTG					
				AAAAT					
				TCTTTT					

				AGGAG					
				ATTTC					
				AATGG					
				ACAAC					
				TCGGC					
				AAAGA					
				GAAAA					
				AATTC					
				AGGAA					
				AACAA					
				CGGGA					
				ATTCA					
				TACAG					
				CTCAC					
				AAGAG					
				GACAA					
				ATAAA					
				AACGG					
				CGAGA					
				GACTG					
				ATAAA					
				CTTCT					
				GTGAA					

				AATTT TGATC TTAAA ATC					
Kiri	Kiri- 1_L Mi		Locusta migratoria	CTGGT TCACT TTAGG ATAGT GCTGT TTTACT ATTAA TCTCA TACTT ACATT ATTAA TTACG GTACC ATATT AGGTT TAAAT TCAAT TTTAT ATAGC	TAGTATT AATCCTA TTAGAGT TAAGTCA ATACCTT TTATTTA CATCCTT TTACATG TAATCAA CCTGTAT AAATTAT GTTATTA ACTGTTT ATCTGTA AAAATCA TTGTTAT GTTATAG GTAACTC ACATGTA	MCPAFAEKDAATPQPAVALPTSGIARATTER TKLHVAHANVQSLQAHLEFRHIFQDLGLHIIL LSETWLKPSITTSSVNLDGYVFLRHDRLNKR GGVGAYIRSELSPTVLHTSNNDETEQVEYMFIE IKPLKEKCLICVLYKPPKVGALDCFETALS DLESTYEHIIIAGDTNINFLFNPSMKFKRMLTSSN LTLLQLAPTHHTADSHTLIDIFATRFDPDRVIRTS QTSAPGMSAHDVIFMTYSLECPKPKPLTTYR DLRRVNLPELTADAXEIVFEPTSDTLPDVNEKL HKITNKLNQLYDKYAPIKTKKVRKPPAPWLTD ELKQLMNLRTAHRTYKLHPTPENHTAYKQK RNRVSQAVRNAKLRYAHS�TDNINRPAVLWK NLRGLGIGKVKTAAEPTVPIEELNQYFTTPTTTI DHQTKQTLIDHITRINNANHDKFCFKHVTGST VKKAIMRITSASTGDDGITIQMIKLIIDPLLPIIT DIFNQSLATSVFPDAWKRLVKPLPKKDIATA PSDYRPICILPALS KALEYIVHDQLTSYLTINNL LDEHQSGFRKYRSTTSALIKVTDDLKQAMDRQ EATIMCFLDFSKAFDTVDFDILLAKLSSLNLSQ	cd0165 0	510	773

			GTTGG	AAGAAAT	GVVQWFRSYLTSRLQCVMSGTEKSQWRKVLS			
			TCTCA	ACTGTAT	GVPQGSVLGPLLFSLYVNDVSSVLSFCKYHMY			
			CGCAT	GTGTAAA	ADDLQLYLSAKPKNLKTAIENFNSDLCALSKW			
			CATTG	ACCCTGG	AQNIGLKLNPSTQAILVGHPRLINHKEYRESQP			
			GCAAT	TCCGATG	SLILNDTPIYFSPSAKSLGVIIDDKLNWAEHTAA			
			TGCTC	TAAGAGA	VCRKASASLHALQRYKKFFPLDLKKKLIQTLVI			
			GCTCT	GGGCCTG	PIIDYSDVILQGLSWESSRRLELIMNACVRYICN			
			ACTAT	ATGGCCC	IRLFDHISPSYKQLSWLRADKRRDFHTLCFLYS			
			CGATA	TAATCTG	LINVHCPSYLSSTLTLLSDQHNRTTRSCQSKTL			
			TTGTT	ATCAGGC	SVPIHHSAGFSKSFVAGTRLWNNLPQYIRELN			
			ATCTTT	TAAATAA	NTASFQRQLMDYLLKQLRQSLIPYGPGTSHPFL			
			ATTTA	ACAAATA	LPDILLPLLSPGTAFLTSFCQNSPCLNQSRYIQ			
			GTGTT	AATAAAT	KYILYLLISMLLLLLLHQQQQKVSQWYC			
			GTTCC	AAATAAA				
			GTGCA					
			GATTA					
			CTTTTA					
			CTGTG					
			CTTGG					
			CGAGA					
			TTCTCC					
			CCCCC					
			CCCTC					

				AGCGC					
				CCCGC					
				CCTCG					
				GCGAA					
				CGGCC					
				GGCGA					
				CCTTG					
				GCTGC					
				AGCGC					
				CTCCC					
				TGTCC					
				CTCCC					
				CTGTC					
				CTGAA					
				CCCAC					
				CTGAC					
				CTGCT					
				TCGCT					
				CTCTCT					
				CTCTG					
				CTGTC					
				CTGCA					
				CTCCT					

				CCCCC GACCC CCCGG ACCGC TAGCC ACCCA GCCAC GCCAC ATTCC A					
CR1	CR1 - 1_L Mi		Locusta migratoria	TAGTT ATTCT CGCAT ATTTG ATAAA TACTG TTTGA TAGCC TTGTCT GTGGC TGGTT TTCAA AACAA	TAGTGGT ACTCAAG GTACCCT CCGCCAC ACACCGT CGGGTGG CTTGCGG AGTATGG TGTAGAT GTAGAGA TG	MVIGCFYRPPGSAAVVAEHLKENLENISSRFPD HVIVLGGDFNLPDIDWETQTFMTGGRDKESSEI VLSALSENYLEQLNREPTRGDNILDLLVTNRPE LFETVNAEQGISDHKAVTASIISAVNRNIKKCR KIFLFGKSDKKRISEYLTAQHRSFISSTDSVEEQ WSKFKAIVQYALDEYVPSKIVRDGKEPPWYN NRIRKLLRKQRELHCKHKHSQSLVDKQKLNE AKCSVRRAMREAFNEFESKVLCTDLAENPKKF WSYVKA VGGSKQNXQTLCDENGIETEDDTLK AELLNVFFQSCFTEEDCTVVP SLDCRTDYKMA DIEIIDRGIEKQLKSLKRGKAAGPDGIPVRFYTE YAKELAPLLAVVYRRSLEERSVPKDWKMAQV IPVFKKGRRTDVQNYRPI SLTSVSCRILEHVLCS	cd0165 0	388	636

				GTCAG	SIMTFLETRNLLCRNQHGFRKRRSCETQLALFV			
				AGACC	HETQRAIDTGSQVDAVFLDFRKAFTDTPHRRL			
				TTACT	MNKVRAYGLSDQLCDWIEEFLDNRMQHVILN			
				ACAGT	GEKSSEVRVMSGVPQGSVVGPLLFTIYINDLA			
				CGTTA	DNIGSSLRLFADDAVVYREVVTMENCIAEQED			
				TCAGT	LHRIDAWCREWQLNLNVDKCNVLRHRKKDP			
				CAGTC	LSFSYTIAGQQLEAVNSINYLGVGIRSDLKWN			
				CAGTG	DHIKSIVGKADARLRFGRILRKCSPKTKEVGY			
				CCAGC	STLVRPQLEYCSPVWDPYQVGLIEEVEKIQRR			
				TGTTG	AARFVTGSFGNRESVTEMIDKLQWCTLQERRS			
				CCAGT	VARYGLLSKFREHTFTESSSILLPPTYISRRDH			
				GTCAC	EDKIREIRAHTTEAYRQSFFPRTIRDWNRREN			
				ATCGT				
				AAACA				
				GCGCT				
				AAGTA				
				TCTAA				
				TTTAA				
				TTTGT				
				GTCAA				
				ATTTA				
				ATTTG				
				CCAGT				

				CTAGT					
				TCTTC					
				ATTTTC					
				GTTAG					
				TACTT					
				CTGTTT					
				TAAGA					
				GGTTT					
				ATTCT					
				CGCGT					
				ACTTG					
				GCAGA					
				TTTCGT					
				TTGTG					
				CTACT					
				GTTTT					
				AATAG					
				TTTATT					
				CTCGC					
				GTATT					
				TGATA					
				AGTAC					
				TGTTG					

				GGCGC					
				GTCAG					
				CAGCA					
				CACAG					
				TCAGC					
				AGCAC					
				CAGCA					
				GAGTT					
				ACAGC					
				TAAGC					
				AGGAG					
				CGAAC					
				TATTG					
				TGACT					
				TATTT					
				GTGGG					
				CTGAG					
				CAGTG					
				TCAAT					
				TTCTAT					
				AAACT					
				CGCGC					
				TAGTA					

				ACAAG					
				ATACT					
				TAACT					
				CTCTA					
				CAGGC					
				AATTG					
				TTGAT					
				TTGTT					
				AGGAT					
				GGATA					
				GGATG					
				TGTGA					
				CTGCT					
				GCGTA					
				CGGAC					
				GCAGG					
				AGGAG					
				CTGGC					
				CGCTG					
				TTCGC					
				GAGCA					
				GCTGA					
				GCGTG					

				CTGTT					
				GGCCG					
				CGGTC					
				AGCCG					
				TCTTC					
				AGGCT					
				GCTGC					
				CTCGG					
				AGTGT					
				TGCGG					
				CAGTG					
				GGGGC					
				TCTGG					
				TGCGT					
				CGCAT					
				GGGAC					
				ACCAC					
				AGGTG					
				TTACA					
				TGCTT					
				CACCC					
				ACGGT					
				CCCTG					

				CTGTC					
				GGGAC					
				ATCTT					
				CGCGC					
				GTGCC					
				GGACG					
				CGGTT					
				GAGCC					
				ACCCT					
				CTCCT					
				CAAGG					
				GGTGT					
				GGCGG					
				GTTCA					
				ACGGC					
				GTTCG					
				CGTCG					
				CTCGG					
				GGCGA					
				AAGGT					
				CAATG					
				TGGAG					
				GCTGG					

				CCGTG					
				CGGCG					
				TCGCC					
				CGCTC					
				TGCCT					
				GTGAG					
				TGGAC					
				ATGTG					
				GCCGC					
				TCCTTC					
				AGCAG					
				GGTCC					
				GAGCA					
				GGCAC					
				ACGGG					
				GGGAG					
				GGGTW					
				TATTA					
				GTGAT					
				TGGGA					
				GCTCC					
				AATGT					
				TAGGC					

				GGGTG					
				ATGGA					
				GCCCC					
				TTAGG					
				GAAAT					
				AGCGG					
				AAAGG					
				TCGGG					
				GAAGA					
				AGGCC					
				AGTGT					
				GCACT					
				CTGTTT					
				GCCGG					
				GGGGT					
				CTCAT					
				CCGAG					
				ATGTG					
				GAGGA					
				GGCCC					
				TTCCG					
				GCGGC					
				GACAG					

				AGAGC					
				ACTGG					
				GTGCA					
				CCCGA					
				CTGCA					
				AATTG					
				TTGCT					
				CATGT					
				CGGCA					
				CCAAT					
				GACTC					
				CTGCC					
				GGCTG					
				GGTTC					
				AGAGG					
				TCTCCT					
				CAGTT					
				CATAC					
				AGGCG					
				GTTGG					
				CGGAG					
				TTGGT					
				GAAGG					

				CTGAA					
				AGCGT					
				CGCTC					
				GCGGG					
				GTGGA					
				ATCTG					
				AGCTA					
				ACTAT					
				TTGTA					
				GTGTC					
				GTTCC					
				CAGAA					
				CCGAT					
				CGCGG					
				TCCTCT					
				GGTTT					
				GGAGC					
				CGAGT					
				GGAAG					
				GCTTA					
				AACCA					
				GAGGC					
				TCAGA					

				CGATT					
				CTGCG					
				GAGAT					
				CTAGG					
				CTGCA					
				AATTT					
				CTCGA					
				CCTCC					
				GCTAT					
				CGGGT					
				GGAGA					
				AATGT					
				AGGGT					
				TCCCC					
				TGAAT					
				AGGTC					
				AGGCG					
				TGCAC					
				TACAC					
				GCAGG					
				AAGCG					
				GCTAC					
				AAGGG					

				TAGCG					
				GAGTA					
				CGTGT					
				GGAGT					
				GCACA					
				TGGGG					
				GTTTTT					
				TAGGT					
				TAGAG					
				AATTC					
				CCTCC					
				CTAGG					
				CCCGA					
				CAAGA					
				CGCCT					
				CCTGA					
				GCCGC					
				GACAA					
				GGTAG					
				CAGTA					
				GGCAG					
				AACAC					
				AACAG					

				GGAAT					
				GAAAA					
				GATCA					
				ATGTA					
				GTAGT					
				AGTTA					
				ACTGC					
				AGGAG					
				CGTCT					
				ATAGA					
				AAGGT					
				CCCAG					
				AACTG					
				CTCTC					
				ATTAA					
				TAAAC					
				GGTCA					
				CAATG					
				CCCAC					
				ATAGT					
				ACTAG					
				GGACG					
				GAAAG					

				TTGGC					
				TGAAA					
				CCAGA					
				TGTAA					
				ACAGT					
				AATGA					
				GATTC					
				TAAAC					
				TCAGA					
				CTGGA					
				ATGTW					
				TACCG					
				CAGAG					
				ATAGG					
				CTGSA					
				CAGTG					
				AAGGG					
				GGAGG					
				CGTGT					
				TTATA					
				GCGAT					
				AAAAA					
				GTGCA					

				CTAGT ATCGA AGGAA ATAGA CGGAG ATCCG AAGTG TGAAA TAATT TGGGT GAAGG TCACG GTAA AGCAG GCTCG AAC					
Dap hne	Dap hne- 46_ LMi		Locusta migratoria	AGTTT TTCTG GMTGG ATCAA ATGAT GTAGC GASAA	TAATTCT GTATATA TGCTGTA TATTTCT ATACATT TGTA ACT TGACTTT	MVSLSNYDWSHIFNNNAHDSANTFFDRFFHTF LNIFETNCPQYKCKVNPKSNSNKH RDKNPWY TAQLANMKSRLLLYRDSYRLQKTDMAKQRYS RAQKDYKHAVNEAKKQHNLVSISSKNKCKK AWTIINSVSKSKQKV DVVISADEFN NF C I KSVN QISENIVRP HETVADLIKCHNVNYTLKDKFLLS EVTPDVVLTIVNNFRPSDSVDIYGISCNMLKKII	cd0165 0	252	527

			ATGAA	GTCAATT	GYIVYPLTKCINKCLIEGVFPDVLKLSRVVPIY			
			GGAAA	GCTGTAA	KKGDRNCPSSYRPISLTPILSKFVLESIIYYQLC			
			AGACT	CAGCTAA	DYLTCCNNIISAVQFGFRKGKSTVHAIDSLIKKV			
			TGTTG	ACAATAA	LNAYEEKKFAQATFCDLSKAFDCVEHISLLKK			
			TTATC	AATTTCA	LVYYGIINSEVDNIFESFLSNRKQIVCIGKQRSQ			
			ATTAA	TTTATTTA	LAEVKCGVPQGSVLGPLLFILMINDLPSYNTH			
			GGAAG	TTTA	SFSMPMIPLESISAQTLICFKLFTILYKYSLLIYA			
			AAGCT		DDTTFFNISSDFDMLQTVTNDTLSQASVWFRA			
			GCAGG		NGFLLNESKTQKIVFSLRDLPVEDFIDSVKFLGI			
			AACTA		VIDSKLSWEPHIKYIXTRLRSRVVYLLRNLKNHV			
			CAAAA		PAAYVRSAYFAFFQSIISYGLLIWGNSSHVQDI			
			TACAA		LVLQKKVIRIITNSDKLAHCKPLFINLKILTVIN			
			AAATC		LYIYTALLHIKSNLSEYQRRGDVHSHGTRNSSH			
			CTCAT		LDIPYYRLSKSLNSYEVVGMKMFNKLPQEWV			
			CTTTTC		EAPLDFLKHKLWCWLAANPCYSLNEFYECKIS			
			AACAC		WYLLGECTIISLMESFMMQCFKLYYFKYLILW			
			CATTT		KTYSDLLIFALCKVICIYALPYVK			
			CGGCA					
			TGATC					
			TTAAA					
			TCCTG					
			GTCAT					
			GTGTG					

				AATGT					
				CTAAG					
				TAACG					
				AGAGT					
				AAATG					
				AAGAA					
				AATTT					
				GTAAA					
				CACGT					
				TAAAT					
				TGTTG					
				CTTTTG					
				TAAAC					
				ATAAT					
				TAACT					
				TAGGT					
				GAAAT					
				ATTTC					
				ATACA					
				GSTCA					
				TKGCT					
				TCCAT					
				CTAAT					

				CACTG					
				GAAAG					
				AAGAT					
				GTGAG					
				TAATG					
				AATCA					
				TAAAT					
				CTAGC					
				CATGA					
				GTTAT					
				TATAC					
				ACAAA					
				TCGAT					
				AGGTC					
				AGATC					
				AATAC					
				CACTT					
				AGATT					
				TAAGG					
				GAAAT					
				TCTTTT					
				TTAGA					
				GGTAG					

				AAAAA					
				GGGCT					
				GAGCC					
				AACTG					
				GTTTG					
				CCCAG					
				ATTCA					
				CACTG					
				GAAGA					
				CCAGA					
				TTAAA					
				GATAA					
				GAATA					
				ATCTA					
				ATAAA					
				TGTTT					
				GTAAT					
				ATTAG					
				AGAGA					
				AATAT					
				AGCAG					
				TAATG					
				ATGAT					

				TGCCC					
				TCATA					
				ACAAC					
				AATTA					
				TTCAT					
				CAACT					
				TTGTC					
				ATTTA					
				AGATA					
				GGCTA					
				TCTCA					
				ATATT					
				GAAGG					
				TATGA					
				GTGGA					
				AAACA					
				CGCAA					
				TTTTA					
				AACGA					
				TTTTGC					
				AAATG					
				AGAAC					
				TTATTT					

				GATGC					
				ATTAT					
				GCTTA					
				AGTGA					
				GTATG					
				GGTGT					
				AAAGG					
				AAATG					
				AAATA					
				TCTTTT					
				CATGT					
				ATTAG					
				ATGAT					
				TTCCA					
				TTTAG					
				TTAAC					
				AGTTT					
				TAGCA					
				GAGAG					
				CAGCA					
				TATTC					
				CTGGG					
				GGGTG					

				TGTCA					
				ATTTA					
				CATTA					
				AGAAA					
				GCATT					
				AATAA					
				ATTGT					
				AGTAG					
				TGAAC					
				TAGAT					
				CTAGG					
				GTTCTT					
				GTGCA					
				GTGAG					
				ATACA					
				TTTTG					
				AAGCC					
				ACAGG					
				TGTAT					
				GTATG					
				GATMA					
				TTTAA					
				AATTG					

				TATTG					
				TAGCT					
				CTTTAT					
				AGATC					
				ACCTA					
				GTGGT					
				GATCC					
				ACTAA					
				TATTTT					
				TAAGA					
				AAAAC					
				TTGAG					
				AATTT					
				ATTCA					
				ATTTCT					
				TTCTC					
				AGACC					
				TCAGT					
				GGAAG					
				AAATA					
				TAGTT					
				AGTGA					
				TTGGT					

				GGTGA					
				CATCA					
				AGCTG					
				TATTT					
				GATGT					
				AAATA					
				AAGAT					
				GCACA					
				CACTG					
				TAAAT					
				ATTTA					
				AAACC					
				GTTAC					
				GACAA					
				TTCAA					
				CTTCA					
				ATTTT					
				ACAAA					
				CTGCA					
				ACAGT					
				CTGCT					
				TGTCTT					
				GACAA					

				CTACT					
				TACAA					
				ATGTA					
				ATAGA					
				GAGTT					
				ACTGT					
				CCTCA					
				GATGT					
				AGCTG					
				TTTTTG					
				ATTTCT					
				CGGAC					
				CATGA					
				CCCAG					
				TTTGG					
				GTAAA					
				AATAA					
				GTACC					
				GATAG					
				GCCTA					
				ATATG					
				CAGTA					
				TAAGG					

				AGTTC AAAGA GCCTA AATAG TCATC ACAAG ACCAA TAACA CAGGA AAAAC TTTCCT AATTC					
Nim b	Nim b- 41_ LMi		Locusta migratoria	CATGG GCTCA GCATT AATTG CTGAG TATGG GCTTG GCGAA CCCGG GGTCC TGAGC	TGACTGG CTCATCC ATTATTA TCCTAGT GATCAGC CAGCCAC ATCTTCT GCTGTGA TTTTTAG CACAGTT GACTATT	MAPTVQPSSVGPKGHQTHSASLKVGGPSFIAPP VPWWGGNTSPPPSGTQVPIPQPEKQQPSPAPLA REEKSVGMLPSPVSKSRLEDPSRSLGDAHTTG RRASRSSSAPETSEKPSQWSKPKEQREDKQKK KRSKKKSLPTALPPSVSQTSDSEDGLEMLASPE DLALTDLPESMDIDATSTKAAAKGDTEAGINS MFMTLMA SQYANSNILQWNCGGYFHRLAELR HLLCQFPVLCVALQETWFPVQRIPALRGYRGY YKNRAAYDRVSGGVCIYVLANVPSAPVPLHT ALEAVAVRIRAGQEITICSLYLPPDKAVSSAEL AALVAQLPPLYLLLGFNAHNPLWGGTEITAR	cd0165 0	849	1106

			TGTGG	TGGTCCA	GRDVEYLLAQLNICLLNNGAATHFSLAHGTYS			
			GCTGT	CTACTTTT	AIDLSLAVQAYFHTLLGGSMTTFVEATISLSCR			
			GTCGC	GTGGTTT	SHSHHLQAACPDGHSVGQTGRLTPRLLQQRSR			
			GCCAG	CCCTTTTC	HTLQLGPWHLLSHRSVPCSPSLLPHLTWRVHD			
			CAGTC	CTTTTAT	DLCGSDHFPLLSLPQSPAGCLPRWSLRRANW			
			TGCTA	ATGTCAT	PAYTSAATVTSPLPDDVDLAVDDVTRAIVSAA			
			CACCG	TTAGCAC	EATVPKSSGLLHRKSVPWWTPEIAEAIRERRR			
			TAAGC	ACACTGC	ALQRHRRHPSLEHLIAFKRLRAQARRLIQQSKR			
			TCTGT	CTTCCTG	DSWERYAATIGSRASPTZVWKQVRRLFQHRPL			
			GCGTG	TACTTCT	AGIPSLHIRGTLSTDPAIAQYMADIMLVARHL			
			CTTCA	CGTGTAC	HLTVRTSAPSKSVLNSGSFPFTCAARKNTMRH			
			ACGAC	TGGCTGT	SVSGSLRVPAPFRASSTGRNPQSSHSGYSVHRP			
			CATCG	GGAGTGC	SCYRAVHGRHYARSSASASYSADFRALKERVE			
			TTCGG	TTCGTCTT	QRQLSFHMRRQEKYNAPFSEWEFESALAACPD			
			CGCGA	CAACTGT	TSPGPDSVHSQFVRHLPAXFQQLLLALYNHLW			
			CGGTG	AAATCAC	VKGAFPSQWRDSIIIPVMKPGADPQVAESYRPI			
			GAATG	ATGTGTG	SLTNVLCKIFEHMSRRLMWILEDRLAPQQ			
			TTGTA	ATCTCTTT	GGFRHGRSAIDNLVSLESAIRTAFTTRREHLVAV			
			CGTTT	GCTGGGA	FFDLRKAYDTTWRHHILATLYDWGLRGPLPTF			
			CAGAG	ACAGGGG	IKNFLSLRTFRVRVGTSLSSTHHIQENGVPQGSV			
			CCCGG	ACTGATA	LSVPLFLVAINGLAAAVGSSVCSSLFADDFCLY			
			GGATC	ACCCAGC	YSSASIGVTERRLQGAIRKAQSWASAHGFTFS			
			TTGGC	CGTTTAG	ASKTRVMHFCRRVHPDPELYLADEQLHVAE			

			TTAAT	TCCCCTTT	THRFLGMLFDNRLTWGPHLRQLKQKCWRHL			
			CGCCC	AAAAACA	NILRCLGHTAWGAGRQTLMHLEFKALILSRDLY			
			AGGTC	ACAACA	GSQVYDSASPSALRILDPIHGGVRLAMGAFR			
			GCGAG		TSPVPSLLAEAGIPPLRFRRQQLASYVAHVHA			
			GACAG		TRLTIQIAGFCFQTPSFPLHNIDPDQVFQSRSGY			
			TASAA		VLSFPHSSPYLYLPHSPANLISLLGPSLAYSSTC			
			ACCTC		NCPAIQRRRFHHHFINSFWTSLARSVMQKSYT			
			TATTA		LMAPRSWHPSPEVSPATVVGFICRPRSCYTPH			
			AAAAA		HPNRRLLEFNTVLPSPQHRPRSGLPPIAVRVRTL			
			AACCT		LSSLESLPLPPAFSCQSHIPPWSIPRLQLHLQLSC			
			CAATC		DPKASVSPHTLHQQFLDILGTFRDAEVIYTDGS			
			TCAAG		QVEGRVGYAYVHGDHIEQCSLLPGCTVFTAEL			
			GTGTG		IAIIRALDHLRTCPRKSYVICSDSLSGLQALDQC			
			CTACG		FPKHHLVRSIQDSVFSLGXSGRSVAFIWTPGHI			
			TGCCG		GIPGNERADQLAKQAIASQRLDIGTPERDIRLG			
			AGGAG		LHHKISMAWNSEWSTLSTPNKLCAIKPSTAAW			
			ATGCA		KTSLRPSRRDSVVLCCLRIGHTWLTHGHLLHR			
			TGGCT		EDPPNCRCGALLTVVHILLDCPLLAALRQSFNL			
			GTTGA		PSALPSLLADDASMVDRLVGFIRAAGFYHMFN			
			GGTTG		GGYLTFRRPPLNSFNLLFLLSCQSYFQLAALVC			
			AACAG		TG			
			TCGCT					
			AGGAG					

				AACTC					
				TGAGG					
				AACCT					
				GCCGC					
				ACCTC					
				AGTTG					
				TATAA					
				GGCTT					
				ACCCA					
				GGCAA					
				GCGGG					
				GCTCT					
				GCCTG					
				GGTGG					
				ACCTT					
				TATAT					
				CCCTA					
				GCTGC					
				TCGTG					
				GGACC					
				AAGAT					
				GAAAC					
				GAGCT					

				ATAAA					
				CTATG					
				CCAAT					
				GGCGT					
				CTCTA					
				AGCGA					
				GGAAG					
				ATCAA					
				ATTTC					
				AATGC					
				CACGA					
				AATAT					
				GACCC					
				CCGAT					
				CGTTC					
				CCTTC					
				CCTTG					
				GTACA					
				TCACG					
				GGAGG					
				AACAT					
				CGGGC					
				TAAAG					

				ATAGG					
				GATGA					
				GTTAT					
				ACTCT					
				CCAAA					
				ATATC					
				TCATA					
				TGCAG					
				CAGAA					
				CAGAC					
				GGGGA					
				MAGTT					
				TCATG					
				ACGAC					
				AAAAC					
				CTCTTT					
				TCTTC					
				GTAGA					
				AAATC					
				TGGAG					
				GATAA					
				GTATG					
				GTGAA					

				GTCGC					
				GGGAA					
				TGTCA					
				AAGAT					
				GAGGA					
				GTGGT					
				TCAGT					
				GCTGA					
				TAGAA					
				GCTGC					
				ATCAT					
				CCACT					
				CAGTC					
				GAAAG					
				CATTG					
				CTTGC					
				TTGCG					
				ACAAG					
				CTGGG					
				TGATG					
				TCCCT					
				GTAAG					
				CATAA					

				CACCC					
				CACAA					
				ATCAC					
				TGAAC					
				ATGGT					
				CCAGG					
				GTCTG					
				ATTTA					
				CCACA					
				GAGAC					
				CTTCTT					
				CTACA					
				GTCTG					
				ATGAC					
				GAGTT					
				ACGCA					
				CCAAC					
				TTAGA					
				GCGAC					
				GGGGC					
				ATCCA					
				CTTCG					
				TCCGG					

				CGCGT					
				GTATA					
				GGGGT					
				CTGAA					
				GGATT					
				GTAGG					
				GTTGC					
				CACCG					
				GTGCT					
				TTCAT					
				CTTGG					
				CCTTT					
				GAGGG					
				CGATA					
				CATTG					
				CCCGA					
				AAAGG					
				TCAAG					
				GTCAT					
				GATTT					
				ACCGG					
				TGTGA					
				CGTAA					

				AGCCT					
				TACAT					
				CCCTC					
				CCCCT					
				ATGTG					
				ATGCT					
				TCAAA					
				TGTTG					
				GAAAT					
				TTGGG					
				CATAT					
				GTCAT					
				CCCGC					
				TGCAM					
				CTCTA					
				GCCCC					
				ACATG					
				TAGGG					
				AATGC					
				GGGTG					
				TCCAT					
				CCCAT					
				CCGGA					

				CGCTC					
				CGTGT					
				ACTCC					
				CCCGC					
				CAATA					
				TGTGT					
				CAACT					
				GTGGG					
				CAGCA					
				CCATC					
				CACCC					
				TGTTC					
				ACCAG					
				ACTGT					
				ACCAT					
				TCTCC					
				GCAAA					
				GAGTG					
				TAAAA					
				TCATG					
				GAAAT					
				TAAGA					
				CCCTG					

				GACAG					
				ACTGA					
				CATAC					
				ACTGA					
				GGCAA					
				GGAAA					
				AAATA					
				TGAAA					
				GACTC					
				CACCC					
				AGTCC					
				GCAGA					
				CAACT					
				TCCTA					
				TGCGA					
				CAGTT					
				ACGTC					
				GTCGC					
				CAGCC					
				ACCCT					
				GGCGA					
				CG					

RTE	RTE - 40_ LMi	Locusta migratoria	CCCAG GGGGC TCCAA CTAGG AGGCG CCCGC GGGAC CCCGA GCTGA GTCTG ACATT GCTTC CACTT ACTTG TGCAG GCCCC TGCTTT TATCTT TCCTG TCGGC CTCCC TTGGC CAACT	TAAAGTA AAGTAAG TA	QWNVKVRQLFHFKGLYISSTGFERKNIYKGTWI SPDGRTVNQIDHVAIEDKLKKWITNVKTARGQ SVVPITFCSKNSKGAECGSDHFLLKIDLRVGW KPKLTKKLQKVVFHDYNKLVTDEGLQDNFRI NLIVLKPWQHCYRNPGNIAATDSLDAAGSLWEGI RDLTETAKEVIPKRMNRKKVWFDXECOLAILE RQRAKQVWIQSARGRYEEGKERFDEARRRTQ ATLRRVKRNYFKNIMVEAQKDFHGNKSKELF RKVKAIKGEGKKHGFIDRNGRVLTNNDIAI RWMEYFDELLNCEEPXETLEFEETVTDPEYP LQHSKKGTCNCGSGIPPTFEEIKKNIKELKNG KSAGEDGIVAEMLSGGDSCIQNIHKLVRKIW QEEVIPEDWKVALICPIHKKNDKMLCDNYRGI ALLNISYKVYSKCVLLSRIKPLAEELIGEYQGG FQPGKSTTDHIFFIKQVCEKMYEYDKEIHLLFV DFQKAYDSIHRTSLFKALIEFGVSKKLVRVLSL CLDDTRAKIKLGNRTTDSFRVTTGLRQGDLS PILFNLVLEKIIREMDSFNHQGIQVANRNITHLA YADDIVLMTHTKENLKCANDDHGKHSKQIW TESEPHKNQIHDSWKSPTPRNPGINCGNNTYK RVKDFKYLGSIISENSSSGMEIKARIQAGSRCL HSIHNILTSRSISRALKRLRYKTLILPVILYGCQC WSTRKQELNRLAVFERRCLRKIFGPVKDNITG	cd0165 0	394	654
-----	------------------------	-----------------------	---	--------------------------	--	-------------	-----	-----

			CTTGTT	EWRRRHNIELEQLYSEPNIVGLMRSRRLEWAG		
			TCTTC	HVVRMSEERWPRVAMEMIPTGKRPLGRPRKR		
			GACCC	WMDGIREDLQKLEAWKNWEQTAKNRNEW		
			CAACA	ALIISRGPLGLIA		
			GTATT			
			AGGTT			
			TCGAG			
			GCTGG			
			AGGGT			
			GTCTTT			
			CATTTT			
			CCAAA			
			CCCTT			
			CTTCT			
			ACAAC			
			TGGAG			
			AGTAA			
			ACTCC			
			AAAAA			
			CAAAA			
			AAATC			
			CTGGT			
			CCTCC			

				AGGTT					
				GGGGG					
				TTGGC					
				AACAG					
				GCTAC					
				CAACC					
				TGTTC					
				CAGAA					
				AACTA					
				ACAAG					
				GTAAA					
				AACGA					
				TATAA					
				CAAGG					
				CCTCG					
				GATAG					
				GACGG					
				ATGAA					
				TGGAA					
				ATCGA					
				CCCAA					
				GCAAG					
				TGAAC					

				GGACA					
				GAATA					
				CTGGA					
				CTACG					
				GATTG					
				GAACC					
				TGGAA					
				TGTTA					
				GAAGC					
				TTGCA					
				TCAAG					
				CAGGA					
				GCAAT					
				ACGGA					
				GTATG					
				ACAAG					
				GGAAG					
				TAGAA					
				AGGTA					
				CAAAA					
				TGGAT					
				GTAGT					
				AGCTC					

				TACAA					
				GAAAT					
				TAGAT					
				GGCCG					
				AACTG					
				CTCAA					
				CCGAT					
				GTGAA					
				AACTC					
				AACAT					
				TCTGT					
				ATGGA					
				GTTTC					
				AGAAG					
				AAACC					
				ATACA					
				CTGGG					
				GACAG					
				GTTTT					
				AAGTA					
				AGGAA					
				AGCCT					
				TATGT					

				AACCA					
				GTACT					
				AGACT					
				TCAGT					
				CAGTA					
				AACTC					
				AAGAA					
				TCTCTT					
				TTTTG					
				ACTAT					
				AGTAC					
				TGGGG					
				GACAC					
				AAGCT					
				TGTCC					
				TGATC					
				AACTG					
				AATTC					
				GAAGG					
				ACGCC					
				TTTTAT					
				GATGA					
				CTGAG					

				ACTGC					
				AAGTG					
				ACAGT					
				GTCCA					
				AAGGA					
				TCACT					
				TAAGG					
				TGCTG					
				TTAGG					
				TGACT					
				TAAAT					
				GCCAA					
				AATTG					
				GAAAA					
				GAACA					
				AATTT					
				TTTTGT					
				GGACT					
				CCATG					
				ATGTA					
				AGCAA					
				TGA					

Dap hne	Dap hne- 51_ LMi	Locusta migratoria	TAATTTT GTACATA ATGTTAA TACATTC TAGTGAA TTTGTGA AAAGTGA CATTTGA TTGTTAT ATACATG TCTTACA CATGACA TAACTTG TATTAAA TTAAATA TGTATAT GTTTGCT CTTGAC ATTAGGA GATAAGT GAAAATG AATATAT AATGCTA	MAPIEKCSGTVLSSKRQKISSSSCKECEKRVMK GIMCVKCNYWLHEKCAKINLKYISDDFPWTC HACLSVETADLVSLVESKNEIHKLLQNDIEALK TEILCLKEENARLLQEKKQCTVGEIHQEKQQDC ANTCAREIQQTKKETRENTYARNTDVVNENIS SAVDVTVNPTCKKPADKYKWKLVTYSKRKEK TTARPTQEDENEFLIIGDSLKKNVAVPNSKIDV RPGIRTQQLGKHFKNLENAKSAEPDCEPKHNY RGV FVHVGTNSLRSSSEEEIINDTRNVIRSARSV YARSKLVISGIIKRRSVSDKYISQINCAIKEQCD SLGAIFIDPNKFLCENCLGKDGLHLNRLGXVTL SRMFTDVCKIIRNEGNGKLRWTDGGDKSDILA EKANFKHHSKVKEIQNRIPEQMPQHLSQQXEN YXKVLHQNIQSIGNKKLEVEILLSELQPDILCIS EHGLTENQIDNLKISDYTLASYFCRSKQKSGG VSIYIKTGKLETANNKGVKLPTVKEKDFEITGV MIEGLSVFCVYRSPSGDISIFLKKICMNMKRNII ICGDINIDMGKGSKIRQDFEELLQYHMEATIS TPTRVTSHSQSVIDNIITNISRSYESHVIQTAIT DHDGQMFYLCRRHGRACETKAKIKEIRVMSE QNINTFRTILSRETWTETLQAQTVNEKYDAFIA TIKFHFEVAFPKVKRQIKVQQQQQWITREVLE QRRKLQELRQVG DVEGYKTLQKTYRKTIRKA	cd0165 0	933	1188
------------	---------------------------	-----------------------	--	---	-------------	-----	------

					TTGACGT ATCCTAT ATTACAC GTACATG TATTGTA CACAATG TAATCCT ACAGGAT CAAATAA AAATTCA ATTCAAT TCAA	KANSINITILNSKNKIKAAWNVINTERKENAGP NKETEIRSMKIDNRIVTDGSKIANILNDNFINVV ENLKLERKSPKKKGTKVTKYSYSTMFRLRPVTE DELRKTIQKLKAKKSAGDDEISNHVIQLVSEQ HTHTISRYSKQLFQRWNFSREIEDKVSNTLTPL VDIANSSFRDGIFPEKLKTSKVVPVYKKGNKD DPNNYRPVSLISGFSKILEMLMYNRLVDYLDK HNILVPSQHGFRMGKSTESAIASLTEYVIQCLD EKRVAAMFLDLSKAFDTIEHDILIKKLENYGI RGLSGEWIESYLCNRKQYVVSIGNSFRSETKTIK YGVPQGSVIGPLLFNLYVNDIGGEEKEKKILYA DDMTVLNHGQNMQQLERSAFISANTTAQWLT ENELVINLKKTMFTVFRNKEKSVQMNLKVND MYLEEVESTKFLGVVDNELKWKEHINYVCK KLSSVIFIMKQLSQYADKQLLCTVYHGLFEPYL QYGVTVWGNSNIQETKKIVHFTKESNQDHTK KKELGDLQCGETQTSKKRKKLFTLQKKAIRIL RRKNSETCRNCFKSLGILTFTSLYIYKTIMYITK EISLYDTNASVHHTNTRGKNEIRNIPHRLTILEK GPRYSGIKFLRNLPKNLRDNIHGNNFSRKLKD YLMEKEFYISIEEYLLN			
Ving i	Ving i-	.	Locusta migratoria	CGTAC CCTGG	TAGCAGA AACTAGC	MDLVLERPHTKYGSAIFVKSGLSIQSTGLTSEN DVEILTVGFQSCTVTSVYKPPNAEFKFVEPLNF	cd0165 0	408	681

28_ LMi		GCGTT	AGCGTCC	QSQPIKVVAGDFNCHSTSWGYREXDKNGEDL			
		AGGAG	TACAGAA	EAWAEQMGLVLIHDPKLPSSFNSGRWRGQYN			
		AAATC	TTAGTCA	PDNIFVSNHIAQQAKKGVLGPLPHTQHRPVTC			
		CTWAC	CCTCTTTC	TVAADVVPETVPFKRRFNFKRAQWEKFSIELD			
		GGCAG	CATATTT	KEVIKLEPSPEMYDSFIQLVRTISRRNIPRGCRT			
		ATCCT	TATGTTT	SYIPGLDKDSKSLQKYEHLFETDPFSEETTQA			
		TCCAT	CTAGTTT	GEELLNAVSAATRXQKWCDMVTQLDMKQNSR			
		CACMA	GTAATTT	KAWKLLKRLNSDPTSAPPSLVNVTPDQIAHQL			
		GGACC	GTGTATT	LLNGKTKRRTAAKKFQRYTDEEDYLGTPFTIE			
		AAAAT	TTAATTG	ELEVVFKSMKNNKAAGLDDLPEQIKMFGPV			
		CAAGA	TATATTT	TKQWILNLLNNCISKLQIPKIWRKAKVVALLK			
		CGTAT	ATTGAAT	PGKEPTDPKHFRPVSLLCQLYKVLRLILNRISE			
		CTCAW	TTTGCAT	HVDKNLINEQSGFRPGKSCCGQVLNLTQYIED			
		CCACA	TCTTGGA	GYERGQVTGVAFIDLSAAYDTVNHKKLMRKV			
		ACGGC	TTGTAAT	YSVTKDYRLTQFIQCLLQNRFFFCNSAHSSSSV			
		TTTTCT	TTGCTCA	CCKIGGFFVTLHNKKSRLWRTLKNGLPQGSVLA			
		CAAAG	ATGTAAT	PLLYNIYTNDQPISSEVRTFIYADDTATAAQGK			
		CMAK	TGTACAT	TFEEVEMKLTAALDGLGKYYDDNHLKPNPGK			
		MTCTC	TGTAAAC	TQVCAFHLRNKEARKKLKIXWRGQQLLHYDT			
		GMAGK	ATATATT	PKYLGVKLDRSLTFKDHCQETKMKVSARNSII			
		AATAT	TAGTGAA	RKLTGSTWGSQPSVLRTSALALCFSAAEYAAP			
		CACTG	TTTGGTG	VWRNSTHAREVDTALNETGRIVTGCLRPTPVR			
		AATAA	TCCCTAG	KIYEIMGIAPPDIRRRTAAETEKKKQETDPRHP			

				AGAAG	ATTGTAA	LFGHEPRKPRLKSRRSFMSTTTSTALRPGTRRT			
				TSMTA	ATTGTTT	ELWQNSSNERAGENIKEEPAEGFHLPYATWKA			
				CTMTC	ATTATAA	LNRLRTGVSRCKTNLRKWGFTNEDELCECGD			
				CGACA	TTGTATA	VQDPQHLLVCRNLPNPCTIEDLHAANDVAIKT			
				GCTGC	TTTATTG	AEFWSLRSI			
				AGTAA	ATCGAAA				
				CACAG	GATATGT				
				CTTGT	CTTAGTC				
				KATAT	CTGCACA				
				TTTACT	CGTAAAT				
				GCTAC	AAATAAA				
				AGGAG	AA				
				ACCCA					
				CAGAG					
				GGCCA					
				CAACA					
				GCAAA					
				GGCCA					
				AAAGT					
				AAATG					
				GA					
Nim b	Nim b-	.	Locusta migratoria	CCCCA GGGGG	TAAGATT TTATTGC	MKRVKLDNGRDTAVSEGRKKNNAPSYDKRTF PSLSTPWEEGQARRLGVKPFPRYLVCTR TDGG	cd0165 0	1012	1291

40_ LMi		CCCAC	TGACGCT	TFTATKPLFFVEHIEDKFGEVETMSKMRTGSLL		
		AGCTC	TGGACAG	IKTVSAAQSAALCACDHLGDVPVSITPHQSLN		
		TTTTGT	GATGGGT	MVQGVIFHRDLLLQTDEELRANLERRGVHFVR		
		GGGTA	GGGTTTG	RVQRGPKDHRVDTGAFILAFEGDTLPEKVKV		
		CGTGT	TGGCCCT	MIYHCCDVKPYVPPPMRCFRCMRFGHMSSRC		
		GTGGC	CCCTGCG	TAVPVCGECGRPLHEGNPCVPPVVCVNCHEHH		
		GCGCA	AGCAGCA	SSRSPECVYKKEKKIQEYKSLDRLTYQEARR		
		CGGGT	GGGTTTG	KFERLHPVQLSTSYASVTSFTPPSSSSHSPILPP		
		CCCCG	GAGCTCT	PPPAPPLPPSQRPSSSQPEKCSPSLALAGEGAP		
		AGCCA	CCCCTCC	SQGSSSRPPDRRSSASSRQPDIAQADNTKVAR		
		TTGCA	CCTCCCC	SSVPDLTDSSSLSGTCPSRPAREKEKEKRKKKQ		
		GCCTT	CTGACGA	KSQDKAPPVVLEAPPLSSQPNNPTELVS DPLFM		
		TCTTCT	CTTCCAT	DVVPSSVTDVDPALPSFVTCISPF AFHMDPRS		
		TACAT	TTCTCCTT	VLIQWNCNGYYRRLLELKHLIASYSAACVVLQ		
		TCCTG	TTAACTT	ESHFTDGHSPTLRGFRAFCRNRVRPLQASGGV		
		GGCTG	GGTTTTA	CTLVRTDVVSSWIPLRTTLEVVAVRVHLHTPL		
		CATTA	ATTCACG	TICNLYLPPDRPPTSAALTALLQQLPPPFLILGD		
		CCTCC	TTGTTTCC	FNAHHLSWGSAYTTSRGRLVDQFIADHDLCLL		
		CCCAG	CCCTTTTT	NDGSPTHFSAAHGTFSSIDLSTSPPPPFSLSFA		
		TCCTC	ACATTTT	GHRMTTFATVTTFRCRSRHLFLHRPFTHFSPPP		
		TCCTC	TAGGCTT	LPFFRWSPHDDLCDSDHFPLILSLPSHWTRDR		
		CTCCC	CTCATTT	IPRWVFRHADWPLYTSQVVFTPSPSDSIDEVIS		
		CATCC	AGGAGCA	DVSNAITRAASIAVPQSTGPPRRRPVPWWNATI		

TAATT	TACACTG	ASAIRERRRALQHLKRHPSRAALVNFKRLRAK		
CCTCC	TTCTCCT	ARYIIKQSKRVCWERFLSSLGSSVPLSRVWAM		
CCGTC	GTATCTG	LRSIQGCQRPSSIPGLALPDGVCTDPLVIAEHLA		
CCTCT	CTTGTC	AHFAKASATASYPVAFRLKKQQAESNLCFTT		
CATCT	TTAATTA	HREEPYNKPFTEWELRLALSSSHDSAPGPDSIH		
CCCCC	ACGGGGA	NQLLKHLSPQQRQYLLQVFNHIWLQGVFPSQW		
CCTCT	TTCTAGA	RDSIVVPVIKPGKDPSSLDSPRPISLTNPCKLL		
CTCGG	TTTGTCG	ERMVARRLCWVLESRGLLSPYQCGFREGRSPI		
TGGTT	CGACGTT	DHLLRLETAIRKAFSKRQHLVAVFFDLRKAYD		
TCCTTT	CCTTGCA	TAWRRHILCTLHDWAYEAPYQFLSVSFYPAAC		
AAAAC	GAATCCC	SVLGLRGSLPIFISQFLSRRMFVRVVGSSFSSPQI		
GGCAC	GCTTTTTT	QENGVPQGSVLSVTFLIAINGLVSSVGSLVTP		
GGCTA	TATTGGT	ALYVDDFCIWYSSHSVASAERQLQGAIIRAST		
TTCCT	TTTAGTC	WALSHGFQFSPSKTRVMHFCRRTTVHPDPELF		
ATGGT	CTGTTCT	LGTQRLEVVAQHRFLGLLFDKKLTWLPHIRQL		
TCCGA	ACTTGAT	KISCMRKLNALRFLAHSSWGADRRTTLIHIYRA		
TGATG	TTTAACC	LIQSRLDYGCQVYGSAAPSALKLLDPVHHCGI		
GTTGT	CCCTTTT	RLATGAFRTSPVDSLLAEAGIPPLQTRRHQLLV		
AGCAA	AGTTCCT	SYAIAIRKFPGHPTYHVLFEGLRTPDTRPRV		
TTTTG	CTATTTT	GLPVGLRLASLCRDFNLPSLDCMRRAIPRTPP		
ATCTT	ACTCTCG	WLVPRPQIRLDLYHAPKASVSPIIFRLLHTVL		
CGTCG	TCACGGT	EEFQGATIFYTDGSKFAGHTGYAFTSPIGVECH		
TTTTCC	CTTAGCC	SLPESCSVFTEALTAISRALHFIKQVALDRVLC		

			TCTGTT	GACCACA	SDSMSSLQAIDRCYSSHPLVIDIHDLLADLRRT			
			TTCGG	CCCTGGC	ACSVAFWVPSHVILGNEQADQLARVAVTP			
			CATAG	TTGCACT	CPFPWVIPGADLRVQMRSLLIQDWNTNWQAT			
			TTCGG	CCATACA	ALSNKLRSIKSTTAAWSSSLRPSRRESTVLCRL			
			TCCCC	GAGGTCA	RIGHSRLTHGFLLRNEPPPHCGCGALQSVAHIL			
			CTCGG	TGGGACT	VECPLLLALRAKYGIPHSLSVLANDSRIADRV			
			GGTTT	GATGACC	LGFIRESGFYSRL			
			GACCT	TAGCTGT				
			CCACA	TTGGTCC				
			AAAAT	CATCCAC				
			TCACA	CCATCAA				
			GTTAC	CCAACCA				
			TAGTG	ACCAAC				
			TGAGC					
			CATTT					
			GGGGA					
			AGAAC					
			TCCCG					
			ACCTA					
			GCCCC					
			TATGG					
			CGTGG					
			ATCCC					

				AATCC					
				TTTCCT					
				TTCCC					
				ACTTC					
				CTTCCT					
				TGCCA					
				TAGCC					
				CCCTC					
				CCCCG					
				CTAGG					
				TCCCC					
				AGCAC					
				GTGTA					
				GCCAG					
				TCCGT					
				GTGGG					
				GGGGG					
				GGGGG					
				GGCTG					
				TTATG					
				TACCC					
				TTATG					
				GCTGA					

				GCCCC					
				CTGAC					
				AACGC					
				AGGGA					
				TCACA					
				CTGCT					
				GGTAC					
				CTGAG					
				CTGCA					
				ACCTC					
				CCCAT					
				GCAGG					
				CCAAG					
				GAGTC					
				GATAC					
				CCGTC					
				CTCGT					
				GGGGT					
				ATCAG					
				GGCTC					
				CCGGC					
				AATGG					
				CCATC					

				GTGCC AGGTG GCCTT TGCTG TGGCT GGGTG GCGCC CATGG GGAGA GCCCC GGATC GGAGT GGGTG GTACC AGGGC GGACG TTTCG CACG					
RTE	RTE - 68_ LMi		Locusta migratoria	GGTAG GGTTT CACCC TATCC CACAA	TGAGGGA AGTTGTG TAAACGT AACCCAT AGATGGC	MDNYRILKGRPAIKVKKTEKTKNVVLILGTAF FIHNSITESIVEFRSPSPRLPVVTLASGNKLYTIF NAHAPINEDNRKNPEKVEEFWSQLEELFRTF DKHVKILMGDFNAQIGKERKFRNIVGKYVAH RRTNRNGERLIDVCESFGMKIMSTQFKRAPKQ	cd0165 0	488	752

			TCTAA	TGTAACG	MKTWRSPNPMLGEFQIDHVATTSPKEIMNVRV		
			ACCCA	AAAATAA	QKGANFDSHYLTRVKVKPIPGNRRRKTTPRIQ		
			CGCAA	ATAAAAA	KYRIEELKISEEIVVKYQHELDTLNSNEWDKM		
			CTACC	A	AQKIQEAAKQTIFLAKKKKHPWWDDECEDAL		
			ATGAT		RKRSEAWNKNWNSTKQNFEEFKTQRRTTAKVF		
			GGCAC		RNAKRKFERDQLKELEENFQKNNTRSFYQTFK		
			AACTA		SKLRGYSSPNICLKMIRANRLLITQHLLFKNDQG		
			CGGAA		KLIMNNKHNCHEHLARYFKELLNCDPPPEKLTf		
			ACTAG		QKPQNKLDPSTPPDKKEIREIIRGLKNNKVSgK		
			TACCC		DSIVAELLKYAGDSFLERLERLFQKIWLtGEIP		
			CAGCC		EEWRDAVIHPLHKKGDKSNVNNYRGISLLPVA		
			GGTTC		YKILAIALKHRIEQHIDQQLGeyQAGFRTGRSC		
			AATCC		AEQIFTIKTLIKTKALAADKKVIMTFVDFKKAY		
			GGGGC		DSVDRDTLIEVLKEFGIDDKTTRLIQQTLTNTR		
			CAAAA		SQIKFLGELSEPFEIKTGLRQGDSLSPILFNCVLE		
			TTTTG		KVIREWRKEMSQFNFPNGIRLGRKQSGLTfEC		
			GTTCA		LAFADDIVFFAENLQIATKQIEVLKETAVKAGL		
			AGTAG		QISFEKTQYMNINATDPVWTMRtKYGDIKRV		
			GGCAA		QRFKYLGEILTPNNNEKAAIEERARKMEIAFRL		
			TGGAT		CQNTYKSKSLSYQAKFRHYSSVVKPECLYASE		
			TCTGG		TIARKGLSDLEKKERKFLRKILGPRKISNEFTFR		
			GGGCC		LRSNQELYQRFENITTTMKKRRLSFYGHIKRM		
			CTACA		KSERLANKILLFQENRKTQVPWVKEVHRDLEK		

				CCATC	CGITAEIIVNRQIFRKKVAEVKDFFEKTRKNR			
				ATGTT	VFSAEERARVSQRMKEYWRKRKEENFRK			
				TGGAT				
				CAGTC				
				AGAGG				
				ATCCT				
				TCATC				
				CCGCA				
				AACTC				
				AGAAC				
				GAAGG				
				AGAAA				
				CATTT				
				CTTTG				
				GTACA				
				CTTAA				
				CGTGA				
				ACACC				
				CTATT				
				GAAAA				
				TTGGA				
				AAACA				
				AAAGG				

				AACTG TCAAA GTACT TGATG AATAC AAAAT AAACT TCTGG CAGTC AAGAA ACAAG ATTCA CAACA GATGA AGTGT CAGCT					
Ving i	Ving i- 3_L Mi		Locusta migratoria	GGGGA CAAGG AAGGA GCCTC CCAGA AGATT GTACC	TGAGCCT GCTATGG ACAATAA TGACGTA TATGTGT GAAAATC ATGAGTA	MMKGPFSEPCLTAISLNIEGLSTEKEQLISNTCK LHSCDVLCLQETHRGPSQRRPRIPGMILVKECP HDQYGSALFIRPGIKVSNIRFNTVNNTEIISVIIG SITVSSIIYKPPGERFAFSQHVCHPQLVMGDFNC HSKVWGYQDTNEDGELLESWAEEQELNLIHD QKQPASFNSARWKRGNPDLIFVSKTLSGSCN KFVLNPIPKTQHRPIMVKFRPVIVPQEVSFKRR	cd0165 0	463	722

				GTACA	TATATGA	FNFKRANWKKFADILDKEIVYLPPNAENYEKF			
				TGTAC	ATGTATA	VEAVKRSARLSIPRGCRTLYIPGISNKMQLYN			
				AGTCG	ATTAGTA	EYIEQYNHDPFSVETIKLGEQLTETISKDRQES			
				GGCGT	CTTAGTT	WKHMLESMDMKHSSRRAWNLLKRLSXDPA			
				CCCCT	TTTATTTG	PKDQVSVTANQIAHQLLLNGKPTRNSKIVTSY			
				GGGCA	TAATCTA	KRPQVDEIPPVPYFTSIELETAISSLKSRKAPGP			
				ACGGT	GTATATG	DDVVNEFIQNIGPIAKEWLLQLYNFCWANKSV			
				AAACA	TAGCCTA	PKLWLKAKVVAVPKPGKSRDDAKNYRPISLL			
				CCGGC	TTAATTA	CCPFKLYERLILNKIGPQLESQIIPQQAGFRPGK			
				AAATC	ACTACAG	SCTGQVLCLTQYIEDGFQKGEKTGVVFDLSA			
				CTTCC	ATGTAGC	AYXTINHKILLENFSKKIFQNEGNYHLTEVVR			
				ACACA	TCAGACA	LISNRRYYVEFQEQRSRWRNQKNGLPQGSVLS			
				CTGAG	CGATTAA	PLLFNLYTNDQPAGPDARSFIYADDVAIACQSR			
				CTACG	ATAAAAA	SISHIERNLSDALEVLTEYYNRNQLRPNPAKTQ			
				TCTAA	TAAA	TCLFHLNNQEAKRELQLEWNSVKLKHCPNPV			
				ACCGA		YLGVTLDRTLSYKSHVEKTRAKIYTRNGIIRKL			
				CATTG		TNSHWGANPEILRSSSLALCFAPAEYACPVWR			
				TATTC		RSTHAHKLDTALNDSCRLITGCLKPTNRDLY			
				AGAAT		MCSGIAPPQIRRTWRWLKEADSARIEDIPCS			
				CTGA		ISQWCKIKIEERFHFHPLNESQSTTRESMWTQ			
						KLEKKKKKNSLQCWNIKEKLPAGSQLEWRVW			
						KSLNRLRCRMGRSKDNLFKWGYAEDELQCG			

						EKQTMTHLLSCPSLPAPCTFQDLWLANDAAV KCAEHWREI			
Dap hne	Dap hne- 28_ LMi		Locusta migratoria	GAAGA TGGCT GCCGA TCAGT ATACT CATTTT GTGCG CGCCG TTGTG AATAG TAAAA ATCGC GTGTT CAAGT GTTCT CTGTG CGAGA AATAC AATGC TATAT GTGAT	TAATTTC CACTGTC TGTGACC GAGACAC CACTAAA ACTGCCT ATTTTAG TTACTTTC ATGCATA ATGTTGT ATGGCAT AGTTTTTC TGGGGAA TCAGCCA CTTTCCA AAAAAAA TATTCAT TGCACAA AAGAGAG CGATCAG GATAATG	MPRKFFRPKMFTHGKPHETVLQVPPNNRFSVF EENTNTVDSFATVLNNERSANTSRKIGKKCV KVNIHADSHGRGLAKILCEDXDVDAYSIVKP GAPLYEIGKNIKDGSPMNSNEYTVIVGGANDV YKNELSSAIQNLNIILESNNSNAMILGIPHRHD LIYESCVNKEIAKANKQFENICNLYPNTQFLPID FLQRDSFTRHGLHLSRKGGKLLCFKIAEMVSA NTKKGSSNSQKTVQTSVPQQPTQSPSAQLYNIN ATTKTPTDDPILPESTNSPPSVDSIEPTSVAAMP GLIVDTTTPASHQEEDIWASRIVPEPMEEASS TPEDSSQLLKIQQFFMVSSQGRKNSVIATPPSK PNVNTKREINPINHPKTLVSVNKLTVFHQNIQG LRSKLDEISINILDKDCIDSADVCFSEHFMKLD TDSLQLDSYNLAALYCRQSLEKGGVVIYVKN KIVHRSLDLSDYCIDQHCEVCVVISTNLSEIT VLSCYKAPAGKICIFLKQIVSVNSYFLQKQRILL IFFAKTKNIIVLGDFNVDYLIDNMGKNDLSHV MSSFNLTAQVDFPTRVTANSISRMTISSPNDNIF LDKSRNPNTIVKRVLNGLSDHDGQLLSIEGVG PLSKNKSYSKTFRAINSQSINTFCDLLQETDWD QIEHLDNVDDKYNAFLNEFICLFEAAFPKRTA	cd0165 0	924	1185

			AATTT	AGCGGAG	KPKHETYCHKNWLTRGIKTSCRRKRNLASYSL			
			ATGCG	TCCAACC	KSCFSLEKEKHYKLYCSILKKVIRAKSNSIKSEI			
			TGCCG	TAGACAT	DVSSNKIKTIWGIIKRETGKIPITENSLEIKKMM			
			TGTCA	TCTGTAG	SFMIMLKKNDVIHDNVEIITDIFNSHFLSSAAQT			
			CATCT	AAACCTT	GKTCSINEAMLFLQKSIKKQSKLTCLLSQCMK			
			TAATA	TTTAGAA	LRLKTKQINMSPVTVHEIEVIIKEMKSKNSVGI			
			AAAAA	AATTGGG	DGISSRILKHICYKPVSKILCHIFNKSIQQGIVPER			
			CAGTG	GATCTGA	LKYAIVKPLFKKGDKTDVTNYRPISLLTIFSKIL			
			GAAAA	CAACTGC	EKL VHKRFIDHLTQYKLLNSSQFGFQKGISTDQ			
			AATGA	ATGTCAA	AIFSFTNQILESFNKLSVPGIFCDLSKAFDSVN			
			ACAGG	TATATCT	HEILLEKAKHYGFIGSAGKWIKS YLTGRKQKV			
			TAAAT	ACTCCCT	MLNNSEGHSVSSSWGSIGCGVPQGSVLGPLL			
			CCTCG	ACTGTGC	LIFINDLPLCTSLKTHFTLFADDT SILINESSNYN			
			TGTAG	TTTATTGT	LEESANIVLREALNWFVINGLALNANKTQIIHF			
			GTCTTT	CAAAAAT	QPSKKMEDPITIKCNGTDLLQAESTKFLGIHIDS			
			GTACA	GTGCACT	NLKWSEHILQLHKKLSSATFAIR			
			AGTCT	TGTACAA				
			TAGTG	AACTAAT				
			ATCAG	GATGTTA				
			CGTGG	CCATACT				
			TAAAC	CACAATA				
			ATGGA	CCAGGAG				
			CATCT	GAAGCTA				

				GCATC	GATCTCC				
				AAAAT	ATTATGA				
				GTGGA	ATTAAAA				
				CGAAA	AACCTAA				
				TTCTG	TACTGTA				
				AGTGA	CAAAAAG				
				ATTAG	GAGTCCA				
				CAACG	CTATATG				
				AAAGA	AGCTGCA				
				AAAAA	AAGTGTT				
				TTATC	CAATGCT				
				AACAT	CTTCCAC				
				CCTGT	TATCCTG				
				GTGAC	AAATCAC				
				GACAT	TTGTCCA				
				AGCAA	TGAGCCT				
				AGTTG	CCCAAAT				
				CGACA	TTAAACA				
				GGAAT	ACAGCTC				
				ATCAT	AAACAAT				
				ATTTTC	ATCTATA				
				AATCA	GATAAGT				
				ACAAA	CCTTTTAT				

				CAGTG	TCAGTTG				
				TGAAT	AAGAATA				
				ACATC	TTTCAAT				
				GCCAA	TTGTGTT				
				ATATA	AACAATT				
				AATTG	TATTTTTT				
				TGCTG	AATTTAC				
				AGTTT	AAGATGT				
				ATTAG	TATTTTGT				
				GCCTA	GTAGCAT				
				AAAAG	GTATTGT				
				TTCAC	AATTTTG				
				GAGTG	TCAATAT				
				GTAAA	ACTATTC				
				CTACA	TTGCTGT				
				CGAAA	GAGATCT				
				CGGTC	ACCTTTTT				
				GCGCA	TATGTAA				
				AGTGC	TTTTATTC				
				CTTCA	TTTAAATT				
				AATAA	TTTTGTGT				
				CCGTT	AATGTAA				
				TCAGT	TTTTTGTA				

				GTCCT	ATTTTAA				
				TGAAG	TGTGTAA				
				AKAAC	TTTTTGTC				
				ACTGA	TATACCT				
				TACTG	TTAATGT				
				TTGAT	GTAATTT				
				TTGTA	GTCTATT				
				TTGAA	GCAAGAG				
				AAACA	GGATTCT				
				AAAGT	GTATGTT				
				TTGTCT	CCTTTTGT				
				ACAAA	AACAATT				
				T	TGACACG				
					TCCTACA				
					TCCAGTG				
					AATGTCT				
					CGCTGTA				
					TTGGATC				
					CATGGGA				
					TTAAATA				
					AATAAAT				
					AAATAA				

Pene lope	Pene lope- 3_L Mi	Locusta migratoria	AG	TAATGTA AACACCA ATCCTGG ACTTCCT TTTCCAC CAATAGG AAGAAAT AAAGCAA ACACCAA TGATAAG AGAGCTG AAACCAC CAATCAT AACTGGC AACATTT AATATTT AAGACAC CAGCTTT CCCTCAT CCAGCAG TAAACGT AACACCC AGAAGAT	SNRAKNICAQASKALLRERIRHIRSSLDSSNRK LIELHLALSSKLAEDWEKVDRLTFQQAERTN DSVATRQMKKFDNLQQKTTGNGQTMDTQRT VVNLSSHELSDAATQILSKGMNFAITPRRVPKE DIIESVEASLRHLPQAEAEKIRQETVRVLAHAK PPKQNISNEERRAIHELKNNPDLVVVQSDKGNA TVVMNTTDYDDRMEKLLSEPIYKKLKGDP KVIKQTQALLKKTRMNFDTARRLNPQVPQAP RIYGVPKIHRESFPLRPVCSIDSPTYRLAKELVP KLQHLVGHTDSYVKDSTHFVALLEEKRISTTD LMVSFDVKSFLTNPVSDTMNILEERVPPDVC DLVQHCLTTTYFRWKGEFYEQTDGVAMGSPL SPVAADIFMEAFEETALQSAPLRPDCWLRYVD DTFVIWPHGEEELQNFHQHLNQQHNKIQTME IEKNGVLPFLDVEVYRKPEGKLGHRVYRKATS TGRYLHATSHHHPTQKKSVLHTLTKRAHRISD EQHLKAELQNIRSIFEANGYGKRMIEQTMTSN DGNKREEQKEEQTIALLPYVKGVTERVGKILR REGIKPIFRSNNKIRDVLGSMKDKTDKLHTAG VYEIGCECGQVYIGETGRPINTRIAEHERYIRLG QHTKSVAEHQEECGKQIGFQEVTVLAKQPYT LKRKIREAIEIAKRPGNMNREDGYRLPASWLP	cd0030 4	380	464
--------------	----------------------------	-----------------------	----	---	---	-------------	-----	-----

					GGCCGCT TGCAATA GTGGCCG AAACGTC GGTGAAA TATTTGA TATATTG ACGATGC GGTCCAA TACCCAG AAGAGTT TTATCAA CAA	AVAALRDDTARSAVHRGPHSPETTAARTQS KQHALAGARHQCYVTATTAPHSPVSTNGKH			
CR1	CR1 - 3_L Mi		Locusta migratoria	CTCAG TGACC GCCGG CCAGT CTGCCAG GCACA GCACTTA CGTGA AATGTGA CAGCA TTTGCAG GCAGC AGTATCT TCTAC ATTGT	TGAAAGT GGTTAAT GAACCCT CTGCCAG GCACTTA AATGTGA TTTGCAG AGTATCT ATGTAGA	MARNCDCCERMEAEVLRLSQLQAALASLAT LETAVGGRRCCGPLAGVQQTPCASVVPSDRPA AVASPLTAREGVAPQSVVDWEVSSARVRRRK TSRGTEVEASPVSLANRFEALSVAEDVPDADG VACPVLAPLPSRSGHSQRVGLLVVGSSNVRRV MEPLRKMAAREGKKSNVHSVCIPGGVIPDVER VLPDAMKSTGCSQLQVVAHVGTNDVCRFGSE EILSGFERLSEVVKTASLACKMKAELTICSIVD RTDCGPLVQSRVEGLNQRLRRFCDRVGCRLD LRQRVVGFRVPLNRSGVHYTQEAATRVAGAV	cd0165 0	871	1141

			AAGGT	TGTAGAT	WRGLGGFLGVRSPLYSGSGYTGSGGCVAWTG			
			AACCG	G	RFFRLEGLGNTQNGFQSQRGQVRHRKKADLG			
			GCTTC		TIGVTVVNCRSCVGKVPQLALIESTEQAQIVIGS			
			TTAAT		ESWLTPPEISSAEIFSKGQMVFRKDRINKKGGGV			
			AGAAG		FIAVTNGLSCSEIEADTSCELVWVEVIPDNRNK			
			CGACT		LIIGSFYRPPDSDDAVAEMFKESLNHISNRYPK			
			TATTTT		HTVIIGGDFNLPSICWHNYVFKTGGKYKTSSEL			
			CATTT		VLNAFSENYFEQLVHEPTRIVSGCENTLDLLAT			
			CGTGT		NNPEIVKSIKTDTGISEHRVVVARLNVEIPKCT			
			GTGCT		KNKRKIYLFNKADKNSLHAFSLDSLQSFKTKN			
			TTCTTT		ASVDQLWLTFKDIISVAIERFIPNKLIKEGADPP			
			TACAG		WYTKQVRSLQKQKTHAKFKKTQNPQDWQ			
			TTCTTC		SFTETRNLARTSMREAYSSYHNEILSRNLAENP			
			AAACT		KKFWSYVKYTSKGTQTIPSLRDSNGNTTDDST			
			TTAGT		AKAELLNTAFRNSFTKEDEVNTPVFNPRTAAS			
			GTGCT		MRNLEVDVPGVVKQLKSLNKNKSSGPDCIPTR			
			CGCGT		FLSEYADAIAPYLAIYNRSLSERSVPKDWKVA			
			GTTTA		QVTPIFKKGNRSDPQNYRPSHSPKLQAISLTSIC			
			TTTATT		SRILEHILCSNIMNYLEENDLLTNSQHGFRKYR			
			AAGAA		SCETQLALYSHEVLSAIDKGSQIDSIFLDFRKAF			
			TCGCG		DTVPHKRLLMKLRAYGISPQLCDWIHDFLSER			
			TTTAG		TQFVVIDGKSSSKTEVISGVPREALNRSDIWS			
			TTGTC		QGSVIGPLLFLVFINDLGDNLNSSLRLFADDAV			

			AGTTC	VYRLVKSTEDQKQLQSDLDKISTWCENWQLT		
			AGTTG	LNKEKCEVIHMTTKRNPLNFNYAINQSNLKVV		
			TTGAT	NSTKYLGVTTITNNLTGENTGRKFYQIPRSYNYE		
			ATTTT	QLDWREHIENIVGKANQRLRFIGRTLRRCTSST		
			GATTT	KETAYITLVRPLLEYCCPLWDPHQVGLTEDIEK		
			ATTAG	VQRRAAFVLSRYGGEASRKCCTNWGGNHEG		
			CTCGC	SEKGSSFCFITLRGGSVTEMLHELGWKSLKER		
			GTATA	RYSLRRDLFAKFQSPTFTSECENILLAPTYLGR		
			TATTTT	HDHPNKIREIRSKTERFRCSFFPRAIREWNGRE		
			TCAGA	KL		
			TTCGC			
			CGTTA			
			ATAAA			
			CAGTT			
			TAGTG			
			TTAGG			
			TTCAT			
			ATTTTT			
			CGAAG			
			CTCAC			
			TGTTT			
			ATCAA			
			ACTTC			

				TGTGT CCTCA GACGT ACTTA CTTGG AACAT ACATA GGTTT TTAGA					
Dap hne	Dap hne- 41_ LMi		Locusta migratoria	T	TAATGTG ACAATCA AAATTTA ATTTATA ACCTACA CTCTTTTA TTGTACA TATTTAT GTTAAAA AGCTAAT CATGTAA CAATATA TATTTTA ATATGTA	EQQCM PQGTC SDDQACPHKSWSVKSQPSFLT TGMNLCNYKPVQRLEVKDCGSM SFFHLNIQSV RNKLPELELFLLEFQCDILCLMSTGHTLSNEHW LKEHEHNMYTPLGYKFLTGF FRSNIEKGGVSIL MKNDLNCKIRIVDLKHLASEGQIEIAA IHLPPS ADTLKSHNF SCLNDKPTRQNACLDNVISNLSR DDVNLNVVKPMLS DHDGILLKLIVPKKDHRPT QPIVNRPV RQLKPINITNL TQRLLTFDWDATV MENTTLNNAFDYPPQLIIMLLIIRPNLFNEYCPP PPSTKV KYKEETKEKLT MVH PHLNSLRILLKIY YEK SCTDSEYRHSYNKLKN IYKKDIISAKKEA NGTYISQATNKCQA AWKIKA EAGLQVVNPPI PRADELNKNFINS GQSAIDHSNLDDNADSFLSS LETELXVDK LKFRWHAIYPKDILDVARSLSNS	cd0165 0	498	740

					ACATTTA TATATGT TAACTTT TGTACCA TAATGTA TGCAATG ACGATGC CTTATAC AACATCA CAGTTGT CCACAGG CAAAATA AATCAAT TTCAATT TCAAT	RTEDIYGLSNFTIKRIIEAICIPLCQLINRMFLEG TYPDCLKQTIVLPIYKKGDRTSPQNYRPIAIVPI TSKIIETIMKNQLNDFFCNSLLNKAQYGFRSGL STVKAVQEVVHSVLHGFENRQHSYATLVDLS KAFDTVSHGILTKKLETYGITEKELALLKSYLA NRNQMVLVNSTNKS DPLKVQYGVPQGSVLGP FLFIVYVNDLPNLLPFKSVMYADDTTFITSHND VNAVKQMNN SMMEKATQWFQENRLTLNNN KTEHILFSLGNINNEQNHPVKLLGIYLD SKLCW NTHTEKLCSKLARVLYLLRKLKDCVSKELML MAYYAFFHAPYIMEYFFGETPLERKKSSSGKR KPSDAFLARTHAYGILCFFPCTLHYGILLWGNS AGAKEVFLWQKKAIRCIFGLSKLT SCKQYFVQ HKILTVPSLYVLKALMHARQNVDEYESRSDV HLYGTRNRNLLDIPWARLSKVKNSFLHMAKIF FNKLPYNIRELHENKFKIVIETWLKEKALYNTN EFIWTENKSCY			
Pene lope	Pene lope- 37_ LMi		Locusta migratoria	TTAAG ACGGC CACTG CACGG AAGAT TATGG	TAAAAGA ATCCGTT GAAATAC GCATTGC GGAAAAT CTGATGA	KLLLNAVIQLLRKGLHYEIPRESGIVVLKADKG NSTVLLPRRLYLEKMYGLLSDTMYRKIEHDPT ARVQRKTCTLLNSSSLAPETSSGPYSTCATEDV HTFEFFLVSSRDIKRLRPHGSVPPRLYGLPKVH KEGIPLRPVSNIDAPTYDLAKHLASLLRPLVG KCTHHIRNSADFINRLGALHLSSDLLVSFDVV	cd0030 4	246	330

			AGAAG	ACCGTGA	SLFTKVPLADSLALIGKRFQSDITGLFRHALSST			
			GGAGC	CAGCGGC	YFMFNQEYFEQTDGVAMGSPLSPLVANFFME			
			CATGC	TACCAGC	DFEERALESAVCKPTVFWRYVDDTFIVWPHGE			
			TATTG	TGAATAA	DKLMEFLHHLNSIHENIRFTXELEKDGCLPFLD			
			TGAGA	TGCGTGG	VLVKRKSDGSVGHSVYRKPTHTDSYLHSSSCH			
			GAGAA	AATCCCG	HPSQTMSVLKTLVHRAHTVSDAESLPKELAHL			
			TTAGT	TCATCTC	RTVFRDNGYSTRQINRAFSKTRNREVDKEEN			
			TATAC	CGAGATT	APAKSLAFLPFVGNVSFKIGRILNNYQVKVIFR			
			AGGCA	TGTTCTA	PPSKISTCWDQDFDLLGSVKDDLRLRKAGIYKI			
			AAGGT	GACGAAG	PCECGMAYIGQTTRTWKNVAQNINVALAFCN			
			TGGAT	ACGCCAG	PASLQLRSIVFPTDIQWSMIKHQFWPQQHPSGT			
			ACATT	AAGACTC	PL			
			TCAGG	CGATAAC				
			AACTG	TGCGGCA				
			TTCCCT	GCGGCAC				
			TGACA	TATCGAA				
			TCGAA	GACAGCT				
			GATTG	GATCCTT				
			CAGCC	GCAGTTC				
			TTTCCC	CACCAGC				
			CCCGG	GAGGGCG				
			GACTG	CTGCCGC				
			GGTTG	CGGGCGG				

				ATGGA	TGTGGTC				
				GTTAC	CTTCTGT				
				TTCGC	CACCGTG				
				CAACT	TCTTCGA				
				TTGAT	CGCATGC				
				TGGGA	GCGGTAA				
				ATTTT	TACCACC				
				GGATG	AGCGCGC				
				CAAGT	TATATAA				
				TCGGG	GGCGGAG				
				ATGTC	CGGAGAG				
				AAAAT	CGTCTTC				
				TGATC	GTCAGTC				
				GCATG	CTTGTTG				
				GTGCC	GCTCACC				
				ACAAC	TGAAGAT				
				AAGCT	GACTGGC				
				TCAAC	AGGTGTC				
				ACAGG	CAGTTGA				
				AAGTT	AATATCG				
				ACTGT	TGGAATG				
				ACAAC	AAGTTTA				
				GTTCC	CGACGAC				

				ATTTT	CGGCTGC				
				AAATT	AAACCCG				
				TCGGA	AAATCTT				
				GAGAA	TTTGAAC				
				TTTGG	A				
				ACGTT					
				GCAAC					
				TGTCT					
				GTTTT					
				GGGCA					
				AAGGA					
				CTGAA					
				GTTTT					
				GTACC					
				TACAC					
				CCAGT					
				GACTG					
				GGTTG					
				AAGAA					
				GCTGT					
				ACGAC					
				ATTCC					
				TATGG					

				CACTG CTGAA GAAAT GGGAA ACTTG CCGTG TTCTCT TGA					
Ving i	Ving i- 9_L Mi		Locusta migratoria	CA	TAATATT TCTGTCT GTTTATA GTTCGTA GTTTACA TGTATGT AAGTAAT CTCTGTA TGTTATA CCTTTAT AAACTTC TGTACAC ATGATGC TTCTGAC	VLCHIFKLFKVLHDRLSKAIDNYLIPQQAGFR PGKSTTSQVINLTQLLEDGYENRKITGVAFIDL TAAYDTVNLRLLMKKLYAITRDFKFTSIIRSLL QNRFRSVEFQGKQSRWRAQKNGLPQGSVLAP LLFNIYTNDQPLPAGTENYIYADDRAMAVQGD SFEEVEKRLTDALDALCQYYADSQKPNPSKT QVCAFHLRNKQANRKL RVSWKDTPLEHCANP KYLGVTLDRALTYKQHCINTKMKVSGRNNIL RKITNSSWGAKPEVVRTTALTVCFSAGEYASS VWYNSAHAKQVDIALNETCRLVTGCLKPTPID KLYHLAGIAPPAVRREAAAYSERARAELSESH LLHNIQPATRRLKSRRSFLRATDNLSDQPGSRV DRWKRSRERHWMPEKEELAPGHNEDWTVWK SLNRLRTNTARSRDNL MKWGYSTVSNLCDCG	cd0165 0	1	232

					ACGAATA AAAAAA	DVQTTQHMYNCPECPSHCTLEDLMLATPNAV DVARFWAKCI			
RTE	RTE - 24_ LMi		Locusta migratoria	GTCTC CCACA AGGCC AACCC CTCGT CCACG TCGTG GGAGA CGGGC AACGG CGCGA AGTAA GGGAT TGTCT GTAAG GGTGA AGTAC AGTGG AGACT ATGCG AGTTC	TGATGAT GATGATG A	MDRRKVKILGLSETKWKGSGKKTLRNGYILH WCGSQTETRNGVALLLHKDVAIHTDIRYVSDR IIQMKLHVDNTYITIFQVYAPQTGCTADEKEQF LENLEDLIVEDNVIIMGDLNAQVGKERQGCEN IIGPHGYGKRNPGEQLLDLCVRNKLIIKNTWY QKRESHKITRYSWDGQHKTVIDYIITDQKASQ TVKVIPSECMDSDHRLLTADLKNVRVPLVKYK KMPKIKTWLLRDPEKKQEYQKKIKDRIPKTEK LGVEEEWKLFKDTLVTEAVHICGKTSAAQREK ETPWWNDRVKTAIKERNVLKKKLDREKQNRV DVRNDTEIERLSQEYRNKKLAVKRLIQNEKDK CWDDFTTKLQEDCQGNKKLLYKLVNDKRYE RTNIVVLETEEGKLLRTESEIRSEMRNHFNSLL NGNVDSSDDSDCPAEETPEVLQPLTWLEVESA LRNMRNGKSPGFDELSSDMMKAAGTPGLQWI YRVLSVIWTNNKIPDDWTKGIIPLFKKGDRKK CGNYRGITLLSHTLKLLEKIIETRLRKIVEPLLE EEQHGFGRKGRSTIDLIFAVRILAEKYWEHGKN LVMVFLDLEKAYDSVPRNKIWKCLEDLKVPN YLIDRVKMLYQTCNSSVQIGNGRSEWFETKRG VQQGSALSPLLFIIVMDRMLKSIKETANESRSL	cd0165 0	497	746

			CGCGA	VFADDVVIWGETEKEVQKSLTVWNNKFKEFN		
			CCGCT	LKISKTKTVEMTINREGTSSNIYLEGTKIKSVSH		
			ACGGC	FQYLGSIKSKDNTIRQEILNRTQKASNFYMQVR		
			AGCTG	NLLWDKKVPQAAKLTLYNTYFTPILYGMETC		
			TGAAG	TLLKKDYSRIQATEMRFIRTMNQSTRRDKIRNE		
			GCCTT	VHRETAGIRVPIANILGKRRLQWYGHVMRMD		
			ACAGG	STRTAKKYFRMEVGGRRPRGRPRRRWIEAVK		
			AACCC	SDIEKNGFNWDDVQDQMLYQDRMRWRALVS		
			TGAAA	HTRETGDGK		
			AGCGG			
			CGGCT			
			AACGG			
			GGCTC			
			TGGTG			
			AAGCC			
			GTGTT			
			AGGCG			
			TAGCT			
			GGCCA			
			GCGGT			
			GGATT			
			AAAGC			
			GTGCT			

				TTCAA					
				GACTA					
				AAACA					
				AGCCT					
				CGGAT					
				GTATA					
				AAAGA					
				AATTA					
				CTGGA					
				CATGA					
				CACCT					
				CTCCA					
				GGGAC					
				CTCAC					
				GAAAG					
				TATGA					
				CGGCC					
				AAAAA					
				CCTGA					
				AGCAT					
				ATTGG					
				AGACS					
				CCTGA					

				AGACA CAAAA ACCAA CAAAG AAACA AAAAC TCAGA CTTGG AACTC TAAAC ATCCT CACGC TCACT GGTAA ATGTG AAGAG CTTGTT GACCT A					
Pene lope	Pene lope- 54_ LMi		Locusta migratoria	CCAGG CACAA CATT CCAGA	TGATGAT GATGGCA CGTTACG CTATCGA	MGIPLRPIVSTIGAPTYQLAKHLTSVLSPQVGK CEHHIKNSAQFVERIRQLRLSNSDLMVSFDVVS LFTRVPLKESLELIEKKFGPSFTKVFNFLTSTY FLFDGEYYEQTDGVPMGSPLSPVIANMYMEEF	cd0030 4	114	198

			CGTGA	AAATTCG	EDKALNTAPLKPTIFYRYVDDTFVIWPHGEDE			
			AAGGG	TTCCACA	LQVFLQHLNSIHSQIQFTMEVEQNGQLPFLDVL			
			CGCGC	ACGATCA	VRRKPDGTLGHSVYRKETHTDQYLHASSCHH			
			TGCGG	ATCCATC	PSQREGVLTTLVHRARVISDEESLKPELDHLKE			
			TCGCT	CGGCTGC	VFQENGYSRQIRRAFCHKSTPQPTPEEDPPAK			
			GCGTG	ACGCCCCG	IAYIPYAGNVSGKIGRILRRQNIKC VYRPPAKIK			
			ATGAT	AGAGATC	SLLGTVKDDLGLRKSGVYHIPCECGKAYIGQTI			
			CCAGA	TTCACCTC	RTIEERCTEHKRHTELQQPTKSAIAEHCLETXH			
			CGTCG	A	KMKYEDTEVLHNTSRYWDCLYKEAIEIRLREN			
			TCGTC		LINKDSGFQLSHAWNPALKFLRSRRKQNTSHS			
			CTGCC		EGRAQTEVRHSSPEQELPPPPQHPGVRTRGQW			
			GGCGG		RGVTAGIYSTAAGPAPQSAGVT			
			ATAAG					
			GGGAA					
			CATAA					
			CCGTT					
			CTGCT					
			AACAG					
			CTGCG					
			GAATA					
			CAAGG					
			AGAAA					
			ATGTA					

				TAAGC					
				TCCTA					
				GATGA					
				CCCTG					
				CTTAC					
				AGGAA					
				GATCA					
				CGACG					
				GATCC					
				AACCA					
				GCCGA					
				ATACA					
				GCGCA					
				AGACA					
				ATTGA					
				ACTTC					
				TTCGG					
				AGTAC					
				TTTCA					
				CAGAC					
				GAAAC					
				AGCTA					
				AGAAA					

				CTCCG TCCTC GTGCG GCCGC ACCAC CAAGA CTCTA GGGCT TCCAA AGGTA CACAA G					
Pene lope	Pene lope- 40_ LMi		Locusta migratoria	AACAC CTACA TACCA AGAAA CAAGA ATAAA TGTTG AACTT TACAA AATTC ATCTG	TAAAACA CACACAC ACACACA CTAAAAC GAAAGCA ACATGAT TCAAAAC TCCCGCC CCCACCA CAGTCAG CTGTTCTT	MSGTKSAEEKTCRRIRNKLKNEGALITKADKG TPWSGCPHHKGRQGNTMVIMKQDDYRNKTL TFIEDNNIHKLTSDPTQRFHKNTQTLLKSITHIF PKQKIKWLKQKNPKAPTLNCLPKIHKEHIPVRP VVNFRNAPTYYLAREIHSYLQKHVFTNNRSI LNTSELIQKIKHIIIPSTATLTSFDITSMYTNIPITE TINIIKQQLQQQTDTPKSHIKETTNILKLITEQN YFTFDNQFYQSQEDGLPMGSPVSGLLANIFLNHI ENKIFDTIVKPKQYKIIYWHRYVDDIICLLDETP SRIQELHNDINNVHPKIKFTIETENDRRLNFLDL SIHRDKNQHKFSIYRKPTTTSTAIHKHSNNPITQ	cd0030 4	247	335

			GAAC	TAACTA	KQAGFQYMLHRLNKTPLSDEQYKNELQTIKQI			
			AGGCA	CTCCCAG	ATENGYDTNIIDKLNHKKAKLKGTTNTQKQT			
			ACCAC	AATTCAA	DQKHNTSTQNNPLNSHTPNHTTQEQKSNKTT			
			ATCTC	CTGCCCA	NKWHTLTNDKATHRIGNILKKQGLKVAVRT			
			TCAGA	CCATCTT	DNSIQRKLKTTNTNTDKHNTSGIYQLTCQDCN			
			CATAC	GTTTTCA	SVYIGQTCRNFNTRYTEHRRALKSNSHSTFAE			
			TCAAT	TAACATA	HLINMNHKPTDINTDLKILKHSKSLPQKLTIEE			
			GATTT	AACATTG	NYHIQKAIERKEIINEYTALCNGTIFSALKELTE			
			AATGA	AGTGAAA	NP			
			ATACA	GTGAAAG				
			GTCAA	TGACACA				
			ACAAA	GTGCCTG				
			ATACA	TCCACAA				
			GAACA	CACGAAA				
			ACTAA	ACAGCGC				
			TCAAA	ATATCCA				
			ATGGG	CCAAAGT				
			AGAAA	AAGTGAC				
			CAACA	AAAAACT				
			ACAAA	TGTTATA				
			AAATA	AACACAT				
			AACAC	AGACATT				
			ACTCC	TTGTTAT				

				TCAAA	AGAAGCA				
				CATAG	GTAGCTA				
				CAACA	CAGTAAA				
				CTGTT	ATCAAAC				
				AAAGA	CAAAACA				
				TAGAA	CAAAAAC				
				ACACA	TACGTCC				
				CAACC	TGTATTTT				
				AAACC	CGTAAAA				
				ACACT	ATCCCGA				
				TTCAC	TATTTAC				
				AAACG	AAGCTAA				
				CTAGT	CCATTCA				
				CAACC	CCCCATT				
				TGACT	AAATACA				
				GACAT	CACGTTT				
				ACAAT	TGGTTTG				
				TCACC	CAGATGA				
				AAACA	CAAAC TG				
				GGAAA	TCAAAAA				
				CAAAC	CTMTCAG				
				CTCTT	TGACGTT				
				AGAAA	TTGTGAG				

				AAGGA	TTGAAGA				
				CCCAA	ACCACAA				
				ATATA	ACACAAC				
				ATGTC	ACATGTG				
				AACAC	GTATTAA				
				ACATA	GGTAGGC				
				TTACA	CTAGCAT				
				CACAG	ACCTATA				
				AACAA	AAAATAA				
				TAGAG	GAAACGA				
				AACCT	TTTATATT				
				TGTCA	AGTCTCA				
				CAGAA	CAATTAT				
				AGTGA	CTGTAGT				
				ATTCA	TTTCACA				
				TAATA	GTTAATG				
				AAACA	TACAAAA				
				ACAAG	AGTAACA				
				AAAGA	ATCATTC				
				AACAC	CATTGTT				
				AAATT	AATTTGA				
				TCAAT	CAGACAA				
				GCAGG	TAAACAT				

				CCTCA CAAGA GAATT AATTG CAGAG GAAAT TAGAT GCATA ATAAA AGAAA ACAGA AATGA CATA	CCCACTG ATGATGC TGCAACG CCAGCGA AACATGT ATGGGAG AAACTAA AAAACAT TTGTGTTT GCTCAAG GCGGACC ATACCTA TAAAGCA AATTTAT TTGTAAG CAAACAC G				
RTE	RTE - 37_ LMi		Locusta migratoria	TATTG GTGGA TTGAG TAACC CTGAA GTGCT	TAAAAAA TTCAGGC ATTTGCA GGATTTG AGGACAG AACTAGA	MMASSLKGSTTPGGNQSTIVKDNVSEPTDSGE LTPTCKRLKSKKVSYFATFNVNSLLKIGKLKH LTDLTNQNIAITALQETRYTDEAAFESEGFRX LKGKPGKRVMKNVPQLGTGFIINRQILGAITDF SSKSSRLSTLTFRGANKMYTMVNVHAPTNETN KTDPEKVEEFWDXDLEETLSKIPTKNVILLGDF	cd0165 0	556	814

			GCTGC	CCGAGAA	NAQLGKERKYRNIIGNYPAHNRTNXNGERLVE		
			CGGTC	CGAAATC	ICKTFDLIMKSTAFKRLPRKQMTWVCQILCKT		
			CCAAC	ATCGTGG	DDMGMPNPLLGEFQLDHVAIPKRAGKEIQNV		
			CCGGA	ACAGAAG	KVLRGANLDSDFLXKIKIKIIPKVHQPRNTIRT		
			TAAAG	AGCGAAG	VKYDTGKLKTSTDFKTSLSQKDYQGWEQLKD		
			GAGGA	AAAGATG	AAIEKRRKAWADWCSHKTLENLNKFQEVKRN		
			GGGTT	CAAAGTA	TAKTIRNLKRKYEQARXTQIEEDFQKNNTRDF		
			GGCAG	TGAGAAT	YRTFKTNLNRYDPPTLSFKNKEGKIVYGEMEN		
			CCAAA	GAAAGAA	SQILAEYFEELLNCEAPADTFSYNDTEPNPDSE		
			GAAGC	TACTGGA	PPTTEEIKNIINTLKNNKAPGEDGIIAELFKYAN		
			CGTAA	AGAGGAG	ENVIIELKQIISKIWEEEKLPEGWTSALVHPIHK		
			AAGAA	GAAGAAC	KRRYIPFTKKGDKTDPGNYRGISLIPVAYKVFS		
			AAGCC	GGAAGAA	KALLNRVEKQLNPQIDEYQGGFRKGRSCAEQI		
			GGGTC	AGGAAGC	LNLKMIMAYQKQRCKKFVLTFVDFKKAYDSV		
			ACATC	TGATGTG	DRQTLFQILKEMKLDKSSSELVRQTLNNTTSK		
			TCAAG	AATTAAC	VKFRGKLSESFKINTGVRQGDGLSPMLFNCVL		
			GCACC	GTGGTCC	DKVIREWRQGMKGVGGIRLGCKKKGIRVDCL		
			TCTGA	AAAGATG	AFADDIVLLSESLEEAVNQTKELQRQAAKTGL		
			GGCTA	GCCGAAA	QISIEKTKYMTNITEAPGWMAIDRGKIQKVEKF		
			ACTAC	CGAAAGA	KYLGEWIDTGLTEKDAMEARINKINLAYKLT		
			CCCAG	AGAAG	MNIYNKKNISINAKLKHYQTVIRPEALYAAETL		
			TTGTA		NLIRTGVIKRLELVERKILRRILGPRRTEDGQFR		
			ATCAC		NRSNQELYSKIEKISNIIKKRRCEFFGHICRMDQ		

				GTTGA CGCCT AGGCG TCAAC CSCCA ATATA AATCA TTCGTT TCAGC		GRLTQQISQFLLYKKSKIRWIQEVQRDLEEMGI RDQEIHN RVIFK			
Dap hne	Dap hne- 29_ LMi		Locusta migratoria	GAAAC CGTTG CAACA AATCG AAAAT TTAGG ACGAT TTGTA GCAGT TTTGT AAACA CACAT TCATT GAAGC	TAGTTGC CCACCCA TTCTACA ATTTAGA AGAATTT CTCAACT GTAATAT AATATTG TAATATT ACTGACA AGCTGTT ATCTCTTT TAAATTA ACAATTT	MSQSPVASSVAVVGNKAAVTVPLREDSPVASS VAAVGNKAAVPVTVPPQZASTATASTAPPNEE AAEKKKRSSREEEVVSSVAGRRKMVPPLRLK DFLVEGSKMKNLEISDLKHPIAPHQNVHSGN SSSELKFATNLLPEHQGIKDKEFFVNDFGLDNN FDIFCFSEHWATENELTLVLDTYNIANSYCRKL HLRGGVAIYTNQSLSYSFTKLDLDYLCEEQNF EAAGIFIVSLKLVIISLYRSPKGIISVFLEKLEML LNVFRESKWLHYDIVIGGDLNADFDVTTNKG VSLNCNDKQRSVTELKNLLRQHNLHSINSRPT RDKACLDNIFVNFKTAEDSCMNCRRMLMYVTV FPFSDHDAVSVNYISRFPDKCDIPVPKVVITRP ITDDKISRFRMALADREWFDQCYGDTHCSLSS DLSAKVVFD RFFKTFLTLFNDSIPLKNAKQHSP	cd0165 0	665	908

				AACCA	ATGTTAT	KKCKVIDKGFKTKVTNRQNPWYTEQLTNLKH			
				GTTCT	CTATGTG	QLMVLYNISNALKTEHARSAYVKCRNEYKKA			
				GCCAG	ACGTTGT	IVEAKKAHNFRSIENSTNKCQAAWKLINSVAK			
				GAGCA	CTATTGC	DPRNKSVMNVSPQKFNDYFIKSVEEVGNAIKPD			
				CTTTA	TGTAATG	IRSSELLSKNVSRRSTTNINVFTFSEVSPSLILKIV			
				CCGGA	GCCTAAT	KQLKSSDSADIYDISCNMLKRVIDCIVYPLTFCI			
				CATGG	GACAATA	NKCISEGYFPDELKVSRRVVPVYKKGDTESSSY			
				ACTGC	AAATTAA	RPISIVPAFSKVLECVIYQQLSVYFENLGLISES			
				ACAGA	TTATTATT	QYGFRKNLSTVDAIDSVMKYLHQVFEDKGFA			
				AACCA	ATT	QVTFCDLSKAFDCVEHTPLLDKLEFYGIRDISL			
				CCTGG		KLLKSYLQNRQVVCIGKEMSSTEQVKVGVP			
				GCAAG		QGSVLGPFLFLIMINDLPSFIKSSTVLYADDTTF			
				GATGT		LHSSNDXSSLKTCVAKTLDQASYWFKANGFL			
				GTTGG		LNETKTQQMVF TLKDKLP SDDPSYVKFLGVYL			
				CCAAA		DDKLSWGQHVKYISGKLSRVIYLLRRLMDYV			
				GAAAT		PEKYVRITYFSFFQSIITYGILWGNSSCVQDILI			
				TCTCC		LQKKAIRVITGSPHKEHCKPLFCKLKVMTVINL			
				ATGTG		YIYYVLIYTKRKL PETKLRENVHYYNTRRNKC			
				ATAAA		IFTSYHRLSKSLNSYELVGINY LISCHNMYKIYP			
				AGCTG		NKHLNKRRCMTG			
				CAGGA					
				ATGAA					
				CAGTC					

				ACGTA					
				AAAAT					
				TGCAC					
				TGCCT					
				GCAAT					
				ATGTT					
				CATTG					
				CCAAT					
				TACAA					
				AMGAA					
				AAGAA					
				GACGG					
				TACCA					
				ACAGC					
				ATATG					
				AAGAA					
				GAAAC					
				ACCTA					
				TGGCT					
				WCATC					
				TAATA					
				AGTAG					
				AAAGT					

				ACATC AGCAC CAGCA CATGA A					
Ving i	Ving i- 49_ LMi		Locusta migratoria	GGGGT ACGAA GAAGG AGCCT CCTAG TAGAC TGTGA ACCAA TAAGC TGAAG GACAC AGTCG GGCGT ACCCT GGAAA CGGTG AAAAA CTTCA	TGATCTG GACACGG AAATGAA TGAATGA	MQAIAEDKRSRWCNLVENLEEITVVQSCREPR MKQNSMEIHKLEWDPTDHKQKNNGVTANQI ASQLLKNGKTGGKKKGKELYPKNKKMMASSI HIHSNRTGRRHSEYETEWPPQYTFTATELEDAI QSMKLNKAAGTDDLRTQIKHFGPKVRNWLL QMINACIKIPKLWRQARVVALLKPGKEIRILKV TDQCHFYAISTRCSWGDPKTMAAGEGGGLTK TWQGDTPKSYRPVSLLCHLYKMLERMILNR VAQIIDGKIIKEQAGFRPGRSCCGQILNLTQHIE DGFERKLITGAAFIDLTAAAYDTVNHRKLLCKL YHLTGDCRLTMVIGSLLQNRRFYVSLNGKNSR WRLQKNGLPQGSVLSPILYNVYTNDQPIQDT RLFVYADDTAVAVQGSSFDEV SQKHPVVLKN LLNTTMPITSIPKTSSGLEELAQYYDANHLKPN PCKTQVCAFHLRNRDAGVDLDVTWQGKKLD TVQHLNTLG Y TIPWGTLDR TLSFKQHCRNTKA KVCSRNNIIRRLTNYEWGAQPSVLRTSALALC VSAAEYAAPVWDASAHTKQVDVSVNETMRIV	cd0165 0	226	489

			CCGGC	TGCLRPTPTGNMYLLAGIAPPRIRRQVAADAE		
			TAATC	RTKKENDYRHPLNGHVTQARRLKSRNSFIART		
			CTTCT	KILDGSQEDNRVCKWKNEQTGLPITPKEEMAP		
			AACAC	GNKLPYTVWRTLNRMRAGVPRCKTNLRRWG		
			CAGAT	LLTNDDDALCECGEVQDEAHLLTCRLLPEPCS		
			CAAGT	FSDLLQANDKAVAVANYWKNKT		
			TAGAG			
			CCAAA			
			AGGAA			
			CATCA			
			AAGAT			
			GCAAC			
			GGCCT			
			TTATC			
			AAGAC			
			CAACC			
			CTCGT			
			GATAA			
			TCTCT			
			ATAAA			
			TATTG			
			AGGGA			
			TTTTCT			

				AGTCA					
				TAAGA					
				GTATT					
				TTATT					
				AGCAG					
				ACATA					
				TGCAA					
				TAAGC					
				ATAAA					
				TGTGA					
				TGTAT					
				TATTG					
				GTGCA					
				GGAAA					
				CACAT					
				CGGGA					
				AACCA					
				CCAGC					
				CACAG					
				ACCTA					
				AAATT					
				GCAGG					
				AATGG					

				AGTTA					
				GTCAT					
				CGAAA					
				GACCT					
				CATGA					
				CAAGT					
				ACGGG					
				AGTGC					
				GATTT					
				TCGTT					
				AGACC					
				TGGGC					
				TCAAT					
				GTAAT					
				ATCAG					
				CTGCA					
				TTGAC					
				AGAGA					
				TAAAT					
				AACAT					
				TGAAG					
				TACTT					
				ACCAT					

				CCGGG					
				CCCAG					
				AAGTT					
				CTCTG					
				TGACG					
				ACTGT					
				GTACA					
				AACCA					
				CCAGA					
				TATGG					
				ACTTT					
				GTTTTTC					
				AATGA					
				ACCTA					
				ACAAT					
				TTTAC					
				ATATG					
				ACCAT					
				ATTAA					
				TTTTGT					
				AATGG					
				GCGAC					
				TTCAA					

				CTGCC					
				ATAGT					
				GAAGC					
				GTGGG					
				GATAC					
				AATGA					
				GACTG					
				ATGAA					
				AACGG					
				GATAC					
				AGCTT					
				GAATC					
				CTGGG					
				CTGAC					
				AGTAA					
				TGACC					
				TGCAA					
				CTTATT					
				CATGA					
				CCCAA					
				AACTC					
				CCAGC					
				ATCGT					

				TCAAT					
				AGCGG					
				AAGAT					
				GGAGG					
				AGGGG					
				ATATA					
				ATCCT					
				GACAA					
				CATCT					
				TTGTG					
				AGTAG					
				GAAGT					
				TGGCA					
				ATGCA					
				GTGTG					
				TGAAG					
				CTTGTT					
				GATCA					
				TCCCA					
				TCCCA					
				AGTAC					
				CCAAC					
				ACAGA					

				CCCAT					
				TATTT					
				GTGTT					
				GTGAA					
				TGCCA					
				TCGTG					
				AAGCC					
				AGATG					
				AGGTG					
				CCCTTT					
				CGAAG					
				GCGGT					
				TTAAT					
				TTCAA					
				AAAAG					
				CAAAC					
				TGGCA					
				ACAAT					
				TTCAA					
				GAGCA					
				ATTGG					
				ATGAA					
				GAGAT					

				TGAGA					
				AAATT					
				CCTCC					
				TCGAC					
				CAGAT					
				GAGTA					
				TGGTA					
				GATTT					
				ATCGA					
				TCTTTG					
				AAAGA					
				GGTTT					
				CAAGG					
				AAAAA					
				TATCC					
				CAAGA					
				GTTGC					
				CACAC					
				CCGTA					
				CATCC					
				AAGGA					
				CTGAC					
				TGGCA					

				GCAGT AAAGC CTTGC TTGAC AGGTA CCAGA AGTTG TTCAA TCAGG ACCCT TTCAG CGAAG ATACG TTGGA GGCTG CGGAA GACCT A					
Nim b	Nim b- 20_ LMi		Locusta migratoria	CCCCT CAGGG GCTCA GCATT AATTT	TAATGGT GGGTTCG TATGCCT GATCTTC AGGCATA	MALPGAKPASLSVXGPNLMLRHMTPRRSLPW PRHGRNAGREIEVRLMPPXTWYVPGLTANRLP PRSPCFLWKTLRISGRWMACQDEIWXCXHX DVVPSPTGTACLVRGGGWHVKMRSGXVLIT TSSPAQSRALLACDKLGDVPVSITPHKSLNMV	cd0165 0	1160	1418

			GCTGA	CACCGCT	QGIIFHKDLLLQXDDELRKNLERRGVHFIRRVH		
			GTACG	TATAGTT	RGPKDNRVATGAFILAFEGDSLPEKVKVMIYR		
			GGCTT	TTAATCT	CDVKPYVPPPMRCFKCWRFGHMSSRCXGNPT		
			GGCGA	GTTGTCA	CRECGRAXHPDAPCAPPPVCVNCGQHHSPCSP		
			ACCCG	CTTTTGTC	XCTVLHKERKIMEYKTLDRLSYSEAKRKYXK		
			GGGTC	TTGTCAG	LHPVSMKTSMPQLHHTASAISATALHCLVFHS		
			CTGAG	TCCTATTT	GPPGXTAHJCPHGLWGHLNLHRPQGPCFGSTT		
			CCGTG	CCGACTG	ISTVGXSVPTPEPESQQPXLAPXARKGSLGSLP		
			GGCTG	GCAGTTT	SKASPADASGWLKXPPTAGRRASRASSVPGDG		
			TGTCG	TAGTATG	SAEPSQHEKPKEXREKKKKKNKSKTKDSL VAT		
			CGCCA	TGCAGAA	VPPXHSTSSSEDEIEILASPKDLAIAESSEAMDV		
			ACAGT	TGGCTGG	EALPQKSATAEDTVXPTAPSKSACPSDQEALG		
			CCGCT	CTCATCC	SASKIGDSRRYCXPYCAVEISMSQRPGGSLILQ		
			ACACC	ATTATTA	WNCGGFFRHLPELQQLLCRFP AFCVALQETWF		
			GTAAG	TCCTAGT	PAQRTPALRGYRGYYKNRTIYDRVSGGVCIYV		
			CTCTG	GATCAGC	LATVPSAAVPLHTTLEAVAVRIRAMQELTICSL		
			TGCGT	CAGCCAA	YLPPDEMVS LTDLAALVAQLPPPFLLLGDFNA		
			GCTTC	CAACATC	HHPLWGGAQTTGRGRDIETLLSQQDICLLNTG		
			AACGA	TGCTGTG	APTHFCVARGTYS AIDL SVCTPSLLPHLSWGGGS		
			CCATC	ATTTTWA	TMTFAGVTISPFFCHSSTWTSGSHPMVSIAAD		
			GTTTCG	GCATAGT	WRGYTSAAVVSVPLPPDIDTAVQQVTA AIVSA		
			GCGCG	TGATTAT	AEETVPITTGSPRRKVVPWWTPEIAEAIRDRRR		
			ACGGT	TTGTTCC	RCNATGAAPQSKIGHPGPSAALQRHRRRRPTIEN		

			GGAAC	ACCTCTA	LIAFKRLRARARRLIRQSKQDCWERYVASIGSR		
			GTTGT	TTGTGGT	ASPSRVWSQVRRFLGYRPPAGIPSLTVDGTLCT		
			ACGTT	TTCCTTTT	DPAVIAECMAEHYASCSATANYSASFRLKQR		
			TCAGA	TCTTTTTA	VERNPLSFHTRRRELYNARLRSGSLTLLSLHAP		
			GCCCG	TGTTGTA	IPLRDQTAYTASSSVTCQRTFITNCSPYLIGSGQ		
			GGGAT	ATTTTCG	GVFSHPSGGIVLLSLSTTGGAKSSIFPHSPPGTV		
			CTTGG	CCCACAC	QCSFTEWEFNTALAACPDTSPPGPDRIHGQFVR		
			CTTAA	TGTCTTC	HLSADFHHKLLALFNRVWTGGVFPSQWRDSIII		
			TCGCC	CTGTACA	PVMKPGADPLAVDSYRPISLTNVLCKLLERMV		
			CGGGT	TCTTGTG	SRRLMWILEDRLAPQQGGFRQGRSTIDNLV		
			CGCGA	TACTGGT	SLESAIRTAFSRCEHLVAVFFDLRRAYDTTWR		
			GGACG	TGCGGGG	HHILATLYDWGLRGPLPIFIKNFLSLRTFRVRV		
			GTACA	TGCTTAT	GTSLSXXRVLENGVPQGSVLSVPLFLIAINGLA		
			AACCT	TTTTAAC	AAVGSSVCCSLFADDFCLYYSSPSIGVAERRLQ		
			CTATA	CGTAAAT	GAIRRAQSWATAHGFTFSASKTRVMHFCRRR		
			AAAAA	GAAATAC	AVHPDPELYLAADQLHVAETHRFLGVLFDR		
			ACCTC	TGTGATC	LTWGPFLRQLKQKCWRHLNILRCLSHTAWGA		
			AATCT	TCTTAGC	DRRTLMHLFKALIQSRLDYGSQVYDSASPSAL		
			CAAGG	TGGGAAC	RILDPI LHGGVRLATGAFTSPVPSLLAEAGIPP		
			TGTGC	AGGGGAC	LRYYRQQLASYIAHVHATPHHPNRRLLPNT		
			TACGT	TGATGAC	DLPSQRRPRSGLPVAVRVHSLSSLESFPLRP		
			GCCGA	CCAGTAG	AFSCLSHIPPWTIPRLQLHLQLSCDPKASVSPQL		
			GGAGA	TTTAGTC	LRHQFLDILGTFRDADVIYTDGSQVDGRVGYA		

				TGCAT	CCCTTTA	YVHGDHTEQRS	LPPGCTVFTAELIAITRALDHL			
				GGCTG	AACACCC	RACPRKSYLICSDSLSALHALDQCFPKHHLVRS				
				TTGAG	AACACCA	IQVTVFSLGKSGRSVAFIWTPGHIGIPGNERAD				
				GTTGA	ACCAACC	QLAKQAIACQRLDIGTPEGDIRLGLHRKIFEAW				
				ACAGT	A	NSEWATMSTPNKLRAIKPSTAAWKTSLRPSRR				
				CGCTA		DSVVLCLRIGHTWLTHGHLLRREEPPECRCG				
				GCGAG		ALLTVVHILLDCPLLAALRQSFNLPTALPSILA				
				CAACC		DDASMADCVLGFIRAAGFYHKI				
				TCTGA						
				GGAAC						
				CTGCC						
				GCACC						
				TCAGT						
				TGTAT						
				AAGGC						
				TTACC						
				CAGGC						
				AAGCG						
				GGGCT						
				CTGCC						
				TGGGT						
				GGACC						
				TTTAG						

				TTCCCT AGCTG CTCGT GGGAC GAGAA TGAAA CGAGC AAATA					
Pene lope	Pene lope- 1_L G		Lottia gigantea	TTAAG TTTAA CCACC TCACC CTCAT CTCCA GCCTG TAAAC CACGA CGTCA TAATA ATGAG AAGGA TAAAT CCAAA	TGACCCC CCTCTCT ATTTTCT AGCTCAT TAAGTAT AATATAT TCTAATC CTTATAT ATAGATG TTTATCTT TTCAGAA CCTATWA ATACTCT TTTCTTTT AGAACTG	MGLPFKLNAGGNALKLGETVVNLSSRKLSPLE ERVLDRLTFIPTSKTLSHEPILSSVEHVNRTFK LNFFFRNRLGRGKRIPFIEKSGWTPPDSAIDSDI LATCDKLHEEISHLPLVNEKHNLTKFERKALN SLKNDQNIIRKSDKGSCCVIQNREDYIFELEKQ LNNANHYRVLDQPIFMETAQKINNILKTLVNK QFISDSQYDYLSAKPEVRQRHIYTLPKIHKNP TEWSLANRIPKGRPIISDCASESYRVSEYIDHFL LPLACKHPSYKDDTTDFVEKIKSLNVPPNARFI SLDVTSMYTNIDNKTGLDAVRKAFLRNP RPDKEILELLELCLTRNDFAFNEKIYLTSGTA MGKRFAPSYANIFMAEWELEISSISPFEPFFYGR FLDDIFIWTESLTNFWQYFEILNNHHPVSKLTA NISEQSMDFLDTTVFKGPSFSQTNKLDIKVFFK PTDTHQLLYKNSFHPKHTFTGIIKSQILRFFRIC	IPR000 477	204	450

			TGGTC	GTAACAA	TQKVDFDAACRLVFSKLRERGYSKRFLRKIKN			
			TATTC	GAAAACT	DTVLDLLKINEPPPPINNFGAIPCRGSRCRTCEH			
			CCACA	TTCATAT	IHMSHYFQNKKNKGKIFEIKERLTCTDKNIY LIS			
			CTCAC	TATAGAC	CKKCDKQYVGETQNALRTRFSGHRFDIIH SKN			
			CAAAC	TTTTGTTT	RPISHHFNLPDHSIEDLLIMPIESIPDSDLIKPKT			
			CAACA	ATTGTAG	KRLQRESYWMDQLQTVAPNGLNIQH DNQMIP			
			GTTAT	ATTACTC	FVATFNHTAAKTSSIVKQYYSELQNRFPNVYT			
			CATTG	GTGGTTG	KRLVIAYKRKNKLSNFLVRGR LDP IESNNT			
			GATCT	TTTTTTTA				
			TCCAA	ACTTAAT				
			CATCA	ACTAACT				
			GCAAC	ATGCTAA				
			ATCCC	CTTCACC				
			GTTCC	TTATCTT				
			TACAT	GTATTGA				
			CCAGA	ACATTTA				
			TACCC	TCTCTTTA				
			AAGCC	TTTGTGA				
			GAATC	CCTTGGC				
			TTACC	CAGTGCT				
			CTGGT	TTACATA				
			GCCCG	GAAGTTT				
			AATTC	TTAGGAA				

				AGCAT	CTTCTTC				
				ATCAA	ACTCTTA				
				TACAT	ACACCAA				
				TACTC	TAACCA				
				TCCAA					
				GATCT					
				CTCCA					
				ACAAC					
				CATTC					
				CGAAA					
				AAGGT					
				GTTAT					
				TTCAT					
				GTTGG					
				CCTAA					
				ATAAT					
				AAACA					
				GGAAT					
				CAGGA					
				GAATC					
				CATAC					
				ACAGA					
				AGAAT					

				GCTGG					
				AACTC					
				CTCAA					
				ACTTG					
				CCAAG					
				TCGAA					
				ATTTC					
				CGTCT					
				GCTAC					
				CCTTT					
				ACGCC					
				ACACA					
				GATTC					
				CTTAC					
				TCAAA					
				GGAGC					
				TACAA					
				TTAGC					
				TGAAC					
				CTCAT					
				TTAGT					
				CAATA					
				ACCTC					

				CAGAA					
				GTTCA					
				ACAGT					
				CTATT					
				ATATC					
				ATAAA					
				CCGTC					
				TACCA					
				ATGCC					
				CATAC					
				TATTC					
				CTCGC					
				TTCAA					
				TGGAG					
				TGGTC					
				AGTGT					
				CATTT					
				CGGAT					
				AACAT					
				CCACT					
				GGATG					
				TCCCC					
				CACTG					

				CAAAA AGTAT TTATTT TTC					
RTE	RTE 1_M T		Medicago truncatula	TAATC CTGCC AAGGC CGGTC CCAAG CCCGG ATAAA AGGAG GGTTG TGTTA GGCCT CGACA GCCAA AGTAA AACTT TGTCG AATCT CTATG ACATG	TGAGTTG GATCCAA ATTTGGT GTATGAT AGAACAC TATGGCG TCATTTG ATCCATG TAGCCGA CCCCACT TAGTGGG ATAAGGC TTGGTTG TTGTTGT	MFSCRFFLIRKSDRKICLDCKVIPGESLTTQHR VMVMDVRVKRRRAKRRSQNGAPRIKWWHLKG EKLRIQHKILEGGFRQPHGSANDMWERMAH EIKKVAKETLGESRGFGPRGKESWWNDSVQ SKVRIKRDCFKDWSRCKNVETWDKYKIARKE AKKAVSEARTQAFEGLYQSLGTKEGEKSIYKL AKGRERKTRDLDQVKCIKDEEGRVLVQERDIK GRWKKYFHNLFNEGYEILPDSNRLDIGEEDRN YNYRRIQEHEVREALKRMSSGKAVGPDNIPI EVWKS LGDRGIVWL TNLFNEIMRTKKMSDEW RRSTLIPIYKNKGDIQNCANYRGIKLMSHTMK LWERVIERRLRKETRVTDNQFGFMPGRSTMEA IYLLRRVMERYRTDKKDLHLVFIDLEKAYDRV PREILWKALEKKGVRIAYIRAIKDMYEGASTS VRTQDGTTEDFPITIGLHQGSTLSPYLFTLVLD VLTEHIQELAPRCMLFADDVVLVGESREEVNG RLETWRQALEAYGFRLSRSKTEYMECNFSGRR SRSTLEVKVGDHIPQVTRFKYLGSFVQNDGEI EADVSHRIQAGWLKWRRASGVLCDKKVPLKL	cd0165 0	317	567

				GACCA	KGKFYRTAVRPALLYGTECWAVKSQHENQVS			
				AATGC	VAEMRMLRWMSGKTRHDIRNDTIRERVGVA			
				AAAAT	PIVEKLVENRLRWFGHVERRPVDAVVRVVDQ			
				AGCGT	MEESQVKRGRGRPRKTIRETIRKGFRGQ			
				GAATG				
				CTAGG				
				TCGTT				
				GCCCG				
				GAAGC				
				AACGC				
				GTTGT				
				ATGGC				
				TCGAG				
				TACAG				
				TGTCA				
				AATGA				
				GCAAG				
				AGCCG				
				CTGCA				
				TCACC				
				ACCTG				
				GATGT				
				AGTGA				

				AAATT					
				AAGCA					
				AGGGT					
				TCCCG					
				CATCT					
				TCGTG					
				AACGG					
				GTGCG					
				GGTAA					
				AGAAG					
				CTAGT					
				TCATG					
				AGAAT					
				AGGAT					
				TCGTTT					
				TGGAA					
				CTTGA					
				AATAT					
				AGGTA					
				CACTT					
				ACAGG					
				AAAAT					
				CTATG					

				GAAGT					
				AGTAG					
				ACACA					
				ATGAC					
				TAGGA					
				GGAAG					
				ATCAA					
				TTTCAT					
				GTGCC					
				TACAA					
				GAAAC					
				TAAGC					
				GGGTG					
				GGGAA					
				AAAGG					
				CGAAA					
				GAATT					
				AGATA					
				GTTCA					
				GGATT					
				TAAGC					
				TTTGG					
				TATAC					

				TGGTG					
				AAGTT					
				AGATC					
				GAGAA					
				ATGGG					
				GTAGG					
				CATCT					
				TTGTG					
				GATAA					
				GGAGT					
				GGAAG					
				AAAGA					
				TATTG					
				TAGAT					
				GTTAA					
				AAGGA					
				TAGGA					
				GATCG					
				GATTA					
				TAGCC					
				CTAAA					
				GATTG					
				TAGTG					

				GAACA					
				AGACA					
				CCTTT					
				AATGT					
				AATTA					
				GCGCT					
				TATGC					
				ACCTC					
				AGGTA					
				GGGTT					
				AGAAG					
				CACAC					
				CTTAA					
				AGTGA					
				AGTTT					
				TGGGA					
				GGAGC					
				TAGAA					
				GGCTT					
				AATTC					
				AGGAT					
				ATCCC					
				GTTAG					

				GAGAA					
				AGATT					
				TTTCTA					
				GGAGG					
				GGATT					
				TAAAT					
				GGGCA					
				TGTAG					
				GGAGT					
				GTTTC					
				GAGAG					
				GTTTC					
				GAGGG					
				TTTGC					
				ATGGG					
				GGGTA					
				TGGTC					
				TCGGG					
				GAGAT					
				TAATG					
				CGGAG					
				GGTAA					
				ATCCA					

				TCTTA GATGT TTCAT CGGCT TTTGA TCTTA CTATA GCTAA TACTT GTTTC AGGAA AAGAG AGGAG CATCT TATCA CTTTC AAGAG TGGGG TCTC					
RTE	Avi RTE _M Un	.	Melopsitta cus undulatus	TCTCTT TCACG GACAC CCCCT	TAAAACC CTTCCCA AATCTTC GTTCGCG	MGIFCCGRLKTKRQRHTPASKGVSRLTLACW NIRTMLDTANSGRPERRSALIAHELSRLNIDIA ALSEVRLHEEGSLKEHGAGYTLYWSGKSKTES HLSGVGFMIKNSIVSKLENLPIGHSDRIISRLPL	cd0165 0	563	847

			TGTCG	AAGTCAA	HNKQHVVLFSIYAPTLQADPAEKDKFYSDLRR			
			TGGGG	GCCATG	LTQKVPIDDKIIILGDFNAQVGKNFEAWKGVL			
			AAAGA		GKHGVGSCNDNGRLLLEFCAEQQLTITNTIFQ			
			GCTTG		QKDSXKTTWMHPRSKHWHLIDYVLVRQRNV			
			CGTGC		RDVCHTRVMPSEAECQTDHRLVRCKLNFHFKP			
			CCTTG		KPKRGGIPRRRLNVNNLQIAAVRDSFQANLQT			
			TGAGG		RLEDHLIDSSPEALWQHIKNSILQSSVESFGFSL			
			TTGGG		KKNKDWFDNNQEVQELLTKKRSAHQAHLA			
			AGCTA		QPLCHVRKAAFRACSKLQQKLRLDIQNKWWL			
			TGCTG		NLAECTQLCADLGDHRGFYEALKAVYGPTYQ			
			GTGGT		VQSPLLSADGQTLLTDKISILNHWSEHFQTLFS			
			GGCAT		ANRVVQGXGIQHILQHLVKNELDIAPTMGETL			
			ATGCC		KAIQQVKTGKAAGVDGIPPEIWKHGGQALHA			
			TCTAG		KFHEL VVRCWEQGKLP SDFRDAIITLYKKKG			
			TAGGG		VKSDCSSYRGITLLSIAGKILARILLNRLVPAIA			
			TCTCC		EELLPE SQCGFRXNRSTTDMVFVLRQLQEKCR			
			CATGC		EQNKGLYVTFVDLTKAFDTVSRKGLWLILERL			
			CAGAC		GCPPKFLKMIILLHEDQQGQVRYGDALSEPFI			
			AGGTC		TXGVKQGCVLAPTLFTIFFSMMLQRAMVDLD			
			TCAAC		EQNGIYIRYRTDGSLFNLRLKAHTKTLNHLV			
			TGAAG		RELLFADDAALVAHTEAALQRLTSCFAEAAEL			
			GGTCA		FGLEVSLKKTVVLYQPAPQEDYHHPHITIGESE			
			GACAA		LKSVQQFSYLGSISSDATIDKEIDNRIAKAYRA			

				AGTGT	FGKLHKRVWCNKHLKKSTKISVYRAIVLSTLL			
				GTCCA	YGSESWVIYRHHLRLLERFHQRCLRTLNIHWT			
				TGGGC	DYVSNVTVLEQAGVTSIEAILLKLQLRWAGHV			
				AGGAT	SRMKDHRLPKIVFYGELATGCRKRSAPKKRYK			
				GGGCT	DSLKQYLSLGHIDYHQWSTLASNQETWRHTIH			
				CACTA	NAVVSFENTRRISLEEKRRRKNRAPPIPPKET			
				GCCTT	FRCAFCNRICLSHIGLFSHQACSKRG			
				TGGGC				
				AACCA				
				TCCTA				
				GGAGA				
				AGGAC				
				AACTC				
				TAACT				
				TCAAA				
				CCCAG				
				GCAGA				
				TGGAG				
				CTCGT				
				TTAGC				
				CCTGT				
				AAGGC				
				CATCC				

				ATCTA					
				AGAGA					
				AGGAT					
				ACTCT					
				AATCA					
				AACCT					
				ATGAC					
				CTGAG					
				GACTT					
				TGCTG					
				TCACC					
				GTCCA					
				AGCTC					
				GCTAG					
				GCCCT					
				GGCAG					
				ATAAA					
				CCTCA					
				GGTGT					
				AAAGG					
				GTGGG					
				GCCAG					
				TTCTG					

				CGCAT					
				GCTGT					
				GCCTC					
				ACCTA					
				AAAAA					
				TCCAC					
				TGCAC					
				AGGCC					
				CGAAG					
				GGTTT					
				ACCCA					
				CACTT					
				GCAAA					
				GCCCT					
				GTAGC					
				GACAG					
				GCGAG					
				GGACG					
				AAACG					
				GCAGG					
				TGATA					
				AGGTG					
				CACTG					

				GAAGY					
				CGCAG					
				ACCCA					
				ACCCT					
				GCATG					
				CAGGC					
				GGTTC					
				AGGAC					
				ATTGG					
				TCGCT					
				CGAGA					
				CTGAC					
				CCAGG					
				AGATG					
				ACAGT					
				CTTGTT					
				TGGCA					
				ACATC					
				CTGAA					
				TGACC					
				AAGCA					
				GCCTT					
				TTTTCA					

				AGGAG AGCAC TGCTT GCTCC ATATG GAGAG AGGCC TAGAA AAGGT GGCCT AACCA AAGCT CATCT CCCCC TCCAC CTGGT TGGCT AGCCG CGGCC A					
RTE	RTE -	.	Mesobuthu s martensii	GCCTT AGAGG AACCG	TGATTAT GATGATG ATGA	MNSSTNLRPSVPGGWNVPADGSKQVYRKAEG AKNHRV SCHGFYLATYNCRTLSTEGK LLEIEE ELKHVKWDILGLSEVRRKGEELLELKSGNLLY	cd0165 0	561	814

2_M Ma	ACTAA	YRG	TND	GRC	SGS	GFL	INK	RIK	DNII	EFN	SSSD	R	V
	GATCA	SRL	V	L	K	I	S	K	R	Y	E	F	Q
	AAACC	EEI	E	K	I	R	G	R	K	Q	Q	F	T
	TAGTG	V	G	K	F	G	E	G	I	R	N	E	R
	TCCAA	G	Q	R	L	V	E	F	A	E	S	E	N
	TCCGT	L	F	I	M	N	S	F	F	K			
	ACCTT	K	N	P	K	R	K	W	T	W	R	N	P
	CACGT	S	G	G	T	A	E	I	D	Y	I	L	A
	GGTGC	S	K	K	Y	L	I	E	D				
	AGCTG	C	T	V	L	N	K	F	N	T	G	S	D
	TTAGT	H	R	M	V	R	C	K	V	H	I	N	V
	ACTCA	K	N	E	R	K	K	M					
	ATGTT	F	K	G	P	R	S	G	I	D	I	G	K
	TCTTTA	L	K	A	N	Q	N	L	F	Q	I	E	L
	ACATG	K	N	R	F	T	E	L	T	N	L	T	E
	AGGAA	D	I	D	I	K	N	L	N	T	N	I	T
	CGGAC	I	L	I	Q	T	A	T	E	V	A	G	K
	ACGGA	Q	R	Q	V	K	N	D	K	L	S	D	K
	TGGGC	T	K	L	L	L	N	K					
	TCACA	R	E	M	K	I	N	G	N	A	V	S	K
	CCAAC	K	I	E	Y	A	E	L	C	K	T	V	R
	CCGCA	K	I	T	D	D	I	R					
	TTGGC	K	Y	N	C	N	L	V	K	K	S	I	E
		E	N	K	G	I	K	S	V	K	K	H	L
		C	C	G	R	K	R	I	F				
		A	L	K	A	K	D	G	K	I	I	K	N
		Q	E	D	I	L	K	R	V	E	E	F	Y
		S	D	L	Y	R	D	T	N				
		N	S	C	L	V	Q	N	S	K	S	E	I
		P	I	L	K	E	E	V	H	L	A	I	K
		K	M	K	K	G	K	A	P				
		G	E	D	G	I	T	V	D	I	L	H	E
		G	G	E	Q	I	E	D	I	L	T	D	L
		F	N	R	C	L	K	Q	K	D			
		V	P	E	D	W	K	N	A	T	V	I	L
		L	H	K	K	G	E	R	E	D	I	S	N
		Y	R	P	I	S	L	L	S	V			
		L	Y	K	V	F	T	R	I	L	A	D	R
		L	K	R	S	L	D	E	A	Q	P	R	E
		Q	A	G	F	R	A	G	Y				
		S	T	M	D	H	I	F	V	L	R	E	I
		V	R	T	N	E	Y	E	I	P	L	C	L
		A	F	I	D	Y	E	K	A	F			
		D	S	V	M	T	S	A	V	L	N	A	L
		A	N	T	D	T	D	S	G	Y	V	D	I
		F	S	K	I	Y	S	Q	A				
		S	A	R	I	R	L	H	K	Y	S	E	K
		I	K	I	G	K	G	V	R	Q	G	D	T
		I	S	P	K	L	F	S	A	V	L		
		E	E	V	I	K	K	M	D	W	D	K	T
		G	L	K	I	D	G	E	Y	L	S	H	L
		R	F	A	D	D	V						
		L	L	A	S	N	P	N	E	L	Q	N	M
		I	Q	S	L	A	T	E	S	K	K	V	G
		L	R	I	N	I	S	K	T	K			
		V	M	Y	N	N	F	A	D	Q	Q	K	I
		K	L	D	N	I	V	L	E	E	V	S	G
		Y	I	Y	L	G	Q	L					
		T	L	D	G	D	L	M	P	E	I	N	R
		R	I	K	L	A	W	S	A	F	G	R	N
		N	S	I	L	K	G	K	M				

				CAGCG TGGTG TGTTTT GGCTA AAACC TTCAA GCATA		PLCLKRKIFDQCILPVLTYGSETWTLTKRAIQR LQTTQRSMERCMLGITRRDRKRNTWVRSQTK VIDVIQKIKKQKWTWAGHIIRINDNRWTTKVL HWIPRDKKRPRKRPHYRWRDEIVNFTSKCWIR EAVDRKNWKLLGETFILQWIDNG			
RTE	RTE - 3_M Ma		Mesobuthu s martensii	AACTC CGCCC AAGAA GCCAT ATAGC TGACTION GGCCA GTCCC AAGCC TGGAT AAAGG AGGAG GGTTG GCGTC AGGAA GGGCA	TAGAAAT GGGAAGG AGCTTGA TGATGAT GATGA	MMSGTHDSNARAFPGVGDAQSLRRPRAVRS RLGLSHQERVRLKKLARRPLLVRASLNVGTM TGRGREIVDMMERRKIDVMCVQETRWKGDK TRELGEGYKIFYSSASKEGRNGVGIITSRKMKE NIIEVSRRNDRIMRMKLDIGETIHVVSAYAPQV GTSMEMKDTFWKTMDEELRQLEEKDRVIIGG DLNGHIGTNKRGIERIHGGHGFGRNQEGEQII DFAVAFDFALANSFFNKKEEHLVTFKSGGNST QIDYFLYRRKHLKEVKNCKVIPSDHVATQHRP MLMDLHLLRPAGREKVKTRVKVRKTKWFKL KDTVLRSRFKETALQRVSMDFEDVQEWWNR NAEVIRCTGREILGETTGKWCWENKESWWWNQ EVQAAVQEKRKAKKKWDSTHAPADLEVFKV RNKQAKKKVAQAKAKACQELYDDLETPEGQ KKIYSIAKHRSQTAKDITHIKQIKNVNGEILRNT GKIQDRWKEYFEELNVENERKYHEDGIPNEG	cd0165 0	567	814

				TCCGG CCGTA AAAAC ACCCT GCCAA GTCTTT		LTRDITKSEVKVALKQMKNGKAVGLDGIPAE AWKALGEEGLDLLWDLGKIYREEEIPDEWRE SIIVPIYKKGKDIQMCGNYRGIKLMSHTMKILE RILDKRLRAEVKIGGQQFGFMPGKSTIDAIFCL RQMMEKFREKQQELHLVFIDMEKAYDRVPRQ EVWRSRLRLNAPEKYVRLIQEMYRKTTTKVR TCLGETSSFEVGVGLHQGSALSPYLFNTVIDVL SEKVREEAPWNMLYADDVVLCAARSRAEVEEK LENWRKALEDRLKISRSTEYLACNIGTQEN EELKLMETTVRKTDHFRYLGSIIVDELSETGRE VNNRIQAGWKNWKEVSGVLCDDKIPVRLKKGK VYRTVVRPAMLYGTEALPLKKS NLQRMKVAE MRMLRWMCGVTRADRIRNEHIQGTVKVTDIA SKIQESRLRWFGHLMRRDEDYTGKRLMNIKV EGRKRGRPKRRWQDIVDDDLRSKGLRREW KDRTQWRKLIKNSDPI			
RTE	RTE - 5_M Ma		Mesobuthu s martensii	GGAAC GTCCA GACAC TATGG GCCCC AGGGA AACTC	TAGCCCC AACGGTT ACCCAAC CATGGGA AGATGAA TGAA	MKNFTFDVLGVSEVRWTGKGELSTGDFIWSG EEKHHFKGVGFLLSKKAQKCLISYNPISSRIISA RFSATPFNISIIHVYAPTTDASDEDIESFYGEIKN TLRETPKKDFTIITGDWNAKIGSNNTGWESAM GKEGYGDRNERGERLLEFATEHNFYICNSRFR QKPCRKWTWTSPDGIHHNMIDLILVQSKWKTS VQNCRTFQSADIQLDHSLVLCNIKIRLKKQSTT	cd0165 0	498	754

				GAACT GTTAC GACAT GAA	SQGTLD FRKRNSNDLKQPPIKQAYEKS LNEQL KKITPTMSLDEHASILESAIKTAIKSSLAVTEKP RKPWITENTLKVANEKRKWKLRHISEENNNI YKAHCIQVKKLAKEDKENWLNEQCEAAENGL KTGKTHQAYKLIKMLRKPFTPKLNILKDDAGK ILQTKTAITNQWTKYCSTLYNGSNNNQSQATD AELEEITPPADEYPCPHILHSEVEAAIRTLKQRK SPGFDEITAEMIKAGGPKLTASLHEL CQKIWLE EKIPTNWGKTILIPKKGDLARCENYRTISLISH LEKVFLIILLRRLQHQL EPYLAEEQAGFRKDRS TVQQILLVRL LAEKA KRKGKKIYNCFIDFQKA FDSINQQRIWAVLASFGVDRKVVKLLQNIYEN TRLAVKTGNDMG TWFKMLIGSRQGDPLSPLL F ITYLERTMDSVTKLNLGV TIQGSII SNLRFADDI NLMNEDLQALQDQATALQASAEKSKLRINTS KTKTMVFGARNIEQQIEINNQPVENTEQFEYLG SIITWDNNCSVEIKNRIAKATGALASLKHIWKS SKITVPTKLRILSVCVFSVLLYAAETWTLKEND KKKLLAFEMRCYRRLLKIKWQDKIRNEEIRNT LNIHTTIDLRQRKLRLFGHICRMDNNRJIHV VTSXMNGKNKQGRPPRAWMD DITDWCKCSA QEAFHRAQNRNQWKTATRTND			
--	--	--	--	--------------------------------	--	--	--	--

RTE	RTE - 1_M Ma		Mesobuthu s martensii	TTCTTC GAGAA GAAAG AACAG TGAAA ATGGA CATGG ATCAG CCCCA ACGGA CGGAC AAANA	TGAAATA AGAAATC TGATGAT GATGATG A	MRLTVXLANSNCTIKDVSVLNQFNTGSDHRLV HAKVKXDIQCERXRLXQHGGQKVKKEEYSLK LQTDSKCSQKTKVXYGIHTMRQCRRNRGNST KKSQRQENIHQNTGADRKEKEDEDEDQIHYPDRI EDEVRKSPRRRQFNILLKEGGVPLYKQLAGIF SDCMRTGQIPDSWKNAVIVPIHKKGDKTDLRN YRPISLLSTIYKVFTIILTRRLEQVLDSNQPREQ AGFRKGYSTMDHIHTINQLKEKCKEYHISFCM AFIDYEKAFDSVSTEAILTALSNOGIETTSTQNX CCKDSGIAPRPPSVRRNLLRIIYEEATAIIRLHK DSRKIKIQRGVRQGDSPKLFTACLEEIFRQIN GWEHKGININGEYLNHLRFADDDIVLFAQTPNE LQEMLSDLKTESEKIGLQINYEKTKVMFNAEVI PQPVTIEGRTIERVEEYKYLQGVITTQGDQMRE IQHRTQLGWHACGRYSNFLRDNRIPICLKRKLF NMCILPVMTYGCETWTLTKKMEQKLATTQRS MERYMLGFTRRDRKTNIWIREQTKVEDILVTV RSSKWHWAGHIARYTDNRRTKRITWQPRQG KRYPGRQYKRWRDDIRKVIGSQWMQIAQXRY IWKQKEAYISSSG	cd0165 0	175	448
RTE	RTE -		Mesobuthu s martensii	TGGGG AAAGC AATGG	TAGGATT GGATTGG AT	MIRSSQAENDARGANPPGGPIADVATGKKDHL SCKDLFTIGTWNVRGLNIGKLEIKREMTRTGL DLLGISELHWTGSGDLKSDDYTVHFSGNDHNR	cd0165 0	537	789

4_M Ma		TGAAC	RRGVAFVANKRLAKYIENARCISDRIMTIRLRG		
		TACCC	KPLGVTIIQVYAPTADADKQDIVDFYTAIQNAL		
		TGTAA	DQTPKKDIIYIMGDFNAKVGDGEDGQVVGRH		
		AAACC	GLGDRNDAGDRLVQFCQENKLRLITNTFFKQH		
		TCACC	KRRLYTWTSPNGQHRNQIDYILCQQRWRGSIT		
		TCGAA	AAKTLPGADCGSDHQLLIARIKIRLCKVKKTV		
		ATTCT	VAKKFDVSRIPTQYTIEVRNRFEALSLDGGQTE		
		ATGGA	ELWAMIKGTIETALEYVPYKMPTKGAVWLTE		
		TCAAG	KTLQIAEKRRREAKAKGEIDKARRLHADFORA		
		ACCAA	RIDKNNALNARCISLEEA FRKGRTRELFATVK		
		AGTGT	QVKTPFSARQSTIKDRDGNELTDQQDIKERWR		
		TATCG	EYTEDLYTSIDDQQNEDQTDQGLQEPDLLESE		
		ATATG	VEWAMRQLPSNKAPGIDGIPAELLRPIPRAL		
		GTGTT	ALCQQIWKSCTWPADWRRSVFVPIPKKGDTR		
		GGACG	DCANYRTIALIPHASKILLKVIQKRLGNVIDREL		
		GTGGG	PDVQAGFRRGRGTRDHIANLRWIMEKAREYQ		
		CCCCC	KKLYMCFIDYTKAFDCVEHSLWVALRELGV		
		GGGCT	PEHLIRLIRSLYQNQEATVRTRYGDTDFWKIKK		
		GGACA	GTRQGCTLSPLFNLYAEVIMRKLDLEESEVG		
		ACATC	VKIGGRTISNLRFADDTLLAESEKDLKRMIM		
		GACCA	KLKNESEKMGLQLNIKKTKIMTTAAKEVHIKI		
		TGCTA	NNEEIEVVDSFIFLGSKVDCNGDWRPEITRRIA		
		CTGGG	LGRTAMVNM DPIWRCKDVSITTKRRLVTAIVF		

				GAAGA GCAAA GG		PIATYGCETWILTKADRRKIDAFELWCWRRLL RIPWTARVTNKSVLERINPDFSLEGKITRQKLT FFGHVMRANSLEKDVMLGRMSGTRPRGRPKT RWLDTIKKDTGKNIEELRELVRDRKTWRDMT HRIAKNRTRLNG			
Loa	Loa- 2_M Ma		Mesobuthu s martensii	T	TAGCCAT CATTGTG GTGGGAA AATTTCC CAGGACT GTATCTG AGGTCCT TCAGACA GCTCTCA ATACAGT CTCTAAG TGGTGTG ATAGGGT TGGATTG TCTGTTA ATCCAGC TAAGACT GTCATTG	YDLALIQEPWHYKGGQIRGLNCKEGKLFTFTSD SGPRTC VFVRNNIN VSPVWDFCSR DITTVRVK LDEGGVSRELIVSSAYFPYDKDEL PPTPEFXLL VEFCQKNHFDLLVGVDANSHHVWWSKDINR RGEYLLEFLMGTNL FVLNKGNEPTFVVVNRRE VIDLTLGTQNISDLIENWHVSGEPSSSDHRYICF NLQGMVPSQV TYRDP RRTN WASFREDLKFNL GGIQSRFCSSVVDLEMTVDQIQQAILLSYHQNCP VRTARSSRQVPWWNEQLGGLRXNVRRLFNKA KRTGDWD TYRSAL TCYNKAIRKAKRSSWKRY CQGI EDIPDVSRLVKIMSSGNFCKVGSIRLPDGS MTSTCKDSLKELFRVHFPGSTLIEGSSCLQVQS DLDGHVYHAKREDWGLVRGVINQSNIRWAVS VFKPFSAGPDGILPVFLQQGIDILASPLCRVFR ACLAFGYIPWSWRQVRVSFIPKPGHSSYTEAK SYRPISLTSFLLKTMERLVDRFVRD GALLKTPL CRSQFAYQSGKSTENALHEVVSLIERAIGYKEI ALGAFLDIQGAFDSTSFEQISLAASRHGVCSTII	cd0165 0	468	649

					TACCTTTT ACACGAA GAAGACA ATTGGCG GGGCTAA TTGAACC ATCATTC TTCGGCA AAAGGTT GCAGTTT TCTACCC AGGTCAA ATACCTT GGTG	RWIGSMLKHRVVYSTLYDEEVKILTDRGCPQG GVLSPLLWCLVVNELLTLHSMGKVSSLLGML MT			
RTE	NA VIR TE1	.	Nasonia vitripennis	AGAAA AGTAA GAAGG AAAGA ACTGG AGAAG AGGCA GAAGA AAGAA	TAAAAGA AAATATA ATACAAA TAGAGAA AAGAGCA AACGAAA GAACGW AAAGAGC AATCGAA	MRAAFDKVNRREMWEMMEKLGLGVRIRNRV KELYRGTRSEIRIGEDKIGGFELHKGVRQGCPL SPTLFNVAMADTEKELSKVQEGGIVLGKKKF WSISYADDVVMVANNATGLKQMLKKFKKIIE RKGLELNTEKSKIMVFKNGGRRKKEDSFFWE KKEIEVVKNFEYLGYYMLKENGKEEAQIKKLKE KASTVMSSMWGLGEELFRDNWELRMRLFDT MVKSVMHEYGAEVWGWKGKEELEKIQRKYM KWVMKLDRTTPEHVLHLETKRYKLETRRRR	NA	NA	NA

			ACGCA	AGGAAAC	AIKYEVKLSKAKEGTLWRECWRTLEKEEEEKE			
			TTTGC	NAGAGCA	RRRGXRRXNREEELNKRGWALSEYHRRLRAG			
			AAAAG	ATAGAGA	EQVWSELERIEKDIQWQKWESEMRDSRQARD			
			GAGGG	AGAGGAA	VKELIKPQGETPEYLRRTGNTGKKGLAMTARF			
			AATGT	AGGAAAA	RMGNEARGYKYWMKEEERICRLCREEEETFE			
			GAAAA	AACAAAY	YIFSRCNRTGSRENNWKNILNGERKNLAKLYR			
			GATCA	AAGGAAA	VIWVRKEKEKEEEX			
			CCAAT	GAWAGA				
			TGAGA	AAGAACC				
			AAGAG	AGACGAA				
			AAGAT	GGAACKA				
			GGAAG	AATAGAG				
			AGGAG	CGCAGGG				
			GCAAT	ACATTTA				
			CGGGT	GAGAGCA				
			CGTTA	GGGGAGA				
			AAGTG	CGACGGA				
			CACTA	ACAGGAG				
			TGAAT	GAACAGT				
			ATAAT	GGATGAA				
			CCTAC	ACAAAAG				
			AAAAA	TCACTAG				
			GTGGA	GWAWGG				

				GAAGA	ACAAGAA				
				TGAGT	AACGATA				
				ATGAA	CAGGTGG				
				AATGG	ARAGAGG				
				ATAAG	AACGTGG				
				GAATA	ACGAGAC				
				GGTAG	AGAAAAG				
				AAAGG	AAAACAC				
				GATAA	GTGGTGA				
				GATGA	GCCACAG				
				TCGAG	ATAGAAA				
				AACTG	AGAGGTT				
				GGAAC	AGTAAAG				
				AGAAA	ATAAAAC				
				TATAT	AAGTAGT				
				TGGAC	GAAAGGA				
				TGGAG	GAAGATA				
				GAAAA	AAAAAGT				
				AATGA	GAAAAGG				
				ACAGA	AGAAGAT				
				AAAAT	GTACGGA				
				AGAGG	GACAAGA				
				AAGGG	AGAAAGA				

				CTAAA	AAGAGAG				
				AGGTA	ATGAGGG				
				TGGAA	AACTGAG				
				GGAAG	AGAGGAG				
				ACTGT	GTRGAAG				
				GAGAA	GATGGAA				
				GAAGG	GATAAAG				
				AGAAT	GCGAGGT				
				AAAGG	TTGAGAG				
				AAAAG	AGAAAAA				
				AATTA	AGAGAAG				
				GAAGA	CCCTGGA				
				GAATG	AAAGAAA				
				CAAGG	ATGCAGG				
				AGGCG	ATGTGAG				
				AGAGA	AGAGGAC				
				AAAGG	GTGGAGT				
				AGGAA	TTTTAAA				
				AGAGA	GGGGAAG				
				GATGG	ATGAAAA				
				AAAGA	ACATGCA				
				AAGAA	GAAAGAG				
				TGGGA	ATAGATG				

				ATGAG	AGATGAG				
				AATGC	AGAAGGG				
				GAAAG	CTTGCAG				
				AAAAG	GTAAGAA				
				AAGAA	AAATAGA				
				AAAAA	GTGGAGA				
				GAAAT	CGGTAAA				
				TGAGG	GAAAAGG				
				AGATA	CAGGAGA				
				TTAGG	TGAGAAG				
				GAAGA	AGGGAGA				
				TAGAA	AGAGTAA				
				AGTCT	TAGGAGA				
				GGAAA	AGAGAGA				
				AAGGG	ATTTTGA				
				GAAGA	GAGAGAG				
				GATAG	AAAGGCT				
				ACAAG	GATGCGG				
				GGTAG	CCAGTAC				
				AGAGA	AGATATA				
				GTCAA	GGAAAGG				
				CCAGG	GGACAGC				
				GGAGA	AGGTGCA				

				GAGAG	AGGTGTG				
				GATAA	GGAAAGA				
				GGGAA	GGCCTGA				
				AGGTA	TAGTACA				
				ACGGA	CCTAAAC				
				AAAGG	AATTGGA				
				AGAGG	CGGAGGA				
				GACAG	GGATTCT				
				AAAGC	ATAAAAG				
				AGATA	ACTGGCT				
				GAAGA	GGAGGAT				
				AATGG	ATAATGG				
				AATGG	AAGGAGT				
				AGGAT	GGAATGG				
				CGAGG	GAGATGG				
				AAGGA	AAGATAC				
				GAGAG	AAGGACG				
				GATAG	GAGGACA				
				GAAAA	CGAAAAA				
				AGAAT	GGTAATT				
				AATAT	TTTAAGA				
				TATAA	GAAARGA				
				TAGCA	AGACATG				

				GGGTG	GAATTGA				
				GAAAG	TATGGGA				
				AAGAA	GAAGAGG				
				AAATG	GCGGAAA				
				GGATA	TAAGAGA				
				GAAAG	AGGAGGC				
				ACGGT	AAAATCA				
				AGAAG	GGCTGGA				
				ACTGG	GCAGTGG				
				ATGAA	CTATCMT				
				AAAAC	TCGATGA				
				TCGAA	GAGAAGG				
				GTATA	GCCAAAG				
				GGTTA	CGCTGAT				
				CGTTG	TATGAGA				
				AGGAA	GAGAAAA				
				AACGG	GGAGAAG				
				TCATT	AGAAGAG				
				GAACG	GATGAGA				
				GGAAG	TCGAGGA				
				ATCTA	GGGCGAG				
				AAGAA	GTCTCGC				
				TAGGG	ACAAAAA				

				GCGGA	AGAATGG				
				ATGTA	GTAGCTG				
				GAAAC	AAATAGA				
				TGGGA	AGGCGAA				
				AGAAA	CTGTACA				
				AAGGA	AGTGGGA				
				GATAT	CGAAAAG				
				CATGA	AAGGAGA				
				GAGAG	GAATTGT				
				AAGGC	AAAAAAA				
				TAGGT	GGAAAAG				
				TACAA	GGATAGA				
				GGGTC	ACKGGGA				
				TGATA	GAAGAGG				
				TATTT	RCAGATA				
				ATAGA	TAAGAAA				
				CAACG	TGGAGGT				
				ATCTA	CGGGGGA				
				ACATG	GTCAGAG				
				GAAAG	GAGGAAG				
				AGAGA	AGGAAGA				
				AGAAA	WTATCAG				
				TAAAG	ACAAAGA				

			AAAAA	TGGAGAG			
			CTGAG	AAAGGAA			
			AGAAT	GAGTAGA			
			ACGCG	GTTTACT			
			ACAAA	CACAAAC			
			GGAAA	ACACATA			
			GAAAA	AACACAC			
			GAAGG	ACAAACA			
			AAAAG	ATCAAAC			
			AGGTC	ACACACA			
			AAGAT	CACACGT			
			AGGAT	ACGACTT			
			ACAAC	ATACACA			
			AAACT	CGCACAC			
			ATGGA	ACACACA			
			TAAAA	CATGCAC			
			GGCGT	AAACRGA			
			AGAGT	CACGAAA			
			ATTTA	ACARGAA			
			TGGAA	AAGMAG			
			GGAAA	GAGARCR			
			GAGAA	CACAYAC			
			CAGGA	ACAYACA			

				ATTTT	CACATRT				
				TCGGA	ACACASG				
				GAAAG	AACATTA				
				AAAAG	CCTACAG				
				GACAG	AGACACA				
				GAGAG	CATAACA				
				GAAGA	CACCAGA				
				AAAGA	TTAGGTT				
				GGGGA	AAAGTAG				
				GCAAA	GATTAAA				
				GGAAC	GCAAGAG				
				TTAAG	GTATGTA				
				GATAA	AATATAC				
				TTAAT	GAATGAA				
				TGGAA	TAAATAT				
				TGTAG	TGTAAAC				
				CGGGA	ACGAATA				
				ATAAA	GTGAAAT				
				GTGAG	ACAATAA				
				TAGAT	AGTTATT				
				GAGAG	ATTATTA				
				GACCT	TTA				
				GGGGC					

				TATCT					
				AAAAG					
				GGTTT					
				GATGT					
				AATTT					
				GTCTG					
				CAGGA					
				AACAT					
				GGCTG					
				GAAGA					
				TAAAG					
				AAGGA					
				AATTT					
				CTGGG					
				AAAAG					
				AAATT					
				AGGAG					
				GTTAT					
				GAAGT					
				CAGAA					
				TTAGA					
				AATGC					
				AAAGA					

				AAGGG					
				GGAGA					
				AAGAG					
				GGAGA					
				GTGAA					
				AGGGG					
				GAATA					
				GTGAT					
				GGCGG					
				TAAGA					
				CAGGG					
				AACGA					
				AAACA					
				GAGAG					
				AATAG					
				AGTGG					
				ATAGA					
				AGAAA					
				AGACA					
				AACGA					
				ATTCA					
				TAGGA					
				GCAAG					

				AATAT					
				TAGGA					
				AGGAG					
				AGATA					
				TATGG					
				TGGAT					
				AGGTA					
				CGGCA					
				TACAT					
				GAGAG					
				AGGAA					
				AAACA					
				AGAAA					
				ACTTT					
				AAGGA					
				GATAG					
				AAGGG					
				ATGAC					
				GGAAA					
				ACGCC					
				AGAGG					
				TGAGA					
				AATTG					

				ATATG					
				GTGCG					
				GAGAC					
				ATGAA					
				TGCAA					
				GAACG					
				GGATA					
				GGAAG					
				GAGGA					
				GGAGT					
				CGACG					
				AGGAA					
				GAAAA					
				CGAAG					
				AGAAC					
				AGGGA					
				ATCGA					
				AAGAT					
				GAAAA					
				GATAA					
				ACAGA					
				GAAGG					
				GGAAG					

				AACTT					
				CTGGA					
				AAAAA					
				CAAGA					
				GAAAT					
				GGGAC					
				ACAGT					
				ATATT					
				GAATG					
				GAAAT					
				ACCGA					
				AGGAG					
				ACAAA					
				ACAGG					
				AGAAT					
				ACACG					
				TACAT					
				AGGAG					
				GAGCA					
				GGATG					
				CTCAG					
				TGATA					
				GATTA					

				TGTGA					
				TAACA					
				AATGA					
				GCAAG					
				GAAAA					
				AGAGA					
				GGTAC					
				AACGC					
				AATAG					
				AGATA					
				GAGCT					
				GGGAA					
				AAACA					
				CTCGG					
				GAAAA					
				CAGAT					
				GAAGA					
				GAAAG					
				ACAAA					
				ATCAC					
				AGAAA					
				GAGCT					
				ATATG					

				GACTG					
				AAAAA					
				GCAAT					
				AGGTC					
				AATTC					
				CGAGA					
				AGAAC					
				TAAAT					
				AAAGG					
				GGAAG					
				AGGAA					
				ACAGA					
				ATGGA					
				CAGAA					
				CTAAA					
				AAGAA					
				AACTA					
				GAAAA					
				AGCAA					
				TCAGA					
				AGGAA					
				AAAGG					
				GAAGC					

				GGCAG					
				AAAAA					
				GAAAG					
				AAAAA					
				CACGT					
				GGTGG					
				GATGA					
				AAAAT					
				GCAGA					
				CTCAG					
				AAAGC					
				AGAGG					
				TAAGG					
				AAGTT					
				AGGCA					
				AGGCG					
				ATTAA					
				GAAGG					
				GAGCG					
				GGAAG					
				GAAAG					
				AATAC					
				GTCAA					

				AGCAA				
				AGAGA				
				GACTG				
				GGGAA				
				TATTA				
				GTAGA				
				AAGAA				
				AAAAG				
				CAGGA				
				AGAAA				
				TGGAA				
				AAAAG				
				AATAA				
				AAGAG				
				GCTGG				
				AGATA				
				GTAAA				
				ACCGG				
				GAGAA				
				AATTC				
				TGGGA				
				GGTGG				
				TAAAA				

				AGTAG					
				GAGAA					
				GAAAA					
				AAGAG					
				GACAA					
				GATGC					
				AGCAG					
				TAAAA					
				TAAAG					
				AAAGA					
				GGAAT					
				GGTTA					
				ACTYA					
				CTTCA					
				AAGGG					
				CAACT					
				AGGAG					
				AAGAG					
				ACAGA					
				GGAAG					
				ATTTG					
				GGAAC					
				AAGCA					

				YACAA					
				GGAGA					
				GGATG					
				AAGAA					
				GGAGA					
				KAGAG					
				AGGAT					
				AAAGA					
				AGRAA					
				GAATA					
				ACCGA					
				AAACG					
				AGGTG					
				AGGAA					
				GGCAA					
				TTAGA					
				AAAAT					
				GAAGA					
				AGGGA					
				AAAGC					
				GCCAG					
				GAGCA					
				GACGG					

				ACTAC				
				AAAAT				
				GAGGT				
				ATGGA				
				TACAC				
				GGGGA				
				GGAAC				
				AGGTA				
				ATAGG				
				AGAAA				
				TAACT				
				AAAAT				
				ACTGA				
				ATAAC				
				ATATG				
				GAAAG				
				GGGGA				
				AAAGT				
				ACCAG				
				AGGAA				
				TGGAA				
				AACGG				
				GACTA				

				ATAAC					
				CCCGC					
				TTTTCA					
				AGAAA					
				GGGAA					
				GAAGG					
				AGGAA					
				GCAAA					
				GAACT					
				ACAGA					
				GGGAT					
				AACTC					
				TCATG					
				GACAC					
				AGGGT					
				ATAAA					
				TTATA					
				CGCAG					
				AAATA					
				ATAAG					
				GGAAA					
				AGATG					
				GAAAG					

				GGAGT					
				TAGAG					
				GGAAG					
				AGAGA					
				TGCTT					
				GACGA					
				TACAC					
				AAATG					
				GGTTT					
				CAGAA					
				AAGGG					
				AGAAG					
				GTACG					
				GCAGA					
				CGCAA					
				TTTAT					
				ATTTT					
				GAGTA					
				AAGCA					
				ATAGA					
				AACTG					
				AACTG					
				GATAA					

				GAAGG GAGGT AAAGT CTACG CGTTTT TCGCG GAT					
CR1	L2- 3_N Vi		Nasonia vitripennis	TGGTG AGTAC AGTTG ACAGC TCTGC TGATC CATCA TTTTTC TGTAC TGCTTT TCTGC TGCTT CTTGC CTTCG GGCCC TCTTCT	CTCGTAC ACACTCC ACGCACA CGCACAC ACCTACT TTCTCGC TCATTCC ACCTTCA CTATCGC CAYACTY TCACACT ATACATA CTCTAAA CATACT CAATCTA TTGTTATT	MEALKTQNVSERISKIEQRLGPLEKRLKALDE LPALKTRIHNAESTITELQAQIQDLSSRSPTMQ QDNGSTVPNTAEICSLRSELAEVKRRQEQTSN CVVVVTGLHYTRETSLHLLAFSVVNALDPTVL RRDVASVRTMGRLDATNSSARGDGRPLPLAV TLSSSALARSIVIAKARKRKLHTSELDATLLEE AKALSPDHQGLININELLPSDVHKLRTTRARLEA KKRQGCRTFVREGRLYTCAATMTASVLRSSPT TPSWRLFPGRPLPTSSNTEATXQYTHQTLSSS AQERSKVKSEYLFCEVSAKGVSPIFVGVVYRP PHAPFFQGSNFIDQLTTHMHNYSTKVIMGDFN SDQLSSSEDAKFIFIDENSLSSVPYGATHHK QGSDTWLDLCLIDEQDRLLSYWKTNSPFINGH DLITATLDVQIPRYVPNTYSYRNFKGISAELR DFLSACDWSSLTSSSLDECISTLNANLTNAINH LAPLRTVTPRRQRHPWFTTVSHVXXXXRILVA	SSF56 672	671	971

			GTTGT	TCTTTTCT	LRDLVSEDRLYRRFRDSRLDSDLRIYRLARD			
			AGATA	TACTTTG	NAHKQVEEARLNYYY SRLSTLTDVAEIWRELE			
			GTTGC	TAATGCT	KLGISATKAPSPSRFTTDELNKHFSISNDPLAP			
			ACGCT	GTAAACT	AVEDYLLTLESLDLPEHFEFSTITESDVLA AVS			
			CCTTTT	TACAATC	HFDTQARGSDGIPQVVISKALPVLAPLLSHIFN			
			TGCAA	GTACAAT	LSLSEPCFPSAWKLSLVRALNKVSSPTALTDYR			
			CCTTC	ATAATGT	PISLLCFLSKALEWL VHRQVSEYLESRLLLDNF			
			ACGTC	ACGCAAA	QTGFRTGHSTHVRVGINKKKVTLLLLFDFSKA			
			GATAC	ATAAAAC	FDTVCHVRLLRKLSTFGFSKQVIRWFASYLTG			
			ATAAA	TAATCT	REQAVVGDNSERFSPRRPAGSVWSLLHCTST			
			TCGCC		TSVSALIPMFILSMRMTCKFTVNATLRSSILY			
			GAAGG		QTRRAFLSSASRRVRLVLAVALYINDIGFCLDS			
			CTGCA		DVSHLIYADDLQIYSQCHLEELDSL SNKMSAN			
			ATTTTT		AERIMGWAAQNRLKLVNKNKTAIVLGSPYYI			
			CTCTG		NALPKVTLINMGVPGQLHLLIWGCQVSFESSV			
			TCATC		RNLGLVLDSKLTWKEHVTQVCKRAHSLMYRL			
			TTTTCT		YFFRKSTNLRLRKHLVQALLFPIIDYCSLVYCD			
			ACTGC		LTQELDLKLQRLVNTGIRYIYGVRRDEXISPYR			
			AAACT		RELQWLTTAGRRKYFTACFLRKMFNSVVP SY			
			TTCTA		VLAFFDFRVTLRPVRGEVTPLEIPAFATETLRN			
			CACGG		SFHISASYLWNNLPSHIRNTSSTVTFRKLA KDH			
			CTCCT		FFQLENT			
			AACCT					

				ACAAA					
				TCTGC					
				TTGGC					
				TTGGC					
				TTGGC					
				TTGTTT					
				CCTCA					
				CTTGC					
				ACGGC					
				GTCTC					
				TTTCTC					
				TCCGC					
				AATCA					
				ACATC					
				GTGGC					
				GACAT					
				CTGCT					
				GCATC					
				GTGGG					
				AGAAC					
				TTTCA					
				AGGCG					
				ACTTC					

				ACACT					
				GATCG					
				AGGTC					
				GGTAC					
				TTTCA					
				CGCTC					
				TTTCC					
				ACTAA					
				CGCTC					
				CATCT					
				AGCCG					
				TAAGA					
				GTCAC					
				AACAC					
				CTACT					
				CTTTC					
				ACTCC					
				GTCGA					
				TAACT					
				TATCA					
				TCCTC					
				CTACA					
				AATAT					

				CTGCT					
				TTACC					
				GACTG					
				CTAGT					
				AAGTT					
				ATCGA					
				CGAAA					
				AATGA					
				GTGCA					
				AACGG					
				CGAGA					
				AGAGT					
				GTGGA					
				AGTGC					
				CTTTTC					
				AGCCA					
				TGCAG					
				CAGCT					
				GTAAA					
				CAACC					
				TATAC					
				AGCAC					
				AAAGA					

				CCCCA					
				AGAAG					
				TGTGA					
				ATTTT					
				GTGGT					
				CTGCA					
				CTACC					
				ACGTC					
				AAGTG					
				CCAAA					
				ACACC					
				GTTAA					
				CATCA					
				AAACT					
				ATCGT					
				GGTGA					
				ACGAC					
				AGGCA					
				GATTA					
				TAGCC					
				TGTCC					
				CTCAT					
				GTGCA					

				GACAA					
				CAGCA					
				ACGCC					
				AAGGG					
				TGCCA					
				AACTG					
				AGGAC					
				TAAAT					
				CCACC					
				TCTGC					
				CACAA					
				CATCA					
				GCAAC					
				GGGAG					
				TAACT					
				GCAAC					
				AACAG					
				CAGCA					
				GGCAA					
				GAATA					
				CTCCC					
				AAACA					
				ACAGC					

				AGCCG				
				GTGAA				
				TAAGA				
				AGACG				
				AGGTC				
				AGCAC				
				CACCA				
				TCAGC				
				CAACA				
				CGTCG				
				AGAAA				
				TCCCT				
				CACCA				
				ACCCG				
				CTCGG				
				CCGGA				
				GTAAC				
				AACAC				
				CTACG				
				CCGTC				
				CGTTG				
				AGGCC				
				GCTTT				

				GAGGG ACATT CGCGG AGCAC TGCTC TCGAC GACAT ACGCA ATGCT CAA					
Loa	Bag gins- 2_N Vi		Nasonia vitripennis	ACACA GATAA ACTTG CATCA TAGCA AAGGC GCCTC GGCCG CACTG GTCAG GCGCA TGGCT AAGAT	GGGATGT CATGACG AAGGGTG GTACAAT GGGTTTA CAGCCTA AGTGCCA GGGATCT CGGAGGT CCCCCCT TCATAGT AATAATA ATAATAA	MPNTRLASRGEIRGLGGAGRIFRDPTCSVPRA CILVKGFDALLLPARCSRDLTAIKIRLPVGEGSE REVVVASGYFPYDSQEEPPPREVQDLVEYCRQ RSIPLILGCDANAAHHIVWGSSDTNGRGDALLQ YLVTTSLCIMNRGREPTFYNSVRSEVIDLTLCT VGMEGWVGSWRVSNEPSLSDHRYIQFEWKER CSETRAFRNPRKTDW AFFRENLRNELQSFKPSF GTTDELHDHWA FELGEIVNISFQRSCPWTIPRGT WGTPWWNRELEDLRRETRRTL NRAKNTRNSI DWRIHREAQRL YKNRINAVRIKGWRDYCEDIE RYPDAARLLRILAKNPEVWLEAIRLPTGEYATS EEECLKLLLEANFPGFRLSHEMGEDESSGRNRQ QRAAWDLAAKVVTPEKVKWAI RN FQPFKAPG	cd0165 0	459	718

				GCACA CAGGC AT	IDGIYPAFLQEGLEELVGPLVKLFRASVALAHV PEIWKTAKVVFIPKTGKPSHTTVKDYRPISLNS FVFKTLERLVDRFIQDDILTVFPLHPNQHAYRA GFSTETALHSAVWRIEEQLERGEVMVGVFLLDI EGAFNCTSIEAVVKEAGLHGMPGPLIKWLQG MLTRRTVISSLGTVTVSGEVSKGCAQGGVVSP TIWCLVANGPLEALNGWGCYAQAYADDFLILI KGSEIGAAMDAMQLALRKVERWCCSTGLSVN PDKVEMVIFSRKYKLDAYRAPKLSGVALQVK DSAKYLGIVLDKKSTWEKHLTAQCDKFLVAL WTCRRAFGSTWGLSPRIILWLYRAVLVPRLAY AALVWWPRAELAGARAAPEKLRGQVLRGAT GAYRTTPTKALGILVKVEPLHLTIIGMAAKAA HRLNAYGQWTQGTRHTRLPGGVGLLPEFSIKT DMMIPRYFFGRRYGVVIPTREEWKARRDSLPG TGDVWYTDGSRAETGTGSGYYCRRDGRGTFF SLGRYATVFQTEIYAILTCAQRNIELGARDRIIT ICSDSQAALRALMAHRTTSRLVWECKVVVNQ LTAHNNKVRLLWVPGHTGIRGNEIADRLAAL GAKHPPIGPEPYTGAARCLLAGEIRDWVEREH TKEWQGTQGCRQAKAVMGQDTNVGWTCKIA GGSRNNSRLLTQIVTGHIRLRYHGLKMGKEAT			
--	--	--	--	----------------------	---	--	--	--

						GVCRWCEGAEETPTHVLTVCPKFAQLRHQCL GELFPTYEEIRALDAGAILTFWRRAGLPA			
R2	R2B _NV i	.	Nasonia vitripennis	GACTA GACTA TGGGT TCAGT CAGTC CCAAA TAGCC GATCC TGGCG CGTCC GGCAG TAATG CCACG TATGA GTCGG TTACC CATCT CTAAA CGCGT AGAGG TGGGG	TGACCTG AACAAAA CGTGTTG TCTTGTCT TGTCTAA AACTATT TATTCGA AATAAGG GGAGGCT AACTGCC TGCAAGT TGAACGC GAAAGTT AGACCTT CCCACCT AAAGCCC AAAAGTG ATCGGGG AATGAAT CCGCGGG TGACCCC	TFAPTHPMVRSGPCRKTKRPGSDYRESLIMDS GNNVASEPRGAVDVTSAAPIGAELNAEPCEGR NQRREAALSAQTRRRNXARRARNAQQADEPG DDEEIETHGPLTIRTXEPMEIVAIKPNQACPK CLQGGTQLLCMGSWELSRHINKEHPSVDVTW VCGACQRRCTTLRSWSCHVLHCKGRQEPKDL PFKCEHCSSLFDSQIGLSQHERHVVHPEVRNDKR AAEANKPKGKSGRRPSIWSEDELLIRELESEY HGARNINEKIAEHFPDRTGRQVSDARRRKDYA ALRGRGGPQGPAEGVEAIEEVDEGEIPEGEELV ATDGAALSGPPENGGSAPAEQVNAPALESSS QQDRECSPA VGSDEQIEDSSDDDEFSDALGEIS LPEPLSVERTTISPPPRDDWKGP MRWEICNASE EAGSYANWVTGLQELVRNNALSEIGLDSL YD QLIQIMRHPSDDNEQDRLQLNARGPPRRGHRK NRRRRRLTAADRKRFAFARCQDLWNNNP KKL AELVIANDLSILQRRQAPGR TETQTLYNELWG RVGPNIEAPRRTEDPIPVSRIFT PITPQEIMGRIR RIKNDSAAGPDGVTKDDL RGRGVSI ALSKLFN SILLAGYYPKAWRENRTTLLPKPEKDPADVKN WRPITISSMVS RVYSGLLDQRVRAVIKQCDRQ	cd0165 0	626	885

