(19) Евразийское патентное ведомство

(21) 202292520 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.12.01

(22) Дата подачи заявки
2021.03.04

(51) Int. Cl. *C12N 15/11* (2006.01)

(54) СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА

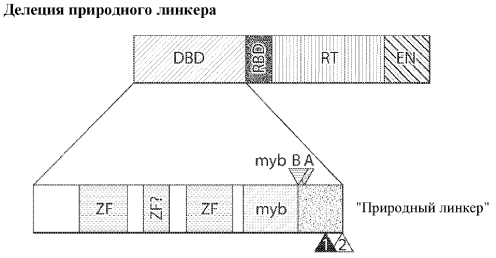
(31) **62/985,291; 63/035,638**
(32) **2020.03.04; 2020.06.05**
(33) **US**
(86) **PCT/US2021/020933**
(87) **WO 2021/178709 2021.09.10**

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

(72) Изобретатель:
Штайнберг Барретт Итан, Ботмер Энн Хелен, Саломон Уильям Эдвард, Щербакова Инна, Котта-Рамузино Чечилия Джованна Сильвиа, Рубенс Якоб Розенблум, Ситорик Роберт Джеймс, Ван Цзы Цзюнь (US)

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

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



202292520 A1

202292520 A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-575577EA/071

СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ МОДУЛИРОВАНИЯ ГЕНОМА

РОДСТВЕННЫЕ ЗАЯВКИ

Настоящая заявка заявляет приоритет по заявке на патент США под регистрационным номером 62/985291, поданной 4 марта 2020 года, и заявке на патент США под регистрационным номером 63/035638, поданной 5 июня 2020 года, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

ПРЕДПОСЫЛКИ ИЗОБРЕТЕНИЯ

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

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Настоящее изобретение относится к новым композициям, системам и способам для изменения генома в одном или нескольких местоположениях в клетке-, ткани- или субъекте-хозяине *in vivo* или *in vitro*. В частности, в настоящем изобретении представлены композиции, системы и способы для введения экзогенных генетических элементов в геном хозяина.

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

1. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где полипептид содержит мутацию, приводящую к инактивации и/или делеции сигнала ядрышковой локализации.

2. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) первый домен связывания ДНК-мишени, например, содержащий первый домен "Zn-палец", (ii) домен обратной транскриптазы, (iii) эндонуклеазный домен

и (iv) второй домен связывания ДНК-мишени, например, содержащий второй домен "Zn-палец", гетерологичный по отношению к первому домену связывания ДНК-мишени; и

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

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

3. Система согласно варианту осуществления 2, где (iii) содержит (iv).

4. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где система способна разрезать первую нить ДНК-мишени по меньшей мере дважды (например, дважды), и

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

5. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где система способна разрезать первую нить и вторую нить ДНК-мишени, и

где расстояние между разрезами является таким же, как расстояние между разрезами, произведенными эндонуклеазным доменом, например, эндонуклеазным доменом встречающейся в природе ретротранспозазы.

6. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где (a), (b) или (a) и (b) дополнительно содержат 5'-UTR и/или 3'-UTR, функционально связанные с последовательностью, кодирующей полипептид, последовательностью гетерологичного объекта (например, кодирующей последовательностью, содержащейся в последовательности гетерологичного объекта) или ими обеими.

7. Система согласно варианту осуществления 6, где 5'-UTR и/или 3'-UTR обеспечивают увеличение экспрессии функционально связанной(связанных) последовательности(последовательностей) на по меньшей мере 10%, 20%, 30%, 40%, 50%, 70%, 70%, 80%, 90% или 100% по сравнению с в остальном аналогичной нуклеиновой кислотой, содержащей эндогенную(эндогенные) UTR, ассоциированную(ассоциированные) с последовательностью гетерологичного объекта, или минимальную 5'-UTR и минимальную 3'-UTR.

8. Система для модификации ДНК, содержащая:

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

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

где:

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

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

9. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, (ii) последовательность гетерологичного объекта и (iii) рибозим, который является гетерологичным по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации.

10. Система согласно варианту осуществления 9, где рибозим является гетерологичным по отношению к (b)(i).

11. Система согласно вариантам осуществления 9 или 10, где РНК-матрица содержит (iv) второй рибозим, например, который является эндогенным по отношению к

(a)(i), (a)(ii), (b)(i) или их комбинации, например, где второй рибозим является эндогенным по отношению к (b)(i).

12. Система согласно вариантам осуществления 9 или 10, где гетерологичный рибозим заменяет рибозим, эндогенный по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации, например, где второй рибозим является эндогенным по отношению к (b)(i).

13. Система для модификации ДНК, содержащая:

необязательно (a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, (ii) последовательность гетерологичного объекта, (iii) 5'-UTR, способную расщепляться с образованием фрагмента и расщепленной РНК-матрицы, где 5'-UTR необязательно представляет собой последовательность, которая связывает полипептид,

где 5'-UTR содержит одну или несколько мутаций (например, по сравнению с 5'-UTR дикого типа, например, описанной в данном документе), которые обеспечивают увеличение аффинности фрагмента к расщепленной РНК-матрице, например, таким образом, что фрагмент гибридизируется с расщепленной РНК-матрицей (например, 5'-UTR расщепленной РНК-матрицы), например, в жестких условиях, например, где жесткие условия включают гибридизацию в 4x хлориде натрия/цитрате натрия (SSC) при приблизительно 65°C с последующей промывкой в 1x SSC при приблизительно 65°C.

14. Система согласно варианту осуществления 13, где РНК-матрица, например, 5'-UTR, содержит рибозим, который расщепляет РНК-матрицу (например, в 5'-UTR).

15. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где (a), (b) или (a) и (b) содержат интрон, который обеспечивает увеличение экспрессии полипептида, последовательности гетерологичного объекта (например, кодирующей последовательности, расположенной в последовательности гетерологичного объекта) или их обоих.

16. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение системы в клетку, где система содержит:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

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

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

17. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где последовательность гетерологичного объекта содержит последовательность, например, ген или его фрагмент, из любой из таблиц 10A-10D или 11A-11G.

18. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где полипептид модифицирован для усиления активности или изменения специфичности; и

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

19. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, и (ii) последовательность гетерологичного объекта, где РНК-матрица содержит одну или несколько химических модификаций, выбранных из дигидроуридина, инозина, 7-метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-O-метилриботимидина, 2'-O-этилриботимидина, 2'-фторриботимидина, C5-пропинилдезоксицитидина (pdC), C5-пропинилдезоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксицитидина, 5-метоксиметилдезоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксицитидина, 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).

20. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен связывания с ДНК-мишенью, (ii) домен обратной транскриптазы, необязательно (iii) эндонуклеазный домен, где полипептид содержит гетерологичный линкер, заменяющий часть (i), (ii) или (iii) или заменяющий эндогенный линкер, соединяющий два из (i), (ii) или (iii); и

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

21. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

(b) РНК-матрицу (или ДНК, кодирующую РНК-матрицу), содержащую (i) последовательность, которая связывает полипептид, (ii) последовательность гетерологичного объекта, (iii) первый домен гомологии, содержащий по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 5'-конце РНК-матрицы, и (iv) второй домен гомологии, содержащий по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 3'-конце РНК-матрицы.

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

23. Система согласно варианту осуществления 22, где активность сигнала ядрышковой локализации снижена на по меньшей мере 50%, 60%, 70%, 80%, 90%, 95% или 99%.

24. Система согласно любому из вариантов осуществления 22 или 23, где полипептид содержит сигнал ядрышковой локализации (NLS), например, эндогенный NLS или экзогенный NLS.

25. Система согласно любым предыдущим вариантам осуществления, где полипептид из (a) содержит домен связывания с ДНК-мишенью (например, эндонуклеазный домен содержит домен связывания с ДНК-мишенью), например, первый домен связывания с ДНК-мишенью, или (a) дополнительно содержит домен связывания с ДНК-мишенью, например, первый домен связывания с мишенью.

26. Система согласно варианту осуществления 25, где:

полипептид из (a) дополнительно содержит второй домен связывания с ДНК-мишенью, например, домен "Zn-палец", который является гетерологичным, например, по отношению к первому домену связывания с ДНК-мишенью или эндонуклеазному домену.

27. Система согласно варианту осуществления 26, где эндонуклеазный домен содержит второй домен связывания с ДНК-мишенью.

28. Система согласно вариантам осуществления 26 или 27, где второй домен связывания с ДНК-мишенью влияет на эндонуклеазную активность полипептида.

29. Система согласно любому из предыдущих вариантов осуществления, где второй домен связывания с ДНК-мишенью влияет на активность внесения полипептидом однонитевых разрывов в ДНК.

30. Система согласно любому из предыдущих вариантов осуществления, где второй домен связывания с ДНК-мишенью связывает локус, представленный в таблице E3.

31. Система согласно любому из предыдущих вариантов осуществления, где локус, представленный в таблице E3, имеет геномный балл, составляющий по меньшей мере 6.

32. Система согласно любому из предыдущих вариантов осуществления, где полипептид из (a) связывается с меньшим количеством последовательностей ДНК-мишеней, чем аналогичный полипептид, который содержит только первый домен связывания с ДНК-мишенью или второй домен связывания с ДНК-мишенью, например, где наличие второго домена связывания с ДНК-мишенью в полипептиде с первым доменом связывания с ДНК-мишенью обеспечивает более точную специфичность полипептида в отношении последовательности-мишени по сравнению со специфичностью полипептида в отношении последовательности-мишени у полипептида, содержащего только первый домен связывания с ДНК-мишенью.

33. Система согласно любому из предыдущих вариантов осуществления, где второй домен связывания с ДНК-мишенью связывается с последовательностью геномной ДНК, которая находится на расстоянии менее 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 или 1 нуклеотида от геномной последовательности, с которой связывается первый домен связывания с ДНК-мишенью.

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

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

36. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит белок CRISPR/Cas, и второй домен связывания с ДНК-мишенью содержит эффекторный домен TAL.

37. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит белок CRISPR/Cas, и второй домен связывания с ДНК-мишенью содержит домен "Zn-палец".

38. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит белок CRISPR/Cas, и второй домен связывания с ДНК-мишенью содержит белок CRISPR/Cas.

39. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит белок CRISPR/Cas, и второй домен связывания с ДНК-мишенью содержит мегануклеазный домен.

40. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит эффекторный домен TAL, и второй домен связывания с ДНК-мишенью содержит домен "Zn-палец".

41. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит эффекторный домен TAL, и второй домен связывания с ДНК-мишенью содержит эффекторный домен TAL.

42. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит эффекторный домен TAL, и второй домен связывания с ДНК-мишенью содержит мегануклеазный домен.

43. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит домен "Zn-палец", и второй домен связывания с ДНК-мишенью содержит домен "Zn-палец".

44. Система согласно любому из предыдущих вариантов осуществления, где первый домен связывания с ДНК-мишенью содержит домен "Zn-палец", и второй домен связывания с ДНК-мишенью содержит мегануклеазный домен.

45. Система согласно любому из предыдущих вариантов осуществления, где второй ДНК-связывающий домен связывается с последовательностью в геномном сайте "безопасная гавань" (GSH) или в геномном сайте Natural Harbor™.

46. Система согласно любому из предыдущих вариантов осуществления, где система способна разрезать первую нить ДНК-мишени и вторую нить ДНК-мишени, например, где разрезы находятся на расстоянии по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 или 200 нуклеотидов друг от друга (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов друг от друга).

47. Система согласно любому из предыдущих вариантов осуществления, где система способна разрезать первую нить ДНК-мишени по меньшей мере дважды (например, дважды), где разрезы находятся на расстоянии по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 или 200 нуклеотидов друг от друга (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов друг от друга).

48. Система согласно любому из предыдущих вариантов осуществления, где разрезы находятся на расстоянии 1-500, 1-400, 1-300, 1-200, 1-100, 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 1-5, 5-500, 5-400, 5-300, 5-200, 5-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30,

5-20, 5-10, 10-500, 10-400, 10-300, 10-200, 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-500, 20-400, 20-300, 20-200, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-500, 30-400, 30-300, 30-200, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-500, 40-400, 40-300, 40-200, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-500, 50-400, 50-300, 50-200, 50-100, 50-90, 50-80, 50-70, 50-60, 60-500, 60-400, 60-300, 60-200, 60-100, 60-90, 60-80, 60-70, 70-500, 70-400, 70-300, 70-200, 70-100, 70-90, 70-80, 80-500, 80-400, 80-300, 80-200, 80-100, 80-90, 90-500, 90-400, 90-300, 90-200, 90-100, 100-500, 100-400, 100-300, 100-200, 200-500, 200-400, 200-300, 300-500, 300-400 или 400-500 нуклеотидов друг от друга.

49. Система согласно любому из предыдущих вариантов осуществления, где расстояние между разрезами является таким же, как расстояние между разрезами, произведенными эндонуклеазным доменом, например, эндонуклеазным доменом встречающейся в природе ретротранспозазы.

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

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

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

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

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

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

54. Система согласно предыдущему варианту осуществления, где 5'-UTR и/или 3'-UTR:

обеспечивают увеличение стабильности, например, периода полужизни, РНК-матрицы, РНК, транскрибированной из (a), или их обеих и/или

обеспечивают увеличение эффективности трансляции последовательности гетерологичного объекта, полипептида или их обоих.

55. Система согласно предыдущему варианту осуществления, где 5'-UTR содержит 5'-UTR из фактора комплемента 3 (C3) или ее функциональный фрагмент или вариант.

56. Система согласно любому из предыдущих вариантов осуществления, где 3'-UTR содержит 3'-UTR из орозомукоида 1 (ORM1) или ее функциональный фрагмент или вариант.

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

i) 5'-UTR обеспечивает увеличение скорости трансляции, например, по сравнению с в остальном аналогичной нуклеиновой кислотой, содержащей эндогенную(эндогенные) UTR, ассоциированную(ассоциированные) с последовательностью гетерологичного объекта, или минимальную 5'-UTR и минимальную 3'-UTR,

ii) 3'-UTR обеспечивает увеличение периода полужизни нуклеиновой кислоты, например, по сравнению с в остальном аналогичной нуклеиновой кислотой, содержащей эндогенную(эндогенные) UTR, ассоциированную(ассоциированные) с последовательностью гетерологичного объекта, или минимальную 5'-UTR и минимальную 3'-UTR, или

iii) как i), так и ii).

58. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит рибозим, который является гетерологичным по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации.

59. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим заменяет рибозим, эндогенный по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации.

60. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит второй рибозим, например, который является эндогенным по отношению к (a)(i), (a)(ii), (b)(i) или их комбинации.

61. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим расположен в 5'-UTR или 3'-UTR РНК-матрицы.

62. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим расположен с 5'-стороны от последовательности гетерологичного объекта или с 3'-стороны от последовательности гетерологичного объекта.

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

64. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим расположен с 5'-стороны от последовательности гетерологичного объекта и производит расщепление с 3'-стороны от гетерологичного рибозима, например, где гетерологичный рибозим представляет собой синтетический или встречающийся в природе рибозим типа "головка молотка".

65. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный рибозим расположен с 3'-стороны от последовательности гетерологичного

объекта и производит расщепление с 5'-стороны от гетерологичного рибозима, например, где гетерологичный рибозим выбран из рибозима из семейства HDV или рибозима типа "топор".

66. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит область гибридизации рибозима, например, матрица с измененным нацеливанием, как, например, посредством плеча гомологии, содержит модифицированную 5'-UTR, содержащую область гибридизации рибозима.

67. Система согласно любому из предыдущих вариантов осуществления, где часть рибозима гибридизируется (например, посредством уотсон-криковского спаривания оснований) с последовательностью с 5'- или 3'-стороны от рибозима.

68. Система согласно любому из предыдущих вариантов осуществления, где последовательность рибозима отличается от своей природной последовательности по меньшей мере 1, 2, 3, 4, 5, 6, 8, 9, 10, 15, 20, 25 или больше парами оснований.

69. Система согласно любому из предыдущих вариантов осуществления, где последовательность рибозима изменена по сравнению с ее природной последовательностью для гибридизации с плечом гомологии, которое расположено с 5'- или 3'-стороны от рибозима-мишени.

70. Система согласно любому из предыдущих вариантов осуществления, где система обеспечивает интеграцию последовательности гетерологичного объекта в геном-мишень с большей эффективностью, чем в остальном аналогичная система, в которой отсутствует гетерологичный рибозим, например, где на по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% или 100% больше клеток демонстрируют интеграцию в присутствии системы, содержащей гетерологичный рибозим, по сравнению с системой, не содержащей гетерологичный рибозим.

71. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит 5'-UTR, способную расщепляться с образованием фрагмента и расщепленной РНК-матрицы.

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

73. Система согласно любому из предыдущих вариантов осуществления, где 5'-UTR содержит одну или несколько мутаций (например, по сравнению с 5'-UTR дикого типа, описанной в данном документе, например, в таблицах 1 или 3 или из белкового домена, указанного в таблице 2).

74. Система согласно любому из предыдущих вариантов осуществления, где одна или несколько мутаций обеспечивают увеличение аффинности фрагмента к расщепленной РНК-матрице, например, таким образом, что фрагмент гибридизируется с расщепленной РНК-матрицей (например, 5'-UTR расщепленной РНК-матрицы) в жестких условиях, например, где жесткие условия для гибридизации включают гибридизацию в 4x хлориде натрия/цитрате натрия (SSC) при приблизительно 65°C с последующей промывкой в 1x SSC при приблизительно 65°C.

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

77. Система согласно любому из предыдущих вариантов осуществления, где интрон функционально связан (например, для распознавания белками клеточного сплайсинга) с последовательностью, кодирующей полипептид, последовательностью гетерологичного объекта (например, кодирующей последовательностью, расположенной в последовательности гетерологичного объекта) или ими обеими.

78. Система согласно любому из предыдущих вариантов осуществления, где интрон расположен в 5'-UTR (например, с 5'-стороны от последовательности гетерологичного объекта).

79. Система согласно любому из предыдущих вариантов осуществления, где интрон расположен в кодирующей последовательности в последовательности гетерологичного объекта.

80. Система согласно любому из предыдущих вариантов осуществления, где интрон расположен в переднем направлении по отношению к кодирующей последовательности в последовательности гетерологичного объекта.

81. Система согласно любому из предыдущих вариантов осуществления, где интрон расположен в заднем направлении по отношению к кодирующей последовательности в последовательности гетерологичного объекта.

82. Система согласно любому предыдущему варианту осуществления, где интрон подвергается сплайсингу после транскрипции РНК-матрицы и перед обратной транскрипцией с праймированием мишенью в ДНК-мишень, например, геномную ДНК.

83. Система согласно любому из предыдущих вариантов осуществления, где интрон подвергается сплайсингу после транскрипции последовательности гетерологичного объекта после того, как последовательность гетерологичного объекта интегрируется в ДНК-мишень, например, геномную ДНК.

84. Система согласно любому из предыдущих вариантов осуществления, где интрон содержит сайт связывания микроРНК.

85. Система согласно любому из предыдущих вариантов осуществления, где эндонуклеазный домен (например, эндонуклеазный домен R2Tg или R2-1_ZA) распознает мотив (например, GG или AAGG, TAAGGT или TTAAGGTAGC), и гетерологичный ДНК-связывающий домен распознает последовательность геномной ДНК, где мотив и последовательность геномной ДНК находятся на расстоянии 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-100, 100-150, 150-200 или 200-250 нуклеотидов друг от друга, где необязательно мотив, распознаваемый эндонуклеазным доменом, содержит 4, 5, 6, 7, 8, 9 или 10 последовательных нуклеотидов из TTAAGGTAGC, AAGGTAGCCAAA или

TAAGGTAGCCAAA, или где мотив, распознаваемый эндонуклеазным доменом, содержит 2 или 3 последовательных нуклеотида из AAGG.

86. Система согласно любому из предыдущих вариантов осуществления, где мотив находится выше последовательности геномной ДНК, например, мотив находится на приблизительно 30-80, 40-70, 50-60 или 55 нуклеотидов выше последовательности геномной ДНК.

87. Система согласно любому из предыдущих вариантов осуществления, где мотив находится ниже последовательности геномной ДНК, например, мотив находится на приблизительно 10-30, 15-25 или 20 нуклеотидов ниже последовательности геномной ДНК.

88. Система согласно любому из предыдущих вариантов осуществления, где мотив имеет ту же ориентацию, что и последовательность геномной ДНК, или имеет обратно комплементарную ориентацию относительно последовательности геномной ДНК.

89. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный ДНК-связывающий домен (например, домен "цинковый палец") является N-концевым или C-концевым относительно эндонуклеазного домена.

90. Система согласно любому из предыдущих вариантов осуществления, где линкер (например, линкер из таблицы 38) расположен между гетерологичным ДНК-связывающим доменом и эндонуклеазным доменом.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

117. 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.

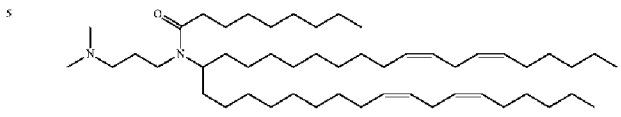
118. Система, содержащая первую липидную наночастицу, содержащую полипептид (или кодирующую его ДНК или РНК) системы Gene Writing (например, описанной в данном документе); и

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

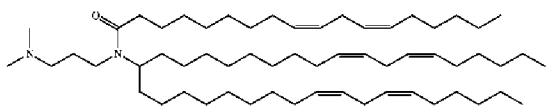
119. Система, набор, полипептид или реакционная смесь согласно любому из предыдущих вариантов осуществления, где система, молекула нуклеиновой кислоты, полипептид и/или ДНК, кодирующая его, составлены в виде липидной наночастицы (LNP).

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

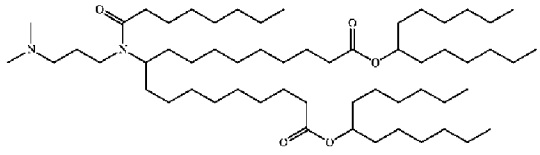
121. LNP согласно любому из предыдущих вариантов осуществления, где катионный липид имеет следующую структуру:



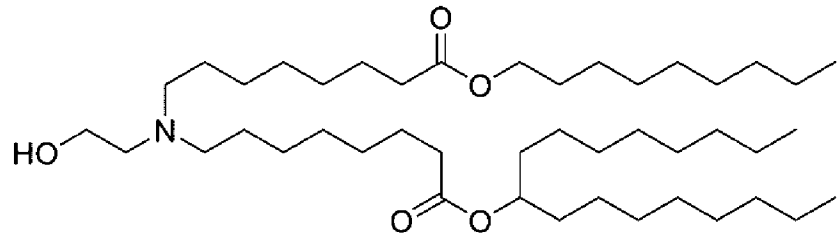
(i),



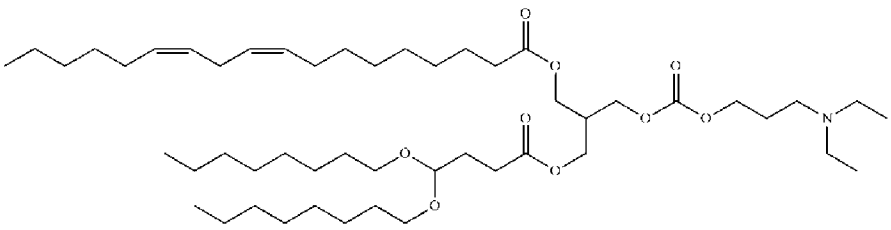
(ii),



(iii),



(vii) или



(ix).

122. 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.

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

124. Система, набор или полипептид согласно варианту осуществления M1, 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.

125. Система, набор или полипептид согласно варианту осуществления M1, где липидная наночастица (или состав, содержащий множество липидных наночастиц) не содержит альдегидов или содержит альдегиды в количестве, меньшем, чем предварительно выбранный уровень.

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

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

128. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где состав на основе липидных наночастиц получен с использованием

одного или нескольких липидных реагентов, характеризующихся общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 3%.

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

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

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

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

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

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

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

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

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

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

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

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

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

141. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов) определяется с помощью жидкостной хроматографии (LC), например, в сочетании с тандемной масс-спектрометрией (MS/MS), например, в соответствии со способом, описанным в примере 26.

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

143. Система, набор или полипептид согласно любому из предыдущих вариантов осуществления, где общее содержание альдегидов и/или количество разновидностей альдегидов определяют посредством выявления одной или нескольких химических модификаций нуклеотида или нуклеозида (например, рибонуклеотида или рибонуклеозида, например, содержащихся в молекуле нуклеиновой кислоты или выделенных из нее, например, как описано в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах, например, как описано в примере 27.

144. Система, набор или полипептид согласно варианту осуществления M21, где химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида выявляют посредством определения присутствия одного или нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 27.

145. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение любой вышеперечисленной системы в клетку, ткань или субъекту, где система осуществляет обратную транскрипцию последовательности РНК-матрицы в нить ДНК-мишени, за счет чего обеспечивается модификация нити ДНК-мишени, и где клетка характеризуется уменьшенной активностью Rad51-опосредованного пути репарации, уменьшенной экспрессией Rad51 или компонента Rad51-опосредованного пути репарации или не содержит функциональный Rad51-опосредованный путь репарации, например, не содержит функциональный ген Rad51, например, содержит мутацию (например, делецию), приводящую к инактивации одной или обеих копий гена Rad51 или другого гена в Rad51-опосредованном пути репарации.

146. Клетка-хозяин (например, клетка млекопитающего, например, клетка человека), содержащая любую вышеперечисленную систему, где клетка-хозяин характеризуется уменьшенной активностью Rad51-опосредованного пути репарации, уменьшенной экспрессией Rad51 или компонента Rad51-опосредованного пути репарации или не содержит функциональный Rad51-опосредованный путь репарации, например, не содержит функциональный ген Rad51, например, содержит мутацию (например, делецию), приводящую к инактивации одной или обеих копий гена Rad51 или другого гена в Rad51-опосредованном пути репарации.

147. Система согласно любому из предыдущих вариантов осуществления, где полипептид связывает промоторную область, 5'-UTR-область, экзон, интрон или 3'-UTR-область последовательности, например, гена или его фрагмента, из любой из таблиц 10A-10D или 11A-11G.

148. Система согласно любому из предыдущих вариантов осуществления, где полипептид дополнительно содержит гетерологичный линкер, заменяющий часть (i) домена связывания ДНК-мишени, (ii) домена обратной транскриптазы, необязательно (iii) эндонуклеазного домена или заменяющий эндогенный линкер, соединяющий два из (i), (ii) или (iii), где линкер необязательно представляет собой линкер из таблицы 38.

149. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i).

150. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (ii).

151. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (iii).

152. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i) и (ii).

153. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i) и (iii).

154. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (ii) и (iii).

155. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (i) и (ii).

156. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (i) и (iii).

157. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (ii) и (iii).

158. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

159. Система согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер содержит по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400 или 500 аминокислот.

160. Способ согласно любому из предыдущих вариантов осуществления, где ткань представляет собой печень, легкое, кожу, мышечную ткань (например, скелетную мышца), глаз или ткань глаза, кровь, клетки крови, иммунные клетки или центральную нервную систему.

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

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

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

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

165. Система согласно любому из предыдущих вариантов осуществления, где (a) содержит РНК, и (b) содержит RNA.

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

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

168. Система согласно любому из предыдущих вариантов осуществления, которая содержит только РНК или которая содержит больше РНК, чем ДНК, в соответствии с соотношением РНК:ДНК, составляющим по меньшей мере 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1 или 100:1.

169. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит открытую рамку считывания в ориентации 5'-3' в РНК-матрице.

170. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит открытую рамку считывания в ориентации 3'-5' в РНК-матрице.

171. Система согласно любому из предыдущих вариантов осуществления, где последовательность, которая связывает полипептид, представляет собой 3'-нетранслируемую последовательность.

172. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит 5'-нетранслируемую последовательность.

173. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица дополнительно содержит промотор, функционально связанный с последовательностью гетерологичного объекта, например, последовательность гетерологичного объекта в некоторых вариантах осуществления может содержать промотор, функционально связанный с последовательностью, такой как последовательность, кодирующая белок.

174. Система согласно любому из предыдущих вариантов осуществления, где промотор расположен между 5'-нетранслируемой последовательностью и последовательностью гетерологичного объекта.

175. Система согласно любому из предыдущих вариантов осуществления, где промотор расположен между 3'-нетранслируемой последовательностью, которая связывает полипептид, и последовательностью гетерологичного объекта.

176. Система согласно любому из предыдущих вариантов осуществления, где 5'-нетранслируемая последовательность представляет собой последовательность из столбца 5 таблицы 3 или последовательность, характеризующуюся по меньшей мере 80% идентичностью с ней.

177. Система согласно любому из предыдущих вариантов осуществления, где 3'-нетранслируемая последовательность представляет собой последовательность из столбца 6 таблицы 3 или последовательность, характеризующуюся по меньшей мере 80% идентичностью с ней.

178. Система согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта содержит фермент, мембранный белок, фактор свертывания крови, внутриклеточный белок, внеклеточный белок, структурный

белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок, запасной белок, иммунорецепторный белок (например, синтетический иммунорецепторный белок, такой как химерный антигенный рецепторный белок (CAR), Т-клеточный рецептор, В-клеточный рецептор), или антитело.

179. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 оснований или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 5'-конце РНК-матрицы.

180. Система согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 оснований или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 3'-конце РНК-матрицы.

181. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение системы согласно любому из предыдущих вариантов осуществления в клетку, ткань или субъекту, за счет чего обеспечивается модификация нити ДНК-мишени.

182. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению по меньшей мере 5 пар оснований экзогенной последовательности ДНК в геном клетки.

183. Способ согласно любому из предыдущих вариантов осуществления, который приводит к добавлению по меньшей мере 100 пар оснований экзогенной последовательности ДНК в геном клетки.

184. Способ согласно любому из предыдущих вариантов осуществления, который приводит к вставке последовательности гетерологичного объекта в ДНК-мишень при среднем числе копий, составляющем по меньшей мере 0,01, 0,05 или 0,5 копии на геном.

185. Способ согласно любому из предыдущих вариантов осуществления, который приводит к тому, что приблизительно 50-100% вставленных последовательностей гетерологичного объекта в ДНК-мишени не являются усеченными.

186. Способ согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота из (а) не интегрируется в геном клетки.

187. Способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 5'-конце РНК-матрицы.

188. Способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит по меньшей мере 5 или по меньшей мере 10 оснований, на 100% идентичных нити ДНК-мишени, на 3'-конце РНК-матрицы.

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

190. Система или способ согласно любому из предыдущих вариантов осуществления, где справедливо одно или несколько из следующего:

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

ii. последовательность гетерологичного объекта содержит тканеспецифический промотор или энхансер;

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

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

v. последовательность гетерологичного объекта кодирует один или несколько интронов;

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

vii. последовательность гетерологичного объекта является отличной от Т-клеточного химерного антигенного рецептора.

191. Система или способ согласно любому из предыдущих вариантов осуществления, где один или оба из домена обратной транскриптазы или эндонуклеазного домена получены из ретротранспозазы птиц, например, имеют последовательность из таблицы 1 или 3 или характеризуются по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

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

193. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид получен из ретротранспозазы птиц, например, ретротранспозазы птиц из столбца 8 таблицы 3 или последовательности, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

194. Система или способ согласно любому из предыдущих вариантов осуществления, где ретротранспозаза птиц представляет собой ретротранспозазу из *Taeniopygia guttata*, *Geospiza fortis*, *Zonotrichia albicollis* или *Tinamus guttatus* или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

195. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид получен из ретротранспозазы из столбца 8 таблицы 3 или последовательности, характеризующейся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

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

197. Система или способ согласно любому из предыдущих вариантов осуществления, где справедливо одно или несколько из следующего:

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

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

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

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

198. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) дополнительно содержит промотор, функционально связанный с последовательностью гетерологичного объекта,

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

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

199. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица (или ДНК, кодирующая РНК-матрицу) дополнительно содержит 5'-нетранслируемую последовательность, которая связывает полипептид, и 3'-нетранслируемую последовательность, которая связывает полипептид, и

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

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

200. Система или способ согласно любому из предыдущих вариантов осуществления, где по меньшей мере один из домена обратной транскриптазы, эндонуклеазного домена или домена связывания с ДНК-мишенью является гетерологичным.

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

идентичную) домену обратной транскриптазы не содержащего LTR ретротранспозона типа апуриновой/апиримидиновой эндонуклеазы (APE), и последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) эндонуклеазному домену не содержащего LTR ретротранспозона APE-типа.

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

203. Система или способ 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.

204. Система или способ 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.

205. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит промотор, функционально связанный с последовательностью гетерологичного объекта.

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

207. Система или способ согласно любому из вариантов осуществления 140-144, где полипептид содержит последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) последовательности под SEQ ID NO: 1016.

208. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит последовательность, на по меньшей мере 80% идентичную (например, на по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичную) последовательности из столбца 8 таблицы 3.

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

210. Система или способ согласно любому из предыдущих вариантов осуществления, где слитая нуклеиновая кислота содержит РНК.

211. Система или способ согласно любому из предыдущих вариантов осуществления, где слитая нуклеиновая кислота содержит ДНК.

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

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

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

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

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

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

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

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

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

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

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

последовательностью полипептида, указанного в таблицах 1-3, или ее доменом обратной транскриптазы или эндонуклеазным доменом.

223. Любая вышеперечисленная система или способ, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), характеризующуюся по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с последовательностью полипептида, указанного в любой из таблиц 1-3, или ее доменом обратной транскриптазы, эндонуклеазным доменом или ДНК-связывающим доменом.

224. Любая вышеперечисленная система или способ, где полипептид содержит последовательность из по меньшей мере 50 аминокислот (например, по меньшей мере 100, 150, 200, 300, 500 аминокислот), характеризующуюся по меньшей мере 80% идентичностью (например, по меньшей мере 85%, 90%, 95%, 97%, 98%, 99%, 100% идентичностью) с аминокислотной последовательностью из столбца 8 таблицы 3 или ее доменом обратной транскриптазы, эндонуклеазным доменом или ДНК-связывающим доменом.

225. Любая вышеперечисленная система или способ, где РНК-матрица содержит последовательность из таблицы 3 (например, одну или обе из 5'-нетранслируемой области из столбца 6 таблицы 3 и 3'-нетранслируемой области из столбца 7 таблицы 3) или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

226. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит последовательность из приблизительно 100-125 п. о. из 3'-нетранслируемой области из столбца 7 таблицы 3, например, где последовательность содержит нуклеотиды 1-100, 101-200 или 201-325 из 3'-нетранслируемой области из столбца 7 таблицы 3, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

227. Любая вышеперечисленная система или способ, где (a) содержит РНК, и (b) содержит РНК.

228. Любая вышеперечисленная система или способ, где (a), (b) или (a) и (b) не содержат ДНК или не содержат более чем 10%, 5%, 4%, 3%, 2% или 1% ДНК по массе или по молярному количеству.

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

230. Любая вышеперечисленная система, которая способна модифицировать ДНК посредством обратной транскрипции с праймированием мишенью.

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

пути эксцизионной репарации нуклеотидов или пути репарации, направляемой гомологией).

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

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

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

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

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

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

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

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

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

241. Способ модификации генома клетки млекопитающего, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления.

242. Способ обеспечения вставки ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления.

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

244. Способ добавления по меньшей мере 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления,

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

245. Способ добавления по меньшей мере 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления, где способ обеспечивает доставку только РНК в клетку млекопитающего.

246. Способ добавления по меньшей мере 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 п. о. экзогенной ДНК в геном клетки млекопитающего, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления, где способ обеспечивает доставку РНК и белка в клетку млекопитающего.

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

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

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

250. Способ модификации генома клетки человека, включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления,

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

где клетка человека не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК, например, как описано в примере 14 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

251. Способ обеспечения добавления экзогенной кодирующей области в геном клетки (например, клетки млекопитающего), включающий приведение клетки в контакт с системой согласно любому из предыдущих вариантов осуществления, где РНК-матрица содержит некодирующую нить экзогенной кодирующей области, где РНК-матрица необязательно не содержит кодирующую нить экзогенной кодирующей области, где доставка необязательно включает невирусную доставку.

252. Способ обеспечения экспрессии полипептида в клетке (например, клетке млекопитающего), включающий приведение клетки в контакт с системой согласно любому

из предыдущих вариантов осуществления, где РНК-матрица содержит некодирующую нить, являющуюся обратно комплементарной по отношению к последовательности, которая будет кодировать полипептид, где РНК-матрица необязательно не содержит кодирующую нить, кодирующую полипептид, где доставка необязательно включает невирусную доставку.

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

254. Способ согласно любому из предыдущих вариантов осуществления, где система работает независимо от ДНК-матрицы.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(b) представляет собой некодирующую последовательность;

(c) представляет собой структурированную РНК или

(d) образует по меньшей мере 1 структуру шпилечной петли.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

284. Любой вышеперечисленный способ, где последовательность гетерологичного объекта содержит:

(a) открытую рамку считывания, например, последовательность, кодирующую полипептид, например, фермент (например, лизосомальный фермент), мембранный белок, фактор свертывания крови, экзон, внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок), внеклеточный белок, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок или запасной белок;

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

(c) акцепторный сайт сплайсинга;

(d) поли(A)-сайт;

(e) сайт эпигенетической модификации или

(f) единицу экспрессии гена.

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

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

287. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень в геноме 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.

288. Любой вышеперечисленный способ, который приводит к тому, что приблизительно 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90% интегрированных элементов в сайте-мишени в геноме не являются усеченными согласно измерению с помощью анализа, описанного в данном документе, например, анализа из примера 6.

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

290. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень в клетке, где вставленная гетерологичная последовательность содержит менее 10%, 5%, 2%, 1%, 0,5%, 0,2% или 0,1% мутаций (например, SNP или одну или несколько делеций, например, усечений или внутренних делеций) по сравнению с гетерологичной последовательностью до вставки, например, согласно измерению с помощью анализа из примера 12 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

291. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень во множестве клеток, где менее чем 10%, 5%, 2% или 1% копий вставленной гетерологичной последовательности содержат мутацию (например, SNP или делецию, например, усечение или внутреннюю делецию), например, согласно измерению с помощью анализа из примера 12 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

292. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в геном клетки-мишени, и где клетка-мишень не демонстрирует повышение уровня экспрессии p53 или демонстрирует повышение уровня экспрессии p53 на менее чем 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии p53 измеряют по уровню белка p53, например, в соответствии со способом, описанным в примере 30 документа PCT/US2019/048607, включенного в данный документ посредством ссылки, или по уровню p53, фосфорилированного по Ser15 и Ser20.

293. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в геном клетки-мишени, и где клетка-мишень не демонстрирует повышение уровня экспрессии каких-либо генов, участвующих в репарации ДНК, и/или генов-супрессоров опухолей, или где повышение уровня экспрессии гена, участвующего в репарации ДНК, и/или гена-супрессора опухолей составляет не более 10%, 5%, 2% или 1%, например, где повышение уровня экспрессии измеряют с помощью секвенирования РНК, например, как описано в примере 14 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

294. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень (например, при числе копий, составляющем 1 вставку или более одной вставки) в приблизительно 1-80% клеток в популяции клеток, приведенной в контакт с системой, например, в приблизительно 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70% или 70-80% клеток, например, согласно измерению с помощью ddPCR одиночных клеток, например, как описано в примере 17 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

295. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень при числе копий, составляющем 1 вставку, в приблизительно 1-80% клеток в популяции клеток, приведенной в контакт с системой, например, в приблизительно 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70% или 70-80% клеток, например, согласно измерению с помощью выделения колоний и ddPCR, например, как описано в примере 18 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

296. Любой вышеперечисленный способ, который приводит к вставке последовательности гетерологичного объекта в сайт-мишень (целевые вставки) с более высокой частотой, чем для вставки в сайт, не являющийся мишенью (нецелевые вставки), в популяции клеток, где соотношение целевых вставок и нецелевых вставок составляет более чем 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 200:1, 500:1 или 1000:1, например,

с использованием анализа из примера 11 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

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

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

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

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

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

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

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

(a) включает введение в нуклеиновую кислоту-матрицу (например, РНК- или ДНК-матрицу) гетерологичной последовательности гомологии, характеризующейся по меньшей мере 50%, 60%, 70%, 80%, 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.

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

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

304. Способ согласно любому из предыдущих вариантов осуществления, где введение согласно (a) включает осуществление мутации одного или нескольких нуклеотидов (например, по меньшей мере 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80,

90 или 100 нуклеотидов) нуклеиновой кислоты-матрицы, в результате чего образуется сегмент нуклеиновой кислоты-матрицы, имеющий последовательность из последовательности гомологии.

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

306. 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.

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

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

309. Способ согласно любому из предыдущих вариантов осуществления, где мотив, распознаваемый эндонуклеазным доменом (например, по меньшей мере 2, 4, 6, 8, 10, 20, 30, 40 или по меньшей мере 50 нуклеотидов, или не более 50, 40, 30, 20, 10, 8, 6, 4 или 2, или менее 3) полипептида Gene Writer, используется в качестве затравки для перенацеливания системы Gene Writing, где ДНК-связывающий домен модифицирован таким образом, что связывание полипептида Gene Writer с новым сайтом-мишенью приводит к правильному расположению эндонуклеазного домена в сердцевинном мотиве для обеспечения эндонуклеазной активности, где необязательно мотив, распознаваемый эндонуклеазным доменом, содержит 4, 5, 6, 7, 8, 9 или 10 последовательных нуклеотидов из TTAAGGTAGC, AAGGTAGCCAAA или TAAGGTAGCCAAA, или где мотив, распознаваемый эндонуклеазным доменом, содержит 2, или 3, или 4 последовательных нуклеотида из AAGG.

310. Способ согласно любому из предыдущих вариантов осуществления, где последовательность AAGG в геноме используется в качестве затравки для перенацеливания системы Gene Writing, где ДНК-связывающий домен модифицирован таким образом, что связывание полипептида Gene Writer с новым сайтом-мишенью приводит к правильному расположению эндонуклеазного домена в мотиве AAGG для обеспечения эндонуклеазной активности.

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

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

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

где:

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

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

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

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

a) получение РНК-матрицы, описанной в любом из предыдущих вариантов осуществления, например, где РНК-матрица содержит (i) последовательность, которая связывает полипептид, содержащий домен обратной транскриптазы и эндонуклеазный домен, и (ii) последовательность гетерологичного объекта; и

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

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

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

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

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

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

317. Система или способ согласно любому из предыдущих вариантов осуществления, где последовательность гетерологичного объекта кодирует фермент (например, лизосомальный фермент), фактор свертывания крови (например, фактор I, II, V,

VII, X, XI, XII или XIII), мембранный белок, экзон, внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок), внеклеточный белок, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок или запасной белок.

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

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

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

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

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

323. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) получены из ретротранспозазы птиц, например, имеют последовательность из таблицы 1 или 3 или из белкового домена, указанного в таблице 2, или характеризуются по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

324. Система или способ согласно любому из предыдущих вариантов осуществления, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, где один или оба из (i) или (ii) получены из ретротранспозазы птиц, и где один или оба из (i) или (ii) дополнительно содержат 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.

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

326. Система или способ согласно любому из предыдущих вариантов осуществления, где нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или нуклеиновая кислота, кодирующая РНК-матрицу, являются отдельными нуклеиновыми кислотами.

327. Система или способ согласно любому из предыдущих вариантов осуществления, где РНК-матрица не кодирует активную обратную транскриптазу, например, содержит инактивированную мутантную обратную транскриптазу, например, как описано в примерах 1 или 2 документа PCT/US2019/048607, включенного в данный документ посредством ссылки, или не содержит последовательность обратной транскриптазы.

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

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

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

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

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

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

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

335. 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).

336. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) первый домен связывания с ДНК-мишенью, например,

содержащий первый домен "Zn-палец", (ii) домен обратной транскриптазы, (iii) эндонуклеазный домен и (iv) второй домен связывания с ДНК-мишенью, например, содержащий второй домен "Zn-палец", гетерологичный по отношению к первому домену связывания с ДНК-мишенью.

337. Полипептид или нуклеиновая кислота, кодирующая полипептид, согласно любому из предыдущих вариантов осуществления, где (iii) содержит (iv).

338. Полипептид или нуклеиновая кислота, кодирующая полипептид, где полипептид содержит (i) домен связывания с ДНК-мишенью, (ii) домен обратной транскриптазы, необязательно (iii) эндонуклеазный домен, где полипептид содержит гетерологичный линкер, заменяющий часть (i), (ii) или (iii) или заменяющий эндогенный линкер, соединяющий два из (i), (ii) или (iii).

339. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i).

340. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (ii).

341. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (iii).

342. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i) и (ii).

343. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (i) и (iii).

344. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, часть (ii) и (iii).

345. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (i) и (ii).

346. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (i) и (iii).

347. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер заменяет, например, путем делеции, эндогенный линкер, соединяющий (ii) и (iii).

348. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с аминокислотной последовательностью SGSETPGTSESATPES (SEQ ID NO: 1023) или GGGS (SEQ ID NO: 1024).

349. Полипептид согласно любому из предыдущих вариантов осуществления, где гетерологичный линкер содержит по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,

15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400 или 500 аминокислот.

350. Нуклеиновая кислота, кодирующая полипептид согласно любому вышеперечисленному варианту осуществления.

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

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

353. Клетка-хозяин, содержащая полипептид согласно любому вышеперечисленному варианту осуществления.

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

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

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

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

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

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

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

361. Любой вышеперечисленный вариант осуществления, где полипептид, домен обратной транскриптазы или ретротранспозаза содержат линкер, содержащий аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с

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

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

363. Любой предыдущий вариант осуществления, где линкер присоединен к остальной части полипептида в положении в ДНК-связывающем домене, РНК-связывающем домене, домене обратной транскриптазы или эндонуклеазном домене.

364. Любой предыдущий вариант осуществления, где линкер присоединен к остальной части полипептида в положении с N-концевой стороны от области альфа-спирали полипептида, например, в положении, соответствующем варианту v1, как описано в примере 26 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

365. Любой предыдущий вариант осуществления, где линкер присоединен к остальной части полипептида в положении с С-концевой стороны от области альфа-спирали полипептида, например, перед РНК-связывающим мотивом (например, в положении -1 относительно РНК-связывающего мотива), например, в положении, соответствующем варианту v2, как описано в примере 26 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

366. Любой предыдущий вариант осуществления, где линкер присоединен к остальной части полипептида в положении с С-концевой стороны от области статистического клубка полипептида, например, с N-концевой стороны относительно ДНК-связывающего мотива (например, ДНК-связывающего мотива c-Myb), например, в положении, соответствующем варианту v3, как описано в примере 26 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

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

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

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

370. Любой вышеперечисленный вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,21, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9 или 1,0 интегрированного элемента/геном.

371. Любой вышеперечисленный вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,085, 0,09, 0,1, 0,15 или 0,2 интегрированного элемента/геном.

372. Любой вышеперечисленный вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,036, 0,04, 0,05, 0,06, 0,07 или 0,08 интегрированного элемента/геном.

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

374. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функциональным фрагментом.

375. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность полипептида R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

376. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функциональным фрагментом.

377. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность полипептида R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

378. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функциональным фрагментом.

379. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность полипептида R2 среднего земляного вьюрка, например, *Geospiza fortis* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

380. Любой предыдущий вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,21 интегрированного элемента/геном.

381. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функциональным фрагментом.

382. Любой вышеперечисленный вариант осуществления, где полипептид содержит аминокислотную последовательность полипептида R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

383. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функциональным фрагментом.

384. Любой вышеперечисленный вариант осуществления, где домен обратной транскриптазы содержит аминокислотную последовательность полипептида R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

385. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичностью последовательности с полипептидом R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функциональным фрагментом.

386. Любой вышеперечисленный вариант осуществления, где ретротранспозаза содержит аминокислотную последовательность полипептида R4 человеческой аскариды, например, *Ascaris lumbricoides* (например, как описано в данном документе), или его функционального фрагмента и дополнительно содержит ряд замен по сравнению с этой природной последовательностью, например, по меньшей мере 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 замен.

387. Любой предыдущий вариант осуществления, где последовательность нуклеиновой кислоты РНК-матрицы или ее часть (например, часть, содержащая по меньшей мере приблизительно 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500 или 4000 нуклеотидов) интегрируется в геномы популяции клеток-мишеней при числе копий, составляющем по меньшей мере приблизительно 0,085 интегрированного элемента/геном.

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

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

390. Любой предыдущий вариант осуществления, где уровень белка p53 определяют в соответствии со способом, описанным в примере 30 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

391. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования p53 в клетке-мишени до

уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80% или 90% от уровня фосфорилирования p53, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

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

393. Любой предыдущий вариант осуществления, где уровень белка p21 определяют в соответствии со способом, описанным в примере 30 документа PCT/US2019/048607, включенного в данный документ посредством ссылки.

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

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

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

397. Любой вышеперечисленный вариант осуществления, где введение системы в клетку-мишень приводит к повышению уровня фосфорилирования Chk1 в клетке-мишени до уровня, составляющего менее чем приблизительно 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%,

60%, 70%, 80% или 90% от уровня фосфорилирования Chk1, индуцируемого путем введения сайт-специфической нуклеазы, например, Cas9, которая нацеливается на тот же геномный сайт, что и указанная система.

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

399. Система для модификации ДНК, содержащая:

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

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

400. Система для модификации ДНК, содержащая:

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

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

где домен RT имеет последовательность из таблицы 1 или 3 или из белкового домена, указанного в таблице 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

411. Система для модификации ДНК, содержащая:

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

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

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

412. Система для модификации ДНК, содержащая:

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

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

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

413. 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.

414. Система для модификации ДНК, содержащая:

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

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

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

415. Система для модификации ДНК, содержащая:

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

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

где (a)(ii) и/или (a)(iii) содержат молекулу TALE; молекулу "цинкового пальца" или молекулу CRISPR/Cas или их функциональный вариант (например, мутантный вариант).

416. Система для модификации ДНК, содержащая:

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

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

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

417. Система для модификации ДНК, содержащая:

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

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

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

419. 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.

420. Система для модификации ДНК, содержащая:

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

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

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

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

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

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

424. 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.

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

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

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

428. 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.

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

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

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

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

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

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

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

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

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

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

439. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит по меньшей мере одну (например, одну или две) гетерологичную последовательность гомологии, характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, содержащейся в молекуле ДНК-мишени (например, геномной ДНК).

440. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где одна из по меньшей мере одной гетерологичной последовательности гомологии расположена на расстоянии приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от 5'-конца молекулы нуклеиновой кислоты-матрицы.

441. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где одна из по меньшей мере одной гетерологичной последовательности

гомологии расположена на расстоянии приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от 3'-конца молекулы нуклеиновой кислоты-матрицы.

442. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии связывается в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов от сайта однонитевого разрыва (например, создаваемого с помощью никазы, например, эндонуклеазного домена, например, как описано в данном документе) в молекуле ДНК-мишени.

443. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется менее чем 50%, 40%, 30%, 20%, 10%, 5%, 4%, 3%, 2% или 1% идентичностью последовательности с последовательностью нуклеиновой кислоты, комплементарной эндогенной последовательности гомологии немодифицированной формы РНК-матрицы.

444. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью молекулы ДНК-мишени, которая отличается от последовательности, связываемой эндогенной последовательностью гомологии (например, замененной гетерологичной последовательностью гомологии).

445. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии содержит последовательность (например, на ее 3'-конце), характеризующуюся по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, расположенной с 5'-стороны от сайта однонитевого разрыва молекулы ДНК-мишени (например, сайта, подвергаемого однонитевому разрыву никазой, например, эндонуклеазным доменом, как описано в данном документе).

446. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии содержит последовательность (например, на ее 5'-конце), подходящую для инициации обратной транскрипции с праймированием мишенью (TPRT).

447. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где гетерологичная последовательность гомологии характеризуется по меньшей мере 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% гомологией с последовательностью, расположенной в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов от сайта-мишени для вставки (например, с 3'-стороны относительно него), например, для последовательности гетерологичного объекта (например, как описано в данном документе), в молекуле ДНК-мишени.

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

449. Система, способ или РНК-матрица согласно любому из предыдущих вариантов осуществления, где молекула нуклеиновой кислоты-матрицы содержит спейсерную последовательность gRNA (например, на расстоянии 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 или 100 нуклеотидов или в их пределах от ее 5'-конца).

450. 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.

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

452. 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.

453. 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.

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

455. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, представляет собой последовательность, например, последовательность UTR, которая связывает домен RT в условиях дикого типа, или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

456. 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.

457. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (iii) последовательность гетерологичного объекта, (iv) 3'-концевой

домен гомологии с мишенью, (i) последовательность, которая связывает сайт-мишень (например, вторую нить сайта в геноме-мишени), и (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида.

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

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

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

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

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

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

464. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность из (ii) специфично связывает домен RT, приведенный в таблице 1 или 3 или указанный в таблице 2, или последовательность домена RT, которая характеризуется по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

465. РНК-матрица согласно любому из предыдущих вариантов осуществления, где последовательность, которая специфично связывает домен RT, представляет собой последовательность из таблицы 1 или 3 или из белкового домена, указанного в таблице 2, или последовательность, характеризующуюся по меньшей мере 70, 75, 80, 85, 90, 95 или 99% идентичностью с ней.

466. РНК-матрица (или ДНК, кодирующая РНК-матрицу), содержащая в направлении от 5' к 3': (ii) последовательность, которая связывает эндонуклеазу и/или ДНК-связывающий домен полипептида, (i) последовательность, которая связывает сайт-мишень

(например, нередактируемую нить сайта в геноме-мишени), (iii) последовательность гетерологичного объекта и (iv) 3'-концевой домен гомологии.

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

468. 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.

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

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

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

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

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

474. Система для модификации ДНК, содержащая:

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

(b) вторую РНК-матрицу (или ДНК, кодирующую вторую РНК-матрицу), содержащую (i) последовательность, которая специфично связывает домен обратной транскриптазы (RT) полипептида (например, полипептида из (a)), (ii) последовательность связывания сайта-мишени (TSBS) и (iii) матричную последовательность RT.

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

476. 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.

477. 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 из белкового домена, указанного в таблице 2, или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ней.

478. 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 из белкового домена, указанного в таблице 2, 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.

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

480. 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.

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

482. Система, способ, набор, РНК-матрица или реакционная смесь согласно варианту осуществления 481, 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.

483. Система, способ, набор, РНК-матрица или реакционная смесь согласно вариантам осуществления 481 или 482, где miRNA представляет собой miR-142, и/или где клетка, не являющаяся мишенью, представляет собой клетку Купфера или клетку крови, например, иммунную клетку.

484. Система, способ, набор, РНК-матрица или реакционная смесь согласно вариантам осуществления 481 или 482, где miRNA представляет собой miR-182 или miR-183, и/или где клетка, не являющаяся мишенью, представляет собой ганглиозный нейрон дорсального корешка.

485. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из вариантов осуществления 481-484, 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.

486. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из вариантов осуществления 481-485, 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.

487. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из вариантов осуществления 481-486, где РНК, кодирующая полипептид из (a), содержит по меньшей мере 2, 3 или 4 сайта связывания с miRNA, например, где сайты связывания с miRNA распознаются одной и той же или разными miRNA.

488. Система, способ, набор, РНК-матрица или реакционная смесь согласно любому из вариантов осуществления 481-487, где РНК, экспрессирующаяся из последовательности гетерологичного объекта, интегрированной в ДНК-мишень, содержит по меньшей мере 2, 3 или 4 сайта связывания с miRNA, например, где сайты связывания с miRNA распознаются одной и той же или разными miRNA.

Определения

Домен. Термин "домен", используемый в данном документе, относится к структуре биомолекулы, которая участвует в осуществлении определенной функции биомолекулы. Домен может содержать непрерывную область (например, непрерывную последовательность) или отдельные несмежные области (например, несмежные последовательности) биомолекулы. Примеры белковых доменов включают без ограничения эндонуклеазный домен, ДНК-связывающий домен, домен обратной транскрипции; примером домена нуклеиновой кислоты является регуляторный домен, такой как домен связывания с фактором транскрипции.

Экзогенный. Используемый в данном документе термин "экзогенный" при использовании его в отношении биомолекулы (такой как последовательность нуклеиновой кислоты или полипептид) означает, что биомолекула была введена в геном, клетку или организм хозяина посредством манипуляций, осуществляемых человеком. Например, нуклеиновая кислота, которую добавляют в существующий геном, клетку, ткань или субъекту с применением методик рекомбинантной ДНК или других способов, является

экзогенной по отношению к существующей последовательности нуклеиновой кислоты, клетке, ткани или субъекту.

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 однонитевому разрыву полипептидом, описанным в данном документе, но обозначения "первая" и "вторая" нити не имеют отношения к порядку, в котором происходят такие однонитевые 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 (мотив C-C) рецептора 5 (CCR5) - ген хемокинового рецептора, известного как корецептор HIV-1; (iii) человеческий ортолог мышиного локуса Rosa26; (iv) локус рДНК; (v) локус альбумина, например, для применения в клетках печени; (vi) локус константной области альфа-цепи Т-клеточного рецептора (TRAC), например, для применения в Т-клетках. Дополнительные сайты GSH известны и описаны, например, в Pellenz et al., электронная публикация от 20 августа 2018 года (https://doi.org/10.1101/396390).

Гетерологичный. Термин "гетерологичный" при использовании для описания первого элемента по отношению ко второму элементу означает, что первый элемент и второй элемент не существуют в природе в описываемом расположении. Например,

гетерологичные полипептид, молекула нуклеиновой кислоты, конструкция или последовательность относятся к (a) полипептиду, молекуле нуклеиновой кислоты или части последовательности полипептида или молекулы нуклеиновой кислоты, не являющимся нативными по отношению к клетке, в которой они экспрессируются, (b) полипептиду, или молекуле нуклеиновой кислоты, или части полипептида или молекулы нуклеиновой кислоты, которые были изменены или подвергнуты мутации по сравнению с их нативным состоянием, или (c) полипептиду или молекуле нуклеиновой кислоты с измененной экспрессией по сравнению с нативными уровнями экспрессии при аналогичных условиях. Например, гетерологичная регуляторная последовательность (например, промотор, энхансер) может применяться для регуляции экспрессии гена или молекулы нуклеиновой кислоты таким образом, чтобы она отличалась от экспрессии гена или молекулы нуклеиновой кислоты, обычно наблюдающейся в природе. В другом примере гетерологичные домен полипептида или последовательность нуклеиновой кислоты (например, ДНК-связывающий домен полипептида или нуклеиновая кислота, кодирующая ДНК-связывающий домен полипептида) могут быть расположены относительно других доменов или могут представлять собой отличную последовательность или происходить из отличного источника по сравнению с другими доменами или частями полипептида или кодирующей его нуклеиновой кислоты. В определенных вариантах осуществления гетерологичная молекула нуклеиновой кислоты может существовать в нативном геноме клетки-хозяина, но может характеризоваться измененным уровнем экспрессии, или иметь отличную последовательность, или и то, и другое. В других вариантах осуществления гетерологичные молекулы нуклеиновых кислот могут не являться эндогенными по отношению к клетке-хозяину или геному хозяина, но вместо этого быть введенными в клетку-хозяина с помощью трансформации (например, трансфекции, электропорации), при этом добавленная молекула может интегрироваться в геном хозяина или может существовать в виде внехромосомного генетического материала как транзиентно (например, мРНК), так и полустабильно в течение более чем одного поколения (например, эписомный вирусный вектор, плазмида или другой самореплицирующийся вектор).

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.

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 однонитевой, 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) выпетливаний. В некоторых вариантах осуществления второй сегмент содержит один или несколько неспаренных нуклеотидов, например, образующих петлю. В некоторых вариантах осуществления третий сегмент содержит один или несколько неспаренных нуклеотидов, например, образующих петлю.

Последовательность структуры "стебель-петля". Используемый в данном документе термин "последовательность структуры "стебель-петля" относится к последовательности нуклеиновой кислоты (например, последовательности РНК), характеризующейся самокомплементарностью, достаточной для образования структуры "стебель-петля", например, имеющей "стебель", содержащий по меньшей мере две (например, 3, 4, 5, 6, 7, 8, 9 или 10) пары оснований, и "петлю", содержащую по меньшей мере три (например, четыре) пары оснований. "Стебель" может содержать ошибки спаривания или выпетливания.

Тканеспецифическая(тканеспецифические) последовательность(последовательности), контролирующие 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 контролирующий 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.

КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

Файл патента или заявки на патент содержит по меньшей мере одно графическое изображение, выполненное в цвете. Копии данного патента или публикации заявки на патент с цветным(цветными) графическим(графическими) изображением(изображениями) будут предоставлены Ведомством по запросу и при уплате соответствующей пошлины.

Фигура 1. Линкерная область на С-конце ДНК-связывающего домена R2Tg может быть усечена и модифицирована. Были сконструированы делеции в природном линкере от домена myb в A или B до положений 1 или 2 наряду с заменой с помощью синтетических линкеров 3GS или XTEN (A). Эффективность интеграции измеряли в клетках HEK293T с помощью ddPCR (B).

Фигура 2. "Посадочные площадки", предназначенные для тестирования мутаций сайтов-мишеней R2Tg Gene Writer.

Фигура 3a. Анализ методом ddPCR, в котором измеряется процентная доля интеграций для всех лентивирусных интегрированных "посадочных площадок" на клетку.

Фигура 3b. Секвенирование ампликонов и анализ методом NGS вставок/делеций, присутствующих в сайтах "посадочных площадок".

Фигура 4. Замена ДНК-связывающего домена ретротранспозазы Gene Writer на ZFP для AAVS1.

Фигура 5. Замена ДНК-связывающего домена ретротранспозазы Gene Writer с активным доменом EN или без него на Cas9 или никазу Cas9 (*=мутантный)

Фигура 6. Слияние AAVS1 ZFP с ретротранспозазой Gene Writer с функциональным ДНК-связывающим доменом или без него.

Фигура 7. Схема внесения однонитевого разреза во вторую нить. (A) Никаза Cas9 слита с белком Gene Writer. Белок Gene Writer вводит однонитевой разрыв в нить ДНК посредством своего EN-домена (обозначается как *), и слитая никаза Cas9 вводит однонитевые разрывы либо в верхнюю, либо в нижнюю нити ДНК (обозначается как X). (B) Gene Writer нацеливается на ДНК посредством своего ДНК-связывающего домена и вводит однонитевой разрыв в ДНК с помощью своего EN-домена (*). Затем никазу Cas9 применяют для создания второго однонитевого разрыва (X) на верхней либо на нижней нитях выше или ниже однонитевого разрыва, введенного с помощью EN.

Фигура 8. Схематическое изображение слияний никаза Cas9-GeneWriter. (A) Схематическое изображение никазы Cas9, слитой с белком Gene Writer. (B) Схематическое изображение gRNA, удлиненной на 3'-конце.

Фигура 9. Схематическое изображение слияний никаза Cas9-GeneWriter. (A) Схематическое изображение никазы Cas9, слитой с белком Gene Writer. (B) Схематическое изображение донорного трансгена, фланкированного UTR, и гомологии с сайтом разреза.

Фигура 10. 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 11. Схематическое изображение мРНК, кодирующей Gene Writer (A). Нативные нетранслируемые области (UTR) были заменены 5'- и 3'-UTR, оптимизированными для экспрессии белка (показаны как 5'-UTRexp и 3'-UTRexp). Экспрессию белка Gene Writer анализировали с помощью анализа HiBit путем исследования экспрессии его HiBit-метки (B).

Фигура 12. 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 13. Доставка системы Gene Writer с применением мРНК, кодирующей полипептид, и плазмидной ДНК, кодирующей РНК-матрицу для ретротранспозиции.

Фигура 14. Схемы иллюстративных стратегий конструирования 5'-UTR. HA=плечо гомологии; K=последовательность Козак; pA=поли(A)-сигнал; AMa=*A. maritima*; Rx=другие виды ретротранспозонов.

Фигура 15. Возможное расположение интрона (или интронов) в РНК-матрице. Интроны показаны изогнутыми линиями. 5'-HA: 5'-плечо гомологии; 3'-HA: 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(A)-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 16. 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(A)-хвостом или без них. Соотношение мРНК Gene Writer и РНК трансгена составляло 1:1.

Фигура 17. Интеграция в геном, выявленная 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 18. Модули, содержащие типичную РНК-матрицу Gene Writer, где отдельные модули можно комбинировать, переупорядочивать и/или исключать для получения матрицы Gene Writer. A=5'-плечо гомологии; B=рибозим; C=5'-UTR; D=последовательность гетерологичного объекта; E=3'-UTR; F=3'-плечо гомологии.

Фигура 19. Модули, содержащие типичную РНК-матрицу Gene Writer, где отдельные модули можно комбинировать, переупорядочивать и/или исключать для получения матрицы Gene Writer. A=5'-плечо гомологии; B=рибозим; C=5'-UTR; D=последовательность гетерологичного объекта; E=3'-UTR; F=3'-плечо гомологии.

Фигура 20. Схема конструкции плазмид с управляющим элементом и трансгеном. Плечи гомологии (HA) и спейсерные последовательности являются вариабельными в данном наборе экспериментов.

Фигура 21. Эффективность интеграции на 3'- или 5'-конце трансгена в тестируемых конструкциях при измерении с помощью цифровой капельной ПЦР. Каждая точка представляет повторность эксперимента. Столбцы представляют собой среднее значение экспериментов в двух повторностях. (A, B) Эффективность интеграции, измеренная в 3'-месте соединения между трансгеном и рДНК хозяина. (C, D) Эффективность интеграции, измеренная в 5'-месте соединения.

Фигура 22. Иллюстративный пример конструкции со сдвигом гомологии, тестируемой в отношении +/-3 п. о. Красным цветом указана гомология с 5'-стороны от сайта однонитевого разрыва дикого типа (WT), и синим цветом указана гомология с 3'-стороны от сайта однонитевого разрыва. В конструкциях со сдвигом на 3'-конце (+) 3'-

гомология начинается намного ниже однонитевого разрыва. Конструкции со сдвигом на 5'-конце (-) включают гомологию от 5'-конца однонитевого разрыва до 3'-плеча гомологии.

Фигура 23. Интеграция на 3'-конце является результатом сдвига 3'-плеча гомологии трансгена. Каждая точка данных представляет повторность, а столбец представляет среднее значение для двух повторностей.

Фигура 24. (A) Временная шкала эксперимента. (B) Схема конфигураций конструкций R2Tg и трансгена. (C) Вестерн-блоттинг против Rad51 демонстрирует уменьшение уровня экспрессии белка Rad51 в день 3.

Фигура 25. 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'- (A) или 5'- (B) месте соединения применяли для оценки эффективности интеграции в день 3.

Фигура 26. (A) Карта последовательности рибозима элемента R2 из *Taeniopygia guttata* (R2Tg) в окружении модулей трансгенной молекулы РНК Gene Writer. Признаки рибозима обозначаются следующим образом: P - область спаренных оснований; P'- нить, комплементарная области спаренных оснований; L - петля в конце P-области; J - нуклеотиды, соединяющие области спаренных оснований. На данной фигуре раскрыта SEQ ID NO: 1592. (B) Предсказание вторичной структуры рибозима R2Tg. Заштрихованным прямоугольником указано предположительное каталитическое положение, которое можно использовать для инактивации рибозима. На данной фигуре раскрыта SEQ ID NO: 1592.

Фигура 27. Карта последовательности рибозима элемента R2 из *Taeniopygia guttata* (R2Tg) в окружении модулей трансгенной молекулы РНК Gene Writer. Признаки рибозима обозначаются следующим образом: P - область спаренных оснований; P'- нить, комплементарная области спаренных оснований; L - петля в конце P-области; J - нуклеотиды, соединяющие области спаренных оснований. На данной фигуре раскрыта SEQ ID NO: 1592.

Фигура 28. Предсказание вторичной структуры рибозима элемента R2 из Taeniopygia guttata. На данной фигуре раскрыта SEQ ID NO: 1592.

Фигуры 29A и 29B 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 GeneWriter может не иметь РНК-связывающего домена или может иметь отдельные домены, которые выполняют функции нескольких доменов, например, домен Cas9 для связывания ДНК и эндонуклеазной активности. Иллюстративные домены, которые могут быть включены в полипептид Gene Writer, включают ДНК-связывающие домены (например, содержащие ДНК-связывающий домен элемента последовательности, указанной в любой из таблиц 1 или 3, или домен, указанный в таблице 2; "цинковый палец"; домен TAL; Cas9; dCas9; никазу Cas9; фактор транскрипции или мегануклеазу), РНК-связывающие домены (например, включающие РНК-связывающий домен белка B-бокса, белка оболочки MS2,

dCas или элемент последовательности, указанной в любой из таблиц 1 или 3, или домен, указанный в таблице 2), домены обратной транскриптазы (например, содержащие домен обратной транскриптазы элемента последовательности, указанной в любой из таблиц 1 или 3, или домен, указанный в таблице 2) и/или эндонуклеазный домен (например, содержащий эндонуклеазный домен элемента последовательности, указанной в любой из таблиц 1 или 3, или домен, указанный в таблице 2; Cas9; никазу Cas9; рестрикционный фермент (например, рестрикционный фермент 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.

На фигурах 30A и B проиллюстрированы мутации ДНК-связывающих мотивов в полипептиде Gene Writer, которые ингибируют интеграцию в нативный сайт. На фигуре 30A показана общая доменная структура ретротранспозазы R2Tg (вверху), включающая ДНК-связывающий домен, содержащий несколько предсказанных ДНК-связывающих элементов (внизу). Два мотива "цинковые пальцы" и мотив c-myb, указанные в белке, были подвергнуты мутации в соответствии с примером 30. На фигуре 30B показано, что активность интеграции для мутантных форм доменов ZF1, ZF2 и c-myb оценивали в клетках HEK293T посредством анализа частоты интеграции в нативный сайт рДНК с применением ddPCR. Каждую отдельную мутантную форму, а также тройную мутантную форму сравнивали с ферментом дикого типа (положительный контроль) и ферментом, инактивированным эндонуклеазой (отрицательный контроль). Данные указывают средние значения для двух повторностей.

Условные обозначения к фигурам: ZF=цинковый палец; myb=c-myb-подобный ДНК-связывающий мотив; RBD=РНК-связывающий домен; RT=домен обратной транскриптазы; EN=эндонуклеазный домен; *=мутантный домен; CNV/геном=среднее количество копий интегрированной ДНК на копию генома.

На фигурах 31A и 31B показано, что сайт расщепления эндонуклеазой для ретротранспозазы может быть выявлен с помощью сигнатуры вставок/делеций. На фигуре 31A показаны предполагаемые места связывания и расщепления в сайте-мишени ретротранспозазы R2Tg. На фигуре 31B показано, что сайт расщепления ретротранспозазы R2Tg был валидирован с помощью анализа изменений генома, обусловленных эндонуклеазной активностью. Плазмидную ДНК, кодирующую ретротранспозазу R2Tg, вводили путем нуклеофекции в клетки U2OS, и через три дня собирали геномную ДНК. Ампликоны сайта-мишени были созданы с применением сайт-специфических праймеров и секвенированы для определения местоположения изменений генома, указывающих на эндонуклеазную активность. Здесь показан график, изображающий частоту вставок (кружки) и делеций (треугольники) на нуклеотид последовательности (ось x). Пик сигнала вставки (горизонтальная линия под фигурой) был локализован в предсказанном динуклеотиде GG.

Условные обозначения к фигурам: ZF="цинковый палец"; myb=c-myb-подобный ДНК-связывающий мотив

На фигурах 32A и B показано определение детерминант последовательности эндонуклеазной активности ретротранспозазы посредством схематического представления изображения "посадочной площадки". На фигуре 33A показано, что лентивирусный вектор экспрессии применяли для клонирования "посадочных площадок", содержащих нативный сайт-мишень ретротранспозазы R2 или сайты, содержащие мутации по сравнению с нативным сайтом. Лентивирусные конструкции были упакованы и использованы для трансдукции клеток U2OS для создания линий клеток с интегрированными в геном "посадочными площадками". "Посадочная площадка" дополнительно содержала кассету с репортерным геном зеленого флуоресцентного белка (GFP) для определения титра. На фигуре 33B показаны последовательности "посадочных площадок", содержащие сайт R2 дикого типа или его мутантные варианты. "Посадочную площадку" нативной последовательности рДНК, содержащую немодифицированную последовательность рДНК (WT_R2Tg), использовали в качестве положительного контроля. Показана серия из 16 "посадочных площадок" с мутантными областями, отмеченным темно-серым цветом, и сайтом расщепления GG, отмеченным светло-серым цветом (слева). График (справа) использовали для визуализации величины изменения каждого сайта-мишени в зависимости от эндонуклеазной активности фермента. Было обнаружено, что мутация динуклеотида AA, прилегающего к динуклеотидному сайту расщепления GG, существенно ухудшает эндонуклеазную активность, поэтому мотив AAGG важен для эндонуклеазной активности R2Tg.

На фигуре 33 показан обзор изображения "посадочной площадки" для перенацеливания полипептида Gene Writer. Схема библиотеки "посадочных площадок", созданной для анализа последовательностей, распознаваемых при перенацеливании R2Tg. Сайт связывания AAVS1-ZF (темно-серый и помеченный AAVS1) использовали для перенацеливания в качестве ДНК-связывающего мотива, и все "посадочные площадки" были встроены в окружение геномной последовательности AAVS1 человека. Последовательность рДНК (черным цветом) добавляли к последовательности AAVS1 различными способами: (категория 1) разная длина последовательности рДНК, (категория 2) разные расстояния между сайтом связывания ZF AAVS1 и последовательностью рДНК, (категория 3) разная ориентация последовательности рДНК относительно сайта AAVS1. Категории 1, 2 и 3 были исследованы комбинаторно, в результате чего были получены "посадочные площадки" с различной длиной последовательностей рДНК и различными расстояниями и ориентациями относительно сайта связывания ZF AAVS. Минимальная последовательность AAGG для расщепления R2Tg сохранялась во всех "посадочных площадках" (черный прямоугольник с белой заливкой). Каждая "посадочная площадка" была разработана с уникальным штрих-кодом на 3'-конце последовательности, чтобы обеспечить вычислительное извлечение и анализ последовательностей "посадочных площадок" из пула.

На фигуре 34 представлено определение на основе секвенирования представленности "посадочных площадок" в пуле U2OS. Пул "посадочных площадок" клеток U2OS был секвенирован и проанализирован для определения представленности штрих-кодов. Примерно 94% "посадочных площадок" были представлены по меньшей мере 10000 ридами (горизонтальная черная полоса). Ось x указывает на принадлежность "посадочной площадки", и ось y показывает общее количество ридов для этого штрих-кода.

На фигурах 35 A и B раскрыто создание сигнатур вставок/делеций в библиотеке "посадочных площадок", позволяющее проводить скрининг химерных полипептидов Gene Writer. На фигуре 35A показана библиотека "посадочных площадок", содержащая различные композиции последовательностей-мишеней рДНК AAVS1 и R2, обработанных полноразмерной ретротранспозазой R2Tg, слитой с "цинковым пальцем", для распознавания последовательности AAVS1. Выполняли секвенирование ампликона, и частоту вставок в сайте-мишени GG (ось Y) наносили на график для каждой "посадочной площадки" (ось X). Иллюстративное количество 230 "посадочных площадок" показано по оси X. Указаны положительные контроли, содержащие 200 нуклеотидов последовательности рДНК, которые демонстрировали ожидаемые сигнатуры вставок в сайте расщепления GG. Отрицательный контроль, в котором отсутствовала какая-либо последовательность рДНК, не содержал никаких вставок. Длины последовательности рДНК, содержащейся в "посадочных площадках", где были обнаружены сигнатуры вставок, указаны и соответствуют 44, 64 и 84 нуклеотидам. На фигуре 35B показано иллюстративное представление конфигураций "посадочных площадок", которые, как обнаружено, содержат сигнатуры эндонуклеазной активности.

На фигурах 36 A и B раскрыто создание сигнатур вставок/делеций в библиотеке "посадочных площадок", позволяющее проводить скрининг химерных полипептидов Gene Writer. На фигуре 36A показана библиотека "посадочных площадок", содержащая различные композиции последовательностей-мишеней рДНК AAVS1 и R2, обработанных полноразмерной ретротранспозазой R2Tg, слитой с "цинковым пальцем", для распознавания последовательности AAVS1. Выполняли секвенирование ампликона, и частоту вставок в сайте-мишени GG (ось Y) наносили на график для каждой "посадочной площадки" (ось X). Иллюстративное количество 230 "посадочных площадок" показано по оси X. Отрицательный контроль, в котором отсутствовала какая-либо последовательность рДНК, не содержал никаких вставок. Длины последовательности рДНК, содержащейся в "посадочных площадках", где были обнаружены сигнатуры вставок, указаны и соответствуют 44, 64 и 84 нуклеотидам. На фигуре 36B показано иллюстративное представление конфигураций "посадочных площадок", которые, как обнаружено, содержат сигнатуры эндонуклеазной активности.

На фигурах 37 A и B описан анализ люциферазной активности для первичных клеток. LNP, составленные в соответствии с примером 38, анализировали в отношении доставки груза в первичные гепатоциты человека (A) и мыши (B) в соответствии с примером 39. В результате анализа люциферазы было выявлено дозозависимую

люциферазную активность в лизатах клеток, что указывает на успешную доставку РНК в клетки и экспрессию люциферазы светлячка из мРНК-груза.

На фигуре 38 показана LNP-опосредованная доставка РНК-груза в печень мыши. LNP, содержащие мРНК люциферазы светлячка, составляли и доставляли мышам iv, а образцы печени собирали и анализировали в отношении люциферазной активности через 6, 24 и 48 часов после введения. Активность репортерного гена в различных составах соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003. Экспрессия РНК была транзиентной, и уровни ферментов возвращались к уровню, близкому к фоновому уровню при обработке средой-носителем, через 48 часов после введения.

Фигура 39. 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 никазы 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;** SEQ ID NO: 468)), который, как было обнаружено, значительно улучшает экспрессию слитого полипептида. Номера образцов соответствуют идентификаторам линкерных последовательностей в таблице 42.

ПОДРОБНОЕ ОПИСАНИЕ

Настоящее изобретение относится к композициям, системам и способам для нацеливания на последовательность ДНК, ее редактирования, модификации или произведения манипуляций с ней (например, вставки гетерологичного объекта в виде последовательности ДНК в сайт-мишень генома млекопитающего) в одном или нескольких местоположениях в последовательности ДНК в клетке, ткани или у субъекта, например, *in vivo* или *in vitro*. Объект в виде последовательности ДНК может включать в себя, например, кодирующую последовательность, регуляторную последовательность, единицу экспрессии гена.

Более конкретно, в настоящем изобретении предусмотрены системы на основе ретротранспозона для вставки последовательности, представляющей интерес, в геном. Настоящее изобретение частично основывается на биоинформатическом анализе для идентификации последовательностей ретротранспозазы и ассоциированных с ней 5'-UTR и 3'-UTR из разнообразных организмов (см. таблицу 3).

Редакторы генома Gene Writer™

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 вариабельную 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.

Обратная транскрипция с помощью ретротранспозонов, не содержащих LTR, происходит посредством уникального процесса, описанного как обратная транскрипция с праймированием мишенью (Luan et al. Cell 72, 595-605 (1993)). Для инициации интеграции эндонуклеазный домен ретротранспозазы создает первый однонитевой разрыв, высвобождая свободный 3'-OH. РНК ретротранспозона, связанная ретротранспозазой с использованием структурных признаков на 3'-конце, затем праймируется сайтом-мишенью с полимеризацией на свободном 3'-OH и применяется в качестве матрицы для обратной транскрипции. В некоторых системах второй однонитевой разрыв намечен во второй нити ДНК, и новый свободный 3'-OH применяется для инициации синтеза второй нити. Некоторые ретротранспозоны, не содержащие LTR, например, R2, как полагают, дополнительно требуют взаимодействия со второй единицей ретротранспозазы на 5'-конце ретротранспозона РНК для этого второго однонитевого разрыва, который активируется при высвобождении 5'-конца (Craig, Mobile DNA III, ASM, ed. 3 (2105)).

Как описано в данном документе, элементы таких не содержащих LTR ретротранспозонов могут быть разбиты на функциональные модули и/или модифицированы длянацеливания на последовательность ДНК-мишени, ее редактирования, модификации или произведения манипуляций с ней, например, для вставки объекта в виде (например, гетерологичной) последовательности нуклеиновой кислоты в геном-мишень, например, геном млекопитающего, посредством обратной транскрипции. Такие разбитые на модули и модифицированные нуклеиновые кислоты, композиции на основе полипептидов и системы описаны в данном документе и называются редакторами генов Gene Writer™. Система редактирования генов Gene Writer™ содержит: (A) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и либо (x) эндонуклеазный домен, который обладает функциональным свойством связывания с ДНК, либо (y) эндонуклеазный домен и отдельный ДНК-связывающий домен; и (B) РНК-матрицу, содержащую (i)

последовательность, которая связывает полипептид, и (ii) гетерологичную вставочную последовательность. Например, белок редактора генома Gene Writer может содержать ДНК-связывающий домен, домен обратной транскриптазы и эндонуклеазный домен. В других вариантах осуществления белок редактора генома Gene Writer может содержать домен обратной транскриптазы и эндонуклеазный домен. В определенных вариантах осуществления элементы полипептида редактора генов Gene Writer™ могут быть получены из последовательностей 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 некоторых вариантах осуществления редактор генома Gene Writer объединен со вторым полипептидом. В некоторых вариантах осуществления второй полипептид получен из не содержащего LTR ретротранспозона APE-типа. В некоторых вариантах осуществления второй полипептид содержит мотив, подобный "цинковому суставу". В некоторых вариантах осуществления второй полипептид является гомологом белков Gag. В некоторых вариантах осуществления второй полипептид обладает специфической активностью связывания в отношении РНК-матрицы. В некоторых вариантах осуществления второй полипептид способствует локализации РНК-матрицы в ядре.

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.

Полипептидный компонент системы редактирования генов Gene Writer

Домен RT

В определенных аспектах настоящего изобретения в основе домена обратной транскриптазы системы Gene Writer лежит домен обратной транскриптазы не содержащего LTR ретротранспозона APE-типа или RLE-типа. 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 некоторых 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 которой указана в таблице 1 или таблице 3, или с пептидом, содержащим домен обратной транскриптазы, указанным 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 получен из RT ретровируса, например, RT HIV-1, RT вируса лейкоза мышей Молони (MMLV), RT вируса миелобластоза 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. *PHK* 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 обратная транскриптаза 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_pepA17), cd01645 (RT_Rtv), cd01646 (RT_бактериальный_ретрон_I), cd01647 (RT_LTR), cd01648 (TERT), cd01650 (RT_nLTR_подобные), cd01651 (RT_интрон_G2), cd01699 (РНК_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 (например, содержащих домены RT) было описано без ограничения теми, которые используются прокариотами (Zimmerly et al. *Microbiol*

Spectr 3(2):MDNA3-0058-2014 (2015); Lampson B.C. (2007) Prokaryotic Reverse Transcriptases. B: Polaina J., MacCabe A.P. (eds) Industrial Enzymes. Springer, Dordrecht), вирусами (Herschhorn et al. Cell Mol Life Sci 67(16):2717-2747 (2010); Menéndez-Arias et al. Virus Res 234:153-176 (2017)) и мобильными элементами (Eickbush et al. Virus Res 134(1-2):221-234 (2008); Craig et al. Mobile DNA III 3rd Ed. DOI:10.1128/9781555819217 (2015)), каждый из который включен в данный документ посредством ссылки.

В некоторых вариантах осуществления домен RT демонстрирует повышенную строгость инициации обратной транскрипции с праймированием мишенью (TPRT), например, по сравнению с эндогенным доменом RT. В некоторых вариантах осуществления домен RT инициирует TPRT, когда 3 нуклеотида в сайте-мишени непосредственно перед однонитевым разрывом первой нити, например, геномная ДНК, праймирующая РНК-матрицу, характеризуются по меньшей мере 66% или 100% комплементарностью по отношению к 3 гомологичным нуклеотидам в РНК-матрице. В некоторых вариантах осуществления домен RT инициирует TPRT при наличии несовпадения менее 5 нуклеотидов (например, несовпадения менее 1, 2, 3, 4 или 5 нуклеотидов) между гомологичной РНК-матрицей и ДНК-мишенью, праймирующей обратную транскрипцию. В некоторых вариантах осуществления домен RT модифицирован таким образом, что строгость в отношении несовпадений при праймировании реакции TPRT увеличивается, например, где домен RT не допускает каких-либо несовпадений или допускает меньшее количество несовпадений в области праймирования по сравнению с доменом RT дикого типа (например, немодифицированным). В некоторых вариантах осуществления домен RT содержит домен RT HIV-1. В вариантах осуществления домен RT HIV-1 инициирует более низкие уровни синтеза даже при несовпадении трех нуклеотидов по сравнению с альтернативным доменом RT (например, как описано в Jamburuthugoda и Eickbush J Mol Biol 407(5):661-672 (2011); включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен RT образует димер (например, гетеродимер или гомодимер). В некоторых вариантах осуществления домен RT является мономерным. В некоторых вариантах осуществления домен RT, например, домен RT ретровируса, в естественных условиях функционирует как мономер или как димер (например, гетеродимер или гомодимер). В некоторых вариантах осуществления домен RT в естественных условиях функционирует как мономер, например, получен из вируса, в котором он функционирует как мономер. Иллюстративные мономерные домены RT, их вирусные источники и ассоциированные с ними сигнатуры RT можно найти в таблице 30, при этом описание сигнатур доменов находится в таблице 32. В некоторых вариантах осуществления домен RT системы, описанной в данном документе, содержит аминокислотную последовательность из таблицы 30 или ее функциональный фрагмент или вариант или последовательность, характеризующуюся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ней. В вариантах осуществления домен RT выбран из домена RT вируса лейкоза мышей (MLV; иногда называемого как MoMLV) (например,

P03355), эндогенного ретровируса свиньи (PERV) (например, UniProt Q4VFZ2), вируса опухоли молочной железы мышей (MMTV) (например, UniProt P03365), вируса обезьян Мейзона-Пфайзера (MPMV) (например, UniProt P07572), вируса лейкоза крупного рогатого скота (BLV) (например, UniProt P03361), вируса Т-клеточного лейкоза человека 1 типа (HTLV-1) (например, UniProt P03362), вируса пенистости человека (HFV) (например, UniProt P14350), вируса пенистости обезьян (SFV) (например, UniProt P23074) или вируса пенистости/синцитиального вируса крупного рогатого скота (BFV/BSV) (например, UniProt O41894) или его функционального фрагмента или варианта (например, аминокислотной последовательности, характеризующейся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ним). В некоторых вариантах осуществления домен RT является димерным в своем естественном функционировании. Иллюстративные димерные домены RT, их вирусные источники и ассоциированные с ними сигнатуры RT можно найти в таблице 31, при этом описание сигнатур доменов находится в таблице 32. В некоторых вариантах осуществления домен RT системы, описанной в данном документе, содержит аминокислотную последовательность из таблицы 31 или ее функциональный фрагмент или вариант или последовательность, характеризующуюся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ней. В некоторых вариантах осуществления домен RT получен из вируса, в котором он функционирует как димер. В вариантах осуществления домен RT выбран из домена RT вируса саркомы/лейкоза птиц (ASLV) (например, UniProt A0A142BKH1), вируса саркомы Рауса (RSV) (например, UniProt P03354), вируса миелобластоза птиц (AMV) (например, UniProt Q83133), вируса иммунодефицита человека I типа (HIV-1) (например, UniProt P03369), вируса иммунодефицита человека II типа (HIV-2) (например, UniProt P15833), вируса иммунодефицита обезьян (SIV) (например, UniProt P05896), вируса иммунодефицита крупного рогатого скота (BIV) (например, UniProt P19560), вируса инфекционной анемии лошадей (EIAV) (например, UniProt P03371) или вируса иммунодефицита кошек (FIV) (например, UniProt P16088) (Herschhorn and Hizi Cell Mol Life Sci 67(16):2717-2747 (2010)) или его функционального фрагмента или варианта (например, аминокислотной последовательности, характеризующейся по меньшей мере 70%, 80%, 90%, 95% или 99% идентичностью с ним). Природные гетеродимерные домены RT могут в некоторых вариантах осуществления также функционировать как гомодимеры. В некоторых вариантах осуществления димерные домены RT экспрессируются в виде слитых белков, например, в виде гомодимерных слитых белков или гетеродимерных слитых белков. В некоторых вариантах осуществления функцию RT в системе выполняют несколько доменов RT (например, как описано в данном документе). В дополнительных вариантах осуществления несколько доменов RT слиты или разделены, например, могут находиться в одном и том же полипептиде или в разных полипептидах.

В некоторых вариантах осуществления Gene Writer, описанный в данном документе, содержит интегразный домен, например, где интегразный домен может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит интегразный домен. В некоторых вариантах осуществления в домене

RT (например, описанном в данном документе) отсутствует интеразный домен или он содержит интеразный домен, который был инактивирован путем мутации или подвергнут делеции. В некоторых вариантах осуществления Gene Writer, описанный в данном документе, содержит домен РНКазы H, например, где домен РНКазы H может быть частью домена RT. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы H, например, эндогенный домен РНКазы H или гетерологичный домен РНКазы H. В некоторых вариантах осуществления в домене RT (например, описанном в данном документе) отсутствует домен РНКазы H. В некоторых вариантах осуществления домен RT (например, описанный в данном документе) содержит домен РНКазы H, который был добавлен, подвергнут делеции, мутации или заменен гетерологичным доменом РНКазы H. В некоторых вариантах осуществления в результате мутации домена РНКазы H образуется полипептид, демонстрирующий более низкую активность РНКазы, например, как определено способами, описанными в Kotewicz et al. Nucleic Acids Res 16(1):265-277 (1988) (включенном в данный документ посредством ссылки во всей своей полноте), например, ниже на по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% или 90% по сравнению с в остальном аналогичным доменом без мутации. В некоторых вариантах осуществления активность РНКазы H устранена.

В некоторых вариантах осуществления домен RT подвергнут мутации для повышения точности по сравнению с в остальном аналогичным доменом без мутации. Например, в некоторых вариантах осуществления мотив YADD (SEQ ID NO: 1547) или YMDD (SEQ ID NO: 1548) в домене RT (например, в обратной транскриптазе) заменен на YVDD (SEQ ID NO: 1549). В вариантах осуществления замена на YADD (SEQ ID NO: 1547), или YMDD (SEQ ID NO: 1548), или YVDD (SEQ ID NO: 1549) 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 домены обратной транскриптазы являются модифицированными, например, посредством сайт-специфической мутации. В некоторых вариантах осуществления домены обратной транскриптазы содержат ряд аминокислотных 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. В вариантах осуществления домен обратной транскриптазы сконструирован таким образом, что он связывается с гетерологичной РНК-матрицей.

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 сформированную ДНК-полимеразу. Также включены ассоциированные сигнатуры RT из баз данных InterPro, Pfam и CD. Хотя сформированная 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	Интрон группы II	CBK92290.1	D4JMT6	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC (SEQ ID NO: 1550)	IPR000477, PF00078, cd01651
TGIRT, trt	Интрон группы II	AAT72329.1	Q6DKY2	MALLERILADRNLITALKRVEANQGAPGIGDVSTDQLRDIYRAHWSTIRAQLLAGTYRPAPVRRVGIPKGPGGTRQLGITPVVDRLIQQIALQELTPIFDPDFSPSSFGFRPGRNAHDAVRQAQGYIQEYGRYVVDMDLKEFFDRVNHDLIMSRVARKVDKKRVLKLIRYALQAGVMIEGVKVQTEEGTQPGGPLSPLLANILLDDLDKELEKRGLKFCYRADDCNIYVSKLRAGQRVKQSIQRFLEKTLKLKVNEEKSVADRPWKRAFGLFSFTPERKARIRLAPRSIQRLKQRIRQLTNPNWSISMPREIHRVNQYVGMWIGYFR	IPR000477, PF00078, cd01651

				LVTEPSVLQTIEGWIRRRLRLCWQL QWKRVRTRIRELRALGLKETAVMEI ANRTKGAWRTTKPQTLHQALGKYT WTAQGLKTSLQRYFELRQG (SEQ ID NO: 1551)	
LtrA	Интрон группы II	AAB06503.1	P0A3U0	MKPTMAILERISKNSQENIDEVFTRL YRYLLRPDIYYVAYQNLYSNKGAST KGILDDTADGFSEEKIKKIIQSLKDGT YYPQPVRRMYIAKKNSKKMRPLGIP TFTDKLIQEAVRIILESIYEPVFEDVS HGFRPQRSCHTALKTIKREFGGARW FVEGDIKGCFDNIDHVTLIGLINLKIK DMKMSQLIYKFLKAGYLENWQYHK TYSGTPQGGILSPLLANIYLHELDKF VLQLKMKFDRESPERITPEYRELHNE IKRISHRLKKLEGEEKAKVLLEYQEK RKRLPTLPCTSQTNKVLKYVRYADD FIISVKGSKEDCQWIKEQLKLFIHNK LKMELSEEKTLITHSSQPARFLGYDI RVRRSGTIKRSGKVKKRTLNGSVEL LIPLQDKIRQFIFDKKIAIQKKDSSWF PVHRKYLIRSTDLEIITIYNSELRGICN YYGLASNFNQLNYFAYLMEYSCLK TIASKHKGTLSKTISMFKDGSGSWGI PYEIKQGKQRRYFANFSECKSPYQFT DEISQAPVLYGYARNTLENRLKAKC CELCGTSDENTSYEIHHVNKVKNLK GKEKWEMAMIAKQRKTLVVCFHCH RHVIHKHK (SEQ ID NO: 1552)	IPR000477, PF00078, cd01651
R2Bm	Ретротранспозон, не содерж	AAB59214.1	V9H052	MMASTALSLMGRCNPDGCTRGKHV TAAPMDGPRGPSSLAGTFGWGLAIP AGEPCGRVCSPATVGFFPVAKKSNK ENRPEASGLPLESERTGDNPTVRGSA GADPVGQDAPGWTCQFCERTFSTNR	IPR000477, PF00078, cd01650

	ащий LTR			GLGVHKRRAHPVETNTDAAPMMVK RRWHGEEIDLLARTEARLLAERGQC SGGDLFGALPGFGRTLEAIKGQRRR EPYRALVQAHLARFGSQPGPSSGGC SAEPDFRRASGAEEAGEERCAEDAA AYDPSAVGQMSPDAARVLSELLEGA GRRRACRAMRPKTAGRRNDLHDDR TASAHKTSRQKRRAEYARVQELYK KCRSRAAAEVIDGACGGVGHSLEE METYWRPILERVSDAPGPTPEALHA LGRAEWHGGNRDYTQLWKPISVEEI KASRFDWRTSPGPDGIRSGQWRAVP VHLKAEMFNAWMARGEIPEILRQCR TVFVPKVERPGGPGEYRPISIASIPLR HFHSILARRLLACCPPDARQRGFICA DGTLENSAVLDAVLGDSRKKLRECH VAVLDFAKAFDTVSHEALVELLRLR GMPEQFCGYIAHLYDTASTTLAVNN EMSSPVKVGRGVRQGDPLSPILFNV VMDLILASLPERVGYRLEMELVSAL AYADDLVLLAGSKVGMQESISAVD CVGRQMGLRLNCRKSAVLSMIPDG HRKKHHYLTERTFNIGGKPLRQVSC VERWRYLGVDFEASGCVTLEHSISS ALNNISRAPLKPQQRLEILRAHLIPRF QHGFVLGNISDDRLRMLDVQIRKAV GQWLRLPADVPKAYYHAAVQDGG LAIPSVRATIPDLIVRRFGGLDSSPWS VARAAAKSDKIRKKLRWAWKQLRR FSRVDSTTQRPSVRLFWREHLHASV DGRELRESTRTPTSTKWIRERCAQIT GRDFVQFVHTHINALPSRIRGSRGRR GGGESSLTCRAGCKVRETTAHILQQ CHRTHGGRILRHNKIVSFVAKAMEE NKWTVELEPRLRTSVGLRKPDIIASR	

				DGVGVIVDVQVVSGQRSLDELHREK RNKYGNHGELVELVAGRLGLPKAE CVRATSCTISWRGVWSLTSYKELRSI IGLREPTLQIVPILALRGSHMNWTRF NQMTSVMGGGVG (SEQ ID NO: 1553)	
LIN E-1	Ретротранспозон, не содержащий LTR	AAC51271.1	O00370	MTGSNSHITILTLNVNGLNSPIKRHR LASWIKSQDPSVCCIQETHLTCRDTH RLKIKGWRKIYQANGKQKKAGVAIL VSDKTDFKPTKIKRDKEGHYIMVKG SIQQEELTILNIYAPNTGAPRFIKQVL SDLQRDLDSHTLIMGDFNTPLSILDR STRQKVNKDTQELNSALHQTDLIDI YRTLHPKSTEYTFFSAPHHTYSKIDH IVGSKALLSKCKRTEIITNYLSDHSAI KLELRIKNLTQSRSTTWKLNNLLLN DYWVHNEMKAEIKMFFETNENKDT TYQNLWDAFKAVCRGKFIALNAYK RKQERSKIDTLTSQLKELEKQEQTHS KASRRQEITKIRAELKEIETQKTLQKI NESRSWFFERINKIDRPLARLIKKKR EKNQIDTIKNDKGDITTDPTEIQTTIR EYYKHLYANKLENLEEMDTFLDTY TLPRLNQEEVESLNRPITGSEIVAIINS LPTKKSPGPDGFTAEFYQRYKEELVP FLLKLFQSIEKEGILPNSFYEASIILIPK PGRDTTKKENFRPISLMNIDAKILNKI LANRIQQHIKKLIHHDQVGFIPGMQG WFNIRKSINVIQHINRAKDKNHVIISI DAEKAFDKIQQPFMLKTLNKLGIDG MYLKIIRAIYDKPTANIILNGQKLEAF PLKTGTRQGCPLSPLLFNIVLEVLAR AIRQEKEIKGIQLGKEEVKLSLFADD MIVYLENPIVSAQNLLKLISNFSKVS GYKINVQKSQAFLYNNNRQTESQIM	IPR000477, PF00078, cd01650

				GELPFTIASKRIKYLGIQLTRDVKDLF KENYKPLLKEIKEDTNKWKNIPCSW VGRINIVKMAILPKVIYRFNAIPIKLP MTFFTELEKTTLKFIWNQKRARIAKS ILSQKNKAGGITLPDFKLYYKATVT KTAWYWYQNRDIDQWNRTEPSEIM PHIYNYLIFDKPEKNKQWGKDSLLN KWCWENWLAICRKLKLDPFLTPYT KINSRWIKDLNVKPKTIKTLEENLGI TIQDIGVGKDFMSKTPKAMATKDKI DKWDLIKLKSFCTAKETTIRVNRQPT TWEKIFATYSSDKGLISRIYNELKQIY KKKTNNPIKKWAKDMNRHFSKEDI YAAKKHMKKCSSSLAIREMQIKTTM RYHLTPVRMAIIKKSGNNRCWRGCG EIGTLVHCWWDCKLVQPLWKSVWR FLRDLELEIPFDPAIPLLGIYPKDYKS CCYKDTCTRMFIAALFTIAKTWNQP NCPTMIDWIKKMWHIYTMEYYAAI KNDEFISFVGTWMKLETIILSKLSQE QKTKHRIFSLIGGN (SEQ ID NO: 1554)	
Penelope	Ретротранспозон, не содержащий LTR	AAL14979.1	Q95VB5	MERSPEPSININGRHAVCTATNMSY AKIKTKYKDSKRTINKFQLTLVKLT KLKSSLKFLLKCRKSNLIPNFIKNLT QHLTILTTDNKTHPDITRTLTRHTHF YHTKILNLLIKHKHNLLQEQTKHMQ KAKTNIEQLMTTDDAKAFFESERNIE NKITTTLKKRQETKHDKLRDQRNLA LADNNTQREWFVNKTKIEFPPNVVA LLAKGPKFALPISKRDFPLLKYIADG EELVQTIKEKETQESARTKFSLLVKE HKTKNNQNSRDRAILDTVEQTRKLL KENINIKILSSDKGNKTVAMDEDEY KNKMTNILDDLCAYRTLRLDPTSRL	IPR000477, PF00078, cd00304

				QTKNNTFVAQLFKMGLISKDERNK MTTTTAVPPRIYGLPKIHKEGTPLRPI CSSIGSPSYGLCKYIIQILKNLTMDSR YNIKNAVDFKDRVNNSQIREEETLV SFDVVSLFPSIPIELALDTIRQKWTKL EEHTNIPKQLFMDIVRFCIEENRYFK YEDKIYTQLKGMPMGSPASPVIADIL MEELLDKITDKLKIKPRLLTKYVDD LFAITNKIDVENILKELNSFHKQIKFT MELEKDGKLPFLDSIVSRMDNTLKI KWYRKPIASGRILNFNSNHPKSMIIN TALGCMNRMMKISDTIYHKEIEHEIK ELLTKNDFPPNIIKTLLKRRQIERKKP TEPAKIYKSLIYVPRLSERLTNSDCY NKQDIKVAHKPTNTLQKFFNKIKSKI PMIEKSNVVYQIPCGGDNNNKCNSV YIGTTKSKLKTRISQHKSDFKLRHQN NIQKTALMTHCIRSNHTPNFDETTIL QQEQHYNKRHTLEMLHIINTPTYKR LNYKTDTENCAHLYRHLLNSQTTSV TISTSKSADV (SEQ ID NO: 1555)	
RT M-ML V	Ретровирус	ADS42990.1	P03355[660-1330]	TLNIEDEHRLHETSKEPDVSLGSTWL SDFPQAWAETGGMGLAVRQAPLIIP LKATSTPVSIKQYPMSQEARLGIKPH IQRLLDQGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPEMGI SGQLTWTRLPQGFKNSPTLFDEALH RDLADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRW LTEARKETVMGQPTPKTPRQLREFL GTAGFCRLWIPGFAEMAAPLYPLTK TGTLFNWGPDQQKAYQEIKQALLTA	IPR000477, PF00078, cd03715

				PALGLPDLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKKLDPVA AGWPPCLRMVAAIAVLTKDAGKLT MGQPLVILAPHAVEALVKQPPDRWL SNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSS LLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHIHGEIYRRR GLLTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMA DQAARKAAITETPDTSTLL (SEQ ID NO: 1556)	
RT RSV	Ретровирус	AAC82561.1	P03354[709-1567]	TVALHLAIPLKWKPDHTPVWIDQWP LPEGKLVALTQLVEKELQLGHIEPSL SCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRG WPLMVLDLKDCFFSIPLAEQDREAF AFTLPSVNNQAPARRFQWKVLPQG MTCSPTICQLVVGQVLEPLRLKHPSL CMLHYMDDLLLAASSHDGLEAAGE EVISTLERAGFTISPDKVQREPGVQY LGYKLGSTYVAPVGLVAEPRIATLW DVQKLVGSLQWLRPALGIPPRLMGP FYEQLRGSDPNEAREWNLDMKMA WREIVRLSTTAALERWDPALPLEGA VARCEQGAIGVLGQGLSTHPRPCLW LFSTQPTKAFTAWLEVLTLLITKLRA SAVRTFGKEVDILLLPACFREDLPLP EGILLALKGFAGKIRSSDTPSIFDIAR PLHVSLKVRVTDHPVPGPTVFTDAS SSTHKGVVVWREGPRWEIKEIADLG ASVQQLEARAVAMALLLWPTTPTN VVTDSAFVAKMLLKMGQEGVPSTA	IPR000477, PF00078, cd01645

				AAFILEDALSQRSAMAAVLHVRSHS EVPGFFTEGNDVADSQATFQAYPLR EAKDLHTALHIGPRALSKACNISMQ QAREVVQTCPHCNSAPALEAGVNPR GLGPLQIWQTDFTLEPRMAPRSWLA VTVDTASSAIVVTQHGRVTSVAVQH HWATAIAVLGRPKAIKTDNGSCFTS KSTREWLARWGIAHTTGIPGNSQGQ AMVERANRLLKDRIRVLAEGDGFM KRIPTSKQGELLAKAMYALNHFERG ENTKTPIQKHWRPTVLTEGPPVKIRI ETGEWEKGWNVLVWGRGYAAVKN RDTDKVIWVPSRKVKPDITQKDEVT KKDEASPLFAG (SEQ ID NO: 1557)	
RT AM V	Ретров ирус	HW6066 80.1	-	TVALHLAIPLKWKPNHTPVWIDQWP LPEGKLVALTQLVEKELQLGHIEPSL SCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRG WPLMVLDLKDCFFSIPLAEQDREAF AFTLPSVNNQAPARRFQWKVLPQG MTCSPTICQLIVGQILEPLRLKHPSLR MLHYMDDLLLAASSHDGLEAAGEE VISTLERAGFTISPDKVQREPGVQYL GYKLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPPRLMGPF YEQLRGSDPNEAREWNLDMKMAW REIVQLSTTAALERWDPALPLEGAV ARCEQGAIGVLGQGLSTHPRPCLWL FSTQPTKAFTAWLEVLTLLITKLRAS AVRTFGKEVDILLLPACFREDLPLPE GILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSS THKGVVVWREGPRWEIKEIADLGAS VQQLEARAVAMALLLWPTTPTNVV TDSAFVAKMLLKMGQEGVPSTAAA	IPR000477, PF00078, cd01645

				FILEDALSQRSAMAAVLHVRSHSEV PGFFTEGNDVADSQATFQAY (SEQ ID NO: 1558)	
RT HIV	Ретровирус	AAB50259.1	P04585[588-1147]	PISPIETVPVKLKPGMDGPKVKQWPL TEEKIKALVEICTEMEKEGKISKIGPE NPYNTPVFAIKKKDSTKWRKLVDFR ELNKRTQDFWEVQLGIPHPAGLKKK KSVTVLDVGDAYFSVPLDEDFRKYT AFTIPSINNETPGIRYQYNVLPQGWK GSPAIFQSSMTKILEPFRKQNPDIVIY QYMDDLYVGSDLEIGQHRTKIEELR QHLLRWGLTTPDKKHQKEPPFLWM GYELHPDKWTVQPIVLPEKDSWTVN DIQKLVGKLNWASQIYPGIKVRQLC KLLRGTKALTEVIPLTEEAELELAEN REILKEPVHGVYYDPSKDLIAEIQKQ GQGQWTYQIYQEPFKNLKTGKYAR MRGAHTNDVKQLTEAVQKITTESIVI WGKTPKFKLPIQKETWETWWTEYW QATWIPEWEFVNTPPLVKLWYQLEK EPIVGAETFYVDGAANRETKLGKAG YVTNRGRQKVVTLTDTTNQKTELQ AIYLALQDSGLEVNIVTDSQYALGII QAQPDQSESELVNQIIEQLIKKEKVY LAWVPAHKGIGGNEQVDKLVSAGIR KVL (SEQ ID NO: 1559)	IPR000477, PF00078, cd01645
Ty1	Ретротранспозон, содержащий LTR	AAA66938.1	Q07163-1[1218-1755]	AVKAVKSIKPIRTTLRYDEAITYNKD IKEKEKYIEAYHKEVNQLLKMKTW DTDEYYDRKEIDPKRVINSMFIFNKK RDGTHKARFVARGDIQHPDTYDSG MQSNTVHHYALMTSLSLALDNNYY ITQLDISSAYLYADIKEELYIRPPPHL GMNDKLIRLKKSLYGLKQSGANWY ETIKSYLIQQCGMEEVRGWSCVFKN SQVTICLFVDDMVLFSKNLNSNKRII	IPR013103, PF07727

				EKLKMQYDTKIINLGESDEEIQYDIL GLEIKYQRGKYMKLGMENSLTEKIP KLNVPLNPKGRKLSAPGQPGLYIDQ DELEIDEDEYKEKVHEMQKLIGLAS YVGYKFRFDLLYYINTLAQHILFPSR QVLDMTYELIQFMWDTRDKQLIWH KNKPTEPDNKLVAISDASYGNQPYY KSQIGNIYLLNGKVIGGKSTKASLTC TSTTEAEIHAISESVPLLNNLSYLIQE LNKKPIIKGLLTDSRSTISIIKSTNEEK FRNRFFGTKAMRLRDEVSGNNLYV YYIETKKNIADVMTKPLPIKTFKLLT NKWIH (SEQ ID NO: 1560)	
Brt	Ретроэлемент, генерирующий разнообразие	NP_958675.1	Q775D8	MGKRHRNLIDQITTWENLLDAYRKT SHGKRRTWGYLEFKEYDLANLLAL QAELKAGNYERGPYREFLVYEPKPR LISALEFKDRLVQHALCNIVAPIFEA GLLPYTYACRPDKGTHAGVCHVQA ELRRTRATHFLKSDFSKFFPSIDRAA LYAMIDKKIHCAATRRLLRVVLPDE GVGIPIGSLTSQLFANVYGGAVDRLL HDELKQRHWARYMDDIVVLGDDPE ELRAVFYRLRDFASERLGLKISHWQ VAPVSRGINFLGYRIWPTHKLLRKSS VKRAKRKVANFIKHGEDESLQRFLA SWSGHAQWADTHNLFTWMEEQYGI ACH (SEQ ID NO: 1561)	IPR000477, PF00078, cd01646
RT86	Ретрон	AAA61471.1	P23070	MKSAEYLNTFRLRNLGLPVMNNLH DMSKATRISVETLRLLIYTADFRYRI YTVEKKGPEKRMRTIYQPSRELKAL QGWVLRNILDKLSSSPFSIGFEKHQSI LNNATPHIGANFILNIDLEDFFPSLTA NKVFGVFHSLGYNRLISSVLTKICCY KNLLPQGAPSSPKLANLICSKLDYRI QGYAGSRGLIYTRYADDLTLSAQSM	IPR000477, PF00078, cd03487

				KKVVKARDFLFSIIPSEGLVINSKKTC ISGPRSQRKVTGLVISQEKVGIGREK YKEIRAKIHHIFCGKSSEIEHVRGWL SFILSVDSKSHRRLITYISKLEKKYGK NPLNKAKT (SEQ ID NO: 1562)	
TERT	Теломеразa	AAG23289.1	O14746	MPRAPRCRAVRSLLRSHYREVLPLA TFVRRLGPQGWRLVQRGDPAAFRA LVAQCLVCVPWDARPPPAAPSFRQV SCLKELVARVLQRLCERGAKNVLAF GFALLDGARGGPPEAFTTSVRSYLP NTVTDALRGSGAWGLLLRRVGDDV LVHLLARCALFVLVAPSCAYQVCGP PLYQLGAATQARPPPHASGPRRRLG CERAWNHSVREAGVPLGLPAPGAR RRGGSASRSLPLPKRPRRGAAPEPER TPVGQGSWAHPGRTRGPSDRGFCV VSPARPAEEATSLEGALSGTRHSHPS VGRQHHAGPPSTSRPPRPWDTPCPP VYAETKHFLYSSGDKEQLRPSFLLSS LRPSLTGARRLVETIFLGSRPWMPGT PRRLPRLPQRYWQMRPLFLELLGNH AQCPYGVLLKTHCPLRAAVTPAAG VCAREKPQGSVAAPEEEDTDPRRLV QLLRQHSSPWQVYGFVRACLRRLVP PGLWGSRHNERRFLRNTKKFISLGK HAKLSLQELTWKMSVRDCAWLRRS PGVGCVPAAEHRLREEILAKFLHWL MSVYVVELLRSFFYVTETTFQKNRL FFYRKSVWSKLQSIGIRQHLKRVQL RELSEAEVRQHREARPALLTSRLRFI PKPDGLRPIVNMDYVVGARTFRREK RAERLTSRVKALFSVLNYERARRPG LLGASVLGLDDIHRAWRTFVLRVRA QDPPPELYFVKVDVTGAYDTIPQDR LTEVIASIIKPQNTYCVRRYAVVQKA	IPR000477, PF00078, cd01648

				AHGHVRKAFKSHVSTLTDLQPYMR QFVAHLQETSPLRDAVVIEQSSSLNE ASSGLFDVFLRFMCHHAVRIRGKSY VQCQGIPQGSILSTLLCSLCYGDMEN KLFAGIRRDGLLLRLVDDFLLVTPHL THAKTFLRTLVRGVPEYGCVVNLRK TVVNFPVEDEALGGTAFVQMPAHG LFPWCGLLLDTRTLEVQSDYSSYAR TSIRASLTFNRGFKAGRNMRRKLFG VLRLKCHSLFLDLQVNSLQTVCTNI YKILLLQAYRFHACVLQLPFHQQVW KNPTFFLRVISDTASLCYSILKAKNA GMSLGAKGAAGPLPSEAVQWLCHQ AFLLKLTRHRVTYVPLLGSLRTAQT QLSRKLPGTTLTALEAAANPALPSDF KTILD (SEQ ID NO: 1563)	
RT Mauriceville	Ретроплазмида	NC_001570.1	Q36578	MPNHRLPNCVSYLGENHELSWLHG MFGLLKRSNPQTGGILGWLNTGPNG FVKYMMNLMGHARDKGDAKEYW RLGRSLMKNEAFQVQAFNHVCKHW YLDYKPHKIAKLLKEVREMVEIQPV CIDYKRVYIPKANGKQRPLGVPTVP WRVYLHMWNVLLVWYRIPEQDNQ HAYFPKRGVFTAWRALWPKLDSQN IYEFDLKNFFPSVDLAYLKDKLMES GIPQDISEYLTVLNRSLVVLTSEDKIP EPHRDVIFNSDGTPNPNLPKDVQGRI LKDPDFVEILRRRGFTDIATNGVPQG ASTSCGLATYNVKELFKRYDELIMY ADDGILCRQDPSTPDFSVEEAGVVQ EPAKSGWIKQNGEFKKSVKFLGLEFI PANIPPLGEGEVKDYPRLRGATRNG SKMELSTELQFLCYLSYKLRIKVLRD LYIQVLGYLPSVPLLRYRSLAEAINE LSPKRITIGQFITSSFEEFTAWSPLKR	cd00304

				MGFFFSSPAGPTILSSIFNNSTNLQEP SDSRLLYRKGSWVNIRFAAYLYSKL SEEKHGLVPKFLEKLREINFALDKVD VTEIDSKLSRLMKFSVSAAYDEVGT LALKSLFKFRNSERESIKASFKQLRE NGKIAEFSEARRLWFEILKLIRLDLF NASSLACDDLLSHLQDRRSIKKWGS SDVLYLKSQRLMRTNKKQLQLDFE KKKNSLKKKLIKRRAKELRDTFKGK ENKEA(SEQ ID NO: 1564)	
RTX	Сконструированная полимераза	QFN49000.1	-	MILDTDYITEDGKPVIRIFKKENGEF KIEYDRTFEPYLYALLKDDSAIEEVK KITAERHGTVVTVKRVEKVQKKFLG RPVEVWKLYFTHPQDVPAIMDKIRE HPAVIDIYEYDIPFAIRYLIDKGLVPM EGDEELKLLAFDIETLYHEGEEFAEG PILMISYADEEGARVITWKNVDLPY VDVVSTEREMIKRFLRVVKEKDPDV LITYNGDNFDFAYLKKRCEKLGINF ALGRDGSEPKIQRMGDRFAVEVKGR IHFDLYPVIRRTINLPTYTLEAVYEA VFGQPKEKVYAEEITTAWETGENLE RVARYSMEDAKVTYELGKEFLPME AQLSRLIGQSLWDVSRSSTGNLVEW FLLRKAYERNELAPNKPDEKELARR HQSHEGGYIKEPERGLWENIVYLDF RSLYPSIIITHNVSPDTLNREGCKEYD VAPQVGHRFCKDFPGFIPSLLGDLLE ERQKIKKRMKATIDPIERKLLDYRQR AIKILANSLYGYYGYARARWYCKEC AESVIAWGREYLTMTIKEIEEKYGFK VIYSDTDGFFATIPGADAETVKKKA MEFLKYINAKLPGALELEYEGFYKR GLFVTKKKYAVIDEEGKITTRGLEIV RRDWSEIAKETQARVLEALLKDGDV	IPR006134, PF00136, cd05536

				EKAVRIVKEVTEKLSKYEVPPEKLVI HKQITRDLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKGSGRIVD RAIPFDEFDPTKHKYDAEYYIEKQVL PAVERILRAFGYRKEDLRYQKTRQV GLSARLKPKGTLEGSSHHHHHH (SEQ ID NO: 1565)	

Таблица 2. Описания сигнатур InterPro, присутствующих в обратных транскриптазах, из таблицы 1

Сигнатура	База данных	Краткое название	Описание
cd00304	CDD	*RT_подобные*	RT_подобные: семейство, подобное обратной транскриптазе (RT, РНК-зависимой ДНК-полимеразе). Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]
cd01645	CDD	*RT_Rtv*	RT_Rtv: обратные транскриптазы (RT) ретровирусов (Rtv). RT катализируют превращение однонитевой РНК в двухнитевую

			вирусную ДНК для интеграции в хромосомы хозяина. Белки этого подсемейства содержат длинные концевые повторы (LTR) и представляют собой многофункциональные ферменты с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). Вирусный РНК-геном входит в цитоплазму в качестве части нуклеопротеинового комплекса, и в цитоплазме организуется процесс обратной транскрипции с образованием линейного ДНК-дуплекса посредством сложного ряда стадий. Этот ДНК-дуплекс является коллинеарным к своей РНК-матрице, но содержит концевые дупликации, известные как LTR, которые отсутствуют в вирусной РНК. Было высказано предположение, что для создания LTR необходимы два специализированных переключателя матрицы, известных в качестве реакций переноса нити или "прыжков". [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]
cd01646	CDD	*RT_бактериальный_ретрон_I*	RT_бактериальный_ретрон_I: обратные транскриптазы (RT) в бактериальных ретротранспозонах или ретронах. Полимеразная реакция этого фермента приводит к образованию уникального комплекса РНК-ДНК, называемого msDNA (многокопийная однонитевая (ss) ДНК), в котором небольшая ssDNA ответвляется от небольшой молекулы ssRNA посредством 2'-5'-фосфодиэфирной связи. RT бактериального ретрона продуцирует кДНК, соответствующую

			лишь небольшой части генома ретрона. [PMID: 1698615, PMID: 16093702, PMID: 8828137]
cd01648	CDD	*TERT*	TERT: теломеразная обратная транскриптаза (TERT). Теломераза представляет собой рибонуклеопротеин (RNP), который синтезирует повторы теломерной ДНК. Субъединица теломеразной РНК обеспечивает матрицу для синтеза этих повторов. Каталитическая субъединица RNP известна в качестве теломеразной обратной транскриптазы (TERT). Домен обратной транскриптазы (RT) расположен в С-концевой области полипептида TERT. Единичные аминокислотные замены в этой области приводят к укорочению теломер и старению. Теломераза представляет собой фермент, который в определенных клетках поддерживает физические концы хромосом (теломеры) во время репликации. В соматических клетках репликация отстающей нити требует постоянного присутствия РНК-праймера на примерно 200 нуклеотидов выше по ходу транскрипции, который является комплементарным матричной нити. Поскольку существует область ДНК, расположенная на расстоянии менее 200 пар оснований от конца хромосомы, где это является невозможным, то хромосома постоянно укорачивается. Однако избыток повторяющейся ДНК на концах хромосом защищает от разрушения кодирующую ген ДНК. Теломераза обычно не экспрессируется в соматических клетках. Было высказано предположение, что экзогенная TERT может увеличить продолжительность жизни клетки или даже привести к ее иммортализации. Однако недавние исследования показали, что ряд

			онкогенов может индуцировать теломеразную активность. В противоположность этому, онкоген c-myc может быть активирован в клетках человека, иммортализованных с помощью TERT. Сравнение последовательностей относит теломеразные белки к семейству RT, но выявляет признаки, отличающие их от родственных белков ретровирусов и ретротранспозонов. [PMID: 9110970, PMID: 9288757, PMID: 9389643, PMID: 9671703, PMID: 9671704, PMID: 10333526, PMID: 11250070, PMID: 15363846, PMID: 16416120, PMID: 16649103, PMID: 16793225, PMID: 10860859, PMID: 9252327, PMID: 11602347, PMID: 1698615, PMID: 8828137, PMID: 10866187]
cd01650	CDD	*RT_nLTR_подобные*	RT_nLTR: ретротранспозон, не содержащий LTR (длинных концевых повторов), и ретровирусная обратная транскриптаза (RT), не содержащая LTR. Это подсемейство содержит как ретротранспозоны, не содержащие LTR, так и ретровирусные RT, не содержащие LTR. RT катализируют превращение однонитевой РНК в двухнитевую ДНК для интеграции в хромосомы хозяина. RT представляет собой многофункциональный фермент с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). [PMID: 1698615, PMID: 10605110, PMID: 10628860, PMID: 11734649, PMID: 12117499, PMID: 12777502, PMID: 14871946, PMID: 15939396, PMID: 16271150, PMID: 16356661, PMID: 2463954, PMID: 3040362, PMID: 3656436, PMID: 7512193, PMID: 7534829, PMID: 7659515, PMID:

			8524653, PMID: 9190061, PMID: 9218812, PMID: 9332379, PMID: 9364772, PMID: 8828137]
cd01651	CDD	*RT_интрон_G2*	RT_интрон_G2: обратные транскриптазы (RT), происходящие из интрона группы II. RT транскрибирует ДНК, используя РНК в качестве матрицы. Белки данного подсемейства обнаружены в бактериальных и митохондриальных интронах группы II. Их наиболее вероятным предком был ретротранспозонный элемент как с gag-подобными, так и с pol-подобными генами. Данное подсемейство белков, по-видимому, захватило последовательности RT из подвижных элементов, в которых отсутствуют длинные концевые повторы (LTR). [PMID: 1698615, PMID: 8828137, PMID: 12403467, PMID: 11058141, PMID: 11054545, PMID: 10760141, PMID: 10488235, PMID: 9680217, PMID: 9491607, PMID: 7994604, PMID: 7823908, PMID: 3129199, PMID: 2531370, PMID: 2476655]
cd03487	CDD	*RT_бактериальный_ретрон_II*	RT_бактериальный_ретрон_II: обратные транскриптазы (RT) в бактериальных ретротранспозонах или ретронах. Полимеразная реакция этого фермента приводит к образованию уникального комплекса РНК-ДНК, называемого msDNA (многокопийная однонитевая (ss) ДНК), в котором небольшая ssDNA ответвляется от небольшой молекулы ssRNA посредством 2'-5'-фосфодиэфирной связи. RT бактериального ретрона продуцирует кДНК, соответствующую лишь небольшой части генома ретрона. [PMID: 1698615, PMID: 8828137, PMID: 11292805, PMID: 9281493, PMID: 2465092, PMID: 1722556, PMID: 1701261, PMID: 1689062]

cd03715	CDD	*RT_ZFREV_подобные*	RT_ZFREV_подобные: подсемейство обратных транскриптаз (RT), обнаруженное в последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ позволяет предположить, что ZFERV принадлежит к отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
cd05536	CDD	*POLBc_B3*	Каталитический домен подсемейства B3 ДНК-полимераз B-типа. Белки архей, участвующие в репликации ДНК, являются сходными с белками эукариот. Некоторые представители архей также имеют несколько ДНК-полимераз семейства B (B1, B2 и B3). До сих пор не была(были) установлена(установлены) какая-либо(какие-либо) конкретная(конкретные) функция(функции) для различных

			представителей ДНК-полимераз B-типа архей. Филогенетический анализ ДНК-полимераз семейства B эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства B архей и эукариот. Структурное сравнение термостабильной ДНК-полимеразы B-типа с ее мезостабильным гомологом позволяет предположить наличие нескольких приспособлений к высокой температуре, таких как более короткие петли, дисульфидные мостики и усиление электростатического взаимодействия в областях контакта субдоменов. [PMID: 10997874, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321]
cd05780	CDD	*ДНК_полимеразы B_Kod1_подобные_экзо*	3'-5'-экзонуклеазный домен ДНК-полимераз семейства B архей со сходством с Kod1 Pyrococcus kodakaraensis, включая полимеразы из Desulfurococcus (Pol D. Tok) и Thermococcus gorgonarius (Pol Tgo). Kod1, Pol D. Tok и Pol Tgo представляют собой термостабильные ферменты, обладающие как полимеразной, так и 3'-5'-экзонуклеазной активностью. Они представляют собой ДНК-полимеразы семейства B. Их амино-концы содержат DnaQ-подобный 3'-5'-экзонуклеазный домен DEDDy-типа, который содержит три мотива последовательности, называемые ExoI, ExoII и ExoIII, со специфическим паттерном YX(3)D в ExoIII. Данные мотивы образуют кластер вокруг активного центра и участвуют в связывании металлов и катализе. Экзонуклеазный домен полимераз семейства B содержит бета-шпилечную структуру, которая играет важную роль в переключении активного центра в случае

			неправильного включения нуклеотидов. Представители этого подсемейства проявляют сходство с эукариотическими ДНК-полимеразами, участвующими в репликации ДНК. Некоторые археи имеют несколько ДНК-полимераз семейства B. Филогенетический анализ ДНК-полимераз семейства B эубактерий, архей и эукариот подтверждает независимые дупликации генов в ходе эволюции ДНК-полимераз семейства B архей и эукариот. [PMID: 18355915, PMID: 16019029, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321, PMID: 9098062, PMID: 12459442, PMID: 16230118, PMID: 11988770, PMID: 11222749, PMID: 17098747, PMID: 8594362, PMID: 9729885]
PF00078	Pfam	*RVT_1*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. [PMID: 1698615]
PF00136	Pfam	*ДНК_полимераза_B*	Данная область ДНК-полимеразы семейства B, по-видимому, состоит из более чем одного структурного домена, возможно, предусматривающего формы активности элонгации, связывания ДНК и связывания dNTP. [PMID: 9757117, PMID: 8679562]
PF07727	Pfam	*RVT_2*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных

			мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Данная запись Pfam включает обратные транскриптазы, не распознаваемые моделью Pfam:PF00078. [PMID: 1698615]
IPR000477	InterPro	*RT_домен*	Применение РНК-матрицы для получения ДНК с целью интеграции в геном хозяина и использования клетки-хозяина является стратегией, применяемой при репликации ретроидных элементов, таких как ретровирусы и бактериальные ретроны. Фермент, катализирующий полимеризацию, представляет собой РНК-направленную ДНК-полимеразу или обратную транскриптазу (RT) (2.7.7.49). Обратная транскриптаза встречается в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II [PMID: 12758069], бактериальные msDNA, гепаднавирусы и каулимовирусы. Ретровирусная обратная транскриптаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Обнаружение ретроэлементов у прокариот поднимает интригующие вопросы, касающиеся их роли у бактерий, происхождения и эволюции обратных транскриптаз, а также того, являются ли бактериальные обратные транскриптазы старшими, чем эукариотические обратные транскриптазы [PMID: 8828137]. Было

			определено несколько кристаллических структур домена обратной транскриптазы (RT) [PMID: 1377403].
IPR006134	InterPro	ДНК-направленная_ДНК_полимераза_B_с несколькими_доменами	ДНК представляет собой биологическую информацию, которая указывает клеткам, как существовать упорядоченно: таким образом, точная репликация является одним из наиболее важных событий в жизненном цикле клетки. Эту функцию выполняют ДНК-направленные ДНК-полимеразы (2.7.7.7) посредством добавления остатков нуклеотидтрифосфата (dNTP) к 5'-концу растущей цепи ДНК, используя в качестве матрицы комплементарную цепь ДНК. Небольшие молекулы РНК обычно применяются в качестве праймеров для элонгации цепи, хотя концевые белки также могут применяться для синтеза цепи ДНК de novo. Несмотря на то, что существует 2 разных способа праймирования, они опосредуются 2 очень сходными классами полимераз, A и B, со сходными способами элонгации цепи. Ряд ДНК-полимераз был сгруппирован под обозначением ДНК-полимераз семейства B. Шесть областей сходства (пронумерованных от I до VI) обнаружены во всех или в подмножестве полимераз семейства B. Наиболее консервативная область (I) включает консервативный тетрапептид с двумя остатками аспартата. Было высказано предположение, что он может участвовать в связывании иона магния. Все последовательности в семействе B содержат характерный мотив DTDS (SEQ ID NO: 1566) и имеют несколько функциональных доменов, включая 5'-3'-домен элонгации, 3'-5'-экзонуклеазный домен [PMID: 8679562], ДНК-связывающий домен и домены связывания как

			для dNTP, так и для пирофосфата [PMID: 9757117]. Данный домен ДНК-полимеразы семейства B, по-видимому, характеризуется более чем одной формой активности, возможно, включающей элонгацию, связывание ДНК и связывание dNTP [PMID: 9757117].
IPR013103	InterPro	*RVT_2*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эта запись включает обратные транскриптазы, не распознаваемые IPR000477 [PMID: 1698615].

Таблица 30. Иллюстративные мономерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	**Номер доступа**	**Организм**	**Последовательность**	**Сигнатуры RT**
Q4VFZ2_9GAMR - только остатки	Q4VFZ2	Эндогенный ретровирус свиней	MGATGQQQYPWTTRRTVDLGVGR VTHSFLVIPECPAPLLGRDLLTKMGA QISFEQGKPEVSANNKPITVLTLQLD DEYRLYSPLVKPDQNIQFWLEQFPQ AWAETAGMGLAKQVPPQVIQLKAS ATPVSVRQYPLSKEAQEGIRPHVQR LIQQGILVPVQSPWNTPLLPVRKPGT NDYRPVQDLREVNKRVQDIHPTVPN PYNLLCALPPQRSWYTVLDLKDAFF CLRLHPTSQPLFAFEWRDPGTGRTG QLTWTRLPQGFKNSPTIFDEALHRDL ANFRIQHPQVTLLQYVDDLLLAGAT KQDCLEGTKALLLELSDLGYRASAK KAQICRREVTYLGYSLRDGQRWLTE ARKKTVVQIPAPTTAKQVREFLGTA	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			GFCRLWIPGFATLAAPLYPLTKEKGE FSWAPEHQKAFDAIKKALLSAPALA LPDVTKPFTLYVDERKGVARGVLTQ TLGPWRRPVAYLSKKLDPVASGWP VCLKAIAAVAILVKDADKLTLGQNI TVIAPHALENIVRQPPDRWMTNARM THYQSLLLTERVTFAPPAALNPATLL PEETDEPVTHDCHQLLIEETGVRKDL TDIPLTGEVLTWFTDGSSYVVEGKR MAGAAVVDGTRTIWASSLPEGTSAQ KAELMALTQALRLAEGKSINIYTDS RYAFATAHVHGAIYKQRGLLTSAGR EIKNKEEILSLLEALHLPKRLAIIHCP GHQKAKDPISRGNQMADRVAKQAA QGVNLLPMIETPKAPEPGRQYTLED WQEIKKIDQFSETPEGTCYTSDGKEI LPHKEGLEYVQQIHRLTHLGTKHLQ QLVRTSPYHVLRLPGVADSVVKHCV PCQLVNANPSRIPPGKRLRGSHPGAH WEVDFTEVKPAKYGNKYLLVFVDT FSGWVEAYPTKKETSTVVAKKILEEI FPRFGIPKVIGSDNGPAFVAQVSQGL AKILGIDWKLHCAYRPQSSGQVERM NRTIKETLTKLTAETGVNDWIALLPF VLFRVRNTPGQFGLTPYELLYGGPPP LVEIASVHSADVLLSQPLFSRLKALE WVRQRAWRQLREAYSGGGDLQIPH RFQVGDSVYVRRHRAGNLETRWKG PYHVLLTTPTAVKVEGISTWIHASHV KPAPPPDSGWKAEKTENPLKLRLHR VVPYSVNNFSS (SEQ ID NO: 1567)	
POL_SFV1 - только остатки	P23074	Вирус пенистости обезьян 1 типа	MDPLQLLQPLEAEIKGTKLKAHWDS GATITCVPEAFLEDERPIQTMLIKTIH GEKQQDVYYLTFKVQGRKVEAEVL ASPYDYILLNPSDVPWLMKKPLQLT	IPR043502, SSF56672, IPR000477, PF00078

			VLVPLHEYQERLLQQTALPKEQKEL LQKLFLKYDALWQHWENQVGHRRI KPHNIATGTLAPRPQKQYPINPKAKP SIQIVIDDLLKQGVLIQQNSTMNTPV YPVPKPDGKWRMVLDYREVNKTIP LIAAQNQHSAGILSSIYRGKYKTTLD LTNGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFTADVVDL LKEIPNVQAYVDDIYISHDDPQEHLE QLEKIFSILLNAGYVVSLKKSEIAQR EVEFLGFNITKEGRGLTDTFKQKLLN ITPPKDLKQLQSILGLLNFARNFIPNY SELVKPLYTIVANANGKFISWTEDNS NQLQHIISVLNQADNLEERNPETRLII KVNSSPSAGYIRYYNEGSKRPIMYV NYIFSKAEAKFTQTEKLLTTMHKGLI KAMDLAMGQEILVYSPIVSMTKIQR TPLPERKALPVRWITWMTYLEDPRI QFHYDKSLPELQQIPNVTEDVIAKTK HPSEFAMVFYTDGSAIKHPDVNKSH SAGMGIAQVQFIPEYKIVHQWSIPLG DHTAQLAEIAAVEFACKKALKISGP VLIVTDSFYVAESANKELPYWKSNG FLNNKKKPLRHVSKWKSIAECLQLK PDIIIMHEKGHQQPMTTLHTEGNNL ADKLATQGSYVVHCNTTPSLDAELD QLLQGHYPPGYPKQYKYTLEENKLI VERPNGIRIVPPKADREKIISTAHNIA HTGRDATFLKVSSKYWWPNLRKDV VKSIRQCKQCLVTNATNLTSPPILRP VKPLKPFDKFYIDYIGPLPPSNGYLH VLVVVDSMTGFVWLYPTKAPSTSAT VKALNMLTSIAIPKVLHSDQGAAFTS STFADWAKEKGIQLEFSTPYHPQSSG KVERKNSDIKRLLTKLLIGRPAKWY	

			DLLPVVQLALNNSYSPSSKYTPHQL LFGVDSNTPFANSDTLDLSREEELSL LQEIRSSLHQPTSPPASSRSWSPSVGQ LVQERVARPASLRPRWHKPTAILEV VNPRTVIILDHLGNRRTVSVDNLKLT AYQDNGTSNDSGTMALMEEDESSTS ST (SEQ ID NO: 1568)	
POL_MPMV - только остатки	P07572	Вирус обезьян Мейзона-Пфайзера	MGQELSQHERYVEQLKQALKTRGV KVKYADLLKFFDFVKDTCPWFPQEG TIDIKRWRRVGDCFQDYYNTFGPEK VPVTAFSYWNLIKELIDKKEVNPQV MAAVAQTEEILKSNSQTDLTKTSQN PDLDLISLDSDDEGAKSSSLQDKGLS STKKPKRFPVLLTAQTSKDPEDPNPS EVDWDGLEDEAAKYHNPDWPPFLT RPPPYNKATPSAPTVMAVVNPKEEL KEKIAQLEEQIKLEELHQALISKLQK LKTGNETVTHPDTAGGLSRTPHWPG QHIPKGKCCASREKEEQIPKDIFPVTE TVDGQGQAWRHHNGFDFAVIKELK TAASQYGATAPYTLAIVESVADNWL TPTDWNTLVRAVLSGGDHLLWKSE FFENCRDTAKRNQQAGNGWDFDML TGSGNYSSTDAQMQYDPGLFAQIQA AATKAWRKLPVKGDPGASLTGVKQ GPDEPFADFVHRLITTAGRIFGSAEA GVDYVKQLAYENANPACQAAIRPY RKKTDLTGYIRLCSDIGPSYQQGLA MAAAFSGQTVKDFLNNKNKEKGGC CFKCGKKGHFAKNCHEHAHNNAEP KVPGLCPRCKRGKHWANECKSKTD NQGNPIPPHQGNRVEGPAPGPETSL WGSQLCSSQQKQPISKLTRATPGSA GLDLCSTSHTVLTPEMGPQALSTGIY GPLPPNTFGLILGRSSITMKGLQVYP	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			GVIDNDYTGEIKIMAKAVNNIVTVS QGNRIAQLILLPLIETDNKVQQPYRG QGSFGSSDIYWVQPITCQKPSLTLWL DDKMFTGLIDTGADVTIIKLEDWPP NWPITDTLTNLRGIGQSNNPKQSSKY LTWRDKENNSGLIKPFVIPNLPVNL WGRDLLSQMKIMMCSPNDIVTAQM LAQGYSPGKGLGKKENGILHPIPNQ GQSNKKGFGNFLTAAIDILAPQQCA EPITWKSDEPVWVDQWPLTNDKLA AAQQLVQEQLEAGHITESSSPWNTPI FVIKKKSGKWRLLQDLRAVNATMV LMGALQPGLPSPVAIPQGYLKIIIDLK DCFFSIPLHPSDQKRFAFSLPSTNFKE PMQRFQWKVLPQGMANSPTLCQKY VATAIHKVRHAWKQMYIIHYMDDIL IAGKDGQQVLQCFDQLKQELTAAGL HIAPEKVQLQDPYTYLGFELNGPKIT NQKAVIRKDKLQTLNDFQKLLGDIN WLRPYLKLTTGDLKPLFDTLKGDSD PNSHRSLSKEALASLEKVETAIAEQF VTHINYSLPLIFLIFNTALTPTGLFWQ DNPIMWIHLPASPKKVLLPYYDAIA DLIILGRDHSKKYFGIEPSTIIQPYSKS QIDWLMQNTEMWPIACASFVGILDN HYPPNKLIQFCKLHTFVFPQIISKTPL NNALLVFTDGSSTGMAAYTLTDTTI KFQTNLNSAQLVELQALIAVLSAFPN QPLNIYTDSAYLAHSIPLLETVAQIK HISETAKLFLQCQQLIYNRSIPFYIGH VRAHSGLPGPIAQGNQRADLATKIV ASNINTNLESAQNAHTLHHLNAQTL RLMFNIPREQARQIVKQCPICVTYLP VPHLGVNPRGLFPNMIWQMDVTHY SEFGNLKYIHVSIDTFSGFLLATLQT	

			GETTKHVITHLLHCFSIIGLPKQIKTD NGPGYTSKNFQEFCSTLQIKHITGIPY NPQGQGIVERAHLSLKTTIEKIKKGE WYPRKGTPRNILNHALFILNFLNLDD QNKSAADRFWHNNPKKQFAMVKW KDPLDNTWHGPDPVLIWGRGSVCV YSQTYDAARWLPERLVRQVSNNNQ SRE (SEQ ID NO: 1569)	
POL_M MTVB - только остатки	P03365	Вирус опухоли молочной железы мышей	MGVSGSKGQKLFVSVLQRLLSERGL HVKESSAIEFYQFLIKVSPWFPEEGG LNLQDWKRVGREMKRYAAEHGTD SIPKQAYPIWLQLREILTEQSDLVLLS AEAKSVTEEELEEGLTGLLSTSSQEK TYGTRGTAYAEIDTEVDKLSEHIYDE PYEEKEKADKNEEKDHVRKIKKVV QRKENSEGKRKEKDSKAFLATDWN DDDLSPEDWDDLEEQAAHYHDDDE LILPVKRKVVKKKPQALRRKPLPPV GFAGAMAEAREKGDLTFTFPVVFM GESDEDDTPVWEPLPLKTLKELQSA VRTMGPSAPYTLQVVDMVASQWLT PSDWHQTARATLSPGDYVLWRTEY EEKSKEMVQKAAGKRKGKVSLDML LGTGQFLSPSSQIKLSKDVLKDVTTN AVLAWRAIPPPGVKKTVLAGLKQG NEESYETFISRLEEAVYRMMPRGEG SDILIKQLAWENANSLCQDLIRPIRK TGTIQDYIRACLDASPAVVQGMAYA AAMRGQKYSTFVKQTYGGGKGGQ GAEGPVCFSCGKTGHIRKDCKDEKG SKRAPPGLCPRCKKGYHWKSECKSK FDKDGNPLPPLETNAENSKNLVKGQ SPSPAQKGDGVKGSGLNPEAPPFTIH DLPRGTPGSAGLDLSSQKDLILSLED GVSLVPTLVKGTLPEGTTGLIIGRSS	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			NYKKGLEVLPGVIDSDFQGEIKVMV KAAKNAVIIHKGERIAQLLLLPYLKL PNPVIKEERGSEGFGSTSHVHWVQEI SDSRPMLHIYLNGRRFLGLLDTGAD KTCIAGRDWPANWPIHQTESSLQGL GMACGVARSSQPLRWQHEDKSGIIH PFVIPTLPFTLWGRDIMKDIKVRLMT DSPDDSQDLMIGAIESNLFADQISWK SDQPVWLNQWPLKQEKLQALQQLV TEQLQLGHLEESNSPWNTPVFVIKK KSGKWRLLQDLRAVNATMHDMGA LQPGLPSPVAVPKGWEIIIIDLQDCFF NIKLHPEDCKRFAFSVPSPNFKRPYQ RFQWKVLPQGMKNSPTLCQKFVDK AILTVRDKYQDSYIVHYMDDILLAH PSRSIVDEILTSMIQALNKHGLVVST EKIQKYDNLKYLGTHIQGDSVSYQK LQIRTDKLRTLNDFQKLLGNINWIRP FLKLTTGELKPLFEILNGDSNPISTRK LTPEACKALQLMNERLSTARVKRLD LSQPWSLCILKTEYTPTACLWQDGV VEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFD LLLQEKEDWPISLLGFLGEVHFHLPK DPLLTFTLQTAIIFPHMTSTTPLEKGI VIFTDGSANGRSVTYIQGREPIIKENT QNTAQQAEIVAVITAFEEVSQPFNLY TDSKYVTGLFPEIETATLSPRTKIYTE LKHLQRLIHKRQEKFYIGHIRGHTGL PGPLAQGNAYADSLTRILTALESAQE SHALHHQNAAALRFQFHITREQAREI VKLCPNCPDWGHAPQLGVNPRGLK PRVLWQMDVTHVSEFGKLKYVHVT VDTYSHFTFATARTGEATKDVLQHL AQSFAYMGIPQKIKTDNAPAYVSRSI	

			QEFLARWKISHVTGIPYNPQGQAIVE RTHQNIKAQLNKLQKAGKYYTPHH LLAHALFVLNHVNMDNQGHTAAER HWGPISADPKPMVMWKDLLTGSWK GPDVLITAGRGYACVFPQDAETPIW VPDRFIRPFTERKEATPTPGTAEKTPP RDEKDQQESPKNESSPHQREDGLAT SAGVDLRSGGGP (SEQ ID NO: 1570)	
POL_MLVMS - только остатки	P03355	Вирус лейкоза мышей Молони	MGQTVTTPLSLTLGHWKDVERIAHN QSVDVKKRRWVTFCSAEWPTFNVG WPRDGTFNRDLITQVKIKVFSPGPHG HPDQVPYIVTWEALAFDPPPWVKPF VHPKPPPPLPPSAPSLPLEPPRSTPPRS SLYPALTPSLGAKPKPQVLSDSGGPL IDLLTEDPPPYRDPRPPPSDRDGNGG EATPAGEAPDPSPMASRLRGRREPP VADSTTSQAFPLRAGGNGQLQYWPF SSSDLYNWKNNNPSFSEDPGKLTALI ESVLITHQPTWDDCQQLLGTLLTGE EKQRVLLEARKAVRGDDGRPTQLP NEVDAAFPLERPDWDYTTQAGRNH LVHYRQLLLAGLQNAGRSPTNLAK VKGITQGPNESPSAFLERLKEAYRRY TPYDPEDPGQETNVSMSFIWQSAPDI GRKLERLEDLKNKTLGDLVREAEKI FNKRETPEEREERIRRETEEKEERRR TEDEQKEKERDRRRHREMSKLLATV VSGQKQDRQGGERRRSQLDRDQCA YCKEKGHWAKDCPKKPRGPRGPRP QTSLLTLDDQGGQGQEPPPEPRITLK VGGQPVTFLVDTGAQHSVLTQNPGP LSDKSAWVQGATGGKRYRWTTDR KVHLATGKVTHSFLHVPDCPYPLLG RDLLTKLKAQIHFEGSGAQVMGPM GQPLQVLTLNIEDEHRLHETSKEPDV	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			SLGSTWLSDFPQAWAETGGMGLAV RQAPLIIPLKATSTPVSIKQYPMSQEA RLGIKPHIQRLLDQGILVPCQSPWNT PLLPVKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYV DDLLLAATSELDCQQGTRALLQTLG NLGYRASAKKAQICQKQVKYLGYL LKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMA APLYPLTKTGTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVD EKQGYAKGVLTQKLGPWRRPVAYL SKKLDPVAAGWPPCLRMVAAIAVL TKDAGKLTMGQPLVILAPHAVEALV KQPPDRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEEGLQH NCLDILAEAHGTRPDLTDQPLPDAD HTWYTDGSSLLQEGQRKAGAAVTT ETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATA HIHGEIYRRRGLLTSEGKEIKNKDEIL ALLKALFLPKRLSIIHCPGHQKGHSA EARGNRMADQAARKAAITETPDTST LLIENSSPYTSEHFHYTVTDIKDLTKL GAIYDKTKKYWVYQGKPVMPDQFT FELLDFLHQLTHLSFSKMKALLERSH SPYYMLNRDRTLKNITETCKACAQV NASKSAVKQGTRVRGHRPGTHWEI DFTEIKPGLYGYKYLLVFIDTFSGWI EAFPTKKETAKVVTKKLLEEIFPRFG MPQVLGTDNGPAFVSKVSQTVADL LGIDWKLHCAYRPQSSGQVERMNR	

			TIKETLTKLTLATGSRDWVLLLPLAL YRARNTPGPHGLTPYEILYGAPPPLV NFPDPDMTRVTNSPSLQAHLQALYL VQHEVWRPLAAAYQEQLDRPVVPH PYRVGDTVWVRRHQTKNLEPRWKG PYTVLLTTPTALKVDGIAAWIHAAH VKAADPGGGPSSRLTWRVQRSQNPL KIRLTREAP (SEQ ID NO: 1571)	
POL_HTL1A - только остатки	P03362	Вирус Т-клеточного лейкоза человека 1 типа	MGQIFSRSASPIPRPPRGLAAHHWLN FLQAAYRLEPGPSSYDFHQLKKFLKI ALETPARICPINYSLLASLLPKGYPGR VNEILHILIQTQAQIPSRPAPPPPSSPT HDPPDSDPQIPPPYVEPTAPQVLPVM HPHGAPPNHRPWQMKDLQAIKQEV SQAAPGSPQFMQTIRLAVQQFDPTA KDLQDLLQYLCSSLVASLHHQQLDS LISEAETRGITGYNPLAGPLRVQANN PQQQGLRREYQQLWLAAFAALPGS AKDPSWASILQGLEEPYHAFVERLNI ALDNGLPEGTPKDPILRSLAYSNAN KECQKLLQARGHTNSPLGDMLRAC QTWTPKDKTKVLVVQPKKPPPNQPC FRCGKAGHWSRDCTQPRPPPGPCPL CQDPTHWKRDCPRLKPTIPEPEPEED ALLLDLPADIPHPKNLHRGGGLTSPP TLQQVLPNQDPASILPVIPLDPARRP VIKAQVDTQTSHPKTIEALLDTGAD MTVLPIALFSSNTPLKNTSVLGAGGQ TQDHFKLTSLPVLIRLPFRTTPIVLTS CLVDTKNNWAIIGRDALQQCQGVL YLPEAKRPPVILPIQAPAVLGLEHLP RPPQISQFPLNPERLQALQHLVRKAL EAGHIEPYTGPGNNPVFPVKKANGT WRFIHDLRATNSLTIDLSSSSPGPPDL SSLPTTLAHLQTIDLRDAFFQIPLPKQ	IPR043502, SSF56672, IPR000477, PF00078

			FQPYFAFTVPQQCNYGPGTRYAWK VLPQGFKNSPTLFEMQLAHILQPIRQ AFPQCTILQYMDDILLASPSHEDLLL LSEATMASLISHGLPVSENKTQQTPG TIKFLGQIISPNHLTYDAVPTVPIRSR WALPELQALLGEIQWVSKGTPTLRQ PLHSLYCALQRHTDPRDQIYLNPSQ VQSLVQLRQALSQNCRSRLVQTLPL LGAIMLTLTGTTTVVFQSKEQWPLV WLHAPLPHTSQCPWGQLLASAVLLL DKYTLQSYGLLCQTIHHNISTQTFNQ FIQTSDHPSVPILLHHSHRFKNLGAQ TGELWNTFLKTAAPLAPVKALMPVF TLSPVIINTAPCLFSDGSTSRAAYILW DKQILSQRSFPLPPPHKSAQRAELLG LLHGLSSARSWRCLNIFLDSKYLYH YLRTLALGTFQGRSSQAPFQALLPRL LSRKVVYLHHVRSHTNLPDPISRLN ALTDALLITPVLQLSPAELHSFTHCG QTALTLQGATTTEASNILRSCHACR GGNPQHQMPRGHIRRGLLPNHIWQG DITHFKYKNTLYRLHVWVDTFSGAI SATQKRKETSSEAISSLLQAIAHLGK PSYINTDNGPAYISQDFLNMCTSLAI RHTTHVPYNPTSSGLVERSNGILKTL LYKYFTDKPDLPMDNALSIALWTIN HLNVLTNCHKTRWQLHHSPRLQPIP ETRSLSNKQTHWYYFKLPGLNSRQ WKGPQEALQEAAGAALIPVSASSAQ WIPWRLLKRAACPRPVGGPADPKEK DLQHHG (SEQ ID NO: 1572)	
POL_FOAMV - только остатки	P14350	Спумаретровирус человека	MNPLQLLQPLPAEIKGTKLLAHWDS GATITCIPESFLEDEQPIKKTLIKTIHG EKQQNVYYVTFKVKGRKVEAEVIA SPYEYILLSPTDVPWLTQQPLQLTIL	IPR043502, SSF56672, IPR000477, PF00078

			VPLQEYQEKILSKTALPEDQKQQLK TLFVKYDNLWQHWENQVGHRKIRP HNIATGDYPPRPQKQYPINPKAKPSI QIVIDDLLKQGVLTPQNSTMNTPVY PVPKPDGRWRMVLDYREVNKTIPLT AAQNQHSAGILATIVRQKYKTTLDL ANGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFTADVVDL LKEIPNVQVYVDDIYLSHDDPKEHV QQLEKVFQILLQAGYVVSLKKSEIG QKTVEFLGFNITKEGRGLTDTFKTKL LNITPPKDLKQLQSILGLLNFARNFIP NFAELVQPLYNLIASAKGKYIEWSEE NTKQLNMVIEALNTASNLEERLPEQ RLVIKVNTSPSAGYVRYYNETGKKPI MYLNYVFSKAELKFSMLEKLLTTM HKALIKAMDLAMGQEILVYSPIVSM TKIQKTPLPERKALPIRWITWMTYLE DPRIQFHYDKTLPELKHIPDVYTSSQ SPVKHPSQYEGVFYTDGSAIKSPDPT KSNNAGMGIVHATYKPEYQVLNQW SIPLGNHTAQMAEIAAVEFACKKAL KIPGPVLVITDSFYVAESANKELPYW KSNGFVNNKKKPLKHISKWKSIAEC LSMKPDITIQHEKGISLQIPVFILKGN ALADKLATQGSYVVNCNTKKPNLD AELDQLLQGHYIKGYPKQYTYFLED GKVKVSRPEGVKIIPPQSDRQKIVLQ AHNLAHTGREATLLKIANLYWWPN MRKDVVKQLGRCQQCLITNASNKA SGPILRPDRPQKPFDKFFIDYIGPLPPS QGYLYVLVVVDGMTGFTWLYPTKA PSTSATVKSLNVLTSIAIPKVIHSDQG AAFTSSTFAEWAKERGIHLEFSTPYH PQSGSKVERKNSDIKRLLTKLLVGRP	

			TKWYDLLPVVQLALNNTYSPVLKY TPHQLLFGIDSNTPFANQDTLDLTRE EELSLLQEIRTSLYHPSTPPASSRSWS PVVGQLVQERVARPASLRPRWHKPS TVLKVLNPRTVVILDHLGNNRTVSID NLKPTSHQNGTTNDTATMDHLEKN E (SEQ ID NO: 1573)	
POL_BLVJ - только остатки	P03361	Вирус лейкоза крупного рогатого скота	MGNSPSYNPPAGISPSDWLNLLQSA QRLNPRPSPSDFTDLKNYIHWFHKT QKKPWTFTSGGPTSCPPGRFGRVPL VLATLNEVLSNEGGAPGASAPEEQP PPYDPPAILPIISEGNRNRHRAWALR ELQDIKKEIENKAPGSQVWIQTLRLA ILQADPTPADLEQLCQYIASPVDQTA HMTSLTAAIAAAEAANTLQGFNPKT GTLTQQSAQPNAGDLRSQYQNLWL QAGKNLPTRPSAPWSTIVQGPAESSV EFVNRLQISLADNLPDGVPKEPIIDSL SYANANRECQQILQGRGPVAAVGQ KLQACAQWAPKNKQPALLVHTPGP KMPGPRQPAPKRPPPGPCYRCLKEG HWARDCPTKATGPPPGPCPICKDPS HWKRDCPTLKSKNKLIEGGLSAPQTI TPITDSLSEAELECLLSIPLARSRPSV AVYLSGPWLQPSQNQALMLVDTGA ENTVLPQNWLVRDYPRIPAAVLGAG GVSRNRYNWLQGPLTLALKPEGPFI TIPKILVDTSDKWQILGRDVPSRLQA SISIPEEVRPPVVGVLDTPPSHIGLEH LPPPPEVPQFPLNLERLQALQDLVHR SLEAGYISPWDGPGNNPVFPVRKPN GAWRFVHDLRATNALTKPIPALSPG PPDLTAIPTHPPHIICLDLKDAFFQIPV EDRFRFYLSFTLPSPGGLQPHRRFAW RVLPQGFINSPALFERALQEPLRQVS	IPR043502, SSF56672, IPR000477, PF00078

			AAFSQSLLVSYMDDILYASPTEEQRS QCYQALAARLRDLGFQVASEKTSQT PSPVPFLGQMVHEQIVTYQSLPTLQI SSPISLHQLQAVLGDLQWVSRGTPTT RRPLQLLYSSLKRHHDPRAIIQLSPE QLQGIAELRQALSHNARSRYNEQEP LLAYVHLTRAGSTLVLFQKGAQFPL AYFQTPLTDNQASPWGLLLLLGCQY LQTQALSSYAKPILKYYHNLPKTSLD NWIQSSEDPRVQELLQLWPQISSQGI QPPGPWKTLITRAEVFLTPQFSPDPIP AALCLFSDGATGRGAYCLWKDHLL DFQAVPAPESAQKGELAGLLAGLAA APPEPVNIWVDSKYLYSLLRTLVLG AWLQPDPVPSYALLYKSLLRHPAIV VGHVRSHSSASHPIASLNNYVDQLL PLETPEQWHKLTHCNSRALSRWPNP RISAWDPRSPATLCETCQKLNPTGG GKMRTIQRGWAPNHIWQADITHYK YKQFTYALHVFVDTYSGATHASAK RGLTTQTTIEGLLEAIVHLGRPKKLN TDQGANYTSKTFVRFCQQFGVSLSH HVPYNPTSSGLDERTNGLLKLLLSK YHLDEPHLPMTQALSRALWTHNQIN LLPILKTRWELHHSPPLAVISEGGETP KGSDKLFLYLLPGQNNRRWLGPLPA LVEASGGALLATDPPVWVPWRLLK AFKCLKNDGPEDAHNRSSDG (SEQ ID NO: 1574)	
O41894_9RETR - только остатки	O41894	Вирус пенистости крупного рогатого скота	MPALRPLQVEIKGNHLKGYWDSGA EITCVPAIYIIEEQPVGKKLITTIHNEK EHDVYYVEMKIEKRKVQCEVIATAL DYVLVAPVDIPWYKPGPLELTIKIDV ESQKHTLITESTLSPQGQMRLKKLLD QYQALWQCWENQVGHRRIEPHKIA	IPR043502, SSF56672, IPR000477, PF00078

			TGALKPRPQKQYHINPRAKADIQIVI DDLLRQGVLRQQNSEMNTPVYPVP KADGRWRMVLDYREVNKVTPLVA TQNCHSASILNTLYRGPYKSTLDLA NGFWAHPIKPEDYWITAFTWGGKT YCWTVLPQGFLNSPALFTADVVDIL KDIPNVQVYVDDVYVSSATEQEHLD ILETIFNRLSTAGYIVSLKKSKLAKET VEFLGFSISQNGRGLTDSYKQKLMD LQPPTTLRQLQSILGLINFARNFLPNF AELVAPLYQLIPKAKGQCIPWTMDH TTQLKTIIQALNSTENLEERRPDVDLI MKVHISNTAGYIRFYNHGGQKPIAY NNALFTSTELKFTPTEKIMATIHKGL LKALDLSLGKEIHVYSAIASMTKLQ KTPLSERKALSIRWLKWQTYFEDPRI KFHHDATLPDLQNLPVPQQDTGKE MTILPLLHYEAIFYTDGSAIRSPKPNK THSAGMGIIQAKFEPDFRIVHLWSFP LGDHTAQYAEIAAFEFAIRRATGIRG PVLIVTDSNYVAKSYNEELPYWESN GFVNNKKKTLKHISKWKAIAECKNL KADIHVIHEPGHQPAEASPHAQGNA LADKQAVSGSYKVFSNELKPSLDAE LEQVLSTGRPNPQGYPNKYEYKLVN GLCYVDRRGEEGLKIIPPKADRVKL CQLAHDGPGSAHLGRSALLLKLQQK YWWPRMHIDASRIVLNCTVCAQTN STNQKPRPPLVIPHDTKPFQVWYMD YIGPLPPSNGYQHALVIVDAGTGFT WIYPTKAQTANATVKALTHLTGTA VPKVLHSDQGPAFTSSILADWAKDR GIQLEHSAPYHPQSSGKVERKNSEIK RLLTKLLAGRPTKWYPLIPIVQLALN NTPNTRQKYTPHQLMYGADCNLPFE	

			NLDTLDLTREEQLAVLKEVRDGLLD LYPSPSQTTARSWTPSPGLLVQERVA RPAQLRPKWRKPTPIKKVLNERTVII DHLGQDKVVSIDNLKPAAHQKLAQ TPDSAEICPSATPCPPNTSLWYDLDT GTWTCQRCGYQCPDKYHQPQCTWS CEDRCGHRWKECGNCIPQDGSSDD ASAVAAVEI (SEQ ID NO: 1575)	

Таблица 31. Иллюстративные димерные ретровирусные обратные транскриптазы и их сигнатуры доменов RT

Название	**Номер доступа**	**Организм**	**Последовательность**	**Сигнатуры RT**
Q83133_AVIMA	Q83133	Вирус, ассоциированный с миелобластозом птиц, 1 типа	RATVLTVALHLAIPLKWKPNHTPVW IDQWPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYRLL HDLRAVNAKLVPFGAVQQGAPVLS ALPRGWPLMVLDLKDCFFSIPLAEQ DREAFAFTLPSVNNQAPARRFQWKV LPQGMTCSPTICQLIVGQILEPLRLKH PSLRMLHYMDDLLLAASSHDGLEA AGEEVISTLERAGFTISPDKVQREPG VQYLGYKLGSTYVAPVGLVAEPRIA TLWDVQKLVGSLQSVRPALGIPPRL MGPFYEQLRGSDPNEAREWNLDMK MAWREIVQLSTTAALERWDPALPLE GAVARCEQGAIGVLGQGLSTHPRPC LWLFSTQPTKAFTAWLEVLTLLITKL RASAVRTFGKEVDILLLPACFREDLP LPEGILLALRGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDA SSSTHKGVVVWREGPRWEIKEIADL GASVQQLEARAVAMALLLWPTTPT NVVTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHVRSH SEVPGFFTEGNDVADSQATFQAYPL	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			REAKDLHTALHIGPRALSKACNISM QQAREVVQTCPHCNSAPALEAGVNP RGLGPLQIWQTDFTLEPRMAPRSWL AVTVDTASSAIVVTQHGRVTSVAAQ HHWATAIAVLGRPKAIKTDNGSCFT SKSTREWLARWGIAHTTGIPGNSQG QAMVERANRLLKDKIRVLAEGDGF MKRIPTSKQGELLAKAMYALNHFER GENTKTPIQKHWRPTVLTEGPPVKIR IETGEWEKGWNVLVWGRGYAAVK NRDTDKVIWVPSRKVKPDITQKDEV TKKDEASPLFAGISDWAPWEGEQEG LQEETASNKQERPGEDTPAANES (SEQ ID NO: 1576)	
POL_ SIVM 1	P05896	Вирус иммунодеф ицита обезьян	MGARNSVLSGKKADELEKIRLRPGG KKKYMLKHVVWAANELDRFGLAES LLENKEGCQKILSVLAPLVPTGSENL KSLYNTVCVIWCIHAEEKVKHTEEA KQIVQRHLVMETGTAETMPKTSRPT APFSGRGGNYPVQQIGGNYTHLPLS PRTLNAWVKLIEEKKFGAEVVSGFQ ALSEGCLPYDINQMLNCVGDHQAA MQIIRDIINEEAADWDLQHPQQAPQ QGQLREPSGSDIAGTTSTVEEQIQW MYRQQNPIPVGNIYRRWIQLGLQKC VRMYNPTNILDVKQGPKEPFQSYVD RFYKSLRAEQTDPAVKNWMTQTLLI QNANPDCKLVLKGLGTNPTLEEMLT ACQGVGGPGQKARLMAEALKEALA PAPIPFAAAQQKGPRKPIKCWNCGK EGHSARQCRAPRRQGCWKCGKMD HVMAKCPNRQAGFFRPWPLGKEAP QFPHGSSASGADANCSPRRTSCGSA KELHALGQAAERKQREALQGGDRG FAAPQFSLWRRPVVTAHIEGQPVEV	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			LLDTGADDSIVTGIELGPHYTPKIVG GIGGFINTKEYKNVEIEVLGKRIKGTI MTGDTPINIFGRNLLTALGMSLNLPI AKVEPVKSPLKPGKDGPKLKQWPLS KEKIVALREICEKMEKDGQLEEAPPT NPYNTPTFAIKKKDKNKWRMLIDFR ELNRVTQDFTEVQLGIPHPAGLAKR KRITVLDIGDAYFSIPLDEEFRQYTAF TLPSVNNAEPGKRYIYKVLPQGWKG SPAIFQYTMRHVLEPFRKANPDVTL VQYMDDILIASDRTDLEHDRVVLQL KELLNSIGFSSPEEKFQKDPPFQWMG YELWPTKWKLQKIELPQRETWTVN DIQKLVGVLNWAAQIYPGIKTKHLC RLIRGKMTLTEEVQWTEMAEAEYEE NKIILSQEQEGCYYQESKPLEATVIK SQDNQWSYKIHQEDKILKVGKFAKI KNTHTNGVRLLAHVIQKIGKEAIVI WGQVPKFHLPVEKDVWEQWWTDY WQVTWIPEWDFISTPPLVRLVFNLV KDPIEGEETYYVDGSCSKQSKEGKA GYITDRGKDKVKVLEQTTNQQAELE AFLMALTDSGPKANIIVDSQYVMGII TGCPTESESRLVNQIIEEMIKKTEIYV AWVPAHKGIGGNQEIDHLVSQGIRQ VLFLEKIEPAQEEHSKYHSNIKELVF KFGLPRLVAKQIVDTCDKCHQKGEA IHGQVNSDLGTWQMDCTHLEGKIVI VAVHVASGFIEAEVIPQETGRQTALF LLKLASRWPITHLHTDNGANFASQE VKMVAWWAGIEHTFGVPYNPQSQG VVEAMNHHLKNQIDRIREQANSVET IVLMAVHCMNFKRRGGIGDMTPAE RLINMITTEQEIQFQQSKNSKFKNFR VYYREGRDQLWKGPGELLWKGEG	

			AVILKVGTDIKVVPRRKAKIIKDYGG GKEMDSSSHMEDTGEAREVA (SEQ ID NO: 1577)	
POL_ RSVP	P03354	Вирус саркомы Рауса	MEAVIKVISSACKTYCGKTSPSKKEI GAMLSLLQKEGLLMSPSDLYSPGSW DPITAALSQRAMILGKSGELKTWGL VLGALKAAREEQVTSEQAKFWLGL GGGRVSPPGPECIEKPATERRIDKGE EVGETTVQRDAKMAPEETATPKTV GTSCYHCGTAIGCNCATASAPPPPY VGSGLYPSLAGVGEQQGQGGDTPPG AEQSRAEPGHAGQAPGPALTDWAR VREELASTGPPVVAMPVVIKTEGPA WTPLEPKLITRLADTVRTKGLRSPIT MAEVEALMSSPLLPHDVTNLMRVIL GPAPYALWMDAWGVQLQTVIAAAT RDPRHPANGQGRGERTNLNRLKGL ADGMVGNPQGQAALLRPGELVAIT ASALQAFREVARLAEPAGPWADIM QGPSESFVDFANRLIKAVEGSDLPPS ARAPVIIDCFRQKSQPDIQQLIRTAPS TLTTPGEIIKYVLDRQKTAPLTDQGI AAAMSSAIQPLIMAVVNRERDGQTG SGGRARGLCYTCGSPGHYQAQCPK KRKSGNSRERCQLCNGMGHNAKQC RKRDGNQGQRPGKGLSSGPWPGPEP PAVSLAMTMEHKDRPLVRVILTNTG SHPVKQRSVYITALLDSGADITIISEE DWPTDWPVMEAANPQIHGIGGGIPM RKSRDMIELGVINRDGSLERPLLLFP AVAMVRGSILGRDCLQGLGLRLTNL IGRATVLTVALHLAIPLKWKPDHTP VWIDQWPLPEGKLVALTQLVEKELQ LGHIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGAPV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			LSALPRGWPLMVLDLKDCFFSIPLAE QDREAFAFTLPSVNNQAPARRFQWK VLPQGMTCSPTICQLVVGQVLEPLR LKHPSLCMLHYMDDLLLAASSHDG LEAAGEEVISTLERAGFTISPDKVQR EPGVQYLGYKLGSTYVAPVGLVAEP RIATLWDVQKLVGSLQWLRPALGIP PRLMGPFYEQLRGSDPNEAREWNLD MKMAWREIVRLSTTAALERWDPAL PLEGAVARCEQGAIGVLGQGLSTHP RPCLWLFSTQPTKAFTAWLEVLTLLI TKLRASAVRTFGKEVDILLLPACFRE DLPLPEGILLALKGFAGKIRSSDTPSI FDIARPLHVSLKVRVTDHPVPGPTVF TDASSSTHKGVVVWREGPRWEIKEI ADLGASVQQLEARAVAMALLLWPT TPTNVVTDSAFVAKMLLKMGQEGV PSTAAAFILEDALSQRSAMAAVLHV RSHSEVPGFFTEGNDVADSQATFQA YPLREAKDLHTALHIGPRALSKACNI SMQQAREVVQTCPHCNSAPALEAG VNPRGLGPLQIWQTDFTLEPRMAPR SWLAVTVDTASSAIVVTQHGRVTSV AVQHHWATAIAVLGRPKAIKTDNGS CFTSKSTREWLARWGIAHTTGIPGNS QGQAMVERANRLLKDRIRVLAEGD GFMKRIPTSKQGELLAKAMYALNHF ERGENTKTPIQKHWRPTVLTEGPPV KIRIETGEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPDITQKD EVTKKDEASPLFAGISDWIPWEDEQ EGLQGETASNKQERPGEDTLAANES (SEQ ID NO: 1578)	
POL_ HV2	P15833	Вирус иммунодеф	MGARGSVLSGKKTDELEKVRLRPG GKKKYMLKHVVWAVNELDRFGLA	IPR043502, SSF56672,

D2		ицита человека 2 типа	ESLLESKEGCQKILKVLAPLVPTGSE NLKSLFNIVCVIFCLHAEEKVKDTEE AKKIAQRHLAADTEKMPATNKPTAP PSGGNYPVQQLAGNYVHLPLSPRTL NAWVKLVEEKKFGAEVVPGFQALS EGCTPYDINQMLNCVGEHQAAMQII REIINEEAADWDQQHPSPGPMPAGQ LRDPRGSDIAGTTSTVEEQIQWMYR AQNPVPVGNIYRRWIQLGLQKCVR MYNPTNILDIKQGPKEPFQSYVDRFY KSLRAEQTDPAVKNWMTQTLLIQN ANPDCKLVLKGLGMNPTLEEMLTA CQGIGGPGQKARLMAEALKEALTPA PIPFAAVQQKAGKRGTVTCWNCGK QGHTARQCRAPRRQGCWKCGKTGH IMSKCPERQAGFLRVRTLGKEASQL PHDPSASGSDTICTPDEPSRGHDTSG GDTICAPCRSSSGDAEKLHADGETTE REPRETLQGGDRGFAAPQFSLWRRP VVKACIEGQSVEVLLDTGVDDSIVA GIELGSNYTPKIVGGIGGFINTKEYK DVEIEVVGKRVRATIMTGDTPINIFG RNILNTLGMTLNFPVAKVEPVKVEL KPGKDGPKIRQWPLSREKILALKEIC EKMEKEGQLEEAPPTNPYNTPTFAIK KKDKNKWRMLIDFRELNKVTQDFT EVNWVFPTRQVAEKRRITVIDVGDA YFSIPLDPNFRQYTAFTLPSVNNAEP GKRYIYKVLPQGWKGSQSICQYSMR KVLDPFRKANSDVIIIQYMDDILIASD RSDLEHDRVVSQLKELLNDMGFSTP EEKFQKDPPFKWMGYELWPKKWKL QKIQLPEKEVWTVNAIQKLVGVLN WAAQLFPGIKTRHICKLIRGKMTLTE EVQWTELAEAELQENKIILEQEQEGS	IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			YYKERVPLEATVQKNLANQWTYKI HQGNKVLKVGKYAKVKNTHTNGV RLLAHVVQKIGKEALVIWGEIPVFHL PVERETWDQWWTDYWQVTWIPEW DFVSTPPLIRLAYNLVKDPLEGRETY YTDGSCNRTSKEGKAGYVTDRGKD KVKVLEQTTNQQAELEAFALALTDS EPQVNIIVDSQYVMGIIAAQPTETESP IVAKIIEEMIKKEAVYVGWVPAHKG LGGNQEVDHLVSQGIRQVLFLEKIEP AQEEHEKYHGNVKELVHKFGIPQLV AKQIVNSCDKCQQKGEAIHGQVNA DLGTWQMDCTHLEGKIIIVAVHVAS GFIEAEVIPQETGRQTALFLLKLASR WPITHLHTDNGANFTSPSVKMVAW WVGIEQTFGVPYNPQSQGVVEAMN HHLKNQIDRLRDQAVSIETVVLMAT HCMNFKRRGGIGDMTPAERLVNMIT TEQEIQFFQAKNLKFQNFQVYYREG RDQLWKGPGELLWKGEGAVIIKVG TEIKVVPRRKAKIIRHYGGGKGLDCS ADMEDTRQAREMAQSD (SEQ ID NO: 1579)	
POL_HV1A2	P03369	Вирус иммунодефицита человека 1 типа	MGARASVLSGGELDKWEKIRLRPG GKKKYKLKHIVWASRELERFAVNP GLLETSEGCRQILGQLQPSLQTGSEE LRSLYNTVATLYCVHQRIDVKDTKE ALEKIEEEQNKSKKKAQQAAAAAG TGNSSQVSQNYPIVQNLQGQMVHQ AISPRTLNAWVKVVEEKAFSPEVIPM FSALSEGATPQDLNTMLNTVGGHQA AMQMLKETINEEAAEWDRVHPVHA GPIAPGQMREPRGSDIAGTTSTLQEQ IGWMTNNPPIPVGEIYKRWIILGLNKI VRMYSPTSILDIRQGPKEPFRDYVDR	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659

			FYKTLRAEQASQDVKNWMTETLLV QNANPDCKTILKALGPAATLEEMMT ACQGVGGPGHKARVLAEAMSQVTN PANIMMQRGNFRNQRKTVKCFNCG KEGHIAKNCRAPRKKGCWRCGREG HQMKDCTERQANFLREDLAFLQGK AREFSSEQTRANSPTRRELQVWGGE NNSLSEAGADRQGTVSFNFPQITLW QRPLVTIRIGGQLKEALLDTGADDT VLEEMNLPGKWKPKMIGGIGGFIKV RQYDQIPVEICGHKAIGTVLVGPTPV NIIGRNLLTQIGCTLNFPISPIETVPVK LKPGMDGPKVKQWPLTEEKIKALVE ICTEMEKEGKISKIGPENPYNTPVFAI KKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDV GDAYFSVPLDKDFRKYTAFTIPSINN ETPGIRYQYNVLPQGWKGSPAIFQSS MTKILEPFRKQNPDIVIYQYMDDLY VGSDLEIGQHRTKIEELRQHLLRWG FTTPDKKHQKEPPFLWMGYELHPDK WTVQPIMLPEKDSWTVNDIQKLVG KLNWASQIYAGIKVKQLCKLLRGTK ALTEVIPLTEEAELELAENREILKEPV HEVYYDPSKDLVAEIQKQGQGQWT YQIYQEPFKNLKTGKYARMRGAHT NDVKQLTEAVQKVSTESIVIWGKIPK FKLPIQKETWEAWWMEYWQATWIP EWEFVNTPPLVKLWYQLEKEPIVGA ETFYVDGAANRETKLGKAGYVTDR GRQKVVSIADTTNQKTELQAIHLAL QDSGLEVNIVTDSQYALGIIQAQPDK SESELVSQIIEQLIKKEKVYLAWVPA HKGIGGNEQVDKLVSAGIRKVLFLN GIDKAQEEHEKYHSNWRAMASDFN	

			LPPVVAKEIVASCDKCQLKGEAMHG QVDCSPGIWQLDCTHLEGKIILVAV HVASGYIEAEVIPAETGQETAYFLLK LAGRWPVKTIHTDNGSNFTSTTVKA ACWWAGIKQEFGIPYNPQSQGVVES MNNELKKIIGQVRDQAEHLKTAVQ MAVFIHNFKRKGGIGGYSAGERIVDI IATDIQTKELQKQITKIQNFRVYYRD NKDPLWKGPAKLLWKGEGAVVIQD NSDIKVVPRRKAKIIRDYGKQMAGD DCVASRQDED (SEQ ID NO: 1580)	
POL_ FIVP E	P16088	Вирус иммунодеф ицита кошек	KEFGKLEGGASCSPSESNAASSNAIC TSNGGETIGFVNYNKVGTTTTLEKR PEILIFVNGYPIKFLLDTGADITILNRR DFQVKNSIENGRQNMIGVGGGKRGT NYINVHLEIRDENYKTQCIFGNVCVL EDNSLIQPLLGRDNMIKFNIRLVMAQ ISDKIPVVKVKMKDPNKGPQIKQWP LTNEKIEALTEIVERLEKEGKVKRAD SNNPWNTPVFAIKKKSGKWRMLIDF RELNKLTEKGAEVQLGLPHPAGLQI KKQVTVLDIGDAYFTIPLDPDYAPYT AFTLPRKNNAGPGRRFVWCSLPQG WILSPLIYQSTLDNIIQPFIRQNPQLDI YQYMDDIYIGSNLSKKEHKEKVEEL RKLLLWWGFETPEDKLQEEPPYTW MGYELHPLTWTIQQKQLDIPEQPTL NELQKLAGKINWASQAIPDLSIKALT NMMRGNQNLNSTRQWTKEARLEV QKAKKAIEEQVQLGYYDPSKELYAK LSLVGPHQISYQVYQKDPEKILWYG KMSRQKKKAENTCDIALRACYKIRE ESIIRIGKEPRYEIPTSREAWESNLINS PYLKAPPPEVEYIHAALNIKRALSMI KDAPIPGAETWYIDGGRKLGKAAKA	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			AYWTDTGKWRVMDLEGSNQKAEIQ ALLLALKAGSEEMNIITDSQYVINIIL QQPDMMEGIWQEVLEELEKKTAIFI DWVPGHKGIPGNEEVDKLCQTMMII EGDGILDKRSEDAGYDLLAAKEIHL LPGEVKVIPTGVKLMLPKGYWGLII GKSSIGSKGLDVLGGVIDEGYRGEIG VIMINVSRKSITLMERQKIAQLIILPC KHEVLEQGKVVMDSERGDNGYGST GVFSSWVDRIEEAEINHEKFHSDPQY LRTEFNLPKMVAEEIRRKCPVCRIIG EQVGGQLKIGPGIWQMDCTHFDGKI ILVGIHVESGYIWAQIISQETADCTV KAVLQLLSAHNVTELQTDNGPNFKN QKMEGVLNYMGVKHKFGIPGNPQS QALVENVNHTLKVWIQKFLPETTSL DNALSLAVHSLNFKRRGRIGGMAPY ELLAQQESLRIQDYFSAIPQKLQAQ WIYYKDQKDKKWKGPMRVEYWGQ GSVLLKDEEKGYFLIPRRHIRRVPEP CALPEGDE (SEQ ID NO: 1581)	
POL_ EIAV Y	P03371	Вирус инфекцион ной анемии лошадей	TAWTFLKAMQKCSKKREARGSREA PETNFPDTTEESAQQICCTRDSSDSK SVPRSERNKKGIQCQGEGSSRGSQPG QFVGVTYNLEKRPTTIVLINDTPLNV LLDTGADTSVLTTAHYNRLKYRGR KYQGTGIIGVGGNVETFSTPVTIKKK GRHIKTRMLVADIPVTILGRDILQDL GAKLVLAQLSKEIKFRKIELKEGTM GPKIPQWPLTKEKLEGAKETVQRLL SEGKISEASDNNPYNSPIFVIKKRSGK WRLLQDLRELNKTVQVGTEISRGLP HPGGLIKCKHMTVLDIGDAYFTIPLD PEFRPYTAFTIPSINHQEPDKRYVWK CLPQGFVLSPYIYQKTLQEILQPFRER	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			YPEVQLYQYMDDLFVGSNGSKKQH KELIIELRAILQKGFETPDDKLQEVPP YSWLGYQLCPENWKVQKMQLDMV KNPTLNDVQKLMGNITWMSSGVPG LTVKHIAATTKGCLELNQKVIWTEE AQKELEENNEKIKNAQGLQYYNPEE EMLCEVEITKNYEATYVIKQSQGIL WAGKKIMKANKGWSTVKNLMLLL QHVATESITRVGKCPTFKVPFTKEQV MWEMQKGWYYSWLPEIVYTHQVV HDDWRMKLVEEPTSGITIYTDGGKQ NGEGIAAYVTSNGRTKQKRLGPVTH QVAERMAIQMALEDTRDKQVNIVT DSYYCWKNITEGLGLEGPQNPWWPI IQNIREKEIVYFAWVPGHKGIYGNQL ADEAAKIKEEIMLAYQGTQIKEKRD EDAGFDLCVPYDIMIPVSDTKIIPTDV KIQVPPNSFGWVTGKSSMAKQGLLI NGGIIDEGYTGEIQVICTNIGKSNIKLI EGQKFAQLIILQHHSNSRQPWDENKI SQRGDKGFGSTGVFWVENIQEAQDE HENWHTSPKILARNYKIPLTVAKQIT QECPHCTKQGSGPAGCVMRSPNHW QADCTHLDNKIILHFVESNSGYIHAT LLSKENALCTSLAILEWARLFSPKSL HTDNGTNFVAEPVVNLLKFLKIAHT TGIPYHPESQGIVERANRTLKEKIQS HRDNTQTLEAALQLALITCNKGRES MGGQTPWEVFITNQAQVIHEKLLLQ QAQSSKKFCFYKIPGEHDWKGPTRV LWKGDGAVVVNDEGKGIIAVPLTRT KLLIKPN (SEQ ID NO: 1582)	
POL_ BIV2 9	P19560	Вирус иммунодеф ицита	MKRRELEKKLRKVRVTPQQDKYYTI GNLQWAIRMINLMGIKCVCDEECSA AEVALIITQFSALDLENSPIRGKEEVA	IPR043502, SSF56672, IPR000477,

		крупного рогатого скота	IKNTLKVFWSLLAGYKPESTETALG YWEAFTYREREARADKEGEIKSIYPS LTQNTQNKKQTSNQTNTQSLPAITT QDGTPRFDPDLMKQLKIWSDATERN GVDLHAVNILGVITANLVQEEIKLLL NSTPKWRLDVQLIESKVREKENAHR TWKQHHPEAPKTDEIIGKGLSSAEQ ATLISVECRETFRQWVLQAAMEVAQ AKHATPGPINIHQGPKEPYTDFINRL VAALEGMAAPETTKEYLLQHLSIDH ANEDCQSILRPLGPNTPMEKKLEAC RVVGSQKSKMQFLVAAMKEMGIQS PIPAVLPHTPEAYASQTSGPEDGRRC YGCGKTGHLKRNCKQQKCYHCGKP GHQARNCRSKNREVLLCPLWAEEPT TEQFSPEQHEFCDPICTPSYIRLDKQP FIKVFIGGRWVKGLVDTGADEVVLK NIHWDRIKGYPGTPIKQIGVNGVNV AKRKTHVEWRFKDKTGIIDVLFSDT PVNLFGRSLLRSIVTCFTLLVHTEKIE PLPVKVRGPGPKVPQWPLTKEKYQA LKEIVKDLLAEGKISEAAWDNPYNT PVFVIKKKGTGRWRMLMDFRELNKI TVKGQEFSTGLPYPPGIKECEHLTAI DIKDAYFTIPLHEDFRPFTAFSVVPV NREGPIERFQWNVLPQGWVCSPAIY QTTTQKIIENIKKSHPDVMLYQYMD DLLIGSNRDDHKQIVQEIRDKLGSYG FKTPDEKVQEERVKWIGFELTPKKW RFQPRQLKIKNPLTVNELQQLVGNC VWVQPEVKIPLYPLTDLLRDKTNLQ EKIQLTPEAIKCVEEFNLKLKDPEWK DRIREGAELVIKIQMVPRGIVFDLLQ DGNPIWGGVKGLNYDHSNKIKKILR TMNELNRTVVIMTGREASFLLPGSSE	PF00078, PF06817, IPR010661

			DWEAALQKEESLTQIFPVKFYRHSC RWTSICGPVRENLTTYYTDGGKKGK TAAAVYWCEGRTKSKVFPGTNQQA ELKAICMALLDGPPKMNIITDSRYAY EGMREEPETWAREGIWLEIAKILPFK QYVGVGWVPAHKGIGGNTEADEGV KKALEQMAPCSPPEAILLKPGEKQN LETGIYMQGLRPQSFLPRADLPVAIT GTMVDSELQLQLLNIGTEHIRIQKDE VFMTCFLENIPSATEDHERWHTSPDI LVRQFHLPKRIAKEIVARCQECKRTT TSPVRGTNPRGRFLWQMDNTHWNK TIIWVAVETNSGLVEAQVIPEETALQ VALCILQLIQRYTVLHLHSDNGPCFT AHRIENLCKYLGITKTTGIPYNPQSQ GVVERAHRDLKDRLAAYQGDCETV EAALSLALVSLNKKRGGIGGHTPYEI YLESEHTKYQDQLEQQFSKQKIEKW CYVRNRRKEWKGPYKVLWDGDGA AVIEEEGKTALYPHRHMRFIPPPDSD IQDGSS (SEQ ID NO: 1583)	
A0A142BKH1_ALV	A0A142BKH1	Вирус лейкоза и саркомы птиц	TVALHLAIPLKWKPDHTPVWIDQWP LPEGKLVALTQLVEKELQLGHIEPSL SCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRG WPLMVLDLKDCFFSIPLAEQDREAF AFTLPSVNNQAPARRFQWKVLPQG MTCSPTICQLVVGQVLEPLRLKHPSL RMLHYMDDLLLAASSHDGLEAAGE EVISTLERAGFTISPDKIQREPGVQYL GYKLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPPRLMGPF YEQLRGSDPNEAREWNLDMKMAW REIVQLSTTAALERWDPALPLEGAV ARCEQGAIGVLGQGLSTHPRPCLWL	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			FSTQPTKAFTAWLEVLTLLITKLRAS AVRTFGKEVDVLLLPACFREDLPLPE GILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSS THKGVVVWREGPRWEIKEIADLGAS VQQLEARAVAMALLLWPTTPTNVV TDSAFVAKMLLKMGQEGVPSTAAA FILEDALSQRSAMAAVLHVRSHSEV PGFFTEGNDVADSQATFQAYPLREA KDLHTALHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAGVNPRGL GPLQIWQTDFTLEPRMAPRSWLAVT VATASSAIVVTQHGRVTSVAARHH WATAIAVLGRPKAIKTDNGSCFTSK STREWLARWGIAHTTGIPGNSQGQA MVERANRLLKDKIRVLAEGDGFMK RIPTGKQGELLAKAMYALNHFERGE NTKTPIQKHWRPTVLTEGPPVKIRIE TGEWEKGWNVLVWGRGYAAVKNR DTDKIIWVPSRKVKPDITQKDELTKK DEASPLFAGISDWAPWKGEQEGL (SEQ ID NO: 1584)	

Таблица 32. Описания сигнатур InterPro, присутствующих в обратных транскриптазах из таблицы 30 (мономерных вирусных RT) и из таблицы 31 (димерных вирусных RT)

Сигнатура	**База данных**	**Краткое название**	**Описание**
cd01645	CDD	*RT_Rtv*	RT_Rtv: обратные транскриптазы (RT) ретровирусов (Rtv). RT катализируют превращение однонитевой РНК в двухнитевую вирусную ДНК для интеграции в хромосомы хозяина. Белки этого подсемейства содержат длинные

			концевые повторы (LTR) и представляют собой многофункциональные ферменты с гибридной активностью РНК-направленной ДНК-полимеразы, ДНК-направленной ДНК-полимеразы и рибонуклеазы (РНКазы H). Вирусный РНК-геном входит в цитоплазму в качестве части нуклеопротеинового комплекса, и в цитоплазме организуется процесс обратной транскрипции с образованием линейного ДНК-дуплекса посредством сложного ряда стадий. Этот ДНК-дуплекс является коллинеарным к своей РНК-матрице, но содержит концевые дупликации, известные как LTR, которые отсутствуют в вирусной РНК. Было высказано предположение, что для создания LTR необходимы два специализированных переключателя матрицы, известных в качестве реакций переноса нити или "прыжков". [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]
cd03715	CDD	*RT_ZFREV_подобные*	RT_ZFREV_подобные: подсемейство обратных

			транскриптаз (RT), обнаруженное в последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ

			позволяет предположить, что ZFERV принадлежит к отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
PF00078	Pfam	*RVT_1*	Ген обратной транскриптазы обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. Обратные транскриптазы встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. [PMID: 1698615]
IPR000477	InterPro	*RT_домен*	Применение РНК-матрицы для получения ДНК с целью интеграции в геном хозяина и использования клетки-хозяина является стратегией, применяемой при репликации ретроидных элементов, таких как ретровирусы и бактериальные ретроны. Фермент, катализирующий полимеризацию, представляет собой РНК-направленную ДНК-полимеразу или обратную транскриптазу (RT) (2.7.7.49).

			Обратная транскриптаза встречается в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II [PMID: 12758069], бактериальные msDNA, гепаднавирусы и каулимовирусы. Ретровирусная обратная транскриптаза синтезируется в качестве части полипротеина POL, который содержит: аспартилпротеазу, обратную транскриптазу, РНКазу H и интегразу. Полипротеин POL подвергается специфическому ферментативному расщеплению с образованием зрелых белков. Обнаружение ретроэлементов у прокариот поднимает интригующие вопросы, касающиеся их роли у бактерий, происхождения и эволюции обратных транскриптаз, а также того, являются ли бактериальные обратные транскриптазы старшими, чем эукариотические обратные транскриптазы [PMID: 8828137]. Было определено несколько кристаллических структур домена обратной транскриптазы (RT) [PMID: 1377403].

IPR043502	InterPro	*Суперсемейство ДНК/РНК-полимераз*	Эта запись представляет суперсемейство ДНК/РНК-полимераз, которое включает ДНК-полимеразу I, обратную транскриптазу, РНК-полимеразу T7, ДНК-полимеразу, обходящую повреждения (Y-семейство), РНК-зависимую РНК-полимеразу и РНК-зависимую РНК-полимеразу dsRNA-фага. Данные ферменты имеют сходную белковую укладку в своем активном центре, которая напоминает субдомен "ладонь" полимераз в форме правой руки. [PMID: 26931141]
SSF56672	Суперсемейство	*ДНК/РНК-полимеразы*	Данное суперсемейство включает ДНК-полимеразы и РНК-полимеразы.
PF06817	Pfam	*RVT_большой палец*	Данный домен известен в качестве домена "большой палец". Он состоит из четырехспирального пучка [PMID: 1377403].
IPR010661	InterPro	*RVT_большой палец*	Данный домен известен в качестве домена "большой палец". Он состоит из четырехспирального пучка. Обратная транскриптаза превращает вирусный РНК-геном в двухнитевую вирусную ДНК. Обратная транскриптаза часто встречается в полипротеине с 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 C-концевые домены Pol, которые включают домен "большой палец", соединительный домен и домен РНКазы H. Существующие данные свидетельствуют о том, что это могут быть весьма консервативные домены. [PMID: 1377403, PMID: 18335052]
PF06815	Pfam	*RVT_соединительный*	Данный домен известен в качестве соединительного домена. Данный домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
IPR010659	InterPro	*RVT_соединительный*	Данный домен известен в качестве соединительного домена. Данный домен находится между доменами "большой палец" и "ладонь" [PMID: 1377403].
cd03715	CDD	*RT_ZFREV_подобные*	RT_ZFREV_подобные: подсемейство обратных транскриптаз (RT), обнаруженное в

			последовательностях, сходных с интактным эндогенным ретровирусом ZFERV у данио-рерио и с RT вируса лейкоза мышей Молони. Ген RT обычно указывает на мобильный элемент, такой как ретротранспозон или ретровирус. RT встречаются в различных мобильных элементах, включая ретротранспозоны, ретровирусы, интроны группы II, бактериальные msDNA, гепаднавирусы и каулимовирусы. Эти элементы можно разделить на две большие группы. Одна группа содержит ретровирусы и ДНК-вирусы, в размножении которых участвует промежуточная РНК. Они сгруппированы вместе с подвижными элементами, содержащими длинные концевые повторы (LTR). Другая группа, также называемая ретротранспозонами поли(А)-типа, содержит грибные митохондриальные интроны и подвижные элементы, в которых отсутствуют LTR. Филогенетический анализ позволяет предположить, что ZFERV принадлежит к

			отдельной группе ретровирусов. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]

Эндонуклеазный домен

В определенных вариантах осуществления эндонуклеазный/ДНК-связывающий домен ретротранспозона APE-типа или эндонуклеазный домен ретротранспозона RLE-типа может использоваться или может быть модифицирован (например, посредством вставки, делеции или замены одного или нескольких остатков) в системе Gene Writer, описанной в данном документе. В некоторых вариантах осуществления эндонуклеазный домен или эндонуклеазный/ДНК-связывающий домен изменен по сравнению с его природной последовательностью таким образом, что он характеризуется измененной частотой использования кодонов, например, улучшенной для применения в клетках человека. В некоторых вариантах осуществления эндонуклеазный элемент представляет собой гетерологичный эндонуклеазный элемент, такой как нуклеаза Fok1, эндонуклеаза, подобная ферменту рестрикции 1 II типа (нуклеаза RLE-типа), или другая эндонуклеаза RLE-типа (также известная как REL). В некоторых вариантах осуществления гетерологичная эндонуклеаза обладает никазной активностью и не образует двухнитевые разрывы. В некоторых вариантах осуществления гетерологичная эндонуклеаза представляет собой нуклеазу, ассоциированную с CRISPR, например, Cas9, или нуклеазу, ассоциированную с CRISPR, с никазной активностью, например, никазу Cas9. Аминокислотная последовательность эндонуклеазного домена системы Gene Writer, описанной в данном документе, может являться на по меньшей мере приблизительно 50%, на по меньшей мере приблизительно 60%, на по меньшей мере приблизительно 70%, на по меньшей мере приблизительно 80%, на по меньшей мере приблизительно 85%, на по меньшей мере приблизительно 90%, на по меньшей мере приблизительно 95%, на по меньшей мере приблизительно 96%, на по меньшей мере приблизительно 97%, на по меньшей мере приблизительно 98%, на по меньшей мере приблизительно 99% идентичной аминокислотной последовательности эндонуклеазного домена ретротранспозона, последовательность ДНК которого указана в таблице 1, 2 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 Fok1 или ее функциональный фрагмент. В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой резольвазу для структур Холлидея или ее гомолог, такой как фермент, разрешающий структуры Холлидея, из Sulfolobus solfataricus -- Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). В определенных вариантах

осуществления гетерологичная эндонуклеаза представляет собой эндонуклеазу из большого фрагмента белка сплайсосомы, такого как Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). Например, полипептид Gene Writer, описанный в данном документе, может содержать домен обратной транскриптазы из ретротранспозона APE- или RLE-типа и эндонуклеазный домен, который содержит Fok1 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 дополнение к однонитевому разрыву в сайте-мишени, который необходим для инициации обратной транскрипции с праймированием мишенью, дополнительная эндонуклеазная активность может быть благоприятной для улучшения разрешения события интеграции (Anzalone et al., Nature 576, 149-157 (2019)). В некоторых вариантах осуществления эндонуклеазный элемент полипептида обеспечивает однонитевой разрыв для инициации обратной транскрипции с праймированием мишенью, и дополнительный гетерологичный домен полипептида обеспечивает дополнительную эндонуклеазную активность. В некоторых вариантах осуществления дополнительная эндонуклеазная активность обеспечивается никазой. В некоторых вариантах осуществления дополнительная эндонуклеазная активность может обеспечиваться гетерологичным ДНК-связывающим элементом, который также обладает эндонуклеазной активностью, например, никазой Cas9. В некоторых вариантах осуществления дополнительная эндонуклеазная активность может содержаться в первом полипептиде Gene Writer. В некоторых вариантах осуществления дополнительная эндонуклеазная активность может быть обеспечена отдельным полипептидом.

В некоторых вариантах осуществления полипептид Gene Writer, описанный в данном документе, содержит эндонуклеазный домен, который предусматривает расщепление в заданном месте в последовательности ДНК-мишени, например, как измерено с использованием анализа из примера 32 в данном документе. В некоторых вариантах осуществления эндонуклеазный домен предусматривает расщепление в сайте GG последовательности ДНК-мишени. В некоторых вариантах осуществления эндонуклеазный домен предусматривает расщепление в сайте AAGG последовательности ДНК-мишени. В некоторых вариантах осуществления последовательность ДНК-мишени, описанная в данном документе, содержит мотив GG или AAGG, например, встречающийся в природе мотив в человеческом геноме.

ДНК-связывающий домен

В определенных аспектах ДНК-связывающий домен полипептида Gene Writer, описанного в данном документе, выбран, разработан или сконструирован для связывания с требуемой последовательностью ДНК-мишени хозяина. В определенных вариантах осуществления ДНК-связывающий домен сконструированной RLE 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 активность расщепления только однонитевой ДНК, например, представляет собой ДНК-никазу, например, представляет собой никазу Cas9. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент с эндонуклеазной активностью замещает эндонуклеазный элемент полипептида. В некоторых вариантах осуществления гетерологичный ДНК-связывающий элемент с эндонуклеазной активностью дополняет эндонуклеазный элемент полипептида, например, вызывает дополнительный однонитевой разрыв в сайте-мишени. 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, описанного в таблице данного документа. 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 полипептид, описанный в данном документе, содержит мутацию в ДНК-связывающем домене. В некоторых вариантах осуществления мутация приводит к снижению или устранению ДНК-связывающей активности ДНК-связывающего домена, например, до менее чем 50%, 40%, 30%, 20%, 10%, 5%, 2% или 1% от активности соответствующей последовательности дикого типа, например, в анализе из примера 30. Мутация может находиться, например, в домене ZF1, домене ZF2 или домене c-Myb. Мутация может представлять собой точечную мутацию. Мутация может находиться

в остатке C (например, замена C на S), например, в остатке C в домене ZF1 или ZF2; в остатке R (например, замена R на A), например, в остатке R в домене c-Myb; или в остатке W (например, замена W на A), например, в остатке W в домене c-myb; или в любой их комбинации. В некоторых вариантах осуществления полипептид, содержащий мутацию в ДНК-связывающем домене, дополнительно содержит гетерологичный ДНК-связывающий домен.

В некоторых вариантах осуществления природная последовательность AAGG в геноме используется в качестве затравки для перенацеливания системы Gene Writing на основе ретротранспозазы R2, где ДНК-связывающий домен подвергнут мутации или заменен гетерологичным ДНК-связывающим доменом таким образом, что связывание полипептида Gene Writer с новым сайтом-мишенью приводит к правильному расположению эндонуклеазного домена в мотиве AAGG для обеспечения эндонуклеазной активности. В некоторых вариантах осуществления последовательность ДНК-мишени, описанная в данном документе, содержит мотив, распознаваемый эндонуклеазным доменом (например, мотив GG или AAGG), например, встречающийся в природе мотив в человеческом геноме. В некоторых вариантах осуществления Gene Writer содержит ДНК-связывающий домен (например, гетерологичный ДНК-связывающий домен), который связывается рядом с мотивом, распознаваемым эндонуклеазным доменом, например, таким образом, что эндонуклеазный домен Gene Writer расположен так, чтобы расщеплять мотив. В некоторых вариантах осуществления ДНК-связывающий домен связывается с сайтом, который находится в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 или 60 нуклеотидов от мотива, распознаваемого эндонуклеазным доменом (например, мотива GG или AAGG). ДНК-связывающий домен может связываться с сайтом, расположенным выше или ниже мотива GG или AAGG. ДНК-связывающий домен может связываться с сайтом, имеющим ту же ориентацию или обратно комплементарную ориентацию по сравнению с мотивом, распознаваемым эндонуклеазным доменом (например, мотивом GG или AAGG). В некоторых вариантах осуществления перенацеленный полипептид Gene Writer содержит (i) эндонуклеазный домен, который распознает мотив, и (ii) гетерологичный ДНК-связывающий домен, который распознает последовательность геномной ДНК. В некоторых вариантах осуществления мотив находится на приблизительно 30-80, 40-70, 50-60 или 55 нуклеотидов выше последовательности геномной ДНК, где необязательно мотив и последовательность геномной ДНК имеют одну и ту же ориентацию. В некоторых вариантах осуществления мотив находится на приблизительно 10-30, 15-25 или 20 нуклеотидов ниже последовательности геномной ДНК, где мотив необязательно имеет обратную ориентацию относительно последовательности геномной ДНК. 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 сконструированная RLE 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 M-область MALAT1 длиной 600 нуклеотидов (описанную в Miyagawa et al., *RNA* 18, (738-751), 2012). В некоторых вариантах осуществления сигнал ядерной локализации получен из длинной некодирующей РНК BORG или представляет собой мотив AGCCC (описанный в Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014)). В некоторых вариантах осуществления последовательность ядерной локализации описана в Shukla et al., *The EMBO Journal* e98452 (2018). В некоторых вариантах осуществления сигнал ядерной локализации получен из не содержащего LTR ретротранспозона, ретротранспозона, содержащего LTR, ретровируса или эндогенного ретровируса.

В некоторых вариантах осуществления полипептид, описанный в данном документе, содержит одну или несколько (например, 2, 3, 4, 5) последовательностей, нацеленных на ядро, например, последовательность ядерной локализации (NLS), например, как описано выше. В некоторых вариантах осуществления NLS представляет собой двухкомпонентную

NLS. В некоторых вариантах осуществления NLS облегчает импорт белка, содержащего NLS, в ядро клетки. В некоторых вариантах осуществления NLS слита с N-концом полипептида Gene Writer, описанного в данном документе. В некоторых вариантах осуществления NLS слита с C-концом полипептида Gene Writer. В некоторых вариантах осуществления NLS слита с N-концом или C-концом домена Cas. В некоторых вариантах осуществления линкерная последовательность расположена между NLS и соседним доменом полипептида Gene Writer.

В некоторых вариантах осуществления NLS содержит аминокислотную последовательность MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 1585), PKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1586), RKSGKIAAIWKRPRKPKKKRKV (SEQ ID NO: 1587), KRTADGSEFESPKKKRKV (SEQ ID NO: 1588), KKTELQTTNAENKTKKL (SEQ ID NO: 1589) или KRGINDRNFWRGENGRKTR (SEQ ID NO: 1590), KRPAATKKAGQAKKKK (SEQ ID NO: 1591) или ее функциональный фрагмент или вариант. Примеры последовательностей NLS также описаны в документе PCT/EP2000/011690, содержание которого включено в данный документ посредством ссылки для раскрытия иллюстративных последовательностей ядерной локализации. В некоторых вариантах осуществления NLS содержит аминокислотную последовательность, раскрытую в таблице 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, внутри пептидных доменов, между пептидными доменами, в C-концевом домене или в комбинации местоположений, чтобы улучшить субклеточную локализацию в ядре. В одном полипептиде можно использовать несколько уникальных последовательностей. Последовательности могут быть в естественных условиях однокомпонентными или двухкомпонентными, например, иметь один или два отрезка из основных аминокислот, или могут использоваться в качестве химерных двухкомпонентных последовательностей. Ссылки на последовательности соответствуют номерам доступа UniProt, за исключением тех случаев, когда они указаны как SeqNLS для последовательностей, полученных с использованием алгоритма предсказания субклеточной локализации (Lin et al BMC Bioinformat 13:157 (2012), включенный в данный документ посредством ссылки во всей своей полноте).

Таблица 39. Иллюстративные сигналы ядерной локализации для применения в системах Gene Writing

Последовательность	**Ссылки на последовательности**	**SEQ ID NO:**
AHFKISGEKRPSTDPGKKAKNPKKKKKKDP	Q76IQ7	1823
AHRAKKMSKTHA	P21827	1824
ASPEYVNLPINGNG	SeqNLS	1825

CTKRPRW	O88622, Q86W56, Q9QYM2, O02776	1826
DKAKRVSRNKSEKKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4	1827
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15	1828
EEQLRRRKNSRLNNTG	G5EFF5	1829
EVLKVIRTGKRKKKAWKRMVTKVC	SeqNLS	1830
HHHHHHHHHHHHQPH	Q63934, G3V7L5, Q12837	1831
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0	1832
HKRTKK	Q2R2D5	1833
IINGRKLKLKKSRRRSSQTSNNSFTSRRS	SeqNLS	1834
KAEQERRK	Q8LH59	1835
KEKRKRREELFIEQKKRK	SeqNLS	1836
KKGKDEWFSRGKKP	P30999	1837
KKGPSVQKRKKT	Q6ZN17	1838
KKKTVINDLLHYKKEK	SeqNLS, P32354	1839
KKNGGKGKNKPSAKIKK	SeqNLS	1840
KKPKWDDFKKKKK	Q15397, Q8BKS9, Q562C7	1841
KKRKKD	SeqNLS, Q91Z62, Q1A730, Q969P5, Q2KHT6, Q9CPU7	1842
KKRRKRRRK	SeqNLS	1843
KKRRRRARK	Q9UMS6, D4A702, Q91YE8	1844
KKSKRGR	Q9UBS0	1845
KKSRKRGS	B4FG96	1846
KKSTALSRELGKIMRRR	SeqNLS, P32354	1847
KKSYQDPEIIAHSRPRK	Q9U7C9	1848
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2	1849

KKVSIAGQSGKLWRWKR	Q6YUL8	1850
KKYENVVIKRSPRKRGRPRK	SeqNLS	1851
KNKKRK	SeqNLS	1852
KPKKKR	SeqNLS	1853
KRAMKDDSHGNSTSPKRRK	Q0E671	1854
KRANSNLVAAYEKAKKK	P23508	1855
KRASEDTTSGSPPKKSSAGPKR	Q9BZZ5, Q5R644	1856
KRFKRRWMVRKMKTKK	SeqNLS	1857
KRGLNSSFETSPKKVK	Q8IV63	1858
KRGNSSIGPNDLSKRKQRKK	SeqNLS	1859
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS	1860
KRKGKLKNKGSKRKK	O15381	1861
KRRRRRRREKRKR	Q96GM8	1862
KRSNDRTYSPEEEKQRRA	Q91ZF2	1863
KRTVATNGDASGAHRAKKMSK	SeqNLS	1864
KRVYNKGEDEQEHLPKGKKR	SeqNLS	1865
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524	1866
KVNFLDMSLDDIIIYKELE	Q9P127	1867
KVQHRIAKKTTRRRR	Q9DXE6	1868
LSPSLSPL	Q9Y261, P32182, P35583	1869
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	Q9GZX7	1870
MPQNEYIELHRKRYGYRLDYHEKKRKKESREAHERSKKAKKMIGLKAKLYHK	SeqNLS	1871
MVQLRPRASR	SeqNLS	1872
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5	1873
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40	1874
PDTKRAKLDSSETTMVKKK	SeqNLS	1875
PEKRTKI	SeqNLS	1876
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5	1877
PGKMDKGEHRQERRDRPY	Q01844, Q61545	1878
PKKGDKYDKTD	Q45FA5	1879

PKKKSRK	O35914, Q01954	1880
PKKNKPE	Q22663	1881
PKKRAKV	P04295, P89438	1882
PKPKKLKVE	P55263, P55262, P55264, Q64640	1883
PKRGRGR	Q9FYS5, Q43386	1884
PKRRLVDDA	P0C797	1885
PKRRRTY	SeqNLS	1886
PLFKRR	A8X6H4, Q9TXJ0	1887
PLRKAKR	Q86WB0, Q5R8V9	1888
PPAKRKCIF	Q6AZ28, O75928, Q8C5D8	1889
PPARRRRL	Q8NAG6	1890
PPKKKRKV	Q3L6L5, P03070, P14999, P03071	1891
PPNKRMKVKH	Q8BN78	1892
PPRIYPQLPSAPT	P0C799	1893
PQRSPFPKSSVKR	SeqNLS	1894
PRPRKVPR	P0C799	1895
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9	1896
PRRVRLK	Q58DJ0, P56477, Q13568	1897
PSRKRPR	Q62315, Q5F363, Q92833	1898
PSSKKRKV	SeqNLS	1899
PTKKRVK	P07664	1900
QRPGPYDRP	SeqNLS	1901
RGKGGKGLGKGGAKRHRK	SeqNLS	1902
RKAGKGGGGHKTTKKRSAKDEKVP	B4FG96	1903
RKIKLKRAK	A1L3G9	1904
RKIKRKRAK	B9X187	1905
RKKEAPGPREELRSRGR	O35126, P54258, Q5IS70, P54259	1906
RKKRKGK	SeqNLS, Q29243, Q62165, Q28685, O18738, Q9TSZ6, Q14118	1907

RKKRRQRRR	P04326, P69697, P69698, P05907, P20879, P04613, P19553, P0C1J9, P20893, P12506, P04612, Q73370, P0C1K0, P05906, P35965, P04609, P04610, P04614, P04608, P05905	1908
RKKSIPLSIKNLKRKHKRKKNKITR	Q9C0C9	1909
RKLVKPKNTKMKTKLRTNPY	Q14190	1910
RKRLILSDKGQLDWKK	SeqNLS, Q91Z62, Q1A730, Q2KHT6, Q9CPU7	1911
RKRLKSK	Q13309	1912
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,	1913
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6	1914
RKRTPRVDGQTGENDMNKRRRK	O94851	1915
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270	1916
RLRFRKPKSK	P69469	1917
RQQRKR	Q14980	1918
RRDLNSSFETSPKKVK	Q8K3G5	1919
RRDRAKLR	Q9SLB8	1920
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922	1921

RRGRKRKAEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475	1922
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1	1923
RRKRSKSEDMDSVESKRRR	Q7TT18	1924
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8	1925
RRPKGKTLQKRKPK	Q6ZN17	1926
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0	1927
RRRGKNKVAAQNCRK	SeqNLS	1928
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9	1929
RRRQKQKGGASRRR	SeqNLS	1930
RRRREGPRARRRR	P08313, P10231	1931
RRTIRLKLVYDKCDRSCKIQKKN RNKCQYCRFHKCLSVGMSHNAI RFGRMPRSEKAKLKAE	SeqNLS	1932
RRVPQRKEVSRCRKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5	1933
RVGGRRQAVECIEDLLNEPGQPL DLSCKRPRP	P03255	1934
RVVKLRIAP	P52639, Q8JMN0	1935
RVVRRR	P70278	1936
SKRKTKISRKTR	Q5RAY1, O00443	1937
SYVKTVPNRTRTYIKL	P21935	1938
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9	1939
TLSPASSPSSVSCPVIPASTDESPG SALNI	SeqNLS	1940
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9	1941
SPKKKRKVE		1942
KRTAD GSEFE SPKKKRKVE		1943
PAAKRVKLD		1944
PKKKRKV		1945

MDSLLMNRRKFLYQFKNVRWA KGRRETYLC		1946
SPKKKRKVEAS		1947
MAPKKKRKVGIHRGVP		1948

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: 1591), где спейсер находится в скобках. Другая иллюстративная двухкомпонентная NLS имеет последовательность PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1593). Иллюстративные NLS описаны в международной заявке WO 2020051561, которая включена в данный документ посредством ссылки во всей своей полноте, в том числе для их раскрытия в отношении последовательностей ядерной локализации.

В определенных вариантах осуществления полипептид системы редактирования генов Gene Writer™ дополнительно содержит последовательность внутриклеточной локализации, например, последовательность ядерной локализации и/или последовательность ядрышковой локализации. Последовательность ядерной локализации и/или последовательность ядрышковой локализации может представлять собой аминокислотные последовательности, которые способствуют импорту белка в ядро и/или ядрышко, где он может способствовать интеграции гетерологичной последовательности в геном. В определенных вариантах осуществления полипептид системы редактирования генов Gene Writer (например, ретротранспозаза, например, полипептид в соответствии с любой из таблиц 1, 2 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, С-конце или во внутренней области полипептида. В некоторых вариантах осуществления используется множество одинаковых или разных сигналов ядрышковой локализации. В некоторых вариантах осуществления сигнал ядерной локализации имеет длину, составляющую менее 5, 10, 25, 50, 75 или 100 аминокислот. Можно использовать разнообразные сигналы ядрышковой локализации полипептида. Например, в Yang et al., *Journal of Biomedical Science* 22, 33 (2015), описан сигнал ядерной локализации, который также функционирует в качестве сигнала ядрышковой локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может также представлять собой сигнал ядерной локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может

перекрываться с сигналом ядерной локализации. В некоторых вариантах осуществления сигнал ядрышковой локализации может содержать отрезок из основных остатков. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть богат аргининовыми и лизиновыми остатками. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, обогащенного в ядрышке. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, обогащенного в локусах рибосомной РНК. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из белка, который связывает рРНК. В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из MSP58. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой однокомпонентный мотив. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой двухкомпонентный мотив. В некоторых вариантах осуществления сигнал ядрышковой локализации может состоять из нескольких однокомпонентных или двухкомпонентных мотивов. В некоторых вариантах осуществления сигнал ядрышковой локализации может состоять из смеси однокомпонентных и двухкомпонентных мотивов. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой два двухкомпонентных мотива. В некоторых вариантах осуществления мотив ядрышковой локализации может представлять собой KRASSQALGTIPKRRSSSRFIKRKK (SEQ ID NO: 1530). В некоторых вариантах осуществления сигнал ядрышковой локализации может быть получен из киназы, индуцирующей ядерный фактор κB. В некоторых вариантах осуществления сигнал ядрышковой локализации может представлять собой мотив RKKRKKK (SEQ ID NO: 1531) (описанный в Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004).

Поскольку эндогенный сигнал ядрышковой локализации может помочь направить полипептид Gene Writer в ядрышко для тех полипептидов, которые получены из ретротранспозонов, в естественных условиях нацеленных на рДНК, например, R1, R2, R4, R8, R9, может быть благоприятной инактивация этого сигнала при перенацеливании на сайт вне рДНК. Эндогенный сигнал ядрышковой локализации (NoLS) можно предсказать 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 путем мутации этих остатков по типу замены на неосновные аминокислоты с простыми боковыми цепями или удаление их из полипептидной цепи 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 доменом обратной транскриптазы и доменом, подобным рестрикционной эндонуклеазе. В некоторых вариантах осуществления предсказанная область NoLS содержит аминокислоты лизин, аргинин, гистидин и/или глутамин, и сигнал ядрышковой локализации инактивируется посредством мутации одного или нескольких из этих остатков по типу замены на аланин и/или удаления их из полипептида.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК, кодирующая полипептид Gene Writer, или ДНК, кодирующая РНК), содержит сайт связывания с микроРНК. В некоторых вариантах осуществления сайт связывания с микроРНК используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, сайт связывания с микроРНК может быть выбран на основании того, что он распознается miRNA, которая присутствует в типе клеток, не являющихся мишенями, но которая отсутствует (или присутствует на сниженном уровне по сравнению с клетками, не являющимися мишенями) в типе клеток, являющихся мишенями. Таким образом, если РНК, кодирующая полипептид Gene Writer, присутствует в клетке, не являющейся мишенью, то она будет связываться с miRNA, а если РНК, кодирующая полипептид Gene Writer, присутствует в клетке-мишени, то она не будет связываться с miRNA (или будет связываться, но на сниженных уровнях по сравнению с клеткой, не являющейся мишенью). Не желая ограничиваться какой-либо теорией, полагают, что связывание miRNA с РНК, кодирующей полипептид GeneWriter, может приводить к снижению продуцирования полипептида GeneWriter, например, посредством разрушения мРНК, кодирующей полипептид, или посредством препятствования трансляции. Соответственно, вставка последовательности гетерологичного объекта в геном клеток-мишеней будет происходить более эффективно, чем в геном клеток, не являющихся мишенями. Система, имеющая сайт связывания с микроРНК в РНК, кодирующей полипептид Gene Writer (или кодируемый в ДНК, кодирующей РНК), может также применяться в комбинации с РНК-матрицей, которая регулируется вторым сайтом связывания с микроРНК, например, как описано в данном документе в разделе под названием "Компонент системы редактирования генов Gene Writer™, представляющий собой РНК-матрицу". В некоторых вариантах осуществления, например, для печеночных показаний, miRNA выбрана из таблицы 4 WO2020014209, которая включена 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, которая включена в данный документ посредством ссылки.

Специалист в данной области может по номерам доступа, представленным в таблицах 1-3, определить последовательности нуклеиновых кислот и соответствующих полипептидов для каждого ретротранспозона и его доменов, например, с помощью стандартных инструментов для анализа последовательностей, таких как средство поиска основного локального выравнивания (BLAST) или CD-Search для анализа консервативных доменов. Другие инструменты для анализа последовательностей известны и могут быть найдены, например, на https://molbiol-tools.ca, например, на https://molbiol-tools.ca/Motifs.htm. SEQ ID NO: 1-112 выравниваются с каждой строкой в таблице 1, и SEQ ID NO: 113-1015 выравниваются с первыми 903 строками в таблице 2.

В таблицах 1-3 в данном документе представлены последовательности иллюстративных транспозонов, в том числе аминокислотная последовательность ретротранспозазы, а также последовательности 5'- и 3'-нетранслируемых областей, обеспечивающие возможность связывания ретротранспозазы с РНК-матрицей, и полная последовательность нуклеиновой кислоты транспозона. В некоторых вариантах осуществления 5'-UTR из любой из таблиц 1-3 обеспечивает возможность связывания ретротранспозазы с РНК-матрицей. В некоторых вариантах осуществления 3'-UTR из любой из таблиц 1-3 обеспечивает возможность связывания ретротранспозазы с РНК-матрицей. Таким образом, в некоторых вариантах осуществления полипептид для применения в любой из систем, описанных в данном документе, может представлять собой полипептид из любой из таблиц 1-3 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 из любой из таблиц 1-3 в данном документе (или последовательность, характеризующуюся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с ними), например, из того же 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 из любой из таблиц 1-3 в данном документе, 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 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, выровнять 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.

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 последовательность ДНК ретротранспозона. В столбце 6 приведена предсказанная 5'-нетранслируемая область, а в столбце 7 приведена предсказанная 3'-нетранслируемая область; обе из них являются сегментами последовательности из столбца 5, которые, как предсказывается, обеспечивают возможность связывания РНК-матрицы с ретротранспозазой из столбца 8. (Понятно, что в столбцах 5-7 показана последовательность ДНК, и что последовательность РНК в соответствии с любым из столбцов 5-7 обычно содержит урацил, а не тимидин.) В столбце 8 приведена предсказанная последовательность ретротранспозазы, кодируемая ретротранспозоном из столбца 5.

Таблица 3. Иллюстративные белки Gene Writer и ассоциированные с ними последовательности и информация

1. Семейство	2. Элемент	3. Номер доступа		4. Организм	5. Последовательность ДНК	6. Предсказанная 5'-UTR	7. Предсказанная 3'-UTR	8. Предсказанная аминокислотная последовательность
R2	R2-1_TG	.		Taeniopygia guttata	GTCTAGTTACAACTGGGCATCGCTGCAGAG ATCGCACCTCCTCGTGGTCCCGCTGGTAGCC CTTCGAAGGGTGACTAAGTCGATCTCTGCC CCAGGTACGGAGCCGTTGGGACTCACCAGT CCAACGTAACTCCTGCCTAAATTCGGTGAA ACAAATTCCTCGGTAAAAAGCCCCATGGCT TCTTGCCCGAAACCTGGCCCCCCGGTTTCAG CAGGGGCAATGAGTTTGGAAAGTGGACTGA CCACCCACTCCGTTCTCGCCATCGAACGTGG TCCCAATTCGTTGGCAAATTCCGGATCAGA CTTTGGGGGGGGGGGTCTGGGGCTACCGTT ACGCCTATTGAGGGTATCGGTCGGCACTCA GACCTCCCGCTCCGACTGGGTAGACCTGGT GTCCTGGAGCCACCCAGGACCCACGTCTAA GTCCCAGCAGGTTGACCTGGTGTCTTTATTT CCTAAACACCGGGTTGACCTGTTATCCAAA AACGACCAGGTAGACCTGGTGGCTCAATTT	GTCTAG TTACAA CTGGGC ATCGCT GCAGAG ATCGCA CCTCCT CGTGGT CCCGCT GGTAGC CCTTCG AAGGGT GACTAA GTCGAT CTCTGC CCCAGG TACGGA	TTCA GGTT ATTT AGAT GCTT AGTT TTTGT ACCT TTCTT GTTTT GTTT AGGA TTTTG ATAG TGTT AGTA TTTTT	MASCPKPGPPVS AGAMSLESGLT THSVLAIERGPN SLANSGSDFGG GGLGLPLRLLRV SVGTQTSRSDW VDLVSWSHPGP TSKSQQVDLVSL FPKHRVDLLSK NDQVDLVAQFL PSKFPPNLAEND LALLVNLEFYRS DLHVYECVHFA AHWEGLSGLPE VYEQLAPQPCV GETLHSSLPRDS ELFVPEEGSSEK

					TTACCATCTAAATTTCCCCCCAATTTGGCAG	GCCGTT	ATAT	ESEDAPKTSPPT
					AAAATGATTTGGCTTTGCTGGTGAACTTAG	GGGACT	TTTTG	PGKHGLEQTGE
					AGTTCTACAGATCGGATTTGCATGTGTATGA	CACCAG	TACG	EKVMVTVPDKN
					GTGTGTTCATTTTGCTGCACATTGGGAGGGA	TCCAAC	ATTG	PPCPCCGTRVNS
					TTAAGTGGTTTGCCTGAGGTGTATGAACAA	GTAACT	CATA	VLNLIEHLKVSH
					CTTGCACCACAACCGTGTGTGGGAGAAACT	CCTGCC	ATGT	GKRGVCFRCAK
					TTACATTCTAGCCTCCCACGAGACAGTGAA	TAAATT	TCTTT	CGKENSNYHSV
					CTGTTTGTGCCTGAAGAGGGGAGCAGCGAG	CGGTGA	TTTAT	VCHFPKCRGPET
					AAGGAGAGCGAGGACGCGCCAAAAACATC	AACAAA	ACAG	EKAPAGEWICE
					TCCTCCGACGCCTGGGAAACATGGTTTGGA	TTCCTC	TTCTG	VCNRDFTTKIGL
					ACAGACTGGGGAGGAAAAAGTGATGGTGA	GGTAAA	TTTTA	GQHKRLAHPAV
					CTGTTCCTGACAAAAATCCACCTTGTCCTTG	AAGCCC	ATAA	RNQERIVASQPK
					CTGTGGTACCCGGGTAAACTCTGTGTTGAAT	C (SEQ	AATA	ETSNRGAHKRC
					CTGATTGAACATCTGAAAGTGTCACACGGG	ID NO:	GACG	WTKEEEELLIRL
					AAAAGGGGGGTTTGTTTTCGGTGTGCAAAA	1140)	ATAG	EAQFEGNKNIN
					TGTGGAAAGGAAAATAGTAACTATCACAGT		CTAG	KLIAEHITTKTA
					GTTGTTTGTCATTTTCCAAAATGCAGGGGTC		AGAC	KQISDKRRLLSR
					CAGAGACGGAGAAAGCCCCAGCTGGGGAG		GTTA	KPAEEPREEPGT
					TGGATTTGTGAGGTATGCAACAGAGATTTT		GGGC	CHHTRRAAASL
					ACAACCAAAATTGGCCTGGGACAACACAAG		AGCC	RTEPEMSHHAQ
					AGATTGGCACACCCAGCAGTGAGAAATCAG		ACAA	AEDRDNGPGRR
					GAAAGGATCGTTGCTTCCCAACCGAAAGAA		GCCA	PLPGRAAAGGR
					ACATCAAATAGAGGTGCTCACAAAAGGTGC		GTTA	TMDEIRRHPDK

					TGGACAAAGGAGGAGGAAGAATTACTAAT		GGTA	GNGQQRPTKQK
					AAGACTGGAGGCTCAGTTCGAGGGAAACAA		GCGG	SEEQLQAYYKK
					AAATATTAATAAGCTTATTGCAGAACACAT		ATAG	TLEERLSAGALN
					AACCACCAAAACAGCTAAGCAGATCAGTGA		TAGG	TFPRAFKQVME
					CAAAAGGCGATTGCTGTCCAGAAAGCCAGC		TAGG	GRDIKLVINQTA
					AGAGGAGCCACGTGAGGAGCCTGGAACGT		AACA	QDCFGCLESISQI
					GTCATCACACCAGGAGAGCAGCTGCGAGCC		GACT	RTATRDKKDTV
					TGAGAACGGAGCCTGAGATGAGTCATCACG		TTTAC	TREKHPKKPFQ
					CCCAGGCAGAGGACAGAGATAATGGACCTG		TATTT	KWMKDRAIKK
					GGAGACGCCCTCTGCCAGGCAGGGCAGCTG		CATA	GNYLRFQRLFY
					CCGGAGGGAGAACAATGGACGAGATAAGA		ACGC	LDRGKLAKIILD
					CGCCACCCTGATAAGGGCAACGGACAGCAG		GTCA	DIECLSCDIPLSE
					AGACCCACCAAGCAAAAATCAGAAGAACA		ATTA	IYSVFKTRWETT
					GCTGCAGGCTTACTATAAAAAGACACTAGA		CCAC	GSFKSLGDFKTY
					GGAACGACTTTCAGCTGGGGCACTTAACAC		CTGA	GKADNTAFRELI
					CTTCCCCCGAGCATTCAAGCAGGTAATGGA		TTTG	TAKEIEKNVQE
					AGGCCGGGATATAAAGCTAGTAATCAATCA		GACC	MSKGSAPGPDGI
					GACAGCGCAGGACTGCTTCGGATGCCTGGA		AATT	TLGDVVKMDPE
					ATCCATAAGCCAAATAAGAACGGCAACCCG		CACG	FSRTMEIFNLWL
					AGATAAAAAGGACACGGTGACCCGGGAGA		GGAT	TTGKIPDMVRG
					AACACCCAAAGAAACCTTTTCAGAAGTGGA		TTGTC	CRTVLIPKSSKP
					TGAAGGACAGAGCAATCAAAAAAGGTAATT		CAAG	DRLKDINNWRPI
					ATCTTCGGTTCCAGCGTTTATTTTATCTTGA		GTGG	TIGSILLRLFSRI

					TAGAGGGAAACTGGCTAAAATCATTTTAGA		ACGG	VTARLSKACPL
					TGATATTGAATGCTTGTCTTGTGACATACCA		GCCA	NPRQRGFIRAAG
					CTCAGTGAAATTTATTCGGTTTTTAAAACAA		CCTTT	CSENLKLLQTII
					GATGGGAAACAACTGGTAGCTTTAAAAGCC		ACTT	WSAKREHRPLG
					TTGGGGACTTTAAAACTTACGGGAAGGCTG		AACC	VVFVDIAKAFDT
					ACAACACTGCCTTCAGAGAATTAATTACGG		CGGA	VSHQHIIHALQQ
					CTAAAGAAATTGAGAAAAATGTGCAGGAA		AAAG	REVDPHIVGLVS
					ATGAGCAAAGGCTCGGCTCCCGGTCCAGAC		GAAC	NMYENISTYITT
					GGGATTACTCTTGGGGACGTCGTAAAGATG		ATAT	KRNTHTDKIQIR
					GATCCCGAGTTTTCCCGGACCATGGAGATTT		ATAA	VGVKQGDPMSP
					TCAATTTATGGTTAACAACTGGTAAAATCCC		TTTAT	LLFNLAMDPLL
					GGACATGGTGAGGGGGTGCAGAACCGTTTT		GTGT	CKLEESGKGYH
					GATTCCAAAATCATCAAAGCCGGATCGTTT		GTTC	RGQSSITAMAFA
					GAAAGACATTAATAACTGGAGACCTATCAC		GATA	DDLVLLSDSWE
					GATCGGTTCCATCTTGCTGAGACTGTTCTCC		AA	NMNTNISILETF
					AGGATTGTAACAGCTAGGCTGAGCAAAGCG		(SEQ	CNLTGLKTQGQ
					TGCCCCCTGAACCCAAGGCAAAGAGGCTTT		ID NO:	KCHGFYIKPTKD
					ATCAGAGCGGCGGGATGCTCTGAAAACTTA		1263)	SYTINDCAAWTI
					AAACTCCTGCAAACTATAATTTGGTCGGCC			NGTPLNMIDPGE
					AAAAGAGAACACAGACCACTGGGTGTTGTA			SEKYLGLQFDP
					TTCGTGGACATCGCCAAGGCTTTTGACACC			WIGIARSGLSTK
					GTAAGCCACCAGCACATCATTCATGCTTTGC			LDFWLQRIDQA
					AGCAAAGAGAGGTGGATCCCCACATCGTCG			PLKPLQKTDILK

					GTCTGGTGAGCAATATGTACGAGAACATCA			TYTIPRLIYIADH
					GTACGTATATCACCACAAAGAGGAACACAC			SEVKTALLETLD
					ACACAGACAAAATCCAGATCCGGGTTGGAG			QKIRTAVKEWL
					TAAAGCAGGGTGACCCGATGTCGCCCCTTT			HLPPCTCDAILY
					TATTTAACCTGGCAATGGACCCTCTATTATG			SSTRDGGLGITK
					CAAGCTGGAAGAGAGTGGCAAAGGATACC			LAGLIPSVQARR
					ACCGAGGACAGAGCAGCATCACAGCGATG			LHRIAQSSDDTM
					GCATTTGCAGACGATCTGGTTTTGCTGAGCG			KCFMEKEKMEQ
					ACTCCTGGGAAAATATGAATACAAATATTA			LHKKLWIQAGG
					GCATACTGGAGACCTTCTGCAATCTGACCG			DRENIPSIWEAP
					GTCTCAAAACACAGGGGCAAAAGTGCCACG			PSSEPPNNVSTN
					GCTTTTACATCAAGCCGACAAAGGACTCTT			SEWEAPTQKDK
					ACACCATCAATGACTGCGCTGCCTGGACTA			FPKPCNWRKNE
					TCAACGGCACACCCCTGAACATGATCGACC			FKKWTKLASQG
					CCGGCGAATCTGAGAAATACCTCGGCCTGC			RGIVNFERDKIS
					AGTTTGACCCGTGGATTGGAATAGCAAGGT			NHWIQYYRRIPH
					CCGGTCTCTCCACAAAACTAGATTTTTGGCT			RKLLTALQLRA
					TCAGCGGATCGATCAAGCACCACTTAAACC			NVYPTREFLAR
					TCTGCAGAAAACTGATATTCTCAAAACATA			GRQDQYIKACR
					CACCATCCCTCGGCTGATCTACATAGCTGAC			HCDADIESCAHII
					CACTCAGAAGTGAAAACTGCACTACTCGAA			GNCPVTQDARI
					ACCCTTGACCAGAAGATCCGGACAGCGGTC			KRHNYICELLLE
					AAGGAATGGCTTCACCTACCTCCGTGCACC			EAKKKDWVVF

					TGCGATGCCATCCTGTACTCGAGCACGAGA GACGGCGGTTTGGGCATCACCAAATTGGCA GGACTGATCCCCAGCGTGCAGGCCCGTAGA CTGCATCGGATCGCACAGTCATCTGACGAT ACGATGAAATGCTTCATGGAAAAAGAGAAA ATGGAACAGCTGCATAAGAAATTGTGGATT CAAGCTGGAGGGGACAGAGAGAACATACC CTCGATTTGGGAAGCACCACCGTCGAGTGA ACCACCAAACAACGTGAGCACAAATTCGGA ATGGGAAGCACCGACCCAGAAAGATAAATT TCCAAAGCCTTGCAATTGGAGGAAAAACGA ATTCAAAAAATGGACCAAATTGGCATCCCA AGGCCGCGGAATTGTAAATTTTGAAAGAGA CAAAATTAGTAACCATTGGATCCAATACTA CAGACGCATACCTCACAGGAAACTCCTCAC TGCACTACAACTCAGGGCCAACGTTTACCC CACGAGAGAATTTCTAGCCAGGGGTAGACA AGACCAATACATCAAGGCGTGTAGGCACTG CGATGCGGACATTGAATCCTGCGCCCACAT CATCGGCAACTGCCCAGTGACACAGGACGC CCGAATCAAGAGGCACAATTACATCTGCGA ACTGCTTCTCGAGGAGGCGAAGAAGAAGGA CTGGGTAGTGTTCAAGGAACCGCACATAAG			KEPHIRDSNKEL YKPDLIFVKDAR ALVVDVTVRYE AAKSSLEEAAA EKVRKYKHLET EVRHLTNAKDV TFVGFPLGARG KWHQDNFKLLT ELGLSKSRQVK MAETFSTVALFS SVDIVHMFASR ARKSMVM (SEQ ID NO: 1016)

					GGATTCCAACAAGGAACTGTACAAACCTGA CCTGATATTTGTGAAGGATGCCCGTGCACTT GTCGTGGATGTGACAGTACGGTATGAAGCA GCCAAATCATCGCTGGAGGAAGCCGCTGCA GAGAAAGTGAGAAAGTACAAACACCTGGA AACGGAAGTAAGACATCTCACGAATGCAAA GGACGTTACTTTTGTGGGCTTTCCCCTAGGA GCGCGGGGGAAATGGCACCAAGATAACTTT AAACTTTTGACTGAGCTTGGCCTCTCCAAAT CGAGGCAAGTGAAAATGGCAGAGACTTTTT CCACAGTAGCGCTCTTTTCATCTGTGGACAT TGTACATATGTTTGCCAGTAGGGCCAGAAA ATCTATGGTTATGTAATTCAGGTTATTTAGA TGCTTAGTTTTTGTACCTTTCTTGTTTTGTTT AGGATTTTGATAGTGTTAGTATTTTTATATT TTTGTACGATTGCATAATGTTCTTTTTTATA CAGTTCTGTTTTAATAAAATAGACGATAGCT AGAGACGTTAGGGCAGCCACAAGCCAGTTA GGTAGCGGATAGTAGGTAGGAACAGACTTT TACTATTTCATAACGCGTCAATTACCACCTG ATTTGGACCAATTCACGGGATTTGTCCAAG GTGGACGGGCCACCTTTACTTAACCCGGAA			

					AAGGAACATATATAATTTATGTGTGTTCGATAAA (SEQ ID NO: 1539)			
R2	R2-1_Gfo	.		Geospiza fortis	AGACTTAAGTGAGTTTGGTTACAACTGGGCATAGCTGCAGAGACCGCGCCTCCTCGCGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAAGTCGGTCTCTGCCCCAGTCCGGGAGTCGATGGGACTCACCAGCCCAACGATTCCTTCCAAAATTTCGGTGAAACAAATTTCTCGGTGCAAGTCGCAAGGCTTGTCACCCGAAACCTAGCCCCCCGGTCGGTCAGGGGCAACGGGTTCGGAAGTGGGATGGCCACCCACCCCGTTCCCGCAGACGAATCCGGCCATGAATCTGATCCATTCCTTGTAGGGAGGAGCTGCGGACAACCGGCACGCCTTACTAGGCAATCGGTTGGCACCCAGACCTCCCGAGATGATATTTTACCATCTAAAACCACCAAATTGACAGAGAATGAATTGGACTTGCTGGTGAACTTTTCTTTAGAATTGTATAGGTCAGATCTGCAGGGATTTGTGCAGGAGGGGATTCATTTTTCTGTGAATAGGGAGGTGTTAGAGGGGTTTCCTGAGGTGTATGAACAACCTGCACCACAACCGGCAGTAGGGGACGATTTAAACACCAGTCTCCCACCGGACAATAATATATGCGTACTTGAGAAGGGTAGCAGTGAAGC	AGACTTAAGTGAGTTTGGTTACAACTGGGCATAGCTGCAGAGACCGCGCCTCCTCGCGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAA (SEQ ID NO: 1141)	GGTAGATAATCTTTGTATAGTGGGGGGGGATCTCATGTACCGGGTTTCTTTTATTTGATTTTCAATAAAACAGACGGTAGCTAGGTTC	VGLCPSPGVDGTHQPNDSFQNFGETNFSVQVARLVTRNLAPRSVRGNGFGSGMATHPVPADESGHESDPFLVGRSCGQPARLTRQSVGTQTSRDDILPSKTTKLTENELDLLVNFSLELYRSDLQGFVQEGIHFSVNREVLEGFPEVYEQPAPQPAVGDDLNTSLPPDNNICVLEKGSSEAVEDGTPEVAHPVPETQGKESPNNIVMVTLPNKNPPCPCCRVRLHSVLALIEHLKGSHGKKRACF

					AGTGGAGGATGGCACACCGGAGGTAGCGC		GCAA	RCVKCGRENFN
					ACCCCGTGCCTGAAACCCAGGGCAAAGAGT		GGCA	YHSTVCHIAKC
					CACCGAATAACATCGTGATGGTAACTCTTC		GCCA	KGPKVEKAPVG
					CCAACAAAAATCCACCATGTCCTTGCTGTA		CAAG	EWICEVCGRDFT
					GGGTCAGACTGCATTCAGTACTGGCTCTGA		CCAA	TKIGLGQHKRL
					TTGAACATCTTAAGGGGTCGCATGGGAAGA		AGAT	AHPLVRNQERID
					AGAGGGCATGCTTTAGGTGTGTCAAGTGTG		AGGT	ASQPKETSNRG
					GGAGGGAGAACTTTAACTATCATAGTACTG		AGGG	AHKRCWTKEEE
					TTTGTCACATCGCAAAATGCAAGGGACCAA		TGCT	EMLIKLEVQFEG
					AAGTTGAGAAGGCCCCAGTGGGAGAGTGG		CATA	HRNINKLIAEHL
					ATCTGTGAGGTATGTGGTAGGGACTTTACA		GTGA	TTKTSKQISDKR
					ACCAAAATCGGCCTGGGACAACATAAAAGA		GTAG	RLLPRKQLTDLS
					TTGGCACATCCCTTGGTTAGAAACCAAGAA		GGAC	KGVAGQKVLDP
					AGGATCGATGCTTCCCAACCGAAGGAGACA		AGTG	GLSHQPQLGVV
					TCAAACAGAGGAGCCCACAAGAGATGTTGG		CCTTT	DNGLGGGHLPG
					ACAAAAGAGGAGGAGGAGATGCTGATAAA		TGAT	GPAAEGRTIEPL
					GTTGGAGGTACAGTTCGAGGGACACAGAAA		TCAC	GHHLDKDNGHR
					CATCAATAAGCTTATCGCGGAACACTTAAC		AACG	EIADQHKAGRL
					AACTAAAACATCCAAACAGATTAGTGATAA		CGTC	QAHYRKKIRKR
					AAGGAGACTATTACCCAGAAAACAATTAAC		AATA	LSEGMISNFPEV
					AGATCTAAGTAAGGGAGTGGCTGGACAGAA		CCAT	FEQLLDCQEAQP
					GGTGCTGGACCCAGGACTGAGTCATCAACC		CTGA	LINQAAQDCFG
					CCAGCTGGGGGTAGTTGACAATGGACTTGG		CACG	CLDSASQIRKAL

					TGGGGGTCATCTGCCAGGGGGGCCAGCTGC		GATA	RKQNTQKDQGD
					TGAAGGAAGAACAATAGAGCCATTAGGAC		CCCTT	QPKRPAQKWM
					ACCACCTTGATAAGGATAACGGTCACCGGG		ACCG	KKRAVKRGHFL
					AAATCGCTGACCAGCACAAGGCAGGGAGG		GACT	RFQKLFHLDRG
					CTGCAGGCCCATTACCGAAAGAAGATAAGG		TGTC	KLAKIILDDVEC
					AAGCGCCTTTCAGAAGGGATGATTAGCAAC		ATGA	LSCDIPPSEIYSV
					TTCCCCGAAGTATTTGAACAACTACTGGACT		TCTCC	FKARWETPGQF
					GCCAGGAAGCACAACCATTGATCAATCAAG		CAGA	AGLGDFEINRKA
					CAGCGCAGGATTGCTTTGGATGCCTGGATT		CTTGT	NNKAFRDLITAK
					CAGCAAGCCAGATAAGGAAGGCGCTCCGA		CCAA	EILKNVREMTK
					AAACAGAACACACAGAAAGACCAGGGGGA		GGTG	GSAPGPDGIALG
					TCAACCCAAAAGACCAGCTCAGAAGTGGAT		GACG	DIRKMDPEYTRT
					GAAAAAAAGAGCAGTTAAGAGGGGTCACTT		GGCC	AELFNLWLTSG
					CCTCCGCTTTCAGAAATTATTTCATCTTGAC		ACCT	EIPDMVRGCRT
					AGGGGGAAATTGGCAAAGATTATTTTGGAC		TTACT	VLIPKSSKPERL
					GACGTAGAGTGTTTGTCCTGTGATATACCAC		TAAC	KDINNWRPITIG
					CCAGTGAAATTTATTCGGTATTCAAAGCCC		CCGG	SILLRLFSRIITAR
					GATGGGAAACACCTGGACAGTTTGCTGGCC		AAAA	LTKACPLNPRQR
					TTGGGGATTTCGAAATTAATAGGAAGGCGA		GGAA	SFISAAGCSENL
					ACAATAAAGCCTTCAGGGACTTAATTACGG		CATA	KLLQTIIRTAKN
					CCAAAGAAATTCTCAAAAATGTGCGGGAGA		TATT	EHRPLGVVFVDI
					TGACCAAGGGCTCGGCCCCAGGTCCAGATG		AATT	AKAFDTVSHQHI
					GGATCGCGCTTGGGGACATCAGGAAGATGG		ATAT	IHVLQRRRVDPH

					ACCCTGAGTACACCCGGACCGCCGAACTCT TCAACTTATGGTTAACATCTGGTGAGATCCC GGACATGGTGAGGGGGTGCAGAACTGTGTT AATCCCCAAATCGTCAAAACCGGAACGCCT GAAGGACATCAATAACTGGAGACCCATCAC GATTGGATCCATCTTGCTGAGACTTTTCTCC AGGATCATAACAGCGAGGTTAACAAAGGCG TGCCCCCTCAACCCTAGGCAAAGAAGCTTC ATCAGTGCGGCAGGATGCTCCGAGAACTTG AAGCTCCTGCAAACCATAATTCGGACTGCT AAAAATGAACACAGACCACTGGGTGTTGTA TTCGTGGACATCGCCAAGGCCTTTGACACC GTGAGCCACCAACACATCATACATGTATTG CAAAGGAGGAGAGTGGACCCCCACATCATT GGATTGGTGAAAAATATGTACAAAGACATC AGTACGGTTATCACCACAAAGAAGAACACA TACACGGACAAAATCCAGATCCAGGTTGGA GTGAAGCAAGGTGATCCGCTTTCGCCCCTTC TATTCAACCTGGCGATGGACCCCCTGTTGTG CAAGCTGGAAGAACACGGCAAAGGATTCCA CCGAGGACAGAGCAAGATAACAGCGATGG CATTCGCTGATGACCTGGTCCTGTTGAGCGA TTCCTGGGAAGACATGAATGCGAACATCAA		GTGT TCGG AAAA (SEQ ID NO: 1264)	IIGLVKNMYKDI STVITTKKNTYT DKIQIQVGVKQ GDPLSPLLFNLA MDPLLCKLEEH GKGFHRGQSKIT AMAFADDLVLL SDSWEDMNANI KILETFCDLTGL KTQGQKCHGFY IKPTKDSYTVNN CAAWTINGTPL NMINPGESEKYL GLQFDPWVGIA KTSLPEKLDFWL ERIDRAPLKPFQ KLDILKTYTIPRL TYVADHSEMKA GALEALDRTIRS AVKDWLHLPSS TCDAILYTSMK DGGLGVTKLVG LIPSVQARRLHR

					GATACTGGAGACCTTCTGCGACCTCACCGG			IAQSPEETMKDF
					TCTCAAAACACAGGGTCAAAAGTGCCACGG			LEKAQMEKMYE
					CTTCTACATCAAGCCTACAAAGGACTCTTAC			KLWVQAGGKR
					ACCGTCAACAACTGCGCTGCGTGGACCATC			KRMPSIWEALPE
					AATGGCACACCCCTGAACATGATCAACCCC			VVPSIDTATTSE
					GGGGAATCAGAGAAATACCTCGGCCTGCAG			WEAPNPKSKYP
					TTTGACCCCTGGGTGGGAATTGCAAAGACC			RPCNWRRKEFK
					AGCCTCCCCGAAAAACTGGACTTCTGGCTC			KWTKLIAQGWG
					GAACGCATTGATCGAGCTCCACTCAAACCA			IRCFKGDKISNN
					TTTCAGAAACTGGACATTCTTAAGACATAC			WIRHYRYIPHRK
					ACCATACCTCGACTGACCTACGTAGCTGAC			LLTAIQLRASVY
					CACTCAGAGATGAAAGCGGGGGCCCTTGAA			PTREFLARGRED
					GCCCTTGACCGGACAATTCGATCGGCGGTC			NCVKSCRHCEA
					AAGGACTGGCTGCACCTACCTTCGAGCACC			AEESCAHIIGMC
					TGTGATGCCATCTTGTACACGAGCATGAAG			PVVRDARIKRH
					GACGGTGGTTTGGGAGTGACCAAATTGGTG			NRICERLMEEAG
					GGACTGATTCCGAGTGTACAAGCCCGGAGG			KRDWTVFQEPH
					CTGCACAGGATTGCGCAGTCACCGGAGGAG			IRDVTKELYKPD
					ACGATGAAAGACTTCCTGGAAAAGGCCCAG			LIFVKEGLALVV
					ATGGAGAAGATGTACGAGAAATTGTGGGTC			DVTIRFESTKTT
					CAAGCTGGAGGGAAAAGAAAGAGGATGCC			LEEAAAEKVNK
					GTCAATTTGGGAAGCGCTCCCGGAGGTTGT			YKHLETEVRNL
					ACCATCCATAGACACAGCCACAACTTCGGA			TNAKDVIFMGFP

					GTGGGAAGCACCGAACCCTAAAAGTAAGTA CCCTAGACCTTGTAATTGGCGCAGAAAAGA ATTTAAAAAGTGGACTAAATTAATAGCCCA GGGCTGGGGAATTAGGTGTTTTAAGGGGGA CAAAATTAGTAACAATTGGATTCGACATTA TAGATACATACCTCACAGGAAACTTCTCAC TGCCATACAGCTCCGGGCCAGTGTGTACCC CACAAGGGAATTTCTCGCGCGGGGGAGGGA AGATAACTGTGTTAAGTCTTGTAGGCACTGT GAGGCGGCAGAGGAGTCCTGTGCCCACATC ATCGGCATGTGTCCAGTCGTGAGGGATGCC CGAATCAAGAGGCACAATCGCATTTGCGAG AGGCTGATGGAGGAGGCGGGGAAGAGGGA CTGGACGGTGTTTCAGGAGCCGCACATAAG GGACGTCACCAAGGAACTGTACAAACCGGA CTTGATATTCGTGAAAGAAGGCCTTGCACTT GTTGTGGATGTTACAATACGGTTCGAGTCA ACCAAGACAACGTTGGAGGAGGCTGCTGCA GAGAAGGTGAACAAGTACAAACATCTGGA GACCGAAGTACGGAACCTCACCAACGCTAA GGACGTTATCTTTATGGGGTTTCCCCTTGGA GCGCGGGGACAATGGTACAATAAGAACTTT GAACTTTTGGACACTCTTGGCCTCCCCAGAT			LGARGQWYNK NFELLDTLGLPR SRQDIIAKTLST DALISSVDIIHMF ASRGRRQHA (SEQ ID NO: 1386)

					CGAGGCAGGACATTATTGCAAAGACTTTAT CCACGGACGCGCTCATTTCATCTGTGGACAT TATACATATGTTTGCCAGTAGAGGCAGAAG ACAGCATGCTTAGGGTAGATAATCTTTGTAT AGTGGGGGGGGATCTCATGTACCGGGTTTC TTTTATTTGATTTTCAATAAAACAGACGGTA GCTAGGTTCGCAAGGCAGCCACAAGCCAAA GATAGGTAGGGTGCTCATAGTGAGTAGGGA CAGTGCCTTTTGATTCACAACGCGTCAATAC CATCTGACACGGATACCCTTACCGGACTTGT CATGATCTCCCAGACTTGTCCAAGGTGGAC GGGCCACCTTTACTTAACCCGGAAAAGGAA CATATATTAATTATATGTGTTCGGAAAA (SEQ ID NO: 1540)			
R2	R2-1_ZA	.		Zonotrichia albicollis	CGACTTGAGAAGGTCTGGTTACAACTGGGC ATAGCTGCAGAGATCGCGCCTCCTCGTGGC CCCGCTGGTAAGCCCTTAACAGGGTGACTA AGTCGATCTCTGCCCCAGTCCAGGAGCCGC TGGGTTTCACCAGCCCAGCGATTCCTTCCAA ATTCGGTGAAACAAATTCCTCGGTAAAAGC CGCGTGGCTTATTGCCTGAAACCTGGCCCCC CGGTTTCAGACAGGGGCAAAGAGTTCGGAA GTGGACTGACCACCCACCCCGAACCCGAGA	CGACTT GAGAAG GTCTGG TTACAA CTGGGC ATAGCT GCAGAG ATCGCG CCTCCT	GTAG TCAC ATTG CACT TTCTG TAAC TTGC ACTG GGTG	NKFLGKSRVAY CLKPGPPVSDRG KEFGSGLTTHPE PESESGHDPTVP NPGPSLGAGEG AQPLPLLRVSVG TQTCEEDFITSRP TKLPGIESELGPL VKFSLEVYRSDL

					GCGAATCTGGTCATGACCCAACTGTCCCAA	CGTGGC	TGGG	KGDVQFEGIHFP
					ATCCTGGTCCGTCTCTTGGAGCGGGGGAAG	CCCGCT	ATGT	DNWGVLEGFPE
					GTGCACAGCCACTACCCTTACTCAGGGTAT	GGTAAG	GGGC	VYEQLAPQPNG
					CGGTGGGCACCCAAACCTGTGAAGAGGACT	CCCTTA	CTGG	GDELNHSLPGD
					TTATAACATCTAGACCAACCAAATTACCCG	ACAGGG	GGTG	REGDVLEKDSSE
					GAATTGAATCAGAATTAGGCCCGCTGGTGA	TGACTA	TGGG	KEKEAAPEALPS
					AGTTTTCTTTAGAGGTTTACAGGTCAGATCT	AGTCGA	TTAT	VQRARSEQLPD
					TAAGGGGGATGTGCAATTTGAGGGGATTCA	TCTCTG	GGGG	NIVKVTVPDKNP
					TTTTCCAGATAATTGGGGGGTACTGGAGGG	CCCCAG	TATA	PCPCCGVRLNSV
					GTTTCCTGAGGTGTACGAACAACTGGCACC	TCCAGG	TATG	LALIEHLKGSHG
					ACAGCCAAACGGGGGAGACGAGTTAAATC	AGCCGC	TGGG	RRRVCFRCAKC
					ATAGTCTCCCAGGGGACAGGGAGGGGGATG	TGGGTT	ATAT	GRENFNHHSTV
					TACTTGAGAAGGATAGCAGCGAAAAGGAG	TCACCA	TCTG	CHYAKCKGPQI
					AAGGAGGCTGCACCAGAGGCATTGCCCTCA	GCCCAG	GTGG	ERPPVGEWICEV
					GTGCAAAGGGCCCGCAGTGAACAGTTGCCA	CGATTC	GAAT	CGRDFTTKIGLG
					GATAACATCGTAAAGGTGACTGTTCCCGAC	CTTCCA	GTCC	QHKRHMHAMV
					AAAAATCCACCATGTCCCTGCTGTGGTGTCC	AATTCG	ATTC	RNQERIDASQPK
					GCTTAAACTCAGTGTTAGCTCTGATTGAACA	GTGA	ACTG	ETSNRGAHKRC
					TCTGAAGGGCTCACACGGGAGGAGGAGGGT	(SEQ ID	TATG	WTKEEEELLMK
					GTGCTTTAGGTGTGCCAAATGTGGGAGGGA	NO: 1142)	CCTA	LEVQFENHKNIN
					GAATTTTAACCACCATAGTACTGTTTGTCAT		TCTTT	KLIAEQLTTKTA
					TACGCAAAGTGCAAAGGTCCACAGATTGAA		TTAA	KQISDKRRMLL
					AGGCCACCAGTGGGAGAGTGGATCTGTGAG		TAAA	KKGRGTTGNLE

					GTATGCGGAAGGGACTTCACGACCAAAATT		AAGA	TEPGMSHQSQA
					GGCCTGGGACAACACAAAAGACATATGCAT		CGGT	KVKDNGLGGDH
					GCAATGGTGAGAAACCAGGAAAGGATCGA		AGCT	LPGGPVVDKGTI
					TGCTTCCCAACCGAAAGAGACATCAAATCG		AGGT	GKPGQHLDTDN
					AGGAGCCCACAAGAGGTGCTGGACGAAGG		TCGC	SHQITAGKKKG
					AGGAGGAAGAACTGCTCATGAAGTTGGAGG		GAAG	GGLQARYRRRI
					TACAGTTTGAGAATCACAAAAACATCAATA		CAGC	MKRLAAGTINIF
					AGCTTATCGCAGAGCAATTAACAACTAAAA		CACA	PKVFKELINDQE
					CAGCTAAACAAATTAGTGATAAAAGGAGAA		AGCC	ARPLINQTTEDC
					TGCTGCTCAAAAAAGGTAGGGGGACAACTG		AATA	FGLLDSACQIRT
					GTAATTTGGAAACAGAGCCTGGGATGAGTC		GCCA	ALREKGKSQEE
					ATCAATCGCAGGCAAAAGTTAAGGACAATG		GTTA	RPRKQYQKWM
					GACTGGGTGGGGACCATCTGCCGGGAGGAC		GGTA	KKRAIKRGDYL
					CAGTTGTCGATAAGGGAACAATAGGGAAGC		GCTC	RFQRLFHLDRG
					CAGGACAACATCTTGACACAGATAACAGCC		ATAG	KLARIILDNTESL
					ATCAAATAACTGCTGGCAAGAAGAAAGGG		TGGG	SCDISPSEIYSVF
					GGAGGGCTGCAGGCTCGTTATAGAAGGAGA		TAGG	KARWETPGHFN
					ATAATGAAACGATTAGCGGCCGGGACAATT		TGAC	GLGDFEIKGKA
					AACATCTTCCCCAAAGTGTTTAAAGAACTG		AGGA	NNKAFRDFITAK
					ATTAACGACCAAGAGGCGAGACCGCTAATC		ACCT	EIEKNVREMSK
					AATCAAACAACAGAAGACTGCTTTGGCCTC		TTGA	GSAPGPDGIALG
					TTGGACTCTGCATGCCAAATTAGAACGGCA		CTCA	DIKKMDPGYSR
					CTCCGGGAGAAGGGCAAATCTCAGGAGGA		GAAC	TAELFNLWLTA

					ACGACCAAGAAAACAGTATCAGAAGTGGAT		GCGT	GDIPDMVRGCR
					GAAGAAGAGAGCGATTAAAAGGGGGGACT		CCAT	TVLIPKSTTPERL
					ATCTCCGCTTCCAGCGATTATTCCATCTAGA		TAAC	KDINNWRPITIG
					CAGGGGGAAACTGGCGAGAATTATCTTGGA		ATCT	SILLRLFSRIITAR
					CAACACTGAGAGCTTGTCTTGCGATATATC		AGAA	MTKACPLNPRQ
					ACCCAGTGAAATTTATTCGGTATTCAAGGC		CGGA	RGFISAPGCSEN
					CAGATGGGAAACACCTGGACACTTCAACGG		CCAA	LKLLQSIIRTAK
					CCTTGGGGACTTTGAAATTAAAGGGAAGGC		ACTT	NEHKPLGVIFVD
					CAACAACAAAGCCTTCAGGGACTTCATCAC		CGGA	IAKAFDTVSHQH
					GGCTAAAGAAATTGAAAAGAACGTGCGGG		CATG	IIHVLQQRRVDP
					AAATGAGTAAGGGTTCGGCGCCAGGTCCAG		CACC	HIVGLVNNMYK
					ATGGGATCGCCCTTGGGGACATCAAGAAGA		GATT	DISTYVTTKKNT
					TGGATCCCGGGTATTCCCGGACCGCCGAGC		AACC	HTDKIQIRVGVK
					TATTCAACTTGTGGCTGACAGCTGGTGACAT		GGAT	QGDPLSPLLFNL
					CCCGGACATGGTGAGGGGGTGCAGGACTGT		TTGTC	AMDPLLCKLEE
					TTTGATCCCGAAATCGACGACACCGGAGCG		CAAG	SGKGFHRGQSSI
					CCTAAAGGACATCAACAACTGGAGACCCAT		GTGG	TAMAFADDLVL
					CACGATTGGTTCCATCTTGCTAAGGCTGTTC		ACGG	LSDSWENMKEN
					TCCAGGATCATAACGGCGAGGATGACTAAG		GCCA	IKILETFCNLTGL
					GCGTGCCCCCTCAACCCGAGACAGAGAGGC		CCTTT	KTQGQKCHGFY
					TTCATCAGTGCGCCGGGATGCTCTGAGAAC		ACTT	IKPTKDSYTINN
					CTGAAACTCCTGCAATCTATAATTCGGACTG		AACC	CPAWTINGTPLN
					CCAAAAATGAGCACAAGCCGCTGGGTGTTA		CGGA	MINPGESEKYLG

					TTTTCGTGGACATTGCTAAGGCTTTTGACAC		AAGG	LQIDPWTGVAK
					CGTGAGCCACCAACACATCATACACGTTTT		GAAC	YDLSTKLKIWLE
					ACAGCAACGGAGGGTTGACCCCCACATTGT		ATAT	SIDRAPLKPLQK
					TGGACTGGTGAACAATATGTACAAGGACAT		ATAG	LDILKTYTIPRLT
					CAGTACGTATGTCACCACAAAGAAGAACAC		TTAT	YLADHSEMKAG
					ACACACGGACAAAATCCAGATCCGGGTTGG		ATGT	ALEALDQQIRTA
					AGTGAAGCAGGGTGACCCACTATCACCCCT		GTTC	VKDWLHLPSCT
					TCTATTCAACTTGGCAATGGACCCCCTGTTG		GTAA	CDAILYVSTRDG
					TGTAAGCTGGAAGAAAGTGGCAAAGGATTC		TA	GLGVTKLAGLIP
					CATCGAGGACAGAGCTCAATAACCGCGATG		(SEQ	SVQARRLHRIAQ
					GCGTTCGCCGACGATCTGGTCTTGTTAAGCG		ID NO:	SPDETMKDFLE
					ACTCCTGGGAGAACATGAAAGAGAACATCA		1265)	KAQMEKMYEK
					AAATACTGGAGACCTTTTGCAATCTCACCG			LWVQAGGKKK
					GTCTCAAAACACAGGGTCAGAAGTGCCACG			GMPSIWEALPM
					GCTTTTACATCAAGCCTACAAAGGACTCTTA			TVPPTNTGNLSE
					CACCATCAACAACTGCCCTGCATGGACCAT			WEAPNPKSKYP
					CAACGGCACACCCCTGAACATGATCAACCC			KPCDWRRKELK
					CGGGGAGTCAGAGAAATACCTCGGCCTGCA			KWTKLESQGRG
					GATCGACCCATGGACTGGAGTAGCAAAATA			VKNFRNDTISND
					CGATCTCTCCACAAAATTGAAAATATGGCT			WIQYYRRIPHRK
					CGAAAGCATTGACCGAGCTCCACTTAAACC			LLTAIQLRANVY
					TCTGCAAAAATTAGACATCCTCAAAACATA			PTREFLARGRGD
					CACCATTCCTCGACTGACCTACCTGGCTGAC			NYVKFCRHCEA

					CATTCAGAGATGAAAGCAGGGGCTCTGGAA GCACTCGACCAGCAGATTCGAACAGCGGTC AAAGACTGGCTGCACCTGCCCTCGTGCACC TGTGATGCCATCTTGTACGTGAGCACGAGG GACGGCGGTTTGGGTGTTACCAAGTTGGCG GGACTGATTCCAAGTGTGCAAGCCCGGAGG CTGCATCGCATTGCGCAGTCGCCGGACGAG ACGATGAAGGACTTCCTAGAGAAGGCGCAG ATGGAGAAGATGTATGAGAAGTTATGGGTT CAAGCTGGAGGCAAAAAGAAGGGGATGCC GTCAATTTGGGAGGCCCTACCGATGACTGT ACCACCCACTAATACAGGTAATCTTTCGGA GTGGGAAGCACCGAACCCCAAAAGTAAGTA CCCAAAACCTTGTGATTGGAGAAGGAAAGA GCTTAAAAAGTGGACAAAATTGGAGTCCCA AGGTCGTGGAGTCAAAAATTTTAGGAATGA TACAATTAGTAACGATTGGATCCAATATTAT AGACGCATACCTCACAGGAAACTCCTCACT GCCATACAACTCAGGGCCAATGTATACCCC ACAAGGGAATTTCTCGCGCGGGGGAGGGGT GATAACTATGTTAAGTTTTGTAGGCACTGTG AAGCGGACCTTGAAACCTGTGGCCATATCA TCGGCTTTTGCCCAGTAACGAAGGACGCCC			DLETCGHIIGFCP VTKDARIKRHN RICDRLCEEAAK REWVVFKEPHL RDATTELFKPDV IFVKEDRALVVD VTVRYESAKTT LEAAAMEKVDK YKHLEAEVKEL TNAKDVVFMGF PLGARGKFYKG NFNLLETLGLPK TRQLSVAKTLST YALMSSVDIVH MFASRSRKPNV (SEQ ID NO: 1387)

					GAATCAAGAGGCACAATCGCATATGCGACA GGCTTTGCGAGGAGGCAGCTAAGAGGGAAT GGGTGGTCTTCAAGGAGCCGCACTTGAGGG ATGCCACCACGGAACTGTTTAAACCGGATG TGATATTCGTGAAAGAGGACCGTGCACTGG TTGTGGATGTGACAGTACGATATGAATCAG CCAAGACAACGCTGGAGGCAGCTGCTATGG AGAAAGTGGACAAGTACAAACATCTGGAG GCAGAAGTGAAGGAACTCACCAACGCAAA GGACGTTGTTTTTATGGGGTTCCCCCTTGGA GCGCGAGGGAAATTCTACAAAGGGAACTTT AACTTGCTAGAGACTCTTGGCCTCCCAAAA ACGAGGCAATTGAGTGTGGCAAAGACTCTA TCCACGTACGCGCTCATGTCATCTGTGGACA TTGTGCATATGTTTGCCAGTAGATCTAGGAA ACCAAATGTCTAGGTAGTCACATTGCACTTT CTGTAACTTGCACTGGGTGTGGGATGTGGG CCTGGGGTGTGGGTTATGGGGTATATATGT GGGATATTCTGGTGGGAATGTCCATTCACT GTATGCCTATCTTTTTAATAAAAAGACGGTA GCTAGGTTCGCGAAGCAGCCACAAGCCAAT AGCCAGTTAGGTAGCTCATAGTGGGTAGGT GACAGGAACCTTTGACTCAGAACGCGTCCA			

					TTAACATCTAGAACGGACCAAACTTCGGAC ATGCACCGATTAACCGGATTTGTCCAAGGT GGACGGGCCACCTTTACTTAACCCGGAAAG GGAACATATATAGTTATATGTGTTCGTAATA (SEQ ID NO: 1541)			
R2	R2Dr	AB0971 26		Danio rerio	AATCCCCCCTACCCAATCCCCCCGTCGTGAC	AATCCC	AAAT	MESTAKGKSYW
					CTCCAGGCCAGGAATCACGAGCGTACGACA	CCCTAC	CCCA	MARRPVEGATE
					GTGGCCATCCGGCAATGACAATAGCGTGAC	CCAATC	GCGG	GSLGRVPFVTRD
					TAACGACAATGAGTCAGATCCATGACCCTT	CCCCCG	GATA	PKRKPEAKRTLT
					GGAGTGGGTTAACCTCCGCCTCTTTAAAAA	TCGTGA	CAGC	HGLGLRECSVV
					CATGGAAAGTACAGCAAAAGGAAAGTCAT	CCTCCA	AAGA	LTRLIEGRRGRD
					ACTGGATGGCCCGTCGCCCAGTAGAAGGTG	GGCCAG	AGGT	HTPSGWNAQRG
					CCACGGAGGGATCTTTGGGTCGGGTCCCTTT	GAATCA	ATCG	MPNDESSVEEPN
					CGTAACGCGAGATCCTAAGCGCAAACCAGA	CGAGCG	GATC	GPIPSNPIPTGTQ
					GGCTAAACGAACACTTACGCATGGCTTAGG	TACGAC	TAAT	ALPEPMADGEQ
					ACTACGAGAATGCTCGGTTGTCTTGACACG	AGTGGC	AAGG	GEHPGVVVTLP
					CCTCATCGAGGGGCGTCGAGGTCGCGATCA	CATCCG	TTGA	LRDLNCPLCGGS
					CACACCATCAGGATGGAACGCACAGCGCGG	GCAATG	GCGA	ASTAVKVQRHL
					CATGCCAAACGACGAAAGCTCGGTCGAGGA	ACAATA	GGAG	AFRHGTVPVRFS
					GCCCAATGGGCCGATACCATCTAACCCCAT	GCGTGA	AGGG	CESCGKTSPGCH
					ACCAACGGGCACCCAAGCCCTGCCTGAACC	CTAACG	TGGA	SVLCHIPKCRGP
					TATGGCGGACGGGGAGCAGGGGGAGCACC	ACAATG	GATC	TGEPPEKVVKCE
					CGGGAGTGGTGGTGACCCTGCCGCTCAGGG	AGTCAG	CTTTG	GCSRTFGTRRAC

					ACTTAAACTGCCCCCTATGTGGCGGGTCGG	ATCCAT	GGGG	SIHEMHVHSEIR
					CGAGCACCGCGGTGAAAGTGCAAAGACACT	GACCCT	GGGT	NRKRIAQDRQE
					TGGCATTTCGCCACGGAACAGTGCCGGTTA	TGGAGT	CGGG	KGTSTDGEGRA
					GATTCAGCTGTGAATCATGTGGAAAAACTT	GGGTTA	CTAA	GVERADAGEGP
					CTCCGGGTTGCCATTCCGTCCTCTGTCACAT	ACCTCC	GTTC	SGEGIPPKRPRR
					TCCGAAATGTCGCGGACCGACAGGCGAGCC	GCCTCT	CCCT	ARTPREPSEPPA
					GCCTGAGAAAGTGGTTAAGTGCGAGGGATG	TTAAAA	CTCG	NPPILSPQPDLPP
					CAGTAGGACGTTTGGCACAAGGAGAGCGTG	AC (SEQ	GGTC	GGLRDLLREVA
					TAGTATACATGAGATGCACGTTCACTCAGA	ID NO:	CTCC	SGWVRAARDG
					AATCCGCAATAGGAAAAGAATTGCTCAAGA	1143)	CACG	GTVIDSVLAAW
					CAGGCAAGAAAAAGGGACCTCGACAGATG		GTGA	LDGNDRLPELV
					GAGAGGGGAGAGCTGGAGTCGAAAGGGCT		CGCT	DAATQRTLQGL
					GACGCTGGGGAAGGTCCCTCTGGGGAAGGG		CTAC	PAGRLARRPATF
					ATCCCCCCTAAACGTCCCAGACGTGCGAGA		CCCT	VAPNRRRGRWG
					ACGCCCAGAGAACCGTCTGAGCCCCCCGCG		CCCT	RRLKLLAKRRA
					AATCCGCCGATTCTCTCGCCACAACCCGATC		CCTC	YHDCQIRFRKDP
					TGCCCCCAGGAGGCCTCCGGGACCTACTCC		GCTC	ARLAANILDGKS
					GGGAGGTGGCCAGTGGGTGGGTAAGGGCA		GTAG	ETSCPINEQAIHE
					GCGAGAGACGGAGGTACGGTGATTGACAGC		AACC	HFRNKWANPSP
					GTGCTCGCAGCATGGTTGGATGGCAACGAT		CAAC	FGGLGRFGTEN
					CGGCTCCCTGAGCTGGTTGACGCGGCGACG		GGTG	RANNAHLLGPIS
					CAAAGGACACTGCAGGGCTTACCTGCAGGG		AACA	KSEVQTSLRNAS
					AGGTTGGCCCGAAGACCCGCAACTTTTGTT		CGGT	NASTPGPDGVG

					GCGCCTAACCGGAGGAGAGGCAGGTGGGG		TGGC	KRDISNWDPECE
					GCGCCGGCTCAAACTGCTCGCTAAGCGCCG		AGGA	TLTQLFNMWWF
					CGCCTACCACGATTGCCAAATTCGGTTCCG		TGAA	TGVIPSRLKKSR
					AAAAGACCCAGCCCGCCTAGCCGCGAACAT		GTGA	TVLLPKSSDPGA
					CCTAGACGGCAAAAGCGAAACAAGTTGCCC		CGTG	EMEIGNWRPITI
					AATCAATGAGCAAGCGATTCATGAGCACTT		AGGG	GSMVLRLFTRVI
					TCGAAACAAATGGGCAAATCCAAGTCCATT		GTAA	NTRLTEACPLHP
					TGGTGGGCTGGGACGATTTGGGACGGAAAA		GACA	RQRGFRRSPGCS
					CAGGGCCAACAACGCCCACCTCCTCGGGCC		TGCG	ENLEVLECLLRH
					AATCTCCAAAAGCGAGGTCCAAACTAGCCT		TACG	SKEKRSQLAVV
					CCGAAATGCATCGAACGCCTCCACACCAGG		TGAG	FVDFAQAFDTV
					CCCAGACGGCGTTGGGAAAAGGGACATTTC		CGCG	SHEHMLSVLEQ
					CAACTGGGATCCTGAGTGTGAGACCCTCAC		CATTT	MNVDPHMVNLI
					TCAGCTGTTTAACATGTGGTGGTTCACAGGT		TTGCT	REIYTNSCTSVE
					GTCATCCCCTCTCGCTTGAAGAAAAGTCGT		GTTCT	LGRKEGPDIPVR
					ACGGTGCTTCTGCCCAAGTCCTCAGACCCA		CTGG	VGVKQGDPLSP
					GGAGCGGAGATGGAGATCGGCAACTGGAG		ACTG	LLFNLALDPLIQ
					ACCAATCACCATCGGGTCGATGGTCTTGCG		GGTT	SLERTGKGCEAE
					GCTTTTCACAAGGGTGATCAATACGAGATT		TCGT	GHKVTALAFAD
					AACGGAAGCCTGTCCGTTGCACCCAAGACA		CCCC	DLALVAGSWEG
					GAGAGGGTTTCGACGAAGCCCCGGGTGTTC		CTCA	MAHNLALVDEF
					GGAGAACCTGGAAGTACTCGAATGTCTCCT		CAAC	CLTTGLTVQPK
					CCGACACTCCAAAGAAAAGCGCAGCCAACT		CATC	KCHSFMVRPCR

					GGCAGTGGTATTCGTCGATTTTGCACAAGC		ACTT	GAFTVNDCPPW
					GTTTGACACCGTCTCTCATGAACACATGCTG		ACAC	VLGGKALQLTNI
					TCAGTCCTTGAGCAGATGAACGTGGATCCC		TATA	ENSIKYLGVKV
					CACATGGTAAATCTGATCCGGGAGATTTAC		GGGG	NPWAGIEKPDLT
					ACAAACAGCTGCACAAGTGTCGAGCTAGGC		CACA	VALDRWCKRIG
					CGGAAAGAGGGACCAGACATCCCAGTGAG		GCGG	KSLLKPSQKVYI
					GGTTGGTGTTAAGCAAGGGGATCCTCTGTC		CTCCT	LNQFAIPRLFYL
					CCCGCTGCTTTTCAACCTGGCTTTGGATCCT		ACCT	ADHGGAGDVM
					CTCATCCAAAGTCTCGAACGCACAGGCAAA		CCCT	LQNLDGTIRKA
					GGGTGTGAGGCCGAAGGTCACAAAGTGACA		CCCT	VKKWLHLPPST
					GCTTTAGCGTTCGCGGATGACCTGGCACTG		ATGA	CNGLLYARNCN
					GTTGCGGGCTCGTGGGAGGGAATGGCACAC		CCCC	GGLGICKLTRHI
					AACCTTGCGCTTGTAGACGAATTCTGCCTAA		CCCTT	PSMQARRMFRL
					CCACCGGCCTCACAGTCCAACCCAAAAAGT		CCCA	ANSSDPLMKAM
					GCCACAGTTTCATGGTCAGGCCCTGCAGAG		TACC	MRGSRVEQKFK
					GTGCCTTCACAGTGAACGACTGCCCCCCAT		GATC	KAWMRAGGEE
					GGGTTCTGGGGGGCAAGGCCCTGCAGCTAA		CATG	SALPRVFGANQ
					CAAACATCGAAAACTCCATCAAATATCTGG		GCTG	YQEGEEVANDL
					GAGTAAAAGTCAATCCTTGGGCGGGGATTG		TTCTA	VPRCPMPSDWR
					AAAAGCCTGACCTTACAGTGGCACTAGACC		GTCT	LEEFQHWMGLP
					GATGGTGCAAGCGCATTGGGAAGTCACTGC		GGAC	IQGVGIAGFFRN
					TCAAACCCTCACAGAAGGTATACATTCTCA		CGAG	RVANGWLRKPA
					ATCAGTTTGCCATCCCGCGACTCTTCTACCT		GGTC	GFKERHYIAALQ

					GGCTGATCACGGTGGGGCCGGCGACGTCAT		GGAC	LRACVYPTLEFQ
					GCTCCAGAACCTGGATGGGACAATCAGGAA		GGGG	QRGRSKAGAAC
					GGCGGTGAAGAAATGGCTGCATCTTCCACC		CATTT	RRCSSRLESSSHI
					GTCAACCTGCAACGGGCTGTTGTATGCCAG		GAAG	LGKCPAVQGAR
					GAACTGTAATGGTGGCCTCGGTATATGCAA		GTAG	IRRHNKICDLLK
					GCTCACTCGGCACATCCCATCAATGCAGGC		CTGG	AEAETRGWEVR
					GAGACGAATGTTCCGCTTGGCCAACTCATC		AATC	REWAFRTPAGE
					GGACCCGTTGATGAAGGCCATGATGCGCGG		CTCC	LRRLDLVLILGD
					CTCCCGAGTCGAACAGAAATTCAAAAAGGC		GCTG	EALVIDVTVRYE
					CTGGATGCGGGCCGGGGGAGAGGAGAGTG		CTGC	FAPDTLQNAGK
					CGCTCCCACGGGTGTTCGGGGCGAATCAGT		GAGC	DKVSYYGPHKE
					ACCAGGAAGGGGAGGAGGTCGCTAACGAT		CTGA	AIARELGVRRV
					CTGGTACCTCGCTGCCCAATGCCGAGCGAT		GGTC	DIHGFPLGARGL
					TGGAGACTGGAAGAATTCCAACACTGGATG		GATG	WLASNSKVLEL
					GGCCTGCCGATCCAGGGTGTGGGTATAGCC		GTTA	MGLSRERVKVF
					GGCTTCTTCAGAAACAGGGTGGCTAACGGA		GAGG	SRLLSRRVLLYS
					TGGCTCAGGAAGCCGGCAGGGTTCAAAGAG		TGAA	IDIMRTFYATLQ
					CGGCACTACATCGCCGCTCTACAACTGCGA		ATAC	(SEQ ID NO:
					GCATGTGTATACCCCACCCTCGAATTCCAGC		TTGG	1388)
					AAAGGGGCAGGAGCAAAGCGGGTGCGGCC		GAGG	
					TGCAGGCGGTGCTCATCCCGGTTGGAATCC		AGAC	
					AGCTCTCACATCCTCGGCAAATGTCCGGCG		ACAG	
					GTGCAGGGAGCCAGAATCAGGCGTCATAAC		CCTC	

					AAAATATGCGACCTCCTGAAGGCCGAAGCC		CGGA	
					GAAACCCGGGGTTGGGAGGTACGCCGGGA		GAGC	
					ATGGGCCTTCAGAACTCCGGCTGGGGAACT		CCCT	
					GAGAAGGCTCGACCTGGTACTCATCCTCGG		CCCG	
					GGATGAGGCATTGGTCATTGACGTCACAGT		GGTG	
					AAGGTACGAGTTCGCTCCGGATACCCTCCA		GTCA	
					GAATGCCGGAAAGGACAAGGTCAGCTACTA		TCAT	
					CGGCCCGCACAAAGAAGCGATCGCTCGGGA		GGCA	
					GCTGGGCGTAAGAAGGGTCGACATACATGG		ACCG	
					GTTTCCGTTGGGTGCACGCGGACTTTGGCTC		GGTG	
					GCCAGCAACTCCAAAGTGCTGGAACTGATG		AAAC	
					GGATTGAGCAGGGAAAGAGTGAAGGTCTTC		CTTA	
					TCCAGACTCTTGAGTCGGAGAGTGCTCCTGT		CGGT	
					ACTCTATCGACATCATGAGGACATTTTACGC		TTCA	
					AACCCTGCAATGAAAATCCCAGCGGGATAC		CTTA	
					AGCAAGAAGGTATCGGATCTAATAAGGTTG		CGAA	
					AGCGAGGAGAGGGTGGAGATCCTTTGGGGG		ACAG	
					GGGTCGGGCTAAGTTCCCCTCTCGGGTCCTC		CACC	
					CCACGGTGACGCTCTACCCCTCCCTCCTCGC		ATAA	
					TCGTAGAACCCAACGGTGAACACGGTTGGC		CAGC	
					AGGATGAAGTGACGTGAGGGGTAAGACAT		GCCG	
					GCGTACGTGAGCGCGCATTTTTGCTGTTCTC		TAAT	
					TGGACTGGGTTTCGTCCCCCTCACAACCATC		AGCG	

					ACTTACACTATAGGGGCACAGCGGCTCCTA		CACC	
					CCTCCCTCCCTATGACCCCCCCTTCCCATAC		GGTG	
					CGATCCATGGCTGTTCTAGTCTGGACCGAG		TGAC	
					GGTCGGACGGGGCATTTGAAGGTAGCTGGA		TACT	
					ATCCTCCGCTGCTGCGAGCCTGAGGTCGAT		GTCC	
					GGTTAGAGGTGAAATACTTGGGAGGAGACA		AGTG	
					CAGCCTCCGGAGAGCCCCTCCCGGGTGGTC		CTGA	
					ATCATGGCAACCGGGTGAAACCTTACGGTT		TATTC	
					TCACTTACGAAACAGCACCATAACAGCGCC		TCAT	
					GTAATAGCGCACCGGTGTGACTACTGTCCA		CTGG	
					GTGCTGATATTCTCATCTGGAGAATACAAC		AGAA	
					ACGGGTAATGGCAGAGTATTCAAAACCCAA		TACA	
					ATGTTTACGATCGACCAACGGAGTCGTTCC		ACAC	
					CTTGCATCTAGGCCGGACCCGAAACTGCCG		GGGT	
					TAATTGCCCGTCCCCAAGGTAGCCTCTTAGA		AATG	
					AAACCGAAGCCCGGTCGGGGCGGTGGTTGC		GCAG	
					GGCGGCGCTGCGGGGGCCTGCTGCTCGGGC		AGTA	
					GGCGTCGGTGTGCCGCGGTGGTTGCGGTGG		TTCA	
					TGCGGCGGGGATCTCGGTCCTTGCGGTGCC		AAAC	
					GCTGTGCCGCCGCGGTCGCGTCGGTGGCGC		CCAA	
					TGGGGTGGTGGCCCGAGTGGCGTCGGCGTG		ATGT	
					CCACTGCCCATAGTCGCCCGCGGGGGCGAC		TTAC	
					CGATCTGGAGGGGCGAGGGGGCTCGCGGG		GATC	

					ACTTTAACGAGAAACGGAACGCAACTTCTC GCATCGCTCCCGGGACTTTCCCCCCTCGTTC AGCCGAGGGATGCCAAAAGGCATGAAAGG TAAGTACCATACCGGTCCGCAAAACTCTCTT CTGACTCGGTTCTCTGTTGGTTTTCTAGAGT AACAACGAGGTGGAGGAGAGGGACATGGC AGGGACTCCCATTCGTGCCAGCGGGTGGGG ACAGATCGAAGGAACGGTTCGAGGGCGTAA CAGACGAGAGGGAATCCGGTCACACATTGA TGCCATGCCTAAATAGGCGAGGTTTGTATTT CTACTTTGTGGGTTCAGTATAGTCGGAGCAT ATGGTCGGTTGTCCCGTTGTTTTCACGGCGG GCAAGCGACTATCATGATAAAGTAGAATGG GAGACGGGCTCCCTGACAAACCCGGAAAGG CGCCCCCCCGTGGTTCGTAGCAGCTGACGG ATCACGCTCGAAGAAAAATGAGTGAGAGG GGACGCCGCAACCAC (SEQ ID NO: 1542)		GACC AACG GAGT CGTT CCCTT GCAT CTAG GCCG GACC CGAA ACTG CCGT AATT GCCC GTCC CCAA GGTA GCCT CTTA GAAA ACCG AAGC CCGG	

							TCGG GGCG GTGG TTGC GGCG GCGC TGCG GGGG CCTG CTGC TCGG GCGG CGTC GGTG TGCC GCGG TGGT TGCG GTGG TGCG GCGG GGAT CTCG	

							GTCC TTGC GGTG CCGC TGTG CCGC CGCG GTCG CGTC GGTG GCGC TGGG GTGG TGGC CCGA GTGG CGTC GGCG TGCC ACTG CCCA TAGT CGCC	

							CGCG GGGG CGAC CGAT CTGG AGGG GCGA GGGG GCTC GCGG GACT TTAA CGAG AAAC GGAA CGCA ACTT CTCG CATC GCTC CCGG GACT TTCCC	

							CCCT CGTT CAGC CGAG GGAT GCCA AAAG GCAT GAAA GGTA AGTA CCAT ACCG GTCC GCAA AACT CTCTT CTGA CTCG GTTCT CTGTT GGTT TTCTA	

							GAGT AACA ACGA GGTG GAGG AGAG GGAC ATGG CAGG GACT CCCA TTCGT GCCA GCGG GTGG GGAC AGAT CGAA GGAA CGGT TCGA GGGC GTAA	

							CAGA CGAG AGGG AATC CGGT CACA CATT GATG CCAT GCCT AAAT AGGC GAGG TTTGT ATTTC TACTT TGTG GGTT CAGT ATAG TCGG AGCA TATG	

							GTCG GTTG TCCC GTTG TTTTC ACGG CGGG CAAG CGAC TATC ATGA TAAA GTAG AATG GGAG ACGG GCTC CCTG ACAA ACCC GGAA AGGC GCCC	

							CCCC GTGG TTCGT AGCA GCTG ACGG ATCA CGCT CGAA GAAA AATG AGTG AGAG GGGA CGCC GCAA CCAC (SEQ ID NO: 1266)	
R2	R2-1_GA	.		Gasterosteus aculeatus	CATATTGGGGTCTCAGGAGGAGACACAGGG TCTGTTGCGGCTCCGGTAAACGGTACCGGA GTCGGTTAAGCATCGTTTGGGCCCGCCTCCA	CATATT GGGGTC TCAGGA	GGAG GGGA GTAG	MLRGGVGTPPA GGAGAVGPGM ASPGGCSVRFSP

					CGTGGTGGTCCGCGGTAACACCAATAGGGT	GGAGAC	GTCT	GGRRLLGHRTG
					GGCTAAGAGGCCCAGTAATTTCCCCGAATT	ACAGGG	CTAC	GLSPSVSWRLK
					GTCTTCCCCCCCGCGCGGGGGGGACCCCCC	TCTGTT	TCTG	RLSVSLRRWSGP
					TTTAGTGTCGGAGCGGTCGCGCCTCCGCGTT	GCGGCT	ACCC	GLLGADGAGGG
					TGGGGTGTCGCAGGCGTGAGCCTTCGTCCC	CCGGTA	GAAG	AAVASPRGTQV
					CTTAAGTTCAGACGGTCCCGGCTTCTTGCCG	AACGGT	GGCC	LGSGAGRRWLG
					GGCCAACCCCCGGTGCAGCGTTCTCCCATG	ACCGGA	CCCC	HGSRGSSPSAAR
					TTGGATCGGCACCCAGCCCCGGGTGCCATG	GTCGGT	CGTTT	GLRRLTVRLKR
					CGAGTTCAGACATTTTGTTTATGTATCGTCT	TAAGCA	CAGA	LSGGLLSPKACR
					GCGTGGTTGACTTGCTAAGCTCATTTCCTCC	TCGTTT	CCTG	DAEEGSSSSPGF
					TCTCACTGCGTCCCCCCAGGTGCTGATCGGT	GGGCCC	ATTCT	RNPKGLGGRGL
					TGAAGAGGATTCGTCGTTGACCTCGGCGGT	GCCTCC	AGGC	TPLGSRRFCRLT
					GAATTTGGGATTGTATTATACAGGTAGGTA	ACGTGG	TACC	VSLNRWRGSLV
					TAGAGGGCGTGCGGATGTTGCGTGGCGGTG	TGGTCC	TGTG	KLNASSRASGR
					TTGGTACTCCCCCGGCTGGGGGAGCGGGTG	GCGGTA	CCTA	RTPVKPACDSR
					CGGTGGGGCCAGGCATGGCCTCGCCGGGTG	ACACCA	ATTG	AGRGSEHAEGG
					GTTGCAGTGTCCGGTTCAGTCCCGGAGGGA	ATAGGG	GGGG	GVSAAPMVLRS
					GGCGACTGCTTGGCCACAGGACTGGAGGGT	TGGCTA	GGTC	RRKLTFSVDGDS
					TGAGTCCCTCCGTGTCCTGGAGGCTCAAGC	AGAGGC	CCAA	NSGDRARSGSV
					GACTGTCTGTCTCTCTGAGGCGCTGGAGCG	CCAGTA	AGAG	SAARPGHLLVD
					GGCCTGGGCTGCTAGGTGCGGATGGTGCGG	ATTTCC	ATGT	GESASSRSGPAG
					GGGGAGGCGCTGCGGTGGCCTCCCCCAGGG	CCGAAT	TGTCT	DARLAGPSTRSR
					GTACGCAGGTCCTGGGAAGTGGGGCCGGGC	TGTCTT	GTTG	RKGCLPPVDFEN

					GTCGGTGGCTTGGGCACGGGTCGCGAGGGT	CCCCCC	TAGA	PKKRTRLMAKM
					CTTCTCCTTCTGCGGCCCGGGGGCTAAGGC	CGCGCG	AGGG	TNGNPTSHVPCP
					GGCTGACGGTACGGTTGAAGCGACTCAGCG	GGGGGG	TTTGC	APCSNGHEGGG
					GTGGCCTGTTGTCCCCTAAGGCGTGTCGGG	ACCCCC	GCCA	RVAVIEGRLPEL
					ATGCGGAAGAAGGAAGCTCCAGCAGCCCA	CTTTAG	CTGA	SGSRISGIQPALP
					GGGTTCCGGAATCCAAAAGGTCTCGGGGGA	TGTCGG	CTGC	VETSFVGQSTGR
					AGGGGGTTGACGCCTCTCGGATCCCGTAGA	AGCGGT	ACGG	GADGDANANSS
					TTTTGTCGGCTGACCGTCTCCCTGAATCGCT	CGCGCC	AAGG	PPSPNLGGSVG
					GGAGGGGCAGTCTGGTGAAGTTGAACGCTA	TCCGCG	GTGG	MVPAVRDGTPP
					GTAGCAGGGCCTCCGGCCGGAGGACCCCTG	TTTGGG	GCCT	LGRPGEDHSREC
					TGAAACCCGCTTGTGACTCTAGAGCCGGAC	GTGTCG	CGAC	AGGNTPLWMLE
					GGGGCTCGGAGCATGCGGAGGGAGGTGGA	CAGGCG	AGGT	DSFRCDYCPREF
					GTGAGCGCTGCACCTATGGTGTTGCGCAGT	TGAGCC	AGGG	GTRAGRSLHMR
					CGGCGTAAGCTCACCTTCTCTGTGGATGGC	TTCGTC	GTTA	RAHLAEYDGAG
					GACTCTAACTCCGGGGATAGGGCCCGGAGC	CCCTTA	CATG	FCWGERLSEFA
					GGGTCCGTCTCTGCAGCCCGTCCTGGCCACT	AGTTCA	ACTC	ATRLWSTEETK
					TGTTGGTGGATGGTGAGAGTGCGTCCTCAA	GACGGT	CGTG	KLAVFCERGVPS
					GATCTGGCCCCGCGGGGGATGCCAGGTTGG	CCCGGC	CTGC	PSECRAIAASLG
					CGGGGCCTTCTACGCGGAGTAGGAGGAAGG	TTCTTG	TCAG	AGKTHHQVRSK
					GTTGCCTTCCCCCGGTCGACTTTGAAAACCC	CCGGGC	CAGA	CRLVFEAIRRRE
					GAAGAAGCGCACACGGTTGATGGCTAAGAT	CAACCC	CCCG	LLEVAAATERLE
					GACGAATGGTAATCCTACCTCGCACGTCCC	CCGGTG	CGCC	KSARRKQPAVP
					TTGCCCTGCCCCGTGCTCAAATGGGCATGA	CAGCGT	TCTG	PAPVHGVRGVL

					AGGAGGTGGGCGAGTTGCGGTGATCGAGGG	TCTCCC	AGAC	RGLLGKRVPRE
					GCGGCTGCCGGAGTTAAGCGGTAGTAGGAT	ATGTTG	CGGG	GGTTGSTSARIV
					CTCTGGAATACAGCCAGCCCTGCCTGTTGA	GATCGG	TAGG	RRDDCRQGAVA
					AACCAGCTTTGTCGGCCAATCGACTGGCCG	CACCCA	GCTA	SASLNLIRRLGR
					GGGCGCGGACGGCGATGCGAATGCGAATA	GCCCCG	CTTG	KATGRSGRRRV
					GTAGCCCGCCTTCTCCTAATCTGGGCGGCTC	GGTGCC	AACA	LGRPPRMDVRR
					GGTTGGGATGGTGCCTGCCGTGCGTGATGG	ATGCGA	AGCG	SVRMRRMRRFL
					TACCCCGCCGCTTGGGCGTCCAGGAGAGGA	GTTCAG	ACGC	YRLARLGWAKL
					TCACTCGCGGGAGTGTGCAGGGGGAAATAC	ACATTT	CCTG	AMFVLDGQMG
					TCCCCTCTGGATGCTGGAGGACAGTTTCCG	TGTTTA	GTGT	ASCPVPLVEVSA
					GTGTGACTACTGTCCTAGGGAATTCGGCAC	TGTATC	ATGT	VFRERWSIVRAF
					AAGAGCGGGGCGCTCGTTGCACATGCGCAG	GTCTGC	CCGT	LGLGQFGGFGT
					GGCTCACCTGGCCGAGTACGACGGGGCAGG	GTGGTT	ATCC	ADNAGFGKLIDP
					TTTCTGTTGGGGTGAACGTCTCAGTGAATTC	GACTTG	TAAC	AEVRAHLQSIKN
					GCCGCTACGCGCCTCTGGTCGACGGAGGAA	CTAAGC	CTGG	RSSPGPDGITKV
					ACCAAAAAGCTGGCCGTGTTTTGTGAGAGG	TCATTT	TTTG	ALSKWDPEGIKL
					GGTGTGCCCTCACCGTCGGAATGCAGAGCC	CCTCCT	GGAA	AHMYSTWLVSA
					ATTGCAGCCTCTCTGGGCGCAGGAAAAACA	CTCACT	AGCC	GIPKVFKKCRTT
					CATCATCAGGTTAGATCGAAGTGTCGACTG	GCGTCC	GATA	LIPKTGDVSLHG
					GTGTTCGAGGCCATTCGGCGGCGTGAATTG	CCCCAG	CCGG	DVGQWRPITIAS
					CTTGAGGTGGCTGCTGCCACGGAGCGTTTG	GTGCTG	CAAT	LVLRLYSRILTE
					GAGAAAAGCGCTAGGCGGAAGCAGCCCGC	ATCGGT	GCCC	RMTVACPSHPR
					CGTACCACCGGCACCCGTACACGGAGTGAG	TGAAGA	GCCA	QRGFIASPGCSE

					AGGGGTCCTGCGGGGCCTACTAGGGAAGCG	GGATTC	CAGG	NLMLLEGCMSL
					GGTGCCGAGAGAGGGTGGTACCACAGGCA	GTCGTT	TGTC	SKAGNGSLAVV
					GCACCTCAGCAAGGATCGTCAGGAGAGACG	GACCTC	GCGC	FVDFAKAFDTV
					ACTGCCGTCAGGGGGCAGTTGCGTCGGCTT	GGCGGT	ACCC	SHEHLLSVLVQ
					CTCTCAATCTGATCAGAAGGCTGGGTCGAA	GAATTT	CACG	KGLDQHMVELI
					AGGCAACGGGCCGCTCCGGCAGGAGACGG	GGGATT	GGAT	KDSYENSVTKV
					GTCCTTGGACGCCCACCCAGGATGGATGTA	GTATTA	GACG	HCQEGCSTDIA
					AGGCGTAGCGTGAGGATGAGGAGGATGCG	TACAGG	TATG	MKVGVKQGDS
					CAGGTTCCTCTATCGGTTGGCCCGGCTGGGC	TAGGTA	GGCC	MSPLLFNLALDP
					TGGGCCAAGTTGGCTATGTTTGTCCTGGACG	TAGAGG	CCGG	LIQQLEREGRGF
					GACAGATGGGGGCGAGCTGCCCCGTTCCAC	GCGTGC	GGGA	PVNGKSITAMAF
					TCGTCGAAGTGTCGGCGGTCTTCCGGGAGA	GG (SEQ	CCTC	ADDLAIVSDSW
					GGTGGAGCATAGTCAGAGCCTTCCTGGGTC	ID NO:	ATGG	EGMRANLDILV
					TGGGTCAGTTCGGGGGCTTCGGGACTGCCG	1144)	ATAC	DFCELTGMRTQ
					ACAACGCAGGATTTGGGAAGCTGATCGATC		TCCA	PSKCHGFLIEKS
					CGGCTGAAGTCAGGGCCCATCTCCAGTCCA		CTGG	GSRSYKVNRCE
					TCAAGAACCGGTCTTCCCCGGGCCCGGATG		ACTT	PWLLNDTALHM
					GCATCACCAAGGTGGCGCTGTCCAAATGGG		GCAC	VGPKESIKYLGV
					ACCCCGAAGGGATTAAATTGGCGCACATGT		AATC	QVNPWTGIFAE
					ACTCAACATGGTTGGTATCGGCAGGCATCC		CTGG	DTVAKLRQWV
					CTAAGGTCTTCAAGAAGTGCAGGACGACAC		TGTA	VAISKTPLRPLD
					TTATCCCAAAGACCGGGGACGTTAGTCTAC		CTGG	KVSLLCQFAVPR
					ATGGTGACGTGGGGCAATGGAGGCCCATAA		ATGC	VIFVADHCMLS

					CCATTGCGTCCCTGGTCCTGAGACTCTATTC		AGCG	AKALTEMDRSIR
					GCGGATCCTGACGGAAAGGATGACAGTGGC		ACGT	QAVKRWLHLAR
					CTGTCCTAGCCACCCGCGCCAGAGGGGCTT		TGGT	CTTNGLLYSRKS
					CATTGCCTCCCCGGGCTGTTCGGAAAACCTC		GACA	SGGLGIPKLSMI
					ATGCTGTTGGAAGGTTGCATGAGTCTCAGC		TAAG	VPAMQARRLLG
					AAGGCAGGAAATGGCTCCCTCGCGGTTGTG		CAAT	LSRSKDETVRW
					TTCGTCGACTTTGCGAAGGCCTTCGATACCG		CGCT	MFLETTDHVAF
					TCTCCCACGAGCACCTCCTGAGTGTTCTGGT		AAGT	ERAWLRAGGSP
					GCAGAAAGGCTTGGACCAACACATGGTGGA		CGGG	DEVPELGPDLVE
					GTTGATCAAGGACTCCTACGAGAACAGCGT		GTAG	GSPAEGNADPV
					GACCAAGGTGCACTGTCAGGAGGGTTGTTC		GGGA	STVRPRKRIVPC
					CACTGACATCGCCATGAAGGTGGGAGTGAA		GGTG	DWRQVEFDRW
					GCAGGGTGACTCCATGTCCCCTCTCCTCTTT		GGGA	AGQLVQGKGIR
					AACCTGGCGCTGGATCCGCTTATCCAGCAA		CCTC	TFEADKISNCWL
					CTTGAACGCGAGGGCCGGGGCTTCCCAGTA		GGCA	YDYPPNKLKPG
					AATGGGAAGTCCATTACTGCGATGGCATTT		CGGC	DFTAAVQLRAN
					GCGGATGACTTGGCCATAGTGAGTGACTCT		TGTA	VYPTRELAGRG
					TGGGAAGGCATGAGAGCCAACCTTGATATC		GGAA	RTDTIDVCCRHC
					CTGGTGGACTTCTGCGAGCTTACTGGAATG		CGGG	GEAPETCWHILA
					CGAACCCAGCCCAGTAAGTGCCACGGGTTC		TGTA	LCPKVKRCRIQR
					CTGATTGAGAAGAGTGGCAGCAGGTCGTAC		TGGG	HHKVCQVLVAE
					AAAGTGAACAGGTGCGAACCGTGGCTGCTG		CTCC	AERHGWEVERE
					AACGACACAGCTCTTCACATGGTCGGGCCT		GGCA	KRWMLPSGECV

					AAGGAATCAATCAAGTACCTGGGCGTCCAG		GCCG	APDLICWLDELA
					GTGAACCCGTGGACAGGGATCTTCGCTGAG		TCGT	LIVDVTVRYEFD
					GATACGGTTGCCAAACTACGACAGTGGGTA		CACT	EESLERARIEKE
					GTTGCAATCTCCAAGACGCCTCTACGTCCGC		CCCA	CKYRPLIPVIRAS
					TTGACAAGGTGTCCCTGTTGTGCCAGTTTGC		TACA	RVQTKKVTVYG
					CGTACCGAGGGTCATCTTCGTGGCTGATCA		ACAC	FPLGARGKWPA
					CTGCATGCTATCTGCGAAGGCCCTGACAGA		AGGG	KNELLLADLGLS
					AATGGATAGGAGCATAAGACAAGCAGTGA		GCTG	KARTRSFAKLLS
					AGAGGTGGTTGCACCTGGCCAGGTGTACCA		CATC	RRVLLHSLDVM
					CGAACGGCCTCCTCTACTCAAGGAAATCCA		CTGG	RTFMR (SEQ ID
					GCGGTGGTCTGGGTATCCCAAAATTGTCGA		TGGC	NO: 1389)
					TGATTGTTCCGGCCATGCAGGCCAGGAGAC		CGGT	
					TCCTGGGCCTGTCCCGTTCTAAGGACGAGA		GCTA	
					CGGTCAGGTGGATGTTTCTGGAGACAACTG		GTTG	
					ATCACGTGGCGTTTGAGAGGGCATGGCTGA		GTTCT	
					GGGCTGGAGGGTCGCCAGATGAGGTACCGG		GGAA	
					AGCTGGGTCCGGATCTGGTGGAGGGCTCCC		GCCC	
					CTGCGGAGGGGAACGCTGACCCTGTCAGCA		GCCC	
					CGGTGAGGCCAAGGAAGCGCATAGTCCCGT		GGGC	
					GTGACTGGCGTCAAGTCGAGTTCGACAGAT		TGGT	
					GGGCCGGTCAATTGGTGCAGGGAAAAGGG		TCGC	
					ATTCGGACGTTCGAAGCGGACAAGATCAGC		AGAA	
					AACTGCTGGTTGTACGACTACCCGCCAAAC		GCAG	

					AAGCTGAAGCCTGGGGATTTTACGGCGGCT		GGTG	
					GTCCAGCTTAGAGCGAACGTTTACCCGACC		CGCC	
					CGGGAGCTAGCGGGTCGCGGAAGGACCGAT		CAGG	
					ACGATAGATGTCTGTTGTCGACACTGTGGG		GTAG	
					GAGGCCCCAGAGACTTGCTGGCACATCCTT		GTTT	
					GCGCTCTGCCCGAAGGTTAAGCGGTGCCGT		GGTA	
					ATTCAGAGGCACCACAAGGTGTGCCAGGTC		TATCT	
					CTCGTCGCGGAGGCTGAGCGCCATGGATGG		GGGT	
					GAAGTGGAAAGGGAAAAGCGCTGGATGCT		CCGG	
					GCCCTCCGGGGAGTGTGTCGCGCCGGACCT		TGCG	
					GATCTGCTGGTTGGATGAGCTGGCGCTCATT		ATAC	
					GTCGATGTGACGGTGAGGTACGAGTTCGAT		CTAT	
					GAGGAGTCGCTAGAACGCGCGCGAATCGAG		CGAT	
					AAGGAATGCAAGTACCGCCCTCTCATTCCA		GGGC	
					GTGATCAGGGCGAGCAGAGTTCAGACGAAG		AGCG	
					AAGGTGACGGTCTATGGCTTCCCTCTGGGA		AGGG	
					GCCAGGGGAAAGTGGCCTGCTAAGAACGA		CCGC	
					GCTGCTGCTCGCCGACCTCGGCCTGAGCAA		CTCG	
					GGCTCGGACTCGGAGTTTTGCTAAACTCCTG		TGAC	
					AGCCGCAGAGTTCTCTTACATTCTCTGGATG		GCGC	
					TTATGAGGACGTTTATGCGTTAAGGAGGGG		TGTG	
					AGTAGGTCTCTACTCTGACCCGAAGGGCCC		TGGA	
					CCCCGTTTCAGACCTGATTCTAGGCTACCTG		GCTG	

					TGCCTAATTGGGGGGGTCCCAAAGAGATGT		GAGC	
					TGTCTGTTGTAGAAGGGTTTGCGCCACTGAC		CGGC	
					TGCACGGAAGGGTGGGCCTCGACAGGTAGG		CTGG	
					GGTTACATGACTCCGTGCTGCTCAGCAGAC		GTAT	
					CCGCGCCTCTGAGACCGGGTAGGGCTACTT		GAAC	
					GAACAAGCGACGCCCTGGTGTATGTCCGTA		AGTT	
					TCCTAACCTGGTTTGGGAAAGCCGATACCG		CTTG	
					GCAATGCCCGCCACAGGTGTCGCGCACCCC		CGGA	
					ACGGGATGACGTATGGGCCCCGGGGGACCT		TGTG	
					CATGGATACTCCACTGGACTTGCACAATCCT		GCGT	
					GGTGTACTGGATGCAGCGACGTTGGTGACA		AGCT	
					TAAGCAATCGCTAAGTCGGGGTAGGGGAGG		AGAT	
					TGGGGACCTCGGCACGGCTGTAGGAACGGG		AGTA	
					TGTATGGGCTCCGGCAGCCGTCGTCACTCCC		CCCG	
					ATACAACACAGGGGCTGCATCCTGGTGGCC		TGGT	
					GGTGCTAGTTGGTTCTGGAAGCCCGCCCGG		TGTG	
					GCTGGTTCGCAGAAGCAGGGTGCGCCCAGG		GGCG	
					GTAGGTTTGGTATATCTGGGTCCGGTGCGAT		TGGT	
					ACCTATCGATGGGCAGCGAGGGCCGCCTCG		GTCG	
					TGACGCGCTGTGTGGAGCTGGAGCCGGCCT		ACCA	
					GGGTATGAACAGTTCTTGCGGATGTGGCGT		AATG	
					AGCTAGATAGTACCCGTGGTTGTGGGCGTG		TTGTC	
					GTGTCGACCAAATGTTGTCCTGTGTGCACAT		CTGT	

					AGGCCAAGGGTTACGTGGGTGGCAGTCAGA AGCACCCGCACCTGGAAGTGATTGCCCCGG GATCCCGGCTCTCTGTGAAGAGCTACCTTG AGGAAAGGTGTTCCGCTGGAACTCAAGACC CTACAGTAGGGGATATCAACTGGCTTTGAG GTGCTGTGATTCCGGAACCAGGGCGAGGGC GAGTACTTAGAGCATGTCCAAAAGCCCGGG GAACGTTCCGGGGGCCTGCTTGGGTCGTTG GACCCACATCCGTAAAACGATGGATCTCGC GTCGGCGCTCGGGAGAACTTCCCGCATGAA CGCTGATTGCATGTGAGAACGCCCCCACGG CGGCGGGGCAGGCGCTCCCCCTGGGTGTAA GGCTCGGGGGGGTCACGGCTCCGCTCTAAA AG (SEQ ID NO: 1543)		GTGC ACAT AGGC CAAG GGTT ACGT GGGT GGCA GTCA GAAG CACC CGCA CCTG GAAG TGAT TGCC CCGG GATC CCGG CTCTC TGTG AAGA GCTA	

							CCTT GAGG AAAG GTGT TCCG CTGG AACT CAAG ACCC TACA GTAG GGGA TATC AACT GGCT TTGA GGTG CTGT GATT CCGG AACC AGGG CGAG	

							GGCG AGTA CTTA GAGC ATGT CCAA AAGC CCGG GGAA CGTT CCGG GGGC CTGC TTGG GTCG TTGG ACCC ACAT CCGT AAAA CGAT GGAT CTCG	

							CGTC GGCG CTCG GGAG AACT TCCC GCAT GAAC GCTG ATTG CATG TGAG AACG CCCC CACG GCGG CGGG GCAG GCGC TCCC CCTG GGTG TAAG	

							GCTC GGGG GGGT CACG GCTC CGCT CTAA AAG (SEQ ID NO: 1267)	
R2	R2_BM	AB0768 41		Bombyx mori	GGGCGATACGCATAATTTTAATTTCCCGATT GAAATCCAGTCGTCTTAATCTGGTGACCAG TGGCGCGGTCACCAGTATAGTGCACAGGAC GTGAATGGCTCCGAGGCTGGCGGAGTCACT CACTATAAGTGTGAGAGACGATGTCCTGTG CCAAGTATACGTCCAACCCTAACGGGTTAA GTGAAATTAGTTGCTCATAACAGGGACGGT GTACCTGTTTGCTCGTGGCTGGCTATCGAAT GGACGGGACCAATACACCCCCCTGTTAGTA ATGGGGTAAGAGAGAGCGGTCTGAAACTAT GGCCGAAATCACGACGCCCCACTCCTACCC ATAACCTGCACGTGGTACCGCCGCACATTG	GGGCGA TACGCA TAATTT TAATTT CCCGAT TGAAAT CCAGTC GTCTTA ATCTGG TGACCA GTGGCG CGGTCA	GCCT TGCA CAGT AGTC CAGC GGTA AGGG TGTA GATC AGGC CCGT CTGTT	MMASTALSLMG RCNPDGCTRGK HVTAAPMDGPR GPSSLAGTFGW GLAIPAGEPCGR VCSPATVGFFPV AKKSNKENRPE ASGLPLESERTG DNPTVRGSAGA DPVGQDAPGWT CQFCERTFSTNR GLGVHKRRAHP

					ACCGATACGGGAGGAGGGGCAGCACTTGA	CCAGTA	TCTTC	VETNTDAAPMM
					ATCACGTAGTCTTGGTGTAGCCATTGCGGG	TAGTGC	CCCG	VKRRWHGEEID
					ACTACAGCCCTCGTAAGTGCCGCCTTAGAA	ACAGGA	GAGC	LLARTEARLLAE
					CGCAACGGGGCAATAGGTGGGCCGGGGCG	CGTGAA	TCGC	RGQCSGGDLFG
					CTAGCGGGGGGGAGTAATCTCCCCTGTTGG	TGGCTC	TCCCT	ALPGFGRTLEAI
					CGTGCACCGCACTGCTCCCACTGGGGGCAG	CGAGGC	TGGC	KGQRRREPYRA
					TGTCATCCGGAAACAGGTGGGCCGGGGCGC	TGGCGG	TTCCC	LVQAHLARFGS
					CACCAGGGGGGAGCAATCCCTCCTGATGAT	AGTCAC	TTAT	QPGPSSGGCSAE
					GGCGAGCACCGCACTGTCCCTTATGGGACG	TCACTA	ATTT	PDFRRASGAEEA
					GTGTAACCCGGATGGCTGTACACGTGGTAA	TAAGTG	AACA	VEERCAEDAAA
					ACACGTGACAGCAGCCCCGATGGACGGACC	TGAGAG	TCAG	YDPSAVGQMSP
					GCGAGGACCGTCAAGCCTAGCAGGTACCTT	ACGATG	AAAC	DAARVLSELLE
					CGGGTGGGGCCTTGCGATACCTGCGGGCGA	TCCTGT	AGAC	GAGRRRACRAM
					ACCCTGTGGTCGGGTTTGCAGCCCGGCCAC	GCCAAG	ATTA	RPKTAGRRNDL
					AGTGGGTTTTTTTCCTGTTGCAAAAAAGTCA	TATACG	AACA	HDDRTASAHKT
					AATAAAGAAAATAGACCTGAAGCCTCTGGC	TCCAAC	TCTA	SRQKRRAEYAR
					CTCCCGCTGGAGTCAGAGAGGACAGGCGAT	CCTAAC	CTGA	VQELYKKCRSR
					AACCCGACTGTGCGGGGTTCCGCCGGCGCA	GGGTTA	TCCA	AAAEVIDGACG
					GATCCTGTGGGTCAGGATGCGCCTGGTTGG	AGTGAA	ATTTC	GVGHSLEEMET
					ACCTGCCAGTTCTGCGAACGAACCTTTTCGA	ATTAGT	GCCG	YWRPILERVSDA
					CCAACAGGGGTTTGGGTGTCCACAAGCGTA	TGCTCA	GCGT	PGPTPEALHALG
					GAGCCCACCCTGTTGAGACCAATACGGATG	TAACAG	ACGG	RAEWHGGNRD
					CCGCTCCGATGATGGTGAAGCGGCGGTGGC	GGACGG	CCAC	YTQLWKPISVEE

					ATGGCGAGGAAATCGACCTCCTCGCTCGCA	TGTACC	GATC	IKASRFDWRTSP
					CCGAGGCCAGGTTGCTCGCTGAGCGGGGTC	TGTTTG	GGGA	GPDGIRSGQWR
					AGTGCTCGGGTGGAGACCTCTTTGGCGCGC	CTCGTG	GGGT	AVPVHLKAEMF
					TTCCAGGGTTTGGAAGAACTCTGGAAGCGA	GCTGGC	GGGA	NAWMARGEIPEI
					TTAAGGGACAACGGCGGAGGGAGCCTTATC	TATCGA	ATCT	LRQCRTVFVPK
					GGGCATTGGTGCAAGCGCACCTTGCCCGAT	ATGGAC	CGGG	VERPGGPGEYRP
					TTGGTTCCCAGCCGGGTCCCTCGTCGGGGG	GGGACC	GATC	ILIASIPLRHFHSI
					GGTGCTCGGCCGAGCCTGACTTCCGGCGGG	AATACA	TTCC	LARRLLACCPPD
					CTTCTGGAGCTGAGGAAGCGGTCGAGGAAC	CCCCCC	GATC	ARQRGFICADGT
					GATGCGCCGAAGACGCCGCTGCCTATGATC	TGTTAG	CTAA	LENSAVLDAVL
					CATCCGCAGTCGGTCAGATGTCGCCCGATG	TAATGG	TCCA	GDSRKKLRECH
					CCGCTCGGGTTCTCTCCGAACTCCTTGAGGG	GGTAAG	TGAT	VAVLDFAKAFD
					TGCGGGGAGAAGACGAGCGTGCAGGGCTAT	AGAGAG	GATT	TVSHEALVELLR
					GAGACCCAAGACTGCAGGGCGGCGAAACG	CGGTCT	ACGA	LRGMPEQFCGYI
					ATTTGCACGATGATCGGACAGCTAGTGCCC	GAAACT	CCTG	AHLYDTASTTL
					ACAAAACCAGTAGACAAAAGCGCAGGGCA	ATGGCC	AGTC	AVNNEMSSPVK
					GAGTACGCGCGTGTGCAGGAACTGTACAAG	GAAATC	ACTA	VGRGVRQGDPL
					AAGTGTCGCAGCAGAGCAGCAGCTGAGGTG	ACGACG	AAGA	SPILFNVVMDLI
					ATCGATGGCGCGTGTGGGGGTGTCGGACAC	CCCCAC	CGAT	LASLPERVGYRL
					TCGCTCGAGGAGATGGAGACCTATTGGCGA	TCCTAC	GGCA	EMELVSALAYA
					CCTATCCTCGAGAGAGTGTCCGATGCACCT	CCATAA	TGAT	DDLVLLAGSKV
					GGGCCTACACCGGAAGCTCTTCACGCCCTA	CCTGCA	GATC	GMQESISAVDC
					GGGCGTGCGGAGTGGCACGGGGGCAATCGC	CGTGGT	CGGC	VGKQMGLRLNC

					GACTACACCCAGCTGTGGAAGCCGATCTCG	ACCGCC	GATG	RKSAVLSMIPDG
					GTGGAAGAGATCAAGGCCTCCCGCTTTGAC	GCACAT	AAAA	HRKKHHYLTER
					TGGCGAACTTCGCCGGGCCCGGACGGTATA	TGACCG	(SEQ	TFNIGGKPLRQV
					CGTTCGGGTCAGTGGCGTGCGGTTCCTGTGC	ATACGG	ID NO:	SCVERWRYLGV
					ACTTGAAGGCGGAAATGTTCAATGCATGGA	GAGGAG	1268)	DFEASGCVTLEH
					TGGCACGAGGCGAAATACCCGAAATTCTAC	GGGCAG		SISSALNNISRAP
					GGCAGTGCCGAACCGTCTTTGTACCTAAGG	CACTTG		LKPQQRLEILRA
					TGGAGAGACCAGGTGGACCGGGGGAATATC	AATCAC		HLIPRFQHGFVL
					GACCGATCTTGATCGCGTCGATTCCCCTGAG	GTAGTC		GNISDDRLRML
					ACACTTTCACTCCATCTTGGCCCGGAGGCTG	TTGGTG		DVQIRKAVGQW
					TTGGCTTGCTGCCCCCCTGATGCACGACAGC	TAGCCA		LRLPADVPKAY
					GCGGATTTATCTGCGCCGACGGTACGCTGG	TTGCGG		YHAAVQDGGLA
					AGAATTCCGCAGTACTGGACGCGGTGCTTG	GACTAC		IPSVRATIPDLIV
					GGGATAGCAGGAAGAAGCTGCGGGAATGT	AGCCCT		RRFGGLDSSPWS
					CACGTGGCGGTGCTAGACTTCGCCAAGGCA	CGTAAG		VARAAAKSDKI
					TTTGACACAGTGTCTCACGAGGCACTTGTCG	TGCCGC		RKKLRWAWKQ
					AATTGCTGAGGTTGAGGGGCATGCCCGAAC	CTTAGA		LRRFSRVDSTTQ
					AGTTCTGCGGCTACATTGCTCACTTATACGA	ACGCAA		RPSVRLFWREH
					TACGGCGTCCACCACCTTAGCCGTGAACAA	CGGGGC		LHASVDGRELR
					TGAAATGAGCAGCCCTGTGAAAGTGGGACG	AATAGG		ESTRTPTSTKWI
					AGGGGTTCGTCAAGGGGACCCTCTGTCGCC	TGGGCC		RERCAQITGRDF
					GATACTCTTCAACGTGGTGATGGACCTCATC	GGGGCG		VQFVHTHINALP
					CTAGCTTCCCTGCCGGAGAGGGTCGGGTAT	CTAGCG		SRIRGSRGRRGG

					AGGTTGGAGATGGAACTTGTGTCCGCTCTG GCCTATGCTGACGACCTAGTCCTGCTTGCGG GGTCGAAGGTAGGGATGCAGGAGTCCATCT CTGCTGTGGACTGTGTTGGTAAGCAGATGG GCCTACGCCTGAATTGCAGGAAGAGCGCGG TTCTGTCTATGATACCGGATGGCCACCGCA AGAAGCATCACTACCTGACTGAGCGAACCT TCAATATTGGAGGTAAGCCGCTCAGGCAGG TGAGTTGTGTTGAGCGGTGGCGATATCTTG GTGTCGATTTTGAGGCCTCTGGATGCGTGAC ATTAGAGCATAGTATCAGTAGTGCTCTGAA TAACATCTCAAGGGCACCTCTCAAACCCCA ACAGAGGTTGGAGATTTTGAGAGCTCATCT GATTCCGAGATTCCAGCACGGTTTTGTGCTT GGAAACATCTCGGATGACCGATTGAGAATG CTCGATGTCCAAATCCGGAAAGCAGTCGGA CAGTGGCTAAGGCTACCGGCGGATGTGCCC AAGGCATACTATCACGCCGCAGTTCAGGAC GGCGGCTTAGCGATCCCATCGGTGCGAGCG ACCATCCCGGACCTCATTGTGAGGCGTTTCG GGGGGCTCGACTCGTCACCATGGTCAGTGG CAAGAGCCGCCGCCAAATCTGATAAGATTC GTAAGAAACTGCGGTGGGCCTGGAAACAGC	GGGGGG AGTAAT CTCCCC TGTTGG CGTGCA CCGCAC TGCTCC CACTGG GGGCAG TGTCAT CCGGAA ACAGGT GGGCCG GGGCGC CACCAG GGGGGA GCAATC CCTCCT G (SEQ ID NO: 1145)		GESSLTCRAGCK VRETTAHILQQC HRTHGGRILRH NKIVSFVAKAM EENKWTVELEP RLRTSVGLRKPD IIASRDGVGVIV DVQVVSGQRSL DELHREKRNKY GNHGELVELVA GRLGLPKAECV RATSCTISWRGV WSLTSYKELRSII GLREPTLQIVPIL ALRGSHMNWTR FNQMTSVMGGG VG (SEQ ID NO: 1390)

					TCCGCAGGTTCAGCCGTGTTGACTCCACAA CGCAACGACCATCTGTGCGCTTGTTTTGGCG AGAACATCTGCACGCATCTGTTGATGGACG CGAACTTCGCGAATCCACACGCACCCCGAC ATCCACAAAGTGGATTAGGGAGCGATGCGC GCAGATAACCGGACGGGACTTCGTGCAGTT CGTGCACACTCATATCAACGCCCTCCCATCC CGCATTCGCGGATCGAGAGGGCGTAGAGGT GGGGGTGAGTCTTCGTTGACCTGCCGTGCT GGTTGCAAGGTTAGGGAGACGACGGCTCAC ATCCTACAACAGTGTCACAGAACACACGGC GGCCGGATTCTACGACACAACAAGATTGTA TCTTTCGTGGCGAAAGCCATGGAAGAGAAC AAGTGGACGGTTGAGCTGGAGCCGAGGCTA CGAACATCGGTTGGTCTCCGTAAGCCGGAT ATTATCGCCTCCAGGGATGGTGTCGGAGTG ATCGTGGACGTGCAGGTGGTCTCGGGCCAG CGATCGCTTGACGAGCTTCACCGTGAGAAA CGTAATAAATACGGGAATCACGGGGAGCTG GTTGAGTTGGTCGCAGGTAGACTAGGACTT CCGAAAGCTGAGTGCGTGCGAGCCACTTCG TGCACGATATCTTGGAGGGGAGTATGGAGC CTGACTTCTTATAAGGAGTTAAGGTCCATA			

					ATCGGGCTTCGGGAACCGACACTACAAATC GTTCCGATACTGGCGTTGAGAGGTTCACAC ATGAACTGGACCAGGTTCAATCAGATGACG TCCGTCATGGGGGGCGGCGTTGGTTGAGCC TTGCACAGTAGTCCAGCGGTAAGGGTGTAG ATCAGGCCCGTCTGTTTCTTCCCCGGAGCTC GCTCCCTTGGCTTCCCTTATATTTAACATCA GAAACAGACATTAAACATCTACTGATCCAA TTTCGCCGGCGTACGGCCACGATCGGGAGG GTGGGAATCTCGGGGATCTTCCGATCCTAA TCCATGATGATTACGACCTGAGTCACTAAA GACGATGGCATGATGATCCGGCGATGAAAA (SEQ ID NO: 1544)			
R2	R8Hm-A	.		Hydra vulgaris	TTCAAGTGGATGAAGCTGGGAAGGTAATCT GTAGTTGGTTGAGTTGGTTGCAGATTACTGC TGTCGATTTTGCTTTCTATTGAAAGCCTGTC TCTACGGGTCCTGAAGCTTGAATTTTGGTAG CTATAGTTTTGTGGGAGGAAAGTGGAATTT TGTACCATCTTTTGTCTCTCGTATCTACTAT AGTAAATCCGGTCATGCAGCCTCTACGCGG CGCAACTAGAAACTTGGATCAGTGATCAAG GCTAATGCATGCCGGGTCTCCTCAGATTAG GAGTATAATACAAATCTGACTTCATCACTA	TTCAAG TGGATG AAGCTG GGAAGG TAATCT GTAGTT GGTTGA GTTGGT TGCAGA TTACTG	TAAA TGCC AAAA GTTG CTTG GGCT AAAT GATA CGTA CGCT	MNLLIVTSSIKES DVPSSGKGGVA VNNITAGASGK DTCVIIHPGTDGI WCCTECVEIHNS GKDLKRHLAKR HPSVTISGYKCN LCPFVSERQLSV GTHLRYCRGVK EVVKREFACAS

					AGAGGCTATGGGGCTAACGATCCTATAGTC	CTGTCG	AGAA	CSFSSDTFSGLQ
					TCGATGAACCTATTGATTGTTACTAGTAGCA	ATTTTG	AAAG	VHMQRKHIAEW
					TAAAAGAAAGTGACGTACCCTCTAGTGGAA	CTTTCT	CGAC	NDQLKEKTEFA
					AGGGGGGTGTAGCAGTCAATAACATAACAG	ATTGAA	TTGCT	WTDRELRELAE
					CAGGAGCTAGTGGAAAAGATACGTGCGTGA	AGCCTG	GCAC	KELTTPSFRYNK
					TCATACACCCAGGTACCGATGGTATTTGGT	TCTCTA	GGAT	IFYAALGTSRTY
					GCTGTACTGAGTGTGTAGAGATACATAACA	CGGGTC	GACG	DAVRKIRYNDR
					GCGGTAAGGATCTGAAACGACATCTTGCAA	CTGAAG	GTTC	YKSAIAEMRSQI
					AACGTCACCCGAGTGTAACGATAAGCGGTT	CTTGAA	ATCA	ADAAAAAQERD
					ACAAATGCAATCTGTGTCCATTTGTTAGTGA	TTTTGG	GAGC	VERGLVSAHSD
					ACGCCAACTAAGTGTGGGGACACATCTGAG	TAGCTA	CCGA	RGKEMLPVVET
					GTACTGCAGAGGCGTAAAAGAAGTGGTTAA	TAGTTT	TATG	KSDIQVNNDIKK
					AAGAGAGTTTGCATGCGCGAGCTGCTCTTTT	TGTGGG	TGCA	DIELTPNSRQKQ
					TCTTCGGATACGTTCTCAGGACTTCAGGTGC	AGGAAA	TGTC	TNLALARPAVIE
					ATATGCAAAGAAAGCATATAGCAGAATGGA	GTGGAA	AAGG	VEEDLGRQDVK
					ACGACCAGCTGAAGGAGAAAACGGAGTTTG	TTTTGT	CGGC	QYLASLRQDDY
					CTTGGACAGACCGAGAATTGAGGGAGCTTG	ACCATC	AGGG	TSPAERSIFAYC
					CTGAGAAGGAACTTACCACTCCTTCCTTCAG	TTTTGT	AGAA	REETNWSATKR
					GTACAACAAAATTTTCTATGCTGCGCTAGGT	CTCTCG	TCAC	QVLKISRTTRGL
					ACCTCCCGGACCTACGACGCTGTGAGGAAA	TATCTA	TAGT	RQPKKVRPFEFP
					ATTCGCTATAATGACAGATACAAATCTGCC	CTATAG	GTAG	EGFKPNRNMRK
					ATTGCTGAAATGCGATCACAGATAGCAGAT	TAAATC	CTGTT	WRKYRFLQECY
					GCGGCTGCCGCTGCACAAGAGAGGGATGTA	CGGTCA	CTTTC	REKRAETVSKIL

					GAGCGGGGTTTAGTTTCAGCACACTCAGAC	TGCAGC	CATT	DGTFIDEPEEEIR
					AGAGGAAAAGAAATGCTCCCTGTTGTTGAA	CTCTAC	ACGA	PELEEVQRMYID
					ACCAAAAGTGATATCCAAGTAAACAACGAT	GCGGCG	CTTA	RLEKRTQLDTT
					ATCAAAAAGGATATTGAATTAACACCGAAT	CAACTA	CGCG	KIVQTDEVFCLQ
					TCAAGACAGAAACAAACTAATCTAGCGCTG	GAAACT	GTTA	SYGRITIGEVRD
					GCAAGGCCAGCTGTAATTGAGGTGGAGGAA	TGGATC	ACGT	ALGASKKDSAS
					GACTTGGGTAGGCAGGATGTGAAACAATAT	AGTGAT	GGCA	GPDGLLLQDVR
					CTCGCATCCCTGCGCCAAGACGACTACACA	CAAGGC	CGAT	RLGPLLLCNIFN
					AGTCCGGCCGAGCGGTCAATCTTTGCATAC	TAATGC	AGAT	MWYLHGIPVEE
					TGCAGGGAGGAAACCAATTGGTCTGCGACA	ATGCCG	TTAC	NRCRTILLYKSG
					AAAAGACAGGTATTAAAGATATCGAGAACT	GGTCTC	ACCA	DRHLASNYRPV
					ACCAGAGGTTTAAGACAACCTAAGAAGGTT	CTCAGA	GGAA	TIGNMLNRLYA
					CGTCCATTTGAGTTTCCGGAAGGGTTCAAA	TTAGGA	ATAA	KIWDKRIRKNV
					CCTAACAGAAATATGAGAAAGTGGAGAAA	GTATAA	TACG	RLHVRQKAFIPV
					GTATAGATTCCTTCAGGAATGCTATAGGGA	TACAAA	TGAA	DGCFENVKTIQC
					AAAGAGAGCTGAGACTGTTAGCAAGATCCT	TCTGAC	GGGT	VLQSYRKRKLE
					GGACGGGACTTTTATCGATGAACCGGAGGA	TTCATC	TCCA	HNVVFIDLAKAF
					AGAGATTAGACCAGAGTTAGAGGAAGTACA	ACTAAG	CCAT	DTVLHDSIRKAL
					ACGTATGTACATTGACCGGCTGGAGAAAAG	AGGCTA	ATAC	WRKGVPSGVVK
					AACTCAGCTGGATACCACGAAGATTGTGCA	TGGGGC	TGGA	VVDSLYAGAVT
					AACAGACGAGGTGTTTTGTCTGCAAAGCTA	TAACGA	GTTT	SISVGKTKTRSIC
					CGGTCGCATTACGATCGGGGAAGTAAGAGA	TCCTAT	AGAT	INSGVKQGCPLS
					TGCACTCGGTGCAAGCAAGAAGGACTCGGC	AGTCTC	CTAT	PLLFNLILDELA

					CTCGGGTCCTGACGGCCTGCTTCTACAGGAT	G (SEQ	GAGG	ERIEATGCGLDL
					GTGAGGAGGCTGGGACCACTATTATTGTGT	ID NO:	GAAA	DGHVLSSMAFA
					AACATCTTTAACATGTGGTACTTACATGGG	1146)	CATTT	DDYVLLAKDSV
					ATCCCTGTGGAAGAAAACAGGTGTCGAACA		GTAA	EMNELIRVCSTF
					ATACTCTTATACAAGAGTGGCGATAGACAT		TAAG	FKEKGLSVNPG
					CTGGCATCAAACTATAGACCTGTGACAATC		TCAG	KCQSLRVLPVK
					GGCAACATGCTGAACAGGCTTTACGCCAAA		TCTG	EKKRSMKVLVR
					ATCTGGGACAAACGGATCCGGAAGAACGTG		GTAA	PHRWWRIKDQD
					CGTCTTCATGTGAGGCAAAAAGCATTTATC		CCTG	VDIPSMTYDSLG
					CCGGTGGATGGGTGCTTTGAGAACGTAAAA		GCGC	KYLGVSIDPTGK
					ACGATCCAATGCGTTCTCCAGTCTTACAGA		CGCT	IALPIEEWKNW
					AAGCGTAAGTTGGAACACAACGTCGTATTT		GTTG	MTKLKECKLKP
					ATTGATCTTGCCAAGGCCTTTGACACGGTCT		AGTC	EQKVKILKEVV
					TGCATGACTCGATAAGGAAAGCGTTGTGGC		AAAT	CSRVNYVLRMS
					GGAAAGGTGTTCCGTCTGGGGTTGTTAAAG		TAAC	ECGISELRSWTR
					TGGTAGACAGCTTATATGCGGGAGCTGTCA		TATG	FVRNWAKNIIHL
					CAAGCATAAGTGTTGGAAAAACGAAAACTC		TCAA	PTWCSSDWIHSI
					GTTCTATATGTATAAACTCTGGAGTCAAGC		TACT	KGLGIPDVSKGI
					AGGGTTGTCCTCTGTCACCTCTTCTATTCAA		CATT	VIQRMRASEKM
					CCTAATACTGGATGAACTAGCGGAGAGGAT		AAGT	STSEDGIVRVVG
					AGAGGCAACCGGCTGCGGGTTAGATCTTGA		TATC	ARLVQKNRVLW
					TGGTCACGTTCTATCATCTATGGCCTTTGCT		GACT	EKAGFEGIELKA
					GACGACTACGTGTTGCTAGCGAAGGACTCC		TTGA	ARRHCEVERLN

					GTGGAGATGAACGAGTTGATAAGAGTGTGT		TATG	NIGNITNGVALK
					AGTACATTCTTCAAAGAGAAAGGCTTATCT		GCAT	TIAAVSSVNRY
					GTAAACCCAGGTAAATGTCAATCGCTAAGA		GGGG	WMIEDNLKSGN
					GTTCTTCCCGTAAAGGAGAAGAAACGGTCA		TGAT	KILVWKAMAGA
					ATGAAGGTCCTTGTTAGACCTCATAGATGG		TCCG	IPTKINLSRGVA
					TGGAGGATAAAAGACCAGGATGTTGACATC		CGTT	DQTLKKCRRCG
					CCATCTATGACATATGACAGCTTAGGAAAA		ATAT	LTAETDGHILAG
					TACCTTGGTGTTTCGATTGACCCAACTGGTA		CAAA	CHTSSDAYSKR
					AGATAGCGCTTCCGATTGAGGAGTGGAAGA		GTCA	HNMLCDKLAKE
					ATTGGATGACCAAGCTAAAAGAGTGTAAGC		AACA	LKLNGGPNRRV
					TCAAGCCCGAGCAGAAAGTTAAAATTCTGA		TGAT	WRERTCFTSTG
					AAGAAGTGGTTTGCTCTCGGGTAAACTACG		GATT	RRYRPDIIVKDD
					TTTTGCGGATGTCAGAGTGTGGCATCAGCG		GCAA	SKITVIDMTCPY
					AACTTCGGAGTTGGACACGATTTGTAAGGA		TGAG	EKSEGHLIQCES
					ATTGGGCGAAAAACATCATTCACTTACCCA		AAAC	AKVTKYEPLKL
					CATGGTGCAGTAGTGACTGGATACACTCGA		TACC	DKYWTRELEGA
					TCAAAGGGTTAGGCATTCCCGACGTTTCGA		ACGC	NGIVAEKVELM
					AGGGAATTGTCATACAACGTATGAGGGCTT		TTGG	GLAIGAIGTIMR
					CGGAGAAAATGTCTACGTCTGAAGACGGTA		TCAC	STLRKLCELKSG
					TAGTCCGCGTGGTTGGTGCACGACTTGTTCA		GTTT	RIVRRLQMIACN
					GAAGAACAGAGTCTTGTGGGAAAAGGCCG		GTGA	NSAQIIKGHLSR
					GTTTCGAAGGTATCGAACTGAAGGCAGCCA		GGAG	ATRRNLR (SEQ
					GGAGGCACTGCGAAGTGGAGAGACTCAAC		AACA	ID NO: 1391)

					AACATTGGTAACATTACCAACGGCGTTGCA		TCTC	
					CTCAAAACTATCGCAGCAGTCTCCTCGGTA		ATTC	
					AATCGGTACTGGATGATTGAAGACAACTTG		AAGC	
					AAATCCGGGAACAAGATTCTCGTTTGGAAA		CTCC	
					GCAATGGCGGGTGCCATTCCAACAAAGATT		CGGA	
					AACCTTTCGCGGGGCGTAGCAGACCAGACC		TGTC	
					CTCAAAAAATGTCGTCGATGCGGTTTAACA		GGCA	
					GCGGAAACGGATGGACACATCTTGGCTGGA		CCCG	
					TGCCATACTAGCAGCGACGCGTACTCAAAA		CTGA	
					CGTCACAACATGCTCTGTGATAAACTCGCC		CATC	
					AAAGAGCTCAAACTCAATGGTGGACCAAAC		TTCTG	
					AGACGTGTGTGGCGCGAGAGGACGTGCTTC		GCTT	
					ACTAGTACAGGCAGGCGATATAGACCTGAC		ATGA	
					ATTATCGTTAAAGATGACAGTAAAATCACA		AAAT	
					GTCATCGATATGACTTGTCCGTATGAGAAA		TTTCA	
					TCAGAAGGACACCTGATCCAATGTGAAAGT		TTAA	
					GCGAAAGTAACTAAATACGAGCCACTCAAG		TTTTT	
					CTAGATAAGTATTGGACTCGAGAACTCGAG		GTAA	
					GGAGCAAATGGTATTGTTGCTGAAAAGGTA		GTCA	
					GAGCTGATGGGATTGGCAATAGGGGCGATC		TGGG	
					GGCACAATCATGCGTAGTACCCTTCGGAAA		CGGC	
					CTCTGTGAGTTAAAGTCGGGCAGGATCGTA		TTGA	
					AGACGTCTACAAATGATTGCTTGTAATAAT		AAGC	

					AGCGCCCAAATTATAAAGGGTCACCTGTCA AGGGCGACTCGGAGGAATTTGCGGTGATAA ATGCCAAAAGTTGCTTGGGCTAAATGATAC GTACGCTAGAAAAAGCGACTTGCTGCACGG ATGACGGTTCATCAGAGCCCGATATGTGCA TGTCAAGGCGGCAGGGAGAATCACTAGTGT AGCTGTTCTTTCCATTACGACTTACGCGGTT AACGTGGCACGATAGATTTACACCAGGAAA TAATACGTGAAGGGTTCCACCATATACTGG AGTTTAGATCTATGAGGGAAACATTTGTAA TAAGTCAGTCTGGTAACCTGGCGCCGCTGTT GAGTCAAATTAACTATGTCAATACTCATTA AGTTATCGACTTTGATATGGCATGGGGTGA TTCCGCGTTATATCAAAGTCAAACATGATG ATTGCAATGAGAAACTACCACGCTTGGTCA CGTTTGTGAGGAGAACATCTCATTCAAGCC TCCCGGATGTCGGCACCCGCTGACATCTTCT GGCTTATGAAAATTTTCATTAATTTTTGTAA GTCATGGGCGGCTTGAAAGC (SEQ ID NO: 1545)		(SEQ ID NO: 1269)	
R2	R8Hm-B	.		Hydra vulgaris	CTTGGGGTCACTGACACATTTTTCGGTAGCC ATAGTTTTTTGAGAGGAAGAGTGGAAGTTT TTCCATGAGTCGTCTCTCGTATAAACTGTGG	CTTGGG GTCACT GACACA	ATGC CCGA GGTA	MSNRITIGDVPS VGKGGLTVNKQ TAGADGAEACV

					TAAATCCGGCCATCCAGCCTCTACGCGGCG	TTTTTC	GTTG	VIHPGAKGIWSS
					CAACTAGAAACTTGGATCAGTGATCAAGGC	GGTAGC	GGAT	PACLRKFTIGKE
					TAATGGATGACGGGACTCCATGGATAAGGA	CATAGT	AATG	LRAHLAQIHKL
					GATATAAAGATCTTATTTGAACGCATCTTAA	TTTTTG	ATGC	APSAVRYRCNK
					GGGGTTATGGGGCTAACACCCCCTTAATTCT	AGAGGA	ACAA	CPYEGDVQLSV
					GGTGCACATTTATTGACCGTTATGAGCAAT	AGAGTG	GCTC	GTHLRYCKGIA
					AGAATCACGATAGGTGATGTACCCTCGGTA	GAAGTT	GTAA	GVVEEKKQFAC
					GGAAAGGGGGGTTTAACTGTCAATAAACAA	TTTCCA	GGCG	AICNFSSDTFSG
					ACAGCAGGAGCTGATGGTGCTGAAGCGTGT	TGAGTC	ACTT	LQVHKQRKHVV
					GTAGTCATACACCCAGGTGCCAAGGGTATT	GTCTCT	GCTG	EWNEQLKEKTE
					TGGTCCTCTCCTGCGTGTTTAAGAAAGTTTA	CGTATA	CACG	FAWTDRELREL
					CGATCGGAAAAGAACTAAGGGCACATTTGG	AACTGT	TATG	AVKEVTIPFSVV
					CTCAAATTCATAAACTTGCACCGAGTGCAG	GGTAAA	CCGC	NTETFAVLDITT
					TTCGGTACAGGTGTAATAAGTGTCCGTATG	TCCGGC	TAAA	RTKDAVRKIRY
					AGGGTGATGTCCAACTCAGTGTGGGAACAC	CATCCA	CGCT	TDRYKSILAEVR
					ATCTGAGGTACTGTAAGGGTATTGCGGGAG	GCCTCT	TAGC	AQVNAVAEEAP
					TGGTGGAGGAGAAAAAGCAATTCGCTTGCG	ACGCGG	TCGA	QASDESQITLLV
					CGATTTGTAATTTCTCTTCGGATACCTTTTC	CGCAAC	TGAG	NTGRGAELQPA
					AGGACTTCAGGTGCATAAGCAAAGAAAGCA	TAGAAA	TGCA	VINITDSIELVTD
					TGTAGTTGAATGGAACGAGCAGCTGAAAGA	CTTGGA	TGTC	VNEVEMVTSNS
					GAAAACGGAGTTTGCTTGGACAGACAGGGA	TCAGTG	AAGA	TNEEQPINAPVE
					ACTGCGGGAGCTGGCGGTTAAGGAAGTAAC	ATCAAG	CGGT	PAVIEADLGRQD
					GATTCCTTTCTCTGTGGTGAATACGGAGACC	GCTAAT	CGGG	AKLYLASLRQS

					TTTGCTGTGCTAGATATTACGACGCGGACTA	GGATGA	AGTA	DCTNASDRWTL
					AGGATGCTGTGAGGAAAATTCGCTACACGG	CGGGAC	TGAT	AYCRGEVDWC
					ATAGATACAAATCTATCCTGGCTGAAGTAC	TCCATG	CAGT	KTKSRLFKVSRH
					GCGCACAAGTTAACGCTGTGGCGGAGGAAG	GATAAG	GGAG	ARGLRQPQRVE
					CGCCGCAAGCTAGTGATGAGAGTCAAATAA	GAGATA	CTGA	NWEFPEGFRPN
					CGCTCTTAGTTAACACAGGCAGGGGAGCAG	TAAAGA	CTTTC	RNLRKWRKYSF
					AATTACAACCTGCTGTGATTAATATAACTG	TCTTAT	CAGA	LQSCYRTKKKE
					ATTCAATTGAATTAGTTACTGATGTCAATGA	TTGAAC	CAAC	TVSKILDGTFKD
					GGTTGAAATGGTAACATCGAATTCAACCAA	GCATCT	TCAC	TPEEEIRPELEEV
					TGAAGAACAGCCTATCAACGCGCCGGTGGA	TAAGGG	GCGG	QRVYVDRLEVR
					ACCGGCTGTAATTGAGGCGGACTTGGGAAG	GTTATG	ATTC	TQLDTTRTVHID
					ACAGGATGCGAAACTATATCTCGCATCGCT	GGGCTA	GCGT	ERFDLVSYGRITI
					GCGTCAAAGCGATTGCACAAACGCATCTGA	ACACCC	GCGG	REVQDAISASKK
					TCGATGGACCCTTGCGTATTGCAGGGGAGA	CCTTAA	TGGA	DASGGPDGLLL
					AGTTGATTGGTGTAAGACGAAAAGCAGGCT	TTCTGG	TACA	QDVKKASPRQL
					TTTCAAAGTATCAAGACATGCCCGGGGTTT	TGCACA	ACAC	CIIFNMWYLHGI
					AAGACAACCTCAAAGGGTGGAGAATTGGG	TTTATT	CTGG	PVVENRCRTILL
					AGTTTCCAGAGGGATTCAGACCTAACAGGA	GACCGT	TATA	HKGGEKHLTSN
					ACCTTCGTAAATGGAGGAAGTATTCATTCTT	T (SEQ ID	ACAT	YRPVTIGNMLN
					GCAAAGTTGCTATAGAACGAAGAAGAAGG	NO: 1147)	ATGA	RVYAKIWDRRI
					AAACTGTTAGTAAGATTCTTGATGGTACTTT		AGGG	RKNLQLHVRQK
					CAAGGACACACCTGAGGAAGAGATTAGGCC		TTCC	AFVPLDGCFEN
					AGAGTTGGAGGAAGTACAACGTGTGTACGT		ATCT	VKTIQCILQSYR

					TGACCGGCTAGAGGTAAGAACTCAGCTGGA		AGTA	RSRREHNVVFV
					TACCACTAGGACAGTGCATATAGACGAAAG		CAGG	DLAKAFDTILHD
					ATTCGATTTAGTAAGCTATGGTCGCATTACG		GATA	SIEKALLRKGIPR
					ATCAGGGAGGTACAAGACGCAATCAGCGCA		ACGA	SVIKVVDSLYAG
					AGCAAGAAGGATGCCTCAGGGGGTCCCGAC		TCCA	AVTSITVGKTKT
					GGCTTGCTCCTACAGGACGTGAAAAAGGCG		TGGG	RPICINSGVKQG
					AGCCCACGCCAATTGTGTATCATCTTTAATA		AGCA	CPLSPLLFNLVID
					TGTGGTACTTGCATGGAATCCCTGTAGTGG		AACT	ELAERLEATGC
					AAAATAGGTGCCGAACAATACTCTTGCATA		AATT	GLDLEGHVISSM
					AGGGTGGCGAGAAGCATCTAACGTCGAACT		AGTT	AFADDYVLLAK
					ACCGACCTGTGACGATCGGCAATATGCTGA		GGAG	DSVEMNVLMN
					ATAGGGTATACGCTAAGATCTGGGACAGAC		GTAA	VCNTFFEEKGLA
					GGATCAGAAAAAACCTGCAACTTCATGTGA		TCCA	VNPAKCQSLRV
					GACAGAAAGCATTCGTCCCGCTGGATGGGT		ACGC	LPVKGKRSMKV
					GCTTTGAGAATGTAAAAACCATCCAATGCA		CGCT	LTRTHRWWKIN
					TTCTCCAGTCTTACAGAAGGAGCAGGCGGG		GTTG	NQDVEIPSMTYE
					AACACAATGTCGTATTTGTCGATCTTGCAAA		AGTC	SVGKYLGVMID
					AGCGTTTGATACGATTTTGCATGATTCGATA		AGTT	PAGKIALPIEEW
					GAGAAAGCATTGCTGAGGAAAGGCATACCG		TTTA	KLWLTRLRECK
					CGAAGTGTGATAAAAGTGGTAGACAGCTTA		ACCG	LKPDQKVKVLK
					TATGCGGGAGCTGTCACGAGCATTACGGTT		CCAG	EVVCARANYVL
					GGGAAAACAAAGACTCGACCTATATGTATA		TCAA	RMSGCGICELRK
					AATTCAGGGGTGAAGCAGGGTTGTCCTCTA		CTCTT	WSRFVRGWVKS

					TCTCCTTTGCTGTTCAATCTAGTAATAGATG		GTAG	IIHFPAWCNSEW
					AACTAGCGGAGAGGCTGGAGGCAACTGGCT		GTTA	MHSSKGLGIPDV
					GCGGTCTTGATCTGGAAGGTCACGTCATTTC		TCGG	VSGIVIQRMRAA
					TTCCATGGCTTTTGCTGATGACTACGTGTTG		TCTTC	EKMAKSTDGVV
					TTGGCGAAAGACTCGGTTGAAATGAACGTG		GGCA	RVVGARIVQTN
					CTAATGAACGTGTGCAATACGTTCTTTGAG		GACC	RVLWKRAGLAG
					GAGAAGGGTTTAGCTGTAAATCCAGCAAAA		TTGG	IELDAARKFCEV
					TGTCAGTCGTTACGCGTTTTGCCTGTAAAAG		ACCG	KRVNKIGNQTN
					GCAAACGGTCCATGAAAGTCCTTACGAGGA		CCTA	GGALKTIAESSV
					CGCATAGATGGTGGAAAATTAATAACCAGG		GCGC	SRHWLLEKNIRP
					ATGTTGAAATCCCATCTATGACATACGAAA		CGGC	GNKILVWKAMA
					GTGTTGGAAAATATCTTGGGGTAATGATTG		CAAC	GVIPTKINLSRG
					ACCCAGCTGGTAAGATTGCTCTTCCGATTGA		AGTT	VADQTLKKCRC
					GGAATGGAAGCTTTGGCTAACTAGGTTAAG		TGTC	CGLTAETDCHIL
					GGAGTGTAAGCTCAAACCTGATCAAAAAGT		GTCG	AGCPTSRDAYS
					GAAGGTGCTGAAAGAGGTAGTTTGTGCCCG		ACTA	KRHNLLCDKLA
					AGCAAACTATGTTCTCCGGATGTCCGGGTG		ACAT	KELRLNGGPSRR
					CGGAATCTGTGAGCTCCGTAAGTGGTCACG		GATG	VWRERMCLSGN
					ATTTGTGAGGGGATGGGTGAAATCCATCAT		ATTT	GRRYKPDIVVK
					TCACTTCCCCGCATGGTGCAATAGCGAATG		GCGA	DDGVITVIDMA
					GATGCATTCGAGCAAAGGCTTAGGCATTCC		GAGA	CPYEKSERHLSQ
					TGATGTAGTGTCAGGAATTGTCATCCAACG		AACC	CEDAKVAKYEP
					AATGAGAGCTGCGGAAAAAATGGCTAAGTC		CACG	LRLDRSWTQEL

					AACAGACGGAGTAGTCCGAGTTGTCGGGGC CCGCATTGTGCAGACAAATAGAGTTTTGTG GAAAAGGGCCGGATTAGCAGGCATAGAACT GGATGCCGCCAGGAAGTTCTGTGAGGTTAA GAGGGTGAACAAAATTGGCAATCAAACCAA TGGAGGCGCCCTCAAGACTATAGCAGAGTC CTCGGTGAGCCGGCACTGGTTATTGGAAAA GAATATAAGACCTGGAAACAAAATTCTAGT TTGGAAGGCAATGGCAGGAGTGATTCCAAC AAAGATCAATCTGTCTAGAGGCGTAGCCGA CCAGACTCTCAAAAAATGTCGGTGTTGTGG TTTAACAGCAGAAACTGATTGTCACATCTTG GCCGGATGTCCTACCAGTCGGGATGCGTAC TCGAAACGTCATAACTTGCTTTGTGATAAAC TCGCCAAAGAGCTAAGACTCAATGGTGGGC CAAGCAGACGGGTGTGGCGCGAGAGGATGT GTCTCTCTGGGAATGGCAGGCGTTATAAGC CCGATATTGTTGTGAAAGATGATGGTGTAA TTACTGTCATCGATATGGCATGTCCGTACGA GAAATCGGAAAGACACCTAAGTCAATGCGA AGATGCAAAAGTTGCTAAGTACGAGCCACT AAGGCTTGATAGGAGTTGGACTCAAGAACT TGAGGGGAATAACGGCAGAAGTGCTAATGA		CTTTG TCAC TTAT GTGA GGAT AAAA TCTCT TGTC CATA TGAT CCTTT GAAG GGAA CAGC GCTTT GAGC TTGCT CGGC GTTG GCAC CTTTA GTCT GTAA	EGNNGRSANEIS VVGIAVGAIGTI TRKTQRILSKLK LAKVGRPLQIIA CNESAQIIRRHL SGSRLRNLR (SEQ ID NO: 1392)

					AATATCAGTTGTAGGGATTGCAGTAGGGGC GATTGGAACAATTACGCGTAAAACCCAGCG GATACTTAGCAAGTTGAAACTGGCCAAGGT CGGAAGACCGTTACAAATAATTGCATGTAA TGAAAGCGCCCAAATTATAAGACGACATCT TTCGGGATCGAGACTTAGAAATTTGCGGTG AATGCCCGAGGTAGTTGGGATAATGATGCA CAAGCTCGTAAGGCGACTTGCTGCACGTAT GCCGCTAAACGCTTAGCTCGATGAGTGCAT GTCAAGACGGTCGGGAGTATGATCAGTGGA GCTGACTTTCCAGACAACTCACGCGGATTC GCGTGCGGTGGATACAACACCTGGTATAAC ATATGAAGGGTTCCATCTAGTACAGGGATA ACGATCCATGGGAGCAAACTAATTAGTTGG AGGTAATCCAACGCCGCTGTTGAGTCAGTT TTTAACCGCCAGTCAACTCTTGTAGGTTATC GGTCTTCGGCAGACCTTGGACCGCCTAGCG CCGGCCAACAGTTTGTCGTCGACTAACATG ATGATTTGCGAGAGAAACCCACGCTTTGTC ACTTATGTGAGGATAAAATCTCTTGTCCATA TGATCCTTTGAAGGGAACAGCGCTTTGAGC TTGCTCGGCGTTGGCACCTTTAGTCTGTAAT		TATTT TCTTG ATAT TATG GACG AAAA AGGT AGTA TGGT TGCA (SEQ ID NO: 1270)	

					ATTTTCTTGATATTATGGACGAAAAAGGTA GTATGGTTGCA (SEQ ID NO: 1546)			
R2	R9Av	GQ3980 57		Adineta vaga	GAAATAGTTTGCAATGGTAGGTGTATGGCG	GAAATA	ACTA	MNLPIREHAVSV
					CCTCTGTGTCTCTCTTTCGCTGGATATAGTT	GTTTGC	GTCT	HNINKFNYLCQL
					TGACGATTTTGTACCAGGTATCTGTTTCTTG	AATGGT	CCTTC	CSKSYDTINSVK
					TGAGTTCAGCACCAGTTTGAACAGGCTTAG	AGGTGT	TTCTA	AHYVACRRQKN
					CGATAGACCTTCGAACTTGAAACACTGTTG	ATGGCG	TTAG	ASSTTAVPTNVI
					TGAAGCTGGCTGGGCCCCTGCAGATTTTCTC	CCTCTG	TCAG	NNNQLAINTNQ
					GATTAGAACGTGAGTGTTACGTCCAGAATG	TGTCTC	TCTA	VISRNPLQCVEC
					ACCCACCAGTGGTTAGTTCTACGTTGCCCTG	TCTTTC	ATTA	LMKQVDFYAKD
					GAAAGGAGAAAAGTTGAGCTAAAATCGCA	GCTGGA	ATTTT	TKALVTHMRTK
					CGGCCTAGTTGTTTATCAAATAGGCACGGT	TATAGT	TCTTA	HAAAYEESKKV
					GAGGAACTCTTCTATGTACCCTGACTAAAG	TTGACG	CATT	ATRRVAWSPDE
					TACTCACTTGTGCGCTGGGTTTGCTCCCCCT	ATTTTG	CTAC	DQILAELEVKLK
					CGCATTGACTTATCTGATCGCACTACCCACC	TACCAG	ATCT	KIQKGQLLSRLV
					AAACGAAACATAAACTTAGCTCGTGGTATC	GTATCT	AGTT	VEYNKCADKSK
					AGTCCACAGCGTGTGCAGTCGGATTCAGGG	GTTTCT	CCAT	APSRSKDAIRTR
					GAGCGTGTTAGTGACAAGCAGGATAATATT	TGTGAG	TATT	RQQHDYKLLLR
					AACATAGTTAATGTTAAGGCGTTCAACATT	TTCAGC	AAAT	SLQSQQPPVGSE
					CCTTATCCAATTGGAAGAGTTGACTGTGAA	ACCAGT	TGGT	DSDSDISSSNNN
					GTTTGTCATGAAGACATTGGACAAATGAAT	TTGAAC	ATGA	PLTTTHNVTPTP
					TTGCCGATTCGAGAGCATGCCGTATCTGTAC	AGGCTT	TCAG	DSSNVVLLIQKI
					ACAATATAAACAAATTTAATTATTTATGCCA	AGCGAT	TGCT	RESVDSIVKITN

					GCTATGTTCTAAGTCTTATGATACTATTAAT	AGACCT	ATCT	LKLNTNMLNAA
					AGTGTTAAAGCTCACTATGTTGCATGCAGA	TCGAAC	CTGC	SAFINQNNNMD
					AGACAGAAGAATGCCTCATCCACAACAGCT	TTGAAA	TACA	PLELSMRGIEED
					GTTCCAACCAATGTCATCAACAACAACCAA	CACTGT	CTCA	VKAIRDKELQKP
					CTTGCTATAAATACTAATCAAGTAATATCA	TGTGAA	ATGC	TRNVPSSTTSRK
					AGAAATCCACTTCAGTGCGTTGAGTGTCTA	GCTGGC	TTAA	PTRNAKRLEKS
					ATGAAACAAGTTGATTTCTATGCTAAAGAT	TGGGCC	TCGT	KKYGYYQHLYY
					ACAAAGGCACTAGTCACGCACATGCGTACT	CCTGCA	ATGT	NNKKKLVAEIL
					AAACATGCTGCTGCCTACGAGGAATCAAAG	GATTTT	TATT	DGETSGAKPPP
					AAAGTCGCAACAAGAAGAGTTGCCTGGAGC	CTCGAT	GACA	MNLVEDYYRNI
					CCTGATGAGGATCAAATTCTTGCTGAACTA	TAGAAC	GTCT	WSRSTIDDSPVN
					GAAGTCAAATTGAAAAAGATACAAAAAGG	GTGAGT	GACA	NIKTVNSDSIFAP
					TCAATTACTTAGTCGTCTTGTCGTTGAATAT	GTTACG	CTTG	ISRDEIKLALSNT
					AATAAATGTGCTGATAAATCGAAAGCTCCT	TCCAGA	ATTA	KKDSAAGPDAV
					TCCAGGTCCAAGGATGCTATTCGTACAAGA	ATGACC	CTCTT	TIKEAKAIIDNL
					CGCCAACAACATGATTACAAACTATTGCTT	CACCAG	ACGA	YVAYNIWLGVQ
					CGCTCACTCCAATCTCAACAACCGCCAGTT	TGGTTA	CATA	GIPEQLKLNKTI
					GGTAGCGAAGACAGTGACAGTGACATATCT	GTTCTA	TGCA	LIPKGNSDLSLL
					TCTAGTAATAACAATCCTTTAACAACAACA	CGTTGC	CTGTT	KNWRPITISSIIL
					CATAATGTCACTCCAACGCCAGATTCATCC	CCTGGA	TGCTT	RVYNRLLAYRM
					AACGTTGTACTACTAATACAAAAGATCCGT	AAGGAG	CAGA	NKIFKTNDKQV
					GAATCTGTAGATTCCATTGTAAAAATAACG	AAAAGT	GAAA	GFKPVNGCGINI
					AACCTCAAATTGAACACGAATATGCTGAAC	TGAGCT	CCAC	SWLHSLLKHAR

					GCAGCAAGTGCGTTCATTAATCAAAATAAC	AAAATC	TGTTC	LNKNSIYACLVD
					AACATGGATCCACTTGAACTATCTATGCGT	GCACGG	ATAT	VSKAFDSVSHQS
					GGTATCGAAGAGGATGTGAAGGCAATTCGA	CCTAGT	AGTG	IVRALTMNGAPS
					GACAAAGAACTTCAGAAACCAACCAGGAA	TGTTTA	AAGT	LLVKLIMDQYT
					CGTTCCTTCTTCAACAACTTCGAGAAAGCCA	TCAAAT	TCCTC	NVNTVITCSGSIS
					ACTCGAAATGCCAAAAGGCTTGAGAAATCA	AGGCAC	AGTT	NKINISSGVKQG
					AAAAAATATGGCTATTATCAACATCTGTAC	GGTGAG	TTCTG	DPLSSLLFNLVI
					TATAATAACAAGAAAAAATTAGTAGCGGAA	GAACTC	TTGA	DELFDVIKDQY
					ATCCTCGATGGCGAAACAAGTGGTGCTAAG	TTCTAT	TATA	GYTIDNIGTTNA
					CCACCTCCAATGAACCTGGTTGAAGATTATT	GTACCC	TTCTT	RCFADDLTLISS
					ATAGAAATATTTGGTCACGTTCTACTATTGA	TGACTA	CTTTC	SRMGMNKLLEL
					TGATTCGCCTGTTAACAATATTAAAACCGTT	AAGTAC	ATTCT	TTKFFKERGLNV
					AATAGTGACTCTATATTTGCTCCAATTTCGC	TCACTT	CGCT	NPSKCMSIGMS
					GTGATGAAATCAAATTAGCATTATCAAATA	GTGCGC	TCTCC	KGYKGKKSKIES
					CGAAAAAGGATTCAGCAGCTGGACCTGACG	TGGGTT	TTTTC	EPLFSITDAQIPM
					CTGTAACAATAAAAGAAGCAAAAGCTATTA	TGCTCC	TACT	LGYIDKTTRYLG
					TTGACAATCTTTATGTTGCATATAATATATG	CCCTCG	GTGT	VNFTSIGAIDAK
					GCTAGGTGTTCAAGGAATTCCTGAACAACT	CATTGA	TCTTT	RIKKDLQDTLD
					GAAATTGAATAAAACTATCTTAATTCCAAA	CTTATC	TTATC	KLEHLKLKAQC
					AGGAAATTCCGATCTTAGTCTACTGAAAAA	TGATCG	AGTT	KMDLLRTYMIP
					CTGGCGACCTATTACAATCTCGTCTATTATC	CACTAC	TTTTG	RFMFQLIHTELY
					CTAAGAGTATACAACAGATTATTAGCATAC	CCACCA	TGGA	PKLLIKMDILIRK
					AGAATGAACAAGATCTTTAAAACTAATGAT	AACGAA	AAAA	LAKRILHLPISTS

					AAACAAGTTGGATTCAAACCTGTTAATGGT	ACATAA	TTGA	SEFFYLPFKEGG
					TGTGGTATTAATATATCTTGGCTTCACTCTC	ACTTAG	GAAT	LQLTSLKEAVGL
					TCTTGAAGCATGCACGCTTAAACAAAAATT	CTCGTG	AAAT	AKIKLHKKIMSS
					CAATATATGCTTGTCTTGTCGATGTGTCTAA	GTATCA	AAAG	NDPMLCYLIESQ
					AGCCTTTGATTCTGTGTCACATCAATCAATA	GTCCAC	T (SEQ	RSRIVEHFMKDL
					GTAAGAGCTCTCACAATGAATGGTGCACCA	AGCGTG	ID NO:	KLGDSLTLNEM
					TCCTTGCTAGTGAAATTAATAATGGATCAAT	TGCAGT	1271)	NNIKECFMKEK
					ATACGAATGTAAATACTGTCATCACATGTTC	CGGATT		RISFAQKIHGVG
					TGGTTCTATATCAAACAAGATAAATATCTCC	CAGGGG		FEVFSSSPLTNQ
					AGTGGTGTCAAGCAAGGTGACCCACTATCT	AGCGTG		WINGEIKTMTTK
					AGCTTGTTGTTCAATCTGGTTATAGATGAAC	TTAGTG		TYINSIKLRTNTL
					TGTTCGATGTAATAAAGGACCAATATGGTT	ACAAGC		ETRVTTSRGLNII
					ATACAATTGATAACATTGGCACCACCAATG	AGGATA		KTCRRCHVADE
					CACGATGCTTCGCCGATGATTTAACACTAAT	ATATTA		SLMHVLQCCSS
					ATCATCATCTAGAATGGGTATGAATAAATT	ACATAG		TKGLRYSRHHKI
					GCTTGAGCTCACCACGAAATTCTTCAAAGA	TTAATG		CAKVANKLVM
					ACGTGGACTAAATGTAAACCCATCAAAGTG	TTAAGG		NGYGVFREKSY
					CATGTCTATTGGCATGTCCAAAGGTTATAA	CGTTCA		PDPNNSGSYLRP
					AGGAAAGAAGAGTAAAATCGAATCTGAAC	ACATTC		DIIAVKNGHVIV
					CACTCTTCTCTATCACCGATGCTCAGATACC	CTTATC		LDVTVVYEVTG
					GATGTTGGGCTATATTGATAAGACAACTCG	CAATTG		ATFINAYQTKIN
					ATATCTCGGTGTAAATTTCACATCTATTGGT	GAAGAG		KYNAIMVQIEQ
					GCCATTGATGCAAAAAGAATCAAAAAAGAC	TTGACT		MFNCVNGELHG

					CTTCAGGACACACTCGATAAGCTTGAACAT CTTAAACTCAAAGCTCAGTGCAAAATGGAT CTCTTACGAACTTATATGATACCAAGATTCA TGTTTCAATTAATTCATACTGAGTTATATCC GAAATTGCTTATTAAAATGGACATCTTAATT AGGAAATTAGCTAAACGAATCCTACATCTG CCCATATCAACGAGTAGTGAATTCTTTTACT TACCCTTCAAAGAAGGAGGTCTTCAACTAA CCTCACTTAAAGAAGCAGTTGGTTTAGCCA AAATAAAATTACACAAGAAGATAATGTCCA GTAATGATCCAATGTTATGCTACTTGATTGA GAGCCAGAGGAGCCGTATTGTCGAACATTT TATGAAAGACCTTAAACTTGGAGATTCTTTA ACATTAAACGAAATGAATAACATCAAAGAG TGCTTCATGAAAGAAAAAAGAATCTCATTT GCTCAAAAAATTCACGGTGTCGGCTTCGAA GTATTCTCATCAAGTCCTTTGACGAACCAAT GGATTAATGGCGAAATTAAGACAATGACAA CTAAAACATACATTAACTCAATTAAACTTA GAACAAATACTCTAGAAACTCGGGTAACAA CATCTCGGGGACTGAACATCATAAAAACAT GTAGAAGATGCCACGTAGCTGACGAAAGTC TCATGCATGTGCTCCAATGTTGCTCTTCTAC	GTGAAG TTTGTC ATGAAG ACATTG GACAA (SEQ ID NO: 1148)		LVIGSRGSIHHS QLHIWHQMGFS SIELKYVAIGCM EDSLRIMSTFSK AIT (SEQ ID NO: 1393)

					CAAAGGTTTACGATACTCTCGTCATCACAA AATATGTGCCAAAGTAGCAAATAAATTGGT AATGAATGGTTATGGTGTATTTCGTGAGAA GAGTTATCCAGATCCAAACAACTCAGGTTC ATACCTTCGACCGGATATAATTGCAGTAAA AAATGGTCATGTTATTGTTCTTGATGTAACG GTTGTGTACGAAGTAACTGGTGCTACGTTTA TTAATGCCTACCAAACAAAAATAAATAAAT ATAATGCGATTATGGTACAAATCGAGCAAA TGTTCAATTGTGTTAATGGTGAATTGCATGG TCTAGTAATTGGATCACGTGGTTCAATTCAT CACAGTCAACTCCACATCTGGCATCAAATG GGATTCTCTTCCATAGAACTTAAATATGTGG CTATAGGATGCATGGAGGATTCGCTCAGAA TCATGTCCACATTCTCAAAAGCTATCACATG AACTAGTCTCCTTCTTCTATTAGTCAGTCTA ATTAATTTTTCTTACATTCTACATCTAGTTCC ATTATTAAATTGGTATGATCAGTGCTATCTC TGCTACACTCAATGCTTAATCGTATGTTATT GACAGTCTGACACTTGATTACTCTTACGACA TATGCACTGTTTGCTTCAGAGAAACCACTGT TCATATAGTGAAGTTCCTCAGTTTTCTGTTG ATATATTCTTCTTTCATTCTCGCTTCTCCTTT			

					TCTACTGTGTTCTTTTTATCAGTTTTTTGTGG AAAAATTGAGAATAAATAAAGT (SEQ ID NO: 1025)			
R2	R2Ol	LC34944 4		Oryzias latipes	CGCACAGGGGACACAGAGCCTGCCCAAGTA CCGCTCCCGAGGGAGCGGGAAACGGGGGG GTGACTATCCCCTGGGGTCCGGCGAGAGCG CTGGTCTACGGACCAGGGGTGGCTGTGGGC AGGCTGCTCCTCAGGCCAGTTGATTAGTTAC GCATGGGCTGTACCTCCACGTGGTCCCGCT GGTAACGACTTGTCGGCTAAATCAGCCCGC CCACCATCTGGGATATGGTTGACCGTCTAA CCCCAGTACTCAGGTCACAAACAAAATGGG AACAGATACAGTGTATGTCGGCCAGGACTA CCCTTCTGGCTTATCAAAACGGGTACCAGC ACGGTTAGTGGCGGGACCGATGCTGCGAGA GCGAAGCTGTCACGCCCATGTGTTTAGGGC TGGACACATGTGGAACTGGCGAACCAGCCT TCCGAGCGGGCGCTGGGACCAGCCCGCTTT GGAGAAGTCTCGGGTCCTAACCCGGTCGGT GGCGACGGCCACCGACCCCGAAATTACCTC TTACCCAGGAAAGTCCGTATCGACAAGTAC GCAGGTTCAGGAGGAGGACTGGTGTAGCCG GGAGAGCGGGTGGATCTCGCCAGGACTTGC	CGCACA GGGGAC GCTG CCTGCC CAAGTA CCGCTC ACAGGG TGAT GAGCGG GAAACG GGGGGG TGACTA TCCCCT GGGGTC TGGCGA GAGCGC TGGTCT ACGGAC CAGGGG TGGCTG TGGGCA	GGGG GACA GCTG GGAG TCTC GGCA TGAT TACA AATC TTGC GCTG CACT CGGA TGTC GTCC CCGT GACG GACA CATT AATC	MGTDTVYVGQ DYPSGLSKRVPA RLVAGPMLRER SCHAHVFRAGH MWNWRTSLPSG RWDQPALEKSR VLTRSVATATDP EITSYPGKSVSTS TQVQEEDWCSR ESGWISPGLAPE EPSVVSEITASM VATMRVATEEV VLEPQPEQVVTI LPEHGRNVPPGL AEQDTASPIEVS VLLPDLAENCPL CGVPSGGLRLL GKHFAVRHAGV PVTYECRKCAW RSPNSHSISCHV

					TCCTGAAGAACCCTCGGTGGTGTCCGAAAT	GGCTGC	CGGA	PKCRGRARMPS
					TACAGCCTCCATGGTAGCGACAATGAGGGT	TCCTCA	AAGC	GDPGIACDLCEA
					AGCAACCGAGGAGGTCGTGCTGGAACCACA	GGCCAG	GAGT	RFATEVGVAQH
					GCCTGAACAGGTCGTCACAATACTGCCGGA	TTGATT	GGTG	KRHVHPVEWNK
					GCATGGTCGAAACGTTCCTCCGGGGCTGGC	AGTTAC	ACTC	VRLERRGARGG
					AGAACAGGACACCGCCAGCCCCATAGAAGT	GCATGG	GCCT	GIKATKLWSVA
					CTCGGTGCTCCTCCCAGACCTCGCTGAGAA	GCTGTA	CAAG	EVETLIRLIREHG
					CTGCCCATTGTGTGGCGTGCCGAGCGGGGG	CCTCCA	(SEQ	DSGATYQLIADE
					CCTACGCTTGCTCGGGAAGCATTTTGCTGTC	CGTGGT	ID NO:	LGRGKTAEQVR
					CGACATGCGGGGGTGCCTGTAACGTATGAG	CCCGCT	1272)	SKKRLLRIDTAS
					TGCCGTAAGTGTGCGTGGCGGAGCCCCAAC	GGTAAC		NSPDDAEVEEER
					AGCCACTCAATCTCGTGTCACGTCCCCAAAT	GACTTG		LESLAVRSSSRS
					GCCGGGGGCGTGCGCGGATGCCCAGTGGCG	TCGGCT		PPSLVATRVREA
					ATCCAGGGATCGCCTGCGATCTCTGTGAAG	AAATCA		VARGESEGGEEI
					CCCGGTTTGCCACGGAGGTTGGGGTCGCCC	GCCCGC		RAIAALIRDVDQ
					AACACAAGCGGCACGTTCATCCGGTGGAGT	CCACCA		NPCLIETSASDII
					GGAACAAGGTGAGGCTGGAAAGGAGAGGT	TCTGGG		SKLGRRVDGPK
					GCGCGCGGAGGGGGAATTAAGGCGACGAA	ATATGG		RPRPVVREQTQE
					GCTCTGGAGTGTAGCGGAGGTAGAGACGCT	TTGACC		KGWVRRLARRK
					AATCCGGCTCATCCGTGAGCACGGAGATTC	GTCTAA		REYREAQYLYS
					AGGTGCCACTTACCAGCTCATTGCCGATGA	CCCCAG		RDQARLAAQIL
					GCTGGGAAGGGGCAAGACGGCCGAACAGG	TACTCA		DGAASQECALP
					TGAGGAGTAAAAAGAGGCTCCTGCGCATAG	GGTCAC		VDQVYGAFREK

					ATACGGCAAGCAATAGCCCAGATGATGCAG AGGTTGAGGAGGAGAGGTTGGAATCTCTGG CGGTTCGGTCCTCGTCACGGTCACCCCCGA GCCTGGTGGCGACCAGGGTCAGGGAGGCAG TTGCCAGGGGTGAATCAGAAGGTGGCGAGG AGATCAGGGCTATTGCTGCTCTCATTAGGG ACGTAGATCAGAATCCTTGTCTGATTGAAA CCTCGGCGTCGGACATCATCTCGAAGCTGG GAAGGAGGGTGGATGGGCCCAAGAGACCC AGGCCCGTTGTCAGAGAACAGACCCAAGAG AAGGGATGGGTAAGGCGGCTTGCCCGGCGG AAAAGGGAGTACAGAGAAGCGCAGTACCT GTACTCAAGGGATCAAGCAAGGCTGGCGGC CCAGATCCTCGATGGTGCCGCCAGCCAGGA ATGCGCCCTCCCGGTGGACCAGGTCTACGG AGCGTTCCGTGAGAAATGGGAAACCGTAGG GCAGTTCCACGGACTTGGTGAGTTCCGGAC GGGTGCACGCGCAGACAACTGGGAGTTCTA CTCTCCAATTCTGGCGGCTGAGGTGAAAGA AAACCTAATGAGAATGGCTAACGGCACGGC CCCGGGACCAGACAGGATAAGCAAAAAGG CTCTGCTTGACTGGGACCCCCGGGGTGAGC AACTGGCACGGCTGTACACGACGTGGCTGA	AAACAA A (SEQ ID NO: 1149)		WETVGQFHGLG EFRTGARADNW EFYSPILAAEVK ENLMRMANGT APGPDRISKKAL LDWDPRGEQLA RLYTTWLIGGVI PRVFKECRTKLL PKSSDPVELQDI GGWRPVTIGSM VTRLFSRILTMR LTRACPINPRQR GFLASSSGCAEN LLIFDEIVRRSRR DGGPLAVVFVD FARAFDSISHEHI LCVLEEGGLDR HVIGLIRNSYVD CVTRVGCVEGM TPPIQMKVGVK QGDPMSPLLFNL AMDPLIHKLETA GTGLKWGDLSI

					TCGGTGGGGTCATACCAAGGGTCTTCAAGG			ATLAFADDLVL
					AGTGCAGGACTAAGCTGCTACCGAAATCCA			VSDSEEGMGRS
					GCGACCCGGTCGAGTTGCAGGACATCGGTG			LGILEKFCQLTG
					GATGGAGGCCGGTGACGATTGGGTCGATGG			LRVQPRKCHGF
					TGACTAGGCTGTTCAGTCGGATTCTAACGAT			FMDKGVVNGC
					GAGGCTAACCCGAGCCTGTCCGATCAATCC			GTWEICGSPIHM
					GAGGCAGCGCGGTTTCTTGGCCTCCTCGAG			IPPGESVRYLGV
					TGGATGCGCGGAAAACCTGTTGATCTTTGA			QVGPGRGVMEP
					CGAGATCGTCAGGCGCTCGAGGCGGGACGG			DLIPTVHTWIERI
					GGGGCCGCTGGCAGTGGTGTTTGTGGACTT			SEAPLKPSQRMR
					TGCGAGGGCCTTTGACTCCATCTCACATGA			VLNSFALPRIIYQ
					ACATATCCTGTGTGTTCTCGAAGAAGGCGG			ADLGKVTVTKL
					GCTTGACAGGCACGTTATCGGGTTGATCCG			AQIDGIVRKAVK
					AAACTCGTACGTGGATTGCGTGACCAGGGT			KWLHLSPSTCN
					GGGTTGTGTCGAGGGCATGACACCACCAAT			GLLYSRNRDGG
					ACAAATGAAGGTTGGAGTGAAGCAGGGAG			LGLLKLERLIPS
					ACCCCATGTCCCCCTTGCTCTTCAACCTGGC			VRTKRIYRMSRS
					TATGGATCCCCTCATCCATAAACTCGAGAC			PDIWTRRMTSHS
					GGCCGGAACTGGACTGAAATGGGGCGATCT			VSKSDWEMLW
					TTCAATCGCCACGCTGGCCTTTGCCGACGAT			VQAGGERGSAP
					CTGGTGCTGGTGAGTGACTCTGAGGAAGGC			VMGAVEAAPTD
					ATGGGGAGGAGTCTCGGGATTTTGGAGAAG			VERSPDYPDWR
					TTTTGCCAACTGACTGGGCTGAGGGTTCAG			REENLAWSALR

					CCCAGGAAGTGTCACGGTTTCTTTATGGAC			VQGVGADQFRG
					AAGGGCGTGGTGAACGGCTGTGGAACCTGG			DRTSSSWIAEPA
					GAAATCTGTGGGTCACCGATCCACATGATT			SVGFAQRHWLA
					CCCCCGGGGGAATCAGTTCGTTATTTGGGA			ALALRAGVYPT
					GTCCAGGTAGGCCCGGGGCGCGGCGTGATG			REFLARGKEKS
					GAACCGGATCTTATCCCTACGGTCCACACG			GAACRRCPARL
					TGGATCGAAAGGATCTCGGAGGCTCCTCTA			ESCSHILGQCPF
					AAGCCCTCACAACGCATGAGGGTTTTGAAC			VQANRIARHNK
					TCATTCGCTCTCCCCCGGATAATTTACCAGG			VCVLLATEAERF
					CCGATCTAGGGAAGGTTACGGTAACCAAAT			GWTVIREFRLED
					TGGCCCAGATAGATGGGATTGTCCGGAAGG			AAGGLKIPDLVC
					CTGTGAAGAAGTGGCTCCATTTGTCACCATC			KKADTVLIVDV
					CACGTGCAATGGACTGCTGTATTCACGGAA			TVRYEMDGETL
					CCGCGACGGTGGTTTGGGCCTCCTAAAGCT			KRAASEKVKHY
					GGAAAGACTAATCCCATCCGTGCGCACGAA			LPVGQQITDKV
					GCGTATCTATCGGATGTCCAGGTCTCCGGAT			GGRCFKVMGFP
					ATCTGGACACGGCGAATGACCAGCCATTCT			VGARGKWPASN
					GTGTCAAAATCTGACTGGGAGATGTTGTGG			NTVLAELGVPA
					GTCCAAGCGGGAGGTGAGAGGGGCAGTGC			GRMRTFARLVS
					ACCTGTAATGGGTGCCGTGGAGGCTGCCCC			RRTLLYSLDILR
					GACCGATGTGGAGAGATCGCCAGACTACCC			DFMREPAGRGT
					AGACTGGCGGCGTGAGGAAAACCTGGCATG			RVALIPAATGAA
					GTCGGCCCTGCGGGTGCAGGGTGTGGGTGC			

					AGACCAGTTTCGAGGCGACAGGACCAGCAG CTCTTGGATCGCCGAGCCCGCTTCGGTTGGG TTCGCGCAGCGCCACTGGTTGGCTGCCCTG GCGCTGAGGGCTGGGGTGTATCCGACTCGG GAGTTTCTGGCTCGGGGTAAGGAAAAGTCA GGAGCAGCTTGCAGACGCTGCCCGGCCAGG TTGGAATCATGTTCACACATACTTGGGCAAT GTCCGTTCGTTCAGGCGAACAGAATTGCGA GGCACAACAAGGTGTGTGTGCTCTTGGCCA CGGAGGCGGAGAGGTTCGGCTGGACGGTAA TAAGGGAGTTCCGTCTTGAGGACGCCGCTG GCGGTCTCAAGATACCCGACCTGGTTTGCA AGAAGGCCGACACAGTTCTCATTGTCGACG TGACCGTCCGGTACGAGATGGATGGAGAGA CGCTAAAAAGGGCCGCATCGGAGAAGGTG AAACACTATCTCCCAGTAGGGCAACAGATA ACGGACAAGGTCGGAGGGCGTTGCTTTAAA GTCATGGGGTTCCCTGTAGGTGCTAGGGGA AAGTGGCCGGCGAGCAACAACACAGTTTTG GCTGAGTTAGGCGTCCCTGCAGGTCGGATG AGGACCTTTGCCAGGCTGGTGAGCCGGAGG ACTCTTCTTTATTCTTTGGATATATTGAGGG ACTTCATGCGTGAGCCGGCCGGCAGGGGAA			N (SEQ ID NO: 1394)

					CTCGGGTTGCTCTCATCCCTGCGGCAACGG GTGCCGCGAATTGAGGGGGACAGCTGGGAG TCTCGGCATGATTACAAATCTTGCGCTGCAC TCGGATGTCGTCCCCGTGACGGACACATTA ATCCGGAAAGCGAGTGGTGACTCGCCTCAA G (SEQ ID NO: 1026)			
R2	R2_LP	AF015814		Limulus polyphemus	TGGGAGGAGACCCAAACTATCCTAGGATGG GGCGGAACCGACCATATGAGCCATATTAAC ATTGCCCACACTATCCTCTGGAGGTACCTCC TCGTGGTACGGCTGGATATAGGTAAATCCT GTAACCAAATCCTCCAACCCGTGAAGGAGA ACACTAAAACCCATATAGTGGCCTCGCCAA CCACTATATGTCCAACGGCAGGAGAAGCTA TCTCCCGGATGGGAAGGAAAACCCTAAACC GTGATGGGAACTTACCGGCCCCATCAGCTA TTGGGTACCCGGTAGGGACTTGCAACCCTA CCCTGTATTTGCATTTTATAGGGAACCGGTC GGCCCTATATCAGAGTAGACCGTTTATTAA ATATGGGTGAAAATATTAACAGTAAAAGCT ATGGTTTGGCGTCCGTGTGGTGCCAGGGCG GCGGCCAAACCCGAGCTACTTGGCACCAAC TGGGGATGGTAGCTTCCGAGCGATTCCCTG GCGACGTGGGACCGATCGACGATGGAGTCC	TGGGAG GAGACC CAAACT ATCCTA GGATGG GGCGGA ACCGAC CATATG AGCCAT ATTAAC ATTGCC CACACT ATCCTC TGGAGG TACCTC CTCGTG GTACGG	ATTTT GTCT CTTTC CCCA ATGA TGTCT ACTA GCAC GCTG CCGA AGCT AGAT AGAT TGAG GAAT CTGC GTAA	GIDGYMFGYAR ASGSTSVSIQSSS MTEGETNERAT PRASDSSSVSIQS SCVTEGECLPPT DNCNPSVENQL PCVTEGRFERVG SLVTVRLPFRKV ACDLCSKEFLTY SKFAVHQANFH NSETQACCTYC GKSDGNHHSIA CHVPKCPWRRT VTFAANLSNFLC DLCNDSFKTKS GLSQHKRHKHP CSRNAERILSLG

					AAACATCCGGAATAGAGGAATTGAGAAATA	CTGGAT	TCTGT	VRTPSARPRQV
					CCTATTCCACCACCGGCTCACATACCCAAG	ATAGGT	AATG	VWSEEETRTLRE
					GTGAACCCGGTGCAACTAGAGTACAACCTA	AAATCC	ATTA	VEVVYSGQKNI
					TCTGTGGCGGTAGGTGCCGAACCACTCAGG	TGTAAC	CGCC	NVLCAGHLPGK
					TGACGGGCTTGTTTATTGATGTCTCCCTACG	CAAATC	TCAT	TSKQVSDKRRD
					AGACACGAATTGTGACAAATCCACTCCGGT	CTCCAA	GGGC	LHRIRSSNVHGT
					GGACAATTACCCGATCTATGAACCTGTTAC	CCCGTG	ATCT	PTTQSRGDPVEQ
					CGATATTAGACAAGAAAATAAAGAACTGAC	AAGGAG	ATCG	VEEYEELDWEG
					AACGCCTAGAGCTTCAGGCAGCATGTCTGT	AACACT	GTAG	MHPFPDPDSKFC
					AAGTATCCAGTCATCGAGCGTGACTGAGGG	AAAACC	CGTC	SYLDQLRDQKG
					CGAAATTGATAATAACTCTGAAACTGAGGA	CATATA	GACC	LTEPVWQEIEIV
					ATTGACGGATATATGTTTGGCTACGCTAGA	GTGGCC	CTGA	AQEWVENLAHV
					GCTTCAGGCAGCACGTCTGTAAGCATCCAG	TCGCCA	CGTT	QSSWNHERTTK
					TCATCGAGCATGACTGAGGGCGAAACTAAC	ACCACT	AAAT	QVPENNTPARRP
					GAAAGGGCCACGCCTAGAGCTTCAGACAGC	ATATGT	TGGG	FKRRLHRVERY
					TCGTCTGTAAGCATCCAGTCATCGTGCGTGA	CCAACG	TAAT	KRFQRMYDLQR
					CTGAGGGTGAATGTCTACCTCCTACAGACA	GCAGGA	AAGA	KRLAEEILDGRE
					ACTGCAACCCGTCTGTAGAGAACCAGTTAC	GAAGCT	AATA	AVTCNLKKEEIK
					CGTGCGTAACTGAGGGTAGGTTTGAACGGG	ATCTCC	TCGA	DHYDQVYGVSN
					TAGGCTCACTGGTGACGGTGCGTCTGCCCTT	CGGATG	(SEQ	DRVSLDDCPRPP
					CAGAAAGGTGGCATGTGACTTGTGTTCTAA	GGAAGG	ID NO:	GANNTDLLKPF
					AGAGTTCTTGACATATTCGAAGTTTGCAGTC	AAAACC	1273)	TPTEVMDSLQG
					CACCAGGCAAACTTCCACAATTCAGAAACT	CTAAAC		MKNGAPGPDKI

					CAGGCATGCTGCACATATTGCGGTAAAAGT	CGTGAT		TLPFLQKRLKNG
					GATGGCAATCATCACTCTATAGCCTGTCAC	GGGAAC		IHVSLANVFNL
					GTTCCGAAATGTCCCTGGCGGCGAACTGTT	TTACCG		WQFSGRIPECM
					ACGTTTGCTGCGAACTTAAGCAATTTCTTGT	GCCCCA		KSNRSVLIPKGK
					GTGATCTTTGCAATGATAGTTTTAAGACCAA	TCAGCT		SNLRDVRNWRP
					ATCAGGGCTTTCGCAACATAAGCGTCATAA	ATTGGG		ITISSIVLRLYTRI
					GCATCCTTGTTCAAGGAATGCTGAACGCAT	TACCCG		LARRLERAVQIN
					CCTTTCTCTTGGAGTCAGGACGCCGTCGGCC	GTAGGG		PRQRGFVPQAG
					CGCCCTCGCCAGGTAGTGTGGTCCGAAGAA	ACTTGC		CRDNIFLLQSAM
					GAAACACGAACCCTCCGGGAAGTGGAAGTA	AACCCT		RRAKRKGTLAL
					GTGTATTCGGGCCAAAAGAACATTAATGTC	ACCCTG		GLLDLSKAFDT
					CTCTGTGCGGGGCATCTACCTGGTAAGACTT	TATTTG		VGHKHLLTSLE
					CCAAACAGGTCTCGGACAAGCGCCGAGACT	CATTTT		RFAVHPHFVRIV
					TGCACAGGATACGGTCTTCTAACGTACATG	ATAGGG		EDMYSGCSTSFR
					GTACACCCACCACTCAGAGTCGTGGAGATC	AACCGG		VGSQSTRPIVLM
					CTGTTGAACAGGTCGAGGAGTACGAGGAGT	TCGGCC		RGVKQGDPMSP
					TGGACTGGGAAGGAATGCATCCTTTTCCCG	CTATAT		ILFNIALDPLLRQ
					ACCCTGACTCTAAGTTTTGCTCGTACCTTGA	CAGAGT		LEEESRGFMFRE
					TCAGCTGAGAGATCAGAAGGGACTCACTGA	AGACCG		GQAPVSSLAYA
					ACCGGTATGGCAGGAGATCGAAATCGTGGC	TTTATT		DDMALLAKDH
					ACAAGAATGGGTAGAAAACCTTGCCCATGT	AAATAT		ASLQSMLGTVD
					TCAATCGTCTTGGAATCATGAGAGAACAAC	GGGTGA		KFCSGNGLGLNI
					CAAGCAGGTGCCAGAAAACAATACACCTGC	AAATAT		AKSAGLLIRGAN

					ACGAAGACCATTTAAAAGGCGTCTCCATCG	TAACAG		KTFTVNDCPSW
					TGTGGAACGTTATAAGCGGTTTCAGAGAAT	TAAAAG		LVNGETLPMIGP
					GTACGACCTCCAGCGAAAGCGCCTGGCTGA	CTATGG		EQTYRYLGASIC
					GGAAATACTAGACGGCCGGGAAGCCGTCAC	TTTGGC		PWTGINSGPVKP
					ATGTAACCTCAAAAAGGAGGAGATCAAAG	GTCCGT		TLEKWIANITES
					ACCACTATGATCAGGTCTACGGTGTGTCAA	GTGGTG		PLKPHQRVDILC
					ATGATAGAGTTTCTCTAGATGACTGCCCCA	CCAGGG		KYALPRLFYQLE
					GGCCACCAGGGGCCAATAACACCGACCTCC	CGGCGG		LGTLNFKELKEL
					TGAAACCGTTTACGCCAACCGAAGTGATGG	CCAAAC		DSMVKQAVKR
					ACTCACTTCAGGGTATGAAGAACGGGGCGC	CCGAGC		WCHLPACTADG
					CTGGCCCTGATAAGATTACCCTACCGTTCCT	TACTTG		LLYSRHRDGGL
					CCAAAAACGTCTTAAAAATGGCATCCATGT	GCACCA		AVVKLESLVPCL
					TTCCTTGGCAAATGTGTTTAACCTTTGGCAA	ACTGGG		KIKTNLRLVHST
					TTCTCGGGTCGCATCCCCGAATGCATGAAG	GATGGT		DPVISSLAESDG
					TCAAATAGGTCAGTCCTCATCCCGAAAGGG	AGCTTC		LVGAIEGIAQKA
					AAGAGCAATCTGCGGGATGTCAGAAACTGG	CGAGCG		GLPIPTPDQRSG
					CGGCCAATCACAATCTCCTCGATTGTGTTGC	ATTCCC		TYHSNWRDMER
					GGCTATACACCAGGATCTTGGCACGCCGTC	TGGCGA		RSWERLALHGQ
					TCGAGCGGGCGGTGCAGATTAATCCCCGAC	CGTGGG		GVELFKGSRSA
					AGCGAGGCTTCGTCCCTCAGGCTGGGTGTA	ACCGAT		NHWLPRPVGMK
					GGGATAATATATTCCTGCTTCAGTCTGCTAT	CGACGA		PHHWVKCLAM
					GAGGAGGGCTAAGCGAAAGGGAACTCTGG	TGGAGT		RANVYPTKRGL
					CTCTGGGGCTTCTTGACTTGTCGAAGGCATT	CCAAAC		SRGNLSKNKDS

					TGACACAGTTGGTCACAAACATCTTCTGAC	ATCCGG		AKCRGCTSMRE
					CAGCCTAGAAAGGTTCGCTGTCCACCCGCA	AATAGA		TLCHLSGQCPKL
					TTTCGTCCGAATTGTGGAGGACATGTACAG	GGAATT		KSMRIRRHNKIC
					TGGTTGTTCGACGTCCTTTCGAGTAGGCAGC	GAGAAA		EHLIAEASFKGW
					CAGTCTACTCGCCCCATCGTTCTGATGAGAG	TACCTA		KVLQEPTLVTD
					GCGTCAAACAAGGGGACCCCATGTCTCCTA	TTCCAC		NGERRRPDLIFH
					TATTGTTCAACATCGCTCTGGACCCTCTTCT	CACCGG		RDDKAVVVDVT
					TCGTCAACTGGAAGAGGAAAGCCGAGGCTT	CTCACA		VRYEISKDTLRE
					TATGTTTAGGGAGGGGCAGGCCCCTGTCTC	TACCCA		AYASKVRRYGC
					ATCTCTAGCATATGCCGATGATATGGCACT	AGGTGA		LTEQIKDLTGAT
					ACTGGCTAAAGATCACGCCAGTCTTCAGTC	ACCCGG		SVVFHGFPMGA
					GATGTTGGGCACTGTGGATAAATTTTGTTCA	TGCAAC		RGAWFPESSDV
					GGGAACGGACTTGGCCTTAACATCGCCAAA	TAGAGT		MADLNIRSKYFE
					AGTGCCGGACTTCTGATTAGGGGAGCGAAT	ACAACC		EFLCRRTILYTL
					AAGACCTTCACTGTCAATGACTGCCCTTCCT	TATCTG		DLLWKSNNEQY
					GGCTAGTAAATGGTGAAACGCTCCCGATGA	TGGCGG		LERLAP (SEQ ID
					TCGGTCCCGAACAAACTTACCGTTATCTTGG	TAGGTG		NO: 1395)
					GGCAAGCATCTGTCCGTGGACTGGGATAAA	CCGAAC		
					CAGCGGGCCTGTTAAACCCACCCTGGAGAA	CACTCA		
					ATGGATAGCCAATATCACAGAGTCTCCCCT	GGTGAC		
					CAAGCCACATCAGAGGGTCGACATACTCTG	GGGCTT		
					TAAGTACGCTTTACCCCGGCTGTTTTACCAA	GTTTAT		
					CTTGAGCTGGGCACTCTGAATTTCAAAGAA	TGATGT		

					CTGAAGGAACTAGACAGCATGGTCAAACAA	CTCCCT		
					GCTGTCAAACGTTGGTGCCATCTACCTGCCT	ACGAGA		
					GTACGGCTGACGGCCTGCTATACTCCCGTC	CACGAA		
					ATCGTGATGGGGGTTTAGCTGTAGTAAAAT	TTGTGA		
					TAGAGTCTCTTGTCCCTTGTCTAAAGATCAA	CAAATC		
					GACAAATCTCAGACTAGTGCATTCGACCGA	CACTCC		
					CCCCGTCATATCATCTTTGGCGGAATCCGAT	GGTGGA		
					GGTTTAGTGGGTGCCATCGAGGGTATTGCT	CAATTA		
					CAAAAGGCTGGGCTTCCGATCCCTACGCCT	CCCGAT		
					GACCAGCGATCTGGAACATATCATTCTAAT	CTATGA		
					TGGAGAGATATGGAAAGGAGAAGCTGGGA	ACCTGT		
					AAGGTTGGCCCTGCACGGGCAAGGTGTGGA	TACCGA		
					GCTCTTCAAAGGCTCAAGATCTGCCAACCA	TATTAG		
					CTGGTTGCCTAGGCCAGTTGGTATGAAGCC	ACAAGA		
					ACACCACTGGGTGAAGTGTCTGGCAATGAG	AAATAA		
					AGCTAATGTATACCCTACAAAAAGAGGCCT	AGAACT		
					CAGTAGAGGGAATCTATCTAAGAACAAAGA	GACAAC		
					TTCCGCCAAGTGTCGGGGATGCACATCAAT	GCCTAG		
					GAGGGAGACCCTATGTCATCTAAGTGGTCA	AGCTTC		
					ATGCCCGAAATTGAAGTCGATGAGAATAAG	AGGCAG		
					GCGCCACAATAAGATCTGTGAGCACTTGAT	CATGTC		
					CGCCGAGGCCAGCTTTAAAGGCTGGAAGGT	TGTAAG		
					TCTGCAAGAGCCTACCTTGGTTACAGACAA	TATCCA		

					TGGTGAACGTCGGCGACCTGATCTGATCTTC CATCGTGATGATAAAGCGGTGGTTGTTGAC GTGACGGTTCGCTACGAAATTTCGAAAGAC ACGTTGAGAGAAGCTTATGCTTCTAAAGTT CGAAGGTATGGATGTTTGACCGAACAAATT AAAGACCTTACAGGGGCTACCTCCGTTGTTT TTCATGGATTTCCAATGGGTGCCCGCGGTGC CTGGTTTCCTGAAAGCTCGGACGTGATGGC CGACCTGAACATTCGGTCAAAATATTTTGA AGAGTTCTTGTGTAGACGCACCATCCTATAT ACACTGGACTTATTATGGAAATCGAATAAC GAACAATATTTAGAAAGGCTTGCACCATAA ATTTTGTCTCTTTCCCCAATGATGTCTACTA GCACGCTGCCGAAGCTAGATAGATTGAGGA ATCTGCGTAATCTGTAATGATTACGCCTCAT GGGCATCTATCGGTAGCGTCGACCCTGACG TTAAATTGGGTAATAAGAAATATCGA (SEQ ID NO: 1027)	GTCATC GAGCGT GACTGA GGGCGA AATTGA TAATAA CTCTGA AACTGA (SEQ ID NO: 1150)		
NeSL	NeSL-1	Z82058		Caenorhabditis elegans	GCTCACTTTCTATCGTGTTAACCGTACGTTT ACACTCCCAGTGAGTGTAATAAAGGTTATT CGATAGAGGGTGTCTCCCTCTTTCTTGGGTA ATTCTTCGGCGGTCCGGGGTCTCTCCCTCGT CTTTTTTTTAAACTTTTCTTTCTCATCCACTC	GCTCAC TTTCTA TCGTGT TAACCG TACGTT	CCTC CAGG GCAC GCCG CACG	MLRRKGRHRM VMVNSVKWQPS AHAEAIGTGKS WAPQRSQASEH GWQSNAMFDPP

					TTTTGCTCCTTTTTACTAACTCTTGTACTCTA	TACACT	CCAA	NRILFARDSWSL
					TAGTCTTTTCTCATCCCCCATCCGCCGTTGG	CCCAGT	AAGT	NQSTHLQNQRS
					GCAAAGTTTATTTACTTTGTTAAATCCATAT	GAGTGT	CCTG	GSGLGIRPGQVR
					TTTATCTCTCTCACCCGTACAGAAAGCGTCT	AATAAA	GCAT	NNMVGGGPHR
					CCTTCTCAAACGCTTTTCTGTACTTTTTCTTA	GGTTAT	AACT	AGDPKRRVELV
					TATTTTCATTAACATATTTTTCCTGTTTATAC	TCGATA	CTGC	SIQGSEVTVRTI
					TAACCTAACCTCCATTGTCAATTACTAACTA	GAGGGT	AAAT	YPSDEIFSCYSKS
					ACTTGTACAACGGATTTCGATGTTGCGCCG	GTCTCC	AACA	CDIKTKAGYGPE
					AAAAGGACGTCACCGAATGGTTATGGTCAA	CTCTTT	TCAA	DLKHLTRHIKNE
					TTCTGTCAAATGGCAACCCAGTGCACATGC	CTTGGG	ACGT	HGLKARWAYQ
					TGAAGCAATTGGAACAGGAAAGTCCTGGGC	TAATTC	CAAT	CGLCNEKSDPSV
					ACCACAGCGGTCCCAGGCATCCGAACACGG	TTCGGC	CAAC	SEGHKWMEAH
					CTGGCAATCAAATGCAATGTTTGATCCCCCC	GGTCCG	TCCA	MVAVHQSSAEK
					AACAGGATTCTCTTCGCCAGAGACTCATGG	GGGTCT	CAAA	RIKSYQKCTGAR
					TCGCTCAACCAATCAACGCATCTTCAAAAT	CTCCCT	CTCTC	VAEQLQAAAPS
					CAAAGGAGCGGATCAGGATTGGGTATAAGA	CGTCTT	CACT	LTVPGKHKSGS
					CCTGGTCAGGTAAGGAACAATATGGTGGGG	TTTTTTA	CTCTT	RDAAKDSMTPT
					GGTGGGCCTCACAGAGCAGGGGACCCAAA	AACTTT	CAAG	KDDDPKTRIYQT
					GCGTCGTGTCGAGCTGGTCAGCATACAAGG	TCTTTCT	TCTTC	RSVVKKSTQKT
					AAGCGAAGTGACCGTCAGAACAATCTACCC	CATCCA	TCGG	AEPTDEGSRGPK
					GTCGGATGAAATATTCAGTTGTTACTCCAA	CTCTTTT	TGCTT	YASIFQKSVKAR
					ATCATGTGATATCAAAACAAAAGCTGGCTA	GCTCCT	CCAA	KSLALLCELSSP
					TGGCCCTGAGGACCTAAAGCACCTGACTCG	TTTTAC	CACC	KPMNPLPTNELT

					TCATATCAAGAACGAGCATGGTCTCAAAGC	TAACTC	ACAA	LKEGNSRELAK
					TCGCTGGGCATATCAATGTGGATTGTGCAA	TTGTAC	TGGT	EEAPSEGIDDIVI
					TGAGAAGTCGGACCCAAGTGTATCGGAAGG	TCTATA	GAAA	IDLDESEESPPRR
					CCACAAATGGATGGAGGCACACATGGTCGC	GTCTTT	GCTC	KRFNTWCLDHE
					CGTTCACCAAAGCTCTGCGGAAAAAAGGAT	TCTCAT	CTTC	SSREAWLDDTAI
					AAAGTCCTATCAGAAATGCACGGGTGCAAG	CCCCCA	ACCT	FWYISYLCRGST
					AGTTGCAGAACAGCTACAAGCTGCTGCTCC	TCCGCC	TTTCC	KYSALDPCLWS
					ATCGCTTACTGTGCCGGGGAAGCACAAATC	GTTGGG	CTCC	MYKVKGSRYIL
					AGGCTCTAGAGACGCTGCCAAAGATTCGAT	CAAAGT	AAAA	DRLESSITYFFPI
					GACACCAACAAAGGATGATGACCCGAAAA	TTATTT	TTCTT	CEEDHWTLLVL
					CCAGGATCTATCAGACACGAAGCGTAGTTA	ACTTTG	CCCA	KDNSYYYANSL
					AAAAGTCGACTCAGAAAACAGCAGAGCCA	TTAAAT	TGTG	HQEPRGPVRDFI
					ACAGATGAAGGGTCTAGAGGCCCAAAGTAC	CCATAT	GGGA	NDSKRARKEFK
					GCATCCATTTTTCAGAAATCCGTCAAAGCA	TTTATC	AGTC	VQVPLQRDSFN
					AGGAAGAGCTTGGCGCTTCTCTGTGAATTA	TCTCTC	CTGTT	CGVHICLMTNSI
					AGCAGCCCTAAGCCTATGAACCCCCTTCCT	ACCCGT	CTTGT	MAGGKWHSEE
					ACAAATGAGCTAACTCTGAAAGAAGGGAAT	ACAGAA	AAGC	DVRNFRKRLKK
					TCAAGAGAGCTCGCCAAAGAGGAAGCACC	AGCGTC	TCTCC	TLQEEGYELYSV
					ATCTGAAGGTATAGACGACATCGTCATCAT	TCCTTC	GGAG	NSLGIPFQAPTT
					CGATCTGGACGAATCGGAGGAGTCGCCACC	TCAAAC	GCTG	EQMDYKETRCK
					CAGAAGGAAACGATTCAACACCTGGTGTCT	GCTTTT	CAAG	RSYASVLTQISP
					GGATCATGAGTCAAGCCGTGAAGCATGGCT	CTGTAC	AGCA	PAKRPDCKPDN
					GGATGACACAGCAATCTTCTGGTACATCTC	TTTTTCT	GAAG	NIFVPTKDCAAE

					CTATCTCTGCAGAGGAAGTACAAAGTACTC	TATATT	AAAT	GNPQEKGRNES
					AGCTTTGGACCCATGCCTCTGGAGTATGTAC	TTCATT	TCTTC	PEEINTEHIVVA
					AAAGTCAAAGGCTCAAGATACATTCTTGAC	AACATA	TTTCT	GKPANNISPRCR
					CGCTTGGAAAGCTCCATCACATATTTTTTCC	TTTTTCC	GACA	STSEMLFEMVK
					CGATATGCGAGGAGGACCATTGGACACTGT	TGTTTA	AGGT	ATTSSGRSSLGT
					TGGTATTGAAAGACAATTCATACTATTATGC	TACTAA	CAGA	MTQDEFIRTSTI
					AAACAGTCTGCACCAAGAGCCACGTGGCCC	CCTAAC	AGGA	AEAVPLMSIKLP
					GGTCAGGGACTTCATCAACGACTCAAAACG	CTCCAT	AGTC	PMELPRKILPPIP
					GGCTCGGAAGGAGTTTAAGGTGCAAGTACC	TGTCAA	CTGTT	PRKPTQTNGGQ
					TCTTCAAAGAGACTCCTTTAACTGTGGAGTG	TTACTA	CTTG	KGKQQRVPTGK
					CACATCTGTCTAATGACCAACTCGATTATGG	ACTAAC	AGGC	PDTLNAKVRNW
					CAGGAGGCAAATGGCACTCTGAAGAAGAC	TTGTAC	GTCC	FNNQLESYAME
					GTCAGAAACTTCAGAAAAAGACTGAAGAA	AACGGA	ATCC	GRSFQRLEWLT
					GACACTCCAGGAAGAAGGCTATGAGCTTTA	TTTCG	CGGG	EVLTASIQKAAA
					CTCGGTCAATAGTCTGGGTATACCATTCCAA	(SEQ ID	CGTC	GDEGIVDIICKR
					GCCCCAACGACTGAGCAAATGGACTACAAA	NO: 1151)	ATAG	NPPLEVAKGEM
					GAAACAAGATGCAAAAGAAGCTATGCCAGT		GAGA	CTQTENKRKTT
					GTTCTTACTCAAATAAGCCCGCCGGCCAAA		GATC	NNAARIADPIQS
					AGGCCGGACTGCAAACCTGACAACAACATA		AGAT	SKGAGDVKASY
					TTCGTACCAACCAAGGATTGTGCTGCCGAA		GCAC	WKERARTYNRII
					GGTAACCCGCAGGAAAAAGGCCGAAATGA		CTTCT	GSKEELCKIPID
					ATCTCCTGAAGAGATCAATACGGAACATAT		AGCA	QLEDFFKKSTSR
					CGTCGTCGCAGGAAAACCTGCAAACAACAT		GGAG	TNVQESIMKEKS

					CAGTCCAAGGTGTCGGAGCACCTCGGAAAT		CTAG	SKIPALKIGNWM
					GCTGTTTGAGATGGTGAAAGCCACAACCAG		AAGG	EKKFIGKEVAFA
					CAGTGGAAGAAGCAGCTTGGGCACCATGAC		GCTG	LRKTKDTAQGA
					GCAGGATGAGTTCATCCGAACCAGCACAAT		CCCT	DGLRYHHLQWF
					CGCCGAGGCAGTTCCCCTAATGAGCATAAA		GTCTT	DPSGELLAKVY
					ACTCCCACCAATGGAGTTGCCAAGGAAAAT		GAGA	NECQRHRKIPKH
					TCTGCCACCAATTCCCCCCAGAAAACCAAC		TCCC	WKEAETILLFKN
					CCAAACCAATGGAGGTCAAAAGGGAAAGC		CACG	GDQSKPENWRPI
					AACAGAGGGTGCCTACAGGAAAACCAGAC		GGGG	SLMPVIYKLYSS
					ACCCTAAATGCTAAAGTCCGGAACTGGTTC		TCAA	LWNRRIRAVPN
					AACAACCAACTTGAGTCGTATGCGATGGAG		TAGA	VLSKCQRGFQE
					GGTCGCAGCTTCCAACGACTGGAATGGCTG		CGGG	REGCNESLAILR
					ACGGAAGTACTCACTGCGTCGATACAAAAA		AGGG	TAIDVAKGKRR
					GCAGCAGCAGGTGATGAAGGAATAGTTGAT		GCTG	NLAVAWLDLTN
					ATTATTTGCAAACGGAACCCGCCACTTGAA		CTGG	AFGSIPHELIEYA
					GTTGCGAAGGGTGAAATGTGCACCCAGACC		CTTTC	LTAYGFPQMVV
					GAAAACAAAAGGAAAACGACCAACAATGC		TCTTT	DVVKDMYQGA
					AGCAAGAATTGCGGACCCAATCCAGAGCAG		TTAA	SMRVKNATEKS
					CAAGGGAGCTGGTGATGTGAAGGCATCGTA		GAGG	DRIPIMSGVKQG
					CTGGAAAGAAAGGGCTCGCACTTACAACAG		AAGC	DPISPTLFNICLE
					GATTATTGGTAGCAAGGAGGAACTCTGCAA		ACCA	TVIRRHLESANG
					AATTCCCATCGATCAACTGGAGGATTTCTTC		ATCC	HQCLKTRIKVLA
					AAGAAATCCACGTCCCGCACCAACGTGCAG		GGAG	FADDMAILTDSP

					GAGTCGATCATGAAGGAGAAAAGCTCCAAA		ATCC	DQLQRELSKLD
					ATTCCTGCTCTCAAGATAGGTAACTGGATG		TTAG	NDCTPLNLIFKP
					GAGAAGAAGTTTATCGGAAAGGAGGTGGC		GGGT	AKCASLVIQKG
					GTTCGCTCTGCGGAAAACAAAAGACACCGC		CAAA	VVRSASIKLKGN
					GCAGGGTGCAGACGGACTGCGATACCACCA		GGAT	AIRCLDENTTYK
					CCTTCAATGGTTTGATCCCAGTGGTGAGTTA		TAAA	YLGVQTGSAARI
					TTGGCGAAGGTATATAACGAGTGCCAACGA		AGGC	SAMDLLEKVTK
					CACAGGAAGATCCCAAAACACTGGAAGGA		AGCA	ELECVVKSDLTP
					GGCCGAGACCATCTTGCTGTTCAAAAATGG		GGTC	PQKLDCLKTFTL
					AGATCAGTCAAAACCAGAAAACTGGCGCCC		CAAT	SKLTYMYGNSIP
					AATTAGCCTGATGCCTGTGATCTACAAACTT		TCTCC	LITEIKMFANIVI
					TACTCCAGTCTGTGGAACCGGAGAATTAGA		TCAC	RGVKVMHRIPV
					GCTGTACCAAATGTGTTGAGCAAATGTCAG		TGAC	RGSPLEYIHLPV
					CGAGGGTTCCAGGAGCGCGAAGGTTGCAAT		TTCG	KDGGLGVACPK
					GAGAGTCTAGCAATACTCAGAACAGCAATC		GTCA	TTCMITFLVSTL
					GACGTGGCCAAAGGAAAACGAAGAAACCT		GAGA	KKLWSDDEYIK
					GGCGGTGGCATGGCTGGATCTGACGAACGC		GGAG	TLFTSLAEEVVK
					GTTTGGATCCATCCCGCACGAATTGATTGA		TCCC	KESKKSTVTMD
					GTACGCGCTGACAGCGTATGGATTTCCGCA		GCCT	DIADYLNVEERI
					AATGGTCGTCGATGTGGTCAAAGATATGTA		TGGA	NRSEFGYNSITR
					CCAGGGAGCATCAATGAGGGTTAAGAACGC		GACC	LRDVMRNLAIT
					GACGGAAAAAAGCGATCGAATCCCAATAAT		TCCC	GDSPLYRLKMV
					GTCTGGGGTGAAACAAGGCGATCCCATTTC		CGGG	VKNGKIALLVQ

					ACCAACACTTTTCAATATATGCCTGGAAACT		GAGG	ATSESMERIYTE
					GTGATTAGAAGACACCTGGAGTCTGCAAAT		TTGCT	EDAKKLQRSLK
					GGTCACCAGTGCCTCAAAACAAGAATTAAG		GAAG	DQVNKALKHRF
					GTACTGGCGTTCGCCGACGACATGGCGATT		AGGC	NTTKVVKSKVV
					TTAACGGATTCCCCCGACCAGCTCCAGCGA		GGAA	RVVQQHPASNR
					GAACTGTCAAAGCTAGACAATGATTGCACG		GCTC	FVTKGGNLSLA
					CCCCTGAATCTTATTTTCAAGCCAGCAAAAT		CTTCT	CHRFVHKARLN
					GTGCATCACTTGTGATCCAAAAAGGAGTTG		AGCA	LLACNYNNYDK
					TGCGGAGCGCATCAATTAAGCTTAAAGGAA		AGAG	SKSKVCRRCGK
					ACGCCATTCGATGCCTTGACGAGAACACCA		CTAG	DLETQWHILQN
					CTTACAAATATTTGGGAGTTCAGACGGGTT		AGGG	CPFGFSKKITER
					CGGCAGCAAGAATTTCAGCAATGGATCTAC		AGTT	HDAVLHKVKTL
					TGGAGAAAGTCACGAAGGAACTTGAATGCG		CCCA	IESGGKKNWTM
					TGGTCAAAAGTGACCTGACGCCGCCGCAAA		GTCC	KIDEELPGFSRL
					AGCTGGACTGTCTTAAAACATTCACGCTGTC		TGAA	RPDICLKSPDEK
					CAAACTGACATACATGTATGGAAATTCCAT		ACCC	QIILADVACPYE
					ACCACTGATCACGGAGATAAAAATGTTTGC		TTGC	HGVEAMERSW
					AAATATCGTCATTCGAGGAGTCAAAGTGAT		GGTT	QAKIDKYETGF
					GCATAGAATCCCAGTCCGAGGGTCACCACT		GATG	AHLRKSGTKLT
					GGAGTACATCCATCTTCCAGTGAAGGATGG		ATGG	VLPIIIGSLGSW
					AGGGCTTGGTGTAGCATGTCCCAAGACAAC		AATG	WKPTGDSLKEL
					CTGCATGATTACGTTCCTTGTCTCTACTCTT		GAAG	GIKGSVINSAIPE
					AAAAAACTCTGGTCAGATGATGAATACATC		AGTA	LCATVLEHSKN

					AAAACATTATTCACATCACTGGCGGAAGAA GTAGTAAAGAAAGAGTCAAAGAAGAGCAC AGTCACTATGGATGATATAGCCGACTATCT CAACGTTGAGGAGAGGATCAATAGGAGCG AATTTGGGTACAATTCCATTACGAGACTGC GGGATGTGATGAGGAACTTGGCCATCACTG GCGACTCCCCACTTTACAGGCTGAAAATGG TAGTAAAGAACGGGAAAATCGCTTTGCTCG TCCAAGCCACAAGCGAAAGCATGGAAAGG ATCTACACGGAAGAAGATGCGAAAAAGCTG CAGCGCTCACTGAAGGATCAAGTGAACAAA GCACTCAAACATCGATTCAACACCACCAAA GTAGTGAAAAGCAAAGTCGTCCGAGTCGTG CAACAGCACCCAGCAAGCAACAGGTTTGTC ACAAAAGGTGGCAACCTGAGCCTTGCATGT CACCGCTTTGTGCATAAAGCACGTCTGAAT CTACTGGCCTGCAACTACAACAACTACGAC AAATCCAAATCAAAAGTCTGTAGGCGTTGT GGGAAGGATCTGGAGACGCAGTGGCACATA CTGCAAAACTGTCCGTTTGGTTTCTCAAAGA AGATCACTGAGAGGCATGATGCCGTCTTGC ACAAGGTCAAAACTCTCATTGAAAGCGGTG GAAAAAAGAATTGGACAATGAAGATTGATG		CTTC GGTA CTGC TCGTT GCTC TCTCT GCGT TTTAC TGCC GAGG GCCG GATT TGCT CGAA TCGC GAAA GGTC TCAA TCGA CCAT TCAA GATG ACGG	TYWNHIFGEAYI PNPMRNGHAKP AGNGWKKERL QKAPVRPTN (SEQ ID NO: 1396)

					AAGAACTTCCAGGATTCAGCAGACTCCGTC		CTTAT	
					CAGATATCTGCCTCAAAAGCCCTGATGAAA		CTAA	
					AACAAATCATCTTGGCAGATGTCGCATGCC		GGTC	
					CATATGAGCATGGAGTAGAAGCGATGGAAA		CGAA	
					GGAGCTGGCAGGCAAAAATCGACAAATAC		AGCA	
					GAGACGGGATTCGCCCACCTGCGGAAATCG		GTTG	
					GGAACCAAGCTGACCGTCCTTCCGATTATA		GGAG	
					ATCGGGTCACTTGGATCATGGTGGAAACCG		AGTA	
					ACAGGTGACAGTCTCAAGGAATTGGGAATC		ACGT	
					AAGGGAAGCGTGATCAACAGTGCCATTCCA		GTTCT	
					GAACTCTGTGCTACTGTTCTCGAACACAGTA		CCTA	
					AGAATACGTACTGGAATCACATCTTCGGTG		CCTTT	
					AAGCGTACATACCAAATCCAATGCGAAACG		CAAG	
					GACACGCAAAACCTGCTGGAAATGGATGGA		TTGA	
					AAAAGGAAAGATTGCAGAAGGCCCCTGTGA		ATGG	
					GGCCTACCAACTAGCCTCCAGGGCACGCCG		TCGTT	
					CACGCCAAAAGTCCTGGCATAACTCTGCAA		TTACT	
					ATAACATCAAACGTCAATCAACTCCACAAA		GTTT	
					CTCTCCACTCTCTTCAAGTCTTCTCGGTGCT		GGGA	
					TCCAACACCACAATGGTGAAAGCTCCTTCA		TAGC	
					CCTTTTCCCTCCAAAATTCTTCCCATGTGGG		TGAC	
					GAAGTCCTGTTCTTGTAAGCTCTCCGGAGGC		TTGA	
					TGCAAGAGCAGAAGAAATTCTTCTTTCTGA		TGCT	

					CAAGGTCAGAAGGAAGTCCTGTTCTTGAGG		AGTA	
					CGTCCATCCCGGGCGTCATAGGAGAGATCA		CGCT	
					GATGCACCTTCTAGCAGGAGCTAGAAGGGC		TCAT	
					TGCCCTGTCTTGAGATCCCCACGGGGGTCA		CTGT	
					ATAGACGGGAGGGGCTGCTGGCTTTCTCTTT		GGAT	
					TTAAGAGGAAGCACCAATCCGGAGATCCTT		GACG	
					AGGGGTCAAAGGATTAAAAGGCAGCAGGT		CTCC	
					CCAATTCTCCTCACTGACTTCGGTCAGAGAG		CCAA	
					GAGTCCCGCCTTGGAGACCTCCCCGGGGAG		GCAG	
					GTTGCTGAAGAGGCGGAAGCTCCTTCTAGC		TCAA	
					AAGAGCTAGAGGGAGTTCCCAGTCCTGAAA		GTAG	
					CCCTTGCGGTTGATGATGGAATGGAAGAGT		ACTT	
					ACTTCGGTACTGCTCGTTGCTCTCTCTGCGT		GAAA	
					TTTACTGCCGAGGGCCGGATTTGCTCGAATC		GGTG	
					GCGAAAGGTCTCAATCGACCATTCAAGATG		CCCT	
					ACGGCTTATCTAAGGTCCGAAAGCAGTTGG		CGCC	
					GAGAGTAACGTGTTCTCCTACCTTTCAAGTT		CTAG	
					GAATGGTCGTTTTACTGTTTGGGATAGCTGA		TTAG	
					CTTGATGCTAGTACGCTTCATCTGTGGATGA		CTCTT	
					CGCTCCCCAAGCAGTCAAGTAGACTTGAAA		AGAC	
					GGTGCCCTCGCCCTAGTTAGCTCTTAGACCT		CTTAT	
					TATGGGTCGCCATGGTTGTGGACGGGTATG		GGGT	
					CTTGCCGGAGCCGAGTCGTGTTTCTTAGAAC		CGCC	

					CAACCTCGACGAGGCGAAAGCTTGCACAAG TTAGCACAATTGTGGTAGGGCCGACTAGAA AATGAGTCCCTTAGGGGGTTACGCCTTGGC GAAAGTGAGGACAATTGGCATTGACGGGTG CTTCGGCACTAGGCAAAGGCGCCACCACAC TGTCCAATCTCTAAAAAGTTCACATTCATCG AAGAACTACCGGAACCAACCACACATGTGT TGAAACCTACACGGTGGAAGGGAAAGGAA AGCTTCGCTGGAACGAAAAGAACGGATAGG TTCCCCTTCTTGATGGCTGTGAGGCTTAGGA TGGACGGGAAGGCCGTGAGGCCTCAGGCGG GTAACTCGGCCAGACGCTAGTTGATCTTCG GATCACGACAGCCCTGGCTAAGAGGAACCC TGGATGGAGTGTGAAGGATGGGCGGGTAGG GGGTTAAGCCTGTTGACAGACCACCGACTG CAGTCACAAAATCAGTGATTATGCGGGTGG ACCAATCTGTTGGCGGGTGTTTCCCTCTACC TGACCCCGCAATATGGTATGTACGATCCTC GGATCTAAAATTCATAATGGCCCACCACAA CCATAAACCTCCCTAGCAGCTGGTGGTCCC GATAATTCGGGTTCTTGCCACTACTGCGACC CAGGCTCGCC (SEQ ID NO: 1028)		ATGG TTGT GGAC GGGT ATGC TTGC CGGA GCCG AGTC GTGT TTCTT AGAA CCAA CCTC GACG AGGC GAAA GCTT GCAC AAGT TAGC ACAA TTGT	

							GGTA GGGC CGAC TAGA AAAT GAGT CCCTT AGGG GGTT ACGC CTTG GCGA AAGT GAGG ACAA TTGG CATT GACG GGTG CTTC GGCA CTAG GCAA	

							AGGC GCCA CCAC ACTG TCCA ATCT CTAA AAAG TTCA CATT CATC GAAG AACT ACCG GAAC CAAC CACA CATG TGTT GAAA CCTA CACG GTGG	

							AAGG GAAA GGAA AGCT TCGC TGGA ACGA AAAG AACG GATA GGTT CCCC TTCTT GATG GCTG TGAG GCTT AGGA TGGA CGGG AAGG CCGT GAGG	

							CCTC AGGC GGGT AACT CGGC CAGA CGCT AGTT GATC TTCG GATC ACGA CAGC CCTG GCTA AGAG GAAC CCTG GATG GAGT GTGA AGGA TGGG	

							CGGG TAGG GGGT TAAG CCTG TTGA CAGA CCAC CGAC TGCA GTCA CAAA ATCA GTGA TTAT GCGG GTGG ACCA ATCT GTTG GCGG GTGT TTCCC	

							TCTA CCTG ACCC CGCA ATAT GGTA TGTA CGAT CCTC GGAT CTAA AATT CATA ATGG CCCA CCAC AACC ATAA ACCT CCCT AGCA GCTG GTGG	

							TCCC GATA ATTC GGGT TCTTG CCAC TACT GCGA CCCA GGCT CGCC (SEQ ID NO: 1274)	
CRE	Cnl1	.		Cryptococcus neoformans	CCCTCTTAATACCCCATAACACATAACAAC CCCCTAATCAACGTTCTCTGCACCTTAAACA CCACCAACATGTCCCTGCAGAGGGCCAAAA ACGCCCGTGGAGATCCTGGTCGGTGCAACC TATGCTCTGCCGACTATAGGGACCTCAAAG ATCATCTCAATAAACAACATTCCACCCATTT CTTCGTCCCCTCCGACCTCCGTGGCTCTTCC CTAGTCGCTTGCCCTCGCTGCGGCACCCCCT GCTCAGCTGGCACTGGTTTATCTCGTCACCA	CCCTCT TAATAC CCCATA ACACAT AACAAC CCCCTA ATCAAC GTTCTC TGCACC	TGAG GAAG AGGA GGTT GGAT TATTT TTTCT TTTCT TTAA	MSLQRAKNARG DPGRCNLCSAD YRDLKDHLNKQ HSTHFFVPSDLR GSSLVACPRCGT PCSAGTGLSRHQ SRYCGLTAPRIR RNRVGNSTNTS RCPPSNTAASPI

					GAGCCGGTATTGCGGTCTCACCGCTCCTCG	TTAAAC	TAAG	VPSPSPERPSPPQ
					AATCCGCCGAAATCGCGTGGGAAACTCAAC	ACCACC	TTGTT	PAEVVASLEPLS
					AAACACATCTCGCTGCCCTCCCTCCAATACT	AAC	TATTT	EAEEVLEVAQV
					GCAGCTTCACCCATCGTTCCTTCGCCTTCCC	(SEQ ID	AAGT	DAETVDTLEGT
					CAGAACGCCCAAGCCCCCCTCAGCCTGCTG	NO: 1152)	AGTT	RRAPESVPRSAE
					AAGTTGTTGCCAGTCTCGAACCATTGTCTGA		TCTTT	EGSTRVRELNM
					AGCCGAGGAGGTGCTGGAGGTCGCCCAGGT		CATT	TAPEEEHRGEEE
					TGATGCCGAGACTGTTGACACGCTGGAAGG		CGGG	SSHTNPTAPAGL
					GACCCGGAGAGCTCCGGAATCCGTTCCGAG		CAAC	ENAVSSTLGPSP
					ATCTGCCGAGGAAGGTAGCACGCGAGTTAG		CCAC	GTLPSLLPSQEC
					GGAGCTAAACATGACAGCGCCGGAGGAGG		ACGA	ANERFLYLAHLP
					AGCATCGTGGGGAGGAGGAGAGTAGTCATA		CAAC	VRSKPLPNNLVT
					CCAACCCAACTGCCCCAGCAGGGCTCGAGA		CCAA	DFMDAAERCAL
					ACGCGGTGAGCTCAACGCTGGGGCCTTCCC		TAAA	AYIAQPSDSTLL
					CTGGGACGTTGCCTTCCTTACTTCCGTCCCA		TTAA	AFLALPKVGLT
					AGAGTGTGCTAACGAAAGATTCCTGTACCT		ACAA	QALAPEQPLRPS
					TGCGCACCTGCCTGTTCGGAGCAAGCCTCT		CGAA	TFLKQFPHIPWP
					GCCCAACAACCTAGTTACCGACTTCATGGA		AAAT	EQPPARRPPSNI
					CGCCGCTGAGCGTTGTGCTCTTGCCTACATT		GCAA	RPDTTKQVIKLV
					GCACAACCCTCGGACTCTACACTGCTGGCA		CCTCT	ENGRLGAAERV
					TTTCTCGCCCTTCCAAAGGTCGGCCTCACCC		ATAA	LEEDASVAELD
					AGGCGCTCGCTCCAGAACAGCCCCTCAGGC		CCC	QGVIDQLITKHP
					CGTCAACCTTCCTTAAGCAGTTCCCGCATAT		(SEQ	KGPSCPFGNAV

					CCCCTGGCCAGAACAGCCACCCGCTCGTCG		ID NO:	GPTPGKAPDIDT
					TCCTCCCAGCAATATTCGTCCAGACACCACC		1275)	IQKALDSFKPDT
					AAACAAGTCATCAAACTCGTTGAGAATGGG			APGVSGWSVPL
					CGCCTAGGTGCGGCAGAGAGGGTGTTGGAG			LKTAAKREPVK
					GAGGATGCTTCAGTAGCCGAACTCGATCAA			QFLQLLCAAIAN
					GGGGTCATCGACCAGCTCATCACCAAGCAC			NTAPGRSMLRT
					CCCAAAGGGCCGTCTTGTCCATTCGGCAAT			SRLIPLKKDDGSI
					GCAGTGGGTCCAACTCCTGGTAAAGCTCCC			RPIAVGELIYRL
					GACATCGACACCATCCAAAAGGCCCTCGAC			CAKALIISHFQP
					TCCTTCAAGCCCGACACAGCACCCGGCGTT			DFLLPFQLGVKS
					AGTGGCTGGTCAGTCCCTCTCTTGAAGACG			IGGVEPIVRLTE
					GCTGCCAAGAGGGAGCCGGTCAAGCAGTTT			RVLEGSAGAEFS
					CTCCAACTCCTCTGCGCCGCCATCGCCAACA			FLASLDASNAFN
					ACACCGCCCCTGGTCGCTCTATGCTCCGCAC			RVDRAEMAAA
					TTCTCGTCTCATCCCCTTGAAGAAGGACGAT			VKTHAPTLWRT
					GGCTCTATCCGACCTATCGCTGTTGGTGAAC			CKWAYGDSSDL
					TTATCTATCGGCTGTGTGCGAAAGCTCTCAT			VCGDKILQSSQG
					CATCTCGCATTTCCAACCCGACTTCCTCCTC			VRQGDPFGPLFF
					CCGTTCCAGCTCGGGGTCAAGTCAATCGGT			SITLRPTLNALS
					GGTGTAGAGCCGATCGTGAGGCTGACAGAG			QSLGPSTQALAY
					AGAGTCTTGGAGGGTTCTGCCGGCGCTGAG			LDDIYLFSNDSQ
					TTCTCCTTTTTAGCCTCGCTCGATGCTTCTA			VLSKTTQFLAD
					ACGCTTTCAACCGTGTAGATAGGGCCGAGA			KQHIIKLNEKKC

					TGGCAGCAGCGGTCAAGACCCATGCGCCGA			KLISFDEIRQEGF
					CGCTTTGGAGGACATGCAAATGGGCCTATG			KMLGTMVGGK
					GCGACTCGTCCGACCTTGTGTGTGGTGACA			EKRAEFLEGRIR
					AAATCCTTCAATCCTCTCAAGGTGTTCGACA			KEMAKVGKLK
					GGGTGACCCCTTTGGCCCTCTCTTCTTCTCG			DLPHQHALLLL
					ATCACCCTCCGACCAACCTTGAATGCCCTCA			RFCIQQNLRHLQ
					GTCAATCGCTAGGTCCGTCTACGCAAGCAC			RSLRSDDLVDL
					TCGCTTACCTCGATGACATCTACCTCTTCTC			WERLDTMLWEE
					AAACGACTCGCAAGTCCTCAGCAAAACTAC			VKRMRMRQRE
					CCAATTCCTCGCCGACAAGCAGCACATCAT			DTAEEEALGRSL
					CAAGCTCAATGAAAAGAAATGCAAGTTAAT			TKLPARLGGLG
					CAGCTTCGATGAGATCAGGCAGGAGGGCTT			LLSFKDVAPLAY
					CAAGATGCTAGGGACGATGGTAGGAGGTAA			RSAAEASDTLLD
					GGAGAAGCGAGCGGAGTTTCTGGAAGGCA			NLGLLSSPEEPP
					GGATTCGGAAGGAAATGGCAAAGGTGGGC			TPIPQRTRCAEL
					AAGCTCAAGGATCTTCCACATCAACACGCG			WESQQEAILHN
					CTCCTTCTATTACGTTTCTGCATTCAGCAAA			LGDTERKRLTE
					ATCTACGACACCTGCAGAGAAGTCTGCGCT			NASRLGRSWLS
					CGGACGACCTTGTAGACCTATGGGAGAGGC			VIPYLQPLRLSN
					TGGACACGATGCTATGGGAGGAGGTGAAAA			VEIASGLHDRTL
					GGATGAGGATGAGGCAGCGGGAGGATACG			VGSSIPVCRFCG
					GCGGAAGAGGAGGCTCTAGGGAGATCGTTG			SDSPLGHDELCR
					ACGAAGCTACCAGCGCGACTGGGCGGACTA			ARNPWTQRRHN

					GGTCTACTTTCCTTCAAAGATGTAGCCCCCC TTGCTTACCGCTCGGCAGCCGAGGCCTCCG ACACTCTCCTCGATAACCTAGGTCTCCTTTC TTCGCCAGAGGAACCTCCAACTCCGATCCC CCAACGAACTCGATGCGCAGAACTCTGGGA ATCGCAACAGGAAGCCATCCTACATAACCT CGGCGACACTGAACGCAAGCGACTCACCGA GAATGCCTCCAGACTCGGCCGAAGTTGGTT ATCAGTTATCCCTTACCTTCAACCCCTGCGC CTTTCCAATGTCGAGATTGCCTCTGGTCTCC ATGACCGCACCCTGGTCGGCTCCTCGATCCC TGTCTGTCGCTTCTGTGGGTCGGACTCACCT TTGGGTCACGACGAGCTTTGCCGCGCCCGC AACCCCTGGACCCAGCGCCGGCACAATGCC ATCAACCGCGTCATTTATCAACACCTCAAA CAAATTCAAGGTGCCACGGTTGAGATTGAG CCCCACACGCTGTCGGGACAAAGGAGAAAC GACCTTCGGGTCAGAGGTTCCAGCGCTCTG GCCTTCACTGACTACGACCTGAAGGTTTACT CCCTCGGGGACCGAGACGCGAGAAGCACCG TCACACCCTGCGCCCCCAACGGCAAGCTGG CCGACTTCTGCTTGGACCGGTGCGTGAACT GGCTCGACAAGGTGGGTCAGGTCGTCTCTA			AINRVIYQHLKQ IQGATVEIEPHT LSGQRRNDLRV RGSSALAFTDY DLKVYSLGDRD ARSTVTPCAPNG KLADFCLDRCV NWLDKVGQVV SKNAPKVTGGV FKPIILSTGGLMS RSTADEWKDW RDAMPVGGFEK MEKRIGVELVK ARARTLVL (SEQ ID NO: 1397)

					AGAACGCTCCGAAGGTCACTGGTGGGGTCT TTAAACCAATCATCCTTTCCACTGGTGGCTT GATGAGCAGGAGCACAGCAGACGAATGGA AGGACTGGAGGGACGCGATGCCGGTGGGG GGGTTCGAGAAAATGGAGAAACGGATTGGT GTCGAGTTAGTAAAGGCAAGGGCGAGGAC GCTGGTCTTATGAGGAAGAGGAGGTTGGAT TATTTTTTCTTTTCTTTAATAAGTTGTTTATT TAAGTAGTTTCTTTCATTCGGGCAACCCACA CGACAACCCAATAAATTAAACAACGAAAAA TGCAACCTCTATAACCC (SEQ ID NO: 1029)			
CRE	CRE-12_CCri	.		Chondrus crispus	ACGCCCCCTATCCATTTCTGCCAGCCTCCCA TCGGCTCGCCGTCTCCGCAACCCCTCTTCCT CGGCTGTACCAGTTCCGCTCCCACAACCTCC CTCGCCACAATGAGTCAACCTAATATTTCGT CCGCTGAGACCCCCTTGTCTCAGCTGCCCAC GCCTGTTCCCACCCCGCCTTCTCCCTCCAAT CCCTCTCTCTCTCTCCCTACTGTGCGTGACC TCCTCCTCTGCCCCATACGCTCCTCCCATGT TTACTCATCCATCCCTTCCTCGTGCTTACAC AGTTTCACGATGCTCCTCATCAAGACTGTCC GCGCTGCGTCGGCCACAATGACCCCAACTG AATCMCATCGCGCATTCATTCATCTACACA	ACGCCC CCTATC CATTTC TGCCAG CCTCCC ATCGGC TCGCCG TCTCCG CAACCC CTCTTC CTCGGC TGTACC	TAAG TCCTT GACG CCTG CCCC GTGA TACA GCAT CGGT ACCC CTAG CATTT	MSQPNISSAETP LSQLPTPVPTPPS PSNPSLSLPTVR DLLLCPIRSSHV YSSIPSSCLHSFT MLLIKTVRAAS ATMTPTESHRAF IHLHILPIAVLRR SFRGETGWRSR TGQHHALRQRI RRASSGRHWAA LWHEALAAHQV

					TTCTTCCCATCGCTGTCTTGCGACGCTCGTT CCGTGGAGAAACCGGATGGCGTTCCCGCAC AGGGCAACATCATGCCCTCCGCCAACGGAT TCGCCGCGCCTCCTCGGGTCGGCATTGGGCT GCCTTGTGGCACGAAGCCCTTGCTGCACAT CAGGTCGACCTTGACTACCGCACGCGTCAC AGCCGTCGTTACCAAGCCTCCGCTACATCG CGCCACCGCATCGGCCGTGCCATGCGTCTG GCCGCCGATGCCCAATATGGACGTGCAATG TCGGCCCTCAAGGCCAAACCGCTGCCCGAT CTACATGCTGCCGCCACCCGCGACACACTC ACCGCGCTTCACCCTCCTCCTGCCAGCCCGG TTCAGCCTCTCTCACCGACTGACCTCCCCCC GGTCCCTGAAATTACGGAAGGTCAAGTCCT CCGAGCGGCACGCGCCCTTAATCCCACATC CGCTGCAGGACCGGACCATCTCTCTCCCCGT ATCCTGCAGCTCTTAGCCCGCACCACTATCA GCCCGGAAGCTGGGGTTACGGGGTTGTCCG CATTGACGAACCTGGTTCGACGTCTCGCCC GAGGTGACATTCCGGATAGAACTGCGCCTC TCCTTGCTGCTGCCACTCTGATCCCCCTCCA ACCCCGCCCTCACAAAATACGGCCGATTGC TGTAGGGCAGGCTTTGCGCCGTCTGGTCAC	AGTTCC GCTCCC ACAACC TCCCTC GCCACA (SEQ ID NO: 1153)	GAAT AAAA AA (SEQ ID NO: 1276)	DLDYRTRHSRR YQASATSRHRIG RAMRLAADAQ YGRAMSALKAK PLPDLHAAATR DTLTALHPPPAS PVQPLSPTDLPP VPEITEGQVLRA ARALNPTSAAG PDHLSPRILQLL ARTTISPEAGVT GLSALTNLVRRL ARGDIPDRTAPL LAAATLIPLQPR PHKIRPIAVGQA LRRLVTKVLLPP AIQDTRDHLLPE QLANSVASGMD AIVHDTRMLMH RHGRNPDYIMV SVDARNAFNTFS RQSLLDRLPLQT PSLARFLNLIYG

					GAAGGTCCTTCTGCCCCCCGCCATCCAGGA			RTVPDLVLPSSP
					CACCCGCGATCACCTTCTCCCAGAACAGCT			RFLMKSQEGTQ
					CGCCAACTCGGTTGCCTCGGGCATGGACGC			QGDPASMLLFSL
					AATCGTCCATGACACGCGCATGCTTATGCA			AIQPLLRRLTRE
					TCGTCACGGTCGAAACCCAGACTACATCAT			CRLDLNRWYAD
					GGTCTCCGTAGACGCGCGGAATGCCTTTAA			DGTLVGPISEVI
					CACCTTCTCACGTCAGTCCCTGCTGGATCGT			KALRILRDDGPQ
					CTCCCTCTGCAGACTCCTTCCCTCGCCCGTT			SGFHVNINKCRA
					TTCTCAATCTAATCTATGGCCGCACCGTTCC			YWPTVMPEKLS
					TGATCTCGTGCTGCCCTCTTCTCCGCGGTTT			ELLRIFPLHVEC
					CTGATGAAAAGTCAGGAGGGCACCCAACAG			GEGGVALLGAP
					GGGGACCCGGCAAGTATGCTCTTATTTTCGC			LGTDAFVRRHL
					TGGCAATCCAGCCGCTCCTGCGTCGTCTCAC			MNKVQSCHASL
					CCGCGAGTGCCGTCTCGACCTGAACCGCTG			SLLDEIPDARTR
					GTACGCGGATGACGGCACTCTGGTCGGGCC			FHLHRVTGSVC
					AATCTCGGAGGTCATCAAGGCACTCCGAAT			KVEHVFRLTPPH
					TCTTCGTGATGACGGCCCGCAGTCCGGATTT			LSLPAATKFDEQ
					CACGTCAATATCAACAAGTGCCGGGCATAC			QIAAYSRLNDV
					TGGCCGACCGTAATGCCAGAAAAGTTGTCC			AVSTSMATQIGL
					GAATTGCTCCGTATCTTCCCCCTTCACGTCG			PFRLGGHGFTPL
					AGTGCGGCGAAGGCGGTGTCGCCTTGCTGG			SPFIHASYAASLI
					GTGCCCCGCTCGGCACAGATGCCTTTGTGC			EAAPVRVKGPH
					GCCGACATCTCATGAACAAGGTTCAATCAT			NPSESFYRRMA

					GCCATGCCTCCCTCAGCCTCCTTGATGAAAT			RRHIVHVLGAL
					TCCCGACGCGCGTACGCGGTTTCACCTTCAC			NPEVRTRGILGT
					CGTGTAACAGGCTCGGTATGCAAAGTCGAG			HSPLGPFEPEAL
					CATGTTTTTCGCCTTACACCTCCCCACCTCT			LSRPERVHHTLI
					CCCTCCCAGCAGCTACAAAATTTGATGAAC			QAMQGATSRLY
					AACAAATCGCTGCTTATTCTCGGTTGAATGA			WEHTAWDLDPL
					TGTGGCCGTTTCCACATCCATGGCTACACAA			PRNHSAASVRR
					ATAGGCCTTCCGTTTCGCCTCGGTGGACACG			RARYNSLRAPG
					GCTTCACCCCACTGTCACCATTCATCCATGC			AASFLCSHPSLT
					GTCCTACGCTGCCAGTTTAATTGAGGCGGC			SRVPSAVWSCM
					ACCTGTTCGTGTGAAGGGCCCACATAACCC			LRRHLDTPVYC
					CTCCGAGTCGTTTTATCGCCGCATGGCCCGT			DSIRPLICSHCCK
					CGTCATATCGTCCACGTACTAGGGGCCTTG			PMDARGDHAAI
					AACCCTGAGGTCCGCACCCGAGGCATTCTT			CRHGFGVVHRH
					GGGACCCATTCTCCCCTCGGACCATTTGAAC			NTVRNLLARHA
					CAGAGGCCCTTTTGTCTAGACCTGAACGCG			FRAAGLCCDLE
					TACACCACACATTAATTCAGGCTATGCAGG			VPSLLPNTANRP
					GGGCCACTTCTCGACTCTACTGGGAACACA			ADILVQPAPPPS
					CCGCGTGGGACCTTGACCCTCTCCCTCGCAA			GALPDRPTAYD
					CCACAGTGCCGCCTCTGTCCGCCGACGTGC			VTVRSPYCRST
					CCGGTACAATTCCCTCCGTGCTCCGGGCGCC			MSLAAKGLAGA
					GCGTCGTTTCTATGCAGCCACCCCTCACTCA			AEAADLDKLRV
					CTTCTCGAGTCCCTTCTGCGGTGTGGTCCTG			HSRTVRDAFHL

					TATGCTACGCCGGCATCTGGATACACCCGT CTACTGTGACTCTATTCGGCCTCTCATATGT TCGCATTGCTGTAAGCCAATGGACGCTCGC GGTGATCATGCTGCAATATGCCGTCATGGC TTCGGCGTCGTTCACCGTCATAACACTGTAC GCAACCTACTCGCCCGTCACGCGTTCCGCG CCGCCGGTCTCTGCTGCGACCTTGAGGTCCC TTCTCTCTTGCCGAATACCGCGAACCGCCCC GCCGATATTCTCGTCCAGCCCGCCCCGCCTC CTTCGGGCGCTCTCCCGGACCGCCCCACTGC GTACGACGTAACCGTTCGTTCCCCCTACTGT CGCTCTACAATGTCTCTCGCTGCGAAAGGC CTCGCGGGTGCAGCGGAAGCTGCTGATTTG GACAAGCTTCGCGTCCATTCCCGTACAGTG CGTGACGCATTTCACCTCCAGCCTGACTCCC CACTCCCTCTACTCGACTGGCACTTTGTCCC GCTCGCATTTGATACTCTCGGCGCGACCAG CTCTCGCACGATGGCAGTCCTTGAGTACCTC GCTCACCGCATTGCCAACCGGACATATTCA TCTTACGGGACCGCCAAGATACGTCTACTA CAACGCATCAGTTTCGCTGTTTGGTCCAGTT TGGCCTCTGCCACCCTTTCCCGTATGCCCTA TCACGGCGCGGCCCTATCGAGCCCCGCCCA			QPDSPLPLLDW HFVPLAFDTLGA TSSRTMAVLEY LAHRIANRTYSS YGTAKIRLLQRI SFAVWSSLASAT LSRMPYHGAAL SSPAQV (SEQ ID NO: 1398)

					AGTGTAAGTCCTTGACGCCTGCCCCGTGAT ACAGCATCGGTACCCCTAGCATTTGAATAA AAAA (SEQ ID NO: 1030)			
CRE	CRE-13_CCri	.		Chondrus crispus	CNCCAGCCAMCGATCCCGCCGCCACTCGCM GCCCGGCCGTCTCGACCGCCACCTCCCCGAS GCCCCAGCCCATCATGGCCTSTWMGCCCCW GATWTCTCCCCCGGGGGCCCCCCCTGCTCC GCTGCGGTACCGGATGTTACAATGTCCACCS CCGCTACCSAAGCSMCASTMGTTKCCGGTC CCCCACCCGATGTCGTCACCCATACGCCKC CGCCTCCCMCACCGGMCGATGAGGGGACC CCCTTCCCSAACCCCTCCCCGAGACATGCAC AGACCSCATGGCACCCCCGGWCCGCACTCT CACCGATKCTGTGGCCGCCCTCCCMACCAC TGCACCCATGCCTCCSCACAACCSCGSMSAG CCCMGCACSCTCTCCAAWGGCCGAAACTSC GCAGCCCACCTCCACATCCGCATGTTTCACC CCTCATCCTCTGCASTGGCCCCCTCCCAACG CCCATGACCCAACCACGCATGAAACGAGCG CTCTCCASAAGCGCCAAGGCGCCCCCTACC AAGCGCCCCTCTGCCTCTCAGGGCCCAGCC GCGTCTTCCCATGACKAATGACCSCGGACG CCCCCACCTCMACCCCCGCGCCCGCCTCCC	CNCCAG CCAMCG ATCCCG CCGCCA CTCGCM GCCCGG CCGTCT CGACCG CCACCT CCCCGA SGCCCC AGCCCA TC (SEQ ID NO: 1154)	TAAT TCAC CTTC ATAT CTGC TAGT GTCT CTGT AAGC GCAC CCCT CATG CATT GATA AAAT TACC CCCC A (SEQ ID NO: 1277)	MAXXPXISPPGA PPAPLRYRMLQ CPPPLPKXXXXP VPHPMSSPIRXR LPHRXMRGPPS XTPPRDMHRPH GTPGPHSHRXC GRPPXHCTHAS XQPRXAXHXLQ XPKLRSPPPHPH VSPLILCXGPLPT PMTQPRMKRAL SXSAKAPPTKRP SASQGPAASSHD XPRTPPPXPPRPP PYRFPPPTLDQH XFALSXAYPHP XPRRPPSPXRXL RHSFPPRFGXQT FSSIPGPRLHSTV

					TACAGGTTCCCTCCCCCWACTCTCGACCAG			LLLIRLVRAATA
					CACTTMTTCGCCCTCTCWTMAGCCTACCCC			ANTPETTTLXSC
					CACCCGMCTCCCMGGCGCCCACCCTCMCCT			TFTCSRLPFFERP
					TKCCGTCSGTTGAGGCACTCCTTTCCTCCCC			SXAXLAGGPRA
					GATTCGGCCSTCAGACATTTTCTTCCATTCC			VNFMLSACXYG
					GGGACCGCGCCTTCATAGCACTGTATTACTT			ERVXDESGXSY
					CTCATCCGCCTCGTCCGCGCCGCTACAGCCG			XXKXHCITSHPP
					CCAACACTCCCGAAACCACCACACTGMATT			RPRRYSRQHSR
					CTTGCACCTTCACCTGCTCCCGACTGCCSTT			NHHTXFLHLHL
					CTTCGAGAGGCCTTCCGKGGCGAGSCTGGC			LPTAXLREAFRG
					TGGAGGTCCTCGCGCGGTCAACTTCATGCTC			EXGWRSSRGQL
					TCCGCTTGCKGATACGGAGAGCGTGTACKG			HALRLXIRRACT
					GACGAGAGTGGGGWCTCTTATGKAAWGAA			GREWGLLXXEA
					GCMCTAGATGCCCACAGCKCCAGGACAGA			LDAHSXRTEWQ
					ATGGCAGCACACGCATGCCCGGCGCCCCTC			HTHARRPSPPVS
					GCCACCCGTTTCCCCATCGGCACGTGCCGCC			PSARAARAMRL
					CGCGCTATGCGCCTTGCCTCCCAAGCTCAAT			ASQAQYGRAMR
					ACGGCCGCGCCATGCGCACATTTACCAACC			TFTNPPLADLND
					CCCCTCTAGCTGACCTCAACGACCCGGCCA			PATMERLQALH
					CGATGGAGCGGCTCCAAGCCCTTCACCCCA			PTPTVPVVPLPP
					CTCCTACCGTGCCCGTCGTGCCCCTGCCACC			SAQPRPPEVTXE
					CTCCGCACAGCCTCGACCACCCGAAGTCAC			AVXRAVRRLNP
					CGMGGAGGCGGTCWTGCGTGCGGTTCGTC			NSAAGPDRMSP

					GCCTCAATCCGAACTCGGCGGCCGGCCCTG			KLLHLLAHTPIS
					ATCGCATGTCCCCGAAATTGCTTCACCTCCT			PEAGVTGLSALT
					GGCTCACACTCCCATAAGCCCAGAAGCGGG			NLVSRLARGSLP
					CGTCACCGGTCTCTCKGCGCTAACCAACCTC			PCTIPLASAATL
					GTCAGCCGCCTGGCTCGCGGCTCCCTCCCAC			LPLQPRPGKIRPI
					CCTGTACGATCCCACTGGCCAGTGCGGCGA			AIGQALRRLVTK
					CACTTCTKCCGTTGCAGCCCCGACCGGGAA			XLLPAAIDDCRD
					AAATACGCCCGATCGCTATTGGGCAAGCCC			HLAPEQXANGIP
					TWCGCCGGCTTGTCACAAAAKTMCTTCTTC			NGIDAIVHDAR
					CTGCCGCCATCGACGACTGTCGGGACCACC			MLVRRHGNDPH
					TTGCTCCCGAACAAMTGGCMAACGGCATAC			YXMVSIDASNA
					CMAACGGCATTGACGCTATCGTACACGACG			FNNFSRQQVLD
					CACGCATGCTAGTACGACGCCACGGTAACG			QLPTRAPSLSRY
					ACCCACACTACKTAATGGTGTCTATTGACG			LDMVYARAPSP
					CTTCCAATGCGTTCAATAATTTCTCACGSCA			LVLPSXPPTILHS
					ACAAGTCCTCGACCAGCTGCCCACTCGAGC			RXGSQQGDPAS
					ACCATCGCTCTCACGATATTTGGATATGGTG			MLLFSLALQPLT
					TACGCACGCGCCCCCTCCCCCCTCGTCTTGC			RLISRECXLXMN
					CTTCATSCCCGCCTACCATACTCCACAGCCG			RWYADDGTIIG
					GGAMGGATCACAACAAGGGGACCCTGCAA			RIDEVXKALDIIT
					GCATGCTCCTTTTCTCGCTTGCCCTCCAGCC			KEGPRFQFFLNP
					GCTCACGCGCCTCATTTCACGTGAGTGTGA			SKTRVFWPSRQ
					MCTWKTAATGAACCGCTGGTATGCGGACG			XDLLSPLMTVG

					ACGGAACTATCATTGGACGGATTGACGAAG			PLRVIDEGGVXL
					TTKCCAAAGCCCTTGATATCATCACTAAAG			LGAPIGSPSXMA
					AGGGGCCCAGGTTCCAATTCTTCCTCAACCC			QYIREXLNTCKT
					TTCGAAGACACGCGTCTTCTGGCCAAGCAG			ALAHLDHIPEAR
					GCAGCMAGACCTCCTCAGCCCGCTCATGAC			MRFHLHRVSAS
					AGTGGGTCCTCTGCGMGTCATCGATGAAGG			ACRLQHLFRLVP
					CGGTGTGGMTCTGCTCGGCGCCCCCATMGG			PDFAXPFAQQFD
					GTCACCAAGCTWTATGGCACAGTACATTCG			RDQLXAYXRFN
					GGAAAAWTTGAACACTTGCAAAACCGCMC			SVTMSPRIVPKY
					TCGCCCATCTCGACCATATCCCCGAGGCCC			GCXFXTXATAS
					GCATGCGCTTTCACCTGCATCGGGTGTCTGC			PHWHLPYTPXT
					TTCTGCATGCCGCTTGCAGCACCTCTTCCGG			LLASSIPLQHGY
					TTGGTCCCCCCGGATTTCGCGWTGCCGTTTG			KVPTFPPSLSISV
					CACAACAATTCGACCGTGACCAACTCGMAG			LHEARCGSFFEI
					CCTATGMGCGCTTTAATAGTGTGACTATGT			YLHSXNPHTSRC
					CGCCAAGAATCGTGCCCAAATACGGCTGCS			DYVAKNRAQIR
					TTTTTCMCACGGWGGCCACGGCCTCACCTC			LXFSHGGHGLTS
					ATTGGCATCTACCATACACGCCTCWTACGC			LASTIHASYAAS
					TGCTAGCCTCATCGATACCGCTCCAGCACG			LIDTAPARLQGP
					GCTACAAGGTCCCCACTTTCCCGCCGTCTCT			HFPAVSQYQRF
					CAGTATCAGCGTTTTGCACGAGGCCCGTTG			ARGPLRVVLRN
					CGGGTCGTTCTTCGAAATTTACCTTCATTCG			LPSFVQPAHFSM
					TKCAACCCGCACACTTCTCGATGACGGAAW			TEXDLGCLEPX

					CGGACCTCGGMTGCCTTGAACCAKCTGCGC			ALLARPERIHTF
					TACTGGCGCGACCTGAACGCATACACACCT			LLQAQYSAAAS
					TTCTACTTCAGGCGCAATACAGTGCAGCAG			SYWQXPLWESF
					CWAGCTCGTACTGGCAAAWACCCCTCTGG			PNPGDHSAASLR
					GAGTCCTTCCCCAACCCTGGTGATCACAGC			KRVRYNSLLAP
					GCAGCCTCGCTACGCAAACGAGTACGCTAC			GATSFLTAHPAA
					AACTCCCTGCTTGCCCCWGGGGCCACCAGT			TSRVHNATWST
					TTTCTCACTGCACACCCSGCCGCCACCTCTC			MLRRHLDAPVT
					GGGTCCACAACGCAACTTGGTCCACGATGT			NDSISPLRCXHC
					TACGTCGGCACCTCGACGCCCCCGTGACCA			SKPMDARGDHA
					ACGATTCCATATCGCCGTTGCGATGTKCTCA			XIXSHGFGTLHR
					CTGCTCCAAGCCTATGGATGCCCGCGGCGA			HNTVRNVLARQ
					CCACGCGWCCATTKGCAGCCACGGGTTTGG			LFRVAGLAYSLE
					TACGTTGCACCGGCATAACACCGTCAGGAA			VPFLIPNTAARP
					CGTCCTCGCCMGGCAGTTATTCCGMGTCGC			ADILVQPPPPAP
					TGGCCTCGCCTACTCGCTCGAAGTACCCTTT			GLPPDXPTAYD
					CTGATTCCCAACACCGCCGCCCGTCCCGCA			VTICSPFRRGML
					GATATTCTCGTCCAACCACCTCCTCCAGCCC			YHAARHRGGAA
					CTGGCCTACCTCCTGACMAACCCACAGCCT			DAASVRKXKAL
					ATGATGTCACGATTTGTAGCCCTTTTCGCCG			ERTIRXALLIED
					CGGAATGTTATACCATGCCGCCCGTCACCG			DNXPPPLDWHF
					CGGCGGAGCCGCCGACGCCGCATCTGTAAG			QPLSFDALGAPS
					GAAGMGCAAAGCCCTCGAGCGCACTATCCG			QSTVHVIEDHA

					CMACGCTCTCCTTATCGAGGACGACAATCS TCCGCCGCCTCTTGACTGGCACTTTCAACCG CTTTCCTTCGACGCWCTGGGMGCCCCCTCT CAGTCTACTGTACACGTTATCGAAGATCAC GCTAAGCTCATGGCCCTCCGCAACTCGTGC ACMATTGCAACTGCCAAATCACGCATCCAA CAACGCCTCAGCTTTGCTATATGGTCCAGTG CTGCCGCCGCTATCCTCTCTCGCCTACCGAC ACACGCCGCGGACATCTCATACCCGATAGA AGTATAATTCACCTTCATATCTGCTAGTGTC TCTGTAAGCGCACCCCTCATGCATTGATAA AATTACCCCCCA (SEQ ID NO: 1031)			KLMALRNSCTIA TAKSRIQQRLSF AIWSSAAAAILS RLPTHAADISYPI EV (SEQ ID NO: 1399)
CRE	CRE-1_ACas	.		Acanthamoeba castellanii	TAACCCTAACCCTCTCCCTCGGCCCCTCTAC CCTAAAGCGCCCTAATCGACCGGCGACGCC CTAATCGCTACCCTCTACGCCCTAATCGACT TTGGCGCCAAAGCGACTTTCCCCGGCCGAT TTTCTTCCTGCCTTTTTCTTTTCTCTCCAAGC GACGCGCCTTTTACTTTGCCGCCGTTCTGTT TTTTCTTTTCTCTTTGCACTTCGCTTCTACTT CACACCTCCTCCTCCTCCTTCTCGACCCGCG CGGCCTCGAGCGACTTGCTGCAGCGGCTCC CGGCCTCCCCCACGCGGCCTGCTACTCCCGC TTTCTAGACGCCCCCGGTCTTGCTCTCAGTC	TAACCC TAACCC TCTCCC TCGGCC CCTCTA CCCTAA AGCGCC CTAATC GACCGG CGACGC CCTAAT	TAAG CCGC GCGA CGAG GACG GCCA GGAC GACC AGGA CGAC GGCG	MATTTISRSPSSS SSSSSARSRASA STSASVASIPRLF RDGRFHCPLAH CQTRTSTWQDL SAHLTRMHDGD VPRDVAAACGI VQCLHEGCRKW FRGAAGLASHR GKARHAPPPAP RAALAVAAVPR

					TCCCGCATCGAAGCGGTAGTCGGGGTACGT	CGCTAC	ACGG	ADSRGRTPAPTP
					GCTCAAGTGACTCAAGCCTCTTTTCAGCCTC	CCTCTA	CGAC	SVAPPXAGPPPR
					GGCGCTCTCTCAATCCGCCTCAGTCTTAGCC	CGCCCT	CACC	AAPRAAPSPLPC
					TTTCAAGTTGCTCGATTACGCTCTCGAATCG	AATCGA	TAGC	PPALPHPPPSASP
					CTCTCTCTCTCAGTCTCAGTCTCAGTCTCAA	CTTTGG	ACCG	PTSSVTSPCSPPT
					TCTCGATCTTGCCTTCGCCTTCGTCTCGACG	CGCCAA	CACG	TPPSQPSPDLFSG
					CCTTGCTCTCGWAATCGCTGCCACTACGTG	AGCGAC	CGCC	FANAPTTPSPPS
					CCAGCTTTTTCGTGCCTTGTCTTCGTGTCGA	TTTCCC	ACGA	TPXSSPAGSPIPA
					CCCGGACCGTTTGCAAGCCCTCGCCTTCGTA	CGGCCG	CATA	ARRFVLPVATPY
					CCCCGCTCTCGTAGCCGTTCTCATCGCTGAA	ATTTTC	TTGTC	PAPAPRANRPKL
					GCGTTCTACGCGCTGGCAGCAAGCCTCGGC	TTCCTG	GCGC	SPVARPFVPKAR
					CCTAGCTTGTAGCGCCGCCGGTGGCCGCTC	CCTTTTT	GCTG	AGAIPEASSPVT
					GCCAACATGGCTACGACGACCATCTCACGA	CTTTTCT	TACA	PQDRAVSRRED
					TCCCCTTCGTCTTCTTCTTCTTCTTCCGC	CTCCAA	GGCG	AAAAPSSAPGL
					TCGCTCGCGGGCGTCAGCTTCCACGTCCGCC	GCGACG	GCTA	GLADEHEDDDT
					TCAGTTGCGTCGATACCCCGCCTCTTCCGCG	CGCCTT	GGTC	YGGDTIALTAPH
					ATGGCCGCTTCCACTGCCCTCTCGCCCACTG	TTACTT	GAGC	APRETRAPFEFE
					TCAGACCCGCACGTCCACATGGCAGGACCT	TGCCGC	CCAG	ACFLEEEAPATA
					CTCCGCGCACCTCACACGCATGCACGACGG	CGTTCT	CCGA	GDLPPYARAFL
					TGACGTGCCCCGTGACGTCGCCGCCGCCTG	GTTTTTT	CCGT	ACPSARLQEIPR
					CGGCATCGTGCAGTGCCTACACGAGGGCTG	CTTTTCT	TCTG	RLKSAWQAAAK
					CCGCAAGTGGTTTCGCGGAGCTGCAGGACT	CTTTGC	AGCC	TIAEAALDCHTA
					TGCCTCTCATAGGGGCAAGGCCCGTCACGC	ACTTCG	TCAG	GDTQGYNAHLR

					CCCGCCACCAGCCCCCCGCGCCGCCCTCGC	CTTCTA	TCGG	LFIELPARGLAV
					CGTCGCCGCCGTGCCCCGCGCGGATTCTCG	CTTCAC	CTTG	PTNCRGAARTK
					CGGTAGGACCCCAGCCCCGACCCCCTCGGT	ACCTCC	AGCC	LQRERLLDIAAG
					AGCCCCGCCCWACGCCGGTCCTCCGCCGCG	TCCTCC	CCCG	RIPAIPDPPCDAP
					AGCTGCGCCACGCGCCGCCCCCAGCCCACT	TCCTTC	GCTT	GADDALRGFPV
					GCCGTGCCCGCCTGCGCTCCCGCACCCGCC	TCGACC	CCCA	SGTTAGDVSND
					CCCCTCTGCCTCTCCTCCCACCTCCAGCGTG	CGCGCG	AGGC	DDSGGVHDRPA
					ACGTCCCCGTGCTCCCCGCCCACGACCCCG	GCCTCG	CTAC	ATASARQAKRL
					CCGTCGCAGCCTTCGCCCGACCTGTTCTCGG	AGCGAC	CGGG	VEQGLSSRALR
					GTTTCGCGAACGCGCCTACCACGCCATCGC	TTGCTG	GCGG	ALERGEPAVAS
					CGCCCTCCACGCCGWCWTCGTCGCCAGCAG	CAGCGG	CTCTT	ADTLGRLEALH
					GCTCGCCCATCCCGGCTGCCAGACGCTTCGT	CTCCCG	TTTCG	PPNPTDRGLWP
					CCTGCCTGTGGCCACGCCCTACCCGGCCCCC	GCCTCC	CCCT	GAPKAAIPRVTA
					GCGCCGCGTGCTAACAGGCCCAAGCTGTCG	CCCACG	GGTT	KHLAQVAKELP
					CCGGTCGCGCGCCCCTTCGTCCCTAAGGCG	CGGCCT	TTTGC	RGSAPGPSGWT
					CGAGCCGGAGCGATACCTGAGGCGTCCTCA	GCTACT	CGGC	FELVQAAIDRQP
					CCTGTGACGCCTCAGGACCGCGCCGTCTCA	CCCGCT	CTGTT	TGTVAAFLIDM
					CGCCGCGAAGACGCCGCCGCCGCCCCTTCG	TTCTAG	TTTTC	AQRALRGTLHW
					TCCGCGCCGGGCCTCGGCCTAGCAGACGAA	ACGCCC	TCTCC	RGLLTASRLVAL
					CACGAGGACGATGACACGTACGGCGGTGAC	CCGGTC	CCCTT	KKPDGGVRPIA
					ACAATCGCGCTCACTGCCCCCCACGCGCCC	TTGCTC	TTCCC	VGEALYRVIGRL
					CGTGAGACCCGCGCCCCCTTCGAGTTCGAG	TCAGTC	CCCTT	VLKADRVMSSA
					GCGTGTTTCCTCGAGGAGGAAGCCCCAGCC	TCCCGC	TTCC	DATQYVGRHQY

					ACCGCCGGCGACCTCCCGCCTTACGCGCGC	ATCGAA	ATTT	GVAYPGGVEAP
					GCCTTCCTCGCTTGCCCGTCAGCCCGTCTCC	GCGGTA	GTAC	VHAVRELHDSG
					AGGAGATCCCGCGCCGCCTCAAGTCCGCGT	GTCGGG	TTAG	QLRAVVSLDWR
					GGCAGGCCGCCGCCAAGACCATCGCGGAAG	GTACGT	TTTTT	NAFNSLDRVHT
					CCGCGCTGGATTGCCACACCGCGGGCGACA	GCTCAA	CCTTC	ALLIADRAPALA
					CGCAGGGCTACAACGCCCACTTGCGGCTCT	GTGACT	GGCC	RLYEWSYREDS
					TCATCGAGCTGCCGGCCCGCGGGCTGGCAG	CAAGCC	GCGG	VLVLPRAFEKA
					TGCCCACCAACTGTCGAGGCGCCGCCCGCA	TCTTTTC	CAGC	GLPASLLSQAGV
					CCAAGCTTCAACGAGAGCGCCTGCTCGACA	AGCCTC	TTGTT	RQGDVLGPLFF
					TCGCCGCCGGAAGGATCCCGGCCATCCCGG	GGCGCT	GCCC	AIGAAPVLDEID
					ACCCGCCGTGCGACGCCCCGGGCGCCGACG	CTCTCA	GGCA	AIPYVTPRAYLD
					ACGCCCTACGCGGTTTCCCCGTATCGGGGA	ATCCGC	TAGT	DIFVTIPHGVTD
					CGACAGCCGGCGACGTCAGCAACGACGACG	CTCAGT	GTTA	AATKAAVAATF
					ACAGCGGAGGTGTGCACGATCGGCCCGCGG	CTTAGC	ATAT	ATAEREGAAAG
					CCACCGCCAGCGCCCGGCAAGCCAAACGGC	CTTTCA	GTTT	LRLNRCKSAVW
					TAGTGGAGCAAGGGCTCTCCTCCCGAGCCC	AGTTGC	AAAA	AADAEALLPPH
					TCCGTGCCCTTGAACGGGGCGAGCCCGCGG	TCGATT	AACG	AAGAREDVESC
					TCGCCTCGGCAGACACCCTCGGGCGTCTCG	ACGCTC	TGTA	APVREGLKILGA
					AGGCGCTCCACCCGCCTAACCCCACCGACA	TCGAAT	AATA	PVGSPAFVAKSL
					GAGGACTATGGCCCGGCGCCCCGAAGGCCG	CGCTCT	AATA	DGIIKRAIGTLDL
					CGATCCCGCGGGTCACCGCCAAGCACCTGG	CTCTCT	ACTG	VADAELPLQHK
					CCCAAGTGGCCAAAGAGCTCCCGCGCGGTA	CAGTCT	TTTA	LVLLRQCVAQIP
					GTGCGCCGGGTCCCTCGGGCTGGACGTTCG	CAGTCT	ACCC	TFWARAVPDAG

					AGCTCGTGCAGGCGGCCATCGACCGCCAAC	CAGTCT	TAAC	PALAVWDTALL
					CCACGGGCACGGTTGCCGCGTTCCTCATCG	CAATCT	CCTA	RRTGALVGLDV
					ACATGGCGCAGAGAGCCCTCCGGGGCACCC	CGATCT	ACCC	RDGSLQADIARL
					TACACTGGCGGGGATTGCTCACCGCCAGCC	TGCCTT	TAA	PVRLGGLGLRS
					GCCTTGTCGCGCTGAAGAAGCCCGACGGCG	CGCCTT	(SEQ	MKDTAPRAFVA
					GTGTACGACCCATCGCCGTAGGCGAGGCCC	CGTCTC	ID NO:	SILFAAALANTR
					TCTATCGCGTCATCGGCCGCCTTGTTCTCAA	GACGCC	1278)	RSELTCSASTAR
					GGCCGACAGGGTGATGTCGAGCGCCGACGC	TTGCTC		RLRAALPELART
					CACGCAATATGTCGGGCGGCACCAGTATGG	TCGWAA		DACNDEAAWR
					CGTGGCCTACCCCGGTGGGGTTGAGGCCCC	TCGCTG		RSIARGVFPDVD
					GGTCCACGCCGTCCGCGAACTGCACGACAG	CCACTA		KLGTTQLQRVL
					CGGCCAGCTCCGAGCGGTCGTCTCGCTCGA	CGTGCC		QGMADSKSAHR
					CTGGCGTAACGCGTTCAACTCGCTCGACCG	AGCTTT		TRRQVPFLFAAV
					CGTGCACACGGCCCTGCTCATCGCCGACCG	TTCGTG		FEDAATPGSGA
					CGCACCCGCTCTCGCGCGACTCTACGAGTG	CCTTGT		WLAAIPSDPTLV
					GTCCTACCGTGAGGACTCAGTCCTCGTGCTG	CTTCGT		LPDAELAEAVRI
					CCGCGCGCGTTCGAAAAGGCGGGGCTGCCG	GTCGAC		KLLTTTANAAG
					GCCTCCCTGCTCTCCCAGGCCGGCGTGCGCC	CCGGAC		VCPACHKTGIDP
					AGGGCGACGTCCTGGGACCCCTCTTCTTCGC	CGTTTG		SHAYTCVSLSHL
					CATCGGCGCTGCCCCGGTCCTCGACGAGAT	CAAGCC		RTARHDVVVRR
					CGACGCCATACCGTACGTGACGCCGCGAGC	CTCGCC		VELACKTEKPV
					GTACCTCGACGACATCTTCGTCACGATACCC	TTCGTA		REHVLAIPPVAP
					CACGGTGTCACGGACGCCGCGACCAAGGCC	CCCCGC		TDNNNNGDEDG

					GCCGTCGCTGCCACCTTCGCTACGGCGGAA	TCTCGT		SPVTTADDNAD
					CGCGAAGGCGCGGCCGCTGGCTTGCGGCTC	AGCCGT		GHAVATKRRPE
					AACCGCTGCAAGTCGGCGGTGTGGGCCGCG	TCTCAT		TRASARAAAAA
					GACGCAGAAGCCCTCCTTCCCCCCCACGCC	CGCTGA		ATAAAAAAIIND
					GCTGGCGCGCGGGAGGACGTCGAGAGCTGC	AGCGTT		NSLLSDDDDDD
					GCACCAGTGCGCGAGGGCCTCAAAATCCTC	CTACGC		DHDDNCHGEER
					GGCGCGCCCGTGGGCTCGCCCGCCTTCGTC	GCTGGC		GEGERNVTCPG
					GCCAAGTCGCTCGACGGCATCATCAAGCGC	AGCAAG		HYTATPFAADD
					GCCATCGGCACACTCGACCTCGTCGCTGAC	CCTCGG		TLDNSDEDNED
					GCCGAGCTACCGCTGCAGCACAAGCTGGTG	CCCTAG		NAHEDDDEDGK
					CTGCTACGGCAGTGCGTGGCCCAGATACCC	CTTGTA		DDNDDDVYNN
					ACGTTCTGGGCCCGCGCCGTGCCCGACGCA	GCGCCG		CNSSSSDGDEGG
					GGCCCGGCCCTCGCCGTCTGGGATACAGCG	CCGGTG		DDLDYEYSDQS
					CTCCTCAGGCGCACGGGCGCGCTGGTCGGA	GCCGCT		VTRSVDAATGE
					CTTGACGTGCGGGACGGGTCCCTGCAAGCC	CGCCAA		SPNPERPTTPTR
					GACATCGCGCGCCTGCCTGTCCGCCTGGGC	C (SEQ		ALLRADLWLPA
					GGTCTCGGCCTCCGTTCAATGAAGGACACG	ID NO:		TSTAVDVMVAA
					GCGCCCCGGGCCTTCGTGGCCTCGATCCTGT	1155)		ACRRSRAKAFD
					TCGCCGCCGCGCTCGCCAACACGCGCCGAT			RAVSRKAAKYG
					CCGAGCTCACGTGCAGCGCCAGTACGGCCC			PAVADGSIAKV
					GACGCCTCAGAGCCGCCCTGCCCGAGCTGG			VPFVVSPFGVLS
					CACGCACCGACGCGTGCAACGACGAAGCCG			RPAKAFLKRAM
					CCTGGCGGCGGTCCATCGCCAGGGGAGTCT			GDTTAAKQAKA

					TCCCCGACGTGGACAAGCTGGGCACCACAC AACTGCAGCGCGTCCTCCAGGGGATGGCGG ACTCCAAGTCCGCCCATCGGACCCGCCGCC AAGTGCCCTTCCTCTTCGCCGCCGTGTTCGA GGACGCCGCCACGCCGGGATCCGGTGCCTG GCTGGCCGCCATACCCTCCGACCCGACCCT CGTCTTGCCAGACGCCGAACTGGCCGAAGC CGTGCGCATTAAGTTGCTCACGACGACGGC CAATGCGGCCGGCGTCTGTCCGGCATGCCA CAAGACCGGCATCGACCCGTCCCACGCGTA CACGTGCGTTTCGCTATCCCATTTGCGCACA GCACGCCACGACGTGGTCGTTCGCCGAGTC GAGCTCGCCTGCAAGACCGAGAAGCCGGTC CGCGAACACGTGCTCGCCATCCCCCCCGTC GCGCCCACCGACAACAACAACAACGGCGAC GAGGACGGCAGCCCAGTCACCACCGCCGAC GACAACGCAGACGGCCACGCGGTCGCGACC AAACGTCGCCCCGAGACCCGCGCTTCAGCC AGAGCCGCCGCCGCCGCCGCCACCGCCGCC GCCGCCGCCGCAATCATCAACGACAACAGC CTCCTGAGCGACGACGACGACGACGACGAC CATGACGACAACTGCCACGGAGAGGAAAG AGGAGAGGGAGAAAGGAACGTCACGTGCC			RLRLAVAAVRG TARLSYAWGAC AALIVGGN (SEQ ID NO: 1400)

					CCGGCCACTACACCGCCACACCCTTCGCCG CGGACGACACGCTCGACAACAGCGACGAG GACAACGAGGACAACGCTCACGAAGACGA TGACGAAGACGGTAAAGATGACAACGACG ACGACGTCTACAACAACTGCAACAGCAGCA GCAGCGACGGGGATGAAGGCGGTGACGAC CTGGACTACGAGTACAGTGACCAGAGCGTC ACTCGAAGCGTCGACGCCGCGACGGGAGAG AGCCCCAACCCCGAGCGCCCTACCACGCCC ACCCGCGCACTACTACGCGCGGACCTGTGG CTACCCGCCACCTCCACCGCGGTGGACGTG ATGGTCGCGGCCGCCTGCCGTCGGTCACGC GCCAAGGCCTTCGACCGAGCCGTCAGCCGC AAGGCCGCGAAATACGGCCCTGCGGTAGCC GACGGCTCGATCGCCAAGGTGGTGCCGTTC GTCGTGTCGCCCTTTGGCGTACTCTCGAGGC CGGCCAAGGCCTTCCTCAAGCGCGCCATGG GCGACACGACGGCGGCCAAACAGGCCAAG GCGCGTCTGCGCCTCGCCGTGGCCGCCGTC CGAGGCACGGCCCGCCTCTCCTACGCCTGG GGCGCCTGCGCCGCCCTCATCGTCGGCGGC AACTAAGCCGCGCGACGAGGACGGCCAGG ACGACCAGGACGACGGCGACGGCGACCAC			

					CTAGCACCGCACGCGCCACGACATATTGTC GCGCGCTGTACAGGCGGCTAGGTCGAGCCC AGCCGACCGTTCTGAGCCTCAGTCGGCTTG AGCCCCCGGCTTCCCAAGGCCTACCGGGGC GGCTCTTTTTCGCCCTGGTTTTTGCCGGCCT GTTTTTTCTCTCCCCCTTTTCCCCCCTTTTCC ATTTGTACTTAGTTTTTCCTTCGGCCGCGGC AGCTTGTTGCCCGGCATAGTGTTAATATGTT TAAAAAACGTGTAAATAAATAACTGTTTAA CCCTAACCCTAACCCTAA (SEQ ID NO: 1032)			
CRE	Cre-1_FCy	.		Fragilariopsis cylindrus	ATCAATCTAATACTGAAGGCAATACCAAAC TCAACCCGAAATCAAAATCGTTAGAATCAA TATACGACCCCCGCTGCTGTACATGTCCAGC CGGATCTCGTTGTAAAGAAGATTGCAGCTG TAAAAAAGTTGGACTTCTTTGTTCTTCCTGT GAAGAAGTTGATTGTGGCTGTTCAAATTCTT TTCATAATAAAGAATTAATGGCTCCACTTCC TTGGAATGCTGCGACATCTTCACCACCGTCA CCTGTCCCATTAACTAACGATAAGAAAAAG GATTCTACTCTACCTACTGCAACATCAAAA AATTTATCCAAAAATAATAATAATAAAAAT AATAATACTAATAGGATTAATAATATTAAA AATAATGATAATACAAATGATGGTTCTAAT	ATCAAT CTAATA CTGAAG GCAATA CCAAAC TCAACC CGAAAT CAAAAT CGTTAG AATCAA TATACG ACCCCC GCTGCT	TAGC ACCA CCAT CTATT CATA TCCA CACA CTGA CCAC CTCC ACCT TCAC AACT	MAPLPWNAATS SPPSPVPLTNDK KKDSTLPTATSK NLSKNNNNKNN NTNRINNIKNND NTNDGSNKINL KLPPAAVKITNP YKNKKKNKKK NNAGKSNPKTN QNPNSSPLSDND DDDTDSSNITIN RRLKFGTDDLA PPNPPSNTNTIGT

					AAGATAAATTTGAAACTGCCCCCGGCAGCT	GTACAT	CCAC	ATAATAATATT
					GTTAAAATCACAAATCCTTATAAGAACAAA	GTCCAG	TCTC	TATAATATTAT
					AAGAAGAACAAGAAGAAGAATAACGCTGG	CCGGAT	AATT	NTTTTTTTTNNT
					AAAGTCGAACCCCAAAACAAACCAAAATCC	CTCGTT	CCCC	TGDNLASNINN
					AAATTCAAGTCCACTTTCGGATAATGATGA	GTAAAG	TGAC	NNNNNNSGSNN
					TGATGATACTGATAGCTCAAACATCACCAT	AAGATT	TACT	SNTNNINNTDG
					TAATAGACGACTAAAATTTGGTACTGATGA	GCAGCT	AAGA	NGSNNRPPPRV
					TTTAGCACCTCCAAACCCACCGTCAAATACT	GTAAAA	AATA	YTVDPRSDLPG
					AATACTATTGGTACTGCTACTGCTGCCACCG	AAGTTG	TTTCA	AEISAANKMLD
					CTGCCACTGCTACAACTACGGCAACCGCAG	GACTTC	TGGT	EVYGDHVHDNP
					CGACTGCTACTACTGCTACCAACACTACTAC	TTTGTT	GGTT	GSHLSGLISSSQ
					TACTACTACTACTACTAATAACACTACTGGA	CTTCCT	ACAT	DQLWQGYFRRL
					GACAATCTTGCAAGTAATATTAATAATAAT	GTGAAG	TGGA	IPHNQSLYDCPK
					AATAATAATAATAATAGTGGCAGTAACAAT	AAGTTG	GGTA	GKLGKDITNEYS
					AGCAACACCAATAATATTAACAATACCGAT	ATTGTG	TCCC	NLFEAIMNGKC
					GGCAATGGTAGTAATAATCGCCCACCTCCC	GCTGTT	AACA	NMEKLLVFPVV
					CGTGTTTACACAGTCGATCCACGAAGCGAC	CAAATT	CCAA	VLQRRHGVTKN
					CTTCCCGGTGCAGAAATCTCTGCTGCAAAT	CTTTTC	GCAC	ADVKRRLLSRL
					AAAATGTTAGATGAAGTATATGGAGACCAT	ATAATA	AAAA	TAWKEGKFKYL
					GTCCATGACAACCCCGGCTCCCATCTCAGC	AAGAAT	TGAA	VEDTHRDLIAK
					GGACTTATTAGCAGCTCTCAAGATCAACTCT	TA (SEQ	CCCA	QSKARGDTTPA
					GGCAGGGTTACTTTCGTCGCCTTATACCTCA	ID NO:	CTAA	HRAKVYSSKLM
					CAACCAATCTCTCTATGACTGCCCTAAGGG	1156)	CCCT	RGHLQSAVNYIT

					AAAACTGGGTAAGGATATAACGAATGAATA		CTCA	DREGGGILYPYD
					TTCAAACTTATTTGAGGCAATCATGAATGG		TCCT	VDEKSGHTVSR
					CAAGTGTAATATGGAGAAGCTACTCGTGTT		ATCC	VLQDKHPSMRD
					TCCAGTAGTGGTTCTACAACGGAGACACGG		ACGG	PGPTAMPAYES
					AGTGACTAAGAATGCCGACGTTAAACGCCG		GGAC	VPELPTLEITAD
					TCTTCTTAGCCGACTCACCGCTTGGAAGGA		CACC	TVEIVAGKLSGG
					AGGCAAATTCAAGTATCTTGTTGAAGACAC		TTTG	AGLSGVDSIQLK
					ACATCGAGATCTTATTGCCAAACAATCCAA		AGCA	HLLLHHGQASQ
					AGCAAGAGGAGATACAACCCCCGCGCACA		GAAC	RLRNVCAKFGR
					GAGCTAAAGTTTACTCAAGTAAACTCATGC		ACCA	WLANEHPPWAS
					GTGGACATCTCCAATCAGCCGTCAACTACA		TTCAT	YRAMLANRLIA
					TCACTGACCGCGAAGGAGGGGGCATCCTTT		ATTA	LDKMPGIRPVGI
					ATCCTTATGACGTCGATGAGAAATCAGGCC		CAAC	GDTWRRFFAKL
					ATACTGTATCAAGAGTGCTACAGGATAAGC		CTTTA	VLAVSMSYATD
					ATCCCAGCATGCGTGATCCTGGTCCCACAG		GCTA	CCGSDQLCAGL
					CCATGCCTGCCTACGAGTCCGTCCCGGAAC		GATT	RAGVDGAIHGL
					TTCCAACACTTGAAATTACAGCTGATACAG		AAGA	SAMWREMESEE
					TTGAGATAGTCGCTGGAAAGCTCAGTGGTG		TAAT	NTGFVLIDADN
					GTGCAGGTCTGAGTGGAGTTGATTCAATAC		TATTT	AFNEVSRINML
					AACTCAAGCACCTCCTCCTCCATCACGGTCA		AGTA	WTIRHEWPAGA
					AGCAAGCCAACGACTGCGCAATGTTTGTGC		CATA	RFAFNCYRHHS
					AAAATTTGGTAGATGGCTTGCCAACGAGCA		TTTTA	LLVVRNPGGKP
					CCCCCCCTGGGCCTCGTACCGTGCCATGCTA		TACT	FTFFSKEGVTQG

					GCAAATAGGCTTATTGCGCTAGACAAAATG CCCGGAATTCGACCAGTCGGTATAGGTGAT ACATGGCGTCGTTTCTTCGCCAAACTTGTTC TAGCAGTCTCTATGTCTTATGCTACTGACTG TTGTGGGTCAGACCAGCTCTGTGCCGGACT AAGAGCCGGAGTTGATGGTGCCATACATGG ACTATCGGCTATGTGGAGGGAGATGGAATC TGAGGAAAACACAGGTTTCGTACTTATTGA CGCAGACAATGCATTCAATGAGGTCTCACG CATCAACATGTTATGGACGATCCGCCACGA ATGGCCTGCTGGAGCTCGATTCGCCTTCAAC TGCTATCGGCACCACAGCCTACTAGTGGTA CGGAATCCAGGCGGGAAACCCTTCACTTTC TTTTCTAAAGAAGGTGTCACACAGGGCGAC CCATTTGCGATGATAGCATATGGTGTCGCTC TCCTACCACTCATCCGCAAACTGAAAGAAT TAAATGTATTATTAGTTCAATCTTGGTATGC AGATGATGCTAGCGCAGCTGGCAAATTTGA TGAAATACTACGCCTTTTTCAAGATTTATTA CGAATGGGACCTGATTTTGGGTACTTTCCTA ATGCATCTAAGAGTATCCTCATCACCCATCC CGACAATGTGGTTGCAGCTCACCACTTCTTC AACGAGACCCATGGCCTAGGTTTCAAGATC		ATTA AAAA AAAA AAAA AAA (SEQ ID NO: 1279)	DPFAMIAYGVA LLPLIRKLKELN VLLVQSWYADD ASAAGKFDEILR LFQDLLRMGPD FGYFPNASKSILI THPDNVVAAHH FFNETHGLGFKI STGSRFLGGFIG DTTSRDEYVSTK IADWIHGTKELA AVARLKYPHAA YTGITKCLQHK WSFTQRVIPGID DLFQPLEDELTN NLLPALFGDPPS TMDDKLRLLTA LPVKHAGLALP NPVTSSATNYK NSTLMSSHLLLA VQGKINFSLQDH RDTCQSSLSASR ELRQTENDSSLT

					AGCACAGGAAGTCGTTTCCTGGGTGGTTTC			NLLAALPPAAA
					ATTGGAGATACCACAAGTCGAGATGAATAC			GQPSTTRAIKRA
					GTATCAACAAAAATCGCCGACTGGATCCAC			GETGLWLTTIPN
					GGCACCAAGGAGCTAGCAGCAGTAGCAAG			HINGNILGCDEFI
					ATTGAAGTATCCACACGCAGCTTACACAGG			DAIRLRYQKVP
					CATTACCAAGTGTTTGCAGCACAAGTGGAG			HNLPAKCDGCG
					TTTTACTCAACGTGTTATTCCTGGCATTGAT			SAFDVGHALQC
					GACCTCTTCCAACCACTGGAGGATGAACTC			KSGGLIIRRHDE
					ACCAATAATTTGCTCCCCGCCCTATTTGGAG			LNLELASLAKM
					ACCCCCCATCCACTATGGATGACAAGCTCA			ALRESAIRAEPEI
					GACTTTTGACCGCTCTGCCAGTCAAACATGC			NPSASIMDSPTTI
					TGGGCTTGCTCTCCCGAATCCAGTTACCTCC			TAIDTNGDRGD
					TCCGCAACCAACTACAAGAATAGCACTCTT			LLIKGFWDNGM
					ATGAGTTCTCATCTTCTGTTGGCTGTTCAAG			DAIIDVRITDTD
					GCAAGATCAACTTCAGTTTACAGGACCACA			AKSYRTRDPKK
					GAGATACCTGTCAATCCTCTCTCTCCGCGTC			VLQSQEKEKKK
					CCGAGAGCTCCGACAAACCGAAAATGATTC			KYLDQCLLQRR
					TTCATTGACCAACCTCCTTGCAGCTCTCCCT			AFTPFVVSVDGL
					CCAGCTGCTGCAGGTCAACCAAGCACAACA			IGYEASNVLKQL
					AGAGCAATCAAGCGTGCTGGGGAAACCGGT			SKRLADKWNKP
					CTTTGGCTTACTACTATCCCTAATCACATCA			YSVTCGIVRSRIS
					ACGGTAACATTCTCGGATGTGACGAATTTA			IACARASNQCLR
					TTGATGCTATTCGATTGAGATACCAAAAAG			GSRIPFKTMSRQ

					TGCCACACAATCTCCCTGCCAAATGTGATG GCTGTGGCTCTGCATTTGATGTAGGGCACG CGCTCCAATGCAAATCCGGGGGCCTAATCA TTAGACGTCATGATGAACTCAATCTTGAGCT TGCATCTTTAGCAAAGATGGCCTTGAGAGA ATCTGCAATACGTGCTGAACCTGAAATCAA CCCCAGCGCCTCTATTATGGATTCTCCCACC ACCATCACAGCCATCGACACAAACGGAGAC CGAGGAGATTTGTTGATCAAGGGCTTTTGG GACAATGGAATGGACGCTATCATCGATGTC AGAATAACAGACACAGATGCCAAATCCTAT CGAACAAGAGACCCAAAAAAAGTCCTACA GTCACAAGAGAAGGAGAAAAAGAAGAAAT ACCTCGATCAATGTCTACTCCAACGTCGAG CCTTTACCCCTTTTGTTGTCTCTGTGGACGG CCTGATTGGTTACGAGGCCAGCAATGTGCT AAAGCAATTATCAAAACGTTTAGCAGATAA ATGGAATAAGCCTTATTCAGTTACATGTGG AATAGTCCGCTCACGTATCAGCATTGCATGT GCGCGAGCTTCCAATCAATGTCTGAGAGGT TCTCGAATACCATTCAAAACAATGAGCAGA CAAATTCAATGGGAGGACGGTGCAGGCGCC GGCCTCTATAGAATTGTCCGCTAGCACCAC			IQWEDGAGAGL YRIVR (SEQ ID NO: 1401)

					CATCTATTCATATCCACACACTGACCACCTC CACCTTCACAACTCCACTCTCAATTCCCCTG ACTACTAAGAAATATTTCATGGTGGTTACAT TGGAGGTATCCCAACACCAAGCACAAAATG AACCCACTAACCCTCTCATCCTATCCACGGG GACCACCTTTGAGCAGAACACCATTCATAT TACAACCTTTAGCTAGATTAAGATAATTATT TAGTACATATTTTATACTATTAAAAAAAAA AAAAAAA (SEQ ID NO: 1033)			
CRE	Cre-1_HM	.		Hydra vulgaris	TTTCTAATGTTACGTGATATGATATGGTTAG TTCATGGTTAGTTTATGTTTATGCTTAGTTT ATGGAAAATCGTTTATTTATGGCACAATATT GTTTGCTGTTTTTAAATTTATGTAACGTGTG CATTTGATGTATATTCTTGAACTTTTTAATC TGAATTTTTACTTGGTTTAATACGTTTATTA TATTCTTCGATTGAGCAATTTATCCTATCAA AGCAATTTATCCTTCGATTCGAGCAATTTAT CCTTCGATTCGAGCAATTTATCCTTCGATTG AGCAATTTATCCTATCAAAATTAGCATATAT ACTGCAATTTTCAAATAATCTACGAAATAA GTTCACTTACTGAAAATCATTAAGTAAAAG AAGAAAGGAAGAAAAAATAAAAATAAAAA GTAGTAAATCCTTTCATAACAATAATCATTC	TTTCTA ATGTTA CGTGAT ATGATA TGGTTA GTTCAT GGTTAG TTTATG TTTATG CTTAGT TTATGG AAAATC GTTTAT TTATGG	TAAC TTGT ATTTT TAAA TTGTT TTATT AGTT T (SEQ ID NO: 1280)	MNMVSICKRCD RSFTTLKGLNIH KGQCKIFVSNTN KQINNVVNNEL TTPNKNKVEINT ILNCDEISVEHY STNTPYLPKINIC ESIIDPNDYLWG HMPFSFLLNHV NTIYDEIVFYHK NLFKVPSGKGG KMFIEELTFWLK QFNNRTKLNGIA MKCFMIVPSLM

					TATTATTAAATTTAAAGGAATATTTTGGTTT	CACAAT		LQKPSIRSKAKE
					TGTACTAAATCATGCGTTCATATTTCACCGA	ATTGTT		HAECLVRRITL
					AGAAGGGGGCTGCTATATTTTTGTTTGAAGT	TGCTGT		WRNGNFSELMR
					TGTTTATCTTAAAACTTTAAACTTGTGTTCA	TTTTAA		EIRYIQSKINTSK
					ACCAACCGTAAACATTAGTTCGCTGTTCGCT	ATTTAT		KKRTFEDISRIFA
					CAAATTATCTACAATATAAAATTTATCAATC	GTAACG		KLMMEGKVAA
					TTTTTTCGTTACGGTAAACAATAAACAATAA	TGTGCA		ALKVLDRESSGI
					AATAACTATAGTTATTTTATTGTTTACCGCA	TTTGAT		LQCSESVLKELK
					TATTGTTTAACTATAGTTAAACAAAGTATTT	GTATAT		SKHPDETPVQD
					GTTTATGGAACATTACCAGTATCTCTTGTTA	TCTTGA		NCLLYGPLQNTP
					AGGTAAACAACAAAACATAGACGGCATCTC	ACTTTT		ECLFDSIDEISIF
					TTTTTAAGGTAATTAAGTATACGGCTAATAA	TAATCT		NSALQTKGSAG
					TAAAAATATACAGCTAATAATAAAATCTTC	GAATTT		PSGMDADLYRR
					AATGAACATGGTTTCTATATGCAAAAGATG	TTACTT		VLCSKCFGPSCK
					TGATCGTAGCTTTACTACCCTTAAGGGACTA	GGTTTA		TLREEIATFTKNI
					AATATTCATAAAGGTCAATGTAAGATCTTT	ATACGT		ATKSYQPDIVQP
					GTTTCCAATACAAATAAACAAATAAACAAT	TTATTA		YIACRLIPLDKN
					GTAGTTAACAATGAATTAACAACACCGAAT	TATTCT		PGIRPIGIGEVLR
					AAAAACAAGGTGGAAATTAATACGATATTA	TCGATT		RIVGKTISHHCQ
					AACTGCGATGAGATATCTGTAGAACACTAT	GAGCAA		KEIKEAAGPLQT
					TCAACCAACACACCTTACTTACCCAAAATA	TTTATC		CAGHGAGAEAA
					AATATTTGTGAATCTATTATAGATCCCAACG	CTATCA		IHAMQKIFHQED
					ACTATCTATGGGGTCATATGCCGTTTAGCTT	AAGCAA		TDGVLLIDARN

					CCTTCTCAACCATGTCAACACAATATACGAT	TTTATC		AFNCLNRSVAL
					GAAATAGTATTTTACCATAAAAACCTTTTTA	CTTCGA		HNIQITCPILAM
					AAGTGCCATCAGGAAAAGGTGGTAAAATGT	TTCGAG		YLVNTYRKPAK
					TTATAGAAGAACTGACCTTTTGGCTAAAAC	CAATTT		LFIYGGETIFSKE
					AGTTTAATAATCGAACCAAATTGAATGGAA	ATCCTT		GTTQGDPLAMP
					TAGCCATGAAATGTTTCATGATAGTCCCTTC	CGATTC		WYSLSTVTIINT
					CCTAATGTTACAGAAGCCCTCAATACGGTC	GAGCAA		LKLVIPDVKQV
					CAAAGCCAAAGAACATGCAGAATGTTTAGT	TTTATC		WLADDATAAG
					AAGACGAATTACATTATGGAGAAACGGGAA	CTTCGA		KLQSLKKWYKC
					CTTTAGTGAATTGATGCGGGAAATTAGATA	TTGAGC		LEDVGGLYGYY
					TATTCAGAGCAAAATTAACACCTCAAAAAA	AATTTA		VNQSKCWLIVK
					GAAAAGGACATTTGAGGATATCTCAAGGAT	TCCTAT		SDNQAEEAKLIF
					ATTCGCAAAACTAATGATGGAAGGTAAAGT	CAAAAT		GNSINITTQGKR
					TGCTGCCGCACTGAAGGTTTTAGATAGAGA	TAGCAT		HLGAALGSEAY
					GTCATCTGGCATCTTGCAATGCTCGGAAAG	ATATAC		KKVYCEDLVSK
					TGTATTGAAAGAATTGAAAAGTAAACACCC	TGCAAT		WSKELNNLCEIA
					AGACGAAACTCCTGTACAAGATAATTGTTT	TTTCAA		TTQPQAAYSAFI
					ACTATACGGCCCGTTACAAAACACTCCAGA	ATAATC		KGYRSKFTYFLR
					ATGTTTATTCGATTCAATTGATGAGATAAGT	TACGAA		TIEAFENFVTPV
					ATATTTAACTCAGCTTTACAGACTAAAGGA	ATAAGT		EKILSEKLLPVLF
					TCTGCAGGTCCTTCTGGAATGGATGCAGAT	TCACTT		GTDCSIIKENRD
					CTTTACCGTCGAGTCCTATGCTCAAAATGTT	ACTGAA		LLALNPSEGGLG
					TTGGACCCTCTTGTAAGACTCTACGAGAAG	AATCAT		ICNLITEAKEQH

					AAATAGCAACATTTACAAAAAATATTGCAA	TAAGTA		TASKKITNLHIK
					CAAAATCCTACCAACCGGATATAGTTCAAC	AAAGAA		SILDQSDVMKE
					CCTACATTGCATGTCGACTAATTCCCTTAGA	GAAAGG		KDDFGKTFSEIK
					CAAAAATCCCGGGATTCGCCCCATAGGAAT	AAGAAA		TKTNMDKSKKK
					TGGGGAAGTGTTACGTAGGATTGTAGGTAA	AAATAA		KEEVKKIHAGLP
					AACCATTAGCCACCATTGTCAAAAAGAAAT	AAATAA		ENLKLLVEQAC
					CAAAGAGGCAGCTGGACCACTACAAACTTG	AAAGTA		DKGASSWLNTL
					CGCAGGACACGGTGCAGGAGCAGAAGCTG	GTAAAT		PIKEQHLDLNKE
					CAATACATGCTATGCAAAAGATATTTCATC	CCTTTC		EFKDALRLRYN
					AGGAAGATACAGATGGTGTTTTGTTAATCG	ATAACA		VPLANLPSYCAC
					ATGCTAGGAACGCGTTTAACTGCCTAAACC	ATAATC		GEKFDELHAMS
					GTTCTGTTGCACTACATAATATACAGATAAC	ATTCTA		CKKGGFVCNRH
					TTGCCCAATCTTAGCTATGTATTTAGTCAAC	TTATTA		DNIRDLLTVCLN
					ACTTACCGTAAACCGGCAAAATTATTCATCT	AATTTA		KVCTDVQAEPH
					ACGGTGGAGAAACTATTTTTTCGAAAGAAG	AAGGAA		LIPLTNEKFNFK
					GCACAACGCAGGGCGATCCCCTCGCCATGC	TATTTT		TANTNDEARLDI
					CATGGTACTCACTTAGCACTGTGACAATCAT	GGTTTT		KAKGFWRKGET
					AAATACATTGAAACTAGTAATTCCTGATGT	GTACTA		AFFDVRVTHVN
					AAAACAAGTATGGTTAGCCGATGATGCTAC	AATCAT		SKSSKKQPTKHI
					CGCTGCAGGAAAATTACAGTCTTTAAAAAA	GCGTTC		FRRHEDAKKRE
					GTGGTATAAATGCCTAGAGGATGTCGGTGG	ATATTT		YLERVLEVEHG
					TTTGTATGGTTATTATGTAAATCAGTCAAAA	CACCGA		TFTPLIFGTNGG
					TGCTGGCTAATAGTAAAATCTGATAACCAA	AGAAGG		FGDECKRFTALL

					GCTGAAGAAGCTAAACTTATATTTGGCAAC TCCATAAATATAACTACTCAGGGAAAAAGG CACTTAGGAGCTGCACTTGGTTCGGAAGCA TACAAAAAAGTGTATTGCGAGGATTTAGTA AGTAAATGGTCTAAAGAACTTAACAATCTC TGCGAAATCGCCACCACGCAACCACAAGCT GCTTATTCAGCTTTTATTAAAGGGTACAGAT CTAAATTCACTTACTTCTTACGCACAATTGA AGCTTTTGAAAATTTCGTAACACCAGTGGA AAAAATTTTATCAGAAAAATTATTACCTGT ATTGTTTGGAACTGATTGTTCTATAATCAAA GAAAATAGGGATTTATTGGCGCTAAATCCA TCGGAAGGAGGACTTGGAATTTGTAACTTA ATAACTGAGGCCAAGGAACAGCATACTGCC TCTAAGAAAATAACTAACTTGCACATAAAA TCAATACTCGATCAGTCAGATGTTATGAAA GAAAAAGATGATTTCGGGAAAACATTTTCA GAAATAAAAACAAAAACAAATATGGATAA ATCTAAAAAAAAAAAAGAAGAGGTTAAAA AAATACATGCAGGACTTCCAGAAAACCTTA AACTTCTGGTTGAACAGGCCTGTGACAAAG GTGCCAGCAGCTGGTTAAACACCTTACCAA TTAAAGAACAACATCTAGATCTGAATAAGG	GGGCTG CTATAT TTTTGTT TGAAGT TGTTTA TCTTAA AACTTT AAACTT GTGTTC AACCAA CCGTAA ACATTA GTTCGC TGTTCG CTCAAA TTATCT ACAATA TAAAAT TTATCA ATCTTT TTTCGT TACGGT AAACAA		AQKLSLKMGER YGAVINWLRTR LSMEITRASLLC LRGSRTPFRHYN TDDVGLENVQC GLI (SEQ ID NO: 1402)

					AAGAGTTTAAGGACGCACTTAGATTGAGAT	TAAACA		
					ATAATGTGCCACTTGCCAATTTACCATCCTA	ATAAAA		
					CTGTGCTTGTGGAGAAAAATTTGACGAGCT	TAACTA		
					ACACGCAATGTCATGCAAAAAAGGTGGCTT	TAGTTA		
					TGTTTGTAACAGACATGATAACATCAGAGA	TTTTATT		
					TTTATTAACTGTTTGCCTAAATAAAGTTTGT	GTTTAC		
					ACTGATGTTCAAGCGGAGCCGCATTTAATT	CGCATA		
					CCATTGACAAATGAAAAATTTAATTTCAAA	TTGTTT		
					ACTGCCAATACCAACGACGAAGCTAGATTG	AACTAT		
					GATATAAAAGCAAAAGGGTTTTGGAGAAAA	AGTTAA		
					GGAGAAACTGCATTTTTTGATGTTAGAGTA	ACAAAG		
					ACGCACGTAAACTCCAAATCCTCCAAAAAA	TATTTG		
					CAACCAACAAAACACATATTCCGTAGGCAT	TTTATG		
					GAAGATGCAAAAAAACGTGAGTATTTAGAA	GAACAT		
					CGAGTTCTAGAGGTTGAACACGGGACATTT	TACCAG		
					ACCCCATTAATTTTTGGTACGAATGGTGGGT	TATCTC		
					TTGGAGACGAATGCAAACGCTTCACGGCAC	TTGTTA		
					TACTCGCACAAAAACTGTCCTTAAAAATGG	AGGTAA		
					GTGAGCGGTACGGAGCTGTTATAAATTGGC	ACAACA		
					TAAGGACACGTCTTTCCATGGAGATTACTA	AAACAT		
					GAGCCTCCCTACTCTGCTTAAGAGGGTCAC	AGACGG		
					GAACCCCATTTAGGCATTATAACACTGACG	CATCTC		
					ATGTTGGCCTGGAAAATGTGCAATGTGGAC	TTTTTA		

					TTATTTAACTTGTATTTTTAAATTGTTTTATTAGTTT (SEQ ID NO: 1034)	AGGTAATTAAGTATACGGCTAATAATAAAAATATACAGCTAATAATAAATCTTCA (SEQ ID NO: 1157)		
CRE	CRE-1_LSa	.		Lactuca sativa	ACATTAAATTAGAGAGGTTGATGTTTCAATGGAAGAAGATGAAATTCCAAGAAGCTATTTTTGTTGCCCACCAAGTGTTTGATAAAATGTCCAAACTAATTTTTCTCTTGTTGCAGCTTTATTGTTCAAGATAATGTAGTTTGCTTAGTTTGAGCGTTCCTTGTGCACACCAACAGTGTGTTGGTGTGCCATTTCCTTTCCTTCCTTTTTAACTATTGCTTCATAGCTTAAGCTTCATCTCGAGGCTTGTTCTCTTGTATGGCTTCTTCTTCTACAAGTTCGAGTGATATTTGTCTGTGCCCGTTCAGAAGCTTCCATTGTTGCCCAAATGGTGAAGTGG	ACATTAAATTAGAGAGGTTGATGTTTCAATGGAAGAAGATGAAATTCCAAGAAGCTATTTTTGTTG	TGAACTATATTTTATATTAAAAAA (SEQ ID NO: 1281)	MASSSTSSSDICLCPFRSFHCCPNGEVGSKGIXRMISHIKRHHLLTEDRKCVLREALSSDVGLFMAVEETLKAFGQWMCGKCMTLHALSRYCHHPDGRVXFVTGADGSSRYIVGILKPSTKESVTNA

					GAAGTAAGGGGATTGKCCGTATGATTTCAC	CCCACC		LGGLVFDVGLL
					ACATCAAAAGGCATCATCTACTTACTGAAG	AAGTGT		DRVFKEPITTVK
					ATCGTAAATGTGTTTTACGTGAAGCTCTTTC	TTGATA		SIPHSCRLAFSQ
					TAGTGATGTTGGTTTATTTATGGCGGTGGAA	AAATGT		ALKTALYKVIA
					GAAACTTTGAAGGCCTTTGGTCAATGGATG	CCAAAC		QPGSVDAWICL
					TGTGGGAAGTGTATGACTTTGCATGCTCTTA	TAATTT		LLLPRCTLQVFR
					GCCGTTATTGTCATCACCCGGATGGTCGTGT	TTCTCTT		PKNRQECRSGN
					GAGKTTTGTTACAGGGGCTGACGGCTCGAG	GTTGCA		RKSLQQSSILKS
					TCGTTACATTGTCGGTATTCTAAAGCCGTCT	GCTTTA		LDTWGKEDGIR
					ACTAAAGAGTCGGTGACAAATGCTCTTGGA	TTGTTC		KLVQNMLDNPE
					GGTTTGGTTTTTGATGTTGGGCTCCTTGATC	AAGATA		VGAMGQGGGIL
					GTGTTTTTAAAGAGCCTATCACTACTGTCAA	ATGTAG		QKESTSSNTNIR
					GAGTATCCCCCATAGTTGTCGCCTTGCTTTC	TTTGCT		QCLRKVADGHF
					TCTCAGGCTTTGAAAACTGCTCTTTACAAGG	TAGTTT		TAAVKVLCSSG
					TGATTGCCCAACCTGGCTCGGTTGATGCATG	GAGCGT		VAPYNGDTIKA
					GATTTGTTTGTTACTTCTTCCTCGCTGCACA	TCCTTG		LEDKHPFRPPPS
					CTGCAGGTGTTTAGGCCCAAAAATAGACAA	TGCACA		MPSPIISEPPLVA
					GAATGTAGGTCTGGGAATAGAAAATCCTTA	CCAACA		DFDCVFGCIKSF
					CAACAAAGCTCCATCCTGAAGTCCTTGGAT	GTGTGT		PKGTSCGRDGL
					ACATGGGGGAAAGAGGATGGTATCAGGAA	TGGTGT		RAQHXLDALCG
					GTTAGTTCAAAATATGTTAGACAATCCCGA	GCCATT		EGSAIATDLIRAI
					GGTTGGGGCCATGGGACAGGGTGGAGGCAT	TCCTTT		TSVVNLWLAGR
					CCTTCAGAAGGAGTCTACATCAAGTAACAC	CCTTCC		CPTILAEFVASA

					CAACATCAGGCAGTGTCTCCGTAAGGTTGC	TTTTTA		PLTPLIKPDNGIR
					AGATGGKCATTTTACCGCAGCAGTGAAAGT	ACTATT		PIAVGTIWRRLV
					GTTATGCTCATCGGGTGTTGCGCCATATAAT	GCTTCA		SKVAMKGVGKE
					GGTGATACTATTAAAGCTTTGGAGGACAAA	TAGCTT		MAKYLNDFQFG
					CACCCTTTCAGGCCACCCCCATCCATGCCGA	AAGCTT		VGVSGGAEVVL
					GCCCCATAATTTCTGAACCTCCCCTTGTAGC	CATCTC		HSANRVLSEHH
					AGACTTTGACTGTGTATTTGGTTGCATCAAA	GAGGCT		ADGSLAMLTVD
					TCCTTCCCTAAAGGAACTTCWTGCGGGAGA	TGTTCT		FSNAFNLVDRSA
					GATGGCTTGAGGGCTCAACACWTACTAGAT	CTTGT		LLHEVKRMCPSI
					GCCCTTTGTGGAGAAGGGTCTGCTATAGCC	(SEQ ID		SLWVNFLYGQA
					ACAGATCTCATACGTGCTATCACTTCAGTGG	NO: 1158)		ARLYIGDQHIWS
					TTAATTTATGGTTAGCGGGAAGATGTCCGA			ATGVQQGDPLG
					CCATTTTGGCAGAGTTTGTTGCATCCGCTCC			PLLFALVLHPLV
					TCTCACGCCTCTGATTAAACCTGACAACGG			HKIRDNCKLLLH
					GATCCGTCCAATTGCAGTAGGCACTATATG			AWYLDDGTVIG
					GAGACGTCTGGTTTCCAAGGTTGCCATGAA			DSEEVARVLNII
					AGGTGTGGGTAAAGAAATGGCCAAGTACCT			RVNGPGLGLEL
					TAATGATTTTCAGTTCGGGGTTGGTGTGTCC			NIKKTEIFWPSC
					GGGGGTGCTGAGGTTGTGTTACACAGTGCC			DGRKLRADLFP
					AATAGGGTGTTGAGTGAACACCACGCTGAT			TDIGRPSLGVKL
					GGGTCTCTTGCAATGCTGACAGTGGATTTCT			LGGAVSRDAGFI
					CGAATGCCTTTAACCTGGTGGATAGATCAG			SGLAMKRAVNA
					CCTTGCTCCACGAGGTTAAGAGGATGTGCC			VDLMGLLPQLC

					CTTCTATTTCTTTGTGGGTGAATTTCTTGTAC			DPQSELLLLRSC
					GGGCAAGCAGCGAGACTTTATATAGGAGAC			MGIAKLFFGLRT
					CAACATATATGGTCTGCCACTGGGGTGCAG			CQPVHIEEAALF
					CAAGGCGACCCCTTGGGCCCTCTTCTTTTTG			FDKGLRRSIEDM
					CCCTCGTTTTGCACCCGCTTGTGCACAAGAT			VVCGGPFFGDIQ
					TAGAGACAATTGTAAGCTCCTTCTCCATGCT			WRLASLPIRFGG
					TGGTATCTAGATGATGGGACTGTCATTGGG			LGLYSAYEVSSY
					GATTCAGAGGAGGTGGCTAGAGTGTTGAAC			AFVASRAQSWA
					ATTATTCGGGTGAATGGTCCAGGCTTGGGT			LQDHILRDSGIC
					CTTGAGTTGAACATCAAGAAAACGGAGATT			GMDSDYLCAMT
					TTTTGGCCCTCCTGTGATGGTAGGAAGCTTC			RLRDTIPGFDCS
					GTGCCGATTTATTCCCAACGGATATAGGGA			GFTNKDTPPKSQ
					GACCTTCTTTGGGGGTGAAGCTCCTTGGGG			KALACALFSKIV
					GGGCTGTTAGCAGAGACGCAGGGTTTATTA			KDMEVDFDMT
					GCGGGCTGGCCATGAAGAGAGCGGTCAATG			VRQKAVFECLR
					CTGTTGATTTGATGGGTCTTCTTCCACAACT			APHAQDFLLTIPI
					ATGTGACCCGCAGAGTGAGCTCCTTTTGCTT			DGLGQHMSPVE
					CGATCATGTATGGGCATTGCAAAACTTTTCT			YRTILRYRLMIP
					TTGGTTTAAGGACATGCCAGCCGGTGCACA			LFPIDEICPVCRK
					TAGAAGAGGCAGCTTTGTTCTTTGACAAAG			ACLDTFGEHAV
					GATTGCGCAGGTCTATCGAGGATATGGTGG			HCRELPGFKYR
					TATGTGGAGGCCCCTTCTTTGGAGACATCCA			HDVVRDVLFDA
					GTGGCGTCTGGCTTCCTTACCTATTCGTTTC			CRRAGISAKKEA

					GGTGGTTTGGGTTTGTACTCGGCATACGAG GTTTCCTCCTACGCATTTGTAGCCTCGAGGG CCCAATCTTGGGCATTACAAGACCACATCTT ACGTGACAGTGGCATATGTGGTATGGACTC TGATTACCTATGTGCTATGACTCGTCTTCGC GATACGATTCCGGGATTCGACTGTAGCGGT TTCACTAATAAGGACACCCCCCCTAAATCC CAAAAAGCATTGGCGTGTGCCCTTTTTAGC AAAATCGTCAAAGATATGGAAGTCGACTTC GACATGACTGTTAGACAGAAAGCAGTTTTT GAGTGTCTTCGGGCACCTCATGCTCAGGATT TTCTGCTAACTATCCCTATTGATGGCCTTGG CCAGCATATGTCTCCTGTGGAGTACCGAAC TATCCTTCGTTACCGCCTCATGATTCCTCTA TTCCCAATTGACGAGATATGCCCAGTTTGCC GCAAGGCATGTTTGGATACCTTTGGGGAAC ATGCGGTTCATTGTAGAGAGCTCCCTGGTTT CAAGTACAGACATGATGTGGTTAGGGATGT TCTCTTTGATGCTTGTCGGCGTGCTGGTATT TCTGCGAAGAAAGAAGCGCCAGTGAACTTT TTGACGGACCCGCAGGATGGAAGATCCACA CTTAGACCGGCTGACATTTTGGTCTTTGGAT GGGTAGGAGGGAAGCACGCGTGTGTGGATC			PVNFLTDPQDG RSTLRPADILVF GWVGGKHACV DLTGVSPLVGLR SGGFTAGHAAL KAAACKVAKHE NACIENQHVFVP FAFDTFGFLAPE AVELLNRVQRV MHSNVISPRSTD VVFKRISFAIQK GLAAQLVARLP SIDMY (SEQ ID NO: 1403)

					TTACTGGGGTCTCTCCTCTCGTCGGTTTGAG GAGCGGGGGTTTCACAGCAGGGCATGCCGC TTTGAAAGCCGCTGCGTGCAAAGTGGCAAA GCACGAGAATGCATGTATAGAAAATCAACA TGTGTTTGTACCTTTTGCATTTGATACATTT GGTTTTCTCGCACCAGAGGCGGTGGAGCTC CTCAACAGAGTCCAACGGGTCATGCATTCT AATGTCATATCTCCTAGATCCACAGATGTTG TTTTCAAAAGAATTAGTTTTGCCATCCAGAA AGGGCTAGCGGCGCAGCTTGTTGCCCGTTT GCCTTCCATCGATATGTATTGAACTATATTT TATATATTAAAAAAA (SEQ ID NO: 1035)			
CRE	Cre-1_MB	.		Monosiga brevicollis	CATCTTGGCGTGAACCACGTTGTCAGACAA AATCTGCAACCCCGCTCTTTGCGGCCCGCGT TTTGGCGGCGCCCTCGCTCCCACCGTGTCCG CTCGCTTGCTCGCTTGCTTGCCCCGCGGACA TGGCCACTGAGTCCGGCGGCGAGGATTCTT GGACCCAGGTCCGCGGTGCTAAACGCCCGA GTGCCGAATCACCTCCAAGCAACACCACCA CCTCGCCTTCCCAAACTCATCGTTCTGCAAA ACACACAAAACATGGCAGCGCTCGCCACGA CCGTAACCATGTTTTCCCTGACCCCATGACC ACCCCGCTTCGCCCTCACGCCCGCCACTCTG	CATCTT GGCGTG AACCAC GTTGTC AGACAA AATCTG CAACCC CGCTCT TTGCGG CCCGCG TTTTGG	TAGG TAGG CACC GTCT CGGG GGTC CCTCT GTGG GGAT CCCT GTGT	MATESGGEDSW TQVRGAKRPSA ESPPSNTTTSPSQ THRSAKHTKHG SARHDRNHVFP DPMTTPLRPHA RHSVPTARASSH VPSTSPAAGATE SSARAVVPAAEP VTRTSNGGGEQ HPIIGNTSNASPR

					TCCCTACCGCCCGTGCCTCGTCTCATGTGCC	CGGCGC	GCAC	TPRTPSSPRSFA
					CTCGACGTCCCCCGCTGCCGGTGCGACCGA	CCTCGC	CTGT	QVAAAMPAAAT
					GTCTTCGGCACGTGCCGTCGTGCCCGCGGC	TCCCAC	CGCT	ATSSAPMTEDLS
					CGAGCCAGTGACCAGGACGTCAAACGGCGG	CGTGTC	CCCT	ASVPSEPNGSGE
					CGGGGAGCAACATCCCATCATCGGAAACAC	CGCTCG	AGGT	QQPSPESTGQTH
					TTCCAATGCTTCTCCCCGCACCCCTCGCACG	CTTGCT	GGTT	HSIPNTPSDFLT
					CCATCCAGCCCTCGCTCCTTTGCTCAAGTTG	CGCTTG	CCTC	MSSDESDSPPRS
					CTGCGGCAATGCCTGCCGCCGCCACTGCCA	CTTGCC	GTTG	TALRAPTPIAPP
					CATCTTCGGCCCCTATGACCGAGGATTTGTC	CCGCGG	TGTCT	AHDGDGDTNGS
					AGCATCGGTGCCCTCTGAGCCAAATGGCAG	AC (SEQ	TTTG	ATPEPLVQSPTP
					CGGGGAGCAACAACCCTCGCCCGAGTCCAC	ID NO:	ATGG	AQMVLPYPSGT
					AGGGCAGACACATCATTCGATTCCTAACAC	1159)	CTTG	QQTHSDPSPPSA
					ACCATCGGATTTTTTGACCATGTCTTCGGAT		ACTT	SPPATTILPAAIS
					GAAAGCGACTCCCCTCCTCGCTCCACCGCA		GTAT	HPVEHSEHANS
					CTCCGCGCGCCCACCCCTATCGCCCCTCCCG		TTTTG	APLGEVSESETH
					CGCATGATGGTGACGGTGACACAAACGGCA		TTTTA	NTAGEHSESEQ
					GTGCCACGCCTGAGCCATTGGTGCAATCAC		ATTTT	DVLLSDPAPPIA
					CTACACCCGCTCAAATGGTGCTGCCATATCC		GCTTT	ANVLDAQRKVL
					ATCGGGTACACAACAAACCCATTCCGATCC		AATT	LKTSGHRQLLA
					CTCTCCGCCCTCTGCTTCACCCCCTGCCACT		TTTGC	CPFGLCKCKGPR
					ACCATTTTGCCCGCTGCCATTTCACATCCTG		TGTA	LDRKAWVNHV
					TCGAACACAGTGAGCATGCAAACTCAGCCC		TTTGT	LREHPYDEQAT
					CACTTGGCGAAGTCAGTGAGAGTGAAACAC		GTGG	DLVKQVMEAKL

					ACAATACAGCGGGCGAACACAGTGAGAGT		TATTT	VAQCNKCHLFF
					GAGCAAGATGTTCTTCTCAGCGATCCCGCTC		TTGCT	EAAGISQHRSRC
					CGCCCATCGCTGCCAACGTGCTGGATGCCC		GAAT	GANLKRATEAL
					AGCGCAAGGTCCTGCTGAAGACATCTGGCC		TTTTG	FHAAGHDLLEI
					ACAGGCAACTCCTCGCCTGCCCATTTGGGCT		TAAG	MRGAWPQQCV
					TTGCAAATGCAAGGGGCCCCGCCTTGACCG		GTCC	GSRISVCELLKL
					CAAAGCCTGGGTCAATCATGTACTACGCGA		TTTGT	ARHPLMQRSRY
					GCACCCCTACGACGAGCAAGCCACTGATCT		ATGA	PSNATETKLMA
					GGTCAAGCAGGTGATGGAGGCCAAATTGGT		TGTCT	ATLSQLYWSAV
					CGCCCAGTGCAACAAGTGTCACCTCTTCTTC		TTGTC	HSDYTAEEREM
					GAAGCTGCTGGTATCAGTCAACACCGCTCC		TTCTG	CWALILALPSML
					CGATGTGGTGCCAATCTGAAGCGAGCGACC		TGGT	LSAPSTALSTIDL
					GAGGCGTTGTTTCATGCGGCTGGACACGAC		CGGT	RNMFHDRLRWL
					CTGCTTGAGATTATGCGTGGCGCTTGGCCCC		TGTTT	VTGQLGRVVDA
					AACAGTGTGTAGGGTCTCGCATCAGTGTCT		TCCTC	MRKAVARKQSR
					GCGAGCTGCTCAAGCTCGCCCGGCATCCAC		AATC	RGQLNAGAGAH
					TGATGCAGCGCAGCCGCTACCCATCCAACG		CGAC	PNDAVDQSLRS
					CCACCGAGACCAAGCTGATGGCTGCCACCC		GTTG	LVRDPDLADEA
					TGAGCCAGCTGTATTGGTCTGCCGTCCACTC		TGTCT	WANHVTNRLNR
					GGACTATACCGCTGAAGAGCGAGAGATGTG		CGTT	GQIAKAFDADK
					CTGGGCTTTGATTTTGGCCTTGCCTAGCATG		GGAT	ARAVIGNSEVQ
					TTGTTGTCTGCTCCCTCGACCGCACTGTCTA		GTGA	AVRDLLVPPGLT
					CGATTGACCTGCGCAATATGTTTCACGATCG		GCGT	PYIASTPASTSTL

					TCTCCGTTGGCTTGTGACGGGGCAACTAGG		GCCG	APATAVSSPTVS
					TCGGGTCGTGGACGCCATGCGCAAGGCAGT		TGGT	FTKGELPKALA
					CGCACGCAAGCAGAGCCGTCGAGGACAGCT		GTTCT	ATKGVTDPYGW
					GAACGCCGGCGCGGGCGCCCACCCGAACGA		TTGT	SGELLASIYRIKE
					CGCAGTCGACCAGAGCCTCAGGTCGCTCGT		GTTT	HFSQVLGPRQGS
					CCGCGACCCGGACCTGGCGGACGAAGCCTG		GTGC	TSDPTAPSDGDA
					GGCAAACCACGTCACGAACCGTCTGAACCG		TGTG	PQGPTTATGGPQ
					AGGTCAGATTGCCAAAGCATTTGATGCCGA		ATGG	VALNKIFHHIAN
					CAAGGCTCGTGCCGTGATTGGTAATTCTGA		CTTGT	NTVPESIRHALC
					GGTTCAGGCCGTGCGCGACCTCTTGGTACC		AGTT	SINYTILEKANG
					GCCCGGGCTGACCCCGTACATTGCTTCGAC		GTGA	KFRPVGTDSIFN
					ACCCGCCTCCACGTCTACACTGGCACCAGC		TGTG	KVVNRALLEQQ
					CACGGCTGTGAGCTCCCCAACCGTGTCCTTC		TGAC	QPHIAHLLQASP
					ACCAAGGGTGAGCTCCCCAAGGCGTTGGCG		TGCC	ELAVGVKDGIS
					GCCACCAAGGGTGTCACCGACCCCTATGGT		TTTTT	AAVGMAFGELQ
					TGGTCTGGTGAGCTTCTTGCCTCCATCTACC		GGGT	ACESTPGWTML
					GCATCAAGGAGCACTTCAGTCAAGTCTTGG		GTCTT	SLDFKSAFNYTD
					GCCCACGCCAGGGTTCTACCAGCGACCCGA		GTGT	RARLHEIVADK
					CTGCTCCTTCTGATGGAGACGCGCCTCAGG		TTGA	VPGLLRAFERH
					GCCCCACCACCGCCACTGGAGGTCCTCAGG		AATG	YARPTTHCIVDK
					TTGCCTTGAACAAGATCTTTCACCACATTGC		GCCG	FFKVIDIDVGQG
					CAACAACACCGTGCCCGAGTCGATTCGACA		TATCT	IVQGNELSPFFF
					TGCCCTTTGCTCCATCAACTACACTATCCTG		CTGG	ALYSCEVLGLL

					GAGAAGGCCAATGGCAAGTTTCGACCCGTG		TTAT	DATTDYRCKVI
					GGCACGGATTCCATCTTCAACAAGGTTGTC		ACTT	KYLDDIVLMGP
					AACCGCGCTCTGCTCGAGCAACAGCAGCCC		GGTC	AEDVAADVEIV
					CATATTGCCCACTTGCTACAGGCCAGTCCA		GTTTT	KARAESAGLHL
					GAGCTGGCCGTCGGAGTCAAGGACGGCATT		GTAC	QPSKSRFYMPR
					TCAGCAGCGGTTGGCATGGCCTTTGGTGAG		GATT	HHSASITAIKSV
					CTTCAAGCCTGTGAGTCTACCCCGGGCTGG		TTTGT	LPDAVRETANT
					ACCATGCTCTCCCTCGATTTCAAGAGTGCCT		TTCTA	GMTVLGTPIGRR
					TCAACTACACCGACCGAGCACGGCTGCACG		TGTG	EWMKKQLNDK
					AGATTGTGGCCGACAAGGTCCCTGGCCTCT		CGTG	AKHIAGKLNDM
					TGCGCGCCTTTGAACGACACTATGCTCGAC		ATTCT	LTTGVSLQALLT
					CCACTACCCACTGCATTGTCGACAAGTTCTT		TCGC	AMQYVPSLINH
					CAAGGTGATTGACATTGATGTTGGCCAAGG		GCTT	LYTLPPSLTSGL
					CATTGTGCAGGGCAACGAGCTATCGCCCTT		GTAC	SELLNRACKDTF
					CTTCTTTGCCCTGTACTCCTGTGAGGTCCTG		TTCTT	VKAFFAKVNLS
					GGTCTCCTCGACGCCACCACTGACTACCGCT		GGCA	APAGAEGHDVT
					GCAAGGTCATCAAGTACCTCGACGACATTG		TGAT	LEQLLEARLFTR
					TACTGATGGGTCCCGCGGAGGACGTGGCGG		AGAA	ANTGGFGLHDL
					CCGACGTGGAGATTGTCAAGGCTCGTGCAG		GCCA	VERGPVAYVCN
					AGTCTGCTGGCCTTCATCTGCAGCCCAGTAA		ATGA	MAKLATRYPRV
					GAGCCGCTTCTACATGCCTCGCCACCATTCG		ATGT	YDRLLEDASRA
					GCTTCCATCACTGCTATCAAGTCTGTATTGC		GTCTT	ADFEAHVQRAG
					CAGATGCCGTGCGCGAGACGGCCAACACGG		GTTCT	FQMATVKDAAT