			AGCTA	AGAGTTG	KGFTEENGCFSNSIQLLDDAVSNAKKAGGVITIL			
			AAGGC	GGTAAAC	DVSKAFDTPHAVIQGCLEKKGIPETVAAYISS			
			CAGGC	CCTTGAA	MYRDCSTAIRTRSGDVKIGMKRGVKQGDPLSP			
			GGTTT	ACGTTGG	LIFNLVLEPLLERLQETSGVEIEGMNLSCAAFA			
			ACCCG	AGAAGCG	DDIVCFANTAPEAGRQLRMVADYLGRLDMSL			
			ACGTC	GAAGAGA	SVSKCIAVEYVPHRKTWYTKNPGLEVNGNAV			
			GAATT	GTCCCGC	PSISPSETFKYLGAKVSPWKGLLEGFESDAFRE			
			TCTCC	CACCGAG	VISRVQRLPLKPMQKVDLLQMYIFPRYTYGLIT			
			AGGTC	CATCGAG	SPPAKAVLKTIDRIIRTRIKEILHLPESVSSFLY			
			TGTGT	TGCTGCG	TPRKQGGLGLLEVEKMLIAALRNGLRARQS			
			CAGTC	GCGCCCG	HDPVTRAAMNSNAADDRKLSYADALRLHWP			
			GACGG	AATGAAA	LTTKELDTYKYQLRLSYAQKWAEQKWQGGQ			
			AATAA	CCGATCG	VEEFAQDPVGNSWLQRYDLLPASRYIDAIKLR			
			AGGTA	CGGATGG	TNTYPTRALMKIIDGRVDSSCRKCQGSSETLG			
			CTACA	TGCAAGT	HILGRCRYTKDKRISRHNEIKDLLKARLAKNH			
			ACATC	CGTAGGA	QVMDEPQITVRGQRFKPDLLVKTNEGRVHVID			
			TACTA	CGGGGCA	VTVRYEHRTYLDEGRTEKIGKYRQILSTLRRD			
			TCTAT	CGACCTA	LHSNAEEVIPIVIGSRGAIPRETRKALSKLGIGK			
			CGGGA	AGCCTCT	SDWLTISLIALRSSLEIVNAFMDD			
			TCGGA	GTCACGG				
			AGACG	CGGCGAA				
			CCTTA	GCCAGGA				
			CAGCG	ATCACCA				

				TTTTCC	TGCAAAG				
				GATTT	GTGTGAA				
				TTGCT	CTGGGGC				
				CTTTG	GGATACC				
				AGCAT	TCCACGG				
				TTTTCT	GGTTTCC				
				TCAAA	CTGGGCA				
				TTGCG	TCGCGCG				
				ATAAC	AGCGATG				
				CGACC	GCCAAAG				
				CGATC	TCCGCTT				
				ACGCG	TCTCAGC				
				GGGCT	TACAAAA				
				TTGAC	CAAAAAT				
				AAAGC	GGTATGA				
				AATGC	GACTTCG				
				GTGGT	TTAACAC				
				CGGTA	TAATTTTT				
				AGATG	CCGAGCC				
				GTTGC	TAGCAGG				
				AATCT	CTCCCTT				
				TTTCC	GACAACG				
				ACCTC	CTTATGA				

				GTTTCT	ATCTGGA				
				TTTAC	AAAGGAC				
				GGAAC	ACAAAGT				
				GAAAG	GGAAAAA				
				CAATG	GCGCTGA				
				CGTGT	TGGTGGA				
				GGGGA	CAAAAGT				
				ACGTT	CAGTTGA				
				AAAAA	GACTTGA				
				CTCCC	TATCAGT				
				TTCAT	TGTTTTG				
				GCATC	ACTAAGA				
				CCAGG	ATTTTATT				
				ATTTA	ATCGTTG				
				TCCTG	ACTTTTA				
				CTTAC	AATATTT				
				TGCAA	TATTATT				
				AGCAA	GACTGTT				
				TGCGT	AATATAC				
				GTGGA	TGACTTG				
				GCGAC	GGACCAA				
				TTTAC	GTCATCT				
				CACGA	CTGTTAC				

				GTCGC	CCGGTAC				
				TCCAC	CGGTTCC				
				CGCAA	TGTCATC				
				AGCAA	AAACCGG				
				TGCGT	AAAGTCC				
				ATCGC	GTCCCAC				
				GCAAA	GTAATGT				
				AGCAA	GGTAGAC				
				TGCGT	GCAGGAG				
				GTGGG					
				GGACT					
				TGTCA					
				AAGAT					
				CCCCC					
				GCCGC					
				AAAGC					
				AATGC					
				GTGTC					
				GGCAC					
				CACGT					
				AGAGC					
				AAAGC					
				GTGTA					

				GGCAG ACTTT GTCAA AAGTA GTTCT GCCGC AAAGC AATGC GTGTG GAGAT CTTCG CCGGT GAAAG CAATA CGTGT GGGCG AACTT AG					
CR1	L2- 5_N Vi	.	Nasonia vitripennis	TGCTG TAACC TAACC TACAA CCTCT	AA	GILPHSPVLPVRMNTPPARPRPRRRELNQQQR QQQIHLPRARLERMCPKRINRHPHRLLAPEP NRLREHRLRLSRRPLRTSSRRWMTFGRSRMRP FRHKTPCLKGSLELRKGLALLQSRLKALDDL ALKTRLNNVESSISELQAQYQELSSRSPAEQQN	cd0165 0	901	1157

			CATGG	ISIAPASEEINNLRSELAEVKRRQERSANNVVVI		
			CTTTTC	SDLAYTQETSLQLLAFTVLAALDSTVLRRDVA		
			TCCTT	TVRTMGRLDATTNAQGDSRFPPPLAVTLSSSA		
			GCTGC	LARSIIVAKARKRKLHTSELDAALLEEARALCP		
			ACGGC	DHQGLININELLPPDIHKLRKARLEAKRRRGC		
			GTCTA	RSFVRDGRIYIRSGEDSERATIISTDAELEAFLA		
			TTTCTC	HTPPPIANTDITHGFFSPYAANSQHGHTLRHTP		
			CTCAA	IIPPKPPSKSSNSSTSASKASLSEGLRVCHFAN		
			AGCGA	SLTGHIEMVRLFLSTRSPFHVIAVTETWLSKIT		
			TCCAC	SIPSLDDYILYRRDRNRNGGGVALYIHHSLTAS		
			GTGGA	VISSSDGEWSGKPGKPEYLFCEISAKGVSPFVG		
			GCCAC	VVYRPPHSPFIQGSNFIEQLTTHMHNYSTKVIM		
			CTGTC	GDFNSDQLSSSEDANFIRAFIDENSLTSVPYGA		
			GGGTA	THHRQDSDTWLDLCLIDEQDCLLSHWKTDTPF		
			CACGG	INGHDLITATLDVQIPRHVPATYSYRNYKGICA		
			AACAC	EKL RDFLSACDWSSFTTSSLDECITILNANLTN		
			CACCA	AINHLAPLKTVTPGRKRHPWFTAALRDLLTER		
			GCAAC	NRLYRRFRKSRLPWDLYFYRIARDDAHKQVE		
			ATCAT	EARLNYYY SRLSTLTDVAEIWRELENLGITTSK		
			CAATC	SPSPARFSSDALNKH FSSISNDPQAPSVEEYLRT		
			ACAGG	LES LDLPEHFIFRAITESDVSAAVSHFNTQARGS		
			TCGGT	DGIPQVVISKALPILAPLLCQIFNLSLSEARFPSA		
			CAATT	WKSSLVRALNKVNSPTALTDYRPISLLCFLSKA		

			CTCAC	LEWLVHKQISEYLETRLYLDFNFQTGFRTGHST			
			CTCTCT	QSGLIKLTDDVRLGIDRKKVTLLLLFDFSKAFD			
			CGATT	TVCHVRLLGKLSSFSGFSKQVIRWLASYLSGRE			
			AGCAC	QAVIGDNNELSTSLPLNTGVPQGSVLGPLLFAL			
			CCCAT	YINDIGFCLDSDVSHLIYADDLQIYSCHLEELDS			
			CTGGC	CSVRMSANAERIMGWAALNRLKLNVLKTKAI			
			TTCGA	VLGSPYYINALPSIANTFINIGGARVDYESSVRN			
			GAGTA	LGLVLDSKL TWKEHVTQICKRAHSLMYRLYFF			
			TTAAG	RKSTNFRLRQHLVQALLFPIIDYCSLAYCDLTQ			
			TCTCT	ELDTKLQRLVNTGIRYIYGVRRDEHISPYRREL			
			AGCTC	QWLTTAGRRKYFMACFLRKIFNTSVPTYILAFF			
			TGTCG	DFHVALRPVRGEVTPLDIPSFKTETL			
			ATAAC				
			TTACC				
			ACCCG				
			CTACT				
			ATCTT				
			ACAAC				
			ACTCT				
			CTACG				
			TTATC				
			GACTG				
			CCAGT				

				AAGTT					
				GTCGA					
				CGGAG					
				ACTTA					
				ACAAT					
				AACAT					
				CGAAA					
				ATGAG					
				CGGAA					
				TTGGT					
				GTGAA					
				GAATG					
				ACGAG					
				TCGCT					
				CCTCA					
				TGTGT					
				GCTCA					
				CTGCA					
				AAAAA					
				ACAAA					
				AAGAG					
				GGAAA					
				ATGAA					

				GACCT GCGAG TTGTG TGACC GGCAG TTCCA CACCA TATGT GTCGC CTTCA GGTCT ATAAT GGTGA						
L1	L1-11_NN	.	Nelumbo nucifera	GGTGG GAGTA ACAGA AGAGA TAGGG GWTGG TGAGG MTGTA GAAGG GTCTA	TAATGAA AGTTACC GTTTCCT ATAAAAA AAA	MKIISWNVRGLGSSRKRSVIREVLRKEKPTIFLI QESKLTEVDRRWVRSVWSARGLSWAFTQSW GASGGIITIWKNEEMDCTEELIGNFTVTIKLRR KEDGFEWALTNVYGPNSYRERDDFWXELAAI RGLWNGPWCLGGDFNVIRFLNETTNPNNHRT SSMRKFNLVDSYELRELPMLQANFTWSNFQE EPLCCKLDRFLFSSDFESAWRGLAVRALPRPTS DHCPVRLSQENLGGGPCPFRFENMWLLHPDFI KKVQLWWEEDQVEFRWKGYKFYKKLQALKA RIKAWNKEEFGRVEERKEWIANRIQEIDGGEA	cd01650	496	761	

			AGGAG	ERRLTVGEREERWKLKQELQEIRKEEISWRQK		
			GAAGG	SRCKWLKEGDGNTKFFHTLANGRKRKNFVGR		
			GGAGM	LEWNGEVITSEEEIARAFVEHFKSIYDSPRRERE		
			GAATC	EIPDLDWLPGNRVLMDQLEKNFEVQEVKDAIF		
			TGGGA	SMERSKAPGPDGFTMAFYQDCWDIIGGDLME		
			ATTAA	VFLEFFERGKINSGANATFIALIPKKVNANHPK		
			TAGTG	DFRPISLIGSMYKVISKVLAEIRKKVLPSLISP		
			ATCCT	QGAFVSGRQIIDQILIANECVDYRKRQGEKGLI		
			TTTCG	CKLDLEKAYDLVDWGFLLEVLRLRGFGERWC		
			AGCAG	MWVKGCISSCHYSILVNGVAKDFFPAKRGLRQ		
			TTGGT	GDPLSPFLFTLIVDVL SVQITRADARNLLRGWR		
			AGAGG	VGREEIVVTHLQFADDLILFCEGDPEELENLKV		
			TGGAA	LVENFEKVSNLRVNWNKSCVAGINRQEVEVA		
			ATTCA	SLARSLGCEVSDWPLKYLGLPLGGNPRSVSFW		
			CTCCA	YPVLERFSKRLAGWKRNYSLGGRITLIKSTLA		
			GGTTG	NLPVYFLSTFKLPVKIANQVEKIMRDFLWEPG		
			GGATC	DTKKDHRVKWEVVSRRKKDKGGLGIGNIRRN		
			AGAAG	AALLGKWLWRFPEQESLWHKVIRSIHGMAA		
			AGGGG	NGWDAGLDPRASRFSPWKDIVGILPVFSPFIRY		
			GCTGA	DACVGNKILFWEDLWIGESPLKDRFPNLYSLS		
			ACTKA	LKKSQFISSFRASGDSWDLGLRRGLRDSEVDS		
			TTGGG	ANDIFSLLSSVFLDHSSPDSRIWAETSTGSFSCS		
			ACTAC	SFFNALVLPSNPVSSFPVKEVWSAPVPPKVRAF		

			TTCTA	CWEVVLQKINTLEVLQRKRPNWCLSPHVCSLC		
			AGGAA	LRNGETVNHLLISCPFSALIWDWFSSLFDCRWT		
			AAGGA	PPTSVADLFDPSFFRQPKKKGRVMWRTTTRAAI		
			GAGGG	LWAVWEERNNRIFQGKEKSAPEVIEGITIKAV		
			AGATT	HWLKTLDLFKDVSIDDLIRGWSILARSSTSSPTI		
			AATAA	QAVWSPPLQCNVKNFDGSSIGNPGPSGIGGV		
			GAGTG	FRDAEGKILALYSGPIGLGDSLRAEVLALLQGL		
			GTGCT	RWAKQLGFQQLEVEGDSKLVVGWLQSNLGT		
			GGAAC	WSLRQHVVRELLSLTRELGAQVYWSRRSANAL		
			TAAAA	ADNLAKQGVNREVFVLCVVSYLLFRGHLLL		
			AGGCA	WKVTVVSSGFI		
			ATTAA			
			AAAAC			
			TTAGC			
			AAGAA			
			TAAGA			
			AAAAA			
			GGACA			
			AGAAG			
			TAACT			
			CAGTT			
			AAAAG			
			CTTCA			

				AACAA					
				GGGGG					
				AGGGA					
				GAGAA					
				TTCTA					
				AAAGA					
				AGAAG					
				AAAAA					
				AGATG					
				AGTGA					
				AGTCG					
				AAAGA					
				CTTAG					
				GTGTG					
				GTGTG					
				GATTA					
				TGAAA					
				GAGGA					
				GGCAC					
				AGGAA					
				GGAAG					
				GAGAG					
				AAAGC					

				TGGGT TAAAT TAAGC TGAGT CGTTT GGGCC TTTTTG CTTAG TTTTCC					
L1	L1- 5_N N	.	Nelumbo nucifera	GAAGT TTGGG GCAAA AGGAG AKAGA AGGAA AGGGG GAAAG TAGTG GTCAA GGGTA GAAGG GAAGT AAGAA	TAACTTA ACTGTTT CTTTTTTC CTTTTCTC TCTTTTG AGGGTTT AAGTTGT ATTCGGA AGATAAG AGTGTTT GTCTTCC TTGCTGT AATTATT CTTGTGA	MDMCGCLQMSMGQMTTEREELWEELADV RG LWDGPWCVGGDFNVIRYPEETNKTNRVT TS MRKFNRFDVDAFELKEIPLAQKTFTWSNLQ ARP ICSKLDRFLFSVDWEEHWGNLTSRALPRV TSD HWPIVLCKEERRSGGPAPFRFENMWLLHP GFT NRVREWWNECNPLVEWEGMRFHKKLQFL KA KIKEWNILEFGRLEERKRDLMDQLESLSKE E RGEFGEEDNELRRGLRRQLEDCLHKEEIS WRQ KSRCKWLQEGDRNTRYFHSLADGRRRRNF ISK LVVDGIELTQQEEIAGACVEHFKSILRRPV SGR VWVEGLDWPDVGGDNRRSLERXFDLEEI KEA VFSMDRTKAPGPDGFTMAFYQDCWDILKE DL LMVFEEFHRRGIVNPGVNSTFIALIPKKKE ASH PRDFRPISLVNSLYKIIAKVLSRLKKVLPD IIEQ	cd0165 0	398	663

			AGAAG	TGTTTCA	NQGAFVSGRQITDQILIANESVESFKKDKKKG		
			TGAGA	ATGAATT	VVCKLDLEKAYDMVDWEFLRKVMTKMGFGA		
			ATGAA	TAATGAA	RWCKWIMGCVRNCHYSVLVNGTVVDFFPATR		
			GATAC	TCTGTTA	GIRQGDPLSPFLFLLVADVLGKLIHRADSRGLI		
			TAAGT	TTGATCA	RGWRVGREDIPLTHLQFADDTILFCEGGNVSFI		
			TGGAA	AAAAAAA	RNLKLVIQLFEQMSNLKVNFGKSCVVGVNLEE		
			TGTGA		GEIRESAEVLGCEVSGWPLSYLGLPLGDNPKR		
			GGGGG		ESFWNPVVERFEKRLVSWKKNYLSLGGRTTLI		
			ATAGG		KATLANLPIYFLSVFKIPXKVSSRLERLMRDFL		
			AAGTG		WEPGEEKRDHRVSWKVVTTRSKNYGGLGLGNI		
			CGAGG		GKRNSALLGKWLWRFPLEQEKLWARIVASIH		
			AAAAG		GMEQNGWDAGIRARTTSTSPWKAIHDLLPVFS		
			AGCAA		PLVRYDLNSGNSIKFWEQIWVGNVPLKDQFPS		
			CAATA		LYALSDFKNEPISSFITPSSSSSGFGWNLGLRRA		
			AAAGA		VRDSLDLSISSLLELLDNVFLDPSKVDSRIWTP		
			GGTAA		TPTGSFSCASFVKVLCQSTSPATSPFPVNEVWC		
			TAAAG		RPIPPRVGSFCWEAYLERFNTGDLLQKKRPNT		
			AGGTT		CLSPHWCVLCKADGETGFHLLVSCQFAKSIWE		
			CGACC		WAFSLFGRRWNPPSDLQSLFSPSFFRWHMKRG		
			CAGAC		RTMWRLRAAVCWSLWEERNRIFSEDERGA		
			ATTGT		EELRAVIHIRAVQWLKSLELFTDYTVEDMLRK		
			ACTGT		WGEIANCTQTKPVIRVSWTPPPAGFLKLNFDG		
			TGCAG		SCRGNPGQSGIGGVFRDENGVISALFSGPIGVG		

				GAGTC	DSLRAEILAVMEGVQKAKEMNVRKLIIEGDSE			
				CAAGA	VVIGWLIKDNCGLWNYXNERRKFRWLVREM			
				GGATG	EIRMSWVKRSANSLADGLAKQGIEREAMFWA			
				GAGTT	KL			
				TAATA				
				ACAGG				
				TGGGC				
				AAGGA				
				GTGTA				
				TGGAA				
				AGCCA				
				GGGGG				
				ATTAG				
				TTGGG				
				TGGCT				
				GCTCC				
				CTCTT				
				GGGGT				
				GCTTC				
				AGGGG				
				GGATA				
				GTGAC				
				AATGT				

				GGAAG AGGGA ATGTG GTTGA AGGCT TGGAG GAACT GATTG GAAGA TACTC AGTTT CCATG AAGTT CAAAC AGCAG GAGG					
L1	L1- 10_ NN		Nelumbo nucifera	GGAGA TTTGT AGGGA AGATT CAGTG GCAGA ATTTG	TAGTTTC ATATCTT TAATAAA GTTTCTT GTTATTG ATCAAAA AAA	MRIXSWNVRGLGSARKRATIKEVXKKFNPDV VLLQESKRMDDFDCRRARSVWRAKGIKWAAAP SWGSSGGIVTLWKEDVLIGLEELIGDHSVSLKF RQVEDGFIEWVLTNVYGPNDYRQRKKLWEELS DIRGLWEGPWCLGGDFNVVRLTEETNNPRQR VTTSMRSFNQFVDXFELKELPMLNNTFTWSNL QSSPICKLDRFLFSTDWEAHWGNLTSRVLPR	cd0165 0	498	763

				TGGGA AAGTT GGGGC GTAAA CTGAA GAACC CAAAT GATTT AAAAG CATTT CTACG AAAAA AGGAG GATAG AAAGN TCGAG AAAAG AAAAT TGAGC GAGGT TGAGA GATTA AGAAG	TSDHWPIVMCKEXLRGDGPAPFRFENMWLTH PGFIEKIKEWWQHYNPPVQWEGMRFHKKLQF LKSQIKEWNRLEFGRVEERKQVLVEQIEELDM KEGRGLFGEEDRERRWSLRKQLEDLLIKEEIS WRQKSRCRWLKEGDRNTGFFHSVANGRRRK NFISKLEVEGTVISHPVEIAEAFVGHFKSIFRRP LNERIWLEGLDWPTLSPENREWLERRFXLEEIK EAIFSMDRTKAPGPDGFTMAFYQDCWGIVEED IGSVFEEFHRRGIINPGLNSTFIALVPKKKEASQ PRDFRPISLVNSLYKIIAKVLSRLKKVLPSTIAP NQGAFVMGRQITDQILIANESLESFRKSKKKG VVCKLDLEKAYDXVDWSFLLEVLSFGFGDS WCXWIWGCIRSCHYSILVNGTVVDFFPASRGI RQGDPLSPFLFTLIADVLGRLIQRADSRNLIK WRVGREEIPLTHLQFADDTILFCEGGDIEWIRN LKLAIQLFERVSNLKVNLDKSSLAGVNLEERE LREYADVLGCATSDWPMAYLGLPLGDNPKKV SFWNPVVERFGKRLASWRKNYLSLGGRLTIR STLANLPVYFLSVFKIPTKVSNTLEKLMRDFL WEPGEEKRDHRVSWEVVTRGKKLGGLGIGNL RKRNMALLGKWLWRFPLERDKLWTKIVASIH GLEDNGWEPGILARTTAASPWSAIHKLLPVFSP CIRYGLNSGNKIKFWEQLWVGDSPLMDQFPSL			
--	--	--	--	---	---	--	--	--

				AGATA TTGAG GGGGA TAAGA AGTTT GGGGC CAGAG GGGAC AGAAG GAAAG GGCCC AAGTA GATTT TGGTG GTGAT AGGGC AGATT CTTGC AAAAT TGTG		YALSDLKNEPISSFLSPSDSPSGFGWNLGLRRAI RDTDL DHILSLLAILENVFLDPXKEDTRFWVPS SSGVFSCASFFETLTQPSDPIGLSFPVKEIWSRPI PPRVGSFCWETFLERINTNDLLQKKRPNSCLLP NRCTLCKADGETGYHLMVTCSVAKSVWEWV FSLFGCVWXP PGSVQSLFSPSFFRWPKKRGKT MWRSIRAAVCWXLWEERNSRIFSDKERTIGNL LDLILLRVAHWLKTMEFSDYSVADXVRNWK EIASSALCKRTCSPSWSPPPSGFIKLNFDGSCXR NPGPSGIGGVFRNEAGAIAMFSGPIGEGDSL R AEILAALEGIQRARELNIRCLVLEGDSEVVIRW LKG GGSGLWSYNNEKNKFKWLSRDIVVHYN WVKRTANSFADGLAKQGVNRDSTFWANFLXF LVFFGGELSLFTSPLYLSFT			
L1	L1- 17_ NN	.	Nelumbo nucifera	AGTGT TTGGA GAGGT	TGAATGT GCAGAGT TTATTCA	MWKKEVLEGKEELIGQFSVSLKFRQVEDGFE WVL TNVYGPNDYRERRQLWEELADIRGLWEG PWCLGGDFNVIRFLEETNSTRQRSTMSMRKFN	cd0165 0	426	691

			AGGGG	GGTTCAT	QFVDAYGLKEIPLSHSCFTWSNLQVNPICSKLD			
			GATCA	GCCGAGT	RFLFSNDWEEHWGSLNSRTLPRVTS DHWPVI			
			ATTGG	GGGCCAA	CKEEVRNAGPTPFRFENMWLGHPGFIKKVKD			
			GCGTT	CAAAAAG	WWLESNPTVKWSGMRFHKKLQHLKGRIKEW			
			TCTGC	GGCCAAG	NLLEFGRLEERKQMVTAQIEALDSKEGRGCFT			
			CCTCA	GCTATGT	EEDKEQRWNLRKQLEDCLLQEEISWRQKSRC			
			TGGGG	GGAGATG	KWLKEGDRNTKFFHAMANGRRRRNYISSLEIE			
			GTCTG	TATGAGG	GATISNQSEIAGAFANHFRSIFRRPLNNRVWIE			
			CAGGG	GCTGCAG	GLEWPTLKLEEREWLERKFXLEEIKEAVFSMD			
			GGTGT	TATGCTG	RSKAPGPDGFTMAFYQDCWEILKEDLRIVFDE			
			TGTGA	GACTGTT	FHSRGVINPGLNSTFIALVPKKRNANHPRDYRP			
			CA	TGGGAGG	ISLVNSLYKIITKVL SIRIRKVLPTIADNQGA FV			
				AGAGGAA	KGRQITDQILIANESLEEFKRSKKKGVVCKLDL			
				CTCACGG	EKAYDMVDWSFLLEV MVNFGFGDRWCKWV			
				ATATTCT	MGCIRKCHYSILINGTVVDFFPALRGIRQGDPL			
				CAGAAAA	SPFLFTMVAEVLGRLISRADSRDLIRGWRVGR			
				GGAGAGA	ESVPLTHLQFADDTILFCEGGDLNRIRNLKLT I			
				ACCTTGG	QLFEQMSNLKVNFEKSCLVGVNMGEGETSEC			
				AAGTAAT	ADILECAVSNWPLNYLGLPLGDNPNQKDEFWIP			
				ATTGGAG	VVERFEKRLATWKKNYLSLGGRITLIRSTLAN			
				GTAATCC	LPVYFLSVFKIPAKVASRLEKQMRDFLWEPGE			
				ATATTAG	EKKDHRVRWEMVCRSKREGGLGIGNIRSRNL			
				GGTGATA	ALLGKWLWRF SLEKDKLWVKIVGSIHG VADN			

					CATTGGC	GWDSGIVARTTRAAPWKAIQAILPLFSPWIRY			
					TGAAAAC	DLHSGNLINFWEHIWIGDSPLKDQFPSLFTLST			
					TTTGGCC	LKNEPVSSFLVPSNSTSGFSWNLGFRREIRDTE			
					TTGTTTG	FDCILSLLSILEGVFLEPTKTDSRIWSLTPSGSFT			
					CAGACTT	CASFFKALTQPPSLGLSFPVKEVWCLPIPPKVG			
					CTCAGTA	SFCWEAFLERVNTCDLLQRKRPNSCLCPHWC			
					GATGAGA	ALCKADGESEYHLLISCRFTKSIWEWCTSLFGC			
					TGGTAAA	RWPPL			
					AAACTGG				
					AAGGAAA				
					TTGCAGA				
					ATTAGCC				
					TCAGCTA				
					AGAAGGC				
					ATCACTC				
					TCTTGCT				
					GGAGTCC				
					CCCACCT				
					AATGGCT				
					TTMTTAA				
					ACTGAAT				
					TTTGATG				
					GCAGCAG				

					TATGGGT				
					AATCCAG				
					GTCCATC				
					TGGAATA				
					GGGGGAG				
					CAATAAG				
					AGATGAA				
					CATGGAA				
					ATCTGTT				
					AGCTATG				
					TTTTCAG				
					GACCAAT				
					TGGGGAA				
					GGAGACT				
					CCTTGAG				
					GGCTGAG				
					ATTCTTG				
					CAGCCCT				
					TGAAGGG				
					GTTCGAA				
					GAGCCAA				
					GGAGATG				
					CAAATTA				

					GAAAGCT				
					GATCATT				
					GAAGGAG				
					ATTCTGA				
					AGTGGTG				
					GTGAACT				
					GGTTAAA				
					AGAGGAT				
					AACTCAA				
					GGATATG				
					GAAGTAC				
					AGCAATG				
					AAAAACG				
					AAAATTT				
					AAGTGGC				
					TGGCTAA				
					GGATGTT				
					GAAATTC				
					AGTACAA				
					ATGGGTT				
					AAGAGAA				
					GTGCAA				
					CTCACTG				

					GCAGATG				
					GTTTAGC				
					TAGGCAA				
					GGGGTTA				
					ATAGAAA				
					CTCTTTGT				
					TTTGGGC				
					TTTATTGT				
					AATTGTA				
					TTTGCTG				
					CTCATAG				
					ACTGTTT				
					AGGGAGC				
					TAAGATT				
					GTTTAGC				
					CCCCTTG				
					CTTTGTA				
					ATTCTTCT				
					TAATTTT				
					AATGAAT				
					ACATTTC				
					ACAGTTA				
					TTGATAA				

					AAAAAAA AAA				
RTE	RTE - 7_N V		Nematoste lla vectensis	TCCTC GATGG GATCG CCACC TTATC GTGGT GAGGG TGTTT GTGTG TCCCA ATGAC CTCTA GAGCT ATGTC GGCGG GAGTA TTTTAC TCCTG GCAGG GACAC CCATG	TAACAGG TAAC	MTRRGQSHPDAPGLTKPLVHPKHAMRIGNWN VRTLYSSGNVAQAAREMGKRDIDIMGISETHW TGQGKMQLRDGETMVYSGREDNIHREGVGIL MSKNAAACLIEWTPWNERMIQARFYSRHIQLT IIHIYAPTEDEHLKDEFYTRLQEVLDNRNQH DMLIITGDMNAKVGDNWDYERVMGKHGLG VRNDNGERLCELCDLNELVITGTLFPHRTIHK TWISPDGKTKNQIDHVLVNRFRNSVKDTRVF RSADVGS DHYLV CATVKLR LRKDPKRKGNIR VKYDTSKLTEKEVRNTFNIKLNRNFQALEQED EEERGE EVVDAKEEVERDFTIMKKALLEVAET VLGKPRKKNKPWISGESWSLIDERQQMNEKIL STRSERVKGQLRKKYREKDREVKRSLKADKK KWLEEIASEAEKAARSQHMKTLYGLTKKLCN ERPRHSTAVLDKDGNNLSKKEEIQRRTWTEHFR EVLNREQPDNPIEPEDEVEFEFGELIEEISVN TFGEVKDTITTLKNGKSPGIDNITAE LLKADPE FSAERVHKLLVKVWRHETIPNDWKKGLIILP KKGNLKECKNSRGITLLSIVGKILGRIVIDRIRN GVDDKKLRKEQAGYRRGRGTTEQVFILRNIEQ VNEWQATLYVNFIDFEKAFDSVHRESLWVIME	cd0165 0	567	824

			CCGAA	KYGVPEKIIRIVRLFYEDFQCAVEDQGETGEWF		
			CAGGT	NIKTGVKQGCNMSGFMFLLVMDWILRRSVGK		
			CGAAA	GENGIRWKFTSKLDDLDFADDIALLSSTKHHIQ		
			GGTAG	EKTAKLDGEARRAGLKINLEKTKMLRINAKQ		
			GGGCC	HGNIVINGKEIEDVSEFTYLGAKVCREGGGMK		
			AGACG	DLRNRLSKARSVYTRLKRIWNSKKIMRRTKIKI		
			AAGAG	YKSLVLSVLLYGCETWKMNNKKDDKLLDVFQ		
			TGATC	QKCLRRILKISWEDHVTNEEVQRRAGVHQLSS		
			CACTG	EVKRRRWKMIGHVLRQGRENDCSVALTWTPE		
			GTCCT	GRRKQGRPKTTRRTVERERGEAGWRNWDE		
			CCAGG	VRSKAADREKWKSTVKALCATRHEEDR		
			TTGGG			
			GGTTG			
			GGCAA			
			AGGGT			
			TGATA			
			ACCCT			
			CTCCC			
			ATAAA			
			AAATA			
			GCTTA			
			TCACA			
			GAAAC			

				CAGAA GCAGA GCAAA TTCAC TTGGG AAGAC CGTGG CTGCA TCTCA TGAAA GAGAT TGT					
CR1	CR1 - 20_ NV		Nematoste lla vectensis	GGTCA GCGGT CGACC AAG	TAATAAT ACCTATT GATATAT ATATGCA TGAAAAG ATATATT TATTTTTA TTTAACA TTTACAT TAACAAT TAACATA	MAANFLTPVFIAFLAILFNQKLHYSPNLTSGIN GFGLNCILANSESILTGLEAFYSHNIRANKMNV GRSSSEHKHLLLCLLLLGGDIQLNPGPKWKFP CGSCNKPVKSNQKGIQCDNCNLWYHTRCCSIN DDLYYTLAASSCTWLCVSCGFPSFSSSLFESSI ELQNSFSSLDLHQDTQNISSSTIRSPETSPLDSS SRNSSNTRKQSKEKLTGMILNCNGLKGTDHIT KFQALINLHDPDFILGTESKLCPDISTYSIFSQN YSVFRRDRNKFGGGVFQAIKSDLVCTEEPDFE VDNCEIIWTSCLKLANRKSIIYMSSYYRPPNQPSD SIDNLNESIRRVFNKVKNFPNIIIGGDFNLGDIE	cd0165 0	707	959

				TTTTTTAC	WNSSVPTPTNPATASLHQKFLQTIDDFSLSQHV			
				TAGTTTC	KTTTRPSSGKTLDLLLTTYPNVIVNTSTASGLS			
				TGATCTT	DHLAVIFEVNLKPSRSLKPPHKVYQYKKANFP			
				TAGGGTG	DLCKFMSDSSTTFFSTNPEKRTVEENWVLFKT			
				CATGCCA	SLSTGISQFIPQKSSRPKYKLPWINTNIKREMRK			
				TGGATGC	KDRMHKKAIKSKNQQHWEAFKRQRNLVSKLI			
				ATGGAGC	KDSHNDYLNNIIGNSLKDNPKKFWSYVRSSKS			
				GGCTGCC	ESIGIPPLKNNDSSVSVTDAHKAETLNTFFHSV			
				AAACTAG	FTQEQLPTPVLGNSPYSSIPDLHISAEGVCKQLS			
				CGGGAGT	ELNPKKACGPDEIPANVLKEVSQSISHWLAFIF			
				TGGCTCA	QQSFDCNKIPSDWSNALITAIHKKGLKSEPANY			
				AGGGTGA	RPISLTSICCKIMEHIILSHMAKHLSSHNILINQQ			
				GGCAAAG	HGFREKFSCETQLISAIHDWAKTINFRGQTDVL			
				ATTCTTT	LLDFSKAFDSVPHQRLLIKLDFYGIRGNMLNWI			
				GTTTCGC	KAFVSNRKQSVSVNGTVSSSKSVVSGVPQGSV			
				GCTTGGG	LGPVLFLLFINDISTSISQSNLRLFADDCVLYREV			
				TTGATTC	ASLQDCQALQQDLDRLSVWSKTWQLKFNVSK			
				TTGTTGG	CYHLGITRKKIPVKFTYSMDGKCITQASSTPYL			
				TTGGCTG	GVTIADNLSWNEHCDNICKKANSTLGLLKRIL			
				TTTCATG	SGCSPEVKDSAYRTLVRPKLEYATSAWNPYTQ			
				CATTTAT	CNINKIEMIQRRAARFVFNDYSRYSSVSPMINS			
				GGCATTC	LGWESLEQRRQLNQAHMFYKIYNGLVGISFPP			
				ACTCTAT				

					TGATGAT	EVTPITRVSRQPNCAPFHQLVTLNDTYKYSFYP			
					AATTTTA	RTIRTWNSINLSTIPDSLAKFKDSLKLVQTN			
					ATTTCAT				
					GTATATA				
					TAATTAA				
					TTATTTA				
					ATTTTAT				
					CAAATAT				
					CTAATT				
					CAATAAT				
					TTTCAAG				
					TAATATA				
					TATTTAT				
					GCTTCAT				
					TTTCAAT				
					TTAACCT				
					AAATTAT				
					TTTTCCG				
					ATGGCCG				
					ATAATTA				
					TTATTATT				
					ATATTTTT				
					CTTTTGT				

					ATTTACA TATTGTA TTTTCTTT TTCAACT CCTAGTA TAGTCTT AACTGTT TTTTAGG AGTATAA CTATTAA AGATTAA AGATTAA AGATTAA AG				
CR1	CR1 - 18_ NV	.	Nematoste lla vectensis	CCAAA ATGGC TGCCA CGGCT TTTGT AGCTC TGTGA ATCTT GTCTC	TAATAGT ATAACGC TATATGT GCTGGCC CTTTGCC TTAATTC AGTCTGT AATACTG CCATGGC	MINNPSNVVPVYFCLPILAFWSFLPCYFHLKAV DRGSHVATLTSEPVASTGWSYSNQELKTIRKS QPKQRLEPKLWQKLGD LGIRKPFRSRRKPRQS RHVCTGLPVVTTVDSTGHPQLSSGCSTNSSSSQ TARFPVIPDTIIANMQSLLPKLDELYAAIMQFP AGIICLTETWLNESHDPDPLIHLPGYLCFRNDRT TSRGGGVCTYINSQYPCRR LPEFEDLSIESLWIS VRPYRLPRSVSTLLIGVIYHPPSNGEMNRQLVD HIDKNCESFLSKHPDGLVVVTGDFNPSSTGIKA	cd0165 0	612	864

				TGTTTT	AATTGAA	QYVTHKTGLTQIINVLTRDTGILDWCLTNKPK			
				CATGT	TAAAACC	YFQQPLQLSKMGSCDHYPILIKPNLSVTNPKAH			
				GTCAT	CCTATCT	KSSVWRRNLRRSNLDAFGRWITQFCWNTILSL			
				TGCTA	ATCTATC	QTAREKFERFYTVITDAVNTYLPLQRVQVCAT			
				GTTGA	TATCTAT	NKPWIGTSLCSLIARRQKALTTLGKDSRLYKEL			
				AGGCG	CTATCTA	RNQVQRECKVCRERYYPNKVAALKESNISRW			
				TTTGA	TCTATCT	WQEVKCLSGSVVSNEWWHQLMDDSTPSVQIL			
				CTGTC	ATCTATC	AERFNEFLLGLTSNFAPISQQVTCSGPVPELLI			
				AAAG	TATCTAT	TPGKAFKALSRIKVKKSSGPDVPVSVIWKEFAP			
					CTATCTA	ELSDVVCNLYNSSLMEGFVPQELKESIVHPVP			
					TCTATC	KCSPPKTLSEDLRPITLTSQLAKVMEGFALDSL			
						FACVKDKLDPKQFSVSGKSTVLALVYFLHCIL			
						ESLERGENYVRIFFADFSKGFDLVDHNVLMSE			
						LDLLGVNEYLQNWIAAFLTARPQRVKISGHTS			
						SPVFPNGGIPQGTKLAPLLFAILVNRLASTWPY			
						RIKYVDDTTIFEVIPRCSPSYLPFIANDICNFATE			
						RGMRLNPKKCRELVINFLQYQPSPVNALQLM			
						GSVIKRVESYKILGVHISDLSWNKHVDYVVK			
						KANKRLYALRLLKKSGVPVQDLIAIYCVLIRSV			
						VEYAGPVWANITGFLVDIVEGIQKRALRIVLPH			
						LRYKEALETSGLQSLADRRRELCISLMTSAKEL			
						EPLRSVITGGVINNVHGYSLRSGSMNTYKHISV			
						TKRFGDFVTYRFL			

L1	L1-17_NS	Nicotiana sylvestris	CTCTC AATAA TGTGA AAACC TTCCTC ACTAC CGCTG CTCCT ATAGA ACACA CCTTC CTTTA CCTCA ATCTC TTCAA CACTT TACAC WCATC ATGTC TGACT CTCAG CAACA MCTGC	TAATATT AATATAT GTTTATC TTTTAAA CCAAAAA AAA	MJKALHGGPWFILSHFLTVRGWEPKFKASSTQ LTYSALWVRLPELPTEFYDLEILSRIGSKLGRLL KIDTCTSSSTRGRYARICIEVPLEKPLKTHLYIG NHRQSLLYEGLNLLCTYCGRFGHPTTSCPEVV RAPNEQNXRXQXESNSTPDTIPSKENNEWKTV TFTKKSIHPPKHLNFDKGKAATMAATPAAQPL XGQSSNPTSAVLGRVTSTRQVXLQGPVIEFNN KFQXLSNAGPSDLQTDHFYXGPIPKNJGTEPLL TQQXAKAHFHTPDKLPTSLPNXLPNDPKPRLS EPNPHLTKSPPLVGYNTFSQDYXTPHZPLPCNH LLTSSTQSSSNSEGPLHILDTLXPTQNKAIXT MHATVTSLSDSRLPQNPKQHDEFPCHEHHHE XPTTLFPSLPNIYXHSIQTXKPNTTLXAEPQXLP NPILNAETNLHRHEPESVSEGKYTNQQPVFSSP SSHNGGXSTASEFITSNLHISAHDHGPPSPILVG NGTAGICSNIHSATQQPRGGVANXKSDSFSRN SDRRNEAGNGSLQSKSMVHDPDSSSNGTPGFQ FHSSAMAI SNGPDESDLGPN SQSP TLLPTSSYG SGLCTLVPATGLIQSTLPLLEPTESSRESHPTKQ SSTITSSISTRTRQRYKRGTD SN SCGDEGDKNRI LVVKRDPILQPQRTVRKKFILKCPPEMVTC SLN FRQTTNPEVGKYAMVIIPTLLQLPLSLDPMTSS VVHPVTQAPMSQPRNPNSYMNFLVWNCRGSH	cd01650	1264	1503
----	----------	----------------------	--	--	--	---------	------	------

				AACAG CTTAA TCAAG TTWCC GACCC TCCGG ATATA GGATC TATGC AAATC GATGA TACAA GTATG CATAC ACAAC CAGCT ACTTT CCTAC AAACT CTACT AACAG ATAAA TACCT	NPEFRRHFRSLLDNHRPTLAILLETHMQDHASL REDFNFGNMAQAPANGLIGGLVVLWNADLIN VTEMRISDQEIHCMVQVSPTKPPWILSAIYASR HYQLRDILWNSLINLHDNTAMPWLIGGDFNEI LESKEKFGGLPMNNNRADKFANFMHXCSSLIDL GYQGSRFTWTNKQRNGVPYTKKWRTILERLD RFLANYDWLNLXPNTTVIHLPRTHSDHCPLLIN LEPPQPIQNRIFRFETMWASHPQFQQLVHATW SNDVPLLATTDNFKRVVTRWNQNTFGNIFQQ KRKLLARLAGLQSSVHYPTSVFLQNLETQTL EYSHILRIEEEFWQLKSRINWLNEGDANTRFYH LSTLHRRRRNRINALQDNVGNWVATTXEIHD IXNYFXDLFQTDQLISENMTGNTYNCEISSDDI LXLXSPLHQDEILTALNSFKAYKAPGPDGLHPI FLPKILDGCRIVLRHTKPDQMGYIQFFYQKY WTDVGDKVSSFCQDVFCSKIIPPELNKTLICLIP KVKHPTRIQQFRPISLCNTVYKLITKIIVNRLKPI ISKIIGPTQTSFQQGKRACDNAIIVQESLNYLAK MKGKYPSMLLKLDLEKAFDRLEWSFIRRALFY FNIPPSMISLIMSCVTTTSSISILINGSKTQYFTPTR GIRQGDPLSPYLFIMCMELFSRSISLSVDYLKW QPIRLSNKGPLLSHLFFADDIILFSKITTKSCHAI DVLNNEFSKYSGQKINFEKSKLFFSKNCSSLIKIL		
--	--	--	--	---	--	--	--

				TCTGC	LLLPLTLKKEPLL VNTWDFLCSKQDPKMQISSS			
				CAWCT	SLIISKINLLDKDFVITSFNIKEGTSFGKYLGFPM			
				CTAAC	FQARPKNADFQFLLDNFKNKL AGWKS RFL TL			
				CAACC	AGRATLIKSTLNQLPNHIMQYIHIPAHILNKMD			
				AAATA	QYQRNFLWGT TLEKKKLHL VRWNEITRPKEA			
				CCTTC	GGLGIQKL RQKNHALLASLAWRLFKTPQAMW			
				CTTAA	ANILLTKYLN TSPNNIHSQIWSNVINGWHQCQ			
				ACACT	LGIKW NIGKGTHIKFWEDPWIGP GTTIRSKIHG			
				CTAGA	PLPLDQQNITISSLINNKGWDWNKLPFELPITLL			
				TCTAG	KVAHNITQIIQATYISKYSNKVDNIFWGLTPTG			
				CAGAT	HFTTRSMYSILEAQHDVSTNSSTKYNWIWKLP			
				GAAGT	VQPKIKGFVWLLHKKRLPTKDYL RHINLIQDA			
				ACTCC	TCQFCHQEDETIRHLFLDCTNAKTFWQDLGIL			
				ATAAC	DFINQKLHYTNSFNWVHDLITCKDFHFPYQMN			
				MCTAC	PATYLPLSLWNIWLARNNNCYKHTMSTASAK			
				TGACG	LTTQQA AKFTYL VAKKHTXVKKRQQNLKWK			
				ATGAC	PPKEHFYKLNTDGAHTTSKAGIGGLIRNSKAE			
				ACATT	WILGFAAAVPXDSSTS AELRALLQGLQLAYSH			
				TATAC	NLVPLEVEVDSQDIITLLYSNNVSFDNLTTDCR			
				CAATT	YFLDRLGKPVVLHAYREQNRVADKLAKAGCN			
				ACATC	FDIHSVVRILEDNPDFAWSVFEQERSAVQGHV			
				AACTG	LEHFRPHSDNYDTVNRYCNSDAVGSSINIGLY			
				ATAAA	NTLAPTTGASACNRMQYRHPPSVRNT			

				TCACG					
				TTTAT					
				ATGAG					
				CCMTG					
				GAAAT					
				ACTCT					
				GTGAT					
				AGTTA					
				AGCTW					
				TTTGG					
				CAAAM					
				GGATT					
				ACACA					
				CCAAA					
				TCATG					
				CATGA					
				TAAAC					
				TGATG					
				ACTCT					
				GTGGA					
				AACCA					
				ACAGA					
				GGTCC					

				TCCTCT CATTG ATTTG GGTTC TGACT TTTTTC TCATT AAGTT CCAAA GAGAA GAAAA T					
L1	L1- 4_M E		Notamacro pus eugenii	AAATT CCTTTT ATGAT ACAAA TATGG TWTTG ATACC TAAAC TAGGA AGAGC	CTTGGTA CTGGCTA AGAAATA GAATGGT AGATCAG TGGAATA GGTTAGG TACACAG AAGTCAA TGACYAT AGTAATC	ERKLYNFPNEYRCKNFKYQKDYNNISQEYIMX KXDLYQECRXGSILGKLXALIISXTKLTKIILSQ MQKKAFDKIHPFLLKTLKSIRINGSFLKIISSIYL KPXASIICNGINMPFQDGNKDVHYHHYYSILY KCLQXKKKKLKELEAKKKQNYHXLQMIYTRI LENQLKNYLKLSTKKLLEIINNFSKVAGYKINP HKSSAFLYITNKTQQQEIEREIPFKXTVDNIKYL GVYLPRQTQELYEHNYKTLFTQIKSDLNNWK NINCSWVGVNIIKMTILPKLIYLFSAIPIKLPKLF YRARKKYQNSSGRTKGQEYQGNKEMLGKVA PPXILNYIIKQQSSKP	NA	NA	NA

				CAAAA	TACTGTT				
				CAGA	TGATAAA				
					CCCAAGG				
					ACCCAG				
					CTTCTGG				
					GATAAGA				
					ACTCACT				
					ATTGAC				
					AAAAACT				
					GCTGGGA				
					AAACTGG				
					AAAACAG				
					TATGGCA				
					GAAACTA				
					GGCATAG				
					ACCAATA				
					CCTGACA				
					CCATATA				
					CCAAAAT				
					AAGGTCC				
					AAATGGG				
					TACATGA				
					TCTAGAT				

					ATAAAGG				
					GTGATAC				
					TATAAAC				
					AAATTAG				
					GGGAGCA				
					AGGAATA				
					GTTTACT				
					TGTCAGA				
					TCTATGG				
					AGAAGGG				
					AAGAATT				
					TATGACC				
					AAACAAG				
					AGATAGA				
					GAACATT				
					ATGAAAT				
					GCAAAAT				
					GGATAAT				
					TTTGATT				
					ACATTAA				
					ATTGAAA				
					AGGTTTT				
					GCACAAA				

					CAAAACC				
					AATGCAA				
					CCAAGAT				
					TAGAAGG				
					AAAGCAG				
					AAAAC TG				
					GGAAAGA				
					ATTTT TA				
					CAGCAAG				
					TGTCTCT				
					GATAAAG				
					GCCTCAT				
					TTCTAAA				
					ATATATA				
					GAGAACT				
					GAGTCAA				
					ATTTACA				
					AGAATAT				
					AAGTCAT				
					TCCCCAA				
					TTGATAA				
					ATGGTCA				
					AAGGATA				

					TGAACAG				
					GCAGTTT				
					TCAGAGG				
					AAGAAAT				
					TAAAGCT				
					ATCTATA				
					GTCATAA				
					AAAAAAT				
					GCTCTAA				
					ATCACTA				
					TTGATTA				
					GAGAGAT				
					GCAAATT				
					AAAACAA				
					CTCTGAG				
					GTACCAC				
					ATCACAC				
					CTATCAG				
					ATTGGCT				
					AACATGA				
					CAAAACA				
					GGAAAAT				
					GATAAAT				

					GTTGGAG				
					AAGATGT				
					GGGAAAG				
					TTGGAAC				
					ACCTAAT				
					TCATTGT				
					TGGTGGA				
					GCTGTGA				
					ACTGATC				
					CAACCAT				
					TCTGGAG				
					AGCAATT				
					TGGAACT				
					ATGCCCA				
					AAGGGCT				
					ATAAAAA				
					TGTGCAT				
					ACCCTTT				
					GATCCAG				
					CAATACT				
					ACTTCTA				
					GGTTGTA				
					TCCCAAA				

					GAGATCA				
					TAAAAAT				
					GGGAAAA				
					GGACCCA				
					CATGTAC				
					AAAAATA				
					TTTATAG				
					CAGCTCT				
					TTTTGTG				
					GTGGCAA				
					AGAACTG				
					GAAATTG				
					AGGGGAT				
					GTCCATC				
					AATTGGG				
					GAATGGC				
					TGAACAA				
					ATTGTGG				
					TATATGA				
					ATGTAAT				
					GGAATAT				
					TATTGTG				
					CTATAAG				

					AAATGAT				
					GAACAGG				
					TAGACTT				
					CAGAAAA				
					ACCTGGA				
					AAGACTT				
					ACATGAA				
					CTGATGC				
					TGAGTGA				
					AGTGAGC				
					AGAACCA				
					GGAGAAC				
					ATTGTAC				
					ACAGTAA				
					CAGCMAC				
					AGTGTGT				
					GAKGACT				
					RTTTTTG				
					ATAGACT				
					TAGCTCT				
					TCTCAGC				
					AATGCAA				
					KGATCCA				

					AAACATT				
					TCCAAAG				
					GACTCAT				
					GATGCAA				
					AATGCTA				
					TCCACAT				
					CCAGAGA				
					AAAAC TG				
					TGGAATC				
					KGAATGC				
					AGATGAA				
					GCAGACT				
					TTTTTTTT				
					TATTCTG				
					TGTCTTCT				
					TTTACAA				
					CATGACT				
					AATATGG				
					AAATATG				
					TTTTACA				
					TGATTGC				
					ACATGTA				
					TAACCTA				

					TATCAAA TTGCTTA CCATCTT AGGGAGG GGGGAGG GGAGGGA GGGAGAA AAATTTG GAACTCA AAATTTT AAATAAA AA				
RTE	RTE - 2_M E		Notamacro pus eugenii	GGTTA CTGCA GCAGG CAAGG AATCC CACAA AAAAT GTACA AAAAC ATCTG CAAAG	TGATGGG CTGGCCA CGTTGTT CGAATGC AAAATGT ATGCTTG CCAAAAA GACTATT TTATGGA GAACTCG CATGGGG	MIPLTISTWNVLTLIDNTKSSRPERGTALVARE LNRYHIQIAALSETRLANEGQLTEVGAGYTFF WSGHSEGECEAGMGFAIKTNLISKLVCLPKG VNDRLMTMLPLAGKHHATIISAYASTMTNPDE VKEKFCEDLETLIINVPKEDKLIILGDFNARVGS NYQTWQGV LGRNGVGN SNSNGHLLLKTCAS HDLLITNTVF XLPKCNKTXWIYPRSKHWHLID YVIVRRRDRQDVRVT KAMCDAECWTDHRLIL SKLNIHIIHQKYCPQGKMTTRRINVNKLELFS EQFVADLEGKLSQHIVGNSGAEKEWAAFRDL VYSTAFAHLGQNT HKHQDWFDENDGEIQKLL	cd0165 0	542	824

					CAGGTGA	NEKRELHRIYQQDSSSTSKKAAFNSIKSKVQA			
					TCACATG	KLREMQDSWLSKKADEIQFYADSNNPKCFYD			
					GTGGTCA	SLKAIYGPKTYGASQLLSADGATLISDKDMILE			
					GAAGAAA	RWAEHFHSVLNRPSINAEIDHLPQVEVNPSL			
					CAATACA	AELPTEEEVLRAIRLLSCGKAPGADSIPAEIYKV			
					AGGACAC	GGPLLIQKLTEIFQVIWQEEVIPQEFKDASIVHL			
					TCTCAAG	YKGKGNRLSCDNHRGISLLVIAGKILARVLLN			
					GTCTCTC	RLILHLEDGHIPESQCGFRKGRGTVMVFAAR			
					TCAAGAA	QLQEKCCQEQNRGLYTTTFVDLTKAFDTVSHGL			
					CTTTGGA	WKIMSKFGCLEKFINIVHQFHDGMFAWVLDN			
					TTTGACT	GQCSRAFPVTSGVKQGCVLAPMLFSMMFSAM			
					GTGCAAC	LTNAFNEDEHSIKVNYHTDGKFFNLKRLQAKT			
					ATGGGAG	KVEGVLVHDFLFADDCALNAASEAEMQQSM			
					ACACTGG	DQFSAACANFGLIINTKKKTQVLHQPPPHHPYVE			
					CACAGGA	PSVTTNGEVLNAVDFKFTYLGSVLSRDVHIDNE			
					CCACTCA	VDTHIARASSVFGRRLRRKVVWERRGIRLTTKLK			
					GCATGGT	VYRAIVLTSLLYACETWTVYQRHARKLNXFH			
					GTGCCCA	LNCLRKILRITWQDKVPDTEVLAQAELPSIQTM			
					CATCAGA	LQRVQL			
					AAAGGTG				
					CTGTACT				
					CTATGAG				
					CAAAGCA				

					GAATTGA				
					GATAGCA				
					CAAAGTA				
					AACACAG				
					GATGTGC				
					AAATTTG				
					GAATATC				
					CACCCCA				
					AATATTC				
					ATATGGA				
					CTATCTG				
					TGCCCAA				
					CCTGTGG				
					TAGAGCA				
					TTCTGAG				
					CTCATAT				
					TGGTCTG				
					ATCAGCC				
					ACAGTCG				
					GACACAC				
					TGAAATT				
					TCACTTT				
					ATCATGG				

					TGATGTC ATTTTGG TCCTCTTT GAAGATG AAGGACA ACAACCA ACCTATT AT				
L2	L2- 1_O L	.	Oryzias latipes	A	TGAGTTG ATGTTTTT ATTGGTT TTATTTTT TTATTGG TTTTATTA TTTTTATT TAATGTA TCGATTT TTATTGA TTTTATTG TGTTTTA GTTTGTT GTTTTGTT TTTATTCT	CLPPSLLVSLSSFELELFELGCSPPVLCVVYRP PKYNKDFLNEFSGLLAEILLKYDRVLIVGDFNI HVCCPDHTMSKDFLSLIDSFNLVQHVSEPTQE RGHTLDLVLTYGLSVSNLVICDAVFSDHMPVL FDITLPCSPCRPRAPARWVRINPSTAALFSASF NQTSVIPQSTCSSTDELSSWLHSTCLTVLDAVA PLKSRQPKVKTEPWLNNVTRAARRECRKAER KFKKDKLQVSLQMLKDSWRHYQATVKEAKR EHLNVILLNRHKPRVLFSTIDAALNAPKSVCV QPTLETCEFLKFFVDKVNNVRALISPPGFDPS VVVSCPVTFDQFEPVTLVLYETVKHLKPSGSP SDAVPPRLFKEVITTLGPAVLAVVNGSLSSGVF PRNCKHATVQPLIKPNLDPSVLSNFRPISKLPF LSKILEKIVYVQLKDFLDENGNEVFQSGFKSL HSTESALLRVFNDILLATDAGDHVVLVLLDLT	cd0165 0	403	669

					GCGTTGT TAAGCAC TTTGGTC AGCTTCT GCTGTTT TAAAAGC GCTTTAT AAATAAA GTTGACT TGACTTG ACTTGAC	AAFDTL DHN ILISRLQQLVGIRGTALQWFRSYL EDRTFCVNLNNVQSSSAPLLYGVPQGSVLGPL LFSLYLLPLGSILRKHNC AFHCYADDTQIYVPI TKKDAHSISPLIKCLEELKIWMALNFLHLNEEK TEVIVFGPSGNSESHPSHTLVTPESHPIALGSLT QSVKSTVTNLGVKMDNGFNLDKQISSVVRSSF FQLRQLAKVKSFLSRKHFETVIHAFVTTRLDY CNSLYFGAPHSSLSRLQLVQNAAARLLTGTWK REHITPVLASLHWLPVHFRVRFKILLFVFKSLN GLAPPYLSSELLHLHSPSRSLRSADQLLLDVPRT RRKLRGDRAFAVAGPSLWNLALPMQVRQARSL NDFKSKLKTHLFSLAFNSV			
L2	L2- 4_O L		Oryzias latipes	AGATT TGTTTT TATTG ACAAT TTTAA TCTTTT CTGGA ACAGA ACCTC ATTCT ATAAA	T	MRGSAGRKANTNWLRLKEAPKPEEAASTTT PPPSFNMALENVRSLSNKTFIINDFIHDYNIDGL FLTETWLSTDASTVLTEASPPNFSFVFSTREGK KGGGTASIMSNSFAVKQITLDEYPSFEHHAFVF NSPPVLCITVYRPPKPSSVFIEEFSEFLASVHTSF NMILITGDFNLHIDGKSDYCTTQFLDLLSSLDF KQHVMQPTHNRGHTLDLVISHGLPTSVSAVV DLAVSDHYCVLFNITGFIKQETSVRTVKKRYL TPEVVEKFITSVRSYSPPVLPAPCDLLVESFNSK LKSSLDSVAPLXSKKIRIKATPPWRTEEIKQLK RRCRSAERLWRKHKLEINYQLYQNQLKLYNQ	IPR000 477	521	582

			ACTGA	TVKQTRNSYFANIISVNKNNPRVLFTTIESLINS		
			TGGGA	NVNKNTMPASASLCEDFADHFRSKIDSVRSNII		
			TTCAC	KGQKSLFIGNDGLILPEEALDSFDLVEAATLGR		
			CCCAG	VFSQVNPTTCFLDPIPTSLFKSFYGSFESEILNIM		
			TCCAC	NSSLQTGIFPAAFKTA VVKPLLKKTNLDP SVFN		
			TGGGC	NYRPVSNLPFLSKVLEKLVFIQLNEFLQNNIFE		
			AGCAA	INQSGFRTNHSTETALLKIVNDXRCA		
			TGTTTT			
			AGCTG			
			CMAAC			
			ATCAG			
			AACT			
			CTGTC			
			CAGAC			
			ATCAC			
			CGGCC			
			CCCAT			
			CATGC			
			CCMCA			
			CCAGT			
			CCAGA			
			CCTCC			
			TCCTC			

				ATCCG					
				TGCAG					
				ACATC					
				GGTCT					
				GTGGA					
				AGTGA					
				CTGAC					
				TGGAT					
				GTGGT					
				ATCAC					
				CCCCC					
				CACAT					
				TCCCC					
				AAAAA					
				ACTGC					
				TTCAA					
				TCGAC					
				AGTGC					
				TTCCTT					
				GCCAC					
				CACCT					
				ACACC					
				CCCCC					

				CAGTC TATCC CATTA AGACC ATTAT TACAG ACAGG TCTTCC					
RTE	RTE - 1_P Po		Papilio polytes		TAAGGGA TAAGGGC AAGTGGA TGATGAT GATTGTA GA	TYVRSAAAGDSDRFGVSVGLHQGSALSPYLFLLVMDALTLNIQEEAPWCMLYADDIVLVGEDEC EVQSRLEEWVRVKLEGAGLKISRTKTEHLFCDF GGPTSFSPIALNGVPLPTCSDFKYLGSLVQSDG GVDRDVKSINAGWMKWRQVTGVTCDPRMP LRLKGQIYKSIIRPVVLYGSECWATKVADERR LHVAEMRMLRWMCVTRLDKVKNVYIRGSL KVAPVVEKLRSKRLAWYGHIMRRNDGDINQK VLRLNVDGYKGRGRPCKVWMDCVKEDMLK KGVDSMTADRVDVWKSSTYCADPP	SSF56 672	18	134
RTE	RTE - 4_P Po		Papilio polytes	ACCGG AGGAC GGGCA GCAGC GTCCC	TGACAAT GATGATG ATGA	MMATRKVRPPVPGGPAVPGYGGSGQGNGGAG GAKNPRLGYGYHTRRPLALATLNTRTLRTDER IAELEELSRRLRWDIIGLSEVRREGEETIILKSGN LFFFREGDQQSQGGVGFIVHKSLVNNVVKIES VSTRVAYLILRISKRYSLKVIQVYAPTSTHSDD	cd0165 0	549	822

			TTTCA	EVETMYEDISRAMHSSKTFYTVVMGDFNAKL		
			AAAAC	GIRDGDELRMGKFGYGRRNHRGHLLAGFMEK		
			GGTGG	EGLYMMNSFFKKQGHRKWTWMSPDGATKNE		
			AGAAG	IDFIMSTKKQIFNDVSVINAVKTGSDHRIVRGT		
			TCTGC	LNIDVKLERCRLMGSTLRPAPVHHNAETFQLE		
			GATCG	LRNQFECLGSSVDDLNDGFVHALQSVGSKFYK		
			CCAAC	THRKSSPKKLSDHTELMKERRDMKLHSSDD		
			CCGCC	AERYRRLNRQISTSIKRDIRRFNTNRVKLAIER		
			TGCCA	NQGSKVFSRDLSIGQSQLLKLKTGEGRVVSSRP		
			AGCGT	ELLSEVERFYGQLYATQEPVTDLVEDPRARLT		
			GGCGA	RHYTEDIPDVSLDEISMALKQLKNNKAPGNDG		
			TTATG	ITAELLRAAGEPALKVLQKLFNSVILEGHTPKA		
			GCAAA	WSRSVVVLFKKGDKALLKNYRPISLLSHVYK		
			TCCCC	LFSRVITNRLARRFDDFQPPEQAGFRKGFSTVD		
			CCAA	HIHTLRQVIQKTEEYNLPLCLAFVDYEKAFDVS		
				ETWAVLQSLQRCRIDYRYIEALRCLYENATMS		
				VRVQSEDTKPIHLQRGVRQGDVISPKLFTAAL		
				DVFKVXKYCHLPETLHRGVGRRFQGPEILVVT		
				YTFLHVTFRVQVIGSLLLLRILIKIGTMLEDLNR		
				ASERVGLRMNMDKTKVMSNVHVSPTPITIGG		
				HTLEVVD EYVYLGQTIQLGRSNFEREVNRRIR		
				LGWAIAXNNRRSGSCGISSRPKTYRSVSRQK		
				SLTSVCCQHTGQRRGHSLRATSGGSKSLKGQ		

						WSGQCSESLCEIASGTMRSAAESVTDIARRIAN LKWRWAGHIARRTDGRWGRRVLEWRPRIGR RSVGRPPARWTDDL VKVAGVHWMRAAQDRA LWQSMGEAYVQQWTTYG			
RTE	RTE - 6_P Po		Papilio polytes	GTACG TGCGA CTCGT GAKTG CTATG TATGA	TAAGGAG ATGATGA TGATGAT GATGTAT G	KSEHSRQVRSRRNRQVQC GCPAPGIGXSPYL FLLVIDALTVEIQEEAPWCMLFADDIVLVGED ALEVQSRLEWRDKLETAGLKISRAKTEQLFC DFGGLSSFIPIALDGAHLPVCSDFRYLGSLIQSD GYIDRDVTNRIKAGWMKWRQVTGVICDPQIPL KLKGKIYKAVVRPVMMYGSECWAVKGTDER RLHVTEMRLRWMCVTRMDKIRNEYIRGSL RVAPIVEKVRSKRLAWYGHVMRRDESHVTRR TLVMSVEGYIGRGRPRKRWMDCVKGDMIKK GVMTEMTYDRVMWRTKTCCADPTGVG	SSF56 672	25	135
RTE	RTE - 19_P Po		Papilio polytes	GGAAA AGAAG GGAGA TATTA AAAAA TAAAA GGAGT TTGAA GCAGG	TGAAAAT GTACATA CATACTT TATTA AAAAAAA CTTTTTAT ATTTAAC TTTAACT ATTTATA	MMRPRSSSLSTKPNQIKQLNTSNQKRNNPNP AGHRGVARPLPPKSMSSGSKLKSPLYLATLNV RTLMEHRLIELEHAISQIKWDIIGLCEVRRSG DEIREYNKYIFYKGEKKGRRGIGFLINKYLKD KIISFEGISDRIAVLNILNGFKHPMSIIQAYAPTE QDTNAEKDKFYEKLTEVIQKSNKNIIVMGDFN AKTGVKSNKDEFVLGKNSCGERNENGQRLIDL CFASNLKIAN SFYKKEKSSGRGCLQTEGTEMK LISFQFLQKRKVKWTWMSPNGRNRNEIDFILT	cd0165 0	581	833

			TATAT	TTGTTATT	RETRILSDVSVINKLNFSTDHRMLRGTLICEDK			
			GTTAC	AATAAGT	NSRKHII SNKNGKTITVPHNILEDIEKTLQDIKK			
			TGAAG	TGTAAAT	QQTIQDKYDFLEKTLKEVNIFGTKRGKRNLN			
			ATTTTC	ATTGTAA	EETKKILEERNELLDNRRENRRHSIANISKKIQSS			
			CGAAA	ATTGTCA	IRKHRKEERLKTLEQIEKTGGMTKGLKQLKE			
			GAAGT	GCGAATA	YSQWIPNIKLOKKKETYRITSRRSSI AKTATEFY			
			ATTAG	AAGGCTT	KQLYTESGESIVASKNII EEGDSNEEEIPTILETE			
			AAAAG	TATCTAT	VAGAIRSQKNDKAPGDDKIPNELLKSTLSSITK			
			CGAAA	CTATCTA	ALTILFNEIINSEQIPVQWTKSTITLLHKKGDKQ			
			AGAAT		DINNYRPISLMSNIYKVFSKVILGRVTQTLDAN			
			TACAA		QPREQAGFRADYSTMDHINVVKQIIEKCNEYG			
			ATAAA		QNYYLAFVDYSKAFDSL SHNYIWKTLASQGIE			
			GGTTA		RKYINIIKEIYKNSTAKIQLEQVGEEFKIERGVR			
			AAGAG		QRDPLSPKIFTAVLENIFRNIDWDKFGININGTK			
			GAGCG		LNHLRFADGIVLFAETPETLRTMLQQLDQESR			
			TGCTA		RAGLSMNPMTKLTMTNAEKDIIELRHEIIEYV			
			AAGGG		DEYIYLGQIISVNQHSKEIDRRISNAWKRYWSL			
			AATAT		KEVVKKKSIPMSIKTKLYNTCILPCLTYGCQT			
			AGCGT		WPSTTKHNLKLANCQRHMERSMLDIKLRNRW			
			ATATA		RNVTIREKNKSKEKKTKVNDVFKKIKKLKWN			
			AAATA		WTGHIIRTNKGKWTKDIMDWYPREGKRQRGR			
			TGATA		QIKRWEDDLPGWRRRLARDRETWKKLGEAY			
			AGATA		VEGQPD TQLDAGK			

				GTAGT					
				CAAAG					
				AAGGC					
				CATTT					
				AGTCA					
				AAGAA					
				AAAAG					
				GAAAA					
				GAGAT					
				CAATC					
				AACAT					
				CACCA					
				GAAAA					
				TCAAA					
				GTTAT					
				GCCAA					
				AAAAC					
				TGGTT					
				AGCAA					
				TAATT					
				CATTG					
				ATAAA					
				GGAAA					

				ATAGG AAAAA CGCTT TCGAC					
CR1	CR1 - 2_P H		Parhyale hawaiensis	CCAGT TACCC TCCGT GGCGG TCGGT TGTGA CGTGT ACGGT GCTCC GCTAG AGTAG ATAAA ATCAG AAAGC TATAG ATTTT CTCGA AAGCA ATAAA	TAACGCC GCCTGGA CCTCGTC CGAAGTA TCATCAA AGGTATC AAAGGTA	MGGNKKGRNKTQFKPVILNSTDDLDQELEPQE TRGVNECTLCLTSKQSKMIQCVLCEENNPHAS CLALNQSGRYHAPVCPSGHGCSVNERSTEKQH WVCTGCRGPTTNATPKVACTKCEFLFCAPCA KLRPGMIDMFNRFPKSFMWNCERCLQVSTQN RKCVASDTKESVGESPKTAGEEPLQKYAETGD HDVKEISYEIAQSTVQKRSSSGGDWISKNEKKP RICKYYAKNQCKYGISGRGCKFFHPKRCPRFM KYGRQGCNSKREC NFFHPWLCQESFRNGTCF DTTCPEIHLKGTIRAAPTAHHRHHAERGENS GRPWQNSNRIQTNKYGENNHQYNVERDITAN DRDMSVYNQDVRRDLQNMQNFLQTLYEKISS METYMEKVQHGGPRAPWNPPRVENPTQLWG DPGTTHQLKTQTAIHHAKQNEHQELKSNKDEI LSKVHVINIRGLLTKNCRKQMV LQDIINDAGS KLLAVTETHLDDAVNDAEVKIENYVLHRSRDR KDRKCGGVALYVRDHLGSHTKQKFSNGVCDC IIVNIEKLRLMCVVCYRPPKCTTEKFAEMLQVI NTHCKEVQSKYQLYITGDFNFPFLKWLTTDLP	cd0165 0	959	1215

			CTAAT	DETGSFVLSRTSSATRDEKDQALSLVRIMEEFS		
			TAAGT	LTQMISSPTRNENVLDLLFVDNAQLIKNIEIQD		
			GTTAA	TTLSDHRLVEITTQLGERDSHRPQKNSQINSISS		
			GCAGT	LRDLNVHDKGINWDGLNAELQSVNWDVLLN		
			GTGCC	NSIESESINETFENILLEKCMKYLSPAKRKFNKS		
			AGCCT	RHHKFRKGLLKKRAKLIKQSQSTKLLQRKSEC		
			GGCTA	KKKLDEIERQLARSHDEEKQEKEKKAIEKIKSN		
			TAAAC	SKYFFAYASHHSSVKSKIGPLVNDTGSIVXDD		
			TCGTA	QEMSEILSRQYQKVFSDKIPNPEIPRIDLTTNYL		
			CGGAC	TDFEFTKLSIELAIDELSGNSAAGPDDVPAKILK		
			CGTGT	MCKLSISYPLYLMWRKSMDDGVVPLTYKHAI		
			GACCA	SPIPKAGNSMDPANYRPVSLTSHVVKTFERVF		
			GCGGG	RKHLVHHLESNNLLTSNQHGFHHRSTTSQLL		
			ACCAG	EHYDSILDLLCQGKSVDVIIYLDFAKAFDKVDH		
			TGACC	GILLKLFNMGIRGKCLRWIEAFLTNRQTQSVIIN		
			ACACC	NKVSQRQPVLSGVPQGTVLGPILFLIMINDMG		
			GTAGC	HQCKFAVPSSFADDTRILAAVQSTADVRNVQL		
			TCGCT	DLTNIFHWTSVNNMTFNESKFELLYGKNFSL		
			TAATA	TESTSYTTTRDGSEIERKQILRDLGVRIQDNATFE		
			GGGGA	DHVARVVTVGKKWSGWVLRTFSSREADTML		
			AGCGT	TLWKQLILPRIEYCSPLWMPSKRKDLERLEAIQ		
			AGACG	RSFTAKIRGMKDLNYHERLAQLQMYSIQRRFE		
				RFTIIYVWKILEGITVNVNGRIQARHSNRIGRTC		

				TCAGT T		EVRIPKTTGGDAIATKYSNSFVVRGARLFNKMP RHIRNVTGTETAKFKHLLDACLLKVPDKPAVS GYVNCGDNSLLATPLVTDCSITSAQDGWRA			
CR1	CR1 - 4_P H		Parhyale hawaiensis	GCGTG GCCTG CGCGT TATCA GGCCG CCCCC ATTGT CCGGC CACCG GACGC CGCTT TGTTT GTATG AGCTG GCTTT GGGTT TGGGG CTCTG GGTAG CCCCT	TGATGGC TCTGTCTT TGCTGGT GCGGGTA TTGGCAG CTTTAAA CGATCTG TACACAG AGCTATT CTGTGAT CTCGGAT CTTCTTTT TTTTCATT CTTTTGTT TTCTTCA CCCGCTG GGCTGGT TGCGTGG GGGTTTG TTAACCT	MVYVYLLFFACLITRYPDVEENPGPRGFGSRS CRILCTNVRGLYGNLAELAVASSNYDLVLCSE TLVSSHRHVSELMIPNFEKPVLRLRDSCPGFRG LAAYVRKGVVVSQRQSEYECPCCEMMVLRVCG SFRNFYLFVYRSPSADDRVFDCLLDAMARIQ SSDRRSAFAFMGDFNCHHVEWLGSGRDCHG HAAFDFAASGCEQLVEGPTHLMGGVLDLLM TDVSDLVGVSVVSPIGSSDHGALSATLQVRQR VPRVTVRREIFLKS RVNWD AVSGAILALPWGV VRESTDPGSTLDRLLSDMVRAYVPSRFITVRTG DEPWFDDACRRAHQLKREAYARWRTSRSAV DFQLFREQRRRAGATYEEAEREHNLRSELLA SSVSAHKWWSTLKS AVLG VSSSV PPLIGTGGD LVADPAAKADLLRYFDGKQSRDRVRIPLACD PEPRFSGFAFRSSEVHRLLLDLDPYGGTDPSGM FPFFLRRVAGVLAAPLASVFRMLRLGSFPEC WRVAHVTPVPKGSVSSRTSDYRPISITPVLSKV YERL VASRLGCFLEKQGVLPPTQYGFRKGLGT CDALLHISHVLQLALDGGHEARLVQLDFSAAF DRVNHLGLLHKLRLMGIGGSVLAVISQFLSDR	cd0165 0	514	772

			GAGTT	CATGGAG	RQCVVVDGKRSEFLDLVSGVPQGSVLGPLLFI			
			GGAGG	TTTCTCC	VYTADLFHLVNNLLVGYADDVTLLAVCDKPS			
			AAATG	GTTCCCA	DRLEVSESIQHDLEMIDSWCSRWGMRLNAGK			
			CCACT	CGTCGCC	TKTMLVSRRTVGPPFPALMLDGVSLPVVDDL			
			CTAAA	AGTCTGG	VILGVNFDPKLTFEPIRSVASRASQKIGLMRR			
			GGCGG	TGTGATT	AWRVFGSVELTRRCFNCFVLPLLEYCSVVWGS			
			ATGAG	TAATAAT	AADCHLCLLDKVVRRCAFMLGGVVPCDLAHR			
			ACCTT	AATAA	RRVGGVCFLYKIYNCPSHPLSEFLPDPFVRVRY			
			TTTGG		TRRTGALHEFALEPLRCRTGQYSRSFFSIVCGC			
			TCCAA		LEWS			
			CGGCA					
			TCCCT					
			GAGAG					
			GAGAT					
			GAACA					
			GGCTG					
			GTGCC					
			ATAGC					
			TGGGT					
			ACTCA					
			GTCTT					
			CCGCA					
			GTCTG					

				CTCCT				
				CTCTTC				
				CTTCC				
				CTCGG				
				GTAGT				
				ACCAC				
				CTGAT				
				GCTCT				
				TTTCCT				
				TTCCTT				
				TCTCTT				
				CTTCCT				
				CTTTTC				
				GGCCA				
				AAGCC				
				CGTAC				
				TGGTG				
				ATGGC				
				AGGGT				
				CTGGG				
				GCGCG				
				TGGAT				
				AGCTG				

				GCCGC TTTCGT TGCTA TCACG ACTGT CGATT TGCTTT TTTGT AATTT TTTTG ATGAT TTCCTT TGCTT					
CR1	CR1 - 3_P H		Parhyale hawaiensis	CAGAA AATAA TCAAC ATTTG ACTGT GGTGT ATATA CAAAT ACAAG TGTAG	TAATGGC AATAGCT CATAGAG GGCCGGG GCGTAAG TTCCGGT GTCTCTTT GGGAATG GGCCTAT TGGTTCG	MKQCDECHSVFPLEFFYYEEVELESCKYCCKI QALQAEIEKLESDREAMQRGDGGEKEKAPRE AHLRLSSPSPWKTVMRNKRQRRSTGAILEQTN HIKVNNRFAAFNDGDDNVATSVRPKIKLKTVN GEQHRGPPQSRDQKKSETSVQPPSHERERNPA PIPGQRAIKTFLIGDQIIKGQYYAFKRGCNDRK VICRPQASTSDATAILKDLRCKKMDHVIIHVG NDLRTYSETQLQLVKDLIAAAKKKSDHVTLTS ILPRLRAGNRLYSRIRYINGQIKREAEXNGLEFI DLFPSFDDDWRLFLPDGVNLNGAGHHVLGEM	cd0165 0	1136	1383

			TGCTA	TCTACTG	LLAAAASETAALEEGLRTRPKTHIEQPRHPAPF			
			AAATG	CCGCCTA	HALPADLPAMEQTATNVPAAVPCNPQISSPSG			
			TGTTC	TCTAATC	MKVGD PKGNEGTHITNSLDTDEEEITPTPDASV			
			GATTA	TAAT	QMSRLGPPPLLPSIVSNLVSTNSNMSEGTDV PQ			
			WACTC		PTNARQTDLDSSNTERGSPVSVSKEKSQSQTPV			
			TTATTC		AKKRGQFPRTSSIPAHEG NRSTRKSGARNIPPK			
			ATTGC		TNSCQPPQTRITRSKANSAGNDLLSARLGGKR			
			TCTCG		PIVRQIAADV KQSDNCQAEIQNISLAXLQT GAL			
			AGTGC		AATDNAFADLT PPIFSQQSDAPSVSVTRRSNAN			
			TACCG		TSVKFLFGNVQSITNKMNELQAIAATLAYDVI			
			CTAGA		ALNETWINLSNKNLEAEVALKGYKLFNIDRPT			
			TATTTT		ETGRGGGSIIYIKEHLNPLCKGFAATKEYEFITT			
			CTCCA		HARLGIRQSLNIMLAYRSPRSTAETDAALYNKI			
			TTGCTT		ENALNTPNETILMGDFNLPQIDWSTLCSPAPGS			
			TTGGT		KLLELLSHHHLTQHVN EPTRYGNILDLVFSTPD			
			TCTTTT		CEIVDLAILDRLGDHNMIEFTLKTGNEFTTNAS			
			GGTTG		RLKL PDYRRANFQGMATEIETTDFETPMRDAS			
			GATTC		VEQCSHILKTSILRAAERYIPNKRVRISNDPLW			
			TGTGG		MNRTVKRCIAGDKCVTKPGSVIQRRRTGPVTS			
			TCGAA		PLDESDSQT LHCR RQVRYKAWKRNPTEENRSS			
			CTTAG		HIRACRDVKKIVRTTKREKEVSVAAMAKRNP			
			AAGAC		KAFYKYVNERR LKKETIGPLKRPDGXITNDLR			
			GCTGA		DKATILNDYFTSVFTVENVTNIPEPAPMQVAPI			

			TCCTG	NTCVFTEIEILQELTHLNVYKSTGPDNLTPRVL			
			AGTTA	REMAPVIVKPLLEIFNRSRLRTGEVPQDWKLAN			
			CTAAG	VPIHKKGDKKSPSNYRPISLTSVVSKTIERLLK			
			TAATA	KRITDHLRMHQLISDTQHGFNRNGRSCLTNLLD			
			TTATA	FFSKVTDIKDEEKGVDLVYLDFQKAFDKVPHQ			
			TTCTAT	RLMKKLEAHGIQGDYLRWIRAWLTGRQQRVK			
			CAAAT	LEGETSDWTQVSSGVPQGSVLGPLLFSIYINDL			
			TTTCCT	DLNIRSKISKFADDTKLCHPATKRAELQQDLD			
			CTATC	HLMEWSTTWQMSFNVDKCSVMHVGLVNPST			
			AATAT	PLKMNGTELA AVTQQRDLGITITNKLQHSQQV			
			TTTTTC	DKSVKMASRVLGFITRNFEYKSKDIIVPLYKAL			
			CATCT	VRPHLEFAVQFWSPHLCKDIDNMERVQRRAT			
			TCCTC	KLIPQLRNKPYECRLKELGLISLEQRRLRGQLIE			
			CTCAT	TFKYLNGFNDVPTTGLFERDLNDRLRNNGQKL			
			TGACT	KHKARRTTLSQQFFPYFIVSIWNRLPSDVVSAE			
			TGCCT	SVNAFKNRLDIHWERHPPQLRTPPPGVASSELL			
			GATCC	LQESQALNGQSFRGHLNPSSSLRSNYTSLSSA			
			TCGCC	YQRL LIVLACIHDRLFGWDSSPPSF SFFFLFLPF			
			AAAGG	FSSLSPVLYILLT			
			GACAA				
			GTATG				
			ACTCC				
			TGAAC				

				TTCCC				
				GAGTG				
				GCTAT				
				AAGTT				
				TGCCT				
				TACTA				
				TAGGT				
				GGGAT				
				TAAAA				
				ATTGT				
				GTGCT				
				TCGGT				
				CTCAA				
				CTTTTT				
				TTTAG				
				TTTTCA				
				AGTAG				
				ATAAC				
				ACCTT				
				TTCAA				
				GAACG				
				CTAAT				
				ATTTTT				

				TTTTTC					
				AAACC					
				TTTCTA					
				GAGAC					
				TAGGA					
				GGTAC					
				ACTAG					
				CGAAT					
				ATTTA					
				TTCTC					
				AAACT					
				CCTTA					
				CGTTTT					
				GGATC					
				TTCAT					
				AAAAT					
				ATTAT					
				TTTTTA					
				TTTCTT					
				TTTTTT					
				CTCTC					
				ATTCA					
				TTTTCT					

				TTTCTA					
				ACAGA					
				CTGTA					
				GGTTT					
				AGTTC					
				AGAAA					
				TGTTCT					
				ACATG					
				AAATG					
				AAAAA					
				AAAGA					
				GCTTG					
				AAACA					
				TATGA					
				TAATT					
				GAGTT					
				CATTC					
				CTAAC					
				ATTGC					
				CAAGA					
				TAAAC					
				AGTTC					
				GGAGC					

				TATAG					
				GCCCG					
				ATGTT					
				CGATC					
				TATTG					
				ATAAT					
				TACCA					
				ATCTT					
				AGCGG					
				CGTGA					
				GGCAA					
				AACAA					
				AGTAA					
				CACTC					
				CAGAG					
				TTCCT					
				GGCAA					
				TTTTTG					
				TTTAA					
				TTCTTG					
				TTTCTC					
				TACTC					
				CTCCC					

				AAAAC					
				TCTCA					
				AACGA					
				CTTCC					
				AATAG					
				TCTTC					
				ATTTT					
				ATTTC					
				AAGAC					
				TTTTA					
				ATACA					
				TCTTAT					
				TGTAG					
				TTTTCT					
				TTTATT					
				TTGAC					
				ATGTT					
				AAAAT					
				GCCTA					
				AWTCG					
				TTGCT					
				GCAAT					
				GTACG					

				TAAAA					
				ATTAG					
				AACTT					
				GATCT					
				TTTTTT					
				GTTCA					
				TCTTTC					
				TAGGA					
				ACATT					
				TGTGT					
				ACGGT					
				GCAAC					
				GGGTT					
				ATATG					
				TTATA					
				TATTA					
				TATAC					
				AGTTC					
				CTTTGT					
				TTACC					
				TCGTC					
				AAGGA					
				GGTTA					

				TGTTTT					
				TCTCG					
				TGTTT					
				GTCTG					
				TCCCTT					
				GTCTG					
				CCTCTT					
				GTAAG					
				TCACC					
				AAGAT					
				TACTC					
				TAAAA					
				CTAAT					
				GAACG					
				GACTC					
				TTATG					
				AAATT					
				TACAG					
				GGATT					
				GTTAG					
				ACATG					
				TTGAC					
				AGTAA					

				CAGAT					
				GATTA					
				ACCTT					
				TGGTA					
				AAAAT					
				CCGTT					
				CACCT					
				GTCTG					
				TCTGC					
				CTCTTT					
				TTTCGT					
				GCGCC					
				AACTC					
				AAAAG					
				GTTCC					
				GGACG					
				GATTT					
				TTATG					
				AAGTT					
				TATAT					
				GATTT					
				GTTGG					
				GTATA					

				AAACG					
				TAGAG					
				TAATT					
				GACCG					
				AAATT					
				TTGGT					
				GATAT					
				GTTTG					
				TTCGA					
				CTCTG					
				TGATG					
				TGTTT					
				GTTCG					
				TCTCT					
				GTGAT					
				GATAT					
				GTTTG					
				TTCGT					
				CTCTTT					
				GTTTCT					
				CGCTG					
				TGTTT					
				GATCT					

				TGATA					
				TTATG					
				CTTGTT					
				AGCAC					
				AATAA					
				CTCGA					
				AAAGC					
				TGTGG					
				ATGGA					
				TCCTC					
				ATGAA					
				ATTTA					
				CAGGA					
				CATGT					
				AGTAA					
				TGTTA					
				AAAGA					
				AATAA					
				AGGAT					
				TAAAC					
				TTTGG					
				TGAGG					
				ATCGG					

				TATGT					
				TCGTTT					
				GGGTG					
				TTGAA					
				CGGAT					
				CTTCA					
				TGAAA					
				ATTAC					
				AGGAA					
				TTTTTT					
				TAATT					
				CTACA					
				GAAAG					
				GCACA					
				GCTGA					
				TTAGA					
				TTTTG					
				GTGTT					
				GATAC					
				AAAGA					
				AATTC					
				ATGGT					
				TTCAG					

				AGCCA					
				ATTTG					
				AATTT					
				TTCGT					
				CACCG					
				TTGCC					
				TTGCTT					
				GGCGG					
				AGGTA					
				TGCGC					
				TCTAG					
				TTTGT					
				AATGA					
				ACTTG					
				CATGT					
				AGGGC					
				ATCTT					
				ATTTT					
				ACTCC					
				TTATC					
				AAATT					
				ATGTC					
				CTTTCT					

				TGCAT					
				GTAGG					
				TTTCTT					
				TTCCTT					
				TAATA					
				CATTG					
				TAACT					
				TACTA					
				TAGCT					
				ATATC					
				TAGGT					
				AGCCT					
				TATTT					
				GCATT					
				TTTTAC					
				TTGTTT					
				CATAG					
				GAATT					
				GCTTG					
				GGGGC					
				AGGAT					
				AACAC					
				GTTAA					

				GATAT					
				TTTCTT					
				TGTTTT					
				TCATC					
				CTTGT					
				GTTTG					
				TAATT					
				TAACT					
				GCCAA					
				GTTCTT					
				ATGTA					
				CACTG					
				TTTAA					
				TCAGT					
				TAAGC					
				ATATA					
				CAGTA					
				TTTGC					
				GATGT					
				ACAGT					
				ATACG					
				GAGGT					
				GCTAC					

				TTTGC					
				ATGAG					
				GAAAC					
				TTATT					
				AACAA					
				AGTTC					
				ATTAA					
				TCTAG					
				TTCAA					
				ACTGT					
				TCAGG					
				CGATT					
				TCGCT					
				CGAAT					
				CTTGG					
				GCCAT					
				AGCTT					
				TAGTA					
				ACGGC					
				CCTAT					
				CGATG					
				GGGCA					
				CGCGT					

				GACGT					
				TTTTGC					
				ATTGA					
				TGCCC					
				CTGTC					
				CCGAA					
				TTCCC					
				GGTAC					
				CGAGT					
				AACTG					
				TTCCA					
				TTTCTG					
				TTTTTC					
				TCCCTT					
				TCCTCT					
				TCTCCT					
				TTCAT					
				GTCCT					
				CTTGT					
				GCCGG					
				CAGCA					
				CAGGC					

				CTACA AT					
Ving i	Ving i- 1_P H		Parhyale hawaiensis	CCACC CGATG CCTGT GCGTC TTCGT CGTGG GTGGA CCCTG CGGGA GCTCT CTGCG CGGCT GCGAC GGGGA CTTCC GCGGG CCAAA CGCAG GCACT CATCG TTGGA	TAACTAC AACAACA ACA	MPCSGLFSAASYHPDWRCPCAPQPPRPPSPPP ASPLSPHPTNAPPLIVMQFNCNGLLHCRSELSA FLHSSNIAIAALQETKLTSSSLPVFPGYTPVRK DRPAANGGGGLLFLIRQDIPYQQATTNHLFPN DHTVEHSGVTVTLCSSSLTLINIYVPPATSCPPG YRLDSSALFADSSADTLIMGDFNAHNTDWYS VTQDARAAARGEESISINNGPYCLLNGDSPT RVPRDGP GSSPDLTVIS AHLALDYSWSTTTLS SDHLPILITAESSDSAPSGPRRTFVNFRKANWE EFTRECEELITSLPLPSSCSGGLRALNRALLSAA KHHIPAGHRKEYIPRLTPDARDAISRRLHRLST SPHDPSLPQLQQEVDRLICGAAQEA WKQTIQD CDPKSNPSKFWSLRLSLSGKQSSVPRNQPI SFG THTFTHNGDIARGFCKQFTATSPPALATPPRLA RRLLRKRHP IAHGFSPFPAQVREAIVASGSST AVGPDQLSILHLRHLGPQGIIYLTRLYNLSLAH ADIPALWKLAVIIAILKPSKPADKGSSYRPISLL CPAAKVLERLLLPFINEAYQPAPHQHGF RPMH STTSALLPLAFQATRGMNKP KPHDRTVALLID FSKAFDTPVTEGLLSSVSGTTLHHNVVRWLAA YLRGRAACCLYNGSRSQFRAVKAGVPQGSVV	cd0165 0	536	795

				GTGGC	SPSLFNLYVSDFPIPRDQLCYPSVTSYADDFTV			
				TACGC	VASHRDPLTAAAALSDHSASVSNWAANKGLT			
				GGGTC	ISTSKSNSTLWTTDTHQSTLDPGITMNDALPL			
				GCTGT	ERRPKILGVTFDTHFTFTPHITSTAARASSRLRL			
				GGCGC	LKALAGTRWGQCKETLLLTYRAFVRPLITYAA			
				TCCCTT	PVWYPGARQSAVRKLQTVQNAALRTATGCLR			
				TCTTCC	MASVDHLHAESKXLRIDQSLDLQCRQYLVSA			
				TCTCTT	RPEHPAHSVVTAQAGPRLIRHNLATKFQPDIN			
				ACTCT	HLLTDNIMPSPEYRSAISSLHRSVQASLDRLP			
				CTTCTC	VNRVLGFAPPEISPDEAVLSRAARSTLSQLRSG			
				CCGCT	FCSALMSYRHRIGLADSDSCPECHRAPHTVNH			
				ATTCTT	LFECPAHPTALLPTSLWDSPLSVTDFLCSLPFS			
				TCTCG	YLPPHARPPPEPPPT			
				CTGTC				
				TCGTG				
				TTTTTT				
				TCCTTC				
				ACGGA				
				CAGGT				
				CCTTTC				
				TGCCG				
				ACGAC				
				CTGCC				

				CGCTC					
				TTGTCT					
				AGGCA					
				GCCTT					
				TTCCTT					
				ACTTG					
				TTTCCC					
				CCTCC					
				CGCTG					
				TCCCTT					
				TTCTCC					
				CCCTA					
				TCCAT					
				TCTAC					
				TAAGC					
				GGTGA					
				TCATC					
				CACCC					
				CAACC					
				CCGGC					
				CCCGC					
				CTGTC					
				CCTGC					

				TCAGT CTGCT CGCGG AACGT CACCC GCCGG GGGTG GTCTTT TCTGT GTTCC GCCTG CGCGC GCTGG GTCCA C					
CR1	CR1 - 2_P Si		Pelodiscus sinensis	GGGTG GTGGT GGTCG GGGAC TCCCT CCTAA GAGGG ACTGA	TAGATGG GATTGGT CCTGCCT AGAGCGG AGGGCTG GACTTGA CGACCTT CTGAGGT	MHRARSEKGGPQFRMEKIVQSTGYLKCLYTN ARSLGNKQEELEALAQAKEYDLGITETWWD DSHDWSAVMEGYRLFRKNRQGRKGGGVALY VREHYDCSELQYKEGEKLVESLWVRFKGANS SDVVVGVCYRPPNQVDEVDEAFFEQLSEASRS QALVLVGDFNHPDICWETNTAVHRQSRKFLE NVGDNFLVQVLKDPTRGRAQLDLLLTNREELI GDVEVGDNLGSSDHEMVDFRILTKGRKESSKI	cd0165 0	502	753

			GCCAT	CTCTTCC	HTLDFKKADFDLSLRDLMERIPWDVNMKGKGV			
			CCCTC	AGTTCTA	QDSWQYFKEALLKAQKETIPTRSKRGKPGRRP			
			TGCCG	TGATTCT	DWLTGEILGVLKHKKEAYRKWKLGMQMTREEL			
			TCCGG	ATGATAA	KCIVRECRGVIRKAKAQMELRLAKDVKDNNK			
			ACCTG	GGGGATG	GFYRHVNNKKKVIREGVQSLMDEGGNLVTEDEV			
			GAATC	GTTGGAT	GKAEVLNAFFASVFTDKVGSWTSVLS DARWE			
			TCGAG	GAAATAA	EDGQPLVGKEQVRNYLEKLVHKSMPDLM			
			AGGTG	CATGATC	HPRVLKELANVIEEPLAIIFEKSWRSGEIPDDW			
			TGCTG	TTGGTAA	KKANVVPFVKKGKKDDPGNYRPVSLTSVPGKI			
			CTTAC	CTAATTG	MEGIIKESILRHLNERKVIRNSQHGFTKGKSCL			
			CGGGA	ACCATTC	TNLISFYDEV TGSVDMGKSVDVIYLD FSKAFD			
			GCTCG	ATTTCCA	TVSHNILASKLRECGLDKWTVRWIERWLEGR			
			CATTC	GTTGGAA	AQRVVINGSMMSGWRSVSSGVPQGSVLGPVLFN			
			AGGAT	ATAGGTC	IFINDLDEGIDCTL SKFADDTKL GGEVDTL KGR			
			GTCAC	AATGGAG	DRVQSDLDKLEDWATRNL MRFNKDKCRVLH			
			TGAGA	GGATGAT	LGRKNPKHSYRLGTNQLSSSSAEKDLGVTVDE			
			GGCTT	AGGAGTT	KLDMSQQCALVAKKANGILGCIKRSIASRSRD			
			ACCAA	GCTATAG	VIIPLYSALVRPHLEYCVQFWAPHYKKDADAL			
			GCTGA	GGAACCT	ERVQRRATKMIRGLEHMTYEERLRDLGLFSLQ			
			TCAAA	TCTGGGT	KRRVRGDLIAAFNFLKGGSKEDGERLFSVTD			
			CCTTC	GTCTGGC	GRTRSNGCLKLWWERSRLDIRKNYFTRRVVKH			
			GGATC	TGGTGAG	WNGLPREVVESPSLEVFKSRLDKALAGLI			
			GCTAC	TCTTGCC				

				CCCTT	CACATGC				
				CCTGC	TCAGGGT				
				TTCTCC	TTAGCTG				
				ATGTG	ATCGCCA				
				GGAAC	TATTTGG				
				TAATG	GGTCAGG				
				ATACG	AAGGAAT				
				GCCAA	TTTCCTCC				
				GAATG	AGGGTAG				
				ACCTT	ATTGGCA				
				GAGCG	GAGGCCC				
				GGTTA	TGGAGGT				
				CTGCG	TTTTCGC				
				GATTA	CTTCCTCT				
				TGTAG	GTAGCAT				
				CGCTG	TGGGCAC				
				GGAAG	AGATCAC				
				AAGGA	AGCTGAA				
				TCCAA	GGACTCT				
				GAATT	CTGCATG				
				CGGAG	TTGGGGC				
				CGCAA	CTTCAAA				
				GTGGT	GTATTTG				

				GTTCT	AAGGCTT				
				CGTCC	CAATATC				
				ATCCT	TGAGACA				
				CCCTG	TAGGTGA				
				TTGAA	GAGGATT				
				GGGAA	ATTCTAA				
				AGGAC	GAGGGGT				
				TGGGC	GGGCGAG				
				GGGGA	ATTCTGT				
				TCGTC	GGCCTGC				
				GAATT	ACTGTGC				
				GAGGA	AGGGGGT				
				CGTAA	CAGACTA				
				ATGCG	GATGATC				
				TGGTT	ATAATGG				
				GCGCA	TCCCTTCT				
				GGTGG	GACCTTA				
				TGTCG	AAGTCTA				
				CAGAG	TGAGTCT				
				AGGGC	ATGAGTC				
				TTTGG	TATGA				
				TTTCTT					
				CGACC					

				ATGGG					
				ACTCT					
				GTTCC					
				GGGCG					
				CAAGG					
				ATTGT					
				TAGGA					
				AGAGA					
				TGGGA					
				TCCAC					
				CTAAC					
				GAAGA					
				GAGGA					
				AGGAG					
				CATCT					
				TCGCG					
				GGCAG					
				GCTTG					
				CAAAC					
				CTAGT					
				GAGGA					
				GGGCT					
				TTAAA					

				CTAGG TTCGT CGGGG GACGG TGACC AAAAC CCTGA GGGGA GTGGG AAAGT CAGAT ACCGG GAAGA A					
RTE	RTE - 14_P Si		Pelodiscus sinensis	CTGGT GCCAA GCTGT ACTGG CTTTC GCCTT TCCCTT GGGTT ACATC	TGATCCT TCGG	MIPGLTEDLRISISNPRKTAVINNELRRLRMDIV ALQETRLSSSGCLREKDFSFFWHGKPPPEETREH GVGFAIRNNLVGSVIPNVGTERMLKIQQTSTV GMANIFSVYAPTLTSAQEDKDKFYDELSIAISE VSSHEPLFILGDFNARVGSDHHSWPSCLGQFG TGKMNENGQRLLLELCSQHDL CVTNTFFNVKP QHKVSWRHPRSKHWHQLDLVLT KRSRLG SVR VTRSYHSADCDTDHLLVCSKVYIQPRRFFRSR KEGRPRIDTGKTRDPSKVCEFVQALEEALPRR	cd0165 0	525	802

			AGTGG	DDTDISKRWEQLRDAIYDTAISIFGKKT VKSAD		
			CGTGG	WFEAHT E V L T P L I E K K R Q A L A A Y R S S P S E E N L		
			AGAGG	QALRFARSQVQQVARRCANDYWLQLCSKIQQ		
			GGGAT	AADTGNMRGMYDGIKQVIGPSQRKTAPLKSA		
			CCTGC	TGEVLKDRTKQMERWVEYYSQLYARETVVTE		
			TGCTG	RALNAIEPLPTMTELDAEPTMEELSDALKALSS		
			GGCAT	GKAPGKDGIPSEILKCANGTLVSELYGILCQCW		
			CAGCC	REGSVPQDMRDANIVTVYKNKGDKSDCNNYR		
			GGTTC	GISLLSTVGKLFARVVLKRLHVLAE RV P E S Q		
			TCAAC	CGFRAERSTIDMVFSVRQLQEKCREQNKPLYI		
			TTACTT	AFIDLT K A F D L V S R K G L F A I L A K I G C P P T L L S I V		
			GCCCA	RAFHDGMRGTIVYDGSTSEPFEILSGVKQGCV		
			GGCCT	LAPTLFGIFFAVLLKYAFGTATEGIFLRTRTDG		
			ACGCC	NLFKLSRLRAKTRVLTKRLREFLFADDA A I A A		
			CTGGA	HTLQELQSLITRFADACMEFGLTISLKKTQVM		
			GAGGA	GQSVGNEPFISVLDYELEV V H E F V Y L G S T I S D N		
			CACTC	LSFDSEINKRIGKAATTFSRLTKRVWANNKLT		
			CGGCA	VHTKVQVYTACVLSTLLYASETWTLC SK Q E Q		
			TCGCT	RLNTFHMRLRRILGISWQDKIPNSAVLERAGT		
			TTCTG	VSMFTLLKQRRMRWLGHVSRMLDGRIPKDLL		
			AGCAG	FGELASGKRPKGRPKLRYKDTCKRDLSALSIS		
			TGACA	ANTWHS L A L D K Y A W K Q G V K E A L A M Y E T T L V		
			CAACA			

				CGGGA	KEAEDKRALRKIRARDTRATSAYCCSQCGKD			
				AGCAG	CHSRIGLHSHRRRCSNQSNRGANP			
				CGGTT				
				ACCGG				
				TATTA				
				AGCCA				
				CTGCA				
				CTTGA				
				TTGGC				
				GTAGA				
				GTGAG				
				GCGCC				
				AGGGG				
				TTGCTT				
				CCGAC				
				GGTGG				
				GAGAG				
				GTCTT				
				CGGAT				
				CTCAT				
				TGGGC				
				AGCTA				
				CCGCC				

				CGCCT					
				CAAGC					
				TGGGC					
				AGCCC					
				CCAGC					
				CAAAT					
				AGGTG					
				CTGCC					
				TCGCC					
				ACGGT					
				CCACT					
				TGCTC					
				CATTG					
				GGTGT					
				GTGGG					
				GCTTA					
				GGGAA					
				TCACC					
				AACAG					
				GTGGA					
				TCGTT					
				AACCT					
				TACAC					

				CAGGC AAAAC AAACA AGCGA CAAAG TAATC CAGCC TTCAG GCTGG GCACA TGGAA TGTA GGACC					
L2	L2-1_P_Si	.	Pelodiscus sinensis	GGTTT AGACA GTTAT GKGCT ATTCC TKTTC CGGGT CCTGT CCCCA GCCTC	TGAGTCT GTGTGTG TGTGTTC CTCCTCT CGAGATG TTTTTAG CAGGTTT TATGTCC CTCTGGG TCTCTTTT	MLLVYRPPRCTADALLELADVSDVALRTPRL LVLGDFNIHAEAVLSGPARDFMDTMTTLGLSQ EITAPTHVAGHTLDLVFWTKGDRVLGIRDVAT LPLSWTDHYLVKFRLSADLPLHRGGGPIKMIR PQRLMDPEGFLRSLRDCSAIMANGPVDALVDT WNTEVARAVDMIAPKRPLRASRAKFAPWFSG ELMAAKRSVRRRLERRWRRTRAESDRLRVRAE LRAYSVAVLTAKKSYFSTRIASAQCRPAELFR VVRGLLQSGTRDEMEEPLAPRCDLFAKHFAD KIARIRLDLDSGCCVSGAQNVL SAPP SLFSLDQ	cd01650	379	640

			AAWAA	GGGGTCT	FQLLQTEDVDKVLGGVRPTTCSLDPCPSWLIK			
			CTCTG	CGTATTT	SSREGMTGWVQGIINASFRSGMVPAALKEAVI			
			GGAAC	TAAATTT	KPLLKKPSLDPRDLNNYRPVANIPFLGKVLER			
			CTTAA	TTATATA	AVALQLQAFLEETDYLDPFQSGFRPGFGTETA			
			CGCCT	TATTTGT	LVALVDDLRRREIDRGASLLILLDLSAAFDTID			
			GKCCA	ATTGTGT	HGILLERLAELGIGGTVLQWFRSYLSGRFQKV			
			TCAGC	TTTTAAA	ELGGCYSTPWRLCYGVPQGSILSPMLFNIYMK			
			CCTGA	AATTTTG	PLGEVIRSFGGGRCHQYADDTQLYFPVTSEPGV			
			GCGTG	TAAGCCG	AATDLNRCLEAVMDWMKSNKLKLNPKTEV			
			CTGTT	CCTCGAA	LWIRRPVRECGVVPVLDGVTPLPLKEQVRSLG			
			GCTGA	TATTTTA	VLLDPFLTFFENQISSVARSAFGQLRLIRQLRPFL			
			ATGCC	AATAGAG	NKHDLATVIHALVTSRLDYCNALYVGLPLKSL			
			AGGTC	AGGCGGG	RKLQLVQNAAARVLTNTRYTEHIMPVLRQLH			
			AGTAT	GTAAAAA	WLPVCFRAQFKVLVMTYKALNGLGPGYLKDR			
			CCCAG	TATTTTA	LLPYEPARGLRSAGEALLRVPLMEIRLTSTRS			
			AAGAC	AATAAA	RAFSAFAPRLWNSLPRDIRMAPTLALFRRQAK			
			AGCTG		ATLFTRAFIGV			
			TCATC					
			CACGA					
			TTTAA					
			TTCTG					
			AATGA					
			AGGGG					

				CTGAC				
				CTGGC				
				TTGTA				
				TCACT				
				GAGAC				
				CTGGC				
				TGAGA				
				GTGGA				
				GGGAG				
				GCGTT				
				GACCT				
				TTCCC				
				AATTG				
				TGCCC				
				GGCTG				
				GGTTC				
				TCAGT				
				ACAGC				
				ATCAG				
				CTGTG				
				GCTTA				
				ATGGT				
				CGGGG				

				AGGTG					
				GGGTG					
				GCAAT					
				TGTCT					
				ATCGA					
				GATAC					
				CATCT					
				CACTG					
				GCTAG					
				GGTCT					
				TGGTC					
				CAACA					
				AAGTG					
				TCAGA					
				TTCGA					
				GTGTC					
				TGTAC					
				TTGAC					
				GTTGG					
				GTGGC					
				CGGGA					
				TAGGT					
				TGGGT					

Pene lope	Pose idon - 2D_ PM	Petromyzo n marinus	TAGCTGC	YLRDFTEKPKSMNPCSHESFKSRINALYGEXFH	cd0030 4	464	548
			CTTATAA	LSVRAFNRQRDKKSQLIETLTFLKRCRDNKVIP			
			TTGTACA	IAFTLTSAAVSSPNSKRILFNASLSLISERIRVLR			
			GCCATTT	FELDSLDKTLLQSHLMLTQKMSPSDWNHVDN			
			CCATTTC	MTAASADSLSAKCKANLIAKFNLLSKKKPEPH			
			TTCCAGT	ITKLDV NKIVINQSNIPISDNLKTVLAKGLNFAV			
			GGGTTGT	TPKHIPVEDIISQVDVALKGIGHIQADVARAEIA			
			ACTCGCT	TILKNAKKPISNISKNERSELANLKKRDDIVILS			
			CTCGTGC	ADKGNATVIMDKNSYNNKIQDLLNDGSYSKIK			
			AGTGCAC	SDPTNSLTIKTNKIITSSTSLIEPTKRPSLKPSAA			
			CACCGCT	RPPRFYGLPKIHKEGTPLRPVSTINSPTYKLAR			
			TCCCCCG	FLSDTLLPFTGKTTSAVLNSTQLVSMVKDIVLS			
			AACCCCC	PDDILVSFDVVS LFTKVPVPDTITIEDLVQHGL			
			TGCAGCA	NPDVPALVRHCLTNTYFTWNGTFYKQNEGTP			
			TAACAAA	MGSPSPVIANVFMEKFESDAIQTSTLRPSLWK			
			AGGCGGC	RYVDDTFVIWPHGLPALNHFLDHINSRHPSIKF			
			TTAACTT	TMEVEKD GKLPFLDVL FERRTDGSLGHSVYRK			
			AGCCCCC	STHTDSYLKPDSHHHPAQKNSLIKTLYHRAHCI			
			CCCCCCC	SDQHHLASELKHVHLALKMNGYSDRTIRQAV			
			ACCCACG	SPTIQKQPPDTPIKIITLPYIAGVTKKISKILNKH			
			TCCCTTC	KDPGPVQYSTQNPTNTKIQVRFNTVHKIRDLP			
			CTGGCTC	SVKDVIPDLQYPGVYKLSCTCGDCYIGETKRHI			
			CTCCCCC	SDRIREHKSDLKHGR TNSSAVADHCFNTAGSH			

				CCCCCT	DIVFSDTKVLCKPPGYYQRFVHEAIQIYKHRH			
				TAACCTG	NFNREDGFKLSNHWKDVINQVKL			
				TATATAA				
				TCTCGGG				
				TGTCTCA				
				GAGCTCT				
				TCACTTG				
				ATACAGC				
				TATTTTG				
				CTGTGAC				
				TGTTCCA				
				CCTGAAG				
				ACGGAAG				
				CTTGTGT				
				TGCCTCC				
				GAAACGT				
				CGTGATT				
				TAATTTA				
				GTTGTTTT				
				CATTTAT				
				TTTATTTT				
				ACCTACG				
				TGACTTT				

					ACCGAAA AAAACCA AAGAGTA TGAATCC TTGCAGT CACGAAA GCTTCAA ATCTAGA A				
Pene lope	Pose idon - 1F_ PM		Petromyzo n marinus		TAACCTG CTGTGAT GTTTCCA TCCTTCCT GCGATTT TAGCTGA CTCTCTG AGTTGCA TTATTAG GTTGGAT GTCTTTG TATTGAC GGGCTGG ATGTCCT	IHKEGTPLRPIVSTINSPTYKLTRFLSETLLPFTG KTTS AVLNSAHLVSLIKDISLSPGDLLVSFDVV SLFTKVPVPDTIAIIEDLVQQGLNXDVPTLALP YQHLLHLEWYILQTRGGYPQWVLPYRQWCA NFGTALPTLTSPGMVHSTNKRRVPPMGSPLSP VVANVFMEKFETDAIQTSTLRPSLWKRYVDDT FVIWPHGQPALNQFFNHINSIHPAIFTMEVEK DGKLPFLDVLIEKRPDGSLGHSVYRKATHTDA YLPKPDSSHHPAQKNSLIKTLHHRAHCISDQHH LASELEHVHQALRMNGYSDRTIRKAVSPAMG RQPSEDPIKTITLPYIAGVTKNISKILNKHKIQV RFNTVHKIRD LIPSVKDVIPDMQYPGVYKLSC GCGSSYIGETKRHVSDRVREHKADLKHGRTNT SAVADHCFNAVGTHEIDFSKTKVLCKPTGYHQ	cd0030 4	156	239

					TCTCGCT	RFVHEAIQIYKHRHNFNREDGYKLSNHWKDVI			
					CAGTCAC	NQVKS			
					CGATGCT				
					GCGACAT				
					ACTAGGC				
					AAGTCAC				
					CGATGCT				
					GCGGCAT				
					ACTAAGC				
					TCGTGCA				
					TTGACAT				
					GCTAATC				
					TCTCACA				
					ATACAAT				
					AGCCACT				
					TCCACAG				
					CCTCAGG				
					GAGACCA				
					TCTCTAG				
					CCCCGCC				
					CTCTTAC				
					CCTGAGG				
					CTGCACC				

					CATCTCA				
					GGCCTAC				
					ATAAGAC				
					CTTGTTTC				
					AGTAGCT				
					TTCTCTCT				
					CTCACCT				
					GAAGACG				
					GTCGCTT				
					GTGTTGC				
					GATCGAA				
					ACGTCGT				
					GATCTAT				
					TAATTTT				
					CTATCTA				
					CGTGACT				
					TTACCGA				
					AAGAACC				
					AAAGAGT				
					ATGGATT				
					CCTGCAG				
					TCACGAA				
					AGCTTCA				

					AATCAAG AT				
Pene lope	Pose idon - 1G_ PM	.	Petromyzo n marinus	AC	TAACTGG CTGTGAT GCTTCCA TCCTACC TGCCGAT GTTAGCA GTTATTT CCGACTC TCTGAGT TACATTA CCTTGTT AAATGTC TTTACAT TAACAGG CTAAATG TCTCACT AACATAG CTGCTGC TCCCTTTT CCTCGAT GCGGCGG	QPPGYENQQYITDSTSLIEPTKRLSLKPSAARPP RFYGLPKIHKDGTPLRPVSTIDSPTYKLARFLS ETLLPFTGKTTS AVLNSAHLVSLIKDISLSPGDL LVSFDVVYLFTKVPVPDTIAIIEDLVQQGLNAD VPALVRHCLTNTYFTWNGTFYKQVEGTPMGS PLSPVVANVFMKFETDAIQASILRPSLWKRY VDDTFVIWPHGLPALNQFFNHINSIHPAIKFTM EVEKDGLPLFLDVLIEKRPDGS LGHCLQESH HGCLSKTSKRDQMAHWDTVYRKATHTDAYL KPDSHHHPAQKNSLIKTLHHRARCISDHHHLA SELEHVHQALKMNGYSDCTIRKAVTPTVGREP PEDPIKTITLPYIAGVTKXISKIMSKHKIQVRFN TVHKIRD LIPSVKDVIPDMQYPGVYKLSCACG SSYIGETKRHVSDRVREHKADLKHGRTNTSAV ADHCFNAVGTHEIDFSKTKVLCKPTGYQQR FV HEAIQIYKHRHNFNREDGYKLSNHWKDVINQ VKS	cd0030 4	155	246

					TGGCGCT				
					CGTTGTT				
					TAATAGG				
					CTGAATG				
					TCTCTCA				
					TTAACAC				
					AGCCGCT				
					GCTGCGG				
					GAGCCGT				
					CTCACAT				
					TGACATG				
					CTAATCT				
					CTAATCT				
					CGAGCAA				
					TACAATG				
					GCTTCTT				
					CAGCACC				
					GTAGCCG				
					TGGGGTT				
					ACATTAA				
					CAGATTG				
					ACATGCT				
					AATCTCT				

					AGCAATA				
					CAATGGC				
					CGAGGTA				
					TACCTAG				
					CTCTGCG				
					ATCTTAC				
					ATTTACT				
					TGCAAAT				
					TCCATCT				
					CTAGCCC				
					CGCCCAC				
					TTACCCT				
					TGGGTTG				
					TACCCAT				
					CTCAAAC				
					CTATATA				
					ACCTCTG				
					GCTGCAG				
					TAGCTCT				
					CTCGCTA				
					TATTTCTT				
					CCTGCCT				
					CAGCAGT				

					CCGCCTG CTGTATT TTTTCAC CTGAAGA CGATTGC TTGTGTT GCGATCG AAACGTC GTGATT TATTTTAT TATTTTTC TATCTAC GTGACTT TACCGAA AGACCCA AAGAGTA TGAATTC CTGCAGT CACGAAA GCTTTAA GTCAA				
Rex 1	Rex 1-	.	Petromyzon marinus	G	TAATTAT TAAGTAA	MAAHTVAAAQCLPGRCLVWLFMYALCLVLC TFCRASNTFSRHDLLKIGLCYERSVTSEFLRLH	cd0165 0	580	832

1_P					TACTGTT	KIPVELARSPGSPWIIIPAGRRIIIIIIIIIERRRQKRG			
M					TACAATT	CRAGALARLRRQPLKPPLPSLFLTNAISLANK			
					GACTGTA	MDELRLQTTANNVVKDSCILLFTETWLHSSIPD			
					AAATAGG	SAIELAGYTAQRHDRTVDSGKSRGGGLCVYV			
					GATGTGC	NNSWCTNTVTVDSHCSPDLEYVTVKCRPIYLP			
					AATGACA	REFTVVMITAVYIPPDANANSAIGLLHGSISIQQ			
					AAATGTA	SKYPDAVQIIAGDFNHADLKAALPKFHQHVKC			
					CCTTTAT	ATRGTKTLDKVYTNIKLGYRARPLPHLGQSDH			
					GTGGGTA	LSLLLIPAYAPIKKTAPTITKTVATWPEGATQQ			
					TGTGGGT	LQDCFDRTNWGVFEHQDLEVFTDSVLCYIKNC			
					ATACTTG	IDTVTVDKRIRVYPNQKPWMTREVQQLKER			
					GGCAGTT	NIAFKSRDKALYSTARANLKRGIKAKADYRR			
					TTTATAT	RIEERLDSNDSRQVWQGVQHLTNYRANLGAA			
					ATAAAAT	DGDTALAEELNLFFARFEVTPPGTASPNPTVHS			
					TTATTGA	SLNLTVEEHEVRRTLRAVNPRKAAGPDGVPGR			
					TTATCTTT	VLRDCADELAGVFTRIFNLSLEQSTVPPCLKSS			
					TTTTATTT	TIVPLPKKTHISSLNDYRPVALTPVVMKCFEKL			
					TATTTTAA	VRSHIATLPRSLDPHQFAYRANRSTEDAVATA			
					TTTCTTGT	LHAALSHLEQQGSYVRMLFVDYSSAFNTILPH			
					TTTATTTA	KLVVKLGLDLGLPHSTCMWIHSFLSGRRQRVR			
					TTTAAAT	VGRHTSTALSLSTGSPQGCVLSPLLYSLYTHDC			
					TTATGAA	APVHHSNTIVKFADDTTVVGLISGGDESAYRD			
					TGATGCA	EVERLTAWCRGNNLLLNNTAKTKELIIDYRRKK			

					CTGACTG GTGAGCA CTTTAAA TTTCGTT GTACTGG TGACAAT GACAATA AAGGATC TATCTAT CTATCT	TDTPPLTINGDCVERVADFRFLGVHIGEGLTWS ANTSELLKKAQQRLYFLRILRKNNITQKLLVSF YRASIESILTYCICVWYTSCTVAQRKALQRVIN TAQKIAGCPLPTLEELHSSRCIKKAQNIKDSSH PGHSLFELLPSGRRFRSVKTRTNRFKHSFYPTAI TTLNTAN			
Pene lope	Pose idon - 1D_ PM	.	Petromyzo n marinus	CCCAG GTCGA CGTAG CCCTS AGAGK CTTGG GACAC ATCCC GGCAG ACGTT SCTAG ATCTG AAATC	TAACTGA CTGTGAT GTTTCCA TCCTACC TGCCTAT GTCAGCT GTTATTT CCGACTC TCTGTGT GTCTTTCT CTCTCTCT CTGTGCA TCTCTCTC	TTYNNXIQLJNDGXYTKINRDPTNRLATKXN NIITNSTSLIDSNKQPSAARPPRFYGXPKI HKEGTPLRPVSTINSPTYKLXXFLPETLLPFTG KTTSAXLNSAHLVSLIKDISLSPDDXLVSFDVV SLFTKVPVPDTIAIIEDLVQQGLNPDVPTLVRX CLTNTYFTWNGTFYKQNEGTPMGSPVVA NVFMEKFETDAIQTSTLRPSLWKRYVDDTFVI WPHGLPALNQFLNHINSIHPSIKFTMEVEKD GK LPFLDV LIEKRPDGSLGHSVYRKATHTDAYLK PDSHHHPAQKNSLIKTLHHRARCISDQHHLAS ELEHVHQALRMNGYS DRTIRKAVSPAMGRQP SEDPTNTITLPYIAGVTKKISKILNKHKIQVRFN TVHKIRD LIPSVKDVIPDMQYPGVYKLSCGCG	cd0030 4	186	270

			GCTGC	CGATCTC	SSYIGETKRHVSDRVREHKADLKHGRTNTSAV			
			TACCC	TACTCCA	ADHCFNAVGSHEIDFSKTKVLCKPTGYQRFV			
			TCAAA	GTGACCC	HEAIQIYKHRHNFNREDGYKLSNHWKDVINQ			
			AACGC	GGGTGCT	VKS			
			TAAGA	GCCGCGG				
			AACCC	GGTCCAC				
			ATTAG	TTTCTGC				
			CAACA	CCCCCCC				
			TCAGC	CCCCCCC				
			AAGGA	CCCTTCT				
			CGAGC	CTCTCCC				
			GTKTT	CCCCTGC				
			GAACT	AGCCGAT				
			CTCTC	CTCTCCA				
			AATTA	GTGCTGC				
			AAAAA	CGCGGTG				
			CAGAG	GTCCACT				
			ATGAC	TTACTCC				
			ATAKT	AGTGACC				
			CATCC	CGGGTGC				
			TCTCA	TGCTGCG				
			GCCGA	GGGTCCA				
			AAGGG	CTTTCTG				

				CAATG	CCCCGCC				
				CCACT	TCCTCCC				
				GTCAT	CCCCCCT				
				CATGG	TCTCTCCT				
				ATAA	CCCCCCC				
					TACTGCC				
					GAGGCCT				
					ATATAAC				
					CTCGTGC				
					CACGGCA				
					GATCTCT				
					CGCCTAC				
					TTCAGCA				
					GTTTTTTC				
					TTGCTGT				
					GGTTGTC				
					CCACCTG				
					ATGACGA				
					TTGCTTG				
					TGTTGCG				
					ATCGAAA				
					CGTCGTG				
					ATCTATT				

					ATTTTCT ACCTACG TGACTTT ACCGAAA GAACCAA AGAGTAT GGATTCC TGCAGTC ACGAAAG CTTCAA TCAAGAT				
Vingi	Vingi-1_PMa		Petromyzon marinus	GGGGAAGAAG	TGACGAT CATTAAG AAAAAAT AAAAACT GTTGCTT TTCACAA TGCCATA CGAAAGA AGAA	MTTCSSLLRAKAQRMSSHRRPNARQRYPGL KSNSTESISQGQSNVNVGPSLRIMQLNIEGLSS AKCDYIAKILQFQDVDVLALEETHLDRQCPRS KIIGYTLIAASHHPKHGLATYVKNDLKDTVQV VPHTSPHYIAIKINDVTVVNVYKPPGMEWPCP ALPQPSHPHIGDFNSHHTLWGYSANDRAGEQ LTSWAAAEDQHLLYDAKDRGTFKSARWKKC YTPDLCFVSRDQNGNPMPSTRTILDDFPNSQH RPVVIHIGLQIPLFRCAPKPRWNLRSAEWVNFT ADLEESVGMIPRAVNYRRFVGLIKAVAKANI ARGFRKVFTPCWTPESQLFRKFNADKKQETA KALLESNLQRRERWRETVEALDFAKSSRKA	cd01650	508	762

						WRLRLMGAAATPAYKPPQVSPDEIARQLINN SRTSIDKSTRRQVYKELYCTLDNCQETSEFSRP FSVEEVDAGIAQLKCGKASGLDDIYPEFIKNLG PRARKWLATFFTDIHSTSIVPKEWKEARIILK PGKDASDAQNYRPISLLSVCYKLMERLLLKRL MPTLEEIVPQEQAGFRTGRNCCDQVLALTSHIE AGFERQLKTATAFIDLSSAYDTVWRKGLLLKL AKAVPCRPTIRLITEMLSNRWFTVHLGEAASSP RRLNNGLPQGSVLAPALFNLYTSDIPATSCRKF TYADDIALATQARKMEDTEAVLEKDLKTLLEE YFKNWRLKPNPSKSVVSAFHLNNRQASKTLN VNFCGAKLRHDEHPRYLGVTLDRTLNYHHHL KQAKGKTKARVNIIQKLAGTTWGSASTLKIS TQALVHSTAAYCAPVWGASPHTNLVDAEINN ALRIISGSLKSTPLPWLPLANIAPANIRRDDAT TREYNKIMKNFTLPLHQDLKEPPRPRLKSRKPF WIRASALAAENKDTATRWSWAAMDVANK NIIHDPTVEVPGFSLPRRLWTTLNRARTQHGRCA AYTFHKWKIKESPNCDCGEVQTIKHITDECSIH AHPGGIAEIHASTTSLDWMRNLCHL			
Pene lope	Pose idon -	.	Petromyzo n marinus	CA	TAAACTA ATTTTCA ATGGTAC	SLSHTSAAVSSPNSKQILFKASLSLISERIRLLRF ELDTLDRTLQAHLLTLTQKMSPSDWNHVDNM TASSADSLSVKCKANLISKFNHLSKKNPVPHIT	cd0030 4	401	485

2_P				AGTATAT	KLDV NKIVINQSSTPISDNLKT V LARGLNFAVT			
M				TTTTTCA	PKHIPVDDIISQVDVALKGIGHIQADVARAEIA			
				AGTAGTG	TILKNAKKPISNISKNERSELANLKKQKKNRSD			
				GTATCCC	IVILSADKGNATVIMDKDSYNNKIQNLLNDGS			
				AAACTTT	YSKIKNDPTNSLTIKTNKIITSSTSLIEPTKRPSL			
				CACAGCA	KPSAARPPRFYGLPKIHKEGTPLRPVSTINSPT			
				GCTTATT	YKLARFLSDTLLPFTGKTSSAVLNSAQLVSMV			
				TCCCTTC	KDVVLSPDDILVSFDVVS LFTKVPVPDTIEIEN			
				ACACCAA	LVQHGLDPDTPALVRHCLTNTYFTWNGTFYK			
				CATTATG	QNEGTPMGSP LSPVIANVFMEKFESDAIQTSTL			
				TAAACTC	RPSLWKRYVDDTFVIWPHGLPALNLFLHHINS			
				CCCCCCC	RHPAIKFTMEVEKD GKL PFLDVL FERRPDGSL			
				CCCCCCC	GHSVYRKPTHTDSYLKPDSHHHPAQNNSLIKT			
				CCGCGCC	LYHRAHCISDQHHLTSELKHVHLALKRNGYSD			
				GGTCCAT	RTIRQAVSPSIQKQPLDTPVK TITLPYIAGVTKK			
				CATGTAC	ISKILNKHKILVRFNTVHKIRD LIPSVKDVIPDL			
				CTCCTAC	QYPGVYKLSCTCGDCYIGETKRHVSDRIREHK			
				GTACATC	SDLKHGR TNSSAVADHCYNAVGDHDIVFSDT			
				ATGCTCA	KVLCKPTGFYQRFVHEAIQIYKHRHNFNREDG			
				TTTTTATG	FKLSNHWKDVLNQTKH			
				CTAATTC				
				CTTTGTC				
				AAGGCGT				

					TTTTGTTT				
					CCTCCCA				
					CACAGCT				
					TAAAGGA				
					CGCGACC				
					TGGGCTT				
					GGCTGTC				
					TTCACTC				
					CAACGCT				
					GTCCTCT				
					GGACAGT				
					CCGGTCC				
					CTTCACC				
					TGATGAA				
					GGCCACT				
					TGTGTTG				
					TGGCCGA				
					AACGTCG				
					TGATTTTT				
					GTTGTTTT				
					ATTTTTTA				
					CCTACGT				
					GACTTTA				

					CCGAATA GAACCAA AGAGTAT GGATCCT TGCAGTC ACGAAAG CTTCAA TCTAGAA				
Pene lope	Pose idon - 1J_P M		Petromyzo n marinus	A	TGACTTG ATTTGTG TTTTCTTG TCAACCT CCACCGT CAGCAGT GTGGGTG TGTGGAG CGAGGTG CAAATGA TTGCCTG GCGGCCC TTTGATT GGTGTAG ACCCACT	KSKFHRLYGEGFHITIRSFNKL RDKKAHLIESL TFLKRCHNSKVIPISFTLTSAAVTSSNAKRILFR ASLALISERIRLLRTELDHVDKTLLKSHLTLTE KMSTSDWEHVDNMAASSAATMSAKCKDNLIS EFNLLSKKKPDPHIKKDVDKVVINLSNSPISDG LKSVLAKGLNFAVTPKHIPVDDIIAQVDVALK GLGHIAADVTRSEIAATLKS AKIPVSNISKDERF ELSQLKKRDDIVILSADKGNATVIMDKTVYND KIQDLLNDGSYTKIKRDPTNSLVTKTNSIITDSA SLIEQTKRFSLKPSAARPPRLYGLPKIHKEGTPL RPIVSTIGSPTYKLARYLSETLQPFTGKTTSSVL NSAHLVSSIKDISLSPGDLLVSFDVVS LFTKVPV PDTIAHIEDLVQQGLNADV PALVRHCLTNTYFT WNGTFYKQTEGTPMGSPV VANIFMENFET DAIQASTLRPSLWKRYVDDTFVIWPHGRPALN	cd0030 4	443	527

					GCGTTGG	QFLNHINSIHPAIKFTMEVEKDGMPLFLDVLIK			
					CGTTGTT	KRPDGSMHSVYRKVTHTDAYLKPDSHHHPS			
					TTCTCCT	QKNSLIKTLHHRAQCISDHEHLTSELKHVHQA			
					ACATTAA	LKANGYSDYTIRKAITPTVRKEPSEDPIKTITLP			
					CATTAGC	YIAGVTKNISKIMNKHKIQVRFNTVHKIRDIPS			
					TGCCCCG	VKDKIPDMQYPGVYKLSCICGSSYIGETKRHV			
					CCCATGC	SDRVREHKADLKHGRTNTSAVADHYFNAAGS			
					TGCTGAC	HEIDFSKTMVLCKPTGYQQRFBVEAIIHYRHNF			
					ACACTTG	NREDGYKLSNHWKDVINHFGS			
					GGCTACG				
					CCCATCT				
					CTAAACC				
					TTTATAA				
					ACAGTTT				
					GCCTTTA				
					GAGGCTA				
					ACCCCCC				
					TTTTCCTG				
					ATGACGA				
					CCGCTTG				
					TGTTGCG				
					ATCGAAA				
					CGTCGTG				

					GTTTTTGT				
					TTTTATTA				
					TATTTAA				
					TTACTGA				
					AAGAACC				
					AAGAATA				
					AGTCTAC				
					ATTATAG				
					ACTGCCA				
					GTACGCT				
					CACTCAG				
					TTTAATA				
					ACCATTG				
					CTACAGT				
					CTACATC				
					ATAGACT				
					GCCAGTA				
					CACTCAC				
					CAGCTTA				
					AAAACCA				
					GGCTACA				
					GTCTACA				
					TTATAGA				

					CTGCCAG				
					TACTCT				
					ACCAAGG				
					TTAATAA				
					CCAGGCT				
					ACAGTCT				
					ACATTAT				
					AGACTGC				
					CAGTACG				
					CTCACCG				
					AGGTAA				
					GAACCAG				
					GCTACAG				
					TCTACAT				
					TATAGAC				
					TGCCAGT				
					AACTCA				
					CCCAGCT				
					TAATAAC				
					CATTGCT				
					ACAGTCT				
					ACATTAT				
					AGACTGC				

					CAGTACA				
					CTCACCA				
					AACTTAA				
					TAACCAG				
					GCTACAG				
					TCTACAT				
					TATAGAC				
					TGCCAGT				
					ACACTCA				
					CTCCGAT				
					TAAGAAC				
					CAGGCTA				
					CAGTCTA				
					CATTAGT				
					GGTTTTTT				
					TGTTTTTA				
					TTTTTATT				
					TTTATTC				
					ACTTTAT				
					TTTTTATT				
					ATTATTTT				
					TCTATCT				
					ACGTGAC				

					TTTACCG AAAGAAC CAAAGA				
Pene lope	Copr ina_ Pc1		Phaneroch aete chrysospor ium	ATGAT CTGTG CATAC AAAAC CTTCG TACTTT TGCCG AGAAA GAGTC CCTTA AGGGG ATTCG TTGCT CCCGA CCTCG GGGAG ACTTTT CTCCA CACCA ATCGC	TAATGCG ATTCGAT AACTGCT GGCCGCA ATGGTCA GAGGATA TTTACTT AGCGGCA TACACTT TGGTTCA AACCCTT AACCTAA GCCCTAA GCCTAAA TATTCAT TAGCGTC TCCCTCA CAGGCCT AAACCCC	MPAQDTTTRRAVRATASSSTGHFSPFTRALSPS QASVVLGEQISFGTFDERCAAALAHCTLLDEVI ATLPKLYQVALSPVLTDIHHDATKRVEIAAQL ERLRQHKAAAGTLPSPKGLGSTPAIQFMKGFL ESQKPDFPGKAPDIVKACAAKLLDGFIEKAA EHAWYESRLQAMEVTRRVLTAAADTHWKDEL EQRYQVLEGFVRPPVEGQVPATVPSAEELHPL KFVTDPMVAQHENIRRSIPRLASTVIMLVVN RALTQQQRNKAQRDLKDQADVEMATDDTAV DSAVKRLEKQVEGLVSGMKALTNGKGVVSFS SSSRTHVEDAHPFVSGQSPSSRPFSLVDRARH EAQGVQLGKRLRQRQDDGHQGPSRRLRCSSG EAGSPQEGQGEGTREGQEDPEEIVAYVRAAPW RYDAPHSPDPVILMLPAPRAIQLLLERVPLEVK RAARFRSRVFLGPHVTVPEHLQMHVSSSLRYL MYSVPSSNVVREAWEDFKDRFRWRVFFTEKLI DSPNDDDPDFDPDYVVPKARKRFDDGVRTTP AYIELGLRKGDKFVDDYISNVMPKATSRSL NLVWLDELKDYLSANKYIVSPSKNLGVCILQ ADWATDQTRTLWNPSPNYRKLGLSLLEADEIM	IPR000 477	664	917

			ATTAC	CTGGAGC	EEKHERIFRAARLARRLGKRQLAEFLVSKIHG			
			AGCCG	GGAGTTG	KDAKTAGYTLGQFYGVPKVHKRPVKMRPIVP			
			ACGAG	TACCCGT	CHSVSQAPSATYVSKMLKPLVSRQPHVIHGSK			
			CGACA	TGAACGG	HLVRQLESLQLDKHRKAWIVSGDIVAFYPNIPL			
			TTTCG	TACAGCC	ENCIRVVTQMWRNSPEAEKLSPEERQLFLMSF			
			CCTCT	AGCAGTT	MTANKRLVIQFGDEYAEQIRGLAMGIACSPDL			
			CGTCG	ATAGAGT	ANLYGSHYENPILARPDMKSKFAFFGRYLDDV			
			GCCCG	ATTTTGTT	LGIVYASSAEEMSANEIRYDGVIEIWSVSE			
			CCAGT	GTACAGC	YHTPFLDLFLYLDPVVEKRLHWKPYSKPLNHRE			
			CCCTT	GTCATTA	RIPWASHHPKDVKKGTFIGEMSRLATLCSKLE			
			CGGGC	TTTAATG	HYRDAISDLNLYLARGYPDDAVRSWVKAHF			
			CTGGC	GATGCAG	SQRWERRLGEPVRDEVFVLKSHFNPAWSAFN			
			GGTTG	TGAGCGC	VHGLSEVVVESWLSSLKLDGANQSQDRTWH			
			CAAGG	CCGCCTT	AAESTDGSELKIFRQTTLGQFQGEQSSLYPGES			
			GGTAT	CGTGGCG	RPEGVAGRVGVANVAKPRVTYDPRDVMDVD			
			ATATC	TGGGCGC	ELVALSEPSFSAHDLSVVWTDNALRVPKSLRIR			
			CCAGT	TTCGGTG	YAMAELRDLADPQVNPALHVPLGVEDQVGDL			
			AGGAA	ATCGCCG	GPASDAESVASGDDGTTNSKPYRICTAKKDSG			
			ACCGC	TGATGGC	IVQEDLDIRSVGFHDRAWLVSRKKIRNLGDIV			
			TGATG	CGCGCAC	NSWKKDIIQAALNSDSVPHMHVDEWQ			
			AGTCT	GTGTCCC				
			TCACG	ACGTGTT				
			ACGAA	CGTCTCG				

				AGCTC	TGGCTGC				
				AGCGC	TAATAGC				
				GACAC	GTCTGTG				
				AACAG	ATCATGT				
				TGTCG	GATCTTC				
				GCTGC	GACCCAG				
				CTTCG	GGCTGCG				
				GTTTTC	GTAACT				
				CTGTG	CTGGATT				
				AAGCA	CGATGCG				
				TTCCG	CATGTTA				
				CGCTC	ATAGCCG				
				GCCGC	GCCCGAA				
				TCACT	TCCGGGT				
				ACGCT	CGGCTCC				
				CGCGT	GAACTGA				
				CTCGC	CTCCGCT				
				GCCCC	AGGTTCG				
				CCAAA	GGAGACT				
				CCCTA	TTCGCCG				
				ATCAA	GGGTGGC				
				ATCGA	TGCGTGC				
				GTAAG	CGGACTT				

				CACTC	CGTCCGG				
				TTCAC	GCACTTT				
				C	GTGGCGT				
					TGATCCG				
					ATCAGAT				
					CGAGATC				
					TCGAGAC				
					TTCGAAA				
					GGTGACT				
					TTTCGGT				
					AGTCCGC				
					TGCCCGC				
					TGGGGTG				
					TTGTTGG				
					CCCCTTC				
					CGCATGA				
					CCTGTAC				
					ATTCAAA				
					ACCTTCA				
					TACTTTC				
					GCCGAGA				
					AAGAGTC				
					CCTTAAG				

					GGGATTT				
					GTTGCTC				
					CCGACCT				
					CGGGGAG				
					ACGTTTC				
					TCCGCGC				
					CAATCGC				
					ATTACAG				
					CCGATGA				
					GAGACAT				
					TTCGCCT				
					CTCGTCG				
					GCCCGCC				
					AGTCCCT				
					TCGGGCC				
					TGGCGGT				
					TGCAAGG				
					GGTATAT				
					ATCCCAG				
					TAGGAAA				
					CCGCTGA				
					TGAGTCT				
					TCACGAC				

					GAAAGCT CAGCGCG ACACAAC AGTGTCG GCTGCCT TCGGTTT TCCTGTG AAGCATT CCGCGCT CGCCGCT CGCTACG CTCGCGT CTCGCGC CCCCCAA ACCCTAA ACCCTAA ACCCTAA ACCCTAA TACCC				
NeS L	Utop ia- 1_PI	.	Phytophth ora infestans	AGCTC GGCCT CGCGG CTGCC	TAAGCTG GTCATCA TGACCGA CAGGGCA	MLADPAALAAGLARAPPPSAPQDPSPAFPAG PAGQNPRAAAPARVEVHTVVAPPGRAGGMLP DPGLVEEPIQATYAHDAAQFECALCPYVAESM AVLVQHRRSAHRGTRFKDIFTSGCQC SLVFYA	cd0165 0	919	1187

			TTCCC	CTACCCA	RIVAASHAVACARRNQRAVPPAPTPVAPTRPE			
			AGGCG	GGTAGGG	ATPQPTGYLAAAMTAAAAAASSDTVVAATN			
			CCGCC	AACCGCC	MQSAVPAAAKTTGLQLVPPELEPALPQRASCH			
			GACTT	CTTAAAA	AGKRRRLNADEAVTPCTPTARVSPQTEVAMA			
			CGCGC	AACCCAG	PHDAPQDDTVLQREAAEPQPDPAATPGAQVQ			
			TCTGG	GAAGACA	RVEDTTAAQDDTVQQDHDADTAQVSPPRRTP			
			CGCGG	CAAACAC	TRWGPRPSSTQEPPMTGEPAAATLAARRPLTP			
			CCCAC	CCTCCAC	AATGTRATRWGPCHRAIGAAAIARLVTGLSTE			
			ACGCC	TTAGTGA	PAQPQRRQPPPPQEPPSQPEPLAAAATAAADIA			
			GCCGC	CATACAT	ATVAADIAAAAAANAAMDVDGGPAADETWLL			
			CGAGC	ATTTTAG	RFDGACRRNPGPGGAGAALFAPSGAVVWTCS			
			CTCCA	CCTAGAT	HYMPSRSETNNTAEYTAALLGVQSAVHHGAS			
			AGCGC	TTCAGTT	HLEVEGDSSLVIAQVKGTAFACRNARLRQLNR			
			GCCCG	ACGGAGA	VRHALRSVDTHKLRHIDRQANAHADRLANRA			
			TTGGC	GGTTACT	LDQRRTSSECGTHGSCMDSCLAVPTALAAQET			
			TTTCG	AACTGGT	PPAAPPSTSATPAEGNAMDDIAAEIARDEGET			
			CAGAC	AAATACG	FPVLPIGPGSAPERQPRLRLRQLSDEERDAAAD			
			GCAGG	AACACAT	ALQELADTMASKIEDADSWTSGEGYISSIPERI			
			GCTCG	ATTCTGT	REVLQPYATAPPQPGRSRRQRRRPPRVTRNQ			
			CGGCG	TCTAATC	REHRLDEALDDMAATQQATPRDQRAVRRARR			
			ACGGC	AGTGTGA	RVGRVRASMAQQELRHEFAKDESKCVAKILK			
			CCTCG	AAACTGG	TASTETAAEDEHPETCPIDAATLHAHFTGVNA			
			CCAGC	TTTTCGC	PRTDFDYDATSGREFRAAMSDLPPATVEIDAF			

			CCCCA	CTTTTGG	DAELTIDEVEDQLTRA AKTSSPGHDGIDYGIYS			
			AGACC	CGGACTT	RFAAQLVPLLHAVFQFCWRHRRVPRLWKVGI			
			CCCCC	TTTCACT	VRLIHKKGDPRQPTNWRPICLQPTIYKLYSGLL			
			TACGA	CGCATTT	AHRLSRWLEGNDRLPMAQKGFRAFNGCHEHN			
			TGTGG	TTGGGCA	FMATTLLDQTRRQHRKLYQVWYDLRNAFGSL			
			CACCA	ATCGTCT	PQQLMWRVLRHLGVDSGFIDRCRDIYRDSAFV			
			CCCGG	GCGGCTA	VANAADGATDPVRQEVGVYQGCPLSPLLFVA			
			CAGGG	GCTTGCT	ALVPLVRRLEKLDGVGVPLADGVRPCTTAYA			
			CGGCC	AGCGGCG	DDLKVFSDSAAGIRKCHDTVAGFLAWTGLRA			
			GGCAG	GACGAGC	NPGKCASLAVTTNARGNPTRDSSMRLEVHDA			
			GCTGC	GGTCTCC	AITTL SLHESYRYLGVGDGYDHVRHRLQLEPK			
			CCGAC	GGGGGCG	LKQLKREAVALLTSG LAPWQVVRALKVYVYP			
			TCGGT	TTCACCT	KVEYALRHLRPLQSQLQAFDRV VSKGLRHLLS			
			ATCGC	TTCCCCC	LPRSATSEVLYAPTSSGGLGLQPLVELHRA LQL			
			CGGGT	GCGAGGC	AHAWQMLH SKDPAIQAVARAQACQVVRKRY			
			GCTAC	CAACTAC	RLQEDHWRGRDDEL VRSFLNSELAASPHAEVL			
			ACTCT	ACCGATC	RRNGDIASLWSDVQRWLRIYHLRFEHCDETEA			
			CAGCC	TTCTCTA	HGPLSFRVPHH NKWLTHKTVLRHVKLHLKIR			
			GCCAC	CACTTTT	HQTRWKGMVDQGKTVRVHGGVGAKFMTTG			
			AGCTC	CTAATTC	AGLSDDDYRFGVKGR LNQVD TNSVLKRKRLR			
			GGGCC	GCCTCCG	AHTTCRDPTCSSAETLAHV LNHCESNMDAIRQ			
			TTGGC	TCTTCGG	RHDDALEQIGSKIR GALDRAKSPTELRLNQTVP			
			GGTCC	TCTTCGG	EYTGAALRPDIVLRNVAAKTMVIADLAVTFED			

				GCCAT	TTGTCGG	QAARARHSSLQLSHDHLTKYQPIVAELQHKG			
				TGGCC	GCTTTTTT	WRVQTAAIVYGTLGSLVQPSNFKAYTEKFKLH			
				CTTGG	CTTTTTG	KREARQLDLQLSSH CIRASHRIWGW HCRQHR			
				AGCTC	ACCAATC	DRQRSGTASRASRGSGGAPRRTSQAPARR			
				GACAG	AGAGCGC				
				CGACA	GCCATGC				
				GCAGC	GCCTCTT				
				GACGA	CTGGCCA				
				CGAGG	ATCAGAG				
				ACGCT	ACCGGGC				
				CAAGA	CCTGTCC				
				CCCCC	TCGGACA				
				ACGCC	GCGAGGC				
				GCCGC	CTCCACG				
				CCCAG	GCCAGCC				
				AACCC	AATCGAG				
				CCAGA	TCTCGGC				
				AGACG	AGCGACG				
				TCGCG	CGTCTTT				
				AGTGT	CTATAGC				
				GCTTG	GCAGCTG				
				CCCCA	ACGAGGC				
				CCCGG	CGATCTG				

				CAGGG	GCGGCCC				
				CAGGC	CCGATTG				
				AGC	GTCCGAC				
					TTTCGGC				
					CAATCAG				
					CGACGAC				
					GAGGGGG				
					CAGGGGT				
					TTACACT				
					TTTGCCC				
					CCGTTTC				
					GACTTCA				
					ACTTCAG				
					GCCAAAA				
					TGGCGAT				
					TTGGACC				
					CTCCACG				
					CGCCGTG				
					CCACTGC				
					TCGGCAC				
					CGGCGGC				
					GATTCAG				
					CGGGTGC				

					AACTTCG				
					GGCACGT				
					GTGCAAC				
					ACATGCA				
					GCGCCCA				
					TTGCACG				
					CCAAGCG				
					GCATCGC				
					GGGACGA				
					CGCCTCG				
					GCCGCCC				
					AAGCGCA				
					GCCCCGC				
					CCTTCCA				
					GCACGAC				
					CTCGCGC				
					CGTTTGG				
					CGGATCG				
					CCATCAA				
					GACGTGC				
					GAGAGCC				
					AGGCGGG				
					GTCGGGC				

					AAAATAT ACTTACT CTAAGTA TGCCCGA ATCCCTG CCCTCTC AGGCTGA ACGCGGC CCCATAC TTGATCT AAGTATG GGAGGAT CCCTGGC CTCTCAG GCTGTAC GCGAGAC CC				
NeS L	Utop ia- 4_PI	AAT U010 0128 1.1	Phytophth ora infestans	AGCTC GGCCT CGCGG CTGCC TTCCC AGGCG	TAAACGG GTCACCT GACCGAC AGGGCAC CACCCAG GTAGGGA	MLADPAALAAGLARAPPPPSAPQDPSPAFPAG PAGQNPRAAAPARVEVHTVVAPPGRAGGMLP DPGLVEEPIQATYAHDAQAQFECALCPYVAESM AVLVQHRRSAHRGTRFKDIFTSGCQC SLVFYA RIVAASHAVACARRNQRAVPPAPTPTVAPTRPE ATPQPTGYLAAAMTAAAAAASSDTVVAATN	cd0165 0	934	1198

			CCGCC	ACCGCCC	MQSAVPAAAKTTGLQLVPPELEPALPQRASCH			
			GACTT	TTTAAAA	AGKRRRLNADEAVTPCTPTARVSPQTEVAMA			
			CGCGC	CCCAGGA	PHDAPQDDTVLQREAAEPQPDPAATQGAQVQ			
			TCTGG	AGACACA	RVEDTTAAQDDTVQQDHDADTAQVSPPR RTP			
			CGCGG	AACACCC	TRWGP RPSSSTQEPSPMTGEP AATLAARRPLTP			
			CCCAC	TCCACAT	AATGTRATR WGPCHRAIGAAA IARLVTGLPTE			
			ACGCC	AGTGACA	PAQPQRRQPPPPQE PPLQPEPQAAAATVAADIA			
			GCCGC	TACATAT	ATVAADIAAAAAANAAMDVDGGPAADETWLL			
			CGAGC	TTTAGCC	RFDGACRRNPGPGGAGAALFAPSGAVVWTCS			
			CTCCA	TAGATTT	HFMP SRSETNNTAEYTALLLGAQSAVHHGAK			
			AGCGC	CAGTTAC	RLNIEGDSHLILSQVRGAFACNNKRLRSLRNR			
			GCCCG	GGAGAGG	VQASLRQLDWYRLQHIDR KANQHADRLANR			
			TTGGC	TTACTAA	ALDLRRTVTECGPHAETRNRCFQTPQPLVEPG			
			TTTCG	CTGGTAC	ETHCVPGSDEVLAANTAMEDATAVPTEDDEA			
			CAGAC	ATAAAAT	EVAARDGGEVFPTIAIGPDSAPAKQPRLRLKKL			
			GCAGG	TACACAT	DEDDFDAAAAAVTRVSEELASKIVDAGDWTS			
			GCTCG	TCTGTTCT	GEGYISAIPERLRAALRPFALPTQPARPQPREPR			
			CGGCG	AATCAGT	MQQPPRRPPRVTRDHLEHRLDEALDTMENVQ			
			ACGGC	GTGAAAA	RSTPQNQKAVRRARRRVGRLRSAMDRTLRK			
			CCTCG	CTGGTTT	KFATHERECVAEILRRASTE EEAANPSQEKCPID			
			CCAGC	TCGCCTT	RATLHEYFTATSTQRTPF DYDSAKGTEFRTFLE			
			CCCCA	TTGGCGG	VMSTPSHETSALTA EPTLDEIEDQLAHVKAGSS			
			AGACC	ACTTTTTC	PGHDGVGYDVYRRFQVQLLPLLHAAFRFCWR			

			CCCC	ACTCGCA	HRRVPALWKVGFVRL LHKKGDPQQPNNWRPI		
			TACGA	TTTTTGG	CLQTAIYKLYSGLLARRLSKFLEANELLPMAQ		
			TGTGG	GCAATCG	KGFRANFGCHEHNFVATTL DQTRRMHRRLY		
			CACCA	TCTGCGG	QVWYDLRNAFGSLPQQLMWGVLRQLGVTEE		
			CCCGG	CTAGCTT	FVARCSGIYEDSYFVVGNASDGATEPVRQEVG		
			CAGGG	GCTAGCG	VYQGCPLSPLLFITALVPLLRALENQDGVGVPL		
			CGGCC	GCGGACG	ADGVRPCATAYADDIKVFCDSATGIQRCHALV		
			GGCAG	AGCGGTC	TRFLEWTGLQANPAKCAFLPVTRSQHSNPTRD		
			GCTGC	TCCGGGG	RDIELRIHGEAIATLGLQESYRYLGVDGDFDHV		
			CCGAC	GCGTACA	RHRLQLEPKLKQIKREAVALLHSELVPWQILK		
			TCGGT	CCTTTCC	ALKVYIYPKVEYALRHLRPLKSQLQGFD SAIV		
			ATCGC	CCCGCGA	RGLRHLLRLPENSHDGLFFSPTSAGGLGLLSLV		
			CGGGT	GGCCAAC	ELHEALQVAHAWQMLH SKDPAIRAIARTQVG		
			GCTAC	TACACCG	QVARKRFKLVEEHWRGREDDLAQRFLNTELA		
			ACTCT	ATCTTCT	ASPHATETRRNGDIGSLWNDVRDTLQTLGLKF		
			CAGCC	CTACACT	AAGDEEEAPGLLQLRVPHHTKWLSHSTVLRH		
			GCTAC	TTTCTAA	VKLHMKLRRMDTWKSKVSQGTTVREHGGVG		
			AGCTC	TTCGCCT	SRFITAGAGLSDAEYRFAIAPRAHLIDTNSTLK		
			GGGCC	TCGTCTT	RRRLRANDTCRAPGCSYTEPPAHILNKCSPNM		
			TTGGC	CGGTCTT	DAIRKRHDDALERIADALRRKVEKSGGRLEVA		
			GGTCC	CGGCTGT	INKTVPEYDGAALRPDIVLRNTETKRAIIADLAI		
			GCTAT	CGGATTT	THENQPTDATTSSALQQSRDNKITKYQTVAAA		
			TGGCC	TTTTCTTT	MMRAGWRVRVTGIVYGS LGSVLPSNFKVYTE		

				CTTGG	TTGACCA	LLALLKRDARRLNRQLSSH CIRASARIWSAHC			
				AGCTC	ATCAGAG	RRHRERQRSGNASRASRGSGGAPRRTSQASAR			
				GACGG	TGCGCCA	R			
				CGACA	TGCGACT				
				GCAGC	CTTCTGG				
				GACGA	CCAATCA				
				CGAGG	GAGACCG				
				ACGCT	GGCCCTG				
				CAAGA	TCCTCGG				
				CCCCC	ACAGCGA				
				ACGCC	GGCCTCC				
				GCCGC	ACGGCCA				
				CCCAG	GCCAATC				
				AACCC	AAGTCTC				
				CCAGA	GGCAGCG				
				AGACG	ACGCGTC				
				TCGCG	TTTCTAT				
				AGTGT	AGCGCAG				
				GCTTG	CTGACGA				
				CCCCA	GGCCGAT				
				CCCGG	CTGGCGG				
				CAGGG	CCCCCGA				
					TTGGTCC				

				CAGGC	GACTTTC				
				AGC	GGCCAAT				
					CAGCGAC				
					GACGAGG				
					GGGCAGG				
					AGTTTAC				
					ACTTTTG				
					CCCCCGT				
					TTCGACA				
					TCAACTT				
					CAGGCCA				
					AAATGGC				
					GATTTCG				
					ACCCTCC				
					ACGCGCC				
					GTGCCAC				
					TGCTCGG				
					CACCGGC				
					GGCGATT				
					CAGCGGG				
					TGCAACT				
					TCGGGCA				
					CGTGTGC				

					AACACAT				
					GCAGCGC				
					CCATTGC				
					ACGCCAA				
					GCGGCAT				
					CGCGGGA				
					CGACGCC				
					TCGGCCG				
					CTCAAGC				
					GCAGCCC				
					CGCCCTT				
					CCAGCAC				
					GACCTCG				
					CGCCGTT				
					TGGCGGA				
					TCGCCAT				
					CAAGACG				
					TGCGAGA				
					GCCAGGC				
					GGGGTCG				
					GGCAAAA				
					TATACTT				
					ACTCTAA				

					GTATGCC				
					CGAATCC				
					CTGCCCT				
					CTCAGGC				
					TGTACGC				
					GGCCCCA				
					TACTTGA				
					TCTAAGT				
					ATGGGAG				
					GATCCCT				
					GGCCTCT				
					CAGGCTG				
					TACGCGA				
					GACCCGT				
					ACGGCCG				
					AATCCCT				
					GGCTTCT				
					CAGCCTG				
					TACGCGG				
					GGCTATA				
					CTTGGTC				
					TAAGTAT				
					GCCCCGG				

					TCGCTGG CCTCTGA GCCCCGTA CACA				
NeS L	Utop ia- 2_PI		Phytophth ora infestans	GCTCG GCCTC GCGGC TGCCT TCCCA GGCGC CGCCG ACTTC GCGCT CTGGC GCGGC CCACA CGCCG CCGCC GAGCC TCCAA GCGCG CCCGT TGGCT	TAGGCGG AAACCAG GCCCAAG ACGGCCG ACAGGGC CCCACCC AGGTAGG GAACCGC CCTAGAA ACCCATT TCGGTGG TCGACTC GAAGGCC TTACCTA TTTTTTCC TTAGACA TTCAATT AGGTAGC GACCAAA	MLADPAALAAGLARAPPPPSAPQDPSPAFFPAG PAGQNPRAAAPARVEVHTVVAPPGRAGGMLP DPGLVDSSPAAATAATPAPVAATATTARAAA RVAVEHHAAHAEPNQEHLPMARVLVEPMQVD ECSSCDRSTLTADDGSGDDVAAPSSMLSNDVA APMDVDSGTSCPPTLQQPLQRPRALHVGSKRR RLDADDGEEAHQLQEEEEAGIHAPALRLSAAS AQPASVLAVYTHNASRFDCTLCAITAGSFASL LTHRNSRHRRTAFLDRFSAGCACGVPFASRLA AARHAQACASLSSAPSAEASSAAGTSSPTADG ADSTVSAVAHAEPGLPHHNDTELTASPPLVSSS DVEVQATKTEATDNRWGAPLPRVLVASRIAG RLAQVPPPRWGPPLPRTTIAARIATRLAATPAP RWDPPPLPRSLVVSRIAARLLPALPDAPACEEEA KDSDTMDWAPTWTNEETKESEPHDEAPGQVD EETIDDADGEWLLRFDGACRANPGPGGAGAA LFKPSGPVVWTCSHYDPSTTATNNTAEYTALL LGARAAADHGVTKLRIEGDSTLVIQQVRGIFA TRSTRLRALRNKVKLELARVGFSLSLHHIDRQA	cd0165 0	1026	1289

TCCGC	TTACAAA	NGHADRLANAGLDRRRRTKLECSVHPDGRGCT		
AGACG	TTTGGTA	NTSVATAAPTAPAAPLPSARPPASTAAPSPDDD		
CAGGG	ACGAGTA	HSDQGDIDDGEVYAAMCISPDAVPHRRPRLRL		
CTCGC	AGCCAAA	RRLTDEESEEGNVVERLAASLA AKIADAPDW		
GGCGA	TGGTAAT	ETAEGYITALPYALYDKLQPYSQSQHQPPRQQ		
CGGCC	ACACAAA	QQQQRQRPRQQQQTQRQRQRCKRGGGSQH		
CTCGC	ACTTTTCT	RQRKTRRRRPVRTRHHREHRIDEALDDLHAL		
CAGCC	GTTCTAA	ESRRPQDRTAISKARRRVGRIRSALDQHQLRH		
CCCAA	TCAGTGT	RFDTDEKACVDGILAAARDKDRAASVTTTAQ		
GACCC	GAAAACT	TAAPPHSAPASAPSSAVDDGICPIPGDLLHAFFT		
CCCCT	GGTTTTCT	DVNTPRTEFDADSPIGARFRELAQLPAAIAAT		
ACGAT	GCCTTTT	ELLMEPPSPDEVEDQLQVRGTSSPGLDGVGY		
GTGGC	GGCGGAC	DVYKTFTQQLLPALHAAFSRCWTDQRPQSW		
ACCAC	TTTTTCAC	KLGVVRLLFKKGDRQDPANWRPICLQQAVYK		
CCGGC	TCGCATT	LYAGILAHRFTRWLDANTRHADAQKGFRVAVN		
AGGGC	TTTGGGC	GCGEHNFLAATLTDNARRRRRELHV VVYDIK		
GGCCG	AATCGTC	NAFGSVPHELLWEVLRRMGVPAQFIACCQGIY		
GCAGG	TGCGGCT	DAAAF TVGNAADGTTAPIQLRLGVFQGCPLSP		
CTGCC	AGCTTGC	HLFTAVISPLLHALKRLPGTG VQLSAVDRPGAS		
CGACT	TAGCGGC	AYADDLKVFSDTKDGITRQHQLVTDFLRW TG		
CGGTA	GGACGAG	MVANPSKCSTMSVQRDNRGVLKTANLTLQLD		
TCGCC	CGGTCTC	GAQIPALGMTEAYAYLGIGDGF DHVRRRVEL		
GGGTG	CGGGGGC	APKLRELKADTTALMQSGLAPWQVVKALKV		

			CTACA	G TTCACC	YIYPRVEYALRHLPFQQQLQGFDRHLARGLR			
			CTCTC	TTTCCCC	HLLRLPTSATTEFLYAPTSRGGLGLLPLTEVHG			
			AGCCG	CGCGAGG	ALQIAHAWQTLHSPDPAIRRIARVQLRQVADA			
			CCACA	CCA ACTA	RHRLDAEHWKERGEELCERLLNSQLG TSAHA			
			GCTCG	CACCGAT	PPKRRNC DIGSLWVDVQRHLSLGLQLQTAPA			
			GGCCT	CTTCTCT	DTHTGAPAQPLQLRVPHHDKWLTHKDVLRHV			
			TGGCG	ACACTTT	KLHIKNNHWHRWTSMRDQGKTARAHGGECS			
			GTCCG	TCTAATT	GFLTQPRGMWEADYRFAVAGR LNQVD TYSVL			
			CCATT	CGCCTCC	KRRRLRSHDRCRQPGCHRAETLAHV LNHC PG			
			GGCCC	GTCTTCG	TMDAVRGRHDGALKRIERELHASATDRRDRV			
			TTGGA	GTCTTCG	ELRVNQTVP SLAGPALRPDLQLYNHTKKTVA			
			GCTCG	GCTGTCG	VVDLAVAFEEQASDDASSSALSLIASHKRAKY			
			ACAGC	GATTTTTT	DRIKRHLERQGWKVHLSALVYGS LGAVASGN			
			GACAG	TCTTTTTG	YQVYTTHLGLLKRDAKRLDRQLSVECIQSSRR			
			CAGCG	ACCAATC	IWNLHCSQHRTRQH QARPSQGPRGSRATETGG			
			ACGAC	AGAGCGC	TPSQTSRR			
			GAGGA	GCCATGC				
			CGCTC	GACTCTT				
			AAGAC	CTGGCCA				
			CCCCA	ATCAGAG				
			CGCCG	ACCGGGC				
			CCGCC	CCTGTCC				
			CCAGG	TCGGACA				

				ACCCC	GCGAGGC				
				CCGCA	CTCCACG				
				GACGT	GCCAGCC				
				CGCGA	AATCAAG				
				GTGTG	TCTCGGC				
				CTTGC	AGCGACG				
				CCCAC	CGTCTTT				
				CCGGC	CTATAGC				
				AAGGC	GCAGCTG				
				AGGCA	ACGAGGC				
				GC	CGATCTG				
					GCGGCCC				
					CCGATTG				
					GTCCGAC				
					TTTCGGC				
					CAATCAG				
					CGACGAC				
					GAGGGGG				
					CAGGGGT				
					TTACACT				
					TTTGCCC				
					CCGTTTC				
					GACTTCA				

					ACTTCAG				
					GCCAAAA				
					TGGCGAT				
					TTCGACC				
					CTCCACG				
					CGCCGTG				
					CCACTGC				
					TCGGCAC				
					CGGCGGC				
					GATTCAG				
					CGGGTGC				
					AACTTCG				
					GGCACGT				
					GTGCAAC				
					ACATGCA				
					GCGCCCA				
					TTGCACG				
					CCAAGCG				
					GCATCGC				
					GGGACGA				
					CGCCTCG				
					GCCGCTC				
					AAGCGCA				

					GCCCCGC				
					CCTTCCA				
					GCACGAC				
					CTCGCGC				
					CGTTTGG				
					CGGATCG				
					CCATCAA				
					GACGTGC				
					GAGAGCC				
					AGGCGGG				
					GTCGGGC				
					AAAATAT				
					ACTTACT				
					CTAAGTA				
					TGCCCCGA				
					ATCCCTG				
					CCCTCTC				
					AGGCTGA				
					ACGCGGC				
					CCCATAC				
					TTGATCT				
					AAGTATG				
					GGAGGGA				

					TCCCTGG CCTCTCA GGCTGTA CGCGAGA CCCGTAC GGCCGAA TCCCCTG GCCTCTC AGCCTGT ACGCGGG GC				
I	I- 2_R Pr		Rhodnius prolixus	CACGT TGTC GAGAC GATCG TGAGG ATGGA TATGG TGGA GTGCC ATATT TATCA GTAAA	TAGGATG CAGCTAA ACTCCAT AGTTGCG ATTTTAG TTCTTAA TACTCA CACAAGT ACTATCA TAACTAC ACTACTA ATTACAC	MWGGHRENLEGRALAEIDSYNLCTLNDGSP TMIPRPGARAQPVDTLCSPDLMSELSWEVIQE PLGSDHLGILVKSFNQRQATSGARIRHPKAGG WNLSKADWEKYRMVLDLKFSTSESTSCYDY LTQEIRSAALGSIPRFPQWLYLNPRAIKPWW DDDCNRVKQESIIAYESYKRAPTLENYVNLNR KLAISRRTYKEKKRIMWYEFCSLDRQTPSNVI WKRIRSLKKGYSPPQVEVTDDGHWVEEFHKS LPDYASQFIDLSNPIGDEDLNXPFLWELKYA LKKGGSTSPGQDGIHYQMLQNLFPALLSLLRI MNEMWVGGAMIESWRTHLIFPLLKPGKDPNS YQSYRPIALSSCVCKTYERMIKHLRLDWWVEE	cd0165 0	340	595

			AAATT	CACTCTA	YNKLAPHQLGFRRHRGTLDCLAAFVNEVHLN			
			GCCTT	AGCTCAC	RTENKPLLAAFVDIHKAYDNVNIAILKDKMLV			
			TCGAC	ATCAAAT	MQVPGRVIVNIINMLQQRQLIIRRGVLVTGPRT			
			CCAAT	AACGAAG	TGIGLPQGAVLSPLLFIYTSDLRDWLVPFIRIV			
			AGTGC	TCACATT	QFADDFTVYTGGKSEENLVEIMGQALHDLNA			
			TGAAT	CAGAGAC	WCNHLGFRISGNKSSVTMF SRKRQNILAGTIQL			
			AATCC	CATTGTG	GPFKVSQSPIVKVLGLYVDHKLTWRPHIKRIIQ			
			ATTAA	TTGTTCT	SCEKALNVRLRYVAHVTYGTDPKIALLLYRNLI			
			GGGGC	AGGAAGG	LSKIDYGCGIYGDACQTSLSGLDKVQYKAIRIC			
			AGAGT	AATGTCA	LGAMRSTPTNVLLVEAKIMPLHIRRELLLGKFL			
			TCAGA	GTACGCA	SKRLIKFKEVSLMRVSELSEVVRSHRFWLKKK			
			TATGT	GCGCTCA	TPLMCEMFRVTEKLYSGLEVLHASPRRSLHYE			
			GGGGC	AGATATC	VSCKESPYVHLDLYDKVKEEWGDLTYLAFIAR			
			TAAAA	CTTATCT	MREIINPYLLFTDGSKSEQGVGCAVLDPQTGS			
			TCCTA	CATGGAA	FRLYKLPNNASIFTAEIVAIWNAVILAEQIQSND			
			TTTAG	GCCACGA	IIVILSDSKSALQKIKRPGDVFDPIIIIEIISHLINA			
			GAGCG	CGCAGAA	MGRHLHLVWLKGHAGIVDNEKVDVLAKRAILL			
			TCAGT	TGTAGTA	GEHTPVLNLADFEGALRKRCYDAWQKLYDLE			
			ATTAC	GAAGAGT	KEIKGKYYADLQPRIPNKPWFVEVLEGERQFY			
			GGTCC	GGCTGGC	TTIMRLRFNHNLTNPHLHKIKLLDSPKCDGDSG			
			TTAGC	ATGACGA	EDGDIDHLLFRCSRYNRNRDKFLKELATHVRF			
			ACTTA	AGAAAGA	PCSAKSILAIMNPDALRIIYKFIRENKIKI			
			CTGCC	AGGGAAA				

				GTCCT	GAGGCTG				
				GGGCA	CGCTATT				
				TAAAG	CCTCAAA				
				CCTCA	TTGAATA				
				GTAGA	TTGGAAA				
				GGAAT	AGATAGG				
				GGAGT	GCTGGTT				
				AATGT	TTTTTCCA				
				ACTAA	CCAACAT				
				ATCAG	ATTGGTA				
				GTTTC	AATTGAG				
				GGGTC	GCGCTGG				
				CTCTT	GAGCCAA				
				ATTCTT	CTGTAAT				
				GGTGG	TGTTCAT				
				CGACT	GCCAATA				
				TTAAT	TGTTGGT				
				GCGCT	TTTTTCCA				
				TCATC	GCCCTCA				
				CA	TATAAAG				
					TAATTGG				
					TTAATGT				
					GCAAACA				

					AGACCAC TATAACA ATAAATA AATAAAT AAA				
RTE X	RTE X- 2_R Pr		Rhodnius prolixus	GAATC CAGAA ATAGA GTATC CAAAG ATAAA TTTATT AATAG TAACG GTTAT AAACT TTTGG ATTTTT TCGAT AGCTT CAGTC TTATA GTCCT	TAGTTTA TGTTTAC TAAAATT TACTTAT TTATTAA AATATAT AATATA	MVSSFRVDHQLYSDHMPITTTLLIHDKQQIYND NTLPLLPKLNWSKQNLQKYQQSISFYLNHIVY KYDTLDAEVNYLNDIIRSSAIGNNNKKGPKVW KADWFDVECDRARRKSFKQLRFRRTNEVED KHLYLKYNFDFKRLCKNKKQKYYVDCARRL YNVKSSKEFWQQVRAFQKRKPWVLGNVSPEQ LFEYFCSLHNSSVTARLVFDVSHRVD AFLDKP FSFDELELVLSKSKENKAPGTD RIC YEFYKF API VFLNRLLLLFNKIFETSKIPASF KESIIFLYKKG NKNSASSYRGISFINVICKLYSGLILHRLNNWL ESNNILKESQAGFRS NYSTIDNLFNL TNIKIKLS KKKQKVYAFFVDLSSAFDSIDRFS LIYKLHRTG LSTKIVTSIKNLLSGTTASVMCRNGLSPNFGTNI GVKQGCLLSPTLFSIFINDLEECLGGGVVISDN CKIKLLAYADDVLLASNPISLQLMINEMESY CYRWKLKINLQKSKIMIFRNGGRPANNEVWY LGTQKVEVVNEFKYLGHILTPTLSYTKHFQNKL SKAKMSLNTIWHHFMDNTSIPFSTKIRIFNSVN	cd0165 0	280	540