					GCATGACGGTCTTGGGAACGCCGATTGGCC		CTTGT	QRPAEIIALRSK
					GTCGCGAGTGGATGAAGAAGCAGCTGAACG		GTTG	AALDDLMAKCA
					ACAAGGCAAAGCACATTGCTGGCAAGCTCA		TTTTG	LDLQQAYLASR
					ATGACATGCTGACGACCGGTGTCTCGCTTC		CGTG	EWGVSTVLTMR
					AGGCCCTCCTCACGGCCATGCAGTACGTGC		CCGT	GRDKLRRLSDT
					CTAGCCTCATCAACCACCTCTACACGCTGCC		CGTG	TFAIAVVSMMG
					CCCAAGTCTCACGTCGGGCTTGTCCGAGCTC		ATTTT	FGLHELINVKPT
					TTGAACCGTGCTTGCAAGGACACCTTTGTCA		GATG	DKCPLCSSKTPQ
					AAGCCTTTTTTGCCAAGGTAAACCTGTCTGC		TCGG	PRLTREHLLTCR
					ACCGGCTGGAGCTGAAGGTCATGACGTTAC		GGTT	PIKRHNALRDE
					GCTGGAACAGCTCCTTGAGGCTCGCCTCTTC		GCAC	MGRLLRYATLS
					ACACGGGCCAACACCGGGGGCTTTGGCCTG		AGCT	HVWVEKSGYN
					CACGACTTGGTTGAGCGCGGTCCGGTGGCT		TTGCT	ANGQSCRIDLHC
					TATGTCTGCAACATGGCCAAGCTGGCCACT		TTCA	RNPFPGGALGPA
					CGCTACCCTCGGGTCTACGATCGACTTTTGG		GCTC	LPDLGIDVTVRT
					AGGATGCATCGAGGGCTGCCGACTTTGAGG		TGAG	AQPPTTSQACIK
					CCCACGTGCAGCGAGCTGGCTTCCAGATGG		GTTC	VGAALRRAEKE
					CCACGGTCAAGGACGCGGCGACCCAGCGAC		AAAC	KRDYYTGFNHG
					CAGCTGAGATCATTGCCCTCCGCTCCAAGG		ACCT	KTLIVPAAMTTT
					CGGCACTGGACGACCTGATGGCCAAGTGCG		AATT	GGFASSFVDLLG
					CGCTGGACCTGCAGCAGGCATATCTGGCCT		TA	QLARCAEARGV
					CACGCGAGTGGGGCGTCAGCACTGTCTTGA		(SEQ	YQPGLDEAFVP
					CCATGCGGGGTCGGGACAAGTTGCGTCGCT			RWKGRFAALVH

					TGAGCGACACGACCTTTGCCATTGCGGTCG TGTCCATGATGGGTTTTGGCCTCCATGAACT CATCAACGTCAAGCCGACGGACAAGTGCCC GCTCTGCAGCAGCAAGACACCTCAGCCGCG ACTGACCCGCGAGCACCTGCTGACCTGCCG TCCCATCAAGCGTCACAACGCCCTTCGCGA CGAGATGGGCCGCCTGCTCAGGTACGCCAC CCTCTCCCATGTCTGGGTGGAAAAGTCTGG CTACAACGCCAACGGTCAGAGCTGCCGCAT CGACCTGCACTGCCGCAACCCCTTTCCCGGC GGTGCTCTGGGCCCAGCTCTGCCCGACCTG GGCATTGACGTGACTGTGCGCACAGCCCAA CCCCCGACCACCTCGCAAGCCTGCATCAAG GTGGGCGCTGCCCTTCGCCGAGCCGAAAAG GAGAAGCGCGACTACTACACCGGTTTCAAC CATGGAAAAACTCTGATCGTCCCTGCGGCG ATGACGACAACCGGTGGGTTCGCCTCCTCC TTTGTGGATCTGCTTGGTCAGCTCGCCCGCT GCGCCGAGGCCCGTGGTGTGTACCAGCCGG GGCTGGATGAGGCCTTTGTTCCTCGGTGGA AGGGTCGCTTTGCGGCGCTGGTCCATCAGA TGAACGCTGACCACATCCAGCGCCACTTTG GCGGTGTCTGCCTGCGCTCGTCGTAGGTAG		ID NO: 1282)	QMNADHIQRHF GGVCLRSS (SEQ ID NO: 1404)

					GCACCGTCTCGGGGGTCCCTCTGTGGGGAT CCCTGTGTGCACCTGTCGCTCCCTAGGTGGT TCCTCGTTGTGTCTTTTGATGGCTTGACTTG TATTTTTGTTTTAATTTTGCTTTAATTTTTGC TGTATTTGTGTGGTATTTTTGCTGAATTTTTG TAAGGTCCTTTGTATGATGTCTTTGTCTTCT GTGGTCGGTTGTTTTCCTCAATCCGACGTTG TGTCTCGTTGGATGTGAGCGTGCCGTGGTGT TCTTTGTGTTTGTGCTGTGATGGCTTGTAGT TGTGATGTGTGACTGCCTTTTTGGGTGTCTT GTGTTTGAAATGGCCGTATCTCTGGTTATAC TTGGTCGTTTTGTACGATTTTTGTTTCTATGT GCGTGATTCTTCGCGCTTGTACTTCTTGGCA TGATAGAAGCCAATGAATGTGTCTTGTTCTC TTGTGTTGTTTTGCGTGCCGTCGTGATTTTG ATGTCGGGGTTGCACAGCTTTGCTTTCAGCT CTGAGGTTCAAACACCTAATTTA (SEQ ID NO: 1036)			
CRE	CRE-2_HMa	.		Hydra vulgaris	AATTTAAAAAAAAAAAATCGTTTATTTATG GCATAATACTGTTTGTAATTTTTGAAAATTC GTGCAACAACTGCAGTTAAATTGAAGAGCT GAAATTTAAGATCTGAGCTTTTCAATCAGA GTTTTTTACCCTAAAACATTAAATTTTATCA	AATTTA AAAAAA AAAAAT CGTTTA TTTATG	TGAG CTCTT ATAA ATTT ATAT	MSSCKVTIPHVC PYCKVELKTICG INRHILKCKKNP LQIPSLQKTNTS LTLEPNTKVIPSI

					TAACAAAAATCGTTCTAATATTATTAAACTT	GCATAA	TATA	TKQNDIIIASTSS
					AAAGAAATTCGTTCTTATATCAAATCTTATT	TACTGT	GCAT	NNLAFNQKKDY
					TCAGTGTTTCACAGACGAAGGGTTTTACTA	TTGTAA	TTTGT	TLTPTYSRKTTP
					GATTTTTATTTTTTCAACTTTTGAATTTGTTT	TTTTTG	TTTA	VSILSSMKMTPI
					ATTATAAAACTGTAAACTAGTGTGCAACCA	AAAATT	(SEQ	SITSHIVRRKLPE
					ACCGTAAAAAWTAGTTAGCTGTTCACCCAA	CGTGCA	ID NO:	LPSQTTNHLFNE
					AATATTATTCAGTATGAAAATATTTAATCTT	ACAACT	1283)	NFINVPFLPEIM
					CTTTATTCGCAGTAAACAATAAAATATCTA	GCAGTT		NHLPVPNNNVM
					GTTAAACAAAATATTTCTTAATAATAAAAA	AAATTG		WGVYSYQQFKL
					ACAAAAACTTTTTCTTAACAAGTACAATGA	AAGAGC		FVDSTYDEIVNY
					GCAGCTGCAAAGTTACTATACCTCATGTGT	TGAAAT		RRNIFNIPSGKA
					GCCCTTATTGTAAAGTAGAACTTAAAACAA	TTAAGA		GKEFIEELTFWL
					TATGCGGAATAAACCGTCACATTTTAAAAT	TCTGAG		RKFNSTSSLNSI
					GCAAAAAGAATCCTTTACAAATACCCAGTC	CTTTTC		ALKVTMILPNLL
					TACAAAAAACTAATACCTCTTTAACACTCG	AATCAG		LQKPSAKSKSKE
					AACCAAATACTAAAGTAATACCCTCAATTA	AGTTTT		HTLCLTRRIDLW
					CAAAACAAAACGATATTATAATAGCATCCA	TTACCC		KKGDTSLLLKE
					CTTCGTCTAACAACTTAGCGTTTAATCAAAA	TAAAAC		VRNIQKKFVNS
					AAAGGACTACACATTAACACCTACATATTC	ATTAAA		KXKRSMDDISRI
					TAGAAAAACGACACCCGTAAGCATACTGTC	TTTTAT		FAKLIMEGKITA
					TTCTATGAAAATGACACCCATAAGTATAAC	CATAAC		ALKFLEKEASSG
					ATCACATATAGTTCGCAGAAAACTACCTGA	AAAAAT		ILPLSDNTLKDL
					GCTTCCTTCTCAAACAACAAATCATTTATTT	CGTTCT		KSKHPEPSRVED

					AATGAGAATTTTATAAATGTTCCCTTCTTGC	AATATT		YSLLFGPIDLIPK
					CTGAAATAATGAACCATCTACCAGTTCCAA	ATTAAA		CFFDCIDEQLVM
					ATAACAATGTCATGTGGGGAGTATACTCAT	CTTAAA		KAAFATKGSAG
					ATCAACAATTTAAATTGTTTGTGGATTCTAC	GAAATT		PSGMDADIYRRI
					CTATGATGAGATCGTAAATTACCGAAGAAA	CGTTCT		LCSKNFIKEGKE
					TATTTTCAACATTCCATCTGGAAAGGCAGGT	TATATC		LRKEIAKMTQN
					AAAGAATTTATAGAGGAGCTAACCTTTTGG	AAATCT		LLTETYEPTFLE
					TTAAGAAAGTTTAATTCCACTTCTAGTTTAA	TATTTC		AFTACRLIPLDK
					ATTCAATCGCGCTGAAAGTTACAATGATTTT	AGTGTT		NPGIRPIGVGEV
					GCCGAATCTTCTTTTGCAAAAACCCTCCGCC	TCACAG		LRRIIGKVISWSF
					AAATCAAAGTCCAAAGAACATACATTATGT	ACGAAG		NSEIKEAAGPLQ
					TTAACTCGTAGGATTGACCTTTGGAAAAAA	GGTTTT		TCAGHGAGAEA
					GGAGATACTAGTTTACTGTTAAAAGAAGTT	ACTAGA		AVHAMKEIFDN
					CGAAATATACAAAAAAAATTTGTAAATTCC	TTTTTAT		VQTDAILLIDAK
					AAAAAKAAAAGATCTATGGACGATATATCT	TTTTTC		NAFNCMNRQVA
					AGAATATTTGCCAAATTAATTATGGAAGGC	AACTTT		LHNIQIICPLISIY
					AAAATCACTGCAGCGCTGAAATTTTTAGAA	TGAATT		LINTYRNPSRLF
					AAAGAGGCATCATCCGGCATACTACCACTA	TGTTTA		VAGGKEISSQEG
					TCAGACAACACATTAAAAGACCTTAAAAGC	TTATAA		TTQGDPLAMPW
					AAACACCCTGAACCCTCCCGAGTAGAAGAT	AACTGT		YSCNTTIIIEHLL
					TATAGCTTACTGTTTGGTCCGATTGATTTAA	AAACTA		VNYPQVKQVW
					TCCCAAAATGTTTCTTCGATTGTATTGATGA	GTGTGC		LADDAAASGSIA
					GCAACTAGTTATGAAAGCAGCATTTGCAAC	AACCAA		NLHSWYQHLID

					TAAAGGATCTGCTGGACCATCAGGAATGGA	CCGTAA		EGCKHGYYVNQ
					TGCCGATATTTATCGCCGCATCTTATGTTCT	AAAWT		SKCWLIVKSPSL
					AAAAACTTCATCAAAGAAGGTAAAGAACTC	AGTTAG		AENAGIVFGKSV
					CGAAAAGAAATTGCTAAAATGACACAAAAC	CTGTTC		NITTEGQRHLGS
					TTACTAACAGAAACATATGAACCAACATTT	ACCCAA		VIGSQNFKNKY
					CTAGAAGCTTTCACTGCTTGTCGATTAATTC	AATATT		CTEKVAKWLTE
					CTCTAGATAAAAATCCAGGTATTAGGCCAA	ATTCAG		LKQLCKVAETQ
					TTGGAGTAGGAGAAGTATTAAGGCGTATCA	TATGAA		PQAAFIAFTKGF
					TAGGTAAAGTAATTAGCTGGAGTTTCAACA	AATATT		RSKFTYFLRTIP
					GTGAGATAAAAGAGGCAGCCGGGCCATTAC	TAATCT		KFEQYLAPVDEI
					AAACATGTGCTGGACATGGGGCAGGAGCCG	TCTTTA		LSHLLLPTLFGK
					AAGCGGCTGTACATGCCATGAAGGAAATAT	TTCGCA		DTPFEDHIRKLF
					TCGACAATGTGCAAACAGATGCAATACTTT	GTAAAC		TLTPRDGGLGIPI
					TGATTGACGCAAAGAACGCTTTTAATTGTAT	AATAAA		LVEEAPHQFLSS
					GAATCGACAAGTCGCCTTACACAACATCCA	ATATCT		VKLTKNLVQQII
					GATCATTTGTCCATTAATTTCAATTTACTTA	AGTTAA		DQDKILKTKNSS
					ATCAATACTTATCGAAATCCATCGAGGCTCT	ACAAAA		GNVLEDLEKILT
					TTGTGGCAGGGGGTAAAGAAATATCATCCC	TATTTC		TDRLKHRKEKII
					AAGAAGGCACAACTCAAGGTGATCCCCTTG	TTAATA		AVDSMQPDSML
					CTATGCCATGGTACTCTTGTAACACCACGAT	ATAAAA		RNIQQTRSECAS
					TATTATAGAACACTTACTTGTAAATTACCCA	AACAAA		TWLNALPLENQ
					CAAGTTAAGCAGGTGTGGTTAGCAGACGAT	AACTTT		GFVLNKEEFRD
					GCTGCAGCTAGTGGAAGCATTGCAAACTTA	TTCTTA		ALCLRYNFDLK

					CATAGCTGGTATCAACACCTTATTGATGAA GGATGTAAACATGGCTACTATGTAAACCAA TCTAAATGCTGGTTAATTGTAAAATCCCCCT CGTTAGCAGAGAATGCAGGCATAGTATTTG GTAAATCGGTCAACATAACTACAGAGGGTC AACGACATTTGGGTTCAGTAATAGGTTCGC AAAATTTTAAGAACAAATATTGCACTGAGA AAGTAGCAAAATGGTTAACCGAGTTAAAAC AACTTTGTAAAGTAGCAGAGACGCAACCAC AGGCCGCTTTTATTGCGTTTACAAAAGGATT TCGTTCAAAATTTACATATTTCCTAAGAACT ATTCCAAAATTTGAACAATATCTAGCGCCC GTAGACGAAATACTTAGTCATTTGTTGTTGC CAACTCTTTTTGGAAAAGATACGCCCTTTGA GGATCACATTAGAAAACTTTTTACATTAACT CCTCGAGATGGAGGATTGGGTATACCTATA CTAGTTGAAGAAGCGCCTCACCAGTTTTTAT CATCTGTTAAATTAACTAAAAATCTTGTACA GCAAATTATAGATCAAGATAAAATTTTAAA AACAAAAAACTCTTCCGGGAATGTTCTAGA AGATCTTGAAAAAATATTAACTACTGACAG ACTTAAGCATCGCAAAGAGAAAATAATTGC AGTTGATTCAATGCAACCAGATTCAATGTT	ACAAGT ACA (SEQ ID NO: 1160)		NIPRICECGEPFN VTHALSCKKGG FISSRHDNIRNLF TTLLKRVCINVQ SEPHLIPLDNEN FYFHTANKSNQ ARLDIKANGFW RNGQTAFFDVR VTHVNSMSNKN LDIAAIFRKHEK EKKREYGERVR EVEHGSLTPLVF GTNGGMGKECH RFVRRLAEKLA EKQNEKYSVVM TWLRTKLSFEIL RSTILCLRGSRTP WTKKNDFEIGV DFKMDALEARI (SEQ ID NO: 1405)

					AAGAAACATACAGCAAACAAGAAGCGAAT GTGCTAGCACCTGGTTAAACGCCTTACCACT AGAAAACCAAGGTTTTGTTTTAAATAAAGA AGAATTTCGAGATGCACTTTGCTTACGTTAT AATTTTGACTTGAAAAATATCCCTCGTATTT GTGAATGCGGAGAACCTTTCAATGTAACTC ATGCATTGTCGTGTAAGAAAGGTGGCTTCA TTTCAAGTCGTCATGATAACATAAGAAATTT GTTCACCACATTGCTTAAAAGAGTGTGCAT AAATGTTCAATCCGAACCACACCTCATACC CCTTGATAATGAGAATTTTTATTTCCATACA GCAAATAAAAGTAACCAAGCTCGTTTAGAT ATTAAAGCAAATGGTTTTTGGCGAAATGGA CAGACAGCATTTTTTGACGTAAGAGTTACG CATGTCAATTCCATGAGCAACAAAAATCTA GATATAGCTGCTATATTCAGAAAACACGAA AAGGAAAAAAAAAAGAGAGTATGGCGAACG AGTTCGTGAAGTCGAACATGGCAGCTTGAC ACCACTAGTGTTTGGTACAAATGGAGGTAT GGGTAAAGAATGTCATCGGTTTGTTAGAAG GCTAGCGGAGAAACTAGCGGAAAAACAAA ATGAAAAATACTCAGTTGTTATGACATGGT TAAGAACAAAGTTATCTTTTGAGATACTTCG			

					CTCAACTATCCTTTGTTTAAGAGGCTCAAGA ACACCCTGGACTAAAAAAAACGATTTTGAA ATCGGTGTTGATTTTAAGATGGATGCTCTAG AAGCAAGAATCTGAGCTCTTATAAATTTAT ATTATAGCATTTTGTTTTA (SEQ ID NO: 1037)			
CRE	MoTeR 1	JQ747487		Magnaporthe oryzae	CCCGAACCCGAACCCAAACCCAAACCCAAA	CCCGAA	TAAT	MVCPTCNGVYA
					CCCAAACCCAAACCCAAACCCAAACCCAAA	CCCGAA	AGGT	DYNDHIRKKHP
					CCCGGAGGGTTCCCAAGTCGCCTAAACCCG	CCCAAA	AACG	DERYTALQLQP
					AAGGGTTTAGGATATTATTTCGTTTATTAGA	CCCAAA	TCCCT	LGLTPCPICKTA
					ATTGGATAATTATTTACCCCTGTTGGACAGG	CCCAAA	ATTTT	CKNDLGVKTHL
					GGGGTTGCAGGGGTTAAATTAAGGTTTTTT	CCCAAA	TGTCT	SKIHKISGASKIS
					ATTATTTATGCGCCGTTTATTTGTTTACCCC	CCCAAA	TTGG	TQPRIRTENTDN
					CCCAAATATTATAAAAGCGCGTTCCATCCTC	CCCAAA	TTTTG	TNSVPTSSFNPV
					TTAGGAAAAGCGAAGCTTTTCCTTGTAAAA	CCCAAA	TTTTT	LPEIQTLTPGLN
					GTCGCTAGACTTTTACTATAAAAGTCGCTAG	CCCAAA	ATCTT	NSRWADNPRKR
					ACTTTTATACCAATCTTTTAACAAAAAGCGT	CCCGGA	TGTTT	RADTPSPTRGRN
					AGCTTTTTGTTGCCAATCTATTAAAAAAAGC	GGGTTC	TTGTT	TRPRRFSYTDID
					GGAGCTTTTTTTAACTTTTTCTTTTTTTTTTT	CCAAGT	TTTGT	LTNDEPADNPR
					TTTTTCTTTTTTTTTTTTTTTTTTCTTTTTTTTT	CGCCTA	TTTCG	ANNPRVNNPRV
					TTTTTTTTTTTTTATATATATTATTATTATTA	AACCCG	TTTTT	NNEPPSSPNSLPS
					TTATTAGCGGTGGGGCTATTTATGCGCTTTA	AAGGGT	GTTTT	ISEFHTPGTLPLT
					ATTTGTGCGGGGCTATTTATGCGCTTTAATT	TTAGGA	TGTTT	NSNISLKDQHDK
					TGTGCGGGGCTATTAATGCGCTTTAACTTTA	TATTAT	TCGTT	ITGPILQKPLIQK

					CAAATTTTATTTATGCGCTTTAATTGCTGCG	TTCGTT	TTTGT	LIEYSKIPIPEHH
					GGCCTGTTAATGCGCTTTAATTTACAAATTT	TATTAG	TTTTT	LHARQAKIFAD
					CATTAATGCGCTTTAACTTTTATATTTACTA	AATTGG	TTTTT	AANRIAKNFIQS
					ATGCGTTATTTATATAATTGCTATTATTATC	ATAATT	GTTTT	PTEKTLFNLLILP
					GTTGCTATTATTATTATTGCTATTATTATCGT	ATTTAC	TGTTT	RIFGIGLINGKVT
					TATTATTATTGCAATTTTATTATATAAACCC	CCCTGT	TTGTT	KIMQNFPSQIPPI
					TCGTTTGTCCCTCGATTTATCCCGTTTCTTTT	TGGACA	TTTGC	PKIDFPSEKTDS
					CCATCCCATCGCGCGTTTTCGTAAGCTTTGG	GGGGGG	CTTTG	DPVLNAKKLLE
					TTTTCGTAGGATTTGCTTTCGTAGGCTTTGC	TTGCAG	TTTTT	KGYIGRAAKAII
					TTTCGTAGGCTTTCGTCAGCTTTTACCTGCT	GGGTTA	GTTTT	DPTPVAPETPES
					TTTATTTTTTCTTTTTCTTTTTATTCCCCCCCC	AATTAA	TATCT	LNILREKHPIGQ
					TTTTTTTTACCTGGTTTATTAGCGGTTTACCT	GGTTTT	TTATT	NNPFNTKSQPIS
					GCTTTTATTACCTGGTTCCCCTTTACCTGTTT	TTATTA	TTTGT	GRQITEKAILLAI
					TATTAGCGGTTTACCTGCTTTTATTACCTGG	TTTATG	TTTTG	SSIGREKAPGLS
					TTCCCCTTTACCTACTTTATAAGCGGTTTAC	CGCCGT	TTTTT	GWTRSLLDAAI
					CTGCTTTTATTACCTGGTTCCCCTTTACCTGT	TTATTT	ACTTT	KIPTQNDVIPAL
					TTTATTAGCGGTTTACCTGCTTTTATTACCT	GTTTAC	GTTTT	RLLTDMIRQGT
					GGTTCCCCTTTACCTGTTTTATTAGCGGTTT	CCCCCC	ATTT	APGRELLCASRL
					ACCAGCTTTTATTACCTGGTTCCCCTTTACC	AAATAT	GTTTT	IGLSKPDGGVRP
					TACTTTATTAGCGGTTTACCCGTTTCTATTA	TATAAA	ATAT	IAVGDLLYKIAF
					GTGGGCATTTATTTCCCGTTTTTATTAGCAG	AGCGCG	TTAC	KAILNTLWSPNC
					TTAAATTTACCCTTTTAAGGTTATTTACCTG	TTCCAT	CTTTT	LLPYQLGVNSIG
					CTTTTATTCACAGGGCACCCCTGTTTTTACT	CCTCTT	GATT	GVEPAIFTLEEAI

					AGCAGTTAAATTTACCTTTTTAAGGTTATTT	AGGAAA	TTTTC	MGPNINGIKSITS
					ACCTGCTTTTATTCACAGGGCACCCCTGTTT	AGCGAA	TATTT	LDLKNAFNSVS
					TTACCAGCAGTTAAATTTACCTTTTTAAGGT	GCTTTT	TTCCC	RAAIASSVAKY
					TATTTACCTGCTTTTATTAACAACCCTTTATT	CCTTGT	ACCC	APTFYRSTCWA
					TTTTCCTATTAACGGGTATTTATTTACCTGTT	AAAAGT	TTATT	YNQPSILITENGS
					TTATTGGAATTCACCCGTTGGACGGCATGGT	CGCTAG	ATTA	VLASAQGIRQG
					TTGCCCAACCTGTAACGGCGTTTACGCCGAT	ACTTTT	TAAC	DPLGPLLFSLAF
					TACAACGACCATATCCGGAAAAAACACCCG	ACTATA	CCCA	RPTLETIQKSLP
					GACGAACGTTATACCGCCCTCCAACTCCAA	AAAGTC	ACCT	YTYIAAYLDDV
					CCATTGGGTTTAACCCCCTGCCCTATATGCA	GCTAGA	ACTA	YILSKTPVKDKI
					AAACCGCTTGCAAAAACGATTTGGGCGTTA	CTTTTA	ATAT	AKIIEKSPFTLNS
					AAACCCACCTATCCAAAATCCACAAAATAT	TACCAA	TTTTT	AKTTETDIDTLK
					CCGGTGCATCGAAAATTTCAACCCAACCGC	TCTTTT	CTTTT	TNGLKTLGSFIG
					GTATACGAACGGAAAATACGGATAATACCA	AACAAA	TTCTT	PTELRKEFLQNK
					ATTCGGTCCCCACGTCGTCGTTTAACCCTGT	AAGCGT	TTTTC	IQNFESSINALK
					CCTTCCCGAAATCCAAACGTTAACCCCGGG	AGCTTT	TTTTT	KLPKQYGLLILR
					GTTAAATAACAGCCGTTGGGCCGATAACCC	TTGTTG	ACGG	KSTQLLLRHLLR
					CAGAAAACGACGGGCCGATACCCCCTCCCC	CCAATC	TTTTA	TLNSQDLWELW
					AACACGGGGTCGGAATACACGCCCACGTCG	TATTAA	TTTTC	EKTDKLIADFVI
					ATTTTCATATACGGATATCGATTTAACAAAC	AAAAAG	CCGT	NLTVTKRKKRPI
					GACGAACCGGCGGATAACCCCAGGGCTAAT	CGGAGC	TTGTT	TDFVTPLITLPIK
					AACCCCAGGGTTAATAACCCCAGGGTTAAT	TTTTTTT	TTTTC	DGGFGLLRHNGI
					AACGAACCCCCCTCCAGCCCAAATTCGTTA	AACTTT	TATTT	AQDIYFAAKDL

					CCTTCGATTTCCGAATTTCACACCCCTGGGA	TTCTTTT	TATTT	TTEIRHKIQRISN
					CCCTACCCCTAACCAATTCGAATATATCGTT	TTTTTTT	GTAC	DFPQNQSPTATE
					AAAAGACCAGCACGACAAAATTACCGGCCC	TTTTTCT	GACA	ILHLLHNGVLAD
					TATATTGCAAAAACCGTTAATCCAAAAATT	TTTTTTT	AAAC	CKNGLTNAQLN
					AATCGAATATTCGAAAATCCCAATCCCAGA	TTTTTTT	CCTT	ALTENASYLGR
					ACACCACCTCCACGCCAGGCAGGCTAAAAT	TTTCTTT	AGCA	KWLNILPIQKSN
					TTTTGCTGACGCCGCAAATCGAATCGCCAA	TTTTTTT	AATA	RLTDWEMAEAV
					AAATTTTATACAAAGCCCAACGGAGAAAAC	TTTTTTT	AGCT	RLRLLAPVKPLT
					ATTATTTAATTTACTTATATTACCCCGCATA	TTTTTAT	TAGA	HPCNHCGNRTNI
					TTCGGTATCGGGTTAATAAACGGAAAAGTA	ATATAT	ATAT	NHEDVCKGAVR
					ACTAAAATAATGCAAAACTTCCCATCCCAA	TATTAT	AATA	KYTARHDQINR
					ATACCCCCTATTCCAAAAATTGATTTTCCAT	TATTAT	AAGC	SFVNSLKSRPEI
					CCGAAAAAAACCGATTCCGACCCGGTTTTAA	TATTAG	GCGA	DVEIEPDLNNEN
					ACGCCAAAAAATTATTGGAAAAAGGGTATA	CGGTGG	ATTA	NVNNANTTTEN
					TTGGCCGTGCGGCAAAGGCTATTATCGATC	GGCTAT	AAA	PTPSPNGQNDTG
					CAACCCCCGTTGCCCCAGAAACCCCGGAAT	TTATGC	(SEQ	CLFTTPIRSGTR
					CGTTAAATATTTTACGGGAAAAACACCCTA	GCTTTA	ID NO:	NGQNGLRADFA
					TTGGCCAAAATAACCCGTTTAATACAAAAT	ATTTGT	1284)	VINGVSKYYYD
					CCCAACCAATATCAGGCAGGCAAATTACCG	GCGGGG		VQIVAINKDSGN
					AAAAAGCTATTTTATTAGCTATTTCGTCTAT	CTATTT		TNPLNTLADAA
					TGGCCGGGAAAAAGCTCCGGGCCTTAGCGG	ATGCGC		NNKRRKYQFLD
					GTGGACGAGATCGTTATTAGATGCAGCCAT	TTTAAT		PFFHPIIISAGGL
					TAAAATACCTACCCAAAACGACGTAATTCC	TTGTGC		MEKDTAQAYK

					GGCTTTACGACTCTTAACGGATATGATTCGC CAGGGTACCGCACCGGGTAGGGAATTATTA TGCGCTTCGCGTTTAATAGGGCTATCCAAAC CCGACGGCGGCGTACGCCCAATAGCCGTTG GGGACCTATTATATAAAATAGCCTTTAAAG CTATTTTAAATACCCTATGGTCCCCAAACTG TTTATTACCTTACCAATTAGGTGTAAATAGT ATAGGTGGCGTCGAACCCGCTATTTTTACCC TCGAAGAGGCTATAATGGGCCCTAATATTA ACGGTATAAAATCGATTACCTCCCTCGATTT AAAAAACGCGTTTAATAGCGTATCCAGGGC TGCAATAGCCTCGTCGGTAGCTAAATACGC ACCAACTTTCTACCGTTCTACCTGTTGGGCC TATAACCAACCTTCGATTTTAATAACGGAA AACGGTTCCGTCCTGGCTAGTGCACAAGGT ATACGCCAAGGCGATCCGTTAGGCCCGTTG TTATTCAGCCTTGCTTTTCGACCTACGTTGG AAACGATCCAAAAATCGCTTCCATATACGT ATATAGCGGCTTATTTGGACGACGTTTATAT TTTATCCAAAACGCCCGTTAAAGATAAAAT AGCCAAAATAATCGAAAAAAGCCCGTTTAC CCTAAATTCCGCCAAAACGACAGAAACGGA TATCGATACGTTAAAAACCAATGGTTTAAA	GGGGCT ATTAAT GCGCTT TAACTT TACAAA TTTTATT TATGCG CTTTAA TTGCTG CGGGCC TGTTAA TGCGCT TTAATT TACAAA TTTCAT TAATGC GCTTTA ACTTTT ATATTT ACTAAT GCGTTA TTTATA TAATTG		QIQKLIGPVAAH WLDTSISLILLRS RTTAAISIAKNR PRA (SEQ ID NO: 1406)

					AACGCTCGGCTCGTTTATTGGACCAACGGA	CTATTA		
					ATTACGGAAGGAATTTTTGCAAAATAAAAT	TTATCG		
					TCAAAATTTCGAATCGTCCATTAACGCCCTG	TTGCTA		
					AAAAAACTCCCTAAACAATACGGATTGCTA	TTATTA		
					ATCTTGCGTAAAAGTACACAATTACTTTTAC	TTATTG		
					GCCATTTGCTCCGTACTTTAAATTCCCAGGA	CTATTA		
					CCTGTGGGAATTATGGGAAAAAACAGATAA	TTATCG		
					ATTAATAGCGGATTTCGTTATAAATTTAACT	TTATTA		
					GTTACAAAACGGAAAAAACGGCCAATTACG	TTATTG		
					GATTTCGTTACGCCGTTAATTACGTTACCTA	CAATTT		
					TAAAGGACGGAGGTTTTGGATTATTACGGC	TATTAT		
					ATAACGGAATAGCCCAAGATATTTATTTTG	ATAAAC		
					CGGCCAAGGATTTAACAACCGAAATTCGGC	CCTCGT		
					ACAAAATCCAACGTATATCCAACGATTTTC	TTGTCC		
					CACAAAATCAAAGCCCTACCGCCACCGAGA	CTCGAT		
					TTTTGCATTTGTTGCATAACGGGGTTTTAGC	TTATCC		
					AGATTGCAAAAACGGGTTAACAAACGCCCA	CGTTTC		
					ATTAAACGCTTTAACCGAAAACGCTAGTTA	TTTTCC		
					TTTAGGTCGAAAATGGCTTAACATTTTACCT	ATCCCA		
					ATCCAAAAATCAAATCGATTAACGGATTGG	TCGCGC		
					GAAATGGCTGAAGCCGTTCGATTAAGATTA	GTTTTC		
					TTAGCCCCGGTTAAACCGTTAACCCACCCCT	GTAAGC		
					GCAACCATTGCGGAAATCGGACCAATATAA	TTTGGT		

					ACCACGAGGACGTTTGCAAAGGTGCCGTAC	TTTCGT		
					GCAAATATACGGCCCGTCACGACCAAATAA	AGGATT		
					ACAGAAGTTTCGTCAATTCGTTAAAAAGTC	TGCTTT		
					GACCAGAAATCGACGTCGAAATCGAACCCG	CGTAGG		
					ATTTAAATAACGAAAATAACGTAAATAACG	CTTTGC		
					CCAATACAACCACCGAAAATCCCACCCCTA	TTTCGT		
					GCCCCAACGGCCAAAACGATACCGGATGCC	AGGCTT		
					TTTTTACAACCCCTATTCGCTCCGGGACCCG	TCGTCA		
					TAACGGCCAAAACGGCCTTAGGGCGGATTT	GCTTTT		
					TGCCGTTATTAACGGCGTATCCAAATATTAT	ACCTGC		
					TACGACGTGCAAATCGTTGCAATTAATAAG	TTTTATT		
					GATTCCGGTAATACAAATCCGTTAAATACG	TTTTCTT		
					TTAGCAGACGCAGCAAATAACAAACGACGT	TTTCTTT		
					AAATACCAATTTTTGGATCCATTTTTCCATC	TTATTC		
					CAATTATAATAAGCGCCGGAGGCCTTATGG	CCCCCC		
					AAAAGGATACAGCACAGGCGTACAAACAA	CTTTTTT		
					ATCCAAAAATTAATAGGCCCCGTTGCGGCC	TTACCT		
					CATTGGTTGGATACGTCGATTTCGTTAATTT	GGTTTA		
					TGTTACGGTCCAGAACGACGGCAGCAATTT	TTAGCG		
					CTATTGCTAAAAACCGCCCTCGTGCGTAAT	GTTTAC		
					AGGTAACGTCCCTATTTTTGTCTTTGGTTTT	CTGCTT		
					GTTTTTATCTTTGTTTTTGTTTTTGTTTTCGTT	TTATTA		
					TTTGTTTTTGTTTTCGTTTTTGTTTTTTTTTTT	CCTGGT		

					GTTTTTGTTTTTGTTTTTGCCTTTGTTTTTGTT TTTATCTTTATTTTTGTTTTTGTTTTTACTTTG TTTTATTTGTTTTATATTTACCTTTTGATTTT TTCTATTTTTCCCACCCTTATTATTATAACCC CAACCTACTAATATTTTTTCTTTTTTCTTTTT TCTTTTTACGGTTTTATTTTCCCGTTTGTTTT TTCTATTTTATTTGTACGACAAAACCCTTAG CAAATAAGCTTAGAATATAATAAAGCGCGA ATTAAAA (SEQ ID NO: 1038)	TCCCCT TTACCT GTTTTA TTAGCG GTTTAC CTGCTT TTATTA CCTGGT TCCCCT TTACCT ACTTTA TAAGCG GTTTAC CTGCTT TTATTA CCTGGT TCCCCT TTACCT GTTTTA TTAGCG GTTTAC CTGCTT TTATTA		

						CCTGGT TCCCCT TTACCT GTTTTA TTAGCG GTTTAC CAGCTT TTATTA CCTGGT TCCCCT TTACCT ACTTTA TTAGCG GTTTAC CCGTTT CTATTA GTGGGC ATTTAT TTCCCG TTTTTAT TAGCAG TTAAAT TTACCC		

						TTTTAA GGTTAT TTACCT GCTTTT ATTCAC AGGGCA CCCCTG TTTTTA CTAGCA GTTAAA TTTACC TTTTTA AGGTTA TTTACC TGCTTT TATTCA CAGGGC ACCCCT GTTTTT ACCAGC AGTTAA ATTTAC CTTTTT		

						AAGGTT ATTTAC CTGCTT TTATTA ACAACC CTTTAT TTTTTCC TATTAA CGGGTA TTTATTT ACCTGT TTTATT GGAATT CACCCG TTGGAC GGC (SEQ ID NO: 1161)		
HERO	HERO-2_BF	.		Branchiostoma floridae	TTTTCAGTCTGGCTCAGCCAGTGACCGCCGG GAAAGTCCGGCTGACTACCACGAATAGGGT GGTGACAGCTGGATAGACAGACGACAGCTC GGAAAGACGGCATTGGGGCAGTATGGGTTG GCACCCCTAACTGCATCTCCCCTAGGAGAG	TTTTCA GTCTGG CTCAGC CAGTGA CCGCCG	TGAT TAAA GACC CGAA ACAC	MNAVCVCGKV CKNQRGLRIHQ TKMACLRRVQA EHRSGAVATTV EPVLSASAPGQT

					CATCCCGCAACACGCTACAAAGAACCACAA	GGAAAG	CCAA	EEDQGPEAPHSA
					AGAGCAATACCCCCAGGGATGCCCGAGAGG	TCCGGC	TGAC	RNLRATPAPPQG
					GGGGGAGGATGAGCATCCCATTCGGACGGT	TGACTA	CCCG	RKSDHHRVKWP
					CCAATCGGTATTGACCCCAGCAAACGGAGA	CCACGA	GGTT	AANSKEWSQFD
					ATCGACAATGAATGCAGTCTGTGTGTGTGG	ATAGGG	CATC	EDVDMILESVSR
					CAAGGTATGTAAGAACCAGAGAGGTTTGAG	TGGTGA	ACTG	GSTDQKLQSMC
					AATCCACCAAACAAAGATGGCCTGCTTAAG	CAGCTG	ATGA	TVIMSMGAERF
					GAGGGTGCAGGCGGAGCACCGCTCAGGGG	GATAGA	TGTG	GTIGQRKPTDT
					CTGTGGCAACCACTGTAGAACCAGTGTTGT	CAGACG	TCCCT	MKPNRREVKIR
					CAGCATCAGCCCCTGGTCAGACGGAGGAGG	ACAGCT	GTTC	QLRQELKSLRRS
					ATCAGGGCCCGGAAGCTCCCCACAGTGCCC	CGGAAA	GCAC	FKASTSGEERAA
					GGAACCTCCGCGCAACGCCTGCCCCTCCAC	GACGGC	TACC	LAELTHHLREKL
					AAGGCAGGAAGTCAGATCATCACCGAGTGA	ATTGGG	AGAG	RTLRRAEWHKK
					AGTGGCCAGCCGCAAACTCCAAGGAGTGGT	GCAGTA	TGTA	KGKERARKRSA
					CGCAGTTTGACGAGGACGTTGACATGATCT	TGGGTT	TTCTA	FITNPFGFTKRLL
					TGGAGTCGGTGTCAAGAGGTAGTACAGACC	GGCACC	GAG	GQKRSGNLTCP
					AAAAGCTTCAGTCCATGTGCACAGTGATTA	CCTAAC	(SEQ	VEEINLHLSNTF
					TGTCCATGGGGGCAGAACGATTTGGCACGA	TGCATC	ID NO:	SDASRDVDLGP
					TTGGGCAGAGGAAACCGACAGACACAATG	TCCCCT	1285)	CPLLVTSPEPEV
					AAGCCAAATCGCCGGGAAGTAAAGATCCGT	AGGAGA		HFDISEPTLKEV
					CAACTGAGGCAGGAGCTAAAGTCGTTGAGG	GCATCC		RETVKAARSSSA
					CGGAGCTTTAAGGCGAGTACGTCGGGAGAG	CGCAAC		PGPSGVVYKVY
					GAGAGAGCTGCTCTTGCAGAGCTCACACAC	ACGCTA		KHCPRLVVRLW

					CACCTTAGGGAGAAGCTTAGGACCCTCAGA	CAAAGA		RILKVVWRRGK
					AGGGCAGAGTGGCACAAGAAGAAGGGTAA	ACCACA		VAADWRQAEG
					AGAAAGAGCCCGGAAGCGCAGTGCTTTCAT	AAGAGC		VWIPKEEESSKV
					CACCAACCCTTTCGGCTTCACCAAGCGACTC	AATACC		DQFRLISLLSVE
					CTAGGGCAGAAGAGGAGTGGGAACCTGAC	CCCAGG		GKIFFKIVAQRLI
					CTGCCCAGTCGAGGAGATCAACCTCCACCT	GATGCC		KYLLDNQYIDTS
					CAGCAATACCTTCAGTGATGCCTCGAGAGA	CGAGAG		VQKGGVPGVPG
					TGTGGATCTTGGTCCTTGCCCTTTGCTGGTG	GGGGGG		CLEHTGVVTQLI
					ACTTCACCTGAGCCGGAAGTGCACTTTGAC	AGGATG		REAKENRGDLA
					ATCTCTGAACCAACTCTGAAGGAGGTCAGA	AGCATC		VLWLDLANAYG
					GAGACAGTCAAGGCGGCGAGGTCCAGTTCG	CCATTC		SIPHKLVETALT
					GCGCCAGGTCCCAGTGGCGTGGTATACAAG	GGACGG		RHHVPESIQNLI
					GTCTACAAACATTGCCCACGGCTTGTGGTG	TCCAAT		LDYYSNFWLRA
					CGCCTCTGGAGGATCCTAAAGGTGGTCTGG	CGGTAT		GSSTATSAWQR
					CGCAGAGGTAAAGTGGCGGCTGATTGGAGG	TGACCC		LEKGIITGCTISV
					CAAGCCGAGGGGGTTTGGATCCCAAAGGAA	CAGCAA		PLFALAMNMIV
					GAGGAGTCAAGTAAGGTAGACCAGTTCCGC	ACGGAG		KGAEAGCRGPV
					TTAATTTCTCTGCTCAGTGTTGAGGGAAAGA	AATCGA		SRSGTRQPPIRA
					TCTTCTTCAAGATTGTGGCCCAGCGTCTAAT	CA (SEQ		FMDDLTVMTAT
					AAAGTACCTTCTGGACAACCAGTATATTGA	ID NO:		VPVCRWLLQGL
					CACATCTGTGCAGAAGGGGGGAGTTCCTGG	1162)		ERLITWARMSF
					TGTCCCAGGATGTCTTGAACACACGGGCGT			KPAKSRSLVLK
					AGTGACCCAGCTCATCCGGGAGGCTAAGGA			KGKVAERFRFT

					GAACAGAGGGGACTTGGCAGTCTTGTGGCT			LGGTQIPTVSEK
					GGATCTCGCGAATGCGTATGGTTCGATCCC			PVKSLGKVFNSS
					CCACAAGCTTGTGGAAACAGCACTGACCAG			LKDTASVQQTR
					ACACCATGTTCCAGAGTCAATTCAGAACCT			SDLTTWLEGIDK
					CATCTTAGATTACTACAGCAACTTCTGGCTA			TGLPGSFKAWM
					AGAGCTGGCTCCAGTACAGCAACTTCAGCA			FQHGVLPRVLW
					TGGCAACGGTTAGAGAAGGGCATCATTACT			PLLVYEVPMTM
					GGATGTACGATTTCAGTGCCCCTCTTTGCAC			VEQLERTISRFL
					TAGCGATGAACATGATTGTTAAAGGAGCGG			RKWLGLPRSLS
					AAGCAGGATGTAGGGGTCCCGTGTCTAGGT			NIALYGRSTKLQ
					CTGGAACCAGGCAGCCGCCGATTCGAGCCT			LPLSGLTEEFKV
					TCATGGACGATCTGACGGTGATGACTGCAA			TRAREVLMYRD
					CAGTCCCGGTGTGTAGATGGCTCCTACAGG			SSDSKVSSAGIH
					GATTAGAGCGTCTCATTACATGGGCACGGA			VRTGRKWKAQE
					TGAGTTTCAAGCCGGCCAAGTCAAGATCTC			AVDQAEARLRH
					TTGTCCTGAAGAAGGGGAAGGTGGCTGAAA			SVLVGSVAVGR
					GGTTCCGTTTCACCCTGGGAGGCACTCAGA			AGLGSCPKPRY
					TTCCCACAGTGTCAGAGAAACCAGTCAAGA			DKVSGKEKRLLI
					GTCTGGGCAAGGTGTTCAACAGCTCTCTGA			QDEIRAGEEEDR
					AGGACACCGCTTCAGTTCAGCAGACTAGGA			RCRMVGMRKQ
					GTGACCTGACAACGTGGCTCGAGGGAATTG			GAWTRWEHAD
					ACAAGACAGGGCTACCTGGTAGCTTCAAGG			SRKVTWPELCR
					CCTGGATGTTCCAGCATGGAGTCTTGCCAA			AEPSRIKFLISSV

					GGGTACTCTGGCCTCTTCTTGTGTACGAGGT			YDVLPSPANLH
					GCCGATGACCATGGTGGAGCAACTGGAGAG			VWGLAETPSCQ
					AACCATCAGCAGGTTCCTTCGCAAATGGTT			LCQRRGTLEHIL
					GGGGCTCCCGAGGTCCTTAAGCAACATTGC			SCCPKALGEGR
					CCTGTACGGTAGATCCACCAAGCTGCAGCT			YRWRHDQVLR
					TCCCTTGAGTGGCCTGACTGAAGAGTTCAA			VLADTVSNAIQS
					GGTTACCCGTGCAAGAGAAGTGTTGATGTA			SRSQQPPKKSIV
					CCGGGACTCCTCAGACTCCAAGGTCTCTTCA			FVRAGEKTRQQ
					GCCGGCATCCATGTCAGGACTGGAAGAAAA			PTSAGGLLSTAR
					TGGAAGGCACAGGAAGCAGTGGATCAGGC			DWQLLVDLGRQ
					AGAGGCAAGGTTGAGACACAGTGTCCTCGT			LKFPEHIVATSL
					GGGGTCCGTGGCAGTAGGACGGGCAGGACT			RPDMVLVSEST
					GGGCAGCTGCCCAAAGCCTCGGTACGACAA			RQVVLLELTVP
					AGTCAGCGGGAAGGAGAAGCGTCTACTGAT			WEERISEANERK
					CCAGGATGAGATAAGGGCTGGGGAAGAGG			RAKYAELVVQS
					AGGATCGGCGATGCAGGATGGTAGGCATGC			QSNGWRARCVP
					GCAAGCAAGGTGCGTGGACTAGGTGGGAAC			VEVGCRGFAGQ
					ATGCTGACTCCCGCAAGGTCACATGGCCAG			SLAYVLKLLGV
					AGTTGTGCAGAGCTGAGCCTTCTCGGATCA			RGFRLRKSIRDIL
					AATTTCTCATCTCTTCAGTGTACGACGTGCT			EAAEKASRWLW
					TCCAAGTCCAGCTAACTTGCATGTCTGGGG			FRRGEPWKPHG
					CTTGGCAGAGACCCCCTCATGCCAACTCTGT			HRSGNDQPRLG
					CAGAGGAGAGGTACCCTTGAACACATTCTC			

					AGTTGTTGTCCGAAAGCACTAGGGGAAGGG AGGTACCGCTGGCGGCATGACCAGGTTCTT AGGGTGTTGGCAGACACAGTTAGCAACGCC ATCCAGAGTAGCAGGAGTCAGCAACCCCCC AAGAAGTCAATTGTCTTTGTCAGGGCCGGA GAGAAAACCCGACAACAACCCACTTCCGCA GGTGGGCTTCTCTCCACTGCTAGAGATTGGC AGCTTCTAGTCGACCTTGGGAGACAGCTCA AGTTTCCAGAACACATTGTAGCCACGTCAC TTCGCCCTGACATGGTACTCGTGTCAGAATC CACCAGACAAGTGGTTCTGCTGGAGCTAAC TGTTCCCTGGGAGGAGCGGATAAGCGAAGC CAACGAGCGGAAGAGGGCGAAGTATGCCG AACTGGTAGTACAAAGCCAGAGTAATGGGT GGAGAGCCCGGTGTGTACCAGTGGAGGTTG GTTGCCGGGGTTTCGCAGGGCAGTCTTTGG CTTATGTGTTAAAACTCCTTGGAGTAAGAG GTTTCCGTCTTCGGAAATCCATCAGGGATAT TCTAGAGGCTGCGGAGAAAGCCTCACGTTG GTTGTGGTTCCGTAGGGGGGAACCGTGGAA GCCACACGGACACAGGTCGGGGAATGATCA ACCTCGGCTGGGTCGCCCGGGCGAGGGTGT ATGGTGATTAAAGACCCGAAACACCCAATG			RPGEGVW (SEQ ID NO: 1407)

					ACCCCGGGTTCATCACTGATGATGTGTCCCT GTTCGCACTACCAGAGTGTATTCTAGAG (SEQ ID NO: 1039)			
HERO	HERO-2_DR	.		Danio rerio	TTCAAGCCTGGCGCAGCCAGTGACTCCTAG	TTCAAG	TGAT	MTHANEQTTNK
					GAATAGACTAGGTGGCAACCAAGAATAGTT	CCTGGC	CAAC	IYVTCICGKLCK
					TGGTCGACTACTGGAGAGACAGTTGACGGC	GCAGCC	CCCG	NHWGLKIHQAR
					ACGGAAAGACGGCACTTGGGACAGTATGGG	AGTGAC	GCTG	MKCLEQESKVQ
					TTAGCACCCCAGCCTGTGTCTTTCGTGAGAG	TCCTAG	GGTC	RTGPEPGETQEE
					AGAACCCAAACAAGCTACGGAAAGCCCCAC	GAATAG	ACCT	PGPEATHRAKSL
					AGAGATATACCCCCAGGAGATCCCGAGAGG	ACTAGG	GGGT	HVPEPQTPSEVV
					GGGGGAGGATGAGATCTCCAATCGGACGGA	TGGCAA	GAGA	QQRIKWPPASK
					TCAAAGGTTAATGACCCATGCAAACGAACA	CCAAGA	GTGT	GSEWLQFDEDV
					GACGACGAACAAAATATATGTGACATGCAT	ATAGTT	ATGA	SNIIQAIAKGDA
					TTGCGGAAAGCTGTGTAAGAACCATTGGGG	TGGTCG	TGTT	DSRLKTMTTIIFS
					CCTAAAAATCCATCAGGCCAGAATGAAATG	ACTACT	GAGA	YALERFGCIEKG
					TTTGGAGCAGGAGAGTAAGGTGCAACGCAC	GGAGAG	GACC	KTKPTTPYTMN
					AGGTCCTGAACCTGGTGAGACGCAGGAGGA	ACAGTT	CGAA	RRATQIHHLRQE
					GCCCGGCCCGGAGGCAACCCACAGAGCCAA	GACGGC	ACAC	LRSLKKLYKKA
					GTCCCTCCATGTACCAGAGCCTCAAACTCC	ACGGAA	TCAA	TDEEKQPLAELK
					AAGCGAAGTAGTTCAACAGCGGATTAAATG	AGACGG	TGAT	NILRKKLMILRR
					GCCCCCAGCCAGCAAAGGAAGTGAGTGGCT	CACTTG	CCCA	AEWHRRRGRER
					GCAGTTCGATGAAGATGTGTCCAACATCAT	GGACAG	GGAT	ARKRAAFITNPF
					TCAAGCCATAGCCAAAGGAGATGCAGATAG	TATGGG	ACAT	GFTKQLLGDKR

					CCGACTCAAAACGATGACTACCATCATCTT	TTAGCA	CACT	SGRLECSIEEVN
					CAGCTATGCTCTAGAAAGATTCGGTTGCAT	CCCCAG	GATG	RFIEETVSDPLRE
					AGAGAAAGGAAAGACCAAGCCCACCACCC	CCTGTG	ATGT	QELEPNKALISP
					CCTACACTATGAACCGTAGGGCTACCCAGA	TCTTTC	GTCC	TPPAREFSLRGP
					TACATCACCTGCGTCAGGAGCTTCGCTCCCT	GTGAGA	CAAA	SLKEVKEIIKAS
					CAAGAAACTGTATAAGAAAGCTACGGATGA	GAGAAC	TGCA	RSASTPGPSGIPY
					GGAGAAGCAACCATTAGCGGAGTTGAAAA	CCAAAC	TCCA	LVYKRCPGLLL
					ACATTTTGCGGAAGAAGCTGATGATCCTAC	AAGCTA	TGAG	HLWKILKVIWQ
					GCAGGGCAGAGTGGCATCGGAGACGAGGG	CGGAAA	ATGT	RGRVAEQWRCA
					CGAGAGAGAGCCAGGAAGCGAGCTGCCTTC	GCCCCA	TTCTT	EGVWIPKEENSK
					ATCACCAATCCCTTTGGCTTCACAAAACAG	CAGAGA	GCAT	NINQFRIISLLSV
					CTGCTCGGGGACAAGCGGAGCGGTCGACTT	TATACC	AA	EGKVFFSIVSRR
					GAATGCTCAATAGAGGAAGTGAATCGCTTC	CCCAGG	(SEQ	LTEFLLENNYID
					ATTGAGGAAACAGTGAGTGATCCACTGAGA	AGATCC	ID NO:	PSVQKGGIPGAP
					GAGCAGGAGCTGGAGCCCAACAAAGCTCTT	CGAGAG	1286)	GCLEHTGVVTQ
					ATCAGCCCCACCCCTCCAGCAAGAGAGTTC	GGGGGG		LIREAHENRGDL
					AGTTTGAGGGGGCCAAGTCTGAAGGAGGTC	AGGATG		VVLWLDLANAY
					AAGGAAATCATTAAGGCATCTCGCTCAGCA	AGATCT		GSIPHKLVELAL
					TCTACTCCAGGCCCTAGTGGCATACCTTACC	CCAATC		HRHHVPSKIKDL
					TTGTCTATAAGCGCTGCCCAGGGCTTCTCCT	GGACGG		ILDYYNNFKMR
					GCATCTGTGGAAGATCTTGAAGGTGATTTG	ATCAAA		VTSGSETSSWHR
					GCAACGAGGAAGAGTTGCTGAGCAGTGGA	GGTTA		IGKGIITGCTISVI
					GGTGTGCCGAGGGAGTGTGGATTCCTAAAG			LFALAMNMVV

					AGGAAAACTCGAAAAACATCAACCAGTTTC	(SEQ ID		KSAEVECRGPLT
					GAATCATCTCTCTATTGAGTGTTGAAGGGA	NO: 1163)		KSGVRQPPIRAY
					AGGTGTTTTTCAGCATCGTCTCACGAAGACT			MDDLTITTTTVP
					GACAGAGTTCCTCCTCGAGAACAATTATAT			GSRWILQGLERL
					TGACCCTTCAGTGCAGAAGGGAGGGATTCC			IAWARMSFKPS
					TGGAGCTCCCGGCTGCTTGGAACACACTGG			KSRSMVLKKGK
					AGTAGTTACACAACTCATCAGAGAGGCCCA			VVDKFHFSISGS
					TGAGAACAGAGGGGACTTGGTTGTCTTGTG			VIPTITEQPVKSL
					GTTGGACTTGGCAAATGCCTATGGGTCCAT			GKLFDSSLKDSA
					ACCCCACAAGCTGGTTGAGCTCGCTCTACA			AIQKSKKELGA
					CCGCCACCACGTTCCTAGTAAGATTAAGGA			WLAKVDKSGLP
					CCTAATTCTGGATTACTACAATAATTTCAAG			GRFKAWIYQHSI
					ATGCGGGTCACATCTGGGTCAGAAACATCA			LPRVLWPLLIYA
					AGCTGGCATCGCATCGGGAAAGGAATAATA			VPMSTVESLERK
					ACAGGCTGCACCATCTCAGTTATTCTTTTCG			ISGFLRKWLGLP
					CTCTCGCCATGAACATGGTGGTCAAGTCAG			RSLTSAALYGTS
					CCGAAGTGGAATGCAGAGGGCCCTTAACTA			NTLQLPFSGLTE
					AGTCAGGTGTGCGACAGCCCCCTATTAGAG			EFMVVRTREAL
					CATATATGGATGACCTTACCATCACAACAA			QYRDSRDGKVS
					CAACGGTCCCAGGGAGCAGGTGGATCTTAC			SACIEVRTGRK
					AAGGACTTGAGAGACTCATCGCCTGGGCTA			WNAGKAVEVA
					GAATGAGTTTTAAGCCCTCCAAGTCTAGGT			ESRLQQKALVG
					CCATGGTGCTGAAGAAGGGGAAAGTGGTTG			TVATGRAGLGY

					ACAAGTTCCATTTTTCCATCTCAGGAAGTGT			FPKTLVSQVKG
					CATCCCAACCATCACGGAGCAACCTGTCAA			KERHHLLQGEV
					GAGTTTGGGGAAGCTCTTTGACTCCAGCCT			RASVEEERVSRV
					AAAAGACTCTGCAGCCATCCAGAAGTCCAA			VGLRQQGAWTR
					AAAAGAACTTGGAGCTTGGCTGGCGAAGGT			WNTLQRRITWA
					TGACAAATCCGGCCTGCCTGGTAGATTCAA			NILQADFQRVRF
					AGCCTGGATCTATCAGCATTCAATTCTGCCC			LVQAVYDVLPS
					CGAGTTTTGTGGCCTCTGCTGATCTATGCAG			PSNLHVWGKNE
					TCCCAATGTCAACAGTTGAGTCCCTAGAAA			TPSCLLCSGRGS
					GGAAGATCAGTGGCTTTCTTCGAAAATGGT			LEHLLSSCPKAL
					TGGGCCTCCCACGCAGTCTTACCAGTGCTGC			ADGRYRWRHD
					ACTATACGGGACAAGTAACACCTTGCAGCT			QVLKAIAASLAS
					ACCATTCAGTGGCCTCACAGAGGAATTCAT			AINTSKNHRAPR
					GGTTGTACGCACCAGAGAAGCCCTACAGTA			KAVHFIKAGEKP
					CAGGGACTCTAGAGATGGCAAGGTGTCATC			RALPQLTTGLLH
					AGCCTGCATCGAGGTGAGGACAGGCAGGA			KASDWQLEVDL
					AATGGAATGCAGGGAAAGCAGTGGAGGTG			GKQLRFPHHIAA
					GCAGAGTCACGCCTGCAACAAAAGGCTCTG			TRLRPDIIAISEA
					GTGGGCACTGTAGCGACAGGCAGAGCGGGC			SRQLIILELTVP
					TTGGGCTATTTTCCAAAGACCTTAGTAAGCC			WEERIEEANERK
					AGGTCAAAGGCAAGGAAAGACACCACCTA			RAKYQELVEEC
					CTCCAGGGAGAGGTTCGAGCAAGTGTGGAG			RERGWRTYYEPI
					GAAGAGAGAGTCAGTAGGGTGGTAGGACTC			EIGCRGFAGRSL

					CGGCAGCAGGGAGCATGGACTAGGTGGAAT ACACTGCAACGTAGGATCACCTGGGCGAAC ATCTTGCAGGCGGATTTCCAACGTGTCCGTT TCCTAGTACAAGCTGTCTACGATGTACTGCC AAGCCCATCAAACCTCCACGTTTGGGGAAA GAATGAGACACCTTCCTGCCTTCTTTGCTCT GGAAGAGGCTCTCTAGAACATCTCCTCAGC AGTTGCCCCAAGGCTCTGGCTGATGGTCGC TATCGTTGGCGCCATGACCAGGTGCTTAAG GCAATTGCTGCGAGCTTAGCTTCAGCCATTA ACACGAGCAAGAACCATCGTGCTCCAAGGA AGGCAGTCCACTTCATCAAAGCTGGAGAAA AACCCCGGGCCCTCCCACAATTAACAACAG GCCTCCTTCACAAAGCCTCGGACTGGCAGC TGGAGGTCGACCTGGGAAAACAGCTGAGGT TTCCTCATCACATCGCTGCAACACGTCTCCG TCCAGACATTATAGCTATCTCAGAAGCTTCA AGACAGCTAATTATTCTGGAGCTTACAGTG CCGTGGGAAGAGCGTATTGAAGAAGCAAAT GAGAGGAAGCGCGCTAAGTACCAGGAATTA GTGGAGGAGTGCAGGGAGAGAGGCTGGAG AACTTACTATGAGCCCATAGAAATTGGATG CAGAGGCTTTGCAGGGCGTTCACTTTGCAA			CKVLSRLGITGV AKKRAIRSASEA AEKATRWLWIK RADPWTAVGTQ VGT (SEQ ID NO: 1408)

					AGTCCTCAGTCGTTTGGGCATTACAGGCGT GGCGAAGAAAAGGGCCATTCGATCCGCAAG CGAAGCCGCAGAGAAGGCCACAAGGTGGC TGTGGATTAAGAGGGCAGATCCGTGGACTG CTGTTGGGACACAAGTCGGGACTTGATCAA CCCCGGCTGGGTCACCTGGGTGAGAGTGTA TGATGTTGAGAGACCCGAAACACTCAATGA TCCCAGGATACATCACTGATGATGTGTCCC AAATGCATCCATGAGATGTTTCTTGCATAA (SEQ ID NO: 1040)			
HERO	HERO-3_BF			Branchiostoma floridae	CTGACCAGCAGACGGGAAGCCCGCGACCAA CTAGTCTCCGCAAATATTGCACACAGGGCG ACCCTATGGAGCTGATTCAGTCAAATTTCCT CTGAGATATACCGATAACTATCTACAGAAA CTGCACAGTTAGTTTGGAAAGAGCTTTTCTA CTGAAAGACAGCAAAATCCGCCACTTTAGA CGAGCGTCAAGACTGCCCTCCCCATAACCA ATATGGCGCTACCTGCTGTACGTTCTGGACC AGCCAGCACCTGGACACTGTTAATCACGCT GGTCATCGTCGCTGCTAAAGGTACAGATGG TTTTATGTCTTTTAAACTGCCACTGCTGTCT ACTGATACCTGGTCTGGGTATAACAATGAT GTGAAAACCCTGCTAGGCCCGCTCCACCAC	CTGACC AGCAGA CGGGAA GCCCGC GACCAA CTAGTC TCCGCA AATATT GCACAC AGGGCG ACCCTA TGGAGC TGATTC	TAGA AACC CACA AGGC TGAG AAAT GTAG AGCA TCTGT ATGG ACAA TATT GATG	MALPAVRSGPA STWTLLITLVIV AAKGTDGFMSF KLPLLSTDTWSG YNNDVKTLLGP LHHELATNEMS PKLAGEGFSDIM CDFMASKPEFSH TTEESHSEGYIS HEPQSLAQVKR LKNKLRKKAFR ADATPEDRKAF RDAIKTYSFMK

					GAACTGGCCACAAATGAAATGTCCCCCAAA	AGTCAA	ATTG	RQQKRKETTKS
					CTAGCTGGGGAGGGATTCAGTGACATCATG	ATTTCC	AAAT	AAHQEKEYHKN
					TGCGACTTTATGGCCAGTAAACCAGAGTTC	TCTGAG	GTTG	FWKFAGKCAKG
					AGCCACACTACCGAAGAAAGTCACTCAGAA	ATATAC	TGAT	QLDIPPVKPAFS
					GGCTATATAAGCCACGAACCACAGTCTCTC	CGATAA	TTTA	VYYANEYYKNK
					GCACAAGTAAAACGCCTGAAAAACAAGCTA	CTATCT	GATC	YSHPTRVDFNK
					CGTAAGAAGGCATTCAGAGCTGACGCAACA	ACAGAA	AAAT	LLWFPHLPVEEQ
					CCTGAGGATCGAAAGGCTTTCAGAGATGCA	ACTGCA	TTAG	LPANSFDMSPVR
					ATTAAAACATACTCCTTCATGAAGCGACAA	CAGTTA	AAAT	PKDIKAVLSKRC
					CAGAAACGAAAGGAAACTACAAAATCGGC	GTTTGG	ATGA	ATSAPGPDGIMY
					AGCACACCAAGAGAAAGAATATCATAAGA	AAAGAG	AAAC	GHLKHLPACHL
					ACTTTTGGAAGTTTGCCGGAAAATGTGCAA	CTTTTCT	CGAA	FLSTLFSKLLES
					AAGGACAGCTCGATATCCCCCCAGTAAAAC	ACTGAA	CTAA	GDPPTSWSSGN
					CGGCATTCTCTGTTTATTATGCAAATGAGTA	AGACAG	ACTA	VSLIHKDGSPEA
					CTACAAAAACAAATACTCACACCCAACCCG	CAAAAT	AATA	AENFRMICLTSC
					TGTTGACTTCAACAAACTGCTCTGGTTTCCT	CCGCCA	TAAT	VSKIFHQILSER
					CATTTGCCGGTGGAGGAACAACTACCTGCG	CTTTAG	GTTTT	WAKYMTCNDLI
					AACTCTTTTGACATGTCACCTGTCAGGCCGA	ACGAGC	TTTTA	DPETQKAFLTGI
					AAGACATTAAGGCAGTCTTATCCAAACGAT	GTCAAG	AAGT	NGCVEHVQVM
					GCGCTACATCTGCACCTGGCCCGGACGGGA	ACTGCC	AATG	REILAHAKKNR
					TCATGTATGGCCACCTCAAGCACCTGCCAG	CTCCCC	ATAA	RTVHITWFDLA
					CTTGTCACCTGTTCCTTAGTACACTGTTCTC	ATAACC	GCAA	DAFGSVEHELIY
					CAAACTGCTTGAGTCCGGAGACCCACCGAC	AAT	TACC	YQMERNGFPPII

					ATCATGGTCATCTGGCAACGTGTCACTTATA	(SEQ ID	CACA	TTYIKNLYSRLK
					CACAAGGATGGTAGTCCAGAAGCTGCCGAA	NO: 1164)	TTGT	GKVKGPGWESD
					AACTTTCGAATGATCTGCCTTACTTCCTGCG		GCAA	PFPFGRGVFQGD
					TCTCCAAGATTTTCCACCAAATACTCTCGGA		TACT	NLSPIIFLTVFQPI
					ACGATGGGCAAAGTACATGACTTGCAATGA		ATCT	LQHLKGVEQQH
					TCTGATAGACCCAGAAACACAAAAGGCATT		ATGT	GYNLNDKHYVT
					CCTGACCGGAATCAACGGCTGTGTGGAGCA		TATG	LPFADDFCLITT
					TGTCCAAGTTATGCGGGAGATCTTAGCACA		TCCTT	NKRQHQKLITQI
					TGCCAAGAAAAACCGCCGAACAGTCCACAT		TGTC	SSNTKSMNLKL
					TACATGGTTTGACCTCGCGGATGCCTTTGGT		CCCC	KPRKCKSMSIVS
					TCTGTAGAACACGAACTGATCTACTACCAG		CTGC	GKPSDISFTIDGD
					ATGGAGAGAAACGGCTTCCCGCCAATTATC		ATGT	PVKTTKDAPEKF
					ACCACGTACATTAAAAACCTGTATTCTCGCC		TTGG	LGGYITFLSKTK
					TGAAAGGGAAAGTGAAGGGTCCAGGCTGG		TCAA	ETYDILAKTIETT
					GAAAGTGATCCGTTCCCGTTCGGAAGAGGA		TAAT	VENINKSAIRNE
					GTGTTCCAAGGAGACAACTTGTCACCCATC		GACC	YKLRVYMEYAF
					ATCTTCCTAACGGTGTTCCAGCCTATTCTAC		ATCG	PSWRYMLMVH
					AGCATCTCAAGGGAGTAGAGCAGCAACATG		TGTC	DLTDTQLQKLD
					GCTACAACCTCAATGACAAGCATTATGTTA		CTGG	SIHTKAIKTWLR
					CACTGCCTTTCGCAGACGACTTTTGTCTCAT		GCTC	MQPSATNAILY
					AACCACAAACAAACGACAGCATCAGAAACT		CGTG	NTRGLNFKSISD
					AATTACTCAAATTTCTTCCAACACAAAGTCA		TACC	LYLEAHALAYS
					ATGAACCTAAAGCTAAAACCACGCAAGTGT		TTTCT	RSVLKADEKVK

					AAGTCTATGTCTATAGTGAGCGGAAAGCCA		TTACT	HALQAKLDRES
					TCGGACATCAGCTTCACAATAGATGGGGAC		ATGA	QWTRKMQKWG
					CCTGTCAAAACGACCAAAGATGCACCGGAG		ATAA	IGKCHTIHQQAI
					AAATTCCTAGGTGGCTACATCACCTTCCTGA		AGAA	HVAKDSEWTSV
					GTAAAACAAAAGAGACCTATGACATCCTAG		TGAT	RKHVKQQVTD
					CAAAGACAATAGAAACGACTGTTGAAAACA		TTTAC	MRHDVWTKHQ
					TAAACAAATCAGCGATAAGGAACGAATACA		TAC	ENLLQQGQMLQ
					AACTCAGGGTTTACATGGAGTACGCCTTCC		(SEQ	LLEEEKCDLTW
					CATCTTGGAGGTACATGCTGATGGTACACG		ID NO:	RSAMYNLPRGIL
					ACCTGACAGACACCCAGCTACAAAAACTCG		1287)	SFAVRASIDALP
					ATTCCATCCACACAAAGGCGATCAAAACAT			TLCNLTTWGKR
					GGCTCAGAATGCAACCTAGTGCAACAAATG			NTDKCKLCGNR
					CAATTCTGTACAACACAAGGGGTCTCAACT			ETLHHVLNHCG
					TCAAAAGCATCTCAGACTTGTACCTAGAAG			VALQQGRYTFR
					CCCACGCTCTGGCCTACAGTAGGTCAGTCCT			HNSVLKHITDTII
					CAAAGCAGATGAGAAGGTAAAACACGCTTT			ESIDTSRINATIY
					ACAAGCCAAACTGGACCGCGAATCGCAATG			ADIQGYTTNGG
					GACTAGGAAAATGCAGAAATGGGGTATTGG			TIPVHTIPTTQKP
					AAAGTGTCACACCATCCACCAGCAAGCCAT			DLIIYLPEQKTL
					CCATGTAGCAAAGGACTCAGAATGGACATC			HIHELTVPFEKNI
					AGTACGCAAACATGTCAAACAACAAGTCAC			KTSHDRKVNKY
					AGATATGCGTCATGACGTCTGGACTAAACA			STLAADLETAGI
					TCAGGAAAACCTTCTACAGCAAGGGCAGAT			SATLTCFEVGSR

					GCTACAACTGCTTGAGGAAGAAAAATGCGA CCTGACATGGCGGTCCGCTATGTACAACCT GCCGAGGGGCATCCTCAGTTTCGCTGTGCG TGCCTCCATCGACGCCCTCCCCACACTCTGT AACCTGACCACCTGGGGAAAACGTAACACT GACAAATGTAAACTGTGTGGCAACCGGGAA ACACTCCACCACGTTCTGAACCACTGCGGT GTCGCTCTCCAACAAGGACGGTACACATTC CGACACAACTCGGTATTGAAGCACATAACG GACACCATCATAGAGTCCATTGACACCTCT CGGATCAACGCCACCATCTATGCGGACATA CAAGGTTACACAACTAACGGAGGTACCATC CCGGTCCATACAATACCCACTACCCAGAAA CCAGACCTGATCATATATTTACCAGAACAG AAGACCCTCCACATCCATGAACTGACTGTA CCCTTTGAAAAGAACATCAAAACAAGTCAT GACCGAAAGGTCAACAAATACAGCACCCTA GCGGCAGATTTAGAAACTGCTGGCATTTCC GCTACACTAACCTGCTTTGAAGTCGGATCA AGGGGACTCGTCACGCCAGAGAACAAGACC AGGCTTAGAACACTGTTCAAAATAGTTAAA GCCAAACCACCGAAGACTCTGTTTACTGAT ATAAGCCGCATTGCGATGTTATCGTCATATG			GLVTPENKTRLR TLFKIVKAKPPK TLFTDISRIAMLS SYAIWNSRHEPY WESETLL (SEQ ID NO: 1409)

					CTATTTGGAACTCACGCCACGAACCGTATT GGGAGTCAGAAACGCTATTGTAGAAACCCA CAAGGCTGAGAAATGTAGAGCATCTGTATG GACAATATTGATGATTGAAATGTTGTGATTT TAGATCAAATTTAGAAATATGAAAACCGAA CTAAACTAAATATAATGTTTTTTTTAAAGTA ATGATAAGCAATACCCACATTGTGCAATAC TATCTATGTTATGTCCTTTGTCCCCCCTGCA TGTTTGGTCAATAATGACCATCGTGTCCTGG GCTCCGTGTACCTTTCTTTACTATGAATAAA GAATGATTTTACTAC (SEQ ID NO: 1041)			
HERO	HERO Dr	.		Danio rerio	AAAGCAGTAGAGATGACGACACATCGCGCA GAAGTTACAACTTCTGGTAAGACGCAGGAG GAGCCAGGCCCGGAGGCAACCCACAGTGCC CAGAGCCTCCTAGTGTCGCCAACACCTGCT GCCGGCCGCTCGCCTGCTACTCAAAGCTGC CCTCAAGTGACAGCAGCTCATAACAGTCCA CAAAGCCCCCAAAGTCAGCAAGTGGCAGTT ACAAGATCTGACTGTGTTCCCTTGGCACAG CCAAGAATCCAGTGGCCCCAATCCTCAAAG AAAGCTGAGTGGCTCCAGTTCGACAAGGAC GTGAATCAGATCCTGGAAGTGACAGGCAAG GGGGGTGTGGACCAGCGACTGTCAACAATG	AAAGCA GTAGAG (SEQ ID NO: 1165)	TAGC ATGC CACT TGGA CACA GGCC GGGG TCTG ATCA GCCT CGGT CGGG	MTTHRAEVTTS GKTQEEPGPEAT HSAQSLLVSPTP AAGRSPATQSCP QVTAAHNSPQS PQSQQVAVTRS DCVPLAQPRIQ WPQSSKKAEWL QFDKDVNQILE VTGKGGVDQRL STMTTLIVNIAA ERFGTVTPKPTP

					ACCACGCTCATAGTGAACATTGCAGCTGAG		TCGC	STYTPSHRVKEI
					CGATTCGGAACTGTGACACCCAAACCCACT		CTGG	KRLRKELKLLK
					CCATCGACATATACTCCAAGCCACAGAGTA		AGGA	RQYKAAGEVER
					AAGGAAATCAAACGTCTCAGGAAAGAACTT		GGGT	AGLEDLRGILRK
					AAGCTACTAAAGAGGCAGTACAAGGCAGC		GTCT	QLVNLCRAEYH
					AGGGGAAGTAGAAAGAGCGGGCCTAGAAG		GTTG	RKRRRERARKR
					ATCTGAGAGGAATCCTGAGGAAACAGCTCG		CAAG	AAFLANPFKLTK
					TGAACCTATGTAGGGCAGAGTATCACAGGA		ACCC	QLLGQKRTGKL
					AGAGGCGGAGAGAGAGAGCAAGGAAAAGG		GAAA	TCSKEAINNHLK
					GCAGCATTTTTGGCCAACCCTTTCAAGTTGA		CACC	ATYSDPNREQPL
					CCAAGCAGCTCCTTGGCCAAAAGAGGACTG		CTGT	GPCGALLTPPEP
					GCAAACTCACCTGCTCCAAGGAGGCTATCA		GAGC	TSEFNMKEPCRS
					ACAATCACCTCAAGGCCACTTATTCTGACCC		CCAG	EVEEVVRRARSS
					GAATAGAGAACAACCCCTGGGGCCTTGCGG		GAAA	SAPGPSGVPYKV
					TGCACTGCTGACACCACCTGAGCCCACATC		CAAC	YKNCPKLLHRL
					AGAGTTCAACATGAAGGAACCCTGCCGGAG		ACTG	WKALKVIWRRG
					TGAAGTAGAGGAAGTGGTGAGGAGAGCAA		ATGA	KIAQPWRYAEG
					GGTCAAGCTCAGCACCAGGCCCAAGCGGAG		TGTG	VYIPKEEKSENI
					TGCCTTACAAGGTATATAAGAACTGCCCAA		TCCA	DQFRVISLLSVE
					AGCTTCTACACAGGCTCTGGAAGGCCCTGA		AGGT	SKIFFSIVAKRLS
					AAGTCATATGGAGAAGAGGGAAGATTGCCC		TGTG	NFLLSNKYIDTS
					AGCCATGGAGGTATGCGGAGGGAGTGTACA		CATC	MQKGGIPGVPG
					TCCCAAAAGAGGAGAAGTCGGAGAACATC		AGGA	CLEHTGVVTQLI

					GACCAGTTTCGAGTCATCTCCTTGCTCAGTG		GATG	REAREGRGDLA
					TGGAGAGCAAAATATTCTTCAGCATTGTGG		TTTCT	VLWLDLTNAYG
					CCAAAAGACTCTCCAACTTCCTATTGAGCA		GTAA	SIPHKLVEVALE
					ATAAATACATCGACACGTCTATGCAGAAGG		C (SEQ	KHHVPQKVKDL
					GAGGCATACCAGGAGTCCCAGGCTGCCTGG		ID NO:	IIDYYSKFSLRVS
					AACACACAGGCGTGGTAACTCAGCTCATTA		1288)	SGQLTSDWHQL
					GGGAGGCAAGAGAAGGCAGGGGGGACCTG			EVGIITGCTISVT
					GCTGTGTTGTGGTTGGATCTCACCAATGCCT			LFALAMNMMV
					ATGGCTCAATACCCCACAAGCTGGTGGAGG			KAAETECRGPLS
					TCGCACTGGAGAAACATCATGTACCCCAGA			KSGVRQPPIRAF
					AGGTGAAAGACCTCATCATCGACTATTACA			MDDLTVTTTSV
					GCAAGTTCAGCTTGAGAGTCTCCTCTGGCC			PGARWILQGLE
					AGTTAACATCAGATTGGCACCAGCTTGAGG			RLVAWARMSFK
					TAGGAATAATCACTGGTTGCACCATCTCAG			PAKSRSLVLRKG
					TGACCCTCTTTGCACTGGCAATGAACATGAT			KVRDEFRFRLG
					GGTCAAAGCAGCTGAGACAGAGTGCAGAG			QHQIPSVTERPV
					GCCCCCTCAGCAAGTCCGGAGTAAGGCAAC			KSLGKAFNCSL
					CTCCCATCAGAGCCTTCATGGACGACCTCA			NDRDSIRETSTA
					CAGTGACAACAACGTCGGTACCAGGAGCAA			MEAWLKAVDK
					GATGGATCCTCCAAGGGTTGGAGAGGCTCG			SGLPGRFKAWV
					TGGCATGGGCACGCATGAGCTTCAAACCTG			YQHGILPRLLWP
					CAAAATCCAGATCCTTGGTGCTTAGGAAAG			LLIYEVPMTVVE
					GCAAAGTCAGAGATGAGTTCCGCTTCAGGC			GFEQKVSSYLRR

					TGGGACAACACCAAATCCCATCAGTCACTG			WLGLPRSLSNIA
					AGAGACCAGTAAAGAGTCTCGGGAAGGCCT			LYGNTNKLKLP
					TTAACTGTAGCCTCAATGACAGAGACTCCA			FGSVREEFIVAR
					TCAGGGAAACCAGCACTGCCATGGAGGCTT			TREHLQYSGSR
					GGTTGAAAGCAGTGGATAAATCAGGGCTCC			DAKVSGAGIVIR
					CTGGAAGATTTAAGGCTTGGGTTTACCAAC			TGRKWRAAEAV
					ATGGAATCCTTCCAAGACTCCTCTGGCCCTT			EQAETRLKHKAI
					GCTAATCTATGAGGTCCCCATGACTGTGGTT			LGAVAQGRAGL
					GAAGGTTTTGAACAAAAGGTGAGCAGCTAT			GSLAATRYDSA
					CTACGCAGATGGCTGGGATTGCCACGCAGC			SGRERQRLVQE
					CTAAGTAACATCGCTCTGTATGGGAACACC			EVRASVEEERTS
					AACAAGCTCAAACTTCCTTTTGGCTCAGTCA			RAVAMRQQGA
					GGGAGGAGTTCATTGTGGCACGGACACGAG			WMKWEQAMER
					AACATCTGCAGTACTCTGGATCCAGAGATG			NVTWKDIWTW
					CGAAAGTGTCCGGGGCAGGGATTGTCATCA			NPLRIRFLIQGV
					GGACAGGGAGAAAGTGGAGGGCAGCAGAG			YDVLPSPSNLYI
					GCAGTCGAACAAGCGGAAACCCGGCTGAA			WGRVETPACPL
					GCACAAGGCCATCCTGGGGGCAGTAGCACA			CSKPGTLEHILSS
					AGGCAGAGCTGGACTTGGGAGCCTAGCAGC			CSKALGEGRYR
					AACCCGATACGACTCGGCCAGTGGGAGGGA			WRHDQVLKSIA
					GAGGCAGAGGCTGGTGCAGGAGGAGGTGC			EAISKGIKDSRY
					GTGCTTCAGTTGAGGAGGAGAGAACCAGCA			RQATAKVIQFIK
					GAGCAGTGGCCATGCGGCAACAAGGTGCCT			EGQRPERTAKN

					GGATGAAGTGGGAGCAGGCGATGGAGCGG AATGTCACCTGGAAGGACATCTGGACATGG AACCCCCTGAGAATCAGGTTCTTGATCCAA GGGGTCTACGACGTTCTTCCCAGCCCATCG AACCTGTACATATGGGGCAGAGTAGAGACA CCTGCATGCCCGCTGTGTTCCAAGCCAGGG ACACTAGAACATATTTTGAGCAGCTGTTCC AAGGCACTAGGTGAAGGTCGGTATCGATGG AGACACGATCAGGTCCTTAAATCCATTGCT GAGGCAATCAGCAAGGGGATCAAGGACAG TCGATACCGCCAAGCCACGGCCAAGGTCAT TCAGTTCATCAAGGAAGGACAAAGGCCAGA GAGAACAGCAAAGAACTGCTCTGCTGGGTT GCTCTCCACGGCCCGAGACTGGGTGATGAC AGTTGATCTTGAGAGGCAGCTAAAGATTCC ACCACACATCACCCAGTCTACGTTGAGACC TGACATAATCTTGGTCTCTGAGGCCACAAA GCAATTAATCCTGCTGGAGCTGACGGTGCC CTGGGAGGAGAGGATGGAGGAGGCTCAGG AGAGAAAGAGGGGAAAATATCAGGAGCTA GTGGAGCAATGTAGGGCGAATGGATGGAG GACCAGGTGCATGCCAGTGGAAGTGGGCAG TAGGGGATTTGCCAGCTACACCCTGAGCAA			CSAGLLSTARD WVMTVDLERQL KIPPHITQSTLRP DIILVSEATKQLI LLELTVPWEER MEEAQERKRGK YQELVEQCRAN GWRTRCMPVEV GSRGFASYTLSK AYGTLGITGTNR RRALSNNVEAA EKASRWLWLKR GEQWGQ (SEQ ID NO: 1410)

					GGCCTATGGTACACTGGGAATAACAGGCAC AAACCGAAGAAGAGCCCTAAGCAACAACG TGGAAGCAGCGGAAAAAGCATCCAGATGG CTCTGGTTGAAGAGGGGGGAACAGTGGGGG CAGTAGCATGCCACTTGGACACAGGCCGGG GTCTGATCAGCCTCGGTCGGGTCGCCTGGA GGAGGGTGTCTGTTGCAAGACCCGAAACAC CCTGTGAGCCCAGGAAACAACACTGATGAT GTGTCCAAGGTTGTGCATCAGGAGATGTTT CTGTAAC (SEQ ID NO: 1042)			
HERO	HEROFr	.		Takifugu rubripes	AGACTAGGTGACAACCAAGAACAGTTWGG TCGACTACTGGAAAGACAGTTGGCAGCTCG GAAAGACGGCACCCGGGACAGTATGGGTTA GCACCCCAGCCTGTATCTTTCGCGAGAAGG AACCCAAACAAGCTACGGAAAGCCCTACAG AGAAACACCCCCAGGAGATCCCGAGAGGG GGGGAGGATGAGATCTCCAATCGGACGGAC CTAACGTTAATGACCCCTGCAATGGAAATG ACTACGACAGTAACATGTATCTGCAGCAAG CTGTGCAAGAACCAGCGTGGCTTAAAGATC CATCAGGCCAGAATGAAATGTCTGGAGCGG GAGGTTGAGGTGCAACGCACAGGTCCTGGA CCTGGTGAGACGCAGGAGGAGCCCGGACA	AGACTA GGTGAC AACCAA GAACAG TTWGGT CGACTA CTGGAA AGACAG TTGGCA GCTCGG AAAGAC GGCACC CGGGAC	TGAT CACC CCGG CTGG GTCG CCTG GGCG AGGG TGTA TGAT GTCG TGAG ACCC	MTPAMEMTTTV TCICSKLCKNQR GLKIHQARMKC LEREVEVQRTGP GPGETQEEPGQE ATHRSQSLHVPE PPNPNRVVQQQ RIKWPPANRRSE WLQFDEDVSNII QATAKGDVDSR LQAISTIIVSYGS ERFGRIEKGNTE TTSYTMNRRSF

					GGAGGCAACCCACAGATCCCAGTCCCTCCA	AGTATG	GAAA	KIHQLRKELRTL
					CGTACCGGAGCCTCCCAACCCTAACAGAGT	GGTTAG	CACC	KKQFKRAXDGD
					AGTTCAACAGCAGCGGATTAAGTGGCCCCC	CACCCC	CTAT	KQALKELYNILR
					AGCAAATAGACGGAGTGAGTGGCTGCAGTT	AGCCTG	GAAC	KKLKTLRRAEW
					TGATGAGGATGTGTCCAACATCATCCAAGC	TATCTT	CCAG	HRRRGRERARK
					CACAGCCAAAGGAGATGTCGACAGCAGACT	TCGCGA	GATA	RAAFIANPFRFS
					CCAGGCGATAAGTACCATCATCGTCAGCTA	GAAGGA	CATC	KQLLGDKRSGR
					TGGCTCAGAAAGATTTGGACGGATCGAGAA	ACCCAA	CTGA	LECSREEVNRFL
					GGGCAACACTGAGACCACCTCTTACACCAT	ACAAGC	CGAT	QNTMSDPLRGQ
					GAACCGCAGGTCCTTTAAGATACACCAACT	TACGGA	GTGT	DLGPNRALISPA
					GCGCAAGGAGCTGCGAACCCTCAAGAAACA	AAGCCC	CCCA	PPSAEFKLAEPS
					GTTCAAGAGAGCTKCTGATGGGGACAAGCA	TACAGA	GTGC	LKEVEEVIKAAR
					AGCTTTAAAAGAGCTGTATAACATCCTGCG	GAAACA	ATCC	SASSPGPSGVPY
					GAAGAAGTTGAAAACTCTCCGCAGAGCAGA	CCCCCA	AGGA	LVYKRCPEILRH
					GTGGCACAGGAGGCGCGGGAGAGAGAGAG	GGAGAT	GATG	LWKALKVIWRR
					CAAGGAAGCGAGCAGCCTTCATTGCCAATC	CCCGAG	TAKC	GRVADQWRCA
					CCTTCCGGTTTTCTAAACAGCTGCTCGGGGA	AGGGGG	TTTA	EGLWIPKEEDSK
					CAAGCGGAGTGGCCGACTTGAGTGCTCAAG	GGAGGA	AGT	NINQFRTISLLSV
					GGAGGAAGTGAATCGCTTCCTCCAAAACAC	TGAGAT	(SEQ	EGKVFFSIVSRR
					CATGAGCGACCCACTGAGGGGTCAAGACCT	CTCCAA	ID NO:	LTEFLLKNNYID
					AGGACCCAACAGAGCGCTCATCAGCCCTGC	TCGGAC	1289)	TSVQKGGIPGVP
					CCCACCATCGGCAGAGTTCAAGCTGGCAGA	GGACCT		GCLEHNGVVTQ
					GCCTAGTTTGAAGGAGGTTGAAGAAGTCAT	AACGTT		LIREAHESKGEL

					CAAGGCAGCCCGTTCTGCATCTTCCCCGGG CCCCAGTGGTGTACCTTACCTCGTCTACAAG CGCTGTCCAGAAATTCTCCGGCATCTGTGG AAGGCCTTGAAAGTGATCTGGCGAAGGGGG AGAGTAGCCGACCAGTGGAGGTGTGCTGAG GGACTTTGGATACCCAAGGAGGAGGACTCG AAAAACATCAACCAGTTTCGGACTATCTCA CTACTGAGTGTGGAAGGGAAGGTGTTTTTT AGCATCGTCTCCCGAAGACTGACCGAGTTT CTCCTCAAGAACAACTACATCGACACTTCA GTGCAGAAGGGTGGGATCCCTGGAGTCCCC GGCTGTCTAGAGCACAATGGTGTAGTCACA CAGCTCATCAGAGAGGCCCATGAGAGCAAA GGAGAACTAGCGGTTTTGTGGTTGGACCTG ACTAACGCCTACGGGTCCATCCCACACAAG CTAGTTGAGCTTGCGCTACACCTACACCATG TTCCCAGTAAGATCAAGGACCTGATTCTGG ATTACTATAATAACTTCAGGCTCAGGGTCA CTTCAGGGTCAGTAACCTCAGACTGGCATC GCCTTGAGAAAGGAATAATAACAGGCTGTA CCATCTCCGTCGTTCTCTTCGTACTGGCGAT GAATATGGTGGTAAAGGCGGCTGAGGTGGA GTGCAGAGGGCCTCTATCCAGATCAGGTGT	A (SEQ ID NO: 1166)		AVLWLDLTNAY GSIPHKLVELAL HLHHVPSKIKDL ILDYYNNFRLRV TSGSVTSDWHR LEKGIITGCTISV VLFVLAMNMV VKAAEVECRGP LSRSGVRQPPIR AYMDDLTVTTT SVPGCRWILQGL ERLILWARMSF KPTKSRSMVLK KGKVVDKFRFSI SGTVIPSITEQPV KSLGKLFDSSLK DTAAIQKSTEEL GGWLTKVDKSG LPGRFKAWIYQ YSILPRVLWPLL VYAVPVTTVESF ERKISSFLRRWL GLPRSLNSAALY

					TCGACAGCCCCCCATAAGAGCCTACATGGA			GTSNTLQLPFSG
					CGACCTTACCGTCACAACAACATCAGTCCC			LTEEFKVARTRE
					AGGGTGTAGGTGGATCTTGCAGGGTTTGGA			ALQYRDSRDCK
					GAGACTCATCCTATGGGCTAGGATGAGTTT			VSSAGIEVKTGR
					TAAGCCCACCAAGTCAAGGTCCATGGTACT			KWKAEKAVXV
					GAAGAAGGGGAAAGTGGTGGACAAATTCC			AESRLRQKALV
					GATTCTCAATCTCAGGAACCGTAATTCCATC			GAVATGRTGLG
					GATCACGGAGCAACCAGTCAAGAGCCTGGG			YFPKTQVSHAR
					AAAGCTCTTTGACTCCAGCCTGAAGGACAC			GKERNHLLQEE
					TGCTGCTATCCAGAAGTCTACGGAAGAGCT			VRAGVEEERVG
					TGGAGGGTGGCTCACTAAGGTGGACAAGTC			RAVGLRQQGA
					TGGCCTGCCTGGTAGATTTAAAGCCTGGAT			WTRWESALQRK
					CTACCAGTACTCCATCCTTCCCAGAGTCCTG			VTWSNIMQADF
					TGGCCTCTCCTCGTGTATGCAGTCCCAGTAA			HRVRFLVAAVY
					CAACAGTGGAATCCTTTGAAAGGAAGATCA			DALPSPANLHA
					GCAGCTTTCTGCGCAGATGGCTGGGTCTTCC			WGKSETPTCSLC
					TCGCAGCCTCAACAGCGCTGCACTGTACGG			SGRGSLEHLLSS
					GACAAGTAACACCCTGCAGCTACCCTTCAG			CPKSLADGRYR
					TGGGCTCACTGAAGAATTTAAGGTGGCACG			WRHDQVLKAV
					CACAAGAGAAGCCCTACAGTACAGAGACTC			AESIALAISTXK
					CAGGGACTGCAAGGTGTCATCAGCCGGGAT			HHHAPKKAISFI
					TGAGGTGAAGACAGGAAGGAAGTGGAAGG			KAGERPRAGPQI
					CAGAAAAGGCAGTGGAKGTGGCTGAGTCAC			TTGLLHTAXDW

					GCCTAAGGCAAAAGGCACTAGTTGGGGCCG TGGCAACAGGAAGAACAGGCTTGGGCTACT TCCCAAAGACCCAAGTCAGCCATGCCCGGG GCAAAGAGAGAAACCACCTACTTCAGGAGG AGGTCCGAGCAGGCGTGGAGGAAGAGCGA GTGGGTAGGGCAGTGGGACTCCGGCAGCAG GGGGCATGGACAAGGTGGGAGAGCGCGTT ACAGCGCAAAGTTACCTGGTCAAACATCAT GCAGGCAGACTTCCACCGCGTCCGGTTCCTT GTGGCGGCAGTCTACGATGCCCTCCCCAGC CCAGCAAACCTCCATGCGTGGGGAAAGAGT GAGACACCCACCTGTTCCCTTTGCTCCGGAA GAGGCTCCCTGGAACATCTCCTTAGCAGCT GCCCAAAGTCCCTGGCTGATGGTCGCTATC GCTGGCGCCACGACCAGGTACTCAAAGCAG TGGCTGAGAGCATAGCCTTGGCCATTAGCA CCASCAAACACCATCATGCTCCGAAGAAGG CAATCTCCTTCATAAAAGCTGGAGAGAGAC CTCGTGCAGGCCCACAGATAACAACGGGAC TCCTCCACACAGCTMCTGATTGGCAACTGC ACGTTGACCTGGGAAAACAACTGATATTCC CCCAGCACATCGCAACAACGTCTCTACGGC CAGACATGATCATCATCTCAGAGGCTTCGA			QLHVDLGKQLIF PQHIATTSLRPD MIIISEASKHLIM LELTVPWEERIE EANERKRAKYQ ELVEECRGRGW RTFYEPIEVGCR GFAGRSLCKAF GRLGVTGTAKK RAIKXASEAAER ATRWXWLKRA DPWVATGTQAG S (SEQ ID NO: 1411)

					AACACCTGATCATGCTGGAGCTTACAGTGC CCTGGGAAGAGCGGATTGAGGAAGCCAAC GAAAGGAAACGTGCCAAGTATCAGGAGCTG GTGGAGGAGTGCAGGGGCAGGGGCTGGAG GACCTTCTACGAGCCCATAGAAGTTGGCTG TAGAGGCTTTGCAGGACGCTCCCTCTGCAA AGCCTTTGGCCGACTGGGAGTCACAGGGAC AGCCAAAAAGAGGGCCATTAAAKCCGCGA GTGAAGCTGCAGAGAGAGCCACGAGGTGGS TGTGGCTKAAAAGGGCAGATCCGTGGGTTG CTACTGGGACACAAGCCGGGTCTTGATCAC CCCGGCTGGGTCGCCTGGGCGAGGGTGTAT GATGTCGTGAGACCCGAAACACCCTATGAA CCCAGGATACATCCTGACGATGTGTCCCAG TGCATCCAGGAGATGTAKCTTTAAGT (SEQ ID NO: 1043)			
HERO	HEROTn	.		Tetraodon nigroviridis	AGATTGGTCTGGCTAAGCCAGTGACGTCCA GGAACAGACTGGCTGACGACCACGAATAGA GTGGTGACAGCTTGGATAGACAGCTGACAG CAGGGAAAGACGGCAACCGGGGCAGGAAG GGCTAGCAACCCAGCCTGCATCTTCCGTGA GGAAGAACCCAAAACTTGCTACGAAGAGCC CGAAGCAAAGATACCCCCAGGGGAGCCCG	AGATTG GTCTGG CTAAGC CAGTGA CGTCCA GGAACA GACTGG	TGAT CACT CCCA GTCG GGTC GCCT GGGT	MATTQASVKPT AVATCVCGKIC KNPRGLKIHQTK MGCLASVQPEQ RARFSLSESREV PARAEPYGPQQP HSPEALGETQEE

					AGAGGGGGGGAGAATGAGCTCCCCAAACG	CTGACG	GAGG	RGQESPHSAQN
					GACGGATAACATGGCAACGACCCAGGCTAG	ACCACG	GGGT	LRAQVAQAPDN
					CGTTAAACCGACAGCGGTTGCCACATGTGT	AATAGA	CTGA	PQHHRRVKWPP
					ATGTGGCAAAATCTGCAAAAACCCACGAGG	GTGGTG	TGTT	ASKVSEWQQLD
					TCTGAAGATCCACCAGACCAAGATGGGGTG	ACAGCT	GAAA	EDLEGILESTAK
					CTTGGCAAGTGTGCAACCAGAGCAGCGCGC	TGGATA	GACC	GGVDRKLQTMT
					AAGGTTCAGCCTCAGCGAGTCGCGGGAGGT	GACAGC	CGAA	TLVISFATERYG
					GCCAGCCAGGGCCGAGCCCTATGGCCCTCA	TGACAG	ACCC	TMEKRAAPEKY
					GCAACCGCATTCTCCTGAGGCCCTTGGTGA	CAGGGA	CCGA	TKNRRAEKISQL
					GACGCAGGAGGAGCGGGGCCAGGAGTCAC	AAGACG	TGAC	RQELRVLKKQF
					CCCACAGTGCCCAGAACCTCCGTGCTCAGG	GCAACC	CCCA	KGASEDQKPGL
					TAGCACAAGCGCCAGACAACCCACAACACC	GGGGCA	GGTA	AELRCTLRKKLL
					ACCGGCGGGTTAAGTGGCCCCCAGCCAGCA	GGAAGG	CTAT	TLRRAEWHRRR
					AAGTGAGCGAGTGGCAGCAGCTTGATGAGG	GCTAGC	CACT	AKERAKKRAAF
					ATTTGGAAGGTATTCTGGAGTCCACCGCAA	AACCCA	GACG	LANPFGFTKQLL
					AAGGTGGAGTAGACAGAAAACTCCAAACA	GCCTGC	ATGT	GQKRSAHLECA
					ATGACCACGCTGGTCATCAGCTTTGCCACC	ATCTTC	GTCC	KEEVDSYLHDT
					GAGAGATATGGTACAATGGAGAAACGCGCT	CGTGAG	AAGA	FSDAERENSLGE
					GCTCCAGAGAAGTACACCAAAAACCGCAGG	GAAGAA	CATG	CRVLISPPEPACS
					GCAGAAAAGATCTCCCAACTGCGGCAGGAA	CCCAAA	CATC	FNTKAPTWKEIQ
					CTTCGGGTCCTGAAAAAGCAGTTCAAGGGC	ACTTGC	AATA	TVVRAARNNSA
					GCCAGCGAGGATCAGAAGCCAGGATTGGCA	TACGAA	GGTG	PGPNGVPYLVY
					GAGCTTCGTTGCACCCTTAGGAAAAAACTG	GAGCCC	TATTT	KRCPKLLARLW

					CTTACCCTTCGCCGAGCAGAGTGGCACCGG AGACGGGCCAAGGAAAGAGCCAAGAAACG CGCTGCATTTTTAGCCAACCCTTTTGGGTTC ACTAAACAACTTTTAGGCCAGAAGCGTAGC GCCCACTTGGAATGTGCAAAAGAGGAGGTT GATTCCTACCTCCACGACACATTCAGTGAC GCAGAACGGGAGAACAGCCTAGGCGAATG TAGAGTGCTGATCAGTCCACCTGAGCCAGC CTGCAGTTTCAACACCAAGGCTCCAACTTG GAAAGAAATCCAAACTGTGGTCAGGGCTGC AAGAAACAACTCAGCTCCTGGACCCAATGG AGTCCCATATCTGGTGTACAAAAGATGCCC CAAACTCCTAGCCCGGCTCTGGAAGATCCT AAGGGTGATCTGGAGAAGGGGGAAGGTCG CCCATCAATGGAGATGGGCGGAAGGGGTGT GGGTTCCGAAGGAGGAGAAGTCAACCTTGA TAGAGCAGTTTAGGACCATCTCACTGCTCA ATGTCGAGGGGAAGATATTCTTTAGTATCCT CTCCCATCGTCTATCAGACTTCCTCCTTAAG AACCAGTACATCGACTCCTCGGTGCAAAAG GGGGGGATCCCTGGGGTACCAGGGTGTTTA GAACACTGTGGCGTGGTGACACAACTAATT AGGGAGGCGCGCGAAGGGAGAGGTAGCCT	GAAGCA AAGATA CCCCCA GGGGAG CCCGAG AGGGGG GGAGAA TGAGCT CCCCAA ACGGAC GGATAA C (SEQ ID NO: 1167)	AGAA ATC (SEQ ID NO: 1290)	KILRVIWRRGKV AHQWRWAEGV WVPKEEKSTLIE QFRTISLLNVEG KIFFSILSHRLSD FLLKNQYIDSSV QKGGIPGVPGCL EHCGVVTQLIRE AREGRGSLAVL WLDLANAYGSI PHKLVEMALAR HHVPGPIKTLIM DYYDSFHLRVT SGSVTSEWHRL EKGIITGCTISVII FALAMNMLAKS AEPECRGPITKS GIRQPPIRAFMD DLTVTTTSVPGC RWILQGLERLM TWARMRFKPGK SRSLVLKAGKV TDRFRFYLGGT

					GGCCGTACTTTGGCTGGACTTAGCTAACGCT			QIPSVSEKPVKS
					TATGGCTCCATACCCCACAAGCTGGTGGAA			LGKMFDGSLKD
					ATGGCATTAGCGAGGCACCATGTCCCAGGC			AASIRETNDQLG
					CCGATCAAGACTCTGATCATGGACTACTAT			HWLTLVDKSGL
					GATAGCTTCCACCTGAGAGTCACGTCAGGC			PGKFKAWVYQH
					AGTGTCACATCTGAATGGCACCGACTAGAG			GILPRILWPLLV
					AAAGGGATCATCACTGGATGCACCATCTCA			YEFPISTVEGLE
					GTGATAATATTCGCCCTGGCCATGAATATG			RRVSSCLRRWL
					CTGGCCAAGTCGGCTGAGCCAGAGTGCAGA			GLPRSLSSNALY
					GGACCCATAACCAAGTCAGGCATTCGCCAG			GNNNKLTLPFSS
					CCCCCCATCAGAGCATTCATGGATGATCTG			LAEEFMVTRAR
					ACAGTAACAACAACGTCAGTTCCAGGGTGC			EVLQYRESKDP
					CGTTGGATCCTCCAGGGCCTGGAGAGGCTT			KVALAGIEVRT
					ATGACTTGGGCCCGTATGCGCTTTAAACCTG			GRRWRAQEAV
					GAAAATCTAGGTCCTTAGTCCTGAAGGCAG			DQAESRLHHKE
					GGAAGGTGACCGACCGCTTCCGCTTCTACC			LVGAVATGRAG
					TGGGAGGCACCCAGATTCCATCAGTCTCTG			LGTTPTTHLSRL
					AGAAACCGGTGAAAAGCCTAGGTAAAATGT			KGKERRDQVQL
					TCGACGGCTCCTTAAAGGATGCCGCTTCCAT			EVRASIEEQRAS
					CAGGGAAACCAATGATCAGCTGGGGCACTG			QWVGLRQQGA
					GCTGACGTTGGTCGATAAGTCAGGTCTTCC			WTRWEEAMAR
					GGGGAAATTCAAGGCATGGGTATACCAGCA			KISWPELWRAEP
					TGGTATCCTACCTAGGATACTGTGGCCACTG			LRIRFLIQSVYD

					CTGGTGTATGAATTTCCAATTTCCACCGTGG			VLPSPSNLFLWG
					AAGGGCTTGAGAGGAGGGTCAGCAGCTGCC			KVESPSCPLCQG
					TCAGGCGTTGGCTGGGACTACCTAGGAGTC			RGTLEHILSSCP
					TGAGCAGCAATGCCCTCTACGGTAACAACA			KALGEGRYRWR
					ACAAGCTGACACTCCCCTTCAGCAGCCTGG			HDQVLKAIAESI
					CAGAGGAATTCATGGTTACCAGAGCTAGGG			SSAMEYSKRLPL
					AAGTTCTCCAGTACAGGGAGTCCAAGGATC			PGRGVRFVRAG
					CCAAGGTAGCTCTTGCCGGCATTGAGGTGC			EQPPPQPRAQPG
					GGACTGGCAGAAGGTGGAGGGCTCAGGAG			LLATARDWQLR
					GCAGTGGACCAGGCAGAATCTCGGCTGCAC			VDLGKQLKFPE
					CACAAAGAGCTTGTGGGAGCCGTGGCGACT			NIVETNLRPDIV
					GGCCGTGCAGGCCTGGGAACAACACCGACC			LHSQSSKQVILL
					ACCCACCTCAGCAGGCTCAAGGGCAAGGAA			ELTVPWEERME
					AGGCGGGATCAGGTCCAACTAGAAGTGAGG			EAYERKAGKYA
					GCCAGTATTGAGGAACAGCGAGCTAGTCAG			ELVEDCRRAGW
					TGGGTGGGGCTGAGGCAGCAAGGCGCTTGG			RSRCLPIEVGGR
					ACTAGGTGGGAAGAGGCCATGGCCAGAAA			GFAGKSLCKAFS
					GATCTCATGGCCTGAGCTGTGGAGGGCTGA			LLGITGMRRRK
					GCCCTTGCGCATCCGCTTCCTTATTCAGTCA			AICAASEAAERA
					GTTTATGACGTCTTGCCCAGCCCATCAAACC			SRWLWIQRDKP
					TCTTCCTGTGGGGCAAGGTGGAATCCCCAT			WTSASWTQAGN
					CATGTCCCTTGTGCCAGGGAAGGGGCACCT			(SEQ ID NO:
					TGGAGCACATCCTCAGCAGCTGTCCCAAAG			1412)

					CACTTGGAGAGGGTCGCTATCGCTGGCGTC ACGACCAGGTGCTGAAGGCAATCGCTGAGT CTATCAGCTCCGCCATGGAGTACAGCAAGC GCCTACCCTTACCGGGACGCGGAGTTAGGT TTGTCAGGGCCGGTGAACAACCTCCTCCCC AACCAAGGGCCCAACCAGGCCTCCTTGCAA CAGCTAGGGACTGGCAACTAAGGGTTGACC TGGGGAAACAATTAAAGTTCCCGGAAAACA TCGTAGAAACCAACCTGAGGCCAGACATTG TTCTGCACTCACAGTCGTCCAAGCAAGTTAT TTTGCTGGAGCTGACTGTGCCCTGGGAGGA GAGAATGGAGGAAGCGTATGAAAGGAAGG CAGGGAAGTACGCTGAGCTGGTGGAGGATT GCCGCAGAGCAGGGTGGCGCAGTAGATGCC TGCCTATAGAGGTTGGGGGTAGGGGCTTTG CAGGGAAGTCACTCTGCAAGGCCTTTAGCC TCCTGGGCATCACAGGCATGCGCAGGAGGA AAGCCATCTGCGCGGCCTCAGAGGCTGCAG AGAGGGCGTCCAGATGGCTGTGGATCCAGC GGGACAAGCCGTGGACGAGCGCTTCTTGGA CACAGGCCGGGAACTGATCACTCCCAGTCG GGTCGCCTGGGTGAGGGGGTCTGATGTTGA AAGACCCGAAACCCCCGATGACCCCAGGTA			

					CTATCACTGACGATGTGTCCAAGACATGCA TCAATAGGTGTATTTAGAAATC (SEQ ID NO: 1044)			
NeSL	LIN9_SM	.		Schmidtea mediterranea	AAACGACATCATGAACGCTTGGCCGCAACA ATCCAGTTATCCCTGCGGTAACATTGTGGA ACTCATAAGACAAGTACTAAAAGAAGAATT AGAAAAATTAGAAGAAAAAATTGAAAATA ATTTATTTATAAAATTTAAAAATTTAAATAA ATTTAAAAATTTAAATTTAAATTTAAATGAA GATAAAAATTTATTTAATCCAATAAATAAT CAAGAAAATCAAGAAAATGATGGATTCAAG ACAATTAAATACTCCAAAAATAAGAAAATA TCAGAACCCAAAAATGACAAACGACATCAT GAAAAGCTACAACTACGCGGTTTTGAGCGA TGTCACGCCTCAAGAAACCACTCAAACAAC AACCCACTTAAATGTCGATATAGACAATGA AACCACCCAACCAAAACAGCCACTTACGAA GTCTGGCAAACCAAAATCTAAACCAATTGC GGTATCATACAAATTTAAAGATGCCACCTT CATCTGGGACACTACCCCACAAACAAATCC ACCAAGAGATTGCACCAAACTTATTGATAA AACAAGACCAAGAAAGACCATCTTCAAAAA ATCAGCATTTCAAAGCTACCTCAAAAAAGA	AAACGA CATCAT GAACGC TTGGCC GCAACA ATCCAG TTATCC CTGCGG TAACAT TGTGGA ACTCAT AAGACA AGTACT AAAAGA AGAATT AGAAAA ATTAGA AGAAAA AATTGA AAATAA	TAAA ATGG CAAA AAGA TATTT CAAG ATGA ATTG TGGA CTCA TCTA AAAA ATGA CCAC CTTG AGTC CAAA TATG CCTA GCTA	MMDSRQLNTPK IRKYQNPKMTN DIMKSYNYAVL SDVTPQETTQTT THLNVDIDNETT QPKQPLTKSGKP KSKPIAVSYKFK DATFIWDTTPQT NPPRDCTKLIDK TRPRKTIFKKSA FQSYLKKELSNE TFVEVKTFLMA THKYRFKDENS RLLAYRIINRYV METANEFKETEF DMARFAKFFTIP ENWLKHLKPYS TATETSPADRIK VQKLVDLTCRY PFKTQEEQTSVA

					ACTGTCCAATGAGACATTTGTGGAAGTAAA	TTTATTT	TCAT	NFLHFFTQRSIIG
					AACCTTCCTCATGGCAACTCACAAATATCGT	ATAAAA	GGTT	ISRDYKFQKFIPF
					TTTAAAGACGAAAACTCAAGACTCTTGGCA	TTTAAA	GCTG	MARKNTRPETT
					TACCGAATAATTAATCGCTATGTCATGGAG	AATTTA	ATGG	STMVTTSPTEQN
					ACAGCAAATGAATTCAAAGAAACCGAATTT	AATAAA	AAAC	RLPMVIITPLEEP
					GACATGGCTCGCTTTGCCAAATTCTTCACAA	TTTAAA	AGTA	KSEHRRPEKRG
					TCCCAGAGAATTGGTTAAAACATCTAAAAC	AATTTA	AGGC	ASNDTIVLSDEE
					CATACTCTACAGCTACCGAAACATCACCGG	AATTTA	ACCT	FPLLKRRTLPTR
					CTGATAGAATAAAAGTACAAAAATTAGTGG	AATTTA	GATA	KSKNPTGAGNV
					ATCTCACATGCAGATACCCATTCAAAACTC	AATGAA	GCTA	PTETECTDEVKF
					AAGAAGAGCAAACAAGTGTAGCAAACTTCC	GATAAA	ACTTT	ILNNEYQIECKE
					TACACTTCTTCACCCAAAGATCAATAATTGG	AATTTA	TCAC	CGKVWENVRN
					AATCTCAAGAGATTATAAATTCCAAAAATT	TTTAAT	TGTG	GLNHLRQKHDF
					TATACCATTTATGGCAAGAAAAAACACCAG	CCAATA	AATA	PNRTDVMVSCV
					GCCGGAGACAACCTCCACTATGGTTACGAC	AATAAT	TCTTC	RCEVPIKGAECV
					TTCTCCAACAGAACAAAACAGACTACCAAT	CAAGAA	AGAT	NHIKNHKKDDK
					GGTAATAATCACACCACTTGAAGAACCAAA	AATCAA	ATTC	EESEAGSLVANT
					AAGTGAACATCGTAGACCAGAGAAAAGAG	GAAA	ACAG	QDIPNESSLSQA
					GCGCAAGCAATGACACAATTGTGCTTAGCG	(SEQ ID	TGAC	AIEVYLRNILKM
					ACGAAGAGTTCCCACTACTTAAAAGGAGAA	NO: 1168)	ACGA	KENQERNIQYLE
					CTCTTCCAACCAGAAAATCCAAAAATCCTA		AAGG	PSTANFLINRNL
					CTGGTGCAGGAAATGTACCWACAGAAACC		ACAC	RAFYQNVKIEKL
					GAATGCACTGATGAAGTTAAATTCATCCTC		CACT	IGWEQVIWLIH

					AACAATGAATACCAAATAGAATGTAAAGAG		AGTA	WNKCHWIVYLA
					TGTGGAAAAGTGTGGGAAAACGTACGAAAT		AAAA	NCDSKTSVILDS
					GGATTAAACCACCTTCGTCAAAAACACGAT		CCAC	DNQMTLQQRCN
					TTCCCAAACCGAACAGATGTTATGGTATCTT		TAGT	IKAKFDKFLEGT
					GCGTAAGATGTGAAGTACCGATCAAAGGAG		TTTTT	FEEKTVLGTLER
					CAGAATGTGTAAATCACATTAAAAATCACA		CTGA	KVPQQPNNFDC
					AAAAAGATGACAAAGAAGAAAGTGAAGCS		CACC	GIYVIQYISDFLK
					GGGAGTCTTGTGGCTAACACTCAAGACATC		TCTTG	DPQRIDYHTPDS
					CCAAATGAAAGTAGCTGACTGTCACAAGCC		CTAC	KRIRKEIGELILE
					GCAATCGAAGTATATCTGAGGAATATTCTG		AAAC	EMKNPASKIKNP
					AAAATGAAAGAAAACCAGGAAAGGAATAT		TCTGT	NKEIQSLLQKFR
					TCAATATCTTGAACCTAGTACTGCGAATTTC		AAAA	LLQINVNDVFH
					CTCATAAATAGGAACCTCAGAGCATTTTAT		ATCA	WFAAEYQKSLP
					CAAAACGTCAAAATCGAAAAGCTTATCGGA		AAAG	KIRTKRDGKLN
					TGGGAACAAGTCATCTGGCTTATACATTGG		GATC	KLSCSYQIQRLF
					AACAAATGTCATTGGATTGTATACCTAGCT		GATA	GLAPKRAVKEIY
					AATTGCGACTCAAAAACCTCTGTTATCTTGG		GGCC	FQETSTADLETR
					ACTCTGACAACCAAATGACATTACAGCAAA		GCGC	VLNEHFKKDES
					GATGTAACATAAAAGCCAAATTTGACAAAT		TTTCA	TMKECKIKNGN
					TCCTAGAAGGTACCTTTGAAGAAAAAACAG		CGGT	HYQDWITKAQI
					TGCTTGGAACCCTAGAAAGAAAAGTTCCTC		CTGT	DNKEILEALKNS
					AGCAACCAAACAACTTCGATTGCGGTATAT		ATTC	TDSAPGEDNIPL
					ATGTGATACAATACATCAGCGACTTTCTTAA		GTAC	RQWIIWNNDGV

					AGACCCACAAAGAATAGATTATCATACACC		TGAA	LFDMFNYIKRTH
					CGACTCCAAAAGAATTAGAAAAGAAATAG		AATC	DIPDMWKNYTT
					GAGAATTAATATTAGAAGAAATGAAAAACC		AAGA	TLLIKPGKSQES
					CTGCCTCAAAAATCAAAAATCCAAACAAAG		TCAA	NIPANWRPISILP
					AAATACAATCTTTACTCCAAAAATTCAGAC		GGAA	TSYRIFMKVLNK
					TACTGCAAATCAATGTGAATGATGTATTCC		GCTTT	RVLEWANRGEL
					ATTGGTTTGCGGCTGAATACCAAAAATCTCT		TCCC	ISKWQKAVDKA
					ACCGAAGATACGTACCAAAAGAGATGGAA		CTTTT	NGCDEHSYVIQ
					AACTGAATAAACTAAGCTGCTCCTATCAAA		AGTC	ALIEKANRSYYK
					TCCAAAGATTATTTGGTCTAGCTCCTAAAAG		AACA	NEQCHLAFLDL
					AGCAGTCAAAGAAATATATTTCCAAGAAAC		CCAG	ADAFGSIPFQVI
					CTCTACAGCAGACTTGGAAACAAGAGTTCT		GTTTC	WHTLKNMGMD
					AAATGAACATTTCAAAAAGGATGAATCAAC		TGTC	EETINLLKEIYK
					GATGAAAGAATGTAAAATAAAAAATGGAA		CTAG	DCSTKYKCGKN
					ACCATTACCAAGACTGGATAACAAAGGCCC		TTGA	ESEKIKITKGVR
					AAATTGATAATAAAGAAATATTGGAAGCCC		GCTT	QGCPLSMTLFSL
					TAAAAAACAGTACAGATTCTGCCCCCGGAG		CCCTT	CIQYLIQGIAEK
					AAGATAACATTCCTCTGAGGCAATGGATAA		GGGA	KKGATIAGQEV
					TCTGGAACAACGACGGTGTCCTCTTTGATAT		CATC	CILAYADDLVIV
					GTTTAACTACATCAAAAGGACACACGATAT		TGCG	ANTAKDMQML
					CCCAGATATGTGGAAAAACTACACCACAAC		TTAC	LTTIENLAKQAD
					ACTACTTATAAAACCCGGAAAAAGCCAAGA		CATTT	LIFKPAKCGYYR
					AAGCAACATCCCCGCTAATTGGAGGCCAAT		GACA	DPRDKKSMMKI

					ATCGATATTGCCAACAAGCTATCGTATATTT		GATG	YGKEISIVDEKN
					ATGAAAGTCCTAAATAAAAGAGTACTAGAA		TACC	VYTYLGVRIGD
					TGGGCTAATAGAGGAGAACTGATATCAAAA		GCCC	TKKKDLNVRFE
					TGGCAGAAAGCCGTAGACAAAGCTAATGGA		CAGT	EVKKKTTAIFKS
					TGTGATGAGCACAGCTATGTCATACAAGCG		CAAA	KLRSDQKLEAY
					CTTATCGAAAAAGCAAACAGAAGCTACTAC		CTCC	NIFCQSKFVYIL
					AAAAACGAGCAATGTCACCTCGCCTTCTTG		CCAC	QGEDIAKTKIET
					GATTTGGCAGATGCTTTTGGAAGCATCCCAT		CTGA	YDEEIKKMIKED
					TCCAAGTAATATGGCATACCCTAAAAAATA		CACT	ILKLQDKSPFTD
					TGGGTATGGATGAGGAAACCATCAACTTGC		GTCC	FVIYSPREKGGL
					TCAAAGAAATCTACAAAGATTGCTCCACAA		TCAA	GITKIIDEQTIQTI
					AATATAAATGTGGAAAGAATGAGTCAGAAA		AACA	NRTAKLLNSSH
					AGATCAAAATTACGAAAGGAGTCCGACAGG		GTTC	RAIRAIIYEELIQ
					GATGCCCATTGTCGATGACCCTCTTCAGCCT		AATT	VANLRGEKEINT
					CTGTATACAATATCTTATACAAGGCATAGC		GCAT	IEEALKWLEGTN
					AGAAAAGAAAAAAGGAGCAACAATTGCAG		CCGA	KYKKNSNAKTT
					GTCAAGAAGTTTGCATATTGGCTTATGCGG		AGAT	WITRVREAFQTL
					ACGACCTAGTAATTGTTGCAAACACAGCAA		CGCA	EKKHKIKVRFVP
					AAGACATGCAAATGCTGTTAACAACAATCG		ATTTT	KENCIGYKIKCD
					AAAATCTGGCAAAACAAGCCGATCTCATAT		TTCA	TQEKIVELDNSK
					TCAAACCGGCAAAATGTGGATATTACAGAG		CTAA	ELSKSLHWMIK
					ACCCAAGAGATAAAAAGTCCATGATGAAGA		AATA	EAYYKEWKALK
					TATATGGCAAAGAAATCAGCATAGTAGACG		AATT	CQGYIISLKTSEF

					AAAAGAATGTTTACACCTACCTAGGTGTAA GAATCGGTGACACAAAGAAAAAAGACCTA AATGTCAGATTCGAAGAGGTCAAAAAGAAA ACGACAGCAATCTTCAAATCGAAATTGCGA AGTGACCAAAAACTAGAGGCATACAACATC TTTTGCCAATCAAAATTTGTGTACATCCTAC AAGGCGAAGATATCGCAAAAACCAAAATTG AAACTTACGACGAAGAAATCAAGAAAATG ATAAAAGAAGATATATTAAAATTACAAGAC AAAAGTCCGTTCACAGACTTCGTTATCTACT CCCCAAGAGAAAAAGGGGGGTTAGGAATA ACAAAGATAATAGATGAACAAACAATTCAA ACTATTAATAGAACGGCAAAACTCCTAAAT AGTAGCCATAGAGCAATCCGGGCTATTATT TATGAAGAGCTAATACAAGTAGCTAACCTA AGAGGAGAAAAAGAAATCAACACCATTGA AGAAGCACTAAAATGGTTGGAAGGTACCAA CAAATACAAAAAGAACTCCAACGCCAAGAC CACCTGGATAACAAGGGTTCGGGAGGCCTT TCAAACTCTAGAAAAGAAACACAAAATCAA GGTTAGATTTGTGCCCAAAGAAAACTGCAT TGGATATAAAATCAAATGCGACACCCAAGA AAAGATAGTGGAGCTTGATAACTCAAAAGA		AACA AAAG TTAA TTAT ACTG CTTC ATTG AGTA AGTA GAAA AACA ATC (SEQ ID NO: 1291)	MEWKMPRGLP DPDWRFLTKVK ANMLDVNMKQ ANQGGRLGSTK CRKCEDKESAS HVINHCASGNW SRVEKHNQVQN ELAKELTKRNIS FEKDSIPKETKE SLRPDLVIRLKD KIMIVDIKCPFD EESAIESARNKN IDKYRELAKEIQ AKTGLQTTVSTF VVCSLGTWDKR NNELLRQMGIR YEESKEMRINMI QKAIHGSRKTY DHHRNFNNG (SEQ ID NO: 1413)

					GTTATCAAAAAGCTTACACTGGATGATAAA AGAGGCATATTATAAAGAATGGAAAGCCCT AAAATGCCAAGGATATATTATAAGCCTAAA AACCTCCGAATTTATGGAGTGGAAAATGCC CAGAGGCCTTCCGGACCCTGATTGGAGATT CCTAACAAAAGTAAAGGCAAATATGTTGGA CGTAAACATGAAACAAGCCAACCAGGGAG GAAGGTTGGGAAGCACAAAATGCCGAAAA TGTGAAGATAAAGAATCGGCAAGCCATGTT ATAAACCACTGTGCCTCAGGTAACTGGAGT AGAGTGGAAAAGCACAACCAGGTGCAAAA TGAGCTAGCAAAAGAACTGACAAAGCGGA ATATCAGCTTCGAAAAGGACAGCATCCCAA AAGAAACAAAAGAGAGCCTAAGACCAGAT TTGGTTATAAGACTCAAAGACAAGATAATG ATAGTGGACATCAAATGCCCATTTGATGAG GAATCTGCTATCGAGAGTGCCAGAAACAAG AACATAGACAAATATCGAGAACTGGCCAAA GAGATCCAAGCAAAAACTGGGTTACAAACA ACAGTCTCAACTTTCGTTGTCTGTTCTTTGG GAACCTGGGATAAGAGGAACAACGAGCTCC TACGGCAGATGGGAATAAGATATGAAGAAT CCAAAGAGATGAGGATCAATATGATCCAAA			

					AAGCCATCCACGGGTCTAGAAAAACCTACG ACCACCACAGAAATTTTAACAATGGTTAAA ATGGCAAAAAGATATTTCAAGATGAATTGT GGACTCATCTAAAAAATGACCACCTTGAGT CCAAATATGCCTAGCTATCATGGTTGCTGAT GGAAACAGTAAGGCACCTGATAGCTAACTT TTCACTGTGAATATCTTCAGATATTCACAGT GACACGAAAGGACACCACTAGTAAAAACC ACTAGTTTTTTCTGACACCTCTTGCTACAAA CTCTGTAAAAATCAAAAGGATCGATAGGCC GCGCTTTCACGGTCTGTATTCGTACTGAAAA TCAAGATCAAGGAAGCTTTTCCCCTTTTAGT CAACACCAGGTTTCTGTCCTAGTTGAGCTTC CCTTGGGACATCTGCGTTACCATTTGACAGA TGTACCGCCCCAGTCAAACTCCCCACCTGA CACTGTCCTCAAAACAGTTCAATTGCATCCG AAGATCGCAATTTTTTCACTAAAATAAATTA ACAAAAGTTAATTATACTGCTTCATTGAGTA AGTAGAAAAACAATC (SEQ ID NO: 1045)			
NeSL	NeSL-1_CBre	.		Caenorhabditis brenneri	AAGGACGCTGGTTTAAGGCCGAATTCGTTC GTTCTTTTTCTGGCGGTCTTGCTTTGAGCTT GGTTTCCGATCCTATGCCCTTGWGCATCAG CGATTGCGTTCATCTGGTCTCGGCCGAAGG	AAGGAC GCTGGT TTAAGG CCGAAT	TAAA CCCA CACG AGAM	MPLXISDCVHLV SAEGDTMNGRS TCGPLSRSSSVV SRSRSSPSPSVPP

					AGATACGATGAACGGGAGGTCCACTTGTGG	TCGTTC	CTAC	HPSPSIGPDTGLS
					GCCATTGTCTCGTTCATCCTCTGTCGTAAGT	GTTCTT	GACG	AGIIGTSRGCSL
					AGGTCGAGGTCTTCCCCTTCCCCCAGTGTTC	TTTCTG	CCAT	WLPEVDNALSQ
					CCCCCCACCCTTCCCCCAGTATCGGTCCAGA	GCGGTC	AAGA	WLRKGLERDHE
					TACAGGATTGTCGGCTGGAATCATCGGCAC	TTGCTT	TCAG	VLVCGFEAAKP
					ATCGAGAGGATGTAGCCTTTGGTTGCCAGA	TGAGCT	GCAT	LSLSKARLLRKT
					GGTGGACAATGCCTTATCACAGTGGCTGAG	TGGTTT	GTAC	PRNTGVVRHILE
					AAAAGGGTTGGAACGAGACCATGAAGTTCT	CCGATC	GGAT	FDGRLVHTNCN
					GGTTTGTGGATTTGAGGCAGCAAAGCCACT	CT (SEQ	GTGA	ETECVLSTLXSX
					GTCACTTTCCAAAGCTAGACTTCTAAGAAA	ID NO:	ATGA	XAVEVVRISLKC
					GACCCCAAGGAACACTGGTGTGGTTAGGCA	1169)	GACT	EPREPCEPKCVL
					CATATTAGAATTTGACGGAAGGTTAGTTCA		GATG	SILCSDKIVXISF
					TACTAACTGTAACGAGACCGAGTGTGTTCTT		AACG	ECETREPFPFFX
					TCTACTTTGTKCAGTGAMGWGGCTGTCGAA		GAAT	DRKFREPIPFVFE
					GTAGTCAGGATATCTCTCAAATGTGAACCC		GAGC	RMYDPRDPIPSFI
					CGTGAACCCTGTGAACCCAAATGTGTTCTTT		ACGT	CWMYDLRQRM
					CTATTTTATGCAGTGATAAGATAGTCWGGA		GCCC	TPGTLPXNPLSX
					TATCATTTGAATGTGAAACWCGTGAACCTT		ATAA	ENKDSWGRPAV
					TTCCTTTCTTCMCGGATCGGAAATTCAGAG		GATC	IKNEIRSMRSYL
					AACCTATTCCTTTCGTTTTCGAGAGGATGTA		GGGT	EENVKENRLNL
					TGACCCAAGAGACCCTATTCCTTCATTTATT		ATKA	LRRLRGGGEGK
					TGTTGGATGTATGACCTGAGACAAAGGATG		AAGA	KMIRKLVAEKK
					ACCCCTGGCACSTTGCCAAGWAATCCCCTT		WCAG	SDTEAVCRILYP

					TCTCMAGAGAACAAAGACAGCTGGGGACG		AGAC	LDDRYECFVDG
					CCCAGCTGTCATAAAGAATGAGATAAGATC		GATC	CETTSTMGYGSS
					TATGAGATCTTATCTCGAAGAGAATGTGAA		CCTA	DLKYMTTHIKK
					GGAAAACCGCCTGAACCTTTTGAGAAGGTT		MCAT	EHGVKVQWTYE
					AAGAGGTGGTGGTGAAGGAAAGAAGATGA		CGGG	CSLCNKQAPFM
					TCAGAAAGTTGGTTGCAGAAAAGAAAAGCG		AAAA	GGAASKWVTA
					ACACAGAGGCTGTCTGCAGGATACTGTACC		CACG	HMATKHTETVK
					CACTTGATGATCGTTATGAGTGTTTTGTTGA		AGTT	LKLKPSISTTAK
					TGGTTGTGAGACAACATCAACGATGGGATA		ATAC	VAAKLDEIAVSL
					CGGGTCTAGTGACCTGAAATACATGACCAC		TGCTT	PKPRQVRVLRD
					ACACATAAAGAAAGAGCATGGTGTGAAAGT		CACT	PDEVKEKVAKP
					CCAATGGACATATGAGTGCTCCCTGTGTAA		GAMC	TLASTREEVKRN
					TAAGCAAGCTCCTTTCATGGGGGGAGCTGC		TCGC	ALRNMAPLVEL
					GTCCAAGTGGGTTACAGCGCACATGGCAAC		TAAG	SSQNQLTGAERP
					AAAGCATACCGAAACGGTGAAGTTGAAGCT		CTCTC	EETSEAMRLEEC
					CAAACCAAGCATCTCGACTACTGCCAAGGT		ATAA	RTPEKIAELEGKI
					TGCTGCGAAGCTAGATGAGATCGCCGTGTC		TGAC	QTRTVTKKLSA
					GCTACCCAAACCGAGACAAGTACGTGTATT		CGAA	LKESMEKRTRE
					GAGAGACCCAGATGAAGTGAAAGAGAAGG		CTTGT	EKVGKPSLAPIH
					TTGCAAAACCAACACTTGCTTCCACGAGAG		TCGC	EEVKKTARRSL
					AAGAAGTGAAGAGAAATGCNTTGCGAAAC		AACT	APLVEPSTFTHL
					ATGGCCCCACTAGTCGAACTGAGTTCTCAG		GCCT	TGASRLQAVRD
					AATCAGTTGACWGGAGCCGAAAGACCTGA		CCTA	AFSKANKDAAA

					GGAAACTAGTGAAGCTATGCGACTCGAGGA		ACCG	KRRSSLAKPARL
					GTGTAGGACTCCAGAGAAGATTGCTGAACT		GGCG	SEIMNTTFTKET
					AGAAGGAAAGATACAGACCCGAACAGTGA		GGTG	VNETKEPVNDT
					CTAAAAAGCTTAGTGCACTGAAAGAGTCAA		TGAG	DESIATIQPQVR
					TGGAAAAGAGAACGAGAGAGGAGAAGGTT		AAGG	VYRFNTWCLDH
					GGGAAACCATCACTTGCTCCAATTCATGAA		GAGG	ETTREAWLTGE
					GAAGTGAAAAAGACTGCAAGACGGAGCTT		TCGC	VVDWFMGKVT
					GGCACCTCTAGTTGAACCGAGTACGTTCAC		CTTG	EKKDQYRVFDS
					TCATTTGACTGGGGCGTCAAGACTTCAGGC		AGGC	LVWSMYKFHG
					TGTTCGTGACGCGTTCTCGAAAGCCAACAA		GGAC	VGYVLDLMRDP
					AGACGCTGCGGCGAAAAGAAGGTCTAGCCT		GCAA	LTYFLPICEHDH
					GGCGAAACCAGCTAGATTATCAGAGATTAT		TGAG	WVLLVIDEKGI
					GAATACCACCTTCACGAAGGAGACGGTAAA		GGAT	WYGDSKGAEPC
					TGAGACGAAAGAACCTGTGAATGATACTGA		GTGT	REIAKFIEETKRE
					CGAGAGTATCGCAACAATCCAGCCACAAGT		GCAG	RRMFPVPVPLQ
					ACGTGTCTACCGGTTTAATACATGGTGTCTC		GTTC	RDGVNCGVHIC
					GATCATGAAACCACGAGAGAAGCCTGGTTA		CCCC	LMVKSIVNGEP
					ACCGGAGAAGTTGTGGATTGGTTCATGGGA		TCTTG	WYTEEEVKVFR
					AAAGTGACTGAGAAGAAAGACCAGTACAG		AGAT	RNVKRGLKEFG
					AGTGTTTGACTCACTTGTATGGTCAATGTAC		CCGA	FELYSERIVYVG
					AAGTTCCATGGTGTAGGGTATGTATTGGAT		AAGT	DDSIKVNDEHD
					CTGATGAGGGATCCTCTAACATACTTCTTAC		CTAA	DDVVFLSEETN
					CAATATGTGAACACGATCACTGGGTTTTGCT		AAGT	NTTFTIEQAEDP

					AGTGATTGATGAGAAAGGAATTTGGTACGG		ACTA	AEEDAQHLESP
					TGACTCGAAAGGTGCAGAACCGTGTAGAGA		GACC	VKPVKLMELKIP
					AATCGCCAAATTCATAGAAGAGACGAAAAG		GAAA	KIEIKKKEIRRKP
					AGAAAGACGAATGTTCCCAGTCCCCGTACC		GATC	KQQIEKKRKVPT
					TCTTCAAAGAGACGGAGTGAACTGTGGTGT		GAGG	GKPDELLVRVR
					ACATATATGTCTAATGGTTAAATCCATCGTG		ACGG	LWLEREVQSYF
					AATGGCGAACCATGGTACACCGAAGAAGA		ACGG	DSGKRFQRLEWI
					AGTGAAAGTGTTCAGAAGAAATGTGAAAAG		GATG	LDVLTAAIHKAT
					AGGTCTGAAAGAATTTGGTTTTGAACTTTAT		GCCG	AGDEQAIERIEK
					TCTGAAAGGATCGTCTATGTCGGAGATGAC		CGAG	RSPPLEVEEGEM
					AGCATAAAAGTGAATGATGAGCATGATGAT		GCAC	STQTEPKKRERK
					GACGTGGTATTCCTCTCGGAGGAGACGAAT		ACGG	EKESGCEMKAS
					AACACTACGTTCACGATCGAGCAAGCAGAA		CGGG	HKEMYFKNRSK
					GATCCGGCTGAAGAGGATGCCCAGCATCTG		TAAC	AFNVIIGKDSKQ
					GAGAGTCCGGTGAAACCTGTAAAGCTCATG		ACAG	CEIPIETLQKFFE
					GAGTTGAAAATTCCAAAGATTGAGATAAAG		CCAG	GTTAETNVPAE
					AAGAAAGAGATTCGGAGAAAACCGAAACA		ATAA	VLKEMGSRLPK
					ACAAATCGAAAAGAAAAGAAAGGTGCCAA		CCTA	LEALDWMEANF
					CAGGGAAACCAGATGAACTGTTGGTCAGAG		GTAG	IESEVSDAMKKT
					TGCGATTATGGTTGGAAAGAGAAGTCCAAT		ATCTT	KDTAPGVDGLR
					CATACTTCGACTCTGGAAAGAGATTCCAAA		CGGA	YHHLKWFDPEY
					GACTGGAGTGGATATTAGATGTCCTCACGG		TCTC	KMLTLLYNECK
					CTGCGATTCACAAGGCTACCGCCGGTGATG		GTCG	NHRKIPSHWKE

					AGCAAGCAATTGAAAGAATTGAGAAGAGA		GCCT	AETILLYKGGDE
					TCACCCCCTTTGGAAGTGGAAGAGGGTGAA		GGAG	TRPDNWRPISLM
					ATGTCTACACAGACAGAACCAAAGAAAAG		ATAT	PTIYKLYSSLWN
					AGAAAGAAAAGAAAAGGAGTCAGGTTGTG		GTGG	RRIRSVGGVMS
					AAATGAAAGCTTCTCACAAGGAGATGTACT		AACC	KCQRGFQEREG
					TCAAAAACCGCTCCAAAGCGTTCAATGTGA		CTGG	CNESIGILRTAID
					TAATTGGAAAAGACTCAAAGCAATGCGAGA		GAAA	VAKGKRRNLSV
					TTCCAATTGAGACCCTGCAAAAGTTCTTTGA		GGAG	AWLDLTNAFGS
					GGGAACAACTGCAGAAACGAATGTGCCAGC		AAAG	VPHELIKSTLES
					AGAAGTGCTGAAAGAGATGGGTTCACGTCT		TTGTT	YGFPEMVTEIV
					GCCAAAGTTGGAGGCGTTGGACTGGATGGA		TGTT	MDMYRGASIRI
					AGCTAATTTCATTGAAAGTGAAGTGTCAGA		GGGC	KSKNEKSEQIVI
					TGCGATGAAAAAGACCAAAGACACCGCTCC		TGGC	KSGVKQGDPISP
					GGGTGTAGACGGACTACGGTATCACCATTT		AAGA	TLFNMCLENVIR
					GAAATGGTTTGATCCAGAGTATAAGATGTT		GTGA	RHLDSASGHRCI
					GACACTTCTCTACAATGAATGTAAGAACCA		AGTT	KTKVKVLAFAD
					TCGAAAGATTCCAAGTCATTGGAAAGAGGC		TGAA	DMAILAENRDQ
					AGAGACGATTCTCCTTTATAAAGGAGGTGA		TGTG	LQTELNKLDKE
					TGAAACGAGGCCCGACAACTGGAGACCTAT		AACC	CESLNLIFKPVK
					AAGTTTGATGCCCACGATCTACAAACTGTA		ACCG	CASLIIERGMVN
					TTCTAGTCTTTGGAACCGAAGAATTAGATC		TCAT	KNAEVVLRGKPI
					AGTTGGTGGTGTGATGAGCAAATGTCAACG		GCAA	RNLDENGSYKY
					AGGTTTCCAAGAGAGAGAAGGGTGTAATGA		CCAC	LGVHTGIATRVS

					AAGCATAGGAATCCTTAGAACGGCTATCGA		TAAA	TMQLLESVTKE
					TGTCGCTAAGGGAAAGAGAAGGAACCTGTC		CCAG	MDLVNQSGMAP
					AGTTGCATGGTTAGACCTTACGAATGCGTTT		TGGC	FQKLDCLKTFVL
					GGTTCAGTACCCCATGAACTGATAAAAAGT		GATG	PKLTYMYANAI
					ACTCTGGAATCGTATGGATTCCCAGAAATG		CGGG	PKLTELKVFANL
					GTGACAGAGATTGTCATGGATATGTACAGA		TGGA	TMRMVKEIHEIP
					GGTGCATCAATCCGAATCAAGAGCAAGAAT		GTCA	IKGSPLEYVQLP
					GAGAAAAGTGAACAGATTGTTATCAAATCT		TCAC	PSQGGLGVACP
					GGAGTGAAGCAGGGAGACCCCATCTCCCCC		AGGA	KITALITFLVNV
					ACGCTGTTCAACATGTGTTTAGAGAATGTG		AAAT	MKKLWSSDSYI
					ATAAGAAGACATCTGGATAGTGCTTCGGGA		GTTTC	RKLYRDYLDEV
					CACAGATGCATAAAGACAAAAGTCAAGGTC		TGTT	AETETGMEEMT
					CTGGCTTTTGCAGATGACATGGCCATACTG		GCTT	KEDIAKYLSGD
					GCCGAGAACAGAGATCAGTTACAAACTGAA		GACT	VPIDKKAFGYNT
					CTAAACAAGTTGGACAAAGAATGTGAATCA		TATC	FTRVRDVCNSLT
					CTAAACCTCATTTTCAAACCGGTAAAGTGT		AGTG	XIXGAPLHKLKI
					GCCAGTTTGATAATTGAGAGAGGAATGGTG		TTTG	VERDGDFAILVQ
					AATAAGAATGCGGAAGTGGTTCTGAGAGGG		ATAT	ATKEGMEKIFTC
					AAGCCAATCAGAAACCTAGATGAGAATGGT		CGCC	AQEKKLQQLLK
					TCCTATAAGTACCTGGGAGTACATACAGGA		CTCA	AEVNTALAHRF
					ATCGCAACAAGAGTTTCAACAATGCAATTG		GGCA	FTEKPVKSAVM
					TTGGAAAGTGTCACGAAAGAAATGGATCTA		CAAG	SVMRQYPQSNA
					GTGAATCAGAGCGGCATGGCTCCGTTTCAA		TATG	FVKNGKNVSIA

					AAACTAGACTGCCTAAAGACGTTTGTTTTGC		AAGG	VHSWIHKARLN
					CGAAACTAACGTACATGTATGCGAACGCAA		CCCC	ALHCNFNTYGE
					TACCTAAGTTAACGGAGCTTAAGGTCTTTGC		CACC	NKSKVCRRCGK
					AAACTTGACGATGAGAATGGTGAAAGAGAT		CACA	DVETQLHILQXC
					TCATGAAATCCCCATCAAGGGATCTCCGTT		TAAA	EYGLPKLINERH
					AGAGTATGTACAGCTACCTCCAAGTCAAGG		CTCC	DAVLHVVRNLI
					CGGATTAGGAGTGGCTTGTCCAAAGATAAC		CTAG	RKGSKKDWKLK
					AGCGTTGATTACCTTCTTGGTCAACGTCATG		CAAC	IDETVSSCNQLR
					AAAAAGCTATGGTCTTCTGACAGTTACATC		TGGT	PDIYMCSPDGKE
					AGAAAACTATACAGGGACTACCTGGATGAA		AGTC	VIMADVTCPYES
					GTCGCAGAGACGGAGACAGGTATGGAAGA		CAGC	GMQAMQESWN
					GATGACGAAAGAAGATATTGCAAAATATCT		AAGC	RKVTKYEGGFS
					GAGTGGTGATGTGCCGATCGACAAGAAGGC		GCTG	HFXKMGKKFTV
					GTTCGGTTACAACACMTTCACAAGAGTAAG		GTWC	LPIVVGSLGTW
					AGATGTGTGCAACAGCCTCACTAASATAGK		TTGCT	WKPTTNSLVQL
					GGGAGCTCCACTGCACAAGTTAAAGATTGT		ACTA	GIEKXTIRRVIPE
					GGAGAGAGACGGTGACTTTGCCATTCTAGT		TTGC	LCSMTMEYSKD
					GCAAGCCACCAAAGAAGGAATGGAGAAAA		GCCC	VYWNHIFGDTF
					TCTTCACCTGTGCTCAGGAGAAGAAACTCC		CAGG	RKPPMRFGVEK
					AACAGCTTCTGAAAGCAGAAGTMAACACG		CTCG	PKGNSWKKEGS
					GCTCTAGCGCACCGTTTCTTCACCGAGAAA		CCC	EPKGAASSD
					CCMGTGAAAAGTGCAGTGATGAGTGTAATG		(SEQ	(SEQ ID NO:
					AGACAGTATCCACAGAGCAATGCCTTTGTG			1414)

					AAGAATGGAAAGAATGTGAGCATTGCTGTC CACTCGTGGATACACAAAGCAAGGTTGAAT GCGCTGCATTGCAACTTCAACACGTACGGT GAAAACAAGTCAAAAGTGTGCCGACGTTGC GGCAAAGACGTGGAAACCCAACTGCACATC CTGCAGWCATGCGAGTACGGGTTACCAAAG CTAATCAACGAAAGACATGATGCGGTGTTA CATGTGGTGAGAAACCTCATCCGCAAAGGC TCAAAGAAAGACTGGAAGCTAAAGATAGAT GAAACTGTGTCAAGTTGTAATCAACTTCGTC CAGACATCTATATGTGTAGCCCAGATGGGA AAGAGGTCATAATGGCAGATGTAACCTGTC CTTATGAATCAGGAATGCAAGCTATGCAAG AGAGTTGGAACCGAAAGGTCACSAAATACG AAGGAGGCTTTAGCCACTTCCAWAAGATGG GAAAGAAATTCACAGTGTTGCCAATAGTGG TTGGATCACTGGGAACGTGGTGGAAACCCA CAACGAACAGCTTAGTTCAACTAGGCATAG AGAAAGASACGATAAGAAGAGTGATCCCCG AGCTGTGCTCAATGACCATGGAATACAGTA AGGATGTCTACTGGAACCATATATTCGGGG ACACCTTCAGGAAGCCACCAATGAGATTTG GTGTAGAGAAGCCAAAGGGTAATAGTTGGA		ID NO: 1292)	

					AGAAGGAAGGCAGCGAGCCGAAAGGTGCT GCTTCCTCCGACTAAACCCACACGAGAMCT ACGACGCCATAAGATCAGGCATGTACGGAT GTGAATGAGACTGATGAACGGAATGAGCAC GTGCCCATAAGATCGGGTATKAAAGAWCA GAGACGATCCCTAMCATCGGGAAAACACG AGTTATACTGCTTCACTGAMCTCGCTAAGC TCTCATAATGACCGAACTTGTTCGCAACTGC CTCCTAACCGGGCGGGTGTGAGAAGGGAGG TCGCCTTGAGGCGGACGCAATGAGGGATGT GTGCAGGTTCCCCCTCTTGAGATCCGAAAG TCTAAAAGTACTAGACCGAAAGATCGAGGA CGGACGGGATGGCCGCGAGGCACACGGCG GGTAACACAGCCAGATAACCTAGTAGATCT TCGGATCTCGTCGGCCTGGAGATATGTGGA ACCCTGGGAAAGGAGAAAGTTGTTTGTTGG GCTGGCAAGAGTGAAGTTTGAATGTGAACC ACCGTCATGCAACCACTAAACCAGTGGCGA TGCGGGTGGAGTCATCACAGGAAAATGTTT CTGTTGCTTGACTTATCAGTGTTTGATATCG CCCTCAGGCACAAGTATGAAGGCCCCCACC CACATAAACTCCCTAGCAACTGGTAGTCCA			

					GCAAGCGCTGGTWCTTGCTACTATTGCGCC CCAGGCTCGCCC (SEQ ID NO: 1046)			
NeSL	NeSL-1_CJap	.		Caenorhabditis japonica	GCGCCCCGGGTTACATTGTCGGGGCCACCT	GCGCCC	TAAA	WRRPAPKQTKN
					TTCTCTTGGAGTAGAGTACAGTCTACTAATT	CGGGTT	AGCC	SSLHHLGHEVK
					TTTTGATAAGCTAGTCGGGTCCGAACCACT	ACATTG	AAAA	RIARLKPGIFEFH
					AGAGTTTGCTTGAAAATGCGTCAAACCAGC	TCGGGG	GCCA	AKPKNSSLHHL
					ATTTTAGAACTCGCCCAAAAGTTCGGCCCC	CCACCT	CGGA	GHGVKRXARLK
					GACCCCCAAACAAATGGGACCTTCTTGACG	TTCTCTT	GCAT	PGIFEFHAKXKN
					ATTTTCCCTGAAAATCGGAGGATGGAATGG	GGAGTA	CGGG	SSLHHLGHEVK
					TCCCCTATTCTTGTAAATAGKACTGTGCAAT	GAGTAC	AAAG	RIARLKPGIFEFH
					ACCCCTTCGTCATCTGTGGGGAACAGATGA	AGTCTA	AAAA	AKPKNSSLHHL
					CACGTGACGTCATCCGTGTAGACGTCACGT	CTAATT	ATGG	GHEVRRNSRLK
					TTTCCCGTGCCTGCGGGAGCCCCCAATCGA	TTTTGA	AAAA	PGIFGFYQKSKN
					GCAATTTTTGCTCTTTTGAGTGTCTGGAACG	TAAGCT	GGAC	SSLHHLGHEVR
					CTTGAAACCCCAGACAAATCAGGCCCAGTC	AGTCGG	TGAA	RIARLKPGILEFH
					GTCGGAAAATTTCTTTTTGAAATTTTTTGGC	GTCCGA	AACG	AKNRIKSGLKVT
					GCCTGCGAAAAAAATTTTTTAACCGCCACA	ACCACT	AGAC	FLSDLXAHAGA
					AACCCCCGGGAGGCGCGGWTAGGGATATC	AGAGTT	TGAA	LACSRFLASTLK
					GATGTCATCGACTCGTCGGTGATCTTTGATT	TGCTTG	AAAT	TEHCRQKSFKPV
					TTCTCTCTGCGTCTCCTATTTTGGAACAGTC	AAAATG	CCCA	GFLLHFLKNSSI
					TCGACCAAAAAACCGGGCCTGGCAACCCAC	CGTCAA	AACA	NEVASLRNVKK
					CGAATCCGGATGTCGGAGGGATTTGGCAAG	ACCAGC	AAAC	XFLEFFSGKPIG
					AAATGTTGGAAATAACGAAATTTCGTTATTT	ATTTTA	AAAT	GMASFSRTKITF

					TCAGCACAATTGTCAAACCGGCAAGAAAAC	GAACTC	CCAA	FKLCLKNFVLSA
					TGGATGGACAAGACACACAATTTACCGGAA	GCCCAA	AACA	ENPPIIRQKTNQ
					ATTGTGCTTGTTACGTCGAATTTCCCAATTT	AAGTTC	AACT	NKASXVQIARG
					TGAAAAAATTCCTCGTTCCACTGGTCGGGA	GGCCCC	GAAA	GHLSDCLPSQK
					CGCGAGGTCAGACGATCTGCACGTCTGAAA	GACCCC	AAAA	MAGVLGRLFLS
					CCCAAAATCTTCGGATTTTATGCAGTAGTGG	CAAACA	AAAA	VQSTLSHRPFDT
					CGCCGCCCGGCTCCCAAACAAACAAAAAAT	AATGGG	AAAA	LLRSDDDKRGR
					TCCTCGTTGCACCATTTGGGGCACGAGGTC	ACCTTC	CAAA	KTIKLQFFIKENL
					AAACGAATTGCACGTTTGAAACCTGGAATT	TTGACG	ACAA	VTPXVARDVKIL
					TTTGAATTCCATGCAAAACCAAAAAATTCC	ATTTTC	AAAC	XKQTKNNSGNS
					TCGTTGCACCATTTGGGGCACGGGGTCAAA	CCTGAA	TGGA	DSNSETKNFSKN
					CGAWTTGCACGTTTGAAACCCGGAATCTTC	AATCGG	CAGA	KVSRQNGPLIGG
					GAATTCCATGCAAAACMAAAAAATTCCTCG	AGGATG	CACT	GNHKKIGENQIT
					TTGCACCATTTGGGGCACGAGGTCAAACGA	GAATGG	GGAA	RTLEIESKSDDN
					ATTGCACGTTTGAAACCCGGAATTTTCGAAT	TCCCCT	ACAG	KVLVLRILYPTN
					TCCATGCAAAGCCAAAAAATTCCTCGTTGC	ATTCTT	TGTC	DWYKCYSQWC
					ACCATCTGGGGCACGAGGTCAGACGAAATT	GTAAAT	AGGC	QHKSLVGYGAH
					CACGTCTGAAACCCGGAATCTTCGGATTCT	AGKACT	AAAG	DLKYLTDHIKST
					ATCAAAAATCAAAAAATTCCTCGTTGCACC	GTGCAA	TCGC	HSKKVEWSYQC
					ATTTGGGGCACGAGGTCAGACGAATTGCAC	TACCCC	CGAT	SICDAKAEGTGT
					GTTTGAAACCCGGAATTCTCGAATTCCATGC	TTCGTC	TATA	KAARWITAHMP
					AAAAAACCGGATAAAATCCGGTCTAAAAGT	ATCTGT	CTGTT	KVHGIEATHRIK
					GACGTTTTTGTCAGATCTTWCTGCTCACGCT	GGGGAA	CCAC	QNSEKTTNVKT

					GGTGCGTTGGCGTGTAGCAGATTTCTGGCA	CAGATG	GCCT	ANSLQEMALSL
					TCGACACTTAAAACGGAGCACTGCCGACAA	ACACGT	TAAA	QKPKNGPKKVV
					AAAAGCTTCAAGCCAGTCGGTTTTCTGCTTC	GACGTC	AGTC	MATSTTPEKKIS
					ATTTCCTAAAAAATTCCTCGATCAACGAGG	ATCCGT	CCGA	ELESKIQTREVA
					TGGCGTCTCTCCGCAACGTCAAAAAAAWKT	GTAGAC	AATG	KQLSALKESAQ
					TTCTTGAATTTTTCTCAGGAAAACCTATCGG	GTCACG	GCGC	KNQQGNKTKNV
					TGGGATGGCTTCTTTCAGTAGAACCAAAAT	TTTTCC	AAAA	KSSLKTIAENTN
					AACTTTTTTCAAACTTTGTTTGAAGAATTTT	CGTGCC	CAAC	ETKKISARKSLI
					GTTTTGTCTGCTGAAAATCCCCCGATAATCC	TGCGGG	CTGA	NYLKPEDVLNHI
					GCCAAAAAACAAACCAAAACAAAGCGAGC	AGCCCC	ATCT	PKEPKPASAKX
					CKTGTCCAAATCGCAAGAGGAGGTCATCTG	CAATCG	ATCT	GLQELTGAQRL
					TCAGACTGTCTKCCGTCCCAAAAGATGGCA	AGCAAT	GAAA	QETRRRFMAGN
					GGAGTTCTCGGACGACTATTTCTCTCGGTTC	TTTTGC	GTGC	RRDSIARRESLS
					AGAGCACTCTCTCGCACCGCCCTTTCGACAC	TCTTTT	TCCA	LGKISNSFKIELK
					GTTATTGCGGAGCGATGACGACAAAAGAGG	GAGTGT	AACC	NAPEKTTLKKP
					GAGGAAAACGATCAAACTCCAGTTTTTTAT	CTGGAA	ACGC	AVTQKQNTSQN
					TAAAGAAAATCTGGTCACACCTKTGGTTGC	CGCTTG	ACAA	VSSSTVVKENKT
					TAGGGACGTGAAAATTTTAAAMAAACAAA	AAACCC	CTCG	GNDVITIDDTET
					CAAAAAACAATTCTGGGAACTCTGATAGCA	CAGACA	GAGA	VKRKINTWCLD
					ACAGTGAAACAAAAAACTTCTCTAAAAATA	AATCAG	AAAT	HESTENAWMAD
					AAGTTTCCAGACAAAATGGCCCATTGATTG	GCCCAG	CAGG	DIIFWYIQKQIEI
					GGGGCGGTAACCACAAAAAAATCGGAGAA	TCGTCG	GACA	SLDNKKFKVIDP
					AACCAAATCACACGCACTTTGGAAATTGAA	GAAAAT	AGTT	LIWTTYRIYGVE

					TCCAAAAGCGATGACAACAAAGTTTTGGTC	TTCTTTT	GCTT	CVQDELVGFEK
					CTCCGAATACTGTACCCAACTAATGATTGGT	TGAAAT	CACG	YFFPICENGHWV
					ACAAGTGTTACTCCCAATGGTGCCAACACA	TTTTTG	CAAC	LLIIDDKRVWYS
					AATCCCTTGTTGGATATGGCGCTCACGATTT	GCGCCT	GGGC	DSLADKPIEVIE
					AAAATACTTGACAGACCACATAAAGTCCAC	GCGAAA	TGGG	DLINKLNRTQG
					TCATTCTAAAAAGGTTGAGTGGTCTTATCAG	AAAATT	ACAG	KFNQTVPKQKD
					TGTAGTATTTGTGACGCAAAAGCCGAAGGT	TTTTAA	GTAC	GFNCGVHVCLV
					ACCGGTACAAAAGCWGCTAGATGGATTAC	CCGCCA	CCCC	AKSVITENFWYT
					AGCCCACATGCCAAAAGTACACGGTATTGA	CAAACC	TCCT	EKDVNDFRKTV
					AGCAACACACAGAATTAAACAAAATTCTGA	CCCGGG	GAAA	KLWLFSEGFELY
					AAAAACAACAAATGTTAAAACTGCGAACAG	AGGCGC	CCGC	SEPYKQIQNKNI
					TCTCCAGGAAATGGCGCTGTCGCTCCAAAA	GGWTA	GAGG	SVNSEKNQISDN
					ACCAAAAAATGGTCCGAAAAAAGTTGTAAT	GGGATA	TTGA	EKNWGDKTQTV
					GGCAACTAGTACGACCCCAGAAAAGAAAAT	TCGATG	GGAT	NESTLKERDEDI
					CTCTGAACTGGAATCAAAAATCCAAACCAG	TCATCG	GGAC	FLLRPHISVGVA
					AGAAGTGGCCAAACAATTGAGCGCTCTGAA	ACTCGT	GGGA	LKTEDEKNQKA
					GGAGTCAGCTCAAAAAAATCAGCAAGGAA	CGGTGA	AGGC	ENLKAPQKLKAI
					ACAAAACAAAAAATGTTAAATCAAGCTTAA	TCTTTG	CGCG	RRLKILKTCLKK
					AAACAATTGCTGAAAACACAAATGAAACM	ATTTTC	AGGC	LTAVKGKPEETE
					AAAAAGATWAGCGCTCGAAAGAGCCTTAT	TCTCTG	TTAT	RAAIPNLMAIKL
					AAACTATCTGAAACCTGAGGATGTGCTCAA	CGTCTC	GGCG	KTPPKVEPVRRN
					TCACATTCCAAAGGAGCCAAAACCAGCTTC	CTATTT	GGTA	PEKGENYXKSQ
					TGCGAAAAKTGGCCTKCAAGAACTGACTGG	TGGAAC	ACTC	PNKKRQIPTGKP

					TGCTCAAAGACTGCAGGAAACMAGAAGAA	AGTCTC	GGTT	DELVKKVREWF
					GGTTTATGGCWGGAAACAGAAGAGATTCA	GACCAA	GGTG	EIQFQAYFEDGK
					ATTGCAAGAAGAGAAAGTCTGTCTCTCGGC	AAAACC	TGCT	SFQRLEWXTGL
					AAAATCTCAAACTCATTTAAAATTGAGCTG	GGGCCT	AGTA	LTAAIHKASAG
					AAAAATGCTCCGGAAAAAACAACTCTTAAA	GGCAAC	GATG	DEQAVGKIIKRC
					AAACCGGCTGTCACTCAGAAACAAAACACG	CCACCG	ATTT	PPLEIEEGEMAT
					AGTCAGAATGTATCTAGTTCTACGGTTGTAA	AATCCG	ATAT	QTETKQKPKNQ
					AAGAGAACAAAACAGGAAATGACGTGATC	GATGTC	CCGA	KSTKGANSSSSI
					ACAATTGATGACACAGAAACTGTTAAAAGA	GGAGGG	CAGC	REAYAENRART
					AAAATAAACACTTGGTGTCTCGACCACGAA	ATTTGG	CCCA	FNKIIGKDDKCE
					TCCACAGAAAATGCGTGGATGGCTGACGAC	CAAGAA	ACTA	IPIEKIEKFFENT
					ATCATATTCTGGTACATCCAGAAACAGATT	ATGTTG	AGAG	TSNTNVPTETLA
					GAAATCAGTTTGGACAATAAAAAGTTCAAA	GAAATA	GAAT	RITSDLPKLEIGS
					GTGATTGATCCACTCATCTGGACCACATATC	ACGAAA	CCTG	WIEEEFREKEVA
					GAATTTATGGTGTCGAATGTGTCCAAGATG	TTTCGT	GGAA	EALKKTKDTAP
					AACTAGTTGGATTTGAAAAATACTTTTTTCC	TATTTT	AGGA	GVDGLRYHHLS
					AATCTGTGAAAATGGTCATTGGGTTTTGCTG	CAGCAC	AAAC	WFDPKXKLLTK
					ATTATCGATGACAAAAGAGTCTGGTACAGT	AATTGT	TTGA	LYNECREHKKIP
					GATTCCCTGGCCGATAAACCAATTGAGGTT	CAAACC	AAAA	GHWKEAETVLL
					ATTGAGGACCTCATAAACAAACTAAATCGA	GGCAAG	GTTTT	YKGGDETQAEN
					ACCCAAGGTAAATTTAACCAAACGGTTCCA	AAAACT	TACA	WRPISLMPTICK
					AAACAAAAAGACGGCTTTAATTGTGGAGTT	GGATGG	GGGC	LYSSLWNKRIKS
					CATGTATGTCTGGTGGCCAAATCCGTTATCA	ACAAGA	TGGT	VTGVLSKCQRG

					CTGAGAACTTTTGGTACACAGAAAAAGACG	CACACA	AATA	FQEREGCNESIAI
					TTAATGACTTCAGAAAAACTGTCAAGCTTT	ATTTAC	GTTC	LRTAIEAAKGTK
					GGCTTTTCAGTGAAGGGTTTGAACTCTATTC	CGGAAA	AGCA	KSLSIAWLDLTN
					AGAGCCGTACAAACAAATCCAAAACAAAA	TTGTGC	CAAT	AFGSVPHESIEA
					ACATTTCCGTTAATTCGGAAAAAAATCAAA	TTGTTA	TGTA	TLIAYGFPGMVT
					TCAGTGATAATGAAAAAAATTGGGGTGATA	CGTCGA	GTCT	EVIKDMYNGASI
					AAACTCAAACTGTGAATGAGAGTACTCTGA	ATTTCC	ACTG	RVKTKNEKSKQI
					AAGAAAGAGATGAAGACATCTTTTTGCTCA	CAATTT	TCTTG	LIKSGVKQGDPI
					GACCACACATCAGTGTTGGAGTTGCTCTCA	TGAAAA	CAAC	SPTLFNICLESVI
					AGACAGAAGACGAGAAAAATCAAAAAGCT	AATTCC	CACA	XRHLKSADGHK
					GAAAACTTGAAAGCCCCACAAAAACTCTGA	TCGTTC	ACAA	CIXSNIKLLAFA
					AACACGGAAGAATTCCAAGTGGACAAAAA	CACTGG	ACCA	DDMAILSDSKT
					CGAGAAACCAGAATCTCCAAATGCCCAGGA	TCGGGA	GTGG	KLQQELQKMDD
					AACTCCAAAAAACGAGCCAAAAATGGTTCC	CGCGAG	TTCTG	DCTPLNLIFKPA
					AAGTCCGAAGAATTCGGAAAAAGAAATTTC	GTCAGA	CGGG	KCASLIIEWGKV
					TACTGAGCTTCTGGATGCTCAAGAAGCCGG	CGATCT	TAGA	QKDQKIKLKGQ
					AGAAGAGCTGAAAAGCGATCCGAAGGCTG	GCACGT	TCAA	FIRSLAEQDTYK
					AAAATCCTGAAAACCTGTCTCAAAAAGCTG	CTGAAA	ACTA	YLGVQTGIETRV
					ACGGCAGTGAAGGGAAAACCGGAAGAGAC	CCCAAA	TAAT	SAMQLMKKTVS
					GGAAAGAGCCGCCATTCCAAACCTCATGGC	ATCTTC	TTGT	ELDKINCSALAX
					AATCAAGCTCAAGACGCCTCCAAAAGTTGA	GGATTT	GTGT	WQKLDAVKTFV
					ACCTGTAAGAAGAAACCCTGAAAAGGGTGA	TATGCA	TTTCT	LPKMTYMYANT
					AAATTACMAAAAAAGTCAGCCAAACAAAA	GTAG	TTTAC	VPKLSELKEFAN

					AGAGACAAATACCAACCGGAAAACCGGAT	(SEQ ID	TTGA	ITMRAIKVMQNI
					GAATTGGTTAAAAAAGTCCGAGAATGGTTT	NO: 1170)	CCCG	PVKGSPLEYVQL
					GAAATTCAATTTCAAGCATATTTTGAGGAC		GGCA	PIGKGGLGVACP
					GGAAAATCCTTCCAGAGGTTAGAGTGGWTG		ACAC	KTTALITYLVST
					ACAGGTTTGCTCACGGCTGCAATTCACAAA		ATTA	MKKLWSTDDYI
					GCTTCGGCTGGAGATGAGCAAGCTGTGGGM		TACC	RKLHTDYLKMV
					AAAATCATCAAACGTTGTCCACCTCTGGAA		ACGT	AIKETKTKEVTL
					ATTGAAGAAGGGGAAATGGCTACCCAAACT		CCAC	EDLASYLSDDK
					GAAACAAAACAAAAACCAAAAAACCAAAA		AAGG	TVCKKAVGYNS
					GAGCACAAAAGGAGCAAATAGTTCCAGCTC		ACGA	FTRVREICKTLS
					AATTCGGGAAGCCTATGCTGAAAACCGAGC		ATTC	KNKGALLSQLKI
					GAGAACCTTCAACAAAATTATTGGAAAAGA		ATAA	IAKDGKLAILVQ
					CGACAAATAGTGTGAAATTCCAATTGAAAA		TGGC	AXKDGKTKIFT
					AATTGAAAAGTTCTTCGAGAACACAACTTC		CCCT	HDHVKTLQKXL
					AAATACCAATGTTCCAACAGAAACACTAGC		CCCT	KKEINEALLHRF
					GAGGATCACTTCTGATCTTCCAAAACTCGA		AAAT	TTEKRVKSEVV
					GATTGGTAGTTGGATTGAAGAAGAGTTCAG		AAAC	RVVQEYPQCNS
					GGAGAAAGAAGTAGCCGAAGCTCTTAAAA		TCCCT	FVRDGGKVSIG
					AAACAAAGGATACTGCCCCAGGTGTAGATG		AGCA	AHRFVHKARLN
					GATTACGGTACCATCATCTGAGCTGGTTTGA		ACTG	LLACNYNTWQD
					TCCAAAAAKGAAACTGCTCACAAAACTGTA		GTGG	AATKQCRRCGY
					CAACGAATGCAGGGAGCACAAGAAAATCC		TCCG	EKETQWHILSSC
					CAGGTCACTGGAAAGAGGCAGAAACTGTAC		GCGA	PKSMGGKITERH

					TCCTCTACAAAGGGGGGGACGAGACGCAGG CCGAGAATTGGCGACCAATCAGTCTCATGC CAACCATCTGCAAGCTATACTCTAGCCTGTG GAACAAAAGAATAAAATCCGTGACAGGTGT TCTGAGCAAATGCCAAAGGGGTTTTCAAGA AAGAGAGGGTTGTAATGAAAGCATTGCAAT TCTCAGAACCGCTATTGAAGCGGCAAAAGG AACAAAAAAGAGCCTGTCAATTGCTTGGTT GGACCTTACCAATGCATTTGGCTCAGTTCCA CACGAATCGATCGAGGCCACACTAATTGCT TACGGTTTTCCGGGAATGGTAACCGAGGTA ATAAAAGACATGTATAATGGCGCATCGATT CGTGTAAAAACAAAAAACGAAAAGAGTAA ACAAATCCTGATTAAATCGGGTGTAAAACA GGGTGATCCAATCTCACCTACTCTTTTCAAC ATTTGCCTTGAAAGTGTCATTMGTCGCCAC CTAAAAAGCGCGGATGGTCACAAATGCATC GAMTCAAACATCAAATTATTGGCGTTTGCC GATGACATGGCAATCCTGTCAGATTCCAAA ACAAAACTCCAACAAGAGTTACAAAAAATG GATGATGACTGTACACCGCTCAACCTTATCT TCAAACCCGCCAAATGTGCAAGTCTGATAA TTGAGTGGGGAAAAGTACAAAAAGATCAA		AGCC GGTT CTTG CCAC TATT GCGC CCCA GGCT CGCC C (SEQ ID NO: 1293)	DSVLKTVKEMI QTGSLKNWKLK LDHELPGSTRLR PDIYLRSPNGSEI ILGDVTIPYEHGI EAMQTAWQKKI EKYEEGFKYLRS TGKKLTIVPIVV GALGSWWKPTT DSLVSLGIDKNT VKRAIPEICSTVL EYSKNIYWNHIF GDSYQKVPMFF GGEKPKGQSWK KVKPPEGKTAS NHEPPG (SEQ ID NO: 1415)

					AAAATAAAACTAAAAGGTCAATTCATCAGA AGTTTGGCCGAACAAGACACCTACAAATAT CTTGGGGTGCAAACTGGCATCGAAACGCGC GTTTCTGCAATGCAACTGATGAAAAAAACK GTCAGCGAGCTTGACAAAATAAATTGCTCT GCACTGGCTCMWTGGCAAAAACTGGACGC AGTAAAAACTTTTGTGCTCCCAAAAATGAC GTACATGTATGCAAATACTGTACCGAAACT CTCCGAGCTCAAAGAGTTCGCAAATATTAC AATGAGAGCAATAAAAGTAATGCAAAACAT TCCAGTAAAAGGTTCACCATTGGAGTATGT ACAGTTACCCATTGGAAAAGGTGGACTAGG AGTGGCATGTCCAAAAACAACTGCGTTGAT AACCTATCTGGTTTCAACAATGAAAAAATT GTGGTCCACTGATGACTATATCAGGAAACT ACACACAGACTACCTGAAAATGGTGGCCAT AAAAGAAACGAAAACAAAAGAGGTCACAC TAGAGGACCTTGCCTCCTACCTAAGTGATG ATAAAACCGTCTGCAAAAAAGCGGTTGGTT ATAATTCATTCACAAGGGTACGAGAAATCT GCAAAACGCTATCAAAAAACAAAGGAGCA CTGTTAAGCCAACTAAAAATCATTGCAAAA GATGGAAAGTTGGCTATTCTGGTACAGGCT			

					STGAAAGATGGCAAAACAAAGATTTTCACG CATGACCACGTGAAAACCTTGCAAAAASTT CTTAAAAAAGAAATAAATGAAGCCCTTCTG CACAGATTCACAACTGAAAAAAGAGTGAAA AGCGAAGTGGTGCGAGTGGTCCAAGAGTAC CCCCAGTGCAACTCCTTTGTCAGAGATGGA GGAAAAGTTAGCATTGGAGCGCATCGCTTT GTGCACAAAGCCAGGTTGAACCTGCTCGCG TGTAATTACAACACGTGGCAGGATGCAGCC ACAAAACAATGCAGAAGGTGTGGATATGAA AAAGAAACCCAATGGCACATCCTCTCATCT TGCCCAAAAAGTATGGGAGGAAAAATAACT GAAAGACACGATTCTGTGTTAAAAACAGTA AAAGAGATGATTCAAACTGGATCTCTCAAA AACTGGAAACTAAAACTTGATCATGAATTG CCAGGATCAACCAGACTTCGCCCGGATATC TATTTGAGAAGCCCAAATGGATCCGAAATA ATTCTTGGCGATGTCACAATCCCGTATGAAC ACGGAATTGAAGCTATGCAAACAGCATGGC AGAAAAAAATTGAAAAATATGAAGAGGGC TTCAAATACCTTCGTTCTACCGGCAAAAAA CTCACAATTGTGCCAATTGTGGTCGGAGCA CTAGGAAGTTGGTGGAAGCCCACAACAGAC			

					AGTCTTGTCAGTCTGGGAATCGACAAAAAT ACTGTAAAAAGAGCTATTCCAGAAATTTGC TCTACAGTACTCGAATACAGTAAAAACATT TACTGGAACCATATATTCGGGGATTCCTACC AAAAAGTACCCATGTTTTTCGGCGGTGAAA AACCAAAGGGGCAAAGTTGGAAGAAAGTG AAGCCTCCTGAAGGCAAAACTGCTTCTAAC CATGAGCCTCCAGGTTAAAAGCCAAAAGCC ACGGAGCATCGGGAAAGAAAAATGGAAAA GGACTGAAAACGAGACTGAAAAATCCCAA ACAAAACAAATCCAAAACAAACTGAAAAA AAAAAAAAAACAAAACAAAAACTGGACAG ACACTGGAAACAGTGTCAGGCAAAGTCGCC GATTATACTGTTCCACGCCTTAAAAGTCCCG AAATGGCGCAAAACAACCTGAATCTATCTG AAAGTGCTCCAAACCACGCACAACTCGGAG AAAATCAGGGACAAGTTGCTTCACGCAACG GGCTGGGACAGGTACCCCCTCCTGAAACCG CGAGGTTGAGGATGGACGGGAAGGCCGCG AGGCTTATGGCGGGTAACTCGGTTGGTGTG CTAGTAGATGATTTATATCCGACAGCCCCA ACTAAGAGGAATCCTGGGAAAGGAAAACTT GAAAAAGTTTTTACAGGGCTGGTAATAGTT			

					CAGCACAATTGTAGTCTACTGTCTTGCAACC ACAACAAACCAGTGGTTCTGCGGGTAGATC AAACTATAATTTGTGTGTTTTCTTTTACTTG ACCCGGGCAACACATTATACCACGTCCACA AGGACGAATTCATAATGGCCCCTCCCTAAA TAAACTCCCTAGCAACTGGTGGTCCGGCGA AGCCGGTTCTTGCCACTATTGCGCCCCAGGC TCGCCC (SEQ ID NO: 1047)			
NeSL	NeSL-1_CRem	.		Caenorhabditis remanei	CGCGAACCAGTCATATGACAGTCTTTATTG ATCGCGGTATAGGCGAGCGAGGCCAGATGG CCGTATGTAGCCTCCACCGTTATTTTTCGTT TTCACCTTTTTCCCCCATCCCCCCGTATGTA AATAATGGATCGTTCGGCGAAAATGGCTGT GGCACAGACAAATCACTGTTGCCCGTCATA GAAGTTGTTGTTCGTGAAGTTAAGATAAAT TGGTCTGAGAATATTTTGGTAGTAGAGTGTC TGATAATGGTAAAGAGCGGAGAAAGAGTC GTTGTAAAGAGACAAAATCTGGAAAAAGTT ATTCAGAATTTGGCAAGAATCAACTCAACT CTATTTTCCAATCTAGGAAATCAGATATTTT GCGTAGTACCCAGAATAAAAGACAGTACCA ATAAAGAGCAGGGATACAGGAAAGAGAAG CAAWTGAAATTCCATGTATCATTCCGAAGT	CGCGAACCAGTCAT (SEQ ID NO: 1171)	TAGC CGAT CGTA AAAG AAAC CGAG CCGT AACA ACAA GCAA AGTA AACA AAAG AAAA ATCA	MTVFIDRGIGER GQMAVCSLHRY FSFSPFSPIPPYV NNGSFGENGCG TDKSLLPVIEVV VREVKINWSENI LVVECLIMVKS GERVVVKRQNL EKVIQNLARINS TLFSNLGNQIFC VVPRIKDSTNKE QGYRKEKQXKF HVSFRSIKSQVP PYLRGGGDVME DTEIRGIRKLEPE

					ATAAAATCCCAAGTTCCACCATATTTGAGA		ATAA	AQLDSSKPLICR
					GGTGGGGGAGATGTAATGGAAGATACAGA		AAAG	VLYPTQGYMYK
					GATAAGAGGTATCAGAAAGTTGGAGCCAGA		GAAG	CFYPKCKGHSN
					GGCTCAGTTAGACAGCTCAAAACCGCTGAT		GTTG	GSTDLRSLKKH
					CTGCAGAGTTCTCTACCCAACGCAAGGTTA		ACCT	MVDKHFTNIEF
					TATGTATAAATGTTTTTATCCAAAGTGTAAA		CAGA	AYKCATCMFLT
					GGACATAGTAATGGATCAACAGATCTGAGA		CCCC	TGKSATALKSIK
					AGTCTGAAGAAACACATGGTGGATAAGCAT		GAGG	AHMASHHKVT
					TTCACGAATATTGAATTTGCATATAAATGTG		AGGG	MEPGKKSLVQK
					CTACGTGTATGTTTTTAACGACTGGGAAATC		AAGA	LNARLEEAAPSL
					GGCCACAGCGTTAAAATCAATAAAGGCACA		GAGA	PMPRNRSKVIQL
					TATGGCAAGTCACCACAAGGTAACGATGGA		CACC	TPEKSISELEKK
					ACCCGGTAAAAAGAGTCTCGTGCAAAAGTT		SAGA	KQTRSVAKQLS
					GAATGCCAGACTCGAAGAAGCTGCTCCATC		AAAA	TLKESAQKKEEE
					ACTTCCAATGCCGAGAAATCGATCAAAGGT		GAGA	VKIAEVKKREPR
					CATACAGTTGACCCCCGAGAAATCGATATC		GACG	LSIIPESNVRRSL
					GGAATTGGAGAAAAAGAAGCAAACTCGTTC		CAGA	AAGLEQCINPEQ
					TGTGGCAAAACAGCTTAGCACACTGAAAGA		GAAA	SVAQRIREKREE
					GTCGGCACAGAAAAAGGAAGAGGAGGTGA		AGGA	YAKASREAAAK
					AGATAGCGGAGGTCAAAAAGAGAGAACCC		GAGA	RRSSLAMKPAR
					CGTCTATCAATAATCCCAGAGTCGAATGTC		CACC	LPDKENEITLQE
					AGGCGAAGTCTGGCGGCAGGACTCGAACAA		TCTC	TKKIDDPIVIDLE
					TGTATAAACCCTGAGCAATCGGTAGCTCAG		ATAA	KECILTTVLQVP

					AGGATAAGAGAAAAAAGAGAAGAATACGC		GGAG	RNQFNSWCLEH
					CAAAGCTTCTAGGGAGGCAGCGGCAAAAA		AGGT	ETTIDAWLTDEV
					GAAGATCGAGTTTGGCAATGAAGCCAGCTA		AGGT	IHMYMCTITENR
					GATTACCAGACAAAGAAAACGAGATTACAC		CAAT	KYFMAIDPVLW
					TCCAGGAAACGAAAAAGATCGATGATCCAA		CCAA	PVYVRNGAEDL
					TCGTTATAGACCTGGAAAAAGAATGTATTC		ATGT	LRRTSCPGTFFF
					TCACTACAGTACTTCAAGTCCCAAGAAACC		AAAC	PICESNHWVLLV
					AGTTCAACTCGTGGTGTCTAGAGCATGAGA		AGAA	IEHDVYWYLDP
					CAACGATTGACGCTTGGTTAACGGATGAGG		AAAA	KGEEPKGNVEIL
					TAATACATATGTACATGTGCACAATAACCG		CCAG	LESMKRKRQYY
					AGAATCGAAAATATTTTATGGCAATCGATC		TGGG	EFPPPSQRDNVN
					CGGTTCTGTGGCCAGTCTATGTGAGAAATG		GAGG	CGVHVCLMAKS
					GAGCAGAGGATCTACTGAGGCGTACTAGTT		AAAG	IVDECGYNWYS
					GCCCAGGAACATTCTTCTTTCCAATTTGTGA		AAAG	EEDVRSFRTNM
					AAGTAATCATTGGGTTCTATTAGTGATAGA		ACTG	KDILKSKGYELC
					ACACGATGTGTATTGGTATCTGGATCCGAA		ATTTC	PEPYNRQNLLKT
					AGGCGAGGAACCAAAAGGAAATGTAGAGA		ACCC	EKQKEVILEEMI
					TTCTTTTAGAGTCCATGAAAAGGAAAAGGC		ACTA	DSFVVEDDMTF
					AGTACTATGAATTCCCACCACCCTCACAGA		AAAT	TVHRDSDHGDD
					GAGATAATGTGAATTGTGGAGTGCATGTCT		GAAT	EVEHLKTIEQEP
					GTCTTATGGCAAAATCAATAGTAGATGAAT		TTGG	ENEISEIENVEGS
					GTGGTTATAATTGGTATTCTGAAGAGGACG		AAAC	VDSVIPKLMEM
					TAAGGTCATTCAGAACCAATATGAAGGACA		AGAA	RVQTPPVINEKR

					TTCTGAAAAGTAAGGGATATGAGTTATGTC		TTTG	GKKRVSAKEKP
					CTGAGCCTTATAATAGGCAAAATTTATTAA		GAAG	RKQKEKEQKVP
					AAACAGAAAAACAAAAGGAAGTTATTCTGG		AGAA	TGKPDELVKRV
					AAGAAATGATCGATTCATTCGTTGTAGAAG		AAGA	RVWFEKEFKSY
					ACGATATGACGTTCACAGTGCATCGGGATT		GAAA	VEDGKSFQRLE
					CTGATCATGGTGATGATGAAGTTGAACATC		GGGA	WXTDVLTAAIQ
					TGAAGACCATTGAGCAGGAACCTGAAAATG		AACC	KASAGDEKAVE
					AAATAAGTGAAATTGAGAATGTAGAGGGAT		TAAA	LIEKRCPPLEXE
					CTGTAGACTCAGTCATTCCAAAGTTGATGG		GAAA	EGEMCTQTEKK
					AAATGAGAGTGCAGACACCTCCAGTGATCA		ATAG	KKPKSGKGNGG
					ATGAAAAAAGAGGTAAAAAGCGAGTATCG		TTCTC	QESMKSLMASY
					GCCAAAGAGAAACCGAGAAAGCAAAAGGA		TTGC	SENRAKTYNRII
					AAAAGAGCAAAAAGTGCCAACAGGAAAAC		CAAA	GKHSKQCEIPIA
					CAGATGAGCTGGTTAAAAGAGTAAGAGTAT		ATTCT	KVQKFFEGTTA
					GGTTTGAGAAAGAATTCAAATCGTATGTGG		GTAG	ETNVPKETLKE
					AAGATGGAAAAAGTTTCCAAAGGTTGGAAT		AGGA	MCSRLPKVEVG
					GGMTAACAGATGTTCTCACTGCAGCAATTC		ATAC	TWIEGEFSESEV
					AGAAGGCGTCAGCCGGAGATGAGAAAGCA		TTTGT	TEALKKTKDTA
					GTAGAACTGATTGAGAAAAGATGTCCACCT		CAAA	PGVDGLRYHHL
					TTGGAAWKCGAGGAGGGTGAAATGTGTAC		ACAT	KWFDPELKMLS
					CCAGACTGAAAAGAAAAAGAAACCAAAAA		GATA	QIYNECREHRKI
					GTGGTAAAGGGAATGGCGGTCAAGAAAGT		GAAA	PKHWKEAETILL
					ATGAAGTCCTTGATGGCCTCATACAGTGAG		CCAG	YKGGDESKXDN

					AACCGAGCCAAAACCTACAATAGAATAATT		TAAT	WRPISLMPTIYK
					GGTAAGCATTCAAAGCAGTGTGAGATCCCA		CTGG	LYSSLWNRRIRA
					ATAGCCAAAGTACAAAAGTTCTTTGAAGGG		TACG	VKGVMSKCQRG
					ACCACTGCCGAGACAAATGTGCCAAAGGAA		AAAG	FQEREGCNESIGI
					ACACTTAAGGAAATGTGTTCACGCCTCCCG		ACAA	LRTAIDVAKGK
					AAAGTTGAAGTGGGAACGTGGATTGAAGGT		GTAA	KRNIAVAWLDL
					GAATTCAGTGAAAGTGAAGTGACTGAAGCA		GACC	TNAFGSVPHELI
					TTGAAGAAGACAAAGGACACAGCACCAGG		TGAA	KETLESYGFPEI
					GGTAGATGGATTAAGGTACCATCACCTGAA		CTGA	VVDVVEDMYR
					ATGGTTTGATCCCGAGTTGAAAATGCTGTC		CAAG	DASIRVTTRTEK
					ACAGATCTATAATGAGTGTAGAGAACACAG		AAGG	SDQIMIKSGVKQ
					AAAAATTCCAAAGCATTGGAAAGAGGCAG		AAGT	GDPISPTLFNMC
					AGACAATTCTTCTCTATAAGGGAGGAGATG		CAGA	LESVIRRHLDRS
					AGTCMAAAMCGGATAATTGGAGGCCTATC		AAGA	VGHRCLKTKIK
					AGTCTGATGCCAACCATCTATAAACTGTATT		AATA	VLAFADDMAVL
					CTAGTCTCTGGAACAGAAGGATTAGAGCAG		CCGC	AESSEQLQKELT
					TGAAAGGGGTGATGAGCAAGTGTCAGAGA		TCAC	AMDADCSALNL
					GGTTTCCAAGAAAGAGAAGGATGTAATGAA		AAAG	LFKPAKCASLIL
					AGTATCGGAATATTGAGAACAGCCATAGAT		CCTG	EKGIVNRLNEV
					GTGGCCAAGGGCAAAAAGAGAAACATAGC		TGAT	VLRGKPIRNLME
					CGTAGCATGGTTAGATCTCACGAATGCCTTT		CGAT	NETYKYLGVQT
					GGATCAGTACCACATGAGCTGATAAAAGAA		TCTCT	GTETRVSIMDHI
					ACTCTGGAATCTTACGGATTTCCAGAAATA		TACC	TEVSREIDLVNM

					GTAGTAGACGTCGTAGAAGACATGTATCGA		TACT	SQLAMHQKLDI
					GATGCATCGATCCGTGTGACGACGCGAACG		GAAC	LKAFILPKMTY
					GAGAAAAGTGATCAGATTATGATCAAGTCA		TTGTT	MYQNTTPKLSE
					GGAGTGAAGCAGGGAGATCCAATCTCGCCT		CTCTT	LKVFANLVMRS
					ACTCTCTTCAACATGTGTCTCGAGAGTGTCA		GGCC	VKEFHNIPLKGS
					TCAGAAGGCATCTCGACAGATCAGTCGGCC		TCGT	PLEYVQLPVGK
					ATCGGTGCCTGAAAACAAAAATAAAAGTAT		AACC	GGLGVACPKNT
					TAGCCTTTGCAGACGATATGGCAGTATTAG		GGCT	ALLTFLVTIMKK
					CAGAAAGTAGTGAACAGTTGCAAAAGGAGT		AAAG	LWSSDSYIRKLY
					TGACAGCTATGGATGCTGACTGCTCAGCAC		GGAG	TDYLEEVAKVEI
					TGAATTTGCTATTCAAACCGGCTAAATGTGC		AAGG	GKFEVNLNDLA
					AAGTCTGATATTGGAAAAAGGAATAGTAAA		AATG	EFLSDERAVDSK
					CAGGTTAAATGAGGTAGTTTTGAGAGGGAA		TTAA	LFGFNAFTRVRE
					ACCGATCAGAAACCTCATGGAAAATGAGAC		TTGG	VVRSLCKNKDS
					CTACAAGTACTTAGGTGTTCAGACAGGTAC		AGAT	PLHSLKIIEREGK
					GGAAACAAGGGTTTCCATAATGGATCATAT		AGAC	LAISVQATEESIE
					AACGGAAGTGTCAAGGGAGATAGATCTAGT		ATAA	KIFTEDQEKKLM
					GAATATGAGTCAACTGGCAATGCACCAGAA		AGAT	YLLKGELNTAL
					ACTAGATATACTCAAAGCCTTCATACTTCCA		AGGT	QHRFFTQKVFKS
					AAGATGACCTATATGTATCAGAACACGACA		GGAG	EVMRVVQQHPQ
					CCTAAACTGTCAGAACTGAAAGTGTTTGCC		TGAA	SNSFVRNGGKM
					AATTTGGTAATGAGGTCAGTGAAGGAATTC		GGTC	SFSAQRFVHPGR
					CACAACATTCCCCTAAAAGGGTCACCGTTG		CTGTT	LNQLPCNYNTW

					GAGTATGTCCAACTTCCCGTAGGAAAAGGA		CTTG	AKGRTKLCRRC
					GGATTAGGAGTGGCATGTCCAAAGAACACA		AAAC	AKNENETQSHIL
					GCCTTATTAACATTCTTGGTAACCATTATGA		TAGG	QVCDYSIGNIIK
					AAAAGTTATGGTCATCTGATAGCTATATCA		AGGA	ERHDAVLYKFR
					GAAAGTTGTATACAGACTACCTAGAGGAGG		ATGT	ELIKRGSKGHW
					TGGCAAAAGTGGAAATTGGAAAGTTCGAGG		GGAA	LERTDRTVPNTG
					TCAACTTGAACGATCTAGCAGAATTCCTAA		AGAG	SQLKPDLYLESP
					GTGACGAAAGAGCAGTCGACAGCAAGTTGT		CAGA	DGKHVILADVT
					TCGGCTTCAATGCGTTCACGAGGGTGAGAG		AGGC	VPYERGIEGMQ
					AAGTGGTGAGGAGTCTCTGTAAGAATAAAG		CGCG	KAWNEKINKYT
					ATTCTCCACTACATAGTCTGAAAATAATTGA		AGGC	DGYKEIFRRQG
					AAGAGAAGGGAAACTTGCCATAAGTGTGCA		TTTA	KSLVVLPLVVGS
					AGCAACCGAAGAAAGTATTGAGAAAATCTT		GACG	LGTWWKPTEES
					CACTGAAGACCAGGAAAAGAAGTTAATGTA		GGTA	LIKLGVEKTTVR
					CCTACTGAAAGGGGAGCTAAATACAGCTCT		ACTC	RIIPETCGMVAE
					CCAGCACAGGTTCTTTACTCAAAAGGTATTC		AGTC	YSKNCYWRHIY
					AAAAGTGAAGTAATGAGAGTGGTTCAACAG		AGTT	GEKYVQTPMIN
					CATCCACAAAGTAACAGTTTTGTCAGAAAT		GCTA	GGKKPEGNDW
					GGTGGAAAAATGAGTTTTTCGGCTCAAAGA		GTGG	KKCEKGIEVPKV
					TTTGTCCACCCAGGAAGACTGAACCAGTTG		TCTTC	TN (SEQ ID NO:
					CCATGTAACTACAACACTTGGGCAAAAGGC		GGAT	1416)
					CGTACGAAGTTGTGTAGAAGGTGTGCAAAG		CCAA	
					AATGAAAATGAGACACAGTCGCATATACTG		CGGC	

					CAAGTGTGTGACTACTCAATAGGAAATATC		TTCG	
					ATAAAGGAAAGACACGATGCAGTTCTTTAT		GACA	
					AAGTTTAGAGAACTCATTAAAAGAGGGTCA		TAGT	
					AAAGGTCATTGGTTAGAGAGAACTGACCGG		GAGG	
					ACAGTACCAAATACTGGATCACAGCTGAAG		AACC	
					CCAGATCTCTATCTGGAAAGCCCAGACGGG		CTGG	
					AAGCATGTGATACTAGCCGATGTGACAGTT		GTAC	
					CCATATGAAAGAGGCATCGAAGGAATGCAA		GGAG	
					AAGGCATGGAATGAGAAAATCAACAAGTAT		AAGA	
					ACTGATGGATATAAAGAAATATTCAGAAGA		AATG	
					CAAGGAAAATCCCTAGTAGTGTTACCATTA		GAAA	
					GTAGTTGGTTCACTGGGAACGTGGTGGAAG		AGAG	
					CCCACGGAGGAAAGTCTGATCAAACTAGGT		ATAG	
					GTTGAGAAGACTACAGTAAGAAGGATAATA		GGCG	
					CCTGAGACGTGTGGAATGGTGGCTGAATAC		GGCA	
					AGTAAGAACTGCTATTGGAGACACATCTAC		AAGG	
					GGTGAAAAGTATGTTCAAACTCCAATGATA		CTAA	
					AATGGAGGAAAAAAGCCTGAAGGAAATGA		GTTC	
					TTGGAAAAAGTGTGAAAAAGGAATAGAAG		ATAC	
					TTCCTAAAGTTACTAATTAGCCGATCGTAAA		ACTG	
					AGAAACCGAGCCGTAACAACAAGCAAAGT		TCAT	
					AAACAAAAGAAAAATCAATAAAAAGGAAG		GCAA	
					GTTGACCTCAGACCCCGAGGAGGGAAGAGA		CCAC	

					GACACCSAGAAAAAGAGAGACGCAGAGAA		TAAA	
					AAGGAGAGACACCTCTCATAAGGAGAGGTA		CCAG	
					GGTCAATCCAAATGTAAACAGAAAAAACCA		TGGG	
					GTGGGGAGGAAAGAAAGACTGATTTCACCC		ATCT	
					ACTAAAATGAATTTGGAAACAGAATTTGGA		GCGG	
					AGAGAAAAGAGAAAGGGAAACCTAAAGAA		GTGA	
					AATAGTTCTCTTGCCAAAATTCTGTAGAGG		ATCA	
					AATACTTTGTCAAAACATGATAGAAACCAG		CTTTC	
					TAATCTGGTACGAAAGACAAGTAAGACCTG		GAAA	
					AACTGACAAGAAGGAAGTCAGAAAGAAAT		AGAA	
					ACCGCTCACAAAGCCTGTGATCGATTCTCTT		GTGA	
					ACCTACTGAACTTGTTCTCTTGGCCTCGTAA		ATGG	
					CCGGCTAAAGGGAGAAGGAATGTTAATTGG		ACGT	
					AGATAGACATAAAGATAGGTGGAGTGAAG		GCTG	
					GTCCTGTTCTTGAAACTAGGAGGAATGTGG		ATGT	
					AAAGAGCAGAAGGCCGCGAGGCTTTAGAC		CTGA	
					GGGTAACTCAGTCAGTTGCTAGTGGTCTTCG		CTTTA	
					GATCCAACGGCTTCGGACATAGTGAGGAAC		AAGA	
					CCTGGGTACGGAGAAGAAATGGAAAAGAG		AGTC	
					ATAGGGCGGGCAAAGGCTAAGTTCATACAC		TGAA	
					TGTCATGCAACCACTAAACCAGTGGGATCT		ATTA	
					GCGGGTGAATCACTTTCGAAAAGAAGTGAA		AAAA	
					TGGACGTGCTGATGTCTGACTTTAAAGAAG		AACA	

					TCTGAAATTAAAAAAACAGATATAAAGGCC CCTCACTATAAACTCCACAGCAACAGGTGG TCCGGCGAGGCCGGTTCTTGCCACCATTGC ACCCCAGGCTCGTC (SEQ ID NO: 1048)		GATA TAAA GGCC CCTC ACTA TAAA CTCC ACAG CAAC AGGT GGTC CGGC GAGG CCGG TTCTT GCCA CCAT TGCA CCCC AGGC TCGT C (SEQ	

							ID NO: 1294)	
NeSL	NeSL-1_TV	.		Trichomonas vaginalis	GGGTGAGTAGTCTAGTGGTATGATTCCTGTTTTGGGTACAGGAGGTCCCGAGAAGCTTCCACTGCAATCGTACGTGTACTGTGGCAACACAGCTATAACAGACAGTTTCACGCCAACCGCGAAAACGATTTTGAAGCCTGAGGAACAAAATTTAGATATCGTTTTGAAAAATATTGCAGCGTTGAATCCAGAAAATTACTCCGACTTAATCAGGAGCCTATCGAAGATGGAGTTCAGATTAGATTACCCGAAAGAAATAGAGAATTACTGGATTTCGGAAAAATTATTTAGCCAATCCATCGCATCATTGCCCATCAGTTTGTTAGTCGCATCCATGTTCTCACCTGAAGACCGTGACTTGAGTACAGAACCGTTCCACTGTAACGCTGATGGCTGTAATTTCCATTGTGACAATTGTGAAAGAATGGTTGAACACATCAGAGAGCACCATAACACTGACCCCATGATCAATACATTTGAAACAACAGAAGACACATTTAGAAGAATAACGGCCATCAAAATAGACAAGACAGGCATCGAAGAACTTAACCCTCTAAAATACAGATGCTCGTATTGCGACGAGTTATTCACCGAAGCAGAAGATCATGCCATCCATATGATTTCACATCTCACAG	GGGTGAGTAGTCTAGTGGTATGATTCCTGTT GTAGTCTAGTGGT (SEQ ID NO: 1172)	TAAG AAGA GATA AGAC GAGT GAGA AGAA CAGA AGCA TAGT AGGA TTGG CAGA GCTT AAGC GATG TCAC TCGG TACG AAAC GTGT	MIPVLGTGGPEK LPLQSYVYCGN TAITDSFTPTAK TILKPEEQNLDI VLKNIAALNPEN YSDLIRSLSKME FRLDYPKEIENY WISEKLFSQSIAS LPISLLVASMFSP EDRDLSTEPFHC NADGCNFHCDN CERMVEHIREH HNTDPMINTFET TEDTFRRITAIKI DKTGIEELNPLK YRCSYCDELFTE AEDHAIHMISHL TEKLSPDISFFFN DILRLYKTIDKP TVQNLFPETQV AIFDTLEETNRF

					AAAAATTATCACCAGATATATCTTTCTTTTT		ACCA	RLIVGREAIETIE
					CAACGACATTTTACGCCTTTACAAAACTATC		AACA	EAFPPSPPGTDR
					GACAAACCAACAGTACAAAATTTATTTCCA		CCGG	KPSIIITDTCQLR
					GAAACACAAGTCGCAATTTTTGACACACTT		ATTC	FVPCMDEPPKG
					GAAGAAACAAACAGATTCAGACTTATCGTA		CGTG	DLGILTLLLRDF
					GGAAGAGAAGCCATAGAAACAATTGAAGA		CTAG	SAHNIPIKSLNN
					AGCATTCCCTCCAAGTCCACCAGGAACAGA		GAAT	KELIADKDIDYS
					TCGGAAACCATCCATAATCATCACAGACAC		CACA	PDFVEGALANA
					CTGTCAACTCAGGTTTGTACCATGCATGGAT		AGCC	EEHDTTNSQNN
					GAACCACCAAAAGGAGATCTCGGAATTCTG		AAAA	NGRYINSAEKLT
					ACTCTACTTTTAAGAGATTTCAGCGCACACA		TAAA	EFLIQCEDYLTNI
					ATATCCCGATTAAATCACTGAACAATAAGG		AGAG	KTLEDLERFYTT
					AACTAATTGCTGATAAAGACATCGATTACA		ACAC	IKDYRVNKEVIA
					GCCCAGATTTTGTCGAAGGAGCTCTAGCCA		CACG	EDTPIFVYFLVE
					ACGCAGAAGAACATGATACAACGAACAGC		AAAA	EGKLPKPGLRCP
					CAGAACAACAATGGAAGATACATTAACTCA		TTACT	LESYEGHEDKA
					GCCGAAAAACTTACAGAATTTTTAATACAA		CACC	FESLRKLCDHFK
					TGTGAAGACTACTTAACGAACATCAAAACA		CTCC	GEIAKTSFDPKV
					CTTGAAGACTTAGAACGTTTCTACACAACG		CTCA	HTIDIWVEFLAQ
					ATTAAAGACTACAGAGTCAACAAAGAGGTT		AACA	AYGTGTFVYKD
					ATCGCCGAAGATACACCAATCTTTGTATATT		GATA	ENGNIDLDTHVF
					TCCTAGTAGAAGAAGGGAAATTACCAAAAC		ATAA	KCPYADCSYTN
					CAGGTCTTAGATGCCCACTTGAATCATACG		TATT	NDRSKLMDHM

					AAGGACACGAAGACAAGGCATTCGAATCAC		AACC	KTKKHAKNVYI
					TGAGAAAACTTTGCGACCACTTCAAAGGAG		TCCC	ERYGFFWGIVIE
					AAATCGCGAAAACGAGCTTTGACCCAAAGG		ATCC	GVNRPKGIVYPT
					TTCACACCATAGACATCTGGGTTGAATTTTT		ATCA	LKDIKEHACRK
					GGCCCAAGCCTATGGCACAGGCACGTTTGT		GTCC	CPEAGCNTYVT
					CTACAAAGATGAAAACGGAAACATCGACCT		GTAT	ELSDIKEHLKKK
					TGATACGCACGTATTCAAATGCCCTTATGCA		GGTC	HKSTTAGVDGEI
					GACTGCTCATACACGAACAACGACAGATCA		TGAT	AHTDATYCWIT
					AAACTCATGGACCACATGAAAACGAAGAA		AACA	KEELDALHAER
					ACACGCCAAGAACGTATACATCGAGAGATA		GACT	ARERAEQVDNT
					CGGCTTCTTTTGGGGTATTGTCATAGAAGGA		AGCA	PVQQIINADNNE
					GTCAACCGACCAAAAGGAATCGTCTACCCG		CCAC	ENNENQEDNGN
					ACACTCAAAGACATCAAAGAACACGCTTGT		ATCC	NEEADALDPPN
					CGCAAATGTCCAGAAGCAGGATGCAACACA		ATGA	NTTETEDEAVH
					TATGTAACAGAATTGAGCGACATCAAAGAA		TACA	AVIINPPATEEEE
					CATCTAAAGAAGAAACATAAGTCTACAACA		CTCA	VAIIAEARRNIPE
					GCAGGAGTAGACGGAGAAATCGCGCACACT		TTGG	LQQAEERGCVT
					GATGCTACATACTGCTGGATTACCAAAGAA		AGTG	PKMTSLVRLKL
					GAACTCGACGCATTACATGCCGAGAGAGCA		AAAA	LKGGGELFNKK
					AGAGAAAGAGCAGAGCAAGTAGACAACAC		CCAC	LTPLATRYAAT
					TCCAGTACAACAGATAATTAATGCTGACAA		CAAC	GNTEADKIKVD
					CAATGAAGAGAACAACGAGAACCAAGAAG		AACA	YLTLKCNAALR
					ACAACGGAAACAACGAAGAAGCAGATGCC		AATC	EMIYTNNHSESK

					CTCGACCCGCCAAATAACACAACAGAGACA		CACC	FMTAENGEDTA
					GAAGATGAAGCGGTTCATGCCGTCATCATC		TAGA	PPPRISEDTRDRI
					AATCCACCAGCAACAGAAGAGGAAGAGGT		CCAA	QKAANEIKGTLI
					AGCCATCATCGCCGAGGCAAGAAGAAACAT		ATCC	KVVKHISHARC
					TCCAGAACTCCAACAAGCAGAAGAGAGAG		TGCC	LKDSTRDDEHN
					GCTGCGTTACACCGAAAATGACATCACTCG		CCAC	KFVEMIAKIKND
					TCCGATTAAAACTATTGAAAGGAGGAGGAG		CTCC	LRDNKFEQYNIE
					AACTTTTCAACAAGAAACTCACTCCATTAG		ACCC	EIFQGPISDQSIL
					CCACAAGATACGCAGCTACAGGAAATACAG		AAGT	NIVNTEDNNEFI
					AAGCAGACAAAATCAAGGTAGATTACTTGA		AGCT	KKMDYINRILGT
					CACTAAAATGCAATGCCGCCTTGAGAGAAA		CGCT	PQDASPYARKK
					TGATCTACACCAATAACCACAGCGAATCAA		TCGC	LQACFADNPTK
					AGTTTATGACAGCAGAAAATGGAGAAGACA		TCGC	TLRNIILADKVP
					CAGCACCACCGCCAAGGATATCGGAAGACA		TCAC	QQSLKPSEYLDY
					CAAGAGATCGCATTCAAAAAGCAGCCAATG		CTAA	YGPQWANEAEG
					AAATAAAAGGAACTCTCATCAAAGTAGTCA		AACT	YENFLHHDYAL
					AACACATAAGTCACGCGAGATGCCTCAAAG		TTGCT	PERYGQVFAND
					ACAGCACGAGAGACGATGAACACAATAAA		CGCT	FLDFMTNESKIIE
					TTCGTCGAAATGATTGCAAAAATCAAAAAC		CGCT	VIRNKNHLSAH
					GATCTCAGAGATAACAAATTCGAACAATAT		TCGC	GLDGIPNSVYM
					AACATTGAAGAAATATTTCAAGGACCGATC		TCGC	LFPVSAAKFLSIL
					TCCGACCAGAGTATTCTCAACATCGTCAAC		TCGT	FRSIIISGHIPDC
					ACGGAGGACAACAACGAATTCATCAAGAA		CTTA	WKLSKTVMLFK

					AATGGATTACATTAACCGAATTCTCGGAAC		ACCC	KDDPSLAKNWR
					ACCACAGGATGCATCACCATATGCAAGGAA		TTTCC	PIGITSCTYRIFM
					GAAGTTACAAGCATGTTTCGCCGATAACCC		GAAT	TLVNKALQMIP
					AACAAAGACTCTCAGAAACATAATCTTAGC		AAAC	MFHAMQKGFV
					CGACAAAGTTCCACAACAATCATTGAAGCC		ACTT	RGATLSEHIAVA
					AAGCGAATACCTTGATTACTACGGACCTCA		ACAA	NEVLCQSTRTQS
					ATGGGCAAACGAAGCTGAAGGCTACGAAA		TTCCC	EMFQTAIDFTNA
					ACTTCCTGCATCATGACTACGCGTTACCGGA		GGCT	FGTVPHQLIFDS
					GAGATATGGCCAAGTTTTCGCAAACGACTT		CGCC	LEAKKVPDSIIN
					CCTCGACTTCATGACAAACGAATCGAAGAT		CCAT	LLKDLYKGART
					CATCGAAGTAATCCGCAACAAGAATCATTT		TTTTT	AIYTRHAHSEIV
					ATCGGCACACGGCCTCGATGGAATTCCGAA		(SEQ	PVRRGVIQGCPL
					CTCAGTTTACATGCTATTCCCAGTCAGCGCC		ID NO:	SPILFNCCLDPLL
					GCAAAATTCCTCAGTATATTATTCAGATCAA		1295)	YAVQRRHFEDG
					TCATCATATCAGGTCACATCCCAGACTGCTG			YRFQDKAGQYS
					GAAGCTCTCCAAGACAGTGATGCTTTTTAA			IAIQAYADDVLV
					GAAGGACGACCCATCGTTAGCAAAGAACTG			ISPTHEGMQRIL
					GAGACCAATCGGCATCACGTCATGCACTTA			NTVDEFQKIAKL
					CAGAATCTTCATGACTTTAGTCAACAAAGC			KVAPQKCVTLA
					GTTACAGATGATCCCAATGTTCCACGCAAT			KTSTAIQPFRIGP
					GCAAAAAGGTTTCGTTCGCGGAGCAACACT			DEIPIKTSMDNIT
					GAGTGAGCACATTGCAGTCGCGAACGAAGT			YLGIPISGTKTSR
					CCTTTGCCAATCAACCAGAACACAGTCTGA			FAAATGILEKVK

					AATGTTCCAAACAGCAATCGATTTCACGAA			AQIRVVFASHLA
					CGCTTTCGGCACAGTTCCTCATCAATTGATC			LSQKIIALRVFIL
					TTTGATTCTCTCGAAGCGAAGAAAGTTCCC			PQLDFYMFHNV
					GATTCGATCATCAATCTGCTCAAGGACCTCT			FRVNDLKATDQ
					ACAAAGGAGCAAGAACGGCTATCTATACAA			MIRGLIDKEAPT
					GACATGCACACTCCGAGATAGTTCCGGTTC			SNIPVSFFYMPK
					GCAGAGGTGTCATCCAAGGCTGTCCACTCA			NKGGFGLVKLE
					GTCCAATCCTCTTCAACTGCTGCTTAGATCC			LRQPQLVLTKFA
					TTTATTATATGCAGTCCAGAGGAGACACTTT			RLWLSQQAETK
					GAGGACGGTTACAGATTCCAAGACAAAGCA			AFFHTMAQEEK
					GGACAGTATTCAATTGCCATTCAAGCTTAC			SFRKVVEDQEN
					GCTGACGACGTTCTAGTCATCTCTCCAACAC			GFLGIKMENGKI
					ATGAAGGAATGCAAAGAATCTTAAACACAG			VQKNERSKRTN
					TAGATGAATTCCAGAAAATTGCGAAACTCA			CFITQAAKAAD
					AAGTTGCACCACAGAAATGCGTCACACTTG			KLEVRFKEWDK
					CCAAAACATCCACTGCAATCCAACCTTTCC			GGIQVRGVGEN
					GCATTGGTCCAGACGAAATCCCAATCAAGA			ATDWYRSKHIG
					CGAGCATGGACAACATCACATATCTTGGAA			QISPLIGRVIQQR
					TACCAATCTCTGGAACAAAGACATCAAGAT			QYEEFKKDETH
					TTGCAGCTGCAACTGGCATTCTGGAAAAGG			SHTFCEPAALAE
					TCAAAGCACAGATCAGAGTCGTCTTCGCGT			SHDIMKRPQAV
					CACATCTCGCTCTCTCTCAGAAGATTATCGC			PNNLYSAAIALR
					TCTCAGAGTCTTCATCTTGCCACAACTTGAC			TNTAPTPANMH

					TTTTACATGTTCCACAACGTATTCAGAGTCA			FHNPEVLANCPL
					ATGACTTGAAAGCGACAGATCAGATGATCC			CGCQSCTLFHTL
					GAGGCCTGATCGACAAAGAAGCGCCGACGT			NMCRNRFSLYK
					CAAACATTCCGGTTTCATTTTTCTACATGCC			WRHNIICDDIYQ
					GAAGAACAAAGGCGGCTTTGGACTCGTTAA			FIHDHYPGVTIK
					ATTGGAACTTCGCCAGCCTCAGCTCGTTCTC			CSARITSDGYQT
					ACTAAATTTGCGAGGTTATGGTTAAGTCAA			TGPELDDTVKD
					CAAGCAGAAACCAAAGCCTTCTTTCACACA			LLPDLVVYDEA
					ATGGCTCAAGAAGAGAAGTCATTCCGCAAG			NKMIKIIEVTCP
					GTCGTCGAAGACCAAGAAAATGGTTTCTTA			YGTDNNVGNSL
					GGCATCAAGATGGAAAACGGCAAAATTGTC			DAAYDKKVNK
					CAGAAGAACGAAAGATCCAAACGCACAAA			YKSLAEQTERLF
					TTGTTTCATCACACAGGCGGCTAAAGCAGC			NWTTTLSIIVVS
					AGACAAACTGGAAGTCAGATTCAAAGAATG			SLGVIPLRTKLD
					GGACAAAGGAGGCATACAAGTCAGAGGTG			ALRISPADHIQL
					TAGGAGAAAATGCAACAGACTGGTACCGCT			LKRLSMHAIAAS
					CGAAACACATCGGCCAAATCTCACCCTTAA			ACIVFEKVPEFF
					TCGGTCGCGTCATCCAACAGAGGCAGTACG			GMRCRPLPGRV
					AGGAGTTCAAGAAAGACGAAACACACTCAC			TAPNAAIPPNNN
					ACACTTTCTGCGAACCAGCAGCGCTAGCGG			ENNNDTDHGQE
					AGTCACACGACATCATGAAGAGACCACAAG			NQQATSEEQPT
					CTGTTCCAAACAACCTCTACTCAGCGGCTAT			NNGNAQEDNGQ
					TGCTCTCCGTACAAACACAGCTCCAACCCC			GEQINNSTEQTIS

					AGCAAACATGCACTTCCACAACCCAGAAGT TTTGGCTAATTGTCCATTGTGCGGATGCCAA TCCTGCACTCTCTTCCACACATTGAACATGT GCAGAAACCGTTTCAGTCTATACAAATGGC GCCACAATATCATATGCGATGACATTTACC AATTCATTCACGATCACTATCCAGGAGTAA CCATCAAATGCTCGGCGAGAATTACAAGTG ACGGCTACCAAACAACAGGCCCAGAGCTCG ACGACACAGTTAAAGATCTCCTCCCAGACC TTGTTGTCTACGATGAAGCGAACAAGATGA TCAAGATCATTGAAGTCACATGCCCTTACG GCACGGACAACAATGTTGGCAACTCTCTTG ACGCGGCATACGACAAAAAGGTTAACAAGT ATAAGAGCCTTGCTGAACAAACAGAGAGAT TATTTAACTGGACCACGACGCTCTCAATTAT CGTAGTCTCATCACTAGGAGTCATCCCTCTC CGTACAAAACTCGACGCATTGAGAATCTCA CCTGCAGATCACATACAGCTACTCAAGAGA CTTTCGATGCACGCGATAGCTGCGAGTGCTT GCATTGTTTTTGAAAAAGTGCCAGAATTCTT CGGTATGCGCTGCCGTCCCCTCCCAGGACG AGTCACAGCTCCCAATGCAGCGATCCCACC AAACAACAATGAAAACAATAACGACACAG			VDQIIEEDAENN AIEQALDQPDED EFLN (SEQ ID NO: 1417)

					ATCATGGTCAGGAGAACCAACAGGCAACCT CTGAAGAGCAACCAACCAACAATGGAAATG CTCAAGAAGACAATGGCCAAGGCGAACAA ATAAATAATTCAACCGAACAAACTATCTCT GTTGATCAAATCATCGAAGAAGATGCTGAG AACAACGCGATAGAACAAGCCTTAGACCAA CCCGATGAGGACGAATTCCTTAACTAAGAA GAGATAAGACGAGTGAGAAGAACAGAAGC ATAGTAGGATTGGCAGAGCTTAAGCGATGT CACTCGGTACGAAACGTGTACCAAACACCG GATTCCGTGCTAGGAATCACAAGCCAAAAT AAAAGAGACACCACGAAAATTACTCACCCT CCCTCAAACAGATAATAATATTAACCTCCC ATCCATCAGTCCGTATGGTCTGATAACAGA CTAGCACCACATCCATGATACACTCATTGG AGTGAAAACCACCAACAACAAATCCACCTA GACCAAATCCTGCCCCACCTCCACCCAAGT AGCTCGCTTCGCTCGCTCACCTAAAACTTTG CTCGCTCGCTTCGCTCGCTCGTCTTAACCCT TTCCGAATAAACACTTACAATTCCCGGCTCG CCCCATTTTTT (SEQ ID NO: 1049)			
NeSL	NeSL-2_CBre	.		Caenorhabditis brenneri	GACTCGCCTTGGGGAAGGTWTTTCAGGGGK SAATTGCCGMAGGCAAGGCAGCCCCCSMM	GACTCG CCTTGG	TAAA CCGG	MRYHXSNXPAX RTSDNXWRSIX

					TAGCTTACAAAGTAAGTACMCATTTTCATTT	GGAAGG	CTCCT	KDVRRPDPSTIE
					CTTGTGAATTCTTTAAACATATTTTTCTTGTT	TWTTTC	CTGG	EKSRYNRSIGIP
					TTTTGATTTCTTTTTTCTCTACCTTCCCCCAA	AGGGGK	GAGG	DSLKXRSSAVRS
					TTCTTCCCCTCATCTTGTGTATACATCCCCCT	SAATTG	AGGT	XSSXPPSGPQDV
					CCTCCAACCAATCAATACATTGACCTCTCTC	CCGMAG	ATGT	RLXNSPSLDDRR
					TTCTGTCAAAAAATCAATACTAGTATATTGT	GCAAGG	CAGA	RLVDCETTLGSY
					CCCTTGTATAGTATTATTTGACGTCGTCTTT	CAGCCC	GGAC	REWTDKPMMG
					GTATTAGGAGTAGGTAACAAWCTGTGTATG	CCSMMT	ATTCT	KMTYAAVTKRA
					GCTTCAAAAAGCATGCACAAACWCCTGTCA	AGCTTA	CCGT	PPRPQTGGARLS
					AAAAGTAWTTCCCATCMTGTGAATAGCTCA	CAAAGT	GGGC	TNLLADEMEIK
					ACGACWKGAAGMCCAATGATATGAGATAT	AAGTAC	GGAT	YRDTNDIRLVID
					CATMGCTCCAACMACCCAGCASCCCGCACC	MCATTT	GGGA	LPNPHLIKCPLC
					TCAGACAATCAMTGGAGATCAATCCMAAA	TCATTT	GGAG	KSCISARGRGAN
					GGACGTCCGTCGCCCAGATCCGTCAACTAT	CTTGTG	TAGG	ALKYMKRHIAD
					CGAGGAGAAAAGCAGGTATAACAGGTCCAT	AATTCT	GTAA	AHHLNADFVYK
					AGGTATTCCAGATTCGCTCAAAGAWCGGAG	TTAAAC	CGAC	CSRCQEHEPENV
					CAGTGCAGTCCGCAGCAKGAGCAGCCMAC	ATATTT	CCGT	CGAKWIVNHLK
					CTCCGTCAGGTCCACAGGACGTCCGTCTCW	TTCTTG	CATT	RVHGYTLEDAV
					CCAATTCGCCATCTCTCGATGATAGGAGAA	TTTTTTG	CTGG	STAKPSTRQQIA
					GGTTAGTWGATTGTGAAACAACACTAGGGT	ATTTCT	ATGC	NAFNDSAPFIDA
					CATACCGCGAATGGACSGATAAACCAATGA	TTTTTCT	CTAA	RKTSDVPEKKSR
					TGGGAAAGATGACGTATGCGGCAGTGACAA	CTACCT	ACCA	EAGLEKFLAPTK
					AAAGAGCGCCCCCAAGACCGCAAACGGGA	TCCCCC	CCAC	SEDTREKTPPST

					GGAGCCCGGTTGAGCACCAATCTCCTAGCA	AATTCT	AATC	RKSSESSEASIQS
					GATGAAATGGAGATAAAGTATCGAGACACC	TCCCCT	TGTC	TIQETLSESSDTL
					AATGACATCCGCCTTGTCATAGACCTTCCCA	CATCTT	AAGG	TVQEIINISSEDE
					ATCCCCACCTCATCAAGTGTCCGCTCTGTAA	GTGTAT	CAAA	MDEEPPKRRVN
					AAGCTGCATAAGTGCGCGGGGAAGAGGTGC	ACATCC	GTGC	VWALIHENGKD
					TAATGCGCTGAAGTACATGAAAAGGCACAT	CCCTCC	CCCA	AWIDSDLMVIFL
					AGCCGACGCCCACCACCTCAACGCCGACTT	TCCAAC	AAAG	ESRARGYESCSII
					CGTCTACAAATGTAGCAGGTGTCAAGAGCA	CAATCA	CACA	DPLNFICTDMSY
					TGAACCAGAAAATGTATGCGGCGCGAAGTG	ATACAT	CGCG	LTTIVRRRMEEG
					GATTGTGAATCATCTCAAAAGAGTACATGG	TGACCT	TGGA	YKKIIFPLCAND
					CTATACTCTAGAAGATGCCGTATCCACAGC	CTCTCT	TCGG	HWTLVTITGSTA
					AAAACCCTCTACAAGGCAGCAGATTGCAAA	TCTGTC	TTTG	TFYDPMGNEPT
					CGCCTTCAACGACTCTGCTCCATTCATAGAC	AAAAAA	GATG	ETVKKMIDELD
					GCCCGGAAAACATCCGATGTGCCAGAGAAG	TCAATA	CCGA	LEMQLAPSNSPR
					AAGAGCAGAGAGGCAGGACTTGAGAAGTT	CTAGTA	CTGA	QRDSWNCGVFV
					CCTGGCCCCTACAAAGTCCGAGGACACAAG	TATTGT	GCCA	MKMAEAYIKDT
					GGAAAAAACCCCGCCCTCCACCAGAAAGTC	CCCTTG	GAGG	QWDLTDVDTDV
					CTCTGAAAGTTCAGAGGCATCAATCCAATC	TATAGT	GCAA	KTFRRSLLTELK
					GACCATCCAAGAGACTCTTTCGGAGTCGTC	ATTATT	AGTC	AKFNIFAEDIQT
					AGACACATTGACCGTCCAAGAAATAATCAA	TGACGT	GAAG	YRPPSRKALTRN
					TATCAGCAGTGAAGATGAAATGGACGAGGA	CGTCTT	GCCG	SQSPVVVCHKC
					GCCACCGAAACGGCGTGTCAATGTCTGGGC	TGTATT	GTAG	SRPATPIQDVSR
					CTTGATCCATGAGAATGGCAAGGACGCCTG	AGGAGT	GCTC	MEVEEAPVLVP

					GATAGACTCAGACTTGATGGTCATATTCCTG	AGGTAA	CCGG	TPEEPPQEWTFV
					GAATCAAGAGCAAGAGGATATGAATCATGC	CAAWCT	CGGG	GKNRKRGVTSR
					AGCATCATAGACCCTCTGAACTTCATTTGCA	GTGTAT	TTGTC	TPNTSPEAKRPA
					CTGACATGTCCTATCTGACCACAATAGTCA	GGCTTC	CGTC	FPPVPLKPSANR
					GAAGGCGCATGGAAGAAGGCTACAAGAAA	AAAAAG	ATAG	WHFPEEETEKM
					ATCATATTTCCATTATGTGCAAATGACCACT	CATGCA	TCAG	EVSSADEVKNST
					GGACACTGGTCACGATAACAGGTAGCACGG	CAAAC	TGGT	PPKPPKIPNLLA
					CCACCTTTTACGATCCAATGGGAAATGAGC	WCCTGT	GCGC	MKIASPVPLKRG
					CAACTGAGACTGTCAAGAAGATGATCGATG	CAAAAA	CTAC	NPSKKHGKGHM
					AGCTCGACCTTGAAATGCAATTAGCCCCGT	GTAWTT	ACCC	MNTARKGPTKK
					CAAACTCTCCTAGACAGAGAGACTCGTGGA	CCCATC	AACT	EMPKGEPANLIV
					ACTGCGGCGTTTTCGTCATGAAAATGGCGG	MTGTGA	GCTA	KIRSWFDEQLK
					AAGCGTACATCAAGGATACGCAATGGGATC	ATAGCT	TGAC	MYKDEGSNLQR
					TCACGGACGTAGACACGGACGTCAAAACGT	CAACGA	ACAC	LTWLSDSLTAAI
					TCAGAAGGAGCCTCCTAACAGAGCTCAAAG	CWKGA	AAGG	GKAFNGNKYIV
					CAAAGTTCAACATCTTTGCCGAGGATATCC	AGMCCA	ACAA	DQIIKRNPPPLVE
					AGACCTACCGGCCACCCTCAAGGAAAGCCT	ATGAT	CCCA	KGAMSTQTSRK
					TAACGAGGAACAGCCAATCGCCTGTCGTTG	(SEQ ID	AAAT	RDEFKPRERMA
					TTTGTCACAAGTGCTCTCGGCCAGCCACACC	NO: 1173)	AAAT	QEPNEPLRIQYA
					GATCCAGGATGTGAGCAGAATGGAAGTGGA		AAGC	KNRQKTFFKIIG
					AGAAGCGCCAGTGCTGGTACCGACTCCTGA		CAAG	KQSEQCTINIET
					AGAGCCTCCACAGGAATGGACCTTCGTCGG		GCGG	VEQHFRKTLKA
					AAAAAACAGAAAGCGTGGTGTGACAAGCC		CGTT	PVVSENAIKTVC

					GAACCCCGAACACGTCGCCGGAAGCCAAGC		AGCT	GSIKKVLMPKTI
					GACCGGCTTTCCCACCAGTACCCCTCAAAC		TCGA	EDPISSVEVKSIL
					CATCAGCCAACAGATGGCACTTTCCAGAAG		GCTA	TKVKDTSPGTD
					AGGAAACTGAAAAGATGGAGGTCTCAAGTG		ACAA	GVKYSNLRWFD
					CCGACGAGGTGAAGAACTCTACCCCTCCAA		GCTC	PEGERLAKLFEE
					AACCACCCAAGATACCTAATCTTCTCGCGA		CCCG	CRKHREIPSHW
					TGAAAATCGCCAGTCCCGTACCTCTGAAGA		AGAG	KEAETILLPKDC
					GAGGAAACCCGTCAAAGAAGCACGGTAAA		GATG	SDEEKKKPENW
					GGACACATGATGAATACAGCGAGAAAGGG		GTTG	RPIALMATIYKL
					TCCGACAAAGAAAGAGATGCCCAAAGGGG		CCAC	YSAVWSRRISG
					AACCAGCGAACCTTATAGTTAAAATCAGAA		AGGG	VQGVISPCQRGF
					GCTGGTTTGATGAGCAACTGAAGATGTACA		CACC	QSLDGCNESIGI
					AGGATGAAGGGTCCAATCTACAAAGACTCA		ATCC	LRMCIDTASVLN
					CATGGTTATCGGACTCTCTGACCGCCGCCAT		TGGG	RNLSCSWLDLT
					CGGAAAAGCATTCAATGGCAACAAATACAT		GAAC	NAFGSVPHELIR
					AGTGGACCAAATCATCAAGAGAAACCCACC		GACC	RSLESFGYPQSV
					ACCACTTGTTGAAAAGGGCGCAATGTCAAC		CGAT	IQIVTDMYKGAT
					ACAGACAAGCCGAAAGAGAGACGAGTTCA		CTTTC	MKVKTADQKT
					AGCCAAGGGAGAGAATGGCCCAAGAGCCC		GGAT	QSIKIEAGVKQG
					AATGAGCCGCTTCGTATTCAATATGCCAAG		GCCC	DPISPTLFNICLE
					AATAGGCAAAAAACGTTCTTCAAGATCATT		AACC	GIIRMHQMREK
					GGGAAACAGTCTGAACAGTGCACCATCAAC		ACCG	GYDCVGHKVRC
					ATTGAGACTGTCGAACAGCACTTCCGAAAA		CCAA	LAFADDLAILTN

					ACACTCAAGGCTCCTGTAGTCTCAGAGAAT		TCTGT	NKDEMQEVIDK
					GCAATTAAAACGGTCTGCGGAAGCATCAAG		CAGG	LDADCRSVSLIF
					AAAGTATTGATGCCAAAGACCATAGAAGAC		CAAC	KPRKCASLTIVR
					CCAATCTCCTCCGTAGAAGTCAAATCCATCT		GTGC	GAVDKYAKIRIN
					TGACGAAAGTGAAAGACACGTCACCAGGA		CCCA	GDAIRTMADRD
					ACAGATGGTGTCAAGTATAGCAATCTACGC		AAAG	TYRYLGVKTGV
					TGGTTCGACCCAGAAGGGGAACGCCTCGCC		CACA	GGRASETEALIQ
					AAACTGTTCGAAGAATGTCGCAAGCATAGA		CGTG	VVKELQKVHET
					GAAATACCCAGCCATTGGAAGGAGGCGGA		CGGA	DLAPHQKLDILK
					GACGATCTTATTGCCAAAGGACTGTTCAGA		GCGG	TFLLPRLQHLYR
					TGAAGAGAAGAAAAAGCCGGAAAATTGGC		TTGG	NATPKLSELREF
					GTCCCATCGCCCTCATGGCAACAATCTATA		ATGC	ENVVMKSVKRY
					AGTTGTATTCAGCAGTGTGGAGCAGAAGAA		CGAC	HNIPIKGSPVEY
					TCTCCGGTGTTCAAGGGGTAATTAGCCCGT		TGAG	VQIPVKKGGLG
					GCCAAAGAGGCTTTCAGTCCCTCGACGGAT		CCAG	VLSPRLTCLITFL
					GCAATGAGTCGATCGGAATATTGCGCATGT		AGGG	TSTLCKLWSDD
					GTATTGACACCGCTTCCGTACTCAATAGGA		CAAA	PFISSIHKDALSR
					ATCTCTCTTGCTCATGGCTTGATCTCACCAA		GTCG	ITVKAMGLTTQS
					CGCCTTCGGGAGCGTTCCTCACGAGCTGAT		TAGG	ATIKETCEYLNT
					AAGGAGATCCCTAGAATCATTCGGATATCC		CCGG	RKAVTKGGYSL
					ACAATCAGTTATCCAAATCGTGACTGACAT		TAGG	FCRMNESLRTLS
					GTACAAGGGAGCAACGATGAAAGTCAAAA		CTCC	VIQGAPLKSMEF
					CGGCAGATCAAAAGACGCAAAGCATCAAA		CGGC	IPVNNEIGIAVQ

					ATAGAAGCGGGGGTGAAACAAGGAGACCC		GGGT	ATKDSEIKVFTK
					CATTTCTCCAACCCTATTCAATATTTGCCTT		TCTCC	ADSLKLMSKLK
					GAAGGCATCATCAGGATGCATCAGATGAGA		GTCG	DLVRSAMLKRF
					GAGAAAGGGTACGATTGTGTCGGGCATAAA		TAGT	LEEKSVKSRVTQ
					GTTCGCTGCCTAGCGTTCGCCGACGACCTTG		CAGT	VLQHHPQSNRF
					CGATTCTAACGAACAACAAAGATGAAATGC		GGTG	VRDGRNCSIAA
					AGGAAGTTATCGACAAGTTGGATGCAGACT		TGCC	QRFVHPARLNL
					GTAGAAGCGTGTCGTTGATCTTTAAACCTA		TACA	LSCNANTYDVN
					GGAAGTGTGCATCTTTGACTATCGTGAGAG		CCTA	HPKGCRRCQAD
					GTGCAGTTGATAAGTATGCAAAGATCAGAA		ACTG	FESQQHILQNCH
					TAAATGGAGACGCGATCAGAACAATGGCGG		CTAT	YSLAGGITQRHD
					ATAGAGACACCTATAGATATCTGGGTGTAA		GACA	RVMNRILQEIGN
					AGACCGGAGTTGGTGGAAGAGCATCGGAA		AGCG	GRKAHYKIMVD
					ACGGAAGCTTTAATTCAGGTGGTCAAGGAG		TATA	METGATRERPDI
					CTCCAAAAGGTCCACGAAACCGACCTGGCT		GGAG	IMEERDGPEVLL
					CCACATCAAAAACTTGACATCCTGAAGACG		GCCC	ADVTVPYENGV
					TTCTTACTGCCAAGACTGCAGCATCTCTACA		GGAA	QAVERAWDKKI
					GAAATGCCACTCCTAAACTGTCAGAGTTGA		AAAC	EKYKHFLDYYR
					GAGAGTTCGAGAACGTTGTTATGAAATCAG		AAGC	KIGKKATILPLV
					TGAAACGGTATCATAACATACCAATAAAGG		CAAG	VGSLGTYWPDT
					GCTCGCCTGTGGAATATGTCCAAATCCCTGT		GCGG	SHSLKMLGLSD
					CAAGAAGGGTGGACTAGGAGTTCTATCTCC		CGTT	GQIRNVIPEICQI
					TCGACTCACATGCCTGATCACTTTCCTTACC		AGCT	ALESSKNIYWK

					TCGACCCTCTGCAAGCTATGGTCCGATGATC		GAGA	HILGDSYKTVEG
					CATTCATATCTTCTATCCACAAAGACGCACT		GCTA	LFCQRNNKEVR
					AAGCAGAATCACAGTGAAAGCGATGGGACT		ACAA	FEGKGEKHHVS
					CACCACTCAAAGTGCCACAATAAAAGAGAC		GCTT	QRFQPLKCEKV
					ATGTGAGTACTTAAACACAAGGAAAGCTGT		CTCG	RTMKSTKEEGR
					CACGAAAGGAGGATATAGTCTATTCTGCCG		TGGA	SRSNAKKGPNW
					CATGAATGAATCTCTCCGCACGCTGTCTGTC		TGGG	RRSKSESDGRSV
					ATCCAAGGTGCTCCACTGAAATCAATGGAA		TGCC	SKGRYWRDPSN
					TTCATCCCGGTGAACAATGAAATCGGTATA		AGAG	KPPHSKMTQSA
					GCGGTACAAGCCACGAAGGATTCCGAGATC		GGCA	LAKR (SEQ ID
					AAAGTCTTCACAAAGGCTGACAGCCTGAAG		CCAT	NO: 1418)
					CTAATGAGTAAGCTAAAAGATCTGGTCAGA		CCTG	
					TCTGCTATGCTCAAACGGTTCCTCGAAGAG		GTGG	
					AAGAGTGTTAAAAGCAGAGTCACCCAAGTA		GTGG	
					CTCCAACACCACCCACAATCCAATAGATTC		ATGG	
					GTCAGGGACGGCCGAAACTGCAGCATAGCA		GGGG	
					GCWCAGAGATTCGTGCACCCKGCCCGTCTG		AGCT	
					AACCTCCTCTCCTGCAACGCCAACACATAC		TGGG	
					GATGTTAACCATCCAAAAGGCTGCAGAAGG		AACG	
					TGCCAGGCTGACTTCGAGTCTCAGCAGCAC		TCCC	
					ATCCTGCAGAACTGTCACTACAGTTTGGCA		GATC	
					GGGGGAATAACCCAGAGACATGACAGAGT		CTTC	
					CATGAACAGGATTCTGCAGGAAATTGGAAA		GGAT	

					CGGGAGAAAAGCTCACTACAAGATAATGGT		GCCC	
					GGATATGGAAACCGGCGCCACAAGAGAAA		AGAC	
					GACCGGATATCATCATGGAGGAAAGGGACG		CACC	
					GTCCAGAAGTGTTACTAGCCGACGTGACAG		GCAA	
					TGCCCTACGAGAATGGGGTTCAAGCGGTTG		TCTGT	
					AGAGGGCGTGGGATAAGAAGATAGAAAAA		CAAG	
					TACAAGCACTTCCTAGATTACTACCGCAAA		GCAC	
					ATCGGAAAGAAGGCTACGATTCTTCCCCTA		CGTG	
					GTAGTCGGTAGCCTAGGAACCTACTGGCCC		CTCC	
					GACACAAGCCACTCACTGAAGATGCTTGGC		AAAA	
					CTTTCCGACGGTCAAATAAGGAATGTTATA		GCAC	
					CCTGAAATCTGCCAAATTGCACTGGAATCC		ACGC	
					TCCAAAAATATCTATTGGAAGCACATTCTC		GCGG	
					GGTGATAGCTACAAAACGGTGGAGGGACTA		GTTG	
					TTTTGTCAGAGGAATAACAAAGAAGTCCGA		GTTT	
					TTCGAAGGAAAAGGTGAAAAACACCACGTG		GGAT	
					TCACAAAGATTCCAACCTCTGAAATGTGAA		GCCG	
					AAGGTGCGTACAATGAAAAGCACAAAAGA		ACTG	
					AGAGGGTAGAAGTAGATCGAATGCCAAGA		AGCC	
					AAGGTCCGAACTGGCGAAGATCAAAAAGC		AGAG	
					GAATCGGACGGAAGGAGTGTGAGTAAAGG		GGCA	
					CAGATACTGGCGAGATCCGTCGAACAAGCC		AAGT	
					GCCACACTCGAAGATGACCCAGTCGGCTTT		CGTA	

					AGCTAAGCGCTAAACCGGCTCCTCTGGGAG		GGCC	
					GAGGTATGTCAGAGGACATTCTCCGTGGGC		GGTA	
					GGATGGGAGGAGTAGGGTAACGACCCGTCA		GGCT	
					TTCTGGATGCCTAAACCACCACAATCTGTCA		CCCG	
					AGGCAAAGTGCCCCAAAAGCACACGCGTGG		GCGG	
					ATCGGTTTGGATGCCGACTGAGCCAGAGGG		GCTC	
					CAAAGTCGAAGGCCGGTAGGCTCCCGGCGG		TCCG	
					GTTGTCCGTCATAGTCAGTGGTGCGCCTACA		TCAT	
					CCCAACTGCTATGACACACAAGGACAACCC		AGTC	
					AAAATAAATAAGCCAAGGCGGCGTTAGCTT		AGTG	
					CGAGCTAACAAGCTCCCCGAGAGGATGGTT		GTGT	
					GCCACAGGGCACCATCCTGGGGAACGACCC		GCCT	
					GATCTTTCGGATGCCCAACCACCGCCAATCT		TCAC	
					GTCAGGCAACGTGCCCCAAAAGCACACGTG		CCAA	
					CGGAGCGGTTGGATGCCGACTGAGCCAGAG		CTGC	
					GGCAAAGTCGTAGGCCGGTAGGCTCCCGGC		TATG	
					GGGTTCTCCGTCGTAGTCAGTGGTGTGCCTA		ACAT	
					CACCTAACTGCTATGACAAGCGTATAGGAG		GCGT	
					GCCCGGAAAAACAAGCCAAGGCGGCGTTA		ACAG	
					GCTGAGAGCTAACAAGCTTCTCGTGGATGG		GAGG	
					GTGCCAGAGGGCACCATCCTGGTGGGTGGA		CCCG	
					TGGGGGGAGCTTGGGAACGTCCCGATCCTT		GAAA	
					CGGATGCCCAGACCACCGCAATCTGTCAAG		AATA	

					GCACCGTGCTCCAAAAGCACACGCGCGGGT		AGCC	
					TGGTTTGGATGCCGACTGAGCCAGAGGGCA		AAGG	
					AAGTCGTAGGCCGGTAGGCTCCCGGCGGGC		CGGC	
					TCTCCGTCATAGTCAGTGGTGTGCCTTCACC		GTTA	
					CAACTGCTATGACATGCGTACAGGAGGCCC		GCAT	
					GGAAAAATAAGCCAAGGCGGCGTTAGCATA		AGGG	
					GGGCTAACAAGCTTCTCGTGGATGGGTGCC		CTAA	
					AGAGGGCACCATCCTGGTGGGTGGATGGGG		CAAG	
					GGAGCTTGGGAACGTCCCGATCGTTCGGAT		CTTCT	
					GCCCAACCACCGCAATCTGCCAGGCAACGT		CGTG	
					GCTTCGGAWGGTCATTGGTTCTAGACTTGT		GATG	
					AATAGACCATTGGCCGGAAGAGCACACGCG		GGTG	
					CGGTTGGTTGGATGCCGACCGAGCCTAGAG		CCAG	
					GGTGCAAACCTGAAGGGCGAGGTCGAAGG		AGGG	
					CCGTGAGGCTCCCGGCGGGAAACTCCGTCA		CACC	
					TAGTTAGTGGTGTGCCTACACCCGACGACT		ATCC	
					ATGACACATAGGAGGAATCCTGATCTGATA		TGGT	
					TGATCATGTATATAGGGAGGGCGAAGGTAA		GGGT	
					ATAGTCAGKGTCAAAGTCCACGTGGCAGCT		GGAT	
					ACTCCCCAGCATAGTAGTGATGCGAGTGGA		GGGG	
					WCCAACTTTGACACTGATGTTCCCTGAGCC		GGAG	
					TGACCCATCTGCACAAATCCAACAGTGTAT		CTTG	
					GATGGCCCACACACTGAGGACGAGTATCAC		GGAA	

					TTGTGATACTCAGAGGTGTCCCCCATGATCA ACCAATATCACAGCTAGCGGACCTACCGTG AGGTAGACCCCCGCCGCTGTAGCAGGCTCG CCTC (SEQ ID NO: 1050)		CGTC CCGA TCGTT CGGA TGCC CAAC CACC GCAA TCTG CCAG GCAA CGTG CTTC GGAW GGTC ATTG GTTCT AGAC TTGT AATA GACC ATTG GCCG	

							GAAG AGCA CACG CGCG GTTG GTTG GATG CCGA CCGA GCCT AGAG GGTG CAAA CCTG AAGG GCGA GGTC GAAG GCCG TGAG GCTC CCGG CGGG	

							AAAC TCCG TCAT AGTT AGTG GTGT GCCT ACAC CCGA CGAC TATG ACAC ATAG GAGG AATC CTGA TCTG ATAT GATC ATGT ATAT AGGG AGGG	

							CGAA GGTA AATA GTCA GKGT CAAA GTCC ACGT GGCA GCTA CTCC CCAG CATA GTAG TGAT GCGA GTGG AWCC AACT TTGA CACT GATG TTCCC	

							TGAG CCTG ACCC ATCT GCAC AAAT CCAA CAGT GTAT GATG GCCC ACAC ACTG AGGA CGAG TATC ACTT GTGA TACT CAGA GGTG TCCC CCAT	

							GATC AACC AATA TCAC AGCT AGCG GACC TACC GTGA GGTA GACC CCCG CCGC TGTA GCAG GCTC GCCT C (SEQ ID NO: 1296)	
NeSL	NeSL-2_CRem	.		Caenorhabditis remanei	CCAACTCTCATCGTATTAACCTACGGTATTC ACTCCTAGTGAGTGTAATAAAGGTTAATTA CGTTTTCTCTTGCMAGAGAAAAAGAAAATT	CCAACT CTCATC GTATTA	TGAA TACC GTCA	MTNVYLKPVND NQTNKTGDNSR NTMSNSQCEMT

					CGAATCCTTTTTGTGTAACTCACAAACTGAC	ACCTAC	GATA	WKPVARTYAQA
					AGAGACCTATCGAATTTCCTTTGTTTCGTAT	GGTATT	AGCC	ASTNPADDKTV
					ATAGGAATAGTCACTCTGGACCACGAAGTG	CACTCC	CCCA	TVLGCKYNLLK
					GACAGTTGTCGGCGGACTTCCAGAGTGGAG	TAGTGA	ACAT	LGNTPQTSKRSP
					AGAAAAGGTGTGAAGAGAGGAGGTCTAGA	GTGTAA	AAAA	PKPSRGGARISS
					AACACTTCGGCTGTCTAGGACCAGTTCCTG	TAAAGG	ATAA	VYTLTDELEITH
					AGTGGAAAGAGGAAGGTCTAGAAACACTTC	TTAATT	AAGT	REEGKITFAIDLP
					GGCTGTCTAGGACCAGTTCGTGAGATCTCTC	ACGTTT	CGGC	NKNNILCPLCRE
					GTGGAGAGTTGAAAACAGTCAGCTGAGGCT	TCTCTT	GTTA	CTQTRGRGSSFT
					ACTGTATTTCTTGATAGCCCCGCCCCCAATC	GCMAG	GCTA	KHMKLHVKEK
					CCCCTCCCCCCCCCCTCGACAGATTTTTCTG	AGAAAA	ACCA	HQLDATFIYKCS
					TTTGACCTCCTGGAATTTGCGAGGAGTGCG	AGAAAA	CTAA	MCNEYEPEKKC
					CGAGAATTTTCGAATTCTTCGCGCGTTTTCT	TTCGAA	ACCG	GTKWIQTHLQK
					CGAAATTTTCCAGAAGATTCGAGCGGAGAA	TCCTTTT	GCTC	VHNYKYDESAI
					TCTTCGAGAAAGTGAGCTGAATTTCGCGCG	TGTGTA	CTCA	VVPVPPNTRQQI
					AATTTTCCGCGATTTTCAAATTATCGATTTT	ACTCAC	TTGG	ANELNNAAPFV
					TGTCGGAAAATTTATTTTCTGGCAAAATTTG	AAACTG	GGGA	DIRKPKAAAVEE
					ATTGAGTTCACGCGGGAGAGAAGGAATTGT	ACAGAG	GAGT	KKTENGALLKF
					TGGAAAAGGGTATTGATTTTTGTGGCGGAG	ACCTAT	ATCA	LTKSNKDDQVK
					GAAACTCCCACTGAATCAATAACTCTCAAA	CGAATT	TTCC	SPSXDIPDAESPE
					GGAGAACTCATCGAACAACCTCGGGTGACC	TCCTTT	GGTG	KETQALTIDPKG
					TGAATCTTGGGCGAAATTTTCGCATTGACAC	GTTTCG	CTCTC	NNSPSKSSIRSSQ
					AAGATAAMACAAATTACTGTKGAAAATAA	TATATA	CGTTT	SSASSVCQEIQEI

					ATCAGAACAAACTGTCAAAAAGAGAGACA	GGAATA	GGGC	ITLSEDEDPKGA
					AAAAGTATTGATTAACAACATCATGACAAA	GTCACT	GGTA	RPKPGINVWSLI
					TGTATATCTTAAGCCTGTGAATGATAACCA	CTGGAC	GGGA	NETGKDAYIDT
					GACTAACAAAACCGGTGATAATTCTAGAAA	CACGAA	GGAG	DIMMAFLKMRV
					TACTATGTCAAATAGTCAATGTGAAATGAC	GTGGAC	TTGG	ENCDSVNIIDPL
					GTGGAAACCTGTAGCCAGAACATATGCTCA	AGTTGT	GTAG	NYQFPARVDLV
					GGCAGCCAGTACTAACCCGGCCGACGACAA	CGGCGG	CGAC	PLIQRNLEDGKK
					AACGGTGACTGTCCTTGGGTGCAAATACAA	ACTTCC	CCGG	RVVFPICADEH
					TCTGCTAAAACTGGGAAATACTCCTCAGAC	AGAGTG	AAGT	WTLLTISNGIAA
					GTCGAAAAGGTCGCCTCCAAAACCATCGAG	GAGAGA	ATGG	FYDPTGSRMSSY
					AGGAGGAGCTCGAATCAGCAGTGTGTATAC	AAAGGT	ATGC	IEELVNELGLIIP
					TCTGACTGATGAGCTGGAGATTACGCACAG	GTGAAG	CCAA	KEQDEQPRQRD
					AGAAGAAGGTAAGATCACATTCGCGATAGA	AGAGGA	CCAC	SYNCGVFVMK
					CCTTCCAAACAAGAATAACATCTTGTGCCC	GGTCTA	CGCA	MAEAFIQDTEW
					GCTGTGTCGGGAGTGCACCCAAACCCGTGG	GAAACA	ATCT	EMEEVEEDVKN
					GAGAGGGTCCAGTTTTACCAAGCATATGAA	CTTCGG	GATC	FRRNLLEELKPN
					ACTCCACGTGAAAGAGAAGCACCAACTTGA	CTGTCT	TGGC	YEIFAEKIKYYN
					TGCCACGTTCATCTACAAGTGTAGTATGTGC	AGGACC	ATTG	SPGKSFAQSRPT
					AACGAGTACGAACCGGAAAAAAAATGCGG	AGTTCC	TGTTT	SRSSQCAVCPTC
					TACGAAGTGGATCCAGACCCACCTTCAAAA	TGAGTG	CGGA	SRSATPMMDVG
					AGTGCACAACTACAAGTATGACGAGTCTGC	GAAAGA	TGGT	NMEVDPVPQQQ
					AATAGTTGTCCCAGTACCACCCAACACAAG	GGAAGG	CTCT	ETPKSREPEQDE
					ACAGCAAATAGCTAATGAGTTGAACAATGC	TCTAGA	GTCT	GWKVVGKARK

					TGCCCCATTCGTTGACATCAGAAAACCGAA	AACACT	CTAG	RGVVTERSPNIS
					AGCTGCTGCTGTTGAGGAGAAGAAGACTGA	TCGGCT	ATCT	PEAKRQFTGPEI
					AAATGGTGCTCTGTTAAAATTCCTGACCAA	GTCTAG	GAAA	KVVSPGKFHPL
					GTCCAATAAGGACGATCAGGTAAAATCCCC	GACCAG	TAGA	VGETEEMEVTC
					ATCGGAWGATATTCCAGATGCGGAAAGCCC	TTCGTG	GCTC	DSPPTKEPTTEP
					TGAAAAAGAAACTCAGGCGCTCACTATCGA	AGATCT	TGGC	KVTPSLPAMKIA
					TCCGAAAGGGAACAACTCACCATCAAAAAG	CTCGTG	CTGA	SPEVTKKQTSKK
					CTCAATAAGATCGAGCCAGTCCTCAGCTTC	GAGAGT	AGAA	KGKYGKKKQXT
					CTCCGTTTGTCAAGAAATCCAGGAAATCAT	TGAAAA	CACA	KKAQPPKGEPT
					CACGTTGAGTGAGGATGAAGACCCAAAAGG	CAGTCA	CGCG	KKAQPKGEPAK
					GGCTCGTCCAAAACCAGGAATCAACGTGTG	GCTGAG	CGGA	LIEQVRTWFDK
					GAGCTTGATAAATGAAACGGGAAAGGATGC	GCTACT	CCGG	QMKSYQEQGSN
					ATACATTGATACAGATATCATGATGGCGTT	GTATTT	TTGG	IQTLTWIADSLT
					CTTGAAGATGAGAGTGGAAAACTGTGACTC	CTTGAT	ATGC	AAIFKANSGNK
					CGTGAACATAATTGATCCACTCAATTACCA	AGCCCC	CGAC	YLVDKITARCPP
					GTTTCCCGCGAGAGTGGACCTAGTCCCACTT	GCCCCC	TCGA	PLLNEGEMATQ
					ATCCAGAGGAATCTGGAAGACGGAAAGAA	AATCCC	TCTG	TSRRTEAVKPK
					AAGAGTCGTGTTTCCGATCTGTGCAGACGA	CCTCCC	GAGG	DRFVKESNEPLR
					ACACTGGACGCTCTTGACCATCTCGAATGG	CCCCCC	GTGC	IQYAKNRAKTF
					AATTGCTGCATTCTATGATCCGACTGGATCG	CTCGAC	AAAC	NVIIGKHSARCEI
					CGAATGAGTAGTTATATTGAAGAGTTGGTG	AGATTT	CTGA	DINVVENHFRQT
					AACGAACTTGGACTGATTATCCCAAAGGAA	TTCTGT	AAGG	LKAQPVTEEAL
					CAGGATGAACAGCCAAGACAAAGAGACAG	TTGACC	GAAA	NTVCSGIKKAK

					CTACAACTGTGGGGTATTTGTGATGAAAAT	TCCTGG	GTTG	VDPSIEGPISSGE
					GGCGGAAGCCTTCATCCAAGATACCGAATG	AATTTG	AAGG	VKAILAKIKDTS
					GGAAATGGAGGAAGTAGAGGAAGACGTGA	CGAGGA	CCGT	PGTDGVKYSDL
					AAAACTTCCGAAGAAATCTCCTTGAAGAAC	GTGCGC	GAGG	KWFDPEGERLA
					TGAAACCCAACTACGAGATATTTGCTGAAA	GAGAAT	CTCCT	LLFDECRQHGKI
					AAATCAAATATTATAACTCTCCGGGAAAAA	TTTCGA	GGCG	PSHWKEAETVL
					GTTTCGCCCAAAGTCGACCCACAAGTCGAA	ATTCTT	GGAA	LPKDCTEEERKK
					GCAGCCAGTGTGCCGTCTGTCCGACGTGCT	CGCGCG	ACTC	PENWRPISLMAT
					CTCGTTCAGCTACACCGATGATGGATGTAG	TTTTCTC	CGTC	VYKLYSSVWNR
					GAAACATGGAAGTGGATCCCGTTCCACAGC	GAAATT	ATAG	RISSVKGVISDC
					AACAAGAGACACCGAAGAGTCGCGAGCCA	TTCCAG	TCAG	QRGFQAIDGCN
					GAACAAGATGAAGGCTGGAAAGTGGTGGG	AAGATT	TGGT	ESIGILRMCIDTA
					AAAGGCTAGAAAGAGGGGAGTGGTAACTG	CGAGCG	GTGC	TVLNRNLSCSW
					AACGATCGCCAAACATCTCTCCAGAAGCCA	GAGAAT	CAAC	LDLTNAFGSVPH
					AGAGACAATTCACTGGTCCAGAGATCAAAG	CTTCGA	ACCC	ELIRRSLAAFGY
					TCGTCTCACCTGGGAAGTTTCACCCACTTGT	GAAAGT	GACG	PESVINIISDMYN
					GGGCGAAACTGAGGAGATGGAGGTGACGT	GAGCTG	ACTA	GSSMRVKTAEQ
					GTGACAGCCCACCAACGAAAGAGCCCACTA	AATTTC	TGAC	KTQNIMIEAGVK
					CGGAACCGAAAGTGACTCCAAGCCTGCCAG	GCGCGA	ATAG	QGDPISPTLFNIC
					CAATGAAAATTGCTAGCCCAGAAGTGACGA	ATTTTC	TTGG	LEGIIRRHQTRK
					AAAAGCAAACGTCAAAGAAGAAGGGAAAG	CGCGAT	AGGA	TGYNCVGNDVR
					TATGGCAAAAAGAAACAGSAGACAAAGAA	TTTCAA	ATCC	CLAFADDLAILT
					AGCTCAGCCGCCGAAAGGGGAGCCAACAA	ATTATC	TGAT	NNQDEMQDVL

					AGAAAGCTCAGCCAAAAGGAGAACCGGCA	GATTTT	CTGA	NQLDKDCRSVA
					AAGCTCATTGAGCAAGTGAGAACTTGGTTT	TGTCGG	TAAT	LIFKPKKCASLTI
					GATAAACAGATGAAATCGTACCAAGAGCAA	AAAATT	AATC	KKGSVDQYARI
					GGTTCTAACATCCAGACACTGACCTGGATT	TATTTT	ATTG	KIHGMPIRTMSD
					GCCGACTCACTCACTGCCGCCATCTTCAAG	CTGGCA	TTCAT	GDTYKYLGVQT
					GCAAACAGTGGAAACAAGTATCTGGTAGAT	AAATTT	ATAA	GNGGRASESESL
					AAGATAACTGCAAGATGCCCACCACCATTG	GATTGA	GGGA	TQIAAELQMVH
					CTGAATGAAGGTGAGATGGCGACGCAGACG	GTTCAC	GGGG	DTDLAPNQKLD
					AGCAGAAGGACAGAAGCGGTGAAACCAAA	GCGGGA	GATG	VLKAFILPRLQH
					AGATCGATTTGTAAAAGAATCTAACGAGCC	GAGAAG	GTAA	MYRNATPKLTE
					GCTCAGAATCCAGTATGCAAAGAACCGAGC	GAATTG	ATAC	LKEFENTVMKS
					AAAGACCTTCAATGTGATAATTGGGAAACA	TTGGAA	CCAG	VKMYHNIPIKGS
					CTCCGCACGATGTGAAATTGATATTAACGT	AAGGGT	GGTC	PLEYVQIPVKNG
					CGTGGAAAACCACTTCAGGCAAACCCTGAA	ATTGAT	CGAA	GLGVMSPRFTC
					AGCACAACCAGTAACAGAAGAAGCATTGA	TTTTGT	ACCA	LITFLASTLFKL
					ATACTGTGTGCAGTGGAATCAAAAAGGCGA	GGCGGA	TCAA	WSDDEYISSIHK
					AAGTTGATCCAAGCATTGAAGGTCCGATCT	GGAAAC	AGCA	KALSRITAKVM
					CGTCAGGAGAAGTGAAAGCGATTCTTGCAA	TCCCAC	GCTA	GLKTQKATLQE
					AGATCAAAGATACCTCTCCCGGAACTGATG	TGAATC	CTGA	QCEYLNTKKAIT
					GAGTGAAGTACAGTGATCTGAAATGGTTCG	AATAAC	CCAG	KGGYSLFSRMN
					ATCCGGAAGGTGAACGTTTGGCGTTGTTGTT	TCTCAA	CATA	EAIRTLSVNLGA
					CGATGAATGTCGACAGCACGGGAAGATTCC	AGGAGA	GTAG	PLKSMQFIPENG
					GAGCCACTGGAAAGAAGCAGAAACTGTTCT	ACTCAT	TGAT	EIALEVQASENS

					GCTACCAAAAGATTGCACTGAAGAGGAAAG	CGAACA	GAAC	QIKVFSKADSM
					AAAGAAGCCAGAGAATTGGAGACCCATCTC	ACCTCG	ACAT	KLVTKLKDLVK
					TCTAATGGCTACTGTATACAAACTCTACTCC	GGTGAC	AGAC	SAMLKNFLENK
					TCAGTCTGGAACAGGAGAATCTCCTCAGTK	CTGAAT	CCTG	KVKSKVVQVLQ
					AAAGGAGTCATCAGTGATTGCCAAAGAGGC	CTTGGG	GGGT	HHPQSNKFVND
					TTCCAGGCGATCGATGGATGCAATGAGTCA	CGAAAT	TCCCT	GKNXSISSQKFV
					ATCGGAATTCTGCGGATGTGCATAGACACA	TTTCGC	GAAC	HPARLSQLVCN
					GCCACAGTTCTCAACCGAAACCTGTCGTGTT	ATTGAC	TCGA	GNSYSKDLPKN
					CATGGTTAGACTTGACGAACGCTTTTGGAA	ACAAGA	CCCA	CRWCGYECESQ
					GCGTGCCCCACGAGTTGATCAGAAGATCAC	TAAMAC	TCTG	AHILQHCTYSLS
					TAGCCGCATTCGGGTATCCTGAATCAGTCAT	AAATTA	CACA	SGITQRHDRVLN
					CAATATAATCAGTGACATGTATAATGGATC	CTGTKG	AACC	RILXEVIKGRKN
					GTCAATGAGAGTCAAGACAGCGGAGCAGA	AAAATA	CACT	NDYYDIMVDTE
					AAACTCAGAACATCATGATTGAAGCTGGAG	AATCAG	TTGT	PGPTRERPDIIMI
					TTAAGCAAGGTGATCCCATCTCGCCAACTCT	AACAAA	ACAA	QKDGPEVLLAD
					ATTCAACATCTGTCTTGAAGGCATAATCCG	CTGTCA	ATGA	VTVPYENGVVA
					AAGGCATCAGACGAGGAAGACAGGTTACA	AAAAGA	ACCA	IEAAWDWKME
					ACTGCGTTGGAAACGACGTACGTTGCCTGG	GAGACA	AACT	KYSHFIDYFARL
					CATTTGCTGACGATCTTGCTATCCTTACCAA	AAAAGT	GATG	GKRAVILPLVVG
					CAACCAGGATGAGATGCAAGATGTGCTCAA	ATTGAT	AAGA	SLGTYWPDTSN
					TCAGCTGGACAAGGACTGTCGTAGTGTTGC	TAACAA	GTTT	SLRMLGLSDGQI
					CCTGATATTTAAGCCAAAGAAGTGTGCTTC	CATC	AATG	RNLIPDISMIALE
					ACTGACGATCAAAAAAGGAAGTGTTGATCA		ATTTC	SSKQIYWRHIFG

					GTATGCAAGAATCAAGATTCATGGAATGCC CATTCGGACTATGTCGGATGGGGATACCTA CAAGTATCTCGGAGTCCAAACCGGAAACGG TGGTAGAGCCTCGGAATCAGAATCCCTGAC TCAGATTGCCGCGGAACTCCAAATGGTCCA TGACACAGACCTGGCGCCGAACCAGAAACT TGATGTGCTGAAGGCATTCATCCTGCCGAG ACTGCAACATATGTACAGAAACGCCACTCC AAAGCTGACGGAGTTAAAGGAGTTTGAGAA CACAGTCATGAAAAGTGTGAAGATGTATCA CAACATCCCGATTAAAGGATCACCACTCGA ATATGTCCAAATTCCAGTAAAGAATGGAGG ACTCGGAGTTATGTCTCCCCGATTCACGTGT CTCATAACGTTCCTGGCGTCCACACTGTTCA AACTGTGGTCAGACGACGAATACATCTCGT CCATCCACAAAAAGGCGTTGAGTAGAATCA CGGCAAAGGTGATGGGACTGAAGACCCAA AAAGCCACGCTCCAAGAGCAGTGCGAGTAC CTGAACACCAAGAAAGCAATCACGAAAGG AGGTTACAGCCTCTTCTCGCGAATGAACGA AGCTATTCGAACGCTCAGTGTCAACCTTGG AGCACCGCTCAAATCAATGCAATTCATTCC GGAAAATGGCGAAATTGCTTTAGAAGTGCA	(SEQ ID NO: 1174)	TTAC ATCA CAGC TAGC GGAC CTAC CGTG AGGT AGAC TCCC GCCG CTGT AGCA GGCT CGCC ATTG (SEQ ID NO: 1297)	DSYRIVSDLYCR KDQQEIRFGDEP MENVQVSDRFQ PFKTREREKKSE EEKKRRSKSKK GKTWRGSKKQT DSRQSGKSNQN QGFQRSVGQGV SR (SEQ ID NO: 1419)

					AGCATCAGAAAACTCACAGATCAAAGTATT CTCGAAAGCTGACAGTATGAAACTGGTGAC AAAGCTGAAAGATCTGGTGAAATCGGCGAT GCTCAAGAACTTCTTGGAAAACAAGAAGGT CAAAAGCAAGGTTGTGCAGGTGCTTCAACA CCACCCACAATCAAACAAATTCGTCAATGA TGGAAAGAACWKCAGCATWTCCTCCCAAA AGTTCGTACACCCAGCACGGCTGAGCCAGC TGGTCTGCAACGGGAACAGCTACAGTAAAG ACCTTCCGAAAAACTGCAGATGGTGCGGCT ACGAATGCGAGTCTCAGGCTCACATCCTCC AGCATTGCACATACAGCCTTTCATCTGGAAT CACCCAGAGGCATGACCGTGTCCTGAACAG GATCTTGCASGAGGTGATAAAAGGCAGAAA AAACAACGACTACTATGACATAATGGTGGA TACGGAGCCCGGACCAACCAGAGAGCGTCC AGATATCATCATGATACAGAAAGATGGTCC GGAAGTCCTACTGGCGGATGTTACGGTACC ATACGAGAATGGAGTTGTTGCGATCGAAGC CGCGTGGGATTGGAAGATGGAGAAGTACAG TCACTTTATTGATTACTTCGCAAGACTGGGA AAGAGAGCAGTAATCCTTCCACTAGTGGTT GGAAGTCTTGGGACCTACTGGCCAGACACC			

					TCAAACTCTCTGAGAATGCTGGGTCTCTCTG ATGGGCAAATCAGAAACCTGATCCCAGACA TCTCCATGATTGCTCTAGAGTCTTCCAAACA AATCTACTGGAGGCATATCTTCGGAGATAG CTACAGAATTGTGAGTGATCTATACTGCAG AAAAGACCAGCAGGAGATCAGATTCGGAG ATGAACCCATGGAAAATGTTCAAGTCTCAG ATCGATTCCAGCCTTTTAAAACAAGAGAGC GTGAGAAGAAATCCGAGGAAGAGAAAAAG AGAAGATCAAAGTCCAAAAAAGGCAAAAC TTGGCGAGGATCCAAAAAACAAACTGATTC CCGGCAATCCGGCAAAAGCAATCAGAATCA GGGCTTCCAAAGAAGCGTTGGACAAGGCGT ATCACGGTGAATACCGTCAGATAAGCCCCC AACATAAAAATAAAAGTCGGCGTTAGCTAA CCACTAAACCGGCTCCTCATTGGGGGAGAG TATCATTCCGGTGCTCTCCGTTTGGGCGGTA GGGAGGAGTTGGGTAGCGACCCGGAAGTAT GGATGCCCAACCACCGCAATCTGATCTGGC ATTGTGTTTCGGATGGTCTCTGTCTCTAGAT CTGAAATAGAGCTCTGGCCTGAAGAACACA CGCGCGGACCGGTTGGATGCCGACTCGATC TGGAGGGTGCAAACCTGAAAGGGAAAGTTG			

					AAGGCCGTGAGGCTCCTGGCGGGAAACTCC GTCATAGTCAGTGGTGTGCCAACACCCGAC GACTATGACATAGTTGGAGGAATCCTGATC TGATAATAATCATTGTTCATATAAGGGAGG GGGATGGTAAATACCCAGGGTCCGAAACCA TCAAAGCAGCTACTGACCAGCATAGTAGTG ATGAACACATAGACCCTGGGGTTCCCTGAA CTCGACCCATCTGCACAAACCCACTTTGTAC AAATGAACCAAACTGATGAAGAGTTTAATG ATTTCTTACATCACAGCTAGCGGACCTACCG TGAGGTAGACTCCCGCCGCTGTAGCAGGCT CGCCATTG (SEQ ID NO: 1051)			
NeSL	NeSL-4_CRem	chrUn		Caenorhabditis remanei	CCCTTTTCTATCGTATTAACTACGATAACCG CTCATTTGAGTGTAAAAAAGGTTCCCCCCTC CTCGCCTGCCTTACCCACGCATCTCTGCCTC TGGGAAGGCGGAGGGTCAACTTGCGGGTCT GTGGATTTCCTTTCCTATCCACCGCCCATAT TCTCTGTCGAAAGCCTACCTAGATCAGCCG GGAGTTTTTCCTATCCCATTCAGGCGATCGC TCAAGGCTGTTTTATCGACACTCCTTCTTGA CAAGTATTTATTTCTTGACAAATTCTATTTT TCCTTTTATCGATTTTCTCTTATTTATCGATT CTTGTGAAAATATGACCAAGACCGAATGGT	CCCTTT TCTATC GTATTA ACTACG ATAACC GCTCAT TTGAGT GTAAAA AAGGTT CCCCCC TCCTCG	TAAC ATGC CTTG GAAG GCAC CACG CCAA AAGT CCTG GCAA CTGA	MTKTEWSWRH RSRSRSVGIVVK IDTSDYANVRV HVAADLSNEDG HTSHNNGIILPIP MKPSVDRFCQIQ YPPRGYYVPHP QSQKGHDAKPS RHWNEEAQPPY YHNNNHGRRGR SAKPSGRRPPRK

					CCTGGCGTCATCGATCTCGCTCCCGCTCTGT	CCTGCC	TTTG	PILQEESLAAHP
					TGGAATCGTTGTGAAAATCGATACAAGCGA	TTACCC	AATA	QIPGDTASAVPL
					CTATGCTAACGTCCGAGTGCATGTCGCGGC	ACGCAT	ATGT	YSDVVNNENKS
					GGACCTTTCCAATGAGGATGGCCACACGAG	CTCTGC	ATAA	QGKPPQGSHRR
					CCACAACAACGGCATCATTCTCCCCATCCC	CTCTGG	AAGT	SGRPGTKPSVPV
					AATGAAGCCCAGCGTCGATCGATTCTGTCA	GAAGGC	AACT	GEAEQETNSRPI
					AATTCAATACCCTCCAAGAGGGTACTATGT	GGAGGG	GGAA	APEPIVKFKHDK
					TCCGCATCCTCAAAGTCAGAAAGGCCATGA	TCAACT	CCAA	HGWTTVQGSHS
					TGCAAAGCCCTCGCGTCATTGGAATGAAGA	TGCGGG	ATGC	SGRPVPKPSVPV
					GGCACAACCTCCCTACTACCACAACAACAA	TCTGTG	CCGA	VSEANRFQLLQE
					TCATGGGAGAAGGGGGCGTTCGGCAAAACC	GATTTC	TAGG	GDFPPLTTSESS
					AAGTGGACGCCGACCCCCACGAAAGCCCAT	CTTTCC	TAGG	QEEIKVPNYQRI
					ACTTCAGGAAGAGTCCCTGGCAGCGCACCC	TATCCA	GCGG	VSPIPLPSEEDSK
					CCAAATACCCGGGGATACTGCGTCAGCGGT	CCGCCC	GAGA	LPTKSNYRAPK
					CCCACTGTACTCCGACGTCGTCAACAATGA	ATATTC	AAAT	GRKSRNYKKPQ
					AAACAAGAGTCAGGGGAAACCACCGCAAG	TCTGTC	GACC	QQNPKKYQQRL
					GGTCCCACAGAAGAAGTGGAAGACCAGGA	GAAAGC	TAGA	PYQPKVNNAPT
					ACAAAGCCCTCTGTTCCGGTTGGTGAGGCA	CTACCT	AAAC	DRMAPEQLKGG
					GAGCAAGAAACGAATTCCCGTCCAATTGCT	AGATCA	ACAA	GGKTAHNDIEE
					CCAGAACCCATCGTGAAATTCAAACACGAT	GCCGGG	AGTC	MEIEEDTDEKIIQ
					AAACACGGGTGGACTACTGTCCAAGGGTCC	AGTTTT	CCAA	VKRIKIVNKLTP
					CACAGTAGTGGAAGGCCGGTACCAAAGCCC	TCCTAT	GCCC	HHFVCMMTYPT
					TCGGTACCGGTGGTTTCAGAGGCAAATCGG	CCCATT	CCGG	DNIYRCFVKGCT

					TTCCAGTTACTCCAGGAAGGGGATTTTCCAC	CAGGCG	ATTC	ATSQGGWGAED
					CCCTTACAACATCCGAATCTTCGCAAGAAG	ATCGCT	GAAA	LKYLTVHIRQEH
					AGATTAAAGTACCGAACTACCAACGAATAG	CAAGGC	GACC	KIKVEWTYECGI
					TGTCACCGATTCCTCTCCCCTCTGAAGAGGA	TGTTTT	TATA	CGDLSGGAGKH
					TAGTAAGTTGCCGACTAAATCAAATTACAG	ATCGAC	GGAA	ISKWIKPHMRK
					AGCGCCCAAGGGACGAAAGAGTCGCAACT	ACTCCT	GTCA	KHNRDAPTNFK
					ACAAGAAGCCACAACAACAAAATCCGAAG	TCTTGA	GTGA	MGSRSSGKPKIT
					AAATATCAGCAGAGGTTACCCTATCAACCC	CAAGTA	ATAG	ELLEESAPSCSN
					AAGGTCAACAATGCTCCGACGGATCGCATG	TTTATTT	AGAG	PRRKTLNQKKT
					GCCCCAGAACAACTCAAAGGAGGAGGAGG	CTTGAC	AAAT	AIITQVTPEKLK
					AAAAACCGCCCACAATGACATTGAAGAGAT	AAATTC	ATCA	TGYQTRSVTKA
					GGAAATTGAGGAAGACACTGACGAGAAGA	TATTTTT	AACA	LSVLKESRQKEL
					TTATCCAAGTGAAACGAATCAAAATCGTCA	CCTTTT	AATC	EVLREEEKANA
					ATAAGCTAACTCCGCATCACTTTGTTTGCAT	ATCGAT	TCAC	KQKSKLHPFFTK
					GATGACGTATCCAACCGACAACATCTATAG	TTTCTCT	CCAT	APHIDGVKPTVR
					ATGCTTCGTCAAAGGTTGCACAGCAACATC	TATTTA	TCAC	RELSKMITPGGE
					ACAAGGTGGTTGGGGAGCAGAGGACCTTAA	TCGATT	AAGG	HKGTKIPMVHT
					GTACCTGACTGTCCATATCAGACAAGAACA	CTTGTG	ACTT	KRGLIQKINRKA
					CAAAATTAAGGTCGAATGGACGTACGAATG	AAAAT	ACTG	KKAKPMHLDES
					CGGGATATGCGGTGACCTATCGGGAGGTGC	(SEQ ID	GTCG	TIIEASQLDVITI
					TGGCAAACATATCAGTAAATGGATCAAACC	NO: 1175)	AGTA	DDDDEDDNMTP
					CCATATGAGGAAGAAACACAATAGAGATGC		GAAA	MRRRFNTWCLD
					CCCAACCAATTTCAAGATGGGTTCAAGAAG		ACAA	HETTQEAWLTD

					TTCAGGTAAACCCAAGATTACTGAACTACT		GCCA	DVINWYLKDLC
					GGAGGAGAGCGCCCCGTCTTGCTCGAATCC		AAAC	FGNEQYMLVDP
					AAGAAGGAAAACCCTCAACCAGAAGAAGA		ATCA	LVWLIYKMGG
					CTGCTATAATCACGCAAGTCACTCCGGAGA		AGCA	MAGVEQRFKSK
					AATTGAAAACGGGCTATCAAACGAGAAGTG		CGAC	KTCLFPICEADH
					TCACGAAGGCTCTCAGCGTCCTGAAAGAGT		GCAA	WILLVFDETNLC
					CACGACAAAAAGAGCTGGAAGTGTTGAGA		AAAG	YANSLGSQPNG
					GAAGAAGAAAAGGCTAACGCTAAACAAAA		GGGT	QVKNFIQQLNR
					GTCTAAACTTCATCCTTTCTTCACCAAAGCC		AACT	KLCSFEKEVPLQ
					CCTCATATAGATGGTGTGAAACCAACAGTA		TTGG	KDSVNCGVHVC
					CGGAGAGAACTATCAAAAATGATTACTCCC		GCAA	LIAKSIVNGQFW
					GGAGGAGAACATAAGGGAACAAAGATACC		CTAA	YDDSDVRTFRT
					AATGGTCCACACCAAGCGCGGTCTCATCCA		TTAA	NAKAALKAQGY
					AAAGATAAACAGAAAAGCTAAAAAGGCTA		CGGA	ELFSEAPKQIEN
					AACCAATGCATCTTGACGAAAGTACCATCA		TACC	PDSSHREDIKEN
					TAGAAGCGTCACAGCTCGACGTCATCACTA		TCCG	SMEMCSESLMI
					TTGACGACGACGACGAAGACGACAACATGA		TGTA	VATPQRSEAPM
					CACCAATGCGAAGAAGATTCAACACTTGGT		TCAG	ELVDTEPSDLES
					GTCTTGACCACGAGACGACTCAAGAAGCAT		GCAA	PKSDRVVYEDCI
					GGTTAACTGACGACGTAATCAATTGGTACT		AGCC	TALSDVSEPRMT
					TGAAAGACCTATGCTTTGGTAACGAACAAT		GCCA	PEKSETPEVPVV
					ACATGCTCGTAGACCCACTAGTATGGCTGA		CCAA	EERDLDWPKLE
					TATACAAGATGGGAGGAATGGCAGGCGTCG		CAGC	SPKSDRVVYED

					AACAAAGGTTCAAAAGCAAGAAGACGTGC		AAAT	CITDLSDVSEQR
					CTATTCCCAATCTGCGAAGCTGACCACTGG		TACT	MTPEKCETPEAP
					ATTCTTCTTGTATTCGATGAGACCAACTTGT		GCCC	LVVECVELERLP
					GCTACGCGAATAGTCTTGGATCCCAACCAA		GATA	KDLPVTDRSTV
					ACGGACAAGTTAAGAACTTCATTCAACAAC		GGTA	VAIPEAVKLEEK
					TCAACCGAAAGCTCTGCAGCTTTGAGAAAG		GGGC	SEVVIPRLMELS
					AAGTTCCACTTCAGAAAGATAGTGTAAACT		GTGA	YTVPPEPSPVVE
					GCGGAGTACATGTCTGCCTGATAGCAAAGT		GAAA	YTQPYTHTHTK
					CAATAGTCAATGGACAATTTTGGTACGATG		ATGA	PKVKATCQMGK
					ATTCAGACGTTCGAACGTTTAGAACCAACG		CCTA	KRKVPTGKPDE
					CCAAGGCGGCTCTGAAAGCCCAGGGCTACG		CAAC	LIQIVRQWFEKE
					AGCTCTTCTCGGAAGCACCAAAACAAATCG		CTCC	FNDYVTEGRNF
					AAAACCCAGACTCCAGCCACAGAGAAGAC		AAGA	QRLEWLTNLLT
					ATCAAGGAGAACAGTATGGAAATGTGTTCG		CCCG	AAIQKASAGDE
					GAATCTTTGATGATCGTTGCGACTCCACAG		AGCC	ETIEKIRKRCPPP
					AGGAGTGAAGCACCTATGGAACTAGTCGAC		CACG	EVRENEMSTQT
					ACTGAGCCTAGTGATCTGGAATCGCCAAAG		GAAT	SQRQKPTTTNQ
					TCAGACAGAGTAGTCTACGAAGACTGCATC		CGAA	KKRSRNTTQSD
					ACAGCTCTATCTGATGTTTCGGAGCCAAGA		AGAC	TQANTYWRNRA
					ATGACTCCAGAAAAGAGCGAAACTCCAGAG		CTAT	KTYNQIIGQDFK
					GTGCCAGTGGTGGAAGAAAGAGATCTGGAT		AGGA	QCDIPIAILEEFY
					TGGCCAAAACTGGAATCGCCAAAGTCAGAC		AGTC	KKTTSVTNVPQ
					AGAGTAGTCTATGAAGACTGCATCACAGAT		AGTG	ETLVKVTSRLPR

					CTGTCTGATGTTTCGGAGCAAAGAATGACT		AATT	LDIGKWIEDPFT
					CCAGAAAAGTGCGAAACCCCAGAAGCGCC		GATG	EQEVFGALKKT
					ATTGGTTGTAGAATGTGTTGAGTTGGAAAG		GAAA	KDTAPGTDGLR
					GCTACCTAAGGATCTGCCAGTCACAGACAG		TACA	YYHLQWFDPDC
					GTCAACTGTCGTGGCAATCCCTGAAGCAGT		AAAC	KMLSSIYNECQH
					AAAACTGGAGGAAAAGTCAGAAGTGGTAA		CAAA	HLKIPAQWKEA
					TTCCACGGCTCATGGAGTTATCATACACCGT		TTTCT	ETILLFKSGDES
					CCCTCCAGAACCCTCTCCAGTGGTTGAATAC		TCCA	KPDNWRPISLMP
					ACCCAACCATACACTCACACTCACACTAAA		TTCA	TIYKLYSSLWNR
					CCAAAGGTCAAAGCTACATGCCAGATGGGA		CAAG	RIRTVKGIMSKC
					AAGAAAAGGAAGGTACCAACTGGGAAACC		GACT	QRGFQEREGCN
					AGACGAACTGATTCAGATTGTGAGACAATG		TACT	ESIGILRSAIDVA
					GTTTGAGAAAGAATTCAACGATTATGTTAC		GGTC	KGKRSHLSVAW
					GGAAGGACGAAACTTTCAACGACTGGAATG		GAGT	LDLTNAFGSVPH
					GCTTACGAACTTACTCACCGCCGCAATACA		AGAG	ELIESTLSAYGFP
					GAAGGCATCAGCTGGTGATGAGGAAACAAT		CACA	EMVVHIVKDMY
					CGAAAAGATTCGAAAGAGATGCCCACCTCC		AGCC	KDASIRVKNRTE
					AGAAGTTAGAGAAAACGAAATGTCCACTCA		AAAA	KSEQIMIKSGVK
					GACATCTCAACGTCAAAAGCCTACCACAAC		TATC	QGDPISPTLFNM
					GAATCAGAAGAAACGCTCTAGAAACACTAC		AAGT	CLETVIRRHLKE
					TCAATCGGATACACAAGCCAACACATACTG		ATGA	SSGHKCIDTRIK
					GCGAAATCGAGCCAAGACATATAATCAAAT		CGCA	LLAFADDMAVL
					CATAGGTCAAGATTTCAAACAGTGTGACAT		AAAA	AESKEQLQKELT

					ACCGATCGCGATACTAGAAGAATTCTATAA		TGGG	EMDEDCTPLNLI
					AAAGACTACCTCAGTGACCAATGTCCCTCA		TAAC	FKPAKCASLIIEF
					GGAAACCCTTGTGAAAGTCACCTCAAGACT		CTTG	GKVRTHEQIML
					ACCAAGGTTAGACATTGGAAAGTGGATCGA		GGCA	KREPIRNLNDDG
					GGATCCGTTCACGGAACAAGAGGTATTTGG		TCCA	TYKYLGVHTGA
					TGCCCTCAAAAAGACAAAAGACACTGCGCC		ATCA	DARTSEEELIISV
					AGGAACAGATGGGCTCAGATACTATCACCT		ACGG	TKEVDLVNRSA
					CCAATGGTTTGATCCCGACTGTAAAATGTTG		ATAC	LTPPQKLDCLKT
					AGTAGCATTTACAATGAATGCCAGCACCAT		CTCT	FTLPKMTYMYA
					CTGAAAATTCCTGCCCAATGGAAAGAAGCT		GCGT	NAIPKLTELSAF
					GAAACAATTCTCCTCTTCAAAAGTGGCGAC		ATCA	ANMVMRGVKII
					GAATCCAAACCAGACAACTGGCGGCCTATA		GGCA	HYIPVRGSPLEYI
					AGTCTCATGCCCACCATCTACAAGCTATACT		AAGT	QIPTGKGGLGVP
					CAAGTCTCTGGAATAGGAGAATACGGACGG		CGCC	CPRITALITFLVS
					TGAAGGGGATTATGAGCAAGTGCCAACGAG		ACCA	TMKKLWSDDEY
					GATTCCAAGAGAGAGAAGGTTGCAATGAGA		AACT	IRKLYNSYLKKV
					GTATCGGAATACTGCGGAGTGCTATTGATG		GTAC	VEAETGIVEVST
					TGGCTAAAGGGAAAAGATCCCACCTGTCCG		TACT	KDLAEYLSNKV
					TTGCATGGCTGGACCTCACCAATGCCTTCGG		CCGA	PSRKHEFGYNC
					TTCAGTACCTCACGAGCTGATTGAAAGCAC		AAAA	YSRIREVCNGLA
					GTTAAGTGCATACGGCTTTCCGGAGATGGT		ACCA	LNQAAPLYKLE
					TGTACACATTGTCAAGGACATGTATAAAGA		AGAA	FIEQDNELAVVV
					CGCTTCCATAAGAGTCAAGAATAGAACGGA		ACAT	QPTEESKERIFT

					GAAAAGTGAGCAGATTATGATAAAATCTGG		GATT	KDHVKKLQSLL
					GGTAAAACAAGGCGACCCTATCTCACCAAC		TTCCC	KASVNDALLHR
					ACTATTCAACATGTGCCTCGAAACGGTGAT		ACTC	FLTTKPVKSEVV
					TAGACGACATCTGAAAGAATCATCAGGTCA		CGTT	QVLQQHPQSNS
					CAAATGCATTGACACCAGAATCAAGCTTCT		AAAG	FVRMGGKVSIS
					TGCATTTGCAGATGATATGGCCGTTCTAGCA		CATC	VHVWIHRSRLN
					GAATCAAAAGAGCAGCTACAAAAGGAGCTT		TCAA	QLTCNYNIFDPK
					ACAGAAATGGATGAAGACTGTACACCTCTC		CCAA	QPKNCRRCGYK
					AACCTAATTTTCAAGCCGGCGAAGTGTGCA		GCTA	NETQWHILQDC
					AGTCTCATCATAGAGTTCGGGAAAGTGAGG		AAGC	TYGWAKLIRER
					ACCCATGAGCAGATCATGTTGAAGCGAGAG		GGTA	HDAVHHKVVT
					CCGATCCGAAACCTCAATGATGACGGAACA		AGGT	MICAGAKKNW
					TACAAGTATCTGGGAGTGCATACGGGAGCA		TATC	GRKIDQELPGFT
					GATGCAAGGACATCAGAAGAGGAGCTGATC		ATGT	SLRPDICLTSPD
					ATTTCTGTAACAAAAGAGGTAGACCTTGTC		CAAA	GKEVIFADVCVP
					AATCGCTCGGCGCTTACGCCACCCCAGAAA		AGGT	YSRTRNIEFAWK
					CTGGACTGTCTTAAGACGTTCACACTCCCAA		GTAG	EKIRKYTEGYSH
					AGATGACCTACATGTATGCCAACGCCATAC		CTAC	LVAQGIKVTVLP
					CAAAACTTACCGAACTTTCAGCGTTCGCTA		AGCA	IAIGSLGTWWTP
					ACATGGTCATGCGAGGAGTCAAGATAATCC		ACCT	TNESLYQLGISK
					ACTATATCCCAGTTAGAGGATCTCCTCTTGA		AAAG	SDIRSAIPLLCST
					ATATATTCAAATTCCGACCGGCAAAGGAGG		CCCG	VMEYSKNAYW
					ACTTGGAGTTCCATGCCCTAGAATCACGGC		AAAG	NHIYGNSYTSVP

					ATTGATTACCTTCCTTGTCTCAACCATGAAG		GTAG	LRYGHQKPDGD
					AAACTGTGGTCTGATGATGAATACATTCGT		GGCC	DWKKELSCEPV
					AAGCTCTACAACTCTTATCTGAAGAAGGTT		GTAT	LALQQ (SEQ ID
					GTGGAGGCGGAAACGGGAATAGTGGAGGT		AAAA	NO: 1420)
					CTCCACAAAGGATCTAGCAGAGTACCTCAG		AGAC	
					CAACAAGGTACCATCCAGAAAGCACGAATT		CTAC	
					CGGGTATAACTGCTACTCGAGGATTCGCGA		ACCC	
					AGTTTGTAATGGGCTAGCTCTCAACCAAGC		TCCA	
					TGCCCCTCTCTACAAACTTGAATTCATCGAA		AGAC	
					CAAGACAATGAGTTAGCAGTTGTTGTCCAG		CTAA	
					CCGACTGAGGAGAGCAAGGAAAGGATTTTC		ACCC	
					ACTAAAGATCATGTGAAAAAGCTCCAGTCG		ACGA	
					CTACTGAAAGCCAGCGTGAATGACGCACTG		ACTC	
					CTACACAGATTCTTGACAACAAAACCCGTC		GAAC	
					AAAAGTGAAGTGGTACAAGTTCTCCAGCAG		GACC	
					CACCCTCAAAGCAACAGCTTCGTCCGAATG		TACA	
					GGAGGTAAAGTAAGTATATCGGTACATGTA		GGAA	
					TGGATCCACAGGTCACGGTTAAACCAACTA		GTCC	
					ACGTGCAATTATAACATCTTTGATCCAAAG		GTGA	
					CAACCGAAAAACTGCCGGAGGTGTGGTTAT		ATGG	
					AAGAACGAGACTCAATGGCACATCCTGCAA		AGAG	
					GACTGCACATATGGCTGGGCTAAACTTATA		AAAT	
					CGAGAAAGACACGATGCCGTACATCACAAG		ATCT	

					GTAGTCACAATGATTTGCGCTGGGGCAAAG		CACC	
					AAGAACTGGGGCCGGAAAATCGACCAAGA		AAAT	
					ACTGCCCGGTTTCACTTCACTCCGTCCAGAC		CTCTT	
					ATTTGTCTGACGAGTCCGGATGGCAAAGAG		CCAT	
					GTTATCTTTGCGGATGTTTGTGTCCCTTACT		TCAC	
					CAAGGACAAGGAACATCGAATTCGCGTGGA		AAAG	
					AAGAGAAAATCCGAAAGTATACAGAAGGA		GCTA	
					TACAGTCATCTTGTTGCACAAGGAATCAAA		ACTG	
					GTGACAGTCCTTCCGATAGCCATAGGATCA		GTCA	
					CTCGGAACTTGGTGGACGCCAACCAACGAA		AGTA	
					AGTCTCTATCAACTGGGTATCAGCAAGAGC		GAGC	
					GATATTCGCAGTGCCATTCCATTACTATGCT		ACAA	
					CTACTGTGATGGAGTATAGTAAGAACGCCT		GCTA	
					ACTGGAATCACATATACGGAAACTCATATA		AGCC	
					CCTCGGTCCCACTGAGATACGGACACCAGA		TCCA	
					AGCCCGATGGAGACGATTGGAAGAAAGAA		AGCA	
					CTGAGTTGCGAACCAGTTCTAGCTCTCCAAC		CGAA	
					AATAACATGCCTTGGAAGGCACCACGCCAA		GTGA	
					AAGTCCTGGCAACTGATTTGAATAATGTAT		TATG	
					AAAAGTAACTGGAACCAAATGCCCGATAGG		GGTA	
					TAGGGCGGGAGAAAATGACCTAGAAAACA		ATTT	
					CAAAGTCCCAAGCCCCCGGATTCGAAAGAC		AGGC	
					CTATAGGAAGTCAGTGAATAGAGAGAAATA		AACC	

					TCAAACAAATCTCACCCATTCACAAGGACT		AATC	
					TACTGGTCGAGTAGAAAACAAGCCAAAACA		AACG	
					TCAAGCACGACGCAAAAAGGGGTAACTTTG		GATA	
					GGCAACTAATTAACGGATACCTCCGTGTAT		CCTC	
					CAGGCAAAGCCGCCACCAACAGCAAATTAC		CGTG	
					TGCCCGATAGGTAGGGCGTGAGAAAATGAC		TATC	
					CTACAACCTCCAAGACCCGAGCCCACGGAA		AGGC	
					TCGAAAGACCTATAGGAAGTCAGTGAATTG		AAAG	
					ATGGAAATACAAAACCAAATTTCTTCCATT		TCGC	
					CACAAGGACTTACTGGTCGAGTAGAGCACA		CACA	
					AGCCAAAATATCAAGTATGACGCAAAAATG		AACA	
					GGTAACCTTGGGCATCCAATCAACGGATAC		CTGT	
					CTCTGCGTATCAGGCAAAGTCGCCACCAAA		ACTA	
					CTGTACTACTCCGAAAAAACCAAGAAACAT		CTCC	
					GATTTTCCCACTCCGTTAAAGCATCTCAACC		GTTA	
					AAGCTAAAGCGGTAAGGTTATCATGTCAAA		CTCC	
					AGGTGTAGCTACAGCAACCTAAAGCCCGAA		CAAA	
					AGGTAGGGCCGTATAAAAAGACCTACACCC		CACA	
					TCCAAGACCTAAACCCACGAACTCGAACGA		TGGA	
					CCTACAGGAAGTCCGTGAATGGAGAGAAAT		TCTCC	
					ATCTCACCAAATCTCTTCCATTCACAAAGGC		TTCTC	
					TAACTGGTCAAGTAGAGCACAAGCTAAGCC		TCAC	
					TCCAAGCACGAAGTGATATGGGTAATTTAG		CAAA	

					GCAACCAATCAACGGATACCTCCGTGTATC AGGCAAAGTCGCCACAAACACTGTACTACT CCGTTACTCCCAAACACATGGATCTCCTTCT CTCACCAAAAAGCTTTATAACCAAGCTAAC GGTGGAAAGGACATCATGTCACGAGGAGTA GCTACAGTAACCTCTCTCTTGAGACTGCAA AGTCGAGGATGGATTGGGAAGGCCGCGAG GCAAAAGGCGGGTAACTCGGCCAGACGCTA GTGATCTTCGGATCCGACAGCCCTGGCCTTA GAGGAACCCTGGGATAAGGAGCACGACGG GAAGGATGTTCCGCAAGGATTTCCCTTCCC ATTAGTCAGGGCTGGCAGTTGGTAATATAG CCTTTCTACACACCACCGTCTTGCACCCACT AAACCAGTGGGATATGCGGGTGGACTCAAT GTAGAAAGGTGTTCCCACTGCCTGACTCGC CAACTTTATATGTCTTGTCAACATAATGGCC CCTCACTATAAACTCCCTAGCAACTGGTGGT CCGGCGAAGCCGGTTCTTGCCACTATTGCG CCCCAGGCTCGCC (SEQ ID NO: 1052)		AAGC TTTAT AACC AAGC TAAC GGTG GAAA GGAC ATCA TGTC ACGA GGAG TAGC TACA GTAA CCTCT CTCTT GAGA CTGC AAAG TCGA GGAT GGAT	

							TGGG AAGG CCGC GAGG CAAA AGGC GGGT AACT CGGC CAGA CGCT AGTG ATCTT CGGA TCCG ACAG CCCT GGCC TTAG AGGA ACCC TGGG ATAA	

							GGAG CACG ACGG GAAG GATG TTCC GCAA GGAT TTCCC TTCCC ATTA GTCA GGGC TGGC AGTT GGTA ATAT AGCC TTTCT ACAC ACCA CCGT CTTG	

							CACC CACT AAAC CAGT GGGA TATG CGGG TGGA CTCA ATGT AGAA AGGT GTTC CCAC TGCC TGAC TCGC CAAC TTTAT ATGT CTTGT CAAC ATAA	

							TGGC CCCT CACT ATAA ACTC CCTA GCAA CTGG TGGT CCGG CGAA GCCG GTTCT TGCC ACTA TTGC GCCC CAGG CTCG CC (SEQ ID NO: 1298)	

NeSL	NeSL-4_SM	.		Schmidtea mediterranea	TTAAATCATTTTTAAATGTGTTTGAATATCT	TTAAAT	TGAG	MNVDLDATIKSI
					TAAATTATCAAATCATATTAATATCAATGCT	CATTTT	TGTG	GMNTKETTYPN
					AAAAAAAAATCGTGCKCATCAGGCGCACGA	TAAATG	CTAC	SQLRVETTPCTS
					AAATAATGGACACAACTCGTCGACCTGCTG	TGTTTG	GAGG	TTIMHASCNTTS
					TCGACTCACAGAGAACCTCAATTTGGAAGA	AATATC	CAGC	TISYSPLPSAVSL
					ATGGGAAGCCTATAATGCTACAATTCCGCC	TTAAAT	GCTG	PESPASSITITTT
					AACCCCTATTTGAATGACAGATAGTCAAAT	TATCAA	GTAA	DDNCDIIETPYP
					ATCAAAAAATATACAAACTGCTGTCAAGCG	ATCATA	TTGC	LPQTNGDLSEIL
					TGACTCACTTCCTTCCAATCGAAAAATAGG	TTAATA	ATCG	KDIEANKDTTM
					AAKATGTAAGAAACATGAAAGTCAAGCTGA	TCAATG	GCGT	SNKVLDCDSDS
					AAAACCAATAATATGTCCTAAAATAAAACA	CTAAAA	TGCA	GDDRDMIIENDR
					ATTTGAAAATATGCAAAAAATACCTATAAA	AAAAAT	GATT	ESDMDLFSQSLL
					ATCACAGCCGAATAAATTCCCATCCGTTCTA	CGTGCK	TGTG	NTNQSDERREK
					AGCAGAAACCGCTACGAACTACTGCAAGAA	CATCAG	TACG	NLTENAPTEITT
					TCGGATCAAGTATATTAATTTCCCCCCMGG	GCGCAC	ATAG	EKSYFDIISKASD
					GGGAAATTAATATACTTGTTAKAAAATTAA	GAAAAT	ATAA	NTTSKKLLNVK
					TTTTTTAATAAAAATAAATAAATCGAATAA	AATGGA	AAAC	NELTAGLPPMPP
					ATATAAAATAAAAATAAATCAAATTAAACT	CACAAC	CAAT	VTNTAKFIRNVR
					TTTATTAACAATAAAATCGCAGTAAGTAAA	TCGTCG	AGTA	PEDIADPTLYRL
					TTTCCACTGTTATTAAATTTAAAACAAAATT	ACCTGC	ATAA	DSRGKLGCRTQ
					CCTTTAAAAATGCCTCTCTTTTTCAGTAATA	TGTCGA	ATGC	YKKPGCGDIAV
					ACACCTTTTCTTGCTTTTATTACTATTTCTTG	CTCACA	TGAG	YDYEAIVEHAA
					TGTACTGTACAAATCGAGCACAGTTATTGC	GAGAAC	CCTA	FIHTIPFNEQNN

					AAATAGGACATAGAAATTCCTTTTTAAGTA	CTCAAT	GCTC	VDCQPCHPKKG
					AATTTAAATCCATGAGAAATAAAATAAAAT	TTGGAA	GCAT	KDVHTIVLIKYA
					CCTTTTGATTCAAAGTTTCTATGTTGCTTTCT	GAATGG	ATCT	DIFNHIEAHSHV
					AATAGAATGGTGTAAGCATTAATGGGTCTT	GAAGCC	AAGC	VQTAITDNMKT
					GATTTTTATAAATTAAATATATTTAATCTAT	TATAAT	CGAA	YLRLTKENXFY
					TAAATTAATATGTTTTTATTAATTATTAATT	GCTACA	AGGC	CSYRNNKKKNK
					TTTATAGTGGGGGGAAATTAATATACTTGA	ATTCCG	AGCA	CKKAFNLESNM
					TCCCAAGAATCAACTGATGATGAAGAATAT	CCAACC	TATA	MDITEHMKTHT
					GTTATTTCAAAATACATACAAGAAGCTGGA	CCTATT	TATG	GYSFDXNLNILC
					AAAAACAAATCAATCGCTACAATGAATGTG	TGAATG	AGAC	YCGIWKPFTELI
					GATCTCGATGCAACAATTAAAAGTATTGGA	ACAGAT	AATT	AHIKTEHLQEYI
					ATGAACACAAAAGAGACGACCTATCCAAAT	AGTCAA	TAAA	NSIPNKENIHNT
					TCACAACTGCGAGTTGAGACGACTCCCTGT	ATATCA	AAAA	TTIVSPLNFAGIL
					ACCTCAACGACTATTATGCATGCATCTTGCA	AAAAAT	AA	ASGETQNIPDEEI
					ACACAACCAGCACTATATCTTACTCTCCATT	ATACAA	(SEQ	IKPRDLPENLAF
					ACCATCGGCTGTGTCACTTCCCGAAAGCCCT	ACTGCT	ID NO:	NRNIENELSWSQ
					GCCTCGTCAATCACAATAACCACAACAGAC	GTCAAG	1299)	HLVKAYIFSYAV
					GATAATTGCGATATTATAGAGACCCCTTAC	CGTGAC		KTSTIFINPYTCN
					CCATTACCTCAAACAAATGGTGACTTGAGT	TCACTT		ALIQCNYKTFFE
					GAAATATTAAAGGATATAGAAGCTAATAAG	CCTTCC		TFPFKDFAKWN
					GACACCACCATGTCGAATAAAGTATTGGAC	AATCGA		EIVLPIHNNTSS
					TGTGACTCTGACAGCGGCGATGATCGGGAC	AAAATA		WSFFFLNKKKR
					ATGATAATAGAAAATGACCGAGAATCTGAC	GGAAKA		VAMIIDPTADDS

					ATGGACCTGTTTTCGCAATCTTTATTGAACA	TGTAAG		HTLHFELATDIL
					CTAATCAATCTGATGAGAGGAGGGAGAAAA	AAACAT		RTILNVQNIFED
					ACTTAACAGAAAATGCTCCAACAGAGATTA	GAAAGT		LNFPLTEVEYPV
					CTACTGAGAAGAGCTACTTTGATATCATCA	CAAGCT		CHEANLSAFXV
					GTAAAGCATCTGATAATACAACCTCTAAGA	GAAAAA		CHFLKCLMSDL
					AACTGCTTAATGTAAAAAACGAATTGACTG	CCAATA		PIDIPDIDHMKET
					CTGGACTACCTCCTATGCCTCCAGTGACCAA	ATATGT		MRPIIRKYNCAK
					TACTGCAAAATTCATTCGAAATGTTCGACCT	CCTAAA		FPESDVRNYRVL
					GAGGATATTGCAGATCCTACCCTATATCGA	ATAAAA		IEDLIYQLNLDTI
					CTTGACAGCAGGGGAAAGCTTGGATGCAGA	CAATTT		TCEEILCEIERIN
					ACWCAATACAAAAAACCCGGATGCGGGGA	GAAAAT		GRLNPKRYFKES
					CATAGCAGTATATGACTATGAGGCGATAGT	ATGCAA		KPKTDIIHLQKK
					TGAACATGCCGCATTTATCCACACAATCCC	AAAATA		KSAELLCVKRL
					ATTTAATGAACAAAATAATGTGGATTGTCA	CCTATA		KFQISQKTEIGKI
					ACCATGCCACCCTAAAAAAGGAAAAGATGT	AAATCA		WENDDVDHRPP
					CCATACAATAGTTCTGATAAAATATGCAGA	CAGCCG		MARFLKTFASQ
					TATCTTTAACCATATTGAAGCCCATAGCCAC	AATAAA		DCPVSNTSSINL
					GTTGTGCAAACCGCGATTACAGATAACATG	TTCCCA		PYYMDTDTDXC
					AAAACCTATCTACGTTTAACAAAGGAAAAT	TCCGTT		TDCENLSHIMK
					TTKTTCTACTGCTCATATCGTAACAACAAAA	CTAAGC		NLDSSAPGMDLI
					AAAAGAATAAATGCAAAAAGGCTTTTAACC	AGAAAC		TGGDWKKISPK
					TTGAATCAAACATGATGGACATAACAGAGC	CGCTAC		HELITAICNCILR
					ACATGAAAACTCATACCGGATACAGTTTCG	GAACTA		NKVCPEKWKLF

					ACMAAAACTTAAACATTCTATGCTATTGTG	CTGCAA		RTVLILKPGKMS
					GTATCTGGAAGCCGTTCACAGAGCTCATTG	GAATCG		ESFRANSWRPL
					CCCACATCAAGACTGAGCATTTGCAAGAAT	GATCAA		AIMDTAYRIFTT
					ATATTAACTCAATACCAAACAAAGAAAATA	GTATAT		LLNNRLLQWIR
					TCCATAATACTACTACCATAGTTTCCCCTCT	TAATTT		NGNLISPNQKAI
					AAACTTTGCTGGGATACTTGCATCTGGCGA	CCCCCC		GIPDGCAEHNAT
					AACTCAAAATATCCCCGATGAAGAAATAAT	MGGGG		LHFAIDRAKRCK
					TAAACCCAGAGATCTGCCAGAAAATCTTGC	GAAATT		TELHIVWLDIAD
					CTTCAACCGAAACATCGAAAATGAATTAAG	AATATA		XFGSLPHDLIWY
					TTGATGGTCGCAGCACTTGGTCAAAGCATA	CTTGTT		TLANMGLKNET
					TATCTTCTCTTATGCTGTTAAAACATCAACC	AKAAAA		LTLIKELYKDVK
					ATCTTCATCAATCCTTATACTTGCAATGCTT	TTAATT		TIFDCQGTLSEP
					TGATCCAGTGCAACTACAAAACTTTCTTTGA	TTTTAA		VPITKGVKQGCP
					AACCTTCCCTTTCAAAGACTTTGCCAAGTGG	TAAAAA		LSMTLFCLSIDYI
					AACGAGATAGTCCTGCCAATTCACAACAAC	TAAATA		LKSILTNYPFLL
					ACTTCTTCTTGGTCCTTCTTCTTCTTGAACAA	AATCGA		HDLNISILAYAD
					GAAAAAACGAGTTGCGATGATTATCGATCC	ATAAAT		DLVLLSDSYLEI
					AACTGCAGATGACAGTCATACCCTGCACTT	ATAAAA		KKSLESTVELAA
					TGAATTGGCTACAGATATCCTAAGGACTAT	TAAAAA		FANLKFKPSKSG
					ACTTAACGTCCAGAATATATTTGAGGACTT	TAAATC		YLSINNVNSDIL
					AAATTTCCCTCTTACTGAGGTCGAATACCCC	AAATTA		KLHLYNEEIPTIS
					GTGTGTCATGAGGCAAACCTTTCCGCATTTT	AACTTT		ENNKYRYLGVD
					MTGTATGCCACTTTCTTAAATGTTTAATGTC	TATTAA		FSYKRNQDVDG

					GGACTTGCCAATTGATATTCCGGATATCGAT	CAATAA		RLGSALALTRSL
					CACATGAAAGAGACWATGAGACCAATTATT	AATCGC		FKSYLHPAQKL
					AGAAAATATAACTGCGCAAAGTTTCCGGAG	AGTAAG		NAYKTFIHSKLI
					AGTGATGTTAGGAATTACCGCGTACTTATC	TAAATT		FSLRNCVIGHRI
					GAGGACCTGATATACCAATTGAACCTTGAC	TCCACT		LDCDRNRVTQG
					ACAATTACTTGTGAGGAAATACTGTGCGAA	GTTATT		REKQLGFDQEIK
					ATCGAAAGAATAAATGGAAGGTTAAATCCC	AAATTT		ALLKTMIGDKF
					AAAAGATATTTTAAAGAGAGTAAACCAAAG	AAAACA		QAXNNYFPYTH
					ACGGATATAATACATCTGCAAAAGAAAAAG	AAATTC		CKLGGLGITSAI
					TCGGCGGAACTCCTCTGTGTTAAAAGATTG	CTTTAA		DEYLIQSITGITR
					AAATTCCAAATCAGTCAGAAAACAGAAATC	AAATGC		LFHSSNLSFRKM
					GGAAAGATATGGGAAAACGACGATGTGGA	CTCTCT		LITELAHSRGGK
					TCACAGACCGCCTATGGCCAGATTCTTGAA	TTTTCA		NFEAGLKWLNC
					GACTTTCGCGAGTCAAGACTGCCCCGTTTCG	GTAATA		EVNKAFPNTSFF
					AATACGTCATCCATAAACCTACCTTACTACA	ACACCT		VKFQKSALALK
					TGGATACTGATACAGATAMGTGTACTGATT	TTTCTT		RKFCICVNLKFV
					GTGAAAATTTGTCGCACATCATGAAGAACT	GCTTTT		EDNFSLEMTYK
					TGGATAGCTCGGCACCTGGAATGGACCTCA	ATTACT		KRTSYVNHQNL
					TTACAGGTGGAGACTGGAAAAAGATCTCCC	ATTTCT		STLSKELHDFVG
					CGAAGCATGAACTGATAACAGCAATCTGCA	TGTGTA		LYYAEQXCQMR
					ATTGTATACTACGAAATAAGGTCTGCCCAG	CTGTAC		VQGHIATAIGDS
					AGAAATGGAAGCTGTTTAGAACAGTTTTAA	AAATCG		ITAKYLIASDILN
					TCCTAAAACCAGGAAAAATGTCCGAGAGTT	AGCACA		DAQYYFLVRAR

					TCAGAGCTAACTCATGGAGACCTCTTGCAA	GTTATT		NNLLNLNYNAY
					TCATGGACACAGCCTATAGAATCTTTACGA	GCAAAT		RLKYNIGTKCRL
					CTCTGCTGAATAACCGCCTGCTGCAATGGA	AGGACA		CHLDEETQAHX
					TCAGGAATGGCAACCTCATAAGCCCGAACC	TAGAAA		FNHCRAKPNAR
					AAAAAGCGATTGGTATACCGGATGGATGTG	TTCCTTT		RVKHENVLVSIV
					CTGAGCATAATGCTACTCTACACTTCGCAAT	TTAAGT		AFLEKIGFEIDVE
					TGACCGAGCTAAACGATGTAAAACTGAACT	AAATTT		KSPKYISIPTKLK
					ACACATTGTTTGGCTCGATATCGCCGATKCA	AAATCC		PDMVIRSKRNK
					TTTGGTTCGCTGCCTCATGACCTGATCTGGT	ATGAGA		DIHVLDLKVPY
					ATACACTGGCTAATATGGGTCTGAAGAATG	AATAAA		DSGEGFEKARE
					AAACACTAACCTTGATTAAGGAACTATATA	ATAAAA		DNYVKYKDLAI
					AGGATGTGAAGACTATCTTCGACTGTCAGG	TCCTTTT		EIGKAFNQKATI
					GAACCTTGTCCGAACCTGTCCCAATTACTAA	GATTCA		SAVVIGCLGTW
					AGGAGTTAAACAAGGTTGCCCATTATCAAT	AAGTTT		DKKNNAALSKI
					GACACTCTTCTGCCTGTCTATTGACTACATT	CTATGT		GLTKTEIISLARI
					CTAAAGTCAATACTGACTAATTATCCCTTCC	TGCTTT		ACPNAVIACYHI
					TTCTTCATGATCTGAACATCAGTATTTTGGC	CTAATA		YREHVSFTKSA
					ATATGCTGATGACTTGGTTCTTCTTTCTGAC	GAATGG		MALPFSLA (SEQ
					TCTTATCTAGAAATCAAAAAATCTTTAGAG	TGTAAG		ID NO: 1421)
					AGTACTGTGGAATTGGCAGCMTTTGCCAAC	CATTAA		
					CTTAAGTTTAAACCTTCGAAGTCTGGATACT	TGGGTC		
					TGTCCATCAACAATGTTAACTCCGATATCCT	TTGATT		
					TAAATTACATCTCTATAATGAGGAGATACC	TTTATA		

					AACGATATCCGAGAATAACAAATACAGATA	AATTAA		
					TCTTGGAGTTGACTTCTCTTACAAAAGAAAT	ATATAT		
					CAGGATGTTGATGGACGACTTGGGTCTGCA	TTAATC		
					CTTGCACTCACCAGATCTCTATTTAAATCAT	TATTAA		
					ACTTGCATCCGGCGCAAAAGCTGAATGCTT	ATTAAT		
					ACAAAACCTTCATCCACTCCAAGCTTATCTT	ATGTTT		
					CTCCTTGCGTAATTGCGTGATAGGTCATAGA	TTATTA		
					ATCCTCGACTGTGATCGGAATAGAGTTACA	ATTATT		
					CAAGGTCGGGAAAAACAGCTGGGCTTTGAT	AATTTT		
					CAGGAAATCAAGGCACTWCTGAAAACCAT	TATAGT		
					GATTGGAGACAAATTTCAGGCAKTAAATAA	GGGGGG		
					CTACTTTCCTTACACTCACTGCAAGCTGGGG	AAATTA		
					GGACTTGGTATAACCTCAGCTATTGATGAA	ATATAC		
					TATTTGATCCAAAGCATTACCGGAATAACA	TTGATC		
					AGATTATTTCACTCATCCAACCTCAGCTTCA	CCAAGA		
					GAAAAATGCTAATCACAGAACTCGCTCATT	ATCAAC		
					CTAGAGGAGGGAAAAACTTTGAAGCGGGG	TGATGA		
					CTAAAATGGCTTAACTGTGAAGTTAACAAG	TGAAGA		
					GCATTCCCCAACACCTCTTTCTTTGTAAAAT	ATATGT		
					TCCAAAAATCGGCACTTGCTCTTAAAAGAA	TATTTC		
					AGTTCTGTATATGCGTTAACCTTAAATTTGT	AAAATA		
					AGAGGACAATTTCTCACTTGAGATGACCTA	CATACA		
					CAAAAAGCGCACTTCTTATGTAAACCATCA	AGAAGC		

					AAACCTCAGCACACTTTCCAAAGAACTCCA CGACTTCGTGGGCCTTTACTATGCWGAGCA ATGWTGTCAAATGAGAGTACAAGGACACA TTGCGACTGCGATCGGGGATAGCATAACAG CTAAATACCTAATAGCTAGTGACATCCTTA ACGACGCACAGTACTACTTCTTGGTACGTG CGAGAAATAATCTTCTGAATCTTAACTACA ATGCGTATCGACTCAAGTATAATATTGGCA CAAAGTGCAGACTTTGCCACCTTGATGAAG AAACTCAGGCCCATSTGTTCAATCACTGCCG TGCCAAACCAAACGCTAGAAGAGTGAAACA CGAAAATGTGCTAGTAAGCATAGTTGCCTT CCTAGAGAAAATTGGATTTGAGATTGATGT GGAAAAATCACCCAAATATATCTCAATACC AACAAAGCTGAAACCTGACATGGTAATTAG GTCTAAGAGGAATAAAGATATACATGTCCT AGACCTAAAAGTGCCATATGACTCAGGAGA AGGCTTTGAAAAAGCGCGGGAAGACAACTA TGTTAAATACAAAGATCTAGCCATTGAAAT TGGAAAGGCATTTAATCAAAAAGCGACTAT ATCTGCTGTGGTGATTGGATGCCTGGGCAC ATGGGACAAGAAGAACAATGCCGCTCTTTC CAAAATCGGGTTGACTAAGACCGAGATCAT	TGGAAA AAACAA ATCAAT CGCTAC A (SEQ ID NO: 1176)		

					ATCTCTTGCCAGGATAGCATGCCCAAATGC GGTAATCGCATGCTATCACATATACCGTGA GCACGTCTCATTTACAAAGAGTGCCATGGC CCTCCCCTTCAGCCTTGCATGAGTGTGCTAC GAGGCAGCGCTGGTAATTGCATCGGCGTTG CAGATTTGTGTACGATAGATAAAAACCAAT AGTAATAAATGCTGAGCCTAGCTCGCATAT CTAAGCCGAAAGGCAGCATATATATGAGAC AATTTAAAAAAAAA (SEQ ID NO: 1053)			
NeSL	R5	AY216701		Girardia tigrina	GTAGGTAACTATGACTGCAAAATAATAATT CTACACCTATTGTTGATAACTCATCTCGTGC GCAAACGGAGCATGTTATTTCTAATCATTTC GTCACACAGGATTCTTCTAATTCTGATAGTA ATATTATAGATAGAGATAGGAACCTTGTTG ATTTAGATGCGTCAATAACTTCTCCTACTAT TATACAGCCAGAGGATAGTAAGATATCTGA GGATGAGGACTTCATCTTAGTCAATAGGAA AAAGAGCAAAAATAAGAAAAAATCTAAGA AAACAACTGAAAATAAAAATGAAATTCCTA TTCAAAAGAGTAAAGATAAGAAAAAGAAG TCTAAAATTAATACCGAAAAACTAACTGAA AATATTACTACTTCTGAAATACCACTTGAAA TTGCTCCTTCCATACCTTTACCTTCAGCAAG	GTAGGT AACTAT GACTGC AAAATA ATAATT CTACAC CTATTG TTGATA ACTCAT CTCGTG CGCAAA CGGAGC ATGTTA TTTCTA	TGAT CCGT GTGT TTGT GTCG TATG ATTG TTTCC GTGT GTGT CTAT ATTTT TCTTT TTTAT	TTGRNLGQWSC YSRSIQQSNYSF KLSSTEVGELVE QSPAPLQSPQFS NNYNNLNINNN LYYSLNTFNQSN NLCCLVNIEFFP TQHLLGDIVNSG CINYMNNYNNF DNINLYINSNVL SYNNYNHSFLA SPYTTNITEHADI NMHVQEVNMQ QDNNTQHAITQ

					TACCTCGGGTTCTCAACAACCGGCCAATCCT	ATCATT	ACTTT	QVSLQATSLQH
					CCAGAAGACGCTACTCTAAGTGATACGGAT	TCGTCA	CAAT	TLDEMIVQFNTA
					CTCTTCCTTACACAGGATGATCCCGATAGTC	CACAGG	TACC	VRLKKKHKVAK
					TTATTCTTTCTGGAAGTACTCAACCAACCTT	ATTCTT	TCGTT	IFRGHNHRKDLP
					TGTTGACCTCAACCCTTCACAGCAATCGGA	CTAATT	GTAA	TLPAREQYKTKP
					ACTTCCTTCAAATACTGACAGCCAAAGATTT	CTGATA	TGTT	KLAIREVLHRKT
					GAGGCGGGTGAAACACCCAAAATCATAACT	GTAATA	ATAA	TATSSPSENAIK
					TCTTACAGGGATGACCTTTTCTACTCTACAG	TTATAG	CTTC	AFFSSYSRPAEL
					TCCTTCACTACAACTCAGATACAGGTTACG	ATAGAG	ATAT	FTGQELLESSWF
					GTATAAGTGTTGACAATGGGGAGCAGAGGT	ATAGGA	GGAA	PVHPEDDFEFRI
					TTCGAATTCTTGCTAGGAATCTTGTCAGGAA	ACCTTG	TATA	PGRDQIAKYIKF
					AACCAAGGATAAGTTCCCCTCTTTATATGCT	TTGATT	TGTA	ASKSAAGLDWI
					GGACAAGTAATTAGACACACAGTCTTCTTC	TAGATG	ATTT	TYEDIKLGDPSG
					AATCACTTCAACCAGGCATACTACGCCAAT	CGTCAA	AGTT	EILQPIFEYIVQN
					AATATAACTGATAGTAAAGGTAATCTAATT	TAACTT	TAGT	NICPSEGKASRTI
					GAGTTTTCTGATGATAAGCCTTTTCAAAGTA	CTCCTA	TTAG	MIPKPGKSDYSD
					TACCGACTGACCCAAAAACTGAACTAGAGC	CTATTA	TTAG	PSSWRPITITSAV
					AAATTAGGAGAGAGAGACAACATCTAGTTG	TACAGC	TTTA	YRLLMKYLTWE
					ATAGAGCTCTTAGACATAATCAGTTACGGG	CAGAGG	GTTT	LYNWILLNQML
					AAACTTATATTTTAAATAAACTTAATAATAA	ATAGTA	AGTT	SRSQKSLGKFEG
					TAATGGGGGGGGTGGCGAACATTTGAAAAG	AGATAT	TAGT	CHDHNAMLNM
					GAAAAAGATCAAAGTCAATACGGATGATGT	CTGAGG	TTAG	LIQDVRRQTNPS
					CTCCAGCAATGATGGAGACAGAAAACATAG	ATGAGG	TTTA	NPINKNKRLYIV

					ACGACAGGAAGAAACCTCGGACAATGGAG	ACTTCA	GTTA	FLDFTNAFGSVP
					CTGCTATTCCCGCTCAATCCAACAATCAAAT	TCTTAG	GTTT	LDTLMYVPQRF
					TACTCCTTTAAACTGAGTTCGACTGAAGTGG	TCAATA	AGTT	GLGTSALTLIKN
					GAGAGCTCGTTGAGCAATCTCCAGCTCCCC	GGAAAA	AGTT	LYLDNYTNVTC
					TTCAGTCGCCTCAGTTCTCTAACAATTATAA	AGAGCA	TAGT	GESKIENVKLNK
					CAATCTAAATATCAACAACAACTTATACTA	AAAATA	TAGT	GVKQGCPLSML
					TAGTCTCAATACTTTTAATCAGTCTAATAAC	AGAAAA	(SEQ	LFNIFINIIIRAIE
					CTTTGCTGTCTTGTTAATATTGAGTTTTTCCC	AATCTA	ID NO:	AMPDVHGYPLG
					AACTCAACACCTTCTTGGTGATATAGTTAAC	AGAAAA	1300)	DMDIRILAYAD
					TCGGGGTGCATAAACTATATGAATAATTAT	CAACTG		DIALISDSHKDL
					AATAACTTTGATAATATTAATTTATATATTA	AAAATA		QEMVYKAEYIG
					ATAGTAATGTATTATCTTACAATAATTATAA	AAAATG		RILGLLFNPSKC
					TCACAGTTTTCTCGCTTCCCCATATACTACT	AAATTC		ALMDIPHDKKR
					AACATCACAGAACATGCAGACATAAACATG	CTATTC		TPPILVNGEMIK
					CACGTGCAAGAAGTTAACATGCAGCAAGAT	AAAAGA		CVGKADPYKYL
					AACAATACACAACATGCTATAACACAACAA	GTAAAG		GTFRSWFRKLDI
					GTCTCTCTACAAGCAACATCTCTGCAACAC	ATAAGA		KELLQMMMDE
					ACGTTGGACGAAATGATAGTCCAGTTTAAC	AAAAGA		TKLITESNLHPH
					ACTGCTGTCAGGTTAAAGAAAAAGCACAAA	AGTCTA		QKIHAYETFIHS
					GTTGCAAAAATCTTTAGGGGACATAATCAT	AAATTA		QLPFHLRHSRIPF
					CGTAAAGACCTTCCAACATTGCCTGCTAGG	ATACCG		SDFITNRKTNKT
					GAACAGTATAAAACTAAACCGAAACTTGCA	AAAAAC		TNNSNDSEKSIQ
					ATTAGAGAGGTACTTCATCGAAAAACAACA	TAACTG		KAYDPESGQLFL

					GCTACGTCTTCCCCTTCTGAAAATGCAATTA	AAAATA		NTFALPSGCAK
					AGGCTTTTTTCTCCTCCTACAGCCGTCCAGC	TTACTA		DFFYITKDAGGP
					TGAACTTTTCACTGGTCAGGAACTTCTTGAA	CTTCTG		QLTSGLDEYLIQ
					TCATCTTGGTTCCCAGTACACCCGGAAGAT	AAATAC		SIMYIFRLLGSE
					GACTTTGAGTTTAGAATTCCGGGTAGAGAC	CACTTG		DPTLNSAIKHDL
					CAAATAGCGAAATACATCAAGTTTGCTAGT	AAATTG		ISHLNLKGFVNI
					AAATCAGCTGCTGGTCTTGACTGGATCACG	CTCCTT		NFSQAISIFNSNF
					TACGAGGATATTAAGTTAGGCGATCCGTCC	CCATAC		TDRTDHFSHLSR
					GGGGAAATTCTCCAACCCATTTTTGAATATA	CTTTAC		TEWARLQLARK
					TAGTACAAAATAACATATGCCCATCCGAGG	CTTCAG		KLKSTLAIQTNV
					GGAAGGCTAGTAGGACCATTATGATTCCCA	CAAGTA		CLINGHLVLTLS
					AACCGGGAAAAAGTGACTATTCAGATCCTT	CCTCGG		LENNVLLIDSKE
					CTTCTTGGCGGCCCATTACAATTACCAGCGC	GTTCTC		KGDVKKIHASL
					AGTATACAGACTTCTCATGAAATATCTTACA	AACAAC		MGFLRLAHLIRL
					TGGGAGCTGTATAACTGGATTCTTCTTAATC	CGGCCA		QKHGWSKLLFS
					AGATGCTGTCCAGGAGCCAAAAGAGTTTAG	ATCCTC		ATTHHEILNKRI
					GGAAGTTTGAGGGATGTCATGATCACAACG	CAGAAG		LNGHVPYKIWY
					CAATGTTGAACATGCTCATCCAAGACGTTA	ACGCTA		FIHRARLGLLPT
					GGAGACAGACCAACCCGTCTAATCCAATCA	CTCTAA		KLFSVSNLCRKC
					ATAAGAATAAGAGGCTATACATAGTCTTCC	GTGATA		GGKKETMSHAL
					TAGACTTTACGAATGCTTTCGGGTCGGTTCC	CGGATC		VNCPMMQTLIN
					GTTAGATACTCTCATGTATGTCCCTCAACGC	TCTTCC		ERHDALEISLVQ
					TTTGGCTTAGGCACCTCTGCTTTAACGCTGA	TTACAC		ILSSKFQGTVIRQ

					TTAAAAACTTATATCTAGATAACTACACAA	AGGATG		KTYVNELRPDIT
					ATGTAACATGTGGGGAAAGCAAAATAGAA	ATCCCG		MESDTQYYLVE
					AACGTAAAATTAAATAAAGGGGTTAAGCAA	ATAGTC		VKCPFDTKMSF
					GGCTGCCCTCTATCTATGCTGCTTTTCAACA	TTATTC		ELRTQQTTDKY
					TTTTTATCAATATTATAATTAGGGCAATAGA	TTTCTG		NIIIEILEDVHPG
					AGCTATGCCAGATGTCCATGGATACCCACT	GAAGTA		KEVRLVTFIVGT
					TGGAGATATGGACATCCGGATACTGGCATA	CTCAAC		LGSWGPQNSDF
					TGCTGATGATATTGCTCTAATATCTGACTCC	CAACCT		LRDLGFSKDEID
					CACAAAGACCTGCAGGAAATGGTCTACAAG	TTGTTG		QVKTRLMLQNI
					GCGGAATATATCGGTCGGATTCTTGGACTA	ACCTCA		NSSCEQWKRFV
					CTCTTCAACCCGTCAAAATGTGCACTTATGG	ACCCTT		QYAPTITPGPIPD
					ACATTCCGCACGACAAGAAGAGGACGCCGC	CACAGC		AESEDDQGTSD
					CTATCCTCGTCAACGGTGAGATGATCAAGT	AATCGG		NGPTAATVQGP
					GTGTTGGAAAGGCCGACCCATACAAATATC	AACTTC		VIGDEEEELQIY
					TTGGAACCTTTAGATCCTGGTTCCGGAAGCT	CTTCAA		DSGLDESSDDEP
					GGATATAAAGGAGCTCCTCCAGATGATGAT	ATACTG		DPDDAELLFTIDI
					GGATGAGACTAAACTCATCACCGAGTCAAA	ACAGCC		EQYLNSVITD
					TCTACATCCTCACCAAAAAATCCACGCGTA	AAAGAT		(SEQ ID NO:
					TGAGACCTTCATTCACAGCCAGCTCCCATTT	TTGAGG		1422)
					CACCTTAGACACAGCCGAATTCCGTTCTCA	CGGGTG		
					GACTTCATAACAAACAGAAAAACAAACAA	AAACAC		
					AACAACAAACAATTCAAACGACTCAGAAAA	CCAAAA		
					ATCTATACAGAAAGCCTACGATCCGGAATC	TCATAA		

					AGGACAATTATTCCTCAACACCTTCGCCCTT	CTTCTT		
					CCAAGTGGATGTGCTAAGGATTTCTTTTACA	ACAGGG		
					TTACAAAAGATGCAGGTGGACCTCAACTCA	ATGACC		
					CAAGCGGACTGGATGAGTACTTAATCCAAT	TTTTCT		
					CAATTATGTACATCTTCCGACTATTGGGCAG	ACTCTA		
					TGAGGACCCCACCTTAAACTCTGCAATAAA	CAGTCC		
					ACATGATCTCATTTCCCACTTAAATTTAAAG	TTCACT		
					GGTTTTGTAAATATTAATTTTTCTCAAGCCA	ACAACT		
					TTTCAATCTTTAATTCAAATTTTACCGACCG	CAGATA		
					AACCGATCACTTTTCACATCTTAGCCGCACT	CAGGTT		
					GAATGGGCAAGACTTCAATTAGCTCGGAAA	ACGGTA		
					AAATTGAAGTCAACCTTAGCCATCCAAACT	TAAGTG		
					AATGTCTGTCTGATAAATGGGCATCTTGTCT	TTGACA		
					TAACTCTTTCGCTAGAAAACAACGTTCTGTT	ATGGGG		
					AATTGATAGTAAAGAAAAGGGGGATGTCAA	AGCAGA		
					GAAGATCCATGCATCCCTCATGGGGTTTCTT	GGTTTC		
					AGGTTAGCTCACCTTATCAGACTGCAAAAA	GAATTC		
					CATGGATGGTCAAAACTGCTCTTCAGTGCG	TTGCTA		
					ACCACTCATCACGAAATACTAAATAAGCGT	GGAATC		
					ATCTTGAATGGTCACGTCCCTTATAAGATTT	TTGTCA		
					GGTACTTTATTCATAGGGCGCGGCTGGGGT	GGAAAA		
					TGTTGCCTACTAAACTCTTTAGTGTTAGTAA	CCAAGG		
					CCTTTGTAGGAAGTGCGGGGGGAAGAAAGA	ATAAGT		

					GACCATGTCGCATGCTTTGGTCAACTGTCCA	TCCCCT		
					ATGATGCAGACCCTCATTAATGAGAGACAT	CTTTAT		
					GATGCTCTTGAAATCTCCCTTGTACAAATTC	ATGCTG		
					TTTCTTCTAAATTTCAGGGTACGGTTATAAG	GACAAG		
					GCAAAAGACCTATGTCAACGAGTTAAGACC	TAATTA		
					CGATATCACAATGGAATCGGATACCCAATA	GACACA		
					TTATCTTGTTGAGGTAAAATGCCCCTTTGAC	CAGTCT		
					ACGAAGATGAGTTTTGAATTGAGAACACAA	TCTTCA		
					CAAACTACTGATAAATACAACATTATTATT	ATCACT		
					GAAATATTAGAAGATGTACACCCTGGGAAG	TCAACC		
					GAGGTGCGCCTTGTTACGTTTATTGTAGGCA	AGGCAT		
					CCTTAGGCTCATGGGGCCCGCAGAACTCGG	ACTACG		
					ACTTTTTGAGAGATCTGGGATTCTCCAAAG	CCAATA		
					ACGAAATCGACCAGGTGAAGACGCGGCTGA	ATATAA		
					TGCTTCAGAATATCAATTCCTCCTGCGAGCA	CTGATA		
					GTGGAAAAGATTTGTGCAATATGCACCCAC	GTAAAG		
					AATTACACCTGGGCCGATTCCAGACGCGGA	GTAATC		
					GAGCGAGGACGATCAGGGGACGAGCGACA	TAATTG		
					ATGGGCCAACAGCTGCTACAGTGCAAGGAC	AGTTTT		
					CGGTGATTGGCGATGAGGAGGAGGAACTTC	CTGATG		
					AAATCTACGATTCCGGCCTTGACGAGTCCA	ATAAGC		
					GCGATGATGAACCCGACCCAGATGATGCTG	CTTTTC		
					AATTACTTTTCACAATTGACATAGAACAAT	AAAGTA		

					ATTTGAATTCTGTGATAACAGACTGATCCGT GTGTTTGTGTCGTATGATTGTTTCCGTGTGT GTCTATATTTTTCTTTTTTATACTTTCAATTA CCTCGTTGTAATGTTATAACTTCATATGGAA TATATGTAATTTAGTTTAGTTTAGTTAGTTT AGTTTAGTTTAGTTTAGTTTAGTTAGTTTAG TTAGTTTAGTTAGT (SEQ ID NO: 1054)	TACCGA CTGACC CAAAAA CTGAAC TAGAGC AAATTA GGAGAG AGAGAC AACATC TAGTTG ATAGAG CTCTTA GACATA ATCAGT TACGGG AAACTT ATATTT TAAATA AACTTA ATAATA ATAATG GGGGGG GTGGCG		

						AACATT TGAAAA GGAAAA AGATCA AAGTCA ATACGG ATGATG TCTCCA GCAATG ATGGAG ACAGAA AACATA G (SEQ ID NO: 1177)		
NeSL	Utopia-1B_CPB	.		Chrysemys picta bellii	GTTTAATTCCTTCTGATGGACATCTGCAACA CCCGTCCTGAAGATGGAGTCTCCTGCWKCC ATTTTTGAAAAAATTGATGCTGCCTTGMAG ATATACTCCGCTGCTGCTGMTTTGGAWGAM AATTCTCTCTCTCTCTCACCTWCAGMTGCAS TCWCGTCAMCTSCTGCTGCTCCTGCGTCCTC TACTCCCCAGAAAACTCAGWGGAAGCCTAT CCCGAAKACCACTCTTGGTGCCTCACGGAA	GTTTAA TTCCTT CTGATG GACATC TGCAAC ACCCGT CCTGAA G (SEQ	TAAC CGAG ACCG CCGA CCAG GGAA ATAA CCCA	MESPAXIFEKID AALXIYSAAAX LXXNSLSLSPXX AXXSXXAAPAS STPQKTQXKPIP XTTLGASRKXR TTXKDEXIXXW XKKAPVDTSXG

					GAMCMGGACCACCASCAAGGATGAAAAMA	ID NO:	CTTCC	RXSTRRTALRDL
					TCAGSASCTGGCKGAAGAAAGCCCCTGTGG	1178)	TTCCC	TSRSXNIXXALQ
					ATACCTCTKCAGGGAGAMCTAGCACCAGAA		TGAC	EEDPRRTPPXSR
					GGACAGCTCTTCGGGACCTCACATCCAGGA		GAAC	DQDAERRPAAP
					GCAGKAATATCWCAMCAGCTCTTCAGGAG		CAAG	EKAATRGAPPTI
					GAGGACCCCCGGAGAACCCCTCCCWCTTCC		GGAC	QDQDADRCPAG
					CGGGACCAGGATGCTGAGCGCCGCCCTGCT		GCAC	RDATGGAPRRP
					GCTCCTGAGAAGGCTGCTACCAGAGGAGCC		CCCA	RTRMLXAAPLG
					CCCCCGACGATCCAGGACCAGGATGCTGAT		CCCA	RMPPEEPPPTTR
					CGCTGCCCTGCTGGGAGGGATGCCACCGGA		TGTA	DQDADRRPAAP
					GGAGCCCCCCGACGACCCAGGACCAGGATG		CTTAT	ERDAPEGTTSST
					CTGANCGCTGCCCCGCTGGGAAGGATGCCG		TCGC	PDPETTYHPPVR
					CCGGAGGAGCCCCCCCCGACGACCCGGGAC		TACA	RRAAPRGTHSX
					CAGGATGCTGACCGCCGCCCCGCTGCTCCA		CCGA	AXDLDAARCPS
					GAGAGGGATGCTCCMGAAGGAACCACCTC		TACT	GQRDIVASESST
					CTCAACCCCGGACCCCGAAACCACTTACCA		GACT	PPGATSPPQASL
					CCCGCCTGTCCGGAGGAGGGCCGCTCCAAG		TGGA	PDXEESPAESAG
					GGGAACGCACTCCGSAGCCCWGGATCTCGA		CTCCT	TTEVRPTEGEAG
					TGCCGCACGCTGTCCTTCCGGGCAAAGAGA		TATA	EDDCIYLQYPXP
					CATCGTGGCCAGTGAGTCCAGCACCCCCCC		CATT	TGLLLCPFCLPX
					AGGAGCGACTTCACCTCCTCAAGCTTCTCTG		CCAT	HGVQTLGALSK
					CCAGACCSAGAGGAATCACCTGCCGAGTCK		GGGT	HVRKAHNKRIA
					GCAGGCACAACAGAGGTCCGCCCCACAGAG		GGCG	FRCSRCDAPFET

					GGTGAGGCAGGGGAAGACGACTGCATCTAC		TACC	QKKCKXHXATC
					CTCCAGTACCCGMTCCCTACAGGCCTCCTC		CGAG	KGPLTTAKVNP
					CTCTGCCCCTTCTGCCTCCCCKTCCATGGAG		CCCA	TDTLRVPTPTPT
					TCCAGACCCTCGGGGCCCTCAGCAAACACG		CTTAT	DGPASAPQPASP
					TTCGKAAGGCCCACAACAAACGGATTGCCT		CCAC	EPQXVRGDQPP
					TCCGGTGTAGCCGCTGCGATGCACCMTTCG		TGAC	TEGSATPASRTD
					AGACTCAAAAGAAATGCAAGTMCCATCAM		ACTTT	DATKRTSPASRI
					GCCACATGCAAGGGACCCCTCACAACCGCG		AAAA	PTLDPAVRGITA
					AAGGTSAACCCCACTGACACCCTGCGGGTT		ACTC	TSQVSDLTRCLS
					CCAACCCCGACCCCCACCGACGGTCCAGCT		TTGC	DLIKTIRHNTDT
					TCAGCACCCCAGCCAGCATCCCCAGAGCCA		ACCC	RRXSAPPQVTSC
					CAGCANGTAAGGGGGGACCAACCGCCAAC		CAAT	RPAVGATSTAP
					CGAGGGAAGCGCAACCCCCGCCTCGAGGAC		CTGG	QAARRDPANGG
					TGACGATGCCACCAAAAGGACCAGCCCCGC		GTCT	ASRSPQIPRPDP
					CTCCAGAATCCCCACGCTGGACCCTGCCGT		ATGC	APGRPNTSSKVT
					GAGGGGGATCACCGCCACCTCSCAGGTCAG		CGGT	QRDSDRQKPHA
					CGACCTCACCAGATGCCTCAGCGACCTCAT		TATG	PPRTPQPDTTRR
					AAAAACCATCCGGCACAACACGGATACGAG		CGAT	RTRTIPSASKHD
					ACGGKGCAGCGCTCCCCCACAGGTMACCTC		ATGT	RAPTKPNTGVSR
					ATGCCGCCCTGCCGTAGGAGCAACTAGCAC		ATGT	TPLPPGRSSAAS
					TGCCCCCCAGGCCGCACGGCGAGACCCAGC		ATCT	ETPRAAPPPPHQ
					CAACGGAGGAGCCTCCCGCAGCCCCCAGAT		CTTC	DPRLKTHLNTAP
					CCCACGGCCGGACCCCGCCCCCGGGAGACC		ATCC	QSEGQQGHRLS

					CAACACCTCCTCCAAGGTTACCCAACGAGA		TTGC	PQHLNPRRQRSR
					CTCCGACCGCCAAAAACCCCATGCCCCACC		AACC	RNDGREQRVAT
					GAGGACCCCCCAGCCGGATACCACCCGCAG		GATA	PWQSAWMEEL
					AAGAACCAGAACCATCCCCAGCGCTTCCAA		CCTG	AKAEDFETFDTL
					ACACGACCGCGCCCCCACAAAGCCCAACAC		TAAT	MDRLTAELSAEI
					CGGTGTTTCCAGAACCCCACTCCCTCCCGGA		CCCT	TARRREPQEASR
					AGATCCAGCGCTGCCTCGGAGACACCGAGA		CATA	ATRRFPAPTRNN
					GCTGCCCCCCCCCCCACCGCACCAGGACCCC		ACCC	TAREGRRGDVG
					CGCCTCAAGACCCACCTGAACACCGCTCCC		AAGC	RRYDPAAASRIQ
					CAGTCCGAGGGACAACAAGGCCACAGACTG		CTGA	KLYRTNRTKAM
					TCTCCGCAGCACCTGAACCCGCGGAGACAA		CCCC	REILDGTSSYCAI
					CGCAGCAGGAGGAACGACGGCCGCGAGCA		AGAT	QPERLYSYFKD
					GAGGGTCGCCACGCCGTGGCAATCCGCCTG		GTAC	VFDHEAQTNLQ
					GATGGAGGAGCTGGCAAAGGCTGAGGACTT		AGTA	RPECLLPLPRINL
					CGAGACCTTCGACACCCTGATGGACAGACT		CCTTC	TEDLERDFSPQE
					GACCGCAGAACTGTCTGCGGAAATTACAGC		CCTCT	VQARLMRTKNT
					CAGAAGGAGGGAACCCCAGGAGGCCTCAC		TAAC	APGKDGIRYHLL
					GGGCCACGCGCAGATTCCCCGCGCCGACCC		TCGT	KKRDPGCLVLA
					GTAACAACACTGCCAGAGAAGGCAGGAGA		GTAT	AIFTKCKQFHRV
					GGGGACGTCGGCCGCCGCTACGATCCGGCA		ATTT	PRSWKKSMTVL
					GCTGCATCCCGTATTCAGAAACTATACAGG		AATT	IHKKGERDDPG
					ACGAACCGGACGAAAGCCATGAGGGAGAT		TTAA	NWRPISLCSTIY
					CCTCGACGGGACCTCCTCCTACTGTGCCATC		ACAT	KLYASCLAARIT

					CAGCCCGAGAGACTCTACTCCTACTTCAAG		TAAC	DWSVCGGAVSS
					GATGTGTTCGACCACGAGGCCCAGACCAAC		TTTA	VQKGFMSCEGC
					TTGCAACGCCCAGAGTGCCTTCTCCCGCTAC		ATAA	YEHNFLLQTAIQ
					CCCGGATCAACCTCACGGAGGACCTGGAGC		AATT	EARRSKRQCAV
					GAGATTTTTCCCCGCAGGAGGTGCAGGCGA		TTTA	AWLDLTNAFGSI
					GGCTGATGAGGACCAAAAACACTGCCCCTG		AA	PHHHIFATLGEF
					GAAAAGATGGCATCCGCTACCACCTGCTGA		(SEQ	GMPETFIQILRD
					AGAAGCGAGACCCCGGCTGCCTGGTGCTTG		ID NO:	LYKDCTTTIRAT
					CTGCCATCTTCACCAAATGCAAGCAGTTTCA		1301)	DGETDAIPIRRG
					TCGCGTTCCCCGCTCCTGGAAAAAGTCCAT			VKQGCPLSPIIFN
					GACCGTGCTCATCCACAAAAAAGGCGAGCG			LAMEPLIRAISS
					AGACGACCCCGGCAACTGGAGGCCCATCTC			GPTGFDLHGKKI
					CCTCTGCTCCACCATCTACAAGCTGTATGCC			SILAYADDLALV
					AGCTGCCTCGCGGCAAGGATCACGGACTGG			ADSSESLQQML
					TCAGTGTGCGGGGGCGCCGTCAGCTCAGTA			DVTSQAAEWM
					CAGAAGGGTTTCATGTCCTGCGAGGGATGC			GLRFNPKKCASL
					TACGAGCACAACTTCCTCCTTCAGACGGCC			HVDGGARALVR
					ATCCAGGAAGCCAGGAGGTCCAAGAGGCA			PSRFLIQGEPMA
					GTGCGCCGTAGCATGGCTTGACCTGACCAA			SLEEGEVYQHL
					CGCCTTCGGGTCCATACCCCACCATCACATC			GTPTGVRVRQT
					TTTGCCACCCTGGGAGAGTTCGGGATGCCA			PEDTIAEILRDA
					GAAACCTTCATCCAGATCCTCCGGGACCTCT			AQIDSSLLAPWQ
					ACAAGGACTGCACCACCACCATCCGCGCCA			KINALNTFLIPRI

					CGGACGGAGAGACGGACGCCATCCCCATCC			SFVLRGSAVAK
					GCCGCGGCGTGAAACAAGGATGCCCCCTCA			VPLNKADSTIRQ
					GCCCCATCATCTTCAACCTGGCCATGGAAC			LVKKWLYLPQR
					CGCTCATCCGAGCCATCTCCAGCGGCCCGA			ASTDIIYISHRQG
					CCGGCTTCGACCTGCACGGCAAGAAAATCA			GANVPRMGDLC
					GCATTCTGGCCTACGCGGACGATCTGGCCC			DVAVMTHAFRL
					TGGTCGCCGACAGCTCGGAAAGCCTCCAGC			LTCPDPTVRSIA
					AAATGCTCGACGTCACCAGCCAAGCCGCCG			QEAVRDVVRKR
					AGTGGATGGGCCTCCGCTTCAACCCCAAAA			IARAPSEQDIAT
					AGTGTGCCTCCCTCCACGTCGACGGTGGCG			YLSGSLEAEFGR
					CCAGGGCGCTGGTCCGGCCATCACGATTCC			EGGDLSSLWSR
					TGATCCAGGGCGAGCCCATGGCCTCCCTCG			ARNASRRLGKRI
					AAGAGGGAGAGGTATACCAACACCTCGGCA			GCCWKWCEER
					CACCCACAGGAGTCCGCGTCCGACAGACCC			RELGILVPRIKTP
					CCGAAGACACCATCGCGGAGATCCTGCGAG			DHTIVTPTARA
					ACGCGGCCCAAATCGACTCCTCCCTGCTCG			MLERTLKDAIRC
					CCCCCTGGCAAAAGATCAACGCCCTCAATA			HYAENLKRKPD
					CCTTCCTGATCCCCCGCATCTCCTTTGTCCT			QGKVFEVSSKW
					CAGGGGATCGGCCGTAGCCAAGGTGCCCCT			DASNHFLPGGSF
					GAACAAGGCCGACAGCACCATCAGGCAGCT			TRFADWRFVHR
					GGTGAAGAAGTGGCTCTACCTTCCCCAGAG			ARLNCVPLNGAI
					GGCCAGCACGGACATCATCTACATTTCCCA			RHGNRDKRCRK
					CAGGCAGGGCGGCGCCAACGTACCTCGGAT			CGYANETLPHV

					GGGTGACCTGTGCGACGTGGCGGTGATGAC CCACGCCTTCCGCCTCCTGACGTGCCCGGAC CCGACGGTGAGGAGCATCGCGCAGGAAGCC GTACGGGACGTGGTCAGGAAACGCATCGCC AGGGCCCCCTCCGAGCAGGACATCGCCACT TACCTCAGCGGCTCCCTGGAGGCTGAGTTC GGGAGAGAGGGGGGAGACCTGTCCTCTCTC TGGTCCCGCGCCCGCAACGCCTCGAGACGC CTGGGTAAGAGGATCGGCTGCTGCTGGAAG TGGTGCGAGGAGCGCCGGGAGCTGGGAATA CTGGTGCCACGCATAAAGACCCCGGACCAC ACCATCGTCACCCCGACCGCCAGAGCTATG CTGGAAAGGACCCTGAAAGACGCCATCCGC TGCCACTATGCCGAGAACCTCAAGCGGAAG CCGGACCAGGGCAAGGTGTTCGAGGTGTCC AGCAAGTGGGACGCCAGCAACCACTTCCTC CCCGGGGGCAGCTTCACCAGGTTCGCCGAC TGGCGGTTCGTCCACAGGGCCCGACTCAAC TGCGTTCCCCTCAACGGAGCCATCCGCCAC GGCAACCGGGACAAGCGCTGCAGGAAGTG CGGCTACGCAAACGAGACCCTGCCCCACGT CCTGTGTGGATGCAAACAGCACTCCGGAGC CTGGCGGCACCGCCACAACGCCATCCAGAA			LCGCKQHSGAW RHRHNAIQNRL VKAIPPSLGKITL DSAIPGTDSRLR PDIVVTDAEKK KVLMVDVTVPF ENRSPAFHEARA RKALKYTPLAE TLRAQGYEVQI HALIVGALGSW DPHNEPVLRAC GVGRRYARLMR QLMVSDTIRWS RDIYTEHITGHR QYHTE (SEQ ID NO: 1423)

					CCGGCTGGTGAAAGCCATCCCGCCGTCCCT GGGGAAGATCACCCTCGACTCCGCCATCCC CGGGACAGACAGCAGACTGCGACCCGACAT CGTCGTGACGGACGCAGAAAAGAAGAAGG TCCTCATGGTAGACGTCACGGTGCCTTTTGA AAACAGGTCACCGGCCTTCCACGAGGCCCG AGCACGGAAGGCGTTGAAGTACACCCCGCT GGCCGAGACCCTGAGAGCCCAGGGCTACGA GGTCCAGATACACGCCCTGATCGTGGGAGC CCTGGGCTCGTGGGACCCCCACAACGAGCC GGTTCTGAGAGCGTGCGGAGTCGGTCGACG CTACGCCCGGCTCATGAGACAGCTCATGGT GTCCGACACCATCAGGTGGTCCAGAGACAT TTATACGGAACACATCACAGGACACCGTCA ATACCACACTGAGTAACCGAGACCGCCGAC CAGGGAAATAACCCACTTCCTTCCCTGACG AACCAAGGGACGCACCCCACCCATGTACTT ATTCGCTACACCGATACTGACTTGGACTCCT TATACATTCCATGGGTGGCGTACCCGAGCC CACTTATCCACTGACACTTTAAAAACTCTTG CACCCCAATCTGGGTCTATGCCGGTTATGCG ATATGTATGTATCTCTTCATCCTTGCAACCG ATACCTGTAATCCCTCATAACCCAAGCCTG			

					ACCCCAGATGTACAGTACCTTCCCTCTTAAC TCGTGTATATTTAATTTTAAACATTAACTTT AATAAAATTTTTAAA (SEQ ID NO: 1055)			
NeSL	Utopia-1_ACa	.		Acanthamoeba castellanii	CCCGTCAAGGGTGCTCCACGAGATCCCTGT CGCTAGCCGACCGGTTTTACCACCCCACCCC GCCCGGACAACCACGGACCCTGCTCCGCAG CAGGACCCCACGCACGATGGCCGCTAAATC CGTCGCCTGCCCTCACGATGGATGCGCCAA CAAGTACGCGTCGGAAGCCTCCCTCCGAAG ACACATTAAGAACAAACACGCTACAGATGA GGAAGGAGATGAGACCTCACACTCCTGTCC CCACTGCCACCGACCTTTCTCCACCGCCCGC GGGCTCAGCGTCCACATTGGCAAATCGCAC CGTCAGGCCCCCCCTGAGCCGACGCGCCCC CCCCCGGCCCCGGCCCCTGCCGATCCCGGC CTCGATCCCGACCCCGGCCCCACCGTGACG CCCCCCAGCCGTGATGACGAAGACCGCGAG GAACCCGACGACGACCCCGTGGAGATCGCG GACCTAAGCTGCCCTCACTGCGCCCAGGCC CTCCCGTCGGCCCACGGCCTCGCCAACCAC CTTCGCGCCTGCAAGGACCACAGGGTCCCC GCCCCTGGAGCACCCCGCTCGGGTCCGCCC AGCTCCAGGTACTGGACTGCTGTCGAGCAC	CCCGTC AAGGGT GCTCCA GCTCCA CGAGAT CCCTGT CGCTAG CCGACC GGTTTT ACCACC CCACCC CGCCCG GACAAC CACGGA CCCTGC TCCGCA GCAGGA CCCCAC GCACG (SEQ ID NO: 1179)	TAAC AACC ATGT ATGG TGAA CCAC ACCT CTCTC GATC TTGT ATTCT GTGA TTGG ACAT CAGA GTTC CTGC GAAG GGAT ACAC	MAAKSVACPHD GCANKYASEAS LRRHIKNKHAT DEEGDETSHSCP HCHRPFSTARGL SVHIGKSHRQAP PEPTRPPPAPAP ADPGLDPDPGPT VTPPSRDDEDRE EPDDDPVEIADL SCPHCAQALPSA HGLANHLRACK DHRVPAPGAPR SGPPSSRYWTA VEHHRYVEAMA RFADHPDLLAR AAAHIGTRTYK QVDSHRTKVIA AEREGRPVRTL DPTMDWRMRP

					CACCGCTATGTGGAGGCCATGGCGCGCTTC		TCTG	YCASTTARWLA
					GCGGATCACCCCGACCTACTTGCGCGCGCG		CCAA	EQGRSPVAPRSP
					GCTGCCCACATCGGGACCCGCACGTACAAA		TCTC	CPEPHAPPPAAA
					CAGGTTGACTCCCACCGCACCAAGGTGATC		GTGG	LLYIPATPPAPTP
					GCGGCGGAGCGCGAGGGCCGCCCTGTCCGC		GTTG	RAPVAPPKLAPP
					ACGCTCGACCCCACGATGGACTGGCGCATG		TAAT	AESTVPATPDGN
					CGGCCCTACTGCGCCAGCACCACGGCCCGG		AAAT	PEAPAPPFSAPG
					TGGCTGGCTGAGCAGGGGCGTAGCCCAGTA		CCAC	PPTPKALPPPPPS
					GCGCCCCGCTCGCCCTGCCCCGAGCCCCAC		ACCT	RRNLRPHLVPK
					GCCCCGCCGCCTGCAGCCGCGCTGCTGTAC		TCAA	DAWQGVADAV
					ATCCCGGCCACGCCCCCCGCGCCAACGCCC		CA	APAASRLLRTPL
					CGTGCCCCAGTGGCGCCTCCCAAGCTTGCG		(SEQ	AHLSTEQWATF
					CCTCCCGCCGAGAGCACCGTGCCCGCCACG		ID NO:	EAALAGLEATL
					CCCGATGGGAATCCGGAGGCGCCAGCACCC		1302)	HHAARSAEAVP
					CCGTTTAGCGCCCCCGGACCTCCCACCCCCA			TRCASRAREDA
					AGGCATTGCCGCCCCCGCCCCCGTCCCGCC			ERQLREARKTR
					GCAACCTGCGCCCTCACCTCGTGCCCAAGG			EIFGKAAALYA
					ATGCTTGGCAGGGGGTCGCCGATGCCGTCG			AGKDPTATIERI
					CCCCTGCCGCCTCGCGCCTCCTGCGCACGCC			PPEVRLHLPTPG
					CCTTGCGCACCTCTCCACCGAGCAGTGGGC			SAEWPARAAAA
					CACGTTCGAAGCCGCCCTCGCCGGCCTCGA			RRVIRRAVARA
					GGCTACGCTCCACCATGCCGCCCGCAGTGC			DRLRKRMGILD
					AGAGGCGGTGCCCACACGCTGCGCTAGCCG			SDRDLQRLFNA

					AGCAAGGGAAGACGCCGAGCGCCAACTCC			NQKKAVRQILA
					GTGAAGCCCGAAAGACGCGTGAGATCTTTG			PSTKAPRCQLDP
					GCAAGGCCGCTGCCCTCTACGCAGCCGGCA			AAVEEAYIQTLA
					AGGACCCCACTGCCACCATCGAGCGCATCC			KPPPIDPSPPWK
					CCCCAGAAGTCCGCCTACACCTGCCAACCC			NSVQWPRPPTA
					CTGGCTCGGCTGAATGGCCCGCCAGGGCGG			ADDGGSPFSVA
					CCGCCGCCCGCAGGGTGATCCGCCGTGCAG			EVRAQLRRLPN
					TCGCGCGAGCGGACCGGTTGCGCAAGCGCA			GSAPGIDGIPYE
					TGGGCATCCTCGATAGCGACCGCGACCTCC			AYKRTKLDATL
					AACGCCTCTTCAACGCTAACCAGAAGAAGG			AHVFEVVRLNA
					CAGTTCGGCAGATCCTCGCCCCGTCCACCA			RLPARWDVART
					AGGCGCCGCGGTGCCAGCTAGACCCAGCCG			VLLYKKGDPND
					CCGTCGAGGAGGCCTACATCCAGACCCTCG			TGNWRPISLQVT
					CCAAGCCGCCGCCGATCGACCCCAGCCCCC			IYKIFTAALSKR
					CGTGGAAGAACTCCGTCCAGTGGCCCCGCC			LISWAGKHNTFS
					CGCCCACTGCCGCCGATGACGGAGGCAGCC			ASQKGFLPAEG
					CCTTCAGCGTCGCCGAGGTCCGGGCCCAGC			CHEHAFVLRSV
					TCCGCCGACTGCCCAACGGGTCCGCCCCAG			LDDARRHKQNV
					GGATCGATGGCATACCGTACGAGGCCTACA			YLAWYDLRNAF
					AGCGTACGAAACTGGACGCCACGCTCGCCC			GSVSHDLIAWC
					ATGTCTTCGAGGTCGTGCGGCTGAATGCGC			AAMLGLPRYLR
					GCCTGCCAGCTCGATGGGATGTGGCGCGCA			DAIGAIYRHSAL
					CGGTCCTGCTCTACAAGAAAGGCGACCCTA			FVQVGDQETTG

					ACGACACCGGCAACTGGCGACCGATAAGCC			VIPMRCGVKQG
					TCCAGGTCACCATCTATAAGATCTTCACGGC			CPLSPLLFNLCV
					CGCCCTGTCGAAGCGGCTCATCTCCTGGGCT			EPALRCLRRTTG
					GGCAAGCACAACACTTTCTCCGCATCGCAG			YKFYGTSITVEG
					AAGGGATTCCTACCGGCCGAAGGCTGCCAC			QAYADDLLTAA
					GAGCACGCGTTTGTCTTGCGAAGCGTGCTT			PSAYHAARQVA
					GACGACGCCCGTCGGCACAAGCAGAACGTG			TIEEWANWAGV
					TACCTTGCCTGGTACGATCTGCGCAACGCCT			SFVVQALSLDAP
					TCGGATCGGTGTCGCACGACCTCATCGCCT			AGKCAALAINFE
					GGTGCGCTGCCATGTTGGGCCTGCCCCGCT			GGLMHSIDPAL
					ACCTCCGGGATGCCATCGGCGCAATCTATC			KVQGAAIPAMS
					GGCACTCAGCGCTCTTCGTCCAAGTTGGGG			RNNVYRYLGVH
					ATCAGGAGACCACCGGCGTCATTCCTATGC			VGLTDALGQAN
					GCTGCGGCGTCAAGCAGGGCTGCCCTCTCA			ELLEKASRDART
					GCCCCCTCCTCTTCAACCTGTGCGTCGAGCC			ICASGLEPWQK
					GGCCCTTCGCTGCCTACGCCGCACCACCGG			VVAIKTFILSRLP
					GTACAAGTTCTACGGCACGTCGATCACCGT			FFFHNGKIQRGR
					CGAGGGCCAGGCCTACGCCGACGACCTGCT			CQQFDRELREN
					CACTGCCGCGCCCTCCGCCTACCATGCGGC			LRAALRLPVCTT
					CCGGCAGGTGGCCACGATCGAGGAATGGGC			NAFFHSRVASG
					CAACTGGGCGGGAGTCTCCTTCGTCGTCCA			GLGILPIAEEQQ
					AGCCCTCTCCCTGGATGCGCCGGCCGGCAA			VYLAAHVFKLL
					GTGTGCCGCCCTCGCGATCAACTTCGAAGG			TSPDLSIRAIARH

					TGGTCTAATGCACTCTATCGACCCTGCCCTC			QLAEVTHARHT
					AAGGTCCAAGGCGCAGCCATCCCGGCCATG			TPVQDGEASPFF
					TCAAGAAACAACGTGTACCGCTACCTCGGA			GWLMRGQEVA
					GTACATGTCGGTCTCACAGATGCGCTCGGC			STTPSGDVSSIW
					CAAGCGAACGAGCTCCTCGAGAAGGCCTCA			FAAAGAYSRMG
					CGCGATGCACGCACGATCTGTGCCTCTGGC			WSVRDALHPTL
					CTCGAACCCTGGCAGAAGGTGGTCGCAATC			TVGPGVQFEGR
					AAGACCTTCATCCTCTCCCGGCTCCCCTTCT			FQRANVIPALRA
					TCTTCCACAACGGGAAGATCCAGAGGGGCC			SAFSRHAVEWS
					GATGCCAGCAATTCGACCGCGAGCTTCGAG			ALRTQGRAAAY
					AAAACCTGCGGGCCGCCCTCCGACTCCCCG			QHAVHPATHH
					TCTGCACCACGAACGCCTTCTTCCATTCCCG			WVHNSAGLTTK
					CGTGGCCTCAGGCGGCCTTGGCATCCTGCC			EYRFAIKCRLGL
					CATCGCGGAAGAACAACAAGTCTACCTGGC			LPTRAAPHHRN
					AGCCCACGTGTTCAAGCTCCTGACTTCGCCA			GPTACRACSYA
					GATCTGTCGATCCGCGCCATCGCCCGACAC			RETANHVLGHC
					CAACTTGCCGAGGTCACCCACGCGCGACAC			PATKAEVIARHN
					ACCACGCCAGTCCAGGACGGCGAAGCGTCA			RICRALAQAAE
					CCCTTCTTCGGATGGCTCATGCGGGGGCAG			ASWTSVLEDVPI
					GAGGTCGCATCAACTACCCCCTCGGGTGAC			PGVDSPLRPDIY
					GTCAGTTCAATCTGGTTCGCAGCTGCAGGC			CSRPGQCAIIEV
					GCCTACTCGAGGATGGGATGGTCAGTCCGC			AVSYEDAFNAS
					GATGCACTCCACCCGACGCTGACAGTTGGT			MEGRAKQKTDK

					CCGGGCGTCCAATTCGAGGGCCGATTCCAA CGTGCCAACGTCATCCCAGCTCTCCGGGCT AGCGCCTTTTCCCGCCATGCTGTGGAATGG AGTGCCCTCCGCACCCAGGGTCGAGCAGCA GCCTACCAACATGCCGTCCACCCTGCAACG CACCACTGGGTCCACAACAGCGCTGGCCTG ACGACCAAGGAGTACCGATTCGCGATCAAG TGTCGATTGGGTCTCCTGCCGACGCGAGCA GCTCCACACCACCGCAATGGGCCAACAGCG TGCAGGGCGTGCTCCTACGCCCGCGAGACG GCCAACCATGTTCTCGGACACTGCCCGGCG ACCAAGGCCGAAGTCATCGCGCGCCACAAC AGGATATGCCGAGCTCTGGCCCAGGCGGCT GAAGCCTCATGGACGTCTGTCCTTGAAGAC GTCCCGATCCCGGGGGTGGACTCCCCCCTA CGACCCGACATCTACTGCTCTCGGCCGGGC CAGTGTGCCATCATCGAGGTCGCGGTCTCCT ACGAGGACGCCTTCAACGCTTCGATGGAGG GCCGGGCGAAGCAGAAGACCGACAAGTAC GCTGGCCTGGCTGCTACCGTCGAGGAGCAG CTGCGGCTCCAAACCCGGCACGCGGCTTTC GTGGTGGGCTTCTCTGGCGTCGTGCTCCCAG CCTCGGTAACCGCTACGGCCACCTCCCTTGA			YAGLAATVEEQ LRLQTRHAAFV VGFSGVVLPAS VTATATSLDLPP KTWNVLLKRCV AASIKGSYTAW RRFRRSTP (SEQ ID NO: 1424)

					TCTCCCCCCCAAAACTTGGAATGTGCTTCTT AAACGTTGTGTTGCTGCCTCAATCAAAGGC AGTTACACAGCGTGGAGAAGATTCCGGCGC TCTACTCCATAACAACCATGTATGGTGAAC CACACCTCTCTCGATCTTGTATTCTGTGATT GGACATCAGAGTTCCTGCGAAGGGATACAC TCTGCCAATCTCGTGGGTTGTAATAAATCCA CACCTTCAACA (SEQ ID NO: 1056)			
NeSL	Utopia-1_AEc	.		Acromyrmex echinatior	GGTGCACAACGGATGCATCATACGTGTACC GGAGCATACGGGCTGTCACGGCGGCTGCAT GCGCGATCTAGCTCGGAGATTTTATTTATTT ATTTATTAATTTATTTATTTATTCATCGAGT GTGAGTGTTCGCGTTTTGCCGAGAAGCGAT TTTCGTTAAGTGATACGCGCCGCGTTCATAG GTTAGGTGTGCAGTGTACGTGGCTGTCGTC GCGAAGATTCTCGAAGATTCTACAAGTTCA AATTTCCGTTAAATTTCGTTAAAGTTCCTAA AACCATCGTGATCGGGAGTGCGTTCCAAAA ATCATCAGTTTCGGCTCGTTCGCAAAACCAC TCCCGGTCGACCCGCGTTCCGAAAACCAGA CAGCCACGTACAAGCAATACGATTGGTCGG TACACTGCCGCGAGTGCGAATAACTATTTA ACTGTTATTATTACTGGAAATTACACTGTAT	GGTGCA CAACGG ATGCAT CATACG TGTACC GGAGCA TACGGG CTGTCA CGGCGG CTGCAT GCGCGA TCTAGC TCGGAG ATTTTA TTTATTT	TAAA TTATT TTGTC TTTGT CTTG GCCC CCCC TTTTT AAAC CAAG CAGG AGAG AGTG GCCC AATG	VCSVRGCRRED SRRFYKFKFPLN FVKVPKTIVIGS AFQKSSVSARSQ NHSRSTRVPKTR QPRTSNTIGRYT AASANNYLTVII TGNYTVFAQWI CYRECTWLLSK FVNFFLTIIGYFF QLRLVVIYEGPV ILDTFSNCGSSLF MRGQXSKALLV RLNRSALAMAD PQVHYIDYPLPP

					TCGCCCAGTGGATCTGTTATCGCGAGTGCA	ATTTAT	CCCA	RVKCVKCFGAE
					CTTGGCTTTTAAGTAAATTTGTTAATTTTTTC	TAATTT	ACTA	GAGKVKGEYSD
					TTAACTATAATTGGATACTTTTTCCAATTGC	ATTTAT	TTAT	PPHLAKHLKKC
					GGCTCGTCGTTATTTATGAGGGGCCAGTKA	TTATTC	ATAT	HPGDTLNYKCSI
					TCTAAGGCCCTTTTAGTCAGACTAAACCGTA	ATCGAG	TAAC	CDLRGTGKYPL
					GTGCACTCGCCATGGCGGATCCACAAGTGC	TGTGAG	TATTT	RDVKAHYAECH
					ACTATATAGACTACCCGCTGCCCCCTAGAG	TGTTCG	ACTG	VSPAVDAAGPS
					TCAAATGCGTAAAATGTTTCGGTGCTGAGG	CGTTTT	TGAT	TRGSLGECSGA
					GGGCAGGCAAAGTAAAGGGCGAATACAGC	GCCGAG	ATTT	GQPTASRAAKA
					GACCCGCCGCATTTAGCAAAACATCTGAAA	AAGCGA	ATTA	TTRLAETVGGT
					AAGTGCCACCCGGGAGACACATTAAATTAT	TTTTCG	TTTG	DKRRAATSGSR
					AAATGCTCAATTTGTGATCTAAGGGGGACC	TTAAGT	ACTG	QLTLPFAATPSP
					GGTAAATACCCCCTTAGAGATGTTAAGGCA	GATACG	TTGG	STAAGEARAPRS
					CATTACGCCGAGTGCCATGTGTCTCCCGCA	CGCCGC	GCGG	XSTTPTSRSPSY
					GTGGATGCGGCGGGTCCAAGCACTCGCGGC	GTTCAT	GCCC	AAVTAGPPSMR
					AGCCTCGGCGAGTGCAGTGGTGCGGGTCAA	AGGTTA	CTCTC	STTTSTTARSKT
					CCGACAGCCAGCCGCGCGGCTAAAGCGACC	G (SEQ	TGCT	VAKGAAPNTTT
					ACGCGATTGGCGGAGACGGTTGGGGGTACG	ID NO:	GGTT	TTTARRSGEAA
					GATAAGCGCCGTGCCGCGACATCGGGATCG	1180)	TTATT	ATRKPPTTATVS
					CGGCAACTCACGCTGCCGTTCGCAGCCACC		TATA	KPRVLSVETVRL
					CCATCGCCATCCACAGCAGCCGGTGAGGCA		TATA	PVDDIQRAGVQ
					AGGGCCCCAAGAAGCGMGTCAACGACACC		TTTTT	NAAKPARAPSR
					GACGAGCAGGTCCCCCTCATATGCGGCAGT		TACT	PPQRTSPEAGGP

					CACTGCGGGCCCGCCATCGATGAGGAGCAC		CGCG	RTTGAKEKCGE
					GACAACTTCCACCACAGCCCGCAGTAAGAC		TACTT	GAYKKLPANSG
					TGTCGCGAAAGGCGCCGCGCCCAACACMAC		TTTGT	NPISTRTRRATS
					GACGACGACAACGGCCAGGAGATCCGGCG		ACTA	VPVEKSEGTAR
					AGGCCGCCGCAACGAGGAAGCCGCCTACGA		CTCT	RERVSPHPPPKG
					CCGCCACGGTGAGTAAACCGCGTGTGTTGT		ATTTT	IDIILSSTSEEEGT
					CGGTAGAAACTGTTAGGTTGCCTGTCGACG		TCTTT	PYQPGGVGRLR
					ACATCCAGCGAGCAGGCGTGCAAAACGCGG		TTATT	LRRKKVTGPPPK
					CCAAACCGGCGCGCGCTCCCTCTCGCCCCC		TTAG	MTPREGVVTRA
					CGCAGAGAACATCACCGGAGGCGGGGGGT		CTAT	RRSTSAPVEKSA
					CCAAGAACAACGGGCGCAAAGGAGAAATG		GCTA	LDARLTALDRTS
					CGGAGAGGGAGCATACAAGAAGTTGCCCGC		TTTTT	SRATGNPTSQIA
					AAACAGCGGCAATCCAATTTCAACCAGGAC		ATCT	GGLYTSRGQPE
					GAGGCGGGCAACTAGCGTGCCGGTTGAGAA		CTTTC	RTPPARLPSLSPT
					GAGCGAAGGCACGGCAAGACGGGAGCGCG		TTTGT	TRGSPSGSLGEI
					TCTCCCCACACCCTCCTCCCAAAGGAATTGA		CTCT	RTPISPATSLPAT
					TATCATCCTATCGTCGACATCGGAGGAGGA		ATTTT	LTTCTVTTTTCG
					GGGCACGCCATACCAGCCCGGCGGCGTGGG		CTTTC	SPITSTGFTGGV
					GAGACTAAGACTAAGGAGGAAAAAGGTGA		TTTTT	GRLITPPSLPQTN
					CCGGACCACCCCCAAAGATGACACCCAGAG		TTCTT	ILPTIGEEGTSPC
					AGGGGGTGGTCACAAGAGCCAGGCGGTCCA		TCCTT	VAVVTTHPRPT
					CCAGCGCTCCCGTCGAGAAGAGTGCCTTGG		TTCTT	GEDAPCEAPQP
					ATGCACGCCTGACGGCTCTGGACCGGACAT		TTCTT	VSDHQRQSIGEP

					CGTCCAGAGCGACAGGCAACCCGACGTCGC		TCTTT	RRDTDTHLACD
					AAATCGCAGGGGGCCTTTACACCAGTAGAG		TATTC	VATGNAHHLHG
					GCCAACCGGAGAGGACGCCCCCTGCGAGGC		TTCTT	DDDHLWKPHNI
					TCCCCAGCCTGTCTCCGACCACCAGAGGCA		TTATT	HGLHRWRGEAD
					GTCCATCGGGGAGCCTAGGCGAGATACGGA		TATCT	NTXEPPPNEHPP
					CACCCATCTCGCCTGCGACGTCGCTACCGG		TTTTT	DHRGGRNVTVR
					CAACGCTCACCACTTGCACGGTGACGACGA		TCTTT	GGRHHPSLGGR
					CCACCTGTGGAAGCCCCATAACATCCACGG		TCTGT	STSPLILPRPTTP
					GCTTCACAGGTGGCGTGGGGAGGCTGATAA		TGTG	EPERGQEERRLE
					CACCWCCGAGCCTCCCCCAAACGAACATCC		GGGC	GAAQPPTTPVVE
					TCCCGACCATCGGGGAGGAAGGAACGTCAC		CCTG	GDNQWDGQWT
					CGTGCGTGGCGGTCGTCACCACCCATCCTTA		ACCG	VSVRRRARRQQ
					GGCGGCAGGAGCACGTCTCCGCTCATCCTA		TCCG	LNDTSPSNSESP
					CCAAGGCCAACGACACCGGAGCCTGAGCGG		AGTG	PTAGPSRSPRIAP
					GGACAAGAGGAGCGGCGGCTAGAAGGCGC		TGAA	LSALIAASTSRH
					GGCGCAGCCACCTACCACACCCGTCGTCGA		TGCC	ETSLNLNCTNG
					GGGGGACAACCAGTGGGATGGCCAGTGGA		GCGA	NICMDRTPPRNI
					CGGTGAGCGTGAGGAGAAGAGCGAGGAGG		AAAA	LPVXAERRRETS
					CAACAACTGAACGATACATCCCCCTCCAAC		CAAT	PQDRVEGDIGY
					TCCGAGTCCCCGCCAACCGCTGGACCATCG		ATTA	GAGKVSAEHPS
					CGTTCGCCGCGCATAGCCCCACTATCTGCGC		TGTTT	APVNVRGVMSR
					TGATAGCGGCGTCGACGAGCCGCCATGAGA		TATA	GRATASSIVPPR
					CCAGCTTAAATCTCAATTGCACGAACGGCA		CGAG	ANRGEGGRQHH

					ATATTTGCATGGACCGAACTCCGCCCCGTA		TGTG	SRRRPDAPVGQP
					ACATTTTGCCGGTARGGGCGGAGCGCCGTC		CATG	SRDHPAPATVA
					GCGAGACATCGCCACAGGATCGCGTGGAGG		TGCG	RQRRRERVAAR
					GAGACATCGGTTATGGTGCTGGAAAGGTAA		TGAT	DALLDRAKDVA
					GTGCCGAACACCCGAGTGCTCCCGTAAATG		ATAT	TIADLEAFAASV
					TCCGTGGTGTGATGTCTCGAGGGAGAGCAA		TTATC	AAFFGEDASAT
					CCGCGTCATCCATCGTGCCACCGCGAGCCA		TATTT	GAAARARDRSV
					ACCGTGGGGAGGGCGGTCGGCAGCATCATA		TATTT	RSREAGARRGV
					GTCGGCGGCGTCCGGACGCTCCTGTCGGTC		TATTT	KGGERPEREGA
					AGCCGTCGCGGGATCACCCGGCGCCTGCGA		ATTT	GRPGSAPADPG
					CTGTCGCGAGGCAGCGTAGGCGTGAGCGGG		ATTA	ASGEARGDWVR
					TGGCCGCCCGCGACGCGCTGCTCGATCGGG		TAAT	EAKRLQALYRA
					CTAAGGACGTCGCTACGATTGCGGATCTGG		TTATT	NRRKAVREVLQ
					AGGCGTTCGCGGCTTCGGTCGCGGCGTTCTT		GCCG	GPADQCQVPKR
					CGGGGAGGATGCATCGGCCACTGGTGCTGC		CGCG	QVQEYFERLYS
					AGCCCGCGCTCGCGATCGTTCGGTACGCTC		CGCT	GGEDLAGAGVE
					ACGGGAGGCGGGTGCGCGTCGGGGGGTGA		CCTC	AERPDPSSPREV
					AGGGAGGTGAGCGTCCGGAGAGAGAGGGC		CGGG	SAVLGPLAEREV
					GCCGGTAGGCCGGGGTCAGCGCCGGCTGAC		ACTTT	DRRLRRMNNSA
					CCCGGAGCGTCGGGAGAAGCACGCGGGGA		TATTC	PGPDGVSYRDL
					CTGGGTGCGCGAGGCCAAACGCTTGCAGGC		GTTG	RGADRGARLLT
					GCTGTACAGGGCGAACCGCCGCAAGGCAGT		ACAA	ALYNICLRLEAV
					GCGAGAGGTGCTCCAGGGACCTGCCGATCA		TACT	PASWKTSNTVLI

					GTGCCAGGTGCCTAAACGTCAGGTCCAGGA		GTGA	HKKGDRGMLEN
					GTACTTCGAGCGGCTGTACAGCGGCGGGGA		TATTT	WRPLALGDTVP
					AGACCTGGCTGGCGCCGGCGTGGAAGCCGA		TTCTG	KLFAALLADRL
					ACGCCCTGACCCCTCGAGTCCGCGTGAGGT		CKCA	TDWAVTRGKLC
					ATCTGCGGTCCTGGGTCCGCTCGCGGAGCG		GGCT	SAQKGFLRDEG
					AGAGGTGGACCGTCGGCTCCGGCGTATGAA		GGGG	CYEHNFVLQEV
					TAACTCTGCGCCGGGTCCCGACGGTGTATC		GGGC	LTHAKRSKRQA
					CTATCGTGACCTCCGTGGGGCGGACCGGGG		TTGC	VVAWLDLSNAF
					AGCGCGGCTCCTCACGGCGCTCTACAACAT		CCCC	GSIPHATIRRALI
					CTGCCTGCGGCTCGAGGCAGTCCCCGCGTC		CAGC	RSAVPRGLIAIW
					CTGGAAGACCTCCAACACTGTGTTGATACA		CCCTT	DSMYDGCTTRV
					CAAGAAAGGAGACCGGGGCATGTTGGAGA		AGTT	RTAEGHTAPIPIR
					ACTGGCGCCCTCTCGCTCTGGGGGACACCG		TTAA	SGVRQGCPLSPII
					TCCCCAAACTCTTCGCCGCGCTCTTGGCCGA		TTGC	FNLAIDSVVRVA
					CCGATTGACCGACTGGGCGGTCACCCGCGG		CTAT	AEXNDGYSLHG
					GAAGCTCTGCTCCGCGCAGAAGGGCTTCCT		GCGG	NTWSALAYADD
					GCGGGACGAGGGGTGCTACGAGCACAACTT		GGGG	IALLAQTPEGME
					CGTCCTGCAGGAGGTCCTGACGCACGCCAA		GGCT	RMLASVEAEAA
					GCGCTCTAAGCGCCAGGCGGTCGTCGCGTG		TTTGT	SVGLRFNPAKC
					GCTGGACCTGTCCAACGCGTTTGGATCGAT		CCCC	ATLHVGAGNGG
					CCCGCACGCGACGATCCGCCGCGCGCTTAT		CGCA	RVLPTSFQIQGE
					AAGATCCGCGGTGCCACGGGGTCTCATAGC		AATG	TINPLAQGESYT
					GATATGGGACTCCATGTACGATGGTTGCAC		TATA	HLGVPTGFSVD

					GACGAGGGTGCGAACCGCCGAGGGTCACAC		TATA	QTPYAAVGDIVS
					AGCACCCATCCCCATCCGGTCGGGCGTCCG		TATA	DLRAVDRSLLA
					TCAGGGTTGTCCGCTAAGCCCTATTATCTTC		TATTT	PWQKIEMLGTFI
					AACCTGGCCATCGACTCGGTCGTCCGTGTG		AGCG	LSRLDFLLRGAR
					GCGGCCGAGWCGAATGACGGGTATTCCCTC		CGCG	VFKGPLTAVDL
					CACGGAAATACCTGGTCGGCATTGGCTTAC		GCTT	NIRRHVKSWLN
					GCGGACGACATCGCACTACTGGCCCAGACG		AGCC	LPQRASAEGVY
					CCCGAGGGGATGGAGAGGATGCTAGCCTCT		GCTTT	MPPRWGGCGLL
					GTGGAGGCGGAGGCAGCGTCGGTGGGGCTG		TGTTT	PLSDLADVLTV
					CGGTTCAACCCTGCAAAGTGTGCCACCCTG		GTAT	AHAYRMLTVRD
					CACGTCGGTGCGGGGAATGGCGGCAGGGTC		TACC	GAVRELAWESL
					CTACCGACGTCATTCCAAATCCAGGGGGAG		CCAG	RGVVGRRIGHA
					ACGATCAACCCCCTGGCTCAGGGTGAGTCG		AGGG	PSCEDIASFLSGS
					TACACCCACCTTGGCGTTCCAACGGGGTTCT		GAAT	LDGRMRGGGEA
					CCGTGGACCAGACGCCCTACGCCGCCGTCG		TGTC	SLWSSARNAAL
					GGGACATCGTCTCGGACCTGCGCGCTGTCG		CCTCT	RQSERLSLRWR
					ACCGCTCACTCCTTGCCCCGTGGCAGAAGA		GGGG	WVEATEEMTLE
					TAGAAATGCTGGGGACCTTCATCCTATCCA		AAAA	CRGPRGAAIKIP
					GGCTTGACTTTCTGCTCCGGGGGGCCAGAG		AAAA	PEARGQVVNRL
					TGTTCAAGGGTCCCCTCACGGCCGTGGACC		TGAT	RSAVAEHYASR
					TTAACATCCGWAGGCATGTTAAATCCTGGC		TGGA	LLSKPDQGKVFE
					TTAACCTCCCTCAGCGAGCAAGCGCGGAGG		AAAA	VSSRSRVSNHFI
					GAGTCTACATGCCGCCCCGTTGGGGGGGAT		TAAA	RGGSFTRFADW

					GTGGACTCCTGCCGCTCTCTGACCTCGCCGA CGTCCTCACGGTTGCCCACGCGTACCGTATG TTAACGGTGCGCGATGGCGCCGTGAGGGAG TTGGCGTGGGAATCGCTGAGGGGAGTGGTT GGGCGCAGGATCGGCCACGCCCCTAGTTGC GAGGATATCGCCTCCTTCCTATCCGGCTCGC TGGATGGAAGGATGAGGGGCGGAGGGGAG GCTTCGCTCTGGTCGAGTGCGCGGAACGCT GCGCTCAGACAGTCCGAGAGGTTGTCCCTG CGTTGGCGGTGGGTCGAGGCCACGGAGGAG ATGACGTTGGAGTGTCGAGGGCCCAGGGGG GCAGCGATTAAGATTCCGCCTGAAGCGCGC GGTCAGGTAGTGAATCGGTTGCGCTCAGCT GTAGCAGAGCACTACGCAAGTAGGTTGCTT AGCAAGCCTGATCAGGGTAAGGTCTTCGAG GTGTCGTCGCGGAGCCGAGTGAGCAATCAC TTTATCCGCGGCGGCAGCTTCACTCGCTTCG CCGACTGGCGCTTTATCCATAAGGCCCGGTT AGATGTTCTTCCTCTCAACGGCGCACGACGT TGGGAGGCCAACGACAAGCGCTGTCGGCGA TGCGGTGAGGTATCGGAGACATTACCCCAT GTGCTCTGTCACTGCGGCATCCACTCCGCCG CGATACAGCTGAGGCACGACGCTGTCCTGC		GTGA GCTA A (SEQ ID NO: 1303)	RFIHKARLDVLP LNGARRWEAND KRCRRCGEVSE TLPHVLCHCGIH SAAIQLRHDAVL HRLWKATRLPG VVRVNQRVEGV SDELGALRPDLV VRHEPSKSVVIC DVTVPFENRWT AFEDARARKIA KYSPLAEELQRR GYRVVVTAFVV GALGSWDPRNE AVLRLLRVGNQ YAAMMRRLIVS DTIRWSRDIYVE HVSGTRQYLAP SRPSGDLATPPR AVRRRWLAEER SAQDAARRGSD SVSVA (SEQ ID NO: 1425)

					ACCGCCTTTGGAAGGCCACTCGCCTTCCAG GGGTAGTGCGGGTTAACCAGCGGGTGGAGG GCGTCAGCGACGAATTGGGGGCGCTACGAC CTGATCTCGTGGTCAGGCACGAGCCCTCCA AAAGTGTCGTCATCTGCGACGTCACGGTGC CATTCGAAAACCGCTGGACCGCTTTCGAGG ACGCCAGGGCGAGGAAAATCGCCAAATACT CGCCTCTGGCAGAGGAGCTACAGCGGCGAG GGTACCGTGTCGTCGTGACGGCCTTCGTCGT CGGCGCCCTCGGCTCGTGGGATCCGAGGAA TGAGGCGGTGTTGAGACTGCTGCGGGTTGG CAACCAGTATGCAGCTATGATGCGGCGCCT CATTGTCTCGGATACCATTCGCTGGTCACGC GACATATATGTGGAGCATGTGTCCGGCACC CGCCAGTACCTGGCTCCTTCCCGTCCCTCTG GGGATCTCGCGACGCCGCCGAGAGCGGTTC GTCGACGCTGGCTCGCCGAGGAGAGAAGCG CACAGGACGCGGCGCGTCGCGGTTCGGATA GTGTGAGTGTCGCGTAAATTATTTTGTCTTT GTCTTGGCCCCCCCTTTTTAAACCAAGCAGG AGAGAGTGGCCCAATGCCCAACTATTATAT ATTAACTATTTACTGTGATATTTATTATTTG ACTGTTGGGCGGGCCCCTCTCTGCTGGTTTT			

					ATTTATATATATTTTTTACTCGCGTACTTTTT GTACTACTCTATTTTTCTTTTTATTTTAGCTA TGCTATTTTTATCTCTTTCTTTGTCTCTATTT TCTTTCTTTTTTTCTTTCCTTTTCTTTTCTTTC TTTTATTCTTCTTTTATTTATCTTTTTTTCTTT TCTGTTGTGGGGCCCTGACCGTCCGAGTGTG AATGCCGCGAAAAACAATATTATGTTTTAT ACGAGTGTGCATGTGCGTGATATATTTATCT ATTTTATTTTATTTATTTATTATAATTTATTG CCGCGCGCGCTCCTCCGGGACTTTTATTCGT TGACAATACTGTGATATTTTTCTGCKCAGGC TGGGGGGGCTTGCCCCCCAGCCCCTTAGTTT TAATTGCCTATGCGGGGGGGGCTTTTGTCCC CCGCAAATGTATATATATATATATTTAGCGC GCGGCTTAGCCGCTTTTGTTTGTATTACCCC AGAGGGGAATTGTCCCTCTGGGGAAAAAAA ATGATTGGAAAAATAAAGTGAGCTAA (SEQ ID NO: 1057)			
NeSL	Utopia-1_AMi	.		Alligator mississippiensis	TGCTGGAAAGACGGAGAACCGCTTCCTTTT TCCCTGCGCCTGGCCTGGTATTGCAGTACCT CCAGGATTAGCGCCAACTAGTCCGGCAGAC TGTCGGAATACAGCAATAGAAAGWGAGCT GACTAGCAGCTTGCTTTCCTTCCTCCGGTGC	TGCTGG AAAGAC GGAGAA CCGCTT CCTTTTT	TGAA CCCC CCCT CTGC ACCA	CHHAGLRPGTP NRTRRPDQTAPL PDPRGHPMPPN RRGSRSRPEEPS RREPPXPRACQG

					AGCATGGGTTCTCGTCAGTCMTGACGGGCT	CCCTGC	GATG	LRVWSPPQQRM
					AGGGAAGGCGGTGCTGCCAGTACGTCCGAA	GCCTGG	GACC	PTPWQTLWLEE
					AGAGTGCCGGTTGCGCAAGCGACCGCGCCA	CCTGGT	TTCA	LSRATTFKAFEA
					CTCAGGTGAGTAGCCAAGGGTCTTACAGTT	ATTGCA	CTTC	SVARLTEELSAA
					CACCGGACCCGAWAACGCGAAAACCCCAA	GTACCT	GAGA	ARPGQPRGGNN
					CTCGGGCTAGTAGCCGAAGACCTGGGTCCC	CCAGGA	GGAT	RPATRRDHRLQ
					CCCCMGGTCAGAGTAGGCGAACGCCWGKG	TTAGCG	TCTTC	PQRRPRRQRYD
					CTCAGAGGACGGAACGCGGAAAACACCCCC	CCAACT	AGCA	PAAASRIQKLYR
					AGGTCCCAAGGACGCCCTGATCCACTGACA	AGTCCG	ATGG	ANRPKAVREILE
					AGAACGCTCGAGGCACGCCAGGAGACCCCC	GCAGAC	ACGA	GPSAFCQVPRET
					AGCTAGGGTGGACCGCCGACTGCAGGTCCG	TGTCGG	CCCC	LFNYFSRVFNPP
					GAGGACCCTCCCAGGAGGGTGGACCAGCGA	AATACA	GCTC	AEAAAPRPATV
					ACCCAAGTTGGCGACGAACCCTGACGCACC	GCAATA	CACC	EALTPVPPAEGF
					CCCCACGATGTCAGGACCCCGACAGGCGGC	GAAAG	CGAA	EDAFTPQEVEAR
					GGTGGACCACTGACCATCGACCGACCCCCA	WGAGCT	GAGG	LKRTRDTAPGR
					GAGGCAGAGAGACTCTCAGAGCCCGGAACC	GACTAG	ACCC	DGIRYSLLKKRD
					CCGGCTGACGAGAGCCGCCTCCCGGCGGAG	CAGCTT	CCGC	PGCLVLSVLFNR
					GACCCCGGAGCCTGAGGATGCCCCCCGGAT	GCTTTC	GATG	CREFRRTPTTW
					GACGGCGGAGCGCCCCGAGCGACAGCGGA	CTTCCT	AGAC	KRAMTVLIHKK
					CCCCTCCGGACCCCCACGGCCCCTCGGTGA	CCGGTG	TCTAT	GDPTDPGNWRPI
					CGATGGCGGGCCCCGAACGACGACGACCCC	CAGCAT	ATGG	ALCSTVAKLYA
					CGGACCCCGGCGGTCCCGAGGACGCCCCCC	GGGTTC	ACTG	SCLAARITDWA
					CCGAGGGTCTCCCCACGCTGGTGGAGGAGC	TCGTCA	AGAC	VTGGAVSRSQK

					CCCGGACCCCCCCGACACCGGACCCCCCCA	GTCMTG	ACTTT	GFMSTEGCYEH
					CGGACGACCCAGGCGAAGGCGTAGACATG	ACGGGC	TTCTT	NFTLQMALDNA
					ACAGCACTCACGTTCCTCCCCTTCCCCCTCC	TAGGGA	CGAA	RRTRKQCAVAW
					CGGCGAAGCTGTTCTGCCCGACCTGCCACC	AGGCGG	CCAC	LDISNAFGSVPH
					CGCCAAGACAGTACAGGTCGCACGGCGACA	TGCTGC	TTCCT	RHIFGTLRELGL
					TGAACAAGCACCTACGGCGCTTCCACCAGC	CAGTAC	CCAC	PDGVIDLVRELY
					TGCGCCTAGCCTTCTACTGCGCCCTCTGCGG	GTCCGA	CATT	HGCTTTVRATD
					CACCGAGTACGAGGCCCTGAAGCTCCTGAA	AAGAGT	GCGG	GETAEIPIRSGVR
					GAACCACCAGAAGGGATGCGAGGGCCACG	GCCGGT	ACCA	QGCPLSPIIFNLA
					GAGCCGAGAGGAGACCCGGCACGCTGGTG	TGCGCA	TTGT	MEPLLRAVAGG
					AGGTCCGCTGCCCCGGCCCGCCGGACCCAG	AGCGAC	AACG	PGGLDLYGQKL
					GCCGCGGTGCGAAGGCCCGCCAGACTGGCC	CGCGCC	GGTT	SVLAYADDLVL
					ACCCCGCCGACAACCCCACCGGACCAGACC	ACTCAG	TGTG	LAPDATQLQQM
					TCCAGGGACCACCCGACGGAGAGACCTGCC	GTGAGT	TGTA	LDVTSEAARWM
					CCAGTGATGCCACCACGCAGGCCTCCGCCC	AGCCAA	TCTAT	GLRFNVAKCAS
					AGGGACCCCCAACCGGACACGACGCCCCGA	GGGTCT	CTTCT	LHIDGRQKSRVL
					CCAGACAGCCCCCCTCCCAGACCCCCGGGG	TACAGT	TTCTC	DSTLTIQGQAM
					CCACCCGATGCCCCCGAACCGCCGGGGATC	TCACCG	TCTC	RHLRDGEAYCH
					CCGGAGCCGCCCGGAGGAACCGAGCCGCCG	GACCCG	AGCG	LGTPTGHRAKQ
					GGAGCCCCCCGKCCCCCGAGCGTGCCAGGG	AWAAC	TCGC	TPEETINGIVQD
					TCTCCGGGTGTGGAGCCCTCCCCAGCAGAG	GCGAAA	GAAC	AHKLDSSLLAP
					GATGCCCACCCCATGGCAAACCCTCTGGCT	ACCCCA	CCCC	WQKIDAVNTFLI
					GGAGGAGCTCTCCCGGGCCACCACCTTCAA	ACTCGG	TCCCT	PRVAFVLRGSA

					GGCCTTCGAGGCCTCGGTGGCCCGGCTCAC	GCTAGT	CCCC	VPKTPLKKADA
					AGAGGAGCTCTCGGCGGCCGCCCGGCCCGG	AGCCGA	TTCCC	EIRRLLKKWLHL
					CCAGCCCCGGGGGGGCAACAACAGACCGG	AGACCT	CTCC	PLRASNEVLHIP
					CGACGCGACGGGACCACAGACTGCAGCCGC	GGGTCC	CCCT	YRQGGANVPRM
					AGAGGCGACCCAGGCGCCAGCGCTACGACC	CCCCCM	CCCC	GDLCDIAVVTH
					CGGCGGCAGCCTCCCGGATCCAGAAGCTGT	GGTCAG	CCCA	AFRLLTCPDXTV
					ACCGGGCCAACCGTCCCAAGGCGGTGAGAG	AGTAGG	CCCC	SIIAASALEETAR
					AGATCCTGGAGGGACCCTCGGCCTTCTGCC	CGAACG	CGGG	KRIGRQPTRRDL
					AGGTCCCCCGGGAGACTCTGTTCAACTACTT	CCWGK	CTTA	ATFLSGSLEGEF
					CAGCAGGGTCTTCAACCCCCCAGCAGAAGC	GCTCAG	GTTG	SRDGGDFASLW
					CGCCGCCCCACGCCCTGCGACCGTCGAAGC	AGGACG	GCTA	SRARNATRRLG
					GCTGACCCCCGTCCCCCCGGCGGAGGGGTT	GAACGC	ACAT	KRIGCAWTWTE
					CGAGGATGCCTTCACGCCGCAGGAAGTGGA	GGAAAA	TGTA	ERRELGVSLQPA
					AGCCCGCCTCAAGAGGACCAGGGACACCGC	CACCCC	TCTCC	PHADRVTVTPR
					CCCCGGCAGGGACGGCATCAGGTACAGTCT	CAGGTC	TGTA	TRTFLERFLKDA
					CCTCAAAAAGCGTGACCCGGGCTGCCTTGT	CCAAGG	ACCT	VRNKYAGDLRA
					TCTTTCTGTTCTCTTCAACAGGTGCAGAGAG	ACGCCC	AGTT	KPDQGKVFDVT
					TTCCGGCGCACGCCCACCACCTGGAAGAGG	TGATCC	GCGT	SKWDSSNHFMP
					GCCATGACGGTCCTCATCCACAAGAAGGGA	ACTGAC	TCCC	SGSFTRFADWRF
					GACCCGACCGACCCGGGCAACTGGAGACCC	AAGAAC	CTCCT	LHRARLNCLPL
					ATCGCCCTGTGCTCCACCGTCGCCAAGCTGT	GCTCGA	CACC	NGAVRFGHRDK
					ACGCCAGCTGCCTGGCGGCCCGCATCACCG	GGCACG	CCCA	RCRRCGYVAET
					ACTGGGCGGTGACCGGCGGGGCCGTCAGCC	CCAGGA	TCCCT	LPHVLCSCKPHA

					GGAGCCAGAAGGGCTTCATGTCGACGGAGG	GACCCC	CTATT	RAWQLCHNAV
					GCTGCTACGAACACAACTTCACCCTCCAGA	CAGCTA	GTTA	QDRLVRAIPAA
					TGGCCCTGGACAATGCCCGGAGGACCAGGA	GGGTGG	GTCC	AGEISVNRTVPG
					AGCAGTGCGCGGTGGCGTGGCTGGACATCT	ACCGCC	CTCG	CESQMRPDIVIT
					CCAATGCCTTCGGCTCCGTGCCCCACCGCCA	GACTGC	CTCG	NEEAKKVVIVD
					CATCTTCGGCACCCTCCGCGAGCTGGGCCT	AGGTCC	GGCG	VTIPFENRRQAF
					ACCGGACGGCGTCATCGACCTGGTGCGAGA	GGAGGA	CTCT	TDARARKREKY
					GCTCTACCACGGCTGCACCACGACCGTCCG	CCCTCC	GTAT	APLADILRGRGY
					CGCCACCGACGGAGAGACCGCGGAGATCCC	CAGGAG	TTCCC	DVTVDALIVGTL
					CATCCGGTCGGGGGTGAGACAGGGCTGCCC	GGTGGA	TACC	GAWDPSNESVL
					CCTCAGCCCCATCATCTTCAACCTGGCCATG	CCAGCG	GGCT	HACRVSRRYAK
					GAACCGCTCCTTCGAGCCGTGGCAGGCGGC	AACCCA	TTGTC	LMRCLMVSDTI
					CCCGGCGGGCTCGACCTGTACGGCCAGAAG	AGTTGG	ATCTT	RWSRDIYVEHIT
					TTGAGCGTCCTGGCCTACGCCGACGACCTC	CGACGA	TTTTG	GHRQYTDPTRR
					GTTCTCCTCGCCCCCGACGCCACCCAGCTGC	ACCCTG	GATT	TAAGPDPEGTA
					AGCAGATGCTGGACGTGACGTCCGAGGCGG	ACGCAC	CACA	(SEQ ID NO:
					CCAGGTGGATGGGCCTGCGCTTCAACGTCG	CCCCCA	ATCC	1426)
					CCAAGTGCGCCTCCCTGCACATCGACGGSA	CGATGT	TAAA	
					GGCAGAAGAGCCGCGTCCTGGACTCCACCC	CAGGAC	CATC	
					TCACGATCCAGGGCCAGGCGATGAGGCACC	CCCGAC	TACT	
					TGCGCGACGGCGAGGCCTACTGCCACCTGG	AGGCGG	AATA	
					GGACGCCCACCGGCCACCGGGCCAAGCAGA	CGGTGG	AAAG	
					CGCCGGAGGAGACCATCAACGGGATCGTGC	ACCACT	TCAA	

					AGGACGCCCACAAGCTGGACTCGTCCCTGC TGGCCCCCTGGCAGAAGATAGACGCGGTGA ACACCTTCCTCATCCCCCGCGTCGCGTTCGT CCTGAGAGGCTCGGCGGTCCCCAAGACCCC CCTCAAGAAGGCGGACGCCGAGATCCGGCG GCTGCTCAAGAAGTGGCTGCACCTGCCGCT GAGGGCCAGCAATGAGGTCCTGCACATCCC CTACCGGCAGGGAGGTGCCAACGTCCCCCG CATGGGAGACCTCTGCGACATCGCGGTGGT CACCCACGCCTTCCGCCTCCTGACCTGCCCG GACSCGACGGTAAGTATCATCGCCGCCAGC GCCCTCGAGGAGACCGCCCGCAAGAGGATC GGGAGGCAGCCCACCAGACGTGACTTGGCC ACCTTCCTCAGCGGCTCGCTGGAGGGCGAG TTCAGCAGAGACGGCGGGGACTTTGCCTCG CTGTGGAGCCGAGCCCGCAACGCCACGCGC CGCCTCGGGAAGCGCATCGGCTGCGCCTGG ACCTGGACCGAGGAGCGCCGGGAGCTGGG AGTCTCCCTGCAACCAGCCCCGCACGCCGA CCGCGTCACCGTGACGCCCCGCACGAGGAC CTTCCTGGAGAGGTTCCTGAAGGACGCCGT CCGAAACAAGTACGCCGGCGACCTGAGGGC CAAACCCGACCAGGGCAAGGTCTTCGACGT	GACCAT CGACCG ACCCCC AGAGGC AGAGAG ACTCTC AGAGCC CGGAAC CCCGGC TGACGA GAGCCG CCTCCC GGCGGA GGACCC CGGAGC CTGAGG ATGCCC CCCGGA TGACGG CGGAGC GCCCCG AGCGAC AGCGGA	TC (SEQ ID NO: 1304)	

					CACCTCGAAGTGGGACTCCAGCAACCACTT	CCCCTC		
					CATGCCCAGCGGGAGCTTCACGCGCTTCGC	CGGACC		
					GGACTGGCGCTTCCTCCACCGCGCCCGCCTC	CCCACG		
					AACTGCCTGCCTCTGAACGGGGCCGTCCGC	GCCCCT		
					TTCGGCCACCGGGACAAGAGGTGCCGACGG	CGGTGA		
					TGCGGCTACGTGGCAGAGACCCTCCCCCAC	CGATGG		
					GTGCTGTGCAGCTGCAAGCCGCACGCCAGA	CGGGCC		
					GCCTGGCAGCTCTGCCACAACGCTGTCCAG	CCGAAC		
					GACCGCCTGGTGAGGGCCATCCCGGCCGCA	GACGAC		
					GCGGGGGAGATCTCCGTGAACCGCACCGTC	GACCCC		
					CCGGGCTGCGAGAGCCAGATGCGCCCCGAC	CGGACC		
					ATCGTCATCACCAACGAGGAGGCCAAGAAG	CCGGCG		
					GTCGTGATCGTGGACGTGACCATCCCCTTCG	GTCCCG		
					AGAACCGGCGCCAAGCCTTCACCGACGCCC	AGGACG		
					GGGCTCGCAAGCGGGAGAAGTACGCCCCGC	CCCCCC		
					TGGCCGACATCCTGAGGGGCCGCGGCTACG	CCGAGG		
					ACGTGACGGTCGACGCGCTCATCGTGGGAA	GTCTCC		
					CGCTCGGAGCCTGGGACCCCAGCAACGAGA	CCACGC		
					GCGTCCTGCATGCCTGCCGCGTCTCCCGCCG	TGGTGG		
					CTACGCCAAGCTGATGCGCTGCCTCATGGT	AGGAGC		
					GTCCGACACCATCCGTTGGTCCCGTGACATC	CCCGGA		
					TACGTGGAGCACATCACGGGCCACCGCCAG	CCCCCC		
					TACACCGACCCCACCAGACGAACCGCCGCC	CGACAC		

					GGACCGGACCCAGAGGGGACCGCCTGAACC CCCCCTCTGCACCAGATGGACCTTCACTTCG AGAGGATTCTTCAGCAATGGACGACCCCGC TCCACCCGAAGAGGACCCCCGCGATGAGAC TCTATATGGACTGAGACACTTTTTCTTCGAA CCACTTCCTCCACCATTGCGGACCATTGTAA CGGGTTTGTGTGTATCTATCTTCTTTCTCTCT CAGCGTCGCGAACCCCCTCCCTCCCCTTCCC CTCCCCCTCCCCCCCCACCCCCGGGCTTAGTT GGCTAACATTGTATCTCCTGTAACCTAGTTG CGTTCCCCTCCTCACCCCCATCCCTCTATTG TTAGTCCCTCGCTCGGGCGCTCTGTATTTCC CTACCGGCTTTGTCATCTTTTTTGGATTCAC AATCCTAAACATCTACTAATAAAAGTCAAT C (SEQ ID NO: 1058)	CGGACC CCCCCA CGGACG ACCCAG GCGAAG GCGTAG ACATGA CAGCAC TCACGT TCCTCC CCTTCC CCCTCC CGGCGA AGCTGT TCTGCC CGACCT GCCACC CGCCAA GACAGT ACAGGT CGCACG GCGACA TGAACA		

						AGCACC TACGGC GCTTCC ACCAGC TGCGCC TAGCCT TCTACT GCGCCC TCTGCG GCACCG AGTACG AGGCCC TGAAGC TCCTGA AGAACC ACCAGA AGGGAT GCGAGG GCCACG GAGCCG AGAGGA GACCCG GCACGC		

						TGGTGA GGTCCG CTGCCC CGGCCC GCCGGA CCCAGG CCGCGG TGCGAA GGCCCG CCAGAC TGGCCA CCCCGC CGACAA CCCCAC CGGACC AGACCT CCAGGG ACCACC CGACGG AGAGAC CTGCCC CAGTGA		

						(SEQ ID NO: 1181)		
NeSL	Utopia-1_CMy	.		Chelonia mydas	CTCTTCTTATGAATACTTGCAACACCTGCAC	CTCTTC	TGAG	MTTKKVLGAST
					TGAAGATGGATTCTCCGGCTGCTATTTTTGA	TTATGA	CCGG	TLQTSSTKGKNS
					AAAACTGATGCTGCTTTGAAGGTGTATTCTG	ATACTT	TACG	GCSKDPLRDAV
					CTGCTGCTACCTTGGAAGGAAATTCTCTCTC	GCAACA	ACAT	PGRSWILRPACR
					TGCTCCTGAGACATCCCCAGCTGCACCGTGT	CCTGCA	CGTG	DITTRRNIPPAPQ
					ACCACCACCACCACTGCTGCTGCTCCACAG	CTGAAG	CATC	QQQPPMESPPTL
					AAGGTTTCTCGGACAATGACTACAAAGAAG	ATGGAT	AACT	QLQDALRRPSPT
					GTCCTAGGTGCCTCCACAACATTACAGACC	TCTCCG	ATGA	PAAAQVADAGG
					AGCAGCACGAAGGGGAAGAACAGTGGCTG	GCTGCT	GAAA	ALAALHTIKRGI
					CTCAAAGGACCCCCTCCGAGATGCTGTTCC	ATTTTT	GGGA	SVDWTSISPKXX
					AGGAAGATCCTGGATTCTGAGGCCGGCCTG	GAAAAA	CTGA	QRXTSASPDACP
					TCGGGACATCACAACCAGAAGGAACATCCC	CTGATG	GAGA	ASETTQRDXRX
					CCCCGCCCCCCAGCAGCAGCAGCCGCCAAT	CTGCTT	CTTTT	LLDARPAGPLDP
					GGAGAGCCCCCCCACTCTGCAGCTGCAAGA	TGAAGG	TCCA	TRPHQDEPASDT
					TGCTCTCAGGCGACCATCTCCCACCCCCGCA	TGTATT	TTGG	ADAAGTPLLQG
					GCTGCCCAGGTCGCTGACGCTGGTGGTGCT	CTGCTG	ACCA	NEDTIYLQYPLA
					CTCGCTGCTCTACACACCATCAAGAGAGGA	CTGCTA	TATG	ADMLICPICSPP
					ATCTCCGTAGACTGGACCAGCATCTCTCCA	CCTTGG	AACT	QSFHLLGVVTR
					AAGASCMCCCAGAGGSTCACCAGCGCCTCG	AAGGAA	GGAA	HLKRCHSKRVA
					CCGGATGCCTGCCCAGCCTCAGAAACCACC	ATTCTC	CCAT	FSCALCSLPFET
					CAGAGGGACCSCAGGWSCCTGCTGGACGCC	TCTCTG	AAAC	QKQCKMHQVA

					CGCCCAGCCGGACCTCTCGACCCTACCCGC	CTCCTG	TCAC	CRKCLKGTTQSP
					CCACACCAGGATGAACCAGCCAGCGATACC	AGACAT	TGAA	APAPSPPAARRP
					GCTGATGCTGCTGGAACCCCCCTGCTGCAA	CCCCAG	CATT	AAPEPQRRKXTS
					GGTAATGAGGACACCATCTACCTGCAGTAT	CTGCAC	AAAT	QAAVKKPAPVA
					CCCCTCGCTGCGGACATGCTCATCTGCCCCA	CGTGTA	CTCA	RPAERDAAIEKV
					TCTGCTCTCCGCCCCAAAGCTTCCACCTCCT	CCACCA	CCAA	PAASGNITQVLA
					CGGTGTCGTCACCAGGCACCTGAAGAGATG	CCACCA	ATGA	SRRPVSPSHVAK
					CCACAGCAAGCGGGTTGCCTTCAGCTGTGC	CTGCTG	GGGT	XISMLRRLSAAS
					CCTCTGCAGCCTGCCCTTCGAGACGCAGAA	CTGCTC	AAAT	PPVQHVPVPRRI
					GCAATGCAAGATGCACCAAGTCGCCTGCAG	CACAGA	CCAT	SAPPRIAARDPV
					GAAATGCCTCAAGGGAACMACACAGTCTCC	AGGTTT	CCTC	AGRASAAPQTA
					TGCCCCGGCTCCCAGCCCTCCTGCTGCACGC	CTCGGA	ATCA	LRTPAAGGASTT
					CGGCCCGCTGCTCCTGAGCCTCAACGAAGA	CA (SEQ	TCGT	PQTALRTPTAGG
					AAGSCGACCTCGCAAGCTGCCGTCAAGAAG	ID NO:	ATCC	ASAMPQTTLPXP
					CCTGCCCCCGTCGCCAGGCCAGCGGAACGG	1182)	ACTC	RRPDWRNQPRS
					GATGCTGCGATCGAGAAGGTACCTGCTGCC		ATTA	HSKAPGLHRQT
					TCGGGGAACATCACCCAGGTCCTCGCCAGC		TACT	DQHGPQVHSAG
					AGGAGGCCCGTCTCACCCTCTCATGTCGCC		CCAC	HCLREISRSSSN
					AAGMAGATCTCCATGCTGAGACGACTCAGT		ACCT	RLGSSHSAAAT
					GCTGCCTCGCCACCTGTCCAGCACGTCCCCG		GAAC	HRRTGGVPATP
					TCCCCAGAAGGATCAGCGCCCCACCGCGCA		ATAG	EPDRVSPTTSNA
					TAGCTGCTCGAGATCCTGTCGCCGGAAGAG		CCAT	XIPPEIPPQHPTE
					CCAGCGCCGCCCCTCAGACCGCCCTGCGAA		TATA	GNPDPRDRRQA

					CTCCAGCCGCCGGAGGAGCCAGCACCACGC		TGAA	DHTAGSEPAPDE
					CTCAGACCGCCCTGCGAACTCCAACCGCCG		CAAC	VEDXEGQRPMV
					GAGGAGCCAGTGCCATGCCTCAGACCACCC		ATAC	RAATPWQTAWT
					TGCCAGMCCCCAGACGTCCAGACTGGAGGA		CCCC	EELQAAASFDDF
					ACCAGCCCCGCAGCCACAGCAAAGCACCGG		ATAT	DLLVDRLTRELS
					GCCTTCATCGCCAGACGGACCAGCACGGCC		CTCA	AEIAPRRSSNQE
					CCCAAGTCCATTCTGCGGGACACTGCCTAC		ATGT	NAPPAHRTPAPN
					GGGAGATCTCACGCTCCAGCAGCAACCGCC		CTGT	HNTTTRGARSR
					TAGGCAGCAGCCACTCATAGAAGGACCGGC		ACTTT	DASRRYDPAAA
					GGTGTCCCAGCAACCCCCGAGCCGGACCGC		GACC	SRIQKLYRANRS
					GTCTCTCCGACCACCAGCAACGCCASCATCC		CGTT	KAMREILDGPSP
					CGCCAGAGATCCCGCCCCAGCACCCAACCG		AACC	YCTIPSERLYSY
					AAGGGAATCCTGACCCACGAGATAGACGGC		TTTTA	FKDVFDRIARND
					AGGCCGACCATACAGCAGGCTCWGAGCCT		CCCC	AQRPECLRPLPR
					GCACCAGACGAGGTCGAGGACCMTGAGGG		CAAT	VDEAGVLETDX
					CCAGCGGCCGATGGTGAGGGCTGCCACNCC		CGGG	TPKEVMARLSK
					GTGGCAGACTGCCTGGACCGAGGAGCTACA		GATA	TKNTAPGKDGIP
					AGCGGCAGCTTCCTTTGACGACTTCGACCTC		TTGC	YSLLKKRDPGC
					CTCGTAGACAGGCTCACCCGAGAACTGTCT		AGAT	LVLATLFNQCK
					GCGGAAATCGCTCCCAGGAGGAGTTCGAAC		TATG	RFCRTPSSWKK
					CAGGAGAACGCCCCGCCTGCCCACAGAACG		TATTC	AMTVLVYKKGE
					CCTGCTCCGAACCACAACACCACCACCAGG		CTTA	RDDPSNWRPISL
					GGAGCCAGAAGTAGAGACGCCAGCCGCCG		CGCC	CSTMYKLYASC

					CTACGATCCAGCAGCGGCTTCAAGGATCCA		ACCC	LASRITEWSVSG
					AAAGCTGTACCGGGCAAACCGCTCCAAGGC		GATC	GAISSIQKGFMS
					CATGAGGGAGATCCTAGACGGGCCCTCGCC		CTAA	CEGCYEHNFVL
					CTACTGCACGATCCCATCTGAGCGTCTCTAC		ACCG	QTTIETARRARR
					AGCTACTTCAAGGATGTATTCGACCGCATA		AATT	QCAVAWLDLA
					GCCCGGAATGACGCGCAGCGCCCAGAGTGC		TCGC	NAFGSMPHHHIF
					CTCCGCCCCCTGCCCCGTGTCGACGAAGCA		ACCC	ATLQEFGMPEN
					GGTGTCCTGGAAACTGACTWTACGCCCAAG		CTTG	FLRVIREVYEGC
					GAAGTGATGGCCAGACTCTCAAAAACAAAA		ATAA	STTIRSVEGETA
					AACACAGCTCCTGGGAAAGACGGCATCCCC		TCTGT	EIPIRSGVKQGC
					TACAGCCTCCTGAAAAAGCGAGATCCCGGC		ACCT	PLSPIIFNLAMEP
					TGCCTGGTCCTCGCCACGCTCTTCAACCAGT		TATTC	LLRAISNGTDGF
					GCAAGCGATTCTGCCGGACTCCCAGCTCCT		CCTG	NLHGERVSVLA
					GGAAGAAGGCCATGACGGTACTGGTGTACA		ATAA	YADDLVLTADD
					AGAAGGGCGAGCGGGATGACCCCAGCAAC		CCAG	PESLQGMLDAT
					TGGAGGCCCATCTCCCTCTGCTCCACGATGT		AAAC	SRAADWMGLRF
					ACAAGCTCTATGCCAGCTGCCTGGCGTCGA		TTCTA	NAKKCATLHID
					GGATCACGGAGTGGTCGGTGAGCGGGGGA		TGCTT	GSKRDSVQTTG
					GCCATCAGCTCCATCCAGAAAGGCTTCATG		AAAC	FQIQGEPVIPLAE
					TCCTGCGAGGGCTGCTACGAACACAACTTC		TCTGT	GQAYQHLGTPT
					GTCCTCCAAACCACCATCGAAACGGCCAGA		ACCG	GFRVRQTPEDTI
					AGGGCGCGGAGGCAGTGCGCGGTAGCGTG		TTTTT	QEILQDAAKIDA
					GCTCGACCTGGCTAACGCCTTTGGGTCCATG		TTTTA	SLLAPWQKINAL

					CCCCACCACCACATCTTTGCCACGCTCCAGG AGTTTGGGATGCCAGAGAACTTCCTTCGTGT GATCCGAGAGGTGTACGAGGGATGCAGCAC CACCATTCGCTCGGTCGAAGGGGAGACCGC CGAGATCCCGATCCGGAGCGGAGTTAAGCA GGGCTGTCCCCTCAGCCCCATCATCTTTAAC CTCGCCATGGAGCCGTTGCTGCGAGCGATC TCCAATGGCACAGATGGCTTCAACCTCCAC GGTGAGAGGGTGAGCGTCCTGGCTTACGCG GATGACCTGGTCCTGACCGCGGATGACCCA GAGAGCCTCCAAGGTATGCTAGATGCCACC AGTCGAGCTGCCGACTGGATGGGGCTCCGC TTCAATGCAAAGAAGTGCGCAACTCTCCAC ATCGACGGCAGCAAAAGGGACTCGGTGCAG ACGACGGGGTTCCAGATCCAGGGCGAGCCC GTCATCCCCCTGGCAGAGGGGCAGGCGTAC CAGCACCTCGGCACGCCGACGGGTTTCCGT GTCCGGCAGACACCCGAGGACACCATCCAG GAGATCTTGCAGGATGCCGCCAAGATCGAC GCCTCCCTGCTAGCACCGTGGCAGAAGATA AACGCCCTGAACACCTTCCTGATCCCCCGC ATCTCGTTCGTCCTAAGGGGATCCGCCGTG GCGAAGGTACCCCTCAACAAGGCAGACAAG		TTTCA ACAT CATC TTAA TAAA ATTA TTAA A (SEQ ID NO: 1305)	NTFLIPRISFVLR GSAVAKVPLNK ADKIVRQLVKK WLFLPQRASNE LVYIAHRHGGA NVPRMGDLCDI AVITHAFRLLTC PDAMVRNIAAN ALHDATKKRIG RAPSNQDIATFL SGSLDGEFGRD GRDIASLWSRA RNATRRLGKRIG CRWEWCEERQE LGVLVPQIRSND NTIVTPSARGML ERTLKAAIHSLY VETLKRKPDQG KAFELTSKWDA SQPLPRRGRLHP FRRLAVHPPCPA QLRPAQRSRPPR EPRQALQEVRLL

					ATCGTCCGGCAGCTGGTGAAGAAGTGGCTG			QRDPAPRPVQL
					TTCCTTCCCCAGAGAGCCAGCAACGAGCTG			QAPLQSLAAAP
					GTCTACATCGCCCACAGGCATGGCGGTGCC			QCHPEPPGESHR
					AACGTCCCCCGCATGGGCGACCTGTGTGAC			TAPGGGRRELR
					ATCGCGGTGATCACCCACGCCTTCCGCCTGC			HPRYPASGTPAN
					TGACGTGTCCCGACGCCATGGTAAGGAACA			HFLAGGGFTRF
					TCGCGGCAAACGCCCTCCATGACGCGACAA			ADWRFIHRARL
					AGAAGCGGATCGGCAGAGCCCCCTCCAACC			NCVPLNGAVRH
					AAGACATCGCCACCTTCCTGAGCGGTTCCCT			GNRDKRCRKCG
					GGATGGCGAATTCGGACGGGACGGGCGCG			YSNETLPHVLCS
					ACATCGCTTCACTGTGGTCCCGCGCTCGCAA			CKPHSRAWQLR
					CGCCACGCGTCGCCTGGGGAAGCGCATCGG			HNAIQNRLVKAI
					CTGCCGCTGGGAGTGGTGCGAGGAGCGCCA			APRLGEVAVNC
					GGAGCTGGGAGTCCTGGTGCCGCAGATCAG			AIPGTDSQLRPD
					GTCCAACGACAACACCATCGTCACCCCGAG			VVVTDEAQKKII
					CGCCAGGGGCATGCTGGAGAGGACCCTGAA			LVDVTVSFENRT
					GGCAGCCATCCACTCACTGTACGTGGAAAC			PAFREARARKLE
					CCTGAAGCGTAAACCGGACCAGGGTAAAGC			KYAPLADTLRA
					CTTCGAACTGACCAGCAAGTGGGACGCCAG			KGYEVQMDALI
					CCAACCACTTCCTCGCCGGGGGCGGCTTCA			VGALGAWDPCN
					CCCGTTTCGCCGACTGGCGGTTCATCCACCG			ERVLRTCGIGRR
					TGCCCGGCTCAACTGCGTCCCGCTCAACGG			YARLMRRLMVS
					AGCCGTCCGCCACGGGAACCGAGACAAGCG			DTIRWSRDIYIE

					TTGCAGGAAGTGCGGCTACTCCAACGAGAC CCTGCCCCACGTCCTGTGCAGCTGCAAGCC CCACTCCAGAGCCTGGCAGCTGCGCCACAA TGCCATCCAGAACCGCCTGGTGAAAGCCAT CGCACCGCGCCTGGGGGAGGTCGCCGTGAA CTGCGCCATCCCCGGTACTGACAGCCAGTT GCGACCTGACGTGGTAGTCACCGACGAGGC CCAGAAAAAGATCATCCTCGTCGACGTCAC GGTCTCCTTTGAGAACAGGACCCCGGCCTT CCGCGAAGCCCGAGCTCGTAAGCTGGAAAA ATACGCCCCCCTGGCCGACACCCTGAGAGC GAAGGGCTACGAGGTGCAGATGGATGCCCT GATCGTCGGAGCCCTGGGCGCTTGGGACCC CTGCAATGAGCGTGTGCTGCGGACCTGTGG GATCGGTCGACGCTACGCACGGCTCATGCG GCGCCTCATGGTCTCGGACACCATCCGATG GTCCAGGGACATCTACATCGAACACATCAC CGGCCACCGACAGTACCAGGAGGTGTGAGC CGGTACGACATCGTGCATCAACTATGAGAA AGGGACTGAGAGACTTTTTCCATTGGACCA TATGAACTGGAACCATAAACTCACTGAACA TTAAATCTCACCAAATGAGGGTAAATCCAT CCTCATCATCGTATCCACTCATTATACTCCA			HITGHRQYQEV (SEQ ID NO: 1427)

					CACCTGAACATAGCCATTATATGAACAACA TACCCCCATATCTCAATGTCTGTACTTTGAC CCGTTAACCTTTTACCCCCAATCGGGGATAT TGCAGATTATGTATTCCTTACGCCACCCGAT CCTAAACCGAATTTCGCACCCCTTGATAATC TGTACCTTATTCCCTGATAACCAGAAACTTC TATGCTTAAACTCTGTACCGTTTTTTTTTATT TCAACATCATCTTAATAAAATTATTAAA (SEQ ID NO: 1059)			
NeSL	Utopia-1_CPB	.		Chrysemys picta bellii	TTTTTTCTGATGCTTGACTGCAAACACCCAT CCAGAAGATGGAATCTCCTGCAGCCATTTTT GAAAAAATTGATGCTGCTTTAAAGATATAC TCCATTCTCCTWKTTTGKAAGAAAACTCTTT TTCAGCTTCAGCTATTCTGTCATCGGCTGCT GCTGTTCCTGCTTCCCAGAAAGCTCAGCMA AAACCTATCCTGAAGACCWCCCTTGGTGCC TCACGGAAGACCCGGASCACCTGCAAGAAC CAAAACATTAGGAGCTGGCTGAAGAAACCC CCCGTGGATACCTCWGCAGGGAGACCTGGS TCCAGMAGGACAKCTCTTCGGGACCTCMCA TCSAGGAGCAAGAATATCTCAACAGCTCTT CAGGAGGGGGACCCCCGGAGAACCCTGCCC GCTTCCCAGAACCAGGATGCTGATCGCCGC	TTTTTTC TGATGC TTGACT GCAAAC ACCCAT CCAGAA GATGGA ATCTCC TGCAGC CATTTT TGAAAA AATTGA TGCTGC TTTAAA	TGAG CCAG AGTG ACAT CGTT CTCC CACT ACGA GAAA GGGA CCAA GTGA CCTTC TCCG	MTQDQDADCCP AGKDATRGAPP MTQDQDADRCP AAPERDAPEGTT SSTPDPKTTYHP AVRRRAARRGM HLRAQDLDAAR CPSGQRDNVAS ESSAPPRATSPP QASLPDPEESPG ESAGTTEIRPTE GEAGEEDRIYLQ YPLPTGLLLCPF CLPVHGVQTLA

					CCCACCGGGAAGGATGCCACCGCAGGAGCC	GATATA	TTGG	ALSKHVRKTYN
					CCCCCAATGACCCAGGACCAGGATGCTGAT	CTCCAT	ATCA	KRIAFRCSRCDL
					TGCTGCCCCGCCGGGAAGGATGCCACCAGA	TCTCCT	TATG	PFETQKKCKFH
					GGAGCCCCCCCGATGACCCAGGACCAGGAT	WKTTTG	AACT	QATCRGPPTTA
					GCTGATCGCTGCCCCGCTGCTCCAGAGAGG	KAAGAA	GGAA	KVNPTDILRVPT
					GATGCTCCGGAAGGAACCACCTCCTCAACC	AACTCT	CCAT	LTPTDDLASAPQ
					CCAGACCCCAAAACTACTTACCACCCGGCT	TTTTCA	AAAC	PASPESQQIRGD
					GTCCGGAGGAGGGCCGCTCGAAGGGGAAT	GCTTCA	TCCCT	QPPTEGSVTPAS
					GCACCTCAGAGCCCAAGATCTCGATGCCGC	GCTATT	GAAC	RTDDATKRTSP
					ACGCTGCCCTTCCGGGCAAAGAGACAACGT	CTGTCA	ATTA	VSRIPTLDPAVR
					GGCCAGTGAGTCCAGCGCCCCCCCAAGAGC	TCGGCT	AATC	GTTATSQVNNL
					GACTTCACCTCCCCAAGCTTCTCTACCAGAC	GCTGCT	TCAC	TRRLSDLIKTIRH
					CCAGAGGAATCACCTGGCGAGTCTGCAGGC	GTTCCT	CAAA	NTDTRRCSAPPQ
					ACAACAGAGATCCGCCCCACGGAGGGTGAG	GCTTCC	TGAG	VTSCRPAVGATS
					GCAGGGGAAGAAGACCGCATCTACCTCCAG	CAGAAA	GGTC	IVPQAARRDPAN
					TACCCGCTCCCTACAGGTCTCCTCCTCTGCC	GCTCAG	AATC	GGASRSPQIPQP
					CCTTCTGTCTCCCCGTCCATGGAGTCCAGAC	CMAAA	CATC	DPAPGRPNTSSK
					CCTCGCGGCCCTCAGCAAACACGTTCGTAA	ACCTAT	CTCA	VTQRASDRQKP
					GACCTACAACAAACGGATTGCTTTCCGGTG	CCTGAA	TCAT	HAPPRTHQPDA
					TAGCCGCTGCGATCTCCCCTTCGAGACCCA	GACCWC	CATA	ARRRTRTIPSAS
					AAAGAAATGTAAGTTTCATCAAGCCACGTG	CCTTGG	TCCA	KHDRAPTKPST
					CAGGGGACCCCCCACGACCGCGAAAGTGAA	TGCCTC	CTCA	GASRTPLPPGRS
					TCCCACTGACATCCTCCGGGTTCCAACCCTG	ACGGAA	TTAT	SAASETPRAALP

					ACCCCCACCGATGATCTGGCTTCAGCACCC	GACCCG	AMTC	TTPGPPPQDPPE
					CAGCCAGCATCCCCAGAGTCACAGCAGATA	GASCAC	CACA	HRSTVRGTTRPQ
					AGGGGGGACCAACCGCCAACTGAGGGAAG	CTGCAA	CCCG	TVPAAPEPAETT
					CGTAACCCCCGCCTCGAGGACTGACGATGC	GAACCA	AACA	QQEERRPRARV
					CACCAAAAGGACCAGTCCCGTCTCCAGAAT	AAACAT	CAGC	ATPWQSAWME
					CCCCACGCTGGACCCTGCTGTGAGGGGGAC	TAGGAG	CACT	ELAKAEDFENF
					CACCGCCACCTCTCAGGTCAACAACCTCAC	CTGGCT	CTAT	DTLMDRLTAEL
					CAGACGCCTCAGCGACCTCATAAAAACCAT	GAAGAA	GAAC	SAEITARRREPQ
					CCGGCACAACACGGACACGAGACGCTGCAG	ACCCCC	TTCAT	EAARATRRFPAP
					CGCTCCCCCACAGGTAACCTCATGCCGCCCT	CGTGGA	ACCC	SRNNTAREGRR
					GCCGTAGGAGCAACTAGCATCGTCCCCCAG	TACCTC	TCAT	GDVGRRYDPAA
					GCTGCACGGCGAGATCCAGCCAACGGAGGA	WGCAG	ATCT	ASRIQKLYRMN
					GCCTCCCGTAGCCCCCAGATCCCACAGCCA	GGAGAC	CAAT	RTKAMREILDG
					GACCCCGCCCCCGGGAGACCCAACACCTCC	CTGGST	GTCT	TSSYCAIQPERL
					TCCAAGGTTACCCAAAGAGCCTCTGACCGC	CCAGMA	GTAC	YSYFKDVFDHE
					CAAAAACCCCATGCCCCACCGAGGACCCAC	GGACAK	TTTG	AQTNLRRPECLS
					CAGCCGGATGCCGCCCGCAGAAGAACCAGA	CTCTTC	ACCC	PLPRIDLTEDLE
					ACCATCCCCAGCGCTTCCAAACACGACCGC	GGGACC	ATCA	RDFSPQEVQARL
					GCCCCGACAAAGCCCAGCACCGGTGCTTCC	TCMCAT	ACCT	SRTKNTAPGKD
					AGAACCCCACTCCCTCCCGGAAGATCCAGT	CSAGGA	TTTAC	GIRYPLLKKRDP
					GCTGCCTCGGAGACACCGAGAGCTGCCCTC	GCAAGA	CCCC	GCLVLAAIFNKC
					CCCACCACACCAGGACCCCCGCCTCAAGAC	ATATCT	AATC	KQFHRVPRSWK
					CCACCTGAACACCGCTCCACAGTCCGAGGG	CAACAG	GGGG	KSMTVLIHKKG

					ACAACAAGGCCGCAAACCGTCCCCGCAGCA	CTCTTC	ATAT	XRDDPGNWRPI
					CCTGAACCTGCAGAGACAACGCAGCAGGAG	AGGAGG	TGCA	SLCSTIYKLYAS
					GAGCGACGGCCACGAGCGAGGGTCGCCAC	GGGACC	GATT	CLAARITDWSV
					GCCGTGGCAATCCGCCTGGATGGAGGAGCT	CCCGGA	ATGT	CGGAVSSVQKG
					GGCAAAGGCTGAGGACTTTGAGAACTTCGA	GAACCC	ATTC	FMSCEGCYEHN
					CACCCTGATGGACAGACTGACTGCAGAACT	TGCCCG	CTCA	FLLQTAIQEARR
					GTCTGCGGAAATTACGGCCAGAAGGAGGGA	CTTCCC	TGCC	SKRQCAVAWLD
					ACCCCAGGAAGCCGCACGGGCCACTCGCAG	AGAACC	ACCT	LTNAFGSIPHHH
					ATTCCCTGCGCCGAGCCGTAACAACACCGC	AGGATG	GATC	IFATLGEFGMPE
					CAGAGAAGGCAGGAGAGGGGACGTCGGCC	CTGATC	TTAA	TFIQILRDLYKD
					GCCGCTACGATCCGGCGGCTGCATCCCGTA	GCCGCC	ACCA	CTTTIRATDGET
					TTCAGAAACTATACAGGATGAACCGGACGA	CCACCG	AACT	DAIPIRRGVKQG
					AAGCCATGAGGGAGATCCTCGACGGGACCT	GGAAGG	TTGC	CPLSPIIFNLAME
					CCTCCTACTGTGCCATCCAGCCCGAGAGGC	ATGCCA	ACCC	PLIRAISSGPTGF
					TCTACTCCTACTTCAAGGATGTGTTTGATCA	CCGCAG	TCGA	DLHGKKLSILAY
					CGAGGCCCAGACCAACTTGCGACGCCCAGA	GAGCCC	TAAT	ADDLVLTADDP
					GTGCCTTTCCCCGCTACCCCGGATCGACCTC	CCCCA	CTGT	ESLQGMLDATS
					ACGGAGGACTTGGAGCGAGATTTTTCCCCG	(SEQ ID	ATGT	RATDWMGLRFN
					CAGGAGGTGCAGGCGAGGCTGTCGAGGACC	NO: 1183)	TATTC	AKKCATLHIDGS
					AAAAACACCGCCCCTGGAAAAGATGGCATC		CCTG	KRDSVQTTGFQI
					CGCTACCCCCTGCTGAAGAAGCGAGACCCC		ATAA	QGEPVIPLAEGQ
					GGCTGCTTGGTGCTCGCTGCCATCTTCAACA		CCAG	AYQHLGTPTGF
					AATGCAAGCAGTTCCATCGCGTTCCCCGCTC		AAAC	RVRQTPEDTIQE

					CTGGAAAAAGTCCATGACCGTGCTCATCCA		TTCTA	ILQDAAKIDASL
					CAAAAAAGGCGAMCGAGACGACCCCGGCA		TGCT	LAPWQKINALN
					ACTGGAGGCCCATCTCCCTCTGCTCCACCAT		CAAA	TFLIPRISFTLRG
					CTACAAGCTGTATGCCAGCTGCCTCGCGGC		CTCT	SAVAKVPLNKA
					AAGGATCACAGACTGGTCAGTGTGCGGGGG		GTTC	DKIIRKLVKKWL
					CGCCGTCAGCTCAGTGCAGAAGGGTTTCAT		ACTA	FLPQRASNELVY
					GTCCTGCGAGGGATGCTACGAGCACAACTT		TTTTT	IAHRHGGANVP
					CCTCCTTCAGACGGCCATCCAGGAGGCCAG		TTTA	RMGDLCDVAVI
					GAGGTCCAAGAGGCAGTGCGCAGTAGCATG		ACAT	THAFRLLTCPDA
					GCTTGACCTGACCAACGCCTTTGGGTCCATA		CATC	TVRNIAANALR
					CCCCACCATCACATCTTTGCCACCCTGGGAG		TTAA	DATEKRIGRAPS
					AGTTCGGGATGCCAGAAACCTTCATCCAGA		TAAA	NQDIATFLSGSL
					TCCTCCGGGACCTCTACAAGGACTGCACCA		ATTTT	DGEFGRDGRDI
					CCACCATCCGCGCCACGGACGGAGAGACGG		TAAA	ASLWSRTRNAT
					ACGCCATCCCCATCCGCCGCGGCGTGAAAC		TCTGT	RRLGKRIGCRW
					AAGGATGCCCCCTTAGCCCCATCATCTTCAA		T (SEQ	EWCEERQELGIR
					CCTGGCCATGGAACCGCTCATCCGAGCCAT		ID NO:	VPQIRSDDNTIV
					CTCCAGCGGCCCGACCGGCTTCGACCTGCA		1306)	TPTARGLLERTL
					CGGCAAGAAACTCAGCATTCTGGCCTACGC			KAAIRSLYVETL
					GGACGATCTGGTCCTGACCGCGGATGACCC			KRKPDQGKAFE
					AGAGAGCCTCCAAGGTATGCTAGATGCCAC			LTSKWDASNHF
					CAGCCGAGCTACTGACTGGATGGGGCTCCG			LDGGGFTRFAD
					CTTCAATGCGAAGAAGTGCGCAACTCTGCA			WRFIHRARLNC

					CATTGACGGCAGCAAAAGGGACTCGGTGCA GACAACGGGGTTCCAGATCCAGGGTGAGCC CGTCATCCCCCTGGCAGAGGGGCAGGCATA CCAGCACCTGGGCACGCCAACAGGGTTCCG TGTCCGGCAGACACCCGAGGACACCATCCA GGAGATCTTGCAGGACGCCGCCAAGATTGA TGCCTCCCTGCTGGCACCGTGGCAGAAGAT AAACGCCCTGAACACCTTCCTGATCCCACG CATCTCGTTCACCCTAAGGGGATCCGCCGT GGCGAAGGTGCCCCTCAACAAGGCAGACAA GATCATCCGGAAGCTGGTGAAGAAGTGGCT GTTCCTTCCCCAGAGAGCCAGCAACGAGCT GGTCTACATCGCCCACAGGCACGGCGGCGC CAACGTCCCCCGCATGGGTGACCTGTGCGA CGTCGCGGTGATCACCCACGCCTTCCGCCTG CTGACATGTCCCGACGCCACGGTGAGGAAC ATTGCGGCGAACGCCCTGCGTGATGCGACA GAGAAGCGGATCGGCAGAGCCCCCTCGAAC CAAGACATCGCCACCTTCCTGAGCGGCTCC CTGGATGGGGAATTCGGACGGGACGGGCGC GACATCGCTTCACTGTGGTCCCGCACTCGCA ACGCCACGCGTCGCCTGGGGAAGCGCATCG GCTGCCGCTGGGAGTGGTGCGAGGAGCGCC			VPLNGAVRHGN RDKRCRKCGYP NETLPHVLCSCK PHSRAWQLRHN AIQNRLVKAIAP RLGEISVNCTIA GTDSQLRPDVV VTDEAQKKIILV DVTVSFENRTPA FREARARKLEK YAPLADTLRAK GYEVQMDALIV GALGAWDPCNE RVLRTCGIGRRY ARLMRRLMVSD AIRWSRDIYIEHI TGHRQYQEA (SEQ ID NO: 1428)

					AGGAGCTGGGAATCCGGGTGCCGCAGATCA GGTCCGACGACAACACCATCGTCACCCCGA CGGCCAGGGGCTTGCTGGAGAGGACTCTGA AGGCCGCCATCCGCTCGCTGTACGTGGAAA CCCTGAAGCGTAAACCGGACCAGGGTAAAG CCTTTGAGTTGACCAGCAAGTGGGACGCCA GCAACCACTTCCTCGACGGGGGCGGCTTCA CCCGTTTCGCCGACTGGCGGTTCATCCACCG TGCCCGGCTCAACTGCGTCCCGCTCAACGG AGCCGTCCGCCACGGGAACCGAGACAAGCG TTGCAGGAAGTGCGGCTACCCCAACGAGAC CCTGCCCCACGTCCTGTGCAGCTGCAAACC CCACTCCAGAGCCTGGCAGCTGCGCCACAA CGCCATCCAGAACCGCCTGGTGAAAGCCAT CGCGCCACGCCTGGGGGAGATCTCCGTGAA CTGCACCATCGCCGGTACCGACAGCCAGCT ACGACCTGACGTGGTCGTCACCGACGAGGC CCAGAAAAAGATCATCCTCGTCGACGTCAC GGTCTCCTTTGAGAACAGGACCCCGGCATT TCGCGAAGCCCGAGCTCGTAAGCTGGAAAA GTACGCCCCCCTGGCTGACACCCTGAGAGC GAAGGGCTATGAGGTGCAGATGGACGCCCT GATTGTCGGAGCCCTGGGCGCCTGGGACCC			

					CTGCAACGAGCGTGTGCTGCGGACCTGCGG GATCGGTCGACGCTACGCACGTCTCATGCG GCGCCTCATGGTCTCAGACGCCATCCGATG GTCCAGGGACATCTACATCGAGCACATCAC CGGCCACCGACAGTACCAGGAGGCGTGAGC CAGAGTGACATCGTTCTCCCACTACGAGAA AGGGACCAAGTGACCTTCTCCGTTGGATCA TATGAACTGGAACCATAAACTCCCTGAACA TTAAATCTCACCAAATGAGGGTCAATCCAT CCTCATCATCATATCCACTCATTATAMTCCA CACCCGAACACAGCCACTCTATGAACTTCA TACCCTCATATCTCAATGTCTGTACTTTGAC CCATCAACCTTTTACCCCCAATCGGGGATAT TGCAGATTATGTATTCCTCATGCCACCTGAT CTTAAACCAAACTTTGCACCCTCGATAATCT GTATGTTATTCCCTGATAACCAGAAACTTCT ATGCTCAAACTCTGTTCACTATTTTTTTTAA CATCATCTTAATAAAATTTTTAAATCTGTT (SEQ ID NO: 1060)			
NeSL	Utopia-1_DYak	.		Drosophila yakuba	AAAGTGTAGTTCTTTTCTGTTTTAGTGTAGT GGGAAGTCTGTTTCTTTTTATTATGTTTTTTA CGAAAAAGTCCTGGTCTTTGAAATTCATTGT CTAAATTTTAAATAAAATTATAAAATTTAA	AAAGTG TAGTTC TTTTCT GTTTTA	TAAA AAAT TAAA ATGC	YAPGYEAAQSP CGREPPRDHHR RPRDACGSSHSP EPCLTTPRLLPE

					AAAGAAAATTAATTAAAGAAGCGATGAAAT	GTGTAG	CTTA	TVSAEPCDDESQ
					ATCTCTGAAATTCAATCAATCAATTAATCAT	TGGGAA	AAAA	RTRYASPHKQA
					GGCGTCTCAGCGAGTGCACGTATTTGCCTA	GTCTGT	TAAA	RTLHDAEPRDA
					CCCCTTCGTGGGACCATTCCGGTGCTCCGTA	TTCTTTT	TAAA	SREHAPSCAEPR
					TGCATGGATGCGTCCGGGATGCATCCCACT	TATTAT	TATA	CHRCQWTHWK
					AGGTCGCTGGGCGAATACGGCACATACGCT	GTTTTTT	TCAA	DCCPHSTNTTD
					GCGGCATATCAGCACATAACCCGGCGCCAC	ACGAAA	AATT	GPEGTDRCADTI
					CCACAAGTGGTTATTACATACCGTTGCCGG	AAGTCC	TAAA	TSPATAACPQRS
					GTCTGTGGCGCTGATATGCCCCGGGGTATG	TGGTCT	AAAA	PCPLGSSNGCDE
					AAGCAGCTCAAAGCCCATGTGGCCGCGAGC	TTGAAA	AAAA	TAPEKRQPAAD
					CACCCCGAGACCACCACAGACGCCCACGGG	TTCATT	CGAG	LVHTAPFAVLV
					ATGCTTGTGGAAGCAGCCACAGCCCCGAAC	GTCTAA	GAAC	RAGPFADLVRA
					CCTGCCTTACCACTCCCCGCCTTCTCCCCGA	ATTTTA	AAAT	GPFADHHQDDD
					GACAGTCAGCGCTGAGCCCTGTGACGACGA	AATAAA	AAAC	PLPHRSGSLGPL
					GAGCCAGCGAACGCGGTACGCCTCTCCCCA	ATTATA	ACAA	CSKQKDPRKTH
					CAAGCAGGCGCGTACTCTGCACGACGCCGA	AAATTT	ATTCT	QHRHSGQAGNQ
					GCCGCGCGATGCTTCCCGGGAACATGCCCC	AAAAAG	GAAA	THTDIPRAAPSR
					ATCCTGCGCTGAGCCCCGTTGTCACAGGTG	AAAATT	GATT	RAAICLMANAA
					CCAGTGGACGCACTGGAAGGACTGTTGCCC	AATTAA	TATA	ATREDLLRAATS
					CCATTCCACCAACACGACAGATGGCCCCGA	AGAAGC	TAAT	LSEMAAANQPT
					GGGCACCGACCGCTGCGCAGACACCATCAC	GATGAA	TTAA	RSPTGGGEPTSQ
					CAGCCCCGCAACGGCCGCCTGCCCGCAACG	ATATCT	AAWA	GRRGPQALADA
					TTCCCCCTGCCCCCTTGGGTCAAGTAACGGG	CTGAAA	TAAA	AKRIQQIYRTNIP

					TGTGACGAGACGGCTCCTGAGAAGCGGCAA	TTCAAT	TCGA	RAMRKVLRTLL
					CCAGCCGCCGATCTCGTCCATACCGCCCCGT	CAATCA	AAAT	TAVFSACLRTG
					TCGCCGTCCTCGTCCGTGCCGGCCCGTTCGC	ATTAAT	AAAT	HVPDLCKKSRT
					CGATCTCGTCCGCGCCGGCCCGTTCGCCGA	CATGGC	GTTG	VLIHKKGDRTD
					CCACCACCAGGACGACGACCCCCTCCCGCA	GTCTCA	AAAA	LSNWRPLSMGD
					CCGGTCTGGGAGTCTCGGCCCCCTCTGCTCC	GCGAGT	CAAA	TIPKLFAAVMA
					AAGCAGAAGGACCCCCGAAAGACCCACCA	GCACGT	AAAA	DRLTAFLTNGG
					GCACCGCCACAGCGGGCAGGCCGGGAACC	ATTTGC	AAAT	RLSEEQKGFLQH
					AAACCCACACGGACATACCCAGAGCCGCCC	CTACCC	AATA	EGCHEHNFVLG
					CCAGCAGAAGGGCCGCAATCTGCCTGATGG	CTTCGT	ATAA	QVLEESRRQGK
					CCAATGCCGCGGCCACCAGGGAGGACCTGC	GGGACC	TAAT	DLVMGWLDLSN
					TGAGGGCCGCCACCAGCCTTTCCGAAATGG	ATTCCG	AAAA	AFGSIPHATIMD
					CGGCCGCGAACCAGCCTACCCGCTCGCCCA	GTGCTC	ACAC	AVAGMGIPSRIR
					CTGGAGGTGGCGAGCCCACCTCACAGGGTA	CGTATG	AATA	TIIHQLATGAAT
					GGCGCGGACCGCAAGCACTGGCAGACGCA	CATGGA	ACAC	TAKTIDGMSEEI
					GCGAAAAGGATCCAACAAATATACAGGACC	TGCGTC	TCAC	PIEAGVRQGCPA
					AACATACCTCGCGCCATGAGAAAAGTCCTG	CGGGAT	CCGG	SPILFNIAIERVL
					AGAACACTGCTCACGGCAGTGTTTAGCGCC	GCATCC	CCTG	RKIKTVNAGYL
					TGCCTGAGGACAGGTCATGTCCCCGATCTG	CACTAG	CCCC	LYGSRISPLAYA
					TGTAAAAAGTCCAGAACGGTCTTAATCCAC	GTCGCT	AGAG	DDLVLIASSPEE
					AAGAAAGGSGACAGAACTGACCTGTCAAAT	GGGCGA	GCAG	MRSLLRAADDA
					TGGAGGCCTCTTTCCATGGGTGACACCATCC	ATACGG	GTAA	AIEAGLHFNPKK
					CCAAATTGTTCGCAGCCGTCATGGCGGACA	CACATA	ACAT	CATLHLTGKKSS

					GGCTGACGGCGTTCCTCACTAACGGAGGAA GGCTCAGCGAGGAGCAGAAGGGCTTCCTCC AGCACGAAGGCTGCCATGAACACAATTTTG TTCTTGGCCAAGTGCTGGAGGAGAGCAGAC GCCAAGGCAAGGACCTCGTCATGGGCTGGC TGGACCTGTCCAACGCGTTCGGGTCGATTCC GCATGCCACCATCATGGACGCGGTCGCCGG TATGGGGATCCCTTCGAGGATCCGGACCAT AATCCACCAGCTGGCCACCGGCGCCGCGAC CACCGCCAAAACCATTGATGGCATGTCGGA AGAGATCCCGATCGAAGCGGGGGTCAGACA GGGCTGCCCAGCCAGCCCAATCCTCTTTAA CATCGCAATAGAGCGGGTACTTCGCAAAAT CAAAACCGTCAACGCGGGGTACCTGCTCTA TGGGAGCCGCATTAGCCCGCTGGCGTACGC CGATGACCTGGTGCTAATTGCGAGCTCCCC AGAGGAGATGAGGTCCTTGCTGCGTGCTGC GGACGACGCCGCAATAGAAGCCGGTCTGCA CTTCAACCCCAAGAAGTGCGCGACCCTACA CCTCACGGGGAAGAAATCCTCGCGGAGGGC AGTGCAGACCGGCTTCCTCGTCCGTGGCAC GCCAATACCGGCCATGACAGAGGGGGATGC CTACGAATACCTCGGCATCCCCCTGGGTTTA	CGCTGC GGCATA TCAGCA CATAAC CCGGCG CCACCC ACAAGT GGTTAT TACATA CCGTTG CCGGGT CTGTGG CGCTGA (SEQ ID NO: 1184)	TTACT GGCC ATAT GGCT TTTTT TTTA A (SEQ ID NO: 1307)	RRAVQTGFLVR GTPIPAMTEGDA YEYLGIPLGLKK NQTPRAAMEAI VGDIAKIDDSLL APWQKIDAART FVAPKLDFVLRS GATLRAPLRHL DTVIKKHIKKW LYLPQRASAEV VYTPLKKGGAG ILPSSILADVLTI AQAHRMVSCPG EVVSRIASEGLR EAVKRKINREPS GDEMAHFLSGS TLSGETASFGDA GFWSRVRMATK RQAVHLGVRW AWRGGELLVES RGQRNRPVATD SNSRSQLIQRLR CAAQDEFLTILI

					AAAAAAAACCAAACACCCAGGGCAGCGAT			NKPDQGKVAKL
					GGAAGCGATAGTTGGGGACATAGCCAAGAT			STLTPVSNAFIR
					AGATGACTCGCTGCTCGCCCCGTGGCAAAA			DGSFTRFADWR
					GATCGACGCGGCCCGCACCTTCGTGGCACC			FIHRARLGVLPL
					GAAGCTTGACTTCGTGCTACGAAGTGGCGC			NGAIRWGSGDK
					CACCTTGCGGGCCCCGCTGCGTCATCTGGAT			RCRVCGYQLES
					ACAGTCATTAAAAAACACATTAAAAAATGG			VPHVLCHCMHH
					CTGTATCTGCCGCAGAGGGCGAGCGCGGAG			SNAMQQRHNA
					GTAGTATACACCCCGCTGAAGAAAGGTGGA			VMDRLAKAGSR
					GCGGGCATACTACCTTCATCTATATTGGCTG			LGTPRVNCRVE
					ATGTCCTAACTATCGCCCAGGCTCACCGCAT			GVAEDMAALRP
					GGTGTCCTGCCCTGGGGAGGTCGTCTCCCG			DLVWRDERSRK
					GATTGCAAGTGAGGGCCTGAGAGAAGCGGT			IVIVDVTVPFEN
					AAAGCGAAAAATAAACCGGGAGCCATCCG			GAEAFDNARGE
					GCGACGAGATGGCCCACTTTCTCTCAGGCT			KEEKYRPLAEA
					CCACTCTATCCGGGGAGACAGCCAGCTTTG			LRAMGYQVKLE
					GCGACGCCGGATTCTGGTCGAGGGTGAGGA			AFIVGALGSWD
					TGGCCACCAAAAGGCAAGCTGTGCATCTGG			PKNERVLKTLG
					GGGTGCGTTGGGCCTGGAGAGGAGGTGAGC			VSRFYAGLMRR
					TACTGGTCGAGAGTAGAGGACAAAGAAACC			LMVADTIRWSR
					GACCAGTGGCCACCGACTCGAACTCCAGGT			DIYVEHVSGIRQ
					CCCAACTCATCCAACGTCTCAGGTGCGCAG			FTLPSGAPSN
					CTCAGGATGAGTTCCTGACCATCCTCATAA			

					ATAAACCCGACCAGGGGAAGGTGGCGAAG CTCTCCACGCTAACCCCAGTCAGCAACGCG TTCATACGCGACGGTAGCTTTACCAGGTTTG CTGACTGGCGGTTTATCCACAGAGCCCGAC TGGGAGTCCTCCCACTCAACGGAGCGATCC GATGGGGCAGCGGCGACAAGCGCTGCCGG GTCTGTGGATATCAGCTGGAGAGCGTTCCA CACGTGTTGTGCCACTGCATGCACCACTCA AACGCAATGCAGCAGAGGCACAACGCGGT GATGGATCGCCTCGCCAAGGCTGGCTCACG GCTGGGGACCCCCAGGGTGAACTGCCGCGT GGAAGGGGTCGCCGAGGACATGGCGGCCCT CAGGCCGGACCTGGTATGGCGCGACGAACG GAGCAGAAAAATCGTCATAGTTGACGTGAC TGTTCCGTTCGAGAACGGGGCTGAAGCGTT TGATAACGCGAGGGGCGAGAAAGAAGAAA AATACCGCCCCCTAGCTGAAGCCCTGCGCG CCATGGGATACCAGGTAAAACTGGAGGCAT TCATTGTCGGAGCCTTGGGCTCGTGGGACC CTAAAAACGAAAGGGTCCTTAAGACTTTGG GTGTCTCCAGGTTTTATGCTGGCCTGATGCG CAGACTGATGGTGGCCGACACCATCAGGTG GTCCCGGGACATTTATGTGGAGCATGTATC			(SEQ ID NO: 1429)

					CGGGATCAGGCAGTTCACCCTGCCAAGTGG AGCTCCCTCCAACTAAAAAATTAAAATGCC TTAAAAATAAATAAATATATCAAAATTTAA AAAAAAAAACGAGGAACAAATAAACACAA ATTCTGAAAGATTTATATAATTTAAAAWAT AAATCGAAAATAAATGTTGAAAACAAAAA AAAAATAATAATAATAATAAAAACACAATA ACACTCACCCGGCCTGCCCCAGAGGCAGGT AAACATTTACTGGCCATATGGCTTTTTTTTT AA (SEQ ID NO: 1061)			
NeSL	Utopia-1_Gav	.		Gavialis gangeticus	CGCTGGAAAGACGGAGAACCGCTTCTTTTT CCTGCGCCCGGCCTGGTATTGCACTTCCTCC AGGACCAGCGCCAACCTAGTCCGGCAGACT GCCGGAATAATAGCCTCAGAAAGAGAGCTG GCTAGCAGCCCTCTTTTCTTTCCTCCGGTGC AGCGTGGGTTCTTGTCAGTCCTGATGGGCTA GGGAAGGCGGTGCCGCCAGTACGTCCGAAA GAGCGCCGGTTGCGCGAGCGACCGCGCCGC TCAGGCGAGTAGCCCAAGGGTCTTACGGTT CGCCGGACCCGATAACGCGAAAGCCCCGAC TCGGGCCAGTAGCCGAAGACCNTGGGCCTC CCTCCCCAGGTCGGAGTAGGCGAACGCCCG TGCTCGGAGGACGGAACGTGGACAAAACAC	CGCTGG AAAGAC GGAGAA CCGCTT CTTTTTC CTGCGC CCGGCC TGGTAT TGCACT TCCTCC AGGACC AGCGCC AACCTA	TGAA CCGC CCCC CCTC CGCG CCAG ACGG ACCT TCAC TTCA CTCC GAGA GGAT	MSGPRQAAADP RPSTDPRRQRDS QSPEPRLTRAAS RRRTPDPEDAPR TTAEHPERRRTP PDPRGPSATTAG PERRRPPDPGGP EDDPPEGLPTLV EEPRTPPTPDPP DGRPRRGCRRG SAHVPPLPPPCE AAVPDLPPAKA VQVAQRHEQTP

					CCCCAGGTCCCAATGACGCCCTGATCCACT	GTCCGG	TCTTC	TALPPAAPSVLL
					GACAAGAACGCTCGAGGCACNCCAGGAGA	CAGACT	GACC	LPLRHRVRGPEA
					CCCCCAGCTAGGGCAGACCGCCGACCACGG	GCCGGA	ACGG	PEEPPQGMPGPR
					GTCGCGGAGGACCCTCCCAGGAGGGTGGAC	ATAATA	ACGA	GREETRHAGEV
					CAGCGAACCCGAGTCGGCGACGAACCCCGA	GCCTCA	CCCC	RRPTTRAAARRP
					CGCACCCCCCCCGCGATGTCGGGACCCCGA	GAAAGA	GCTC	ARPAAPPATPPD
					CAGGCGGCGGCGGACCCCCGGCCATCGACC	GAGCTG	CACC	QTSGDRPTERPA
					GACCCCCGGAGGCAGAGAGACTCTCAGAGC	GCTAGC	CGAA	PATPPRRSAPRD
					CCGGAACCCCGGCTGACGAGAGCCGCCTCC	AGCCCT	GAGG	PRPDVTPRPDGP
					CGGCGGAGGACCCCGGACCCCGAGGACGCC	CTTTTCT	ACCC	PPGPPGPPDAPD
					CCCCGGACGACGGCGGAGCACCCCGAGCGA	TTCCTC	CCGC	PPRIPEPPGEPEP
					CGGCGGACTCCTCCGGACCCCCGCGGNCCC	CGGTGC	GATG	PGALQLPSVPGS
					TCGGCGACGACGGCGGGCCCCGAGCGGCGA	AGCGTG	AGAC	PGAETSAQQRM
					CGNCCCCCGGACCCCGGCGGTCCCGAGGAC	GGTTCT	TCTAT	PTPRQALWLEEL
					GACCCCCCCGAGGGCCTCCCCACNCTGGTG	TGTCAG	ACGG	SRATAFEAFEAS
					GAGGAGCCCCGAACCCCCCCGACACCGGAC	TCCTGA	ACTG	VARLTEELSAA
					CCCCCCGACGGACGACCCAGGCGAGGGTGC	TGGGCT	AGGC	ARPGQPRRGAD
					AGACGCGGCAGCGCTCACGTTCCTCCCCTTC	AGGGAA	ACTT	NGPTTRRDHRP
					CCCCTCCCTGCGAAGCTGCTGTGCCCGACCT	GGCGGT	CCTTC	QPQRRPRRQRY
					GCCACCCGCCAAGGCAGTACAGGTCGCACA	GCCGCC	GAAC	DPAAASRIQKLY
					ACGACATGAACAAACACCTACGGCGCTTCC	AGTACG	CACT	RANRPKAAREIL
					ACCAGCTGCGCCTAGCGTTCTACTGCTCCCT	TCCGAA	TCCTC	EGPSAFCQVPRE
					CTGCGGCACCGAGTACGAGGCCCTGAAGCT	AGAGCG	CACC	TLFNYFSRVFNP

					CCTGAAGAACCACCACAAGGTATGCCAGGG	CCGGTT	ATTG	PAEAAAPRPAT
					CCCCGGGGCCGAGAGGAGACCCGGCACGCT	GCGCGA	CGGA	VEALTPVPPAEG
					GGTGAGGTCCGCCGCCCCACGACCCGGGCC	GCGACC	CCAT	FEEAFTPREVEA
					GCGGCGCGAAGGCCCGCCAGACCGGCCGCC	GCGCCG	TGTA	RLKRTRDTAPG
					CCGCCGGCGACCCCACCGGACCAGACCTCC	CTCAGG	ACGG	RDGIRYGLLKK
					GGGGACCGCCCGACGGAGAGACCCGCCCCG	CGAGTA	GTTT	RDPGCLVLSVLF
					GCGACGCCACCACGCAGGTCTGCACCCAGG	GCCCAA	GTGT	NRCREFRRTPAA
					GACCCCCGACCGGACGTGACGCCCCGACCG	GGGTCT	GTAT	WKRAMTVLIHK
					GACGGCCCCCCTCCCGGACCCCCGGGGCCG	TACGGT	CTAT	KGDPTDPGNWR
					CCCGACGCCCCCGACCCGCCGAGGATCCCG	TCGCCG	CTCCT	PIALCSTVAKLY
					GAGCCGCCCGGNGAGCCCGAGCCGCCGGG	GACCCG	TTCTC	ASCLAARITDW
					AGCCCTCCAGCTCCCGAGCGTGCCGGGGTC	ATAACG	TCTC	AVTGGAVSRSQ
					TCCGGGTGCGGAGACCTCCGCACAGCAGAG	CGAAAG	AGCG	KGFMSTEGCYE
					GATGCCCACCCCGCGGCAAGCCCTCTGGCT	CCCCGA	TCGC	HNFTLQMALDN
					GGAGGAGCTCTCCCGGGCCACCGCCTTCGA	CTCGGG	GAAC	ARRTRKQCAVA
					GGCCTTCGAGGCCTCGGTGGCCCGGCTCAC	CCAGTA	CCCC	WLDISNAFGSVP
					GGAGGAGCTCTCGGCGGCCGCCCGGCCCGG	GCCGAA	TCCC	HRRIFGTLRELG
					CCAGCCCCGGAGGGGCGCCGACAACGGACC	GACCNT	CCAC	LPDGVIDLVREL
					GACGACGCGACGAGACCACAGACCGCAGC	GGGCCT	CCCC	YHGCTTTVRAT
					CGCAGAGGCGACCCAGGCGCCAGCGCTACG	CCCTCC	CACC	DGETAEIPIRSG
					ACCCGGCGGCAGCCTCCCGGATCCAGAAGC	CCAGGT	CCCG	VRQGCPLSPIIFN
					TGTACCGGGCCAACCGCCCCAAGGCGGCGA	CGGAGT	GGCT	LAMEPLLRAVA
					GAGAGATCCTGGAGGGACCCTCGGCTTTCT	AGGCGA	TAGT	GGPGGLDLYGQ

					GCCAGGTCCCCCGGGAGACTCTGTTCAACT	ACGCCC	TGGC	KLSVLAYADDL
					ATTTCAGCAGGGTCTTCAACCCCCCGGCAG	GTGCTC	TAAC	VLLAPDATQLQ
					AAGCCGCCGCCCCACGCCCCGCGACCGTCG	GGAGGA	ATTG	QMLDVTSEAAR
					AAGCGCTGACCCCCGTCCCCCCGGCAGAGG	CGGAAC	TATCT	WMGLRFNVAK
					GGTTCGAGGAGGCCTTCACGCCGCGGGAAG	GTGGAC	CCTG	CASLHIDGRQKS
					TGGAAGCCCGCCTGAAGAGGACCAGGGAC	AAAACA	TAAC	RVLDSTLTIQGQ
					ACCGCCCCCGGCAGGGACGGCATCAGGTAC	CCCCCA	CTAG	AMRHLRDGEAY
					GGTCTCCTNAAGAAACGTGACCCGGGCTGC	GGTCCC	TCGC	CHLGTPTGHRA
					CTCGTTCTTTCTGTTCTCTTCAACAGGTGCA	AATGAC	GTTC	KQTPEETINGIV
					GAGAGTTCCGGCGCACGCCCGCCGCCTGGA	GCCCTG	CCCT	QDAHKLDSSLL
					AGAGGGCCATGACGGTCCTCATCCACAAGA	ATCCAC	CCTC	APWQKIDAANT
					AGGGAGACCCGACCGACCCGGGCAACTGG	TGACAA	ACCC	FLIPRVAFVLRG
					AGACCCATCGCCCTGTGCTCCACCGTGGCC	GAACGC	CCAT	SAVPKTPLKKA
					AAGCTGTACGCCAGCTGCCTGGCGGCCCGC	TCGAGG	CCCT	DAEIRRLLKKW
					ATCACCGACTGGGCGGTGACCGGCGGGGCC	CACNCC	CTATT	LHLPLRASNEVL
					GTCAGCCGGAGCCAGAAGGGCTTCATGTCG	AGGAGA	GTTA	HIPYRQGGANV
					ACGGAGGGCTGCTACGAACACAACTTCACC	CCCCCA	GTCC	PRMGDLCDIAV
					CTCCAGATGGCCCTGGACAACGCCCGGAGG	GCTAGG	CTCG	VTHAFRLLTCPD
					ACCAGGAAGCAGTGCGCGGTGGCGTGGCTG	GCAGAC	CTCG	ATVSIIAASALEE
					GACATCTCCAACGCCTTCGGCTCCGTGCCCC	CGCCGA	GGCG	TARKRIARQPTG
					ACCGCCGCATCTTCGGCACCCTCCGCGAGC	CCACGG	ATCT	RDLATFLSGSLE
					TGGGCCTACCGGACGGCGTCATCGACCTGG	GTCGCG	GTAT	GEFGRDGGDFA
					TGCGAGAGCTCTACCACGGNTGCACCACGA	GAGGAC	TTCCC	SLWSRARNATR

					CCGTCCGCGCCACCGACGGAGAGACCGCGG	CCTCCC	TATC	RLGKRIGCAWT
					AGATCCCCATCCGGTCGGGGGTGAGGCAGG	AGGAGG	GGCT	WTEECRELGVS
					GCTGCCCCCTCAGCCCCATCATCTTCAACCT	GTGGAC	TTGTC	LQPAPHADRVT
					GGCCATGGAACCGCTCCTTCGAGCCGTGGC	CAGCGA	ATCTT	VTPRTRTFLERF
					GGGCGGCCCCGGCGGGCTCGACCTGTACGG	ACCCGA	TTTTC	LKDAVRNKYAG
					CCAGAAGTTGAGCGTCCTGGCCTACGCCGA	GTCGGC	TGGA	DLRAKPDQGKV
					CGACCTCGTCCTCCTCGCCCCCGACGCCACC	GACGAA	TTCCC	FDVTSKWDASN
					CAGCTGCAGCAGATGCTGGACGTGACGTCC	CCCCGA	GATC	HFMPSGSFTRFA
					GAGGCGGCCAGGTGGATGGGCCTGCGCTTC	CGCACC	CTAA	DWRFLHRARLN
					AACGTCGCCAAGTGCGCCTCCCTGCACATC	CCCCCC	ACAT	CLPLNGAVRFG
					GACGGCAGGCAGAAGAGCCGCGTCCTGGAC	GCG	TTACT	HRDKRCRRCGY
					TCCACCCTCACGATCCAGGGCCAGGCGATG	(SEQ ID	AATA	AAETLPHVLCSC
					AGGCACCTGCGCGACGGCGAGGCCTACTGC	NO: 1185)	AAAG	KPHARAWQLRH
					CACCTGGGGACGCCCACCGGCCACCGGGCC		TCAA	NAVQDRLVRAI
					AAGCAGACGCCGGAGGAGACCATCAACGG		TCTGT	PAAAGEISVNRT
					GATCGTGCAGGACGCCCACAAGCTGGACTC		TCTTT	VPGCESQMRPDI
					GTCCCTGCTGGCCCCCTGGCAGAAGATAGA		(SEQ	VITNEEAKKVVI
					CGCGGCGAACACCTTCCTCATCCCCCGCGTC		ID NO:	VDVTIPFENRRQ
					GCGTTCGTCCTGAGAGGCTCGGCGGTCCCC		1308)	AFTDARARKRE
					AAGACCCCCCTCAAGAAGGCGGACGCCGAG			KYAPLADTLRG
					ATCCGGCGGCTGCTCAAGAAGTGGCTGCAC			RGYDVTVDALI
					CTGCCGCTGAGGGCCAGCAACGAGGTCCTG			VGTLGAWDPSN
					CACATCCCCTACCGGCAGGGAGGCGCCAAC			ESVLRACRVSRR

					GTCCCCCGCATGGGAGACCTCTGCGACATC GCGGTGGTCACCCACGCCTTCCGCCTCCTGA CCTGCCCGGACGCGACGGTAAGTATCATCG CCGCCAGCGCCCTCGAGGAGACCGCCCGCA AGAGGATCGCGAGGCAGCCGACCGGACGN GACTTGGCCACCTTCCTCAGCGGCTCGCTGG AGGGCGAGTTCGGCCGAGACGGCGGGGACT TTGCCTCGCTGTGGAGCCGAGCCCGCAACG CCACGCGCCGCCTCGGGAAGCGCATCGGCT GCGCCTGGACCTGGACCGAGGAGTGCCGGG AGCTGGGAGTCTCCCTGCAACCAGCCCCGC ACGCCGACCGCGTCACCGTGACGCCCCGCA CGAGGACCTTCCTGGAGAGGTTCCTGAAGG ACGCCGTCCGAAACAAGTACGCCGGCGACC TGAGGGCCAAACCCGACCAGGGCAAGGTCT TCGACGTCACCTCGAAGTGGGACGCTAGCA ACCACTTCATGCCCAGCGGGAGCTTCACGC GCTTCGCGGACTGGCGCTTCCTCCACCGCGC CCGCCTCAACTGCCTGCCTCTGAACGGGGC CGTGCGCTTCGGCCACCGGGACAAGAGGTG CCGACGGTGCGGCTACGCGGCAGAGACCCT CCCCCACGTGCTGTGCAGCTGCAAGCCGCA CGCCAGAGCCTGGCAGCTCCGCCACAACGC			YAKLMRCLMVS DTIRWSRDIYVE HITGHRQYSDPT RRAAAGPDPEG TA (SEQ ID NO: 1430)

					TGTCCAGGACCGCCTGGTGAGGGCCATCCC GGCCGCGGCGGGGGAGATCTCCGTGAACCG CACCGTCCCGGGCTGCGAGAGCCAGATGCG ACCCGACATAGTCATCACCAACGAAGAGGC CAAGAAGGTCGTGATCGTGGACGTCACCAT CCCCTTCGAGAACCGGCGCCAAGCCTTCAC CGACGCCCGGGCTCGCAAGCGGGAGAAGTA CGCCCCGCTGGCCGACACCCTGAGGGGCCG CGGCTACGACGTGACGGTCGACGCGCTCAT CGTGGGAACGCTCGGAGCCTGGGACCCCAG CAACGAGAGCGTCCTGCGTGCCTGCCGCGT CTCCCGCCGCTACGCCAAGCTGATGCGCTG CCTCATGGTGTCCGACACCATCCGTTGGTCC CGCGACATCTACGTGGAACACATCACGGGC CACCGCCAGTACTCCGACCCCACCAGACGA GCCGCCGCCGGACCGGACCCGGAGGGGACC GCCTGAACCGCCCCCCCTCCGCGCCAGACG GACCTTCACTTCACTCCGAGAGGATTCTTCG ACCACGGACGACCCCGCTCCACCCGAAGAG GACCCCCGCGATGAGACTCTATACGGACTG AGGCACTTCCTTCGAACCACTTCCTCCACCA TTGCGGACCATTGTAACGGGTTTGTGTGTAT CTATCTCCTTTCTCTCTCAGCGTCGCGAACC			

					CCCTCCCCCACCCCCCACCCCCGGGCTTAGT TGGCTAACATTGTATCTCCTGTAACCTAGTC GCGTTCCCCTCCTCACCCCCATCCCTCTATT GTTAGTCCCTCGCTCGGGCGATCTGTATTTC CCTATCGGCTTTGTCATCTTTTTTCTGGATTC CCGATCCTAAACATTTACTAATAAAAGTCA ATCTGTTCTTT (SEQ ID NO: 1062)			
NeSL	Utopia-1_LV	AGCV01358106		Lytechinus variegatus	ATCTACTATCATGTCTTGTCCAAGAGAGGG AAGCGATCACCTCGGTCCTGATCCTGAGAC ACCCGCCCTCCATCAGGGTTCTGACATCCG GGTTACCAGTTCTCGCCTTCGAGGTTCCCGC GGAAAGAGTTCTCGCCAACCAAGCTCCCGA CACCAAGTTCCTGCCAGCGAGGCTTCCGCC ACCGCCCAGCAGACTGCCGCCAACGAGTGC CAGGTGTGTGGATCTTCCTTCGCCACCTCCA GTGGACTCCGCCGCCACATGGCCAGGCTTC ATCGAGCTGCCTCTGCGGATCCTGAGGGTG CTGCGCCGGCTTCCATCACAGAGATTTTCGA CTACCCCTTGCCTTCCCGGTGGAAATGCTCG GCATGCTCGGAGAACTTTTTCAACCAGCAG ACCCTCAAGCGACACCAGACCAGGCATCAT CCAGCTACCACCTTCGCGTATGCCTTTCGGT GTTCGTCATGCCGGTCCGAGTTCGACTCAGC	ATCTAC TATC (SEQ ID NO: 1186)	TGAA TAGC ATTT ATAT TGTG TTCC AAAC AACA TACT CATT ATTA TATCT AAAC ATTTT TTTTT CTGTT	MSCPREGSDHL GPDPETPALHQG SDIRVTSSRLRG SRGKSSRQPSSR HQVPASEASAT AQQTAANECQV CGSSFATSSGLR RHMARLHRAAS ADPEGAAPASIT EIFDYPLPSRWK CSACSENFFNQQ TLKRHQTRHHP ATTFAYAFRCSS CRSEFDSARRAA NHWQVHKKERS QLSGTEPQASSQ

					ACGGAGGGCTGCGAACCATTGGCAGGTCCA		CCTG	ARVSMAHSPPPL
					CAAGAAGGAGCGATCTCAACTCTCTGGCAC		ACAA	PNTSWAELASN
					CGAGCCCCAGGCCTCTTCCCAAGCCAGAGT		TCTA	PAEIPSFVWESPP
					TAGCATGGCTCATTCTCCTCCACCTCTGCCC		CGTA	KNRPSVEEFGSS
					AACACTTCTTGGGCGGAGCTCGCCTCGAAT		AAGT	LPTDVTMMSQS
					CCTGCCGAGATACCTTCCTTCGTCTGGGAGT		CTGC	PPPQVQSSPVPA
					CTCCTCCCAAGAACCGCCCCTCGGTTGAGG		TAAC	LTPLSPAATASS
					AGTTCGGTTCGTCTCTGCCAACTGATGTTAC		CAAC	SPPGAARQLTPP
					GATGATGTCTCAAAGTCCTCCACCGCAGGT		TGGC	TQTNTPVTQRA
					ACAGTCGTCTCCTGTCCCTGCTCTGACTCCT		ATGA	RLQPEADVVPEL
					CTTTCACCCGCTGCCACTGCCTCCAGTTCTC		TGAA	PPSVTEHPVSDA
					CTCCAGGGGCTGCAAGGCAGCTGACCCCTC		ATAA	QHWVDAVSSAS
					CTACACAGACTAACACCCCAGTCACCCAGA		GATA	DWSEFEAVCDQ
					GGGCTCGCCTGCAACCTGAAGCAGACGTCG		AAAT	FVIHAVAVSRPN
					TACCTGAACTCCCTCCTTCAGTCACCGAGCA		CCCC	LARPQQQDRQR
					CCCTGTGTCTGACGCTCAACACTGGGTTGAT		TTAC	SGDHPPRQQRG
					GCTGTATCCTCTGCATCAGATTGGTCTGAGT		ACAT	QHRPTFDVREAS
					TTGAAGCAGTATGTGATCAATTTGTCATCCA		TAAT	RIQKLYRTSKKR
					CGCTGTTGCTGTTTCCCGTCCCAATCTTGCT		TTCTT	AIRHILKEKSPSF
					CGACCCCAGCAGCAAGATAGGCAGAGATCT		GTCA	SGSESDVLDFFR
					GGTGACCACCCTCCTAGACAGCAAAGAGGT		CATC	EVYSAKEVDEE
					CAGCATCGACCAACCTTCGATGTCCGTGAG		ATAA	AVGKLASSLFD
					GCAAGTAGAATCCAGAAGCTCTATCGTACC		TGCTT	VPQGDDSATSLS

					AGCAAGAAAAGAGCCATCAGACACATACTG		TGTC	LPTSAKEIGARL
					AAAGAGAAATCACCTTCCTTCTCTGGTTCCG		AAAG	SRMTNSAPGKD
					AGTCAGACGTCTTAGACTTCTTCCGCGAGGT		CAAT	RLEYRHIRRADG
					GTATTCTGCTAAAGAAGTTGACGAGGAAGC		GTCC	SFSISEAIFNKCL
					AGTTGGTAAACTAGCATCCTCGCTCTTCGAT		TACA	AEGRIPAPWKT
					GTCCCTCAAGGTGATGACTCTGCGACATCTC		TAAT	ASTILLHKAGPT
					TGTCTCTGCCCACGTCAGCGAAGGAGATCG		ATCT	DDPANFRPIALQ
					GAGCAAGGCTGTCAAGGATGACAAACTCTG		CGAT	SCLYKLFMAVL
					CCCCCGGGAAGGATCGCTTGGAGTACAGAC		GTCA	ADRLTKWACEN
					ACATTCGACGTGCGGACGGGTCCTTCAGCA		CCCC	QYLSPEQKSARP
					TCTCTGAGGCCATCTTTAACAAATGCCTGGC		AATT	CEGCFEHSFLLS
					TGAAGGTCGGATCCCAGCTCCTTGGAAGAC		AATT	AALKDCRRNQK
					AGCATCTACCATCCTACTTCACAAGGCTGG		TTAC	TICIGWLDLRNA
					CCCCACGGATGATCCCGCCAACTTCCGCCC		ATCC	FGSIPHPVIKIVL
					AATCGCCTTACAGTCATGTCTCTACAAGCTT		TTCG	SSLGVPDSLVTL
					TTTATGGCTGTACTTGCGGACCGGCTGACCA		GTAA	LMDAYNGASTS
					AGTGGGCCTGTGAGAACCAGTACCTCAGCC		CCTTT	FTLTGGQTDTVP
					CCGAGCAGAAGTCCGCTCGCCCCTGCGAGG		ATAC	IRSGVKQGCPMS
					GGTGCTTCGAGCACTCCTTCCTTCTCTCAGC		CGTT	PILFNLAIELIIRA
					TGCCCTGAAGGACTGCAGGAGAAACCAGAA		GGAT	VKKNASDNHLG
					GACCATCTGCATCGGTTGGTTGGACCTTAG		CAAC	VTVQGKNLSILA
					GAATGCATTTGGAAGCATTCCTCATCCTGTC		ATAT	YADDLVLLSRD
					ATCAAGATCGTCCTGTCCAGTCTGGGTGTCC		ATGA	TEGLQSLLQVA

					CTGATTCGCTTGTTACCCTCCTCATGGATGC		TTTGT	GSSASTLQMQF
					CTACAATGGTGCGTCAACCTCGTTCACGCTG		AAAA	KPQKCATLTLD
					ACCGGGGGCCAGACCGACACCGTACCCATC		CTGTT	CKRGTNVRQSA
					AGATCAGGGGTGAAGCAAGGCTGCCCGATG		ATTTC	HHIQGAAIPSLT
					TCCCCAATCCTCTTCAACCTGGCCATCGAAC		TGAG	EEERYRYLGVPI
					TTATCATCAGGGCAGTCAAGAAGAATGCAT		TTTTT	GLPRLTSLQESS
					CAGACAACCATCTCGGAGTGACTGTCCAGG		TCTAT	RKLSSDIETISSS
					GCAAGAACCTCTCCATCCTGGCCTATGCTG		GCTA	LLAPWQKLDAI
					ATGACCTAGTGCTGCTCAGCCGAGACACTG		ATAA	KTFVIPILQYTLR
					AAGGCCTCCAATCCCTCCTTCAAGTTGCTGG		A (SEQ	ATEYLKSDLKPL
					CTCTTCTGCATCTACCCTTCAGATGCAGTTT		ID NO:	RAAIIKHVKKIC
					AAGCCCCAGAAGTGTGCAACACTCACCCTT		1309)	HLPVRSSNAFVF
					GACTGCAAGCGTGGTACCAATGTTAGGCAG			ASRPSGGLAFVD
					TCTGCTCACCATATCCAAGGGGCTGCCATCC			PGVDADILVVT
					CATCGCTTACTGAAGAGGAGCGTTACCGCT			QAVRTLASDDD
					ACCTTGGGGTCCCCATTGGACTCCCAAGGC			TVRAVALGQLT
					TCACAAGTCTACAGGAGAGCTCTAGGAAGC			SVVHRTVHSAPS
					TGTCGTCAGACATCGAGACCATTTCATCCTC			DDCIDKFLSGSS
					CCTTCTAGCGCCTTGGCAGAAACTGGATGC			EGPLANSGNSG
					AATCAAGACCTTTGTGATACCCATCCTTCAG			QASSLWSRTRA
					TATACCTTGCGGGCCACAGAATACCTCAAG			ASRRLKIRIVGA
					TCCGACTTGAAACCCCTACGAGCAGCCATT			SSGDIKVESGGR
					ATCAAGCACGTGAAGAAGATCTGCCATCTT			AIPSKKVTAGLR

					CCGGTAAGATCATCCAACGCCTTCGTCTTCG			SDHHNEMSEKL
					CGTCCCGACCCTCCGGTGGTTTGGCTTTTGT			RSLPDQGKVAR
					TGACCCAGGAGTGGATGCCGACATCCTTGT			ALSLDSFANATS
					CGTAACCCAAGCCGTCCGTACCCTGGCATC			WLTSGSFIRFCD
					GGATGATGACACGGTGAGAGCCGTTGCTCT			WRFIHRARLNC
					AGGACAGCTGACCTCAGTGGTCCACAGGAC			LPTNAAVRRWK
					GGTCCATTCTGCACCATCAGATGACTGCATC			QNANTKCRRCD
					GACAAGTTCTTGTCCGGAAGCAGTGAGGGC			HQLETLPHIINN
					CCTCTGGCAAACTCAGGGAACAGTGGCCAG			CRPNMVPIRRRH
					GCATCATCTCTTTGGTCACGTACAAGGGCTG			NSIQERLVKAIH
					CATCCCGTCGCCTGAAGATAAGGATCGTCG			YGDIYQDQHVP
					GAGCCTCGAGCGGCGACATCAAGGTCGAGA			GDPNPRERPDIT
					GTGGAGGCCGTGCCATCCCCTCCAAGAAAG			VVEGNKVTIVDI
					TGACTGCAGGTCTCCGCTCCGACCATCACA			TIPFDNGPDALS
					ACGAGATGTCAGAGAAGCTGCGGTCCCTTC			TAANAKVMKY
					CAGACCAGGGCAAGGTAGCCAGGGCACTCT			DTLRQELASRG
					CCTTGGATAGTTTCGCCAACGCAACATCAT			MDVEVHAFVIG
					GGCTGACATCCGGGAGTTTCATCCGATTTTG			SLGSWHGDNER
					CGACTGGAGGTTCATCCACCGTGCTCGCCT			VLGRLGISRRYR
					AAACTGTCTACCTACCAACGCTGCTGTCCG			TLMRRLCCIDAI
					GCGGTGGAAACAGAACGCCAACACCAAGT			KGSRDIYIEHVT
					GCCGAAGGTGTGATCACCAGTTGGAGACGT			GHRQY (SEQ ID
					TGCCCCACATTATTAACAACTGCCGTCCCAA			NO: 1431)

					CATGGTACCGATCCGCCGGAGGCACAACTC CATCCAAGAACGCCTGGTGAAGGCGATTCA CTACGGTGACATCTACCAGGACCAGCATGT CCCCGGCGACCCGAACCCAAGGGAGCGACC CGACATCACAGTGGTTGAGGGTAACAAGGT GACCATTGTCGACATCACCATCCCATTCGAC AATGGTCCGGATGCCCTCTCCACGGCTGCC AACGCCAAAGTCATGAAATATGACACACTG CGACAAGAGCTGGCCAGCCGTGGTATGGAT GTCGAGGTCCACGCTTTTGTAATCGGCTCCC TTGGGTCGTGGCACGGGGACAACGAGCGTG TTCTGGGACGTCTAGGTATCTCTCGCCGATA CAGAACACTGATGAGGCGCCTGTGCTGCAT AGATGCGATCAAAGGCAGCCGGGATATCTA TATCGAACATGTTACCGGTCACCGGCAGTA TTGAATAGCATTTATATTGTGTTCCAAACAA CATACTCATTATTATATCTAAACATTTTTTTT TCTGTTCCTGACAATCTACGTAAAGTCTGCT AACCAACTGGCATGATGAAATAAGATAAAA TCCCCTTACACATTAATTTCTTGTCACATCA TAATGCTTTGTCAAAGCAATGTCCTACATAA TATCTCGATGTCACCCCAATTAATTTTACAT CCTTCGGTAACCTTTATACCGTTGGATCAAC			

					ATATATGATTTGTAAAACTGTTATTTCTGAG TTTTTTCTATGCTAATAAA (SEQ ID NO: 1063)			
NeSL	Utopia-1_NVit	.		Nasonia vitripennis	CCATTCCTTCGTACGGGTTTTCGTGCCGGCA TAGCCGGGTGGGAGACTCGCGCGGGGGAG GTCATATCTCACCACCATCCTCGAGCTTTTT GCTGCACCTGATCCGGGMTAACTGGCCGAG AAGTTAAATGTATTACTGTTAATGTACTCAT GGAGCAACAACCACACACCAAAGCGATCCG GGAGGGGGACTTCATCGTGATACTCCTCCC GCAAAGTGACGACGAGACACTTTGCTGTCC ACTCTGCGTGGGACGCGGACGGTACTCCGG CAAGACGCGGGTTGAGTGTCTCAATAGACA TGTCAAAGAGGTGCACCCAGACCTAACCAC CACCTTTAGGTGCTGGGGGTGCGGGTTCGC GGCTCCCGGGGATAAGAAATATCCCCGGAA GATCGTGACCCAGCACTGCGCCACCTGCGT CCCCGAGGTTTCGTCCGCGCCCTCGGGCAG AGTGGACGGAGAGAGGAGAGTTAACACCC GCCGTCGGTTGGGTATTGCCGCGGCAACCG AGGCCAGTCCCGTGAGACGTACTCGTCGTA ACGGCCTGGCCTCACCTCCAGTCGAGCAGA ACATCAGCCAAAGCGCGGCTCCTCCGGAAC CCGCAAGGGTGCCACAACATCCGGAGATAG	CCATTC CTTCGT ACGGGT TTTCGT ATAGCC GCCGGC GGGTGG GAGACT CGCGCG GGGGAG GCTCGA ACCTGA (SEQ ID NO: 1187)	TAGT GGGG CCAT AACA CCTA GGCC CCAC AGTG TGGC GATG TCCA TGTG TGTC GTCC TTACT TATTT ATTT ATTT GTCC TGTC GAGC	SGXTGREVKCIT VNVLMEQQPHT KAIREGDFIVILL PQSDDETLCCPL CVGRGRYSGKT RVECLNRHVKE VHPDLTTTFRC WGCGFAAPGDK KYPRKIVTQHC ATCVPEVSSAPS GRVDGERRVNT RRRLGIAAATEA SPVRRTRRNGL ASPPVEQNISQS AAPPEPARVPQH PEIVALGESADD EVFRSPVNSPPR DWRAAAPQQA ASSSPXAVPGIT AATPSNTTRTGN GSAXSILAEHPIP

					TTGCGCTGGGAGAGTCAGCGGATGACGAAG		TGTTT	APPPTNTTEANG
					TATTTCGATCTCCGGTGAACTCTCCCCCAAG		AATC	RADIPRSGRAPP
					AGACTGGAGGGCTGCGGCGCCCCAACAGGC		TATTC	PGXQAARRRAP
					GGCTTCCTCCTCTCCTCSGGCGGTGCCCGGT		ATAT	TTEQRRIVGLLE
					ATCACTGCTGCAACCCCGTCCAACACCACC		GTAT	AATGREQLEEA
					AGGACAGGGAATGGCTCGGCANACTCCATC		GTGT	TTQAMLFLARL
					CTAGCCGAGCACCCGATACCTGCACCTCCG		TGTG	TGRRPEPRNAIR
					CCAACCAACACCACGGAGGCGAATGGTCGT		TGTG	PGXRQRHPAQG
					GCCGACATACCCCGTTCTGGGCGGGCTCCG		TCGT	DVQAQAPDRIX
					CCGCCCGGTCNACAAGCCGCCCGTAGAAGG		CCCT	EAKKLQRLYRT
					GCTCCCACGACAGAACAACGTCGCATTGTG		CGTA	SKKRAVQKILA
					GGACTCCTCGAGGCGGCCACTGGTCGGGAG		GCAG	GPXMNCQIDKN
					CAGCTTGAGGAGGCAACGACTCAAGCTATG		TTTTA	TITAHFVELAAR
					CTATTCCTGGCCAGATTGACCGGGCGGCGG		TTCC	RDGGEDWPDVF
					CCTGAACCCCGGAACGCCATTCGGCCGGGA		GTCC	DREEPTAASGEA
					MCGCGGCAACGTCATCCAGCTCAGGGGGAT		AACC	LCTPITREEVFR
					GTCCAAGCGCAGGCGCCTGATCGGATTCAS		AGAG	RLKGRNNTSPGP
					GAGGCCAAGAAACTCCAGCGTCTCTACCGA		GTCG	DGITYRDLAKA
					ACATCCAAGAAGCGGGCAGTMCAGAAGAT		ACCA	XPGAHVLAALY
					TTTGGCAGGTCCTNATATGAACTGCCAAAT		ATAT	NXIWRIEATPAL
					CGACAAGAATACCATCACAGCCCACTTTGT		TAAA	WGVSNTTLIYK
					GGAGCTGGCTGCCCGCAGAGACGGTGGAGA		ATAA	KGDAMDISNWR
					GGATTGGCCCGATGTCTTCGATCGGGAAGA		GCAT	PISLGDTVPKLF

					GCCCACGGCGGCGAGCGGCGAGGCCCTTTG		GGCT	AAILADRIKRW
					TACCCCTATCACAAGGGAGGAAGTGTTTCG		TGAA	AVANGRYSASQ
					ACGGCTCAAGGGGAGGAACAACACCTCACC		GCAG	KGFLEFEGCYEH
					TGGGCCGGACGGAATCACCTACAGGGACCT		GCCA	NFVLQEAIREAK
					CGCCAAGGCCSATCCAGGGGCCCATGTGCT		AGCG	GGRKELVVAWL
					CGCCGCCCTCTATAACASCATCTGGCGCATT		CCGT	DLASAFTSVPHS
					GAGGCGACGCCTGCCTTGTGGGGGGTGTCC		GTTCT	SILQALEGHGLP
					AACACMACCCTKATATACAAGAAAGGCGA		AACC	SKARNIISSLYTG
					CGCCATGGACATCTCCAACTGGAGACCGAT		CCGT	MTTRFHTAEGP
					CTCCTTGGGAGATACGGTCCCCAAGCTATTT		TTTA	TDPILIQSGVRQ
					GCCGCGATCCTGGCCGATAGAATCAAAAGG		GGGG	GCPLSPDVFNLT
					TGGGCCGTGGCCAACGGGCGCTACAGTGCC		AAGT	LEVVLREIQRTG
					TCCCAGAAGGGCTTCTTGGAGTTCGAGGGA		TACTT	EGYTIEGRRISH
					TGCTATGAACATAACTTCGTGCTCCAGGAG		AACC	LAYADDVAILA
					GCCATTCGGGAGGCAAAAGGCGGGAGAAA		TAAA	DSPAGMRRLLF
					GGAGTTGGTGGTGGCCTGGCTGGACTTAGC		AATA	AAERGARAVGL
					CAGTGCCTTTACATCCGTGCCCCACTCCTCT		CAAC	TFNPAKCATLHI
					ATCCTGCAGGCGCTCGAGGGACATGGCCTA		TTTTC	AGRGEEAVRPT
					CCAAGCAAAGCGAGAAACATTATCTCCTCG		C (SEQ	EFSVQGTPVRAL
					CTCTACACAGGCATGACCACCCGCTTCCAC		ID NO:	ASGEAYEHLGIP
					ACGGCGGAGGGACCCACCGACCCTATTCTC		1310)	TGYQVRQTPINT
					ATCCAGTCGGGGGTCAGACAGGGTTGCCCA			LRDLLADIGSID
					TTAAGCCCCGACGTGTTCAACCTTACCCTGG			RSLLAXWQKLD

					AAGTTGTCCTACGGGAAATACAACGGACGG			AVGTFLLPRLDF
					GGGAGGGCTACACGATCGAGGGCAGGCGT			TMQGAHIDKGF
					ATTTCCCATCTCGCCTACGCCGATGATGTCG			LTEADKIIKKAA
					CGATCCTGGCTGACTCCCCGGCTGGAATGA			KSWLSLPQRAS
					GGAGGCTGCTTTTTGCAGCAGAACGGGGAG			AELVFLPPSQGG
					CTCGGGCGGTCGGCCTCACCTTCAACCCAG			GGLLTVAHSYK
					CGAAGTGCGCTACACTGCACATAGCGGGAC			MLYSSDVTVSTI
					GAGGKGAGGAGGCAGTTAGACCGACTGAG			AGSTLRRTVSER
					TTCTCCGTTCAAGGGACACCTGTCCGTGCCC			LKKRASNIDIAR
					TTGCTTCCGGGGAGGCCTACGAACATCTTG			FLSGDLDLPRST
					GCATCCCCACAGGCTACCAAGTCAGGCAGA			SPSTFWTKVRSA
					CCCCAATCAACACCTTGCGGGATCTCCTTGC			ALRIKTKLGLR
					TGACATCGGCTCCATCGACCGGTCCTTACTG			WSWCQGGEVLL
					GCGSCCTGGCAGAAGCTGGATGCAGTGGGA			MACGDPRAPGT
					ACCTTCCTACTGCCCCGGCTTGATTTCACCA			RVSPQTKHLVTT
					TGCAGGGCGCCCATATCGACAAGGGCTTCT			SLRRCLNRHYA
					TGACCGAGGCGGACAAGATCATCAAAAAA			ESLLAKKDQGK
					GCTGCAAAATCTTGGCTAAGCCTTCCTCAGC			VFEVTRRSGQSN
					GAGCGAGCGCGGAGCTCGTGTTTCTACCTC			HFLRSGSFTRFC
					CCTCGCAAGGGGGAGGGGGTTTGCTCACGG			DWRFIHRARLD
					TCGCTCACAGCTATAAAATGCTGTACTCGA			VLPLNAAKRWQ
					GCGATGTCACCGTGAGCACCATAGCGGGAT			RGMDKRCRRCG
					CCACACTGAGGAGGACCGTTTCGGAGAGGC			SDLETLPHVLSH

					TTAAAAAGCGGGCGTCTAACATCGATATCG CCCGCTTCCTCTCCGGAGATCTTGATCTGCC CCGGTCGACGAGCCCGTCGACATTCTGGAC CAAGGTTCGGAGCGCGGCCCTGAGAATTAA GACCAAACTCGGGCTGCGCTGGAGCTGGTG TCAGGGGGGGGAAGTACTCCTGATGGCCTG TGGGGATCCACGTGCCCCCGGAACAAGAGT CTCGCCGCAGACAAAACATCTTGTCACCAC GAGTCTGCGGCGCTGCCTGAATCGGCACTA TGCGGAGAGTCTCCTTGCCAAAAAAGATCA AGGCAAGGTCTTTGAGGTTACCCGGAGGAG CGGACAATCCAACCACTTCCTCCGGAGTGG CAGCTTCACTCGCTTCTGCGACTGGCGCTTT ATCCACCGCGCCCGTCTCGACGTTCTCCCCC TCAATGCTGCCAAACGGTGGCAACGAGGGA TGGACAAGAGATGTCGCAGATGCGGGTCAG ACCTGGAGACTCTCCCGCATGTGCTGAGTC ACTGCGGCCCGCACTCGGCGGCGAGACAAA AGAGGCACAACAACATCCAGGACAGGTTGG TGAAGGCAGCCTCGAGATGTCCGGGGACCA TTAGTGTCAACCAGACCGTTGTGGGCGTCC GCGGCCCCGACGCGGCATTAAGACCKGACA TTGTCGTTCGCGACGATGTCAACCGCCGCGT			CGPHSAARQKR HNNIQDRLVKA ASRCPGTISVNQ TVVGVRGPDAA LRPDIVVRDDV NRRVTIVDVAV PFENRLEAFDGV REAKIAKYTPLA RQLTDSGYTVT VEAFVVGALGA WDPRNERVLSL LSISRYYAILMR RLMVSDTIRWS RDIYVEHVSGIR QYRE (SEQ ID NO: 1432)

					CACCATCGTCGATGTGGCCGTGCCGTTCGA AAACCGCCTCGAGGCGTTCGACGGCGTTCG KGAGGCCAAGATTGCCAAATATACACCTTT GGCCAGGCAACTCACGGACAGCGGTTACAC CGTAACGGTGGAGGCCTTTGTAGTGGGCGC CCTCGGCGCGTGGGACCCGCGAAATGAACG GGTCCTGAGCCTCCTCAGCATTAGTCGTTAT TATGCGATACTGATGAGGAGGCTCATGGTT AGTGACACAATTCGCTGGTCCCGGGACATC TACGTGGAACATGTGTCGGGCATTCGGCAG TACCGGGAATAGTGGGGCCATAACACCTAG GCCCCACAGTGTGGCGATGTCCATGTGTGT CGTCCTTACTTATTTATTTATTTGTCCTGTCG AGCTGTTTAATCTATTCATATGTATGTGTTG TGTGTGTCGTCCCTCGTAGCAGTTTTATTCC GTCCAACCAGAGGTCGACCAATATTAAAAT AAGCATGGCTTGAAGCAGGCCAAGCGCCGT GTTCTAACCCCGTTTTAGGGGAAGTTACTTA ACCTAAAAATACAACTTTTCC (SEQ ID NO: 1064)			
NeSL	Utopia-1_PCa	.		Phytophthora capsici	GCCCGCCGGTGGAGTAGCCATGTTGGCCAC CACCGCCCAAGTCTCCGCCGCAGCTGCGAC TGCTGCTGCTCATCGAGTCGCAGTAGTCGC	GCCCGC CGGTGG AGTAGC	TAGA CGGC ACAG	MVVSRITARLEA TPAPRWDPPLPR RVIASRIADRLV

					AGCTGCACAAGCACCACCTCCTGCCGGACG	CATGTT	TTCTG	PATAPCRSALNA
					CGCCGCCGTGGAGCACCACGCGCGCGCTGA	GGCCAC	GCCC	AFPSPSRDTVTE
					GCCGTACCAAGACCAAGGNTTCCAGGCTCG	CACCGC	ACGT	SFTQEDRQLEPL
					CGCGCGCGTGGMGCTCCCAGCCAGCGACGC	CCAAGT	AGGC	TRHVDEETKDS
					GTTCAGCAGCAGGGTCGGCTGCGGCAGCTC	CTCCGC	CGAA	ELPGRAPTVLDE
					GACGCTGCACGGGGACGGCAGCGACGGCA	CGCAGC	AGGG	ESKDNDATAGE
					GGCACCAAGACCACGACGACCAAGCCGTCG	TGCGAC	CCCC	WLLRFDGACRA
					CTGCCGTCCCCATGGACGTCGACCAAGGTG	TGCTGC	ACCC	NPGPGGAGAVL
					CTCGGTGGCCGACAGCGGATCAMCACCCGC	TGCTCA	ATGT	FNPSGAASWTC
					TGTCGCCGCCACCGGAGCTCCGAGTCGGCG	TCGAGT	AGGG	SHFMPGATETN
					GCAAGCGCCGCCGCCTGAACGACGGCGACG	CGCAGT	AACC	NTAEYTAFLLG
					ACGAAGACATCCGCGAGCTGGCCGAGCTTC	AGTCGC	GCCC	ARAAADHGATL
					TGCTGCCCGACGAGGAGGAGGCCGACGACC	AGCTGC	TCGG	LRVQGDSQLVL
					ACAAACCAGCGCCCAGGTTACCCGCGACCA	ACAAGC	GAAA	RQVKGIYGAKS
					GCGCTCATCCGGCCTCCGTCCTCGCTGTGTA	ACCACC	TCCA	TRLRRLRDAVR
					CGCGCACAACGCGCAGCGCTTCAACTGCAC	TCCTGC	TTCC	AELARVGQFSL
					GTTGTGCGTGTACACGGCTGCCAGCTTCGCT	CGGACG	GGTG	HHIDRQDNAHA
					GCTCTTACGCGACACAGGGACTCTCGGCAC	CGCCGC	TTCG	DRLANRALDMK
					CGGCGCGTGACCTTCCTGGACAGGTTCTCG	CGTGGA	ACTG	STLVECATHPGR
					GCGGGTTGCGCGTGCGGCAAACCTTTTGCC	GCACCA	AAGA	NACTTTLTTSAA
					TCGAGGCTGGCCGCAGCAAGACACGCACAA	CGCGCG	GATG	AESPASPPPVGA
					ACGTGCGCCAGCCTCAGCACCACACTGGTC	CGCTGA	CTCCT	RDTPMADAGEE
					GCGGTTTCGACGACAGGAGGAGCATCAAGC	GCCGTA	TCGC	RLADVDDGEVY

					CACACTGTCGTCGGAGCCAACACCACCGTC	CCAAGA	CTTG	AAMRLGPGEVP
					GCCACGGCGGTCACCGCCGAACCCCCCCTG	CCAAGG	ACGG	ERRPRLRLRQLS
					CTCCACCATCAAGCCTCGGAACTCACTGTG	NTTCCA	AGGT	DEELEAASEMV
					CCCCCCCCCACGTGTGAGTTCCCCCGACGTC	GGCTCG	ACAT	ERLGAALSAKIT
					GATGTGCAGCTGCACAGTCCGCCACAGGAA	CGCGCG	CTCG	DAEDWASAEGY
					GATCAGCACGAGGACGCCACCCANCACCCG	CGTGGM	ACAG	ITALPYMLYDKL
					GAGAGCACGCAACACCAACCTCCTGAGGCA	GCTCCC	TCGA	QSYSQAPRGPQ
					ACCCGCTGGGGTTCGCCGCTCGCGCCCACG	AGCCAG	ACTT	QPVLTRSPRGDD
					CTCGTTGCCTCCAGGATTGCTCAGCGACTCG	CGACGC	CAAC	RPASSEPNASST
					GCGAGCTGGAACCTCCACGCTGGGGCCCAC	GTTCAG	TCGC	TGGVASEHQPR
					CATTACCCCGCGCGATGGTTGTTTCGCGAAT	CAGCAG	AACA	RRRRRGRRKGR
					TACAGCACGGCTTGAAGCTACACCTGCTCC	GGTCGG	TATC	RQRRNPRRSGR
					TCGATGGGATCCGCCACTCCCACGCCGCGT	CTGCGG	CGAT	EGATGGHQQHK
					CATCGCCTCCCGCATTGCCGACCGGCTTGTA	CAGCTC	ACAG	KHKPRPPRETQH
					CCTGCGACGGCACCGTGTCGCTCGGCCCTG	GACGCT	TTAC	HREHRLDEALD
					AACGCTGCATTCCCTTCTCCATCGCGAGACA	GCACGG	AAAC	ELHALERTDPH
					CGGTGACGGAATCGTTCACCCAGGAGGACA	GGACGG	CACA	NRPAIAKARRR
					GGCAACTGGAACCCCTGACGCGGCACGTGG	CAGCGA	GTTA	VGRIRSAINQQL
					ACGAAGAGACGAAGGACAGCGAGCTGCCG	CGGCAG	GATA	LRHKFDTDEKA
					GGCAGGGCGCCGACTGTACTCGACGAGGAG	GCACCA	TCAG	CVDGILSTARAE
					TCCAAGGACAATGACGCCACTGCCGGTGAG	AGACCA	ATAG	RAARAATPSPPA
					TGGCTCCTCCGGTTCGACGGAGCTTGTCGA	CGACGA	GAAC	SGAPTTTVSAPG
					GCGAACCCAGGCCCAGGCGGAGCAGGTGCC	CCAAGC	CTTCC	AIVTNDDGTCPI

					GTGCTATTCAACCCTAGCGGTGCAGCGTCA	CGTCGC	TTTA	PSDKLWRHFDA
					TGGACCTGCTCGCACTTTATGCCGGGCGCC	TGCCGT	GGAA	VNTPRLDFDAE
					ACGGAGACAAACAACACCGCCGAGTACAC	CCCCAT	GCTA	APGSAAFRAAM
					GGCATTCCTCCTAGGCGCCAGAGCTGCTGC	GGACGT	ACGG	DHLPAATRLLD
					AGATCACGGTGCGACGCTACTTCGCGTCCA	CGACCA	GTAC	LLKEAPSTDEIE
					GGGGGACAGCCAGCTCGTACTCAGGCAGGT	AGGTGC	ACTG	TQLQHVKASSSP
					CAAGGGCATTTACGGTGCGAAGAGCACCCG	TCGGTG	GATG	GLDGVGYDVYK
					TCTACGGAGACTACGCGACGCTGTTCGTGC	GCCGAC	GTAA	RFTIQLLPVLRA
					CGAACTGGCTCGGGTGGGGCAATTCTCCCT	AGCGGA	ATAC	AFRCCWLYKKV
					CCACCACATCGACAGGCAGGACAATGCGCA	TCAMCA	ACAT	PQSWKLGVVRL
					CGCAGACCGCCTTGCGAACAGAGCGCTCGA	CCCGCT	ACAT	LHKKGPREDPA
					CATGAAGAGCACACTAGTGGAATGCGCAAC	GTCGCC	TTC	NWRPICLQQAIY
					GCACCCCGGCCGGAACGCCTGCACCACGAC	GCCACC	(SEQ	KLYTGILARRLT
					ACTGACCACCAGTGCTGCCGCTGAATCACC	GGAGCT	ID NO:	RWMDANDRHA
					GGCATCACCCCCGCCGGTGGGAGCTCGAGA	CCGAGT	1311)	PGQKGFRAVNG
					CACGCCGATGGCCGATGCCGGCGAAGAGCG	CGGCGG		CGEHNFLAATLI
					CCTCGCAGACGTTGACGATGGTGAAGTTTA	CAAGCG		DNARRKHRPLY
					CGCGGCGATGCGGCTTGGCCCAGGCGAAGT	CCGCCG		EVWYDFRNAFG
					TCCTGAACGACGGCCACGACTCCGCCTCCG	CCTGAA		SVPFALLWDSL
					CCAGCTGAGCGACGAGGAACTGGAAGCAG	CGACGG		QRLGVPPDYVD
					CGAGCGAGATGGTGGAGCGACTGGGCGCCG	CGACGA		MCKGLYNQASF
					CCTTATCGGCCAAGATCACAGACGCGGAAG	CGAAGA		VVGNAVDGSTA
					ACTGGGCGAGTGCGGAGGGCTACATCACGG	CATCCG		PVEQRVGVFQG

					CCCTACCGTACATGCTGTACGACAAGCTTC	CGAGCT		CPLSPQLFNAAI
					AGTCGTACTCACAGGCACCGCGTGGACCCC	GGCCGA		SPLLYALRRLPD
					AACAACCGGTCTTGACCCGGTCCCCGCGCG	GCTTCT		TGVQLSSVDRP
					GCGACGATCGCCCAGCTTCCTCGGAGCCCA	GCTGCC		GASAYADDLKIF
					ACGCATCATCCACGACTGGAGGGGTGGCAT	CGACGA		SGTKAGITQQHE
					CGGAGCACCAACCGCGGCGCCGCCGACGAC	GGAGGA		LVATFLRWTGM
					GAGGACGACGCAAGGGCCGACGACAGCGC	GGCCGA		QANPAKCRSMG
					CGCAACCCGCGCCGATCGGGACGCGAGGGA	CGACCA		VRRNTNGAVEA
					GCTACAGGTGGGCATCAACAACACAAGAAA	CAAACC		DNVHLELDDTPI
					CACAAACCACGCCCACCCCGCGAGACGCAG	AGCGCC		PSMTHMQSYTY
					CATCATCGGGAACACCGTCTCGACGAGGCT	CAGGTT		LGIGDGFDHVR
					CTCGATGAGCTCCACGCCCTCGAGCGTACC	ACCCGC		RRVELAPKLKTL
					GATCCTCACAACCGGCCAGCAATTGCCAAG	GACCAG		KHDVTALVESG
					GCCCGCCGCCGCGTCGGTCGCATCCGGTCT	CGCTCA		LAPWQVVKAV
					GCCATCAACCAGCAGCTACTCCGCCACAAG	TCCGGC		KVYLYPRVEYA
					TTCGACACGGACGAGAAGGCCTGTGTCGAC	CTCCGT		LRHLRPDDQHL
					GGCATCCTGTCTACAGCCCGCGCCGAACGC	CCTCGC		ESFDLHLRRGLR
					GCTGCACGAGCAGCAACGCCGTCGCCACCA	TGTGTA		HLLRLPKNATN
					GCATCAGGAGCGCCCACCACGACAGTCTCG	CGCGCA		EFFYAPVSRGGL
					GCGCCCGGTGCCATCGTGACCAACGATGAT	CAACGC		GLLPLVELHAAL
					GGCACGTGCCCCATCCCGAGTGACAAGCTT	GCAGCG		QIAHGWQTLHS
					TGGCGACACTTCGACGCCGTGAACACACCG	CTTCAA		PDPAIRRVAREQ
					CGCCTTGACTTCGATGCGGAGGCACCCGGG	CTGCAC		LYQIADARHRL

					AGTGCGGCCTTCCGAGCGGCGATGGACCAC	GTTGTG		DKDHWPHRREE
					CTGCCCGCGGCGACACGGTTGCTGGACCTG	CGTGTA		LCELLLNGELGT
					CTCAAAGAGGCGCCGAGCACGGACGAGATC	CACGGC		SAHAPPKRRNG
					GAGACGCAACTCCAGCACGTGAAGGCCTCC	TGCCAG		DIGSLWVDVRK
					TCCAGCCCCGGCCTCGACGGCGTCGGCTAC	CTTCGC		NLKTFGLKVAT
					GACGTGTACAAGCGCTTCACGATACAGTTG	TGCTCT		APANQETGVPA
					CTGCCTGTGCTGCGGGCGGCGTTCCGGTGCT	TACGCG		QPLQLRVPHHA
					GCTGGCTGTATAAGAAGGTTCCTCAGAGCT	ACACAG		EWLDHGNVLRH
					GGAAGCTCGGCGTGGTGCGCTTGCTGCACA	GGACTC		VKLHIKNLHWQ
					AGAAGGGCCCGCGTGAGGACCCTGCGAACT	TCGGCA		TWCALSDQGKT
					GGAGACCCATCTGTCTCCAGCAAGCCATCT	CCGGCG		ARVHGGVGSAF
					ACAAACTATACACGGGCATCCTCGCGCGGC	CGTGAC		LTRPRGMWESD
					GACTAACCCGGTGGATGGACGCGAATGACC	CTTCCT		YRFAVAARLNV
					GCCACGCACCCGGTCAGAAGGGATTCCGAG	GGACAG		VDTVNTLSRRR
					CGGTGAACGGCTGCGGGGAACACAACTTCC	GTTCTC		LRAHDRCRYPA
					TCGCAGCTACGCTCATCGACAACGCCCGCC	GGCGGG		CRWKETLAHVL
					GCAAGCATCGGCCACTGTACGAAGTCTGGT	TTGCGC		NHCPGTMDAVR
					ACGACTTCCGTAACGCGTTCGGCAGCGTGC	GTGCGG		GRHDDALKEIE
					CGTTCGCGTTGCTGTGGGATTCGCTACAGCG	CAAACC		HTLRASSGDRRE
					CCTCGGTGTCCCGCCCGACTACGTGGACAT	TTTTGC		LRVNQTVPGLP
					GTGCAAGGGACTCTACAACCAAGCCTCGTT	CTCGAG		GPPLRPDIQVYN
					CGTCGTCGGGAACGCAGTGGATGGCAGCAC	GCTGGC		HDKRTVAVVDL
					AGCGCCGGTAGAGCAGCGCGTCGGCGTTTT	CGCAGC		AVAFDEQPSEDP

					TCAAGGCTGTCCGCTCAGCCCCCAGCTCTTT	AAGACA		ESSGLAKAVQIK
					AACGCCGCCATCAGCCCACTGCTCTATGCG	CGCACA		KAKYAGIKEHL
					CTGCGCCGGCTACCAGATACCGGCGTGCAG	AACGTG		ENQGWKVHLSA
					CTCTCCAGTGTTGACCGCCCTGGTGCCTCAG	CGCCAG		IVYGSLGSVAAS
					CGTACGCCGACGACCTTAAGATTTTCAGCG	CCTCAG		NHKVYTEHLGL
					GCACCAAGGCGGGCATCACGCAGCAACACG	CACCAC		LKRDAKRLDRQ
					AGCTGGTGGCGACGTTCCTGCGGTGGACCG	ACTGGT		LSSACIQSSRRI
					GGATGCAGGCCAACCCTGCCAAGTGCAGGT	CGCGGT		WNFHCAKHRAR
					CCATGGGCGTGCGGCGCAACACCAACGGM	TTCGAC		QHEHQAPPSQA
					GCCGTCGAGGCAGACAACGTCCACCTCGAG	GACAGG		TRGRRVTETGG
					CTCGACGATACGCCGATCCCCTCCATGACA	AGGAGC		NPSRTDRR (SEQ
					CACATGCAATCCTACACGTACCTGGGGATC	ATCAAG		ID NO: 1433)
					GGCGACGGCTTCGATCACGTGCGCCGCCGC	CCACAC		
					GTCGAGCTCGCTCCCAAGCTGAAGACTCTC	TGTCGT		
					AAGCACGACGTGACGGCGCTCGTGGAGTCC	CGGAGC		
					GGCCTTGCGCCATGGCAGGTGGTGAAGGCC	CAACAC		
					GTGAAGGTGTACCTCTACCCTCGGGTAGAG	CACCGT		
					TACGCCCTCCGGCACCTCCGACCCGATGAT	CGCCAC		
					CAGCACCTGGAGAGCTTCGACCTCCACCTT	GGCGGT		
					CGCCGCGGCCTCCGCCACCTGCTTCGCCTTC	CACCGC		
					CCAAGAACGCCACGAACGAATTCTTCTACG	CGAACC		
					CGCCGGTGTCGCGCGGTGGTTTGGGGCTGC	CCCCCT		
					TGCCGCTGGTGGAACTGCACGCAGCACTGC	GCTCCA		

					AAATCGCGCACGGATGGCAGACGCTCCACT	CCATCA		
					CACCCGACCCAGCCATCCGTCGCGTCGCGC	AGCCTC		
					GCGAGCAGCTCTACCAGATTGCCGACGCGC	GGAACT		
					GGCACCGGCTCGACAAGGACCACTGGCCAC	CACTGT		
					ACCGGAGAGAGGAACTTTGCGAGCTCCTCC	GCCCCC		
					TCAACGGGGAGCTCGGGACGTCGGCCCACG	CCCCAC		
					CACCGCCCAAGCGACGCAACGGTGACATCG	GTGTGA		
					GCTCTCTGTGGGTCGACGTCCGCAAGAACC	GTTCCC		
					TCAAGACGTTTGGACTGAAGGTGGCCACAG	CCGACG		
					CGCCGGCCAACCAGGAGACCGGTGTACCAG	TCGATG		
					CCCAGCCGCTGCAGCTCCGCGTGCCTCACC	TGCAGC		
					ACGCCGAGTGGCTGGACCACGGCAACGTGC	TGCACA		
					TGAGGCACGTGAAGCTACACATCAAGAACC	GTCCGC		
					TCCACTGGCAAACCTGGTGCGCCCTAAGTG	CACAGG		
					ACCAAGGGAAGACCGCACGGGTGCACGGC	AAGATC		
					GGTGTTGGCAGCGCATTCCTGACCCGACCT	AGCACG		
					CGCGGCATGTGGGAAAGCGACTACCGCTTC	AGGACG		
					GCGGTGGCCGCTCGCCTCAACGTGGTCGAC	CCACCC		
					ACCGTCAACACACTGTCGCGCCGGCGACTC	ANCACC		
					CGCGCCCACGACCGGTGTCGGTACCCGGCC	CGGAGA		
					TGCAGGTGGAAGGAGACTCTGGCACACGTC	GCACGC		
					CTCAACCACTGCCCGGGTACCATGGACGCG	AACACC		
					GTACGTGGCCGGCACGATGACGCGCTCAAG	AACCTC		

					GAAATCGAACACACGCTCCGCGCGTCGTCG GGCGACCGCCGCGAGCTGCGTGTTAACCAA ACGGTGCCTGGCCTGCCCGGGCCACCCCTG CGCCCCGACATCCAGGTCTACAACCATGAC AAACGCACGGTCGCAGTTGTAGACCTGGCT GTGGCGTTCGATGAGCAGCCGAGTGAGGAC CCAGAGAGCTCGGGGTTGGCCAAGGCAGTC CAGATCAAGAAAGCCAAGTACGCGGGCATC AAGGAGCATCTCGAGAACCAGGGGTGGAA GGTTCATCTCTCGGCCATCGTGTACGGCTCG CTTGGCTCGGTGGCGGCCAGCAACCACAAG GTGTACACTGAACACCTTGGGTTGCTCAAG CGCGATGCAAAGCGGCTGGACCGGCAGCTC TCGTCGGCATGCATCCAGTCCAGCCGCCGC ATCTGGAATTTCCACTGCGCCAAGCACCGA GCACGCCAGCACGAGCACCAAGCACCACCA AGCCAAGCGACACGGGGCAGGAGGGTGAC GGAAACCGGGGGGGAATCCGTCGCGCACCGA CCGCCGCTAGACGGCACAGTTCTGGCCCAC GTAGGCCGAAAGGGCCCCACCCATGTAGGG AACCGCCCTCGGGAAATCCATTCCGGTGTT CGACTGAAGAGATGCTCCTTCGCCTTGACG GAGGTACATCTCGACAGTCGAACTTCAACT	CTGAGG CAACCC GCTGGG GTTCGC CGCTCG CGCCCA CGCTCG TTGCCT CCAGGA TTGCTC AGCGAC TCGGCG AGCTGG AACCTC CACGCT GGGGCC CACCAT TACCCC GCGCG (SEQ ID NO: 1188)		

					CGCAACATATCCGATACAGTTACAAACCAC AGTTAGATATCAGATAGGAACCTTCCTTTA GGAAGCTAACGGGTACACTGGATGGTAAAT ACACATACATTTC (SEQ ID NO: 1065)			
NeSL	Utopia-1_PI	.		Phytophthora infestans	AGCTCGGCCTCGCGGCTGCCTTCCCAGGCG	AGCTCG	TAAG	MLADPAALAAG
					CCGCCGACTTCGCGCTCTGGCGCGGCCCAC	GCCTCG	CTGG	LARAPPPPSAPQ
					ACGCCGCCGCCGAGCCTCCAAGCGCGCCCG	CGGCTG	TCAT	DPSPAFPAGPAG
					TTGGCTTTCGCAGACGCAGGGCTCGCGGCG	CCTTCC	CATG	QNPRAAAPARV
					ACGGCCCTCGCCAGCCCCCAAGACCCCCCC	CAGGCG	ACCG	EVHTVVAPPGR
					TACGATGTGGCACCACCCGGCAGGGCGGCC	CCGCCG	ACAG	AGGMLPDPGLV
					GGCAGGCTGCCCGACTCGGTATCGCCGGGT	ACTTCG	GGCA	EEPIQATYAHDA
					GCTACACTCTCAGCCGCCACAGCTCGGGCC	CGCTCT	CTAC	AQFECALCPYV
					TTGGCGGTCCGCCATTGGCCCTTGGAGCTCG	GGCGCG	CCAG	AESMAVLVQHR
					ACAGCGACAGCAGCGACGACGAGGACGCT	GCCCAC	GTAG	RSAHRGTRFKDI
					CAAGACCCCCACGCCGCCGCCCCAGAACCC	ACGCCG	GGAA	FTSGCQCSLVFY
					CCAGAAGACGTCGCGAGTGTGCTTGCCCCA	CCGCCG	CCGC	ARIVAASHAVA
					CCCGGCAGGGCAGGCAGCATGCTCGCCGAC	AGCCTC	CCTT	CARRNQRAVPP
					CCAGCAGCGCTGGCAGCCGGTCTGGCTCGC	CAAGCG	AAAA	APTPVAPTRPEA
					GCTCCGCCGCCCCCATCGGCACCCCAGGAC	CGCCCG	AACC	TPQPTGYLAAA
					CCAAGCCCCGCCTTCCCTGCTGGCCCCGCCG	TTGGCT	CAGG	MTAAAAAASSD
					GCCAGAACCCTCGCGCCGCCGCCCCAGCCC	TTCGCA	AAGA	TVVAAATNMQS
					GTGTTGAAGTCCACACAGTCGTGGCACCAC	GACGCA	CACA	AVPAAAKTTGL
					CCGGCAGGGCGGGCGGCATGCTGCCCGACC	GGGCTC	AACA	QLVPPELEPALP

					CGGGGTTGGTTGAAGAACCTATCCAAGCGA	GCGGCG	CCCT	QRASCHAGKRR
					CGTACGCGCACGACGCTGCCCAGTTCGAGT	ACGGCC	CCAC	RLNADEAVTPC
					GTGCGTTGTGCCCGTACGTCGCTGAGAGCA	CTCGCC	TTAG	TPTARVSPQTEV
					TGGCAGTGCTCGTTCAGCACCGACGGTCGG	AGCCCC	TGAC	AMAPHDAPQDD
					CTCACCGCGGCACCCGCTTCAAGGACATCT	CAAGAC	ATAC	TVLQREAAEPQP
					TCACCAGCGGGTGCCAGTGCAGCCTAGTGT	CCCCCC	ATAT	DPAATPGAQVQ
					TTTACGCGCGCATTGTTGCCGCCAGCCATGC	TACGAT	TTTA	RVEDTTAAQDD
					AGTGGCCTGTGCCAGGCGCAACCAGCGAGC	GTGGCA	GCCT	TVQQDHDADTA
					CGTGCCACCAGCTCCGACCCCTGTTGCCCCC	CCACCC	AGAT	QVSPPRRTPTRW
					ACGCGGCCCGAGGCTACCCCACAACCTACT	GGCAGG	TTCA	GPRPSSTQEPSP
					GGGTACCTCGCCGCGGCGATGACTGCCGCT	GCGGCC	GTTA	MTGEPAATLAA
					GCGGCAGCCGCATCAAGCGACACCGTAGTG	GGCAGG	CGGA	RRPLTPAATGTR
					GCAGCCGCCACCAACATGCAGAGCGCCGTG	CTGCCC	GAGG	ATRWGPCHRAI
					CCAGCCGCCGCCAAGACCACTGGGCTTCAG	GACTCG	TTACT	GAAAIARLVTG
					CTCGTCCCCCCCGAGCTGGAGCCTGCCCTTC	GTATCG	AACT	LSTEPAQPQRRQ
					CGCAGCGTGCTTCTTGCCATGCAGGCAAAC	CCGGGT	GGTA	PPPPQEPPSQPEP
					GCCGTCGGCTCAATGCCGACGAGGCGGTGA	GCTACA	AATA	LAAAATAAADI
					CACCGTGTACGCCTACGGCGCGGGTGTCGC	CTCTCA	CGAA	AATVAADIAAA
					CGCAAACAGAAGTGGCGATGGCGCCTCACG	GCCGCC	CACA	AANAAMDVDG
					ACGCACCCCAGGACGACACCGTCTTACAGC	ACAGCT	TATTC	GPAADETWLLR
					GTGAGGCCGCCGAGCCACAGCCAGACCCAG	CGGGCC	TGTTC	FDGACRRNPGP
					CAGCAACGCCGGGTGCCCAAGTGCAGCGAG	TTGGCG	TAAT	GGAGAALFAPS
					TCGAAGACACAACTGCGGCGCAGGATGACA	GTCCGC	CAGT	GAVVWTCSHY

					CTGTGCAGCAGGACCACGACGCCGACACTG	CATTGG	GTGA	MPSRSETNNTAE
					CCCAGGTCTCGCCGCCCCGGCGCACACCCA	CCCTTG	AAAC	YTALLLGVQSA
					CACGCTGGGGACCAAGACCAAGCTCAACTC	GAGCTC	TGGT	VHHGASHLEVE
					AAGAGCCTAGCCCCATGACAGGCGAGCCGG	GACAGC	TTTCG	GDSSLVIAQVKG
					CTGCCACCCTGGCAGCGCGAAGACCACTGA	GACAGC	CCTTT	TFACRNARLRQ
					CTCCTGCCGCAACGGGCACTCGTGCCACGC	AGCGAC	TGGC	LRNRVRHALRS
					GGTGGGGGCCATGTCACCGCGCGATCGGGG	GACGAG	GGAC	VDTHKLRHIDR
					CCGCAGCCATCGCACGACTTGTCACGGGCC	GACGCT	TTTTT	QANAHADRLAN
					TTTCAACGGAGCCAGCGCAGCCTCAGCGAC	CAAGAC	CACT	RALDQRRTSSEC
					GGCAGCCACCGCCACCACAAGAACCTCCAT	CCCCAC	CGCA	GTHGSCMDSCL
					CGCAACCAGAGCCGCTAGCAGCAGCAGCAA	GCCGCC	TTTTT	AVPTALAAQET
					CAGCCGCAGCAGACATAGCAGCAACAGTTG	GCCCCA	GGGC	PPAAPPSTSATP
					CAGCAGACATTGCGGCAGCGGCAGCGAACG	GAACCC	AATC	AEGNAMDDIAA
					CAGCAATGGACGTGGATGGTGGACCGGCAG	CCAGAA	GTCT	EIAARDEGETFP
					CCGACGAGACATGGCTGTTGCGCTTTGACG	GACGTC	GCGG	VLPIGPGSAPER
					GCGCATGCCGACGGAACCCAGGACCTGGTG	GCGAGT	CTAG	QPRLRLRQLSDE
					GAGCAGGTGCCGCACTCTTTGCGCCCAGCG	GTGCTT	CTTG	ERDAAADALQE
					GTGCCGTGGTGTGGACGTGCTCGCACTACA	GCCCCA	CTAG	LADTMASKIED
					TGCCAAGCCGCTCCGAAACGAACAACACCG	CCCGGC	CGGC	ADSWTSGEGYIS
					CCGAGTACACGGCGTTACTGCTCGGCGTAC	AGGGCA	GGAC	SIPERIREVLQPY
					AGAGCGCAGTCCACCACGGTGCCTCGCACC	GGCAGC	GAGC	ATAPPQPGRSRR
					TGGAAGTCGAGGGCGACAGCAGCCTGGTAA	(SEQ ID	GGTC	QQRRRPPRVTR
					TCGCTCAGGTAAAGGGCACCTTCGCATGCA	NO: 1189)	TCCG	NQREHRLDEAL

					GGAATGCGCGGCTCCGCCAACTCCGCAACC		GGGG	DDMAATQQATP
					GAGTGCGTCACGCGCTCCGCTCCGTGGACA		CGTT	RDQRAVRRARR
					CGCACAAACTCCGGCACATCGACCGCCAGG		CACC	RVGRVRASMAQ
					CCAACGCGCATGCAGATCGCCTTGCCAACC		TTTCC	QELRHEFAKDES
					GCGCGCTAGACCAGCGGAGGACCAGCAGC		CCCG	KCVAKILKTAST
					GAGTGCGGGACGCATGGCTCGTGCATGGAC		CGAG	ETAAEDEHPETC
					TCCTGCCTCGCCGTACCCACGGCCCTGGCG		GCCA	PIDAATLHAHFT
					GCGCAAGAAACGCCACCGGCAGCACCACCC		ACTA	GVNAPRTDFDY
					AGCACCAGTGCCACTCCAGCCGAAGGTAAT		CACC	DATSGREFRAA
					GCCATGGACGACATCGCAGCCGAGATTGCT		GATC	MSDLPPATVEID
					GCGCGTGACGAGGGCGAGACCTTCCCAGTC		TTCTC	AFDAELTIDEVE
					CTGCCAATTGGCCCAGGCTCAGCCCCAGAA		TACA	DQLTRAAKTSSP
					CGGCAACCCCGACTCCGTCTGCGCCAACTT		CTTTT	GHDGIDYGIYSR
					AGCGATGAAGAGCGCGACGCAGCGGCAGA		CTAA	FAAQLVPLLHA
					TGCCCTGCAGGAGCTCGCCGATACCATGGC		TTCG	VFQFCWRHRRV
					GAGTAAGATCGAGGATGCTGACAGCTGGAC		CCTC	PRLWKVGIVRLI
					CTCCGGCGAAGGATACATCAGCTCCATCCC		CGTC	HKKGDPRQPTN
					CGAGCGTATCCGAGAGGTACTGCAGCCATA		TTCG	WRPICLQPTIYK
					CGCTACGGCCCCGCCGCAGCCGGGACGCTC		GTCTT	LYSGLLAHRLSR
					TCGCCGGCAACAACGCCGTCGACCACCACG		CGGT	WLEGNDRLPMA
					GGTGACGCGCAACCAGCGCGAACATCGACT		TGTC	QKGFRAFNGCH
					CGACGAAGCTTTGGACGACATGGCAGCCAC		GGGC	EHNFMATTLLD
					ACAGCAAGCGACCCCGCGAGACCAGCGGG		TTTTT	QTRRQHRKLYQ