				TAACG GTAGA TTTGA AAATG ATTTG GAAGG CGAGT TCACC TACAT TAGCG AGGTT GGCTG TTCGG TAATT GACTA TTCCT GCATT AGTAT TAAAG GTATT GAC		RSMMCYGAQVWGWQRFVEVERLFYFFIKKLF KSPKTTPSAFIHTETGIDPIYLYTFEMHFKYLVR VLQLPDNRYPKILCHEIISKILWFLEIKNLITKF NLPNDLNWQNIGNFTLWEANFNILIKSLQADM NNTFQLKLFNSGSHPLYKNLQYNLGDNPYLS SHPFHLIRMIFKVRGDLLPLNYKFYKNRDLDC LMCASGEVEDLFHFLAVCPALSNMREQWFGS RTLPAQHVAAYLNGKDWEALAKYCLAALSRR ALLT			
RTE X	RTE X-		Saccoglossus		TAGTACT GTAAGTT	LNNKVNVEDKSFSQTIVNDISSMESLDDYKHII NQDFTNKEVLQHILLLKSGKASGTDNISNEM	cd0165 0	96	361

	18_S K		kowalevsk ii		TTTGTTGT TCTGAAT GTTTTGTT TTGTTATT GTTATTTT TGGACTT TACTCAA GATTTCA TTGGTAC GTTCCAA ATAAAAA ATTCTAT CTATCTA TCTATCT AT	KHGAPFIINSLTKMFTQVLNSGIMPLHWNSSLII PIFKSGNPKDANNYRGIAISSCVSKLYTMMLNT RLTKYLQETNQIVDNQFGFKKNHRTTDNIYIL KSLINKYVNCKLKKSGNYLFSCFIDFSKAFDSV WREGLLYKLLRLGVGGNFYKAIAIKAIYKNTKY SIKLSAGITPYFNSCRGVKQGCNLSQFLNIFVN GICDDVKNFDPVDLNGCLKLNCLLYADDLLLFS TTQEGLOKCIDAVQNYCDNWHLKVNINKSKIII FNRLGKNLENKFQFYLDATKLQCVNYNYLG IVFQTNGKFHNAMSRLQDKAQKALFKMYKM LWISEECYSVSLAKKLFNSIIKPTALYGADVWG GVMLNIYKGS GKNKCTVDNCNFETKLFSDKLI PEKLHIKFCRVLGVHRKASNDPCRGELGQYP LILDICVSVMKYYYQRLVCKKKTILSLMQPMKN PKLLCMTHGLTLYKNYKKNNSIIDAAYEESKT LVHDPWLNFIQKLLVITDLTNSDPHQNNFCDIF KEKLKCLYRMKWKNLLNNCKLDFYRCAKE NFEFEHYLEKIKNKNHRIA YTKFRTSAHFLPIES GRHKGILKENRLCNLHECNEVGDEKH YILYCN NCQIKEQRSRFLAALYIYAPKLEDFSMDELFP II VSSNDFFILSIFAKFIFYILDMF			
RTE X	RTE X-		Saccogloss us	CC	TAATAGT TGTTAGT	CINDSLLET FVSKHDIVCLLETWLDYTD SVSVP NGYLYFRVDRKKHKNARKASGGMIIFYKSKF	cd0165 0	551	812

5_S	kowalevsk	GTTATCT	RKGINKFISKYKESLWLKLDKHFFGLENDLYIN		
K	ii	GTTTTAA	STYIPPIGSSYYNSSDLTGEDIFNIIAEESLHYSN		
		CATTTTA	HGNVCILGDLNSRTVNLNDTLDDINISDIDIPYT		
		TTGTATTT	ELMDNNRQSMDTMNNKYGQYLCDLCLTTDM		
		TTTATAT	IIANGRIVGDLMGKITCFQYNGVSQVDYCVVQ		
		CTGTATT	KSFLPKINYLKINELSHISDHCSLSLAINTNYVN		
		TCTTGTT	EHSIDISLTPYKKYNWNNSSKENFVTQLNSKDS		
		ATTGTGT	VDKMNTFLDNHFNLEESDKAISDFNDIMHFTA		
		TGTTGTT	DKALKFTAISRKFSRKNKRKNKQKWFDNDL		
		GTTTTTG	AHVRKCLKDISRLVDKFPFDQNLRMQRMSINK		
		ATTATTG	RYKKLVKLKKNKFKEKLLHDLNMNLETKNPSE		
		ACCATGG	FWDIIKQIKEYNDLDTSKHPPSIASTEWTNYFS		
		TTAATGT	KLYNISTTDSNSVLDSEVTNFVDNMDHNATSN		
		ATTCCCA	PDLKEISIAEIIKAIKKTKV NKSSGPDSILNEMF		
		ATTCATT	KCYSPLLLACISKLFNHILSSGCFPVHWSTSYL		
		AATTTGA	VPIHKS GHIWDVN NYRGIALSSATSKIFNTIIHI		
		GTCAAAT	RIQHFLDKNNFFQDTQFGFRKNHRTTDNVYIL		
		AAAATAT	KSLINKYTNKVGKHLYACFVDLRKAFDSVW		
		ACTATAC	REGLLFKLANAGVTGKIIHLIKSMYSNTKYAIK		
		TATACTA	TNQGYCKPFMSSSGVKQGCNLSPLLFNIIYINDL		
		TACTATA	SESLVQTEPVRLHDMSFNCLLFADDIVILSNSG		
		CTA	SGLQDSLNLFN DYCKKWKLTVNIKKTKVIIFN		
			KRGRKINDIFFLGDEQIEVVSSYTYLGITFQNSG		

						NFTVAVNNLVKKAQRALGVLKQTLNLIAS PINAIKLFRSSVMPILLYGSEIWA AFNFKPNITP ELFIQKLFSDSSKIESIHIKFCQKLLGVHKRASNI AVRAELGSYPLITEICTQMYKFWLRVKNKECS NIVHCAYQSECTINSNSWANFVSKLLEITRLNE TDHNANIFRNNLKRIYYEKIWLPTLSKKEHNK LRIYNKYKSDYEFEPYLNKINNLNHRVSVTKF RISAHKLPIEKLRYQKIPSEQRFC TLCS DQSVGD EMHALFDCKLDEIVKARSYFWNTLSLYNDNIL KLPKEDMFVHIMGIKDSFIANS LGKFINTILQC NT			
RTE	RTE - 1_S K		Saccoglossus kowalevskii	AACAG TGGCT ATGCA ATAAA CAGTG GCTAT GCAAT GAATA GTGGC TGTGC AATGA ACAGT	TGAATCC GGCCACC CTCGTCT TCGGGGA CCAAACC GGATTTA AAACTCT GGACTAC TACAT	RGEGSNSKKERC IQENAKSGTEENKVRYKRLK KQAKKVVA KAMREENE KVLSELKENPNKVY KIVKAMKGDGRDVTGGRCMRGSDGKLNFSER DRSKVWKEHMEKIMNEENEWDQDVEADLVE GPVERVSREEVVKAI RDMKTGKAAGPSEVSAE MIFASGDIGIGVIMELCQRVLDGRGMPEEWEV SVVVPIFKGKGDAKSCGAYRGVKLLEHAMKI VERVIVNRIRKLVEVDEM QFGFMPGKG TIDAV FILRRMQEEYRAKGKKLYMCFVDLEKAFDRV PRKVM EWAMRKRGIPEAMVRAVMSLYEGAK TRVRVGSELSEFEVKVGVHQGSVLSPLIFAVV IDVVTESAREGLMREMLYADDLVLMDDTMEG	cd0165 0	188	431

			GGCTA	LREKFVKWKKAFESKGLKVNLGKTKVMVSGT		
			TGCAA	EEKALVSKVDPCGMCGKRVKVNSVRCVKCG		
			TAAAC	KWIHGRCAKIKRVTQRLARDFACGRCKNGVG		
			AGTGG	RVVEQEEKLCDEVESVKEFTYLGDKVSVGGG		
			CTATG	CEAAVTARARFAWVKFRECGELLYGKRFPLE		
			CAATG	MKGIVYKSCVRSALLYGSETWCLSENELGILR		
			AATAG	RTERAMVRVMCGVKLMDRKITEDLMQMLGL		
			TGGCT	KEAVDRLAKANGVRWYGHVLRREDGHVLR		
			GTGCA	ALDLEVSGPRKRGRPKKTWKMQVEEESGKVG		
			ATGAA	MRMRDALNRAKWREGVRAIATSLR		
			CAGTG			
			GCTAT			
			GCAAT			
			AAACA			
			GTGGC			
			TATGC			
			AATGA			
			ACAGT			
			GGCTA			
			TGCAA			
			TAAAC			
			AGKGG			
			CAGAG			

				GAGAC ACAGG CTGGT GGAAT GA					
RTE X	RTE X- 10_S K		Saccoglossus kowalevskii	AAAGC ACAGA TACCA TATAA TTAAA AAACA ATATA CTAG	TAATTGT AATACGT ATTTTAT ATATACC CTTTGAT TTGTTTGT AATTTTG AAACTTT TGTTGTTT TTGTTTAT TTGTGTTT GTTTTTTC TGATCTG TATCCTT ACCATAC AGCACCC ACTGATG GTGCAGT	INKNKKNFFKNEMINKLETCDKTDQNEYWSLL KQLRCQQNSLNPDESTINSDEWLDHFSKLYNNN NNNADFKQSLKLEDTKLSNHELNNVISYNEI KKAIKNLKNKKSTAEDSISNEMLKTKNSLLIPC IQKLFNFILESGTYPTSWKSTILVPIHKSGDRSN PSNYRAIAISSCLSKLFTGIMNCRLTNHLEKNK LMNEVQFGFRKNRRTSDAIFLLKTAVEKTLYG KKETGAKCKRLYSCFVDLKKAFDSVSRDALFL KLQKTGINGNFYNIKSMYTDQEYKVKLNSGIT KGFHPTSGVKQGCNISPTLFNIFMNDLPGIFGE NVDKPIMIDTFPLNCIQYADDLAISESVDGLQ NCLNKLNDYCTKWYLVNVNRKTKVMIFNNS NRLIKNMNFKINNEEITSVKEFCYLGIFTPNG KFKINQSNLKNKAMKALFSIRHTILNHKMISTK LSLKLFDQLISPILLYGSEIWANEITNENNYLEQ ACNHFYRFILGVSKVTPLIGIRGEIGRIPIRIKAT LSLIKYWHRLSLLPENHILKSALKETVSSNSSW IKHVKTILGHTDLSEVNKIYKVGNI AELTKLLE	cd0165 0	150	418

					TGTGTAA TAAAGTT ATTATTA TTATT	EKLKNKYIQTWYQQLSHVPKNGKSGNKLRTY NLFKMEYEFEGYLTEIQDFNLRRMVTKFRLSD HNLEIELGRRKFPRMEANYRLCKKCNTGTVED EKHFILDCCFFTQERNTLLEDCKLEMCSTTNRN LIFTCIINSSTALAKYLKNTIYKT			
RTE X	RTE X- 6_S K		Saccogloss us kowalevsk ii	AAGTT AAATC TAGTG AGTAC TTATG GTGTA AAATT GATAG TAATT TAG	TAGCAAC TCCATCA AATTAGT CTGCAGG TAGTTTT AATGCAA TGGCATA TTTAGTA GCTTAGT AGTTTTA TTGTTAT ACTAATT TTCATGT ATTTTTTC CTCCATT CTTAGTT ACATGTA TTCAATG	IFIYVCLYPPENSVIHNRQDIDIFEKLSQDVIKYS NIGEVLLMGDFNARTGTLADHIINDNSTYMPIP DNYVDDEYNRNNKDQKVNSFGKKLLELNK NNNLLILNGRTPGDLLGKFTCYNFAGGASTVD YALVSKSFLSNIQYFRVNEINNLSDHCAISLAIT TRDLLSSDNNEIKCTPLPDRYLWDTKSIKRYQ DALHSTEIINKLVSFNYENNNIXNSVTELTNIIC NAADNSLKKIKHKISSSNKMKNTRQKWFDD DCWNLRRKKFREAAKLVEKYPKDCHIRGQFMV LRKKYNKILKIKRKQYKQNLWDSVTSQESNPK QYWKNFKIFSNDTSISANNDIIQADEWYNHF KNLNNLKNSKTCKYDSDFQNKIINETREMETN KEINKELNEPITIGEVKAAVKKLKLNKSSGEDK IINELIKYGQNYLMEPITAVFNNVLKSGTFPKG WKGSIIVPIHKKGSRSDPDNFRGVALSSCLGKL FTATLNNRLIKYLEKNNLYSPYQFGFRRNRRT TDCLFILKSIVDKYIHSEYSKNSKSKKLYACFV DYKKAFDSMWRDGLYWKLQKYGITGNIYNLI	cd0165 0	458	726

				TTTTATTC	KNMYSGVSYKVKTSKGLTNEFDSFMGVKQGC			
				GTGATTC	NISPTLFSLFINDLPEKIKSANGTPPELNSQKLN			
				TGGTGGA	CILFADDVLILSQTQEGMQCSIDALAKYSNSW			
				TTGAACT	KLSVSLNKTCLMVFNKRNHTLNKKSVMFYEGN			
				GTAATTC	EIATVDSFNYLGIVFTPNGKFKTASNTLKEKA			
				CGTCTTG	MKAMFPLKQSLTTL SHYKISTALNLFDKLIVPII			
				CTTTGTA	NYGCEVWGEESAKHSNIIESVQFAFCRFLLGVS			
				CTAAACA	KNSTKNAILGELGRYPLNVKIETQMLKYWLRL			
				ATAACAT	KSPVIPNIVATAFMEINTWKTGWVERIEEITKN			
				TAGCTTA	NNLSELSKQHEKKIGEVLRENYVKTWRRLEFN			
				TGTTTAA	DSRMNGATNKLTSYRKFKTTFNMENYLNDIK			
				TGTGCTA	XLKLRRFLCKFRISDHKLLIETGRQKKIPRNERI			
				TTTGTGA	CKLCTDDKIEDEYHFLIGCSKYDIPRNKLFKYI			
				TATTAGT	QGKKSDFLKMNEDEKFRLIMTSFAGIETAQFIH			
				TGTGTGA	TSMNLRDET			
				AGTATTA				
				TTTCTGTT				
				CTTGTTTT				
				ACAACCA				
				TAAAATT				
				ACTCTCT				
				GTAGTAG				
				ATTTTCG				

					TGTTGAA TAAAATT ATTATTA TTATTA				
RTE X	RTE X- 13_S K		Saccogloss us kowalevsk ii	AT	TAATTAA TATAACT TATTAAG ATACCTT TTCTGTTT TGTTTTTT CTAACCA TTTCTTGT TCTTCGTT TTTATTTG ATTTTATT ATGTTTT CACTWTG TCTTCATT TGTATTTT TCGTCTT GCCATTC TATATGT ATTTGAA	TDSQDINTNXKHESNILNKLSTDSKELNEPISQL EIQKAINNLKNGKSPGKDNISNEMIKAGRHTLII PVQKLFFNMILTSGLFPTNWNTGIITPVFKKGDK SDPSNYRGITLLNTLSKLFTAVLKQRLLIKYIDV NNLIVKEQIGFIPKFRTSDHIFVLKQILAKYLEK GKKVFACFIDFKKAFDSVWHQGLFHKLAQNGI TGKFYTIIRSMYSESLSCVKTKNGLTSEFPVKR GIRQGDGLSPLLFIYVNDIKNEIEQSNNDPPPP SLTVSQLSCMLYADDLLLLSVSESLQNALNS LGTYCKKWKLEINTTKSKIMIFNKMKGHIKKH KFCIDKQELENVPTYSYLGIKLSASGKFYKAQE DLANRAMKACFKLKQQLFSETNIPIKIYLMKF DSVIKPVLLYCCEIWSNENINIKNLSNPNLME KLHSKFCKFLLKVNCKASNIACLELGRYPLA FSSIVQTIKYWLRISDKPCNTLIYEAFLRGKDM DSKGIKNWVSHAKNIFNITNMDHMLDTVTAD KKQTIITSLESNLKSMFKNHFKETLYNDNKGK NRKNKLRTYRMIKHNFDFEKYLEDVKNPSHRS ALTKLRISAHTLQIEAGRYINQTVEERKCHYCQ	cd0165 0	89	353

					ACTTTAT TGTATTT KAAAAAT CCAGATC CTGAACA GGATATG ATTTCAA TAAAGGC ATTACTT ACT	EKQIEDESHFLLKCSLYADLRNKFLHTTKVITTI DKSDIENTVDILKTKDKQSLIHLAKYIFECFNL RNDTHKQKIPAAL			
RTE X	RTE X- 22_S K		Saccogloss us kowalevsk ii	A	TGAATTT AATGAAC TACACGA TTGTATT GTTGATA TATTATTT CAAGCTC TCTTTCAT AACCTTT ATTTGTG AATGTAA TCGTTAT TATTTAA	LYNXYLISIIDKGHRDILGVRFENKLTGFTFLVF SCYLPPENSPWGRDANAFFSHLLSQLYMTDDV DTIIICGDLNARLGNLNDNIDVVDNIRNRHVID TKINQHGHSFCEFLQESKLCILNGRFDRHQNDF TFISSRGRSVVDYVCVPQDHLSYCNNFNVITM SSLVEKYNLYDALTDRSKLPDHSLLIFDFLDRP LGSIVEQSLPNDINIGNNRHSATVRKYCLNQTP INFLESEVAREALLHIISKIEICRETQQEIDEIYD DLCKCILNEMNSTIPNYNCSSGARKRFRHHKP YWNDELYSMWRVMRERERSYTRFGGSKNIRT ILRNDFKAARRAFDKKLRCERNYRNDLVSSI EKVCTENPHQFWEHINKLGPVKRKDIPMEVYD DGGHITSNHNIVLNKWKHDFSSLYNQSDLDGF	cd0165 0	514	778

					GGTGACT TATAAGT CCACTGG GCTGGGC AGTTCTC TGAGCTG CAGGTCA CTATAAT AAATTAC TTACTTA CTTACT	DANFYEHILHHKLVLENDLDDPLYDSNSELNK NFSTVEIKTVVAKAKSGKAVGVDKIPYEVLKN DSSVAILTHLFQLCFDSAKVPCIWKKALICPILK VRSSDPRVPMNYRGISIQSTLSKLYSALLNNRI TRFLENNNILVDEQNGFRKDRNCEDHIYILTSII RNRISHDLSTFAAFIDLRKAFDSVDRDALLYKL LINGINGKMYKSIRAMYSDTNSSVKLNDFKTD WFNTEQGVRRQGDNLSPITLFSIFINDLAVEINLK QKGITLDNNVNVSILLYADDIVLLADNEIDMQ MMLSHLNSWCEKWKIRVNNVKSKMLHFRKA SEARSAFQFCIGDDVLDYTDKYRYLGIVLDEF VDYSVTAGIAADSASRALGALFVKFKTLKNLG FSTFTKLYDASVAPILDYCSGVWGFSEKCE MIQNRALRFFLGTHRFSPPLAVKGDGTGWVSCR DRRWAHMLVRLWKRLTELDDNRLVKRVFLWD HSLNYYNNWSSEIYQIMELVDLNYEHLIDVNIH ALRENIARRSESSWSESVSHKPKLRVYATFKF EYKPEKYVLINLSRGQRSLLAQLRMGILPLRIE TGRFRHEPVSERICKFCDANVVEDEMHFLFDC QQYNELRHEYLDLNEITGISREQILVRLFSGPI RKLALFIEKAFYKRRNLLYV			
Crack	Crack-		Salmo salar	CTCTC CCGAG	TGATGTA TTGTTGT	MTPACFVDYKLSLPNRGTDYVCSHTRDSDSL VTQTFGLPSYSNHLSPALYSCYLMKLLYNMIL	cd0165 0	627	884

3_S Sa		CTTTC	CTCTACC	LPSDAHSDVINQHCRALPVLCRALIVLLLMISG		
		GCTCG	TTCTTGC	NVHVHPGPSTVASPNSDLCSDICFTDFCSRKSL		
		CCTCC	CCTTTGT	GFLHVNTRSLLPKMDQLKVWVHSSNPDLVIT		
		AGCAC	GCTGTTG	ETWLRKSVLNTDVNLSGYNLFRQDRSSKGGG		
		ACACA	TCTGTGC	VAIFTKDHLQCSVVSTKSVPKQFDLLVLSIKLS		
		CACTG	CCAATAA	NSSLLTVAGCYRPPSAPACTLPALSSLLAPYTK		
		TTGGG	TGTTTGT	SEFVLLGDLNWDMLKPPDQVLKQWDSLNLISQ		
		GCGAG	ACCATGT	IITNPTRYDSKHPEKATLLDVILTNNPDYQSG		
		TGACT	TTTGTGC	VFCNDLSDHCFACVRNGCSVKRPVLICHRRL		
		CTGAA	TGCTACC	LKNFNEQAFLHELASVKWYRISLIPSVEDAWT		
		CTTAC	ATGTTGT	FFFDIFSGIVNKHAPIKKMRIKNRFSWFDRDL		
		AATGG	TGTTATG	AELLHLKNCIWRKARHTHTQADWLSFRQMRN		
		CTAGC	TTGTGTT	KCTQAIRKAKVSYFKEQFSLCGSNPKKFWKTV		
		CCCAA	GCTACCA	KDLENKPSSSQLPMSLVNDDVVVTDKKHMAE		
		GTCGT	TGCTGTG	LFNHFFIKSGFLFDSAMPPCPSNISSSPTPSNATI		
		TATAT	TTGTCAT	PDASPSFSPAPLQSFSLQAVTESEVLKELLKLDP		
		ATTTA	GTGTTGC	KKTSGSDGLDPFFFKVAAPIIAKPISNLFNLSLL		
		ATTTTT	TGCCTTG	SGEVPIAWKAATVRPLFKGGDQADPNCYRPISI		
		TCTTGT	CTATGTG	LPCLSKVLEKL VNNQLTGFLDVYSILSGMQSG		
		GCATT	TTGTCTT	FRSGYGCVTATLKV LNDVTIALDSKQCCAAIFI		
		TTGCT	AGGTCTC	DLAKAFD TVDHSILVGRLRSIGVSEGLAWFA		
		ATTTG	TTATGTG	NYLSQRVQCIKSENLLSQPLPVTKGVPQGSILG		
		CCTC	TGTTGCT	PTLFSIYINNIAQAVGSSLIHLYADDTVLYSAGP		

				CTCTTGTT	SLDFVLNALQQSFLSVQQAFSTLNLVLNTSKT			
				GTGATGT	KVMWFGKKNAPLPTGVITTSEGLELEVVTSYK			
				GTGTTTT	YLGWLDGALSFSQHISKLQAKVKSRLGFLYR			
				GTCCTAT	NRSSFTPAAKLTLIQMTILPMLDYGDIIYRSAG			
				ATTTATA	KGALERLDVLYHSAIRFATNAPYRTHHCTLYS			
				TTGTATT	SVNWSSLYTRRKTHWLMLIYKTLLGLTPPYLR			
				CATTTTTT	YLLQPSSSTYNTRSASHILLKVPKAHTSLGRSS			
				AATCCCA	FQFAAASDWNELQQTLKLDSFISISSFKDSIMD			
				GCCCCCG	TLTDSCGCFV			
				TCCCGCA				
				GGAGGCC				
				TTTTGCCT				
				TTTGGTA				
				GGCCGTC				
				ATTGTAA				
				ATAAGAC				
				TTTGTTCT				
				TAACTGA				
				CTTGCCT				
				AGTTAAA				
				TAAAGGT				
				TAAATAA				
				A				

Tx1	Tx1- 2B_ SSa	Salmo salar	GT	TAATAAC AAGATAA ACAATGT AAAACAT ACGGGGT CTGTAAA ATGTATA TAGGTTC AGAACTT TTGTGAA ATAGCAC AGTTACA AATAGAA ATCAAAC TGGATGG ACATCAG AAATAGA GGAAGGA CTAAAAA CAAACAA AAAATAA CTATTGT AAAATAG	QRNSKGVMVLINSNFDPNVQIVQTDHQGRWII LNMLLDNKQIWLINLYGPNNDPSFFENIYKN LSTLQATLDSIIMVGDFNTVLNTSMDRKGNHT TNYHPQALKEIMNVMDILELVDIWRLKYPDLV RYTWRRNLNQASRLDYFLMPFSLAPKVKVLIGD RMRSDDHHIIGIYITLTEFPRGRGYWKFNQSLLD DKLFRTRTEEFITDFFRHNIGTADPLIVWDTFK CAFRGHAIQYSSIKQKQFRSKESILTKEIEGLTV QLDSNKNGTIEAQNKLEEKQKEMEELIQERSS VIYYKNKANWMEYGEKCTKFFFNLQYRNATK KMYKCYQKNVLKLVTDNGVTHDSPNXILKEE VKYFKNMFSFQAPPSPLTETNCMDFFPNNNVK LTSVQKDSCEGQITEEELDAIGAFKDGKTPGL DGIPVEVYKTFFDILKGPLLACFNHSYINGRLS DTQQEGLISLLLKQDPSGIYKDPVHLKNWRPL TLQCCDAKILAKCLAHRIKKVLSDIHPNQTFG LHGRYIGDNIRQVLETIEYYEISGTPGLVFIADF EKAFDKVRLEFIYKCLEYFNFGESLIKWVKVM YSNPRCKIVNNGYISESFKLSRGVKQGCPLSAY LFIIAIEMLAVKIRSNNNIKGLEIRGLKTKVSLY ADDSCFLLKPQLES LHGLIEDLDTFAILSGLKP NYDKCTILRIGSQKNANFTLPCSLPIKWSDGDV DILGIQIPKERNDLTPIHFYRKLAKIDKILLPWK	cd0165 0	457	723
-----	--------------------	----------------	----	---	--	-------------	-----	-----

				ACTGTGT	GKYLSICGKITLINSLVISQFTYLLMVLPTPSDL			
				CTGTAAA	LFKLYEQKIFNFIWNGKPKIKRAYLYNEYEF			
				ATGTGTA	GGQKLNIALDLSLKASVIPKLYLNPWFSSK			
				TAAGATG	LVRMSHPMFKKGLFPFIQITPAHFRLFEKEIISKI			
				AATAAAC	SLFFKQALESWLQFQFNPPEKTEQIIQQILWLN			
				TGAAGGT	SNIVIDKKKIFEEMFKKGILVNDIINRTGGVMS			
				AAAAGCC	HMQLTQTYGNVCSTQNYNQLIAALPQKWKRQ			
				GAAGTGT	VEGEKSKELVCRPCIKHEHKWLKKSVINKNYQ			
				TTATTAG	FHLRTKKLTAVPYKLQNSWEEIFDVPIPHMV			
				TTTACTC	YELIRKTTTPDSKLQIFQFKLLYKILATNRMLYI			
				CAATTGG	WGIQSSQLCRFCCEEAEGLDHLFWYCPYVARF			
				GGGATCG	WSQVQEWLKNCNICLELTLQIAILGDLKSHSQ			
				GTGGTAG	SINNIIILAKMFIFNLQSVEAMRIERFNCFKHH			
				GGTTTGC	STVEKYMARNRNPKWMMMLRDRWEGLNGAEG			
				GGGGAAT	WD			
				AATAATA				
				AAGGTAT				
				ATTCTTT				
				AAAAAAG				
				TATGTAT				
				GTCTATA				
				TAGGTAT				
				GTGTATG				

					TATATAT GTGTATA TGTATGC ATGCGTG TATGGAT ATATATA TTTACCC CAAAAAT ATATGGG GGATTGG AAATGAT GCAGACA ATTACAT TGATGGA AGCAATA				
Tx1	Tx1- 2_S Sa		Salmo salar	CTGTA TCCGG TACAA AGACA ACAAA GGAAG GATGG TAGGC	TGAGAGA TAGAATG ATGGTCA AAGATAA AAGTAAA AATTTAA AAAAGTA AAAAAAA	MVRGEISIASYN CNGLADYKKRRSVFTWLKEK EYNIYCLQETHSTSLDEVAWKKEWGGEIIFCH GQRNSKGVMILINKNFDLNVQIIRNDSQGRWIL LNMKVDEKEIWLINLYGPNQDDPHFFENIYTN LLNLQATNDLIIMVGDYNTVLSTSM DRKGNHS TNYHHRALKEITNIMDTLEIVDIWRLKNPDLV RYTWRRLNQASRLDYFLVSFSLASKVKKVLIG DRMRSDHHLIGHITLIDFPRGRGYWKFNQSLL	cd0165 0	517	783

			GTGCC	CATAATA	EDNLFLTKTKFITEFFQYNIGSANPLIVWDTFK			
			TTTATC	AAAAGTA	CTFRGHSIQYSSIIKKQFLAKETRLTREIHELIVQ			
			ATTGT	CATTTGA	VDSNKNDTTEIQNKLEEKQKELEELIQERSNVI			
			GGACA	ATGACAC	YYKNKANWMEYGEKCTKFFLNQYRNANKK			
			AACTC	TAAGTGG	NLQKLVTEDGVIYDSPNXILKEEANYFRQMFS			
			TTTAT	CAGTGTT	FPSHPLPLNEDYGKEFFPNNIKNGKLTNVQKD			
			AGATG	TTTACAA	QCEGQITEEELFEAIKSFQSGKTPGLDGIPVEVY			
			GACAG	CTAATGC	QVFFDILKAPLLDCFNYSYRNGSLSGTQQEGLI			
			CTATT	CGGTTTG	SLLLKQDPDGKYKDPVYLKNWRPLTLQCCDA			
			CCGAG	CCTGAGG	KILAKCIALRIKRVLPGIVHPDQTGFLHGRYIG			
			ACAGC	CTGATGC	DNIRQLLEIIHHETYKKPGMVFIADFEKAFDK			
			TCCAT	CGTGCCG	VRLDFIYKCLDFFNFGNSLIKWVKIMYSNPRC			
			AACAC	GTGTTTG	KIVNNGYFSESEFELSRGVKQGCPLSPYLFVMAI			
			CGTGG	TACACGT	EMLAIKIRSNNNIRGLEIQGLKTKVSMYADDSS			
			CTGTA	GCATATA	FMLSPQARSLQCLIKDLDNFSVLSGLKPNYDK			
			CTAAA	CACACAC	CTILRIGSLKNTTFTLPCSLPIKWADGEVDILGI			
			TTCTA	TCTCATT	HITKDINKLSTMNFNRKLVKIDKILQPWRGKY			
			CAGGG	CAAATAA	LSIYGKIALINSLVISQFTHLLMALPTPDDSSFFK			
			ACTTC	ACAC	SYEQKIFRFIWDAPDKIKHAYLYNEYELGGL			
			AGGTT		RLLNIKALNLSLKASLIQKFYLNPKWFSSRLIR			
			ACGAA		KAHPLFKNGLFAFVQIAMSHFRLIENYTFFKVS			
			AAAAA		VFFKQALQSWLQFQFHPPEKIEQILWLNSNVL			
			GAAGG		VDKIPVFMGKMFEEKGILFLNDIVNWNGRVMSE			

				TATGG	MELSELYGKVCSIQEYNQLITALPQKWRRRVA			
				GAACT	AGGGRELVCLPNIKDQNWRRNKN SINRKVYQ			
				GTAAA	FHLRTRMLTTVPYRLQNSWEEIFDVPIPWYRV			
				AATAT	YELIYKTTQDSRLRAFQLKLLYRILATNKMLNI			
				CTGAA	WGIKSSKLCRFCCEDTESIDHLFWYCPQVACF			
				CATTA	WSQVQEWLKMHSIDLKLTLEIVLLGDLERPGQ			
				TGAAG	SITNILILLVKVFIFNSQSVDSIRLDRLKLYVKH			
				TGGAT	HSIVERYVLRRNPKWVASRDRWDGLREAEGW			
				ATGAA	YVELETSGSGVAV			
				CACAC				
				ACGTA				
				CTCAA				
				TTACA				
				TGCTT				
				GCTTA				
				CACAT				
				ACGGA				
				CTTAT				
				ACATA				
				CACAC				
				ACACA				
				CCTGA				
				TCTCG				

				CTCTCT					
				TCTCTC					
				ACTCT					
				CTCTTT					
				CTCTCT					
				TCTCTC					
				TCTCTT					
				CTCTC					
				CCTCT					
				CTCTCT					
				ATTGT					
				CGAGA					
				ACTGT					
				TTTAT					
				AATTT					
				AATGT					
				GTTTA					
				TTGTTT					
				GTCTA					
				TCTTTT					
				TTATG					
				TATTT					
				GTCTA					

				CAAGT					
				TTATA					
				TATGT					
				TTACA					
				ACAAG					
				CAAGG					
				GGAAG					
				ATATC					
				AGAAT					
				AGGGG					
				CATTG					
				GGGAA					
				TAATA					
				ACATA					
				TTGTA					
				CGGAA					
				TACAT					
				TTGTA					
				CTGTA					
				GTGAA					
				GCCGT					
				CTCTA					
				GAGTA					

				ATGCA					
				AGGTC					
				TCTTTC					
				ATGAT					
				CATGT					
				GAAAC					
				GTCAA					
				TTGCT					
				ATGGA					
				AATGC					
				TGGGA					
				TGTGT					
				TTTCA					
				ATTAT					
				GTTAA					
				AGGTA					
				ATTAT					
				GATGC					
				AACTA					
				TGATG					
				GTGAT					
				ACGAT					
				ATGGA					

				TTACA					
				ATTAT					
				CATAT					
				GAATA					
				TTAAG					
				GCTAT					
				ACGCA					
				CTACA					
				TGTTA					
				CACAC					
				ATAGA					
				CACTC					
				AAACA					
				TGCAA					
				ATTGG					
				CAGAG					
				TTATT					
				ATTTA					
				TTTGG					
				ACATC					
				ACACA					
				TTATA					
				GTCTT					

				ACTCT					
				TATAC					
				AGACA					
				TCATA					
				CACAC					
				TCTCA					
				TTATA					
				TCATA					
				TCCAA					
				TATCA					
				TACAC					
				ATCAA					
				CCTAC					
				TCAGA					
				TTTGCT					
				ACACA					
				CATAA					
				TATCTT					
				GTCTC					
				AATTT					
				TCTAA					
				CATCT					
				GTGGC					

				ATTGG					
				CTCAC					
				AGGCA					
				AACAG					
				GCAAG					
				GGAAT					
				AAAAC					
				AAATA					
				TTGCT					
				AGAAC					
				CTATTT					
				CTTGA					
				ATTCT					
				AAATA					
				TCAGA					
				CATTT					
				GAAAG					
				CTCTA					
				GTTTT					
				GACCT					
				TCATA					
				ATACC					
				TCTGA					

				TGGCA GTAAC TTTAC AAGCA CTAAT TATGG TCAAT TTAGA CTGAG AATAG TATCT AGTT					
L1	L1- 5_S Sa		Salmo salar	TAAAA AAAAT GTATT TTTTAT TATTA TTATTT TTTTTG ATTGG GGGGT ACGTT CAATT	TGAATTT GCACTTT TTAACGC ACTCTCA ATTGTMA TATGGTA ATCTACC TTTGGGC AGCATGT GCTGGCC CCGCCAC	WFGALSYLFSDLHAMQATIGTTNERRAGLTFT SWNVKGLNEPIKRGKVL AHLKALSSDIIFLQET HLKNNSHNRLKCKWVGQVYHSNFS AKTRGT AILVRKGIPFLHKTTIVDKEGRYVIVIGEIHSTS VTLLNIYGPNIDNPSFFKRV LALIPDISHTNLVI GGDLNCVLDQYLD RSSTRRTPTSYSSEFLNTYI KNSNLFDIWRISNPTGREYSFYSHVHN VYTRID YFLVDAKLLPYTCNVRYHDIIISDHSPLTFSLRL GDIVPNERVWRLNPQLL TEPTFCEHLKDQITFF FDTNDNTETSPALLWETLKAYLRGCIISFQTAR SRQNRGELVELEGQIHLLDRENASHPSMEKHK	cd0165 0	535	793

			TAKCA	CCGAGCA	KITSLKFEYNQILSAKIAKSFLYAKQKYFEFGD			
			AAATC	GTTGGGA	KPHKLLARQLRKNVSDRMIHKVKSASGELLSS			
			CTTGTT	GGGGAAT	PKDINDRFRQFYETLYTSKADPNPLIMQNFLED			
			CCSAT	AGGGTGG	CNLPALNQEDSNFLNKEISLEEIRETIKSLKSGK			
			CTCCT	AATAAGG	TPGPDGYPGEFYKTFSNMLSPYLHKMLVQAN			
			AACCC	GGGGGAT	EDGALPSTLDEAFITVIHKKGKDPEEVGSYRPIS			
			ACAAT	AAAGGGG	LLNTDQKILAKTLANRLSTLIGKLVHSDQTGFI			
			ATGTA	GGTATAA	PNRNSFFNLRRLFNIMYSQRLPNVDLAVISLDA			
			ATTGT	GTGGGGG	EKAQDQVEWPYLFKVLQKCNIGDGFIKWIQLL			
			ACTGC	GGTTACA	YRNPCARILTNQSLSPRFNLYRGTRQGCALSP			
			TGTCT	TCGACCC	MLFALHIEPLAQAIRSHAAIHGYNTKDTL NKISL			
			ATGTC	TTTTTTCT	YADDILLYVTEPQDSIPAILXVINLFGTFSGYRI			
			GGTTG	TTCTACC	NWNKSELMPIRLQNTSWLEHLPVKLSSEKFTY			
			CTGAT	TGTCCTT	LGIVVTKQYSLLFKENFPSLMQKLKANIQFWR			
			GGTTT	GTTCTGT	TLPISSLGRINAIKMVFLPQLLYLYQNIPVFIPKS			
			ATGTT	ATGTCTT	FHKQLDSIINPFIWDYKTHRIGKKHLCKSKMEG			
			TAGAC	GTTTTGT	GLSLPNFIFYWAANLRAVTFLDDAPPPSSW			
			CGGCA	AATATTT	LSMEREECHPFSIGAVILSPVNLEMSLYRNNPII			
			TTGCTT	GTTTTGT	HSTVRIWKQIKVHFELRPISFMLPVARNPSFAP			
			AGCAA	GCCCAGA	SNLDNTFERWGELGISTIGDLYIGGTFASFESLR			
			TGTAA	ACTCTGG	ETYNLPRSNNFFRYLQIRDYVRKHLPTFGNAKPS			
			TCTTGT	GTCTTTT	MFDGCIKICPTSDKLISRIYDAFQSVSAPSTDAI			
			GATGC	GGTTAAT	KAKWEEELGTDISVADWEDSLEYIHTCSINSR			

			TATAT	GTCTGTT	HRLIQFKVLHRLHYSKTKLHRIFPGTSPMCDKC			
			CTTGC	CTCTTAT	QAEQGTLLHCFALCSNLYGYWCGIFGVLSEVL			
			TTATCT	GTTTAGA	ETSIDPDPLLIILGVSDSLNGLTNPQKQLISYSLI			
			CAGGG	TAAGAAC	SAKKLILLFWKKREAPSTKLWLSELANTVHLE			
			ATTGA	TGTAAAC	RIRYILNNKLSTFDQIWQPFLSYLDQAAL			
			CCCAA	TTGTCTTT				
			GATGA	TATACCT				
			MGCTT	ATACACC				
			AAGTG	ATTCTTG				
			CGCAA	TGTATGT				
			TATGT	GTTCAAT				
			TTAAG	AAAAAAT				
			TTGCA	ATTTGAA				
			ACTTA	CAAAAAA				
			TTCTGT					
			TTAGG					
			TTTAA					
			TAGAT					
			GCTAA					
			TTGAA					
			CACTT					
			TAATC					
			GCGGG					

				AGCCT					
				CCTCA					
				GTTCA					
				GATGT					
				TAGTA					
				GAAGT					
				TCTAG					
				GATAG					
				TGAGA					
				GGTGA					
				ACGTA					
				CAGTT					
				GGGAT					
				TTTATT					
				TTTGTT					
				TTACC					
				TCTTGT					
				TCGAG					
				GTCGC					
				GCCGT					
				GCTTTT					
				TTGGC					
				ATCGG					

				CCAGA				
				CAATG				
				TGTTT				
				ATTAT				
				TTTTAT				
				TTTCAT				
				GTTTTT				
				TTCTCT				
				CCTAT				
				TTGTG				
				TATGC				
				CTGTT				
				AAATG				
				TGGGT				
				TGATG				
				GGGGG				
				GGTAA				
				ATGTT				
				GGGGA				
				AAGTA				
				GGGGG				
				ACAGG				
				GTGGA				

				GAGTA AAGGT GCTTG CAGGG GGGGA GGGGT GGGTT GGATA CTGCT CGGGT GTAGG TGCTA CATAT GCTGA TCGTG A					
Crack	Crack-10_Sa		Salmo salar	ATGCT TTTTTT TGCTA TCGTA CCTAA CATTTT AACGG	TAGCTGT GTATTGC CGCTTTC CATGTTG TCTGTTG TCTGTAG CTTGTGA	MSLRXXXYFAMIVSYVAALLIISVRPCWDDTI NLLPENIYTKHIGSPIHRSGLYTAYHNVLNNGQ VYLAPYSTEPLRRDKHKSNIYRKLLNYLFTTLL LSGDVQLNPGPNITEPAILTGVESRGWPRPLIV AAVEVGECYXPMVSLDSPGSRIDFDSSNIPKPL YAIPDFASGTQAVLISSPHPVQAGGIVNSATEL PLIKTKQNGLNPAVRKHKRCNFFQCVNHSRVI	cd01650	727	983

			TTTAC	GGTGTGG	WDPRAKPKGLLGHLNIRSVIPKSDQIQHLLTD		
			AMCGA	AAACACT	SNLDFLCLSETWLHKNSPCAALIVPGYNVFRR		
			CAGGA	TTGTTGTT	DRIEGRGGGLMIYIKEHIRCKQIEWSCDNELEC		
			GCTTT	TTTATGG	IGLNVTLSPQMSFTLIGMYRPPSTKSVFFDQFN		
			GTATT	ATTTTGT	NMLRECDFGKEVILMGDFNINYEDKSCRKTLK		
			TCATT	CTTGTTG	RITNTFDLTQLVKGPTRVTCCKSTQIDLVFSNK		
			CATTTT	CTTTTTGT	PERVTKSFNMVTGLSDHNLTLIARKLSKSRFKL		
			TTACS	TCTATGT	STVRKPDQLRIPKSELNYFENAINGINWNDLLS		
			AGGAA	TGCTCTG	YTDVEADSEVFLSTIQATINGFLKKIKSKPGQK		
			CG	TCTGTAT	STLPWLNGEIWKLMKERDYALKTALKSKLEH		
				GCTACAT	DRRRFTMLRNKVMKEIRQAKANFFINIIGEAK		
				CTTGCTT	GNSKLIWENLKKLTGKDHSNTAKRLEIMVNN		
				GTCCTAT	NLTQDAVEIAIAFNSYFIDSVRVLTQNPSTGLL		
				GTTGCTC	GSVLVNDAAQPVFIIREVSESKVNKVISSLKNSK		
				TGTCTGT	SKDVFGLDSTFLKNYKESLIGPITKVTNTSIGLG		
				ATGCTAT	VFPRVWKSAIITAIFKSGDPADVSNYRPISILPV		
				GTCTTGC	VSKVVEKCVAEQLISHLNNSPFTLHSMQFGFR		
				TTGTCCT	AKHSTETANCFLLENVKSKMDKGGVVGAVFL		
				ATGTTAC	DLRKAFTVTNHEILITKLSKFNFSPDALRWMK		
				TCTGCGT	SYLEGRTQCVRVSNELSPTLSYDVGVPPQGSILG		
				GTGCTCA	PLLFSLYINDLPSVCTGSEVQMYADDTVIYVH		
				CTGTTCA	AKSKQQAQELTTVMVQVTKWLSDSLHLN		
				ATGATTG	VKKTVCMMFTKRATDATEPDVYVSGEKLQVV		

					TCTATAT TGTAATT GTTTTTA ATAACCT GCCCAGG GACTGCG GTTGAAA ATTAGCC GGCTGGC TAAAACC GGCACTT TACTGA AACGTTG ATTAATG TGCACTG TCCCTGT AAAAATA AAATAAA CTCAAAC TCAA	SDFKYLGIILDSNLSFKKHVKKVIQITKFNLNF RFIRNCLTTEVAKLYFKSMILPHLTYCLTSWAQ ACYTTLKPIQSVYKQALKVLDRKPNSHHHCYI LRKHELLSWENLVQYTDACL VFKILNGLAPPP LSIFVKQKTQTYGSRSTRSAMRGDCIVPLRKST FSKSAFSVRASHVWNTLPSDTHNCTTYHTFTK CMKTWLKVNQTCEHNP			
Rex 1	Rex 1-		Salmo salar	AGATG GCGCC GACAG	TAGCCGG CTACCAC CCGGTTA	LLDIRAAVTHQHYDQEYDFPELDPLFVPPRAIE LIPEAAPRRRRRRRGIRSGLLVRLRRRAHHPPL PSILLANVQSLDNKVDELRRARISFQRDIRDCNIL	cd0165 0	539	848

4_S Sa		ACATG	CTCAACC	CFTESWLSPDILSPSIQPAGFSVHRADRNKELS		
		GCAGC	CTGCACC	GKKKGGGVCFMINYSWCDCDNVQKLKSFCSP		
		TCTGC	TTAGAGG	DLEYLTIKCRPYLPPREFSSVIVTAVYIPPQADT		
		TTCTA	CTGCTGC	TTALKELHWTLCLETTYPEAAAFIVAGDFNKA		
		GCTCC	CCTATAT	NLRKTLPKFYQHIDCSTRAAKTLDHCYSPFRD		
		TAAGC	ACATAGA	AYKALPRPPFGKSDHDSILLLLPSYRQKLKQEV		
		AACTT	CTGGAAT	VLRSIQRWSDQSESMQLQDCFDHADWDMFRVA		
		TGCAG	CACTGGC	SENNIDVYTDVTEFIRKCIQDVVPTVTIKTYP		
		TATTT	CACTTTA	NQKPWIDGSIRAKLKARTTAFNHGKVTGNMV		
		GTTTTT	ATAATGG	EYKQCSYSLRKAQAKRQYRDKVESQFNGS		
		TTGTG	AACACTA	DTRRMWQGLQTITDYKGKTSHVADTDVLLPD		
		TGTTA	GTCACTT	KLNTFFARFEDNTVPPTRPAPKDCGLSFSVAD		
		TTTCTT	TAATAAT	VSKTFKRVNPRKAAGPDGIPSRVLRACADQLA		
		ACATT	GTTTACA	GVFTDIFNLSLSQSAVPTCFKMSTIVPVPKKAK		
		ATTAG	TACTGCT	VTELNDYRPVALTSVIMKCFERLVKDHITSTLP		
		CCCAG	TTACTCA	DTLDPLQFAYRPNRSTDDAIAIALHTALSHLDK		
		AACGT	TCTCATA	RNTYVRMLFIDYSSAFNTIVPSKLIKL GALGLN		
		TTTTTG	TGTATAT	PALCNWVLDLFTGRPQVVKGNGNTSTSLILNT		
		TGTTA	ACTGTAT	GAPQGCVLSPLLYSLFTHDCVATHASNSIIKFA		
		TTACA	TCTATTCT	DDTTVVGLITNNDDETAYREEVRALAEWCQEN		
		TACAG	ACTGTAT	NLSLQRQQNEGADRGLQETAEGARPYPHRRD		
		CCGGA	TTTAGTC	RSGEGGKLQVPRRTHHGPWRSGARKITSPSNV		
		AATAA	AATGCCA	NKTKELIVDFRKQQREHAPIHIDGTAVEKVESF		

				CTCCGAC	KFLGVHITDDLKWSTRHTDSVVKKAQQRLFNL			
				ATTGCTC	RRLKKFGLAPKTLTNFYRCTIESILSGCITAWY			
				GTCCTAA	GNCTARNRRALQRVVRSRQAQRITGGTLPALQDT			
				TATTTAT	YSTRCHRKAKKIIKDINHPSHGLFTPLSSRRRG			
				ATATTTC	QYRCIKAGTERLKNSFYLKAIRLLNSHH			
				TTAATTC				
				CATTCTTT				
				TACTTTT				
				AGATTTG				
				TGTGTAT				
				TGTTGTG				
				AATTGTT				
				AGATACT				
				ACTGCAC				
				TGTTGGA				
				GCTAGGA				
				ACACAAG				
				CATTTCG				
				CTACACC				
				CGCAATA				
				ACATCTG				
				CTAAATA				
				TGTGTAT				

					GTGACCA ATAAAAT TTGATTT GATTTGA G				
Rex 1	Rex 1- 3B_ SSa		Salmo salar	GCGCC GGAGG AGATG GCTGC CGTTTT ACGGK CTCCT AACCA ATTGT GCTAT TATGT GGGGT TTTTTT CGCGT TATTT GTAAC TTATTT	TAGTCAC TTTAACC ATACCTA CATGTAC ATACTAC CTCAATC AGCCTGA CTAACCG GTGTCTG TATATAG CCTCGCT ACTCTTA TTTTCAA TGTCTTTT TACTGTT GTTTTATT TCTTTACT TACCTAT	CFCHRILRQKRASGYQDS DHSPRIRQRFFLQQQ QAGLTGYSPNTXQGRHPRYSQEEATTRGRRA GCLVRTRRRRVGKLPLPSVLLANVQSLDNKLD EVRSRISYQRDIKNCNILCFTESWLNDMDIQL AGYTLHRQDRTAHSGKTRGGGLCIFVNNSWC TKSKEVSRFCSPEVEYIVINCRPHYLPREFSAIL FVAVYLPPTDAGTKTALSQLYKDISKQENAH PEAALLVAGDFNAGKLKSVLPNFYQHVKCAT RGKKILDHLYSTHRDAYKALPRPPFGKSDHNSI LLIPAYKQKLKQEAPVTRSIRKWSDEADAKLQ DCFAXTDWNMFRDSSDGIEEYTTSTVTGFINKCI EDVVPTVTVRTYPNQKPWITGNIRTELKGRAA AFKERDSNPEAYKKSRYALRRTIKQAKRQYRA KIESYYTGSDARRMWQGLQTITDYKGKHSREL PSDTSPLDELNHFYARFEASNTEACMRAPAVP DDCVITLSVADVSKTFKQVNIHKAAGPDGLPG RVLACADQLAGVFTDIFNMSLTESVIPTCFKQ TTIVPVPKNTKATCLNDYRPVALTSVAMKCFE	cd0165 0	549	801

				TGTAC ATAA	TATTACA CACATAC TTACCTTT TTTTCCG CACTATT GGTTAGA GCCTGTA AGTAAGC ATTTAC TGTAAGG TCTACAC CTGTTGT ATTCGGC GCACGTG ACAAATA AACTTTG ATTTGAT T	RLVMAHINTIIPETLDPLQFAYRPNRSTDDAISI ALHTALSHLDKRNTYVRMLFIDYSSAFNTIVPS KLITKLRTLGLNTSLCNWILDFLTGRPQVVRV GSNTSATLILNTGAPQGCVLSPLLYSLFTHDCM ARHDSNTIIKFADDTTVVGLITDNDDETAYREEV RDLAVWCQDNNLSLNVTKTKEMIVDYRKRRRA EHAPILIDGAVVEQVESFKFLGVHITNKLSWSK HTKTVVKRARQSLFPLRRLKRFGMGPEILKRF YSCTIESILTGCITAWYGNCASDRKALQRVVR TAQYITGAKLPAIQDLYTRRCQRKALKIVKDPS HPSHRLFSLLPHGKRYRSAKSRTKRLNSFYYPQ AIRLLNRDQKAPQQLPPSHKTPEQVIKWLPGL FALCAPPPSSTLCLSYMHSFNYTFMYILPQLA RPTSAPAHWLTGLSALCPTTRQPLFLRCCYSLF IIYA			
Crack	Crack- 2B_ SSa		Salmo salar	GCCCA CAGGA CTGGA AAACC CACMA	TGACCTA GATAGTT TGTGTGT ATGCATT GATATGT	MFFSLALFALFHIMSISGKLPRKGLKIAHINICS LRNKVHEINNLLTSDNIHILAISETHLDNSFDDT SVAIQGYNIYRRDRNAYGGGVAVYIQSHIPVM LREDLMSSVIEVLWLQVHLAHLKPFLLGCCYR PPSANSQYLNMCMLDSVCDVNREYVFLGD	cd0165 0	557	816

			CCKGC	AGGCTAC	LNIDWFSSSCPLKRKLLTVTSACNLVQVINQPT			
			CCAGG	GTGTGCC	RVFTNTTGTRSSTCIDHIFTNAVELCSKAVSVPI			
			TGATA	TTTTTAA	GCS DHNIVAISRKAKVPTAGPKIVYKRSYKRFC			
			GGCCC	AAATGTA	CDSYVDDVKNICWSDVINEEHPDAALDEFMK			
			AGGCC	TGTAGTT	LLLPIIDKHAPVKKLTVRTVKAPWIDEELKNC			
			GAGGT	CGTCGAG	MVERDGAKGVANKSGCTSDWLTYCKLRNYV			
			WGCTK	CTGTTCTT	TKLNKKKKKLYYEAKINDIKNDGKKTLEYFK			
			TTCTG	GTCTAWT	XNYGQKDKFNSIFHRIRWLIHHKTIRMMEKKL			
			GAWAG	GATGTTT	WSTLNXIMGRKTNSTPSFIESDGLFITKPFDA			
			AGCCA	TGTATTA	NYFNDYFIGKVGKLRQEMPTTNSEQLYSCIKK			
			AGAAC	TGTCATT	QIMKEKQCKFEFCKVSVGEVEKLLLSINNDKP			
			TTGWG	CTGTATT	PGIDNLDGKLLRMVADSIATPICHIFNLSLEESL			
			AGGAA	ATGTTTC	CPQAWREAKVIPLPKSGKAAFTGSNSRPISLLP			
			CGTAT	ATGTTTT	ALSKLLEKIVFDQIQCYFSVNKLTTDFQHAYRE			
			ATCTT	GTGTGGA	GHSTCTALTQMTDDWLKEIDNKKIVGAVLLD			
			CCTCA	CCCCAGG	FSAAFDDIHDHNNLLKKRMICYGFSTSAISWIQSY			
			ACGAG	AAGAGTA	LSNRTQRVFFNGSFSNVKHKVCGVPQGSSSLGP			
			GACTG	GCTGCTG	LLFSIFTNDLPLALNKACVSMYADDSTIYASAT			
			TCCTG	CTTTTGC	TANEVTETLNKELQSVLEWVASNKLVLNISK			
			AAGCT	AACAGCT	KSIVFGTNHSLSSRPQLNLMNGVAVEQVEET			
			GTGCG	AATGGGG	KLLGVTLDCKLSWSKHIDSMVVKMGRGLAVI			
			CCAGA	ATCCTAA	KRCSAFLTPHKKQVLQALVLSNLDYCPVW			
			AGAGG	TAAAATA	SSAARKDLVKLQLAQNRAARFALNCNQRADI			

			AAAGA	CCAAATA	NTMHASLSWLRVEERLTASLLLFIRNINVLKIP			
			GCTGA	CCAAA	NCLHSQLTHSSDHTYPTRHATRGLFTVPKSR			
			TCCCA		TNSRKRTVLYRALIAWNFLPSHIAQINSKPGFK			
			GCCAT		KQIKQHLAAQRLSPI			
			GAAAG					
			CTGCC					
			AGAGT					
			GCGTG					
			GGGAC					
			ATTGC					
			KTACA					
			TCCGC					
			TACGA					
			CAGGC					
			TCATT					
			GTCCA					
			CCCTC					
			CCTCC					
			CAGAA					
			GCCTG					
			GAAGG					
			GATGA					
			GAGAG					

				CCAAG CCTAT GGGTT CGTAG CCTCA ACCCC ACACA CACAC ACACA CACAC ACCAA TTGAT TAATG GACTG CTGAA TGT					
Rex 1	Rex 1- 3D_ SSa		Salmo salar	TGGCG CCGGA GAAGA AGGCA GACGT TTTAC GTGTC	TAGTCAA ATGGCTA CCCAGAC TATTTGC ATTGCCC CCCTCCT CCACACC	MISCRPHYLPREFSSVFFVAVYIPPQSEAGTKT ALNELYSAISKQENAHPEAALLVAGDFNAGKL KSVLPNFYQHVKCATRGKKTLDHLYSTHRDA YKALPRPPFGKSDHNSILLIPAYKQKLKQEAPV TRSIKKWSDEADAKLQDCFASTDWNMFRDSS DGIEEYTTSVIGFINKCIDDVVPTVTVRTPNQ KPWITGNIRTELKARAAAFKERDSNPEAYKKS	cd0165 0	372	624

			CCCAA	ACTGCCA	RYALRRTIKQAKRQYRTKIESYYTGSDARRM			
			CCGAT	CTCTCTG	WQGLQTITDYKGKHSRELPSTSLPDELNYFY			
			TGTGT	TTGTCAT	ARFEANNTETCMRAPAVPEDCVITLSAADVSK			
			TTTTTT	CTATGCA	TFKQVNIHKAAGPDGLPGRVLRACADQLASVF			
			GTTTCG	TAGTCAC	TDIFNLSLSESVIPTCFKQTTIVPVPKNTKVTCL			
			TTTATT	TTTAATT	NDYRPVALTSVAMKCFERLVMAHINTIIPGTL			
			TGCGT	AACTCTA	DPLQFAYRPNRSTDDAISIALHTALSHLDKRNT			
			TGTTT	CCTACAT	YVRMLFIDYSSAFNTIVPSKLITKLRTLGLNTSL			
			GTAAC	GTACATA	CNWILDFLTGRPQVVRVGNNTSATLTLNTGAP			
			TTATTT	CTACCTC	QGCVLSPLLYSLFTHDCMARHDSNTIHKFADDT			
			TTTTAC	AACTAAC	TVVGLITDNDDETAYREEVRDLAVWCQDNNLS			
			TTATTT	CGGTGCC	LNVIKTKEMIVDYRKKRTEHAPILIDGAAVEQ			
			TGTAC	CCCGCAC	VESFKFLGVHITNKLTSKHTKTVVKRARQNL			
			ATAAT	ATTGACT	FPLRRLKRFGMGPQILKRFYSCTIESILTGCITA			
			GTTGC	CTGTACC	WYGNCASDRKALQRVVRTAQYITGAKLPAI			
			CGCTA	GGCACCC	QDLYTRRCQRKALKIVKDSSHPSHRLFSLPH			
			CCGTC	CCCTGTA	GKRYRSAKSRSKRLNSFYPPQAIRLLNI			
			TCTTAT	TATAGTC				
			GACCG	TCGCTAT				
			AAAAT	TGTTATTT				
			AACTT	TACTGCT				
			CTGGA	GCTCTTT				
			CATCA	AATTACT				

				GGACT	TGTTACT				
				GCGAT	TTTATCTC				
				TACTC	TTATTCTT				
				ACCAC	ATCCGTA				
				GGACT	TTTTTTTA				
				GGCAG	AACTGCA				
				AATCC	CTGTTGG				
				TTTTTT	TTAGGGG				
				TCCTTT	CTCGTAA				
				AACGA	GTAAGCA				
				GTCTG	TTTCACT				
				ACGAG	GTAAGGT				
				CCCGA	CTACACC				
				CGCGA	TGTTGTA				
				ACGAT	TTCGGCG				
				ATACT	CATGTGA				
				GCTTT	CTAATAC				
				CTCGG	AATTTGA				
				GAACA	TTTTTGAT				
				GGCCC	TTGATTT				
				AGATC	G				
				CCCGT					
				GATTT					

				GCGTG					
				AAGAG					
				GAGGC					
				GGAGA					
				AAAAG					
				GGGCC					
				AGAGG					
				GCGGG					
				CTGCC					
				TTCTG					
				AGAAT					
				TCGTA					
				GGCGA					
				TCGAA					
				TAAAC					
				CCCCA					
				CTTCCT					
				TCCAT					
				TCTGC					
				TAGCA					
				AACGT					
				GCAAT					
				CTTTG					

				GASAA					
				TAAAA					
				TAGAT					
				GACCT					
				ACGCG					
				GAAGA					
				TTAAA					
				CTACC					
				AACGG					
				GACAT					
				TCAAA					
				ACTGT					
				AATAT					
				CTTAT					
				GCTTC					
				ACGGA					
				GTCGT					
				GGCTG					
				AACGA					
				CGACA					
				CTATC					
				AACAT					
				ACAGC					

				TGGCT					
				GGTTA					
				TACGC					
				TGTAC					
				CGGCA					
				GGATA					
				GAACA					
				GCGGC					
				GTCTG					
				GTAAG					
				ACAAG					
				GGGCG					
				GCGGA					
				CTATG					
				TATTTT					
				TGTAA					
				ATAAC					
				AGCTG					
				GTGCA					
				CGATA					
				TCTAA					
				GGAAG					
				TCTCG					

				AGCTA TTGCT CGCCT GAGGT AGAGT ATCTC					
Crack	Crack-13_Sa		Salmo salar	ATCCA AGATG GTAGC CAATC AGTCT GTCCG TTTGT ATGTG TGATC CTTGA GATCA GAGTT TTAAG AGTTT TTATG TTTTA ATGAC	TAGGCTG GATCATG TTGTTGT ATTGTAT TGTTGTA TGTTTTA ATTATGT AATGTAT TGATTGT TGCTGCC TTCTTGG CCAGGTC TCCCTTG AAAAAGA GACTCTG GGTCTCA ATGGGCT	MYLLLCLLFTQCHPGLNRLNDQWVSELSSPYA LQFANTPTRPSCQYHQELGLSLQLHLYHSQTQ LDLHLMAYFQNLKLLNYFFSVSKDSVRFIDNG AGSTKSSRIGQLRYIVMLLLISGNVRPNPGPV DTMQSLPTPYDFKNRAGFGLLHVNVRSFLPKI DMIRIWAKTTNADVMVLSETWLKKSVSNNNIS IDGYTVFRTDRMSKGGGVAIYVKSEFNTSEILS ITKAKHFELLAVKVNVEYKDSHITVVGCYRPPS ASGDALNSLSVDLHKLNDSEFIILGDLNWDML TSVSDSFKELCDSLNL AQLINAPTRPNPRAPNK STLLDIL TNTPHKYTSTGIFCNDVSDHCAIACV RNTKMTKAKPRYIFKRHFRQFDEQAFLHDLYH NIDRVSLIPDVDTAWDYFYMKFVSICDKHAPV KKFRISGRDNPWFSDNL AELIRKRNISWAQAR HTNAPVDWASFRALRNKCTGLIRKSKSDYYL NAV TENLNNPTKFWKLIKSVSGSNVSSGLPDH LMMDSNEVKDKAHIVGVFNKHFISAGSVFDN	cd01650	635	891

			CCTGA	TTTCCTG	GGAQASNVSSVTVNYDIGPHVNHFNFEFVSYT			
			TCCTG	GTAAAT	EVYKALKAIDTKKSAGPDNLDPYLLKIAAGIIT			
			TTTAC	AAAGGT	EPVAHIFNLSLLTNSIPSIWKSAYVLPLLKGGDP			
			AGAAG	AAATAAA	SDANNYRPISKLPILAKVYESLVNSQLKNFFIE			
			CAAAT	AA	KNILSGVQSGFRSGHSTTTAAMAVANDIINAL			
			ACTTTT		DKKQHCAALFVDLSKAFDSVDHELLLARLRNI			
			TATGT		GLSEGAVNWFRNYLSDRITQCVYTDNHKSSFL			
			GAAGG		EINRGVPQGSILAPVLFSIFINDLGNGMQPAKL			
			TGGAT		HLYADDTVIYSCAPSLVQAVEELQTAFQSLQA			
			GCTGT		SLYGLKLVLNVQKTKFMTFTRARTLPENVISIV			
			ATATA		TSGGLSIEKVSSYKYLGIWLDDKLSFKVHVDN			
			ACAAC		LVTKLKCLKLGFYFRNKACFPLMARKKL VQAT			
			TATTA		FLSVIDYGDLLYMHAASSVLQRLDSVYHASLR			
			TTTTTA		FITNAKSLTHHCTLYQMVGWTSLYMRRKIHL			
			GATCT		YVFIYKALLGKLPLYLCSLVSFTTSSYHTRSAR			
			CAGTC		WLLKVPRTFTVLGKTAFSSCAPDAWNSLQS			
			TCTATT		MLHLDMLVPLNEFKILMGDSVTEECKCF			
			TTACG					
			AAATT					
			TACTT					
			GTACA					
			AAGAT					
			TTCGG					

				ATTGA					
				GAAGG					
				CCTTT					
				ATTTTT					
				TTCTA					
				CACCT					
				TTCTG					
				GAGTT					
				GGAGC					
				TCGCC					
				GAGTC					
				AGGAG					
				CGTCT					
				GTCTG					
				CGCGC					
				AAGGC					
				CTGCT					
				ACGCG					
				AGCGC					
				AGGTG					
				CGCAC					
				TGAAG					
				GCTAG					

				GAGCG					
				GAGAA					
				GCGGT					
				TATCG					
				GATTG					
				GAAAA					
				CCTGA					
				TGCAC					
				CGTGG					
				GCGGT					
				GGGAA					
				AATCT					
				ACGTT					
				GTTCT					
				ACACA					
				ACGTC					
				TACTG					
				GTCGT					
				TGCAC					
				ACCTG					
				TGCTG					
				AAATA					
				CCTCT					

				AAAAA CCGGC ATT					
R2	PER ERE -9	BN00 0800	Schistoso ma mansoni	ATCTC ACGTT TTAAT TTATTT TTGAA CTACT GCAGT CTGAG TGCTT CTAAC GACCC GAAGG CTCAG AAACT ACCCA CTTCTT GAACT GCTAC TTTTTG CTGTTT	TAACGGC TGAACGA ATAGCCC CCTTCAC TCTTAGA CATTCCC CCACTGT TGTTGCT TATCTTC ATGCTCT TGTGTTA ATTGACT GCTCTCT TCTGGGT TGACGTC TGATTGT CTCTCTCT CTTTCCA TATTGCT TGCTCTG	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSST CLICFAIFPTHNILLSHATAIHHISCPPTPVQDGS QQMSCVLCAAFFSSNRGLTQHIRHRHISEYNE LIRQRIAVQPTSRIWSPFDDASLLSIANHEAHRF PTKNDLCQHISTILTRRTAEAVKRRLHLQWS RSPTAITTSSNNHTITDIPNTEARYIFPVDLDEHP PLSDATTPNASTHPLPELLVILTPLPSPTRLQNIS ESQTSHESNKNSMHTPPTYACDPDET LGATPSS TIPSCFHSYQDPLAEQRGKLLRASASLLQSSCT RIRSSSLLAFLQNESTLMDEEHVSTFLNSHAEF VFPRTWTPSRPKHP SHAPANVSRKKRRKIEYA HIQRLFHHRPKDASNTVLDGRWRNPYVANHS MIPDFDCFWTTVFTKTNSPDSREITPIIPMTPSLI DPILPSDVTWALKEMHG TAGGIDRLTSYDLMR FGKNGLAGYLNMLLALAYLPTNLSTARVTFVP KSSSPVSPEDFRPISVAPVATRCLHKILAKRWM PLFPQERLQFAFLNRDGC FEAVNLLHSVIRHVH TRHTGASFALLDISRAFDTVSHDSIIRA AKRYG APELLCRYLN NYRRSTSCVNRTELHPTCGVK QGDPLSPLLFIMVLDEVLEGLDPMTHLTV DGE	cd0165 0	492	750

			ATCCA	CCCGCTT	SLNYIAYADDLVVFAPNAELLQRKLDRI SILLH			
			CAACA	ACTTCCA	EAGWSVNPEKSRTL D LISGGH SKIT ALSQTEFTI			
			ACAGT	ATAGTTG	AGMRIPPLSAADTFDYLGIKFNFKGRC PVAHID			
			TGTGA	TCATATT	LLNNYLTEISCAPLKPQ QRMKILKDNLLPRLLY			
			TTCTAT	ATGTCTT	PLTLGIVHLKTLK SMDRNIHTAIRKWLRLPSDT			
			TCTCC	TGTTTAC	PLAYFHS PVAAGGLGILHLSSSVPFHRRKRLET			
			ANATA	TTGCCAT	LLSSPNRLLHKLPT SPTLASYSHLSQLPVRIGHE			
			TTCCTT	GTCTAAC	TVTSREEASNSWVRRLHSSCDGKG LLLAPLST			
			GTGCT	GACAATT	ESHAWLRYPQSIFPSVYINAVKLRG GLLSTKVR			
			TTTGTC	ACTTTAT	RSRGGRVTNGLNCRGGCAHHETIHHILQH CAL			
			AACAT	CTACCTT	THDIRCKRHNELCNLVAKKLRRQKIHFLQEPCI			
			TATTCT	AGTTTGT	PLEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTH			
			ATACC	CCTCTTG	ASRLKISKYGNERTVASIKRFLTSSGYIITSVR			
			AACTG	GTTTCGA	QTPVVLTFRGILDRASSQSLRRLCFSSRDLGDL			
			TACCA	TTGCCTT	CLSAIQGSIKIYNTYMRGT			
			CCTAC	CATATGT				
			TTCTTC	TCATGGC				
			ATCTC	GGAATCT				
			ACGTT	GATGTTT				
			TTAAT	ATAATGA				
			TCTGG	CTATTCC				
			TCTTAT	TATTACC				
			TTTCTC	ACCACTA				

				ATCAT	CAACTAC				
				TAGTC	TATTATT				
				ACGGA	ATTTTCA				
				GAGGG	TTACTAT				
				CCTAT	TAACATT				
				GAACG	ATTATAA				
				GTCCG	ACATTAT				
				TGACG	TACTATT				
				CGAAA	ATTATTA				
				TTCTAT	TTACTAT				
				CCGCG	TATTACT				
				ATTTC	TCTACAA				
				GACCT	TTAATAT				
				CTCCT	TATGGCT				
				GCTAG	ACTCCTC				
				TGGTC	TCAGCAC				
				CCCGA	ACCAATA				
				AGTAC	AAATATC				
				GGTTC	AATCAAA				
				CTCTG	CATCTCA				
				GCCTG	ATTATAT				
				TCAGT	CCACCTA				
				TGTGT	TTAAACT				

				TAAAA CTATA TAATA ACG	CTCTCTA TTTCCCCT GAGTTAT AAACTTA CAATTCA GTCTAAC CGAATAT CTCTCTTT TACAAAT CTTAAGT ATGTAAT TTTGTGC CAAACCC ATTGGG TCTGTAC AATTTGA TACTTAA AAATAAA TGTTATT AGCC				
R2	R2S m-A	.	Schistoso ma mansoni	ATGTT TTAAT TTATTT	TAGCCCC CTTCACT CTTAGAC	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSST CLICFAIFPTHNILLSHATAIHHISCPPTPVQDGS QQMSCVLCAAFFSSNRGLTQHIRHRHISEYNE	cd0165 0	492	749

			TTGAA	ATTCCCC	LIRQRIAVQPTSRIWSPFDDASLLSIANHEAHRF			
			CTACT	CACTGTT	PTKNDLYQHISTVLTRRTAEAVKRRLHLQWS			
			ACTGT	GTTGCTT	RSPTAITTSSNNHTTTDIPNTEARYIFPVDLDEH			
			CTGAG	ATCTTCA	PPLSDATTPDASTHPLPELLVILTPLPSPTRLQNI			
			TGCTT	TGTTTTTG	SESQTSHESNRNSMHTPPTYACDSDESLGVTPS			
			CTTAC	TGTTAAT	STIPSCFHSYRDPLAEQRSKLLRASASLLQSSCT			
			AACCT	TGACTGC	RIRSSSLLAFLQNASTLMDEEHVSTFLNSHGEF			
			GAAGG	TCTCTTCT	VEPRTWTPSRPKHPSHAPANVSRKKRRKIEYA			
			CTCAG	GGGTTGA	HIQTLFHHRPKDAANTVLDGRWRNPYVANHS			
			AAACT	TGTCTGA	MIPDFDCFWTTVFTKTNPSDSREITPIIPMTPSLI			
			ACCCA	TTGTCTCT	DPILPSDVTWALKEMHG TAGGIDRLTSYDLMR			
			CTTTTT	CTCTCTTT	FGKNGLAGYLNMLLALAYLPTNLSTARVTFVP			
			GCTGT	CCATATT	KSSSPVSPEDFRPISVAPVATRCLHKILAKRWM			
			TTATC	GCTTGCT	PLFPQERLQFAFLNRDGC FEAVNLLHSVIRHVH			
			CACAA	CTCCCCG	TRHAGASFALLDISRAFDTVSHDSIIRAAKRYG			
			CAACA	CTTACTT	APELLCRYLN NYRRSTSCVNRTELHPTCGVK			
			GTTGT	CCAATAG	QGDPLSPLL FIMVLDELLEGLDPMTHL TVDGE			
			GAATC	TTGTCAT	SLNYIAYADDLVVFAPNAELLQRKLD RISLLLH			
			TATTCT	ATTATGT	EAGWSINPEKSR TLDLISGGH SKITALSQTEFTI			
			CCAAA	CTTTGTTT	AGMRIPPLSAADTFDYLG IKS NFKGRC PVAHID			
			TATTC	ACTTGCC	LLNNYLTEISCA PLKPQ QRMKILKDNLLPRLLY			
			CTTGT	ATGTCTA	PLTLGIVHLKTLK SMDRNIHTAIRKWLRLPSDT			
			GCTTTT	ACGACAA	PLAYFHSPVAAGGLGILHLSSSVPFHRRKRLET			

			GTCAA	TTACTTT	LLSSPNRLLHKLPTSPTLASYSHLSQLPVRIGHE			
			CATTA	ATCTACC	TVTSREEASNSWVRRLHSSCDGKGLLLAPLST			
			TTCTAT	TTAGTTG	ESHAWLRYPQSIFPSVYINAVKLRGGLLSTKVR			
			ACCAA	GTCCTCT	RSRGGRVTNGLNCRGGCAHHETIHHILQHCAL			
			CTGTA	TGGTTTG	THDIRCKRHNELCNLVAKKLRRQKIHFLQEPCI			
			CCACC	GTTGCCT	PLEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTH			
			TACTT	TCATGTG	ASRLLKISKYGNERTVASIKRFLTSSGYIITSVR			
			CTTCA	TTCATGG	QTPVLTFRGILERASSQSLRRLCFSSRDLGDL			
			TCTCA	CGGAATC	LSAIQGSIKIYNTYMRGTQRLNE			
			CGTTTT	TGATGTT				
			AATTC	TATAATG				
			TGGTC	ACTATTC				
			TAATT	CTACTAC				
			TTCTC	CACCATT				
			ATCAT	ACAACATA				
			TAGTC	TTATTATT				
			ACGGA	ATCACTA				
			GAGGG	TTATTAA				
			CCTAT	CATTATT				
			GAACG	ATTACTT				
			GTCCG	CTACAAT				
			TGACG	TAGTATT				
			CGAAA	ATGGCTA				

				TTCAA	CTCCTTTC				
				TCCAC	AGCACAC				
				GAATT	CAATAAA				
				CGTCC	ATCTCAA				
				TCTTCT	TCAAACA				
				GCTAG	TCTCACT				
				TGGTC	TATTAAA				
				CCCGA	CTCTCTA				
				AATAC	TTTCCCCT				
				GGTTC	TCGTTAT				
				CTCTG	AAACTTA				
				GCCTG	CAATTCA				
				TCAGT	GTTTAAC				
				TGTGT	CGAATAT				
				TAAAA	CTCTCTTT				
				CTATA	TACAAAT				
				TAATA	CTTAAGT				
				ACG	ATGTAAT				
					TTTGTGC				
					CAAGCCC				
					ATTTGGG				
					TCTGTAC				
					AATTTGA				

					TACTTAA AAATAAA TGTTAT				
Pene lope	Pene lope- 1_S M		Schmidtea mediterranea	TATAC AAACA CCGAT AAACT ATACC TTCCC AATCC CGATC CAGAA TTAAA TTCAA CACTT TCGAA GGTTT GTCGA AAGAT T	TA ACTCT GATACTT TTACTTA AAGATTA AAACAAT TTTATAG CAAAAAA GGAAAAC CTTCAGG AATGAAA ATCAGAG TAGAATT GTATTTA GGTCTTG AAGTTGT TTGTGGA GGGTTGT TTGAAGG CTTTTTTG TTGCGTT	MDVVEKEGKRKPYKSNLSNNEQGVLKALKTK PYVYLPSDKGGEFCVIEKSRYCEAAAYKHLSDS NTYKKISRMSPNTIETKINKVWKDIAAKMNLK KRLTKSFVTKNSDLPNFYHLIKTHKDGPSIKIRP IVSNTNGPTTKTSWLLSKILKPLLNNVPYHLQS SDCLMARLNKTGKTTHSEFKYPFSLDVVALYT SIDPSEAIECLRRRLNLLIKEMNTENIINLLT VILENTYFTFDQNIYKQIKGLPMGSSISGLLAIF MDYLERQTLRTIQSIGFYARYVDDIFITTKSKK EAEEIFHIFNEAHPDIKFEIEYPDENGSLNLLDL RIKIDVDTPTYGFYKKSARKEIFMHQRSALPYL QKVNVIKNEKQRIANRCTEGKENYKQQLXFD KVLRRNGYHKEDIERMNSKPKTPMDRTQKKK KYFYLNIPFISDTLDFKIRKELNHTGLPIRISHKS WTLRQALANNTKKENICTMENCQTPKKICFRK NVVYCITCNACKEIYIGSTLRNLHQRIKEHLDP AGKSSVYRHLQQCQPRKERKSISIKILAGDNIQ KNLRFKEAILLDKHPSINTKEEISEIASLICC	SSF56 672	144	314

					TTTAAGA				
					GGAAAGC				
					TTTTTCAT				
					ATACATC				
					TTTCATA				
					AGACTTT				
					AAAAACT				
					CAGGAAC				
					TATTTATT				
					AATTTGC				
					TCTGAAT				
					TTTATAA				
					AAAACCT				
					TCTTTAG				
					TTTTAATT				
					TTATTAA				
					GAAATTT				
					GAAAGAC				
					AAAAAAA				
					AATGACG				
					TTAGAAG				
					CAGTTAT				
					TTAACAT				

					AAGATAT				
					AGAAAAC				
					ATTTTAA				
					AATTTTA				
					AAGTTTT				
					ATATAAA				
					AAAATTA				
					AAGTTTA				
					TAAANAT				
					TAAATTA				
					ATAAATG				
					TTAAATT				
					AAACAAA				
					AAAGAGA				
					GGAAAAA				
					AATTTAT				
					TTTAAAA				
					ATTTTAA				
					ATTGTAT				
					ATTTTAT				
					AAAAGAT				
					ATTTAAA				
					TTTTTAA				

					AATTTTA				
					GGTTTAA				
					AATGTTT				
					TAAAATT				
					AAATTTA				
					AACCAAT				
					CAAAAAG				
					CAAAAAA				
					GTNAATT				
					TTAATGA				
					AAATCAG				
					AAGGTAT				
					TTCGGCA				
					TTTTTCTC				
					ATTGTAT				
					ATAGTAG				
					TTCCGAA				
					AGAAACT				
					GAAGAGG				
					ATTATTT				
					AATAATC				
					GAAACTT				
					TTTTCAA				

					AAATATA ATTATTA CTTCGCA GCAAGGG CTTATTTT GTTAAAT AATAATA ATAAT				
Петр отра нспо зон, не соде ржа щий LTR	LIN 10_S M	.	Schmidtea mediterranea	TTTAC ACATT GACAT TCGAT AGCAC CGGGC CGAAA TCCGG ATCTG ATAAG TTATTT TATTC AAAAC AACCA TTRACTT	TAACTTC TA	MSKNGTDPHLELCYKAHVFLNGSKEIPLSTA VGHWREIDSTGIFDSDFSELMEQIIAKPFKTEGL QLKIENAIKTFDKFIADAVNPDKLINHKHLNFF YNQLYYNPLKNFNFSIPIKSILKNNNIPSHRNFF NNEESTPLNKKSISFKNISHSKPFSYLTALTNNL KYENIDNILISIRNCTDSFFINLNICFFELKHLIK NNFINYFNDYFYIFYLKIQKFNNKKLQKSNYY QNNIPLLFKISSLTNDFFINIENYFSKIKTKRNIN LFLNDFLXILKIKINILLFKYSXFKKLNPKNKHE KISTRSSXASGRLALAVTPSSPLDLTINTTKQTT TDINTDTTNTSHKFSNNRYSFLDQNEPLDLSI VKHPHKNTNIPSTVKNTYNPLNKQKPFHKTTT KTIFSNAQNNIKKHNIIDTSKRPHFSXPIPKSCNT DIKYEIPPIEKNTPKLNDIIXSNITFNPKTPNIQQ NISLPKFILPSNLHNNTTNVDLDAESTLVNPHF	cd0165 0	1274	1556

				TATTTT CCAGA TAATA ATGAA TTAAA TGCTG ATCCT ATCTTT GAGAA AAGTT TGCAA GAATT ACGTC TT	QLNFNLKPCPNLFIPKXDLPXNLNMEIDQKLLL TNTKASDETISYSPITQTQTDESVLTHKLIELISI NTQANRIIEQINSIDIDLFSQSPHSYEIPNTETKEI FTVYNNLDIAKKRVEIIIIEEDILKTKTQTNDIP MINFNNNYINNNSLFGNILLSNESVENLFRDISD NDSIISKSPNKKINNQDFLKKTVKINKTMFNL NDGYLICIGGKCEKKKIGIHNDNIIHRKTNHS DTIDKNTTFLCLCEGSVKRKTILLNKIIQHTET FHSEKHFPVTREELNNKNKILELXMNTEGTLN CRYTNIKKHCQNDLTYKSTYAQIEKHIAKKHK RKTCNINEIICHCGETLEHHNIEHHNKHIGVLP AEDKLRMIPDNIPVSVKDIGNDISILSTFFCMHL LPQSQQLLYNHTLLKPSYKQNFKTFFETFPFSQ YPNWNEIILPFKTSTVNWAVFYFNKLKAFATII DPLSPNSLQKHRKIADTILNTIKNIQNIYPSNIN LTELNYPMCANTNESGFFICHIIQCISINHNYITI PIISEIRKRLDNLINKVIIDRAPIDNILKYTNILNN IKILQNNNSENPDGLAEINRIHECIFPIKHKIKK XLSLIHLLLLKIIKLNNKXXFNPPSTIIKNNKIE LTYKLRFNFNINYKSTINKIFNNSINTLSHPDRI ILNFEREIIINDIKFKYKLPYYPENFIHKPTTIEEI TNTFKIVDLSSSPGKDQISYLDWKKIDPDNSILC SLFNNILFKSITPLKWKTFKTKLIKPGKENSSH		
--	--	--	--	---	--	--	--

					DLSSWRPLAILDTSYRLFATLINRRLLEWIKLG NLININQKAIGCPDGCAEHNSVLLLLNERAVR MGTNINIVWLDLADAFGSISHDLIWNVLTNMG INNNTIKLFKNMYNNCNTYYECENNHYNFLKI CIIIVTLIMNVKTTTTKTIPINNGVKQGCPMSMT LFCLAINFIITNILKDHPFIINNYNVSILAYADDI VLISDNRIKIAKILTDIIKYTSDVKLKFRPTKCGF YENNRNHDEKPLKIYNEIPIIDENNIYKYLGVG FGGKSNHNIDNILEKALTDAEATFLSDLHPSQK LFAYKTSIHSRLIFPFRTCIVKHMILDCDRNRIV DNRTKQLGFDQKIKRLIKNSIGDIFQNISNNFIY AHCSMGGLGIVPSIDEYLIQSTIQLTRLLMSSD VTMRDFIIGELIECVRARFNYLENEIEMAMKW LNGDIDDEKVGKPCFLKFKKSVYRLKIKFSIIC KLTYIDNIFSLEISKNDFKIVINNENIDNASKIM HILTGEWYALEWYKMTHQGHIAETVGKSQFS KHLIKSTTLGDNQYFFLIKARNNSLNLNSNPSN HKDKKDVCNRLCRIQWETQAHILNHCTLSQN ARRNKHNHVLEIVHNFFASKNYDIDIEKSPPNV DTTLRPDLIKTDDKNKIKLSLVDIKVPYDNET NFINARKQNLDKYRNLAJETGKDNCKAAVY TIIVGSLGSDWKENNKDLQKLGLGVNEIGEMA RMCIKAAIEASYMIFMNHVTNNK		
--	--	--	--	--	--	--	--

RTE	RTE - 1_Si t	Setaria italica	AACTC CGCCT	TGATTTG TGGCTCT TGTGTTG TTTCAAC TCTAGCC TACCCCA ACTTGCT TGGGACT GAAAGGC TATGTTG TTGTTGT	MLSQHRSRQALVKEEGCDRLGEPTLNLAILME MKPKRNPLGRNPLSDAPYRNPGMVLNGQGP RLPLGDAPCRDLGMVSSEQGSGRHFLSGALHR RPGVVKNEQGSSHLSSRRVRRARKLVEPTRIRV GSWNVGSLTGKLRELVDVATRRRVNLCVQE TKWKGQKAKEVDNTGFKLWYTGTAAANRNGV GVLIDKSLKDGVDVRRQGDRIILVRLVVGDM VLNVISAYAPQVGHDESAKRLFWEDLDSLVRT VPSSEKLFIGGDLNGHVGTTTASAGFEAVHGGFG YGSRNQEGEEVLDFAVAFDLMIANITFFRKRRS HLVTFSQGQHYSDIDFILTRREDKRACLDCKVI PGECVVSQHKLVVADFRFQMRARRDKQAKIA RTKWWKLKEETSEVFKERVKEGPWKEGEDA NNMWEKMANCIRKVASEVLGVTKGGRGVDK DTWWWNEEVQRTIKEKKECYRRLFHDRSVDN IEKYKEAKKIAKRAVSVAKGRAYEDLYQRLG TKEGEKDIYRMAKVRERKTRDFNQVKCIKDE MENILVKEDEIKHRWQEYFDKLFNGENENTTL QLDDSFDDTNRHFVRKIQESEVREALKRMKSG KAMGPDGIPIKVWRCLGDIAIVWLTKLFNHIFR SNKMPEEWRRSILVPIYKNKGDIQSCTNYRGIK LMSHTMKLWERVIEHRLRRITRVSMNQFGFM PGRSTMEAIFLIRQVMERYREKKKDLHMFID	cd0165 0	641	890
-----	-----------------------	--------------------	----------------	---	---	-------------	-----	-----

						LEKAYDKIPRNVMMWWALDKHKVPTKYVGLIK DMYNNVVTRVRTSDGDTDDFPIRIGLHQGSAL SPYLFALVMDEVTRDIQGDIPWCMLFADDVVL VDESRAGVNRKLELWRETLESKGFRLSRTKTE YMRCDFGTTTHEEGDVSLESQVVPRKDTFRYL GSMLQRDGDIDEDVSHRIKAGWMKWRQASGI LYDKRVPLKLKGKFYRTAIRPAMLYGAECWP TKRRHVQQISVAEMRLRWICGHTRRDRVRN EDIRDLRGVAPIEEKLIQHRLRWFGHVQRRPPE APVRSGTLRRDSNVKRGRGRPKLTWEEAIKGD LKGWNIPKDLALDRSAWKTAHVPEP			
Pene lope	Pene lope- 5_Sl n		Solenopsis invicta	TTTTTG TGTTCT CATCT TTTTTT GTCTG GTCTT GATCA TTTGT GGTGA TGGGT GTCAT TTTTGT	C	KTLLKVETNPXTSTNNKWVFNISKTEIPPSVSN LLQLGGNFCLPPNKKTSITEFIKDIESNRCIQKE RGEIRNSVIPLLHNFLHNNSSLDEVEKELLRMK RATSEFCKSNPNVIFTRADKGNITVAMDRLDYI NNMEQLLHDKETYSIKNNPIKKVEKDLNTTIK RWYDXDFISKQSYFSLHSSDSTIPKAYGLPKVH KENFPFRIIVSSVNTALYPIAKFLHKKIINNSLPFD SRQVYNSFDIAKSLSGMKIREEYCLLSLDAKSL FTNIPLDLAMDGIRKRWPSIENNTKIPFDEFIFA LKFILSSTFFTENN VFYKQTFGTPMGSPLSPIA NIVLHDLEEKALKSINLEIPFYRYVDDIVMAA PPNSLNFILETFNSFHNRLQFTIELEQDRSLSFL	SSF56 672	215	444

			ATTGA	DLSLKITDNTIHDWFHKKTFSGRYLSFFSHHPIC		
			ATAGA	HKIGTIYNMVDRAALLSHPTFQQKNLELIIGLL		
			TTTTTT	LKNGYPLEFIFKYIKIRIKNLFNTKMAPIITNDQ		
			ATTCTT	DQTKKDENTKKFIVVPYIRNISEKTTSLLDKNT		
			ATTTT	FTFGYRCINKINNFIVHKDKTERDCSSNVVYK		
			AATAT	IFCNNCDATYVGQTKRQLRTRAKEHINNIKLD		
			ATTTG	PSKHSVITNHRIFFDHSFNWKDIKILDTEENYQ		
			AAAAT	KRLISEMIHIKQQTNGLNSQNDTDLLHDSYFCL		
			AAATT	LDSLKDNTNI		
			CTAGA			
			GGATA			
			TCCGT			
			TTTAT			
			AGAAG			
			TAGTC			
			CAATA			
			ATTAA			
			TTCCA			
			AATTT			
			TTTTGT			
			TGGAA			
			AGTAG			
			GATGA			

				GATAG					
				CAATA					
				GCGCT					
				CTATC					
				CACCA					
				TGTTA					
				TAGAT					
				TGTTC					
				CAATT					
				TTATG					
				GCATA					
				TTGGG					
				TGGTG					
				AGAAA					
				AAAAC					
				GATAA					
				ATAAC					
				GACCA					
				GAGAA					
				AGTTT					
				TTTTAT					
				GGAAC					
				CAATC					

				AATGA					
				TGATT					
				GTATT					
				GTCGG					
				TTATTT					
				TTAAT					
				GACAG					
				GTCCA					
				AGAAA					
				CTTAG					
				GCTAC					
				GGTCC					
				TGTTC					
				CAATT					
				CAATG					
				GTGAA					
				TTGTA					
				ATCTA					
				TTATG					
				AAAAC					
				TGTTA					
				AATGT					
				CTCGA					

				GGATA					
				AAATT					
				GAGTG					
				AGTTC					
				GGAGG					
				TGCTG					
				CCATA					
				ACAAT					
				GTCGT					
				CGACA					
				TAACG					
				ATAAT					
				AGAAC					
				GGTAT					
				TTCTA					
				GATTT					
				ATTGA					
				CTTTA					
				ACGCT					
				TTTTCT					
				TCTAA					
				ACCGT					
				GCAAT					

				ACGAT ATTCG CTATA ATGGG TGATA AGGGC GATCC CATTG GTGTA CCAAA AGTCT GTTTG TAA					
R1	R1- 2_S1 n	.	Solenopsis invicta	GGACT TTGCG GCGGG GCTCG GCCTG GTCCT TCTGA ACCGG GGCCG GGTCA	TAGGCCC TCAAGCC CTGCCGG GGGCGCC CATCGCT AGGGCGC CCCTGGC GAGCCGA GTCGGCC GCCAAGG	MMSRACDASMPRSRPMARRAAYWWSAELAE LRRATVAARRAHTRAKRRGMEDALNNAAV LREARKALRTAIARAKAAAWDELLSELDRDP WGRPYKIVMKRLRPWAPPVSETLDPQFLGEV VDKLFPTRGRDIPEKTGPPPGWTEEELEVS RHE MVAAFARITGRKAKAPGPDGIHAKVWARASSI IGEALRGPTYTKCLREGTFPRRWKRARLVLLKK EGKPAESPSAYRPLCMLDDAGKIFERIIANRIIR HMSRNGPDLSPGQYGFREARSTVDAIRQVRSL SESITGDGKVALAISLDIANAFNSIPWGEVVAA	cd0165 0	210	472

			GCACC	CCGACCG	LSEHFRLPPYLVAIVRDYFRDRVLEFRDKTGLQ			
			TACGT	GGTCCGG	QRRDMSCGVPQGSVLGPLLWNATYDKVLRG			
			GGGGG	GGGGCTC	ALPPGCHVVGYADDTFVVAGGASWVIAIAMG			
			CCGGG	CGGAAGT	NWAVNCVVGSIKSLGLRVAPEKTEAIFHDGS			
			GGGGA	ATGCTGC	SRVPPRAFVRVDNTRVRVGATLKYLGLLLDG			
			GTCCA	ACGCGTT	RWAFVEHFDKLAQRLGRRTDALLGLMPNLRG			
			TCGTC	CCCGGAG	PRVGARRAYAYAAMAGALYGAPVWFREALA			
			GACGT	CCCCCA	SRRIRDVLHAVQRRLARRIVRAYRTAPSAATM			
			GACGT	GTCATGG	VLAGLPPEFTADALAWSYARAKAVRLQGGV			
			GGGCC	ACCAGGG	VTPRVAALLRERARRRAMRAWRESLANDPPA			
			TCCCC	CAGGACA	AGSRILEAVLPHLDEWVGRPWGGLSYRTTQV			
			CGGCG	CCCGGAG	LTGHGCFGEYLCRIRKEPTTQCHHCEADRDSA			
			CCCTT	AGGTTTT	QHTLEDCPAWEELRGVLRREEIGDDLRLAISQ			
			AACAG	AGTGGGT	MVSRERAWVAVSSFCEQVMRQKEEAETVRER			
			GGTAA	AGGCCCT	RPGGRMPPGEGNDGRGGGDGGHDPPWLPPRP			
			AGGGA	CGCCGGC	PRRRRPAPRPPGPAARLXKGGGGPADQPRSP			
			TGGCA	ACGTCGC	PSPQLRRRGATAPIWRGGGDPPPLRLPAPPSGH			
			AGTGG	CGGCGGG	AAAGGGPLTPASPARQTRPDGPGGATLRHNAE			
			AGACC	GGAGAGC	SLPPLPRGGGRGKGLLSPPQGRKXGRAAVSA			
			GGCTC	CCCACAT	ARRGPGLRRGPDTPGLYPQQHQGKRRGRDGV			
			CTTGG	AACCACC	PLPRPRAD			
			AAGAA	AGGCCTC				
			CTGTC	CCCCGGG				

				TGATC	GCCTGGG				
				ACAGA	GTATGCG				
				CTTAT	GTAACGC				
				CTCCA	ATTTCCT				
				TGGAG	CTCCGTG				
				CTGGT	AACAAAA				
				CTCCC	AAAA				
				CCCCC					
				CCCCC					
				CCGCG					
				GAGGT					
				GCGTG					
				AGCAA					
				CGCCG					
				CCGCA					
				ANCGG					
				GCCGG					
				AAANC					
				GATGG					
				GTCCT					
				GAAGA					
				AACTC					
				GACCC					

				CGAAG CCCTC GAGGT CAGTC TCCTC GCCTC CACTT GGCCG GAGGG AGGAG CTGAC CTCGG AGCCG AGGAG GAGGC GGAGC GCCTA TGCGA G					
RTE	RTE l_Sa r	.	Sorex araneus	ATCCA CTACC CAGCC ACCGC	TA	MVTPLRAIGGRP KSLRPSRRARQATESIPPARQ SLASYPSRIRYANNSNNNWP HSPVTERAPNTA ALRKMSKERLKKSRGRTRL PKRRKTKMIVCTF NARTLASEASIEDLMVQAQKIKYDIIGLTETRR	cd0165 0	583	837

			CACCC	HQSHHAIFDTGEELFLGTCDNRGVGGVGVLV		
			TCCAG	NTNLAMSIDSFECLETTRIGRLRLNRCGSLPAVSI		
			GTCGC	FVVYAPTSNYDEEEIERFYMELEKFKEDHTF		
			TCTCC	YKIIVGDFNAKIGPRRSPEELHIGTHGLEWNDQ		
			GGACA	GERLSEFIMSTKTIHGNSQFQKAESKRWTWES		
			CTTGG	PGGQFHNEIDHIIFNRRFCLTDVAVVPKFQTGS		
			GCCGA	DHRLLRKAFYFTERGEKSAKFKSKKATPRMTT		
			ACCTC	NWELFGTIAATWEDAVLDNIDEEYDRLVQHL		
			ACGCA	HVCAKNAESEKATNRRLSSETLELIRQRGLAR		
			CGAGT	ASGNHKL TSELTKLCRDAIKEDLKERRAAVLA		
			GAAGC	SAAEAGKSIRNARLSFANYKTKMTALRRPDGS		
			CCCGC	ITSSRKAMEKIIHDFYSDLFDSHVHLPTYQIPQD		
			AAAGC	GYVVPNVLPSEIRHAISLVKTRTAPGPDKVRPE		
			CAGGT	HLKNLPPVLINTLARLFTRYLSECKVPSQWKTS		
			AAGCA	RTVLLYKKGDIHDIGNYRPICLLSVVYKLFTRV		
			GAACT	ILNRIGRTLDEGQPCEQAGFRRGFSTIDHIHTVT		
			TTGCG	KLIEVSREFKMPLCLTFIDLKKAFDSEVETAEVIE		
			ACCTT	ALAKQGVQTQYIKILRELYCGFTTRISPFIKEVI		
			GAGCT	IDVKRGVRQGDITSPKLFSAALENIMRRLEWE		
			CTAGG	GMGVKIDGRQLHHLRFADDIVLITPNISQAAQ		
			TTCAA	MLADFDHECGKVGLQLNLNKTMTFMKNELVP		
			CGAGA	EAPFALNGTNISECSSYVYLGREINMRNDLVPE		
			CCTGG	LRRRKRAAWNAFKSVEEVVKRTKNLRLRAHL		

				AAACA	FDSTVLPALTYASETWALRKQDENAIRVSQRG			
				AGATC	IERAMLGISRLTQVREGIRSSDLRRRSKIRDAVS			
				CTCTG	FAKASKIRWAGHVMRFRDDRWTTRAVIDWIPR			
				TCTGC	DVRRPRGRPPTRWSD			
				CTCCT				
				CCAAT				
				CTCCG				
				GCATC				
				TGAGG				
				GAGGG				
				GTCCA				
				ACCCA				
				GACGC				
				CCCAC				
				CCTAC				
				CCAAG				
				ATAAA				
				TACCT				
				CACTG				
				GCCGA				
				CCCCC				
				ACTAC				
				ATCCA				

				CTACC					
				CAGCC					
				ACCGC					
				CACCC					
				TCCAG					
				GTCGC					
				TCTCC					
				GGACC					
				GCGGA					
				CACAC					
				CATAA					
				AACAC					
				TTAAT					
				ACCCA					
				TTATTC					
				CCGCA					
				CCATA					
				TTTGA					
				TAGGG					
				TTCAA					
				ATTTG					
				GTGTG					
				AACC					

Ving i	Ving i- 4_S Pur	Strongyloc entrotus purpuratus	GGGGC	TGAAGTG	MKMPNRAASQSNSAEGPLSRPALTVLSWNAE	cd0165 0	476	731
			ACGCA	GCCGGAG	GISDAKEEVLAHMCKQMNCVLCLQETHRGS			
			ACAAG	TGGCAAT	DNPRPKIKGMKLAIERPHRQYGSAIFMNEKIRP			
			AGTCC	GCCTAAA	DSTSYTHNNDIEVLNVTIKGLTISSYKPPNISFS			
			TATCT	TCGAATG	YPSTTTNRHIVIGDYNHSTQWGYDTTGDDGA			
			CGCAC	CGGACGA	LVETWAQNNDLKLHDPKLPKSFASARWKRR			
			CCCGG	GAATGGG	YNPDLVFCSRLLAPTSRKFAEPIRSQHRPIGI			
			GGCAA	GCGGTGT	QLTAVITPNKTPFKRRFNFAKADWPAFTEHLD			
			CGTGT	AAGGTCA	HALTHLPADPKNYESFIDIVKKVSRTHIPRGCR			
			CTGAC	ATATGGC	TTYIPGIDEESSQLLANYTTAYDTDAFSED TINL			
			ACGGC	AAGATGG	GEELISRISQRHRQMWHMIETDLRHSSRKA			
			CGGTT	GGCTATG	WATIRKLGIDSTRAQQHYNVTADDEVANQLLK			
			GTAGC	CCCATGA	NGKAPKTTNSHRHKLPQVDQNMSLFTRPISEE			
			AACTA	AGGAGAT	ELQKGISTLHVNKAAGIDDIRSEDIKHMGSSVA			
			CTTCG	ACTCTCT	QKWLLDMFNNCIENTCVPKQWLRTKVVA ILK			
			AAGAT	GCGACTG	PGKDSNEAKNFRPISLLCQTYKLLERLILNRIGP			
			CAAGA	CGGAGAG	HVDSQLNCEQAGFRPGKSTTGQLLNLTQHIED			
			TTAAG	ACTCAGA	GFQNNLITGAVFVDLSAAYDTVNHRR LQTKL			
			ATTAA	CGATGGA	HDITNDLKLTRFIQKMMQNRRFFVELDGQKSR			
			GATTG	CCACCTC	WRKQKNGLPQGGVLAPTLYNIYTADIPSTENT			
			AGAAT	TTGCAGT	RNFIYADDTCITAQHSTFEA VELHLNTALASLS			
			AATAA	GCGGCGG	EYYRANYLRANPSKTQVCVFHLKNREAKREL			
			TAATT	AACGGGT	KLKWNNVQLNHCEHPVYLG VKLDRTL SFKKH			

				TAGAC	TGCACAG	IEKTIGKVESRNAIIGKLASSHYGADPKTVQSA			
				TTAGA	ATGAAGA	AVALCMSSAEYACPVWSRSAHVKKLDTTLNN			
				CTTAG	CCTGTCC	TYRKITGCLKPTKVEEIYHLAGIGSPNQRRNT			
				AGTAA	CTTGTTA	CLKERMKQATDDRHMLHGYLQPHKRLISRHS			
				TATTA	ATGAAAC	FLDSVPPPEQNPPPSAKPVLIKGLT			
				AGATT	GGCCATT				
				ACGAT	GCAGCAA				
				TAAGA	TCGAGCG				
				TCAAG	ATGGAAG				
				ACCAA	AACTTAA				
				CATAA	TTTAATT				
				CATA	AATTTAC				
					TTCTAAG				
					TTAAGGA				
					ACTTTAC				
					ATTACTA				
					TCTTGAT				
					CTAAGGA				
					GACACGA				
					ACGAACG				
					AACG				

Rex 1	Rex 1- 1_S P	Strongyloc entrotus purpuratus	TGAAT	TGATTGT	MDTYKVYVLFYIALCIIIIGLRGHSSYSKHETAF	cd0165 0	658	909
			GCGCC	ATTTAAT	VPVRAPSRAHKQLHRSGYHKTPIKYYANASAT			
			TAGCA	ATACTCT	FTLPIICCGDVESNPGPFELTSCNTNKRYKYTR			
			GGCAG	GAAACAA	GELVNINSNSCGKSLRLPVQTWKTISELEIGKE			
			CCGTC	ATTAAAT	FKTRRGCRAGVRKLKHFKHPIPIILPTARTKLP			
			TCTGTT	TGTCAAA	KKYPRFPRRRCLTSVPVAPQSGCDLPSFLLSNV			
			TTTGCT	TTACAAT	RSACNKFDELSAVISYLHVDIAAITETWFRNDC			
			CTGAA	TTCATAT	PLDNFSIPGYYLCTKSRTNRIGGGVAIYVRELL			
			TTTTAC	TCCTTAG	NPSPADVQVPINLEVTWVKIRPPRLPRSISCIYC			
			CNTGT	TATATAA	AVVYYPQPDVELETELIDHLETTIDVILTNHTD			
			TTTAC	TTAATTC	AGFVILGDFNRMNIDPVLLTHSFRQVVD C PTR			
			CACTG	ATAAATT	GAHILDKIITNIHQFYDKPQICSPIGMSDHNSIM			
			ATAAT	TGAGGTC	WFPLKIKPKLPNTTRYVFNRPM T DSAIRAFGQ			
			CTGAG	TGATTTTT	WISSYDWPEVTEDVDVNVRCGSFYSTLNNAIN			
			TGTAT	TACCAAG	IYFPVRKVLLHFRDKP W MTGEIKSLIHKRQIAF			
			GCTGT	TTAGTCA	HQKGILWKRLRNKVNSSIQKAKVKYYRSRIQ			
			TTTATT	ATGTATT	KLKKSNSPRWYQEIRILTKGSRQSLNIPVDGVD			
			GTTGA	TTTTATG	PSERDKIANAIN EHLISVSSDLQPLNLAALPAYL			
			TCAAC	ACATGAA	PSRSLPEVYPWQVNKVL SKIKCSTANGPDNISA			
			TGTCA	GAATGCG	RLIRLFSVELSGPLTYIINASFSGKVPASWKCA			
			TCTTAT	ATGCATA	IVVPIPKVNPARI D KLRPISLTDHFAKVIEGFIAK			
			GTACT	TCTATAT	WMYQDIAQHIDPCQFGNQKRLSTSHCLINILH			
			TGTGA	TTATATTT	YVFLNAEVPKTSSTILLTDFQKAFDRIDHTVAI			

			CACTG	TAATATC	KKLIEIGVRPAVVYWISDFLSERKQRVRYHSDL			
			CATGA	TCTAATG	SEWCQINAGVPQGTKLGPLIFLVMINDFKSSES			
			CATTTT	TTCGCCT	IHMFKYVDDLTLVECRKTNQISNMQEAVECLS			
			TGAGG	TATAATA	LWADINKMQLNPTKCIKMDICFSRVPVNQSSII			
			TGATG	TTCATAT	LENCVLSTVSQVKLLGVTIQNDLKWEAHVRDI			
			TATTT	TTGTCTT	EKRANAKLYMLRSLGKFGLSHEDLITIFIGYIRP			
			ATATA	GCAATGT	ILEYATQVWSGGLTSAQKQRIERLQKRALRIIF			
			TTAGA	ATCTTTG	GKDYHSYANALTKSKLQTLESRRNDLCVTFFK			
			TACTG	TATTTTAT	KTLSCTDQFKQFIPQRAHNRTLRRRTNKIPEPRC			
			GAAAC	CAGCATG	KTKRMQNSPIPDLIRLYNKSFL			
			TGTTTT	TACATAA				
			TCTAA	TTGTTTTTC				
			ATTTT	CCTCTCT				
			GATTC	GTCCATC				
			AACGG	TGTCCCC				
			ATGTG	TTTCTCCC				
			TACAC	CCCTCTC				
			AGTAC	TCTCTCTC				
			ATGCA	TCTCTCTC				
			AGTAA	TCTCTCTT				
			AGCCA	CTCTCTCT				
			TTTTAC	CTCTCTTT				
			ACTGG	TCTCTCTT				

				GCTGT	TCTTTCTC				
				TACAC	TCTCTCCT				
				ATGTA	GCTTTCC				
				CTATA	CATCTCT				
				CTGCG	CTTTCTTT				
				TGGAT	TTTAACT				
				GTGTG	CAAATTT				
				AACGT	TTAAATT				
				CGTTC	GAAATAT				
				GTGTG	TGTATAA				
				TGACT	GATTAAT				
				GATTT	GGTATAT				
				GCCTA	AGTATAT				
				TCCGT	ACACGCT				
				GGTGT	CTTCATA				
				GACAT	ATTATGT				
				TCATC	AGATTGT				
				TACAT	AGTGTGA				
				ACAAT	CATCTTG				
				GAATA	AAATGTA				
				GTCTT	TACTTGC				
				GATAC	TTTTTTCC				
					AGCAATT				

				GCACT G	TACTCAT GTGCAAT TCAGTTA CTTTTAC TGTGATA TCTCTTG AAATGTA TAATTGT CTTCACC ATCAATG TTTATTC ATGCGCA ATTCAGT TACTTGT AACTGCG ATACATG GAAATGA ATCAATA AATACCA TTTATCT ATCTATC TATCTAT CTATCA				
--	--	--	--	------------	--	--	--	--	--

RTE	RTE - 6_S P	Strongyloc entrotus purpuratus	GGTAT GCGGC TCTTTA TGTAG GTAAC CCCTTT TAATT AGCTC AGGGA AAATC CACCT CCCCT GCCTT GCATA GATGG CATTG GAAGG TGTAG CGAGC CTCCA TTGTA GCTAT ACTTT	MMRQTNKNGYRPGCRPGEQEP RPPVTNQA EV GKLNTGFKSKSRIKIIDCKAATLISSFNVRSLKS KEKMGE LTA LAEEKRIDILCIQEHHIYHENVDI QMKDMGKGWVLAMSSGTKNTVNSTVGGVGI LMSPFAYKSLEYLEKVSSRILLAKFSGNPAPTV ITCYSPTNCSPDNEAVNFYSLLSDVIKKIPQHN MKIICGDMNAQVSSADSKGFTFGNRTNRNGQ YLLNMTFECDLIACNTRFQKAKEKLWTFTYPN KSQAQLDYILVNKKWRNSVKDCQAYNSFGPL NSDHRIVSTKMQLSLRSNKKVKAKSNLPHDW QLFKSDCSIKEAYTVEVNNRFEALYQDIENPTA NEIYSALIEAHEGAATKYVPKRKGKKKRV PWE NEKIREKRKCLKVAYNKKNGAPTAEHAETLD KAKKDLQDAYDEEQSNYIAQKCTEIEIGAVQQ QTRKVWDIVKEITGKXSTPVDQLRGKTXQERL NQWKDHFSNLLGQPPPDNNNPVSKFTIAELKV AVTSLSNNKCPGVDGIPAEVWKSGALDQQ LLE VCNRTL IHGEKPD LWSRSIIPV P KKGDLSTPQN YRGIALTPIAAKVFNKMLLNRI RPXLEPLLRKN QNGFXPGRSTVXHILTLRRLVEGIKAKHLTAV LTFVDFXKAFDSINRKKMLKILXAYGIPHTIVT AVGV MYKDTTAQVRSPDGETDSFTILAGVLQ GDTLAPYLFVIALDYALRTAXEGFEDLGFTLQ	cd0165 0	560	825
-----	----------------------	--------------------------------------	---	---	-------------	-----	-----

				GAATT TATTCT TATTA GCAAG CTAAG AAACA CTGAA		ERKSSRYPAVMITDTDFADDIALISDDLEKAQE LLXQVESAAASQVGLQINSTKTEFMLRNLREGEI CTADXTKSKQVQNFQYLGSWVISKQDLNIRK AKAWTALNKLTAIWKS NLGRKLKVRFFT VET VLLYGSEAWTLTSSLEKQLDGCYTRLLRAALN VNWRDHMTNIELYRDITXISARLRDRRLRFSG HCWRSKSE			
RTE	RTE - 3_S P	048B 05	Strongyloc entrotus purpuratus	GGAAT ACCAT CAGAG TTGGA ACATG GAATG TGAGA ACACT GTATC AGAGC GGCAA GCTGG AAAAT GTAA ACAAG AGATG	AGAGAGA GA	MVITNTWFKEHPRRVYTWKSPGDIVRNQIDYI AINSRFKRAVKQSKTYPGADCGSDHVPVICTL QCKLKKLKRPKVVKILDFKQLSKPDIRQEFTV QVKNRFELLIDEGEENTWETMRNILVETAEY VPTKERKRRNKWMTEEILLMMKDRQKIGNRQ CKKYRDLDKQIRQKCNEAKESWLNEQCVEIE NQFDKNSNVYQRINEISGKKSGCSNPGCIK GK DGSILFEMKDKMSRWTQYIGELFEDSREAMPS FADEVKGEKILKSEVRWAMSKMKKNKAAGP DGVVTEMFEALEEYGV DRLTEIINKIYDDGTFP DELGRSIFITLPKKPGA VDCEQYRTISLICHITKI ILRILLRARSRITPMVGREQFGFVKDAGTRNA IFVVRNITERAVEMQKDVYMC FIDYSKAFDKV KHSELFEE LKKLDIYRKDLRL LQSLYWNQFAC IRVEGECSEYINIRRGVRQGCVMSPDLFNMYSE IILRELEDENGIQVGGHSITNIRYADDTVLFSES	cd0165 0	321	578

			ATGAG	EKDLQRLN V VVKESERMGLEINYTKTKCLV		
			GCTAA	VSKKSSPKCNLKLKEQIIKQVSSFNYLGSMIAE		
			ATGTC	DARCETEIKRRIALAKSAFSKLGKILTNRTMSM		
			AACAT	RTRIRVLNCYIYPVLMYGSETWTLASDTKKRL		
			ATTGG	ESCEMWFLRRMMKIPWTDKVSNEEVLTRAGV		
			GACTA	QRKMIREMLIRQLKFMGHITRKEGLENLALTG		
			AATGA	KIVGRRSRGRRRILWMDSLQKLLEERG VKELG		
			AACAA	VVLIQKARSRGLWNSMIAKVKRYGT		
			GATGG			
			CTGAA			
			TAACG			
			GCGAG			
			TTTGT			
			GATAG			
			ATGAG			
			TACAG			
			ATTGA			
			TTTAT			
			GCAGG			
			AGGTG			
			AAAAG			
			CATGA			
			AAGAG			

				GTGTG					
				GGGTT					
				ACTGC					
				TGGAT					
				CCAGA					
				TATCT					
				CAAAG					
				AGTGT					
				GTTGG					
				GACAC					
				TGGTG					
				TGTAA					
				GTGAT					
				AGGGT					
				GCTCC					
				TAGTG					
				AAATT					
				ACAGG					
				GACAC					
				CCTTTT					
				AACAT					
				CTCAA					
				TCATA					

				GTGGT					
				TTATG					
				CACCA					
				ACGGC					
				AGATA					
				GTGCT					
				GACGA					
				AGTTA					
				TTGAC					
				ACCTT					
				TTATG					
				ATTCC					
				TTAGA					
				TAAAG					
				CTAAA					
				TCTCA					
				GTGCA					
				AATCG					
				AATGA					
				CATTG					
				TCATT					
				GTCCT					
				GGGAG					

				ACCTG					
				AACGC					
				TAAAG					
				TCGGT					
				CATGG					
				TACAG					
				AGGGC					
				AAGCA					
				GTAGG					
				TCCGC					
				ATGGA					
				ATTGG					
				TGAAA					
				GGAAT					
				GAACG					
				AGGGG					
				ACAGA					
				TGGGT					
				TCAGT					
				GGTGT					
				GAGGG					
				GGAGA					
				AC					

RTE	RTE - 8_S P	Strongyloc entrotus purpuratus	TTCTCT ACTAT AATCA GTGTT TTTATC ACAGC ACTGA ATAGA CATAA TCAAT AAAGT AAAGT AATAT TTGGT TGGTT CAATT AGTAC TCTTAT CATAA GAGTA ATTGT AATCG CCTCA	AAATGAA GGCGGTC ATTGTCG CTTCGCG ACAGACA GCTACTA CTACTA	MRNSMKLHTGSVVFRV NKDRAVRSGLPSTVP SMLGXKQQQQKDDTSGITTM TFGQWNVRTLL DREGTGRPERRTALVAMELEKYRIDIAALSET RFSESGRLDDMGYTFFWSGKPDGVRREAGVG FAIRKEILAKLTEMPKAVNDRIMTMTVPLTKK GNATLISVYAPTMTSPDET KENFYNQLRRTL R DTPRTDKLILMGDFNARVGRDYEKWQSVIGR HGIGKCNTNGELL LALCSEFELL LTNVFKQRE EHKTTWMHPRSKHWHLIDYIITRQRDRMDVH STRTMRGANCLTDHQLLRSKIAFALRSKRRKQ GATKLAKLNTDRLHVNTQRLNLEQDMEKALT NWGVQEGMTVDQTWASLSKV VYETASNTLG KPERKHQDWFDPEDESELLILMKKRDETHQRVL QTRSTRSTTTAYKEACRHLQRYTRKMKTWW EKKAEDLQRSADNNDMKS FYTGLKEVWGPQT RSTVHLKSLDGNTTFSDNKRVLERWSQHFETL LNQPGDIEPAARDRINQRPVVTALDDIPTRKEL NLAIASLGDGKAPGVDGIPSEIWKHGGATLTN SLLMLIQQAWAEGSVPQEWKDANIVTIFKKGD RTQCGNYRGISLLSIAGKVFARILLNRLNAHITP DILPETQCGFRNNRSTVDMIFCLRQLQEK CIEH NRSLYVVFVDFTKAFDTVGR TGLWQLLRKYG CPEKFTGMIESLHTGMMAKVKEAGETSDSFPV	cd0165 0	589	869
-----	----------------------	--------------------------------------	--	--	---	-------------	-----	-----

			TTCTTT	SNGVKQGCVLAPTLFSIFLSAMLEEAFADFEEG			
			TATTA	VYIQRQDADLFNVAHFKAKTKSTQILVRELL			
			AATGT	FADDSALVAHTPEQMQHVIDVFSSASKKFGLQ			
			TAATT	INIKKTEVLFQPGTDHREEEEILVDGTALNQVD			
			GTTTTT	DFTYLGSTISKDGRIDSELVKRMSKASSAFGRL			
			GCCAT	RTRLWNNHHVSTRVKCKIYRAIVLSTLLYGAE			
			GCTTA	SWTLYKSQVKKLHAFMMRHLRAIMRISWKDK			
			GTTGT	VTNKEVLERANLPSMEDLLIRKNLRWTGHVIR			
			AACCA	MPSERLPKQVLFSQLPAGERGIGRPRLRYKDTI			
			CTGCC	KRNLKRRQIETKTWTTAAGQRAVWRTAVK			
			TTGTG				
			GTTCA				
			GGTTC				
			TCTTTA				
			GAATC				
			TAACT				
			AAGGC				
			TAAGG				
			GAGCA				
			AACCC				
			TAACA				
			GAAAA				
			CCATG				

				CCACC					
				GTGAC					
				AGCAT					
				TGAGG					
				GCGAT					
				TGACG					
				ATACG					
				GAACT					
				CCCGA					
				CATCC					
				TTCGA					
				CCAGT					
				CACTG					
				CAGGG					
				TCTAG					
				TGGTT					
				CACAC					
				CCGTC					
				CTGGC					
				TTGTC					
				CTCGC					
				CGGTG					
				TGAAA					

				CTGCC ACCAG TTCCC CGAGA GAGTG CCGTT AGTAA CTTGC AAGAG CTACC GGTAC TTATC CACCT TTGCA AGCAG CCTTC ATT					
Dap hne	Dap hne- 11_S Pu		Strongyloc entrotus purpuratus	TAGTA TGTGA CAATT ATCAT GATGG AAAAT	TAGTATA ACAAAAC CGTGACT TATTTATT CTTTATC ATGCAAC	MNSHAHTPDILVDDINESKYYTPQEFKNSFSY KPDNISLLHINSRSLNKNFEYVENLLHSLNNFA FSVIGISETWLHQNSPDMFNLPNYKLIRADRKG RRGGGVAFYITQNLNFKIRSDIKLTQAESLFIEI NNSNVKNVIIGLIYRPPDSNCELFCDELDFYLH KIGDENKHNFILGDFNINFSPSSDNSNSTNCMQ	cd0165 0	531	787

			TCGAT	TGCTTAC	LMHSYGFFSIINKPTRINPHSSTQIDNIFSNVHN			
			TGTCC	TGCACCT	NKTIGGILCSEVSDHLPFLTCECKLSCSKSFDE			
			AAATG	GCACTTG	LYRKESKRNIELLKQDLFFEGWAEVYNVNDV			
			GGTAA	CACCGCA	NTAYKHFSNKLQYYEYKNIPLGKMSNKRNP			
			AGCAT	CCCTCTT	RNPWITKGLLKSIQTRNRLYKLHLRNPTDTNL			
			TTGAA	GATCACT	NTYKIYRNKLTKLIRTSRRLHFSNKIKKASSNT			
			TCTTG	CAGTTTT	TATWKIIEVMGKKTNPPPTDNITLNGTKIHDP			
			AAAGA	TCGAAGA	SLHANSFNSFFTNIPELASKINRTHAHFKDYL			
			ATTGT	GAGTTAT	SNQNQESLFLNPTNPSEIINITQSLNNSKSHGFD			
			ACGGT	GTCGCCA	RISMSLLKEIMYPLAKPLTHIFNHSLSQGVFPDL			
			TGTCA	ATTCTTTT	LKIAKVNPIKKDDDPHEISNYRPISLLPSISKVLE			
			CTAAT	TACGAGG	KIVHNRLHKFITKHKILNSNQYGFRKNYSTDL			
			CCTAA	CCGCCAT	ALIQIYDKITSAIANKEHVIGIFCDLSKAFDTLN			
			ATTGT	CCATCTC	HAILLSKLSHYGIRGQPLLWFKDYLTNRRQYV			
			CATGA	CAGCCTA	TFNSYDSPELLIQCGVPQGSILGPLLFLLYVNDI			
			TTGTA	ACCGCTC	INTSPLLSFILFADDTNIFYSHKDLNSLNNTLNY			
			TTCCTT	ACGATGG	EINKVSDWFKSNKLSLNTKKTHFMYFKHHSH			
			CCAGG	TGGTCTC	NTETPIYIKIDGMLLEKKTNSKFLGVLIDETLT			
			GGACA	CTGCGTT	WNDHLHQVTMCISKSIGIISKLKFLLPATLFL			
			TGCAG	CTAAATT	LYNSLVLPYITYCNSVWANCSTKINSIFKLQK			
			TAGAA	ACCTCTC	KAIRICTGSHFLAHTDPLFYKLKTLKIFDLNTT			
			CTTTGT	TAATTTA	QVAVIMFKYINNLLPPSFDNMFRFNNSVHSYS			
			ATCAT	TTGAAAA				

			GCAGG	TATTTCA	TRISGNFHLTNPKLSITHKSVRHSGPDIWNDFP			
			AATGT	TATTATG	DNIKSCSTIYSLKATLKRQIIQSYAPHSLN			
			ATTCTT	TTATTGT				
			TACCC	ATGCTTC				
			TCAAA	GTACATG				
			TGGGC	TTTATCG				
			TTGAG	TATGATA				
			ACAAG	TAATTGT				
			TATTC	ACATGAT				
			GTTAT	TTTTTATT				
			CCTTT	TAATTAC				
			GATCT	TTTTGTA				
			TGACG	AACATTT				
			ACTGG	TATTCAT				
			GATTA	TGTATTT				
			TCAAG	GTTATTG				
			ACCGA	TTCGAAT				
			AGTCC	TTTTGTA				
			TCACT	CTTTGAA				
			CTTTAT	TTTGAAT				
			CAATC	TTGTATA				
			ATTAT	TTTTTGA				
			TATCT	ACGCAGA				

				GTATA	AATAAAA				
				GCATT	TCAATCA				
				CCAAA	AACAAAC				
				AATTT	AAACAAA				
				GAGTC					
				TATCT					
				GTTAT					
				ATTGT					
				AATTG					
				ACTGA					
				ACTTTT					
				TTTTGT					
				TTTTTG					
				AGCGG					
				GAATG					
				TCAAA					
				ACCCT					
				CTTTG					
				CCATT					
				TAATA					
				TTATT					
				GTACT					
				CCGAA					

				ATGAC					
				GTTAT					
				TGCTA					
				TTAAG					
				CATCA					
				TCTCA					
				GAGTC					
				TGCTT					
				ACATG					
				CATAC					
				ATTTT					
				CTCCT					
				CCGCA					
				CAACA					
				AATCT					
				CTTTGT					
				TTTGC					
				AAAAC					
				TCCTG					
				TGTCA					
				ATCAC					
				AGAAT					
				CCATT					

				GTTAT					
				AGACT					
				GAACT					
				AGTTT					
				TGATG					
				AGCGG					
				GAATG					
				TCAAA					
				TCCCT					
				CTTTTC					
				CATTT					
				AATAA					
				TCGTA					
				CTCCG					
				ATTTG					
				TCATT					
				ATTGT					
				TATCA					
				AGCAT					
				TTTGTT					
				TCTTCC					
				CAGCC					
				CCTCT					

				ATTTT					
				GCGTG					
				TCAAA					
				AATAC					
				ACCCT					
				TTACT					
				CGGGC					
				GGTCG					
				GGCTC					
				CGCTC					
				TGTTG					
				CTCCT					
				AAGCA					
				ATTAT					
				TGTAA					
				CTGTG					
				ACCAA					
				GATGG					
				GTCTT					
				ATTGT					
				ATTTTC					
				ATTTG					
				TATTA					

				TATTTT					
				GAAGT					
				TACCC					
				GCACA					
				CTTGC					
				ACTTC					
				GTACA					
				CACAC					
				AAACG					
				CACCA					
				GCATC					
				ACACA					
				TGCAG					
				TTGTTC					
				ATCCT					
				CCAAA					
				TTTTCG					
				GTACT					
				TTACA					
				TAATA					
				CGTAT					
				CGTTA					
				A					

CR1	CR1 - 31_S P	Strongyloc entrotus purpuratus	TAAATTC CGGCAAC GCTGT AGTAG TTTACTG TCTAC GCACTAT ATGTG GTATATA TACTG TACACCT ACNCT CTCACCT ACTGA GCACACC TCTGTT TGGACAT CCTGC CCCCATG CAGCA TCGCCAC AGTAA ACACATG TGTAT CACTCGC GTATT ACACATC ATTAG TGCAAGA TAGTA ATCTTCA GTTCTT GCAGTGA CATAA AGGAGGC TTAGA AGATTAT TATTG TTTGAAG TGTA AAGAAGA AGAAGAA	MGRKPKYPCGVCSRAVTTRTTGVQCEGTCLK WFHAECIGLTSTEYVAFDSSSIVWICNHCGAQ NLAHSCSSLQLXPD TNRFGLDPDGQNVISEE NSNSCRPTXDSSNPSSPGSPAYSSSPQNIQRRS KHKPRPRVLRFLNVNCQSIRAKMQPFSILVHQ HNPDIVIGTESWLS DGINSNEVFLDEQYSILRR DRSNDPHGGVFIATKTDLIVTREEELETDCIL WCRIEMAGSKTLRVAAYYRPKEDDEVSLAEL QRSLSRLDRRHLVVL AGDFNLPGWKWDGPYV GPCAYPALHHQFGEIINDHGLTQLVKEPTRQA NTLDLILISNPTLVNSVKVVP GISDHNCPVADID MSPRRYQPKRKVPMYKKADWEGFAELMQH VADYICKSSSASVHDLWQLFKSGIEGGIQTfV PHKMLKQKQDLPWITSGIRKLIYKRDRHLHKQV TRLRKQNMLVPFPLETELQHLKRLIQAESRKA YWSHLESIFTADGPNEYEGMKRFWKfVKNNR TDRCGIHTLKTDEHTITSPQDKAEALNKQFKM AFTNETDVPADLLPETS PYDSMPDIHISTEGIVK MLKKLKVRKAPGPDGIPALVLRKLSRVVAPSL RAIFDVSYRTSEVPDDWRLANVVP IYKKGDKT DPENYRPISLTSIICKMMEHIFASNIMGHARRNS ILRPLQHGFQAKKSCELQLTGfVTDLLNNMED NNQTDIVITDFSKAFDKVGHQRLLKKMEYYG	cd0165 0	632	884
-----	-----------------------	--------------------------------------	---	--	-------------	-----	-----

					GAAATAT A	VRGHNIDWISSFLSNRKQQVLDGHTSSTVEV QSGVPQGSVLGPCLFLFYINDLPNQLSTKVRLF ADDTVLYHTIQTPEDAELLQQDLDQLCGWSSK WMMALNPKKCQVITVSRKKTKYVHRYSLNG EILESVSAAKYLGITLTADLNWTQHVSATCKK ANNTLAFLRRNLRVNSTRLKTTAYQALVRPLL EYGSVWDPHTIRCVNQLEMVQRRARYVLN RYHHTSSVSDMLSQLGWTSLQKRREHQRLTL LYKMHYSLVDFDIHHYIQPATVVRTTRSSRPY CYRVMQSRTEQHIHSFFPRTVRQWNALPLEVV LAPSVEAFQRQLKTGGK			
Ving i	Ving i- 3_S Pur		Strongyloc entrotus purpuratus	AACTA AAGCA ACAAT TTGAA GTCTA TCATC GCGAT CTTAT CGCCA GTAGG CCTAC ATCTC	TAGATGA TGTA TGTACTC CCTGATA TTCTTTGC ACGAACG AACGAA	MMADTSTCVLQLNAEGPYPLKARSDYPHIAK ENWVNVICLQETHCETEEQLHIEGFDLICFIPH KKHGIASYVKQGMKASTTGRSPVNSALQWST FEAMGTHFCNLKPPVDFVEILFPKMPRNSFV CGDFNSRSTAWGYPXTNKNGLIFEWLESQEL VILYDSKDNPTFYSGSXKSWTCPDLSLVSSNIR HTCSRTVLPKFPRSGHCPIIISDIKLQLSNLPPK RWNFRKADWPKFQELTNNKLTSRLPDPSTSNV NDSYSAFCSMHLHQAALSIPIRGRRHPYIPCWD EECDRLYESCAKSKTIEEREAVGNDLLJSLGEK RRIRWEEEVSEINFTHSSRKAWTTINKLTGRSS APKQCPVTANSIASVLVNNGKWQDKTTEAKA	cd0165 0	473	728

				TAGCC CTCTA TCATC AGAAG AGTTT CCTAC GCAGT TCCTG GAAGC ACGTG TCCTTT TCAGG ACAAG CCATC AAG	HSRSVNSEIKNLLKSTPESTSMSEAF TLDEVET ALKSIKSGKAPGNDNIHTEFXXNMSRKATSWL QAFLSTCMTTSTIPSKWKTAKVIAILKPKKPAD EPKSYRPISLLSHIYKLLERLLLARIVDTVEEKL PTTQAGFRKGKSTTDQVVRLVNNIETAFQKKQ KFGAVLVDLTAAFDTVWHRGLYLKTLQTIPD VRIVKFIMLLVQDRSFVLET SNGECSRKRKL RN GVPQGSVLAPILFNIYVSDMPKGECMQYSYAD DTALGFDNHSFEAIKTSLEKDLKMLAEFFNRW HLKMSIPKTVSSVFHLANRCANRELNISLDGSR LNFVSTPKYLGVTLD RSLTFHEHAKATAQKTQ ARVSLKKLAGTGWGANFTTLRTSTLALAFST AEYAAPAWSHSTHVHKVDVVLNDAMRLISGT LTATPVSNLPILSGIPPAQLRRDAQYTKLAQKH DRDKCLVPAPSAFEGQRLPRRH FATQAELPL EKPPSTLATWIAKRWSDLQWNSSTTNLKEFIP KPCAKPSGSDLPRKAWVNLNRIRTGVGRTQYF LHKIGAXSSPNCACGEIQTMDHVIDDCPIYKSP HGASGLRELDDETICWLNKDLPI			
CR1	CR1 - 24_S P		Strongyloc entrotus purpuratus	TAGAATA CCCGACG CAGAACT GAGCCAG	TDTDRGKQSKFKRKL RDRKLQIFYTNCD SLLN KRNELLLLLQSHD VDLVLLTEVFPKNHQNPLT HRDFILEGYDLHCNLEDGNRGVAIYIARDLYH LVGECDLQHSFHESVWILITLRGSDKLLIGCMY	cd0165 0	509	757

					ACACTTA	RSPSLDNHEELKRAFTQVTSINSSHVLIAGDFN			
					CCAAGAA	YPNIDWKLGRARSHTEQEFIE TVADCYLHQHI			
					GACACCC	GEHTRYREGQRSSMLDLVFTNEDGMINHIDYV			
					AACTACG	PGIGRSDHIGILFGLTVYTPDTHPTPHLQFYRG			
					AGAAGGA	QYDDMCKELDQIEWESELENKSVEDAWEFFL			
					AGAAAAT	VTLTEKMEKFIPKSKPKHRRKKQRWMTREAL			
					GAAATTG	RKQSQKHRAWKEAQRTKHPVDIMRARRLAK			
					AACTGGA	QLQTLTKQLCRNFEKDLAMNVKSNPKAFWRVY			
					AGCCCGA	TNSKLKTSCRLGELRTPDGERISTDKGKADAL			
					CTGCTGA	NTFFATTFTDEDKGHIPKLPPRHKGETLTSISIS			
					GAACACA	EDDVYKKLVHLSKYKCAGPDGIHPCVLKETA			
					GCATTCT	ASVSEPLSIIYNASLKEGKLPSAWKVSHVVPVH			
					TCCAGAA	KKASKMDVKNYRPISLTSVPCKILESIRDEIV			
					TCTCGGG	AHLMENDLLCDEQHGFVPGRSCVTQLLTVME			
					TAAAGGT	AWTEMLDKGTPIDVVYLDFKKA FDSVPHLRL			
					AAAGGTA	LSKLETYGICGQIKEWIADFLQGRCQVRVND			
					A	KLSDWTKVRSGVPQGSVLGPTMFVLCINDLPD			
						GLQGSVKIYADDTKLYGKASTKDDLCSIQSDL			
						DALMEWSNTWQMHFNTQKCKVLHIGPGNVK			
						HIYALGDDVLEATSEEKDLGILIDDELKFHNHV			
						AYAVKKANGVLSCIRRSFECRDRSTIPKLYKG			
						LVRPLIDYGDSIWYPRFKTDSKTIERVQKRATK			
						MIYDIRNRPYIERLRALNLP SLEYRRRRRGDMIQ			

						VFKIIHGLERIPVDTFFEHSTSNTRGHSLKIVKQ RCSTKIRQDAFSQRVINDWNTLP EEVVTATTV NAFKARLDRVWKDKRFCSPYE			
RTE	RTE - 16_S P		Strongyloc entrotus purpuratus	GACCA TATTTT TGTTTT GAAAT CCCTG CTTGA CATTT ACTTG TCTAA ACGGA AGAGA ATTTA CTGCG CTTTTA TAGAC TATAG AAAAG CTTTTG AT	TACAGTC TTTCTTTA GGTCAAG TATAGTT TTGATGG TTTGTTTG TTCTTTTG TTCATTTT TTTTTTTT TTGTTAG TATACTT TTTGTTTT GTTTGTC TTATTAC ATTATAT TCCATTA TGTTTTTG TTTGTC AATGGAC GATTATG	MVDRVSLWHKLLRSGVDGKVFTVIKNMYSQ AKSCVIVDGMSEAFQCDVGVRQGENLSPLL AIYLNDFEDFLSKHFKGLKFLVENVCADILDKE VNKDLFLQLYTLLYADDTLILADTAEEELQVAL NALSQYCKIWKLSVNAMKTKVLIFSRGKVRKF PVFRFGEDEIDVVDDFVYLG VQVNYNGSFKK AISKQVSQARKALFSMSLKVQRLKLPLDIQCEL FDQLVLPILY GCEVWGYEDLSQIEIFHRKFLRS ALCVNKFTPDCMIYGETGRAWLANSIKHRMV SFWLRIVKGSPSKYSVKFYFLQRLISQNVDMST SFKWLS CIEKILNHAGLGIVWLKQGEGFSNSWI LSTLKLRLNDISKQDVMATIWTNSVCLNYRIF KEDCLFERYLVLLSKKDRITLSRFRCSRHLPI NNNRFKEAVNEISIECPCCDMHESGDEFHYVF VCPFFRKERELYLGRKVIVKRPCTSHMLNLFSC SGIGQLKKLVFLRIIMSYFARLYTCKSEQSNH MSPIKRNVTTRSGRLIKOPTILDL	NA	NA	NA

					GCGTATT				
					TTGTTAG				
					TCTGCAT				
					ATATCTT				
					AACAATC				
					CAATCAC				
					AACTCAC				
					TTACATG				
					TACAAAA				
					ATGAAAC				
					TTATAAT				
					TTCTTTTT				
					TAGTTAA				
					GACAATA				
					ATGTATT				
					TAAAAAA				
					AAAAAAT				
					TGTAAAA				
					TTATTTTT				
					TGTTTTTA				
					TCTTCTT				
					TTTCTTCG				
					TGTTTGT				

					ATTTAGA				
					TTGTTCA				
					TGTGGTA				
					TTTTCCA				
					CTTTTCAT				
					GTATGCT				
					TGCCATA				
					TTGTATA				
					ATATTGT				
					ATTTATG				
					TTTAGCA				
					TAGGCTA				
					GGTTTAG				
					ATCATTC				
					CTATTTA				
					TCGTTAT				
					ACTTTTA				
					TATTGTT				
					GTTTTGTT				
					GTATTCT				
					GTATTCT				
					TTGTATT				
					ATTAAGC				

					TATGTTT GTATTCT TTATTAT GTATCAT TATCATG ACATGTA CATGTGT TGTACTC TCATAAC CCAGAAG GGGGTCG AATGAGT CAATAAA TTCTATTC TATTCTA TTC				
CR1	CR1 - 5_S P		Strongyloc entrotus purpuratus		TGAACCT TGTATAA ACTCTAA TAGAAGG GGCAAAC AGACTTT AGTATGA	PSQRVNIELCCEITVSADKLAPNMLDNFGIIV LSVIASLPHFIAESRPALLAGVTMEFQLVQNSID TERYSDQYRYTKDELLSLRNTSHSRLPDNVFA RIKELGIDKRRGCRAAGSHKWRKIAVVEGRRRR TPKPPGVRAPQLIPIKTITTVSTDRQSQGNLPTI YMCNPRSLNNKLDEFSVVVHELCCDVCVVTE TWFRADQPESDFAVDDYLVFAKSRDSRRGGG	cd0165 0	633	885

				CTGATTT	VAIYAKPESMPKTVNFLPPHPHHEVIWLQLRP			
				TAAATGA	KRLPRDISCLYVAAVYSPPDNDCVNDLIQYLV			
				TATTTGC	NAIDDIKTRHGDAGVVILGDFNRADTSELTSTC			
				ATGTTTTT	NLRQVVRKATRGTAILDLILTDIHQWYSEPEIL			
				AATTACT	SPLGLGGDHNAVLWKPCLNKTNSKNKAVRKF			
				CACTTTG	VRPMKDSAIREFGQWISQHTWDDVMASSSAQ			
				AAATTAA	EKCDKLYASLQNAMDEYLPKSVKMHPRDK			
				AATAATG	PWVTPYLKSLIGQRQKAFSNDRTEEWKRLRN			
				TTTTTTAA	KVQNKKTELKKNFYKDRVLKLKRENPSWYK			
				AACATAT	HLKIMTSSKTASNIIPESPELDDLAIADC VNDH			
				AGTTTGG	FANVSQDIPELDLQSLPAYLPTPKPQLTVQPHE			
				GATTATT	VYKELNRIKPGKSSGPDGIPARIVKEFAVELAS			
				ACTATTG	PLCEVLNSSYHSGSVPSQWKGAIIPKEKAA			
				CCTATTT	SSIDKQRPISLTDHF AKIAEVFIYRKLLADVRHQ			
				CCAATGC	LDPKQFGARAGRSTTHCLVEIYNFLCKRAELPS			
				ATGAAAC	NVTSIVSTDFSRAFD RVDHTILIRKLITLEVHPW			
				CTTCGGG	VIVWICSFLTDRQNQVRYHGELSSTRSMTAGV			
				CGTAGAC	PQGTRLGPLLFLVLIDDALRESEVERWKYVDD			
				GCGGAGG	MSLASAEPSNQHGQFMQNVLVNLEEWCNQN			
				AGTGGTC	NVRLNPKKC NIMRVTF SKTPLPDRDFSLCNDV			
				TATTGTT	LQEVSAVKLLGVILQRDLKWN DHVSVTISRAS			
				AGTGAGG	SRLYMLRRWKEFHLPAADLVT VYVGFVRPLL			
				AGGATGT	EYACPVWHSSLTCAQSAALERIQKRALRLIFPG			

				CTGTTGT	TSYNTSLDLARIVDLETRRKTL	CINFATKASRS		
				TGTTTTCT	EVGSDWFTENNPT	RNL	RKNKRYREPR	CRTNR
				TTGTTCA	YQNSPIPFMTKLLN	FN		
				TCCCCTC				
				ACTTCAA				
				TAGATAA				
				CCTACCG				
				TGCCTGC				
				GCCTGCC				
				AACATGG				
				TTATGTTT				
				TAGGATC				
				TTGTTTAT				
				TACATCC				
				CAGCCTT				
				TTTAAAT				
				AGATTTA				
				ATTCACC				
				TTAGCTA				
				AACTGGA				
				TCATGTT				
				TTTTTTTA				
				CTCTTTA				

					CTAATTA				
					ACTGGAT				
					AAATCAC				
					TCGAGCA				
					TTATAAT				
					ACATGTA				
					CCATATT				
					TAAAGCA				
					TGACCTA				
					TAGTATT				
					TATTCAT				
					TCACAAA				
					AAATGTG				
					ATTTCAT				
					AGAAAAT				
					TGAACGA				
					TTAATGC				
					TAAACAT				
					AGATTTA				
					TATTTTGT				
					TAATTTT				
					CCACACG				
					CAACACT				

					ACTATTT				
					TCAGAGT				
					TGAAATG				
					TTCTAAA				
					CCAGTTA				
					ATTGCCA				
					TGTACAG				
					TATTACG				
					CATGTAT				
					TTCTTGTT				
					AAGTATG				
					TAATGTT				
					TTAATCA				
					TGTAAAA				
					GTTATCA				
					TTTTCTTA				
					TACTCCC				
					TCCCCCT				
					TTCTTTCT				
					CTCCTCC				
					GCCATAC				
					CTGTCCT				
					TTTGTATT				

					CATTATT				
					TCACTGT				
					GAATTGT				
					GCTGTAT				
					ATTGTAT				
					GCATGGG				
					TGTTGAT				
					CCGTTGC				
					GTCGTGC				
					TATTCAT				
					CAGGGGG				
					GTATATG				
					ATCTTAC				
					TGTTTCG				
					TGCCCCA				
					CTGGTGA				
					CTGGGGT				
					GGCGCAG				
					CGGTTCG				
					ACGCCGG				
					TGCCAAG				
					AGTTGCA				
					TTATTAT				

					CTATAGT				
					AGTTTAT				
					TATTAGT				
					AAGTGCA				
					ATTTTAC				
					CTTGTA				
					TAATTAA				
					TTCTGTA				
					ATCCATT				
					GCGCAAT				
					TTAGTTTT				
					TAGCTAC				
					GAAGCAT				
					GGAATTA				
					ATGAACA				
					TTTATTGT				
					TTACATG				
					TTTGATT				
					GATTTGA				
					ATAAACC				
					ATTTATC				
					TATCTAT				

					CTATCTA TCTA				
Реп отра нспо зон, не соде ржа щий LTR	BRI DGE 1_T N		Tetraodon nigroviridi s		TGATGAA GCCTTTT GGATAGA AGGCGAA ACGTTTT CCAAAAA GAAAATT GAAAAGT CCAGATG AACACAG AATCTTT TCTCGGA TAACTAT GACCTGG ATGACTG AGAACCT ACACAGA CA	LTKTEKDLLSRGLNSAVTPEELPVVDLITATET AIRNNKLPETEAQIRLKVSAALANAKAPASNI TAQEKRALASLAKDKDITILPADKGRCTVVLN TTDYDSKILSLLGDSNTYEKLKRDPTSTYKKK VIDLLQKLEKEQAIDKPQYYRLYPGEATPCFY GLPKIHKEGTPLRPVSSINSVTYNISKFLASILA PMVGNTPHHIKNSQDFAEKVSNLQLESGETMV SFDVTSLFTCIPTSEATEAIHKRLLLDKNLQERT TLSPAQICTMLDLCLNTTYFQFRESFYRQKHG CAMGSPVSPIVANLYMEEVEKRALLSFTGAAP SRWFRYVDDTWVKIQTKEVEKFTTHLNQTD FVKFTREDVKGDSLAFDCEIRIEEDRNLSIEIY RKPTHTDQYLLFDSHHPLEHKLGVIKTLQHRA TKVPTTSHGRKKEQDHLTTALKTCGYPDWAF TKTRKKRDPSNEEEKDKRHSVSIPYMAGVSEK FRRILQKHNPVQFKPSNLTQRKLVHPKDKTPR HKQSNVVYAVQCQQECRDLYIGETKQPLHKR MAQHRRATTSGQDSAVHLHLKESGHSFEDNQ VRVLAREDWFERGVKEAIVKMEKPSLNRG GGLRHLSPTYNAVLHSFHKQKPKHTHSFTRA GVTSPGDPAVKRKTSQTETRSTTLPTTLPTTPE	cd0030 4	296	379

						TSIQTTGILMQRLSVDISMPMVQKGYIFKLCTE PDQN			
R2	R2- 1_T Gut	.	Tinamus guttatus	CTGGG GACCG TGGTT ACAAC CCGGG CTTAG CTGCA GAGAC AGTAC CTCCC CGTGG TTCCC GCCGG ACCCC GTAAC ATCGG GTGAC TGAAT CTGTC TCTGC CCCGG	TAGGGGG CTTGGCA TTTCTCAT TGCCTGC TCCTGAA AGGATAT GGGTCCT GCGTCGC GTGGTAG GCAGACC CATTCGT CCGAGTA GGGGGCT TGGCAGT NTCCATT GCCTGTG CCCGAAA GGACGTG GGTCATC TGGTCTG TCTGCCT	MGSWIVNFVSVATQTGEFPVDTARRAPVPVTS YPESECHXPLPLTFCNSDVTIWGGVRPEPVDCL GDLPEXYDALPGVAGPREXVGGSPPGEGVRSP GIASPSGTAVQHDFGSPILVPGAEEAEVSTPVV KVPQDHPACPCCGTRVVKVTALSEHLRRAHG RKRVLFCQSRCGRMNEKHHSIACHFPKCRGPP VEEGPLGAPEWCCEECGQKFNTKSGLSQHKRS VHPLTRNVERIEAARPKGKGKRGAAHKGCWTE AEVAQLIELEGRFKNQRFINKLIAEHLPSKSAK QISDKRRQLAAATKTSSPEKRVTSSTSGESSPE VEKVEGIKREYRRRVGEWLCAGSLXDQTSFQ KILEDVESGSEIVTGPLEELASFARGKLAAARV RHRKHPAEAVPAREEQRWMKRRVGRRGLY LRFQRLFALDRRKLAGIILDDVESIKCPLPMEE VADVFRRRWEEVAPFTGSGSFRSLGKADNGA FKPMISAKEVMKNVXEMSRRSAXGPDGLSLR DLMKIDPQGSMAELFNLWLLAGRVPDQVKA GRTVLIPKSADPGKIGNIDNWRPITIGSVXLRM FSRILSARLRRACPINRRQRGFIAAPGCSENKLK LQALIKSAKRDRHTLGVVVFVDLAKAFDSVNH QHIFQVLVQKGVDGHIIDILRDLYTNAGTYLES	cd0165 0	542	806

			GAGTA	ACACCTC	GSQRSGFIKILRGVKQGDPLSPILFNLALDPLLC			
			GTTCC	TCTAGAC	RLEDRLGYKYGDQQIXSLAFADDLALLSDS			
			TCCTT	TTGTAAC	WEGMQQSIRVVEEFCQRTGLRVQAPKCHGFLI			
			GCCCT	ATCTAGT	RPTKESYTINDCDPWTIADMQLDMIDPGSSEK			
			ATTGA	CTGTCAA	YLGLGIDPWIGLSRPELSEVLTRWVKNIGGAPL			
			CCAGC	CAAGATC	KPLQKVDILRSYALPRLLFIADHAGLSATCLHS			
			GGTCG	AAAATTC	LDLSIRSAVKGWLHLPPSTCDAIYVSYKDGGL			
			CCGGC	TTCACAC	GLPRLASLIPNVQARRLVRIAQSEDDVIRSVVL			
			TGCTC	AGACGAC	QEGIQEEIRKVWISAGGRPEKVPSVTGEFPVME			
			AATAG	CGAGCTT	AQAADSEALSEWERRAPRTIYPIPCWKREME			
			TATTCT	GCTCAGT	NWTNLKSQGHGIRNFENDRISNDWLLHYGRIP			
			AGGCG	CTTCCTG	HRKLITAIQLRANVYPTREFLARGLGEGAPRG			
			TGAAA	TACCCGC	CRHCPAEWESCSHIIGYCPAVQEARIKRHNDIC			
			TATAG	AGAATTT	GVLAEEARKLGWVIFIEPHLRDNTNELFKPDL			
			CGATA	TGCTCTT	VLVKGSCAKVVDVTIRYESGLTTLSDAAAEKA			
			GTCCT	GCTCTCC	RKYQHLAGEVRALTSATTVDLGFPIGARGK			
			AGTGG	TTTGGCT	WYVGNNGLLSDLGFSTSRVVRIARALSKKALL			
			TTGTCT	GTGTCCT	SSVDIIHIFASRARQAQTSE			
			TACTG	GGACGTG				
			GGCCA	GGACTAT				
			TAGCC	TCCATCT				
			CCTTG	CGTCCCA				
			CTTCA	AATGCCG				

				GGGGT CATTC GCGAA GTCTC TCAGG AGAAC TGGGG GTGGT GTTCTT CTGGG TATAG CTAAA CCCCC TAGAC TGTGT CCGAT CC	CGTCCAA TTATACC GGATTTG ACAAAGC GGACGGC CCGCTTT ATAAGCC GGAAAAG GTGCCTT GTAAAAT TGCAAGG TTCATTA AATAG				
CR1	CR1 - 3_T Ca	.	Tribolium castaneum	ATAAT CTTTA ACAAA AATTT GCTTT AAAAT	AA	MSVCYKCKSSVVNSIACLNCKKAYHPSCLRAF KNYSAIGQDLTSSISACPCCDKLKSQNPTSPTL NISTNTIFEDLGILKQETDLIKKTLDDLKTFITES PTTVCTFDKIDEYVRKSFDDLESKLISLLKDTL NREIEKLKNSIKNEVIAEISTNLNRTNDNIKTAN SSDLTLKSTNSKNVSTQSVSKTNRYKVRIFAD	cd0165 0	915	1166

			AAAAT	QKGRHCIDLLETYPVSKYVIQSVVKPYSSLQE		
			GTTTG	LLKTENNNEQLGNNDFFVIYAGNHEAQKGKIP		
			AAAAA	DENTLNELINHKFLNTNIILIGCPYFVNRPIINN		
			CCGAA	FVYELNCNFYNVAKNNKNIYYLDSNDKIFDQF		
			ATTAA	KTDNFVINKAVKRCIIKNLGLITECQDHTNRS		
			ACTAA	AVEQNPVATFLDTSLPSVDKSDSQSFLSVFLNR		
			ATTAA	FVCLQPSHSKISISINKKTVNSFHGIYQNVRLK		
			TTGTG	SKSVTFFNSFLSTSDEFDLIVITESWLDDTVLDG		
			TCTAA	ELFSDNYVFRDRDLRHLGIHRGGGVLI AVN		
			CACTT	NKYNCRQLDISNISRRSSPAIDLVAIKINVTSYQ		
			TAGTA	HLFVFVVIYPPSISFSDFDNFFDLLNMFDEINND		
			GAGGA	NVLFLGDFNTPLFSGFSKNNFSISLKHFA SFHN		
			AAAAG	FNQFNNIRNNRNILDLIFSNMDCAVSKSSFSF		
			GAGAA	VPEDPFHPSLEFTLDIVAKTPDKFYMNRDHVA		
			GACAT	FNFKKANFDCLYTTLQVVDWSELEKFSDVNE		
			AAAAG	ACRVFYETLYNCFELSVPKLFCRKNYRTFPTW		
			TCACC	FNTDIINCLRRKSNILKKYRRTKNMHFYEEFKK		
			AAAGT	LRSLSKSLITTAYSQYIGHMENTISKDPKKFWS		
			CTTCA	FVQAKKRNSRIPGVL YHNGNAISNPNDIVNSFA		
			GAGTC	KFFKSVFSDMTLTSDSPSMMTNLNSFITVNEIN		
			GTTTTT	EHEIIRVLKRSKDTLTAGCDGIPSFLLRDCAHIF		
			AGTAC	SKPLVYIFNLILQTSTFPNIWKIAHVCPVFKTGD		
			TAAAT	ITQIQNYRPISILCNFSKVFESIYDRIYSSVKRFI		

			TTGCTT	SPSQHGFVEKRSTVTNLACFTQFISDALDRKGQ		
			ATATG	VDVIYMDFKKAFDQIDHLLLLLKLDQYGFSGS		
			CAAGC	LLSLIKSYLANREQRVKYRNYISDSYVATSGVP		
			GCTTA	QGSNLGPLLFLLFINDICGSLSTCAKLLFADDL		
			ATTAA	KIYTEIKSIEDCLTLQNNINAVVKWCNENRLYL		
			GGTAA	NPSKCNVMSYTKKREFLEFVYDISSVTLHRTFII		
			CAGTT	KDLGVIFDTELAFSEHIRDVTARAIKSYGFIYR		
			TTTTG	NCRDFKNLSVMKTLFFSLVRSKLEYGALIWHP		
			GTTTCT	IYKIHIDQLENIQRRFLKFLVFIIDGNYPPIRGYDQ		
			TGCTT	NLLLNRFGLQSLQFRRICIIKFLYNLIN		
			ATATC			
			TCGAA			
			AACAG			
			TTACT			
			CCTAT			
			CAATT			
			TTTATC			
			TTTTTA			
			CTAAA			
			ATTAA			
			AGCTG			
			ATAAA			
			ATTTC			

				CTACA					
				AAATA					
				GTTAT					
				TTGCA					
				TTTTTC					
				ATGTA					
				GGACT					
				AACCA					
				TAAGC					
				GAGAT					
				ATAAG					
				TTTGA					
				AAATA					
				ATTAA					
				TATTT					
				AAAAA					
				AAGTG					
				CTTTA					
				ACAGA					
				AAATG					
				CCTTG					
				TGATT					
				TAAAA					

				TACAC					
				TTAAT					
				TTATT					
				GGGTC					
				CAATA					
				CTTTAT					
				TATTT					
				ACTTT					
				ATTTTT					
				TTGGT					
				GCTAT					
				GCTAG					
				GCTTG					
				TGCTA					
				ATGTG					
				CTATA					
				AATAA					
				GGTGT					
				TACAA					
				TTTTGT					
				GCTGT					
				AAATA					
				AGGTG					

				TTGTG					
				CATTTT					
				GTGCT					
				ATAAA					
				CGTCA					
				GTTTTT					
				TAATT					
				TTTCGT					
				TAATT					
				ATTTG					
				TGGTA					
				ATAAC					
				TTAAT					
				AAGTG					
				ACTAA					
				GGTTT					
				AAGGT					
				TATGT					
				TAAAC					
				TGTAG					
				TCAAT					
				AGCGT					
				TTTGC					

				ATTTCT				
				TTTGG				
				ACTTT				
				CAAAC				
				ATTTA				
				TACTTT				
				ACACA				
				AAGAG				
				TTATA				
				CTGAT				
				ATAAA				
				TGACG				
				TGATC				
				AATGC				
				TCGTA				
				CTATA				
				ACCAA				
				TAATC				
				CAATT				
				TTGTG				
				CTACT				
				TTTTAT				
				GAACA				

				TTAGT					
				TTTTAC					
				ATGTA					
				AATTT					
				CATAT					
				CAGAG					
				GCGTA					
				CGCTT					
				CTAGA					
				ATTGC					
				TGATA					
				ATAAT					
				TATTA					
				TTCTAT					
				CCTCT					
				GTTCC					
				TTTTGT					
				GTCCT					
				TTCCG					
				TGGCC					
				TACCT					
				GTGCT					
				TGTTTT					

				TTTATT TTGAG TCACT ACGAA CGGCC GCGTA GTGTG AAGCG TGTGT CCTTG TCTTTA TTCCCT TGCCT GATTG TATTT ATCAA CTCAC TCATA					
CR1	CR1 - 1_T Ca		Tribolium castaneum	ATTCT AAGCA TTGTA AAAGT ATTTTT	CTTATGT CTTAAAA TTATTTTT ATGTTAT TGTTATTT	MALRAARKNYYGQRIQNADNKS KCAWNIVN KICGRSKKCCEIQIEGDPKQIVNDINQYFINSAP FVRSKSNNTISSNAKTMYXYPVTSQEIITHIQN LKNKKSAGFDEISNNIIKYCNEEISAPLSFVIQN SLDYGIFPEQLKRAQVLPIYKKGDPTLISNYRPI	cd0165 0	148	408

			TATCA	CTGTGTA	SLSSFSKIFETVISSRITNFLMKFQLLNSSQHG			
			TATAA	TAATACT	YIKGKSVDSAIFSFTNALLDKLEEKKVVLGIFL			
			TATGT	GACGTTT	DFSKAFDCLDHNILISKLKKEYGIRGIMLKLITSY			
			ATGAA	TTCGTAT	LNNRYQQVTITKNSTVYTSNKLLIKQGVPQGS			
			ATGGT	CAATTTT	LGPLFFVIYINDLLNILTVSDNNSAVNFADDTNI			
			TCAAG	GTTACTA	LIWGENMTDTMKEAKICMNKVINWTKENKLV			
			AAAAC	CATCTGT	LNFEKTYGIHFAICNHNTEKPNKIVISSKEIELI			
			ACAAG	ATAAATT	EKTNFLGICIDHKL SWTNHIDNLLKKLNSVHFT			
			AATAA	GATAGAA	IRVLKKYINIEQLRIIFYANFQSLLSYGIIFWGQS			
			CCGCA	ATAAATT	GKISKVFVAQKQVMRTMFNLKYNETCRNIFKS			
			CGCAG	ATTATTA	NNFLT VTGLYIYRLLLFFFKQNLLFNVTNDHK			
			TAAAA	TTATTA	YDTRKVDLEYPRHKLTLTEKNSYYAGIKLYNS			
			CAAGA		LPVHLRTPCNLNLFKKKIFDYILDLEPYTLNEF			
			ATAGA		YNRQVF			
			CAATG					
			TTTTTA					
			CAAAT					
			ATGAC					
			CTTTA					
			ATAAC					
			TTTAA					
			AATAG					
			AAGTT					

				CTTGA					
				ACCAC					
				ACATA					
				TCGGA					
				TCACA					
				TGGGT					
				ATACA					
				GCTAA					
				CTACG					
				AGCAT					
				AAAAA					
				ATAAT					
				ACTAA					
				TAATA					
				CTGAT					
				GCAGT					
				AAAAA					
				ACGTT					
				AGAAT					
				TTTAA					
				ATGTA					
				TATGC					
				ATGTA					

				ATCGC				
				AATGA				
				TATCA				
				ATGCT				
				CAGTG				
				GAGCA				
				ATTTT				
				ATAAA				
				TATTTT				
				CAATG				
				ATAAA				
				TTTAA				
				CGAAT				
				GTTTTC				
				CTCTA				
				AAACA				
				AATAA				
				AAGAA				
				AAAAA				
				AACAA				
				GAAAA				
				TATTT				
				ATAAA				

				AATGA					
				TCCTA					
				AAGTA					
				AAATA					
				TTATA					
				AATAT					
				TATCT					
				AGATT					
				TGTTA					
				TCTGT					
				GATAA					
				GTAAG					
				TGTGA					
				CGATA					
				AATAT					
				GTTAT					
				TCTTTA					
				TAAGT					
				ATTTG					
				AAAAA					
				AAATA					
				CGAC					

NeS L	NeS L- 1_T V	Trichomonas vaginalis	GGGTG AGTAG TCTAG TGGT	TAAGAAG AGATAAG ACGAGTG AGAAGAA CAGAAGC ATAGTAG GATTGGC AGAGCTT AAGCGAT GTCACTC GGTACGA AACGTGT ACCAAAC ACCGGAT TCCGTGC TAGGAAT CACAAGC CAAAATA AAAGAGA CACCACG AAAATTA CTCACCC TCCCTCA	MIPVLGTGGPEKLPLQSYVYCGNTAITDSFTPT AKTILKPEEQNLDIVLKNIAALNPENYSDLIRSL SKMEFRLDYPKEIENYWISEKLFSQSIASLPISL LVASMFSPEDRDLSTEPFHCNADGCNFHCDNC ERMVEHIREHHNTDPMINTFETTEDTFRRITAI KIDKTGIEELNPLKYRCSYCDELFTAEDHAIH MISHLTEKLSPDISFFFNDILRLYKTIDKPTVQN LFPETQVAIFDTLEETNRFRLLIVGREAIETIEEAF PPSPPGTDRKPSIIITDTCQLRFVPCMDEPPKGD LGILTLLLRFSAHNIPIKSLNNKELIADKDIDY SPDFVEGALANAEHDTTNSQNNNGRYINSAE KLTEFLIQCEDYLTNIKTLEDLERFYTTIKDYR VNKEVIAEDTPIFVYFLVEEGKLPKPLRCPL SYEGHEDKAFESLRKLCDFHFKGEIAKTSFDPK VHTIDIWVEFLAQAYGTGTFVYKDENGIDLD THVFKCPYADCSYTNNDRSKLMDHMKTKKH AKNVYIERYGFFWGIVIEGVNRPKGIVYPTLKD IKEHACRKCPEAGCNTYVTELSDIKEHLKKKH KSTTAGVDGEIAHTDATYCWITKEELDALHAE RARERAEQVDNTPVQQIINADNNEENNENQED NGNNEEADALDPPNNTTETEDAEVHAVIINPP ATEEEEVAIIEARRNIPELQQAEEERGCVTPKM TSLVRLKLLKGGGELFNKKLTPLATRYAATGN	cd0165 0	1049	1310
----------	-----------------------	--------------------------	---------------------------------	---	--	-------------	------	------

				AACAGAT	TEADKIKVDYLTCLKNAALREMIYTNNHSESK			
				AATAATA	FMTAENGEDTAPPPRISEDTRDRIQKAANEIKG			
				TTAACCT	TLIKVVKHISHARCLKDSTRDDEHNKFVEMIA			
				CCCATCC	KIKNDLRDNKFEQYNIEEIFQGPISDQSILNIVN			
				ATCAGTC	TEDNNEFIKKMDYINRILGTPQDASPYARKKL			
				CGTATGG	QACFADNPTKTLRNIILADKVPQQSLKPSEYLD			
				TCTGATA	YYGPQWANEAEGYENFLHHDYALPERYGQVF			
				ACAGACT	ANDFLDFMTNESKIIEVIRNKNHLSAHGLDGIP			
				AGCACCA	NSVYMLFPVSAAKFLSILFRSIIISGHIPDCWKL			
				CATCCAT	SKTVMLFKKDDPSLAKNWRPIGITSCTYRIFMT			
				GATACAC	LVNKALQMIPMFHAMQKGFVRGATLSEHIAV			
				TCATTGG	ANEVLCQSTRTQSEMFQTAIDFTNAFGTVPHQ			
				AGTGAAA	LIFDSLEAKKVPDSIINLLKDL YKGARTAIYTR			
				ACCACCA	HAHSEIVPVRRGVIQGCPLSPILFNCCLDPLLYA			
				ACAACAA	VQRRHFEDGYRFQDKAGQYSIAIQAYADDVL			
				ATCCACC	VISPTHEGMQRILNTVDEFQKIAKLKVAPQKC			
				TAGACCA	VTLAKTSTAIQPFRIGPDEIPIKTSMDNITYLGIP			
				AATCCTG	ISGTKTSRFAAATGILEKVKAQIRVVFASHLAL			
				CCCCACC	SQKIIALRVFILPQLDFYMFHNVFRVNDLKATD			
				TCCACCC	QMIRGLIDKEAPTSNIPVSFFYMPKNKGGFGLV			
				AAGTAGC	KLELRQPQLVLTKFARLWLSQQAETKAFFHT			
				TCGCTTC	MAQEEKSFRKVVEDQENGFLGIKMENGKIVQ			
				GCTCGCT	KNERSKRTNCFITQAAKAADKLEVRFKEWDK			

					CACCTAA AACTTTG CTCGCTC GCTTCGC TCGCTCG TCTTAAC CCTTTCC GAATAAA CACTTAC AATTCCC GGCTCGC CCCATTT TTT	GGIQVRGVGENATDWYRSKHIGQISPLIGRVIQ QRQYEEFKKDETHSHTFCEPAALAE SHDIMKR PQAVPNNLYSAAIALRTNTAPTPANMHFHNPE VLANCPLCGCQSC TLFHTLNMCRNRFSLYKW RHNIICDDIYQFIHDHYPGVTIKCSARITSDGYQ TTGPELDDTVKD LLPDLVVYDEANKMIKII EVT CPYGTDNNVGN SLDAAYDKKVNKYKSLAEQT ERLFNWTTTLSIIVVSSLGVIPLRTKLDALRISP ADHIQLLKRLSMHAIAASACIVFEKVPEFFGMR CRPLPGRVTAPNAAIPPNNNNENNNDTDHGQEN QQATSEEQPTNNGNAQEDNGQGEQINNSTEQT ISVDQIIEEDAENNAIEQALDQPDEDEFLN			
Ingi	Ingi	.	Trypanoso ma brucei	CCCTG GCG	TAATAAA CGAATAA CAAAAAT CAAAACA AAAGGAT TGCTAAT TGACATC TTTTGGA GAGTCCG GGGTGGG	MPATSTWCQGPVPRIIGGSQEPA AFLSWG TLL CSGYGIIQHRDQQRLAGTPFFICRSLGTCQRAIS SIIRTKMLLSGDVEENPGPSLRGMQWNCAGLS QGKRLALHKTLVDERIAFCLLSETRMTPGEAA CFSVAGYQHHGIARNCKGGGV SILVREDLPVE TGMAVVGRIEQVHATIHLARGTALT VTSAYIP PKHTFTATDLDTLLTTDGAQLIGADANAHALA WDRASPPNTKGETLTQWCIDNQFLVCNTGECT RYARHHGESTPDVTL SRNCTVYTWTSLYSPDS DHHHIFFDVIVGDDTDALSCPRLRKPMYAWLK	cd0165 0	550	805

				GGGCTTC	ADWRNFRLKVDELCKRIGREKNVNTLEQKLSS		
				TCGCCCC	AIRIATKVSVPRGCRATPPHWTPELAKLDEEIA		
				ATCTGCT	GCGPSHRREKL VATRKQILDRTTKKRWSTLCS		
				GTATTCC	RLAVSDRCSWHIVKKVYAPRPLTTPAVLVDN		
				GTTCAAC	AAITDYRQAERFSKLYSSRARRHPDSHPPAPIK		
				TGCGGAG	TIASEFSPITMAELRRSIKLLPSGSAAGPDCLYN		
				CTACAAC	EALQHLGRTALNVVLRLFNESLRTGVVPPAW		
				AAAAATT	KTGVIIPIKAGKKAEDLDSYRPVTLTSCCLKV		
				ATAGAGG	MERIIAARLRDTVESQLTPQQSGFRPGCSTLEQ		
				GTGTGTT	LLHVRAALCRPTHQYRTGAVFVDYEKAFTDV		
				AGGCTGA	DHDKIAREMHRMKVSPHIVKWCVSFLSNRTG		
				ATAAAAA	RVRFKEKLSRSRTFERGVPQGTVLGPIMFIIVM		
				AGGGAGA	NSLSQRLAEVPLLQHGFFADDLTLLARHTERD		
				CTCTGCC	VINHTLQCGLNVVLQWSKEYFMSVNVAKTKC		
				ACAGTCG	TLFGCIERHPLTLQLDGERIGADRTPKLLGVTF		
				CCAGACC	QCLQGMATHAAETRRKMDFRLLQIAAISAST		
				GATAGCA	WGPRRQVLRAFYLALVQAHTMYGIEVWYWD		
				TCTCAGG	ASERSRDLLAAQHKASRIIAGIPHGTRKEDSL		
				GCTCTAC	LEANLLPLKTTTLVRSMKFMLMCESRGGCLRR		
				GGTGATG	SAEEVYHSKHPVRALHSRIMRSYPHLRIEPREH		
				GCTGATG	PLETSTLRHSCRPLFHTQIKPVCADDPDDVKRE		
				GCCGCGC	ASEKWIARHFARRGKEPPRREHYELWTDGSVS		
				CAGTGGG	LGEKSGAAALLYRNNTLICAPKTGAGELSCSY		

					GGGAAAC TCTCACG AAGGCAC GAAGAAA ATTCCAA AAAAAAA AA	RAECVALEIGLQRLLKWLPAYRSTPSRLSIFSD SLSMLTALQTGPLAVTDPILRRLWRLLLQVQR RKIRIRLQFVFGHCGVKRNEVCDEMAKKAAD LPQLRDTWIPDIIAYAKRVL RSEEVHENTHRFG ITGNHFPTKHKEELTREEETALARFRVGSSRHY GWMLRKINPSVPPQCRWCNPQHAAIGPTIQT PTVATR TLQRTSEPTKCTECDATYQCRSSAVT H MVNKHG FVRADALRRIKYGDATPAVDIPPEP PPVVAIVPLPSSTRVPMRPQVLHCTLCASKFAV PGRLLHHLRTIHGIGSGSCRVKRGRENEDSXQ GDGRAPAAPASQDTRKLPFQCDLCEASFGTRS SLSLHKKFKHK SIVTEDGT VVVVQFPRKRARE GNVDVPGKGEVQCGVCQKVLSCRDSLIRHCK AFHKGEGVELKCSKSTKLCATDSPHTNTSMLV CPTCGRQCASKTGLTLHQKKMHGMKVERAV TSQRGDCEETSLHLMSCPGLKELRVRF AVEGV CVQDVCFSKRLAQFLIAVERSRPQVKSSPKVT VIQPTLPSATSPLIAECGASRKSTGRRNSPDTLR DRGGMQSTSNRNRKKRGRE			
Rex 1	Rex 1- 1_X L		Xenopus laevis	CACCC TCCCC CAGGG AAGAC	TGACCTT ACTATAT ACCACAA CGTTACA	MDVGTTKVQVYNRQTLLRYSNHAYLTIKDGL LHTLRDLGLLREPDLLTAVSPGANGGRRSTRK RCTRTRKRGKRAGVRARLRTNPSRPALPSILLS NVCSIENKMDYIRLQQT TQREFRDCCMFVFTE	cd0165 0	563	816

			ACATC	CACCCTT	TWLSERVPDAAIQLDGLVAVRADRVAALSGK			
			TGCAA	GTATAAA	NRGGGLCVYINTKWCRNSVLVSSHCSPLVEFV			
			GGGCT	TAAATAG	VVKCRPFYLPREIASIHIGVYIPPSANTKDALW			
			CCCAG	CCATTGG	ELYSAITDIQNTHPDGLIIIAGDFNHANLRSVLP			
			GGCAC	ATACAAT	KFHQHVEFATRGPNTLDQVYTNISSAYRAEPR			
			TGTTA	TGTTCTCT	PHFGYSDHISVMLIPAYRPLVRHSKPVLKQVK			
			AAATC	ATGAGCA	TWPAGAMTALQDCFECTDWNMFREAATYS			
			TGCAC	CATCTGC	YINLDEYTSSVISYITKCIDDVTVSKCITTRLNQ			
			CCATC	ACTTTAT	KPWFTGEVRVLLRARDSAFREGDEVALRTAR			
			CGTGC	CATGTTC	AKLSRAIRVAKRAYTQRIHSFFQENGDTRRMW			
			TTTAC	TTTATCA	QGIQAITNYRSTPSACDSASIADALNDFYARF			
			AAGTC	TGTTCTT	EAQNNTGMRKTITLPSDQALRLTVADVRLMTL			
			CCCTT	ACTACTA	ERVNPRKAAGPDNIPGRVLRECADQLADVFTD			
			CTGCC	TCATATC	IFNISLNTSTVPKCFKTTTIVPVPKSSVSCLND			
			AACAT	ACAATGA	YRPVALTPIIMKCFEKLVKKHIKTLLPPALDPL			
			CTCCC	TACACCC	QFAYRQNRSTDDAIATTLHLALSHLDKKDTYV			
			ACACC	ATCATGT	RMLFIDFSSAFNTIIPQHLEKLSLLGLNTSLCS			
			AGCAG	CTGACGT	WVLDFLTDRPQSVRIGNIISRTTTLSTGAPQGC			
			CCTAA	TTAGCAT	VLSPLLFTLLTHDCAATHRTNHIKFADDTTVV			
			AAGGG	TATTTAC	GLISNNDSPYREEVQRLTDWCRANNLSLNVD			
			ACATC	TTAACAT	KTKEMVVDFRKTRSDHLPLYIDGSNVEIVKST			
			CGCTG	ATTACAT	KFLGVHLAENLTWSHNTSYLAKKAQQRLYFL			
			CCTAA	CTGCTGA	RKLRKSHLPPSILKTFYRGTIESILSSCITVWAG			

			AGGAG	GTGTGCA	NCTVMERKTLQRIVKTAEKIIGVSLPSITDIYTT			
			GAAAA	CTGTCTT	RCIRKATSIVADPTHPSHSLFTLLPSGKRYRSIR			
			ACTGG	GTCCTGT	ALTTRLCNSFFPQAIRLLNTQGPDRSLSHSHSI			
			CTCTCT	CCCTGCA				
			AGCAC	CTGTGCT				
			ACACA	GTCCTGT				
			GAGCC	CTTGCAG				
			GGAGA	GAGTTGC				
			ACAAA	ATATTTTT				
			ACGGG	GCACTTG				
			GCCTA	CACTTTTT				
			CCTCC	GTCTGTT				
			TCGAC	GTCCGGT				
			AGAAG	ACATCTA				
			CCGGG	TGTTGTG				
			GACTC	TGTAGCA				
			GAGTA	CCATGGT				
			GTAGA	CCTAGAG				
			GGTAC	GAACGTT				
			CCGGG	GTTTCGT				
			ACCGA	TTCACTG				
			GCTGG	TGTACTG				
			GGCCT	TACAAAC				

				GAACT	GTATATG				
				GCAGG	GATGAAA				
				GCAGA	TGACAAT				
				CTGCT	AAACTCA				
				GGCCT	CTCTTGA				
				GGTGA	CTCTTGA				
				CGTGG	CTTG				
				GGCAC					
				CTCGG					
				GGGAG					
				ACTGT					
				GGGCC					
				TAGTG					
				GCGGA					
				CAAGA					
				TGGCG					
				GCGCG					
				ACAGT					
				AGATT					
				GCAGC					
				GGCCT					
				CTCCG					
				GATCA					

				CAAGG TGCTTT CTTTTT TGTGT TTTTTG TGCTTT TTTACT TTGTTT TTTGT GAAGC T					
L2	L2- 3_X L	.	Xenopus laevis	CCCCA GACCT GTGTG CAGCA GCTCC CAGCT TTTACT CTCCC TATTT ATTAA GACCT CTTTA	TGAACTG GCACTTA TATAACA GTGTAAC AAACTGT AACCCAC AGCACCT TAATACC CCTCTGA ATTGTGT CTGTATG TTACCCT	MAKLIIFAVFACMVLPKVVTSFNRPIVCQYSIYI SSSLLPSTYSSSYTLNYLQHLCPPVTNKTRICR HFKSSSHIFSLFYLLLLAAGDISPNPGPCSIPTYI QPRFPANLRKSSRSSTFSNLIQIPTLPSSCLPFSC ALLNTRSLTNKSTLIHDLFALIHLVFLHQKPGS HRMIPPLRLHSLMAVSLSLTPPEPLDVVGELEF SSRCRFKPIPIPPAFSFNSFEVHAIHIFHPLSVRV AVIYRPPSASSSAAFLSDFDTWLSFFLCTDSPAI ILGDFNFHIDDASQPWPSRFLRLTSSFELQQWT NTPTHKDGHFLLDLVFTKKLSLSGFNSEPFPLSD HNLVTFSSISHSPCPPASSPTVLVRDTRDIDL PAL AQSFRRSSLISLKEMSDPDTLVSEYNHILSSTIGS	cd0165 0	664	922

			GGGGA	CCCATT	FAPLQAKRTRALNPRPWLNSQTRFLRSCTRSA			
			GAGCA	AGACTGT	ERMWKKSRQTQADFIHYKFLLSCLNSALSKANK			
			TTTGCT	AAGCCCT	STTVPLIHKPDSYAPVRDLLNVCGRNPVRKQT			
			GGGGT	ACGGGGT	LFTINFYCPASILPCQKPTRVLQYPYKQSQIQPT			
			GAGCA	AGGGTCC	TPVFSIQYFLHPSSTVSVNQTHSPQDFADFFKD			
			TTTGC	TCTAGCC	KVESIRNQIPTSTNVNQLLLKPSPSACLNSFPPV			
			AAGGG	TTTTGTTT	TMSEVSKLLLSSPLTTCPLDPMPSLLKHCVAE			
			TGCTA	CCTTGAC	LTPLLTHIFNSSLASGTFPSPFKQACVKPILKKA			
			TTCGA	ACTGAGC	TLDPACLSNYRPVSLLPLASKLLERIVFSRITNF			
			AGTGA	ACTTAAT	LHTHDLLDPLQSGFRPGHSTETALCRVTNDLQ			
			TTCATT	CTGTATT	VAKAKGHFSILLLLDLSSAFDTVDSLLMQILY			
			TACAA	GTAATTA	SLGIRNQAASWLSSYLSNRSFTVSYDNKTSSPV			
			GTGAT	TATTTTAT	PLNVGVPQGSVLGPLLFSLYTLSLGDLIHSFGF			
			CTGGA	ATTTATG	NYHLYADDTQIYLSTPSLTAETEAQISNCLLAIS			
			AGAGT	TGAATTG	NWMNQRHLKLNLTKTIELIIFPPKPGPTPLFTISI			
			TTAAC	TATTTCT	DGMLINPVNSARCLGVIFDSSLFSFDHINTTVK			
			AAGGA	AATAATG	TCHFFLRNIAKIRPFLSPDTSKTLMHALILSRLD			
			GTTAT	CACTTAT	YCNLLLTGLPNSHLSPLQSVLNTAARIILLSSKR			
			GAATT	TGTTACT	VQAPLLKSLSWLPIKQRITYKLLLITFKALHSS			
			GACAT	TTTTATTC	APNYISSLVSLYVPARNLRSSQSNRLVVPPTTT			
			CTCTCT	CAATGAC	AVSRIKPFSLATPYIWNAPESLRRESSFNVFKT			
			TCCTCT	CCCTTGT	KLKDYLWEHL DNT			
			TCTGTT	TTGTTAC				

				TATTA	TACTAAT				
				CAGGT	TTATTGTT				
				AAATC	TTGCTGT				
				TGCTA	ACAGTGC				
				AAACT	TTTGCCC				
				GCTAT	TCAAGGA				
				TCCCA	GCGCTAT				
				TCTCT	ACAAATA				
				GCTGT	AAAATAT				
				GGCCC	ACATACA				
				TGTTG	TACATAC				
				TCTGC	ATAC				
				ATTTT					
				GCACC					
				TTTATT					
				GGTAA					
				ATCTG					
				CTAAA					
				ACTGC					
				TATGC					
				ATTCC					
				CTTCC					
				ATCTC					

				TGCTG					
				TGGCC					
				CTGTT					
				GTCTG					
				CATTTT					
				TGCAC					
				CATTT					
				ATTGG					
				TAAAT					
				CTGCT					
				AAAAC					
				TGCTA					
				TTGCA					
				TTCCTT					
				CCATC					
				TCTGC					
				TGTGG					
				CCCTG					
				TTGTCT					
				GCATT					
				TTTGC					
				ACCAT					
				TTATT					

				GGTAA					
				ATCTG					
				CTAAA					
				ACTGC					
				TATTTC					
				CATCT					
				GCATT					
				ACACC					
				TACAC					
				TGTTT					
				AGTGA					
				GTTAT					
				ATTTA					
				TGCAT					
				TGTTC					
				ATTTG					
				CCACT					
				TCTAA					
				ACTAG					
				TCACA					
				TAATA					
				TTGAT					
				TATAA					

				GTTCT					
				CCCTG					
				TAGTC					
				TTGTA					
				ACTTG					
				TACAT					
				TTTAC					
				GACCC					
				CCTAT					
				AAACT					
				CCTCT					
				CCAAA					
				CAAAC					
				TTTGT					
				AAGTC					
				AGGTT					
				ATCTC					
				TGTAT					
				TCATC					
				CTGAG					
				TCTTG					
				CTCTG					
				CTCTC					

				ACTCT					
				CTGTC					
				ACAGC					
				TGGGT					
				GATCA					
				GTTTG					
				GTGTT					
				TGATC					
				AATTA					
				GAATC					
				AGTTG					
				TGATT					
				CCCCA					
				GACCT					
				GTGTG					
				CAGCA					
				GCTCC					
				CAGCT					
				TCTAC					
				TCTCC					
				CTATTT					
				ATTAA					
				GACCT					

				CTTTA					
				GGGGA					
				GAGCA					
				TTTGCT					
				GGGGT					
				GAGCA					
				TTTGC					
				AAGGG					
				TGATT					
				CGTTT					
				CGAAG					
				TGATT					
				CATTT					
				ACAAG					
				TGATC					
				TGGAA					
				GAGTT					
				TAACA					
				AGGAA					
				AGCTA					
				CAAGG					
				AGAGA					
				CAACA					

				CAGAG GACTC CATT AGGAC ATAAC ACTGC TACCG GCTAG AAACT GCA					
Pene lope	Pene lope- 5_X T		Xenopus tropicalis	TTGTTT TAGAC TTAAT TCTAC TAG	TAACACC TTCATCA ATGAGAA TCTTGGT ACCATTG TGATTGG ATCATAC CTTCTTA CACCTGT CAGTTTC AGCTAAC ACCTAAC TGAACAA	ILYIMTWMNENLHRHIATHFGENTLKL VREYERTARKLADYRNHLRFNLR CRYQGITPCSLRLGSSVKGHR AKVILQKAQKQLNERNVRQTNFS IEVLVHKFDNAKQRLTAKLPVLT LQRRVVEFTERAQLAQHAKGKER QINKFTSLLSRTNNTQRTDKST GRKEDENKTCQNIKDLWVKNLSD RELTEPEKDVLA KGLNF AVAP RYPVVD FITATESSIHN NKIP VDEAENIRLKVSAALANAKAPPS NLSLQENRALTSLAKDSSVTIL PADKGRCTVVLNTSDYHAKVST LLNDSDTYEQLKRDPTSNYKKK VIDCLQHLEKEKVISPALYRRL YPGEATPCLYGLPKIHKEGAPLR PIVSSINCVTYNIAKYVANILAP LVGNTVHHIQNSMDFVKKVKEL KLEIDDTMVSY	cd0030 4	482	563

					GTTCAAT	DVTSLFTCIPTAE AIDTVRKRLKQDTT LSSRTK			
					GGAACCA	LTPEQVCSLLDLCLNTTYFKHKEVFYRQKHGC			
					TTGTGAT	AMGSPVSPIVANLYMEEVEKKALDTFKGTTPS			
					TGGATGT	HWFRYVDDTWVKIKEREVPAFTKHINSVDKNI			
					CTGTGAC	TFTREDVKDSKLAFLDCAIVIKEGRDL D IEVYR			
					TCTACTA	KPTH TDQYLLFDSHHPLEHKLGVIRTLHHRAE			
					CCATCAA	TVASNAEAKEKEYKHLKGALKTCGYPDWAFI			
					GAGTTTA	KTKSKTNRKTRPTNRREEHSRRNNIVIPYVAGT			
					AATACCG	SEKLRRIFNKHRI PVFFKPSNTLRQKL VHPKDP			
					GGGAATT	TPKHMKS NVVYAVQCSEECSDLYIGETKQPLC			
					CCCTACC	KRMAQHRRANSSGQDSAVYLHLKEKGHSFED			
					AGTCATT	SNVHILDREDRWFERGVKEAIYASLEKPSLNR			
					TGAACTG	GGGLRHRLSPTYNGALTSLPRQFHNSSH FQSC			
					AAGAAGC	TLKD			
					CACTCGG				
					ATGAGTG				
					GTGAAAC				
					GTTTTCA				
					AGAAAAA				
					CTCAGAA				
					AAGTCCA				
					GTTGTTTT				
					AGACTTA				

					ATTCTAC TAGATAC TGTATAT CATGACC TGGATGA ATGAGAA TCTTCAT AGACATA				
Pene lope	Pene lope- 6_X T		Xenopus tropicalis	A	TAACACC TGCTTTG AAGAGTT GCATGAT ACTTGGG TAATCCT TTCAAAG TAACTCC ACAGCAT TCACACC TCTGTGA GTTTCAG ATGTCAC CTTTCTG AAGAGTT	KGHRASNILQKAQKHLLNERVRQINFTIEALK QKIEQLKQSLTKQLPDVTLERVVQFIEHAQMA QHIKSKERQISKFTSLLSRSSGLKRTEELTWRQ KDETSKNTKDTWVKNLSDRVLTQPEKDV LAK GLNYAVTPQHIPPVELITATESAINNNHIKVEE AEQLRLKVSAALSSARAPPSNLTTQERRAL TSL SKDPNITILPADKGRCTVVLNTTDYHCKMTSL LSDSNTYEPLRRDPSSSYKKKVIDCLQQLEKEE AIDRALYHRLYPREATPCLYGLPKIHKDGAPL RPIVSSINSVTYSIAKYLANILAPLVGKTVHHIQ NAKEFVTKIQGVKLEAEETMVSVDVTSLFTCIP TTEATETVRNRLQKDNTLSSRTKLSPNQVCLL LDLCLNTTYFKYKDQFYRQKHGCAMGSPVSPI VANLYMEEVERKALLTFNGTTPSHWFRYVDD TWVKIRSNEVA AFSEHINSVDNNIKFTREDVHE	cd0030 4	413	496

				ACATAGT	NKLAFLDCLISIQEGGILKTEVYRKPTHTDQYL			
				GCCATTG	LFDSHHPLEHKLGVIRTLHHRAESVTTDTESKD			
				TGATTGG	KECKHLRGALKACGYPDWAFVKTRATKPNRN			
				ATTAGTG	TKRNNRPEAERRRNIVIPYVAGVSEKLRRIFNK			
				GAAGAGT	HHIPVFFKPSNTLRQKL VHPKDPTPKEKQSNV			
				ATATATG	VYAVQCSEECTDLYIGETKQLLSKRMAQHRR			
				CTGAGGA	ANTTGQDSAVFLHLKDKGHSFENSINVQILDKE			
				CTTCCCA	DRWFERGVKEAIHVKVEKPSLNRGGGLRHHL			
				TACCAGT	SATYNAVLTSVPRQFQNSSHIHSCNSNK			
				CAGTTGA				
				ACTGAAG				
				AAGCTGC				
				TCGGATG				
				AGTAGTG				
				AAACGTC				
				TTCAATG				
				ATTACTA				
				AACAAGT				
				CCAGTTG				
				CTTTAAA				
				TTTATTTA				
				TACTAGA				
				TAT				

L2	CR1 -L2- 1_X T	Xenopus tropicalis	AGTCC	TAAAACA	MSLLLLVLSFMFVLPKVTSSFSPPADCPYSITIS	cd0165 0	616	874
			TGTGT	TTATTTTG	SSLLHSTHCCSYNLYTFLKNLAPPTNTSQKHN			
			TCAGG	CCTAGTC	NFKSFSHLLSLSLLLLLAAGDISPNPGPCRIPISY			
			AGTGG	CTGCGCT	RPRNASLLVKPSQTANLSNLIQIPTLSSSYLPFS			
			CCCAG	TAAGGGC	CALLNTRSVTNKAALIHDLFCSSSFHLFAITET			
			CTTGC	AAATGCC	WLSQNDNAIEAALSYGGLSFSHTPRITGRGGG			
			TCTCC	CATACCT	VGVLLSPHCRFKLIPIPTFRFPSFEVHAVQLFH			
			CTATA	GGTGCAC	PLSVRVAVVYRPPSSKSLATFLSDFESWLSFFL			
			TCTTA	TCTTACC	CSDTPAILGDFNCHIDDPSPQPWPSRFLHLTSSF			
			AGAAC	TTCCAAT	DLRQWTNTATHKDGHFLDLVFSKNLDLSAFN			
			CTCAT	TTGTGCC	SEPFPLSDHHLISFSISHVSPSPSTNAQPKTLIRN			
			TTGGG	TGTATGT	TQRVDITALSNSLSSSLSSFCASSDPELLVNTYN			
			AAAGC	TACCCAA	SILSSAINTHAPLQPQHSRAHNPRPWLNPQTRF			
			AGCGC	CCACTTA	LRSCARSAERIWRKSRVKADFLHYKFLMVCL			
			AAGGG	GATTGTA	NTALSQAKQVYYNTLVNNNKSNNPRRLFSIFNA			
			TAGCA	AGCTCTA	LLHPSPNESIISEHTPQDFADFFKSKVDSIRKQF			
			GCGCA	CGGGGCA	SQPSDTGNLNLPKSPLSVFSSFNLVTESEVSQLL			
			CAAGT	GGGACCT	RSSPPTTCSLDPMPSLLKQCAPALTPALTHIFN			
			GAATC	CCTTCCT	SSLASGTFPSLFKQACVKPILKKATLDPSCLSN			
			CTTCC	ACTGTGT	YRPVSLLPLASKLLERLVFSRVTNFLCTHNLLD			
			AAAGT	CTCATAC	PMQSGFRPAHSTETALCRVANDLQTAKAKGR			
			GAATC	CACATGG	YSVLILLDLSSAFDTVDSVLMQILYSLGIRDQ			
			ATTCA	CACTTAT	AASWFSSYLSNRSFSVALANKSSTPVPLSVGVP			

			CAAGT	TCCCTGT	QGSVLGPLLFSLYTLSLGDLISSFGLKYHLYAD			
			GAACT	GCATTTA	DIQIYLDTPALTTDVQTQIGNCLLAISSWMNQR			
			GGAAA	TATATAT	HLKLNLAKTELMVFPPKPGSPPTITIDGMTIN			
			AGGAG	TTATTGT	PVNSARCLGVIFDQSLSFSNHINNTAKTCRFFL			
			TTTAA	ATTTATTT	RNIAKIRPFLSQVTAKTLIHALILSRDLYCNLLL			
			CAAGG	ATTATAA	TGLPDSHLSPLQSILNSAARILLSPKREPVQPH			
			AGTTA	CACTTGT	LSSLAWLPVKQRIAYKILLTFKALHSSAPHYI			
			GGAAT	CCTCCCT	SSLVSLHVPGRLLRSSLRLRLFTPSTPTALSRLK			
			TGACA	GTGTGTA	PFYLAAPYLWNSIPESLRREHSPTLFKKKLSCY			
			CCCTC	ATTTTGT	LLEH			
			TCTCTT	ATTTTGT				
			CTGCT	AAGATTG				
			TATTCT	TACAGCG				
			ACAGG	CTGCGTA				
			TAATT	CCCTTGT				
			GCTGC	GGCGCTT				
			ACTAT	TATAAAT				
			TTCCA	AAAGTTA				
			TCTCC	TACATAC				
			ACTGT	ATACATA				
			GGCCC	CATACAT				
			TGAAT	ACATACA				
			ACTGC					

				ATTTT					
				GCTAG					
				TTTATT					
				GGGTA					
				ATTGC					
				TGCAC					
				TATTTC					
				CATCT					
				TCACT					
				GTGGC					
				CCTGA					
				ATACT					
				GCATT					
				TTGCT					
				AGTTT					
				ATTGG					
				GTAAT					
				TTAGC					
				TAGTT					
				CTATTT					
				CCATT					
				CTCAA					
				TTCCC					

				AATAT					
				TGCAT					
				GCCCC					
				CTATT					
				CCTGT					
				GCTCC					
				ACCAA					
				TTTGTT					
				GATTT					
				TATGA					
				CCCCC					
				ATCGT					
				AAACT					
				CCTCT					
				AAACA					
				AACTT					
				TGTAG					
				TTCAG					
				TGTAT					
				CTTTG					
				GGCTG					
				ATCTT					
				CACTT					

				GTGCA					
				CTTCTC					
				TCATC					
				CTGGG					
				ACTCA					
				GCTGA					
				TTGAT					
				TTAAT					
				TTGTT					
				ATTTA					
				ACTAG					
				CTGCA					
				ATCAG					
				CTCTG					
				ATTCC					
				CCAGT					
				CCTGT					
				GTTCA					
				GGAGT					
				GGCCC					
				AGCTT					
				GCTCT					
				CCCTA					

				TATCTT					
				AAGAA					
				CCTCA					
				TTTGG					
				GAAAG					
				CAGCG					
				CGGGT					
				AGCAG					
				CGCAC					
				AAGTG					
				AATCC					
				TTCCA					
				AAGTG					
				AATCA					
				TTCAC					
				AAGTG					
				AACTG					
				AAAAA					
				GGAGT					
				TTAAC					
				AAGGA					
				GTTCA					
				ACAAG					

				AGGAG CAACA GGAGG AGCAA CAGCA GAAGA CCGGC CAGCC TCTGT CCATT TGAAT CGGGA CCCAG TTTACT GGATA TACCT GCATA					
L1	LIN E1- 3_Z M	.	Zea mays	GGGAA ACCAT AACTG TGGAA CAATT GCTTA	TAGAGTT TTTATAG ATGAGTC AGTCTAT GCTAGTT TCGTTCA	MDPEKIIVWNVRGLNSLARQNSVRELLNCYG ADIICLQETKMAQPSSRVLLSMLGSDFTGSLAL PSNGSRGGILIAWRRHLGITGATRDIHSISVQF CSVEGHQWWLTCVYGPQSNEAKIEFLQELRHI RAACQGPWMITGDFNLILHAEDKNNSHINRA MMGRFKKLVDLALKEIPLHGRKFTWSNQQN	cd0165 0	500	762

			ATCCC	GGCCGCC	QPTLVKLDRVLC SIDWEELFPGAVLQSAASED			
			TTTGCT	GTAAGGC	SDHCPLLLGFKDCKGVKRRFHFESFWTKIEGF			
			AGTGT	GTCAGTT	HEVVHNAWNSVPVGHCPFSTFDLKLKGTA KA			
			TCACC	GTACCCC	LQGWSSKKVGH IETQLHLARDLIHQLEIAQDG			
			AGGTT	GGTCTAT	RILSVLERWLLSKLKKLALALASTKR TIARLRS			
			CATCC	TTTAGAC	RISWLKEGDANSKLFHSFARYRKKKNSVSKLT			
			TGACA	CTTGTTT	VGNQVLTAHSDIAAAVDQFYFDLIGASKDRNI			
			CCCTG	ATTTCTTC	SINLEALETAGTPFPNLIELDSPITEQEVCNAINS			
			GAGCT	TTTTAAT	LPSDKSPGPDGFTGRFYKICWSIIKEDLMAAVI			
			AAAGG	ATAATGA	AIWNRKFDNFHKLNSAFITLIPKKDNADQVKD			
			ATGTG	TACGCAG	FRPISLVHSFAKLITKILASRLASRLNDLVSPNQ			
			GCTGT	CTCTCCT	NAFIKGRFIQDNFMLVQQTAKLLHQQRHPCLM			
			TTACC	GCGTGTT	LKLDITKAFDSVSWPFLIEVMQNLGFGQIWRDI			
			GATGT	CGAGAAA	ISGLLLTSSTRVLVNGHPGSIISHRRGLRQGDPL			
			TCTGC	AAAAAAA	SPMLFVLVMDVLGQLVSKADEEGLLQPLSTRQ			
			TTGGT		LQHRVSLYADDVVLFINPVAEDMTMISDILQL			
			GTTTA		FGEASGLYNNNQKSSVYPIRCHEDNLEIIQQFW			
			GACCC		PCEISSFPCKYLGLPLSLHKLRRREQVQPIVDKIA			
			GTCCT		NQLPGWKADLLTKAGRSVLVQSVLTGVLIYS			
			TGGTG		VMAVDLPSTTLKDIDKIRRGFLWRGRKNAMG			
			CCTGC		GHC A VAWGSVCRPRELGGLGIFSITELSWALK			
			TTCAA		MRWAWLQKTEPGRP WASLPLHTPKKIREFLQ			
			GGGAG		VALYSEIGNGASTLFWSDRWLDGQRIGDIAPR			

				CTTTG	LIETIPMKLINKRTVQEAIEDNGWINDIPGSLSV			
				GATCG	GALADFLRIWDLVAHVLDLDPHKEDKHIFRLAA			
				TCGAA	NGKYSTKAAYEGLFIGSVEFEPFERIWKTWAP			
				CCTCC	PKCRFFLWLVAHKKCWTADRLEKRGLDHPEK			
				CTTTG	CPLCEQERETIDHLLVNCVFSRECWFLLLRQFG			
				CCGTT	LQGLAPQPTDINFMEWWHQANEAIQGSFRDG			
				GATGG	LNSLIMLGAWILWNHRNRCIFDGLSPNVANFLI			
				GATGT	HVGDERRLWETAGAKGLTSLVASLSHVA			
				CTTCT				
				GGCAG				
				GAGAA				
				CGCTG				
				ATGTA				
				TCCAA				
				TCCGC				
				ATCAC				
				ATTCT				
				CGTCT				
				GCCCC				
				ACGGA				
				GGGAT				
				GGATC				
				CGATC				

				CTAAG					
				CTTCA					
				TGCTA					
				TTACTT					
				CGGGC					
				GGGGA					
				TGGTT					
				CTGGT					
				CACGT					
				GCAGC					
				AGGGA					
				TCCCC					
				TCCTTT					
				TTCAA					
				GACCT					
				CCCCA					
				TTCGA					
				GTGGG					
				AGAAG					
				GGGAG					
				AACAT					
				GGTCA					
				TTCGG					

				GACGT					
				CTCTC					
				AGTTC					
				ATGAG					
				CGTTT					
				GGGCC					
				CAAGA					
				GATGA					
				GAACA					
				TGCAG					
				CAGCG					
				GAGCA					
				TTAGC					
				TCCGG					
				GGTGG					
				CAGTT					
				TCAGA					
				TTTGA					
				GGAGA					
				ACTTT					
				GGAAT					
				CTGAT					
				GGGGG					

				GACCT					
				TGTCT					
				CCTAG					
				GCTGG					
				TAGAG					
				ATGAC					
				TGATG					
				ACATG					
				GTTCTT					
				GTTTC					
				ACCGG					
				TGAAG					
				GCGAA					
				GCAGC					
				TGAAA					
				GTTTA					
				TTTTCG					
				TCAAA					
				GATTC					
				AAGAA					
				TGTCC					
				TGAAC					
				CAGGA					

				GGAAG					
				CTGGA					
				ACTCC					
				CAATT					
				CTGTC					
				ACTGT					
				GCTTC					
				CTCCA					
				TCCCC					
				CACAG					
				ACGTG					
				GCAGC					
				TGCTG					
				ACACA					
				GATCC					
				ATCTG					
				AGTAT					
				CCTGT					
				TTCAG					
				ATGTT					
				GACGC					
				TGTGT					
				TGGCA					

				ACTCC					
				GTCTA					
				GAACA					
				TTACC					
				TCCCT					
				CCCTG					
				TCTGA					
				TGTGG					
				CAGCA					
				GCACA					
				CGCGG					
				CTCCG					
				TCGGA					
				GGCTA					
				ATGCC					
				TTAGA					
				CGTTG					
				ATGCT					
				GAGTT					
				GGCTG					
				CTCCA					
				TCTGA					
				GGAGG					

				ACATT					
				TCAGC					
				AACAA					
				CACAT					
				GCTGC					
				GTTGG					
				CTGCT					
				CCATC					
				TGAGG					
				ACCCT					
				CCAGC					
				AACAA					
				CACAA					
				GATTC					
				TGTGG					
				AGATT					
				TGTGC					
				TCAAG					
				AGCAA					
				TTCAT					
				CAGCA					
				AATTG					
				TCAAA					

				CCAGA				
				TCACA				
				CGCGT				
				ATTAC				
				CAATC				
				CCAAA				
				ATCTG				
				TCAGA				
				CGTCA				
				GTCCT				
				CTATC				
				CCCCT				
				GGCAG				
				TTCCA				
				AGACG				
				CAGCC				
				GACGT				
				GTGGC				
				GGGGG				
				TTGGA				
				GTTGA				
				ATTTA				
				AAATT				

				CAGGA					
				TTGGG					
				GAAAA					
				AGATC					
				TAAGA					
				AACAA					
				GCTAT					
				GGTGG					
				CTCTA					
				AATAT					
				TATTG					
				AGGAC					
				AATGA					
				AGGAT					
				TCAGC					
				CGTCA					
				AGCTT					
				TAGAT					
				GACTA					
				CGCCA					
				AGCTC					
				TTCAA					
				GAATA					

				CTCTTT				
				CTCGT				
				AGTCA				
				TGTGC				
				AGGCT				
				CTTGC				
				TGCCC				
				TTTTTG				
				GGTGG				
				TCTGT				
				CCCAG				
				ATAAT				
				TTAGA				
				ATATA				
				TAACT				
				GGGAC				
				TGGGG				
				ATATT				
				ACCGC				
				AACAA				
				GCTTA				
				TTTTA				
				GCTGG				

				CGATG TTGCT GATAT TGTTT GTCGA TTTAC CTCTC ATTGT TTCTA C					
RTE	RTE 1_Z M	.	Zea mays	AATGG CAAGG GTCTT CGCAT CTGTC TCAAC GGGTG CGAGG GGTGG AAGCT AGTTG AGCCG ACTAG	TGATTAG GCCTTTT CCCTTTT AGTGCTC TCAAAAG CTTTTCCC CTCCCCT TTCTCTCT CTTTCTCT TTTTGGT GACATCT TGTTGGG TTTCAAC	MPLSESEKRKFWEDLDGMVRAVPTNEKLFIGG DLNGHVGSTNVGYELAHGGFGYGSRNQGED LDFAVAYNLVIANTFFRKRDShLVTFSSGHRSS QIDFVLTRREDKQACLDCKVIPGESVVPQHNL VVADFRFRIRTHRDKQAKIARTKWWKLKGET SEVFRERVFVEGAWSEEEDANNMWVKMATCI RKVASEVFGVTKGSIGEPKDTWWWTEDVQKA IKEKKECYRSLFHDRSAVNIERYKVAKKTAKR AVSEAKGRAYDDL YRRLSTKEGEKDVYKMA RIRERKTRDLNQVKCIKDEMDQLLVKGQDIKQ RWQRYFDNLFNGENETMDTQLDDSFDDL NRC FVRRIQESEVKEALKRMKGGKAMGPDGIPIEV WKCLGDIAIVWLTKLFNHIFRSNKMPDEWRRS	cd0165 0	409	658

			GATCC	TCTAGCC	TLVPIFKNKGDIQSCTNYRGIKLMSHTMKLWE			
			GTGTA	TACCCCA	RVIEHRLRGMTHITMNQFGFMPGRSTMEAIFLI			
			GGTAG	ACTTGCT	RQVMERYKEQKKDLHMFIDLEKAYDKIPRN			
			TTGGA	TGGGACT	LMWWALDKHKVPTKYVTLIKDMYDKVVTSV			
			ACGTA	AAAAGGC	RTTDGDTNVFPINIGLHQGSALSPYLFALVIDE			
			GGGTC	TATGTTG	VTRDIQGDIPWCMLFADDVVLVDESREGVNR			
			CCTTA	ATGTTGA	KLELWRQTLESKGFRISRTKTEYMRCDFGTTIS			
			CAGGT	TGTTGAT	EDGDVSLGGQVVPKKDTFRYLGSMQLQRDGDI			
			AAGCT	GTT	DEDVSHRIKAGWMKWRQASGVLCDKRVPQK			
			AAGAG		LKGKFYRTTIRPAMLYGAECWPTKRRHIQQLS			
			TTAGT		VAEMRMLRWICGHTRLDRVRNDDIRDRLGVA			
			CGATG		PIEEKLIQHRLRWFGHVHRRPPEAPVHRGIIRR			
			CAGCG		DNNVKRGRGRPNLTWEEAIKRDLKEWNIPME			
			AGCAG		LCLDRSAWKEAIVPEP			
			GAGGC					
			GTGTT					
			AGTAT					
			CGACA					
			AAATG					
			GAAGG					
			GGCAA					
			AAAGC					
			GAAGG					

				AGGTG					
				GAGAA					
				CACTG					
				GTTTA					
				CTTTG					
				GTACT					
				CAGGA					
				ACAGT					
				GGCAA					
				ACAAA					
				AATGA					
				GTAGG					
				TATCC					
				TGATT					
				GATAA					
				GACCC					
				TCAAG					
				GATGG					
				AGTGG					
				TGGAC					
				GTCAA					
				AAGGC					
				AGGGA					

				GATAG GATTA TCCTA GTCAA ACAGT TTGGG GATGT AGTCT TGAAC ATTAT AAGTG CAT					
L1	LIN E1- 43_ ZM		Zea mays	GTCTA GTCAG CAGAA TCCCC CAACC CTAGT GAAAC TTGAC AGAGT ATTTA TGCAC	TAGCTAG CATTGTG GTAATTC GTGTTAG TTTCTTCT TTAGNCA TGGATGG GTTTCTG TATGCCT GTTACTG GCCGCCG	MTIVWQEGAGFILSPFGLSLRDFWRLFRWPGIL FQLCSVLLQRLNSKLKATARGLQAWSAKKVG HISASLALARVLLHQLEIAQDSRGLSPAELWFL LCLKKHSLALSSLQRTIARLRSRIGWLREGDAN TKLFHMHSRFRKKKNFVPKLVSNGAVLTSHD AKADLVNEFYGNLLGQCSDREQTISLESLGIPC HDLSALDAPFTEKEVWDITLHLPSDKAPGPDG FTGKFYRVCWEIHKADVMVFSAVGSRRFANF QILNTAYITLIPKFEGADQVKNFRPISLVHSVAK LLTKLLANRLAGRLHDMVSPNQSAFIKGRFIQ DNYMLVQQTTRYLHQQKQPRMLLKLDISKAF	cd0165 0	263	526

				ACTGG	TAGGCGC	DSVSWPFLLEVLQHLGFGQVWRDIISGLLWTS			
				ACTGG	CTCTTTTG	SSQVLLNGLPGQPILHRRGLRQGDPLSPFLFILV			
				GTGGA	TTTTCTTC	MDTLGFLINKAAQDGLLQPLAPRALQHRVSLY			
				TTTGT	TGTAAC	ADDVVIFLRPAAMDINVILDCLQLFGDASGLR			
				TCCAA	GTCCTTC	TNVQKSSVYPIRCGDQDLALLQNMLPCELAEF			
				ATTCC	TTCTTTA	PCRYLGLPLAIKKLSKIQVQPIIDKIANQLSGW			
				TTGCTT	ATATAAT	KAELLTKAGRKVHVQHVLTMGMVIYLA MAVD			
				CAGAG	ATCGCGC	LPAWALK AIDKIRRGFLWRGRRDVKGGHCHV			
				TTCTG	AGTTCTC	AWGIVCRPRELGGLGISSMKELAWSLRMRWL			
				CCTCN	CTGCGCT	WLAKTEPSRPWASLAIQVPKKVKDFFSMAMQ			
				AATGA	TTCGAGA	TELGNGATTTFFWTDRWLSGHRIAELVPRLFGII			
				CTCAG	AAAAAAA	PKRRVSKRTVVD A ISNQSWISDIQGALSVGIIN			
				ACCAC	AA	EFLILWDLIDSIVLRPQVEDKHFFRLAANGKFS			
				TGCCC		SKEAYRGFFLGSSEFEPFHRIWKS WAPPKTKFF			
				CCTCC		MWLVAHRKVWTADRLQKRGM DHPERCPLCD			
				TTTTG		QDQETLDHMLIGCVFAREFWFKLLLQVNLQFL			
				GGCCT		APQSEDSAFLEWWRTLCTKVSGIARDGLNSLV			
				GA		ILGVWTLWKQRNGCVFDNKSPSIADAIRRVEL			
						EVHLWEMAGAKKLSLLTAPIPGIPAS			
L1	LIN E1- 15_ ZM	.	Zea mays		TGATGAT	MAKKKRRKAPWAEAPSSPRAPAPVAHGPGGV	NA	NA	NA
					AGGCCTC	RSGLVGLPSTVGFAAQGKCAPAPFASSVSWTG			
					ATTTGGA	WSQGISPLAPTLAHSSRVLLPAATSDAAVSKSR			
					ATTGTCA	LGFPDFVGSFSFPHCAAFLSRLWSLPVSSRSD			

				GGGTTTA	FHGGWCVTSLRPPVPEEEWPRLGSGVVRSSKN			
				GGGAATA	LGEMDPKNKNIPNPSWDRWGQKDSAQGSQS			
				GTTCTAA	WNRNRNLSWRLKPQIGKSDPQEGLMSSSSGD			
				GGTTCAG	GIQSGNPILSPSSLKKEIDLARPAFCQKCGVEG			
				TTTCTGA	HHARNCFNALWCDICRKDTHVTARCVLPKQS			
				AAGAACT	KPNMPIVGMAADGLGFYLSHFAKPISKKPKRV			
				CATTAGA	FIGLVKVIEGLVSAEDLAKDFGFHFPWGKTWK			
				GATGAGA	ATKCHSGFLMQFPSQERLDEMISFPELKMMS			
				AAGTTGA	GAKIVVVPWSSRAKPKSRLHTAWIVAENVPEE			
				CTTTATT	LQNYQSICEIGSMIGAVEEVDLMSLDS DIVRF			
				GGTTTGC	KVHIKSVAMIPPVVEVAVKPFLYDIYFRIESISD			
				AAGAAAC	EGWNDETINLGKRASVDIHGINDPVFDKSGKK			
				CCAAAAA	PRNDDEVGDEDPXPNFKVGESSSQGKESFKKS			
				AGTCATT	EVFMDSSKDKSGDGMRLLEDLSDEKVDFDPNE			
				TTAGTGA	DDLSSQEEFAKDMEEMHTDFQAKVGEVL			
				TTCTTGG	SIPLTKDEKQDXVGKKSSSEILMSTSEGIRKSSRL			
				TTGACTT	EKNEDIKVADKAISRAEAKDAFLNKGMMNNP			
				CGCTGGC	FSLLNASNEDLFDISNKLGVCLGSSDSSAMANL			
				TGGTAGT	DLIRDLEMSRKILAIQSCGSKKNEEPIVDDINEV			
				AAACCTT	DSSVHADSEKDIDYDLLDVMVLRKGRKIKHR			
				TTGCCTG	KKQQKKASPKIRYTPNKKREAENFXLPDPF			
				GTTTTCT				
				GCCCCGC				

					CTAATGG				
					TAGATCT				
					GGTGGTT				
					TATTGGT				
					GGGTTTT				
					AATACTG				
					ATGTGTT				
					TGATGTG				
					AGAGAAA				
					ATGAAAT				
					TGGTGAA				
					TTTATGA				
					TTCGCAC				
					CTTGATT				
					TTTCATA				
					GAGAAAA				
					AAATCTC				
					ATCTGGA				
					ACTTTGT				
					TAATGTG				
					TATGGTG				
					CTGCTCA				
					GAAGGAG				

					AACAAAT				
					GCAGATT				
					CTTGAGT				
					GAGCTTT				
					CGTCTTTT				
					TGCTCGA				
					GAAGTCG				
					TGCTCCT				
					TTGCTTG				
					TTGGTGG				
					TGATTTT				
					AATATTA				
					TTAGAAA				
					AGCAGAR				
					GAGAAAA				
					ACAAGCC				
					TGGTGGT				
					GTTAATA				
					AGTGGAG				
					CTTCCTTT				
					TTAACAG				
					TATCATT				
					GAGCAAA				

					ATGGGCT				
					TGTGGAG				
					TTTGA CT				
					TGAACAA				
					TAGATTG				
					TTCACTT				
					GGTCTAA				
					TAATCGG				
					ATTGACC				
					CTACTTTT				
					GAAAAGT				
					TGGATAG				
					ATTTCTG				
					GCTAGTC				
					CTGAGTG				
					GGATCTT				
					GCTTATA				
					ATAATAT				
					CAATGTT				
					CGTGGTT				
					TGAATAG				
					ATCCTTTT				
					CTGATCA				

					TGTCCTT				
					TATGCAT				
					TCAAACG				
					GATGTTA				
					TCCCTAC				
					TGGCAGT				
					AGAGTTT				
					TCAAATA				
					TGAGTTA				
					TGTTGGA				
					AGGTTAG				
					ACCAGAT				
					TTTCACA				
					AAAAAGT				
					GATAGAC				
					AATTGGT				
					CTATTCC				
					TGTGAAG				
					GCTAAAG				
					GTAGTAT				
					TGACATT				
					TGGAAAA				
					CGAAAAT				

					TAAACGT				
					TTGAAAC				
					AAATGCT				
					CAAAGGT				
					TGGAATA				
					TTAATGT				
					TGAAGGG				
					CATTACA				
					AAAAATG				
					AAAAAAG				
					AATTGTT				
					GGCAAAA				
					ATTGACA				
					TCTTTGA				
					CAAAATC				
					AGTGAAG				
					TGGTTGG				
					TCTTTCA				
					GAAAATG				
					ACAGGAT				
					GGAGAAA				
					CTTGACC				
					TGGAATT				

					GACTCTC				
					AAAAAAT				
					TAGTTGA				
					TGAGGAG				
					GGAATTA				
					AGCTCAA				
					ACAGAGA				
					GCTCGTG				
					ATAAATT				
					CATTTTG				
					GATGGCG				
					ATGAGAA				
					TTCTAAA				
					TATTTTC				
					ATCTTTT				
					AGCCAAG				
					TGTAAGA				
					GAAGGAA				
					GCTTAAA				
					ATTGTGA				
					CTCTTAC				
					TCATGAA				
					GACAAAG				

					TTGCTCA				
					AGATGAT				
					GATGAAA				
					TTAATCA				
					TTTGGCG				
					ACTTCTTT				
					TTACCAA				
					AATCTGT				
					TTGGTCC				
					TTCTCAA				
					GAATCTA				
					GTATTTC				
					TTTGAAT				
					CATTTGA				
					ATATGAA				
					AATGTTG				
					GATGATA				
					CTGACAG				
					AGATTGC				
					CTGACCA				
					GACCGTT				
					TGACATG				
					GAGGAAA				

					TTAGAGA				
					AGTGGTG				
					TTTTCTAT				
					GAAACAT				
					AATAGAG				
					CTCCTGG				
					TCCTGAT				
					AGCATCC				
					CAGCTGA				
					ATTTTTTC				
					AGGAGTT				
					TTGGGAC				
					ACGATTA				
					AAAATGA				
					TCTGTTT				
					AACCTCT				
					TTCAAGA				
					TTTTCAT				
					GATGGCA				
					CTCTTAA				
					CATTGAG				
					AGACTTA				
					ATTTTGG				

					AATGGTG				
					ACTCTTA				
					TTCCCAA				
					AGTTCCT				
					AATGCTA				
					TTGACAT				
					TAAAGCT				
					TTCAGAC				
					CTATTTG				
					CTTGCTT				
					AATGTGT				
					GTTACAA				
					GATTATT				
					ACCAAAG				
					TTTGGAC				
					CAACAGA				
					TTGGCCA				
					GTTGTAT				
					CAATAAT				
					GTGATTA				
					GTGATTT				
					TCAGTAT				
					GGTTTTA				

					TCAAAGG				
					TAAATAC				
					ATCATGG				
					ATGGAGT				
					GGTCTCT				
					CTCCATG				
					AGATTAT				
					TCATGAG				
					GTGAAAA				
					AGAAGAA				
					GCAGAGT				
					GGGGTGA				
					TTTTTAA				
					GGTGGAT				
					TTTGAGA				
					AAGCTTA				
					TGACAAG				
					GTTAATT				
					GGAACCT				
					TCTTCAT				
					CATATGA				
					TGATTAA				
					AAAAGGT				

					TTTGGTG				
					ACAAATG				
					GTGTGAC				
					TGGGTTA				
					TGAGAAC				
					TGTTAGA				
					GGTGGTA				
					AAGTGGC				
					CATCAAG				
					ACAAATG				
					ACAGGGT				
					TGGTCCC				
					TATTTTA				
					CCACTCA				
					CAAAGGT				
					GTCAGAC				
					AAGGTGA				
					TCCTTTCT				
					CCCCCTG				
					CTTTTTA				
					ATATTGC				
					TGCTGAT				
					GGGCTTG				

					CTTGTAT				
					GATTCAA				
					AAAGCCA				
					AAGATGA				
					AGGTATT				
					ATTAGGG				
					GACTCAT				
					TCCTCAT				
					ATCATCC				
					CTAATGG				
					TTGTTGTT				
					GTTTGCA				
					ATATGCT				
					GATGATA				
					CCATTTTT				
					TTGCTTC				
					AAGATGA				
					TCTGGAG				
					GGAGCCA				
					GAAATCT				
					TAAGTTT				
					ATTCTTT				
					GCTTATT				

					TGAACAG				
					ATGTCTG				
					GTCTCAA				
					AATCAAT				
					TTTCATA				
					AAAGTGA				
					AATTTTC				
					TGTTTAG				
					GTGAAGC				
					TGTTGAT				
					AAGGAAA				
					TGTGGTA				
					TGCTGAC				
					ATTTTTA				
					CCTGCCC				
					TTATAAT				
					TGTCTTC				
					CTATGAA				
					ATATTTG				
					GGGGTTC				
					CGATTGA				
					TGACAAA				
					AAACTTT				

					GCAAGTC				
					CTTGTGG				
					TTACCCA				
					CAGTAGA				
					GAAAGTT				
					GAAAAAT				
					GGGTTCT				
					TGGCAAG				
					GAAAATT				
					CCTTAGT				
					TTGGGTG				
					GTCGGTT				
					AGTCTTG				
					ATTAATA				
					GCAGTCT				
					TACCAAT				
					GTCCCTT				
					TATACAT				
					GCTCTCG				
					ATGTATA				
					AAGCTCC				
					CAAGAAT				
					ATTTTGA				

					AGAAAAT				
					GGATCTC				
					TTTAGAA				
					AAAGGCT				
					TTTGTGG				
					CAAGGTG				
					GCCATAC				
					TACAAAA				
					AAATATC				
					ATTTGGC				
					TGATTGG				
					AATATGG				
					TTTGTTCT				
					CCGAGAG				
					GTCAGGG				
					TGGCCTT				
					GGGGTCT				
					TGGATCT				
					GCACAAG				
					ATGAATG				
					AGGCGCT				
					TCTTGCT				
					AAATGGA				

					TTTGGA				
					GTTGGA				
					AATACGA				
					ATGGGCT				
					WTGGCAG				
					ACCATTA				
					TCAGACA				
					TAAATAT				
					GTTAAAG				
					GAAGGCC				
					TATTATTT				
					CATTAAA				
					AAAAGAC				
					AAAGTGA				
					TTCTCAC				
					TTTTGGA				
					AAGGTAT				
					TCTGGAT				
					GCTCGGG				
					ACAATTT				
					TTACAAA				
					TATTGTA				
					AAAAAAG				

					GTAGGCA				
					ATGGCAA				
					TAATACC				
					AGCTTTT				
					GGCACAA				
					TATTTGG				
					TGTGATT				
					TAGAACC				
					TCTGTCTT				
					CTAAATA				
					TCCTAGT				
					TTGTTTG				
					ATATTGC				
					TTATGAT				
					AAAGATA				
					TCACAGT				
					GGAAAAA				
					GCCTTTT				
					CTTCTGA				
					TTTTAGA				
					GCTGTGA				
					ATTTTAG				
					AAGAAGG				

					ATTTATG				
					GTGCTTT				
					GGAAGAT				
					GATTATA				
					ATAACCT				
					TATTACA				
					CAGTGTA				
					ATAATAT				
					TACTATT				
					TCTGAGG				
					AGGATGA				
					TAAATCA				
					GTTTGGC				
					TTCTGGG				
					TAATAAA				
					GGTTTTT				
					CTGTTAA				
					TTCTTTTT				
					ACAAAGA				
					AATCAAA				
					TGTGCTC				
					AGATTCC				
					TTTGCCTT				

					CTAGTTT				
					CTTGTGG				
					AAGACTA				
					GACTCCC				
					ACATAAA				
					ATCAAAG				
					CTTTTTTG				
					TGGCTTG				
					TTATGCA				
					TAAAAAG				
					ATTTTAA				
					CCAAGGA				
					TAATTG				
					TTCAAAA				
					AAAATTG				
					GAAGGGC				
					AATCTGG				
					ATTGTAC				
					TTTTTGTG				
					GATTTTC				
					AGAATCC				
					ATAGATC				
					ATTGTTT				

					TTCAGTG				
					CTCGGTT				
					GCTAGAT				
					TTACTTG				
					GAGAATA				
					GTCTACT				
					CGGCTCT				
					TGGCTTA				
					ATTTCTA				
					TTCCCAA				
					CAGTGTG				
					GATGATC				
					TGTTTGG				
					CTCTTGG				
					ATTAATA				
					GTTTTAA				
					CAAAACT				
					GAAAAGA				
					ATCTGGT				
					CCTTTTGG				
					GTTGTGG				
					AGCTGTC				
					ATTTGGG				

					CGATCTG				
					GCGATGT				
					CGCAATG				
					ATTGTTG				
					CTTTGGA				
					ACTGACA				
					GAATTGA				
					TGATCCG				
					ACTAATG				
					TGATTTT				
					CTCTTGTT				
					GCTATTG				
					GCTTGAT				
					ACTTGGG				
					CTATCCG				
					TCAGAAA				
					AAGAAGG				
					AAAGAAA				
					ATTGGTG				
					GAACAAG				
					GAAGCCT				
					CAAAATC				
					AGAAGGA				

					TGGCGTG				
					TGAGATC				
					TACAACA				
					AGACTCG				
					TGGTTGG				
					AAGCCAG				
					TCGACAG				
					ACGGATC				
					CAATGAA				
					GATTCTG				
					CTGTGTT				
					TGGCTAT				
					TTTGGTT				
					CATCAAC				
					TTTGTGTT				
					TGCTGTA				
					ATCCCCC				
					CTCGCCT				
					CGGTAGC				
					TTCAGGC				
					TACTGAA				
					AACTTTG				
					GGCTATA				

					TCGGGGY				
					TGAAGTT				
					GTTTAGT				
					TTCGTGC				
					TACTCTA				
					GTTTGAT				
					CTTTGCG				
					ATCTTTG				
					TTTTTGGT				
					CAGCTAT				
					GTGCTGA				
					TTCTGTG				
					CTGCTTC				
					TGGTAGT				
					TTGGTTC				
					G TTCAGG				
					TTTAGTT				
					CCGTCGT				
					AGCTGAT				
					AGTTGTG				
					TGTCTTG				
					TCGGCTT				
					TTAAATT				

					CTGGTTG				
					GCTCTAT				
					GGTCTTG				
					TAAGAAC				
					TTGTGTC				
					TCTTTCTA				
					TTTCGTA				
					ATGAAAT				
					AGGGGGT				
					TATCCCC				
					TCTGTCA				
					AAAAAA				

В таблице 44 представлены домены ретровирусной обратной транскриптазы для использования в полипептидах Gene Writer. Ферменты обратные транскриптазы дикого типа собирали и упорядочивали по приоритету в соответствии с описаниями в данном документе (см. пример 33). В столбце "Тип" указано, соответствует ли последовательность последовательности дикого типа ("корневой") или содержит мутации, которые могут улучшать активность фермента ("производный").

Таблица 44. Домены ретровирусной обратной транскриптазы для использования в полипептидах Gene Writer. В некоторых вариантах осуществления система или способ, описанные в данном документе, предусматривают домен обратной транскриптазы, имеющий аминокислотную последовательность, соответствующую домену обратной транскриптазы из таблицы 44, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней, или его функциональный фрагмент.

Название	Название_вируса	ID в UniProt	Тип	Пептид
----------	-----------------	--------------	-----	--------

AVIRE _P0336 0	AVIRE	P03360	корнево й	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWA EINPPGLASTQAPIHVQLLSTALPVRVRQYP ITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKR VETIHPTVP NPYTLLSLLPPDRIWYSVLDLKD AFFC IPLAPESQLIF AFEW ADAEEGESGQLTWTRL PQGFKNS PTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKK AQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQP LYAATRGGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVL TQAL GPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPD KWL TNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL TSTRPDLTDQ PLAQAEATLFTDGSSYIRDGKRYAGAAVVTLD SVIWAEPLIGTSAQKAELIALTKALEWSKD KSVNIYTDSRYAFATLHVHGMIYRERGLL TAGGKA IKN APEILALLTAVWLPKRVAVMHCKG HQQDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE _P0336 0_3mut	AVIRE	P03360	произво дный	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWA EINPPGLASTQAPIHVQLLSTALPVRVRQYP ITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKR VETIHPTVP NPYTLLSLLPPDRIWYSVLDLKD AFFC IPLAPESQLIF AFEW ADAEEGESGQLTWTRL PQGFKNS PTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKK AQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQP LYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVL TQAL GPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPD KWL TNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL TSTRPDLTDQ PLAQAEATLFTDGSSYIRDGKRYAGAAVVTLD SVIWAEPLIGTSAQKAELIALTKALEWSKD KSVNIYTDSRYAFATLHVHGMIYRERGWLT AGGKA IKN APEILALLTAVWLPKRVAVMHCKG HQQDDAPTSTGNRRADEVAREVAIRPLSTQATIS

AVIRE _P0336 0_3mut A	AVIRE	P03360	произво дный	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWAEINPPGLASTQAPIHVQLLSTALPVRVRQYP ITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLL PVRKSGTSEYRMVQDLREVNKR VETIHPTVP NPYTLLSLLPPDRIWYSVLDLKD AFFC IPLAPESQLIF AFEWADAEEGESGQLTWTRL PQGFKNS PTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKK AQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQP LYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVL TQAL GPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPD KWL TNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL TSTRPDLTDQ PLAQAEATLFTDGSSYIRDGKRYAGAAVVTLD SVIWAEPLIGTSAQKAELIALTKALEWSKD KSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKA IKNAP EILALLTAVWLPKRVA VMHCKG HQBDDAPTSTGNRRADEVAREVAIRPLSTQATIS
BAEV M_P10 272	BAEV M	P10272	корнево й	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPM SLEAHMGIRQHIIKFLELGVL RPCRSPWNTPLL PVKKPGTQDYRPVQDLREINKRTVDIHPTVPN PYNLLSTLKPDISWYTVLDLKD AFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRL PQGFKN SPTLFDEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPT K KACTQGTRHLLQELGEKGYRASAK KAQICQTKVTYLG YILSEGKRWLTPGRIETVARIPPPRNPREVREFLG TAGFCRLWIPGFAELAA PLYALTKESTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAGVLTQKLG PWKRVPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPD RWITNARLTHYQALLLDTDRVQFGPPVTLN PATLLVPENQPSPHDCRQVLAETHGTREDLKD QELPDADHTWYTDGSSYLD SGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELS KGKKANIYTDSRYAFATAH THGSIYERRGLLTSEGKEIKNKA EIALLKALFLPQEVAIIHCPGH QKGQDPVAVGNRQADRVARQAAMA EVLTLATEPDNTSHIT

BAEV M_P10 272_3 mut	BAEV M	P10272	произво дный	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPM SLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLL PVKKPGTQDYRPVQDLREINKRTVDIHPTVPN PYNLLSTLKPDISWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKN SPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAK KAQICQTKVTYLG YILSEGKRWLTPGRIETVARIPPPRNPREVREFLG TAGFCRLWIPGFAELAA PLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDT SKPFTLFLDERQGIAGVLTQKLG PWKR PVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPD RWITNARLTHYQALLLDTDRVQFGPPVTLN PATLLVPENQPSPHDCRQVLAETHGTREDLKD QELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPG TSAQKAELIALTKALELS KGKKANIYTDSRYAFATAHTHGSYERRGWL TSEGKEIKNKA EIALLKALFLPQEVAIIHCPGH QKGQDPVAVGNRQADRVARQAAMA EVLTLATEPDNTSHIT
BAEV M_P10 272_3 mutA	BAEV M	P10272	произво дный	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPM SLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLL PVKKPGTQDYRPVQDLREINKRTVDIHPTVPN PYNLLSTLKPDISWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKN SPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAK KAQICQTKVTYLG YILSEGKRWLTPGRIETVARIPPPRNPREVREFLGKAGFCRLFIPGFAELAA PLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDT SKPFTLFLDERQGIAGVLTQKLG PWKR PVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPD RWITNARLTHYQALLLDTDRVQFGPPVTLN PATLLVPENQPSPHDCRQVLAETHGTREDLKD QELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPG TSAQKAELIALTKALELS KGKKANIYTDSRYAFATAHTHGSYERRGWL TSEGKEIKNKA EIALLKALFLPQEVAIIHCPGH QKGQDPVAVGNRQADRVARQAAMA EVLTLATEPDNTSHIT

BLVA U_P25 059	BLVA U	P25059	корнево й	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTL PTPGGLQPHRRFAWRVLPQGFINSFALFERALQEPLRQVSAAFSQSLLVS YMDDILYVSPTEEQ RLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQ TVLGDLQWVSRGTPTRRPLQLLYSSLKGIDDPRAIIHLSPEQQQGIAELRQALSHNARSRYNE QEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSY AKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQF SPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPLNI WVDSKYLYSLLRTLVLGAWLQPDVPVPSYALLYKSLLRHPAIFVGHVRS SHSSASHPIASLNNYV DQL
BLVA U_P25 059_2 mut	BLVA U	P25059	произво дный	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTL PTPGGLQPHRRFAWRVLPQGFINSFALFQALQEPLRQVSAAFSQSLLVS YMDDILYVSPTEEQ RLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQ TVLGDLQWVSRGTPTRRPLQLLYSSLPIDDPRAIIHLSPEQQQGIAELRQALSHNARSRYNEQ EPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSYA KTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSP EPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPLNIWV DSKYLYSLLRTLVLGAWLQPDVPVPSYALLYKSLLRHPAIFVGHVRS SHSSASHPIASLNNYVDQL
BLVA U_P25 059_2 mutB	BLVA U	P25059	произво дный	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRVTNALTKPIPALSPGPPDLTAPPTHLP HIICLDLKDAFFQIPVEDRFRSYFAFT LPTPGGLQPHRRFAWRVLPQGFINSFALFQALQEPLRQVSAAFSQSLLVS YMDDILYVSPTEE QRLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQL

				QTVLGDLQWVSRGTPPTRPLQLLYSSLKPIDDPRAIHLSPEQQQGIAELRQALSHNARSRYNE QEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSY AKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQF SPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPLNI WVDSKYLYSLLRTLVLGAWLQPDVPVSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLNNYV DQL
BLVJ_ P03361	BLVJ	P03361	корнево й	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTL PSPGGLQPHRRFAWRVLPQGFINSFALFERALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQ RSQCYQALAAARLDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQA VLGDLQWVSRGTPPTRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQ EPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAK PILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDP IPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVD SKYLYSLLRTLVLGAWLQPDVPVSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
BLVJ_ P03361 _2mut	BLVJ	P03361	произво дный	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTL PSPGGLQPHRRFAWRVLPQGFINSFALFNRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQ RSQCYQALAAARLDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQA VLGDLQWVSRGTPPTRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQ EPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAK PILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDP

				IPAALCLFSDGATGRGAYCLWKDHLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVD SKYLYSLLRTWVLGAWLQPDVPVSYALLYKSLLRHPAIVVGHVRSASSHPIASLNNYVDQL
BLVJ_ P03361 _2mutB	BLVJ	P03361	произво дный	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIALSPGPPDLTAPPTHPPHICLDLKDFAFFQIPVEDRFRFYLSFTL PSPGGLQPHRRFAWRVLPQGFINSFALFQRALQEPLRQVSAAFSQSLLVSYMDDILYASPTTEEQ RSQCYQALAAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQA VLGDLQWVSRGTPPTRRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQ EPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAK PILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDP IPAALCLFSDGATGRGAYCLWKDHLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVD SKYLYSLLRTWVLGAWLQPDVPVSYALLYKSLLRHPAIVVGHVRSASSHPIASLNNYVDQL
FFV_O 93209	FFV	O93209	корнево й	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEPPVRQQNVTTIHGTQEGDVYYVNLK IDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLF EKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKE STMNTPVYPVPKPNRWRMVLDRVAVNKVTPLIAVQNNQHSYGILGSLFKGRYKTTIDLNGF WAHPVPEDYWITAFTWQKQYCWTVLPQGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYIS HDSEKEHLEYLDILFNRLKEAGYIISLKKSNANSIVDFLGFQITNEGRGLTDTFKEKLENITAPT TLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETITKLNGAEYLQG RKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTLEKLLTTVHKGLLKALD LSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPVAV DTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHT AQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKW KSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFVH

FFV_O 93209- Pro	FFV	O93209	произво дный	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPH KIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVL DYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQY CWTVLPQGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYI ISLKKSNANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIPDFTELI APLYALIPKSTKNYVPWQIEHSTTLETITLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYN EGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQT AKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAIT SPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILV VTDSNYVAKAYNEELD VWASNGFVNNRKKPLKHISKWKS VADLKRLRPDVVVTHEPGHQKL DSSPHAYGNNLADQLATQASFKVH
FFV_O 93209- Pro_2m ut	FFV	O93209	произво дный	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPH KIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVL DYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQY CWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYI ISLKKSNANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIPDFTELI APLYALIPKSPKNYVPWQIEHSTTLETITLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYN EGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQT AKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAIT SPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILV VTDSNYVAKAYNEELD VWASNGFVNNRKKPLKHISKWKS VADLKRLRPDVVVTHEPGHQKL DSSPHAYGNNLADQLATQASFKVH

FFV_O 93209- Pro_2m utA	FFV	O93209	произво дный	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPH KIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVL DYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDL SNGFWAHPIVPEDYWITAFTWQGKQY CWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYI ISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGKLN FARNFIPDFTEL IAPLYALIPKSPKNYVPWQIEHSTTLETITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYY NEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQ TAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPVDTGKDNKKHPSNFQHIFYTDGSAI TSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILV VTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKS VADLKRLRPDVVVTHEPGHQKL DSSPHAYGNNLADQLATQASFKVH
FFV_O 93209_ 2mut	FFV	O93209	произво дный	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEETVRQQNVTTIHGTQEGDVYYVNLK IDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLF EKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKE STMNTPVYPVPKPNGRWRMVL DYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDL SNGF WAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYIS HDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPT TLKQLQSILGKLN FARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETITKLNGAEYLQG RKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALD LSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPV DTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHT AQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKW KS VADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH

FFV_O 93209_ 2mutA	FFV	O93209	произво дный	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLK IDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLF EKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKE STMNTPVYPVPKPNGRWRMVL DYRAVNKV TPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGF WAHPIVPEDYWITAFTWQ GKQYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYIS HDSEKEHLEYLDILFNRLKEAGYIISLKKSNIAN SIVDFLG FQITNEGRGLTDTFKEKLENITAPT TLKQLQSILGKLN FARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETITKLNGAEYLQG RKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALD LSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLP AV DTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHT AQFAEIAAFEFA LKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKW KSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FLV_P 10273	FLV	P10273	корнево й	TLQLEEEYRLFEP ESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQY PMPHEAYQG IKPHIRMLDQGILKPCQSPWNTPLL PVKKPGTEDYRPVQDLREV NKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKD AFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQG FKNSPTLFD EALHSDLADFRVRYPALVLLQYVDDLLLAATRTECLEGTKALLET LGNKG YRA SAKKAQICLQEV TYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLG TAGYCRLWIPGF AELAAPLYPLTRPGTLFQWGTEQQLA FEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLV QKLG PWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLT LGQPLTILTSHPVEALVR QPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDC LQILAETHGTRP DLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALK MAEGKKLTVYTDSRYAFATTHVHGEIYRRRGLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCP GHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP

FLV_P 10273_ 3mut	FLV	P10273	произво дный	TLQLEEEYRLFEPSTQKQEMDIWLKNFPQAWAETGGMGT AHCQAPVLIQLKATATPISIRQY PMPHEAYQGIKPHIRMLDQGILKPCQSPWNTPLL PVKKPGTEDYRPVQDLREVNKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKD AFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQG FKNSPTL FNEALHSDLADFRVRYPALVLLQYVDDLLAAATRTECLEGTKALLET LGNKG YRA SAKKAQICLQEV TYLGYS LKDGQRWLT KARKEAILSIPVPKNSRQVREFLG TAGYCRLWIPGF AELAAPLYPLTRPGTLFQWGTEQQ LAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLV QKLG PWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLT LGQPLTILT SHPVEALVR QPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRP DLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPG TSAQRAELIALTQALK MAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHC PGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P 10273_ 3mutA	FLV	P10273	произво дный	TLQLEEEYRLFEPSTQKQEMDIWLKNFPQAWAETGGMGT AHCQAPVLIQLKATATPISIRQY PMPHEAYQGIKPHIRMLDQGILKPCQSPWNTPLL PVKKPGTEDYRPVQDLREVNKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKD AFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQG FKNSPTL FNEALHSDLADFRVRYPALVLLQYVDDLLAAATRTECLEGTKALLET LGNKG YRA SAKKAQICLQEV TYLGYS LKDGQRWLT KARKEAILSIPVPKNSRQVREFLG KAGYCRLFIPGFA ELAAPLYPLTRPGTLFQWGTEQQ LAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQ KLGPW KRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLT LGQPLTILT SHPVEALVRQ PPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPD LTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPG TSAQRAELIALTQALK MAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHC PGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP

FOAM V_P14 350	FOAM V	P14350	корнево й	MNPLQLLQPLPAEIKGTKLLAHWN SGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFK VKGRKVEAEVIASPYEYILLSPTDVPWL TQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTL FVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPESYWLT AFTWQGGKQYCWTRL PQGFLNSPALFTADVVDLLKEIPNVQVYVDDIY LSHDDPKEHVQQLEKVFQILLQAGYVVS LKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNI TPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNT ASNLEERLPEQRLVIK VNTSPSAGYVRY YNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHK ALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPEL KHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIP LGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLK HISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAM V_P14 350-Pro	FOAM V	P14350	произво дный	VPWL TQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIR PHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMV LDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLT AFTWQGGKQ YCWTRL PQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAG YVVS LKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNF AELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIK VNTSPSAGYV RY YNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSM TKI QKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYT DGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIP GPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGIS LQIPVFILKGNALADKLATQGSYVVN

FOAM V_P14 350- Pro_2m ut	FOAM V	P14350	произво дный	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIR PHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMV LDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLT AFTWQGKQ YCWTRL PQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAG YVVS LKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNF AELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYV RYYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSMTKI QKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYT DGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIP GPVLVITDSFYVAESANKELPYWKSNGFVNKKKPLKHISKWK SIAECLSMKPDITIQHEKGIS LQIPVFILKGNALADKLATQGSYVVN
FOAM V_P14 350- Pro_2m utA	FOAM V	P14350	произво дный	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIR PHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMV LDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLT AFTWQGKQ YCWTRL PQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAG YVVS LKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGKLN FARNFIPNF AELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYV RYYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSMTKI QKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYT DGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIP GPVLVITDSFYVAESANKELPYWKSNGFVNKKKPLKHISKWK SIAECLSMKPDITIQHEKGIS LQIPVFILKGNALADKLATQGSYVVN

FOAM V_P14 350_2 mut	FOAM V	P14350	произво дный	MNPLQLLQPLPAEIKGTKLLAHWN SGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFK VKGRKVEAEVIASPYEYILLSP TDVPWL TQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTL FVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGV LTP QNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPESYWLT AFTWQGKQYCWTRL PQGFLNSPALFNADVVDLLKEIPNVQVYVDDIY LSHDDPKEHVQQLEKVFQILLQAGYV VSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNI TPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNT ASNLEERLPEQRLVIK VNTSPSAGYVRYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHK ALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPEL KHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIP LGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLK HISKWKSIAECLSMKPDITIQHEKGISLQIPVFI LKGNALADKLATQGSYVVN
FOAM V_P14 350_2 mutA	FOAM V	P14350	произво дный	MNPLQLLQPLPAEIKGTKLLAHWN SGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFK VKGRKVEAEVIASPYEYILLSP TDVPWL TQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTL FVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGV LTP QNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPESYWLT AFTWQGKQYCWTRL PQGFLNSPALFNADVVDLLKEIPNVQVYVDDIY LSHDDPKEHVQQLEKVFQILLQAGYV VSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNI TPPKDLKQLQSILGKLN FARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNT ASNLEERLPEQRLVIK VNTSPSAGYVRYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHK ALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPEL KHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIP

				LGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLK HISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
GALV _P2141 4	GALV	P21414	корнево й	VLNLEEEYRLHEKVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSYTWYSVLDLKD AFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGF KNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRV SAKKAQLCQREV TYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGTAGFCRLWIPGF ASLAAPLYPLTKESIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLT LGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRR DLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALR LAEGKNINIYTD SRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPG HQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV _P2141 4_3mut	GALV	P21414	произво дный	VLNLEEEYRLHEKVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSYTWYSVLDLKD AFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGF KNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRV SAKKAQLCQREV TYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGTAGFCRLWIPGF ASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLT LGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRR DLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALR

				LAEGKNINIYDSRYAFATAHIHGAIYKQRGWLTSAAGKDIKNKEEILALLEAIHLPRRVAIHCP GHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV _P2141 4_3mut A	GALV	P21414	произво дный	VLNLEEEYRLHEKVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSYTWYSVLDLKD AFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGF KNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRV SAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGKAGFCRLFIPGF ASLAAPLYPLTKPSIPFIWTEEHQQA FDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLTLGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRR DLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALR LAEGKNINIYDSRYAFATAHIHGAIYKQRGWLTSAAGKDIKNKEEILALLEAIHLPRRVAIHCP GHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
HTL1A _P0336 2	HTL1A	P03362	корнево й	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWKVLPQGFKNSPTLFEMQLAHILQPIRQAF PQCTILQYMDDILLASPSHEDLLLLSEAT MASLISHGLPVSENKTQQTPTGIKFLGQIISP NHLTYDAVPTVPIRSRWALPELQALLGEIQWVS KGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIML TLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNI STQTFNQFIQTS DHPSVPILLHHSR FKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVII NTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFLD SKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALL ITPVLQL

HTL1A _P0336 2_2mut	HTL1A	P03362	производный	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWKVLPQGFKNSTPLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEA TMASLISHGLPVSENKTQQTPGTIKFLGQIISP NHL TYDAVPTVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHH NISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSP VIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIF LDSKYLHYLRTLALGTFQGRSSQAPFQALLPRLLSRKV VYLHHVRSHTNLPDPISRLNALTDA LLITPVLQL
HTL1A _P0336 2_2mut B	HTL1A	P03362	производный	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWKVLPQGFKNSTPLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEA TMASLISHGLPVSENKTQQTPGTIKFLGQIISP NHL TYDAVPTVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHH NISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSP VIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIF LDSKYLHYLRTLALGTFQGRSSQAPFQALLPRLLSRKV VYLHHVRSHTNLPDPISRLNALTDA LLITPVLQL
HTL1C _P1407 8	HTL1C	P14078	корнево й	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWRVLPQGFKNSTPLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEA

				TMASLISHGLPVSENKTQQTPGTIKFLGQIISP NHLTYDAVPKVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHH NISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPV IINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFL DSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDAL LITPVLQL
HTL1C _P1407 8_2mut	HTL1C	P14078	произво дный	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEA TMASLISHGLPVSENKTQQTPGTIKFLGQIISP NHLTYDAVPKVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHH NISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPV IINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFL DSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDAL LITPVLQL
HTL1C _P1407 8_2mut B	HTL1C	P14078	произво дный	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFI HDLRATNSLTIDLSSSSPGPPDLSSPPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYG PGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEA TMASLISHGLPVSENKTQQTPGTIKFLGQIISP NHLTYDAVPKVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHH

				NISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPV IINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFL DSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDAL LITPVLQL
HTL1L _P0C21 1	HTL1L	P0C211	корнево й	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHD RATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGT RYAWKVLPQGFKNSTPLFEMQLASILQPIRQAFPQCIVLQYMDDILLASPSPELQQLSEATMA SLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGT PTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTL TGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHNNISI QTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTA PCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKY LYHYLRTLALGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPI L
HTL1L _P0C21 1_2mut	HTL1L	P0C211	произво дный	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHD RATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGT RYAWKVLPQGFKNSTPLFQMQLASILQPIRQAFPQCIVLQYMDDILLASPSPELQQLSEATMA SLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGT PTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTL TGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHNNISI QTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTA PCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKY

				LYHYLRTLAWGTFQ GKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPI L
HTL1L _P0C21 1_2mut B	HTL1L	P0C211	произво дный	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDL RATNSLTVDLSSSSPGPPDLSSPPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGT RYAWKVLPQGFKNSTPLFQMQLASILQPIRQAFPPQCIVLQYMDILLASPSPEDLQQLSEATMA SLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGT PTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTL TGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISI QTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIINTA PCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKY LYHYLRTLAWGTFQ GKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPI L
HTL32 _Q0R5 R2	HTL32	Q0R5R2	корнево й	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALAKHIEPYQGPNNPIFPVKKPNGKWRFIHDR ATNSVTRDLASPSGPPDLTSLPQGLPHLRTIDLDAFFQIPLPTIFQPYFAFTLPQPNNGPGTR YSWRVLPQGFKNSTPLFEQQLSHILTPVRKTFPNSLIIQYMDILLASPAGELAALTDKVTNAL TKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTP VLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGT TAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQA LTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAP CLFSDGSASKAAFIWDRQVIHQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLI GHLRRMALGAFFGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPL LPL

HTL32 _Q0R5 R2_2m ut	HTL32	Q0R5R2	произво дный	GLEHLPPPPEVSQFPLNPERLQAL TDLVSRAL EAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLR ATNSVTRDLASPSGPPDL TSLPQGLPHLRTIDL TDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTR YSWRVLPQGFKNSPTLFQQQLSHIL TPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNAL TKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTP VLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGT TAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQA LTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAP CLFSDGSASKAAFIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLI GHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAP LLPL
HTL32 _Q0R5 R2_2m utB	HTL32	Q0R5R2	произво дный	GLEHLPPPPEVSQFPLNPERLQAL TDLVSRAL EAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLR ATNSVTRDLASPSGPPDL TSPQGLPHLRTIDL TDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTR YSWRVLPQGFKNSPTLFQQQLSHIL TPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNAL TKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTP VLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGT TAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQA LTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAP CLFSDGSASKAAFIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLI GHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAP LLPL
HTL3P _Q4U0 X6	HTL3P	Q4U0X6	корнево й	GLEHLPPPPEVSQFPLNPERLQAL TDLVSRAL EAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLR ATNSL TRDLASPSGPPDL TSLPQDLPHLRTIDL TDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTR YSWRVLPQGFKNSPTLFEQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNAL

				TKEGLPMSLEKTQATPGSIHFLGQVISPCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTP VLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGTT AVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALT YYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLF SDGATSKAAFILWDKQVIHQVLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIGH LRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLP L
HTL3P _Q4U0 X6_2m ut	HTL3P	Q4U0X6	произво дный	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLR ATNSLTRDLASPSGPPDLTSLPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTR YSWRVLPQGFKNSTPLFQQQLSHILAPVRKAFPNSLIQYMDDILLASPALRELTALTDKVTNA LTKEGLPMSLEKTQATPGSIHFLGQVISPCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGT PVLRSSHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGT TAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQAL TYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCL FSDGATSKAAFILWDKQVIHQVLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIG HLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPL LPL
HTL3P _Q4U0 X6_2m utB	HTL3P	Q4U0X6	произво дный	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLR ATNSLTRDLASPSGPPDLTSPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTR YSWRVLPQGFKNSTPLFQQQLSHILAPVRKAFPNSLIQYMDDILLASPALRELTALTDKVTNA LTKEGLPMSLEKTQATPGSIHFLGQVISPCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGT PVLRSSHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGT TAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQAL

				TYYLHTSDQSSVAILLQHSRHFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCL FSDGATSKAAFILWDKQVIHQVQLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIG HLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPL LPL
HTLV2 _P0336 3	HTLV2	P03363	корнево й	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRA TNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYA WTVLPQGFKNSPTLFEQQLAAVLNPMRKMFPSTIVQYMDDILLASPTNEELQQLSQLTLQAL TTHGLPISQEK TQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTL TELQVILGEIQWVSKGTPIL RKHLQSLYSALHGYRDPRACITLTPQQLHALHAIQQALQHNCRGRLNPALPLLGLISLSTSGTT SVIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMSKQAL CDFLRNSPHPSVGILIHMMGRFHNLSQPSGPWKTLHLPTLLQEPRLLRPIFTLSPVVLDTAPC LFS DGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKGELLALICGLRAAKPWPSLNIFLDSKY LIKYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLV PL
HTLV2 _P0336 3_2mut	HTLV2	P03363	произво дный	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRA TNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYA WTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPSTIVQYMDDILLASPTNEELQQLSQLTLQAL TTHGLPISQEK TQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTL TELQVILGEIQWVSKGTPIL RKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQQALQHNCRGRLNPALPLLGLISLSTSGTTS VIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMSKQALC DFLRNSPHPSVGILIHMMGRFHNLSQPSGPWKTLHLPTLLQEPRLLRPIFTLSPVVLDTAPCLF SDGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKGELLALICGLRAAKPWPSLNIFLDSKYLI

				KYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLVP L
HTLV2 _P0336 3_2mut B	HTLV2	P03363	произво дный	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRA TNAITTTLTSPSPGPPDLTSPPTALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYA WTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPSTIVQYMDDILLASPTNEELQQLSQLTLQAL TTHGLPISQEKTTQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTLTELQVILGEIQWVSKGTPIL RKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQQALQHNCRGRLNPALPLLGLISLSTSGTTS VIFQPKQNWPLAWLHTPHPPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFFHNMSKQALC DFLRNSPHPSVGILIHMMGRFHNLSQPSGPWKTLHLPTLLQEPRLLRPIFTLSPVVLDTAPCLF SDGSPQKAAYVLWDQTLQQDITPLPSHETHSAQKGELLALICGLRAAKPWPSLNIFLDSKYLI KYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLVP L
JSRV_ P31623	JSRV	P31623	корнево й	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKL SAAQQLVQEQLRLGHIEPSTSAWNSPIF VIKKKSGKWRLQLDLRKVNETMMHMGALQPGLPTPSAIPDKSYIIVIDLKDCFYTIPLAPQDCK RFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFQLYLVHYMDDILL AHTDEHLLYQAFSILKQHLSLNLVIADDEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHLKTL NDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLEGRTALQSIEEAI RQQQITYC DYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYVELVAKIIAKGRHEAIQYFGM EPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYP SDKLLQF ASSHAFIFPKIVRRQPIPE ATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNLF TDSSYVVGAL QMIETVPIIGTTSPEVLNLF TLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLT KQVFF QS

JSRV_ P31623 _2mutB	JSRV	P31623	произво дный	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKL SAAQQLVQEQLRLGHIEPSTSAWNSPIF VIKKKSGKWRL LQDLRKVNETMMHMGALQPGLTPSPIPKSYIIVIDLKDCFYTIPLAPQDCK RFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFQQLYL VHYMDDILL AHTDEHLLYQAFSILKQHL SLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQL VKLQTDHLKTL NDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPTLSLEGRTALQSIEEAI RQQQITYC DYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYEYELVAKIIAKGRHEAIQYFGM EPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYP SDKLLQF ASSHAFIFPKIVRRQPIPE ATLIFTDGSSNGTAALIINHQTYYAQT SFSSAQVVELFAVHQALLTVPTS FNLFTDSSYVVGAL QMIETVPIIGTTSPEVLNLF TLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLT KQVFF QS
KORV_ Q9TTC1	KORV	Q9TTC1	корнево й	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVY PWTTKRLLKIGKQVTHSFLVIPECAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLN LEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMS KEAREGIRPHIQRFDLGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKR VQDIHPTVPN PYNLLSSLPPSHTWYSVLDLKD AFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRL PQGFKN SPTLFDEALHRDLASFALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLQLSKLGYRVSA KKAQLCREEVTYLG YLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLG TAGFCRLWIPGFAS LAAPLYPLTREKVPFTWTEAHQEAFGR IKEALLSAPALALPDLTKPFALYVDEKEGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLT LGQNVLVIAPHNLESIVR QPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLR DQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEG TSAQKAELIALTQALRLA EGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGH QRGTDPVATGNRKAD EAAKQAAQSTRILTETTKN

KORV _Q9TT C1-Pro	KORV	Q9TTC1	произво дный	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVDIHTVPNPYNLLSSLPPSHTWYSVLDLKDAFFC LKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVM LQYVDDLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWL TPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAF GRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW PTCLKAIAAAVALLLKDADKLTGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNER VSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQ AGAAIVDNKRTVWASNLPEG TSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIIY KQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIHCPGHQRGTDVPVATGNRKADEAAKQAAQ STRILTETTKN
KORV _Q9TT C1- Pro_3m ut	KORV	Q9TTC1	произво дный	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVDIHTVPNPYNLLSSLPPSHTWYSVLDLKDAFFC LKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVM LQYVDDLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWL TPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAF GRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW PTCLKAIAAAVALLLKDADKLTGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNER VSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQ AGAAIVDNKRTVWASNLPEG TSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIIY

				KQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIHCPGHQRGTDPVATGNRKADEAAKQAAQ STRILTETTKN
KORV _Q9TT C1- Pro_3m utA	KORV	Q9TTC1	произво дный	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRQVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAAFFC LKLHPNSQPLFAFEWRDPEKGNTGQLTWTRL PQGFKN SPTLFNEALHRDLASFRALNPQVVM LQYVDDLLVAAPTYRDCKEGTRRLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWL TPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAF GRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW PTCLKAIAAVALLLKDADKLTGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNER VSFAPPAILNPATLLPVESDDTPIHCSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQ AGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTD SRYAFATAHVHGAIIY KQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIHCPGHQRGTDPVATGNRKADEAAKQAAQ STRILTETTKN
KORV _Q9TT C1_3m ut	KORV	Q9TTC1	произво дный	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVY PWTTKRLKIGQKQVTHSFLVIPECAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLN LEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMS KEAREGIRPHIQRFDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRQVQDIHPTVPN PYNLLSSLPPSHTWYSVLDLKDAAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRL PQGFKN SPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLQELSKLGYRVSA KKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLG TAGFCRLWIPGFAS LAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTGQNVLVIAPHNLESIVR

				QPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLR DQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEG TSAQKAELIALTQALRLA EGKSINIYTDSRYAFATAHVHGAIYKQRGWLT SAGKDIKNKEEILALLEAIHLPKRVAIIHCPGH QRGTD PVATGNRKADEAAKQAAQSTRILTETTKN
KORV _Q9TT C1_3m utA	KORV	Q9TTC1	произво дный	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVY PWTTKRLKIGQKQVTHSFLVIPECAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLN LEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMS KEAREGIRPHIQRFDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVDIHTVPN PYNLLSSLPPSHTWYSVLDLKD AFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKN SPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRLLQELSKLGYRVSA KKAQLCREEVTYLG YLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFAS LAAPLYPLTRPKVPFTWTEAHQEA FGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLK DADKLT LGQNV LVIAPHNLESIVR QPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLR DQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEG TSAQKAELIALTQALRLA EGKSINIYTDSRYAFATAHVHGAIYKQRGWLT SAGKDIKNKEEILALLEAIHLPKRVAIIHCPGH QRGTD PVATGNRKADEAAKQAAQSTRILTETTKN
MLVA V_P03 356	MLVA V	P03356	корнево й	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHT VPNPYNLLSGLPPSHRWYTVL DLKD AFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQG FKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELD CQQGTRALLTLGNLGYRAS AKKAQLCQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKGT LFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG

				VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAET HGTRPDLTDQIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIA LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVA V_P03 356_3 mut	MLVA V	P03356	произво дный	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSİKQY PMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLTLGNLGYRAS AKKAQLCQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAET HGTRPDLTDQIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIA LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPK RLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVA V_P03 356_3 mutA	MLVA V	P03356	произво дный	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSİKQY PMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLTLGNLGYRAS AKKAQLCQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAV

				EALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETETEVIVARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVB M_Q7S VK7_3 mutA_ WS	MLVB M	Q7SVK7	корнево й	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPPQGFKNSPTLTFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASKKAQICQKQVKYLGILLREGQRWLTEARKETVMGQPVPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETETEVIVAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVB M_Q7S VK7	MLVB M	Q7SVK7	произво дный	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPPQGFKNSPTLTFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASKKAQICQKQVKYLGILLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH

				GTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETETEVWAGALPAGTSAQRAELIAL TQALKMAEGKRLNVYTDSRYAFATAHIGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRL SIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVB M_Q7S VK7_3 mut	MLVB M	Q7SVK7	произво дный	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYP MSHEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASA KKAQICQKQVKYLGILLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLWIPGFA EMAAPLYPLTKPGTLFSWGPDQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH GTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETETEVWAGALPAGTSAQRAELIAL TQALKMAEGKRLNVYTDSRYAFATAHIGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVB M_Q7S VK7_3 mutA_ WS	MLVB M	Q7SVK7	корнево й	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYP MSHEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASA KKAQICQKQVKYLGILLREGQRWLTEARKETVMGQPVPKTPRQLREFLGKAGFCRLFIPGFA EMAAPLYPLTKPGTLFSWGPDQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH GTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETETEVWAGALPAGTSAQRAELIAL

				TQALKMAEGKRLNVYTDSRYAFATAHGHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVB M_Q7S VK7	MLVB M	Q7SVK7	произво дный	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYP MSHEARLGKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREV NKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGF KNSPTLFDEALHRDLADFRIQH PDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASA KKAQICQKQVKYLG YLLREGQRWL TEARKETVMGQPV PKTPRQLREFLGTAGFCRLWIPGFA EMAAPLYPLTKTGTLFSWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH GTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIAL TQALKMAEGKRLNVYTDSRYAFATAHGHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRL SIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVB M_Q7S VK7_3 mut	MLVB M	Q7SVK7	произво дный	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYP MSHEARLGKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREV NKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQH PDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASA KKAQICQKQVKYLG YLLREGQRWL TEARKETVMGQPV PKTPRQLREFLGTAGFCRLWIPGFA EMAAPLYPLTKPGTLFSWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH GTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIAL

				TQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVC B_P083 61	MLVC B	P08361	корнево й	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHDCLDILAEA HGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVC B_P083 61_3m ut	MLVC B	P08361	произво дный	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKPGTLFNWGPDPQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHDCLDILAEA HGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI

				ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGN SAEARGNRMADQAAREVATRETPETSTLL
MLVC B_P083 61_3m utA	MLVC B	P08361	произво дный	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTL FNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLGKAGFCRLFIPGFA EMAAPLYPLTKPGTLFNWGPDQKQAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDREVQFGPVVALNPATLLPLPEEGLQHDCLDILAEAH GTRSDLM DQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTEVIWARALPAGTSAQRAELIA LTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGN SAEARGNRMADQAAREVATRETPETSTLL
MLVF 5_P268 10	MLVF 5	P26810	корнево й	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQG FKNSPTL FDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGLCRLWIPGF AEMAAPLYPLTKTGTLFKWGPDQKQAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDREVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTEVIWAKALPAGTSAQRAELIAL

				TQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF 5_P268 10_3m ut	MLVF 5	P26810	произво дный	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGLCRLWIPGF AEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIAL TQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF 5_P268 10_3m utA	MLVF 5	P26810	произво дный	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGLCRLFIPGF AEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIAL

				TQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF F_P268 09	MLVF F	P26809	корнево й	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLG I KPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFD EALHRDLADFR IQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKGT LFEWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIA LTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVF F_P268 09_3m ut	MLVF F	P26809	произво дный	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLG I KPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFR IQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKPGTLFEWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIA

				LTQALKMAEGKKLVYTDSTRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVF F_P268 09_3m utA	MLVF F	P26809	произво дный	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLG I KPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDILAEAH GTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIA LTQALKMAEGKKLVYTDSTRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVM S_P033 55_PL V919	MLVM S	P03355	корнево й	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLG I KPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNC LDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEV I WAKALPAGTSAQRAELI

				ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFE
MLVM S_P033 55	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3m ut	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKPGTLFNWGPDPQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI

				ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3m utA_W S	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLNFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEVIAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_PL V919	MLVM S	P03355	корнево й	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLNFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEVIAKALPAGTSAQRAELI

				ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFE
MLVM S_P033 55	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEVIAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3m ut	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQG FKNSPTLFDALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETEVIAKALPAGTSAQRAELI

				ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3m utA_W S	MLVM S	P03355	произво дный	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLD TDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVR D_P11 227	MLVR D	P11227	корнево й	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIKQYP MSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQGLREVNKRVEDIHPT VPNPYNLLSGLPTSHRWYTVL DLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQG FKNSPTLFD EALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRAS AKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPRF AEMAAPLYPLTKTGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQAMLLD TDRVQFGPVVALNPATLLPLPEEGAPHDCLEILAET HGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIA

				LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGLLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVR D_P11 227_3 mut	MLVR D	P11227	произво дный	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYP MSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQGLREV NKRVEDIHPT VPNPYNLLSGLPTSHRWYTVL DLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQG FKNSPTLFNEALHRGLAD FRIQH PDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRAS AKKAQICQKQVKYLGYLLREGQRWL TEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPRF AEMAAPLYPLTKPGTLFNWGP DQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQAMLLD TDRVQFGPVVALNPATLLPLPEEGAPHDCLEILAET HGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIA LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGWL TSEGREIKNKSEILALLKALFLPK RLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVR D_P11 227_3 mutA	MLVR D	P11227	произво дный	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYP MSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQGLREV NKRVEDIHPT VPNPYNLLSGLPTSHRWYTVL DLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQG FKNSPTLFNEALHRGLAD FRIQH PDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRAS AKKAQICQKQVKYLGYLLREGQRWL TEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPRFA EMAAPLYPLTKPGTLFNWGP DQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV LTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQAMLLD TDRVQFGPVVALNPATLLPLPEEGAPHDCLEILAETH GTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIAL

				TQALKMAEGKRLNVYTDSRYAFATAHGHGEIYKRRGWLTSEGREIKNKSEILALLKALFLPKR LSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MMTV B_P033 65_WS	MMTV B	P03365	корнево й	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFADQI SWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQ DLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKR PYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILT SMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRNLNDFQKLLGNINW IRPFLKLTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCIL KTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKV QFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANG RSVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKI YTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQ QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRNLNDFQKLLGNIN WIRPFLKLTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSAN

				GRSVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRT KIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65-Pro	MMTV B	P03365	произво дный	GRDIMKDIKVRMLTDSRDDSQDLMIGAIESNLFADQISWKSQPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQLDLRAVNATMHDMGALQPGLPSPVAVP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLRTLNDQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTR KLTPACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPK VITPYDIFCTQLIIKGRHRSEKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLP KDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAIEIVA VITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHT GLPGPLAQGNAYADSLTRILT
MMTV B_P033 65- Pro_2m ut	MMTV B	P03365	произво дный	GRDIMKDIKVRMLTDSRDDSQDLMIGAIESNLFADQISWKSQPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQLDLRAVNATMHDMGALQPGLPSPVAVP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLRTLNDQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRK LTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVI TPYDIFCTQLIIKGRHRSEKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPK DPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAIEIVAVI TAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGL PGPLAQGNAYADSLTRILT

MMTV B_P033 65- Pro_2m utB	MMTV B	P03365	произво дный	GRDIMKDIK VRLMTDSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQLDLRAVNATMHDMGALQPGLPSPVAPP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRK LTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVI TPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPK DPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVI TAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGL PGPLAQGNAYADSLTRILT
MMTV B_P033 65_2m ut	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIK VRLMTDSPDDSQDLMIGAIESNLFAD QISWKS DQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL QLDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNIN WIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSAN GRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPR TKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIK VRLMTDSPDDSQDLMIGAIESNLFAD

65_2m utB				QISWKS DQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL QDLRAVNATMHDMGALQGPLSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNIN WIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTF TLQTAIIFPHMTSTTPLEKGIVIFTDGSAN GRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRT KIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65_2m utB_WS	MMTV B	P03365	произво дный	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFADQI SWKS DQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL DLRAVNATMHDMGALQGPLSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRP YQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTS MIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNINWI RPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILK TEYTPACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTF TLQTAIIFPHMTSTTPLEKGIVIFTDGSANR SVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIY TELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033	MMTV B	P03365	произво дный	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFADQI SWKS DQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL

65_2m ut_WS				DLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKR PYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILT SMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRRTLNDFQKLLGNINW IRPFLKLTTGELKPLFEILNPDSNPISTRKLTPACKALQLMNERLSTARVKRLDLSQPWSLCILK TEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIGRHRSELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGR SVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNL YTDSKYVTGLFPEIETATLSPRTKIY TELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65_WS	MMTV B	P03365	корнево й	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFADQI SWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQ DLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKR PYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILT SMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRRTLNDFQKLLGNINW IRPFLKLTTGELKPLFEILNGDSNPISTRKLTPACKALQLMNERLSTARVKRLDLSQPWSLCIL KTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIGRHRSELFSKDPDYIVVPYTKV QFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANG RSVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNL YTDSKYVTGLFPEIETATLSPRTKI YTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQ QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK

				RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNIN WIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIHKGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLL TFTLQTAIIFPHMTSTTPLEKGIVIFTDGSAN GRSVTYIQGREPIIKENTQNTAQQA EIVA VITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPR T KIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65-Pro	MMTV B	P03365	произво дный	GRDIMKDIK VRLMTDSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQDLRAVNATMHDMGALQPGLPSPVAVP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTR KLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPK VITPYDIFCTQLIHKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLP KDPLL TFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVA VITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPR TKIYTELKHLQRLIHKRQEKFYIGHIRGHT GLPGPLAQGNAYADSLTRILT
MMTV B_P033 65- Pro_2m ut	MMTV B	P03365	произво дный	GRDIMKDIK VRLMTDSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQDLRAVNATMHDMGALQPGLPSPVAVP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLR TLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRK LTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVI

				TPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPK DPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVI TAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGL PGPLAQGNAYADSLTRILT
MMTV B_P033 65- Pro_2m utB	MMTV B	P03365	произво дный	GRDIMKDIKVRMLMTDSPDDSQDLMIGAIESNLFADQISWKSQDPVWLNQWPLKQEKLQALQQ LVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQLDLRAVNATMHDMGALQPGLPSPVAPP KGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYL GTHIQGDSVSYQKLQIRTDKLRRTLNDQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRK LTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVI TPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPK DPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVI TAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGL PGPLAQGNAYADSLTRILT
MMTV B_P033 65_2m ut	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLMTDSPDDSQDLMIGAIESNLFAD QISWKSQDPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQL DLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRRTLNDQKLLGNIN WIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSAN

				GRSVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPR KIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65_2m utB	MMTV B	P03365	произво дный	WVQEISDSRPMLHIYLNRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVA RSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFAD QISWKSQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL QDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFK RPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEIL TSMIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRNLNDFQKLLGNIN WIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCI LKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIGRHRSKELFSKDPDYIVVPYTK VQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIFPHMTSTTPLEKGIVIFTDGSAN GRSVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPR KIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65_2m utB_W S	MMTV B	P03365	произво дный	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSPDDSQDLMIGAIESNLFADQI SWKSQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL DLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRP YQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTS MIQALNKHGLVVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRNLNDFQKLLGNINWI RPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILK TEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIFPHMTSTTPLEKGIVIFTDGSANGR

				SVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIY TELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65_2m ut_WS	MMTV B	P03365	произво дный	VQEISDSRPMLHIYLNRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARS SQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRMLTDSRDDSQDLMIGAIESNLFADQI SWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLQ DLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKR PYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILT SMIQALNKHGLVSTEKIQKYDNLKYLGTTHIQGDSVSYQKLQIRTDKLRNLDFQKLLGNINW IRPFLKLTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILK TEYTPTACLWQDGVVEWHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGR SVTYIQGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIY TELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MPMV _P0757 2	MPMV	P07572	корнево й	LTAADILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLAEGHITESSPWNT IFVIKKKSGKWRLQDLRAVNATMVLGMALQPGLPSPVAIPQGYLKIIDLKDCFFSIPLHPSDQ KRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDI LIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKL QTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKGDSDPNSHRSLSKEALASLEKVETAIAEQ FVTHINYSPLIFLIFNTALTPTGLFWQDNPMWIHLASPCKVLLPYYDAIADLIILGRDHSSKY FGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPNKLIFCKLHTFVFPQIISKTPL NNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLA HSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPPIAQGNQRADLATKIVA SNINT

MPMV _P0757 2_2mut	MPMV	P07572	произво дный	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQL EAGHITESSPWNTPIFVIKKKSGKWRLQLDLRAVNATMVL MGALQPGLPSPVAIPQGYLKI IIDLKDCFFSIPLHPSDQKRFAFSLPSTNFK EPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDI LIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKL QTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKPDSDPNSHRSL SKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPI MWIHLPASPKKVLLPYYDAIADLIILGRDH SKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYP PNKLIQFCKLHTFVFPQIISK TPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGP I AQGNQRADLATKIVASNINT
MPMV _P0757 2_2mut B	MPMV	P07572	произво дный	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQL EAGHITESSPWNTPIFVIKKKSGKWRLQLDLRAVNATMVL MGALQPGLPSPVAPPQGYLKI IIDLKDCFFSIPLHPSDQKRFAFSLPSTNFK EPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDI LIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKL QTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKPDSDPNSHRSL SKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPI MWIHLPASPKKVLLPYYDAIADLIILGRDH SKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYP PNKLIQFCKLHTFVFPQIISK TPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGP I AQGNQRADLATKIVASNINT
PERV_ Q4VFZ	PERV	Q4VFZ2	корнево й	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQR LIQQGILVPVQSPWNTPLL PVRKPGTNDYRPVQDLREVNKR VQDIHPTVPPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFK

2_3mut A_WS				NSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASA KKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFAT LAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNTVIAPHALENIVR QPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKD LTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQAL RLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIHC PGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
PERV_ Q4VFZ 2	PERV	Q4VFZ2	произво дный	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKRVDIHP TVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQ GFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPG FATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARG VLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNTVIAPHALEN IVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV RKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAI IHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut	PERV	Q4VFZ2	произво дный	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKRVDIHP TVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQ GFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR

				ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPG FATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARG VLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNITVIAPHALEN IVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV RKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLA IHCPCGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut A_WS	PERV	Q4VFZ2	корнево й	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPL SKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKRVDIHTVP NPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPPQGFK NSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASA KKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFAT LAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNITVIAPHALENIVR QPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKD LIDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQAL RLAEKKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIHC PGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
PERV_ Q4VFZ 2	PERV	Q4VFZ2	произво дный	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKRVDIHP TVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPPQ GFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPG

				FATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARG VLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNITVIAPHALEN IVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV RKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAI IHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut	PERV	Q4VFZ2	произво дный	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKRVDIHP TVPNPYNLLCALPPQRSWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRL PQ GFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLESDLGYR ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPG FATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARG VLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTGQNITVIAPHALEN IVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV RKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA GREIKNKEEILSLLEALHLPKRLA IIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
SFV1_ P23074	SFV1	P23074	корнево й	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTF KVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKEL LQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQG VLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLD LTNGFWAHPITPESYWLTAFTWQGKQYCWTRL PQGFLNSPALFTADVVDLLKEIPNVQAYVD DIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLN

				ITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIIISVLNQA DNLEERNPETRLIHKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLI KAMDLMGQEILVYSPIVSMTHKIQRTPPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQI PNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKS SHSAGMGIAQVQFIPEYKIVHQWSIPLGD HTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSK WKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_ P23074 -Pro	SFV1	P23074	произво дный	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRI KPHNIATGTLAPRPQKQYPINPKAKPSIQIVDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAF TWQGK QYCWTRLPPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAG YVVS LKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPN YSELVKPLYTIVANANGKFISWTEDNSNQLQHIIISVLNQADNLEERNPETRLIHKVNSSPSAGYIR YYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLMGQEILVYSPIVSMTHKIQ RTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTE DVIAKTKHPSEFAMVFYTD GSAIKHPDVNKS SHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGP VLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQ PMTTLHTEGNNLADKLATQGSYVVH
SFV1_ P23074 - Pro_2m ut	SFV1	P23074	произво дный	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRI KPHNIATGTLAPRPQKQYPINPKAKPSIQIVDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAF TWQGK QYCWTRLPPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAG YVVS LKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPN YSELVKPLYTIVAPANGKFISWTEDNSNQLQHIIISVLNQADNLEERNPETRLIHKVNSSPSAGYIR

				YYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQ RTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTD GSAIKHPDVNKS HSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGP VLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHHQ PMTTLHTEGNNLADKLATQGSYVVH
SFV1_ P23074 - Pro_2m utA	SFV1	P23074	произво дный	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRR KPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTVPVVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGK QYCWTRLPPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAG YVVSLLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGKLNFAFNFI PNIYSELVKPLYTIVAPANGKFISWTEGNSNQLQHISVLNQADNLEERNPETRLIIVNSSPSAGYIR YYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQ RTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTD GSAIKHPDVNKS HSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGP VLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHHQ PMTTLHTEGNNLADKLATQGSYVVH
SFV1_ P23074 _2mut	SFV1	P23074	произво дный	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTF KVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKEL LQKLFLKYDALWQHWENQVGHRRKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQG VLIQQNSTMNTVPVVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLD LTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPPQGFLNSPALFNADVVDLLKEIPNVQAYVD DIYISHDDPQEHLEQLEKIFSILLNAGYVVSLLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLN ITPPKDLKQLQSILGKLNFAFNFIPIYSELVKPLYTIVAPANGKFISWTEGNSNQLQHISVLNQ

				DNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLI KAMD LAMGQEILVYSPIVSMTKIQRTP LPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQI PNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGD HTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSK WKSIAECLQLKPDIIIMHEKGHQQPM TTLHTEGNNLADKLATQGSYVVH
SFV1_ P23074 _2mut A	SFV1	P23074	произво дный	MDPLQLLQPLEAEIKGTKLKAHWNSGATITC VPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTF KVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKEL LQKLFLKYDALWQHWENQVGHRRIKPHNIATGT LAPRPQKQYPINPKAKPSIQIVIDDLKQG VLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLD LTNGFWAHPITPESYWLTAFTWQ GKQYCWTRL PQGFLNSPALFNADVVDLLKEIPNVQAYVD DIYISHDDPQEHL EQLEKIF SILLNAGYVVS LKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLN ITPPKDLKQLQSILGKLN FARNFIPNYSELVKPLYTIVAPANGKFISWTE DNSNQLQHISVLNQA DNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLI KAMD LAMGQEILVYSPIVSMTKIQRTP LPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQI PNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGD HTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSK WKSIAECLQLKPDIIIMHEKGHQQPM TTLHTEGNNLADKLATQGSYVVH
SFV3L _P2740 1	SFV3L	P27401	корнево й	MDPLQLLQPLEAEIKGTKLKAHWNSGATITC VPQAFLEEEVPIKNIWIKTIHGEKEQP VYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQ SLFLKYDALWQHWENQVGHRRIKPHHIATGT VNPRPQKQYPINPKAKASIQT VINDLLKQGV L IQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSN GFWAHSITPESYWLTAFTWL GQQYCWTRL PQGFLNSPALFTADVVDLLKEVPNVQVYVDDIY ISHDDPREHLEQLEKVFSLLLNAGYVVS LKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNIT

				PPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQQLQNIISMLNSAEN LEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNIVYTKAEVKFTNTEKLLTTIHKGLIK ALDLGMGQEILVYSPIVSMTKIQTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVP TVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKS HNAGMGIAQVQFKPEFTVINTWSIPLGDHT AQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L _P2740 1-Pro	SFV3L	P27401	произво дный	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIK PHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGGQ YCWTRLPPQGFLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAG YVVS LKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFARNFIPNF SELVKPLYNIIATANGKYITWTTDNSQQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIR FYNEFAKRPIMYLNIVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQT TPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGS AIKHPNVNKS HNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPV LIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIHEKGHQPTAS TFHTEGNNLADKLATQGSYVVN
SFV3L _P2740 1- Pro_2m ut	SFV3L	P27401	произво дный	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIK PHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGGQ YCWTRLPPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAG YVVS LKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFARNFIPNF SELVKPLYNIIATAPGKYITWTTDNSQQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIR

				FYNEFAKRPIMYLNLYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQK TPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGS AIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVA AVEFACKKALKIDGPV LIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWWSIADCIQLKPDIIIHEKGHQPTAS TFHTEGNNLADKLATQGSYVVN
SFV3L _P2740 1- Pro_2m utA	SFV3L	P27401	произво дный	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIK PHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGGQ YCWTRLPPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAG YVVSLLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGKLN FARNFIPNF SELVKPLYNIIATAPGKYITWTTDNSQQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIR FYNEFAKRPIMYLNLYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQK TPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGS AIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVA AVEFACKKALKIDGPV LIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWWSIADCIQLKPDIIIHEKGHQPTAS TFHTEGNNLADKLATQGSYVVN
SFV3L _P2740 1_2mut	SFV3L	P27401	произво дный	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQ SLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGV L IQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSN GFWAHSITPESYWLTAFTWLGGQYCWTRLPPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIY ISHDDPREHLEQLEKVFSLLL NAGYVVSLLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNIT PPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQQLQNIISMLNSAEN

				LEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYYVYTKAEVKFTNTEKLLTTIHKGLIK ALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHDKTLPELQQVP TVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKS HNAGMGIAQVQFKPEFTVINTWSIPLGDHT AQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L _P2740 1_2mut A	SFV3L	P27401	произво дный	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQ SLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGV LQQNSIMNTPVYPVPKPDGKWRMVLVDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDSL N GFWAHSITPESYWLTAFTWLGQQYCWTRL PQGFLNSPALFNADVVDLLKEVPNVQVYVDDIY ISHDDPREHLEQLEKVFSLLLNAGYVVSLLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNIT PPRDLKQLQSILGKLN FARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQQLQNIISMLNSAEN LEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYYVYTKAEVKFTNTEKLLTTIHKGLIK ALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHDKTLPELQQVP TVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKS HNAGMGIAQVQFKPEFTVINTWSIPLGDHT AQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFVCP _Q8704 0	SFVCP	Q87040	корнево й	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFK VKGRKVEAEVIASPYEYILLSPDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKA LFTKYDNLWQHWENQVGHRRIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT P QNSTMNTPVYPVPKPDGRWRMVLVDYREVNKTIPLTAAQNAQHSAGILATIVRQKYKTTLDLAN GFWAHPITPDSYWLTAFTWQGKQYCWTRL PQGFLNSPALFTADAVDLLKEVPNVQVYVDDIY LSHDNPHEHIQQLEKVFQILLQAGYVVSLLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNV

				TPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLN YVFSKAELKF SMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLP ELKHIPDVYTSSIPPLKHPSQYEGVFCTDGS AIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q8704_0-Pro	SFVCP	Q87040	производный	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLT AFTWQGKQYCWTRL PQGFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLN YVFSKAELKF SMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLP ELKHIPDVYTSSIPPLKHPSQYEGVFCTDGS AIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q8704_0-Pro_2mut	SFVCP	Q87040	производный	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLT AFTWQGKQYCWTRL PQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAG

				YVRYYNESGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLP ELKHIPDVYTSSIPPLKHPSQYEGVFC TDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKV PGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQ PINTSIHTEGNALADKLATQGSYVVN
SFVCP _Q8704 0- Pro_2m utA	SFVCP	Q87040	произво дный	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIR PHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT PQNSTMNTPVYPVPKPDGRWRMV LDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPDSYWLTAF TWQGK QYCWTRL PQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQA GYVVS LKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVT PPKDLKQLQSILGKLN FARNFIP NFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAG YVRYYNESGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLP ELKHIPDVYTSSIPPLKHPSQYEGVFC TDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKV PGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQ PINTSIHTEGNALADKLATQGSYVVN
SFVCP _Q8704 0_2mut	SFVCP	Q87040	произво дный	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFK VKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKA LFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT P QNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPDSYWLTAF TWQGKQYCWTRL PQGFLNSPALFNADAVDLLKEVPNVQVYVDDI YLSHDNPHEHIQQLEKVFQILLQAGYVVS LKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLN VTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNT

				ASNLEERLPDQRLVIKVENTSPSAGYVRYYNESGKKPIMYLN YVFSKAELKFSMLEKLLTTMHK ALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPEL KHIPDVYTSSIPPLKHPSQYEGVFCTDGS AIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPL GHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKH ISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLATQGSYVVN
SFVCP _Q8704 0_2mut A	SFVCP	Q87040	произво дный	MNPLQLLQPLPAEVKGTKLLAHWN SGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFK VKGRKVEAEVIASPYEYILLSP TDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQ LKA LFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLT P QNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLAN GFWAHPITPDSYWLTAFTWQ GKQYCWTRL PQGFLNSPALFNADAVDLLKEVPNVQVYVDDI YLSHDNPHEHIQQLEKVFQILLQAGYV VSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLN VTPPKDLKQLQSILGKLN FARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNT ASNLEERLPDQRLVIKVENTSPSAGYVRYYNESGKKPIMYLN YVFSKAELKFSMLEKLLTTMHK ALIKAMD LAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPEL KHIPDVYTSSIPPLKHPSQYEGVFCTDGS AIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPL GHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKH ISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLATQGSYVVN
SMRV H_P03 364	SMRV H	P03364	корнево й	PRSRADIPVPHADKISWKITDPVWVDQWPLTYEKT LA AIALVQEQLAAGHIEPTNSPWNTPIFII KKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYF AFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLAC DSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPF SYLGFELHHQQVFTPRVCLKTDHLKTLND FQKLLGDIQWLRPYLKLPTSALVPLNNILKGDPNPLSVRALTPEAKQSLALINKAIQNQSVQQIS YNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYP SLLAMLIHKGRY

				TGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSPDVIQFAKLHQFIFPKI TKCAPIQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTDS AYIAQSVPLLETVPFIKSSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGLAEGNALADAA TQIFPIISD
SMRV H_P03 364_2 mut	SMRV H	P03364	произво дный	PRSRaidIPVPHADKISWKITDPVWVDQWPLTYEKTAAIALVQEQLAAGHIEPTNSPWNTPIFII KKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYF AFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLAC DSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPF SYLGFELHHQQVFTPRVCLKTDHLKTLND FQKLLGDIQWLRPYLKLPTSALVPLNNILKPDNPLSVRALTPEAKQSLALINKAIQNQSVQQIS YNLPLVLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSELLAMLIKGRY TGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSPDVIQFAKLHQFIFPKI TKCAPIQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTDS AYIAQSVPLLETVPFIKSSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGLAEGNALADAA TQIFPIISD
SMRV H_P03 364_2 mutB	SMRV H	P03364	произво дный	PRSRaidIPVPHADKISWKITDPVWVDQWPLTYEKTAAIALVQEQLAAGHIEPTNSPWNTPIFII KKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAPPLNYHKIVIDLKDCFFTIPLHPEDRPY FAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLA CDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPF SYLGFELHHQQVFTPRVCLKTDHLKTLN DFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDNPLSVRALTPEAKQSLALINKAIQNQSVQQI SYNLPLVLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSELLAMLIKGR YTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSPDVIQFAKLHQFIFPK ITKCAPIQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTD

				SAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPPFIGHLRAHLNLPGPLAEGNALADA ATQIFPIISD
SRV1_ P04025	SRV1	P04025	корнево й	LTAAIDMLAPQQCAEPITWKSDEPVWVDQWPLTSEKLAAAQQLVQEQL EAGHITESNSPWNT PIFVIKKKSGKWRLQLDLRAVNATMVLMGALQPGLPSPVAIPQGYLKIIDLKDCFFSIPLHPSD QKRFAFSLPSTNFKPEMQRFQWKVLPQRMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMD DILIAGKDGQQVLQCFDQLKQELTIAGLHIAPEKIQQLQDPYTYLGFELNGPKITNQKAVIRKDKL QTLNDFQKLLGDINWLRPYLKLTTADLKPLFDTLKGDSNPNSHRSL SKEALALLDKVETAIAE QFVTHINYSLPLMFLIFNTALTPTGLFWQNNPIMWVHLPASPKKVLLPYDAIADLIILGRDHSK KYFGIEPSVIIQPYSKSQIDWLMQNTEMWPIACASYVGILDNHYPNKL IQFCKLHAFIFPQIISK TPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSA YLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAHGNQKADLAT KTVASNINT
SRV1_ P04025 _2mutB	SRV1	P04025	произво дный	LTAAIDMLAPQQCAEPITWKSDEPVWVDQWPLTSEKLAAAQQLVQEQL EAGHITESNSPWNT PIFVIKKKSGKWRLQLDLRAVNATMVLMGALQPGLPSPVAPPQGYLKIIDLKDCFFSIPLHPSD QKRFAFSLPSTNFKPEMQRFQWKVLPQRMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMD DILIAGKDGQQVLQCFDQLKQELTIAGLHIAPEKIQQLQDPYTYLGFELNGPKITNQKAVIRKDKL QTLNDFQKLLGDINWLRPYLKLTTADLKPLFDTLKGDSNPNSHRSL SKEALALLDKVETAIAE QFVTHINYSLPLMFLIFNTALTPTGLFWQNNPIMWVHLPASPKKVLLPYDAIADLIILGRDHSK KYFGIEPSVIIQPYSKSQIDWLMQNTEMWPIACASYVGILDNHYPNKL IQFCKLHAFIFPQIISK TPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSA YLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAHGNQKADLAT KTVASNINT

SRV2_ P51517	SRV2	P51517	корнево й	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIIESNSPWNT PIFVIKKKSGKWRLQLDLRAVNATMVLMGALQPGLSPSPAIPQGYFKIVIDLKDCFFTIPLQPV DQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYM DDILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRD KLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRSLSEAALASLQKVETAIA EQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPCKVLLPYDAIADLIILGRDN SKKYFGLEPSTIIQPYSKSIHWLMQNTETWPIACASYAGNIDNHYPNKLIFCKLHAVVFPRI ISKTPLDNALLVFTDGSSTGIAAYTFEKT TVRFKTSHTSAQLVELQALIAVLSAFPHRALNVYTD SAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGLSQGNHITDLAT KVVATTLTT
SRV2_ P51517 _2mutB	SRV2	P51517	произво дный	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIIESNSPWNT PIFVIKKKSGKWRLQLDLRAVNATMVLMGALQPGLSPSPAAPPQGYFKIVIDLKDCFFTIPLQPV DQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYM DDILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRD KLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRSLSEAALASLQKVETAIA EQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPCKVLLPYDAIADLIILGRDN SKKYFGLEPSTIIQPYSKSIHWLMQNTETWPIACASYAGNIDNHYPNKLIFCKLHAVVFPRI ISKTPLDNALLVFTDGSSTGIAAYTFEKT TVRFKTSHTSAQLVELQALIAVLSAFPHRALNVYTD SAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGLSQGNHITDLAT KVVATTLTT
WDSV _O9281 5	WDSV	O92815	корнево й	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPL ISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPSLH WFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFSQALYQSLHKIKFKISS

				EICIMDDVLIASKDRDTNLKDTAVMLQHLLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKI LPDRKVTVSQFQQPTTIRQIRAFGLVGYCRHWIPEFSIHSKFLEKQLKKDTAEPFQLDDQQVE AFNKLKHAITTAAPVLVVPDPAKPFQLYTSHSEHASIAVLTKKHAGRTRPIAFLSSKFDAIESGLP PCLKACASIHRSLTQADSFILGAPLIYTTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVA CSAVSPAHL YMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPMTLFSDGSYTTGRGGA AV VMHRPVTDDFIIHHQQPGGASQAELLAALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLW MHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAV FLVQR
WDSV _O9281 5_2mut	WDSV	O92815	произво дный	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPL ISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTA VVASPTTVLSNLAPSLH WFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKIS SEICIMDDVLIASKDRDTNLKDTAVMLQHLLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGR KILPDRKVTVSQFQQPTTIRQIRAFGLVGYCRHWIPEFSIHSKFLEKQLKPDTAEPFQLDDQQV EAFNKLKHAITTAAPVLVVPDPAKPFQLYTSHSEHASIAVLTKKHAGRTRPIAFLSSKFDAIESGL PPCLKACASIHRSLTQADSFILGAPLIYTTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFV ACSAVSPAHL YMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPMTLFSDGSYTTGRGGA A VVMHRPVTDDFIIHHQQPGGASQAELLAALAAACHLATDKTVNIYTDSRYAYGVVHDFGHL WMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNA VFLVQR
WDSV _O9281 5_2mut A	WDSV	O92815	произво дный	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPL ISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTA VVASPTTVLSNLAPSLH WFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKIS SEICIMDDVLIASKDRDTNLKDTAVMLQHLLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGR

				KILPDRKVTVSQFQQPTTIRQIRAF LGKVG YCRHFIPEFSIHSKFLEKQLKPDTAEPFQLDDQQV EAFNKLKHAIT TAPVLVVPDPAKPFQLYTS HSEHASIAVL TQKHAGRTRPIAFLSSKFDAIESGL PPCLKACASIHRS LTQADSFILGAPLIYTT HAICTLLQRDRSQLVTASRFSKWEADLLRPELTFV ACSAVSPAHL YMQSCENNIPPHDCVLLTHTISRPRPDLSDLPDPDMTLFSDGSYTTGRGGAA VVMHRPVTDDFIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHL WMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNA VFLVQR
WMSV _P0335 9	WMSV	P03359	корнево й	VLNLEEEYRLHEKPVPSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQRFDLGLVLPQCSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSHTWYSVLDLKD AFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGF KNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRV SAKKAQLCQKEVTYLG YLLKEGKRWLTPARKATVMKIPPTTPRQVREFLGTAGFCRLWIPGF ASLAAPLYPLTKESIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLT LGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLN PATLLPVESEATPVHRCSEILAEETGTRR DLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQAL RLAEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCP GHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV _P0335 9_3mut	WMSV	P03359	произво дный	VLNLEEEYRLHEKPVPSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQRFDLGLVLPQCSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSHTWYSVLDLKD AFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGF KNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRV SAKKAQLCQKEVTYLG YLLKEGKRWLTPARKATVMKIPPTTPRQVREFLGTAGFCRLWIPGF

				ASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLTGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRR DLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQAL RLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAAGKDIKNKEEILALLEAIHLPKRVAIIHC PGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV _P0335 9_3mut A	WMSV	P03359	произво дный	VLNLEEEYRLHEKVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYP MSKEAREGIRPHIQRFDLGLVLCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTV PNPYNLLSSLPPSHTWYSVLDLKD AFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGF KNSPTLFNEALHRDLAPFRALNPQVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRV SAKKAQLCQKEVTYLG YLLKEGKRWLTPARKATVMKIPPPTTPRQVREFLGKAGFCRLFIPGF ASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPTCLKAVA AVALLLKDADKLTGQNVTVIASHSLESIV RQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRR DLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQAL RLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAAGKDIKNKEEILALLEAIHLPKRVAIIHC PGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
XMRV 6_A1Z 651	XMRV 6	A1Z651	корнево й	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLG IKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRA SAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPG FAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK

				GVL TQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHA VEALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEKEAPHDCLEILAE THGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGLLTSEGREIKNKNEILALLKALFLP KRLSIIHCPGHQKGN SAEARGNRMADQAAREAAMKAVLETSTLL
XMRV 6_A1Z 651_3 mut	XMRV 6	A1Z651	произво дный	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRA SAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG TAGFCRLWIPG FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVL TQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHA VEALVKQPPDRWLSNARMTHYQAMLLDTRVQFGPVVALNPATLLPLPEKEAPHDCLEILAE THGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLP KRLSIIHCPGHQKGN SAEARGNRMADQAAREAAMKAVLETSTLL
XMRV 6_A1Z 651_3 mutA	XMRV 6	A1Z651	произво дный	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRA SAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG KAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV

			EALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDCLEILAET HGTRPDLTDQIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIA LTQALKMAEGKKLVYTDSDRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPK RLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL
--	--	--	--

В некоторых вариантах осуществления домены обратной транскриптазы являются модифицированными, например, посредством сайт-специфической мутации. В некоторых вариантах осуществления домены обратной транскриптазы сконструированы таким образом, что они имеют улучшенные свойства, например, обратная транскриптаза SuperScript IV (SSIV), полученная из RT MMLV. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он характеризуется более низкой частотой ошибок, например, как описано в WO2001068895, включенной в данный документ посредством ссылки. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он является более термостабильным. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он является более процессивным. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он характеризуется толерантностью к ингибиторам. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он является более быстрым. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он характеризуется лучшей толерантностью к модифицированным нуклеотидам в РНК-матрице. В некоторых вариантах осуществления домен обратной транскриптазы может быть сконструирован таким образом, что он вставляет модифицированные нуклеотиды ДНК. В некоторых вариантах осуществления домен обратной транскриптазы сконструирован таким образом, что он связывает РНК-матрицу. В некоторых вариантах осуществления одна или несколько мутаций выбраны из D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K или D653N в домене RT обратной транскриптазы вируса лейкоза мышей или соответствующей мутации в соответствующем положении другого домена RT. В некоторых вариантах осуществления одна или несколько мутаций выбраны согласно описанному в WO2018089860A1, включенной в данный документ посредством ссылки (например, мутация C952S, и/или C956S, и/или C952S, C956S (двойная мутация), и/или C969S, и/или H970Y, и/или R979Q, и/или R976Q, и/или R1071S, и/или R328A, и/или R329A, и/или Q336A, и/или R328A, R329A, Q336A (тройная мутация), и/или G426A, и/или D428A, и/или G426A, D428A (двойная мутация), и/или любая их комбинация; где положения указаны относительно SEQ ID NO: 52 из WO2018089860A1), в домене RT ретротранспозазы R2Bm или соответствующей мутации в соответствующем положении другого домена RT.

В некоторых вариантах осуществления домен RT обладает корректирующей активностью. В некоторых вариантах осуществления домен RT эволюционировал от ДНК-зависимой ДНК-полимеразы и приобрел РНК-зависимую ДНК-полимеразную активность. Синтетическая эволюционировавшая корректирующая RT, известная как ксенополимераза обратной транскрипции (RTX, GenBank: QFN49000.1), была ранее получена путем использования ДНК-зависимой ДНК-полимеразы (KOD, GenBank: ABN15964.1) и отбора

по РНК-зависимой ДНК-полимеразной активности (Ellefson et al 2016). В некоторых вариантах осуществления сконструированная RT может содержать сигнатуры ДНК-зависимой ДНК-полимеразы, полученные из фермента дикого типа, например, IPR006134, PF00136, cd05536.

В некоторых вариантах осуществления домен обратной транскрипции осуществляет распознавание и выполняет обратную транскрипцию только определенной матрицы. В некоторых вариантах осуществления матрица содержит специфические последовательности. В некоторых вариантах осуществления матрица содержит включение UTR, которая обеспечивает ассоциацию нуклеиновой кислоты с доменом обратной транскриптазы (например, нетранслируемую область (UTR) из ретротранспозона, например, 3'-UTR ретротранспозона R2).

Записывающий домен может также обладать ДНК-зависимой ДНК-полимеразной активностью, например, обладать ферментативной активностью, способной обеспечивать записывание ДНК в геном из последовательности ДНК-матрицы. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность обеспечивается ДНК-полимеразным доменом в полипептиде. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность обеспечивается доменом обратной транскриптазы, который также способен к ДНК-зависимой полимеризации ДНК, например, к синтезу второй нити.

В некоторых вариантах осуществления записывающий домен (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 записывающего домена.

В отличие от других типов механизмов обратной транскрипции, например, ретровирусных RT и ретротранспозонов, содержащих LTR, обратная транскрипция в ретротранспозонах, не содержащих LTR, таких как R2, выполняется только на РНК-матрицах, содержащих специфические последовательности распознавания. Ретротранспозазе R2 требуется, чтобы ее матрица содержала минимальную 3'-UTR-область для инициации TPRT (Luan and Eickbush Mol Cell Biol 15, 3882-91 (1995)). В некоторых вариантах осуществления полипептид Gene Writer получен из ретротранспозазы с требуемым связывающим мотивом, и РНК-матрица содержит указанный связывающий мотив, так что специфическая ретротранспозиция происходит только с желаемой матрицей. В некоторых вариантах осуществления полипептид Gene Writer получен из ретротранспозона, выбранного из таблицы 3, и 3'-UTR в РНК-матрице содержит 3'-UTR из того же ретротранспозона из таблицы 3.

14. 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 использовании 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.

15. Эндонуклеазный домен

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 FokI, эндонуклеаза, подобная ферменту рестрикции I II типа (нуклеаза RLE-типа), или другая эндонуклеаза RLE-типа (также известная как REL). В некоторых вариантах осуществления гетерологичная эндонуклеаза обладает никазой 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%, по меньшей мере приблизительно 60%, по меньшей мере приблизительно 70%, по меньшей мере приблизительно 80%, по меньшей мере приблизительно 85%, по меньшей мере приблизительно 90%, по меньшей мере приблизительно 95%, по меньшей мере приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98%, по меньшей мере приблизительно 99% идентичной аминокислотной последовательности эндонуклеазного домена ретротранспозона, последовательность ДНК

которого указана в таблице 1 или 3. Средний специалист в данной области способен идентифицировать эндонуклеазные домены 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 FokI 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* -- SsoI 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 одностранных ДНК, например, только нуклеазной активностью, например, является нуклеазой 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 FokI или его функциональный фрагмент. В еще нескольких других вариантах осуществления домены гомологичной эндонуклеазы 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 эндонуклеазный домен обладает нуклеазной активностью и не образует двухнитевые разрывы. В некоторых вариантах осуществления эндонуклеазный домен образует одностранные разрывы с большей частотой, чем двухнитевые разрывы, например, по меньшей мере 90%, 95%, 96%, 97%, 98% или 99% разрывов являются одностранными 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 обнаруживаемых уровнях.

В некоторых вариантах осуществления эндонуклеазный домен обладает нуклеазной активностью, которая обеспечивает внесение одностранных разрывов в ДНК сайта-мишени в первой нити; 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 нуклеазной активностью, которая обеспечивает внесение одностранных разрывов в ДНК сайта-мишени в первой нити и не обеспечивает внесение одностранных разрывов в ДНК сайта-мишени во второй нити. 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 одностебельные разрывы в ДНК сайта-мишени в нити, содержащей сайт РАР (например, и не вносит одностебельные разрывы в ДНК сайта-мишени в нити, не содержащей сайт РАР). В качестве дополнительного примера, 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 одностебельные разрывы в ДНК сайта-мишени в нити, не содержащей сайт РАР (например, и не вносит одностебельные разрывы в ДНК сайта-мишени в нити, содержащей сайт РАР).

В некоторых других вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение одностебельных разрывов в ДНК сайта-мишени в первой нити и второй нити. 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 одностебельный разрыв в первой нити ДНК. ДНК сайта-мишени теперь содержит две разные последовательности для ДНК первой нити: одна соответствует исходной геномной ДНК, и вторая соответствует той, которая полимеризована из последовательности гетерологичного объекта. Считается, что две разные последовательности уравниваются друг с другом, где сначала первая из них гибридизируется со второй 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 одностебельного разрыва во вторую 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 одностебельный 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 одностебельного разрыва в первой нити.

В качестве альтернативы или дополнительно, не желая ограничиваться какой-либо теорией, полагают, что дополнительный одностебельный разрыв во второй нити может способствовать синтезу второй нити. В некоторых вариантах осуществления, где Gene Writer™ вставил или заменил часть первой нити, необходим синтез новой последовательности, соответствующей вставке/замене во второй нити.

В некоторых вариантах осуществления полипептид содержит один домен, обладающий эндонуклеазной активностью (например, один эндонуклеазный домен), и

указанный домен вносит односторонний разрыв как в первой нити, так и во второй нити. 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 одностороннего разрыва в первую нить, и дополнительную gRNA, которая направляет внесение одностороннего разрыва во вторую нить. В некоторых вариантах осуществления полипептид содержит множество доменов, обладающих эндонуклеазной активностью, и первый эндонуклеазный домен вносит односторонний разрыв в первой нити, и второй эндонуклеазный домен вносит односторонний разрыв во второй нити (необязательно первый эндонуклеазный домен не вносит (например, не может вносить) односторонний разрыв во второй нити, а второй эндонуклеазный домен не вносит (например, не может вносить) односторонний разрыв в первой нити).

В некоторых вариантах осуществления эндонуклеазный домен способен вносить односторонний разрыв в первую нить и вторую нить. В некоторых вариантах осуществления односторонние разрывы в первой и второй нитях происходят в одном и том же положении в сайте-мишени, но в противоположных нитях. В некоторых вариантах осуществления односторонний разрыв второй нити происходит в смещенном местоположении, например, выше или ниже первого одностороннего разрыва. В некоторых вариантах осуществления эндонуклеазный домен образует делецию сайта-мишени, если односторонний разрыв второй нити находится выше одностороннего разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен образует дупликацию сайта-мишени, если односторонний разрыв второй нити находится ниже одностороннего разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен не образует дупликации и/или делеции, если односторонние 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 одностороннего 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 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 FokI.

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 (SEQ ID NO: 1577), 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 (SEQ ID NO: 1577) с одной копией мотива LAGLIDADG (SEQ ID NO: 1577) обычно образуют гомодимеры, тогда как представители с двумя копиями мотива LAGLIDADG (SEQ ID NO: 1577) 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, сливают 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 предусматривает домен CRISPR/Cas (также называемый в данном документе CRISPR-ассоциированным белком). В некоторых вариантах осуществления ДНК-связывающий домен предусматривает домен CRISPR/Cas. В некоторых вариантах осуществления домен CRISPR/Cas предусматривает белок, участвующий в системе коротких палиндромных повторов, регулярно расположенных группами (CRISPR), например, белок Cas, и необязательно связывает направляющую РНК, например, одиночную направляющую РНК (sgRNA).

Системы CRISPR представляют собой адаптивные системы защиты, изначально обнаруженные у бактерий и архей. В системах CRISPR используются РНК-направляемые нуклеазы, называемые CRISPR-ассоциированными эндонуклеазами или эндонуклеазами "Cas" (например, Cas9 или Cpf1), для расщепления чужеродной ДНК. Например, в типичной системе CRISPR/Cas эндонуклеаза направляется на нуклеотидную последовательность-мишень (например, сайт в геноме, который подлежит редактированию последовательности) с помощью специфичных в отношении последовательности некодирующих "направляющих РНК", которые нацеливаются на одно- или двухнитевые последовательности ДНК. Были идентифицированы три класса (I-III) систем CRISPR. В системах CRISPR II класса используется одна эндонуклеаза Cas (вместо нескольких белков

Cas). Одна система CRISPR II класса содержит эндонуклеазу Cas II типа, такую как Cas9, CRISPR-ПНК ("crRNA") и транс-активирующую crRNA ("tracrRNA"). crRNA содержит "направляющую ПНК", в типичном случае последовательность ПНК длиной приблизительно 20 нуклеотидов, которая соответствует последовательности ДНК-мишени. В системе дикого типа и в некоторых сконструированных системах crRNA также содержит область, которая связывается с tracrRNA с образованием частично двухнитевой структуры, которая расщепляется ПНКазой III, что приводит к образованию гибрида crRNA/tracrRNA. Затем гибрид crRNA/tracrRNA направляет эндонуклеазу Cas9 для распознавания и расщепления последовательности ДНК-мишени. Последовательность ДНК-мишени, как правило, прилегает к "мотиву, прилежащему к протоспейсеру" ("PAM"), который является специфичным для определенной эндонуклеазы Cas; однако, последовательности PAM встречаются по всему рассматриваемому геному. Эндонуклеазы CRISPR, идентифицированные у различных видов прокариот, имеют уникальные требования к последовательности PAM; примеры последовательностей PAM включают 5'-NGG (*Streptococcus pyogenes*), 5'-NNAGAA (CRISPR1 *Streptococcus thermophilus*), 5'-NGGNG (CRISPR3 *Streptococcus thermophilus*) и 5'-NNNGATT (*Neisseria meningitidis*). Некоторые эндонуклеазы, например, эндонуклеазы Cas9, ассоциированы с G-богатыми сайтами PAM, например, с 5'-NGG, и осуществляют расщепление ДНК-мишени с образованием "тупых" концов в местоположении на 3 нуклеотида выше сайта PAM (с 5'-стороны от него). Другая система CRISPR II класса содержит эндонуклеазу Cpf1 V типа, которая имеет меньший размер, чем Cas9; примеры включают AsCpf1 (из *Acidaminococcus* sp.) и LbCpf1 (из *Lachnospiraceae* sp.). Cpf1-ассоциированные комплексы CRISPR подвергаются процессингу в зрелые crRNA без потребности в tracrRNA; другими словами, система Cpf1 в некоторых вариантах осуществления содержит только нуклеазу Cpf1 и crRNA для расщепления последовательности ДНК-мишени. Эндонуклеазы Cpf1, как правило, ассоциированы с T-богатыми сайтами PAM, например, с 5'-TTN. Cpf1 также может распознавать мотив PAM 5'-СТА. Cpf1, как правило, расщепляет ДНК-мишень посредством введения смещенного или ступенчатого двухнитевого разрыва с образованием выступающего 5'-конца длиной 4 или 5 нуклеотидов, например, посредством расщепления ДНК-мишени с образованием смещенного или ступенчатого разреза, охватывающего 5 нуклеотидов, расположенного на 18 нуклеотидов ниже сайта PAM (с 3'-стороны от него) на кодирующей нити и на 23 нуклеотида ниже сайта PAM на комплементарной нити; при этом выступающий конец длиной 5 нуклеотидов, который образуется в результате такого смещенного расщепления, обеспечивает возможность более точного редактирования генома за счет вставки ДНК посредством гомологичной рекомбинации по сравнению со вставкой в ДНК при расщеплении с образованием тупых концов. См., например, Zetsche et al. (2015) *Cell*, 163:759-771.

Различные CRISPR-ассоциированные (Cas) гены или белки могут использоваться в технологиях, предусмотренных настоящим изобретением, и выбор белка Cas будет зависеть от конкретных условий способа. Конкретные примеры белков Cas включают системы

класса II, в том числе Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1 или C2C3. В некоторых вариантах осуществления белок Cas, например, белок Cas9, может быть получен из любого из разнообразных видов прокариот. В некоторых вариантах осуществления конкретный белок Cas, например, конкретный белок Cas9, выбран для распознавания конкретной последовательности мотива, прилегающего к протоспейсеру (PAM). В некоторых вариантах осуществления ДНК-связывающий домен или эндонуклеазный домен содержит последовательность, нацеливающуюся на полипептид, такой как белок Cas, например, Cas9. В определенных вариантах осуществления белок Cas, например, белок Cas9, может быть получен из бактерий или архей или синтезирован с применением известных способов. В определенных вариантах осуществления белок Cas может быть получен из грамположительных бактерий или грамотрицательных бактерий. В определенных вариантах осуществления белок Cas может быть получен из бактерий рода *Streptococcus* (например, *S. pyogenes* или *S. thermophilus*), *Francisella* (например, *F. novicida*), *Staphylococcus* (например, *S. aureus*), *Acidaminococcus* (например, *Acidaminococcus* sp. BV3L6), *Neisseria* (например, *N. meningitidis*), *Cryptococcus*, *Corynebacterium*, *Haemophilus*, *Eubacterium*, *Pasteurella*, *Prevotella*, *Veillonella* или *Marinobacter*.

В некоторых вариантах осуществления белок Cas требует, чтобы мотив, прилегающий к протоспейсеру (PAM), присутствовал в последовательности ДНК-мишени или прилегал к ней для того, чтобы белок Cas мог связываться и/или функционировать. В некоторых вариантах осуществления PAM представляет собой следующее или содержит, в направлении от 5'-к 3'-концу, следующее: NGG, YG, NNGRRT, NNNRRT, NGA, TYCV, TATV, NTTN или NNNGATT, где N означает любой нуклеотид, Y означает C или T, R означает A или G, а V означает A, C или G. В некоторых вариантах осуществления белок Cas представляет собой белок, указанный в таблице 4. В некоторых вариантах осуществления белок Cas содержит одну или несколько мутаций, обеспечивающих изменение его PAM. В некоторых вариантах осуществления белок Cas содержит мутации E1369R, E1449H и R1556A или аналогичные замены аминокислот, соответствующих указанным положениям. В некоторых вариантах осуществления белок Cas содержит мутации E782K, N968K и R1015H или аналогичные замены аминокислот, соответствующих указанным положениям. В некоторых вариантах осуществления белок Cas содержит мутации D1135V, R1335Q и T1337R или аналогичные замены аминокислот, соответствующих указанным положениям. В некоторых вариантах осуществления белок Cas содержит мутации S542R и K607R или аналогичные замены аминокислот, соответствующих указанным положениям. В некоторых вариантах осуществления белок Cas содержит мутации S542R, K548V и N552R или аналогичные замены аминокислот, соответствующих указанным положениям.

16. Таблица 4. Белки CRISPR/Cas, виды и мутации

Название	Фермент	Вид	Количество а. к.	PAM	Мутации для изменения	Мутации для обеспечения
----------	---------	-----	------------------	-----	-----------------------	-------------------------

					распознавания PAM	каталитическ ой инертности
<i>FnCas9</i>	Cas9	<i>Francisella novicida</i>	1629	5'-NGG-3'	Wt	<i>D11A/H969A/N995A</i>
<i>FnCas9 RHA</i>	Cas9	<i>Francisella novicida</i>	1629	5'-YG-3'	E1369R/E1449H/R1556A	<i>D11A/H969A/N995A</i>
<i>SaCas9</i>	Cas9	<i>Staphylococcus aureus</i>	1053	5'-NNGRRT-3'	Wt	<i>D10A/H557A</i>
<i>SaCas9 KKH</i>	Cas9	<i>Staphylococcus aureus</i>	1053	5'-NNNRRT-3'	E782K/N968K/R1015H	<i>D10A/H557A</i>
<i>SpCas9</i>	Cas9	<i>Streptococcus pyogenes</i>	1368	5'-NGG-3'	Wt	<i>D10A/D839A/H840A/N863A</i>
<i>SpCas9 VQR</i>	Cas9	<i>Streptococcus pyogenes</i>	1368	5'-NGA-3'	D1135V/R1335Q/T1337R	<i>D10A/D839A/H840A/N863A</i>
<i>AsCpf1 RR</i>	Cpf1	<i>Acidaminococcus sp. BV3L6</i>	1307	5'-TYCV-3'	S542R/K607R	<i>E993A</i>
<i>AsCpf1 RVR</i>	Cpf1	<i>Acidaminococcus sp. BV3L6</i>	1307	5'-TATV-3'	S542R/K548V/N552R	<i>E993A</i>
<i>FnCpf1</i>	Cpf1	<i>Francisella novicida</i>	1300	5'-NTTN-3'	Wt	<i>D917A/E1006A/D1255A</i>
<i>NmCas9</i>	Cas9	<i>Neisseria meningitidis</i>	1082	5'-NNNGA TT-3'	Wt	<i>D16A/D587A/H588A/N611A</i>

В некоторых вариантах осуществления белок Cas является каталитически активным и разрезает одну или обе нити сайта ДНК-мишени. В некоторых вариантах осуществления за осуществлением разреза в сайте ДНК-мишени следует образование изменения, например, вставки или делеции, например, посредством механизма клеточной репарации.

В некоторых вариантах осуществления белок Cas модифицирован таким образом, чтобы в нем была деактивирована или частично деактивирована нуклеаза, например, Cas9 с дефицитом нуклеазы. При том, что Cas9 дикого типа приводит к образованию двухнитевых разрывов (DSB) в специфических последовательностях ДНК, которые служат мишенью для gRNA, доступен целый ряд эндонуклеаз CRISPR с модифицированными функциональными свойствами, например: "никазная" версия Cas9, которая была частично деактивирована, образует только одонитевой разрыв; каталитически неактивная Cas9 ("dCas9") не разрезает ДНК-мишень. В некоторых вариантах осуществления связывание dCas9 с последовательностью ДНК может препятствовать транскрипции в этом сайте за счет создания стерического препятствия. В некоторых вариантах осуществления связывание dCas9 с якорной последовательностью может препятствовать формированию и/или сохранению геномного комплекса (например, ASMC) (например, уменьшать или предупреждать его). В некоторых вариантах осуществления ДНК-связывающий домен содержит каталитически неактивную Cas9, например, dCas9. Из уровня техники известно много каталитически неактивных белков Cas9. В некоторых вариантах осуществления dCas9 содержит мутации в каждом эндонуклеазном домене белка Cas, например, мутации D10A и H840A или N863A. В некоторых вариантах осуществления каталитически неактивный или частично неактивный домен CRISPR/Cas содержит белок Cas, содержащий одну или несколько мутаций, например, одну или несколько мутаций, приведенных в таблице 4. В некоторых вариантах осуществления белок Cas, описанный в конкретной строке таблицы 4, содержит одну, две, три или все мутации, приведенные в той же строке таблицы 4. В некоторых вариантах осуществления белок Cas, например, не описанный в таблице 4, содержит одну, две, три или все мутации, приведенные в строке таблицы 4, или соответствующую мутацию в соответствующем сайте в этом белке Cas.

В некоторых вариантах осуществления каталитически неактивный, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию D11 (например, мутацию D11A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию H969 (например, мутацию H969A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию N995 (например, мутацию N995A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, содержит мутации в одном, двух или трех положениях из D11, H969 и N995 (например, мутации D11A, H969A и N995A) или аналогичные замены аминокислот, соответствующих указанным положениям.

В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию D10 (например, мутацию D10A) или аналогичную замену аминокислоты, соответствующей

В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию D16 (например, мутацию D16A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию D587 (например, мутацию D587A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию H588 (например, мутацию H588A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, или частично деактивированный белок Cas9 содержит мутацию N611 (например, мутацию N611A) или аналогичную замену аминокислоты, соответствующей указанному положению. В некоторых вариантах осуществления каталитически неактивный белок Cas9, например, dCas9, содержит мутацию D16 (например, D16A), мутацию D587 (например, D587A), мутацию H588 (например, H588A) и мутацию N611 (например, N611A) или аналогичные замены аминокислот, соответствующих указанным положениям.

В некоторых вариантах осуществления ДНК-связывающий домен или эндонуклеазный домен могут содержать молекулу Cas, содержащую gRNA (например, нуклеиновую кислоту-матрицу, например, РНК-матрицу, содержащую gRNA) или связанную с ней (например, ковалентно).

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas *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 Gene Writer может содержать белок Cas, который указан в таблице 40А. Предсказанные или валидированные мутации никазы для установления никазной активности в белке Cas, который показан в таблице 40А, производятся на основании сигнатуры мутации SpCas9(N863A). В некоторых вариантах осуществления система, описанная в данном документе, содержит белок Gene Writer из таблицы 3 и белок Cas из таблицы 40А. В некоторых вариантах осуществления белок или домен из таблицы 3, 41 или 44 слит с белком Cas из таблицы 40А.

Таблица 40А. Белки CRISPR/Cas, виды и мутации

Вариант	Исходный хозяин	Последовательность белка	Мутация никазы
Nme2 Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGV RVFERAEVPKTGDSLAMARRLARSVRRLTRRRRAHLLRAR RLLKREGVLQAADFDEGLIKSLPNTPWQLRAAALDRKLT LEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAN NAHALQTGDFRTPAELALNKFEEKESGHIRNQRGDYSHTFSR KDLQAEILLLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALS GDAVQKMLGHCTFEPAPKAANKNTYTAERFIWLTCLNNLRI LEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDT AFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKK SPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALL KHISFDKQVQISLKALRRIVPLMEQGKRYDEACAEIYGDHY GKKNTEEKIYLPPPADEIRNPVVLRLALSQARKVINGVVRRY GSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKF REYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVR NEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQT PYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDE DGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQ QKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFF AQEVMIRVFGKPDGKPEFEEADTPEKLRRTLLAEKLSSRPEAV HEYVTPLFVSRAPNRKMSGAHKDTLRSKRFRVKHNEKISVK	N611 A

		RVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGN AKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKN AYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAE NILPDIDCKGYRIDDSYTFCSLHKYDLIAFQKDEKSKVEFA YYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQV NELGKEIRPCRLKKRPPVR	
PpnCa s9	Pasteu rella pneum otropic a	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTF ERA EVAKTGESLALSRRRLARSSRRLIKRRAERLKKAKRLK AEKILHSIDEKLPINWQLRVKGLKEKLERQEWAAVLLHLS KHRGYLSQRKNEGKSDNKELGALLSGIASNHQMLQSSEYR TPAEIAVKKFQVEEGHIRNQRGSYTHTF SRLDLLAEMELLFQ RQAELGNSYTSTTLLENLTALLMWQKPALAGDAILKMLGK CTFEPSEYKAAKNSYSAERFVWLTKLNNLRILENGTERALN DNERFALLEQPYEKSCLTYAQVRAMLALSDNAIFKGVRYL GEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNEKGN SDLLDEIGTAFSLYKTDDDICRYLEGKLPERVNLALLENLNF DKFIQLSLKALHQILPLMLQGQRYDEAVSAIYGDHYGKKST ETTRLLPTIPADEIRNPVVLRTLQARKVINAVVRLYGGSPARI HIETAREVGKSYQDRKKLEKQQEDNRKQRESAVKKFKEMF PHFVGEPKGDILKMRLYELQQAKCLYSGKSLELHRLLEKG YVEVDHALPFSRTWDDSFNNKVLVLANENQNKGNLTPYE WLDGKNNSERWQHVFVRVQTSGFSAKKQRILNHKLDEK GFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNG QITALLRHRWGLQKVREQNDRHHALDAVVVACSTVAMQQ KITRFVRYNEGNVFSGERIDRETGEIPLHFSPWAFFKENVEI RIFSENPKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTG QGHMETVKSARKLNEGLSVLKVPLTQLKLSDLERMVNRDR EIALYESLKARLEQFGNDPAKAFAPFYKKGGALVKAVRLE QTQKSGVLVRDGNGVADNASMVRVDVFTKGGKYFLVPIYT WQVAKGILPNRAATQGKDENDWDIMDEMATFQFSLCQND LIKLVTKKKTIFGYFNGLNRATSNINIKEHDLKSKGKLGiy LEVGVKLAISLEKYQVDELGKNIRPCRPTKRQHVR	N605 A
SauCa s9	Staphy lococc	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV	N580 A

	us aureus	EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIEVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIIEAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE RIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV KVKSSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANAD FIFKEWKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGP VIKKIKYYGNKLN AHL DITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI SNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVMIDI TYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVK SKKHPQIIKKG	
SauCa s9- KKH	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIEVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIIEAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE	N580 A

		RIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV KVKSSINGGFTSFLRRKWKFKKERNKGYKHHAE DALIIANAD FIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETE QEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGP VIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI SNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVMIDI TYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEV KSKKHPQIIKKG	
SauriC as9	Staphy lococ us auricul aris	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFP EADSENNNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMI DLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLH NISVSMGDEEQDNE LSTKQQLQKNAQQLQDKYVCELQLER LTNINKVRGEKNRKFTEDFVKEVKQLCETQRQYHNIDDQFI QQYIDL VSTRREYFEGPGNGSPYGWDGDLKWEKLMGRC TYFPEELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEY YEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKSG KPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEI SIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVID ELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSP VVKRAFIQSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFIN KLQKQNEATRKKIEQLLAKYGNTNAKYMIEKIKLHDMQEG KCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLV KQSENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKD RISKKKRDMLEERDINKFEVQKEFINRNLVDTRYATRELSN LLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGY KHHAEDALVIANADFLFKTHKALRRTDKILEQPGLEVNDTT VKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKNR QLINDTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTER PQKILMYQHDPKTFEKLMTILNQYAEAKNPLAAYYEDKGE	N588 A

		YVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFREFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPDKYEAQKQKKIKESDLFVGsfyyNDLIMYEDELFRVIGVNSDINNLEVELNMVDITYKDFCEVNNVTGEKRIKKTIGKRVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	
SauriC as9- KKH	Staphy lococ us auricul aris	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMI DLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGDEEQDNElstKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRfKTEDFVKEVKQLCETQRQYHNIDDQFI QQYIDL VSTRREYFEGPGNGSPYGWDGDLKWEKLMGRC TYFPEELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEY YEKYHIIENVFKQKKNPtlKQIAKEIGVQDYDIRGYRITKSG KPQFTSfKL YHDLKNIFEQAKYLEDVEMlDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVID ELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSP VVKRAFIQSIKVINAVINRFGLPEDIIIElAREKNSKDRRKFIN KLQKQNEATRKKIEQLLAKYGNTNAKYMIEKIKLHDMQEG KCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLV KQSENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKD RISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATRELSN LLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGY KHHAEDALVIANADFLFKTHKALRRTDKILEQPGLEVNDTT VKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNR KLINDTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTER PQKILMYQHDPKTFEKLMTILNQYAEAKNPLAAYYEDKGE YVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFREFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPDKYEAQKQKKIKESDLFVGsfyKNDLIMYEDELFRVIGVNSDINNLEVELNMVDITYKDFCEVNNVTGEKHIKKTIGKRVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	N588 A
ScaCa s9- Sc++	Strept ococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRliYLALAHIIKFR	N872 A

		GHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVD AKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFS AAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDL ALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKR SGK LATEEEFYKFIKPILEKMDGAEELLAKLNRDDLRLKQR TFDNGSIPHQIHLKELHAILRRQEETFYPFLKENREKIEKILTR IPYYVGPLARGNSRF AWLTRKSEEAITPWNFEEVVDKGASA QSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKY VTERM RKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEIIGVEDRFNASLGT YHDLLKIIKDKDFLDNEENE DILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR HYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQ LIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIIKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRER KKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVEN RGKSDNVPSEE VVKMKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRD KNDKPIREVKVITLKS KLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNG ETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSK ESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKV EKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDI KKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQH LVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEK YILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGAS GGFTFLDLVDVKQGRRLRYQTVTEVLDATLIYQSITGLYETRT DLSQLGGD	
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD	N863A

		AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAQLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFK EDIQAQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDE LVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPPKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLG GD	
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSVDKLFQILVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP	N863A

		NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLEPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASQAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESIRPKRNS DKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEVGKSK KLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVL SAYNKHARDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVL DATLIHQ SITGLYETRIDLSQLG GD	
SpyCa s9- SpRY	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF	N863 A

		LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESIRPKRNS DKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIQISEFSKRVI LADANLDKVL SAYNKHARDKPIREQAENIIHLFTLTRLGAPRA FKYFDTTIDPKQYRSTKEVLDATLIHQSI TGLYETRIDLSQLG GD	
St1Cas 9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAEN NLVRR TNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLD DAS DDGNSSVG DY AQIVKENS KQLETKTPGQIQ LERYQTYGQLR GDFTVEKD GKKHRLIN VFPTSAYRSEALRILQTQQEFNPQIT DEFINRYLEILT GKRKY YHGP GNEKSR TDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPT ETKKLSKEQKNQIINYVKNEKAMGPAKL FKYIAKLLSCDVA	N622 A

		DIKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLD KLAYVLTNLNTEREGIQEALEHEFADGSFSQKQVDELVQFRK ANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGK QKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAI KEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAA MLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC LYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYA TANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLNKK KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAV DALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDE YKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSKFNKISD ATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKI YKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEV PCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYSKLGNIH DITPKDSNNKVVLQSVSPWRADVFNKTTGKYEILGLKYA DLQFEKGTGTYKISQEKYNDIKKKEGVDSSEFKFTLYKND LLLVKDTETKEQQLFRFLSRTMPKQKHVELKPYDKQKFE GGEALIKVLGNVANSQGCKKGLGKSNISYKVRTDVLGNQH IIKNEGDKPKLDF	
BlatCa s9	Brevib acillus lateros porus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKN GEALAVPRREARSSRRRLRRKKHRIERLKHMFVRNGLAVDI QHLEQTLRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIHLA QRRGFQSNRKTDGSSSEDGQVLVNVVTENDRLMEEKDYRTVA EMMVKDEKFSDHKRNKNGNYHGVVSRSSLLVEIHTLFTETQ RQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTF LPKEKRAPKASWHFQYFMLLQTINHIRTNVQGTRSLNKEEI EQVVNMALTKSKVSYHDTRKILDLSEEYQFVGLDYGKEDE KKKVESKETIIKLDDYHKLNKIFNEVELAKGETWEADDYDT VAYALTFFKDDDEDIRDYLNQKYKDSKNRLVKNLANKEYTN ELIGKVSTLSFRKVGHLCLKALRKIIPFLEQGMTYDKACQAA GFDFQGISKKKRSVVLVIDQISNPVVNRALTQTRKVINALI KKYGSPETIHETARELSKTFDERKNITKDYKENRDKNEHAK KHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISEFERLK ESGYTEVDHIIPYSRSMNDSYNNRVLVMTRENREKGNQTPF	N607 A

		EYMGNDTQRWYEFQRVTTNPQIKKEKRQNLLLKGFTNRR ELEMLERNLNDTRYITKYL SHFISTNLEFSPSDKKKKVVNTS GRITSHLRSRWGLEKNRGQNDLHHAMDAIVIAVTSDFSFIQQ VTNYYKRKERRELNGDDKFPLPWKFFREEVIARLSPNPKEQI EALPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVV GKTKEGKNITAKKTALVDISYDKNGDFNMYGRETDPATYE AIKERYLEFGGNVKKAFSTDHLHKPKKDGTGKPLIKSVRIME NKTLVHPVNGKGGVVYNSSIVRTDVFQRKEKYLLPVYVT DVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLYPNDLIFI RQNPKKKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIE FIIHDGSYYAKGVGVQNLDCEKYQVDILGNYFKVKGEKRL ELETSDSNHKGKDVNSIKSTSR	
cCas9- v16	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHNVNEV EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIENVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIHKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE RIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRLVDTRYATRGLMNLLRSYFRVNNLDV KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIINANAD FIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGP VIKKIKYYGNKLNALHDITDDYPNSRNKVVKL SLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI	N580 A

		SNQAEFIASFYKNDLIKINGEL YRVIGVNSDKNNLIEVNMIDI TYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEV KSKKHPQIIKKG	
cCas9- v17	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIENVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE RIEEIIRTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV KVKSSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANAD FIFKEWKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTL YSTRKDD KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGP VIKKIKYYGNKLNALHDITDDYPNSRNKVVKL SLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKLKKI SNQAEFIASFYKNDLIKINGEL YRVIGVNNSTRNIVELNMIDI TYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEV KSKKHPQIIKKG	N580 A
cCas9- v21	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV	N580 A

		KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIENVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE RIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIINANAD FIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGP VIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI SNQAEFIASFYKNDLIKINGELYRVIGVNSDDRNIIELNMIDIT YREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKS KKHPQIIKKG	
cCas9- v42	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSEL SGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVR GSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLET RRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIENVF KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVY HDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNE RIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL LNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER	N580 A

		DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV KVKSSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIINANAD FIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEI FITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQ TYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGP VIKKIKYYGNKLNNAHLDDITDDYPNSRNKVVKLSLKPYPFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI SNQAEFIASFYKNDLIKINGELYRVIGVNNNRLNKIELNMIDI TYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEV KSKKHPQIIKKG	
CdiCa s9	Coryn ebacte rium diphth eriae	MKYHVGIDVGTFVGLAAIEVDDAGMPIKTLVSHIHDSG LDPDEIKSAVTRLASSGIARRTRRLYRRKRRRLQQLDKFIQR QGWPVIELEDYSDPLYPWKVRaelAASYIADEKERGEKLSV ALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRAS GQVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDY AREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDP LQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVE EKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTATMTD DGERAGARPPHTD TNRSIVNSRIAPLVDWWKTASALEQHA MVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSL HLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFIEPS WTPPTPRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPER VIIIEHVREGFVTEKRAREMDGDMRRRAARNAKL FQEMQEK LNVQGKPSRADLWRYQSVQRQNCQAYCGSPITFSNSEMD HIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPF AIWAKNTS IEGVSVKEAVERTRHWVTD TGMRSTDFKKFTKAVVERFQR ATMDEEIDARSMESVAWMANELRSRVAQH FASHGTTVRV YRGS LTAEARRASGISGKLKFFDGVGKSRLDRRHHAIDAAV IAF TSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAE HRAAWRVWCQKMEKLSALLTEDLRDDR VVMSNVRLRL GNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTR EPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIA LRGGYAELGSSFH HARVYKITSGKKPAFAMLRVYTIDLLPY RNQDLFSVELKPQTMSMRQA EKKLRDALATGNAEYLGWL	H573 A (альт ернат ивная)

		VVDDEL VVDTSKIATDQVKAVEAELGTIRRWVRVDGFFSPSK LRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSNGN VTVVRDLSLGRVRLESTAHLPVTWKVQ	
CjeCa s9	Camp ylobac ter jeuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGES LALPRRLARSARKRLARRKARLNHLKHLIANEFKLNIEDYQ SFDESLAKAYKGSISPYLRFALNELLSKQDFARVILHIAK RRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLY KEYFQKFKEKSKEFTNVRNKKESYERCIAQSFLKDELKLIFK KQREFGFSFSKKFEEVLSVAFYKRALKDFSHLVGNCSFFTD EKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNA LLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTIFYEFKK YKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYD LNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEA CNELNLKVAINEDKKDFLPAFNETYYKDEVTPVVLRAIKE YRKVLNALLKKYVGKVKINIELAREVGKNHSQRAKIEKEQ NENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAY SGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNVLVFTKQ NQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDK NYKDKEQKNFKDRNLNDTRYIARLVNNTKDYLDLPLSD DENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDR NNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAK KISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSG ALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVK NGDMFRVDIFKHKKTNKFYAVPIYTMDFAKVLPNKAVAR SKKGEIKDWILMDENYEFCSLYKDSLILIQTKDMQEPEFVY YNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAK SIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	N582 A
GeoCa s9	Geoba cillus stearot hermo philus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQT GESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKDEL DKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRR GFKSNRKSERSNKENSTMLKHIEENRAILSSYRTVGEMIVKD PKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSC TEEFENEYITIWASQRPVASKDDIEKKVGFCFTEPKEKRAPK ATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQ KNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQENIRFLE	N605 A

		LDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDD ADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTK FGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQK TMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGGSPVSIHIEL ARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNP TGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVI PYSRSLDDSYTNKVLVL TRENREKGNRIPAAYLGVGTERWQ QFETFVL TNKQFSKKKRDRLLRLHYDENEETEFKNRNLNDT RYISRFFANFIREHLKFAESDDKQKVYTVNGRVT AHLRSRW EFNKNREESDLHHA VDAVIVACTTPSDIAKVTA FYQRREQN KELAKKTEPHFPQPWPHFADEL RARLSKHPKESIKALNLGN YDDQKLESLQPVFVSRMPKRSVTGA AHQETLRRYVGIDERS GKIQT VVKTKLSEIKLDASGHFPMYGKESDPRTYE AIRQRLL EHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPL NDGKTVA YNSNIVRVDVFEKDGYVCVPVYTMDIMKGILP NKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTV KTAAGEEINVKDV FVYYKTIDSANGGLELISHDHRFSLRGV GSRTLKRFEKYQVDVLGNIYKVRGEKRVGLASSAHSKPGK TIRPLQSTRD	
iSpyM acCas 9	Strept ococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLI ALSLGLTP NFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLF LAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS	N863 A

		RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEIQTVGQNGGLFDDNPK SPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQ LIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTL VDIGDGIKRLWASSKEIHKGNQLVVSCKSQILLYHAHHLDS DLSNDYLQNHNNQFDVLFNEIISFSKKCKLGKEHIQKIENVY SNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQ YKGKKDYILPCTEGTLIRQSITGLYETRVDL SKIGEDSGGSG GSKRTADGSEFES	
NmeC as9	Neisse ria menin gitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGV RVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTR RLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTP LEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAG NAHALQTGDFRTPAELALNKFEKESGHIRNQSRDYSHTFSR KDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALS GDAVQKMLGHCTFEPAPKAANKNTYTAERFIWLTKLNNLRI LEQGSERPLTDTERATLMDEPYRKSCLTYAQARKLLGLEDT AFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKK SPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLK HISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYG KKNTTEKIYLPPIPADEIRNPVVLRLALSQARKVINGVVRRYG SPARIHIETAREVGKSKDRKEIEKRQEENRKDREKAAAKFR EYFPNFVGEPKSKDILKLRLYEQQHKGKCLYSGKEINLGRLE KGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPY EYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDG FKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQ	N611 A

		ITNLLRGFWGLRKVRAENDRRHHALDAVVVACSTVAMQQKI TRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQ EVMIRVFGKPDGKPEFEEADTLEKLRTLAEKLSSRPEAVHE YVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVP LTQLKLDLEKMVNREREPKLYEALKARLEAHKDDPAKAF AEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIA DNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQ GKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFQYFAS CHRG TGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDE LGKEIRPCRLKKRPPVR	
ScaCa s9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNR KSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRY LQEIFANEMAKLDDSFQRLSEESFLVEEDKKNERHPIFGNLA DEVAYHRNYPTIYHLRKKLADSPEKADLRILIYLAHIIKFR GHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVD AKGILSARLSKSKRLEKLIIVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFS AAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDL ALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGIGIKHRKRT TKLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLRLKQRT FDNGSIPHQIHLKELHAILRRQEEFYPLKENREKIEKILTFRI PYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASA QSFIERMTNFDEQLPNKKVLPKHSLLEYEFTVYNELTKVKY VTERM RKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE DILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRR HYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFGFSNRNFMQ LIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRER KKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVEN RGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKA ERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRD KNDKPIREVKVITLKS KLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKS	N872 A

		EQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNG ETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSK ESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKV EKGKAKKLKSVKVLVGITIMEKGSYEKDPGFLEAKGYKDI KKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQH LVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEK YILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGAS GGFTFLDLVDVKQGRRLRYQTVTEVLDATLIYQSITGLYETRT DLSQLGGD	
ScaCa s9- HiFi- Sc++	Strept ococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSSKKFKVLGNTNR KSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRY LQEIFANEMAKLDDSFQRLYESFLVEEDKKNERHPIFGNLA DEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFR GHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVD AKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFS AAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDL ALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKR SGKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLRLKQR TFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFR IPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASA QSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKY VTERM RKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE DILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRR HYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNANFMQ LIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIIKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRER KKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFIDDSIDNKVLTRSVEN RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRD KNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNG	N872 A

		ETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSK ESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKV EKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDI KKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQH LVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEK YILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGAS GGFTFLDLVDVKQGRRLRYQTVTEVLDTLIYQSITGLYETRT DLSQLGGD	
SpyCa s9- 3var- NRRH	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHL GELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKT YAHLFDDKVMKQLKRLRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE LVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIASEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKGNS	N863 A

		DKLIARKKDWDPPKKYGGFNSPTAAYSVLVVAKEVGKSK KLKSVKELLGITIMERSSSFENPIGFLEAKGYKEVKKDLIIL PKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYL ASHYEKLKGGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGVPA AFKYFDTTIDKKRYTSTKEVLDTLHQSIITGLYETRIDLSQL GGD	
SpyCa s9- 3var- NRTH	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAHAMIKFR GHFLIEGDLNPDNSVDKLFQILVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHL GELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTIAHLFDDKVMKQLKRLRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDL LVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKGNS DKLIARKKDWDPPKKYGGFNSPTVAAYSVLVVAKEVGKSK	N863 A

		KLLSVKELLGITIMERSSSFENPIGFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLA SHYEKLKGSSSEDNKQKQLFVEQHKHYLDEIIEQISEFSKRVL ADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGASAA FKYFDTTIGRKLYTSTKEVLDATLIHQ SITGLYETRIDLSQLG GD	
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDL TLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHL GELHAILRRQGDFYPFLKDNREKIEKILTFRIPIYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVL PKHSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKV VDE LVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKGNS DKLIARKKDWD PKKYGGFNSPTVAYSVLVVAKVEKGKSK KLLSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKL	N863A

		PKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTINRKQYNTTKEVLDTLIRQSITGLYETRIDLSQL GGD	
SpyCa s9- HF1	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAQLSKDQTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSK KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA	N863 A

		SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVLDATLIHQ SITGL YETRIDLSQLG GD	
SpyCa s9- QQR1	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKV VDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNA VVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWD PKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKL PKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL	N863 A

		ADAQLDKVLSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTFKQKQYRSTKEVLDTLHQSIITGLYETRIDLSQL GGD	
SpyCa s9- SpG	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPPKKYGGFLWPTVAYSVLVVAKEVGKSK KLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHRRDKPIREQAENIIHLFTLTNLGAPA	N863 A

		AFKYFDTTIDRKQYRSTKEVL DATLIHQ SITGLYETRIDL SQL GGD	
SpyCa s9- VQR	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNE MAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVDKLFQQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASASQSFIERMTNFD KNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWD PKKYGGFVSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAA	N863 A

		FKYFDTTIDRKQYRSTKEVL DATLIHQ SITGLYETRIDLSQLG GD	
SpyCa s9- VRER	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNE MAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLA LAHMIKFR GHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTA LIKKYPKLESEFVYG DYKVYDVRKMIAKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWD PKKYGGFVSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKL PKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKR VIL ADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAA	N863 A

		FKYFDTTIDRKEYRSTKEVLDATLIHQSI TGLYETRIDLSQLG GD	
SpyCas9- xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNE MAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP NFKSNFDLAEDTKLQLSKDTYDDDLNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLG ELHAILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEKVVDKGASASQSFIERMTNFDK NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK LINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDI QKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVK VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVHDHVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPPKYGGFDSPTVAYSVLVVAKVEKGKSKKL KSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLAS HYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVL ADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAA	N863 A

		FKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLG GD	
SpyCa s9- xCas- NG	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNE MAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLA LAHMIKFR GHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDTKLQLSKDTYDDDL DNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLG ELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDK NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG VEDRFNASLGT YHDL LKIIKD KDFLDNEENEDILEDIVLTLTL FEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRK LINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDI QKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDDELVK VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP SEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESIRPKRNSD KLIARKKDWD PKKYGGFVSPTVAYSVLVVAKVEKGKSKKL KSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFK	N863 A

		YFDTTIDRKVYRSTKEVL DATLIHQ SITGLYETRIDLSQLGG D	
St1Cas 9- CNRZ 1066	Strept ococcus thermo philus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAEN NLVRR TNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDDAS DDGNSSVG DY AQIVKENS KQLETKTPGQIQLERYQTYGQLR GDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQIT DEFINRYLEILTGKRKYHGP GNEKSR TDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPT ETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLD KLAYVLT LNTEREGIQEAL EHEFADGSFSQKQVDEL VQFRK ANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGK QKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAI KEYGDFDNIVIE MARETNEDDEKKA IQKI QKANKDEKDAA MLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC LYTGKTISI HDLINNSNQFEVDHILPLSITFDDSLANKVLVYA TANQEK GQRTPYQALDSMDDAWSFRELKAFVRESKTL SNK KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAV DALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDE YKESVFKAPYQH FVDTLKSKEFEDSILFSYQVDSKFN RKISD ATIIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKI YKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKE VPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGN PIDITPENSKNKVVLQSLKPWR TDVYFNKATGKYEILGLKY ADLQFEKGTGTYKISQEKYNDIKKKEGVDS DSEFKFTLYKN DLLLVKDTETKEQQLFRFLSRTL PKQKH YVELKPYDKQKFE GGEALIKVLGNVANGGQCIKGLAKSNISYKVRTDVLGNQH IINKNEGDKPKLDF	N622 A
St1Cas 9- LMG1 831	Strept ococcus thermo philus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAEN NLVRR TNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDDAS DDGNSSVG DY AQIVKENS KQLETKTPGQIQLERYQTYGQLR GDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQIT	N622 A

		DEFINRYLEILTGKRKYHGPNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPT ETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLD KLAYVLTNTEREGIQEALEHEFADGSFSQKQVDELVQFRK ANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGK QKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAI KEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAA MLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC LYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYA TANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSENK KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAV DALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDE YKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSKFNKISD ATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKI YKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKE VPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYDYDSKLLGN PIDITPENSKNKVVLQSLKPWRDVFYFNKNTGKYEILGLKY ADLQFEKKTGTYKISQEKYNGIMKEEGVDSSEFKFTLYKN DLLLVKDTETKEQQLFRFLSRTMPNVKYYVELKPYSKDKFE KNESLIEILGSADKSGRCIKGLGKSNISYKVRTDVLGNQHII KNEGDKPKLDF	
St1Cas 9- MTH1 7CL39 6	Strept ococcus thermo philus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAEN NLVRRITNRQGRRLARRKKHRRVRLNRLFEEGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDAS DDGNSSVGDYAQIVKENSQKLETKTPGQIQLERYQTYGQLR GDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQIT DEFINRYLEILTGKRKYHGPNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPT ETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLD KLAYVLTNTEREGIQEALEHEFADGSFSQKQVDELVQFRK ANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGK QKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAI	N622 A

		KEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAA MLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC LYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYA TANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLNKK KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAV DALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDE YKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSKFNKISD ATYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKI YKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEV PCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNI DITPKDSNNKVVLQSLKPWRDVFYFNKNTGKYEILGLKYSD MQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDL LLLKDSengeQILLRFTSRNDTSKHVELKPYNRQKFEGSEY LIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHIIKNE GDKPKLDF	
St1Cas 9- TH147 7	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAEN NLVRRTNRQGRRLARRKKHRRVRLNRLFEESSGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDAS DDGNSSVGDYAQIVKENSQKLETKTPGQIQLERYQTYGQLR GDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQIT DEFINRYLEILTGKRKYHGPNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPT ETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLD KLAYVLTNTEREGIQEALHEFADGSFSQKQVDELVQFRK ANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGK QKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAI KEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAA MLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC LYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYA TANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLNKK KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAV DALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDE	N622 A

	YKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSKFNRKISD ATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKI YKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEV PCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNI DITPKDSNNKVVLQSLKPWRDVFYFNKNTGKYEILGLKYSD MQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDL LLLKDSENGEQILLRFTSRNDTSKHVELKPYNRQKFEGSEY LIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNE GDKPKLDF	
--	--	--

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен 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/Cpf1, 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/Cpf1, 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/Cpf1, 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/Cpf1, 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 (НураCas9), 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 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 содержит spCas9, spCas9-VRQR (SEQ ID NO: 1578), spCas9-VRER (SEQ ID NO: 1579), xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER (SEQ ID NO: 1580), spCas9-LRKIQK (SEQ ID NO: 1581) или spCas9-LRVSQL (SEQ ID NO: 1582).