					CTGTGCGCCGTGCGCGGCGACGTGTGGGCC		TCTTT	VWYDLRNAFGS
					GGGTGCGAGCCTCAATGGCACAACAGGAGC		TTGA	LPQQLMWRVLR
					TGCGCCACGAGTTTGCCAAGGACGAGTCTA		CCAA	HLGVDSGFIDRC
					AGTGTGTGGCCAAGATCCTGAAGACAGCGT		TCAG	RDIYRDSAFVVA
					CGACAGAGACGGCAGCGGAGGACGAGCAC		AGCG	NAADGATDPVR
					CCGGAGACGTGCCCCATTGATGCAGCTACG		CGCC	QEVGVYQGCPL
					CTCCACGCCCATTTCACGGGAGTCAACGCG		ATGC	SPLLFVAALVPL
					CCAAGGACGGACTTCGACTACGATGCCACC		GCCT	VRRLEKLDGVG
					AGCGGCCGGGAGTTTCGAGCGGCGATGAGC		CTTCT	VPLADGVRPCT
					GACCTTCCGCCAGCCACCGTGGAGATTGAT		GGCC	TAYADDLKVFS
					GCGTTCGACGCGGAGCTGACGATCGACGAG		AATC	DSAAGIRKCHD
					GTGGAGGACCAGCTCACACGCGCGGCCAAG		AGAG	TVAGFLAWTGL
					ACCTCAAGCCCCGGCCACGACGGCATCGAC		ACCG	RANPGKCASLA
					TACGGCATCTACAGCCGCTTCGCCGCGCAG		GGCC	VTTNARGNPTR
					CTCGTTCCGCTCCTCCACGCGGTGTTCCAGT		CTGT	DSSMRLEVHDA
					TTTGCTGGAGACACCGTCGTGTGCCAAGAC		CCTC	AITTLSLHESYR
					TGTGGAAGGTGGGCATCGTACGCCTCATCC		GGAC	YLGVGDGYDHV
					ACAAGAAGGGCGACCCAAGGCAGCCGACT		AGCG	RHRLQLEPKLK
					AACTGGAGGCCGATCTGCCTTCAGCCCACG		AGGC	QLKREAVALLT
					ATCTACAAGCTCTACAGCGGGCTGCTGGCG		CTCC	SGLAPWQVVRA
					CATCGGCTCTCACGCTGGCTGGAAGGTAAC		ACGG	LKVYVYPKVEY
					GACCGCCTGCCGATGGCCCAGAAGGGGTTC		CCAG	ALRHLRPLQSQL
					AGGGCCTTCAACGGGTGCCACGAGCACAAT		CCAA	QAFDRVVSKGL

					TTCATGGCTACCACGTTGCTGGACCAGACC		TCGA	RHLLSLPRSATS
					CGCCGGCAGCACCGAAAGCTGTACCAAGTG		GTCT	EVLYAPTSSGGL
					TGGTACGACCTTCGCAACGCCTTCGGCTCGC		CGGC	GLQPLVELHRA
					TTCCTCAGCAGCTCATGTGGCGCGTGCTTCG		AGCG	LQLAHAWQML
					CCACCTTGGCGTCGACTCAGGCTTCATCGAC		ACGC	HSKDPAIQAVA
					CGCTGCAGGGATATCTACAGGGACTCGGCC		GTCTT	RAQACQVVRKR
					TTCGTGGTGGCGAACGCGGCGGACGGTGCG		TCTAT	YRLQEDHWRGR
					ACGGACCCGGTGCGCCAAGAGGTGGGCGTA		AGCG	DDELVRSFLNSE
					TACCAAGGCTGCCCACTCAGCCCGCTCCTGT		CAGC	LAASPHAEVLR
					TCGTCGCGGCGCTGGTGCCTCTCGTGCGCCG		TGAC	RNGDIASLWSD
					TCTCGAGAAACTCGACGGCGTTGGTGTGCC		GAGG	VQRWLRIYHLR
					ACTGGCGGACGGTGTGCGGCCCTGCACCAC		CCGA	FEHCDETEAHGP
					GGCCTACGCCGATGACCTCAAGGTGTTCAG		TCTG	LSFRVPHHNKW
					CGACAGCGCCGCCGGCATCCGCAAGTGCCA		GCGG	LTHKTVLRHVK
					CGACACAGTGGCTGGCTTCCTCGCGTGGAC		CCCC	LHLKIRHQTRW
					AGGGCTGCGCGCTAACCCTGGCAAGTGTGC		CGAT	KGMVDQGKTV
					CAGCCTCGCAGTGACGACGAACGCCCGGGG		TGGT	RVHGGVGAKF
					CAACCCGACTAGAGACAGCAGCATGCGTCT		CCGA	MTTGAGLSDDD
					CGAGGTCCACGACGCCGCCATCACCACGCT		CTTTC	YRFGVKGRLNQ
					CAGCCTCCACGAAAGCTACCGCTACCTGGG		GGCC	VDTNSVLKRKR
					CGTGGGTGATGGATATGACCACGTGCGCCA		AATC	LRAHTTCRDPTC
					TCGTCTCCAGCTCGAGCCAAAACTCAAGCA		AGCG	SSAETLAHVLN
					GCTCAAGCGGGAGGCGGTGGCGCTGCTGAC		ACGA	HCESNMDAIRQ

					GTCGGGTTTGGCACCGTGGCAGGTGGTGCG		CGAG	RHDDALEQIGSK
					CGCGCTCAAGGTGTACGTCTACCCCAAGGT		GGGG	IRGALDRAKSPT
					GGAGTACGCGCTGCGCCACCTGCGCCCGCT		CAGG	ELRLNQTVPEYT
					GCAGTCGCAACTGCAGGCGTTCGACCGTGT		GGTT	GAALRPDIVLRN
					CGTCTCCAAGGGCTTGCGACACCTGCTGAG		TACA	VAAKTMVIADL
					CCTGCCTCGCTCAGCCACCAGCGAGGTGCT		CTTTT	AVTFEDQAARA
					GTACGCCCCGACGTCCAGCGGAGGCCTGGG		GCCC	RHSSLQLSHDH
					CCTGCAGCCGCTCGTGGAGCTGCACCGCGC		CCGT	KTLKYQPIVAEL
					CCTACAGCTCGCCCACGCGTGGCAGATGCT		TTCG	QHKGWRVQTA
					GCATTCCAAGGACCCAGCCATCCAAGCAGT		ACTT	AIVYGTLGSVQP
					GGCAAGAGCGCAGGCTTGCCAGGTGGTGCG		CAAC	SNFKAYTEKFKL
					CAAGCGCTACCGCCTGCAAGAAGACCATTG		TTCA	HKREARQLDLQ
					GCGCGGGCGCGATGACGAGCTCGTGCGCAG		GGCC	LSSHCIRASHRI
					CTTCCTGAACTCCGAGCTCGCCGCCTCGCCA		AAAA	WGWHCRQHRD
					CACGCTGAGGTCCTCCGTCGCAACGGTGAC		TGGC	RQRSGTASRASR
					ATTGCCTCGCTGTGGTCCGACGTGCAGCGG		GATT	GSGGAPRRTSQ
					TGGCTGCGCATCTACCACCTGCGGTTCGAG		TGGA	APARR (SEQ ID
					CACTGCGACGAGACAGAGGCCCATGGCCCG		CCCT	NO: 1434)
					CTCAGCTTCCGCGTGCCGCACCACAACAAG		CCAC	
					TGGCTTACGCACAAGACTGTGCTGCGGCAC		GCGC	
					GTGAAGCTACACCTGAAGATCCGACACCAG		CGTG	
					ACCCGATGGAAGGGCATGGTGGACCAGGGC		CCAC	
					AAGACGGTGCGCGTGCACGGTGGGGTGGGC		TGCT	

					GCCAAGTTCATGACGACCGGAGCGGGCCTG		CGGC	
					TCGGACGACGACTACCGCTTCGGGGTGAAG		ACCG	
					GGCCGCCTCAACCAAGTCGACACGAACTCG		GCGG	
					GTGCTCAAGCGCAAACGACTGCGGGCACAC		CGAT	
					ACGACCTGCAGGGACCCAACCTGCTCCTCG		TCAG	
					GCTGAGACGCTGGCACATGTACTCAACCAC		CGGG	
					TGCGAGTCCAACATGGACGCCATTCGCCAG		TGCA	
					CGTCACGACGACGCCCTGGAGCAGATTGGG		ACTT	
					TCTAAGATCCGCGGAGCGCTCGACCGCGCC		CGGG	
					AAGTCTCCCACAGAGCTGCGCCTGAACCAG		CACG	
					ACTGTACCCGAGTACACAGGCGCAGCTCTA		TGTG	
					CGCCCAGACATTGTGCTTAGAAACGTGGCC		CAAC	
					GCCAAGACGATGGTGATCGCCGACTTGGCG		ACAT	
					GTCACGTTCGAAGACCAAGCGGCTCGCGCT		GCAG	
					CGTCACTCCTCGCTGCAGCTCAGCCACGAC		CGCC	
					CACAAGACTCTCAAGTACCAACCAATCGTG		CATT	
					GCGGAGCTGCAGCACAAGGGATGGCGAGT		GCAC	
					GCAGACTGCTGCCATAGTGTACGGCACGCT		GCCA	
					GGGCTCGGTGCAGCCCAGCAACTTCAAGGC		AGCG	
					CTACACTGAGAAGTTCAAGCTTCACAAGCG		GCAT	
					CGAGGCTAGGCAGCTGGATCTCCAGCTGTC		CGCG	
					TAGCCACTGCATTCGCGCCAGCCACCGCAT		GGAC	
					CTGGGGCTGGCACTGCCGACAGCACCGCGA		GACG	

					TCGGCAACGAAGCGGCACTGCCTCGCGAGC		CCTC	
					GTCGCGTGGGTCTGGGGGGGCCCCGCGGCG		GGCC	
					CACATCGCAGGCACCGGCACGGCGATAAGC		GCCC	
					TGGTCATCATGACCGACAGGGCACTACCCA		AAGC	
					GGTAGGGAACCGCCCTTAAAAAACCCAGGA		GCAG	
					AGACACAAACACCCTCCACTTAGTGACATA		CCCC	
					CATATTTTAGCCTAGATTTCAGTTACGGAGA		GCCC	
					GGTTACTAACTGGTAAATACGAACACATAT		TTCC	
					TCTGTTCTAATCAGTGTGAAAACTGGTTTTC		AGCA	
					GCCTTTTGGCGGACTTTTTCACTCGCATTTT		CGAC	
					TGGGCAATCGTCTGCGGCTAGCTTGCTAGC		CTCG	
					GGCGGACGAGCGGTCTCCGGGGGCGTTCAC		CGCC	
					CTTTCCCCCGCGAGGCCAACTACACCGATCT		GTTT	
					TCTCTACACTTTTCTAATTCGCCTCCGTCTTC		GGCG	
					GGTCTTCGGTTGTCGGGCTTTTTTCTTTTTGA		GATC	
					CCAATCAGAGCGCGCCATGCGCCTCTTCTG		GCCA	
					GCCAATCAGAGACCGGGCCCTGTCCTCGGA		TCAA	
					CAGCGAGGCCTCCACGGCCAGCCAATCGAG		GACG	
					TCTCGGCAGCGACGCGTCTTTCTATAGCGCA		TGCG	
					GCTGACGAGGCCGATCTGGCGGCCCCCGAT		AGAG	
					TGGTCCGACTTTCGGCCAATCAGCGACGAC		CCAG	
					GAGGGGGCAGGGGTTTACACTTTTGCCCCC		GCGG	
					GTTTCGACTTCAACTTCAGGCCAAAATGGC		GGTC	

					GATTTGGACCCTCCACGCGCCGTGCCACTG CTCGGCACCGGCGGCGATTCAGCGGGTGCA ACTTCGGGCACGTGTGCAACACATGCAGCG CCCATTGCACGCCAAGCGGCATCGCGGGAC GACGCCTCGGCCGCCCAAGCGCAGCCCCGC CCTTCCAGCACGACCTCGCGCCGTTTGGCG GATCGCCATCAAGACGTGCGAGAGCCAGGC GGGGTCGGGCAAAATATACTTACTCTAAGT ATGCCCGAATCCCTGCCCTCTCAGGCTGAA CGCGGCCCCATACTTGATCTAAGTATGGGA GGATCCCTGGCCTCTCAGGCTGTACGCGAG ACCC (SEQ ID NO: 1066)		GGGC AAAA TATA CTTA CTCT AAGT ATGC CCGA ATCC CTGC CCTCT CAGG CTGA ACGC GGCC CCAT ACTT GATC TAAG TATG GGAG GATC CCTG	

							GCCT CTCA GGCT GTAC GCGA GACC C (SEQ ID NO: 1312)	
NeSL	Utopia-1_PMi	.		Patiria miniata	CTGATGTGGATACCTTGGAATTACTCAACC GTGTCGGAGTCTTTTGTCTTTTGCGCCCAAC ACCTCATGGATACCATGCTTGTCGCTGGAG CGACGTTACAAGCGTGAGGGCGCCCTCCAT GCCGGACAGCTGGTCTGTGCCATGTGCTTG AAGTCGTTCTCCTCGACTTCGGGCTTGCGGC GGCACATGGCCCGGCTGCACCGGCAGCCTT CTCCCGACGCCTCCACGCCCTCGACTATGAC GGAAGTGTTCCCGTACCCCCTTCCGAAGGT GTGGCCGTGCGTGGTGTGCCGCGAGAACTT CTATCACAATCAAACCTTGAAGCGTCACCA GAAGAATTTTCACCCCAATGTTGACCTCACC ACCGTCTACCAGTGCTCAGTGTGCGGGCAG GAGTTTGTCACCGGCAGGAAAGCATCCTTC	CTGATG TGGATA CCTTGG AATTAC TCAACC GTGTCG GAGTCT TTTGTC TTTTGC GCCCAA CACCTC ATGGAT ACCATG CTTGTC	GATT AGCG AACA CTAA TATC CTGC CMTA GACG TGAT TGCT AATC CGCA AACC AACC	MCLKSFSSTSGL RRHMARLHRQP SPDASTPSTMTE VFPYPLPKVWP CVVCRENFYHN QTLKRHQKNFH PNVDLTTVYQC SVCGQEFVTGR KASFHFKVHRR MSASAIPTPSAM PSSPMDLIRGLV GEPLPPSPARTPP PLARYISPAPRSS FSPPWNPSPPPR

					CATTTCAAAGTTCATCGGCGGATGTCAGCC	GCTGGA	GGAT	SPTPLPRPLTPPP
					AGCGCCATCCCTACCCCGTCTGCAATGCCCT	GCGACG	CTAC	RSPSPPPRSPTPP
					CCTCACCGATGGATTTGATCCGAGGCCTCGT	TTACAA	AGCC	PPVTLTTAPVTE
					CGGGGAACCCCTGCCTCCTAGCCCTGCCCG	GCGTGA	TGAA	PAVPVALTTAQ
					TACACCGCCACCACTTGCCCGATACATCTCC	GGGCGC	CACT	VTEPSAPAVHTA
					CCAGCGCCCCGGTCTTCTTTTTCACCACCCT	CCTCCA	GAAC	APVTLSNAPVTE
					GGAATCCCTCCCCACCGCCCCGGTCGCCCA	TGCCGG	TTTA	PATPATDPATPV
					CCCCACTGCCACGGCCACTCACTCCACCGC	ACAGCT	ATCTT	TRLHSPVTHISC
					CACGATCACCGAGCCCGCCGCCCCGGTCGC	GGTCTG	CACC	SISFTASHAPYSC
					CAACACCACCTCCCCCGGTCACCCTCACCA	TGCC	CATG	AAPTSPSVYACS
					CCGCCCCGGTCACCGAGCCCGCCGTCCCGG	(SEQ ID	TCAC	PRRRQCSSTIAA
					TCGCCCTCACCACCGCCCAGGTCACCGAGC	NO: 1190)	ATCT	VCNSEASSGNPC
					CCTCCGCCCCGGCCGTCCACACCGCCGCCC		GGAC	LLALPVHRHHLP
					CGGTCACCCTCTCCAACGCCCCGGTCACCG		ACTA	DTSPQRPGLLFH
					AGCCCGCCACCCCGGCCACCGATCCTGCCA		GGTT	HPGIPPHRPGRP
					CCCCGGTCACCAGACTCCACTCACCGGTCA		TTTGC	PHCHGHSLHRH
					CTCACATCTCATGCTCGATCTCCTTCACGGC		TCTGT	DHRARRPGRQH
					CAGCCACGCCCCCTATTCCTGTGCAGCCCCG		TTGT	HLPRSPSPPPRSP
					ACATCTCCGTCCGTCTACGCCTGCTCACCCA		GTTTT	SPPSRSPSPPPRS
					GGCGTAGACAATGCAGTTCCACCATCGCAG		CCTG	PSPPPRPSTPPPR
					CAGTCTGCAATTGACGTGTGGCTGGCCGAG		CCTTT	SPSPTPRSPSPPP
					CTGTCACGCTCGGCTGACTTCGAGTCTTTTG		WCMT	RPPILPPRSPDST
					AGGATGTGTGTGATCGGTTTGTGGAGTTTGC		TGGA	HRSLTSHARSPS

					CGCCGCTGAGGGTCGAAACAACGGACGCCC		WCTT	RPATPPIPVQPR
					AGCCCGCCCAGCGCACCAACCGCCCCGTGA		TCGC	HLRPSTPAHPGV
					CCGCGGGAACCAGGGACCTCGCCCCCAACG		MCTG	DNAVPPSQQSAI
					TCCCCCTCGTCCCCACCGCCCCGGGCTTGGC		GAAT	DVWLAELSRSA
					CCGGAGTTTGATGCACAGGAGGCCAGTCGT		TTATT	DFESFEDVCDRF
					CTGCAAAAGCTGTATCGGACCAGCAAGAAG		TGTC	VEFAAAEGRNN
					CGCGCCATCAGGACTATCTTGACCGGCTCT		GCTT	GRPARPAHQPPR
					GATGTCCGATACAGCGGCTACCGAGGTCCT		GGAT	DRGNQGPRPQR
					CACTAGTGGATTTCTTTCACTTGAAGTATTT		TATTT	PPRPHRPGLGPE
					TCCGCTCGTGAGGTCGACACCGACACCATT		TTTTT	FDAQEASRLQK
					GCCACGGATACCAGCCTCCTGTTTCCGAACT		CTCA	LYRTSKKRAIRT
					CCGCCCAAGCCCGTGAGTCCGGCCAAGACC		CAAT	ILTGSDVRYSGY
					TGCTTCGCCCAGTCACCCAGCGTGAAGTCTC		TTGG	RGPHPALMSDT
					CCTCCGTCTTGGCCGGATGTCCAACTCCGCA		ATCT	AATEVLTSGFLS
					CCAGGGAAGGACCGGTTGGAGTACCGTCAC		ATTTT	LEVFSAREVDTD
					ATCCGCCAGGTTGATGGCGCATTCCGCGTC		CGTTT	TIATDTSLLFPNS
					ACCCTGGAAATCTTCAACCGGTGCCTGCGT		GTTC	AQARESGQDLL
					GAAAGCCGGGTCCCATCCTCCTGGAAGACA		ACTT	RPVTQREVSLRL
					GCGACCACCGTGCTGATCCACAAGAAGGGG		CGAA	GRMSNSAPGKD
					GATGCCACCGACCCGGCCAACTTCCGCCCC		CTCT	RLEYRHIRQVD
					ATCGCCCTCCAGTCCTGCCTCTACAAGCTGT		AGCT	GAFRVTLEIFNR
					TGATGGCGATCTTGTCCGACCGTGTCACCAC		GCCC	CLRESRVPSSWK
					ATGGGCCCTGGACAACGATCTCATCAGCAG		TTTCT	TATTVLIHKKGD

					CTCGCAGAAGAGCGCCCGCCCCGGTGAAGG		TCGG	ATDPANFRPIAL
					TTGCTACGAACACACATTTCTTCTCTCCACC		ACAC	QSCLYKLLMAIL
					GTCGTGAAGGATGCCCGGCGTAACCAGAAA		TGAA	SDRVTTWALDN
					AACATGTACGCCGCCTGGCTGGATTTACAA		CTTTA	DLISSSQKSARP
					AATGCATTCGGCAGCATCCCCCACGACGCA		ATCTT	GEGCYEHTFLLS
					ATGTTCACCGTGTTAACCTCCATCGGGGCCC		CGCC	TVVKDARRNQK
					CGGAGGGCCTGGTCTCCCTGGTGCGCGATG		ATGG	NMYAAWLDLQ
					TGTACACTGACGCCTCCACCGACTTCGTCAC		CTGT	NAFGSIPHDAMF
					ACCCACCGGCCGCACTGCTGCAGTCCCGAT		CAGT	TVLTSIGAPEGL
					TCACTCAGGCGTGAAGCAGGGCTGCCCGAT		CGCC	VSLVRDVYTDA
					CTCTCCAGTGCTTTTCAACCTGACGCTGGAA		GGTT	STDFVTPTGRTA
					CTTATTATCCGAGCAGTGAATGCTTCAGCG		CACT	AVPIHSGVKQG
					ACTCGGGACCGCAGTGCGCCGGTGGTCCAT		TGCT	CPISPVLFNLTLE
					GGCCAGGCCGTCCCCATACTTGCGTATGCT		GCGG	LIIRAVNASATR
					GACGACCTGGTCATACTCAGCCGATCCAGT		TGGG	DRSAPVVHGQA
					GACGGCCTGCAGAGCCTCCTGACAACAGCA		ATCC	VPILAYADDLVI
					TCCATCATGGCCACCAAGATCCAGCTGAAG		TGTT	LSRSSDGLQSLL
					TTCAAGCCCGCCAAGTGCGCTTCCCTCTCCC		GTGA	TTASIMATKIQL
					TCGAGTGCCGACGGGGGACCAAGGTACGCC		TAAT	KFKPAKCASLSL
					CTCTTGAGTTCAACGTGCAGGACAAGATCA		CCCC	ECRRGTKVRPLE
					TCCCCGCCCTTACAGAGGAGCAGCACTACC		GTGC	FNVQDKIIPALT
					GGTACCTTGGTGTCCCCATCGGCCTCTACCG		ATTG	EEQHYRYLGVPI
					CACCGACGACAGCCTGGAGACCTTGGTGGC		CCCA	GLYRTDDSLETL

					CAAGATGACTGATGACATCCAGCGGATTGA		TGGA	VAKMTDDIQRI
					CTCCTCACTCCTCGCTCCCTGGCAGAAACTG		TTTAT	DSSLLAPWQKL
					GATGCCATCCGGACCTTCGTACAGCCCTGC		TTCC	DAIRTFVQPCLA
					CTAGCCTACACCCTGCGGGCAGGGGACTGT		GCCT	YTLRAGDCAKK
					GCGAAGAAACACCTCAAGAGACTGCGGGG		TAGT	HLKRLRGQLVK
					ACAGCTGGTGAAGACAGCCCGCAAAGTATG		TGTTC	TARKVCNLPTR
					CAACCTGCCGACTCGGGCCACCACAAACTA		CTAA	ATTNYIFADRRA
					CATCTTTGCTGATCGTAGGGCGGGTGGCTTG		CCTT	GGLGFIDPNVDA
					GGTTTCATCGACCCAAACGTCGATGCCGAT		GGAT	DIQIITQAVRML
					ATTCAAATCATCACGCAGGCCGTACGCATG		TTATT	SSPDDITRAIATG
					CTGTCTTCACCCGACGACATCACCAGAGCC		GCTG	QLSSVVHRTIHR
					ATCGCCACCGGCCAGCTGTCTTCCGTGGTCC		TGGG	APTQEETDEFLS
					ACCGCACCATCCACCGAGCACCGACACAGG		TGAT	ASMEGDFANSG
					AGGAGACGGACGAGTTCTTGTCAGCATCAA		GCCC	NSGQASSLWSR
					TGGAGGGGGACTTTGCCAACAGTGGGAACA		GGGT	ARAAARRLKVT
					GCGGCCAGGCCTCTTCATTGTGGTCCAGAG		TTTGT	ISGSLSGSVITKS
					CACGTGCGGCGGCCCGGCGACTGAAAGTTA		TTAC	TENREMAAKSIT
					CCATCAGCGGCAGCCTTTCCGGCTCAGTGA		ATCG	TALRAQSRAHY
					TCACCAAGTCAACAGAGAACCGGGAAATGG		GGAT	THRLLSLPDQGK
					CTGCCAAGTCCATCACTACCGCCCTGCGTGC		CCCG	VGQSLNQDQYM
					TCAGTCCCGCGCCCACTACACCCACCGGCT		CTGC	NSSSWMSSGSYI
					GCTGTCCCTGCCCGACCAGGGGAAAGTTGG		GGCT	LFCDWRFIHRAR
					CCAGTCCCTCAACCAAGACCAATACATGAA		CGGT	LNTLPTNATAQ

					CTCGTCCTCGTGGATGTCCTCCGGTAGCTAC		GTTC	RWKPNTSPACR
					ATCCTGTTCTGTGACTGGAGATTCATCCATC		CATG	RCQHPQETLPHI
					GGGCCCGTCTCAACACCCTTCCGACCAACG		CGAC	LNHCPPNMVPIR
					CAACGGCGCAGCGATGGAAACCCAACACCA		TGGC	RRHNLVQQRIVS
					GTCCCGCCTGTCGCCGTTGCCAGCACCCGC		AGCC	AVRHGRVFVDQ
					AAGAGACCCTGCCCCACATCCTCAACCACT		CCTTT	HVPEDPNPRERP
					GCCCGCCCAACATGGTACCCATCCGCCGCA		GTTT	DITVVEGDKVTI
					GACACAACCTCGTCCAGCAACGCATCGTGT		ACTC	IDVCCPFDNGRD
					CTGCGGTACGCCACGGGCGGGTGTTCGTCG		TSGA	ALMTAAAAKET
					ACCAGCACGTCCCTGAGGATCCCAACCCCC		CTCT	KYADLKQALVA
					GGGAGCGCCCGGACATCACGGTCGTTGAGG		ATTC	AGKDVEVFGFA
					GCGACAAAGTCACCATCATCGATGTATGTT		ATTG	VGSLGSWLPSN
					GCCCCTTCGACAACGGCCGTGATGCCCTGA		TTGT	ERALRRLGIAKR
					TGACCGCCGCAGCCGCCAAGGAGACCAAGT		ATTTC	FRTLMRKLLRID
					ATGCCGACCTGAAGCAGGCGCTGGTGGCTG		TTCC	AIKGSRDVYIEH
					CGGGGAAAGACGTCGAGGTGTTCGGATTCG		ACAC	MCGHRQYT
					CCGTTGGATCGCTAGGCTCCTGGTTACCTAG		TGGC	(SEQ ID NO:
					CAACGAGCGGGCACTTCGACGTCTCGGCAT		CAGT	1435)
					CGCCAAGCGCTTCCGGACTCTGATGCGCAA		GATC	
					ACTCCTGCGCATCGACGCCATCAAGGGATC		ACAC	
					CCGGGACGTGTACATCGAGCACATGTGTGG		TGCT	
					CCATCGCCAATACACCGATTAGCGAACACT		GTGT	
					AATATCCTGCCMTAGACGTGATTGCTAATC		TTCCC	

					CGCAAACCAACCGGATCTACAGCCTGAACA		GGGA	
					CTGAACTTTAATCTTCACCCATGTCACATCT		AGAT	
					GGACACTAGGTTTTTGCTCTGTTTGTGTTTT		ATCC	
					CCTGCCTTTWCMTTGGAWCTTTCGCMCTGG		TCTG	
					AATTTATTTGTCGCTTGGATTATTTTTTTTCT		CGGT	
					CACAATTTGGATCTATTTTCGTTTGTTCACT		TTTCA	
					TCGAACTCTAGCTGCCCTTTCTTCGGACACT		CGCT	
					GAACTTTAATCTTCGCCATGGCTGTCAGTCG		CTGG	
					CCGGTTCACTTGCTGCGGTGGGATCCTGTTG		GTGT	
					TGATAATCCCCGTGCATTGCCCATGGATTTA		CTCC	
					TTTCCGCCTTAGTTGTTCCTAACCTTGGATT		CCGG	
					TATTGCTGTGGGTGATGCCCGGGTTTTGTTT		GCAA	
					ACATCGGGATCCCGCTGCGGCTCGGTGTTC		CGCA	
					CATGCGACTGGCAGCCCCTTTGTTTACTCTS		CTGG	
					GACTCTATTCATTGTTGTATTTCTTCCACAC		TTGCT	
					TGGCCAGTGATCACACTGCTGTGTTTCCCGG		TGCT	
					GAAGATATCCTCTGCGGTTTTCACGCTCTGG		GCGC	
					GTGTCTCCCCGGGCAACGCACTGGTTGCTTG		CATC	
					CTGCGCCATCACCCTTTTCGTTTATATTCAT		ACCC	
					TTTCAGTCTGCCGTTATCTTGGCCAGCGCTC		TTTTC	
					ATTCTTTTGTGATGGCCGTGGACTGACCCTC		GTTT	
					TGCGGTTTTCTGCGGTCASCACTCTCGGTGA		ATAT	
					ATGCTGTGCCACCATTTTTCATTTGTTTACTT		TCATT	

					TTTCAGCTACAATTATCCTGGCCAGCTTTCA CTCTTTTGTGATGGCCGTGGACGGACCCTCT GCTGGTTTTCACGCTCCGGGTTTGTCTGCGG TCAGCACCCTTGGTGAGTGCTGGTCGTTTGT TTGCTCATTTTGCTTAGTTCACCATTATCTTT MTTCTCTTTTGTWTGGTTTTCCTAGCGGTTG TCTGGGAGTTGAGCTGCAGTTGTCTGGTCTT GGTTTTACCCCCATTTGTTTTCTTTTAACGC GGGGCGTATTGCCTTGACCGGCCGTCTCAG CTTTTCTCCTAGAGCAACCTTCCGTTCATCC AACTTTTAGTTTTCTCAGTTCTTGGCCATTC CGGTTGGTTAATTTTTATTTATACTTAATTTT ATGTTTACATTTTCTGGTTGGAGACCATTTT AGCTTGTTTTAATAGCTTTTCTTCTTTAATTA ATACCCTCTGCCATTGAGGGTTTTTATTACT ATTAATTTTGTTTACTCTTTGTAACTTGTTTG ATTGAATATTTTAATAAACCAC (SEQ ID NO: 1067)		TTCA GTCT GCCG TTATC TTGG CCAG CGCT CATT CTTTT GTGA TGGC CGTG GACT GACC CTCT GCGG TTTTC TGCG GTCA SCACT CTCG GTGA ATGC	

							TGTG CCAC CATTT TTCAT TTGTT TACTT TTTCA GCTA CAAT TATC CTGG CCAG CTTTC ACTC TTTTG TGAT GGCC GTGG ACGG ACCC TCTG CTGG TTTTC	

							ACGC TCCG GGTT TGTCT GCGG TCAG CACC CTTG GTGA GTGC TGGT CGTTT GTTT GCTC ATTTT GCTT AGTT CACC ATTA TCTTT MTTC TCTTT TGTW	

							TGGT TTTCC TAGC GGTT GTCT GGGA GTTG AGCT GCAG TTGTC TGGT CTTG GTTTT ACCC CCAT TTGTT TTCTT TTAA CGCG GGGC GTAT TGCC TTGA	

							CCGG CCGT CTCA GCTTT TCTCC TAGA GCAA CCTTC CGTT CATC CAAC TTTTA GTTTT CTCA GTTCT TGGC CATT CCGG TTGG TTAA TTTTT ATTT ATAC	

							TTAA TTTTA TGTTT ACAT TTTCT GGTT GGAG ACCA TTTTA GCTT GTTTT AATA GCTTT TCTTC TTTA ATTA ATAC CCTCT GCCA TTGA GGGT TTTTA TTACT	

							ATTA ATTTT GTTT ACTC TTTGT AACT TGTTT GATT GAAT ATTTT AATA AACC AC (SEQ ID NO: 1313)	
NeSL	Utopia-1_PS	.		Phytophthora sojae	AGACGAGCAACGCGCTGGGGCCCAAGACCT GGCACGAACGACACTGCCCAGGCTGTGAAC GACGAGCACGCTGCTAACCCGGCAGCGCAC CGGCCTCTGGGCTCCGCTGCACCGGTTACC GGTGCAACACGGTGGGGTCCACGTCACGGC GCGATCGGGGCAGCGGCTGTAGCTCGCCTG CTCACAGGCCTACCCACGGCACCAGCACCA	AGACGA GCAACG CGCTGG GGCCCA AGACCT GGCACG AACGAC	TGAA GCTG CACA AGCG CGAG GCTC GACA	MDVDGGPAMPE PWVLRFDGACR RNPGPGGAGAA LFKPCGTVAWT CSHYMPNSSET NNTAEYTALLL GVQSAVHHGAS

					GCTACGCGCCGGCCTGCTTCGGCTCGGCGC	ACTGCC	GCTG	HLEIEGDSHLVV
					TGCCCAGACCCGCCCGCTCCCCCCGCGGCC	CAGGCT	GACC	AQVKGTFACRN
					ACGACGACAGCCCCCGATGCGATGGACGTC	GTGAAC	TTCA	PRLRQLRNRVR
					GACGGAGGGCCGGCTATGCCCGAGCCGTGG	GACGAG	GCTG	HALRAVTSLTL
					GTGCTCCGCTTCGACGGCGCCTGCCGCCGG	CACGCT	TCGA	KHIDRKANAHA
					AACCCGGGACCTGGTGGAGCTGGTGCCGCC	GCTAAC	GCCA	DRLANRALDLK
					TTGTTCAAGCCGTGCGGGACCGTGGCGTGG	CCGGCA	CTGC	RSLAECGEHQG
					ACGTGCTCGCACTACATGCCAAACAGCAGC	GCGCAC	ATCC	AMESCLHMNPA
					GAGACCAACAACACCGCTGAATACACGGCG	CGGCCT	GCGC	AQRQREQPAPP
					CTGCTGCTTGGCGTGCAGAGTGCCGTCCAC	CTGGGC	CAGC	ARPACAPTRAES
					CACGGCGCCTCGCACCTGGAGATCGAAGGT	TCCGCT	CACC	ASDHDEDIDAEI
					GACAGCCACCTCGTCGTAGCGCAGGTCAAG	GCACCG	GCAC	AARDGGEAFPT
					GGCACCTTCGCCTGCAGGAACCCGCGCCTC	GTTACC	CTGG	LPIGPGTAPARQ
					CGCCAGCTGCGCAACCGCGTCCGCCACGCC	GGTGCA	GGCT	PRLRLRQLTEDE
					CTCCGCGCCGTGACGTCCTTAACGCTCAAG	ACACGG	GGTA	QEAAASALQAM
					CACATCGACCGGAAGGCGAACGCGCACGCG	TGGGGT	CTGC	AEELACKIEDAD
					GACAGACTGGCAAACCGGGCGCTGGACCTG	CCACGT	CGGC	SWTSGDGYISAI
					AAGCGTTCGCTGGCTGAGTGCGGCGAGCAC	CACGGC	GCCA	PSRIRQLLQPFT
					CAAGGGGCGATGGAGAGCTGCCTCCACATG	GCGATC	CCGC	AAQPHPRPPLQQ
					AATCCTGCTGCACAACGCCAGCGCGAGCAA	GGGGCA	GAAG	QRQRPPRVTRT
					CCCGCACCTCCTGCCCGCCCCGCGTGCGCA	GCGGCT	GACA	QREHRLDEALD
					CCCACGCGCGCCGAGTCTGCCAGCGATCAT	GTAGCT	ACGG	EMAAVQQERPT
					GACGAGGACATCGACGCAGAGATTGCCGCG	CGCCTG	AGCG	SRSAVRRARRR

					CGCGACGGCGGCGAGGCGTTCCCCACGCTG	CTCACA	GCAA	VGRIRASMRQQ
					CCTATCGGGCCCGGGACGGCTCCAGCGCGG	GGCCTA	CGCC	QLRHDFARNES
					CAGCCGCGCCTTCGACTGCGGCAGCTGACC	CCCACG	TCGC	KCVEDILRAASA
					GAAGACGAGCAGGAAGCGGCGGCGTCGGC	GCACCA	GAGC	ETAAEEHPETCP
					CCTGCAGGCCATGGCGGAGGAGCTAGCGTG	GCACCA	GCCG	IDSGTLHEHFTA
					CAAGATCGAAGACGCAGACAGCTGGACCTC	GCTACG	CGTG	VNSPRINFLPDE
					GGGCGATGGATACATCAGCGCCATCCCCAG	CGCCGG	GGTC	ACGALFREAMA
					TCGCATCCGGCAGTTGCTGCAGCCCTTCACA	CCTGCT	TGGG	DVGTPQERRSA
					GCAGCCCAGCCCCACCCTCGGCCTCCACTG	TCGGCT	GGGG	LTDELTMDEVE
					CAACAGCAGCGACAGCGACCGCCCCGGGTG	CGGCGC	GCCC	DQLMQAATNSS
					ACGCGCACTCAGCGCGAGCATCGACTCGAC	TGCCCA	GCGG	PGHDGVGYDIY
					GAGGCCCTCGACGAGATGGCGGCGGTGCAG	GACCCG	CGCA	KKFAAQLVPLL
					CAGGAGCGACCCACCAGCAGATCGGCCGTC	CCCGCT	CATC	HAAFQSCWRHH
					CGCCGTGCGCGGCGCCGCGTCGGCCGGATC	CCCCCC	GCAG	RVPALWKVGFV
					CGCGCCTCGATGCGCCAACAGCAGCTCCGC	GCGGCC	GCTC	RLIHKKGDPNDP
					CACGACTTCGCTCGCAACGAGAGCAAGTGC	ACGACG	GGGC	ANWRPICLQTAI
					GTCGAGGACATTCTGCGTGCTGCGTCGGCG	ACAGCC	ACGG	YKLYSGLLARR
					GAGACCGCAGCGGAGGAGCACCCGGAGAC	CCCGAT	CGGT	LSAYLEANGLLL
					GTGCCCCATCGACTCGGGCACGCTCCACGA	GCG	AAGC	MAQKGFRAYN
					GCACTTCACGGCGGTGAACTCGCCCCGCAT	(SEQ ID	TGGT	GCHEHNFVATT
					CAACTTCCTACCAGATGAAGCTTGCGGGGC	NO: 1191)	CATTT	LLDQTRRMRRR
					GTTGTTCCGCGAGGCGATGGCCGACGTCGG		GACC	LYQVWYDLRN
					GACACCGCAGGAGCGGCGGAGCGCGCTCAC		AACA	AFGSVHQDML

					GGACGAGCTGACGATGGACGAGGTGGAGG		GGGC	WYVLRLLGVER
					ACCAGCTCATGCAAGCGGCCACGAACTCCA		ACTA	AFVERCDDIYED
					GCCCCGGCCACGACGGCGTGGGCTACGACA		CCCA	SYFVVGNAADG
					TCTACAAGAAGTTCGCGGCGCAGCTGGTCC		GGTA	ATEPVRQEVGV
					CCCTCCTCCACGCAGCCTTCCAGTCCTGTTG		GGGA	YQGCPLSPLLFI
					GCGCCACCACCGGGTGCCTGCCCTGTGGAA		ACCG	AALVPLLRALE
					GGTGGGGTTCGTCCGCCTCATCCACAAGAA		CCCTT	KLDGVGVALAD
					GGGCGACCCCAACGACCCAGCGAACTGGCG		CAAA	GVRPCTTAYAD
					GCCGATCTGCCTGCAGACCGCGATCTACAA		AACC	DLKVFSDSAAGI
					GCTCTACAGCGGTCTGCTGGCGCGTCGGCT		CAGG	TRCHAVVEKFL
					GTCGGCGTACCTGGAGGCGAACGGGCTGCT		AAGA	EWTVLQANPGK
					GCTCATGGCGCAGAAGGGGTTCAGGGCGTA		CACA	CAFLAVTRNAR
					CAATGGTTGCCACGAGCACAACTTCGTGGC		AACA	GNPAHDKDMKL
					AACCACGCTCCTGGACCAAACCCGACGCAT		CCCT	SLHDEEVSSIKL
					GCGCCGCAGGCTGTACCAGGTCTGGTACGA		CCCTT	HDSYRYLGVGD
					CCTGCGGAACGCGTTCGGGTCCGTTCACCA		TAGT	GFDHVRHRLQL
					GGACATGCTGTGGTATGTGCTCCGCCTGCTC		GACA	EPKLQQIKREAV
					GGCGTCGAGCGTGCGTTCGTCGAGCGCTGC		TGCA	ALMQSGLAPWQ
					GACGACATCTACGAGGACTCCTACTTCGTC		TATTT	VVKALKTYVYP
					GTCGGCAACGCGGCCGACGGAGCGACGGA		TAGC	KVEYALRHLRP
					ACCAGTGCGGCAGGAGGTGGGCGTGTACCA		CTAC	LQSQLQGFDRV
					AGGCTGTCCGCTCAGCCCCCTCCTGTTCATC		ATTTC	VAKGLRHLLRL
					GCCGCGCTGGTGCCGTTGCTCCGAGCATTG		AGTT	PRSATNEVLYAP

					GAGAAGCTCGACGGAGTGGGCGTCGCGTTG		ACGG	TSSGGLGLQPLV
					GCGGACGGGGTGCGGCCATGCACCACCGCC		AGAG	EMHRALQIAHA
					TACGCCGACGACCTCAAGGTGTTCAGCGAC		GTTA	WQMLHSKDPAI
					AGCGCGGCCGGCATCACGCGCTGCCACGCT		CTAA	REVARAQVWQ
					GTCGTTGAGAAGTTCCTGGAGTGGACGGTG		CTGG	VARKRHRLREE
					CTGCAGGCCAATCCAGGCAAGTGCGCCTTC		TAAA	HWRERDDELVR
					CTGGCCGTGACACGCAACGCGCGCGGCAAC		CACG	CFLNSELAASPH
					CCCGCCCACGACAAGGACATGAAGCTGTCG		AACA	AEALRRHGDIGS
					CTCCACGACGAGGAGGTCTCAAGCATCAAG		CACA	LWSDVQRWLRI
					CTGCACGACAGCTACCGCTACCTCGGCGTG		T (SEQ	YHLSLVVQDDR
					GGCGACGGCTTCGATCACGTGCGCCACCGT		ID NO:	NGLDPLGLRVP
					CTCCAGCTAGAGCCCAAGCTACAGCAGATC		1314)	HHAKWLDHKS
					AAGCGGGAGGCGGTGGCGCTGATGCAGTCG			VLRHVKLHLKIR
					GGGTTGGCGCCGTGGCAGGTGGTGAAGGCG			HQTRWKGLAD
					CTCAAGACCTACGTCTACCCCAAGGTCGAG			QGKTVRAHGGV
					TACGCGCTCCGCCACCTCCGACCGCTCCAGT			GAKFMSTWAGL
					CCCAACTGCAGGGCTTCGACCGTGTGGTAG			SDDDYRFGVKA
					CCAAGGGCCTGCGGCACCTGCTACGCCTGC			RLNQIDTNAVL
					CCCGCTCGGCAACCAACGAGGTGCTCTACG			KRKRLRSHKTC
					CACCAACGTCGAGCGGGGGCTTGGGGCTCC			RDPTCSSAETLA
					AGCCGCTGGTGGAAATGCACCGCGCGCTGC			HVLNHCESNMD
					AGATCGCCCATGCGTGGCAGATGCTGCACT			AIRQRHDDALE
					CGAAGGACCCAGCGATCCGGGAAGTGGCGC			QIGSKIRNALKR

					GCGCACAGGTGTGGCAGGTGGCCCGCAAGC GTCACCGCCTCCGCGAAGAACACTGGCGGG AGCGTGACGACGAGCTCGTGCGGTGCTTCC TCAACTCCGAGCTCGCCGCCTCGCCCCACG CGGAGGCGCTGCGCCGCCACGGCGACATCG GGTCACTGTGGTCCGACGTACAGCGGTGGC TGCGCATCTACCACCTCAGCCTGGTGGTGC AGGACGACCGCAACGGCCTGGACCCGCTCG GCCTCCGCGTGCCGCACCACGCCAAGTGGC TCGACCACAAGAGCGTGCTGCGGCACGTCA AGTTGCACCTCAAGATCCGCCACCAGACCC GATGGAAGGGTCTGGCGGACCAGGGCAAG ACGGTGCGGGCCCACGGCGGTGTCGGTGCC AAGTTCATGTCGACATGGGCGGGCCTGTCG GACGACGACTACCGCTTCGGCGTGAAGGCC CGCCTCAACCAGATCGACACGAACGCCGTG CTCAAGCGCAAGCGGCTGCGGTCGCACAAG ACCTGCAGGGACCCGACCTGCTCCTCGGCC GAGACGCTGGCCCATGTCCTCAACCATTGC GAGTCCAACATGGACGCAATCCGCCAGCGT CATGACGACGCCTTGGAGCAGATCGGGTCC AAGATCCGCAACGCGCTCAAGCGCGGCAAG TCGACCGCGGAGCTGCGCCTGAACCAGACG			GKSTAELRLNQ TVPEYTGAALRP DIVLRIVAAKK MVIADLAVTFEE HAAGARHSSLQ LSHDHKTLKYQ PIVAELQLKGW QVQTAAIVYGS LGSVQPSNSTPT RKS (SEQ ID NO: 1436)

					GTACCCGAGTACACGGGCGCAGCTCTGCGG CCAGACATCGTACTGCGGATCGTGGCCGCC AAGAAGATGGTGATCGCCGACTTGGCGGTC ACGTTCGAAGAGCACGCGGCCGGCGCCCGC CACTCCTCGCTGCAGCTCAGCCACGACCAC AAGACACTCAAGTACCAACCGATCGTGGCG GAGCTGCAGCTCAAGGGGTGGCAGGTGCAG ACCGCGGCGATCGTGTACGGGTCGCTGGGC TCGGTGCAGCCCAGCAACTCAACACCTACA CGGAAAAGTTGAAGCTGCACAAGCGCGAG GCTCGACAGCTGGACCTTCAGCTGTCGAGC CACTGCATCCGCGCCAGCCACCGCACCTGG GGCTGGTACTGCCGGCGCCACCGCGAAGGA CAACGGAGCGGCAACGCCTCGCGAGCGCCG CGTGGGTCTGGGGGGGGCCCGCGGCGCACA TCGCAGGCTCGGGCACGGCGGTAAGCTGGT CATTTGACCAACAGGGCACTACCCAGGTAG GGAACCGCCCTTCAAAAACCCAGGAAGACA CAAACACCCTCCCTTTAGTGACATGCATATT TTAGCCTACATTTCAGTTACGGAGAGGTTAC TAACTGGTAAACACGAACACACAT (SEQ ID NO: 1068)			

NeSL	Utopia-1_PU	ADOS01001321		Pythium ultimum	GCGGTGTACGCGCACAACGCCGCGCTCTTC	GCGGTG	TGAT	MGTQSARERGA
					GAGTGCACGTTGTGCGCGCACACCGCGCGG	TACGCG	GCGG	PSAPHSHTLGPR
					GATCTCGCCGCGCTCCAGCAGCATCGGCGC	CACAAC	GTCA	TPPRPPACSKHG
					TCCGCGCACCGCAGCGTCCGCTTTGTGGATC	GCCGCG	TATT	ELESAAGGRDG
					ACTTCCACAGCGGATGCGCGTGCGGCGTGA	CTCTTC	GACC	QCSDGAERERD
					GCTTCCACTCGCGTGCGGCGGCAACCAAGC	GAGTGC	GAAA	AERDIRANERDC
					ACGCGCGCGAATGTCCAGAGAGCGCGTTCT	ACGTTG	GGGC	NGDGDGDDADS
					CGGTCGCCGCCGCCGCGCGCACTGCAGCGG	TGCGCG	ACCA	DSDDRNDARRR
					CCAACACCGCAGGTATGTCTCTCGGCGCCG	CACACC	TCCA	SRRPRATATTTT
					ACGAACGCGACCACCTCGCGTCCGTCGGCG	GCGCGG	CGTA	SAPTTTTTTTS
					CCTTGCATGATGTTGCATCCCCCGCTTTTGG	GATCTC	GGAC	ATTSATTPATDS
					CAACATTTTGCCGGTTGCGTTTGCGACCGCG	GCCGCG	ACCG	SPWVLRFDGAC
					GCAGACGCATCAAGCGCCACCGTGATCGCA	CTCCAG	CCCT	RRNPGPGGAGA
					GACGCAGCCATGCAGCACAGTGCTGTGCCC	CAGCAT	CAAA	ALFEPGGAVVW
					TCTGCTGCCGCCCAATCCCCTCGGCGTGCGC	CGGCGC	AACC	TVSHYLPGSETN
					ACGTCCCCCCCGTGCCGCGCGCCACCACGA	TCCGCG	CAGT	NTAEYSAMLLG
					CACCATCCGCGCTGCGGATTGGTGGCAAAC	CACCGC	TCAG	VRSAIHHGATRL
					GCCGCCGCCTGAACGACGACGGCGACAACG	AGCGTC	TTTAT	RVEGDSHLALS
					AAAACAGCGACGGCCGCGACGCCGACATCG	CGCTTT	TGAC	QVRGTFACTNR
					AGATGCGCGCCGACGACACCGACGCACCAG	GTGGAT	ACCC	RLRKLRNRVQA
					CGCCGACCAACCCCGCGACCAGTGCGGCTG	CACTTC	TCCA	ALRELGDYRLV
					CAACGCCCGCGCGCACAGCACCAGCACCAA	CACAGC	CTTA	HIDRQANAHAD
					CGGATGCCGCGACCGCGCGCCGCCGCCACA	GGATGC	GTGA	RLANRALDLRK

					CGCGATGGGGACCCAGAGTGCTCGTGAACG	GCGTGC	CATG	TKVDCGPHATT
					TGGCGCGCCTTCCGCGCCGCACTCACACAC	GGCGTG	CATA	TDACVQPAEILA
					GCTGGGGCCCCGCACTCCGCCGCGACCACC	AGCTTC	TTCA	PTARLSSSSSSSS
					AGCATGCAGCAAACATGGAGAACTCGAATC	CACTCG	AACC	SSSSDEPMPGLE
					AGCTGCTGGTGGCCGCGACGGACAATGCAG	CGTGCG	GATA	EPAADDETDAD
					CGATGGCGCGGAACGCGAGCGCGACGCCG	GCGGCA	CATA	AEADIAMRDGG
					AGCGCGACATCCGCGCCAACGAGCGCGACT	ACCAAG	TTCGT	EIFPTLQIGPGSA
					GCAACGGCGACGGCGACGGCGACGACGCC	CACGCG	TAGG	PAQQPRLRLRQL
					GACAGCGACAGCGACGACCGCAACGACGC	CGCGAA	AGAG	SDDESEAAARTL
					GCGACGACGCAGCAGACGACCGCGAGCGA	TGTCCA	GTTA	EHFANDMASKI
					CGGCGACAACAACGACGAGCGCGCCGACC	GAGAGC	CTAA	ADADDWRSGEG
					ACGACCACGACCACGACAACGAGCGCAAC	GCGTTC	CTGG	YISAIPVRLRELL
					GACGAGCGCGACGACGCCGGCGACAGACA	TCGGTC	TAAT	APYAVPIRSPPR
					GCTCCCCCTGGGTCTTGCGCTTTGACGGCGC	GCCGCC	ATAT	NASSRPPRPQSR
					ATGTCGCCGCAACCCAGGCCCCGGCGGCGC	GCCGCG	CACC	PPRPPRVTRHQR
					CGGTGCTGCGCTCTTCGAGCCAGGTGGCGC	CGCACT	ATTTC	EHRLDEALDDL
					CGTGGTGTGGACGGTGTCTCACTACCTCCCG	GCAGCG	(SEQ	AAAQRSTSTDQ
					GGCAGTGAGACCAACAACACCGCCGAGTAC	GCCAAC	ID NO:	RSIRNARRRVGR
					TCGGCGATGCTGCTCGGGGTGCGCAGCGCC	ACCGCA	1315)	IRTAQAQSDLRS
					ATCCACCACGGCGCGACCCGACTCCGCGTC	GGTATG		QFATNERACVE
					GAGGGCGACAGCCACCTCGCGCTCTCGCAG	TCTCTC		SILRAAKPDGTE
					GTGCGCGGCACCTTCGCGTGCACGAACCGG	GGCGCC		PQASAGTCPIDR
					CGTCTGCGCAAGCTGCGCAACCGAGTGCAA	GACGAA		ATLHAHFAGVN

					GCGGCACTGCGCGAGCTCGGCGACTACCGG	CGCGAC		TPRERFDFDDAL
					CTCGTGCACATCGACCGGCAAGCGAACGCG	CACCTC		GADFRAALDVL
					CATGCCGACCGCCTCGCGAACCGCGCCCTC	GCGTCC		PPPDQAADAFA
					GACCTCCGCAAGACCAAGGTCGACTGCGGG	GTCGGC		DELSLGEVEDQL
					CCCCACGCAACGACCACCGACGCGTGCGTG	GCCTTG		DRVVASSSPGLD
					CAGCCGGCGGAAATCCTGGCGCCAACAGCC	CATGAT		GVGYDVFKRFR
					CGCCTGTCCAGCTCCAGCTCCAGCTCCAGCT	GTTGCA		LQLLPLLHAAY
					CCAGCTCCAGCGACGAGCCGATGCCGGGAC	TCCCCC		QCCWRHRRVPA
					TCGAGGAGCCAGCTGCGGACGACGAGACCG	GCTTTT		TWKVGLVRLLH
					ACGCTGACGCCGAAGCCGACATCGCAATGC	GGCAAC		KKGDRAEPNNW
					GCGATGGAGGCGAGATTTTTCCGACGCTCC	ATTTTG		RPICLQQAIYKV
					AGATCGGCCCCGGCTCGGCGCCAGCACAAC	CCGGTT		YSGLLARRLSR
					AACCGCGGCTCCGCCTCCGCCAGCTCTCGG	GCGTTT		WLEANERFTTA
					ACGACGAGTCCGAGGCAGCCGCGCGCACGC	GCGACC		QKGFREFNGCH
					TCGAGCACTTCGCCAACGACATGGCGAGCA	GCGGCA		EHNFVASSLLDQ
					AGATTGCGGACGCTGACGACTGGCGTTCCG	GACGCA		TRRLHRKLYAV
					GGGAGGGCTACATCAGTGCCATCCCTGTGC	TCAAGC		WYDLRNAFGSM
					GCCTCCGTGAACTCCTCGCACCATATGCGGT	GCCACC		PQPLMWRVLAR
					GCCAATCAGGAGCCCTCCACGGAACGCCAG	GTGATC		LGVDTAFLQRC
					CTCCCGTCCCCCGCGCCCGCAATCTCGTCCG	GCAGAC		EDIYADSFFVVG
					CCTCGCCCGCCGCGAGTGACTCGACACCAG	GCAGCC		NAADGATDPVR
					CGCGAACACCGACTGGATGAGGCGCTCGAC	ATGCAG		QEVGVYQGCPL
					GACCTGGCCGCTGCGCAGCGCTCGACATCG	CACAGT		SPLLFISALIPLL

					ACCGACCAGCGCTCCATCCGCAACGCCCGG	GCTGTG		RALQRLPGVGV
					CGTCGAGTCGGCCGCATCCGCACGGCACAA	CCCTCT		PLADGVRPCTT
					GCCCAATCCGACCTCCGCAGCCAGTTCGCC	GCTGCC		AYADDLKVFSD
					ACGAACGAGCGCGCCTGCGTCGAGAGCATC	GCCCAA		SAAGIQQCHGT
					CTCCGCGCCGCCAAACCCGACGGCACCGAG	TCCCCT		VARFLRWTGLR
					CCTCAAGCAAGCGCGGGCACGTGCCCCATC	CGGCGT		ANASKCALLPV
					GACCGTGCGACGCTCCACGCGCATTTCGCG	GCGCAC		TTTARGNPAIDD
					GGCGTCAACACGCCGCGCGAGCGCTTCGAC	GTCCCC		TLQLELHGDAIA
					TTTGACGACGCGCTGGGAGCCGACTTTCGC	CCCGTG		RLTLQSSYAYLG
					GCCGCGCTCGACGTTCTGCCACCGCCGGAC	CCGCGC		VGDGFDHVQHR
					CAAGCCGCCGATGCGTTCGCCGACGAGCTC	GCCACC		VQLAPKLAELK
					TCGCTGGGCGAGGTCGAGGACCAGCTCGAC	ACGACA		RDAVALLRSGL
					CGCGTCGTCGCCTCCTCCAGTCCCGGCCTGG	CCATCC		APWQVLKAIKV
					ACGGAGTGGGCTACGACGTGTTCAAGCGCT	GCGCTG		YLYPRIEYALRH
					TTCGGCTGCAGCTCCTGCCGCTCCTCCACGC	CGGATT		LRPLQSQLEGFD
					TGCATACCAGTGCTGCTGGCGCCACCGCCG	GGTGGC		RAVAKGFRHLL
					CGTGCCTGCAACGTGGAAGGTGGGATTGGT	AAACGC		RLPANATNELL
					GCGCCTGCTGCACAAGAAAGGCGATCGCGC	CGCCGC		YAPVSSGGLGLL
					TGAGCCAAACAACTGGCGGCCGATCTGCTT	CTGAAC		PLVELHKALQIA
					GCAGCAGGCCATCTACAAGGTCTACTCTGG	GACGAC		HGWQMLHSKD
					GCTGCTGGCGCGCCGACTGTCGCGGTGGCT	GGCGAC		AAVQAIARAQV
					CGAGGCGAACGAACGCTTCACCACCGCCCA	AACGAA		RQVVQKRYTLD
					GAAAGGTTTCCGGGAGTTCAACGGGTGCCA	AACAGC		ADHWQGRDDE

					CGAGCACAATTTCGTGGCATCCTCGCTCCTC	GACGGC		LVQLFLNSELAA
					GACCAGACGCGTCGGCTCCATCGGAAGCTG	CGCGAC		SPHATIKRRNGD
					TACGCTGTCTGGTACGACTTGCGGAACGCC	GCCGAC		IGSLWSDVQRH
					TTCGGTTCAATGCCGCAGCCCCTGATGTGGC	ATCGAG		LKTLQLRLETRE
					GCGTCCTCGCTCGCCTCGGAGTCGACACTG	ATGCGC		PTADAPDSPNGL
					CGTTCCTCCAGCGGTGCGAGGACATCTACG	GCCGAC		LHLRVPHHRKW
					CGGACTCGTTTTTCGTCGTCGGCAACGCAGC	GACACC		LSHKTVLRHMK
					CGACGGTGCCACGGACCCCGTCCGACAGGA	GACGCA		LHIRLCHKHKW
					GGTCGGCGTCTACCAAGGATGCCCGCTGAG	CCAGCG		QSMSDQGRTVR
					CCCGCTGCTGTTCATCTCAGCATTGATTCCT	CCGACC		AHGQAGSHFVS
					CTCCTGCGGGCTCTGCAGCGGCTACCAGGC	AACCCC		RGVGLWDADY
					GTCGGCGTTCCCCTGGCGGACGGAGTCCGT	GCGACC		RFALQARLNQL
					CCGTGTACGACCGCGTACGCGGACGACCTC	AGTGCG		DTNSTLKRRRQ
					AAGGTGTTCAGCGACAGTGCCGCCGGCATC	GCTGCA		RTNATCRAPNC
					CAGCAATGCCACGGCACTGTCGCTCGATTC	ACGCCC		SRTETLAHVLN
					CTCCGATGGACGGGGCTGCGCGCGAACGCG	GCGCGC		HCETNMDVIRQ
					AGCAAGTGTGCGCTGCTGCCAGTGACAACC	ACAGCA		RHDGALEQIGA
					ACTGCGCGCGGCAACCCCGCCATCGACGAC	CCAGCA		AINAAIKGRRTD
					ACGCTGCAGCTCGAGCTGCACGGCGACGCC	CCAACG		TEVRLNQTVPEF
					ATCGCCCGCCTCACCCTCCAATCGAGCTATG	GATGCC		NGPAWRPDIQV
					CCTACCTTGGGGTGGGTGATGGCTTCGACC	GCGACC		RDARSKTMVIA
					ACGTGCAGCACCGCGTGCAGCTCGCGCCCA	GCGCGC		DLAITFEDQPND
					AGCTCGCAGAGCTGAAGCGCGACGCTGTGG	CGCCGC		QSASSSLQHSRE

					CACTGCTGCGGTCGGGGTTGGCGCCTTGGC AGGTACTGAAGGCCATCAAGGTCTATCTGT ACCCGCGCATCGAGTACGCGCTGCGCCACC TCCGACCGCTGCAGTCGCAGCTCGAGGGGT TTGATCGCGCGGTAGCGAAGGGGTTCCGAC ACCTCCTCCGTCTACCCGCGAACGCCACCA ACGAGCTGCTCTACGCGCCTGTGTCGAGCG GCGGCCTGGGGCTTCTGCCACTCGTGGAGC TCCACAAAGCTCTGCAGATTGCCCACGGGT GGCAGATGCTCCACTCGAAAGACGCCGCCG TGCAGGCGATCGCGCGCGCACAGGTCCGCC AAGTGGTCCAAAAGCGCTACACGCTGGACG CGGACCACTGGCAGGGACGAGACGACGAG CTCGTGCAGCTCTTTCTCAACAGTGAGCTCG CCGCCTCGCCCCACGCGACCATCAAGCGGC GCAACGGAGACATTGGTTCGCTCTGGAGCG ATGTCCAGCGCCACCTGAAGACGCTGCAGC TCCGGCTCGAGACGCGCGAGCCGACTGCCG ACGCGCCTGACTCCCCCAACGGCCTCCTCC ACCTCCGTGTGCCGCACCACCGCAAATGGC TCAGCCACAAAACGGTGTTGCGGCACATGA AGCTCCACATCCGCTTGTGCCACAAGCACA AGTGGCAGAGCATGAGCGACCAGGGGCGC	CACACG CG (SEQ ID NO: 1192)		HKIAKYQPIAAA LERQGWRVHTS AIVYGSLRSVHP SNFTVYTELLGL LKRDARRLNTT LSCHCIRSSRRV WNWHCGQHRA RQHQRCQEGRA HGSGGNQRAEG GTATT (SEQ ID NO: 1437)

					ACGGTGCGAGCGCACGGACAGGCTGGCAGC CACTTTGTGTCGCGCGGGGTGGGCCTATGG GACGCGGACTACCGCTTCGCGCTCCAGGCC CGCCTCAACCAACTCGACACCAACTCGACA CTCAAGCGTCGCCGCCAGCGCACCAATGCC ACCTGCCGCGCGCCGAACTGCTCGCGCACC GAGACACTGGCGCACGTCCTCAATCATTGC GAAACCAACATGGACGTGATCCGCCAGCGT CACGACGGCGCGCTCGAGCAGATTGGTGCG GCGATCAACGCTGCCATCAAGGGCCGCCGC ACCGACACCGAGGTCCGACTCAACCAGACA GTTCCAGAGTTCAACGGACCGGCGTGGCGG CCCGACATCCAGGTGCGCGACGCGCGCTCC AAGACCATGGTGATCGCGGACCTCGCGATC ACCTTCGAGGACCAGCCAAACGACCAGTCC GCCAGCTCGTCGCTCCAACACAGCCGCGAA CACAAGATCGCGAAATACCAGCCAATCGCG GCGGCACTGGAGCGGCAAGGGTGGCGGGT ACACACAAGTGCCATTGTGTACGGGTCGCT ACGCTCGGTCCACCCCAGCAACTTCACCGT GTACACGGAGTTGCTGGGACTGCTCAAGCG CGACGCCCGTCGCCTGAACACAACGCTGTC GTGCCACTGCATCCGCAGCAGCCGACGCGT			

					GTGGAACTGGCACTGCGGCCAGCACAGAGC ACGACAGCATCAGCGATGCCAAGAAGGCCG CGCGCACGGGTCCGGGGGGAACCAGCGCGC GGAAGGGGGCACGGCAACAACATGATGCG GGTCATATTGACCGAAAGGGCACCATCCAC GTAGGACACCGCCCTCAAAAACCCAGTTCA GTTTATTGACACCCTCCACTTAGTGACATGC ATATTCAAACCGATACATATTCGTTAGGAG AGGTTACTAACTGGTAATATATCACCATTTC (SEQ ID NO: 1069)			
NeSL	Utopia-1_SM	.		Strigamia maritima	GGAGTGTTCTTTTCGGAGACGCCGCCTACTT TAGAGGAGAGAATCCCCACGGGCATCCTCA TTTGATCTGATCCATCGAGTATCTGCGAATA GTCGGCGCACTCCTTTTGCCATGATCCCGGG GGTCTCATGGTAAAAAGGTTTGTGGCACGG CTTAGTTGACGCCCCTCTTCCACGTCACTCG GCCTGCATCGATCGACTGCTCTCTCCTCTTC CTCCCTCCCTCCTCTTCACGCTCTCTTCACGT GACCCCTTCCCATCCCGCCCCTCGGCTTTTG GCAAAGATCTGTGTGTCCTCCAAAGCACCC ATCTACCATTTGCTCGAGTTGCGATTGGTCG AAGCTGCCACGCCACTCGTCTACTCTGCCTC TCTGACCCCTTCCCCCCTCTCTCCCTCTGCC	GGAGTG TTCTTTT CGGAGA CGCCGC CTACTT TAGAGG AGAGAA TCCCCA CGGGCA TCCTCA TTTGAT CTGATC CATCGA	TGAT GGGA GAGT GAGG AATC TTCTC CACT GTGC AAAA CCAT ACAG TCAG AAGA	MATVRLKYPYP PEGILCGPCAAN TNAPQTRPYSD KSGLAKHLKLY HKATLVVECRH CGHEESDLRKM KKHISTNHPVAA AAAPTVPPRLGP TAPPPPRVILRPR FIPRPRTPSPSSSS SSASSPASSRRS VSLPPASPPVSS ASSPAARSGRNS

					GCCCTCGCCTCTCGGGTCCCTTCCCATCTCC	GTATCT	TGCT	PDSQGTAPVTPI
					CTCCTCCCATCCACGTGCTCCCTCCTCTCGT	GCGAAT	AACT	GTVRNSPAGSPA
					CACGTGATCGTTGCGGCTCGACATGGCCAC	AGTCGG	ACTA	LSYSTASPIASTI
					GGTCAGACTCAAGTACCCCTATCCTCCCGA	CGCACT	GTTT	TTPRHLSPASPA
					GGGTATCTTGTGTGGACCGTGTGCCGCCAA	CCTTTT	GATA	LSAGPGSLGASP
					CACGAACGCCCCTCAGACGCGGCCATACTC	GCCATG	CCCT	PVSPTAATVPPA
					CGATAAGTCGGGGTTGGCCAAGCATCTAAA	ATCCCG	GTGC	PPATVPAVMAA
					GCTCTATCACAAAGCCACTCTAGTGGTGGA	GGGGTC	CCCC	TVPFVAATTVPS
					GTGTCGTCACTGTGGTCACGAGGAGTCTGA	TCATGG	TGCA	VGSSTVPQRPAG
					TCTGAGGAAGATGAAGAAGCACATCTCGAC	TAAAAA	ATGT	PRRPPPFPIDDWI
					GAACCACCCCGTCGCTGCGGCCGCCGCGCC	GGTTTG	CCCG	GRIARVSSLPEL
					TACCGTACCCCCTCGCCTGGGTCCCACCGCC	TGGCAC	CGTG	DAVSRLLEDEV
					CCTCCTCCGCCCCGGGTTATCCTCCGCCCTC	GGCTTA	TCGT	VKRRPPDPNARP
					GCTTCATCCCACGGCCTCGGACTCCCTCTCC	GTTGAC	ACCC	ASLHPTRRPPPP
					TTCCTCCTCCTCGTCATCGGCGTCGTCTCCG	GCCCCT	AAGC	STRPRPCHGTGG
					GCCTCTTCCAGAAGGTCTGTATCCCTTCCCC	CTTCCA	CCGG	TSVSLAALTSSCI
					CTGCGAGTCCCCCCGTCTCCTCCGCTTCCTC	CGTCAC	CTGG	REDHRGSLPLLC
					CCCGGCCGCTCGCTCGGGTCGGAATTCGCC	TCGGCC	CATT	TSGWDSPWPLS
					GGACTCGCAGGGGACGGCTCCGGTCACTCC	TGCATC	GAGA	PSPSSHCPSSNPC
					GATCGGTACCGTACGGAACTCGCCGGCCGG	GATCGA	CACA	PSSSTPPSSLLGP
					TAGCCCTGCCCTCAGCTACTCGACGGCCTCC	CTGCTC	TTAG	PRHSHLTWRGS
					CCGATCGCCTCTACCATCACGACGCCACGC	TCTCCT	GCTC	GSTTPSHRHCSR
					CATCTCTCGCCTGCGTCCCCCGCTCTCTCTG	CTTCCT	TCGC	ARYHHLLWRLP

					CCGGGCCTGGCTCTCTCGGCGCGTCTCCGCC	CCCTCC	TCCC	YPSSPRLPYPPS
					CGTCTCTCCCACGGCCGCTACAGTACCCCCC	CTCCTC	CCGT	ALARYPSVPPAL
					GCTCCTCCAGCTACCGTACCCGCAGTGATG	TTCACG	ATAC	DDPLPSLSTIGSE
					GCGGCTACCGTACCCTTCGTCGCCGCGACT	CTCTCT	TCTC	GSPECHLCRNW
					ACCGTACCCTCCGTCGGCTCTAGCACGGTA	TCACGT	ATAA	TPSQGCSRMKW
					CCCCAGCGTCCCGCCGGCCCTCGACGACCC	GACCCC	TTTCG	SRDAPLTPTPDP
					CCTCCCTTCCCTATCGACGATTGGATCGGAA	TTCCCA	TGTA	PHSIRHAALPLH
					GGATCGCCCGAGTGTCATCTCTGCCGGAAT	TCCCGC	CGCT	LLALVPAMGPA
					TGGACGCCGTCTCAAGGTTGCTCGAGGATG	CCCTCG	AATC	ELQSLWRRSRPR
					AAGTGGTCAAGAGACGCCCCCCTGACCCCA	GCTTTT	CTAC	AFAKITEGASPF
					ACGCCAGACCCGCCTCACTCCATCCGACAC	GGCAAA	CCTA	CALPVGTVHGH
					GCCGCCCTCCCCCTCCATCTACTCGCCCTCG	GATCTG	CCCC	FLQVHQATAHL
					TCCCTGCCATGGGACCGGCGGAACTTCAGT	TGTGTC	TCCCT	PTPVPLPPLRPPR
					CTCTCTGGCGGCGCTCACGTCCTCGTGCATT	CTCCAA	TTGA	SSDPLVTPISPGE
					CGCGAAGATCACCGAGGGAGCCTCCCCCTT	AGCACC	CCAC	VLDRLRRATDT
					TTGTGCACTTCCGGTTGGGACAGTCCATGGC	CATCTA	TCAC	APGPDTIIYSEW
					CACTTTCTCCAAGTCCATCAAGCCACTGCCC	CCATTT	CCAA	RAIDPTGRLLSS
					ATCTTCCAACCCCTGTCCCTCTTCCTCCACT	GCTCGA	CCAT	LFQKVQTFGAPT
					CCGCCCTCCTCGCTCCTCGGACCCCCTCGTC	GTTGCG	GTGT	RWKESTTTLIHK
					ACTCCCATCTCACCTGGAGAGGTTCTGGATC	ATTGGT	ATAG	GGDHTAMSSWR
					GACTACGCCGAGCCACCGACACTGCTCCCG	CGAAGC	CTGT	PIALLSTVAKIY
					GGCCAGATACCATCATCTACTCTGAGTGGA	TGCCAC	GCTG	GSILSHRLTTWA
					GGGCCATCGACCCGACTGGACGACTCCTCT	GCCACT	GTGA	VQNGRLSLSQK

					CCTCTCTCTTCCAGAAGGTCCAGACCTTCGG	CGTCTA	TCCC	GFLPFRGCLDQN
					GGCGCCAACCCGATGGAAGGAGAGTACGA	CTCTGC	GGGG	YLVQSCLQDAR
					CCACACTCATCCACAAGGGTGGTGACCACA	CTCTCT	CGGT	RNKKTLSLAFLD
					CAGCGATGTCCTCTTGGCGTCCAATCGCCCT	GACCCC	TATTC	LKNAFGSIPHLTI
					CCTCTCCACAGTGGCCAAGATCTATGGGAG	TTCCCC	ACTG	RHSLEWLGLAP
					CATCCTGTCTCATCGGCTGACGACCTGGGCC	CCTCTC	GTTA	SSIDILEASFLGS
					GTCCAGAATGGTCGTCTCTCGCTCTCGCAGA	TCCCTC	TCAT	STRVRTETGLTP
					AGGGATTCCTGCCATTTCGAGGATGCCTCG	TGCCGC	ATCA	PISLDTGVVQGA
					ACCAGAACTACCTTGTTCAGTCGTGCCTCCA	CCTCGC	TTCTA	PLSPILFNLAIEP
					AGACGCCCGCCGCAACAAGAAGACACTCTC	CTCTCG	AAAT	LLRTVPSAHSGF
					ACTCGCCTTCCTCGATCTGAAGAATGCCTTT	GGTCCC	GATC	SLHGHVVSVVA
					GGGAGCATCCCCCACCTGACCATTCGTCAC	TTCCCA	TTTG	YADDLAILAPST
					TCCCTCGAGTGGCTCGGACTGGCCCCTTCAT	TCTCCC	ATCT	PALQSQLDAISG
					CGATCGACATTCTTGAAGCGTCATTCCTGGG	TCCTCC	CTCA	MADWAGLSFNP
					ATCGTCGACTCGCGTTCGAACCGAGACTGG	CATCCA	ATTA	AKCATVTLTGK
					ACTCACTCCCCCCATCAGCCTCGACACTGG	CGTGCT	ACAA	DNSRDTLSLQGS
					AGTGGTCCAAGGAGCTCCCCTCTCCCCCATC	CCCTCC	CTAA	PVPSISDGDAYK
					CTCTTCAATCTAGCCATCGAACCCCTCCTCC	TCTCGT	CTTAT	HLGVPTGTTTFP
					GCACGGTCCCCTCTGCCCATTCTGGATTTTC	CACGTG	TTCTG	SGTDAIKKMTT
					TCTCCATGGCCATGTCGTCTCTGTGGTCGCC	ATCGTT	TTTGT	DLQAIDHSDLAP
					TACGCGGATGACCTCGCCATCCTCGCCCCAT	GCGGCT	TTCAT	AQKLDALRTFI
					CTACTCCCGCTCTACAATCCCAATTGGATGC	CGAC	TGTTT	MPQLSFHLSHGS
					TATCTCTGGCATGGCAGATTGGGCGGGCCT		CACC	VPKAPLTQLDK

					CTCATTCAATCCGGCCAAGTGCGCGACCGT CACTCTCACTGGGAAGGATAACTCTCGTGA TACGCTGTCCCTGCAAGGCTCCCCCGTCCCT TCCATCTCTGACGGCGATGCATACAAGCAC CTTGGAGTCCCCACTGGCACTACTACCTTCC CCAGTGGCACCGATGCAATCAAGAAGATGA CGACGGATCTACAGGCCATAGACCACTCCG ACCTCGCACCGGCCCAGAAGCTAGATGCCC TCCGCACTTTCATCATGCCCCAGCTCTCCTT CCATCTCTCTCACGGCTCAGTCCCCAAAGCC CCCCTCACTCAACTGGACAAGAAGATCAAG CGGGCTGCCAAGCACTGGCTGTTCCTTCCTC AACGAGCCTCGAACGAAATACTGTACATGT CCCACTTGCACGGAGGGCAATCCCTACTCC CCCTCTCGGTGCTAGCGGACATCGGACAAG TGACTCACGCCGTCGCTCTTCTTCAATCAAG GGATCCAGCCGTGGCTGACCTCGCTCTCCG GACTTGCCGAGAAGTCGCATCGAAACGGGC AAAGAAGACGGTCAACGGCCCCGAGTTGGC CCAGTACCTCTCCGGGTCCACTGACGGAAT CTACTGCACCCCGACTTCTGACATCCCGAGT CTCTGGACGACTGCCCGAGCTGCAACACGC CGACTCTCCTCCACTCTGCCACTCACGTGGA	(SEQ ID NO: 1193)	TCGT AAGA GGAA GTTC ATTG TGCG ATAA ATCA A (SEQ ID NO: 1316)	KIKRAAKHWLF LPQRASNEILYM SHLHGGQSLLPL SVLADIGQVTH AVALLQSRDPA VADLALRTCRE VASKRAKKTVN GPELAQYLSGST DGIYCTPTSDIPS LWTTARAATRR LSSTLPLTWTSP LPSGVPFLSINGS PLSPFRVQSTLT NAIRHNHLSTLI AKRDQGNSYRT SHDPDPSNYWV KGGDFLRFCDW RFIHRARLNLLP VNGARRWDAN SIKTCRRCGAPN ETLAHVLNVCP VGLPEMKKRHD AIHARIKKALRP

					CATCCCCCCTTCCCTCTGGCGTCCCCTTCCT TTCTATCAATGGGTCCCCCCTCTCCCCATTC CGAGTACAGTCCACCCTCACCAATGCCATT AGACATAACCACCTCTCCACTCTCATTGCCA AACGGGACCAGGGAAATTCCTACCGTACCT CACACGACCCCGACCCCTCTAACTACTGGG TCAAGGGAGGAGACTTCCTCCGATTCTGTG ACTGGCGATTCATACACAGAGCCCGACTGA ACCTCCTCCCAGTGAACGGGGCGCGCCGAT GGGACGCAAACAGCATCAAGACGTGCCGGC GCTGTGGTGCGCCCAATGAGACACTCGCCC ACGTGCTCAACGTCTGCCCTGTCGGACTGCC GGAGATGAAGAAGAGGCACGATGCAATTC ATGCCCGCATCAAGAAAGCCCTCCGACCCT CTCCCCATACGGTAGTACACCACGACCGGA CAGTTCCCGGATGCGGCCCCCTGAGACCGG ACATTCTGAGAATCAGCGAGAGAGACAAAT CTGTTGCCATTGTAGATATCCACGTTCCCTT CGACAATGGCACTGATGCGGTTGAGCGAGC TCACGAGACAAAGCGGGCGAAGTACGAACT GATCCGGCGGCACTACGAACACCAAGGGTA TCGAGTCACGTTTGACTCACTAGTAGTCACA GCGCTTGGTAGACTGTGGCGGGGGAGTGAG			SPHTVVHHDRT VPGCGPLRPDIL RISERDKSVAIV DIHVPFDNGTDA VERAHETKRAK YELIRRHYEHQG YRVTFDSLVVT ALGRLWRGSEA ALQALQISSQYS KLLRKLLVADAI HGSRNVYAHH MTGMVM (SEQ ID NO: 1438)

					GCAGCGCTACAAGCGCTCCAGATCTCGTCA CAGTATAGTAAGCTCCTCCGTAAGTTACTCG TGGCTGATGCCATCCACGGCTCGAGAAACG TGTACGCCCATCATATGACCGGTATGGTTAT GTGATGGGAGAGTGAGGAATCTTCTCCACT GTGCAAAACCATACAGTCAGAAGATGCTAA CTACTAGTTTGATACCCTGTGCCCCCTGCAA TGTCCCGCGTGTCGTACCCAAGCCCGGCTG GCATTGAGACACATTAGGCTCTCGCTCCCCC GTATACTCTCATAATTTCGTGTACGCTAATC CTACCCTACCCCTCCCTTTGACCACTCACCC AACCATGTGTATAGCTGTGCTGGTGATCCC GGGGCGGTTATTCACTGGTTATCATATCATT CTAAAATGATCTTTGATCTCTCAATTAACAA CTAACTTATTTCTGTTTGTTTCATTGTTTCAC CTCGTAAGAGGAAGTTCATTGTGCGATAAA TCAA (SEQ ID NO: 1070)			
NeSL	Utopia-1_SP	AAGJ02140537		Strongylocentrotus purpuratus	AAGGCTCAAACCAGGCTGCCAACCAAGCTG CCAGCTTAGGCACCAACCAAGCTGCCAGTC CAGGCTCCAACCAGGCTGCCCACCAATCTG CCAGCTTAGGCTCCAACCAAGCTGCCCACC TAGCTGCCAGCTTAGGCACCAACCAAGCTG CCAGCCGAGGATCCAACCAGGCTGCCCACC	AAGGCT CAAACC AGGCTG CCAACC AAGCTG CCAGCT	TGAC GAAA TGTTC AATA TATG ACAT	MSHSITEVFDYP LPSRWKCTVCL ENFFNQQTLKR HQARHHQTTSF LYVFRCSACQA EFDSARKASNH

					TAGCTGCCAGCTTAGGCTCCAACCAAGCCG	TAGGCA	TCTA	WQSHKRKPILSQ
					CCAGCCCAGGCTCCAACCAGGTTGCCACCC	CCAACC	ATGT	PAVNEIPSSGLD
					GAAAGTCTGCCAGTTTAGGCACCAACCAAG	AAGCTG	TCAT	PSPPRSRPPVEVI
					CTGCCAGCCGAGGCTCCAACCAGGCTGCCA	CCAGTC	GTAT	GSSFPDDVSMLS
					CTCGAGGCTCCAACCAGGCTTTCACCCGAA	CAGGCT	TTTTT	EPSTPSTSLQMD
					GCATTACCCAGGCTGCCCGTTTAGGCACAA	CCAACC	GTTT	PEVVHPPSRSISF
					ACCAGGCTGCCAGCCGAGGCTCCAACCAAG	AGGCTG	GCTT	SPMHLSPTQPAS
					CAACCAGCCCAGGTTCCAATCAAGCTGCCA	CCCACC	GACA	PIQIGVEVSFNSS
					GCCGAGGCTCCAAGTAAGCTGCCAGCCGAG	AATCTG	AAGC	SSLQMDPEVVQ
					GCTCCAACCAGGCTGCCACTCGAGGCTCCA	CCAGCT	TAGA	PPSPSISFSPMHL
					ACCAGGCTTTCACCCGAAGCATTACCCAGG	TAGGCT	TGAA	SPTQPASPIQIGV
					CTGCCCGTTTAGGCACAAACCAAGCTGCCA	CCAACC	TATA	EVSFNSSGSLQM
					GCCGAGGCTCCAACCAAGCAACCAGCCCAG	AAGCTG	TTCCC	DPEVVQPPSPSIS
					GTTCCAATCAAGCTGCCAGCCGAGGCTCCA	CCCACC	TGTC	FSPMHLSPTQPV
					ACCAAGCTGCCAGCCGAGGCTCCAACCAGG	TAGCTG	CTAC	SPIQIGIEVSFTSS
					CTGCCCACCAAGCTGCCAACTTAGGCTCCA	CCAGCT	TACA	SPLQMDPEVVQ
					ACCAAGCTGCCAGCCCAGGCTCCAACCAGG	TAGGCA	TACA	PPSPSMSYSPMH
					TTGCCACCCGAATCTCTGCCAGTTTAGGCAC	CCAACC	TCTC	LSLTQPDSPIPVD
					CAACCAAGCTGCCAGCCTAGGCTCCAACCA	AAGCTG	GATG	IDVIPAAEVPLP
					GGCTGCCACTCGAGGCTCCAACCAGGCTTT	CCAGCC	TGCC	DIEIPPSPDRHPA
					CACCCGAAGCATCACCCAGGCTGCCCTTTT	GAGGAT	CTTG	AEVPLPDIEIPPS
					AGGCACAAATCAAGCTGCCAGCCGAGGCTC	CCAACC	CTAG	PDRHPVAEVPLP
					CAACCAAGCAACCAGCCCAGGTTCCAATGA	AGGCTG	ACGA	DIEIPPSPDRHPQ

					AGCTACCAGCCGAGGCTCCAACCAAGCTGT	CCCACC	TCCA	SPPRPVMMEQP
					CAGCCGAGGCTCCAACCAGGCTGCCCACCA	TAGCTG	TTGG	VHTPPPADTQQ
					AGCTGCCAGCTTAGGTACCAACCAAGCTGC	CCAGCT	TAAC	ANGPQHWVTVL
					CAGCCCAGGCTCCAACCAGGTTGCCACCCG	TAGGCT	CATG	ANATNWEDFGR
					AAACTCTGCCAGTTTAGGCACCAACCAAGC	CCAACC	ATAT	VCVEFANHAVE
					TGCCAGCCGAGGCTCCAACCAGGCTGCCAC	AAGCCG	CAAT	AARSRQDAPQV
					TCGAGGCTCCAACCAGGCTTTCACCCCAAG	CCAGCC	GGAT	RPAAQRQPRRPT
					CATCAACCAGGCTGCCCGTTTAGGCACAAA	CAGGCT	TAAC	RPRQPTFDVREA
					CCAAGCTGCCAGCCGAGGCTCCAACCAGGC	CCAACC	ACAT	SRLQKLYKRSK
					CTCCAACCAAGCTACCAGCCGAAGCTCTGC	AGGTTG	GATC	KRAVRHILRDD
					CAGATTAGGCACCAACCAAGCTGCCACCGT	CCACCC	TGTA	APSFSGSNEQLL
					TAGAGGCCCAGAGCCCACCAATCCCATAAC	GAAAGT	AACA	DYFKEIYAPPEI
					GTGTGAGATATGTGAAGCTTCCTTCCACACC	CTGCCA	ATAT	DENRAQQLAES
					TCCGCCGGTCTCCGTCGCCACACGGCCAGG	GTTTAG	ATGA	LFTDLEEAKESA
					CTTCATCGCACCACTGCTGAACACACTGAC	GCACCA	TTTAC	AALMSPISQQEI
					GACAGCATGAGCCATTCCATCACCGAGGTA	ACCAAG	ACAA	STRLSRMSNSAP
					TTCGACTATCCGTTGCCCTCCCGTTGGAAGT	CTGCCA	TGTA	GKDRLEYRHIR
					GCACAGTGTGCTTAGAGAACTTTTTCAATCA	GCCGAG	TTTAT	QADGACRVTHI
					GCAGACCCTCAAGCGCCACCAGGCCCGGCA	GCTCCA	TGTTT	MFNRCLQEHRIP
					CCACCAAACGACCTCCTTCCTCTACGTCTTT	ACCAGG	CAAT	SAWKEATTILIH
					CGATGCTCAGCCTGCCAGGCTGAGTTTGATT	CTGCCA	AAAT	KSGTTDDPANF
					CTGCTCGTAAGGCTTCTAATCACTGGCAGA	CTCGAG	CTGTT	RPIALQSCLYKL
					GCCACAAGAGGAAACCAATCCTCTCACAAC	GCTCCA	CTTTT	FMGILSDRMTQ

					CAGCAGTGAATGAAATCCCTTCTTCGGGTCT	ACCAGG	CACT	WACNHNLLSPE
					AGATCCATCCCCTCCAAGGTCTCGTCCACCG	CTTTCA	TTAA	QKSARPCEGCH
					GTCGAGGTGATTGGTTCTTCTTTTCCAGATG	CCCGAA	TACA	EHTFLLSSVIKD
					ATGTGTCTATGCTGTCAGAACCATCTACCCC	GCATTA	TGAA	TKRNQKTANIA
					CTCTACATCACTACAAATGGACCCAGAGGT	CCCAGG	ACAT	WLDLRNAFGSIP
					GGTGCATCCTCCGTCACGATCCATATCCTTT	CTGCCC	GACT	HQAIHAVLTTIG
					TCGCCCATGCATCTTAGCCCCACTCAGCCAG	GTTTAG	GTGC	APVSLVMLLKD
					CCAGCCCGATTCAAATAGGCGTTGAGGTTT	GCACAA	CTTCT	TYTGASTSFLST
					CCTTTAACTCTTCCAGCTCACTTCAGATGGA	ACCAGG	CCAA	SGETDPIQIQSG
					CCCAGAGGTCGTGCAGCCCCCGTCACCGTC	CTGCCA	CTGG	VKQGCPMSAILF
					CATTTCCTTCTCACCCATGCATCTTAGCCCC	GCCGAG	AGAC	NLTIELIIRAVKK
					ACTCAGCCCGCCAGCCCGATTCAAATAGGC	GCTCCA	CTAC	KATDDGLGLVV
					GTTGAGGTTTCCTTTAACTCTTCCGGCTCAC	ACCAAG	ATAA	HGQRLSIMAYA
					TTCAGATGGACCCAGAGGTGGTACAGCCCC	CAACCA	TTTGT	DDLVLMSKTPE
					CGTCACCGTCCATCTCCTTCTCACCCATGCA	GCCCAG	TAAA	GLDAILSVASEQ
					TCTTAGCCCCACCCAACCAGTCAGCCCGATT	GTTCCA	ATGA	AETLRLAFKPTK
					CAAATAGGCATTGAGGTTTCCTTTACCTCTT	ATCAAG	TATA	CASLSLSCRHGT
					CCAGCCCACTTCAGATGGACCCCGAGGTAG	CTGCCA	AATA	SVLPREYTVQG
					TGCAGCCCCCGTCACCCTCCATGTCCTACTC	GCCGAG	TTTCG	HLMPALDEEEQ
					ACCCATGCATCTTAGCCTCACTCAGCCAGA	GCTCCA	AAGA	YRYLGVPFGLPR
					CAGCCCCATCCCAGTAGACATTGATGTCAT	AGTAAG	TGAA	FTNLKDLIGKLK
					ACCAGCGGCTGAGGTCCCCTTGCCTGATATT	CTGCCA	ATTA	GNIETIASSLLAP
					GAGATTCCTCCTTCTCCGGATAGACATCCAG	GCCGAG	TTATT	WQKLDAIKTFV

					CGGCTGAGGTCCCCTTGCCTGATATTGAGAT	GCTCCA	AATA	QPGLSFVLRAA
					TCCTCCTTCTCCGGATAGACATCCAGTGGCT	ACCAGG	A (SEQ	DYLKSDLRSLKS
					GAGGTCCCCTTGCCTGATATTGAGATTCCTC	CTGCCA	ID NO:	AITTNVKKICQL
					CTTCTCCGGATAGACATCCACAATCGCCTCC	CTCGAG	1317)	PLRAANAYIFAA
					ACGACCAGTGATGATGGAGCAGCCAGTCCA	GCTCCA		KESGGLAFIDPN
					TACACCACCGCCGGCCGATACACAGCAGGC	ACCAGG		VDADIQVITQAV
					TAATGGTCCACAGCATTGGGTGACTGTTCTA	CTTTCA		RVLSSDDEVVQ
					GCCAACGCCACCAATTGGGAAGACTTCGGG	CCCGAA		TIATSQLKSVVH
					AGAGTCTGTGTAGAGTTTGCTAATCATGCTG	GCATTA		RTIHAVPTEEDI
					TAGAAGCAGCCCGGTCTCGGCAAGATGCAC	CCCAGG		DNYLSGSNEGL
					CTCAAGTAAGGCCTGCTGCACAGAGACAGC	CTGCCC		LANSGNSGQAS
					CTCGCAGACCGACTAGACCTCGTCAACCCA	GTTTAG		SLWSRTRSAAR
					CGTTTGATGTCAGGGAGGCAAGCCGTTTGC	GCACAA		RLHLTLRATTSG
					AGAAGCTCTACAAACGCAGCAAGAAAAGA	ACCAAG		TVVVNQQADID
					GCGGTTCGGCACATTCTAAGAGATGACGCC	CTGCCA		HTRDILPASITRG
					CCCTCCTTCTCTGGCAGCAATGAGCAACTGC	GCCGAG		LRLIQRTTNAEK
					TTGACTACTTCAAAGAAATCTATGCTCCTCC	GCTCCA		LKSLPDQGKVA
					CGAGATTGATGAAAATAGAGCACAACAGTT	ACCAAG		RSLSNDPFANGS
					GGCAGAGTCTCTCTTTACCGATCTTGAGGA	CAACCA		SWHATGKFIRFC
					AGCGAAGGAGTCTGCGGCCGCCCTAATGTC	GCCCAG		DWRFIHRARLN
					TCCTATATCTCAGCAAGAGATCAGTACACG	GTTCCA		CLPTNVATKRW
					CCTTTCTCGAATGAGCAATTCTGCCCCTGGA	ATCAAG		KANANGKNGH
					AAGGACCGCCTCGAGTATAGACATATTCGA	CTGCCA		QQETLPHVLNH

					CAAGCTGATGGTGCATGCCGTGTTACCCAC	GCCGAG		CLPNMVPIRRRH
					ATCATGTTCAACAGGTGCCTACAGGAACAC	GCTCCA		DNIQQRLVTAIR
					CGTATACCATCAGCATGGAAGGAAGCTACA	ACCAAG		HGDVFVNQHVP
					ACCATCCTCATTCACAAGTCCGGTACAACC	CTGCCA		GDPNPRERPDIT
					GATGACCCAGCAAACTTCCGCCCGATCGCA	GCCGAG		VIEGNKVTVIDIS
					CTCCAGTCATGCTTGTACAAACTCTTCATGG	GCTCCA		VPFDNGPNACT
					GCATCCTCTCAGACAGAATGACACAGTGGG	ACCAGG		TAAQAKVEKYS
					CGTGCAATCACAACCTTCTCAGCCCAGAGC	CTGCCC		ALRQALRDMGR
					AGAAGAGCGCTCGCCCATGTGAGGGATGTC	ACCAAG		DVEVHGFIVGA
					ATGAACACACCTTTCTTCTGTCGTCTGTCAT	CTGCCA		LGTWHQGNERA
					CAAAGATACCAAGCGAAACCAGAAGACTG	ACTTAG		LGRLGVSRWYR
					CCAACATCGCCTGGCTAGATCTTAGGAACG	GCTCCA		TLMRKLCCIDAI
					CCTTCGGAAGCATCCCCCACCAGGCTATCC	ACCAAG		QASRDIWVEHV
					ATGCAGTCCTGACAACGATAGGAGCACCTG	CTGCCA		TGHRQYE (SEQ
					TCAGCCTGGTTATGCTCCTGAAGGACACAT	GCCCAG		ID NO: 1439)
					ACACGGGAGCATCCACATCCTTTCTTAGCA	GCTCCA		
					CGAGTGGAGAGACTGACCCAATACAGATCC	ACCAGG		
					AGTCCGGTGTTAAACAAGGATGCCCGATGT	TTGCCA		
					CTGCCATCCTCTTCAACCTAACCATTGAACT	CCCGAA		
					TATCATCAGAGCCGTCAAGAAGAAGGCAAC	TCTCTG		
					GGATGATGGCCTCGGGCTTGTGGTGCATGG	CCAGTT		
					ACAACGCCTATCCATCATGGCCTACGCTGA	TAGGCA		
					CGACCTCGTCCTGATGAGCAAGACACCAGA	CCAACC		

					AGGCCTGGATGCCATTTTGTCTGTGGCATCC	AAGCTG		
					GAGCAGGCAGAAACCCTACGTCTGGCATTC	CCAGCC		
					AAGCCTACCAAGTGTGCCTCCCTCTCACTCT	TAGGCT		
					CATGTAGACACGGTACATCTGTCCTTCCACG	CCAACC		
					CGAATACACCGTCCAAGGTCACCTGATGCC	AGGCTG		
					AGCTCTCGACGAGGAAGAGCAGTACCGCTA	CCACTC		
					CCTCGGAGTTCCCTTCGGCTTACCTCGTTTC	GAGGCT		
					ACAAATCTCAAGGACCTGATCGGCAAGCTG	CCAACC		
					AAGGGAAACATCGAGACCATTGCATCATCA	AGGCTT		
					TTGCTGGCCCCATGGCAAAAGTTGGATGCA	TCACCC		
					ATCAAAACCTTCGTCCAACCAGGTCTCAGTT	GAAGCA		
					TCGTACTGCGGGCAGCGGACTACTTGAAGT	TCACCC		
					CGGACTTGCGTTCTCTGAAATCTGCCATCAC	AGGCTG		
					CACCAACGTCAAAAAGATATGCCAGCTTCC	CCCTTT		
					CCTTCGGGCGGCCAACGCATACATCTTCGC	TAGGCA		
					GGCTAAGGAATCGGGTGGTCTAGCCTTCAT	CAAATC		
					AGACCCCAACGTAGATGCGGACATCCAAGT	AAGCTG		
					CATCACTCAAGCCGTTCGTGTCCTGTCGTCT	CCAGCC		
					GACGATGAAGTCGTTCAGACAATAGCAACA	GAGGCT		
					TCCCAACTCAAGTCGGTTGTTCATCGCACAA	CCAACC		
					TCCACGCAGTCCCTACTGAAGAGGACATCG	AAGCAA		
					ACAACTACTTGTCTGGCAGCAACGAGGGCC	CCAGCC		
					TGCTTGCCAATTCGGGTAATAGCGGCCAAG	CAGGTT		

					CTTCTTCTCTGTGGTCTCGAACCAGGTCAGC	CCAATG		
					CGCAAGAAGACTTCATCTTACCCTCAGAGC	AAGCTA		
					CACCACCAGTGGTACAGTTGTTGTCAACCA	CCAGCC		
					GCAGGCAGACATCGACCACACAAGGGACAT	GAGGCT		
					TCTACCAGCATCAATAACAAGAGGACTGAG	CCAACC		
					GCTCATCCAACGAACCACAAATGCTGAGAA	AAGCTG		
					ACTCAAGTCCCTACCAGACCAAGGGAAAGT	TCAGCC		
					TGCCAGATCCCTCAGCAACGATCCATTCGCT	GAGGCT		
					AACGGCTCTTCTTGGCATGCAACTGGAAAG	CCAACC		
					TTCATCCGCTTCTGCGATTGGCGGTTCATTC	AGGCTG		
					ACCGAGCCAGACTGAACTGCCTCCCCACGA	CCCACC		
					ATGTAGCCACGAAGAGGTGGAAGGCAAAT	AAGCTG		
					GCGAATGGCAAGAATGGCCATCAACAAGA	CCAGCT		
					GACACTGCCTCACGTACTGAATCATTGCCTA	TAGGTA		
					CCCAACATGGTCCCCATCAGACGCCGCCAT	CCAACC		
					GACAACATCCAACAACGACTGGTAACAGCC	AAGCTG		
					ATCAGACATGGCGATGTCTTTGTCAATCAA	CCAGCC		
					CATGTACCGGGAGACCCCAACCCAAGAGAA	CAGGCT		
					CGCCCAGATATCACTGTTATCGAGGGAAAC	CCAACC		
					AAAGTCACAGTCATCGATATTTCTGTCCCCT	AGGTTG		
					TCGACAATGGTCCCAACGCATGCACCACAG	CCACCC		
					CTGCACAAGCCAAAGTCGAAAAGTACTCTG	GAAACT		
					CCCTCCGTCAAGCCTTGAGAGATATGGGAA	CTGCCA		

					GAGACGTCGAAGTCCATGGCTTCATAGTTG GTGCCTTAGGAACCTGGCACCAGGGCAACG AACGTGCTCTCGGACGCCTGGGTGTCAGTC GCTGGTACAGAACACTGATGAGAAAGCTGT GCTGCATTGATGCTATCCAAGCTTCCAGAG ACATTTGGGTTGAACATGTCACTGGACATC GACAGTACGAGTGACGAAATGTTCAATATA TGACATTCTAATGTTCATGTATTTTTTGTTTG CTTGACAAAGCTAGATGAATATATTCCCTGT CCTACTACATACATCTCGATGTGCCCTTGCT AGACGATCCATTGGTAACCATGATATCAAT GGATTAACACATGATCTGTAAACAATATAT GATTTACACAATGTATTTATTGTTTCAATAA ATCTGTTCTTTTCACTTTAATACATGAAACA TGACTGTGCCTTCTCCAACTGGAGACCTACA TAATTTGTTAAAATGATATAAATATTTCGAA GATGAAATTATTATTAATAA (SEQ ID NO: 1071)	GTTTAG GCACCA ACCAAG CTGCCA GCCGAG GCTCCA ACCAGG CTGCCA CTCGAG GCTCCA ACCAGG CTTTCA CCCCAA GCATCA ACCAGG CTGCCC GTTTAG GCACAA ACCAAG CTGCCA GCCGAG GCTCCA ACCAGG		

						CCTCCA ACCAAG CTACCA GCCGAA GCTCTG CCAGAT TAGGCA CCAACC AAGCTG CCACCG TTAGAG GCCCAG AGCCCA CCAATC CCATAA CGTGTG AGATAT GTGAAG CTTCCT TCCACA CCTCCG CCGGTC TCCGTC		

						GCCACACGGCCAGGCTTCATCGCACCACTGCTGAACACACTGACGACAGC (SEQ ID NO: 1194)		
NeSL	Utopia-1_TSP	.		Trichinella spiralis	TTTCTGGTATGAATCCCAAGCGGATTCGTTACGAAATTTGCATAAGTTTTTGAAAAAATAGGCATTTGGTCGAGTGCTCGCACCACCATTTGTCGCGGGTCGTCCTGATATTGCACTACATTCAGGAACGGCCTATTCCCTTCGGGGAATTGTGTTTTAGGAATTGGAATCGGTTTGGTTACGATCGGTCGAGTGGTTCGTGAGATCGAGTGACAGCCGGGTGGCAGCGACAATGTGTAGTGCAAAGACACCAGCCTTGAAGTCGGGTCGCCGTCGCGGAAAAGAGGTCAACTATGAAGGCCAAATTGTACGGGTAGAAAGACGGCGTGGATCTCGCAGTAGTACGAGCGCAACGGACCTTGG	TTTCTGGTATGAATCCCAAGCGGATTCGTTA	TGAT	MCSAKTPALKS
						GTATGAATCCCAAGCGGATTCGTTA	CCGT	GRRRGKEVNYE
						ATCCCAAGCGGATTCGTTA	CCCG	GQIVRVERRRGS
						AGCGGAAGCGGATTCGTTA	AACC	RSSTSATDLGTR
						TTCGTTAGCGGATTCGTTA	AACG	MVTRGRKKLME
						ACGAAAAGCGGATTCGTTA	GAAC	ASVREAGHHGG
						TTTGCAAGCGGATTCGTTA	CACA	ESASTVDVDVV
						TAAGTTAGCGGATTCGTTA	TTGC	ESKKITGKTARR
						TTTGAAAGCGGATTCGTTA	CGCA	NRRAPSGDGKR
						AAAATAAGCGGATTCGTTA	TGAC	RESCGAECGQA
						GGCATTAGCGGATTCGTTA	TTCG	VCGNAVADRSE
						TGGTCGAGCGGATTCGTTA	ATTTC	ASSPRTPNVSKS

					TACGAGGATGGTTACGCGTGGGCGAAAGAA	AGTGCT	GCTTT	GRDKCGQPTIK
					GCTGATGGAGGCGTCGGTTCGTGAAGCGGG	CGCACC	TGCT	ASTPSPPKRKPT
					TCACCACGGAGGTGAATCGGCGAGTACTGT	ACCATT	CTTTT	TSSSPRTPCLSK
					GGACGTCGACGTGGTGGAATCGAAGAAAAT	TGTCGC	TGTA	RGARSKIPSTPD
					TACGGGTAAAACGGCGCGCCGTAATAGAAG	GGGTCG	TTTA	TPSTSGGSGKQR
					AGCGCCAAGCGGTGATGGAAAAAGACGAG	TCCTGA	ATTTT	VLVSPLLRTEKL
					AGAGCTGTGGTGCGGAATGTGGCCAAGCCG	TATTGC	GCTA	PDLEVLQRTEEQ
					TATGCGGTAACGCTGTTGCGGACCGTAGTG	ACTACA	TTAA	VTVRATFPIAQA
					AGGCCAGTTCGCCCCGCACTCCCAACGTTA	TTCAGG	CAAT	VVCPLGCEKPY
					GTAAATCTGGTCGTGACAAGTGTGGTCAAC	AACGGC	TCAG	TAVRPDGQFAH
					CCACTATAAAAGCCTCCACACCATCACCAC	CTATTC	TTTGT	QTLTRHFMRVH
					CGAAAAGAAAACCGACCACTTCGAGTAGCC	CCTTCG	TAAC	NCHSVQWHYRC
					CACGTACGCCATGCCTCTCCAAGCGCGGAG	GGGAAT	TGTTT	RNCNTDFLPAD
					CGAGATCCAAAATTCCGTCCACGCCAGATA	TGTGTT	TGTA	HRYPLRVVNTH
					CACCATCGACTTCTGGCGGTTCAGGTAAGC	TTAGGA	TTCAT	VRSCVSRWEITR
					AACGTGTTTTGGTGTCGCCGTTGCTTAGGAC	ATTGGA	TTGA	KLGESEDLHGV
					GGAAAAATTGCCCGACCTCGAAGTTCTACA	ATCGGT	AGAT	RCDLCDYVGVS
					AAGAACCGAGGAACAGGTGACTGTACGGG	TTGGTT	CCAA	KRAVGLHRRRH
					CAACTTTCCCAATCGCCCAGGCGGTGGTCT	ACGATC	ATAA	ANENIMQNTGT
					GCCCACTGGGATGCGAGAAACCGTACACGG	GGTCGA	AAC	AAQIEALSKQV
					CCGTGAGACCAGACGGTCAATTTGCGCACC	GTGGTT	(SEQ	GEIRVAGDYSQF
					AAACGCTGACTCGTCATTTCATGCGAGTCC	CGTGAG	ID NO:	KFGKRVRQYVA
					ACAATTGTCACAGCGTCCAGTGGCATTACC	ATCGAG	1318)	PTQRRDGLDEEE

					GGTGTCGCAATTGTAATACGGATTTCTTGCC	TGACAG		VHEAEEEEVPAE
					GGCAGATCACCGGTATCCCCTCCGAGTTGT	CCGGGT		SRTILGEPSTAT
					GAACACCCACGTTCGCAGCTGTGTTAGCCG	GGCAGC		AIGAEEISATGP
					GTGGGAGATAACCCGGAAGTTGGGAGAGA	GACA		VRADTAAQQMI
					GCGAAGATTTGCATGGAGTTCGTTGCGATC	(SEQ ID		CRIGQWCVWPQ
					TTTGCGATTACGTAGGCGTGTCGAAAAGAG	NO: 1195)		DYHSIPAPQCWT
					CCGTCGGATTGCATCGCCGCCGGCACGCCA			DTLMDLMIEQIV
					ACGAGAACATTATGCAAAATACGGGTACGG			LQRYPDGAGVS
					CCGCTCAAATCGAGGCCTTGAGCAAACAAG			VMSCSAVSAAI
					TCGGCGAAATACGTGTAGCAGGCGACTATA			HHEISAEFAAQV
					GTCAGTTTAAGTTTGGTAAACGGGTTCGGC			MSSHDASLYCII
					AATACGTGGCACCAACGCAGAGGCGAGAC			PVNVRNHWQMI
					GGGTTGGACGAAGAGGAGGTTCATGAAGCC			VLDVAERVVHY
					GAAGAAGAGGAGGTGCCAGCCGAATCACG			YCSLREHNTVV
					TACCATCCTAGGTGAACCCAGCACCGCAAC			LSSLLSLVELSG
					TGCTATAGGAGCTGAGGAGATCTCTGCAAC			KHTGCTSWKIET
					GGGTCCAGTTCGAGCAGACACAGCTGCTCA			HDGAPVQTNAF
					GCAGATGATCTGCCGGATCGGGCAATGGTG			DCGPFSCLFLKH
					TGTCTGGCCGCAGGACTACCATTCGATCCC			LLHGIDMNFGD
					GGCACCTCAGTGCTGGACGGACACGTTGAT			RESAALRTDLKF
					GGACCTCATGATTGAACAAATCGTTCTGCA			MIDAVSTPVVPA
					GCGCTATCCCGATGGTGCGGGTGTGAGCGT			TDKLKKKPDGS
					GATGAGCTGCTCAGCAGTAAGCGCAGCCAT			ATQLTQFQQKF

					ACACCACGAGATATCTGCAGAGTTTGCCGC			LSASDNWQSPD
					TCAGGTCATGTCTTCACATGATGCCAGTTTG			VDLQAVYDEVV
					TATTGCATCATTCCTGTGAATGTACGTAATC			ESIVSGHNEPSS
					ATTGGCAGATGATTGTGCTGGATGTGGCTG			RNTSRLQKKSPG
					AACGAGTAGTCCATTACTACTGTTCTCTCCG			KGGKGQVRRRS
					CGAGCATAACACTGTTGTTTTATCGTCTCTG			AVTRDPAWLKS
					TTATCGCTAGTGGAGCTGAGTGGTAAGCAC			ASAVQKAFNSA
					ACTGGCTGCACCTCGTGGAAGATCGAGACT			PARTVNAILRRP
					CACGATGGCGCGCCGGTGCAAACGAATGCG			NACPSFTATQV
					TTCGATTGTGGTCCGTTTTCGTGCTTGTTTTT			ADHYFNLRPAV
					GAAGCACCTGCTGCACGGCATCGACATGAA			TSLAPEVIDILPP
					TTTCGGTGATCGGGAGTCGGCGGCTTTGCG			PATDHSMLVAE
					CACCGACCTGAAGTTTATGATCGATGCGGT			LSESEVWEKMQ
					ATCCACTCCCGTCGTTCCGGCGACGGATAA			KAPNSAPGADRI
					GCTAAAGAAAAAACCTGACGGCTCAGCGAC			TIRMVRMADPG
					TCAGTTGACCCAGTTCCAGCAGAAGTTCCT			AMILTRFYRACL
					GAGCGCTTCTGACAACTGGCAATCTCCTGA			LRKWVPLQWK
					TGTCGACTTGCAAGCGGTGTATGACGAAGT			QSVCKLLYKDG
					GGTGGAGAGTATTGTGTCCGGACACAATGA			DKERLANWRPI
					ACCGAGCTCTCGTAACACCAGCCGTTTGCA			ALEPVLQRVLSA
					AAAGAAAAGTCCGGGTAAAGGAGGTAAAG			VVASRVTNWAR
					GACAGGTTCGCCGACGCAGTGCGGTAACGC			ANGLISLEAQKG
					GTGACCCGGCATGGTTGAAATCTGCGTCAG			FQPADGTSEHNF

					CGGTTCAAAAGGCGTTCAATTCGGCACCGG			VMEVAIQEARR
					CCCGTACTGTGAATGCAATCCTCCGGCGTCC			TNAQLAISWLDI
					GAATGCCTGTCCTTCGTTTACAGCCACGCAA			SNAFGTVSHQLL
					GTGGCAGATCATTATTTCAATCTCAGACCG			FSLLERYGLDPT
					GCGGTTACCTCGCTTGCACCCGAGGTTATTG			FTSFIQNLYKDA
					ACATCTTACCTCCGCCCGCCACTGACCACTC			TIVVKGANGTH
					GATGCTTGTTGCGGAGTTGAGTGAGTCGGA			VTARWSVGVRQ
					GGTCTGGGAAAAGATGCAGAAAGCACCCA			GDPCSGILFCLF
					ACTCTGCGCCTGGTGCGGACAGGATCACCA			VEPLLRSVLPSL
					TTCGTATGGTGCGGATGGCAGATCCAGGTG			PCEAETTAVNV
					CGATGATTCTCACCAGGTTCTACCGTGCCTG			LGQPITALAYAD
					TCTACTAAGAAAATGGGTGCCCCTGCAATG			DIALFAPSIGVM
					GAAGCAATCCGTGTGCAAGCTCCTTTATAA			QQQLCKIQGMA
					GGACGGGGATAAGGAGCGCCTCGCCAACTG			SAMGFRFNPKK
					GCGGCCTATCGCCCTGGAGCCCGTGTTACA			CASLYLNRAVV
					GCGGGTGCTCAGTGCGGTAGTCGCCAGCAG			NAATFTISGEEIP
					AGTGACGAACTGGGCGAGGGCAAACGGGTT			ALVHGDTFRYL
					GATTTCCTTGGAAGCCCAGAAAGGCTTCCA			GVAAGLGKPQT
					GCCAGCTGATGGTACCTCCGAGCACAATTT			PFSLLRENLREA
					CGTCATGGAGGTTGCAATTCAAGAGGCAAG			ELIFRSKLAPWQ
					ACGCACTAATGCACAGCTTGCCATCAGCTG			KMDAYRTYVLP
					GTTAGACATCTCCAACGCCTTTGGCACAGT			RLTFQLMIAKFN
					GTCGCACCAATTGTTGTTTTCGTTGTTGGAG			NIKQSAGQYDR

					CGCTATGGGCTGGACCCGACTTTTACCAGTT			AILRLVKRCFQL
					TTATCCAGAATTTGTATAAAGACGCGACCA			PVETSTDFIRAP
					TTGTAGTCAAAGGCGCCAATGGAACCCACG			RQCGGLGVPSL
					TTACCGCGCGCTGGTCAGTCGGAGTACGCC			RELYATAKVSR
					AAGGTGACCCTTGCAGTGGTATCTTGTTTTG			ALKMLWSPCRV
					TTTGTTTGTTGAGCCACTCTTGCGGTCAGTC			VSSLAASQLQR
					TTACCATCTTTACCGTGTGAGGCGGAAACG			VASAYFAKRLR
					ACTGCGGTTAACGTGTTAGGACAGCCGATC			DVEAADLSTFM
					ACAGCGCTGGCTTACGCTGATGACATCGCG			NAARSTPLDRSG
					CTGTTCGCCCCATCCATCGGGGTTATGCAAC			YPTCLWMDARK
					AACAGCTGTGCAAAATCCAGGGCATGGCCA			QMSYLTKVAGV
					GCGCCATGGGCTTTCGGTTCAACCCGAAGA			DCYFLVGEAGT
					AGTGCGCGTCCCTGTACTTGAATCGTGCAGT			SFFIRNGLGQTV
					AGTGAATGCCGCAACATTCACGATAAGCGG			SVLSPLRKNKV
					GGAGGAAATACCTGCTCTCGTACACGGGGA			MSVLGGAIQTR
					CACTTTTCGCTATCTTGGAGTAGCCGCTGGG			HLDAWLQCKRQ
					CTGGGTAAACCCCAGACTCCTTTCTCATTAC			GKTASCIVLDRS
					TGCGTGAGAACCTCCGAGAAGCGGAACTTA			SSRFITTGRYTSF
					TATTCCGGTCAAAGCTGGCTCCGTGGCAGA			AAMRFALPARL
					AGATGGATGCGTACCGGACTTACGTACTGC			DLLPCRARSSM
					CACGTTTGACATTTCAGCTTATGATTGCTAA			RSYQNCRRCGY
					ATTTAACAACATCAAGCAATCGGCCGGGCA			DRETLPHILQHC
					GTATGACCGCGCAATATTGAGGCTCGTGAA			RQFSAPAYQAR

					GCGTTGCTTCCAGCTTCCGGTGGAGACGTCC			HDAVQGRLETV
					ACCGACTTCATTAGAGCCCCCAGACAGTGT			MRRRFPNLRVN
					GGTGGTCTGGGCGTGCCATCATTGCGTGAG			RALPEIGSNKRP
					TTGTACGCAACCGCAAAGGTTTCGCGGGCA			DLVVVDEEKRL
					CTCAAGATGCTCTGGTCTCCCTGTCGGGTAG			VILLDIAIVFENT
					TCTCGAGCCTGGCAGCCAGTCAGCTTCAGC			AAAFVDARTRK
					GAGTTGCGTCCGCGTACTTTGCGAAGAGGT			WAHYEKEILAY
					TGAGAGATGTTGAGGCGGCTGACCTCTCTA			RLRGYSVTYDAI
					CCTTCATGAACGCCGCGCGATCTACGCCGC			VVGALGTWDPK
					TAGATCGGTCAGGGTATCCGACTTGCTTGTG			NDAILKRIGVVS
					GATGGACGCACGCAAGCAGATGTCGTACCT			QRYLRLMKVLV
					GACAAAGGTAGCCGGCGTTGACTGTTACTT			VSEMLEHSSRIY
					CCTCGTTGGCGAGGCTGGGACCTCTTTCTTC			RKHLGLRDLLP
					ATCCGGAATGGCCTCGGTCAGACGGTCTCA			DTGTKRRPVGT
					GTTCTAAGCCCCTTGAGGAAGAACAAAGTT			TETDPPGGDLRQ
					ATGAGCGTTTTAGGGGGGGCGATCCAAACT			KKRNTISARASG
					CGCCATCTGGACGCCTGGCTCCAATGCAAG			GKCLERRFTSPV
					CGCCAAGGTAAGACTGCGAGCTGCATTGTG			GTPSQRGELQC
					CTCGACCGGTCATCCTCCCGTTTTATAACGA			QPCPGPRRPALA
					CCGGACGGTACACCAGCTTTGCAGCTATGC			GIAPNPPSLQPR
					GCTTTGCGTTGCCCGCTCGCCTGGATTTGTT			KPPPRQHQKPVT
					ACCTTGCCGGGCGAGATCCTCTATGCGTTCG			KSTAH (SEQ ID
					TACCAGAACTGTCGGCGGTGCGGTTACGAC			NO: 1440)

					CGAGAGACCCTGCCGCACATCCTTCAGCAC TGTCGGCAGTTCAGCGCCCCCGCTTACCAG GCCCGGCATGACGCAGTTCAGGGACGCCTG GAAACGGTCATGCGTAGGCGGTTCCCCAAC CTCCGCGTAAACCGGGCTCTACCGGAGATT GGGTCGAACAAGCGGCCTGATCTAGTGGTC GTTGACGAGGAAAAGCGGTTAGTGATTCTG CTTGATATCGCAATCGTGTTTGAAAACACTG CCGCTGCCTTCGTCGATGCCCGGACCAGGA AGTGGGCGCACTATGAAAAGGAGATCCTCG CCTACCGGCTACGAGGATACTCCGTGACGT ACGACGCCATCGTCGTCGGTGCTCTGGGGA CGTGGGACCCGAAGAACGATGCCATCCTGA AAAGGATCGGTGTCGTCTCCCAAAGGTACC TACGGCTCATGAAGGTTCTCGTGGTGTCCG AGATGTTGGAACATTCTTCACGGATCTACA GAAAACACCTCGGCCTGCGAGACCTTCTCC CGGACACCGGCACGAAACGGCGACCGGTTG GCACGACGGAGACAGACCCACCCGGTGGTG ATCTCCGTCAGAAGAAGCGCAATACCATCA GCGCTAGGGCATCAGGTGGGAAATGCCTGG AACGGCGGTTCACGTCACCCGTCGGGACGC CCTCGCAGCGTGGTGAGCTGCAGTGCCAAC			

					CGTGCCCTGGCCCTAGAAGACCCGCCCTCG CTGGAATAGCGCCCAACCCGCCTAGTCTCC AGCCCAGAAAACCACCCCCGCGGCAACACC AGAAGCCGGTTACGAAGTCTACAGCGCATT GATCCGTCCCGAACCAACGGAACCACATTG CCGCATGACTTCGATTTCGCTTTTGCTCTTTT TGTATTTAATTTTGCTATTAACAATTCAGTT TGTTAACTGTTTTGTATTCATTTGAAGATCC AAATAAAAC (SEQ ID NO: 1072)			
NeSL	Utopia-2_CMy	.		Chelonia mydas	GGCAGAAACTGCACSTTCTAGAAGACTCAC TGCCTATCCTGAGGAAGACTACCGCTTTGG AGATGGATTCTACTGCCGCTGCTTCAAAAG AAAATCTTCATGCTGCTTCGGAGGCTCCAG GACAGASTGAGAAGATGATCGCTCCCTCGC CGATTCCACAGAAACCTTCAACTGCCGCTC GGACCGCTGCTGCTCCACGGAGCGCCCATC GGGGAACAGCCTCTAAGAGACCCTCAGCGC TTCCAGGACATCGCAGATGAGCAGCATCGA TTGGAAAAGCAGCGCCTCCCTGAAGAAAAC CCGTGGAGATGCTGCTGCAGGGAWATCTTG CACCTCGAGGACGGCCTCCCAGGACATCGC AGCCAGGACAAACATCATCTCGCCTGCTCT TCAGGASAAGGATGCCMSAAGAACTTMTCC	GGCAGA AACTGC ACSTTC TAGAAG ACTCAC TGCCTA TCCTGA GGAAGA CTACCG CTTTGG AGATGG ATTCTA CTGCCG CTGCTT	TGAG CTGG AGTG CCGA TGAG AAGC GCAG TCGG GAAA GTAA CTGA AATA CTTTC CTCA	MLQLRLPTPQTL RLLHPSQLPQSH STKRWSNIYEER ARAPKDTTIERV SAASKIPKLDPA KRRIGAPLQLM QGNSISRQLSAS SQYVQHNAVVR RVSAPPHTGNFT GSCRRKLALPH RPPSETQPAESY LHCWTTQRDPT LADQHGLQDHP TGLPPEGSHCIK

					CACCTCCTCMACTGCCCAGGATCCTSATGCT	CAAAAG	TGGA	DPRQNSSTEGQR
					GGTCGTCGTCCTGCTGTGCTGGAAGMKACC	AAAATC	TTGT	RSGNSQARRIPK
					CCAACTGGCGAGACCTCAGAAACCACCCAG	TTCATG	ATTTT	RASTAASKTVLP
					MWGGACCKCAACATCTCACTAGATGCCTGC	CTGCTT	CTAA	KRTSAALKSVR
					CCAGCCGAACCTCTCCATGCWACTCTCCCA	CGGAGG	ATGG	EDTALVLEDPA
					GAGCAACSAGAACCATCCAGMGAATCCGCT	CTCCAG	ACAA	KWSSQHREGXR
					GATATGACTGAAGCCMATCCAACAGAGGG	GACAGA	CCTA	QQANPTAVFQP
					AGAAGGAAAGGAGAATGACTGCATCTATCT	STGAGA	CCTA	EPAEIEQQPXVR
					CCAGTATCCCCTCCCTACGGACACGCTCCTC	AGATGA	ATTCT	AATPWQAAWM
					TKCCCCTTCCGCTATCCGAGGGTTCCAGTAC	TCGCTC	CAAT	EELARTASFXDF
					ATTGGCAGTCTCAGCAMACACCTCAAGAGA	CCTCGC	TACT	DLLVDRLTKDL
					ATCCATASCAAGCGGATCACCTTCCGGTGTG	CGATTC	GAGG	SAEIVSGRKGTQ
					CCCTCTSCGACCTGCCTTTCGAGACGCAGAT	CACAGA	GACA	ENTPTAHRQNQ
					GAAATGTAAGTCTCATCAAGTCACCTGCAA	AACCTT	ATCT	NNMREARRRNI
					AGGACATCTCGAACTGGAAGAGTCCAACTT	CAACTG	CCAC	SRCYDPAAASRI
					TACCAGTCTATGTTGCCGCCACCCCATCTCT	CCGCTC	TCATT	QKLYRSNRPKA
					GCTCCGAAAGCAGAAACACCACATGCTCCA	GGACCG	GATA	MREILDGPSSYC
					GCTCCGGCTTCCGACTCCCCAAACACTCCG	CTGCTG	TATTT	AIPSERLFLYFK
					GCTACTGCATCCCAGCCAACTCCCCCAGAG	CTCCAC	TGCTT	GVFDRVAQND
					CCACAGCACGAAGAGGTGGTCCAACATCTA	GGAGCG	TCCA	MQRPECLXPXP
					TGAGGAAAGGGCCCGGGCGCCCAAGGATA	CCCATC	CAAC	RVDYAEDXEQD
					CTACCATTGAAAGAGTAAGTGCTGCCTCAA	GGGGAA	CAAA	FTSWEVEARLT
					AAATCCCTAAGCTGGACCCTGCCAAGAGGA	CAGCCT	TCTCT	KTKNTAPGKDG

					GGATCGGCGCTCCCTTGCAGCTTATGCAAG	CTAAGA	GTAC	IRYNFLKKRDPG
					GCAACTCCATCTCCAGACAGCTTAGTGCTTC	GACCCT	AACT	CLVLTAIFNKCK
					CTCGCAGTATGTCCAGCACAACGCCGTCGT	CAGCGC	TTTCA	QFRRTPSSWKKS
					CAGAAGGGTCAGCGCCCCTCCACACACGGG	TTCCAG	TGAG	MMVLVYKKGK
					TAACTTCACGGGATCCTGTCGTCGGAAGCT	GACATC	TGAT	QDNPNTRRPISL
					AGCACTACCTCACAGGCCACCCAGCGAGAC	GCAGAT	GTAC	CSTMYKLYASC
					CCAGCCAGCGGAGAGCTATCTCCACTGCTG	GAGCAG	CCGA	LAARITDWSVN
					GACCACGCAACGGGACCCCACCCTAGCGGA	CATCGA	GTAC	GGAISSIQKGFM
					CCAGCACGGCCTCCAAGACCACCCAACGGG	TTGGAA	TTGG	SCKGCYEHNFV
					ATTGCCCCCAGAAGGCTCACACTGCATCAA	AAGCAG	ATTCT	LQTAIHMARRA
					AGACCCCAGACAGAACTCATCCACAGAAGG	CGCCTC	AATA	WRQCAIAWLDL
					GCAGCGTCGCTCCGGAAACTCCCAAGCAAG	CCTGAA	TCTA	ANAFGSMPHQH
					ACGCATCCCTAAAAGGGCCAGTACCGCTGC	GAAAAC	AACT	IFDMLREFGMPE
					CTCTAAAACTGTCCTTCCTAAAAGGACCAG	CCGTGG	GTAT	NFLQLVRELYE
					TGCTGCCCTAAAATCCGTAAGGGAAGACAC	AGATGC	TGTT	GCTTTICSMEGE
					TGCCCTCGTTCTGGAAGATCCGGCCAAATG	TGCTGC	AAAT	TPEIPIRSGVKQG
					GAGTTCTCAACACCGGGAGGGAAAWAGGC	AGGGA	CTATT	CPLSPIVFNLAM
					AACAGGCTAATCCCACAGCAGTTTTCCAGC	WATCTT	CACC	EPLIRAISSGLGG
					CTGAGCCTGCGGAGATTGAGCAGCAGCCCW	GCACCT	TAAA	FDLYDNRVNILA
					CGGTGAGGGCCGCCACACCATGGCAAGCCG	CGAGGA	TTTG	YADDLVLIADN
					CTTGGATGGAGGAGTTGGCGAGGACAGCCT	CGGCCT	GGTT	PESLQQMLDITS
					CCTTCGAKGACTTTGACCTCCTTGTAGACAG	CCCAGG	ATTG	QAANWMGLRF
					ACTCACTAAAGACCTGTCTGCAGAAATTGT	ACATCG	CTGA	NARKCASLHID

					ATCTGGGAGGAAGGGAACTCAGGAGAACA	CAGCCA	TTAT	GSRRDSVQATSF
					CACCAACTGCKCACAGACAGAACCAAAACA	GGACAA	GTAC	QIQGEPMIFLED
					ACATGAGGGAAGCCAGGAGAAGGAACATC	ACATCA	TCTAT	GQAYQHLGTPT
					AGTCGCTGCTATGACCCAGCAGCAGCATCC	TCTCGC	GTAT	GFRVQQTPEDTI
					AGGATCCAGAAGCTCTACCGGTCAAACCGC	CTGCTC	CATA	AEILRDVARIDS
					CCTAAGGCTATGAGAGAGATCCTCGATGGG	TTCAGG	TGAC	SLLAPWQKINAL
					CCCTCATCCTACTGCGCAATCCCGTCCGAGA	ASAAGG	TTTTA	NTFLIPRISFVLR
					GACTATTCTTGTACTTCAAGGGGGTGTTTGA	ATGCCM	AAAA	GSAMVKVPLNK
					CCGTGTAGCTCAGAACGACATGCAGCGTCC	SAAGAA	CAAA	ADNTIRQLVKK
					AGAGTGCCTTCAKCCTTWACCCCGGGTTGA	CTTMTC	CTTTG	WMFLPQRASNE
					CTATGCAGAGGACCWGGAACAAGACTTTAC	CCACCT	TATTT	LVYISHRQGGA
					ATCATGGGAAGTAGAGGCCAGACTTACAAA	CCTCMA	GTGG	NVPRMGDLCDV
					GACCAAAAACACAGCCCCAGGAAAAGATG	CTGCCC	ATAA	AVITHAFRLLTC
					GCATTCGCTACAACTTCCTGAAAAAACGAG	AGGATC	TCTA	PDAMVRNIAES
					ACCCCGGTTGCCTGGTACTAACTGCCATTTT	CTSATG	AGCA	ALQDAVKKRIA
					CAACAAATGCAAACAGTTCCGCCGTACTCC	CTGGTC	CTAT	RTPSNQDVATY
					CAGCTCCTGGAAGAAATCTATGATGGTGCT	GTCGTC	ACCC	LSGSLEGEFGRD
					CGTCTACAAGAAAGGCAAGCAAGACAACCC	CTGCTG	AGAT	GGDFASLWTRA
					CAACACTCGGAGACCCATCTCTCTCTGTTCC	TGCTGG	GTAC	RNATRRLEKRIG
					ACCATGTACAAACTGTATGCCAGCTGCCTT	AAGMK	AGAC	CHWTWCEERQE
					GCGGCTAGGATCACAGACTGGTCGGTGAAC	ACCCCA	ACTC	LGVLVPQVKNT
					GGAGGAGCCATCAGCTCCATCCAGAAAGGC	ACTGGC	TTTTC	DHTIITPRARTM
					TTCATGTCATGCAAAGGCTGCTACGAACAC	GAGACC	CCAA	LERTLKDAIRCQ

					AACTTTGTCCTTCAGACTGCCATCCACATGG	TCAGAA	CCTA	YVENLKRKPDQ
					CCAGGAGGGCATGGAGGCAATGTGCCATAG	ACCACC	TGTA	GKAFEVTCKWD
					CATGGCTCGACCTGGCTAATGCTTTTGGATC	CAGMW	TTAT	ASNHFLPGGSFT
					GATGCCCCACCAGCACATTTTTGACATGCTG	GGACCK	ATTTT	RFADWRFIHRA
					CGAGAGTTCGGGATGCCTGAAAACTTTCTC	CAACAT	TTTA	RLNCVPLNGAV
					CAACTGGTCCGGGAACTGTATGAAGGCTGC	CTCACT	ACAT	RHGNRDKRCRK
					ACCACCACCATCTGCTCCATGGAAGGAGAG	AGATGC	TAGC	CGYANETLPHV
					ACACCCGAAATTCCTATCCGTAGCGGCGTG	CTGCCC	TTTA	LCSCKPHSRAW
					AAGCAAGGCTGCCCCCTGAGCCCCATCGTT	AGCCGA	ATAA	QLRHNAIQDRL
					TTCAACTTAGCCATGGAGCCGCTCATTCGA	ACCTCT	AATT	ARAIPPPVGKVA
					GCGATCTCCAGCGGTCTAGGCGGTTTCGAT	CCATGC	TTTA	VNSAIPGTDSQL
					CTGTATGACAACAGAGTGAATATCCTGGCC	WACTCT	AA	RPDIVITNEDRK
					TACGCAGACGATCTGGTCCTGATCGCAGAC	CCCAGA	(SEQ	KIIMVDVTVPFE
					AACCCCGAGAGTCTCCAACAAATGCTTGAC	GCAACS	ID NO:	NRTPAFHDARA
					ATCACCAGCCAGGCTGCCAACTGGATGGGA	AGAACC	1319)	RKVEKYAPLAE
					CTCCGCTTCAATGCCAGGAAGTGTGCATCC	ATCCAG		TLRAKGYQVQT
					CTGCATATCGATGGCAGTAGAAGGGATTCG	MGAATC		HALIVGALGAW
					GTCCAGGCGACGTCTTTTCAGATCCAGGGT	CGCTGA		DPSNERVLRECG
					GAACCCATGATCTTCCTCGAGGACGGGCAA	TATGAC		IGQRYARLMRQ
					GCGTACCAACATCTCGGCACGCCGACGGGT	TGAAGC		LMVSDAIRWSR
					TTCCGCGTCCAGCAGACACCTGAGGATACC	CMATCC		DIYIEHITGHRQ
					ATCGCGGAGATCCTACGAGATGTGGCCAGG	AACAGA		YQEG (SEQ ID
					ATCGACTCCTCCCTACTGGCACCATGGCAG	GGGAGA		NO: 1441)

					AAGATCAACGCCTTGAACACCTTCCTGATC	AGGAAA		
					CCCCGTATCTCATTCGTCCTGAGGGGATCTG	GGAGAA		
					CCATGGTGAAGGTACCTCTGAACAAGGCAG	TGACTG		
					ACAACACCATCCGGCAGCTGGTGAAGAAGT	CATCTA		
					GGATGTTTCTTCCCCAGAGGGCCAGCAATG	TCTCCA		
					AACTGGTGTACATCTCGCACAGGCAGGGCG	GTATCC		
					GTGCCAACGTCCCTCGAATGGGTGATCTGT	CCTCCC		
					GCGACGTTGCCGTGATCACTCATGCCTTCCG	TACGGA		
					CCTTCTGACGTGCCCGGATGCCATGGTGAG	CACGCT		
					GAACATCGCGGAGAGTGCTCTGCAGGATGC	CCTCTK		
					CGTCAAGAAGCGAATCGCCAGGACCCCCTC	CCCCTT		
					CAACCAAGATGTTGCCACCTACCTGAGCGG	CCGCTA		
					CTCGCTGGAAGGCGAATTTGGAAGAGACGG	TCCGAG		
					GGGAGACTTTGCTTCGCTCTGGACTCGTGCC	GGTTCC		
					CGCAATGCTACGCGACGACTGGAGAAGCGT	AGTACA		
					ATCGGCTGCCACTGGACGTGGTGCGAGGAA	TTGGCA		
					CGCCAGGAGCTGGGAGTCCTGGTGCCACAG	GTCTCA		
					GTGAAGAACACAGATCACACTATCATCACT	GCAMAC		
					CCGAGAGCTAGAACCATGCTGGAGAGGACC	ACCTCA		
					CTGAAGGATGCCATCCGCTGCCAATACGTG	AGAGAA		
					GAAAACCTGAAACGGAAGCCGGACCAGGG	TCCATA		
					CAAGGCATTCGAGGTGACGTGCAAGTGGGA	SCAAGC		
					TGCCAGCAACCACTTCCTCCCCGGGGGCAG	GGATCA		

					CTTCACCCGATTTGCCGACTGGAGGTTCATC	CCTTCC		
					CACAGGGCCCGGCTCAACTGCGTCCCGCTG	GGTGTG		
					AACGGAGCCGTCCGCCACGGGAATCGGGAC	CCCTCT		
					AAGCGATGCAGGAAGTGCGGCTATGCCAAC	SCGACC		
					GAGACGCTACCCCACGTCCTGTGTAGCTGC	TGCCTT		
					AAGCCCCATTCCAGAGCCTGGCAGCTGCGA	TCGAGA		
					CACAACGCCATCCAAGATCGCCTGGCCAGA	CGCAGA		
					GCCATCCCGCCACCCGTGGGGAAGGTTGCC	TGAAAT		
					GTAAACTCCGCCATCCCCGGAACCGACAGC	GTAAGT		
					CAACTGCGACCGGACATCGTCATCACCAAC	CTCATC		
					GAGGACCGGAAGAAGATCATCATGGTGGAC	AAGTCA		
					GTCACAGTGCCCTTCGAGAACAGGACCCCA	CCTGCA		
					GCCTTCCACGATGCCCGAGCTCGAAAGGTG	AAGGAC		
					GAGAAATATGCCCCTCTGGCCGAAACCTTG	ATCTCG		
					AGAGCTAAGGGTTACCAGGTCCAGACACAC	AACTGG		
					GCGCTGATCGTCGGAGCCTTGGGCGCATGG	AAGAGT		
					GACCCCAGTAACGAGCGAGTGCTGAGAGAA	CCAACT		
					TGCGGAATCGGTCAACGCTACGCTCGGCTG	TTACCA		
					ATGCGGCAACTCATGGTGTCAGATGCCATC	GTCTAT		
					AGGTGGTCGAGGGACATCTACATAGAACAC	GTTGCC		
					ATCACCGGACATCGGCAATACCAGGAGGGA	GCCACC		
					TGAGCTGGAGTGCCGATGAGAAGCGCAGTC	CCATCT		
					GGGAAAGTAACTGAAATACTTTCCTCATGG	CTGCTC		

					ATTGTATTTTCTAAATGGACAACCTACCTAA TTCTCAATTACTGAGGGACAATCTCCACTCA TTGATATATTTTGCTTTCCACAACCAAATCT CTGTACAACTTTTCATGAGTGATGTACCCGA GTACTTGGATTCTAATATCTAAACTGTATTG TTAAATCTATTCACCTAAATTTGGGTTATTG CTGATTATGTACTCTATGTATCATATGACTT TTAAAAACAAACTTTGTATTTGTGGATAATC TAAGCACTATACCCAGATGTACAGACACTC TTTTCCCAACCTATGTATTATATTTTTTTAAC ATTAGCTTTAATAAAATTTTTAAA (SEQ ID NO: 1073)	CGAAAG CAGAAA CACCAC (SEQ ID NO: 1196)		
NeSL	Utopia-2_PCa	.		Phytophthora capsici	ACCGCCCAAGTCTCCACCGCAGCTGCGACT GCTGCTGCTCATCGAGCCGCAGTAGTCGCA GCTGTACAAGCACCACCTCCTACCGGACGC GCCGTCGTGGAGCACCACGCGCGCGCTGAG CCGCNTCAAGACCAAGAATTCCAGGCTCGC GCGCGCGTGGCGCTTCAAGATGGCGACGCG ACCAGCAGGAGTGGCCGCTCTACTGACGGC GGAGACGAGGCCAAGACCGAGGACAGCGA CGCGGACGCAGCCAAGCGCCAAGCCGCTCA AGCAGCACCGAGACCCCATACCCMGGACA CCGCGCCGTACGACGGCCACAGGGACTCGG	ACCGCC CAAGTC TCCACC GCAGCT GCGACT GCTGCT GCTCAT CGAGCC GCAGTA GTCGCA GCTGTA	TAAG CTGG TCATT TGAC CGAC AGGG CACT ACCC AGGT AGGG AACC	MQDMEEELLLD VEMETETTEPQT STAXDATTTTD RPTRWGPHPRA VAAAAIAQLVT GEXAXPALPSR QDRRPAPRSHAP TRSRWGPRHQA VGAAAIASLAT GLPASAAPVSRA TKHGEGRRRLQ

					TGCTGGCTGTGTACGCACACAACGCACCTG	CAAGCA	GCCC	TRWGPRVSIPRA
					CATTCACCTGCGCGTTGTGTGTGTACACAGC	CCACCT	TTCA	ARRPGSRWGPP
					ACGCAACTTCGCCGAGCTGACCAAGCATCG	CCTACC	GAAA	RAAGASGQLPA
					CCATGCAGCGCATCGTCACACCCGCTTCGT	GGACGC	CCCA	SASGATGQLPEH
					GGATCACTTCCACAGCGGGTGCACGTGTGG	GCCGTC	GGAA	VEAITTTPRVAS
					CATCGGCTTCCAGTCGCGCGCGGCAGCTAC	GTGGAG	GACA	DADEGPTPPDP
					GCGACACGCTCAAGCCTGTGCAGACAGCAC	CACCAC	CAAA	WILRFDGACRR
					ACACGCCACCGTAGCTGCCTCGCGCGACCC	GCGCGC	CACC	NPGPGGAGAAL
					GGCMNCCASTGCSSCCSGNGCCGGMAGCGA	GCTGAG	CTCC	FKPSGAVVWTC
					SGAGGAGGACCNCGCACCCCCCGGTCCTCT	CCGCNT	CTTTA	SHYMPSSNETN
					SSTCGCGGCAGCATCTGCAGCTGCAGCAGC	CAAGAC	GTGA	NTAEYTALLLG
					CGCATCAAGCGCCACTCCAGCTGCAGATAC	CAAGAA	CATA	VQSAVHHGATR
					TGCCACCACGCAGAGCGCCGTGCCCATCGC	TTCCAG	CATA	LDIEGDSSLVIA
					TACTCCTGGGCCCCAGTACGCCCCCCACGT	GCTCGC	TTTTA	QVKGTFACRNA
					GCTGGAGCCACCTCCAGAGCTCCGAGTTTC	GCGCGC	GGCT	KLRQLRNRVRH
					CGGCAAACGCCGMCGCCTCAACACGCCGAT	GTGGCG	ACAT	ALRSVEKYTLR
					CGACCTGCAGCCGCTGGACGTGGACGCGCT	CTTCAA	TTCA	HIDRKANAHAD
					GATGCAGGACATGGAGGAGGAGCTGCTCCT	GATGGC	GTTA	RLANRALDRRS
					GGACGTGGAGATGGAGACGGAGACGACGG	GACGCG	CGGA	SSSECEPHGSCM
					AACCGCAGACCTCGACGGCCAMGGACGCC	ACCAGC	GAGG	ERCCGTDTTPA
					ACCACCACCACGGACAGGCCAACACGATGG	AGGAGT	TTACT	VQGPTPQAAAA
					GGCCCCCACCCCAGAGCAGTGGCGGCTGCA	GGCCGC	AACT	VPVQVWPQWQ
					GCCATMGCGCAGCTGGTGACCGGCGAGSCA	TCTACT	GGTA	RQTMVAWTTSH

					GCAGMTCCAGCTCTGCCTTCGAGACAGGAC	GACGGC	AATA	GGRCRDCSTRC
					CGCCGCCCAGCACCACGATCCCACGCACCC	GGAGAC	AAAA	GRSLPSLAHRPR
					ACCCGGTCTCGATGGGGACCGCGCCATCAA	GAGGCC	GCAC	LSPRRQPRLRLR
					GCTGTCGGTGCCGCAGCCATCGCGTCGCTA	AAGACC	TTT	QLSDEERDXAA
					GCTACAGGACTACCGGCCTCTGCGGCACCC	GAGGAC	(SEQ	DALQELSDVMA
					GTGTCTCGCGCAACCAAGCACGGGGAAGGA	AGCGAC	ID NO:	SKIVDADSWDT
					CGACGCCGTCTCCAGACACGATGGGGACCT	GCGGAC	1320)	GEGYISSIPERIR
					CGGGTGAGCATCCCACGAGCCGCACGCCGT	GCAGCC		EVLQPYTTRPPR
					CCAGGATCGAGGTGGGGACCTCCACGCGCG	AAGCGC		PGHQQQQRRRP
					GCAGGGGCCTCTGGCCAACTTCCTGCCAGC	CAAGCC		PRVTRNQREHR
					GCCTCCGGGGCCACTGGCCAACTTCCGGAG	GCTCAA		LDEALDDMQAT
					CATGTGGAGGCCATCACCACGACACCACGT	GCAGCA		QQAAPRDQRAI
					GTGGCATCGGACGCTGATGAAGGACCGACA	CCGAGA		HRARRRVGRVR
					CCGCCGGATCCGTGGATACTCCGCTTCGAC	CCCCAT		ASMAKQELRQA
					GGCGCATGCAGGCGTAACCCCGGACCGGGC	ACCCMG		FAKDESKCVSKI
					GGTGCAGGTGCTGCATTGTTTAAGCCGAGT	GACACC		LAGASAETAAE
					GGCGCCGTGGTGTGGACGTGCTCCCACTAC	GCGCCG		EHVDECPIDAAT
					ATGCCAAGCAGCAACGAAACCAACAACACC	TACGAC		LHAHFTGTNAP
					GCCGAGTACACGGCACTGCTGCTGGGCGTA	GGCCAC		RTDFDYDAACG
					CAGAGCGCGGTCCACCACGGTGCCACTCGC	AGGGAC		QEFRGALDSMQ
					CTAGACATCGAGGGCGACAGCAGCCTGGTG	TCGGTG		PPTVATDAFEEE
					ATTGCCCAGGTCAAGGGCACATTCGCGTGC	CTGGCT		LTIDEVEDQLTR
					AGGAACGCGAAACTCCGGCAGCTCCGCAAC	GTGTAC		AAKTSSPGHDGI

					CGCGTCCGGCATGCGCTCCGCTCCGTGGAG	GCACAC		GYDIYSRFAAQL
					AAGTACACCCTCCGCCACATCGACCGCAAG	AACGCA		VPLLHAAYQFC
					GCCAACGCCCACGCAGACCGTCTGGCCAAC	CCTGCA		WLHRRVPALW
					CGCGCGCTGGACCGGCGGTCGAGCAGCAGC	TTCACC		KLGIVRLIHKKG
					GAGTGCGAGCCACACGGCTCGTGCATGGAG	TGCGCG		DPMQPTNWRPI
					CGATGCTGTGGCACCGACACGACACCAGCG	TTGTGT		CLQPAIYKIYSG
					GTGCAAGGACCGACACCGCAAGCAGCAGC	GTGTAC		LLARRLSRWME
					AGCAGTGCCAGTGCAGGTGTGGCCCCAGTG	ACAGCA		QNQRLPMAQKG
					GCAGCGGCAGACGATGGTAGCGTGGACGAC	CGCAAC		FRAFNGCHEHN
					GAGCCATGGAGGTCGATGCCGAGATTGCAG	TTCGCC		FVATTLLDQTRR
					CACGAGATGCGGGCGAAGCCTTCCCAGTCT	GAGCTG		SHRRLYQVWYD
					TGCCCATCGGCCCAGGCTCAGCCCCAGGCG	ACCAAG		LRNAFGSLPQQL
					GCAACCTCGGCTTCGACTGCGCCAACTCAG	CATCGC		MWSVLRHLGV
					CGACGAGGAACGCGACRMAGCAGCGGACG	CATGCA		DASFIARCKNIY
					CCCTCCAGGAGCTTAGCGACGTCATGGCCA	GCGCAT		QDSAFVVANAV
					GCAAGATCGTTGACGCAGACAGTTGGGACA	CGTCAC		DGATDPVRQEV
					CCGGCGAAGGCTACATCAGCTCCATACCGG	ACCCGC		GVYQGCPLSPLL
					AGCGCATCCGAGAGGTGCTGCAGCCGTACA	TTCGTG		FISALVPLIRRLE
					CCACGCGACCACCACGGCCTGGGCACCAGC	GATCAC		KLDGVGVPLAE
					AGCAGCAACGGCGACGGCCACCACGTGTGA	TTCCAC		GVRPCATAYAD
					CACGCAACCAGCGCGAACACCGGCTAGACG	AGCGGG		DIKVFSDSAAGI
					AAGCGCTGGACGACATGCAAGCCACGCAGC	TGCACG		RKCHDAVTRFL
					AAGCAGCACCGCGGGACCAACGGGCGATA	TGTGGC		EWTGLRANPGK

					CATCGAGCGCGGCGTCGCGTCGGGCGTGTC	ATCGGC		CASLAVTTNAR
					CGGGCATCGATGGCCAAGCAAGAGCTACGG	TTCCAG		GNPVRDDGVHL
					CAGGCGTTCGCCAAGGACGAGTCCAAGTGT	TCGCGC		ELQGEVIAPLSL
					GTGTCCAAGATATTGGCCGGTGCCTCGGCG	GCGGCA		HDSYRYLGVGD
					GAGACAGCGGCGGAAGAGCACGTCGACGA	GCTACG		GFDHVRHRLQL
					GTGCCCGATTGACGCCGCCACGCTGCATGC	CGACAC		EPKLQQIKREAV
					CCACTTCACAGGCACCAATGCGCCACGGAC	GCTCAA		ALMQSGLAGW
					GGACTTCGACTACGACGCTGCATGCGGTCA	GCCTGT		QVVKALKTFVY
					AGAATTCCGAGGAGCGCTGGACAGCATGCA	GCAGAC		PKVEYALRHLR
					GCCGCCGACCGTTGCGACCGACGCGTTCGA	AGCACA		PLQSQLQGFDR
					GGAAGAGTTGACGATCGACGAGGTGGAGG	CACGCC		AVVRGLRHLLR
					ACCAGCTTACACGCGCGGCCAAGACCTCCA	ACCGTA		LPQSATTEFFYT
					GCCCGGGACATGACGGCATCGGCTACGACA	GCTGCC		PTSGGGLGLQSL
					TCTACAGTCGCTTCGCCGCGCAGCTCGTGCC	TCGCGC		VEMHQALQVA
					ACTGCTCCACGCCGCGTACCAGTTCTGCTGG	GACCCG		HAWQMLHSKD
					CTGCATCGGCGCGTGCCAGCCCTGTGGAAG	GCMNCC		AAVVAVAKEQV
					TTGGGCATCGTCCGCCTCATCCACAAGAAG	ASTGCSS		CQVARKRYRLQ
					GGCGACCCGATGCAGCCAACGAACTGGCGA	CCSGNG		EEHWRGRGDEL
					CCGATCTGTCTGCAGCCCGCGATTTACAAG	CCGGMA		VRLFLNSELAAS
					ATCTACAGCGGACTGCTGGCGCGTCGGCTG	GCGASG		PFADCLRRNGDI
					TCGCGCTGGATGGAGCAGAACCAGCGACTC	AGGAGG		GSLWTDVQRTL
					CCCATGGCACAGAAGGGGTTCCGAGCGTTC	ACCNCG		RLHHLSLTAQD
					AATGGTTGCCACGAACACAATTTCGTGGCA	CACCCC		DRDGQDPLALR

					ACCACGCTCCTGGATCAGACGCGCCGGTCG	CCGGTC		VPHHTKWLDHK
					CACCGCAGGCTGTACCAAGTCTGGTACGAC	CTCTSST		TVLRHVKLHMK
					CTGCGGAACGCGTTCGGGTCACTGCCGCAG	CGCGGC		IRHQTRWKGLV
					CAGCTCATGTGGAGCGTGCTCCGCCACCTC	AGCATC		DQGKTVRVHGG
					GGCGTCGACGCAAGCTTCATCGCCCGGTGC	TGCAGC		LGAKFVSTGAG
					AAGAACATCTACCAAGACTCGGCTTTCGTG	TGCAGC		LSDDAYRFGVK
					GTGGCGAACGCGGTGGACGGCGCGACGGAT	AGCCGC		ARLNQVDTNAV
					CCTGTCCGCCAGGAGGTGGGCGTGTACCAA	ATCAAG		LKRKRLRSSKTC
					GGTTGTCCCCTGAGCCCCCTCCTCTTCATCT	CGCCAC		RDPTCSSAETLA
					CGGCGCTTGTGCCGCTCATTCGCCGCCTCGA	TCCAGC		HALNHCASNMD
					GAAGCTCGACGGCGTAGGGGTGCCGTTGGC	TGCAGA		AIRQRHDDALE
					CGAGGGTGTACGACCGTGTGCCACCGCCTA	TACTGC		QIGSKIRGALER
					CGCCGACGACATCAAGGTGTTCAGTGACAG	CACCAC		AKSTTELRLNQT
					CGCCGCTGGCATCCGTAAATGCCACGACGC	GCAGAG		VPEYTGAALRP
					TGTCACTCGTTTCCTCGAGTGGACAGGCCTG	CGCCGT		DIVLRNVAAKK
					CGGGCAAACCCCGGCAAGTGTGCCAGCCTA	GCCCAT		MVIADLAVTFE
					GCGGTGACGACAAACGCCCGCGGCAACCCG	CGCTAC		DHAAGARHSSL
					GTCCGCGACGACGGCGTGCACCTAGAGCTG	TCCTGG		QLSHDHKTLKY
					CAAGGCGAAGTCATCGCCCCCCTCAGTCTC	GCCCCA		QPIVAELRVQG
					CACGATAGCTATCGCTACCTGGGTGTGGGC	GTACGC		WQVQTAAIVYG
					GATGGCTTCGACCATGTGCGCCACCGCCTC	CCCCCA		SLGSVQPSNFKT
					CAGCTTGAACCCAAGCTCCAACAGATCAAG	CGTGCT		YTEKLKLHKRE
					CGAGAGGCGGTGGCGCTGATGCAGTCCGGA	GGAGCC		ARQLDLQLSSH

					CTGGCGGGGTGGCAGGTGGTGAAGGCGCTC AAGACCTTCGTGTACCCCAAGGTCGAGTAC GCGCTGCGCCACCTACGACCACTCCAGTCC CAGCTCCAAGGCTTCGACCGCGCCGTCGTC CGTGGCCTGCGCCACCTACTGCGCCTCCCCC AGTCCGCCACCACCGAGTTCTTCTACACACC AACGTCTGGAGGTGGGCTGGGGCTGCAGTC CCTCGTGGAGATGCATCAGGCGCTGCAGGT GGCGCATGCCTGGCAGATGCTCCATTCCAA GGACGCGGCCGTCGTGGCCGTTGCGAAGGA GCAGGTCTGCCAGGTAGCACGCAAGCGCTA CCGCCTGCAGGAAGAACATTGGCGCGGGCG TGGCGACGAGCTCGTGAGGCTCTTCCTCAA CTCCGAGCTCGCCGCCTCGCCCTTCGCTGAC TGCCTCCGCCGCAACGGCGACATCGGATCC CTCTGGACGGACGTGCAGCGCACGCTCCGC CTCCATCACCTCAGCCTGACGGCTCAAGAC GACAGAGATGGCCAAGACCCGCTTGCCCTT CGCGTGCCGCACCACACCAAGTGGCTCGAC CACAAGACTGTGCTGCGGCACGTGAAGTTG CACATGAAGATCCGCCACCAGACCCGATGG AAGGGGCTAGTGGACCAGGGCAAGACGGT ACGCGTGCACGGTGGCCTCGGTGCGAAGTT	ACCTCC AGAGCT CCGAGT TTCCGG CAAACG CCGMCG CCTCAA CACGCC GATCGA CCTGCA GCCGCT GGACGT GGACGC GCTG (SEQ ID NO: 1197)		CIQASHRIWGW HCRRHREGQRS GNTSRASRGSG GTPRRTSQVRA RR (SEQ ID NO: 1442)

					CGTGTCGACCGGGGCGGGGCTGTCGGACGA CGCATACCGGTTCGGCGTGAAGGCCCGCCT CAACCAGGTCGACACGAACGCCGTGCTCAA GCGCAAGCGGCTGCGGTCCAGCAAGACCTG CAGAGACCCGACCTGCTCCTCGGCCGAGAC GCTGGCTCATGCTCTCAATCACTGCGCGTCC AACATGGACGCCATCCGCCAGCGTCATGAC GACGCCTTAGAGCAGATCGGGTCCAAGATC CGCGGCGCGCTTGAGCGAGCCAAGTCGACC ACAGAGCTGCGCCTGAACCAGACGGTGCCT GAGTATACAGGCGCAGCTCTGCGACCCGAC ATCGTCCTGCGGAACGTGGCCGCCAAGAAG ATGGTGATCGCCGACTTGGCAGTCACGTTC GAGGACCACGCGGCCGGCGCTCGTCACTCC TCGCTGCAGCTCAGCCACGATCACAAGACA CTCAAGTACCAGCCAATCGTGGCAGAGCTG CGGGTGCAGGGGTGGCAGGTGCAGACCGCG GCGATTGTGTACGGGTCGCTGGGCTCGGTG CAGCCCAGCAACTTCAAGACGTACACCGAG AAGTTGAAGTTGCACAAGCGCGAGGCTCGA CAGCTGGATCTTCAACTGTCGAGCCACTGC ATTCAAGCCAGCCACCGCATCTGGGGCTGG CATTGCCGGCGCCATCGCGAAGGACAACGA			

					AGCGGCAATACCTCGCGAGCATCGCGTGGG TCTGGGGGGACCCCGCGGCGCACATCGCAG GTAAGGGCACGGCGATAAGCTGGTCATTTG ACCGACAGGGCACTACCCAGGTAGGGAACC GCCCTTCAGAAACCCAGGAAGACACAAACA CCCTCCCTTTAGTGACATACATATTTTAGGC TACATTTCAGTTACGGAGAGGTTACTAACT GGTAAATAAAAAGCACTTT (SEQ ID NO: 1074)			
NeSL	Utopia- 2_PI	.		Phytophthor a infestans	GCTCGGCCTCGCGGCTGCCTTCCCAGGCGC CGCCGACTTCGCGCTCTGGCGCGGCCCACA CGCCGCCGCCGAGCCTCCAAGCGCGCCCGT TGGCTTCCGCAGACGCAGGGCTCGCGGCGA CGGCCCTCGCCAGCCCCCAAGACCCCCCCT ACGATGTGGCACCACCCGGCAGGGCGGCCG GCAGGCTGCCCGACTCGGTATCGCCGGGTG CTACACTCTCAGCCGCCACAGCTCGGGCCTT GGCGGTCCGCCATTGGCCCTTGGAGCTCGA CAGCGACAGCAGCGACGACGAGGACGCTC AAGACCCCCACGCCGCCGCCCCAGGACCCC CCGCAGACGTCGCGAGTGTGCTTGCCCCAC CCGGCAAGGCAGGCAGCATGCTCGCCGACC CAGCAGCGCTGGCAGCCGGTCTGGCTCGCG	GCTCGG CCTCGC GGCTGC CTTCCC AGGCGC CGCCGA CTTCGC GCTCTG GCGCGG CCCACA CGCCGC CGCCGA GCCTCC AAGCGC	TAGG CGGA AACC AGGC CCAA GACG GCCG ACAG GGCC CCAC CCAG GTAG GGAA CCGC	MLADPAALAAG LARAPPPPSAPQ DPSPAFPAGPAG QNPRAAAPARV EVHTVVAPPGR AGGMLPDPGLV DSSPAAATAATP APVAATATTAR AAARVAVEHHA HAEPNQEHLPM ARVLVEPMQVD ECSSCDRSTLTA DDGSGDDVAAP SSMLSNDVAAP

					CTCCGCCGCCCCCATCGGCACCCCAGGACC	GCCCGT	CCTA	MDVDSGTSCPP
					CAAGCCCCGCCTTCCCTGCTGGCCCCGCCG	TGGCTT	GAAA	TLQQPLQRPRAL
					GCCAGAACCCTCGCGCCGCCGCCCCAGCCC	CCGCAG	CCCA	HVGSKRRRLDA
					GTGTTGAAGTCCACACAGTCGTGGCACCAC	ACGCAG	TTTCG	DDGEEAHQLQE
					CCGGCAGGGCGGGCGGCATGCTGCCCGACC	GGCTCG	GTGG	EEEAGIHAPALR
					CGGGGTTGGTTGACTCAAGCCCAGCAGCGG	CGGCGA	TCGA	LSAASAQPASVL
					CTACCGCAGCTACTCCAGCTCCGGTAGCTG	CGGCCC	CTCG	AVYTHNASRFD
					CTACTGCTACAACTGCTCGCGCTGCTGCTCG	TCGCCA	AAGG	CTLCAYTAGSF
					CGTCGCCGTGGAGCACCACGCGCACGCTGA	GCCCCC	CCTT	ASLLTHRNSRHR
					GCCGAACCAAGAACATCTACCGATGGCTCG	AAGACC	ACCT	RTAFLDRFSAGC
					CGTGCTCGTGGAGCCCATGCAGGTCGACGA	CCCCCT	ATTTT	ACGVPFASRLA
					GTGCAGCAGCTGCGACCGCAGCACTTTGAC	ACGATG	TTCCT	AARHAQACASL
					AGCCGACGACGGGAGCGGTGACGACGTTGC	TGGCAC	TAGA	SSAPSAEASSAA
					AGCCCCAAGCTCCATGCTCAGCAACGACGT	CACCCG	CATT	GTSSPTADGADS
					CGCCGCCCCCATGGACGTTGACTCCGGCAC	GCAGGG	CAAT	TVSAVAHAEPG
					CAGTTGCCCTCCAACGTTGCAGCAGCCGCT	CGGCCG	TAGG	LPHHNDTELTAS
					CCAGCGACCGCGTGCGCTGCACGTAGGCAG	GCAGGC	TAGC	PPLVSSSDVEVQ
					CAAGCGCCGCCGACTCGACGCAGATGACGG	TGCCCG	GACC	ATKTEATDNRW
					TGAAGAAGCTCACCAGCTGCAAGAAGAGG	ACTCGG	AAAT	GAPLPRVLVASR
					AGGAGGCTGGCATTCATGCCCCAGCGCTCA	TATCGC	TACA	IAGRLAQVPPPR
					GGCTATCCGCGGCCAGCGCTCAACCGGCCT	CGGGTG	AATT	WGPPLPRTTIAA
					CCGTCCTCGCCGTGTACACGCACAACGCGT	CTACAC	TGGT	RIATRLAATPAP
					CGCGGTTCGACTGCACGCTGTGCGCGTACA	TCTCAG	AACG	RWDPPLPRSLV

					CGGCCGGCAGCTTCGCGTCACTTCTGACGC	CCGCCA	AGTA	VSRIAARLLPAL
					ACAGGAACTCTCGGCACCGCCGCACCGCAT	CAGCTC	AGCC	PDAPACEEEAK
					TCCTCGACCGATTCTCGGCGGGTTGCGCGTG	GGGCCT	AAAT	DSDTMDWAPT
					TGGCGTGCCATTCGCCTCGAGGCTGGCTGC	TGGCGG	GGTA	WTNEETKESEP
					AGCAAGACACGCGCAAGCGTGCGCCAGCCT	TCCGCC	ATAC	HDEAPGQVDEE
					CTCGAGCGCACCATCGGCCGAGGCTTCGTC	ATTGGC	ACAA	TIDDADGEWLL
					GGCAGCAGGAACATCAAGCCCCACTGCCGA	CCTTGG	AACT	RFDGACRANPG
					CGGAGCCGACTCCACCGTCTCAGCCGTGGC	AGCTCG	TTTCT	PGGAGAALFKP
					TCACGCTGAGCCCGGCCTGCCCCACCACAA	ACAGCG	GTTCT	SGPVVWTCSHY
					CGACACGGAGCTCACAGCGTCCCCCCCGCT	ACAGCA	AATC	DPSTTATNNTAE
					GGTGAGCTCCTCCGATGTTGAAGTGCAGGC	GCGACG	AGTG	YTALLLGARAA
					AACAAAGACCGAAGCCACCGACAACCGCTG	ACGAGG	TGAA	ADHGVTKLRIE
					GGGCGCCCCGCTTCCGCGCGTGCTGGTCGC	ACGCTC	AACT	GDSTLVIQQVRG
					GTCTCGAATCGCCGGGCGCCTCGCTCAGGT	AAGACC	GGTT	IFATRSTRLRAL
					GCCACCCCCACGCTGGGGCCCCCCGCTGCC	CCCACG	TTCG	RNKVKLELARV
					GCGCACCACCATCGCAGCCCGCATTGCGAC	CCGCCG	CCTTT	GSFSLHHIDRQA
					GAGGCTCGCAGCCACGCCTGCCCCCCGCTG	CCCCAG	TGGC	NGHADRLANAG
					GGACCCACCGCTGCCACGAAGCCTGGTTGT	GACCCC	GGAC	LDRRRTKLECSV
					GTCACGCATCGCCGCCCGCCTGCTACCCGC	CCGCAG	TTTTT	HPDGRGCTNTS
					GCTACCAGACGCGCCTGCATGCGAGGAAGA	ACGTCG	CACT	VATAAPTAPAA
					GGCCAAGGACAGCGACACGATGGACTGGG	CGAGTG	CGCA	PLPSARPPASTA
					CGCCAACTTGGACAAACGAGGAGACCAAG	TGCTTG	TTTTT	APSPDDDHSDQ
					GAAAGCGAGCCACATGACGAAGCGCCAGG	CCCCAC	GGGC	GDIDDGEVYAA

					GCAGGTCGACGAGGAGACCATCGACGACGC	CCGGCA	AATC	MCISPDAVPHRR
					CGACGGCGAGTGGCTTTTGCGCTTCGACGG	AGGCAG	GTCT	PRLRLRRLTDEE
					TGCTTGCCGTGCAAACCCAGGCCCAGGCGG	GCAGC	GCGG	SEEAGNVVERL
					CGCCGGCGCAGCTCTCTTCAAGCCCAGCGG	(SEQ ID	CTAG	AASLAAKIADAP
					CCCCGTAGTGTGGACTTGTTCCCACTACGAC	NO: 1198)	CTTG	DWETAEGYITA
					CCGAGCACCACGGCAACGAACAACACAGC		CTAG	LPYALYDKLQP
					AGAGTACACTGCGTTGCTGCTCGGGGCCAG		CGGC	YSQSQHQPPRQ
					GGCCGCCGCCGACCACGGCGTCACGAAGCT		GGAC	QQQQQRQRPRQ
					GCGCATCGAGGGCGACAGCACGCTGGTCAT		GAGC	QQQTRQRRQRR
					CCAGCAAGTGCGGGGCATCTTCGCGACGAG		GGTC	CKRGGGSQHRQ
					AAGCACGCGGCTCCGAGCACTGCGCAACAA		TCCG	RKTRRRRPPRVT
					GGTCAAGCTCGAGCTCGCGCGCGTGGGGAG		GGGG	RHHREHRIDEAL
					CTTCTCGCTCCACCACATCGACCGGCAGGC		CGTT	DDLHALESRRP
					GAACGGTCACGCCGACCGCCTTGCCAACGC		CACC	QDRTAISKARRR
					CGGCCTCGACCGACGCCGCACCAAGCTCGA		TTTCC	VGRIRSALDQH
					GTGCAGTGTGCACCCCGATGGCAGGGGCTG		CCCG	QLRHRFDTDEK
					CACCAACACCTCCGTGGCGACCGCAGCGCC		CGAG	ACVDGILAAAR
					TACTGCTCCTGCAGCGCCCTTGCCCTCAGCC		GCCA	DKDRAASVTTT
					CGGCCTCCTGCATCTACGGCCGCTCCGTCTC		ACTA	AQTAAPPHSAP
					CTGATGACGACCACAGCGACCAAGGCGACA		CACC	ASAPSSAVDDGI
					TCGACGACGGCGAGGTGTACGCCGCAATGT		GATC	CPIPGDLLHAFF
					GCATCAGCCCCGATGCTGTGCCGCACCGCC		TTCTC	TDVNTPRTEFDA
					GACCGCGCCTCCGCCTACGGCGCTTGACGG		TACA	DSPIGARFREAL

					ACGAAGAGTCCGAGGAGGCTGGCAACGTCG		CTTTT	AQLPAAIAATEL
					TGGAGAGGTTGGCAGCGTCGCTGGCGGCGA		CTAA	LMEPPSPDEVED
					AGATTGCAGATGCGCCAGACTGGGAGACGG		TTCG	QLQRVRGTSSPG
					CGGAGGGCTACATCACGGCCCTACCGTACG		CCTC	LDGVGYDVYKT
					CCCTCTACGACAAGCTGCAGCCGTACTCCC		CGTC	FTQQLLPALHA
					AGTCCCAGCACCAACCTCCACGTCAACAGC		TTCG	AFSRCWTDQRV
					AGCAGCAGCAGCGGCAACGCCCACGGCAG		GTCTT	PQSWKLGVVRL
					CAGCAGCAAACACGGCAGCGTCGACAGCG		CGGC	LFKKGDRQDPA
					GAGATGCAAGCGCGGCGGTGGCAGCCAAC		TGTC	NWRPICLQQAV
					ACAGGCAGCGGAAGACGCGCCGCCGACGG		GGAT	YKLYAGILAHRF
					CCCCCACGGGTGACACGCCACCACCGCGAG		TTTTT	TRWLDANTRHA
					CACCGGATCGATGAAGCCCTCGACGACCTC		TCTTT	DAQKGFRAVNG
					CACGCGCTCGAGAGCAGGCGCCCTCAGGAC		TTGA	CGEHNFLAATL
					AGGACGGCCATCAGCAAGGCCAGGCGTCGA		CCAA	TDNARRRRREL
					GTCGGTCGCATCCGCTCCGCCCTTGACCAGC		TCAG	HVVWYDIKNAF
					ACCAATTGCGCCACCGCTTCGACACCGACG		AGCG	GSVPHELLWEV
					AGAAGGCCTGCGTGGACGGCATCCTGGCCG		CGCC	LRRMGVPAQFI
					CCGCGCGCGACAAGGACAGGGCGGCATCA		ATGC	ACCQGIYDAAA
					GTGACGACAACAGCGCAAACAGCGGCTCCG		GACT	FTVGNAADGTT
					CCTCATTCTGCACCAGCTTCAGCTCCTTCGT		CTTCT	APIQLRLGVFQG
					CGGCGGTGGACGACGGAATCTGCCCGATTC		GGCC	CPLSPHLFTAVIS
					CGGGAGACTTGCTCCATGCCTTCTTCACAGA		AATC	PLLHALKRLPGT
					CGTGAACACGCCGCGCACGGAGTTCGACGC		AGAG	GVQLSAVDRPG

					CGACTCGCCCATCGGAGCCCGCTTTCGGGA		ACCG	ASAYADDLKVF
					GGCGCTTGCTCAGCTGCCCGCTGCCATCGCT		GGCC	SDTKDGITRQHQ
					GCCACGGAGCTACTCATGGAGCCGCCGTCG		CTGT	LVTDFLRWTGM
					CCGGACGAGGTTGAGGACCAGTTGCAACGT		CCTC	VANPSKCSTMS
					GTACGGGGTACGTCCAGCCCCGGCTTGGAC		GGAC	VQRDNRGVLKT
					GGAGTGGGGTACGACGTCTACAAGACCTTC		AGCG	ANLTLQLDGAQI
					ACACAGCAACTCCTACCGGCCCTGCACGCC		AGGC	PALGMTEAYAY
					GCTTTCTCCAGGTGTTGGACTGATCAGCGCG		CTCC	LGIGDGFDHVR
					TCCCACAGAGCTGGAAGCTCGGAGTCGTGC		ACGG	RRVELAPKLREL
					GGCTCCTGTTCAAGAAAGGAGATCGGCAGG		CCAG	KADTTALMQSG
					ACCCCGCCAACTGGAGGCCGATCTGCCTCC		CCAA	LAPWQVVKALK
					AGCAAGCCGTCTACAAGCTGTACGCCGGTA		TCAA	VYIYPRVEYALR
					TCCTGGCCCACCGGTTCACACGCTGGCTCG		GTCT	HLRPFQQQLQG
					ACGCCAACACCCGGCACGCTGACGCACAGA		CGGC	FDRHLARGLRH
					AGGGATTCCGCGCGGTTAACGGCTGCGGGG		AGCG	LLRLPTSATTEF
					AACACAATTTCCTCGCAGCCACGCTGACGG		ACGC	LYAPTSRGGLGL
					ACAACGCCCGCCGCCGACGCCGCGAGCTGC		GTCTT	LPLTEVHGALQI
					ACGTGGTCTGGTACGACATCAAGAACGCGT		TCTAT	AHAWQTLHSPD
					TCGGCAGTGTGCCCCACGAACTGCTGTGGG		AGCG	PAIRRIARVQLR
					AGGTGCTGAGGCGCATGGGAGTGCCGGCGC		CAGC	QVADARHRLDA
					AGTTCATCGCGTGCTGCCAAGGGATCTACG		TGAC	EHWKERGEELC
					ACGCGGCAGCTTTCACGGTCGGGAACGCAG		GAGG	ERLLNSQLGTSA
					CAGACGGCACGACGGCCCCAATTCAGCTGC		CCGA	HAPPKRRNCDIG

					GCTTGGGCGTGTTTCAGGGATGTCCCTTGAG		TCTG	SLWVDVQRHLR
					CCCCCACCTCTTCACGGCGGTGATCAGCCC		GCGG	SLGLQLQTAPA
					ACTCCTCCACGCACTCAAGCGGCTCCCTGG		CCCC	DTHTGAPAQPL
					CACGGGGGTACAGCTCTCAGCCGTCGACCG		CGAT	QLRVPHHDKWL
					GCCAGGTGCGTCAGCATATGCCGACGACCT		TGGT	THKDVLRHVKL
					CAAGGTCTTCAGTGACACCAAGGATGGCAT		CCGA	HIKNNHWHRWT
					CACGCGGCAGCACCAGCTCGTCACCGACTT		CTTTC	SMRDQGKTARA
					CCTGCGCTGGACAGGCATGGTGGCAAACCC		GGCC	HGGEGSGFLTQP
					GTCCAAGTGCAGCACCATGTCTGTCCAGAG		AATC	RGMWEADYRF
					GGACAACCGCGGCGTTTTGAAGACCGCCAA		AGCG	AVAGRLNQVDT
					CCTCACGCTCCAGCTGGACGGCGCCCAAAT		ACGA	YSVLKRRRLRS
					CCCAGCACTCGGGATGACCGAGGCGTACGC		CGAG	HDRCRQPGCHR
					GTACCTGGGCATAGGCGACGGATTCGACCA		GGGG	AETLAHVLNHC
					CGTACGCCGCCGCGTCGAGTTGGCCCCCAA		CAGG	PGTMDAVRGRH
					GCTCCGCGAACTGAAGGCAGACACGACGGC		GGTT	DGALKRIERELH
					GCTGATGCAGTCGGGGCTGGCGCCGTGGCA		TACA	ASATDRRDRVE
					AGTTGTGAAGGCCCTCAAGGTTTACATCTA		CTTTT	LRVNQTVPSLA
					CCCACGCGTGGAATATGCCCTCCGCCACCT		GCCC	GPALRPDLQLY
					CCGACCGTTTCAGCAGCAGCTCCAAGGCTT		CCGT	NHTKKTVAVVD
					TGACCGACACCTCGCGCGCGGGCTGCGTCA		TTCG	LAVAFEEQASD
					CCTTCTCCGCCTGCCAACATCAGCCACCACC		ACTT	DASSSALSLIAS
					GAGTTCCTGTATGCACCAACGTCGCGAGGA		CAAC	HKRAKYDRIKR
					GGTTTGGGGCTGCTACCGCTTACAGAGGTT		TTCA	HLERQGWKVHL

					CACGGTGCCCTTCAGATTGCGCACGCCTGG		GGCC	SALVYGSLGAV
					CAGACCCTGCACTCGCCGGACCCCGCCATC		AAAA	ASGNYQVYTTH
					CGCCGCATTGCCCGCGTGCAGCTTCGCCAA		TGGC	LGLLKRDAKRL
					GTCGCCGACGCCCGGCACCGCTTGGACGCC		GATT	DRQLSVECIQSS
					GAGCACTGGAAGGAGCGGGGGGAGGAGCT		TCGA	RRIWNLHCSQH
					GTGCGAGCGGCTGCTAAACTCGCAGCTCGG		CCCT	RTRQHQARPSQ
					GACATCGGCCCACGCGCCTCCCAAACGACG		CCAC	GPRGSRATETG
					CAATTGCGACATCGGCTCGCTGTGGGTCGA		GCGC	GTPSQTSRR
					TGTCCAGCGCCACCTCCGCTCTCTCGGCCTA		CGTG	(SEQ ID NO:
					CAGCTACAGACAGCCCCCGCGGACACGCAC		CCAC	1443)
					ACTGGCGCACCAGCTCAGCCTCTCCAGCTTC		TGCT	
					GCGTGCCGCACCACGACAAGTGGCTGACGC		CGGC	
					ACAAGGACGTGCTGCGGCACGTGAAGCTGC		ACCG	
					ACATCAAGAATAACCACTGGCACCGCTGGA		GCGG	
					CATCCATGCGGGACCAAGGGAAGACGGCCC		CGAT	
					GCGCCCACGGCGGTGAAGGCAGCGGCTTCC		TCAG	
					TTACGCAGCCGCGAGGCATGTGGGAAGCCG		CGGG	
					ACTACCGGTTCGCGGTGGCGGGTCGTCTTA		TGCA	
					ACCAAGTGGACACGTACAGCGTACTGAAGC		ACTT	
					GCCGGCGCCTGCGCTCGCACGACAGATGCC		CGGG	
					GCCAGCCAGGATGCCATCGTGCCGAAACAC		CACG	
					TGGCTCACGTGCTAAACCACTGTCCCGGCA		TGTG	
					CGATGGACGCGGTCCGAGGGCGGCACGACG		CAAC	

					GAGCGCTCAAGCGCATCGAGCGTGAGCTCC		ACAT	
					ACGCGTCGGCCACGGACCGCCGGGACCGCG		GCAG	
					TCGAGCTCCGAGTGAACCAGACCGTGCCGT		CGCC	
					CACTCGCCGGCCCCGCGTTACGGCCCGATC		CATT	
					TCCAGCTCTACAACCACACCAAGAAGACGG		GCAC	
					TGGCGGTGGTCGACCTGGCCGTGGCGTTTG		GCCA	
					AGGAGCAGGCGAGTGACGACGCGAGCAGC		AGCG	
					TCGGCGCTGTCCCTTATCGCCAGCCACAAG		GCAT	
					CGAGCCAAGTACGACCGCATCAAGCGACAC		CGCG	
					CTCGAGCGCCAAGGATGGAAGGTACACCTC		GGAC	
					TCGGCGCTCGTGTACGGGTCGCTTGGGGCG		GACG	
					GTCGCTAGTGGCAACTACCAGGTGTACACC		CCTC	
					ACACACCTGGGGCTACTCAAGCGCGATGCA		GGCC	
					AAGCGGCTGGACCGGCAGCTGTCTGTCGAA		GCTC	
					TGCATCCAGTCCAGCCGCCGCATCTGGAAT		AAGC	
					CTACACTGCAGCCAGCACCGGACTCGCCAA		GCAG	
					CACCAGGCGAGGCCCAGCCAAGGACCAAG		CCCC	
					AGGCAGCCGGGCGACGGAGACCGGGGGGA		GCCC	
					CTCCGTCCCAGACAAGCCGCCGCTAGGCGG		TTCC	
					AAACCAGGCCCAAGACGGCCGACAGGGCC		AGCA	
					CCACCCAGGTAGGGAACCGCCCTAGAAACC		CGAC	
					CATTTCGGTGGTCGACTCGAAGGCCTTACCT		CTCG	
					ATTTTTTCCTTAGACATTCAATTAGGTAGCG		CGCC	

					ACCAAATTACAAATTTGGTAACGAGTAAGC		GTTT	
					CAAATGGTAATACACAAAACTTTTCTGTTCT		GGCG	
					AATCAGTGTGAAAACTGGTTTTCGCCTTTTG		GATC	
					GCGGACTTTTTCACTCGCATTTTTGGGCAAT		GCCA	
					CGTCTGCGGCTAGCTTGCTAGCGGCGGACG		TCAA	
					AGCGGTCTCCGGGGGCGTTCACCTTTCCCCC		GACG	
					GCGAGGCCAACTACACCGATCTTCTCTACA		TGCG	
					CTTTTCTAATTCGCCTCCGTCTTCGGTCTTCG		AGAG	
					GCTGTCGGATTTTTTTCTTTTTGACCAATCA		CCAG	
					GAGCGCGCCATGCGACTCTTCTGGCCAATC		GCGG	
					AGAGACCGGGCCCTGTCCTCGGACAGCGAG		GGTC	
					GCCTCCACGGCCAGCCAATCAAGTCTCGGC		GGGC	
					AGCGACGCGTCTTTCTATAGCGCAGCTGAC		AAAA	
					GAGGCCGATCTGGCGGCCCCCGATTGGTCC		TATA	
					GACTTTCGGCCAATCAGCGACGACGAGGGG		CTTA	
					GCAGGGGTTTACACTTTTGCCCCCGTTTCGA		CTCT	
					CTTCAACTTCAGGCCAAAATGGCGATTTCG		AAGT	
					ACCCTCCACGCGCCGTGCCACTGCTCGGCA		ATGC	
					CCGGCGGCGATTCAGCGGGTGCAACTTCGG		CCGA	
					GCACGTGTGCAACACATGCAGCGCCCATTG		ATCC	
					CACGCCAAGCGGCATCGCGGGACGACGCCT		CTGC	
					CGGCCGCTCAAGCGCAGCCCCGCCCTTCCA		CCTCT	
					GCACGACCTCGCGCCGTTTGGCGGATCGCC		CAGG	

					ATCAAGACGTGCGAGAGCCAGGCGGGGTCG GGCAAAATATACTTACTCTAAGTATGCCCG AATCCCTGCCCTCTCAGGCTGAACGCGGCC CCATACTTGATCTAAGTATGGGAGGGATCC CTGGCCTCTCAGGCTGTACGCGAGACCCGT ACGGCCGAATCCCCTGGCCTCTCAGCCTGT ACGCGGGGC (SEQ ID NO: 1075)		CTGA ACGC GGCC CCAT ACTT GATC TAAG TATG GGAG GGAT CCCT GGCC TCTC AGGC TGTA CGCG AGAC CCGT ACGG CCGA ATCC CCTG GCCT	

							CTCA GCCT GTAC GCGG GGC (SEQ ID NO: 1321)	
NeSL	Utopia-2_PR	.		Phytophthora ramorum	TCAAGCCCCGCCGCCAAGCCCAGCTGCGGC TGTTGCCGCCCCTCCAGCAGCAGCAGTCGC AGCTGAACCTACAGCCCCTCCTGCTGATCG CGCCGCCGTGGAGCCCCGCGCGCGCCGA GCCGCCCCAAGAACAAGCACCCCCAGCTAG CGCGCGCGTGGAGCCCATGTCGGGCGACGT GGTCAGCAGCGACGGCAGCAGCAGGACGA CGGACGCGAGCGGCGACGGCGACGACGGC GCAGGCAGCAGCGATGCAGCAGGCGACGTC GGCGTCGTCGCCATGGACGTTGACCAAGGT GCCCGCCGACAGCAGCCGCCATGGCAGCGC GTCGGCGGCAAGCGTCGACGTCTCAACGAC GTCGACGACGAAGACACCCGCGAGCTGGCC GAGCTCCTGCTCGAGGAGGAGGACGAGGCT GGGGACCACGCTCCAGCGCCTAGGCTATCC	TCAAGC CCCGCC TTGG GTTG CTCA AGCG TGAT GCGA AGCG GCTG GACC GGCA GCTC TCGG TGGC	TGAG CACC DGAGSSDAAGD VGVVAMDVDQ GARRQQPPWQR VGGKRRRLNDV DDEDTRELAELL LEEEDEAGDHA PAPRLSAASARP ASVLSVYAHNA QRFQCTLCTYT AASFASLKRHR DSRHRRTAFLD RFSAGCACGAPF ASRLAAANHAH	MSGDVVSSDGS SRTTDASGDGD DGAGSSDAAGD VGVVAMDVDQ GARRQQPPWQR VGGKRRRLNDV DDEDTRELAELL LEEEDEAGDHA PAPRLSAASARP ASVLSVYAHNA QRFQCTLCTYT AASFASLKRHR DSRHRRTAFLD RFSAGCACGAPF ASRLAAANHAH

					GCGGCCAGCGCTCGCCCGGCCTCCGTCCTCT	CGCCGC	GTGC	ACASLNRTLSV
					CCGTGTACGCGCACAATGCGCAGCGCTTCC	CGTGGA	ATCC	AATPAAGELSPT
					AGTGCACGCTGTGCACGTACACGGCCGCCA	GCCCCG	AGTC	AGAANATVKAA
					GCTTCGCGTCGCTCAAGCGACACCGGGACT	CGCGCG	CAGC	TVTPDSPRQDPP
					CCCGGCACCGCCGCACCGCCTTCCTCGACC	CGCCGA	CGCC	ELAASPPLASSP
					GGTTCTCGGCGGGATGTGCGTGCGGTGCGC	GCCGCC	GCAT	DVAVQAADMQ
					CCTTCGCGTCGAGGCTGGCTGCTGCGAACC	CCAAGA	CTGG	APTRWDPPLPRT
					ACGCACACGCGTGCGCCAGCCTCAACCGCA	ACAAGC	AACC	LVATRVASRLT
					CCTTATCGGTAGCGGCTACGCCGGCAGCAG	ACCCCC	TGCA	DLTPPRWGPPLP
					GGGAACTGAGCCCCACTGCTGGTGCAGCCA	AGCTAG	CTGC	RATVVSRIAARL
					ACGCCACCGTCAAGGCGGCCACCGTCACGC	CGCGCG	AGCC	EAAPTPRWGPPL
					CCGACTCCCCCCGCCAAGACCCCCCGGAGC	CGTGGA	AGCA	PRVVVASRIAER
					TCGCAGCGTCCCCCCCGCTAGCGAGCTCCC	GCCC	CCGC	LAPPELAADDET
					CCGACGTTGCTGTGCAGGCCGCCGACATGC	(SEQ ID	GCGC	KDGEEDQSFTEP
					AGGCCCCGACTCGTTGGGACCCGCCCCTCC	NO: 1199)	GCCA	VAAARSXGGED
					CCCGCACGCTGGTGGCGACACGGGTCGCCT		GCAC	ANGEWLLRFDG
					CCCGCCTCACCGACCTGACACCGCCGCGGT		CAAG	ACRANPGPGGA
					GGGGACCGCCGCTGCCTCGAGCCACGGTGG		CACC	GAALFKPSGPV
					TGTCTCGCATCGCGGCGAGACTCGAGGCCG		AGGG	VWTCSHYMPSS
					CACCAACGCCCCGTTGGGGCCCTCCACTTCC		GGAA	SETNNTAEYTAL
					ACGCGTCGTGGTTGCGTCCMGGATCGCCGA		GTCG	LLGMRAAADHG
					ACGCTTAGCTCCGCCCGAGCTGGCAGCTGA		GGCG	ATRVHVEGDST
					CGACGAGACCAAGGACGGCGAGGAAGACC		GCGG	LVIQQVRGIFAT

					AGAGCTTTACGGAACCTGTGGCTGCCGCCC		AGAC	RSTRLRGLRKSV
					GGTCMGAKGGCGGCGAGGACGCGAACGGC		CGGG	KAEMARMEHV
					GAGTGGCTGCTTCGCTTCGACGGCGCCTGC		GGGA	TLHHIDRQANG
					CGAGCCAACCCAGGCCCCGGCGGTGCGGGC		CTCC	HADRLANAALD
					GCAGCGCTGTTCAAGCCCAGCGGCCCGGTG		GCCG	RRKTKLECGLH
					GTGTGGACCTGCTCCCACTATATGCCGAGC		CGCA	PDGQGCSSTAA
					AGCAGCGAGACCAACAACACGGCCGAGTA		CCGG	TTAVPSVVPDRP
					CACGGCTCTGCTGCTCGGCATGCGGGCTGC		CCGC	PSSTAAAPTPSA
					GGCGGACCATGGCGCTACGCGCGTGCACGT		CGCT	EPDETEQGDIDD
					CGAAGGTGACAGCACCCTGGTGATCCAGCA		AGAC	GEVYAAMCIGP
					AGTGCGGGGCATCTTCGCTACCCGTAGCAC		GGCA	DSIPERRPRLRL
					CCGGCTCCGTGGACTGCGCAAGAGCGTGAA		CACA	RQLSETEEEEAG
					GGCGGAGATGGCGCGGATGGAGCATGTCAC		GGCC	AIVERLAATLAG
					GCTCCACCACATCGACCGCCAAGCCAACGG		CACA	KITDASDWATA
					CCATGCCGACCGCCTCGCCAACGCCGCCCT		GCGG	EGYITALPYTLY
					CGACCGCCGCAAGACCAAGCTGGAGTGCGG		CCGA	DKLQPFAQHRH
					ACTCCACCCTGACGGCCAAGGCTGCTCCAG		CAGG	QPRPQHRQQPQ
					CACCGCCGCGACGACGGCGGTCCCCTCTGT		GCCA	RDPPLGTHDGD
					GGTGCCGGACCGGCCTCCCTCGTCAACCGC		CACC	HGQPSTSRSRRR
					TGCCGCGCCAACGCCAAGCGCCGAACCCGA		CAGG	RRRAKDRLRRR
					CGAGACCGAGCAGGGCGACATCGACGACG		TAGG	PPRITRHHREHR
					GCGAGGTCTACGCCGCCATGTGCATTGGCC		GAAC	LDEALDDLRAV
					CGGACTCCATACCGGAGCGCCGACCACGCC		CGCC	EHASPHDRPAV

					TCCGCCTACGGCAGCTGTCCGAGACCGAGG		CTCA	ARARRRVGRVN
					AGGAGGAGGCGGGTGCCATTGTGGAGAGG		AACC	SAIAQQQLRHKF
					CTGGCGGCAACGTTGGCGGGCAAGATCACG		CCGC	DKDEKACVDGI
					GACGCGAGTGACTGGGCGACGGCGGAGGG		CGGT	LAAARASRGLA
					CTACATCACTGCCCTCCCCTACACTCTATAC		ACAT	TPSASASRHPPP
					GACAAACTGCAGCCGTTCGCCCAGCACCGC		TATG	VPSTAADDGSCP
					CACCAGCCTCGCCCACAGCACCGGCAGCAA		GTCC	IPSDELHAFFTA
					CCCCAGCGCGACCCACCCCTCGGCACCCAC		GACA	VNTPAGTFEPM
					GACGGGGACCACGGCCAGCCGAGCACCAG		CCTA	APVGAPFRSAV
					CCGCAGCCGTCGCCGTCGCCGCCGCGCCAA		TGAG	AHLPAATSQPEL
					GGATCGCTTACGCCGACGCCCACCGCGTAT		GTGC	LSDAPTTDDIED
					TACCCGCCACCACCGTGAGCACCGCCTCGA		AACC	QLQRARGSSSPG
					CGAGGCTCTGGACGACCTGCGGGCGGTGGA		TGTA	LDGVGYDIYKA
					GCACGCCAGTCCCCACGACCGGCCGGCCGT		CACA	FAAQLLPALHA
					CGCTCGGGCCCGGCGGCGCGTTGGCCGCGT		AGTT	AFACCWRHKQV
					CAACTCCGCCATTGCCCAGCAACAGCTGCG		ACAC	PQSWKVGVVRL
					ACACAAGTTTGACAAGGACGAGAAGGCGTG		ACCA	LFKKGERTEPAN
					CGTGGATGGAATCTTGGCCGCGGCCCGAGC		CATA	WRPICLQQAIYK
					TTCTCGCGGGTTAGCAACCCCATCGGCCTCA		GCGA	LYAGVLARRLT
					GCCTCGAGGCATCCACCGCCGGTACCCTCC		CTAC	RWLDANGRHA
					ACCGCTGCTGATGACGGCTCCTGCCCCATCC		CAGG	DTQKGFRAMNG
					CCAGCGATGAGCTCCACGCCTTCTTCACGG		TATTT	CGEHNFLAATL
					CCGTGAACACGCCGGCGGGCACCTTCGAGC		ACTA	VDQARRKRREL

					CCATGGCACCAGTGGGTGCCCCGTTCCGCT		CCTG	HVVWYDFANAF
					CCGCCGTCGCCCACCTCCCCGCCGCGACAT		GAAG	GSVPHDLLWEA
					CGCAGCCTGAGTTGCTCTCGGACGCGCCGA		CCAA	LERQGVPSPFIA
					CCACGGACGACATCGAGGACCAACTGCAGC		GGAT	CCRGLYADAVF
					GTGCCCGTGGCAGCTCCAGCCCTGGCCTGG		TAAC	TVGNAADGTTA
					ACGGCGTGGGGTACGACATTTACAAGGCGT		CGGT	PIALRVGVFQGC
					TCGCGGCCCAGCTCCTGCCGGCCCTGCACG		CGGT	PLSPHLFTAAIAP
					CAGCATTCGCGTGCTGCTGGCGCCACAAGC		AATA	LLHALKRLPDT
					AAGTCCCGCAGAGCTGGAAGGTGGGGGTGG		CACA	GVQLSRVDCLG
					TGCGCCTCCTGTTCAAGAAGGGGGAGCGAA		TAAC	ASAYADDLKIFS
					CCGAGCCGGCGAACTGGCGCCCGATCTGCC		TTT	GTEGGTKRQHA
					TGCAGCAGGCAATCTACAAGCTGTACGCCG		(SEQ	LVADFLRWTGM
					GGGTGCTAGCTCGCCGACTCACCCGCTGGC		ID NO:	RANPAKCCTMS
					TGGACGCCAACGGCCGCCACGCAGACACGC		1322)	VQRDTRGVLKA
					AGAAGGGATTCCGGGCGATGAACGGCTGCG			CNLGLQLDGAPI
					GGGAACACAATTTCCTCGCAGCCACGCTGG			PALTMSASYAY
					TCGACCAGGCCCGCCGCAAGCGACGAGAGC			LGIGDGFDHVR
					TCCACGTCGTGTGGTACGACTTCGCCAACG			RRIELAPKLQEL
					CCTTCGGCAGCGTGCCACATGACCTGCTGT			KHDATALLQSG
					GGGAGGCGCTCGAGCGGCAGGGCGTCCCCA			LAPWQVVKAV
					GCCCCTTCATTGCTTGCTGCCGGGGCCTCTA			KVYLYPRVEYA
					CGCGGATGCCGTCTTCACAGTCGGCAACGC			LRHLRPFHQQLE
					AGCGGACGGCACCACGGCACCGATCGCGCT			GFDRHLVRGLR

					ACGGGTGGGGGTCTTCCAGGGATGTCCGCT			HLLRLPANATT
					GAGCCCCCACCTCTTCACGGCCGCGATCGC			AFFYAPVSRGGL
					CCCGCTGCTCCACGCGCTCAAGCGCTTACC			GLLPLTELHAAL
					AGACACGGGCGTACAGCTGTCTCGTGTCGA			QVAHGWQMLN
					CTGCCTCGGTGCCTCCGCGTACGCGGACGA			SKDPAIRRIARV
					CCTCAAGATATTCAGCGGCACCGAGGGCGG			QLRQIADARHRI
					GACCAAGAGACAGCACGCGCTCGTGGCGGA			DAQAWQDREEE
					CTTCCTGCGCTGGACCGGCATGCGCGCGAA			LAQLLLNSQLG
					TCCGGCGAAGTGCTGCACCATGTCGGTCCA			ASTGAPPKRRN
					GCGCGACACACGTGGAGTGCTCAAGGCCTG			GDIGSLWVDVQ
					CAACCTCGGACTGCAGCTGGATGGCGCCCC			RHLRHLSLKLET
					GATCCCCGCGCTGACCATGAGCGCCTCGTA			APACAETGTAA
					CGCCTACCTGGGGATCGGGGATGGCTTCGA			AMLQLRVPHHD
					CCATGTGCGCCGCCGCATCGAGCTCGCGCC			KWLDHKTVLRH
					CAAACTCCAGGAGCTCAAGCACGATGCGAC			VKLHYKNKHW
					GGCGCTGCTGCAGTCCGGACTGGCCCCATG			ARWAAMXDQG
					GCAGGTAGTGAAGGCGGTGAAGGTGTACCT			KTARTHGGAGS
					CTACCCTCGGGTGGAGTACGCCCTTCGCCA			GFLTRPRGMWE
					CCTTCGCCCCTTCCACCAGCAGCTGGAGGG			ADYRFAVAARL
					CTTCGACAGACACCTCGTGCGCGGCCTCCG			NQLDTHSVLKR
					CCACCTCCTCCGCCTGCCCGCCAACGCCAC			RRLRXHDRCRQ
					GACGGCATTCTTCTACGCGCCGGTGTCGCG			PGCTQGGDAGA
					CGGCGGGCTGGGCCTGCTGCCCCTGACGGA			RAQPLRRHHGR

					GCTGCACGCGGCGCTACAAGTCGCGCACGG ATGGCAGATGCTGAACTCGAAGGACCCTGC CATCCGACGCATCGCGCGCGTGCAGCTCCG GCAAATCGCCGACGCGCGCCACCGGATCGA CGCTCAAGCGTGGCAAGACCGGGAGGAAG AGCTGGCGCAGCTCCTCCTGAACTCGCAGC TCGGGGCGTCCACCGGCGCACCCCCGAAGC GCCGCAACGGCGACATCGGCTCGCTCTGGG TGGACGTCCAGCGCCACCTCCGCCATCTCA GCCTCAAGCTGGAGACCGCACCGGCGTGTG CGGAGACCGGCACCGCCGCGGCCATGCTGC AGCTTCGCGTGCCGCACCACGACAAGTGGC TAGACCACAAGACTGTGCTGCGGCACGTGA AGCTGCACTACAAGAACAAGCACTGGGCCA GGTGGGCGGCCATGMGSGACCAAGGCAAG ACGGCACGCACGCATGGGGGCGCGGGGAG CGGCTTCCTCACCCGGCCACGAGGCATGTG GGAAGCCGACTACCGCTTCGCGGTGGCCGC CCGCCTGAACCAGCTCGACACGCACAGCGT SCTGAAGCGCCGGCGCCTCCGGSMSCACGA CCGCTGTCGACAGCCGGGCTGCACCCAGGG SGGAGACGCTGGCGCACGTGCTCAACCACT GCGCCGGCACCATGGACGCGGTCCGCGGCC			GPRPPRRRPQAH RARAARVVAGR PGPRRAPGQPD GAVARRPRATA RPPAVQPHQED GGGGRPGRGVR GAGERRPGELG AGTHRRTQARE VCRRQATPRAP RVEGPPLGARV RLARRGAGRQP QGAY (SEQ ID NO: 1444)

					GCCACGACGACGCCCTCAAGCACATCGAGC GCGCGCTGCACGCGTCGTCGCCGGGCGGCC AGGACCGCGTCGAGCTCCGGGTCAACCAGA CGGTGCCGTCGCTCGCCGGCCCCGCGCTAC GGCCCGACCTCCAGCTGTACAACCACACCA AGAAGACGGTGGCGGTGGTCGACCTGGCCG TGGCGTTCGAGGAGCAGGCGAGCGACGACC CGGAGAGCTCGGCGCTGGCACGCATCGCCG CACACAAGCGCGCGAAGTATGCCGGCGTCA AGCGACACCTCGAGCGCCAAGGGTGGAAG GTCCACCTCTCGGCGCTCGTGTACGGCTCGC TCGGCGCGGTGCCGGCCGGCAACCACAAGG TGCTTACTGAGCACCTTGGGTTGCTCAAGCG TGATGCGAAGCGGCTGGACCGGCAGCTCTC GGTGGCGTGCATCCAGTCCAGCCGCCGCAT CTGGAACCTGCACTGCAGCCAGCACCGCGC GCGCCAGCACCAAGCACCAGGGGGAAGTC GGGCGGCGGAGACCGGGGGGACTCCGCCG CGCACCGGCCGCCGCTAGACGGCACACAGG CCCACAGCGGCCGACAGGGCCACACCCAGG TAGGGAACCGCCCTCAAACCCCGCCGGTAC ATTATGGTCCGACACCTATGAGGTGCAACC TGTACACAAGTTACACACCACATAGCGACT			

					ACCAGGTATTTACTACCTGGAAGCCAAGGA TTAACCGGTCGGTAATACACATAACTTT (SEQ ID NO: 1076)			
NeSL	Utopia-2_PU	.		Pythium ultimum	TGACTGGTGTTTGATCACGATCAATGAGGT	TGACTG	TAAG	MDYDDSEFFDAI
					GATTAACATGAGCCGGAGCAGGCCCCTTAC	GTGTTT	CGGG	CIPDEDADVLDD
					ACGCTGGTGCTGTAATGGTTCAGGAATGCT	GATCAC	GGGT	GDEGDEGGNDD
					CTTATGAGTAACTCCACAGTATAATTTTTGT	GATCAA	CCAG	ESSEPLPLAITNA
					TAGCGGGGGTGAGCGCGTGCGCCGCCCCCA	TGAGGT	ACCC	PSAPLHATMLC
					CCTTTCTCTTTTCGTTTGTATTTGGTTCCTCT	GATTAA	CACA	GTVTQPWLLRF
					GGAGCGACTTCCGTGGCTTTTGTCGCGCGCT	CATGAG	AGAG	DGACRRNPGPA
					ATCGCGCCCGGCATGAGGATCGTCGCTCCT	CCGGAG	AGAA	GAGATLTRPNGI
					GACGCCATGTGCGCCAGTATATCGGAGCCC	CAGGCC	GCAG	ILWTHYRYIPDK
					CGTGCTCCGCTTGGGTGTGTGTTCCGCCACG	CCTTAC	GAAT	TATNNVAEYEA
					ACGCCACGCGATTTGCGTGCACGGCATGCG	ACGCTG	CATG	LLDGLRCAAHH
					CTGGCGACTTCCGCTCAGTGCGCGAACTCG	GTGCTG	GTCC	GVKHLRIEGDSN
					CGGTGCATCGTCGACGTGCGCACCGGTCTC	TAATGG	GCAT	LVIEQVKGIFAC
					TGGCCTTCCAGGACTGGTATGAGTCGTCGT	TTCAGG	GGAC	STSLRPRRDQVR
					GCGCGTGCAGAGCGCTCTTCACCGCCCGTC	AATGCT	CAGT	EILRHFETYSFR
					TCGACGCGTTCGTTCATTCGAGCCGTTGCGA	CTTATG	AGGG	HIDRALNRQAD
					CCACAATGCGAACCAATCAGCGAACGCCGT	AGTAAC	CACG	RLANQALDLLK
					CCCCTCGTCCGCCGCCTCGACAACGAGAGC	TCCACA	CTCC	TVSVCALSQTR
					GGACGCGGATGCGACGGCCCCCCCTCCCGC	GTATAA	ACAA	VQDDTGAAHGC
					TATCTCTCTTCCGCGGACGGGCTTCCCCCGC	TTTTTGT	AGGT	WHWTPPDASPT

					CCCTGTCCTAGTTGCGCCGCTCTCCTTCTGG	TAGCGG	TATC	DDASTSILTQDV
					ATGCGGCAACTTGGACGACCCACGTGCACG	GGGTGA	GCCC	PVPMDIDDYDP
					CGTGCGCTGCCGCCAGTGCGACGGCGTTGA	GCGCGT	TCAA	DEPMNAADDPV
					TCCAGCGTCATCTACACGCCGTTCCTCCCGT	GCGCCG	ACCC	SINAEREGGTVY
					ATCGACAGCAACCGATCTCAACGGCGATGA	CCCCCA	ATCA	PVLRLGPNVVPE
					TAGCCAACCTCCAGACCCGCGTGCCAACAT	CCTTTC	CACG	RQKRLQIPWLPP
					CCAACAGCTTCCAGCCATCCATTCGACAAC	TCTTTTC	AAGG	REMQKLEKKIE
					GACAACGACAACGACAACGACGACGACCC	GTTTGT	ATCT	VLGETFASRIRD
					GCTGCTCGACGACGTGCTCATGGACTACGA	ATTTGG	AAAA	APDWFSAEGYIT
					CGACTCGGAGTTTTTCGACGCCATCTGCATT	TTCCTC	AGAA	ALTSELATLIRQ
					CCAGACGAAGACGCTGACGTGCTCGACGAT	TGGAGC	AGCA	STAATTGPNAA
					GGTGACGAAGGTGACGAAGGTGGTAACGA	GACTTC	ACAA	RPCERTISKEKR
					CGACGAGTCGTCCGAGCCACTGCCTCTAGC	CGTGGC	TCGA	RARRTTPLQRAL
					GATCACGAACGCTCCAAGCGCTCCCTTGCA	TTTTGT	AATA	AEAKHELQIIQP
					CGCGACCATGCTCTGCGGCACAGTTACACA	CGCGCG	GTAA	DASRKSVRKAA
					GCCATGGCTCCTGCGCTTCGACGGCGCGTG	CTATCG	ATAG	RRVKRISQAQQ
					CCGACGCAACCCAGGACCCGCCGGCGCCGG	CGCCCG	TAAA	RHDLRRLFSTNE
					CGCCACGCTCACGCGGCCCAATGGTATCAT	GCATGA	TAGC	RRCVEKILRDPP
					CCTATGGACGCACTACAGGTATATCCCGGA	GGATCG	TTAG	VGPSSTSSSLPA
					CAAGACGGCGACCAACAACGTTGCCGAGTA	TCGCTC	AAAG	TDDDRCTIDPAD
					CGAAGCGCTGCTTGACGGGCTGCGCTGCGC	CTGACG	GTCA	LFAYFQTQATAP
					GGCGCACCACGGCGTCAAACATCTGCGTAT	CCATGT	ACAT	TNFDFDDEGGE
					TGAAGGTGACTCCAACCTCGTCATCGAGCA	GCGCCA	GCGA	LFRSVLDELPRA

					AGTCAAAGGCATCTTCGCGTGCTCCACATC	GTATAT	AATG	DQEVHLLEDEIT
					GCTGCGCCCGCGCCGCGACCAAGTCCGCGA	CGGAGC	CATG	RDEIEDQLSRISK
					GATCCTGCGCCACTTCGAGACCTACAGCTTC	CCCGTG	AGGA	STAPGLDGITNA
					CGCCACATCGACCGCGCACTTAATCGACAA	CTCCGC	ACGT	VYVRFKLQLLD
					GCCGACCGCCTGGCCAACCAGGCACTCGAT	TTGGGT	AAGA	ALQAAFNACWR
					CTCCTCAAGACGGTCTCCGTGTGCGCGCTCT	GTGTGT	TGGT	YNRVPSMWKA
					CTCAAACTCGCGTCCAAGACGATACCGGTG	TCCGCC	AATA	AFVRLIYKKGN
					CCGCACATGGGTGCTGGCACTGGACGCCGC	ACGACG	CATTT	RAVPSNWRPICL
					CAGATGCGTCGCCAACCGACGATGCATCCA	CCACGC	TTAT	QQTVYKLYTAIL
					CGTCGATACTCACGCAGGACGTTCCAGTCC	GATTTG	A (SEQ	ASRLQRWMDA
					CCATGGATATCGATGATTACGACCCCGACG	CGTGCA	ID NO:	NARFTMSQKGF
					AGCCCATGAATGCGGCCGATGATCCAGTCA	CGGCAT	1323)	RAFNGCHEHNF
					GCATTAATGCTGAACGAGAAGGTGGTACTG	GCGCTG		VATCLHDQTRR
					TCTACCCAGTGCTTCGCCTCGGTCCCAACGT	GCGACT		LRKKLAIVWYD
					CGTCCCCGAACGACAAAAGAGGTTGCAAAT	TCCGCT		LRNAFGSLPHEY
					CCCGTGGCTCCCGCCGCGTGAGATGCAAAA	CAGTGC		LWRVLARLGMP
					ACTCGAAAAGAAGATCGAAGTGCTGGGGG	GCGAAC		PQFVARVRQLY
					AGACTTTTGCGTCGCGCATTCGTGACGCGCC	TCGCGG		ADASFTVESRD
					CGATTGGTTCTCTGCAGAAGGCTACATCAC	TGCATC		GTTDPVQLERG
					GGCTCTAACGTCCGAGCTTGCAACGCTCATT	GTCGAC		VYQGCPLSPYLF
					CGCCAATCAACGGCGGCGACGACGGGACCG	GTGCGC		IAALIPLVRALH
					AATGCAGCACGACCCTGCGAGCGCACCATC	ACCGGT		KLKDQHGIVLA
					AGCAAGGAAAAGCGGCGCGCGCGAAGAAC	CTCTGG		PGVTDCVSAYA

					GACACCGTTACAACGCGCGCTCGCGGAGGC	CCTTCC		DDIKIFARSGTG
					AAAGCACGAGCTCCAGATCATTCAGCCTGA	AGGACT		AKALHEIVVRFL
					TGCCTCGCGCAAGAGCGTGAGGAAGGCTGC	GGTATG		SWTNMAANPA
					ACGACGAGTCAAACGGATCTCCCAAGCGCA	AGTCGT		KCALMVTDGAR
					GCAGCGCCACGACCTGCGTCGCCTCTTCTCG	CGTGCG		GGDDTDASMTL
					ACCAACGAGCGTCGTTGTGTCGAGAAAATC	CGTGCA		SIEGETIPRLTGK
					CTTCGTGATCCCCCAGTGGGGCCATCGAGT	GAGCGC		EGYVYLGVEDG
					ACCAGCTCCAGCTTGCCAGCCACCGACGAT	TCTTCA		LAHERRATCLR
					GACCGCTGTACCATTGACCCAGCCGACCTG	CCGCCC		DSLKAASADVV
					TTCGCCTACTTCCAGACGCAAGCCACGGCC	GTCTCG		RLLRSDLAPWQI
					CCAACGAACTTCGACTTTGACGACGAAGGC	ACGCGT		VRAIKSHVLSRF
					GGCGAACTGTTCCGCTCGGTGCTCGACGAA	TCGTTC		DYVLRHLRPFLS
					CTGCCGCGCGCCGACCAAGAGGTTCACCTT	ATTCGA		LFDGFDKMLVR
					CTTGAGGACGAGATCACTCGGGACGAGATC	GCCGTT		GIKRLCQLPQTA
					GAAGACCAGCTCTCGCGCATCTCGAAATCG	GCGACC		TSEFLFSPTSAG
					ACGGCACCGGGACTCGACGGCATCACGAAC	ACAATG		GLGFLPLKELFA
					GCTGTGTACGTGCGCTTCAAGCTACAGCTG	CGAACC		ALQIVHALQML
					CTTGATGCGCTACAAGCTGCATTCAATGCAT	AATCAG		HSKDANVRAIA
					GTTGGCGATACAATCGTGTTCCCAGCATGT	CGAACG		RHQALQVVRKR
					GGAAGGCGGCGTTTGTGCGTCTGATCTACA	CCGTCC		YALQSDHWSDR
					AAAAGGGCAATCGCGCTGTGCCAAGCAATT	CCTCGT		EEELLEEFFNGT
					GGCGGCCGATCTGCCTACAGCAGACAGTGT	CCGCCG		LERSPFALAKKV
					ACAAACTATACACAGCCATCCTCGCCTCGC	CCTCGA		SGDIASLWTDV

					GACTGCAGAGATGGATGGACGCGAACGCGC	CAACGA		RVNLTKYGLKF
					GGTTCACGATGTCGCAAAAAGGATTTAGAG	GAGCGG		GEAHGRRLQPL
					CGTTTAACGGTTGCCACGAACACAATTTTGT	ACGCGG		VSHTDKQLAPQ
					GGCTACCTGTCTACACGACCAAACACGACG	ATGCGA		QWASAIKTHMR
					CCTTCGAAAGAAGCTTGCGATCGTGTGGTA	CGGCCC		LRHLKRWTTLV
					TGATCTCCGGAACGCATTCGGCTCGCTCCCG	CCCCTC		DQGKTARMHER
					CATGAGTATTTGTGGCGAGTACTGGCGCGC	CCGCTA		IGSAFLTRPSGV
					CTTGGCATGCCACCGCAGTTTGTCGCAAGA	TCTCTC		YDASYAFAVRA
					GTGAGGCAGCTGTACGCGGATGCAAGCTTC	TTCCGC		RLNQVDTRSAL
					ACCGTGGAATCCCGTGACGGGACGACAGAC	GGACGG		KRKRIVNNSRCR
					CCCGTCCAGCTGGAGCGTGGCGTATATCAA	GCTTCC		VSGCSELETLAH
					GGGTGTCCTCTGAGTCCATACTTGTTTATCG	CCCGCC		VLNHCRFGSDSI
					CCGCACTCATTCCGCTGGTGCGCGCACTCCA	CCTGTC		RARHAETLLLIK
					CAAGCTCAAAGACCAGCACGGCATCGTGCT	CTAGTT		TTMERELTRPGR
					GGCTCCCGGCGTCACCGACTGCGTCTCGGC	GCGCCG		QHQRLLVDATV
					TTATGCCGACGACATCAAGATTTTCGCTCGA	CTCTCC		PEARDPVPSNDA
					TCGGGAACCGGCGCCAAAGCCCTCCACGAG	TTCTGG		AESNIGAMAPSI
					ATCGTGGTGCGATTTCTGTCCTGGACGAAC	ATGCGG		SHLRPDIQLYDN
					ATGGCAGCGAACCCGGCCAAGTGCGCGCTG	CAACTT		KTMEAVIIDLAV
					ATGGTGACCGATGGCGCTCGTGGCGGTGAC	GGACGA		AFEDQSTDDAA
					GACACCGATGCGAGCATGACGCTATCTATC	CCCACG		SSSSFARVKGVK
					GAAGGTGAGACGATCCCACGCCTCACCGGC	TGCACG		TKKYEVIKQFLE
					AAAGAAGGCTACGTGTATCTCGGAGTGGAA	CGTGCG		YKGYTVHVAAL

					GATGGGCTCGCGCATGAGCGCAGAGCAACG	CTGCCG		VYGSLGSVDTG
					TGCCTGCGCGACTCGCTCAAGGCAGCGAGC	CCAGTG		NFAVYTERLGL
					GCGGATGTCGTCCGCCTGCTTCGCAGCGAC	CGACGG		RKGAVRRLECS
					CTCGCACCTTGGCAAATCGTGCGTGCGATC	CGTTGA		LSARHINFAHR
					AAGTCGCACGTGCTCTCGCGTTTCGATTATG	TCCAGC		MWRRHAIAHTT
					TACTCCGACACTTGCGCCCGTTCCTTAGTCT	GTCATC		GLRLIGTNSVQQ
					CTTTGACGGCTTCGACAAGATGCTGGTGCG	TACACG		QGVQRAPAEKQ
					CGGAATCAAGAGGCTGTGCCAGCTGCCGCA	CCGTTC		QHQRPVQRPSR
					AACAGCCACGTCCGAGTTTTTGTTCTCGCCC	CTCCCG		AQAPRDQPSQQ
					ACCTCGGCGGGTGGCCTCGGCTTCCTGCCA	TATCGA		QQSQQSFQQQS
					CTGAAGGAGCTTTTTGCGGCGCTGCAGATT	CAGCAA		QQSQQSQQSQQ
					GTGCACGCACTTCAAATGCTGCACTCGAAG	CCGATC		SQQSRQRHAPTP
					GACGCCAACGTCCGTGCCATTGCACGGCAC	TCAACG		TPVPVPVPVPVS
					CAAGCGCTGCAAGTGGTGCGTAAGCGCTAC	GCGATG		TPTLTPTPTPTRR
					GCGCTGCAGTCTGACCATTGGAGCGACCGC	ATAGCC		PKPAPSSTQPQQ
					GAAGAGGAACTTCTCGAAGAGTTCTTTAAC	AACCTC		GAPAQRRQQRE
					GGCACGCTGGAGCGCTCGCCCTTCGCCCTC	CAGACC		QKKQPACRRRH
					GCAAAGAAAGTGTCAGGGGACATCGCGTCA	CGCGTG		ATAVRETPPAAP
					CTCTGGACGGATGTACGAGTGAATCTCACC	CCAACA		TAARTATPTARP
					AAGTACGGACTCAAATTCGGCGAGGCGCAC	TCCAAC		TATTRSTSTTRS
					GGACGGCGACTGCAGCCGCTCGTCAGTCAC	AGCTTC		TATATSTTRSSA
					ACTGACAAGCAGCTTGCACCTCAACAATGG	CAGCCA		PTSRPSAPRPRSS
					GCTAGTGCAATCAAGACGCACATGCGCCTG	TCCATT		APRPRSSAPTTR

					CGCCACCTCAAGCGATGGACCACCCTGGTC GACCAGGGTAAGACAGCCCGGATGCATGAG CGCATCGGAAGTGCGTTTCTGACAAGGCCG TCTGGAGTGTACGACGCCAGCTATGCGTTC GCGGTTCGAGCAAGACTAAACCAAGTCGAT ACGCGCAGCGCATTGAAGCGCAAGCGCATC GTGAACAATAGCCGCTGCCGTGTGAGTGGA TGCAGCGAGCTGGAAACGCTCGCCCACGTA CTGAATCACTGCCGATTCGGGTCAGACTCG ATACGAGCACGACACGCAGAGACACTCTTG CTGATCAAGACGACCATGGAACGTGAGCTC ACGCGGCCAGGGCGGCAGCACCAGCGCCTC CTGGTCGACGCGACAGTACCCGAAGCTCGT GATCCCGTACCAAGTAATGACGCTGCTGAG AGCAACATTGGCGCAATGGCACCCAGCATC TCCCACTTGCGCCCCGACATTCAACTCTACG ACAACAAGACAATGGAAGCCGTGATCATCG ACCTCGCGGTTGCGTTCGAAGACCAGTCAA CGGACGACGCCGCATCCAGTTCCAGTTTCG CGCGCGTCAAGGGTGTGAAGACAAAGAAAT ATGAGGTAATCAAGCAATTTCTGGAATACA AAGGCTACACTGTGCATGTAGCAGCGCTAG TCTATGGTTCACTAGGCTCCGTTGACACCGG	CGACAA CGACAA CGACAA CGACAA CGACGA CGACCC GCTGCT CGACGA CGTGCT C (SEQ ID NO: 1200)		SAAPTTRSAAPT AIATTSVYKSRR TTNAISGATTRR SASSKRAPMQP RTALTPTQQQQ RQ (SEQ ID NO: 1445)

					CAACTTTGCCGTGTATACGGAACGTCTCGGT CTCCGCAAGGGCGCAGTGCGGCGACTGGAG TGCTCGCTCTCGGCGCGCCACATCAACTTCG CGCACCGTATGTGGCGACGCCACGCGATCG CCCACACGACCGGTTTGCGCCTTATCGGCA CGAACAGCGTGCAACAGCAGGGTGTGCAGC GCGCCCCGGCAGAGAAGCAACAACATCAGC GACCAGTGCAGCGACCCAGCCGCGCGCAAG CTCCACGTGATCAGCCATCGCAGCAACAGC AGTCCCAGCAGTCGTTCCAGCAGCAGTCCC AGCAGTCCCAGCAGTCCCAGCAGTCCCAGC AGTCCCAGCAGTCGCGTCAACGCCACGCCC CAACTCCTACCCCAGTCCCAGTCCCAGTCCC AGTTCCAGTATCGACACCGACACTAACACC GACACCGACACCGACACGCCGGCCCAAGCC GGCTCCGTCCTCGACTCAACCCCAGCAAGG AGCTCCAGCACAGCGACGACAACAACGAG AGCAGAAAAAGCAACCTGCGTGCCGCCGTC GACACGCAACGGCAGTGAGAGAGACACCG CCAGCAGCTCCGACAGCAGCTCGGACAGCA ACTCCGACAGCGAGACCGACAGCGACAACG AGATCAACATCGACAACGAGATCAACAGCG ACAGCGACATCGACAACGAGATCATCAGCA			

					CCGACATCGAGACCATCAGCACCGAGACCG AGATCATCAGCACCGAGACCGAGATCATCA GCACCGACAACGAGATCAGCAGCTCCGACA ACGAGATCAGCTGCACCGACAGCGATAGCG ACAACCAGCGTTTACAAAAGTCGACGCACG ACCAACGCCATCTCTGGCGCAACGACGCGA CGCAGTGCGTCATCCAAGCGTGCGCCAATG CAGCCCCGAACTGCATTGACGCCAACACAA CAACAACAACGACAGTAAGCGGGGGGTCC AGACCCCACAAGAGAGAAGCAGGAATCAT GGTCCGCATGGACCAGTAGGGCACGCTCCA CAAAGGTTATCGCCCTCAAACCCATCACAC GAAGGATCTAAAAAGAAAGCAACAATCGA AATAGTAAATAGTAAATAGCTTAGAAAGGT CAACATGCGAAATGCATGAGGAACGTAAGA TGGTAATACATTTTTATA (SEQ ID NO: 1077)			
NeSL	Utopia-3_PI	.		Phytophthora infestans	GTACGGCCGAATCCCTGGCCTCTCAGCCTGT ACGCGGGGCTATACTTGGTCTAAGTATGCC CCGGTCGCTGGCCTCTGAGCCCGTACACAG CTCGGCCTCGCGGCTGCCTTCCCAGGCGCC GCCGACTTCGCGCTCTGGCGCGGCCCACAC GCCGCCGCCGAGCCTCCAAGCGCGCCCGTT GGCTTCCGCAGACGCAGGGCTCGCGGCGAC	GTACGG CCGAAT CCCTGG CCTCTC AGCCTG TACGCG GGGCTA	TAGA CGGC AAAG TTCTG GCCC GCGT AGGC	MPRSLASEPVHS SASRLPSQAPPT SRSGAAHTPPPS LQARPLASADA GLAATALASPQ DPPYDVAPPGR AAGRLPDSVSPG

					GGCCCTCGCCAGCCCCCAAGACCCCCCCTA	TACTTG	CGAA	ATLSAATARAL
					CGATGTGGCACCACCCGGCAGGGCGGCCGG	GTCTAA	AGGG	AVRHWPLELDS
					CAGGCTGCCCGACTCGGTATCGCCGGGTGC	GT (SEQ	CCCC	DSSDDEDAQDP
					TACACTCTCAGCCGCCACAGCTCGGGCCTT	ID NO:	GCCC	HAAAPGPPADV
					GGCGGTCCGCCATTGGCCCTTGGAGCTCGA	1201)	AGGT	ASVLAPPGRAGS
					CAGCGACAGCAGCGACGACGAGGACGCTC		AGGG	MLADPAALAAG
					AAGACCCCCACGCCGCCGCCCCAGGACCCC		TAAC	LARAPPPPSAPQ
					CCGCAGACGTCGCGAGTGTGCTTGCCCCAC		GCCC	DPSPALPAGPAG
					CCGGCAGGGCAGGCAGCATGCTCGCCGACC		TCGG	QNPRAAAPARV
					CAGCAGCGCTGGCAGCCGGTCTGGCTCGCG		GAAA	EVHTVVAPPGR
					CTCCGCCGCCCCCATCGGCACCCCAGGACC		ACCA	AGGMLPDPGLV
					CAAGCCCCGCCCTCCCTGCTGGCCCCGCCG		TTCTG	DSSPAAATAATP
					GCCAGAACCCTCGCGCCGCCGCCCCAGCCC		GTGT	APVAATATTAR
					GTGTTGAAGTCCACACAGTCGTGGCACCAC		TTGG	VAVEHHAHAEP
					CCGGCAGGGCGGGCGGCATGCTGCCCGACC		CTGTT	NQEHLPMARVL
					CGGGGTTGGTTGACTCAAGCCCAGCAGCGG		TTTCA	VEPMQVDECSS
					CTACCGCAGCTACTCCAGCTCCGGTAGCTG		ACAG	CDRSTLTADDG
					CTACTGCTACAACTGCTCGCGTCGCCGTGG		TCGA	SGDDVAAPSSM
					AGCACCACGCGCACGCTGAGCCGAACCAAG		ACTT	LSNDVAAPMDV
					AACATCTACCGATGGCTCGCGTGCTCGTGG		CAAC	DSGTSCPPTLQQ
					AGCCCATGCAGGTCGACGAGTGCAGCAGCT		TCGG	PLQRPRALHVGS
					GCGACCGCAGCACTTTGACAGCCGACGACG		AGCA	KRRRLNADDGE
					GGAGCGGTGACGACGTTGCAGCCCCAAGCT		TATC	EAHQLQEEEEA

					CCATGCTCAGCAACGACGTCGCCGCCCCCA		AGAT	GIHAPALRLSAA
					TGGACGTTGACTCCGGCACCAGTTGCCCTCC		ACAC	SAQPASVLAVY
					AACGTTGCAGCAGCCGCTCCAGCGACCGCG		TTACT	THNASRFDCTLC
					TGCGCTGCACGTAGGCAGCAAGCGCCGCCG		CACA	AYTAGSFASLKT
					ACTCAACGCAGATGACGGTGAAGAAGCTCA		CATTT	HRNSRHRRTAF
					CCAGCTGCAAGAAGAGGAGGAGGCTGGCA		AGAT	LDRFSAGCACG
					TTCATGCCCCAGCGCTCAGGCTATCCGCGG		ATCA	VPFASRLAAAR
					CCAGCGCTCAACCGGCCTCCGTCCTTGCCGT		GATA	HAQACASLSSAP
					GTACACGCACAACGCGTCGCGGTTCGACTG		GGGA	LAEASSAAGASS
					CACGCTGTGCGCGTACACGGCCGGCAGCTT		ACCT	HTVDGADSTVS
					CGCATCACTGAAGACGCACAGGAACTCTCG		TTATT	AAGHAEPDLPR
					GCACCGCCGCACCGCATTCCTCGACCGATT		AGGG	HNATELTASPPL
					CTCGGCGGGTTGCGCGTGTGGCGTGCCCTTC		AGAT	VSSTDVEVQAT
					GCCTCGAGGCTGGCTGCAGCAAGACACGCG		AACG	ETEATENRWGT
					CAAGCGTGCGCCAGCCTCTCGAGCGCACCA		GGTA	PLPRVLVASRIA
					TTGGCCGAGGCTTCGTCGGCAGCAGGAGCA		CACC	GRLAQVPPPRW
					TCAAGCCACACTGTCGACGGAGCCGACTCC		GGAT	GPPLPRTTIAGRI
					ACCGTCTCAGCCGCGGGTCACGCTGAGCCC		GGTA	ATRLAATPAPR
					GACCTGCCCCGCCACAACGCCACGGAGCTC		AATA	WSPPLPRSLVAS
					ACAGCGTCCCCCCCGCTGGTGAGCTCCACC		TACA	RIAGRLLPALPD
					GATGTTGAAGTGCAGGCAACAGAGACCGAG		AAAC	APACEDEAKDS
					GCCACCGAGAACCGCTGGGGCACTCCACTT		CTTCT	DEMDWEASEPH
					CCGCGCGTGCTGGTCGCGTCTAGGATCGCC		CTGTT	VEAPGPVDEETI

					GGGCGCCTCGCTCAGGTGCCACCCCCACGC		CTAA	DDADGEWLLRF
					TGGGGCCCCCCGCTACCGCGCACCACCATC		TCAG	DGACRANPGPG
					GCGGGCCGCATTGCGACGAGGCTCGCCGCC		TGTG	GAGAALFKPSG
					ACACCTGCCCCCCGCTGGAGCCCACCGCTT		AAAA	PVVWTCSHYDP
					CCACGAAGCCTGGTTGCGTCACGCATCGCC		CTGG	STTATNNTAEYT
					GGCCGCCTGCTACCCGCTCTGCCAGACGCG		TTTTC	ALLLGARAAAD
					CCTGCATGCGAGGACGAGGCCAAGGACAGC		GCCT	HGVTKLRVEGD
					GACGAAATGGACTGGGAGGCGAGCGAGCC		TTTG	STLVIQQVRGIF
					GCACGTCGAAGCGCCAGGGCCGGTCGACGA		GCGG	ATRSTRLRALRN
					GGAGACCATCGACGACGCCGACGGCGAGTG		ACTTT	KVKLELARVGS
					GCTCCTGCGCTTTGACGGTGCCTGCCGTGCC		TTCA	FSLHHIDRQANG
					AACCCAGGCCCAGGCGGCGCCGGCGCAGCT		CTCG	HADRLANAGLD
					CTCTTCAAGCCAAGCGGTCCCGTAGTGTGG		CATTT	RRRTQLECSVHP
					ACTTGTTCCCACTACGACCCGAGCACCACG		TTGG	DGRGCTNTSVA
					GCAACGAACAACACAGCAGAGTACACTGCG		GCAA	TAAPTASAAPST
					TTGCTTCTCGGGGCCAGGGCCGCCGCCGAC		TCGT	PTRPPATTAAPF
					CACGGCGTCACGAAGCTGCGCGTCGAGGGC		CTGC	HSDQGHIDEDD
					GACAGCACGCTGGTCATCCAGCAAGTGCGG		GGCT	ERRADIDDGEIY
					GGCATCTTCGCGACGAGAAGCACGCGGCTC		AGCT	APMTLGPDEVP
					CGAGCACTGCGCAACAAGGTCAAGCTCGAG		TGCT	ARRPRLRLRQLS
					CTCGCGCGCGTGGGGAGCTTCTCGCTCCAC		AGCG	DEELEAAGAIVE
					CACATCGACCGGCAGGCGAACGGTCACGCC		GCGG	RLSASLSAKITD
					GACCGCCTTGCCAACGCCGGCCTCGACCGA		ACGA	AEDWGTAEGYI

					CGCCGCACCCAGCTCGAGTGCAGTGTGCAC		GCGG	TALPHLLYDKLL
					CCCGATGGTAGGGGCTGCACCAACACCTCC		TCTCC	PYSRTAPRHQRP
					GTAGCGACCGCAGCGCCTACAGCTTCTGCA		GGGG	PRPSRNQQDHP
					GCACCCTCGACCCCAACCCGGCCTCCTGCA		GCGT	QPRRDQPQRNV
					ACTACGGCCGCTCCGTTTCACAGCGACCAA		TCAC	DEQQHAESQQG
					GGCCACATCGACGAGGACGACGAACGTCGT		CTTTC	EDQRQQQPPTR
					GCCGACATCGACGACGGTGAAATCTACGCC		CCCC	RRRRRGKRRGR
					CCAATGACGCTCGGGCCAGACGAAGTGCCT		GCGA	RQRRHPRQPGQ
					GCTCGGCGACCGCGCCTCCGACTCCGCCAA		GGCC	STRASQQSRQRR
					CTCAGCGACGAAGAACTCGAGGCCGCTGGT		AACT	PRPPRVTRHHRE
					GCCATCGTGGAAAGGCTCAGCGCATCGTTG		ACAC	HRIDEALDELHT
					TCAGCCAAGATCACCGACGCAGAAGACTGG		CGAT	LERARPQDRSAI
					GGGACGGCAGAAGGCTACATTACGGCCCTG		CTTCT	DKARRRVGRVR
					CCTCACCTGCTGTACGACAAGCTCCTACCGT		CTAC	GAINQHLLRHRF
					ACTCGCGGACAGCGCCCCGTCATCAGCGGC		ACTTT	DTDEKACVADI
					CACCCCGGCCATCGCGCAACCAGCAAGACC		TCTA	LEKAHAARAAR
					ACCCGCAGCCGCGACGCGACCAACCGCAGC		ATTC	TAQAAGAATST
					GCAACGTGGACGAACAGCAGCATGCGGAGT		GCCT	GGAATAPTQQA
					CACAGCAGGGTGAAGACCAGCGACAGCAG		CCGT	ATSALGDADDG
					CAGCCGCCGACGCGCCGCCGACGTCGGCGA		CTTC	TCPILADELWQY
					GGGAAGCGTAGAGGCCGCCGACAGCGACG		GGTC	FTGTNTPRWEF
					ACACCCTCGCCAGCCAGGACAGAGTACCAG		TTCG	NPATPVGEAFRT
					AGCCTCCCAGCAAAGCCGGCAGCGCCGACC		GTTG	AMARLPPATRL

					CCGACCCCCGCGCGTTACCCGCCACCACCG		TCGG	RELLTEAPTADE
					CGAGCACCGCATCGACGAGGCGTTGGATGA		GCTTT	IETQLQHVRGSS
					GCTCCATACACTGGAGCGGGCGCGGCCACA		TTTCT	SPGLDGIGYDVY
					GGACCGATCGGCGATTGACAAGGCGAGACG		TTTTG	QRFAQQLLPVLT
					GCGCGTCGGACGTGTCCGCGGGGCCATCAA		ACCA	ASFKRCWTAKM
					CCAACACCTGCTGCGCCACCGGTTTGACAC		ATCA	VPQSWQVGVVR
					CGACGAGAAGGCGTGCGTGGCGGACATCTT		GAGC	LLYKKGAHDDP
					GGAGAAAGCGCACGCGGCGCGAGCGGCAC		GCGC	ANWRPICLQQAI
					GTACGGCACAGGCGGCGGGCGCAGCAACGT		CATG	YKLYTGVLARR
					CTACAGGAGGAGCAGCGACAGCACCAACA		CGAC	LVRWLDVNDRH
					CAGCAAGCAGCCACGTCAGCCCTTGGTGAT		TCTTC	APGQKGFRAVN
					GCCGACGACGGAACCTGCCCCATCCTGGCC		TGGC	GCGEHNFLAAT
					GACGAGCTCTGGCAGTACTTCACCGGCACC		CAAT	LIDQARRKRRSL
					AACACGCCGCGCTGGGAGTTCAACCCCGCG		CAGA	YEVWYDFRNAF
					ACACCCGTCGGCGAGGCGTTCCGTACGGCG		GACC	GSVPFQLLWDS
					ATGGCAAGGCTTCCTCCGGCGACGAGGCTC		GGGC	LQRLGAPADFID
					CGCGAACTGCTGACAGAAGCCCCGACTGCG		CCTG	MCKGLYHQAAF
					GACGAAATCGAGACCCAGCTACAGCACGTG		TCCTC	VIGNAADGPTA
					CGGGGATCCTCGAGCCCAGGACTCGATGGC		GGAC	AIRQQVGVFQG
					ATCGGCTACGACGTCTACCAGCGCTTCGCA		AGCG	CPLSPQLFNVAI
					CAGCAACTGCTCCCGGTGCTCACCGCTTCGT		AGGC	SPLLFALRRLPE
					TCAAGCGATGCTGGACAGCCAAGATGGTGC		CTCC	TGVQLSGDDRV
					CACAGAGCTGGCAGGTGGGGGTCGTGCGAC		ACGG	GVSAYADDLKT

					TGCTGTACAAGAAGGGCGCGCATGACGACC		CCAG	FSSTKAGATKQ
					CGGCAAACTGGCGGCCCATTTGCCTCCAGC		CCAA	HELVAAFLAWT
					AAGCCATCTACAAGCTTTACACAGGTGTAC		TCAA	GMKANAAKCSS
					TTGCTCGGCGACTCGTGCGGTGGCTGGACG		GTCT	MGVRRNSNGAT
					TGAACGACAGGCACGCACCAGGACAGAAG		CGGC	EADNLDLALDG
					GGGTTCAGGGCAGTGAACGGCTGCGGGGAG		AGCG	TPIPSMTHMQSY
					CACAACTTCCTCGCAGCCACGCTCATCGAC		ACGC	TYLGIGDGFDH
					CAAGCACGTCGCAAGCGCCGATCCCTATAC		GTCTT	VHRRIELAPKLK
					GAGGTATGGTACGACTTCAGGAACGCCTTC		TCTAT	TLKQDTTALLES
					GGCAGCGTGCCATTCCAGCTGCTGTGGGAC		AGCG	GLAPWQVVKA
					TCGCTGCAGCGCCTCGGCGCTCCCGCAGAC		CAGC	VKVYLYPRVEY
					TTCATCGACATGTGCAAGGGGCTGTACCAC		TGAC	ALRHLRPEDQLL
					CAGGCCGCCTTCGTCATCGGGAACGCCGCC		GAGG	ESFDLHLRAGLR
					GACGGACCGACAGCAGCAATCCGGCAGCA		CCGA	HLLRLPKNANN
					GGTTGGCGTCTTCCAAGGGTGCCCACTGAG		TCTG	DFFYSPVSRGGL
					CCCCCAGCTCTTCAACGTCGCGATCAGCCC		GCGG	GLLPLVELHAAL
					GCTGCTCTTCGCGCTCCGCAGGCTGCCGGA		CCCC	QIAHGWQMLNS
					GACCGGGGTGCAGCTGTCCGGCGACGATCG		CGAT	TDPATRRIAREQ
					TGTTGGAGTTTCGGCTTATGCGGACGACCTC		TGGT	LHQIADARHRL
					AAGACCTTCAGCAGCACCAAGGCGGGCGCC		CCGA	DKAHWKERGD
					ACCAAGCAACACGAGCTCGTCGCCGCGTTC		CTTTC	ELCQLFLNLDLG
					CTGGCTTGGACTGGCATGAAAGCCAACGCA		GGCC	TSAHAPPKRRN
					GCCAAGTGCAGCAGCATGGGCGTACGGCGC		AATC	CDIGSLWVDVR

					AACAGCAACGGCGCGACAGAGGCGGACAA		AGCG	KNLQAFGLKLE
					TCTGGACCTCGCGCTCGACGGCACTCCGAT		ACGA	TAPADAESGTPA
					CCCCAGCATGACTCACATGCAATCGTACAC		CGAG	LPLQLRVPHHEK
					CTACCTGGGGATCGGAGACGGTTTCGACCA		GGGG	WLTHRDVLRHV
					TGTGCACCGCCGCATCGAGCTCGCGCCCAA		CAGG	KQHLKNKHWR
					GCTGAAGACGCTCAAGCAGGACACGACGGC		GGTT	AWCAFQDQGRT
					GCTACTCGAGTCCGGCCTCGCGCCTTGGCA		TACA	ARAHGGVGSSFI
					GGTGGTGAAGGCCGTGAAGGTTTACCTTTA		CTTTT	TRPRGMWESDY
					CCCTCGTGTCGAGTACGCGCTTCGGCATCTG		GCCC	RFAVAARLNMV
					AGACCCGAAGACCAGCTGCTGGAGAGCTTC		CCGT	DTSATLARRRLR
					GACCTTCACCTCCGCGCCGGCCTTCGCCACC		TTCG	AHDRCRYPGCR
					TGCTCCGCCTGCCCAAGAACGCCAACAACG		GCTT	WKESLEHVLNH
					ACTTCTTCTACTCGCCGGTCTCGCGCGGTGG		CAAC	CPGTMDAVRGR
					CTTGGGACTCCTGCCGCTGGTGGAGCTGCA		TTCA	HDGVLREIEHAL
					CGCGGCGCTACAAATCGCGCACGGGTGGCA		GGCC	RAPSGARRELR
					GATGCTTAACTCGACGGACCCAGCCACCCG		AAAA	VNQTVPGLPGP
					ACGCATCGCGCGCGAGCAGCTCCACCAGAT		TGGC	ALRPDIQVYNH
					TGCGGATGCAAGGCACCGCCTTGACAAGGC		GATT	DQRTVAVVDLA
					ACACTGGAAAGAGCGTGGCGATGAGCTCTG		TGGA	VAFDRQDRDDP
					CCAGCTCTTCCTTAACTTGGACCTCGGGACG		CCCT	ETSGLAKAAAE
					TCGGCCCACGCACCGCCCAAACGCCGCAAC		CCAC	KKAKYTGIQRH
					TGCGACATCGGCTCGCTGTGGGTCGACGTC		GCGC	LERQGWKVHLS
					CGCAAGAACCTGCAGGCCTTTGGGCTCAAG		CGTG	ALVYGSLGSVA

					CTCGAGACCGCCCCGGCCGACGCTGAGTCC		CCAC	PNNYKVYTEHL
					GGCACACCCGCCCTGCCGCTGCAGCTTCGC		TGCT	GLLKRDAKRLD
					GTGCCACACCACGAGAAGTGGTTGACGCAC		CGGC	RTLSVACIQSSR
					CGCGACGTGCTGCGGCACGTGAAGCAGCAC		ACCG	RIWNLHCAKHR
					CTGAAGAACAAACATTGGCGGGCGTGGTGT		GCGG	ARQHQTPSQSR
					GCCTTCCAAGACCAAGGCAGGACGGCACGG		CGAT	GRRVTETGGAP
					GCACACGGTGGGGTCGGCAGCAGCTTCATC		TCAG	SRTDRR (SEQ ID
					ACCCGCCCCCGTGGCATGTGGGAGAGCGAC		CGGG	NO: 1446)
					TACCGCTTTGCGGTGGCCGCTCGCCTCAATA		TGCA	
					TGGTGGACACGAGTGCCACTCTCGCGCGCC		ACTT	
					GACGCCTCCGCGCCCATGACCGCTGCCGCT		CGGG	
					ACCCGGGCTGCCGCTGGAAGGAGTCACTGG		CACG	
					AGCACGTGCTCAACCACTGTCCGGGCACGA		TGTG	
					TGGACGCCGTGCGTGGCCGCCACGACGGTG		CAAC	
					TGCTCCGAGAGATTGAGCACGCGCTCCGTG		ACAT	
					CACCATCGGGAGCACGCAGAGAGCTGCGCG		GCAG	
					TGAACCAGACGGTTCCGGGCCTACCCGGCC		CGCC	
					CCGCACTACGCCCCGACATCCAGGTCTACA		CATT	
					ACCACGACCAACGCACGGTCGCAGTGGTGG		GCAC	
					ACCTGGCGGTGGCCTTCGACAGGCAGGACC		GCCA	
					GCGACGACCCAGAGACCTCGGGCCTGGCCA		AGCG	
					AAGCCGCGGCGGAGAAGAAGGCCAAGTAC		GCAT	
					ACTGGCATCCAGCGACACCTCGAGCGCCAA		CGCG	

					GGGTGGAAGGTCCACCTTTCCGCGCTCGTCT		GGAC	
					ACGGGTCACTGGGCTCGGTGGCCCCCAACA		GACG	
					ACTACAAAGTCTACACGGAGCACCTGGGGC		CCTC	
					TCCTCAAGCGTGACGCCAAGCGGCTGGACC		GGCC	
					GGACGCTGTCGGTCGCGTGCATCCAGTCCA		GCTC	
					GCCGCCGCATCTGGAACCTGCACTGCGCCA		AAGC	
					AGCATCGGGCGCGGCAGCACCAGACGCCAT		GCAG	
					CCCAGTCAAGGGGCAGGCGGGTGACGGAG		CCCC	
					ACCGGGGGGGCTCCGTCGCGCACCGATCGC		GCCC	
					CGCTAGACGGCAAAGTTCTGGCCCGCGTAG		TTCC	
					GCCGAAAGGGCCCCGCCCAGGTAGGGTAAC		AGCA	
					GCCCTCGGGAAAACCATTCTGGTGTTTGGCT		CGAC	
					GTTTTTCAACAGTCGAACTTCAACTCGGAGC		CTCG	
					ATATCAGATACACTTACTCACACATTTAGAT		CGCC	
					ATCAGATAGGGAACCTTTATTAGGGAGATA		GTTT	
					ACGGGTACACCGGATGGTAAATATACAAAA		GGCG	
					CCTTCTCTGTTCTAATCAGTGTGAAAACTGG		GATC	
					TTTTCGCCTTTTGGCGGACTTTTTCACTCGC		GCCA	
					ATTTTTGGGCAATCGTCTGCGGCTAGCTTGC		TCAA	
					TAGCGGCGGACGAGCGGTCTCCGGGGGCGT		GACG	
					TCACCTTTCCCCCGCGAGGCCAACTACACC		TGCG	
					GATCTTCTCTACACTTTTCTAATTCGCCTCC		AGAG	
					GTCTTCGGTCTTCGGTTGTCGGGCTTTTTTCT		CCAG	

					TTTTGACCAATCAGAGCGCGCCATGCGACT		GCGG	
					CTTCTGGCCAATCAGAGACCGGGCCCTGTC		GGTC	
					CTCGGACAGCGAGGCCTCCACGGCCAGCCA		GGGC	
					ATCAAGTCTCGGCAGCGACGCGTCTTTCTAT		AAAA	
					AGCGCAGCTGACGAGGCCGATCTGGCGGCC		TATA	
					CCCGATTGGTCCGACTTTCGGCCAATCAGC		CTTA	
					GACGACGAGGGGGCAGGGGTTTACACTTTT		CTCT	
					GCCCCCGTTTCGGCTTCAACTTCAGGCCAAA		AAGT	
					ATGGCGATTTGGACCCTCCACGCGCCGTGC		ATGC	
					CACTGCTCGGCACCGGCGGCGATTCAGCGG		CCGA	
					GTGCAACTTCGGGCACGTGTGCAACACATG		ATCC	
					CAGCGCCCATTGCACGCCAAGCGGCATCGC		CTGC	
					GGGACGACGCCTCGGCCGCTCAAGCGCAGC		CCTCT	
					CCCGCCCTTCCAGCACGACCTCGCGCCGTTT		CAGG	
					GGCGGATCGCCATCAAGACGTGCGAGAGCC		CTGT	
					AGGCGGGGTCGGGCAAAATATACTTACTCT		ACGC	
					AAGTATGCCCGAATCCCTGCCCTCTCAGGCT		GGCC	
					GTACGCGGCCCCATACTTGACCTAAGTATG		CCAT	
					GGAGGATCCCTGGCCTCTCAGGCTGTACGC		ACTT	
					GRGGACCAAGTACAGCCGAATCCCTGGCCT		GACC	
					CTCAGCCTGTACGCGGGGCTATACTTGGTCT		TAAG	
					AAGTATGCCCCGGTCGCTGGCCTCTGAGCC		TATG	
					CGTACGC (SEQ ID NO: 1078)		GGAG	

							GATC CCTG GCCT CTCA GGCT GTAC GCGR GGAC CAAG TACA GCCG AATC CCTG GCCT CTCA GCCT GTAC GCGG GGCT ATAC TTGG TCTA AGTA	

							TGCC CCGG TCGC TGGC CTCT GAGC CCGT ACGC (SEQ ID NO: 1324)	
NeSL	Utopia-3_PR	.		Phytophthora ramorum	TGCGCGGCAGACCAAGACGCGCAGCCAACA ACAGACCGTGCAGCGTGCGGGTGGAAAGCG CCGCCGCCTGAACGCTGGTGACGATGAAGA CCAGCGAGAGCTGGCCGAGCTCCTGCTCGT GGACGAGGACAAGGCTGGCGCCGAACACC CCGCGCTCAGGCTGCCCACGGCCAGCGCTC ATCCGGCCTCCGTCCTCTCCGTGTACGCGCA CGCTGCAACTCGCTTCGACTGCACGCTGTGC ACGTACACGGCTGCCAGCCTCGCTTCGCTC AAGCGCCACCGCTCGTCTCGGCACCGACGC ACGGCCTTCCTCGACAAGTTCTTGGCGGGCT GCGCGTGCGGCACGCCCTTCGCATCGAGGT	TGCGCG GCAGAC CAAGAC GCGCAG CCAACA ACAGAC CGTGCA GCGTGC GGGTGG AAAGCG CCGCCG CCTGAA	TAGA CGGC AACA CTCT GGCC CGCG TCGG CCGA AAGG GCCC CACC CACG	MQDQVDAEQQ ARNRWGPPLPR PLVASRVAARL GEVPPPRWGPPL PRGVVVSRIAAR LEAVPVPRGGPP LPRSFVATRIAD RLAPPSPDLSLL DEEMKESEPPDP THHSADEDSTD AETADAVMEPA FVSDPPTATPRE

					TGGCCGCAGCCAGACACGCGCAAGCGTGCG	CGCTGG	TAGG	WRLQFDGACRG
					CCAACCTCTGCACCACCTCGGCGACGACTT	TGACGA	GAAC	GPNPGGAGALL
					CGACGGCAGCAAAGGCATCAAGCCCCACTG	TGAAGA	CGCC	YNPEGAVVWTG
					CTGCCGGAGGCAGACCCACCGTCCGTGCAG	CCAGCG	CTCG	SHYMPGAKETN
					TGGTCACCGCCGCGCCCGACCTGCCCCGCC	AGAGCT	GGAA	NSAEYTALLIGA
					AGTATCCCTCGGAGCTCGTTGCGTCCCCCCC	GGCCGA	ACCC	RAAADHGARQL
					GCAGCCGAGCTCCACCAACGTTGCAATGCA	GCTCCT	AGTC	RIEGDSLLVIRQ
					GGACCAGGTCGACGCAGAGCAGCAGGCGC	GCTCGT	TGGT	VKGLYATKSTR
					GAAATCGCTGGGGCCCACCGCTCCCTCGAC	GGACGA	GTTC	LRQLRNAVRHE
					CGCTGGTGGCCTCCCGGGTCGCAGCCCGAC	GGACAA	GGCC	LARVGQHSLHHI
					TCGGCGAGGTACCGCCGCCGCGATGGGGCC	GGCTGG	AAGA	DRQGNAFADRL
					CACCACTTCCGCGAGGGGTGGTGGTGTCCC	CGCCGA	AATG	ANRALDLKSDK
					GTATTGCGGCCCGGCTGGAAGCCGTCCCAG	ACACCC	CACC	VECKEHPVAGA
					TCCCACGCGGGGGACCCCCGCTCCCACGCA	CGCGCT	ACCA	CTTCMGSPSAGP
					GCTTCGTCGCGACCAGGATCGCAGACCGCC	CAGGCT	CCAC	PATPPPTTADIE
					TCGCTCCCCCCTCCCCGGACCTGTCGCTGCT	GCCCAC	GGCG	MADAGSDDELR
					GGACGAGGAGATGAAGGAGAGCGAGCCGC	GGCCAG	GAGG	ADIDDGEVYAP
					CTGACCCGACGCACCACTCAGCGGACGAGG	CGCTCA	TGCA	MRLEPGVIPTRR
					ATTCCACGGACGCAGAGACGGCGGACGCAG	TCCGGC	TTTCG	SRLRLRQLTDDE
					TGATGGAGCCCGCGTTCGTCAGCGACCCAC	CTCCGT	ACAG	MEAAGEVVERL
					CCACGGCGACACCCCGGGAGTGGCGCCTCC	CCTCTC	TCGA	SAGLSAKIADAD
					AATTCGATGGAGCTTGCCGTGGCGGGCCCA	CGTGTA	ACTT	DWETAEGYITA
					ACCCCGGTGGTGCTGGCGCACTGCTCTACA	CGCGCA	CAAC	LPYMLYDKLQQ

					ACCCCGAGGGTGCTGTCGTGTGGACGGGTT	CGCTGC	CCGC	YTQVRHGTARS
					CGCACTACATGCCGGGCGCCAAGGAGACGA	AACTCG	CACA	PAPHPQRRDVQ
					ACAACTCGGCCGAGTACACGGCCCTCCTGA	CTTCGA	TATC	GQVETHREPRH
					TCGGCGCCCGCGCAGCAGCAGACCACGGTG	CTGCAC	GGAT	ETIGQPDQPGEP
					CACGACAGCTGCGCATCGAGGGCGACAGCC	GCTGTG	ATAG	SPTRRRRRGKR
					TGCTGGTCATTCGTCAAGTCAAGGGATTGT	CACGTA	TTAC	KGRRQRRHPRR
					ACGCCACCAAGAGCACTCGGCTGCGACAGC	CACGGC	AGCT	TNCGGGGRQQR
					TCCGGAACGCCGTTCGCCACGAGCTGGCTC	TGCCAG	CTAG	KQRHPRPPRGTR
					GGGTGGGGCAGCACTCGCTCCACCACATCG	CCTCGC	TTAG	HHREHRIDEAID
					ACCGGCAAGGCAACGCGTTCGCGGACCGCC	TTCGCT	ACAT	ELHALERARPQ
					TCGCCAACCGGGCGCTCGACCTGAAGAGCG	CAAGCG	CGGA	ARPAIAKARRR
					ACAAAGTCGAGTGCAAGGAACACCCCGTGG	CCACCG	TAGG	VGRIRSAIDQQL
					CCGGCGCGTGCACGACATGCATGGGTTCGC	CTCGTC	AACT	LRHRFDTAEKE
					CCTCGGCGGGCCCGCCAGCGACACCGCCCC	TCGGCA	TCTTA	CVDGILAAART
					CTACCACTGCCGACATCGAGATGGCGGACG	CCGACG	GAAA	ARDARTTVRAA
					CGGGGAGCGACGACGAGCTCCGAGCAGAC	CACGGC	ATTA	AATGTTATPETA
					ATCGACGACGGTGAAGTGTATGCTCCGATG	CTTCCT	ACGG	VTSGTEQQDDN
					CGCTTAGAGCCCGGCGTGATCCCGACCCGC	CGACAA	GTAT	GTCPIPSEVLWR
					CGCTCGCGCCTCCGCCTTCGCCAGCTGACCG	GTTCTT	ACCG	HFDSVNTPQRDF
					ACGATGAGATGGAGGCAGCTGGTGAGGTAG	GGCGGG	GATG	DPEAPEGAAFRS
					TCGAGCGCTTGAGTGCGGGCCTGTCGGCCA	CTGCGC	GTAA	AMARLPAATRF
					AGATCGCGGACGCCGACGACTGGGAGACG	GTGCGG	ATAA	MELLKEEPSTDG
					GCGGAAGGCTACATCACAGCCCTCCCGTAC	CACGCC	AATA	IEVQLQHASSTS

					ATGCTCTACGACAAGCTGCAGCAGTACACT	CTTCGC	AAAA	SPGLDGVGYDV
					CAGGTGCGGCATGGCACGGCTCGTTCACCA	ATCGAG	CTTC	YKRFASQLLPVL
					GCGCCCCACCCACAGCGCCGAGACGTCCAA	GTTGGC	(SEQ	KAAFKCCWTHK
					GGCCAGGTGGAAACCCACCGGGAGCCGCG	CGCAGC	ID NO:	QVPQSWKLGVV
					GCACGAGACCATTGGTCAGCCAGACCAGCC	CAGACA	1325)	RLLYKKGDRED
					AGGTGAGCCGTCACCGACGCGCCGCCGCCG	CGCGCA		PANWRPICLQQ
					CCGAGGAAAGCGTAAAGGCCGCCGCCAGC	AGCGTG		AIYKIYTGVLAR
					GACGTCATCCCCGACGGACCAATTGTGGTG	CGCCAA		RLTRWQDANDR
					GTGGTGGCCGTCAGCAGCGGAAGCAACGGC	CCTCTG		HAPGQKGFRPV
					ATCCACGCCCACCGCGCGGGACACGCCACC	CACCAC		NGCGEHNFLAA
					ACAGGGAGCACCGCATCGACGAAGCCATCG	CTCGGC		MLIDHARRKHR
					ACGAACTGCACGCGCTGGAGCGTGCCCGTC	GACGAC		PLYEVWYDFRN
					CCCAAGCTCGCCCAGCCATTGCCAAAGCTC	TTCGAC		AFGSVPLGLLW
					GCCGCCGAGTTGGCCGGATCCGATCGGCGA	GGCAGC		DALERTGVPAE
					TCGATCAGCAGCTTCTGCGCCACCGCTTCGA	AAAGGC		YIAAVQGLYDH
					CACCGCCGAGAAGGAGTGCGTGGATGGCAT	ATCAAG		AAFMVGNAVD
					CTTGGCAGCAGCTCGCACGGCACGCGACGC	CCCCAC		GSTAPILQRVGV
					ACGGACAACAGTGAGGGCGGCCGCGGCAA	TGCTGC		FQGCPLSPPLFS
					CAGGGACGACAGCGACTCCAGAGACGGCA	CGGAGG		AAISPLLHALQL
					GTGACGAGCGGTACGGAGCAGCAGGACGA	CAGACC		LPSSGVQLSGDD
					TAACGGCACCTGCCCCATCCCGAGCGAGGT	CACCGT		RPGVSAYADDL
					GCTGTGGAGACACTTCGACAGCGTCAATAC	CCGTGC		KTFSGTKAGVT
					GCCCCAGAGGGACTTCGACCCGGAGGCTCC	AGTGGT		EQHELVAMFLR

					CGAGGGTGCTGCGTTCCGTTCTGCGATGGC	CACCGC		WTGMAANPAK
					CCGGCTGCCAGCAGCCACGCGATTCATGGA	CGCGCC		CRSMGVRRNGN
					GCTGCTGAAGGAAGAGCCGTCCACGGACGG	CGACCT		GAIEADHLELAL
					CATCGAGGTCCAGCTGCAGCACGCGAGCAG	GCCCCG		DDTPIPTLTHLQ
					CACCTCCAGCCCGGGCCTCGACGGCGTCGG	CCAGTA		SYTYLGIGDGFD
					CTACGACGTGTACAAGCGCTTCGCGTCGCA	TCCCTC		HVRRRVALAPK
					GCTGCTGCCGGTGCTGAAGGCGGCTTTCAA	GGAGCT		LKLLKQDATAL
					GTGCTGCTGGACGCACAAGCAGGTGCCCCA	CGTTGC		MESGLAPWQVV
					GAGCTGGAAGCTCGGCGTGGTGCGCCTCCT	GTCCCC		KAVKGYLYPRV
					GTACAAGAAGGGCGACCGTGAAGACCCGG	CCCGCA		EYALRHLRPDD
					CGAACTGGAGGCCCATCTGCCTCCAGCAAG	GCCGAG		QLLESFDLHLRR
					CCATCTACAAGATCTACACGGGGGTGCTCG	CTCCAC		GLRHLLRLPKSA
					CGCGGCGCCTCACCCGCTGGCAGGACGCCA	CAACGT		NNDFVYAPVSR
					ACGACCGCCACGCCCCGGGCCAGAAGGGCT	TGCA		GGLGFLPLVELH
					TCCGGCCGGTGAATGGCTGCGGGGAACACA	(SEQ ID		AALQIAHGWQM
					ACTTCCTCGCGGCCATGCTCATCGACCACGC	NO: 1202)		INSPDPAIRRIAR
					GCGCCGCAAGCACCGGCCACTGTACGAGGT			EQLHQVADARH
					GTGGTACGACTTCCGGAACGCTTTCGGAAG			RLDKDHWKQR
					CGTACCCCTCGGGCTGCTGTGGGACGCCCT			GDELCELLLNGE
					GGAGCGGACAGGCGTTCCCGCCGAGTACAT			LGTSAHAPPKRR
					CGCTGCCGTCCAGGGCTTGTACGACCACGC			NGDIGSLWVDV
					AGCGTTCATGGTGGGCAACGCGGTGGACGG			RKNLKAFGLKL
					CAGCACGGCGCCGATCCTGCAGCGTGTCGG			ATAPADPESGAP

					AGTGTTCCAGGGGTGCCCGCTGAGCCCACC			AKPLQLCVPHH
					GCTGTTCTCGGCGGCGATCAGCCCCCTCCTC			AEWLDHRNVLR
					CACGCGTTGCAGCTACTGCCCAGCTCAGGA			HVKQHMKNKR
					GTACAGCTCTCGGGTGACGACCGCCCGGGT			WRAWCSHVDQ
					GTCTCCGCGTACGCCGACGACCTCAAGACC			GRTARAHGGVG
					TTCAGCGGCACCAAGGCGGGCGTCACCGAG			SGFLTRPRGMW
					CAGCACGAGCTAGTGGCCATGTTCCTGCGA			ESDYRFAVAAR
					TGGACCGGCATGGCGGCGAACCCGGCCAAG			LNMLDTVNVLA
					TGCAGGTCGATGGGCGTGCGCCGCAACGGC			RRRLRAHDRCR
					AACGGCGCCATCGAGGCGGACCACCTGGAG			HPGCRWKETLA
					CTGGCGCTGGACGACACGCCGATCCCTACG			HVLNHCPGTMD
					TTGACGCACCTGCAGTCGTACACGTACCTG			SIRGRHDDALKE
					GGGATCGGCGACGGCTTCGACCACGTGCGC			IERTLHASSGDR
					CGCCGCGTCGCGCTCGCGCCCAAGCTCAAG			QGRTELRTNQT
					TTGCTCAAGCAGGACGCGACGGCGCTGATG			VPGLAGPALRP
					GAGTCCGGCCTCGCGCCGTGGCAGGTGGTG			DLQVYNHDQRT
					AAGGCCGTGAAGGGGTACCTCTACCCTCGG			VAVVDLAIAFD
					GTGGAGTACGCCCTCCGGCACCTGCGACCC			EQPRDDPESSGL
					GACGACCAGCTCCTGGAGAGCTTCGACCTC			AKAAAEKKAKY
					CACCTCCGCCGCGGCCTCCGCCACCTGCTCC			AGIKRHLERQG
					GCCTCCCCAAGAGCGCCAACAACGACTTCG			WKVHLSALVYG
					TGTACGCGCCGGTGTCGCGCGGGGGCCTGG			SLGSVAPSNYK
					GCTTCCTGCCGCTGGTGGAGCTGCACGCGG			VYTEHLGLLKR