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен 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 имеет не более 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 отличий (например, мутаций) по сравнению с любыми аминокислотными последовательностями, описанными 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 (SEQ ID NO: 1023)
Иллюстративный линкерный мотив	(SGGS) _n (SEQ ID NO: 1583)
Иллюстративный линкерный мотив	(GGGS) _n (SEQ ID NO: 1584)
Иллюстративный линкерный мотив	(GGGGS) _n (SEQ ID NO: 1535)
Иллюстративный линкерный мотив	(G) _n
Иллюстративный линкерный мотив	(EAAAK) _n (SEQ ID NO: 1534)
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 <i>Streptococcus iniae</i>	Эталонная последовательность в NCBI: NC_021314.1
Cas9 из <i>Belliella baltica</i>	Эталонная последовательность в NCBI: NC_018010.1
Cas9 из <i>Psychroflexus torquis</i>	Эталонная последовательность в NCBI: NC_018721.1
Cas9 из <i>Streptococcus thermophilus</i>	Эталонная последовательность в NCBI: YP_820832.1
Cas9 из <i>Listeria innocua</i>	Эталонная последовательность в NCBI: NP_472073.1
Cas9 из <i>Campylobacter jejuni</i>	Эталонная последовательность в NCBI: YP_002344900.1
Cas9 из <i>Neisseria meningitidis</i>	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 Paracubacteria])
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 (<i>Bacillus hisashii</i>)	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	

В некоторых вариантах осуществления полипептид Gene Writing имеет эндонуклеазный домен, содержащий нуклеазу Cas9, например, Cas9 H840A. В вариантах осуществления Cas9 H840A имеет следующую аминокислотную последовательность.

Нуклеаза Cas9 (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHER
HPIFGNIVDEVAYHEKYPTIYHLRKKLVSTDKADLRLLIYLAHMIKFRGHFLIEGDLN
PDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGITALIKKYPKLESEFV
YGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKKNPIDFLEAKGY
KEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 1585)

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):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLILAATSEL
DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLYGILLKEGQRWLTEARKETVMG
QPTPKTPRQLREFLGTAFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQ
ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPPWRPVAAYLSKKLDPVAAGWP

PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL
 LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT
 DGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLVY
 TDSRYAFATAHIIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIHCPGHQKGH
 SAEARGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1586)

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, например, RT M-MLV, например, содержащий следующую последовательность:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKAT
 STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLVPKKPGTNDYRPVQDL
 REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
 PEMGISGQLTWTRLPGGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLAAATSEL
 DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLYGILLKEGQRWLTEARKETVMG
 QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQ
 ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTKQLGPWRRPVAYLSKKLDPVAAGWP
 PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL
 LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT
 DGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLVY
 TDSRYAFATAHIIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIHCPGHQKGH
 SAEARGNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 1548)

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:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKAT
 STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLVPKKPGTNDYRPVQD
LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF
EWRDPEMGISGQLTWTRLPGGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDL
LLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLYGILLKEGQRWLT
 EARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPD
 QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTKQLGPWRRPVAYLSK
 KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN
 ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP
 LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK
MAEGKKLVYTDDSRYAFATAHIIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLS
IIHCPGHQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 1587)

Сердцевинная RT (жирный шрифт), аннотирована согласно вышеуказанному
PHКаза 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):

TLNIEDEYRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPGGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLALLAATSEL
DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLYGYLLKEGQRWLTEARKETVMG
QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQ
ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPPWRPVAYLSKKLDPVAAGWP
PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL
LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT
DGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVY
TDSRYAFATAHNGEIIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIHCPGHQKGH
SAEARGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1588)

В некоторых вариантах осуществления полипептид Gene Writer может содержать линкер, например, пептидный линкер, например, линкер, описанный в таблице 38. В некоторых вариантах осуществления полипептид Gene Writer содержит гибкий линкер между эндонуклеазой и доменом RT, например, линкер, содержащий аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1589). В некоторых вариантах осуществления домен RT полипептида Gene Writer может быть расположен с С-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 38. Иллюстративные линкерные последовательности

Аминокислотная последовательность
GGS
GGS GGS
GGS GGS GGS
GGS GGS GGS GGS
GGS GGS GGS GGS GGS
GGS GGS GGS GGS GGS GGS
GGGS
GGGS GGGGS
GGGS GGGGS GGGGS
GGGS GGGGS GGGGS GGGGS
GGGS GGGGS GGGGS GGGGS GGGGS
GGGS GGGGS GGGGS GGGGS GGGGS GGGGS
GGG
GGGG
GGGGG
GGGGGG
GGGGGGG
GGGGGGGG
GSS
GSS GSS
GSS GSS GSS
GSS GSS GSS GSS
GSS GSS GSS GSS GSS
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
GGGSSEAAAK
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
 SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
 SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
 HLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
 WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEG
 MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
 TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
 RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQ

VSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ
 KGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
 NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
 AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
 KLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLES
 EFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
 GETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
 WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEA
 KGYKEVKKDLIIKLPKYSLENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
 KLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVLSAYNKHRRDKPIR
 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQL
 GGD (SEQ ID NO: 1590)

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 шпильчатых петель, например, 1, 2 или 3 петли для ассоциации матрицы с нуклеотидным доменом Cas9. В некоторых вариантах осуществления каркас gRNA в направлении от 5' к 3' несет последовательность

GTTTGTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
 AAGTGGGACCGAGTCGGTCC (SEQ ID NO: 1591).

(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 односторонний разрыв, 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 одностороннего разрыва первой нити, например, 0-50, 50-100, 100-200 нуклеотидов от одностороннего 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, 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 SGGSSGGSSGSETPGTSESATPSSGGSSGGSS (SEQ ID NO: 1589).

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.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит эффекторную молекулу TAL. Эффекторная молекула TAL, например, эффекторная молекула TAL, которая специфично связывается с последовательностью ДНК, обычно содержит множество эффекторных доменов TAL или их фрагментов и необязательно одну или несколько дополнительных частей встречающихся в природе эффекторов TAL (например, N- и/или С-конец множества эффекторных доменов TAL). Многие эффекторы TAL известны специалистам в данной области и коммерчески доступны, например, от Thermo Fisher Scientific.

Встречающиеся в природе TALE представляют собой природные эффекторные белки, секретируемые многочисленными видами бактериальных патогенов, включая патоген растений *Xanthomonas*, который модулирует экспрессию генов в растениях-хозяевах и способствует бактериальной колонизации и выживанию. Специфичное связывание эффекторов TAL определяется центральным доменом повторов из tandemно расположенных почти идентичных повторов, обычно состоящих из 33 или 34 аминокислот (вариабельных диостатков повторов, домена RVD).

Представители семейства эффекторов TAL различаются в основном количеством и порядком своих повторов. Количество повторов обычно находится в диапазоне от 1,5 до 33,5 повтора, а С-концевой повтор обычно короче по длине (например, приблизительно 20 аминокислот) и обычно называется "полуповтором". Каждый повтор эффектора TAL обычно характеризуется корреляцией "один повтор - одна пара оснований", при этом разные типы повторов демонстрируют разную специфичность в отношении пары оснований (один повтор распознает одну пару оснований в последовательности гена-мишени). Как правило, чем меньше количество повторов, тем слабее взаимодействия белок-ДНК. Было показано, что количество 6,5 повтора достаточно для активации транскрипции репортерного гена (Scholze et al., 2010).

Вариации от повтора к повтору происходят преимущественно в положениях аминокислот 12 и 13, которые поэтому были названы "гипервариабельными" и которые отвечают за специфичность взаимодействия с последовательностью промотора ДНК-мишени, как показано в таблице 5, в которой указаны иллюстративные вариабельные диостатки повторов (RVD) и их соответствие основаниям-мишеням нуклеиновых кислот.

17. Таблица 5 - RVD и специфичность в отношении оснований нуклеиновых кислот

Мишень	Возможные комбинации аминокислот RVD												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

Соответственно, можно модифицировать повторы эффектора TAL для нацеливания на определенные последовательности ДНК. Дополнительные исследования показали, что RVD NK может нацеливаться на G. Сайты-мишени эффекторов TAL также, как правило, содержат T, фланкирующий 5'-основание, на которое нацеливается первый повтор, но точный механизм этого распознавания неизвестен. На сегодняшний день известно более 113 эффекторных последовательностей TAL. Неограничивающие примеры эффекторов TAL из *Xanthomonas* включают Nah2, Nah3, Nah4, AvrXa7, AvrXa10 и AvrBs3.

Соответственно, эффекторный домен TAL эффекторной молекулы TAL, описанной в данном документе, может быть получен из эффектора TAL любого вида бактерий (например, *Xanthomonas*, таких как африканский штамм *Xanthomonas oryzae* pv. *Oryzae* (Yu et al. 2011), *Xanthomonas campestris* pv. *raphani*, штамм 756C и *Xanthomonas oryzae* pv. *oryzicola*, штамм BLS256 (Bogdanove et al. 2011)). В некоторых вариантах осуществления эффекторный домен TAL содержит домен RVD, а также фланкирующую(фланкирующие) последовательность(последовательности) (последовательности с N-концевой и/или C-концевой стороны от домена RVD) также из встречающегося в природе эффектора TAL. Он может содержать больше или меньше повторов, чем RVD встречающегося в природе эффектора TAL. Эффекторная молекула TAL может быть разработана для нацеливания на данную последовательность ДНК, исходя из приведенного выше кода и других кодов, известных из уровня техники. Количество эффекторных доменов TAL (например, повторов (мономеров или модулей)) и их специфическую последовательность можно выбрать, исходя из желаемой последовательности ДНК-мишени. Например, эффекторные домены TAL, например, повторы, могут быть удалены или добавлены, чтобы соответствовать специфической последовательности-мишени. В одном варианте осуществления эффекторная молекула TAL по настоящему изобретению содержит от 6,5 до 33,5 эффекторного домена TAL, например, повтора. В одном варианте осуществления эффекторная молекула TAL по настоящему изобретению содержит от 8 до 33,5 эффекторного домена TAL, например, повтора, например, от 10 до 25 эффекторных доменов TAL, например, повторов, например, от 10 до 14 эффекторных доменов TAL, например, повторов.

В некоторых вариантах осуществления эффекторная молекула TAL содержит эффекторные домены TAL, которые соответствуют последовательности ДНК-мишени до абсолютного совпадения. В некоторых вариантах осуществления допускается ошибка спаривания между повтором и парой оснований-мишеней в последовательности ДНК-мишени, если это позволяет осуществлять функцию полипептида, содержащего эффекторную молекулу TAL. Как правило, связывание TALE обратно коррелирует с количеством ошибок спаривания. В некоторых вариантах осуществления эффекторная молекула TAL полипептида по настоящему изобретению содержит не более 7 ошибок спаривания, 6 ошибок спаривания, 5 ошибок спаривания, 4 ошибок спаривания, 3 ошибок спаривания, 2 ошибок спаривания или 1 ошибки спаривания и необязательно не содержит ошибок спаривания с последовательностью ДНК-мишени. Не желая ограничиваться какой-

либо теорией, в целом полагают, что чем меньше количество эффекторных доменов TAL в эффекторной молекуле TAL, тем меньшее количество ошибок спаривания будет допустимо и все еще будет позволять осуществлять функцию полипептида, содержащего эффекторную молекулу TAL. Считается, что аффинность связывания зависит от суммы совпадающих комбинаций повтор-ДНК. Например, эффекторные молекулы TAL, имеющие 25 или более эффекторных доменов TAL, могут допускать до 7 ошибок спаривания.

В дополнение к эффекторным доменам TAL, эффекторная молекула TAL по настоящему изобретению может содержать дополнительные последовательности, полученные из встречающегося в природе эффектора TAL. Длина С-концевой(концевых) и/или N-концевой(концевых) последовательности(последовательностей), включенных с каждой стороны от части, представляющей собой эффекторный домен TAL эффекторной молекулы TAL, может варьироваться и может быть выбрана специалистом в данной области, например, на основании исследований Zhang et al. (2011). Zhang et al. охарактеризовали ряд усеченных на С-конце и N-конце мутантных форм в белках на основе TAL-эффекторов, полученных из *Нах3*, и идентифицировали ключевые элементы, которые способствуют оптимальному связыванию с последовательностью-мишенью и, таким образом, активации транскрипции. В целом было обнаружено, что транскрипционная активность обратно коррелирует с длиной N-конца. Что касается С-конца, был идентифицирован важный элемент из ДНК-связывающих остатков в пределах первых 68 аминокислот последовательности *Нах3*. Соответственно, в некоторых вариантах осуществления первые 68 аминокислот с С-концевой стороны от эффекторных доменов TAL встречающегося в природе эффектора TAL содержатся в эффекторной молекуле TAL. Соответственно, в варианте осуществления эффекторная молекула TAL содержит 1) один или несколько эффекторных доменов TAL, полученных из встречающегося в природе эффектора TAL; 2) по меньшей мере 70, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260, 270, 280 или больше аминокислот из встречающегося в природе эффектора TAL с N-концевой стороны от эффекторных доменов TAL и/или 3) по меньшей мере 68, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260 или больше аминокислот из встречающегося в природе эффектора TAL с С-концевой стороны от эффекторных доменов TAL.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен представляет собой или содержит молекулу с "Zn-пальцами". Молекула с "Zn-пальцами" содержит белок с "Zn-пальцами", например, встречающийся в природе белок с "Zn-пальцами" или сконструированный белок с "Zn-пальцами" или его фрагмент. Многие белки с "Zn-пальцами" известны специалистам в данной области и коммерчески доступны, например, от Sigma-Aldrich.

В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит не встречающийся в природе белок с "Zn-пальцами", который сконструирован таким образом, что он связывается с выбранной последовательностью ДНК-мишени. См., например, Beerli, et al. (2002) *Nature Biotechnol.* 20:135-141; Pabo, et al. (2001) *Ann. Rev. Biochem.* 70:313-340;

Isalan, et al. (2001) *Nature Biotechnol.* 19:656-660; Segal, et al. (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo, et al. (2000) *Curr. Opin. Struct. Biol.* 10:411-416; патенты США №№ 6453242; 6534261; 6599692; 6503717; 6689558; 7030215; 6794136; 7067317; 7262054; 7070934; 7361635; 7253273; и публикации заявок на патент США №№ 2005/0064474; 2007/0218528; 2005/0267061, все из которых включены в данный документ посредством ссылки во всей своей полноте.

Сконструированный белок с "Zn-пальцами" может иметь новую специфичность связывания по сравнению со встречающимся в природе белком с "Zn-пальцами". Способы конструирования включают без ограничения рациональный дизайн и различные типы отбора. Рациональный дизайн включает, например, использование баз данных, содержащих последовательности нуклеотидных триплетов (или квадруплетов) и отдельные аминокислотные последовательности "Zn-пальцев", в которых каждая последовательность нуклеотидного триплета или квадруплета соответствует одной или нескольким аминокислотным последовательностям "цинковых пальцев", которые связывают конкретную последовательность триплета или квадруплета. См., например, патенты США №№ 6453242 и 6534261, включенные в данный документ посредством ссылки во всей своей полноте.

Иллюстративные способы отбора, в том числе фаговый дисплей и двугибридные системы, раскрыты в патентах США №№ 5789538; 5925523; 6007988; 6013453; 6410248; 6140466; 6200759 и 6242568; а также международных патентных публикациях №№ WO 98/37186; WO 98/53057; WO 00/27878 и WO 01/88197, а также GB 2338237. Кроме того, усиление специфичности связывания белков с "цинковыми пальцами" описано, например, в международной патентной публикации № WO 02/077227.

Кроме того, как раскрыто в этих и других литературных источниках, домены "цинковые пальцы" и/или белки с несколькими "цинковыми пальцами" могут быть связаны друг с другом с помощью любых подходящих линкерных последовательностей, в том числе, например, линкеров длиной 5 или больше аминокислот. См. также патенты США №№ 6479626; 6903185 и 7153949 в отношении иллюстративных линкерных последовательностей длиной 6 или больше аминокислот. Описанные в данном документе белки могут содержать любую комбинацию подходящих линкеров между отдельными "цинковыми пальцами" белка. Кроме того, усиление специфичности связывания связывающих доменов "цинковых пальцев" описано, например, в совместной международной патентной публикации № WO 02/077227.

Белки с "Zn-пальцами" и способы разработки и конструирования слитых белков (и полинуклеотидов, кодирующих их) известны специалистам в данной области и подробно описаны в патентах США №№ 61400815; 789538; 6453242; 6534261; 5925523; 6007988; 6013453 и 6200759; международных патентных публикациях №№ WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 и WO 03/016496.

Кроме того, как раскрыто в этих и других литературных источниках, белки с "Zn-пальцами" и/или белки с несколькими "Zn-пальцами" могут быть связаны друг с другом, например, в виде слитого белка, с помощью любых подходящих линкерных последовательностей, в том числе, например, линкеров длиной 5 или больше аминокислот. См. также патенты США №№ 6479626; 6903185 и 7153949 в отношении иллюстративных линкерных последовательностей длиной 6 или больше аминокислот. Молекулы с "Zn-пальцами", описанные в данном документе, могут содержать любую комбинацию подходящих линкеров между отдельными белками с "цинковыми пальцами" и/или белками с несколькими "Zn-пальцами" в молекуле с "Zn-пальцами".

В определенных вариантах осуществления ДНК-связывающий домен или эндонуклеазный домен содержит молекулу с "Zn-пальцами", содержащую сконструированный белок с "цинковыми пальцами", который связывается (специфичным для последовательности образом) с последовательностью ДНК-мишени. В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит один белок с "Zn-пальцами" или его фрагмент. В других вариантах осуществления молекула с "Zn-пальцами" содержит множество белков с "Zn-пальцами" (или их фрагментов), например, 2, 3, 4, 5, 6 или больше белков с "Zn-пальцами" (и необязательно не более 12, 11, 10, 9, 8, 7, 6, 5, 4, 3 или 2 белков с "Zn-пальцами"). В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит по меньшей мере три белка с "Zn-пальцами". В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит четыре, пять или шесть "пальцев". В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит 8, 9, 10, 11 или 12 "пальцев". В некоторых вариантах осуществления молекула с "Zn-пальцами", содержащая три белка с "Zn-пальцами", распознает последовательность ДНК-мишени, содержащую 9 или 10 нуклеотидов. В некоторых вариантах осуществления молекула с "Zn-пальцами", содержащая четыре белка с "Zn-пальцами", распознает последовательность ДНК-мишени, содержащую 12-14 нуклеотидов. В некоторых вариантах осуществления молекула с "Zn-пальцами", содержащая шесть белков с "Zn-пальцами", распознает последовательность ДНК-мишени, содержащую 18-21 нуклеотид.

В некоторых вариантах осуществления молекула с "Zn-пальцами" содержит двурукий белок с "Zn-пальцами". Двурукие белки с "цинковыми пальцами" представляют собой белки, в которых два кластера белков с "цинковыми пальцами" разделены промежуточными аминокислотами, так что два домена "цинковые пальцы" связываются с двумя прерывистыми последовательностями ДНК-мишеней. Примером двурукого типа связывающего белка с "цинковыми пальцами" является SIP1, где кластер из четырех белков с "цинковыми пальцами" расположен на амино-конце белка, а кластер из трех белков с "Zn-пальцами" расположен на карбоксильном конце (см. Remade, et al. (1999) EMBO Journal 18(18):5073-5084). Каждый кластер "цинковых пальцев" в этих белках способен связываться с уникальной последовательностью-мишенью, а расстояние между двумя последовательностями-мишенями может составлять множество нуклеотидов.

18. 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 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 определенных вариантах осуществления ДНК-связывающий домен полипептида 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 сохраняет частичную эндонуклеазную активность для расщепления ssDNA, например, обладает никазой

активностью. 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, домена TAL, домена ZF, домена Myb, их комбинаций или их множеств. В определенных вариантах осуществления гетерологичный ДНК-связывающий домен представляет собой ДНК-связывающий домен ретротранспозона или вируса, описанного в таблице 1 или таблице 3. Средний специалист в данной области способен идентифицировать ДНК-связывающие домены на основании гомологии по отношению к другим известным ДНК-связывающим доменам с помощью таких инструментов, 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 полипептид может связываться с одной или с более чем одной последовательностью ДНК хозяина.

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 сигнал ядерной локализации имеет длину, составляющую 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 М-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). 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 С-концом полипептида Gene Writer. В некоторых вариантах осуществления NLS слита с N-концом или С-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 MDSLLMNRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 1592), PKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1593), RKSGKIAAIWKRPRKPKKKRKV (SEQ ID NO: 1594), KRTADGSEFESPKKKRKV (SEQ ID NO: 1595), KKTTELQTTNAENKTKKL (SEQ ID NO: 1596) или KRGINDRNFWRGENGRKTR (SEQ ID NO: 1597), KRPAATKKAGQAKKKK (SEQ ID NO: 1598) или ее функциональный фрагмент или вариант. Иллюстративные последовательности 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 39. 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 С-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 Bioinform 13:157 (2012), включенный в данный документ посредством ссылки во всей своей полноте).

Таблица 39. Иллюстративные сигналы ядерной локализации для применения в системах Gene Writing

Последовательность	Ссылки на последовательности
AHFKISGEKRPSTDPGKKAKNPKKK KKKDP	Q76IQ7
AHRAKKMSKTHA	P21827
ASPEYVNLPIGNG	SeqNLS
CTKRPRW	O88622, Q86W56, Q9QYM2, O02776
DKAKRVSRNKSEKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGG4
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15
EEQLRRRKNSRLNNTG	G5EFF5
EVLKVIRTGKRKKKAWKRMVTKV C	SeqNLS
HHHHHHHHHHHHQH	Q63934, G3V7L5, Q12837
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0
HKRTKK	Q2R2D5
IINGRKLKLLKSRRRSSQTSNNSFTS RRS	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
KKYENVVIKRSRKRGRPRK	SeqNLS
KNKKRK	SeqNLS
KPKKKR	SeqNLS
KRAMKDDSHGNSTSPKRRK	Q0E671
KRANSNLVAAYEKAKKK	P23508
KRASEDTTSGSPPKSSAGPKR	Q9BZZ5, Q5R644
KRFKRRWMVRKMKTCK	SeqNLS
KRGLNSSFETSPKKVK	Q8IV63
KRGNSSIGPNDLSKRKQRKK	SeqNLS
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS
KRKGKLKNKGSKRKK	O15381
KRRRRRRRREKRKR	Q96GM8
KRSNDRTYSPEEEKQRRRA	Q91ZF2
KRTVATNGDASGAHRAKKMSK	SeqNLS
KRVYNKGEDEQEHLPGGKKR	SeqNLS
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524
KVNFLDMSLDDIIYKELE	Q9P127
KVQHRIAKKTTRRRR	Q9DXE6

LSPSLSPL	Q9Y261, P32182, P35583
MDSLLMNRRKFLYQFKNVRWAKG RRETYLC	Q9GZX7
MPQNEYIELHRKRYGYRLDYHEKK RKESREAHERSKKAKKMIGLKAK LYHK	SeqNLS
MVQLRPRASR	SeqNLS
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40
PDTKRAKLDSETTMVKKK	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
RKAGKGGGGHKTTKKRSAKDEKV P	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
RKKSIPLSIKNLKRKHKRKKNKTR	Q9C0C9
RKLVPKNTKMKTKLRTNPY	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,
RKRSPKDKKEKDLGAGKRRKT	Q7RTP6
RKRTPRVGQTGENDMNKRRRK	O94851
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270
RLRFRKPKSK	P69469
RQQRKR	Q14980
RRDLNSSFETSPKKVK	Q8K3G5
RRDRAKLR	Q9SLB8
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922
RRGRKRKA EKQ	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
RRRREGPRARRR	P08313, P10231
RRTIRLKLVDKCDRSCKIQKKNRN KCQYCRFHKCLSVGMSHNAIRFGR MPRSEKAKLKAE	SeqNLS
RRVPQRKEVSRCRKRK	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	
MDSLLMNRRKFLYQFKNVRWAKG RRETYLC	
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, имеющая последовательность KR[PAATKKAGQA]KKKK (SEQ ID NO: 1598), где спейсер заключен в скобки. Другая иллюстративная двухкомпонентная NLS имеет последовательность PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1600). 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 или 3 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 κВ. 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 пониженном уровне по сравнению с клеткой, не являющейся мишенью). Не желая ограничиваться какой-либо теорией, полагают, что связывание miRNA с РНК, кодирующей полипептид Gene Writer™, может приводить к снижению выработки полипептида Gene Writer™, например, посредством разрушения мРНК, кодирующей полипептид, или посредством препятствования трансляции. Соответственно, в таких вариантах осуществления Gene Writer будет вносить добавления/редактирования в геном клеток-мишеней более эффективно, чем он редактирует геном клеток, не являющихся мишенями, например, последовательность гетерологичного объекта будет вставляться в геном клеток-мишеней более эффективно, чем в геном клеток, не являющихся мишенями, или вставка или делеция образуются более эффективно в клетках-мишенях, чем в клетках, 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". В некоторых вариантах осуществления, например, для печеночных показаний, miRNA выбрана из таблицы 4 WO2020014209, включенной в данный документ посредством ссылки.

В некоторых вариантах осуществления ДНК, кодирующая полипептид Gene Writer, содержит промоторную последовательность, например, тканеспецифическую промоторную последовательность. 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. Система, содержащая последовательность тканеспецифического промотора в ДНК полипептида, 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 ДНК, кодирующей полипептид Gene Writer, также может применяться в комбинации с ДНК, кодирующей РНК-матрицу, управляемой тканеспецифическим промотором, например, для достижения более высоких уровней РНК-матрицы в клетках-мишенях, чем в клетках, не являющихся мишенями. В некоторых вариантах осуществления, например, для печеночных показаний, тканеспецифический промотор выбран из таблицы 3 WO2020014209, включенной в данный документ посредством ссылки.

Специалист в данной области может по номерам доступа и/или последовательностям, представленным в таблицах 1 и 3, определить последовательности нуклеиновых кислот и соответствующих полипептидов для каждого ретротранспозона или вируса 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>.

В таблицах 1 и 3 в данном документе представлены последовательности иллюстративных транспозонов или вирусов, в том числе аминокислотная(аминокислотные) последовательность(последовательности) ретротранспозазы, обратной транскриптазы, ДНК-связывающего домена и/или эндонуклеазного домена; последовательности 5'- и 3'-нетранслируемых областей, обеспечивающие возможность связывания полипептида, например, ретротранспозазы, с РНК-матрицей, и/или полная последовательность нуклеиновой кислоты транспозона. В некоторых вариантах осуществления 5'-UTR,

содержащаяся или указанная в таблицах 1 или 3, обеспечивает возможность связывания полипептида, например, ретротранспозазы, с РНК-матрицей. В некоторых вариантах осуществления 3'-UTR, содержащаяся или указанная в таблицах 1 или 3, обеспечивает возможность связывания полипептида, например, 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 1 или 3 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, содержащейся или указанной в таблицах 1 или 3 (или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними), например, из того же транспозона, 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, содержащейся или указанной в таблицах 1-3, например, сегмент полной последовательности транспозона, 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 система или способ, описанные в данном документе, предусматривают 3'-UTR, 5'-UTR или их обе из ретротранспозона из таблицы 3. В некоторых вариантах осуществления 3'-UTR, 5'-UTR или они обе имеют последовательность, содержащую часть полной последовательности ДНК ретротранспозона, показанную в столбце 5 таблицы 3 международной заявки PCT/US2019/048607, которая включена в данный документ посредством ссылки во всей своей полноте, включая таблицу 3. В некоторых вариантах осуществления последовательность нуклеиновой кислоты или аминокислотная последовательность характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью в таблице 3 из PCT/US2019/048607.

В некоторых вариантах осуществления система или способ, описанные в данном документе, предусматривают последовательность нуклеиновой кислоты или аминокислотную последовательность ретротранспозона, описанную в таблице 1 или таблице 2 международной заявки PCT/US2019/048607, которая включена в данный документ посредством ссылки во всей своей полноте, включая таблицы 1 и 2. В некоторых вариантах осуществления последовательность нуклеиновой кислоты или аминокислотная последовательность характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью ретротранспозона, описанной в указанной таблице 1 или таблице 2 из PCT/US2019/048607.

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. Специалист в данной области может по номерам доступа, представленным в данном документе, выравнивать 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.

19. 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. Иллюстративные последовательности транспозаз птиц в таблице 3 включают последовательности из *Taeniopygia guttata* (зебровой амадины; название транспозона R2-1_TG), *Geospiza fortis* (среднего земляного вьюрка; название транспозона R2-1_Gfo), *Zonotrichia albicollis* (белошейной воробьиной овсянки; название транспозона R2-1_ZA) и *Tinamus guttatus* (белогорлого тинаму; название транспозона R2-1_TGut).

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 или 3 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 PCT/US2019/048607, которые включены в данный документ посредством ссылки.

В некоторых вариантах осуществления анализ включает стадии: (1) введения клеток НЕК293Т 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 PCT/US2019/048607, включенной в данный документ посредством ссылки.

В некоторых вариантах осуществления полипептид GeneWriter™ 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 в клетке, описанной в данном документе (например, клетке, содержащей гетерологичную последовательность в сайте-мишени для вставки), содержится 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 PCT/US2019/048607, включенной в данный документ посредством ссылки. 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 транспозона дикого типа, например, из *Taeniopygia guttata*), 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 PCT/US2019/048607, которая включена в данный документ посредством ссылки. 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 PCT/US2019/048607.

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 PCT/US2019/048607, включенной в данный документ посредством ссылки. 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 PCT/US2019/048607 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.

20. 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.

21. *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).

22. Дополнительные домены

Полипептид Gene Writer™ обладает функциями, необходимыми для связывания последовательности ДНК-мишени и нуклеиновой кислоты-матрицы (например, РНК-матрицы), внесения одностороннего разрыва в сайт-мишень и записи (например, обратной

транскрипции) матрицы в ДНК, что приводит к модификации сайта-мишени. В некоторых вариантах осуществления к полипептиду могут быть добавлены дополнительные домены для повышения эффективности процесса. В некоторых вариантах осуществления полипептид Gene Writer™ может содержать дополнительный домен лигирования ДНК для присоединения ДНК, полученной в результате обратной транскрипции, к ДНК сайта-мишени. В некоторых вариантах осуществления полипептид может содержать гетерологичный РНК-связывающий домен. В некоторых вариантах осуществления полипептид может содержать домен, обладающий 5'-3'-экзонуклеазной активностью (например, где 5'-3'-экзонуклеазная активность приводит к усилению репарации изменения сайта-мишени, например, в пользу изменения относительно исходной геномной последовательности). В некоторых вариантах осуществления полипептид может содержать домен, обладающий 3'-5'-экзонуклеазной активностью, например, корректирующей активностью. В некоторых вариантах осуществления записывающий домен, например, домен RT, обладает 3'-5'-экзонуклеазной активностью, например, корректирующей активностью.

В некоторых вариантах осуществления полипептид не содержит домен РНКазы Н. В некоторых вариантах осуществления полипептид содержит домен РНКазы Н, эндогенный по отношению к одному из других доменов. В некоторых вариантах осуществления полипептид содержит домен РНКазы Н, гетерологичный по отношению к другим доменам. В некоторых вариантах осуществления полипептид содержит инактивированный эндогенный домен РНКазы Н.

В некоторых вариантах осуществления Gene Writer, описанный в данном документе, содержит полипептид, ассоциированный с направляющей РНК (gRNA). В определенных вариантах осуществления gRNA содержится в молекуле нуклеиновой кислоты-матрицы. В других вариантах осуществления gRNA отделена от молекулы нуклеиновой кислоты-матрицы. В некоторых вариантах осуществления, где gRNA содержится в молекуле нуклеиновой кислоты-матрицы, молекула нуклеиновой кислоты-матрицы дополнительно 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). В вариантах осуществления спейсер gRNA содержит последовательность, 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 спейсер gRNA направляет активность домена Cas (например, Cas9) на последовательность нуклеиновой кислоты, содержащуюся в молекуле нуклеиновой кислоты-мишени. В некоторых вариантах осуществления, где gRNA содержится в молекуле нуклеиновой кислоты-матрицы, молекула нуклеиновой кислоты-матрицы дополнительно содержит сайт связывания праймера (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 3'-конца). В вариантах осуществления сайт связывания праймера содержит последовательность нуклеиновой кислоты,

характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с последовательностью нуклеиновой кислоты, расположенной на 5'-конце (например, в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40 или 50 нуклеотидов) сайта одностороннего разрыва в молекуле нуклеиновой кислоты-мишени. В вариантах осуществления связывание сайта связывания праймера с молекулой нуклеиновой кислоты-мишени приводит к праймированию TPRT.

23. Компонент системы редактирования генов Gene Writer™, представляющий собой нуклеиновую кислоту-матрицу

Системы Gene Writer™, описанные в данном документе, могут осуществлять модификацию в сайте-мишени ДНК хозяина с использованием последовательности нуклеиновой кислоты-матрицы. В некоторых вариантах осуществления системы Gene Writer, описанные в данном документе, осуществляют транскрипцию последовательности РНК-матрицы в сайты-мишени ДНК хозяина с помощью обратной транскрипции с праймированием мишенью (TPRT). 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. Система Gene Writer™ также может удалять последовательность из генома-мишени или вводить замену, используя последовательность объекта. 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 система Gene Writer содержит молекулу нуклеиновой кислоты-матрицы (например, РНК или ДНК). В некоторых вариантах осуществления молекула нуклеиновой кислоты-матрицы содержит 5'-концевую область гомологии и/или 3'-концевую область гомологии. В некоторых вариантах осуществления 5'-концевая область гомологии содержит последовательность нуклеиновой кислоты, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% сходством 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'-концевая область 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'-концевая область гомологии является гетерологичной по отношению к остальной части молекулы нуклеиновой кислоты-матрицы. В некоторых вариантах осуществления 3'-концевая область гомологии является гетерологичной 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'-концевая область гомологии связывается в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов от сайта одностороннего разрыва в молекуле нуклеиновой кислоты-мишени. В некоторых вариантах осуществления связывание 3'-концевой области гомологии с молекулой нуклеиновой кислоты-мишени позволяет инициировать обратную транскрипцию с праймированием мишенью (TPRT), например, при этом 3'-концевая область гомологии выступает в качестве праймера для TPRT. 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) содержит последовательность гетерологичного объекта. В некоторых вариантах осуществления последовательность гетерологичного объекта может быть

транскрибирована с помощью домена RT полипептида Gene Writer™, например, с обеспечением таким образом введения изменения в сайт-мишень в геномной ДНК. В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину по меньшей мере 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 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 1,000 нуклеотидов (нт) или имеет длину по меньшей мере 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. В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину не более 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 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, 1000 или 2000 нуклеотидов (нт) или имеет длину не более 20, 15, 10, 9, 8, 7, 6, 5, 4 или 3 тысяч оснований. В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину 30-1000, 40-1000, 50-1000, 60-1000, 70-1000, 74-1000, 75-1000, 76-1000, 77-1000, 78-1000, 79-1000, 80-1000, 85-1000, 90-1000, 100-1000, 120-1000, 140-1000, 160-1000, 180-1000, 200-1000, 500-1000, 30-500, 40-500, 50-500, 60-500, 70-500, 74-500, 75-500, 76-500, 77-500, 78-500, 79-500, 80-500, 85-500, 90-500, 100-500, 120-500, 140-500, 160-500, 180-500, 200-500, 30-200, 40-200, 50-200, 60-200, 70-200, 74-200, 75-200, 76-200, 77-200, 78-200, 79-200, 80-200, 85-200, 90-200, 100-200, 120-200, 140-200, 160-200, 180-200, 30-100, 40-100, 50-100, 60-100, 70-100, 74-100, 75-100, 76-100, 77-100, 78-100, 79-100, 80-100, 85-100 или 90-100 нуклеотидов (нт) или имеет длину 1-20, 1-15, 1-10, 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 2-20, 2-15, 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4, 2-3, 3-20, 3-15, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-20, 4-15, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-20, 5-15, 5-10, 5-9, 5-8, 5-7, 5-6, 6-20, 6-15, 6-10, 6-9, 6-8, 6-7, 7-20, 7-15, 7-10, 7-9, 7-8, 8-20, 8-15, 8-10, 8-9, 9-20, 9-15, 9-10, 10-15, 10-20 или 15-20 тысяч оснований. В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30 или 10-20 нуклеотидов, например, имеет длину 10-80, 10-50 или 10-20 нуклеотидов, например, имеет длину приблизительно 10-20 нуклеотидов. В некоторых вариантах осуществления РНК-матрица содержит последовательность, указанную в таблице 43, или последовательность с по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

В определенных вариантах осуществления нуклеиновая кислота-матрица содержит индивидуально адаптированную последовательность РНК-матрицы, которую 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 транскены, 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 RLE, являться сконструированной таким образом, чтобы она содержала гетерологичные кодирующие последовательности, или представлять собой комбинации вышеперечисленного.

Нуклеиновая кислота-матрица (например, РНК-матрица) может характеризоваться некоторой степенью гомологии с ДНК-мишенью. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью нуклеиновой кислоты-матрицы (например, РНК-матрицы) может выступать в качестве области отжига с ДНК-мишенью, так что ДНК-мишень расположена таким образом, чтобы праймировать обратную транскрипцию нуклеиновой кислоты-матрицы (например, РНК-матрицы). В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, 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 больше оснований с точной гомологией с ДНК-мишенью на 3'-конце РНК. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, 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 больше 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 нуклеиновой кислоты-матрицы (например, РНК-матрицы). В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, 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 не содержащего LTR ретротранспозона, например, не содержащего LTR ретротранспозона, описанного в данном документе, например, не содержащего LTR ретротранспозона из таблицы 1 или 3.

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 шпильчатые 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, гибридизирующей 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 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 нуклеиновая кислота-матрица (например, 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 не содержащего LTR ретротранспозона, например, не содержащего LTR ретротранспозона, описанного в данном документе.

Нуклеиновая кислота-матрица (например, РНК-матрица) системы обычно содержит последовательность объекта (например, последовательность гетерологичного 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, 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, вторую 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 с мишенью. 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. Например, в некоторых вариантах осуществления первая РНК содержит первый конъюгирующий домен, а вторая РНК содержит второй конъюгирующий домен, и первый и второй конъюгирующие домены способны гибридизоваться друг с другом, например, в жестких условиях. В некоторых вариантах осуществления жесткие условия гибридизации включают гибридизацию в 4х хлориде натрия/цитрате натрия (SSC) при приблизительно 65°C с последующей промывкой в 1х SSC при приблизительно 65°C.

В некоторых вариантах осуществления последовательность объекта может содержать открытую рамку считывания. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, 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, CTGACCCTTCTCTCTCTCCCCAGAG (SEQ ID NO: 1601) (из гена HBB человека) и TTTCTCTCCCAAG (SEQ ID NO: 1602) (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(А)-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 В (HPRE) или вируса гепатита сурков (WPRE).

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 ее активности, например, может препятствовать вставке последовательности гетерологичного объекта в геном. Соответственно, система будет редактировать геном клеток-мишеней более эффективно, чем она редактирует геном клеток, не являющихся мишенями, например, последовательность гетерологичного объекта будет вставляться в геном клеток-мишеней более эффективно, чем в геном клеток, не являющихся мишенями, или вставка или делеция образуются более эффективно в клетках-мишенях, чем в клетках, 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 GeneWriter™". В некоторых вариантах осуществления, например, для печеночных показаний, miRNA выбрана из таблицы 4 WO2020014209, включенной в данный документ посредством ссылки.

В некоторых вариантах осуществления последовательность объекта может содержать некодирующую последовательность. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать регуляторный элемент, например, промоторную или энхансерную последовательность или сайт связывания с miRNA. В некоторых вариантах осуществления интеграция последовательности объекта в сайт-мишень приведет к повышению экспрессии эндогенного гена. В некоторых вариантах осуществления интеграция последовательности объекта в сайт-мишень приведет к снижению экспрессии эндогенного гена. В некоторых вариантах осуществления

нуклеиновая кислота-матрица (например, 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 ТАТА. В некоторых вариантах осуществления промотор содержит элемент распознавания В. В некоторых вариантах осуществления промотор имеет один или несколько сайтов связывания для факторов транскрипции.

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 GeneWriter™, например, как описано в данном документе. Система, содержащая последовательность тканеспецифического промотора в РНК-матрице, также может применяться в комбинации с ДНК, кодирующей полипептид GeneWriter™, управляемой тканеспецифическим промотором, например, для достижения более высоких уровней белка GeneWriter™ в клетках-мишенях, чем в клетках, не являющихся мишенями. В некоторых вариантах осуществления, например, для печеночных показаний, тканеспецифический промотор выбран из таблицы 3 WO2020014209, включенной в данный документ посредством ссылки.

В некоторых вариантах осуществления система Gene Writer, например, ДНК, кодирующая полипептид Gene Writer, ДНК, кодирующая РНК-матрицу, или ДНК или РНК, кодирующие последовательность гетерологичного объекта, разработана таким образом, что один или несколько элементов функционально связаны с тканеспецифическим промотором, например, промотором, который является активным в Т-клетках. В дополнительных вариантах осуществления промотор, активный в Т-клетках, является неактивным в других типах клеток, например, В-клетках, НК-клетках. В некоторых вариантах осуществления промотор, активный в Т-клетках, получен из промотора гена, кодирующего компонент Т-клеточного рецептора, например, TRAC, TRBC, TRGC, TRDC. В некоторых вариантах осуществления промотор, активный в Т-клетках, получен из промотора гена, кодирующего компонент специфичного для Т-клеток белкового кластера

дифференцировки, например, CD3, например, CD3D, CD3E, CD3G, CD3Z. В некоторых вариантах осуществления специфичные для Т-клеток промоторы в системах Gene Writer обнаруживают посредством сравнения общедоступных данных об экспрессии генов по типам клеток и выбора промоторов из генов с повышенной экспрессией в Т-клетках. В некоторых вариантах осуществления промоторы можно выбирать в зависимости от желаемой широты экспрессии, например, промоторы, которые активны только в Т-клетках, промоторы, которые активны только в НК-клетках, промоторы, которые активны как в Т-клетках, так и в НК-клетках.

В некоторых вариантах осуществления РНК-матрица содержит последовательность микроРНК, последовательность siRNA, последовательность направляющей РНК, последовательность РНК, взаимодействующей с Piwi.

В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) содержит сайт, который координирует эпигенетическую модификацию. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) содержит инсульт хроматина. Например, нуклеиновая кислота-матрица (например, РНК-матрица) содержит сайт для CTCF или сайт, являющийся мишенью метилирования ДНК.

В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, 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 не кодирующую последовательность, такую как последовательность, кодирующая микроРНК).

В некоторых вариантах осуществления последовательность объекта из нуклеиновой кислоты-матрицы (например, РНК-матрицы) вставляется в геном-мишень в эндогенном интроне. В некоторых вариантах осуществления последовательность объекта из нуклеиновой кислоты-матрицы (например, 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 нуклеиновой кислоты-матрицы (например, РНК-матрицы) вставляется в геном-мишень в геномном сайте "безопасная гавань", таком как AAVS1, CCR5, ROSA26 или локус альбумина. В некоторых вариантах осуществления Gene Writer применяют для интеграции CAR в локус константной области α -цепи Т-клеточного рецептора (*TRAC*) (Eyuquem et al Nature 543, 113-117 (2017)). В некоторых вариантах осуществления Gene Writer применяют для интеграции CAR в локус константной области β -цепи Т-клеточного рецептора (*TRBC*). Многие другие сайты "безопасная гавань" были определены с помощью вычислительных подходов (Pellenz et al Hum Gen Ther 30, 814-828 (2019)) и могут применяться для интеграции, опосредованной Gene Writer. В некоторых вариантах осуществления последовательность объекта из

нуклеиновой кислоты-матрицы (например, РНК-матрицы) добавляется в геном в межгенной или внутригенной области. В некоторых вариантах осуществления последовательность объекта из нуклеиновой кислоты-матрицы (например, 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 нуклеиновой кислоты-матрицы (например, 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 нуклеиновой кислоты-матрицы (например, 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.).

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 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™).

Можно использовать альтернативные ориентации мотивов РНК-матрицы, например, чтобы ограничить интеграцию сайта-мишени с желаемой генетической полезной нагрузкой. В некоторых вариантах осуществления домены ассоциации полипептида могут быть расположены с 5'-стороны от желаемой последовательности-матрицы. Например, последовательность гетерологичного объекта может быть расположена ниже 5'-UTR и 3'-UTR, придавая ориентацию 5'-3' 5'-UTR-3'-UTR-(последовательность гетерологичного объекта). В других вариантах осуществления выше последовательности гетерологичного объекта добавляется только 3'-UTR. Например, это придает ориентацию 5'-3' 3'-UTR-(последовательность гетерологичного объекта). В определенных вариантах осуществления кодирующая полипептид область и последовательность гетерологичного объекта могут быть закодированы в одной и той же молекуле, но при этом 5'-UTR (например, 5'-UTR из ретротранспозона R2) находится между двумя областями, например, придавая ориентацию 5'-3' (последовательность, кодирующая полипептид)-5'-UTR-(последовательность гетерологичного объекта).

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 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 40 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 40. Белки CRISPR/Cas, виды и мутации

Вариант	Исходный хозяин	Последовательность белка	Мутация
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGV RVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARR LLKREGVLQAADFDEGLIKSLPNTPWQLRAAALDRKLTPL WSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNA HALQTGDFRTPAELALNKFEEKESGHIRNQRGDYSHTFSRKDL QAEILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTCLNNLRILEQ GSRPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFK GLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNL SSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFD KFBVQISLKLRRIVPLMEQGKRYDEACAEIYGDHYGKKNT KIIYLPPIPADEIRNPVVLRLALSQARKVINGVVRRYGSPAR TAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFV GEPKSKDILKLRLYEQQHKGKCLYSGKEINLVLNKGVEID HALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKD NSREWQEFKARVETSRFPRSKKQRILLQKFDGDFKECNLND TRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFW GLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEM NAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKP DGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAP NRKMSGAHKDTLRSKRFBKHNEKISVKRVWLTEIKLADLE NMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFY KKGGLVKAVRVEKTQESGVLLNKKNAITYADNGDMVRVD VFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSY	N611A

		TFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWH DKGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPV R	
PpnC as9	Pasteu rella pneum otropic a	MQNNPLNYILGLDLGASIGWAVVEIDEESSPIRLIDVGVRTFE RAEVAKTGESLALSRRRLARSSRRLIKRRAERLKKAKRLLKAE KILHSIDEKLPINWQLRVKGLKEKLERQEWAAVLLHLSKHR GYLSQRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEI AVKKFQVEEGHIRNQRGSYTHTFSRDLLAEMELLFQRQAEL GNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEP EYKAAKNSYSAERFVWLTKLNNLRILENGTERALNDNERFA LLEQPYEKSCLTYAQVRAMLALSDNAIFKGVRYLGEDKKT V ESKTTLIEMKFYHQIRKTLGSAELKKEWNEKGNLSDLLDEIG TAFSLYKTDDDICRYLEGKLPERVLNALLNENLNFDFIQLSLK ALHQILPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLPTIP ADEIRNPVVLRLTLQARKVINAVVRLYGSPARIHIETAREVG KSYQDRKKLEKQQEDNRKQRESAVKKFKEMFPHFVGEPKG KDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDHALP FSRTWDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSER WQHFVVRVQTSGFSAKKQRILNHLKDEKGFIERNLNDTRY VARFLCNFIADNMLLVGKGKRNVFASNGQITALLRHRWGLQ KVREQNDRHHALDAVVVACSTVAMQQKITRFVRYNEGNVF SGERIDRETGEIPLHFPSPWAFFKENVEIRIFSENPKLELENRL PDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSARLN EGLSVLKVPLTQLKLSLDMVNRDREIALYESLKARLEQFG NDPAKAFAPFYKKGGALVKAVRLEQTQKSGVLVRDGNV ADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNRAATQ GKDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTIFGYFNG LNRATSNINIKEHDLKSKGKLGILEVGVKLALSLEKYQVD ELGKNIRPCRPTKRQHVR	N605 A
SauC as9	Staphy lococ cus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA	N580 A

		YNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTS FLRRKWFKKERNKGYKHHAEADALIIANADFIFKEWKKLDK AKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIKDF KDYKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNG LYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQYG DEKNPLYKYYEETGNLYTKYSKKDNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYNNDLI KINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP RIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
SauC as9- KKH	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTS	N580 A

		FLRRKWKFKKERNKGYKHHAEDALIIANADFIKWKLDK AKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIKDF KDYKYSHRVDKKPNRKLINDTL YSTRKDDKGNTLIVNNLNG LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYG DEKNPLYKYEEETGNLYTKYSKKDNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLI KINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
Sauri Cas9	Staphy lococ us auricul aris	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPE ADSENNNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMIDL NNVPKSTDPTYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNIS VSMGDEEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNI NKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQYI DLVSTRREYFEGPGNGSPYGWDGDLKWKYEKLMGRCTYFP EELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEYEYKYH IIENVFKQKKNP TLKQIAKEIGVQDYDIRGYRITKSGKPQFTSF KLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALD QLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPEN QMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSI KVINAVINRFGLPEDIIELAREKNSKDRRKFINKLQKQNEATR KKIEQLLAKYGNTNAKYMIEKIKLHDMQEGKCLYSLEAIPLE DLLSNPTHYEV DHIIPRSVSFDNSLNNKVLVKQSENSKKGNR TPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEE RDINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAV KVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANA DFLFKTHKALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFE TPKQVKNIKQFRDFKYSHRVDKKPNRQLINDTL YSTREIDGE TYVVQTLKDL YAKDNEKVKKLFTERPQKILMYQHDPKTFEK LMTILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKI KYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFREFDIYKCEQG YKMVSIGYLDVLKKNYYPKDKYEA EKQKKKIKESDLFV GSFYNDLIMYEDELFRVIGVNSDINN LVELNMVDITYKDFC EVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQ LIFKRGEL	N588 A

Sauri Cas9- KKH	Staphy lococ us auricul aris	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPE ADSENNNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMIDL NNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNIS VSMGDEEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNI NKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQYI DLVSTRREYFEGPGNGSPYGWDGDLKWEKLMGRCTYFP EELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEYEYKYH IENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKSGKPQFTSF KLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALD QLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPEN QMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSI KVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATR KKIEQLLAKYGNTNAKYMIEKIKLHDMQEGKCLYSLEAIPLE DLLSNPTHYEV DHIIPRSVSFDNSLNNKVLVKQSENSKKGNR TPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEE RDINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAV KVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANA DFLFKTHKALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFE TPKQVKNIKQFRDFKYSHRVDKKPNRKLINDTLYSTREIDGE TYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEK LMTILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKI KYIDKKLGSYLDVSNKY PETQNKLVKLSLKSF RFDIYKCEQG YKMVSIGYLDVLKKDNYYYIPKDKYEA EKQKKKIKESDLFV GSFYKNDLIMYEDELFRVIGVNSDINNLEVELNMVDITYKDFC EVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQ LIFKR GEL	N588 A
ScaC as9- Sc++	Strept ococcu s canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRK SIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQ EIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLADE VAYHRNYPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHF LIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGI LSARLSKSKRLEKLI AVFPNEKKNGLFGNIIALALGLTPNFKS NFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKN LSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKT LVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGLAT	N872 A

		EEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIP HQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGPL ARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTN FDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPE FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEII GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSR KMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEE IEKAQVSGQGDSLHEQIADLAGSPAIIKKGILQTVKIVDELVKV MGHKPENIVIAMARENQTTTKGLQQSRERKKRIEEGIKELES QILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFI KRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLK SKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGITALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANGAIRKRPLIETNGETGEVVWNKEKDFATV RKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGI TIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENG RRRMLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNN LGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFA VSDSILLSNSFVSLLKYTSFGASGGFTFLDLVDKQGRLRYQTV TEVLDATLIYQSITGLYETRTDLSQLGGD	
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT	N863A

		RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
SpyC as9- NG	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLLIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV	N863 A

		DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHHPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGF VSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASARF LQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVE QHKHYLDEIIQISEFSKRVLADANLDKVL SAYNKHRRDKPIR EQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDT LIHQSTGLYETRIDLSQLGGD	
SpyC as9- SpR Y	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLLIYLAHAMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASASQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER	N863 A

		LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPPKYGGF LWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAK QLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPI REQAENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
St1C as9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENN LVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINL NPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDD GNSSVG DY AQIVKENS KQLETKTPGQIQLERYQTYGQLRGDF TVEKDGGKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI NRYLEILTGKRKYHGP GNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKL SKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYR IDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEAL EHEFADGSFSQKQVDEL VQFRKANSSIFGK GWHNFSVKLMMEIPELYETSEEQMTILTRLGKQKTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVI EMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTL SNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR	N622 A

		GQFTSQLRRHWGIEKTRDITYHHHAVDALIIAASSQLNLWKK QKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDT LKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGKDKA DETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQT FEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYS KKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWR ADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIK KKEGVDSSEFKFTLYKNDLLL VKDTETKEQQLFRFLSRTMP KQKHVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLG KSNISYKVRTDVLGNQHIIKNEGDKPKLDF	
BlatC as9	Brevib acillus lateros porus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKN GEALAVPRREARSSRRRLRRKKHRIERLKHMFVRNGLAVDI QHLEQTLRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIHLAQ RRGFQSNRKTDGSSSEDGQVLVNV TENDRLMEEKDYRTVAE MMVKDEKFSDHKRNKNGNYHGVVSRSSLLVEIHTLFETQRQ HHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPK EKRAPKASWHFQYFMLLQTINHIRTNVQGTRSLNKEEIEQV VNMALTKSKVSYHDTRKILDLSEEYQFVGLDYGKEDEKKKV ESKETIIKLDDYHKLNKIFNEVELAKGETWEADDYDTVAYAL TFFKDDIEDIRDYLQNKYKDSKNRLVKNLANKEYTNELIGKV STLSFRKVGHL SLKALRKIIPFLEQGMTYDKACQAAGFDFQG ISKKKRSVVLVIDQISNPVVRAL TQTRKVIN ALIKKYGSPE TIHIETARELSKTFDERKNITKDYKENRDKNEHAKKHLSELGI INPTGLDIVKYKLWCEQQGRCMYSNPISFERLKESGYTEVD HIIPYSRSMNDSYNNRVL VMTRENREKGNQTPFEYMGNDTQ RWYEFQRVTTPNQIKKEKRQNLLLKGFTNRRELEMLERNL NDTRYITKYL SHFISTNLEFSPSDKKKKVVNTSGRITSHLSR WGLEKNRGQNDLHHAMDAIVIAVTSDSFIQQVTNYYKRKER RELNGDDKFPLPWKFFREEVIARLSPNPKEQIEALPNH FYSED ELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITA KKTALVDISYDKNGDFNMYGRETD PATYEAIKERYLEFGGN VKKAFSTDLHKPKKDGTGGLIKSVRIMENKTLVHPVNKGK GVVYNSSIVRTDVFQRKEKYLLPVYVTDVTKGKLPNKVIV AKKGYHDWIEVDDSFTFLFSLYPNDLIFIRQNP KKKISLKKRI ESHSISDSKEVQEIHAYYKGVDSSTA AIEFIIHDGSYYAKGVG	N607 A

		VQNLDCEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDV NSIKSTSR	
cCas 9-v16	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYEYEFQIENNVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNNLLRSYFRVNNLDVKVKSINGGFTS FLRRKWKFKKERNKGYKHHAE DALIANADFIKWKKLDK AKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDF KDYKYSHRVDKKPNRKLINDTL YSTRKDDKGNTLIVNNLNG LYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQYG DEKNPLYKYEEETGNLYTKYSKKDNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLI KINGELYRVIGVNSDKNNLIEVNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580 A
cCas 9-v17	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYEYEFQIENNVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI	N580 A

		TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTS FLRRKWFKKERNKGYKHHAEDALIANADFIFKEWKKLDK AKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDF KDYKYSHRVDKKPNRKLINDTL YSTRKDDKGNTLIVNNLNG LYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQYG DEKNPLYKYEEETGNLYTKYSKKDNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLI KINGEL YRVIGVNNSTRNIVELNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
cCas 9-v21	Staphy lococ cus aureus	MKRNILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRHRIQRVKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYEKFQIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVSTGKPEFTNLKVYHDIKDI TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTS FLRRKWFKKERNKGYKHHAEDALIANADFIFKEWKKLDK AKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDF	N580 A

		KDYKYSHRVDKKPNRKLINDTL YSTRKDDKGNTLIVNNLNG LYDKDNDKLLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYG DEKNPLYKYEEETGNLYLTKYSSKDNNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLLKKISNQAEFIASFYKNDLI KINGEL YRVIGVNSDDRNIELNMIDITYREYLENMNDKRPPH IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
cCas 9-v42	Staphy lococ us aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVEN NEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEE DTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEIIEAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEII RTT GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYE VDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNNLSYFRVNNLDVKVKSINGGFTS FLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDK AKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDF KDYKYSHRVDKKPNRKLINDTL YSTRKDDKGNTLIVNNLNG LYDKDNDKLLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYG DEKNPLYKYEEETGNLYLTKYSSKDNNGPVIKKIKYYGNKLNA HLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKN LDVIKKENYYEVNSKCYEEAKKLLKKISNQAEFIASFYKNDLI KINGEL YRVIGVNNNRLNKKIELNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580 A
CdiC as9	Coryn ebacte rium	MKYHVGIDVGTF SVGLAAIEVDDAGMPIKTL SLVSHIHDSGL DPDEIKSAVTRLASSGIARRTRRL YRRKRRRLQQLDKFIQRQ GWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVA	H573 A (альт

	diphtheriae	LRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRASG QPVPETATVGQMVTLCELGTKLRGEGGVLSARLQQSDYAR EIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQ PGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKN LVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTATMTDDGE RAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKA LSNAEVDDFDSPEGAKVQAFFADLDDDDVHAKLDSLHLPVGR AAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRI GEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIHVREG FVTEKRAREMDGDMRRRAARNAKLFQEMQEKLNVQGKPSR ADLWRYQSVQRQNCQCA YCGSPITFSNSEMDHIVPRAGQS TNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAV ERTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARS MESVAWMANELRSRVAQHFA SHGTTVRVYRGS LTAEARRA SGISGKLKFFDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLA VRSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKM EKLSALLTEDLRDDR VVVM SNVRLRLGNGSAHKETIGKLSK VKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPER HIRVNGTHVYAGDNIGLFPVSAGSIALRGGY AELGSSFHAR VYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSM RQAEKKLRDALATGNAEYLGWL VVDEL VVDTSKIATDQV KAVEAELGTIRRWVRVDGFFSPSKLRLRPLQMSKEGIKKESAP ELSKIIDRPGWLPAVNKLFSDGNVTVVRDSLGRVRLESTAH LPVTWKVQ	ернат ивная)
CjeC as9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGES LALPRRLARSARKRLARRKARLNHLKHLIANEFKLN YEDYQ SFDESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAK RRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYK EYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKK QREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDE KRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNAL LNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGT YFIEFKKY KEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLN QNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNE LNLKVAINEDKKDFLPAFNETYKDEV TNPVVLRAIKEYRK	N582 A

		VLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENY KAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKI KISDLQDEKMLEIDHIYPYSRSFDDSYMNVLVFTKQNEKL NQTPFEAFGNDSAKWQKIEVLAKNLPKKQKRILDKNYKDK EQKNFKDRNLNDTRYIARLVLNYTKDYLDLPLSDDENTKL NDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHA IDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKN KRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRK EEEFYQSYGGKEGVKALELGKIRKVNGKIVKNGDMFRVDI FKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWI LMDENYEFCSLYKDSLILIQTKDMQEPEFVYNAFTSSTVSL IVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEK YIVSALGEVTKAEFRQREDFKK	
GeoC as9	Geoba cillus stearot hermo philus	MRYKIGLDIGITSVGWAVMNLDPRIEDLGVRIFDRAENPQTG ESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELD KLFEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGF KSNRKSERSNKENSTMLKHIEENRAILSSYRTVGEMIVKDPK FALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEE FENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYT FQSFIWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKIT YHDIRTLLHLPDDTYFKGIVYDRGESRKQENENIRFLELDAYH QIRKAVDKVYGGKSSSFLPIDFDTFGYALTLFKDDADIHSY LRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHL SLK ALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIP PIANPVVMRALTQARKVVNAIIKKYGPVSIHIELARDLSQTF DERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKF KLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDS YTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTN KQFSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANF IREHLKFAESDDKQKVYTVNGRVT AHLRSRWEFNKNREESD LHHA VDAVIVACTTPSDIAKVTA FYQRREQNKELAKKTEPHF PQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQPV FVSRMPKRSVTGA AHQETLRRYVGIDERSGKIQT VVKTLSE IKLDASGHFPMYGKESDPRTYE AIRQRLLEHNNDPKKAFQEP LYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVA YNSNIVR	N605 A

		VDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKE MTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEINVKDVFVY YKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLG NIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	
iSpy Mac Cas9	Strept ococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHHPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEIQTVGQNGGLFDDNPKSPLEVTSPSKLVPLKKELNP KKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKF LRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKG NQLVVSKKSQILLYHAHHLSDLSNDYLNHNQQFDVLFNE IISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFT	N863 A

		QLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGL YETRVDLSKIGEDSGGSGGSKRTADGSEFES	
Nme Cas9	Neisse ria menin gitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGV RVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRR LLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLE WSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNA HALQTGDFRTPAELALNKFEEKESGHIRNQRSDYSHTFSRKDL QAEILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA VQKMLGHCTFEPAPKAANKNTYTAERFIWLTCLNNLRILEQ GSRPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFK GLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNL SPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDK FVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEK IYLPPIPADEIRNPVVLRLALSQARKVINGVVRRYGSPARIHIET AREVGKSFKDRKEIEKRQEENNRKDREKAAAKFREYFPNFVG EPKSKDILKLRLYEQQHGKCLYSGKEINLGRLENEKGYVEIDH ALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNS REWQEFKARVETSFRFPRSKKQRILLQKFDEEDGFKERNLNDTR YVNRFLCQFVADRMRLTGKGGKRVFASNGQITNLLRGFWG LRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPD GKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPN RKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLDLEKM VNREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGN RTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFE KGDKYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSF NFKFSLHPNDLVEVITKKARMFGYFASCHRG TGNNIRIHDLD HKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPV R	N611 A
ScaC as9	Strept ococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRK SIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQ EIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLADE VAYHRNYPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHF LIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGI LSARLSKSKRLEKLIIVFPNEKKNGLFGNIIALALGLTPNFKS	N872 A

		NFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKN LSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKT LVRQQFPEKYAEIFKDDTKNGYAGYVVGIGIKHRKRTTKLATQ EEFYKFIKPILEKMDGAEELLAKLNRDDLRLKQRTFDNGSIPH QIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGPLA RGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNF DEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERM RKPEF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIG VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRK MINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEI EKAQVSGQGDSLHEQIADLAGSPAIIKKGILQTVKIVDELVKV MGHKPENIVIAMARENQTTTKGLQQSRERKKRIEEGIKELES QILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFI KRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLK SKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANG EIRKRPLIETNGETGEVVWNKEKDFATV RKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKEVGKAKKLKSVKVLVGI TIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENG RRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNN LGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFA VSDSILLSNSFVSLKYTSFGASGGFTFLDLVDKQGRLRYQTV TEVL DATLIYQSITGLYETRTDLSQLGGD	
ScaC as9- HiFi- Sc++	Strept ococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRK SIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQ EIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLADE VAYHRNYPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHF LIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGI LSARLSKSKRLEKLI AVFPNEKKNGLFGNIIALALGLTPNFKS NFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKN LSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKT	N872 A

		LVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGLAT EEEFYKFIKPILEKMDGAEELLAKLNRDDLRLKQRTFDNGSIP HQIHLKELHAILRRQEEFYPLKENREKIEKILTRIPYYVGPL ARGNSRFAWLTRKSEEAITPWNFEVVDKGASAQSFIERMTN FDEQLPNKKVLPKHSLLEYEFTVYNELTKVKYVTERM RKPE FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEII GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSR KMINGIRDKQSGKTILDFLKSDGFSNANFMQLIHDDSLTFKEE IEKAQVSGQGDSLHEQIADLAGSPAIIKKGILQTVKIVDELVKV MGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELES QILKENPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFI KRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLK SKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANG EIRKRPLIETNGETGEV VWNKEKDFATV RKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLLSVKVLVGI TIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENG RRRMLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNN LGYIEQHREEFK EIFEKIIDFSEKYILKNKVNSNLKSSFDEQFA VSDSILLSNSFVSLLKYTSFGASGGFTFLDLVDKQGRRLRYQTV TEVL DATLIYQSITGLYETRTDLSQLGGD	
SpyC as9- 3var- NRR H	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLV DSTDKADLR LIYLA LAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADFLAAKNL SDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKAL VRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILR	N863 A

		RQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGGHKKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGF NSPTAAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEK NPIGFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAG VLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARDKPI REQAENIIHLFTLTNLGVPAAFKYFDTTIDKKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
SpyC as9- 3var- NRT H	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKAL VRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILR RQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL	N863 A

		PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGGHKKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGF NSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEK NPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAS VLHKGNELALPSKYVNFLYLASHYEKLKGSSDNKQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHDKPI REQAENIIHLFTLTNLGASAAFKYFDTTIGRKL YTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
SpyC as9- 3var- NRC H	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLV DSTDKADLRLLIYLA LAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAQLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKAL VRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILR RQGDFYPFLKDNREKIEKILTFRIPYVVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASASQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL	N863 A

		GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGGHKKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPPKYGGF NSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAG VLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLD ATLIRQSITGLYETRIDLSQLGGD	
SpyC as9- HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVSTDKADLRLLIYLALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG	N863 A

		KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSIITGLYETRIDLSQLGGD	
SpyC as9- QQR 1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIV	N863 A

		IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKYGGF DSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAR ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADAQLDKVLSAYNKHDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTFKQKQYRSTKEVLD ATLIHQSTGLYETRIDLSQLGGD	
SpyC as9- SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHHPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP	N863 A

		QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF LWPTVAYSVLVVAKEVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAK QLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
SpyC as9- VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIV IEMARENQTTQKGQKNSRERMKRIEELGKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR	N863 A

		QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF VSPTVAYSVLVVAKEVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	
SpyC as9- VRE R	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLV DSTDKADLR LIYLA LAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR KDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEF	N863 A

		VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF VSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAR ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDA TLIHQ SITGLYETRIDLSQLGGD	
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLR LIYLA LAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALV RQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRR QEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDS LHEHIANLAGSPAIIKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWR QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRK DFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFV YGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVN	N863A

		IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSSFEN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGV LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKHDKPIR EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DAT LIHQSITGLYETRIDLSQLGGD	
SpyC as9- xCas- NG	Strept ococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDTKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALV RQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRR QEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDS LHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWR QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRK DFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFV YGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVN IVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPPKKYGGFV SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSSFEN	N863 A

		PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFL QKGNELALPSKYVNFLYLASHYEKLKGSPEDEKQQLFVEQ HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIRE QAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVL DATL IHQSITGLYETRIDLSQLGGD	
St1C as9- CNR Z106 6	Strept ococcus thermo philus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENN LVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINL NPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDASDD GNSSVG DY AQIVKENS KQLETKTPGQIQLERYQTYGQLRGDF TVEKDGGKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI NRYLEILTGKRKYHGP GNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKL SKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYR IDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEAL EHEFADGSFSQKQVDEL VQFRKANSSIFGK GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVI EMARETNEDDEKKAIQKI QKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTL SNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKK QKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTL KSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKD ETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTF EKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRKYS KKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPW RTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNDI KKKEGVDS DSEFKFTLYKNDLLL VKDTETKEQQLFRFLSRTL PKQKH YVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLA KSNIS IYKVRTDVLGNQHIIKNEGD KPKLDF	N622 A
St1C as9-	Strept ococcus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENN LVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINL NPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDASDD	N622 A

LMG 1831	thermo philus	GNSSVGDYAIQIVKENSQLETKTPGQIQLERYQTYGQLRGDF TVEKDGGKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI NRYLEILTGKRKYYHGPNGNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKL SKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYR IDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEALEHEFADGSFSQKQVDEL VQFRKANSSIFGK GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVI EMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKK QKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFDLTL KSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGKDKKD ETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHPQTF EKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRKYS KKNGNGPEIKSLKYYDSKLLGNPIDITPENSNNKVVQLSLKPW RTDVYFNKNTGKYEILGLKYADLQFEKKTGTYSQEKYNGI MKEEGVDSSEFKFTLYKNDLLL VKDTETKEQQLFRFLSRT MPNVKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLG KSNISYKVRTDVLGNQHIIKNEGDKPKLDF	
St1C as9- MTH 17CL 396	Strept ococcus thermo philus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENN LVRRTNRQGRRLARRKKHRRVRLNRLFEEGLITDFTKISINL NPYQLRVKGLTDELSNEELFIALKNMVKHRSYLLDDASDD GNSSVGDYAIQIVKENSQLETKTPGQIQLERYQTYGQLRGDF TVEKDGGKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI NRYLEILTGKRKYYHGPNGNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKL SKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYR IDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEALEHEFADGSFSQKQVDEL VQFRKANSSIFGK GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK	N622 A

		TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVI EMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKK QKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDT LKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKA DETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQT FEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYS KKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWR TDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIK VREGVDENSEFKFTLYKNDLLLLKDSengeQILLRFTSRNDT SKHYVELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSN ISIIKVRTDVLGNQHIIKNEGDKPKLDF	
St1C as9- TH14 77	Streptococcus thermophilus	MSDLVGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENN LVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINL NPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDD GNSSVGDYAQIVKENSQKLETKTPGQIQLERYQTYGQLRGDF TVEKDGGKHLINVFPTSAYRSEALRILQTQQEFNPQITDEFI NRYLEILTGKRKYHGPNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKL SKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYR IDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEALHEFADGSFSQKQVDELVQFRKANSSIFGK GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVI EMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKK QKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDT	N622 A

	LKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKA DETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQT FEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYS KKGNNGPEIKSLKYYDSKLGHNHIDITPKDSNNKVVQLSLKPWR TDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIK VREGVDENSEFKFTLYKNDLLLLKDSENGINEQILLRFTSRNDT SKHYVELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSN ISIIKVRTDVLGNQHIIKNEGDKPKLDF	
--	--	--

В таблице 9В представлены параметры для определения необходимых компонентов для разработки gRNA и/или РНК-матриц для применения вариантов Cas, перечисленных в таблице 3А, 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 одностороннего разреза ssDNA в мишени является важным для разработки 3'-концевой области РНК-матрицы, которую необходимо отжечь с последовательностью непосредственно с 5'-стороны от одностороннего разреза с целью инициации обратной транскрипции с праймированием мишенью.

Таблица 9В. Параметры для определения необходимых компонентов для разработки gRNA и/или РНК-матриц для применения вариантов Cas, перечисленных в таблице 9А, для Gene Writing

Вариант	PAM	Разрез	Ярус	Спейсер (min)	Спейсер (max)	crRNA	Тетрапетля	tracrRNA
Nme2 Cas9	NNN NCC	-3	1	22	24	GTTGT AGCTC CCTTT CTCAT TTCG	GAAA	CGAAATGAGA ACCGTTGCTA CAATAAGGCC GTCTGAAAAG ATGTGCCGCA ACGCTCTGCC CCTTAAAGCTT CTGCTTTAAG

								GGGCATCGTT TA
PpnCa s9	NNN NRT T		1	21	24	GTTGT AGCTC CCTTTT TCATT TCGC	GAAA	GCGAAATGAA AAACGTTGTT ACAATAAGAG ATGAATTTCTC GCAAAGCTCT GCCTCTTGAA ATTTCGGTTTC AAGAGGCATC TTTTT
SauCa s9	NNG RR;N NGR RT	-3	1	21	23	GTTTT AGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAATGCCG TGTTTATCTCG TCAACTTGTTG GCGAGA
SauCa s9- KKH	NNN RR;N NNR RT	-3	1	21	21	GTTTT AGTAC TCTGT AAT	GAAA	ATTACAGAAT CTACTAAAAC AAGGCAAAT GCCGTGTTTAT CTCGTCAACTT GTTGGCGAGA
Sauri Cas9	NNG G	-3	1	21	21	GTTTT AGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAATGCCG TGTTTATCTCG TCAACTTGTTG GCGAGATTTTT
Sauri Cas9- KKH	NNR G	-3	1	21	21	GTTTT AGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAATGCCG TGTTTATCTCG TCAACTTGTTG GCGAGATTTTT

ScaCa s9- Sc++	NNG	-3	1	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCa s9	NGG	-3	1	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCa s9-NG	NG (NGG =NG A=N GT> NGC)	-3	1	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
SpyCa s9- SpRY	NRN >NY N	-3	1	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
StlCa s9	NNA GAA W>N NAG GAW =NN GGA AW	-3	1	20	20	GTCTT TGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
BlatC as9	NNN NCN	-3	1	19	23	GCTAT AGTTC	GAAA	GGTAAGTTGC TATAGTAAGG

	AA> NNN NCN DD> NNN NC					CTTAC T		GCAACAGACC CGAGGCGTTG GGGATCGCCT AGCCCGTGTTT ACGGGCTCTC CCCATATTCA AAATAATGAC AGACGAGCAC CTTGGAGCAT TTATCTCCGAG GTGCT
cCas9 -v16	NNV ACT; NNV ATG M;N NVA TT;N NVG CT;N NVG TG;N NVG TT	-3	2	21	21	GUCUU AGUAC UCUG	GAAA	CAGAAUCUAC UAAGACAAGG CAAAAUGCCG UGUUUAUCUC GUCAACUUGU UGGCGAGAUU UUUUU
cCas9 -v17	NNV RRN	-3	2	21	21	GUCUU AGUAC UCUG	GAAA	CAGAAUCUAC UAAGACAAGG CAAAAUGCCG UGUUUAUCUC GUCAACUUGU UGGCGAGAUU UUUUU
cCas9 -v21	NNV ACT; NNV ATG	-3	2	21	21	GUCUU AGUAC UCUG	GAAA	CAGAAUCUAC UAAGACAAGG CAAAAUGCCG UGUUUAUCUC

	M;N NVA TT;N NVG CT;N NVG TG;N NVG TT							GUCAACUUGU UGGCGAGAUU UUUUU
cCas9 -v42	NNV RRN	-3	2	21	21	GUCUU AGUAC UCUG	GAAA	CAGAAUCUAC UAAGACAAGG CAAAAUGCCG UGUUUAUCUC GUCAACUUGU UGGCGAGAUU UUUUU
CdiCa s9	NNR HHH Y;NN RAA AY		2	22	22	ACUGG GGUUC AG	GAAA	CUGAACCUCA GUAAGCAUUG GCUCGUUUC AAUGUUGAUU GCUCCGCCGG UGCUCUUAU UUUUAAGGGC GCCGGC
CjeCa s9	NNN NRY AC	-3	2	21	23	GTTTT AGTCC CT	GAAA	AGGGACTAAA ATAAAGAGTT TGCGGGACTC TGCGGGGTTA CAATCCCCTA AAACCGCTTTT TT
GeoC as9	NNN NCR AA		2	21	23	GUCAU AGUUC CCCUG A	GAAA	UCAGGGUUAC UAUGAUAAGG GCUUUCUGCC UAAGGCAGAC

								UGACCCGCGG CGUUGGGGAU CGCCUGUCGC CCGCUUUUGG CGGGCAUUCC CCAUCCUU
iSpyM acCas 9	NAA N	-3	2	19	21	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
NmeC as9	NNN NGA YT;N NNN GYT T;NN NNG AYA; NNN NGT CT	-3	2	20	24	GTTGT AGCTC CCTTT CTCAT TTCG	GAAA	CGAAATGAGA ACCGTTGCTA CAATAAGGCC GTCTGAAAAG ATGTGCCGCA ACGCTCTGCC CCTTAAAGCTT CTGCTTTAAG GGGCATCGTT TA
ScaCa s9	NNG	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
ScaCa s9- HiFi- Sc++	NNG	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC

SpyCa s9- 3var- NRR H	NRR H	-3	2	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
SpyCa s9- 3var- NRT H	NRT H	-3	2	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
SpyCa s9- 3var- NRC H	NRC H	-3	2	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
SpyCa s9- HF1	NGG	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCa s9- QQR1	NAA G	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC

SpyCas9-SpG	NGN	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCas9-VQR	NGA N	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCas9-VRER	NGC G	-3	2	20	20	GTTTT AGAGC TA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTATC AACTTGAAAA AGTGGCACCG AGTCGGTGC
SpyCas9-xCas	NG;G AA;G AT	-3	2	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTA AGAGC TATGC TG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC
StlCas9-	NNA CAA	-3	2	20	20	GTCTT TGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG

CNRZ 1066								AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
StlCa s9- LMG1 831	NNG CAA	-3	2	20	20	GTCTT TGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
StlCa s9- MTH1 7CL3 96	NNA AAA	-3	2	20	20	GTCTT TGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
StlCa s9- TH14 77	NNG AAA	-3	2	20	20	GTCTT TGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT

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 РСТ 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).

24. 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'-GTTTGTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGGACCGAGTCGGTCC-3' (SEQ ID NO: 1591)) 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 одностороннего разрыва в первой нити (например, непосредственно с 3'-стороны от одностороннего разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от одностороннего 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 одностороннего разрыва в первой нити (например, непосредственно с 5'-стороны от одностороннего разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от одностороннего разрыва в первой нити).

Способы и композиции для модифицированной РНК (например, gRNA или РНК-матрицы)

В некоторых вариантах осуществления РНК-компонент системы (например, РНК-матрица или gRNA) 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'-О-метил (2'-O-Me), 2'-О-(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-матрицы (например, та их часть, которая связывает сайт-мишень) или направляющие РНК, описанные в данном документе, содержат любую из последовательностей, показанных в таблице 4 WO2018107028A1, включенной 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-матрицы (например, та их часть, которая связывается с сайтом-мишенью) или gRNA содержат 2'-О-метил- (2'-О-Me-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит 2'-О-(2-метоксиэтил-) (2'-О-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'-О-метил- (2'-О-Me-), 2'-О-(2-метоксиэтил-) (2'-О-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'-О-метил- (2'-О-Me-), 2'-О-(2-метоксиэтил-) (2'-О-MOE-) и/или 2'-фтор- (2'-F-) модифицированных нуклеотидов. Модификация конца может включать в себя фосфотиоатную (PS), 2'-О-метил- (2'-О-Me-), 2'-О-(2-метоксиэтил-) (2'-О-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'-О-Me-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'-ОМе, в положениях, где 2'-ОН 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 Kosak 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.

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 линейаризованной ДНК-матрицы, где UTP необязательно замещен 1-метилпсевдо-UTP. В некоторых вариантах осуществления транскрипт содержит 5'- и 3'-UTR, например, GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1603) и UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCC AGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1604), 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(Г)-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 одонитевых сегментов РНК друг с другом.

Одним из примеров блокатора концов, который можно применять в сочетании, например, с РНК-лигазой T4, является средство для дидезокси-терминации. РНК-лигаза T4, как правило, катализирует АТР-зависимое лигирование фосфодиэфирных связей между 5'-фосфатным и 3'-гидроксильным концами. В некоторых вариантах осуществления при применении РНК-лигазы T4 на концах, подвергаемых лигированию, должны присутствовать подходящие концы. Один из способов блокирования РНК-лигазы T4 на конце предусматривает отсутствие правильного формата конца. Обычно концы сегментов РНК с 5-гидроксильной или 3'-фосфатной группой не будут выступать в качестве субстратов для РНК-лигазы T4.

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-аминокси). В некоторых вариантах осуществления лигирование молекул РНК с использованием реакции клик-химии является предпочтительным, 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 Хьюстона - 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, карбонил/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 акридидом 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 T4 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.

25. Состав нуклеиновых кислот-матриц

В некоторых вариантах осуществления нуклеиновая кислота-матрица представляет собой РНК-матрицу. В некоторых вариантах осуществления РНК-матрица содержит один или несколько модифицированных нуклеотидов. Например, в некоторых вариантах осуществления РНК-матрица содержит один или несколько дезоксирибонуклеотидов. В некоторых вариантах осуществления области РНК-матрицы заменены нуклеотидами ДНК, например, для повышения стабильности молекулы. Например, 3'-конец матрицы может содержать нуклеотиды ДНК, тогда как остальная часть матрицы содержит нуклеотиды РНК, которые могут быть подвергнуты обратной транскрипции. Например, в некоторых вариантах осуществления последовательность гетерологичного объекта в основном или полностью состоит из нуклеотидов РНК (например, по меньшей мере 90%, 95%, 98% или 99% нуклеотидов РНК). В некоторых вариантах осуществления один или оба из 3'-UTR и 3'-концевого домена гомологии с мишенью в основном или полностью состоят из нуклеотидов ДНК (например, по меньшей мере 90%, 95%, 98% или 99% нуклеотидов ДНК). В других вариантах область матрицы для записи в геном может содержать нуклеотиды ДНК. В некоторых вариантах осуществления нуклеотиды ДНК в матрице копируются в геном с помощью домена, способного к ДНК-зависимой ДНК-полимеразной активности. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность обеспечивается ДНК-полимеразным доменом в полипептиде. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность обеспечивается доменом обратной транскриптазы, который также способен к ДНК-зависимой полимеризации ДНК, например, к синтезу второй нити. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность обеспечивается ДНК-полимеразой. В некоторых вариантах осуществления ДНК-зависимая ДНК-полимеразная активность, обеспечиваемая ДНК-полимеразным доменом в полипептиде, не способна обеспечивать активность обратной транскрипции. В некоторых вариантах осуществления молекула-матрица состоит только из нуклеотидов ДНК. В некоторых вариантах осуществления ДНК-матрица полимеризуется в геном с помощью ДНК-полимеразы. В некоторых вариантах осуществления матрица, которая состоит из нуклеотидов ДНК, содержит модифицированные нуклеотиды ДНК. В некоторых вариантах осуществления матрица, которая состоит из нуклеотидов ДНК, содержит модифицированный остов.

Нуклеотиды, составляющие матрицу системы Gene Writer™, могут быть природными или модифицированными основаниями или их комбинацией. Например, матрица может содержать псевдоуридин, дигидроуридин, инозин, 7-метилгуанозин или другие модифицированные основания. В некоторых вариантах осуществления матрица

может содержать нуклеотиды типа запертых нуклеиновых кислот. В некоторых вариантах осуществления модифицированные основания, используемые в матрице, не ингибируют обратную транскрипцию матрицы. В некоторых вариантах осуществления модифицированные основания, используемые в матрице, могут улучшать обратную транскрипцию, например, специфичность или точность.

26. 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.

27. *Полнотид Gene Writer*

28. *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_VM В. 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.

29. 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_VM 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.

30. 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 одностороннего 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 односторонних 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 односторонний разрыв в ДНК *in vitro*. В вариантах осуществления односторонний разрыв приводит к открытому основанию. В вариантах осуществления открытое 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 открытых оснований (например, обнаруживаемый 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 односторонний разрыв в ДНК в клетке. В вариантах осуществления эндонуклеазный домен способен вносить односторонний разрыв в ДНК в клетке HEK293T. В вариантах осуществления неисправленный односторонний разрыв, который подвергается репликации в отсутствие Rad51, приводит к увеличению частоты NHEJ в месте одностороннего разрыва, что может быть обнаружено, 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 с⁻¹ M⁻¹ *in vitro*. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью, составляющей более чем приблизительно 10^5 , 1×10^6 , 1×10^7 или 1×10^8 с⁻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 с⁻¹ M⁻¹. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью в клетках, составляющей более чем приблизительно 10^5 , 1×10^6 , 1×10^7 или 1×10^8 с⁻¹ M⁻¹.

31. 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 НЕК293Т), 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 обнаружения 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 НЕК293Т). В вариантах осуществления частоту специфического 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).

В некоторых вариантах осуществления домен обратной транскриптазы может содержать мутацию, например, как указано в таблице 45. В вариантах осуществления мутация приводит к модификации, например, увеличению, стабильности и функциональных свойств домена RT. В некоторых вариантах осуществления мутация приводит к модификации, например, увеличению, процессивности и аффинности к матрице домена RT. В некоторых вариантах осуществления мутантный домен RT может демонстрировать по меньшей мере 5-кратное, по меньшей мере 10-кратное, по меньшей мере 15-кратное, по меньшей мере 20-кратное, по меньшей мере 25-кратное, по меньшей мере 30-кратное, по меньшей мере 40-кратное, по меньшей мере 45-кратное, по меньшей мере 50-кратное, по меньшей мере 55-кратное, по меньшей мере 60-кратное, по меньшей мере 65-кратное, по меньшей мере 70-кратное, по меньшей мере 80-кратное, по меньшей мере 100-кратное увеличение процессивности по сравнению с немутантным доменом RT. В вариантах осуществления мутантный домен RT может демонстрировать по меньшей мере 5-кратное, по меньшей мере 10-кратное, по меньшей мере 15-кратное, по меньшей мере 20-кратное, по меньшей мере 25-кратное, по меньшей мере 30-кратное, по меньшей мере 40-кратное, по меньшей мере 45-кратное, по меньшей мере 50-кратное, по меньшей мере 55-кратное, по меньшей мере 60-кратное, по меньшей мере 65-кратное, по меньшей мере 70-кратное, по меньшей мере 80-кратное, по меньшей мере 100-кратное увеличение аффинности к матрице по сравнению с немутантным доменом RT. В некоторых вариантах осуществления мутантный домен RT может содержать одну или несколько мутаций, выбранных из D200N/T330P/L603W, T306K, W313F, L139P, E607K.

В таблице 45 раскрыты мутации, улучшающие свойства различных обратных транскриптаз. Ключевые мутации, которые, как ожидается, будут наиболее действенными, были применены в группах ретровирусов. Сохранение последовательности в группе вирусов при одной из этих ключевых мутаций приводило к внедрению мутации в этой группе (см. пример 33, фигуры 36 А и В). Положения в последовательности относятся к положениям в RT MMLV. В некоторых вариантах осуществления домен RT, описанный в данном документе, содержит мутацию, описанную в таблице 45.

Таблица 45. Перечень иллюстративных мутаций домена RT

Группа	L139	D200	T306	W313	T330	L603	E607
Гамма		D200N	T306K	W313F	T330P	L603W	
Эпсилон		D200N	T306K	W313F	T330P	L603W	
Дельта	L139P	D200N	X	X	T330P	L603W*	X
Бета	L139P	X	X	X	T330P	X	X
Спума		D200N	T306K	X	T330P	L603W	

32. Продукты слияния Cas-RT

В некоторых вариантах осуществления полипептид Gene Writer содержит домен RT, слитый с молекулой Cas. В некоторых вариантах осуществления молекула Cas представляет собой DBD и/или эндонуклеазный домен полипептида Gene Writer. В некоторых вариантах осуществления домен RT содержит Cas9. В некоторых вариантах осуществления Cas9 может содержать мутацию, например, раскрытую в таблице 40A. В таблице 46 раскрыт перечень иллюстративных слитых белков Cas-RT.

В некоторых вариантах осуществления молекула Cas в полипептиде Gene Writer обладает активностью, аналогичной активности в остальном аналогичной молекулы Cas, которая не слита с доменом RT. В некоторых вариантах осуществления активность составляет по меньшей мере 40%, 50%, 60%, 70%, 80% или 90% от активности в остальном аналогичной молекулы Cas. В некоторых вариантах осуществления молекула Cas в полипептиде Gene Writer может характеризоваться активностью образования вставок/делеций, составляющей по меньшей мере 40%, 50%, 60%, 70%, 80% или 90% от активности в остальном аналогичной молекулы Cas, которая не слита с доменом RT, например, в анализе согласно примеру 32.

В некоторых вариантах осуществления полипептид Gene Writer содержит аминокислотную последовательность в соответствии с таблицей 46 ниже или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней. В некоторых вариантах осуществления нуклеиновая кислота, кодирующая полипептид Gene Writer, содержит последовательность нуклеиновой кислоты в соответствии с таблицей 47 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

Таблица 46. Перечень иллюстративных полипептидов Gene Writer, содержащих продукты слияния Cas-RT

Название	Домен Cas	Линкер	Ретровирус-источник RT	Полипептид-источник RT	Последовательность полипептида Gene Writer
Cas-RT(MMLV)	Cas9 (N863A)	SGGSS GGSSG SETPG TSESA TPESS GGSSG GSS	Вирус лейкоза мышей Молони (MMLV или MLVMS)	P03355	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKF KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQ EIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTF RIPPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIAMARENQTTQKGQKNSRERMKRIIEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDN KVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA

				ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLV VAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHL FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQL GGDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSTLNIEDEYRLHETSKEP DVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS IKQYPMSQ EARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPEMGISGQLTWTRL PQGFKN SPTL FNEALHRDLADFRIQH PDLILLQY VDDL LLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG Y LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMA APLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDE KQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTK DAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLD TDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTW YTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLK
--	--	--	--	---

					ALFLPKRLSIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIEN SSPSGGSKRTADGSEFEPKKKRKV
Cas- RT(PERV)	Cas9 (N863A)	SGSS GGSSG SETPG TSESA TPESS GGSSG GSS	Эндоген ный ретровир ус свиней (PERV)	Q4VFZ2	MAPKKKRKVGIHGVPAAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFR IPYYVGPLARGNSRF AWMTRKSEETITPWNFEVVDKGASAQSFIERMTNF DKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK IHKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSL TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMG RHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN KVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSCLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKL ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG

					EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKK LKSVKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHL FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQL GGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSLDDEYRLYSPLVKPDQ NIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEA QEGIRPHVQRLIQQGILVPVQSPWNTPLLVRKPGTNDYRPVQDLREVNKR VQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPGTGRTGQLTWTRL PQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYV DDL LAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSL RDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPL YPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDAD KLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPA ALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSS YVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGK SINIYTDSRYAFATAHVHGA IYKQRGWLT SAGREIKNKEEILSLLEALHLPK RLAIIHCPGHQKAKDPISRG NQMADRVAKQAAQGVNLLP
Cas-RT(MLV BM)	Cas9 (N863A)	SGGSS GGSSG SETPG	Вирус лейкоза мышей	Q7SVK7	MAPKKKRKVG IHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

		TSESA TPESS GGSSG GSS	(MLVBM)	LRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF DKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLK IHKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMG RHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN KVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKL ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKK LKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRRDKPIREQAENIIHL
--	--	--------------------------------	-------------	---

				FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQL GGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSSLGIEDEYRLHETSTEPD VSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIIQQYPM SHE ARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQH PDLILLQY VDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYL LREGQRWLTEARKETVMGQPVPKTPRQLREFLGKAGFCRLFIPGFAEMAA PLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEK QGYAKGVLTQKLGWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKD AGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQ FGPVVALNPATLLPLPEEGAPHDCLEILAE THGTRPDLTDQPIPDADHTWYT DGSSFLQEGQRKAGAAVTTE TEVIWAGALPAGTSAQRAELIALTQALKMA EGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALF LPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
Cas-RT(MMTVB)	Cas9 (N863A)	SGGSS GGSSG SETPG TSESA TPESS GGSSG GSS	Вирус опухоли молочно й железы мышей (MMTVB)	P03365 MAPKKKRKVGIIHGVPAAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRILIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN

REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFR
 IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF
 DKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
 VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
 IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL
 TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMG
 RHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
 QLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
 KVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA
 ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
 VITLKSKLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVG TALIKKYPKL
 ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS
 KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
 LKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENG
 RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
 EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHL
 FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQL
 GGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSVQEISDSRPM LHIYLN
 RRFLGLLDTGADKTCIAGRDPANWPIHQTESSLQGLGMACGVARSSQPL
 RWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIK VRLMTD SPDDSQDL MIGAI
 ESNLFADQISWKS DQPVWLNQWPLKQEK LQALQQLVTEQLQLGHLEESNS

					PWNTPVFVIKKKSGKWRLQLDLRAVNATMHDMGALQPGLPSPVAVPKG WEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKN SPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQA LNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRRTLNDFAQ KLLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNE RLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITP YDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGF LGEVHFHLPKDPLLTFTLQTAIFPHMTSTTPLEKGIVFTDGSANGRSVTYI QGREPIIKENTQNTAQQAIEIVAVITAFEEVSQPFNL YTDSKYVTGLFPEIETA T LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLT RILTA
Cas- RT(MPM V)	Cas9 (N863A)	SGGSS GGSSG SETPG TSESA TPESS GGSSG GSS	Вирус обезьян Мейзона- Пфайзер а (MPMV)	P07572	MAPKKRKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLV DSTKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF DKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK

				IHKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQL KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMG RHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN KVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKK LKS VKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHL FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQL GGDSGGSSGGSSGSETPGTSESATP ESSGGSSGSSSTA AIDILAPQQCAEPIT WKSDEPVWVDQWPLTNDKLAAAQQLVQE QLEAGHITESSSPWNTPIFVIK KKSGKWRL LQDLRAVNATMVL MGALQPGLPSPVAIPQGYLKI IIDLKDCFF SIPLHPSDQKRFAFSLPSTNFK EPMQRFQWKVLPQGMANSPTLCQKYVATA IHKVRHAWKQMYIIHYMDDILIAGKD GQQVLQCFDQLKQELTAAGLHIAP EKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWL RP YLKLTTGDLKPLFDTLKGSDPNSHRSLSKEALASLEKVETAIAEQFVTHIN YSLPLIFLIFNTALTPTGLFWQDNPMWIHL PASPKKVLLPYYDAIADLIILG
--	--	--	--	---

					RDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYP PN KLIQFCKLHTFVFPQIISK TPLN NALLVFTDGSSTGMAAYTLTDTTIKFQ TNL NSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISE TAK LFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLTKIVASNIN TN
--	--	--	--	--	---

Таблица 47. Иллюстративные последовательности мРНК, кодирующие полипептид
Gene Writer

Название	мРНК (от 5' к 3')	Хвост
Cas9-RT(MMLV)	AGGAAAUAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAG AGCCACCAUGAAACGGACAGCCGACGGAAGCGAGUUCGAG UCACCAAAGAAGAAGCGGAAAGUCGACAAGAAGUACAGCA UCGGCCUGGACAUCGGCACCAACUCUGUGGGCUGGGCCGUG AUCACCGACGAGUACAAGGUGCCCAGCAAGAAAUUCAAGG UGCUGGGCAACACCGACCGGCACAGCAUCAAGAAGAACCUG AUCGGAGCCCUGCUGUUCGACAGCGGCGAAACAGCCGAGGC CACCCGGCUGAAGAGAACCGCCAGAAGAAGAUACACCAGAC GGAAGAACCGGAUCUGCUAUCUGCAAGAGAUUCUUCAGCAA CGAGAUGGCCAAGGUGGACGACAGCUUCUCCACAGACUG GAAGAGUCCUCCUGGUGGAAGAGGAUAAGAAGCACGAGC GGCACCCCAUCUUCGGCAACAUCGUGGACGAGGUGGCCUAC CACGAGAAGUACCCACCAUCUACCACCUGAGAAAGAAACU GGUGGACAGCACCGACAAGGCCGACCUGCGGCUGAUCUAUC UGGCCCUGGCCCACAUGAUCUAGUUCGGGGCCACUUCUG AUCGAGGGCGACCUGAACCCCGACAACAGCGACGUGGACAA GCUGUUCAUCCAGCUGGUGCAGACCUACAACCAGCUGUUCG AGGAAAACCCCAUCAACGCCAGCGGCGUGGACGCCAAGGCC AUCCUGUCUGCCAGACUGAGCAAGAGCAGACGGCUGGAAA AUCUGAUCGCCCAGCUGCCCGGCGAGAAGAAGAAUGGCCU GUUCGGAAACCUGAUUGCCCUGAGCCUGGGCCUGACCCCA ACUUCAAGAGCAACUUCGACCUGGCCGAGGAUGCCAAACU GCAGCUGAGCAAGGACACCUACGACGACGACCUGGACAACC UGCUGGCCCAGAUCGGCGACCAGUACGCCGACCUGUUUCUG GCCGCCAAGAACCUGUCCGACGCCAUCCUGCUGAGCGACAU CCUGAGAGUGAACACCGAGAUACCAAGGCCCCCUGAGCG CCUCUAUGAUCUAGAGAUACGACGAGCACCACCAGGACCUG ACCCUGCUGAAAGCUCUCGUGCGGCAGCAGCUGCCUGAGAA GUACAAAGAGAUUUUCUUCGACCAGAGCAAGAACGGCUAC GCCGGCUACAUCGACGGCGGAGCCAGCCAGGAAGAGUUCU ACAAGUUCAUCAAGCCCAUCCUGGAAAAGAUGGACGGCAC	(A)80

CGAGGAACUGCUCGUGAAGCUGAACAGAGAGGACCUGCUG
 CGGAAGCAGCGGACCUUCGACAACGGCAGCAUCCCCACCA
 GAUCCACCUGGGAGAGCUGCACGCCAUUCUGCGGCGGCAGG
 AAGAUUUUUACCCAUUCCUGAAGGACAACCGGGAAAAGAU
 CGAGAAGAUCCUGACCUUCCGCAUCCCCUACUACGUGGGCC
 CUCUGGCCAGGGGAAACAGCAGAUUCGCCUGGAUGACCAG
 AAAGAGCGAGGAAACCAUCACCCCCUGGAACUUCGAGGAA
 GUGGUGGACAAGGGCGCUUCCGCCCAGAGCUUCAUCGAGC
 GGAUGACCAACUUCGAUAAGAACCUGCCCAACGAGAAGGU
 GCUGCCCAAGCACAGCCUGCUGUACGAGUACUUCACCGUGU
 AUAACGAGCUGACCAAAGUGAAAUACGUGACCGAGGGAAU
 GAGAAAGCCCGCCUUCUGAGCGGCGAGCAGAAAAAGGCC
 AUCGUGGACCUGCUGUUCAAGACCAACCGGAAAGUGACCG
 UGAAGCAGCUGAAAGAGGACUACUUCAAGAAAAUCGAGUG
 CUUCGACUCCGUGGAAAUCUCCGGCGUGGAAGAUCGGUUC
 AACGCCUCCUGGGCACAUACCACGAUCUGCUGAAAAUUAU
 CAAGGACAAGGACUUCUGGACAAUGAGGAAAACGAGGAC
 AUUCUGGAAGAUUUCGUGCUGACCCUGACACUGUUUGAGG
 ACAGAGAGAUGAUCGAGGAACGGCUGAAAACCUAUGCCCA
 CCUGUUCGACGACAAAGUGAUGAAGCAGCUGAAGCGGCGG
 AGAUACACCGGCUGGGGCAGGCUGAGCCGGAAGCUGAUCA
 ACGGCAUCCGGGACAAGCAGUCCGGCAAGACAAUCCUGGA
 UUUCUGAAGUCCGACGGCUUCGCCAACAGAAACUUCAUGC
 AGCUGAUCCACGACGACAGCCUGACCUUUAAGAGGACAU
 CCAGAAAGCCCAGGUGUCCGGCCAGGGCGAUAGCCUGCACG
 AGCACAUUGCCAAUCUGGCGGCAGCCCCGCCAUUAAGAAG
 GGCAUCCUGCAGACAGUGAAGGUGGUGGACGAGCUCGUGA
 AAGUGAUGGGCCGGCACAAAGCCCGAGAACAUUCGUGAUCGA
 AAUGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAG
 AACAGCCGCGAGAGAAUGAAGCGGAUCGAAGAGGGCAUCA
 AAGAGCUGGGCAGCCAGAUCUGAAAGAACACCCCGUGGA
 AAACACCCAGCUGCAGAACGAGAAGCUGUACCUGUACUACC
 UGCAGAAUGGGCGGGGAUAUGUACGUGGACCAGGAACUGGA
 CAUCAACCGGCUGUCCGACUACGAUGUGGACCAUAUCGUGC
 CUCAGAGCUUUCUGAAGGACGACUCCAUCGACAACAAGGU

GCUGACCAGAAGCGACAAGGCCCGGGCAAGAGCGACAAC
 GUGCCCUCCGAAGAGGUCGUGAAGAAGAUGAAGAACUACU
 GGCGGCAGCUGCUGAACGCCAAGCUGAUUACCCAGAGAAA
 GUUCGACAAUCUGACCAAGGCCGAGAGAGGCGGCCUGAGC
 GAACUGGAUAAGGCCGGCUUCAUCAAGAGACAGCUGGUGG
 AAACCCGGCAGAUACAAAAGCACGUGGCACAGAUCCUGGAC
 UCCCGGAUGAACACUAAGUACGACGAGAAUGACAAGCUGA
 UCCGGGAAGUGAAAGUGAUCACCCUGAAGUCCAAGCUGGU
 GUCCGAUUUCCGGAAGGAUUUCCAGUUUUACAAAGUGCGC
 GAGAUCAACAACUACCACCACGCCCACGACGCCUACCUGAA
 CGCCGUCGUGGGAACCGCCUGAUCAAAAAGUACCCUAAGC
 UGGAAAGCGAGUUCGUGUACGGCGACUACAAGGUGUACGA
 CGUGCGGAAGAUGAUCGCCAAGAGCGAGCAGGAAAUCGGC
 AAGGCUACCGCCAAGUACUUCUUCUACAGCAACAUCAUGA
 ACUUUUUCAAGACCGAGAUUACCCUGGCCAACGGCGAGAU
 CCGGAAGCGGCCUCUGAUCGAGACAAACGGCGAAACCGGG
 GAGAUUCGUGUGGGAUAAGGGCCGGGAUUUUGCCACCGUGC
 GGAAAGUGCUGAGCAUGCCCCAAGUGAAUAUCGUGAAAAA
 GACCGAGGUGCAGACAGGCGGCUUCAGCAAAGAGUCUAUC
 CUGCCCAAGAGGAACAGCGAUAAGCUGAUCGCCAGAAAGA
 AGGACUGGGACCCUAAGAAGUACGGCGGCUUCGACAGCCCC
 ACCGUGGCCUAUUCUGUGCUGGUGGUGGCCAAAGUGGAAA
 AGGGCAAGUCCAAGAAACUGAAGAGUGUGAAAGAGCUGCU
 GGGGAUCACCAUCAUGGAAAGAAGCAGCUUCGAGAAGAAU
 CCAUCGACUUUCUGGAAGCCAAGGGCUACAAAGAAGUGA
 AAAAGGACCUGAUCAUCAAGCUGCCUAAGUACUCCCUGUU
 CGAGCUGGAAAACGGCCGGAAGAGAAUGCUGGCCUCUGCC
 GGCGAACUGCAGAAGGGAAACGAACUGGCCCUGCCCUCCAA
 AUAUGUGAACUUCUGUACCUGGCCAGCCACUAUGAGAAG
 CUGAAGGGCUCCCCCGAGGAUAAUGAGCAGAAACAGCUGU
 UUGUGGAACAGCACAAAGCACUACCUGGACGAGAUCAUCGA
 GCAGAUCAAGCAGUUCUCCAAGAGAGUGAUCUGGCCGAC
 GCUAAUCUGGACAAAGUGCUGUCCGCCUACAACAAGCACCG
 GGAUAAGCCCAUCAGAGAGCAGGCCGAGAAUAUCAUCCAC
 CUGUUUACCCUGACCAAUUCUGGGAGCCCCUGCCGCCUCAA

GUACUUUGACACCACCAUCGACCGGAAGAGGUACACCAGCA
 CCAAAGAGGUGCUGGACGCCACCCUGAUCCACCAGAGCAUC
 ACCGGCCUGUACGAGACACGGAUCGACCUGUCUCAGCUGGG
 AGGUGACUCUGGAGGAUCUAGCGGAGGAUCCUCUGGCAGC
 GAGACACCAGGAACAAGCGAGUCAGCAACACCAGAGAGCA
 GUGGCGGCAGCAGCGGCGGCAGCAGCACCCUAAAUAUAGA
 AGAUGAGUAUCGGCUACAUGAGACCUCAAAAGAGCCAGAU
 GUUUCUCUAGGGUCCACAUGGCUGUCUGAUUUUCCUCAGG
 CCUGGGCGGAAACCGGGGGCAUGGGACUGGCAGUUCGCCA
 AGCUCCUCUGAUCAUACCUCUGAAAGCAACCUCUACCCCCG
 UGUCCAUA AAAACAAUACCCCAUGUCACAAGAAGCCAGACU
 GGGGAUCAAGCCCCACAUAACAGAGACUGUUGGACCAGGGA
 AUACUGGUACCCUGCCAGUCCCCCUGGAACACGCCCCUGCU
 ACCCGUUAAGAAACCAGGGACUAAUGAUUAUAGGCCUGUC
 CAGGAUCUGAGAGAAGUCAACAAGCGGGUGGAAGACAUCC
 ACCCCACCGUGCCCAACCCUUAACAACCUCUUGAGCGGGCUC
 CCACCGUCCCACCAGUGGUACACUGUGCUUGAUUUAAAGG
 AUGCCUUUUUCUGCCUGAGACUCCACCCACCAGUCAGCCU
 CUCUUCGCCUUUGAGUGGAGAGAUCCAGAGAUGGGAAUCU
 CAGGACAAUUGACCUGGACCAGACUCCCACAGGGUUUCAA
 AAACAGUCCCACCCUGUUUAAUGAGGCACUGCACAGAGACC
 UAGCAGACUUCCGGAUCCAGCACCCAGACUUGAUCCUGCUA
 CAGUACGUGGAUGACUUACUGCUGGCCGCCACUUCUGAGC
 UAGACUGCCAACAAGGUACUCGGGGCCUGUUACAAACCCUA
 GGGAACCUCGGGUUAUCGGGCCUCGGCCAAGAAAGCCCAA
 UUUGCCAGAAACAGGUCAAGUAUCUGGGGUUAUCUUCUAAA
 AGAGGGUCAGAGAUGGCUGACUGAGGCCAGAAAAGAGACU
 GUGAUGGGGCAGCCUACUCCGAAGACCCUCGACAACUAAG
 GGAGUUCCUAGGGAAGGCAGGCUUCUGUCGCCUCUUAUC
 CCUGGGUUUGCAGAAAUGGCAGCCCCCUGUACCCUCUCAC
 CAAACCGGGGACUCUGUUUAAUUGGGGGCCAGACCAACAA
 AAGGCCUAUCAAGAAAUCAAGCAAGCUCUUCUAACUGCCCC
 AGCCCUGGGGUUGCCAGAUUUGACUAAGCCCUUUGAACUC
 UUUGUCGACGAGAAGCAGGGCUACGCCAAAGGUGUCCUAA
 CGCAAAAACUGGGACCUUGGCGUCGGCCGGUGGCCUACCUG

UCCAAAAAGCUAGACCCAGUAGCAGCUGGGUGGCCCCCUUG
CCUACGGAUGGUAGCAGCCAUUGCCGUACUGACAAAGGAU
GCAGGCAAGCUAACCAUGGGACAGCCACUAGUCAUUCUGG
CCCCCAUGCAGUAGAGGCACUAGUCAACAACCCCCCGAC
CGCUGGCUUCCAACGCCCCGGAUGACUCACUAUCAGGCCU
GCUUUUGGACACGGACCGGGUCCAGUUCGGACCGGUGGUA
GCCCUGAACCCGGCUACGCUGCUCACUGCCUGAGGAAGG
GCUGCAACACAACUGCCUUGAUUAUCCUGGCCGAAGCCCACG
GAACCCGACCCGACCUAACGGACCAGCCGCUCCCAGACGCC
GACCACACCUGGUACACGGAUGGAAGCAGUCUCUUACAAG
AGGGACAGCGUAAGGCGGGAGCUGCGGUGACCACCGAGAC
CGAGGUAAUCUGGGCUAAAGCCCUGCCAGCCGGGACAUCCG
CUCAGCGGGCUGAACUGAUAGCACUCACCCAGGCCCUAAAG
AUGGCAGAAGGUAAGAAGCUAAAUGUUUAUACUGAUAGCC
GUUAUGCUUUUGCUACUGCCCAUAUCCAUGGAGAAUAUA
CAGAAGGCGUGGGUGGCUCACAUCAGAAGGCAAAGAGAU
AAAAAUAAAGACGAGAUUCUUGGCCCUACUAAAAGCCCUCU
UUCUGCCCAAAAGACUUAAGCAUAAUCCAUGGUCCAGGACA
UCAAAAGGGACACAGCGCCGAGGCUAGAGGCAACCGGAUG
GCUGACCAAGCGGCCCCGAAAGGCAGCCAUCACAGAGACUCC
AGACACCUCUACCCUCCUCAUAGAAAAUUCAUACCCUCUG
GCGGCUCAAAAAGAACCGCCGACGGCAGCGAAUUCGAGCCC
AAGAAGAAGAGGAAAGUCUGAUUAAUUAAGCUGCCUUCUG
CGGGGCUUGCCUUCUGGCCAUGCCCUUCUUCUCUCCCUUGC
ACCUGUACCUCUUGGUCUUUGAAUAAAGCCUGAGUAGGAA
GUCUAG

Cas9- RT(PERV)	AGGAAAU AAGAGAGAAAAGAAGAGUAAGAAGAAAUUAAG AGCCACCAUGGCUCCCAAAAAGAAAAGGAAGGUGGGCAUU CACGGCGUGCCUGCGGCCGACAAAAAGUACAGCAUCGGCCU UGAUUAUCGGCACCAAUAGCGUGGGCUGGGCCGUUAUCACA GACGAAUACAAGGUACCCAGCAAGAAGUUCAAGGUGCUGG GGAAUACAGACAGGCACUCUAUCAAGAAAAACCUUAUCGG GGCUCUGCUGUUUGACUCAGGCGAGACCGCCGAGGCCACCA GGUUGAAGAGGACCGCAAGGCGAAGGUACACCCGGAGGAA GAACAGGAUCUGCUAUCUGCAGGAGAUUCUUCAGCAACGAG AUGGCCAAGGUGGACGACAGCUUCUCCACAGGCUGGAGG AGAGCUUCCUUGUCGAGGAGGAUAAGAAGCACGAACGACA CCCCAUCUUCGGCAACAUAUGUCGACGAGGUCGCUUAUCACG AGAAGUACCCACCAUCUACCACCUGCGAAAGAAAUUGGU GGAUAGCACCGAUAAAGCCGACUUGCGACUUAUCUACUUG GCUCUGGGCGACAUGAUUAAGUUCAGGGGCCACUUCCUGA UCGAGGGCGACCUUAACCCCGACAACAGUGACGUAGACAA AUUGUUCAUCCAGCUUGUACAGACCUAUAACCAGCUGUUC GAGGAAAACCCUAUUAACGCCAGCGGGGUGGAUGCGAAGG CCAUAUCUAGCGCCAGGCUGAGCAAAAGCAGGCGCUUGGA GAACCUGAUAGCCCAGCUGCCCGGUGAAAAGAAGAACGGC CUCUUCGGUAAUCUGAUUGCCCUGAGCCUGGGCCUGACCCC CAACUUCAAGAGCAACUUCGACCUGGCAGAAGAUGCCAAG CUGCAGUUGAGUAAGGACACCUAUGACGACGACUUGGACA AUCUGCUCGCCCAAUCGGCGACCAGUACGCUGACCUGUUC CUCGCCGCCAAGAACCUUUCUGACGCAAUCCUGCUUAGCGA UAUCCUUAAGGGUGAACACAGAGAUACCAAGGCCCCCUGA GCGCCAGCAUGAUCAAGAGGUACGACGAGCACCAUCAGGA CCUGACCCUUCUGAAGGCCCUGGUGAGGCAGCAACUGCCCG AGAAGUACAAGGAGAUUUUUUCGACCAGAGCAAGAACGG CUACGCCGGCUACAUCGACGGCGGAGCCAGCCAAGAGGAGU UCUACAAGUUCAUCAAGCCCAUCCUGGAGAAGAUGGAUGG CACCGAGGAGCUGCUGGUGAAGCUGAACAGGGAAGAUUUG CUCCGGAAGCAGAGGACCUUUGACAACGGUAGCAUCCCCA CCAGAUCCACCUGGGCGAGCUGCACGCAAUACUGAGGCGAC AGGAGGAUUUCUACCCCUUCCUCAAGGACAAUAGGGAGAA	(A)80
--------------------------------	---	--------------

AAUCGAAAAGAUUCUGACCUUCAGGAUCCCCUACUACGUG
GGCCCUUCUUGCCAGGGGCAACAGCCGAUUCGCUUGGAUGAC
AAGAAAGAGCGAGGAGACCAUCACCCCCUGGAACUUCGAG
GAAGUGGUGGACAAAGGAGCAAGCGCGCAGUCUUUCAUCG
AACGGAUGACCAAUUUCGACAAAAACCUGCCUAACGAGAA
GGUGCUGCCCAAGCACAGCCUGCUUUACGAGUACUUCACCG
UGUACAACGAGCUCACCAAGGUGAAAU AUGUGACCGAGGG
CAUGCGAAAACCCGCUUUCUGAGCGGCGAGCAGAAGAAG
GCCAUCGUGGACCUGCUGUUAAGACCAACAGGAAGGUGA
CCGUGAAGCAGCUGAAGGAGGACUACUUCAAGAAGAUCGA
GUGC UUUGAUAGCGUGGAAAU AAGCGGCGUGGAGGACAGG
UUCAACGCCAGCCUGGGCACCUACCACGACUUGUUGAAGAU
AAUCA AAGACAAGGAUUUCCUGGAUAAUGAGGAGAACGAG
GAUAUACUCGAGGACAUCGUGCUGACUUUGACCCUGUUUG
AGGACCGAGAGAUGAUUGAAGAAAGGCUCAAAACCUACGC
CCACCUGUUCGACGACAAAGUGAUGAAACAACUGAAGAGA
CGAAGAUACACCGGCUGGGGCAGACUGUCCAGGAAGCUC A
UCAACGGCAUUAGGGACAAGCAGAGCGGCAAGACCAUCCU
GGAUUUCCUGAAGUCCGACGGCUUCGCCAACCGAAACUUCA
UGCAGCUGAUUCACGAUGACAGCUUGACCUUCAAGGAGGA
CAUCCAGAAGGCCAGGUUAGCGGCCAGGGCGACUCCUGC
ACGAACAUAUUGCAAACCUGGCAGGCUCCCCUGCGAUCAAG
AAGGGCAUACUCGACACCGUUAAGGUUGUGGACGAAUUGG
UCAAGGUCAUGGGCAGGCACAAGCCCGAAAACAUAGUU AU
AGAGAUGGCCAGAGAGAACCAGACCACCCAAAAGGGCCAG
AAGAACAGCCGGGAGCGCAUGAAAAGGAUCGAGGAGGGUA
UCAAGGAACUCGGAAGCCAGAUCCUCA AAGAGCACCCCGUG
GAGAAUACCCAGCUCCAGAACGAGAAGCUGUACCUGUACU
ACCUGCAGAACGGCAGGGACAUGUACGUUGACCAGGAGUU
GGACAUCAACAGGCUUUCAGACUAUGACGUGGAUCACAUA
GUGCCCCAGAGCUUUCUUA AAGACGAUAGCAUCGACAACA
AGGUCCUGACCCGCUCCGACAAAGCCAGGGGCAAAAGCGAC
AACGUGCCAAGCGAAGAGGUGGUUAAAAAGAUGAAGAACU
ACUGGAGGCAACUGCUCAACGCGAAAUUGAUCACCCAGAG
AAAGUUCGAUAACCUGACCAAGGCCGAGAGGGGCGGACUC

UCCGAACUUGACAAAGCGGGCUUCAUAAAGAGGCAGCUGG
 UCGAGACCCGACAGAUACGAAGCACGUGGCCCAAUCCUC
 GACAGCAGAAUGAAUACCAAGUACGAUGAGAAUGACAAAC
 UCAUCAGGGAAGUGAAAGUGAUUACCCUGAAGAGCAAGUU
 GGUGUCCGACUUUCGCAAAGAUUUCAGUUCUACAAGGUG
 AGGGAGAUCAACAACUACCACCAUGCCCACGACGCAUACCU
 GAACGCCGUGGUCGGCACCGCCCUGAUUAAGAAGUAUCCA
 AAGCUGGAGUCCGAAUUUGUCUACGGCGACUACAAAGUUU
 ACGAUGUGAGGAAGAUGAUCGCUAAGAGCGAACAGGAGAU
 CGGCAAGGCCACCGCUAAGUAUUUCUUCUACAGCAACAUCA
 UGAACUUUUUCAAGACCGAGAUACACUUGCCAACGGCGA
 AAUCAGGAAGAGGCCGCUUAUCGAGACCAACGGUGAGACC
 GGCGAGAUUCGUGUGGGACAAGGGCAGGGACUUCGCCACCG
 UGAGGAAAGUCCUGAGCAUGCCCCAGGUGAAUAUUGUGAA
 AAAAACUGAGGUGCAGACAGGCGGCUUUAGCAAGGAAUCC
 AUCCUGCCCAAGAGGAACAGCGACAAGCUGAUCGCCCGGAA
 GAAGGACUGGGACCCUAAGAAGUAUGGAGGCUUCGACAGC
 CCCACCGUAGCCUACAGCGUGCUGGUGGUCGCGAAGGUAG
 AGAAGGGGAAGAGCAAGAAACUGAAGAGCGUGAAGGAGCU
 GCUCGGCAUAACCAUCAUGGAGAGGUCCAGCUUUGAGAAG
 AACCCCAUUGACUUUUUGGAAGCCAAGGGCUACAAAGAGG
 UCAAAAAGGACCUGAUCAUCAAAACUCCCCAAGUACUCCCUG
 UUUGAAUUGGAGAACGGCAGAAAGAGGAUGCUGGCGAGCG
 CUGGGGAACUGCAAAAGGGCAACGAACUGGCGCUGCCCAG
 CAAGUACGUGAAUUUUCUGUACCUGGCGUCCCACUACGAA
 AAGCUGAAAGGCAGCCCCGAGGACAACGAGCAGAAGCAGC
 UGUUCGUGGAGCAGCACAAGCAUUACCUGGACGAGAUAAU
 CGAGCAAUCAGCGAGUUCAGCAAGAGGGUGAUUCUGGCC
 GACGCGAACCUGGAUAAGGUCCUCAGCGCCUACAACAAGCA
 CCGAGACAAACCCAUCAGGGAGCAGGCCGAGAAUAUCAUA
 CACCUGUUCACCCUGACAAAUCUGGGCGCACCUGCGGCAUU
 CAAAUACUUCGAUACCACCAUCGACAGGAAAAGGUACACU
 AGCACUAAGGAGGUGCUGGAUGCCACCUUGAUCCACCAGU
 CCAUUACCGGCCUGUAUGAGACCAGGAUCGACCUGAGCCAG
 CUUGGAGGCGACUCUGGAGGAUCUAGCGGAGGAUCCUCUG

GCAGCGAGACACCAGGAACAAGCGAGUCAGCAACACCAGA
 GAGCAGUGGCGGCAGCAGCGGCGGCAGCAGCCUGGACGAC
 GAGUACAGACUGUAUAGCCCUCUGGUGAAGCCAGAUCA
 ACAUUCAGUUCUGGCUGGAACAGUUUCCACAGGCCUGGGC
 CGAAACAGCCGGAUUGGGCCUGGCCAAGCAGGUGCCUCCUC
 AGGUGAUCCAGCUGAAGGCCAGCGCCACACCUGUGUCCGUG
 CGGCAGUACCCUCUGUCCAAGGAGGCUCAGGAGGGCAUCA
 GACCUCACGUCCAGCGGCUGAUCCAGCAGGGGAUCCUGGUG
 CCCGUGCAAAGCCCUUGGAACACCCCUUUCUGCCCGUGAG
 AAAACCCGGCACAAACGACUACCGGCCUGUGCAGGACCUGA
 GAGAAGUGAACAAGCGGGUGCAGGACAUCCACCCACAGU
 GCCAAAUCCUUACAACCUGCUUUGUGCCCUGCCCCCCCAGC
 GCAGCUGGUACACCGUUCUGGACCUGAAAGAUGCCUUUUU
 CUGUCUGAGACUUAUCCUACAAGCCAGCCCCUGUUCGCCU
 UCGAGUGGCGGGAUCCUGGCACCGGCCGGACAGGCCAGCUG
 ACAUGGACCAGACUGCCUCAGGGCUUCAAGAACAGCCCUAC
 CAUCUUCAACGAGGCCCUGCACAGAGACCUUGCCAACUUCA
 GAAUCCAACACCCACAGGUGACCCUGCUCCAGUACGUGGAU
 GACCUGCUGCUGGCCGGCGCCACAAAACAAGAUUGCCUGGA
 AGGCACCAAGGCCCUUCUGCUGGAGCUGAGCGACCUGGGA
 UAUCGGGGCCUCUGCUAAGAAAGCUCAGAUCUGCAGGAGAG
 AGGUGACCUACCUGGGCUACUCUCUGAGAGAUGGCCAAAG
 AUGGCUGACCGAGGCCAGAAAGAAAACCGUGGUGCAAUC
 CCCGCUCCUACAACAGCCAAGCAGGUUAGAGAGUUCCUGGG
 AAAGGCUGGAUUUUGCAGACUGUUAUCCAGGCCUUUGCC
 ACCCUGGCGCGCCCUUGUACCCCUAGACCAAACCUAAGGG
 CGAGUUCAGCUGGGCCCCAGAGCACCAGAAGGCAUUCGACG
 CGAUCAAGAAGGCUCUGCUGUCUGCCCCUGCCCUGGCUCUG
 CCCGACGUGACAAAGCCCUUACCCUGUACGUGGACGAACG
 GAAGGGCGUGGCUAGAGGGCGUUCUGACCCAGACCCUGGGU
 CCUUGGAGAAGGCCUGUGGCCUACCUCAGUAAGAAGCUGG
 AUCCUGUGGCCUCUGGCUGGCCUGUGUGCCUGAAGGCCAUC
 GCCGCCGUGGCCAUUCUGGUCAAGGAUGCCGAUAAGCUGA
 CCCUAGGCCAGAAUAUCACCGUGAUCGCCCCUCACGCCUC
 GAGAACAUCGUGCGGCAGCCUCCCGACAGAUGGAUGACCA

ACGCCAGAAUGACCCACUACCAGAGCCUGUUGCUGACCGAG AGAGUGACCUUCGCCCCUCCAGCUGCCCUGAAUCCCGCCAC UCUGCUGCCCGAGGAAACCGACGAGCCUGUGACCCACGACU GCCACCAGCUGCUGAUCGAGGAAACCGGCGUCAGAAAGGA CCUGACAGAUAUCCCUCUGACCGGAGAGGUGCUGACAUGG UUCACCGACGGCAGCAGCUACGUCGUGGAAGGCAAGCGGA UGGCCGGCGCCGCUUGGUCGACGGCACAAGAACCAUCUGG GCUUCCAGCCUGCCUGAGGGCACCAGCGCCCAGAAGGCCGA GCUGAUGGCCCUCACACAGGCCCUGCGGCUGGCUGAGGGCA AAAGCAUCAACAUCUAUACAGACAGCCGUUACGCCUUCGCC ACAGCGCACGUGCACGGCGCCAUCUACAAGCAGAGAGGAU GGCUGACCUCUGCCGGAAGAGAAAUCAAGAACAAGGAAGA AAUCCUGAGCCUGCUGGAAGCCCUGCAUCUCCCAAAGAGAC UGGCCAUCAUCCACUGCCCCGGCCACCAGAAGGCCAAAGAC CCUAUCAGCAGAGGCAACCAGAUGGCCGACCGGGUGGCCAA GCAAGCCGCCCAAGGCGUGAAUCUGCUGCCUUAGUUAUU AAGCUGCCUUCUGCGGGGCUUGCCUUCUGGCCAUGCCCUUC UUCUCUCCCUUGCACCUGUACCUCUUGGUCUUUGAAUAAA GCCUGAGUAGGAAGUCUAG
--

33. В некоторых вариантах осуществления слитый белок может содержать молекулу Cas, например, мутантный Cas9, например, Cas-нуклеазу, содержащую мутацию, ингибирующую (например, инактивирующую) один активный сайт эндонуклеазы, например, нуклеазу Cas9 Cas9(N863A). В некоторых вариантах осуществления слитый белок содержит пептидный линкер, например, гибкий пептидный линкер, богатый глицином и серином, например, линкер, раскрытый в таблицах 38 и/или 42, например, линкер 10 из таблицы 42. В некоторых вариантах осуществления слитый белок содержит домен RT, например, домен RT, содержащий последовательность из таблицы 1, таблицы 3, таблицы 30, таблицы 31, таблицы 41, таблицы 44, таблицы 50 или ее фрагмент или вариант. В некоторых вариантах осуществления слитый белок Cas-RT (или нуклеиновую кислоту, кодирующую его) составляют с gRNA. В некоторых вариантах осуществления длина линкера составляет от 2 до 40 аминокислот, от 5 до 30 аминокислот, от 5 до 20 аминокислот, от 10 до 20 аминокислот или от 10 до 15 аминокислот. В некоторых вариантах осуществления слитые белки Cas-RT обладают ДНК-связывающей активностью, сходной с таковой у молекулы Cas, которая не слита с доменом RT. В некоторых вариантах осуществления Cas-RT может содержать домен RT, содержащий мутацию. В вариантах осуществления мутантный домен RT демонстрирует повышенную

процессивность и аффинность к матрице по сравнению с немутантным доменом RT. 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, дубликаций 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.

34. Внесение одностороннего разрыва во вторую нить

В некоторых вариантах осуществления система Gene Writer, описанная в данном документе, обладает уникальной активностью, которая обеспечивает внесение одностороннего разрыва в первую нить и вторую нить ДНК-мишени. Как обсуждается в данном документе, не желая ограничиваться какой-либо теорией, полагают, что внесение одностороннего разрыва в первую нить ДНК сайта-мишени обеспечивает наличие 3'-ОН, который может использоваться доменом RT для обратной транскрипции последовательности РНК-матрицы, например, последовательности гетерологичного объекта. Не желая ограничиваться какой-либо теорией, полагают, что введение дополнительного одностороннего разрыва во вторую 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 односторонний разрыв во второй нити производится тем же эндонуклеазным доменом (например, уникальным доменом), что и односторонний разрыв в первой нити. В некоторых вариантах осуществления один и тот же полипептид Gene Writer осуществляет как односторонний разрыв в первой нити, так и односторонний разрыв во второй нити. В некоторых вариантах осуществления полипептид Gene Writer™ содержит домен CRISPR/Cas, а дополнительный односторонний разрыв во второй нити направляется дополнительной нуклеиновой кислотой,

например, содержащей вторую gRNA, направляющую домен CRISPR/Cas для внесения одностороннего разрыва во вторую нить. В других вариантах осуществления дополнительный односторонний разрыв во второй нити производится другим эндонуклеазным доменом (например, никасным доменом), нежели односторонний разрыв в первой нити. В некоторых вариантах осуществления этот другой эндонуклеазный домен расположен в дополнительном полипептиде (например, система по настоящему изобретению дополнительно содержит дополнительный полипептид), отдельном от полипептида Gene Writer. В некоторых вариантах осуществления дополнительный полипептид содержит эндонуклеазный домен (например, никасный домен), описанный в данном документе. В некоторых вариантах осуществления дополнительный полипептид содержит ДНК-связывающий домен, например, описанный в данном документе.

В данном документе предполагается, что положение, в котором происходит односторонний разрыв во второй нити относительно одностороннего разрыва в первой нити, может влиять на степень, в которой происходит одно или несколько из следующих событий: получение желаемых модификаций ДНК с помощью Gene Writing, появление нежелательных двусторонних разрывов (DSB), появление нежелательных вставок или появление нежелательных делеций. Не желая ограничиваться какой-либо теорией, полагают, что внесение одностороннего разрыва во вторую нить может происходить в двух основных ориентациях: односторонние разрывы внутрь и односторонние разрывы наружу.

В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь домен RT полимеризуется (например, с использованием РНК-матрицы (например, последовательности гетерологичного объекта)) в направлении от одностороннего разрыва во второй нити. В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь местоположение одностороннего разрыва в первой нити и местоположение одностороннего разрыва во второй нити расположены между первым сайтом PAM и вторым сайтом PAM (например, в ситуации, в которой оба односторонних разрыва производятся полипептидом (например, полипептидом Gene Writer), содержащим домен CRISPR/Cas). В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь местоположение одностороннего разрыва в первой нити и местоположение одностороннего разрыва во второй нити находятся между сайтами, где полипептид и дополнительный полипептид связываются с ДНК-мишенью. В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь местоположение одностороннего разрыва во второй нити находится на той же стороне, что и сайты связывания полипептида и дополнительного полипептида, относительно местоположения одностороннего разрыва в первой нити. В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь местоположение одностороннего разрыва в первой нити и местоположение одностороннего разрыва во второй нити расположены между сайтом PAM и сайтом на расстоянии от сайта-мишени.

Пример системы Gene Writer, обеспечивающей ориентацию одностороннего разрыва внутрь, содержит полипептид Gene Writer, содержащий домен CRISPR/Cas, РНК-матрицу,

содержащую gRNA, которая направляет внесение одностороннего разрыва ДНК сайта-мишени в первую нить, и дополнительную нуклеиновую кислоту, содержащую дополнительную gRNA, которая направляет внесение одностороннего разрыва в сайте на расстоянии от местоположения первого одностороннего разрыва, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва расположены между сайтами PAM для сайтов, к которым две gRNA направляют полипептид Gene Writer. В качестве дополнительного примера, другая система Gene Writer, которая обеспечивает ориентацию одностороннего разрыва внутрь, содержит полипептид Gene Writer, содержащий молекулу с "цинковыми пальцами" и первый сигнальный домен, где молекула с "цинковыми пальцами" связывается с ДНК-мишенью таким образом, что направляет первый сигнальный домен для одностороннего разрыва первой нити сайта-мишени; дополнительный полипептид, содержащий домен CRISPR/Cas, и дополнительную нуклеиновую кислоту, содержащую gRNA, которая направляет дополнительный полипептид для одностороннего разрыва сайта на расстоянии от ДНК сайта-мишени во второй нити, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва находятся между сайтом PAM и сайтом, с которым связывается молекула с "цинковыми пальцами". В качестве дополнительного примера, другая система Gene Writer, которая обеспечивает ориентацию одностороннего разрыва внутрь, содержит полипептид Gene Writer, содержащий молекулу с "цинковыми пальцами" и первый сигнальный домен, где молекула с "цинковыми пальцами" связывается с ДНК-мишенью таким образом, что направляет первый сигнальный домен для одностороннего разрыва первой нити сайта-мишени; дополнительный полипептид, содержащий эффекторную молекулу TAL и второй сигнальный домен, где эффекторная молекула TAL связывается с сайтом на расстоянии от сайта-мишени таким образом, что направляет дополнительный полипептид для одностороннего разрыва второй нити, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва находятся между сайтом, с которым связывается эффекторная молекула TAL, и сайтом, с которым связывается молекула с "цинковыми пальцами".

В некоторых вариантах осуществления в ориентации одностороннего разрыва наружу домен RT полимеризуется (например, с использованием РНК-матрицы (например, последовательности гетерологичного объекта)) в направлении к одностороннему разрыву во второй нити. В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь в случае, когда как первый, так и второй односторонние разрывы производятся полипептидом, содержащим домен CRISPR/Cas (например, полипептидом Gene Writer), первый сайт PAM и второй сайт PAM расположены между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити. В некоторых вариантах осуществления в ориентации одностороннего разрыва внутрь полипептид (например, полипептид Gene Writer™) и дополнительный полипептид связываются с сайтами в ДНК-мишени между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити. В некоторых

вариантах осуществления в ориентации одностороннего разрыва внутрь местоположение одностороннего разрыва во второй нити находится на противоположной стороне от сайтов связывания полипептида и дополнительного полипептида относительно местоположения одностороннего разрыва в первой нити. В некоторых вариантах осуществления в ориентации внутрь сайт PAM и сайт на расстоянии от сайта-мишени расположены между местоположением одностороннего разрыва в первой нити и местоположением одностороннего разрыва во второй нити.

Пример системы Gene Writer, обеспечивающей ориентацию одностороннего разрыва наружу, содержит полипептид Gene Writer, содержащий домен CRISPR/Cas, РНК-матрицу, содержащую gRNA, которая направляет внесение одностороннего разрыва ДНК сайта-мишени в первую нить, и дополнительную нуклеиновую кислоту, содержащую дополнительную gRNA, которая направляет внесение одностороннего разрыва в сайте на расстоянии от местоположения первого одностороннего разрыва, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва расположены вне сайтов PAM для сайтов, к которым две gRNA направляют полипептид Gene Writer (т. е. сайты PAM находятся между местоположением первого одностороннего разрыва и местоположением второго одностороннего разрыва). В качестве дополнительного примера, другая система Gene Writer, которая обеспечивает ориентацию одностороннего разрыва наружу, содержит полипептид Gene Writer, содержащий молекулу с "цинковыми пальцами" и первый сигнальный домен, где молекула с "цинковыми пальцами" связывается с ДНК-мишенью таким образом, что направляет первый сигнальный домен для одностороннего разрыва первой нити сайта-мишени; дополнительный полипептид, содержащий домен CRISPR/Cas, и дополнительную нуклеиновую кислоту, содержащую gRNA, которая направляет дополнительный полипептид для одностороннего разрыва сайта на расстоянии от ДНК сайта-мишени во второй нити, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва находятся вне сайта PAM и сайта, с которым связывается молекула с "цинковыми пальцами" (т. е. сайт PAM и сайт, с которым связывается молекула с "цинковыми пальцами", находятся между местоположением первого одностороннего разрыва и местоположением второго одностороннего разрыва). В качестве дополнительного примера, другая система Gene Writer, которая обеспечивает ориентацию одностороннего разрыва наружу, содержит полипептид Gene Writer, содержащий молекулу с "цинковыми пальцами" и первый сигнальный домен, где молекула с "цинковыми пальцами" связывается с ДНК-мишенью таким образом, что направляет первый сигнальный домен для одностороннего разрыва первой нити сайта-мишени; дополнительный полипептид, содержащий эффекторную молекулу TAL и второй сигнальный домен, где эффекторная молекула TAL связывается с сайтом на расстоянии от сайта-мишени таким образом, что направляет дополнительный полипептид для одностороннего разрыва второй нити, где местоположение первого одностороннего разрыва и местоположение второго одностороннего разрыва находятся вне сайта, с которым связывается эффекторная молекула TAL, и сайта, с которым связывается молекула с "цинковыми пальцами" (т. е. сайт, с которым связывается

эффекторная молекула TAL, и сайт, с которым связывается молекула с "цинковыми пальцами", находятся между местоположением первого однострессового разрыва и местоположением второго однострессового разрыва).

Не желая ограничиваться какой-либо теорией, полагают, что для систем Gene Writer, в которых предусмотрен однострессовый разрыв второй нити, в некоторых вариантах осуществления предпочтительна ориентация однострессового разрыва наружу. Как описано в данном документе, однострессовый разрыв в ориентации внутрь может приводить к более высокому количеству двухстрессовых разрывов (DSB), чем однострессовый разрыв в ориентации наружу. DSB могут быть распознаны путями репарации DSB в ядре клетки, что может приводить к нежелательным вставкам и делециям. Ориентация однострессового разрыва наружу может обеспечивать снижение риска образования DSB и, соответственно, меньшее количество нежелательных вставок и делеций. В некоторых вариантах осуществления нежелательные вставки и делеции представляют собой вставки и делеции, не кодируемые последовательностью гетерологичного объекта, например, вставка или делеция, образующиеся в пути репарации двухстрессовых разрывов, не связанные с модификацией, кодируемой последовательностью гетерологичного объекта. В некоторых вариантах осуществления желаемая модификация с помощью Gene Writing включает изменение ДНК-мишени (например, замену, вставку или делецию), кодируемое последовательностью гетерологичного объекта (например, которое достигается посредством записи последовательности гетерологичного объекта в сайт-мишень с помощью Gene Writer). В некоторых вариантах осуществления однострессовый разрыв в первой нити и однострессовый разрыв во второй нити представлены в ориентации наружу.

Кроме того, расстояние между однострессовым разрывом в первой нити и однострессовым разрывом во второй нити может влиять на степень, в которой происходит одно или несколько из следующих событий: получение желаемых модификаций ДНК с помощью Gene Writing, появление нежелательных двухстрессовых разрывов (DSB), появление нежелательных вставок или появление нежелательных делеций. Не желая ограничиваться какой-либо теорией, полагают, что преимущество однострессового разрыва во второй нити - смещение репарации ДНК в сторону включения последовательности гетерологичного объекта в ДНК-мишень - увеличивается по мере уменьшения расстояния между однострессовым разрывом в первой нити и однострессовым разрывом во второй нити. Однако считается, что риск образования DSB также увеличивается по мере уменьшения расстояния между однострессовым разрывом в первой нити и однострессовым разрывом во второй нити. Соответственно считается, что количество нежелательных вставок и/или делеций может увеличиваться по мере уменьшения расстояния между однострессовым разрывом в первой нити и однострессовым разрывом во второй нити. В некоторых вариантах осуществления расстояние между однострессовым разрывом в первой нити и однострессовым разрывом во второй нити выбирают, чтобы сбалансировать преимущества смещения репарации ДНК в сторону включения последовательности гетерологичного объекта в ДНК-мишень и риск образования DSB и нежелательных делеций и/или вставок. В некоторых вариантах

осуществления система, где односторонний разрыв в первой нити и односторонний разрыв во второй нити находятся на по меньшей мере пороговом расстоянии друг от друга, характеризуется повышенным уровнем желаемых результатов модификации с помощью Gene Writing, сниженным уровнем нежелательных делеций и/или сниженным уровнем нежелательных вставок по сравнению с в остальном аналогичной системой с ориентацией одностороннего разрыва внутрь, где первый односторонний разрыв и второй односторонний разрыв находятся на менее чем пороговом расстоянии друг от друга. В некоторых вариантах осуществления пороговое(пороговые) расстояние(расстояния) указаны ниже.

В некоторых вариантах осуществления первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии по меньшей мере 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 друг от друга. В некоторых вариантах осуществления первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии не более 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, 200 или 250 нуклеотидов друг от друга. В некоторых вариантах осуществления первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии 20-200, 30-200, 40-200, 50-200, 60-200, 70-200, 80-200, 90-200, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 20-190, 30-190, 40-190, 50-190, 60-190, 70-190, 80-190, 90-190, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 20-180, 30-180, 40-180, 50-180, 60-180, 70-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 20-170, 30-170, 40-170, 50-170, 60-170, 70-170, 80-170, 90-170, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 20-160, 30-160, 40-160, 50-160, 60-160, 70-160, 80-160, 90-160, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 20-150, 30-150, 40-150, 50-150, 60-150, 70-150, 80-150, 90-150, 100-150, 110-150, 120-150, 130-150, 140-150, 20-140, 30-140, 40-140, 50-140, 60-140, 70-140, 80-140, 90-140, 100-140, 110-140, 120-140, 130-140, 20-130, 30-130, 40-130, 50-130, 60-130, 70-130, 80-130, 90-130, 100-130, 110-130, 120-130, 20-120, 30-120, 40-120, 50-120, 60-120, 70-120, 80-120, 90-120, 100-120, 110-120, 20-110, 30-110, 40-110, 50-110, 60-110, 70-110, 80-110, 90-110, 100-110, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 20-90, 30-90, 40-90, 50-90, 60-90, 70-90, 80-90, 20-80, 30-80, 40-80, 50-80, 60-80, 70-80, 20-70, 30-70, 40-70, 50-70, 60-70, 20-60, 30-60, 40-60, 50-60, 20-50, 30-50, 40-50, 20-40, 30-40 или 20-30 нуклеотидов друг от друга. В некоторых вариантах осуществления первый односторонний разрыв и второй односторонний разрыв находятся на расстоянии 40-100 нуклеотидов друг от друга.

Не желая ограничиваться какой-либо теорией, полагают, что для систем Gene Writer, в которых предусмотрен односторонний разрыв второй нити и выбрана ориентация одностороннего разрыва внутрь, может быть предпочтительным увеличение расстояния между односторонним разрывом в первой нити и односторонним разрывом во второй нити. Как описано в данном документе, ориентация одностороннего разрыва внутрь может приводить к более высокому количеству DSB, чем ориентация одностороннего разрыва наружу, и может приводить к более высокому количеству нежелательных вставок и делеций, чем ориентация

однонитевого разрыва наружу, но увеличение расстояния между однонитевыми разрывами может нивелировать это увеличение количества DSB, нежелательных делеций и/или нежелательных вставок. В некоторых вариантах осуществления ориентация однонитевого разрыва внутрь, где первый однонитевой разрыв и второй однонитевой разрыв находятся на по меньшей мере пороговом расстоянии друг от друга, характеризуется повышенным уровнем желаемых результатов модификации с помощью Gene Writing, сниженным уровнем нежелательных делеций и/или сниженным уровнем нежелательных вставок по сравнению с в остальном аналогичной системой с ориентацией однонитевого разрыва внутрь, где первый однонитевой разрыв и второй однонитевой разрыв находятся на менее чем пороговом расстоянии друг от друга. В некоторых вариантах осуществления пороговое расстояние указано ниже.

В некоторых вариантах осуществления однонитевой разрыв в первой нити и однонитевой разрыв во второй нити представлены в ориентации внутрь. В некоторых вариантах осуществления однонитевой разрыв в первой нити и однонитевой разрыв во второй нити представлены в ориентации внутрь, и однонитевой разрыв в первой нити и однонитевой разрыв во второй нити находятся на расстоянии по меньшей мере 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450 или 500 нуклеотидов друг от друга, например, по меньшей мере 100 нуклеотидов друг от друга (и необязательно не более 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130 или 120 нуклеотидов друг от друга). В некоторых вариантах осуществления однонитевой разрыв в первой нити и однонитевой разрыв во второй нити представлены в ориентации внутрь, и однонитевой разрыв в первой нити и однонитевой разрыв во второй нити находятся на расстоянии 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 100-150, 110-150, 120-150, 130-150, 140-150, 100-140, 110-140, 120-140, 130-140, 100-130, 110-130, 120-130, 100-120, 110-120 или 100-110 нуклеотидов друг от друга.

35. 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 РСТ № РСТ/US 2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11

марта 2010 г.; международной заявке согласно РСТ РСТ/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 РСТ РСТ/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 РСТ РСТ/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 РСТ РСТ/US2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной заявке согласно РСТ РСТ/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 РСТ РСТ/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: (а) 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 определенных 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 $5 \cdot 10^5$ клеток/мл, приблизительно 10^6 клеток/мл, приблизительно $5 \cdot 10^6$ клеток/мл, приблизительно 10^7 клеток/мл, приблизительно $5 \cdot 10^7$ клеток/мл, приблизительно 10^8 клеток/мл, приблизительно $5 \cdot 10^8$ клеток/мл, приблизительно 10^9 клеток/мл, приблизительно $5 \cdot 10^9$ клеток/мл, приблизительно 10^{10} клеток/мл или приблизительно $5 \cdot 10^{10}$ клеток/мл.

36. 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 определенных 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](http://epd.epfl.ch/index.php)).

Таблица 33. Иллюстративные клеточно- или тканеспецифические промоторы

Промотор	Клетки-мишени
Промотор B29	В-клетки
Промотор CD14	Моноцитарные клетки
Промотор CD43	Лейкоциты и тромбоциты
Промотор CD45	Гемопозитические 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	Мегакарициты
Промотор ICAM-2	Эндотелиальные клетки
Промотор INF-бета	Гемопозитические клетки
Промотор Mb	Мышечные клетки
Промотор Nphs1	Подоциты
Промотор OG-2	Остеобласты, одонтобласты
Промотор SP-B	Легкое
Промотор Syn1	Нейроны
Промотор WASP	Гемопозитические клетки
Промотор SV40/bAlb	Печень
Промотор SV40/bAlb	Печень
Промотор SV40/Cd3	Лейкоциты и тромбоциты
Промотор SV40/CD45	Гемопозитические клетки
Промотор NSE/RU5'	Зрелые нейроны

Таблица 34. Дополнительные иллюстративные клеточно- или тканеспецифические промоторы

Промотор	Описание гена	Специфичность гена
APOA2	Аполипопротеин А-II	Гепатоциты (из предшественников гепатоцитов)
SERPINA1 (hAAT)	Ингибитор пептидаз SERPIN, клада А (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	Гомеобокс 1 поджелудочной железы и двенадцатиперстной кишки	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 (αMHC)	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 А (также называемый предсердным натрийуретическим фактором)	Предсердная специфичность во взрослых клетках
Slc8a1 (Ncx1)	Семейство 8 переносчиков растворенных веществ (натрий/кальцевый обменник), представитель 1	Кардиомиоциты из ранних стадий развития
Промоторы, специфичные для ЦНС		
Промотор	Описание гена	Специфичность гена
SYN1 (hSyn)	Синапсин I	Нейроны

GFAP	Глиальный фибрилярный кислый белок	Астроциты
INA	Белок промежуточных филаментов нейронов альфа-интернексин (α- 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	Трансглутаминаза 3	Кератиноциты из зернистого слоя
Промоторы, специфичные для иммунных клеток		
Промотор	Описание гена	Специфичность гена
ITGAM (CD11B)	Интегрин, α-М (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.

В иллюстративных целях 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 (CamKII α) (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 SM22 α (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 SM22 α 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-фосфодиэстеразы

(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 фосфолипидкиназы (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 фосфолипидкиназы, 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 Е (АроЕ) или его функциональный фрагмент, например, для управления экспрессией в печени. В некоторых вариантах осуществления используются две копии энхансера АроЕ или его функционального фрагмента. В некоторых вариантах осуществления энхансер АроЕ 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 (МСК), промотор гена десмина млекопитающих (DES), промотор гена тяжелой цепи α -миозина (α -MHC) или промотор гена сердечного тропонина Т (сTnT). Другие примеры промоторов включают промотор гена бета-актина, коровий промотор вируса гепатита В, Sandig et al., *Gene Ther.*, 3:1002-9 (1996); промотор гена альфа-фетопroteина (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 Т-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 одонитевую 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 gAAV, 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 созревaniem 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 РСТ № 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 данным документом, 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, кодирующую 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	uguaguguiiucsuacuuuaugga
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-5p	uuuggcaaugguagaacucacacu
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-3p	ugguucuaagacuugccaacua
Ганглиозные нейроны дорсального корешка	miR-183	hsa-miR-183-5p	uauggcacugguagaauucacu
Ганглиозные нейроны дорсального корешка	miR-183	hsa-miR-183-3p	gugaauuaccgaagggccauaa
Гепатоциты	miR-122	hsa-miR-122-5p	uggagugugacaauagguguuug
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-НТ)-отвечающий интеин, молекулу iCas (например, iCas9); 4-НТ-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 ВНЗ, 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 биопереключателю), белки AcrIIA2 и AcrIIA4, а также BRD0539.

В некоторых вариантах осуществления интеин, отвечающий на малые молекулы (например, 4-гидрокситамоксифен (4-НТ)-отвечающий интеин), встроен в определенные сайты молекулы Cas (например, Cas9). В некоторых вариантах осуществления встраивание 4НТ-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 4НТ-отвечающего средства, противодействующего CRISPR-Cas9 (например, arC9 или dC9). В некоторых вариантах осуществления dCas9 может обеспечивать репрессию функции Cas9, зависящую от дозы 4-НТ. В некоторых вариантах осуществления arC9 может обеспечивать контроль функции Cas9, зависящий от дозы 4-НТ. 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 ВНЗ к N- и C-концам модифицированного Cas9.BCL. В некоторых вариантах осуществления взаимодействие между пептидами BCL-xL и ВНЗ может удерживать Cas9 в неактивном состоянии. В некоторых вариантах осуществления малая молекула (например, A-385358 (A3)) может нарушать взаимодействие между пептидами BCL-xL и ВНЗ 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.

37. Самоинактивирующиеся модули для регулирования активности Gene Writer

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 по 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 односторонний 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 (рапалогам), 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 и обеспечивая агрегацию полипептида каспазы. При агрегации полипептид каспазы индуцирует апоптоз. Количество рапамицина или аналога рапамицина, вводимого пациенту, может варьироваться; если желательно удаление более низкого уровня клеток посредством апоптоза, 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 римиудуцида 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 определенных 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.

38. *Разрешение событий Gene Writing™*

После записи нуклеиновой кислоты-матрицы в сайт-мишень могут осуществляться дополнительные формы активности для повышения общей эффективности включения. В некоторых вариантах осуществления односторонний разрыв может быть инициирован в геноме на незаписываемой нити ДНК, чтобы стимулировать копирование вновь записанной ДНК во вторую нить. В некоторых вариантах осуществления односторонний разрыв может находиться в пределах по меньшей мере 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 оснований от сайта-мишени. В некоторых вариантах осуществления этот второй односторонний разрыв осуществляется тем же полипептидом, который осуществляет запись. В других вариантах осуществления второй односторонний разрыв может осуществляться дополнительным полипептидом, отвечающим за никазную активность, например, никазой Cas9.

Для некоторых систем Gene Writer™ после процесса записи может оставаться 3'-флэп, содержащий вновь записанную ДНК, которая должна вытеснить фланкирующую последовательность-мишень для отжига со второй геномной нитью и завершения редактирования. В некоторых вариантах осуществления 3'-флэп разработан таким образом, чтобы он имел усиленную способность к внедрению нити. В некоторых вариантах осуществления добавляется 5'-3'-эксонуклеазная активность для расщепления открытого 5'-конца вытесненной цепи. В некоторых вариантах осуществления добавляется ДНК-лигазная активность для завершения реакции. В некоторых вариантах осуществления экзонуклеазные и/или лигазные формы активности необязательно предусмотрены в полипептиде Gene Writer™. В некоторых вариантах осуществления экзонуклеазные и/или лигазные формы активности необязательно предусмотрены отдельно от полипептида Gene Writer™.

Исходя из опубликованного механизма действия ретротранспозонов, не содержащих LTR, полученные из них системы Gene Writing™ могут не требовать добавления дополнительных функций для разрешения события записи. В некоторых вариантах

осуществления система может приводить к полной записи, не требующей эндогенных факторов хозяина. В некоторых вариантах осуществления система может приводить к полной записи, не требующей репарации ДНК. В некоторых вариантах осуществления система может приводить к полной записи, не вызывая ответ на повреждение ДНК.

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; или gRNA) может содержать немодифицированные или модифицированные азотистые основания. Встречающиеся в природе РНК синтезируются из четырех основных рибонуклеотидов: АТР, СТР, УТР и ГТР, 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 РСТ/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 ЕР 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-модифицированных 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, C5-пропинилдезоксцитидина (pdC), C5-пропинилдеоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдеоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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 6, одну или несколько химических модификаций остова из таблицы 7, один или несколько химически модифицированных кэпов из таблицы 7. 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 6. 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 7. В качестве альтернативы или в комбинации, нуклеиновая кислота может содержать один или несколько модифицированных кэпов, например, описанных в данном документе, например, в таблице 8. 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.

39. Таблица 6. Модифицированные нуклеотиды

ридин	тил-6-тиогуанозин
5-азауридин	-диметил-6-тиогуанозин
ридин	дин-4-оновый рибонуклеозид
псевдоуридин	5-азауридин
псевдоуридин	уридин
оксиуридин	псевдоуридин

илуридин	псевдоуридин
боксиметилуридин	илуридин
боксиметилпсевдоуридин	пинилпсевдоуридин
пинилуридин	ил-1-деазапсевдоуридин
пинилпсевдоуридин	1-метил-1-деазапсевдоуридин
инометилуридин	оксипсевдоуридин
инометилпсевдоуридин	1-тиофосфат)-аденозин
инометил-2-тиоуридин	1-тиофосфат)-цитидин
инометил-4-тиоуридин	1-тиофосфат)-гуанозин
илуридин	1-тиофосфат)-уридин
илпсевдоуридин	1-тиофосфат)-псевдоуридин
1-метилпсевдоуридин	метиладенозин
1-метилпсевдоуридин	метилцитидин
ил-1-деазапсевдоуридин	метилгуанозин
1-метил-1-деазапсевдоуридин	метилуридин
роуридин	метилпсевдоуридин
ропсевдоуридин	метиинозин
дигидроуридин	иладенозин
дигидропсевдоуридин	илтио-N6-метиладенозин
оксиуридин	илтио-N6 изопентениладенозин
окси-4-тиоуридин	илтио-N6-(цис-
оксипсевдоуридин	гидроксиизопентенил)аденозин
окси-2-тиопсевдоуридин	етил-N6-треонилкарбамоиладенозин
цитидин	дроксинорвалилкарбамоиладенозин
ризоцитидин	илтио-N6-
илцитидин	гидроксинорвалилкарбамоиладенозин
етилцитидин	ибозиладенозин (фосфат)
милцитидин	-диметиинозин
тилцитидин	-диметилцитидин
оксиметилцитидин	етил-2'-О-метилцитидин
илпсевдоизоцитидин	дин
лоцитидин	илгуанозин
лопсевдоизоцитидин	О-диметилгуанозин

илтиoadенин	бамоилметилуридин
оксиаденин	бамоилметил-2'-О-метилуридин
ин	боксиметиламинометилуридин
илинозин	боксиметиламинометил-2'-О-метилуридин
ин	боксиметиламинометил-2-тиоуридин
озин	О-диметилцитидин
загуанозин	боксиметилуридин
за-8-азагуанозин	О-диметиладенозин
гуанозин	О-2'-триметиладенозин
7-деазагуанозин	диметилгуанозин
7-деаза-8-азагуанозин	1,7-триметилгуанозин
илгуанозин	-диметилуридин
7-метилгуанозин	илдигидроуридин
илинозин	мил-2'-О-метилцитидин
оксигуанозин	-диметилгуанозин
илгуанозин	етилвиозин
етилгуанозин	юзин
1-диметилгуанозин	етиладенозин
огуанозин	
ил-8-оксогуанозин	
ил-6-тиогуанозин	

40. Таблица 7. Модификации остова

2'-О-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

41. Таблица 8. 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

m7,2'dGppppG

m7,3'OmeGppppG

m7,3'dGppppG

42. Получение композиций и систем

Специалисту в данной области будет понятно, что способы разработки и конструирования конструкций на основе нуклеиновых кислот и белков или полипептидов

(таких как системы, конструкции и полипептиды, описанные в данном документе) являются стандартными в данной области техники. 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 конструирования, 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 селективируемый 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-лактамам 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 селективируемый 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 селективируемый маркер (например, ген устойчивости к антибиотикам), последовательность, регулируемую 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. Векторы экспрессии 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 векторы и векторы экспрессии 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 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 частично 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. Настоящее изобретение дополнительно 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:

(i) длину РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-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, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы 5'-кэп, например, является ли этот кэп 7-метилгуанозинowym 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, содержат 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, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активность РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу;

(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, имеют 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, содержат 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), аминизомасляной кислоты, дегидроаланина, цистатинина, лантиниона, дженколовой кислоты, диаминопимелиновой кислоты, гомоаланина, норвалина, норлейцина, гомонорлейцина, гомосерина, O-метилгомосерина и O-этилгомосерина, этионина, селеноцистеина, селеногомоцистеина, селенометионина, селеноэтионина, теллурицистеина или теллуromетионина) в полипептиде, первом полипептиде или втором полипептиде, например, содержат 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, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактными (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, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей полипептид, первый полипептид или второй полипептид; 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 контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

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, имеют 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(А)-хвоста в мРНК, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствие, отсутствие и/или тип 5'-кэпа на мРНК, например, содержат 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 О-Ме-m⁷G;

(iv) присутствие, отсутствие и/или тип одного или нескольких модифицированных нуклеотидов (например, выбранных из псевдоуридина, дигидроуридина, инозина, 7-метилгуанозина, 1-N-метилпсевдоуридина (1-Ме-Ψ), 5-метоксиуридина (5-МО-U), 5-метилцитидина (5mC) или запертого нуклеотида) в мРНК, например, содержат 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, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% мРНК интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность; или

(vi) активность мРНК в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей мРНК.

Кольцевые РНК в системе Gene Writing

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 интрона в молекулу РНК (или ДНК, кодирующую молекулу РНК), что приводит к циркуляризации РНК, и сконструированная таким образом circRNA может характеризоваться повышенными продуцированием белка и стабильностью (Wesselhoeft et al. *Nature Communications* 2018). 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, 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, линейаризуется (например, в клетке-хозяине) 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 интрона в молекулу РНК (или ДНК, кодирующую молекулу РНК), что приводит к циркуляризации РНК, и сконструированная таким образом circRNA может характеризоваться повышенным 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 полипептид Gene Writer™ кодируется в виде circRNA. Хотя в определенных вариантах осуществления нуклеиновая кислота-матрица представляет собой ДНК, такую как ssDNA, в некоторых вариантах осуществления она может быть представлена в виде РНК, например, с обратной транскриптазой.

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 линейаризации 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 линейаризуется в ядре клетки-мишени. В некоторых вариантах осуществления линейаризация 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 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 содержит 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, tРНК, 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.

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 некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид, или нуклеиновую кислоту-матрицу, или и то, и другое), доставляемая в клетки, разработана в виде миникольца, где последовательности остова плазмиды, не относящиеся к Gene Writer™, удаляются перед введением в клетки. Например, в миникольце может отсутствовать бактериальная точка начала репликации и селектируемый маркер. В некоторых вариантах осуществления миникольцо не содержит никакой бактериальной последовательности. Было показано, что миникольца приводят к более высокой эффективности трансфекции и экспрессии генов по сравнению с плазмидами с остовами, содержащими бактериальные части (например, бактериальную точку начала репликации, кассету селекции по устойчивости к антибиотикам), и использовались для увеличения эффективности транспозиции (Sharma et al Mol Ther Nucleic Acids 2013). В некоторых вариантах осуществления ДНК-вектор, кодирующий полипептид Gene Writer™, доставляется в виде миникольца. В некоторых вариантах осуществления ДНК-вектор, содержащий нуклеиновую кислоту-матрицу Gene Writer™ (например, РНК-матрицу), доставляется в виде миникольца. В некоторых вариантах осуществления бактериальные части фланкированы сайтами рекомбинации, например, сайтами attP/attB, loxP, FRT. В некоторых вариантах осуществления добавление когнатной рекомбиназы приводит к внутримолекулярной рекомбинации и вырезанию бактериальных частей. В некоторых вариантах осуществления сайты для рекомбиназы распознаются рекомбиназой phiC31. В некоторых вариантах осуществления сайты для рекомбиназы распознаются рекомбиназой Cre. В некоторых вариантах осуществления сайты для рекомбиназы распознаются рекомбиназой FLP. В дополнение к плазмидной ДНК, миникольца можно получить путем вырезания необходимой конструкции, например, кассет экспрессии транспозазы или

кассеты экспрессии РНК-матрицы, из остова вируса. Ранее было показано, что вырезание и циркуляризация донорной последовательности из каркасной последовательности вируса может иметь важное значение для эффективности интеграции, опосредованной транспозазой (Yant et al *Nat Biotechnol* 2002). В некоторых вариантах осуществления миникольца вначале составляют, а затем доставляют к клеткам-мишеням. В других вариантах осуществления миникольца образуются внутриклеточно из ДНК-вектора (например, плазмидной ДНК, гAAV, scAAV, ceDNA, ДНК "Doggybone") посредством совместной доставки с рекомбиназой, что приводит к вырезанию и циркуляризации нуклеиновой кислоты, фланкированной сайтом распознавания рекомбиназы, например, нуклеиновой кислоты, кодирующей полипептид Gene Writer™, нуклеиновой кислоты-матрицы (например, РНК-матрицы) или нуклеиновой кислоты, кодирующей ее, или и того, и другого.

Для оптимизации экспрессии белка может быть полезно предоставить настраиваемые контрольные элементы, которые можно применять для модулирования активности белка. В некоторых вариантах осуществления настраиваемая система может содержать по меньшей мере один эффекторный модуль, отвечающий на по меньшей мере один стимул. Система может представлять собой без ограничения систему дестабилизирующего домена (DD). Данная система дополнительно описана в РСТ/US2018/020704, а также в предварительной заявке на патент США № 62/320864, поданной 11 апреля 2016 г., № 62/466596, поданной 3 марта 2017 г., и в международной публикации WO2017/180587 (содержание каждой из которых включено в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления настраиваемая система может содержать первый эффекторный модуль. В некоторых вариантах осуществления эффекторный модуль может содержать первый элемент ответа на стимул (SRE), функционально связанный с по меньшей мере одной полезной нагрузкой. В одном аспекте полезная нагрузка может представлять собой иммунотерапевтическое средство. В одном аспекте второй SRE композиции может отвечать на по меньшей мере один стимул или взаимодействовать с ним. В некоторых вариантах осуществления первый SRE может содержать дестабилизирующий домен (DD). DD может быть получен из исходного белка или из мутантного белка, имеющего одну, две, три или больше аминокислотных мутаций по сравнению с исходным белком. В некоторых вариантах осуществления исходный белок может быть выбран без ограничения из белка FKBP человека, содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 3; DHFR человека (hDHFR), содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 2; DHFR из *E. coli*, содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 1; PDE5, содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 4; PPAR-гамма, содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 5; CA2, содержащего аминокислотную последовательность РСТ/US2018/020704 под SEQ ID NO: 6; или NQ02, содержащего аминокислотную последовательность РСТ/US2018/020704

под SEQ ID NO: 7. В некоторых вариантах осуществления к полипептиду Gene Writer применяются настраиваемые контрольные элементы, так что, например, DD и стимул можно применять для модулирования эффективности интеграции матрицы. В некоторых вариантах осуществления настраиваемые контрольные элементы применяются к одному или нескольким пептидам, закодированным в последовательности гетерологичного объекта матрицы, так что, например, DD и стимул можно применять для модулирования активности полезной нагрузки, интегрированной в геном. В определенных вариантах осуществления полезная нагрузка, содержащая DD, может представлять собой терапевтический белок, например, функциональную копию подвергнутого эндогенной мутации гена. В определенных вариантах осуществления полезная нагрузка, содержащая DD, может представлять собой гетерологичный белок, например, CAR.

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;

(с) 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 (СМС)

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) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа 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, содержат ли

по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу;

(vii) длины 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, имеют 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) присутствия, отсутствия и/или типа посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержат 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) присутствия, отсутствия и/или типа 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), аминокислоты, дегидроаланина, цистатинина, лантинина, дженколовой кислоты, диаминопимелиновой кислоты, гомоаланина, норвалина, норлейцина, гомонорлейцина, гомосерина, O-метилгомосерина и O-этилгомосерина, этионина, селеноцистеина, селеногомоцистеина, селенометионина, селеноэтионина, теллурицистеина или теллуροметионина) в полипептиде, первом полипептиде или втором полипептиде, например, содержат 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) стабильности полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактными (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) активности полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей полипептид, первый полипептид или второй полипептид; или

(xii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или 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 контаминации пирогенами, вирусами, грибами, бактериальными патогенами и/или 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.

43. 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 геномом человека, 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 белки включают белковый иммунного рецептора, например, синтетический иммунорецепторный белок, такой как химерный антигенный рецепторный белок (CAR), Т-клеточный рецептор, В-клеточный рецептор или антитело.

Система Gene Writing™ может применяться для модификации иммунных клеток. В некоторых вариантах осуществления для модификации Т-клеток можно применять систему Gene Writing™. В некоторых вариантах осуществления Т-клетки могут включать любую субпопуляцию Т-клеток, например, CD4+, CD8+, гамма/дельта, необученные Т-клетки, стволовые Т-клетки памяти, центральные Т-клетки памяти или смесь субпопуляций. В некоторых вариантах осуществления система Gene Writing™ может применяться для доставки или модификации Т-клеточного рецептора (TCR) в Т-клетке. В некоторых вариантах осуществления система Gene Writing™ может применяться для доставки по меньшей мере одного химерного антигенного рецептора (CAR) в Т-клетки. В некоторых вариантах осуществления система Gene Writing™ может применяться для доставки по меньшей мере одного CAR в естественные клетки-киллеры (NK). В некоторых вариантах

осуществления система Gene Writing™ может применяться для доставки по меньшей мере одного CAR в естественные Т-клетки-киллеры (NKT). В некоторых вариантах осуществления система Gene Writing™ может применяться для доставки по меньшей мере одного CAR в клетку-предшественника, например, в клетку-предшественника Т-, NK- или NKT-клеток. В некоторых вариантах осуществления клетки, модифицированные по меньшей мере одним CAR (например, CAR-Т-клетки, CAR-NK-клетки, CAR-NKT-клетки), или комбинация клеток, модифицированных по меньшей мере одним CAR (например, смесь CAR-NK/Т-клеток), применяется для лечения состояния, которое определено в целевом ландшафте CAR-терапии в MacKay, et al. Nat Biotechnol 38, 233-244 (2020), включенном в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления иммунные клетки содержат CAR, специфичный к опухолевому или патогенному антигену, выбранному из группы, состоящей из AChR (фетального ацетилхолинового рецептора), ADGRE2, AFP (альфа-фетопротеина), BAFF-R, BCMA, CAIX (карбоангидразы IX), CCR1, CCR4, CEA (раково-эмбрионального антигена), CD3, CD5, CD8, CD7, CD10, CD13, CD14, CD15, CD19, CD20, CD22, CD30, CD33, CLLI, CD34, CD38, CD41, CD44, CD49f, CD56, CD61, CD64, CD68, CD70, CD74, CD99, CD117, CD123, CD133, CD138, CD44v6, CD267, CD269, CDS, CLEC12A, CS1, EGP-2 (эпителиального гликопротеина-2), EGP-40 (эпителиального гликопротеина-40), EGFR(HER1), EGFR-VIII, EpCAM (молекулы адгезии эпителиальных клеток), EphA2, ERBB2 (HER2, рецептора 2 эпидермального фактора роста человека), ERBB3, ERBB4, FBP (фолат-связывающего белка), рецептора Flt3, рецептора фолиевой кислоты α, GD2 (ганглиозида G2), GD3 (ганглиозида G3), GPC3 (глипикана-3), GPI00, hTERT (теломеразной обратной транскриптазы человека), ICAM-1, интегрин αVβ3, рецептора интерлейкина 6, IL13Ra2 (рецептора интерлейкина-13 субъединицы α-2), κ-цепи, KDR (рецептора домена вставки киназы), LeY (антигена Y системы Льюис), L1CAM (молекулы клеточной адгезии LI), LILRB2 (рецептора B2, подобного лейкоцитарному иммуноглобулину), MART1, MAGE-A1 (ассоциированного с меланомой антигена A1), MAGE-A3, MSLN (мезотелина), MUC16 (муцина 16), MUC1 (муцина I), лигандов KG2D, NY-ESO-1 (антигена рака/яичка), PRI (протеиназы 3), TRBC1, TRBC2, TFM-3, TAC1, тирозиназы, сурвивина, hTERT, онкофетального антигена (h5T4), p53, PSCA (простатического антигена стволовых клеток), PSMA (простатического специфического мембранного антигена), hROR1, TAG-72 (опухолеассоциированного гликопротеина 72), VEGF-R2 (фактора роста эндотелия сосудов R2), WT-1 (опухолевого белка Вильмса) и антигенов HIV (вируса иммунодефицита человека), гепатита В, гепатита С, CMV (цитомегаловируса), EBV (вируса Эпштейна-Барр), HPV (вируса папилломы человека).

В некоторых вариантах осуществления иммунные клетки, например, Т-клетки, NK-клетки, NKT-клетки или клетки-предшественники, модифицируют *ex vivo* и затем доставляют пациенту. В некоторых вариантах осуществления система Gene Writer™ доставляется одним из способов, упомянутых в данном документе, и иммунные клетки,

например, Т-клетки, NK-клетки, NKT-клетки или клетки-предшественники, модифицируются *in vivo* у пациента.

В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, доставляется в ткань или клетку головного мозга, мозжечка, надпочечников, яичников, поджелудочной железы, парашитовидной 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, парашитовидной 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, кожного, 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 Writing может применяться для осуществления вставки, делеции, замены или их комбинации в клетке, ткани или у субъекта. 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 система Gene Writing может применяться для осуществления множественных модификаций (например, множественных вставок, делеций или замен и всех их комбинаций) клетки-мишени либо одновременно, либо последовательно. В некоторых вариантах осуществления система Gene Writing может применяться для дополнительной модификации уже модифицированной клетки. В некоторых вариантах осуществления система Gene Writing может применяться для модификации клетки, отредактированной с помощью дополнительной технологии, например, клетки с отредактированным геном, например, клетки с одним или несколькими нокаутами CRISPR. В некоторых вариантах осуществления ранее отредактированная клетка представляет собой Т-клетку. В некоторых вариантах осуществления предыдущие модификации включают нокауты генов в Т-клетке, например, эндогенного TCR (например, TRAC, TRBC), HLA класса I (B2M), PD1, CD52, CTLA-4, TIM-3, LAG-3, DGK. В некоторых вариантах осуществления система Gene Writing применяется для вставки TCR или CAR в ранее модифицированную Т-клетку.

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 использоваться 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 использоваться 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 использоваться 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 tandemных копий тринуклеотидного повтора CAG, тогда как пациенты с болезнью Хантингтона часто имеют > 40 копий, что может привести, например, к образованию удлиненного и дисфункционального белка гентингина. В

некоторых вариантах осуществления Gene Writer корректирует экспансию повтора, например, распознавая ДНК на конце области повторов и внося одонитевой 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 одонитевой 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 система Gene Writer, описанная в данном документе, применяется для лечения показаний, указанных в любой из таблиц 9-12. Например, 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 9-12, например, у субъекта, имеющего соответствующие показания, указанные в любой из таблиц 9-12. В некоторых вариантах осуществления клетка представляет собой клетку печени, а сайт-мишень представляет собой ген из таблицы 9, например, у субъекта, имеющего соответствующие показания, указанные в таблице 9. В некоторых вариантах осуществления клетка представляет собой HSC, а сайт-мишень представляет собой ген из таблицы 10, например, у субъекта, имеющего соответствующие показания, указанные в таблице 10. В некоторых вариантах осуществления клетка представляет собой клетку ЦНС, а сайт-мишень представляет собой ген из таблицы 11, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11. В некоторых вариантах

осуществления клетка представляет собой клетку глаза, а сайт-мишень представляет собой ген из таблицы 12, например, у субъекта, имеющего соответствующие показания, указанные в таблице 12. В некоторых вариантах осуществления сайт-мишень находится в кодирующей области гена. В некоторых вариантах осуществления сайт-мишень находится в промоторе. В некоторых вариантах осуществления сайт-мишень находится в 5'-UTR или 3'-UTR гена из любой из таблиц 9-12. В некоторых вариантах осуществления сайт-мишень находится в интроне или экзоне гена. 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 дублирована 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.

44. Таблица 9. Показания и генетические мишени, например, в печени

Заболевание	Пораженный ген
Острая интермиттирующая порфирия	HMBS
Дефицит альфа-1-антитрипсина (ААТ)	SERPINA1
Дефицит аргиназы	ARG1
Дефицит аргининосукцинатлиазы	ASL
Дефицит карбамоилфосфатсинтетазы I	CPS1
Дефицит цитрина	SLC25A13
Цитруллинемия I типа	ASS1
Синдром Криглера-Найяра (гипербилирубинемия)	UGT1A1
Болезнь Фабри	GLA
Семейная гиперхолестеринемия 4 типа (гомозиготная семейная холестеринемия)	LDLRAP1
Глутаровая ацидурия I типа	GCDH
Глутаровая ацидурия II типа (множественный дефицит ацил-СоА-дегидрогеназы)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
Болезнь накопления гликогена IV типа	GBE1
Гемофилия А	F8
Гемофилия В	F9
Наследственный гемохроматоз	HFE

Гомоцистинурия	CBS
Болезнь мочи с запахом кленового сиропа (MSUD)	Тип Ia: BCKDHA Тип Ib: BCKDHB Тип II: DBT
Метилмалоновая ацидемия (дефицит метилмалонил-КоА-мутазы)	MMUT
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS
MPS3 (синдром Санфилиппо)	Тип IIIa: SGSH Тип IIIb: NAGLU Тип IIIc: HGSNAT Тип IIIId: GNS
MPS4	Тип IVA: GALNS Тип IVB: GLB1
MPS6	ARSB
MPS7	GUSB
Дефицит орнитинтранскарбамилазы	OTC
Фенилкетонурия (дефицит фенилаланингидроксилазы)	PAH
Поликистозная болезнь печени	PCLD1: PRKCSH PCLD2: SEC63 PLCD3: ALG8 PCLD4: LRP5
Болезнь Помпе	GAA
Первичная гипероксалурия 1 типа (оксалоз)	AGXT
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
Дефицит пируваткарбоксилазы	PC
Тирозинемия I типа	FAH
Болезнь Вильсона	ATP7B

45. Таблица 10. Показания и генетические мишени для HSC

Заболевание	Пораженный ген
Адренолейкодистрофия (CALD)	ABCD1
Альфа-маннозидоз	MAN2B1
Анемия Даймонда-Блекфена	
Врожденная амегакариоцитарная тромбоцитопения	MPL
Врожденный дискератоз	TERC
Анемия Фанкони	FANC
Болезнь Гоше	GBA
Глобоидно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Гемофагоцитарный лимфогистиоцитоз	PPRF1; STX11; STXBP2; UNC13D
Злокачественный детский остеопетроз - аутосомно- рецессивный остеопетроз	Вовлечены многие гены
Метахроматическая лейкодистрофия	PSAP
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS
MPS7	GUSB
Муколипидоз II типа	GNPTAB
Болезнь Ниманна-Пика А и В	SMPD1
Болезнь Ниманна-Пика С	NPC1
Болезнь Помпе	GAA
Дефицит пируваткиназы (PKD)	PKLR
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB

46. Таблица 11. Показания и генетические мишени для ЦНС

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 окулоmotorной апраксией)	SETX
Болезнь Тея-Сакса	HEXA
X-сцепленная адренолейкодистрофия	ABCD1

47. Таблица 12. Показания и генетические мишени для глаза

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 (GUCY2D)	GUCY2D
Цистиноз, глазной ненефропатический	CTNS
"Сотовидная" ретиальная дистрофия Дойна (DHRD)	EFEMP1
Семейный окулолептоменингеальный амилоидоз	TTR
Кератит-ихтиоз-глухота (KID)	GJB2
Решетчатая дистрофия роговицы I типа	TGFBI
Макулярная дистрофия роговицы (MCD)	CHST6
Дистрофия роговицы Месманна	KRT12, KRT3
Оптическая атрофия	OPA1
Пигментный ретинит (AR)	USH2A
Пигментный ретинит (AD)	RHO
Дистрофия глазного дна Сорсби	TIMP3
Болезнь Штаргардта	ABCA4

48. Дополнительные 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, ВН4-дефицит в следствие частичного дефицита PTS, ВН4-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 моямая 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 Corrick, 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 порфибилиногенсинтазы; эпизодическая атаксия 2 типа; трихо-рино-фаланговый синдром 3 типа; прогрессирующая семейная сердечная блокада типа IV; восприимчивость к глиOME 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- Sjögren; 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; неклеточная 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; МҮН-ассоциированный полипоз; недостаточность 6-putmvoyl-тетрагидроптеринсинтазы; синдромы Алажиля 1 и 2; лимфангиомиоматоз; мышечно-глазное заболевание головного мозга; расстройства, связанные с WFSI; 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 li; 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\ха8vre; 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 Е; атаксия задних столбов спинного мозга с пигментным ретинитом; точечная хондродисплазия 1, Х-сцепленная рецессивная и 2 Х-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, 2I, 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 остеолит, 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 Х-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 поведенческими расстройствами; синдром удлиненного интервала QT; дикарбоновая аминокислотурия; 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 IIВ и IIА; семейный гиперинсулинизм; мегалэнцефальная лейкоэнцефалопатия с подкорковыми кистами тип 2а; 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; цереоидный липофусциноз нейронов 2, 6, 7 и 10; врожденная генерализованная липодистрофия 2 типа; недостаточность фруктозо-бисфосфатазы; 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; полимикрия, асимметричная, двусторонняя лобно-теменная; спинальная мышечная атрофия, дистальная, аутосомно-рецессивная, 5; метилмалоновая ацидурия из-за недостаточности метилмалонил-КоА-мутаза; 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 1А 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 Темтаи; 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;ocereбелло-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 центрального 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 интермиттирующая 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 X-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, 3а (с или без экстраокулярного поражения), 3б; синдром Прадера-Вилли; злокачественная меланома; синдром Блума; заболевание Дарье, сегментарное; многоочаговая остеолитическая нефропатия; гемохроматоз 1, 2В 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 X-хромосомой; неонатальная гипотония; синдром Орста-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 X-хромосомой) и 2; гиперхолестеринемия и гиперхолестеринемия, аутосомно-рецессивная; новообразование яичника; инфантильный ганглиозидоз GM1; синдромная X-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; дисплазмиогенемия; 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-кистевое типа, Седагатского типа, с колбочко-палочковой дистрофией, и Козловского типа; синдром ихтиоза недоношенных; синдром Стиклера 1 типа; очаговый сегментарный гломерулосклероз 5; дефицит 5-оксипролиназы; синдромальная микрофтальмия 5, 7 и 9; ювенильный полипоз/синдром наследственной геморрагической телеангиэктазии; недостаточность бутирил-КоА-дегидрогеназы; сахарный диабет взрослого возраста у молодых, тип 2; ушная отсталость, синдромальная, типа Класа-Йенсена, сцепленная с X-хромосомой; глухота, кохлеарная, с миопией и умственными нарушениями, без вестибулярного поражения, аутосомно-доминантная, X-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; гипериммуноглобулинемия 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, C1, 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, 11m; 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 X-хромосомой; недостаточность аденилаткиназы; X-сцепленный тяжелый комбинированный иммунодефицит; копропорфирия; амилоидная кардиомиопатия, связанная с транстиретином; гипокальциемия, аутосомно-доминантная 1; синдром Бругада; врожденный миастенический синдром, чувствительный к ацетазоламиду; первичная гипомagneмия; 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 удлинённого 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 удлинённого интервала QT 1, 2, 2/9, 2/5, (дигенный), 3, 5 и 5, приобретенный, предрасположенность к; синдром Андермана; дистрофия колбочек сетчатки 3Б; эритропозитическая 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; недостаточность пируваткиназы эритроцитов; дисплазия Шнекенбекена; пируэтная желудочковая тахикардия; дистальная миопатия типа Марксбери-Григгса; недостаточность УДФ-глюкозо-гексозо-1-фосфат уридилтрансферазы; внезапная сердечная смерть; синдром Ней-Лаксова 1; атрансферринемия; гиперпаратиреоз 1 и 2; злокачественная меланома кожи 1; симфалангизм, проксимальный, 1b; прогрессирующая псевдоревматоидная дисплазия; болезнь Верднига-Гоффмана; ахондрогенез 2 типа; голопрозэнцефалия 2, 3, 7 и 9; хаболечение Шиндлера, тип 1; цереброретинальная микроангиопатия с кальцинатами и кистами; гетеротаксия, висцеральная, сцепленная с X-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, X-сцепленная), 5 (допа-чувствительный тип), 10, 12, 16, 25, 26 (миоклоническая); эпифизарная дисплазия, множественная, с миопией и кондуктивной глухотой; неспецифическое нарушение проводимости сердца; бронхотические синдромы 2 и 3; нарушение биогенеза пероксисом 14В, 2А, 4А, 5В, 6А, 7А и 7В; 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 3а, 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 X-хромосомой; синдром истощения митохондриальной ДНК 11, 12 (кардиомиопатический тип), 2, 4В (тип MNGIE), 8В (тип MNGIE); брахидактилия при гипертонии; плоская роговица 2; синдром Аарскога; множественная эпифизарная дисплазия 5 или доминантная; эндотелиальная дистрофия роговицы 2 типа; недостаточность аминоксиллазы 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 X-хромосомой; синдром Перро 4; синдром Джервелла и Ланге-Нильсена 2; наследственные неполипозные колоректальные новообразования; синдром Робина, 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 р450 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 X-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 гамартонных 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 парализациями и апраксией; семейная гемиплегическая мигрень 1 и 2 типов; вентрикуломегалия при кистозной болезни почек; эластическая псевдоксантома; гомотистеинемия в следствии недостаточности MTHFR, недостаточности CBS и гомотистинурия, реагирующая на пиридоксин; дилатационная кардиомиопатия 1A, 1AA, 1C, 1G, 1BB, 1DD, 1FF, 1HH, II, 1KK, 1N, 1S, 1Y и 3B; недостаточность гуанинооксидазы 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; гипомagneмизм, 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), гидроксизин-дефицитный, 4-й тип, вариант 4-го типа и обусловлен дефицитом тенастина-X; миопия 6; плоский таз; семейный холодовой аутовоспалительный синдром 2; пороки развития сердца и крупных сосудов; болезнь фон Виллебранда типа 2М 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, мегаконияльного типа; множественные атрезии желудочно-кишечного тракта; синдром МакКьюна-Олбрайта; синдром ногтя-надколенника; синдром нейроакантоцитоза Мак-Леода; системный вариабельный иммунодефицит 9; частичная недостаточность гипоксантингуанинфосфорибозилтрансферазы; 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 II и тип III; врожденная парамиятония фон Эйленбурга; синдром Дебюкуа; недостаточность карнитинпальмитоилтрансферазы I, II, III (позднее начало) и IV (инфантильный); вторичный гипотиреоз; нижнечелюстно-лицевой дизостоз, по типу Treacher Collins, аутосомно-рецессивный; синдром Каудена 1; синдром Ли-Фраумени 1; недостаточность аспарагинсинтетазы; «сотовидная» ретиальная 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; пилломатриксомы; цианоз, переходящий неонатальный и атипичный нефропатический;

взрослая атаксия с окуломоторной апраксией; гемангиома, капиллярная инфантильная; РС-Кба; 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 X-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/BBV; муковисцидоз; дистрофия боуеновой мембраны II типа Тилля-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 X-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; С-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, 1B, 1C, 1E, 1F и 2A; судороги; рак щитовидной железы, фолликулярный; синдром летальной врожденной

контрактуры 6; дистальная наследственная моторная нейронопатия типа 2В; 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 пуриноклеозидфосфорилазы; 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 червopodobного 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; метакхроматическая лейкодистрофия; недостаточность лизосомной кислой липазы; синдром 3 М 2; врожденный амавроз Лебера 11, 12, 13, 16, 4, 7 и 9; мандибулоакральная дисплазия с липодистрофией типа А или В, атипичная; синдромы Мейера-Горлина 4; гипотрихоз 8 и 12; синдром короткого интервала QT 3; эктодермальная дисплазия 1 ib; анонихия; псевдогипопаратиреоз 1А 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 arov32; синдром Коккейна типа А; тяжелый неонатальный гиперпаратиреоз; расстройство, подобное атаксии-телеангиэктазии; синдром Пендреда; I система групп крови; семейная доброкачественная пузырьчатка; висцеральная гетеротаксия 5, аутосомная; нефрогенный несахарный диабет, нефрогенный несахарный диабет, X-сцепленный; миопатия с миниядрами 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 гамартоты 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 ПА и ПВ; 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 Тей-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-бронхио-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 I 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 Англи-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, В-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 E1-альфа; рак толстой кишки; синдром Нанса-Хорана; заболевание Сандхоффа, взрослый и детский типы; артрогрипоз, почечная дисфункция, синдром холестаза; аутосомно-рецессивная гипофосфатемическая болезнь костей; «Сотовидная» ретикулярная дистрофия Дойна; спиноцеребеллярная атаксия 14, 21, 35, 40 и 6; деменция с тельцами Леви; митохондриальное заболевание, связанное с RRM2B; Броуди миопатия; синдром мегалэнцефалии-полимикригии-полидактилии-гидроцефалии 2; синдром Ушера, типы 1, IB, ID, 1G, 2A, 2C и 2D; тип гипокальцификации и тип гипоматурии, ПА1 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 нефрофтиз; синдром ювенильного полипоза; сенсорная атаксическая невропатия, дизартрия и офтальмопарез; недостаточность 3-гидроксиацил-КоА-дегидрогеназы; карцинома парашитовидной железы; X-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 птериgiumов, 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 трихотриодистрофия; узловый 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, 31С, 38, 40, 8, обусловленный дефектом $cd3$ -дзета, с гипериммуноглобулинемией М 1 и 2 и Х-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, 4А и 4В; и семейная гипертрофическая кардиомиопатия 1, 2, 3, 4, 7, 10, 23 и 24.

49. 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 дублирована 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.

50. 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 силвиева водопровода	L1CAM
Синдром Ли	SURF1
Метахроматическая лейкодистрофия (ARSA)	ARSA
MPS 2 типа	IDS
MPS 3 типа	SGSH
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

Заболевание	Пораженный ген
Х-сцепленная адренолейкодистрофия	ABCD1

51. 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

52. 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 удлинённого интервала QT 1 типа (Романо-Уорда)	KCNQ1

53. 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
Глободно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Гемофагоцитарный лимфогистиоцитоз	PPRF1; 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 А и В	SMPD1
Болезнь Ниманна-Пика С	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
Х-сцепленная агаммаглобулинемия	BTK
Х-сцепленный SCID	IL2RG

54. Таблица 17. Заболевания почек и пораженные гены

Заболевание	Пораженный ген
Синдром Альпорта	COL4A5
Аутосомно-доминантный поликистоз почек (PKD1)	PKD1
Аутосомно-доминантный поликистоз почек (PKD2)	PKD2
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

55. Таблица 18. Заболевания печени и пораженные гены

Заболевание	Пораженный ген
Острая интермиттирующая порфирия	HMBS

Синдром Алажилля	JAG1
Дефицит альфа-1-антитрипсина	SERPINA1
Дефицит карбамоилфосфатсинтазы I	CPS1
Цитруллинемия I типа	ASS1
Синдром Криглера-Найяра	UGT1A1
Болезнь Фабри	LPL
Синдром семейной хиломикронемии	GLA
Болезнь Гоше	GBE1
GSD IV	GBA
Гем А	F8
Гем В	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 ОТС	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	ATP7B
Болезнь накопления гликогена типа 1a	G6PC
Болезнь накопления гликогена типа IIIb	AGL
Изовалериановая ацидемия	IVD
Болезнь Вольмана	LIPA

56. 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 В (SP-B) (нарушение метаболизма легочного сурфактанта 1)	SFTPB

57. 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 парамииотония	SCN4A
Миотония Томсена	CLCN1
Миопатия VCP (IBMPFD) 1	VCP

58. 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 узловый 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

59. 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 (SEQ ID NO: 1605); 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

60. 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).

61. *Таблица 23. Показания, гены и причинные патогенные мутации*

Заболевание	Ген	Патогенная мутация[#]
Ахроматопсия	CNGB3	1148delC
Дефицит альфа-1-антитрипсина	SERPINA1	E342K
Дефицит альфа-1-антитрипсина	SERPINA1	E342K
Дефицит альфа-1-антитрипсина	SERPINA1	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

Заболевание	Ген	Патогенная мутация[#]
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	Q100R
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

Заболевание	Ген	Патогенная мутация[#]
Болезнь Дарье	ATP2A2	N767S
Болезнь Дарье	ATP2A2	N767S
Буллезный узловый эпидермолиз	LAMB3	R42X
Буллезный узловый 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-СоА-дегидрогеназы	GCDH	R138G
Дефицит глутарил-СоА-дегидрогеназы	GCDH	M263V
Дефицит глутарил-СоА-дегидрогеназы	GCDH	R402W
Глициновая энцефалопатия	GLDC	A389V
Глициновая энцефалопатия	GLDC	G771R
Глициновая энцефалопатия	GLDC	T269M
Гемофилия А	F8	R2178C
Гемофилия А	F8	R550C
Гемофилия А	F8	R2169H
Гемофилия А	F8	R1985Q
Гемофилия В	F9	T342M
Гемофилия В	F9	R294Q
Гемофилия В	F9	R43Q
Гемофилия В	F9	R191H
Гемофилия В	F9	G106S
Гемофилия В	F9	A279T

Заболевание	Ген	Патогенная мутация[#]
Гемофилия В	F9	R75*
Гемофилия В	F9	R294*
Гемофилия В	F9	R379Q
Врожденный дефицит антитромбина I типа	SERPINC1	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	HPRT1	R51*
Синдром Леша-Нихена	HPRT1	R170*
Поясно-конечностная мышечная дистрофия типа 2D	SGCA	Arg77Cys
Синдром Марото-Лами (MSPVI)	ARSB	Y210C
Средиземноморский вариант дефицита G6PD	G6PD	S188D
Дефицит ацил-СоА-дегидрогеназы со средней длиной цепи	ACADM	K329E
Дефицит ацил-СоА-дегидрогеназы со средней длиной цепи	ACADM	K329E

Заболевание	Ген	Патогенная мутация [#]
Дефицит ацил-СоА-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 А	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)

Заболевание	Ген	Патогенная мутация[#]
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11	E297G
Пропионовая ацидемия	PCCB; PCCA	c.1218_1231del14ins1 2
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 А (MPSIIIA)	SGSH	R74C
Синдром Санфилиппо А (MPSIIIA)	SGSH	R245H
Синдром Санфилиппо В (MPSIIIB)	NAGLU	R297*
Синдром Санфилиппо В (MPSIIIB)	NAGLU	Y140C
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	Q3*
Серповидно-клеточная анемия	HBB	E6V

Заболевание	Ген	Патогенная мутация[#]
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
Х-сцепленная миотубулярная миопатия	MTMI	c.1261-10A>G
Х-сцепленный ретиношизис	RS1	R102W

Заболевание	Ген	Патогенная мутация [#]
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.

62. 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.

63. *Таблица 24. Показания, гены, компенсаторные редактирования и иллюстративные признаки конструкций*

Заболевание	Ген	Изменение нуклеотида [#]
Дефицит альфа-1-антитрипсина	SERPINA1	F51L
Дефицит альфа-1-антитрипсина	SERPINA1	M374I
Дефицит альфа-1-антитрипсина	SERPINA1	A348V/A347V
Дефицит альфа-1-антитрипсина	SERPINA1	K387R
Дефицит альфа-1-антитрипсина	SERPINA1	T59A
Дефицит альфа-1-антитрипсина	SERPINA1	T68A
ATTR-амилоидоз	TTR	A108V
ATTR-амилоидоз	TTR	R104H
ATTR-амилоидоз	TTR	T119M
Кистозный фиброз	CFTR	R555K
Кистозный фиброз	CFTR	F409L

Заболевание	Ген	Изменение нуклеотида[#]
Кистозный фиброз	CFTR	F433L
Кистозный фиброз	CFTR	H667R
Кистозный фиброз	CFTR	R1070W
Кистозный фиброз	CFTR	R29K
Кистозный фиброз	CFTR	R553Q
Кистозный фиброз	CFTR	I539T
Кистозный фиброз	CFTR	G550E
Кистозный фиброз	CFTR	F429S
Кистозный фиброз	CFTR	Q637R
Серповидно-клеточная анемия	HBB	A70T
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.

64. 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).

65. *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
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

Заболевание	Ген	Изменение нуклеотида [#]
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
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

Заболевание	Ген	Изменение нуклеотида [#]
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	С. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	с. -90 связывание с BCLIIA
Серповидно-клеточная анемия	HBG1/2	с. -90 связывание с BCLIIA
Серповидно-клеточная анемия	HBG1/2	С. -202 C>T, -201 C>T, -198 T>C, -197 C>T, -196 C>T, -195 C>G
Серповидно-клеточная анемия	HBG1/2	С. -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.

66. 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).

67. 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-паллидолуизийская атрофия	ATN1	CAG	Кодирующая область

Заболевание	Ген	Причинный повтор	Местоположение повтора
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	CCCCGCCCGG CG (SEQ ID NO: 1606)	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

Заболевание	Ген	Причинный повтор	Местоположение повтора
Спиноцеребеллярная атаксия 12	PPP2R2B	CAG	Промотор, 5'-UTR?
Спиноцеребеллярная атаксия 17	TBP	CAG	Кодирующая область
Синдромальная/несиндромальная X-сцепленная умственная отсталость	ARX	GCG	Кодирующая область

68. 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 одонитевого 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 одонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по одонитевому разрыву первой нити.

69. Таблица 27. Иллюстративные заболевания, ткани, гены, патогенные мутации, последовательности РНК-матрицы и последовательности gRNA для второго одонитевого разрыва

Заболевание	Ткань	Ген	Мутация	РНК-матрица	gRNA для второго одонитевого разрыва
Альфа-1-антитрипсин	Печень	SERPINA1	PiZ	TCCCCTCCAGGCCGTGCATAGTT T TAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCTcGTCGATGGTCAGCAC AGCCTTATGCACGGCCTGGA (SEQ ID NO: 1607)	TTTGGT GAACT TGACC TCGG (SEQ ID NO: 1608)

Заболевание	Ткань	Ген	Мутация	РНК-матрица	gRNA для второго однострессового разрыва
Кистозный фиброз	Легкие	CFT R	Дельта-F508	ACCATTAAGAAAATATCATGTT TTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCACCAaagATGATATTTTC TTTA (SEQ ID NO: 1609)	AaagAT GATAT TTTCTT TAA (SEQ ID NO: 1610)
Серповидно-клеточная анемия	HSC	HBV	HbS	GTAACGGCAGACTTCTCCACGTT TTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCGACTCCTGaGGAGAAGT CTGCC (SEQ ID NO: 1611)	TGGTG AGGCC CTGGG CAGGT (SEQ ID NO: 1612)
Болезнь Вильсона	Печень	ATP7 B	H1069Q	TTGGTGACTGCCACGCCCAAGTT TTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCACA _c CCCTTGGGCGTGG CAGTC (SEQ ID NO: 1613)	GGCCA GCAGT GAACA cCCCT (SEQ ID NO: 1614)
ARVC	Сердце	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTT TTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCTTCACtGAACCAGCAGT GTTCC (SEQ ID NO: 1615)	TTGGT TGAAA ATGAT TTTGT (SEQ ID NO: 1616)

Заболевание	Ткань	Ген	Мутация	РНК-матрица	gRNA для второго одонитевого разрыва
Синдром удлиненного интервала QT 1 типа	Сердце	KCN Q1	P343S	CCAGGGAAAACGCACCCACGGT TTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCTCcCAGCGGTAGGTGCC CCGTGGGTGCGTTTTTC (SEQ ID NO: 1617)	CTCCT TCTTT GCGCT CcCAG (SEQ ID NO: 1618)
Муколипидоз IV типа	ЦНС	MCO LN1	406-2A>G	GCCCTCCCCTTCTCTGCCCAGTT TTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGGACCGAGT CGGTCCGGTACtGTGGGCAGAGA AGGGG (SEQ ID NO: 1619)	TCAGG CAACG CCAGG TACtG (SEQ ID NO: 1620)

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 одонитевого 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 одонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по одонитевому разрыву первой нити.

70. Таблица 35. Иллюстративные матрицы для Gene Writing и последовательностей gRNA для второго одонитевого 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 одноразового разрыва
миотоническая дистрофия 1 типа	DMPK	NC_000019.10 (45769709..45782490, комплементарная нить)	CTG	3'-UTR	CTCGAAGGGTC CTTGTAGCCGTT TTAGAGCTAGA AATAGCAAGTT AAAATAAGGCT AGTCCGTTATCA ACTTGAAAAAG TGGGACCGAGT CGGTCCGTGAT CCCCcagcagcagc agcagcagcagcagcag cagcagcagcagcagca gcagcagcagcagcag CATTCCCGGCTA CAAGGACCCT (SEQ ID NO: 1621)	ATCA CAGG ACTG GAGC TGGG (SEQ ID NO: 1622)
миотоническая дистрофия 2 типа	CNBP	NC_000003.12 (129167827..129183896, комплементарная нить)	CCTG	Интрон 1	ACCACTGCACT CCAGCCTAGGT TTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCGTGTC TGTCTGTCTGTC	GCCT CAGC CTCC TGAG TAGC (SEQ ID NO: 1624)

					TGTCTGTCTGTC TGTCTGTCTGTC TGcctgcctgcctgcctg cctgcctgcctgcctggct gcctgtctgcctgtctgcct gcctgcctgcctgcctgcc tgctgTCTGTCTC ACTTTGTCCCCT AGGCTGGAGTG CA (SEQ ID NO: 1623)	
Синдром ломкой X- хромосом ы с умственно й отсталость ю	FMR1	NC_000023.11 (147911919..1 47951127)	CGG	5'-UTR	GGGGGCGTGCG GCAGCGCGGGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCTGCG GGCGCTCGAGG CCCAGccgcccgcgc cgccgcccgcgcgcgc cctccgcccgcgcgcgc ccgcccgcgcgcgcgcC GCTGCCGCACG (SEQ ID NO: 1625)	GCTC AGAG GCGG CCCT CCAC (SEQ ID NO: 1626)
Атаксия Фридрейх а	FXN	NC_000009.12 (69035752..69 079076)	GAA	Инtron	CAGGCGCGCGA CACCACGCCGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG	CGCT TGAG CCCG GGAG GCAG (SEQ ID NO: 1628)

					TCGGTCCAACC CAGTATCTACTA AAAAATACAAA AAAAAAAAAAAA AAgaagaagaagaaga agaaAATAAAGA AAAGTTAGCCG GGCGTGGTGTC GCGC (SEQ ID NO: 1627)	
Болезнь Хантингто на	HTT	NC_000004.12 (3074681..324 3960)	CAG	Кодирую щая область	GGCGGCTGAGG AAGCTGAGGGT TTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCAGTCC CTCAAGTCCTTC cagcagcagcagcagca gcagcagcagcagcagc agcagcagcagcagcag cagcagcaacagccgcc accgccgccgccgccgc cgccgcctcctCAGCT TCCTCAG (SEQ ID NO: 1629)	CGCT GCAC CGAC CGTG AGTT (SEQ ID NO: 1630)

Спиноцере беллярная атаксия 1	ATXN 1	NC_000006.12 (16299112..16 761490, комплементар ная нить)	CAG	Кодирую щая область	TGAGCCCCGGA GCCCTGCTGGTT TTAGAGCTAGA AATAGCAAGTT AAAATAAGGCT AGTCCGTTATCA ACTTGAAAAAG TGGGACCGAGT CGGTCCACAAG GCTGAGcagcagca gcagcagcagcagcagc agcagcagcagcatcagc atcagcagcagcagcagc agcagcagcagcagcagc cagcagcagCACCT CAGCAGGGGCTC CGGG (SEQ ID NO: 1631)	TCCA GTTC TCCG CAGA ACAC (SEQ ID NO: 1632)
-------------------------------------	-----------	---	-----	---------------------------	---	---

71. 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, обратная 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.

72. 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	CA13415 37
Сарграмостим	Иммунодепрессивные средства	CA13411 50
Секретин	Средства диагностики	
Пегилированный интерферон альфа-2b	Иммунодепрессивные средства	CA13415 67
Аспарагиназа	Антинеопластические средства	
Тиротропин альфа	Средства диагностики	US58405 66
Антигемофильный фактор	Коагулянты и тромботические средства	CA21246 90
Анакинра	Противоревматические средства	CA21419 53
Грамицидин D	Антибактериальные средства	
Внутривенный иммуноглобулин	Иммунологические факторы	
Анистреплаза	Фибринолитические средства	
Инсулин регуляр	Противодиабетические средства	
Тенектеплаза	Фибринолитические средства	CA21296 60
Менотропины	Средства для улучшения фертильности	
Интерферон гамма-1b	Иммунодепрессивные средства	US69366 95
Интерферон альфа-2a, рекомбинантный		CA21726 64
Коагуляционный фактор VIIa	Коагулянты	
Опрелвекин	Антинеопластические средства	
Палифермин	Средства против мукозита	
Глюкагон рекомбинантный	Гипогликемические средства	
Альдеслейкин	Антинеопластические средства	
Ботулотоксин типа B	Антидистонические средства	

Омализумаб	Противоаллергические средства	CA21138 13
Лютропин альфа	Средства для улучшения фертильности	US57672 51
Инсулин лиспро	Гипогликемические средства	US54749 78
Инсулин гларгин	Гипогликемические средства	US74766 52
Коллагеназа		
Расбуриказа	Супрессоры подагры	CA21759 71
Адалимумаб	Противоревматические средства	CA22434 59
Имиглюцераза	Ферментативные заместительные средства	US55498 92
Абциксимаб	Антикоагулянты	CA13413 57
Ингибитор альфа-1-протеиназы	Ингибиторы сериновых протеиназ	
Пегаспаргаза	Антинеопластические средства	
Интерферон бета-1a	Антинеопластические средства	CA13416 04
Пегадемаза крупного рогатого скота	Ферментативные заместительные средства	
Сывороточный альбумин человека	Заменители сыворотки	US67233 03
Эптифибатид	Ингибиторы агрегации тромбоцитов.	US67066 81
Иодированный сывороточный альбумин	Средства диагностики	
инфликсимаб	Антиревматические средства, противовоспалительные средства, нестероидные, дерматологические средства,	CA21062 99

	желудочно-кишечные средства и иммунодепрессивные средства	
Фоллитропин бета	Средства для улучшения фертильности	US7741268
Вазопрессин	Антидиуретические средства	
Интерферон бета-1b	Адьюванты, иммунологические и иммуносупрессивные средства	CA1340861
Интерферон альфакон-1	Противовирусные средства и иммунодепрессивные средства	CA1341567
Гиалуронидаза	Адьюванты, анестетики и пермеабилзирующие средства	
Инсулин свиной	Гипогликемические средства	
Трастузумаб	Антинеопластические средства	CA2103059
Ритуксимаб	Антинеопластические средства, иммунологические факторы и антиревматические средства	CA2149329
Базиликсимаб	Иммунодепрессивные средства	CA2038279
Myotom b	Иммунологические факторы и иммунодепрессивные средства	
Digoxin Immune Fab (овец)	Антидоты	
Ибритумомаб		CA2149329
Даптомицин		US6468967
Тозитумомаб		
Пегвисомант	Гормонозаместительные средства	US5849535
Ботулотоксин типа А	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		CA13391 98
Альглоцераза	Ферментативные заместительные средства	
Капромаб	Индикаторы, реагенты и диагностические средства	
Ларонидаза	Ферментативные заместительные средства	
Урофоллитропин	Средства для улучшения фертильности	US57670 67
Эфализумаб	Иммунодепрессивные средства	
Сывороточный альбумин	Заменители сыворотки	US67233 03
Хориогонадотропин альфа	Средства фертильности и гонадотропины	US67066 81
Антитимоцитарный глобулин	Иммунологические факторы и иммунодепрессивные средства	
Филграстим	Иммунодепрессивные средства, антинейтропенические средства и гемопозитические средства	CA13415 37
Фактор свертывания ix	Коагулянты и тромботические средства	
Бекаплермин	Средства, индуцирующие ангиогенез	CA13408 46
Агалсидаза бета	Ферментативные заместительные средства	CA22654 64
Интерферон альфа-2b	Иммунодепрессивные средства	CA13415 67

Окситоцин	Утеротоники, антиколитические средства и средства для индукции родов	
Энфувиртид	Ингибиторы слияния HIV	US64754 91
Паливизумаб	Противовирусные средства	CA21976 84
Даклизумаб	Иммунодепрессивные средства	
Бевацизумаб	Ингибиторы ангиогенеза	CA22863 30
Арцитумомаб	Средства диагностики	US84200 81
Арцитумомаб	Средства диагностики	US77901 42
Экулизумаб		CA21890 15
Панитумумаб		
Ранибизумаб	Офтальмологические средства	CA22863 30
Идурсульфаза	Ферментативные заместительные средства	
Альглокозидаза альфа	Ферментативные заместительные средства	CA24164 92
Эксенатид	Гипогликемические средства	US68727 00
Мекасермин		US56818 14
Прамлинтид		US56864 11
Галсульфаза	Ферментативные заместительные средства	
Абатацепт	Антиревматические средства и иммунодепрессивные средства	CA21105 18

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	US69770 77
Пертузумаб	Моноклональные антитела	CA23765 96
Рилонацепт	Иммунодепрессивные средства	US58440 99
Деносумаб	Средства для сохранения плотности костей и моноклональные антитела	CA22572 47
Лираглутид		US62683 43
Голимумаб	Антипсориатические средства и моноклональные антитела, и ингибитор TNF	
Белатацепт	Антиревматические средства и иммунодепрессивные средства	
Бусерелин		
Велаглуцераза альфа	Ферменты	US71382 62
Тесаморелин		US58613 79
Брентуксимаб ведотин		
Талиглуцераза альфа	Ферменты	
Белимумаб	Моноклональные антитела	
Афлиберцепт	Антинеопластические средства и офтальмологические средства	US73067 99
Аспарагиназа <i>erwinia chrysanthemi</i>	Ферменты	
Окриплазмин	Офтальмологические средства	
Глюкарпидаза	Ферменты	
Тедуглутид		US57893 79
Раксibaкумаб	Антиинфекционные средства и моноклональные антитела	

Цертолизумаб пегол	Ингибитор TNF	CA23802 98
Инсулин-изофан	Гипогликемические средства и антидиабетические средства	
Эпоэтин дзета		
Обинутуумаб	Антинеопластические средства	
Фибринолизин, он же плазмин		US32341 06
Фоллитропин альфа		
Ромиплостим	Колонiestимулирующие факторы и тромбопоэтические средства	
Луцинактант	Легочные сурфактанты	US54079 14
Тализумаб	Иммунодепрессивные средства	
Алискирен	Ингибитор ренина	
Экстракт пыльцы амброзии		
Секукинумаб	Ингибитор	US20130 202610
Соматотропин рекомбинантный	Гормонозаместительные средства	CA13264 39
Дротрекогин альфа	Антисептические средства	CA20368 94
Алефацепт	Дерматологические и иммунодепрессивные средства	
Липопротеин OspA	Вакцины	
Урокназа		US42580 30
Абареликс	Антитестостероновые средства	US59688 95
Серморелин	Гормонозаместительные средства	
Апротинин		US51985 34

Гемтузумаб озогамидин	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		
Культированные аутологичные хондроциты		
Берактант		
Блинатумомаб	Антинеопластические средства, иммунодепрессивные средства, моноклональные антитела, антинепластические 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		US20140 55370
Эволокумаб	Средства, модифицирующие липиды, простые; Сердечно-сосудистая система	
Концентрат фибриногена (человека)		
Филграстим-sndz		
Внутренний фактор желудка		
Иммуноглобулин к гепатиту В		
Кальцитонин человека		
Человеческий противостолбнячный иммуноглобулин		
Человеческий иммуноглобулин к вирусу бешенства		
Человеческий иммуноглобулин Rho(D)		
Гиалуронидаза (человеческая рекомбинантная)		US77674 29
Идаруцизумаб	Антикоагулянт	
Человеческий иммуноглобулин	Иммунологические факторы; иммунодепрессивные средства; Противоинфекционные средства	
Ведолизумаб	Иммунодепрессивное средство; антинеопластическое средство	US20121 51248
Устекинумаб	Дерматологическое средство, иммунодепрессивное средство; антинеопластическое средство	
Туроктоког альфа		

Очищенное производное белка	туберкулиновое	
Симоктоког альфа	Антигеморрагические средства: фактор свертывания крови VIII	
Силтуксимаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства	US76121 82
Себелипаза альфа	Ферменты	
Сакрозидаза	Ферменты	
Рамуцирумаб	Антинеопластические средства и иммуномодулирующие средства	US20130 67098
Концентрат протромбинового комплекса		
Порактан альфа	Легочные сурфактанты	
Пембролизумаб	Антинеопластические средства и иммуномодулирующие средства	US20121 35408
Пегинтерферон бета-1a		
Офатумумаб	Антинеопластические средства и иммуномодулирующие средства	US83378 47
Обилтоксаксимаб		
Ниволумаб	Антинеопластические средства и иммуномодулирующие средства	US20131 73223
Нецитумумаб		
Метрелептин		US20070 099836
Метоксиполиэтиленгликоль-эпоэтин бета		
Меполизумаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства, ингибиторы интерлейкина	US20081 34721
Иксекизумаб		
Инсулин свиной	Гипогликемические средства, антидиабетические средства	
Инсулин Деглудек		

Инсулин из говядины		
Тиреоглобулин	Гормональная терапия	US50990 01
Человеческий противосибирязвенный иммуноглобулин	Производное плазмы крови	
Антиингибиторный коагулянтный комплекс	Факторы свертывания крови, антигемофильное средство	
Антитимоцитарный глобулин (лошадиный)	Антитело	
Антитимоцитарный глобулин (кролика)	Антитело	
Бродалумаб	Антинеопластические средства и иммуномодулирующие средства	
Ингибитор эстеразы C1 (рекомбинантный)	Кровь и органы кроветворения	
Каникумаб	Антинеопластические средства и иммуномодулирующие средства	
Хорионический гонадотропин (человека)	Гормоны	US67066 81
Хорионический гонадотропин (Рекомбинантный)	Гормоны	US57672 51
Человеческий фактор свертывания крови X	Факторы свертывания крови	
Динутуксимаб	Антитело, иммунодепрессивное средство, антинеопластическое средство	US20140 170155
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	CA21493 29
Ленограстим	Антинеопластические средства и иммуномодулирующие средства	

Пеглотиказа	Ферменты	
Протамина сульфат	Антагонисты гепарина, гематологические средства	
Белок S человека	Антикоагулянтный белок плазмы	
Сипулеуцель-Т	Антинеопластические средства и иммуномодулирующие средства	US81531 20
Соматропин рекомбинантный	Гормоны, гормонозаместители и антагонисты гормонов	CA13264 39, CA22525 35, US52887 03, US58497 00, US58497 04, US58980 30, US60042 97, US61528 97, US62350 04, US68996 99
Сусоктоког альфа	Факторы свертывания крови, Антигеморрагические средства	
Тромбомодулин альфа	Антикоагулянтное средство, антиагрегантное средство	

73. 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	TNF α	Болезнь Крона
Трастузумаб	HER2	Рак молочной железы
Алемтузумаб	CD52	Хронический миелоидный лейкоз
Адалимумаб	TNF α	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	TNF α	Болезнь Крона
Устекинумаб	IL-12/23	Псориаз
Канакинумаб	IL-1 β	Синдром Макла-Уэльса
Голимумаб	TNF α	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 меланوما
Брентуксимаб ведотин	CD30	Лимфома Ходжкина, системная анапластическая крупноклеточная лимфома
Пертузумаб	HER2	Рак молочной железы
Трастузумаб эмтанзин	HER2	Рак молочной железы

Раксibaкyмaб	PA <i>B. anthracis</i>	Сибирeязвeннaя инфeкция
Обинyтyзyмaб	CD20	Хрoничeский лимфоцитарный лейкоз
Силтyксимaб	IL-6	Болeзнъ Кастлeманa
Рамyцирyмaб	VEGFR2	Рак жeлyдкa
Вeдoлизyмaб	Интeгрин $\alpha4\beta7$	Нeспeцифичeский язвeнный колит, болeзнъ Крoнa
Блинaтyмoмaб	CD19, CD3	Острый лимфoбластный лейкоз
Нивoлyмaб	PD-1	Мелaнoмa, нeмeлкoккeтoчный рак лeгкoгo
Пeмбpoлизyмaб	PD-1	Мелaнoмa
Идaруцизyмaб	Дaбигaтpaн	Устрaнeниe дaбигaтpaн-индyцирoвaннoй антикoагуляци
Нeцитyмyмaб	EGFR	Нeмeлкoккeтoчный рак лeгкoгo
Динyтyксимaб	GD2	Нeйрoбластoмa
Сeкyкинyмaб	IL-17 α	Псoриaз
Мeпoлизyмaб	IL-5	Тяжeлaя эoзинoфильнaя астмa
Алирoкyмaб	PCSK9	Высoкoe сoдeржaниe хoлeстeринa
Эвoлoкyмaб	PCSK9	Высoкoe сoдeржaниe хoлeстeринa
Дaрaтyмyмaб	CD38	Мнoжeствeннaя миeлoмa
Элoтyзyмaб	SLAMF7	Мнoжeствeннaя миeлoмa
Иксeкизyмaб	IL-17 α	Псoриaз
Рeслизyмaб	IL-5	Астмa
Оларaтyмaб	PDGFR α	Сaркoмa мягких ткaнeй
Бeзлoтoксyмaб	Энтeрoтoксин В <i>Clostridium difficile</i>	Пpeдyпpeждeниe рeцидивa инфeкции <i>Clostridium difficile</i>
Атeзoлизyмaб	PD-L1	Рак мoчeвoгo пyзыря
Обилтoксaксимaб	PA <i>B. anthracis</i>	Пpeдyпpeждeниe ингaляциoннoй фoрмy сибирскoй язвы
Инoтyзyмaб oзoгaмицин	CD22	Острый лимфoбластный лейкоз
Брoдaлyмaб	IL-17R	Блaшeчный псoриaз
Гyсeлькyмaб	IL-23, p19	Блaшeчный псoриaз
Дyпилyмaб	IL-4R α	Атoпичeский дeрмaтит
Сaрилyмaб	IL-6R	Рeвмaтoидный артрит
Авeлyмaб	PD-L1	Кaрцинoмa из кeтoк Мeркeлa

Окрелизумаб	CD20	Рассеянный склероз
Эмицизумаб	Фактор IXa, X	Гемофилия А
Бенрализумаб	IL-5R α	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	IFN γ	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	CD79 β	Диффузная крупноклеточная В-клеточная лимфома
Бролуцизумаб	VEGF-A	Макулярная дегенерация
Кризанлизумаб	P-селектин	Серповидно-клеточная анемия

74.

75. Способы модификации растений

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.

76. А. Доставка в растение

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 подвита *japonica*) или в часть растения (например, в каллус растения) посредством агробактерий. В некоторых вариантах осуществления системы и способы, описанные в данном документе, могут применяться в растениях путем замены гена растения (например, гена гигромицинофосфотрансферазы (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.

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).

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, имеющих 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.

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.

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.

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.

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.

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%).

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).

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.

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.

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.

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.

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.

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.

77. В. Способы применения

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.

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.

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.

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 как 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.

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.

78. C. Растения

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.

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, ликопиты, 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, клевер, 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, канولا, клещевина, кукуруза, хлопчатник, крамбе, лен, семена 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.

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 esculentum*, *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 rimpaii*, *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.

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.

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.

79. 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 неделяющуюся клетку, например, неделящийся фибробласт или неделяющуюся 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 PCT/US2019/048607, который включен в данный документ посредством ссылки. Специалисту в данной области будет понятно, что компоненты системы Gene Writer™ могут 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 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.

Для доставки можно использовать различные наночастицы, такие как липосомы, липидные наночастицы, катионные липидные наночастицы, ионизируемые липидные наночастицы, полимерные наночастицы, наночастицы золота, дендримеры, наночастицы циклодекстрина, мицеллы или комбинации вышеизложенного.

Липидные наночастицы являются 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 (см., например, патентную заявку WO 2020014209, идеи которой, касающиеся разработки, получения и применения фузосом, включены в данный документ посредством ссылки).

В некоторых вариантах осуществления белковый(белковые) компонент(компоненты) системы Gene Writing™ могут быть предварительно ассоциированы с нуклеиновой кислотой-матрицей (например, РНК-матрицей). Например, в некоторых вариантах осуществления полипептид Gene Writer™ может быть сначала объединен с нуклеиновой кислотой-матрицей (например, РНК-матрицей) с образованием

рибонуклеопротеинового (RNP) комплекса. В некоторых вариантах осуществления RNP может быть доставлен в клетки посредством, например, трансфекции, нуклеофекции, вируса, везикулы, LNP, экзосомы, фузосомы.

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).

80. 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 одонитевой или содержит одонитевой сегмент, например, представляет собой одонитевую ДНК или содержит одонитевой 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 одонитевой или содержит одонитевой сегмент, например, представляет собой одонитевую ДНК или содержит одонитевой 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 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) (cactcctccccatcctctcctctgtccctctgtccctctgaccctgcactgtcccagcacc (SEQ ID NO: 1633)) или орозомукоида 1 (ORM1) (caggacacagccttgatcaggacagagacttgggggccatcctgcccctccaacccgacatgtgtacctcagcttttccctcacttgcat

caataaagcttctgtgtttggaacagctaa (SEQ ID NO: 1634)) (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 (SEQ ID NO: 1603), и/или 3'-UTR содержит UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCC AGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA

(SEQ ID NO: 1604), 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 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'-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAUAUAAGAGCCACC-3' (SEQ ID NO: 1603). В некоторых вариантах осуществления 3'-UTR мРНК-компонента (или транскрипта, полученного из ДНК-компонента) системы содержит последовательность 5'-UGAUAUAAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA-3' (SEQ ID NO: 1604). 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 Т 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 Т7, Т3 или SP6. Вышеуказанная 5'-UTR начинается с GGG, что является подходящим сайтом начала для оптимизации транскрипции с применением РНК-полимеразы Т7. 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.

81. 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 анелловир, применение которого описано в WO2018232017A1, которая включена в данный документ посредством ссылки во всей своей полноте.

82. 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 со следующим: 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 хелперными 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 не кодирующие 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 спейсерную ДНК 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 одностранных 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, фланкируемые 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 шпильчатые структуры, 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, высвобождении из генома хозяина или плазмиды и инкапсулировании 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 шпильчатые структуры. См., например, WO2012123430.

Общепринято, что вирионы AAV с капсидами получают путем введения плазмиды или плазмид, кодирующих геном gAAV или scAAV, белки Rep и белки Cap (Grimm et al., 1998). При введении этих хелперных плазмид 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 паковую клетку в сочетании с хелперными 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, обратную комплементарную необходимой последовательности нуклеиновой кислоты, так что эти два компонента могут располагаться внахлест и гибридизоваться 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 обратных комплементарных модуля, которые могут гибридизоваться 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 линейную дуплексную ДНК с закрытыми 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 концевой 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 сеDNA 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 сеDNA не содержит какого-либо капсидного белка, например, не упакована в инфекционную частицу AAV. В некоторых вариантах осуществления сеDNA составляют в LNP (см., например, WO2019051289A1).

В некоторых вариантах осуществления сеDNA-вектор состоит из двух самокомплементарных последовательностей, например, асимметричных, или симметричных, или по сути симметричных ITR, определенных в данном документе, фланкирующих указанную кассету экспрессии, где сеDNA-вектор не ассоциирован с капсидным белком. В некоторых вариантах осуществления сеDNA-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 концевой разрешения (trs) AAV или функциональный вариант RBE. См., например, WO2019113310.

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 эпизотически 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 крупных физиологических регуляторных элементов могут быть осуществлены, например, посредством разделения белка(ов), который(ые) 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 инъектировать 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, AAV12, 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 <i>et al.</i> , <i>Mol. Ther.</i> 18 , 118-25 (2010) 2. Ginn <i>et al.</i> , <i>JHEP Reports</i> , 100065 (2019) 3. Paulk <i>et al.</i> , <i>Mol. Ther.</i> 26 , 289-303 (2018). 4. L. Lisowski <i>et al.</i> , <i>Nature</i> . 506 , 382-6 (2014). 5. L. Wang <i>et al.</i> , <i>Mol. Ther.</i> 23 , 1877-87 (2015). 6. Hausl <i>Mol Ther</i> (2010)

Легкое	AAV (AAV4, AAV5, AAV6 ¹ , AAV9, H22 ²) Аденовирус (Ad5, Ad3, Ad21, Ad14) ³	1. Duncan <i>et al.</i> , <i>Mol Ther Methods Clin Dev</i> (2018) 2. Cooney <i>et al.</i> , <i>Am J Respir Cell Mol Biol</i> (2019) 3. Li <i>et al.</i> , <i>Mol Ther Methods Clin Dev</i> (2019)
Кожа	AAV (AAV6 ¹ , AAV-LK19 ²)	1. Petek <i>et al.</i> , <i>Mol. Ther.</i> (2010) 2. L. Lisowski <i>et al.</i> , <i>Nature</i> . 506 , 382-6 (2014).
HSC	Аденовирус (HAd5/35 ⁺⁺)	Wang <i>et al.</i> <i>Blood Adv</i> (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 обнаружения. 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 (гНСП) в фармацевтической композиции меньше или равняется 100 нг/мл гНСП на 1×10^{13} vg/мл, например, меньше или равняется 40 нг/мл гНСП на 1×10^{13} vg/мл или 1-50 нг/мл гНСП на 1×10^{13} vg/мл. В некоторых вариантах осуществления фармацевтическая композиция содержит менее 10 нг гНСП на $1,0 \times 10^{13}$ vg или менее 5 нг гНСП на $1,0 \times 10^{13}$ vg, менее 4 нг гНСП на $1,0 \times 10^{13}$ vg или менее 3 нг гНСП на $1,0 \times 10^{13}$ 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 $1,2 \times 10^6$ 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 $5,0 \times 10^5$ пг на 1×10^{13} vg, менее $2,0 \times 10^5$ пг на $1,0 \times 10^{13}$ vg, менее $1,1 \times 10^5$ пг на $1,0 \times 10^{13}$ vg, менее $1,0 \times 10^5$ пг hcDNA на $1,0 \times 10^{13}$ vg, менее $0,9 \times 10^5$ пг hcDNA на $1,0 \times 10^{13}$ vg, менее $0,8 \times 10^5$ пг hcDNA на $1,0 \times 10^{13}$ vg или любое промежуточное значение концентрации.

В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции меньше или равняется $1,7 \times 10^5$ пг/мл на $1,0 \times 10^{13}$ vg/мл, или 1×10^5 пг/мл на $1,0 \times 10^{13}$ vg/мл, или $1,7 \times 10^6$ пг/мл на $1,0 \times 10^{13}$ vg/мл. В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции составляет менее $10,0 \times 10^5$ пг на $1,0 \times 10^{13}$ vg, менее $8,0 \times 10^5$ пг на $1,0 \times 10^{13}$ vg или менее $6,8 \times 10^5$ пг на $1,0 \times 10^{13}$ vg. В вариантах осуществления фармацевтическая композиция содержит менее 0,5 нг на $1,0 \times 10^{13}$ vg, менее 0,3 нг на $1,0 \times 10^{13}$ vg, менее 0,22 нг на $1,0 \times 10^{13}$ vg или менее 0,2 нг на $1,0 \times 10^{13}$ vg или любое промежуточное значение концентрации бычьего сывороточного альбумина (BSA). В вариантах осуществления содержание бензоназы в фармацевтической композиции составляет менее 0,2 нг на $1,0 \times 10^{13}$ vg, менее 0,1 нг на $1,0 \times 10^{13}$ vg, менее 0,09 нг на $1,0 \times 10^{13}$ vg, менее 0,08 нг на $1,0 \times 10^{13}$ 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 $5,0 \times 10^{13}$ vg/мл, от $1,2$ до $3,0 \times 10^{13}$ vg/мл или от $1,7$ до $2,3 \times 10^{13}$ 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 осмоляльность 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 $5,0 \times 10^{13}$ vg/мл, от 1,0 до $4,0 \times 10^{13}$ vg/мл, от 1,5 до $3,0 \times 10^{13}$ vg/мл или от 1,7 до $2,3 \times 10^{13}$ 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 $1,0 \times 10^{13}$ vg, менее чем приблизительно 30 пг/г (ppm) цезия, приблизительно 20-80 ppm полоксамера 188, менее чем приблизительно 0,22 нг BSA на $1,0 \times 10^{13}$ vg, менее чем приблизительно $6,8 \times 10^5$ пг остаточной плазмидной ДНК на $1,0 \times 10^{13}$ vg, менее чем приблизительно $1,1 \times 10^5$ пг остаточной hcDNA на $1,0 \times 10^{13}$ vg, менее чем приблизительно 4 нг гНСП на $1,0 \times 10^{13}$ 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 $1,7 \times 10^{13}$ до $2,3 \times 10^{13}$ vg/мл, инфекционный титр от приблизительно $3,9 \times 10^8$ до $8,4 \times 10^{10}$ ME на $1,0 \times 10^{13}$ vg, содержание общего белка приблизительно 100-300 пг на $1,0 \times 10^{13}$ vg, средняя выживаемость > 24 дней у мышей A7SMA с дозой вирусного вектора приблизительно $7,5 \times 10^{13}$ 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 $\pm 20\%$, $\pm 15\%$, $\pm 10\%$ или $\pm 5\%$ 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](https://doi.org/10.1038/s41573-019-0012-9), включая таблицу 1, которая включена посредством ссылки во всей своей полноте.

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-С может быть слит с С-концевой частью второго домена, описанного в данном документе, для присоединения N-концевой части к С-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 С-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-С). Например, у цианобактерий DnaE, каталитическая α-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-С".

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 С-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-С). 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 Ppr8 (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(1):446-461, включенном в данный документ посредством ссылки. Способы разработки и применения интеинов известны из уровня техники и описаны, например, в WO2020051561, WO2014004336, 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 парDNAbp. 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, нуклеазный 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
GGTAACATTTATACTCAGCCAGTTGCCAGTGGCACGACCGGGGAGAGCAGGAAGT
ATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAAT
TTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGG
ACCTCATGCGAGTTGACAACCTTCCTAAT (SEQ ID NO: 1637)

Белок интеин-N DnaE:

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEV
FEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1638)

ДНК интеина-C DnaE:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATA
TTGGAGTCGAAAGAGATCACAACCTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT
(SEQ ID NO: 1639)

Интеин-C:

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1640)

ДНК Cfa-N:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTAT
TGGAAAGATTGTCTGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATG
GTTTCGTTTACACACAGCCCATTTGCTCAATGGCACAATCGCGGCGAACAAGAAGTAT
TTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTC
ATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGA
TCTCAAACAAGTGGATGGATTG CCA (SEQ ID NO: 1641)

Белок Cfa-N:

CLSYDTEILTVEYGFLLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVF
EYCLEDGSIIRATKDHKFMTTDQGMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 1642)

ДНК Cfa-C:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGG
AAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGG
AGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ
ID NO: 1643)

Белок Cfa-C:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLKNGLVASN
(SEQ ID NO: 1644)

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(тетрадеканоиокси)пропил-1-0-(ω-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 стерина, 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 стерина включают фитостерины, в том числе описанные в Eygeris et al (2020), [dx.doi.org/10.1021/acs.nanolett.0c01386](https://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 быть катионным липидом, ионизируемым катионным липидом, например, катионным липидом, который может существовать в положительно заряженной или нейтральной форме в зависимости от pH, или аминокислотным липидом, 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 аминокислотных групп, которые несут положительный заряд. В некоторых вариантах осуществления липидная частица содержит катионный липид в составе с одним или несколькими нейтральными липидами, ионизируемыми аминокислотными 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 pK_a (например, более высоким, чем первое эффективное значение pK_a) 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), (iii), (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.

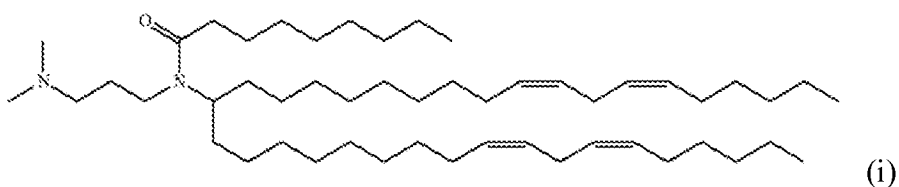
Иллюстративные ионизируемые липиды, которые можно применять в составах на основе липидных наночастиц, включают без ограничения липиды, перечисленные в таблице 1 документа WO2019051289, включенного в данный документ посредством ссылки. Дополнительные иллюстративные липиды включают без ограничения одну 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-с 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 WO2013/016058; A из WO2012/162210; 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 WO2009/132131; 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.

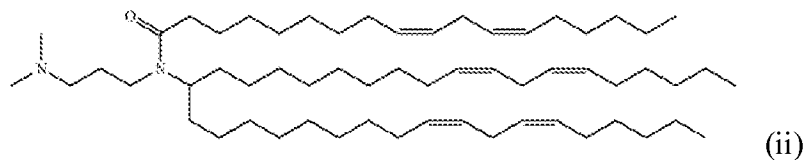
В некоторых вариантах осуществления ионизируемый липид представляет собой MC3 (6Z,9Z,28Z,31Z)-гептатриаконта-6,9,28,31-тетраен-19-ил-4-(диметиламино)бутаноат (DLin-MC3-DMA или MC3), например, описанный в примере 9 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей 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)метилпропилотдаека-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[а]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).

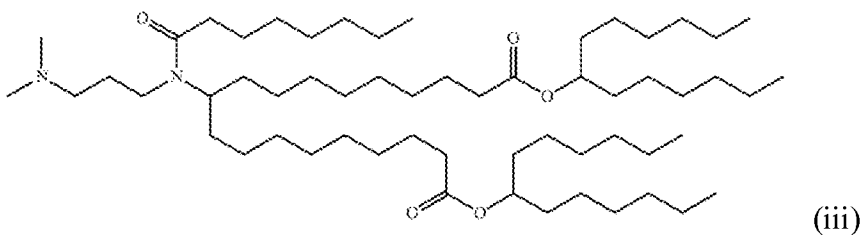
Некоторые неограничивающие примеры липидных соединений, которые можно применять (например, в комбинации с другими липидными компонентами) для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), включают



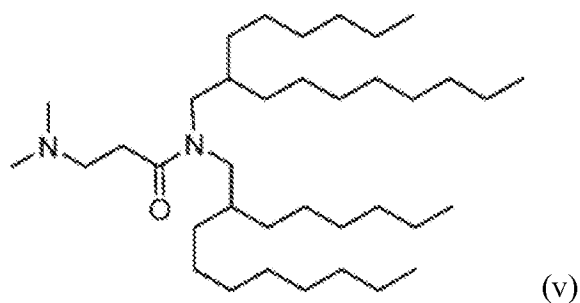
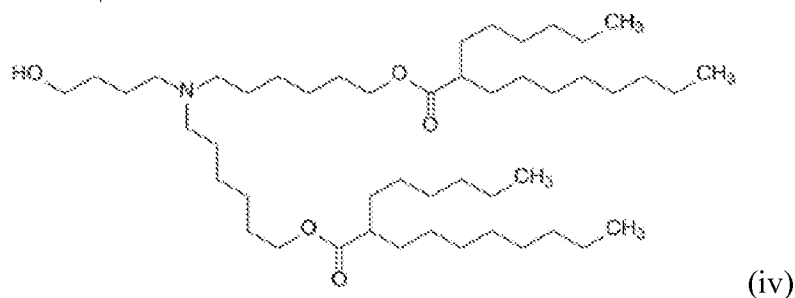
В некоторых вариантах осуществления LNP, содержащую формулу (i), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



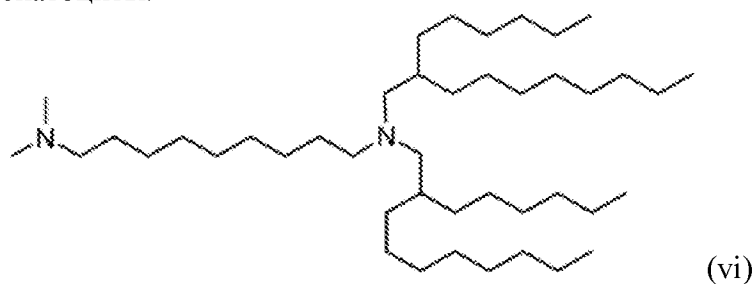
В некоторых вариантах осуществления LNP, содержащую формулу (ii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



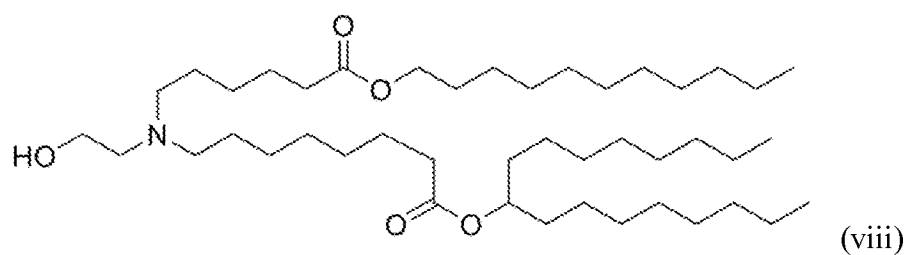
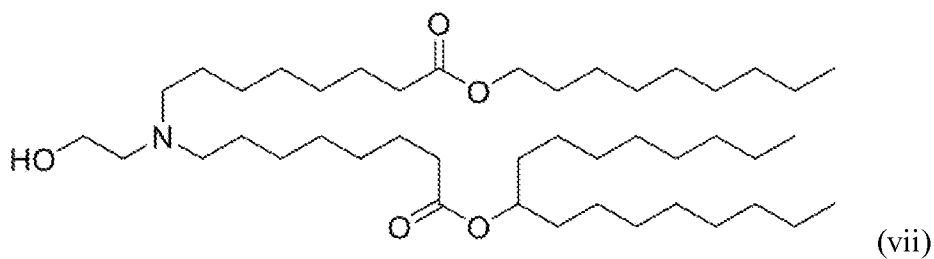
В некоторых вариантах осуществления LNP, содержащую формулу (iii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



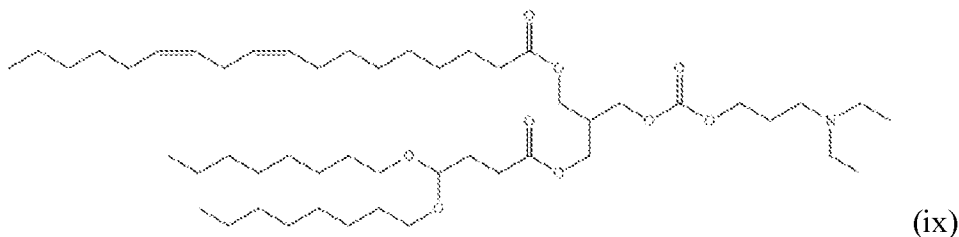
В некоторых вариантах осуществления LNP, содержащую формулу (v), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



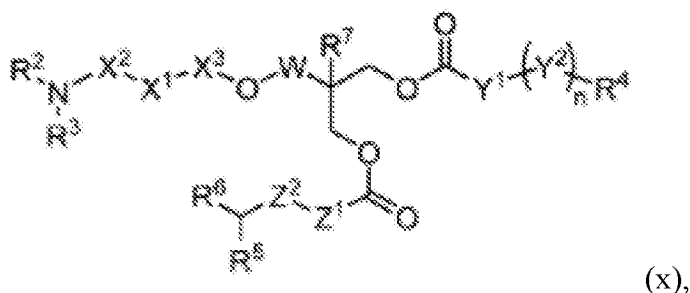
В некоторых вариантах осуществления LNP, содержащую формулу (vi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



В некоторых вариантах осуществления LNP, содержащую формулу (viii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

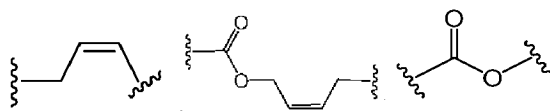


В некоторых вариантах осуществления LNP, содержащую формулу (ix), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



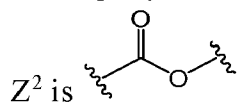
где

X^1 представляет собой O, NR^1 или прямую связь, X^2 представляет собой C2-5-алкилен, X^3 представляет собой C(=O) 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),

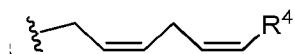
n равняется 0-3, R^4 представляет собой C1-15-алкил, Z^1 представляет собой C1-6-алкилен или прямую связь,



(в любой ориентации) или отсутствует, при условии, что если Z^1 представляет собой прямую связь, то Z^2 отсутствует;

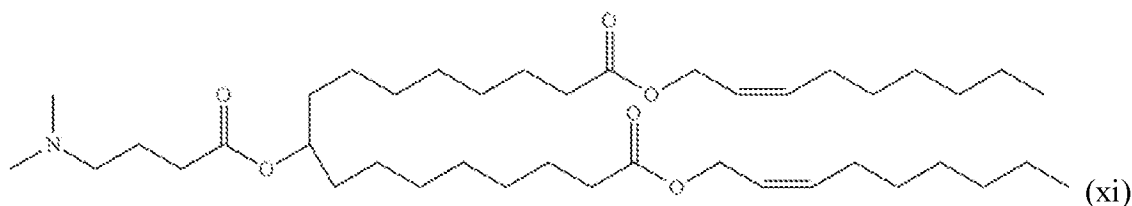
R^5 представляет собой C5-9-алкил или C6-10-алкокси, R^6 представляет собой C5-9-алкил или C6-10-алкокси, W представляет собой метилен или прямую связь, и R^7 представляет собой H или Me или их соль, при условии, что если R^3 и R^2 представляют

с собой C2-алкилы, X^1 представляет собой O, X^2 представляет собой линейный C3-алкилен, X^3 представляет собой C(=O), Y^1 представляет собой линейный C1-алкилен, $(Y^2)_n-R^4$ представляет собой

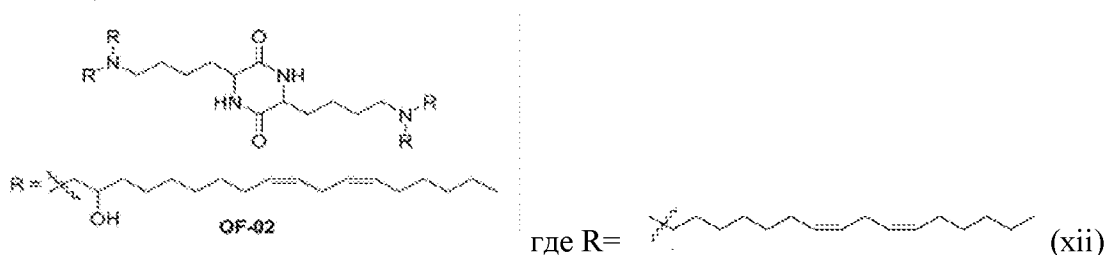


, R^4 представляет собой линейный C5-алкил, Z^1 представляет собой C2-алкилен, Z^2 отсутствует, W представляет собой метилен, и R^7 представляет собой H, то R^5 и R^6 не представляют собой Cx-алкокси.

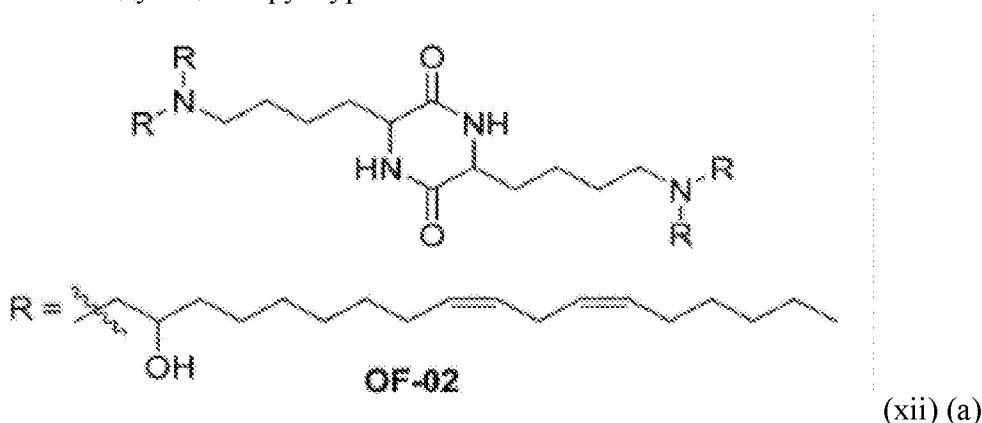
В некоторых вариантах осуществления LNP, содержащую формулу (xii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

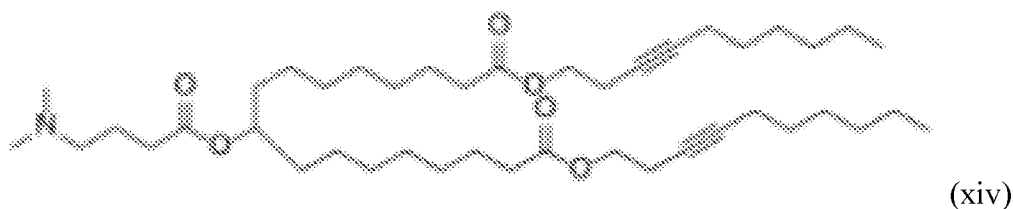
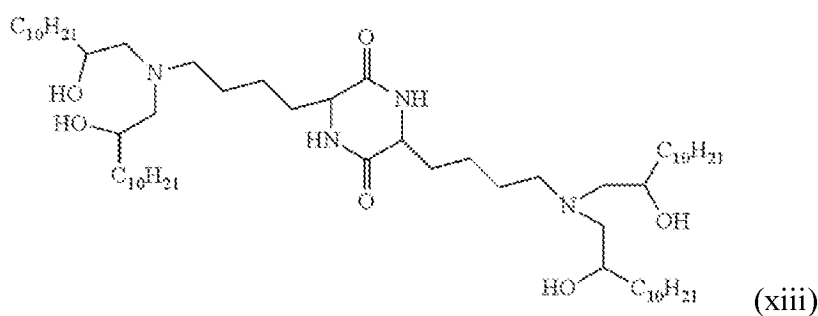


В некоторых вариантах осуществления LNP, содержащую формулу (xi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

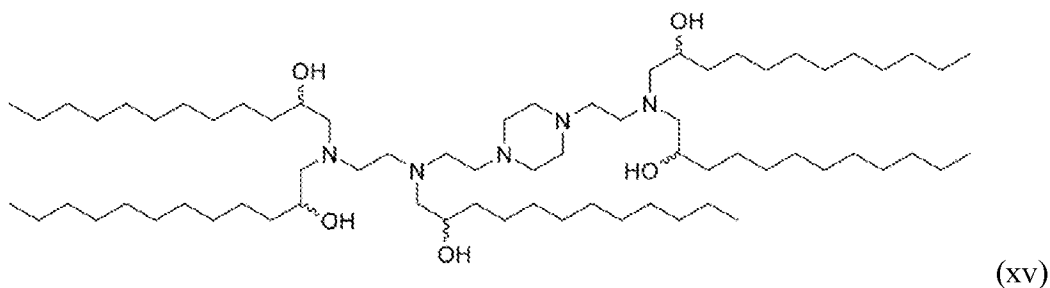


В некоторых вариантах осуществления липид формулы (xii) может быть представлен следующей структурой:



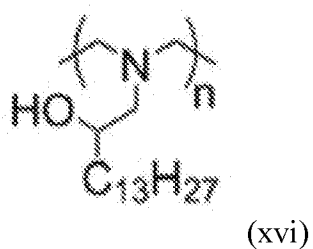


В некоторых вариантах осуществления LNP содержит соединение формулы (xiii) и соединение формулы (xiv).

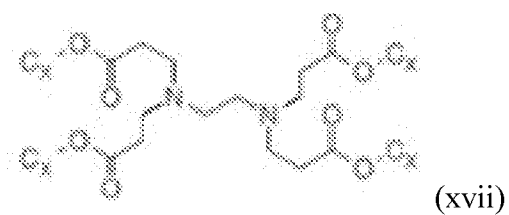


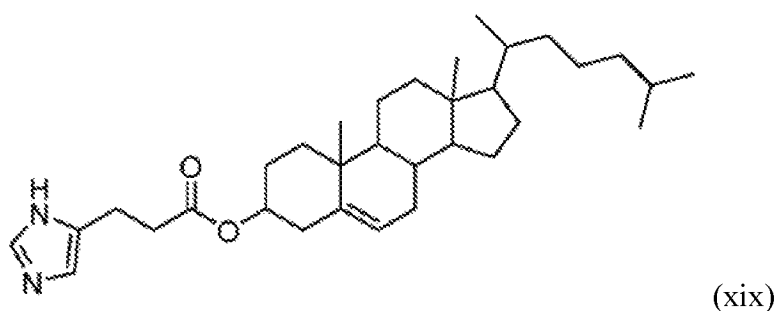
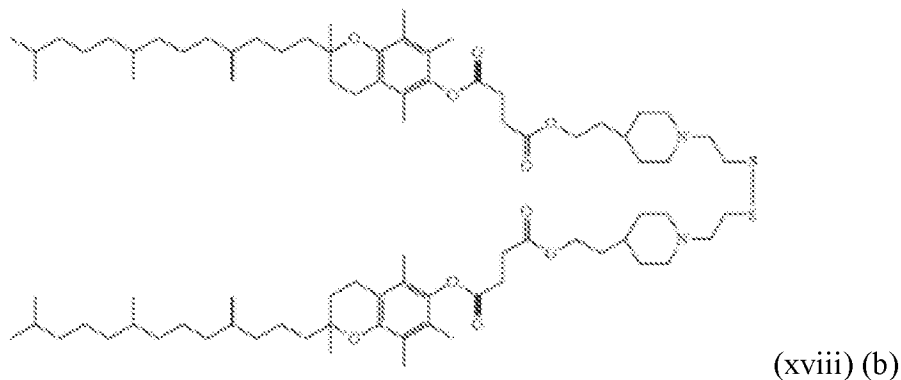
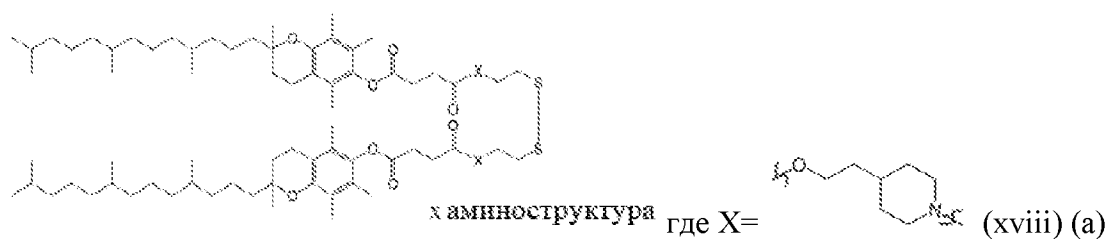
В некоторых вариантах осуществления LNP, содержащую формулу (xv), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

Сердцевинный PEI₆₀₀

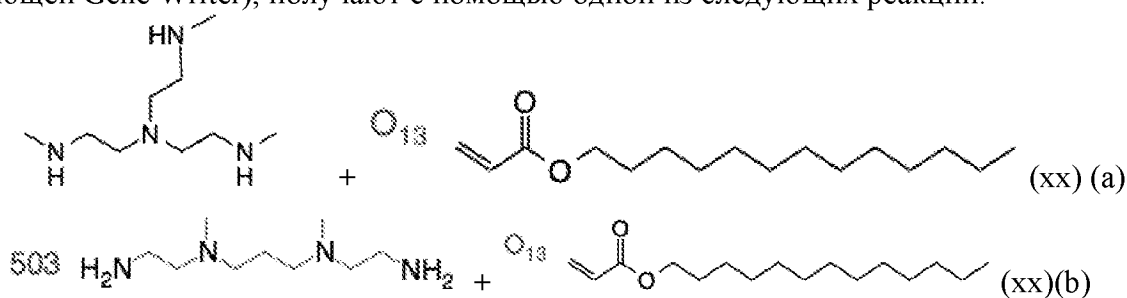


В некоторых вариантах осуществления LNP, содержащую состав формулы (xvi), применяют для доставки композиции Gene Writer, описанной в данном документе, в эндотелиальные клетки легкого.





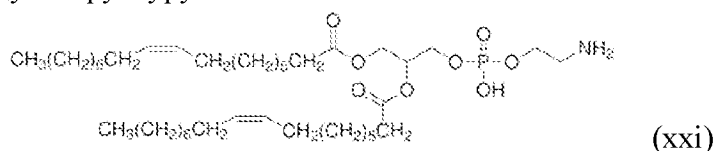
В некоторых вариантах осуществления липидное соединение, применяемое для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), получают с помощью одной из следующих реакций:



Иллюстративные некатионные липиды включают без ограничения дистеароил-sn-глицерофосфоэтаноламин, дистеароилфосфатидилхолин (DSPC), диолеоилфосфатидилхолин (DOPC), дипальмитоилфосфатидилхолин (DPPC), диолеоилфосфатидилглицерин (DOPG), дипальмитоилфосфатидилглицерин (DPPG), диолеоилфосфатидилэтаноламин (DOPE), пальмитоилолеоилфосфатидилхолин (POPC), пальмитоилолеоилфосфатидилэтаноламин (POPE), 1,2-диолеоил-sn-глицеро-3-фосфоэтаноламин (DOPE), диолеоилфосфатидилэтаноламин-4-(N-малеимидометил)-циклогексан-1-карбоксилат (DOPE-Mal), дипальмитоилфосфатидилэтаноламин (DPPE),

димиристоилфосфоэтаноламин (DMPE), дистеароилфосфатидилэтаноламин (DSPE), монометилфосфатидилэтаноламин (такой как 16-О-монометил-PE), диметилфосфатидилэтаноламин (такой как 16-О-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](https://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:



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.

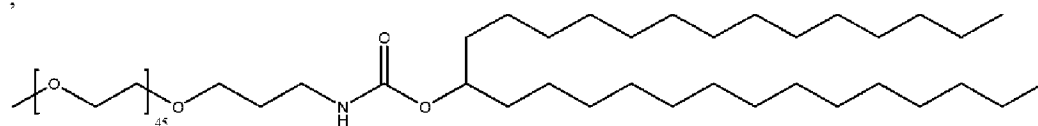
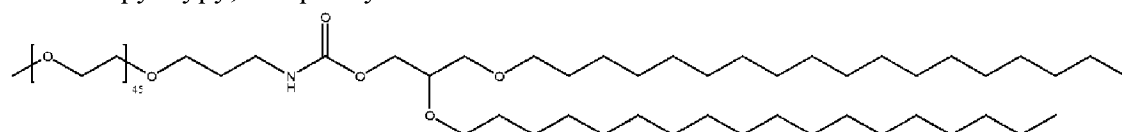
В некоторых аспектах липидная наночастица может дополнительно содержать компонент, такой как стерин, для обеспечения целостности мембраны. Одним иллюстративным стеринном, который можно применять в липидной наночастице, является холестерин и его производные. Неограничивающие примеры производных холестерина включают полярные аналоги, такие как 5 α -холестанол, 5 β -копростанол, холестерил-(2'-гидроксид)-этиловый эфир, холестерил-(4'-гидроксид)-бутиловый эфир и 6-кетохолестанол; неполярные аналоги, такие как 5 α -холестан, холестенон, 5 α -холестанон, 5 β -холестанон и холестерилдеканат; и их смеси. В некоторых вариантах осуществления производное холестерина представляет собой полярный аналог, например, холестерил-(4'-гидроксид)бутиловый эфир. Иллюстративные производные холестерина описаны в публикации согласно РСТ 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.

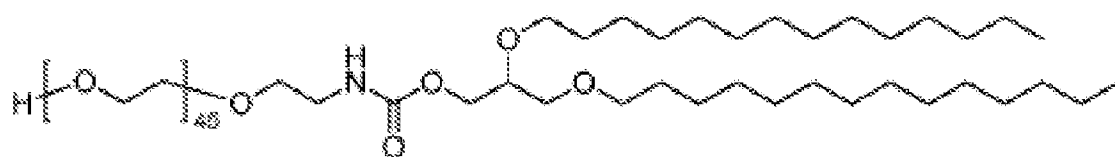
В некоторых вариантах осуществления липидная наночастица может содержать полиэтиленгликоль (PEG) или конъюгированную молекулу липида. Обычно они применяются для ингибирования агрегации липидных наночастиц и/или обеспечения стерической стабилизации. Иллюстративные конъюгированные липиды включают без ограничения конъюгаты PEG-липид, конъюгаты полиоксазолин (POZ)-липид, конъюгаты полиамид-липид (такие как конъюгаты АТТА-липид), конъюгаты катионный полимер-липид (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), PEG-сукцинатдиацилглицерин (PEG-DAG) (например, 4-0-(2',3'-ди(тетрадеканилокси)пропил-

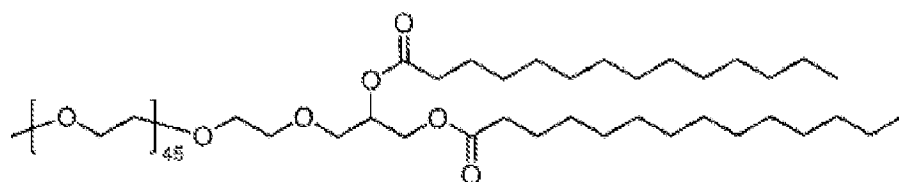
1-0-(w-метокси(полиэтокси)этил)бутандиоат (PEG-S-DMG)), PEG-диалкоксипропилкарбам, натриевую соль N-(карбонилметоксиполиэтиленгликоль 2000)-1,2-дистеароил-sn-глицеро-3-фосфоэтаноламина, 1,2-димиристоил-sn-глицерин, метоксиполиэтиленгликоль (DMG-PEG-2K) или их смесь. Дополнительные иллюстративные конъюгаты 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:



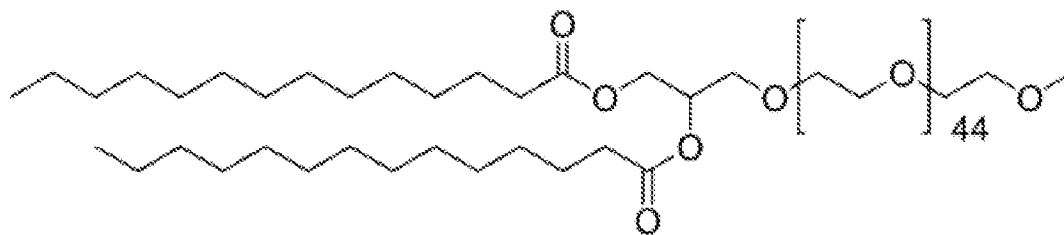
(xxii),



(xxiii),



(xxiv) и



(xxv).

В некоторых вариантах осуществления вместо PEG-липида также можно применять 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 АТТА-липид) и конъюгаты катионный полимер-липид (GPL).

Иллюстративные конъюгированные липиды, т. е. конъюгаты PEG-липид, конъюгаты (POZ)-липид, конъюгаты АТТА-липид и катионный полимер-липид, описаны в патентных заявках согласно РСТ 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, некаатионного 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% некаатионного 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% некаатионного 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% неcatiонного 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% неcatiонного 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% неcatiонного 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% неcatiонного 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% неcatiонного 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% неcatiонного 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% неcatiонных 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, неcatiонный 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 неcatiонного 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/неcatiонный 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 40. В некоторых вариантах осуществления

содержание реакционноспособных примесей (например, альдегидов) и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяют посредством обнаружения 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 обнаружения 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 41. В вариантах осуществления химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида обнаруживают посредством определения присутствия одного или нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 41.

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 управляет ассоциацией с тканями, где клетки экспрессируют рецептор, и доставкой 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, зависящих 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 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)пропил)октадека-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]-диоколана (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, мПНК Cas9, описаны в WO2019067992 и WO2019067910, обе 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 (ПНК). 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.

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, номера доступа последовательностей, указанные в данном документе, в том числе в любой таблице, 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.

83. 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.

84. Пример 1. Gene Writer™, обеспечивающий замену нуклеотидов в геномной ДНК для коррекции мутации, обуславливающей дефицит альфа-1-антитрипсина в клетках человека

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 (E342K), распространенным аллелем, ассоциированным с дефицитом альфа-1-антитрипсина. Полипептид Gene Writer™ использует нуклеазу 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 она содержит дополнительные нуклеотиды в необходимом положении таким образом, что обратная транскрипция РНК-матрицы приводит к образованию новой нити ДНК, содержащей замену.

Для создания трансверсии в пораженном гене *SERPINA1* человека, которая восстанавливает триплет GAG, кодирующий глутамат, у здоровых пациентов, применяют полипептид Gene Writer™ со специфической нуклеиновой кислотой-матрицей, которая кодирует каркас gRNA для связывания полипептида, спейсер для хоуминга полипептида, домен гомологии с мишенью для настройки TPRT и последовательность матрицы для обратной транскрипции, которая включает необходимую замену. Иллюстративная РНК-матрица несет последовательность (1)TCCCCTCCAGGCCGTGCATA(2)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)TcGTCGATGGTC AGCACAGCCTTAT(4)GCACGGCCTGGA (SEQ ID NO: 1607), где числа используются для обозначения модулей матрицы в порядке (5'-3') (1) спейсер gRNA; (2) каркас gRNA; (3) последовательность гетерологичного объекта и (4) 3'-концевой праймирующий домен гомологии, а строчная буква "с" указывает положение на матрице, несущее нуклеотидную замену, которая должна быть записана в сайт-мишень для коррекции мутации E342K. Иллюстративная gRNA для обеспечения второго однонитевого разрыва, как описано в вариантах осуществления данной системы, содержит спейсерную последовательность TTTGTTGAACTTGACCTCGG (SEQ ID NO: 1608) и направляет нуклеазу Cas9 для

осуществления одностороннего разрыва во второй нити сайта-мишени в пределах области гомологии. В некоторых вариантах осуществления данный второй односторонний разрыв обеспечивает улучшение эффективности редактирования.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени, после чего геномную ДНК экстрагируют из клеток. 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.

85. Пример 2. Gene Writer™, обеспечивающий короткую вставку в геномную ДНК для коррекции CFTR

В этом примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки короткого отрезка из нуклеотидов.

В этом примере полипептид Gene Writer™ и записывающая матрица представлены в виде ДНК, трансфицированной в клетки HEK293T, которые имеют мутацию CFTR дельта-F508 - распространенный аллель, ассоциированный с кистозным фиброзом. Полипептид Gene Writer™ использует 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 короткую вставку.

Для создания короткой вставки в пораженном локусе CFTR человека, которая восстанавливает триплет ТТТ, кодирующий фенилаланин 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 вставку из 3 нуклеотидов.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени. После инкубационного периода 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.

86. Пример 3. Gene Writer™, обеспечивающий делецию геномной ДНК для коррекции мышечной дистрофии Дюшенна (DMD)

В этом примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности путем делеции нуклеотидов.

Одной из наиболее распространенных мутаций, обнаруживаемых у пациентов с DMD, является делеция, которая приводит к удалению экзона 50 в палочковидном домене дистрофина, в результате чего экзон 51 помещается вне рамки с предшествующими экзонами. Такая мутация приводит к продуцированию усеченного дистрофина, что приводит к патологическим эффектам заболевания. Для облегчения течения заболевания оставшаяся часть из общего количества 79 экзонов, акцепторный сайт сплайсинга удаляется из экзона 51, что приводит к восстановлению полноразмерного белка, и этот подход известен как пропуск экзона.

В этом примере полипептид Gene Writer™ и записывающая матрица представлены в виде РНК, нуклеофицированной в клетки, содержащие делецию в экзоне 50, что приводит к получению усеченного продукта дистрофина, как описано выше. Клетки-мишени представляют собой полученные от пациента iPSC, содержащие мутацию, либо синтетически сконструированные с использованием CRISPR-Cas для создания делеции. Полипептид Gene Writer™ использует Cas9 в качестве 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 делецию в необходимом положении таким образом, что обратная транскрипция РНК-матрицы приводит в результате к образованию новой нити ДНК, в которой отсутствуют удаленные нуклеотиды.

Для создания короткой делеции, при которой удаляется 5'-акцепторный сайт сплайсинга экзона 51, 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 5-нуклеотидную делецию, проксимальную относительно однонитевого разрыва, индуцированного полипептидом Gene Writer™, который содержит акцепторный сайт сплайсинга.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени. После инкубационного периода 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. Анализ белка с помощью вестерн-блоттинга используется для дополнительного подтверждения экспрессии восстановленного дистрофина по сравнению с усеченным дистрофином, продуцируемым в неотредактированных клетках.

87. Пример 4. Gene Writer™ позволяет осуществлять вставку крупного участка в геномную ДНК

В этом примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки крупного отрезка из нуклеотидов.

В этом примере полипептид Gene Writer™, gRNA и матрица для записи представлены в виде ДНК, трансфицированной в клетки HEK293T. Полипептид Gene Writer™ использует нуклеазу 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 односторонний разрыв в сайте-мишени, а затем выполняет TPRT на матрице, иницируя реакцию с использованием праймирующих областей на матрице, которые комплементарны последовательности, непосредственно примыкающей к сайту одностороннего 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 инкубационного периода 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.

88. Пример 5. Gene Writer производит редактирование без включения связывающих последовательностей из РНК-матрицы

В этом примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки генетической полезной нагрузки, которая не приводит к вставке дополнительной последовательности из молекулы-матрицы.

В этом примере полипептид Gene Writer™ и матрица для записи представлены в виде ДНК, трансфицированной в клетки HEK293T. Полипептид Gene Writer™ использует нуклеазу Cas9 в качестве как ДНК-связывающей, так и эндонуклеазной функциональной единицы. Матрица для записи сконструирована так, что она характеризуется гомологией с

последовательностью-мишенью и в то же время содержит генетическую полезную нагрузку (например, единицу экспрессии гена GFP) в необходимом положении таким образом, что обратная транскрипция РНК-матрицы приводит в результате к образованию новой нити ДНК, содержащей необходимую вставку.

Чтобы выполнить специфическую вставку генетической полезной нагрузки без встраивания посторонних мотивов матрицы (например, мотива связывания белка), расположение молекулы РНК-матрицы является таким, что последовательности связывания белка (например, UTR) являются концевыми относительно последовательностей гомологии, используемых для записи новой полезной нагрузки в геномный сайт-мишень.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и превращения геномной ДНК-мишени. После инкубационного периода 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.

89. Пример 6. Редактирование генома с помощью Gene Writer™ в присутствии ингибиторов репарации ДНК

В этом примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки генетической полезной нагрузки, которая не приводит к вставке дополнительной последовательности из молекулы-матрицы.

В этом примере с помощью экспериментов тестируют эффект различных путей репарации ДНК в отношении Gene Writing™ посредством применения ингибиторов путей репарации ДНК или линий клеток, дефектных по путям репарации ДНК. При применении ингибиторов путей репарации ДНК вначале проводили анализ жизнеспособности клеток PrestoBlue для определения токсичности ингибиторов и необходимости применения какой-либо нормализации для следующих анализов. SCR7 представляет собой ингибитор NHEJ, который применяется в серии разведений при доставке Gene Writing™. Белок PARP представляет собой ядерный фермент, который в виде гомодимеров связывается как с одонитевыми, так и с двухнитевыми разрывами. Таким образом, его ингибиторы применяются при тестировании соответствующих путей репарации ДНК, в том числе пути репарации посредством гомологичной рекомбинации и пути эксцизионной репарации оснований. Процедура эксперимента является такой же, как и для SCR7. Линии клеток с дефицитом сердцевинных белков пути эксцизионной репарации нуклеотидов (NER) использовали для тестирования эффекта NER в отношении Gene Writing™. После доставки системы Gene Writing™ в клетку применяли ddPCR для оценивания ретротранспозиции в условиях ингибирования путей репарации ДНК. Также проводили анализ методом секвенирования для оценивания того, играют ли определенные пути репарации ДНК роль

в изменении места соединения интегрированного элемента. В некоторых вариантах осуществления Gene Writing™ в геном не будет уменьшаться вследствие нокдауна путей репарации ДНК, что позволяет предположить, что система не использует пути репарации клеток-хозяев для интеграции ДНК. В некоторых вариантах осуществления Gene Writing™ в геном не будет уменьшаться более чем на 50% вследствие нокдауна путей репарации ДНК, что позволяет предположить, что система не зависит от путей репарации клеток-хозяев для интеграции ДНК.

90. Пример 7. Внутренние делеции Gene Writer, демонстрирующие модульность домена белка.

В данном примере описаны делеции в полипептиде Gene Writer, при которых сохраняется функциональность и которые дополнительно демонстрируют модульность ДНК-связывающего домена.

В данном примере проводили серию экспериментов для тестирования активности различных мутантных ретротранспозаз, а также для получения структурных данных о модульности белка. В данном эксперименте тестировали удаление полипептидного отрезка после мотива с-myb в ДНК-связывающем домене (DBD) и замену его гибким линкером (фигура 8a). Удаленный отрезок полипептида называют "природным линкером", поскольку он представляет собой промежуточную область между ДНК-связывающими мотивами и РНК-связывающим доменом. Удаленная область полипептида охватывает следующее, начинаясь на N-концевой стороне либо в местоположении А (предсказанный статистический клубок, следующий за мотивом с-myb), либо в местоположении В (конец предсказанной альфа-спирали, которая содержит часть мотива с-myb), и удаленная область заканчивается либо в местоположении v1 (альфа-спиральная область R2Tg, которая предшествовала предсказанному -1 РНК-связывающему мотиву), либо в местоположении v2 (С-концевая сторона альфа-спиральной области R2Tg, которая предшествовала предсказанному -1 РНК-связывающему мотиву). Вместо удаленного отрезка полипептида, "природного линкера", присутствует любой из двух линкеров (линкера А, XTEN: SGSETPGTSESATPES (SEQ ID NO: 1023), и линкера В, 3GS: GGS (SEQ ID NO: 1024)). Для каждой из этих мутантных ретротранспозаз, содержащих различные удаленные области (местоположение А - v1, местоположение А - v2, местоположение В - v1 или местоположение В - v2), они были заменены либо линкером А, либо линкером В с помощью ПЦР на ДНК-плазмиду, экспрессирующую R2Tg, что таким образом дает следующие последовательности: с-mybА - v1 заменен линкером 3GS (SEQ ID NO: 1024), с-mybА - v2 заменен линкером 3GS (SEQ ID NO: 1024), с-mybА - v1 заменен линкером XTEN, с-mybА - v2 заменен линкером XTEN, с-mybВ - v1 заменен линкером 3GS (SEQ ID NO: 1024), с-mybВ - v2 заменен линкером 3GS (SEQ ID NO: 1024), с-mybВ - v1 заменен линкером XTEN, с-mybВ - v2 заменен линкером XTEN, как показано в таблице E1 ниже. Вставку линкеров проверяли посредством секвенирования по Сэнгеру и ДНК-плазмиды очищали для трансфекции.

91. Таблица E1. Аминокислотные последовательности мутантных форм R2Tg с линкерами вместо области "природного линкера", которая находится между ДНК-связывающим доменом (DBD) и РНК-связывающим доменом.

N-концевой ДНК-связывающий домен выделен курсивом, а линкер, соединяющий с остальной частью белка, выделен полужирным шрифтом и подчеркнут.

Мутантная форма R2Tg	Аминокислотная последовательность	SEQ ID NO:
R2Tg с делецией природного линкера, участок "местоположение с-мыбA - v1" заменен линкером 3GS	MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEPGTCH HTRRAA<u>GGGS</u>CFGCLESISQIRTATRDKKDTVTRKHPKKPFQKWMKDR AIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGS FKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVV KMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRDKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIWSAK REHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENIST YITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGF YIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYADHSEVKTALLETLDQKI RTAVKEWLHLPCTCDAILYSTRDGGGLGITKLGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSEPPNNVST NSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNH WIQYYRRIPHRKLLTALQLRANVYPTREFLAGRQDQYIKACRHCDADIES CAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYK PDLIFVKDARALVVDVTVRYEAAKSSLEEEAAEKVRKYKHLETEVRHLTNA KDVTFGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSV DIVHMFASRARKSMVM	1646

R2Tg делецией природного линкера, участок "местоположение c- mybA - v2" заменен линкером 3GS	MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNSLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPCRCGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLSRKPAEEPREEPGTCH HTRRAA<u>GGGS</u>TATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQ RLFYLDRGKLAKIILDDIECLSDIPLSEIYSVFKTRWETTGSFKSLGDFKTYG KADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTM EIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFS RIVTARLSKACPLNPRQGRGFIARAAGCSENKLLQTIWSAKREHRPLGVVVFV DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTD KIQIRVGVKQGDPMSPLLFNLAMDPDLLCKLEESGKGYHRGQSSITAMAFA DDLVLSDSWENMNTNISILETFCNLTGLKTQGGKCHGFYIKPTKDSYTIN DCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRI DQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHL PPCTCDAILYSSSTRDGGGLGITKLGLIPSVQARRLHRIAQSSDDTMKCFME KEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHR KLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARA LVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFFVGFPLG ARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRAR KSMVM	1647
R2Tg делецией природного линкера, участок "местоположение c-	MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNSLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPCRCGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE	1648

mybA - v1" заменен линкером XTEN	LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREPGTCH HTRRAA<u>SGSETPGTSESATPES</u>CFGCLESISQIRTATRDKKDVTREKHPKK PFQKWMKDRAIKKGNLYRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSV FKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPG PDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKP DRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSEN KLLQTIIWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVG LVSNNMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDPDLLCK LEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLK TQGGQKCHGFYIKPTKDSYINDCAAWTINGTPLNMIDPGESEKYLGLQFD PWIGIARSGSLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVK TALLETLDQKIRTAVKEWLHLPPCTDAILYSSSTRDGGGLGITKLAGLIPSVQA RRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP SSEPPNNVSTNSEWEAPTQKDKFKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAGRQDQYIKA CRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKDWVVFKEP HIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHL ETEVRLHTNAKDVT FVGFLGARGKWHQDNFKLLTELGLSKSRQVKMAE TFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg c делецией природного линкера, участок "местополо жение c- mybA - v2" заменен линкером XTEN	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREPGTCH HTRRAA<u>SGSETPGTSESATPE</u>STATRDKKDVTREKHPKKPFQKWMKDR AIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGS FKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVV KMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIIWSAK REHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENIST	1649

		YITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGGQKCHGF YIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKI RTAVKEWLHLPPCTCDAILYSTRDGGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVST NSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNH WIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIES CAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWWVFKEPHIRDSNKELYK PDLIFVKDARALVVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNA KDVTFGVPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSV DIVHMFASRARKSMVM	
R2Tg делецией природного линкера, участок "местополо жение с- mybB - v1" заменен линкером 3GS	с	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSQSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGKNKINKLIAEHITTKTAKQISDKRRLLGGGSCFGCLESISQIRT ATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKII LDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITA KEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIP DMVRGCRTVLIPKSSKPDRDKDINNWRPITIGSILLRLFSRIVTARLSKACPL NPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ IIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGD PMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTLGLKTQGGQKCHGFYIKPTKDSYTINDCAAWTINGTPL NMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDI LKTYTIPRLIYIADHSEVKTALLETLDQKIRTAKEWLHLPPCTCDAILYSTR DGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWI QAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRK NEFKKWTKLASQGRGIVNFERDKISNHWIYQYYRRIPHRKLLTALQLRANVY	1650

		PTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAK SSLEEAAAEKVRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKL LTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg делецией природного линкера, участок "местополо жение с- mybB - v2" заменен линкером 3GS	c	MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGKNINKLIAEHITTKTAKQISDKRRLLGGGSTATRDKKDTVTR EKHPKPPFQKWMKDRAIKKGNYLRFQRLFYLD RGLAKIILDDIECLSCDIP LSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMS KGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLI PKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAA GCSENLLKLTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVG VKQGDPMSPLLFNLA MDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILET FCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSGSLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYI ADHSEVKTALLETLDQKIRTAVKEWLHLP PCTCDAILYSSSTRDGGGLGITKLA GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENI PSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTCLA SQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKD WVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAAE KVRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKLLTELGLSKS RQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	1651
R2Tg делецией природного линкера,	c	MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK	1652

участок "местополо жение с- mybB - v1" заменен линкером XTEN	<i>TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLSGSETPGTSESATPESC FGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQRL FYLDGRGLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKA DNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEI FNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRFLSRI VTARLSKACPLNPRQRGFIRAAGCSENKLLQTIIWSAKREHRPLGVVFVDI AKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKI QIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFAD DLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTIND CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPP CTCDAILYSSSTRDGGGLGITKLGLIPSVQARRLHRIAQSSDDTMKCFMEKE KMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKD KFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKL LTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQD ARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDIFVKDARALV VDVTVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVTFGFPLGAR GKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKS MVM</i>	
R2Tg c делецией природного линкера, участок "местополо жение с- mybB - v2" заменен линкером XTEN	<i>MASCPKPGPPVSAGAMSLESLTTHSVLAIERGPNSLANSGSDFGGGGLG LPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDL LSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAH WEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPK TSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVS HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLSGSETPGTSESATPEST ATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDGRGLAKII LDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITA KEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIP</i>	1653

DMVRGCRTVLIPKSSKPDRCLKDINNWRPITIGSILLRLFSRIVTARLSKACPL NPRQRGFIRAAGCSENKLLQTIWSAKREHRPLGVVFDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGD PMSPLLNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPL NMIDPGESEKYLGLQFDPWIGIARSGSLSTKLDFWLQRIDQAPLKPLQKTDI LKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSSTR DGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWI QAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRK NEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVDVTVRYEAAK SSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFFVGFLGARGKWHQDNFKL LTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	
--	--

Клетки HEK293T помещали в 96-луночные планшеты и выращивали в течение ночи при 37°C, 5% CO₂. Клетки HEK293T трансфицировали плазмидами, экспрессирующими R2Tg (дикого типа), мутантную форму эндонуклеазы R2 и мутантные формы с природными линкерами. Трансфекцию проводили с помощью реагента для трансфекции Eugene HD в соответствии с рекомендациями производителя, при этом в каждую лунку вводили 80 нг плазмидной ДНК и 0,5 мкл реагента для трансфекции. Все трансфекции проводили в двух повторностях, и перед экстракцией геномной ДНК клетки инкубировали в течение 72 ч.

Активность мутантных форм измеряли с помощью анализа методом ddPCR, в котором количественно определяли число копий интегрированного R2Tg путем измерения количества ампликонов с 3' соединениями (фигура 8b).

Делеции, которые начинаются после статистического клубка, следующего за ДНК-связывающим мотивом c-myb (местоположение A, c-mybA), хорошо переносятся, при этом активность интеграции является близкой к таковой R2Tg дикого типа. Конечная точка делеции природной линкерной области почти одинакова как для местоположения v1 (N-конец по отношению к альфа-спирали, предшествующей -1 РНК-связывающему мотиву), так и для местоположения v2 (C-конец по отношению к альфа-спирали, предшествующей -1 РНК-связывающему мотиву). При делеции, начинающейся в местоположении A и заканчивающейся в местоположении v1 или v2, замена данного полипептидного отрезка линкером XTEN (SEQ ID NO: 1023), по-видимому, обеспечивает сохранение наибольшей степени активности, тогда как замена линкером 3GS (SEQ ID NO: 1024) характеризуется снижением активности интеграции на примерно 50%. При делециях природного линкера, которые начинаются в местоположении B (c-mybB), данные конфигурации демонстрируют более заметное снижение активности интеграции по сравнению с диким типом или

делециями в местоположении A (с-mybA). Различие в активности может быть связано со структурой белка в зависимости от положения делеции, которая создает неоптимальную трехмерную структуру ретротранспозазы из-за местоположения линкера, длины линкера или комбинации аминокислот линкера, которые не являются оптимальным для соединения местоположения B с местоположениями v1 или v2. Даже несмотря на то, что N-концевое местоположение делеции природного линкера mybB является субоптимальным, C-конец делеции характеризовался наилучшей переносимостью в v2 с любым из линкеров 3GS (SEQ ID NO: 1024) или линкера XTEN и, по-видимому, является предпочтительным местоположением для наличия полипептида, предшествующего области RBD-1.

92. Пример 8. Определение специфичности эндонуклеазного домена Gene Writer в отношении мишени

В данном примере описано применение специализированной геномной "посадочной площадки" в клетках человека, чтобы определить, существует ли требование в отношении последовательности для расщепления мишени и последующей интеграции с помощью системы Gene Writer.

В этом примере были созданы клеточные линии, имеющие "посадочные площадки" или стабильные интеграции, которые имитируют область рДНК, содержащую положение R2, к которому ретротранспозазы R2 направляются для ретротранспозиции (см. фиг. 9). Интегранты или "посадочные площадки" были разработаны таким образом, чтобы они содержали последовательность области дикого типа внутри и вблизи сайта R2, находящегося в рДНК, мутацию последовательности длиной 12 п. о. в сайте расщепления R2 и вблизи него, или мутацию последовательности длиной 75 п. о. в сайте расщепления R2 или вблизи него (таблица E2). ДНК для данных различных "посадочных площадок" химически синтезировали и клонировали в вектор pLenti-N-tGFP. Клонированные "посадочные площадки" в лентивирусном векторе экспрессии подтверждали, и их последовательность верифицировали посредством секвенирования "посадочных площадок" по Сэнгеру. Плазмидами с верифицированной последовательностью (9 мкг) вместе со смесью для упаковки в лентивирус (9 мкг, полученной от Biosettia) трансфицировали с применением Lipofectamine 2000™ в соответствии с инструкциями производителя в паковую линию клеток LentiX-293T (Takara Bio). Трансфицированные клетки инкубировали при 37°C, 10% CO₂ в течение 48 часов (включая одну смену среды через 24 ч) и среду, содержащую вирусные частицы, собирали из чашки для культивирования клеток. Собранную среду фильтровали через фильтр с размером пор 0,2 мкм для удаления клеточного дебриса и готовили для трансдукции клеток U2OS. Среду, содержащую вирус, разбавляли в DMEM и смешивали с полибреном с получением серии разбавлений для трансдукции клеток U2OS, где конечная концентрация полибрена составляла 8 мкг/мл. Клетки U2OS выращивали в среде, содержащей вирусы, в течение 48 часов, и затем открепляли с использованием свежей среды. Открепленные клетки выращивали до достижения конфлюэнтности и эффективность трансдукции при различных разбавлениях вируса измеряли по экспрессии GFP с помощью проточной цитометрии и

обнаружения с помощью ddPCR интегрированного в геном лентивируса, который содержал GFP и различные "посадочные площадки" рДНК (WT, мутация длиной 12 п. о. или мутация длиной 75 п. о.). Положительная по GFP линия клеток, полученная с использованием разбавления вирусной среды 1:10 (>99% GFP+), была выбрана для последующих экспериментов и криоконсервирована.

Для тестирования того, могут ли мутации в положении расщепления R2 и вблизи него влиять на активность системы Gene Writer, управляющий элемент R2Tg Gene Writer вместе с плазмидой, экспрессирующей молекулу трансгена Gene Writer, посредством электропорации вводили в линии клеток с различными "посадочными площадками". С целью тестирования того, влияет ли последовательность в сайте расщепления и вблизи него на активность интеграции полипептидной последовательности Gene Writer, плечи гомологии молекулы-матрицы Gene Writer были разработаны таким образом, чтобы они характеризовались 100% гомологией с последовательностью длиной 100 п. о., расположенной слева (модуль А молекулы Gene Writer) от положения расщепления, и последовательностью длиной 100 п. о., расположенной справа (модуль F молекулы Gene Writer) от положения расщепления, для каждой из "посадочных площадок". Изменения в плечи гомологии плазмиды экспрессии молекулы-матрицы Gene Writer вводили с помощью ПЦР и подтверждали посредством секвенирования по Сэнгеру. Осуществляли нуклеофекцию либо 73 нг плазмид экспрессии управляющего элемента R2Tg WT Gene Writer, либо плазмид экспрессии управляющего элемента R2Tg с мутантным эндонуклеазным доменом Gene Writer с применением программы нуклеофекции DN100 в каждую из линий клеток U2OS с различными "посадочными площадками" (WT, с мутантной последовательностью длиной 12 п. о. или с мутантной последовательностью длиной 75 п. о.) с 177 нг плазмид, которые экспрессировали молекулы-матрицы Gene Writer, которые характеризовались 100% гомологией либо с "посадочной площадкой" WT, либо с "посадочной площадкой" с мутантной последовательностью длиной 12 п. о., либо с "посадочной площадкой" с мутантной последовательностью длиной 75 п. о. После нуклеофекции клетки выращивали при 37°C, 10% CO₂ в течение 3 дней перед лизисом клеток и экстракцией геномной ДНК. С помощью ddPCR экстрагированную gDNA оценивали на предмет интеграции молекулы-матрицы Gene Writer в сайт "посадочной площадки". Активность внесения одонитевых разрывов в ДНК измеряли посредством обнаружения вставок, делеций и/или комбинации как вставок, так и делеций на "посадочной площадке" с помощью анализа последовательности нового поколения, осуществляемого в отношении ампликона, который получали из "посадочной площадки", находящейся в gDNA.

Активность интеграции R2Tg Gene Writer значительно снижается, если область расщепления была подвергнута мутации, где интеграция молекулы-матрицы Gene Writer не происходит ни в линии клеток с "посадочной площадкой" с 12 п. о., ни в линии клеток с "посадочной площадкой" с 75 п. о. (фигура 10a). Более того, интеграция не обнаруживается при использовании молекул-матриц Gene Writer, которые содержат плечи гомологии,

соответствующие "посадочным площадкам" с мутантными последовательностями длиной 12 п. о. или длиной 75 п. о. Чтобы исключить, что потеря активности интеграции происходит из-за несовместимых плеч гомологии, активность внесения одонитевых разрывов в ДНК измеряли с помощью NGS-анализа "посадочной площадки". Активность внесения одонитевых разрывов не зависит от "молекулы-матрицы" Gene Writer, поскольку управляющий элемент R2Tg WT Gene Writer характеризовался сопоставимыми показателями вставок/делеций на "посадочной площадке" WT при использовании молекулы-матрицы Gene Writer WT, с мутантной последовательностью длиной 12 п. о. или мутантной последовательностью длиной 75 п. о. (фигура 10b). В случае "посадочных площадок" с 12 п. о. и 75 п. о., независимо от молекулы-матрицы Gene Writer, нуклеофекцию которой осуществляли совместно с последовательностью R2Tg WT Gene Writer, не было показано наличия каких-либо ридов выше фонового уровня, которые бы содержали вставки/делеции. Уровень вставок/делеций был аналогичен таковому при использовании управляющего элемента/матрицы Gene Writer с мутациями в последовательности эндонуклеазы.

93. Таблица E2. Иллюстративные "посадочные площадки"

Название последовательно сти "посадочной площадки"	Последовательность 5'→3'
	(<u>рДНК</u> , подчеркнуто; область расщепления , выделено полужирным шрифтом; <i>подвергнутая мутации последовательность</i> , выделено полужирным курсивом)
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACGCGAT GTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAATGAAGCG CGGGTAAACGGCGGGAGTAACTATGACTCTCTT AAGGTAGCCAAATGCCTC GTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCT ACCTACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCGTCGTTTT ACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACCTAATCGCCTTGCA GCACATCC (SEQ ID NO: 1654)
Мутантная форма 12 п. о.	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACGCGAT GTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAATGAAGCG CGGGTAAACGGCGGGAGTAACTATGACTCTCTT TCCAATATGATTTGCCTCG TCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTA CCTACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCGTCGTTTTAC AACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACCTAATCGCCTTGCA GCACATCC (SEQ ID NO: 1655)

Мутантная форма 75 п. о.	gctcacacaggaaacagctatgaccatgattacgccaagctggtgacgcatgtgatttctgccagt gctctgaatgtcaaagtgaagaaattcaatgaagcgcggttaaaccggcgaggtaactatgactctc ttccaatatgattccacccatggcaaatccatggcaccgtcaaggctgagaacgggaagctt gtcatcaatggaaactatccagcgaaaccacagccaagggaattcactggccgtcgtttacaac gtcgtgactgggaaaaccctggcggttacccaactaatcgcttgagcacatcc (SEQ ID NO: 1656)
-----------------------------	---

В некоторых вариантах осуществления Gene Writer получен с использованием ретротранспозазы с некоторым уровнем специфичности в отношении последовательности-мишени в эндонуклеазном домене. Таким образом, желательным может являться перенацеливание Gene Writer на местоположение в геноме, которое характеризуется гомологичностью с природной последовательностью-мишенью, распознаваемой эндонуклеазным доменом, называемой мотивом распознавания эндонуклеазой (ERM). В некоторых вариантах осуществления данная последовательность, являющаяся вторичной мишенью, может содержаться в области, окружающей сайт однокитевого разрыва. В конкретных вариантах осуществления последовательность из 13 нт (TAAGGTAGCCAAA (SEQ ID NO: 1657)) на основе сайта однокитевого разрыва элемента R2, например, R2Tg, применяют для поиска в геноме человека подходящих местоположений для перенацеливания Gene Writer, в котором гетерологичный ДНК-связывающий домен сконструирован таким образом, чтобы он обеспечивал перемещение Gene Writer к эндогенному ERM для направления эндонуклеазной активности и последующей ретротранспозиции РНК-матрицы. В некоторых вариантах осуществления сайт в геноме человека характеризуется 100% идентичностью с по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 или 13 нуклеотидами в мотиве из 13 нуклеотидов. В дополнительных вариантах осуществления сайт в геноме человека, содержащий ERM, выбран из таблицы E3, и продукт слияния ДНК-связывающего домена, например, ZF, TAL или dCas9, со специализированной gRNA предназначен для обеспечения перемещения полипептида к данному сайту (например, см. пример 9). В предпочтительных вариантах осуществления сайт в геноме имеет балл по безопасной гавани, составляющий по меньшей мере 5, 6, 7, 8, как определено в Pellenz et al Hum Gene Ther 30, 814-282 (2019) и показано в таблице E3. В некоторых вариантах осуществления РНК-матрица (или ДНК, кодирующая РНК-матрицу) разработана таким образом, что плечи гомологии соответствуют фланкирующим геномным последовательностям, окружающим ожидаемый сайт однокитевого разрыва в новой мишени.

94. *Таблица E3. Сайты в геноме человека, соответствующие отрезку длиной 13 нуклеотидов вблизи сайта однокитевого разрыва, осуществляемого элементами R2*

В геноме человека осуществили поиск последовательности, характеризующейся 100% идентичностью полной последовательности совпадения из 13 нуклеотидов или 12 последовательным нуклеотидам ("Совпадение"). Местоположение в хромосоме и координаты начала и конца представлены для каждого совпадения. Балл("Балл")

представляет собой метрику, с помощью которой каждый сайт оценивают в отношении восьми желательных характеристик безопасной гавани.

Хромосома	Начало	Конец	Источник	Совпадение	Балл
chr06	123749082	123749094	NC_000006.12	13	8
chr02	5035294	5035305	NC_000002.12	12	8
chr02	145760352	145760341	NC_000002.12	12	8
chr02	147034635	147034624	NC_000002.12	12	8
chr02	181792104	181792115	NC_000002.12	12	8
chr03	34017171	34017182	NC_000003.12	12	8
chr03	74784684	74784695	NC_000003.12	12	8
chr03	110093351	110093362	NC_000003.12	12	8
chr06	14459104	14459093	NC_000006.12	12	8
chr06	119620936	119620947	NC_000006.12	12	8
chr06	145123473	145123462	NC_000006.12	12	8
chr07	12024654	12024665	NC_000007.14	12	8
chr07	52001436	52001447	NC_000007.14	12	8
chr07	115339421	115339410	NC_000007.14	12	8
chr08	126384299	126384310	NC_000008.11	12	8
chr12	84083562	84083573	NC_000012.12	12	8
chrX	117646432	117646421	NC_000023.11	12	8
chr02	106547509	106547521	NC_000002.12	13	7
chr02	226038592	226038604	NC_000002.12	13	7
chr03	102522532	102522520	NC_000003.12	13	7
chr03	110933592	110933604	NC_000003.12	13	7
chr03	119752575	119752563	NC_000003.12	13	7
chr03	172868603	172868615	NC_000003.12	13	7
chr03	191985222	191985210	NC_000003.12	13	7
chr05	6213503	6213515	NC_000005.10	13	7
chr05	58295578	58295566	NC_000005.10	13	7
chr05	129844500	129844512	NC_000005.10	13	7
chr06	1454372	1454360	NC_000006.12	13	7
chr06	48973921	48973909	NC_000006.12	13	7
chr08	18663054	18663066	NC_000008.11	13	7
chr08	93499020	93499032	NC_000008.11	13	7

chr08	119753973	119753985	NC_000008.11	13	7
chr09	86856907	86856919	NC_000009.12	13	7
chr12	29955571	29955583	NC_000012.12	13	7
chr12	118529104	118529092	NC_000012.12	13	7
chr13	65656029	65656041	NC_000013.11	13	7
chr22	34266611	34266623	NC_000022.11	13	7
chrX	26651640	26651628	NC_000023.11	13	7
chrX	119194351	119194363	NC_000023.11	13	7
chrX	139180620	139180608	NC_000023.11	13	7
chr01	106465846	106465857	NC_000001.11	12	7
chr02	964160	964171	NC_000002.12	12	7
chr02	40018947	40018936	NC_000002.12	12	7
chr02	62845403	62845392	NC_000002.12	12	7
chr02	64834920	64834909	NC_000002.12	12	7
chr02	67969608	67969619	NC_000002.12	12	7
chr02	76183118	76183129	NC_000002.12	12	7
chr02	81819286	81819297	NC_000002.12	12	7
chr02	119597238	119597249	NC_000002.12	12	7
chr02	122897376	122897365	NC_000002.12	12	7
chr02	123603423	123603412	NC_000002.12	12	7
chr02	144644206	144644217	NC_000002.12	12	7
chr02	145221757	145221746	NC_000002.12	12	7
chr02	158367531	158367520	NC_000002.12	12	7
chr02	160092083	160092072	NC_000002.12	12	7
chr02	192245037	192245048	NC_000002.12	12	7
chr02	195223552	195223563	NC_000002.12	12	7
chr02	200351999	200351988	NC_000002.12	12	7
chr02	237068525	237068514	NC_000002.12	12	7
chr03	18724351	18724340	NC_000003.12	12	7
chr03	23969399	23969388	NC_000003.12	12	7
chr03	25177339	25177350	NC_000003.12	12	7
chr03	34880863	34880852	NC_000003.12	12	7
chr03	66233879	66233890	NC_000003.12	12	7
chr03	74527939	74527950	NC_000003.12	12	7

chr03	98583025	98583014	NC_000003.12	12	7
chr03	99278452	99278463	NC_000003.12	12	7
chr03	116060228	116060239	NC_000003.12	12	7
chr03	139468578	139468589	NC_000003.12	12	7
chr03	140064054	140064043	NC_000003.12	12	7
chr03	140438138	140438127	NC_000003.12	12	7
chr03	152457330	152457341	NC_000003.12	12	7
chr03	160950736	160950725	NC_000003.12	12	7
chr03	167207758	167207769	NC_000003.12	12	7
chr03	167722472	167722483	NC_000003.12	12	7
chr03	180475661	180475672	NC_000003.12	12	7
chr04	121590786	121590775	NC_000004.12	12	7
chr04	133719599	133719588	NC_000004.12	12	7
chr05	11564132	11564121	NC_000005.10	12	7
chr05	11970221	11970210	NC_000005.10	12	7
chr05	32814431	32814420	NC_000005.10	12	7
chr05	38003029	38003018	NC_000005.10	12	7
chr05	39758118	39758129	NC_000005.10	12	7
chr05	41221615	41221604	NC_000005.10	12	7
chr05	74838717	74838728	NC_000005.10	12	7
chr05	86444529	86444518	NC_000005.10	12	7
chr05	86617117	86617106	NC_000005.10	12	7
chr05	89438360	89438349	NC_000005.10	12	7
chr05	108102395	108102406	NC_000005.10	12	7
chr05	110231750	110231761	NC_000005.10	12	7
chr05	113996496	113996485	NC_000005.10	12	7
chr05	117233050	117233039	NC_000005.10	12	7
chr05	121622921	121622932	NC_000005.10	12	7
chr05	122520704	122520693	NC_000005.10	12	7
chr05	142330490	142330479	NC_000005.10	12	7
chr05	156359105	156359094	NC_000005.10	12	7
chr06	19187842	19187831	NC_000006.12	12	7
chr06	41103469	41103458	NC_000006.12	12	7
chr06	49856872	49856883	NC_000006.12	12	7

chr06	54896309	54896298	NC_000006.12	12	7
chr06	64416107	64416096	NC_000006.12	12	7
chr06	104438997	104439008	NC_000006.12	12	7
chr06	109349688	109349677	NC_000006.12	12	7
chr06	110631149	110631160	NC_000006.12	12	7
chr06	114751383	114751372	NC_000006.12	12	7
chr06	116514339	116514350	NC_000006.12	12	7
chr06	121606126	121606137	NC_000006.12	12	7
chr06	126139788	126139777	NC_000006.12	12	7
chr06	130852449	130852460	NC_000006.12	12	7
chr06	136843057	136843068	NC_000006.12	12	7
chr06	155559692	155559681	NC_000006.12	12	7
chr06	158099232	158099221	NC_000006.12	12	7
chr07	24339208	24339219	NC_000007.14	12	7
chr07	39519736	39519725	NC_000007.14	12	7
chr07	72228733	72228722	NC_000007.14	12	7
chr07	82955239	82955228	NC_000007.14	12	7
chr07	104990180	104990191	NC_000007.14	12	7
chr07	107698186	107698197	NC_000007.14	12	7
chr07	111002449	111002438	NC_000007.14	12	7
chr07	115191852	115191841	NC_000007.14	12	7
chr07	115572755	115572766	NC_000007.14	12	7
chr08	21126162	21126151	NC_000008.11	12	7
chr08	25928055	25928066	NC_000008.11	12	7
chr08	45036535	45036524	NC_000008.11	12	7
chr08	45248484	45248473	NC_000008.11	12	7
chr08	45502096	45502085	NC_000008.11	12	7
chr08	45763984	45763973	NC_000008.11	12	7
chr08	53335054	53335043	NC_000008.11	12	7
chr08	55581238	55581227	NC_000008.11	12	7
chr08	63169546	63169557	NC_000008.11	12	7
chr08	66553887	66553898	NC_000008.11	12	7
chr08	86283378	86283367	NC_000008.11	12	7
chr08	113060704	113060715	NC_000008.11	12	7

chr08	114537195	114537206	NC_000008.11	12	7
chr08	114886082	114886071	NC_000008.11	12	7
chr08	127206415	127206426	NC_000008.11	12	7
chr08	133590421	133590410	NC_000008.11	12	7
chr08	135425161	135425150	NC_000008.11	12	7
chr10	7034863	7034852	NC_000010.11	12	7
chr10	124304797	124304786	NC_000010.11	12	7
chr10	131502495	131502506	NC_000010.11	12	7
chr11	5557203	5557192	NC_000011.10	12	7
chr11	31242576	31242565	NC_000011.10	12	7
chr11	31419537	31419526	NC_000011.10	12	7
chr11	33169689	33169678	NC_000011.10	12	7
chr11	55948947	55948958	NC_000011.10	12	7
chr11	85654901	85654890	NC_000011.10	12	7
chr11	92588724	92588735	NC_000011.10	12	7
chr11	105227927	105227938	NC_000011.10	12	7
chr11	106302916	106302927	NC_000011.10	12	7
chr11	110594096	110594085	NC_000011.10	12	7
chr11	125228337	125228326	NC_000011.10	12	7
chr12	8769205	8769194	NC_000012.12	12	7
chr12	30438984	30438973	NC_000012.12	12	7
chr12	33556205	33556216	NC_000012.12	12	7
chr12	39419629	39419640	NC_000012.12	12	7
chr12	88509246	88509257	NC_000012.12	12	7
chr12	92878719	92878708	NC_000012.12	12	7
chr12	107034133	107034144	NC_000012.12	12	7
chr12	109455187	109455176	NC_000012.12	12	7
chr13	88827671	88827660	NC_000013.11	12	7
chr14	48126959	48126948	NC_000014.9	12	7
chr20	851155	851166	NC_000020.11	12	7
chr22	17367227	17367238	NC_000022.11	12	7
chrX	20420571	20420582	NC_000023.11	12	7
chrX	22545830	22545819	NC_000023.11	12	7
chrX	28899719	28899708	NC_000023.11	12	7

chrX	33013952	33013963	NC_000023.11	12	7
chrX	103880419	103880408	NC_000023.11	12	7
chrX	105612448	105612459	NC_000023.11	12	7
chrX	107929443	107929454	NC_000023.11	12	7
chrX	112571297	112571286	NC_000023.11	12	7
chrX	127584278	127584267	NC_000023.11	12	7
chrX	130836800	130836811	NC_000023.11	12	7
chrX	140554200	140554211	NC_000023.11	12	7
chrX	146697583	146697572	NC_000023.11	12	7
chr01	97461481	97461469	NC_000001.11	13	6
chr01	104600535	104600547	NC_000001.11	13	6
chr02	12589473	12589461	NC_000002.12	13	6
chr02	187173643	187173631	NC_000002.12	13	6
chr03	29181713	29181701	NC_000003.12	13	6
chr04	32171549	32171537	NC_000004.12	13	6
chr04	116646904	116646916	NC_000004.12	13	6
chr04	164205821	164205833	NC_000004.12	13	6
chr04	170244792	170244780	NC_000004.12	13	6
chr05	15818439	15818451	NC_000005.10	13	6
chr05	174059501	174059489	NC_000005.10	13	6
chr06	94936128	94936140	NC_000006.12	13	6
chr06	98142018	98142006	NC_000006.12	13	6
chr06	151814731	151814719	NC_000006.12	13	6
chr07	14490189	14490201	NC_000007.14	13	6
chr07	53075165	53075153	NC_000007.14	13	6
chr07	87815318	87815306	NC_000007.14	13	6
chr07	103485572	103485584	NC_000007.14	13	6
chr08	35951572	35951560	NC_000008.11	13	6
chr08	39327231	39327219	NC_000008.11	13	6
chr08	69690270	69690258	NC_000008.11	13	6
chr08	117816166	117816154	NC_000008.11	13	6
chr08	123134654	123134666	NC_000008.11	13	6
chr09	68817454	68817466	NC_000009.12	13	6
chr09	68894040	68894028	NC_000009.12	13	6

chr09	80470190	80470178	NC_000009.12	13	6
chr10	1642234	1642222	NC_000010.11	13	6
chr10	73077072	73077060	NC_000010.11	13	6
chr10	110134589	110134601	NC_000010.11	13	6
chr11	9150979	9150991	NC_000011.10	13	6
chr11	9153635	9153623	NC_000011.10	13	6
chr11	13413693	13413705	NC_000011.10	13	6
chr11	41773900	41773912	NC_000011.10	13	6
chr11	77886545	77886557	NC_000011.10	13	6
chr11	79988166	79988154	NC_000011.10	13	6
chr11	108008162	108008150	NC_000011.10	13	6
chr12	30847723	30847735	NC_000012.12	13	6
chr12	86693014	86693026	NC_000012.12	13	6
chr12	122128926	122128914	NC_000012.12	13	6
chr13	29622714	29622726	NC_000013.11	13	6
chr14	39336522	39336510	NC_000014.9	13	6
chr15	94819443	94819431	NC_000015.10	13	6
chr17	10951262	10951250	NC_000017.11	13	6
chr19	30854506	30854518	NC_000019.10	13	6
chr20	42688485	42688473	NC_000020.11	13	6
chrX	38138789	38138801	NC_000023.11	13	6
chrX	86361231	86361243	NC_000023.11	13	6
chrX	107051786	107051798	NC_000023.11	13	6
chrX	109054235	109054247	NC_000023.11	13	6
chr01	32830533	32830522	NC_000001.11	12	6
chr01	56714138	56714127	NC_000001.11	12	6
chr01	79950536	79950547	NC_000001.11	12	6
chr01	81600862	81600851	NC_000001.11	12	6
chr01	88351333	88351344	NC_000001.11	12	6
chr01	100720346	100720335	NC_000001.11	12	6
chr01	103153587	103153598	NC_000001.11	12	6
chr01	163679268	163679279	NC_000001.11	12	6
chr01	178138239	178138228	NC_000001.11	12	6
chr01	202443386	202443375	NC_000001.11	12	6

chr01	214381798	214381787	NC_000001.11	12	6
chr01	239483920	239483909	NC_000001.11	12	6
chr02	5995932	5995921	NC_000002.12	12	6
chr02	14869774	14869785	NC_000002.12	12	6
chr02	37466261	37466250	NC_000002.12	12	6
chr02	38845623	38845634	NC_000002.12	12	6
chr02	38849877	38849866	NC_000002.12	12	6
chr02	52660534	52660523	NC_000002.12	12	6
chr02	55372861	55372872	NC_000002.12	12	6
chr02	62005199	62005210	NC_000002.12	12	6
chr02	70287567	70287556	NC_000002.12	12	6
chr02	79359701	79359712	NC_000002.12	12	6
chr02	84655638	84655627	NC_000002.12	12	6
chr02	126324776	126324787	NC_000002.12	12	6
chr02	149537132	149537143	NC_000002.12	12	6
chr02	169529510	169529521	NC_000002.12	12	6
chr02	175817135	175817124	NC_000002.12	12	6
chr02	180079693	180079682	NC_000002.12	12	6
chr02	206324011	206324000	NC_000002.12	12	6
chr02	206814054	206814043	NC_000002.12	12	6
chr02	224807794	224807805	NC_000002.12	12	6
chr02	229238864	229238853	NC_000002.12	12	6
chr02	236280053	236280064	NC_000002.12	12	6
chr03	154343	154354	NC_000003.12	12	6
chr03	8511973	8511984	NC_000003.12	12	6
chr03	16880365	16880376	NC_000003.12	12	6
chr03	18087857	18087846	NC_000003.12	12	6
chr03	47168148	47168137	NC_000003.12	12	6
chr03	47937628	47937617	NC_000003.12	12	6
chr03	48992978	48992989	NC_000003.12	12	6
chr03	82163078	82163067	NC_000003.12	12	6
chr03	103449909	103449898	NC_000003.12	12	6
chr03	120049593	120049604	NC_000003.12	12	6
chr03	143783076	143783065	NC_000003.12	12	6

chr03	149601763	149601752	NC_000003.12	12	6
chr03	167891194	167891183	NC_000003.12	12	6
chr03	181054638	181054627	NC_000003.12	12	6
chr03	191545181	191545170	NC_000003.12	12	6
chr03	197899144	197899155	NC_000003.12	12	6
chr04	668375	668364	NC_000004.12	12	6
chr04	19382020	19382031	NC_000004.12	12	6
chr04	19484541	19484552	NC_000004.12	12	6
chr04	26997338	26997349	NC_000004.12	12	6
chr04	55658608	55658619	NC_000004.12	12	6
chr04	70437852	70437841	NC_000004.12	12	6
chr04	79981798	79981809	NC_000004.12	12	6
chr04	94968197	94968208	NC_000004.12	12	6
chr04	102674459	102674470	NC_000004.12	12	6
chr04	124485434	124485445	NC_000004.12	12	6
chr04	126123159	126123148	NC_000004.12	12	6
chr04	137124764	137124753	NC_000004.12	12	6
chr04	160702860	160702849	NC_000004.12	12	6
chr04	167052375	167052386	NC_000004.12	12	6
chr04	179139043	179139032	NC_000004.12	12	6
chr04	179161408	179161397	NC_000004.12	12	6
chr04	187143772	187143761	NC_000004.12	12	6
chr05	10200709	10200720	NC_000005.10	12	6
chr05	33225853	33225842	NC_000005.10	12	6
chr05	76255175	76255186	NC_000005.10	12	6
chr05	82855245	82855256	NC_000005.10	12	6
chr05	84139572	84139561	NC_000005.10	12	6
chr05	88198462	88198473	NC_000005.10	12	6
chr05	102501084	102501073	NC_000005.10	12	6
chr05	109583817	109583806	NC_000005.10	12	6
chr05	128180682	128180671	NC_000005.10	12	6
chr05	136190403	136190392	NC_000005.10	12	6
chr05	154189555	154189566	NC_000005.10	12	6
chr05	171957271	171957282	NC_000005.10	12	6

chr05	175317578	175317567	NC_000005.10	12	6
chr06	4853151	4853162	NC_000006.12	12	6
chr06	16133021	16133032	NC_000006.12	12	6
chr06	26103447	26103436	NC_000006.12	12	6
chr06	35947570	35947581	NC_000006.12	12	6
chr06	68279419	68279430	NC_000006.12	12	6
chr06	79806546	79806557	NC_000006.12	12	6
chr06	85260407	85260418	NC_000006.12	12	6
chr06	136633874	136633885	NC_000006.12	12	6
chr06	137931054	137931043	NC_000006.12	12	6
chr06	139739984	139739973	NC_000006.12	12	6
chr06	140341418	140341429	NC_000006.12	12	6
chr06	145869806	145869795	NC_000006.12	12	6
chr06	146731539	146731528	NC_000006.12	12	6
chr06	168728425	168728436	NC_000006.12	12	6
chr07	51771646	51771635	NC_000007.14	12	6
chr07	137082304	137082315	NC_000007.14	12	6
chr07	141052267	141052278	NC_000007.14	12	6
chr08	17556548	17556537	NC_000008.11	12	6
chr08	30097319	30097308	NC_000008.11	12	6
chr08	68502659	68502670	NC_000008.11	12	6
chr08	86697209	86697198	NC_000008.11	12	6
chr08	91622182	91622171	NC_000008.11	12	6
chr08	92498179	92498168	NC_000008.11	12	6
chr08	124481608	124481597	NC_000008.11	12	6
chr08	129563081	129563092	NC_000008.11	12	6
chr08	131305462	131305451	NC_000008.11	12	6
chr09	14627274	14627285	NC_000009.12	12	6
chr09	15151836	15151847	NC_000009.12	12	6
chr09	22322306	22322295	NC_000009.12	12	6
chr09	23783142	23783153	NC_000009.12	12	6
chr09	26318093	26318104	NC_000009.12	12	6
chr09	31054959	31054970	NC_000009.12	12	6
chr09	79007585	79007596	NC_000009.12	12	6

chr09	88239264	88239253	NC_000009.12	12	6
chr09	96543680	96543669	NC_000009.12	12	6
chr09	99112802	99112813	NC_000009.12	12	6
chr09	123836553	123836564	NC_000009.12	12	6
chr10	33633573	33633562	NC_000010.11	12	6
chr10	65551995	65551984	NC_000010.11	12	6
chr10	66717930	66717941	NC_000010.11	12	6
chr10	74291798	74291787	NC_000010.11	12	6
chr10	82621770	82621781	NC_000010.11	12	6
chr10	91090519	91090530	NC_000010.11	12	6
chr10	99682921	99682910	NC_000010.11	12	6
chr10	107653284	107653273	NC_000010.11	12	6
chr10	127387876	127387887	NC_000010.11	12	6
chr11	10330421	10330410	NC_000011.10	12	6
chr11	21052051	21052062	NC_000011.10	12	6
chr11	56948810	56948799	NC_000011.10	12	6
chr11	91992913	91992902	NC_000011.10	12	6
chr11	96712150	96712139	NC_000011.10	12	6
chr11	99478699	99478710	NC_000011.10	12	6
chr11	103284503	103284514	NC_000011.10	12	6
chr11	110624774	110624763	NC_000011.10	12	6
chr11	118226686	118226697	NC_000011.10	12	6
chr11	121927186	121927175	NC_000011.10	12	6
chr11	127371998	127372009	NC_000011.10	12	6
chr12	21742376	21742387	NC_000012.12	12	6
chr12	33375091	33375102	NC_000012.12	12	6
chr12	79305333	79305322	NC_000012.12	12	6
chr12	87018030	87018041	NC_000012.12	12	6
chr12	97027085	97027074	NC_000012.12	12	6
chr12	97030674	97030685	NC_000012.12	12	6
chr12	97794786	97794775	NC_000012.12	12	6
chr12	99326334	99326345	NC_000012.12	12	6
chr12	100617295	100617284	NC_000012.12	12	6
chr12	106997614	106997603	NC_000012.12	12	6

chr12	114419769	114419758	NC_000012.12	12	6
chr13	29428703	29428714	NC_000013.11	12	6
chr13	34838980	34838991	NC_000013.11	12	6
chr13	68672648	68672637	NC_000013.11	12	6
chr13	68677576	68677565	NC_000013.11	12	6
chr13	79534292	79534303	NC_000013.11	12	6
chr13	83374368	83374357	NC_000013.11	12	6
chr13	91208120	91208131	NC_000013.11	12	6
chr13	92057240	92057251	NC_000013.11	12	6
chr13	105912154	105912165	NC_000013.11	12	6
chr14	37970959	37970948	NC_000014.9	12	6
chr14	40492006	40491995	NC_000014.9	12	6
chr14	44782915	44782926	NC_000014.9	12	6
chr14	48758306	48758317	NC_000014.9	12	6
chr14	88004548	88004537	NC_000014.9	12	6
chr15	56610753	56610764	NC_000015.10	12	6
chr15	70757589	70757578	NC_000015.10	12	6
chr15	96964230	96964219	NC_000015.10	12	6
chr16	66442829	66442818	NC_000016.10	12	6
chr16	74623964	74623975	NC_000016.10	12	6
chr16	75189302	75189291	NC_000016.10	12	6
chr17	9332911	9332900	NC_000017.11	12	6
chr18	32474384	32474373	NC_000018.10	12	6
chr18	34128952	34128963	NC_000018.10	12	6
chr18	55039826	55039815	NC_000018.10	12	6
chr18	78931519	78931508	NC_000018.10	12	6
chr19	31065225	31065236	NC_000019.10	12	6
chr19	32434028	32434017	NC_000019.10	12	6
chr19	51221292	51221303	NC_000019.10	12	6
chr20	1361969	1361958	NC_000020.11	12	6
chr20	4448895	4448906	NC_000020.11	12	6
chr20	13696489	13696478	NC_000020.11	12	6
chr20	20275384	20275395	NC_000020.11	12	6
chr20	26367536	26367525	NC_000020.11	12	6

chr21	37223237	37223248	NC_000021.9	12	6
chr21	46496495	46496484	NC_000021.9	12	6
chr22	39560335	39560346	NC_000022.11	12	6
chrX	986645	986656	NC_000023.11	12	6
chrX	5921242	5921253	NC_000023.11	12	6
chrX	6765829	6765840	NC_000023.11	12	6
chrX	15504137	15504126	NC_000023.11	12	6
chrX	22546280	22546269	NC_000023.11	12	6
chrX	41199361	41199372	NC_000023.11	12	6
chrX	43885293	43885282	NC_000023.11	12	6
chrX	67874307	67874296	NC_000023.11	12	6
chrX	110216026	110216037	NC_000023.11	12	6
chrX	110566890	110566879	NC_000023.11	12	6
chrX	111357390	111357379	NC_000023.11	12	6
chrX	150589443	150589454	NC_000023.11	12	6
chr01	23207589	23207577	NC_000001.11	13	5
chr01	25897408	25897420	NC_000001.11	13	5
chr01	65491478	65491490	NC_000001.11	13	5
chr01	154831168	154831180	NC_000001.11	13	5
chr02	35254361	35254349	NC_000002.12	13	5
chr02	207969171	207969159	NC_000002.12	13	5
chr03	185371630	185371642	NC_000003.12	13	5
chr04	46469891	46469879	NC_000004.12	13	5
chr04	105058847	105058835	NC_000004.12	13	5
chr04	124730032	124730044	NC_000004.12	13	5
chr04	158619352	158619364	NC_000004.12	13	5
chr06	85949972	85949960	NC_000006.12	13	5
chr06	109604972	109604960	NC_000006.12	13	5
chr10	59089285	59089273	NC_000010.11	13	5
chr10	99263586	99263598	NC_000010.11	13	5
chr11	96315922	96315934	NC_000011.10	13	5
chr15	33186727	33186715	NC_000015.10	13	5
chr15	87091718	87091706	NC_000015.10	13	5
chr16	16972153	16972165	NC_000016.10	13	5

chr16	59986446	59986458	NC_000016.10	13	5
chr18	12587445	12587457	NC_000018.10	13	5
chr18	78691060	78691048	NC_000018.10	13	5
chr19	39627504	39627492	NC_000019.10	13	5
chr19	54674561	54674573	NC_000019.10	13	5
chr20	30512867	30512855	NC_000020.11	13	5
chr20	45173430	45173442	NC_000020.11	13	5
chr21	35062647	35062659	NC_000021.9	13	5
chrX	77412877	77412889	NC_000023.11	13	5
chrX	130349739	130349727	NC_000023.11	13	5
chr01	8663054	8663065	NC_000001.11	12	5
chr01	26335998	26336009	NC_000001.11	12	5
chr01	42582606	42582595	NC_000001.11	12	5
chr01	47032830	47032819	NC_000001.11	12	5
chr01	69196253	69196264	NC_000001.11	12	5
chr01	70300023	70300034	NC_000001.11	12	5
chr01	82771042	82771053	NC_000001.11	12	5
chr01	100102957	100102946	NC_000001.11	12	5
chr01	107996202	107996213	NC_000001.11	12	5
chr01	162211653	162211642	NC_000001.11	12	5
chr01	208646365	208646354	NC_000001.11	12	5
chr01	215734460	215734449	NC_000001.11	12	5
chr01	234143991	234144002	NC_000001.11	12	5
chr01	241045297	241045286	NC_000001.11	12	5
chr02	140780861	140780872	NC_000002.12	12	5
chr02	149162575	149162586	NC_000002.12	12	5
chr02	162692841	162692852	NC_000002.12	12	5
chr02	222738270	222738259	NC_000002.12	12	5
chr03	67248099	67248110	NC_000003.12	12	5
chr03	174292637	174292648	NC_000003.12	12	5
chr04	12331297	12331308	NC_000004.12	12	5
chr04	21504937	21504948	NC_000004.12	12	5
chr04	43962965	43962976	NC_000004.12	12	5
chr04	57433948	57433937	NC_000004.12	12	5

chr04	85682861	85682872	NC_000004.12	12	5
chr04	106114290	106114301	NC_000004.12	12	5
chr04	113028283	113028294	NC_000004.12	12	5
chr04	151151805	151151794	NC_000004.12	12	5
chr04	152051162	152051173	NC_000004.12	12	5
chr04	179052931	179052920	NC_000004.12	12	5
chr05	6661409	6661420	NC_000005.10	12	5
chr05	93549147	93549158	NC_000005.10	12	5
chr05	148916732	148916721	NC_000005.10	12	5
chr05	153193520	153193531	NC_000005.10	12	5
chr05	169165696	169165685	NC_000005.10	12	5
chr06	99056822	99056833	NC_000006.12	12	5
chr07	21203640	21203651	NC_000007.14	12	5
chr07	27364344	27364355	NC_000007.14	12	5
chr07	45331667	45331656	NC_000007.14	12	5
chr08	28102047	28102036	NC_000008.11	12	5
chr08	64148089	64148078	NC_000008.11	12	5
chr08	121058238	121058249	NC_000008.11	12	5
chr08	134902692	134902681	NC_000008.11	12	5
chr09	26814924	26814935	NC_000009.12	12	5
chr09	35739632	35739643	NC_000009.12	12	5
chr09	77017601	77017612	NC_000009.12	12	5
chr09	83041777	83041788	NC_000009.12	12	5
chr09	87072669	87072658	NC_000009.12	12	5
chr09	134613617	134613628	NC_000009.12	12	5
chr10	7938397	7938408	NC_000010.11	12	5
chr10	59688277	59688266	NC_000010.11	12	5
chr10	91834373	91834384	NC_000010.11	12	5
chr10	106036664	106036653	NC_000010.11	12	5
chr11	1648239	1648228	NC_000011.10	12	5
chr11	28286474	28286485	NC_000011.10	12	5
chr11	59609982	59609993	NC_000011.10	12	5
chr11	82154773	82154784	NC_000011.10	12	5
chr12	56884436	56884425	NC_000012.12	12	5

chr12	65309897	65309908	NC_000012.12	12	5
chr12	70312802	70312791	NC_000012.12	12	5
chr12	108169798	108169809	NC_000012.12	12	5
chr13	41643771	41643782	NC_000013.11	12	5
chr13	43730188	43730177	NC_000013.11	12	5
chr13	66772070	66772081	NC_000013.11	12	5
chr13	67266239	67266250	NC_000013.11	12	5
chr13	70438394	70438405	NC_000013.11	12	5
chr13	72462904	72462915	NC_000013.11	12	5
chr13	73589220	73589209	NC_000013.11	12	5
chr13	114256981	114256970	NC_000013.11	12	5
chr14	53548116	53548105	NC_000014.9	12	5
chr14	91128016	91128005	NC_000014.9	12	5
chr15	55623598	55623609	NC_000015.10	12	5
chr15	59650410	59650421	NC_000015.10	12	5
chr15	67895787	67895798	NC_000015.10	12	5
chr15	75030887	75030898	NC_000015.10	12	5
chr15	80376611	80376600	NC_000015.10	12	5
chr17	2259971	2259960	NC_000017.11	12	5
chr17	13599804	13599793	NC_000017.11	12	5
chr17	49970374	49970385	NC_000017.11	12	5
chr17	74411987	74411998	NC_000017.11	12	5
chr18	6692184	6692173	NC_000018.10	12	5
chr18	26936361	26936372	NC_000018.10	12	5
chr18	32164785	32164796	NC_000018.10	12	5
chr18	57372141	57372152	NC_000018.10	12	5
chr18	76028676	76028665	NC_000018.10	12	5
chr18	79860251	79860240	NC_000018.10	12	5
chr20	2767508	2767497	NC_000020.11	12	5
chr20	32334864	32334853	NC_000020.11	12	5
chr20	42969400	42969411	NC_000020.11	12	5
chr21	15405882	15405871	NC_000021.9	12	5
chr21	27128817	27128828	NC_000021.9	12	5
chr21	27724878	27724889	NC_000021.9	12	5

chr21	33775512	33775523	NC_000021.9	12	5
chr22	40201219	40201208	NC_000022.11	12	5
chrX	24583713	24583724	NC_000023.11	12	5
chrX	53003928	53003939	NC_000023.11	12	5
chrX	75537169	75537180	NC_000023.11	12	5
chrX	91187582	91187593	NC_000023.11	12	5
chr01	237603124	237603136	NC_000001.11	13	4
chr02	132279864	132279852	NC_000002.12	13	4
chr02	176672291	176672279	NC_000002.12	13	4
chr04	47096940	47096952	NC_000004.12	13	4
chr05	170123837	170123825	NC_000005.10	13	4
chr10	97944808	97944796	NC_000010.11	13	4
chr10	114226626	114226614	NC_000010.11	13	4
chr13	67884795	67884783	NC_000013.11	13	4
chr14	59591410	59591398	NC_000014.9	13	4
chr16	3659076	3659088	NC_000016.10	13	4
chr18	25418784	25418772	NC_000018.10	13	4
chrX	45634061	45634049	NC_000023.11	13	4
chr01	3217976	3217987	NC_000001.11	12	4
chr01	92837827	92837816	NC_000001.11	12	4
chr01	112701651	112701662	NC_000001.11	12	4
chr01	166000671	166000660	NC_000001.11	12	4
chr01	178801277	178801288	NC_000001.11	12	4
chr02	177290177	177290166	NC_000002.12	12	4
chr02	218084695	218084706	NC_000002.12	12	4
chr02	236494650	236494639	NC_000002.12	12	4
chr04	42894460	42894471	NC_000004.12	12	4
chr04	66200304	66200315	NC_000004.12	12	4
chr06	35644009	35643998	NC_000006.12	12	4
chr06	35671520	35671531	NC_000006.12	12	4
chr09	95179956	95179945	NC_000009.12	12	4
chr09	122078420	122078431	NC_000009.12	12	4
chr09	132891241	132891252	NC_000009.12	12	4
chr09	134244101	134244112	NC_000009.12	12	4

chr10	46934395	46934384	NC_000010.11	12	4
chr10	48117437	48117448	NC_000010.11	12	4
chr10	102716315	102716304	NC_000010.11	12	4
chr12	31614069	31614080	NC_000012.12	12	4
chr13	18693215	18693204	NC_000013.11	12	4
chr14	30845671	30845660	NC_000014.9	12	4
chr14	94062711	94062722	NC_000014.9	12	4
chr17	10363532	10363543	NC_000017.11	12	4
chr17	59667014	59667025	NC_000017.11	12	4
chr17	68278027	68278038	NC_000017.11	12	4
chr18	44686796	44686785	NC_000018.10	12	4
chr18	55570049	55570060	NC_000018.10	12	4
chr20	37099530	37099519	NC_000020.11	12	4
chr21	14473970	14473981	NC_000021.9	12	4
chr21	28191101	28191112	NC_000021.9	12	4
chr01	85362838	85362827	NC_000001.11	12	3
chr14	106817445	106817434	NC_000014.9	12	3
chr17	12074729	12074718	NC_000017.11	12	3
chr21	8217645	8217657	NC_000021.9	13	2
chr21	8400683	8400695	NC_000021.9	13	2
chr21	8444915	8444927	NC_000021.9	13	2
chr01	65227883	65227872	NC_000001.11	12	2
chr17	31347890	31347879	NC_000017.11	12	1

95. Пример 9. Перенацеливание Gene Writer на геномный сайт "безопасная гавань".

В данном примере описан Gene Writer, содержащий гетерологичный ДНК-связывающий домен, который перенаправляет свою активность на геномный сайт "безопасная гавань".

В данном примере полипептидная последовательность Gene Writer изменена таким образом, что ее природный ДНК-связывающий домен заменен, подвергнут мутации/инактивирован и/или соединен с другой полипептидной последовательностью, которая может перенаправить систему Gene Writer в другое местоположение в геноме, которое не является ее эндогенным или природным сайтом связывания. В некоторых случаях полипептидная последовательность, которая перенаправляет систему Gene Writer к неприродному местоположению в геноме, также могут присоединять к любому модулю полипептидной последовательности Gene Writer и/или вставлять в него.

В некоторых вариантах осуществления полипептидная последовательность, применяемая для перенаправления системы Gene Writer к неприродной мишени в геноме, кодирует следующее: "цинковый палец", серию смежных, регулярно или неравномерно расположенных "цинковых пальцев", эффектор, подобный активатору транскрипции (TALE), серию смежных, регулярно или неравномерно расположенных эффекторов, подобных активатору транскрипции (TALE), Cas9, Cas9 с мутациями по остаткам в его каталитическом домене, которые обеспечивают инактивацию эндонуклеазной активности, специфичной в отношении двухнитевой ДНК (называемую каталитически неактивной Cas9 (dCas9)), Cas9 с точечной мутацией или несколькими точечными мутациями в одном каталитическом домене, обеспечивающими преобразование эндонуклеазы Cas9 в способную расщеплять только одну нить двухнитевой ДНК (называемую нисказой Cas9) (см. фиг. 12).

В некоторых вариантах осуществления полипептидная последовательность, применяемая для перенаправления системы Gene Writer, нацелена на местоположение геномной "безопасной гавани" (например, сайт AAVS1 на хромосоме 19 человека) (Pellenz, S., et. al., Human Gene Therapy, 30(7), 814-828, 2019), см. фигуры 11 и 13. В дополнительных вариантах осуществления полипептидная последовательность, применяемая для перенаправления полипептидной последовательности Gene Writer, применяется в сочетании с нуклеиновой кислотой, которая нацеливается на местоположение "безопасной гавани" в геноме (например, полипептидная последовательность для каталитически мертвого Cas9 вместе с единичной направляющей РНК, нацеливающейся на сайт AAVS1 в хромосоме 19).