					CGCTGCAGATTGCGCACGGGTGGCAGATGA TCAACTCGCCCGACCCAGCCATCCGGCGCA TCGCGCGCGAGCAGCTTCACCAGGTTGCCG ACGCGCGGCACCGGCTCGACAAGGACCACT GGAAGCAGCGCGGTGACGAGCTGTGCGAGC TCCTCCTCAACGGGGAGCTCGGGACGTCGG CCCACGCGCCACCCAAACGGCGCAACGGCG ACATTGGCTCGCTGTGGGTCGACGTCCGCA AGAATCTCAAGGCCTTCGGGCTGAAGCTCG CAACCGCGCCCGCCGACCCGGAGTCCGGCG CACCTGCCAAGCCGCTGCAGCTTTGCGTGC CGCACCACGCTGAGTGGCTGGACCACCGCA ACGTGCTGCGGCACGTGAAGCAGCACATGA AGAACAAGCGCTGGCGGGCCTGGTGCAGCC ACGTGGACCAGGGGCGGACGGCGCGCGCG CACGGAGGCGTCGGCAGCGGTTTCCTGACG AGACCGCGGGGCATGTGGGAGAGCGACTAC CGCTTCGCGGTGGCCGCTCGCCTCAACATG CTTGACACGGTGAACGTGCTCGCGCGCAGA CGCCTCCGCGCCCACGACCGTTGCCGTCAC CCGGGCTGTCGGTGGAAGGAGACGTTGGCA CACGTCCTCAACCACTGCCCGGGTACGATG GACAGCATCCGCGGCCGCCACGACGACGCG			DAKRLDRQLSV ACIQSSRRIWNL HCAQHRARQHQ DQPAPRGRRVT ETGGTPSRTDRR (SEQ ID NO: 1447)

					CTCAAGGAAATCGAGCGCACGCTCCACGCG TCGTCGGGCGACCGCCAGGGCCGCACCGAG CTCCGCACGAACCAGACCGTGCCGGGCCTG GCCGGGCCAGCTCTGCGCCCCGACCTCCAG GTCTACAACCACGACCAACGCACGGTCGCG GTGGTGGACCTGGCCATCGCGTTCGACGAG CAGCCGCGCGACGACCCAGAGAGCTCCGGG CTGGCGAAGGCGGCGGCCGAGAAGAAGGC CAAGTATGCGGGCATCAAGCGACACCTCGA GCGCCAAGGGTGGAAGGTCCACCTCTCGGC GCTCGTGTACGGCTCGCTGGGCTCGGTGGC ACCCAGCAACTACAAGGTGTACACCGAGCA CCTGGGGCTGCTCAAGCGGGACGCCAAGCG GCTGGACCGGCAGCTGTCGGTGGCGTGCAT CCAGTCCAGCCGCCGCATCTGGAACCTGCA CTGCGCCCAGCACCGCGCCAGGCAGCACCA AGACCAACCAGCCCCAAGGGGCAGGCGGG TGACGGAGACCGGGGGGACTCCGTCGCGCA CCGATCGCCGCTAGACGGCAACACTCTGGC CCGCGTCGGCCGAAAGGGCCCCACCCACGT AGGGAACCGCCCTCGGGAAACCCAGTCTGG TGTTCGGCCAAGAAATGCACCACCACCACG GCGGAGGTGCATTTCGACAGTCGAACTTCA			

					ACCCGCCACATATCGGATATAGTTACAGCT CTAGTTAGACATCGGATAGGAACTTCTTAG AAAATTAACGGGTATACCGGATGGTAAATA AAATAAAAACTTC (SEQ ID NO: 1079)			
NeSL	Utopia-4_PI	AATU01001281.1		Phytophthora infestans	AGCTCGGCCTCGCGGCTGCCTTCCCAGGCG	AGCTCG	TAAA	MLADPAALAAG
					CCGCCGACTTCGCGCTCTGGCGCGGCCCAC	GCCTCG	CGGG	LARAPPPPSAPQ
					ACGCCGCCGCCGAGCCTCCAAGCGCGCCCG	CGGCTG	TCAC	DPSPAFPAGPAG
					TTGGCTTTCGCAGACGCAGGGCTCGCGGCG	CCTTCC	TTGA	QNPRAAAPARV
					ACGGCCCTCGCCAGCCCCCAAGACCCCCCC	CAGGCG	CCGA	EVHTVVAPPGR
					TACGATGTGGCACCACCCGGCAGGGCGGCC	CCGCCG	CAGG	AGGMLPDPGLV
					GGCAGGCTGCCCGACTCGGTATCGCCGGGT	ACTTCG	GCAC	EEPIQATYAHDA
					GCTACACTCTCAGCCGCTACAGCTCGGGCC	CGCTCT	CACC	AQFECALCPYV
					TTGGCGGTCCGCTATTGGCCCTTGGAGCTCG	GGCGCG	CAGG	AESMAVLVQHR
					ACGGCGACAGCAGCGACGACGAGGACGCT	GCCCAC	TAGG	RSAHRGTRFKDI
					CAAGACCCCCACGCCGCCGCCCCAGAACCC	ACGCCG	GAAC	FTSGCQCSLVFY
					CCAGAAGACGTCGCGAGTGTGCTTGCCCCA	CCGCCG	CGCC	ARIVAASHAVA
					CCCGGCAGGGCAGGCAGCATGCTCGCCGAC	AGCCTC	CTTTA	CARRNQRAVPP
					CCAGCAGCGCTGGCAGCCGGTCTGGCTCGC	CAAGCG	AAAC	APTPVAPTRPEA
					GCTCCGCCGCCCCCATCGGCACCCCAGGAC	CGCCCG	CCAG	TPQPTGYLAAA
					CCAAGCCCCGCCTTCCCTGCTGGTCCCGCCG	TTGGCT	GAAG	MTAAAAAASSD
					GCCAGAACCCTCGCGCCGCCGCCCCAGCCC	TTCGCA	ACAC	TVVAAATNMQS
					GTGTTGAAGTCCACACAGTCGTGGCACCAC	GACGCA	AAAC	AVPAAAKTTGL
					CCGGCAGGGCGGGCGGCATGCTGCCCGACC	GGGCTC	ACCC	QLVPPELEPALP

					CGGGGTTGGTTGAAGAACCTATCCAAGCGA	GCGGCG	TCCA	QRASCHAGKRR
					CGTACGCGCACGACGCTGCCCAGTTCGAGT	ACGGCC	CATA	RLNADEAVTPC
					GTGCGTTGTGCCCGTACGTCGCTGAGAGCA	CTCGCC	GTGA	TPTARVSPQTEV
					TGGCAGTGCTCGTTCAGCACCGACGGTCGG	AGCCCC	CATA	AMAPHDAPQDD
					CTCACCGCGGCACCCGCTTCAAGGACATCT	CAAGAC	CATA	TVLQREAAEPQP
					TCACCAGCGGGTGCCAGTGCAGCCTAGTGT	CCCCCC	TTTTA	DPAATQGAQVQ
					TTTACGCGCGCATTGTTGCCGCCAGCCATGC	TACGAT	GCCT	RVEDTTAAQDD
					AGTGGCCTGTGCCAGGCGCAACCAGCGAGC	GTGGCA	AGAT	TVQQDHDADTA
					CGTGCCACCAGCTCCGACCCCTGTTGCCCCC	CCACCC	TTCA	QVSPPRRTPTRW
					ACGCGGCCCGAGGCTACCCCACAACCTACT	GGCAGG	GTTA	GPRPSSTQEPSP
					GGGTACCTCGCCGCGGCGATGACTGCCGCT	GCGGCC	CGGA	MTGEPAATLAA
					GCGGCAGCCGCATCAAGCGACACCGTAGTG	GGCAGG	GAGG	RRPLTPAATGTR
					GCAGCCGCCACCAACATGCAGAGCGCCGTG	CTGCCC	TTACT	ATRWGPCHRAI
					CCAGCCGCCGCCAAGACTACTGGGCTTCAG	GACTCG	AACT	GAAAIARLVTG
					CTCGTCCCCCCCGAGCTGGAGCCTGCCCTTC	GTATCG	GGTA	LPTEPAQPQRRQ
					CGCAGCGTGCTTCTTGCCATGCAGGCAAAC	CCGGGT	CATA	PPPPQEPPLQPEP
					GCCGTCGGCTCAATGCCGACGAGGCAGTGA	GCTACA	AAAT	QAAAATVAADI
					CACCGTGTACGCCTACGGCGCGGGTGTCGC	CTCTCA	TACA	AATVAADIAAA
					CGCAAACAGAAGTGGCGATGGCGCCTCACG	GCCGCT	CATT	AANAAMDVDG
					ACGCACCCCAGGACGACACCGTCCTACAGC	ACAGCT	CTGTT	GPAADETWLLR
					GTGAGGCCGCCGAGCCACAGCCAGACCCAG	CGGGCC	CTAA	FDGACRRNPGP
					CAGCAACGCAGGGTGCCCAAGTGCAGCGAG	TTGGCG	TCAG	GGAGAALFAPS
					TCGAAGACACAACTGCGGCGCAGGATGACA	GTCCGC	TGTG	GAVVWTCSHFM

					CTGTGCAGCAGGACCACGACGCCGACACTG	TATTGG	AAAA	PSRSETNNTAEY
					CCCAGGTCTCGCCGCCCCGGCGCACACCGA	CCCTTG	CTGG	TALLLGAQSAV
					CACGCTGGGGACCAAGACCAAGCTCAACTC	GAGCTC	TTTTC	HHGAKRLNIEG
					AAGAGCCTAGCCCCATGACAGGCGAGCCGG	GACGGC	GCCT	DSHLILSQVRGA
					CTGCCACCCTGGCAGCGCGAAGACCACTGA	GACAGC	TTTG	FACNNKRLRSL
					CTCCTGCCGCAACGGGCACTCGTGCCACGC	AGCGAC	GCGG	RNRVQASLRQL
					GGTGGGGGCCATGTCACCGCGCGATCGGGG	GACGAG	ACTTT	DWYRLQHIDRK
					CCGCAGCCATCGCACGACTTGTCACGGGCC	GACGCT	TTCA	ANQHADRLANR
					TTCCAACGGAGCCAGCGCAGCCTCAGCGAC	CAAGAC	CTCG	ALDLRRTVTEC
					GGCAGCCACCGCCACCACAAGAACCTCCTT	CCCCAC	CATTT	GPHAETRNRCF
					TGCAGCCAGAGCCACAAGCAGCAGCAGCA	GCCGCC	TTGG	QTPQPLVEPGET
					ACAGTCGCAGCAGACATCGCAGCAACAGTA	GCCCCA	GCAA	HCVPGSDEVLA
					GCAGCAGACATTGCGGCAGCAGCAGCGAAC	GAACCC	TCGT	ANTAMEDATAV
					GCAGCAATGGACGTGGATGGTGGACCGGCA	CCAGAA	CTGC	PTEDDEAEVAA
					GCCGACGAGACATGGCTGTTGCGCTTTGAC	GACGTC	GGCT	RDGGEVFPTIAI
					GGCGCATGCCGACGGAACCCAGGACCTGGT	GCGAGT	AGCT	GPDSAPAKQPRL
					GGAGCAGGTGCCGCACTCTTTGCGCCCAGC	GTGCTT	TGCT	RLKKLDEDDFD
					GGTGCCGTAGTATGGACGTGCTCGCACTTC	GCCCCA	AGCG	AAAAAVTRVSE
					ATGCCGAGCCGCTCCGAAACGAACAACACC	CCCGGC	GCGG	ELASKIVDAGD
					GCCGAGTACACGGCGCTACTGCTCGGCGCA	AGGGCA	ACGA	WTSGEGYISAIP
					CAGAGCGCAGTCCACCACGGCGCCAAACGC	GGCAGC	GCGG	ERLRAALRPFAL
					TTGAATATCGAGGGCGATAGCCACCTGATC	(SEQ ID	TCTCC	PTQPARPQPREP
					CTCTCGCAAGTGCGCGGTGCCTTCGCGTGC	NO: 1203)	GGGG	RMQQPPRRPPR

					AACAACAAGCGGTTGCGGTCTCTCCGAAAC		GCGT	VTRDHLEHRLD
					CGAGTCCAGGCCTCTCTTCGGCAGCTCGACT		ACAC	EALDTMENVQR
					GGTACCGCTTGCAGCACATCGACCGCAAGG		CTTTC	STPQNQKAVRR
					CCAACCAGCATGCCGACCGCCTCGCGAACC		CCCC	ARRRVGRLRSA
					GCGCCCTGGACCTGCGCCGCACGGTGACCG		GCGA	MDRTRLRKKFA
					AGTGCGGCCCACACGCGGAGACTCGCAATC		GGCC	THERECVAEILR
					GGTGCTTTCAGACACCGCAGCCCCTCGTCG		AACT	RASTEEAANPSQ
					AGCCCGGCGAGACGCATTGTGTACCGGGGT		ACAC	EKCPIDRATLHE
					CTGATGAGGTCCTCGCTGCTAATACCGCGA		CGAT	YFTATSTQRTPF
					TGGAGGATGCGACCGCTGTGCCTACCGAAG		CTTCT	DYDSAKGTEFR
					ACGACGAGGCTGAGGTTGCAGCGCGAGACG		CTAC	TFLEVMSTPSHE
					GGGGTGAGGTGTTTCCAACCATCGCCATCG		ACTTT	TSALTAEPTLDE
					GGCCTGACTCCGCCCCTGCCAAGCAACCAC		TCTA	IEDQLAHVKAG
					GGCTTCGCCTCAAGAAGCTGGACGAAGACG		ATTC	SSPGHDGVGYD
					ATTTCGATGCGGCTGCCGCCGCCGTGACGC		GCCT	VYRRFQVQLLP
					GTGTGTCCGAGGAGCTGGCAAGCAAGATCG		TCGT	LLHAAFRFCWR
					TGGACGCTGGCGACTGGACATCAGGTGAAG		CTTC	HRRVPALWKVG
					GCTACATCAGCGCAATTCCTGAACGACTTC		GGTC	FVRLLHKKGDP
					GAGCAGCGCTCCGGCCATTTGCCCTTCCGA		TTCG	QQPNNWRPICL
					CGCAGCCAGCACGACCACAGCCCCGGGAGC		GCTG	QTAIYKLYSGLL
					CTCGTATGCAGCAGCCACCCCGCCGGCCAC		TCGG	ARRLSKFLEANE
					CCCGCGTCACACGCGACCACCTGGAGCACA		ATTTT	LLPMAQKGFRA
					GGCTTGACGAGGCGCTCGACACCATGGAGA		TTTCT	FNGCHEHNFVA

					ATGTGCAGCGCTCCACGCCGCAAAACCAGA		TTTTG	TTLLDQTRRMH
					AGGCAGTGCGCAGGGCCCGGCGCCGTGTCG		ACCA	RRLYQVWYDLR
					GGCGGCTTCGTTCGGCGATGGACCGGACCA		ATCA	NAFGSLPQQLM
					GACTGCGGAAGAAGTTCGCCACACACGAAC		GAGT	WGVLRQLGVTE
					GGGAGTGCGTGGCCGAGATCCTGCGACGTG		GCGC	EFVARCSGIYED
					CTTCAACGGAGGAGGCGGCGAACCCCAGTC		CATG	SYFVVGNASDG
					AGGAGAAATGTCCTATCGACCGCGCCACCC		CGAC	ATEPVRQEVGV
					TCCACGAGTACTTTACCGCCACCAGCACCC		TCTTC	YQGCPLSPLLFIT
					AGCGGACACCATTCGACTACGACAGCGCCA		TGGC	ALVPLLRALEN
					AAGGCACAGAGTTCCGCACGTTCCTGGAGG		CAAT	QDGVGVPLADG
					TGATGTCGACGCCTAGCCACGAGACAAGCG		CAGA	VRPCATAYADD
					CGTTGACCGCAGAGCCTACGCTGGACGAGA		GACC	IKVFCDSATGIQ
					TCGAAGACCAGCTCGCCCACGTCAAAGCCG		GGGC	RCHALVTRFLE
					GCTCAAGTCCAGGGCATGATGGGGTCGGGT		CCTG	WTGLQANPAKC
					ACGACGTGTACCGCCGCTTCCAGGTTCAGC		TCCTC	AFLPVTRSQHSN
					TGCTGCCCCTCCTGCATGCAGCCTTCCGGTT		GGAC	PTRDRDIELRIH
					TTGCTGGCGCCACCGCCGCGTGCCTGCCCTG		AGCG	GEAIATLGLQES
					TGGAAGGTCGGCTTCGTTCGCCTGCTGCAC		AGGC	YRYLGVGDGFD
					AAGAAAGGTGACCCGCAGCAGCCAAACAA		CTCC	HVRHRLQLEPK
					CTGGCGACCAATCTGCCTACAGACCGCCAT		ACGG	LKQIKREAVALL
					CTACAAGCTCTACAGCGGCCTCCTCGCCCGT		CCAG	HSELVPWQILKA
					CGACTGTCCAAGTTCCTGGAAGCCAACGAG		CCAA	LKVYIYPKVEY
					CTGTTGCCCATGGCACAGAAGGGCTTCCGA		TCAA	ALRHLRPLKSQL

					GCGTTCAACGGTTGCCACGAGCACAACTTC		GTCT	QGFDSAIVRGLR
					GTGGCAACCACGCTCCTGGACCAGACGCGC		CGGC	HLLRLPENSHDG
					CGCATGCATCGCAGGCTGTACCAGGTCTGG		AGCG	LFFSPTSAGGLG
					TACGACCTGCGCAATGCTTTCGGGTCGCTGC		ACGC	LLSLVELHEALQ
					CGCAGCAGCTCATGTGGGGCGTCTTACGCC		GTCTT	VAHAWQMLHS
					AGCTTGGGGTGACAGAGGAGTTTGTGGCTC		TCTAT	KDPAIRAIARTQ
					GCTGCAGTGGCATCTACGAGGACTCATACT		AGCG	VGQVARKRFKL
					TCGTGGTCGGCAACGCGAGCGACGGAGCGA		CAGC	VEEHWRGREDD
					CGGAGCCAGTGCGCCAGGAGGTGGGAGTAT		TGAC	LAQRFLNTELA
					ACCAGGGCTGCCCGCTCAGCCCCCTCCTCTT		GAGG	ASPHATETRRN
					TATCACTGCGCTCGTGCCACTCCTCCGGGCG		CCGA	GDIGSLWNDVR
					CTAGAGAACCAGGATGGCGTTGGCGTCCCG		TCTG	DTLQTLGLKFA
					CTCGCAGACGGAGTACGCCCGTGCGCCACC		GCGG	AGDEEEAPGLL
					GCGTACGCGGACGACATCAAGGTCTTCTGT		CCCC	QLRVPHHTKWL
					GACAGCGCGACGGGCATCCAGCGCTGCCAC		CGAT	SHSTVLRHVKL
					GCCCTCGTGACGCGCTTCCTGGAGTGGACT		TGGT	HMKLRRMDTW
					GGGCTGCAGGCCAACCCGGCCAAGTGCGCG		CCGA	KSKVSQGTTVR
					TTCCTGCCGGTGACACGCAGCCAGCACAGC		CTTTC	EHGGVGSRFITA
					AACCCCACACGCGACCGAGACATCGAGCTT		GGCC	GAGLSDAEYRF
					CGCATCCACGGCGAGGCCATCGCCACCCTC		AATC	AIAPRAHLIDTN
					GGACTTCAAGAAAGCTACCGGTACCTGGGG		AGCG	STLKRRRLRAN
					GTTGGCGACGGCTTCGATCACGTGCGACAC		ACGA	DTCRAPGCSYTE
					CGCCTCCAGCTCGAACCCAAGCTGAAGCAG		CGAG	PPAHILNKCSPN

					ATCAAGCGAGAGGCGGTGGCACTGCTGCAC		GGGG	MDAIRKRHDDA
					TCGGAGCTGGTGCCGTGGCAGATTCTCAAG		CAGG	LERIADALRRKV
					GCTCTGAAGGTGTACATCTACCCGAAGGTA		AGTT	EKSGGRLEVAIN
					GAGTACGCCCTCAGACACCTGCGGCCGCTC		TACA	KTVPEYDGAAL
					AAGTCACAGCTTCAGGGCTTCGACAGCGCC		CTTTT	RPDIVLRNTETK
					ATCGTGCGCGGCCTGCGCCACCTGCTCCGC		GCCC	RAIIADLAITHEN
					CTGCCGGAGAACAGCCACGACGGCCTCTTC		CCGT	QPTDATTSSALQ
					TTCTCACCCACCTCGGCTGGGGGACTGGGA		TTCG	QSRDNKITKYQT
					CTGCTGTCGCTCGTCGAGCTGCACGAAGCG		ACAT	VAAAMMRAGW
					CTGCAGGTGGCACACGCGTGGCAAATGCTC		CAAC	RVRVTGIVYGSL
					CACTCCAAGGACCCGGCCATCCGTGCCATC		TTCA	GSVLPSNFKVYT
					GCGCGCACGCAGGTGGGACAAGTGGCGCGC		GGCC	ELLALLKRDAR
					AAGCGGTTCAAGCTGGTCGAGGAGCACTGG		AAAA	RLNRQLSSHCIR
					CGCGGACGGGAAGACGACCTCGCACAGCGC		TGGC	ASARIWSAHCR
					TTCCTCAACACGGAGCTCGCCGCCTCGCCCC		GATT	RHRERQRSGNA
					ACGCCACGGAGACACGCCGCAACGGCGAC		TCGA	SRASRGSGGAPR
					ATCGGTTCCCTATGGAACGACGTCCGCGAC		CCCT	RTSQASARR
					ACGCTGCAGACCCTCGGACTCAAGTTCGCC		CCAC	(SEQ ID NO:
					GCAGGCGACGAGGAGGAAGCACCTGGACT		GCGC	1448)
					ACTGCAGCTCCGCGTGCCGCACCACACCAA		CGTG	
					GTGGCTCAGCCACTCGACGGTGTTGAGGCA		CCAC	
					CGTGAAGCTGCACATGAAACTGCGCCGCAT		TGCT	
					GGACACGTGGAAGTCCAAGGTCAGTCAGGG		CGGC	

					CACGACGGTGCGAGAGCACGGCGGCGTGG		ACCG	
					GCTCACGCTTCATCACGGCGGGTGCAGGCC		GCGG	
					TGTCCGACGCCGAGTATCGCTTCGCTATTGC		CGAT	
					GCCCCGCGCCCACCTGATTGATACCAACAG		TCAG	
					CACCCTCAAGCGCCGTCGACTCCGGGCCAA		CGGG	
					CGACACCTGCCGGGCTCCGGGATGCTCCTA		TGCA	
					CACGGAACCCCCCGCCCACATCCTGAACAA		ACTT	
					ATGCAGTCCCAACATGGACGCGATCAGGAA		CGGG	
					ACGCCACGACGATGCTCTCGAGCGTATCGC		CACG	
					TGACGCCCTGCGGCGCAAGGTGGAGAAGTC		TGTG	
					CGGCGGCAGGCTCGAGGTGGCCATCAACAA		CAAC	
					GACGGTACCGGAGTACGACGGCGCGGCCCT		ACAT	
					GCGCCCCGACATCGTACTCCGGAACACGGA		GCAG	
					GACCAAGCGTGCCATCATCGCCGACCTGGC		CGCC	
					GATCACCCACGAGAACCAGCCGACCGACGC		CATT	
					GACCACGTCTTCAGCGCTACAGCAAAGCCG		GCAC	
					TGACAACAAGATTACCAAGTACCAGACGGT		GCCA	
					CGCGGCGGCAATGATGCGCGCCGGTTGGCG		AGCG	
					GGTTCGCGTGACGGGCATCGTCTACGGCTC		GCAT	
					GCTCGGCTCAGTCCTGCCGAGCAACTTCAA		CGCG	
					GGTCTACACCGAGCTGCTGGCGCTGCTCAA		GGAC	
					GCGGGACGCCCGCCGCCTTAACCGCCAGCT		GACG	
					CTCCAGCCACTGCATCCGGGCCAGCGCGCG		CCTC	

					GATATGGAGCGCGCACTGCCGACGACACCG		GGCC	
					TGAGCGGCAACGAAGCGGCAATGCCTCGCG		GCTC	
					AGCGTCGCGTGGGTCCGGGGGGGCCCCGCG		AAGC	
					GCGCACATCGCAGGCATCGGCACGGCGGTA		GCAG	
					AACGGGTCACTTGACCGACAGGGCACCACC		CCCC	
					CAGGTAGGGAACCGCCCTTTAAAACCCAGG		GCCC	
					AAGACACAAACACCCTCCACATAGTGACAT		TTCC	
					ACATATTTTAGCCTAGATTTCAGTTACGGAG		AGCA	
					AGGTTACTAACTGGTACATAAAATTACACA		CGAC	
					TTCTGTTCTAATCAGTGTGAAAACTGGTTTT		CTCG	
					CGCCTTTTGGCGGACTTTTTCACTCGCATTT		CGCC	
					TTGGGCAATCGTCTGCGGCTAGCTTGCTAGC		GTTT	
					GGCGGACGAGCGGTCTCCGGGGGCGTACAC		GGCG	
					CTTTCCCCCGCGAGGCCAACTACACCGATCT		GATC	
					TCTCTACACTTTTCTAATTCGCCTTCGTCTTC		GCCA	
					GGTCTTCGGCTGTCGGATTTTTTTCTTTTTGA		TCAA	
					CCAATCAGAGTGCGCCATGCGACTCTTCTG		GACG	
					GCCAATCAGAGACCGGGCCCTGTCCTCGGA		TGCG	
					CAGCGAGGCCTCCACGGCCAGCCAATCAAG		AGAG	
					TCTCGGCAGCGACGCGTCTTTCTATAGCGCA		CCAG	
					GCTGACGAGGCCGATCTGGCGGCCCCCGAT		GCGG	
					TGGTCCGACTTTCGGCCAATCAGCGACGAC		GGTC	
					GAGGGGGCAGGAGTTTACACTTTTGCCCCC		GGGC	

					GTTTCGACATCAACTTCAGGCCAAAATGGC GATTTCGACCCTCCACGCGCCGTGCCACTGC TCGGCACCGGCGGCGATTCAGCGGGTGCAA CTTCGGGCACGTGTGCAACACATGCAGCGC CCATTGCACGCCAAGCGGCATCGCGGGACG ACGCCTCGGCCGCTCAAGCGCAGCCCCGCC CTTCCAGCACGACCTCGCGCCGTTTGGCGG ATCGCCATCAAGACGTGCGAGAGCCAGGCG GGGTCGGGCAAAATATACTTACTCTAAGTA TGCCCGAATCCCTGCCCTCTCAGGCTGTACG CGGCCCCATACTTGATCTAAGTATGGGAGG ATCCCTGGCCTCTCAGGCTGTACGCGAGAC CCGTACGGCCGAATCCCTGGCTTCTCAGCCT GTACGCGGGGCTATACTTGGTCTAAGTATG CCCCGGTCGCTGGCCTCTGAGCCCGTACAC A (SEQ ID NO: 1080)		AAAA TATA CTTA CTCT AAGT ATGC CCGA ATCC CTGC CCTCT CAGG CTGT ACGC GGCC CCAT ACTT GATC TAAG TATG GGAG GATC CCTG GCCT	

							CTCA GGCT GTAC GCGA GACC CGTA CGGC CGAA TCCCT GGCT TCTC AGCC TGTA CGCG GGGC TATA CTTG GTCT AAGT ATGC CCCG GTCG CTGG	

							CCTCTGAGCCCGTACACA (SEQ ID NO: 1326)	
NeSL	Utopia-4_PR	.		Phytophthora ramorum	CCAGGAATCACCCCCGCCGCCCCCAAGCGCTCCGCCGCCGAGCCCAGCTGCGGCTGCTGCCGCCCCACTGGTGACGGCAGTCGCAGCTGCCCCTGCCACCTCTCCTGTTGGCCGCGCCGCCGTGGAGCCCCGCGCGCGCGCCGAGCCGCCCCAAGAACAAGCACCCCCAGCTAGCGCGCGCGTGGAGCCCATGTCGGGCGACGTGGTCAGCAGCGACGGCAGCAGCAGGACGACGGACGCGAGCGGCGACGGCGACGACGGCGCAGGCAGCAGCGATGCAGCAGGCGATGTCGGCGTCGTCGCCATGGATGTTGACCAAGGTGCCCGCCGACAGCAGCCGCCATGGCAGCGCGTCGGCGGCAAGCGTCGACGTCTCAACGACGTCGACGACGAAGACACCCGCGAGCTGGCCGAGCTCCTGCTCGAGGAGGAGGACGAGGCTGGGGACCACGCTCCAGCGCCTAGGCTATCCGCGGCC	CCAGGATCACCGATCACCCGCCACGGCATCCCAGCGGCCCCGCCGCCGAGCGCCACAGGCGGCCACACCAGGAAGCCGCTGCGGCTGCTGCCCGCCGCCCCGCCAAGGCCGCCGCCGAGGCCGCCCCCAGCTGCGGCCCCGCCGCCG

					AGCGCTCGCCCGGCCTCCGTCCTCTCCGTGT	CACCTC	CCGG	AAPAAGELSPTA
					ACGCGCACAATGCGCAGCGCTTCCAGTGCA	TCCTGT	CCGG	GAANATVKAAT
					CGCTGTGCACGTACACGGCCGCCAGCTTCG	TGGCCG	TACA	VTPDSPRQDPPK
					CGTCGCTCAAGCGACACCGGGACTCCCGGC	CGCCGC	TTAT	LAATPPLASSAL
					ACCGCCGCACCGCCTTCCTCGACCGGTTCTC	CGTGGA	GGTC	VVDPDHAEQQA
					GGCGGGATGTGCGTGCGGCGTGCCCTTCGC	GCCCCG	CGAC	RERWGPPLPRTL
					GTCGAGGCTGGCTGCTGCGAACCACGCACA	CGCGCG	ACCT	VAGRVAARLSE
					CGCGTGCGACAGCCTCAACCGCACCTTCTC	CGCCGA	ATGA	VPAPRWGPPLPR
					GGTAGCGGCTGCGCCGGCAGCAGGGGAATT	GCCGCC	GGTG	GVVAFRIGHRV
					GAGCCCCACTGCTGGTGCAGCCAACGCCAC	CCAAGA	CAAC	LPPEMTSDEETK
					TGTCAAGGCGGCCACTGTCACGCCCGACTC	ACAAGC	CGGT	DDSSVQDGDRQ
					CCCCCGCCAAGACCCCCCGAAGCTCGCAGC	ACCCCC	ACAC	DYPVAAMDVDS
					GACCCCCCCCGCTAGCGAGCTCCGCTCTTGTC	AGCTAG	AAGT	GMSGEWLLRFD
					GTCGACCCGGACCACGCCGAGCAACAAGCC	CGCGCG	TACA	GACRANPGPGG
					CGCGAGCGCTGGGGCCCGCCCCTACCCAGG	CGTGGA	CACC	AGAALSQPDGS
					ACGCTGGTGGCAGGCCGTGTTGCAGCCCGC	GCCC	ACAT	VVWTCSHYMPS
					CTCAGCGAGGTTCCAGCTCCCCGCTGGGGA	(SEQ ID	AGCG	SSETNNTAEYTA
					CCCCCACTGCCACGCGGCGTGGTGGCATTC	NO: 1204)	ACTA	LLLGTRAAADH
					CGGATCGGCCACCGCGTGCTCCCGCCGGAG		CCAG	GTTTLRVEGDST
					ATGACGAGCGACGAGGAGACGAAGGACGA		GTAT	LVIQQVRGIFAT
					CTCGAGTGTCCAGGACGGCGACCGCCAGGA		TTACT	RSVTLRHLRDQ
					CTACCCGGTGGCCGCGATGGACGTGGACAG		ACCT	VKLELARVGRF
					CGGCATGAGCGGCGAGTGGCTGCTGCGCTT		GGAA	SLHHIDRQANA

					CGACGGCGCCTGCCGCGCCAACCCTGGCCC		GCCA	HADRLANRALD
					AGGTGGCGCCGGCGCAGCGCTGTCCCAGCC		AGGA	LRRTVSECGVHP
					CGACGGCTCCGTTGTATGGACTTGCTCCCAC		TTAA	DGNGCTPTAIDD
					TACATGCCGAGCAGCAGCGAAACCAACAAC		CCGG	RPLAPTQQPPDA
					ACCGCCGAATACACGGCGCTGCTGCTCGGT		TCGG	PPPPPAADIEME
					ACGCGAGCCGCCGCCGACCACGGAACCACG		TAAT	DPDDEDLADID
					ACCCTGCGAGTGGAGGGCGACAGCACCTTG		ACAC	DGEVYAAMRV
					GTGATCCAGCAGGTGCGTGGCATCTTTGCC		ATAA	GPNATPQRRRR
					ACGCGGAGCGTCACTCTCCGCCACCTCCGC		CTTT	GRSGTAKKHRR
					GACCAAGTCAAGCTGGAACTCGCCCGGGTG		(SEQ	QRPPRVTRHHR
					GGGAGATTTTCGCTCCACCACATTGATCGCC		ID NO:	EHRLDEALDDL
					AAGCCAATGCCCACGCCGACCGCCTCGCCA		1327)	HAVERSTPSDRT
					ACCGAGCCCTCGACCTCCGGCGGACAGTGT			TVRRARRRVGR
					CGGAGTGCGGCGTCCACCCCGACGGGAACG			VNSAIEQQRLRH
					GCTGCACGCCCACGGCCATCGACGACCGAC			RFDTDEKACVT
					CCCTAGCGCCGACTCAGCAGCCACCTGACG			DILAKACATREA
					CCCCGCCACCGCCGCCCGCCGCCGACATTG			ARTTASGGDPP
					AGATGGAGGACCCGGACGACGAGGACCTC			AGPATPAAGSA
					GCCGACATCGACGACGGCGAGGTCTACGCG			DDGTCPILGEEL
					GCCATGCGCGTTGGCCCGAATGCCACTCCG			WRFFDSVNTPR
					CAGCGACGACGCCGTGGACGCTCAGGCACC			QEFAPDAPVGA
					GCGAAGAAGCACCGCCGCCAGCGCCCACCC			AFRSALARLPAA
					CGGGTGACCCGACACCACCGCGAACACCGC			TSCKELLTAAPS

					CTCGACGAGGCACTCGATGACCTCCATGCA			AGEVEDQLQHV
					GTGGAGCGCTCCACCCCCAGCGACCGCACC			RGASSPGLDGV
					ACCGTCCGCCGGGCACGACGCCGCGTTGGA			GYDVYQHFAAQ
					CGCGTCAATTCGGCCATCGAACAGCAGCGC			LLPALTAAFKAC
					CTCCGCCACCGCTTCGACACGGACGAGAAG			WTAKRVPQSW
					GCATGTGTCACCGACATCCTGGCGAAGGCG			KLGVVRLLHKK
					TGCGCGACGCGAGAGGCAGCGAGGACCAC			GAREDPANWRP
					GGCCTCCGGCGGTGATCCACCTGCAGGCCC			ICLQQAIYKLYT
					GGCGACGCCAGCGGCAGGCAGCGCAGACG			GLLARRLVRWL
					ACGGCACCTGCCCCATCTTGGGAGAAGAAC			DANDRHAPGQK
					TCTGGCGGTTTTTCGACAGCGTCAACACGCC			GFRAVNGCGEH
					GCGCCAAGAATTCGCCCCGGACGCCCCGGT			NFLAATLVDQA
					GGGTGCCGCCTTCCGCTCGGCGTTGGCACG			RRKRRTLFEVW
					CCTACCGGCCGCGACCAGTTGCAAGGAGCT			YDFRNAFGSVPF
					GCTGACTGCGGCGCCGTCGGCCGGCGAAGT			ALLWDALARLG
					GGAAGACCAGCTACAACACGTCCGTGGCGC			VPDDYVTMCKG
					CTCCAGCCCTGGACTCGACGGCGTCGGGTA			LYESAAFVVGN
					CGACGTGTACCAGCACTTCGCGGCGCAGCT			AIDGTTDPIALR
					CCTGCCGGCACTGACCGCGGCATTCAAGGC			VGVFQGCPLSPQ
					ATGCTGGACGGCGAAGAGAGTGCCTCAGAG			LFNAAISPLLFA
					CTGGAAGCTCGGGGTGGTGCGGCTGCTCCA			LQRLPATGVQLS
					CAAGAAGGGCGCGCGCGAGGACCCGGCAA			GDDCPGASAYA
					ACTGGAGGCCGATCTGCCTCCAGCAAGCCA			DDLKIFSGTEDG

					TTTACAAGCTCTACACGGGACTCCTCGCGC			IKRQHALVADFL
					GGCGGCTGGTGCGTTGGCTAGACGCCAACG			RWTGMAANPN
					ACCGCCACGCTCCAGGCCAGAAGGGCTTCC			KCCTMSVQRDG
					GAGCGGTGAACGGCTGCGGGGAGCACAACT			RGVLKTDDLQL
					TCCTCGCAGCCACGCTCGTCGACCAAGCCC			DLAGTPIPALSM
					GCCGCAAGCGCCGCACCCTCTTCGAAGTCT			SASYTYLGIGDG
					GGTACGATTTTCGCAACGCGTTCGGCAGCG			FDHVRRRVELA
					TGCCATTCGCGCTGCTGTGGGACGCGCTGG			PALKQLKDDAT
					CGCGGCTCGGCGTGCCCGACGACTACGTTA			TLLQSGLAPWQ
					CCATGTGCAAGGGGCTCTACGAGAGCGCGG			VVKAVKTYLYP
					CGTTTGTCGTCGGGAACGCTATCGACGGCA			RVEYALRHLRPF
					CGACGGACCCAATTGCGCTGCGGGTGGGCG			QQQLEGFDRHL
					TCTTCCAAGGGTGCCCACTGAGCCCCCAAC			ARGLRHLLRLP
					TCTTCAACGCTGCCATTAGCCCGCTGCTCTT			GNATAECFYAP
					CGCGTTGCAGCGGCTGCCGGCCACGGGCGT			VSRGGLGLLPLT
					GCAGCTGTCGGGCGATGACTGCCCTGGTGC			ELHAALQVAHG
					CTCGGCTTATGCCGACGACCTCAAGATTTTC			WQLLNSKDPAI
					AGTGGCACCGAGGACGGCATCAAGCGACA			RRIARVQLRQIA
					GCACGCGCTCGTGGCCGACTTCTTGCGCTG			DARHRIDSRAW
					GACCGGCATGGCGGCGAACCCCAACAAGTG			EGRDEELCELLL
					CTGCACCATGTCGGTCCAGCGCGACGGACG			NSQLGTSPDAPP
					CGGCGTGCTCAAGACTGACGACCTCCAGCT			KRRNGDIGSLW
					AGACCTGGCTGGCACGCCGATCCCCGCGCT			VDVQRHLRTLG

					GTCCATGAGCGCTTCGTACACCTACCTGGG			LKFATAPACAD
					GATCGGCGATGGATTCGATCACGTGCGCCG			AGSAATTLQLR
					CCGCGTCGAGCTCGCTCCAGCACTCAAGCA			VPHHDKWLDHR
					GCTTAAGGACGACGCGACGACGCTGCTGCA			TVLRHVKLHVK
					GTCGGGCCTCGCCCCGTGGCAGGTGGTGAA			HRHWSKWAAM
					GGCGGTGAAGACGTACCTCTACCCCCGGGT			RDQGKTARAHG
					GGAGTACGCTCTCCGCCACCTTCGCCCCTTC			GAGSGFLTRPRG
					CAACAGCAGCTTGAGGGCTTCGACCGGCAC			MWEADYRFAV
					CTCGCGCGCGGCCTCCGCCACCTCCTGCGA			AARLNQLDTHS
					CTGCCGGGCAACGCCACGGCGGAGTGCTTC			VLKRRRLRAHD
					TACGCTCCGGTGTCGCGCGGCGGGCTGGGC			HCRQPGCSRAE
					CTGCTGCCACTGACGGAGCTGCACGCGGCG			TLAHVLNHCAG
					CTCCAAGTCGCGCACGGATGGCAGCTGCTG			TMDAVRGRHD
					AATTCGAAGGACCCTGCCATCCGGCGCATC			DALKHIERALH
					GCGCGCGTGCAGCTCCGGCAGATCGCCGAC			ASSPGGQDRVE
					GCGCGCCACCGTATCGACAGCCGCGCATGG			LRVNQTVPSLA
					GAGGGGCGCGACGAGGAGTTGTGCGAGCTG			GPALRPDLQLY
					CTGCTCAACTCGCAGCTCGGGACGTCGCCC			NHTKKMVAVV
					GACGCGCCACCGAAGCGTCGGAACGGCGAC			DLAVAFEEQAS
					ATCGGTTCTCTTTGGGTCGACGTCCAGCGCC			DDPESSALARIA
					ACCTCCGCACGCTCGGCCTCAAGTTCGCGA			AHKRAKYAGV
					CCGCGCCGGCCTGCGCGGACGCCGGCTCTG			KRHLERQGWKV
					CGGCGACCACGTTGCAGCTCCGCGTGCCGC			HLSALVYGSLG

					ACCACGATAAGTGGTTGGACCATCGCACTG TCCTGCGGCACGTGAAACTGCACGTGAAGC ACCGGCACTGGTCCAAGTGGGCGGCCATGC GCGACCAAGGCAAGACGGCCCGCGCCCATG GCGGCGCGGGGAGCGGCTTCCTCACCCGGC CACGAGGCATGTGGGAAGCCGACTACCGCT TCGCGGTGGCCGCCCGCCTGAACCAGCTCG ACACGCACAGCGTGCTGAAGCGCCGGCGTC TCCGAGCCCACGACCACTGCAGGCAGCCGG GCTGCTCCAGGGCGGAGACGCTGGCGCACG TGCTCAACCACTGCGCCGGCACCATGGACG CGGTCCGCGGCCGCCACGACGACGCCCTCA AGCACATCGAGCGCGCGCTGCACGCGTCGT CGCCGGGCGGCCAGGACCGCGTCGAGCTCC GGGTGAACCAGACGGTGCCGTCGCTCGCCG GCCCCGCGCTACGGCCCGACCTCCAGCTGT ACAACCACACCAAGAAGATGGTGGCGGTGG TCGACCTGGCCGTGGCGTTCGAGGAGCAGG CGAGCGACGACCCGGAGAGCTCGGCGCTGG CACGCATCGCCGCACACAAGCGCGCGAAGT ATGCCGGCGTCAAGCGACACCTCGAGCGCC AAGGGTGGAAGGTCCACCTCTCGGCGCTCG TGTACGGCTCGCTCGGCGCGGTGCCGGCCG			AVPAGNHKVLT EHLGLLKRDAK RLDRQLSVACIQ SSRRIWNLHCSQ HRARQHQAPGG SRAAETGGTPPR TGRR (SEQ ID NO: 1449)

					GCAACCACAAGGTGCTTACTGAGCACCTTG GGTTGCTCAAGCGTGATGCGAAGCGGCTGG ACCGGCAGCTCTCGGTGGCGTGCATCCAGT CCAGCCGCCGCATCTGGAACCTGCACTGCA GCCAGCACCGCGCGCGCCAGCACCAAGCAC CAGGGGGAAGTCGGGCGGCGGAGACCGGG GGGACTCCGCCGCGCACCGGCCGCCGCTAG ACGGCACACAGGCCCACAGCGGCCGACAG GGCCACACCCAGGTAGGGAACCGCCTTCAA ACCCGGCCGGTACATTATGGTCCGACACCT ATGAGGTGCAACCGGTACACAAGTTACACA CCACATAGCGACTACCAGGTATTTACTACCT GGAAGCCAAGGATTAACCGGTCGGTAATAC ACATAACTTT (SEQ ID NO: 1081)			
NeSL	Utopia-5_PI	.		Phytophthora infestans	CTCAAGCCTAGCAGCGGCTACCGCAGCTAC TCCAGCTCCGGTAGCTGCTACTGCTACAACT GCTCGCGCTGCTGCTCGCGTCGCCGTGGAG CACCACGCGCACGCTGAACCGAACCAAGAA CATCTACCGATGGCTCGCGTGCTCGTGGAG CCCATGCAGGTCGACGAGTGCAGCAGCTGC GACCGCAGCACTTTGACAGCCGACGACGGG AGCGGTGACGACGTTGCAGCCCCAAGCTCC TTGAACAGCAACGACGTCGCCGCCCCCATG	CTCAAG CCTAGC AGCGGC TACCGC AGCTAC TCCAGC TCCGGT AGCTGC TACTGC	TGAC GCAC CGTG ACAT AGTG CGGC ACGT GAAG ATGC	MARVLVEPMQV DECSSCDRSTLT ADDGSGDDVAA PSSLNSNDVAAP MDVDSGTDCPP ALQQPPQRPRAL HVGSKRRRLDA DDEEEARQLQE EEEAGIHAPALR

					GACGTTGACTCCGGCACTGATTGCCCTCCA	TACAAC	ACAT	LSAASAQPASVL
					GCGTTGCAACAGCCGCCCCAGCGACCGCGT	TGCTCG	GAAG	AVYTHNASRFD
					GCGCTGCACGTAGGCAGCAAGCGCCGCCGA	CGCTGC	CTCC	CTLYAYTAGSF
					CTCGACGCAGATGATGAGGAAGAAGCTCGC	TGCTCG	GACA	ASLKTHRNSRH
					CAGCTGCAAGAAGAGGAGGAGGCTGGCAT	CGTCGC	CTGG	RRTAFLDRFSAG
					CCATGCCCCAGCGCTCAGGCTATCCGCGGC	CGTGGA	GCCA	CACGVPFASRL
					CAGCGCTCAACCGGCCTCCGTCCTCGCCGT	GCACCA	AGTG	AAARHAQACAS
					GTACACGCACAACGCGTCGCGGTTCGACTG	CGCGCA	GGCG	LSSAPLAEASSA
					CACGCTGTACGCGTACACGGCCGGCAGCTT	CGCTGA	GCCA	AGASSHTVDEA
					CGCATCACTGAAAACGCACAGGAACTCTCG	ACCGAA	TGCG	DSTVSAAGHTEP
					GCACCGCCGCACCGCATTCCTCGACCGATT	CCAAGA	CGAC	DLPRHNATELT
					CTCGGCGGGTTGCGCGTGTGGCGTGCCCTTC	ACATCT	CAAG	ASPPLVSSPDVE
					GCCTCGAGGCTGGCTGCAGCAAGACACGCG	ACCG	GCAA	VQAPETEATEN
					CAAGCGTGCGCCAGCCTCTCGAGCGCACCA	(SEQ ID	GACA	RWGTPLPRVLV
					TTGGCCGAGGCTTCGTCGGCAGCAGGAGCA	NO: 1205)	GCTC	ASRIAGRLAQVP
					TCAAGCCACACTGTCGACGAAGCCGACTCC		GTGC	PPRWGPPLPRTT
					ACCGTCTCAGCCGCGGGTCACACCGAGCCC		ACAT	IAGRIATRLAAT
					GACCTGCCCCGCCACAACGCCACGGAGCTC		GGTG	PAPRWDPPLPRS
					ACAGCGTCCCCCCCGCTGGTGAGCTCCCCC		GGGT	LVVSRIAARLLP
					GATGTTGAAGTGCAGGCACCAGAGACAGAG		TGGT	ALPDAPACEEEA
					GCCACCGAGAACCGCTGGGGCACTCCACTT		AGTG	KDSDTMDWAPT
					CCGCGCGTGCTGGTCGCGTCTAGGATCGCC		GCTT	WTNEETKDSEP
					GGGCGCCTCGCTCAGGTGCCACCCCCACGC		CCTC	HDEAPGQVDEE

					TGGGGCCCCCCGCTACCGCGCACTACCATC		ACAC	TIDDADGEWLL
					GCGGGCCGCATTGCGACGAGGCTCGCCGCC		GGCC	RFDGACRANPG
					ACACCTGCCCCCCGCTGGGACCCACCGCTG		GCGA	PGGAGAALFKP
					CCACGAAGCCTGGTTGTGTCACGCATCGCC		GGCC	SGPVVWTCSHY
					GCCCGCCTGCTACCCGCGCTACCAGACGCG		TGTG	DPSTTATNNTAE
					CCTGCATGCGAGGAAGAGGCCAAGGACAG		GGAA	YTALLLGARAA
					CGACACGATGGACTGGGCGCCAACTTGGAC		GCCG	ADHGVTKLRVE
					AAACGAGGAGACCAAGGACAGCGAGCCGC		ACTA	GDSTLVIQQVRG
					ATGACGAAGCGCCAGGGCAGGTCGACGAG		CCGG	IFATRSTRLRAL
					GAGACCATCGACGACGCCGACGGCGAGTGG		TTCG	RNKVKLELARV
					CTTTTGCGCTTCGACGGTGCTTGCCGTGCAA		CGGT	GSFSLHHIDRQA
					ACCCAGGCCCAGGCGGCGCCGGCGCAGCTC		GGCC	NGHADRLANAG
					TCTTCAAGCCCAGCGGCCCCGTAGTGTGGA		GGCC	LDRRRTKLECSV
					CTTGTTCCCACTACGACCCGAGCACCACGG		GCTT	HPDGRGCTNTS
					CAACGAACAACACAGCAGAGTACACTGCGT		AAAC	VATAAPTAPAA
					TGCTGCTCGGGGCCAGGGCTGCCGCCGACC		CAGG	PLPPARPPATTA
					ACGGCGTCACGAAGCTGCGCGTCGAGGGCG		TAGA	APSHDDDHSVQ
					ACAGCACGCTGGTCATCCAGCAAGTGCGGG		CACG	GDIDDGEVYAA
					GCATCTTCGCGACGAGAAGCACGCGGCTCC		CACA	MCIGPDAVPHR
					GAGCACTGCGCAACAAGGTCAAGCTCGAGC		GTGT	RPRLRLRHLTDE
					TCGCGCGCGTGGGGAGCTTCTCGCTCCACC		CCTC	ESEEAGDVVER
					ACATCGACCGGCAGGCGAACGGCCACGCCG		AAGC	LAASLAAKIAD
					ACCGCCTCGCCAACGCCGGCCTCGACCGAC		GCCG	APDWETAEGYI

					GCCGCACCAAGCTCGAGTGCAGTGTGCACC		GCGC	TALPYALYDKL
					CCGATGGCAGGGGCTGCACCAACACCTCTG		CTCC	QPYSQAQPQPPS
					TGGCGACCGCAGCGCCTACTGCTCCTGCAG		GAGC	QQQQQQQQRPR
					CGCCCTTGCCCCCAGCCCGGCCTCCTGCAAC		ACAT	QQQQTRQRRQR
					TACAGCCGCTCCGTCTCATGATGACGACCA		GACA	RGKRGGGSQRR
					CAGCGTTCAAGGCGACATCGACGATGGCGA		GGTG	QRKTRRRRPPR
					AGTATACGCTGCAATGTGCATCGGCCCCGA		CAGA	VTRHHREHRIDE
					TGCGGTGCCGCACCGCCGACCGCGCCTCCG		CACC	ALDDLHAIESRR
					CCTACGGCACTTGACGGACGAGGAGTCCGA		CAGG	PQDRTAISKARR
					GGAGGCTGGCGACGTCGTGGAGAGGCTGGC		ATGC	RVGRIRSALDQH
					AGCGTCGCTGGCGGCGAAGATTGCAGACGC		ACGC	QLRHRFDTDEK
					GCCCGACTGGGAGACGGCGGAGGGCTACAT		GCTC	VCVDAILAGAR
					CACAGCCCTACCGTACGCTCTCTACGACAA		CGAG	ASQGATTAPPSA
					GCTGCAGCCGTACTCCCAGGCCCAGCCTCA		ACGC	TTDPPAPMDDS
					GCCTCCAAGCCAACAGCAGCAGCAGCAGCA		TGGC	RCPIPGDDLWRF
					GCAGCGACCACGGCAGCAGCAGCAAACAC		GCAT	FDSVNTPRRSFD
					GGCAGCGTCGACAGCGGAGAGGCAAGCGC		GTGC	AEAPDGAAFRE
					GGCGGTGGCAGCCAACGCAGGCAGCGGAA		TTAA	AMACLPAATRA
					GACGCGCCGCCGACGGCCCCCGCGGGTGAC		CCAC	QELLTEAPTVDE
					ACGCCACCACCGCGAGCACCGGATCGATGA		TGCG	VEDQIQHARASS
					AGCCCTCGACGACCTCCACGCGATCGAGAG		ACGG	SPGLDGVGYDIY
					CAGGCGCCCTCAAGACAGGACGGCCATCAG		AACC	KQFAAHLLPAL
					CAAGGCCAGGCGTCGAGTCGGCCGCATCCG		ATGG	HTAFVCCWNHK

					CTCCGCCCTTGACCAGCACCAATTGCGCCA		ACGC	RVPQSWKLGVV
					CCGCTTCGACACCGATGAGAAGGTGTGCGT		AGTC	RLLHKKGDRQD
					GGACGCCATCCTGGCTGGGGCGCGAGCGAG		CGTG	PANWRPICLQQT
					CCAAGGAGCGACAACGGCACCACCGTCGGC		GCCG	IYKLYAGILSRR
					AACCACGGACCCCCCAGCTCCGATGGACGA		GCCA	FVRWLDANARH
					CAGCAGGTGCCCCATTCCCGGCGACGATCT		TGAC	AEAQKGFRAMN
					ATGGAGGTTTTTCGACAGCGTTAACACGCC		GCCG	GCGEHNFLAAT
					ACGGCGGTCGTTCGACGCAGAGGCACCAGA		CACT	LVDHARRKRKE
					CGGAGCTGCCTTCCGCGAAGCGATGGCGTG		CAAG	LHVVWYDLAN
					CTTGCCCGCCGCCACGAGAGCCCAGGAACT		ATTA	AFGSVPHDLLW
					GCTGACGGAGGCGCCTACGGTGGACGAGGT		TTGA	ETLARQGVPPTF
					TGAAGACCAGATTCAGCACGCCCGCGCCAG		GCGT	VDCCRGIYSDA
					CTCCAGCCCCGGCCTCGACGGCGTTGGTTA		GCGC	AFTIGNAADGTT
					CGACATCTACAAGCAGTTCGCCGCCCACCT		TCCTC	APIRLRVGVFQG
					CCTGCCAGCACTGCACACAGCGTTCGTGTG		GCAT	CPLSPHLFTAAI
					CTGCTGGAACCACAAGCGGGTGCCTCAGAG		CGTC	APLLHALKRLPV
					CTGGAAGCTCGGCGTGGTGCGCCTTCTCCA		GGCC	TGVQLTGVDRP
					CAAGAAAGGCGATCGACAGGATCCTGCCAA		GACC	GAAAYADDLKT
					CTGGCGGCCGATCTGCTTGCAGCAGACCAT		AGCA	FSSSVDGIKRQH
					CTACAAGCTCTATGCCGGCATCCTGTCACGC		GGAC	ELVATFLRWTG
					CGGTTCGTGCGCTGGCTGGACGCAAACGCT		CGTG	MAANLSKCSAM
					CGTCACGCAGAGGCGCAGAAAGGATTCCGG		CTGA	SVQRDSRGVLK
					GCGATGAACGGCTGCGGGGAGCACAACTTC		GCTC	TGDLCLKLNDA

					CTCGCGGCCACGCTCGTCGACCACGCCCGC		CGCG	EIPALSMTASYA
					CGCAAGCGCAAGGAGCTGCACGTCGTGTGG		TGAA	YLGIGDGFDHV
					TACGACCTTGCCAACGCCTTCGGCAGTGTG		CCAG	RRRLELAPMMK
					CCACACGACTTGCTATGGGAAACGCTGGCG		ACCG	QLKHDATALMQ
					CGCCAAGGCGTACCGCCCACGTTCGTGGAC		TGCC	SGLASWQVVKA
					TGCTGCCGCGGGATCTACAGTGATGCAGCC		GTCA	VKVYLYPRIEYA
					TTCACCATCGGGAACGCCGCCGACGGCACC		CTCG	LRHLRPFKQQLE
					ACAGCGCCCATCAGGCTTCGCGTGGGAGTA		CCGG	AFDEHLRRGLR
					TTCCAGGGCTGCCCGCTCAGCCCCCACCTCT		CCCC	HLLRLPTNATSA
					TCACCGCCGCAATCGCCCCGCTGCTCCATGC		GCGC	FFSAPTSRGGLG
					GCTGAAGAGGCTGCCGGTCACTGGAGTGCA		TACG	LLPLTELHAALQ
					ACTCACCGGCGTGGACCGTCCCGGAGCTGC		GCCC	IAHGWQILNSPD
					TGCCTACGCGGACGACCTCAAGACGTTCAG		GACC	GATQRIAREQLR
					TAGCTCCGTGGACGGAATCAAGAGGCAGCA		TTCA	EIPDARHRLDTA
					CGAGCTCGTGGCGACCTTCCTACGATGGAC		GCTC	HWRNRDAELCE
					CGGCATGGCGGCGAACCTTTCCAAGTGCAG		TACA	LLLNSQLGQSSH
					CGCCATGTCGGTCCAGCGCGACAGCCGGGG		ACCA	APPKRRNCDIGS
					CGTGCTCAAGACAGGTGACCTATGCCTCAA		CACC	LWIDIRRQLGTL
					GCTCAACGACGCAGAGATCCCAGCGCTGAG		AAGA	GLKFETAPGRRS
					CATGACGGCCTCCTACGCGTACCTCGGCAT		AGAC	HQPARPAIAAFA
					CGGGGACGGCTTCGACCACGTGCGCCGCCG		GGTG	CRTTTSG (SEQ
					CCTCGAGCTCGCGCCAATGATGAAGCAGCT		GCGG	ID NO: 1450)
					CAAGCACGACGCGACGGCGCTGATGCAGTC		TGGT	

					GGGGCTGGCCTCATGGCAAGTGGTGAAGGC		CGAC	
					GGTGAAGGTATACCTCTACCCCCGCATAGA		CTGG	
					GTACGCCCTTCGCCACCTTCGCCCGTTCAAA		CCGT	
					CAGCAATTAGAAGCTTTCGACGAGCACCTC		GGCG	
					CGCCGCGGGCTCCGCCACCTGCTGCGACTA		TTGA	
					CCGACGAACGCCACCAGCGCCTTCTTCTCA		GGAG	
					GCGCCGACATCGCGCGGCGGGCTGGGCCTT		CAGG	
					CTGCCGCTGACGGAGCTGCACGCGGCGCTT		CGAG	
					CAGATCGCCCATGGGTGGCAGATACTCAAC		TGAC	
					TCCCCAGACGGCGCCACCCAGAGGATTGCC		GACG	
					CGCGAGCAGCTCCGGGAGATTCCAGACGCG		CGAG	
					CGCCACCGCCTGGACACTGCCCACTGGAGG		TAGC	
					AACCGCGACGCAGAGCTGTGCGAGCTGCTG		TCGG	
					CTCAACTCGCAGCTCGGACAGTCCAGCCAC		CACT	
					GCCCCACCCAAGCGGCGCAACTGCGACATC		GTCC	
					GGATCCCTTTGGATCGACATTCGGCGCCAG		CGGA	
					CTCGGCACGCTCGGGCTCAAATTCGAGACC		TCGC	
					GCGCCTGGCCGACGCAGTCACCAACCAGCC		CAAC	
					CGCCCAGCCATTGCAGCTTTCGCGTGCCGC		CACA	
					ACCACGACAAGTGGCTGACGCACCGTGACA		AGCG	
					TAGTGCGGCACGTGAAGATGCACATGAAGC		AGCC	
					TCCGACACTGGGCCAAGTGGGCGGCCATGC		AAGT	
					GCGACCAAGGCAAGACAGCTCGTGCACATG		ACGA	

					GTGGGGTTGGTAGTGGCTTCCTCACACGGC		CCAC	
					CGCGAGGCCTGTGGGAAGCCGACTACCGGT		ATCA	
					TCGCGGTGGCCGGCCGCTTAAACCAGGTAG		AGCT	
					ACACGCACAGTGTCCTCAAGCGCCGGCGCC		ACAC	
					TCCGAGCACATGACAGGTGCAGACACCCAG		CTCG	
					GATGCACGCGCTCCGAGACGCTGGCGCATG		AGCG	
					TGCTTAACCACTGCGACGGAACCATGGACG		CCAA	
					CAGTCCGTGGCCGGCCATGACGCCGCACTC		GGAT	
					AAGATTATTGAGCGTGCGCTCCTCGCATCGT		GGAA	
					CGGCCGACCAGCAGGACCGTGCTGAGCTCC		GGTA	
					GCGTGAACCAGACCGTGCCGTCACTCGCCG		CACC	
					GCCCCGCGCTACGGCCCGACCTTCAGCTCT		TCTC	
					ACAACCACACCAAGAAGACGGTGGCGGTG		GGCA	
					GTCGACCTGGCCGTGGCGTTGAGGAGCAGG		CTCG	
					CGAGTGACGACGCGAGTAGCTCGGCACTGT		TGTA	
					CCCGGATCGCCAACCACAAGCGAGCCAAGT		CGGG	
					ACGACCACATCAAGCTACACCTCGAGCGCC		TCGC	
					AAGGATGGAAGGTACACCTCTCGGCACTCG		TTGG	
					TGTACGGGTCGCTTGGGGCGGTCGCTAGTG		GGCG	
					GCAACTACCAGGTGTACACCACACACCTGG		GTCG	
					GGCTACTCAAGCGCGATGCAAAGCGGCTGG		CTAG	
					ACCGGCAGCTGTCTGCCTAATGCATCCAGT		TGGC	
					CCAGCCGCCGCATCTAGAATCTACACTGCA		AACT	

					GCCAGCACCGGACTCGCCAACACCAGGCGA GGCCCAGCCAAGGACCAAGAGGCAGCCGG GCGACGGAGACCGGGGGGACTCCGTCCCAG ACAAGCCGCCGCTAGGCGGAAACCAGGCCC AAGACGGCCGACAGGGCCCCACCCAGGTAG GGAACCGCCCTAGAAACCCATTTCGGTGGT CGACTCGCAAGCCTTACCTATATTTTAGACG TAGCGACCAAATTACAAATTTGGTAACGAG TAAGCCAAATGGTAATACACAAAACTTT (SEQ ID NO: 1082)		ACCA GGTG TACA CCAC ACAC CTGG GGCT ACTC AAGC GCGA TGCA AAGC GGCT GGAC CGGC AGCT GTCT GCCT AATG CATC CAGT CCAG CCGC	

							CGCA TCTA GAAT CTAC ACTG CAGC CAGC ACCG GACT CGCC AACA CCAG GCGA GGCC CAGC CAAG GACC AAGA GGCA GCCG GGCG ACGG AGAC	

							CGGG GGGA CTCC GTCC CAGA CAAG CCGC CGCT AGGC GGAA ACCA GGCC CAAG ACGG CCGA CAGG GCCC CACC CAGG TAGG GAAC CGCC CTAG	

							AAAC CCAT TTCG GTGG TCGA CTCG CAAG CCTT ACCT ATAT TTTA GACG TAGC GACC AAAT TACA AATT TGGT AACG AGTA AGCC AAAT GGTA	

							ATAC ACAA AACT TT (SEQ ID NO: 1328)	
NeSL	Utopia-5_PR	.		Phytophthora ramorum	CCAGGCATCACCCCCGCCGCCCCCAAGCGC TCCGCCGCCGAGCGCAGCTGCGGCTGCTGC CGCCCCACTGGTGACGGCAGTCGCAGCTGC CCCTGCCACCTCTCCTGTTGGCCGCGCCGCC GTGGAGCCCCGCGCGCGCGCCGAGCCGCCC CAAGAACAAGCACCCCCAGCTAGCGCGCGC GCGGAGCCCATGTCGGGCGACGTGGTCAGC AGCGACGGCAGCAGCAGGACGACGGACGT GAGCGGCGACGGCGACGACGGCGCCGACG GCGCAGGCAGCAGCGATGCAGCAGGCGAT GTCGGCGTCGTCGCCATGGATGTTGACCAA GGTGCCCGCCGACAGCGGCCGCCATGGCAG CGCGTCGGCGGCAAGCGTCGACGTCTCAAC GACGTCGACGACGAAGACACCCGCGAGCTG GCAGAGCTCCTGCTCGAGGAGGAGGACGAG GTTGGGGCCCAAGCACCAGCGCTCAGGCTA	CCAGGC ATCACC CCCGCC GCCCCC AAGCGC TCCGCC GCCGAG CGCAGC TGCGGC TGCTGC CGCCCC ACTGGT GACGGC AGTCGC AGCTGC CCCTGC	TGAT GCCC GCCA ACTA CAAG GTGC TTACT GAGC ACCT TGGG CTGC TCAA GCGT GATG CGAA GCGG	MSGDVVSSDGS SRTTDVSGDGD DGADGAGSSDA AGDVGVVAMD VDQGARRQRPP WQRVGGKRRRL NDVDDEDTREL AELLLEEEDEVG AQAPALRLFAA SAHPASVLSVY AHNAQRFVCTL CAYTAASFASL KRHRDSRHRRV SFVDKFSAGCA CGTPFGSRLAAA RHAQACASLSIP

					TTCGCGGCCAGCGCTCACCCGGCCTCCGTCC	CACCTC	CTGG	RTVTAPAAAGD
					TCTCCGTGTACGCACACAATGCGCAGCGCT	TCCTGT	ACCG	LSPTATGANAT
					TCGTGTGCACGCTGTGCGCGTACACGGCCG	TGGCCG	GCAG	ASAAATSPDLPR
					CCAGCTTCGCGTCGCTCAAGCGACACCGGG	CGCCGC	CTGT	PASPELAASPPQ
					ACTCACGACACCGGCGCGTGTCCTTCGTGG	CGTGGA	CGGT	TSPFDVAIQADA
					ACAAGTTCTCGGCGGGCTGCGCGTGCGGCA	GCCCCG	GGCG	AEQTAWTRWDP
					CGCCCTTCGGCTCGAGGCTGGCAGCGGCGA	CGCGCG	TGCA	PLTRAAVAARV
					GACATGCGCAAGCGTGCGCCAGCCTCTCGA	CGCCGA	TCCA	ASRLAVVPAPR
					TCCCACGCACGGTAACAGCTCCGGCGGCAG	GCCGCC	GTCC	WGPPLSRTLVAS
					CAGGAGATTTGAGCCCCACTGCCACGGGAG	CCAAGA	AGCC	RIAARLDAQTSR
					CCAACGCCACCGCCTCGGCGGCCGCAACTT	ACAAGC	GCCG	WGPPLPPAMVA
					CGCCCGATTTGCCCCGCCCAGCTTCCCCGGA	ACCCCC	CATC	SRIASRLAAMPA
					GCTCGCTGCGTCCCCCCCGCAGACGAGCCC	AGCTAG	TGGA	PRWGPPLLRTVI
					CTTTGACGTTGCCATACAGGCAGACGCTGC	CGCGCG	ACCT	ASRIADRLLPPE
					AGAGCAGACAGCGTGGACGCGCTGGGACCC	CGCGGA	GCAC	LAADEETKDDD
					TCCGCTCACCCGCGCTGCGGTGGCAGCCCG	GCCC	TGCG	VHMDNAASVD
					TGTCGCCTCCCGCCTCGCCGTGGTTCCCGCG	(SEQ ID	CGCA	VDEESEVADVV
					CCGCGATGGGGGCCGCCGCTGTCAAGGACG	NO: 1206)	GCAT	MTDHDGEWLLR
					CTGGTGGCCTCCAGGATTGCAGCACGCCTG		CGAG	FDGACRANPGP
					GACGCGCAGACGTCACGATGGGGACCGCCC		CACG	GGAGAALFKPS
					CTGCCACCGGCGATGGTGGCGTCCCGGATT		GCAG	GPVVWACSRY
					GCGTCGAGACTCGCCGCCATGCCGGCACCC		CACC	MPSSSATNNTAE
					CGCTGGGGCCCGCCGCTGCTGCGGACCGTG		AGGG	YTALLLGARAA

					ATCGCATCCCGCATCGCCGACCGACTGCTC		CCAA	ADHGATHLRVE
					CCACCGGAGTTGGCGGCCGACGAGGAGACG		GCGC	GDSTLVIQQVRG
					AAGGATGACGACGTCCACATGGACAACGCG		CAAG	IFAARSTRMRAL
					GCCAGCGTGGACGTCGACGAGGAGTCCGAG		GGGC	RNQVQSELARV
					GTCGCCGACGTGGTCATGACCGACCACGAC		AGTC	GSFSLHHIDRQD
					GGCGAATGGCTGTTGCGCTTCGACGGCGCC		GGGC	NAHADRLANRA
					TGTCGTGCCAACCCTGGCCCAGGCGGCGCC		GGCG	LDLRRTVIECGI
					GGAGCCGCTCTGTTCAAGCCCAGCGGCCCC		GAGA	HCDGVGCTATT
					GTAGTGTGGGCATGCTCGCGCTACATGCCG		CCGG	TEVQSSSAPEIPT
					AGCAGCAGCGCGACCAACAACACGGCCGA		GGGG	RPVADDHDEHE
					GTACACGGCGCTGCTGCTCGGCGCTAGGGC		ACTC	VVDVVDVCGVC
					CGCTGCCGACCATGGAGCCACGCACCTGCG		CGCC	GVCGDRGTCGV
					AGTGGAAGGCGACAGTACGCTTGTGATCCA		ACAG	CDVSGDIDDGE
					GCAGGTGCGTGGCATCTTCGCAGCGCGCAG		AGCG	VYAAMRTGPDA
					CACCCGGATGAGGGCGCTGCGCAACCAGGT		GCCG	VPARRPRLRLRK
					GCAGTCCGAGCTCGCGCGGGTGGGCAGCTT		CCGC	LTDEEQEEAGTL
					CTCGCTCCACCACATCGACAGGCAAGACAA		TAAG	AERLGATLAAKI
					CGCACACGCGGACCGCCTTGCCAATCGCGC		CGGA	ADARDWESAEG
					GCTGGACCTGCGCCGGACCGTGATAGAGTG		CATC	YITALPYLLYDK
					CGGCATCCACTGTGATGGTGTTGGCTGCAC		GGGC	LLPYSQGPARSL
					CGCCACCACGACGGAGGTACAGTCATCATC		CCGT	PVRQHQRQQQQ
					AGCGCCCGAGATACCGACCCGGCCGGTTGC		AGCG	PDGQFQRPTQSR
					TGACGACCATGATGAACATGAAGTTGTTGA		GCCG	SAARRQRRQRH

					TGTTGTTGATGTTTGTGGAGTTTGTGGAGTT		ACAG	RARRRPPRVTR
					TGTGGCGACCGTGGAACTTGTGGAGTCTGT		GGCC	HHREHRLDEAL
					GACGTTAGTGGAGATATCGATGATGGCGAG		ACAC	DDLHAVERATP
					GTCTACGCAGCCATGCGTACGGGACCTGAC		CCAT	SDRRSIRRARRR
					GCTGTTCCTGCGCGCCGACCACGCCTCCGCT		GTAG	VGRVNSAVEQQ
					TGCGGAAACTCACTGACGAAGAGCAAGAA		GGAA	RLRHHFDTNEK
					GAGGCAGGCACGCTGGCGGAACGGCTTGGA		CCGC	GCVEILLAKARA
					GCGACGCTGGCGGCCAAGATTGCAGACGCA		CCTCT	QRSTTVARTAV
					CGCGACTGGGAGTCGGCGGAGGGATACATC		AAAC	GEPNSGAAEDD
					ACAGCCCTCCCTTATCTACTCTACGACAAAT		CCGC	GTCPIPSERLHR
					TGTTGCCGTACTCCCAGGGCCCAGCTCGCTC		CCGG	HFTEVNTPGSSF
					CTTGCCGGTCCGCCAGCACCAACGGCAGCA		TACA	DAMAPVGAPFR
					GCAGCAGCCGGACGGACAATTCCAGCGACC		TTAT	AALAHLPAATE
					CACCCAGTCCCGATCTGCGGCTCGTCGACA		GGTC	ASELLTEAPTPD
					GCGACGCCAGCGCCACCGGGCACGTCGCCG		CGAC	EIEDQLQRAKGT
					ACCGCCACGCGTGACTCGTCACCACCGTGA		ACCT	TSPGLDGVGYD
					ACACCGACTAGACGAAGCGCTTGACGACCT		ATGA	VYKAFSTQLLPV
					CCACGCGGTGGAGAGAGCTACGCCCAGCGA		GGTG	LHAAFQCCWQH
					CCGCCGCTCCATTCGCCGAGCCCGCCGCCG		CAAC	HRVPQSRKQGIV
					GGTCGGCCGCGTTAATTCGGCCGTCGAACA		CGGT	RLLYKKGPRED
					ACAACGGCTCCGGCACCACTTCGATACGAA		ACAC	PANWRPICLQQ
					CGAGAAGGGCTGCGTGGAGATCCTACTGGC		AAGT	VIYKTYAGVLA
					CAAGGCGCGTGCACAGAGGTCGACGACGGT		TACA	RRFTRWLAANG

					GGCCCGGACGGCAGTTGGAGAGCCGAACA		CACC	RHADAQKGFRT
					GCGGTGCGGCCGAGGACGACGGCACGTGCC		ACAT	VNGCGEHNFLA
					CCATCCCGAGCGAGCGCCTTCACCGCCATTT		AGCG	STLIDHARRSRR
					CACCGAGGTGAACACGCCCGGGTCCAGCTT		ACTA	ELHMVWYDLK
					TGATGCGATGGCCCCCGTGGGTGCACCGTT		CCAG	NAFGSVPQELL
					CCGTGCGGCACTTGCGCACCTCCCCGCGGC		GTAT	WEVLQRMGVPP
					CACAGAGGCCTCGGAGCTGCTGACAGAGGC		TTACT	AFVEVCQGLYQ
					GCCGACGCCCGATGAGATCGAGGATCAGCT		ACCT	DAAFTVGNAAD
					GCAGCGAGCCAAGGGAACGACCAGCCCTG		GGAA	GPTDLVRQLVG
					GCCTGGACGGAGTGGGCTACGACGTCTACA		GCCA	VFQGCPLSPHLF
					AGGCCTTCTCCACCCAGTTACTGCCGGTGCT		AGGA	TAAISPLLHALD
					GCACGCGGCATTCCAGTGCTGCTGGCAGCA		TTAA	RLKDTGVRLSA
					CCACCGTGTGCCGCAGAGCAGGAAGCAAGG		CCGG	DDRPGASAYAD
					CATCGTGAGGCTCCTGTACAAAAAGGGCCC		TCGG	DLKIFSGTADGV
					GCGCGAAGACCCCGCGAACTGGAGACCTAT		TAAT	KRQHALVADFL
					CTGCCTCCAGCAAGTGATCTACAAGACCTA		ACAC	RWTGMVANPN
					CGCCGGGGTCCTCGCGCGGAGGTTTACGCG		ATAA	KCCTMSVQRDG
					CTGGTTGGCTGCAAACGGCCGCCATGCCGA		CTTT	RGVLKACDLEL
					CGCCCAGAAGGGGTTCCGCACTGTAAATGG		(SEQ	QLDGARIPSLIM
					CTGCGGTGAGCACAATTTTCTCGCGTCGAC		ID NO:	NASYAYLGTGD
					GCTGATCGATCATGCGCGCCGCAGCCGCAG		1329)	GFDHVRRRIELV
					AGAGCTCCACATGGTCTGGTACGACCTGAA			PALMQLKDDAT
					GAACGCCTTCGGCAGCGTGCCCCAGGAGTT			ALLQSGLAPCQ

					GCTGTGGGAGGTGCTGCAGCGCATGGGTGT			VVKAVKTYLFP
					CCCGCCAGCCTTTGTGGAGGTCTGCCAGGG			RVEYALRHLRPF
					GCTGTACCAGGATGCTGCCTTCACCGTCGG			QQQLEGFNRHL
					CAACGCTGCTGACGGCCCGACCGATCTCGT			VCGLRHLLRLP
					GCGGCAGCTCGTGGGCGTGTTTCAAGGATG			VSATTSFFFVPV
					CCCGCTGAGCCCCCACCTGTTCACTGCGGC			SRGGLGLLPLTE
					GATCAGCCCGCTCCTCCACGCGCTGGACCG			LHAAPADRRPA
					ACTCAAGGACACGGGCGTGCGGCTGTCAGC			PPPRPRPLEGAG
					GGACGACCGTCCCGGTGCCTCGGCCTACGC			GENMRAADQLA
					CGACGACCTCAAGATCTTCAGTGGCACCGC			ARDVGPRPTQA
					GGACGGCGTTAAGCGACAGCACGCGCTCGT			PQRRHRLVVGR
					GGCCGACTTCCTGCGCTGGACCGGCATGGT			RPAPPPRTRPQA
					GGCGAACCCCAACAAGTGCTGCACCATGTC			RDRAGVRGDRH
					GGTCCAGCGCGACGGACGCGGCGTGCTGAA			RGGDAAASRAA
					GGCCTGCGACCTGGAGCTGCAGCTGGACGG			PREVAGPPHGPA
					AGCCCGGATCCCCTCGCTGATCATGAACGC			AGEAAHEEQAL
					GTCGTACGCGTACCTGGGGACAGGCGACGG			AAVGRDEGPRQ
					CTTTGACCACGTGCGCCGCCGCATCGAGCT			DRPHPWWCRER
					CGTCCCGGCGCTCATGCAGCTCAAGGACGA			LPHAASRPMGD
					TGCGACGGCACTGCTCCAGTCGGGGCTGGC			RLPLCSGGSAQP
					GCCGTGTCAGGTGGTGAAGGCGGTGAAGAC			AGHAQRAEAPA
					TTACCTCTTCCCGCGGGTGGAGTACGCACTC			PPRARPVSTAGL
					CGCCACCTCCGCCCCTTTCAGCAGCAACTG			LPCRDTGTRAES

					GAAGGCTTTAATCGCCACCTCGTTTGCGGCC TTCGCCACCTGCTCCGACTTCCGGTCAGCGC CACGACCAGCTTCTTCTTCGTGCCGGTGTCG CGCGGCGGGCTAGGCCTGTTGCCACTGACG GAGCTGCACGCAGCTCCGGCTGATCGCCGA CCGGCGCCACCGCCTAGACCCCGACCACTG GAAGGAGCGGGAGGAGAAAATATGCGAGC TGCGGATCAACTCGCAGCTCGGGACGTCGG CCCACGCCCCACCCAAGCGCCGCAACGGCG ACATCGGCTCGTTGTGGGTCGACGTCCAGC GCCACCTCCGAGGACTCGGCCTCAAGCTAG AGACCGCGCCGGCGTGCGCGGAGACCGGCA CCGCGGAGGCGACGCTGCAGCTTCGCGTGC CGCACCACGAGAAGTGGCTGGACCACCGCA CGGTCCTGCGGCAGGTGAAGCTGCACATGA AGAACAAGCATTGGCAGCGGTGGGCCGGG ATGAAGGACCAAGGCAAGACCGCCCGCACC CATGGTGGTGCAGGGAGCGGCTTCCTCACG CGGCCTCGCGGCCTATGGGAGACCGACTAC CGCTTTGCAGTGGCGGGTCGGCTCAACCAG CTGGACACGCACAGCGTGCTGAAGCGCCGG CGCCTCCGCGCGCACGACCAGTGTCGACAG CCGGGCTGCTCCCGTGCCGAGACACTGGCA			LRRHHGRGPRPP RRCAQDHRARA HRVVRAARTAP SSGSTRPCPRSP APRCGPTSSSST TPRRRWRWSTW PWRSRSRRATTP RALRWRASPRT SEPSTPASSGTSS AKGGRSTSRRSC TARWAR (SEQ ID NO: 1451)

					CACGTGCTGAATCACTGCGCCGGCACCATG GACGCGGTCCGCGGCCGCCACGACGATGCG CTCAAGACCATCGAGCGCGCGCTCATCGCG TCGTCCGGGCGGCCAGGACCGCGCCGAGCT CCGGGTCAACCAGACCGTGCCCTCGCTCGC CGGCCCCGCGCTGCGGCCCGACCTCCAGCT CTTCAACCACACCAAGAAGACGGTGGCGGT GGTCGACCTGGCCGTGGCGTTCGAGGAGCA GGCGAGCGACGACACCGAGAGCTCTGCGCT GGCGCGCATCGCCGCGCACAAGCGAGCCAA GTACGCCGGCATCAAGCGGCACCTCGAGCG CCAAGGGTGGAAGGTCCACCTCTCGGCGCT CGTGTACGGCTCGTTGGGCGCGGTGATGCC CGCCAACTACAAGGTGCTTACTGAGCACCT TGGGCTGCTCAAGCGTGATGCGAAGCGGCT GGACCGGCAGCTGTCGGTGGCGTGCATCCA GTCCAGCCGCCGCATCTGGAACCTGCACTG CGCGCAGCATCGAGCACGGCAGCACCAGGG CCAAGCGCCAAGGGGCAGTCGGGCGGCGG AGACCGGGGGGACTCCGCCACAGAGCGGCC GCCGCTAAGCGGACATCGGGCCCGTAGCGG CCGACAGGGCCACACCCATGTAGGGAACCG CCCTCTAAACCCGCCCGGTACATTATGGTCC			

					GACACCTATGAGGTGCAACCGGTACACAAG TTACACACCACATAGCGACTACCAGGTATT TACTACCTGGAAGCCAAGGATTAACCGGTC GGTAATACACATAACTTT (SEQ ID NO: 1083)			
NeSL	YURE Ci	.		Ciona intestinalis	ATCAACCCACTACTACACCCTCTACAAAGA ACCCACTACAGCAAGATCTTGGCGACCAAC TACACCTGCAGCCTCCCGTCGACTCTTCACC ACCTGCGCACCTCCCTCCGACTCCAAGCGG CAGACAGTGGCACCACCAGCACTCCAAGGG TCCACAGCCAGAAGAGTGCACCCCCCACTG CTAGGAAGAGCTAGGCCATTGCCCGCTGTG GGCCTTGTGCTGGCTGTCTGCCGCTGATTGG CACGTTAGGGAGTGTGCGCTGAGCACGATA GGGAGGCTGCTGAAGCCGAAGAACCCGCGC CGGGGACTTTAGGACTTTAACTAAAATCTG TAACACTAAAATTGGAATACTGGATTACTA TTGGATTACTACTGGACAATAACCTGATCTA CAACAACCAATTGGATTACAGAACCAAATT ACAATCTCAACTACAAACAAAAAGTGAGTC CCCGGCCGGGTTTCTATATTACCCAATTACT GTTTTTACAATTTTAAAATTTTATAGTATTTT ACTAATTTTTCCGCCCCGCTAGCACTTAAAT TGGCACCCCCCCCCCCCSTCCAAAAAAAAA	ATCAAC CCACTA CTACAC CCTCTA CAAAGA ACCCAC TACAGC AAGATC TTGGCG ACCAAC TACACC TGCAGC CTCCCG TCGACT CTTCAC CACCTG CGCACC TCCCTC CGACTC	TAGC CAAA AGAT TTGG TGTTC TTGG CGAG CTGC AGTC CCGG CAGA GTAG GCGA GATC GTGA GCTC TACA CCGC TTCA	MATSSSSVSSGN VQTEVRCVYHG KGDLLLECPVA HCPSIHPTVATIT KHLKKHHTPQF EQITTKNLTITYT CSQCSFSTTGLT QHHISKHYKTC KGVGAVQEGNK GRFCCPACGTR WALLCKARHHF NNVHFEYDTPPI AAFSGTPYKLK KRKFTIINKALT YSCPLPLNQLLC PLWSCSLTILNK PLSSVQQETAHG DGSQGQSYVPT QLRQVLRARCH

					TAGATAACCCTCCACCACCTAAACCCCGGT	CAAGCG	CGAT	CGNPPIGKGHW
					CACCCCCTCAGAACTAATTGGAACCCGGAT	GCAGAC	TTGA	ASCQGKRPLSSP
					CTTTTCCCTCTATCAAATTTCCGGTTCAGTTT	AGTGGC	CGAG	KGGRSSPTPPAN
					GGCTTAATTACCTAAAACCCGTCTTTATTTT	ACCACC	TTCTT	LTLHFLNYLPFQ
					ATCCCTTTATTGCCCCTTAAACCAAATTATT	AGCACT	CTTG	LPSQSSSPQSSTL
					CAAATTGAAATCCGGCCAAATTGCCTTGCT	CCAAGG	AGTC	DPTACKARVPIP
					GAACCACCATTTTTGTTTTTCTTATAAAAGT	GTCCAC	GATC	SFLRGDCEVTFF
					AAATTTTTCAACAACCCAAAATCATAAAAT	AGCCAG	CCCA	IIPSVNFYRPYLS
					TAAAACTGTTATCTGATTAGTCTAAATTCTG	AAGAGT	ACGG	YPLRMFWRNRT
					TTATTGACAAAAACCTCCATACCAACCTTA	GCACCC	CAAC	SSGHCSLHRVV
					ATCTTATCATTATCTTCACCCAGACTGCTCA	CCCACT	GCAC	RGFVRERLVPHP
					ATCTAGTTCTTTTTTCAATCAATAAGGAGGT	GCTAGG	TCCA	RSKSPARTPLEF
					GTTTTCTTTGGTCTGTTGTTTTATAGTTTACC	AAGAGC	AATT	LCEFRLAGVFPD
					GTTATAATAGTCTGTTGTCCTATAGTTTTCC	TAGGCC	TAAC	PGKVASLRPVPA
					GTGGTGTATTTCTGGTCTACTATAGTTGGTG	ATTGCC	GACC	PLTLCLSPPVAG
					CCTCTAAAACATTATTATACCGAAGAGTGC	CGCTGT	GCAT	PMISCEDHSAPP
					TAGCGATTTCTATTATTAACTCACCCACCAG	GGGCCT	GGCG	SVRSSSPIPNSPA
					TACTGCTGTGCATCCACGCAGCACCACGCC	TGTGCT	ACGC	SVSSVEAHLSDL
					TGCCATCACGTACAGTGCGCTTGCTGCCTTG	GGCTGT	CACT	LDKVSSGELRPL
					TGTTTTGTTTTGCAGACACCACTGGACGATA	CTGCCG	GATC	SPTLPSSGFFGPL
					GATGGCGACTTCGTCTTCCTCCGTTTCATCG	CTGATT	TCCC	LPPTPPPRPTPSA
					GGGAATGTTCAGACCGAGGTGCGTTGCGTC	GGCACG	AGTC	EKASPSGLSYLP
					TACCACGGAAAAGGAGACTTGTTGCTGGAA	TTAGGG	ATCC	CREVKIASIARPS

					TGCCCAGTTGCCCATTGCCCTAGCATACACC	AGTGTG	GTGC	PASQRVGCDAD
					CCACAGTTGCCACAATCACAAAACACTTAA	CGCTGA	TCCG	RTGPSLNPNYQ
					AAAAACATCACACACCTCAATTCGAACAGA	GCACGA	AGGG	QTSPPSTPSFSPI
					TAACCACTAAAAATCTCACAATAACATATA	TAGGGA	CACT	VRPPKFPRSGAK
					CATGCAGTCAATGTAGTTTCTCCACCACCGG	GGCTGC	TAAG	VNSKSKPPGVRP
					TCTCACACAACACCACATATCCAAACACTA	TGAAGC	GGGA	RRAKPIEPGTES
					CAAAACATGCAAAGGCGTGGGCGCTGTTCA	CGAAGA	ACCG	ASPVDVDTISSS
					GGAAGGCAACAAGGGTCGATTCTGCTGCCC	ACCCGC	ATTC	VQEPCTPENRTP
					TGCATGCGGTACCAGGTGGGCTCTTTTGTGT	GCCGGG	ACTG	EFFYERKWLVSI
					AAGGCACGGCACCATTTTAATAATGTCCAC	GACTTT	ATGC	LNIHEREGSNFF
					TTCGAGTATGACACCCCTCCAATTGCCGCTT	AGGACT	TGCA	QFNRDLEYWTQ
					TCTCTGGCACACCCTATAAACTAAAAAAAC	TTAACT	AGTT	LLSGSQKGGRA
					GGAAATTTACAATTATTAATAAGGCCCTAA	AAAATC	GGCG	KRASYNRGAAN
					CATATTCATGCCCTCTCCCCCTCAACCAACT	TGTAAC	GGGG	QAMKNRDSGRK
					ACTTTGCCCTTTGTGGTCGTGCTCCCTCACG	ACTAAA	CGCA	DFDPRPVAGHSS
					ATCTTAAATAAACCTCTTTCAAGCGTGCAGC	ATTGGA	CTCA	GGGTELGSRPR
					AGGAAACGGCACATGGCGACGGCTCACAA	ATACTG	CCAA	YPKGARLRADF
					GGCCAGAGTTATGTTCCAACTCAGTTGCGG	GATTAC	TGGA	WRDMKGTVRK
					CAGGTGCTCAGAGCCCGTTGCCACTGCGGA	TATTGG	AGAT	LLDGSNGERRC
					AACCCTCCAATTGGAAAGGGCCACTGGGCA	ATTACT	GAGG	GIPLDIIERKFRQ
					TCTTGCCAAGGGAAGCGACCACTCTCTTCA	ACTGGA	GAGC	VSMPGWIDHRR
					CCTAAGGGTGGGCGCAGTTCGCCCACTCCA	CAATAA	CCAA	YAAGASPSLVT
					CCTGCTAATCTAACCCTCCACTTCCTTAACT	CCTGAT	CGAT	QAETDVAITSEE

					ATCTCCCCTTCCAACTTCCCTCACAAAGCTC	CTACAA	GCCC	VEAVLSGLNVQ
					CTCCCCCCAAAGTAGCACCCTCGACCCTAC	CAACCA	CGAA	SAPGSDGLSYRF
					GGCCTGTAAAGCTCGAGTCCCCATCCCCTCT	ATTGGA	CTAC	WKGLDPSGRLL
					TTCCTACGCGGCGACTGTGAAGTCACCTTTT	TTACAG	CGCA	SCLFEIVRRHGRI
					TCATCATCCCCTCCGTTAATTTTTATAGACC	AACCAA	CCGT	PGAWPTCSVILL
					CTACCTCTCCTACCCCCTCAGGATGTTCTGG	ATTACA	CCCC	CKDAQGDVQD
					CGTAACCGAACATCCTCCGGTCACTGTTCCC	ATCTCA	AGCA	VGNWRPITICRT
					TACACCGGGTCGTTCGAGGCTTTGTCCGAG	ACTACA	CCGT	LYKLYAAVIAR
					AGCGACTCGTCCCCCATCCTCGGTCGAAAT	AACAAA	GTCC	RIQTWAKQGGV
					CGCCGGCGAGGACTCCCCTTGAATTCCTCTG	AAGTGA	TTGT	LSRLQKGFMPV
					CGAGTTCCGGCTCGCGGGGGTCTTCCCCGA	GTCCCC	GAGT	EGVFEHVFMLD
					TCCTGGGAAGGTCGCGTCCCTTCGACCCGT	GGCCGG	AATT	TVLSDAKLRRK
					ACCAGCTCCCCTTACGCTTTGCCTTTCTCCA	GTTTCT	ATTTT	NLLAVFLDVRN
					CCAGTCGCTGGTCCAATGATTTCCTGTGAGG	ATATTA	CACT	AFGSVRHECLL
					ACCACAGCGCCCCCCCTTCCGTTCGCTCGAG	CCCAAT	AATC	KVLRHFDAPHY
					CTCCCCCATTCCTAACTCCCCAGCCTCCGTA	TACTGT	CAGG	LVELVRDIYTGA
					AGCTCCGTGGAGGCGCACTTGTCTGACTTGT	TTTTAC	GACG	TCRVRSSVGETG
					TAGACAAAGTGTCTTCCGGAGAACTGCGGC	AATTTT	GGGC	DIPCDRGVRQG
					CGCTGTCCCCCACCCTGCCTTCTTCAGGCTT	AAAATT	CATG	YPLSGILFNLVT
					TTTTGGCCCTCTGTTGCCACCCACTCCCCCT	TTATAG	AACT	EVLIPGLSAGND
					CCCCGACCTACCCCCTCGGCCGAAAAGGCA	TATTTT	GAAC	GYRMACLGGKL
					TCTCCTTCAGGCCTTTCCTACCTACCCTGCA	ACTAAT	TATG	TQVLAYADDLV
					GAGAAGTAAAAATCGCGTCCATTGCACGGC	TTTTCC	TCTG	VVTENRDQMLR

					CATCCCCTGCCTCACAGAGGGTCGGCTGCG	GCCCCG	CACG	QLGVCEEFGRW
					ACGCGGACCGAACAGGGCCTTCGCTGAACC	CTAGCA	CTGC	AGLAFNQRKCG
					CTAATTATCAGCAAACCTCCCCCCCGTCTAC	CTTAAA	CCCG	LIGWRTLRGGR
					TCCGAGTTTCTCCCCCATAGTTAGACCCCCT	TTGGCA	CCTA	RVALEDPLLLN
					AAATTCCCCCGGTCGGGCGCCAAAGTAAAT	CCCCCC	GCCT	GVEIPLLRPGEH
					TCCAAGTCCAAACCCCCCGGCGTCCGACCA	CCCCCC	CGGC	YKYLGAMTGV
					CGCCGTGCCAAGCCAATTGAGCCAGGGACC	STCCAA	CACA	MSVPRTGSQLIK
					GAGTCGGCATCACCCGTTGATGTCGACACT	AAAAAA	AATA	DFRARLQRLFTS
					ATTAGTAGTTCGGTTCAGGAACCTTGTACTC	ATAGAT	AATA	FLTPHQKLIALK
					CAGAAAACAGAACCCCGGAATTTTTTTATG	AACCCT	GTCA	RFLLPSLSFHLR
					AGCGAAAGTGGCTCGTGAGTATTTTAAATA	CCACCA	AGCC	VRPIARSELIALD
					TTCACGAGCGGGAGGGATCCAATTTTTTTCA	CCTAAA	GGGG	RRVRECLRVAF
					ATTTAACCGGGATCTGGAGTATTGGACGCA	CCCCGG	CGTA	RLTKPSCQAVFH
					GTTGCTTTCAGGCTCCCAGAAGGGAGGACG	TCACCC	CTAC	TPTDMDGLGVP
					CGCGAAGCGTGCGTCCTACAATCGTGGTGC	CCTCAG	CAAT	SVCSESSILTIAQ
					GGCGAACCAAGCGATGAAAAACCGCGATTC	AACTAA	TGAT	GFKVLTSPDGTV
					CGGCCGTAAGGATTTCGATCCCCGGCCGGT	TTGGAA	TCGG	SATASARVKLY
					AGCAGGCCACAGTAGTGGAGGTGGCACCGA	CCCGGA	ATTT	AAKFGGLTEAG
					GCTCGGTTCGCGGCCACGATACCCTAAGGG	TCTTTTC	GGCG	PSDWARYLSGD
					CGCACGTCTTCGTGCCGACTTCTGGAGGGA	CCTCTA	GGGC	DVNGNSTRKPG
					CATGAAAGGGACGGTCCGTAAGTTACTTGA	TCAAAT	CCAT	ANLPSGLWTRV
					TGGTTCGAACGGTGAGCGTCGCTGCGGGAT	TTCCGG	CACG	RCASRQLGAVW
					TCCCCTCGATATCATAGAGCGGAAATTCCG	TTCAGT	CGCC	RVCPENGITVRV

					GCAAGTATCCATGCCCGGCTGGATAGACCA	TTGGCT	TCCG	RNSVITSRDRRK
					CAGACGGTACGCGGCCGGCGCGTCCCCGTC	TAATTA	CCCC	LIRSFHDCSNQQ
					CCTAGTAACCCAAGCTGAAACGGATGTTGC	CCTAAA	CACA	WKEQWMQHPN
					CATCACCAGTGAGGAAGTTGAGGCGGTCCT	ACCCGT	CACA	QEKTAAAHMA
					CAGTGGCCTAAACGTTCAGTCCGCTCCGGG	CTTTAT	AAAA	YADANRWVKQ
					CTCGGATGGCCTCTCCTATCGATTTTGGAAG	TTTATC	CACT	PSVMEPHTYFFA
					GGCCTCGATCCTAGTGGTCGATTGTTGTCAT	CCTTTA	TTTA	LRARLNLLPTRV
					GCCTTTTCGAAATCGTTAGGAGGCATGGTC	TTGCCC	AACT	SRAIYSRDQHPD
					GTATTCCGGGCGCGTGGCCGACTTGTAGCG	CTTAAA	CTTC	ILCRRCGASVES
					TCATCCTGCTCTGTAAAGATGCTCAGGGTG	CCAAAT	GGTT	LPHVLNHCPPN
					ACGTTCAGGACGTGGGAAACTGGAGACCGA	TATTCA	CCCC	MSIILGRHNLVL
					TCACCATCTGCCGCACTCTGTACAAATTGTA	AATTGA	AACC	QEVLNAVDKTQ
					CGCGGCTGTGATCGCCAGACGTATTCAGAC	AATCCG	ACTA	FKEISVDRTVPE
					TTGGGCGAAACAGGGCGGGGTGCTCTCGCG	GCCAAA	CAAC	HMSETGEALRP
					GCTCCAGAAAGGGTTCATGCCAGTTGAGGG	TTGCCT	AAAG	DIVARRNDGSV
					TGTGTTTGAGCACGTCTTCATGTTGGATACC	TGCTGA	CGAG	VVVDVACPFDQ
					GTCCTTTCCGACGCCAAGCTGAGGCGCAAG	ACCACC	CGGC	KANFDEAAKRK
					AATCTTCTTGCCGTCTTCCTTGACGTTCGGA	ATTTTT	CCCT	LLKYDKLCCNIA
					ATGCGTTCGGAAGCGTCCGGCATGAATGCT	GTTTTT	CTTA	ASTGKPVECHSI
					TGCTCAAAGTTCTTCGGCACTTCGATGCTCC	CTTATA	GATC	VVGSLGSLAEG
					CCATTACCTGGTTGAACTCGTGAGGGATATT	AAAGTA	CAAT	LSTSLRALGITD
					TACACCGGTGCGACCTGCCGTGTTAGATCTT	AATTTT	TTTA	FARSKLVACHQ
					CGGTTGGTGAGACCGGTGATATACCCTGTG	TCAACA	AAAT	

					ATCGAGGTGTCAGGCAGGGCTATCCCCTGT	ACCCAA	TTTA	G (SEQ ID NO:
					CTGGCATTCTTTTCAACCTGGTAACGGAGGT	AATCAT	AATC	1452)
					GCTAATACCGGGGCTCAGTGCCGGGAATGA	AAAATT	AGTG	
					TGGCTACCGGATGGCTTGCTTAGGTGGTAA	AAAACT	CACT	
					GCTCACTCAAGTCCTGGCCTACGCTGACGA	GTTATC	CAAC	
					TCTGGTGGTAGTCACGGAAAACCGTGATCA	TGATTA	TTTTT	
					GATGCTACGTCAGCTGGGTGTGTGTGAGGA	GTCTAA	TACG	
					ATTCGGCCGTTGGGCAGGCCTTGCATTCAAT	ATTCTG	TGTT	
					CAACGGAAGTGTGGCCTGATCGGCTGGAGG	TTATTG	GTGT	
					ACGTTGCGGGGAGGGAGGAGGGTCGCACTG	ACAAAA	TTTTT	
					GAGGATCCTCTGCTCCTTAACGGAGTAGAG	ACCTCC	GTAT	
					ATCCCCCTGTTGCGGCCCGGTGAACACTAT	ATACCA	TTTTT	
					AAATACCTTGGCGCCATGACTGGCGTCATG	ACCTTA	CCCA	
					AGTGTTCCCCGCACCGGATCCCAGCTGATC	ATCTTA	CAGA	
					AAGGACTTCCGTGCAAGGCTCCAACGACTG	TCATTA	TTTTG	
					TTTACATCGTTCCTCACACCACACCAGAAGC	TCTTCA	TATTT	
					TGATCGCTCTGAAACGGTTTCTGCTGCCAAG	CCCAGA	TTAT	
					TCTCAGCTTTCATCTGAGGGTTCGTCCGATT	CTGCTC	ATTA	
					GCGCGGTCTGAGCTGATCGCGCTCGACCGT	AATCTA	TATTT	
					CGCGTGCGCGAGTGCCTTAGGGTGGCGTTT	GTTCTT	TATA	
					CGCTTGACCAAGCCCTCGTGTCAGGCTGTGT	TTTTCA	TACA	
					TTCATACACCGACTGACATGGATGGTCTGG	ATCAAT	CAAC	
					GTGTCCCGTCGGTATGCAGTGAGAGCAGCA	AAGGAG	ACTA	

					TACTGACGATCGCCCAGGGGTTCAAGGTGT	GTGTTT	TTTTT	
					TAACGTCACCCGACGGCACAGTGTCCGCGA	TCTTTG	ATAC	
					CGGCTTCGGCGCGCGTGAAGCTCTACGCCG	GTCTGT	ACTA	
					CTAAATTTGGCGGCCTCACTGAAGCAGGTC	TGTTTT	CCTT	
					CTTCGGACTGGGCGCGTTATCTATCTGGTGA	ATAGTT	GCAC	
					TGACGTCAACGGGAATAGCACTCGCAAACC	TACCGT	TGTC	
					CGGCGCTAACCTGCCTTCTGGTCTGTGGACT	TATAAT	CCAC	
					CGCGTCCGCTGTGCCAGCAGGCAGTTAGGT	AGTCTG	TTTTT	
					GCTGTGTGGCGCGTTTGCCCGGAAAACGGC	TTGTCC	GTAA	
					ATCACTGTCCGGGTGCGTAATTCTGTGATTA	TATAGT	TTATC	
					CCTCCCGAGACCGACGGAAGCTCATCAGAT	TTTCCG	ACCT	
					CCTTCCACGATTGCTCTAATCAGCAGTGGA	TGGTGT	TTTAC	
					AGGAGCAGTGGATGCAGCACCCCAACCAGG	ATTTCT	CTTTT	
					AGAAGACCGCTGCAGCGCATATGGCTTATG	GGTCTA	ATGC	
					CCGATGCGAACCGCTGGGTGAAGCAGCCGT	CTATAG	CGCT	
					CGGTTATGGAACCTCATACATACTTTTTTGC	TTGGTG	CGCT	
					ATTGAGGGCGCGCTTGAACCTGTTACCCAC	CCTCTA	GAGC	
					TAGGGTTTCACGAGCTATTTATAGCAGGGA	AAACAT	TCCTT	
					CCAACACCCGGACATTTTATGCCGACGTTGT	TATTAT	GCAC	
					GGTGCGTCAGTTGAGTCCCTCCCGCATGTAC	ACCGAA	GGAT	
					TAAACCATTGTCCTCCCAATATGAGCATAAT	GAGTGC	GACC	
					ACTCGGGAGGCATAATTTGGTACTGCAGGA	TAGCGA	AGAC	
					GGTTCTGAACGCGGTAGACAAAACCCAGTT	TTTCTA	AAAC	

					CAAAGAGATATCCGTTGACCGAACGGTCCC	TTATTA	TTTTA	
					CGAGCACATGTCGGAGACAGGTGAGGCCTT	ACTCAC	TAAA	
					GCGGCCTGACATAGTGGCACGGCGGAATGA	CCACCA	ATTA	
					CGGTTCCGTCGTGGTCGTCGACGTGGCTTGC	GTACTG	TAAC	
					CCGTTCGATCAAAAGGCGAACTTTGATGAG	CTGTGC	ATTG	
					GCGGCAAAACGGAAGCTCCTAAAGTACGAC	ATCCAC	TTTTA	
					AAGCTTTGTTGCAATATCGCTGCCTCCACAG	GCAGCA	ATTG	
					GTAAACCTGTGGAGTGTCATTCGATAGTCG	CCACGC	TCGC	
					TGGGCTCCTTGGGGAGCCTCGCGGAGGGTC	CTGCCA	GGGG	
					TGTCCACCTCCCTCCGGGCTTTGGGAATAAC	TCACGT	TATC	
					CGATTTTGCTCGCTCTAAGCTTGTCGCTTGC	ACAGTG	AGTG	
					CATCAAGGGTAGCCAAAAGATTTGGTGTTC	CGCTTG	GCGC	
					TTGGCGAGCTGCAGTCCCGGCAGAGTAGGC	CTGCCT	CCCC	
					GAGATCGTGAGCTCTACACCGCTTCACGAT	TGTGTT	TTGC	
					TTGACGAGTTCTTCTTGAGTCGATCCCCAAC	TTGTTTT	GGCC	
					GGCAACGCACTCCAAATTTAACGACCGCAT	GCAGAC	GTGA	
					GGCGACGCCACTGATCTCCCAGTCATCCGT	ACCACT	AAGC	
					GCTCCGAGGGCACTTAAGGGGAACCGATTC	GGACGA	CCTTT	
					ACTGATGCTGCAAGTTGGCGGGGGCGCACT	TAG	TCAC	
					CACCAATGGAAGATGAGGGAGCCCAACGAT	(SEQ ID	CTTA	
					GCCCCGAACTACCGCACCGTCCCCAGCACC	NO: 1207)	ATTC	
					GTGTCCTTGTGAGTAATTATTTTCACTAATC		GCCC	
					CAGGGACGGGGCCATGAACTGAACTATGTC		CTTA	

					TGCACGCTGCCCCGCCTAGCCTCGGCCACA		GTCC	
					AATAAATAGTCAAGCCGGGGCGTACTACCA		AAAT	
					ATTGATTCGGATTTGGCGGGGCCCATCACG		TTTTC	
					CGCCTCCGCCCCCACACACAAAAACACTTT		CCCA	
					TAAACTCTTCGGTTCCCCAACCACTACAACA		TTGC	
					AAGCGAGCGGCCCCTCTTAGATCCAATTTT		CTGT	
					AAAATTTTAAATCAGTGCACTCAACTTTTTT		AAAA	
					ACGTGTTGTGTTTTTTGTATTTTTTCCCACAG		GTGC	
					ATTTTGTATTTTTATATTATATTTTATATACA		GTTG	
					CAACACTATTTTTATACACTACCTTGCACTG		CCGT	
					TCCCACTTTTTGTAATTATCACCTTTTACCTT		CGAC	
					TTATGCCGCTCGCTGAGCTCCTTGCACGGAT		TCGA	
					GACCAGACAAACTTTTATAAAATTATAACA		ACTC	
					TTGTTTTAATTGTCGCGGGGTATCAGTGGCG		GTCC	
					CCCCCTTGCGGCCGTGAAAGCCCTTTTCACC		TTTTT	
					TTAATTCGCCCCTTAGTCCAAATTTTTCCCC		ACCG	
					ATTGCCTGTAAAAGTGCGTTGCCGTCGACTC		CCTCT	
					GAACTCGTCCTTTTTACCGCCTCTCTGTTAC		CTGTT	
					TGAACTATGACCAGCTTGGCTGTTGAAGTC		ACTG	
					GGCTTTAATTCGCCGGCTTCACATTATATTT		AACT	
					TTTTGTTGGGTAATCTGTTTTTTTTCATATAT		ATGA	
					AAAACCATTCGGCTTAAAATCACCAATCCC		CCAG	
					CATCATCCACCCTGGTCACCCATTGAAACAT		CTTG	

					CTCTTTTTAGTCAATTATTTTTTCAGATTACC CCTCCCAAATTGTCATATAGTTTAACACCCC GTTTGCCAAGTTGGTCTTTTCCCCATCCCCT TCTGTCCTTCCGGTAATCCCATAACATTTTA TTCTTTAATTCGGCCCCATAAATAGTGTAAT CTTAAGACGAAGTCCCCAGAGGCCCGGTCC CCCCCTTCTTTGTCATGGACCGGGCCAGCCC CCCTGTTACCACAACAACCCACCATTTTATT CTTTCTTTCTTTTTATTAGTATTTATTTATAT TGCCGCCCAACTGTCGTGTCGTGGCGCAGG GGGGTTCCCTCTGTCGGCCCAGTGGACGAC CGTCTAAAAAACAGGCACAGGCAGCAGGC ACTTATCACCCGGTACCTCCGGGTACCGGA GGCTGGTTTTCAGCGCACTGTACGTGATGG CCAATTTTATTCATTGCATTTTATCCGCGTC GTGGTGTTTGCGTGGATGCATTAATAAAAG ATGAAATTCC (SEQ ID NO: 1084)		GCTG TTGA AGTC GGCT TTAA TTCG CCGG CTTC ACAT TATA TTTTT TTGTT GGGT AATC TGTTT TTTTT CATA TATA AAAC CATT CGGC TTAA AATC	

							ACCA ATCC CCAT CATC CACC CTGG TCAC CCAT TGAA ACAT CTCTT TTTA GTCA ATTA TTTTT TCAG ATTA CCCC TCCC AAAT TGTC ATAT AGTT	

							TAAC ACCC CGTTT GCCA AGTT GGTC TTTTC CCCA TCCC CTTCT GTCC TTCC GGTA ATCC CATA ACAT TTTAT TCTTT AATT CGGC CCCA TAAA TAGT	

							GTAA TCTTA AGAC GAAG TCCC CAGA GGCC CGGT CCCC CCCTT CTTTG TCAT GGAC CGGG CCAG CCCC CCTG TTAC CACA ACAA CCCA CCAT TTTAT	

							TCTTT CTTTC TTTTT ATTA GTAT TTATT TATA TTGC CGCC CAAC TGTC GTGT CGTG GCGC AGGG GGGT TCCCT CTGT CGGC CCAG TGGA CGAC CGTC	

							TAAA AAAC AGGC ACAG GCAG CAGG CACT TATC ACCC GGTA CCTC CGGG TACC GGAG GCTG GTTTT CAGC GCAC TGTA CGTG ATGG CCAA TTTTA	

							TTCAT TGCA TTTTA TCCG CGTC GTGG TGTTT GCGT GGAT GCAT TAAT AAAA GATG AAAT TCC (SEQ ID NO: 1330)	
R2	PERER E-9	BN0008 00		Schistosom a mansoni	ATCTCACGTTTTAATTTATTTTTGAACTACT GCAGTCTGAGTGCTTCTAACGACCCGAAGG CTCAGAAACTACCCACTTCTTGAACTGCTAC TTTTTGCTGTTTATCCACAACAACAGTTGTG ATTCTATTCTCCANATATTCCTTGTGCTTTTG	ATCTCA CGTTTT AATTTA TTTTTG AACTAC	TAAC GGCT GAAC GAAT AGCC	MPVSTGAETDIT SSLPIPASSIVSP NYTLPDSSSTCLI CFAIFPTHNILLS HATAIHHISCPPT

					TCAACATTATTCTATACCAACTGTACCACCT	TGCAGT	CCCTT	PVQDGSQQMSC
					ACTTCTTCATCTCACGTTTTAATTCTGGTCTT	CTGAGT	CACT	VLCAAAFSSNR
					ATTTTCTCATCATTAGTCACGGAGAGGGCCT	GCTTCT	CTTA	GLTQHIRHRHIS
					ATGAACGGTCCGTGACGCGAAATTCTATCC	AACGAC	GACA	EYNELIRQRIAV
					GCGATTTCGACCTCTCCTGCTAGTGGTCCCC	CCGAAG	TTCCC	QPTSRIWSPFDD
					GAAGTACGGTTCCTCTGGCCTGTCAGTTGTG	GCTCAG	CCAC	ASLLSIANHEAH
					TTAAAACTATATAATAACGATGCCGGTCTC	AAACTA	TGTT	RFPTKNDLCQHI
					AACCGGCGCAGAAACTGACATAACCTCTTC	CCCACT	GTTG	STILTRRTAEAV
					TTTGCCTATTCCTGCATCCTCAATCGTCTCG	TCTTGA	CTTAT	KRRLLHLQWSR
					CCAAACTACACACTCCCTGATTCCTCTTCAA	ACTGCT	CTTC	SPTAITTSSNNH
					CCTGCCTTATATGTTTCGCTATCTTCCCCAC	ACTTTT	ATGC	TITDIPNTEARYI
					CCACAACATACTCCTCTCCCATGCCACTGCA	TGCTGT	TCTTG	FPVDLDEHPPLS
					ATCCACCATATTTCTTGTCCTCCTACTCCAG	TTATCC	TGTT	DATTPNASTHPL
					TGCAAGACGGTTCTCAGCAGATGTCTTGTGT	ACAACA	AATT	PELLVILTPLPSP
					TCTTTGCGCCGCCGCTTTTTCATCTAACAGG	ACAGTT	GACT	TRLQNISESQTS
					GGACTAACACAACACATTCGCCACCGGCAC	GTGATT	GCTC	HESNKNSMHTP
					ATCTCCGAATATAACGAACTAATCAGACAA	CTATTC	TCTTC	PTYACDPDETLG
					CGAATTGCAGTGCAGCCGACGTCTCGCATA	TCCANA	TGGG	ATPSSTIPSCFHS
					TGGTCACCATTCGATGATGCTTCTTTACTAT	TATTCC	TTGA	YQDPLAEQRGK
					CCATCGCTAACCATGAAGCCCATAGATTCC	TTGTGC	CGTC	LLRASASLLQSS
					CCACGAAGAATGACCTATGCCAACATATCA	TTTTGT	TGAT	CTRIRSSSLLAFL
					GCACCATACTAACACGCAGGACGGCAGAAG	CAACAT	TGTCT	QNESTLMDEEH
					CCGTCAAACGCCGACTCCTCCACCTACAGT	TATTCT	CTCTC	VSTFLNSHAEFV

					GGTCGAGATCACCAACAGCGATTACTACCT	ATACCA	TCTTT	FPRTWTPSRPKH
					CTTCGAATAATCACACAATCACAGACATCC	ACTGTA	CCAT	PSHAPANVSRK
					CCAATACCGAGGCCCGATATATTTTTCCGGT	CCACCT	ATTG	KRRKIEYAHIQR
					AGACCTAGACGAACATCCACCATTGTCTGA	ACTTCT	CTTG	LFHHRPKDASN
					TGCCACAACCCCCAACGCATCGACACATCC	TCATCT	CTCT	TVLDGRWRNPY
					ACTCCCAGAACTCCTTGTCATCTTGACACCG	CACGTT	GCCC	VANHSMIPDFD
					CTTCCATCCCCGACTAGACTACAAAACATA	TTAATT	GCTT	CFWTTVFTKTN
					TCCGAATCACAGACCTCCCATGAATCTAAT	CTGGTC	ACTT	SPDSREITPIIPM
					AAGAACTCAATGCATACACCGCCAACGTAT	TTATTTT	CCAA	TPSLIDPILPSDV
					GCCTGCGATCCGGATGAGACACTAGGGGCT	CTCATC	TAGT	TWALKEMHGT
					ACTCCCTCATCAACTATTCCCTCATGCTTCC	ATTAGT	TGTC	AGGIDRLTSYDL
					ACAGTTATCAGGACCCCCTAGCTGAACAAA	CACGGA	ATAT	MRFGKNGLAGY
					GAGGCAAACTCCTGAGGGCATCCGCCAGCC	GAGGGC	TATG	LNMLLALAYLP
					TACTACAAAGCAGTTGTACTCGCATACGGT	CTATGA	TCTTT	TNLSTARVTFVP
					CCTCCAGCCTGCTCGCCTTCCTCCAAAACGA	ACGGTC	GTTT	KSSSPVSPEDFR
					ATCCACATTAATGGACGAGGAACACGTGTC	CGTGAC	ACTT	PISVAPVATRCL
					CACCTTCCTCAATAGTCATGCAGAATTCGTC	GCGAAA	GCCA	HKILAKRWMPL
					TTCCCTAGAACATGGACCCCATCCCGACCC	TTCTAT	TGTCT	FPQERLQFAFLN
					AAACACCCCTCCCACGCCCCAGCTAATGTTT	CCGCGA	AACG	RDGCFEAVNLL
					CTAGGAAGAAAAGGAGGAAAATAGAGTAC	TTTCGA	ACAA	HSVIRHVHTRHT
					GCACACATCCAGAGACTCTTCCACCACCGT	CCTCTC	TTACT	GASFALLDISRA
					CCCAAAGATGCCTCCAACACCGTTCTAGAC	CTGCTA	TTATC	FDTVSHDSIIRA
					GGTCGGTGGAGAAACCCCTATGTCGCAAAC	GTGGTC	TACC	AKRYGAPELLC

					CATTCAATGATTCCAGACTTCGACTGCTTCT	CCCGAA	TTAG	RYLNNYYRRST
					GGACAACAGTCTTTACTAAAACAAATTCCC	GTACGG	TTTGT	SCVNRTELHPTC
					CAGACAGCCGGGAGATTACTCCAATCATCC	TTCCTC	CCTCT	GVKQGDPLSPLL
					CTATGACTCCCTCTCTCATTGACCCGATCCT	TGGCCT	TGGT	FIMVLDEVLEGL
					CCCCTCTGACGTCACATGGGCGCTGAAAGA	GTCAGT	TTCG	DPMTHLTVDGE
					AATGCATGGCACGGCCGGTGGGATTGATCG	TGTGTT	ATTG	SLNYIAYADDL
					TCTGACATCGTACGATCTGATGAGATTCGG	AAAACT	CCTTC	VVFAPNAELLQ
					GAAGAATGGTCTTGCTGGATATCTCAACAT	ATATAA	ATAT	RKLDRISILLHE
					GCTACTCGCTCTTGCATACCTTCCCACTAAT	TAACG	GTTC	AGWSVNPEKSR
					CTTTCAACAGCACGGGTAACTTTCGTCCCCA	(SEQ ID	ATGG	TLDLISGGHSKIT
					AGTCATCAAGTCCTGTGTCACCTGAGGACTT	NO: 1208)	CGGA	ALSQTEFTIAGM
					CCGTCCCATCAGTGTCGCTCCAGTAGCCACT		ATCT	RIPPLSAADTFD
					AGGTGCCTGCACAAAATTCTAGCTAAGAGA		GATG	YLGIKFNFKGRC
					TGGATGCCGCTCTTTCCACAGGAACGACTTC		TTTAT	PVAHIDLLNNYL
					AGTTCGCTTTCCTAAACCGAGATGGATGCTT		AATG	TEISCAPLKPQQ
					TGAAGCAGTTAATCTTCTGCACTCGGTCATA		ACTA	RMKILKDNLLPR
					CGGCACGTCCACACCCGCCATACAGGAGCA		TTCCT	LLYPLTLGIVHL
					TCCTTCGCCCTGCTCGACATATCACGGGCCT		ATTA	KTLKSMDRNIH
					TTGACACTGTATCACATGACTCCATCATCAG		CCAC	TAIRKWLRLPSD
					AGCGGCGAAAAGATATGGGGCACCTGAACT		CACT	TPLAYFHSPVAA
					GTTATGCCGCTACCTCAATAACTATTACCGA		ACAA	GGLGILHLSSSV
					CGTTCAACCAGCTGCGTCAACCGCACTGAA		CTAC	PFHRRKRLETLL
					TTGCATCCTACGTGTGGGGTGAAGCAAGGA		TATT	SSPNRLLHKLPT

					GACCCCCTGTCGCCACTCCTCTTCATCATGG		ATTA	SPTLASYSHLSQ
					TTCTCGACGAAGTACTGGAAGGTCTAGATC		TTTTC	LPVRIGHETVTS
					CAATGACCCACCTAACAGTTGATGGAGAGA		ATTA	REEASNSWVRR
					GCTTGAACTACATAGCTTATGCTGACGATCT		CTATT	LHSSCDGKGLLL
					CGTAGTTTTCGCTCCAAATGCGGAACTCCTT		AACA	APLSTESHAWL
					CAACGGAAACTCGATCGGATCTCCATACTT		TTATT	RYPQSIFPSVYIN
					CTACACGAGGCTGGATGGTCGGTTAACCCT		ATAA	AVKLRGGLLST
					GAAAAAAGCCGGACCCTGGACCTAATCTCT		ACAT	KVRRSRGGRVT
					GGTGGCCATTCCAAAATCACAGCGCTCTCT		TATT	NGLNCRGGCAH
					CAGACAGAATTCACCATCGCGGGGATGCGT		ACTA	HETIHHILQHCA
					ATACCACCGCTTTCTGCCGCCGACACCTTCG		TTATT	LTHDIRCKRHNE
					ACTACCTGGGTATCAAATTCAACTTCAAGG		ATTA	LCNLVAKKLRR
					GCCGATGCCCAGTGGCCCATATTGACTTATT		TTACT	QKIHFLQEPCIPL
					GAACAACTACCTCACGGAAATATCGTGCGC		ATTA	EKTYCKPDFIIIR
					TCCACTTAAGCCGCAGCAGCGCATGAAGAT		TTACT	DSIAYVLDVTVS
					CTTGAAAGATAATCTACTCCCTCGACTCCTA		TCTA	DDGNTHASRLL
					TACCCCCTGACTCTAGGAATAGTACACCTG		CAAT	KISKYGNERTVA
					AAAACCCTGAAGTCAATGGACCGAAATATC		TAAT	SIKRFLTSSGYII
					CACACGGCCATAAGGAAATGGTTGCGGCTA		ATTA	TSVRQTPVVLTF
					CCCTCCGACACCCCGCTAGCATATTTTCACT		TGGC	RGILDRASSQSL
					CACCCGTCGCTGCCGGAGGCCTAGGGATCC		TACT	RRLCFSSRDLGD
					TCCATCTGTCCTCATCGGTTCCATTCCACCG		CCTCT	LCLSAIQGSIKIY
					TCGAAAACGTCTAGAAACCCTCCTATCTTCA		CAGC	

					CCGAACCGCCTACTGCACAAGTTGCCAACT		ACAC	NTYMRGT (SEQ
					TCCCCAACACTAGCTTCTTATTCACACCTTA		CAAT	ID NO: 1453)
					GTCAACTGCCAGTTCGAATTGGGCACGAGA		AAAA	
					CCGTAACGTCTAGAGAAGAGGCTTCCAACA		TATC	
					GCTGGGTGAGACGATTACATTCGTCCTGCG		AATC	
					ACGGGAAGGGACTACTCCTAGCACCACTAA		AAAC	
					GCACCGAGTCCCATGCATGGCTGCGCTACC		ATCT	
					CCCAGTCTATTTTTCCATCTGTTTACATCAA		CAAT	
					CGCCGTTAAATTACGAGGTGGCTTACTATCC		TATA	
					ACCAAAGTCAGGAGATCTCGCGGAGGTAGA		TCCA	
					GTGACGAATGGCCTGAACTGTCGAGGCGGT		CCTA	
					TGCGCCCATCATGAAACGATCCACCACATT		TTAA	
					CTGCAACATTGCGCGCTCACCCACGACATC		ACTC	
					AGATGCAAACGCCATAACGAACTATGCAAC		TCTCT	
					CTTGTGGCAAAGAAACTGCGTAGGCAAAAA		ATTTC	
					ATCCATTTCTTACAGGAGCCCTGCATTCCTC		CCCT	
					TAGAAAAAACCTACTGCAAACCTGATTTTA		GAGT	
					TAATTATACGTGACTCAATTGCTTATGTTCT		TATA	
					AGACGTCACTGTATCGGACGACGGAAACAC		AACT	
					CCACGCCAGCCGCCTGTTAAAAATATCAAA		TACA	
					ATACGGCAATGAGCGAACCGTCGCGTCGAT		ATTC	
					CAAGCGTTTCCTCACATCCAGTGGATATATC		AGTC	
					ATTACCAGTGTTCGACAAACACCAGTCGTC		TAAC	

					CTTACATTCAGAGGTATTCTGGATAGAGCA		CGAA	
					AGTTCACAATCCCTACGACGCCTATGTTTTT		TATCT	
					CATCCCGTGACCTCGGTGACCTTTGCCTGAG		CTCTT	
					TGCGATTCAAGGCTCAATTAAAATATATAA		TTAC	
					TACCTATATGAGAGGAACCTAACGGCTGAA		AAAT	
					CGAATAGCCCCCTTCACTCTTAGACATTCCC		CTTA	
					CCACTGTTGTTGCTTATCTTCATGCTCTTGT		AGTA	
					GTTAATTGACTGCTCTCTTCTGGGTTGACGT		TGTA	
					CTGATTGTCTCTCTCTCTTTCCATATTGCTTG		ATTTT	
					CTCTGCCCGCTTACTTCCAATAGTTGTCATA		GTGC	
					TTATGTCTTTGTTTACTTGCCATGTCTAACG		CAAA	
					ACAATTACTTTATCTACCTTAGTTTGTCCTC		CCCA	
					TTGGTTTCGATTGCCTTCATATGTTCATGGC		TTTG	
					GGAATCTGATGTTTATAATGACTATTCCTAT		GGTC	
					TACCACCACTACAACTACTATTATTATTTTC		TGTA	
					ATTACTATTAACATTATTATAAACATTATTA		CAAT	
					CTATTATTATTATTACTATTATTACTTCTACA		TTGA	
					ATTAATATTATGGCTACTCCTCTCAGCACAC		TACTT	
					CAATAAAATATCAATCAAACATCTCAATTA		AAAA	
					TATCCACCTATTAAACTCTCTCTATTTCCCC		ATAA	
					TGAGTTATAAACTTACAATTCAGTCTAACCG		ATGT	
					AATATCTCTCTTTTACAAATCTTAAGTATGT		TATT	
					AATTTTGTGCCAAACCCATTTGGGTCTGTAC		AGCC	

					AATTTGATACTTAAAAATAAATGTTATTAGC C (SEQ ID NO: 1085)		(SEQ ID NO: 1331)	
R2	R2-1_BTe	.		Bombus terrestris	TCAATAGTTACTCGGGGGCAGGCGGGATAT TGGTCTTGCCTTGCCCAAGTCACACTCCTAC CTCCTCGTGGTACCGCCGGTAACACGCGCA CGTCCACATCAGCGAGGGGCGTACTCCCCC GGATGTGGCGGCGCGTGGCTAAACGGAGTG TGGCGACGAAGGAGCGAAAGACTAACAAC TATAACGGTCTTCCGTAACGGCTACTTGGA GCCGTGAATAATGGAGCCTATATTAAACCC TGGAACTCGTTCCTTCGTTCTGTTGACGACT GGAACGGCAAAGGACATGATTTGGATAACA ATTGGAAACTTAATATCGAATTCACAATTA AATTGCAAAATTCGATAATAATACGAATAG TGCCTCGGACGCAGCGCCTCTCAGTCCCGG AGGCGCTGTGGCGGATTTATCCGCTTCAGA GGGGACTACGGATAACGACCAGGCTATGTC CCCGGCTATGAGTTTGATNACGGTGCCGCT CGTGGGCAATCGGGTCGCCTGTCCAGNATG TGAGAAAAGGGAGGCTAACCTTTTCTTCTT AAATCTGTCGGACCTGGACAGGCACCTAAC ACAACACCACCCGGATGCCCCGATCTNTTG	TCAATA GTTACT CGGGGG CAGGCG GGATAT TGGTCT TGCCTT GCCCAA GTCACA CTCCTA CCTCCT CGTGGT ACCGCC GGTAAC ACGCGC ACGTCC ACATCA GCGAGG GGCGTA CTCCCC	TGAA AACA CGAT AACG ATTA TGAA TCAA AATA AGAA AGTA AACA TCCC AGAA ATTG TCTA CGTC TTATT TGTT ATCT ATTT	IAKFDNNTNSAS DAAPLSPGGAV ADLSASEGTTD NDQAMSPAMSL XTVPLVGNRVA CPXCEKREANLF FLNLSDLDRHLT QHHPDAPIXWS CIDCAKCFPKLH GARCHIPKCGG ASSQARTGEFQC EACPMSFGSRR GLSTHERHAHP AVRNIKRRGAD PPEENTKSWKV EXVARLKGLWE IFKDHKYPNKEI SKFLTTKTVDX XKYQRKKLNLI GXESPQEATSLA

					GTCNTGTATCGACTGCGCGAAATGCTTCCC	CGGATG	ATTT	TEGGCDLVSSG
					GAAGCTTCATGGGGCTAGATGCCACATACC	TGGCGG	GTTT	NASFGSPVGRNE
					TAAGTGTGGTGGCGCTAGCAGCCAGGCCAG	CGCGTG	A (SEQ	NEEELIHEWKLS
					GACAGGNGAGTTCCAGTGCGAGGCCTGTCC	GCTAAA	ID NO:	LKNEINKPTEVP
					CATGAGCTTCGGGTCGCGCAGAGGGCTATC	CGGAGT	1332)	PILKEVYNRLML
					CACCCACGAACGGCATGCGCACCCTGCCGT	GTGGCG		IWEEHQDDRDS
					CAGAAATATTAAAAGAAGGGGAGCGGACC	ACGAAG		LTESLDHFIRTA
					CCCCAGAAGAAAATACAAAATCTTGGAAAG	GAGCGA		LYELINKINKNQ
					TAGAANNGGTAGCACGCTTGAAGGGGCTAT	AAGACT		TDLKTKRAAKT
					GGGAAATATTCAAAGACCATAAATATCCCA	AACAAC		KSPKNNRNSRK
					ACAAAGAAATCAGCAAATTCCTCACCACNA	TATAAC		RFSYARCQELFH
					AAACGGTNGATCANATNAAATATCAAAGAA	GGTCTT		ECPRRLADAVV
					AAAAGCTAAACTTAATTGGTGNGGAAAGCC	CCGTAA		NNDQAYLEPAR
					CCCAAGAGGCTACCTCACTGGCAACAGAGG	CGGCTA		QPPGSEEVRGLY
					GAGGGTGCGATCTCGTTAGTTCAGGCAATG	CTTGGA		EKLWGQVGSTY
					CCAGTTTTGGCTCGCCTGTAGGCCGCAACG	GCCGTG		VPAPVTRVPKLS
					AGAACGAAGAAGAGCTCATCCACGAATGG	AATAAT		LSEIFPPIAAEDV
					AAGCTCTCGCTAAAAAACGAAATAAATAAG	GGAGCC		GERIGKIRKKAA
					CCAACCGAGGTGCCCCCTATTTTAAAGGAG	TATATT		AGPDGLQRDHL
					GTTTACAATCGGCTGATGTTAATCTGGGAA	AAACCC		TIPGLPIIMAKIY
					GAGCACCAAGATGATCGAGACTCTCTAACG	TGGAAC		NILVYCSYFPSA
					GAGAGTCTCGATCACTTTATACGCACAGCT	TCGTTC		WKENRTTLIPKI
					CTATACGAGCTTATAAACAAAATAAATAAA	CTTCGT		NKPCSLVENWR

					AACCAAACGGATTTAAAAACTAAAAGAGCA	TCTGTT		PITISPILGRIFSSI
					GCTAAAACAAAAAGCCCAAAAAATAATAG	GACGAC		IDGRIRRGTVLN
					AAACTCCAGGAAGCGATTCTCATACGCTCG	TGGAAC		MRQKGFTSENG
					TTGCCAGGAGTTATTTCACGAATGCCCAAG	GGCAAA		CKINIELLNSAL
					GAGACTGGCTGATGCCGTGGTCAATAACGA	GGACAT		NYSKRNSGGIFT
					TCAGGCATATCTCGAACCAGCCAGGCAACC	GATTTG		IVDISKVFDTVP
					TCCCGGATCAGAGGAAGTGAGGGGGCTCTA	GATAAC		HAALKPCLAKK
					CGAAAAGCTGTGGGGACAAGTGGGTTCCAC	AATTGG		GVPALIVDLIDE
					GTACGTCCCGGCTCCAGTTACGAGAGTCCC	AAACTT		MYKNVKTTIKT
					CAAACTATCCCTATCCGAGATCTTCCCGCCG	AATATC		KDGGVEIMIRRG
					ATAGCGGCTGAGGATGTGGGGGAGAGAATC	GAATTC		VKQGDPLSPLLF
					GGCAAAATAAGAAAGAAAGCTGCGGCAGG	ACAATT		NLCLEPLLEEIEE
					ACCGGATGGATTACAAAGAGATCACCTAAC	AA (SEQ		QASGINVSEHRK
					TATCCCCGGCCTGCCTATCATAATGGCGAA	ID NO:		VSVLAFADDIVL
					AATCTATAATATACTCGTATATTGTTCTTAT	1209)		LGADAREAQHQ
					TTTCCCTCCGCATGGAAGGAGAATAGAACA			INVLTDYLQSLM
					ACTCTTATTCCAAAAATAAATAAGCCATGC			MNLSIEKCQTFE
					AGCCTGGTCGAGAACTGGCGACCAATTACT			VVAKKDTWFIK
					ATTAGTCCGATTCTAGGCCGAATCTTCTCCT			EPGLKIGNQIMP
					CCATTATCGACGGGAGGATAAGAAGAGGCA			TVDPDEAFKYL
					CTGTATTAAATATGAGGCAGAAGGGCTTTA			GAKIGPWKGVH
					CATCTGAAAACGGATGTAAAATTAACATCG			CGVIVPELLSVV
					AATTATTAAATTCTGCCTTAAACTATAGTAA			KRVRKLSLKPG

					AAGAAATAGCGGCGGGATATTCACTATCGT			QKVELLTKYIFP
					GGATATCTCGAAAGTTTTTGATACAGTGCCT			RYIYHLLVSPPS
					CATGCAGCCTTAAAACCTTGTCTGGCAAAA			DTVLKLLDSEV
					AAGGGTGTGCCCGCCCTTATCGTCGATTTGA			RQEVKTILHLVP
					TCGACGAAATGTATAAAAACGTAAAAACTA			STATGFFYTPKA
					CTATAAAAACTAAAGATGGCGGGGTCGAGA			CGGIGIPRFEHII
					TTATGATCCGCCGAGGAGTTAAGCAAGGCG			KLGTLKSAIKIA
					ACCCCCTATCGCCCTTACTCTTCAATTTATG			NSIDPAVAGLID
					CTTGGAGCCACTGCTGGAAGAGATCGAGGA			DAAIKKLKQTA
					GCAAGCCAGTGGAATAAATGTTAGCGAACA			NSLRINWPASLE
					TCGTAAAGTATCAGTTCTGGCCTTCGCAGAC			DIEKARKRLRKE
					GATATCGTTCTACTTGGAGCGGACGCGAGG			HISQWADLKCQ
					GAAGCGCAACACCAAATAAACGTACTTACC			GQGVPDFIKNKT
					GACTACCTGCAAAGCCTCATGATGAATCTA			GNLWLEDHSLL
					TCAATTGAAAAATGCCAAACCTTCGAGGTT			KPSRLIDALRLR
					GTGGCCAAAAAGGATACCTGGTTCATTAAG			TNTFGTRSVLAR
					GAACCTGGACTTAAAATCGGGAATCAAATA			ADKNIDVTCRR
					ATGCCTACTGTAGATCCCGACGAGGCTTTC			CRAQPETLGHIL
					AAATACCTGGGCGCCAAAATTGGTCCCTGG			GLCQHTKGLRIK
					AAAGGCGTTCATTGTGGTGTAATTGTTCCAG			RHDEVKSLLEG
					AACTTCTGAGCGTGGTGAAAAGGGTGAGGA			RLKSKKNNEVF
					AACTCTCCCTCAAGCCGGGCCAGAAGGTGG			VEPTIKAGGSLF
					AACTTCTAACTAAATACATCTTCCCCCGCTA			KPDLVIKNGERV

					CATTTACCATCTACTTGTAAGTCCGCCAAGC GATACCGTCCTTAAACTACTAGACAGCGAG GTCAGACAAGAAGTTAAAACTATCTTACAT CTTGTGCCTTCCACTGCCACAGGCTTCTTTT ACACTCCAAAGGCCTGTGGAGGAATAGGAA TACCGAGGTTTGAACATATAATCAAACTCG GTACCCTTAAAAGTGCAATAAAAATCGCAA ACTCGATCGATCCAGCGGTCGCTGGCCTAA TCGATGATGCAGCCATAAAAAAGTTAAAGC AAACGGCCAACTCCTTGCGGATCAATTGGC CAGCCTCCTTGGAGGATATTGAAAAAGCTC GTAAACGTCTAAGGAAAGAGCATATCAGCC AATGGGCTGATTTAAAATGCCAGGGACAAG GCGTTCCTGATTTTATCAAAAATAAAACTG GCAACCTGTGGCTTGAGGACCATAGTCTAC TCAAGCCATCGAGACTCATCGATGCCCTTC GATTAAGAACAAATACCTTTGGTACAAGAT CGGTGCTGGCACGGGCCGACAAAAATATTG ATGTAACATGTCGAAGATGTCGTGCCCAGC CCGAGACCCTTGGACACATACTCGGGCTAT GTCAGCACACTAAAGGCTTAAGAATCAAAA GGCACGATGAGGTTAAATCCCTCCTCGAAG GAAGATTAAAAAGTAAAAAGAACAACGAA			LVVDVTVRYEN KNYLALAEKEKI EKYRPCLRALK EIFNAKGGEILP VVLGSRGTITPN TEKVLKRLGIAN NDIKTILLNVLR SSIELCNIFIDD (SEQ ID NO: 1454)

					GTATTTGTAGAGCCGACGATTAAGGCCGGA GGCAGTTTATTTAAACCAGACCTCGTAATTA AAAACGGGGAAAGGGTTCTCGTGGTCGACG TAACTGTCCGCTACGAGAACAAAAACTACC TGGCCTTAGCCGAAAAAGAAAAGATAGAG AAGTATCGGCCATGCCTAAGGGCATTAAAA GAAATATTTAACGCCAAAGGAGGAGAGATT CTTCCGGTGGTCTTGGGTAGTAGAGGCACT ATCACGCCTAACACCGAGAAAGTCCTCAAG CGATTGGGAATCGCTAACAATGATATAAAG ACAATACTCTTAAATGTATTAAGAAGCTCG ATAGAGCTGTGTAATATATTCATTGATGACT GAAAACACGATAACGATTATGAATCAAAAT AAGAAAGTAAACATCCCAGAAATTGTCTAC GTCTTATTTGTTATCTATTTATTTGTTTA (SEQ ID NO: 1086)			
R2	R2-1_Crp	.		Crocodylus porosus	AGGCGTCTCCTTTAAGGGCAACGGTCTGGT TACGCGGTCGCAGCAGKCTTGCKCCAGGTA CCTCCWCGTGGTTCCCGCCGGGTGCCGMAG MCCCAGGSCTGTCGGTAGCTCGATCCTGGC ACAGTAMGGCCAGGGGAGTKCTTCCTTGCT GCWGGTGCCCCACAAGCGTKTGGCAGSMC MCCTGCTTCWTCGCAAMAATMASAGKGTS	AGGCGT CTCCTT TAAGGG CAACGG TCTGGT TACGCG GTCGCA	TGAA TCCC ACTC TGGG GACC CCCA AAAA	VPPGAEARGRY HHPRXEXARQG EPPSXRVFLVXL PDSNPPCPICGD HVXXXSVLALH CVEGHXWAXV QYQCTHCGILC

					MTCAGTAGTCGGCCCCGCCGCTAGCCAAAA	GCAGKC	TTAG	HIPRCQGRVXEX
					CTGTTCGCCACSCAGTTACAGATGGTGCCTG	TTGCKC	AAAA	TGKDXXCPECP
					TGCTGACCKGKKGCCCCGTGGGCTCGGWGG	CAGGTA	ACCC	ASFDEKAGLSQ
					TGCAGGGGWGGCCGGCCCATGGCTGGGCC	CCTCCW	AAAA	HKRHTVTXSXE
					AGACSTGGGCCKTGGAGCCCGCTCCCAMCC	CGTGGT	CAGT	RVAGXLLRAXL
					CAGAGTTCCCCCTCAAMGCCGCCAGGKCAG	TCCCGC	TGTG	RHGCWSVEEEE
					MAMCAKCCGGGGAGKGMCACGGCCCCGGC	CGGGTG	TTTA	TLTRLDAMFXG
					CGTGAGTCCCCCCCGGTGCWGAGGCCCGGG	CCGMAG	AGTG	ARNINQLIAAEX
					GCCGATACCACCACCCCCGGTGMGAGMGC	MCCCAG	TGTTC	VSKMXKQISDK
					GCCCGCCAGGGTGAACCACCCTCCAMCAGG	GSCTGT	TTGTT	WRXLXLXPEQT
					GTTTTCCTGGTGASGCTGCCTGACTCAAACC	CGGTAG	TGTC	TXGGXAESASV
					CKCCGTGCCCGATTTGCGGGGACCACGTGS	CTCGAT	CCTTT	VXXESMTPEME
					GTWAACWCTCCGTGCTGGCCCTCCACTGCG	CCTGGC	GGCT	AQSPAXPPGKIR
					TGGAGGGCCACGWGTGGGCGSATGTGCAGT	ACAGTA	TCAC	KIFTGQDGHAG
					ACCAGTGCACCCATTGTGGCATCCTCTGCCA	MGGCCA	CTCC	GXAWENQEDFH
					CATCCCACGTTGCCAGGGGAGGGTGCMKGA	GGGGAG	AAGT	WTRWARRWLK
					GCWGACCGGGAAGGACTKGGMTTGTCCTG	TKCTTC	TGCG	RGQXLSDKVQE
					AGTGCCCTGCCTCCTTCGATGAGAAAGCWG	CTTGCT	ATCC	VLGXWVEGQPR
					GCCTATCGCAGCACAAGCGGCACACGGTAA	GCWGGT	CCCA	IXAWVDXVSLD
					CACAMAGTGWKGAACGGGTTGCAGGKWAC	GCCCCA	TCTCC	VLTLFLGVPPGP
					CTGTTGAGGGCTKGTTTAAGGCACGGGTGC	CAAGCG	CCTG	QRAPSKKGPXE
					TGGTCGGTGGAAGAAGAGGAAACTCTCACC	TKTGGC	CGCT	GGKPTSWMNKR
					CGCCTWGACGCGATGTTCCSTGGCGCCCGG	AGSMC	GTCTT	AIKWGTFLRYQ

					AACATCAACCAGCTGATCGCTGCTGAGMTG	MCCTGC	TCTG	HLFGANRKLLA
					GTWTCGAAGATGCSTAAGCAGATCAGCGAC	TTCWTC	AATG	AXILDGAXRNQ
					AAATGGAGGTWGCTCKSTCTGWGTCCTGAG	GCAAM	ACCA	XTLLLEEVXQX
					CAGACCACATWGGGTGGTGAKGCTGAGTC	AATMAS	GTGG	YXGKWEAEPPF
					KGCCTCCGTGGTGSAGGWWGAGTCCATGA	AGKGTS	TGTT	EGLGRFGXXRD
					CTCCGGAGATGGAAGCCCAAAGCCCAGCTG	MTCAGT	GAGG	VDSFAFEALITX
					AMCCGCCTGGGAAAATCAGGAAGATTTTCA	AGTCGG	CTGG	EEAVKHMMXM
					CTGGACAAGATGGGCACGCCGGTGGCTGAA	CCCCGC	TGTG	AXXSAPGPDKL
					GAGGGGGCAGGKTCTGTCAGACAAGGTGCA	CGCTAG	ACCT	TLRDLRRADPE
					AGAAGTCCTGGGCKCGTGGGTGGAGGGTCA	CCAAAA	CGGT	GDALAELFSLW
					ACCCAGGATTCMTGCCTGGGTCGACKCMGT	CTGTTC	CACC	XITGVVPDRLKE
					CTCCCTTGATGTTTTGACTTTGTTCTTGGGG	GCCACS	TCCA	XQXVLIPKAVDF
					GTGCCCCCGGGACCGCAGAGGGCTCCGAGC	CAGTTA	AGGC	EKLRQLGNWRP
					AAGAAGGGGCCCAMGGAGGGTGGCAAACC	CAGATG	CAAG	ITIXSIVLQLXSR
					WACGTCCTGGATGAACAAACGTGCCATCAA	GTGCCT	TGCC	VLTARLTAACPI
					ATGGGGCACATTCCTTCGGTACCAGCACCT	GTGCTG	CTGG	XPHQQGFISAPX
					GTTTGGTGCCAACAGGAAGCTCCTGGCGGC	ACCKGK	CCCC	CAENLKXPELIX
					CMTCATCCTGGATGGTGCTGASCGTAATCA	KGCCCC	GAGT	RKVKXDRRPLG
					GTGSACCCTCCTGCTWGAGGAGGTCTTKCA	GTGGGC	AGGA	VAFVDXARAFD
					GGMCTACCKAGGAAAGTGGGAGGCCGAGC	TCGGWG	CCAG	SVSHDXISWVL
					CGCCTTTCGAGGGTCTTGGACGGTTTGGGST	GTGCAG	GTGG	KAKGVDQHIVN
					CCAMMGGGATGTGGATAGCTTCGCATTCGA	GGGWG	CCCA	LIEDSYQKVTM
					GGCCCTGATCACTCKGGAGGAGGCTGTCAA	GCCGGC	GCTT	RVQVFSGSTPPIS

					GCACATGATGGMAATGGCCWMSAAMTCGG	CCATGG	GCTG	IKXGVKQGDPM
					CTCCAGGTCCGGATAAGCTCACCCTGAGAG	CTGGGC	GGCA	SPLLFNIAMDPLI
					ATCTGCGCCGCGCKGACCCCGAGGGAGATG	CAGACS	CCCG	XKLKTVRQGVK
					CTCTTGCCGAACTGTTCAGCCTGTGGMTGA	TGGGCC	TCAC	VGSASLTTLAFA
					TCACTGGCGTGGTCCCGGACAGACTCAAGG	KTGGAG	CGCC	DGLXLLXDSWE
					AGTKTCAGTSTGTGCTGATACCAAAGGCAG	CCCGCT	CAGG	GMQHNITTSXTP
					TGGACTTTGAGAAGCTGAGGCAGCTGGGCA	CCCAMC	GCAA	GRACNTTSRHP
					ACTGGCGCCCCATCACTATCGSGTCCATTGT	CCAGAG	ATGG	RGLLQPHGPTSA
					KTTGCAGCTATTMTCACGGGTGCTGACCGC	TTCCCC	AAGG	TXKMXGVLLES
					ACGGCTCACCGCAGCWTGCCCCATAASCCC	CTCAAM	GATC	XMRLLYGEQLR
					CCACCAGCAGGGKTTCATMTCGGCACCKKG	GCCGCC	TATC	GLEDXRPXXHD
					GTGTGCCGAAAACCTGAAGMTGCCTGAGCT	AGGKCA	CTGA	AXARRADTISGL
					CATCTTSCGSAAGGTGAAGCAMGACAGGCG	GMAMC	CCAC	EXRSLGWDXQT
					GCCGCTGGGGGTGGCATTCGTGGATMTAGC	AKCCGG	TACC	RFGYATXLLAR
					SAGGGCSTTCGACTCGGTATCGCACGACMA	GGAGKG	AGGC	EDGDCXAQTNA
					CATTTCWTGGGTCCTAAAGGCCAAAGGGGT	MCACGG	TAAG	EALCWXSXPFP
					GGACCAACACATCGTGAATCTCATWGAAGA	CCCCGG	TGTG	GCAXRPXYANX
					TTCATACCAGAAGGTTACCATGAGAGTACA	CCGTGA	GTGC	GWVASEALDSM
					AGTGTTCAGTGGTTCCACCCCWCCCATCAG	(SEQ ID	TGCC	SRRXVKEWFHL
					CATCAAAWCCGGGGTCAAGCAAGGCGACC	NO: 1210)	TAGC	PACTDXLLXSR
					CGATGTCMCCCCTSTTGTTCAACATTGCGAT		CTGC	HRDGGLGLLRL
					GGACCCCCTGATAGSCAAGCTCAAGACAGT		CGTA	ARXXLAAXVRR
					CAGGCAGGGAGTGAAAGTCGGGAGTGCTTC		AGGT	PIRVATSSDEVT

					CCTGACCACCCTGGCATTTGCGGATGGCCTC		CAAG	RKVSYACGISDE
					GWMCTGCTGAKTGACTCCTGGGAGGGCAT		CGCC	VERLXLAXGGD
					GCAACACAACATCACGACATCCTAGAGGAC		CTGC	XSNVPRFEDPXA
					TTCTGCAACCTCACGGGCCTACGAGTGCAA		TGCC	PKSXXVQGPHE
					CCCAMAAAATGKCAGGGGTTCTTCTTGAAT		GCTC	AAQETPRVVRT
					CCGAMATGCGACTCCTWTACGGTGAACAA		GGGT	QAIPWPSNWRA
					CTGCGMGGCCTGGAAGATWGCCGGCCGTG		AKCA	EEHSKWAQLSC
					MSATCACGATGCTSGGGCCAGGCGAGCCGA		GTCC	QGERVELFCND
					CACGATATCTGGGCTTGAAKGTCGGTCCCT		TCGTT	PVSNGWINSRG
					GGGTTGGGATWGGCAAACCAGATTTGGGTA		CACT	QLAERLWIMAL
					CGCAACTAAKCTCCTGGCTCGAGAGGATGG		CGTC	KLRSNIYPTREF
					GGACTGCCWCGCTCAAACCAATGCAGAAG		CCTC	LGRGQAGTNIG
					CTCTTTGCTGGTSCAGTATKCCATTCCCAGG		CTAG	CRHCTHPRETLG
					CTGCGCTMTCAGGCCGWTTTATGCGAACMC		TACC	HILGICPAMQEA
					CGGTTGGGTGGCCTCGGAGGCTCTGGATTC		CTCC	RILRHNKLCKIL
					CATGAGCCGAAGAAAKGTGAAGGAGTGGTT		GCMT	AAEGKNCEWTV
					CCACCTACCCGCCTGTACCGACKGCCTCCTC		CTGC	YYEPHLHNAAG
					CASTCCCGCCACCGGGATGGGGGCCTKGGG		GCTT	ELRKPDLIFVRD
					TTGCTGCGCCTGGCGAGAGMCWTTCTGGCA		CTGC	GTALVVDITVW
					GCGSAAGTGAGGAGGCCGATTCGTGTAGCC		TATC	YEGGPATLLSTT
					ACCTCATCTGATGAAGTCACGCGGAAAGTA		CACT	AEKATKYLDLN
					TCCTACGCATGTGGGATAAGTGATGAAGTG		GTGG	TQIQELTGAEQV
					GAGCGKCTCTGKTTGGCMAGWGGGGGAGA		CCGG	TYFGFPIGARGK

					CATSTCCAATGTCCCGAGGTTCGAGGATCCT		WGAT	WHADNWRVLS
					GAWGCCCCGAAGTCTSCAKGGGTGCAGGGC		GCCG	ELGLSNSRKERV
					CCCCATGAGGCTGCCCAGGAGACACCACGT		AGGW	TRLLSWRALLG
					GTTGTCCGGACACAGGCGATCCCCTGGCCC		TGSW	SVDMVNIFVSK
					TCCAACTGGAGAGCCGAGGAACACTCCAAA		MCAW	HRQENLLDEHC
					TGGGCCCAACTCAGCTGTCAGGGAGAGAGG		CCTC	TPAEQVVSSYAS
					GTGGAGTTATTCTGCAATGATCCAGTCAGC		GACA	(SEQ ID NO:
					AATGGCTGGATCAACAGTCGGGGACAACTG		CCTC	1455)
					GCGGAACGCCTCTGGATCATGGCCTTAAAA		CAGG	
					CTGCGTTCGAACATTTATCCCACCAGGGAG		GCCA	
					TTTCTTGGGAGAGGCCAGGCAGGTACCAAC		AGCG	
					ATAGGTTGCCGGCACTGTACACACCCTCGT		CCTT	
					GAAACATTAGGGCATATCCTGGGCATATGT		GGCC	
					CCTGCCATGCAGGAGGCACGGATCCTCCGG		MCAG	
					CACAACAAGCTGTGCAAGATCCTGGCAGCT		GTAG	
					GAGGGCAAAAACTGTGAGTGGACAGTGTAT		GACC	
					TATGAGCCGCACCTACATAACGCAGCTGGG		TGGC	
					GAACTGCGTAAACCTGACCTCATTTTTGTCC		ACCT	
					GAGACGGTACCGCTCTGGTTGTGGATATCA		GCCC	
					CTGTGTGGTATGAAGGGGGCCCCGCAACCC		AGGG	
					TCCTGTCCACCACTGCAGAAAAGGCCACAA		GGCC	
					AATACCTGGATCTGAACACACAGATCCAGG		AGAC	
					AGCTCACAGGGGCTGAGCAGGTCACCTACT		ACTG	

					TTGGCTTCCCTATAGGAGCCAGGGGAAAGT		CCTG	
					GGCATGCTGACAATTGGCGAGTACTGTCTG		CGGC	
					AACTGGGATTGTCCAACTCCCGGAAGGAGC		AAGG	
					GGGTCACACGGCTCCTGTCGTGGCGAGCAC		GAAA	
					TGCTCGGGTCAGTGGACATGGTGAACATCT		GGGA	
					TTGTATCTAAACACAGGCAGGAAAACCTAT		GCCG	
					TGGATGAGCACTGCACCCCAGCTGAGCAAG		TCCCT	
					TTGTGAGTTCTTATGCGAGTTGAATCCCACT		GACC	
					CTGGGGACCCCCAAAAATTAGAAAAACCCA		GTTA	
					AAACAGTTGTGTTTAAGTGTGTTCTTGTTTG		CCGG	
					TCCCTTTGGCTTCACCTCCAAGTTGCGATCC		GCTT	
					CCCATCTCCCCTGCGCTGTCTTTCTGAATGA		GATG	
					CCAGTGGTGTTGAGGCTGGTGTGACCTCGG		GTGC	
					TCACCTCCAAGGCCAAGTGCCCTGGCCCCG		TGGC	
					AGTAGGACCAGGTGGCCCAGCTTGCTGGGC		TAGC	
					ACCCGTCACCGCCCAGGGCAAATGGAAGGG		CCGC	
					ATCTATCCTGACCACTACCAGGCTAAGTGT		CGTA	
					GGTGCTGCCTAGCCTGCCGTAAGGTCAAGC		TGTC	
					GCCCTGCTGCCGCTCGGGTAKCAGTCCTCGT		AAGC	
					TCACTCGTCCCTCCTAGTACCCTCCGCMTCT		ACTC	
					GCGCTTCTGCTATCCACTGTGGCCGGWGAT		CACA	
					GCCGAGGWTGSWMCAWCCTCGACACCTCC		GCTG	
					AGGGCCAAGCGCCTTGGCCMCAGGTAGGA		CTCG	

					CCTGGCACCTGCCCAGGGGGCCAGACACTG		AGTT	
					CCTGCGGCAAGGGAAAGGGAGCCGTCCCTG		CGCT	
					ACCGTTACCGGGCTTGATGGTGCTGGCTAG		GGCT	
					CCCGCCGTATGTCAAGCACTCCACAGCTGC		TCAC	
					TCGAGTTCGCTGGCTTCACCTTCATCCCWCC		CTTC	
					TAGTGCCTTCCGCCTCTGCGCTATTTTCGTC		ATCC	
					CCGACTCGTACCTCCCCACCTCTGCGCTAC		CWCC	
					MGCTATCCCTGMAAGGACCAAGTGGGCAG		TAGT	
					GGGSGTTCGCCCCCCGCCCGCAGGAGGCTC		GCCT	
					GGCGTATCCGTGGCKTCWTGCCTCCACCGT		TCCG	
					CTTGTGCCGCTAGAGGGGTACCTCMGAGAC		CCTCT	
					CGGCGCAACACGACCTTGACGSTTAGACAG		GCGC	
					TAGGGTGMAACAKCCCTGCTGCAGGCSTGA		TATTT	
					AGGGCCAAACGGCTGTGCCATGAGAGGGG		TCGT	
					AACCTTGAAGACCGGGSCAGTCAGCCAGTT		CCCG	
					AGTCAGTTGGGCGAAACAATCCCAGCTGCA		ACTC	
					GGCCCAMMAGGGCTGTCAGGTGAGGGGGT		GTAC	
					ATCCCCAWCCACCCCCCGCCGCCGACTACG		CTCC	
					GAGGCAKGAAGTCCCTAGTGACTTCKGACC		CCAC	
					CCCACGTCTTGTGCCGGGAGAGGGGAACCT		CTCT	
					TGAAGATCGGGGCAAGCCGCACTTGATAGT		GCGC	
					TAGCCAGTCKGGTGAAACAATCCCAGCTGC		TACM	
					GGGTCCGAAAGGGCCGACTWCCAGGCGAG		GCTA	

					GGGGGCCTGCGGAAAMCCCCCTCCATGGTA		TCCCT	
					CGGAGKTCTGGCGTCCTMACCGACWCCTTG		GMAA	
					CCACCAACGTCTTGTGCCGGGAGAGGGGAA		GGAC	
					CCTTGAAGATCGGGSCAAGCTGCACTTGAT		CAAG	
					GGTTAGTCAGTCGGGTGAAATAATCCCAGC		TGGG	
					WGCCCCCGCTGTGACTGCTAAGMCWGGTC		CAGG	
					CCCAAGGGGCATGAGGCATSTGCGCTGAGC		GGSG	
					CGGSAGGGGTGACACMCGGCGATCGGCGC		TTCG	
					AGCACAKASTGAAGGGAGGCACTTGCTGAG		CCCC	
					ACTGCTTCTGAGGCCCCAGACTTGGGGTGG		CCGC	
					TGCAGCCTTGTCTGGGGTATGGTACAGCAC		CCGC	
					CCTACTGCTCCCTTTGGKCAGCAGAATTCGT		AGGA	
					CCCGACCTCTTACCCACCCGAGTCTGCGCCT		GGCT	
					TGTTCCGCTATCCTGCATCTCCGATCCACCT		CGGC	
					CGCTGTCTCCCCGCTGCGCTGCTTTTCTCTC		GTAT	
					AAGTGGGTTAAATCTTGTCATGATTACCTCC		CCGT	
					CACGTTTCCGCTCAAGGGCAATGCCCAAMA		GGCK	
					TGACGGGGATCGCTGGTGCATGGCAGTCAT		TCWT	
					GAGACCATCCGGACCCTCCGGTGGTCGCTA		GCCT	
					TAGTCATTTTKTGTTGCATGGGGCATSCTGA		CCAC	
					GTCACTTAACCGAAAGACTCWAAATAACTC		CGTC	
					AAAAGAGGKAMCCTCTGSGGTTCGGTAAA		TTGT	
					(SEQ ID NO: 1087)		GCCG	

							CTAG AGGG GTAC CTCM GAGA CCGG CGCA ACAC GACC TTGA CGSTT AGAC AGTA GGGT GMAA CAKC CCTG CTGC AGGC STGA AGGG CCAA ACGG	

							CTGT GCCA TGAG AGGG GAAC CTTG AAGA CCGG GSCA GTCA GCCA GTTA GTCA GTTG GGCG AAAC AATC CCAG CTGC AGGC CCAM MAGG GCTG	

							TCAG GTGA GGGG GTAT CCCC AWCC ACCC CCCG CCGC CGAC TACG GAGG CAKG AAGT CCCT AGTG ACTT CKGA CCCC CACG TCTTG TGCC GGGA	

							GAGG GGAA CCTT GAAG ATCG GGGC AAGC CGCA CTTG ATAG TTAG CCAG TCKG GTGA AACA ATCC CAGC TGCG GGTC CGAA AGGG CCGA CTWC	

							CAGG CGAG GGGG GCCT GCGG AAAM CCCC CTCC ATGG TACG GAGK TCTG GCGT CCTM ACCG ACWC CTTG CCAC CAAC GTCTT GTGC CGGG AGAG	

							GGGA ACCT TGAA GATC GGGS CAAG CTGC ACTT GATG GTTA GTCA GTCG GGTG AAAT AATC CCAG CWGC CCCC GCTG TGAC TGCT AAGM CWGG	

							TCCC CAAG GGGC ATGA GGCA TSTGC GCTG AGCC GGSA GGGG TGAC ACMC GGCG ATCG GCGC AGCA CAKA STGA AGGG AGGC ACTT GCTG AGAC	

							TGCTT CTGA GGCC CCAG ACTT GGGG TGGT GCAG CCTT GTCT GGGG TATG GTAC AGCA CCCT ACTG CTCC CTTTG GKCA GCAG AATT CGTC CCGA	

							CCTCT TACC CACC CGAG TCTG CGCC TTGTT CCGC TATC CTGC ATCT CCGA TCCA CCTC GCTG TCTCC CCGC TGCG CTGC TTTTC TCTC AAGT GGGT	

							TAAA TCTTG TCAT GATT ACCT CCCA CGTTT CCGC TCAA GGGC AATG CCCA AMAT GACG GGGA TCGC TGGT GCAT GGCA GTCA TGAG ACCA TCCG	

							GACC CTCC GGTG GTCG CTAT AGTC ATTTT KTGT TGCA TGGG GCAT SCTG AGTC ACTT AACC GAAA GACT CWAA ATAA CTCA AAAG AGGK AMCC	

							TCTGS GGTT CGGT AAA (SEQ ID NO: 1333)	
R2	R2-1_DWi	.		Drosophila willistoni	GAAGCTGGGTCGGATGAGCGCAGAAGGGG TGTTCTTTGGAACACTGTAATTCATAAGTCG TAAGTCTGATCAAGTCGACTCGAAACCTCC TCGTGGTGTTTCCTGGGTGCTGTTGAGTTCC TAGTCTCTAGGTTCTTTTCAGTAGCTAATTC GAGCGGCGAAGCAACTCTTGGGGTTACCGG CCCCTTGAGCCAAGAAGCGTTGGTACAGAA TCAAATAATAATAGTCCTCGGAGCAATATC ACTATCACTTCAGCGACTTCACGTCCTGGAG ACCAACCGAGAGAGGCTATAGCAGTGGTAA ATCTCGCGGGAGAGATTCCCTGTGCAGTAT GCGGGCGCCTCTTCAATACTAGAAGGGGGC TCGGTGTACACATGTCACATCAACACAAAG ACGAACTAGATACGCAACGTCAGCGTGAAG ATGTAAAACTCCGATGGAGCGAGGAAGAA GCGTGGATGATGGCGAGAAAGGAGGTGGA	GAAGCT GGGTCG GATGAG CGCAGA AGGGGT GTTCTT TGGAAC ACTGTA ATTCAT AAGTCG TAAGTC TGATCA AGTCGA CTCGAA ACCTCC TCGTGG	TAGA TGTA CTAA CCTCT AGCT TTTCT TATA CTTTT GCCT GCTA CCTT GGCA TTAC ATCT AAAA AGGT	FERRSNSWGYR PLEPRSVGTESN NNSPRSNITITSA TSRPGDQPREAI AVVNLAGEIPCA VCGRLFNTRRG LGVHMSHQHKD ELDTQRQREDV KLRWSEEEAW MMARKEVELEA SGNLRFPNKKL AEVFTHRSSEAI KCFRKRGEYKA KLEQIRGQSTPT PEALDSITSQPRP SLLERNHQVSSS

					GCTCGAAGCAAGTGGTAATTTGAGATTTCC	TGTTTC	ACAA	EAQPINPSEEQS
					TAATAAGAAGCTAGCGGAAGTATTTACTCA	CTGGGT	ACAT	NWEIMRILQGY
					CCGTAGCTCCGAAGCAATTAAATGTTTTCG	GCTGTT	CGCA	RPVECSPRWRA
					GAAGAGGGGTGAATATAAGGCAAAACTGG	GAGTTC	TTGTC	QVLQTIVDRAQ
					AGCAGATCAGAGGGCAATCTACTCCCACCC	CTAGTC	ATAA	AVGKETTLQCL
					CAGAAGCGTTGGACTCTATTACCTCACAGC	TCTAGG	AGAG	SNYLLEVFPLPN
					CTCGCCCTAGTTTACTCGAGCGAAACCACC	TTCTTTT	GTGG	EPHTIGRSNLRR
					AAGTATCATCGTCGGAAGCGCAACCAATCA	CAGTAG	TTTTA	PRTRRQLRQQE
					ATCCATCAGAAGAACAGTCGAACTGGGAAA	CTAA	GTAC	YAQVQRRWDK
					TCATGCGGATACTACAGGGCTATCGCCCCG	(SEQ ID	GTAG	NTGRCIKSLLDG
					TAGAATGTAGTCCCCGGTGGAGAGCCCAGG	NO: 1211)	GCGC	TDESVMPNQEI
					TCTTGCAAACTATCGTAGATAGGGCGCAGG		TGTG	MEPYWKQVMT
					CCGTAGGGAAGGAAACCACTCTCCAATGCT		GGAC	NPSTCSCDNTRF
					TATCCAACTATCTCCTGGAAGTATTTCCATT		TTCAT	RMEHSLETVWS
					ACCAAACGAACCACACACCATCGGTCGGAG		TGTC	AITPRDLRENKL
					CAATTTGCGAAGACCTCGAACTAGGAGACA		CCGG	KLSSAPGPDGIT
					GTTAAGACAACAAGAGTACGCACAGGTTCA		TGAT	PRTARSVPLGIM
					GCGTCGTTGGGATAAGAATACTGGGAGATG		GCAG	LRIMNLILWCGK
					CATTAAATCCTTGCTTGATGGAACAGATGA		TGAA	IPFSTRLARTIFIP
					GTCGGTTATGCCAAACCAAGAGATAATGGA		TCGT	KTVTANRPQDF
					ACCCTATTGGAAACAAGTAATGACGAATCC		GCAT	RPITVPSVLVRQ
					CAGCACATGCTCTTGCGATAACACAAGATT		ACGA	LNAVLASRLAS
					CCGTATGGAACATTCGCTTGAGACGGTTTG		GATT	KVNWDPRQRGF

					GTCAGCGATAACGCCACGCGACCTGAGGGA		GTCC	LPTDGCADNAT
					AAATAAGTTAAAGTTGTCAAGTGCTCCGGG		AGTA	LVDLILREHHKR
					TCCTGACGGTATCACTCCAAGAACAGCCAG		GTTG	WKSCYLATVDV
					GAGTGTACCCTTAGGCATTATGCTACGCAT		GTTG	SKAFDSVSHQAI
					AATGAACCTGATTCTCTGGTGCGGCAAAAT		CTCG	IKTLQAYGAPTN
					ACCATTCTCTACCCGACTGGCCAGAACTATC		TATCT	FVSFIEEQYKGG
					TTCATTCCGAAGACTGTGACGGCAAATCGA		TTAG	GTSLNGAGWSS
					CCGCAAGACTTTCGTCCAATAACAGTCCCCT		AAGA	EVFIPARGVKQG
					CGGTTTTGGTCAGGCAATTAAACGCTGTTCT		TTTCC	DPLSPLLFNLIID
					GGCTTCTCGATTGGCTTCTAAAGTCAACTGG		TTCCT	RLLRSYPREIGA
					GATCCAAGGCAGCGCGGTTTCCTACCTACC		CGGC	KVGNTMTSAAA
					GATGGGTGTGCTGATAATGCGACGTTGGTT		GATC	FADDLVLFAETP
					GATCTCATTTTGCGGGAGCACCATAAACGG		AAAA	MGLQTLLDTTV
					TGGAAGTCATGTTACCTTGCGACGGTGGAT		AAAA	GFLASVGLSLNA
					GTCAGCAAGGCTTTTGACTCAGTATCACAC		AAAA	DKCFTVSIKGQA
					CAGGCCATTATCAAGACTTTACAGGCCTAT		AAAA	KQKCTVVERRS
					GGTGCTCCAACAAACTTTGTCAGCTTCATAG		AAAA	FCVGERECPSLK
					AAGAACAGTATAAGGGCGGCGGAACCTCCC		AAA	RTEEWKYLGIRF
					TCAATGGGGCAGGATGGAGTTCAGAGGTGT		(SEQ	TADGRARYSPA
					TTATACCCGCGCGGGGCGTTAAGCAAGGTG		ID NO:	DDLGPKLLRLTR
					ACCCTCTGTCTCCACTATTATTTAATCTTAT		1334)	APLKPQQKLFAL
					CATTGATAGATTACTTAGGTCCTACCCCAGA			RTVLIPQLYHQL
					GAGATTGGTGCCAAAGTCGGAAATACCATG			TLGSVMIGVLR

					ACAAGCGCGGCAGCGTTCGCGGATGATCTG			KCDRLVRQFVR
					GTGCTATTTGCGGAAACTCCGATGGGGCTT			RWLDLPLDVPV
					CAAACATTGTTGGATACCACGGTAGGCTTC			AYFHAPHTCGG
					CTAGCCTCCGTGGGACTCTCCCTTAATGCTG			LGIPSIRWIAPML
					ATAAGTGCTTCACTGTCAGTATAAAGGGGC			RLKRLSNIKWPH
					AAGCCAAGCAGAAGTGTACTGTCGTAGAAC			LEQSEVASSFID
					GACGGAGCTTTTGTGTAGGTGAGCGCGAGT			DELQRARDRLK
					GTCCTTCATTGAAGCGTACTGAAGAGTGGA			AENVQLCSRPEI
					AGTATTTAGGTATCCGGTTCACTGCGGATG			DSYFANRLYMS
					GGCGGGCTCGGTATAGTCCAGCAGACGACC			VDGCGLREAGH
					TCGGTCCGAAGCTGTTAAGATTAACAAGAG			YGPQHGWVSQP
					CCCCTCTGAAACCACAACAGAAGTTATTTG			TRLLTGKEYLH
					CACTTAGGACTGTCCTTATCCCACAACTCTA			GVKLRINALPSK
					TCACCAACTAACACTTGGGAGTGTGATGAT			SRTTRGRHELER
					AGGCGTCCTAAGAAAATGTGACAGATTGGT			RCRAGCDAPET
					ACGGCAATTCGTAAGGAGATGGTTAGATCT			TNHILQKCYRTH
					CCCACTGGATGTACCAGTTGCTTACTTTCAC			GRRVARHNSVV
					GCCCCCCACACTTGTGGGGGTCTCGGGATT			NAVKRGLERKG
					CCGTCAATTAGATGGATAGCACCGATGCTG			CVVHVEPSLQC
					CGTCTGAAGCGATTGAGCAATATTAAATGG			DSGLNKPDLVGI
					CCCCACCTCGAACAATCCGAGGTAGCTAGC			RQNHIYVIDVQV
					TCTTTCATTGACGACGAATTGCAAAGGGCT			VTDGHSLDQAH
					CGAGATAGATTAAAGGCGGAAAATGTGCAG			QRKVERYDRAD

					CTGTGTTCGCGTCCAGAGATTGACTCGTATT TCGCAAATAGATTGTACATGTCTGTTGATGG TTGCGGTCTCCGTGAAGCAGGTCATTATGG CCCGCAACATGGATGGGTGAGTCAGCCCAC GCGCTTGCTAACAGGAAAGGAATATTTGCA CGGTGTCAAATTGCGGATAAATGCCCTACC CTCGAAGTCTCGTACGACGAGGGGAAGGCA CGAATTGGAGAGACGGTGTCGTGCAGGATG TGATGCTCCCGAGACAACAAACCACATCTT GCAAAAATGCTATCGTACGCATGGGAGGCG GGTAGCTAGACACAACAGCGTAGTAAATGC CGTCAAGCGGGGACTTGAACGGAAAGGCTG CGTTGTCCATGTCGAACCAAGTCTGCAATG CGACTCGGGCTTAAATAAACCGGACCTGGT GGGAATCCGACAGAATCACATTTATGTGAT AGACGTTCAGGTTGTGACAGACGGACATTC CTTAGACCAAGCGCACCAGCGCAAGGTCGA AAGGTACGACAGAGCTGACATAAGATCACA AATGCGGCGATTTTTCGGAGCGACAGGTGA AATCGAGTTTCATTCCGTTACACTCAACTGG AGAGGAATCTGGAGTGGTCAGTCGGTAAAA CGATTGATTGCAAAGGATCTCCTCATCGCTG AAGATACCAAACTCATCAGCGTCAGAGCAG			IRSQMRRFFGAT GEIEFHSVTLNW RGIWSGQSVKR LIAKDLLIAEDT KLISVRAVNGG VTSFKYFMYCA GYTRS (SEQ ID NO: 1456)

					TAAATGGCGGAGTGACGTCCTTCAAATATT TCATGTATTGTGCTGGGTATACTCGAAGCTA GATGTACTAACCTCTAGCTTTTCTTATACTT TTGCCTGCTACCTTGGCATTACATCTAAAAA GGTACAAACATCGCATTGTCATAAAGAGGT GGTTTTAGTACGTAGGCGCTGTGGGACTTC ATTGTCCCGGTGATGCAGTGAATCGTGCAT ACGAGATTGTCCAGTAGTTGGTTGCTCGTAT CTTTAGAAGATTTCCTTCCTCGGCGATCAAA AAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1088)			
R2	R2-1_Gav	.		Gavialis gangeticus	AGGCATCTCCTTKAAGGGTAATGGTCTGGT TACATGGTCATAGCAGGTTTGTGTCAGGTA CCTCCCAGTGGTTCCCGCCGGGTGSCAMAG CCCCAGGGCTGTCGGTAGCTCGATCCTGGT ACAGTACGGCCAAGGGAGTTCTTCCTTGCT GTCGGGTGCCTCGCAAGCACKTGGCAGCCC CAATCGCTTCATTGCGAAAAACACAAACGT CCTAAGGGGATGATCAGCTAGTCAGTTCTG CCGCTAGCCAAAACTGTTTGCCACCCAGTT ACAGATAGCGTCTGTGCTGACCAGCTGCCC CGCGGGCTTGGGGTGCAGTGGAGGCCGGCC CGTGGCCAGGCCGGACACGGGCCGTGGAGC	AGGCAT CTCCTT KAAGGG TAATGG TCTGGT TACATG GTCATA GCAGGT TTGTGT CAGGTA CCTCCC AGTGGT	TGAG AAGT TGCG AGTT CTTAT GCAA GTTG AATA CCAC TCTK GKGA CCCC	PAAPRAWGAVE AGPWPGRTRAV EPAPSPESSPSEA ARAAPAGEGHG PGHESPSVQRPE ADTTAPGVSAPT REGEPPSTRVFL VRLPDSNPPCPI CRDHVGKPSAL ALHCVESHAWA DVQYQCTHCKK VSANKHSILCHI

					CTGCTCCCAGTCCAGAGAGTTCCCCCTCGG	TCCCGC	AAAA	PCCQGRVPEWT
					AAGCTGCCAGGGCAGCACCAGCCGGGGAG	CGGGTG	AAW	GKDWACPECPA
					GGCCACGGCCCCGGTCATGAGTCCCCCTCG	SCAMAG	WAAA	SFNKKVGLSQH
					GTGCAGAGGCCTGAGGCCGATACCACTGCC	CCCCAG	CCCC	KRHVHPVTRNV
					CCTGGTGTGAGCGCGCCCACCAGGGAGGGT	GGCTGT	AAAA	ERVAGSLSRAG
					GAACCACCCTCCACCAGGGTTTTCCTGGTG	CGGTAG	CAGT	LRPQTRRGCWS
					AGGCTGCCTGATTCAAACCCGCCGTGCCCA	CTCGAT	TGTG	VEEEETLTCLDA
					ATTTGCAGGGACCATGTGGGTAAACCCTCC	CCTGGT	TTTA	MFRGARNINQLI
					GCGCTGGCCCTCCACTGCGTGGAGAGCCAT	ACAGTA	AGTG	AAEMVTKMPK
					GCGTGGGCGGATGTGCAATACCAGTGCACC	CGGCCA	TGTTC	QISDKRRQLGLC
					CATTGCAAAAAGGTCAGTGCTAACAAACAC	AGGGAG	TTGTT	PEQTTLGGDAES
					AGCATCCTCTGCCACATCCCATGTTGCCAGG	TTCTTC	CGTC	TSVVEEESMTPE
					GGAGGGTGCCCGAGTGGACCGGGAAGGAC	CTTGCT	CCTTT	METQSPINPPGK
					TGGGCTTGTCCTGAGTGCCCTGCCTCCTTCA	GTCGGG	GGCT	IRKILAQRARRW
					ATAAGAAAGTTGGCCTATCGCAGCACAAGC	TGCCTC	TCAC	LKKGQGLSDKV
					GGCACGTGCATCCGGTAACACGTAATGTGG	GCAAGC	CTCSA	REVLGAWVEGQ
					AACGGGTTGCAGGGAGCCTATCGAGGGCTG	ACKTGG	AGTT	PRIHAWVDSVSL
					GTTTAAGGCCCCAGACCAGACGCGGGTGCT	CAGCCC	GCGA	DVLTLFLGVPSG
					GGTCGGTGGAAGAAGAGGAAACTCTCACCT	CAATCG	TCCC	PQRAPNKKRPK
					GCCTAGACGCGATGTTCCGTGGTGCCCGGA	CTTCAT	CCCA	EGGKPTSWMNK
					ACATCAACCAGCTGATCGCTGCTGAAATGG	TGCGAA	TCTCC	CAVKWGTFLRY
					TAACGAAGATGCCTAAGCAGATCAGTGACA	AAACAC	CCTG	QHLFGANRKLL
					AACGGAGGCAGCTCGGTCTGTGTCCTGAGC	AAACGT	CGCT	VAIVLDGADRN

					AGACCACATTGGGTGGTGATGCTGAATCGA	CCTAAG	GCCT	QCTLLLEEVFQA
					CCTCCGTGGTGGAGGAAGAGTCCATGACTC	GGGATG	TTCA	YREKWGLEEVL
					CGGAGATGGAAACCCAAAGCCCAATTAACC	ATCAGC	GAAC	RAYRGKWEVES
					CGCCTGGGAAAATCAGGAAGATCCTGGCCC	TAGTCA	GGCC	SFEGLGRFGVRR
					AGAGGGCACGCCGGTGGCTGAAGAAGGGG	GTTCTG	GGTG	DADNFAFKALIT
					CAGGGTCTGTCGGACAAGGTGCGAGAAGTC	CCGCTA	GTGT	PEEVVKHMMA
					CTGGGCGCATGGGTGGAGGGTCAACCCAGG	GCCAAA	CGAG	MASKSAPGPDK
					ATTCATGCCTGGGTCGACTCAGTCTCCCTTG	ACTGTT	GCTG	LTLRDLRRADPE
					ATGTTTTGACTTTGTTCTTGGGGGTGCCCTC	TGCCAC	GCGC	GDALAELFSLW
					AGGACCGCAGAGGGCTCCGAACAAGAAGA	CCAGTT	GACC	LITGTVPDGLKE
					GGCCCAAGGAGGGTGGCAAACCAACGTCCT	ACAGAT	TCGG	CRSVLIPKTVDR
					GGATGAACAAATGTGCCGTCAAATGGGGCA	AGCGTC	TCAC	EKLGQLGNWRP
					CATTCCTTCGGTACCAGCACCTGTTTGGTGC	TGTGCT	CTCC	ITIGSIVLRLFSR
					CAACAGGAAGCTCCTGGTGGCSATCGTCCT	GA (SEQ	AAGG	VLTARLAAACPI
					GGACGGCGCTGACCGTAATCAGTGCACCCT	ID NO:	CCAA	NPRQRGFIAAPG
					CCTGCTAGAGGAGGTCTTCCAGGCCTACCG	1212)	GTGC	CAENLKVLELLL
					AGAAAAGTGGGGGCTAGAGGAAGTCCTTCG		CCTG	RKRKRDRQPLG
					GGCCTACCGAGGAAAGTGGGAGGTAGAGTC		GCCC	VVFVDLARAFD
					ATCTTTTGAGGGTCTCGGACGGTTCGGGGTC		CGAG	SVSHDHISWVL
					CGCCGGGATGCGGATAACTTCGCATTCAAG		TAGG	KAKGVDEHIVN
					GCCCTGATCACTCCTGAGGAGGTTGTCAAA		ACTG	LIEDSYQKVTTR
					CACATGATGGCAATGGCCTCGAAATCGGCT		AGTG	VQVFNGVTPPISI
					CCAGGTCCGGATAAGCTCACCCTGAGAGAT		GCCC	KTGVKQGDPMS

					CTGCGCCGCGCTGACCCCGAGGGAGATGCT		AGCT	PLLFNIAMDPLI
					CTTGCCGAACTCTTCAGCCTGTGGCTGATTA		CGCT	AKLETDGQGVK
					CCGGCACGGTCCCGGACGGACTCAAGGAGT		GGGC	VGSASLTTLAFA
					GTCGGTCTGTGTTGATACCCAAAACGGTGG		ACCC	DDLVLLSDSWE
					ACCGGGAGAAGTTGGGCCAGCTGGGCAACT		GTCA	GMLKNISILEDF
					GGCGCCCTATCACGATTGGGTCCATCGTTTT		CCAT	CNLTGLRVQPK
					ACGGCTATTCTCACGAGTGCTAACCGCACG		CTGG	KCQGFFLNPTCD
					GCTCGCCGCAGCATGTCCCATCAACCCCCG		GGCA	SFTVNNCEAWK
					TCAGAGGGGTTTCATAGCGGCGCCGGGGTG		AATG	IAGREITMLGPG
					CGCCGAGAACCTGAAGGTGCTTGAGCTTCT		GAAG	ESTRYLGLNVGP
					CTTGCGGAAGAGGAAGCGAGACAGGCAGC		GGAT	WVGIDKPDLGT
					CGTTGGGTGTGGTATTTGTGGATCTAGCGA		CTGT	QLSSWLERIGTA
					GGGCGTTCGATTCGGTGTCACACGATCACA		CCTG	PLKPMQKLSLL
					TTTCTTGGGTCCTAAAGGCCAAAGGGGTGG		ACCA	VQYAIPRLNYQ
					ACGAACACATCGTGAATCTCATCGAAGATT		CTAC	ADYAGIGRVAL
					CTTACCAGAAGGTTACCACGAGAGTACAAG		CAGG	EALDSMNRRKV
					TGTTCAATGGCGTCACCCCTCCTATCAGCAT		CTAA	KEWFHLPACTS
					CAAAACCGGGGTTAAGCAAGGCGACCCGAT		GTGT	DGLLHSRHRDG
					GTCCCCCCTCTTGTTCAACATTGCGATGGAC		GGTG	GLGLPRLAKAIP
					CCCCTGATAGCGAAGCTCGAGACAGACGGA		CGGC	EAQVRRLIRVAT
					CAGGGAGTAAAAGTCGGGAGTGCCTCCCTG		CTAG	SSDEVTRKVSY
					ACCACCCTGGCCTTCGCGGATGATCTCGTCC		CCTG	ACGISDEVERL
					TGCTTAGCGACTCTTGGGAGGGCATGCTGA		CCGT	WLARGGDMSSV

					AGAACATCAGCATCCTAGAGGACTTCTGCA		AAGG	PRFEDPEAPRSP
					ACCTCACGGGCCTACGAGTGCAACCCAAAA		TCAA	GVQGPCEAAQEI
					AATGTCAGGGGTTCTTCTTGAATCCGACATG		GCGC	PSVVRKLAIPRP
					CGACTCCTTTACGGTGAACAACTGCGAGGC		CCTG	SNWRSKKHSKW
					CTGGAAGATAGCCGGCCGTGAGATCACGAT		CTGC	AQLSCQGEGME
					GCTCGGACCAGGCGAGTCGACACGATATCT		CACT	LFCNDPVSNGW
					GGGCTTGAATGTCGGTCCTTGGGTTGGGAT		CAGG	NNSRGQLAEHL
					CGACAAACCAGATTTGGGTACGCAACTAAG		TATC	QIVALKLRSNIY
					CTCCTGGCTCGAGAGGATAGGGACTGCTCC		AGTC	PTREFLGRSQAS
					ACTCAAACCGATGCAGAAGCTCTCTTTGCT		CTCG	TNVGCRHCTHP
					GGTGCAGTATGCCATACCCAGGCTGAACTA		TTCA	HETLGHILGICP
					TCAGGCCGATTACGCGGGCATCGGCAGGGT		CTTGT	AVQEARIIRHNK
					GGCCTTGGAGGCTCTGGATTCTATGAACCG		CCCT	LCKILAAEGKKC
					GAGAAAGGTAAAGGAGTGGTTCCATCTTCC		CCTA	EWTVYYELQLL
					CGCCTGTACCTCGGACGGTCTCCTCCACTCC		GTAC	NAAGELCKPDLI
					CGTCACCGTGACGGGGGTCTTGGGTTACCG		CCTCT	FVRDGTXLVVN
					CGTCTGGCGAAAGCCATTCCGGAAGCGCAA		GCCT	VTVGYEGGPAX
					GTGAGGAGGCTGATCCGCGTAGCCACTTCA		CTGC	LLSTAAEKATK
					TCTGATGAAGTCACGCGGAAAGTATCCTAC		TCTTT	YLDXNAQIQELT
					GCGTGTGGGATAAGTGACGAAGTGGAGCGG		TGCT	GAEQVTYFGFPI
					CTCTGGTTGGCGAGGGGTGGGGACATGTCC		ATCC	GARGKWHADN
					AGTGTACCGAGGTTCGAGGATCCTGAGGCC		ACTA	XRVLSELGLSNS
					CCGAGGTCTCCGGGGGTGCAGGGCCCCTGC		TGGC	RKERVARLLXW

					GAGGCTGCCCAGGAGATTCCGAGCGTAGTC		CAGT	RALLGSVDMVN
					CGGAAGCTTGCGATCCCCCGGCCCTCCAAC		GATG	IFASKHRQENLS
					TGGAGATCCAAGAAACACTCCAAATGGGCC		TTGA	DXALSPS (SEQ
					CAACTCAGCTGTCAGGGAGAGGGGATGGAG		GGTT	ID NO: 1457)
					TTATTCTGCAATGATCCAGTCAGCAATGGCT		GGTG	
					GGAACAACAGTCGGGGACAACTGGCGGAA		CATC	
					CACCTCCAGATCGTGGCCTTAAAACTGCGTT		CTTG	
					CAAACATTTATCCCACCAGGGAGTTTCTTGG		GTCA	
					GAGAAGCCAGGCAAGTACCAATGTAGGTTG		CCTC	
					CCGGCATTGTACACACCCTCATGAAACACT		CAGG	
					AGGGCATATCTTGGGCATATGCCCTGCCGT		GCCA	
					GCAGGAGGCACGGATCATCCGGCATAACAA		AGCG	
					GCTGTGCAAGATCCTAGCAGCTGAGGGCAA		CCTT	
					AAAGTGTGAGTGGACAGTGTATTATGAACT		GGCC	
					GCAACTGCTTAACGCTGCAGGGGAACTGTG		ACAG	
					TAAACCTGACCTCATTTTTGTCCGAGACGGT		GTAG	
					ACCGMTCTGGTTGTGAATGTCACTGTGGGG		GACC	
					TACGAAGGGGGCCCCGCAWCCCTCCTATCC		TGGC	
					ACCGCTGCAGAAAAGGCCACAAAATACCTG		ACCT	
					GATMTGAACGCACAGATCCAGGAGCTCACA		GCCC	
					GGGGCTGAGCAGGTCACCTACTTTGGCTTC		AGGG	
					CCTATWGGAGCCAGGGGAAAGTGGCATGC		GGCC	
					TGACAACTGKCGAGTACTGTCTGAACTGGG		AGAC	

					ATTGTCCAACTCCCGGAAGGAGCGGGTCGC		ACTG	
					ACGGCTCCTGTMGTGGCGAGCACTGCTCGG		CCTG	
					GTCAGTGGACATGGTGAACATCTTTGCATCT		TGGC	
					AAGCACAGGCAGGAAAACCTATCGGATGA		AAGG	
					MGCACTGAGCCCCAGCTGAGAAGTTGCGAG		GAAA	
					TTCTTATGCAAGTTGAATACCACTCTKGKGA		GGGA	
					CCCCAAAAAAWWAAACCCCAAAACAGTTG		GCCG	
					TGTTTAAGTGTGTTCTTGTTCGTCCCTTTGG		TCCCT	
					CTTCACCTCSAAGTTGCGATCCCCCCATCTC		GACC	
					CCCTGCGCTGCCTTTCAGAACGGCCGGTGG		GTTA	
					TGTCGAGGCTGGCGCGACCTCGGTCACCTC		CCAG	
					CAAGGCCAAGTGCCCTGGCCCCGAGTAGGA		GCTT	
					CTGAGTGGCCCAGCTCGCTGGGCACCCGTC		GAGA	
					ACCATCTGGGGCAAATGGAAGGGATCTGTC		TGGT	
					CTGACCACTACCAGGCTAAGTGTGGTGCGG		GCTG	
					CCTAGCCTGCCGTAAGGTCAAGCGCCCTGC		GCTA	
					TGCCACTCAGGTATCAGTCCTCGTTCACTTG		GCCC	
					TCCCTCCTAGTACCCTCTGCCTCTGCTCTTTT		ACCA	
					GCTATCCACTATGGCCAGTGATGTTGAGGTT		TATG	
					GGTGCATCCTTGGTCACCTCCAGGGCCAAG		TCAA	
					CGCCTTGGCCACAGGTAGGACCTGGCACCT		GCAC	
					GCCCAGGGGGCCAGACACTGCCTGTGGCAA		TCCA	
					GGGAAAGGGAGCCGTCCCTGACCGTTACCA		CAGC	

					GGCTTGAGATGGTGCTGGCTAGCCCACCAT		TGCTT	
					ATGTCAAGCACTCCACAGCTGCTTGAGTTTG		GAGT	
					TTGGCTTCACCTTCATCCCACCTAGTGTCTT		TTGTT	
					CTGCCTCTGCACTATTTTCATCCCAACTCGT		GGCT	
					ACCTCCCCATCTCTGCGCTCCTGCTATCCCT		TCAC	
					GCAAGGACCAAGTAGGCAGGGGGGTTCATC		CTTC	
					CCCCTACCTGCAGGAGACTCAGCATATCCA		ATCC	
					TGACTTCTTGCCTCCACCGTCTTGTGGCGCT		CACC	
					AGAGGGGTACCTCAGAGACCGGCACAACAT		TAGT	
					GACCTTGACGGTTAGACAGTAGGGTCAAAC		GTCTT	
					AACCCTGCTGCAGGCCCAAAGGGCCAACAG		CTGC	
					CTGTGCCACGAGAGGGGAACCTTGAAGACT		CTCT	
					GGGGCAGTCTGACCATGCTGGTTAGTCAGT		GCAC	
					TGGGTGAAATAATCCCAGCTGCAGGCCCAA		TATTT	
					AAGGGCTGACAGTCAGGTGAGGGGGTATCT		TCAT	
					CCATCTGCTCCCCACTGCCAACTACGGAGG		CCCA	
					CATGAAGTCCGTAGTGACTTCTGACCCCCA		ACTC	
					CGTCTTGTGCCATGAGAAGGGAACCTTGAA		GTAC	
					GATTGGGACAAACCGCACTTGAAAGTTACT		CTCC	
					CAGCCGGGTGAAAATAAGTCCCAGTTGCGG		CCAT	
					GCCCCTCGGGGCTGACAGTCAGGTGAGGAG		CTCT	
					GGCTGCAAAGCCCATCTCCTGACTCCAGAG		GCGC	
					GCCTGGCGTCCTAACCGACTTCTTGCCACCA		TCCT	

					ATGTCTTGCGCCAGGAGAGGGCAACCTTGA		GCTA	
					AGATCGGGGCAAGCCGCACTTGATAGTTAG		TCCCT	
					CCAGTCGAGTGAAACAATCTCAGCTGCGGG		GCAA	
					TCCGAAAGGACTGACTTCCAGGCGAGGGGG		GGAC	
					GGGCCTGCGGAAAACCCCCTCCATGGTACG		CAAG	
					GAGGTCTGGCATCCTAACCGACACCTTGCC		TAGG	
					ACCAATGTCTTGTGCCAGGAGAGGGGAACC		CAGG	
					TTGAAGACTGGGGCAAGCCGCAGTTGATGG		GGGG	
					TTAGTCAGTCGGGTGAAATAATCCCGGCTG		TTCAT	
					CACCCTGCTGTGACTGCTAAGCCCGGTCCCC		CCCC	
					AAGGGGCATGAGGCATGTGCGCTGAGACGG		CTAC	
					GAGGGGTGACATCTGGCGATCAGCACAGCA		CTGC	
					CAGACTGAAGGGAGGCACTTGCCGAGAATG		AGGA	
					CTTCTGAGGCCCCAGACTTGGGGTGGTGCA		GACT	
					GCTTTGTCTCGTGTATAGTACAGCACCCTAC		CAGC	
					TGCTCCCTTTGGGCAGCAGAATTTGTCCTGA		ATAT	
					CCTCTTACCCACCCGAGTCTGCGCTTTTGTT		CCAT	
					CCACCTCGCTGTCTCCCTGCTGTGCTGTTTT		GACT	
					TCTCTCAAGTGGGTTAAATCTCAACATGATT		TCTTG	
					ATCTCCCACGTTTCCGCTCAAGGGCAATGCC		CCTC	
					CAACATGACGGAGATCGTTGGTGCATGGTA		CACC	
					GTCACGAGACCATCCGGACCCTCCAGTGGT		GTCTT	
					CGCTATAGTCATTTTGTGTTGCATGGGGCAT		GTGG	

					GCTGAGTCACTTAACCGAAAGACTGTAAAT AACTCAAAAGAGGTACCCTCCGGGGTTCGG TAAA (SEQ ID NO: 1089)		CGCT AGAG GGGT ACCT CAGA GACC GGCA CAAC ATGA CCTT GACG GTTA GACA GTAG GGTC AAAC AACC CTGC TGCA GGCC CAAA GGGC CAAC	

							AGCT GTGC CACG AGAG GGGA ACCT TGAA GACT GGGG CAGT CTGA CCAT GCTG GTTA GTCA GTTG GGTG AAAT AATC CCAG CTGC AGGC CCAA	

							AAGG GCTG ACAG TCAG GTGA GGGG GTAT CTCC ATCT GCTC CCCA CTGC CAAC TACG GAGG CATG AAGT CCGT AGTG ACTT CTGA CCCC CACG	

							TCTTG TGCC ATGA GAAG GGAA CCTT GAAG ATTG GGAC AAAC CGCA CTTG AAAG TTACT CAGC CGGG TGAA AATA AGTC CCAG TTGC GGGC CCCT	

							CGGG GCTG ACAG TCAG GTGA GGAG GGCT GCAA AGCC CATC TCCT GACT CCAG AGGC CTGG CGTC CTAA CCGA CTTCT TGCC ACCA ATGT CTTG	

							CGCC AGGA GAGG GCAA CCTT GAAG ATCG GGGC AAGC CGCA CTTG ATAG TTAG CCAG TCGA GTGA AACA ATCT CAGC TGCG GGTC CGAA AGGA	

							CTGA CTTCC AGGC GAGG GGGG GGCC TGCG GAAA ACCC CCTC CATG GTAC GGAG GTCT GGCA TCCT AACC GACA CCTT GCCA CCAA TGTCT TGTG	

							CCAG GAGA GGGG AACC TTGA AGAC TGGG GCAA GCCG CAGT TGAT GGTT AGTC AGTC GGGT GAAA TAAT CCCG GCTG CACC CTGC TGTG ACTG	

							CTAA GCCC GGTC CCCA AGGG GCAT GAGG CATG TGCG CTGA GACG GGAG GGGT GACA TCTG GCGA TCAG CACA GCAC AGAC TGAA GGGA GGCA	

							CTTG CCGA GAAT GCTT CTGA GGCC CCAG ACTT GGGG TGGT GCAG CTTTG TCTC GTGT ATAG TACA GCAC CCTA CTGC TCCCT TTGG GCAG CAGA	

							ATTT GTCC TGAC CTCTT ACCC ACCC GAGT CTGC GCTTT TGTTC CACC TCGC TGTCT CCCT GCTG TGCT GTTTT TCTCT CAAG TGGG TTAA ATCT CAAC	

							ATGA TTATC TCCC ACGT TTCC GCTC AAGG GCAA TGCC CAAC ATGA CGGA GATC GTTG GTGC ATGG TAGT CACG AGAC CATC CGGA CCCT CCAG	

							TGGT CGCT ATAG TCATT TTGT GTTG CATG GGGC ATGC TGAG TCAC TTAA CCGA AAGA CTGT AAAT AACT CAAA AGAG GTAC CCTC CGGG GTTC	

							GGTAAA (SEQ ID NO: 1335)	
R2	R2-1_IS	.		Ixodes scapularis	GTTCCAAAGGAAGGCACTCCTTTGGTTCGT GATGAGATGTTCATGGTGCTTGCCTAGCTG GAGAAATCCGACTCACACCTGCACGTGGTC CCTGCCGCCTGCCAGTATGCCGAGGAAACG GGTGCAACTTAATCCGTGGATACTGGTAGC AACGTGAGCAACGGTACGGTCCTTCGCGGA CCACCCTGGGCGTTCGGGTTGCCAGCCCGTT CGCCCGAAATATCTTGGCCCTGAAACTAAA AGAAAAATGCAATGCACCAGCCGACTGGCT GATGCACCAAGATTTGCCCGAGTGGGCGTC GAGGGTGAAGGTGTCGGTGCGTCTGGTAAC GGCACTGATGCGCAGTTATGGTATGGCTGC ACGGGCTGTGACGAAGCCTTCTCGTCCCTG CGAGGACTGAGAATTCATGCGGCCCAAAAG AAACATGGAAACCAAGATGGCCTTCTCCGC CTGCCGGCGGGACGGCCCCGAAAACGACGA GTGGGGAAGAGCACCACAGCGGGTGCTTCG GACCGGGTGACCACGGATCCAGTGCCTGCT	GTTCCA AAGGAA GGCACT CCTTTG GTTCGT GATGAG ATGTTC ATGGTG CTTGCC TAGCTG GAGAAA TCCGAC TCACAC CTGCAC GTGGTC CCTGCC GCCTGC CAGTAT	TAGT GTGA CGGA GTCC TCAA GCCC CCAC AAGT GCCT GCCA GGTG GCAG GAAA GGGC AACT ACTG GTGA GCGA	MQCTSRLADAP RFARVGVEGEG VGASGNGTDAQ LWYGCTGCDEA FSSLRGLRIHAA QKKHGNQDGLL RLPAGRPRKRR VGKSTTAGASD RVTTDPVPAPVP ESPGLLPGLPGP SLPGCSDLPPGV LPGGWSASPGPL SWPPSLDAGPLP GPSRVSPGPSRP SPGKPTGPPSLD AGPLPGPSRVSP GPSRPSPGKPPG TPEPLPGSPGGR

					CCAGTTCCTGAATCTCCTGGGCTGCTGCCTG	GCCGAG	CCCA	RGVSPGQPGSRT
					GGCTACCTGGACCATCGCTGCCTGGGTGCT	GAAACG	AGCA	DPSSSAGAGHFV
					CGGACCTGCCGCCTGGGGTGCTGCCTGGAG	GGTGCA	AGGC	CPQCSRAFSSKI
					GGTGGTCTGCATCACCTGGGCCTCTCTCCTG	ACTTAA	GGAG	GMSQHQKHAHL
					GCCTCCTTCCCTGGATGCCGGGCCTCTGCCT	TCCGTG	CCAA	EEYNAGINITRT
					GGACCTTCAAGAGTATCACCTGGACCTTCA	GATACT	GACC	KARWDPEETYL
					AGACCTTCGCCCGGGAAGCCGACCGGGCCT	GGTAGC	AAGC	LARLEATLNPD
					CCTTCCCTGGATGCCGGGCCTCTGCCTGGAC	AACGTG	TGGA	HKNINQTLHAA
					CTTCAAGAGTATCACCTGGACCTTCAAGAC	AGCAAC	GCCA	LPRGSCRTLESI
					CTTCGCCCGGGAAGCCGCCCGGGACGCCTG	GGTACG	AGAG	KAHRKQAAYRD
					AGCCGCTGCCTGGATCTCCTGGCGGTCGGC	GTCCTT	CAAC	LVTSLRSARESS
					GCGGGGTGTCCCCGGGACAGCCCGGGTCAC	CGCGGA	TCCA	EAQHVPDRPLET
					GGACCGACCCCTCAAGCTCTGCTGGCGCCG	CCACCC	GGAG	PEPQTPANPQRD
					GACACTTCGTATGCCCGCAGTGCAGCAGAG	TGGGCG	GCAG	SKQAVIEALQSL
					CCTTCTCAAGCAAGATTGGCATGTCTCAAC	TTCGGG	GGGT	IGRAPPGSFQGA
					ACCAAAAACATGCCCACCTCGAAGAATACA	TTGCCA	GGAT	RLWDIARQATR
					ACGCGGGCATTAACATCACCCGTACCAAAG	GCCCGT	ATCA	GTNILPLLNSYL
					CCCGGTGGGACCCCGAGGAGACCTATCTTC	TCGCCC	AGAG	RDVFTLPTKPTR
					TGGCCCGCCTGGAGGCCACCCTCAACCCAG	GAAATA	CAAC	KKPAVRPARSR
					ACCACAAGAACATCAACCAGACGCTGCACG	TCTTGG	CCCA	RKQKKQEYART
					CCGCGCTGCCCCGCGGTTCCTGTCGAACCCT	CCCTGA	AGGG	QDLFRKKQSDC
					GGAGAGCATCAAGGCCCACCGAAAGCAGG	AACTAA	ACAC	ARAVLDGPTSSS
					CGGCTTACAGGGACCTGGTGACGAGCCTGC	AAGAAA	AGAC	VPGTGAFLQTW

					GGTCAGCCAGGGAGAGCAGCGAGGCGCAG	A (SEQ	CACG	REIMTGPSPALE
					CACGTTCCGGACCGGCCCCTGGAGACCCCG	ID NO:	GGCA	APPLPTRGEVDL
					GAGCCCCAGACACCAGCGAACCCTCAAAGA	1213)	ACTA	FFPATAQEIQSA
					GACTCGAAGCAGGCAGTCATCGAAGCGCTG		CTGG	EIAVNSAAGPDG
					CAATCCCTCATCGGCCGAGCACCACCAGGC		TGAG	FSARLLKSVPAL
					TCCTTCCAGGGAGCGCGTCTCTGGGACATC		CGCC	LLRVMVNLLLL
					GCGAGGCAAGCCACAAGGGGGACGAACAT		CAAG	VRRVPAALRDA
					CCTCCCACTCCTGAACAGCTACCTGAGGGA		ACAG	RTTFIPKVPDAV
					TGTTTTCACCCTCCCCACAAAGCCAACAAG		GGGT	DPSQFRPITVAS
					AAAGAAACCTGCAGTGCGGCCCGCCCGGAG		GGAT	VLQRLLHRILAK
					CCGCAGAAAACAAAAGAAACAAGAGTATG		ATTA	RALEAIPLNFRQ
					CCAGAACACAAGATCTATTCAGGAAAAAGC		AGAA	RAFQPVDGCAE
					AGTCCGACTGTGCCAGGGCGGTCCTGGACG		CAGC	NIWLLSTALNEA
					GCCCCACGTCGTCATCGGTCCCTGGAACGG		CCCA	RTRRRPLHMAS
					GCGCCTTCCTGCAAACCTGGCGAGAGATCA		CAAA	VDLTKAFDRVT
					TGACGGGGCCCAGCCCTGCACTCGAGGCAC		GTGT	TDAILRGARRA
					CGCCTCTACCTACCCGGGGGGAAGTCGACC		TACC	GLSGEFIGYLKE
					TGTTCTTCCCGGCGACGGCGCAAGAGATCC		TATA	LYTTSRTLLQFQ
					AGAGCGCTGAGATAGCCGTCAACTCGGCTG		TTAA	GESLLVEPTTGV
					CTGGACCCGACGGGTTCTCAGCCCGTCTCCT		CAAT	RQGDPLSPILFN
					CAAGTCCGTCCCGGCCCTCCTCCTAAGGGTC		AAAG	LVLDEYLSSLDP
					ATGGTTAACCTTCTGCTCCTCGTCCGACGTG		TTGA	DISFVSGDLRLD
					TCCCGGCGGCCCTCCGGGACGCGAGAACGA		AGCC	AMAFADDLIVF

					CCTTCATCCCGAAGGTCCCCGATGCAGTGG		TCAA	ASTPAGLQDRL
					ACCCCTCCCAATTTCGCCCAATAACGGTGG		CCAC	DALVEFFDPRGL
					CCTCCGTTCTCCAGCGCCTACTACATCGCAT		GCAT	RVNVKKSFTLSL
					CCTGGCCAAGAGGGCGCTGGAGGCCATTCC		TGCG	QPGRDKKVKVV
					CCTCAACTTTCGACAAAGAGCCTTTCAGCC		GGTT	CDQIFTIGGTPLP
					GGTGGATGGCTGTGCCGAGAATATATGGCT		AGAT	ASKVATPWRYL
					GCTGTCCACCGCGCTCAACGAGGCAAGAAC		GGCG	GMTFTPQGSINK
					CAGACGGCGCCCGCTACACATGGCGAGCGT		TGGC	GTSEQLDLLLTR
					CGACCTAACCAAGGCATTCGACCGGGTCAC		TTGG	TSKAPLKPQQRL
					CACGGATGCCATCCTGAGGGGCGCAAGGCG		CCCG	VVLRNYLLPRL
					CGCCGGGCTGTCCGGGGAGTTCATCGGATA		CCGC	YHRLVLGPWSA
					CCTGAAGGAGCTCTACACAACATCCAGGAC		CATG	ALLLKMDTTIRG
					CCTCCTGCAGTTCCAGGGAGAGAGCCTGCT		ATGA	AIRRWMDLPHD
					TGTCGAACCCACGACCGGCGTGCGACAGGG		GCTG	TPLGFFHAPVTE
					CGACCCACTGTCGCCCATCCTCTTCAACCTG		GAAC	GGLGINSLRASIP
					GTCCTGGACGAGTACCTCTCCTCCCTGGACC		CCTC	AMVLQRLDGLH
					CGGACATCTCCTTCGTCTCGGGCGACTTGCG		CACC	FSTHPGAEVAIQ
					CCTCGATGCGATGGCATTCGCTGACGACTT		TGGT	LPFLTGLHRRAE
					GATCGTCTTTGCCTCAACCCCAGCCGGCCTG		GGGC	AAAQYQGQRLL
					CAGGATCGGCTCGATGCCCTCGTCGAGTTCT		CGCA	SKADVHRMWS
					TCGACCCAAGGGGGCTCAGGGTGAACGTGA		CGAG	ARLHGSCDGRP
					AGAAGAGCTTCACGCTATCGCTGCAGCCGG		ACCA	LRESKRVPAAH
					GACGAGACAAGAAGGTCAAGGTGGTGTGTG		CCGG	RWAAEGTRLLS

					ACCAGATCTTCACCATCGGAGGAACCCCAC		CTCTT	GRDFISITKLKIN
					TCCCAGCCTCCAAGGTCGCAACCCCTTGGC		TCTA	ALPTLERTSRGQ
					GCTACCTGGGGATGACCTTCACCCCCCAGG		CTAA	HKDIQCRAGCQ
					GCTCAATCAACAAGGGCACCAGCGAGCAGT		GGCC	AVESLGHVLQA
					TGGACCTACTGCTCACGAGAACCAGTAAGG		GGTC	CHRGHRGRIRR
					CCCCCCTCAAGCCACAACAGAGGCTGGTGG		TCCG	HDNIARYVCGR
					TCTTAAGAAACTACCTGCTCCCGAGGCTAT		TGAC	LTQIGWAVKWE
					ACCATCGCCTCGTGCTTGGACCTTGGTCGGC		TGCG	PHYSVAGRTLK
					CGCCCTCCTACTGAAGATGGACACCACCAT		GTTG	PDIVAHRGAETV
					TCGAGGAGCCATTAGACGCTGGATGGATCT		GGAT	VLDAQVVGTSM
					CCCGCACGACACACCGCTGGGTTTCTTCCAC		AAAC	RLGFHHAQKKE
					GCCCCAGTAACGGAGGGAGGCCTAGGAATC		TCCA	KYSLPDLLHQV
					AACTCCCTGCGAGCATCAATTCCAGCCATG		AGCA	CEGRRDAARVS
					GTGCTCCAACGGCTGGATGGACTTCACTTC		CTGA	TITLNFRGVWAP
					AGCACGCATCCCGGAGCTGAGGTCGCCATC		GCGG	ESAQDLKSLGLT
					CAGCTGCCGTTCCTGACAGGACTCCATCGA		TAAA	DNDLKLLTVRC
					AGAGCGGAAGCGGCGGCCCAATACCAGGG		AAAA	LQGGAQCFRLH
					ACAGAGACTACTGTCCAAAGCGGACGTCCA		AAAA	RRMTTVVKATG
					CCGGATGTGGAGCGCAAGACTCCACGGGAG		AAAA	DEANALPAHSG
					CTGCGACGGAAGACCCCTTCGGGAGTCCAA		AAAA	LPPTQLGGRTLG
					GAGAGTGCCGGCTGCCCATCGTTGGGCCGC		AAAA	PSAHNQSARTT
					GGAAGGCACCAGACTACTCTCGGGAAGGGA		AAAA	(SEQ ID NO:
					CTTCATCTCGATCACGAAACTCAAGATAAA		(SEQ	1458)

					CGCGCTACCTACACTCGAGCGCACCAGCCG GGGCCAGCACAAGGACATCCAGTGCAGAGC TGGCTGCCAGGCTGTTGAATCCCTGGGCCA CGTCCTACAAGCTTGCCATCGAGGACACCG TGGCCGAATCCGGCGGCATGATAACATTGC CCGCTACGTCTGCGGCCGACTGACCCAGAT TGGCTGGGCGGTGAAGTGGGAGCCCCACTA CTCTGTCGCTGGAAGGACCCTCAAACCTGA CATCGTTGCCCATCGTGGAGCCGAGACTGT CGTGCTCGACGCCCAGGTCGTCGGCACCAG CATGCGACTGGGCTTCCACCACGCTCAAAA GAAAGAAAAGTACTCTCTCCCAGACCTCCT CCACCAAGTCTGCGAGGGACGGAGAGACGC AGCCCGGGTGTCAACAATCACCCTCAATTTT CGAGGTGTTTGGGCACCTGAGAGCGCCCAG GACCTGAAGTCCCTGGGCCTGACGGACAAC GACCTAAAGCTTCTCACCGTCCGCTGCCTCC AGGGCGGCGCGCAGTGTTTCCGGCTGCACC GCCGAATGACCACCGTGGTGAAGGCCACGG GCGATGAAGCCAACGCCCTCCCCGCCCATT CGGGCTTGCCGCCAACACAGCTTGGTGGCC GAACCCTGGGTCCCTCTGCCCACAATCAGA GTGCAAGAACTACTTAGTGTGACGGAGTCC		ID NO: 1336)	

					TCAAGCCCCCACAAGTGCCTGCCAGGTGGC AGGAAAGGGCAACTACTGGTGAGCGACCCA AGCAAGGCGGAGCCAAGACCAAGCTGGAG CCAAGAGCAACTCCAGGAGGCAGGGGTGG ATATCAAGAGCAACCCCAAGGGACACAGAC CACGGGCAACTACTGGTGAGCGCCCAAGAC AGGGGTGGATATTAAGAACAGCCCCACAAA GTGTTACCTATATTAACAATAAAGTTGAAG CCTCAACCACGCATTGCGGGTTAGATGGCG TGGCTTGGCCCGCCGCCATGATGAGCTGGA ACCCTCCACCTGGTGGGCCGCACGAGACCA CCGGCTCTTTCTACTAAGGCCGGTCTCCGTG ACTGCGGTTGGGATAAACTCCAAGCACTGA GCGGTAAAAAAAAAAAAAAAAAAAAAAA AAA (SEQ ID NO: 1090)			
R2	R2-1_MLe	.		Mnemiopsis leidyi	TGGGGGCCCCTTGGACTTGCTCCCTGGGGC AGGACACCAGTGAAAGGAGATCCTCAAGAC AGGACAGAGAGACAGGCACACCAACCCTTC GAACCTGAGGAGACACCCAGATCTGGTTAC CCCATTCCTGTAACCATGGTTGCTCTCCGTG CGCTGTTAAGGAAACCCAGCCTAGTACCCT CGGGGAAAGGTTGCGTAGTACTTAGCAGTG TGCGGATCAAACCTCTACCGGTCTCTCTAGC	TGGGGG CCCCTT GGACTT GCTCCC TGGGGC AGGACA CCAGTG AAAGGA	TAGT GACG AGAA AAGT CGCT TTATC CTTA CATA	MSNTSHSKLNL KMDNKLKTSLE TPSGVRADSIITR VRTSSNRGEHSN GVTYPRCEQGV APLDTHGGICDA PPQVTVPATETD KQKKCEYCEFT

					GATGAAAGTTTCTCCGACTGGAACTTGAGG	GATCCT	ACAG	YLKPRQIGTHM
					GACTGGCTAGCCCAGCTAGCCTGTAGAGCA	CAAGAC	TGTG	RKRHPQEWNDI
					ATGCGTATACGATGCCTTGGCGACAATGGC	AGGACA	ATAG	KRTKFLSEKRQ
					GACCGCTGCTTAGCAGACGGAGGTTAGTGA	GAGAGA	TCAT	KRWLDEDFELL
					AAGGGCGACTTGCTGTTCATAGTCACGTGA	CAGGCA	CCTT	CIGQEEYLVLSSI
					GTCGTCTAGAAACTGCACCGATGCTGCCCT	CACCAA	ATCA	GKQGKGINQYI
					CTGTTCCAGGAGAAGGACAGTGGAGGACTA	CCCTTC	CAAC	QTKYFPTLSTDA
					AATCGTAGCGCGAGCGGTGTTATGTCGAAC	GAACCT	TGTC	IKSQRKSRRFSE
					ACTAGCCACTCGAAACTTAACCTCAAAATG	GAGGAG	CTGG	YSEKRSRELQPC
					GATAATAAACTCAAGACTAGTCTCGAGACT	ACACCC	CGAA	NTSSDPEELPNE
					CCCTCCGGAGTCCGAGCAGATTCGATAATA	AGATCT	CAAA	AVTENSPLSFDP
					ACGCGGGTGAGGACCAGCAGTAACCGTGGT	GGTTAC	ATGG	LDRDVVKKISSK
					GAACACAGCAATGGTGTCACGTATCCCCGG	CCCATT	CAGA	DHGDQILLVQE
					TGCGAACAGGGGGTTGCGCCTCTGGATACG	CCTGTA	GGGA	HLINGRYQEAN
					CATGGGGGGATATGTGACGCTCCTCCACAG	ACCATG	TAGT	TLAKAIFEKLSG
					GTGACGGTTCCCGCCACTGAGACGGATAAA	GTTGCT	ACTC	KFPNLKTGDHR
					CAGAAAAAATGTGAATACTGCGAGTTCACA	CTCCGT	GGTC	PGKQQTARKVG
					TACCTGAAACCGAGACAGATAGGAACCCAC	GCGCTG	CAAC	KKRVRGSGKKL
					ATGAGAAAACGGCATCCCCAGGAATGGAAC	TTAAGG	CAGA	SPSKQNRRELYA
					GACATCAAACGGACCAAATTCCTGAGCGAA	AAACCC	AGGA	IVQKQWRTKKR
					AAGAGACAAAAACGTTGGCTTGACGAAGAC	AGCCTA	AGCC	SKVINQILTGNL
					TTCGAGCTTCTCTGCATAGGCCAAGAGGAA	GTACCC	CACA	NKEQSYTHTPD
					TACCTTGTGCTATCCTCGATTGGCAAACAGG	TCGGGG	CGAT	QLAQFWSTLFG

					GCAAAGGGATCAACCAGTATATTCAAACCA	AAAGGT	GCCA	RVSPRDDRPINH
					AGTACTTCCCAACCCTTAGTACAGATGCTAT	TGCGTA	AACT	RRSVIPELDKPL
					TAAATCTCAAAGGAAATCCAGACGGTTTTC	GTACTT	TGCT	SVEEVEAALKG
					AGAGTACAGCGAAAAGAGATCACGTGAACT	AGCAGT	GTAA	AKDAATGIDGV
					CCAACCTTGCAATACCAGCAGCGACCCCGA	GTGCGG	CCCA	PISHLKHLGSAA
					AGAACTTCCGAATGAAGCGGTCACCGAAAA	ATCAAA	CGTG	LTILYNGLYVTG
					CAGTCCGCTCTCTTTTGATCCACTTGACCGA	CCTCTA	AGCG	SIPDPWKRARTI
					GATGTTGTAAAGAAAATAAGCTCCAAAGAT	CCGGTC	GAAA	LIPKSNPPASPG
					CACGGCGATCAGATCCTCTTAGTACAGGAG	TCTCTA	ACAT	DYRPISISSYFYR
					CACCTCATCAACGGACGGTATCAAGAAGCG	GCGATG	CCTC	IYTSAISKRLASA
					AACACGTTGGCCAAAGCTATATTCGAGAAG	AAAGTT	CGAG	VSLDDRQKGFIK
					CTCTCCGGTAAATTCCCGAACCTGAAGACC	TCTCCG	TATA	EDGIRDNLSLID
					GGCGACCATCGGCCTGGTAAACAACAAACA	ACTGGA	GTAT	TLINETKAGSKS
					GCAAGGAAAGTCGGTAAGAAAAGAGTTAG	ACTTGA	GATG	LFMTFMDVKKA
					AGGAAGCGGTAAGAAGCTATCCCCCTCTAA	GGGACT	GAAG	FDSVSHYAIARS
					ACAGAATAGACGCGAGCTATACGCGATAGT	GGCTAG	GATA	LEWAGVPDGM
					ACAAAAGCAGTGGCGCACTAAGAAAAGAT	CCCAGC	CAGG	RSVIADLYQDCT
					CTAAGGTGATCAACCAGATACTGACAGGTA	TAGCCT	ATGT	TDICGRSVKVTR
					ATCTAAACAAAGAACAATCTTATACACATA	GTAGAG	GGAC	GVKQGDPLSSTL
					CACCCGATCAGCTAGCGCAATTCTGGAGTA	CAATGC	CGCT	FNLVIEMVMSN
					CGCTGTTTGGCAGAGTCAGTCCGAGAGACG	GTATAC	CTAG	VPERLGIQFQGH
					ATCGACCGATTAACCACCGGCGGTCAGTTA	GATGCC	GCGG	RLFYLAFADDL
					TCCCGGAGCTAGACAAACCCCTCTCTGTTG	TTGGCG	CGGA	VLLTRGPTANQ

					AGGAGGTCGAAGCTGCGCTCAAAGGGGCCA	ACAATG	CGGT	KLVSLVHEQLA
					AAGATGCTGCAACAGGCATTGACGGAGTCC	GCGACC	GTAG	RVGLELHPGKC
					CAATCTCCCACCTGAAGCACCTAGGCAGTG	GCTGCT	ACGG	KSIAIMADPKRK
					CAGCTCTAACAATCTTGTACAACGGATTGT	TAGCAG	CGAC	TTFVDQGSSVLI
					ATGTGACCGGGTCGATTCCTGACCCCTGGA	ACGGAG	CTTA	GGEPVSSLGPQE
					AAAGAGCAAGAACCATACTGATCCCAAAGT	GTTAGT	CCAT	WYKYLGIKLGS
					CAAATCCACCAGCTTCACCTGGGGATTATA	GAAAGG	AGCT	GGMPQGIYRDQ
					GACCTATCTCGATCAGCTCGTATTTCTATAG	GCGACT	GCGA	LADLLAKTDSA
					AATCTACACGAGCGCGATCAGCAAAAGACT	TGCTGT	CTGC	PLKPQQRLYILR
					TGCTTCTGCCGTCAGCCTTGACGACCGTCAG	TCATAG	TGTG	SHILPKFNHRLM
					AAAGGATTCATCAAAGAAGATGGAATACGA	TCACGT	CGAT	FERVTCQTLEGL
					GACAACTTATCATTGATTGATACTCTCATCA	GAGTCG	CCGG	DKLIRTHVRKW
					ATGAGACGAAAGCAGGCTCGAAATCCCTAT	TCTAGA	AACG	LKLPKDTPGPAF
					TCATGACCTTCATGGACGTCAAGAAAGCTT	AACTGC	GGCC	YADKGSGGLGL
					TCGACTCCGTGTCGCATTACGCAATTGCCAG	ACCGAT	TTTCT	ITLRYRVPLLKL
					ATCTCTAGAATGGGCTGGCGTCCCCGATGG	GCTGCC	CACT	RRHKKMADSPD
					AATGAGAAGTGTCATAGCCGATCTATACCA	CTCTGT	TACC	PVIRLIPNAEPTI
					GGACTGCACCACTGATATCTGCGGCAGATC	TCCAGG	TCAA	SLLARWTKMCS
					CGTCAAAGTGACCAGAGGCGTGAAGCAGG	AGAAGG	GCAC	LYGKQYQHKSE
					GAGACCCCCTAAGCTCCACTCTGTTCAACCT	ACAGTG	CGTG	LSKIIRDKYWTM
					GGTAATAGAAATGGTGATGTCAAACGTCCC	GAGGAC	TCGG	CDGKGLRTEVP
					CGAACGTTTGGGAATTCAATTCCAGGGACA	TAAATC	ATCG	PDTAKKTLSLLF
					CAGACTGTTCTATCTGGCTTTTGCAGACGAC	GTAGCG	CGGG	EDRTPLKPGQLI

					CTAGTCCTCCTGACGAGAGGACCCACAGCT AACCAGAAGCTAGTATCCTTAGTCCATGAA CAACTAGCAAGAGTTGGACTTGAGCTACAT CCGGGAAAGTGCAAGTCAATCGCTATCATG GCTGACCCCAAGAGGAAAACCACTTTCGTA GATCAGGGATCCAGTGTTCTCATTGGAGGC GAACCCGTCTCTTCGCTCGGTCCGCAAGAA TGGTACAAATACCTTGGGATTAAACTGGGA TCAGGAGGCATGCCCCAGGGTATTTACCGA GACCAACTAGCAGACCTTCTAGCGAAGACA GACAGTGCCCCCCTGAAACCCCAACAAAGA TTGTATATACTCAGATCGCATATACTACCGA AGTTCAACCATAGACTAATGTTCGAGAGAG TAACGTGCCAAACACTAGAGGGTCTTGACA AACTGATCAGGACCCACGTCAGAAAGTGGT TGAAATTACCCAAAGACACGCCTGGTCCAG CCTTCTACGCTGATAAAGGTTCAGGCGGAC TCGGACTGATAACTCTCAGATACCGAGTTC CCCTCCTAAAACTTAGGAGACACAAAAAGA TGGCGGACTCACCTGATCCCGTTATCAGGTT GATACCAAACGCGGAACCGACCATCTCCTT GTTAGCTAGATGGACCAAGATGTGCTCCCT CTATGGCAAACAGTACCAGCACAAGTCCGA	CGAGCG GTGTT (SEQ ID NO: 1214)	GTGG GCGC GGGA TTCCT GCTG GGAA GTTG TCTG CACC CAAG TCGT CAGC TAAC TGTC GGTA CTTAT ACGC ACCC AGAG CCAT GCTC TTAC CAGG	GAIGVRLNTLGT PARNNRAKGYS PEANICDKCPGN RQATLGHISQTC PATHGRRVKRH DKIVNRIAKALK ERGSVKNILTEP HLRHDKLPLRK PDLIVHTEKSVEI IDVQVVADQGIS RHEDEDQQKKI VKYDVDGYKR AAYKMLGIDYG SIPCNVSAFTIT WRGNLAPHSLK LASRLQFSPVLK YIVADSLVDTW GAFLIWGKTS (SEQ ID NO: 1459)

					ACTCTCTAAGATAATTAGAGATAAATACTG		CCAC	
					GACAATGTGCGACGGGAAAGGCTTGAGGAC		TTGA	
					AGAGGTGCCACCAGATACGGCTAAGAAAAC		CGGC	
					ACTAAGCCTGCTATTCGAGGACCGGACTCC		GCTA	
					ACTCAAACCAGGACAACTAATTGGAGCTAT		CACC	
					CGGGGTCCGCCTAAATACCCTTGGTACCCC		GGTC	
					CGCTCGTAACAATCGAGCTAAAGGCTACAG		TGTA	
					TCCTGAGGCTAATATTTGTGACAAATGTCCA		GGAG	
					GGCAACAGGCAGGCCACTCTTGGCCACATC		GGTA	
					TCGCAAACTTGTCCGGCGACTCACGGGAGG		TCTC	
					AGAGTCAAACGACACGACAAGATTGTGAAC		AGCG	
					AGAATCGCAAAAGCTCTAAAAGAACGAGG		ACTT	
					CTCCGTCAAGAACATACTTACCGAACCCCA		GGAC	
					CCTCAGGCACGATAAACTGCCCCTACGGAA		AAAA	
					ACCCGACCTAATAGTACATACGGAAAAGTC		TGGA	
					GGTCGAAATCATTGATGTCCAGGTCGTTGC		TCTTG	
					AGACCAGGGCATCTCGAGGCACGAAGATGA		ACTG	
					AGACCAACAAAAGAAGATCGTCAAATATGA		CCTG	
					CGTGGATGGATACAAGAGAGCGGCCTACAA		AGGG	
					AATGCTAGGGATTGACTACGGGAGCATCCC		CGCT	
					CTGCAATGTAAGTGCGTTCACTATCACCTGG		TATC	
					CGAGGAAACCTTGCCCCCCATTCACTCAAA		GAGT	
					CTGGCGAGTCGTCTCCAGTTTTCACCCGTGC		GACC	

					TCAAATATATCGTGGCTGACTCCCTAGTTGA		CAGA	
					CACATGGGGTGCGTTCCTGATATGGGGAAA		GTAA	
					GACGTCTTAGTGACGAGAAAAGTCGCTTTA		GCTG	
					TCCTTACATAACAGTGTGATAGTCATCCTTA		GTAG	
					TCACAACTGTCCTGGCGAACAAAATGGCAG		GAAG	
					AGGGATAGTACTCGGTCCAACCAGAAGGAA		AATC	
					GCCCACACGATGCCAAACTTGCTGTAACCC		TTGC	
					ACGTGAGCGGAAAACATCCTCCGAGTATAG		AGTT	
					TATGATGGAAGGATACAGGATGTGGACCGC		GAGA	
					TCTAGGCGGCGGACGGTGTAGACGGCGACC		GGGC	
					TTACCATAGCTGCGACTGCTGTGCGATCCG		TAGT	
					GAACGGGCCTTTCTCACTTACCTCAAGCACC		AGGG	
					GTGTCGGATCGCGGGGTGGGCGCGGGATTC		CCAG	
					CTGCTGGGAAGTTGTCTGCACCCAAGTCGT		CCTCT	
					CAGCTAACTGTCGGTACTTATACGCACCCA		GCTC	
					GAGCCATGCTCTTACCAGGCCACTTGACGG		GTCA	
					CGCTACACCGGTCTGTAGGAGGGTATCTCA		ACTG	
					GCGACTTGGACAAAATGGATCTTGACTGCC		TCCT	
					TGAGGGCGCTTATCGAGTGACCCAGAGTAA		AGGC	
					GCTGGTAGGAAGAATCTTGCAGTTGAGAGG		GTTA	
					GCTAGTAGGGCCAGCCTCTGCTCGTCAACT		ACGT	
					GTCCTAGGCGTTAACGTGTGCCTCCCTATGA		GTGC	
					GGTAACCGTGAAGTATTACTTGTTTCCCTGT		CTCC	

					AGAGACCATATAACTAGCTGAGAAACGCCC TGCGGGTTAATATACAGTTACCGCCGTCCTG CATCCCCGAGTTGTGACCAGCGTACAGCCA TGTACCACGGCACTTCCAGCATTCTCTGGTC TAAGAATGTATCAGCTGGCCCGCCGAAAGA TAGAGCGACCCCGCCTCTATCACTAGAGTA AGCAGGCACGCGAAATATATGCAGGACGCC TTGGCTCAGTAGCTCTGGTACTGAGTATAGC ATGGCTGAACGCCCCTTTAAGTTCGGTGGC CCAGAGACTCTCCGAGACTTCCCAGTCCTG GAGAGGGTACACCTTTACCAGAACTTTCTG TGGTCGGTTG (SEQ ID NO: 1091)		CTAT GAGG TAAC CGTG AAGT ATTA CTTGT TTCCC TGTA GAGA CCAT ATAA CTAG CTGA GAAA CGCC CTGC GGGT TAAT ATAC AGTT ACCG CCGT	

							CCTG CATC CCCG AGTT GTGA CCAG CGTA CAGC CATG TACC ACGG CACT TCCA GCAT TCTCT GGTC TAAG AATG TATC AGCT GGCC CGCC GAAA	

							GATA GAGC GACC CCGC CTCT ATCA CTAG AGTA AGCA GGCA CGCG AAAT ATAT GCAG GACG CCTT GGCT CAGT AGCT CTGG TACT GAGT ATAG	

							CATG GCTG AACG CCCC TTTA AGTT CGGT GGCC CAGA GACT CTCC GAGA CTTCC CAGT CCTG GAGA GGGT ACAC CTTTA CCAG AACT TTCTG TGGT	

							CGGTTG (SEQ ID NO: 1337)	
R2	R2-1_PM	.		Petromyzon marinus	CTATTAATGGGATGAAGAAGGGGGACACGA GTTTGTGTGTGCATCCAGTTTCCATGGTGCA TGCAGGAGTGGTGGTTTAAATGGCGAGACT CTACAGGGCTTCCATGGCTACACGGGATGC AAGGCATCAGACATTTTGGCACAGGCAATC CTTTTGGTCTCTACCGCAATCATGTCTTAGA CCTCAGTAGCGACCACTACAACCACAGTGG TGACTGCTGTTGAGTGAAGGACGACTGAGC GCTGGATAACAACTTTCTTGCGTGGCCCAA CATCGAAGCAACCACTTCGGAGCTGGCACA AGGCAAGAGGGCAGCCCAAGGTGTGAATC ATCTCAACTTCACTGCAGGAAGAAATGCTG TGCAAGGATGAGTGTGAACGACACCAACGG GATTGTTGCTGACCAGGAGGTGCCAACCAA ATTTGAATGGATTGACTTTGGGCCTGGTTTC TCCTGCGTGTATTGCACGGAAAAACAAGTG GCTACACGTGTGGCCGTCGTGTCCTGGGGTT TCGCAACACAACTCCACAAGATCGACAACT	CTATTA ATGGGA TGAAGA AGGGGG ACACGA GTTTGT GTGTGC ATCCAG TTTCCA TGGTGC ATGCAG GAGTGG TGGTTT AAATGG CGAGAC TCTACA GGGCTT CCATGG	TAAT TTAA GGTA AAAT CGTG GGAT TGTTT TGAT GGCA ATCT GCCT AGTC GCGG CCTTC CATTT TGGG TAGG CAGC	MNERLTDELTT EFILSDMFLWDY PCTDQNKCYPC NLVFLDHRTWS SHMARVHPHAN KTYKCRICNRTA DSIHKIASHYGR TCKSLIGKTNAI TTTIDETLFSCLH CSRGFTTKTGLG VHTRRTHPTEHE AILQQNTPGRKV RWGEEEVEIMA HKEAQQKDEDI NMNQLIQNSVM PHRTLEAIKGKR RNIKYKELVRTL KETTYKVENQC

					ATGAGGATGACAATGTACTTAAAGAACAAA	CTACAC	AGAC	LVNLVLPTTSEI
					GAGACTGACGCCAAAGGGGATTTAAACCGC	GGGATG	CCAT	TTTPSEGDQPAI
					CAAATCGTACATTGGGTCTCACTACAATTTT	CAAGGC	CTAT	RAEKEQSPTAAE
					TTTACGTGTATTTATTTTCCTAAGTGTCTGT	ATCAGA	ATAA	DLQVIINDLKSQ
					ACTTGCCATTCTTCGCTGCTTTTTCTGCATTA	CATTTT	CAAA	NFSHNQALLLL
					ATTGCATATCGTATGCAAATAAGCGAATTA	GGCACA	CTAC	NSHVEKFLNRS
					ACCACCACCGTGCAACTATATGCAGATGTT	GGCAAT	TTTGC	KPIKRKDHVNQ
					ACAGCTGAGCCCTCTATCATACCGGTGTACT	CCTTTT	CTTTC	QEIDENRHRRQS
					AATCTGGTATGGTGTTGGCATGCTATGCTTG	GGTCTC	ATAG	KQTKYRRYQYL
					CGTAACGACCTTTGCTGATTGGTTCAGTCGG	TACCGC	GGGT	YHTNKKALLDEI
					CTGATGGTGGGTTCAGGCGAAACATTTGTA	AATCAT	ACCC	TSDRSGPSIYPTE
					TATTGGTTTAATCAAACCGAAACACTAAAA	GTCTTA	GACC	ESIRGTFVTLFES
					TTTTGAACACAGTTTTCCATTACACCAGTTG	GACCTC	CTAC	NSPPDNIPSKLK
					TATTGCTAGAAGTGCAAATCGAAGGAGTCA	AGTAGC	CAAC	NDQSCIDIVKAI
					ATTTTGACCGACGATTAGCTGCCGATGTGC	GACCAC	TTTCG	TLDELIKTLAIM
					GGTGAAAAAGCTGATCACAATAGCATACAC	TACAAC	GGGA	KDKSPGQDNITL
					TTGGGCCGACAACCCCGTGTGCTATAAACG	CACAGT	AGTA	SDLRTLPIKYLL
					TAAGTCGCGAATTATAAAGAAAACAAACCG	GGTGAC	AAAG	DILNIILYIQDIPQ
					GACGGACTACTCGGTGACGAACTAACATCG	TGCTGT	AAA	IWKQHRTRLIPK
					CTCATGAATGAGCGATTAACAGACGAGCTG	TGAGTG	(SEQ	TKEELEKPSNW
					ACTACGGAATTTATCCTTTCGGACATGTTTT	AAGGAC	ID NO:	RPITISSIVIRLLH
					TATGGGACTACCCATGCACAGATCAGAACA	GACTGA	1338)	KILSYRLGQQLK
					AATGTTATCCATGCAATCTTGTTTTCCTAGA	GCGCTG		LNYRQKAFLPV

					CCACAGAACTTGGTCATCACATATGGCACG	GATAAC		DGCFENSALLHF
					GGTACATCCACATGCAAACAAAACGTATAA	AACTTT		IIHNARQKHENT
					ATGTCGAATTTGTAATCGCACAGCAGATAG	CTTGCG		QIVSIDLSKAFDS
					CATACACAAGATAGCGTCACACTACGGAAG	TGGCCC		VSHESIIRALNRF
					AACTTGCAAAAGTTTAATAGGTAAAACTAA	AACATC		NLSKESITYLTNI
					TGCTATAACCACCACAATTGATGAAACACT	GAAGCA		YKCNLTDIVFGS
					ATTTAGTTGTTTACATTGCAGCAGAGGTTTT	ACCACT		TIMRNINLKRGV
					ACTACGAAAACAGGTTTAGGGGTACATACT	TCGGAG		KQGDPLSPLLFN
					AGACGAACTCATCCGACAGAACATGAGGCT	CTGGCA		MIMDELLDNLP
					ATACTACAGCAAAACACACCAGGAAGGAA	CAAGGC		TYIGVNVGNQK
					AGTTAGATGGGGAGAAGAAGAGGTAGAAA	AAGAGG		VNSMMFADDLI
					TTATGGCCCATAAAGAAGCCCAACAGAAGG	GCAGCC		LFAETECGMNK
					ATGAGGACATAAACATGAATCAACTAATAC	CAAGGT		LLDITTKFLDDR
					AGAACTCAGTTATGCCACACAGAACGCTAG	GTGAAT		HLKININKCNSL
					AAGCGATTAAAGGGAAGCGGAGAAATATC	CATCTC		RFIKYGKQKTFS
					AAGTATAAGGAATTGGTAAGGACTTTGAAA	AACTTC		VATTSSYFINNE
					GAAACTACCTATAAGGTAGAAAATCAATGC	ACTGCA		PINPVSYVKGFK
					CTTGTTAACTTAGTTTTACCGACAACATCGG	GGAAGA		YLGIEFDPRGKR
					AAATAACAACTACACCTTCGGAAGGAGATC	AATGCT		SISCNLLAAMLN
					AGCCAGCAATAAGGGCCGAAAAAGAACAA	GTGCAA		KLTRAPLKPEKK
					TCACCGACAGCAGCTGAGGATCTTCAGGTC	GGATGA		VYLINNNLIPRII
					ATAATTAACGATCTAAAGAGCCAGAATTTT	GTGTGA		HQLVLGKVTKG
					AGCCACAATCAGGCGTTACTGCTACTCAAT	ACGACA		LLMSLDSEIRKT

					TCTCATGTAGAAAAGTTTTTAAATCGAAGT	CCAACG		VKLLLRLPHDTP
					AAACCAATTAAAAGGAAAGATCACGTAAAC	GGATTG		DSFFYTSVSNGG
					CAACAGGAGATAGATGAGAATAGGCATCG	TTGCTG		MGIRNLCDSVA
					AAGACAATCAAAGCAAACTAAATACAGGA	ACCAGG		LSIINRHNKLITS
					GATATCAATATTTATACCATACGAACAAGA	AGGTGC		DDLVIRALSQQS
					AAGCTCTATTAGACGAGATTACTTCAGATA	CAACCA		YTIATLKQAHII
					GATCGGGGCCAAGTATATACCCAACTGAGG	AATTTG		AGSKFPSKSLNQ
					AAAGCATACGGGGAACATTCGTTACTTTAT	AATGGA		NKWSNKLYQTT
					TCGAGTCAAACTCTCCTCCAGATAATATACC	TTGACT		DGRGLVYCQSQ
					CTCTAAATTAAAAAACGACCAATCCTGCAT	TTGGGC		TENNSWITGNH
					CGATATCGTAAAAGCAATCACCTTAGATGA	CTGGTT		RTIKSYNYIDMV
					GTTGATTAAGACCCTAGCAATTATGAAGGA	TCTCCT		KLRINALPTKSR
					TAAGTCACCTGGACAGGACAACATTACTCT	GCGTGT		CNRGTLETKQC
					TAGCGATCTTAGGACTTTACCAATAAAATA	ATTGCA		RFKCRSINNQISE
					TTTACTAGATATCTTAAATATCATCCTTTAC	CGGAAA		ETLAHILQKCDR
					ATACAGGATATACCACAAATATGGAAACAG	AACAAG		SHYSRIARHDSL
					CACAGAACAAGACTTATCCCGAAAACTAAA	TGGCTA		VQFLATAAQKL
					GAGGAATTAGAAAAACCCTCAAATTGGAGA	CACGTG		NWEVIKEPTLPS
					CCCATAACCATCTCATCAATTGTAATTAGGC	TGGCCG		DTNKAKPDLILV
					TATTACATAAAATTTTAAGTTATCGTCTAGG	TCGTGT		RDSHVLIVDVA
					ACAGCAATTAAAGCTTAATTACAGGCAGAA	CCTGGG		VPWESRSLAHA
					AGCATTCCTCCCGGTAGACGGATGTTTCGA	GTTTCG		YDFKVKKYATD
					AAATAGTGCATTACTACACTTCATCATACAC	CAACAC		KKMQAYLKTIY

					AACGCCAGGCAAAAGCACGAAAACACGCA AATAGTGTCAATAGACCTCAGTAAGGCATT CGATTCTGTCAGCCACGAATCGATTATTAG AGCCTTAAACCGATTTAACTTATCAAAGGA ATCCATAACGTACTTGACCAACATCTATAA GTGTAATCTAACTGATATTGTATTTGGATCG ACAATAATGCGTAACATAAATCTAAAAAGA GGCGTAAAGCAAGGAGATCCACTTTCACCG TTACTATTTAACATGATTATGGATGAATTAT TAGATAACTTGCCGACATATATAGGAGTTA ATGTAGGAAATCAGAAAGTAAATTCTATGA TGTTTGCAGACGACCTTATCCTATTTGCAGA AACGGAATGTGGCATGAATAAACTCTTAGA TATAACTACTAAATTCCTCGATGACAGACA CTTGAAAATAAATATAAACAAATGCAATTC GTTAAGATTTATCAAGTACGGCAAACAGAA GACATTTAGTGTTGCAACGACATCATCGTA CTTTATAAATAACGAACCCATTAATCCGGT ATCATATGTAAAGGGATTCAAATATCTAGG CATTGAATTTGACCCAAGGGGAAAACGATC TATAAGCTGTAACCTGCTCGCAGCAATGTT AAACAAACTGACCAGAGCACCGTTAAAGCC AGAAAAGAAAGTATATTTAATCAATAACAA	AACTCC ACAAGA TCGACA ACTATG AGGATG ACAATG TACTTA AAGAAC AAAGAG ACTGAC GCCAAA GGGGAT TTAAAC CGCCAA ATCGTA CATTGG GTCTCA CTACAA TTTTTTT ACGTGT ATTTAT TTTCCT AAGTGT		PEKEIRTEALIIS ARGGWCALNN MVTKKVGLSSA WVKLALIKVME GSVKIWRSWSK G (SEQ ID NO: 1460)

					TTTAATACCTCGTATTATTCATCAATTGGTC	CTGTAC		
					CTCGGAAAAGTTACCAAGGGTTTATTGATG	TTGCCA		
					TCACTTGATTCTGAAATTAGGAAAACAGTA	TTCTTC		
					AAGCTTCTGCTCAGGTTGCCACACGATACG	GCTGCT		
					CCCGACAGTTTCTTTTATACATCAGTATCCA	TTTTCT		
					ACGGAGGAATGGGTATAAGAAATTTATGCG	GCATTA		
					ACTCAGTTGCACTATCTATAATAAACAGAC	ATTGCA		
					ACAACAAATTGATAACTTCAGATGATCTAG	TATCGT		
					TAATAAGAGCATTATCACAACAATCATACA	ATGCAA		
					CTATTGCAACGTTAAAACAGGCCCATATCA	ATAAGC		
					TTGCAGGCTCCAAATTTCCTTCGAAATCTTT	GAATTA		
					AAATCAGAACAAATGGTCAAATAAACTATA	ACCACC		
					TCAAACAACAGATGGTCGGGGGTTGGTATA	ACCGTG		
					CTGCCAATCTCAAACAGAAAACAATTCATG	CAACTA		
					GATAACAGGGAATCATAGAACAATAAAATC	TATGCA		
					GTATAATTACATAGACATGGTTAAACTAAG	GATGTT		
					GATTAATGCACTACCGACTAAATCGAGATG	ACAGCT		
					CAATCGAGGGACGTTAGAGACCAAGCAATG	GAGCCC		
					TAGATTTAAATGTCGAAGTATTAACAACCA	TCTATC		
					AATTTCAGAGGAAACATTGGCACATATCTT	ATACCG		
					GCAAAAGTGTGATCGCAGTCATTATTCAAG	GTGTAC		
					AATCGCAAGGCATGATTCTTTGGTGCAATTT	TAATCT		
					CTGGCAACGGCCGCACAAAAACTAAACTGG	GGTATG		

					GAAGTGATCAAAGAACCCACTTTACCGAGC GATACAAATAAGGCAAAACCGGACTTAATT TTAGTAAGAGACTCTCATGTCTTGATAGTAG ATGTGGCAGTTCCGTGGGAGTCTCGATCATT GGCACATGCATACGATTTTAAGGTGAAAAA ATACGCTACTGACAAAAAAATGCAAGCATA TTTAAAAACTATATATCCGGAAAAAGAAAT TAGAACGGAGGCTTTAATCATATCTGCACG TGGGGGCTGGTGCGCTTTAAATAATATGGT AACAAAAAAGGTGGGATTGTCAAGTGCATG GGTAAAATTAGCATTGATCAAGGTCATGGA GGGTTCCGTAAAGATATGGCGCTCTTGGAG CAAAGGATAATTTAAGGTAAAATCGTGGGA TTGTTTTGATGGCAATCTGCCTAGTCGCGGC CTTCCATTTTGGGTAGGCAGCAGACCCATCT ATATAACAAACTACTTTGCCTTTCATAGGGG TACCCGACCCTACCAACTTTCGGGGAAGTA AAAGAAA (SEQ ID NO: 1092)	GTGTTG GCATGC TATGCT TGCGTA ACGACC TTTGCT GATTGG TTCAGT CGGCTG ATGGTG GGTTCA GGCGAA ACATTT GTATAT TGGTTT AATCAA ACCGAA ACACTA AAATTT TGAACA CAGTTT TCCATT ACACCA		

						GTTGTA TTGCTA GAAGTG CAAATC GAAGGA GTCAAT TTTGAC CGACGA TTAGCT GCCGAT GTGCGG TGAAAA AGCTGA TCACAA TAGCAT ACACTT GGGCCG ACAACC CCGTGT GCTATA AACGTA AGTCGC GAATTA		

						TAAAGA AAACAA ACCGGA CGGACT ACTCGG TGACGA ACTAAC ATCGCT C (SEQ ID NO: 1215)		
R2	R2-1_SM	.		Schmidtea mediterranea	CAGTGCTATTCGAATGTCAATGTGAAGAAA TTCAACTAAGCTCTGGTTAACGGCGGGAGT AACTATGACTCTCTTAAGGAATTAAGAATTT ACCTGCCKTAKTAAAARTGAAATCMGTTGT TCATWGCAAGTGGTATTGTACACCTTCCCG CGGTGCTAGTCGTTTAAAACTAAGTTACAA ACCACGAGGGGCGTCCTGACGGACTGSAAA AGCATTGAGRGTCMTGAAGAGAGGCTCTTA TTGTACGAATCTCTTCAACGATCGAAGTCTG GACCGATATGAGAACTAATACATTAGTTGA CAGGTGAAAAATACTGTTGATTACTTAGTTC TCAGTCATGTGGTATATTGCCAGTCAATTAC	CAGTGC TATTCG AATGTC AATGTG AAGAAA TTCAAC TAAGCT CTGGTT AACGGC GGGAGT AACTAT GACTCT	GGCC ACGC GCGT CGTC CTTGT TTGA TCAC TAGT GGAT CAAC CTTC GACT	MKKVLNNETEK LPGSNLTFMCGF CDREFDTARGR GVHESRGHLVE RDAAVQSRVKA VVSKKYYYSNE EDVALAKMQLX HADLAKSEXLE AMYLALGKGRT REAIEQHIRKSL RYKGVLEEQRK LLETARGNVRQ

					TACATWAATATTAGTGTGGCTCTCAAAGGA	CTTAAG	CCCC	NNVGVPASNAT
					ACACGATTGRTCGGCAGTCCAATGCGCGAC	GAATTA	GGAA	KNLQRFLESLPL
					TGGCGGGCTTGTTGTTTGCATTTGTTACCGG	AGAATT	CTGT	GTNRREERLDRI
					CTACTTGAAAAGGTTATATATAGCAGACGC	TACCTG	GGGA	IRSNSIESQRLEL
					TTAAAGCGCGACTGTAATTTACATYTCATTG	CCKTAK	GTGG	IHYCNDMCQDF
					CCCAGTATTTGTCTTTTGTCAGATTTAGCAA	TAAAAR	CGGA	VQLDCQXNPIN
					AATTTCATATTTTGTTAATTACCTTAACTGG	TGAAAT	AGAA	AIRRRNPKRLSK
					TTAAACGATCCCATAATTGCTTGCAATTATT	CMGTTG	AGGC	KQLKRAKFSAL
					ATAAAGTAATTCAGGTAAAAATTACATATC	TTCATW	CAGA	QRLWIRDRKAA
					TGGCTGATCCTGCCAGTAGTCATTTTACTTC	GCAAGT	GGAT	AQLVLKDKLDS
					CGCCGCGCTATAAAACAGTTTAAAAACTGA	GGTATT	GTCC	LLSNKEDSKDL
					ATAGGAATCAAAAAGAACATGGCAAGCGA	GTACAC	TGAA	GSYWQQVFERE
					CTATATGTAACTGGGCATTCAACATTCCCTA	CTTCCC	ACCA	SELDRRPIPQVV
					TTACATATGGTGGTGCCTGGGGTCTGTTTTA	GCGGTG	TATA	ENEELNSPVLEK
					TATAATGGGTACCCGGGAAGTGGATCTGTA	CTAGTC	TTTAT	EVEWAVKNIKK
					TCACCAGTCATGGTGCCATATCTTTKGATAA	GTTTAA	TTAT	STAAGPDGLTA
					AGATACAGTTTAAAACTGCGATGATACTAA	AACTAA	AGAA	LALKKIPYSELV
					TAGAGATCCTCTTAGACCTTCGTAAAGAAG	GTTACA	GTTTT	KLFNIILLVGFLP
					TGGGGATTGATGACATTAGCATTGGAAGAA	AACCAC	ACTT	DVLKNSRTILIPE
					TTAAATCTCCAAGGAAATGGAGTAACTTCA	GAGGGG	CATC	VDNPQGGGDYR
					ATGAAGTCCCACAACCCCGTTGAAGGGCTG	CGTCCT	CTATT	PISINSVLTRTLN
					GGTTCGAGTATCGAGAGAAAACTCTAAATT	GACGGA	TACG	KILAKRVSEGDF
					CTCTTCGGTTMTGTCCAACGGAGGGGACAT	CTGSAA	TATTT	GINGQKGFKSV

					TACTGTAAAATATCCTCTAAAAACAACTAT	AAGCAT	CAGT	DGCLENLATVE
					GAAAAAAGTCTTAAATAATGAAACCGAAAA	TGAGRG	ATGA	SILADARMKKN
					ATTACCGGGAAGTAATCTAACTTTCATGTGC	TCMTGA	AAAT	KLAVVFLDMSK
					GGGTTCTGTGATCGGGAATTCGACACGGCW	AGAGAG	GAGT	AFDSVNHESIVR
					AGGGGCAGAGGAGTGCACGAGAGTAGAGG	GCTCTT	AAAG	AGEIKGYPKLL
					TCATTTAGTTGAGCGGGATGCGGCGGTTCA	ATTGTA	TTCTC	MTYVKECYNDA
					GAGCAGAGTGAAAGCCGTGGTGAGTAAAA	CGAATC	GACT	TTNVAGVTAKF
					AGTATTATTATAGTAACGAAGAGGATGTGG	TCTTCA	CGAT	NRGVKQGDPLS
					CATTAGCGAAAATGCAGTTASAGCATGCAG	ACGATC	GAGT	PALFNNVIDLAI
					ATCTGGCCAAAAGCGAACWATTAGAAGCC	GAAGTC	TGGG	ERVSGTGIGYN
					ATGTATCTTGCATTGGGAAAGGGAAGAACT	TGGACC	GGCA	MGGKKYSVVA
					CGTGAAGCCATAGAGCAACACATAAGGAA	GATATG	ACCA	YADDLVLFGES
					ATCGTTACGTTATAAGGGGGTCCTTGAAGA	AGAACT	TTGG	REGLQIALTALL
					ACAGCGAAAGCTCCTTGAGACAGCAAGGGG	AATACA	GGGT	EELKLNGLTPNP
					AAATGTTCGGCAAAATAACGTGGGTGTGCC	TTAGTT	CCTG	AKSASLTFERSG
					AGCTAGCAATGCCACTAAAAATCTGCAAAG	GACAGG	AAGA	PHWFASTDTVT
					ATTCTTAGAATCGTTACCCTTGGGAACGAAT	TGAAAA	GAGG	ALGDQIPAMGNI
					AGGCGCGAGGAACGATTGGATAGGATTATC	ATACTG	CTCTC	ETYKYLGIKFNS
					CGATCTAACTCGATCGAAAGCCAAAGACTT	TTGATT	ACTG	CGVVKGSLPGIY
					GAATTGATCCACTATTGTAACGATATGTGTC	ACTTAG	TAAA	TKKLELISKAPL
					AAGACTTTGTGCAACTAGACTGTCAAARGA	TTCTCA	AAAT	KPQQRLAMLTD
					ACCCCATCAATGCTATAAGGCGCAGAAATC	GTCATG	CTCTT	FLIPGVLHQAVF
					CGAAAAGACTATCGAAAAAGCAGTTGAAG	TGGTAT	CGTG	GQTNAGDLRSL

					AGAGCTAAATTTAGCGCTCTTCAACGGCTTT	ATTGCC	TCTGT	DKRTRRAVRSW
					GGATAAGGGATCGGAAAGCTGCTGCACAGT	AGTCAA	TTATT	CHLPSDTSTAFI
					TAGTGTTGAAGGATAAGCTTGATAGTTTGCT	TTACTA	CCTA	HAKAKDGGMGI
					CAGCAATAAAGAGGATTCCAAGGATTTGGG	CATWAA	GGCA	PSIRAEVQFGKL
					ATCGTATTGGCAACAGGTCTTCGAACGTGA	TATTAG	CNTG	DRFGKLPNVKD
					GTCCGAATTAGACCGCAGACCCATACCACA	TGTGGC	CTGC	ERSKVLADNAHI
					AGTGGTGGAAAACGAAGAGTTAAATTCCCC	TCTCAA	ATTA	KKKMLEKLGVG
					GGTATTAGAGAAGGAAGTAGAGTGGGCTGT	AGGAAC	TGAA	IPIKGVRCKNKL
					CAAAAACATTAAGAAGTCCACTGCCGCAGG	ACGATT	GCGG	EFYNKMREELIK
					ACCAGACGGGTTAACGGCACTTGCCTTGAA	GRTCGG	WGAA	SNDGIGLKEASL
					GAAAATACCGTATTCCGAGCTAGTCAAACT	CAGTCC	AGTA	VPSANTWLKLS
					ATTTAATATAATACTGTTGGTGGGATTCCTA	AATGCG	AAGT	DLHMSGRTFVG
					CCTGATGTATTAAAAAATAGTAGAACTATC	CGACTG	TAGA	CLKTRGNLMAT
					CTAATACCCGAAGTGGATAATCCCCAAGGG	GCGGGC	GCTG	VTRTSRGGQNP
					GGYGGTGATTATAGACCGATTTCGATCAAT	TTGTTG	AGAG	GIELNCKKGCQ
					TCAGTGCTCACTAGAACACTAAATAAGATC	TTTGCA	ATAG	YQGSLNHIVQK
					CTAGCGAAACGAGTCTCGGAAGGTGATTTT	TTTGTT	GTAC	CPVVKGLRIKRH
					GGTATCAATGGTCAAAAAGGATTCAAAAGT	ACCGGC	TTGCT	DEVVKYVEEITK
					GTAGATGGTTGTCTAGAGAATCTAGCAACA	TACTTG	GCAT	KAGWSATMEPII
					GTTGAATCAATTTTGGCCGATGCTAGAATG	AAAAGG	TATG	PFEGSHRKPDLV
					AAGAAAAATAAGCTTGCGGTAGTATTCTTG	TTATAT	AAGC	LVRGDLGKVVD
					GATATGAGTAAAGCTTTTGATTCTGTAAACC	ATAGCA	GGAG	IQIVSDHCGLDE
					ACGAGAGTATTGTTAGAGCTGGAGAAATCA	GACGCT	AAAG	KNSCKIGKYDN

					AAGGTTATCCCAAACTATTAATGACGTATG	TAAAGC	GCCT	DIIRNYVRGLGP
					TGAAGGAGTGTTATAACGACGCTACTACGA	GCGACT	CGAA	SRVEVAAITLN
					ACGTCGCAGGTGTAACAGCCAAGTTTAACC	GTAATT	TAAA	WRGVWSRDSFN
					GGGGAGTTAAACAAGGTGATCCTCTTTCGC	TACATY	TAGG	LIKRLGMTEMD
					CGGCGTTATTTAATAACGTAATAGACCTTGC	TCATTG	GTGT	AKIISMRVLAST
					AATTGAGCGAGTTTCCGGTACTGGAATTGG	CCCAGT	TAGA	AKMFKTCKKVL
					ATATAATATGGGCGGTAAGAAGTATTCTGT	ATTTGT	GTTA	EPVCRTKTADC
					GGTCGCGTATGCTGACGATCTTGTTCTATTC	CTTTTG	TTGA	DGYGPEETSARP
					GGTGAGTCGAGAGAGGGGTTGCAAATAGCC	TCAGAT	TGGA	CHELNLKESSGT
					CTGACGGCGCTACTGGAGGAATTAAAGCTA	TTAGCA	GAGT	* (SEQ ID NO:
					AACGGTCTGACACCCAATCCAGCGAAAAGT	AAATTT	ATAC	1461)
					GCTTCGTTGACATTTGAGAGATCAGGGCCA	CATATT	TAGT	
					CATTGGTTCGCAAGTACCGATACCGTTACA	TTGTTA	AAGC	
					GCACTAGGAGATCAAATACCGGCCATGGGT	ATTACC	TTAA	
					AACATCGAAACCTATAAATACTTAGGAATC	TTAACT	GCTG	
					AAGTTTAATTCGTGCGGAGTGGTTAAAGGG	GGTTAA	CGCC	
					AGTCTACCCGGGATATACACCAAGAAATTG	ACGATC	TCGC	
					GAGTTAATCTCTAAGGCTCCTTTGAAGCCGC	CCATAA	GCGG	
					AACAGCGGTTAGCTATGCTAACCGATTTCTT	TTGCTT	TGCC	
					GATTCCTGGAGTATTACACCAAGCCGTGTTT	GCAATT	CAAA	
					GGACAGACAAATGCAGGGGACCTGCGCAGT	ATTATA	AATA	
					CTCGATAAACGGACGAGAAGAGCGGTGAG	AAGTAA	TACTT	
					ATCTTGGTGTCACCTTCCCTCTGATACGTCA	TTCAGG	AATG	

					ACGGCATTCATCCATGCTAAAGCTAAGGAT	TAAAAA	AGAG	
					GGAGGAATGGGAATTCCATCTATAAGAGCC	TTACAT	CAAT	
					GAAGTCCAATTTGGAAAATTGGATAGATTT	ATCTGG	AACT	
					GGAAAGTTACCCAATGTCAAAGATGAACGA	CTGATC	CAAG	
					TCKAAAGTTTTGGCCGATAATGCTCACATTA	CTGCCA	GNGA	
					AAAAGAAGATGTTAGAGAAATTGGGAGTG	GTAGTC	GTTT	
					GGGATCCCAATCAAAGGAGTGCGTTGTAAG	ATTTTA	AATT	
					AACAAACTCGAGTTCTACAACAAGATGCGT	CTTCCG	CATA	
					GAAGAGTTGATTAAGTCGAATGATGGCATT	CCGCGC	TGCG	
					GGCTTAAAAGAGGCATCTTTGGTTCCCTCTG	TATAAA	CATG	
					CGAACACTTGGTTAAAACTGAGTGACCTAC	ACAGTT	CGGC	
					ATATGAGCGGTCGCACATTTGTAGGATGTC	TAAAAA	ACCA	
					TTAAAACACGGGGAAACCTTATGGCTACCG	CTGAAT	AGGT	
					TAACAAGAACTAGCAGAGGAGGTCAGAAC	AGGAAT	GCTG	
					CCGGGTATAGAGTTGAACTGTAAAAAGGGA	CAAAAA	AATG	
					TGCCAATATCAGGGAAGTCTGAACCATATA	GAACAT	GCAT	
					GTCCAAAAGTGCCCAGTAGTGAAAGGGTTG	GGCAAG	CGAT	
					AGGATAAAACGACATGATGAAGTCGTTAAG	CGACTA	TAAA	
					TATGTGGAAGAGATCACGAAAAAGGCTGGA	TATGTA	CCTCT	
					TGGTCTGCAACAATGGAACCAATAATTCCG	ACTGGG	CCTG	
					TTCGAGGGGTCACACCGAAAACCGGATTTA	CATTCA	TTGT	
					GTTCTGGTCCGGGGCGATCTCGGGAAAGTT	ACATTC	AGAA	
					GTGGATATTCAAATAGTTTCTGATCACTGCG	CCTATT	GCAG	

					GTCTGGATGAAAAAAATAGTTGTAAGATCG	ACATAT	GTCA	
					GTAAGTATGATAATGACATCATACGAAATT	GGTGGT	TAAA	
					ATGTAAGGGGGCTAGGACCATCGAGAGTGG	GCCTGG	TGGA	
					AGGTAGCGGCGATAACCCTTAATTGGCGAG	GGTCTG	GGRG	
					GCGTGTGGAGTAGAGACTCATTCAATCTCA	TTTTAT	GGCA	
					TTAAGAGATTAGGAATGACAGAAATGGATG	ATAATG	ACCA	
					CGAAAATCATTTCTATGAGAGTGCTGGCGA	GGTACC	CTGA	
					GCACTGCGAAGATGTTTAAGACGTGTAAAA	CGGGAA	AACT	
					AAGTGTTAGAGCCCGTTTGCAGAACCAAGA	GTGGAT	TATG	
					CCGCTGATTGTGATGGTTACGGACCGGAGG	CTGTAT	AGCC	
					AAACCTCCGCCCGACCATGTCATGAGCTTA	CACCAG	AGAA	
					ACTTGAARGAGAGTTCGGGGACGTAAGGCC	TCATGG	GAAG	
					ACGCGCGTCGTCCTTGTTTGATCACTAGTGG	TGCCAT	CTTA	
					ATCAACCTTCGACTCCCCGGAACTGTGGGA	ATCTTT	ACTA	
					GTGGCGGAAGAAAGGCCAGAGGATGTCCTG	KGATAA	CAWA	
					AAACCATATATTTATTTATAGAAGTTTTACT	AGATAC	AGTT	
					TCATCCTATTTACGTATTTCAGTATGAAAAT	AGTTTA	TTAG	
					GAGTAAAGTTCTCGACTCGATGAGTTGGGG	AAACTG	GCAA	
					GCAACCATTGGGGGTCCTGAAGAGAGGCTC	CGATGA	TTACT	
					TCACTGTAAAAAATCTCTTCGTGTCTGTTTA	TACTAA	GAAC	
					TTCCTAGGCACNTGCTGCATTATGAAGCGG	TAGAGA	GGAG	
					WGAAAGTAAAGTTAGAGCTGAGAGATAGG	TCCTCT	TTAA	
					TACTTGCTGCATTATGAAGCGGAGAAAGGC	TAGACC	CTGTT	