96. Таблица Е4. Перенацеленные конструкции Gene Writer. Показанные примеры предназначены для перенацеливания полипептидной последовательности R2Tg Gene Writer на сайт AAVS1 с применением доменов ZF или Cas9

Название полипептида Gene Writer	Полипептидная последовательность (полипептидная последовательность для перенацеливания, выделено курсивом; <u>линкер</u> , подчеркнуто, выделено полужирным шрифтом)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный в местополо	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSGGGSTATRDKKDTVTRKHPKPFQKWMKDRAIKKGNYLRFQRLFYLDRGK</i> <i>LAKIILDDIECLSCDIPLSEIYSVKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV</i> <i>QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSK</i> <i>PDRLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIWS</i> <i>AKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT</i> <i>HTDKIQIRVGVKQGDPMSPLLFNAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLL</i>

жении v2 DBD R2Tg с помощью линкера 3GS	SDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMI DPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYADH SEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSTRDGGGLITKLAGLIPSVQARRLHRI AQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLT ALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPDIFVKDARALVVDVTVRYEAAKSSLEEAEEK VRKYKHLETEVRHLTNAKDVTFGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFS TVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1658)
Специфичн ый в отношении AAVS1 левый ZFP, прикрепле нный в местополо жении v2 DBD R2Tg с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSSGSETPGTSESATPE</i>STATRDKKDVTREKHPKPFQKWMKDRAIKKGNLYR FQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAF RELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMV RGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGC SENKLLQTIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNM YENISTYITTKRNTHTDKIQRVGVKQGDPMSPLLFNLAMDPDLLCKLEESGKGYHRGQSSI TAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDI LKTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSTRDGGGLITKLA GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQ DARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDIFVKDARALVVDVTVRYE AAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFGFPLGARGKWHQDNFKLLTELGLS KSRQVKMAETFS TVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1659)
Специфичн ый в отношении AAVS1 левый ZFP, прикрепле нный в	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSGGGSC</i>FGCLESISQIRTATRDKKDVTREKHPKPFQKWMKDRAIKKGNLYR FQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAF RELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMV RGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGC SENKLLQTIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNM

местополо жении v1 DBD R2Tg с помощью линкера 3GS	YENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSI TAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDI LKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSTRDGGGLGITKLA GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQ DARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYE AAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKLLTELGLS KSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1660)
Специфичн ый в отношении AAVS1 левый ZFP, прикрепле нный в местополо жении v1 DBD R2Tg с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH TKIHLRGSSGSETPGTSESATPES</i> CFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMK DRAIKKGNLYRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVKTRWETTGSFKSLGDFK TYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLW LTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPR QRGFIRAAGCSENKLLQTIWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVPD HIVGLVSNM YENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGK GYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSTRD GGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFE RDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARAL VVDVTVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1661)
Специфичн ый в отношении AAVS1 правый	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYL DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT KIHLRGSGGGST</i> ATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRFQRLFYLD RGKL AKIILDDIECLSCDIPLSEIYSVKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSK

ZFP, прикрепле нный в местополо жении v2 DBD R2Tg с помощью линкера 3GS	PDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIWS AKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT HTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLL SDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMI DPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYADH SEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGGGLITKLAGLIPSVQARRLHRI AQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQQYRRIPHRKLLT ALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPDIFVKDARALVVDVTVRYEAAKSSLEEAAAEK VRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFS TVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1662)
Специфичн ый в отношении AAVS1 правый ZFP, прикрепле нный в местополо жении v2 DBD R2Tg с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYL</i> <i>DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT</i> <i>KIHLRGSSGSETPGTSESATPES</i> TATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVR GCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCS ENLKLQTIWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSIT AMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAA WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDIL KTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGGGLITKLA GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQ DARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDIFVKDARALVVDVTVRYE AAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLS KSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1663)
Специфичн ый в отношении AAVS1	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYL</i> <i>DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT</i> <i>KIHLRGSGGGS</i> CFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR

правый ZFP, прикрепле нный в местополо жении v1 DBD R2Tg с помощью линкера 3GS	ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVR GCRTVLIPKSSKPDRLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRGFIRAAGCS ENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPDLLCKLEESGKG YHRGQSSIT AMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYINDCAA WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDL KTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGGLGITKLA GLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQ DARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYE AAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLS KSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1664)
Специфичн ый в отношении AAVS1 правый ZFP, прикрепле нный в местополо жении v1 DBD R2Tg с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYL V DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT KIHRLGSS</i> <u>SGSETPGTSESATPES</u> <i>CFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKD RAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWL TTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQ RGFIRAAGCSENKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPH IVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPDLLCKLEESGKG YHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKD SYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPL KPLQKTDLKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDG GLGITKLGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSI WEAPPSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFER DKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHI IGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALV VDVTVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNF KLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1665)</i>
Специфичн ые в отношении	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH TKIHRLGSS</i> <u>SGSETPGTSESATPES</u> <i>GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHT</i>

AAVS1 левый и правый ZFP (разделенн ые линкером XTEN), прикрепле нные в местополо жении v2 DBD R2Tg с помощью линкера 3GS	<i>TGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEK</i> <i>PFACDICGRKFAQGYNLAGHTKIHLRGSGGGSTATRDKKDVTREKHPKKPFQKWMKD</i> <i>RAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT</i> <i>YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWL</i> <i>TTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQ</i> <i>RGFIRAAGCSENKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPH</i> <i>IVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPDLLCKLEESGKG</i> <i>YHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKD</i> <i>SYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPL</i> <i>KPLQKTDILKTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSTRDG</i> <i>GLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSI</i> <i>WEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFER</i> <i>DKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHI</i> <i>IGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARALV</i> <i>VDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNF</i> <i>KLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1666)</i>
Специфичн ые в отношении AAVS1 левый и правый ZFP (разделенн ые линкером XTEN), прикрепле нные в местополо жении v2 DBD R2Tg с помощью	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSSGSETPGTSESATPESGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTH</i> <i>TGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEK</i> <i>PFACDICGRKFAQGYNLAGHTKIHLRGSSGSETPGTSESATPESTATRDKKDVTREKHP</i> <i>KKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWET</i> <i>TGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEF</i> <i>SRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTAR</i> <i>LSKACPLNPRQRGFIRAAGCSENKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHII</i> <i>HALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAM</i> <i>DPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQG</i> <i>QKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKL</i> <i>DFWLQRIDQAPLKPLQKTDILKTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPPC</i> <i>TCDAILYSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLW</i> <i>IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKL</i> <i>ASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC</i> <i>RHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYK</i>

линкера XTEN	PDLIFVKDARALVVDVTVRYEAAKSSLEEEAAAEKVRKYKHLETEVRHLTNAKDVT FVGFP L GARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1667)
Специфичн ый в отношении AAVS1 левый ZFP, прикрепле нный к N- концу R2Tg с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSSGSETPGTSESATPES</i> ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLA NSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKH RVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAHWEGLS GLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTG EEKVMVTVDPKNPPCPCGTRVNSVLNIEHLKVSHGKRGVCFRCAKCGKENSNYHSVV CHFPCRCRGPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAAEPR EETPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRH PDKGNGQQRPTKQKSEELQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQT AQDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLD R GKLAKIILDDIECLSDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEK NVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKS SKPDRCLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTII WSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTK RNTHTDKIQIRVGVKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSITAMAFADDL VLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPL NMIDPGESEKYLGLQFDPWIGIARSG LSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYI ADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGG LGITKLAGLIPSVQARR LHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTN SEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHR KLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNY ICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEEAA AEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAE TFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1668)
Специфичн ый в отношении AAVS1	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLV</i> <i>DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT</i> <i>KIHLRGSSGSETPGTSESATPES</i> ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLAN SGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV

правый ZFP, прикрепле нный к N- концу R2Tg с помощью линкера XTEN	DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVECVHFAAHWEGLSGL PEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEE KVMVTPDKNPPCPCGTRVNSVLNIEHLKVSHGKRGVCFRCAKCGKENSNYHSVVC HFPKCRGPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPR EEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRLPLGRAAGGRTMDEIRRH PDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFQVMEGRDIKLVINQT AQDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRFQRLFYLDR GKLAKIILDDIECLSDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEK NVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKS SKPDRCLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTII WSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTK RNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDL VLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPL NMIDPGESEKYLGLQFDPWIGIARSLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYI ADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARR LHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTN SEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHR KLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNY ICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVDVTVRYEAAKSSLEEAA AEKVRKYKHLETEVRHLTNAKDVTFGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAE TFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1669)
Специфичн ые в отношении AAVS1 левый и правый ZFP, прикрепле нные к N- концу R2Tg с помощью	MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH TKIHLRGSS<u>SGSETPGTSESATPES</u>GIIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTH TGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEK PFACDICGRKFAQGYNLAGHTKIHLRGSS<u>SGSETPGTSESATPES</u>ASCPKPGPPVSAGAMS LESGLTTHSVLAIERGPNSLANS GSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSH PGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYR SDLHVECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESE DAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNIEHLKVSHGKRG VCFRCAKCGKENSNYHSVVC HFPKCRGPETEKAPAGEWICEVCNRDFTTKIGLGQHKRL AHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTK

линкера XTEN	TAKQISDKRRLLSRKPAEEPREPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRR PLPGRAAAGGRTMDEIRRHDPKNGGQQRPTKQKSEEQQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMK DRAIKKGNLYRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFK TYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLW LTTGKIPDMVRGCRTVLIPKSSKPDRDKDINNWRPITIGSILLRFSRIVTARLSKACPLNPR QRGFIRAAGCSENKLLQTIWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDP HIVGLVSNMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDPDLLCKLEESGK GYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDLKTYTIPRLIYADHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSTRD GGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFE RDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1670)
Специфичн ый в отношении AAVS1 левый ZFP, прикрепле нный к N- концу R2Tg, содержаще го мутации инактиваци и DBD, с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL</i> <i>TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH</i> <i>TKIHLRGSSGSETPGTSESATPES</i>ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLSLA NSGSDFGGGGLGLPLRLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKH RVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAHWEGLS GLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTG EEKVMVTVDPKNPPSPSSGTRVNSVLNIEHLKVSHGKRGVCFRCAKCGKENSNYHSV CHFPKCRGPETEKAPAGEWISEVSNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKACATKEEEEELLIRLEAQFEGNKNINKLIAEHITTKAKQISDKRRLLSRKPAEEPRE EPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRH DPKNGGQQRPTKQKSEEQQLQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQTA QDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSS KPDRDKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIIW

	SAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRN THTDKIQIRVGVKQGDPMSPLLFNAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVL LSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLN MIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYA DHSEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRL HRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNS EWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCADIESCAHIIGNCPVTQDARIKRHNYI CELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAA AEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAE TFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1671)
Специфичн ый в отношении AAVS1 правый ZFP, прикрепле нный к N- концу R2Tg, содержаще го мутации инактиваци и DBD, с помощью линкера XTEN	<i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYL</i> <i>DHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHT</i> <i>KIHLRGSS</i><u>SGSETPGTSESATPES</u><i>ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLAN</i> <i>SGSDFGGGGLGLPLRLLRVSVGTQTSRSDWDLVSWSHPGPTSKSQQVDLVSLFPKHRV</i> <i>DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGL</i> <i>PEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEE</i> <i>KVMVTPDKNPPSPSSGTRVNSVNLNIEHLKVSHGKRGVCFRCAKCGKENSNYHSVCH</i> <i>FPKCRGPETEKAPAGEWISEVSNRDTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNR</i> <i>GAHKACATKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEP</i> <i>GTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRHDPK</i> <i>GNGQQRPTKQKSEELQAYYKKTLEERLSAGALNTFPRAFQVMEGRDIKLVINQTAQD</i> <i>CFGCLESISQIRTATRDKKDVTREKHPKPFQKWMKDRAIKGNYLRFQRLFYLDRGKL</i> <i>AKIILDDIECLSCDIPLSEIYSVKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV</i> <i>QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSK</i> <i>PDRLKDINNWRPITIGSILLRFSRIVTARLSKACPLNPRQRFIRAAGCSENKLLQTIIWS</i> <i>AKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT</i> <i>HTDKIQIRVGVKQGDPMSPLLFNAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVL</i> <i>SDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMI</i> <i>DPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADH</i> <i>SEVKTALLETLDQKIRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRI</i> <i>AQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW</i> <i>EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLT</i>

	ALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAAEK VRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFS TVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1672)
Специфичн ые в отношении AAVS1 левый и правый ZFP, прикрепле нные к N- концу R2Tg, содержаще го мутации инактиваци и DBD, с помощью линкера XTEN	MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFARSDHL TTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGH TKIHLRGSSGSETPGTSESATPESGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTH TGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEK PFACDICGRKFAQGYNLAGHTKIHLRGSSGSETPGTSESATPESASCPKPGPPVSAGAMS LESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSH PGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYR SDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESE DAPKTSPPTPGKHGLEQTGEEKVMVTVDPKNPPSPSSGTRVNSVLNIEHLKVSHGKRG VCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGEWISEVSNRDFTTKIGLGQHKRL AHPAVRNQERIVASQPKETSNRGAKACATKEEEELLIRLEAQFEGNKNINKLIAEHITTKT AKQISDKRRLLSRKPAEEPREPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRP LPGRAAAGGRTMDEIRRHDPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNFTP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKPPFQKWMK DRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVKTRWETTGSFKSLGDFK TYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLW LTTGKIPDMVRGCRTVLIPKSSKPDR LKDINNWRPITIGSILLR LFSRIVTARLSKACPLNPR QRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDP HIVGLVSNMYENISTYITTKRNTHTDKIQRVGVKQGDPMSPLLFN LAMDPLLCKLEESGK GYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRD GGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWT KLASQGRGIVNFE RDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVT FVG FPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFS TVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1673)

Cas9	S.	MAPKKRKRKVGIVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIK
pyogenes,		KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSSFFHRLEESFL
прикрепле		VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHF
нная	в	LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
местополо		GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFL
жении	v2	AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQ
DBD R2Tg с		SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
помощью		GELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF
линкера		EEVVDKGASAQSFIERMNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
XTEN33aa		FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
		KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR
		LSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHE
		HIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM
		KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIV
		PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
		KAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV
		SDFRKDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVVRKMIA
		KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
		VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
		VAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIILPKYSLFELENG
		RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII
		EQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
		KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD <u>SGGSSGGSSGSETPGTSESATPESSGG</u>
		<u>SSGGS</u> STATRDKKDTV TREKHPKKPFQKWMKDRAIKKGN YLRFQRLFYLD RGLAKIILD
		DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK
		GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKD
		INNWRPITIGSILLR LFSRIVTARLSKACPLNPRQRGFIRAAGCSEN LKLLQTIIWSAKREHR
		PLGVVFVDIAKAFTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI
		RVGVKQGDPMSPLLFN LAMP LLCKLEESGKG YHRGQSSITAMAFADDLVLLSDSWEN
		MNTNISILETFCNL TGLKTQGQKCHGFYIKPTKDSY TINDCAAWTINGTPLNMIDPGESE
		KYLGLQFDPWIGIARSG LSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA
		LLETLDQKIRTAVKEWLHLP PCTCDAILYSTRDGG LGITKLAGLIPSVQARRLHRIAQSSD
		DTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ

		KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEA AAEKVRKYK HLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1674)
Cas9 <i>pyogenes</i> , содержаща я каталитиче ские мутации (dCas9), прикрепле нная в местополо жении v2 DBD R2Tg с помощью линкера XTEN33aa	S.	MAPKKRKRKVG I HGVPAADKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHF LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII KDKDFLDNEENEDILEDIVLT LTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM KRIE EG I KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIV PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV SDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLIILPKYSLFELENG RKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD <u>SGGSSGGSSGSETPGTSESATPESSGG</u> <u>SSGGS</u> STATRDKDTVTREKHPKPFQKWMKDRAIKKGN YLRFQRLFYLD RGLAKIILD DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPDRLKD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIIWSAKREHR PLGVV FVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI

	RVGVKQGDPMSPLLFNAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSLSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPCTCDAILYSSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSD DTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYRRIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCADADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVDVTVRYEAAKSSLEEEAAEKVRKYK HLETEVRHLTNAKDVTFFVGFLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1675)
Мутантная форма никазы Cas9 D10A <i>S. pyogenes</i> , прикрепле нная в местополо жении v2 DBD R2Tg с помощью линкера XTEN33aa	MAPKKRKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM KRIEEDIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGELSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV SDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKDLIIKLPKYSLEFELNG RKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD <u>SGGSSGGSSGSETPGTSESATPESSGG</u>

	SSGGSTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILD DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIWSAKREHR PLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSGSLSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGG LGITKLAGLIPSVQARRLHRIAQSSD DTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEA AAEKVRKYK HLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1676)
Мутантная форма никазы Cas9 N863A <i>S. pyogenes</i> , прикрепле нная в местополо жении v2 DBD R2Tg с помощью линкера XTEN33aa	MAPKKKRKVG I HGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHF LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFWMTRKSEETITPWNF EEVVDKGASAQSFIERMNTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK

	VLSMPQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDR KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD <u>SSGSSSGSSGSETPGTSESATPESSGG</u> <u>SSGSS</u> TATRDKKD TVTREKHPKKPFQKWMKDRAIKKGN YLRFQRLFYLD RGLAKIILD DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIWSAKREHR PLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFN LAMDPLLCKLEESGKG YHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSY TINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSG LSTKLDFWLQRIDQAPLKPLQKT DILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGG LGITKLAGLIPSVQARRLHRIAQSSD DTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEA AAEKVRKYK HLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1677)
Мутантная форма никазы Cas9 D10A <i>S. pyogenes</i>, прикрепле нная в местополо жении v2 DBD R2Tg, содержаще й мутацию EN, с	MAPKKKRKVGIHGVPAAADKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHF LIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLP GEKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRF AWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLT LTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFL KSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERM

помощью линкера XTEN33aa	<i>KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV</i> <i>PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT</i> <i>KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV</i> <i>SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIA</i> <i>KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK</i> <i>VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV</i> <i>VAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKDLIIKLPKYSLFELNG</i> <i>RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII</i> <i>EQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDR</i> <i>KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD</i><u>SGGSSGGSSGSETPGTSESATPESSGG</u> <u>SSGGS</u><i>STATRD KDTVTREKHPKPFQKWMKDRAIKKGN YLRFQRLFYLD RGLAKIILD</i> <i>DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK</i> <i>GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKD</i> <i>INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGCSENKLLQTIWSAKREHR</i> <i>PLGVVFDIAKAFTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI</i> <i>RVGVKQGDPMSPLLFNAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN</i> <i>MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE</i> <i>KYLGLQFDPWIGIARSGSLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA</i> <i>LLETLDQKIRTAVKEWLHLPCTCDAILYSSTRDGG LGITKLAGLIPSVQARRLHRIAQSSD</i> <i>DTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ</i> <i>KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLR</i> <i>ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA</i> <i>KKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRYEAAKSSLEEA AAEKVRKYK</i> <i>HLETEVRHLTNAKDVT FVG FPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF</i> <i>SSVDIVHMFASRARKSMVM (SEQ ID NO: 1678)</i>
Мутантная форма никазы Cas9 N863A <i>S. pyogenes</i> , прикрепле нная в местополо	<i>MAPKKKRKVG I HGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK</i> <i>KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHRLEESFL</i> <i>VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHF</i> <i>LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLP</i> <i>GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFL</i> <i>AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQ</i> <i>SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL</i> <i>GELHAILRRQEDFY PFLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF</i>

жении v2 DBD R2Tg, содержаще й мутацию EN, с помощью линкера XTEN33aa	EEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV SDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKPKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD <u>SGGSSGGSSGSETPGTSESATPESSGG</u> <u>SSGGS</u>STATRDKKDTV TREKHPKPFQKWMKDRAIKKGNLRFQRLFYLD RGKLAKIILD DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRKLD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGCSENKLLQTIIWSAKREHR PLGVVFDIAKAFTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSLSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYADHSEVKTA LLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGG LGITKLAGLIPSVQARRLHRIAQSSD DTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRRIPIHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRYEAAKSSLEEA AEKVRKYK HLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1679)
Cas9 pyogenes, прикрепле	S. MAPKKKRKVG I HGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFL VEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHF

нная	в	LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLP
местополо		GEKKNGLFGNLIASLSGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFL
жении	v2	AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
DBD	R2Tg,	SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
содержаще		GELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF
го мутацию		EEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPA
EN,	с	FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
помощью		KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR
линкера		LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE
XTEN33aa		HIANLAGSPAIKKGILQTVKVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM KRIEEDIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKDLIIKLPKYSLFELENG RKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKKHYLDEII EQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD <u>SGGSSGGSSGSETPGTSESATPESSGG</u> <u>SSGGS</u>STATRDKKDTV TREKHPKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILD DIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRKLD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTIIWSAKREHR PLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFN LAMP LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSGLSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGLGITKLAGLIPSVQARRLHRIAQSSD DTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYK

	HLETEVRHLTNAKDVTFTVGFPGLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1680)
--	--

97. Пример 10. Инактивация эндогенного сигнала ядрышковой локализации в Gene Writer

В данном примере описан Gene Writer, в котором эндогенный сигнал ядрышковой локализации был инактивирован для уменьшения уровня внутриклеточного нацеливания белка на ядрышко.

В данном примере сигнал ядрышковой локализации (NoLS) ретротранспозазы предсказан с помощью вычислений с использованием опубликованного алгоритма, тренированного на валидированных белках, которые локализуются в ядрышке (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388-7399 (2010)). Предсказанная последовательность NoLS определяется на основании как аминокислотной последовательности, окружения аминокислотной последовательности, так и предсказанной вторичной структуры ретротранспозазы. Идентифицированная последовательность, как правило, богата основными аминокислотами (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388-7399 (2010)), и если эти остатки подвергнуть мутации, обеспечивающей замену на неосновные аминокислоты с простыми боковыми цепями или удаление их из полипептидной цепи, то это может предотвратить локализацию в ядрышке (Yang, C. P., et. al., Journal of Biomedical Science, 22(1), 1-15. (2015), Martin, R. M., et. al., Nucleus, 6(4), 314-325 (2015)). В некоторых вариантах осуществления последовательность NoLS расположена в аминокислотной области ретротранспозазы, которая находится между полимеразным мотивом обратной транскриптазы и мотивом, подобным рестрикционной эндонуклеазе. Предсказанная область NoLS содержит аминокислоты лизин, аргинин, гистидин и/или глутамин, при этом сигнал ядрышковой локализации инактивируют посредством мутации одного или нескольких из этих остатков по типу замены на остаток аминокислоты аланин и/или удаления из полипептидной цепи ретротранспозазы одной или нескольких из данных аминокислот. В некоторых вариантах осуществления аминокислотная последовательность управляющего элемента Gene Writer R2Tg, находящаяся выше RLE, подвергнута мутации таким образом, что остатки лизина (K) заменены на остатки аланина (A), например, предсказанный NoLS R2Tg (аминокислоты 1128-1154 из полипептидной последовательности) (APTQKDKFPKPCNWRKNEFKKWTCLAS (SEQ ID NO: 1681)) подвергается мутации по 1, 2, 3, 4, 5, 6 или 7 остаткам с образованием инактивированного NoLS (APTQADAFFAPCNWRANEFAAWTALAS (SEQ ID NO: 1682)).

98. Пример 11. Применение внесения однострессового разрыва во вторую нить в системе Gene Writer

В данном примере описана система Gene Writer, в которой ретротранспозиция сочетается с целенаправленной активностью внесения однострессового разрыва во вторую нить для увеличения эффективности событий интеграции. Однострессовый разрыв второй нити может быть получен с помощью (1) никазы Cas9, слитой с системой Gene Writer, в

которой Gene Writer вводит один одонитевой разрыв с помощью своего эндонуклеазного домена (EN), а слитая никаза Cas9 размещает другой одонитевой разрыв либо на верхней, либо на нижней нити ДНК (фиг. 7A), или (2) системы Gene Writer, в которой активный домен EN вводит одонитевой разрыв, а никаза Cas9 вводит второй одонитевой разрыв либо на верхней, либо на нижней нити ДНК, выше или ниже от индуцированного Gene Writer одонитевого разрыва (фиг. 7B).

В первой части данного примера никаза Cas9 слита с белком Gene Writer (фиг. 7A). Cas9 нацелена на последовательность ДНК с помощью gRNA. Белок Gene Writer вводит одонитевой разрыв в ДНК с помощью своего домена EN, а дополнительный одонитевой разрыв создается за счет никазной активности Cas9. Этот дополнительный одонитевой разрыв может быть нацелен на верхнюю или нижнюю нити ДНК, окружающие введенный с помощью Gene Writer одонитевой разрыв (фиг. 8A). Конструкции, разработанные и протестированные, включают (см. схематическую фигуру 14A):

Cas9-N863A-R2tg (RBD*, RT, EN);

Cas9-H840A-R2tg (RBD*, RT, EN);

Cas9-D10A-R2tg (RBD*, RT, EN);

dCas9- R2tg (RBD*, RT, EN).

ДНК-связывающий домен представляет собой никазу Cas9, которая направляет молекулу Gene Writer к ДНК-мишени посредством gRNA. РНК-связывающий домен (RBD) в данном наборе конструкций Gene Writer инактивирован точечной мутацией (RBD*). В качестве донора для вставки конструкции, в которых РНК-связывающий домен R2Tg является неактивным, используют gRNA, удлиненную на своем 3'-конце, чтобы включить донорную последовательность для модификации генома (фиг. 14B). Такие модификации включают замены нуклеотидов, делеции нуклеотидов и вставки нуклеотидов. В данной первой серии экспериментов вышеуказанные конструкции-R2Tg(RBD*, RT, EN) и dCas9-R2Tg(RBD*, RT, EN), являющиеся продуктами слияния с gRNA-матрицей, удлиненной на 3'-конце, нацеливающейся на локус AAVS1, доставляют в клетки U2OS посредством нуклеофекции в буфере SE с применением программы DN100. Применяемые gRNA включают gRNA для каждой конструкции, которые нацелены либо на нижнюю, либо на верхнюю нить ДНК. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят секвенирование ампликона с последующим вычислительным анализом с применением CRISPResso (инструмент для анализа вставок/делеций). Вставки, делеции или замены нуклеотидов, опосредованные gRNA, удлиненной на 3'-конце, наблюдаются при доставке dCas9-R2Tg(RBD*, RT, EN), и их частота увеличивается при доставке конструкций никаза-Cas9-R2Tg(RBD*, RT, EN).

Во второй части данного примера никаза Cas9 слита с белком Gene Writer (фиг. 7A). Cas9 нацелена на последовательность ДНК с помощью gRNA. Белок Gene Writer вводит одонитевой разрыв в ДНК с помощью своего домена EN, а дополнительный одонитевой разрыв создается за счет никазной активности Cas9. Такой дополнительный одонитевой разрыв может быть нацелен на верхнюю или нижнюю нити ДНК, окружающие

однонитевой разрыв, введенный Gene Writer (фиг. 7A). В отличие от конструкций, перечисленных выше, РНК-связывающий домен R2Tg является активным (фиг. 15A), а матрица, используемая для модификации генома, представляет собой трансген, фланкированный областями UTR (фиг. 15B). Конструкции включают (см. схематическую фигуру 15A):

Cas9-N863A-R2tg (RBD, RT, EN);

Cas9-H840A-R2tg (RBD, RT, EN);

Cas9-D10A-R2tg (RBD, RT, EN);

dCas9- R2tg (RBD, RT, EN).

Для трансгена, фланкированного UTR, необходимо наличие плеч гомологии в сайте введения однонитевого разрыва. Перечисленные выше конструкции подвергают нуклеофекции в 200 тыс. клеток U2OS с помощью gRNA, нацеливающейся на локус AAVS1, с использованием импульсного кода DN100, чтобы определить сайт введения однонитевого разрыва для обеспечения точной разработки плеч гомологии для трансгенной ДНК-донора. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят секвенирование ампликона с последующим вычислительным анализом с применением CRISPResso в качестве инструмента для анализа вставок/делений. Сайт однонитевого разрыва, осуществляемого доменом EN, идентифицируют по вставкам/делениям, которые домен EN осуществляет в сайте AAVS1. Плечи гомологии длиной 100 п. о., фланкирующие сайт однонитевого разрыва, осуществляемого EN, разрабатывают и включают в трансген. Для достижения модификации генома, осуществляют нуклеофекцию слитых конструкций на основе Cas9-R2Tg, перечисленных выше, в клетки U2OS вместе с gRNA, нацеливающейся либо на верхнюю, либо на нижнюю нить локуса AAVS1, и соответствующими трансгенами, характеризующимися гомологией с ранее определенным сайтом однонитевого разрыва. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят ddPCR для обнаружения интеграции трансгена в сайте AAVS1. Интеграции наблюдаются при доставке dCas9-R2Tg(RBD, RT, EN), и их частота увеличивается при доставке конструкций Никаза-Cas9-R2Tg(RBD, RT, EN).

В другом примере белок Gene Writer нацелен на ДНК посредством своего ДНК-связывающего домена (фиг. 7B). Белок Gene Writer вводит однонитевой разрыв ДНК в нити ДНК. Кроме того, никазу Cas9 применяют для создания второго однонитевого разрыва либо на верхней, либо на нижней нитях ДНК, выше или ниже первого однонитевого разрыва. В этом примере плазида Gene Writer, нацеленная на сайт AAVS1 (фиг. 16A) и с фланкированным UTR трансгеном с гомологией к сайту AAVS1 (фиг. 16B), подвергается нуклеофекции в 200 тыс. клеток U2OS с использованием импульсного кода DN100. Следующие конструкции Cas9 трансфицировали вместе с плазмидами Gene Writer (фиг. 16C):

Cas9-N863A

Cas9-H840A

Cas9-D10A

dCas9

Нуклеофекцию всех конструкций на основе Cas9 осуществляют совместно с gRNA, нацеливающимися на локус AAVS1, находящийся либо на верхней, либо на нижней нитях, выше или ниже одонитевого разрыва, введенного Gene Writer. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят ddPCR для обнаружения интеграции трансгена в сайте AAVS1. Интеграции наблюдаются при доставке dCas9, и их частота увеличивается при доставке конструкций на основе никазы-Cas9.

99. Пример 12. Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR

В данном примере описано применение гетерологичных UTR для усиления внутриклеточной экспрессии полипептида Gene Writer.

В этом примере полипептид Gene Writer экспрессировали из мРНК (фигура 17). В плазмидной матрице для получения мРНК нативные UTR ретротранспозона заменяли с помощью UTR, оптимизированных для экспрессии белка (C3 5'UTR и ORM 3'UTR из Asrani et al., RNA Biology 15, 756-762 (2018) или 5' и 3' UTR из Richter et al., Cell 168, 1114-1125 (2017)). Плазмида содержала промотор T7, за которым следовала 5'UTR, кодирующая последовательность ретротранспозона, 3'UTR, линкер 3GS (SEQ ID NO: 1024), сигнал ядерной локализации SV40 (NLS), линкер XTEN, последовательность HiBit и поли(A)-хвост длиной 96-100 нуклеотидов (SEQ ID NO: 1683). Плазмиду линейаризовали с помощью ферментативной рестрикции с получением тупого конца или выступающего 5'-конца ниже поли(A)-хвоста и применяли для транскрипции *in vitro* (IVT) с применением полимеразы T7 (NEB). После стадии IVT РНК обрабатывали ДНКазой I (NEB). После стадии замены буфера проводили ферментативную реакцию кэпирования с применением фермента кэпирования вируса коровьей оспы (NEB) и 2'-О-метилтрансферазы (NEB) в присутствии GTP и SAM (NEB). Кэпированную РНК концентрировали и заменяли буфер. 50000 клеток HEK293T трансфицировали с помощью 0,5 мкг мРНК Gene Writer в присутствии или в отсутствие РНК-матрицы в молярном соотношении 1:1 с применением системы трансфекции Neon (1150 В на импульс, 20 мс на импульс, 2 импульса в наконечниках по 10 мкл в 96-луночном формате). РНК-матрица была транскрибирована *in vitro* из плазмиды, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК").

После трансфекции клетки HEK293T выращивали в течение 5 часов перед анализом экспрессии Gene Writer с помощью исследования экспрессии его HiBit-тега, используя стандартный протокол <https://www.promega.com/-/media/files/resources/protocols/technical-manuals/500/nano-glo-hibit-lytic-detection-system-technical-manual.pdf?la=en>. Было

обнаружено, что экспрессия белка значительно улучшается при использовании 5'- и 3'-UTR_{exp} из C3-ORM по сравнению с использованием нативных UTR из R2Tg (фигура 17).

Интеграцию генома анализировали через 3 дня после трансфекции с использованием 3'-ddPCR (фигура 18).

100. **Пример 13. Улучшенные компоненты Gene Writer для смешанной доставки РНК и ДНК**

В данном примере описаны улучшения молекулы РНК, кодирующей полипептид Gene Writer, которые обеспечивают усиление экспрессии и повышение эффективности ретротранспозиции при использовании с матрицей Gene Writer, кодирующейся в плазмидной ДНК.

В этом примере полипептидный компонент системы Gene Writer экспрессируется из мРНК, описанной в примере 12 ("Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR"). Плазмидную матрицу синтезировали таким образом, чтобы репортерный ген (eGFP) был фланкирован нетранслируемыми областями (UTR) R2Tg и содержал последовательность длиной 100 п. о., гомологичную его рДНК-мишени. Экспрессия матрицы управлялась промотором CMV млекопитающих. Авторы настоящего изобретения вводили плазмиду в клетки HEK393T с помощью реагента для трансфекции FuGENE® HD. Клетки HEK293T высевали в 96-луночные планшеты при 10000 клеток/луночка за 24 часа до трансфекции. В день трансфекции 0,5 мкл реагента для трансфекции и 80 нг ДНК смешивали в 10 мкл Opti-MEM и инкубировали в течение 15 минут при комнатной температуре. Затем смесь для трансфекции добавляли в среду с посеянными клетками. Клетки отделяли и применяли для электропорации 0,5 мкг мРНК на лунку с применением системы трансфекции Neon (1150 В на импульс, 20 мс на импульс, 2 импульса в наконечниках по 10 мкл в формате 96 лунок).

Клетки HEK293T трансфицировали следующими тестируемыми средствами:

1. мРНК, кодирующая полипептид, описанный выше;
2. Плазида, кодирующая РНК-матрицу, описанную выше;
3. Комбинация 1 и 2. Плазмиду предварительно подвергали липофекции за 24 ч до трансфекции мРНК, описанной выше.

После трансфекции клетки HEK293T культивировали в течение 1-3 дней и затем анализировали в отношении сайт-специфического редактирования генома. Геномную ДНК выделяли из каждой группы клеток HEK293.

Проводили ddPCR для подтверждения интеграции и оценки эффективности интеграции. Зонды и праймеры Taqman разрабатывали, как описано в PCT/US 2019/048607, для амплификации ожидаемого продукта, простирающегося от 5'-конца к 3'-концу мест соединения интегрированного элемента. Результаты анализа числа копий методом ddPCR (по сравнению с эталонным геном RPP30) показаны на фигуре 19. Интеграция в геном в присутствии мРНК и плазмиды с матрицей достигала среднего числа копий, составляющего 0,683 интегранта/геном при нацеливании на место соединения на 3'-конце, и составляющего 0,249 интегранта/геном при нацеливании на место соединения на 5'-конце. Трансфекция только мРНК приводила к среднему числу копий, составлявшему 0,002 интегранта/геном, по сравнению с 0,0004 интегранта/геном при трансфекции только плазмидой.

101. Пример 14. Улучшенные компоненты Gene Writer для доставки на основе РНК

В данном примере описаны улучшения молекулы РНК, кодирующей полипептид Gene Writer, которые обеспечивают усиление экспрессии и повышение эффективности ретротранспозиции при совместной доставке с РНК-матрицей Gene Writer.

В этом примере полипептидный компонент системы Gene Writer экспрессируется из мРНК, описанной в примере 12 ("Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR"). Плазмидная матрица для получения РНК-матрицы содержала промотор Т7, за которым следовал репортерный ген, экспрессирующий IRES (eGFP), фланкированный нетранслируемыми областями (UTR) R2Tg, и последовательность длиной в 100 п. о., гомологичная его рДНК-мишени. Матрицу плазмиды линейаризовали с помощью ферментативной рестрикции с получением тупого конца или выступающего 5'-конца ниже последовательности РНК-матрицы и применяли для транскрипции *in vitro* (IVT) с применением РНК-полимеразы Т7 (NEB). После стадии IVT РНК обрабатывали ДНКазой I (NEB) и либо ферментативно полиаденилировали поли(А)-полимеразой (NEB), либо нет. После стадии замены буфера проводили ферментативную реакцию кэпирования с применением фермента кэпирования вируса коровьей оспы (NEB) и 2'-О-метилтрансферазы (NEB) в присутствии GTP и SAM (NEB). Кэпированную РНК концентрировали и заменяли буфер. 50000 клеток HEK293T котрансфицировали с помощью 0,5-1 мкг мРНК GeneWriter и РНК-матрицы при молярных соотношениях от 1:4 до 1:12. Систему трансфекции Neon применяли для трансфекции РНК (1150 В на импульс, 20 мс на импульс, 2 импульса в наконечниках по 10 мкл в 96-луночном формате).

После трансфекции клетки HEK293T культивировали в течение по меньшей мере 1 дня и затем анализировали в отношении сайт-специфического редактирования генома. Геномную ДНК выделяли из каждой группы клеток HEK293.

Проводили ddPCR для подтверждения интеграции и оценки эффективности интеграции. Зонды и праймеры Taqman разрабатывали, как описано в PCT/US 2019/048607, для амплификации ожидаемого продукта, простирающегося от 5'-конца к 3'-концу мест соединения интегрированного элемента. Среднего числа копий, составляющего 0,498 интегрантов/геном, достигали в присутствии 0,5 мкг мРНК и при молярном соотношении мРНК Gene Writer и РНК-матрицы, составляющем 1:8, если РНК-матрица была ферментативно полиаденилированной, по сравнению со средним числом копий, составляющим 0,031 интегрантов/геном, если РНК трансгена не была полиаденилированной.

102. Пример 15. Gene Writer, которые обеспечивают доставку генетического груза, содержащего интроны

В данном примере описано применение системы Gene Writer для интеграции генетического груза, содержащего интроны, посредством применения доставки на основе РНК для настройки экспрессии гена, представляющего интерес, из его недавно введенного геномного локуса.

В данном примере в технологии Gene Writing применяется РНК-матрица, кодирующая белок, представляющий интерес, содержащий его нативные или ненативные интроны. Например, интрон 6 гена триозофосфатизомеразы (TPI) (Nott et al., 2003) будут применять в качестве одного из ненативных интронов в данных экспериментах.

Сообщалось, что присутствие интронов в геномной копии гена и их удаление посредством сплайсинга влияет почти на все аспекты экспрессии гена, включая скорость его транскрипции, процессинг мРНК, экспорт, локализацию в клетке, трансляцию и распад (рассмотрено в Shaul International Journal of Biochemistry and Cell Biology 91B, 145-155 (2017)). Интроны могут быть вставлены в разные части РНК-матрицы (фигура 21), и в зависимости от расположения интрона их роль в экспрессии генов может различаться.

Интрон в 5'-UTR_{exp}, расположенный рядом с сайтом начала транскрипции, вводит активирующие модификации хроматина (Bieberstein et al., Cell Reports 2, 62-68 (2012)), улучшает точность распознавания сайта начала транскрипции и облегчает рекрутинг PolII (Laxa et al., Plant Physiology 172, 313-327 (2016)), увеличивает скорость инициации транскрипции (Kwek et al., Nature Structural Biology 9, 800-805 (2002)) и элонгации (Lin et al., Nature Structural and Molecular Biology 15, 819-826 (2008)), а также улучшает продуктивную элонгацию в смысловой ориентации относительно антисмысловой ориентации (Almada et al., Nature 499, 360-363 (2013)).

Интрон в 3'-UTR_{exp} ограничивает экспрессию мРНК до одной молекулы белка на мРНК: комплекс соединения экзонов (EJC), оставленный сплайсосома ниже стоп-кодона, распознается механизмом нонсенс-опосредованного распада (NMD), и поэтому мРНК помечается для удаления в конце первого раунда трансляции (Zhang et al., RNA 4, 801-815 (1998)).

Однако возможность применения интронов в терапевтическом гене может быть ограничена сплайсингом, который происходит до интеграции матрицы. Например, интрон с прямой ориентацией будут подвергаться сплайсингу, если РНК-матрица кодируется и доставляется в ДНК-плазмиде, поскольку транскрипция в том же направлении дает РНК-матрицу, которая будет подвергаться сплайсингу до интеграции, таким образом, не встраивая интрон в геном. Кроме того, конструкции на основе лентивируса, предназначенные для доставки трансгена, должны кодировать последовательность с интроном в обратной ориентации, поскольку процесс вирусной упаковки приведет к сплайсингу интрона и отсутствию интрона в упакованных вирусных частицах (Miller et al. J Virol 62, 4337-45 (1988)). Однако также считается, что обратная ориентация приводит к снижению титра и трансдукции вируса (Uchida et al., Nat Commun 10, 4479 (2019)). Стоит отметить, что, поскольку матрицу Gene Writer можно создать посредством транскрипции *in vitro* и доставить непосредственно в виде РНК, то проблемы сплайсинга желаемых интронов перед интеграцией можно избежать. Таким образом, в некоторых вариантах осуществления матрица Gene Writer может содержать один или несколько интронов в одинаковой смысловой ориентации с транскриптом, при этом такую матрицу создают с помощью IVT и доставляют в клетку-мишень в виде РНК.

Интрон в любом расположении, показанном на фигуре 21, будет привлекать snRNP U1, который защищает мРНК от преждевременного расщепления и полиаденилирования (Kaida et al., Nature 468 664-681 (2010); Berg et al., Cell 150, 53-64 (2012)). Кроме того, EJC взаимодействует с компонентами комплекса TREX (транскрипция-экспорт) и обеспечивает увеличение скорости экспорта мРНК из ядра в цитоплазму в 6-10 раз по сравнению с конструкциями без интронов (Valencia et al., PNAS 105, 3386-3391 (2008)). Было также продемонстрировано, что связывание белка, связывающего полипиримидиновый тракт, белка-регулятора сплайсинга, опосредует значительное увеличение периода полужизни подвергнутых сплайсингу транскриптов (Lu & Cullen, RNA 9, 618-630 (2003); Millevoi et al., Nucleic Acid Research 37, 4672-4683 (2009)). Было показано, что эффективность трансляции мРНК увеличивается в присутствии белков SR (богатые серином и аргинином белки, участвующие в сплайсинге РНК) (Sanford et al., Genes & Development 18, 755-768 (2004); Sato et al., Molecular Cell 29, 255-262 (2008)) и белков EJC и их периферических факторов (Nott et al., Genes & Development 18, 210-222 (2004)).

В этом примере как РНК-матрица, несущая интрон или интроны, так и полипептид Gene Writer доставляются в клетки в виде транскрибированных *in vitro* экпированных РНК, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"). Через один-три дня после трансфекции оценивают экспрессию GOI и его интеграцию в геном. В некоторых вариантах осуществления интеграция в геном и/или экспрессия белка будут выше для РНК-матрицы, содержащей интроны.

103. Пример 16. Разработка 5' UTR ретротранспозона для улучшения эффективности интеграции

В данном примере описана делеция, замена или мутация 5'UTR ретротранспозона для повышения эффективности интеграции.

Область 5'UTR ретротранспозонов, не содержащих LTR, выполняет несколько функций, включая саморасщепляющую рибозимную активность, которая была показана в некоторых элементах и предсказана в дополнительных ретротранспозонах (см. модули В и С на фигурах 27-28) (Ruminski et al. J Biol Chem 286, 41286-41295 (2011)). Предполагается, что рибозимная активность обеспечивает расщепление РНК внутри или выше 5'UTR. Либо увеличение, либо ограничение данной активности и структурного компонента 5'UTR может повышать эффективность ретротранспозиции. Предсказание структуры рибозима R2Tg представлено на фигуре 29.

Для оценки конструирования 5'UTR были разработаны конструкции для усиления или ослабления этих активностей (фигура 20). В случае (А) природный 5'UTR R2Tg применяют для интеграции в транс-положении, как и в предыдущих экспериментах. Случай (В) иллюстрирует делецию 5'UTR. (С) и (D) представляют случаи, когда 5'UTR от исходных видов (в данном случае R2Tg из *T. guttata*) были заменены на 5'UTR ретротранспозона от других видов. Случай (С) представляет собой пример, в котором 5'UTR из *A. maritima* R2 заменили на 5'UTR из R2Tg. Случай (D) представляет собой общий случай, в котором могут быть заменены UTR от дополнительных видов ("Rx"), например, от *B. mori*, *D. ananasse*, *F.*

auricularia, *L. polyphemus*, *N. Giraulti* или *O. latipes*, или из ретротранспозона, выбранного из приведенной в данном документе таблицы, или любого ретротранспозона из таблиц 1-3 документа PCT/US2019/048607, включенного в данный документ посредством ссылки. Случай (E) представляет замену рибозима, такого как рибозим типа "головка молота", например, RiboJ (Lou et al Nat Biotechnol 30, 1137-1142 (2012)). Случай (F) представляет собой инактивацию 5'UTR R2Tg посредством точечных мутаций, например, 75C>T в 5'UTR (фигура 20B, положение обозначено заштрихованным прямоугольником). Ожидается, что последовательности 5'UTR будут модульными для любой последовательности вставки, опосредованной ретротранспозоном.

Каждый случай оценивают, как и в предыдущих примерах, с помощью трансфекции плазмиды с полипептидом Gene Writer с плазмидой с матрицей и оценки частоты интеграции с помощью ddPCR. В некоторых вариантах осуществления замена или мутация 5'-UTR приводит к увеличению эффективности интеграции.

104. Пример 17. Модификация 5'- и 3'-концов РНК-компонентов Gene Writer для улучшения стабильности РНК

В данном примере описано добавление некодирующих последовательностей к 5'- и 3'-концам РНК с целью улучшения стабильности в клетке млекопитающего.

Распад эукариотических РНК в клетках в основном осуществляется экзорибонуклеазами. В данном примере период полужизни РНК продлевают за счет введения защитных последовательностей и/или модификаций на их 5'- и 3'-концах. Наиболее распространенным природным способом защиты концов РНК является введение структуры 5'-кэп и 3'-поли(А)-хвоста. В этом примере полипептидный компонент системы Gene Writer экспрессируется из мРНК, описанной в примере 12 ("Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR"). Плазмидная матрица для получения РНК-матрицы содержала промотор Т7, за которым следовал репортерный ген, экспрессирующий IRES (eGFP), фланкированный нетранслируемыми областями (UTR) R2Tg, и последовательность длиной в 100 п. о., гомологичная его рДНК-мишени. Плазмидную матрицу линеаризовали с помощью ферментативной рестрикции с получением тупого конца или выступающего 5'-конца ниже последовательности РНК-матрицы и применяли для транскрипции *in vitro* (IVT) с применением полимеразы Т7 (NEB). После стадии IVT РНК обрабатывали ДНКазой I (NEB) и либо ферментативно полиаденилировали поли(А)-полимеразой (NEB), либо нет. После этапа замены буфера проводили или не проводили реакцию ферментативного кэпирования, приводящую к структуре кэпа 1, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"). РНК-матрицу концентрировали и заменяли буфер. 50000 клеток НЕК293Т котрансфицировали с помощью 0,5 мкг мРНК Gene Writer и РНК-матрицы при молярных соотношениях от 1:1 до 1:8 с применением системы трансфекции Neon (1150 В на импульс, 20 мс на импульс, 2 импульса в наконечниках по 10 мкл в 96-луночном формате)

После трансфекции клетки НЕК293Т культивировали в течение 1-3 дней и затем анализировали в отношении сайт-специфического редактирования генома. Геномную ДНК выделяли из каждой группы клеток НЕК293.

Проводили ddPCR для подтверждения интеграции и оценки эффективности интеграции. Зонды и праймеры Taqman разрабатывали, как описано в PCT/US 2019/048607, для амплификации ожидаемого продукта, простирающегося через 3'-конец мест соединения интегрированного элемента. Интеграция генома была улучшена при использовании матрицы с ферментативным экпированием и поли(А)-хвостом (фигура 22).

Среднего числа копий, составляющего 0,498 интегрантов/геном, достигали в присутствии 0,5 мкг мРНК и при молярном соотношении мРНК и РНК-матрицы, составляющем 1:8, если РНК-матрица была ферментативно полиаденилированной, по сравнению со средним числом копий, составляющим 0,031 интегрантов/геном, если РНК трансгена не была ферментативно полиаденилированной.

Модификации РНК на 3'-конце

Сообщалось, что взаимодействия между поли(А)-хвостом, длиной меньше 15-20 нт, и поли(А)-связывающим белком (PABP) дестабилизируются, что приводит к быстрому разрушению РНК (Chang et al., *Molecular Cell* 53, 1044-1052 (2014); Subtelny et al., *Nature* 508, 66-71 (2014)). Для определения подходящей длины поли(А)-хвоста РНК-матрицы, авторы настоящего изобретения будут тестировать ее длину, составляющую 30, 40, 50, 60, 70, 80, 90 и 100 нуклеотидов. Матрицы для IVT будут получать с помощью ПЦР с применением праймеров, кодирующих поли(А)-хвосты вышеуказанной длины. IVT, обработку ДНКазой I и экпирование Gene Writer и РНК-матрицы будут проводить, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"). Через один-три дня после трансфекции будут анализировать интеграцию генома. В некоторых вариантах осуществления уровень интеграции в геном будет выше для матрицы РНК с поли(А)-хвостом подходящей длины.

В клетке разрушение РНК инициируется укорачиванием ее поли(А)-хвоста деаденилазами. Поскольку деаденилазы представляют собой 3'-5'-экзорибонуклеазы, предпочитающие поли(А)-отрезки, для защиты поли(А)-хвоста от его укорочения были предложены концевые остатки, представляющие собой уридин, цитидин и чаще всего гуанин, обнаруживаемые в природных поли(А)-хвостах многих мРНК (Chang et al., *Molecular Cell* 53, 1044-1052 (2014)). Авторы настоящего изобретения проанализируют Gene Writer и РНК-матрицы с кодируемым поли(А)-хвостом с концевыми G или C, или нерегулярно расположенными остатками G или C (аналогично тем, которые применяли в Lim et al., *Science* 361, 701-704 (2018)) в соответствии с описанным ранее.

Было описано, что у некоторых РНК развились альтернативные пути защиты своих 3'-концов. Сообщается, что специфическая структура "стебель-петля" длиной 16 нуклеотидов, фланкированная непарными 5 нуклеотидами с обеих сторон, защищает 3'-конец мРНК, кодирующей гистон H2a.X (Mannironi et al., *Nucleic Acid Research* 17, 9113-9126 (1989)). Было показано, что гетерологичная мРНК, оканчивающаяся гистоновой

структурой "стебель-петля", регулируется клеточным циклом (Harris et al., *Molecular Cellular Biology* 11, 2416-2424 (1991); Stauber et al., *EMBO Journal* 5, 3297-3303 (1986)). Структура "стебель-петля" распознается и защищается белком, связывающим структуру "стебель-петля" (SLBP). Белок накапливается незадолго до того, как клетки вступают в S-фазу, и быстро разрушается в конце S-фазы (Whiffeld et al., *Molecular Cellular Biology* 20, 4188-4198 (2000)). Элемент "стебель-петля" будут вставлять в 3'-конец мРНК Gene Writer и РНК-матриц и тестировать, как описано выше, в отношении индукции событий интеграции в геном, специфичных для клеточного цикла.

Некоторые вирусные и длинные некодирующие РНК в результате эволюции защищают свои 3'-концы с помощью тройной спиральной структуры (Brown et al., *PNAS* 109, 19202-19207 (2012)). Кроме того, сообщается, что структурные элементы тРНК, Y-РНК и РНК Vault (рассмотренные в Labno et al., *Biochemica et Biophysica Acta* 1863, 3125-3147 (2016)) продлевают период полужизни данных некодирующих РНК. Авторы настоящего изобретения вставят структуры для защиты 3'-конца РНК-матриц и проверят их эффективность в системе Gene Writing, описанной выше.

Наконец, авторы настоящего изобретения включают dNTP, 2'-О-метилированные NTP или фосфотиоат-NTP в 3'-конец РНК транскриптов для увеличения периода полужизни данных молекул за счет защиты 3'-конца РНК от экзорибонуклеаз. Авторы настоящего изобретения включают одиночные модифицированные нуклеотиды или их отрезки путем удлинения 3'-конца РНК с помощью ДНК-полимераз (например, фрагмент Кленова), способных удлинять последовательность РНК посредством добавления модифицированных нуклеотидов (Shcherbakova & Brenowitz, *Nature Protocols* 3, 288-302 (2008)).

Химическая модификация одного нуклеотида 3'-конца РНК может быть осуществлена посредством первого окисления 3'-конца сахара рибозы перйодатом натрия с образованием реакционноспособного альдегида с последующей конъюгацией альдегид-реакционноспособного модифицированного нуклеотида.

В качестве альтернативы ДНК-лигазы T4 или РНК-лигазы T4 могут применять для лигирования с помощью шунта (Moore & Query, *Methods in Enzymology* 317, 109-123 (2000)) отрезков модифицированных нуклеотидов с 3'-концом РНК.

Возможно также химическое лигирование двух фрагментов. Фосфодиэфирная связь между двумя субстратами РНК может быть образована либо посредством активации сложной фосфомоноэфирной группы с применением реакционноспособного имидазолида, либо с помощью конденсирующего реагента, такого как цианогенбромид. Недостатком химического лигирования является то, что оно также может привести к образованию сложной 2'-5'-фосфодиэфирной связи вместе с образованием желаемых 3'-5' фосфодиэфирных связей.

Модификации РНК на 5'-конце

В дополнение к структуре кэпа 1, описанной в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"), будут исследованы другие группы

защиты 5'-конца. В частности, авторы настоящего изобретения будут применять гиперметилованные (Wurth et al. *Nucleic Acid Res* 42, 8663-8677 (2014)), фосфотиоатные (Kuhn et al., *Gene Therapy* 17, 961-971 (2010)), NAD⁺-производные (Kiledjian, *Trends in Cell Biology* 28, 454-464 (2018)) и модифицированные (например, биотинилированные: Bednarek et al., *Phil Trans R Soc B* 373, 20180167 (2018)) аналоги кэпа для котранскрипционного кэпирования.

Авторы настоящего изобретения также будут метить 5'-конец РНК с помощью 5'-[γ-тио]трифосфата для создания реакционноспособной группы серы, и химически модифицировать 5'-конец с помощью защитных модификаций, применяя галоацетамидное производное модифицированной группы.

Предлагаемые модификации для защиты 3'- и 5'-концов РНК будут введены в РНК-матрицы и/или мРНК Gene Writer (если они совместимы с трансляцией). Эффективность интеграции РНК с геномом будет протестирована, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК").

105. Пример 18. Применение модифицированных оснований РНК в системе Gene Writer

В данном примере описаны системы Gene Writer, содержащие модифицированные основания РНК для потенциального улучшения признаков системы, например, повышения эффективности интеграции, снижения клеточного ответа на чужеродные нуклеиновые кислоты. Для полипептида Gene Writer предлагаемые модификации, относящиеся к кодирующей области, являются совместимыми с трансляцией. Для РНК-матрицы предлагаемые модификации совместимы с обратной транскрипцией.

В данном примере мРНК, кодирующую полипептид Gene Writer, транскрибировали *in vitro* со 100% заменой соответствующего гНТФ на один из модифицированных гНТФ: псевдоурин (Ψ), 1-N-метилпсевдоурин (1-Me-Ψ), 5-метоксиурин (5-MO-U) или 5-метилцитидин (5mC). В противном случае получение, очистку РНК и трансфекции ими клеток проводили, как описано в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"). Способность генной интеграции модифицированных мРНК сравнивали с таковой способностью немодифицированной мРНК (G0) с использованием ddPCR, при этом все полипептидные мРНК были соединены с немодифицированной РНК-матрицей (фигура 23). Интеграцию обнаруживали, в случае когда полипептид кодировали с применением каждого модифицированного гНТФ, при этом наиболее высокий сигнал исходил от 5-MO-U, а наиболее низкий от 5mC. Это демонстрирует, что полипептидный компонент Gene Writer является функциональным при экспрессии из мРНК, содержащей модифицированные основания.

Дополнительно в данном примере описана модульность молекулы-матрицы Gene Writer, где она состоит из всех или подмножества иллюстративных модулей, перечисленных на фигуре 6 и проиллюстрированных на фигуре 5. Отдельные модули могут быть получены посредством химического синтеза или синтеза *in vitro* в виде непрерывной молекулы нуклеиновой кислоты или в виде отдельных фрагментов, которые впоследствии

объединяются вместе. Отдельные модули молекулы-матрицы Gene Writer могут представлять собой химически модифицированные нуклеиновые кислоты, могут состоять частично или полностью из компонентов, отличных от нуклеиновых кислот, характеризоваться измененным порядком расположения и/или быть пропущенными при образовании молекулы-матрицы Gene Writer.

В некоторых вариантах осуществления молекулу-матрицу Gene Writer (все модули, A-F) синтезируют с помощью транскрипции *in vitro*, где 0-100% замен соответствующего rNTP (аденозин, цитидин, гуанозин и/или уридин) осуществляется одним или несколькими модифицированными rNTP (модификация основания или рибозы), например, 5'-гидроксилом, 5'-фосфатом, 2'-О-метилом, 2'-О-этилом, 2'-фтором, риботимидином, C-5-пропинил-dC (pdC), C-5-пропинил-dU (pdU), C-5-пропинил-C (pC), C-5-пропинил-U (pU), 5-метил-C, 5-метил-U, 5-метил-dC, 5-метил-dU-метокси, (2,6-диаминопурин), 5'-диметокситритил-N4-этил-2'-дезоксцитидином, C-5 пропинил-fC (pfC), C-5-пропинил-fU (pfU), 5-метил-fC, 5-метил-fU, C-5-пропинил-mC (pmC), C-5-пропинил-fU (pmU), 5-метил-mC, 5-метил-mU, LNA (запертой нуклеиновой кислотой), MGB (средством, связывающимся с малой бороздкой), псевдоуридином (Ψ), 1-N-метилпсевдоуридином (1-Me-Ψ), 5-метоксиуридином (5-MO-U). В случае модифицированных нуклеотидов в данном варианте осуществления используют включение посредством реакции транскрипции, в которой применяется природная или мутантная полипептидная последовательность РНК-полимеразы, которая легко включает модифицированные нуклеотиды в транскрипт РНК, получаемый *in vitro* (Padilla, R., *Nucleic Acids Research*, 30(24), 138e - 138, 2002; Ibach, J., et. al., *Journal of Biotechnology*, 167(3), 287-295, 2013; Meyer, A. J., et. al., *Nucleic Acids Research*, 43(15), 7480-7488, 2015). Модифицированная молекула-матрица Gene Writer обычно полностью или частично совместима с активностью обратной транскриптазы полипептидной последовательности Gene Writer; для модулей или частей модулей молекулы-матрицы Gene Writer, применяемой в качестве матрицы для обратной транскрипции, предпочтение отдается модификациям, совместимым с обратной транскрипцией (Motorin et al., *Methods in Enzymology* 425 21-53, 2007; Mauger et al., *PNAS* 116, 24075-24083, 2019). Системы Gene Writer с молекулами-матрицами, содержащими модифицированные rNTP, тестируют, как описано выше и в примере 14 ("Улучшенные компоненты Gene Writer для доставки на основе РНК").

В некоторых вариантах осуществления химическим путем синтезируют отдельные модули, содержащие модифицированные нуклеотиды, например, 5'-гидроксил, 5'-фосфат, 2'-О-метил, 2'-О-этил, 2'-фтор, риботимидин, C-5-пропинил-dC (pdC), C-5 пропинил-dU (pdU), C-5 пропинил-C (pC), C-5 пропинил-U (pU), 5-метил-C, 5-метил-U, 5-метил-dC, 5-метил-dU-метокси, (2,6-диаминопурин), 5'-диметокситритил-N4-этил-2'-дезоксцитидин, C-5 пропинил-fC (pfC), C-5 пропинил-fU (pfU), 5-метил-fC, 5-метил-fU, C-5-пропинил-mC (pmC), C-5-пропинил-fU (pmU), 5-метил-mC, 5-метил-mU, LNA (запертую нуклеиновую кислоту), MGB (связывающее вещество малой бороздки) псевдоуридин (Ψ), 1-N-метилпсевдоуридин (1-Me-Ψ), 5-метоксиуридин (5-MO-U), где отдельные модули затем

лигируют ферментативными (например, лигирование с помощью шунта с применением ДНК-лигазы T4, Moore, M. J., & Query, C. C. *Methods in Enzymology*, 317, 109-123, 2000) или химическими способами (например, Fedorova, O. A., et. al., *Nucleosides and Nucleotides*, 15(6), 1137-1147, 1996) для создания полной молекулы-матрицы Gene Writer.

Примером модифицированной молекулы-матрицы Gene Writer является молекула, где модули A и F представляют собой по 100 нуклеотидов химически синтезированной РНК с цитидиновыми и уридиновыми нуклеотидами, содержащими модификации в виде 2'-O-метилрибозы, и при этом модуль A содержит (3) фосфотиоатные связи между первыми 3 нуклеотидами на 5'-конце и модуль F содержит (3) фосфотиоатные связи между последними 3 нуклеотидами на 3'-конце модуля. Модули B-E синтезируют с помощью транскрипции *in vitro* с применением РНК-полимеразы (RNAP), например, T7 RNAP, T3 RNAP или SP6 RNAP (NEB), или ее производных, которые обладают улучшенными свойствами, например, повышенной точностью, повышенной процессивностью или повышенной эффективностью включения модифицированных нуклеотидов. Модуль A лигируют к 5'-концу молекулы модуля B-E, транскрибируемого *in vitro*, при этом модуль F лигируют к 3'-концу молекулы модуля B-E, транскрибируемого *in vitro*, посредством лигирования с помощью шунта (описанного в Moore, M. J., & Query, C. C. *Methods in Enzymology*, 317, 109-123, 2000). Такую полностью собранную РНК-матрицу (все модули, A-F) затем применяют с полипептидом Gene Writer (или нуклеиновой кислотой, кодирующей полипептид) в клетке-мишени для оценки интеграции в геном, как в предыдущих примерах. В некоторых вариантах осуществления модификации РНК не снижают эффективность интеграции более чем на 50%, например, как измерено с помощью ddPCR. В некоторых вариантах осуществления модификации РНК улучшают эффективность интеграции, например, как измерено с помощью ddPCR. В некоторых вариантах осуществления модификации РНК улучшают реакцию обратной транскрипции, например, улучшают процессивность или точность, измеренные с помощью секвенирования событий интеграции.

106. Пример 19. Матрицы Gene Writer, которые не содержат UTR

В данном примере описана конфигурация молекулы-матрицы Gene Writer, которая приводит к исключению UTR, так что эти области, применяемые в ретротранспозиции, не интегрируются в клетку-хозяина.

В данном примере авторы настоящего изобретения описывают изменение расположения, пропуск и/или замену модулей UTR молекулы-матрицы Gene Writer (фигуры 5 и 6), приводящие к тому, что управляющий элемент Gene Writer не включает модули UTR в геном в качестве части ретротранспозиции. В некоторых вариантах осуществления модули молекулы-матрицы Gene Writer для 5'- и 3'-UTR (модули B+C и E молекулы-матрицы Gene Writer) перемещают к концам молекулы, так что их функция взаимодействия с управляющим элементом Gene Writer не изменяется, но плечо гомологии теперь расположено рядом с последовательностью гетерологичного объекта (модуль D), где комплементарность плеч гомологии действует в качестве праймера для обратной

транскрипции. В некоторых случаях модули В и/или С пропущены в молекуле-матрице Gene Writer, при этом модуль Е следует за модулем F.

Дополнительными примерами не включения UTR в геном являются удаление модулей В и С из молекулы-матрицы Gene Writer, изменение положения модуля F (3'-плечо гомологии) таким образом, чтобы он следовал за модулем D (последовательностью гетерологичного объекта), и замена модуля Е на связывающий лиганд, такой как биотин. Данная молекула-матрица Gene Writer в данном случае будет состоять из модуля А (5'-плеча гомологии) - модуля D (последовательности гетерологичного объекта) - модуля F (3'-плеча гомологии) - модуля Е, состоящего из биотина. Полипептидную последовательность управляющего элемента Gene Writer будут модифицировать для включения аминокислотной последовательности мономерного стрептавидина. Данный пример иллюстрирует, как целесообразно опосредование ассоциации, опосредованной компонентом, отличным от нуклеиновой кислоты, молекулы-матрицы Gene Writer с полипептидной последовательностью управляющего элемента Gene Writer.

107. Пример 20. Gene Writer могут интегрировать генетический груз независимо от пути репарации, направляемого гомологией.

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации путем гомологичной рекомбинации является ингибированным.

В данном примере клетки U2OS обрабатывали с помощью 30 пмоль (1,5 мкМ) ненацеленной контрольной siRNA (Ctrl) или siRNA к Rad51, основному компоненту пути репарации путем гомологичной рекомбинации. SiRNA доставляли совместно с плазмидой с управляющим элементом R2Tg и плазмидой с трансгеном в *транс*-положении (см. схему конфигурации управляющего элемента и трансгена на фиг. 24). В частности, в сочетании с трансгеном применяли плазмиду, экспрессирующую R2Tg, контрольный R2Tg с мутацией в домене RT или контрольный R2Tg с мутацией, обеспечивающей инактивацию эндонуклеазы (фиг. 25 А, В). Всего 250 нг ДНК-плазмид с молярным соотношением управляющего элемента к трансгену, составляющим 1:4, вместе с 30 пмоль siRNA подвергали нуклеофекции в 200 тысяч клеток U2OS, ресуспендированных в 20 мкл буфера для нуклеофекции SE с применением импульсного кода DN100. Белковые лизаты, собранные на 3-й день, показали отсутствие Rad51 в условиях обработки siRad51 (фиг. 25С). gDNA экстрагировали в день 3, и проводили анализы ddPCR для обнаружения интеграции трансгена в локусе рДНК. Результаты анализа числа копий методом ddPCR (по сравнению с эталонным геном RPP30) показаны на фигуре 26. Отсутствие Rad51 приводит к снижению примерно на 20% опосредованной R2Tg интеграции трансгена в локус рДНК как в 3'-, так и в 5'-месте соединения (фиг. 26), что указывает на то, что опосредованная R2TG вставка трансгена не полностью зависит от присутствия пути гомологичной рекомбинации и может происходить в отсутствие эндогенного пути HR. В некоторых вариантах осуществления независимость от HR позволяет системе Gene Writing работать в клетках и тканях с эндогенно низкими уровнями HR, например, в печени, головном мозге, сетчатке, мышечной, костной, нервной ткани, клетках в фазе G₀ или G₁, неделящихся клетках,

стареющих клетках, терминально дифференцированных клетках. В некоторых вариантах осуществления независимость от HR позволяет системе Gene Writing работать в клетках или у пациентов или в тканях, содержащих клетки с мутациями в генах, участвующих в пути HR, например, BRCA1, BRCA2, P53, RAD51.

108. Пример 21. Gene Writer могут интегрировать генетический груз независимо от пути репарации посредством однострессовой матрицы

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации посредством однострессовой матрицы (SSR) является ингибированным.

В таком примере путь SSR будут ингибировать с помощью siRNA к основным компонентам пути: FANCA, FANCD2, FANCE, USP1. Также будут включены контрольные siRNA 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 обнаруживают по отсутствию основных компонентов пути SSR.

109. Пример 22. Системы Gene Writer с усиленной активностью в отношении клеток-мишеней по сравнению с клетками, не являющимися мишенями

В данном примере описывают включение регуляторных последовательностей в системы Gene Writer с целью снижения интеграционной активности в клетках, не являющихся мишенями.

В данном примере генетическая регуляция осуществляется за счет (i) применения тканеспецифических промоторов для усиления экспрессии и интеграции компонентов в клетках-мишенях и (ii) применения miRNA-связывающих сайтов для уменьшения интеграции в клетки, не являющиеся мишенями, которые характеризуются увеличенными эндогенными уровнями соответствующих miRNA. Применяемые клетки-мишени, представляют собой гепатоциты человека, а клетки, не являющиеся мишенями, представляют собой гемопоэтические стволовые клетки (HSC). Управляющим элементом интеграции в данном случае является плазида, кодирующая полипептид Gene Writer (например, ретротранспозаза R2Tg), управляемая различными промоторами и с скремблированными или специфическими miRNA-связывающими сайтами после кодирующей последовательности. Матрицу для интеграции кодируют в плазмидной ДНК, так что транскрипция приводит к образованию последовательности гетерологичного объекта, фланкированной последовательностями гомологии и UTR. Последовательность гетерологичного объекта может содержать репортерный ген, который управляется

различными промоторами и со скремблированными или специфическими miRNA-связывающими сайтами после кодирующей последовательности. Применяемый в данном документе контрольный промотор представляет собой CMV, и контроль для miRNA-связывающего сайта представляет собой случайно скремблированную версию сайта связывания для miR-142. Применяемый в данном документе специфический в отношении ткани-мишени промотор представляет собой ApoE.HCR.hAAT, который экспрессируется в клетках печени, при этом нецелевой тканеспецифический miRNA-связывающий сайт является комплементарным с miR-142 (uguaguguuuuissuacuuuauugga (SEQ ID NO: 1684)), который экспрессируется в HSC.

Нуклеофекцию клеток-мишеней и клеток, не являющихся мишенями, осуществляют с использованием комбинации полипептида Gene Writer (1) и матрицы (2), выбранных из следующего.

Конструкции полипептида Gene Writer (1):

- a. неспецифический управляющий элемент: CMV-R2Tg;
- b. неспецифический инактивированный управляющий элемент: CMV-R2Tg(EN*);
- c. тканеспецифический управляющий элемент: ApoE.HCR.hAAT-R2Tg-miR142;
- d. тканеспецифический инактивированный управляющий элемент: ApoE.HCR.hAAT-R2Tg(EN*)-miR142.

Конструкции матрицы Gene Writer (2):

- a. неспецифический трансген: CMV-gfp;
- b. тканеспецифический трансген: ApoE.HCR.hAAT-gfp-miR142.

Клетки инкубируют в течение по меньшей мере трех дней и затем оценивают эффективность интеграции и экспрессию репортера. В случае эффективности интеграции выполняют ddPCR для количественной оценки среднего числа интеграций на геном для каждого образца. В некоторых вариантах осуществления соотношение между эффективностью интеграции в клетках-мишенях, и клетках, не являющихся мишенями, выше при применении матрицы в паре с тканеспецифическим управляющим элементом (1a) по сравнению с неспецифическим управляющим элементом (1c). Для оценки экспрессии репортера клетки анализируют с помощью проточной цитометрии для обнаружения флуоресценции GFP и с помощью RT-qPCR для обнаружения транскрипции. В некоторых вариантах осуществления соотношение между флуоресценцией в клетках-мишенях и клетках, не являющихся мишенями, выше при применении управляющего элемента в паре с тканеспецифической кассетой трансгена (2b) по сравнению с неспецифической кассетой трансгена (2a). В некоторых вариантах осуществления соотношение между уровнями транскриптов в клетках-мишенях и клетках, не являющихся мишенями, выше при применении управляющего элемента в паре с тканеспецифической кассетой трансгена (2b) по сравнению с неспецифической кассетой трансгена (2a). В некоторых вариантах осуществления комбинация тканеспецифического управляющего элемента (1a) с тканеспецифической кассетой трансгена (2b) приводит к максимальному соотношению транскрипции или экспрессии между клетками-мишенями и клетками, не являющимися

мишенями. В качестве альтернативы, скрининговый анализ можно проводить в той же линии клеток, искусственно экспрессирующей или не экспрессирующей данную miRNA, например, клеткой-мишенью для скрининга является клетка HEK293T, а клетка, не являющаяся мишенью, имитируется путем введения сверхэкспрессии miR-142 в клетки HEK293T.

110. **Пример 23. Коррекция дефицита альфа-1-антитрипсина с помощью липидных наночастиц, содержащих Gene Writer**

В данном примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности в одном нуклеотиде *in vivo*. Более конкретно, полипептид Gene Writer™ и матрицу для записи доставляют в клетки печени мыши с помощью липидных наночастиц для коррекции мутации SERPINA1 PiZ, обуславливающей дефицит альфа-1-антитрипсина.

Составление LNP (система LNP-INT01), несущих Cas9 и gRNA, и обработка ими мышинных моделей описаны в Finn et al. *Cell Reports* 22:2227-2235 (2018), способы из которой включены в данный документ посредством ссылки.

Кэпированную и полиаденилированную мРНК полипептида Gene Writer, содержащую N1-метил-псевдо-U, создают посредством транскрипции *in vitro* с применением матрицы линейаризованной плазмидной ДНК и РНК-полимеразы T7. мРНК полипептида очищают от фермента и нуклеотидов с применением набора MegaClear Transcription Clean-up в соответствии с протоколом производителя (ThermoFisher). Концентрацию транскрипта определяют с помощью измерения поглощения света при 260 нм (Nanodrop), и транскрипт анализируют посредством капиллярного электрофореза с помощью TapeStation (Agilent). РНК-матрицу, содержащую последовательность, корректирующую мутацию, также получают с помощью транскрипции и трансляции *in vitro* с применением аналогичных способов. В данном примере РНК-матрица содержит последовательность, показанную в примере 1.

LNP составлены с соотношением амина и РНК-фосфата (N:P), составляющим 4,5. Липидные компоненты наночастиц растворяют в 100% этаноле при следующих молярных соотношениях: 45 мол. % липида LP01, 44 мол. % холестерина, 9 мол. % DSPC и 2 мол. % PEG2k-DMG. Груз РНК (молярное соотношение мРНК полипептида и РНК-матрицы, составляющее 1:40) растворяют в 50 мМ ацетатного буфера (pH 4,5), в результате чего концентрация груза РНК составляет примерно 0,45 мг/мл. LNP составляют посредством микрофлюидного смешивания растворов липидов и РНК с применением настольного прибора Precision Nanosystems NanoAssembler в соответствии с протоколом производителя. После смешивания LNP собирают и разбавляют в PBS (примерно 1:1), и затем оставшийся буфер заменяют на PBS (избыток объема образца в 100 раз) в течение ночи при 4°C при осторожном перемешивании с применением 10 кДа Slide-a-Lyzer G2 Dialysis Cassette (ThermoFisher Scientific). Полученную смесь затем фильтруют с применением стерильного фильтра с размером пор 0,2 мкм. Фильтрат хранят при температуре от 2°C до 8°C. Составы с несколькими дозами можно изготовить с применением 25 мМ цитрата, буфера для груза

с 100 mM NaCl (pH 5) и осуществить замену буфера посредством TFF на Tris-солевой сахарозный буфер (TSS) (5% сахарозы, 45 mM NaCl и 50 mM Tris). Составленные LNP имеют средний размер, составляющий 105 нм. Эффективность инкапсуляции определяют с помощью анализа на основе ribogreen (Leung et al., 2012). Размер частиц и полидисперсность измеряют методом динамического рассеяния света (DLS) с применением прибора для DLS Malvern Zetasizer.

Мыши NSG-PiZ, несущие человеческий аллель SERPINA1 PiZ (E342K), получены из лаборатории The Jackson Laboratory. Для оценки способности системы Gene Writing редактировать мутантный аллель *in vivo*, LNP вводили через латеральную хвостовую вену в дозе 3 мг/кг в объеме 0,2 мл на животное. Животных, обработанных вспомогательными средствами, применяют в качестве отрицательных контролей во всех исследованиях. Животных подвергают эвтаназии в различные моменты времени путем обескровливания посредством пункции сердца под анестезией изофлураном. В некоторых вариантах осуществления животных подвергают эвтаназии через одну неделю после обработки для анализа на Gene Writing. Ткань печени собирают из средней или левой латеральной доли у каждого животного для экстракции и анализа ДНК.

Для NGS-анализа эффективности редактирования праймеры для ПЦР разработаны к последовательностям вблизи сайта-мишени, и область, представляющую интерес, амплифицируют из экстрагированной геномной ДНК. Дополнительную ПЦР проводят в соответствии с протоколами производителя (Illumina) с добавлением соответствующих химических веществ для секвенирования, после чего ампликоны секвенируют на Illumina MiSeq. Риды, получаемые в результате секвенирования, выравнивают с эталонным геномом мыши после исключения тех, которые имеют низкие баллы качества. Полученные файлы, содержащие риды, сопоставляют с картой эталонного генома (файлы BAM), где выбирают риды, перекрывающие область-мишень, представляющую интерес, и рассчитывают соотношение числа ридов дикого типа и числа ридов, содержащих реверсивную мутацию SERPINA1, кодирующуюся в РНК-матрице. Процентную долю редактирования (например, "эффективность редактирования" или "процент редактирования") определяют в виде отношения общего числа ридов реверсивной последовательности к общему числу ридов последовательности.

В некоторых вариантах осуществления данный пример повторяют с дополнительными группами мышей, и режим с повторным введением доз применяют для анализа свойств системы, характеризующихся зависимостью "доза-эффект". В данных экспериментах мышей распределяют по группам для еженедельного введения доз до 4 недель, при этом эвтаназию и анализ тканей, описанные в данном документе, проводят каждую неделю. В некоторых вариантах осуществления мыши, которые получают большее количество доз состава на основе LNP, демонстрируют более высокую эффективность Gene Writing, измеренную с помощью секвенирования, например, мыши, получающие 2 дозы с интервалом в одну неделю, которых анализируют на третьей неделе, показывают более высокую долю ридов исправленных генов с помощью NGS образцов ткани печени, по

сравнению с мышами, получавшими разовую дозу и проанализированными на третьей неделе. При применении введение доз таким образом может позволять осуществлять настройку терапевтического вмешательства после оценки ответа пациента на одну или несколько доз.

111. Пример 24. Применение Gene Writing для борьбы с болезнями экспансии повторов

В данном примере описано применение системы редактирования генов Gene Writer™ для лечения болезни экспансии повторов посредством перезаписи нормального числа повторов в локус. Более конкретно, полипептид Gene Writer™ и матрицу для записи доставляют в ЦНС мыши посредством вектора AAV для сброса повторов CAG в HTT в соответствии с специализированной РНК-матрицей для лечения болезни Хантингтона. У здоровых людей обычно содержится от 10 до 35 повторов CAG в гене гентингина (*HTT*), в то время как люди с болезнью Хантингтона могут иметь от 36 до более 120 повторов.

В данном примере РНК-матрица предназначена для исправления области повторов CAG гена HTT с помощью кодирования последовательности с 10 такими повторами и гомологией с фланкирующей последовательностью-мишенью для полной записи по локусу-мишени. Можно сконструировать множество примеров таких РНК-матриц, при этом иллюстративная РНК, кодируемая ДНК, содержит последовательность (1)GGCGGCTGAGGAAGCTGAGG(2)GTTTTAGAGCTAGAAATAGCAAGTTAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)AGTCCCTCAAG TCCTTCcagcagcagcagcagcagcagcagcagcagccgccaccgccgccgccgccgccgcctcct(4)CAGCTTCC TCAG (SEQ ID NO: 1685), где числа используются для обозначения модулей матрицы в порядке (5'-3'): (1) спейсер gRNA, (2) каркас gRNA, (3) последовательность гетерологичного объекта, (4) 3'-концевой праймирующий домен гомологии, при этом коррекция повторов кодируется в (3). За областью повтора CAG следует короткая область повтора, кодирующая 11 остатков пролина (8 остатков кодируются триплетами CCG). Не желая ограничиваться какой-либо теорией, данную область включают в (3) для того, чтобы поместить (4) в более уникальную область для предотвращения неправильного праймирования. Иллюстративная gRNA для обеспечения второго одностороннего разрыва, как описано в вариантах осуществления данной системы, содержит спейсерную последовательность CGCTGCACCGACCGTGAGTT (SEQ ID NO: 1630) и направляет нуклеазу Cas9 для осуществления одностороннего разрыва во второй нити сайта-мишени в пределах области гомологии. В некоторых вариантах осуществления данный второй односторонний разрыв обеспечивает улучшение эффективности редактирования.

С целью доставки полной системы Gene Writing в ЦНС в данном примере Gene Writer разделен на два генома AAV, при этом первый кодирует домен нуклеазы Cas9, слитый с интеином-N расщепленной пары интеинов (интеин-N DnaE: CLSYETEILTVEYGLLPKIVEKRIECTVYSVDNNGNIYTPVAQWHDRGEQEVFEYCL EDGSLIRATKDNKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1638)), и второй кодирует домен RT, слитый с интеином-C расщепленной пары интеинов (интеин-C

DnaE, MIKIATRKYLKGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1640)), и РНК-матрицу. Два полипептидных компонента экспрессируются с промотора полимеразы II, например, промотора, специфичного для нейронных клеток, описанного в данном документе, и РНК-матрица и gRNA для обеспечения второго одностороннего разрыва экспрессируются с промотора полимеразы III, например, промотора U6. При совместном инфицировании клетки два полипептидных компонента воссоздают полный полипептид Gene Writer с N-концевой Cas9 и C-концевой RT, и РНК-матрица экспрессируется и обратно транскрибируется в локус-мишень. Для достижения доставки к клеткам ЦНС (в частности, в хвостатого ядра и скорлупы базальных ганглиев) в данном документе применяют псевдотипированную систему gAAV2/1, где ITR AAV2 применяют для упаковки описанных нуклеиновых кислот в частицы с капсидом AAV1. Применяемые в данном документе протоколы получения AAV, введения в виде инъекций мышам и сбора соответствуют идеям из Monteys et al. *Mol Ther* 25(1):12-23 (2017).

Мыши FVB-Tg(YAC128)53Hay/J получены из лаборатории The Jackson Laboratory. У таких трансгенных мышей экспрессируется полноразмерный белок гентингтин человека с ~118 повторами глутамина (тринуклеотидные повторы CAG) и развивается гиперкинез в возрасте трех месяцев. В возрасте 8 недель мышей обрабатывают комбинацией вируса gAAV2/1-Cas9 и вируса gAAV-MMLV_RT/hU6templateRNA в соотношении 1:1. Для инъекций gAAV мышей анестезируют изофлураном и 5 мкл смеси gAAV вводят унилатерально в правое полосатое тело со скоростью 0,2 мкл/мин. Через три недели мышей умерщвляют и собирают ткань мозга для экстракции геномной ДНК и NGS-анализа.

Для NGS-анализа эффективности редактирования разрабатывают праймеры для ПЦР, фланкирующие сайт-мишень, и область, представляющую интерес, амплифицируют из экстрагированной геномной ДНК. Дополнительную ПЦР проводят в соответствии с протоколами производителя (Illumina) с добавлением необходимых химических веществ для секвенирования, после чего ампликоны секвенируют на Illumina MiSeq. Риды, получаемые в результате секвенирования, выравнивают с эталонным геномом мыши после исключения тех, которые имеют низкие баллы качества. Полученные файлы, содержащие риды, сопоставляют с картой эталонного генома (файлы BAM), где выбирают риды, перекрывающие область-мишень, представляющую интерес, и рассчитывают соотношение числа ридов аллеля заболевания (>35 CAG-повторов) и числа ридов восстановленного аллеля (10-35 CAG-повторов). Процентную долю редактирования (например, "эффективность редактирования" или "процент редактирования") определяют в виде отношения общего числа ридов восстановлений, указанных выше, к общему числу ридов последовательности.

112. Пример 25. Доставка системы Gene Writing с помощью средств доставки, представляющих собой LNP и AAV

В данном примере описано применение системы редактирования генов Gene Writer™ для изменения геномной последовательности в одном нуклеотиде *in vivo*. Более конкретно, полипептид Gene Writer™ и матрицу для записи доставляют в клетки печени

мышь с помощью комбинации липидных наночастиц (мРНК, кодирующую полипептид) и AAV (ДНК, кодирующую РНК-матрицу) для коррекции мутации SERPINA1 PiZ, обуславливающей дефицит альфа-1-антитрипсина.

Кодирующую полипептид Gene Writer мРНК, к которой добавлен кэп и хвост, получают путем транскрипции *in vitro* и вводят в состав с LNP-INT01, как описано в примере 23, но без совместного составления с РНК-матрицей.

В данном примере РНК-матрица кодируется в виде ДНК и доставляется посредством AAV. Идеи Cunningham et al. *Mol Ther* 16(6):1081-1088 (2008) описывают применение rAAV2/8 с промотором альфа-1-антитрипсина человека (hAAT) и двумя копиями области контроля в печени из энхансера аполипопротеина E (ApoE) для эффективной трансдукции и стимуляции экспрессии груза в печени молодых мышей. Соответственно, rAAV2/8.ApoE-hAAT.PiZ (rAAV2/8.PiZ), описанный в данном документе, содержит описанный выше AAV и промоторную систему, управляющую экспрессией РНК-матрицы для коррекции мутации PiZ, в дополнение к направляющей введение второго однокитевого разрыва gRNA, управляемой промотором U6 (последовательности РНК, ранее описанные в примере 1).

Мышь NGS-PiZ, несущие человеческий аллель SERPINA1 PiZ (E342K), получены из лаборатории The Jackson Laboratory. Для оценки активности системы Gene Writing редактировать мутантный аллель *in vivo*, 8-недельным мышам внутрибрюшинно вводили ~10¹¹ vg rAAV2/8.PiZ для экспрессии РНК-матрицы, и через латеральную хвостовую вену вводили составленные LNP в дозе 3 мг/кг в объеме 0,2 мл на животное для экспрессии полипептида Gene Writer. Животных подвергают эвтаназии в различные моменты времени путем обескровливания посредством пункции сердца под анестезией изофлураном. В некоторых вариантах осуществления животных подвергают эвтаназии через одну неделю после обработки для анализа на Gene Writing. Ткань печени собирают из средней или левой латеральной доли у каждого животного для экстракции и анализа ДНК.

Для NGS-анализа эффективности редактирования праймеры для ПЦР разработаны к последовательностям вблизи сайта-мишени, и область, представляющую интерес, амплифицируют из экстрагированной геномной ДНК. Дополнительную ПЦР проводят в соответствии с протоколами производителя (Illumina) с добавлением необходимых химических веществ для секвенирования, после чего ампликоны секвенируют на Illumina MiSeq. Риды, получаемые в результате секвенирования, выравнивают с эталонным геномом мыши после исключения тех, которые имеют низкие баллы качества. Полученные файлы, содержащие риды, сопоставляют с картой эталонного генома (файлы BAM), где выбирают риды, перекрывающие область-мишень, представляющую интерес, и рассчитывают соотношение числа ридов дикого типа и числа ридов, содержащих реверсивную мутацию SERPINA1, кодирующуюся в РНК-матрице. Процентную долю редактирования определяют в виде отношения общего числа ридов реверсивной последовательности к общему числу ридов последовательности.

113. Пример 26. Применение системы Gene Writer™ для доставки терапевтического гена в печень в мышинной модели химерной печени человека.

В данном примере описана система редактирования генома Gene Writer™, доставляемая в печень *in vivo* для интеграции и стабильной экспрессии генетической полезной нагрузки. В частности, LNP применяют для доставки системы Gene Writing, способной интегрировать полную кассету экспрессии ОТС для лечения дефицита ОТС в гуманизированной мышинной модели.

В данном примере систему Gene Writing применяют для лечения дефицита ОТС в гуманизированной мышинной модели, в которой гепатоциты человека, полученные от пациентов с дефицитом ОТС, привиты в мышиную модель (Ginn et al JHEP Reports 2019). Иллюстративная система Gene Writing для интеграции большой нагрузки содержит систему обратной транскриптазы, направленную на Cas9, в которой используется высокопроцессивная обратная транскриптаза, например, MarathonRT. Иллюстративный компонент, представляющий собой РНК-матрицу, содержит, от 5'- к 3'-концу: (1) спейсер gRNA, гомологичный сайту "безопасной гавани" AAVS1, (2) каркас gRNA, (3) последовательность гетерологичного объекта и (4) 3'-концевую область гомологии с мишенью для отжига с геномной ДНК непосредственно перед одонитевым разрывом первой нити для праймирования TPRT последовательности гетерологичного объекта. Иллюстративной последовательностью для (1) является GGGGCCACTAGGGACAGGAT (SEQ ID NO: 1689). Область (2) несет каркас gRNA, описанный в данной заявке, обычно содержащий последовательность GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGGACCGAGTCGGTCC (SEQ ID NO: 1591). В данном примере (3) содержит полную кассету экспрессии ОТС, где последовательность, кодированная для печени, кодирующая ОТС человека (UniProt P00480), функционально ассоциирована с промоторной системой ApoE.hAAT, описанной в примере 25. Иллюстративной последовательностью для (4) является CTGTCCCTAGTG (SEQ ID NO: 1690). Иллюстративная последовательность дополнительного спейсера gRNA для создания одонитевого разрыва второй нити для повышения эффективности интеграции представляет собой AGAGAGATGGCTCCAGGAAA (SEQ ID NO: 1691).

Самкам мышей *Fah^{-/-}Rag2^{-/-}Il2rg^{-/-}* (FRG) возрастом от восьми до 12 недель прививают гепатоциты человека, выделенные от доноров детского возраста или приобретенных у Lonza (Базель, Швейцария), как описано ранее (Azuma et al Nat Biotechnol 2007). У привитых мышей циклически включают и выключают наличие 2-(2-нитро-4-трифторметилбензоил)-1,3-циклогександиона (NTBC) в питьевой воде для стимулирования репопуляции печени. Кровь собирают каждые две недели и в конце эксперимента для измерения уровней альбумина человека, используемого в качестве маркера для оценки уровня приживления, в сыворотке крови с помощью твердофазного иммуноферментного анализа (ELISA; Bethyl Laboratories, Inc., Монтгомери, Техас). Через одиннадцать недель после прививки мышей обрабатывают с помощью Gene Writer™, состав которого приведен в примере 23. Для осуществления лечения LNP вводят через латеральную хвостовую вену в дозе 3 мг/кг в объеме 0,2 мл на животное.

После инъекции вектора мышей подвергают циклическому введению NTBC еще в течение следующих 5 недель перед эвтаназией. ДНК и РНК впоследствии экстрагируют из лизатов печени стандартными способами. Экспрессию ОТС впоследствии анализируют посредством проведения RT-qPCR на выделенных образцах РНК с применением праймеров, специфических в отношении последовательности. Уровни ОТС человека также измеряли в плазме крови на протяжении всего эксперимента с применением набора ОТС ELISA для человека (например, набора Aviva Systems Biology ОТС ELISA Kit (Human) (OKCD07437)) в дни -7, 0, 2, 4, 7, 14, 21, 28 и 35 после инъекции в соответствии с протоколом, рекомендованным производителем.

Для анализа эффективности редактирования проводят анализ ddPCR с применением пары праймеров, которые отжигаются либо с прохождением через 5'-концевое место соединения, либо через 3'-концевое место соединения интеграции, при этом один праймер каждого набора отжигается с последовательностью гетерологичного объекта, а другой праймер отжигается с соответствующей областью сайта AAVS1 в геноме. Результаты анализа нормализуются к эталонному гену для подсчитывания числа интеграций сайта-мишени на геном.

Для анализа интеграций в сайте-мишени выполняют секвенирование с длинным ридом, проходящее через сайт интеграции. Разрабатывают праймеры для ПЦР, фланкирующие сайт-мишень, и область, представляющую интерес, амплифицируют из экстрагированной геномной ДНК. Дополнительную ПЦР проводят в соответствии с протоколами производителя (PacBio) с добавлением необходимых химических веществ для секвенирования, после чего ампликоны секвенируют посредством PacBio. Риды, получаемые в результате секвенирования, выравнивают с эталонным геномом мыши после исключения тех, которые имеют низкие баллы качества. Полученные файлы, содержащие риды, сопоставляют с картой эталонного генома (файлы BAM), где выбирают риды, содержащие последовательность вставки относительно эталонного генома, для дополнительного анализа с целью определения полноты интеграции, определенной в данном примере как содержащая полный промотор и кодирующую последовательность ОТС.

114. Пример 27. Gene Writer для интеграции CAR в Т-клетки ex vivo

В данном примере описана доставка системы редактирования генома Gene Writer™ в Т-клетки ex vivo для интеграции и стабильной экспрессии генетической полезной нагрузки. В частности, LNP применяют для доставки системы Gene Writing, способной интегрировать химерный антигенный рецептор (CAR) в локус *TRAC* для создания CAR-Т-клеток для лечения В-клеточной лимфомы.

В данном примере система Gene Writing содержит полипептид Gene Writing, например, домен нуклеазы Cas9 и домен обратной транскриптазы R2Tg, описанные в данном документе, gRNA для направления активности нуклеазы в локус-мишень, и РНК-матрицу, содержащую в направлении от 5'- к 3'-концу следующее:

- (1) последовательность гомологии с сайтом-мишенью длиной 100 нуклеотидов на 3'-конце одонитевого разрыва первой нити;
- (2) 5' UTR из R2Tg;
- (3) последовательность гетерологичного объекта;
- (4) 3' UTR из R2Tg;
- (5) последовательность гомологии с сайтом-мишенью длиной 100 нуклеотидов на 5'-конце одонитевого разрыва первой нити;

где (3) содержит кодирующую последовательность для CD19-специфичной молекулы CAR Hu19-CD828Z (Genbank MN698642; Brudno et al. *Nat Med* 26:270-280 (2020)). Gene Writer в этом примере направляют к 5'-концу первого экзона *TRAC* с помощью нацеленной gRNA, например, TCAGGGTTCTGGATATCTGT (SEQ ID NO: 1692), чтобы поместить груз под эндогенный контроль экспрессии из этого локуса, одновременно разрушая эндогенный TCR, как указано в Eyquem et al. *Nature* 543:113-117 (2017). Все данные три компонента (мРНК полипептида, gRNA и матрица) предусматривают РНК, которую синтезируют посредством транскрипции *in vitro* (например, мРНК полипептида, РНК-матрица) или посредством химического синтеза (gRNA).

Состав на основе LNP, применяемый в данном примере, был проверен и валидирован для доставки в Т-клетки *ex vivo*, как описано в Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020), которая включена в данный документ посредством ссылки во всей своей полноте. В частности, применяли состав на основе LNP C14-4, содержащий холестерин, фосфолипид, связанный с липидом PEG и ионизируемый липид C14-4 (фигура 2C из Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020)) для инкапсуляции всех трех РНК-компонентов при молярном соотношении мРНК полипептида:gRNA:РНК-матрица, составляющем приблизительно 1:40:40.

В Т-клетках могли проводить дополнительные изменения с целью улучшения активности CAR-Т-клеток в отношении их когнатной мишени. В некоторых вариантах осуществления второй состав на основе LNP C14-4, который описан, содержит предварительно образованный комплекс RNP Cas9/gRNA, где gRNA нацелена на экзон 1 *Pdcd1* для инактивации PD-1, что может усиливать противоопухолевую активность CAR-Т-клеток за счет нарушения этой ингибиторной контрольной точки, которая в противном случае может вызвать подавление клеток (см. Rupp et al. *Sci Rep* 7:737 (2017)). Таким образом, применение обоих составов на основе наночастиц позволяет осуществлять нацеливание на лимфому посредством предоставления нагрузки в виде антител к CD19, одновременно повышая эффективность за счет отключения ингибиторной контрольной точки PD-1. В некоторых вариантах осуществления клетки можно обрабатывать наночастицами одновременно. В некоторых вариантах осуществления клетки можно обрабатывать наночастицами на отдельных стадиях, например, сначала доставлять RNP для отключения PD-1, и затем обрабатывать клетки наночастицами, несущими CAR к CD19. В некоторых вариантах осуществления второй компонент системы, который обеспечивает улучшение эффективности Т-клеток, может приводить к нокауту PD-1, TCR, CTLA-4, HLA-

I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS, PDC1, CISH, TRAC или их комбинаций. В некоторых вариантах осуществления отключение PD-1, TCR, CTLA-4, HLA-I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS, PDC1, CISH или TRAC может быть предпочтительно, например, с применением siRNA, нацеливающейся на PD-1. В некоторых вариантах осуществления нацеливания siRNA на PD-1 могут достигать с применением самодоставляющейся системы RNAi, описанной в Ligtenberg et al. *Mol Ther* 26(6):1482-1493 (2018) и в WO2010033247, включенных в данный документ посредством ссылки во всей своей полноте, в которых описаны обширные химические модификации siPHK, придающие полученным гидрофобно модифицированным молекулам siPHK способность проникать во все типы клеток *ex vivo* и *in vivo*, и способность достичь длительного отключения специфического гена-мишени без каких-либо дополнительных составов или методик доставки. В некоторых вариантах осуществления один или несколько компонентов системы можно доставлять с помощью других способов, например, с помощью электропорации. В некоторых вариантах осуществления посредством нокина в клетки вводят дополнительные регуляторы для обеспечения сверхэкспрессии для контроля опосредованных Т-клетками и НК-клетками иммунных ответов и контроля поглощения макрофагами, например, PD-L1, HLA-G, CD47 (Han et al. *PNAS* 116(21):10441-10446 (2019)). Нокин можно осуществить путем применения дополнительной системы Gene Writing с матрицей, несущей кассету экспрессии для одного или нескольких таких факторов (3) с нацеливанием на локус безопасной гавани, например, AAVS1, например, с использованием gRNA GGGGCCACTAGGGACAGGAT (SEQ ID NO: 1689), чтобы нацелить полипептид Gene Writer на AAVS1.

LNP применяют для обработки первичных Т-клеток, активированных Dynabeads, при соотношении CD4⁺:CD8⁺ 1:1 при общей концентрации мРНК 450 нг/мкл. Полученные популяции Т-клеток анализируют в отношении интеграции, экспрессии и эффекта. Для оценки интеграции применяют ddPCR с праймерами, обеспечивающими получение ампликона, простирающегося от интегрированного CAR до фланкирующей геномной последовательности TRAC. Сравнение сигнала с эталонным геном (например, RPP30) позволяет количественно определить среднее число копий на геном и эффективность интеграции. Для анализа экспрессии применяют проточную цитометрию с иммунологическими зондами для оценки уровня и процента клеток, проявляющих поверхностную экспрессию CAR. Чтобы проанализировать активность CAR-Т-клеток, обработанные клетки оценивают посредством анализа уничтожения раковых клеток при совместном культивировании. Посредством конструирования клеток Nalm6 ALL для экспрессии люциферазы уничтожение раковых клеток можно оценить по изменению люминесценции после совместного культивирования с CAR-Т-клетками по сравнению с сигналом от клеток отдельно, Nalm6 Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020). Таким образом, систему Gene Writing могут применять для создания CAR-Т-клеток *ex vivo* с желаемой цитотоксической активностью.

115. Пример 28. Gene Writer для интеграции CAR в Т-клетки *in vivo*

В данном примере описана система редактирования генома Gene Writer™, доставляемая в Т-клетки *in vivo* для интеграции и стабильной экспрессии генетической нагрузки. В частности, нацеленные наночастицы применяют для доставки системы Gene Writing, способной интегрировать кассету экспрессии химерного антигенного рецептора (CAR) в локус Rosa26 мыши для создания CAR-Т-клеток в мышинной модели.

В данном примере система Gene Writing содержит полипептид Gene Writing, например, домен нуклеазы Cas9 и домен обратной транскриптазы R2Tg, описанные в данном документе, gRNA для направления активности нуклеазы в локус-мишень, и РНК-матрицу, содержащую в направлении от 5'- к 3'-концу следующее:

- (1) последовательность гомологии с сайтом-мишенью длиной 100 нуклеотидов на 3'-конце одностороннего разрыва первой нити;
- (2) 5' UTR из R2Tg;
- (3) последовательность гетерологичного объекта;
- (4) 3' UTR из R2Tg;
- (5) последовательность гомологии с сайтом-мишенью длиной 100 нуклеотидов на 5'-конце одностороннего разрыва первой нити,

где (3) содержит кодирующую последовательность для CD19-специфического m194-1BBz CAR, управляемого промотором EF1a (Smith et al. *Nat Nanotechnol* 12(8):813-820 (2017)). Gene Writer в этом примере направляется к мышинному локусу Rosa26 с помощью gRNA, например, АСТССAGТСТТТСТАГААГА (SEQ ID NO: 1693) (Chu et al. *Nat Biotechnol* 33(5):543-548 (2015)). Получение молекул РНК осуществляют согласно примерам, приведенным в данном документе, например, путем транскрипции *in vitro* (например, мРНК полипептида Gene Writer, РНК-матрицы) и путем химического синтеза (например, gRNA). Модификации РНК-компонентов системы описаны в других местах. Для мРНК Gene Writer последовательность дополнительно содержит 5'-UTR (например, GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1603)) и 3'-UTR (например, UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1604)), фланкирующие кодирующую последовательность. Показано, что данная комбинация 5'-UTR и 3'-UTR приводит к хорошей экспрессии функционально связанной ORF (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

С целью добиться доставки конкретно к Т-клеткам, создают нацеленные LNP (tLNP), которые несут конъюгированные mAb к CD4. См., например, Ramishetti et al. *ACS Nano* 9(7):6706-6716 (2015). В качестве альтернативы, конъюгирование mAb к CD3 могут применять для нацеливания как на CD4⁺, так и на CD8⁺ Т-клетки (Smith et al. *Nat Nanotechnol* 12(8):813-820 (2017)). В других вариантах осуществления наночастица, применяемая для доставки к Т-клеткам *in vivo*, представляет собой ограниченную наночастицу, в которой отсутствует нацеливающий лиганд, как указано в Lokugamage et al. *Adv Mater* 31(41):e1902251 (2019).

tLNP можно получить путем предварительного получения смеси нуклеиновых кислот (например, смесь мРНК полипептида:gRNA:РНК-матрицы при молярном соотношении, составляющем 1:40:40) со смесью липидов (холестерин, DSPC, PEG-DMG, Dlin-MC3-DMA и DSPE-PEG-малеимид) с последующим химическим конъюгированием с желаемым восстановленным с помощью DTT mAb (например, антителом к CD4, например, клоном YTS.177) с функциональной группой малеимида на LNP. См., Ramishetti et al. ACS Nano 9(7):6706-6716 (2015).

Мышам C57BL6/J в возрасте от шести до 8 недель внутривенно вводят составленный LNP в дозе 1 мг РНК/кг веса тела. Кровь собирают через день и через три дня после введения в пробирки для сбора, покрытые гепарином, и лейкоциты выделяют посредством центрифугирования по плотности с применением Ficoll-Paque PLUS (GE Healthcare). Через пять дней после введения животных подвергают эвтаназии и собирают кровь и органы (селезенку, лимфатические узлы, клетки костного мозга) для анализа Т-клеток. Экспрессию CAR к CD19 обнаруживают с помощью FACS с применением специфической иммунологической сортировки. Интеграцию положительных клеток подтверждают с помощью ddPCR в отсортированной популяции, где применяются праймеры, фланкирующие место соединения интегрированного элемента, например, один праймер пары отжигается с интегрированным грузом, а другой праймер отжигается с геномной ДНК из сайта-мишени Rosa26.

116. Пример 29. Оценка расстояния и ориентации РАМ между первым и вторым односторонними разрывами для уменьшения образования не обусловленных матрицей вставок/делений во время использования Gene Writing

В этом примере показано, как можно оптимизировать размещение второго одностороннего разрыва, используемого в системе Gene Writing, чтобы (1) увеличить частоту внедрения желаемого редактирования с помощью полипептида системы Gene Writer с РНК-матрицей, а также (2) уменьшить нежелательные вставки и/или делеции, которые могут возникнуть как побочный продукт второго одностороннего разрыва.

Иллюстративная система Gene Writing может внедрить желаемую геномную модификацию (например, вставку, делецию или точечную мутацию) с использованием 1) РНК-матрицы, которая содержит gRNA и последовательность гетерологичного объекта, содержащую желаемую геномную модификацию, и 2) полипептида Gene Writing, содержащего нуклеазу Cas9 (например, Cas9 N863A), слитую с обратной транскриптазой (RT) (например, доменом RT из MMLV). В указанной иллюстративной системе Gene Writing продукт слияния Cas9-RT вводит первый односторонний разрыв, который высвобождает доступный 3'-ОН для инициации реакции обратной транскриптазы с использованием РНК-матрицы в качестве матрицы для обратной транскрипции с праймированием мишенью. Размещение второго одностороннего разрыва рядом с первым односторонним разрывом, но на противоположной нити от него усиливает внедрение желаемой модификации генома.

В этом эксперименте вставка из 3 нуклеотидов (СТТ) направлена на локус HEK3. РНК-матрица для вставки содержит (1) спейсер gRNA, гомологичный сайту HEK3, (2)

каркас gRNA, (3) последовательность гетерологичного объекта, содержащую вставку СТТ, и (4) 3'-концевую область гомологии с мишенью для отжига с геномной ДНК непосредственно перед одонитевым разрывом первой нити для обеспечения обратной транскрипции с праймированием мишенью последовательности гетерологичного объекта. Используемая последовательность РНК-матрицы представляет собой (5'-3') GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCA<AAG>CGTGCTCAGTCTG (SEQ ID NO: 1694), где "<>" используется для обозначения последовательности вставки.

Кроме того, был разработан набор gRNA для второго одонитевого разрыва, нацеленный на одонитевой разрыв противоположной нити ДНК в качестве первого одонитевого разрыва, который размещал второй одонитевой разрыв либо выше, либо ниже местоположения желаемой вставки СТТ на различных расстояниях в диапазоне от 26 до 257 п. о. Вышерасположенный второй одонитевой разрыв создает набор одонитевых разрывов с ориентацией внутрь с сайтами РАМ наружу (без включения РАМ), а нижерасположенный второй одонитевой разрыв создает набор одонитевых разрывов с ориентацией наружу с сайтами РАМ внутри одонитевых разрывов (с включением РАМ), как описано в данном документе. gRNA для второго одонитевого разрыва были разработаны с использованием веб-инструмента и перечислены в таблицах E5 и E6. Расстояние между двойными одонитевыми разрывами указывает на расстояние между первым одонитевым разрывом, направляемым РНК-матрицей, и вторым одонитевым разрывом, направляемым gRNA для второго одонитевого разрыва, а ориентация РАМ (например, "с включением РАМ" и, следовательно, ориентация наружу или "без включения РАМ" и, следовательно, ориентация внутрь) предусмотрена относительно первого одонитевого разрыва, как показано на фигуре 31.

117. Таблица E5. Нацеливание gRNA на второй одонитевой разрыв выше первого одонитевого разрыва в ориентации "без включения РАМ"

Ориентация РАМ относительно первого одонитевого разрыва	Ориентация	Расстояние между двойными одонитевыми разрывами (нт)	Последовательность sgRNA	Последовательность РАМ
наружу	антисмысловая	28	TGGGCCCCAAGGATTGACCC (SEQ ID NO: 1695)	AGG

наружу	антисмысл овая	33	CCCAAGGATTGACCCAGGCC (SEQ ID NO: 1696)	AGG
наружу	антисмысл овая	34	CCAAGGATTGACCCAGGCCA (SEQ ID NO: 1697)	GGG
наружу	антисмысл овая	38	GGATTGACCCAGGCCAGGGC (SEQ ID NO: 1698)	TGG
наружу	антисмысл овая	108	GCAGAAATAGACTAATTGCA (SEQ ID NO: 1699)	TGG
наружу	антисмысл овая	109	CAGAAATAGACTAATTGCAT (SEQ ID NO: 1700)	GGG
наружу	антисмысл овая	120	TAATTGCATGGGCGTTTCCC (SEQ ID NO: 1701)	TGG
наружу	антисмысл овая	121	AATTGCATGGGCGTTTCCCT (SEQ ID NO: 1702)	GGG
наружу	антисмысл овая	136	TCCCTGGGATCCCTGTCTCC (SEQ ID NO: 1703)	AGG
наружу	антисмысл овая	161	TCTCTCATCCATGCCTTTCT (SEQ ID NO: 1704)	AGG
наружу	антисмысл овая	197	CCCTTGCTTAAACTCTCCA (SEQ ID NO: 1705)	AGG
наружу	антисмысл овая	222	TCTCATGCCAAGCTCCCTGC (SEQ ID NO: 1706)	AGG
наружу	антисмысл овая	232	AGCTCCCTGCAGGACATCCC (SEQ ID NO: 1707)	AGG
наружу	антисмысл овая	240	GCAGGACATCCCAGGCCCTC (SEQ ID NO: 1708)	TGG
наружу	антисмысл овая	241	CAGGACATCCCAGGCCCTCT (SEQ ID NO: 1709)	GGG
наружу	антисмысл овая	255	CCCTCTGGGACAGCAGCTCA (SEQ ID NO: 1710)	CGG
наружу	антисмысл овая	256	CCTCTGGGACAGCAGCTCAC (SEQ ID NO: 1711)	GGG

118. Таблица Е6. Нацеливание gRNA на второй одонитевой разрыв ниже первого одонитевого разрыва в ориентации "с включением PAM"

Ориентация РАМ относительно первого одонитевого разрыва	Ориентация	Расстояние между двойными одонитевыми разрывами (нт)	Последовательность sgRNA	Последовательность РАМ
внутри	антисмысловая	26	GACGCCCTCTGGAGGAAGCA (SEQ ID NO: 1712)	GGG
внутри	антисмысловая	27	CGACGCCCTCTGGAGGAAGC (SEQ ID NO: 1713)	AGG
внутри	антисмысловая	34	TGTCCTGCGACGCCCTCTGG (SEQ ID NO: 1714)	AGG
внутри	антисмысловая	37	AGCTGTCCTGCGACGCCCTC (SEQ ID NO: 1715)	TGG
внутри	антисмысловая	63	GCACATACTAGCCCCTGTCT (SEQ ID NO: 1716)	AGG
внутри	антисмысловая	90	GTCAACCAGTATCCCGGTGC (SEQ ID NO: 1717)	AGG
внутри	антисмысловая	96	AAACTTGTC AACCAGTATCC (SEQ ID NO: 1718)	CGG
внутри	антисмысловая	133	CCAGGGACCTCCCTAGGTGC (SEQ ID NO: 1719)	TGG
внутри	антисмысловая	139	CCCCTTCCAGGGACCTCCCT (SEQ ID NO: 1720)	AGG
внутри	антисмысловая	150	GGTGAGGCTGGCCCCTTCCA (SEQ ID NO: 1721)	GGG
внутри	антисмысловая	151	TGGTGAGGCTGGCCCCTTCC (SEQ ID NO: 1722)	AGG
внутри	антисмысловая	162	CCCTCCTCTCCTGGTGAGGC (SEQ ID NO: 1723)	TGG
внутри	антисмысловая	166	AGGTCCCTCCTCTCCTGGTG (SEQ ID NO: 1724)	AGG

внутри	антисмысл овая	171	GGGCCAGGTCCCTCCTCTCC (SEQ ID NO: 1725)	TGG
внутри	антисмысл овая	186	GAGCTCGACCCTGAAGGGCC (SEQ ID NO: 1726)	AGG
внутри	антисмысл овая	191	CTGTTGAGCTCGACCCTGAA (SEQ ID NO: 1727)	GGG
внутри	антисмысл овая	192	TCTGTTGAGCTCGACCCTGA (SEQ ID NO: 1728)	AGG
внутри	антисмысл овая	233	GCTGAAAGCCACTGGGCTCT (SEQ ID NO: 1729)	GGG
внутри	антисмысл овая	234	TGCTGAAAGCCACTGGGCTC (SEQ ID NO: 1730)	TGG
внутри	антисмысл овая	240	TGCAGGTGCTGAAAGCCACT (SEQ ID NO: 1731)	GGG
внутри	антисмысл овая	241	ATGCAGGTGCTGAAAGCCAC (SEQ ID NO: 1732)	TGG
внутри	антисмысл овая	257	TTGATCTCTGATTTTCATGC (SEQ ID NO: 1733)	AGG

Для проведения эксперимента 200000 клеток U2OS в 20 мкл буфера SE подвергали нуклеофекции с помощью 800 нг плазмиды, кодирующей полипептид Gene Writer™ (N863ACas9-RT), 200 нг РНК-матрицы и 83 нг gRNA для второго одностороннего разрыва, перечисленных в таблицах E5 и E6. Система нуклеофекции Lonza Amara используется с кодом нуклеофекции DN100. После нуклеофекции к клеточной суспензии добавляли 80 мкл среды DMEM+10% FBS, и клетки помещали в 24-луночный планшет с 500 мкл DMEM+10% FBS. Геномную ДНК экстрагировали в день 3 после нуклеофекции.

Чтобы проанализировать экстрагированную ДНК на наличие желаемой вставки СТТ, проводили секвенирование ампликона, как описано в данном документе, путем амплификации локуса НЕК с использованием праймеров, окружающих первый односторонний разрыв. Ожидаемый ампликон длиной 300-350 п. о. затем секвенировали на Illumina MiSeq. Частоту желаемых вставок ССТ определяли с использованием последовательности операций вычислительного анализа CRISPResso (Clement et al. Nat Biotechnol 37(3):224-226 (2019)).

Для измерения нежелательных вставок и/или делеций, возникающих как побочные продукты реакции, проводили амплификацию длинных фрагментов с праймерами, расположенными > 1,5 т. п. о. выше и ниже сайта первого одностороннего разрыва, с получением ампликона > 3 т. п. о. Этот ампликон секвенировали с использованием

секвенирования с длинными ридами (например, PacBio) и анализировали на наличие вставок и делеций, возникающих в результате двойного одностороннего разрыва.

В некоторых вариантах осуществления реакция с использованием gRNA для второго одностороннего разрыва, которая делает разрез ниже первого одностороннего разрыва и обеспечивает ориентацию "с включением PAM" или наружу, приводит к меньшему количеству непреднамеренных мутаций (например, мутаций в сайте-мишени, отличных от целевой вставки CTT) по сравнению с gRNA, располагающейся выше первого одностороннего разрыва на аналогичном расстоянии, но обеспечивающей ориентацию "без включения PAM" или внутрь, как измеряется способами, описанными в данном документе. В других вариантах осуществления gRNA для второго одностороннего разрыва, которая делает разрез выше первого одностороннего разрыва и обеспечивает ориентацию "без включения PAM" или внутрь, приводит к меньшему количеству нежелательных мутаций (например, мутаций в сайте-мишени, отличных от целевой вставки CTT), когда расстояние между первым и вторым односторонними разрывами составляет по меньшей мере 100 нуклеотидов, по сравнению с gRNA для второго одностороннего разрыва, которая обеспечивает расстояние между первым и вторым односторонними разрывами менее 100 нуклеотидов, при измерении способами, описанными в данном документе.

Таким образом, в некоторых вариантах осуществления предпочтительной конструкцией для gRNA для второго одностороннего разрыва является конструкция, приводящая к 1) ориентации "с включением PAM" или наружу или 2) ориентации "без включения PAM" или внутрь, где первый и второй односторонние разрывы отделены по меньшей мере 100 нуклеотидами (фигура 32).

119. Пример 30. Разработка и экспрессия в клетках человека систем Gene Writing с использованием различных продуктов слияния Cas-RT

В этом примере описывается конструирование и экспрессия полипептидов Gene Writing, содержащих продукты слияния доменов Cas и обратной транскриптазы, в клетках млекопитающих. В данном документе было показано, что полипептиды Gene Writing с этими доменами обеспечивают точную сайт-специфическую модификацию ДНК-мишени из молекулы РНК-матрицы. В данном документе описывается экспрессия библиотеки доменов для создания новых систем, которые могут иметь различные функциональные характеристики. Более конкретно, в данном документе описаны слитые белки, содержащие 1) Cas-нуклеазу, содержащую мутацию, инактивирующую один активный сайт эндонуклеазы, например, нуклеазу Cas9 (N863A); 2) пептид-линкер для соединения функциональных белковых доменов, например, последовательность из таблицы 38 или 42, например, SGGSSGGSSGSETPGTSESATPESSGGSSGGSS; и 3) обратную транскриптазу (RT), например, домен RT, описанный в настоящей заявке, например, домен RT, содержащий последовательность из таблицы 1, таблицы 3, таблицы 30, таблицы 31 или таблицы 41, или ее производное, которое можно использовать в таком анализе, совокупно называемые в этом примере как Cas-RT. Соответственно, слитые белки Cas-RT собираются

на плазмиде и доставляются совместно с экспрессионной плазмидой с одной направляющей РНК (sgRNA) для валидации экспрессии системы в клетках человека.

Полипептиды Gene Writer, полученные из продуктов слияния доменов Cas-RT, проанализированных в данном документе, содержат: (1) сигнальный домен Cas9 дикого типа или Cas9(N863A); (2) пептидный линкер (SGGSSGGSSGSETPGTSEATPSSGGSSGGSS); (3) домены RT, выбранные из таблицы 1 и таблицы 30, взятые из различных источников; а также (4) по меньшей мере один сигнал ядерной локализации. Клетки U2OS или HEK293T трансфицировали путем нуклеофекции Lonza Amaxa в количестве 250000 клеток на лунку с помощью ~800 нг плазмиды продукта слияния Cas9(N863A)-RT с 200 нг плазмиды sgRNA. Для оценки уровня экспрессии продуктов слияния Cas9-RT клеточные лизаты собирали в день 2 после трансфекции и анализировали с помощью вестерн-блоттинга с использованием первичного антитела к Cas9. Несколько продуктов слияния Cas9-RT продемонстрировали заметную экспрессию белка (фигура 33), что свидетельствует об уровнях экспрессии, достаточных для активности Gene Writing. Примечательно, что для различных конструкций Cas9-RT наблюдается широкий диапазон уровней экспрессии, демонстрирующий влияние конструкции продукта слияния и выбора RT на уровень экспрессии Cas-RT в клетках.

120. Пример 31. Улучшение экспрессии продуктов слияний Cas-RT посредством выбора линкера

В данном примере демонстрируется оптимизация продуктов слияния Cas-RT для улучшения экспрессии белка в клетках млекопитающих. Как описано в примере 30, конструирование новых продуктов слияния Cas-RT посредством простой замены новых функциональных доменов может привести к низкой или умеренной экспрессии полипептида Gene Writer. Таким образом, в данном документе предполагается, что модифицированные конфигурации продуктов слияния могут быть привлекательными в контексте различных доменов. Не желая ограничиваться примером, одним из таких подходов к улучшению экспрессии и стабильности новых продуктов слияния является применением библиотеки линкеров. В данном документе последовательность пептидного линкера между доменами Cas и RT продукта слияния Cas-RT варьируется с использованием библиотеки последовательностей линкера. Более конкретно, линкеры из таблицы 42 использовали для создания новых вариантов слитой конструкции Cas9-RT, ранее демонстрировавшей низкую экспрессию белка (см. пример 30 и фигуру 33), и доставляли в клетки человека для скрининга на улучшенную экспрессию белка Cas-RT.

Набор из 22 пептидных линкеров (таблица 42) с различной степенью длины, гибкости, гидрофобности и вторичной структуры впервые применяли для создания вариантов слитого белка Cas-RT посредством замены исходного линкера (см. пример 30). Клетки HEK293T трансфицировали с помощью электропорации 250000 клеток/лунку с помощью ~800 нг каждой слитой плазмиды Cas9-RT вместе с 200 нг плазмиды с одиночной направляющей РНК. Для оценки уровня экспрессии продуктов слияния Cas9-RT клеточные лизаты собирали в день 2 после трансфекции и анализировали с помощью вестерн-

блоттинга с использованием первичного антитела к Cas9. Линкер 10, указанный в таблице 38, значительно улучшил экспрессию продукта слияния Cas-RT (фигура 34), демонстрируя потенциально сильное влияние последовательности пептидного линкера на экспрессию Cas-RT.

Таблица 42. Пептидные последовательности, применяемые в качестве линкеров между доменами Cas и RT в полипептидах Gene Writer, содержащих продукты слияния Cas-RT

№	Последовательность линкера	Примечания
1	GGG	Короткий
2	GGGGG	Гибкий, короткий
3	(GGGGG) ₂	Гибкий
4	(GGGGG) ₃	Гибкий, длинный
5	(GGGGG) ₄	Гибкий, очень длинный
6	(G) ₆	Гибкий
7	(G) ₈	Гибкий
8	GSAGSAAGSGEF	Гибкий
9	(GSSGSS)	Средний
10	(GSSGSS) ₂	Средний, гибкий
11	(GSSGSS) ₃	Средний
12	SGSETPGTSESATPES	XTEN
13	(EAAAK)	Жесткая спираль, короткий
14	(EAAAK) ₂	Жесткая спираль, средний
15	(EAAAK) ₃	Жесткая спираль, длинный
16	PAP	Жесткий, короткий
17	PAPAP	Жесткий, короткий
18	PAPAPAPAP	Жесткий, средний
19	A(EAAAK) ₄ ALEA(EAAAK) ₄ A	Жесткий, очень длинный со спиралью
20	GGGGG(EAAAK)GGGGG	Гибкий - спираль - изгиб
21	(EAAAK)GGGGG(EAAAK)	Спираль - изгиб - спираль
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	Гибкий - XTEN - гибкий

121. **Пример 32. Cas-опосредованная расщепляющая активность Gene Writer, содержащих продукты слияния Cas-RT**

Этот пример демонстрирует способность продуктов слияния Cas-RT сохранять функциональные свойства белковых доменов. В частности, при анализе клеток, обработанных полипептидами Gene Writer, содержащими компетентный в отношении расщепления домен Cas (кливазу), связывание ДНК может быть прочитано с помощью анализа сайта-мишени, чтобы продемонстрировать активность Cas в качестве составной части продуктов слияния. В данном документе такие продукты слияния кливазы Cas-RT, в которых оба активных сайта нуклеазы являются функциональными, например, Cas9 (дикий тип)-RT, были совместно доставлены на плазмидных векторах вместе с плазмидой экспрессии sgRNA для нацеливания Cas на сайт AAVS1 в клетках человека. Анализ образования вставок/делеций в предполагаемом сайте расщепления в AAVS1 продуктами слияния кливазы Cas-RT функционировал как показатель как ДНК-связывающей активности, так и эндонуклеазной активности, тем самым подтверждая эффективное нацеливание на ДНК с помощью продуктов слияния Cas-RT.

Продукты слияния Cas-RT с полностью функциональными эндонуклеазными доменами, например, содержащие Cas9 дикого типа с обоими интактными активными сайтами нуклеазы, например, Cas9(N863), были получены из слитых белков Cas-RT, описанных в данном документе, например, содержащих никазу Cas9, например, Cas9(N863A), чтобы повысить чувствительность обнаружения связывания и расщепления ДНК. Поскольку интактная нуклеаза Cas9 может разрезать обе нити, вызывая событие двухнитевого расщепления в геноме, репарация этих сайтов генерирует более высокий сигнал мутации (вставка/делеция), чем репарация одонитевого разрыва ДНК. Таким образом, частоту образования вставок/делеций в продуктах слияния сравнивали с частотой для неслитых Cas9 дикого типа, чтобы оценить сохранение функциональных свойств Cas в качестве составной части новых продуктов слияния Cas-RT.

Клетки U2OS или HEK293T трансфицировали путем нуклеофекции Lonza Amaxa в количестве 250000 клеток на лунку с помощью ~800 нг плазмиды продукта слияния Cas9(WT)-RT вместе с 200 нг плазмиды sgRNA для получения gRNA, нацеленной на Cas9 в AAVS1 (таблица 43, gRNA P7). Для оценки ДНК-связывающей и расщепляющей активности продуктов слияния кливазы Cas9-RT геномную ДНК (gDNA) собирали в день 3 после трансфекции. Паттерны вставок/делеций в gDNA анализировали путем секвенирования ампликонов в локусах, на которые нацелена sgRNA. Результаты секвенирования анализировали с помощью последовательности операций CRISPResso2 (Clement et al Nat Biotechnol 37(3):224-226 (2019)). Все протестированные продукты слияния кливазы Cas-RT показали образование вставок/делеций, соизмеримое с их соответствующими уровнями экспрессии белка (фигура 33), что указывает на то, что Cas-опосредованная ДНК-связывающая активность сохраняется в продуктах слияния Cas-RT (фигура 35).

122. Пример 33. Gene Writer, содержащие продукты слияния Cas-RT с различными доменами RT, обеспечивают точное редактирование в клетках человека

Этот пример демонстрирует способность нескольких протестированных продуктов слияния Cas-RT программируемым образом внедрять в геномной ДНК в клетках человека. Более конкретно, домен обратной транскриптазы продуктов слияния Cas-RT, например, домен RT, описанный в этой заявке, варьировали для определения способности редактирования генома у продуктов слияния Cas-RT с использованием новых комбинаций RT. РНК-матрицы совместно доставляли на плазмидных векторах вместе с экспрессионными плазмидами Cas-RT в клетки человека для определения активности перезаписи у продуктов слияния Cas-RT.

Для создания библиотек доменов для полипептидов Gene Writer были выбраны эффекторные белки Cas; см. таблицу 37 и таблицу 40A. Можно было определить (таблица 40A) дополнительные домены Cas9, которые были дополнительно отобраны для использования в полипептидах Gene Writer, описанных в данном документе, в качестве признаков, включающих требования РАМ к последовательности-мишени, предсказанные мутации для придания нисказной активности (например, D10A, H840A или N863A для SpCas9) и признаки gRNA, включающие одиночную направляющую композицию, например, конкретные параметры спейсера и последовательности каркаса gRNA для обеспечения связывания полипептида с когнатным ферментом Cas. Линкерные последовательности для соединения доменов Cas и RT были собраны на основе поиска разнообразия по длине, гибкости и составу для оптимизации слитых белков (таблицы 38 и 42). Оптимизация экспрессии продукта слияния путем скрининга линкеров дополнительно описана в примере 31. Домены обратной транскриптазы были извлечены из различных источников с использованием литературы и сигнатур белковых доменов RT, как предусмотрено в этой заявке, в том числе из ретротранспозонов, не содержащих LTR, ретротранспозонов LTR, интронов группы II, элементов, генерирующих разнообразие, ретронов, теломераз, ретроплазмид, ретровирусов и полимераз с эволюционировавшей РНК-зависимой ДНК-полимеразной активностью (например, в таком анализе можно использовать домен RT, содержащий последовательность из таблицы 1, таблицы 3, таблицы 30, таблицы 31, таблицы 44 или таблицы N41 или ее производное).

В частности, для оценки использования новых доменов RT в качестве составной части полипептида системы Gene Writer для успешного редактирования генома было выбрано подмножество иллюстративных доменов RT из ретровирусов для слияния с нисказой Cas9(N863A). Вкратце, сначала была создана, а затем упорядочена по приоритету база данных POL-белков Retroviridae (см. The UniProt Consortium Nucleic Acids Res 47(D1):D506-D515 (2019); Mitchell et al. Nucleic Acids Res 47(D1):D351-D360 (2019)). Не желая ограничиваться таким примером, полагают, что ретровирусные RT из родов *Betaretrovirus*, *Deltaretrovirus*, *Gammaretrovirus* и *Spumavirus* могут функционировать как мономерные белки (см., например, таблицу 1 из Herschhorn et al Cell Mol Life Sci 67(16):2717-2747 (2010)), и, таким образом, это может быть преимущественным для использования в слитом белке, как описано в данном документе. Выбор ретровирусных мономерных последовательностей RT, полученных в результате анализа с этими

критериями, показан в таблице 44. Дополнительно, мутации, которые, как было показано, стабилизируют домены RT, как описано в этой заявке и в литературе (таблица 45) (Anzalone et al. *Nat Biotechnol* 38(7):824-844 (2020); Baranauskas et al. *Protein Eng Des Sel* 25(10):657-668 (2012); Arezi and Hogrefe *Nucleic Acids Res* 37(2):473-481 (2009); Yasukawa et al. *J. Biotechnol* 150(3):299-306 (2010); результаты которых, поскольку они относятся к улучшению стабильности и функции RT, включены в данный документ во всей своей полноте), были проанализированы в отношении применения к доменам-кандидатам RT (положения, представленные в данном документе, приведены на основе аминокислотной последовательности RT MMLV в качестве эталонной). Например, RT MMLV с мутационным профилем L139P/D200N/T330P/L603W/E607K показал примерно 65-кратное увеличение процессивности и 48-кратное увеличение аффинности к матрице (Baranauskas et al. *Protein Eng Des Sel* 25(10):657-668 (2012)) и повышение эффективности первичного редактирования геномной ДНК в 1,6-5,1 раза с мутационным профилем D200N/T306K/W313F/T330P/L603W (Anzalone et al. *Nat Biotechnol* 38(7):824-844 (2020)). Из этих исследований был идентифицирован ключевой набор D200N/T330P/L603W, и выравнивание доменов RT из ретровирусных родов, описанных в данном документе, было использовано для предсказания того, где находятся соответствующие консервативные аминокислоты (фигура 36A). Дополнительные мутации T306K и/или W313F также применяли там, где это было уместно, и L139P и/или E607K использовали, когда нельзя было применить ни одну из мутаций набора T306K/W313F (фигура 36 B). Были созданы продукты слияния никазы Cas9 с этими доменами RT дикого типа или мутационными вариантами с потенциально улучшенной активностью, а иллюстративные продукты слияния описаны в таблице 46.

Для создания точных редактирований с использованием продуктов слияния Cas-RT Gene Writer были сконструированы РНК-матрицы в качестве матриц обратной транскрипции с внесением редактирования в сайт-мишень генома с помощью домена RT. РНК-матрицы были разработаны так, чтобы содержать (i) спейсерную последовательность gRNA для направления Cas-RT в область-мишень, например, последовательность, комплементарную 20-нуклеотидной последовательности в локусе НЕК3; (ii) последовательность связывания праймера, способная к комплементарному спариванию оснований с одной нитью ДНК с одонитевым разрывом для обратной транскрипции с праймированием мишенью; (iii) последовательность гетерологичного объекта, обеспечивающую матрицу для обратной транскрипции, которая дополнительно содержит предполагаемую конечную последовательность-мишень; и (iv) последовательность каркаса gRNA для связывания с доменом Cas9 слитого полипептида Cas9-RT. Конструкции, используемые в данном документе, специально располагались в ориентации от 5' к 3' (i), (iv), (iii), (ii). РНК-матрицы, кодируемые на плазмидах, клонировали таким образом, чтобы экспрессия управлялась промотором U6, а терминация транскрипции контролировалась 7-нуклеотидным поли(Т)-отрезком, следующим за последовательностью связывания

праймера на 3'-конце кассеты РНК-матрицы. Составы матриц описаны в таблице 43 (матрицы P1, P2, P3).

Клетки U2OS или HEK293T трансфицировали путем электропорации 250000 клеток/лунка с ~800 нг экспрессионной плазмиды продукта слияния Cas9-RT(MMLV), 200 нг экспрессионной плазмиды РНК-матрицы и 83 нг экспрессионной плазмиды дополнительной gRNA для второго одностороннего разрыва (2gRNA P5) (таблица 43). Для оценки способности продуктов слияния Cas-RT к редактированию генома геномную ДНК (gDNA) собирали в день 3 после трансфекции. Частоту преднамеренных (точное и безрубцовое редактирование, как было разработано) по сравнению с непреднамеренными (любые непреднамеренные изменения последовательности-мишени) редактированиями ("коэффициент активности") в локусах-мишенях анализировали с помощью секвенирования ампликонов. Используемый в данном документе термин "секвенирование ампликонов сайта-мишени" включает использование сайт-специфичных праймеров для ПЦР-амплификации сайта-мишени, секвенирование ампликонов на Illumina MiSeq, а также обнаружение и определение характеристик событий редактирования с использованием последовательности операций CRISPResso2 (Clement et al Nat Biotechnol 37(3):224-226 (2019)). Несколько продуктов слияния Cas-RT продемонстрировали заметную активность редактирования генома, при этом несколько продуктов слияния Cas-RT имели коэффициенты активности ~3 (фигура 37), что демонстрирует, что различные продукты слияния Cas-RT, выведенные из описанных в данном документе доменов обратной транскриптазы, могут эффективно и точно вносить редактирования в человеческий геном.

Таблица 43. Список РНК-матриц и gRNA, используемых в некоторых примерах

Название	Описание	спейсер	каркас	RT+вставка	PBS	РНК-матрица
Матрица P1	HEK3_8P BS_10RT(CTT в 1)	GGCCCAG ACTGAGC ACGTGA	GTTTTAGA GCTAGAA ATAGCAA GTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG AGTCGGTG C	TCTGCCAT CAAAG	CGT GCT CA	GGCCCAGACT GAGCACGTGA GTTTTAGAGC TAGAAATAGC AAGTTAAAAT AAGGCTAGTC CGTTATCAAC TTGAAAAAGT GGCACCGAGT CGGTGCTCTG CCATCAAAGC GTGCTCA

Матр ица P2	HEK3_13 PBS_10R T(CTT В 1)	GGCCCAG ACTGAGC ACGTGA	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG AGTCGGTG C	TCTGCCAT CAAAG	CGT GCT CAG TCT G	GGCCCAGACT GAGCACGTGA GTTTTAGAGC TAGAAATAGC AAGTTAAAAT AAGGCTAGTC CGTTATCAAC TTGAAAAAGT GGCACCGAGT CGGTGCTCTG CCATCAAAGC GTGCTCAGTC TG
Матр ица P3	HEK3_17 PBS_10R T(CTT В 1)	GGCCCAG ACTGAGC ACGTGA	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG AGTCGGTG C	TCTGCCAT CAAAG	CGT GCT CAG TCT GGG CC	GGCCCAGACT GAGCACGTGA GTTTTAGAGC TAGAAATAGC AAGTTAAAAT AAGGCTAGTC CGTTATCAAC TTGAAAAAGT GGCACCGAGT CGGTGCTCTG CCATCAAAGC GTGCTCAGTC TGGGCC
Матр ица P4	HBB_13P BS_10RT(Т на А В 4)	GCATGGT GCACCTG ACTCCTG	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG	AGACTTCT CCACAG	GAG TCA GGT GCA C	GCATGGTGCA CCTGACTCCT GGTTTTAGAG CTAGAAATAG CAAGTTAAAA TAAGGCTAGT CCGTTATCAA CTTGAAAAAG TGGCACCGAG

			AGTCGGTG C			TCGGTGCAGA CTTCTCCACA GGAGTCAGGT GCAC
2gRN A, P5	HEK3_+9 0	GTCAACC AGTATCC CGGTGC	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG AGTCGGTG C	NA	NA	NA
2gRN A, P6	HBB_+72	GCCTTGA TACCAAC CTGCCCA	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT GGCACCG AGTCGGTG C	NA	NA	NA
gRN A P7	g19_AAV S1	GTCCCCT CCACCCC ACAGTG	GTTTTAGA GCTAGAA ATAGCAA GTTAAAAT AAGGCTA GTCCGTTA TCAACTTG AAAAAGT	NA	NA	NA

			GGCACCG AGTCGGTG C			
--	--	--	--------------------------	--	--	--

123. Пример 34. Мультиплексирование системы Gene Writer для одновременного редактирования нескольких локусов в клетке человека

Этот пример демонстрирует использование системы Gene Writer для редактирования нескольких сайтов в геноме. В некоторых приложениях может быть очень полезно иметь возможность сконструировать несколько участков генома, например, для коррекции нескольких генетических мутаций или для оптимизации сконструированной клетки для клеточной терапии путем выполнения нескольких одновременных модификаций *ex vivo* или *in vivo*. В этом примере используется система из 3 плазмид, содержащая: 1) экспрессионную плазмиду полипептида Gene Writer, например, плазмиду, кодирующую нуклеазу Cas9, слитую с обратной транскриптазой (Cas-RT); 2) плазмиду с матрицей, например, плазмиду, кодирующую экспрессионную кассету для РНК-матрицы, которая определяет сайт генома и редактирование для внедрения в этот сайт; и 3) экспрессионную плазмиду gRNA для второго одностороннего разрыва, например, плазмиду, кодирующую дополнительную последовательность gRNA для направления одностороннего разрыва второй нити для Cas9 в местоположение, проксимальное к сайту-мишени.

В этом примере на два геномных локуса, ген HBB и локус HEK3 человека, нацеливание производилось с использованием gRNA, содержащей спейсерные последовательности с идентичностью по отношению к этим сайтам, чтобы определить способность нацеливаться на несколько локусов параллельно. Для оценки нацеливания на любой локус по отдельности или на оба одновременно клетки обрабатывали различными составами плазмид с матрицей, чтобы обеспечить нацеливание на: 1) только HEK3, 2) только HBB или 3) как HBB, так и HEK3.

В частности, 800 нг плазмиды, кодирующей продукт слияния Cas9-RT(MMLV) (таблица 46), 200 нг плазмиды, кодирующей HEK3-модифицирующую матрицу (матрица P2, таблица 43) и/или плазмиды, кодирующей HBB-модифицирующую матрицу (матрица P4, таблица 43), и 83 нг плазмиды, кодирующей gRNA для второго одностороннего разрыва HEK3 (2gRNA P5, таблица 43) и/или плазмиды, кодирующей gRNA для второго одностороннего разрыва HBB (2gRNA P6, таблица 43), вводили путем нуклеофекции с использованием программы DS_150 в клетки HEK293T. После нуклеофекции клетки выращивали при 37°C, 5% CO₂ в течение 3 дней перед лизисом клеток и экстракцией геномной ДНК. Праймеры, специфичные для каждого локуса, использовали для амплификации области, а ампликоны секвенировали с использованием Illumina MiSeq. Показатели частоты точной коррекции и вставок/делеций анализировали с использованием последовательности операций CRISPResso2 (Clement et al. Nat Biotechnol 37(3):224-226

(2019)) для определения эффективности Gene Writing. В таблице 51 перечислены компоненты системы Gene Writer, используемые в этом примере.

Таблица 51.

Название	Описание	спейсер	каркас	RT+вставка	PBS	PHK-матрица
Матрица P1	HEK3_8P BS_1 ORT(CTT в 1)	GGCCCA GACTGA GCACGT GA	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	TCTGC CATCA AAG	CGTGCTC A	GGCCCAGACTGAGCACGTG AGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAAGGCTA GTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG TCTGCCATCAAAGCGTGCTC A
Матрица P2	HEK3_13 PBS_10RT (CTT в 1)	GGCCCA GACTGA GCACGT GA	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	TCTGC CATCA AAG	CGTGCTC AGTCTG	GGCCCAGACTGAGCACGTG AGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAAGGCTA GTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG TCTGCCATCAAAGCGTGCTC AGTCTG
Матрица P3	HEK3_17 PBS_10RT (CTT в 1)	GGCCCA GACTGA GCACGT GA	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	TCTGC CATCA AAG	CGTGCTC AGTCTG GGCC	GGCCCAGACTGAGCACGTG AGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAAGGCTA GTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTG TCTGCCATCAAAGCGTGCTC AGTCTGGGCC
Матрица P4	HBB_13P BS_1	GCATGG TGCACC	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA	AGACT TCTCC ACAG	GAGTCA GGTGCA C	GCATGGTGCACCTGACTCCT GGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAAGGCTA

	ORT(T на A в 4)	TGACTC CTG	GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC			GTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGC AGACTTCTCCACAGGAGTCA GGTGCAC
2gRNA , P5	HEK 3_+9 0	GTCAAC CAGTAT CCCGGT GC	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	NA	NA	NA
2gRNA , P6	HBB _+72	GCCTTG ATACCA ACCTGC CCA	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	NA	NA	NA
gRNA P7	g19_ AAV S1	GTCCCC TCCACC CCACAG TG	GTTTTAGAGCT AGAAATAGCA AGTTAAAATAA GGCTAGTCCGT TATCAACTTGA AAAAGTGGCA CCGAGTCGGT GC	NA	NA	NA

При независимом тестировании обе мишени показали высокую степень точной коррекции: примерно 36% редактирование в НЕК3 и 23% редактирование в НВВ (фиг. 38). Дополнительно при одновременном нацеливании было достигнуто примерно 34% редактирование сайтов-мишеней НЕК3 и 14% редактирование сайтов-мишеней НВВ с

точной коррекцией, обеспечиваемой соответствующими РНК-матрицами. Кроме того, вставки и делеции наблюдались с низкой частотой во всех условиях, при этом вставки для каждого локуса достигали сходных уровней при тестировании по отдельности или в комбинации. Хотя это и не является явной целью этого примера, отсутствие увеличения частоты вставок/делеций во время одновременного редактирования является положительным индикатором потенциального увеличения количества локусов, на которые можно нацеливаться параллельно, без ущерба для точности каждого отдельного редактирования.

124. Пример 35. Доставка систем Gene Writer без использования ДНК посредством нуклеофекции клеток человека

В этом примере описывается применение системы Gene Writer для редактирования генома в клетках человека посредством доставки РНК-компонентов, например, мРНК, кодирующей полипептид Gene Writer, и РНК-матрицы. Не желая ограничиваться какой-либо теорией, ожидают, что способность доставлять только РНК-компоненты в отсутствие ДНК будет давать этой системе основные преимущества, включая снижение иммуногенности и клеточной токсичности, связанной с обнаружением ДНК в цитоплазме, и доступность описанных в данном документе систем липидных наночастиц, большинство из которых оптимизированы для доставки РНК, что может позволить обойти проблемы, связанные с вирусной доставкой терапевтических нуклеиновых кислот (например, производственные проблемы, ранее существовавший иммунитет, иммуногенный ответ на вирусные белки). Снижение клеточной токсичности за счет использования системы РНК может быть особенно важно для модификации более чувствительных типов клеток, таких как первичные клетки. Дополнительно, нуклеофекция может быть эффективным методом доставки этих систем в клетки пациента, например, для конструирования клеток *ex vivo*. Таким образом, очень важно продемонстрировать способность системы Gene Writing функционировать должным образом при доставке в виде только РНК и в отсутствие ДНК. В частности, этот пример демонстрирует доставку Gene Writing, состоящей только из РНК, для модификации генома клеток HEK293T. Чтобы продемонстрировать, что Gene Writing на основе РНК не ограничивается одной композицией, используются две версии слитого полипептида Cas-RT, которые содержат домен RT, полученный либо из вируса мышинного лейкоза Молони (Cas9-RT (MMLV)), либо из эндогенного ретровируса свиней (Cas9-RT(PERV)) (таблица 46).

мРНК, кодирующие полипептид Gene Writer (1), были получены с использованием транскрипции *in vitro*, управляемой полимеразой T7. Как правило, плазмиды, кодирующие конструкции мРНК, содержат кассету транскрипции, содержащую следующие компоненты: промотор T7, 5'-UTR, кодирующую последовательность Gene Writer (никазу Cas9, слитую с обратной транскриптазой с помощью пептидного линкера и дополнительно содержащую сигнал ядерной локализации), 3'-UTR и поли(А)-хвост из 80 нуклеотидов. В этом примере молекулы РНК были получены с использованием немодифицированных нуклеотидов из линеаризованной плазмидной матрицы. мРНК, кодирующие Cas9-

RT(MMLV) или Cas9-RT(PERV) (таблица 47), были экпированы котранскрипционно с помощью CleanCap AG (TriLink BioTechnologies).

РНК-матрицы Gene Writer (2), кодирующие геномные редактирования, были созданы путем химического синтеза и очищены путем стандартного обессоливания. Первые и последние три основания каждой РНК-матрицы содержали 2'-О-метилфосфотиоатные модификации. РНК-матрицы различной длины были сконструированы для введения различных мутаций в локус HEK3 человека (таблица 48).

Где указано, gRNA для второго однокитевого разрыва (3) были созданы путем химического синтеза и содержали следующие модификации последовательности: mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

Для анализа РНК систем Gene Writing, описанных в данном документе, клетки HEK293Т высевали за 2 дня до нуклеофекции для получения 70-80% конfluence в день нуклеофекции. РНК смешивали в соответствии со следующими комбинациями: i) только мРНК Cas9-RT (1); ii) мРНК Cas9-RT (1), РНК-матрица (2) и gRNA для второго однокитевого разрыва (3); или iii) мРНК Cas9-RT (1) и РНК-матрица (2). Смеси РНК содержали 4,5 мкг мРНК Cas9-RT (1), РНК-матрицу (2) в конечной концентрации 5 мкМ и gRNA для второго однокитевого разрыва (3) в конечной концентрации 1,3 мкМ. Смеси подвергали нуклеофекции в примерно 200000 клеток с использованием 96-луночной системы Lonza Amaxa Nucleofector Shuttle в соответствии с протоколами производителя. Затем клетки лизировали, и собирали геномную ДНК через 72 часа после нуклеофекции. Библиотеки для секвенирования ампликонов были получены с использованием праймеров для амплификации по сайту-мишени, и было выполнено секвенирование Illumina. Показатели частоты точной коррекции и вставок/делеций анализировали с использованием последовательности операций CRISPResso2 (Clement et al. Nat Biotechnol 37(3):224-226 (2019)).

В этих экспериментах была достигнута примерно 20% активность точной записи с помощью Cas9-RT (MMLV) с использованием матрицы 1 (таблица 48). Падение активности наблюдалось для матриц длиной более 120 нуклеотидов; матрица 4, которая кодировала то же редактирование, что и матрица 1, но с добавлением 20 нуклеотидов на 3'-конце матрицы RT, демонстрировала примерно 3,1-кратное падение активности точной записи и примерно 2,4-кратное падение соотношения точных коррекций и вставок/делеций (фигура 39). Использование мРНК, кодирующей Cas9-RT, с разными UTR и подходами к экпированию давало сходные уровни активности, хотя наблюдалось небольшое увеличение в случае с мРНК-5 (таблица 49; фигура 40). Нуклеофекция всех РНК Gene Writer Cas9-RT(PERV) с помощью матрицы 1 и gRNA для второго однокитевого разрыва дополнительно приводила к эффективности точной записи, составляющей примерно 7% (фигура 41). В экспериментах этого примера добавление gRNA для второго однокитевого разрыва приводило к увеличению активности записи.

В таблице 48 представлены последовательности молекул РНК-матриц, используемых в примерах Gene Writing, состоящей только из РНК. Спейсерная последовательность каждой РНК-матрицы, описанная в данном документе, сохранялась постоянной и состояла из 20 нуклеотидов (5'-GGCCCAGACTGAGCACGTGA-3'), на 100% идентичных сайту-мишени в локусе HEK3 человека (также известном как LINC01509) (последовательность соответствует NC_000009.12:107422339..107422358, сборка GRCh38.p13). РНК-матрица обычно содержит компоненты, показанные в таблице, так что спейсер+каркас+RT+редактирование+PBS+хвост могут дать полную молекулу.

Таблица 48. РНК-матрицы, используемые в различных примерах, раскрытых в данном документе

Название	Описание	спейсер	каркас	RT+редактирование	PBS	Хвост	Объединенная РНК-матрица	Последовательность РНК	Длина
Матрица 1	HEK3_13PBS_10RT(CTT в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	TCTGCC ATCAAA G	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC GGTGCTCTGCCAT CAAAGCGTGCTCA GTCTGTTTT	mr*mr*mC*rCrCrArGrArCrUrGrArGrCrArCrGrUrGrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrCrUrGrCrCrArUrCrArArArGrCrGrUrGrCrUrCrArGrUrCrUrGrU*mU*mU*mU	126
Матрица 2	HEK3_13PBS_10RT(10 нт в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA	TCTGCC ATCACA TG TAGT TG	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC GGTGCTCTGCCAT	mr*mr*mC*rCrCrArGrArCrUrGrArGrCrArCrGrUrGrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrCr	133

			CCGAGTCG GTGC				CACATGTAGTTGC GTGCTCAGTCTGT TTT	UrGrCrCrArUrCrArCrArUrGrU rArGrUrUrGrCrGrUrGrCrUrCr ArGrUrCrUrGrU*mU*mU*mU	
Матр ица 3	HEK3_13P BS_10RT(2 0 нт в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	TCTGCC ATCACA TGTAGT TGAGGT CAATGA	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC GGTGCTCTGCCAT CACATGTAGTTGA GGTCAATGACGTG CTCAGTCTGTTTT	mr*mr*mC*rCrCrArGrArCrUrG rArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArA rUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrC rGrUrUrArUrCrArArCrUrUrGr ArArArArArGrUrGrGrCrArCrC rGrArGrUrCrGrGrUrGrCrUrCr UrGrCrCrArUrCrArCrArUrGrU rArGrUrUrGrArGrGrUrCrArAr UrGrArCrGrUrGrCrUrCrArGrU rCrUrGrU*mU*mU*mU	143
Матр ица 4	HEK3_13P BS_30RT(CTT в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA	GGAAGC AGGGCT TCCTTTC CTCTGC CATCAA AG	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC	mr*mr*mC*rCrCrArGrArCrUrG rArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArA rUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrC rGrUrUrArUrCrArArCrUrUrGr ArArArArArGrUrGrGrCrArCrC	146

			AAGTGGCA CCGAGTCG GTGC				GGTGCGGAAGCA GGGCTTCCTTTCC TCTGCCATCAAAG CGTGCTCAGTCTG TTTT	rGrArGrUrCrGrGrUrGrCrGrGr ArArGrCrArGrGrGrCrUrUrCrC rUrUrUrCrCrUrCrUrGrCrCrAr UrCrArArArGrCrGrUrGrCrUrC rArGrUrCrUrGrU*mU*mU*mU	
Матр ица 5	HEK3_13P BS_30RT(1 0 нт в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	GGAAGC AGGGCT TCCTTTC CTCTGC CATCAC ATGTAG TTG	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC GGTGCGGAAGCA GGGCTTCCTTTCC TCTGCCATCACAT GTAGTTGCGTGCT CAGTCTGTTTT	mr*mr*mC*rCrCrArGrArCrUrG rArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArA rUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrC rGrUrUrArUrCrArArCrUrUrGr ArArArArArGrUrGrGrCrArCrC rGrArGrUrCrGrGrUrGrCrGrGr ArArGrCrArGrGrGrCrUrUrCrC rUrUrUrCrCrUrCrUrGrCrCrAr UrCrArCrArUrGrUrArGrUrUrG rCrGrUrGrCrUrCrArGrUrCrUr GrU*mU*mU*mU	153

Матр ица 6	HEK3_13P BS_30RT(2 0 нт в 1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	GGAAGC AGGGCT TCCTTTC CTCTGC CATCAC ATGTAG TTGAGG TCAATG A	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTA GAGCTAGAAATAG CAAGTTAAAATAA GGCTAGTCCGTTA TCAACTTGAAAAA GTGGCACCGAGTC GGTGCGGAAGCA GGGCTTCCTTTCC TCTGCCATCACAT GTAGTTGAGGTCA ATGACGTGCTCAG TCTGTTTT	mr*mr*mC*rCrCrArGrArCrUrG rArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArA rUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrC rGrUrUrArUrCrArArCrUrUrGr ArArArArArGrUrGrGrCrArCrC rGrArGrUrCrGrGrUrGrCrGrGr ArArGrCrArGrGrGrCrUrUrCrC rUrUrUrCrCrUrCrUrGrCrCrAr UrCrArCrArUrGrUrArGrUrUrG rArGrGrUrCrArArUrGrArCrGr UrGrCrUrCrArGrUrCrUrGrU* mU*mU*mU	163
---------------	------------------------------------	--------------------------------------	--	--	-------------------------------	------	---	--	-----

Таблица 49. Различное получение и составы мРНК полипептидов Gene Writer, используемые в различных примерах

Название	Транскрипци онная матрица	Кэпирование	Поли(А)	Модифициро ванные NTP	5'-UTR	3'-UTR
мРНК-1	ПЦР-ампликон	CleanCap (AG); котранскрипци онный	Добавлено во время амплифика ции	Без обработки	AGGAAATAAGA GAGAAAAGAAG AGTAAGAAGAA ATATAAGAGCC ACC	GCTGGAGCCTCGGTGGCCATG CTTCTTGCCCCTTGGGCCTCC CCCCAGCCCCTCCTCCCCTTC CTGCACCCGTACCCCGTGGT CTTTGAATAAAGTCTGA

мРНК-2	ПЦР-ампликон	CleanCap (AG); котранскрипци онный	Добавлено во время амплифика ции	5moU	AGGAAATAAGA GAGAAAAGAAG AGTAAGAAGAA ATATAAGAGCC ACC	GCTGGAGCCTCGGTGGCCATG CTTCTTGCCCCTTGGGCCTCC CCCCAGCCCCTCCTCCCCTTC CTGCACCCGTACCCCCGTGGT CTTTGAATAAAGTCTGA
мРНК-3	ПЦР-ампликон	Ферментативн ый, 2'-О- метилованн ый (кэп 1); посттранскрип ционный	Добавлено во время амплифика ции	Без обработки	GGGAAATAAGA GAGAAAAGAAG AGTAAGAAGAA ATATAAGAGCC ACC	GCTGGAGCCTCGGTGGCCATG CTTCTTGCCCCTTGGGCCTCC CCCCAGCCCCTCCTCCCCTTC CTGCACCCGTACCCCCGTGGT CTTTGAATAAAGTCTGA
мРНК-4	ПЦР-ампликон	Ферментативн ый, 2'-О- метилованн ый (кэп 1); посттранскрип ционный	Добавлено во время амплифика ции	5moU	GGGAAATAAGA GAGAAAAGAAG AGTAAGAAGAA ATATAAGAGCC ACC	GCTGGAGCCTCGGTGGCCATG CTTCTTGCCCCTTGGGCCTCC CCCCAGCCCCTCCTCCCCTTC CTGCACCCGTACCCCCGTGGT CTTTGAATAAAGTCTGA
мРНК-5	Линеаризованн ая плазмида	CleanCap (AG); котранскрипци онный	Кодируемы й плазмидой	Без обработки	AGGAAATAAGA GAGAAAAGAAG AGTAAGAAGAA ATATAAGAGCC ACC	GCTGCCTTCTGCGGGGCTTGC CTTCTGGCCATGCCCTTCTTCT CTCCCTTGACCTGTACCTCT TGGTCTTTGAATAAAGCCTGA GTAGGAAGTCTA

125. Пример 36. Использование модифицированных нуклеотидов в системе Gene Writer, состоящей только из РНК

В этом примере описывается применение системы Gene Writer для редактирования генома в клетках человека посредством доставки РНК-компонентов, например, мРНК, кодирующей полипептид Gene Writer, и РНК-матрицы. Дополнительно к демонстрации системы без использования ДНК в примере 35 в этом примере описывается включение модифицированных нуклеотидов, например, 5-метоксиуридина, в мРНК, кодирующую полипептид Gene Writer, и включение модифицированных нуклеотидов, например, 2'-О-метилфосфотиоата, в РНК-матрицу Gene Writer.

мРНК, кодирующие полипептид Gene Writer (1), были получены с использованием управляемой полимеразой T7 транскрипции *in vitro* ампликона, полученного из плазмиды с помощью ПЦР. Плазида, кодирующая конструкцию мРНК, содержала кассету транскрипции, содержащую следующие компоненты: промотор T7, 5'-UTR, кодирующую последовательность Gene Writer (никазу Cas9, слитую с обратной транскриптазой с помощью пептидного линкера и дополнительно содержащую двухкомпонентный NLS SV40) и 3'-UTR. Добавляли компонент в виде поли(А)-хвоста таким образом, чтобы он кодировался в ампликоне, служащем матрицей для транскрипции РНК. В этом примере молекулы мРНК получали путем включения в реакцию транскрипции одного модифицированного нуклеотида - 5-метоксиуридина (5moU). мРНК, кодирующая Cas9-RT(MMLV) (таблица 47), была экпирована либо котранскрипционно с помощью CleanCap AG (TriLink BioTechnologies), либо посттранскрипционно посредством ферментативного экпирования (2'-О-метилованный, кэп 1) (таблица 49).

РНК-матрицы Gene Writer (2), кодирующие геномные редактирования, были созданы путем химического синтеза и очищены путем стандартного обессоливания. Первые и последние три основания каждой РНК-матрицы содержали 2'-О-метилфосфотиоатные модификации. В данном документе для введения вставки СТТ в локус НЕК3 человека использовали матрицу 1 (таблица 48).

Где указано, gRNA для второго одностороннего разрыва (3) были созданы путем химического синтеза и содержали следующие модификации последовательности: mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

Для анализа РНК систем Gene Writing, описанных в данном документе, клетки НЕК293Т высевали за 2 дня до нуклеофекции для получения 70-80% конфлюэнтности в день нуклеофекции. РНК смешивали в соответствии со следующими комбинациями: i) только мРНК Cas9-RT (1); ii) мРНК Cas9-RT (1), РНК-матрица (2) и gRNA для второго одностороннего разрыва (3); или iii) мРНК Cas9-RT (1) и РНК-матрица (2). Смеси РНК содержали 4,5 мкг мРНК Cas9-RT (1), РНК-матрицу (2) в конечной концентрации 5 мкМ и gRNA для второго одностороннего разрыва (3) в конечной концентрации 1,3 мкМ. Смеси подвергали нуклеофекции в примерно 200000 клеток с использованием 96-луночной

системы Lonza Amaxa Nucleofector Shuttle в соответствии с протоколами производителя. Затем клетки лизировали, и собирали геномную ДНК через 72 часа после нуклеофекции. Библиотеки для секвенирования ампликонов были получены с использованием праймеров для амплификации по сайту-мишени, и было выполнено секвенирование Illumina. Показатели частоты точной коррекции и вставок/делеций анализировали с использованием последовательности операций CRISPResso2 (Clement et al. Nat Biotechnol 37(3):224-226 (2019)).

В этих экспериментах была достигнута примерно 20% точность записи с использованием системы Gene Writing, состоящей только из РНК, которая содержала модифицированные нуклеотиды (5moU) в мРНК, кодирующей полипептид Gene Writer (фигура 42). Примечательно, что включение модифицированного нуклеотида 5moU не приводило к наблюдаемому ингибирующему эффекту в отношении эффективности записи. Аналогичная эффективность была получена в способах экпирования мРНК, анализируемых в данном документе (см. таблицу 49). Небольшое снижение эффективности наблюдалось в отсутствие gRNA для второго одностороннего разрыва (фигура 42).

126. Пример 37. Доставка систем Gene Writer без использования ДНК путем трансфекции клеток человека на основе липидов

В этом примере описывается применение системы Gene Writer для редактирования генома в клетках человека посредством доставки РНК-компонентов, например, мРНК, кодирующей полипептид Gene Writer, и РНК-матрицы. Не желая ограничиваться какой-либо теорией, ожидают, что способность доставлять только РНК-компоненты в отсутствие ДНК будет давать этой системе основные преимущества, включая снижение иммуногенности и клеточной токсичности, связанной с обнаружением ДНК в цитоплазме, и доступность описанных в данном документе систем липидных наночастиц, большинство из которых оптимизированы для доставки РНК, что может позволить обойти проблемы, связанные с вирусной доставкой терапевтических нуклеиновых кислот (например, производственные проблемы, ранее существовавший иммунитет, иммуногенный ответ на вирусные белки). Снижение клеточной токсичности за счет использования системы РНК может быть особенно важно для модификации более чувствительных типов клеток, таких как первичные клетки. Реагенты для трансфекции на основе липидов могут использоваться непосредственно для конструирования клеток *ex vivo*, а наночастицы на основе липидов подходят для доставки РНК *in vivo* в клетки пациента. Таким образом, очень важно продемонстрировать способность системы Gene Writing функционировать должным образом при доставке в виде только РНК и в отсутствие ДНК. В частности, этот пример демонстрирует доставку системы Gene Writing, состоящей только из РНК, для модификации генома клеток HEK293T с использованием реагентов для трансфекции на основе липидов Lipofectamine 3000 и MessengerMAX (Invitrogen). Чтобы продемонстрировать, что Gene Writing на основе РНК не ограничивается одной композицией, используются две версии слитого полипептида Cas-RT, которые содержат

домен RT, полученный либо из вируса мышинного лейкоза Молони (Cas9-RT (MMLV)), либо из эндогенного ретровируса свиней (Cas9-RT(PERV)) (таблица 46).

мРНК, кодирующие полипептид Gene Writer (1), были получены с использованием транскрипции *in vitro*, управляемой полимеразой T7. Как правило, плазмиды, кодирующие конструкции мРНК, содержат кассету транскрипции, содержащую следующие компоненты: промотор T7, 5'-UTR, кодирующую последовательность Gene Writer (никазу Cas9, слитую с обратной транскриптазой с помощью пептидного линкера и дополнительно содержащую сигнал ядерной локализации), 3'-UTR и поли(A)-хвост из 80 нуклеотидов. В этом примере молекулы РНК получали с использованием немодифицированных нуклеотидов либо из линейаризованной плазмидной матрицы, либо с использованием ПЦР-ампликона транскрипционной кассеты, описанной выше. мРНК, кодирующая Cas9-RT(MMLV), была экспонирована либо котранскрипционно с помощью CleanCap AG (TriLink BioTechnologies), либо посттранскрипционно посредством ферментативного экспонирования (2'-О-метилованный, кэп 1) (таблица 49). мРНК, кодирующая Cas9-RT(PERV), была получена из плазмидной матрицы и котранскрипционно экспонирована с помощью CleanCap AG (TriLink BioTechnologies) (таблица 47).

РНК-матрицы Gene Writer (2), кодирующие геномные редактирования, были созданы путем химического синтеза и очищены путем стандартного обессоливания. Первые и последние три основания каждой РНК-матрицы содержали 2'-О-метилфосфотиоатные модификации. В данном документе для введения вставки СТТ в локус НЕК3 человека использовали матрицу 1 (таблица 48).

Где указано, gRNA для второго одностороннего разрыва (3) были созданы путем химического синтеза и содержали следующие модификации последовательности: mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

Для анализа описанных в данном документе систем РНК Gene Writing примерно 50000 клеток НЕК293Т высевали в 24-луночные планшеты за 1 день до липофекции. РНК смешивали в соответствии со следующими комбинациями: i) только мРНК Cas9-RT (1); ii) мРНК Cas9-RT (1), РНК-матрица (2) и gRNA для второго одностороннего разрыва (3); или iii) мРНК Cas9-RT (1) и РНК-матрица (2). Смеси РНК содержали 0,45 мкг мРНК Cas9-RT (1), РНК-матрицу (2) в конечной концентрации 2,5 пМ и gRNA для второго одностороннего разрыва (3) в конечной концентрации 1,0 пМ. РНК смешивали со средой Opti-MEM (Gibco) и реагентом Lipofectamine 3000 или MessengerMAX (Invitrogen) и добавляли к клеткам. Затем клетки лизировали, и собирали геномную ДНК через 72 часа после нуклеофекции. Библиотеки для секвенирования ампликонов были получены с использованием праймеров для амплификации по сайту-мишени, и было выполнено секвенирование Illumina. Показатели частоты точной коррекции и вставок/делеций анализировали с использованием последовательности операций CRISPResso2 (Clement et al. Nat Biotechnol 37(3):224-226 (2019)).

В этих экспериментах была достигнута активность точной записи величиной до примерно 17% с использованием системы Gene Writing, состоящей только из РНК, доставляемой путем трансфекции на основе липидов, что приближается к эффективности, аналогичной нуклеофекции (фигура 43 В; см. пример 35 для нуклеофекции). Также использовали Lipofectamine 3000 (фигура 43А). В отличие от нуклеофекции (пример 35), не наблюдалось заметного снижения при использовании матрицы 4, длиннее на 20 нуклеотидов, по сравнению с матрицей 1 (таблица 48; фигура 43С). Дополнительно, липофекция Gene Writer Cas9-RT(PERV), состоящей только из РНК, с помощью матрицы 1 приводила к точной записи желаемого редактирования с эффективностью примерно 3,5% (фигура 44).

127. Пример 38. Gene Writing на основе РНК позволяет точно редактировать первичные Т-клетки без использования ДНК

В этом примере описывается использование системы Gene Writer на основе слитого полипептида Cas9-RT для геномного редактирования последовательностей ДНК-мишеней. Более конкретно, в этом примере описывается нуклеофекция системы, состоящей только из РНК, в первичные CD4⁺ Т-клетки для Gene Rewriting в первичных клетках человека, например, в качестве средства демонстрации системы Gene Rewriter для применения *ex vivo*.

Описанная в данном документе система, состоящая только из РНК, содержала: 1) мРНК, кодирующую полипептид Gene Writer, например, РНК, кодирующую слитый полипептид Cas9-RT, в качестве управляющего элемента для запрограммированного редактирования гена посредством процесса нацеленного внесения однонитевого разрыва и обратной транскрипции, как описано в настоящем изобретении; 2) молекулу РНК-матрицы, например, РНК, содержащую (i) спейсерную последовательность gRNA для направления управляющего элемента в область-мишень, например, последовательность, комплементарную 20-нуклеотидной последовательности в локусе HEK3; (ii) последовательность связывания праймера, способную к комплементарному спариванию оснований с одной нитью ДНК с однонитевым разрывом для обратной транскрипции с праймированием мишенью; (iii) последовательность гетерологичного объекта, обеспечивающую матрицу для обратной транскрипции, которая дополнительно содержит предполагаемую конечную последовательность-мишень; и (iv) последовательность каркаса gRNA для связывания с доменом Cas9 слитого полипептида Cas9-RT; и 3) необязательную дополнительную gRNA для обеспечения внесения однонитевого разрыва второй нити вблизи сайта-мишени, например, РНК, содержащую (i) спейсерную последовательность для нацеливания управляющего элемента с индукцией второго однонитевого разрыва на противоположной нити от первого однонитевого разрыва, направляемого РНК-матрицей, в сайте, проксимальном к сайту-мишени (например, в пределах 50-150 нуклеотидов от первого однонитевого разрыва); и (ii) каркасную последовательность gRNA, опосредующую ассоциацию управляющего элемента с доменом Cas9. В этом примере слитый полипептид Cas-RT (1) (таблица 46) содержит никазу Cas9(N863A), слитую с

доменом обратной транскриптазы MMLV. Используемые в данном документе РНК-матрицы (2) специально расположены в ориентации от 5' к 3' (i), (iv), (iii), (ii), как указано в описании и подробно описано в таблице 48 и примере 35.

Для доставки системы Gene Writer на основе РНК в первичные CD4⁺ Т-клетки человека и валидации экспрессии белка 1000000 клеток (CD4⁺ Т-клетки периферической крови человека, № по каталогу Lonza 2W-200) стимулировали с помощью CD3/CD28 в течение двух дней, а затем нуклеофицировали с помощью 0, 2,5, 5 или 10 мкг мРНК, кодирующей полипептид Cas-RT, с использованием 96-луночной системы Nucleofector Shuttle (Lonza) с программой нуклеофекции EO-115 в соответствии с протоколами производителя. Через один день после нуклеофекции эффективность доставки оценивали с помощью иммуноблоттинга с антителом Cas9 (Cell Signaling) для измерения экспрессии белка у полипептида Gene Writer из нуклеофицированной мРНК (фигура 45А).

Затем первичные Т-клетки CD4⁺ человека подвергали нуклеофекции с помощью: (1) 5 мкг мРНК полипептида Gene Writer (контроль только с записывающим доменом); (2) 5 мкг мРНК полипептида Gene Writer и 5 мкМ РНК-матрицы, например, одной из шести РНК-матриц из таблицы 48, которые нацелены на один и тот же сайт локуса HEK3, но отличаются результатом редактирования или конструкцией; или (3) 5 мкг мРНК полипептида Gene Writer, 5 мкМ РНК-матрицы, например, одной из шести РНК-матриц из таблицы 48, и 2,075 мкМ дополнительной gRNA для создания одностороннего разрыва второй нити, например, gRNA для второго одностороннего разрыва, нацеленная на последовательность на 108 нуклеотидов выше сайта-мишени HEK3, описана в примере 35. Через три дня после нуклеофекции клетки собирали для изучения 1) жизнеспособности клеток после доставки РНК с помощью системы Gene Writer и 2) эффективности редактирования в сайте-мишени генома. Для оценки жизнеспособности клеток процентную долю живых клеток измеряли с помощью проточной цитометрии после окрашивания клеток флуоресцентным красителем для выявления живых/мертвых клеток (BioLegend). Жизнеспособность клеток была сравнима в экспериментальных условиях и в отсутствие нуклеофекции (необработанный контроль) (фигура 45 В). Для оценки эффективности редактирования с помощью систем Gene Writing геномную ДНК анализировали методом секвенирования ампликонов на основе ПЦР, описанным в примере 35. Эффективность желаемого редактирования (точной записи) достигала примерно 6,3% при использовании матрицы 1 (таблица 48) с мРНК полипептида Gene Writer (фигуры 46 А и В). В данном документе добавление gRNA для второго одностороннего разрыва (фигура 46 В) приводило к сходным уровням редактирования. Таким образом, этот пример демонстрирует использование систем Gene Writing для высокоспецифичного редактирования в первичных Т-клетках и дополнительно показывает успешное применение в этих клетках Gene Writing, состоящей только из РНК без использования ДНК.

128. Пример 39. Обнаружение интеграции, опосредованной ретротранспозазой, в клетках человека

В этом примере описывается идентификация ретротранспозонов, демонстрирующих функциональные свойства в клетках человека. Путем анализа нативных или модифицированных ретротранспозонов на активность 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, как описано в этом примере, используется для обеспечения активности обратной транскриптазы в полипептиде Gene Writer, например, в качестве RT слитого полипептида Cas-RT. 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 50, содержащую кодон-оптимизированную последовательность, кодирующую ретротранспозазу человека, слитую с меткой HiBit для обнаружения 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 48). В данном документе репортер GFP кодируется в отсутствие промотора, управляющего его экспрессией, чтобы избежать потери сигнала из-за токсичности GFP (см. фигуру 48).

Чтобы доставить двухплазмидную систему в клетки U2OS, около 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 антисмысловый интрон, как описано ниже. Экспрессию белка-управляющего элемента измеряли с помощью биolumинесцентного 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 подвергается сплайсингу во время транскрипции, таким образом, только репортерная ДНК, полученная из транскрипта путем обратной транскрипции, не будет иметь последовательность интрона (фигура 48). Для ограничения обнаружения 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 50, анализировали согласно этому примеру для определения активности в клетках человека. Анализ эффективности интеграции 163 систем ретротранспозонов-кандидатов с помощью ddPCR показан на фигуре 49. 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.

129. Пример 40. 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 tandemной масс-спектрометрией (МС/МС) 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 обнаружение 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.

130. Пример 41. 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 обнаружения химических модификаций, обусловленных контаминацией альдегидами липидных реагентов (см., например, пример 40).

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 обнаруживают 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 42. Gene Writer™ позволяет осуществлять вставку крупного участка в геномную ДНК

В этом примере описано применение 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 нуклеазу Cas9 в качестве 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 односторонний разрыв в сайте-мишени, а затем выполняет TPRT на матрице, иницируя реакцию с использованием праймирующих областей на матрице, которые комплементарны последовательности, непосредственно примыкающей к сайту одностороннего 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 инкубационного периода 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.

131. Пример 43. Gene Writer могут интегрировать генетический груз независимо от пути репарации посредством односторонней матрицы

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации посредством односторонней 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 обнаруживают по отсутствию основных компонентов пути SSTR.

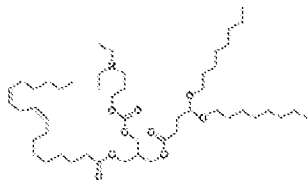
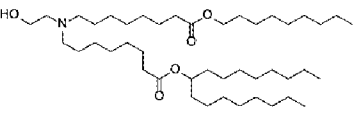
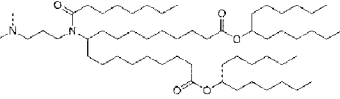
132. Пример 44. Состав на основе липидных наночастиц, инкапсулирующих мРНК люциферазы светлячка

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(А)-хвост. мРНК синтезировали в стандартных условиях для транскрипции РНК-полимеразой T7 *in vitro* с экпированием во время транскрипции, однако с 100% заменой нуклеотидтрифосфата UTP на N1-метилпсевдоуридинтрифосфат в реакции. Очищенную мРНК растворяли в 25 мМ цитрате натрия, рН 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 NanoAssembler™ 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% буфере сахарозы 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 44 (формула (ix), (vii) и (iii))

ID ЛИПИДА	Химическое название	Молекулярная масса	Структура
LIPIDV00 3	(9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-(((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат	852,29	
LIPIDV00 4	Гептадекан-9-ил 8-((2-гидроксиэтил)(8-(нонилокси)-8-оксооктил)амино)октаноат	710,18	
LIPIDV00 5		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 А2**.

Таблица А2. 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, обнаруженного, 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 45. 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 1х среду Е Уильяма без фенолового красного и инкубировали при 37°C с 5% CO₂. Через 4 часа среду заменяли поддерживающей средой (1х среда Е Уильяма без фенола, содержащая Hepatocyte Maintenance Supplement Pack

(ThermoFisher Scientific)) и клетки выращивали в течение ночи при 37°C с 5% CO₂. 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 активности люциферазы. 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 активности люциферазы.

Для проведения анализа активности люциферазы клеточные лизаты в буфере для пассивного лизиса размораживали, переносили в круглодонный 96-луночный микротитрационный планшет и центрифугировали при 15000g при 4°C в течение 3 минут с удалением клеточного дебриса. Концентрацию белка измеряли для каждого образца с использованием набора для анализа белка Pierce™ BCA Protein Assay Kit (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 активности люциферазы проводили в 96-луночных микротитрационных планшетах 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. 50А и В, и указывают на успешную LNP-опосредованную доставку РНК в первичные клетки в культуре. Как показано на фиг. 50А, LNP, составленные в соответствии с примером 44, анализировали в отношении доставки груза в первичные гепатоциты человека (А) и мыши (В) в соответствии с примером 45. 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 46. 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 44, и тестировали *in vitro* (пример 45) 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 активности люциферазы для каждого образца определяли концентрацию белка в гомогенатах печени с использованием набора для анализа белка Pierce™ 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. 51). Как показано на фиг. 51, LNP, содержащие мРНК люциферазы светлячка, составляли и доставляли мышам iv, и образцы печени собирали и анализировали в отношении активности люциферазы через 6, 24 и 48 часов после введения. Активность репортерного гена в различных составах соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003. Экспрессия РНК была транзистентной, 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 одностороннего 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, с учетом того, 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, коррекцию 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, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нуклеазный 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 домен RT содержит последовательность из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

2. 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, нуклеазный 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 домен RT содержит последовательность из таблицы 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2,

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.

3. 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, нуклеазный 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, 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 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.

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, нуклеазный 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 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.

5. Система по любому из предыдущих пунктов, 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.

6. 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, нуклеазный 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 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.

7. 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, нуклеазный 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, 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 (a)(ii) и/или (a)(iii) содержит домен TAL; домен "цинковый палец" или домен CRISPR/Cas, выбранный из таблицы 3, или их функциональный вариант (например, мутантную форму).

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 (например, спейсер CRISPR), 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 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.