					CTCGAATAAATAGGGTGTTAGAGTTATTGA	TTCGTA	AGTT	
					TGGAGAGTATACTAGTAAGCTTAAGCTGCG	AAGAAG	AACA	
					CCTCGCGCGGTGCCCAAAAATATACTTAAT	TGGGGA	CTAC	
					GAGAGCAATAACTCAAGGNGAGTTTAATTC	TTGATG	CATG	
					ATATGCGCATGCGGCACCAAGGTGCTGAAT	ACATTA	TAGT	
					GGCATCGATTAAACCTCTCCTGTTGTAGAA	GCATTG	TGTTT	
					GCAGGTCATAAATGGAGGRGGGCAACCACT	GAAGAA	ATAA	
					GAAACTTATGAGCCAGAAGAAGCTTAACTA	TTAAAT	AGCA	
					CAWAAGTTTTAGGCAATTACTGAACGGAGT	CTCCAA	AATA	
					TAACTGTTAGTTAACACTACCATGTAGTTGT	GGAAAT	TCAG	
					TTATAAAGCAAATATCAGGTTTCAGTCTATA	GGAGTA	GTTTC	
					TACTAAAAGTATTTTTTGATACCGTGGTATA	ACTTCA	AGTC	
					TAGGCAACTAGTTAGGAAATAGTAAGGGAT	ATGAAG	TATA	
					GACGCATTGTCTCTCTTTATGGTACTGAGGA	TCCCAC	TACT	
					AACTTATCGACTCGCGAGGGATGAAACCCG	AACCCC	AAAA	
					TAAACCGATCGATYTAGCCTATAAGTACCA	GTTGAA	GTAT	
					GCGACAGTTAAACCATCTTACGCGAGGGGT	GGGCTG	TTTTT	
					AAAACCTGAGGACCGATTATGGTATAACTT	GGTTCG	GATA	
					CTCAAGATTAGCACAAAATGCGAGTGCAAC	AGTATC	CCGT	
					TTGAGGAGGAGGATTTGAGTGTTAATTCAT	GAGAGA	GGTA	
					AATGTACTAATCTAATTAAACTGTGACGGG	AAACTC	TATA	
					AATTGCAGCTTCGGCTGTAATTACTTTGAGG	TAAATT	GGCA	
					CCTATCACGGATTGTAAGGAACATATTGAC	CTCTTC	ACTA	

					ACCGTAAGTCTAACGTGTTCCCGATTTCCAA CCAGGTCATATGAAGGGCTGCCCTTGATAA GGCGGATTTGACCCAATTCTTCATATGAGA GGCTTATTCCAGCCTTCCCGTAGTACCGTGA GGTTTTCCCGCCTCGAACGGAACAATGTTG CAGGGTAATTAAGTACATCGGGCTATATMG CGATATTTAACGTTTTA (SEQ ID NO: 1093)	GGTTMT GTCCAA CGGAGG GGACAT TACTGT AAAATA TCCTCT AAAAAC AACT (SEQ ID NO: 1216)	GTTA GGAA ATAG TAAG GGAT GACG CATT GTCT CTCTT TATG GTAC TGAG GAAA CTTAT CGAC TCGC GAGG GATG AAAC CCGT AAAC CGAT CGAT	

							YTAG CCTA TAAG TACC AGCG ACAG TTAA ACCA TCTTA CGCG AGGG GTAA AACC TGAG GACC GATT ATGG TATA ACTT CTCA AGAT TAGC ACAA	

							AATG CGAG TGCA ACTT GAGG AGGA GGAT TTGA GTGT TAAT TCAT AATG TACT AATC TAAT TAAA CTGT GACG GGAA TTGC AGCT TCGG CTGT	

							AATT ACTTT GAGG CCTA TCAC GGAT TGTA AGGA ACAT ATTG ACAC CGTA AGTC TAAC GTGT TCCC GATT TCCA ACCA GGTC ATAT GAAG GGCT	

							GCCC TTGA TAAG GCGG ATTT GACC CAAT TCTTC ATAT GAGA GGCT TATTC CAGC CTTCC CGTA GTAC CGTG AGGT TTTCC CGCC TCGA ACGG AACA	

							ATGT TGCA GGGT AATT AAGT ACAT CGGG CTAT ATMG CGAT ATTT AACG TTTTA (SEQ ID NO: 1339)	
R2	R2-1_SP	.		Strongyloce ntrotus purpuratus	TCTCGCGACGCGTTCTTCTGCCTGATGAAGT CACGTCAGGTAATAGACTTAGAAGGTTGAT GAGCGTTCCTCTCCTGGACCGGGGGTGAGG ATGTGTTCTACTGAATCATCGTTCCGGTTGT GAGGTCCGCTGCAAATAGGCCTTGGGGTGG TCTACCTCCGGGTCGCTACTCCTTTTGAGCT AGTCGATCCAGGTGAGAGTCGGGGAAGCCC	TCTCGC GACGCG TTCTTCT GCCTGA TGAAGT CACGTC AGGTAA	TAAA CCTT GCCT CCCC GGGC CCCC CTCA	MENSFAWEGTS SAEGRTTVEDSP SSSDDFVSNVGF KVAKADPTVWE EANMSEDNTIIE DPPSSSDDFANN VGFKVTRADPT

					ACTTAGGTGGGCCAGCTAAGCAGATCACCC	TAGACT	GTGA	AWEEASTSTETE
					CCCCAGCACGACAGTGCGCTGTAATATAGC	TAGAAG	CTAA	DLPSSSNFIDNV
					CTGTTGAGAGTGCACCCATTTATACTGTTAT	GTTGAT	GACA	ETQIDMAGPTA
					AAAAGTATAAATGGTTCTTAGCTAACCTAT	GAGCGT	ACTTT	WEDADSNEDNI
					CCAATGATTTTTATTGTTATGTATAATACTG	TCCTCT	CACC	DEGTPNNINNNL
					ACCACTTTAGTCTAGTGTCATGTAAGGATCC	CCTGGA	GTAA	AIVRGRADAYA
					GAGACAATAGCCTATGTCTCTAGTCTAGCA	CCGGGG	TAAT	CSCCERNFISLK
					ATAACACAAGAGGGATAATCCACCTCCCTA	GTGAGG	CATA	AIGTHLKETHNK
					CTCCAGCCACCACCTTTACCTTTCTTATCTT	ATGTGT	TATTT	KVVFECAKCQH
					CCATATAGATGCCAACCAGACGGGTCGATG	TCTACT	GTAT	TFVKAHGLACH
					GGCAGTGAGAAACAAGAATGGGGATAATTT	GAATCA	ACCA	VPKCKEDADTP
					GGAATTTGACATTCTTTCCTAAAGCAGACCT	TCGTTC	TGTA	MLNRLLHGCGE
					GAAGGGTTAAGAGTCACAAGCGGAGTCCGG	CGGTTG	TTAA	CGLAFNTRRGLS
					ACTGTCTTTATAAACAGACAATATTTTCTTA	TGAGGT	TCTA	QHERHRHPSAC
					TTTTACCACCTATATGGGGATATTTCCCAAC	CCGCTG	GGTA	LSTRRRSRLDGI
					TCGTAATATAGGGCCCACTGTCAAGTGGCT	CAAATA	ACAA	ARKKSLRNRRDI
					GGTATAAGTTACCCTGTGGGTAACAAATAA	GGCCTT	TTGA	WTNDEIRLLKQ
					ATTCAAACAAAGAATGGAAAACTCCTTTGC	GGGGTG	AAGT	LMIQYEHAKKIN
					GTGGGAGGGTACCAGTTCTGCTGAAGGCAG	GTCTAC	AACA	IKIAEHFNHKNA
					AACAACAGTTGAGGATTCACCCTCAAGCAG	CTCCGG	TTGA	KQVMHKRRSLR
					TGACGATTTTGTTAGCAATGTCGGATTCAAA	GTCGCT	ACCG	EKDMALGAPHD
					GTAGCTAAGGCGGATCCCACTGTGTGGGAG	ACTCCT	TATTC	APPPLAEEPIIEV
					GAAGCCAATATGTCAGAAGACAATACAATA	TTTGAG	ATAC	VEGAREELEQA

					ATTGAGGACCCACCATCGAGTAGCGACGAC	CTAGTC	TTTA	PVDVILPDLEAL
					TTTGCTAACAATGTCGGCTTCAAAGTAACTA	GATCCA	ATGA	TVNDGRGGGSP
					GGGCAGATCCCACTGCGTGGGAGGAAGCCA	GGTGAG	GTAC	VLTEGGESTRD
					GTACATCTACAGAGACTGAGGATTTACCCT	AGTCGG	AATA	MEENGGTDTRS
					CGAGCAGTAATTTTATTGACAATGTCGAGA	GGAAGC	GTGA	PSPREERAGSTP
					CCCAAATAGATATGGCGGGCCCCACTGCGT	CCACTT	AGGG	WERGWQPRVD
					GGGAGGACGCCGATAGCAATGAAGATAAC	AGGTGG	AACT	RGRGEYKGYGG
					ATAGATGAGGGCACACCCAATAATATTAAT	GCCAGC	GTAT	ERGDRHTVSLS
					AATAATTTAGCAATAGTCAGAGGGAGGGCG	TAAGCA	ATTA	QRGESGVDPLTP
					GATGCCTATGCATGCTCGTGTTGCGAACGG	GATCAC	CATA	GGVVEDYDDSY
					AACTTCATTTCTCTCAAGGCAATCGGTACTC	CCCCCC	CCTC	LEDYFPGWDED
					ACCTCAAGGAAACCCATAACAAAAAGGTTG	AGCACG	GAGA	EHMHIIGRLDLS
					TCTTTGAATGTGCCAAATGTCAACACACCTT	ACAGTG	TAGA	DESEQGEAAVSP
					CGTTAAAGCACACGGACTAGCGTGCCACGT	CGCTGT	GGTT	RLGSFGDLLEEV
					CCCCAAATGCAAAGAGGACGCCGACACACC	AATATA	TTTGT	NAMEGKDNLSE
					TATGCTTAATAGACTGCTCCACGGCTGTGG	GCCTGT	ACCT	ALAETLGLVLH
					AGAGTGCGGTCTTGCATTCAATACGAGAAG	TGAGAG	TAAG	EGHRVEYIKEK
					AGGGCTGTCTCAACACGAGCGCCATAGGCA	TGCACC	GGTT	MNINVKQMATE
					CCCGTCGGCGTGTCTAAGTACTAGGAGAAG	CATTTA	CGTG	ILAHGANKGNP
					GTCACGGCTAGATGGTATAGCTAGGAAAAA	TACTGT	AGAA	KRRKEAVAAKR
					ATCATTGAGGAATAGGAGGGATATTTGGAC	TATAAA	TCCA	NPGTRLDRAQR
					GAATGATGAGATCCGTCTACTCAAACAACT	AGTATA	TACT	DNQAKAKAKEK
					CATGATACAGTACGAGCATGCTAAAAAGAT	AATGGT	GCAT	KRIFSETQTQYK

					AAACATTAAGATAGCGGAACATTTCAACCA	TCTTAG	AAAG	KNPHRLVEKLL
					TAAGAATGCAAAACAAGTCATGCATAAACG	CTAACC	GGGT	DGKGDERCSVS
					TAGATCCTTAAGGGAAAAGGACATGGCCTT	TATCCA	TAAC	LEVIQRTYMNRF
					AGGTGCTCCACATGATGCACCTCCGCCATT	ATGATT	TTGT	SRESKEVDIGAY
					GGCAGAAGAACCGATAATAGAAGTAGTTGA	TTTATT	AACT	VDPETVEDNQGI
					GGGTGCGAGAGAAGAGTTAGAACAGGCTCC	GTTATG	CAAC	VDPISKAEIERAI
					TGTAGATGTAATCCTTCCAGATCTTGAAGCA	TATAAT	CCCA	STTKKGSAPGPD
					CTCACGGTTAATGATGGGAGAGGGGGTGGC	ACTGAC	GGGA	GVTYDALKAYG
					AGCCCCGTGTTGACCGAGGGAGGGGAGAGT	CACTTT	GACA	NCQLYLLIMYN
					ACAAGGGATATGGAGGAGAACGGGGGGAC	AGTCTA	AGCC	TWLAMGKVPSE
					CGACACACGGTCTCCCTCTCCCAGAGAGGA	GTGTCA	GACA	AKTYRSILIPKG
					GAGAGCGGGGTCGACCCCCTGACTCCGGGG	TGTAAG	AATT	QGDPMDINNFR
					GGAGTGGTAGAAGATTATGATGACTCGTAC	GATCCG	CGGC	PLTLANVISRLY
					TTAGAAGACTACTTCCCTGGGTGGGATGAG	AGACAA	ATTA	SKILTRRLDGAV
					GATGAGCACATGCATATTATTGGGAGGCTC	TAGCCT	CTAT	SVCPRQRGFTH
					GACCTAAGTGACGAGTCGGAACAAGGGGA	ATGTCT	GTGA	KASIEDNTLILRE
					GGCTGCTGTCAGTCCAAGACTAGGCTCCTTC	CTAGTC	GGGT	LIMKSKRNKECL
					GGAGACCTATTAGAAGAAGTTAATGCAATG	TAGCAA	CATA	AVVLLDLAKAF
					GAAGGCAAGGATAACCTGTCTGAGGCACTG	TAACAC	AGTG	DTVSHDLIIKAL
					GCCGAAACCCTGGGCCTTGTGTTACACGAA	AAGAGG	TTAA	RRHRVHEHLISV
					GGTCATAGAGTAGAATACATTAAAGAGAAA	GATAAT	AGGA	IMDLYEGGTTSF
					ATGAACATAAATGTTAAACAAATGGCAACA	CCACCT	CCCA	TTDEGTTCPIAIR
					GAAATATTAGCCCACGGGGCAAATAAGGGA	CCCTAC	TTGT	SEVKQGDSLSPV

					AACCCCAAACGGAGAAAGGAGGCGGTCGC	TCCAGC	ATAT	LFNLALDPLLAT
					CGCCAAGCGGAATCCCGGTACGAGATTAGA	CACCAC	AACT	LEQRGKGVEIG
					TCGAGCCCAGAGGGACAACCAAGCGAAGG	CTTTAC	TGTA	GHTFVSLAYAD
					CCAAGGCCAAGGAGAAAAAGAGAATCTTCT	CTTTCTT	AATG	DTALVSSSHLD
					CCGAAACACAGACCCAATATAAGAAGAATC	ATCTTC	TACA	MTANLDITVEY
					CCCACAGACTCGTGGAAAAACTTCTTGATG	CATATA	GAGA	LNATGLSLNVR
					GGAAGGGTGACGAGCGGTGCTCTGTCAGCC	GATGCC	ATTC	KCQGFLLTPINK
					TTGAAGTGATCCAAAGGACCTACATGAACA	AACCAG	GAGT	SFLVNEAESWV
					GGTTCTCAAGAGAAAGCAAAGAGGTTGACA	ACGGGT	TGAG	VEREAIPWVEPG
					TAGGAGCCTATGTTGACCCCGAGACTGTGG	CGATGG	TTCG	DTAKYLGVQVG
					AGGATAACCAAGGGATAGTGGACCCCATCA	GCAGTG	AAGA	PWSRPWPSIQPV
					GCAAGGCTGAAATAGAAAGGGCCATCTCTA	AGAAAC	ATAA	IKRLTAYCESID
					CCACGAAAAAGGGATCCGCACCTGGACCGG	AAGAAT	ACAA	KAALKPRQRIHI
					ACGGGGTCACCTATGATGCCCTAAAGGCGT	GGGGAT	ACAA	LTTYIAPRIAFEI
					ATGGAAACTGTCAATTATACTTATTAATAAT	AATTTG	AAAA	AEGGYSTLVDC
					GTACAACACATGGCTGGCCATGGGGAAGGT	GAATTT	GAAA	RGGIQYTRIREV
					CCCATCAGAGGCAAAAACCTACAGGTCAAT	GACATT	ACAA	DMTIRNYVRKW
					TCTAATCCCAAAAGGTCAAGGGGACCCCAT	CTTTCC	AGAG	LFLPACLSNSFL
					GGACATCAACAATTTTCGTCCATTAACACTC	TAAAGC	AATT	YTRRGEGGLGL
					GCCAATGTAATCAGCAGATTATATTCCAAA	AGACCT	CGTT	VSFYDYVPTER
					ATCTTGACCAGGAGGCTGGACGGGGCGGTA	GAAGGG	CGTC	MRKLVRVCDSE
					AGTGTCTGTCCGAGGCAAAGAGGCTTCACT	TTAAGA	CATA	DPVIAGAAASLG
					CACAAAGCTAGTATTGAAGACAATACACTC	GTCACA	GAAG	LRERAAKISAQT

					ATCCTACGGGAACTCATCATGAAGAGTAAG	AGCGGA	CGGA	GLPVPVKPKGA
					AGAAATAAAGAATGCCTTGCAGTGGTACTT	GTCCGG	AAAA	HNAWRKVQKK
					CTTGACTTAGCCAAGGCTTTCGACACCGTTT	ACTGTC	GCGC	KWKAQPTQGK
					CCCACGATCTCATCATCAAGGCCCTGAGGA	TTTATA	AAAA	GVSCYQHRLGN
					GGCATCGGGTGCATGAACATCTCATTAGCG	AACAGA	AGGC	KWLGAPSFLTE
					TCATAATGGACCTGTACGAAGGGGGAACTA	CAATAT	GTCC	NDYIWAIKLRTN
					CATCTTTCACGACAGATGAGGGCACCACCT	TTTCTT	TAAC	LVPTREAMGRGI
					GCCCCATCGCGATCCGGAGCGAAGTTAAGC	ATTTTA	CTCA	IGRNQVECRHC
					AAGGAGACTCTTTGTCGCCGGTGTTATTCAA	CCACCT	CGGT	HTTIETMGHISG
					CTTGGCTCTCGACCCGCTACTAGCTACACTC	ATATGG	CTAA	YCQMVLDIRLIR
					GAGCAAAGAGGGAAAGGTGTTGAGATCGG	GGATAT	CCCA	HNRICKALIKAA
					GGGACACACGTTCGTCAGCCTGGCCTATGC	TTCCCA	GAAA	TATGLRVTEEPR
					GGATGACACGGCACTCGTCTCGTCCAGCCA	ACTCGT	CATTT	IVGTDGKNYLP
					CCTGGACATGACCGCAAACCTAGACATAAC	AATATA	AAGG	DLIFSAGAGEPC
					TGTAGAATACCTCAATGCAACGGGACTGAG	GGGCCC	GGGG	YVVDPTVVWD
					TCTCAATGTGAGGAAATGTCAAGGCTTCTT	ACTGTC	AAGG	DDPKNLREAWR
					ACTCACCCCCATCAATAAGTCCTTCTTAGTC	AAGTGG	AGCT	GKVRKYTPIIPA
					AATGAAGCCGAGAGCTGGGTAGTGGAACG	CTGGTA	CATT	VEAMLHPSSVQI
					GGAAGCTATCCCGTGGGTTGAACCTGGGGA	TAAGTT	ACTC	FGFVCGARGTW
					CACAGCTAAATATCTGGGGGTACAAGTGGG	ACCCTG	CAAT	CPMNDDIAKIVG
					TCCGTGGTCACGTCCGTGGCCATCCATTCAA	TGGGTA	CCAA	LKNSGISRTLQI
					CCAGTAATTAAACGCCTCACAGCCTACTGT	ACAAAT	AACC	VLCDTIRMVKA
					GAGAGCATCGACAAAGCGGCGTTAAAGCCA	AAATTC	ACGT	

					AGACAGAGAATACACATTCTAACGACCTAT ATTGCCCCTAGGATTGCCTTTGAGATTGCAG AGGGGGGATACAGTACACTCGTATAAGGGA GGTAGACATGACAATCAGAAACTACGTCAG GAAATGGCTATTCTTGCCCGCCTGCCTTTCA AACTCCTTCCTTTACACAAGGAGAGGAGAG GGTGGTCTTGGGCTGGTAAGCTTTTACGACT ACGTACCGACTGAACGTATGAGGAAGCTAG TTAGGGTGTGTGATTCTGAAGATCCGGTTAT AGCAGGGGCAGCAGCCTCGCTAGGGTTAAG GGAGAGAGCTGCGAAGATAAGCGCTCAGA CAGGGTTGCCAGTTCCAGTTAAACCCAAAG GGGCACACAATGCATGGAGAAAAGTCCAG AAGAAAAAGTGGAAAGCTCAACCCACCCA GGGGAAGGGAGTTAGTTGTTACCAACACAG GTTGGGCAACAAATGGTTAGGTGCCCCCAG CTTCCTCACTGAAAATGACTACATTTGGGCC ATTAAACTGAGGACAAACCTCGTCCCTACT AGGGAAGCTATGGGGAGAGGTATCATAGG GAGGAATCAAGTGGAGTGCAGACACTGCCA TACAACAATAGAAACAATGGGACACATTTC TGGATACTGCCAGATGGTCCTGGATATTCGT TTGATTAGACACAACCGTATATGCAAAGCC	AAACAA AGA (SEQ ID NO: 1217)	TCCCT AGCC CAAA AGCG CGCT TCAA TGAG GGAG CCGA TATT GGAG GACA GAGT TCAT CAAA AGCC ACCT GTAA GTAG GCCC ATTTC TGCC AAGG	FMAR (SEQ ID NO: 1462)

					CTGATCAAGGCTGCTACGGCTACAGGACTC		ACAT	
					AGAGTTACGGAAGAACCACGAATAGTGGGT		GCGC	
					ACGGATGGGAAAAACTATTTGCCAGATCTT		AAAA	
					ATTTTCAGTGCGGGAGCTGGGGAGCCATGT		GGAA	
					TATGTAGTTGATCCAACAGTGGTGTGGGAT		GCAG	
					GATGATCCCAAGAACCTGAGGGAGGCCTGG		ATTA	
					AGGGGCAAGGTGAGGAAATACACACCAAT		TCAA	
					CATTCCGGCGGTTGAAGCGATGCTTCACCC		ACCA	
					ATCATCAGTTCAAATATTTGGATTTGTATGT		GTCA	
					GGCGCAAGGGGAACATGGTGCCCAATGAAT		AACA	
					GATGACATAGCCAAGATTGTCGGTTTAAAG		AGCA	
					AACTCAGGAATTAGCAGGACACTACAAATA		CAAA	
					GTCCTTTGCGACACCATAAGAATGGTCAAG		CATT	
					GCCTTCATGGCCAGATAAACCTTGCCTCCCC		GGGG	
					GGGCCCCCCTCAGTGACTAAGACAACTTTC		GGAT	
					ACCGTAATAATCATATATTTGTATACCATGT		GGGA	
					ATTAATCTAGGTAACAATTGAAAGTAACAT		TAAC	
					TGAACCGTATTCATACTTTAATGAGTACAAT		CCCG	
					AGTGAAGGGAACTGTATATTACATACCTCG		GAAA	
					AGATAGAGGTTTTTGTACCTTAAGGGTTCGT		GAGA	
					GAGAATCCATACTGCATAAAGGGGTTAACT		GGGA	
					TGTAACTCAACCCCAGGGAGACAAGCCGAC		TCTTT	
					AAATTCGGCATTACTATGTGAGGGTCATAA		AGAT	

					GTGTTAAAGGACCCATTGTATATAACTTGTA AATGTACAGAGAATTCGAGTTGAGTTCGAA GAATAAACAAACAAAAAAGAAAACAAAGA GAATTCGTTCGTCCATAGAAGCGGAAAAAG CGCAAAAAGGCGTCCTAACCTCACGGTCTA ACCCAGAAACATTTAAGGGGGGAAGGAGCT CATTACTCCAATCCAAAACCACGTTCCCTAG CCCAAAAGCGCGCTTCAATGAGGGAGCCGA TATTGGAGGACAGAGTTCATCAAAAGCCAC CTGTAAGTAGGCCCATTTCTGCCAAGGACA TGCGCAAAAGGAAGCAGATTATCAAACCAG TCAAACAAGCACAAACATTGGGGGGATGGG ATAACCCCGGAAAGAGAGGGATCTTTAGAT AGTGGATGGAAGGGGTGGGACGTTGAATCC AACCATGCCGTTTTTATGTTCCCGATAAAGA AGGATAAGGTCACTCCAGCCTGACACACAA AGTGGGGTAAAAGAACTCCGCTCGTACGGA CTCCAAATAGAA (SEQ ID NO: 1094)		AGTG GATG GAAG GGGT GGGA CGTT GAAT CCAA CCAT GCCG TTTTT ATGT TCCC GATA AAGA AGGA TAAG GTCA CTCC AGCC TGAC ACAC AAAG	

							TGGG GTAA AAGA ACTC CGCT CGTA CGGA CTCC AAAT AGAA (SEQ ID NO: 1340)	
R2	R2-1_SSa	AGKD01072455		Salmo salar	AATCTTTAACCCCGGACTCTTGGGGTTCTTA CGACTCTGTATGAGGAACAGTCGAAGAGAG GGCGCTACCAATCCAAGTATATGTCCCAAG AGGGCTGGGACAGGGTGGAAGAGTGCACCT CGCGATCTGGGGCAGGGAAGGAATGGAGA GAAGTCGAAGAAGGCTTGTAGAGAAGGGG CTCTCCTAGATCCTAACCTGTATGACGCCCG TAAAACGGTGACCCCAGTAGCGAATAAAGG AGGCAGGTGACAATAGAGGGCAGGGCCGA CTTCCCAGGTTTACATTGTTGTACTTGTCAA	AATCTT TAACCC CGGACT CTTGGG GTTCTT ACGACT CTGTAT GAGGAA CAGTCG AAGAGA	TAGA TCCG TTTGT TATG ATTG GAGG GAGC CTGC CGAG TGGT	MSGKRIVEMSG CDEKICQNKHC LKRRWAWISGP KGETSPPRKRGT CENVSFQDKSH ASDPDPLKAPEA REDAGSVAPQW VGEIKTPSLTSR DGVSEVVLPPQP VHAEGVSPASD

					CATAAAGAGGTGTCTCAATAGTTTGAATCA	GGGCGC	ATGA	SKDKATKITLLIS
					ACAAGGGAGAGGAATACCGACCTGCTCCCT	TACCAA	GCGC	LPVCDLRCGRC
					TGGGGCGGGGGTACTGGTCTTAGCCCGGTT	TCCAAG	TCCA	ERPLETVGKAV
					CCCCGCAAGTTTCCTTTGCCTGGGATGTGCC	TATATG	ACTA	RHFAVAHPTVS
					TGACTGGCTCCATCCCCTTTCCCCATTAGGC	TCCCAA	TTGA	VVFKCQKCEKS
					ACGGCTAGATGACGCACCGATGGGCGGGTG	GAGGGC	ACCC	SKNSHSISCHIPK
					TGTAGGTCGCTACCGAAGGGGACTGGGGGT	TGGGAC	ATAT	CKGMTETRTDV
					GTCCGGTGAACCAGGACTTCCCAAAATGGT	AGGGTG	GATT	EGDHGCDHCQE
					CTCACATTTTTAAGCGGCTTGAGTATCGCCC	GAAGAG	CCCG	KFTTAMGLTQH
					AGTATCCTCGCGCGGCACTGGGAACCCAGT	TGCACC	AGGC	KRHRHIVQYCK
					CAACCGCTCTGTGCCCCGGCGCAGGCGGGG	TCGCGA	CTGG	EKEGEMTARRK
					GTTTAATGTCTCCCCGGCTTCACCGGCGCTT	TCTGGG	CCAG	GEVEAVKWSE
					CGGCGACGACGCAGAGGAGCACCCGGAGG	GCAGGG	ACGC	WEESEVARLSD
					CCCCCATGAACTTAAACCAACCTATCTTGA	AAGGAA	CTAG	GLAGLKMINRRI
					AATATGGCCTCTCGTTCGGGTGAAGGGCAG	TGGAGA	ATGC	ADSLGTGKTAE
					GTGGGAAGAGAGGGCTGCCTCACGATAAAC	GAAGTC	CTGC	QVRQKRRRMRP
					ACCTAGTCAATAGCCAGTCGGGAAAAATGT	GAAGAA	CACA	EKVRCDKPKEA
					GGAATGTTAGGACAGGGAGGTAAGGGAGG	GGCTTG	ATTG	KDKSNLIKMLSI
					CGGCTTTTTCGTAAGGCTCCTTCAACCCCTA	TAGAGA	AACG	PSATPTPQTGLK
					CCTGTAGTCCACCTATTGCAGGTGTTGACAA	AGGGGC	CAGC	GFLLGELNGVA
					CATGCAAGATGACCTGCCTCGTTACGGGTC	TCTCCT	CCTA	TKGEVQIGGVTL
					GCGTATCATTGCTACAGGTCGTGTGCCGCTT	AGATCC	GCTT	SLRGVEQDSALL
					CTAAGAGGATAGTAAGGAGAGGTTATAGGG	TAACCT	GCTA	NTSALELQRLLG

					AGGTCCTGTTAGGGCTTCCTCAACCCCTCTC	GTATGA	GGAG	GRAGSANPLSL
					TATGCGATTCCTTACAGGAGTGGATCGAGA	CGCCCG	ATCC	QRERETTLPSER
					AGTCCCGGACGTAATACACCCTGGAGGTAA	TAAAAC	ATAG	RKTKQGEYRRV
					GGGAGTGGCCTTCTAGTAGGGCTGCTTCAA	GGTGAC	GAAC	QKMFRSNEKKI
					CCCCTCTGATGGGAGTGTACCGGGAACCTC	CCCAGT	TGGC	AKYILDGNGDG
					GACTTGTAAGCACAGGTTAGTATGGGAGCA	AGCGAA	CTAT	EAASPPLEIALA
					GAAGGGGGAGCCGTAATGGGCTTCTCTTTC	TAAAGG	GGGG	FKSRWEEVETF
					ACCTGCTTACATAATACCTGTGGTGCATGTA	AGGCAG	CGTC	HGLGQFYSRGE
					TCTAGGTCTTGGCGGGAGAGTACTGAGAGA	GTGACA	ATGA	ADGVVFRSLISM
					CAAGGTTGAGACCCCAAGATTGGGTCTCCC	ATAGAG	CGGT	SEVCENLGAIKN
					TAGCCTCTATAGCTGCGACTCTTAGCGGGG	GGCAGG	TGAA	NTAAGPDGITKP
					ATATGGAGTAACATGTACCAAGGGAGTGAA	GCCGAC	GTTC	ALLEWDPTGAK
					TAATAAAGGAATTGACGGGGTACAAGTGAC	TTCCCA	CTCC	LAAIFSIWLTSG
					TGTTGGCCCGAATCCTAGCCACTTGATGACC	GGTTTA	ATAG	TLPGPFKKCRTT
					AGGGATATATTACAGACTGAGAGCGAATCT	CATTGT	CGTG	LIPKTDDPILLTQ
					AGAGACGTGAGAATAAAGTGAGAATTGGTT	TGTACT	CTTG	VAGWRPLTIGS
					GAGCGAATACAGAGGAAGATGAGCGGAAA	TGTCAA	GGAG	VVLRLYSRILTH
					GAGAATTGTTGAGATGAGCGGTTGCGATGA	CATAAA	GGGA	RLERACPINPRQ
					GAAAATCTGCCAAAATAAGCATTGCTTGAA	GAGGTG	CGAC	RGFISSPGCSENL
					GCGCAGATGGGCGTGGATTTCCGGCCCAAA	TCTCAA	AATG	MILGGLIKRSWA
					GGGGGAGACATCTCCTCCTCGCAAGAGAGG	TAGTTT	ACGA	KGERLAVVLVD
					AACTTGCGAAAACGTATCCTTCCAGGATAA	GAATCA	GTCA	FARAFDSVSHSH
					ATCTCATGCCTCGGACCCAGATCCTCTCAAA	ACAAGG	TGAC	ILEILRQRGLDE

					GCCCCGGAAGCGAGAGAGGACGCGGGTTC	GAGAGG	GTAC	HIIGIVGDSYTD
					GGTGGCGCCTCAGTGGGTTGGCGAAATAAA	AATACC	CGAG	VTTTITVSGEQS
					GACACCTAGCCTAACATCGCGGGATGGGGT	GACCTG	AGAA	PPIDMRVGVKQ
					TAGTGAAGTCGTGCTGCCACCACAGCCAGT	CTCCCT	CCCC	GDPMSPLLFNLA
					TCATGCAGAGGGAGTATCCCCAGCCAGCGA	TGGGGC	AACC	LDPMIDTLERYG
					CTCAAAGGACAAGGCTACGAAGATCACCCT	GGGGGT	CAGG	LGYRMGEQQIT
					GCTGATATCACTCCCCGTCTGTGATTTGAGA	ACTGGT	TTGG	ALAFADDLVLV
					TGCGGACGATGTGAAAGACCATTAGAGACC	CTTAGC	GGGA	SDSWEGMACNI
					GTTGGGAAGGCGGTGAGACACTTTGCTGTG	CCGGTT	GAGA	RILEEFCRLTGL
					GCTCATCCGACGGTGTCGGTGGTTTTTAAGT	CCCCGC	GCCA	RIQPRKCHGFLI
					GTCAAAAGTGTGAGAAGAGCAGCAAGAAT	AAGTTT	GCAA	QKIQRARSVNLC
					AGCCACTCCATCTCCTGTCATATCCCTAAAT	CCTTTG	GAGC	KPWIVCGEELH
					GTAAGGGAATGACAGAGACCCGGACGGAT	CCTGGG	GGAG	MVGPEESVSYL
					GTGGAAGGTGATCACGGCTGTGATCATTGC	ATGTGC	ATGC	GMKVSPWHGIM
					CAGGAGAAGTTTACAACGGCTATGGGGCTG	CTGACT	TTGG	EPDPVERLCNWI
					ACTCAACACAAGAGACACAGACACATCGTC	GGCTCC	TATA	SSIGRSPLKPSQK
					CAGTATTGTAAGGAGAAGGAGGGGGAGAT	ATCCCC	CCAA	VRMLNVYAAPR
					GACAGCAAGAAGGAAGGGTGAAGTCGAAG	TTTCCC	GCTA	MTYQADHGGL
					CAGTCAAATGGAGCGAATGGGAAGAAAGT	CATTAG	GCAG	GPIVLNVLDGMI
					GAGGTGGCAAGGTTGAGCGATGGACTGGCT	GCACGG	AGAG	RKAVKVWLHLP
					GGGCTAAAAATGATCAACAGGCGAATCGCA	CTAGAT	AGGG	LCTCDGLLYSRC
					GATAGCCTGGGGACTGGAAAAACTGCGGAA	GACGCA	TTGA	QDGGLGIVKLA
					CAGGTGAGGCAGAAAAGACGTAGAATGAG	CCGATG	AGAG	CQIPSIQARRVY

					ACCTGAGAAGGTACGGTGTGACAAACCTAA	GGCGGG	GATG	RLWHSKEAITR
					GGAGGCAAAAGATAAGAGCAATCTTATCAA	TGTGTA	ACTA	VVTRRTVEAEE
					AATGCTGTCCATACCGAGTGCAACACCAAC	GGTCGC	CTGG	YRGMWLRAGG
					ACCCCAAACTGGGCTCAAAGGATTCCTCCT	TACCGA	GCTC	SEAGLPPLEDRE
					TGGAGAACTAAATGGGGTTGCTACCAAAGG	AGGGGA	AGAG	EGAVQCTDTAG
					TGAAGTACAGATTGGGGGAGTTACGTTGTC	CTGGGG	TCAT	SVKPKNPVIPDW
					CCTAAGGGGGGTAGAGCAAGACTCAGCCTT	GTGTCC	CTCA	RRAEFLKWQNL
					GCTCAATACGTCGGCCCTTGAACTGCAACG	GGTGAA	CCCT	TAQGVGVQVFG
					ATTGTTGGGAGGGAGAGCGGGGAGCGCGA	CCAGGA	AAAA	GDKNSNHWMA
					ATCCGTTAAGCTTACAGAGAGAACGTGAGA	CTTCCC	GGCG	NPETLGSKERHY
					CCACATTGCCAAGTGAAAGAAGGAAAACG	AAAATG	GTGG	IAGLQLRANVYP
					AAACAAGGAGAGTACCGGAGGGTACAAAA	GTCTCA	GGCA	TREALSRGRPDL
					GATGTTTCGAAGCAATGAAAAGAAAATTGC	CATTTT	TCGG	PKVCRQCLAGT
					TAAATACATTCTAGATGGCAATGGAGATGG	TAAGCG	TTGA	ESCAHILGQCPA
					AGAAGCGGCCTCCCCGCCACTCGAGATCGC	GCTTGA	ACAC	VKDSRIRRHHKL
					TTTGGCGTTCAAAAGTAGGTGGGAGGAAGT	GTATCG	CTAC	CDLLASEAESAG
					GGAAACATTTCATGGGCTTGGCCAGTTCTA	CCCAGT	CCAT	WTVIREMCCRT
					CTCGAGGGGGGAAGCGGATGGTGTGGTCTT	ATCCTC	ACCG	RAGALRRPDLV
					CCGGTCACTTATCTCAATGAGCGAAGTATG	GCGCGG	GGAT	FVKTGFALVVD
					TGAAAACCTGGGGGCAATTAAAAACAACAC	CACTGG	GGGA	VTVRYEMAYDT
					AGCAGCTGGGCCAGATGGGATAACAAAACC	GAACCC	GGTG	LMGAAAEKVAR
					GGCATTGCTTGAATGGGATCCCACTGGTGC	AGTCAA	GTAG	YTPITPYVAMTL
					GAAACTGGCCGCTATCTTTTCCATATGGTTG	CCGCTC	GCCG	KARRVKVFGFP

					ACATCGGGCACCCTGCCTGGGCCTTTTAAG	TGTGCC	AAAA	LGARGKWPGSN
					AAGTGCAGAACAACCTTAATACCCAAGACC	CCGGCG	AGAA	DRLLKAMGVGG
					GATGACCCGATTTTACTCACCCAGGTGGCT	CAGGCG	CAGG	GRRKQLAKLFS
					GGGTGGAGGCCTCTCACTATCGGGTCGGTG	GGGGTT	AAGA	RRALLYSLDVL
					GTTTTGAGGCTATACTCTCGCATCCTGACAC	TAATGT	TGGT	RDFYRAEGETG
					ACAGGCTGGAGCGGGCGTGTCCCATTAACC	CTCCCC	GGAG	DLDDESVDDHL
					CGCGCCAGAGAGGATTCATTTCCTCACCTG	GGCTTC	TAAG	(SEQ ID NO:
					GGTGTTCGGAAAACTTGATGATCTTGGGAG	ACCGGC	TTGA	1463)
					GTCTAATCAAGAGAAGTTGGGCGAAGGGCG	GCTTCG	GAGC	
					AGAGGCTGGCGGTAGTGTTGGTTGACTTTG	GCGACG	GGTT	
					CGCGTGCGTTTGACTCTGTGAGCCACTCGCA	ACGCAG	GCTC	
					CATCTTGGAGATACTCAGACAGAGAGGGCT	AGGAGC	GGGA	
					TGATGAACATATCATCGGAATCGTAGGTGA	ACCCGG	AGTT	
					CTCGTACACCGATGTAACGACCACAATTAC	AGGCCC	ATGT	
					AGTCAGTGGGGAGCAGTCCCCTCCCATTGA	CCATGA	TGTG	
					CATGAGGGTAGGGGTTAAGCAGGGAGACCC	ACTTAA	ATAA	
					GATGTCTCCGCTACTGTTTAATCTAGCCCTG	ACCAAC	CTCC	
					GACCCAATGATCGACACCCTTGAACGCTAC	CTATCT	ATTA	
					GGCTTGGGGTACAGGATGGGCGAGCAGCAG	TGAAAT	AGGC	
					ATCACGGCCCTAGCTTTTGCTGACGATCTGG	ATGGCC	CGGT	
					TTCTGGTGAGCGACTCGTGGGAGGGCATGG	TCTCGT	GGGC	
					CGTGCAATATCCGTATTCTGGAAGAATTTTG	TCGGGT	ATGG	
					TCGACTGACTGGGCTGAGGATTCAGCCTAG	GAAGGG	TGCG	

					GAAGTGTCATGGTTTCCTCATACAGAAGAT	CAGGTG	GATA	
					TCAGAGGGCGAGATCGGTAAACCTCTGTAA	GGAAGA	ATGG	
					GCCCTGGATAGTGTGTGGTGAAGAACTACA	GAGGGC	AAAC	
					TATGGTCGGGCCGGAAGAGTCGGTTTCCTA	TGCCTC	TATA	
					CCTGGGTATGAAGGTGAGCCCATGGCATGG	ACGATA	AAAA	
					CATTATGGAGCCAGATCCTGTCGAACGACT	AACACC	CAAT	
					CTGTAACTGGATCAGTTCGATTGGGCGGTC	TAGTCA	AAAA	
					ACCGCTGAAGCCTTCTCAGAAGGTGAGGAT	ATAGCC	AGAA	
					GTTGAATGTTTATGCTGCCCCGAGGATGACT	AGTCGG	AGAC	
					TATCAGGCGGATCATGGCGGGCTGGGGCCA	GAAAAA	CAAA	
					ATTGTCCTGAATGTACTCGATGGGATGATC	TGTGGA	AAAA	
					AGGAAAGCAGTGAAGGTGTGGCTACACCTT	ATGTTA	TGTTC	
					CCGCTGTGTACCTGTGATGGGCTACTTTACT	GGACAG	TGTT	
					CTAGATGCCAGGACGGTGGACTGGGCATAG	GGAGGT	ATGA	
					TAAAATTGGCTTGTCAAATCCCTTCTATCCA	AAGGGA	TGCC	
					AGCTAGAAGGGTCTACCGCCTGTGGCATTC	GGCGGC	TTAC	
					TAAGGAAGCCATAACGCGGGTAGTCACCCG	TTTTTC	ACAT	
					AAGGACGGTCGAAGCGGAAGAGTACCGTG	GTAAGG	GTCT	
					GGATGTGGCTGAGAGCCGGTGGGAGTGAGG	CTCCTT	GGGA	
					CAGGTTTGCCTCCCCTGGAAGATAGGGAAG	CAACCC	GACC	
					AAGGTGCTGTACAGTGTACAGACACTGCCG	CTACCT	CCAT	
					GTTCGGTGAAGCCGAAAAACCCAGTCATTC	GTAGTC	AAGG	
					CCGATTGGAGGCGAGCTGAGTTCCTCAAGT	CACCTA	GTCT	

					GGCAAAACCTGACAGCGCAAGGGGTTGGA	TTGCAG	CCCC	
					GTGCAGGTCTTTGGCGGTGACAAAAACAGC	GTGTTG	TTAT	
					AATCACTGGATGGCGAATCCGGAAACGTTG	ACAACA	ACTT	
					GGATCGAAAGAGCGCCACTATATTGCAGGT	TGCAAG	CACT	
					CTACAGTTGAGGGCCAATGTATATCCAACT	ATGACC	GGGA	
					CGCGAGGCACTGTCCAGGGGCAGGCCGGAC	TGCCTC	AACC	
					TTACCTAAAGTCTGCCGGCAATGCCTAGCA	GTTACG	CCAT	
					GGAACTGAGTCTTGCGCGCATATTCTCGGG	GGTCGC	AAGG	
					CAGTGTCCTGCAGTGAAGGATAGCCGCATC	GTATCA	GTAT	
					AGGCGGCATCATAAACTGTGTGACCTGCTA	TTGCTA	CCCC	
					GCAAGTGAAGCCGAAAGCGCCGGATGGAC	CAGGTC	CTAT	
					CGTCATCAGAGAAATGTGTTGCAGAACTCG	GTGTGC	ATTT	
					TGCCGGAGCTTTGCGGCGTCCAGACCTGGT	CGCTTC	ACTG	
					GTTTGTGAAAACCGGTTTTGCTTTGGTGGTG	TAAGAG	GGAG	
					GATGTTACTGTGCGGTACGAGATGGCCTAT	GATAGT	ACCC	
					GATACGCTCATGGGTGCGGCTGCCGAGAAA	AAGGAG	CATA	
					GTTGCTCGGTACACCCCAATTACTCCATATG	AGGTTA	AGGG	
					TTGCGATGACCCTGAAGGCAAGGAGAGTCA	TAGGGA	TCTCC	
					AGGTGTTTGGCTTTCCACTGGGAGCCCGAG	GGTCCT	CCCT	
					GCAAATGGCCGGGAAGCAACGACCGGCTGC	GTTAGG	ATAG	
					TGAAGGCTATGGGTGTTGGTGGCGGCAGGA	GCTTCC	ATGT	
					GGAAACAGCTGGCCAAGTTGTTTAGCAGGA	TCAACC	AGAG	
					GGGCGCTCCTGTACTCCTTAGATGTCCTTAG	CCTCTC	CGTA	

					GGACTTCTACCGGGCGGAGGGAGAAACGG	TATGCG	AGGG	
					GGGACTTGGATGATGAGAGCGTGGATGATC	ATTCCT	GTCT	
					ATCTATAGATCCGTTTGTTATGATTGGAGGG	TACAGG	CCAA	
					AGCCTGCCGAGTGGTATGAGCGCTCCAACT	AGTGGA	AGTA	
					ATTGAACCCATATGATTCCCGAGGCCTGGC	TCGAGA	CCGG	
					CAGACGCCTAGATGCCTGCCACAATTGAAC	AGTCCC	CCGA	
					GCAGCCCTAGCTTGCTAGGAGATCCATAGG	GGACGT	TATG	
					AACTGGCCTATGGGGCGTCATGACGGTTGA	AATACA	GCCT	
					AGTTCCTCCATAGCGTGCTTGGGAGGGGAC	CCCTGG	TATG	
					GACAATGACGAGTCATGACGTACCGAGAGA	AGGTAA	GCAA	
					ACCCCAACCCAGGTTGGGGGAGAGAGCCAG	GGGAGT	ACTC	
					CAAGAGCGGAGATGCTTGGTATACCAAGCT	GGCCTT	TGGT	
					AGCAGAGAGAGGGTTGAAGAGGATGACTA	CTAGTA	GGTA	
					CTGGGCTCAGAGTCATCTCACCCTAAAAGG	GGGCTG	GGGA	
					CGGTGGGGCATCGGTTGAACACCTACCCAT	CTTCAA	CAAG	
					ACCGGGATGGGAGGTGGTAGGCCGAAAAA	CCCCTC	GAGG	
					GAACAGGAAGATGGTGGAGTAAGTTGAGA	TGATGG	TAAG	
					GCGGTTGCTCGGGAAGTTATGTTGTGATAA	GAGTGT	GGCA	
					CTCCATTAAGGCCGGTGGGCATGGTGCGGA	ACCGGG	GTGC	
					TAATGGAAACTATAAAAACAATAAAAAGA	AACCTC	CAAC	
					AAGACCAAAAAAATGTTCTGTTATGATGCC	GACTTG	CCCT	
					TTACACATGTCTGGGAGACCCCATAAGGGT	TAAGCA	ACTT	
					CTCCCCTTATACTTCACTGGGAAACCCCATA	CAGGTT	GATC	

					AGGGTATCCCCCTATATTTACTGGGAGACC	AGTATG	GGGA	
					CCATAAGGGTCTCCCCCTATAGATGTAGAG	GGAGCA	CCAT	
					CGTAAGGGGTCTCCAAAGTACCGGCCGATA	GAAGGG	CCAG	
					TGGCCTTATGGCAAACTCTGGTGGTAGGGA	GGAGCC	GGAA	
					CAAGGAGGTAAGGGCAGTGCCAACCCCTAC	GTAATG	TGCC	
					TTGATCGGGACCATCCAGGGAATGCCATCC	GGCTTC	ATCC	
					TCCCGCGAAGGTGATGTGGTGAGGTAAGGG	TCTTTC	TCCC	
					GGGAGCCCGTCTTCGAGTTTCCCCAACCCCT	ACCTGC	GCGA	
					ACCCACAGGTGAGAGGAGGAGAAGAGGAA	TTACAT	AGGT	
					TCTGTCCCCAACGGGAGGAGGGTGAGGTGT	AATACC	GATG	
					AAGGGGGAGACCTTCTAGTAGGGTCTTCTC	TGTGGT	TGGT	
					AGTCGCCTGACGTCCTGACTGTGGGGTGGA	GCATGT	GAGG	
					TCAGTACCCTACAGGTGAGACCGGTGAGGT	ATCTAG	TAAG	
					AAGGGTGTGGCCCTCTTGAGGGCTGCGCCA	GTCTTG	GGGG	
					ACCCCTACTCGAGGTAACCTGAGGGAGTGG	GCGGGA	GAGC	
					TGGAATGGCGGCATGTTAGTGCTGGGACTT	GAGTAC	CCGT	
					GATTGCGAGGGTTTAATGAGAGTGGCCTGC	TGAGAG	CTTC	
					TGAGAGCAACACTTGTGGTGCTTAAAGCGG	ACAAGG	GAGT	
					GGCGGCCCATGACCACCGTGAGATAGGACA	TTGAGA	TTCCC	
					CTGCACAGTGCAGCCATGAGGTTCCTGGAG	CCCCAA	CAAC	
					GATGATGCGATGAGGTGGGGGCCTCATCAG	GATTGG	CCCT	
					CCCCTCCTGGCAGGGCGTCGGCCAGGGAAA	GTCTCC	ACCC	
					CTAAATGTCTCTAGCATGTCAGTGCAGTGA	CTAGCC	ACAG	

					GGTAAGGGGAGAGCACTCTAGTAGGGCTCT	TCTATA	GTGA	
					TCCAACCCCTACCTGTAGGTCACCTGGTCCA	GCTGCG	GAGG	
					GGTGTCGATGATGTGAAAACAAGAGCTACT	ACTCTT	AGGA	
					TTGGTACCGGTCTGTTGCAAAAAGGGTTCT	AGCGGG	GAAG	
					GCAGAGGACGACGGCTATCCCTATCGGGAG	GATATG	AGGA	
					GGAATAGTCGGTCCCAGGTAGTGGAAAATG	GAGTAA	ATCT	
					GGGCTTTCCACTGAGCATGAAAATGTGGTA	CATGTA	GTCC	
					GAGGTTGCGTCCAACCCAATGATTTGCAGC	CCAAGG	CCAA	
					AGAGCTCTTGGACACGAAGTCTGTATAGTC	GAGTGA	CGGG	
					CCATGCAGGCAGCCAACCAGAGAATGGTGG	ATAATA	AGGA	
					CAAGACCCCAGCTCCGTATGGGAGGGGAGG	AAGGAA	GGGT	
					GCCAAGATATACGGAACGGCTGCTAAAGCG	TTGACG	GAGG	
					TTCTGCCGGTGTCAGTCTAATCACAGACAG	GGGTAC	TGTA	
					CTGTGACGAAACAAAGTATGGGTTCCGACA	AAGTGA	AGGG	
					TGCTTGGTCAGCTCTTAGCCGCAAGGCTTAA	CTGTTG	GGAG	
					ATCGAACGCAGCCCGCCGAGAGTGAACATT	GCCCGA	ACCT	
					AAACGGGGATGGAATGTGTCTAGCGGTTAC	ATCCTA	TCTA	
					GTACTACCAGGGCTCAGGTTCGCCTGAGCC	GCCACT	GTAG	
					GAGGCTCTACACGTCATGGTGGGAGTTCTC	TGATGA	GGTC	
					CCCACGCTCGTGAGGGCATGTAGTGGGATG	CCAGGG	TTCTC	
					GCATGTGGCGGACCATCAGCTGGCACTACC	ATATAT	AGTC	
					AGGCCTCGGGCTTGCCCGAGTGCGGGACCT	TACAGA	GCCT	
					CACACATTGTAGGTGTGCTTGTCCCCCCTAC	CTGAGA	GACG	

					GTTCGAAGACTTGAGGCGGAGAATACTCAT AGGCCCCACGGCAAAGGGACACAACACGG AGGCTTGTGTCCGACGAGCCGTGGACTCCT ATAGACAGCCCGGGATATCACTGGGCACGC TCATACTGAAGAAATTCGATGAACCGGGCC TACCGGAGCAAATGCACTCTAATCGCCTTT GTGGGCGACTGTGGCCCCCTCATGCGAGTG AGGAATATCATAAACTGCAATGGTTCAAAA AGTGATTCCTATGGCTCGTCGGGGAGGGCT GACTGGGGCAAGCAAATGATTGAAAGGGG AAGAACCTTTTTCAACTGTTTCTTGCCAAGC CCGGTTGATGGTGGCGCTAGTAATTGCGAC GGGAAAATGCGGTTTAAGTCTCCGAAGTAG TGCGTAGCACCGGATGTCGACCGGGTGTAA AAGCCCTTCGTAAAGTCCCTGGGGAGGTCA GTCCTGGGGCTACTGATGCGCAGTATGTAA TTCGCAGAATAGGGCCATCGATACCGCCTG CGTGACTCGACTGGGTTTCCACTTGAGGAT ATCCGACCGTAGCGTGCACCCTCTTGTAGTT GCGCCGGAAACGGCTGTGTTCCCTCACGTA TGTGAGGAAACTCAACAATGTGAGTGGGTA AACGGCGGGACGAACTATGGCTCTCGT (SEQ ID NO: 1095)	GCGAAT CTAGAG ACGTGA GAATAA AGTGAG AATTGG TTGAGC GAATAC AGAGGA AG (SEQ ID NO: 1218)	TCCT GACT GTGG GGTG GATC AGTA CCCT ACAG GTGA GACC GGTG AGGT AAGG GTGT GGCC CTCTT GAGG GCTG CGCC AACC CCTA CTCG AGGT	

							AACC TGAG GGAG TGGT GGAA TGGC GGCA TGTT AGTG CTGG GACT TGAT TGCG AGGG TTTA ATGA GAGT GGCC TGCT GAGA GCAA CACT TGTG	

							GTGC TTAA AGCG GGGC GGCC CATG ACCA CCGT GAGA TAGG ACAC TGCA CAGT GCAG CCAT GAGG TTCCT GGAG GATG ATGC GATG AGGT GGGG	

							GCCT CATC AGCC CCTC CTGG CAGG GCGT CGGC CAGG GAAA CTAA ATGT CTCT AGCA TGTC AGTG CAGT GAGG TAAG GGGA GAGC ACTC TAGT	

							AGGG CTCTT CCAA CCCC TACC TGTA GGTC ACCT GGTC CAGG TGTC GATG ATGT GAAA ACAA GAGC TACTT TGGT ACCG GTCT GTTG CAAA AAGG	

							GTTCT GCAG AGGA CGAC GGCT ATCC CTAT CGGG AGGG AATA GTCG GTCC CAGG TAGT GGAA AATG GGGC TTTCC ACTG AGCA TGAA AATG TGGT	

							AGAG GTTG CGTC CAAC CCAA TGAT TTGC AGCA GAGC TCTTG GACA CGAA GTCT GTAT AGTC CCAT GCAG GCAG CCAA CCAG AGAA TGGT GGCA	

							AGAC CCCA GCTC CGTA TGGG AGGG GAGG GCCA AGAT ATAC GGAA CGGC TGCT AAAG CGTT CTGC CGGT GTCA GTCT AATC ACAG ACAG CTGT	

							GACG AAAC AAAG TATG GGTT CCGA CATG CTTG GTCA GCTC TTAG CCGC AAGG CTTA AATC GAAC GCAG CCCG CCGA GAGT GAAC ATTA AACG	

							GGGA TGGA ATGT GTCT AGCG GTTA CGTA CTAC CAGG GCTC AGGT TCGC CTGA GCCG AGGC TCTA CACG TCAT GGTG GGAG TTCTC CCCA CGCT	

							CGTG AGGG CATG TAGT GGGA TGGC ATGT GGCG GACC ATCA GCTG GCAC TACC AGGC CTCG GGCT TGCC CGAG TGCG GGAC CTCA CACA TTGT	

							AGGT GTGC TTGTC CCCC CTAC GTTC GAAG ACTT GAGG CGGA GAAT ACTC ATAG GCCC CACG GCAA AGGG ACAC AACA CGGA GGCT TGTG TCCG	

							ACGA GCCG TGGA CTCCT ATAG ACAG CCCG GGAT ATCA CTGG GCAC GCTC ATAC TGAA GAAA TTCG ATGA ACCG GGCC TACC GGAG CAAA TGCA	

							CTCT AATC GCCT TTGT GGGC GACT GTGG CCCC CTCA TGCG AGTG AGGA ATAT CATA AACT GCAA TGGT TCAA AAAG TGAT TCCT ATGG CTCG	

							TCGG GGAG GGCT GACT GGGG CAAG CAAA TGAT TGAA AGGG GAAG AACC TTTTT CAAC TGTTT CTTG CCAA GCCC GGTT GATG GTGG CGCT AGTA	

							ATTG CGAC GGGA AAAT GCGG TTTA AGTC TCCG AAGT AGTG CGTA GCAC CGGA TGTC GACC GGGT GTAA AAGC CCTTC GTAA AGTC CCTG GGGA	

							GGTC AGTC CTGG GGCT ACTG ATGC GCAG TATG TAAT TCGC AGAA TAGG GCCA TCGA TACC GCCT GCGT GACT CGAC TGGG TTTCC ACTT GAGG	

							ATAT CCGA CCGT AGCG TGCA CCCT CTTGT AGTT GCGC CGGA AACG GCTG TGTTC CCTC ACGT ATGT GAGG AAAC TCAA CAAT GTGA GTGG GTAA	

							ACGG CGGG ACGA ACTA TGGC TCTC GT (SEQ ID NO: 1341)	
R2	R2-1_TCas	.		Tribolium castaneum	AGTCATAGAGCCAGAACCTCCTCGTGGTCC CGCTGGGCACAGGGATTAATTTTTCTGTGGC AAATTTGACTGGCTTCAGAGAGCGTTTTTCG AAGTGGWCTGTGTGACTGCGTTCCCCCCTT AGTTGCTATWTCCGCTKMGATTAACATCTC ACCTCGACGTWTAAGATCATTATGTCGAGG CGCCCGGGGAAGTCTAATGAGCCCCCGGTT CGTAGTAGGGCTATGGGTTTAACCACTTTAT CTGGCACCAAGACATCAAATTCTGGTGCCC AGGGGCCCTCTACCAGTGCCCCCATGCAGA ACATGGCTGGGGGTTTTGTTTGTGACTGCGG CCGGAGCTATGCCTTGAAGACATCTCTGGC CCGACACAAGAAGGAGTGCGGAAAGAACA	AGTCAT AGAGCC AGAACC TCCTCG TGGTCC CGCTGG GCACAG GGATTA ATTTTT CTGTGG CAAATT TGACTG GCTTCA	TGAA AAGA GTGG CAGT TGTG GAGA CTTCC TTCG GACG TCGG GTCG GATT TTCG	MSRRPGKSNEPP VRSRAMGLTTL SGTKTSNSGAQ GPSTSAPMQNM AGGFVCDCGRS YALKTSLARHK KECGKNNAECR WCGTRFNTLAG TRQHERKAHFV QYQSDLAKALP QPESELMEKIAI VEARSXNGIFYK EMMASTGLTHQ

					ATGCGGAGTGTCGATGGTGCGGGACAAGGT	GAGAGC	GACG	QVRSRREKPEY
					TTAACACCTTGGCTGGCACTCGTCAGCATG	GTTTTT	CCAG	KGFLERARRSLA
					AGCGAAAGGCTCACTTTGTCCAGTATCAGT	CGAAGT	GGTA	QTNIRAGSISPAS
					CTGATCTGGCAAAGGCGCTCCCCCAACCTG	GGWCTG	CCTC	TXAGSLESASPK
					AGTCGGAGCTTATGGAGAAAATCGCCATTG	TGTGAC	CACC	AGCSSSASPGPT
					TGGAGGCCAGGTCCAKTAACGGCATCTTCT	TGCGTT	GCTG	TRSRAPTKGVPX
					ATAAGGAGATGATGGCCTCCACGGGTCTGA	CCCCCC	GGTT	RSSNSARIVVEA
					CCCACCAACAGGTCAGATCAAGGAGAGAG	TTAGTT	CACA	QVHTRAPPNTG
					AAACCCGAATATAAGGGCTTCCTGGAGCGA	GCTATW	AACT	ETEVALRESRRT
					GCACGCAGGAGCCTGGCCCAGACCAATATC	TCCGCT	AGGC	VPRLGXNPSRPC
					AGGGCTGGCTCCATCTCACCAGCGAGTACG	KMGATT	GAAC	GISPLMAIAIDED
					AWCGCAGGCTCTTTGGAGTCTGCGTCTCCG	AACATC	ATCT	SVLGGLRVQAG
					AAAGCGGGGTGTAGTAGCAGTGCCTCGCCC	TCACCT	GCCG	PSPTAVHSVEAF
					GGACCGACCACCAGGTCCAGGGCACCTACC	CGACGT	ATAC	PGTSSMTPMET
					AAGGGTGTGCCCCWACGAAGCTCTAATTCG	WTAAG	CCTCT	DRVHNKSGIDPI
					GCGAGAATTGTGGTGGAGGCCCAAGTCCAT	ATCATT	TTAG	LEHNGTRQVRR
					ACCCGGGCGCCACCCAATACCGGGGAAACG	(SEQ ID	GTCA	EESSTREDPVEQ
					GAAGTGGCTCTTCGTGAGTCACGCCGAACC	NO: 1219)	TAGG	WSPNYPKTPVT
					GTGCCCCGGCTCGGGCMTAATCCATCTAGG		ACCA	MPNITTTADAX
					CCCTGTGGTATCTCTCCTTTGATGGCGATTG		CATG	XTSYNRTPQTLP
					CCATAGATGAAGACTCTGTGTTGGGCGGGC		TCTCT	GNRRRRSRSLPP
					TACGAGTGCAGGCCGGGCCAAGTCCCACCG		GCAC	VQRKSASDXLE
					CGGTGCACTCCGTCGAAGCTTTCCCCGGAA		GAGA	SVDSLGPWAVF

					CGTCCTCGATGACCCCGATGGAGACCGATC		TTAA	LQDQVDAGSLS
					GGGTACATAATAAGTCCGGTATAGACCCCA		CCCA	GNDSLADLVRV
					TTCTCGAACATAATGGGACACGGCAGGTTC		(SEQ	ALTKSDRGVLN
					GAAGGGAGGAGTCTTCCACACGGGAAGACC		ID NO:	DAVNRYLAQRA
					CCGTCGAGCAATGGTCTCCCAACTATCCTA		1342)	ESLRIRKRGSKG
					AAACGCCTGTCACTATGCCCAACATTACAA			KRKSKTGRHYG
					CTACTGCTGATGCTWCGTKCACCTCGTACA			QTTSGSGQRAA
					ACCGAACACCTCAGACACTCCCTGGAAACC			LFKKHQDLFLK
					GGAGGCGACGCAGTCGCAGTCTACCACCAG			NRRGLAETILSG
					TGCAACGGAAATCCGCTTCCGACGAWCTCG			KEDFGPRPEPPV
					AGTCCGTTGATTCGCTGGGGCCCTGGGCGG			TSVEEFYGGIFE
					TCTTCTTGCAAGATCAAGTAGACGCCGGGT			SPSPPDNEPFEV
					CACTAAGCGGCAACGACTCGCTGGCGGATC			RATGVEDPPPLT
					TGGTGAGGGTGGCTCTCACTAAGTCCGATC			SPWTKRTLRSPC
					GCGGTGTCCTTAATGATGCCGTGAATCGGT			YWRRRPPPTYIT
					ATCTGGCCCAAAGGGCGGAGAGCTTGAGGA			MDEIKAARAGW
					TACGTAAGCGAGGCAGTAAGGGCAAAAGA			QISAPGSDQIPV
					AAGAGTAAGACTGGTCGCCACTATGGCCAA			AAVKTMSELEL
					ACAACATCTGGCAGTGGTCAAAGAGCTGCG			AILFNIILFRNVQ
					CTCTTTAAAAAGCACCAGGACCTCTTTCTGA			PSAWGVLRTTL
					AAAACAGGCGAGGATTGGCGGAGACTATCT			VPKDGDLRNPA
					TGTCGGGGAAGGAGGATTTTGGTCCTAGAC			NWRPITISSALQ
					CTGAACCTCCCGTAACAAGTGTGGAAGAGT			RLLHRVLAARL

					TCTACGGCGGCATTTTCGAGTCGCCTTCTCC			SKLVSLSSSQRG
					TCCCGATAACGAACCCTTCGAAGTCCGTGC			FTEIDGTLANAL
					TACTGGCGTCGAAGACCCCCCCCCCACTTAC			ILHEYLQYRRQT
					ATCACCATGGACGAAATAAAGGCGGCAAG			GRTYXVVSLDV
					AGCCGGTTGGCAGATTTCAGCGCCTGGATC			RKAFDTVSHCS
					TGACCAAATACCGGTCGCGGCCGTCAAAAC			VSRALGRFGIPS
					TATGTCCGAGCTGGAACTGGCGATACTGTT			VIREYILATFGA
					CAATATTATATTGTTCAGAAACGTCCAGCCT			QTTIKCGSVTTR
					TCAGCCTGGGGGGGTACTTCGTACGACGCTG			PIRMLRGVRQG
					GTTCCCAAAGATGGGGATCTTAGAAATCCG			DPLSPVLFNLVM
					GCCAACTGGCGGCCCATTACWATCAGTTCT			DELLEKVNEKY
					GCCCTGCAGCGTCTTTTGCATAGAGTATTGG			EGGSLQSGERC
					CGGCCCGTCTTTCGAAGTTGGTCTCGTTGAG			AIMAFADDLILI
					TTCGTCTCAGCGGGGCTTTACCGAGATTGAC			ADRDQDVPAMF
					GGTACTCTAGCGAATGCGCTGATACTGCAT			DDVSTFLERRG
					GAATATTTGCAGTATCGCAGGCAGACTGGC			MSVNPAKCRAL
					CGCACCTACCASGTCGTGTCTCTGGACGTCA			IAGAVSGRSVVR
					GAAAGGCGTTTGACACGGTTTCCCACTGCT			TGSSYKIHNTPIP
					CTGTTAGTAGAGCCCTGGGTCGTTTTGGTAT			NVDALDAFKYL
					CCCCTCTGTTATCAGGGAATACATCCTGGCC			GLEFGHKGVER
					ACATTTGGCGCACAGACGACCATAAAATGT			PTIHNLSVWLNN
					GGGAGCGTTACGACAAGACCGATAAGGATG			LRRAPLKPDQK
					CTCCGTGGAGTGAGACAGGGTGACCCTCTT			CLFIRQYVIPRLL

					AGCCCGGTCCTATTCAATCTTGTGATGGATG			YGMQNPQVTSR
					AGCTTCTGGAAAAGGTCAATGAGAAATACG			VLREADRLIRRH
					AGGGCGGCTCACTGCAAAGCGGCGAACGCT			LKTYYHLNVHT
					GTGCCATCATGGCGTTTGCTGACGACCTCAT			PDSLIHASVSDG
					CCTGATTGCTGACCGAGATCAGGATGTACC			GLGIMELRKAIP
					CGCAATGTTCGATGATGTTTCGACATTTCTC			RIFLGRLVKLLN
					GAGCGGCGCGGGATGTCTGTCAACCCTGCC			KNKDSVLSSVL
					AAGTGCAGAGCGTTGATTGCGGGCGCGGTG			QSNRVRTLMGK
					AGCGGGAGAAGTGTTGTGAGGACAGGCAG			LSTMAGEVPEST
					CTCTTATAAGATCCATAACACGCCGATTCCA			FWRNQIASGPLS
					AATGTCGACGCCCTAGATGCATTTAAGTAT			KGLEQAAEDSA
					CTGGGGCTAGAATTTGGCCACAAAGGCGTA			SRLWISEKPSGW
					GAGCGCCCCACAATCCACAACCTCTCCGTC			SGRDHVRAVQL
					TGGCTTAACAACCTCAGACGAGCCCCTCTC			RTGNLPTKAIPS
					AAACCGGACCAAAAATGCTTATTCATCCGG			VPVGQRRCRHG
					CAATATGTCATCCCTCGATTGCTCTACGGCA			CACDESISHVLQ
					TGCAAAATCCTCAAGTTACCTCTAGGGTACT			MCPLTHADRIRR
					TCGAGAAGCTGACCGTCTGATAAGGAGACA			HDEVVKKVARH
					CTTAAAGACATACTACCATCTTAATGTCCAT			CTSRGWTVEVE
					ACTCCAGATTCTCTTATTCACGCGAGTGTGA			PHIRSRCGRLFK
					GTGACGGAGGTCTGGGTATAATGGAATTGC			PDLAVHQPGGAI
					GTAAAGCAATTCCAAGAATTTTCCTAGGAC			VIADVQVSWDS
					GGCTAGTTAAGCTTCTCAACAAAAACAAAG			ESLTVPYERKRA

					ATTCGGTCTTAAGTTCGGTCCTGCAGTCTAA CCGTGTGAGAACACTAATGGGCAAATTATC AACCATGGCGGGTGAAGTACCGGAGTCGAC GTTTTGGAGAAATCAGATTGCCTCGGGTCC CCTTTCGAAGGGTCTCGAACAGGCGGCCGA GGACTCCGCAAGTCGACTCTGGATCAGCGA GAAACCTTCCGGTTGGTCGGGTAGAGATCA TGTCCGTGCAGTCCAACTTCGCACGGGCAA CCTCCCGACCAAAGCCATTCCATCGGTGCC GGTTGGTCAGAGACGCTGCAGACACGGATG TGCCTGCGATGAATCGATATCGCATGTGCT GCAGATGTGCCCCCTGACGCACGCTGATCG GATTCGCAGGCACGATGAAGTGGTTAAGAA AGTAGCCCGCCATTGCACCTCACGCGGTTG GACGGTGGAAGTCGAGCCACACATTCGCTC TAGGTGTGGTAGGCTTTTTAAGCCTGACCTG GCAGTCCATCAGCCAGGCGGCGCCATTGTG ATAGCGGATGTCCAGGTATCATGGGATTCC GAAAGCTTGACGGTACCTTACGAACGCAAG CGAGCTAAGTACGACGTGCCCCAATTCCAT CAGGCGGCACAACATGCATGGCCGGGTAAG GCACTTACGTTTGCCCCTGTAATAGTGGGCG CCCGGGGCATATGGCCCCGGATAAACAACG			KYDVPQFHQAA QHAWPGKALTF APVIVGARGIWP RINNDRSAALQI PPVVRRACVNS VVKWGSSIHAT FMRSVWANRLN PRPLRA (SEQ ID NO: 1464)

					ACCGCTCTGCGGCGTTGCAGATTCCTCCGGT CGTGCGACGAGCTTGCGTAAATTCGGTTGT CAAGTGGGGGTCGTCGATCCATGCTACCTTT ATGAGGTCAGTATGGGCAAATCGACTCAAC CCCCGACCCTTAAGGGCTTGAAAAGAGTGG CAGTTGTGGAGACTTCCTTCGGACGTCGGG TCGGATTTTCGGACGCCAGGGTACCTCCAC CGCTGGGTTCACAAACTAGGCGAACATCTG CCGATACCCTCTTTAGGTCATAGGACCACAT GTCTCTGCACGAGATTAACCCA (SEQ ID NO: 1096)			
R2	R2-1_TGut	.		Tinamus guttatus	CTGGGGACCGTGGTTACAACCCGGGCTTAG CTGCAGAGACAGTACCTCCCCGTGGTTCCC GCCGGACCCCGTAACATCGGGTGACTGAAT CTGTCTCTGCCCCGGGAGTAGTTCCTCCTTG CCCTATTGACCAGCGGTCGCCGGCTGCTCA ATAGTATTCTAGGCGTGAAATATAGCGATA GTCCTAGTGGTTGTCTTACTGGGCCATAGCC CCTTGCTTCAGGGGTCATTCGCGAAGTCTCT CAGGAGAACTGGGGGTGGTGTTCTTCTGGG TATAGCTAAACCCCCTAGACTGTGTCCGATC CATGGGGTCCTGGATCGTGAATTTCGTTTCG GTGGCGACTCAGACGGGAGAATTCCCTGTG	CTGGGG ACCGTG GTTACA ACCCGG GCTTAG CTGCAG AGACAG TACCTC CCCGTG GTTCCC GCCGGA CCCCGT	TAGG GGGC TTGG CATTT CTCA TTGC CTGC TCCT GAAA GGAT ATGG GTCC	MGSWIVNFVSV ATQTGEFPVDT ARRAPVPVTSYP ESECHXPLPLTF CNSDVTIWGGV RPEPVDCLGDLP EXYDALPGVAG PREXVGGSPPGE GVRSPGIASPSG TAVQHDFGSPIL VPGAEAAEVSTP VVKVPQDHPAC

					GATACGGCCAGGAGGGCACCTGTGCCGGTA	AACATC	TGCG	PCCGTRVVKVT
					ACATCATACCCTGAGTCGGAATGCCACNTA	GGGTGA	TCGC	ALSEHLRRAHG
					CCGTTGCCCCTGACATTTTGTAACTCGGATG	CTGAAT	GTGG	RKRVLFQCSRC
					TGACTATTTGGGGAGGGGTTCGCCCTGAAC	CTGTCT	TAGG	GRMNEKHHSIA
					CGGTGGACTGCTTGGGWGATCTTCCRGAGG	CTGCCC	CAGA	CHFPKCRGPPVE
					YGTATGATGCACTCCCAGGGGTGGCTGGGC	CGGGAG	CCCA	EGPLGAPEWCC
					CTCGGGAAYCGGTGGGTGGGAGCCCGCCGG	TAGTTC	TTCGT	EECGQKFNTKS
					GRGAAGGGGTCAGGTCGCCAGGGATTGCGT	CTCCTT	CCGA	GLSQHKRSVHP
					CRCCCTCTGGTACTGCGGTCCAACATGATTT	GCCCTA	GTAG	LTRNVERIEAAR
					TGGGAGTCCCATCCTCGTACCGGGAGCCGA	TTGACC	GGGG	PKGKGKRGAHK
					AGCCGCCGAGGTNTCTACCCCGGTAGTGAA	AGCGGT	CTTG	GCWTEAEVAQL
					GGTTCCNCAAGACCATCCAGCATGTCCNTG	CGCCGG	GCAG	IELEGRFKNQRFI
					CTGTGGTACGAGGGTGGTGAAAGTAACGGC	CTGCTC	TNTC	NKLIAEHLPSKS
					GTTGTCAGAACATCTTAGGAGGGCCCACGG	AATAGT	CATT	AKQISDKRRQL
					TCGGAAACGGGTCCTATTTCAGTGCTCCCG	ATTCTA	GCCT	AAATKTSSPEKR
					ATGCGGGAGGATGAATGAGAAACATCATAG	GGCGTG	GTGC	VTSSTSGESSPE
					CATCGCGTGCCATTTCCCGAAGTGCCGGGG	AAATAT	CCGA	VEKVEGIKREYR
					GCCCCCAGTTGAGGAGGGTCCCCTGGGTGC	AGCGAT	AAGG	RRVGEWLCAGS
					ACCCGAGTGGTGCTGTGAGGAGTGCGGGCA	AGTCCT	ACGT	LXDQTSFQKILE
					GAAATTTAACACCAAAAGCGGCCTGTCTCA	AGTGGT	GGGT	DVESGSEIVTGP
					GCACAAAAGATCTGTGCACCCACTTACGAG	TGTCTT	CATC	LEELASFARGKL
					GAATGTGGAACGGATAGAGGCAGCTCGTCC	ACTGGG	TGGT	AAARVRHHRKH
					GAAAGGAAAAGGGAAGCGTGGTGCCCACA	CCATAG	CTGT	PAEAVPAREEQ

					AAGGCTGTTGGACCGAGGCGGAGGTGGCCC	CCCCTT	CTGC	RWMKRRVGRR
					AGCTGATTGAACTGGAGGGGAGATTCAAGA	GCTTCA	CTAC	GLYLRFQRLFAL
					ACCAGCGATTTATCAACAAGCTGATCGCGG	GGGGTC	ACCT	DRRKLAGIILDD
					AGCATTTGCCATCGAAATCGGCGAAGCAGA	ATTCGC	CTCT	VESIKCPLPMEE
					TCAGCGATAAGAGAAGGCAGCTGGCGGCA	GAAGTC	AGAC	VADVFRRRWEE
					GCGACCAAGACATCGTCGCCCGAGAAGAGG	TCTCAG	TTGT	VAPFTGSGSFRS
					GTAACGTCATCAACGAGTGGGGAGTCCTCC	GAGAAC	AACA	LGKADNGAFKP
					CCTGAAGTGGAAAAGGTGGAGGGTATCAAG	TGGGGG	TCTA	MISAKEVMKNV
					AGAGAATATAGAAGGCGTGTTGGAGAGTGG	TGGTGT	GTCT	XEMSRRSAXGP
					TTGTGCGCTGGGTCCCTGMAGGACCAGACN	TCTTCT	GTCA	DGLSLRDLMKI
					TCGTTCCAGAAGATCTTGGAAGATGTGGAG	GGGTAT	ACAA	DPQGSRMAELF
					AGCGGCTCCGAGATTGTCACCGGTCCGCTG	AGCTAA	GATC	NLWLLAGRVPD
					GAGGAACTGGCCTCCTTTGCGAGGGGGAAG	ACCCCC	AAAA	QVKAGRTVLIPK
					CTCGCGGCAGCTAGAGTGCGACATCATCGT	TAGACT	TTCTT	SADPGKIGNIDN
					AAGCACCCAGCTGAGGCCGTGCCTGCGCGA	GTGTCC	CACA	WRPITIGSVXLR
					GAGGAGCAGAGGTGGATGAAGCGCAGGGT	GATCC	CAGA	MFSRILSARLRR
					GGGTCGTCGGGGCTTGTACCTCAGGTTCCA	(SEQ ID	CGAC	ACPINRRQRGFI
					GCGGCTTTTTGCNTTGGATCGCAGRAAGCTT	NO: 1220)	CGAG	AAPGCSENLKLL
					GCTGGGATCATCCTCGACGATGTCGAGTCC		CTTG	QALIKSAKRDH
					ATCAAGTGCCCCCTTCCGATGGAAGAAGTC		CTCA	RTLGVVFVDLA
					GCTGACGTCTTCAGGAGAAGGTGGGAGGAG		GTCTT	KAFDSVNHQHIF
					GTCGCCCCCTTTACCGGCTCGGGCTCATTCC		CCTG	QVLVQKGVDGH
					GAAGTTTGGGGAAGGCTGACAACGGTGCCT		TACC	IIDILRDLYTNAG

					TCAAGCCCATGATCTCCGCYAAGGAGGTCA		CGCA	TYLESGSQRSGF
					TGAAGAACGTCRCGGAGATGTCTCGACGCT		GAAT	IKILRGVKQGDP
					CCGCGCYGGGWCCCGAYGGCCTCTCCCTGC		TTTGC	LSPILFNLALDPL
					GGGATCTGATGAAGATTGATCCCCAGGGCA		TCTTG	LCRLEDRGLGY
					GCCGCATGGCTGAATTGTTCAACCTGTGGCT		CTCTC	KYGDQQIXSLAF
					GTTGGCAGGACGGGTCCCGGACCAAGTGAA		CTTTG	ADDLALLSDSW
					GGCGGGCCGAACGGTCCTGATCCCNAAGTC		GCTG	EGMQQSIRVVE
					GGCCGATCCCGGGAAGATCGGGAACATTGA		TGTC	EFCQRTGLRVQ
					CAACTGGCGGCCCATCACCATCGGGTCCGT		CTGG	APKCHGFLIRPT
					TATNCTCAGAATGTTCTCTCGGATTTTGAGC		ACGT	KESYTINDCDP
					GCTAGACTGCGGCGAGCATGCCCCATTAAT		GGGA	WTIADMQLDMI
					AGAAGGCAAAGGGGGTTTATAGCAGCCCCT		CTATT	DPGSSEKYLGLG
					GGCTGCTCGGAAAATCTGAAGCTTCTCCAG		CCAT	IDPWIGLSRPELS
					GCGCTCATCAAGAGCGCGAAACGAGATCAT		CTCG	EVLTRWVKNIG
					AGGACCCTTGGAGTCGTGTTTGTCGATTTGG		TCCC	GAPLKPLQKVDI
					CTAAGGCCTTCGACTCCGTGAACCACCAGC		AAAT	LRSYALPRLLFI
					ATATCTTCCAAGTCCTGGTCCAGAAGGGTG		GCCG	ADHAGLSATCL
					TCGATGGGCATATTATCGACATCCTAAGAG		CGTC	HSLDLSIRSAVK
					ACCTGTATACCAACGCTGGAACGTATCTGG		CAAT	GWLHLPPSTCD
					AGTCAGGTTCCCAGCGATCGGGATTTATTA		TATA	AIIYVSYKDGGL
					AGATCCTCAGGGGAGTGAAACARGGGGACC		CCGG	GLPRLASLIPNV
					CACTCTCTCCCATCCTGTTCAATCTTGCATT		ATTT	QARRLVRIAQSE
					AGATCCTTTGCTGTGCCGCCTGGAAGATCG		GACA	DDVIRSVVLQEG

					GGGCCTCGGTTATAAGTATGGAGACCAACA		AAGC	IQEEIRKVWISA
					AATAWCATCGTTGGCATTTGCGGATGATCT		GGAC	GGRPEKVPSVT
					CGCCCTGCTCAGCGACTCTTGGGAGGGCAT		GGCC	GEFPVMEAQAA
					GCAGCAGAGTATTCGGGTGGTAGAGGAATT		CGCT	DEALSEWERRA
					TTGTCAACGGACCGGGCTGCGGGTTCAAGC		TTAT	PRTIYPIPCKWR
					GCCGAAATGCCACGGGTTTTTGATCAGGCC		AAGC	KREMENWTNLK
					AACTAAAGAGTCATATACCATCAATGACTG		CGGA	SQGHGIRNFEND
					TGACCCGTGGACGATTGCAGATATGCAATT		AAAG	RISNDWLLHYG
					GGATATGATCGATCCGGGCAGTTCCGAGAA		GTGC	RIPHRKLITAIQL
					GTATCTTGGCCTAGGGATAGACCCATGGAT		CTTGT	RANVYPTREFL
					TGGTCTATCGAGACCGGAACTGTCCGAGGT		AAAA	ARGLGEGAPRG
					GCTGACCCGCTGGGTGAAGAACATCGGGGG		TTGC	CRHCPAEWESC
					CGCCCCTTTGAAGCCACTCCAGAAGGTGGA		AAGG	SHIIGYCPAVQE
					CATCTTGAGGAGCTACGCCCTNCCAAGGCT		TTCAT	ARIKRHNDICGV
					GCTGTTCATTGCGGATCACGCAGGCCTGAG		TAAA	LAEEARKLGWV
					CGCCACCTGTTTGCATTCCCTGGACCTTTCG		TAG	IFIEPHLRDNTNE
					ATAAGATCTGCCGTCAAGGGCTGGTTACAT		(SEQ	LFKPDLVLVKGS
					CTACCGCCTAGTACGTGTGACGCTATTATTT		ID NO:	CAKVVDVTIRY
					ACGTCAGCTACAAGGACGGCGGGCTGGGTC		1343)	ESGLTTLSDAAA
					TTCCCCGTCTGGCGAGCCTAATTCCAAATGT			EKARKYQHLAG
					ACAGGCTCGCAGGTTGGTGCGGATCGCCCA			EVRALTSATTV
					ATCGGAGGATGATGTCATCAGGAGTGTGGT			DFLGFPIGARGK
					ACTCCAGGAGGGTATCCAGGAGGAGATCCG			WYVGNNGLLSD

					GAAGGTCTGGATCTCGGCTGGGGGGCGACC GGAGAAGGTTCCATCTGTGACGGGGGAGTT CCCAGTGATGGAGGCTCAGGCGGCTGACGA GGCCCTATCCGAGTGGGAGAGGCGAGCTCC ACGAACCATCTATCCCATTCCCTGTAAGTGG AGGAAGAGAGAAATGGAGAATTGGACCAA TCTAAAATCGCAAGGCCACGGGATTCGGAA TTTTGAAAATGACCGAATCAGTAATGATTG GCTCCTGCATTATGGCCGCATTCCCCACCGC AAACTAATAACAGCTATCCAGTTGCGGGCC AATGTCTATCCCACWCGGGAGTTTTTGGCC CGCGGCCTGGGCGAGGGCGCACCCAGGGG ATGTAGGCACTGTCCCGCGGAGTGGGAATC TTGTTCCCACATAATTGGCTACTGCCCGGCT GTCCAGGAGGCCAGGATCAAAAGGCATAAT GACATCTGTGGTGTGCTGGCTGAAGAGGCN AGAAAGCTGGGATGGGTGATATTTATAGAG CCCCATCTCAGAGATAACACCAATGAGCTC TTCAAGCCAGATTTGGTTTTGGTGAAGGGA TCCTGTGCGAAGGTAGTGGATGTAACCATC CGCTACGAGAGTGGGTTAACCACCTTGAGT GACGCCGCGGCAGAAAAGGCTAGGAAGTA TCAACATCTGGCAGGGGAGGTGCGGGCCCT			LGFSTSRVVRIA RALSKKALLSSV DIIHIFASRARQA QTSE (SEQ ID NO: 1465)

					AACATCGGCCACTACTGTAGACTTCCTGGG TTTCCCTATTGGCGCTAGAGGGAAGTGGTA CGTTGGTAATAATGGACTCCTTTCCGACCTT GGGTTCTCCACTAGCCGTGTAGTGCGGATA GCGAGGGCCCTCTCTAAAAAGGCTCTCCTA TCGTCCGTGGACATTATACATATTTTTGCGT CTCGCGCTAGACAGGCCCAAACGTCCGAGT AGGGGGCTTGGCATTTCTCATTGCCTGCTCC TGAAAGGATATGGGTCCTGCGTCGCGTGGT AGGCAGACCCATTCGTCCGAGTAGGGGGCT TGGCAGTNTCCATTGCCTGTGCCCGAAAGG ACGTGGGTCATCTGGTCTGTCTGCCTACACC TCTCTAGACTTGTAACATCTAGTCTGTCAAC AAGATCAAAATTCTTCACACAGACGACCGA GCTTGCTCAGTCTTCCTGTACCCGCAGAATT TTGCTCTTGCTCTCCTTTGGCTGTGTCCTGG ACGTGGGACTATTCCATCTCGTCCCAAATGC CGCGTCCAATTATACCGGATTTGACAAAGC GGACGGCCCGCTTTATAAGCCGGAAAAGGT GCCTTGTAAAATTGCAAGGTTCATTAAATA G (SEQ ID NO: 1097)			
R2	R2-1_TSP	.		Trichinella spiralis	CTCCTGACTAACCTGATTTCGTCCGTGCGGC GGCGTTTTCTTTTCGCTCTCCGCTCGTCGAA	CTCCTG ACTAAC	TGAG GTTTT	MSNRLANTAAA GGVPEKTSGTL

					ATTTGCTGTAGTTGATTCGCTTTTCTTTGCGT	CTGATT	TGTTT	DIPGQPSSSGEK
					TTTCTTCTACTTTCGCAGTTTTTTCTGCATTG	TCGTCC	TCTTT	RAISYPGPFGCN
					CCACGATGTCAAACCGCCTTGCCAATACTG	GTGCGG	TTTCC	SCSFTSTTWLSL
					CTGCGGCTGGTGGGGTTCCAGAGAAAACCT	CGGCGT	TTTTA	ELHFKSVHNIRD
					CGGGAACTTTAGACATTCCTGGCCAACCCT	TTTCTTT	CCAT	FVFLCSKCKKS
					CTTCATCCGGTGAAAAGCGTGCGATCTCTTA	TCGCTC	TCTTG	WPSINSVASHYP
					CCCTGGTCCATTCGGTTGCAATTCGTGTTCG	TCCGCT	TTCC	RCKGSVKAAVV
					TTTACGAGTACGACTTGGCTCTCATTGGAAT	CGTCGA	ATTG	PTSLANTCTTCG
					TGCATTTTAAAAGCGTCCATAATATTCGTGA	AATTTG	TTGTT	SSFGTFSGLQLH
					CTTCGTCTTCCTCTGCTCTAAATGTAAAAAA	CTGTAG	ATTT	RKRAHPDVFAA
					AGCTGGCCATCGATCAACTCCGTAGCTAGC	TTGATT	GCTTT	SCSKKTKARWS
					CATTACCCTCGGTGCAAAGGTAGCGTCAAG	CGCTTT	AATC	NDEFTLLARLEA
					GCTGCAGTTGTTCCTACATCTTTGGCGAATA	TCTTTG	CTGT	GLDPACKNINQ
					CGTGCACCACGTGCGGCTCAAGCTTCGGTA	CGTTTT	ATTTT	VLAERLMEYNIT
					CTTTCAGTGGTCTTCAACTCCATCGGAAAAG	CTTCTA	ACCG	RGVEMIKGQRR
					AGCACATCCGGACGTTTTTGCTGCTTCTTGT	CTTTCG	CCGG	KDQYKALVRQL
					AGCAAAAAAACGAAGGCGCGTTGGTCTAAC	CAGTTT	CAAT	RSNSETQQCVG
					GACGAATTTACCCTTCTGGCGAGACTCGAA	TTTCTG	TCCA	LAGSMDSNVPA
					GCAGGTCTGGATCCAGCCTGTAAAAACATT	CATTGC	TTGTT	NDTSSSVASEVS
					AACCAAGTACTAGCGGAAAGGTTAATGGAG	CACG	ATTA	ITYPEYGAVMSC
					TATAACATCACCAGAGGCGTAGAAATGATA	(SEQ ID	TTACT	DLIKEATGMAIV
					AAAGGCCAACGTAGAAAAGATCAGTACAA	NO: 1221)	GTTA	DINELQSNLRKA
					AGCGCTCGTTCGTCAACTCCGGTCAAATTCT		CTGTT	FLSGRKLPMKF

					GAAACACAGCAATGTGTAGGTTTAGCCGGA		ATTA	HGARETAQKKM
					AGTATGGATTCGAACGTACCGGCCAACGAT		TTGTT	ANPRVAKFKRF
					ACATCGTCTTCCGTTGCATCAGAGGTCAGC		ACTA	QRLFRSNRRKL
					ATTACGTACCCTGAGTACGGGGCCGTGATG		TTGTT	ASHIFDKASLEQ
					TCGTGCGACCTAATTAAAGAAGCGACTGGT		TTTAC	FGGSIDEASDHL
					ATGGCCATAGTTGACATCAACGAGTTGCAA		TTTTA	EKFLSRPRLESD
					AGCAACTTACGAAAAGCCTTCTTGTCCGGC		CTTA	SYSVISGDKSIG
					CGCAAGCTTCCCATGAAGTTCCATGGAGCG		CTAC	VAHPILAEEVEL
					CGTGAAACCGCCCAGAAGAAAATGGCCAAC		TGTT	ELKASRPTAVGP
					CCCCGTGTTGCGAAATTCAAGCGTTTCCAAC		ATTA	DGIALEDIKKLN
					GGTTGTTTCGAAGCAACAGGAGGAAACTGG		TACTT	TYDIASLFNLWL
					CCAGCCACATCTTCGACAAAGCCTCACTGG		TAAT	KAGDLPASVKA
					AGCAATTCGGTGGCAGCATCGATGAGGCAT		TCGTT	SRTIFLPKSDGTT
					CTGACCATTTAGAAAAGTTCCTCTCCCGGCC		AACT	DISNCRPITIASA
					AAGATTGGAGTCCGATTCTTATTCCGTGATA		TACG	MYRLFSRIITRR
					AGCGGTGATAAGTCAATCGGAGTTGCACAT		TTATT	LAARLELNVRQ
					CCAATTTTGGCCGAGGAGGTGGAATTGGAA		GTTA	KAFRPEMNGVF
					TTAAAAGCCTCCCGACCAACCGCTGTTGGT		CCAC	ENSAILYALIKD
					CCGGATGGAATTGCACTGGAAGACATTAAA		TACTT	AKVRSREICVTT
					AAACTCAATACTTACGACATAGCCAGTCTTT		ACTTT	LDLAKAFDTVP
					TCAACCTCTGGCTAAAAGCTGGCGACCTAC		GCTC	HSRILRALRKNN
					CCGCATCGGTGAAAGCCAGTAGAACCATCT		TCTC	VDPESVDLISKM
					TTTTGCCCAAAAGCGACGGCACCACCGACA		GCAA	LTGTTYAEIKGL

					TATCGAACTGTCGGCCAATCACAATCGCAT		ACGT	QGKLIPIRNGVR
					CCGCCATGTATCGGCTGTTCAGCAGAATAA		TCGTT	QGDPLSPLLFSL
					TAACGCGACGTCTGGCAGCCAGGTTGGAAT		GTTG	FIDEIIGRLQACG
					TGAACGTGCGGCAAAAAGCGTTCCGGCCTG		TTTCT	PAYDFHGEKICI
					AAATGAACGGCGTATTCGAGAACTCCGCCA		TTTG	LAFADDLTLVA
					TTTTATACGCCCTCATCAAGGATGCTAAGGT		GACC	DSAAGMKILLK
					CAGGTCAAGGGAAATTTGCGTAACTACGCT		AGGT	AACDFLEESGM
					CGACCTTGCCAAGGCCTTTGACACGGTGCC		TTAG	SLNAEKCRTLCI
					CCACTCACGCATTTTACGAGCCCTGAGGAA		AGAA	TRSPRSRKTFVN
					AAATAATGTCGACCCGGAATCCGTCGACCT		ATCG	PAAKFIISDWKT
					GATTTCGAAAATGTTAACGGGTACGACTTA		CACG	GISSEIPSLCATD
					TGCAGAAATAAAAGGGCTCCAGGGCAAACT		CACA	TFRFLGHTFDGE
					TATACCCATTCGCAATGGAGTCAGGCAAGG		GCGG	GKIHIDTEEIRSM
					TGACCCCTTGTCGCCCCTATTATTTAGTCTA		AACT	LKSVKSAPLKPE
					TTTATAGACGAGATAATAGGTCGCCTACAA		GGAC	QKVALIRSHLLP
					GCCTGCGGCCCTGCCTACGATTTCCATGGCG		CGCT	RLQFLFSTAEAD
					AAAAAATTTGCATCCTGGCTTTCGCCGATG		TAAG	SRKAWLIDSIIR
					ATCTGACGCTGGTGGCTGACAGCGCAGCTG		CCAG	GCVKEILHSVKA
					GTATGAAGATCCTTCTAAAAGCGGCTTGTG		AAAT	GMCTDIFYIPSR
					ACTTCCTGGAGGAATCTGGAATGTCACTTA		AGTA	DGGMGFTSLGE
					ATGCAGAGAAATGCCGCACTCTCTGTATTA		AAGT	FSLFSRQKALAK
					CAAGATCTCCCCGAAGCCGCAAGACTTTCG		AACA	MAGSSDPLSKR
					TCAACCCAGCTGCCAAATTCATCATCAGCG		A (SEQ	VAEFFIERWNIA

					ATTGGAAAACGGGTATCAGCTCAGAAATCC CCTCCCTGTGTGCGACGGACACCTTTCGTTT CCTGGGGCACACCTTCGATGGAGAAGGAAA GATCCACATCGATACGGAGGAAATTCGATC CATGCTCAAATCGGTGAAGTCAGCTCCACT GAAACCGGAACAGAAGGTGGCTTTGATACG GTCACACCTTCTTCCCCGCCTTCAGTTCCTG TTTTCTACAGCTGAAGCTGACAGCCGGAAA GCCTGGTTGATCGATTCCATCATCAGGGGG TGTGTGAAGGAGATCTTGCACTCAGTGAAA GCTGGTATGTGCACTGATATCTTTTACATAC CCTCTAGAGACGGTGGAATGGGATTTACTT CCCTCGGGGAGTTTTCTCTTTTCAGCAGGCA GAAGGCACTCGCCAAGATGGCTGGATCGTC GGACCCCCTCTCGAAACGGGTTGCTGAATT CTTCATCGAAAGGTGGAACATCGCCCGTGA CCCGAAAGTCATTGAAGCTGCTCGGCGCGT CTACCAGAAAAAACGGTACCAACGCTTTTT CCAGACGTACCAGAGCGGTGGATGGAATGA ATTTTCGGGAAACACTATTGGGAACGCCTG GTTGACAAACGGCCGTGCCCGCGGAAGAAA TTTCATAATGGCTGTGAAATTCCGTTCCAAC ACCGCAGCCACCCGGGCCGAAAACCTACGA		ID NO: 1344)	RDPKVIEAARRV YQKKRYQRFFQ TYQSGGWNEFS GNTIGNAWLTN GRARGRNFIMA VKFRSNTAATR AENLRGRPGTK ECRFCKSATETL AHICQRCPANH GLVIQRHDAVV TFLGEVARKEG YQVMIEPKVSTP VGALKPDLLLIK ADTAFIVDVGIA WEGGRPLKLVN KMKCDKYKTAI PAILETFHVGHA ETYGVILGSRGC WLKSNDKALAS IGLNITRKMKEH LSWLTFEIIFITQI SRIYNSFMKK

					GGCCGCCCCGGCACGAAAGAATGCCGGTTT TGCAAGAGTGCCACCGAAACTTTGGCACAC ATTTGCCAGAGGTGTCCGGCAAATCACGGC TTGGTTATCCAGCGCCATGACGCAGTCGTA ACATTCCTGGGGGGAAGTGGCGCGGAAGGAA GGTTACCAGGTCATGATAGAGCCTAAGGTG TCAACCCCGGTCGGCGCGCTCAAGCCCGAC CTCCTACTCATCAAAGCCGACACTGCATTCA TTGTGGATGTAGGCATTGCATGGGAAGGTG GACGCCCACTAAAGCTGGTCAACAAAATGA AATGTGACAAGTACAAAACTGCCATCCCGG CAATTTTGGAAACATTTCACGTTGGCCATGC TGAGACGTACGGCGTTATTCTGGGCAGCCG CGGATGCTGGCTCAAGAGCAACGACAAGGC GTTGGCATCAATTGGGCTCAATATCACACG GAAGATGAAAGAACACCTGAGCTGGTTGAC GTTTGAAATTATATTTATAACTCAAATAAGC CGGATTTATAACTCATTCATGAAAAAATGA GGTTTTTGTTTTCTTTTTTCCTTTTACCATTC TTGTTCCATTGTTGTTATTTGCTTTAATCCTG TATTTTACCGCCGGCAATTCCATTGTTATTA TTACTGTTACTGTTATTATTGTTACTATTGTT TTTACTTTTACTTACTACTGTTATTATACTTT			(SEQ ID NO: 1466)

					AATTCGTTAACTTACGTTATTGTTACCACTA CTTACTTTGCTCTCTCGCAAACGTTCGTTGT TGTTTCTTTTGGACCAGGTTTAGAGAAATCG CACGCACAGCGGAACTGGACCGCTTAAGCC AGAAATAGTAAAGTAACAA (SEQ ID NO: 1098)			
R2	R2-1_TUr	Каркас_6		Tetranychus urticae	CTCTCTTATTTTAACATATTCGATGTACTCG TACATTGAATATGCTTTTATTTTTTTTCAAA GTTTTTTGGGTGCATACCCCTGGAAAATTCT GAGATGTATAAATCTCCCATCAGCTTTGGCT GAAACGTTGGCTAAGTTTTGTAGGTTGTTTG CCCCCTACTACTTAGTCGCAAATGGTATTTG CTAACAGTTGTTAAATTGTTACATTTACAAG TCCTATCCAGTGCCTCCTCGTGGCGCTACCC GGTAACACTTAGAGTAATCTGAGTGGCTAA ACTGGAAGGGCGGAAAATGCAAACAGGCG GTTGGTAGATGCTTCGGCATTTTGCCAAAA ATCCACGGCTTTTTAGCCCAACAACATCAG GGTGATGGACCCGCCAGCTTGTGGTCAGGA TCCCATCCATGAATAAAGCATGGCTCTGCTT CTGGTGCATCCTCAACGGGATCGGCTTCGG CTGGATGTAAGTCTTGCGGAGGCATGTGCA TCTTGGGTGGCCTAACAAGCCATTCAAGAG	CTCTCT TATTTT AACATA TTCGAT GTACTC GTACAT TGAATA TGCTTT TGCTTT TATTTTT AGTTTT TTCAA TTGGGT GCATAC CCCTGG AAAATT CTGAGA TGTATA	TGAT GTAT CCCTT CAAT ATAT TGTA ATCC TCATT CGTC CCTA TCCTT TCATT TGAT AAAA GAAA CTTTG TTGCT	MCILGGLTSHSR EGGLSRGSSQLK TVKPQNEEDNG TTQLKAGSADSF PRPSGDLNPEEP LSIDICPVCFRQ MKSYLGVRVH MQKMHLEEYN ASIPDPVVSHTR WSDEEAAQLAF TEAKIEVDKLLP RGKGINKFLLEL LPGRTLESIKSH RKRQSHKDLVR KYVKEFVDTLA ADNDDDTIICQD NGDIFNDPIVGA

					AAGGAGGGCTCTCGCGAGGGAGTTCTCAAC	AATCTC	CCTTT	TDSQSETETVAD
					TTAAGACCGTTAAGCCACAAAACGAAGAAG	CCATCA	TAAT	PAEFKTFIELAD
					ATAATGGAACTACCCAGTTGAAGGCTGGGA	GCTTTG	AGTT	DPTKPKVVAKL
					GTGCGGACTCTTTCCCAAGGCCAAGTGGTG	GCTGAA	GGTC	RNLIKDKPKSEI
					ATTTGAACCCAGAGGAACCACTGAGCATTG	ACGTTG	CCTC	LGSDILVRILRRT
					ATATCTGTCCAGTTTGTTTCAGACAGATGAA	GCTAAG	CTGT	LHGLPVEDELD
					GAGTTATTTAGGCGTTAGAGTGCATATGCA	TTTTGT	CCCTT	QYLEVYFTGKIK
					GAAGATGCACTTAGAAGAATATAATGCGTC	AGGTTG	TTCTG	QRRSKTQTALS
					TATTCCGGACCCAGTCGTTTCCCACACTAGG	TTTGCC	GAAC	KKQIKQRDYGR
					TGGTCCGATGAAGAGGCAGCACAATTAGCT	CCCTAC	CTGG	LQELYSRSRKRC
					TTTACTGAAGCTAAGATTGAGGTCGACAAG	TACTTA	TTGT	ANEILNPTSMSG
					CTTTTACCTCGTGGCAAAGGCATAAATAAG	GTCGCA	ATCG	GFGHQELSEFW
					TTCCTGTTAGAGCTTTTGCCTGGTCGAACCT	AATGGT	ATTA	TKTFGPDEQPTL
					TAGAGTCTATCAAGTCTCACAGAAAAAGAC	ATTTGC	TTGA	GEVEIIPKENCW
					AATCGCATAAAGACTTGGTTCGCAAATATG	TAACAG	AAGT	WDIFSPISSDEIK
					TAAAGGAGTTCGTTGATACTTTAGCTGCAG	TTGTTA	TGCA	ASYPSIGKAAGP
					ATAACGATGACGATACCATCATCTGTCAAG	AATTGT	ATAA	DNFSAYQLRKV
					ATAATGGTGACATTTTTAATGACCCCATCGT	TACATT	ACGG	PVWHLECLYNIF
					TGGAGCAACTGATTCTCAGTCTGAGACTGA	TACAAG	ATTT	AFYKDIPSRLKD
					AACTGTGGCTGATCCTGCTGAATTCAAGAC	TCCTAT	AA	AKTILIPKKDNA
					TTTTATTGAGTTAGCAGACGATCCCACAAA	CCAGTG	(SEQ	ESPGDFRPITLSS
					ACCAAAAGTCGTAGCAAAACTCAGAAACTT	CCTCCT	ID NO:	IITRHFHKILATR
					GATAAAAGATAAACCAAAGAGCGAGATCCT	CGTGGC	1345)	VNNFVRFHPMQ

					TGGAAGTGACATTCTTGTACGAATCCTGCG	GCTACC		RGFIQSDGCLEN
					AAGAACTCTTCATGGCTTGCCTGTTGAGGA	CGGTAA		TALIQTVIREAK
					CGAGTTGGACCAGTACTTGGAGGTTTACTTC	CACTTA		VRRKQVHITFCD
					ACTGGGAAGATAAAGCAGAGACGTTCTAAG	GAGTAA		VRKAFDSVRYD
					ACACAAACTGCCTTATCTAAGAAACAGATT	TCTGAG		SIIAAIAKKGAP
					AAGCAGAGAGATTATGGACGTCTTCAGGAG	TGGCTA		GSFIMYLSNLYR
					TTATACTCCCGGTCTAGAAAAAGGTGTGCA	AACTGG		GNKTTLLTAGG
					AATGAGATCTTGAACCCTACTTCTATGAGTG	AAGGGC		ETRITPTRGVRQ
					GTGGTTTTGGTCATCAGGAGCTATCCGAATT	GGAAAA		GDPLSPILFNCV
					CTGGACAAAGACGTTCGGACCAGATGAGCA	TGCAAA		MDQILTALPSRT
					ACCTACTCTGGGCGAGGTGGAGATCATTCC	CAGGCG		GFTLSAGDESVN
					GAAGGAGAACTGTTGGTGGGATATTTTCTC	GTTGGT		VNCLAFADDIILI
					GCCAATATCTTCAGACGAAATCAAAGCCTC	AGATGC		SKTKNGHQELL
					ATACCCCAGTATAGGGAAAGCTGCTGGTCC	TTCGGC		DVTQRILKENGL
					AGATAATTTCTCTGCTTATCAACTTAGAAAA	ATTTTG		DLNPDKCCSLSL
					GTTCCGGTTTGGCATCTGGAATGTCTCTATA	CCAAAA		IPHSKTKKIKVV
					ATATCTTTGCTTTCTACAAGGACATACCGAG	ATCCAC		RADFVVNGVKV
					TAGACTTAAAGACGCGAAGACAATTTTGAT	GGCTTT		RSMSIGDSTCYL
					ACCAAAGAAGGATAACGCTGAAAGTCCAG	TTAGCC		GVSINVTGQVAP
					GTGATTTCCGCCCTATCACCTTGAGCTCAAT	CAACAA		VKMYQALCEKL
					TATCACTCGACATTTTCACAAGATTCTAGCG	CATCAG		DSAAIKPHQRLY
					ACGAGAGTGAATAATTTTGTTCGATTCCATC	GGTGAT		ILKHFVITKMFH
					CTATGCAAAGAGGCTTCATTCAATCTGATG	GGACCC		PLILSTIAAHKIK

					GTTGTTTGGAGAACACTGCGCTCATACAGA	GCCAGC		NLDLISRRYVRK
					CGGTCATTAGAGAAGCCAAAGTGCGTAGGA	TTGTGG		WLHLPHDCGSG
					AGCAAGTGCATATCACCTTCTGTGATGTGC	TCAGGA		MIHAKVSDGGL
					GAAAAGCGTTTGACAGCGTCCGGTATGACT	TCCCAT		GVPLLFRTIADL
					CGATCATAGCAGCTATTGCAAAGAAAGGTG	CCATGA		KVRRKEKLQVH
					CACCAGGAAGTTTCATCATGTACCTTAGTA	ATAAAG		ENPIFRILAKLST
					ACTTATATAGAGGAAATAAGACTACGCTTT	CATGGC		VSKELENCKKIA
					TGACTGCTGGCGGTGAAACAAGAATAACTC	TCTGCT		SKTTDIQEKTFK
					CAACACGGGGCGTTAGACAAGGAGATCCGT	TCTGGT		EMLATYDGLSL
					TATCTCCGATACTTTTTAATTGCGTGATGGA	GCATCC		KEARAVPEVHK
					CCAGATTCTCACGGCTCTTCCTTCGAGAACT	TCAACG		WVDSYDKRYKF
					GGTTTTACGTTATCTGCTGGTGACGAATCAG	GGATCG		AGRDFVQVIQA
					TTAACGTTAACTGTTTGGCGTTTGCTGATGA	GCTTCG		RFNALPTRSRV
					TATAATTTTAATCTCAAAGACCAAGAATGG	GCTGGA		WRGRGADEKSL
					ACATCAAGAACTCCTTGATGTAACACAAAG	TGTAAG		RCRAGCNARET
					AATTTTAAAGGAGAATGGCCTCGATCTCAA	TCTTGC		LNHVSQSCFRTH
					TCCTGATAAGTGCTGTTCGCTAAGCTTGATA	GGAGGC		RVRTARHDKIL
					CCGCACAGCAAAACAAAGAAAATCAAAGT	(SEQ ID		DFICERLDVVGV
					GGTTCGTGCTGACTTCGTGGTGAACGGAGT	NO: 1222)		KYVREKPISFPG
					CAAAGTGCGATCAATGAGTATTGGTGATTC			KKLIPDLIVENT
					AACTTGTTACCTAGGTGTCTCAATAAATGTA			DQALVLDLQIV
					ACTGGCCAAGTAGCACCAGTTAAAATGTAT			GDNSELPLDERG
					CAGGCTCTGTGTGAGAAACTTGATAGTGCT			KNKVIKYNCSE

					GCTATTAAGCCTCATCAGAGACTGTACATA TTGAAGCATTTCGTAATAACGAAGATGTTTC ACCCGCTGATTTTGTCAACTATTGCGGCACA CAAGATAAAGAATCTGGATTTGATTAGCAG ACGTTATGTCAGGAAGTGGTTACACTTACCT CATGATTGCGGATCTGGTATGATTCATGCCA AAGTAAGCGATGGAGGATTGGGTGTTCCAT TACTGTTCAGGACAATTGCTGATCTCAAAGT TCGGAGAAAAGAAAAGCTGCAGGTCCATGA AAACCCAATATTTAGAATATTGGCAAAGCT ATCTACAGTAAGCAAGGAGTTGGAAAATTG TAAAAAGATAGCTTCCAAGACGACAGACAT TCAAGAAAAGACTTTCAAGGAAATGCTTGC GACTTATGATGGTTTGTCATTGAAAGAAGC TCGAGCGGTACCTGAGGTGCACAAATGGGT GGACAGCTATGATAAGCGTTATAAGTTTGC AGGCCGAGATTTCGTTCAAGTTATACAGGC TCGTTTCAATGCGTTACCAACGAGATCTCGA GTATGGCGGGGCAGAGGTGCTGATGAAAAA TCGTTACGTTGCAGAGCAGGCTGCAACGCT AGGGAAACGTTGAATCACGTTTCACAATCT TGCTTCAGGACGCATCGAGTACGAACAGCT AGACACGACAAAATCCTGGATTTTATTTGT			MQELYKRKKKT LAVKALTLHYK GLMAPETSNILR SFGFKSKDLEK MAYMALFGTV AAWGIFNRSTET MRSVANWPRPE EL (SEQ ID NO: 1467)

					GAAAGGCTGGACGTAGTCGGTGTTAAATAC GTACGAGAAAAGCCAATCAGCTTTCCTGGC AAAAAGTTGATTCCTGATTTGATTGTGGAA AACACGGACCAAGCTCTTGTGCTCGATCTC CAGATTGTTGGCGATAACAGTGAACTTCCA CTTGATGAAAGAGGCAAGAACAAGGTTATC AAATACAATTGTTCGGAGATGCAGGAATTG TACAAGAGGAAGAAAAAGACTTTGGCGGTT AAGGCACTCACGCTACATTATAAAGGGTTG ATGGCACCGGAAACAAGCAATATCCTAAGG TCATTCGGATTCAAGTCTAAGGATTTGGAA AAGATGGCGTATATGGCGTTGTTTGGTACG GTGGCTGCTTGGGGCATTTTCAATCGCTCGA CGGAGACAATGCGTTCAGTCGCTAACTGGC CTAGACCTGAGGAATTATGATGTATCCCTTC AATATATTGTAATCCTCATTCGTCCCTATCC TTTCATTTGATAAAAGAAACTTTGTTGCTCC TTTTAATAGTTGGTCCCTCCTGTCCCTTTTCT GGAACCTGGTTGTATCGATTATTGAAAGTT GCAATAAACGGATTTAA (SEQ ID NO: 1099)			
R2	R2-2_DWi	.		Drosophila willistoni	GAAGCTGGGAAGCTGGGTCGGATGAGCGCA GAAGGGGTGTTCTTCGGAGCACTGTAATTC ATAAGTCGTAAGTCTGATCAAGTCGACTCG	GAAGCT GGGAAG CTGGGT	TAGA TGTA CTAA	FERRSNSWGYQ NLEPSNVGQDM NTVPRINNTTTT

					GAACCTCTTCGTGGTGTTTCCTGGGTGCTGT	CGGATG	CCTCT	PATSRPGDQPRE
					TGAGTTCCTAGTCTCTAGGTTCTCTCCAGTA	AGCGCA	AGTT	AIAVVNLAGEIP
					GCTAATTCGAGCGGCGAAGCAACTCTTGGG	GAAGGG	TCTCT	CAVCGRLFNTR
					GTTACCAGAACCTTGAGCCAAGTAACGTTG	GTGTTC	ATAC	RGFGVHMSHQH
					GTCAAGATATGAATACTGTGCCTCGGATAA	TTCGGA	TTTTG	KDELDTQRQRE
					ACAACACAACAACAACTCCAGCGACTTCCC	GCACTG	CCTG	DVKLRWSEEEA
					GTCCTGGAGACCAACCGAGAGAGGCTATAG	TAATTC	CTAC	WMMARKEVEL
					CAGTGGTAAATCTCGCGGGAGAGATTCCCT	ATAAGT	CTTG	EASGNLRFPNK
					GTGCAGTATGTGGGCGCCTCTTCAATACTA	CGTAAG	GCAT	KLAEVFTHRSSE
					GAAGGGGGTTTGGGGTTCATATGTCACATC	TCTGAT	TACA	AIKCFRKRGEYK
					AACACAAAGACGAACTAGATACGCAACGTC	CAAGTC	TCTA	AKLEQIRGQSTP
					AGCGTGAAGATGTAAAACTCCGATGGAGTG	GACTCG	AAAA	TPEALDSITSQPR
					AGGAAGAAGCGTGGATGATGGCGAGAAAG	GAACCT	GGTA	PSLLERNHQVSS
					GAGGTGGAGCTCGAAGCAAGTGGTAATTTG	CTTCGT	CAAA	SEAQPINPSEEQS
					AGATTTCCTAATAAGAAGCTAGCGGAAGTA	GGTGTT	CATC	NWEIMRILQGY
					TTTACTCACCGTAGCTCCGAAGCAATTAAAT	TCCTGG	GCAT	RPVECSPRWRA
					GTTTTCGGAAGAGGGGTGAATATAAGGCAA	GTGCTG	TGGC	QVLQTIVDRAQ
					AACTGGAGCAGATCAGAGGGCAATCTACTC	TTGAGT	AAAA	AVGKETTLQCL
					CCACCCCAGAAGCGTTGGACTCAATTACCT	TCCTAG	AGAG	SNYLLEVFPLPN
					CACAGCCTCGCCCTAGTTTACTCGAGCGAA	TCTCTA	GTGG	EPHTIGRSNLRR
					ACCACCAAGTATCATCGTCGGAAGCGCAAC	GGTTCT	TTTTA	PRTRRQLRQQE
					CAATCAATCCATCAGAAGAACAGTCGAACT	CTCCAG	GTAC	YAQVQRRWDK
					GGGAAATCATGCGGATACTACAGGGCTATC	TAGCTA	ATAG	NTGRCIKSLLDG

					GCCCCGTAGAATGTAGTCCCCGGTGGAGAG	A (SEQ	GCGC	TDESVMPNQEI
					CCCAGGTCTTGCAAACTATCGTAGATAGGG	ID NO:	TGTG	MEPYWKQVMT
					CGCAGGCCGTAGGGAAGGAAACCACTCTCC	1223)	GGAC	NPSTCSCENTRF
					AATGCTTATCCAACTATCTCCTGGAAGTATT		TTCAT	RMEHSLETVWS
					TCCATTACCAAACGAACCACACACCATCGG		TGTC	AITPRDLRENKL
					TCGGAGCAATTTGCGAAGACCTCGAACTAG		CCGA	KLSSAPGPDGIT
					GAGACAGTTAAGACAACAAGAGTACGCAC		TGAT	PRTARSVPLGIM
					AGGTTCAGCGTCGTTGGGATAAGAATACTG		GCAG	LRIMNLILWCGK
					GGAGATGCATTAAATCCTTGCTTGATGGAA		CGAA	IPFSTRLARTIFIP
					CAGATGAGTCGGTTATGCCAAACCAAGAGA		TCGT	KTVTANRPQDF
					TAATGGAACCCTATTGGAAACAAGTAATGA		GCAT	RPITVPSVLVRQ
					CGAATCCCAGCACATGCTCTTGCGAAAACA		ACGA	LNAVLASRLAS
					CAAGATTCCGTATGGAACATTCGCTTGAGA		GATT	KVNWDPRQRGF
					CGGTTTGGTCAGCGATAACGCCACGCGACC		GTCC	LPTDGCADNAT
					TGAGGGAAAATAAGTTAAAGTTGTCAAGTG		AGTA	LVDLILREHHKR
					CTCCGGGTCCTGACGGTATCACTCCAAGAA		GTTG	WKSCYLATVDV
					CAGCCAGGAGTGTACCCTTAGGCATTATGC		GTTG	SKAFDLVSHQAI
					TACGCATAATGAACCTGATTCTCTGGTGCG		CTCG	IKTLQAYGAPTN
					GCAAAATACCATTCTCTACCCGACTGGCCA		TATCT	FVSFIEEQYKGG
					GAACTATCTTCATTCCGAAGACTGTGACGG		TTAG	GTSLNGAGWSS
					CAAATCGACCGCAAGACTTTCGTCCAATAA		AAGA	EVFIPARGVKQG
					CAGTCCCCTCGGTTTTGGTCAGGCAATTAAA		TTTCC	DPLSPLLFNLIID
					CGCTGTTCTGGCTTCTCGATTGGCTTCTAAA		TTCCT	RLLRSYPREIGA

					GTCAACTGGGATCCAAGGCAGCGCGGTTTC		CGGC	KVGNTMTSAAA
					CTACCTACCGATGGGTGTGCTGATAATGCG		GATC	FADDLVLFAETP
					ACGTTGGTTGATCTCATTTTGCGGGAGCACC		AAAA	MGQTLLDTTLG
					ATAAACGGTGGAAGTCATGTTACCTTGCGA		NAAA	FLASVGLSLNAD
					CGGTGGATGTCAGCAAGGCTTTTGACTTAG		AAAA	KCFTVSIKGQAK
					TATCACACCAGGCCATTATCAAGACTTTAC		AAAA	QKCTVVERRSF
					AGGCCTATGGTGCTCCAACAAACTTTGTTA		AAAA	CVGERECPSLKR
					GCTTTATAGAAGAACAGTATAAGGGCGGCG		AAA	TEEWKYLGIRFT
					GAACCTCCCTCAATGGGGCAGGATGGAGTT		(SEQ	ADGRAQYSPAD
					CAGAGGTGTTTATACCCGCGCGGGGCGTTA		ID NO:	DLGPKLLRLTRA
					AGCAAGGTGACCCTCTGTCTCCACTATTATT		1346)	PLKPQQKLFAH
					TAATCTTATCATTGATAGATTACTTAGGTCC			RTVLIPQLYHQL
					TACCCCAGAGAGATTGGTGCCAAAGTCGGA			TLGSVMIGVLG
					AATACCATGACAAGCGCGGCAGCGTTCGCG			KCDRLVRQFVR
					GATGATCTGGTGCTATTTGCGGAAACTCCG			RWLDLPLDVPV
					ATGGGGCAAACATTGTTGGATACCACGCTA			AYFHAPHTCGG
					GGCTTCCTAGCCTCCGTGGGACTCTCCCTTA			LGIPSIRWIAPML
					ATGCTGATAAGTGTTTCACTGTCAGTATAAA			RLKRLSNIKWPH
					GGGGCAAGCCAAGCAGAAGTGTACTGTCGT			LEQSEVASSFID
					AGAACGACGGAGCTTTTGTGTAGGGGAGCG			DELQRARDRLK
					CGAGTGTCCTTCATTGAAGCGTACTGAAGA			AENVQRCSRPEI
					GTGGAAGTATTTAGGTATCCGGTTCACTGC			DSYFANRLYMS
					GGATGGGCGGGCTCAATATAGTCCAGCAGA			VDGCGLREAGH

					CGACCTCGGTCCGAAGCTGTTAAGATTAAC AAGAGCCCCTCTGAAACCACAACAGAAGTT ATTTGCACATAGGACTGTCCTTATCCCACAA CTCTATCACCAACTAACACTTGGGAGTGTG ATGATAGGCGTCCTAGGAAAATGTGACAGA TTGGTACGGCAATTCGTAAGGAGATGGTTA GATCTCCCACTGGATGTACCAGTTGCGTACT TTCACGCCCCCCACACTTGTGGGGGTCTCGG GATTCCGTCAATTAGATGGATAGCACCGAT GCTTCGTCTGAAGCGATTGAGCAATATTAA ATGGCCCCACCTCGAACAATCCGAGGTAGC TAGCTCTTTCATTGACGACGAATTGCAAAG GGCTCGAGATAGATTAAAGGCGGAAAATGT GCAGCGGTGTTCGCGTCCAGAGATTGACTC GTATTTCGCAAATAGGTTGTACATGTCTGTT GATGGTTGCGGTCTCCGTGAAGCAGGTCAT TATGGCCCGCAACATGGATGGGTGAGTCAG CCCACGCGCTTGCTAACAGGAAAGGAATAT TTGCACGGTGTCAAATTGCGGATAAATGCC CTACCCTCGAAGTCTCGTACGACGAGGGGA AGGCACGAATTGGAGAGACGGTGTCGTGCA GGATGTGATGCTCCCGAGACAACAAACCAC ATCTTGCAAAAATGCTATCGTACGCATGGG			YGPQHGWVSQP TRLLTGKEYLH GVKLRINALPSK SRTTRGRHELER RCRAGCDAPET TNHILQKCYRTH GRRVARHNSVV NAVKRGLERKG CVAHVEPSLQC DSGLNKPDLVGI RQNHIYVIDVQV VTDGHSLDQAH QRKVERYDRAD IRSQMRRFFGVT GEIEFHSVTLNW RGIWSGQSVKR LIAKDLLIAEDT KLISVRAVNGG VTSFKYFMYCA GYTRS (SEQ ID NO: 1468)

					AGGCGGGTAGCTAGACACAACAGCGTAGTA AATGCCGTCAAGCGGGGACTTGAACGGAAA GGCTGCGTTGCCCATGTCGAACCAAGTCTG CAATGCGACTCGGGCTTAAATAAACCGGAC CTGGTGGGAATCCGACAGAATCACATTTAT GTGATAGACGTTCAGGTTGTGACAGACGGA CATTCCTTAGACCAAGCGCACCAGCGCAAG GTCGAAAGGTACGACAGAGCTGACATAAGA TCACAAATGCGGCGATTTTTCGGAGTGACA GGTGAAATCGAGTTTCATTCCGTTACACTCA ACTGGAGAGGAATCTGGAGTGGTCAGTCGG TAAAACGATTGATTGCAAAAGATCTCCTCA TCGCTGAAGATACCAAACTCATCAGCGTCA GAGCAGTAAACGGCGGAGTGACGTCTTTTA AATATTTCATGTATTGTGCTGGGTATACTCG AAGCTAGATGTACTAACCTCTAGTTTCTCTA TACTTTTGCCTGCTACCTTGGCATTACATCT AAAAAGGTACAAACATCGCATTGGCAAAAA GAGGTGGTTTTAGTACATAGGCGCTGTGGG ACTTCATTGTCCCGATGATGCAGCGAATCGT GCATACGAGATTGTCCAGTAGTTGGTTGCTC GTATCTTTAGAAGATTTCCTTCCTCGGCGAT			

					CAAAANAAAAAAAAAAAAAAAA (SEQ ID NO: 1100)			
R2	R2-2_PM	.		Petromyzon marinus	CGGTGCGTTCCCTTGGGTAAGGAACACGAG	CGGTGC	TGAG	RPQTKLMTDKL
					TCTTAGTGGCCTTGACCTCCACGTGGTCCCG	GTTCCC	AATA	KFSSQLARGLA
					CTGGTAACATCATCTCTTGATGATGGCTAAC	TTGGGT	ATGA	KQRAMDGARV
					AAGGCTAATGCACCCATTCCATCTCCTATCT	AAGGAA	GGTG	GDPPITVRPTET
					CGCATGGAGGCCGCTATGCGTGATTACTAG	CACGAG	CTAA	DLCNTEGSWGR
					AGGCCACAAACAAAACTTATGACTGACAAA	TCTTAG	CCTC	RPMKLLFVSVST
					CTTAAGTTCTCAAGCCAGTTGGCACGAGGC	TGGCCT	CCTG	QTQNEDALWAS
					CTGGCAAAACAACGTGCTATGGATGGCGCT	TGACCT	GGCC	DVAKPMASRSA
					CGGGTTGGCGATCCACCCATTACAGTTAGA	CCACGT	TGAC	LKMTSIPSMTFH
					CCCACAGAAACCGATCTGTGCAACACTGAG	GGTCCC	CAAA	NSSLEKEEEMN
					GGTTCATGGGGACGCCGTCCTATGAAACTC	GCTGGT	CCCA	YDFYEQIKSLVE
					TTGTTTGTCTCGGTGTCAACCCAGACACAGA	AACATC	GAAC	SDDSSDDFTEDD
					ATGAAGATGCCCTCTGGGCATCTGATGTTG	ATCTCT	ACAT	EDVEESFLDISA
					CTAAACCTATGGCGTCTAGGTCGGCGCTAA	TGATGA	CACT	EEPVLGKFPIDT
					AAATGACGAGTATACCTTCCATGACCTTCC	TGGCTA	GGCC	KGTITVVLPSLE
					ATAACTCGTCCTTGGAAAAGGAAGAGGAGA	ACAAGG	AAGA	YICVICKQHMG
					TGAACTACGATTTTTACGAACAGATTAAAA	CTAATG	TGAT	KASELVAHFNIK
					GTCTAGTTGAGTCGGATGACTCTTCAGATG	CACCCA	TTCCC	HRDIPLVFKCAK
					ACTTTACAGAGGATGATGAGGATGTGGAGG	TTCCAT	GCAG	CDKTNSNHRSIA
					AGTCCTTCCTCGACATATCGGCTGAGGAAC	CTCCTA	CACG	CHAPKCGGIKLT
					CCGTGTTGGGAAAGTTTCCCATTGACACCA	TCTCGC	TTGCT	EESLPMVCECC

					AGGGAACTATCACGGTTGTACTGCCTAGTC	ATGGAG	TTTCT	QARFATLSGLSQ
					TGGAGTATATATGCGTTATCTGTAAACAAC	GCCGCT	CTCTC	HKRHAHPVTRN
					ACATGGGCAAAGCATCCGAACTTGTTGCAC	ATGCGT	GATA	EERIKDGIKGTS
					ATTTTAACATCAAACACAGAGACATTCCTCT	GATTAC	CCCG	QRGVHRSCWSL
					GGTGTTTAAGTGCGCTAAATGTGACAAAAC	TAG	AGAA	KEVEQLALLEL
					CAACTCAAACCACCGATCGATTGCCTGTCA	(SEQ ID	TGTTC	QFQGKKNINKII
					TGCCCCCAAATGTGGGGGAATAAAGTTAAC	NO: 1224)	TGCG	AEALGTKTNKQ
					TGAGGAAAGCTTACCAATGGTCTGTGAATG		GAAC	VSDKRRDLSKK
					TTGCCAGGCACGCTTTGCGACTCTAAGTGG		TACG	TGAPMSDSLHFS
					CCTTTCGCAGCACAAGAGGCATGCTCACCC		CAGT	SRPLETLSPPPN
					AGTCACCCGTAATGAGGAAAGGATTAAAGA		CTAT	VTTGTSSILAQA
					TGGTATAAAGGGTACCTCGCAGAGAGGGGT		GAAC	AERLTNENSGTL
					ACACCGTAGCTGCTGGTCTTTGAAGGAAGT		AGTC	EKPAMEAIKAW
					AGAACAGCTGGCCCTTCTAGAGTTGCAGTT		ACAG	LNGEGQHDALV
					TCAGGGGAAAAAGAATATCAATAAGATCAT		ACAA	ETATALMLCPM
					TGCTGAAGCGCTTGGGACTAAGACCAACAA		CCTCT	RLVKNKGKRSK
					GCAAGTCTCTGACAAGAGACGGGACCTAAG		GATC	PENDIIKPRILPT
					TAAAAAGACAGGGGCCCCCATGTCAGACAG		CAAG	RSWMKKRAEK
					CTTACATTTTTCTTCTAGGCCTCTTGAGACA		(SEQ	RGSFMKHQKLF
					TTGTCTCCCCCACCAAATGTAACAACGGGG		ID NO:	FKNRSLLASLVL
					ACTTCATCCATACTCGCTCAAGCAGCTGAG		1347)	DGTERHECRIPN
					CGGCTTACGAATGAGAATTCTGGGACCCTG			ADVYRFYCEKW
					GAAAAGCCTGCAATGGAGGCAATAAAGGCT			EKVLPFNGLGQF

					TGGCTTAACGGCGAGGGCCAACATGATGCC			KSSGVANNEYF
					CTCGTAGAAACTGCCACAGCATTGATGCTTT			EPLISVEEVQTAI
					GTCCGATGAGATTGGTGAAAAACAAAGGCA			RAIKPTSAAGPD
					AACGTTCAAAACCCGAGAACGACATTATTA			GLTRAAICAADP
					AACCCAGGATATTACCCACACGATCTTGGA			EGRTLTALFNA
					TGAAGAAGAGAGCGGAAAAACGAGGAAGC			WMITGIIPKELK
					TTCATGAAGCACCAGAAGCTCTTCTTTAAG			KNRTILIPKVMD
					AACCGCTCTCTTCTTGCGTCCTTAGTCCTGG			DEKLKELGNWR
					ATGGCACTGAACGTCATGAATGCCGAATCC			PITIGSMILRLFS
					CGAACGCAGATGTATATCGTTTTTACTGCGA			RIMTARLARACP
					AAAATGGGAGAAGGTGTTGCCATTCAATGG			LNPRQRGFIAAS
					CCTGGGCCAATTTAAGTCATCAGGTGTTGC			GCSENLKVLQD
					AAATAACGAATACTTTGAGCCCCTAATTTC			LMRHAKKLHRP
					GGTGGAGGAAGTTCAGACTGCCATACGGGC			LAVMFIDIAKAF
					CATTAAACCAACGTCAGCAGCTGGGCCAGA			DSVSHAHILWV
					TGGCCTAACAAGGGCTGCAATCTGTGCTGC			LRHKKVDEHVV
					CGACCCCGAGGGTCGGACACTGACAGCCCT			GIIQNAYDRCTT
					ATTTAATGCATGGATGATTACAGGAATTATT			SFKSNGESTREIS
					CCCAAAGAGTTGAAAAAGAATAGGACGATT			IRVGVKQGDPM
					CTTATTCCTAAGGTTATGGACGATGAAAAG			SPLLFNLAMDPL
					CTGAAAGAATTGGGGAACTGGAGACCAATA			ICTLESHGVGYS
					ACGATTGGTTCAATGATTCTGAGATTATTTT			IDTDHVTALAFA
					CCAGAATAATGACTGCACGTCTTGCTCGTG			DDLVLVSESWV

					CTTGTCCCTTAAACCCAAGGCAGCGTGGTTT			GMAANLAILESF
					TATAGCGGCATCTGGCTGCTCTGAAAATCTT			CGLSGLEVQAR
					AAGGTGCTACAGGACCTTATGAGACACGCT			KCQGFMISPTKD
					AAGAAATTGCACAGGCCGTTGGCTGTCATG			SYTVNNCDPWT
					TTCATCGACATAGCGAAAGCTTTTGACTCG			IKNKDVHMIQP
					GTTTCGCATGCTCATATTTTATGGGTGTTAA			DESTKYLGLKIC
					GGCACAAGAAAGTAGATGAACACGTGGTG			PWTGIIRSDLHV
					GGCATCATCCAGAACGCCTACGATCGGTGT			QLKTRISKIDEA
					ACGACCTCGTTCAAAAGCAATGGCGAGTCG			PLKPTQKVELLN
					ACTCGAGAAATTAGCATACGTGTTGGTGTC			AYALPRLLYPA
					AAACAGGGTGACCCCATGTCACCCCTGCTC			DHSDCKQSTLR
					TTCAATCTTGCCATGGACCCTTTGATATGCA			VLDQEIIKAVKG
					CCCTAGAGTCACACGGAGTTGGGTACTCCA			WLHLPASTCDG
					TTGATACCGACCACGTGACAGCTCTTGCGTT			LLYARARDGGL
					TGCTGATGATTTGGTGTTGGTGAGCGAATCT			AILKLENAIPSV
					TGGGTTGGTATGGCCGCCAATCTAGCGATC			QVRRLQRIANSS
					TTGGAATCATTTTGTGGGCTATCGGGATTGG			DAIARNIASSQG
					AGGTTCAGGCCAGAAAGTGCCAGGGCTTCA			VEEEYRSLWVR
					TGATAAGCCCAACCAAAGATTCATATACGG			AGGDSEAIPTFF
					TGAACAACTGCGACCCATGGACTATCAAAA			LRGSESKEPVYP
					ATAAAGATGTCCATATGATCCAACCTGATG			RPCDWRKRESR
					AATCAACGAAATACCTTGGTCTAAAAATTT			RRCEKPVQGRGI
					GCCCTTGGACTGGCATTATACGGTCGGATCT			VNFAQDRISNA

					ACATGTTCAACTAAAGACACGGATCTCGAA			WLGPRCGFKQC
					AATCGATGAGGCGCCTCTGAAACCGACTCA			FFIAALQLRANI
					GAAGGTCGAACTCCTCAATGCCTACGCCCT			YPTRESINRGRD
					ACCCAGATTATTGTACCCTGCTGACCACTCG			GASRSCRKCSA
					GACTGCAAGCAATCAACTCTCCGTGTGTTG			RLESLSHILGQC
					GACCAAGAAATAATAAAGGCGGTAAAAGG			PAVQKFKDCAT
					ATGGCTCCATCTTCCCGCGTCAACCTGTGAC			QTEAEMVPHGP
					GGGCTGTTGTACGCCAGAGCCCGAGACGGA			AGNALPGWSLS
					GGCCTTGCCATCTTGAAACTGGAAAATGCA			RTFLVNVPQYK
					ATTCCTTCGGTTCAAGTTAGAAGGCTGCAA			NSRIARRNKISDI
					CGTATTGCAAACTCCTCTGACGCTATCGCTC			LADEAARLGW
					GAAACATTGCGTCCTCGCAGGGTGTGGAGG			WVYKEPRFTSE
					AAGAGTACCGAAGTCTGTGGGTACGGGCAG			AGELRKPALVF
					GGGGTGACAGCGAAGCAATCCCAACGTTCT			AKGEEALVIDVT
					TTCTCAGGGGTTCGGAATCAAAAGAGCCCG			VRFELSRKTSSE
					TGTATCCGAGACCCTGCGATTGGAGGAAAC			AASHQVAYYTP
					GCGAATCTCGGAGACGGTGCGAAAAGCCGG			PCDQVKVLTKA
					TTCAAGGAAGGGGCATTGTAAACTTTGCGC			SNVTFFGFQVG
					AAGATAGAATCAGTAATGCATGGTTGGGGC			ARGKVAP (SEQ
					CACGGTGCGGCTTTAAACAATGCTTCTTTAT			ID NO: 1469)
					CGCAGCATTACAATTAAGGGCAAATATTTA			
					CCCAACAAGAGAAAGCATAAACAGAGGCA			
					GAGATGGTGCCTCACGGTCCTGCAGGAAAT			

					GCTCTGCCAGGCTGGAGTCTCTCTCGCACAT TCTTGGTCAATGTCCCGCAGTACAAAAATTC AAGGATTGCGCGACGCAATAAGATCAGCGA CATTCTAGCTGACGAAGCGGCGAGACTGGG CTGGTGGGTGTACAAAGAGCCACGGTTCAC ATCTGAAGCCGGAGAGCTAAGGAAACCTGC CCTTGTGTTTGCCAAAGGTGAGGAAGCGCT TGTTATTGATGTCACCGTCCGGTTTGAGCTC TCGAGGAAAACCTCATCAGAGGCTGCCTCG CACCAAGTTGCGTACTACACCCCCCCTTGTG ATCAAGTCAAAGTGCTGACGAAGGCAAGCA ATGTCACATTCTTTGGATTCCAGGTTGGGGC AAGAGGGAAAGTGGCCCCTTGAGAATAATG AGGTGCTAACCTCCCTGGGCCTGACCAAAC CCAGAACACATCACTGGCCAAGATGATTTC CCGCAGCACGTTGCTTTTCTCTCTCGATACC CGAGAATGTTCTGCGGAACTACGCAGTCTA TGAACAGTCACAGACAACCTCTGATCCAAG (SEQ ID NO: 1101)			
R2	R2-2_SMed	.		Schmidtea mediterranea	AGTCATAGGGTGAACTGCAATTCTGACACG ATGACCGAGCTGTGTCAGTTTGCAGCTAGT CGCTAAAGACTCGATCAGTCCGCCAAGTGA GGTGGCCGGGTATCTGCAGCACTAGAGCCA	AGTCAT AGGGTG AACTGC AATTCT	TAGA AGGG AAAC AAAG	RKELVTIKNLFE ESGATAPAPVPL EVAVEVHQSSS VPEITDESTTTQ

					CTGGTATCAAGAGCAGAGATACGCGAGTGG	GACACG	GAAA	EGSYSEPPIHRC
					AAGTTGAGTACGACTACCTTCACGGGGTCC	ATGACC	AACG	ENCGREFRTRA
					TCCTGATAACCACAGTGGACTGTGGGAACT	GAGCTG	AAAT	GVQQHRRKAHT
					AAATGTGTGCTCAGCGTTCCCTACTTTCTCG	TGTCAG	GACT	NEFMEEKEKAA
					TAGGGTAAAGGGTATGATAACCCAGAGAAT	TTTGCA	GGAA	PTKKLRWTDEE
					ATCCCATGGGAGATATCCATGGAAAAAGCA	GCTAGT	ACTA	KEILIESEIKIIKE
					CCACGTTAGACAATCCGATGGTCTAACTCG	CGCTAA	TGAA	GSLKEQHEINKI
					GCTCCGAGGGGCTAACTATCCCAAAGGGCT	AGACTC	GGAT	LASRMPGRSQD
					TAAAGAAAAGAACTGGTTACAATCAAAAAT	GATCAG	ATAG	GIAKIRQKQEHK
					CTATTTGAAGAGTCAGGTGCCACCGCGCCT	TCCGCC	CTGA	AEIQRRLHGTVT
					GCACCAGTTCCCCTAGAAGTTGCAGTTGAG	AAGTGA	AAGC	TNETRGNRTSEI
					GTACACCAGTCATCGAGTGTTCCCGAGATA	GGTGGC	CGCA	TEPIRSLPINTKT
					ACCGACGAATCTACGACTACTCAGGAAGGA	CGGGTA	AGGA	WSEDEMKRML
					AGCTATTCAGAACCACCGATACACCGATGT	TCTGCA	AGGC	AEEVKLRTKNE
					GAGAACTGTGGAAGAGAATTCAGAACAAG	GCACTA	TAAG	KDINKKLAEIFP
					AGCAGGAGTTCAACAGCACCGGAGAAAAG	GAGCCA	TCCT	NRTMGSIKSKRT
					CTCACACCAACGAGTTTATGGAGGAGAAAG	CTGGTA	GAAA	KDKDYQDLVKL
					AGAAGGCAGCTCCAACCAAGAAGCTTCGAT	TCAAGA	CCGA	TMQTISENPDNE
					GGACAGACGAAGAAAAAGAGATCCTCATC	GCAGAG	TCTA	TDFNTSNTENNS
					GAAAGCGAGATAAAGATCATCAAGGAAGG	ATACGC	CATC	TDAEKEVKNYL
					ATCCCTTAAAGAACAACACGAAATAAACAA	GAGTGG	TTCG	NMLLLTINEEE
					GATTCTAGCTTCAAGAATGCCTGGAAGATC	AAGTTG	ATCC	WLTSTLKEAAT
					ACAAGACGGAATAGCCAAAATTCGCCAGAA	AGTACG	CAAG	LALQGKKTEAS

					ACAAGAGCACAAGGCTGAAATACAGAGAA	ACTACC	AGGA	EKLNEYASKTLF
					GGCTACATGGGACCGTCACTACCAACGAAA	TTCACG	ACTG	PGLKITNQTRKR
					CCAGAGGAAACAGAACCAGCGAAATTACC	GGGTCC	TGGG	EKKISKRETRRQ
					GAACCAATAAGAAGCTTACCAATAAACACC	TCCTGA	TTAA	EYAEIQKLYKK
					AAAACCTGGAGCGAAGACGAAATGAAAAG	TAACCA	GCTT	NISSAAEKAING
					AATGCTAGCCGAAGAAGTGAAGTTGAGAAC	CAGTGG	GAGC	KWSIKPEEEYHN
					CAAGAATGAAAAGGATATCAACAAGAAGC	ACTGTG	CGAC	NKDLIKAWKPIL
					TAGCAGAAATATTCCCGAACAGAACAATGG	GGAACT	GGAA	EAPPFSDCRPIEN
					GGTCCATAAAGAGCAAAAGGACGAAAGAT	AAATGT	AAAG	IKEMDYALMEIS
					AAGGACTACCAGGATTTGGTAAAGCTAACA	GTGCTC	CGAA	TAEIFLAIRAMG
					ATGCAAACAATCAGCGAGAACCCAGACAAT	AGCGTT	TGCA	KTAPGPDGIKYS
					GAAACAGACTTCAATACCAGCAACACGGAA	CCCTAC	TGTT	KLKKNIQSMAIL
					AACAACAGCACTGATGCAGAGAAAGAAGT	TTTCTC	AGAC	FNTCLLTSFLPLP
					GAAGAACTACCTCAATATGCTACTACTGAC	GTAGGG	GACG	LKIARTILIPKQE
					CATCAACGAGGAAGAATGGTTGACATCCAC	TAAAGG	AGGT	NPGILDYRPLTI
					TCTGAAGGAAGCAGCAACACTAGCACTACA	GTATGA	ACAG	ASVVTRVFHSIL
					AGGAAAGAAAACTGAAGCATCTGAAAAGC	TAACCC	TCAC	AKKLDNNAQLS
					TTAACGAGTATGCAAGCAAAACGCTGTTCC	AGAGAA	CTCCT	QRQKGFRKCDG
					CTGGATTGAAAATAACTAATCAGACGAGGA	TATCCC	CGTG	VAENIVILETILT
					AACGAGAAAAGAAAATATCAAAAAGAGAA	ATGGGA	GTAT	NSRSEKRPLCM
					ACCAGAAGGCAAGAATACGCTGAAATACA	GATATC	TTGG	AFVDLRKAFDS
					GAAACTCTACAAAAAGAATATTTCAAGCGC	CATGGA	CGGG	VGHESIIRGAKR
					AGCCGAAAAAGCAATCAATGGGAAATGGTC	AAAAGC	CAAT	VGVPPMLLEYIS

					CATAAAACCCGAAGAAGAATACCACAACA	ACCACG	GCTC	SSYQNASTNLFG
					ATAAGGATCTCATTAAAGCATGGAAACCAA	TTAGAC	ACTA	EILNSRRGVRQG
					TACTAGAGGCACCTCCATTCAGTGACTGTA	AATCCG	AATT	DPLSPILFNFVID
					GGCCCATCGAAAACATCAAGGAAATGGACT	ATGGTC	AACT	EALENLNRNIGY
					ACGCTTTAATGGAGATCAGCACGGCAGAAA	TAACTC	GTGA	LLKEEKVSCLAF
					TCTTCCTCGCAATAAGAGCCATGGGGAAGA	GGCTCC	GTAG	ADDIVLIAETKG
					CTGCACCAGGACCTGATGGCATTAAATATT	GAGGGG	CTGA	GLENHIEKLLEK
					CAAAGCTCAAGAAAAATATCCAATCAATGG	CTAACT	GAAC	LNGAGLELNAS
					CAATATTATTCAACACATGCCTACTAACGA	ATCCCA	TGTA	KCATLMVMKN
					GCTTCCTGCCACTCCCATTGAAGATAGCAA	AAGGGC	TGTG	GKEKSTYISTKA
					GGACAATCCTGATACCTAAACAAGAGAATC	TTAA	TATC	IKIKENDIPTMK
					CAGGTATCCTTGACTATCGACCACTAACAA	(SEQ ID	ATGA	ATETYKYLGLQ
					TAGCCTCAGTGGTGACAAGAGTGTTCCACA	NO: 1225)	AAAA	MGFKAREQNAN
					GTATCCTTGCGAAGAAGCTCGACAACAATG		AAAA	EVITEGLENITR
					CCCAATTAAGTCAACGACAAAAAGGATTTC		(SEQ	APLKPQQRIHIL
					GAAAATGTGATGGAGTTGCGGAAAATATAG		ID NO:	RDFLIPRLIHKLV
					TAATACTCGAAACTATATTAACCAACAGCA		1348)	LGRVAKKSLKRI
					GAAGTGAAAAGAGACCGCTCTGTATGGCCT			DQNIRKKVRNW
					TCGTTGACTTAAGAAAAGCATTCGATTCTGT			LHLPKDTTAAFI
					GGGACATGAGTCTATCATCAGAGGAGCAAA			HADAGDGGLGV
					AAGAGTTGGAGTGCCACCAATGTTGCTCGA			PALEHTIPLLKR
					GTACATTTCKTCAAGTTACCAGAATGCGTCT			ERITNLRKSNDP
					ACTAACTTGTTCGGCGAAATACTCAACTCG			VTKECLRMEYT

					AGAAGAGGAGTCAGGCAAGGCGACCCTCTG			KQVLGKWSRPT
					AGCCCTATTCTCTTCAATTTTGTTATCGATG			KIGETLATNKSQ
					AAGCTCTAGAAAACCTCAACAGGAATATTG			LKEAFRKQMLIT
					GATATCTACTGAAGGAAGAAAAAGTGAGTT			LDGKGLKDHHE
					GCCTAGCTTTCGCGGACGACATAGTCCTGA			TPTIHKWIRRGE
					TAGCTGAGACAAAAGGAGGCCTAGAGAATC			NMTGKQFITAV
					ATATCGAGAAACTATTAGAGAAGCTGAATG			KIRGNLVATKSR
					GGGCCGGTCTCGAGTTGAACGCCTCGAAAT			NSRGRPEQEKLC
					GCGCAACACTGATGGTGATGAAAAACGGAA			EAQCGRPDSLG
					AGGAAAAATCAACGTATATATCAACAAAAG			HILQGCWRTHG
					CAATCAAAATCAAAGAAAATGACATTCCAA			MRVERHNNICR
					CAATGAAAGCCACAGAAACGTACAAATATC			RIKAIMKGKESE
					TCGGATTGCAAATGGGTTTCAAAGCTAGAG			VVEEPRLQTNE
					AACAGAATGCTAATGAGGTTATTACAGAAG			GLRKPDLLICHK
					GACTGGAGAATATAACAAGAGCACCACTGA			GKIIICDAQVVA
					AGCCCCAGCAGAGGATACATATCCTACGAG			DSSNCSLESENQ
					ACTTCCTTATACCAAGATTAATACACAAATT			RKIDYYKKDSV
					GGTATTAGGAAGAGTGGCCAAGAAGTCATT			VSEARKLIGRVD
					AAAAAGAATCGACCAGAATATAAGAAAGA			EDIIIMAVTFNW
					AAGTGAGGAATTGGCTACATCTCCCTAAAG			RGAISKTSIRDL
					ACACGACAGCAGCATTCATACACGCTGATG			DMLLDIKSKEVI
					CAGGAGATGGAGGGCTTGGAGTACCAGCGT			KMSRKIIRDNSI
					TAGAACACACAATTCCTCTACTGAAAAGAG			MVEMHRNRTEK

					AAAGAATAACTAATCTAAGAAAATCCAATG ATCCAGTTACCAAAGAATGCCTGAGAATGG AGTACACCAAACAAGTACTGGGAAAATGGA GTAGACCAACTAAAATTGGAGAAACTCTGG CTACCAACAAAAGCCAACTAAAAGAAGCAT TCAGAAAACAGATGTTAATAACGCTAGATG GAAAAGGGCTAAAAGATCACCACGAAACG CCCACTATCCACAAATGGATCAGAAGAGGA GAGAACATGACCGGCAAACAGTTTATCACA GCAGTTAAAATAAGAGGAAACCTTGTGGCA ACTAAGTCAAGAAATAGCAGAGGGAGACC CGAACAAGAGAAACTATGTGAAGCTCAATG TGGACGACCAGATAGCCTGGGACACATATT ACAAGGTTGCTGGAGAACACATGGTATGCG AGTGGAAAGGCATAACAACATTTGTCGCAG AATTAAAGCAATAATGAAAGGAAAGGAAA GCGAAGTAGTCGAAGAACCAAGACTACAA ACGAATGAAGGTCTCAGAAAGCCTGACTTA CTGATCTGTCACAAAGGTAAAATAATAATA TGCGATGCACAAGTAGTGGCAGATAGCTCG AACTGCAGTCTTGAAAGCGAAAACCAAAGA AAGATAGATTACTACAAGAAAGATTCAGTA GTATCAGAAGCAAGAAAACTTATCGGACGT			RR (SEQ ID NO: 1470)

					GTCGACGAAGATATAATTATAATGGCAGTG ACCTTTAACTGGAGAGGAGCCATCTCAAAA ACATCAATAAGAGATTTGGACATGCTTCTA GATATAAAATCAAAAGAAGTAATTAAAATG TCAAGGAAAATAATCAGAGATAATAGCATC ATGGTGGAGATGCACAGAAACCGGACTGAG AAAAGGAGATAGAAGGGAAACAAAGGAAA AACGAAATGACTGGAAACTATGAAGGATAT AGCTGAAAGCCGCAAGGAAGGCTAAGTCCT GAAACCGATCTACATCTTCGATCCCAAGAG GAACTGTGGGTTAAGCTTGAGCCGACGGAA AAAGCGAATGCATGTTAGACGACGAGGTAC AGTCACCTCCTCGTGGTATTTGGCGGGCAAT GCTCACTAAATTAACTGTGAGTAGCTGAGA ACTGTATGTGTATCATGAAAAAAAAA (SEQ ID NO: 1102)			
R2	R2- 2_TCas	.		Tribolium castaneum	TGGAAGACCCCGCCCATGAGGCTTGGAGAG TGTGATCCTGATCACACTTGAAAAGTTATGC TGAGTACGTCGTGAGAGTCGGTAACTGTCC CAGGATGGTCTGGGATAGGCTAAACCTCAG CAGGGGAAAGTTGTAGGGGCCTGCCACCCC TACACTTTTATAGATATGGCATTCGATACCT CAAATAGAGCCTCGGACTTGGAGGAGCATG	TGGAAG ACCCCG CCCATG AGGCTT GGAGAG TGTGAT CCTGAT	TGAT GCTC CTTTG GTTTT ACCA TCTGT GGGG	MKSRSFRRIGDC AAGSSRRGVRL TGKAGREGRFA ASPHLSPRYLAG SVSGNVPSVPPG PGLGAGAPAFA AGRNADGGPAQ

					GTTCCCCTCCTCCTCGTACTAGACCTGGAAC	CACACT	GCAT	NPCPYCARSFTT
					CAACGGTCTTGACAACCCCATTGGACCTAC	TGAAAA	CGGT	ANGRGLHIRRA
					GGGAGCGGACCATGCTCATGGACATGGATT	GTTATG	CCTC	HPDEANNAIDIE
					CCGAAGACGAAGCGGGGGAACACGGACCC	CTGAGT	ACGG	RIHARWSHEET
					CCCGCCGATAATGCTCACTTAACGTCAGGC	ACGTCG	TTTCC	AMMARLEAGAI
					GAACCCATCGAAATCATCTTGATGTTACCCT	TGAGAG	TCGG	QRGGVRFMNQF
					TTCAAAGCAGGTCATGCGGCATATGTCTCA	TCGGTA	GTTTC	LVPRMPGRTLE
					ATGCCGGAAAGGGTGGCTCCCGGGCGTAAC	ACTGTC	CTATT	AVKSKRRDATY
					GGACATCTATCGCTGATGGAASCTAGCTCG	CCAGGA	GTTTT	KALVQRFLQAP
					CTCTTGAAGAACGGCGGACGGGGACTTGGA	TGGTCT	TCCT	QINLPELRDGDA
					ATCGTGGTGTGGTTCTGATGTAAGTCCTGAA	GGGATA	AAAC	PRQPDPQQENPP
					ATTATGGCGTGATGGCCCGCCCGCCCGACC	GGCTAA	CCGA	EPPSFDGAIRGA
					GGAGGGACTTAGAACCCCCTTCCGCGAGGG	ACCTCA	CAAG	VADLVGGVDW
					TCCTGTCTGTAGGTCCWCCATCTCCGTAAA	GCAGGG	GAGC	QRLGFQGDRLC
					ACGAGTTGGAGGAAACCGCAGACGGGGAT	GAAAGT	CCTTT	NIARRACDGGD
					GAAGAGTCGTTCGTTCCGTCGCATCGGGGA	TGTAGG	GGCC	VSGQLLGWLRD
					TTGCGCAGCTGGGTCAAGTCGGCGGGGTGT	GGCCTG	CTCCT	VFPVKRVSTRG
					TCGGCTAACGGGCAAGGCTGGGCGTGAGGG	CCACCC	CCTT	DQSDLDVDGAL
					GAGGTTTGCAGCTTCCCCTCACTTAAGCCCA	CTACAC	AAAC	VSRRTARTREY
					CGTTATCTTGCAGGCTCGGTTTCAGGTAATG	TTTTAT	ACCT	ARVQELYRKEP
					TTCCTTCAGTTCCCCCCGGCCCCGGCCTGGG	AGATAT	CTCCT	KACLARILGDRR
					GGCCGGGGCTCCTGCATTTGCTGCTGGTCG	GGCATT	TCAT	EGANRAPNRDP
					GAATGCCGATGGGGGGCCCGCCCAAAATCC	CGATAC	CCTST	AFIDFWRGVFSE

					GTGCCCGTATTGTGCACGGTCGTTTACTACC	CTCAAA	TAGT	ASAEVEGWAEE
					GCCAATGGCCGCGGGTTGCACATACGGAGG	TAGAGC	CCAT	VSDHGELARRV
					GCGCATCCAGACGAGGCTAATAACGCCATA	CTCGGA	TCCG	WDPISVEEVGRS
					GACATCGAACGCATTCATGCCCGCTGGTCA	CTTGGA	GCTA	RVRNGAAPGPD
					CACGAGGAGACAGCCATGATGGCCAGGCTT	GGAGCA	AAAT	GIAVSVWNKLP
					GAGGCCGGTGCCATCCAGCGAGGGGGTGTA	TGGTTC	GATG	PEAAALLFNVLL
					AGGTTTATGAACCAGTTTCTGGTTCCGCGCA	CCCTCC	AAGA	LGRCLPAELTRT
					TGCCGGGCAGGACTCTGGAAGCGGTGAAAA	TCCTCG	CCGA	RTVFIPKTDAPR
					GTAAGCGCAGAGATGCGACTTACAAAGCAC	TACTAG	GGAG	TPADYRPISIASV
					TGGTCCAGAGGTTTCTGCAGGCGCCGCAAA	ACCTGG	TGTC	VARHFHRVLSA
					TAAACCTCCCCGAACTGCGTGATGGTGATG	AACCAA	ACTC	RVQRIPDLFTKY
					CGCCTCGTCAGCCTGATCCTCAACAGGAAA	CGGTCT	TCTTG	QRGFLSGVDGIA
					ACCCTCCCGAGCCCCCTTCTTTCGATGGGGC	TGACAA	GCGG	DNLSVLDTMLT
					GATCAGGGGCGCTGTCGCCGACCTTGTCGG	CCCCAT	GGTT	MSRRCCKHLHL
					AGGGGTCGACTGGCAGCGCCTCGGGTTTCA	TGGACC	AACC	AALDVSKAFDT
					GGGAGACCGTCTCTGTAACATCGCGAGACG	TACGGG	CGTC	VSHFAIVRACRS
					GGCTTGCGATGGGGGGGATGTTTCGGGCCA	AGCGGA	CAAG	IFGSAETVLEEG
					GTTGCTGGGCTGGCTGCGGGATGTCTTCCCC	CCATGC	TGTA	GRRHFVQVRXG
					GTCAAGAGGGTTAGTACTAGAGGGGATCAG	TCATGG	AATG	VRQXDPLSPLLF
					AGCGACCTAGATGTCGACGGCGCATTGGTG	ACATGG	TGAC	NLVLDRALKRL
					AGCAGGCGCACGGCGAGGACAAGGGAGTA	ATTCCG	CTCG	STDVGFRLTDAT
					CGCTCGGGTGCAGGAGCTGTACCGGAAGGA	AAGACG	CCAT	KVTALAFADDV
					ACCTAAGGCGTGTTTGGCTCGAATCCTTGGT	AAGCGG	TCGG	VLCATTARGLQ

					GATCGTCGAGAGGGGGCCAATCGGGCTCCT	GGGAAC	GCTC	TNLDVLEAELRL
					AATCGCGATCCCGCATTCATCGATTTTTGGC	ACGGAC	TGAT	AGLLLNPNKCQ
					GAGGCGTCTTTTCGGAGGCGTCCGCTGAGG	CCCCCG	A (SEQ	ALSLVASGRDH
					TCGAGGGATGGGCGGAGGAGGTCTCCGACC	CCGATA	ID NO:	KVKLVTKPTFK
					ATGGTGAGCTGGCACGCCGGGTTTGGGACC	ATGCTC	1349)	VGQNTIHQVDA
					CCATCTCTGTCGAGGAGGTGGGAAGGTCTC	ACTTAA		SSIWKYLGIQFR
					GGGTCCGGAATGGTGCAGCGCCGGGGCCCG	CGTCAG		GSGMCGCGSEG
					ACGGAATAGCGGTTTCCGTGTGGAACAAGC	GCGAAC		VAAGLKRITCAP
					TTCCGCCGGAGGCCGCTGCCCTTCTTTTCAA	CCATCG		LKPQQRMHLLR
					TGTGCTGCTGTTGGGGAGATGTCTGCCTGCA	AAATCA		VFFLPKFYHAW
					GAACTCACCCGCACGCGAACGGTCTTTATT	TCTTGA		TFGRLNAGVLR
					CCAAAGACCGACGCACCGCGAACACCTGCA	TGTTAC		RLDVVVRTSVR
					GACTATCGGCCGATTTCGATCGCGTCGGTA	CCTTTC		TWLRLPHDIPVG
					GTCGCACGGCACTTCCACCGTGTCTTATCCG	AAAGCA		YFHAPTKSGGL
					CACGTGTTCAACGCATTCCCGACTTGTTCAC	GGTCAT		GIPQLSRFIPFLR
					AAAGTATCAAAGGGGCTTCTTAAGCGGCGT	GCGGCA		LKRFDRLGRSA
					GGACGGGATCGCAGATAATCTCTCGGTCCT	TATGTC		VDYVRECAFTDI
					TGACACAATGCTGACGATGTCCCGGAGATG	TCAATG		ADRKIRWCRER
					CTGCAAGCATTTGCACTTGGCCGCGTTAGA	CCGGAA		LSGIVDQVAGG
					CGTGTCAAAAGCATTTGACACTGTCTCTCAC	AGGGTG		RDALDAYWTA
					TTTGCCATAGTGCGTGCTTGCCGGTCCATAT	GCTCCC		QLHQSVDGRAL
					TTGGGAGTGCGGAGACCGTGTTGGAGGAGG	GGGCGT		RESASVASSTQ
					GCGGGAGGAGGCACTTCGTGCAAGTGCGGY	AACGGA		WLRCSTRAIPAS

					GTGGGGTGAGGCAGKGAGACCCACTCTCTC	CATCTA		DWLHYTAVHIG
					CTCTCCTCTTTAACCTGGTCTTGGACCGGGC	TCGCTG		ALPSRVRTSRGR
					TCTGAAGAGGCTGTCCACCGATGTGGGCTT	ATGGAA		RGGQDVSCRGG
					TCGTCTAACGGATGCGACCAAAGTTACGGC	SCTAGC		CLLDETPAHCIQ
					ACTCGCTTTTGCGGACGACGTCGTTCTGTGC	TCGCTC		VCHRTHGGRVL
					GCGACCACCGCTAGAGGACTCCAGACCAAT	TTGAAG		RHDAIAKRISAD
					CTGGATGTTCTCGAAGCGGAGCTTCGCCTA	AACGGC		LMELGWIVTRE
					GCCGGTCTCTTGTTGAACCCCAACAAGTGTC	GGACGG		VSFRTTAGVFRP
					AGGCCTTATCCTTGGTGGCTTCCGGACGGG	GGACTT		DMVAVKEGVT
					ACCACAAAGTGAAGCTCGTGACAAAACCCA	GGAATC		VILDVQIVSPAP
					CCTTTAAGGTGGGACAAAACACGATTCACC	GTGGTG		TLDEAHRRKVA
					AAGTGGATGCGTCCAGTATTTGGAAGTACC	TGGTTC		KYRDRADLARY
					TGGGAATCCAGTTTAGGGGTAGTGGGATGT	TGATGT		LAEAAVARGRA
					GCGGATGCGGTTCGGAGGGTGTGGCTGCTG	AAGTCC		PPANIRFASATIS
					GGTTGAAGCGGATCACTTGTGCTCCGCTGA	TGAAAT		WRGVWSAESV
					AGCCTCAGCAGAGGATGCACCTGCTGCGAG	TATGGC		GSLRELGLSARH
					TCTTTTTCTTGCCCAAGTTTTATCATGCTTGG	GTGATG		FDRYTTMALCG
					ACATTTGGGAGACTCAATGCGGGTGTACTG	GCCCGC		SWRNWVRFNAS
					CGCCGTCTGGACGTTGTGGTCCGAACTTCG	CCGCCC		TASRMGRGRGD
					GTTCGGACCTGGCTACGCCTCCCCCACGAT	GACCGG		ASPRRHENQQ
					ATACCGGTGGGATACTTTCACGCTCCTACG	AGGGAC		(SEQ ID NO:
					AAGTCCGGGGGGCTGGGGATCCCGCAGTTG	TTAGAA		1471)
					TCACGATTCATCCCGTTCTTGCGATTAAAGC	CCCCCT		

					GGTTCGACCGGTTGGGGCGTTCCGCGGTTG ACTATGTCCGAGAGTGTGCATTTACGGACA TTGCGGATCGAAAGATCCGATGGTGCCGTG AGCGCCTCTCGGGCATAGTGGACCAAGTCG CGGGTGGGCGGGACGCGCTTGATGCATATT GGACAGCGCAGCTTCATCAGTCGGTGGACG GAAGAGCACTAAGGGAGTCCGCTTCGGTGG CTTCCAGCACGCAGTGGCTCCGCTGTAGCA CACGTGCAATTCCGGCCTCTGACTGGCTGC ATTACACGGCTGTCCATATTGGCGCCCTACC ATCGCGTGTTCGCACATCGCGTGGGAGGCG GGGCGGCCAGGATGTCTCTTGTCGGGGTGG ATGTCTTCTGGACGAAACTCCGGCACACTG CATCCAGGTTTGCCACCGCACCCACGGGGG GCGGGTGTTGCGACACGACGCCATCGCCAA GCGCATTTCTGCCGACCTGATGGAACTTGG ATGGATCGTAACGAGGGAGGTTTCCTTCCG GACGACTGCGGGTGTTTTCCGGCCGGACAT GGTCGCGGTAAAGGAGGGTGTTACCGTGAT CTTGGACGTGCAGATTGTTTCACCGGCCCCG ACACTGGATGAGGCCCATCGGCGAAAGGTT GCCAAGTATCGGGACAGGGCTGACCTGGCC AGGTACTTGGCTGAGGCTGCGGTTGCTCGT	TCCGCG AGGGTC CTGTCT GTAGGT CCWCCA TCTCCG TAAAAC GAGTTG GAGGAA ACCGCA GACGGG G (SEQ ID NO: 1226)		

					GGGCGTGCGCCGCCTGCGAATATACGATTT GCCAGCGCGACCATCTCTTGGCGAGGCGTT TGGAGTGCTGAGTCGGTGGGGTCCCTGCGC GAGCTGGGGCTCTCCGCTCGGCACTTCGAT CGCTATACAACGATGGCACTCTGCGGTTCA TGGAGAAATTGGGTGCGGTTCAATGCGTCC ACGGCCTCTCGGATGGGTCGTGGACGCGGC GACGCGAGTCCTCGGAGGCACGAAAATCAG CAGTGATGCTCCTTTGGTTTTACCATCTGTG GGGGCATCGGTCCTCACGGTTTCCTCGGGTT TCCTATTGTTTTTCCTAAACCCGACAAGGAG CCCTTTGGCCCTCCTCCTTAAACACCTCTCC TTCATCCTSTTAGTCCATTCCGGCTAAAATG ATGAAGACCGAGGAGTGTCACTCTCTTGGC GGGGTTAACCCGTCCAAGTGTAAATGTGAC CTCGCCATTCGGGCTCTGATA (SEQ ID NO: 1103)			
R2	R2-7_MR	.		Megachile rotundata	TCTAGTTAGCAAGCGGCCCCCTCTAATCGG GTCCCGCCACGAGTACCTTCGGGGAAACCA AGAGCCGCCTCTGCGAACCTACCTCCGCAT TGGGTTGCCGGCCCGGTGCGGTTGTAATTC AGTGGGCCCAAATTCACAAGGAAAAGAGA AAAAGAATTGTCGGTTGGCCCTTAGGGCAT	TCTAGT TAGCAA GCGGCC CCCTCT AA (SEQ	TAAA TGTC GAAC CGAA TTTTG GGTA	SGPATSTFGETK SRLCEPTSALGC RPGAVVIQWAQ IHKEKRKRIVG WPLGHLGSPTSL KLRHPRLQAKRI

					CTCGGATCACCGACCTCACTGAAACTTCGA	ID NO:	ACGT	VPVLAELMQCL
					CACCCGCGCCTGCAAGCGAAGCGCATAGTT	1227)	GCAC	CARHVSGRSPQ
					CCAGTGCTTGCGGAGTTGATGCAATGTTTGT		CCCA	KSAWAFTLEER
					GTGCTCGACATGTCAGCGGCCGTTCGCCAC		CCAT	TSACGVDAMFV
					AAAAATCGGCTTGGGCGTTCACGTTAGAAG		CCTT	CSTCQRSFATKI
					AGCGCACGAGTGCTTGCGGAGTTGATGCAA		AATC	GLGVHVRRAHV
					TGTTTGTGTGCTCGACATGTCAAAGGTCGTT		GGCA	EVANAAISVERV
					CGCCACAAAAATCGGCTTGGGCGTTCACGT		GCAC	KDRWSEEERRI
					TAGAAGAGCGCACGTGGAAGTCGCAAACGC		GCAA	MAAVEVRGVLS
					TGCCATATCAGTGGAGCGCGTTAAGGATCG		TAAA	GARFINEYIMSH
					TTGGTCGGAGGAGGAGCGAAGAATAATGGC		GCCG	LQTSRTLESVKG
					TGCCGTCGAGGTTCGAGGAGTACTCTCTGG		TGGG	TRKNPKYKELV
					GGCTCGTTTTATAAACGAGTACATTATGAGT		CAGT	ATLLEEARTSVR
					CATCTCCAGACCAGCCGGACCCTCGAATCT		GGTT	EESPRSAVNDSA
					GTCAAAGGAACTCGGAAAAACCCCAAATAC		TTAG	TQPSGPSDTRSL
					AAGGAACTTGTCGCAACCTTACTGGAAGAG		TGGG	RTEHLFTESTEP
					GCACGAACCAGTGTGCGGGAGGAGAGTCCT		TAGT	FEHRIRELIGDLE
					CGTAGTGCGGTCAACGATAGCGCCACCCAA		CATT	GVTDFRAELLVS
					CCATCAGGCCCCTCTGATACACGAAGCCTG		AGGA	IAEQQLQGDEV
					CGCACCGAACACCTGTTCACTGAAAGCACC		GTCC	AESLTRWLGEV
					GAGCCTTTCGAACATCGGATCCGAGAGTTG		CACA	FKPENQQQQVQ
					ATCGGGGATCTTGAGGGAGTGACCGACTTC		GTAC	RKRRRQRKAPV
					AGAGCTGAGTTGCTGGTTTCCATTGCGGAA		CCAG	SGQLPKWRERR

					CAGCAGCTTCAAGGCGATGAGGTGGCGGAA		CGAA	RDYAAMQTLFH
					TCCCTGACTCGGTGGCTTGGCGAGGTTTTTA		CATC	RNPSLAAGRVL
					AACCCGAGAACCAACAACAACAGGTTCAGC		TTAG	DGKNESRPPDLP
					GTAAGCGTCGGCGACAAAGAAAGGCCCCTG		TGGG	EMTAFWEPILTE
					TCAGTGGGCAGCTGCCAAAGTGGCGCGAGA		TCTG	QSAEHRAVGPA
					GGCGAAGGGATTATGCAGCGATGCAGACTC		CGTA	SEKSELCSVWGP
					TCTTCCATCGGAACCCCAGTCTTGCTGCTGG		AACG	VEKEELLSSVPP
					TCGTGTGCTGGACGGCAAGAACGAGTCCCG		CATTT	LDTAVGPDGVT
					GCCCCCGGATCTGCCAGAAATGACAGCGTT		CCAC	ARQWRAVLPAV
					TTGGGAACCGATACTCACCGAGCAATCTGC		TGCC	RALLYNIILKRG
					CGAACATCGGGCAGTTGGGCCGGCTTCCGA		TATC	SFPASMLESRTV
					GAAATCGGAGTTGTGCTCAGTATGGGGTCC		CTCC	FLPKKQHSVNP
					AGTAGAGAAAGAAGAACTTCTCTCCTCCGT		GGGA	ADFRPISIASVV
					GCCGCCGTTGGACACGGCTGTTGGCCCTGA		AAAA	VRQLHKILAMR
					CGGTGTCACCGCCAGACAATGGAGAGCGGT		AAAA	LRRTNLVDERQ
					TTTACCGGCCGTTAGAGCTCTCTTGTATAAC		AAAA	RCMDDGCAENI
					ATCATCCTCAAGAGAGGCTCTTTCCCCGCCT		AAAA	TVLASLLDDAR
					CGATGCTGGAAAGCAGAACCGTGTTCCTGC		AAAA	HGLKELHLVSL
					CGAAGAAGCAACATAGTGTCAATCCTGCTG		AAAA	DCAKAFDSVSH
					ACTTTCGACCCATCAGCATAGCGTCGGTCGT		AAAA	HAIDATLKECGL
					CGTCCGTCAATTACACAAGATCTTGGCCAT		AAAA	PAGFVQYISRTY
					GCGCCTTCGCCGTACGAACCTCGTTGACGA		AAAA	SDSSTRLEVGRN
					AAGGCAACGATGCATGGATGACGGATGTGC		AAAA	RSEPIKTNRGVR

					CGAGAACATCACGGTGTTGGCCTCGCTGCT GGACGATGCGAGGCATGGTCTGAAGGAGCT GCATCTGGTTTCGTTGGATTGTGCTAAGGCG TTCGACAGCGTTAGTCACCATGCGATCGAC GCAACACTAAAAGAATGCGGTCTGCCCGCG GGATTTGTGCAATACATCTCCCGCACATATT CCGACAGTTCCACCAGGCTCGAGGTCGGAA GGAATCGGTCGGAGCCAATAAAGACCAACC GGGGAGTGCGGCAGGGTGATCCCCTGTCAA CCCTCATCTTCTGTCTGTGCTTCGATAGAGT CGCGCGTACCCTCTCACCACACATTGGGTAT GACCTTAATAACACCCGCATCAGTACTCTG CTTTACGCTGATGATGCTTTCCTTGTCTCGA CGACTGCTCCGGGGATGAACATCCTGCTCC GGAGTGTGGAGGAAAGTGCCGGTGAGGTG GGGCTTTCTTTTAACACCTCGAAATGCTCAG CCCTCTCCTTAATTCCATCTGGAAAGGAAA AGAAAATGAAGGTCGGAACGACTCCGACGT TCAAAACGTCGCAGGGTTTTATTACCCAGA TAACGCCCTCGCAAGAGTGGAGGTATCTAG GAGTCGACTTCCAATACTCCGGACCGAAAA AAGCAAGTCGGAGTCTCAAAATCGAGCTGG AACGCATTTCGAAGGCCCCCCTCAAGCCGC		AAAA (SEQ ID NO: 1350)	QGDPLSTLIFCL CFDRVARTLSPH IGYDLNNTRIST LLYADDAFLVS TTAPGMNILLRS VEESAGEVGLSF NTSKCSALSLIPS GKEKKMKVGTT PTFKTSQGFITQI TPSQEWRYLGV DFQYSGPKKAS RSLKIELERISKA PLKPQQRLLILR VYLLPRYYHHL VLSRTTLGHLRG LDLQVRAAVRR WLSLPRDIPIAY FHTTAKEGGLG LPAFETSIPCLM LARLRSMETSTC KAARAAVQGF WVQKRIHWATA ALTKNGEALTC

					AACAGCGGCTTTTAATCCTCAGAGTCTACCT			KADVDRWWAS
					GCTGCCGCGATATTACCATCATCTCGTACTC			RLHKSVDGREL
					TCCAGGACAACCCTTGGTCACCTTAGGGGC			RECSGVGSSST
					CTGGACCTGCAAGTGCGCGCGGCTGTTCGT			WVNSALNITGR
					AGGTGGCTGTCTCTGCCGCGCGACATCCCC			DYVQYHHVRIN
					ATTGCCTACTTCCACACAACAGCGAAGGAA			SLPTRIRTSRGV
					GGGGGTCTGGGTCTTCCGGCGTTTGAGACG			RREGMEVTCRA
					TCCATTCCATGCCTGATGTTAGCCCGCTTGA			GCQVTETAAHV
					GGTCGATGGAAACCTCGACTTGCAAGGCGG			IQSCHRTHGGRI
					CTAGGGCGGCGGTTCAAGGCTTCTGGGTGC			LRHNAVCKVLA
					AGAAACGCATCCATTGGGCTACGGCCGCTC			SGLRDKGWEVR
					TCACTAAGAACGGAGAAGCCCTAACATGCA			EEPKLRTRQGLR
					AGGCGGACGTCGATCGGTGGTGGGCCAGTA			KPDIVAIKDGVA
					GGCTCCACAAGTCAGTGGATGGTCGCGAAC			RVIDAQVVSGS
					TGCGTGAGTGCTCAGGCGTTGGCTCGTCCA			GPLDEAHETKR
					GCACCTGGGTAAATTCTGCGCTTAACATCA			KYYSDNGDVTA
					CCGGCAGAGATTACGTGCAGTATCATCACG			AIARECNIAPSN
					TCCGCATCAATTCTCTCCCGACTCGCATTAG			VAYSSCTISWRG
					GACATCTCGGGGTGTGCGAAGAGAGGGGAT			VWSPRSAADLL
					GGAGGTCACGTGCCGCGCTGGGTGTCAGGT			QVGLSKKLLGFI
					AACTGAGACCGCGGCGCATGTCATCCAGTC			TLRVLRGSHLN
					TTGCCACCGCACACATGGAGGTCGGATTCT			WTRWNKMTTM
					TCGACACAATGCCGTGTGCAAAGTGCTGGC			RVHHQRTGIG

					GTCCGGCCTTCGCGACAAGGGATGGGAGGT GCGTGAAGAGCCAAAACTGCGAACGCGAC AGGGTCTGAGGAAGCCGGACATAGTGGCGA TAAAGGACGGCGTCGCAAGGGTGATAGACG CACAGGTCGTGAGTGGATCTGGACCGCTTG ACGAGGCGCATGAGACCAAGCGCAAGTATT ATTCCGACAATGGCGACGTAACTGCTGCGA TTGCTCGTGAGTGCAACATTGCTCCATCGAA TGTTGCATACTCGTCATGCACTATCTCTTGG AGGGGTGTGTGGTCGCCTCGGTCCGCTGCA GACCTACTACAGGTGGGTCTGTCCAAAAAG CTCTTGGGTTTCATCACCCTAAGGGTCTTGA GAGGCAGTCATCTAAATTGGACACGGTGGA ACAAAATGACAACGATGCGCGTACACCATC AGCGAACAGGTATCGGGTAAATGTCGAACC GAATTTTGGGTAACGTGCACCCCACCATCCT TAATCGGCAGCACGCAATAAAGCCGTGGGC AGTGGTTTTAGTGGGTAGTCATTAGGAGTC CCACAGTACCCAGCGAACATCTTAGTGGGT CTGCGTAAACGCATTTCCACTGCCTATCCTC CGGGAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1104)			(SEQ ID NO: 1472)

R2	R2Amel	.		Apis mellifera	TGGTAATCAAATGCCTCGCTATTTTAGTAGC	TGGTAA	TGAT	MSSNEEGASDT
					GGTAGCGCTCCGCCCGCGCAGGAACCATTG	TCAAAT	CGTT	GAPGPGVPVAD
					ACGCCGCCGTAGTGTGGGTGATTTTATATCC	GCCTCG	AAAA	VSAADGRATYD
					AACCAATCACGTCAACTACGATCATTTGTA	CTATTT	GTAA	DHGMSTDYEKQ
					ATCACCGACGGTACTTGGTAGGGGTACCAC	TAGTAG	AAAT	TIELPLNGQIQCL
					ATGGGCATTCTTGCTCATTCCACAACGCCGC	CGGTAG	CTATT	WCHIEGRNQRF
					CTCCATCATGGCAACAATTTAAAATATATAT	CGCTCC	TATTT	LQESQYLKHKD
					AAATTCTTAAGGTTTGACCGTATTCATATAT	GCCCGC	ATTTT	TQHPKGEIIWRC
					ATATATATATATTTAATATTACAACCATAAA	GCAGGA	TATTC	AACQKEFEKLH
					TCTTATATCGAGCCTTCTATTTGGTCTCAAA	ACCATT	CTAT	GCRCHLPKCKG
					AGCAATACGTTGTCAGATCTTGTAGAACAT	GACGCC	ATTA	RKEAKGVAKFK
					CAGGAGTGAGCGGTGCGCTGTGGTATCCGT	GCCGTA	TAAC	CDSCEESFLTQR
					GCTTTGTGCCGCGGCGACAAACCAATACGC	GTGTGG	ACAT	GLSMHELHRHP
					TGCTGCCTGTCGCAAAGCAATACGCTGCTG	GTGATT	TATTT	AIRNLKRTQGTS
					ATTCTCGGATGCGGGTGTCGACGGTCACGC	TTATAT	ATTT	RGNTRPINRASV
					AAAGCGATACGCTGGTGGGGTTTCAAAACA	CCAACC	ATTT	WSKEETDLLIKL
					ATACGGCGCTGGTGCTAAAAAGCATTATGC	AATCAC	ACTT	NERYKHLKQPN
					CGCTAACGGCTGGATTGTCGATCGCCGCTG	GTCAAC	ATTG	VALKEYFPDKT
					CGGGGGCTAGTGGCGCACCCAGAGAGGTGC	TACGAT	TTTTA	LKQISDKRRLLP
					GACGCGCAAGCATTGGTTCTGTGCGAAGCG	CATTTG	AAGA	VQEPEDVATTD
					GAGTTCTTGAGAGTAATGGTTGCTGGGGGC	TAATCA	TGAC	ETGPPPSDSSEES
					ACAAAGCGCAACATATAGCCTCTTATGCCT	CCGACG	GAAG	IYESATEDEGGG
					CAAGTCGTAGTTCGTACCTCCACGTGGTCCC	GTACTT	CCGC	DMQQTAPNDS

					GCTGGAATGCCTATCGACTCCTCCCCGGAG	GGTAGG	AAGG	WKEPFIQSIRTN
					GATCATAGAGTTCGAAACCGGCTACGGCGA	GGTACC	CCAA	HLEEEDSLRKVE
					GGCAAGGGCGGTGAGGTGCACACCGATGG	ACATGG	TCCA	EAIERMAMNEG
					GGAGCAGCGACCCCACCTACCCTTAGCTAA	GCATTC	AATT	VTEQEVGTLLE
					GAGAGCAGGCGATCCGCCAACTGTCAGCAC	TTGCTC	TAAC	QFVDSLTQSPTT
					GAAATAAACTAATCATATGTATACGAGGGA	ATTCCA	AAAA	ERKGSRRKSQK
					GAATTTACAACGGGTACCTTGTGCCCGAAC	CAACGC	GAAC	TTKRKTTHNNR
					CGCCTGTAGGTATCACCTACAGGTGTTAAA	CGCCTC	GAGA	KKFLYAKHQEL
					ATGAATCTGATAGCTGGCGGATCGTCGACC	CATCAT	CTAC	YKKSPRRLLELA
					CTCTTTGATGGCTCTGCGCCAACGACTGGA	GGCAAC	TGGT	LSGESSSGREVV
					AAGAATAGGAACGGAAGTCTAATGGAAGG	AATTTA	CGAC	NLPEADSVGPLY
					AAAGTGTCGGGAGCACTATAAATTCCCAAA	AAATAT	ATTA	KSLWGQIGPEKT
					GAAGAAAAGAAAAGAAAAAAAATAAAAAA	ATATAA	AAAA	HRNQPMCNNID
					CCCAAATTAAATGTCGAGTAACGAAGAAGG	ATTCTT	GACG	MSEIWTPIALES
					AGCCTCGGACACGGGGGCCCCCGGCCCTGG	AAGGTT	AAGC	LVEKFKKIKSDT
					GGTCCCCGTGGCGGACGTGTCCGCTGCGGA	TGACCG	AGCT	AAGADQIKKFH
					TGGAAGGGCCACTTATGACGACCATGGCAT	TATTCA	GCCA	LRKKGALHVFA
					GTCCACGGACTACGAGAAACAAACCATCGA	TATATA	GCTG	KLCNLLMLHRI
					ACTGCCCCTGAACGGGCAGATACAATGCTT	TATATA	ATAA	YPAQWKTNRTT
					ATGGTGCCACATTGAAGGAAGGAACCAAAG	TATATT	ACAA	LIPKPGKSAEEV
					ATTCCTCCAGGAAAGCCAATACTTAAAACA	TAATAT	CAGA	ENWRPITIGSLL
					CAAAGACACGCAACACCCTAAAGGCGAAAT	TACAAC	GCCC	GRIYSAMIDRKL
					AATATGGAGGTGCGCTGCATGCCAGAAAGA	CATAAA	GTCT	RSKIKQHIRQKG

					ATTCGAGAAACTCCATGGCTGCAGGTGTCA	TCTTAT	CGGC	FTQEDGCKNNIA
					CTTACCCAAGTGTAAGGGACGCAAAGAAGC	ATCGAG	CTTTA	ILSSALTKMKED
					CAAGGGTGTTGCCAAATTTAAATGCGACTC	CCTTCT	CACC	SGGIITIIDISKAF
					CTGCGAGGAATCGTTCCTAACGCAAAGAGG	ATTTGG	GAGC	DTVPHGEISQSL
					ATTGTCTATGCACGAGCTACATAGACATCC	TCTCAA	GGTG	MNKGVPSPICEY
					AGCGATTAGGAACTTGAAAAGAACACAAG	AAGCAA	CAAG	IQKMYIGCKTIIY
					GCACCAGTCGGGGAAATACCAGACCAATCA	TACGTT	TCCT	CRDKKTLPVDIL
					ACAGAGCCTCGGTCTGGTCCAAAGAGGAAA	GTCAGA	GACG	RGVKQGDPLSP
					CGGACCTCTTGATAAAACTTAACGAGCGCT	TCTTGT	TACT	LLFNLIIDPIIGTL
					ACAAACACTTAAAACAGCCGAACGTAGCGC	AGAACA	ATTG	DETTEGIKLENE
					TCAAAGAATATTTCCCCGACAAGACACTAA	TCAGGA	TACG	NISVLAFADDLV
					AACAAATCAGCGACAAAAGAAGGCTCTTGC	GTGAGC	TCTA	LLAKDKETADK
					CCGTTCAGGAACCCGAAGACGTGGCCACAA	GGTGCG	GGGC	QNRLINEYLDDL
					CTGATGAAACGGGACCTCCTCCTTCCGACTC	CTGTGG	GCGG	KMKVSAEKCTT
					ATCGGAAGAGAGCATATACGAATCGGCCAC	TATCCG	GGCA	FEIKRQNKTWFL
					GGAGGACGAAGGAGGAGGAGATATGCAAC	TGCTTT	GATT	GDPQLTLGQQRI
					AAACGGCTCCAAACGATAGCTGGAAGGAGC	GTGCCG	CTAC	PYADPEAAIKYL
					CGTTTATACAAAGTATAAGAACAAACCACC	CGGCGA	CGTG	GTNFNPWRGLC
					TCGAAGAGGAAGACTCCCTTCGAAAGGTGG	CAAACC	TAGA	KTSIKEIIDAART
					AAGAAGCCATCGAAAGAATGGCTATGAATG	AATACG	ATCT	VKQLKLKPHQK
					AAGGGGTAACTGAACAAGAGGTGGGCACC	CTGCTG	GGGG	INLIRTYLLPRYI
					CTTCTTGAACAATTTGTCGACTCCCTAACTC	CCTGTC	CGAC	HKLVANPPPLGT
					AATCCCCAACAACGGAAAGAAAGGGGAGC	GCAAAG	GCCT	LDLIDKELKTIIK

					CGACGTAAGAGTCAAAAGACTACAAAAAG	CAATAC	CCGC	EILHLHPSTTDG
					AAAGACCACCCATAACAATAGAAAAAAGTT	GCTGCT	GAGG	LIYTDKSHGGLG
					CTTATATGCCAAACACCAGGAGCTCTATAA	GATTCT	CACT	IQRVANIVKLAK
					AAAGAGCCCACGAAGGCTTCTGGAGTTGGC	CGGATG	CCCT	LKHSILMTRSED
					GTTATCGGGTGAGTCTAGCAGTGGCAGAGA	CGGGTG	GGAC	NAVKIALNGQE
					AGTGGTTAATCTCCCTGAGGCCGACTCAGT	TCGACG	AACG	GMVKRYATSIG
					GGGTCCACTATATAAAAGTCTATGGGGCCA	GTCACG	TACG	LQWPCGIEEIEE
					AATAGGCCCGGAAAAAACTCACAGAAACC	CAAAGC	CTAA	TRKKLKRADTN
					AACCTATGTGCAACAATATCGATATGAGCG	GATACG	AGCG	KWKTLISQGQGI
					AAATTTGGACTCCAATCGCCCTGGAGAGCC	CTGGTG	TACG	KEFFGDKTGNA
					TTGTCGAAAAATTCAAAAAGATAAAGTCCG	GGGTTT	GCTA	WLYNPEMLRPS
					ACACCGCAGCCGGCGCGGACCAGATAAAG	CAAAAC	AGTG	RYLDALKLRTN
					AAATTCCACCTGAGAAAGAAAGGGGCACTA	AATACG	CGCC	TYGTKAALHRA
					CACGTATTCGCCAAACTGTGTAACCTCCTCA	GCGCTG	TCCC	KRDIDINCRRCG
					TGCTGCACCGAATATACCCAGCACAGTGGA	GTGCTA	GAAA	VQVETLGHILGL
					AAACCAACCGAACCACGCTTATTCCCAAAC	AAAAGC	GGGT	CTHTKNKRIKR
					CGGGGAAGAGCGCGGAAGAGGTTGAGAAC	ATTATG	CCCC	HDEICDLIAKNV
					TGGAGACCAATCACCATCGGGTCTCTGCTG	CCGCTA	GTTC	SKEYVIFREPEV
					GGAAGAATATATTCGGCTATGATCGACCGT	ACGGCT	CTAA	EVNGDRRKPDM
					AAATTACGGTCGAAAATAAAGCAGCACATA	GGATTG	TTTTT	VIKDHDKVYVV
					AGACAGAAGGGGTTTACACAGGAGGATGG	TCGATC	CCGA	DVTVRYENNDS
					CTGTAAAAATAATATAGCCATTCTCAGTAG	GCCGCT	GCCC	LNKAYKEKENK
					TGCCTTAACCAAAATGAAAGAGGACTCAGG	GCGGGG	GCGG	YKETAEIMRRD

					TGGAATCATAACCATAATAGACATTTCCAA	GCTAGT	GCAG	LKAKESRVLPV
					AGCCTTCGACACGGTTCCCCACGGCGAAAT	GGCGCA	ATCT	VIGSRGAVPRAT
					AAGCCAAAGTCTGATGAACAAAGGAGTCCC	CCCAGA	CGTG	IENLKVLGLQTK
					ATCGCCCATATGCGAATACATTCAAAAAAT	GAGGTG	GCAG	HALTASLIALRS
					GTACATAGGTTGTAAAACTATTATATATTGC	CGACGC	TGAC	SIEMANEFLDYD
					AGAGACAAGAAAACACTGCCAGTGGACAT	GCAAGC	GCTA	HTT (SEQ ID NO:
					ACTGAGAGGAGTCAAACAGGGAGACCCGCT	ATTGGT	GAAA	1473)
					ATCGCCACTGCTTTTCAACTTGATAATAGAT	TCTGTG	GTTA	
					CCCATAATAGGGACACTGGACGAGACGACG	CGAAGC	AGTC	
					GAGGGCATTAAATTAGAAAACGAGAACATT	GGAGTT	CGCG	
					TCAGTTCTCGCCTTCGCCGACGACCTTGTCC	CTTGAG	GACA	
					TTTTGGCGAAAGACAAAGAAACAGCCGATA	AGTAAT	TATA	
					AGCAAAATCGGCTCATCAATGAATATCTGG	GGTTGC	AAAT	
					ACGACCTGAAAATGAAAGTATCCGCCGAAA	TGGGGG	TACA	
					AATGCACAACCTTCGAAATCAAACGGCAGA	CACAAA	GCCT	
					ACAAAACGTGGTTCCTAGGAGACCCACAGT	GCGCAA	TAAA	
					TGACGTTGGGTCAGCAACGTATCCCGTATG	CATATA	TAAT	
					CCGACCCAGAAGCAGCAATCAAATACCTAG	GCCTCT	GAAC	
					GAACCAACTTCAATCCATGGAGAGGGTTGT	TATGCC	CCCA	
					GCAAAACCTCGATAAAAGAAATCATCGATG	TCAAGT	CGAA	
					CGGCTAGAACTGTCAAACAGCTGAAACTTA	CGTAGT	GGAG	
					AGCCGCATCAAAAAATCAACCTTATAAGAA	TCGTAC	GTAT	
					CCTACCTCTTGCCAAGATACATACATAAATT	CTCCAC	CCTC	

					GGTGGCAAATCCTCCCCCTCTGGGGACTCT AGACCTAATCGATAAAGAGCTCAAAACTAT AATAAAGGAAATATTGCACCTCCATCCGTC CACCACGGACGGACTAATATACACTGATAA GAGCCATGGCGGTCTAGGGATCCAGCGGGT GGCGAACATAGTCAAGCTGGCCAAACTAAA ACATAGTATACTAATGACAAGGTCAGAGGA TAATGCCGTCAAGATAGCACTTAACGGGCA AGAGGGAATGGTGAAAAGATACGCCACGTC CATAGGCCTACAATGGCCATGTGGAATAGA AGAAATCGAGGAAACGCGTAAAAAACTCA AGAGGGCGGATACAAACAAATGGAAAACT TTAATTTCGCAAGGACAAGGCATAAAAGAA TTTTTCGGGGATAAAACCGGGAATGCCTGG TTATACAACCCCGAAATGCTGAGACCGTCT CGATACCTGGACGCACTAAAACTGAGAACA AATACATATGGCACAAAAGCAGCACTCCAC AGAGCGAAAAGAGACATAGACATAAACTG TCGGAGATGCGGCGTCCAGGTGGAAACCCT AGGACATATATTGGGACTATGCACCCATAC AAAGAACAAAAGAATAAAAAGACACGACG AAATCTGCGATCTTATCGCAAAGAATGTCT CTAAAGAATACGTGATATTTAGGGAACCAG	GTGGTC CCGCTG GAATGC CTATCG ACTCCT CCCCGG AGGATC ATAGAG TTCGAA ACCGGC TACGGC GAGGCA AGGGCG GTGAGG TGCACA CCGATG GGGAGC AGCGAC CCCACC TACCCT TAGCTA AGAGAG CAGGCG	GAAA TTCC GCCA CGAT CCTTC TGAT CGTA GGCG CAAA ACA (SEQ ID NO: 1351)	

					AAGTAGAGGTAAACGGTGACAGACGTAAA	ATCCGC		
					CCAGATATGGTCATAAAAGACCATGACAAG	CAACTG		
					GTATACGTCGTGGACGTCACCGTAAGATAC	TCAGCA		
					GAAAACAATGATTCCCTAAACAAGGCCTAC	CGAAAT		
					AAAGAAAAAGAAAACAAATACAAAGAGAC	AAACTA		
					AGCGGAAATTATGAGAAGAGACTTGAAAGC	ATCATA		
					AAAAGAGAGCAGAGTTCTGCCAGTGGTTAT	TGTATA		
					CGGGAGCAGAGGGGCAGTGCCCCGAGCCA	CGAGGG		
					CTATAGAAAATCTAAAAGTCCTAGGGCTTC	AGAATT		
					AAACAAAACATGCCCTGACGGCTTCGCTCA	TACAAC		
					TAGCCCTCCGATCGTCGATCGAAATGGCAA	GGGTAC		
					ATGAATTCCTGGACTACGATCACACTACGT	CTTGTG		
					GATCGTTAAAAGTAAAAATCTATTTATTTAT	CCCGAA		
					TTTTATTCCTATATTATAACACATTATTTATT	CCGCCT		
					TATTTACTTATTGTTTTAAAGATGACGAAGC	GTAGGT		
					CGCAAGGCCAATCCAAATTTAACAAAAGAA	ATCACC		
					CGAGACTACTGGTCGACATTAAAAAGACGA	TACAGG		
					AGCAGCTGCCAGCTGATAAACAACAGAGCC	TGTTAA		
					CGTCTCGGCCTTTACACCGAGCGGTGCAAG	AATGAA		
					TCCTGACGTACTATTGTACGTCTAGGGCGCG	TCTGAT		
					GGGCAGATTCTACCGTGTAGAATCTGGGGC	AGCTGG		
					GACGCCTCCGCGAGGCACTCCCTGGACAAC	CGGATC		
					GTACGCTAAAGCGTACGGCTAAGTGCGCCT	GTCGAC		

					CCCGAAAGGGTCCCCGTTCCTAATTTTTCCGAGCCCGCGGGCAGATCTCGTGGCAGTGACGCTAGAAAGTTAAGTCCGCGGACATATAAAATTACAGCCTTAAATAATGAACCCCACGAAGGAGGTATCCTCGAAATTCCGCCACGATCCTTCTGATCGTAGGCGCAAAACA (SEQ ID NO: 1105)	CCTCTTTGATGGCTCTGCGCCAACGACTGGAAAGAATAGGAACGGAAGTCTAATGGAAGGAAAGTGTCGGGAGCACTATAAATTCCCAAAGAAGAAAAGAAAAAGAAAAAAATAAAAACCCAAATTAA		

						(SEQ ID NO: 1228)		
R2	R2B_N Vi	.		Nasonia vitripennis	GACTAGACTATGGGTTCAGTCAGTCCCAAA	GACTAG	TGAC	TFAPTHPMVRS
					TAGCCGATCCTGGCGCGTCCGGCAGTAATG	ACTATG	CTGA	GPCRKTKRPGS
					CCACGTATGAGTCGGTTACCCATCTCTAAAC	GGTTCA	ACAA	DYRESLIMDSGN
					GCGTAGAGGTGGGGAGCTAAAGGCCAGGC	GTCAGT	AACG	NVASEPRGAVD
					GGTTTACCCGACGTCGAATTTCTCCAGGTCT	CCCAAA	TGTT	VTSAAPIGAELN
					GTGTCAGTCGACGGAATAAAGGTACTACAA	TAGCCG	GTCTT	AEPCEGRNQRR
					CATCTACTATCTATCGGGATCGGAAGACGC	ATCCTG	GTCTT	EAALSAQTRRR
					CTTACAGCGTTTTCCGATTTTTGCTCTTTGA	GCGCGT	GTCT	NXARRARNAQQ
					GCATTTTTCTTCAAATTGCGATAACCGACCC	CCGGCA	AAAA	ADEPGDDEEIET
					GATCACGCGGGGCTTTGACAAAGCAATGCG	GTAATG	CTATT	HGPLTIRTXEPM
					TGGTCGGTAAGATGGTTGCAATCTTTTCCAC	CCACGT	TATTC	EIVAIAKNPQAC
					CTCGTTTCTTTTACGGAACGAAAGCAATGC	ATGAGT	GAAA	PKCLQGGTQLL
					GTGTGGGGAACGTTAAAAACTCCCTTCATG	CGGTTA	TAAG	CMGSWELSRHI
					CATCCCAGGATTTATCCTGCTTACTGCAAAG	CCCATC	GGGA	NKEHPSVDVTW
					CAATGCGTGTGGAGCGACTTTACCACGAGT	TCTAAA	GGCT	VCGACQRRCTT
					CGCTCCACCGCAAAGCAATGCGTATCGCGC	CGCGTA	AACT	LRSWSCHVLHC
					AAAAGCAATGCGTGTGGGGGACTTGTCAAA	GAGGTG	GCCT	KGRQEPKDLPF
					GATCCCCCGCCGCAAAGCAATGCGTGTCGG	GGGAGC	GCAA	KCEHCSLSFDSQ
					CACCACGTAGAGCAAAGCGTGTAGGCAGAC	TAAAGG	GTTG	IGLSQHERHVHP
					TTTGTCAAAAGTAGTTCTGCCGCAAAGCAA	CCAGGC	AACG	EVRNDKRAAEA
					TGCGTGTGGAGATCTTCGCCGGTGAAAGCA	GGTTTA	CGAA	NKPKGKSGRRP

					ATACGTGTGGGCGAACTTAGACCTTCGCGC	CCCGAC	AGTT	SIWSDEDLLLIR
					CCACACATCCCATGGTGAGGAGCGGCCCAT	GTCGAA	AGAC	ELESEYHGARNI
					GTCGAAAGACTAAGCGCCCCGGTAGTGACT	TTTCTC	CTTCC	NEKIAEHFPDRT
					ACCGTGAAAGTCTAATAATGGACAGCGGAA	CAGGTC	CACC	GRQVSDARRRK
					ATAACGTTGCCTCGGAGCCGAGGGGAGCTG	TGTGTC	TAAA	DYAALRGRGGP
					TGGATGTGACCTCAGCAGCTCCAATCGGGG	AGTCGA	GCCC	QGPAEGVEAIEE
					CGGAGTTAAACGCCGAACCTTGCGAAGGTC	CGGAAT	AAAA	VDEGEIPEGEEL
					GCAACCAAAGGAGGGAGGCTGCCTTAAGTG	AAAGGT	GTGA	VATDGAALESG
					CTCAAACACGCCGGCGAAATNCGGCCCGCC	ACTACA	TCGG	PPENGGSAPAEQ
					GAGCTCGGAATGCCCAACAGGCTGACGAGC	ACATCT	GGAA	VNAPALESSSQQ
					CCGGCGATGATGAGGAAATAGAGACACAC	ACTATC	TGAA	DRECSPAVGSDE
					GGGCCTCTAACTATCCGGACGNCGGAGCCG	TATCGG	TCCG	QIEDSSDDDEFS
					ATGGAGATTGTCGCAATAGCGAAAAACCCA	GATCGG	CGGG	DALGEISLPEPLS
					CAGGCCTGTCCCAAATGCCTGCAGGGAGGT	AAGACG	TGAC	VERTTISPPPRD
					ACCCAACTTCTCTGCATGGGCAGCTGGGAA	CCTTAC	CCCA	DWKGPMRWEIC
					CTAAGCAGGCACATTAATAAAGAACATCCG	AGCGTT	GAGT	NASEEAGSYAN
					TCAGTCGACGTGACCTGGGTGTGCGGTGCT	TTCCGA	TGGG	WVTGLQELVRN
					TGTCAAAGGCGCTGCACAACGCTCAGGTCG	TTTTTG	TAAA	NALSEIGLDSLY
					TGGAGCTGTCATGTTCTNCACTGTAAGGGG	CTCTTT	CCCTT	DQLIQIMRHPSD
					CGACAAGAACCAAAGGATCTGCCGTTCAAA	GAGCAT	GAAA	DNEQDRLQLNA
					TGTGAGCATTGCAGCTTGTCGTTTGACTCGC	TTTTCTT	CGTT	RGPPRRGHRKN
					AAATCGGACTCTCTCAGCACGAGAGGCATG	CAAATT	GGAG	RRRRRLTAADR
					TCCATCCAGAGGTGCGAAACGATAAGCGCG	GCGATA	AAGC	KRFAFARCQDL

					CGGCAGAGGCCAATAAGCCAAAGGGCAAG	ACCGAC	GGAA	WNNNPKKLAEL
					AGTGGCCGTAGGCCTTCTATATGGTCCGAC	CCGATC	GAGA	VIANDLSILQRR
					GAAGACTTGCTGCTCATCCGGGAATTAGAG	ACGCGG	GTCC	QAPGRTETQTL
					AGCGAATACCACGGAGCTCGAAATATCAAT	GGCTTT	CGCC	YNELWGRVGPN
					GAAAAAATAGCTGAACATTTCCCGGATAGA	GACAAA	ACCG	IEAPRRTEDPIPV
					ACAGGCAGACAGGTGTCGGACGCCCGGAG	GCAATG	AGCA	SRIFTPITPQEIM
					GCGTAAGGATTATGCCGCGCTCCGAGGGAG	CGTGGT	TCGA	GRIRRIKNDSAA
					AGGAGGCCCGCAAGGCCCAGCAGAAGGAG	CGGTAA	GTGC	GPDGVTKDDLR
					TCGAGGCCATCGAAGAGGTAGACGAAGGC	GATGGT	TGCG	GRGVSIALSKLF
					GAAATCCCTGAGGGGGAAGAGCTGGTCGCC	TGCAAT	GCGC	NSILLAGYYPKA
					ACCGATGGCGCTGCGTTGGAAAGCGGCCCC	CTTTTC	CCGA	WRENRTTLLPK
					CCGGAGAATGGAGGGAGCGCACCCGCAGA	CACCTC	ATGA	PEKDPADVKNW
					ACAAGTCAATGCGCCCGCGCTGGAAAGCAG	GTTTCT	AACC	RPITISSMVSRV
					TTCTCAGCAAGATCGAGAGTGCAGCCCGGC	TTTACG	GATC	YSGLLDQRVRA
					AGTGGGGTCCGATGAACAAATCGAGGACAG	GAACGA	GCGG	VIKQCDRQKGF
					CAGTGACGACGACGAATTCAGCGACGCATT	AAGCAA	ATGG	TEENGCFSNIQL
					AGGAGAAATATCACTCCCAGAACCTCTCTC	TGCGTG	TGCA	LDDAVSNAKKA
					GGTTGAACGCACAACAATCTCACCACCTCC	TGGGGA	AGTC	GGVITILDVSKA
					TCGAGATGACTGGAAAGGCCCAATGAGGTG	ACGTTA	GTAG	FDTVPHAVIQGC
					GGAGATTTGCAATGCGAGCGAAGAAGCCGG	AAAACT	GACG	LEKKGIPETVAA
					AAGTTACGCGAACTGGGTGACTGGACTGCA	CCCTTC	GGGC	YISSMYRDCSTA
					GGAGCTAGTCAGGAACAATGCGCTGAGTGA	ATGCAT	ACGA	IRTRSGDVKIGM
					AATAGGACTAGACTCCCTGTATGACCAGCT	CCCAGG	CCTA	KRGVKQGDPLS

					CATCCAGATTATGCGGCACCCTTCCGATGA	ATTTAT	AGCC	PLIFNLVLEPLLE
					CAACGAACAGGATCGCCTTCAATTGAACGC	CCTGCT	TCTGT	RLQETSGVEIEG
					TAGAGGCCCCCCACGAAGGGGCCACCGCAA	TACTGC	CACG	MNLSCAAFADD
					GAACCGACGGCGCCGTCGTCTCACGGCCGC	AAAGCA	GCGG	IVCFANTAPEAG
					TGATCGAAAGCGGTTTGCCTTTGCCAGGTG	ATGCGT	CGAA	RQLRMVADYLG
					CCAAGATCTTTGGAACAACAACCCAAAGAA	GTGGAG	GCCA	RLDMSLSVSKCI
					GCTAGCCGAGTTAGTGATTGCCAATGACCT	CGACTT	GGAA	AVEYVPHRKTW
					GTCCATTCTCCAAAGGCGCCAAGCGCCAGG	TACCAC	TCAC	YTKNPGLEVNG
					TAGAACGGAGACACAGACTCTGTACAACGA	GAGTCG	CATG	NAVPSISPSETFK
					GCTGTGGGGGAGGGTCGGACCTAATATCGA	CTCCAC	CAAA	YLGAKVSPWKG
					AGCGCCAAGGCGCACCGAAGACCCGATACC	CGCAAA	GGTG	LLEGFESDAFRE
					CGTATCGAGGATCTTCACTCCGATCACTCCC	GCAATG	TGAA	VISRVQRLPLKP
					CAAGAGATAATGGGCAGAATCAGGCGAATC	CGTATC	CTGG	MQKVDLLQMYI
					AAAAACGACTCGGCAGCGGGTCCTGACGGG	GCGCAA	GGCG	FPRYTYGLITSPP
					GTAACGAAGGACGACCTGAGAGGAAGAGG	AAGCAA	GATA	AKAVLKTIDRII
					AGTCAGCATAGCCCTCTCCAAGCTGTTCAA	TGCGTG	CCTC	RTRIKEILHLPES
					CTCGATCCTGCTAGCGGGTTACTACCCAAA	TGGGGG	CACG	VSSSFLYTPRKQ
					GGCATGGAGAGAGAACAGAACAACCCTTCT	ACTTGT	GGGT	GGLGLLEVEKM
					GCCGAAGCCAGAAAAAGATCCTGCTGACGT	CAAAGA	TTCCC	VLIAALRNGLRA
					TAAGAACTGGCGGCCCATTACCATCAGCTC	TCCCCC	TGGG	RQSHDPVTRAA
					AATGGTTAGTCGAGTCTACTCAGGCTTGCTT	GCCGCA	CATC	MNSNAADDRLK
					GACCAGCGAGTGAGGGCCGTCATTAAGCAG	AAGCAA	GCGC	SYADALRLHWP
					TGTGATCGGCAGAAAGGATTCACGGAGGAA	TGCGTG	GAGC	LTTKELDTYKY

					AATGGCTGTTTCAGCAACATACAGTTGTTG	TCGGCA	GATG	QLRLSYAQKWA
					GATGACGCCGTATCGAACGCAAAGAAAGCG	CCACGT	GCCA	EQKWQGQGVEE
					GGCGGTGTCATTACTATCTTGGATGTTTCGA	AGAGCA	AAGT	FAQDPVGNSWL
					AAGCATTCGACACTGTCCCGCATGCCGTGA	AAGCGT	CCGC	QRYDLLPASRYI
					TCCAAGGGTGCTTGGAGAAAAAAGGAATCC	GTAGGC	TTTCT	DAIKLRTNTYPT
					CCGAAACCGTGGCCGCCTATATCTCGAGCA	AGACTT	CAGC	RALMKIIDGRVD
					TGTATCGCGACTGCTCCACTGCAATCCGAA	TGTCAA	TACA	SSCRKCQGSSET
					CGAGGAGCGGGGACGTAAAGATTGGAATG	AAGTAG	AAAC	LGHILGRCRYTK
					AAGAGAGGAGTCAAGCAGGGGGATCCCCT	TTCTGC	AAAA	DKRISRHNEIKD
					GTCACCTCTCATTTTCAATCTGGTTCTCGAA	CGCAAA	ATGG	LLKARLAKNHQ
					CCTCTATTAGAACGATTGCAAGAGACAAGT	GCAATG	TATG	VMDEPQITVRG
					GGAGTGGAAATCGAAGGCATGAATCTCTCG	CGTGTG	AGAC	QRFKPDLVVKT
					TGCGCGGCTTTCGCAGACGACATAGTATGT	GAGATC	TTCGT	NEGRVHVIDVT
					TTTGCGAATACAGCCCCCGAGGCGGGAAGG	TTCGCC	TAAC	VRYEHRTYLDE
					CAGCTACGGATGGTGGCGGATTATCTGGGC	GGTGAA	ACTA	GRTEKIGKYRQI
					CGACTCGATATGAGTCTTTCAGTGTCAAAGT	AGCAAT	ATTTT	LSTLRRDLHSNA
					GTATAGCTGTAGAGTATGTCCCCCACAGGA	ACGTGT	TCCG	EEVIPIVIGSRGA
					AGACCTGGTACACTAAAAACCCAGGCCTCG	GGGCGA	AGCC	IPRETRKALSKL
					AGGTGAACGGTAATGCCGTTCCGAGCATCT	ACTTAG	TAGC	GIGKSDWLTISLI
					CACCTAGTGAGACGTTCAAGTACCTCGGGG	(SEQ ID	AGGC	ALRSSLEIVNAF
					CAAAGGTCTCTCCCTGGAAGGGGCTGCTCG	NO: 1229)	TCCCT	MDD (SEQ ID
					AAGGCTTCGAATCTGACGCGTTCAGGGAAG		TGAC	NO: 1474)
					TCATATCCCGCGTCCAAAGACTGCCGTTGA		AACG	

					AGCCCATGCAAAAGGTGGACCTTCTACAGA		CTTAT	
					TGTATATCTTTCCGAGGTACACCTATGGGTT		GAAT	
					GATAACATCGCCTCCGGCGAAGGCAGTCTT		CTGG	
					AAAGACTATCGACCGGATCATAAGAACGAG		AAAA	
					AATCAAGGAGATCCTCCACCTGCCAGAATC		GGAC	
					GGTAAGTAGCAGTTTTCTCTACACGCCGAG		ACAA	
					GAAGCAGGGTGGATTGGGGCTCCTTGAAGT		AGTG	
					GGAGAAGATGGTGCTGATAGCCGCTCTTCG		GAAA	
					GAACGGCTTGAGAGCCCGTCAATCCCACGA		AAGC	
					TCCGGTCACACGCGCGGCCATGAACTCGAA		GCTG	
					CGCAGCCGACGATCGGCTAAAATCTTACGC		ATGG	
					CGATGCTCTAAGACTACACTGGCCACTAAC		TGGA	
					AACCAAGGAGCTAGATACTTATAAGTATCA		CAAA	
					GCTTCGCCTGAGCTATGCCCAGAAATGGGC		AGTC	
					TGAGCAAAAATGGCAAGGCCAGGGGGTCG		AGTT	
					AGGAGTTCGCACAAGATCCCGTCGGCAACT		GAGA	
					CATGGCTGCAGCGCTATGATCTTCTGCCCGC		CTTG	
					GTCAAGGTACATCGATGCTATCAAGCTACG		ATAT	
					AACAAACACGTACCCGACGCGAGCACTAAT		CAGT	
					GAAGATCATAGATGGACGTGTTGATAGCTC		TGTTT	
					TTGCCGAAAGTGCCAAGGCAGTAGCGAGAC		TGAC	
					CCTTGGTCATATACTTGGCAGATGCCGGTAT		TAAG	
					ACTAAGGATAAGCGAATAAGCCGGCACAAT		AATT	

					GAAATCAAAGACCTCCTCAAGGCTCGTCTA		TTATT	
					GCCAAAAACCATCAGGTTATGGACGAGCCG		ATCG	
					CAGATAACGGTCCGAGGCCAGAGGTTTAAA		TTGA	
					CCCGATCTCGTCGTGAAAACGAATGAGGGA		CTTTT	
					AGGGTGCACGTAATCGACGTAACTGTCCGC		AAAT	
					TACGAGCACAGAACCTATCTGGATGAGGGC		ATTTT	
					CGTACTGAGAAAATTGGCAAATATCGCCAA		ATTA	
					ATTCTCAGCACGCTTCGGCGAGATCTGCACT		TTGA	
					CGAACGCCGAGGAGGTCATTCCAATTGTAA		CTGTT	
					TCGGGTCGAGAGGTGCAATTCCAAGGGAAA		AATA	
					CGCGGAAAGCCCTCTCGAAACTCGGAATTG		TACT	
					GCAAGAGTGATTGGCTTACGATCTCACTAA		GACT	
					TAGCGCTGCGTAGCTCGCTAGAGATCGTCA		TGGG	
					ACGCCTTCATGGATGACTGACCTGAACAAA		ACCA	
					ACGTGTTGTCTTGTCTTGTCTAAAACTATTT		AGTC	
					ATTCGAAATAAGGGGAGGCTAACTGCCTGC		ATCT	
					AAGTTGAACGCGAAAGTTAGACCTTCCCAC		CTGTT	
					CTAAAGCCCAAAAGTGATCGGGGAATGAAT		ACCC	
					CCGCGGGTGACCCCAGAGTTGGGTAAACCC		GGTA	
					TTGAAACGTTGGAGAAGCGGAAGAGAGTCC		CCGG	
					CGCCACCGAGCATCGAGTGCTGCGGCGCCC		TTCCT	
					GAATGAAACCGATCGCGGATGGTGCAAGTC		GTCA	
					GTAGGACGGGGCACGACCTAAGCCTCTGTC		TCAA	

					ACGGCGGCGAAGCCAGGAATCACCATGCAA AGGTGTGAACTGGGGCGGATACCTCCACGG GGTTTCCCTGGGCATCGCGCGAGCGATGGC CAAAGTCCGCTTTCTCAGCTACAAAACAAA AATGGTATGAGACTTCGTTAACACTAATTTT TCCGAGCCTAGCAGGCTCCCTTGACAACGC TTATGAATCTGGAAAAGGACACAAAGTGGA AAAAGCGCTGATGGTGGACAAAAGTCAGTT GAGACTTGATATCAGTTGTTTTGACTAAGA ATTTTATTATCGTTGACTTTTAAATATTTTAT TATTGACTGTTAATATACTGACTTGGGACCA AGTCATCTCTGTTACCCGGTACCGGTTCCTG TCATCAAACCGGAAAGTCCGTCCCACGTAA TGTGGTAGACGCAGGAG (SEQ ID NO: 1106)		ACCG GAAA GTCC GTCC CACG TAAT GTGG TAGA CGCA GGAG (SEQ ID NO: 1352)	
R2	R2Ci-B	AB0971 22		Ciona intestinalis	CGACGGTGAACCACCTTGTCGCGGTGTAAG AGCTTTAGTGTCTCGAACAAGAAATAGCTT GTGTGCTGTCCTTCTGGGCGGTGCACATACT TCTTAACCTCCCGAGGCCATGCCGGCGGGG GCTTTAGCCCCCGGCAGGTTTTACCATGCCG GACGGGTTCGAGAGGTAGAGGCCAAACTAA GAGTTCACCAGCAGACTTCGCACGCGGCTG GCCACTGGCCGAAGTTTAAACAACAGGGCC GCATCTTCCCAAACTCAATATATGGTGTTAA	CGACGG TGAACC ACCTTG TCGCGG TGTAAG AGCTTT AGTGTC TCGAAC AAGAAA	TGAC AGTA ATAT GAAA ACAT CACA TCTG ACCG GCAC	MGEWPWVSWS LTVLVEKWRPF TILQPYPMPGQL RVDVYLPRKTS YLMDKNIYENT TSPGGGPLCGEK THRSDVIIPPPGF APSTDTASNTLG ENVDASATTSSA

					GTGAACCGTGCCGATGGGTGAATGGCCATG	TAGCTT	AGAA	NPLSQEPGWCES
					GGTAAGTTGGTCTCTAACGGTCCTTGTGGA	GTGTGC	TCAC	CSKLFKSQRGLR
					GAAGTGGCGACCCTTCACAATACTTCAGCC	TGTCCT	CATG	VHQRSKHPELY
					TTACCCTATGCCCGGGCAACTTAGGGTGGA	TCTGGG	CCGT	HSQNQPLPRSKA
					CGTATATCTCCCGCGGAAAACATCTTATCTC	CGGTGC	AATG	RWSDEEMVIFA
					ATGGATAAAAATATATACGAAAACACTACT	ACATAC	CACC	REEIANRKIRFIN
					TCTCCAGGAGGGGGACCGCTTTGCGGTGAG	TTCTTA	CAAC	QHLHKVFPHRT
					AAGACGCATCGTTCCGATGTGATCATCCCTC	ACCTCC	TAAG	LESIKGLRGKNV
					CACCGGGGTTCGCCCCGTCGACGGACACTG	CGAGGC	GATT	RYARIMADLEA
					CTTCCAATACATTAGGGGAAAATGTGGATG	CATGCC	CCAA	EMTSQPEAATSL
					CATCTGCAACCACTAGCAGTGCCAACCCAT	GGCGGG	TGGG	CTETSENLASSN
					TAAGTCAGGAACCAGGATGGTGCGAATCTT	GGCTTT	TAAA	VLPQTRGWAEN
					GTTCCAAGCTGTTCAAATCACAAAGAGGAC	AGCCCC	AAAA	LVENIDTAHLAN
					TTAGAGTCCACCAAAGATCAAAACATCCAG	CGGCAG	AAAA	LGPLSQFEPGKP
					AGTTGTACCATTCTCAGAATCAACCGCTTCC	GTTTTA	AAAA	SSSTKEAINTEY
					GCGATCTAAAGCACGATGGAGTGATGAGGA	CCATGC	AAAA	NDWISKWLPSG
					AATGGTTATCTTTGCTCGCGAAGAAATTGC	CGGACG	AAAA	AAHRERRANPP
					AAATAGGAAGATCCGCTTTATCAATCAGCA	GGTTCG	AAAA	STKLNARATRR
					CTTACACAAAGTGTTTCCACACAGAACGCT	AGAGGT	AAAA	LQYSRIQNLYKL
					AGAGTCAATTAAAGGGTTGAGGGGAAAAA	AGAGGC	AAAA	NRSACAQEVLS
					ACGTAAGATACGCTCGCATCATGGCTGACC	CAAACT	AAAA	GAWKVQSGELN
					TTGAGGCGGAAATGACCTCTCAGCCCGAAG	AAGAGT	AA	LKEVQPFWEKM
					CAGCCACAAGCTTGTGCACCGAAACATCAG	TCACCA	(SEQ	FRKESAKDRRK

					AAAATCTCGCTTCTAGCAATGTTCTGCCCCA	GCAGAC	ID NO:	PKPTGEVLWGL
					AACAAGAGGATGGGCTGAAAATCTCGTCGA	TTCGCA	1353)	MEPLTIAEVGST
					AAATATTGACACAGCACATCTGGCAAACCT	CGCGGC		LKSTTPSAPGPD
					TGGGCCGTTATCCCAGTTTGAACCTGGTAA	TGGCCA		KLTLDGVKRIPI
					ACCAAGCAGCAGTACAAAAGAGGCTATCAA	CTGGCC		AELVSHYNLWL
					CACCGAGTACAACGACTGGATCAGCAAATG	GAAGTT		YAGYQPEGLRE
					GCTTCCCTCTGGTGCTGCTCATAGAGAAAG	TAAACA		GITTLIPKIKGTR
					AAGAGCAAATCCACCCTCAACTAAGCTAAA	ACAGGG		DPAKLRPITVSS
					TGCTAGAGCAACAAGGAGGCTACAATACAG	CCGCAT		FICRIFHRCLAQ
					TAGGATCCAGAATCTCTACAAATTAAATCG	CTTCCC		RMETSLPLGERQ
					TTCTGCCTGCGCCCAAGAAGTATTAAGCGG	AAACTC		KAFRKVDGICH
					GGCTTGGAAGGTCCAGTCGGGCGAACTGAA	AATATA		NIWSLRSLIHNS
					CTTGAAAGAGGTCCAACCGTTTTGGGAAAA	TGGTGT		KDNLKELNITFL
					AATGTTCCGGAAAGAATCCGCCAAGGACAG	TAAGTG		DVRKAFDSISHK
					GCGCAAGCCCAAGCCCACTGGTGAAGTACT	AACCGT		SLGIAAARLGLP
					CTGGGGACTGATGGAGCCTCTAACAATTGC	GCCG		PPLITYISNLYPN
					CGAGGTGGGATCCACTCTGAAGTCAACGAC	(SEQ ID		CSTKLKVNGKIS
					ACCATCAGCCCCAGGACCCGACAAGTTAAC	NO: 1230)		KPIEVRRGVRQG
					ACTAGATGGCGTAAAGAGGATACCCATTGC			DPLSPLLFNAVM
					TGAACTAGTATCCCATTACAACCTATGGTTG			DWALSELDPRV
					TACGCTGGCTACCAACCAGAGGGACTGCGC			GVQIGEQRINHL
					GAAGGAATAACCACCCTAATTCCGAAAATT			AFADDIILVSST
					AAGGGTACGCGTGACCCGGCTAAGCTACGC			KIGMVSSINTLS

					CCAATTACAGTGTCTAGCTTTATTTGCCGCA			RHLAKSGLEISA
					TCTTCCACCGATGTTTGGCACAACGAATGG			GKEGKSASMAI
					AAACCTCTCTTCCATTGGGCGAAAGACAGA			VVDGKKKMWT
					AAGCTTTCCGCAAGGTCGATGGCATCTGTC			VDPLPRFKVNSQ
					ATAATATTTGGTCCCTAAGGTCTCTAATACA			KIPALSITQQYK
					CAATAGCAAAGACAACTTAAAAGAACTCAA			YLGINIDAQGAR
					TATAACATTTTTGGACGTACGAAAAGCATTT			NDAARILTEGLA
					GACTCCATCTCGCATAAATCACTGGGCATA			ELSRAPLKPQQR
					GCTGCAGCCCGCCTTGGACTTCCCCCACCCT			LYLLRVHLLPKL
					TGATAACATACATCTCCAACCTTTACCCAAA			QHGLVLSSCAK
					CTGTTCTACCAAATTAAAAGTAAACGGAAA			RALTYLDKSVR
					AATATCAAAACCAATTGAAGTTCGTCGAGG			SAIRRWLTLPKD
					AGTCCGACAAGGAGACCCACTCTCCCCGCT			TPTAFYHAKAC
					ACTGTTTAATGCAGTAATGGATTGGGCCCT			DGGLGITRLEHT
					ATCTGAACTGGACCCAAGAGTGGGAGTACA			IPILKRNRMMKL
					AATTGGGGAACAACGAATAAACCATCTAGC			TLSEDPVIMELV
					ATTCGCCGATGATATTATTTTGGTCTCTTCA			KLTYFTNLLHK
					ACAAAGATCGGAATGGTGTCCTCCATCAAT			YSNVKLLNSWP
					ACGTTGAGTCGCCATCTGGCAAAATCAGGC			VTDKDSLARAE
					CTCGAAATTAGCGCTGGGAAAGAAGGAAA			ASMLHTSVDGR
					ATCAGCTAGTATGGCCATCGTTGTTGATGG			GLSNCSDVPRQS
					AAAGAAAAAGATGTGGACAGTAGATCCCCT			DWVTNGASLLS
					TCCTAGGTTCAAAGTAAACAGCCAGAAAAT			GRDFIGAIKVRG

					TCCGGCACTGAGTATTACACAACAGTATAA GTATCTGGGGATAAACATAGATGCCCAAGG GGCAAGAAATGATGCAGCTCGAATCCTTAC AGAGGGTCTAGCAGAATTATCTCGTGCCCC ACTGAAGCCACAACAGCGACTATATCTGCT TCGGGTGCATCTTCTTCCGAAACTGCAACAC GGACTGGTTTTAAGTAGTTGTGCAAAAAGA GCGCTAACATACCTGGACAAAAGTGTAAGG TCTGCAATAAGAAGATGGCTTACACTTCCG AAGGACACACCAACAGCCTTTTACCATGCG AAAGCATGTGACGGTGGACTAGGCATTACA CGTTTGGAACACACAATACCAATACTCAAA AGAAATAGAATGATGAAGTTGACATTGTCA GAGGATCCAGTGATCATGGAACTTGTAAAA TTGACATACTTTACAAATTTACTACACAAAT ACTCAAATGTGAAGCTGCTTAACTCCTGGC CAGTAACAGACAAAGACAGTCTGGCCCGTG CCGAAGCCTCCATGCTTCACACCTCTGTGGA TGGGCGTGGACTAAGCAATTGCAGTGATGT CCCGCGCCAAAGTGACTGGGTCACAAATGG CGCGTCGCTACTTAGTGGTCGTGACTTTATT GGAGCAATAAAAGTGAGGGGCAACCTCTTG CCAACAAAGGTTAGTGCGGCACGTGGACGG			NLLPTKVSAAR GRQREITCDCCR RPESLGHILQTC PRTWGPRISRHD SLLKRVRNQAC LKNWTPIIEPSIP TNIGLRRPDLVL AKGNIAFLVDA TVVADNANMQ LQHEAKVEKYN NSDIKEWIKVHC PGVDEVRVTSLT ANWRGCLYGGS ASFLTEDLGLPK AELSLLSAKINE KGYYLWCAHY RGTARLWNRPL RS (SEQ ID NO: 1475)

					CAACGGGAGATAACATGCGACTGCTGCAGA CGACCGGAGTCACTGGGCCACATCCTGCAG ACCTGCCCTAGAACATGGGGACCGAGAATC TCCAGGCATGACTCCCTGTTGAAGCGAGTT CGGAATCAGGCATGCTTAAAAAACTGGACC CCGATAATAGAGCCATCGATACCAACCAAC ATAGGACTTCGTCGGCCAGACCTCGTTTTGG CCAAAGGTAACATCGCCTTTCTTGTAGATGC CACCGTGGTTGCAGACAATGCCAACATGCA GTTGCAACACGAAGCAAAAGTGGAAAAGT ACAACAACTCAGATATAAAGGAATGGATAA AAGTACACTGCCCAGGGGTTGATGAAGTGA GGGTAACGTCACTAACTGCGAACTGGCGCG GTTGCTTGTATGGAGGTTCAGCATCCTTCCT AACAGAAGACCTGGGACTGCCTAAAGCAGA GTTAAGTCTGCTCAGTGCAAAGATCAATGA AAAAGGATACTACCTTTGGTGTGCTCACTA CAGGGGAACAGCAAGGTTGTGGAATCGACC GTTAAGGTCCTGACAGTAATATGAAAACAT CACATCTGACCGGCACAGAATCACCATGCC GTAATGCACCCAACTAAGGATTCCAATGGG TAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAA (SEQ ID NO: 1107)			

R2	R2C_NGi	.		Nasonia girault	CGGGTTCCCCCGACTTCGGCTTGCCGTGGTC	CGGGTT	TAGC	WVTSPRRPRYV
					TGGGGCTCACTGCTTTTTGTGGAGTCATGGT	CCCCCG	GGAC	GPQKKKASDGN
					TACATGGTGACCCTGGTTCCTCGCACCCCCG	ACTTCG	TGGA	DGRAAARAEPT
					CTGGAAACTATCTGGGGAGGCCATGATTGG	GCTTGC	CTGT	NPGGPDRADDD
					GTAACGATAAAGGTCCTGGTCGTGTCCTCCT	CGTGGT	CTGG	EGDVKFWCEFP
					GAGATAGGCTGAATGGGTCACTAAGTGGCA	CTGGGG	AGGA	GCDRFFMTRSG
					CCTAATGGGTTACCAGCCCTAGGCGGCCGA	CTCACT	GTGT	RGLHHKKGHPD
					GATACGTTGGTCCACAGAAAAAGAAAGCCT	GCTTTT	TTAA	WNDQRNLAGK
					CGGATGGAAATGACGGACGAGCCGCTGCTC	TGTGGA	CTCG	QHRKEIWSEEER
					GTGCCGAACCAACGAATCCGGGCGGACCAG	GTCATG	GGTT	LLLAKKEAELAI
					ACCGCGCTGACGACGACGAAGGGGATGTTA	GTTACA	CTCA	SGARFINVELRD
					AATTCTGGTGTGAATTTCCAGGTTGTGATCG	TGGTGA	TGGG	FTARSLDAIKGQ
					CTTCTTTATGACCAGGAGCGGTAGAGGCCT	CCCTGG	AACC	RKRPDYKILVEK
					CCATCACAAGAAAGGCCACCCTGATTGGAA	TTCCTC	CGAC	FVRELRVRGIRQ
					TGATCAGAGAAACCTGGCCGGAAAGCAACA	GCACCC	AACG	GVASRSQQARA
					CCGAAAAGAGATATGGTCGGAGGAAGAAC	CCGCTG	TTGTT	MAVAGAPAATS
					GTCTCCTGCTTGCCAAAAAAGAGGCGGAGC	GAAACT	ATCTT	SGAPPVATQPPP
					TTGCCATCAGTGGAGCTAGGTTTATTAACGT	ATCTGG	GTAT	SGRVLRSQVVE
					AGAGTTGCGTGATTTTACAGCGCGCTCCCTA	GGAGGC	GACA	APAMEIPVAESE
					GACGCTATCAAGGGCCAGCGAAAGAGACCC	CATGAT	ATTC	GDSSGDELFEDV
					GACTATAAGATCTTAGTCGAGAAATTTGTC	TGGGTA	ATAA	EPVRLSDLPPDR
					AGGGAGTTAAGGGTTAGAGGCATTCGTCAA	ACGATA	AAAA	FTIYFAGLEIPGT
					GGAGTGGCCTCGCGGAGTCAACAAGCTCGC	AAGGTC	AAAA	EDIYAHRLHTIC

					GCGATGGCGGTGGCAGGAGCTCCTGCAGCG	CTGGTC	AAAA	LMTTWRTKEEV
					ACGTCCTCGGGGGCACCACCTGTCGCGACT	GTGTCC	AAAA	RLELGLFLKDLF
					CAACCACCACCATCAGGTCGCGTACTAAGA	TCCTGA	AAAA	PSKGSQERPERT
					TCTCAGGTCGTTGAAGCACCAGCGATGGAG	GATAGG	AAAA	NLPDPRNRIERR
					ATCCCCGTGGCAGAGTCGGAAGGTGACTCC	CTGAAT	AA	RGEYKKCQDLW
					TCGGGGGACGAGCTGTTTGAGGATGTCGAG	GGGTCA	(SEQ	RRNKSTCVQRIL
					CCCGTGCGATTGTCCGACCTACCCCCTGACA	CTAAGT	ID NO:	KEDLSQGECLPR
					GGTTTACGATATACTTTGCTGGGCTTGAAAT	GGCACC	1354)	ELMEPFWNATF
					ACCCGGCACCGAAGATATATATGCCCACAG	TAA		TQNPGTAPVLPP
					GCTCCATACCATCTGCCTGATGACAACGTG	(SEQ ID		PTEVYSSVWEPI
					GCGAACCAAAGAAGAGGTAAGATTAGAAC	NO: 1231)		RPENIKGNYPPQ
					TTGGCCTCTTTTTGAAAGATTTGTTCCCGAG			NTAAGIDGLTV
					TAAGGGCAGTCAAGAACGCCCGGAGAGAA			GDLKGVSREML
					CCAACCTGCCGGACCCGAGAAATCGGATCG			ARIFNLFMWCG
					AGAGGCGGAGGGGGGAGTACAAAAAATGC			KLPEHLCASRTI
					CAGGATCTATGGCGACGAAATAAGTCAACC			LLPKKPGAKVP
					TGTGTTCAGCGGATCCTTAAAGAGGATCTTT			GEFRPITVTSVLI
					CGCAGGGTGAATGTTTGCCTCGAGAGCTGA			RTFHKVLAERL
					TGGAGCCCTTTTGGAATGCGACTTTCACCCA			KVVPLDPRQRG
					GAATCCTGGCACGGCTCCGGTGCTCCCTCCT			FRESDGCAENV
					CCCACGGAGGTTTATTCTAGTGTTTGGGAGC			MLLDMTIRYHH
					CCATTCGGCCCGAGAATATCAAGGGCAACT			ERRRKMFLALL
					ATCCGCCGCAGAACACCGCGGCAGGGATAG			DMAKAFDSVSF

					ACGGACTGACAGTGGGTGACCTGAAAGGGG			ESMREVLTTKGI
					TGTCGCGGGAGATGCTGGCCAGAATTTTTA			PTPFIEYFMTHL
					ACTTATTCATGTGGTGCGGCAAACTGCCAG			EDSFTVLQHGN
					AGCACCTTTGTGCCTCACGCACAATTCTCCT			WQSGKIHPTCG
					GCCCAAGAAACCTGGGGCGAAAGTCCCCGG			VKQGDPLSPPIF
					CGAATTCAGGCCTATCACCGTGACATCCGT			NFIMDEMLKRL
					CCTCATCCGGACCTTTCACAAGGTTCTGGCC			PKEIGVNLDGLF
					GAAAGACTGAAGGTTGTCCCTCTTGACCCC			VNAMAFADDLS
					CGCCAAAGAGGCTTCAGAGAGTCCGATGGA			LVANTEQGLQIL
					TGTGCAGAGAACGTGATGCTACTGGACATG			IDEATSFLGLCG
					ACCATCCGGTACCACCACGAGCGGCGCAGA			LRANPNKCVTL
					AAGATGTTCTTGGCCCTGCTAGACATGGCT			AIKTIPKEKKTAI
					AAGGCATTTGACTCGGTCTCTTTCGAGTCCA			DPSSHFRIGNAV
					TGCGGGAGGTTTTGACTACTAAAGGCATAC			IPSLKRTDEWVY
					CAACGCCATTTATTGAGTATTTTATGACGCA			LGIKFNSNGRLIS
					CTTGGAGGATAGTTTTACTGTTCTTCAGCAT			DAKPKLIKDLEL
					GGTAACTGGCAATCGGGGAAAATCCACCCA			LTKAPLKPQQRL
					ACATGTGGTGTGAAGCAAGGCGATCCACTG			WALKVIVIPGIL
					TCTCCGCCTATCTTCAACTTCATCATGGATG			YRGTLGSSTAG
					AAATGTTGAAGAGGTTGCCTAAGGAAATCG			YLRSLDCVIRAY
					GGGTTAACTTGGACGGGTTATTTGTTAATGC			VRRWLRLPGDC
					TATGGCATTTGCGGATGACCTGAGCCTTGTT			PNGYFHAAVAD
					GCCAATACCGAACAAGGTCTGCAGATCCTC			GGLGVHPIRYK

					ATAGATGAAGCTACTTCCTTTCTGGGGCTCT			AMVDRLARLRK
					GTGGACTCCGCGCCAATCCCAATAAGTGCG			LEKSAYITGPEA
					TCACCCTAGCAATTAAGACCATCCCGAAGG			ARYLQRQVSIAE
					AGAAAAAGACGGCCATTGACCCCTCATCAC			NRLRDGANRIM
					ATTTTAGGATAGGTAATGCGGTGATCCCCTC			SDASMLREFLRE
					GTTGAAGAGGACAGATGAGTGGGTGTACTT			LLYKSFDGRPLE
					AGGGATCAAATTTAATTCAAATGGTCGCCT			NSSKVPGQHRW
					TATCTCTGACGCAAAACCCAAGCTCATAAA			VEEPTRFLSGAD
					AGATCTTGAGCTACTAACTAAGGCACCACT			YMNCIRARIAAL
					CAAACCACAGCAAAGACTGTGGGCGCTTAA			PTAARCARGRL
					GGTGATTGTCATTCCGGGCATCCTTTACAGA			KDKHCRAGCGN
					GGTACCCTGGGGAGCAGCACCGCAGGCTAC			VETLNHVLQFC
					CTACGCTCTCTTGACTGTGTAATAAGGGCCT			HRTHGTRIGRH
					ATGTTCGGCGATGGCTACGTCTCCCTGGAG			DAVVKYVVGGL
					ATTGCCCGAATGGGTATTTTCATGCAGCGGT			KKRGYAVKEEP
					TGCGGATGGAGGGCTGGGAGTTCACCCCAT			KIVLQDVVYKP
					ACGATACAAGGCGATGGTAGATCGCCTTGC			DMVATKEGKTL
					CCGGCTCCGAAAATTAGAGAAATCCGCGTA			ILDAQVLGDQR
					CATCACGGGGCCTGAAGCCGCACGTTATCT			DMRLAHEDKLR
					TCAAAGACAAGTTTCTATCGCCGAAAATAG			KYGAPEFKRKIR
					GCTCCGAGATGGGGCCAACCGCATTATGAG			SETGSATIKSLS
					TGATGCGAGTATGCTAAGGGAGTTCCTTCG			VTLSWRGLWGP
					GGAGCTTCTGTACAAGTCCTTTGATGGTCGT			DSVKGLLEEGVI

					CCCCTGGAAAATTCCAGCAAAGTACCAGGT CAGCACCGCTGGGTCGAGGAGCCAACCCGT TTCCTATCCGGGGCGGACTATATGAATTGTA TACGTGCGAGGATCGCAGCTCTTCCGACTG CAGCCAGGTGTGCTAGGGGACGTCTCAAAG ACAAGCATTGCCGGGCAGGTTGCGGAAATG TGGAGACGCTTAACCACGTCTTGCAATTCTG CCACCGTACCCATGGCACTCGCATTGGACG CCATGATGCGGTTGTAAAGTATGTTGTAGG AGGACTCAAGAAGAGGGGGTACGCAGTGA AAGAAGAGCCGAAAATCGTCTTACAGGATG TGGTGTACAAACCTGATATGGTTGCGACCA AGGAAGGGAAAACACTCATTCTGGACGCTC AGGTTCTAGGCGACCAGCGTGATATGAGAC TGGCACATGAAGATAAGCTCCGTAAGTATG GGGCCCCAGAATTTAAACGAAAGATCAGGA GTGAGACGGGGTCGGCAACCATTAAGTCCT TGTCGGTTACATTGAGCTGGCGAGGGTTGT GGGGACCTGACTCAGTAAAGGGGCTCCTCG AAGAGGGAGTGATTCTAAAGAAGGACCTTA AGATCTTGTCCACAAGAGTACTTATAGGAG CTTTGGCAGGCTGGAGAAGGTTTAATGAAA GGACGAGTATGGCAACATCTGGAAGAAGA			LKKDLKILSTRV LIGALAGWRRF NERTSMATSGR REEVTTRMVRR WKRRERVGVG (SEQ ID NO: 1476)

					GAAGAGGTTACAACAAGGATGGTGAGAAG GTGGAAGAGAAGAGAGAGGGTCGGTGTTG GTTAGCGGACTGGACTGTCTGGAGGAGTGT TTAACTCGGGTTCTCATGGGAACCCGACAA CGTTGTTATCTTGTATGACAATTCATAAAAA AAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1108)			
R2	R2La	JN937617		Lepidurus arcticus	GGGGTAGCAATTGATCGATTCCCGCTTCCTC GTGGCGCTACCCTGGGTAATACTATGAGGA ATTGATCACACCGTAGCGAACGTCATCAGT CACAGCTGCACGAATCCAGATAGAAATATA ACAGACGAGTAATTCTTTTAAAAGCTCGTC AGTAATCTTCCCAGATGTCTGGAAAGAGCA GCAAACCGAGAACTGTATCCAGTGGATCCA GTTCCCAAGAGACTCCTCCAAGTGGGTCCA ACGCATGTGATATATGCGGGAAATGCTTCA TGAAACCGGTCGGTTTGTCCAGAGTCCATC CTAGTCAATATCATGCACGCTTAGAAAAGA ACCAACCTAAAGCGAAGAAGTTCAGATGGA CTGATGAGGATTTGTATTTTTTAGCGAAAAA AGAAGCGGAACTTTTACACTTGGGCAGCAT TAAATTTGTGAATAAAGAACTGGCAGAATT CTTTCCGGAGAAATCTGTCGACCAAATCAG	GGGGTA GCAATT GATCGA TTCCCG CTTCCT CGTGGC GCTACC CTGGGT AATACT ATGAGG AATTGA TCACAC CGTAGC GAACGT CATCAG TCACAG	TGAT AATC GCTC CATC CTGC AACT AATT ATGA ATGC AAAT CTGTT AAGT GACA TTAG TGAT ACTT	MSGKSSKPRTVS SGSSSQETPPSGS NACDICGKCFM KPVGLSRVHPSQ YHARLEKNQPK AKKFRWTDEDL YFLAKKEAELL HLGSIKFVNKEL AEFFPEKSVDQI RGQRRSETYKQ QVLSIHSELLKL QTVADSPPPSRIP AKEVSAWLDFF LALPKTKNKFSE DKLDQLIRTAQ DGTLILDDLDLY

					AGGCCAAAGACGTTCTGAGACGTACAAACA	CTGCAC	ACCT	LREVLVQPTSQG
					GCAAGTTTTATCTATCCATTCTGAACTGCTC	GAATCC	GATA	EKQAKLLPPPKS
					AAGTTGCAGACTGTAGCCGATTCTCCTCCCC	AGATAG	CTTA	SREKRDREYAR
					CATCGAGAATCCCGGCGAAAGAAGTCTCCG	AAATAT	CCCT	AQNLYRKNKTA
					CTTGGCTTGACTTTTTTCTGGCCCTTCCCAA	AACAGA	GGTA	CVNAILDGNKK
					AACCAAGAATAAATTTAGTGAAGACAAGTT	CGAGTA	TTTAT	CENKIPDIDDFW
					AGACCAACTCATCCGTACCGCTCAAGATGG	ATTCTT	TTGA	KTIFESHSPPDAE
					TACTCTCATCCTGGATGATCTAGATTTATAT	TTAAAA	CCTA	PVCYVVDEEPT
					CTTAGAGAAGTTCTAGTTCAACCCACTAGTC	GCTCGT	TACTT	NIWSWISFFEMN
					AGGGAGAGAAACAAGCGAAGTTGCTCCCAC	CAGTAA	ACCC	HNYPDSSTSPGP
					CCCCCAAATCCAGTAGAGAGAAAAGAGACC	TCTTCC	TGGT	DGVTARMLRSIP
					GCGAATATGCCCGTGCCCAGAATTTATATC	CAG	ATCT	ARVLNKLLNLL
					GGAAGAATAAGACCGCTTGTGTGAACGCCA	(SEQ ID	ACCT	LFIEDLPAVFKC
					TCCTTGATGGTAACAAGAAGTGCGAAAATA	NO: 1232)	GACA	HRTVLIPKIDNP
					AAATACCTGACATTGACGATTTCTGGAAGA		TATA	TSPGEFRPITISSI
					CCATATTTGAGTCCCATTCTCCCCCAGATGC		TTTAT	VVRQLNKIIAAR
					AGAACCGGTGTGTTATGTCGTTGACGAGGA		CTAA	VSEGVPINPRQK
					GCCGACAAATATTTGGAGTTGGATTTCATTT		CCAC	AFRQIDGCAEN
					TTTGAGATGAATCATAACTACCCCGATTCGT		CTAC	VFLLDFILRDAK
					CTACCTCACCCGGTCCTGATGGCGTGACTGC		CTAT	TKIKSLSLATVDI
					TCGCATGCTCCGTTCCATCCCAGCACGTGTC		GATG	KKAFDSVSHHSI
					CTAAATAAATTGTTAAACCTTTTACTCTTTA		ACTC	FRAIRGARCPEN
					TTGAAGACTTACCTGCAGTGTTCAAGTGCC		CCGC	LVNYIQNSYSGC

					ATCGTACCGTGTTGATACCGAAAATCGACA		GGAA	TTQISVGGSISAS
					ATCCGACATCACCAGGCGAATTTCGTCCCA		ACTC	KIPMNRGVKQG
					TCACGATTTCCTCCATAGTTGTTCGGCAGCT		TCAC	DPLSPVLFNLVI
					GAACAAAATCATCGCTGCTCGCGTCTCAGA		TTAC	NEIIRKLPASIGY
					GGGCGTTCCCATAAATCCTCGCCAGAAGGC		CTTAT	PINSELSINCIAY
					CTTCCGGCAAATTGATGGATGCGCTGAGAA		TACC	ADDLILVTNTRE
					TGTTTTCCTTTTGGATTTCATTCTTCGTGACG		CACT	GLKLLLGLLNEE
					CAAAAACGAAAATTAAGTCCCTTTCGTTGG		TGGT	LPKRGLELNASK
					CTACTGTCGACATTAAGAAAGCATTTGATA		CTTTT	CFGLSLTALGKL
					GTGTGTCACATCATTCCATCTTCCGAGCGAT		ATTTT	KKTHLCTSDQL
					TAGAGGAGCGAGATGCCCTGAGAATCTAGT		CTCG	DLHGTLIKNLTA
					GAATTATATCCAGAACAGCTACTCGGGTTG		TTCCT	EESWVYLGVPF
					TACGACCCAAATTTCTGTTGGCGGCTCTATA		TATT	SHIGRSKSFSPDL
					TCAGCATCAAAAATCCCGATGAACCGTGGT		ACTTT	EALLNKLQKSPL
					GTCAAACAAGGGGACCCACTATCTCCTGTC		GTTC	KLQQKLFALRV
					CTCTTTAATTTAGTGATAAACGAAATTATTC		CTTTG	YLIPRLLHGLVL
					GCAAACTGCCGGCCTCGATTGGATACCCAA		GTGT	SRVAIGELKIMD
					TCAATAGTGAGCTTTCGATAAACTGTATTGC		AGGG	KLILKHLRVWL
					TTATGCGGATGACCTGATACTAGTTACAAA		TTCTC	RLPKDTPLGFFY
					TACTAGAGAGGGCCTGAAGCTACTACTCGG		TGGT	SPVKLGGLGIKN
					CCTATTGAATGAAGAGCTCCCTAAGCGAGG		TTTTG	LRTNVLKCRKQ
					ATTAGAATTGAATGCATCTAAGTGTTTCGGC		GAAC	RIERMLVSPDDV
					TTATCCTTAACAGCACTCGGGAAATTGAAG		GGCT	VRLVAESEIFLK

					AAGACTCACCTGTGTACCAGTGACCAACTC		TCCTT	ETDKLKDLLTIN
					GATTTACATGGGACTCTAATTAAGAATCTTA		AGCC	GMCLDXRNVPR
					CCGCTGAGGAATCCTGGGTTTATTTGGGTGT		GGAA	TGKNNKFWSER
					TCCTTTCTCACACATTGGGCGCTCTAAGAGC		TTTTG	LYTSFDGKTLA
					TTCAGTCCAGATCTTGAAGCTCTTTTAAATA		TCTG	YSEYFTQGGGW
					AACTCCAGAAATCGCCGCTTAAACTGCAAC		ATGT	IREDKILQPAHV
					AAAAGCTTTTTGCTTTGCGAGTTTATCTAAT		ATCTT	FAECIKLRINALP
					TCCTAGACTGCTCCACGGTTTAGTCCTATCA		GCTT	TKSRVAHGRPT
					AGAGTGGCCATTGGCGAATTAAAAATAATG		GTGT	KDRSCRAGCLD
					GATAAGTTAATTTTGAAACACTTACGCGTAT		CCTT	VQKVPTIETINHI
					GGCTTCGTTTACCGAAAGATACCCCTTTGGG		GAAA	AQVCPRTHGARI
					ATTTTTCTACTCACCGGTGAAATTGGGCGGT		TATA	KRHDRLVQFLS
					TTGGGGATTAAAAACTTACGTACAAACGTT		CGAC	LNLRKNPKRNV
					TTGAAGTGCAGGAAACAACGCATTGAGCGA		CCAG	LVEYNFRTVAGI
					ATGCTAGTCTCCCCTGATGATGTGGTGCGCT		GCTT	RKPDIIVIEDTRA
					TGGTTGCAGAGTCGGAGATTTTCCTGAAAG		GCGT	VILDVQVVGDSS
					AGACGGACAAGTTAAAGGACCTACTTACTA		CATTT	NLEMEYLEKSR
					TTAACGGAATGTGCCTAGATTANCGGAATG		AGGC	KYSNDANFINAL
					TGCCTAGAACGGGAAAAAATAACAAGTTTT		TCTG	QKLYPTVTNLTF
					GGAGTGAAAGGCTATATACGTCATTTGATG		GGAA	HAVTFNNRGLIA
					GTAAAACACTTGCCTATAGCGAGTACTTTA		A (SEQ	KSTVAALRMLG
					CTCAAGGCGGCGGCTGGATCCGTGAAGATA		ID NO:	VPPRCIMILCVIS
					AGATCCTTCAACCTGCCCATGTTTTCGCTGA		1355)	LEKTLEVWRMF

					ATGCATCAAGTTACGTATTAATGCTTTACCC ACTAAATCTAGAGTTGCACATGGCCGTCCA ACTAAAGATCGTTCCTGTCGTGCCGGATGTC TTGACGTTCAAAAAGTGCCAACTATCGAAA CAATCAATCATATAGCACAAGTTTGTCCTCG GACGCATGGTGCAAGAATAAAAAGACACG ATCGGCTCGTTCAATTCCTTTCTCTCAACCT TAGGAAAAATCCAAAAAGAAATGTTTTAGT GGAGTACAATTTCCGTACAGTCGCTGGTAT CCGTAAACCGGATATCATTGTTATTGAAGA TACCCGTGCAGTTATCCTTGACGTCCAAGTT GTTGGTGATTCGTCCAATTTGGAGATGGAA TACCTGGAGAAATCCAGGAAATATTCTAAT GACGCTAACTTTATCAATGCGCTTCAAAAA CTTTATCCAACGGTCACTAACCTAACTTTTC ATGCAGTTACCTTCAACAATAGAGGCCTGA TTGCCAAATCCACAGTGGCGGCCCTCCGGA TGCTCGGCGTCCCACCTAGATGCATAATGA TCCTTTGTGTGATTAGCCTTGAGAAGACCCT GGAAGTCTGGAGGATGTTTAACCAATCAAC TGCTTCCGCTCGGAAATGATAATCGCTCCAT CCTGCAACTAATTATGAATGCAAATCTGTTA AGTGACATTAGTGATACTTACCTGATACTTA			NQSTASARK (SEQ ID NO: 1477)

					CCCTGGTATTTATTTGACCTATACTTACCCT GGTATCTACCTGACATATATTTATCTAACCA CCTACCTATGATGACTCCCGCGGAAACTCTC ACTTACCTTATTACCCACTTGGTCTTTTATTT TCTCGTTCCTTATTACTTTGTTCCTTTGGTGT AGGGTTCTCTGGTTTTTGGAACGGCTTCCTT AGCCGGAATTTTGTCTGATGTATCTTGCTTG TGTCCTTGAAATATACGACCCAGGCTTGCGT CATTTAGGCTCTGGGAAA (SEQ ID NO: 1109)			
R2	R2LcA	.		Lepidurus couesii	TTTGGGGTAGCAATTGATCGATTCCCGCCTC CTCGTGGCGCTACCCCGGGATAGCCTCAAA GAAATTTGACGGTAAAGCAAAGAGGAATTG ATCACCCAAGGCAGTACATCGGCCTTCCTG CAGGAGCTCTGATAAAGATATTAGTGAGTT ATTCTGTTGAAGCTCGCTATTTCATTCCCCT GATGTCAGAGGAAAGCAGACCAAAGCAGA CGGCCTCCAAGCGTGGAGCCGCCGTGGAGA AGACTATGATGAGTGGGACCTACGTGTGTA CTTTATGTGGTAGGAGTTTTGAGAAGTCGGT GGGTTTGTCACTCCACACGAACCGGATGCA CCCTGAAGCTTACAATAAGCTGAAGGAGGC TAAGAAACCTGTTTTGAAGAAGGCGCGATG GTCAGAGGAGGAGGTTTTCCTTCTAGCGCA	TTTGGG GTAGCA ATTGAT CGATTC CCGCCT CCTCGT GGCGCT ACCCCG GGATAG CCTCAA AGAAAT TTGACG GTAAAG CAAAGA	TAGT GCTT GAGT GATG CCTA TCCTT TCTTT GATT AACT CTTA CCAT ATAC TTAC CAGT	MSEESRPKQTAS KRGAAVEKTM MSGTYVCTLCG RSFEKSVGLSLH TNRMHPEAYNK LKEAKKPVLKK ARWSEEEVFLL AQKEAELSFIGG IKFMNIELHKIFP ERELEGIKGQRK NPTYKAQVVSL LAEIRESKANDS SSSSSSSSSCDSA SLGISNWLEFLL

					GAAGGAAGCGGAGCTCTCATTCATTGGAGG	GGAATT	TCTTA	ALPKTSNQFQEG
					AATTAAGTTCATGAACATCGAGCTGCATAA	GATCAC	CCCG	RLDRLISDALRG
					AATTTTTCCAGAGCGTGAGTTGGAGGGAAT	CCAAGG	TACTT	VDVLENLDAYL
					CAAAGGCCAAAGAAAGAACCCAACGTACA	CAGTAC	ACCC	LEVFAKPMAQN
					AAGCGCAAGTCGTTAGCTTATTAGCTGAGA	ATCGGC	TGTA	PCPKPPPPAKNS
					TCCGAGAGTCTAAGGCCAATGATTCCTCTTC	CTTCCT	TACTT	RERRDREYSRV
					ATCTTCCTCATCCTCCTCCTCTTGTGATTCG	GCAGGA	ACCT	QNFYKKNRSACI
					GCCTCTTTGGGCATTAGCAATTGGCTTGAAT	GCTCTG	GTGT	NSILDGNTRSQN
					TTCTGCTAGCTCTCCCGAAGACCTCCAACCA	ATAAAG	GCGT	VIPGLTKFWTET
					ATTCCAGGAAGGTAGGCTAGATAGGCTAAT	ATATTA	ACCT	FEKNSPPDDEAP
					CTCTGATGCCCTGCGGGGTGTTGATGTGCTA	GTGAGT	GTGT	DQFVADEPRDM
					GAAAACCTTGATGCTTACCTTCTCGAGGTTT	TATTCT	ACTT	YKWITFYEMSQ
					TCGCTAAGCCCATGGCGCAAAATCCTTGTC	GTTGAA	GTCC	DYLDSSTAPGV
					CAAAGCCCCCACCCCCTGCGAAGAACAGTC	GCTCGC	TTTA	DGFSAKQLRSM
					GTGAGAGAAGGGATAGAGAATACTCTAGA	TATTTC	GCCG	SPRVLNKILNLL
					GTTCAGAATTTTTATAAGAAGAACCGTTCA	ATTCCC	CCTT	LLSENLPNSFKM
					GCGTGCATTAATTCCATTTTAGACGGGAAC	CTG	GTGT	HKTVLIPKIDDP
					ACCCGGTCTCAAAATGTGATACCTGGCTTG	(SEQ ID	TTTTA	KSPGDFRPITISP
					ACGAAGTTTTGGACCGAGACGTTTGAGAAG	NO: 1233)	CCAT	VLARLLNKILAA
					AACTCGCCACCTGATGATGAAGCCCCTGAC		TGGT	RLSKLVPISQRQ
					CAGTTCGTTGCGGACGAGCCTAGAGACATG		ACTT	KAFLPVDGCGE
					TATAAGTGGATAACTTTTTATGAGATGTCTC		ACCT	NIFLLDYILRSSK
					AAGATTATCTTGATTCTTCCACTGCTCCTGG		TGTG	KSSKSVAMAVL

					AGTGGATGGTTTCTCTGCCAAACAGCTCCG		TGGT	DVKKAFDSVSH
					GTCCATGTCTCCCAGGGTGCTAAATAAGAT		TGCC	HSILRALNEAKC
					ATTGAACCTTTTACTTCTTAGTGAGAACTTA		CGAT	PINFINFVRNSY
					CCTAATTCGTTTAAGATGCATAAGACAGTTC		ACTT	DGCTTKLTCGG
					TTATACCGAAGATTGATGATCCAAAGAGTC		ACCT	TSFPDSVRMNR
					CTGGGGATTTTAGACCGATCACGATTTCCCC		TGTA	GVKQGDPLSPV
					TGTCCTTGCTCGCCTCCTTAACAAGATTCTC		TTTGC	LFNLIIDSAIRKL
					GCTGCTAGATTGAGCAAACTAGTCCCCATT		CTTGT	PDSIGYVIRDGL
					AGTCAGCGTCAAAAGGCTTTTTTGCCTGTTG		AATT	KINCLAYADDLI
					ATGGCTGTGGGGAGAATATTTTCCTTTTGGA		CTGC	LVASSRAGLKTL
					TTATATTCTTCGTAGCTCTAAGAAAAGTAGT		ATGA	LNIVAEHLSLRG
					AAATCAGTGGCGATGGCAGTTTTGGATGTT		TATTT	LDLNAAKCHGL
					AAAAAGGCATTTGATAGTGTATCGCATCAC		ATTG	SIIASGKAKTTY
					TCCATCCTTAGAGCCCTTAATGAGGCCAAG		TGTA	VSAADSLDLDG
					TGTCCTATTAACTTTATTAACTTCGTCAGGA		GGTT	QPIKNLGVLDT
					ATAGTTATGATGGATGCACTACCAAACTTA		CCTG	WTYLGIPFSHLG
					CTTGTGGGGGTACTTCCTTTCCGGACTCGGT		ATGC	RAEKVSPDLTNL
					ACGCATGAACCGAGGAGTGAAACAGGGAG		TTAC	LNKLQKAPLKL
					ATCCCCTATCGCCTGTGCTCTTTAATTTGAT		CTGA	QQKLYAVRNFV
					CATTGATAGTGCAATTAGGAAGTTACCTGA		TTTGT	IPRALHGLILSKT
					CTCTATCGGCTATGTGATAAGGGATGGCTT		CCCC	NLKELNTLDRAI
					AAAGATTAATTGCCTGGCATACGCTGATGA		CTCA	RVFLRTLLYLPK
					CCTGATTTTGGTAGCTTCAAGTCGTGCTGGC		TCAT	DTPLGFFHSPIKS

					CTTAAGACCTTACTGAATATTGTTGCTGAGC		CTTTA	GGLGITCFRTSV
					ATCTCTCGCTACGCGGGTTGGATTTGAATGC		GTTTC	LKCRLQRIARM
					TGCTAAATGTCATGGTCTGTCAATTATCGCC		GTTCT	RSSCDGVIQAVA
					AGTGGTAAGGCCAAAACTACTTACGTTAGT		ATTTC	ESDIFADEYAKL
					GCTGCGGATTCCCTAGATCTTGACGGTCAA		ACTC	RDLIRINGNVLD
					CCTATAAAGAATTTGGGAGTGCTGGACACG		CATT	TTESIKRYWAQ
					TGGACCTACCTCGGGATTCCATTTTCCCATC		ATGG	RLHSSVDGKTL
					TGGGTCGAGCTGAGAAAGTTTCTCCTGATTT		AGTT	AYMDYFPQGNL
					GACCAACTTACTTAACAAACTTCAGAAAGC		CCGT	WMSEDKVSQRS
					TCCCCTCAAACTCCAGCAGAAGTTATATGCT		TTGTT	YVFADCVKLRI
					GTACGTAACTTTGTTATCCCTAGAGCACTCC		TTTTG	NAIPTRVRVSRG
					ATGGCCTCATTCTATCGAAGACAAACCTAA		GTGG	RPNKEMCCRAK
					AAGAATTAAACACGCTTGATAGAGCAATTA		AGGT	CFDSQRMPAFES
					GAGTTTTTCTACGTACCTTGCTTTACTTACC		ACAG	LNHITQVCPRTH
					TAAAGATACGCCACTGGGATTTTTCCATTCC		CACC	GSRIQRHDKIAK
					CCTATCAAGTCCGGAGGACTTGGGATTACG		CTTTA	FLFKNLNNCPSR
					TGTTTTAGAACATCTGTCCTTAAGTGCCGGC		AGCT	SVLYEPHFVTVD
					TACAGAGAATAGCCCGAATGCGTAGTTCCT		GGAA	GLRKPDIIIYDDS
					GTGATGGAGTTATACAGGCAGTGGCGGAAT		TTGA	HMVVLDVQVVS
					CTGATATCTTCGCTGATGAATATGCCAAGTT		GTGA	DSANLEKEFECK
					GCGGGACCTTATTAGGATAAACGGCAATGT		GTTT	AKKYANDVALR
					GCTCGACACGACCGAGAGTATTAAGAGGTA		ATGT	SAMLIKYPFIKSF
					TTGGGCTCAACGTCTTCACTCTTCGGTAGAT		ACTTT	SFVAATYNNRG