9. 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) последовательность, которая специфично связывает домен RT, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

10. Система по п. 9, где РНК-матрица дополнительно содержит последовательность, которая связывает (a)(ii) и/или (a)(iii).

11. 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 с мишенью.

12. 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, вторую 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 с мишенью.

13. 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.

14. РНК-матрица по п. 13, дополнительно содержащая (v) последовательность, которая связывает эндонуклеазный и/или ДНК-связывающий домен полипептида (например, того же полипептида, который содержит домен RT).

15. РНК-матрица по любому из п. 13 или п. 14, где домен RT содержит последовательность, выбранную из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, которая характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

16. РНК-матрица по любому из п. 13 или п. 14, где домен RT содержит последовательность, выбранную из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, 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.

17. РНК-матрица по любому из пп. 13-16, где последовательность из (ii) специфично связывается с доменом RT.

18. РНК-матрица по любому из пп. 13-17, где последовательность, которая специфично связывает домен RT, представляет собой последовательность, например, последовательность UTR, из таблицы 1 или из домена из таблицы 2 или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

19. 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.

20. 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 эндонуклеазный и/или ДНК-связывающий домен полипептида.

21. Система или РНК-матрица по любому из предыдущих пунктов, где РНК-матрица, первая РНК-матрица или вторая РНК-матрица содержит последовательность, которая специфично связывает домен RT.

22. Система для модификации ДНК, содержащая:

(a) первую РНК-матрицу (или ДНК, кодирующую первую РНК-матрицу), содержащую (i) последовательность, которая связывает эндонуклеазный домен, например, нуклеазный домен, и/или ДНК-связывающий домен (DBD) полипептида, и (ii) последовательность, которая связывает сайт-мишень (например, вторую 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 гетерологичного объекта и (iii) 3'-концевой домен гомологии с мишенью.

23. Система по п. 14, где нуклеиновая кислота, кодирующая первую РНК-матрицу, и нуклеиновая кислота, кодирующая вторую РНК-матрицу, являются двумя отдельными нуклеиновыми кислотами.

24. Система по п. 14, 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.

25. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нуклеазный домен, где домен RT имеет последовательность из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними.

26. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нуклеазный домен, где домен RT имеет последовательность из таблицы 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2, 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.

27. 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 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними; 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, нисказный домен.

28. 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 1 или 3 или последовательность домена обратной транскриптазы из таблицы 2, 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; 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, нисказный домен.

29. Система по любому из п. 27 или п. 28, где нуклеиновая кислота, кодирующая первый полипептид, и нуклеиновая кислота, кодирующая второй полипептид, являются двумя отдельными нуклеиновыми кислотами.

30. Система по любому из п. 27 или п. 28, 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.

31. Реакционная смесь, содержащая:

клетку и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов.

32. Реакционная смесь, содержащая:

ДНК, содержащую сайт-мишень, и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов.

33. Набор, содержащий:

систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов;

инструкции по применению системы, полипептида, РНК-матрицы или ДНК, кодирующей их; а также

одну или обе из клетки или ДНК, содержащих сайт-мишень.

34. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен TAL.

35. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен "цинковый палец".

36. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен CRISPR/Cas.

37. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен представляет собой нуклеазный домен.

38. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен содержит домен CRISPR/Cas.

39. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен CRISPR/Cas содержит домен или полипептид из таблицы 4 или его функциональный вариант (например, мутантную форму).

40. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен CRISPR/Cas содержит домен или полипептид из рода/вида из таблицы 4.

41. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен содержит нуклеазу типа II (например, FokI), резольвазу для структур Холлидея или нуклеазу для двухнитевой ДНК, содержащую изменение, которое приводит к устранению ее способности разрезать одну нить (например, превращающее нуклеазу для двухнитевой ДНК в нуклеазу).

42. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT содержит обратную транскриптазу или ее функциональный фрагмент или вариант, выбранный из таблицы 1 или 3, или последовательность домена обратной транскриптазы из таблицы 2.

43. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT содержит одну или несколько мутаций (например, вставку, делецию или замену) по сравнению со встречающимся в природе доменом RT, или доменом RT или функциональным фрагментом, выбранным из таблицы 1 или 3, или последовательностью домена обратной транскриптазы из таблицы 2, например, по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 мутаций.

44. Система, набор, полипептид или реакционная смесь по п. 43, где одна или несколько мутаций выбраны из D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K или D653N в домене RT обратной транскриптазы вируса лейкоза мышей или соответствующей мутации в соответствующем положении другого домена RT.

45. Система, набор, полипептид или реакционная смесь по п. 43, где одна или несколько мутаций выбраны из WO2018089860A1, включенной в данный документ посредством ссылки (например, мутация C952S, и/или C956S, и/или C952S, C956S (двойная мутация), и/или C969S, и/или H970Y, и/или R979Q, и/или R976Q, и/или R1071S, и/или R328A, и/или R329A, и/или Q336A, и/или R328A, R329A, Q336A (тройная мутация), и/или G426A, и/или D428A, и/или G426A, D428A (двойная мутация), и/или любая их комбинация;

где положения указаны относительно SEQ ID NO: 52 из WO2018089860A1), в домене RT ретротранспозазы R2Bm или соответствующей мутации в соответствующем положении другого домена RT.

46. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с С-концевой стороны от DBD в полипептиде.

47. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с С-концевой стороны от никазного домена в полипептиде.

48. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с N-концевой стороны от DBD в полипептиде.

49. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с N-концевой стороны от никазного домена в полипептиде.

50. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где полипептид содержит линкер, например, расположенный между доменом RT и доменом DBD или доменом RT и никазным доменом.

51. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где длина линкера составляет 2-50, например, 2-30, аминокислот.

52. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где линкер представляет собой гибкий линкер, например, содержащий остатки Gly и/или Ser.

53. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью комплементарен последовательности, прилегающей к сайту, подлежащему модификации системой, или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной последовательности, прилегающей к сайту, подлежащему модификации системой.

54. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью имеет длину более 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, 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 (например, имеет длину 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 или 30 нуклеотидов).

55. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью имеет длину не более 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, 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.

56. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту,

подлежащему модификации системой, за исключением положения или положений, подлежащих модификации.

57. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением положений, кодирующих последовательность, подлежащую вставке в сайт.

58. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением того, что последовательность гетерологичного объекта не содержит нуклеотидов, кодирующих последовательность, подлежащую делеции в сайте.

59. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта имеет длину более 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 или 30 нуклеотидов (например, имеет длину 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 или 30 нуклеотидов).

60. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта имеет длину не более 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 или 30 нуклеотидов.

61. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к замене по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидами нуклеотидов не из сайта-мишени.

62. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к вставке 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 или 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 в сайт-мишень.

63. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к делеции 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, 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов.

64. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

65. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), имеет длину по меньшей мере 45, 50,

55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 85 или 80 нуклеотидов).

66. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид, связывает эндонуклеазный домен и/или домен DBD.

67. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), содержит gRNA.

68. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), имеет длину по меньшей мере 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21 или 20 нуклеотидов), например, имеет длину 17, 18, 19, 20, 21, 22, 23 или 24 нуклеотида.

69. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень, комплементарна второй нити сайта-мишени или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной второй нити сайта-мишени.

70. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

71. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, дополнительно содержащие gRNA, нацеленную на вторую нить, которая направляет эндонуклеазный домен (например, никазный) домен для осуществления однонитевого разрыва второй нити (например, в геноме-мишени).

72. Система, набор, РНК-матрица или реакционная смесь по п. 71, где РНК-матрица дополнительно содержит gRNA, нацеленную на вторую нить.

73. Система, набор, РНК-матрица или реакционная смесь по п. 71, где gRNA, нацеленная на вторую нить, расположена в нуклеиновой кислоте, отдельной от РНК-матрицы.

74. Система, набор, РНК-матрица или реакционная смесь по любому из пп. 71-73, где gRNA направляет эндонуклеазный домен (например, никазный домен) для осуществления однонитевого разрыва второй нити (например, в геноме-мишени) в сайте, который находится 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, однонитевого разрыва в первой нити).

75. Система, набор, РНК-матрица или реакционная смесь по любому из пп. 71-74, где gRNA специфично связывается с первой нитью.

76. Система, содержащая:

мРНК, кодирующую полипептид или систему по любому из предыдущих пунктов, и

РНК-матрицу по любому из предыдущих пунктов.

77. Система по п. 76, где мРНК, кодирующая полипептид или систему по любому из предыдущих пунктов, и РНК-матрица по любому из предыдущих пунктов расположены в разных молекулах нуклеиновой кислоты.

78. Система, содержащая молекулу РНК, содержащая:

РНК-матрицу (или РНК, кодирующую РНК-матрицу) по любому из предыдущих пунктов и

последовательность, кодирующую систему или полипептид по любому из предыдущих пунктов.

79. Система по п. 78, где молекула РНК содержит участок внутренней посадки рибосомы, например, функционально связанный с последовательностью, кодирующей систему или полипептид.

80. Система по любому из п. 78 или п. 79, где молекула РНК содержит сайт расщепления, например, расположенный между РНК-матрицей (или РНК, кодирующей РНК-матрицу) и последовательностью, кодирующей систему или полипептид.

81. Система или полипептид по любому из предыдущих пунктов, где полипептид содержит расщепленный интеин, например, два или несколько (например, все) из домена RT, DBD, эндонуклеазного (например, никазного) домена или их комбинации транслируются как отдельные белки, которые объединяются в один полипептид путем сплайсинга белков.

82. Липидная наночастица (LNP), содержащая систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, по любому из предыдущих пунктов.

83. Система, содержащая:

первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы или полипептид по любому из предыдущих пунктов; и

вторую липидную наночастицу, содержащую РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу по любому из предыдущих пунктов.

84. Вирус, вирусоподобная частица, фузосома или виросома, содержащие систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, по любому из предыдущих пунктов.

85. Система, содержащая:

первые вирус, вирусоподобную частицу, фузосому или виросому, содержащие полипептид (или кодирующую его ДНК или РНК) системы или полипептид по любому из предыдущих пунктов; и

вторые вирус, вирусоподобную частицу, фузосому или виросому, содержащие РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу по любому из предыдущих пунктов.

86. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы имеет длину более 100, 125, 150, 175 или 200 нуклеотидов или имеет длину по меньшей мере 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 имеет длину менее 15, 10, 5 или 20 тысяч оснований или имеет длину менее 500, 400, 300 или 200 нуклеотидов).

87. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит поли(А)-хвост (например, поли(А)-хвост, который имеет длину по меньшей мере 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов).

88. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит:

5'-кэп, например: 7-метилгуанозиновый кэп (например, кэп O-Me-m⁷G); гиперметилованный 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)), и/или

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-метилованных NTP или фосфотиоат-модифицированных NTP; химической модификации одного нуклеотида (например, окисления 3'-концевой рибозы до реакционноспособного альдегида с последующей конъюгацией альдегид-реактивного модифицированного нуклеотида) или химического лигирования с другой молекулой нуклеиновой кислоты.

89. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где РНК-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, C5-пропинилдезоксцитидина (pdC), C5-пропинилдеоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдеоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-

цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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).

90. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит один или несколько модифицированных нуклеотидов.

91. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где после теста на стабильность по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы остаются интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов или длиной по меньшей мере 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).

92. Система, набор или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 1% сайтов-мишеней являются модифицированными после анализа системы на активность.

93. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где система, полипептид, РНК-матрица и/или ДНК, кодирующая их, составлены в виде липидной наночастицы (LNP).

94. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где ДНК, кодирующая систему, полипептид и/или РНК-матрица упакованы в вирус, вирусоподобную частицу, виросому, липосому, везикулу, экзосому или LNP.

95. Система, набор, РНК-матрица, полипептид или реакционная смесь по п. 94, где ДНК, кодирующая систему, РНК-матрица или полипептид упакованы в аденоассоциированный вирус (AAV).

96. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где система, РНК-матрица, полипептид, липидная наночастица (LNP), вирус, вирусоподобная частица или виросома свободны или по сути свободны от контаминации пирогенами, вирусами, грибами, бактериальными патогенами и/или белками клетки-хозяина.

97. Вирус, вирусоподобная частица или виросома, содержащие:
систему, РНК-матрицу или полипептид по любому из предыдущих пунктов или ДНК, кодирующую любое из них, и
капсидный белок аденоассоциированного вируса (AAV).

98. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома по любому из предыдущих пунктов, где система, РНК-матрица и/или полипептид активны в ткани-мишени и менее активны (например, не активны) в ткани, не являющейся мишенью.

99. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома по п. 98, дополнительно содержащие одну или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных для ткани-мишени, где одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных для ткани-мишени, находятся в функциональной ассоциации с РНК-матрицей, полипептидом или нуклеиновой кислотой, кодирующей его, или ими обоими.

100. Способ изготовления РНК-матрицы, включающий:

(a) получение РНК-матрицы по любому из предыдущих пунктов и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа одного или нескольких модифицированных нуклеотидов (например, выбранных из дигидроуридина, инозина, 7-метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-О-метилриботимидина, 2'-О-этилриботимидина, 2'-фторриботимидина, C5-пропинилдезоксцитидина (pdC), C5-пропинилдезоксйуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдезоксйуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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 РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу; или

(vii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли РНК-матрица от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

101. Способ изготовления системы для модификации ДНК, включающий:

(a) получение системы для модификации ДНК по любому из предыдущих пунктов и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа 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, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу;

(vii) длины 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, имеют 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) присутствия, отсутствия и/или типа посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% полипептида, первого полипептида или второго полипептида, выбранную посттрансляционную модификацию;

(ix) присутствия, отсутствия и/или типа одной или нескольких искусственных, синтетических или неканонических аминокислот в полипептиде, первом полипептиде или втором полипептиде, например, содержат 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) стабильности полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактными (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) активности полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей полипептид, первый полипептид или второй полипептид; или

(xii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

102. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с:

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

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

103. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с системой, полипептидом, РНК-матрицей или ДНК, кодирующей их, по любому из предыдущих пунктов;

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

104. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с:

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

за счет чего обеспечивается модификация сайта-мишени в геномной ДНК в клетке, где клетка является эуплоидной, не является иммортализованной, является частью ткани, является частью организма, является первичной клеткой, является неделящейся, является гаплоидной (например, клеткой зародышевой линии), является нераковой полиплоидной клеткой или получена от субъекта, у которого имеется генетическое заболевание.

105. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту:

(а) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нисказный домен; и

(b) РНК-матрицы (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;

за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

106. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту системы, полипептида, РНК-матрицы или ДНК, кодирующей их, по любому из предыдущих пунктов;

за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

107. Способ по любому из п. 105 или п. 106, где заболевание или состояние, ассоциированное с генетическим дефектом, представляет собой показание, указанное в любой из таблиц 9-12, и/или где генетический дефект представляет собой дефект гена, указанного в любой из таблиц 9-12.

108. Способ по любому из пп. 105-107, где субъект является пациентом-человеком.

По доверенности

Формула изобретения

1. 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, нуклеазный 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 домен RT содержит последовательность из таблицы 30 или 44 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

2. 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, нуклеазный 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 домен RT содержит последовательность из таблицы 30 или 44,

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.

3. 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, нуклеазный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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, 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 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.

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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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 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.

5. Система по любому из предыдущих пунктов, 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.

6. 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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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 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.

7. 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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; и

(b) 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 (a)(ii) и/или (a)(iii) содержит домен TAL; домен "цинковый палец" или домен CRISPR/Cas, выбранный из таблицы 4, или их функциональный вариант (например, мутантную форму).

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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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, 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 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.

9. 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; где домен RT содержит последовательность из таблицы 30 или 44; 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) последовательность, которая специфично связывает домен RT, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

10. Система по п. 9, где РНК-матрица дополнительно содержит последовательность, которая связывает (a)(ii) и/или (a)(iii).

11. 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, где домен RT содержит последовательность из таблицы 30 или 44,

(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 с мишенью.

12. 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, нуклеазный домен; где домен RT содержит последовательность из таблицы 30 или 44;

(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, вторую 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 с мишенью.

13. 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 ретровирусного полипептида, (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии с мишенью.

14. РНК-матрица по п. 13, дополнительно содержащая (v) последовательность, которая связывает эндонуклеазный и/или ДНК-связывающий домен полипептида (например, того же полипептида, который содержит домен RT).

15. РНК-матрица по любому из п. 13 или п. 14, где домен RT содержит последовательность, выбранную из таблицы 30 или 44, или последовательность, которая характеризуется по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

16. РНК-матрица по любому из п. 13 или п. 14, где домен RT содержит последовательность, выбранную из таблицы 30 или 44, 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.

17. РНК-матрица по любому из пп. 13-16, где последовательность из (ii) специфично связывается с доменом RT.

18. РНК-матрица по любому из пп. 13-17, где последовательность, которая специфично связывает домен RT, представляет собой последовательность, например, последовательность UTR, из таблицы 30 или 44 или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

19. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен ретровирусного 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.

20. 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 эндонуклеазный и/или ДНК-связывающий домен ретровирусного полипептида.

21. Система или РНК-матрица по любому из предыдущих пунктов, где РНК-матрица, первая РНК-матрица или вторая РНК-матрица содержит последовательность, которая специфично связывает домен RT.

22. Система для модификации ДНК, содержащая:

(a) первую РНК-матрицу (или ДНК, кодирующую первую РНК-матрицу), содержащую (i) последовательность, которая связывает эндонуклеазный домен, например, нуклеазный домен, и/или ДНК-связывающий домен (DBD) полипептида, и (ii) последовательность, которая связывает сайт-мишень (например, вторую 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

гетерологичного объекта и (iii) 3'-концевой домен гомологии с мишенью; где домен RT полипептида содержит последовательность из таблицы 30 или 44.

23. Система по п. 14, где нуклеиновая кислота, кодирующая первую РНК-матрицу, и нуклеиновая кислота, кодирующая вторую РНК-матрицу, являются двумя отдельными нуклеиновыми кислотами.

24. Система по п. 14, 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.

25. 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 30 или 44 или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

26. 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 30 или 44, 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.

27. 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 30 или 44 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.

28. 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 30 или 44, 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; 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, нисказный домен.

29. Система по любому из п. 27 или п. 28, где нуклеиновая кислота, кодирующая первый полипептид, и нуклеиновая кислота, кодирующая второй полипептид, являются двумя отдельными нуклеиновыми кислотами.

30. Система по любому из п. 27 или п. 28, 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.

31. Реакционная смесь, содержащая:

клетку и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов.

32. Реакционная смесь, содержащая:

ДНК, содержащую сайт-мишень, и любую систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов.

33. Набор, содержащий:

систему, полипептид, РНК-матрицу или ДНК, кодирующую их, по любому из предыдущих пунктов;

инструкции по применению системы, полипептида, РНК-матрицы или ДНК, кодирующей их; а также

одну или обе из клетки или ДНК, содержащих сайт-мишень.

34. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен TAL.
35. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен "цинковый палец".
36. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где DBD содержит домен CRISPR/Cas.
37. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен представляет собой никазный домен.
38. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен содержит домен CRISPR/Cas.
39. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен CRISPR/Cas содержит домен или полипептид из таблицы 4 или его функциональный вариант (например, мутантную форму).
40. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен CRISPR/Cas содержит домен или полипептид из рода/вида из таблицы 4.
41. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где эндонуклеазный домен содержит нуклеазу типа II (например, FokI), резольвазу для структур Холлидея или нуклеазу для двухнитевой ДНК, содержащую изменение, которое приводит к устранению ее способности разрезать одну нить (например, превращающее нуклеазу для двухнитевой ДНК в никазу).
42. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT содержит обратную транскриптазу или ее функциональный фрагмент или вариант, выбранный из таблицы 30 или 44.

43. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT содержит одну или несколько мутаций (например, вставку, делецию или замену) по сравнению со встречающимся в природе доменом RT или доменом RT или функциональным фрагментом, выбранным из таблицы 30 или 44, например, по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 мутаций.

44. Система, набор, полипептид или реакционная смесь по п. 43, где одна или несколько мутаций выбраны из D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K или D653N в домене RT обратной транскриптазы вируса лейкоза мышей или соответствующей мутации в соответствующем положении другого домена RT.

45. Система, набор, полипептид или реакционная смесь по п. 43, где одна или несколько мутаций выбраны из WO2018089860A1, включенной в данный документ посредством ссылки (например, мутация C952S, и/или C956S, и/или C952S, C956S (двойная мутация), и/или C969S, и/или H970Y, и/или R979Q, и/или R976Q, и/или R1071S, и/или R328A, и/или R329A, и/или Q336A, и/или R328A, R329A, Q336A (тройная мутация), и/или G426A, и/или D428A, и/или G426A, D428A (двойная мутация), и/или любая их комбинация; где положения указаны относительно SEQ ID NO: 52 из WO2018089860A1), в домене RT ретротранспозазы R2Bm или соответствующей мутации в соответствующем положении другого домена RT.

46. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с С-концевой стороны от DBD в полипептиде.

47. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с С-концевой стороны от никазного домена в полипептиде.

48. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с N-концевой стороны от DBD в полипептиде.

49. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где домен RT расположен с N-концевой стороны от никазного домена в полипептиде.

50. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где полипептид содержит линкер, например, расположенный между доменом RT и доменом DBD или доменом RT и никазным доменом.

51. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где длина линкера составляет 2-50, например, 2-30, аминокислот.

52. Система, набор, полипептид или реакционная смесь по любому из предыдущих пунктов, где линкер представляет собой гибкий линкер, например, содержащий остатки Gly и/или Ser.

53. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью комплементарен последовательности, прилегающей к сайту, подлежащему модификации системой, или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной последовательности, прилегающей к сайту, подлежащему модификации системой.

54. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью имеет длину более 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, 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 (например, имеет длину 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 или 30 нуклеотидов).

55. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где 3'-концевой домен гомологии с мишенью имеет длину не более 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, 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.

56. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением положения или положений, подлежащих модификации.

57. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением положений, кодирующих последовательность, подлежащую вставке в сайт.

58. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта комплементарна сайту, подлежащему модификации системой, за исключением того, что последовательность гетерологичного объекта не содержит нуклеотидов, кодирующих последовательность, подлежащую делеции в сайте.

59. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта имеет длину более 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 или 30 нуклеотидов (например, имеет длину 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 или 30 нуклеотидов).

60. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта имеет длину не более 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 или 30 нуклеотидов.

61. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к замене по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидами нуклеотидов не из сайта-мишени.

62. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к вставке 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 или 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 в сайт-мишень.

63. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта приводит к делеции 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, 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов.

64. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность гетерологичного объекта отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

65. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), имеет длину по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 85 или 80 нуклеотидов).

66. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид, связывает эндонуклеазный домен и/или домен DBD.

67. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), содержит gRNA.

68. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), имеет длину по меньшей мере 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120 или 130 нуклеотидов (и необязательно имеет длину не более 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30,

29, 28, 27, 26, 25, 24, 23, 22, 21 или 20 нуклеотидов), например, имеет длину 17, 18, 19, 20, 21, 22, 23 или 24 нуклеотида.

69. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень, комплементарна второй нити сайта-мишени или содержит не более 1, 2, 3, 4 или 5 ошибок спаривания с последовательностью, комплементарной второй нити сайта-мишени.

70. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), отделена от последовательности, которая связывает полипептид (например, которая связывает эндонуклеазный домен и/или домен DBD), по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 или 30 нуклеотидами.

71. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, дополнительно содержащие gRNA, нацеленную на вторую нить, которая направляет эндонуклеазный домен (например, никазный) домен для осуществления одонитевого разрыва второй нити (например, в геноме-мишени).

72. Система, набор, РНК-матрица или реакционная смесь по п. 71, где РНК-матрица дополнительно содержит gRNA, нацеленную на вторую нить.

73. Система, набор, РНК-матрица или реакционная смесь по п. 71, где gRNA, нацеленная на вторую нить, расположена в нуклеиновой кислоте, отдельной от РНК-матрицы.

74. Система, набор, РНК-матрица или реакционная смесь по любому из пп. 71-73, где gRNA направляет эндонуклеазный домен (например, никазный домен) для осуществления одонитевого разрыва второй нити (например, в геноме-мишени) в сайте, который находится 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, одонитевого разрыва в первой нити).

75. Система, набор, РНК-матрица или реакционная смесь по любому из пп. 71-74, где gRNA специфично связывается с первой нитью.
76. Система, содержащая:
мРНК, кодирующую полипептид или систему по любому из предыдущих пунктов,
и
РНК-матрицу по любому из предыдущих пунктов.
77. Система по п. 76, где мРНК, кодирующая полипептид или систему по любому из предыдущих пунктов, и РНК-матрица по любому из предыдущих пунктов расположены в разных молекулах нуклеиновой кислоты.
78. Система, содержащая молекулу РНК, содержащая:
РНК-матрицу (или РНК, кодирующую РНК-матрицу) по любому из предыдущих пунктов и
последовательность, кодирующую систему или полипептид по любому из предыдущих пунктов.
79. Система по п. 78, где молекула РНК содержит участок внутренней посадки рибосомы, например, функционально связанный с последовательностью, кодирующей систему или полипептид.
80. Система по любому из п. 78 или п. 79, где молекула РНК содержит сайт расщепления, например, расположенный между РНК-матрицей (или РНК, кодирующей РНК-матрицу) и последовательностью, кодирующей систему или полипептид.
81. Система или полипептид по любому из предыдущих пунктов, где полипептид содержит расщепленный интеин, например, два или несколько (например, все) из домена RT, DBD, эндонуклеазного (например, никазного) домена или их комбинации транслируются как отдельные белки, которые объединяются в один полипептид путем сплайсинга белков.

82. Липидная наночастица (LNP), содержащая систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, по любому из предыдущих пунктов.

83. Система, содержащая:

первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы или полипептид по любому из предыдущих пунктов; и

вторую липидную наночастицу, содержащую РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу по любому из предыдущих пунктов.

84. Вирус, вирусоподобная частица, фузосома или виросома, содержащие систему, РНК-матрицу, полипептид (или РНК, кодирующую его) или ДНК, кодирующую систему, РНК-матрицу или полипептид, по любому из предыдущих пунктов.

85. Система, содержащая:

первые вирус, вирусоподобную частицу, фузосому или виросому, содержащие полипептид (или кодирующую его ДНК или РНК) системы или полипептид по любому из предыдущих пунктов; и

вторые вирус, вирусоподобную частицу, фузосому или виросому, содержащие РНК-матрицу (или кодирующую ее ДНК) системы или РНК-матрицу по любому из предыдущих пунктов.

86. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы имеет длину более 100, 125, 150, 175 или 200 нуклеотидов или имеет длину по меньшей мере 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 имеет длину менее 15, 10, 5 или 20 тысяч оснований или имеет длину менее 500, 400, 300 или 200 нуклеотидов).

87. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит поли(А)-хвост (например, поли(А)-хвост, который имеет длину по меньшей мере 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов).

88. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит:

5'-кэп, например: 7-метилгуанозиновый кэп (например, кэп O-Me-m7G); гиперметилованный 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)), и/или

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-метилованных NTP или фосфотиоат-модифицированных NTP; химической модификации одного нуклеотида (например, окисления 3'-концевой рибозы до реакционноспособного альдегида с последующей конъюгацией альдегид-реактивного модифицированного нуклеотида) или химического лигирования с другой молекулой нуклеиновой кислоты.

89. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где РНК-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, C5-пропинилдезоксцитидина (pdC), C5-пропинилдеоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдезоксцитидина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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).

90. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы содержит один или несколько модифицированных нуклеотидов.

91. Система, набор, РНК-матрица или реакционная смесь по любому из предыдущих пунктов, где после теста на стабильность по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы остаются интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов или длиной по меньшей мере 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).

92. Система, набор или реакционная смесь по любому из предыдущих пунктов, где по меньшей мере 1% сайтов-мишеней являются модифицированными после анализа системы на активность.

93. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где система, полипептид, РНК-матрица и/или ДНК, кодирующая их, составлены в виде липидной наночастицы (LNP).

94. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где ДНК, кодирующая систему, полипептид и/или РНК-матрица упакованы в вирус, вирусоподобную частицу, виросому, липосому, везикулу, экзосому или LNP.

95. Система, набор, РНК-матрица, полипептид или реакционная смесь по п. 94, где ДНК, кодирующая систему, РНК-матрица или полипептид упакованы в аденоассоциированный вирус (AAV).

96. Система, набор, РНК-матрица, полипептид или реакционная смесь по любому из предыдущих пунктов, где система, РНК-матрица, полипептид, липидная наночастица (LNP), вирус, вирусоподобная частица или виросома свободны или по сути свободны от контаминации пирогенами, вирусами, грибами, бактериальными патогенами и/или белками клетки-хозяина.

97. Вирус, вирусоподобная частица или виросома, содержащие:

систему, РНК-матрицу или полипептид по любому из предыдущих пунктов или ДНК, кодирующую любое из них, и

капсидный белок аденоассоциированного вируса (AAV).

98. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома по любому из предыдущих пунктов, где система, РНК-матрица и/или полипептид активны в ткани-мишени и менее активны (например, не активны) в ткани, не являющейся мишенью.

99. Система, набор, РНК-матрица, полипептид, вирус, вирусоподобная частица или виросома по п. 98, дополнительно содержащие одну или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных для ткани-мишени, где одна или несколько первых тканеспецифических последовательностей, контролирующих экспрессию, специфичных для ткани-мишени, находятся в функциональной ассоциации с РНК-матрицей, полипептидом или нуклеиновой кислотой, кодирующей его, или ими обоими.

100. Способ изготовления РНК-матрицы, включающий:

(a) получение РНК-матрицы по любому из предыдущих пунктов и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа одного или нескольких модифицированных нуклеотидов (например, выбранных из дигидроуридина, инозина, 7-

метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-О-метилриботимидина, 2'-О-этилриботимидина, 2'-фторриботимидина, C5-пропинилдезоксцитидина (pdC), C5-пропинилдезоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксцитидина, 5-метоксиметилдезоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксцитидина, C5-пропинил-f-цитидина (pfC), C5-пропинил-f-уридина (pfU), 5-метил-f-цитидина, 5-метил-f-уридина, C5-пропинил-m-цитидина (pmC), C5-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 РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу; или

(vii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли РНК-матрица от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

101. Способ изготовления системы для модификации ДНК, включающий:

(a) получение системы для модификации ДНК по любому из предыдущих пунктов и

(b) анализ одного или нескольких из:

(i) длины РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеют 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) присутствия, отсутствия и/или длины поли(А)-хвоста в РНК-матрице, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99%

присутствующей РНК-матрицы поли(А)-хвост (например, поли(А)-хвост длиной по меньшей мере 5, 10, 20 или 30 нуклеотидов);

(iii) присутствия, отсутствия и/или типа 5'-кэпа в РНК-матрице, например, содержат 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) присутствия, отсутствия и/или типа 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, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

(v) стабильности РНК-матрицы (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% РНК-матрицы интактными (например, длиной более 100, 125, 150, 175 или 200 нуклеотидов) после теста на стабильность;

(vi) активности РНК-матрицы в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей РНК-матрицу;

(vii) длины 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, имеют 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) присутствия, отсутствия и/или типа посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержат ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% полипептида, первого полипептида или второго полипептида, выбранную посттрансляционную модификацию;

(ix) присутствия, отсутствия и/или типа одной или нескольких искусственных, синтетических или неканонических аминокислот в полипептиде, первом полипептиде или втором полипептиде, например, содержат 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) стабильности полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остаются ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактными (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) активности полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, являются ли модифицированными по меньшей мере 1% сайтов-мишеней после анализа на активность системы, содержащей полипептид, первый полипептид или второй полипептид; или

(xii) присутствия, отсутствия и/или уровня одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от контаминации пирогенами, вирусами, грибами, бактериальными патогенами или белками клетки-хозяина.

102. 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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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;

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

103. Способ модификации сайта-мишени в геномной ДНК в клетке, при этом способ включает:

приведение клетки в контакт с системой, полипептидом, РНК-матрицей или ДНК, кодирующей их, по любому из предыдущих пунктов;

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

104. 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, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; 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;

за счет чего обеспечивается модификация сайта-мишени в геномной ДНК в клетке, где клетка является эуплоидной, не является иммортализованной, является частью ткани, является частью организма, является первичной клеткой, является неделящейся, является гаплоидной (например, клеткой зародышевой линии), является нераковой полиплоидной клеткой или получена от субъекта, у которого имеется генетическое заболевание.

105. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту:

(a) полипептида или нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит (i) домен обратной транскриптазы (RT), (ii) ДНК-связывающий домен (DBD) и (iii) эндонуклеазный домен, например, нисказный домен; где домен RT содержит последовательность из таблицы 30 или 44; и

(b) РНК-матрицы (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;

за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

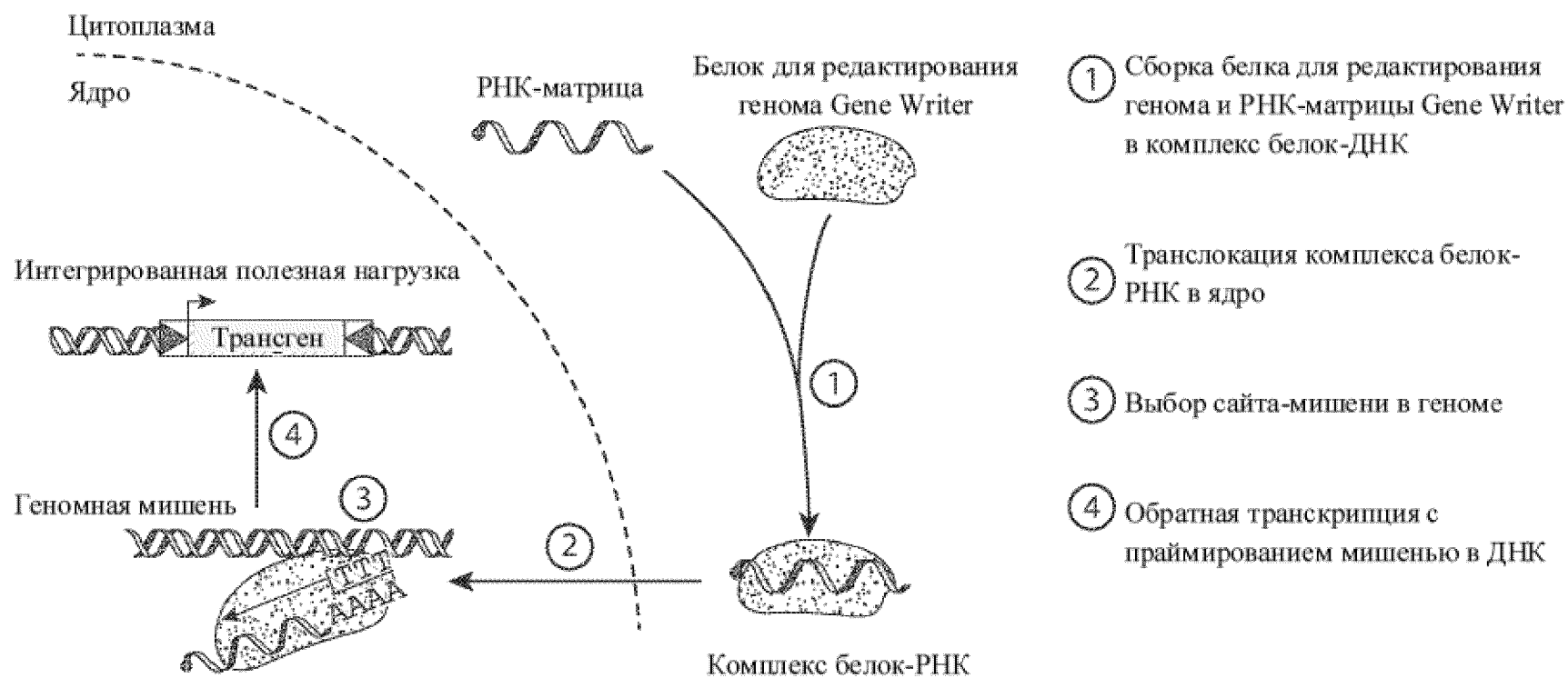
106. Способ лечения субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом, включающий:

введение субъекту системы, полипептида, РНК-матрицы или ДНК, кодирующей их, по любому из предыдущих пунктов;

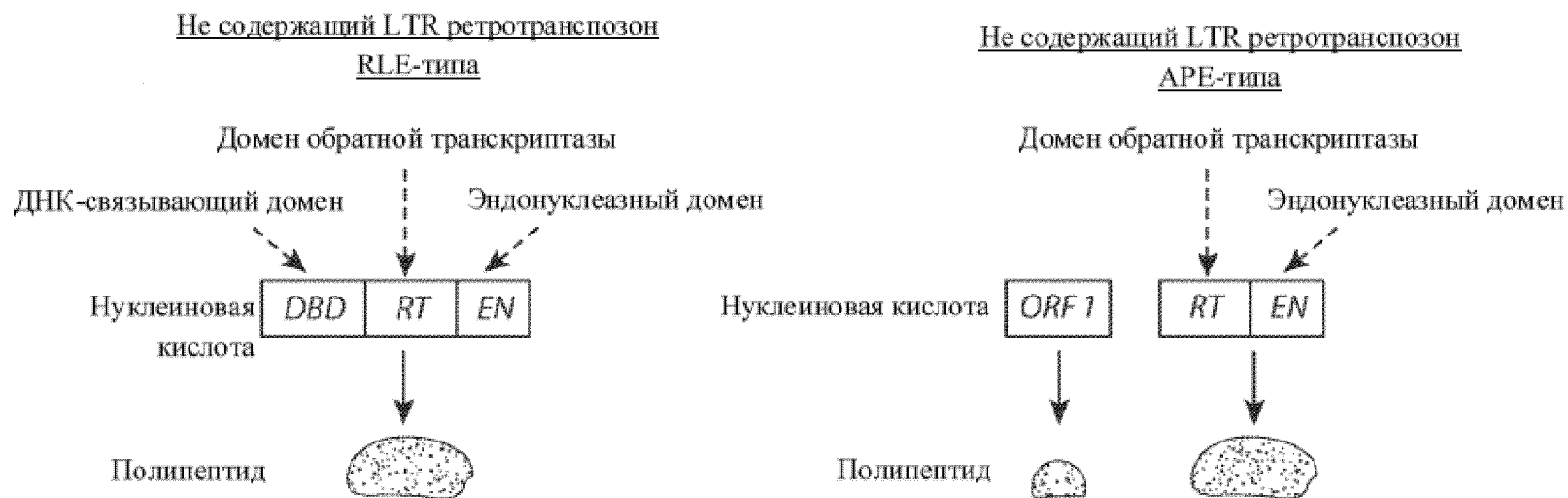
за счет чего обеспечивается лечение субъекта, у которого имеется заболевание или состояние, ассоциированное с генетическим дефектом.

107. Способ по любому из п. 105 или п. 106, где заболевание или состояние, ассоциированное с генетическим дефектом, представляет собой показание, указанное в любой из таблиц 9-12, и/или где генетический дефект представляет собой дефект гена, указанного в любой из таблиц 9-12.

108. Способ по любому из пп. 105-107, где субъект является пациентом-человеком.

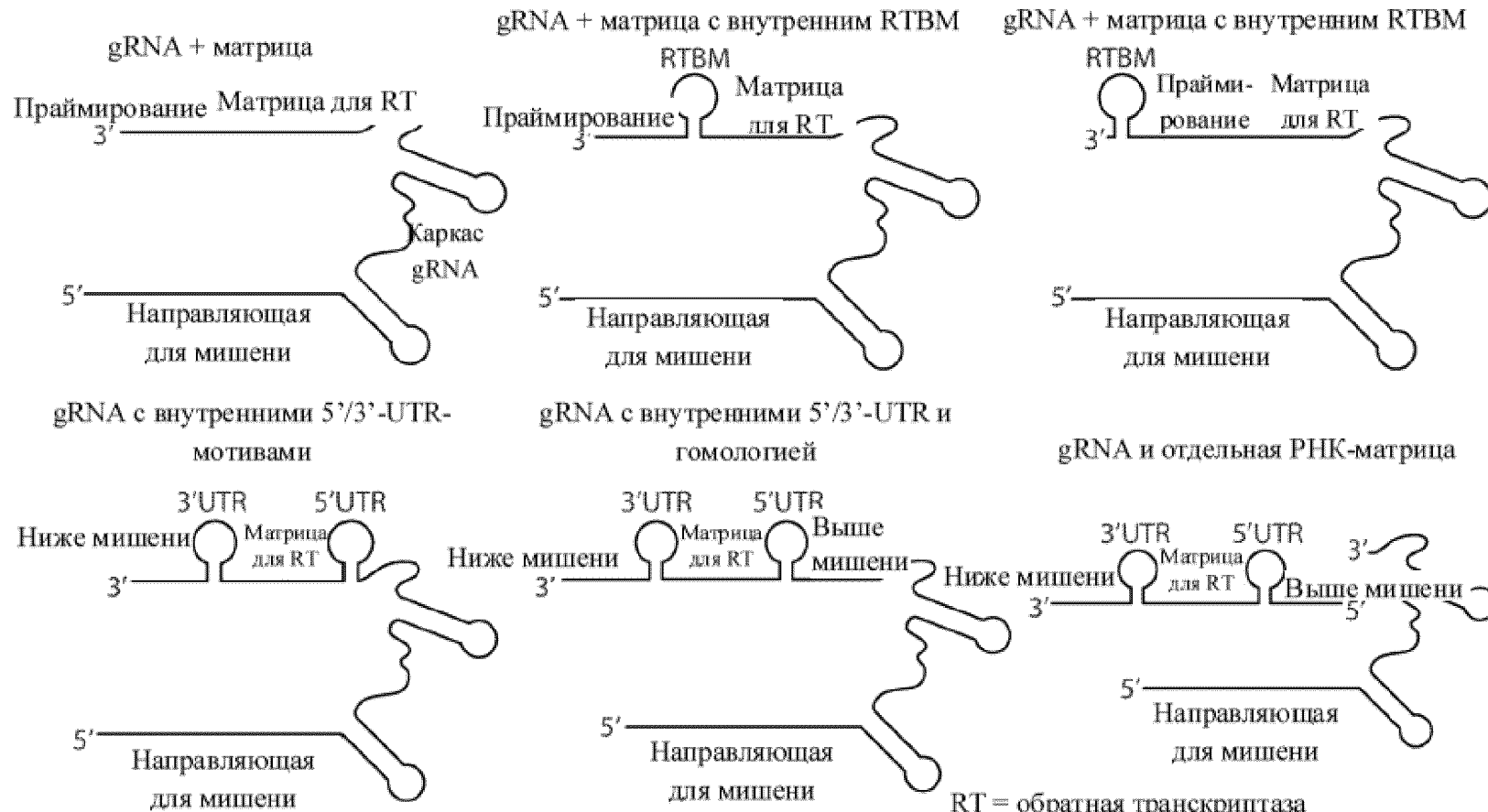


ФИГ. 1



ФИГ. 2

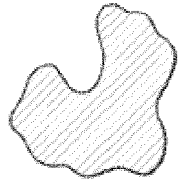
Примеры структуры РНК-матрицы Gene Writer



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



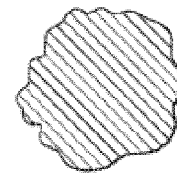
РНК-связывающий

Белок В-бокса

Белок оболочки

MS2

Белок dCas



Обратная транскриптаза

RTn

Другая RT

Пептид, содержащий домен RT



Эндонуклеаза

RTn

Cas9, никаза Cas 9

Рестрикционный фермент
(например, типа II, например,

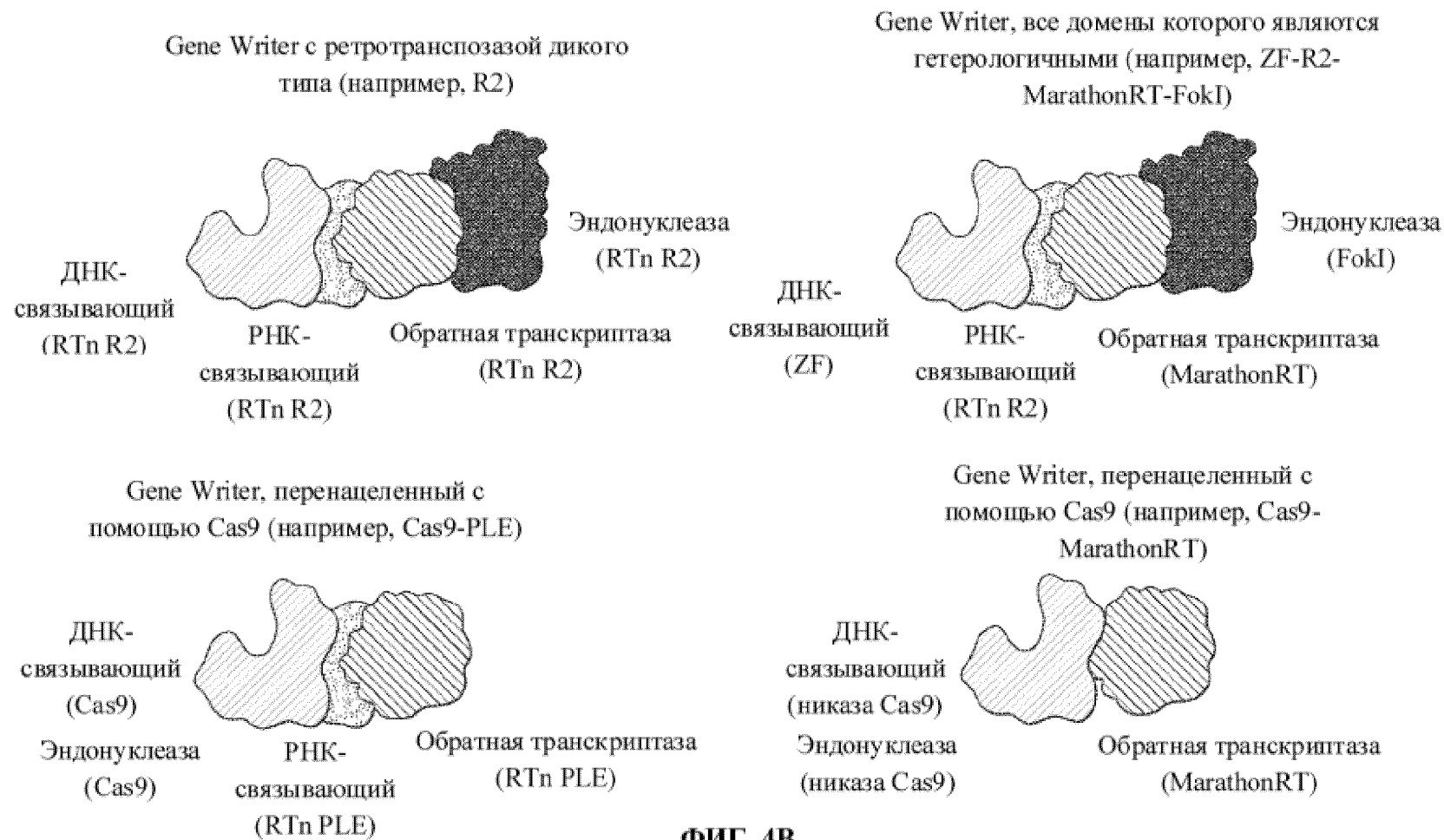
FokI)

Мегануклеаза

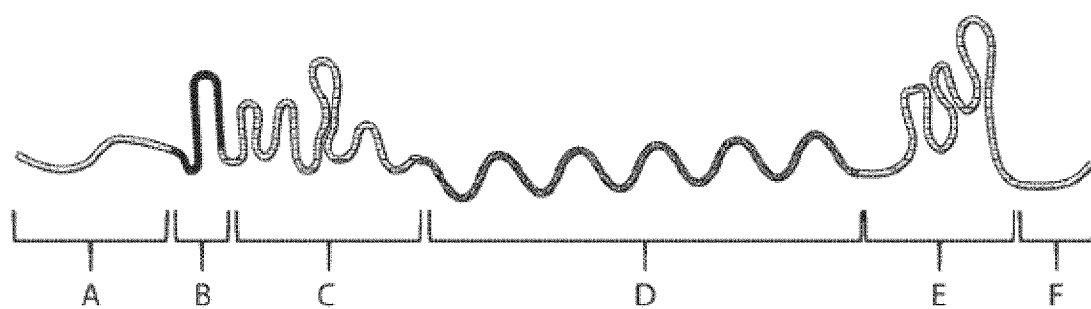
Резольваза для структур Холлидея

RLE

ФИГ. 4А



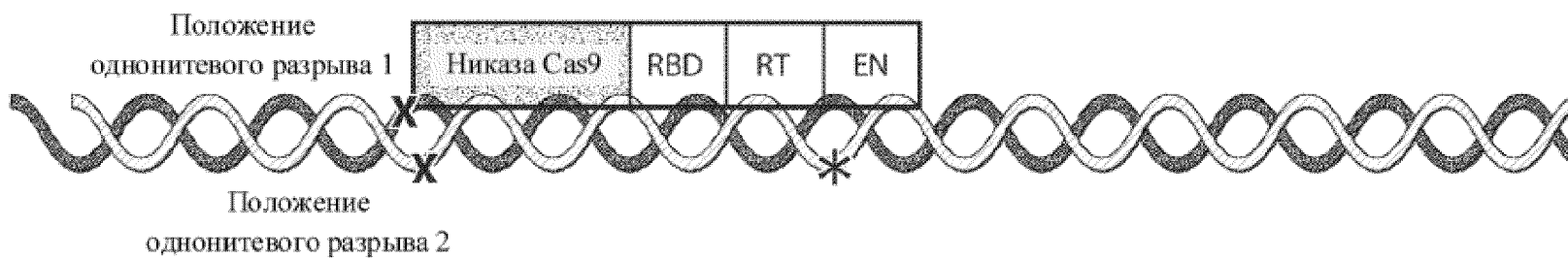
ФИГ. 4В



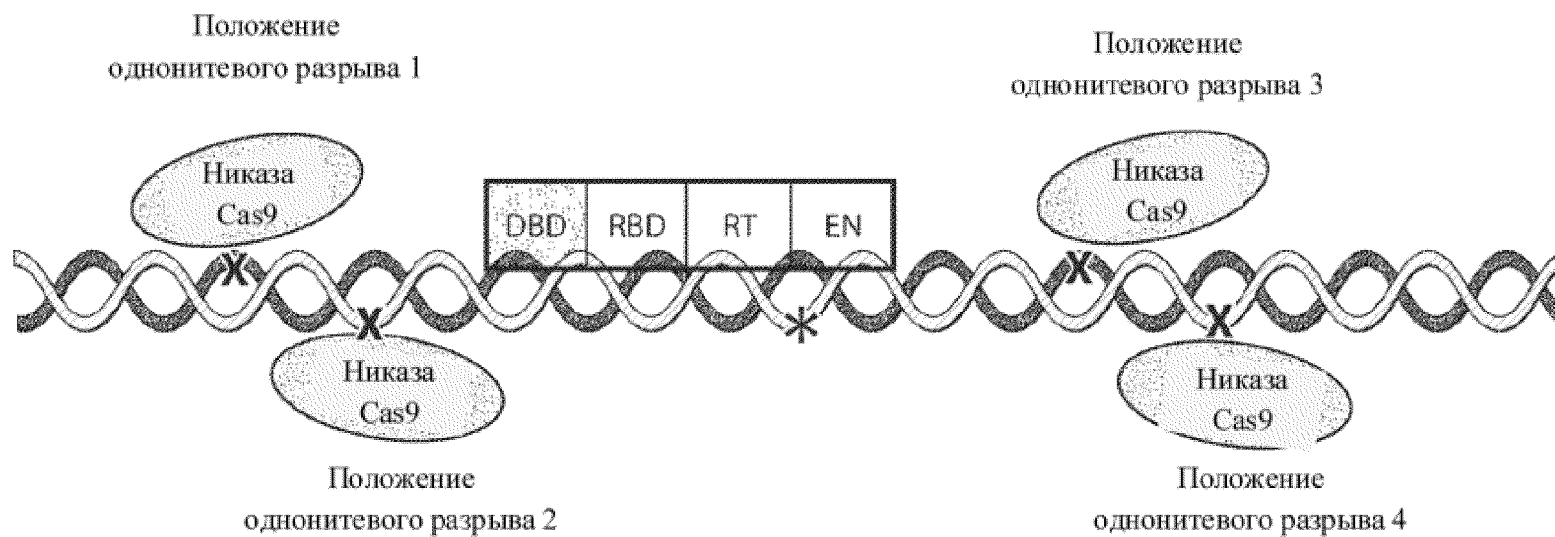
ФИГ. 5

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 односторонний 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 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 односторонний разрыв в ДНК-мишени

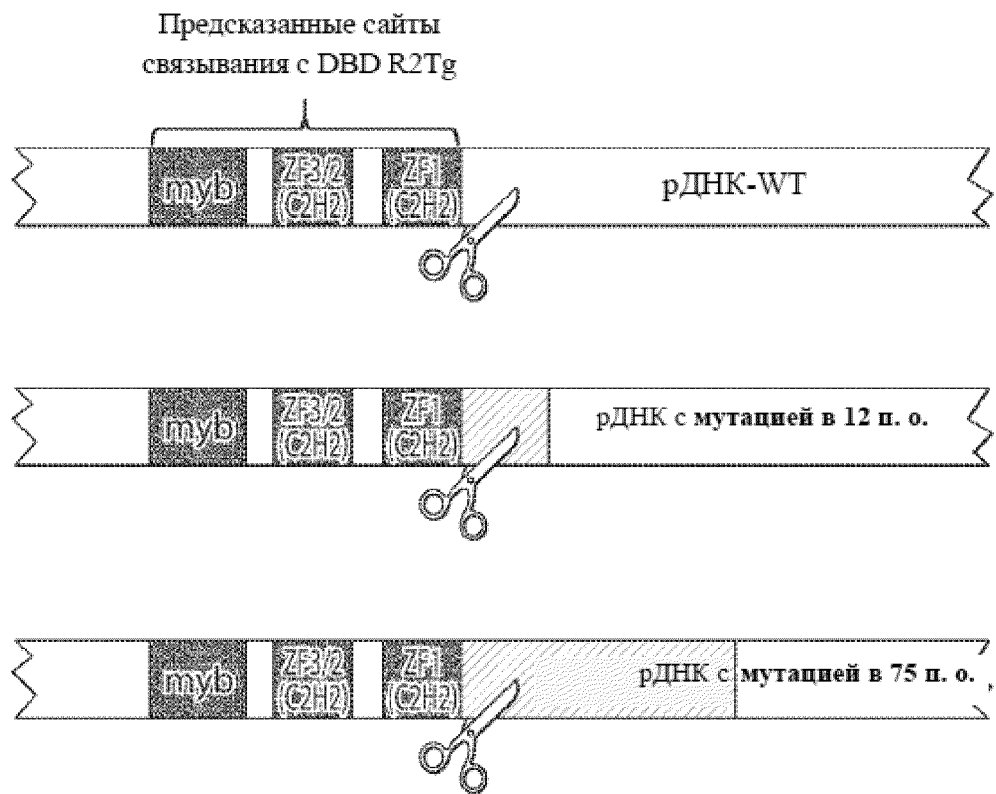
ФИГ. 6



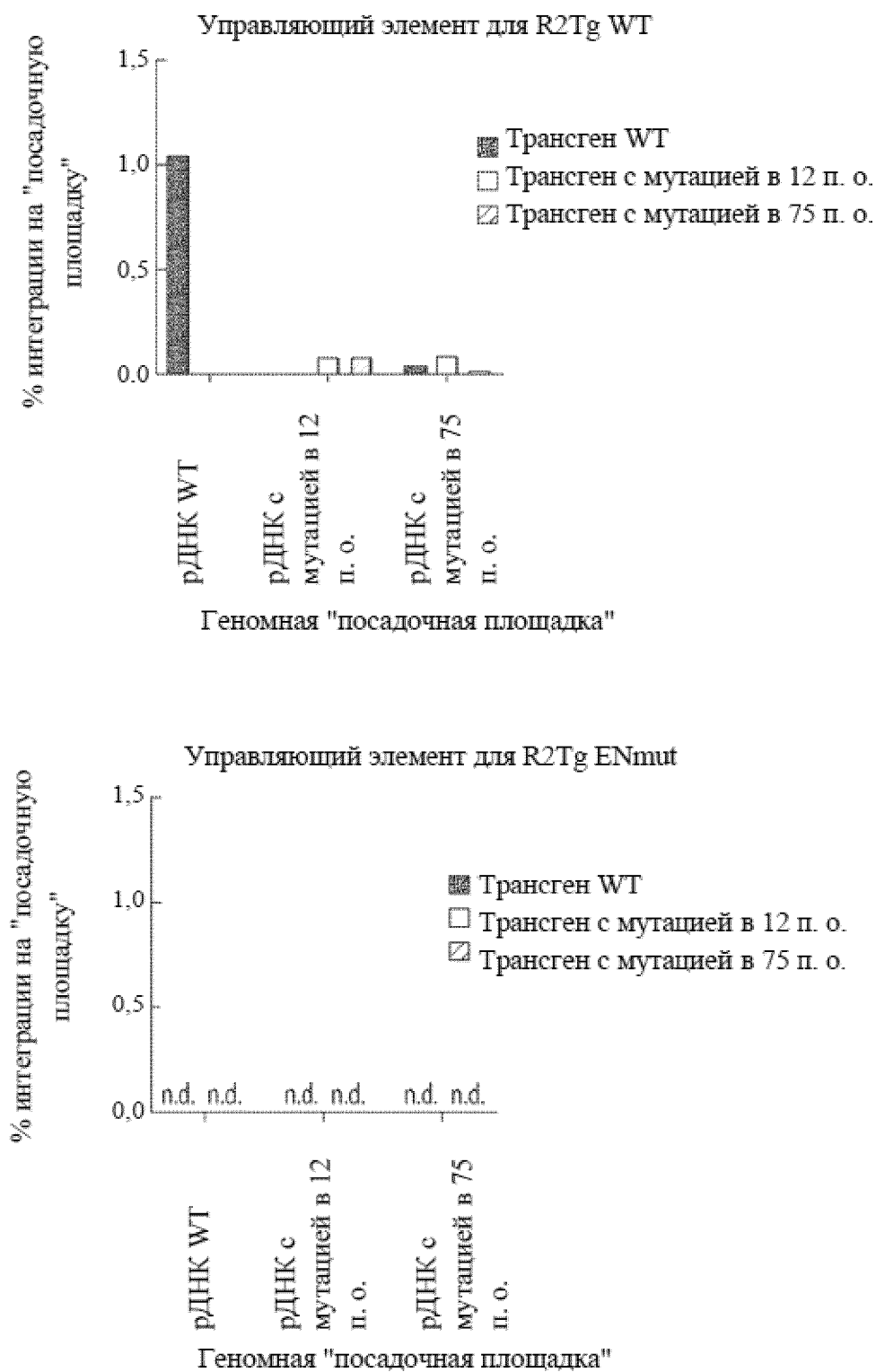
ФИГ. 7А



ФИГ. 7В

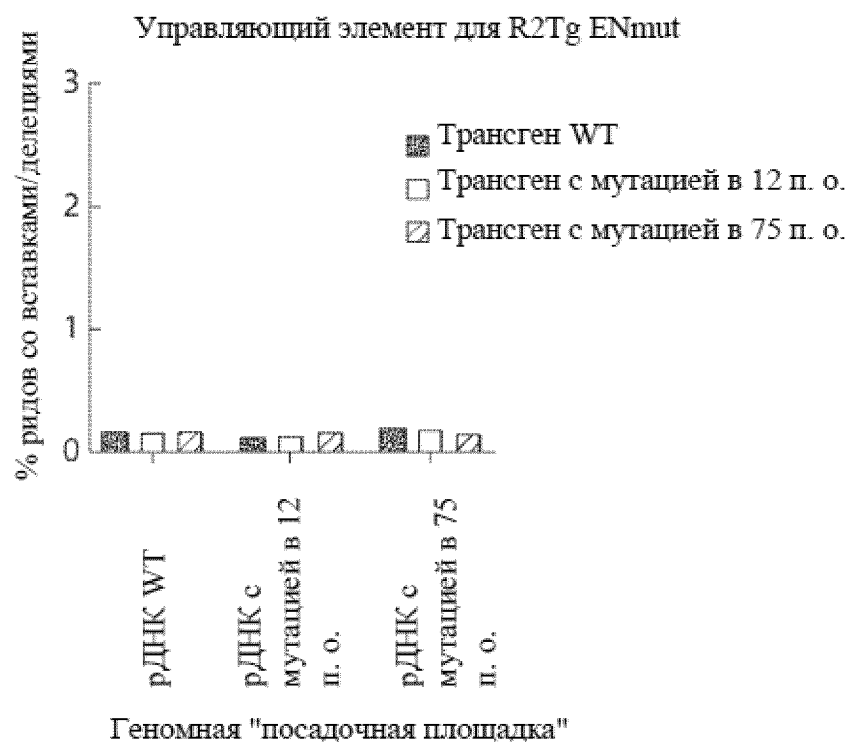
Фигуры, изображающие "посадочные площадки"**ФИГ. 9**

Интеграция в "посадочные площадки" с мутантным сайтом R2



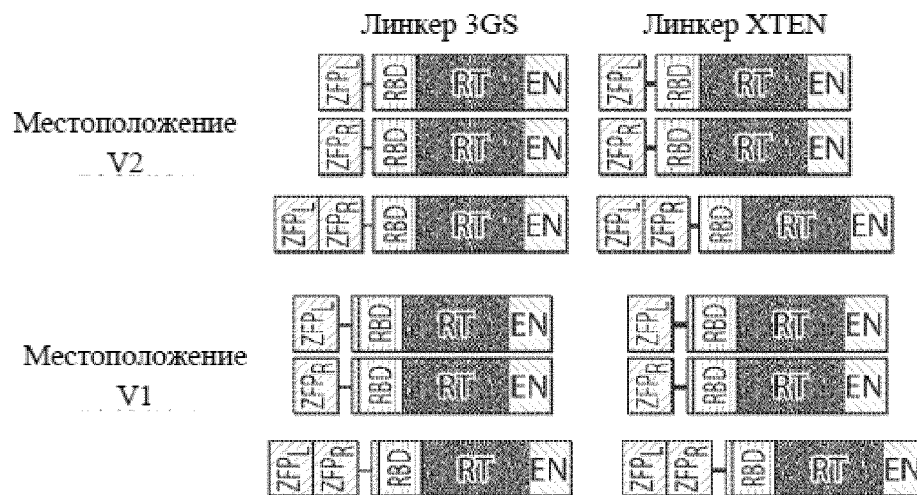
ФИГ. 10А

Вставки/делеции (внесение однонитевых разрывов) в "посадочных площадках" с мутантным сайтом R2



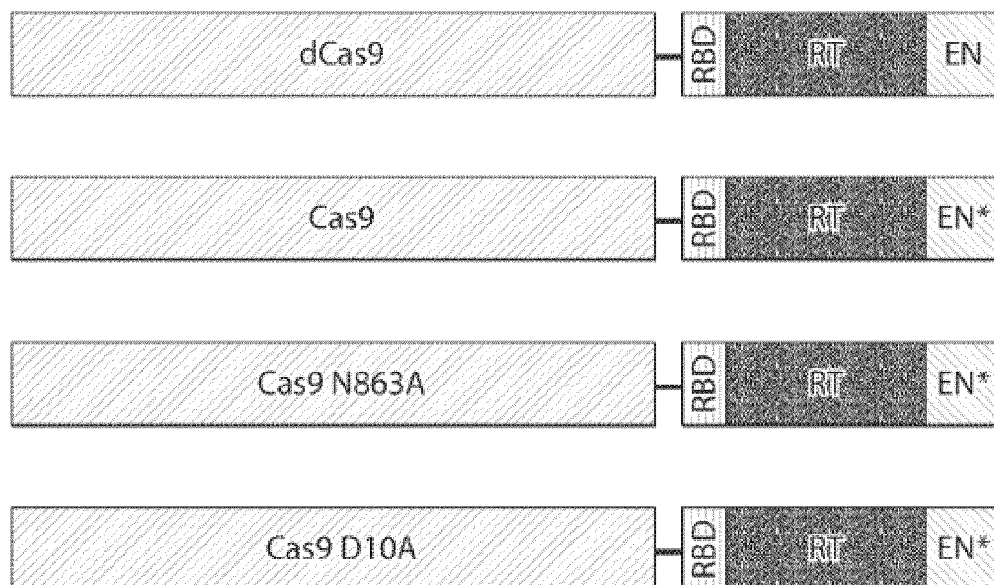
ФИГ. 10В

Примеры перенацеливания полипептида Gene Writer путем замены ДНК-связывающего домена



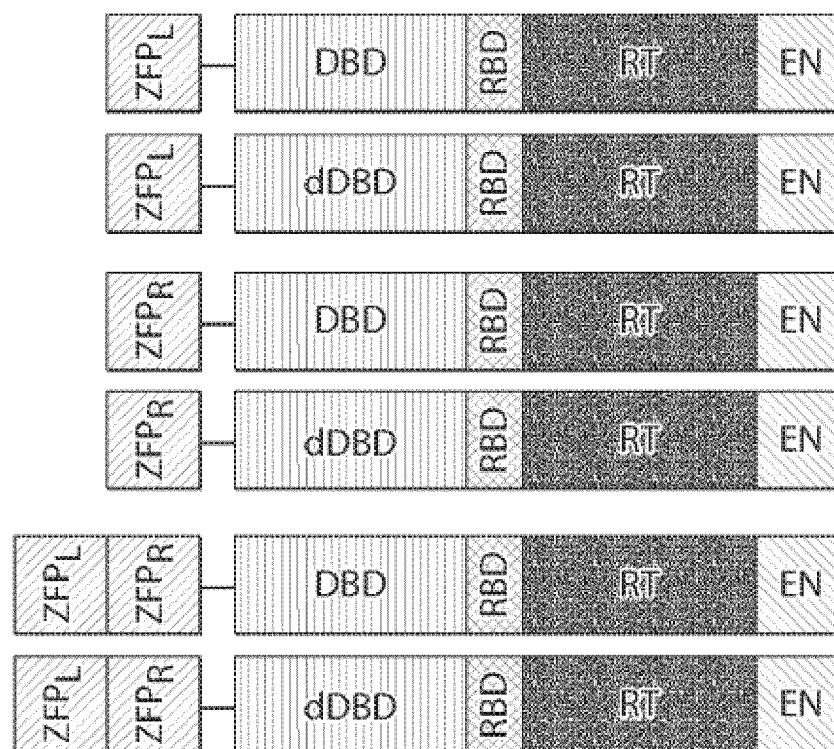
ФИГ. 11

Примеры перенацеливания полипептида Gene Writer путем замены ДНК-связывающего домена



ФИГ. 12

**Примеры перенацеливания полипептида Gene Writer путем
слияния с концом**



ФИГ. 13

Внесение одонитевого разрыва во вторую нить: слияние Cas9-Gene Writer с удлиненной на 3'-конце донорной gRNA

Слитые конструкции R2Tg-Cas9

Gene Writer Cas9-N863A-R2tg(RBD*, RT, EN)



Gene Writer Cas9-N840A-R2tg(RBD*, RT, EN)



Gene Writer Cas9-D10A-R2tg(RBD*, RT, EN)

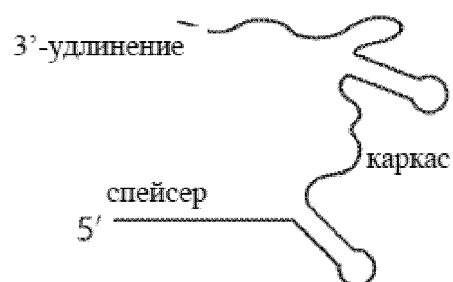


Gene Writer dCas9-R2tg(RBD*, RT, EN)



ФИГ. 14А

Донор: gRNA, удлиненная на 5'-конце



ФИГ. 14В

**Внесение одонитевого разрыва во вторую нить: продукты слияния
Cas9-Gene Writer с трансгеном, фланкированным UTR**

Gene Writer N863A-R2tg(RBD, RT, EN)



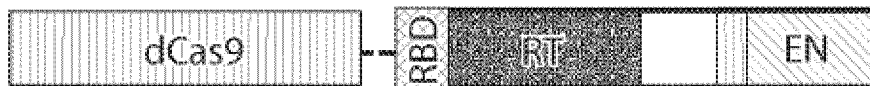
Gene Writer N840A-R2tg(RBD, RT, EN)



Gene Writer D10A-R2tg(RBD, RT, EN)



Gene Writer dCas9-R2tg(RBD, RT, EN)



ФИГ. 15А

Донор: UTR-трансген



ФИГ. 15В

Внесение однонитевого разрыва во вторую нить: нуклеаза Cas9

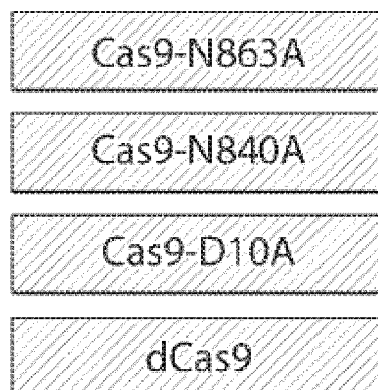


ФИГ. 16А

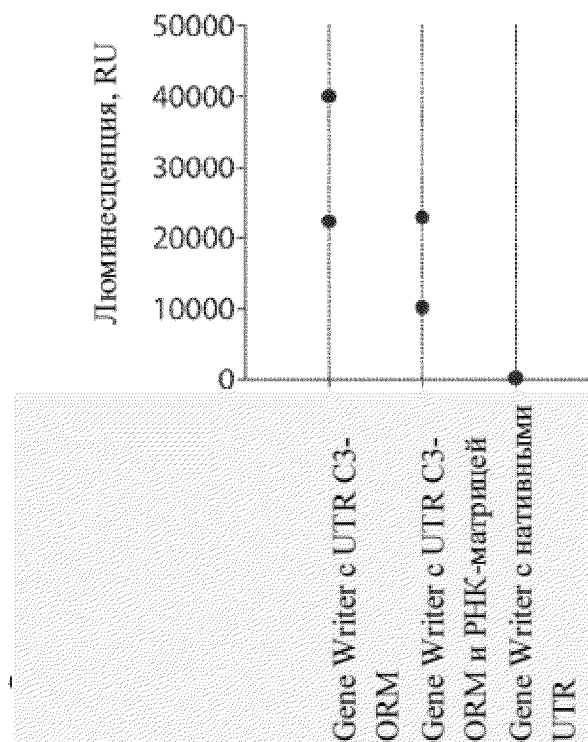
Донор: UTR-трансген



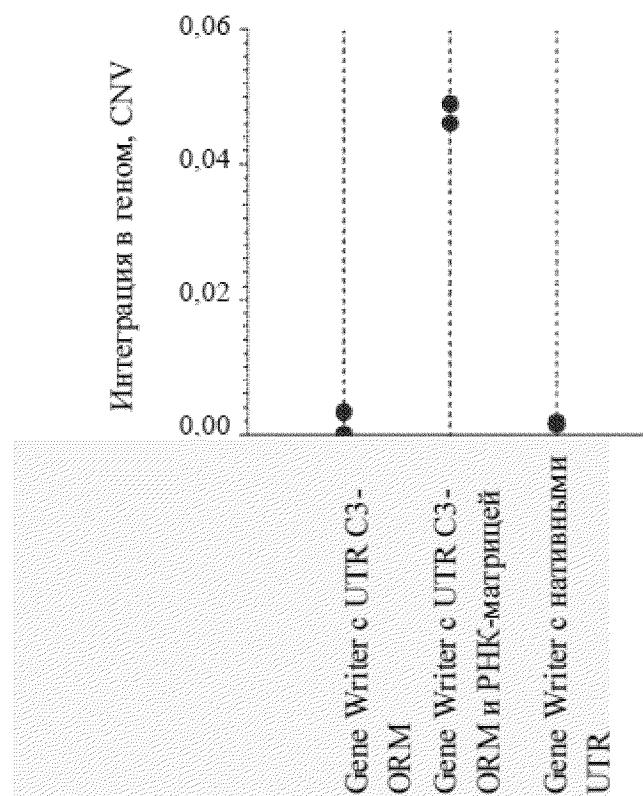
ФИГ. 16В



ФИГ. 16С

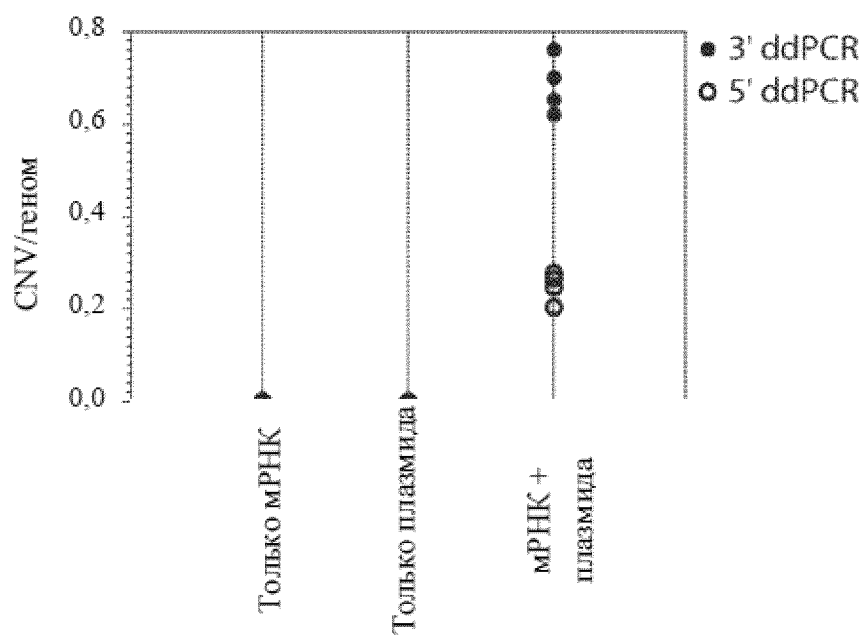
Улучшенная экспрессия полипептида Gene Writer**ФИГ. 17А****ФИГ. 17В**

Улучшенная экспрессия полипептида Gene Writer



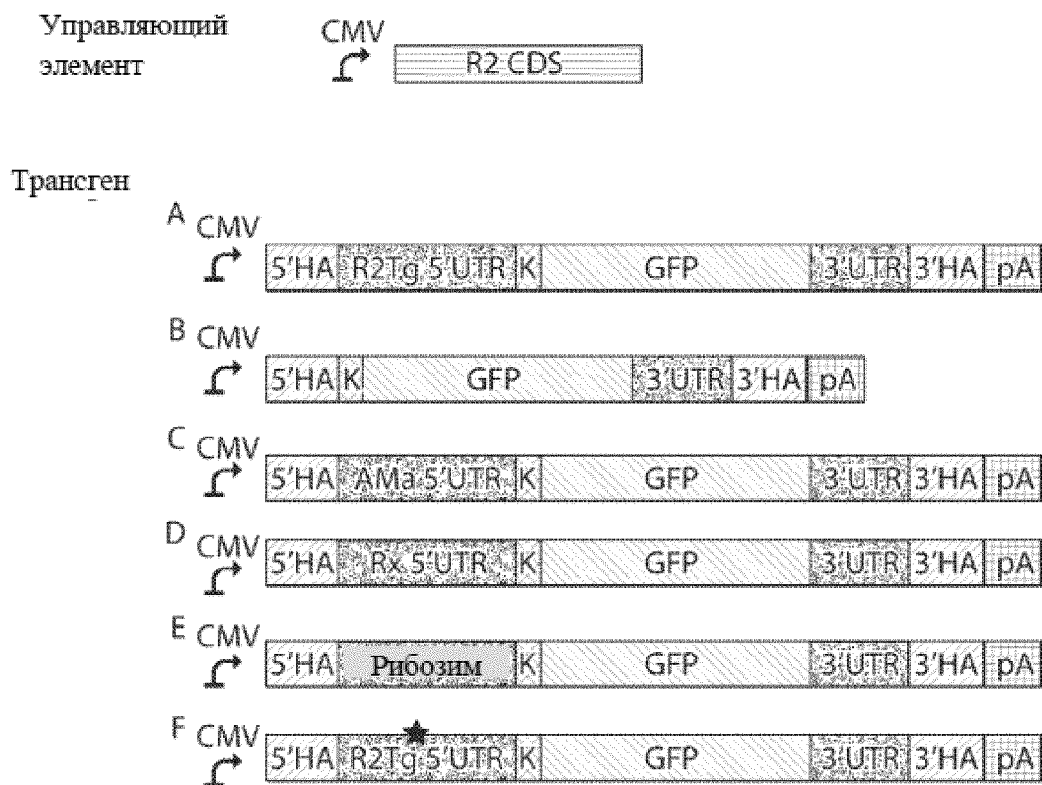
ФИГ. 18

Пример смешанной доставки с помощью РНК/ДНК



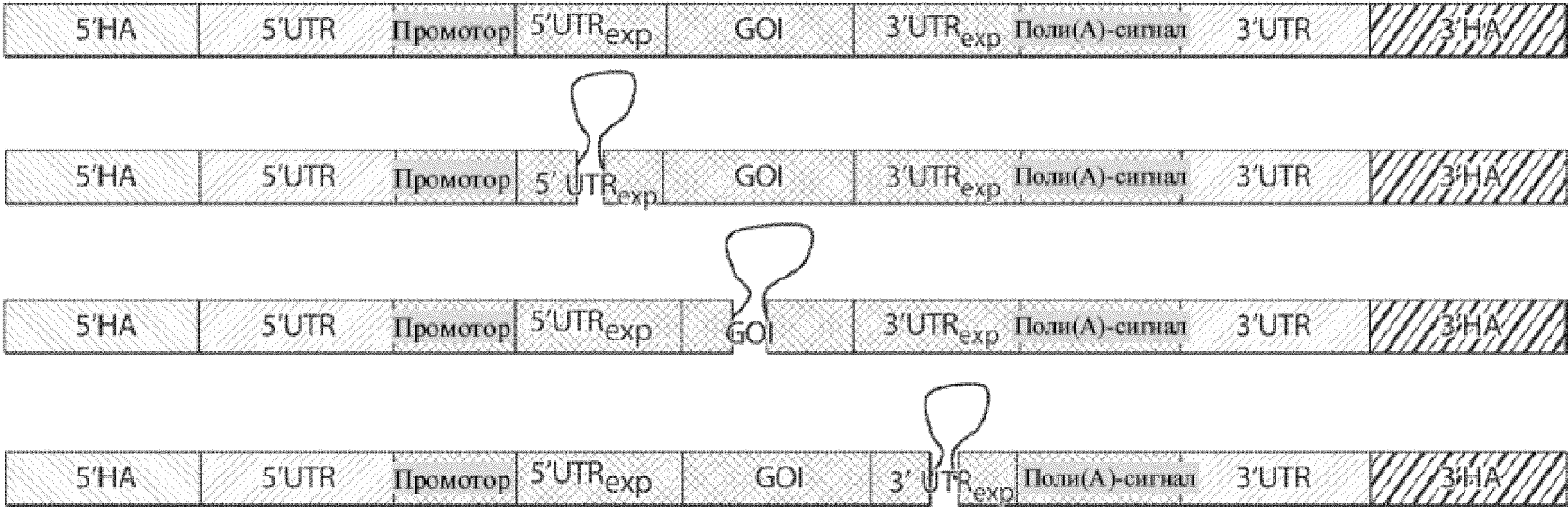
ФИГ. 19

Конструирование 5'-UTR

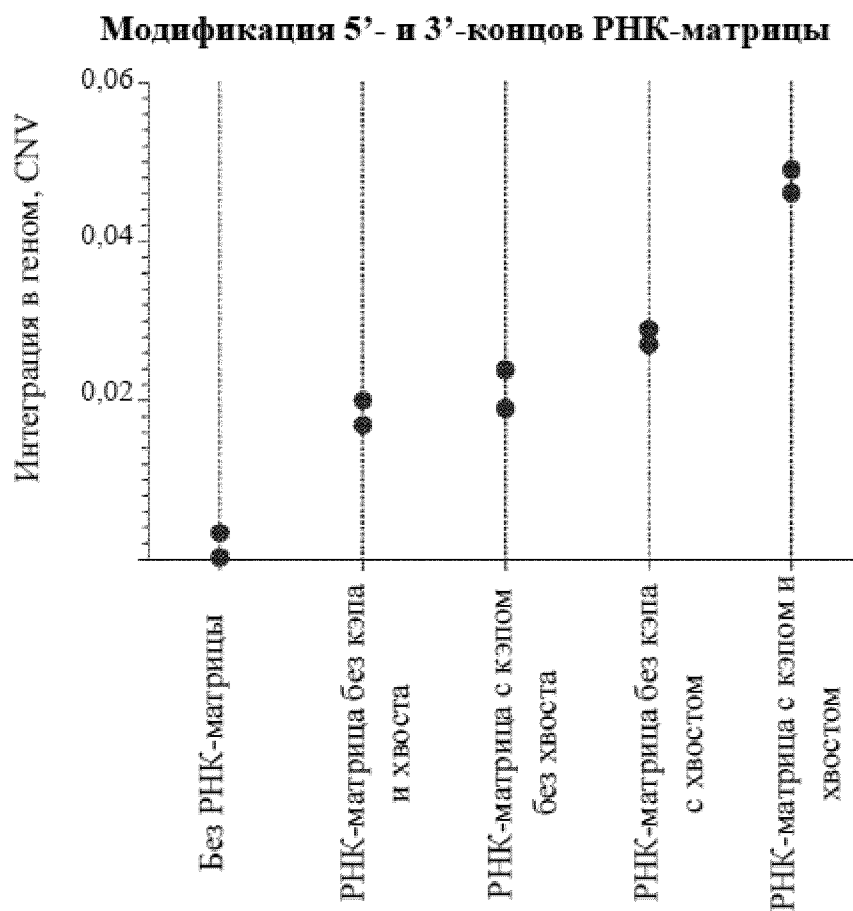


ФИГ. 20

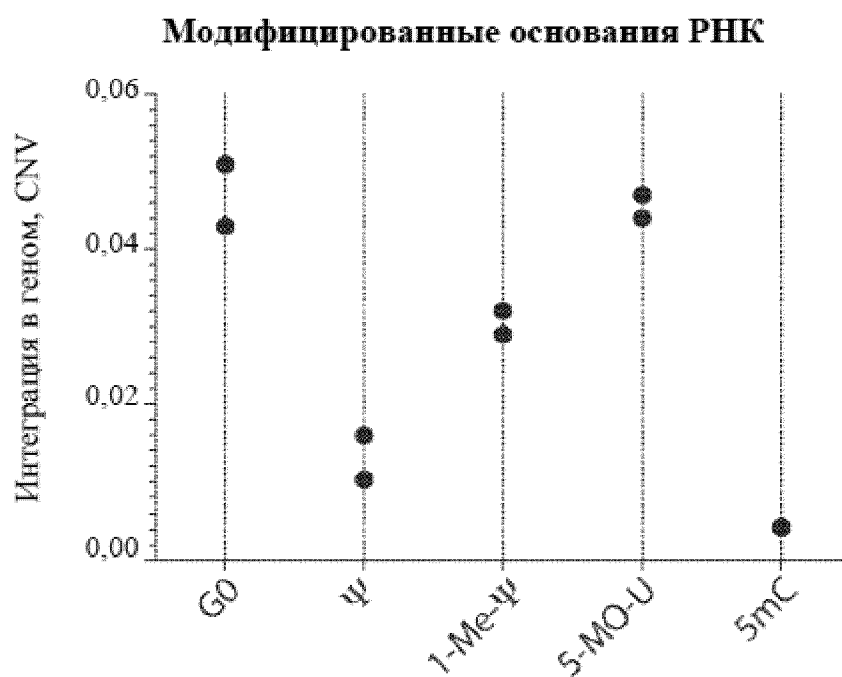
Интроны в РНК-матрице



ФИГ. 21

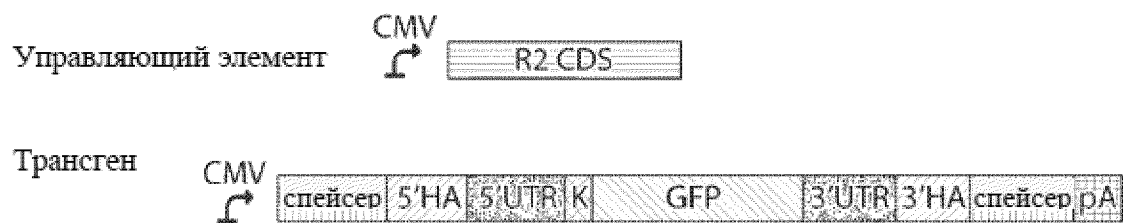


ФИГ. 22



ФИГ. 23

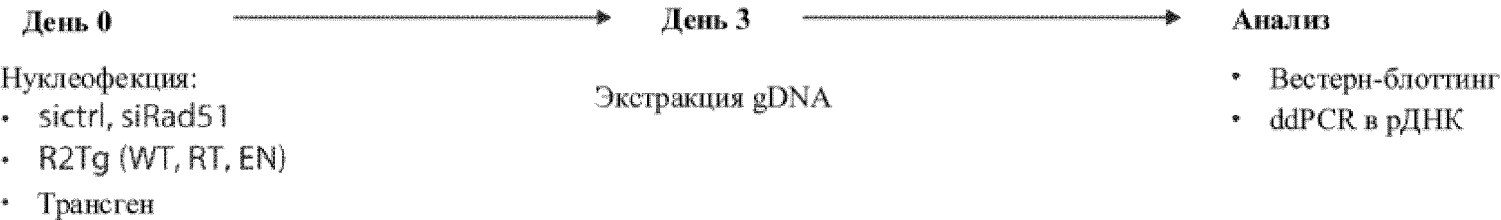
Управляющий элемент и трансген для экспериментов с НА



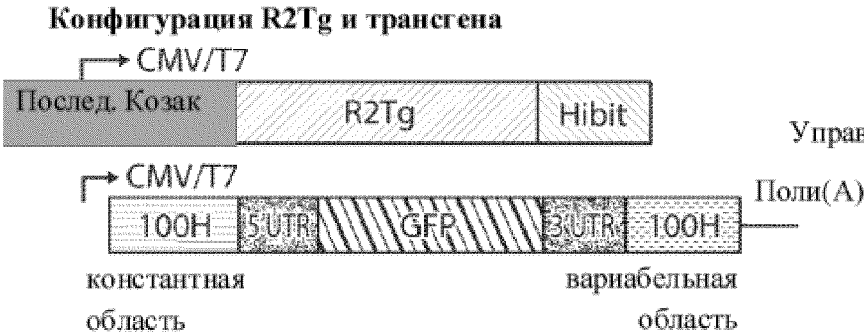
ФИГ. 24

Gene Writing при отсутствии Rad51

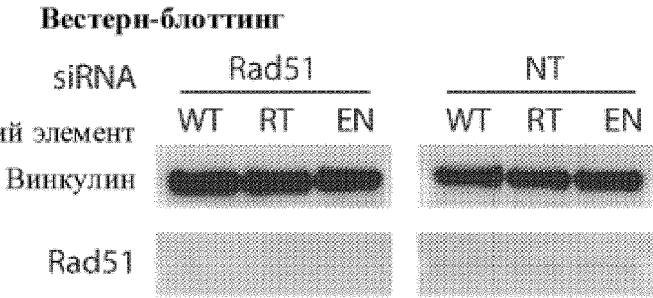
Схема эксперимента



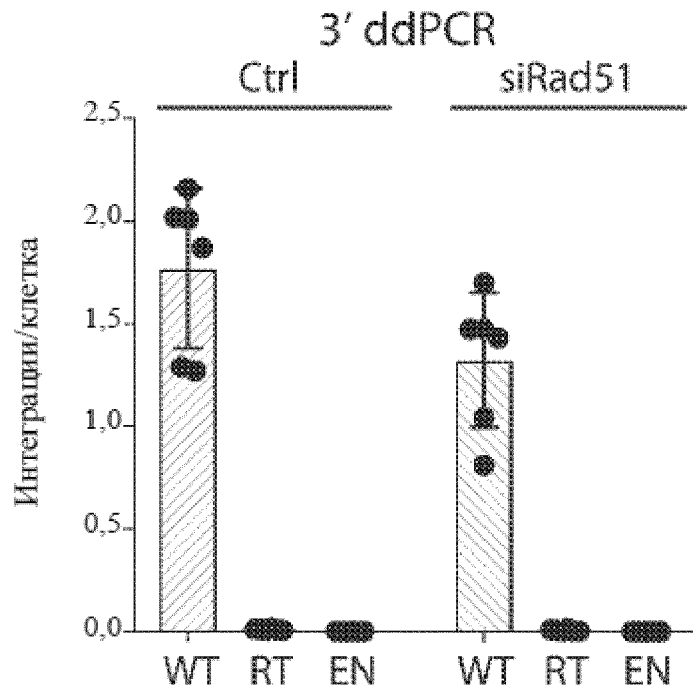
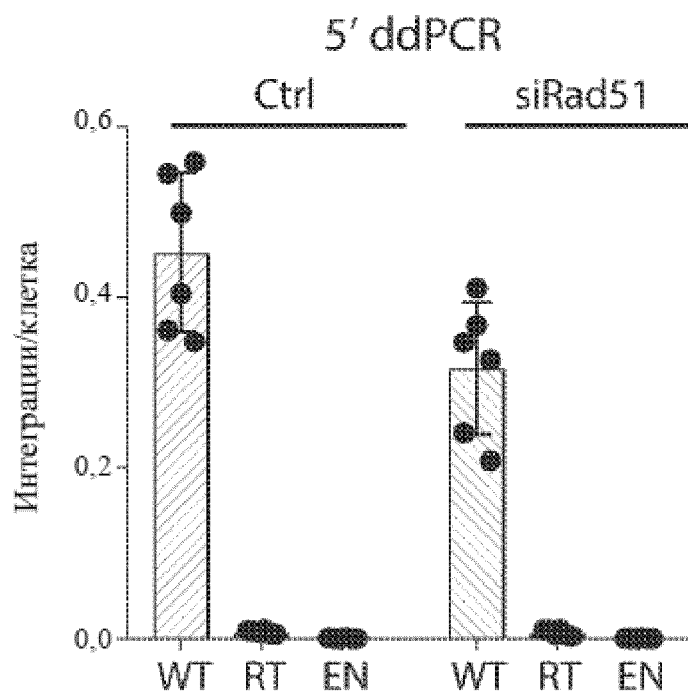
ФИГ. 25А

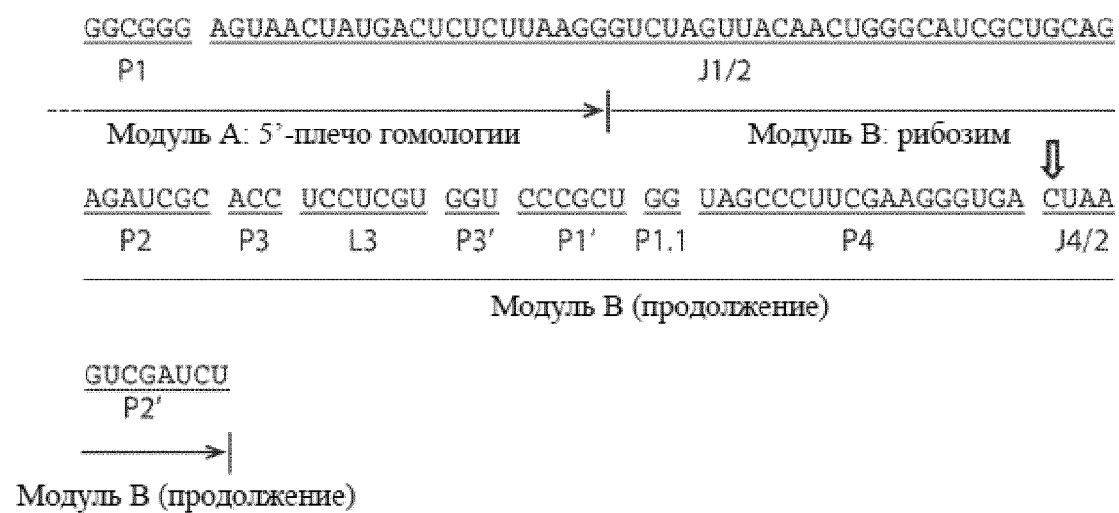


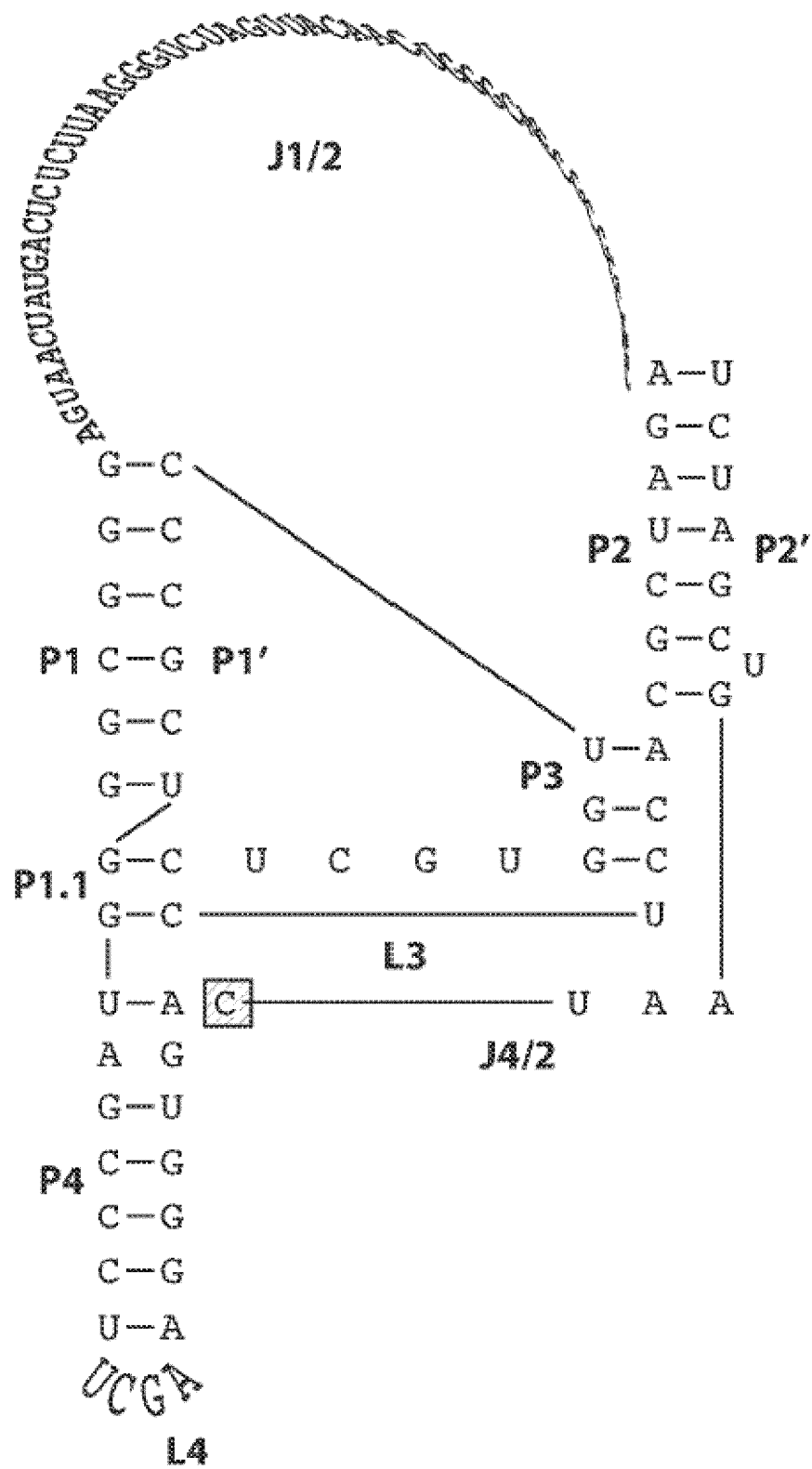
ФИГ. 25В



ФИГ. 25С

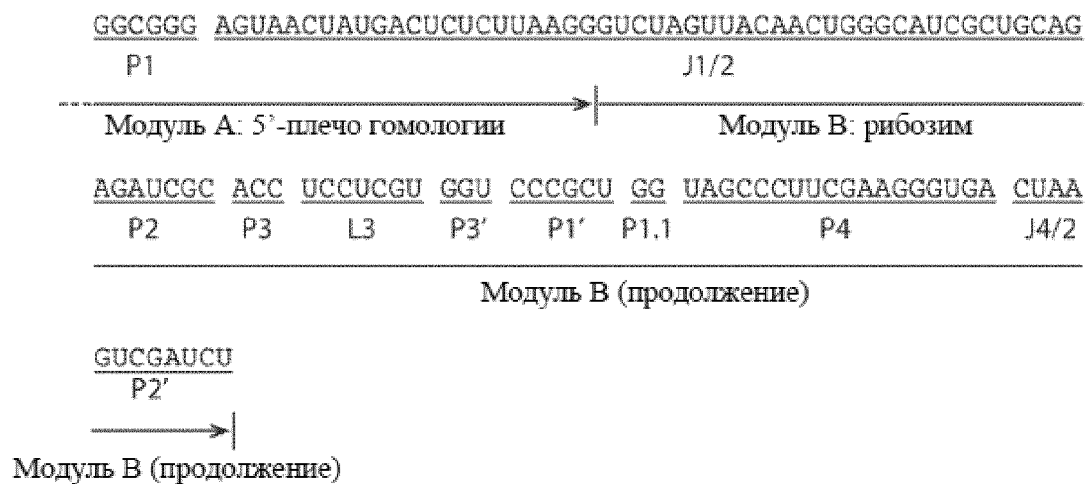
Gene Writing при отсутствии Rad51**ФИГ. 26А****ФИГ. 26В**

Рибозим R2Tg**ФИГ. 27А**



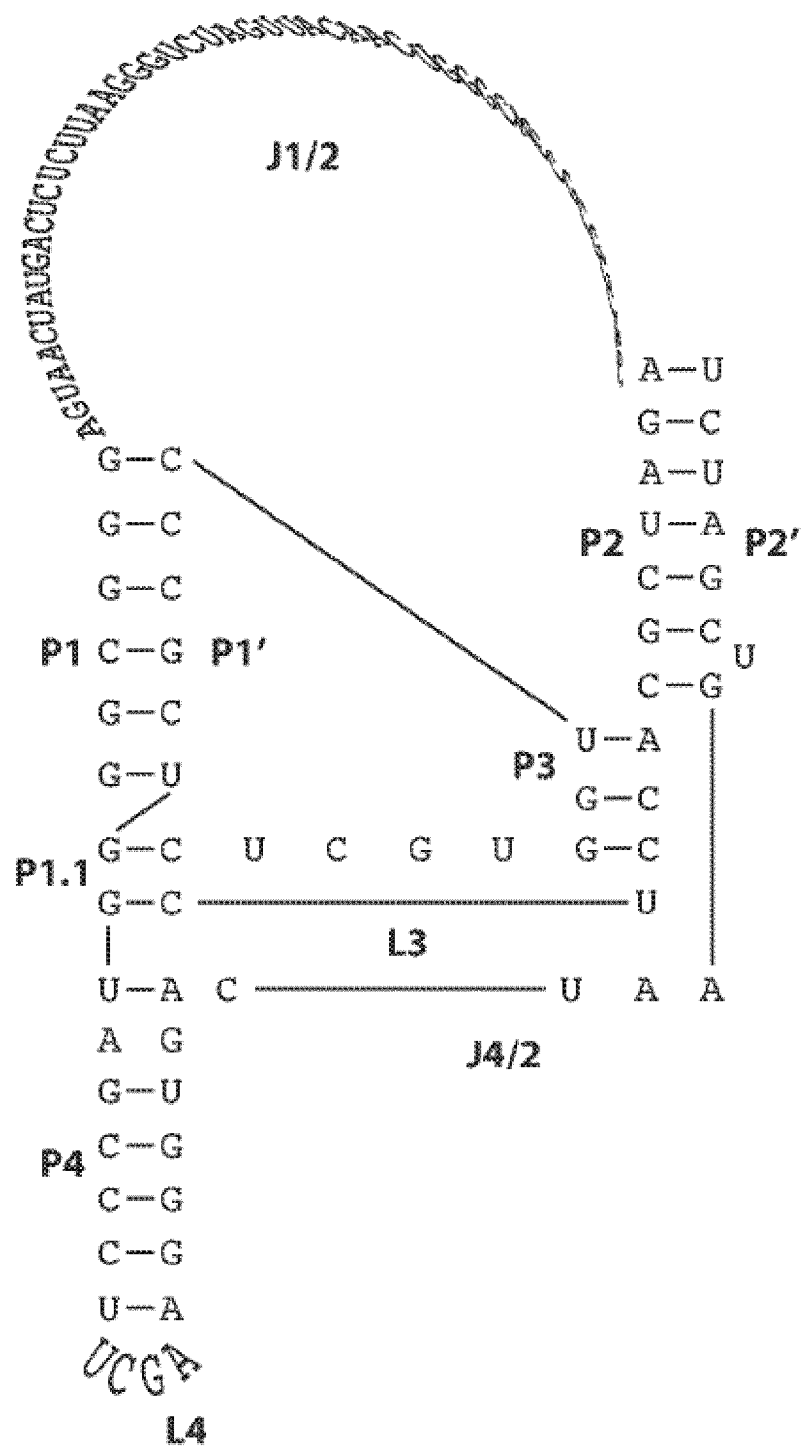
ФНГ. 27В

Последовательность рибозима R2Tg



ФИГ. 28

Предсказанная вторичная структура рибозима R2Tg



ФИГ. 29

5' -XXXXXXXXCATCATCATCATCATCATCATCATCATCATXXXXXXXX-3'
3' -YYYYYYYYGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAYYYYYYYY-5'

5' -XXXXXXXXCATCATCATCATCATCATCATCATCATCATXXXXXXX-3'
3' -YYYYYYYGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTA-3'

РНК-матрица

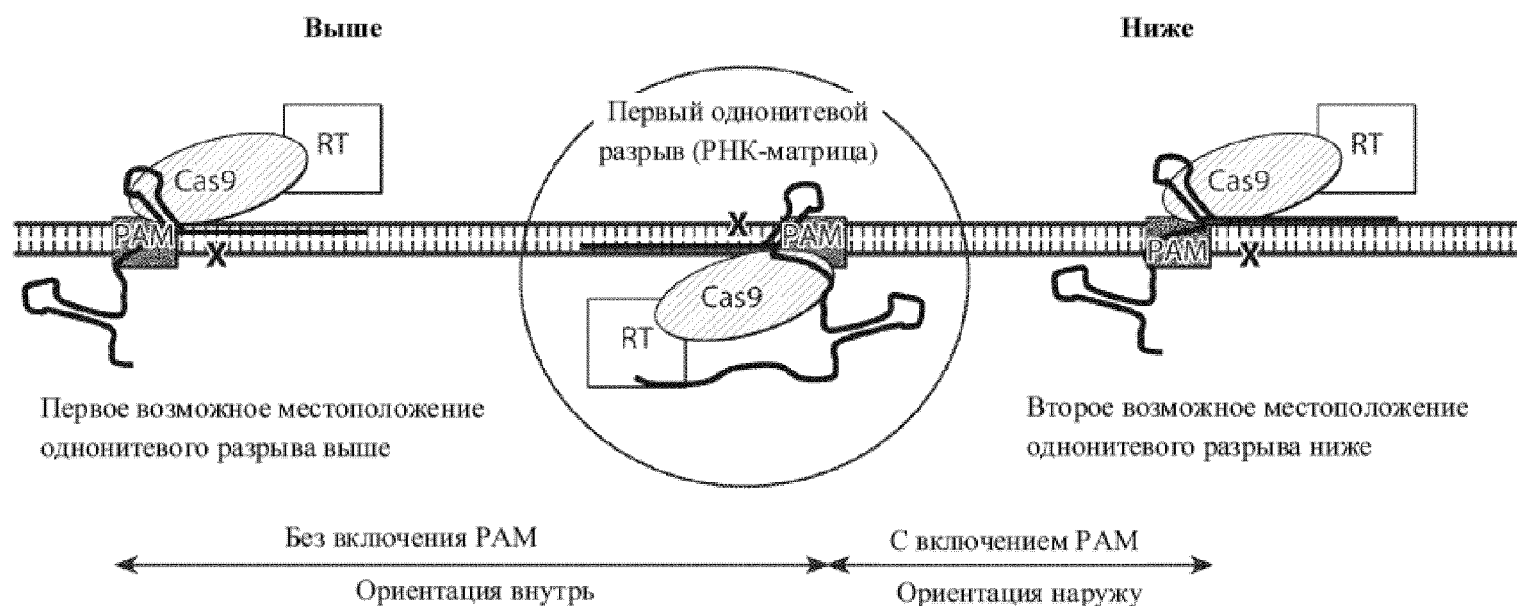
AGTAGTAGTAGTAGTAGTAGTA
←TPRT
5' -XXXXXXXXCATCATCATCATXXX-3'
GTAYYYYYYY-5'

[illegible]

5'-XXXXXXXXCATCATCATCATXXXXXXXX-3'
3'-YYYYYYYGTAGTAGTAGTAYYYYYYYY-5'

ФИГ. 30

**Ориентации "с включением PAM" и "без включения PAM" для
внесения второго одностороннего разрыва**

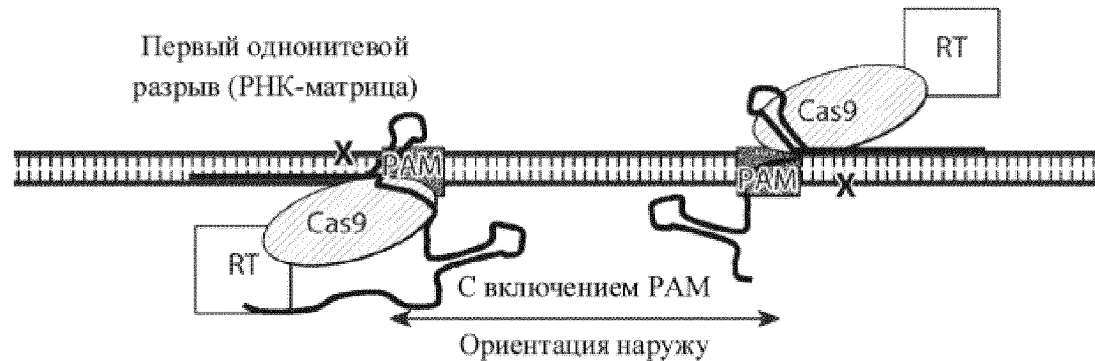


РНК-матрица направлена на локус НЕК для введения вставки СТТ, опосредованной полипептидом Gene Writer. Первый односторонний разрыв указан красным X. Для усиления внедрения желаемого редактирования второй односторонний разрыв может быть размещен на противоположной нити от первого одностороннего разрыва. Второй односторонний разрыв может быть размещен выше или ниже первого одностороннего разрыва. В этом примере ориентация PAM представляет собой "без включения PAM", если нацеливание на второй односторонний разрыв происходит выше, или "с включением PAM", если нацеливание на второй односторонний разрыв происходит ниже относительно первого одностороннего разрыва.

ФИГ. 31

Оптимизация ориентации PAM и расстояния между одонитевыми разрывами для снижения частоты нежелательных результатов редактирования

Ориентация "с включением PAM" приводит к снижению частоты нежелательных результатов редактирования, обусловленных появлением двойных одонитевых разрывов

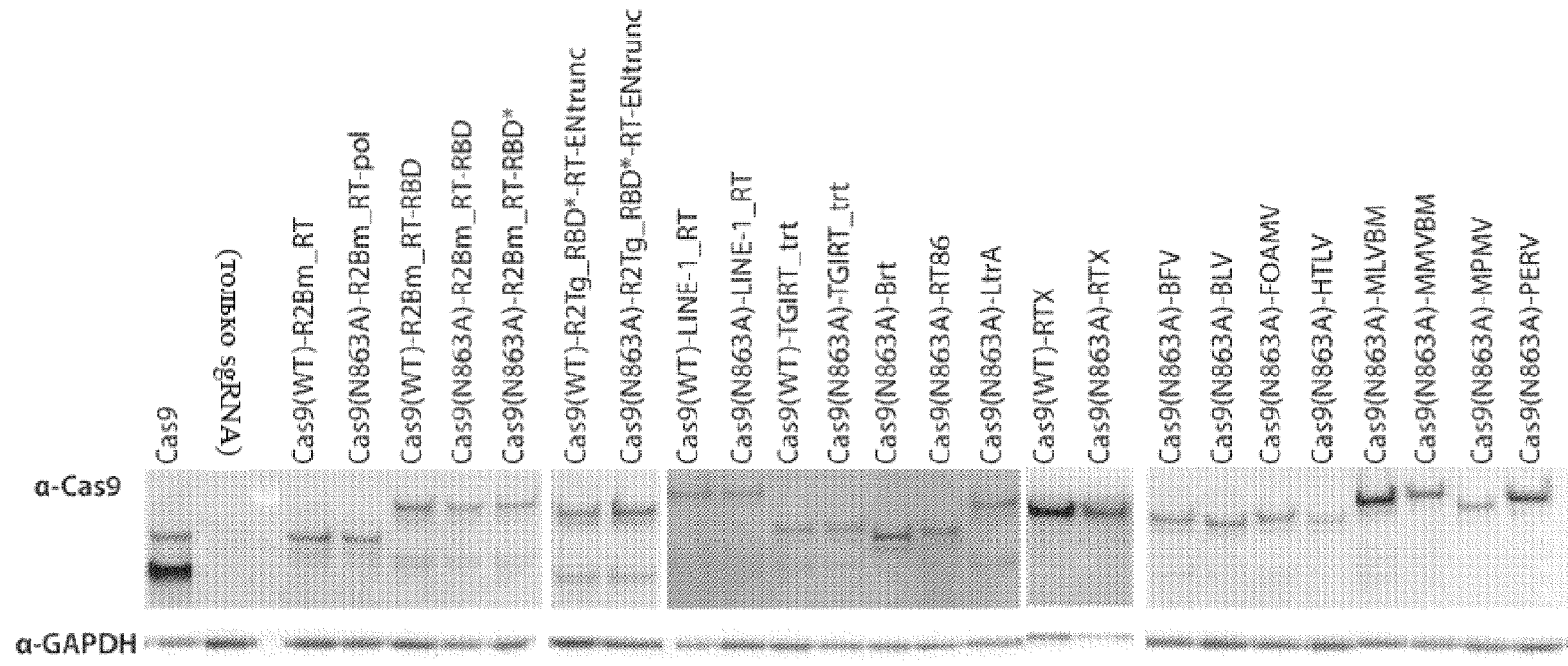


ФИГ. 32А

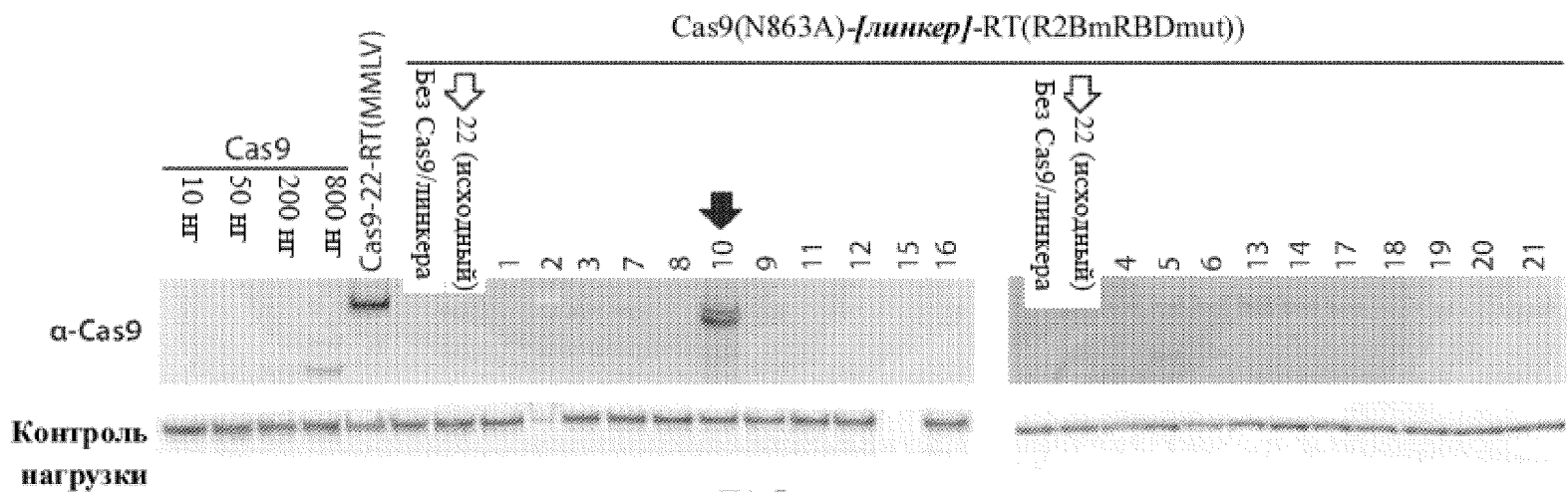
Ориентация "без включения PAM" приводит к снижению частоты нежелательных результатов редактирования, если два одонитевых разрыва разделены увеличенным расстоянием



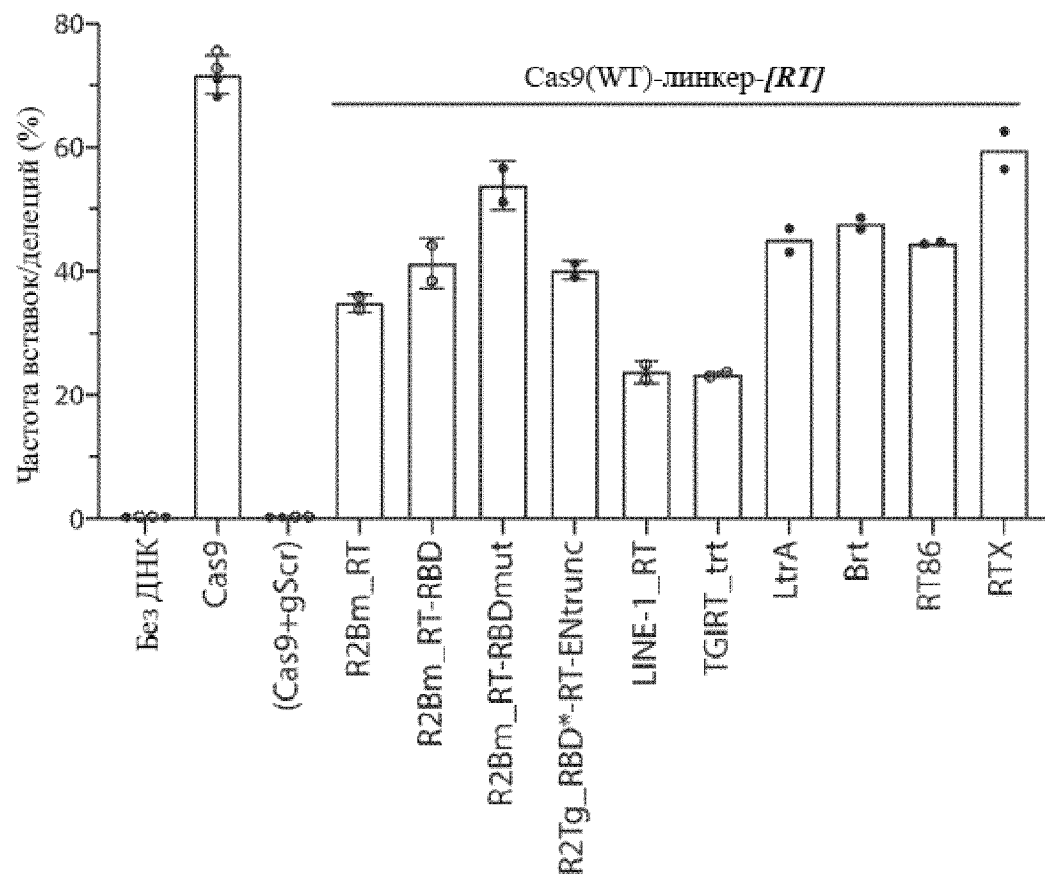
ФИГ. 32В



ФИГ. 33



ФИГ. 34



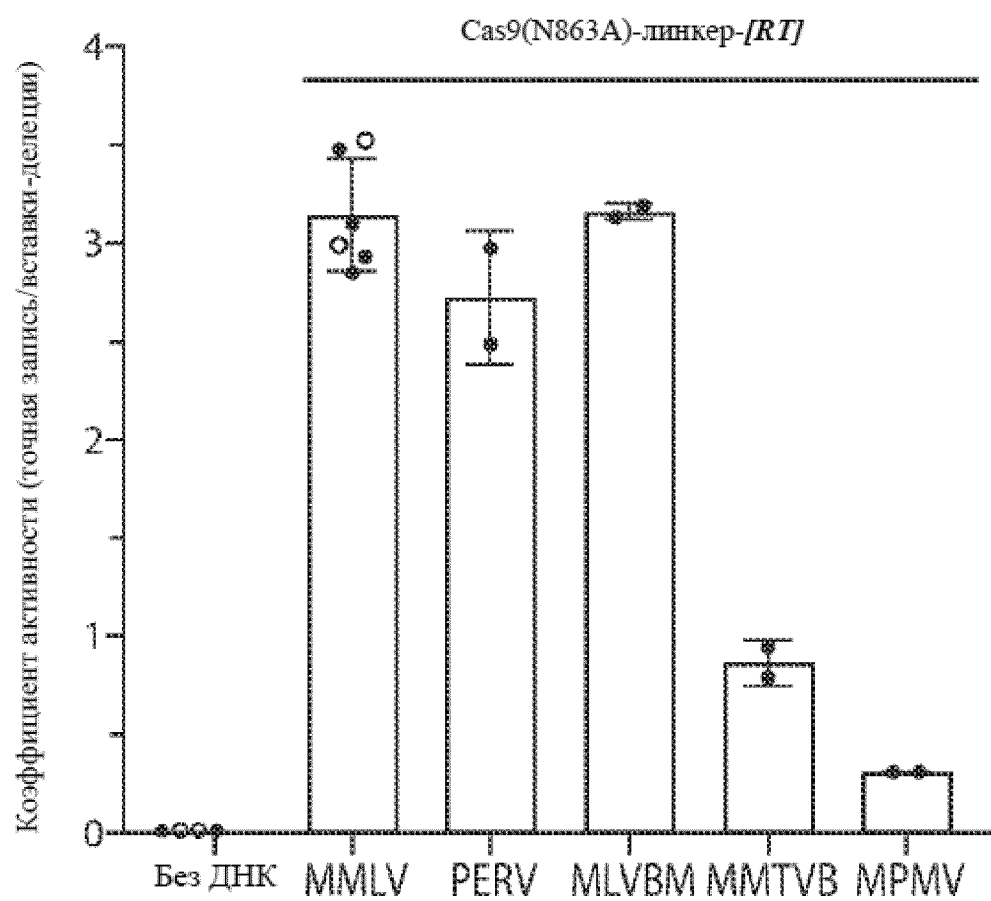
ФИГ. 35

	D200N	T330P	L603W
Консенсусная	WTVLPQGFKNSTLFDALARDLADFRIGHP	WIPGFAELLXPLYPLLKGTDP--KISLXPE	FATAEXLG-----G-LTSAGKXIRNKXELL
Гамма	..R.....H..... ..R.....H..... ..R.....N..QG..LD.. ..R.....I.....H..N.. ..S.....H.....	MAA....T..G.....LPNWC.DMAA....T..G.....LPNWC.DAQ....AATF..GN....LVNGEKT..AA....T..EKGE---F..WA..MAA....T..G.....LP..G..D	IH.EIYRRR.L...E..E...D...VH.EIYRRR.L...E..RE...N...I.VH.MIYRER.L...AG..A...AP...VH.AIKQOR.L...RE...E...S...IH.EIYRRR.L...E..RE...S...
Эпсилон	IH....SO..YOS.HKIKPKIS	E.SIESKF.EKO..ND..A--EPFO.DDO	YGVV.DY.HLMMHR.FV....TP....HK...F
Дельта	..R.....I...A..ER..OEP.ROVSAAFS ..R.....EMO..HI..QPI.QAF.. ..K.....EMO..HI..QPI.QAF..EQO..AV..NPM.KMF..	GT.TTRRP.QL..SS...ID..RAI.H.S.. GT.TLRQP.HS..CA.QRH...RDO.Y.N.S GT.TLRQP.HS..CA.QRH...RDO.Y.N.S GT.TLRKH.OS..SA.H.YR..RAC.T.T.O	YSLIRT.....VLG.WLPDPVPSYA YHYLRT.A-----LGTFOGRSSQAPFO YHYLRT.A-----LGTFOGRSSQAPFO IKYL.S.A-----IG.FLGTSAHOTLO
Бета	..K....MA....CKIV.TAIHKV.HAWK ..K....M....CKFVDKAILTV.DKYQ ..K....MA....CKIV.AAI.FV.A.W..	YLKLTIGD.K..FDT...DS..NSHR..SK FLKLTIGE.K..FEI..N..DSN..ISTRK.T.. YLKLTISA.V..NMI...DPN.LSVRA.T.. F..NYS..VK...TIVPANG--KF..WTED F..N....VQ...N..IPSAKG--KY.EWSE.	AHSIPL.E-----TVAQI.H.SETAKLF TGLFPEIE-----ATLSPRTKIYTE.. BOSVRL.E-----TVRFI.SST.ATPLF AES.NKELPYWKS.N.FVNNKK.PLRHVSQWK AES.NKELPYWKS.N.FVNNKK.PL.HISKWK
Спума	..R.....L...A...PA---VV.LLKEI. ..R.....L...A...PA---VV.LLKEI.		

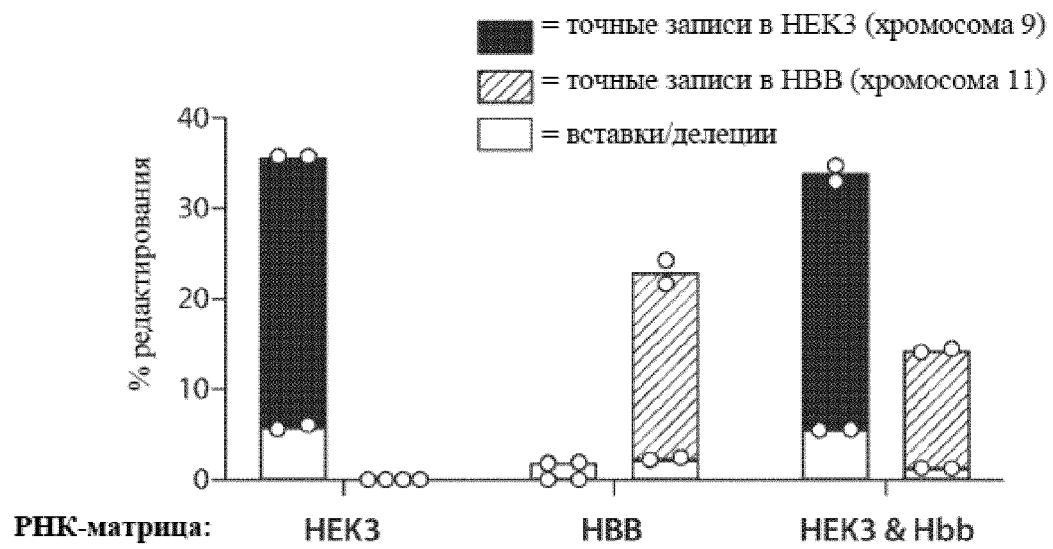
ФИГ. 36A

	T306K & W313F Anzalone et al.	L139P Baranauskas et al.	E607K Baranauskas et al.
Консенсусная	PKTLRQLOEFLGTIGWCRWLWIPGFAELLYP	DIHATVPGPPBLLS-LPSSLXWYTVLDLKD	EXLG-----G-LTSAGKXIRNKXELLALLK
Гамма	...P...R....A.F.....MAA.. ...P...R....A.F.....MAA.. ...K..VR....Y.....AQ.. ..T.AK.VR....A.F.....T.AA.. ...P...R....A.F.....MAA.. ..T.I...IRA...LV.Y..H...E.SIESKF	...P...N.YN...G....HQ..... ...P...N.YN...G....HQ..... T..P...N.YT...L...DRI..S..... ...P...N.YN..CA...QRS..... ...P...N.YN...G....HQ..... PLT.V.AS.TTV..N.A...H.F..I..SN.	..IH.EIYRRR.L...E..E...D... ..VH.EIYRRR.L...E..RE...N... ..VH.MIYRER.L...AG..A...AP...T ..VH.AIKQOR.L...RE...E...S...E ..IH.EIYRRR.L...E..RE...S... ..DP.HLMMHR.FV....TP....HK...EY...
Эпсилон	..IS..H...TV..DLO.VSNGT.TTRRP.QL RWA.PE..AL..E.Q.VSNGT.TLRQP.HS RWA.PE..AL..E.Q.VSNGT.TLRQP.HS OW..TE..VI..E.Q.VSNGT.LLKE.OS LQ..NDF.KL..D.N.L.HYKLTIGD.K.. LR..NDF.KL..D.N.I.HYKLTIGE.K.. L..NDF.KL..D.Q.J.HYKLTISA.V.. ..D.K...SI..LLNFA.NF..NYS..VK.. ..D.K...SI..LLNFA.NF..N....VQ..	P.P.LS...D.TA..I..TH.PRIYC..... ..LSSSS...D.S...T.T.AHLQTI... ..LSSSS...D.S...T.T.AHLQTI..R.. TLTSFS...D.T...T.A.PHLQTI..T.. LMG.LQ..L.SPVA..I..QGYL..KIII...C ..MG.LQ..L.SPVA..V..KGWE..IIII...Q.C PMG.LQ..L.SPVA..I..LMV..HKI..I...C L.A.QNQHSAQI..SIVRG-KYK.T...TNG LTA.QNQHSAQI..ATIVRG-KYR.T...ANG	RT.....VLG.WLPDPVPSYAL.Y.. RT.A-----LGTFOGRSSQAPFO...P RT.A-----LGTFOGRSSQAPFO...P ..S.A-----IG.FLGTSAHOTLO.A.P PL.E-----TVAQI.H.SETAKLFLOCO PEIE-----ATLSPRTKIYTE.KH.O PE.E-----TVRFI.SST.ATPLFSK.O NKELPYWKS.N.FVNNKK.PLRHVSQWKIAE NKELPYWKS.N.FVNNKK.PL.HISKWKIAE
Дельта			
Бета			
Спума			

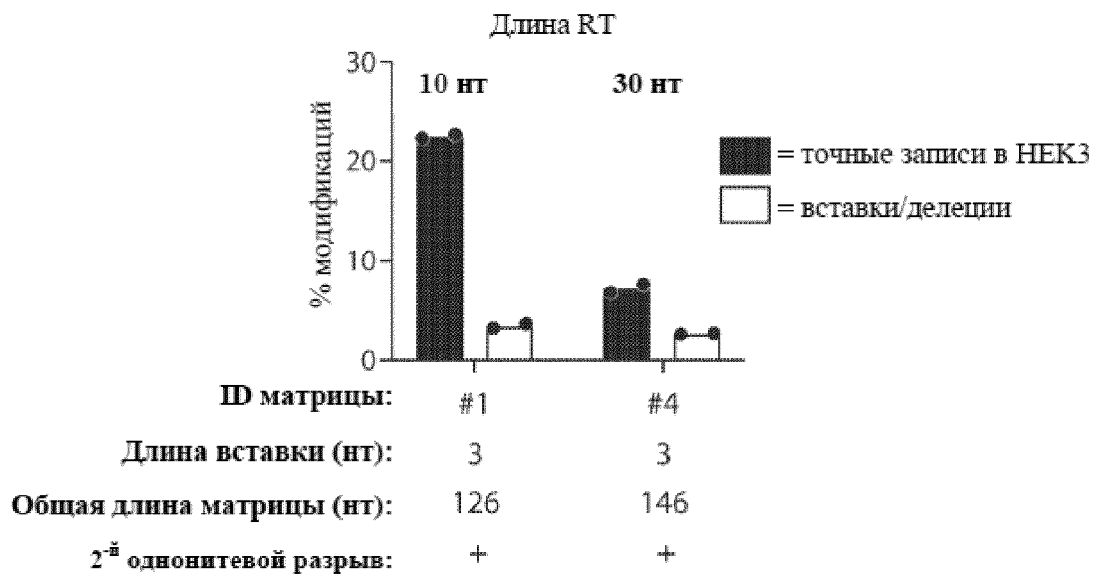
ФИГ. 36B



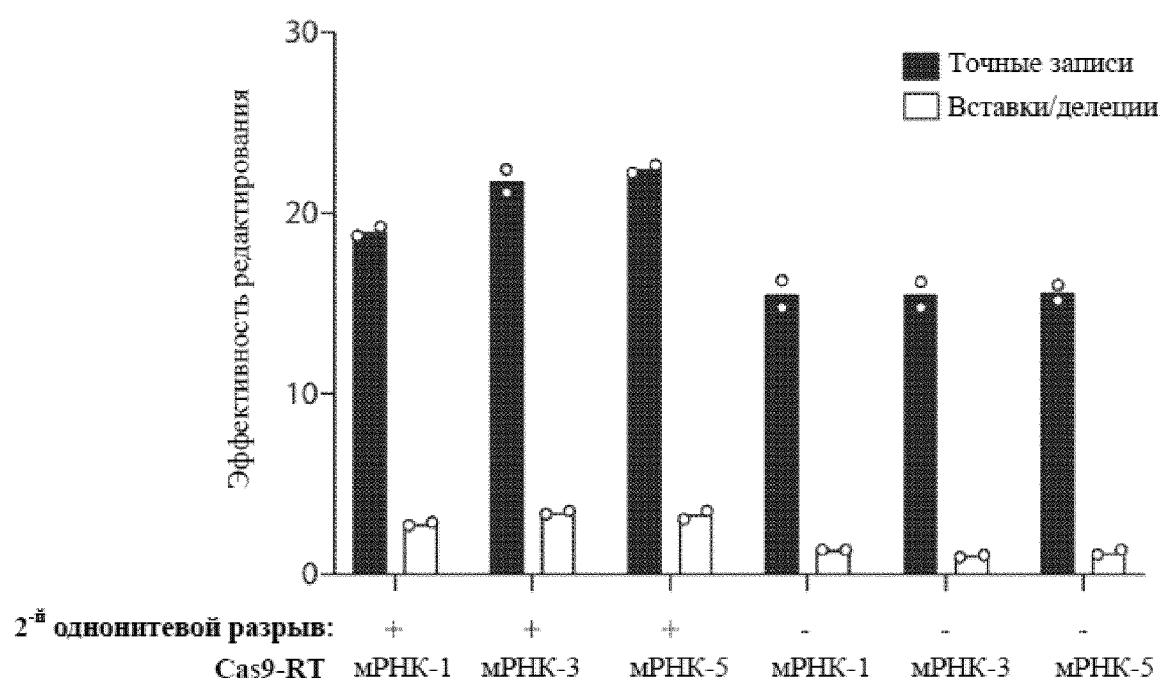
ФИГ. 37



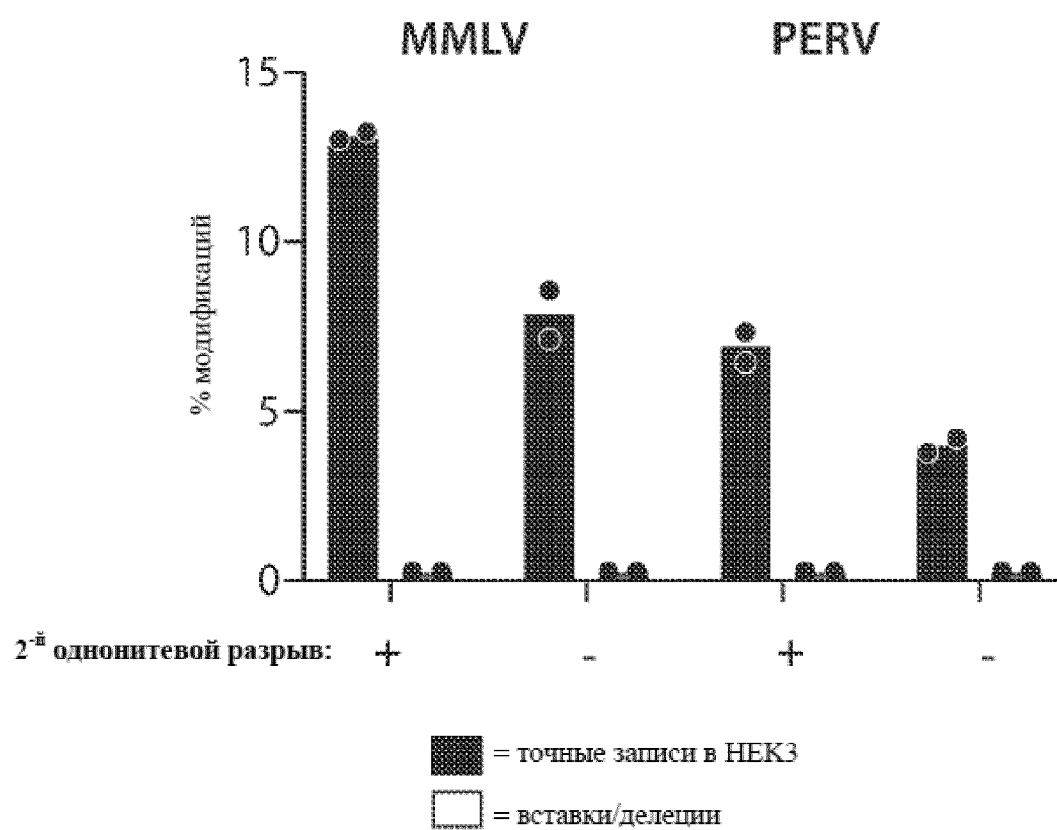
ФИГ. 38



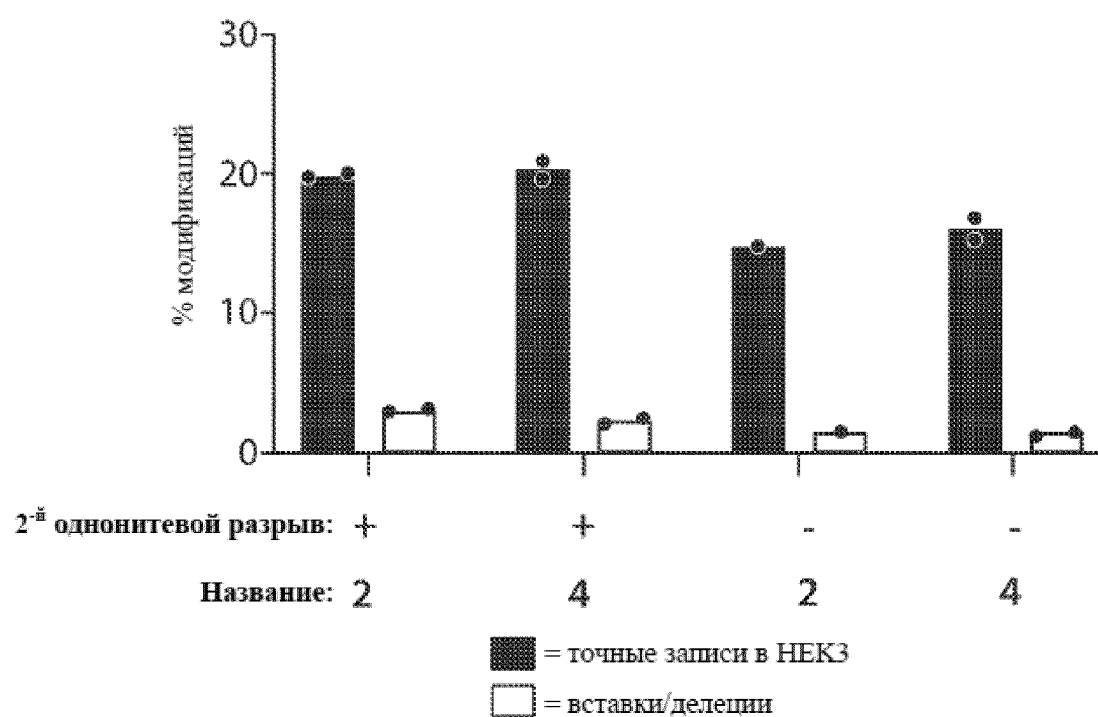
ФИГ. 39



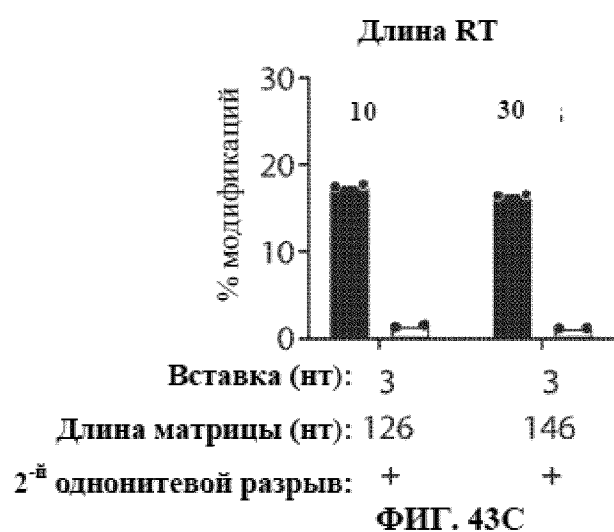
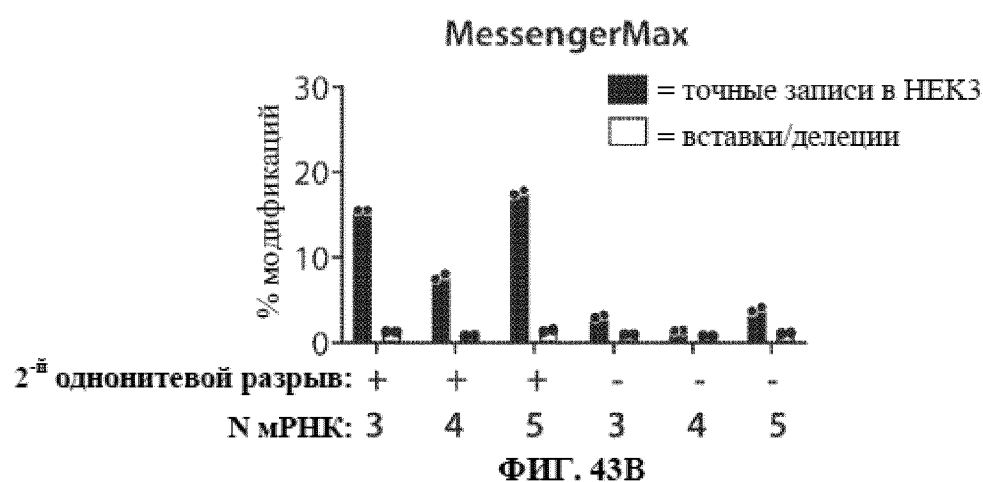
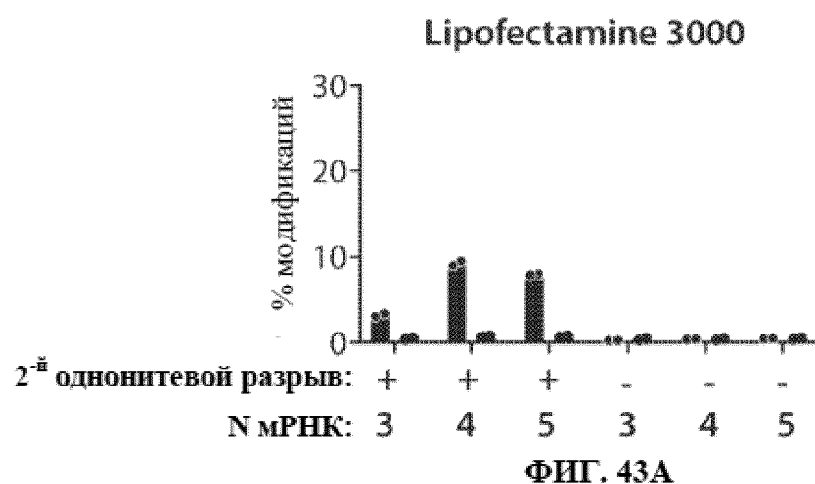
ФИГ. 40

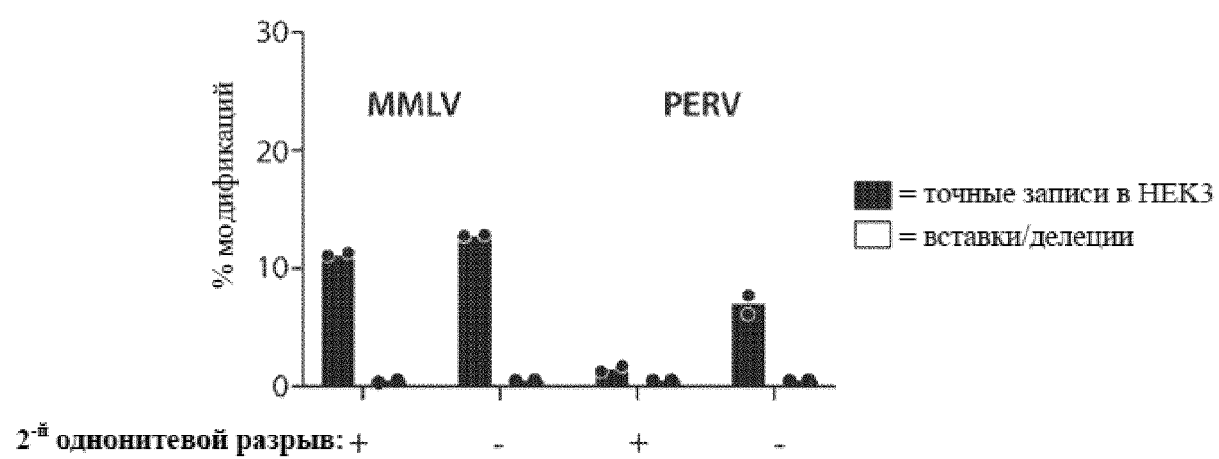


ФИГ. 41



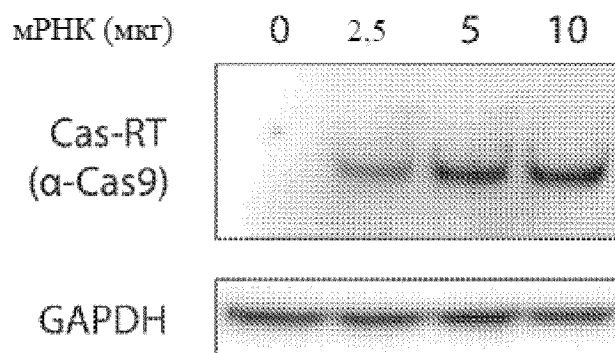
ФИГ. 42





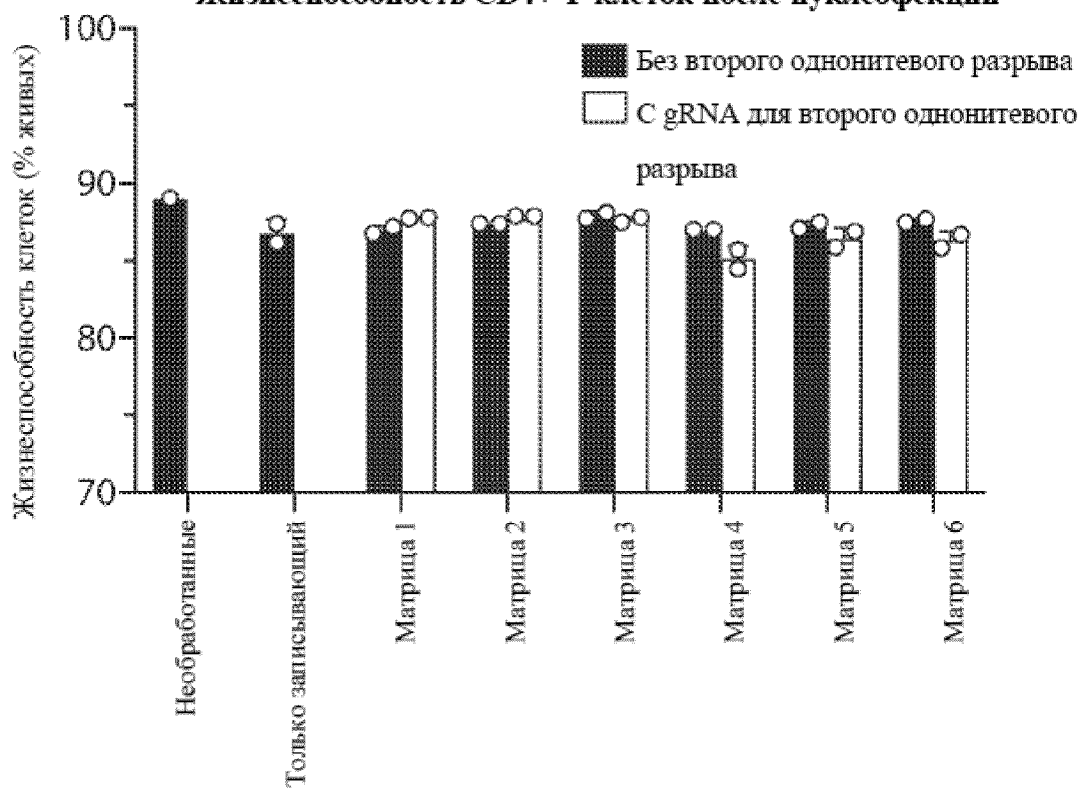
ФИГ. 44

Экспрессия полипептида Gene Writer

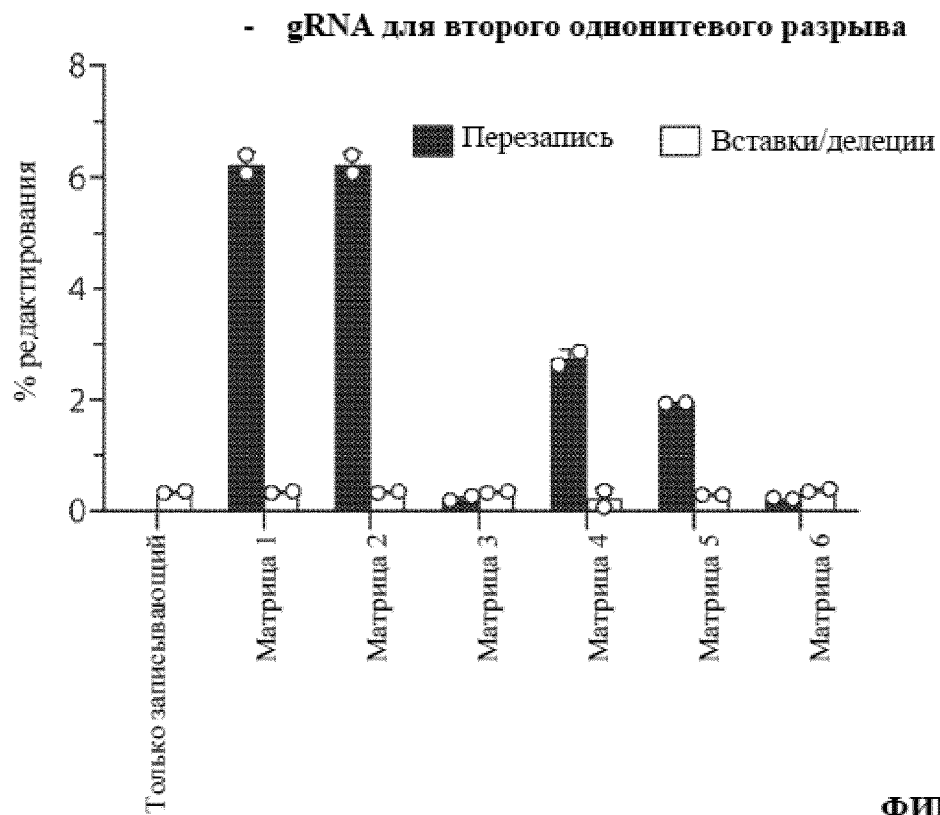


ФИГ. 45А

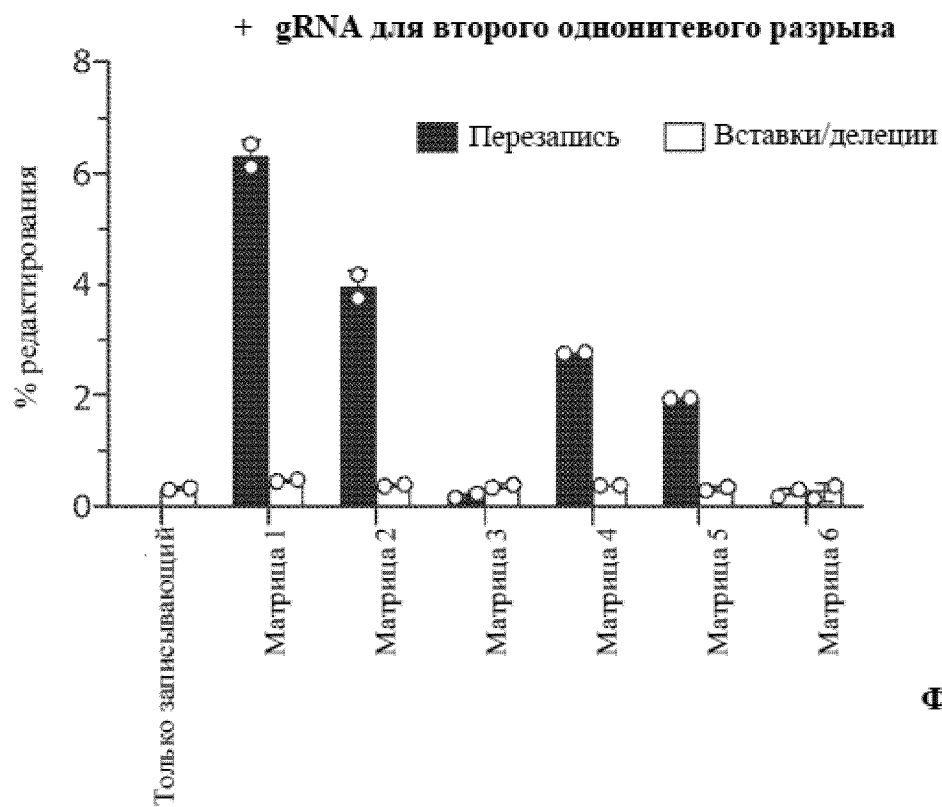
Жизнеспособность CD4⁺ Т-клеток после нуклеофекции



ФИГ. 45В

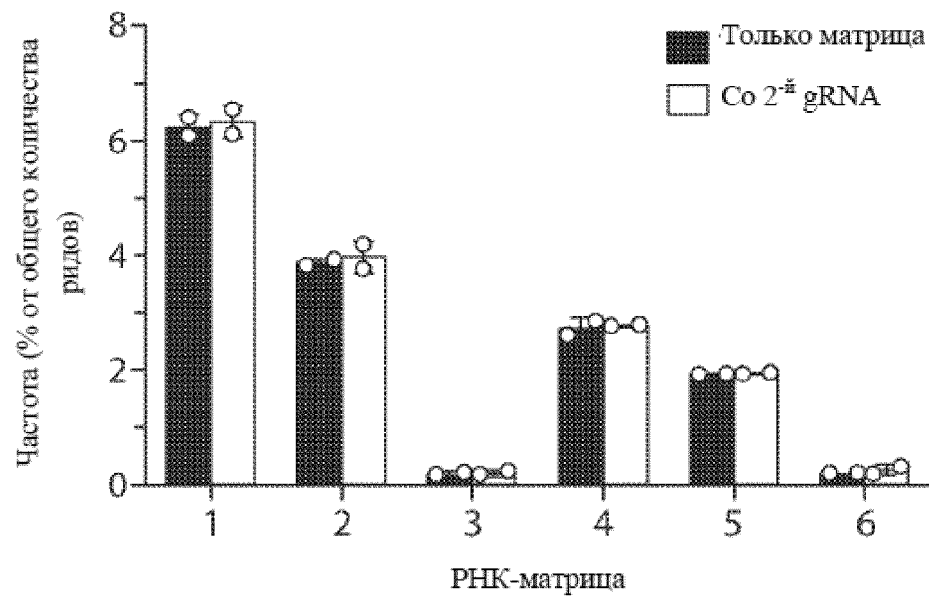


ФИГ. 46А



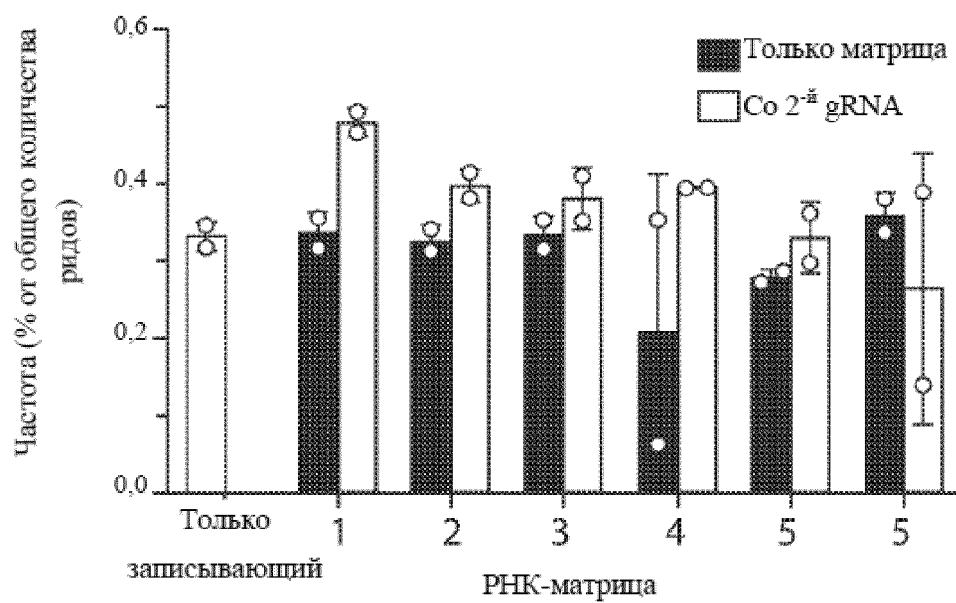
ФИГ. 46В

Точная запись

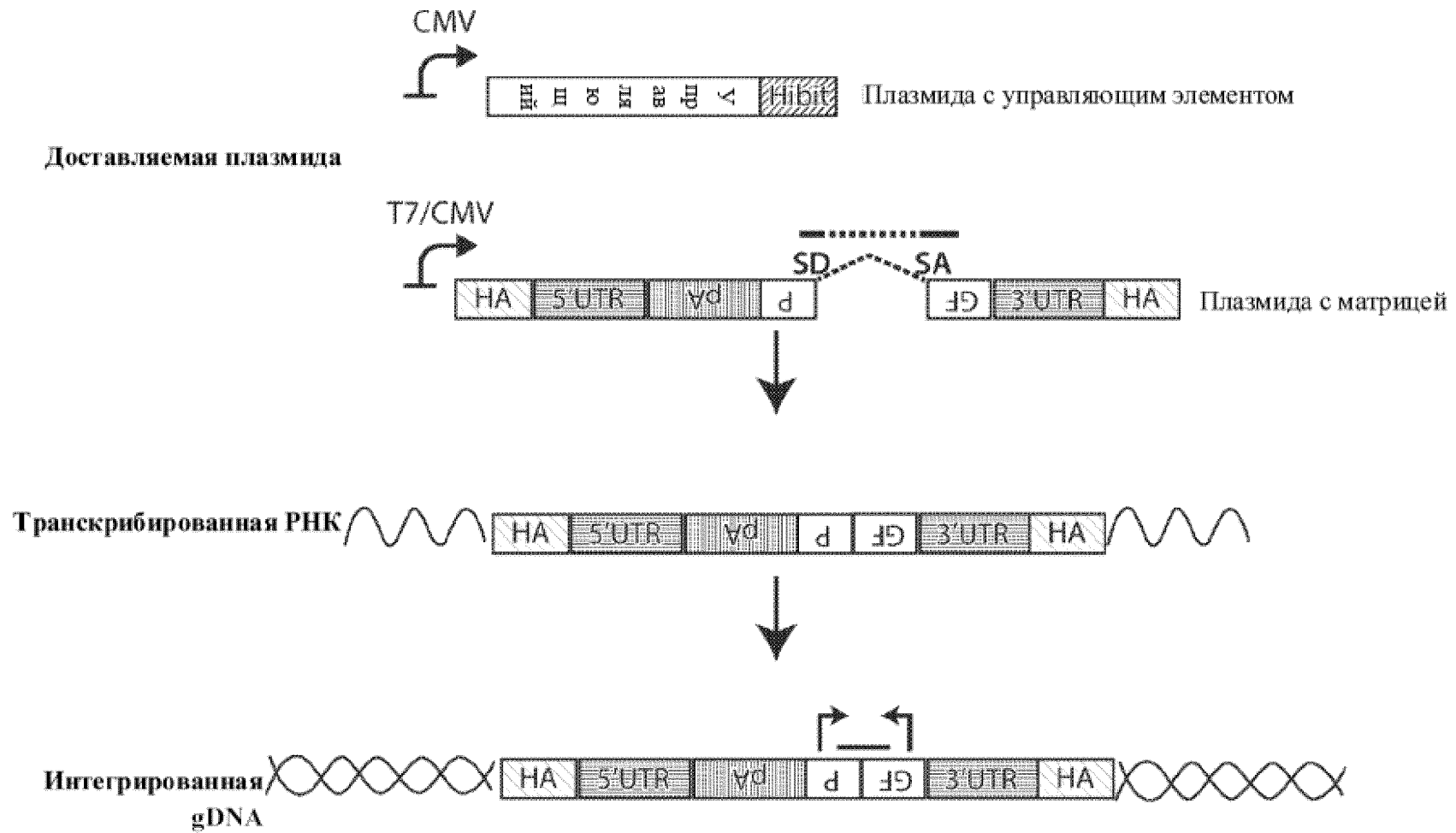


ФИГ. 47А

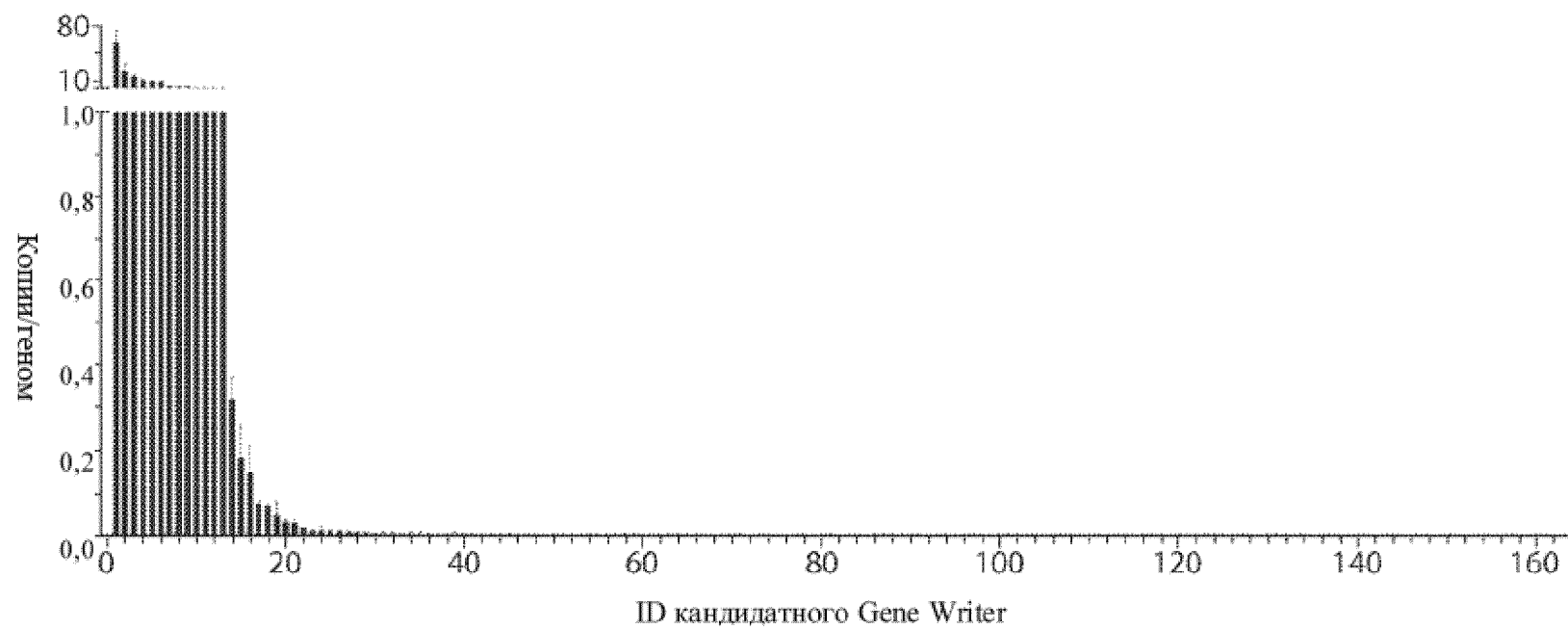
Неточные редактирования



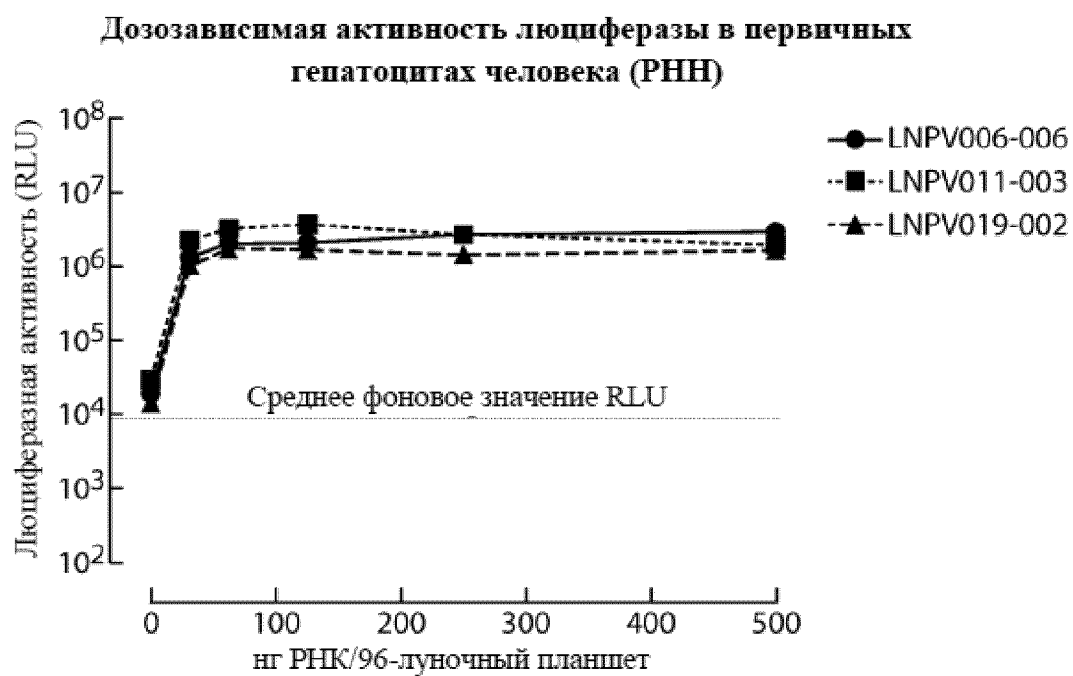
ФИГ. 47В



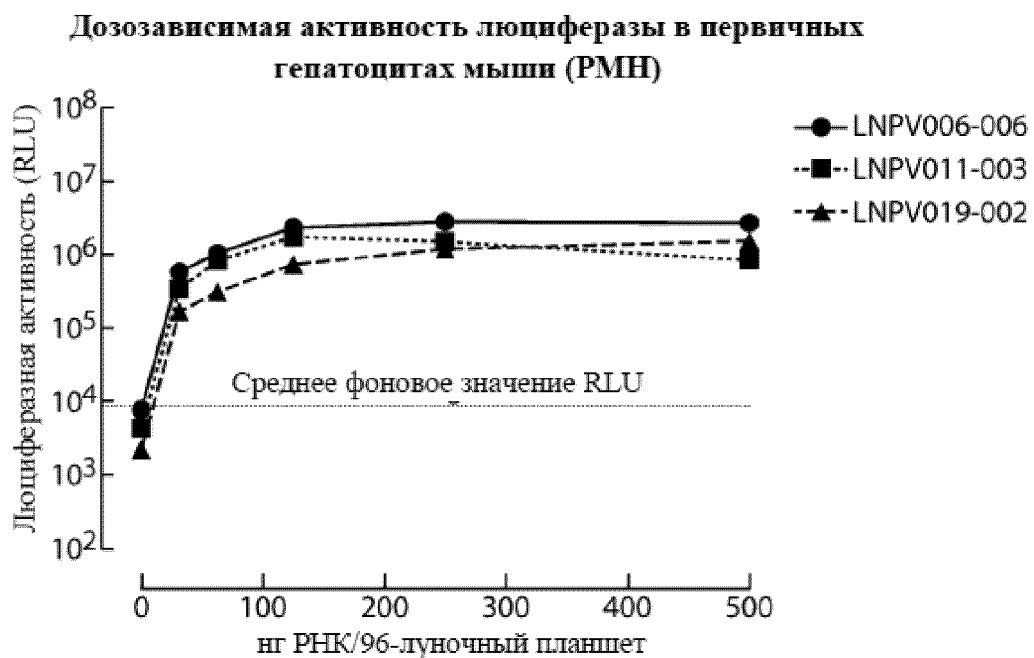
Фигура 48



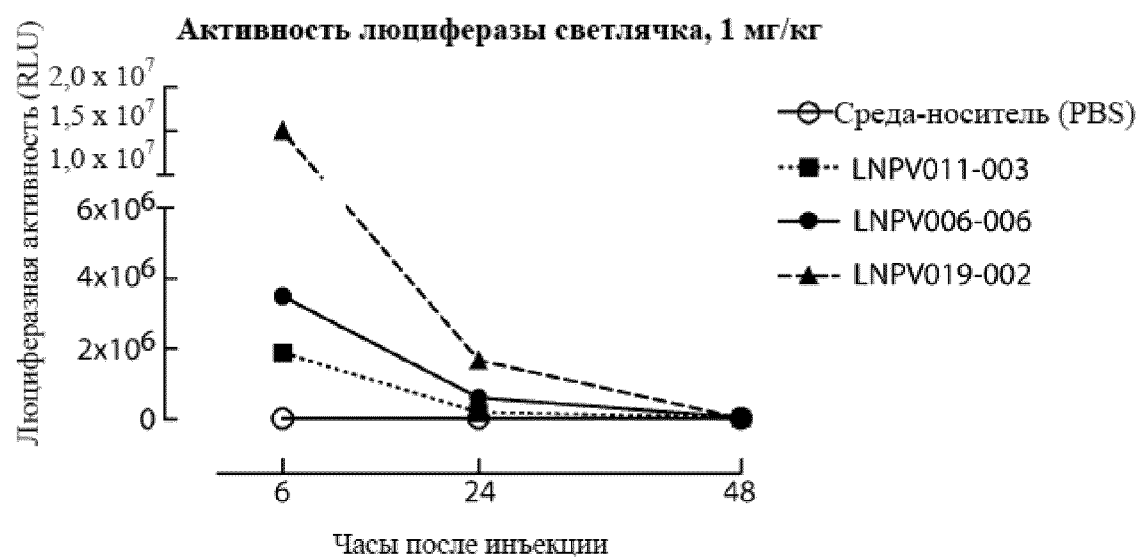
Фигура 49



ФИГ. 50А



ФИГ. 50В



ФИГ. 51