					GGCAAGACTTTGGCCTACATGGATTATTTCC CTCAGGGTAATCTTTGGATGTCAGAGGATA AAGTGAGCCAGCGATCTTACGTCTTTGCTG ACTGTGTCAAGTTACGGATTAATGCCATCCC AACTCGTGTCCGTGTTTCTCGTGGTCGACCT AATAAGGAAATGTGCTGTAGAGCGAAATGT TTTGATTCCCAGAGAATGCCAGCGTTTGAAT CCTTGAACCACATAACGCAAGTGTGTCCGC GCACTCATGGCTCGCGAATCCAGCGGCATG ATAAGATAGCCAAGTTCCTTTTTAAGAACCT GAACAATTGCCCATCTCGCTCGGTTCTGTAT GAACCCCATTTTGTGACGGTCGATGGCCTG CGGAAGCCTGACATTATCATTTACGATGAT AGCCACATGGTGGTTCTCGATGTCCAGGTG GTTTCCGACTCTGCTAACCTGGAGAAGGAA TTTGAGTGCAAGGCGAAGAAATATGCCAAT GATGTTGCACTTCGATCTGCAATGCTAATTA AATACCCATTCATTAAGTCATTTTCTTTCGT TGCAGCTACCTACAACAACCGTGGTCTGAT AGCTAAATCCTCTGTCCAGGTCCTTCGCCAG CTGGGTCTCTCGCCGCGCTCCATAATGGTCT CCATCCTCATCTGCCTTGAGGGTACTTTGGA AACATGGCGCATCTTCAATCAGTCTACAAT		GGAT GGTT GTAA TAAA CTAC CCGG AGGC AT (SEQ ID NO: 1356)	LIAKSSVQVLRQ LGLSPRSIMVSIL ICLEGTLETWRI FNQSTMNAH (SEQ ID NO: 1478)

					GAATGCCCATTAGTGCTTGAGTGATGCCTAT CCTTTCTTTGATTAACTCTTACCATATACTT ACCAGTTCTTACCCGTACTTACCCTGTATAC TTACCTGTGTGCGTACCTGTGTACTTGTCCT TTAGCCGCCTTGTGTTTTTACCATTGGTACT TACCTTGTGTGGTTGCCCGATACTTACCTTG TATTTGCCTTGTAATTCTGCATGATATTTAT TGTGTAGGTTCCTGATGCTTACCTGATTTGT CCCCCTCATCATCTTTAGTTTCGTTCTATTTC ACTCCATTATGGAGTTCCGTTTGTTTTTTGG TGGAGGTACAGCACCCTTTAAGCTGGAATT GAGTGAGTTTATGTACTTTGGATGGTTGTAA TAAACTACCCGGAGGCAT (SEQ ID NO: 1110)			
R2	R2LcB	JN937619		Lepidurus couesii	TTTTGGGGTAGCAATTGATCGATTCCCGCCT CCTCGTGGCGCTACCCTGGGATAACCTCAA AGAAATTTGACGGTAAAGCTAAGAGGAATT GATCACACCGTGACGAATATCATCAGTCAC AGCTGCACGAATCCAGATAGATATATAACA GGCGAGTAATTCTTTTCGAAGCTCGTCAGTA ATCTTCCCAGATGTCTGGAAAGAGCAGCAA ACCGAGAACTGTATCCAGCGGATCCAGTTC CCAAGAGACTCCTCCAAGTGGGTCCAACGC ATGTGATATATGCGGGAAGTGCTTCATGAA	TTTTGG GGTAGC AATTGA TCGATT CCCGCC TCCTCG TGGCGC TACCCT GGGATA ACCTCA	TGAT AATC GCTC CATC CTGC AACT AATT ATGA ATGC AAAT	MSGKSSKPRTVS SGSSSQETPPSGS NACDICGKCFM KPVGLSLHMSK VHPTQYHARLE KNQPKAKKFRW TDEDLYFLAKK EAELLLLGGIKF MNKELAEFFPE KSVDQIKGQRRS

					ACCAGTTGGTTTATCACTACACATGTCCAAA	AAGAAA	CTGTT	ETYKQQVVSIHS
					GTCCATCCTACTCAGTACCATGCACGCTTGG	TTTGAC	AAGT	ELLKLQAVADS
					AAAAGAATCAACCGAAAGCCAAGAAATTTA	GGTAAA	GACA	PPPSRIPAKEVSA
					GATGGACGGATGAAGATTTGTATTTTTTAGC	GCTAAG	TTAG	WLDFLLALPKT
					GAAGAAGGAAGCTGAACTTTTACTTTTGGG	AGGAAT	TGAT	KNKFSEDKLDQ
					TGGCATTAAGTTCATGAACAAGGAACTGGC	TGATCA	ACTT	LIRTAQEGTPVL
					AGAATTCTTTCCGGAGAAATCTGTTGACCA	CACCGT	ACCT	NDLDLYLREVL
					AATCAAAGGCCAAAGACGTTCCGAGACATA	GACGAA	GATA	VQPTRQGERQA
					CAAACAGCAAGTTGTATCAATTCATTCTGA	TATCAT	CTTA	KPLPPPKSSREK
					ACTGCTTAAGTTGCAGGCTGTAGCTGATTCT	CAGTCA	CCCT	RDREYARVQNF
					CCTCCACCATCGAGAATTCCAGCGAAAGAA	CAGCTG	GGTA	YRKNKTACVNA
					GTCTCGGCCTGGCTTGATTTCCTTCTGGCCC	CACGAA	TTTAT	ILDGNKKCENKI
					TTCCCAAAACCAAGAATAAATTTAGTGAGG	TCCAGA	TTGA	PDIDEFWKAIFE
					ACAAGTTAGACCAACTTATACGTACCGCGC	TAGATA	CCTA	SQSPPDAEPVSY
					AAGAAGGCACTCCGGTCCTAAACGATCTAG	TATAAC	TACTT	VVDEEPKNIWS
					ATTTATATCTAAGAGAAGTTCTAGTTCAGCC	AGGCGA	ACCC	WISFFEMNRNYP
					CACAAGACAGGGAGAGAGACAGGCGAAGC	GTAATT	TGGT	DTSTSPGPDGVT
					CTCTCCCACCACCCAAATCCAGTCGAGAGA	CTTTTC	ATCT	ARMLRSIPARVL
					AAAGAGACAGAGAATATGCTCGTGTCCAGA	GAAGCT	ACCT	NKLLNLLLFIED
					ATTTTTATCGCAAGAATAAAACCGCATGTG	CGTCAG	GACA	LPAVFKCHRTV
					TGAACGCCATCCTTGATGGAAATAAGAAGT	TAATCT	TATA	LIPKVDNPALPG
					GCGAAAACAAAATACCTGACATTGACGAAT	TCCCAG	TTTAT	EFRPITISSIIVRQ
					TCTGGAAGGCCATATTTGAATCTCAATCTCC		CTAA	LNKIIAARVSEG

					TCCTGATGCAGAACCAGTGAGTTACGTTGT	(SEQ ID	CCAC	VPINPRQKAFRQ
					AGACGAGGAACCGAAGAATATTTGGAGTTG	NO: 1234)	CTAC	IDGCAENVFLLD
					GATTTCGTTCTTCGAAATGAATCGTAACTAC		CTAT	FILRDAKTKIKS
					CCTGATACGTCGACCTCACCCGGTCCAGAC		GATG	LSLATVDIKKAF
					GGCGTGACGGCTCGTATGCTCCGTTCCATCC		ACTC	DSVSHHSIFRAIR
					CTGCGCGTGTCCTAAATAAATTGTTAAACCT		CCGC	GARCPENLVNYI
					ATTACTCTTTATTGAAGACTTACCTGCAGTG		GGAA	QNSYSGCTTQIS
					TTCAAGTGCCATCGTACCGTTCTGATACCTA		ACTC	VGGSISTTKILM
					AAGTTGACAATCCAGCATTACCGGGAGAAT		TCAC	NRGVKQGDPLS
					TTCGGCCTATCACGATTTCTTCCATAATTGT		TTAC	PVLFNLVINEIIR
					TCGGCAGCTGAATAAAATCATCGCTGCTCG		CTTAT	KLPASIGYPINSE
					CGTCTCAGAGGGCGTTCCCATAAATCCTCGT		TACC	LSINCIAYADDLI
					CAGAAGGCCTTTCGGCAAATAGATGGATGT		CACT	LVANTREGLKL
					GCTGAGAATGTTTTCCTTTTAGATTTCATTC		TGGT	LLNLLNEELPKR
					TTCGTGACGCAAAAACGAAAATTAAGTCCC		CTTTT	GLELNASKCFGL
					TTTCGTTGGCTACTGTCGACATTAAGAAAGC		ATTTT	SLTALGKLKKT
					ATTTGACAGTGTGTCACATCATTCCATTTTC		CTCG	HLCTSDQLDLH
					CGAGCGATTAGAGGAGCGAGATGCCCTGAG		TTCCT	GTLIKNLTAEES
					AATTTGGTGAATTATATCCAGAACAGCTATT		TATT	WVYLGVPFSHI
					CGGGTTGTACGACCCAAATTTCCGTAGGCG		ACTTT	GRSKSFSPDLEA
					GCTCTATATCTACAACAAAAATCCTGATGA		GTTC	LLNKLQKSPLKL
					ACCGTGGTGTAAAACAAGGGGACCCACTAT		CTTTG	QQKLFALRVYLI
					CTCCTGTCCTCTTTAACTTAGTGATAAACGA		GTGT	PRLLHGLVLSRV

					AATTATTCGCAAGCTGCCGGCCTCGATTGG		AGGG	AIGELKIMDKLI
					ATATCCAATCAATAGTGAGCTCTCGATAAA		TTCTC	LKHLRVWLRLP
					CTGTATTGCCTATGCGGATGACCTGATACTA		TGGT	KDTPLGFFYSPV
					GTTGCAAATACTAGAGAGGGTCTGAAGCTA		TTTTG	KLGGLGIKNLRT
					CTACTCAACCTATTGAATGAAGAGCTCCCT		GAAC	NVLKCRKQRIER
					AAGCGAGGATTAGAATTGAATGCATCTAAG		GGCT	MLVSPDDVVRL
					TGTTTCGGCTTATCCTTAACAGCACTCGGGA		TCCTT	VAESEIFLKETD
					AATTGAAGAAGACTCACCTGTGTACCAGTG		AGCC	KLKDLLTINGM
					ACCAACTCGATTTACATGGGACTCTAATTA		GGAA	CLDXRNVPRTG
					AGAATCTTACCGCTGAGGAATCCTGGGTTT		TTTTG	KNNKFWSERLY
					ATTTGGGTGTTCCTTTCTCACACATTGGGCG		TCTG	TSFDGKTLAYSE
					CTCTAAGAGCTTCAGTCCAGATCTTGAAGCT		ATGT	YFTQGGGWIRE
					CTTTTAAATAAACTCCAGAAATCGCCGCTTA		ATCTT	DKILQPAHVFAE
					AACTGCAACAAAAGCTTTTTGCTTTGCGAGT		GCTT	CIKLRINALPTKS
					TTATCTAATTCCTAGACTGCTCCACGGTTTA		GTGT	RVAHGRPTKDR
					GTCCTATCAAGAGTGGCCATTGGCGAATTA		CCTT	SCRAGCLDVQK
					AAAATAATGGATAAGTTAATTTTGAAACAC		GAAA	VPAIETINHIAQV
					TTACGCGTATGGCTTCGTTTACCGAAAGATA		TATA	CPRTHGARIKRH
					CCCCTTTGGGATTTTTCTACTCACCGGTGAA		CGAC	DRLVQFLSLNLR
					ATTGGGCGGTTTGGGGATTAAAAACTTACG		CCAG	KNPKRNVLVEY
					TACAAACGTTTTGAAGTGCAGGAAACAACG		GCTT	NFRTVAGIRKPD
					CATTGAGCGAATGCTAGTCTCCCCTGATGAT		GCGT	IIVIEDTRAAILD
					GTGGTGCGCTTGGTTGCAGAGTCGGAGATT		CATTT	VQVVGDSSNLE

					TTCCTGAAAGAGACGGACAAGTTAAAGGAC CTACTTACTATTAACGGAATGTGCCTAGATT ANCGGAATGTGCCTAGAACGGGAAAAAAT AACAAGTTTTGGAGTGAAAGGCTATATACG TCATTTGATGGTAAAACACTTGCCTATAGCG AGTACTTTACTCAAGGCGGCGGCTGGATCC GTGAAGATAAGATCCTTCAACCTGCCCATG TTTTCGCTGAATGCATCAAGTTACGTATTAA TGCTTTACCCACTAAATCTAGAGTTGCACAT GGCCGTCCAACTAAAGATCGTTCCTGTCGT GCCGGATGTCTTGACGTTCAAAAAGTGCCA GCTATCGAAACAATCAATCATATAGCACAA GTTTGTCCTCGGACGCATGGTGCAAGAATA AAAAGACACGATCGGCTCGTTCAATTCCTTT CTCTCAACCTTAGGAAAAATCCAAAAAGAA ATGTTTTAGTGGAGTACAATTTCCGTACAGT CGCTGGTATCCGTAAACCGGATATCATTGTT ATTGAAGATACCCGTGCAGCTATCCTTGAC GTCCAAGTTGTTGGTGATTCGTCCAATTTGG AGATGGAATACCTGGAGAAATCCAGGAAAT ATTCTAATGACGCTACTTTATCAATGCGTAT CAATGCGCTTCAAAAACTTTATCCAACGGT CACTAGCCTAACTTTTCATGCAGTTACCTTC		AGGC TCTG GGAA A (SEQ ID NO: 1357)	MEYLEKSRKYS NDATLSMRINA LQKLYPTVTSLT FHAVTFNNRGLI AKSTVAALRML GVPPRCIMILCVI SLEKTLEVWRM FNQSTASARK (SEQ ID NO: 1479)

					AACAATAGAGGCCTGATTGCCAAATCCACA GTGGCGGCCCTCCGGATGCTCGGCGTCCCA CCTAGATGCATAATGATCCTTTGTGTGATTA GCCTTGAGAAGACCCTGGAAGTCTGGAGGA TGTTTAACCAATCAACTGCTTCCGCTCGGAA ATGATAATCGCTCCATCCTGCAACTAATTAT GAATGCAAATCTGTTAAGTGACATTAGTGA TACTTACCTGATACTTACCCTGGTATTTATT TGACCTATACTTACCCTGGTATCTACCTGAC ATATATTTATCTAACCACCTACCTATGATGA CTCCCGCGGAAACTCTCACTTACCTTATTAC CCACTTGGTCTTTTATTTTCTCGTTCCTTATT ACTTTGTTCCTTTGGTGTAGGGTTCTCTGGT TTTTGGAACGGCTTCCTTAGCCGGAATTTTG TCTGATGTATCTTGCTTGTGTCCTTGAAATA TACGACCCAGGCTTGCGTCATTTAGGCTCTG GGAAA (SEQ ID NO: 1111)			
R2	R2Nvec-A	.		Nematostella vectensis	GGTTGGGGCCTTCTCGTGGCGGAGTCGTGA GTAAGGGGTATAGGGGTAAGGGACACCAC GGACCGAGAACGGTTACCGCTCAAGGCGAG TGGTGGAAGGCATAAAATCGTAACGCCGCC CTCCGACCTGCTCCTGAAACTAATGCCAAC CAACTGACTGTGGGGCTAACCTCCCCAGAG	GGTTGG GGCCTT CTCGTG GCGGAG TCGTGA GTAAGG	TGAT GGTG GGTT ACTC GCCT CTGT	MLRGTGNMND KRDGSATADPT SALLGAVGDGS LVCNLCGLACK SRGGLSIHRRSK HATVYHAERQP

					TCAGGATGCTGAGGGGTACTGGGAATATGA	GGTATA	GTAA	APRAKARWTND
					ATGACAAGAGGGACGGATCGGCCACTGCTG	GGGGTA	CAGG	EMILVARKQIAS
					ATCCTACCTCAGCGCTTTTGGGGGCTGTCGG	AGGGAC	CAAA	EKSRCSAVVEG
					TGATGGCAGCCTGGTATGCAATCTGTGTGG	ACCACG	TGAA	MREAVPHRTFD
					CCTTGCTTGCAAGTCTCGTGGGGGGCTCTCG	GACCGA	AGCT	AVKSLKTKNRN
					ATTCATCGAAGGAGCAAGCATGCCACGGTG	GAACGG	GCGC	YTRILEQIRAECS
					TACCATGCGGAAAGGCAGCCCGCTCCGCGG	TTACCG	AAGC	EEEVIESGVLKD
					GCTAAGGCGAGATGGACGAATGATGAAATG	CTCAAG	AGTC	RTENVCVQTTS
					ATACTTGTGGCCAGAAAACAAATAGCTTCG	GCGAGT	GATG	NVPGSAGRAAS
					GAAAAGAGTAGGTGTAGCGCAGTAGTTGAA	GGTGGA	AGCC	VELEGNIQVGH
					GGGATGCGAGAAGCAGTTCCACATAGAACT	AGGCAT	AAAG	QLAQKTMAGN
					TTTGACGCCGTCAAAAGTCTCAAAACCAAG	AAAATC	CCGC	NSRKQPANHTN
					AATAGGAACTACACACGGATTCTGGAGCAG	GTAACG	ACAG	WAEFNIEEGNIT
					ATTCGCGCTGAATGCTCCGAGGAAGAGGTG	CCGCCC	CCCC	LRKSKRKANGM
					ATTGAATCAGGAGTCCTTAAGGATAGGACG	TCCGAC	CGAC	PDATHRPGPPTV
					GAGAACGTGTGTGTGCAAACAACATCTAAC	CTGCTC	TGGG	DSLKHPVCLLQ
					GTTCCCGGTTCGGCAGGACGGGCAGCATCG	CTGAAA	GTAC	GAADKRDEPHT
					GTGGAGCTAGAGGGTAATATCCAGGTGGGC	CTAATG	AGGG	VEQLYYNIEEG
					CATCAACTCGCCCAAAAAACCATGGCAGGC	CCAACC	CAGC	MPLAEEQQWSE
					AATAATAGTAGGAAACAGCCTGCCAATCAT	AACTGA	CCTG	KLFDAIDSSLLS
					ACCAACTGGGCTGAGTTCAATATTGAGGAG	CTGTGG	GGCT	VEVELGRIVPGC
					GGAAATATTACCCTCCGAAAAAGCAAAAGA	GGCTAA	ATGC	PDEETRQLIDRE
					AAGGCTAATGGTATGCCCGATGCGACCCAC	CCTCCC	CCGA	FLDFIHSYSREK

					CGCCCCGGCCCACCGACTGTCGACAGTCTT	CAGAGT	AGTC	PPQRGLAKSKPP
					AAGCACCCAGTATGCTTGTTGCAGGGTGCA	CAGG	TTTTG	PKGPKSLRRQQ
					GCCGACAAAAGGGACGAACCGCACACAGT	(SEQ ID	ACAG	YRQLQRLWDKN
					GGAACAATTGTATTATAACATTGAGGAAGG	NO: 1235)	GTCC	RSAAAEQALTG
					AATGCCTCTTGCCGAAGAGCAACAGTGGTC		AAAC	KWQEVRTAAG
					AGAAAAGCTTTTCGATGCAATCGACTCATC		TTCA	VPLSLMEVPWR
					TTTGTTGTCTGTCGAGGTCGAGCTGGGGAG		CCCTT	EIFETPSTTDVRE
					AATCGTTCCTGGTTGCCCCGACGAAGAGAC		GCCG	PAPAGPVLWQL
					CCGACAGCTCATTGATAGAGAATTCTTGGA		CCAG	LRPVTIAEVEDA
					TTTTATCCATTCCTATTCGAGGGAAAAGCCG		TAGG	ISSKKSASGPDG
					CCACAAAGAGGTCTCGCTAAGTCCAAACCA		GCAC	VPCAALQTMGA
					CCACCCAAAGGACCAAAGAGCCTCAGACGC		CAGG	ASLAAHFNLWL
					CAGCAATATCGGCAACTCCAGAGATTGTGG		GCCA	LAGTQPKRLTEC
					GACAAAAACCGATCCGCAGCAGCCGAACA		GGTG	RTIFVPKEVNTH
					GGCCCTAACAGGTAAATGGCAAGAAGTTCG		CAGG	LPLHHRPITIGSV
					GACTGCTGCTGGGGTCCCATTATCGTTGATG		TGGG	VVRLFHQILGKP
					GAGGTCCCGTGGCGGGAGATTTTCGAAACA		CGCT	MEAVLPLGSGQ
					CCATCTACGACGGACGTCCGAGAGCCGGCC		TTGTT	RGFRKGDGICQ
					CCAGCTGGGCCGGTTCTGTGGCAGCTTCTCA		CTATT	NIWLLHTLIRRS
					GACCAGTGACTATCGCAGAAGTAGAGGACG		TGGT	TDLLRPLKLVFL
					CAATTTCCTCCAAGAAATCAGCTAGCGGTC		TTCG	DVKKAFDSVSH
					CAGATGGAGTGCCTTGTGCCGCCCTTCAAA		CTTA	ESLLIAAKRLGV
					CCATGGGCGCCGCCTCCCTAGCAGCTCACTT		ATTTT	PGPLLTYINELY

					TAACTTGTGGCTGTTGGCCGGTACTCAGCCC		TGTT	SRSETVFEVGGE
					AAACGCCTAACAGAATGCCGAACAATATTT		AAAT	SSGSVKVSQGV
					GTGCCTAAGGAGGTAAATACTCACCTGCCC		TTTTC	KQGDPLSSTLFN
					CTCCATCATAGGCCAATCACTATCGGCTCG		CCGT	CVIDWAVSDLD
					GTGGTTGTAAGACTCTTCCACCAGATTCTGG		GCGC	PHIGVLLGESRV
					GAAAGCCAATGGAAGCAGTACTGCCATTGG		CCAC	SFLAYADDLVL
					GGTCGGGCCAGCGCGGTTTTCGAAAGGGTG		CCTTT	LSETEAALTSQL
					ACGGAATATGCCAGAACATCTGGCTCCTCC		TTAA	NSIEKSLAHCGL
					ATACTCTAATCAGACGCAGCACTGATTTGCT		CCCTT	KLSTGDSGKSAS
					GAGGCCACTCAAGCTGGTCTTCTTGGATGT		TTAA	LNIVIDGKAKR
					GAAAAAGGCCTTCGATTCGGTTAGTCACGA		CACC	WVVNPTPFLRA
					ATCCCTTTTAATTGCGGCGAAGCGTCTCGGA		AATT	SGGEIRSLVANE
					GTGCCCGGGCCGCTCCTAACGTACATCAAT		TTTTT	TYKYLGINIGAQ
					GAACTGTACTCTCGTTCGGAGACAGTCTTTG		TACA	GVKAAEYNAFK
					AAGTTGGGGGCGAGAGCAGTGGCTCTGTTA		ACCC	EALDNLSRAPLK
					AAGTATCCCAAGGGGTAAAGCAAGGGGATC		TCTA	PQQRLFLLKTYL
					CATTGAGTTCGACGCTCTTCAACTGCGTCAT		CTCA	LPQLHHSLVLSR
					TGACTGGGCGGTCTCAGATCTCGACCCCCA		ATAA	TTGKLLNSLDAL
					CATAGGCGTCTTACTAGGGGAGTCTCGAGT		TCCA	VRKAVRGWLKL
					CTCCTTTCTAGCTTACGCTGATGACCTGGTT		AATG	PHDTHRAFFYA
					CTCCTATCGGAGACCGAGGCTGCACTGACA		AATA	HQADGGLNVPS
					AGCCAGCTGAACTCAATTGAAAAGAGTCTA		AAAG	LYHLIPLLRRSR
					GCGCACTGTGGTCTCAAGCTCAGTACGGGG		CGGC	YERLTRVEDPEI

					GACAGTGGTAAATCTGCTTCGCTGAACATC		ATAA	REVSRTDYFKR
					GTGATTGATGGTAAAGCTAAGCGGTGGGTT		CAGG	VLGAAAAATTV
					GTAAACCCTACCCCGTTCCTACGTGCGTCTG		TGAA	AGHRIDSKVTLR
					GTGGCGAAATTCGATCTCTCGTCGCAAATG		CAC	LAWREALYASA
					AGACGTACAAGTACTTGGGCATTAACATCG		(SEQ	DGRGLSQCPLVP
					GAGCCCAGGGCGTGAAAGCTGCTGAATACA		ID NO:	EVHSWVTDVSG
					ATGCATTCAAAGAAGCACTCGATAACTTGT		1358)	LQTGSQYISAVR
					CTCGAGCTCCACTTAAGCCGCAGCAGAGGT			LRGALLPTAVR
					TATTCCTGCTGAAGACCTATCTGTTGCCGCA			KSRGRGGVNSN
					ACTACACCACTCACTGGTATTGTCAAGGAC			CDCCGRGQPEF
					AACTGGCAAACTATTGAACAGCCTCGATGC			LGHVLQTCPRT
					TCTCGTAAGAAAAGCTGTACGGGGATGGCT			WGSRISRHNHV
					TAAATTGCCACACGACACTCATAGGGCCTT			LSLIAKACRSRR
					CTTTTATGCTCACCAGGCCGACGGGGGACT			WQVLEEPIIQTP
					CAATGTCCCAAGTCTGTACCATCTAATACCT			AQLLKPDLVIW
					CTCCTAAGGAGGAGTAGGTATGAACGCTTG			NHQAAYVVDVS
					ACGAGGGTCGAAGACCCTGAGATAAGGGA			VPGDNTPLSTCH
					AGTGTCCCGAACCGACTACTTCAAAAGAGT			NRKVAYYSGES
					CTTGGGTGCCGCGGCGGCGGCCACCACGGT			VREWVRSKTGH
					TGCAGGGCATCGAATTGATTCCAAAGTAAC			NPTVSSVVINWR
					GTTGCGCCTAGCTTGGCGCGAGGCCCTTTAC			GAMAKESYRLL
					GCTTCAGCTGACGGGCGGGGTCTCTCCCAA			TKDLRLAKTLPR
					TGCCCGTTGGTACCCGAGGTTCACTCCTGGG			LLVLRVLEGGH

					TTACCGATGTTTCCGGTCTTCAGACCGGATC GCAGTACATCTCTGCTGTTCGCTTAAGGGG GGCCTTGCTTCCCACGGCGGTAAGGAAATC GAGGGGTCGCGGGGGTGTTAACTCGAATTG CGACTGTTGCGGTCGTGGACAACCAGAGTT TCTTGGGCACGTTCTCCAGACGTGTCCTCGG ACCTGGGGGTCTCGCATCAGTCGTCACAAC CACGTCCTCTCCCTTATCGCTAAAGCGTGCC GTTCGCGAAGATGGCAGGTGCTGGAAGAGC CGATCATCCAGACTCCTGCGCAGTTGCTCA AGCCAGACTTGGTGATCTGGAACCATCAAG CGGCGTACGTCGTTGACGTTAGCGTACCGG GTGATAACACCCCGCTCAGTACGTGTCACA ACAGGAAAGTCGCATATTACTCTGGTGAGA GCGTTCGTGAATGGGTGCGTTCCAAGACTG GGCATAATCCTACAGTGTCCTCTGTAGTGAT TAACTGGCGCGGTGCCATGGCAAAGGAGTC CTACCGCCTTTTGACGAAAGACCTTCGTCTA GCGAAGACCCTGCCTCGTCTGCTGGTACTA AGAGTTCTGGAAGGCGGCCATGGAATCTGG CTCAATTTCCATCGGTCGACCTTCGCCGTGG GAGTAACTTGATGGTGGGTTACTCGCCTCTG TGTAACAGGCAAATGAAAGCTGCGCAAGCA			GIWLNFHRSTFA VGVT (SEQ ID NO: 1480)

					GTCGATGAGCCAAAGCCGCACAGCCCCCGA CTGGGGTACAGGGCAGCCCTGGGCTATGCC CGAAGTCTTTTGACAGGTCCAAACTTCACCC TTGCCGCCAGTAGGGCACCAGGGCCAGGTG CAGGTGGGCGCTTTGTTCTATTTGGTTTCGC TTAATTTTTGTTAAATTTTTCCCGTGCGCCC ACCCTTTTTAACCCTTTTAACACCAATTTTTT TTACAACCCTCTACTCAATAATCCAAATGA ATAAAAGCGGCATAACAGGTGAACAC (SEQ ID NO: 1112)			
R2	R2Sm-A	.		Schistosoma mansoni	ATGTTTTAATTTATTTTTGAACTACTACTGT CTGAGTGCTTCTTACAACCTGAAGGCTCAG AAACTACCCACTTTTTGCTGTTTATCCACAA CAACAGTTGTGAATCTATTCTCCAAATATTC CTTGTGCTTTTGTCAACATTATTCTATACCA ACTGTACCACCTACTTCTTCATCTCACGTTT TAATTCTGGTCTAATTTTCTCATCATTAGTC ACGGAGAGGGCCTATGAACGGTCCGTGACG CGAAATTCAATCCACGAATTCGTCCTCTTCT GCTAGTGGTCCCCGAAATACGGTTCCTCTG GCCTGTCAGTTGTGTTAAAACTATATAATAA CGATGCCGGTCTCAACCGGCGCAGAAACTG ACATAACCTCTTCTTTGCCTATTCCTGCATC	ATGTTT TAATTT CCCC ATTTTT TTCA GAACTA CTCTT CTACTG AGAC TCTGAG ATTC TGCTTC CCCC TTACAA ACTG CCTGAA TTGTT GGCTCA GCTT GAAACT ATCTT ACCCAC CATG TTTTTG TTTTT	TAGC CCCC TTCA CTCTT AGAC ATTC CCCC ACTG TTGTT GCTT ATCTT CATG TTTTT	MPVSTGAETDIT SSLPIPASSIVSP NYTLPDSSSTCLI CFAIFPTHNILLS HATAIHHISCPPT PVQDGSQQMSC VLCAAAFSSNR GLTQHIRHRHIS EYNELIRQRIAV QPTSRIWSPFDD ASLLSIANHEAH RFPTKNDLYQHI STVLTRRTAEAV

					CTCAATCGTCTCGCCAAACTACACACTCCCT	CTGTTT	GTGT	KRRLLHLQWSR
					GATTCCTCTTCAACCTGCCTTATATGTTTCG	ATCCAC	TAAT	SPTAITTSSNNH
					CTATCTTCCCCACCCACAACATACTCCTCTC	AACAAC	TGAC	TTTDIPNTEARYI
					CCATGCCACTGCAATCCACCATATTTCTTGT	AGTTGT	TGCT	FPVDLDEHPPLS
					CCTCCTACTCCAGTGCAAGACGGTTCTCAGC	GAATCT	CTCTT	DATTPDASTHPL
					AGATGTCTTGTGTTCTTTGCGCCGCCGCTTT	ATTCTC	CTGG	PELLVILTPLPSP
					TTCATCTAACAGGGGACTAACACAACACAT	CAAATA	GTTG	TRLQNISESQTS
					TCGCCACCGGCACATCTCCGAATATAACGA	TTCCTT	ATGT	HESNRNSMHTP
					ACTAATCAGACAACGAATTGCAGTGCAGCC	GTGCTT	CTGA	PTYACDSDESLG
					GACGTCTCGCATATGGTCACCATTCGATGAT	TTGTCA	TTGTC	VTPSSTIPSCFHS
					GCTTCTCTACTATCAATCGCTAACCATGAAG	ACATTA	TCTCT	YRDPLAEQRSK
					CCCATAGATTCCCCACGAAGAATGACTTAT	TTCTAT	CTCTT	LLRASASLLQSS
					ACCAACACATCAGCACTGTATTAACACGCA	ACCAAC	TCCA	CTRIRSSSLLAFL
					GGACGGCAGAGGCCGTCAAACGCCGACTCC	TGTACC	TATT	QNASTLMDEEH
					TCCACCTACAGTGGTCCAGATCACCCACAG	ACCTAC	GCTT	VSTFLNSHGEFV
					CGATTACTACCTCTTCGAATAATCACACAAC	TTCTTC	GCTC	FPRTWTPSRPKH
					CACAGACATCCCCAATACCGAGGCCCGATA	ATCTCA	TCCC	PSHAPANVSRK
					TATTTTTCCGGTAGACCTAGACGAACATCCA	CGTTTT	CGCT	KRRKIEYAHIQT
					CCATTGTCTGATGCCACAACCCCCGACGCA	AATTCT	TACTT	LFHHRPKDAAN
					TCGACACATCCACTCCCAGAACTCCTTGTCA	GGTCTA	CCAA	TVLDGRWRNPY
					TCTTGACACCGCTTCCATCCCCGACTAGACT	ATTTTC	TAGT	VANHSMIPDFD
					ACAAAACATATCCGAATCACAGACCTCCCA	TCATCA	TGTC	CFWTTVFTKTN
					TGAATCCAATAGGAACTCAATGCATACACC	TTAGTC	ATAT	SPDSREITPIIPM

					GCCAACGTATGCCTGCGATTCGGATGAGTC	ACGGAG	TATG	TPSLIDPILPSDV
					ACTAGGGGTTACTCCCTCATCAACTATCCCC	AGGGCC	TCTTT	TWALKEMHGT
					TCATGCTTCCACAGTTATCGGGACCCCCTAG	TATGAA	GTTT	AGGIDRLTSYDL
					CTGAACAAAGAAGCAAACTCCTGAGGGCAT	CGGTCC	ACTT	MRFGKNGLAGY
					CCGCCAGCCTACTACAAAGCAGTTGTACTC	GTGACG	GCCA	LNMLLALAYLP
					GCATACGGTCCTCCAGCCTGCTCGCCTTCCT	CGAAAT	TGTCT	TNLSTARVTFVP
					CCAAAACGCATCCACATTAATGGACGAGGA	TCAATC	AACG	KSSSPVSPEDFR
					ACACGTGTCCACCTTCCTCAATAGTCATGGA	CACGAA	ACAA	PISVAPVATRCL
					GAATTCGTCTTCCCTAGAACATGGACCCCAT	TTCGTC	TTACT	HKILAKRWMPL
					CCCGACCCAAACACCCCTCCCACGCCCCAG	CTCTTC	TTATC	FPQERLQFAFLN
					CTAATGTTTCTAGGAAGAAAAGGAGGAAAA	TGCTAG	TACC	RDGCFEAVNLL
					TAGAGTACGCACACATCCAGACACTCTTCC	TGGTCC	TTAG	HSVIRHVHTRH
					ACCACCGTCCCAAAGATGCCGCCAACACCG	CCGAAA	TTGG	AGASFALLDISR
					TTCTAGACGGTCGGTGGAGAAACCCCTATG	TACGGT	TCCTC	AFDTVSHDSIIR
					TCGCAAACCATTCAATGATTCCAGACTTCG	TCCTCT	TTGG	AAKRYGAPELL
					ACTGCTTCTGGACAACAGTCTTTACTAAAAC	GGCCTG	TTTG	CRYLNNYYRRS
					AAATTCCCCAGACAGCCGGGAGATTACTCC	TCAGTT	GTTG	TSCVNRTELHPT
					AATCATCCCTATGACTCCCTCTCTCATTGAC	GTGTTA	CCTTC	CGVKQGDPLSP
					CCGATCCTCCCCTCTGACGTCACATGGGCGC	AAACTA	ATGT	LLFIMVLDELLE
					TGAAAGAAATGCATGGCACGGCCGGTGGGA	TATAAT	GTTC	GLDPMTHLTVD
					TTGATCGTCTAACATCGTACGATCTGATGAG	AACG	ATGG	GESLNYIAYAD
					ATTCGGGAAGAATGGTCTTGCTGGATATCT	(SEQ ID	CGGA	DLVVFAPNAEL
					CAACATGCTACTCGCTCTTGCATACCTTCCC	NO: 1236)	ATCT	LQRKLDRISLLL

					ACTAATCTCTCAACAGCACGGGTAACTTTC		GATG	HEAGWSINPEKS
					GTCCCCAAGTCATCAAGTCCTGTGTCACCTG		TTTAT	RTLDLISGGHSK
					AGGACTTCCGTCCCATCAGTGTCGCTCCAGT		AATG	ITALSQTEFTIAG
					AGCCACTAGGTGCCTGCACAAAATTCTAGC		ACTA	MRIPPLSAADTF
					AAAGAGATGGATGCCGCTCTTTCCACAGGA		TTCCT	DYLGIKSNFKGR
					ACGACTTCAGTTCGCTTTCCTAAACCGAGAT		ACTA	CPVAHIDLLNNY
					GGATGCTTTGAAGCAGTTAATCTTCTGCACT		CCAC	LTEISCAPLKPQ
					CGGTCATACGGCACGTCCACACCCGCCATG		CATT	QRMKILKDNLL
					CAGGAGCATCCTTCGCCCTGCTCGACATATC		ACAA	PRLLYPLTLGIV
					ACGGGCCTTTGACACTGTATCACATGACTCC		CTATT	HLKTLKSMDRN
					ATCATCAGAGCGGCGAAAAGATATGGGGCA		ATTA	IHTAIRKWLRLP
					CCTGAACTGTTATGCCGCTACCTCAATAACT		TTATC	SDTPLAYFHSPV
					ATTACCGACGTTCAACCAGCTGCGTCAACC		ACTA	AAGGLGILHLSS
					GCACTGAATTGCATCCTACGTGTGGGGTGA		TTATT	SVPFHRRKRLET
					AGCAAGGAGACCCCCTGTCGCCACTCCTCT		AACA	LLSSPNRLLHKL
					TCATCATGGTTCTCGACGAATTACTGGAAG		TTATT	PTSPTLASYSHL
					GTCTAGATCCAATGACCCACCTAACAGTTG		ATTA	SQLPVRIGHETV
					ATGGAGAGAGCTTGAACTACATAGCTTATG		CTTCT	TSREEASNSWV
					CTGACGATCTCGTAGTTTTCGCTCCAAATGC		ACAA	RRLHSSCDGKG
					AGAACTCCTTCAACGGAAACTCGATCGGAT		TTAG	LLLAPLSTESHA
					CTCCCTACTTCTACACGAGGCTGGATGGTCG		TATT	WLRYPQSIFPSV
					ATTAACCCTGAAAAAAGCCGGACCCTGGAC		ATGG	YINAVKLRGGL
					CTAATCTCTGGTGGCCATTCCAAAATCACA		CTAC	LSTKVRRSRGG

					GCGCTCTCTCAGACAGAATTCACCATCGCG		TCCTT	RVTNGLNCRGG
					GGGATGCGTATACCACCGCTTTCCGCCGCC		TCAG	CAHHETIHHILQ
					GACACCTTCGACTATCTGGGTATCAAATCC		CACA	HCALTHDIRCKR
					AACTTCAAGGGCCGATGCCCAGTGGCCCAT		CCAA	HNELCNLVAKK
					ATTGACTTATTGAACAACTACCTCACGGAA		TAAA	LRRQKIHFLQEP
					ATATCGTGCGCTCCACTTAAGCCGCAGCAG		ATCT	CIPLEKTYCKPD
					CGCATGAAGATCTTGAAAGATAATCTACTC		CAAT	FIIIRDSIAYVLD
					CCTCGACTCCTCTACCCCCTGACTCTAGGAA		CAAA	VTVSDDGNTHA
					TAGTACACCTGAAAACCCTGAAGTCAATGG		CATC	SRLLKISKYGNE
					ACCGAAATATCCACACGGCCATAAGGAAAT		TCAC	RTVASIKRFLTS
					GGTTGCGGCTACCCTCCGACACCCCGCTAG		TTATT	SGYIITSVRQTPV
					CATATTTTCACTCACCCGTCGCTGCCGGAGG		AAAC	LTFRGILERASS
					CCTAGGGATCCTCCATCTGTCCTCATCGGTT		TCTCT	QSLRRLCFSSRD
					CCATTCCACCGTCGAAAACGTCTAGAAACC		ATTTC	LGDLCLSAIQGS
					CTCCTATCTTCACCGAACCGCCTACTGCACA		CCCTT	IKIYNTYMRGTQ
					AGTTGCCAACTTCCCCAACACTAGCTTCTTA		CGTT	RLNE (SEQ ID
					TTCACACCTTAGTCAACTGCCAGTTCGAATT		ATAA	NO: 1481)
					GGGCACGAGACCGTAACGTCTAGAGAAGA		ACTT	
					GGCTTCCAACAGCTGGGTGAGACGATTACA		ACAA	
					TTCGTCCTGCGACGGGAAGGGACTACTCCT		TTCA	
					AGCACCACTAAGCACCGAGTCCCATGCATG		GTTT	
					GCTGCGCTACCCCCAGTCTATTTTTCCAAGT		AACC	
					GTTTACATCAACGCCGTTAAATTACGAGGT		GAAT	

					GGCTTACTATCCACCAAAGTCAGGAGATCT		ATCT	
					CGCGGAGGTAGAGTGACGAATGGCCTGAAC		CTCTT	
					TGTCGAGGCGGTTGCGCCCATCATGAAACA		TTAC	
					ATCCACCACATTCTGCAACATTGCGCGCTCA		AAAT	
					CCCATGACATCAGATGCAAACGCCATAACG		CTTA	
					AACTATGCAACCTTGTGGCAAAGAAACTGC		AGTA	
					GTAGGCAAAAAATCCATTTCTTACAGGAGC		TGTA	
					CCTGCATTCCTCTAGAAAAAACTTACTGCA		ATTTT	
					AACCTGATTTTATAATTATACGTGATTCAAT		GTGC	
					TGCTTATGTTCTAGACGTCACTGTATCGGAC		CAAG	
					GACGGAAACACCCACGCCAGCCGCCTGTTA		CCCA	
					AAAATATCAAAATACGGCAATGAGCGAACC		TTTG	
					GTCGCATCGATCAAGCGATTCCTCACATCC		GGTC	
					AGTGGATATATCATTACCAGTGTTCGACAA		TGTA	
					ACACCAGTCCTTACATTCAGAGGTATTCTGG		CAAT	
					AGAGAGCAAGTTCACAATCCCTACGACGCC		TTGA	
					TATGTTTTTCGTCCCGTGACCTCGGTGACCT		TACTT	
					TTGCCTGAGTGCGATTCAAGGCTCAATTAA		AAAA	
					AATATATAATACCTATATGAGAGGAACCCA		ATAA	
					ACGGCTGAACGAATAGCCCCCTTCACTCTT		ATGT	
					AGACATTCCCCCACTGTTGTTGCTTATCTTC		TAT	
					ATGTTTTTGTGTTAATTGACTGCTCTCTTCTG		(SEQ	
					GGTTGATGTCTGATTGTCTCTCTCTCTTTCC			

					ATATTGCTTGCTCTCCCCGCTTACTTCCAAT AGTTGTCATATTATGTCTTTGTTTACTTGCC ATGTCTAACGACAATTACTTTATCTACCTTA GTTGGTCCTCTTGGTTTGGTTGCCTTCATGT GTTCATGGCGGAATCTGATGTTTATAATGAC TATTCCTACTACCACCATTACAACTATTATT ATTATCACTATTATTAACATTATTATTACTT CTACAATTAGTATTATGGCTACTCCTTTCAG CACACCAATAAAATCTCAATCAAACATCTC ACTTATTAAACTCTCTATTTCCCCTTCGTTAT AAACTTACAATTCAGTTTAACCGAATATCTC TCTTTTACAAATCTTAAGTATGTAATTTTGT GCCAAGCCCATTTGGGTCTGTACAATTTGAT ACTTAAAAATAAATGTTAT (SEQ ID NO: 1113)		ID NO: 1359)	
R2	R2Tc	EU854578		Triops cancriformis	TTTTTGGTCTGGCATTTGATCGTTTCCGCCT CCTCGTGGCGCCAGACTGGGTAAGCTGATT TATCAGTGAGCTAAGAGAAACGATCACCGC AGGAGTCCATCTACCTGCGCTGCACGTGAT TTCATTGTGCACTTGGCGAGTTATCCCTTGT GGAGCTCGCTCGTTAGCTTTTGAGATGTCTC AAAAGCGCAGACCGGAGAAAGCAGTCCCC GACGAGGGAGCGACTGCTCATGATGTTGCG	TTTTTG GTCTGG CATTTG ATCGTT TCCGCC TCCTCG TGGCGC CAGACT	TAGA TGAC TGCC CTAC CCCTT TGCT GCCG AAGA	MSQKRRPEKAV PDEGATAHDVA QPDKSKCSVCG ETFKGPASVTM HMVKKHPVEFN ELKMAKKPVPK KVRWSEEEIFQL ARTEAELTLQG

					CAACCGGATAAGAGTAAGTGTTCTGTTTGT	GGGTAA	ACTA	VRFINVELQKIFP
					GGGGAGACTTTTAAGGGTCCTGCTTCCGTG	GCTGAT	CTGA	AREIEGIKGQRK
					ACAATGCACATGGTAAAGAAGCACCCAGTC	TTATCA	AGAC	LAKYKELVKDQ
					GAATTTAATGAGCTTAAGATGGCCAAGAAG	GTGAGC	TATT	LDEIGRAPNPPE
					CCTGTTCCGAAGAAAGTACGATGGTCTGAG	TAAGAG	GAAC	QEIGEDVPSPFK
					GAGGAGATCTTCCAACTTGCCAGGACAGAG	AAACGA	TAAC	AWLELLLALPK
					GCCGAACTTACCCTTCAGGGAGTTCGGTTTA	TCACCG	CAAC	TPNDFLEHKLD
					TTAATGTCGAGCTGCAGAAGATCTTCCCAG	CAGGAG	TGTG	NIIVQALKEDVN
					CACGCGAGATCGAGGGCATCAAAGGCCAG	TCCATC	TTAA	SDQVFNDLNSY
					AGGAAACTGGCTAAGTACAAAGAGCTCGTT	TACCTG	GTAA	LKLILEPSGRAK
					AAAGATCAGCTGGACGAAATCGGCCGTGCT	CGCTGC	GAAC	SVPGEIIHGDPSG
					CCTAACCCCCCTGAGCAAGAGATCGGAGAG	ACGTGA	TAAT	SAKTSVTKAPKP
					GATGTGCCTAGCCCGTTCAAGGCATGGCTG	TTTCAT	GCCT	ATVSSSRKKRR
					GAACTTCTGCTTGCGCTCCCCAAGACCCCCA	TGTGCA	CTTTT	DAEFARIQRLYR
					ATGATTTTCTAGAACACAAACTTGACAACA	CTTGGC	TCCCT	KNRTSCINTILD
					TTATTGTTCAAGCGCTGAAGGAAGACGTGA	GAGTTA	GCAT	GNTREHEAPKN
					ACTCCGATCAGGTATTCAATGACCTTAACTC	TCCCTT	GTAT	MEGFWREIFERE
					CTATTTAAAGCTGATTCTTGAGCCGAGTGGT	GTGGAG	CCCC	SPDDPDDPDIFL
					AGAGCCAAGTCAGTTCCCGGGGAAATTATA	CTCGCT	TGAT	EEEASDIWKYIS
					CACGGTGATCCCAGTGGGTCTGCTAAGACC	CGTTAG	CAGT	FYEMCNLYPPPS
					TCAGTAACCAAGGCTCCTAAGCCTGCTACA	CTTTTG	GACT	TAPGPDGFSSKD
					GTTTCTTCTAGCAGGAAAAAACGCAGAGAT	AG (SEQ	TATTT	LRRMTPRVLNKI
					GCCGAGTTTGCTAGGATCCAGCGTCTCTATC		TCTTT	LNLLLHLRDLPQ

					GGAAAAATCGGACATCATGTATAAATACAA	ID NO:	TCCT	ILKSHRTVLIPKT
					TCCTTGATGGCAATACAAGAGAGCATGAGG	1237)	GTTG	DLPTKPGDFRPI
					CCCCTAAGAACATGGAAGGATTTTGGAGGG		CGCC	TISNILVRHLNKI
					AAATATTCGAGAGGGAATCCCCGGATGACC		CTTTG	LANRVSHLIPIN
					CCGATGATCCCGACATTTTCCTTGAGGAAG		TTTA	ERQKAFLPIDGC
					AAGCCAGTGATATTTGGAAGTACATCTCTTT		GTTA	AENIFTLDFILHH
					CTATGAGATGTGCAACCTTTATCCGCCTCCC		TTTCC	ARTKIKSLSMAI
					TCCACTGCTCCTGGCCCCGATGGGTTTTCCT		TTTA	LDISKAFDSVSH
					CTAAGGACCTCCGGCGGATGACGCCAAGAG		ATTA	HSIFRALREARC
					TCCTCAACAAAATTTTAAACCTACTCCTTCA		CTAG	PIGFIKFIENCYG
					TTTGAGAGATCTTCCTCAGATCCTGAAATCC		AATT	GCFTKLFCGGV
					CATCGAACTGTACTAATACCAAAAACCGAT		ATCTT	KYPSEVSMNRG
					TTGCCTACTAAACCCGGGGACTTTCGTCCCA		TTCGT	VKQGDPLSPVLF
					TAACGATCTCAAATATACTAGTCAGACATC		TCTCC	NLVIDGLIRQIPS
					TTAATAAGATTCTGGCTAATAGAGTGAGCC		GTCT	ALGFNVSDQVK
					ACCTTATACCAATCAATGAGAGGCAGAAAG		AATT	VSCIAYADDLILI
					CTTTCCTACCAATTGATGGCTGCGCAGAAA		GCTTT	ATTRAGLKTLL
					ATATTTTCACTTTGGACTTCATACTCCATCA		TTCG	DLTNSYLAKRG
					TGCGCGGACCAAAATTAAATCGCTGTCAAT		GTGT	LSLNPDKCSALS
					GGCCATTCTAGATATTAGTAAGGCATTTGAT		AGGA	IVASGKQKLVYI
					AGCGTCTCGCATCATTCCATCTTCCGCGCTT		ACGG	ASSEHFDLAGQ
					TGCGAGAGGCTCGATGCCCCATTGGTTTTAT		CTAC	KMRNLNVGDS
					TAAATTCATTGAGAATTGTTATGGTGGATGT		CTAA	WRYLGIQFSHL

					TTTACCAAACTTTTCTGTGGTGGCGTGAAAT		AGCT	GRAEKVTPDLT
					ACCCGAGTGAGGTTTCTATGAACCGAGGAG		GGAA	CLINRLQKAPLK
					TAAAGCAGGGGGATCCGCTGTCACCAGTCC		GTGG	LQQKLYALRIYL
					TATTTAACTTAGTCATTGATGGTCTGATCAG		GAAG	IPRLIHGLTLSKT
					GCAAATTCCATCCGCTCTTGGATTCAACGTA		TGTTT	NLGELKTLDKLI
					AGTGATCAGGTGAAAGTCAGCTGTATCGCC		TCAA	RKYIRAWLHLP
					TACGCTGATGACCTGATCCTTATTGCGACCA		TGTA	DDTPMGYFYTP
					CAAGGGCTGGCCTGAAGACCCTGCTTGACT		CTTTG	LKAGGLGLPSLR
					TGACTAACTCCTACCTAGCAAAACGGGGAC		TGAT	LVILNNRLERIL
					TTTCGCTCAACCCGGACAAATGCTCGGCCCT		TATA	RMKASQDIIVRT
					ATCGATTGTGGCATCCGGCAAGCAGAAGCT		GAAA	IAESETLGVEIRK
					TGTATATATTGCGTCTTCCGAGCATTTTGAT		TATA	LHDLLSIDGTIL
					CTCGCAGGTCAGAAAATGCGCAATTTAAAC		TGAC	DTSVKIHSFWAE
					GTCGGTGACTCTTGGAGATACCTGGGTATTC		CCGA	RLYSSYDGKCL
					AATTCTCGCATCTCGGTAGAGCCGAAAAGG		GGTG	CNSANFPPGNK
					TCACTCCAGACCTAACATGTCTAATTAATAG		CATT	WIGEDSLNQRS
					ACTTCAAAAGGCCCCTTTAAAGCTTCAACA		GTTT	HIFADCLKLRIN
					GAAGCTTTATGCTCTCCGAATCTATCTTATA		GGCA	ALPTRSRTARGR
					CCCAGGCTTATCCATGGCCTCACCCTCTCTA		TTTCC	PLKDKPCRAGC
					AGACCAACCTCGGAGAGCTTAAAACGCTGG		TCGA	RNSDGVKVIETL
					ACAAGCTAATCCGCAAGTACATCCGCGCCT		GAAA	NHITQVCERTHG
					GGCTCCATCTCCCAGATGACACTCCGATGG		(SEQ	ARVKRHDRLVD
					GCTATTTCTATACCCCCTTAAAAGCTGGTGG			FAVKGLQRPHR

					ACTTGGCCTTCCTTCACTCCGTCTAGTGATC CTTAACAATCGACTTGAACGGATACTAAGG ATGAAGGCTTCTCAAGACATTATTGTCCGC ACTATTGCCGAATCGGAGACGTTGGGGGTC GAGATTCGAAAGCTTCACGATCTGTTGTCC ATTGATGGAACAATTCTGGACACGAGCGTT AAGATCCATAGCTTCTGGGCAGAAAGACTC TACTCGTCCTATGATGGTAAATGTCTTTGCA ATTCTGCAAATTTTCCCCCCGGAAACAAGT GGATCGGTGAAGATTCTCTGAACCAACGCT CGCATATCTTCGCTGATTGCCTAAAGCTCCG GATAAATGCGCTCCCAACACGTTCCAGAAC AGCACGTGGAAGGCCACTAAAAGATAAACC ATGTAGAGCTGGATGTCGGAACAGTGATGG TGTAAAGGTCATTGAGACACTGAATCACAT TACACAGGTGTGTGAGAGAACCCACGGAGC CCGCGTCAAGCGGCATGATCGCCTAGTGGA CTTCGCGGTGAAAGGACTACAAAGACCCCA TCGTGTGGTTCTAAAGGAGCCACATTACAA GACAGTAAACGGTGTTCGGAAACCGGACAT CGTCATTAAAATCCCCGACCACACATACAT TTGTGATTTCCAAGTTGTATCTGATACCTCT TGTTTAGAGTTGGAGTTCCGGAAGAAAGCC		ID NO: 1360)	VVLKEPHYKTV NGVRKPDIVIKIP DHTYICDFQVVS DTSCLELEFRKK ALKYAEDKGLC DQLTRDHPGELS FTAITFNTRGLIA KSSVTALRKLG MPPRSIMTLQKI CMEGSLEIWRIF NQTTAMARN (SEQ ID NO: 1482)

					CTCAAGTATGCGGAGGACAAAGGCCTCTGC GATCAACTTACCCGCGACCATCCTGGGGAG CTCTCATTCACCGCAATTACTTTCAATACAA GAGGGCTTATAGCCAAGTCTTCAGTCACCG CCCTCCGGAAATTGGGGATGCCACCCAGAT CTATCATGACCCTCCAGAAAATATGTATGG AGGGTTCTCTGGAGATCTGGAGGATTTTTA ACCAGACTACTGCTATGGCCCGCAATTAGA TGACTGCCCTACCCCTTTGCTGCCGAAGAAC TACTGAAGACTATTGAACTAACCAACTGTG TTAAGTAAGAACTAATGCCTCTTTTTCCCTG CATGTATCCCCTGATCAGTGACTTATTTTCT TTTCCTGTTGCGCCCTTTGTTTAGTTATTTCC TTTAATTACTAGAATTATCTTTTCGTTCTCC GTCTAATTGCTTTTTCGGTGTAGGAACGGCT ACCTAAAGCTGGAAGTGGGAAGTGTTTTCA ATGTACTTTGTGATTATAGAAATATATGACC CGAGGTGCATTGTTTGGCATTTCCTCGAGAA A (SEQ ID NO: 1114)			
R2	R2_DAn	.		Drosophila ananassae	AGAATATGGATTTGATTGTGCAGAGGGGGT GCTATACCGTAACTCGTAAGCCATGCAATC AGATCAAGTCGACTCAAAACCTCCTCGTGG TATTCTCTGGGTGCCAGTATTTACTGGTAGC	AGAATA TGGATT TGATTG TGCAGA	TAGC CAAT GCAC GGGT	FERRKDPWGYR PPGTLKQIGATE NNEPRNLNRFV RGESTASSLEST

					TGATTTGAGCGGCGAAAGGATCCTTGGGGT	GGGGGT	TCCA	QFGTSAEVNLA
					TACCGGCCCCCTGGAACCTTAAAACAAATT	GCTATA	GATT	GRVPCTICEMTF
					GGTGCAACTGAAAATAACGAGCCTCGGAAC	CCGTAA	AAGC	SSKRGLGVHMS
					CTAAATCGTTTTGTAAGAGGAGAATCCACG	CTCGTA	TTGCT	HRHKDDLDAQR
					GCTTCCAGCCTGGAGAGCACACAATTTGGA	AGCCAT	GCCG	LRVDKKARWSE
					ACCAGTGCAGAGGTTAACCTTGCAGGGAGG	GCAATC	AAGC	EETLMMARKEV
					GTGCCCTGTACGATATGTGAAATGACGTTC	AGATCA	ATAC	ELAASGVRFLN
					AGCTCCAAGAGGGGTTTGGGCGTTCACATG	AGTCGA	CATC	KKLAEIFTHRSA
					TCACATCGGCACAAAGACGATCTTGATGCA	CTCAAA	AAAA	DAISSYRKRSEY
					CAACGTCTTCGTGTCGATAAAAAGGCAAGG	ACCTCC	TCGG	KAKLEQIRGQSV
					TGGTCAGAGGAAGAAACCTTGATGATGGCG	TCGTGG	CATA	PTPEAEEINTTQ
					AGAAAGGAGGTTGAGCTTGCAGCAAGTGGT	TATTCT	AAAT	RRPSNSEQNRRV
					GTACGATTTCTTAATAAGAAGCTAGCGGAG	CTGGGT	TCGC	PRSEGGPIAPTE
					ATTTTCACCCACCGCAGTGCCGATGCGATAT	GCCAGT	TTAA	QTNNEILRVLQG
					CTTCGTATCGGAAGAGGAGTGAGTACAAGG	ATTTAC	TAAA	LAPVVCLPRWR
					CAAAACTAGAGCAGATAAGGGGGCAATCC	TGGTAG	GGAG	AEVLQNIVDNA
					GTTCCCACCCCAGAAGCAGAAGAAATCAAC	CTGA	GTGG	QVSGQETTLQSL
					ACCACACAGCGCCGCCCTAGTAATAGCGAG	(SEQ ID	TTTTA	SSYLMEIFPPRN
					CAAAACCGACGAGTACCAAGATCAGAAGG	NO: 1238)	GTAC	EPHILTRPRTEPR
					GGGACCAATCGCACCAACCGAACAGACGA		GTAG	NMRQRRRQQY
					ACAACGAAATCCTTAGGGTACTACAGGGTC		GCGT	ARVQRNWDKH
					TAGCACCTGTAGTATGCTTACCCCGGTGGA		CCCG	PGRCIKSLLEED
					GAGCCGAGGTCCTGCAAAATATCGTAGATA		GGAC	DESVMPNQEVM

					ATGCGCAGGTCTCGGGACAGGAAACCACTC		TTGTC	EPYWRRVMTQP
					TCCAAAGTTTATCCAGTTATCTCATGGAAAT		TCGG	SSSSIKRDMFNM
					TTTTCCGCCACGGAATGAACCGCACATTCTG		GATG	EHSLERVWSAV
					ACGAGGCCCCGAACGGAGCCTCGAAATATG		AATC	NQRDLRATKVK
					AGACAACGTAGAAGGCAGCAGTACGCGAG		GTGC	LSSSPGPDGITPK
					GGTTCAGCGTAACTGGGATAAACATCCGGG		ATGC	TARSVPEGIMLR
					GCGATGCATAAAGTCCCTACTGGAGGAAGA		GTAT	IMNLILWCGNLP
					TGATGAGTCGGTGATGCCAAACCAGGAGGT		AATT	YSIRLARTIFIPK
					CATGGAGCCATATTGGAGACGGGTAATGAC		GGGA	KATANQPQDYR
					TCAGCCTAGCTCAAGCTCGATAAAACGCGA		TCGA	PISVPSVIVRQLN
					CATGTTTAACATGGAGCATTCACTCGAGAG		TAAC	AILASRLSAAIN
					GGTATGGTCCGCTGTGAACCAGCGCGATCT		AAAT	WDTRQRGFLPT
					TAGGGCCACAAAAGTCAAATTATCTAGTTC		ACCA	DGCADNTTIVD
					TCCAGGCCCGGACGGGATCACTCCAAAAAC		ACTA	LVLREHHKRFK
					TGCCAGGAGTGTCCCCGAAGGCATTATGCT		AGTT	SCYIGTLDVSKA
					TCGCATAATGAACTTGATCCTCTGGTGCGG		ATTA	FDAVAHEAVYN
					GAATTTGCCGTACTCTATCCGTCTGGCCCGA		CTAA	TLASYGAPKGFI
					ACCATCTTCATTCCGAAGAAGGCGACGGCA		TATA	NYLRKAYEGGG
					AATCAACCGCAAGACTATCGTCCTATTTCA		TCGA	TMLAGNGWVSE
					GTCCCCTCGGTTATAGTTAGGCAACTAAAT		AATA	AFIPARGVKQG
					GCCATTTTGGCTTCCCGGTTGAGCGCAGCCA		CATA	DPLSPILFNLVID
					TCAACTGGGACACGCGTCAGCGAGGGTTCC		AATA	RLLRSLPSEIGA
					TACCTACCGATGGGTGTGCTGATAATACGA		TCCC	KVGNAMTNAA

					CGATTGTTGATTTAGTTTTGAGGGAACATCA		GTCC	AFADDIVLFAET
					TAAGCGATTTAAATCGTGCTACATCGGGAC		TTAC	PMGLQKLLDTT
					CCTCGATGTTAGTAAGGCCTTTGATGCTGTA		GTAT	VCFLSSVGLTLN
					GCTCACGAAGCGGTCTACAACACATTGGCT		CTTTG	TDKCFTVSIKGQ
					TCATATGGTGCCCCGAAAGGCTTCATCAAC		AAGA	AKQKCTVVERR
					TACTTACGGAAGGCGTACGAGGGCGGCGGC		TTTCC	SFLIGGRECPSL
					ACAATGCTCGCTGGGAACGGGTGGGTTTCA		ATCC	KRTDEWKYLGI
					GAGGCGTTCATTCCTGCCCGAGGAGTGAAG		TCAG	KFTAEGRARYD
					CAGGGTGACCCTCTGTCTCCCATACTATTCA		CGAA	PAEDLGPKLLRL
					ACTTGGTCATTGACCGGTTGCTTAGGTCCTT		CAAA	TRAPLKPQQKLF
					ACCCAGTGAGATTGGTGCCAAAGTCGGAAA		AAAA	ALRTVLIPQLYH
					TGCCATGACAAACGCAGCAGCATTCGCAGA		AAAA	KLTLGSVTIGVL
					TGATATAGTCCTTTTTGCGGAAACTCCGATG		AAA	KKFDKLVRYTA
					GGGCTTCAGAAATTGTTGGACACCACCGTT		(SEQ	RKWLGLPVDVP
					TGTTTCCTTTCCTCGGTGGGTCTCACCCTTA		ID NO:	VSFFHAPHKSGG
					ATACTGATAAATGTTTCACGGTCAGTATTAA		1361)	LGLPSLRWTAP
					GGGGCAAGCCAAACAAAAGTGTACCGTCGT			MLRLKRLSNIK
					CGAACGGCGAAGCTTCTTGATTGGCGGGCG			WPHLERSEVASS
					CGAGTGTCCTTCATTGAAGCGTACTGACGA			FVEEEMRRARD
					GTGGAAGTACTTAGGGATTAAATTCACTGC			RLQAGSEELLTR
					GGAGGGGCGGGCCCGGTACGATCCAGCAG			SQVDSYLANRL
					AGGACCTCGGTCCAAAGCTGTTGAGATTGA			HMSVDGCGLRE
					CTCGGGCCCCCCTGAAACCACAACAGAAGT			AERFAPQHGWV

					TATTTGCCCTTCGGACCGTCCTTATCCCACA ACTCTATCACAAGCTGACCCTTGGGAGTGT GACGATAGGCGTTCTGAAGAAATTTGACAA ATTAGTTCGATATACCGCACGGAAGTGGTT GGGGCTTCCGGTGGACGTACCAGTTTCTTTT TTCCATGCCCCCCACAAGAGTGGGGGTCTC GGGTTACCATCTCTAAGATGGACAGCACCA ATGCTTCGACTAAAGCGATTGAGCAACATA AAATGGCCTCACCTCGAGCGATCCGAGGTA GCCAGCTCTTTCGTGGAGGAAGAAATGCGG AGGGCCCGGGATAGGCTTCAGGCTGGAAGT GAAGAACTGTTAACCCGTTCGCAGGTAGAT TCGTACTTGGCAAATAGATTGCACATGTCTG TTGATGGTTGCGGGCTCCGTGAAGCAGAGC GTTTTGCTCCGCAACACGGGTGGGTTAGTC AGCCCACGCGTTTGCTAACAGGAAAGGAAT ATACTGATGGAATCAAACTGCGGATAAATG CCCTACCCTCAAGGTCTCGTACTACGAGGG GAAGGCACGAATTGGAGAGACGGTGCCGTG CAGGATGTGATGCTCCCGAAACAACAAATC ACATCTTGCAGCAATGCTATAGAACTCACG GGAGGAGGATAGCTCGGCACAACGGCGTA GTAAATTTTCTCAAGCGGGGACTTGAGCGA			SQPTRLLTGKEY TDGIKLRINALP SRSRTTRGRHEL ERRCRAGCDAP ETTNHILQQCYR THGRRIARHNG VVNFLKRGLER RGCVVHVEPSL QGETGLNKPDL VAIRQNRIYVID TQIVTDGHSLDQ AHQRKVGKYDT PDIRTNLRRSFG AFDIEFHSATVN WRGIWSGQSVK RLIASDLLSSGD SNIISVRVISGGL WSWRQFMYLS GYTRDWT (SEQ ID NO: 1483)

					AGAGGCTGCGTCGTTCATGTTGAACCAAGT CTGCAGGGCGAAACCGGACTGAATAAACCT GACCTGGTGGCTATCCGACAAAATCGCATT TATGTGATTGACACTCAGATTGTGACTGAC GGACATTCTCTCGACCAAGCGCACCAGCGT AAGGTCGGGAAGTACGATACACCGGACATA CGGACGAATTTGCGGAGATCTTTCGGTGCC TTTGACATTGAGTTCCATTCCGCCACTGTGA ACTGGAGGGGAATATGGAGTGGTCAATCAG TAAAACGGTTGATCGCTTCAGACCTCCTCA GCTCTGGTGATAGCAATATCATCAGTGTCC GGGTAATCAGTGGTGGTCTCTGGAGCTGGC GGCAGTTCATGTATCTGTCGGGGTACACTC GCGATTGGACTTAGCCAATGCACGGGTTCC AGATTAAGCTTGCTGCCGAAGCATACCATC AAAATCGGCATAAAATTCGCTTAATAAAGG AGGTGGTTTTAGTACGTAGGCGTCCCGGGA CTTGTCTCGGGATGAATCGTGCATGCGTATA ATTGGGATCGATAACAAATACCAACTAAGT TATTACTAATATATCGAAATACATAAATATC CCGTCCTTACGTATCTTTGAAGATTTCCATC CTCAGCGAACAAAAAAAAAAAAAA (SEQ ID NO: 1115)			

R2	R2_DM	X51967		Drosophila melanogaster	TTGGGGATCATGGGGTATTTGAGAGCAGAG	TTGGGG	TAGC	MTTRPSVDIFPE
					GGGGAGTATTCTTCTGTAATTCGTAAGTCAT	ATCATG	TAAA	DQYEPNAAATL
					ATCATATGATGTGCGGAAGGGGAATTTTAC	GGGTAT	TCGTT	SRVPCTVCGRSF
					TCTGTAACTCACAAGTCTCTCCTTTACTCAA	TTGAGA	TGGT	NSKRGLGVHMR
					GTCGACTCAAAACCTCCTCGTGGTGGTCCC	GCAGAG	TCAA	SRHPDELDEERR
					GGTAATGCTAAACTCGTTTAGCAGCTAATTT	GGGGAG	AACA	RVDIKARWSDE
					GAGCGGAAAAACTTTTCCGATGGGCTGGTT	TATTCT	TTTGC	EKWMMARKEV
					CCCCAGAGGAAATTTATTCATATTGGAACT	TCTGTA	TTGCT	ELTANGCKHIN
					ACAAGCACAAATAACGAGCCTCGGATACCT	ATTCGT	GTCTT	KQLAVYFANRS
					TTACACAATCTGATGACGACCCGACCCTCC	AAGTCA	GGCA	VEAIKKLRQRG
					GTGGATATCTTCCCGGAGGACCAATATGAA	TATCAT	TAAC	DYKEKIEQIRGQ
					CCAAACGCAGCGGCTACTCTATCTAGGGTT	ATGATG	ATCA	SALAPEVANLTI
					CCCTGCACAGTATGTGGCCGGTCCTTTAACA	TGCGGA	ATAA	RRRPSRSEQDHQ
					GCAAGAGAGGACTCGGTGTTCACATGCGAT	AGGGGA	AGGC	VTTSETTPITPFE
					CTCGGCACCCAGACGAACTTGATGAAGAAC	ATTTTA	ATAA	QSNREILRTLRG
					GTCGACGTGTCGATATAAAGGCAAGATGGA	CTCTGT	ACAT	YSPVECHSKWR
					GTGATGAAGAGAAGTGGATGATGGCGAGA	AACTCA	CGCA	AQELQTIIDRAH
					AAGGAGGTTGAGCTCACAGCAAATGGATGT	CAAGTC	AAAT	LEGKETTLQCLS
					AAACACATAAACAAGCAACTAGCGGTGTAT	TCTCCT	AATG	LYLLGIFPAQGV
					TTTGCAAACCGCAGCGTCGAAGCCATCAAA	TTACTC	GTTA	RHTLTRPPRRPR
					AAGCTAAGACAGAGGGGCGATTATAAGGA	AAGTCG	TAAT	NRRESRRQQYA
					GAAAATAGAGCAGATAAGAGGGCAATCAG	ACTCAA	TAAA	VVQRNWDKHK
					CTCTCGCCCCGGAAGTTGCTAATCTAACCAT	AACCTC	TGGC	GRCIKSLLNGTD

					AAGGCGCCGCCCTAGTAGAAGTGAGCAAGA	CTCGTG	TATG	ESVMPSQEIMVP
					CCACCAAGTAACAACATCGGAAACAACTCC	GTGGTC	AGGA	YWREVMTQPSP
					AATCACTCCCTTCGAACAGTCGAACAGGGA	CCGGTA	TGGT	SSCSGEVIQMDH
					AATTTTGCGGACACTACGCGGGTATAGCCC	ATGCTA	TTTA	SLERVWSAITEQ
					CGTAGAATGCCATTCCAAATGGAGAGCCCA	AACTCG	GTAC	DLRASRVSLSSS
					AGAGTTGCAAACTATCATTGATAGGGCACA	TTTAGC	GTAG	PGPDGITPKSAR
					TCTCGAGGGAAAGGAAACCACTCTCCAATG	AGCTAA	GCGT	EVPSGIMLRIMN
					CTTATCGCTATATCTCCTGGGAATTTTTCCG	TTTGAG	TGCG	LILWCGNLPHSI
					GCACAGGGTGTACGACACACACTGACGAGA	CGGAAA	GAAC	RLARTVFIPKTV
					CCTCCTCGGAGACCTCGGAACAGGAGAGAA	AACTTT	TTCG	TAKRPQDFRPIS
					AGCAGAAGGCAGCAGTACGCTGTCGTCCAG	TCCGAT	GTTC	VPSVLVRQLNAI
					CGTAACTGGGATAAGCATAAAGGAAGATGC	GGGCTG	ATAT	LATRLNSSINWD
					ATCAAGTCCTTGCTAAATGGAACTGATGAG	GTTCCC	AGAG	PRQRGFLPTDGC
					TCGGTAATGCCAAGCCAAGAAATAATGGTT	CAGAGG	CAAT	ADNATIVDLVL
					CCCTACTGGAGAGAAGTAATGACTCAGCCT	AAATTT	GAAT	RHSHKHFRSCYI
					AGCCCAAGCTCTTGCAGTGGAGAAGTGATA	ATTCAT	CGTG	ANLDVSKAFDS
					CAAATGGATCACTCGCTTGAGAGGGTATGG	ATTGGA	CATG	LSHASIYDTLRA
					TCTGCTATTACAGAGCAGGACCTTCGGGCG	ACTACA	CTAG	YGAPKGFVDYV
					TCAAGAGTCTCATTATCCTCGTCTCCGGGGC	AGCACA	GAAA	QNTYEGGGTSL
					CTGACGGGATAACTCCAAAATCTGCCAGGG	AATAAC	ACTG	NGDGWSSEEFV
					AGGTGCCGTCAGGTATTATGCTGCGCATAA	GAGCCT	ACCA	PARGVKQGDPL
					TGAACCTAATTCTATGGTGCGGTAATCTACC	CGGATA	CACA	SPILFNLVMDRL
					ACACTCCATACGACTGGCCAGAACCGTCTT	CCTTTA	CAGT	LRTLPSEIGAKV

					CATCCCGAAGACGGTGACGGCGAAGCGACC GCAAGACTTTCGTCCAATATCAGTGCCTTCA GTCCTGGTAAGACAGCTAAATGCAATATTG GCAACCCGGTTGAACTCATCAATCAATTGG GACCCGCGCCAGCGGGGCTTCTTACCAACC GACGGATGCGCCGATAATGCGACGATAGTC GACTTAGTCTTGAGGCATAGCCATAAGCAC TTTAGATCTTGCTACATCGCAAATTTAGATG TAAGCAAGGCATTTGATTCTCTATCACATGC ATCTATATACGACACCTTACGTGCTTATGGT GCGCCAAAGGGCTTCGTTGACTACGTACAG AACACGTACGAGGGCGGTGGTACCAGTCTC AATGGGGACGGTTGGAGTTCAGAGGAATTC GTCCCTGCTAGAGGAGTGAAGCAGGGTGAC CCTTTGTCTCCTATTCTATTTAACTTGGTAAT GGACAGGTTACTTAGAACCTTACCCAGCGA AATTGGTGCCAAAGTCGGAAATGCCATTAC TAACGCGGCCGCGTTTGCAGATGATTTGGT ACTATTTGCGGAAACTCGGATGGGGCTTCA AGTATTGTTGGACAAGACGTTGGATTTTCTA TCTATCGTCGGCCTCAAACTTAATGCCGACA AATGTTTTACCGTTGGCATTAAGGGCCAGC CGAAACAGAAGTGTACCGTGTTAGAGGCAC	CACAAT CTG (SEQ ID NO: 1239)	GTTG GCAG ACCT AGTA TCTTT CGAA GATT TCCA TACC TCCG CGAT CAAA AAAA AAAA AAAA AAAA AAA (SEQ ID NO: 1362)	GNAITNAAAFA DDLVLFAETRM GLQVLLDKTLD FLSIVGLKLNAD KCFTVGIKGQPK QKCTVLEAQSF YVGSSEIPSLKR TDEWKYLGINF TATGRVRCNPA EDIGPKLQRLTK APLKPQQRLFAL RTVLIPQLYHKL ALGSVAIGVLRK TDKLIRYYVRR WLNLPLDVPIAF VHAPPKSGGLGI PSLRWVAPMLR LRRLSNIKWPHL TQNEVASSFLEA EKQRARDRLLA EQNELLSRPAIE KYWANKLYLSV DGSGLREGGHY

					AGAGCTTCTACGTAGGCTCGAGTGAGATTC CATCACTGAAGCGCACGGACGAGTGGAAGT ACTTAGGCATCAACTTCACTGCAACCGGGA GGGTTCGATGCAATCCGGCCGAGGACATTG GTCCAAAGCTACAAAGATTGACAAAGGCCC CCCTCAAACCACAACAGAGGTTGTTCGCCC TTCGGACTGTCCTTATCCCACAGCTCTACCA CAAGTTAGCCCTTGGGAGTGTGGCGATAGG CGTCCTAAGAAAAACTGATAAACTTATAAG ATATTATGTGCGAAGATGGCTAAATCTTCC GCTGGATGTGCCGATAGCATTTGTTCATGCA CCCCCAAAAAGTGGAGGTCTCGGAATTCCA TCACTAAGATGGGTAGCTCCAATGTTAAGG CTAAGACGCTTGAGTAACATTAAATGGCCT CACCTCACGCAAAACGAGGTAGCCAGCTCT TTCCTCGAAGCCGAAAAACAACGGGCCCGA GATAGATTATTAGCTGAACAAAATGAACTG TTATCGCGTCCGGCAATAGAAAAATATTGG GCGAACAAGTTGTACCTCTCAGTTGATGGT AGCGGACTCCGTGAAGGCGGCCATTATGGC CCGCAACACGGGTGGGTTAGTCAACCCACG CGTTTATTAACAGGAAAGGAATATATGGAC GGTATTCGTCTGCGGATAAATGCCCTACCC			GPQHGWVSQPT RLLTGKEYMDG IRLRINALPTKSR TTRGRHELERQ CRAGCDAPETT NHIMQKCYRSH GRRVARHNCVV NRIKRGLEERGC VVIVEPSLQCES GLNKPDLVALR QNHIDVIDTQIV TDGHSMDDAHQ RKINRYDRPDIR TELRRRFEAAG DIEFHSATLNWR GIWSGQSVKRLI AKGLLSKYDSHI ISVQVMRGSLG CFKQFMYLSGFS RDWT (SEQ ID NO: 1484)

					ACAAAGTCTCGTACTACAAGGGGAAGGCAC GAATTGGAACGACAGTGTCGTGCAGGATGT GATGCTCCCGAAACAACAAACCACATAATG CAAAAATGCTACCGCTCGCATGGGAGGCGG GTAGCTAGACACAACTGCGTAGTAAATCGA ATCAAGCGGGGACTTGAGGAGAGAGGCTGC GTGGTCATTGTTGAACCAAGTCTGCAGTGC GAATCCGGCCTTAATAAACCAGACCTGGTG GCACTACGACAAAATCACATTGATGTGATC GACACACAAATTGTGACAGACGGACACTCT ATGGATGATGCGCACCAGCGCAAAATCAAT AGATACGACAGACCGGACATACGAACTGAA TTGCGTCGCAGATTCGAAGCCGCAGGTGAC ATTGAATTCCATTCTGCCACCCTGAACTGGA GGGGGATCTGGAGTGGTCAATCCGTTAAAA GATTGATAGCAAAGGGTCTCCTCAGCAAAT ATGATAGTCATATCATTAGCGTCCAGGTTAT GAGAGGCAGTCTCGGTTGTTTTAAACAGTT CATGTACCTGAGCGGGTTTTCCCGAGATTG GACTTAGCTAAATCGTTTGGTTCAAAACATT TGCTTGCTGTCTTGGCATAACATCAATAAAG GCATAAACATCGCAAAATAATGGTTATAAT TAAATGGCTATGAGGATGGTTTTAGTACGT			

					AGGCGTTGCGGAACTTCGGTTCATATAGAG CAATGAATCGTGCATGCTAGGAAAACTGAC CACACACAGTGTTGGCAGACCTAGTATCTTT CGAAGATTTCCATACCTCCGCGATCAAAAA AAAAAAAAAAAAAAAAA (SEQ ID NO: 1116)			
R2	R2_DPe	.		Drosophila persimilis	AAGATATGGATCTGAATAATAGCGTAGAAG GGGAGTCATTCCGTAATTCGTAAATCGTAA AAATCAGATCAAGTTGATTCAAGACCTCCT CGTGGTATCTTCTGGATGCTATTAGACTGAA GTTCTTTTGGTCTAATAGTAACTAACTTGAA CAGCGAAACAGTCCTATGGGGTTGCCAGCC CCTTGGACAGTTCAGTTTGATTGGCACTAAT ATGCAAAATACAACGCCTCGGATAATAAAC ACTAATTCGTTGACGAACCAAATCCCTACG GTCTCTAGCCTAGGGGCCCAATCTGAACAT AGTGCACAGGTTAACCCAAACAGTGGTTAC CAATGCACGATATGTGAATCGTCTTTCCGTA GCAAAAGCGGACTAGGCGTTCACATGTCAC GTCGGCACAAGGACGAGTTTGATCAACTTC GTCTGCGTACCGACCGTAAGGCACAATGGA GTGAGGAAGAGTTGAGTATGATGGCAAGAA AAGAGATCGAGCTCGCAGCAAATGGAGAA AGATATCTAAATAAGAAGCTAGCGGAAGTA	AAGATA TGGATC TGAATA ATAGCG TAGAAG GGGAGT CATTCC GTAATT CGTAAA TCGTAA AAATCA GATCAA GTTGAT TCAAGA CCTCCT CGTGGT ATCTTC TGGATG	TAGC CTAT ACAC TATG TTGG AGAG AAGA CGCT TGCT ACCT AGGC AAAA TGTG AAAT TAGG TATA AACA TCGT	SSFGLIVTNLNS ETVLWGCQPLG QFSLIGTNMQNT TPRIINTNSLTNQ IPTVSSLGAQSE HSAQVNPNSGY QCTICESSFRSKS GLGVHMSRRHK DEFDQLRLRTD RKAQWSEEELS MMARKEIELAA NGERYLNKKLA EVFTNRSVDAIK KCRQRERYKTKI EQLKGQAVPLP EALESETIQRRPS IRERDLLVTPPN TLGTTPTELSNS

					TTTACGAACCGTAGCGTCGACGCTATCAAG	CTATTA	GGTT	EILAVLQGYPPV
					AAATGTCGACAGAGGGAGAGATATAAGAC	GACTGA	GTAA	VCNDQWRVEVL
					CAAAATCGAACAGCTAAAGGGTCAAGCTGT	(SEQ ID	AACT	QSIVDGAQASG
					TCCTCTCCCAGAAGCATTAGAATCTGAAAC	NO: 1240)	TGAG	KEITLQRLSTYL
					CATACAGCGCCGCCCTAGTATACGCGAGCG		GTGG	MEVFPSQNDRPI
					AGATCTCCTAGTAACGCCACCTAACACTCTC		GTTTT	QTRPPRRPRNRR
					GGAACCACTCCAACCGAACTGTCGAACAGT		TAGT	QGRRQQYALTQ
					GAAATCCTGGCAGTACTACAGGGGTACCCA		ACGT	RNWDKHKGRCI
					CCTGTAGTATGCAATGACCAATGGAGAGTT		ATGC	KAILDGTEGTAT
					GAGGTTTTGCAATCCATCGTAGATGGTGCG		GTGA	MPSQGIMGSYW
					CAGGCCTCGGGTAAGGAAATTACTCTTCAG		TTACT	RQVMTQTSPTY
					CGCTTGTCTACTTACCTTATGGAAGTATTTC		TCGT	SGTNTTFRTEHP
					CCTCACAGAATGACCGCCCCATTCAAACGA		AATC	LEGVWSPITLGD
					GACCTCCACGGAGACCTCGTAATAGGAGAC		ATGA	LRVHRVSLTKSP
					AAGGTAGGAGACAGCAGTACGCCTTAACAC		ATCG	GPDGITPRTVRSI
					AGCGTAACTGGGACAAGCACAAAGGTCGTT		TGCA	PSGVMLRIMNLI
					GTATAAAAGCCATTTTGGATGGAACTGAGG		TGCT	LWCGKLPVSIRQ
					GGACAGCAACTATGCCAAGTCAAGGTATCA		AGTG	ARTIFIPKVGNA
					TGGGGTCCTATTGGAGACAAGTCATGACAC		GGGT	SRPQDFRPITVQ
					AAACAAGCCCAACATATAGTGGTACGAACA		TTGG	SVMVRILNAILA
					CCACGTTCCGGACGGAACACCCACTTGAAG		CCTC	SRLTSSVDWDP
					GGGTTTGGTCCCCGATAACACTAGGGGACC		CACT	RQRGFLPTDGC
					TAAGGGTACACAGAGTGTCATTGACGAAAT		AGTA	ADNTTIVDLILR

					CTCCAGGACCTGATGGAATTACTCCAAGAA CTGTCAGGAGTATTCCGTCAGGAGTTATGCT TCGCATAATGAACCTGATACTTTGGTGCGG AAAGTTGCCTGTCTCCATCCGACAGGCACG AACCATCTTCATTCCGAAGGTGGGGAATGC TTCTCGACCGCAAGACTTTCGTCCAATTACG GTACAATCTGTTATGGTAAGGATTTTAAATG CCATTTTGGCTTCCCGGTTGACCTCATCAGT CGACTGGGATCCGCGTCAGCGAGGTTTCCT TCCAACCGACGGATGTGCCGATAATACGAC GATAGTCGACTTAATCTTAAGGGATCACCA TAAACGTTGTAAATCACTTTATATCGCAACT TTAGATATAAGCAAAGCATTTGACTCGGTG TCTCATGCAGCAGTTAGCGCCACTCTAACTG CATATGGTGCCCCTAAAGAATTCGTTGACT ACGTACAAAATTCGTACGAGGTCTGTGGCA CAACGCTCAATGGGGACGGATGGAGATCAG AGGAATTCATACCTGCTCGAGGTGTCAGAC AGGGTGACCCGCTATCTCCCATAATATTCA ACTTGATCATCGATCAGTTGCTTAGGTCCTA CCCCAATGAGATTGGTGCCACAATCGGTGA TCACACAACAAACGCGGCCGCGTTCGCAGA TGATATTGTCTTATTTGCGGAAACTCGTTTA		TCTTT GAAG ATTTT CCTTC CTCA GCGA TCAA AAAA AA (SEQ ID NO: 1363)	DHHKRCKSLYI ATLDISKAFDSV SHAAVSATLTA YGAPKEFVDYV QNSYEVCGTTL NGDGWRSEEFIP ARGVRQGDPLS PIIFNLIIDQLLRS YPNEIGATIGDH TTNAAAFADDI VLFAETRLGLQT MLDTTVDFLSS VGLTLNSDKCFT VGIKGQPKQKC TVVIPETFRIGSR SCPALKRTDEW KYLGITFTAQGR TRYSPADDLGP KLLRLTRSPLKP QQKLFALRTVLI PQLYHKLTLGS VMIGVLRKCDIL VRSTVRKWLGL

					GGCCTTCAAACAATGCTAGACACGACTGTC			PLDVSTAFFHAP
					GATTTTCTATCTTCAGTCGGGCTTACCCTTA			HTYGGLGIPSVR
					ACTCGGATAAATGTTTTACAGTTGGAATAA			WVAPMLRMKR
					AGGGGCAACCGAAACAGAAGTGTACTGTGG			LSNIKWAHLAQ
					TCATCCCAGAGACCTTCCGTATCGGTTCGCG			SEAASSFLTDEL
					CTCGTGTCCTGCATTGAAGCGCACAGACGA			NKARGRTLAGL
					GTGGAAGTATTTAGGCATAACATTCACTGC			NELTSRTEIETY
					ACAAGGGAGGACCAGGTACAGTCCAGCCG			WANRLYMSVD
					ACGACCTAGGTCCGAAGCTGTTGAGGCTGA			GRGLREAGLFRP
					CAAGGTCCCCCCTAAAACCACAACAGAAAT			QHGWVCQPTRL
					TGTTCGCACTCAGAACAGTTCTTATCCCACA			LTGQDYRNSIKL
					ACTTTACCATAAGCTGACCCTAGGTAGTGT			RINALPSRSRTT
					GATGATAGGTGTTCTGAGAAAGTGTGACAT			RGRNELERQCR
					ACTGGTACGTTCGACCGTAAGGAAGTGGTT			AGCDAPETTNHI
					AGGGCTTCCTCTGGACGTGTCAACTGCATTC			LQNCYRTHGRR
					TTCCATGCTCCTCATACTTATGGGGGCCTCG			VARHNCVVNNL
					GAATCCCTTCAGTTCGTTGGGTAGCGCCAAT			KRILEEKGHTVH
					GCTACGTATGAAAAGATTGAGCAATATTAA			VEPSLQLETSVS
					GTGGGCCCACCTCGCGCAATCCGAGGCGGC			KPDLVCIRDNH
					CAGCTCATTTCTTACCGACGAATTGAATAA			ACVIDAQIITDG
					GGCCCGGGGTAGAACTCTGGCTGGACTGAA			LFLDDVHHRKV
					TGAGTTGACATCGCGTACAGAGATCGAAAC			EKYKRPEVISAL
					GTACTGGGCGAACAGGTTGTATATGTCTGTT			RREFGVSGNVE

					GATGGTCGCGGCTTACGTGAAGCGGGACTT TTTCGTCCCCAACACGGCTGGGTGTGTCAGC CCACGCGTTTGCTAACAGGTCAAGATTACC GAAACAGTATCAAGCTGCGAATAAATGCCC TACCATCGAGGTCTCGTACCACGAGGGGCA GAAATGAATTGGAACGGCAATGTCGTGCAG GTTGTGATGCTCCCGAAACAACAAACCACA TCCTGCAGAATTGTTACCGTACGCATGGGA GGCGGGTAGCAAGACATAACTGTGTAGTCA ATAACCTTAAGAGGATTCTTGAGGAGAAGG GCCACACAGTACACGTCGAACCAAGTTTGC AGCTGGAAACCTCGGTAAGTAAACCAGACC TGGTGTGTATCCGTGACAATCACGCTTGCGT GATTGATGCGCAGATTATAACGGATGGACT GTTTCTCGACGATGTGCACCATCGCAAAGTT GAGAAATATAAAAGACCGGAAGTTATATCT GCACTGCGGAGAGAATTCGGAGTGTCGGGC AACGTCGAAGTCCTATCCGCGACGTTAAAC TGGCGTGGGATCTGGAGCAATCAATCCGTT AGAAGATTGATAGCAAAGGGTCTCATCTCA TCCGGTGACAGCAATGTCATTAGCGCCAGA GTGGTAACAGGCGGACTATATTGCTTCAGA CAGTTCATGTATCTCGCAGGTTACACTCGAG			VLSATLNWRGI WSNQSVRRLIA KGLISSGDSNVIS ARVVTGGLYCF RQFMYLAGYTR DWT (SEQ ID NO: 1485)

					ATTGGACTTAGCCTATACACTATGTTGGAG AGAAGACGCTTGCTACCTAGGCAAAATGTG AAATTAGGTATAAACATCGTGGTTGTAAAA CTTGAGGTGGGTTTTTAGTACGTATGCGTGA TTACTTCGTAATCATGAATCGTGCATGCTAG TGGGGTTTGGCCTCCACTAGTATCTTTGAAG ATTTTCCTTCCTCAGCGATCAAAAAAAA (SEQ ID NO: 1117)			
R2	R2_DPs	.		Drosophila pseudoobscura	CAATTGGAAAGATATGGGTCTGAATAATAG CGTAGAAGGGGAGTCATTCCGTAATTCGTA AATCGTAAAAATCAGATCAAGTTGATTCAA GACCTCCTCGTGGTATCTTCTGGATGCTATT AGACTGAAGTTCTTTTGGTCTAATAGTAACT AACTTGAACAGCGAAACAGTCCTATGGGGC TGCCAGCCCCTTGGACAGTTCAGTTTGATTG GCACTAATATGCAAAATACAACGCCTCGGA TAATAAACACTAATTCGTTGACGAACCAAA TCCCTACGGTCTCTAGCCTAGGGGCCCAATC TGAACATAGTGCACAGGTTAACCCAAACAG TGGTTACCAATGCACGATATGTGAATCGTCT TTCCGTAGCAAAAGCGGACTAGGCGTTCAC ATGTCACGTCGGCACAAGGACGAGTTTGAT CAACTTCGTCTGCGTACCGACCGTAAGGCA	CAATTG GAAAGA TATGGG TCTGAA TAATAG CGTAGA AGGGGA GTCATT CCGTAA TTCGTA AATCGT AAAAAT CAGATC AAGTTG ATTCAA	TAGC CTAT ACAC TATG TTGG AGAG AAGA CGCT TGCT ACCT AGGC ATAA TGTG AAAT TAGG	SSFGLIVTNLNS ETVLWGCQPLG QFSLIGTNMQNT TPRIINTNSLTNQ IPTVSSLGAQSE HSAQVNPNSGY QCTICESSFRSKS GLGVHMSRRHK DEFDQLRLRTD RKAQWSEEELS MMARKEIELAA NGERYLNKKLA EVFTNRSVDAIK KCRQRERYKTKI EQLKGQAVPLP

					CAATGGAGTGAGGAAGAGTTGAGTATGATG	GACCTC	TATA	EALESETIQRRPS
					GCAAGAAAAGAGATCGAGCTCGCAGCAAA	CTCGTG	AACA	IRERDLLVTPPN
					TGGAGAAAGATATCTAAATAAGAAGCTAGC	GTATCT	TCGT	TLGTTPTELSNR
					GGAAGTATTTACGAACCGTAGCGTCGACGC	TCTGGA	GGTT	EILAVLQGYPPV
					TATCAAGAAATGTCGACAGAGGGAGAGATA	TGCTAT	GTAA	VCNDQWRVEVL
					TAAGACCAAAATCGAACAGCTAAAGGGTCA	TAGACT	AACT	QSIVDGAQASG
					AGCTGTTCCTCTCCCAGAAGCATTAGAATCT	GA (SEQ	TGAG	KEITLQRLSTYL
					GAAACCATACAGCGCCGCCCTAGTATACGC	ID NO:	GTGG	MEVFPSQNDRPI
					GAGCGAGATCTCCTAGTAACGCCACCTAAC	1241)	GTTTT	QTRPPRRPRNRR
					ACTCTCGGAACCACTCCAACCGAACTGTCG		TAGT	QGRRQQYALTQ
					AACAGGGAAATCCTGGCAGTACTACAGGGG		ACGT	RNWDKHKGRCI
					TACCCACCTGTAGTATGCAATGACCAATGG		ATGC	KAILDGTEGTAT
					AGAGTTGAGGTTTTGCAATCCATCGTAGAT		GTGA	MPSQGIMGSYW
					GGTGCGCAGGCCTCGGGTAAGGAAATTACT		TTACT	RQVMTQTSPTY
					CTTCAGCGCTTGTCTACTTACCTTATGGAAG		TCGT	SGTNTTFRTEHP
					TATTTCCCTCACAGAATGACCGCCCCATTCA		AATC	LEGVWSPITLGD
					AACGAGACCTCCACGGAGACCTCGTAATAG		ATGA	LRVHRVSLTKSP
					GAGACAAGGTAGGAGACAGCAGTACGCCTT		ATCG	GPDGITPRTVRSI
					AACACAGCGTAACTGGGACAAGCACAAAG		TGCA	PSGVMLRIMNLI
					GTCGTTGTATAAAAGCCATTTTGGATGGAA		TGCT	LWCGKLPVSIRQ
					CTGAGGGGACAGCAACTATGCCAAGTCAAG		AGTG	ARTIFIPKVGNA
					GTATCATGGGGTCCTATTGGAGACAAGTCA		GGGT	SRPQDFRPITVQ
					TGACACAAACAAGCCCAACATATAGTGGTA		TTGG	SVMVRILNAILA

					CGAACACCACGTTCCGGACGGAACACCCAC		CCTC	SRLTSSVDWDP
					TTGAAGGGGTTTGGTCCCCGATAACACTAG		CACT	RQRGFLPTDGC
					GGGACCTAAGGGTACACAGAGTGTCATTGA		AGTA	ADNTTIVDLILR
					CGAAATCTCCAGGACCTGATGGGATTACTC		TCTTT	DHHKRCKSLYI
					CAAGAACTGTCAGGAGTATTCCGTCAGGAG		GAAG	ATLDISKAFDSV
					TTATGCTTCGCATAATGAACCTGATACTTTG		ATTTT	SHAAVSATLTA
					GTGCGGAAAGTTGCCTGTCTCCATCCGACA		CCTTC	YGAPKEFVDYV
					GGCACGAACCATCTTCATTCCGAAGGTGGG		CTCA	QNSYEVCGTTL
					GAATGCTTCTCGACCGCAAGACTTTCGTCCA		GCGA	NGDGWRSEEFIP
					ATTACGGTACAATCTGTTATGGTAAGGATTT		TCAA	ARGVRQGDPLS
					TAAATGCCATTTTGGCTTCCCGGTTGACCTC		AAAA	PIIFNLIIDQLLRS
					ATCAGTCGACTGGGATCCGCGTCAGCGAGG		AAAA	YPNEIGATIGDH
					TTTCCTTCCAACCGACGGATGTGCCGATAAT		AAAA	TTNAAAFADDI
					ACGACGATAGTCGACTTAATCTTAAGGGAT		AAAA	VLFAETRLGLQT
					CACCATAAACGTTGTAAATCACTTTATATCG		AAAA	MLDTTVDFLSS
					CAACTTTAGATATAAGCAAAGCATTTGACT		AA	VGLTLNSDKCFT
					CGGTGTCTCATGCAGCAGTTAGCGCCACTCT		(SEQ	VGIKGQPKQKC
					AACTGCATATGGTGCCCCTAAAGAATTCGT		ID NO:	TVVIPETFRIGSR
					TGACTACGTACAAAATTCGTACGAGGTCTG		1364)	SCPALKRTDEW
					TGGCACAACGCTCAATGGGGACGGATGGAG			KYLGITFTAQGR
					ATCAGAGGAATTCATACCTGCTCGAGGTGT			TRYSPADDLGP
					CAGACAGGGTGACCCGCTATCTCCCATAAT			KLLRLTRSPLKP
					ATTCAACTTGATCATCGATCAGTTGCTTAGG			QQKLFALRTVLI

					TCCTACCCCAATGAGATTGGTGCCACAATC			PQLYHKLTLGS
					GGTGATCACACAACAAACGCGGCCGCGTTC			VMIGVLRKCDIL
					GCAGATGATATTGTCTTATTTGCGGAAACTC			VRSTVRKWLGL
					GTTTAGGCCTTCAAACAATGCTAGACACGA			PLDVSTAFFHAP
					CTGTCGATTTTCTATCTTCAGTCGGGCTTAC			HIYGGLGIPSVR
					CCTTAACTCGGATAAATGTTTTACAGTTGGA			WVAPMLRMKR
					ATAAAGGGGCAACCGAAACAGAAGTGTACT			LSNIKWAHLAQ
					GTGGTCATCCCAGAGACCTTCCGTATCGGTT			SEAASSFLTDEL
					CGCGCTCGTGTCCTGCATTGAAGCGCACAG			NKARGRTLAGL
					ACGAGTGGAAGTATTTAGGCATAACATTCA			NELTSRSEIETY
					CTGCACAAGGGAGGACCAGGTACAGTCCAG			WANRLYMSVD
					CCGACGACCTAGGTCCGAAGCTGTTGAGGC			GRGLREAGLFRP
					TGACAAGGTCCCCCCTAAAACCACAACAGA			QHGWVCQPTRL
					AATTGTTCGCACTCAGAACAGTTCTTATCCC			LTGQDYRNGIK
					ACAACTTTACCATAAGCTGACCCTAGGTAG			LRINALPSRSRT
					TGTGATGATAGGTGTTCTGAGGAAGTGTGA			TRGRNELERQC
					CATACTGGTACGTTCGACCGTAAGGAAGTG			RAGCDAPETTN
					GTTAGGGCTTCCTCTGGACGTGTCAACTGCA			HILQNCYRTHG
					TTCTTCCATGCTCCTCATATTTATGGGGGCC			RRVARHNCVVN
					TCGGAATCCCTTCAGTTCGTTGGGTAGCGCC			NLKRILEEKGHT
					AATGCTACGTATGAAAAGATTGAGCAATAT			VHVEPSLQLETS
					TAAGTGGGCCCACCTCGCGCAATCCGAGGC			VSKPDLVCIRDN
					GGCCAGCTCATTTCTTACCGACGAATTGAAT			HACVIDAQIITD

					AAGGCCCGGGGTAGAACTCTGGCTGGACTG AATGAGTTGACATCGCGTTCGGAGATCGAA ACGTACTGGGCGAACAGGTTGTATATGTCT GTTGATGGTCGCGGCTTACGTGAAGCGGGA CTTTTTCGTCCCCAACACGGCTGGGTGTGTC AGCCCACGCGTTTGCTAACAGGTCAAGATT ACCGAAACGGTATCAAGCTGCGAATAAATG CCCTACCATCGAGGTCTCGTACCACGAGGG GCAGAAATGAATTGGAACGGCAATGTCGTG CAGGTTGTGATGCTCCCGAAACAACAAACC ACATCCTGCAGAATTGTTACCGTACGCATG GGAGGCGGGTAGCAAGACATAACTGTGTAG TCAATAACCTTAAGAGGATTCTTGAGGAGA AGGGCCACACAGTACACGTCGAACCAAGTT TGCAGCTGGAAACCTCGGTAAGTAAACCAG ACCTGGTGTGTATCCGTGACAATCACGCTTG CGTGATTGATGCGCAGATTATAACGGATGG ACTGTTTCTCGACGATGTGCACCATCGCAA AGTTGAGAAATATAAAAGACCGGAAGTTAT ATCTGCACTGCGGAGAGAATTCGGAGTGTC GGGCAACGTCGAAGTCCTATCCGCGACGTT AAACTGGCGTGGGATCTGGAGCAATCAATC CGTTAGAAGATTGATAGCAAAGGGTCTCAT			GLFLDDVHHRK VEKYKRPEVISA LRREFGVSGNV EVLSATLNWRGI WSNQSVRRLIA KGLISSGDSNVIS ARVVTGGLYCF RQFMYLAGYTR DWT (SEQ ID NO: 1486)

					CTCATCCGGTGACAGCAATGTCATTAGCGC CAGAGTGGTAACAGGCGGACTATATTGCTT CAGACAGTTCATGTATCTCGCAGGTTACACT CGAGATTGGACTTAGCCTATACACTATGTTG GAGAGAAGACGCTTGCTACCTAGGCATAAT GTGAAATTAGGTATAAACATCGTGGTTGTA AAACTTGAGGTGGGTTTTTAGTACGTATGC GTGATTACTTCGTAATCATGAATCGTGCATG CTAGTGGGGTTTGGCCTCCACTAGTATCTTT GAAGATTTTCCTTCCTCAGCGATCAAAAAA AAAAAAAAAAAAAAAAAA (SEQ ID NO: 1118)			
R2	R2_DSe	.		Drosophila sechellia	GGGATCAGGGGTAATTGCGAGCAGAGGGG GAGTATTTTTCTGTAATTCGTAAGTCATATC ATATGGTGTGCGGAAGGGGAATTTTACTCT GTAACTCACAAGTCTCTCCTTTACTCAAGTC GACTCAAAACCTCCTCGTGGTGGTCCCCGG TAATGCTAAACTTGTTTAGCAGCTAATTTGA GCGGCAAAACTTTTCCGATGGGTTGGTTCCC CAGAGGAAATTTATTCATATTGGAACTACA AATAGAAATAACGAGCCTCGGAGCAATTTA CGCAATCTGATGACGACCCGACCCTCCGTG GACATCTTCCCGGAGGACCAATATGAACCA	GGGATC AGGGGT AATTGC GAGCAG AGGGGG AGTATT TTTCTG TAATTC GTAAGT CATATC ATATGG	TAGC TAAA ACGT TTGG TTCA AAAC ATTT GCTT GCTG TCTTG GCAT	FERQNFSDGLVP QRKFIHIGTTNR NNEPRSNLRNL MTTRPSVDIFPE DQYEPNAAATL SRVPCTVCGRSF NSKRGLGVHMR SRHPDELDEERR RVDIKARWSEE EKWMMARKEV ELTANGHKHIN

					AACGCAGCGGCTACTCTATCTAGGGTACCC	TGTGCG	AACA	KQLAVYFANRS
					TGCACAGTATGTGGCCGGTCCTTTAACAGC	GAAGGG	TCAA	VEAIKKLRQRG
					AAGAGAGGACTCGGTGTTCACATGCGATCT	GAATTT	TAAA	DYKEKIEQIRRQ
					CGGCACCCAGACGAACTTGATGAAGAACGT	TACTCT	GGCA	SALVPEVANLTI
					CGACGTGTCGATATAAAGGCAAGATGGAGT	GTAACT	TAAA	RRRPSRSEQNHQ
					GAGGAAGAGAAGTGGATGATGGCGAGAAA	CACAAG	CATC	VTTSETTPITPFE
					GGAGGTTGAGCTCACAGCAAATGGACATAA	TCTCTC	GCAA	QSNREILRTLRG
					ACACATAAACAAGCAACTAGCGGTGTATTT	CTTTAC	ATAA	YSPVECHSKWR
					TGCAAACCGCAGCGTCGAAGCCATCAAAAA	TCAAGT	TGGT	AQELQTIIDRAE
					GCTAAGACAAAGGGGCGATTATAAGGAGA	CGACTC	AATA	LEGKETTLQCLS
					AAATAGAGCAGATAAGAAGGCAATCAGCTC	AAAACC	TATA	LYLLGIFPAQGV
					TCGTCCCGGAAGTTGCAAATCTAACCATAA	TCCTCG	AATT	RHTLTRPPRRPR
					GGCGCCGCCCTAGTAGAAGTGAGCAAAACC	TGGTGG	GGCT	NRRESRRQQYA
					ACCAAGTAACAACATCAGAAACAACTCCAA	TCCCCG	ATGA	VVQRNWDKHK
					TCACACCCTTCGAACAGTCGAACAGGGAAA	GTAATG	GGAT	GRCIKSLLNGTD
					TTTTGCGGACACTGCGTGGGTATAGCCCCGT	CTAAAC	GGTT	ESVMPSQEVMV
					AGAATGCCATTCCAAATGGAGAGCCCAAGA	TTGTTT	TTAG	PYWREVMTQPS
					GCTACAAACGATCATTGACAGGGCAGAGCT	AGCAGC	TACG	PSSCSREVIQMD
					CGAGGGAAAGGAAACCACTCTCCAATGCTT	TAA	TAGG	HSLERVWSAITE
					ATCGCTATATCTCCTAGGAATTTTTCCGGCA	(SEQ ID	CGTT	HDLRASRISLSSS
					CAGGGTGTACGTCACACTCTGACGAGACCT	NO: 1242)	GCGG	PGPDGITPKTAR
					CCTCGGAGGCCTCGAAATAGGAGAGAAAGC		AACT	EVPSGIMLRIMN
					AGAAGGCAGCAGTATGCTGTCGTCCAGCGT		TCGG	LILWCGNLPHSI

					AACTGGGATAAGCATAAAGGAAGATGCATC		TTCA	RLARTVFIPKTV
					AAGTCCTTGCTAAATGGAACTGATGAGTCG		GATA	TAKRPQDFRPIS
					GTAATGCCAAGCCAAGAAGTAATGGTTCCT		GAGC	VPSVLVRQLNAI
					TACTGGAGAGAAGTAATGACTCAGCCTAGC		AATG	LATRLNSSINWD
					CCAAGCTCTTGCAGTAGAGAAGTGATACAA		AATC	PRQRGFLPTDGC
					ATGGATCACTCGCTTGAGAGGGTTTGGTCT		GTGC	ADNATIVDLVL
					GCTATTACAGAGCACGACCTTCGGGCATCA		ATGC	RHSHKHFRSCYI
					AGAATCTCATTATCTTCATCTCCGGGGCCTG		TAGG	ANLDVSKAFDS
					ACGGGATAACTCCAAAAACAGCCAGGGAG		AAAC	LSHASIYDTLRA
					GTGCCGTCAGGTATTATGTTGCGAATAATG		TGAA	YGAPKGFVDYV
					AACCTAATTCTATGGTGCGGCAATCTACCA		GTGT	QNTYEGGGTSL
					CACTCTATCCGACTGGCCAGAACCGTCTTCA		TGAC	NGDGWSSEEFV
					TCCCGAAGACGGTGACGGCGAAGCGACCGC		AGAC	PARGVKQGDPL
					AAGACTTTCGTCCAATATCGGTGCCTTCAGT		CTAG	SPILFNLVMDRL
					CCTGGTAAGACAGCTAAATGCAATATTGGC		TATCT	LRNLPSEIGARV
					AACCCGGTTGAACTCATCAATCAATTGGGA		TTCG	GNAITNAAAFA
					CCCGCGCCAGCGGGGCTTCTTACCTACCGA		ATAG	DDLVLFAETRM
					CGGATGTGCCGATAACGCGACGATAGTTGA		ATTTC	GLQVLLDRTLD
					CTTAGTCTTGAGGCATAGCCATAAGCACTTT		CATA	FLSLVGLKLNA
					AGATCTTGCTACATTGCAAATTTAGATGTAA		CCTC	DKCFTVGIKGQP
					GCAAGGCATTCGATTCTTTATCGCATGCATC		CGCG	KQKCTVLEAQS
					TATATATGACACCTTACGTGCTTATGGTGCG		ATCA	FYVGSREIPSLK
					CCAAAGGGCTTCGTTGACTACGTACAGAAT		AAAA	RTDEWKYLGIN

					ACGTACGAGGGTGGCGGTACCAGTCTCAAT		AAAA	FTATGRVRCNP
					GGGGACGGTTGGAGTTCAGAGGAATTCGTC		AAAA	AEDIGPKLQRLT
					CCTGCTAGAGGAGTGAAGCAGGGTGACCCT		AAAA	KAPLKPQQRMF
					TTGTCTCCTATTCTATTTAACTTGGTAATGG		AAAA	ALRTVLIPQLYH
					ACAGGTTACTTAGAAACCTACCCAGCGAAA		(SEQ	KLALGSVAIGIL
					TTGGTGCCAGAGTCGGAAATGCCATTACTA		ID NO:	RKTDKLIRYYV
					ACGCGGCCGCGTTTGCAGATGATTTGGTAC		1365)	RRWLNLPLDIPI
					TATTTGCTGAAACTCGAATGGGACTTCAAG			AFIHAPPKSGGL
					TATTGTTGGACAGAACGTTGGATTTTCTATC			GIPSLRWVAPM
					TCTCGTCGGCCTCAAACTTAATGCCGACAA			LRLRRLSNIKWP
					ATGTTTTACCGTTGGCATTAAGGGCCAGCC			HLTQNEVASSFL
					GAAACAGAAGTGTACCGTGCTAGAGGCACA			EAEKQRARDRL
					GAGCTTTTACGTAGGCTCAAGGGAGATTCC			LAEQNELLSRPA
					ATCATTGAAGCGAACGGACGAGTGGAAGTA			IEKYWANKLYL
					CTTAGGCATCAACTTCACTGCAACCGGGAG			SVDGSGLREAG
					GGTTCGATGCAATCCGGCCGAGGACATTGG			HWGPQHGWVN
					TCCGAAGCTACAAAGATTGACAAAGGCCCC			QPTRLLTGKEYI
					CCTCAAACCACAACAGAGGATGTTCGCCCT			DGIRLRINALPT
					TAGGACTGTCCTTATCCCACAGCTCTATCAC			KSRTTRGRHELE
					AAGTTAGCCCTTGGGAGTGTGGCGATAGGC			RQCRAGCDAPE
					ATCCTACGAAAAACTGACAAACTAATACGA			TTNHIMQKCYR
					TACTATGTGCGAAGATGGCTAAATCTTCCG			SHGRRIARHNC
					CTGGATATACCGATAGCATTCATTCACGCA			VVNRIKRGLEER

					CCCCCAAAAAGTGGAGGTCTCGGAATTCCA TCACTTAGATGGGTAGCTCCAATGTTAAGG CTAAGACGTTTGAGTAATATTAAATGGCCT CACCTCACGCAAAACGAGGTAGCCAGCTCT TTCCTCGAAGCAGAAAAACAACGGGCCCGA GATAGATTATTAGCAGAACAAAATGAATTG TTATCGCGTCCGGCAATAGAAAAATATTGG GCGAACAAATTGTACCTCTCAGTTGATGGT AGCGGACTCCGTGAAGCAGGCCATTGGGGC CCGCAACACGGGTGGGTTAATCAACCCACG CGTTTACTAACAGGAAAGGAATATATAGAC GGTATTCGTCTGCGGATAAATGCCCTACCC ACGAAGTCTCGTACTACAAGGGGAAGGCAC GAATTGGAACGACAGTGTCGTGCAGGATGT GACGCTCCCGAAACAACAAACCATATTATG CAAAAATGTTACCGATCGCATGGGAGGCGC ATAGCTAGACACAACTGCGTAGTAAATCGA ATCAAGCGGGGACTTGAGGAGAGAGGCTGC GTAGTCATTGTTGAACCAAGTCTGCAGTGC GAATCCGGTCTTAATAAACCAGACCTGGTG GCACTACGACAAAATCACATTGATGTGATC GACATACAAATTGTGACAGACGGACACTCT ATGGATGATGCACACCAGCGCAAAATCAAT			GCVVIVEPSLQC ESGLNKPDLVA LRQNHIDVIDIQI VTDGHSMDDAH QRKINRYDRPDI RTELRRRFEAAG DIEFHSATLNWR GIWSGQSVKRLI AKGLLSKYDSHI ISVQVMRGSLG CFKQFMYLSGFS RDWT (SEQ ID NO: 1487)

					AGATACGACAGACCGGACATACGAACTGAG TTGCGTCGCAGATTCGAAGCCGCAGGTGAC ATTGAATTCCATTCTGCCACCCTGAACTGGA GGGGGATCTGGAGTGGTCAATCCGTTAAAA GATTGATAGCGAAGGGTCTCCTCAGCAAAT ATGATAGTCATATCATTAGCGTCCAGGTTAT GAGAGGCAGTCTCGGTTGTTTTAAACAGTT CATGTACCTGAGCGGGTTTTCCCGAGATTG GACTTAGCTAAAACGTTTGGTTCAAAACAT TTGCTTGCTGTCTTGGCATAACATCAATAAA GGCATAAACATCGCAAATAATGGTAATATA TAAATTGGCTATGAGGATGGTTTTAGTACGT AGGCGTTGCGGAACTTCGGTTCAGATAGAG CAATGAATCGTGCATGCTAGGAAACTGAAG TGTTGACAGACCTAGTATCTTTCGATAGATT TCCATACCTCCGCGATCAAAAAAAAAAAAA AAAAAAAA (SEQ ID NO: 1119)			
R2	R2_DSi	.		Drosophila simulans	GGGATCTGGGGTAATTGCGAGCAGAGGGGG AGTATTTTTCTGTAATTCGTAAGTCATATCA TATGGTGTGCGGAAGGGGAATTTTACTCTG TAACTCACAAGTCTCTCCTTTACTCAAGTCG ACTCAAAACCTCCTCGTGGTGGTCCCCGGT AATGCTAAACTTGTTTAGCAGCTAATTTGAG	GGGATC TGGGGT AATTGC GAGCAG AGGGGG AGTATT	TAGC TAAA ACGT TTGG TTCA AAAC	TCLAANLSGKN FSDGLVTQRKFT HIGTTNTNNEPR ISLHNLMTTRPS VDIFPEDQYEPN AAATLSRVPCT

					CGGCAAAAACTTTTCCGATGGGCTGGTTAC	TTTCTG	ATTT	VCGRSFNSKRG
					CCAGAGGAAATTTACTCATATTGGAACTAC	TAATTC	GCTT	LGVHMRSRHPD
					GAACACAAATAACGAGCCTCGGATATCTTT	GTAAGT	GCTG	ELDEERRRVDIK
					ACACAATCTGATGACGACCCGACCCTCCGT	CATATC	TCTTG	ARWSEEEKWM
					GGATATCTTCCCGGAGGACCAATATGAACC	ATATGG	GCAT	MARKEVELTAN
					AAACGCAGCGGCTACTCTATCTAGGGTTCC	TGTGCG	AACA	GHKHMNKQLA
					CTGCACAGTATGTGGCCGGTCCTTTAACAG	GAAGGG	TCAA	VYFANRSVEAIK
					CAAGAGAGGACTCGGTGTTCACATGCGATC	GAATTT	TAAA	KLRQRGDYKEK
					TCGGCACCCAGACGAACTTGATGAAGAACG	TACTCT	GGCA	IEQIRGQSALVP
					TCGACGTGTCGATATAAAGGCAAGGTGGAG	GTAACT	TAAA	EVANLTIRRRPS
					TGAGGAAGAGAAGTGGATGATGGCGAGAA	CACAAG	CATC	RSEQNHQVTTSE
					AGGAGGTTGAGCTCACAGCAAATGGACATA	TCTCTC	GCAA	TTPITPFEQSNRE
					AACACATGAACAAGCAACTAGCGGTGTATT	CTTTAC	AATA	ILRTLRGYSPVE
					TTGCAAACCGCAGCGTCGAAGCCATCAAAA	TCAAGT	ATGG	CHSKWRAQELQ
					AGCTAAGACAGAGGGGCGATTATAAGGAG	CGACTC	TTAT	TIIDRAELEGKE
					AAAATAGAGCAGATAAGAGGGCAATCTGCT	AAAACC	ATAT	TTLQCLSLYLLG
					CTCGTCCCGGAAGTTGCAAATCTAACCATA	TCCTCG	AAAT	IFPAQGVRHTLT
					AGGCGCCGCCCTAGTAGAAGTGAGCAAAAC	TGGTGG	GGCT	RPPRRPRNRRES
					CACCAAGTAACAACATCAGAAACAACTCCA	TCCCCG	ATGA	RRQQYAVVQRN
					ATCACTCCCTTCGAACAGTCGAACAGGGAA	GTAATG	GGAT	WDKHKGRCIKS
					ATTTTGCGGACACTGCGTGGGTATAGCCCC	CTAA	GGTT	LLNGTDESVMP
					GTAGAATGCCATTCCAAATGGAGAGCCCAA	(SEQ ID	TTAG	SQEVMVPYWRE
					GAGCTACAAACGATCATTGACAGGGCAGAG	NO: 1243)	TACG	VMTQPSPSSCSG

					CTCGAGGGAAAGGAAACCACTCTCCAATGC		TAGG	EVIQMDHSLER
					TTATCGCTATATCTCCTGGGAATTTTTCCGG		CGTT	VWSAITEHDLR
					CACAGGGTGTACGACACACGCTGACGAGAC		GCGG	ASRISLSSSPGPD
					CTCCTCGGAGACCTCGGAATAGGAGAGAAA		AACT	GITPKSAREVPS
					GCAGAAGGCAGCAGTATGCTGTCGTCCAGC		TCGG	GIMLRIMNLILW
					GTAACTGGGATAAGCATAAAGGAAGATGCA		TTCA	CGNLPHSIRLAR
					TCAAGTCCTTGCTAAATGGAACTGATGAGT		GATA	TVFIPKTVTAKR
					CGGTAATGCCAAGCCAAGAAGTAATGGTTC		GAGC	PQDFRPISVPSVL
					CCTACTGGAGAGAAGTAATGACTCAGCCTA		AATG	VRQLNAILATRL
					GCCCAAGCTCTTGCAGTGGAGAAGTGATAC		AATC	NSSINWDPRQR
					AAATGGATCACTCGCTTGAGAGGGTTTGGT		GTGC	GFLPTDGCADN
					CTGCTATTACGGAGCATGACCTTCGGGCGT		ATGC	ATIVDLVLRHSH
					CAAGAATCTCATTATCTTCATCTCCGGGGCC		TAGG	KHFRSCYIANLD
					TGACGGGATAACTCCAAAATCTGCCAGGGA		AAAA	VSKAFDSLSHAS
					GGTGCCGTCAGGTATTATGTTGCGAATAAT		CTGA	IYDTLRAYGAPK
					GAACCTAATTCTATGGTGCGGTAATCTACC		CCAC	GFVDYVQNTYE
					ACACTCTATCCGACTGGCCAGAACCGTCTTC		ACGC	GGGTSLNGDGW
					ATCCCGAAAACGGTGACGGCGAAGCGACCG		AGTG	SSEEFVPARGVK
					CAAGACTTTCGTCCAATATCGGTGCCTTCAG		TTGG	QGDPLSPILFNL
					TCCTGGTAAGACAGCTAAATGCCATATTGG		CAGC	VMDRLLRNLPS
					CAACCCGGTTGAACTCATCAATCAATTGGG		CCTA	EIGAKVGNAITN
					ACCCGCGCCAGCGGGGCTTCTTACCTACCG		GTAT	AAAFADDLVLF
					ACGGATGTGCCGATAATGCGACGATAGTTG		CTTTC	AETRMGLQVLL

					ACTTAGTCTTGAGGCATAGCCATAAGCACT		GATA	DKTLDFLSLVGL
					TTAGATCTTGCTACATAGCTAATTTAGATGT		GATT	KLNADKCFTVGI
					AAGCAAGGCATTCGATTCTTTATCGCATGC		TCCA	KGQPKQKCTVL
					ATCTATATATGACACCTTACGTGCTTATGGT		TACC	EAQSFYVGSREI
					GCGCCAAAGGGCTTCGTTGACTACGTACAG		TCCG	PSLKRTDEWKY
					AATACGTACGAGGGTGGCGGTACCAGTCTC		CGAT	LGINFTATGRVR
					AATGGGGACGGTTGGAGTTCAGAGGAATTC		CAAA	CNPAEDIGPKLQ
					GTCCCTGCTAGAGGAGTGAAGCAGGGTGAC		AAAA	RLTKAPLKPQQ
					CCTTTGTCTCCTATTCTATTTAACTTGGTAAT		AAAA	RMFALRTVLIPQ
					GGACAGGTTACTTAGAAACCTACCCAGCGA		AAAA	LYHKLALGSVAI
					AATTGGTGCCAAAGTCGGAAATGCCATTAC		AAAA	GVLRKTDKLIRY
					TAACGCGGCCGCGTTTGCAGATGATTTGGT		AAAA	YVRRWLNLPLD
					ACTATTTGCTGAAACTCGAATGGGACTTCA		(SEQ	VPIAFIHAPPKSG
					AGTATTGTTGGACAAAACGTTGGATTTTCTA		ID NO:	GLGIPSLRWVAP
					TCTCTCGTCGGCCTCAAACTTAATGCCGACA		1366)	MLRLRRLSNIK
					AATGTTTTACCGTTGGCATTAAGGGCCAGC			WPHLTQNEVAS
					CGAAACAGAAGTGTACCGTGCTAGAGGCAC			SFLEAEKQRAR
					AGAGCTTCTACGTAGGCTCGAGGGAGATTC			DRLLAEQNELLS
					CATCACTGAAGCGAACGGACGAGTGGAAGT			RPAIEKYWANK
					ACTTAGGCATCAACTTCACTGCAACTGGGA			LYLSVDGSGLRE
					GGGTTCGATGCAATCCGGCCGAGGACATTG			AGHWGPQHGW
					GTCCAAAGCTACAAAGATTGACAAAGGCCC			VNQPTRLLTGK
					CCCTCAAACCACAACAGAGGATGTTCGCCC			EYIDGIRLRINAL

					TTAGGACTGTCCTTATCCCACAGCTCTATCA CAAGTTAGCCCTTGGGAGTGTGGCGATAGG CGTCCTACGAAAAACTGACAAATTAATAAG ATATTATGTGCGAAGATGGCTAAATCTTCC GCTGGATGTGCCGATAGCATTCATTCATGC ACCCCCAAAAAGTGGAGGTCTCGGAATTCC ATCACTAAGATGGGTAGCTCCAATGTTAAG GCTAAGACGATTGAGTAATATTAAATGGCC TCACCTCACGCAAAACGAGGTAGCCAGCTC TTTCCTCGAAGCCGAAAAACAACGGGCCCG AGATAGATTATTAGCAGAACAAAATGAATT GTTATCGCGTCCGGCAATAGAAAAATATTG GGCGAACAAATTGTACCTCTCAGTTGATGG CAGCGGACTCCGTGAAGCGGGCCATTGGGG ACCGCAACACGGGTGGGTTAATCAACCCAC GCGTTTACTAACAGGAAAGGAATATATAGA CGGTATTCGTCTGCGGATAAATGCCCTACCC ACGAAGTCTCGTACTACAAGGGGAAGGCAC GAATTGGAACGACAGTGTCGTGCAGGATGT GACGCTCCCGAAACAACAAACCACATAATG CAAAAATGTTACCGATCGCATGGGAGGCGC GTAGCTAGACACAACTGCGTAGTAAATCGA ATCAAGCGGGGACTTGAGGAGAGAGGCTGC			PTKSRTTRGRHE LERQCRAGCDA PETTNHIMQKC YRSHGRRVARH NCVVNRIKRGL EERGCVVIVEPS LQCESGLNKPDL VALRQDHIDVID IQIVTDGHSMDD AHQRKINRYDR PDIRTELRRRFE AAGDIEFHSATL NWRGIWSGQSV KRLIAKGLLSKY DSHIISVQVMRG SLGCFKQFMYL SGFSRDWT (SEQ ID NO: 1488)

					GTGGTCATTGTTGAACCAAGTCTGCAGTGC GAATCCGGCCTTAATAAACCGGACCTGGTG GCACTACGACAAGATCACATTGATGTGATC GACATACAAATTGTGACAGACGGACACTCT ATGGATGATGCACACCAGCGCAAAATCAAT AGATACGACAGACCGGACATACGAACTGAA TTGCGTCGCAGATTCGAAGCCGCAGGTGAC ATTGAATTCCATTCTGCCACCCTGAACTGGA GGGGGATCTGGAGTGGTCAATCCGTTAAAA GATTGATAGCAAAGGGTCTCCTCAGCAAAT ATGATAGTCATATCATTAGCGTCCAGGTTAT GAGAGGCAGTCTCGGTTGTTTTAAACAGTT CATGTACCTGAGCGGGTTTTCCCGAGATTG GACTTAGCTAAAACGTTTGGTTCAAAACAT TTGCTTGCTGTCTTGGCATAACATCAATAAA GGCATAAACATCGCAAAATAATGGTTATAT ATAAATGGCTATGAGGATGGTTTTAGTACG TAGGCGTTGCGGAACTTCGGTTCAGATAGA GCAATGAATCGTGCATGCTAGGAAAACTGA CCACACGCAGTGTTGGCAGCCCTAGTATCTT TCGATAGATTTCCATACCTCCGCGATCAAA AAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1120)			

R2	R2_DYa	.		Drosophila yakuba	GGGAACATGGGGTAAAGGTGAGTAGAGGG	GGGAAC	TAGC	FERRIFPKGLVP
					GGAGTATTTTTTATACTCTGCAACTCATAAG	ATGGGG	TTAA	LTKDNHIGTTNL
					TCTTGCCTTTACTCAAGTCGACTCAAAACCT	TAAAGG	AACG	QNEPRIFTNDLL
					CCTCGTGGTGTTTCCCGGTAATGTTAAACTT	TGAGTA	TTTG	TTRPSVDHVPED
					GTTTAGCAGCTAATTTGAGCGGCGAATCTTT	GAGGGG	GTTC	QYEPNAAATLS
					CCGAAAGGGTTGGTTCCCCTGACGAAGGAT	GAGTAT	ACAT	RVPCTVCDRSFN
					AATCATATTGGTACCACAAATTTACAAAAC	TTTTTAT	ACAT	SKRGLGVHMRS
					GAGCCTCGGATATTTACTAATGATCTGTTGA	ACTCTG	CTGC	RHPDELDEERRR
					CGACCCGACCCTCCGTGGATCACGTCCCGG	CAACTC	CTGC	VDIKARWSEEE
					AGGACCAATATGAACCAAACGCAGCGGCTA	ATAAGT	TGCC	KWMMARKEVE
					CTCTATCAAGGGTTCCCTGCACAGTATGTGA	CTTGCC	TTGG	LMANGFKHINK
					CCGGTCCTTTAACAGTAAGAGAGGACTCGG	TTTACT	CACA	QLAVYFANRSV
					TGTTCACATGCGATCTCGGCACCCAGACGA	CAAGTC	ATAT	EAIKKLRQRGD
					ACTTGATGAAGAACGTCGACGTGTCGATAT	GACTCA	CAAA	YKEKIEQIRGQS
					AAAAGCAAGGTGGAGTGAGGAAGAGAAGT	AAACCT	AAGG	ALAPEVANLTIR
					GGATGATGGCGAGAAAGGAGGTCGAGCTC	CCTCGT	CATA	RRPSRSEQDHQ
					ATGGCAAATGGTTTTAAACACATAAACAAG	GGTGTT	AACA	VPTSEASPITPLE
					CAACTAGCGGTGTATTTTGCAAACCGTAGC	TCCCGG	TCGC	QSNREILRTLRG
					GTCGAAGCCATTAAAAAGCTGAGACAGAGG	TAATGT	ACAT	YSPVVCPSKWR
					GGCGATTATAAGGAGAAAATAGAGCAGAT	TAAACT	ATTG	AQELQTIIDRAE
					AAGAGGGCAATCCGCTCTCGCCCCAGAAGT	TGTTTA	GTTA	FEGKETTLQCLS
					TGCTAATCTAACCATAAGGCGCCGCCCTAG	GCAGCT	TTTAC	LYLQGIFPVQGV
					TAGAAGTGAGCAAGACCACCAAGTACCAAC	AA (SEQ	GGCT	RHTLTRPPRRPR

					GTCAGAAGCATCTCCAATCACTCCGCTCGA	ID NO:	ATGA	NRRESRRQQYA
					ACAGTCGAACAGGGAAATTTTGCGGACGCT	1244)	GGAT	VIQRNWDKHKG
					GCGTGGGTATAGCCCCGTAGTATGCCCTTCC		GGTT	RCIKSLLNGTDE
					AAATGGAGAGCCCAAGAACTACAAACTATC		TTAG	SVMPSREFMEP
					ATTGATAGGGCGGAATTTGAGGGAAAGGAA		TACG	YWREVMTQPSP
					ACCACTCTCCAATGCTTATCGCTCTACCTCC		TAGG	SSCNGEVIRTDH
					AGGGAATTTTTCCGGTACAGGGTGTACGAC		CGTT	SLETVWSAITEQ
					ACACGCTGACGAGGCCTCCTCGGAGACCTC		GCGG	DLRASRVSLSSS
					GGAATAGGAGGGAAAGCAGAAGGCAGCAG		AACT	PGPDGVTPKTA
					TACGCTGTCATCCAGCGAAACTGGGATAAG		TCGG	REVPSGIMLRIM
					CATAAAGGAAGGTGCATTAAGTCCCTGCTT		TTCG	NLILWCGNLPHS
					AATGGAACTGATGAATCGGTAATGCCAAGC		GATA	IRLARTIFIPKTV
					CGAGAATTTATGGAGCCCTACTGGAGAGAA		GAGC	TAKRPQDFRPIS
					GTAATGACTCAGCCTAGCCCAAGCTCTTGC		AATG	VPSVLVRQLNAI
					AATGGAGAAGTGATTCGTACGGATCACTCG		AATC	LATRLTSSIDWD
					CTTGAGACGGTATGGTCTGCAATAACGGAA		GTGC	PRQRGFSPTDGC
					CAAGACCTTAGGGCATCAAGAGTTTCATTA		ATGC	ADNATIVDLVL
					TCTTCATCTCCGGGGCCTGACGGGGTAACTC		TAGG	RHSHKYFKSCYI
					CAAAAACTGCCAGGGAGGTGCCGTCAGGTA		AACT	ANLDVSKAFDS
					TTATGCTACGCATAATGAACCTAATTCTATG		GACC	LSHAAIYGTLRA
					GTGCGGTAACTTACCTCATTCTATCCGACTG		AAAT	YGAPKGFVDYV
					GCCAGAACCATCTTCATCCCGAAGACGGTG		AACA	QKTYEGGGISLN
					ACGGCAAAGCGACCGCAAGACTTTCGTCCA		GCAG	GEGWCSEEFVP

					ATATCGGTGCCTTCCGTCCTGGTAAGACAG		CCCT	ARGVKQGDPLS
					CTAAATGCCATCTTGGCAACCCGATTGACCT		AGTA	PILFNLVIDRLLR
					CATCAATCGATTGGGACCCGCGCCAGCGGG		TCTTT	ALPSEIGTKVGN
					GCTTCTCACCAACCGACGGTTGCGCCGATA		CGAA	AMINAAAFADD
					ATGCGACGATAGTTGACTTAGTCCTGAGGC		GATT	LVLFAETRMGL
					ATAGCCATAAGTACTTTAAATCTTGCTACAT		TCCA	QTLLDKTVDFLS
					CGCCAACTTAGATGTTAGCAAGGCATTTGA		TACC	TVGLKLNADKC
					CTCATTGTCACATGCAGCAATATATGGGAC		TTTGC	FTVGIKGQPKQK
					ATTACGAGCTTATGGTGCGCCGAAGGGTTT		GATC	CTVLEAQSFCV
					TGTTGACTATGTACAGAAGACGTACGAGGG		AAAA	GSREIPTLKRTD
					AGGTGGTATCAGTCTCAACGGGGAAGGTTG		AAAA	EWKYLGIHFTAS
					GTGTTCAGAGGAATTCGTGCCTGCTAGAGG		AAAA	GRVRCNPAEDIG
					AGTGAAGCAGGGCGACCCTTTGTCCCCCAT		AAAA	PKLQRLSEAPLK
					TCTATTTAACTTGGTCATCGACCGGTTACTT		AAA	PQQRLFALRTVL
					AGAGCCCTACCTAGCGAGATTGGTACCAAG		(SEQ	IPQLYHKLSLGS
					GTCGGAAATGCCATGATAAACGCTGCTGCA		ID NO:	VTIGVLRKTDKL
					TTTGCAGATGATTTGGTACTATTTGCGGAAA		1367)	IRFYVRRWLNLP
					CTCGGATGGGACTTCAAACTTTGTTGGACA			SDVPIAFVHAPP
					AGACTGTGGACTTTTTATCCACCGTCGGCCT			KCGGLGIPSLRW
					TAAACTTAATGCTGATAAGTGCTTTACTGTC			VAPMLRLRRLS
					GGTATTAAGGGACAGCCGAAACAGAAGTGT			NIKWPHLVQSEE
					ACTGTGCTAGAGGCACAGAGCTTCTGCGTA			ASSFIEAEKQRA
					GGCTCGAGAGAGATTCCAACACTGAAGCGT			RGRLIAEQNELL

					ACTGACGAGTGGAAGTATCTCGGTATACAT			SRPAIEKYWAN
					TTCACTGCAAGTGGGAGGGTTCGATGCAAT			RLYLSVDGGGL
					CCGGCAGAGGACATTGGTCCAAAGCTACAA			REAGHYGPQHG
					AGATTGTCAGAGGCCCCCCTTAAGCCACAA			WVSQPTRLLTG
					CAGAGGTTGTTCGCCCTTCGGACTGTCCTGA			KEYLDGIRLRIN
					TCCCACAACTCTATCACAAGTTATCCCTTGG			ALPTKSRTTRGR
					GAGTGTGACGATAGGCGTCTTACGAAAGAC			HELERQCRAGC
					TGACAAGCTAATACGTTTCTATGTGCGAAG			DAPETTNHIMQ
					ATGGCTAAATCTTCCGTCGGATGTGCCGAT			KCYRSHGRRVA
					AGCATTCGTTCATGCCCCCCCAAAATGTGG			RHNCVVNRIKR
					GGGTCTCGGAATTCCATCACTAAGATGGGT			GLEERGCVVIAE
					AGCACCAATGTTACGACTTAGACGATTGAG			PSLQCESGLNKP
					CAACATAAAATGGCCCCACCTCGTACAATC			DLVVLRQNHID
					CGAGGAAGCCAGCTCCTTCATCGAAGCGGA			VIDVQVVTDGH
					AAAACAAAGGGCCCGAGGTAGATTGATAGC			SMDEAHQRKIN
					TGAACAAAATGAATTGTTATCGCGTCCGGC			RYDRPDIRTELR
					AATAGAAAAGTATTGGGCGAACAGGTTGTA			RRFEAAGDIEFH
					CCTCTCCGTTGATGGTGGCGGACTCCGTGA			SATLNWRGIWS
					AGCGGGCCACTATGGTCCCCAACACGGGTG			GQSVKRLIAKGL
					GGTTAGTCAGCCCACGCGTTTACTAACAGG			LSKYDSHIISVQ
					AAAGGAATATTTAGACGGTATTCGGCTGCG			VMRGSLGCFRQ
					GATAAATGCCCTACCCACAAAGTCTCGCAC			FMYLSGFSRDW
					TACGAGGGGAAGGCACGAATTGGAGAGGC			

					AGTGTCGTGCAGGATGTGATGCTCCCGAAA CAACAAACCACATTATGCAAAAATGCTATC GTTCGCATGGGAGGCGAGTAGCTAGACACA ACTGCGTAGTAAATCGAATCAAGCGGGGAC TTGAGGAGAGAGGCTGCGTGGTCATTGCCG AACCAAGTCTGCAGTGCGAATCCGGCCTTA ATAAACCGGACCTGGTGGTACTCCGACAAA ATCACATTGATGTGATTGACGTGCAGGTTGT GACAGACGGACATTCTATGGACGAAGCGCA TCAGCGCAAAATCAATAGATACGACAGACC GGACATACGAACTGAATTGCGTCGCCGATT CGAAGCCGCAGGTGACATCGAGTTTCATTC TGCCACCCTAAACTGGAGGGGGATTTGGAG TGGTCAATCCGTAAAACGATTGATTGCGAA GGGTCTCCTCAGCAAATATGATAGTCATAT CATTAGCGTCCAGGTTATGAGAGGCAGTCT TGGCTGTTTTAGACAGTTCATGTACCTTAGC GGGTTTTCTCGTGATTGGACGTAGCTTAAAA CGTTTGGTTCACATACATCTGCCTGCTGCCT TGGCACAATATCAAAAAGGCATAAACATCG CACATATTGGTTATTTACGGCTATGAGGATG GTTTTAGTACGTAGGCGTTGCGGAACTTCG GTTCGGATAGAGCAATGAATCGTGCATGCT			T (SEQ ID NO: 1489)

					AGGAACTGACCAAATAACAGCAGCCCTAGT ATCTTTCGAAGATTTCCATACCTTTGCGATC AAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1121)			
R2	R2_KF	GU9495 58		Kalotermes flavicollis	GAGAGGTCACTGTTGCTGATCAGCTGGACA CCTAGCTGACTGCGTGCATGCGTGCACGCC GCGCGCCGCCTCCTCGTGGTACCGCTGGTA ACGCAGGGTCACTGAGACTACCTGCGGCTA AAGGCGCGCGCGAGGGATTGTAAGTCCACC ACCTCCACGTGGTTCCCCGCGGGCAACGGC ATAGTTCATCTGCCGGCCAAGCGTGCCGTTC CCTCGATCCTCCTCTGATTAGGGATAGGAG GGGGGCGGTGGCTCCCGCCGACGACCTGGC AAACACCTTGGACACGCTAAGATAATAGGC CGTCCTCCGGGCCGGCCGAATCATAAGCAC ACAATGGAAGTCCCTCTAACATCAAGCCTC GGATCTGGGGCAACCCAAGCTCCCGGGACC CCAGAATTATTAGGGGAGCACACCGTTGAA CGACCTGGCCTTGATCAAGGACATTCTTATG GTTTGTTAATGGATGACGTAGAGCTGCCGG TGAGGTTACCATTTTTTGGGCCTCTCATATG CCCAGGATGCAGGACGCTGCTCACCTCTGA GGAAACGATTAGCAGCCATCATAGGAGGGT	GAGAGG TCACTG TTGCTG ATCAGC TGGACA CCTAGC CGTGCA GCGTG CACGCC GCGCGC TCACTG TTGCTG GCCTCC GGCCGA ATCATA GCCTCC CTCGTG ATCCGG CCGAAT GCCTCC

					CCACCCTGATGCCCGGACGAGGTGGGTCTG	CGGCTA	GTTA	DASQQGLGPND
					CTATGGTTGTGATTCCCCATTTATGACATAC	AAGGCG	TCCG	GNRGIVPSGQSS
					CGGGCCATTAAATGTCACTTACCTAAATGCT	CGCGCG	GGGC	PLFLEGSDAEGD
					CTGGCAGGAAGGTAGTGACAGGGGATCACA	AGGGAT	AAGT	EDVFNVLVPPTL
					TCTGTAATGGCTGCACAAAGCGGTTTGAGT	TGTAAG	CACT	GGLEPAGQVHS
					CTCAGAGAGGTCTCTCACTGCATAAAAGGA	TCCACC	AGCA	LSEGETSPLVGE
					GAGCCCATCCGGGTCTTCGGAACGAGGAAA	ACCTCC	ATAT	ADPCFMGGTPS
					TGCTTGAGCCTCCTGTGAGAGCTGAGCGAC	ACGTGG	GGGA	AGEASGSTLLGP
					GCCCTAACGCACATAAATCCTCAATATGGA	TTCCCC	CTTCT	DPTPADGYSLV
					GCATTGATGAGATTCGTATCCTGGAGCAAT	GCGGGC	CCGG	RKDLQLSVQTSP
					ATGAGGCTGCCTACGTCGGCGATCTCCACA	AACGGC	GTCC	LLAVGSVGTES
					TCAACATGAAGATAGCTGCTCACCTGCCGT	ATAGTT	GTGC	VQFERGVLSCG
					TCAAAACCAATAAACAAGTGAGCAACTATC	CATCTG	GGTC	TPPEFLHPEQFA
					GGAACGATCGCCGTAAAAAGTCCAGAACGG	CCGGCC	CTTCC	HCANNDPVLNA
					CGACTGACGCGAGCCAACAGGGACTTGGTC	AAGCGT	AACA	SEEQVHAPLGEE
					CCAATGACGGCAATAGAGGAATTGTCCCCA	GCCGTT	TGAG	ANDLPDNNHPS
					GCGGGCAGTCATCCCCGTTATTTTTAGAGG	CCCTCG	CTGG	ELGVDPEDPTCS
					GCTCTGACGCCGAGGGGGACGAGGATGTTT	ATCCTC	ACGT	PATEQVQPSSEE
					TTAATGTCCTAGTGCCGCCCACCCTAGGGG	CTCTGA	AGTC	EADDPFAQFKA
					GTCTTGAACCTGCTGGACAGGTTCATTCACT	TTAGGG	CACT	WRRRVASYALK
					CTCTGAGGGGGAAACTTCGCCGCTGGTCGG	ATAGGA	CTAT	IETGVLPAQVDD
					CGAAGCGGATCCCTGTTTTATGGGTGGCAC	GGGGGG	GACT	LLRRLRDGDTQ
					GCCGTCTGCCGGAGAGGCCTCGGGCTCAAC	CGGTGG	TGAA	SKVTCEEVEEV

					CTTGTTGGGTCCGGATCCCACCCCGGCAGA	CTCCCG	CGAT	VLSLTRTILGGT
					TGGATATTCTCTGGTCAGGAAAGACTTACA	CCGACG	ACGG	APKKRVEGRTK
					GCTCAGTGTGCAGACATCCCCGTTATTAGC	ACCTGG	GGGC	WTYKSRTNHEA
					AGTGGGCTCTGTCGGCACTGAGTCTGTGCA	CAAACA	GTAT	RKRIMYARCQD
					ATTTGAACGTGGGGTGTTATCGTGCGGTAC	CCTTGG	CTCC	LYRRRPQRPVER
					GCCACCAGAATTTTTACACCCGGAACAGTT	ACACGC	CCCG	AVGYQAEESLL
					TGCACATTGTGCAAACAATGACCCGGTTTT	TAAGAT	GAAG	DNQDERPSHGA
					GAACGCGTCAGAGGAGCAGGTGCACGCCCC	AATAGG	AGGT	FETFYTGLWGK
					ACTAGGGGAAGAAGCCAATGACCTCCCTGA	CCGTCC	CGCC	SGQCNITMPPGV
					CAACAACCATCCGTCGGAGCTGGGGGTTGA	TCCGGG	TAGG	PRHTGHVLREV
					CCCCGAAGACCCGACGTGTAGTCCCGCAAC	CCGGCC	CGAC	TPKDIYSRLRKL
					GGAACAGGTCCAACCCTCTTCTGAGGAAGA	GAATCA	TTAG	KKDYAPGPDGV
					AGCCGATGACCCGTTTGCGCAATTCAAGGC	TAAGCA	AA	TKLKVQSMGAY
					GTGGCGAAGAAGGGTGGCTTCCTATGCCTT	CACA	(SEQ	PSLLAKVYNLV
					GAAGATCGAGACCGGGGTTCTTCCAGCCCA	(SEQ ID	ID NO:	MLTGYFSSCWK
					AGTAGATGACCTCCTTCGGAGACTGAGGGA	NO: 1245)	1368)	EHKTSLIPKDRG
					TGGGGACACACAGTCTAAGGTAACGTGCGA			SPMDVSNWRPIT
					GGAGGTGGAGGAAGTGGTTTTATCCCTCAC			IGSLLSRIYTGLI
					TAGAACGATACTGGGTGGCACGGCACCCAA			ERRLRTVSDIHQ
					AAAGCGCGTTGAGGGTCGAACGAAGTGGAC			RQVGFMPVNGC
					GTACAAAAGTCGAACCAACCATGAGGCTCG			AANLFIFDECIR
					CAAGCGTATAATGTACGCTAGATGCCAGGA			QAKKEGTIVGSL
					CTTATATCGCAGGCGGCCTCAGCGACCTGT			IDVAKAFDTVPH

					GGAGAGGGCCGTCGGATACCAGGCGGAGG			EAILRALSSQGV
					AATCTCTTCTTGACAACCAAGATGAGAGAC			DEHTMAHIRDM
					CATCCCACGGAGCCTTTGAAACCTTCTATAC			YSGIRTRINGKG
					TGGGCTCTGGGGTAAGTCCGGTCAATGCAA			SDIPLVRGVKQG
					TATCACAATGCCTCCTGGTGTGCCTCGGCAC			DPLSPMLFNMV
					ACCGGGCATGTACTGAGGGAAGTTACGCCA			MDPLIRDLQRK
					AAGGACATCTACTCGAGGCTGAGGAAACTA			GFRIGGHEIGAL
					AAGAAAGACTATGCACCCGGCCCGGATGGC			AFADDIVLLADS
					GTAACTAAATTGAAGGTGCAGAGCATGGGA			IDGAQDHVDQV
					GCATATCCCTCACTTCTTGCAAAGGTATATA			GRYMNKLGMT
					ATCTGGTGATGCTAACCGGTTACTTCTCCAG			LNPRKSSSFLITA
					TTGTTGGAAGGAACATAAGACCTCTCTAAT			MRKTWICRDPG
					TCCTAAGGATAGAGGAAGCCCAATGGATGT			LSIGETKVPGAR
					CAGTAATTGGAGACCGATTACGATTGGCTC			PSSALKYLGVN
					CTTACTCAGTCGGATTTATACGGGTCTCATA			YTLSEGLESGAL
					GAACGGAGGCTGAGGACAGTTTCTGACATC			IDKLMQAVNRA
					CACCAAAGGCAGGTAGGGTTTATGCCCGTC			RGLALKPLQKV
					AATGGTTGTGCCGCCAACTTGTTTATATTTG			NLILERIIPKFLY
					ATGAATGCATCCGACAAGCCAAAAAAGAG			GIILGGPSLTRLH
					GGGACAATTGTCGGCAGTCTTATAGACGTC			AADKCVRMAV
					GCCAAAGCATTTGATACCGTACCCCATGAA			KEILHLHPSTTD
					GCCATCCTACGGGCGTTGAGCTCGCAAGGT			HVLYARKKDGG
					GTAGATGAGCACACGATGGCTCACATAAGA			MGIPRLAHLVRL

					GACATGTATTCCGGTATCCGTACGCGGATC			ASLRSGLALLAS
					AACGGCAAAGGGAGCGATATCCCTCTGGTG			GDVAVQAAGM
					AGAGGAGTGAAGCAAGGTGATCCTCTATCC			AGDLEGRCKKV
					CCAATGCTCTTTAATATGGTGATGGACCCTC			ANDLRLNWPVT
					TCATACGAGACCTACAGAGGAAGGGCTTCA			LRDVVRASNKF
					GAATTGGCGGCCACGAGATAGGGGCGTTGG			KSQESKDWERL
					CCTTTGCGGATGACATCGTCTTGTTGGCGGA			ASQGHGVKDFR
					CTCCATTGATGGGGCCCAGGATCATGTAGA			NDRLGNCWLYD
					TCAGGTTGGGCGCTACATGAATAAACTGGG			PTVLSSSRYTDA
					CATGACCCTCAATCCACGAAAATCGTCCTC			LRLRTNTFGVN
					CTTCCTCATCACTGCGATGAGGAAGACCTG			VALRRADKDLE
					GATATGTAGGGACCCGGGCTTGTCAATAGG			VNCRRCHGKPE
					GGAGACTAAGGTGCCTGGCGCCAGACCCTC			TLGHVLGECVA
					CTCTGCACTTAAATATCTCGGAGTAAACTAC			GKGMRIQRHDK
					ACACTCTCGGAAGGCCTGGAGAGTGGTGCC			MAAFVATKCEE
					TTGATAGACAAGCTCATGCAGGCCGTTAAC			KGYQTTREQLFS
					AGAGCACGAGGTTTGGCCCTTAAACCGCTG			IEQGKLKPDLVV
					CAAAAGGTCAACCTCATACTTGAACGAATA			IDGERALIVDVT
					ATCCCCAAATTTTTGTATGGTATCATCCTTG			VRFESGNALSRG
					GCGGGCCTAGTTTGACTAGGCTGCATGCAG			ASEKIEKYQPLA
					CCGACAAGTGTGTCAGGATGGCTGTAAAGG			DYFVSQGAVRE
					AGATCCTGCACCTCCATCCTAGTACAACAG			ANVLPIVVGSRG
					ACCATGTACTATATGCCAGAAAAAAGGATG			AITQATLKSLAT

					GGGGAATGGGGATTCCCCGACTGGCTCACC TTGTCCGCCTAGCCAGTCTCAGGTCGGGGTT AGCTCTGTTAGCGAGTGGGGACGTTGCCGT TCAAGCAGCGGGGATGGCAGGCGACCTGGA AGGCCGCTGCAAGAAAGTCGCAAATGATTT GCGGCTTAATTGGCCGGTGACGCTGAGAGA TGTGGTGAGAGCCAGCAATAAATTTAAATC TCAAGAATCAAAAGATTGGGAGCGTCTGGC CTCCCAGGGTCACGGAGTCAAGGACTTCCG GAATGACCGTCTTGGGAATTGCTGGCTCTAT GATCCGACGGTGTTGTCCAGCAGCAGATAC ACTGATGCTCTCAGGCTGAGGACTAACACG TTCGGCGTCAACGTTGCCCTAAGGAGGGCC GACAAGGACCTTGAGGTCAACTGTCGGCGA TGCCACGGGAAACCTGAAACCTTAGGGCAC GTGCTCGGAGAGTGTGTGGCAGGTAAGGGT ATGAGGATACAACGACACGACAAGATGGC GGCCTTTGTGGCTACAAAATGTGAAGAAAA GGGGTACCAGACGACCCGCGAACAGCTATT CTCCATCGAGCAGGGAAAACTCAAACCTGA CCTCGTCGTAATAGACGGGGAAAGGGCTCT CATTGTCGACGTAACAGTCCGCTTTGAGAG TGGGAACGCCCTCTCACGGGGAGCCTCTGA			LGLDVERVGKY LAICAVASSVEI ACMHLDYT (SEQ ID NO: 1490)

					GAAGATCGAAAAGTATCAGCCTTTGGCTGA CTATTTTGTCTCTCAGGGGGCTGTTCGAGAA GCCAACGTGTTACCGATTGTTGTCGGTTCCA GGGGAGCAATCACCCAGGCCACATTGAAGT CCCTAGCTACGCTAGGGTTAGATGTGGAAC GTGTTGGCAAGTACCTCGCCATCTGTGCCGT TGCCTCCAGCGTTGAGATCGCATGCATGCA CCTGGATTACACATAATGTCCCTTTTGGCTT GCCCCCACCTGCTTAAAGGAACTGGCAGGA AAGAGAGTGATCCGTGCCATAGAAATATGG TTATCCGGGGCAAGTCACTAGCAATATGGG ACTTCTCCGGGTCCGTGCGGTCCTTCCAACA TGAGCTGGACGTAGTCCACTCTATGACTTG AACGATACGGGGGCGTATCTCCCCCGGAAG AGGTCGCCTAGGCGACTTAGAA (SEQ ID NO: 1122)			
R2	R2_RL	GU949555		Reticulitermes lucifugus	CTCACTGCTGTCATGCCATTGTTAATGCGTT GGGTGATGGCGGGTGATGGGAGACGAGTTA CAGCAGAGCTGGCTTCACGGGGGGGGCTGT AGTTGCCCGAACCAGTCTGCTTCTGAGTGCC TCCACGTGGCCTCGCTGGAAACGTCGGAGC TGCTTACGGCTAACCCGTAGACAACTCCGG CGGCCAAACTCAGAAAACGGCCTTACTACC	CTCACT GCTGTC ATGCCA TTGTTA ATGCGT TGGGTG ATGGCG	TGAG TTAG ATAT ATAT GACG ACTG ATTCT	MMADYNNSVD HALEDNTRLIFA RDAVLARVCGP FDNLECGLCGV LLTSLQGVREHC HRAHHNLDLTF QCTKCDKGFSS

					AGGGTCATCCCTCGCGGCGGCGCTAAGACC	GGTGAT	TCCCT	YRGICCHFSKCI
					TTACTGTATGTACTACGGACTACAGTTGAGT	GGGAGA	AACT	GARISVSEGPLS
					AGCGCGGGAGAGCACTTGATGTGAGGGTAG	CGAGTT	AAAA	CSECEREFDSKR
					CAACTTGTTGCTACTTCGACTGTCTCCCTGA	ACAGCA	TATA	ALSTHERHMHP
					AGATTTCCGAAGGGGTGCCGCAAGTCCATG	GAGCTG	ATTT	GIRNAKRLKDF
					GGGGCGGTCCTGCTGTGGGAGTTATCTTGC	GCTTCA	GGAA	NPRGGGGKTIH
					ACCCGGGAAGCCCTCAGTGGATAAATACTC	CGGGGG	TGTG	GNTKWTEEEVQ
					CAAAATGGCTTAGCCACCTCACGTGGAGAG	GGGCTG	CATT	LLVSLSKRFEGY
					AGGGGAATTAGGGGTTTACTCCTAAGGGCC	TAGTTG	ACTT	KSINKEISLILTS
					GATGCCTCCAATCGGATTTCCCCGGGCAAC	CCCGAA	ATAC	KTCKQISDKRRY
					GGCTATTATCAGCCGGCTAAGTTCGGACCC	CCAGTC	TTTAT	LNLHNGNGGLA
					TCATCTACTGATGGGATACTCTCTCTCCCCC	TGCTTC	TGTA	AAEAVLEFCEDS
					CGCCTGCGTGATTAGTTGGGAATGCACGCG	TGAGTG	TTTAT	HPEVTESDGAV
					GGCGGGTGATTCTGACAAGCCCAGCCCAAA	CCTCCA	TGAC	LSEIMDEECHQS
					AGTTCTTACCACGACGTACCCGGGGGTGTC	CGTGGC	ATCC	SVTMRSSIVHGD
					ACCCGAACTAATATATTTTCCGCCGATCGGT	CTCGCT	GTAG	IGREVQGKELVR
					TCGCTGTGGGGGGATGTCGGGGCCTGGTGG	GGAAAC	TAGT	IPPDNSVMGNC
					CCGGGAACCGTCAGCCGCTCTTTGAGTGTC	GTCGGA	CTGTT	VVLLRKLATDK
					CTGCGTAACCCGGAGGCGGTGACGTCGAAG	GCTGCT	TAGT	RSDLDLSKDKEL
					GGCTAGGCTAGCACGGTCCAGACCCCTGAA	TACGGC	GGAT	RIDIEKATKANR
					CGCCTGGGGAGACGGGCCCACCACCTGAGT	TAACCC	TTAG	ESADGRVIQSVA
					AGGAGTGGGTCCTTTACCTCATTTGAGGTGT	GTAGAC	ATCG	GNEIDPDTFQW
					CTCCTCGTTCTTGTAGGGGGCGAAGGACGA	AACTCC	TTAA	KELLLGQVRGFP

					GGAATGCATCCGTCCCTCCAGGATGCTGGT	GGCGGC	CCAT	RVDENSELFDLD
					GGTTTCCGTCTGGTGGGCTCATGCTGAAAC	CAAACT	TCTGT	DKLTKELSSDSP
					GCGCAGGTGCCGATTTGTCGAAGAGGACAT	CAGAAA	GACG	VWNDNCELIVS
					ATGGGGTAACCCATAGAACCTAGGCAGGAG	ACGGCC	TCGC	DLCQVLCKKKY
					TAATCCCTTAGCTGGGGGGGTCAGCTGGTG	TTACTA	TGAT	ELGRQHHVRKG
					GGTCCTGTCAATTTATCCTCCCCTCCATGCC	CCAGGG	GTCT	KRHRGIHHKRE
					AGAGCCGGTCCGAGGTTAGAGACGGACTCA	TCATCC	ATAT	KFRECQKIFRKS
					TTTTCTCCTTTTTATATGTCATGATGGCCGA	CTCGCG	GTAC	PRKLAEYLYRD
					TTATAATAATTCTGTTGATCATGCACTTGAG	GCGGCG	CAGT	KDLSHISKDAST
					GATAATACTAGATTGATATTTGCGCGTGAC	CTAAGA	GGCA	PQGIEQYYSQL
					GCAGTACTTGCGAGAGTATGTGGACCTTTT	CCTTAC	GAAT	WGEPELLESNTI
					GATAACTTGGAGTGCGGGTTGTGTGGGGTG	TGTATG	GCTG	EEKLPSSSLFDC
					CTCCTGACGTCCCTTCAGGGCGTTCGAGAG	TACTAC	ACTA	LPPITPEEVEGRI
					CATTGCCACCGGGCGCACCATAACTTGGAT	GGACTA	AGAA	HKIRPSSAPGLD
					TTGACTTTTCAATGTACTAAATGCGATAAAG	CAGTTG	CAAA	GVRKIHLVGKGI
					GGTTTTCCTCGTATAGGGGAATTTGTTGTCA	AGTAGC	ATAA	TLVLVKLYNLLF
					TTTTTCAAAGTGTATAGGAGCTCGTATATCT	GCGGGA	TTTTA	LTGGYPECWKR
					GTTTCAGAGGGCCCCCTTAGTTGTTCAGAAT	GAGCAC	AACA	NRTVFIPKIGKD
					GTGAAAGGGAATTTGACTCTAAAAGAGCTC	TTGATG	ATAG	LSEVGGWRPLTI
					TTTCAACCCACGAGAGGCATATGCACCCAG	TGAGGG	ACAG	GSLLARMYSAF
					GGATTAGGAATGCTAAGAGGCTCAAAGATT	TAGCAA	CACT	LERRIRRVTSLS
					TTAATCCGCGGGGGGGGGGTGGTAAAACTA	CTTGTT	AAGA	LSQRGFTNIQGC
					TTCACGGCAATACTAAGTGGACTGAAGAGG	GCTACT	AATT	HVNLTILKEGIR

					AGGTCCAGTTGCTTGTATCACTTAGCAAGA	TCGACT	CTGC	QAKVKNGGVIV
					GATTTGAGGGGTATAAGTCTATCAATAAAG	GTCTCC	AAAC	SVDIEKAFDTIP
					AAATTAGCCTTATTTTAACATCAAAAACTTG	CTGAAG	TGGA	HSVIFSRLASQG
					TAAACAAATTTCTGATAAACGTCGCTATTTA	ATTTCC	ACGT	VPPLLRKIISNM
					AATTTGCATAATGGAAATGGGGGGCTAGCT	GAAGGG	GGCC	YKDVYTVIEGQ
					GCAGCGGAAGCTGTATTAGAATTTTGTGAA	GTGCCG	CACG	CIPIKRGVKQGD
					GATAGTCACCCTGAAGTTACTGAGAGCGAT	CAAGTC	GTAG	PLSPLLFNIAIDP
					GGGGCGGTGCTTTCTGAAATCATGGATGAG	CATGGG	GCCA	VLRSLEEFQGGL
					GAATGTCATCAGTCATCGGTAACTATGAGG	GGCGGT	ATAT	PLGNSAIKILAF
					TCAAGCATCGTACATGGCGACATTGGTCGT	CCTGCT	ACCG	ADDIILGASSAG
					GAGGTTCAAGGGAAGGAATTGGTCCGGATC	GTGGGA	GAAA	QAQQMVDMLGI
					CCGCCTGATAATAGTGTAATGGGGAATTGT	GTTATC	GGGA	GLTSCGLGVSH
					GTTGTTCTACTTAGGAAACTTGCTACTGACA	TTGCAC	AAAT	RKCFGFQIVNKN
					AAAGAAGTGATTTGGACTTGAGTAAGGATA	CCGGGA	GACA	KTWTIVDPMITL
					AAGAGTTGCGCATTGACATTGAGAAGGCGA	AGCCCT	CATC	NGSSLPFSGPED
					CGAAAGCTAACAGAGAAAGCGCAGACGGT	CAGTGG	CCCC	RLPYLGVDINP
					AGGGTAATTCAATCTGTGGCTGGTAATGAG	ATAAAT	CTTA	WDRKSRYDAG
					ATCGATCCCGATACTTTTCAATGGAAAGAA	ACTCCA	AA	QRLISAAKRGSQ
					TTGCTGCTTGGGCAGGTACGGGGGTTTCCA	AAATGG	(SEQ	LSLKPQQKINLIT
					CGAGTTGATGAGAACTCTGAATTATTTGATT	CTTAGC	ID NO:	AFLLPKFLYILIE
					TGGATGACAAATTGACTAAGGAATTATCTT	CACCTC	1369)	DPPSPAYLKSID
					CCGACAGCCCCGTGTGGAATGACAACTGCG	ACGTGG		HDLRQIYKNILH
					AACTTATTGTGTCGGACTTATGTCAGGTCTT	AGAGAG		LPNCVSTAFMY

					ATGCAAAAAGAAATATGAGTTAGGTAGGCA	GGGAAT		SPKRDGGLGLPR
					GCATCACGTTAGGAAAGGAAAGAGGCATCG	TAGGGG		LSCLVPLAHLKA
					GGGTATACATCATAAGAGGGAGAAATTTAG	TTTACT		GIKLGSLQDSLV
					GGAATGCCAGAAGATATTTAGAAAGTCCCC	CCTAAG		REITTSDRFVRT
					CAGAAAGCTTGCCGAATATCTCTACCGTGA	GGCCGA		MGSVAHSLWAS
					TAAGGACCTGTCTCATATCAGTAAAGACGC	TGCCTC		WSLTLQDIYKL
					TAGCACTCCGCAGGGTATAGAGCAATACTA	CAATCG		KSALKRREAKA
					TTCTCAGTTATGGGGGGAACCTGAATTACT	GATTTC		WESCVSQGQGA
					AGAATCCAACACAATTGAAGAAAAGTTACC	CCCGGG		AQFRGDSIGNN
					GTCCAGTTCCTTATTTGACTGTCTTCCCCCC	CAACGG		WLHNPGTYRPG
					ATCACGCCGGAGGAGGTTGAGGGAAGAATC	CTATTA		QYIEALKLRANL
					CATAAGATTCGACCGTCTTCCGCCCCAGGG	TCAGCC		TGVRVNLKRSG
					TTGGATGGGGTACGCAAGATCCATCTTGTC	GGCTAA		YNVPITCRFCKD
					GGTAAAGGGATAACCCTGGTCCTTGTTAAA	GTTCGG		IPETQAHVLGLC
					CTCTACAACTTGCTATTTTTGACTGGGGGAT	ACCCTC		PKTKGMRIQRH
					ATCCTGAGTGTTGGAAGCGGAACAGGACCG	ATCTAC		DSIVNRVRDKL
					TCTTTATACCGAAGATTGGGAAAGATCTGA	TGATGG		KTKSPVALMHE
					GTGAGGTGGGAGGATGGAGACCGTTGACAA	GATACT		QNFTVEEGQVF
					TTGGTTCCCTATTAGCTCGGATGTATTCAGC	CTCTCT		KPDIVTILGEVG
					CTTCCTAGAGAGGAGGATTAGGAGGGTGAC	CCCCCC		YVIDVTVRYDD
					CTCCTTGTCCTTGTCGCAGAGGGGCTTTACG	GCCTGC		RDYIKDASVEKI
					AATATCCAAGGTTGTCACGTTAATTTGACTA	GTGATT		RKYEALKGYLK
					TCCTCAAAGAAGGCATTCGACAGGCCAAAG	AGTTGG		DLYPQLNKVEV

					TCAAGAACGGGGGAGTTATAGTCTCAGTAG ACATAGAGAAGGCTTTTGACACAATCCCTC ATTCTGTAATTTTCTCCCGCTTGGCGTCTCA AGGTGTTCCTCCCCTTCTTAGGAAAATTATT TCCAATATGTATAAGGATGTCTACACTGTG ATTGAGGGGCAGTGTATTCCAATAAAGAGA GGAGTTAAACAAGGGGATCCTCTTTCTCCA CTACTTTTCAATATTGCCATCGATCCTGTGT TGCGCAGTTTGGAGGAGTTTCAGGGGGGTC TCCCACTTGGTAACTCGGCTATTAAGATCTT GGCCTTTGCTGATGATATTATATTAGGAGCT AGTTCGGCGGGTCAGGCGCAGCAGATGGTG GATATGTTGGGAATAGGACTTACTTCGTGT GGGTTGGGAGTATCCCATAGAAAATGTTTT GGATTTCAAATCGTGAATAAAAACAAGACC TGGACAATTGTTGACCCGATGATCACACTC AATGGTTCTTCATTACCCTTCTCAGGGCCAG AGGACAGGTTGCCTTATCTCGGGGTGGATA TAAACCCATGGGATAGGAAATCACGGTATG ACGCTGGACAGAGACTTATTTCGGCGGCCA AACGAGGCTCTCAACTATCCCTTAAGCCTC AGCAAAAGATAAATTTAATAACAGCTTTTC TTTTGCCAAAGTTTCTGTACATTCTGATTGA	GAATGC ACGCGG GCGGGT GATTCT GACAAG CCCAGC CCAAAA GTTCTT ACCACG ACGTAC CCGGGG GTGTCA CCCGAA CTAATA TATTTT CCGCCG ATCGGT TCGCTG TGGGGG GATGTC GGGGCC TGGTGG CCGGGA		LPLVFGSRGAVP GSTVHNMGLLG FTKREMVHISRK VIADSLIISNFLE VY (SEQ ID NO: 1491)

					AGATCCGCCTTCGCCTGCCTATCTGAAATCC	ACCGTC		
					ATCGATCATGACCTGAGGCAGATATATAAG	AGCCGC		
					AATATTTTGCACCTCCCTAACTGTGTGTCAA	TCTTTG		
					CCGCCTTTATGTACTCTCCCAAGAGGGATG	AGTGTC		
					GTGGTTTGGGGCTACCGCGTCTGTCCTGTCT	CTGCGT		
					AGTTCCACTTGCCCATTTAAAAGCCGGGAT	AACCCG		
					CAAGCTCGGTTCCCTCCAAGATTCTTTGGTT	GAGGCG		
					AGGGAAATTACCACCTCAGACAGGTTTGTA	GTGACG		
					CGTACAATGGGTTCGGTAGCTCACTCCCTTT	TCGAAG		
					GGGCATCATGGTCTCTAACGTTGCAGGATA	GGCTAG		
					TCTATAAACTTAAGTCCGCACTGAAACGTA	GCTAGC		
					GGGAAGCTAAGGCATGGGAGAGTTGTGTGT	ACGGTC		
					CTCAGGGACAAGGAGCAGCGCAATTTCGCG	CAGACC		
					GGGACTCAATTGGAAACAACTGGCTCCATA	CCTGAA		
					ATCCGGGAACCTATCGACCTGGACAGTATA	CGCCTG		
					TCGAAGCACTTAAGCTTCGTGCTAATTTAAC	GGGAGA		
					AGGGGTTAGAGTCAACTTGAAGCGTTCAGG	CGGGCC		
					CTATAATGTTCCAATAACTTGTAGGTTCTGT	CACCAC		
					AAGGATATACCTGAAACGCAGGCTCATGTT	CTGAGT		
					CTTGGACTTTGTCCAAAGACCAAGGGAATG	AGGAGT		
					AGAATACAAAGGCACGATTCAATTGTGAAT	GGGTCC		
					AGGGTGCGAGATAAACTTAAGACTAAGTCG	TTTACC		
					CCCGTTGCTCTCATGCATGAGCAAAACTTCA	TCATTT		

					CCGTTGAGGAAGGCCAGGTGTTCAAACCAG ATATAGTGACGATTTTGGGAGAGGTGGGTT ACGTAATTGATGTGACAGTACGGTATGACG ATAGAGATTACATCAAGGATGCCTCTGTGG AAAAAATCAGGAAATACGAGGCCCTTAAAG GGTACCTGAAGGATTTATATCCACAGCTGA ATAAGGTAGAGGTTTTGCCTCTTGTCTTTGG TAGCAGAGGGGCCGTGCCAGGTAGTACTGT TCACAACATGGGATTGTTAGGTTTTACGAA ACGTGAAATGGTCCACATCTCTCGCAAAGT TATAGCCGACTCACTGATAATCTCTAATTTT TTAGAGGTTTATTGAGTTAGATATATATGAC GACTGATTCTTCCCTAACTAAAATATAATTT GGAATGTGCATTACTTATACTTTATTGTATT TATTGACATCCGTAGTAGTCTGTTTAGTGGA TTTAGATCGTTAACCATTCTGTGACGTCGCT GATGTCTATATGTACCAGTGGCAGAATGCT GACTAAGAACAAAATAATTTTAAACAATAG ACAGCACTAAGAAATTCTGCAAACTGGAAC GTGGCCCACGGTAGGCCAATATACCGGAAA GGGAAAATGACACATCCCCCCTTAAA (SEQ ID NO: 1123)	GAGGTG TCTCCT CGTTCT TGTAGG GGGCGA AGGACG AGGAAT GCATCC GTCCCT CCAGGA TGCTGG TGGTTT CCGTCT GGTGGG CTCATG CTGAAA CGCGCA GGTGCC GATTTG TCGAAG AGGACA TATGGG GTAACC		

						CATAGA ACCTAG GCAGGA GTAATC CCTTAG CTGGGG GGGTCA GCTGGT GGGTCC TGTCAA TTTATC CTCCCC TCCATG CCAGAG CCGGTC CGAGGT TAGAGA CGGACT CATTTT CTCCTT TTTATA TGTC		

						(SEQ ID NO: 1246)		
R2	R2_RU	GU949554		Reticulitermes urbis	CACACTGCTGTCATGCCATTGTTAATTCGTT	CACACT	TGAG	MMADYNNSVD
					GGGTGATGGCGGGTGATGGGAGACGAGTTA	GCTGTC	TTAG	HALEDDTRFIFA
					CAGCAGAGCTGGCTTCACGGGGGGGCTGTA	ATGCCA	ATAT	RDSVLARVCGH
					GTGGCCCGAACCAGTCTGCTTCTGAGTGCC	TTGTTA	ATAT	FDNLKCELCGV
					CCCAAGTGGCCTCGCTGGAAACGTCGGAGC	ATTCGT	AACG	LLTSLQGVREHC
					TGCTTACGGTTAACCCGTAGACAACTCCGG	TGGGTG	ACTG	HRSHHNLDLTF
					CGGCCAAACTCAGAAAACGGCCTTACTACC	ATGGCG	ATTCT	QCTKCDKGFSS
					AGGGCCATCCCTTGCGGTTGCGCTAAGACC	GGTGAT	TCCCT	YRGICCHFSKCR
					TTACTGTATGTACTACGGACTACAGTTGAGT	GGGAGA	AACT	GARISVSEGPLS
					AGCGCGGGAGAGCACTCGATGTGAGGGTAG	CGAGTT	AAAA	CSECERKFDSKR
					CAACTTGTTGCTTCTTCGACTGTCTCCCTGA	ACAGCA	TATA	ALSTHERHMHP
					AGAGTTCCGAAGAGGTGCCGCAAGTCCATG	GAGCTG	ATTT	GIRNAKRLKDF
					GGGGCGACCCGGCTGTGGGAGTTATCCTGC	GCTTCA	GGAA	NPRGGGKTIHG
					ACCCGGGAAGCCCTCAGTGGATAAATACTC	CGGGGG	TGTG	NTKWTEEEVQL
					CAAAATGGCTTAGCCACCTCACGTGGAGAG	GGCTGT	CATT	LVSLSKRFEGYK
					AGGGGAATTAGGGGTTTACTCCTAAGGGCC	AGTGGC	ACTT	SINKEISLILTSK
					GATGCCTCCAATCGGATTTCCCAGGGCAAC	CCGAAC	ATAC	TCKQISDKRRYL
					GGCTATTATCAGCCGGCTAAGTTCGGACCC	CAGTCT	TTTAT	NLHNGNGGLAA
					TCATCTACCGATGGGATACTCTCTCTCCCCC	GCTTCT	TGTA	AEAVLVFCDDS
					CGCCTGCGTGATTAATTGGGAATGCACGAG	GAGTGC	TTTAT	HLEVTDSDGAV
					GGCGGGTGATTCTGACAAGCCCAGCCCAAA	CCCCAA	TGAC	LSEIMDEEYYQS

					AGTTCTTACCACGACGTACCCGGGGGTGTC	GTGGCC	ATCC	SLTMRSSIVHGD
					ACCCGAACTAATATATTTTCCACCGATCGGT	TCGCTG	GTAG	IGREVQGKDLV
					TCGCTGTGGGGGGATGTCGGGGCCTGGTGG	GAAACG	TAGT	RIPPDNSVMGNC
					CCAGGAACCGTCAGCCGCTCTTTGAGTGTC	TCGGAG	CTGTT	VVLLRKLATEK
					CTGCGTAACCCGGAGGCGGTGACGTCGAAG	CTGCTT	TAGT	RSDLDLSKDKEL
					GGCTAGGCTAGCACGGTCCAGACCCCTGAA	ACGGTT	GGAT	RIDIEKATKANR
					CGCCTGGGGAGACGGGCCCACCACCTGTGT	AACCCG	TTAG	ESADGRVIQSVA
					AGGAGTGGGTCCTTTACCTCATTTGAGGTGT	TAGACA	ATCG	DNEIDPDTFQW
					CTCCTCGGTCTAGTAGGGGGCGAAGGACGA	ACTCCG	TTAA	KELLLGQVRGFP
					GGGATGCATCCGTCCCTCCAGGATGCTGGT	GCGGCC	CCAT	RVDENSELFDLD
					GGTTTCCGTCTGGTGGGCTCATGCTGAAACC	AAACTC	TCTGT	DKLTSELSSDSP
					CGCAGGTGCCGATTTGTCGAAGAGGACATA	AGAAAA	GACG	VWNDNCELIVS
					TGGGGTAACCCATAGAACCTAGGCAGGAGT	CGGCCT	TCGC	DLCHVLCKNKY
					AATCCCTTAGCTGGGGGGGTCAGCTGGTGG	TACTAC	TGAT	ELGRQHHVRKG
					GTCCTGTCAATTTATCCTCCCCTCCATGCCA	CAGGGC	GTCT	KRHRGIHHKRE
					GAGCCGGTCCGAGGTTAGAGACGGACTCAT	CATCCC	ACAT	KFRECQKIFRKS
					TTTCTCCTTTTTATATGTCATGATGGCCGAT	TTGCGG	GTAC	PRKLAEYLYRD
					TATAATAATTCTGTTGATCATGCACTTGAGG	TTGCGC	CAGT	KDLSHISKDVST
					ATGATACTAGATTTATATTTGCGCGTGATTC	TAAGAC	GGCA	PQGIEQYYSQL
					AGTACTTGCGAGAGTATGTGGACATTTCGA	CTTACT	GAAT	WGKPELLESNT
					TAACTTGAAGTGTGAGTTATGTGGAGTGCT	GTATGT	GCTG	TEEMLPSSSLFD
					CCTTACGTCCCTTCAGGGCGTTCGGGAGCAT	ACTACG	ACTA	CLPPITPEEVEGR
					TGCCACCGGTCGCACCATAACTTGGATTTG	GACTAC	AGAA	IHKIRPSSAPGLD

					ACTTTTCAATGTACTAAATGCGATAAAGGG	AGTTGA	CAAA	GVGKIHLVGKGI
					TTTTCCTCGTATAGGGGAATTTGTTGTCATT	GTAGCG	ATAA	TLVLAKLYNLLF
					TTTCGAAGTGTAGAGGAGCTCGTATATCTGT	CGGGAG	TTTTA	LTGGYPECWKR
					TTCAGAGGGCCCCCTTAGTTGTTCAGAATGT	AGCACT	AACA	NRTVFIPKIGKD
					GAGAGGAAATTTGACTCTAAAAGAGCTCTT	CGATGT	ATAG	LSEVGGWRPLTI
					TCAACCCACGAGAGGCATATGCACCCAGGG	GAGGGT	ACAG	GSLLARMYSAF
					ATTAGGAATGCTAAGAGGCTCAAAGATTTT	AGCAAC	CACT	LERRIRRVTSLSS
					AATCCGCGGGGGGGTGGTAAAACTATTCAC	TTGTTG	AAGA	SQRGFTNIQGCH
					GGCAATACTAAGTGGACCGAAGAGGAGGTC	CTTCTT	AATT	VNLTILKEGIRQ
					CAATTGCTTGTATCACTTAGCAAGAGATTTG	CGACTG	CTGC	AKVKNGGVIVS
					AAGGGTATAAGTCTATCAATAAAGAAATTA	TCTCCC	AAAC	VDIEKAFDTIPH
					GCCTTATTTTAACATCAAAAACTTGTAAACA	TGAAGA	TGGA	SVIFSRLASQGV
					AATTTCTGATAAACGTCGCTATTTAAATTTG	GTTCCG	ACGT	PPLLRKIISNMY
					CATAATGGAAATGGGGGGCTAGCGGCAGCG	AAGAGG	GGCC	KDVYTVIEGQCI
					GAAGCTGTATTAGTATTTTGTGATGATAGTC	TGCCGC	CACG	PIKRGVKQGDPL
					ACCTTGAAGTTACTGACAGCGATGGGGCGG	AAGTCC	GTAG	FPLLFNIAIDPVL
					TGCTTTCTGAAATCATGGATGAGGAATATT	ATGGGG	GCCA	RSLEEFQGGLPL
					ATCAGTCATCGCTAACTATGAGGTCAAGCA	GCGACC	ATAT	GNSAIKILAFDD
					TCGTACATGGCGACATTGGTCGTGAGGTTC	CGGCTG	ACCG	DIILGASSAGQA
					AAGGGAAGGATTTGGTCCGGATCCCGCCTG	TGGGAG	GAAA	QQMVDMLGIGL
					ATAATAGTGTAATGGGGAATTGTGTTGTTCT	TTATCC	GGGA	TSCGLGVSHRK
					ACTTAGAAAACTTGCTACTGAAAAAAGAAG	TGCACC	AAAT	CFGFQIVNKNKT
					TGATTTGGACTTGAGTAAGGATAAAGAGTT	CGGGAA	GACA	WAIVDPMITLN

					GCGCATTGACATTGAGAAGGCGACGAAAGC	GCCCTC	AATC	GSSLPFSGPEDR
					TAACAGAGAAAGCGCAGATGGTAGGGTAAT	AGTGGA	CCCC	LPYLGVDTNPW
					TCAATCTGTGGCTGATAATGAGATCGATCC	TAAATA	CTTA	DRKSRYDAGQR
					CGATACTTTTCAATGGAAAGAATTGCTTCTT	CTCCAA	AA	LISAAKRGSQLS
					GGGCAGGTACGGGGGTTTCCACGAGTTGAT	AATGGC	(SEQ	LKPQQKINLITTF
					GAGAACTCTGAATTATTTGATTTGGATGAC	TTAGCC	ID NO:	LLPKFLYILIEDP
					AAATTGACTAGTGAATTATCTTCCGACAGC	ACCTCA	1370)	PSPAYLKSIDHD
					CCCGTGTGGAATGACAACTGCGAACTTATT	CGTGGA		LRQIYKNILHLP
					GTGTCGGACTTATGTCACGTCTTATGCAAAA	GAGAGG		NCVSTAFMYSP
					ATAAATATGAGTTAGGTAGGCAGCATCACG	GGAATT		KRDGGLGLPRL
					TTAGGAAAGGAAAGAGGCATCGGGGTATAC	AGGGGT		SCLVPLAHIKAG
					ATCATAAGAGGGAGAAATTTAGGGAATGCC	TTACTC		IKLGSLQDSLVR
					AGAAGATATTTAGAAAGTCCCCCAGAAAGC	CTAAGG		EITTSDRFVRTM
					TTGCCGAATATCTCTACCGCGATAAGGACC	GCCGAT		GSVSHSIGASWP
					TGTCTCATATCAGTAAAGACGTTAGCACTCC	GCCTCC		LTLQDIYKLKSA
					GCAGGGTATAGAGCAATACTATTCTCAGTT	AATCGG		LKRREAKAWES
					ATGGGGGAAACCTGAATTACTAGAATCCAA	ATTTCC		CVSQGQGAAQF
					CACAACTGAAGAAATGTTACCTTCCAGTTC	CAGGGC		RGDSIGNNWLH
					CTTATTTGACTGTCTTCCTCCCATCACGCCG	AACGGC		NPGTFRPGQYIE
					GAGGAGGTTGAGGGGAGAATCCATAAGATT	TATTAT		ALKLRANSTGV
					CGACCGTCTTCCGCCCCAGGGTTGGATGGG	CAGCCG		RVNLKRSGYNV
					GTAGGCAAGATCCATCTTGTCGGTAAAGGG	GCTAAG		PITCRFCKDIPET
					ATAACCCTGGTCCTTGCTAAACTCTACAACT	TTCGGA		QAHVLGLCPKT

					TGCTATTTTTGACTGGGGGATATCCTGAGTG	CCCTCA		KGMRILRHDSIV
					TTGGAAGCGGAACAGGACCGTCTTCATACC	TCTACC		NRVRDKLKTKS
					GAAGATCGGGAAAGATCTGAGTGAGGTGG	GATGGG		PVALMHEQNFT
					GAGGATGGAGACCGCTGACAATTGGCTCCC	ATACTC		VEEGQVFKPDIV
					TATTAGCTCGGATGTATTCAGCCTTCCTAGA	TCTCTC		TILGEVGYVIDV
					GAGGAGGATTAGAAGGGTGACTTCCTTGTC	CCCCCG		TVRYEDRDYIK
					CTCGTCACAGAGGGGCTTTACGAATATTCA	CCTGCG		DASVEKIRKYEA
					AGGTTGTCACGTTAATTTGACTATCCTCAAA	TGATTA		LKGYLKDLYPQ
					GAAGGCATTCGACAGGCCAAAGTTAAGAAC	ATTGGG		LNKVEVLPLVF
					GGGGGAGTTATAGTTTCAGTAGACATAGAG	AATGCA		GSRGAVPGSTV
					AAGGCTTTTGACACAATCCCCCATTCTGTAA	CGAGGG		HNMGLLGFTKR
					TTTTTTCCCGCTTGGCGTCTCAAGGTGTTCC	CGGGTG		EMVHISRKVITD
					TCCCCTTCTTAGGAAAATTATTTCCAATATG	ATTCTG		SLIIISNFLEVY
					TATAAGGATGTTTACACCGTGATTGAGGGA	ACAAGC		(SEQ ID NO:
					CAGTGTATTCCAATAAAGAGAGGAGTTAAA	CCAGCC		1492)
					CAAGGGGATCCTCTTTTTCCACTACTTTTCA	CAAAAG		
					ATATTGCCATCGATCCTGTGTTGCGCAGTTT	TTCTTA		
					GGAGGAGTTTCAGGGGGGTCTCCCACTTGG	CCACGA		
					TAACTCGGCTATTAAGATCTTGGCCTTTGAT	CGTACC		
					GATGATATTATATTAGGAGCTAGTTCGGCG	CGGGGG		
					GGTCAGGCGCAGCAGATGGTGGATATGTTG	TGTCAC		
					GGAATAGGACTTACTTCGTGTGGGTTGGGA	CCGAAC		
					GTATCCCATAGAAAATGTTTTGGATTTCAAA	TAATAT		

					TCGTGAATAAAAACAAGACCTGGGCAATTG	ATTTTC		
					TTGACCCAATGATCACACTCAATGGTTCTTC	CACCGA		
					ATTACCCTTCTCAGGGCCAGAAGACAGGCT	TCGGTT		
					GCCTTATCTCGGGGTGGATACAAACCCATG	CGCTGT		
					GGATAGGAAATCACGGTATGACGCTGGACA	GGGGGG		
					GAGACTTATTTCGGCTGCCAAACGAGGCTC	ATGTCG		
					TCAACTATCCCTTAAGCCTCAGCAAAAGAT	GGGCCT		
					AAATTTAATAACAACTTTTCTTTTGCCGAAG	GGTGGC		
					TTTTTGTACATTCTGATTGAAGATCCGCCTT	CAGGAA		
					CGCCTGCCTATCTGAAATCCATCGATCATGA	CCGTCA		
					CCTGAGGCAGATATATAAGAATATTTTGCA	GCCGCT		
					CCTCCCTAACTGCGTGTCAACCGCCTTTATG	CTTTGA		
					TACTCTCCCAAGAGGGATGGTGGCTTGGGG	GTGTCC		
					CTACCGCGTCTGTCCTGTCTTGTTCCTCTTG	TGCGTA		
					CCCATATAAAAGCCGGGATCAAGCTTGGTT	ACCCGG		
					CCCTCCAAGATTCTTTGGTTAGGGAAATTAC	AGGCGG		
					CACCTCAGACAGGTTTGTACGTACGATGGG	TGACGT		
					TTCGGTATCTCACTCCATTGGGGCATCATGG	CGAAGG		
					CCTCTAACGTTGCAGGATATCTATAAACTTA	GCTAGG		
					AGTCCGCACTGAAACGTAGGGAAGCTAAGG	CTAGCA		
					CATGGGAGAGTTGTGTGTCTCAGGGACAAG	CGGTCC		
					GAGCAGCGCAATTTCGAGGGGACTCAATTG	AGACCC		
					GAAACAACTGGCTCCATAATCCGGGAACCT	CTGAAC		

					TTCGACCTGGACAGTATATCGAAGCACTTA	GCCTGG		
					AGCTTCGTGCTAATTCAACAGGGGTTAGAG	GGAGAC		
					TCAACTTGAAGCGTTCAGGCTATAATGTTCC	GGGCCC		
					AATAACTTGTAGGTTCTGTAAGGATATACCT	ACCACC		
					GAAACGCAGGCTCATGTTCTTGGACTTTGTC	TGTGTA		
					CAAAGACCAAGGGAATGAGAATACTAAGG	GGAGTG		
					CACGACTCAATTGTGAATAGGGTGCGAGAT	GGTCCT		
					AAACTTAAGACTAAGTCGCCCGTTGCTCTC	TTACCT		
					ATGCATGAGCAAAATTTCACTGTTGAGGAA	CATTTG		
					GGCCAGGTGTTCAAACCAGATATAGTGACG	AGGTGT		
					ATTTTGGGAGAGGTGGGTTACGTAATTGAT	CTCCTC		
					GTGACGGTACGGTATGAAGATAGAGATTAC	GGTCTA		
					ATCAAGGATGCCTCTGTGGAAAAAATCAGG	GTAGGG		
					AAATACGAGGCCCTTAAAGGGTACTTGAAG	GGCGAA		
					GATCTGTATCCACAGCTGAATAAGGTAGAG	GGACGA		
					GTATTGCCTCTTGTCTTTGGTAGCAGAGGGG	GGGATG		
					CCGTGCCAGGTAGTACTGTTCACAACATGG	CATCCG		
					GATTGTTAGGTTTTACGAAACGTGAAATGG	TCCCTC		
					TCCACATCTCTCGCAAAGTTATTACCGACTC	CAGGAT		
					ACTGATAATAATCTCTAATTTTTTAGAGGTT	GCTGGT		
					TATTGAGTTAGATATATATAACGACTGATTC	GGTTTC		
					TTCCCTAACTAAAATATAATTTGGAATGTGC	CGTCTG		
					ATTACTTATACTTTATTGTATTTATTGACAT	GTGGGC		

					CCGTAGTAGTCTGTTTAGTGGATTTAGATCG TTAACCATTCTGTGACGTCGCTGATGTCTAC ATGTACCAGTGGCAGAATGCTGACTAAGAA CAAAATAATTTTAAACAATAGACAGCACTA AGAAATTCTGCAAACTGGAACGTGGCCCAC GGTAGGCCAATATACCGGAAAGGGAAAAT GACAAATCCCCCCTTAAA (SEQ ID NO: 1124)	TCATGC TGAAAC CCGCAG GTGCCG ATTTGT CGAAGA GGACAT ATGGGG TAACCC ATAGAA CCTAGG CAGGAG TAATCC CTTAGC TGGGGG GGTCAG CTGGTG GGTCCT GTCAAT TTATCC TCCCCT CCATGC CAGAGC		

						CGGTCC GAGGTT AGAGAC GGACTC ATTTTC TCCTTTT TATATG TC (SEQ ID NO: 1247)		
R2	RaR2	FJ461304		Rhynchosciara americana	CAGAACGTGGAGAAACGGAATAACTACCCA GATCCGTTGGTTAACCGGTGGCAAAGTTAA TCAAGGTTGCCATAGGCTTAATAACCCTAT GGAAATGTTTCCACACACCTCCACGTGGTG CCTGCCGGAAATTGTTCTAGGGTGAACAAG CTAAGTTGTGAGAAACGGGCTCCACCACAA TATGGAGCCTGCCAGGGCGCGAGACTCAGG ACTCTCCATGTACAAAGTGGTTAGTTGCAA AAAGAGTGCGCCTAGCATGACTGATAATTT TTCACTGAAATAACGTTGAACTTTATCTGTG TCATGTGCACAACACTATGGTGTCTGATCA AGCACCATCAGTGGTGGACCTGCTAATGTA TTAGTAGAACGTGTCCAGGCGATAATGCAC	CAGAAC GTGGAG AAACGG AATAAC TACCCA GATCCG TTGGTT AACCGG TGGCAA AGTTAA TCAAGG TTGCCA TAGGCT	TAGA AATG TGTG CGAT AAGG TGTG AATA GAAG GGTT CACC AAGA GGGA GACC	MSNYNETNTSG GDNPRMATQTT GSLSSGPINQHT CELCCRTFGTRA GLGQHVRKTHPI ESNQSINVERKK RRWSPEEIRRM ANMEAQATINNI KHLTQYLATYL PQRTLNAIKGRR RDAEYKELVTGI IANLRSNSSTQQ TNQVCNESEMS

					ACACGGCCTCCGGGCCCATCGCTTTTTTTGA	TAATAA	TAGT	QRSKILQSIRESV
					GATTCCCTAGAAACTTCAGTGTGTGCGACA	CCCTAT	TTGG	RDLRSRRNKYA
					ACTGTATAACCCATAAGGGATGGACAAAGG	GGAAAT	ACCT	KALQELGEAAL
					TTATACTAGGGGGTAAAAACCCTAATCGGC	GTTTCC	CAGA	CGKMLNEEQLI
					TAATGGCAAATGGGATGTAGAAATGCCAAA	ACACAC	AATG	HCIKSMFNTAK
					GATTACTCGCACCGAATAATGGTGGCCGAA	CTCCAC	GGGT	CPKGPRFRKTAT
					AAGCGGGTAATCGAATGAAATGGTAATGCT	GTGGTG	CATA	HSGTNKQQRQQ
					GTAGCGGAAACATGATCACATTCTGTGACA	CCTGCC	GGAG	RYARVQKLYK
					GTAAACCATTAGACCTAGGGGGAACTATGA	GGAAAT	TGAT	MNRKVAAKMV
					TTAACAAGATACCAGCTTACATGGAAGCAA	TGTTCT	AGGT	LEETDKIQIKLP
					TGAAATGTAAGTCACAGTAGTGATAAGTGG	AGGGTG	TGTA	DHDPMFKFWES
					TGAAGAGTCTTGTAATCACCGTAACTAGGC	AACAAG	AAGC	EFKEGEGMPER
					CAGGTTCTGGGGATGCCATGAACTTAGGGG	CTAAGT	CGTT	MPKDLKESPDL
					GAGTATGGTTAGCAGATCTACCAGCTAACA	TGTGAG	GGGG	KAIWDPVTEEE
					CTATTACTGATAATATGTAAGCCGCAGTAG	AAACGG	AATC	VRKAKVANNTA
					CGCTAAGTGGTGTACAGATTTGCAACCACC	GCTCCA	CGGC	AGPDGIQPKSW
					GTAACTAAGTTCTGTTTCGATGGACTAGGG	CCACAA	TACA	NRISLKYKTLIY
					GGAATCATGATTAACAAGATACCAACTTAC	TATGGA	CATG	NLLLYYEKVPH
					ATGGAAGTTATGAAATGTAAGTCACAGTAG	GCCTGC	GTAT	KLKVSRTVFIPK
					TGATAAGTGGTGAAGAGTCTTGTAATCACC	CAGGGC	CTCA	KKDGSSDPGEFR
					GTAACTAGGCCAGATGTAGTCAAAGCACAT	GCGAGA	GGAG	PLTICSVVLRGF
					GTTTAGGGGGAACAAGGTTAACATTGGTAA	CTCAGG	CCAT	NKILVQRLVSLY
					AAGACCAATGCAACCTCCGTAACTAAACGT	ACTCTC	TCAT	KYDERQTAYLPI

					GAAGGCACAAAACTAAAAGTCCAGTGGATG	CATGTA	GCGC	DGVGTNIHVLA
					ACAGGTGAGGTCATCACTGGACCCAAATGT	CAAAGT	TGAT	AILNDSNTKLSE
					TTTAAGCTCATCATAACAACACGGTGAAAA	GGTTAG	CTCA	LHVALLDITKAF
					ATCCAGCATTTATTTGCCTGATTGAGTAGCT	TTGCAA	TTAA	NRLHHTSIIKSL
					TCCACACTATTCCAAAGCCGAACCTATCTG	AAAGAG	GGCG	VGKGFPYGFITFI
					GGTTTTTCTTGAAAGGCCGTATAGGGCATAT	TGCGCC	TAAT	RRMYTGLQTM
					GTCGAGAAATAAGTCCAAGGTGAGGTAGTG	TAGCAT	AAAC	MQFEGHCKMTQ
					TGGCCCTGTACCCAGGGGTAAGGTACTATA	GACTGA	TGTG	VNRGVYQGDPL
					CGGTCGAGTGGCTCAGTAGGCCTAACTAGC	TAATTT	AAAC	SGPIFLLAIEKGL
					CACTGAGTCACTATAATGACTAGTTATGAG	TTCACT	AGAT	QALDKEVGYDI
					CAACTACAATGAAACAAACACGAGTGGTGG	GAAATA	CCTG	GDVRVNAGAY
					GGATAACCCTCGGATGGCCACTCAAACCAC	ACGTTG	ATAA	ADDTDLVAGTR
					TGGATCGCTATCCAGTGGACCAATTAACCA	AACTTT	TGCC	LGLQDNINRFSS
					GCACACATGTGAACTATGTTGTAGGACATT	ATCTGT	GTGC	TIKQVGLEVNPR
					TGGTACACGTGCCGGATTGGGACAACATGT	GTCATG	TACC	KSMTLSLVPSGK
					GAGAAAAACCCATCCAATTGAGTCCAACCA	TGCACA	AAAT	EKKMKVETGKP
					AAGCATTAATGTAGAACGAAAGAAAAGAA	ACACTA	GATG	FRANDVPLKELS
					GATGGTCACCTGAAGAGATAAGACGTATGG	TGGTGT	TAAC	INDFWRYLGISY
					CGAATATGGAAGCCCAAGCTACCATAAACA	CTGATC	GAGG	TNEGPERLSLTIE
					ACATCAAACACTTAACTCAATACCTTGCAA	AAGCAC	CGGA	QDLERLTKAPL
					CTTACCTTCCTCAAAGAACTTTGAACGCTAT	CATCAG	AATA	KPQQRIHMLNA
					TAAAGGAAGGAGAAGAGATGCTGAGTATA	TGGTGG	AAAT	YVIPKYQDKLV
					AAGAACTGGTGACCGGCATTATAGCAAATC	ACCTGC	TAAT	LSKTTAKGLKR

					TTCGGAGTAATTCCAGTACACAACAAACCA	TAATGT	CTGG	TDRQIRQYVRR
					ATCAGGTCTGTAATGAAAGTGAAATGTCAC	ATTAGT	GGCG	WLKLPHDVPIA
					AACGCAGTAAGATACTGCAAAGTATACGAG	AGAACG	TTCTG	YLHAPVKSGGL
					AAAGCGTACGTGACTTGAGAAGTAGACGTA	TGTCCA	CGGA	NIPCLQYWIPLL
					ATAAATATGCGAAAGCATTACAAGAGCTTG	GGCGAT	ATGA	RVNRVNKITESQ
					GGGAAGCAGCATTGTGCGGTAAAATGTTGA	AATGCA	CTAC	RSVLAAVGKTA
					ACGAAGAACAGCTAATACATTGTATAAAGT	CACACG	TAAA	LLTSTVYKCNQS
					CAATGTTCAACACCGCTAAGTGTCCGAAAG	GCCTCC	TATA	LATLGGNPTML
					GACCAAGATTTCGCAAAACTGCAACCCACA	GGGCCC	GCGA	AYRTYWEKELY
					GTGGAACAAATAAGCAACAACGGCAACAA	ATCGCT	TGCT	AKVDGKDLQNA
					CGATATGCTAGGGTGCAGAAGTTGTACAAA	TTTTTTG	ATAT	RDDKASTRWNG
					ATGAACCGAAAGGTTGCCGCAAAAATGGTG	AGATTC	ATAC	MLHSDISGEDYL
					TTAGAAGAAACGGATAAGATACAGATAAA	CCTAGA	AAAC	NYHKLRTNSVP
					GTTACCGGACCATGATCCGATGTTCAAGTTT	AACTTC	GACT	TKVRTARGRPQ
					TGGGAAAGTGAATTTAAGGAGGGCGAAGG	AGTGTG	GATG	KETSCRGGCKST
					CATGCCGGAAAGGATGCCCAAGGATCTGAA	TGCGAC	GTAA	ETLQHVVQQCH
					GGAATCGCCCGACTTAAAAGCGATATGGGA	AACTGT	CACC	RTHGGRTLRHD
					TCCGGTAACTGAAGAGGAAGTTAGAAAGGC	ATAACC	GGCC	RIVGLLQHELRR
					AAAAGTGGCGAATAACACCGCGGCAGGTCC	CATAAG	TTA	DYNVLAKQELK
					TGATGGAATTCAGCCGAAATCATGGAACAG	GGATGG	(SEQ	TGIGLRKPDLVL
					GATAAGTTTAAAATATAAGACGTTAATATA	ACAAAG	ID NO:	IKDDTAHIVDVQ
					CAACTTACTCCTGTACTATGAAAAAGTTCCA	GTTATA	1371)	VARCSKLNESH
					CATAAGCTTAAGGTGTCACGAACTGTGTTT	CTAGGG		VRKRSKYDKKE

					ATTCCTAAGAAAAAAGATGGATCCAGTGAT	GGTAAA		IEVEVKSRYRVS
					CCTGGTGAATTCCGACCTCTTACAATATGTT	AACCCT		KVMYEACTISY
					CAGTGGTCCTTAGAGGTTTCAATAAGATCCT	AATCGG		KGIWDKQSVMS
					TGTGCAGAGGCTTGTATCCCTATATAAATAT	CTAATG		MRRLGVSEYCL
					GATGAAAGGCAAACAGCTTATTTGCCTATA	GCAAAT		FKIVTSTLRGTW
					GATGGTGTGGGCACGAACATACACGTGCTT	GGGATG		LCWKRFNMITS
					GCGGCGATACTCAATGACTCGAACACAAAA	TAGAAA		VRS (SEQ ID NO:
					CTGAGTGAACTGCATGTAGCGCTATTGGAC	TGCCAA		1493)
					ATTACGAAAGCATTCAATCGCCTACATCAC	AGATTA		
					ACATCGATAATCAAATCCTTAGTGGGAAAG	CTCGCA		
					GGGTTCCCTTATGGATTTATTACTTTCATCA	CCGAAT		
					GAAGAATGTATACGGGCTTACAAACGATGA	AATGGT		
					TGCAATTTGAGGGCCACTGCAAAATGACAC	GGCCGA		
					AAGTGAACAGAGGAGTCTACCAAGGAGAC	AAAGCG		
					CCTTTAAGCGGGCCAATCTTTTTGTTGGCTA	GGTAAT		
					TTGAAAAAGGCTTACAAGCTCTTGATAAGG	CGAATG		
					AAGTTGGCTATGACATAGGGGATGTGAGAG	AAATGG		
					TCAATGCAGGTGCTTACGCAGATGATACAG	TAATGC		
					ACTTGGTTGCAGGAACGAGACTTGGCTTGC	TGTAGC		
					AAGACAACATAAACAGGTTTTCAAGTACTA	GGAAAC		
					TTAAACAGGTTGGACTTGAGGTAAATCCAA	ATGATC		
					GAAAGTCAATGACATTGTCTCTGGTACCTTC	ACATTC		
					GGGGAAAGAGAAGAAAATGAAAGTAGAAA	TGTGAC		

					CTGGCAAACCTTTCAGAGCGAACGATGTTC	AGTAAA		
					CTCTGAAGGAACTGTCAATTAATGATTTTTG	CCATTA		
					GAGATACCTGGGCATAAGCTATACGAATGA	GACCTA		
					AGGGCCTGAAAGACTGAGTCTTACAATAGA	GGGGGA		
					GCAAGATCTAGAACGCCTAACTAAAGCTCC	ACTATG		
					ATTAAAGCCACAACAACGGATACATATGCT	ATTAAC		
					TAACGCATATGTAATACCTAAATATCAAGA	AAGATA		
					CAAACTTGTCTTGAGTAAAACCACGGCAAA	CCAGCT		
					AGGGCTGAAGCGAACAGATAGGCAGATAA	TACATG		
					GGCAATATGTAAGACGATGGCTAAAGTTAC	GAAGCA		
					CACATGATGTTCCAATTGCATATTTGCATGC	ATGAAA		
					ACCAGTTAAATCTGGTGGCCTAAACATTCC	TGTAAG		
					ATGCTTGCAATACTGGATACCATTACTTAGA	TCACAG		
					GTAAACAGAGTCAATAAGATTACAGAAAGT	TAGTGA		
					CAACGCAGTGTACTTGCAGCGGTTGGTAAA	TAAGTG		
					ACAGCCCTGCTAACGTCAACTGTTTATAAAT	GTGAAG		
					GTAATCAGTCTCTAGCAACATTGGGTGGAA	AGTCTT		
					ACCCAACAATGCTTGCATATCGAACGTATT	GTAATC		
					GGGAAAAAGAGCTATATGCAAAGGTCGATG	ACCGTA		
					GGAAAGACCTACAAAATGCGCGGGATGATA	ACTAGG		
					AGGCTTCGACGAGGTGGAATGGCATGTTGC	CCAGGT		
					ATAGTGACATTAGTGGAGAAGACTATTTGA	TCTGGG		
					ATTATCATAAGCTTCGAACAAATAGTGTAC	GATGCC		

					CCACTAAAGTAAGAACGGCAAGAGGACGTC	ATGAAC		
					CTCAAAAGGAAACGTCTTGCAGAGGTGGTT	TTAGGG		
					GTAAAAGCACGGAAACGCTGCAACATGTAG	GGAGTA		
					TTCAGCAGTGTCATAGAACACATGGGGGAA	TGGTTA		
					GAACTCTAAGACATGACAGGATTGTGGGTC	GCAGAT		
					TTCTCCAACATGAACTACGCCGAGACTATA	CTACCA		
					ATGTCTTAGCTAAACAAGAACTAAAAACGG	GCTAAC		
					GCATAGGGTTGCGCAAACCGGATTTAGTTC	ACTATT		
					TTATAAAAGATGACACTGCACACATCGTTG	ACTGAT		
					ACGTGCAAGTCGCAAGGTGTAGTAAACTGA	AATATG		
					ATGAAAGCCATGTAAGGAAGCGATCCAAAT	TAAGCC		
					ACGACAAGAAAGAAATAGAGGTAGAAGTA	GCAGTA		
					AAAAGTAGATATCGTGTAAGTAAGGTGATG	GCGCTA		
					TATGAAGCTTGCACGATATCATACAAAGGT	AGTGGT		
					ATTTGGGATAAGCAGAGCGTGATGAGCATG	GTACAG		
					AGAAGACTTGGAGTAAGTGAATACTGTCTT	ATTTGC		
					TTCAAGATTGTCACATCAACACTAAGAGGA	AACCAC		
					ACATGGCTTTGTTGGAAACGGTTCAATATG	CGTAAC		
					ATCACTAGTGTTCGATCATAGAAATGTGTG	TAAGTT		
					CGATAAGGTGTGAATAGAAGGGTTCACCAA	CTGTTT		
					GAGGGAGACCTAGTTTGGACCTCAGAAATG	CGATGG		
					GGGTCATAGGAGTGATAGGTTGTAAAGCCG	ACTAGG		
					TTGGGGAATCCGGCTACACATGGTATCTCA	GGGAAT		

					GGAGCCATTCATGCGCTGATCTCATTAAGG CGTAATAAACTGTGAAACAGATCCTGATAA TGCCGTGCTACCAAATGATGTAACGAGGCG GAAATAAAATTAATCTGGGGCGTTCTGCGG AATGACTACTAAATATAGCGATGCTATATA TACAAACGACTGATGGTAACACCGGCCTTA (SEQ ID NO: 1125)	CATGAT TAACAA GATACC AACTTA CATGGA AGTTAT GAAATG TAAGTC ACAGTA GTGATA AGTGGT GAAGAG TCTTGT AATCAC CGTAAC TAGGCC AGATGT AGTCAA AGCACA TGTTTA GGGGGA ACAAGG TTAACA		

						TTGGTA AAAGAC CAATGC AACCTC CGTAAC TAAACG TGAAGG CACAAA ACTAAA AGTCCA GTGGAT GACAGG TGAGGT CATCAC TGGACC CAAATG TTTTAA GCTCAT CATAAC AACACG GTGAAA AATCCA GCATTT		

						ATTTGC CTGATT GAGTAG CTTCCA CACTAT TCCAAA GCCGAA CCTATC TGGGTT TTTCTT GAAAGG CCGTAT AGGGCA TATGTC GAGAAA TAAGTC CAAGGT GAGGTA GTGTGG CCCTGT ACCCAG GGGTAA GGTACT		

						ATACGG TCGAGT GGCTCA GTAGGC CTAACT AGCCAC TGAGTC ACTATA ATGACT AGTT (SEQ ID NO: 1248)		
R2	YURE-2_Cis	.		Ciona savignyi	CCAAAATTACTTCCAGCACCTCCACAGCAG ACGAACGAAGAAAGAAGACTTACGAAATA AGAATAAGCAGTTTAAAGACGAAGCAGAC GAACACCACTCCACCAACGACGCTCTGCAG CTACACCACCACCATCTCGCCGCAACAAGA AGAATTTTCCGCTGCTCTCCACTCTGCTCCA ACACCACCTCTCCTGCTCTGTGGACTGCTGC CTTGCTGCTGGACCAACCTCTACCCGAAGG AACCCTTCGAACCCAGCAGCAAGGTACGTG TCACCACCTCTCCACAGAGCCAAGGCCAGA GTGGATAGAGCAGCGGCCTCTACAACCAAG	CCAAAA TTACTT CCAGCA CCTCCA CAGCAG ACGAAC GAAGAA AGAAGA CTTACG AAATAA GAATAA	TAAG ATCC GCGG CTGT GGCG CCGA ACGA GCAC CTGC CCAT TCTTC	MAGHKITMSEG KLLEVAVRYGG VRNVSYECPVP DCTKTFSQANN LIRHLNNFGNTK HRAHNFTYFFTC EKCKIQIHSNTK HNISNHYKQCC ATGGGPSCETG QYFCPACEQAG LGNLESALRHFQ

					TTCTCCACTCCGACGACAAAACACCTGCCTC	GCAGTT	TTGT	SSHPEFNLPPRS
					GTGGCCAGAGTCCTGCCGAAGAAGAATCTT	TAAAGA	AGGG	QFSKSHPNSYTL
					CGACCCCAACACCGTCTCGGCAGGGCTCCA	CGAAGC	ACTTT	SLKPKDHLMKIL
					AGCAACGACATCAGCCACGAGTGCCCACCG	AGACGA	TTCA	YSGPLTPGQLVC
					GAGTGGACGCGGCGAACCCGAGGACCGCTG	ACACCA	CCCT	PIKICLRSSAARL
					TCGCCAAGAACAATAACATCACCGCCTGCA	CTCCAC	CACT	FHDVSKLRKHM
					GCAGCTACCAAACAAAGGTTAGTCTCCTAC	CAACGA	CCCC	LVDHNRTLVYE
					CTCACCTACAAACTCGTCTGAATAGACGCC	CGCTCT	CCAA	TTCGKCLRPVD
					CCCGCGTGGGAGCTAACCTTGGTAGCTAGC	GCAGCT	TAGT	TSKNMRKTTSH
					TGCAGTGCCCTGCGACAGCGGCCTCGAAGA	ACACCA	TTTTT	FEKCSGESFISSP
					GCTGCAGAGCGTCAGCCTGCTTCGACCTCG	CCACCA	TTCGT	SPIPQKTYKLDL
					CTTGTTCTCATTTCAACCTACTCCGTCCTGT	TCTCGC	TTTTT	PSTSTPPPRKSPK
					GGAATCAAAAGAGCCCCACTAACAATTACA	CGCAAC	CGTTT	LQPYKPIRTFKN
					TTCATAAAATCTAGCAAGAACGAAGAGAAG	AAGAAG	TTTCA	PLTKSSQSKSDN
					CGACCAACTTTTAATCCATAACTTTTAGATC	AATTTT	CCCC	PPKPTPFFSPRTL
					TTTTTATTTATTACTGTTTTTAAGCCCTAAGC	CCGCTG	CACC	ERSASWPALSEV
					ATATTGCCTTTTTTTAGATCTTATAATTATA	CTCTCC	CCAC	VDPLPKLKEKHP
					AAAATAGATTCAAAGTTAACACCACCAGGC	ACTCTG	TCGC	SLPCALDKCPPS
					CGCTACAGAGCATTTTATTTAATCAATTTGT	CTCCAA	CTCT	PRIKPSTLVPPCH
					TACCGACCTCCTGCTGCTTCTTTTTACTTTCT	CACCAC	GGGC	TANNSPKPTSPE
					CCAGACTACTACCCGGATACAACCCTTGGA	CTCTCC	TGCA	SPSTLKPLPRPIR
					AACGAGAGGAATGGCCGGTCACAAGATCAC	TGCTCT	CATC	PSKPLEDWLTV
					AATGTCGGAGGGGAAGCTGCTGGAGGTTGC	GTGGAC	CCAC	RSVGPDREIVLN

					TGTACGGTATGGTGGTGTGAGGAATGTCTC	TGCTGC	ACGT	IGPRPRPGPAAG
					GTACGAGTGTCCAGTCCCGGATTGCACAAA	CTTGCT	AGGG	SRTTSPPSTAPA
					AACCTTCTCACAGGCAAATAACTTGATACG	GCTGGA	ACCT	KRVAANPIAAPL
					GCATTTAAATAACTTTGGCAATACAAAACA	CCAACC	GTTT	SGEPGATLDCG
					CCGAGCACACAATTTCACCTATTTTTTCACC	TCTACC	ATAT	QTGRKVQPPKK
					TGTGAGAAGTGCAAAATCCAAATTCACAGC	CGAAGG	TATTT	RPTESAGSLPPP
					AACACAAAACATAATATTTCAAATCATTAC	AACCCT	GCCT	AEPATDLLTGRE
					AAACAGTGTTGTGCAACCGGCGGCGGACCC	TCGAAC	TTTAT	GLARLVEEYHL
					TCGTGCGAGACGGGTCAATACTTCTGCCCT	CCAGCA	ATGT	SGDFGAFCRDLE
					GCCTGCGAGCAAGCAGGGCTGGGGAACCTG	GCAAGG	ACCA	RWTALSSTNRR
					GAGTCGGCACTCCGCCATTTTCAATCCTCCC	TACGTG	CTTTT	PKPRRGRYNRG
					ACCCTGAATTTAATCTACCCCCCCGGTCTCA	TCACCA	TAAA	AAARATRNRGR
					ATTTTCAAAATCTCATCCCAATAGTTATACC	CCTCTC	TATA	DDRQDPQDRDD
					CTATCCTTAAAACCTAAAGACCACCTTATG	CACAGA	TTTTT	QGGPGPVTCGR
					AAAATACTTTACAGTGGACCACTGACTCCG	GCCAAG	GTAC	PQRYKRAAALR
					GGGCAATTGGTATGCCCCATTAAGATCTGT	GCCAGA	CCCA	SAFGRDMKATV
					CTCAGGTCCTCAGCTGCACGTCTTTTCCATG	GTGGAT	CAAG	RRIIDGERGDAR
					ACGTGTCCAAGCTGAGGAAGCACATGTTGG	AGAGCA	ATGC	CEIDPKTIEGRFR
					TTGATCACAACCGAACCCTTGTGTACGAGA	GCGGCC	TTTTC	DELSPPVREGPE
					CAACATGTGGCAAATGCTTGCGGCCTGTCG	TCTACA	GCCA	CSLPPWMAEAQ
					ACACCTCGAAAAATATGCGGAAAACGACGT	ACCAAG	AAAA	AGEHAPSNDSQ
					CGCACTTCGAAAAGTGTTCTGGCGAATCCTT	TTCTCC	AAAA	PGDAYDGPITAL
					TATCTCATCCCCCTCACCCATACCCCAAAAG	ACTCCG	AAAT	EVEMVLSTLNV

					ACTTATAAACTCGACTTACCCTCAACCAGC	ACGACA	TTTTG	GSAPGSDGLSY
					ACCCCTCCCCCCCGTAAATCTCCCAAATTAC	AAACAC	TATC	GFWRALDPKGL
					AACCCTACAAACCAATTCGAACCTTTAAGA	CTGCCT	ACAT	VLSELFEVCRIE
					ACCCCCTCACCAAGTCCAGTCAATCCAAAT	CGTGGC	TTTTA	RRVPGPWKSSR
					CCGACAACCCACCCAAACCTACCCCTTTTTT	CAGAGT	TATTT	VTLICKDAEGDL
					CTCGCCAAGAACACTTGAGAGGTCGGCGTC	CCTGCC	TGTA	DDLGNWRPISIC
					GTGGCCCGCATTGAGCGAGGTTGTCGACCC	GAAGAA	AAAC	QTVYKIYAAVL
					ACTGCCCAAGCTCAAGGAAAAACACCCCTC	GAATCT	ACAG	ARRLQSWALDG
					GCTGCCATGCGCCCTGGATAAGTGCCCACC	TCGACC	ATTTT	GVISRSQKGFMP
					CTCACCCCGGATCAAACCCTCGACCCTAGT	CCAACA	TATA	FEGVYEHVFLL
					CCCTCCCTGCCATACAGCAAACAATAGTCC	CCGTCT	AACT	DSVVADARATR
					CAAACCAACTTCACCCGAATCACCCTCTAC	CGGCAG	TTGC	RSLAVCWLDLR
					CCTAAAACCCTTACCCCGCCCCATCCGACC	GGCTCC	ACTA	NAFGSVDHTTIV
					ATCTAAACCCTTGGAGGATTGGTTAACTGT	AAGCAA	TTTTT	EALSRFGAPAGL
					GCGGAGCGTGGGACCGGATCGGGAGATCGT	CGACAT	ATAT	VEMISDIYTGGS
					CCTGAACATCGGTCCGAGACCCCGACCCGG	CAGCCA	AAAC	CRIRTRAGFTPDI
					CCCCGCAGCAGGCTCCAGGACAACATCTCC	CGAGTG	TTCG	PVGRGVRQGCP
					CCCGTCAACGGCACCCGCGAAAAGGGTTGC	CCCACC	CACT	LSGIIFNLVMEV
					TGCAAACCCAATCGCAGCACCCTTGAGCGG	GGAGTG	TATTT	LLRGVEANNAC
					GGAGCCGGGTGCGACCCTTGATTGCGGCCA	GACGCG	AAAA	GYRLSCAGGAS
					GACCGGGCGAAAGGTTCAGCCGCCAAAAA	GCGAAC	TGAA	VRVLAYADDVA
					AGAGACCGACAGAGAGTGCTGGGTCCCTCC	CCGAGG	TCGC	LVGSSRAEXKIQ
					CTCCTCCGGCCGAGCCAGCCACCGATCTCCT	ACCGCT	ATCTT	LGVCERFAAWA

					TACAGGGAGAGAGGGGCTGGCGAGGCTGG	GTCGCC	TTTTA	GFSFNNKKCAA
					TTGAGGAATATCATCTGTCGGGGGACTTTG	AAGAAC	TATA	MVLKHQRGGRR
					GCGCTTTCTGTCGGGATCTCGAGCGATGGA	AATAAC	CACC	LLDSAPLRLCGE
					CGGCGCTTTCGTCGACCAACCGAAGGCCAA	ATCACC	AACA	EVAILGPDSFYK
					AGCCGAGGCGCGGCAGGTACAACCGCGGTG	GCCTGC	CAAA	YLGAHTGYGRQ
					CGGCTGCCCGTGCCACCAGGAACCGGGGCC	AGCAGC	CAGG	TGGQLVDRVER
					GAGATGATCGTCAAGATCCACAAGATCGAG	TACCAA	ATGT	QVVRLFTSFLTP
					ACGATCAAGGAGGCCCCGGTCCGGTGACAT	ACAAAG	GCAG	TQKLSALKRIVL
					GCGGGCGGCCGCAAAGATACAAGCGGGCC	GTTAGT	CTCA	PAMSFHLRVRP
					GCCGCGCTGAGGTCGGCCTTCGGGAGGGAC	CTCCTA	GGGG	CAEGHLRRLDN
					ATGAAGGCGACCGTCCGTCGCATCATAGAC	CCTCAC	AACC	TVRRCVKTALR
					GGAGAACGGGGCGACGCTCGCTGTGAAATC	CTACAA	AATC	LPKGSCRAFFHT
					GACCCCAAAACCATCGAGGGTCGATTTCGC	ACTCGT	CTGC	SPDAGGLGITSV
					GACGAGCTGTCGCCGCCGGTTCGGGAAGGG	CTGAAT	GTCC	VAECDILTVTQA
					CCGGAGTGCTCGTTGCCCCCTTGGATGGCTG	AGACGC	CTCCT	FKMLSSPDHLVS
					AAGCCCAGGCAGGTGAGCATGCGCCATCGA	CCCCGC	AGCG	LVAKGRLGMHA
					ATGATAGCCAACCCGGCGATGCATACGATG	GTGGGA	GCGG	ARMGRSETASA
					GGCCAATCACGGCGCTGGAGGTGGAGATGG	GCTAAC	GAGG	CAMADYLSGDS
					TCCTCAGTACCCTGAACGTCGGCTCTGCCCC	CTTGGT	GGCG	VMGHXSWKTG
					GGGATCGGACGGCCTCTCCTATGGATTTTG	AGCTAG	CCCA	YRMPADLWTAT
					GAGAGCTCTAGACCCGAAAGGACTCGTCCT	CTGCAG	CCTA	RAASRRLSLRFS
					TTCGGAGCTCTTCGAAGTCTGTAGGATCGA	TGCCCT	CCCC	PQPQGEFGLESG
					GCGCCGGGTCCCGGGGCCATGGAAGAGCAG	GCGACA	CACG	TFKIAPRERRSL

					CCGGGTCACCCTCATCTGCAAAGATGCAGA	GCGGCC	CTCCT	TRRLHHRQNLW
					GGGTGACCTCGACGATTTGGGGAACTGGCG	TCGAAG	CTTG	WRNQWAALPN
					GCCCATCTCCATCTGCCAGACGGTGTATAA	AGCTGC	AGCA	QGKTVAAHSAY
					GATCTACGCCGCCGTGTTGGCCCGTCGACT	AGAGCG	CCAA	AASNNWVKGPS
					GCAGAGCTGGGCGCTTGACGGTGGCGTCAT	TCAGCC	CAGG	SLAPQALFFGLK
					CTCCCGAAGTCAAAAGGGCTTCATGCCGTT	TGCTTC	GACT	ARLNQMPTRSV
					CGAGGGGGTATACGAGCATGTATTCCTCCT	GACCTC	CCCTT	KACYSRAPNYD
					CGACTCGGTCGTCGCTGACGCGCGCGCGAC	GCTTGT	CCGG	KSCRRCGAEVE
					CCGGAGGTCGCTGGCCGTGTGCTGGCTNGA	TCTCAT	AGCC	TLPHVLNHCPKS
					CCTCCGGAACGCGTTCGGCAGCGTCGACCA	TTCAAC	CCTG	MKSILERHDSVL
					CACCACCATAGTCGAGGCGTTGTCTCGCTTC	CTACTC	CACC	AEVLAAIPRGTF
					GGCGCCCCCGCGGGCCTCGTGGAGATGATA	CGTCCT	CTCA	ASVDVDRTSRE
					TCGGACATCTATACGGGTGGGTCCTGCCGA	GTGGAA	ACTTT	HFRRVGEALRP
					ATTAGAACTCGTGCGGGATTCACCCCGGAC	TCAAAA	TCTTA	DIVARRHDGSV
					ATCCCTGTCGGTCGTGGAGTGCGACAGGGG	GAGCCC	TTTTT	VVADVTCPFESC
					TGTCCCTTGTCCGGTATCATCTTCAACCTCG	CACTAA	AAAA	ASALDTAAARKI
					TGATGGAGGTCCTGCTGCGGGGCGTCGAAG	CAATTA	AAAA	EKYDQLCANLR
					CGAACAACGCTTGTGGATACCGGCTCTCCT	CATTCA	AAAA	QLYRKPVESHA
					GCGCCGGCGGCGCGTCCGTCAGGGTGCTCG	TAAAAT	TCAT	LVVGSLGSWGR
					CGTACGCCGACGATGTGGCTCTGGTGGGCT	CTAGCA	ATAT	TNNTALAALGIR
					CTTCCAGGGCCGAGATNAAGATCCAGCTGG	AGAACG	ATTG	GAVRSRLAKQL
					GTGTGTGCGAGCGGTTTGCCGCCTGGGCCG	AAGAGA	ATCTT	VNLSVEGSHNI
					GTTTTTCTTTTAACAACAAGAAGTGCGCCGC	AGCGAC	GACG	WLRWSGGIPKD

					CATGGTACTGAAACATCAGAGAGGGGGTCG	CAACTT	ACGG	LVR (SEQ ID NO:
					GAGGCTCTTGGACTCGGCGCCTCTTCGCCTT	TTAATC	GGGC	1494)
					TGCGGTGAAGAGGTGGCGATCCTGGGCCCC	CATAAC	TACA	
					GACTCCTTCTACAAGTACCTGGGGGCGCAT	TTTTAG	TTCA	
					ACCGGCTACGGGCGGCAAACGGGTGGACA	ATCTTT	GCCC	
					GCTTGTCGATCGAGTCGAAAGGCAGGTCGT	TTATTT	CCAA	
					CAGGCTGTTCACCTCGTTCCTCACCCCCACC	ATTACT	AAAC	
					CAAAAACTCTCGGCCCTGAAGAGAATAGTC	GTTTTT	CCAC	
					CTGCCCGCCATGAGCTTCCATCTCCGGGTCC	AAGCCC	CCAC	
					GGCCCTGCGCCGAGGGGCATCTCCGGCGCC	TAAGCA	CATC	
					TTGACAACACAGTGAGGCGCTGTGTGAAGA	TATTGC	CCCA	
					CAGCGTTGCGCCTGCCGAAGGGGTCGTGTC	CTTTTTT	ACGA	
					GGGCTTTTTTCCACACGTCCCCCGACGCAGG	TAGATC	GTGC	
					GGGCCTGGGGATCACGTCGGTTGTCGCCGA	TTATAA	CGGG	
					GTGCGACATACTGACGGTGACCCAGGCCTT	TTATAA	GCAT	
					CAAGATGCTGTCTTCACCGGACCATCTCGTC	AAATAG	TGAA	
					TCGCTAGTTGCCAAGGGCCGTCTGGGGATG	ATTCAA	GAGC	
					CACGCCGCGCGCATGGGCCGGTCCGAGACG	AGTTAA	TCCG	
					GCGTCCGCTTGCGCCATGGCGGACTACCTG	CACCAC	GCAC	
					AGCGGGGACTCGGTAATGGGGCACANGTCG	CAGGCC	AATT	
					TGGAAAACCGGATACAGGATGCCGGCCGAT	GCTACA	AGCA	
					CTCTGGACAGCCACTCGAGCTGCCAGCCGG	GAGCAT	CTTA	
					CGCCTGTCCCTCCGGTTTTCCCCGCAGCCCC	TTTATTT	GCTT	

					AAGGCGAATTCGGCCTCGAGTCGGGCACTT TCAAGATCGCCCCCAGGGAGCGGCGCTCCT TGACCCGAAGGCTGCACCACAGGCAAAACC TGTGGTGGCGGAACCAGTGGGCGGCCCTTC CCAACCAAGGGAAGACCGTCGCCGCCCACT CCGCCTACGCAGCCTCCAACAACTGGGTCA AGGGCCCGTCCTCCCTGGCCCCTCAGGCCCT GTTCTTCGGCCTTAAAGCGCGGCTGAACCA GATGCCAACGCGCTCGGTCAAGGCCTGCTA CTCAAGGGCGCCGAACTACGACAAGTCCTG TCGTAGGTGTGGCGCGGAGGTGGAGACCCT CCCGCACGTGCTGAATCATTGCCCCAAGTC CATGAAATCGATCTTGGAGCGGCATGATTC GGTGCTTGCGGAGGTCCTCGCCGCCATCCCT CGCGGCACATTCGCCAGTGTCGACGTCGAC AGGACGTCCCGAGAACATTTCCGGCGAGTG GGCGAAGCTCTGCGACCCGACATAGTCGCC CGTCGACATGACGGATCCGTCGTGGTCGCA GACGTGACGTGCCCATTCGAGTCCTGCGCG TCGGCCCTCGATACGGCGGCCGCGCGAAAA ATCGAAAAATACGACCAGCTGTGTGCGAAC TTGCGGCAGTTATACCGTAAGCCCGTCGAG TCGCACGCACTGGTCGTAGGTTCCTTGGGC	AATCAA TTTGTT ACCGAC CTCCTG CTGCTT CTTTTT ACTTTC TCCAGA CTACTA CCCGGA TACAAC CCTTGG AAACGA GAGGA (SEQ ID NO: 1249)	ATTT ATTA TTTTT TGTC AACA TTTTT GTTTT TTCA AATT TTTTC ACCC CTCA CCCC CACC CTAA TAGG TCCCT CGGG CTTG GGCC CCTTT TTCGT GCTC	

					AGCTGGGGCAGGACCAACAACACTGCTCTG		GAGA	
					GCTGCACTCGGAATCCGAGGCGCGGTTCGC		AGCG	
					TCGAGGTTGGCCAAGCAATTGGTCAACCTC		TCAC	
					AGCGTCGAGGGTAGCCACAACATCTGGCTA		ATCG	
					CGCTGGTCCGGCGGCATCCCAAAGGACCTG		CCCC	
					GTCAGATAAGATCCGCGGCTGTGGCGCCGA		ACTG	
					ACGAGCACCTGCCCATTCTTCTTGTAGGGAC		ACCA	
					TTTTTCACCCTCACTCCCCCCAATAGTTTTTT		CGAC	
					TTCGTTTTTTCGTTTTTTCACCCCCACCCCAC		CTTCC	
					TCGCCTCTGGGCTGCACATCCCACACGTAG		CCGA	
					GGACCTGTTTATATTATTTGCCTTTTATATG		CATT	
					TACCACTTTTTAAATATATTTTTGTACCCCA		GGAG	
					CAAGATGCTTTTCGCCAAAAAAAAAAAATT		TCCTT	
					TTTGTATCACATTTTTATATTTTGTAAAACA		GGCG	
					CAGATTTTTATAAACTTTGCACTATTTTTAT		TCTCC	
					ATAAACTTCGCACTTATTTAAAATGAATCGC		CAGG	
					ATCTTTTTTATATACACCAACACAAACAGG		TCGA	
					ATGTGCAGCTCAGGGGAACCAATCCTGCGT		AACA	
					CCCTCCTAGCGGCGGGAGGGGCGCCCACCT		GTCC	
					ACCCCCACGCTCCTCTTGAGCACCAACAGG		CAAG	
					GACTCCCTTCCGGAGCCCCTGCACCCTCAAC		TGAT	
					TTTTCTTATTTTTAAAAAAAAAAAATCATAT		AGCA	
					ATATTGATCTTGACGACGGGGGCTACATTC		CCTA	

					AGCCCCCAAAAACCCACCCACCATCCCCAA CGAGTGCCGGGGCATTGAAGAGCTCCGGCA CAATTAGCACTTAGCTTATTTATTATTTTTT GTCAACATTTTTGTTTTTTCAAATTTTTTCAC CCCTCACCCCCACCCTAATAGGTCCCTCGGG CTTGGGCCCCTTTTTCGTGCTCGAGAAGCGT CACATCGCCCCACTGACCACGACCTTCCCC GACATTGGAGTCCTTGGCGTCTCCCAGGTC GAAACAGTCCCAAGTGATAGCACCTAATGC TCGACTTGTTTCGGCCTGGGCCGCCGAGGA TTCCCAGAACGACCATTCTTCTAAATAATAT TTATATTTCAGAATAAAACTATATATATATC GTTGGCGGGACTTGTCCCGCCTCGATACCG AGTGCTGCAGAGCGGCAAAATAAAGAAGA AACCGACGTCGCTCTGCAGCCAAGGACCAC CCAAACTCAAGCCAGCACCGTCGACAACCA ACATCCTCAAGTCGGCGGTTGCTGGAACAA CTCATAACATCTTCAANATAAATTATCACCC TGTGCAGCAGGAGGCCGTGCTTTTAAAACT ACTCTGTAGTGGCTCATGATAATATTTCGCT CCTTTTTTGCCCCGTGTAAACTTAGTNGATG CGAATAAAATCAGTTGAATCA (SEQ ID NO: 1126)		ATGC TCGA CTTGT TTCG GCCT GGGC CGCC GAGG ATTC CCAG AACG ACCA TTCTT CTAA ATAA TATTT ATAT TTCA GAAT AAAA CTAT ATAT ATAT	

							CGTT GGCG GGAC TTGTC CCGC CTCG ATAC CGAG TGCT GCAG AGCG GCAA AATA AAGA AGAA ACCG ACGT CGCT CTGC AGCC AAGG ACCA CCCA	

							AACT CAAG CCAG CACC GTCG ACAA CCAA CATC CTCA AGTC GGCG GTTG CTGG AACA ACTC ATAA CATC TTCA ANAT AAAT TATC ACCC TGTG	

							CAGC AGGA GGCC GTGC TTTTA AAAC TACT CTGT AGTG GCTC ATGA TAAT ATTTC GCTC CTTTT TTGC CCCG TGTA AACT TAGT NGAT GCGA ATAA	

							AATC AGTT GAAT CA (SEQ ID NO: 1372)	
R4	Dong	.		Bombyx mori	GCTAGCTCCCTAAAATCCTACCTTACGTCCG AGGCGAACATCTGTCCACGTGGGGAGCGGA AACGCGTACTATCGAAACTTACGCGGCTAA CAAGGTAAAGGTAACCCATTAATATGGAGA CAAGACTAAAAAGAAAATTGAGAGGGCCG CTTCCCGGGGGCGATCGCCGGGGCACACCT GGAGCTGGCGCTGGGTGTTCCAGCATAGGA TCCGTCAGCGGCGGAGAGTTGAGCAGGCGG GCTCTCGCCGGGGATTTACAACCCGAAAAT GCTACAGCCGCCAACAGAAATCACGATGTA GAAAGTAGGAGCAATAGCCCATGTGAACCT TACAGCCCGAGTACCGGTTCATACAACCCC TCGGTACAATCATCACCATCATCCTCGGGTC ATAGAGGCTCACCAACGTCGACTATGGATG TTGCGCAGGGGGCGAATCTTTTTACCTGCGT CCACCAAAGCTGGCAAAACAAGGGGCCGC	GCTAGC TCCCTA AAATCC TACCTT ACGTCC GAGGCG AACATC TGTCCA CGTGGG GAGCGG AAACGC GTACTA TCGAAA CTTACG CGGCTA ACAAGG	TAAA AACT AGCA TAAT TATT AACT CATA ACTA ATGT ATAT TACTT GGCC AAAA GCCC GTAT ATAC	MLRRGRIFLPAS TKAGKTRGRMK WSREVNLFIMR TYYYVTKLETD LTIYRKKLHEHF SLKYPNVIISQQ RISDQKRAIERN KLLSQETLDRLK EEVRKQLEDEQ TNNVENEKLNS ETYSHEYTTLTP QTILTKKTQQHT NIISSTQTSHSST QTESITLLLENE VDILNTNPTEGA TQTQEVKDKFE

					ATGAAATGGAGCAGAGAAGTAAACCTATTT	TAAAGG	AGTT	TNLTMYSGMDP
					ATCATGCGCACCTACTACTACGTCACTAAAT	TAACCC	CCAC	KARPPLPKLKYS
					TAGAAACTGATCTGACTATTTACAGAAAAA	ATTAAT	CGGC	SKLNELIRLFNN
					AGCTGCATGAACATTTTTCACTAAAATACCC	ATGGAG	TCTGT	DILVDYISPDTQ
					TAATGTAATAATTTCACAGCAAAGAATATC	ACAAGA	CGAC	LSDVHTLTYCT
					TGACCAAAAAAGGGCAATAGAAAGAAATA	CTAAAA	AGAC	AVTISEQLKYKII
					AACTACTATCTCAAGAAACCCTAGACCGAC	AGAAAA	TGAA	AIEGNARHKKN
					TAAAAGAGGAAGTGAGAAAACAGCTCGAA	TTGAGA	CTGA	FKPPWQQRLEK
					GACGAACAAACCAATAATGTTGAAAATGAA	GGGCCG	GAAA	DIAKLRADIGKL
					AAATTAAATTCAGAAACATATTCACACGAA	CTTCCC	GGGG	TQYINNNRSKK
					TATACGACACTTACTCCACAAACTATTTTAA	GGGGGC	AAAC	VVQSVEQIFKNT
					CGAAGAAAACACAGCAACATACAAATATTA	GATCGC	ATAT	KIHTSHENGNK
					TCTCTAGCACACAGACCTCTCATTCATCCAC	CGGGGC	GGAA	KSQEFLDTLKQ
					TCAAACCGAATCAATAACTCTGCTGTTAGA	ACACCT	ATAA	KLALKAHRLKR
					AAATGAAGTCGATATTTTAAATACTAACCC	GGAGCT	TAAT	YNNSQKRKNEN
					CACAGAAGGGGCAACACAGACACAAGAAG	GGCGCT	AA	TIFLTNEKLFYR
					TTAAAGATAAATTTGAAACGAATTTAACAA	GGGTGT	(SEQ	NLIKPKTDRDNS
					TGTACTCAGGCATGGACCCTAAAGCAAGGC	TCCAGC	ID NO:	NIDIPTAEQLEM
					CGCCATTGCCTAAACTTAAATATAGCTCTAA	ATAGGA	1373)	YWARLWENSA
					GCTAAATGAACTGATACGTCTATTTAATAAT	TCCGTC		KHNDKANWITE
					GACATACTTGTAGATTACATCTCACCAGAC	AGCGGC		EKERWDTIEEM
					ACACAACTATCAGATGTACATACATTAACA	GGAGAG		QFDDVTEEEITTI
					TATTGCACCGCCGTAACTATTTCGGAACAA	TTGAGC		TARLHNWKSPG

					CTAAAATATAAAATTATAGCAATAGAAGGA	AGGCGG		IDKIHNFWFKKL
					AACGCGAGACATAAAAAAAACTTCAAACCA	GCTCTC		ICLHKTIAKNLT
					CCGTGGCAACAAAGATTGGAGAAGGATATA	GCCGGG		DIISGNQSIPEFIA
					GCAAAATTGAGAGCAGATATTGGTAAACTG	GATTTA		TGITYMIPKGDF
					ACCCAATACATAAATAATAATAGATCTAAA	CAACCC		SIEASQYRPITCL
					AAAGTAGTCCAAAGCGTTGAACAAATATTT	GAAAAT		PTIYKILTTVITK
					AAAAACACTAAAATACACACATCACACGAA	GCTACA		KINSHIEHNNILA
					AATGGCAATAAAAAATCTCAAGAATTTTTA	GCCGCC		EEQKGCRRGHM
					GACACACTGAAACAAAAATTAGCTCTAAAA	AACAGA		GCKEQLIIDSTIM
					GCCCACAGACTAAAAAGATATAACAACTCA	AATCAC		KHATTKNRNLH
					CAGAAACGAAAGAATGAGAACACTATATTT	GATGTA		CTYIDYKKAFDS
					CTAACAAATGAAAAACTATTCTACAGAAAC	GAAAGT		IPHSWLIQVLEIY
					CTAATAAAGCCGAAAACTGATCGAGACAAT	AGGAGC		KINPIIISFLRNIM
					AGTAATATAGATATACCAACAGCAGAACAA	AATAGC		THWQTTLKLKN
					TTGGAAATGTATTGGGCTAGGTTATGGGAA	CCATGT		PPNFVTTRQIAIK
					AATAGTGCAAAACATAATGACAAAGCAAAT	GAACCT		KGIYQGDSLSPL
					TGGATTACTGAAGAAAAAGAAAGATGGGAT	TACAGC		WFCLALNPLSH
					ACAATAGAAGAAATGCAATTTGATGACGTG	CCGAGT		QLHNDRAGYRI
					ACAGAGGAAGAAATAACTACTATAACAGCT	ACCGGT		KQQDNTETIISH
					AGACTACATAACTGGAAATCCCCAGGTATA	TCATAC		LIYMDDIKLYAK
					GATAAAATCCATAATTTTTGGTTTAAAAAA	AACCCC		NDKEMKKLIDT
					CTAATTTGCTTACACAAAACAATAGCCAAA	TCGGTA		TTIFSNDISMQF
					AATCTAACAGATATTATCTCTGGAAATCAA	CAATCA		GLDKCKTVHIIK

					AGTATTCCCGAATTCATAGCGACAGGAATC	TCACCA		GKVQPGDYTID
					ACTTATATGATACCAAAAGGTGACTTCTCTA	TCATCC		DTNTITAMEPSD
					TAGAAGCATCCCAATATCGACCAATTACAT	TCGGGT		LYKYLGFQQLK
					GCCTTCCGACTATTTACAAAATTTTAACAAC	CATAGA		GLDHITIKQSLTS
					AGTTATTACAAAGAAAATAAATTCACATAT	GGCTCA		EYKKRINAICKT
					AGAACACAATAATATCTTAGCTGAAGAACA	CCAACG		KLSGKHLIKALN
					GAAAGGGTGTAGACGAGGCCACATGGGCTG	TCGACT		TYAIPILTYSFGII
					CAAGGAACAGCTAATTATAGACTCAACCAT	ATGG		KWSKTDIEQIER
					CATGAAACACGCCACCACAAAAAATAGAA	(SEQ ID		ITRTTLTKHNNL
					ATTTACACTGTACATATATTGACTACAAAA	NO: 1250)		HPKSAIERLTIKR
					AAGCTTTTGATAGCATCCCACATTCATGGCT			QDGGRGMIDIW
					GATCCAAGTCCTAGAAATCTACAAAATTAA			HLWRKQIHSLK
					CCCTATAATAATAAGCTTCCTACGCAATATC			TFFYIKSDLSEIH
					ATGACACATTGGCAAACCACACTTAAATTA			RAIAQNDNNYT
					AAAAACCCTCCTAATTTTGTAACAACACGA			PLNLKQKELIDN
					CAAATAGCCATAAAAAAGGGTATTTACCAA			TENLRNRNPQK
					GGTGATTCTCTCAGCCCTTTGTGGTTTTGCC			DMEENWKKKA
					TCGCCTTGAACCCACTATCCCATCAATTGCA			LHGRHPHDLSQ
					TAATGACCGGGCGGGATACCGCATTAAACA			SHIDSKASNMW
					ACAAGATAACACCGAAACAATAATATCACA			LKTGSLFPETEG
					CCTGATTTATATGGACGACATTAAATTATAC			FLIAIQDQVINTK
					GCAAAAAATGACAAAGAAATGAAAAAGTT			NYRKYIIKDPTIR
					AATAGATACTACCACGATATTCAGCAACGA			DDKCRKCNTQP

					CATCAGTATGCAATTTGGACTTGATAAATGT AAAACCGTACATATAATAAAAGGAAAAGTC CAACCCGGTGATTATACAATAGATGACACA AACACAATAACGGCGATGGAACCAAGTGAC CTTTATAAATATCTGGGCTTTCAGCAGCTCA AAGGACTCGATCATATAACAATAAAGCAAT CATTAACTTCAGAGTACAAAAAACGTATCA ATGCCATTTGCAAAACGAAATTATCTGGAA AACATCTTATAAAAGCACTGAACACCTATG CAATACCCATTCTAACCTACTCATTTGGAAT AATAAAATGGAGCAAAACTGACATAGAAC AAATAGAACGCATAACAAGGACTACATTAA CAAAACACAATAATCTTCATCCAAAATCTG CAATAGAAAGATTGACAATTAAAAGACAAG ACGGGGGTAGAGGCATGATAGATATTTGGC ATCTATGGCGTAAACAAATACACAGCTTAA AAACATTTTTCTACATAAAATCAGATTTAAG TGAAATTCACAGAGCCATAGCACAAAATGA TAACAACTACACACCGCTAAATCTCAAACA AAAAGAACTAATAGATAATACAGAAAACCT AAGAAATAGAAACCCACAAAAAGACATGG AAGAAAACTGGAAGAAAAAAGCGCTACAT GGACGACACCCTCATGACCTAAGCCAATCT			ETIQHITGACSTL TQTDYTHRHNQ LANIIHQQLALK HKLIQNTNTPYY NYKPQTVLEND SCKLYYDRAILT DRTIHYNRPDIT LQDKNNKVTYII DIAVPNTHNIQK TFTEKMTKYTE LKEEIVRIWKQK KAYIVPIIISTTG VVPNHIHNSLKL LDLKDNIFISLQ KAAILNTCRIVR KFMQLEENQTY YTQ (SEQ ID NO: 1495)

					CACATAGACAGCAAGGCATCAAACATGTGG CTCAAAACAGGAAGTCTGTTCCCCGAAACG GAAGGATTTTTAATTGCCATACAGGACCAA GTAATAAACACAAAAAATTACAGAAAATAT ATTATTAAAGATCCCACTATTAGAGACGAT AAATGCCGCAAATGCAACACCCAGCCAGAA ACCATACAGCACATAACTGGAGCATGTTCA ACCCTTACACAGACAGATTACACTCACAGA CACAACCAACTTGCTAATATTATCCATCAAC AACTGGCTCTCAAACATAAATTGATACAAA ATACAAACACACCGTACTACAATTATAAAC CACAAACCGTTCTTGAAAATGACTCCTGTA AACTTTATTATGATCGCGCTATTCTTACCGA TAGGACGATTCACTACAATAGACCGGATAT CACTTTACAAGATAAAAACAATAAAGTCAC CTACATTATTGACATTGCAGTCCCGAATACC CACAACATTCAAAAAACGTTTACAGAAAAG ATGACAAAATACACAGAACTTAAAGAAGA AATAGTTAGAATTTGGAAACAGAAAAAAGC ATACATAGTCCCAATAATAATCTCAACCAC TGGAGTTGTCCCAAACCACATCCACAACAG CTTAAAGCTTCTAGATTTAAAAGATAACAT ATTTATTTCACTACAAAAGGCAGCTATCCTA			

					AATACATGCAGAATAGTGAGAAAGTTCATG CAGCTTGAAGAAAACCAAACTTACTACACG CAATAAAAACTAGCATAATTATTAACTCAT AACTAATGTATATTACTTGGCCAAAAGCCC GTATATACAGTTCCACCGGCTCTGTCGACA GACTGAACTGAGAAAGGGGAAACATATGG AAATAATAATAA (SEQ ID NO: 1127)			
R4	DongAG	AB097127		Anopheles gambiae	GAAGGCTAACCACAATAATGGAGACACGAT CTATGAGAAAAAGGACCACACGATTGCCCG AGGAGGGAGCCCCTACTGGAGCTGGGCCTG GGACGGGAGACAGAGCAAGCATCCAGCGG CTGGAAGATGAAATGGTGCAAGAGCGTTCT TTCAGTCAACGGGCTCTGCCCGTACCACGA ACGCAAAACAGAAACGGCAGCCCTATAAAT CACCAAGGCAATGCTGCATCTGCTAATGTT GCTGTGGCTGATAGACAACAGTCACTCATT TTGGCAGGAGGCCGACGGCAGAGGATTATG TGGACGAGAGAGATGAACCATTACGTGATC CGTTGCTACTATGTTTACACGAGGATGGAG ACGGACATGCCCGGCAGAGTGAAGATGCTG GGTATGTTCAATGACCGTTTCCCTCGGTTTG CGCATCAGCTTGACCTGAGCAAGTTGTATA TACGGCAGCGAGCTATTATTTTGCCTGAGG	GAAGGCTAACCACAATA (SEQ ID NO: 1251)	TAAC ATCC GGTG CAAA CTCA TTAA CATT AAGA AAAG AGAG AGGA GAAA TGAG AATG AGAT TCATT	METRSMRKRTT RLPEEGAPTGAG PGTGDRASIQRL EDEMVQERSFS QRALPVPRTQN RNGSPINHQGN AASANVAVADR QQSLILAGGRRQ RIMWTREMNHY VIRCYYVYTRM ETDMPGRVKML GMFNDRFPRFA HQLDLSKLYIRQ RAIILPEELEFIK LEVRREFGEEEA GWRESSRISARL

					AACTGGAGTTCATCAAGCTGGAGGTGCGGA		CACC	NTIDQNTSRASE
					GGGAATTTGGAGAGGAAGAGGCAGGCTGG		TTTG	DRDLDEPTAPGL
					CGCGAGTCGAGTAGGATTTCTGCTAGGCTT		GCAT	SVDIQHQMATA
					AATACAATAGACCAAAATACATCAAGGGCG		TTGA	VTQFHGTDPLSR
					AGTGAGGATCGTGATTTGGATGAACCCACC		ATAG	HRLPKLHYSYR
					GCTCCAGGATTGTCAGTGGATATCCAACAC		CCCG	LKTAVSIINQDV
					CAGATGGCCACAGCTGTTACGCAGTTCCAT		GGGT	LPQYLDSVGSIE
					GGGACAGACCCCTTGTCTCGTCACCGACTA		AGGT	DLQLIVYSAAV
					CCAAAACTGCATTATTCTTACCGCCTGAAA		GAAA	AVVRTLWLRTY
					ACAGCAGTAAGCATCATAAACCAAGATGTT		AGTT	PQGDSEGRPCSK
					CTACCTCAGTATTTGGATAGCGTAGGGAGC		CCCA	AEKPAWMRRLE
					ATTGAGGATCTGCAGTTAATTGTGTATTCGG		GCAT	NRINATRTKIGR
					CTGCGGTGGCTGTTGTACGAACGCTATGGTT		ATTG	MQEYQRGNSSM
					GCGGACCTATCCGCAAGGAGACAGCGAAG		CTGA	KVVRQIAEMVK
					GTCGACCATGCTCCAAGGCTGAAAAGCCCG		GAAG	PKELRDLTDANI
					CCTGGATGCGACGTCTAGAAAACCGGATCA		TGAC	TEVLDIHLQRLS
					ACGCAACACGGACAAAGATTGGTCGAATGC		AAAA	ALAKRLRRYAE
					AGGAATATCAACGGGGAAATTCATCTATGA		TTCG	CSKRKEQNRMF
					AGGTGGTACGTCAGATTGCTGAAATGGTTA		GATA	NINEREFYNWIR
					AACCTAAAGAACTACGAGACCTCACTGATG		ATAA	NDKPNFREGLP
					CCAACATAACGGAGGTACTCGACATCCATT		TAAT	DIGDFTQFWAN
					TACAACGGTTGAGTGCCCTTGCAAAACGAT		AATA	LCEKPVQHNSE
					TACGACGTTATGCTGAATGCTCGAAGCGGA		ATAA	GMRLAEDERFS

					AAGAACAAAATCGAATGTTCAACATTAACG		TAAT	DGIEDMPVLVV
					AGAGAGAATTTTACAACTGGATCCGAAATG		AATA	NAQDIREATQY
					ATAAGCCCAATTTTAGAGAAGGGCTCCCGG		ATAA	TRNGAAPGPDF
					ATATTGGCGACTTTACACAGTTTTGGGCCAA		TAAT	VYNFWYKKLITI
					TCTATGTGAGAAACCTGTCCAACACAACAG		AATA	HEQIAACFNTVL
					CGAAGGAATGAGGTTAGCAGAAGATGAGC		ATAA	EDSRKLPKFITG
					GCTTCAGTGATGGTATCGAAGACATGCCCG		TAAT	GVTYFLPKDQN
					TGCTAGTTGTGAATGCTCAAGACATACGTG		AATA	TKNPAKYRPLT
					AGGCAACGCAGTACACCAGGAATGGAGCTG		ATAT	CLSNLNKVLSSV
					CACCAGGACCCGATTTTGTATACAATTTTTG		GCAT	ITQKVKDHCDT
					GTATAAAAAGCTAATCACAATCCATGAGCA		AATA	NNVMTEEQTGR
					GATAGCGGCATGCTTCAATACGGTGTTGGA		ATA	RKNTQGCKDQV
					AGATTCGAGAAAACTACCAAAATTTATCAC		(SEQ	IIDAVIVGQAAK
					CGGGGGAGTTACTTACTTTCTACCAAAAGA		ID NO:	KQRNLDMAYID
					TCAAAACACAAAAAATCCTGCGAAGTATAG		1374)	YKKAYDSVPHS
					ACCACTTACCTGTCTTTCTAACTTAAACAAA			YLLKVLQLYKV
					GTGCTGTCGTCAGTGATAACGCAGAAAGTG			DGNVIKLMQHA
					AAAGATCATTGCGATACCAACAACGTAATG			MGMWSTSLHVT
					ACCGAAGAACAGACAGGACGTCGAAAAAA			DGKVVLRSRSL
					CACGCAAGGCTGTAAAGACCAGGTCATTAT			NIRRGIFQGDTF
					TGATGCAGTCATTGTTGGTCAAGCAGCCAA			STLWFCLAMNP
					GAAACAAAGAAATCTGGATATGGCATACAT			LSRTLNQQCNF
					CGATTACAAGAAGGCGTATGATTCAGTACC			GYLLKSEEISTRI

					CCATTCATACCTTCTTAAGGTACTCCAGTTG			THTFFMDDLKL
					TACAAAGTAGACGGGAATGTCATCAAGCTG			FAETVQKMHHL
					ATGCAGCACGCGATGGGTATGTGGAGTACA			LKNVQGFSNDIK
					TCTCTACACGTTACCGACGGAAAAGTTGTA			MEFGIGKCRSIH
					CTACGGTCAAGATCACTCAATATCAGGAGG			LHRGQVLDADS
					GGTATTTTCCAAGGTGACACCTTTAGTACGC			FRANEQEEIRHM
					TGTGGTTTTGTCTAGCTATGAACCCGCTTAG			VQGETYKFLGF
					CAGAACACTCAACCAGCAATGCAACTTTGG			LQLRGIHYAVIK
					GTATTTACTCAAAAGTGAAGAAATAAGCAC			KELQDKFLHRV
					GAGAATCACCCACACCTTCTTTATGGATGA			SCILKSFLSVGN
					CTTGAAGCTGTTCGCAGAAACAGTACAGAA			KVKAINTFAVA
					GATGCACCACCTGTTGAAGAACGTGCAGGG			LLTYSFGVMKW
					ATTCAGCAACGACATTAAAATGGAATTTGG			SNTDLEALERTI
					TATCGGTAAATGTCGATCAATTCATCTACAC			RVVSTKHQMRH
					CGAGGTCAAGTATTGGATGCCGATAGCTTC			PKASVERVILPR
					CGTGCCAACGAACAAGAGGAAATCCGCCAC			KIGGVGIIDIQAL
					ATGGTTCAAGGTGAAACTTACAAGTTCCTC			CISQIHQLRSYF
					GGTTTCCTGCAGCTGAGGGGTATTCACTATG			VESQNRHELYR
					CAGTGATCAAGAAAGAGCTACAGGACAAGT			TVYKADHGLSA
					TCTTACATCGTGTTAGCTGTATCCTGAAGAG			LHLAQQDYQLN
					CTTTTTGTCAGTCGGCAACAAGGTGAAAGC			CNIKTVDGKGA
					AATAAACACATTTGCGGTGGCTCTGTTGAC			TWKQKELHGTH
					CTACAGCTTTGGAGTAATGAAATGGTCTAA			THQLNLEHIDKV

					TACTGACTTGGAAGCGTTGGAGCGAACAAT TCGTGTGGTTTCCACTAAGCACCAAATGCGT CACCCAAAAGCGTCCGTCGAGAGAGTAATC CTGCCACGAAAAATAGGAGGGGTAGGAATC ATTGATATTCAGGCACTTTGTATTTCTCAGA TCCATCAGCTGCGAAGTTACTTCGTGGAAA GCCAAAACCGACATGAATTATACCGCACTG TGTATAAAGCAGATCACGGATTAAGCGCCC TGCATCTAGCGCAGCAAGATTACCAGCTGA ATTGCAACATAAAAACCGTCGATGGAAAAG GCGCAACGTGGAAACAGAAGGAGTTACATG GGACGCACACCCATCAACTGAATCTGGAAC ATATCGACAAAGTGTCATCTAGCACTTGGC TTGTGAGGTGTGACCTTTTCTGTGAGACAGA AGGTTTCATGGTAGCCATCCAAGACCGGGT AATTGCGACGTGGAACTATCGGCGGTGTAT ATTGCGTGAAGACGTGGAGGACCGATGCAG AAAGTGCAACTCAGGAGGAGAATCGATTGA GCATGTCATTGCCGGCTGTCCAGTGCTAGCT GGGTCAGCGTATCTCGATCGCCACAACGAC GTTGCCAAAATTGTTCACCAGCAGCTTGCA CTGAGGCACAAGTTGGTAGAGCGATTTTTA CCCTGCTACCGATACCTCCCAGATCCGGTCC			SSSTWLVRCDLF CETEGFMVAIQ DRVIATWNYRR CILREDVEDRCR KCNSGGESIEHV IAGCPVLAGSAY LDRHNDVAKIV HQQLALRHKLV ERFLPCYRYLPD PVQENDCIKLY WDREIITDILIRA NRPDILVYEKRK KRATIDIDIAVTL DHNVQTTFSTK VMKYHDLAEEL KQTWYLEDIRIV PVIISATGIVPMA LLRSLDELELQR ELPRIQKAVILR TCSTLRRFLNPY N (SEQ ID NO: 1496)

					AGGAAAATGATTGCATAAAGCTGTATTGGG ATCGCGAAATTATAACGGACATCCTCATCC GTGCCAATAGGCCAGACATCTTAGTCTACG AGAAAAGAAAGAAACGAGCGACCATCGAC ATCGACATTGCTGTAACGTTAGACCATAAT GTTCAGACAACATTTTCCACCAAGGTGATG AAGTATCATGATCTGGCAGAGGAGTTGAAG CAGACGTGGTATCTGGAGGATATCCGCATT GTTCCGGTAATCATCTCGGCGACCGGAATT GTACCTATGGCCCTCTTACGTTCCCTGGACG AGCTCGAACTGCAGAGAGAACTACCCAGGA TTCAGAAGGCGGTGATTCTTCGAACATGTA GCACTTTAAGAAGGTTCCTGAATCCCTATA ACTAACATCCGGTGCAAACTCATTAACATT AAGAAAAGAGAGAGGAGAAATGAGAATGA GATTCATTCACCTTTGGCATTTGAATAGCCC GGGGTAGGTGAAAAGTTCCCAGCATATTGC TGAGAAGTGACAAAATTCGGATAATAATAA TAATAATAATAATAATAATAATAATAATAA TAATAATAATAATATGCATAATAATA (SEQ ID NO: 1128)			

R4	R4-1_BX	CADV01008175		Bursaphelenchus xylophilus	GGGATCCTGGGTTCCTACTACCTCGCTCCAC	GGGATC	TAAG	MTCNNAVVFPP
					CTCCTCGCGATGGATCCTGGGGAAGTCTCC	CTGGGT	AAAA	ADGNPAGTADR
					GGACTGAGCTAAGAGAGCGTTAAAGTAGAG	TCCTAC	GCAT	NFAIRFPSSEPPG
					GGTGACGGCGTAAGTACCTCCAAGTTGCGG	TACCTC	GAAA	PSGIRPSEPLDGR
					TGGAGCGGAACATCTACTCTTCGGAGAGAG	GCTCCA	TAAT	TGIGDVEHAQA
					GGGAAGCTCTATGGCGGCGTTAGAAAGGTT	CCTCCT	AAGA	GNGGFLVDVLE
					GGACTACGGCAACGCCAGGGAGATGGGGA	CGCGAT	AATC	YKEAHRYGSKC
					AGGTTCATCAGGTGATACTAGTTCGCTACTG	GGATCC	AGAT	EFCYVQTKGTV
					TCATTCGATGTATCCGGAACCATACTCGCCA	TGGGGA	AAGA	CSKPRTDAWLK
					AGTTGTGAACTATGTGAAAGTCTGGATCCA	AGTCTC	ATAA	CEILFLLHHAYT
					ATCCAAGACCACGGGGCGCAATTAAAAGGT	CGGACT	CAAG	ANQNKSIELAES
					GTGAAGCAGCTTGCTGGTGATCACCACGGT	GAGCTA	AATA	AFRRAGITRRSK
					GGTACCTACACCGGCGGGGGAACATCTTTA	AGAGAG	CTAA	ATIAKRWSLIQR
					ATGCCGAGATGACCGCGCGAATAATAGGGA	CGTTAA	TAAG	GKGTDYKEYW
					GCCGGCAAAAGCCGAGCGTAAGTGGAAAG	AGTAGA	TATA	DEYFEKFRYEC
					GATACAGAAATTTGCTCAAAAGTATACCAA	GGGTGA	TCAT	NPTPIVRRKRNR
					CCCGCCACTTATAAAAGACCGAGAAAAGGG	CGGCGT	GTAA	LAAGLQSPSSVP
					TACCACTAGAGCTCATTATAAAACATCTAT	AAGTAC	CTAT	NGYEFERKRTC
					GACATGTAATAATGCTGTGGTCTTTCCGCCA	CTCCAA	GACA	ETPLDTKASSLP
					GCCGATGGTAACCCCGCTGGGACTGCGGAT	GTTGCG	AAAA	LICNLLTGIVGV
					AGGAATTTTGCTATTAGGTTCCCCTCCAGTG	GTGGAG	GAAC	ENVEENMSVEC
					AACCTCCAGGACCCTCTGGAATACGACCTT	CGGAAC	GCAC	TEPKELSGTANS
					CTGAGCCTCTGGACGGTAGAACCGGGATTG	ATCTAC	CAAT	SVPGLAEGVYE

					GCGACGTTGAGCACGCCCAGGCGGGAAATG	TCTTCG	AAGA	RRHNNVNEPAA
					GAGGATTCCTGGTTGATGTTCTGGAATATA	GAGAGA	ACAT	GCPQDVPVANN
					AAGAAGCACATCGCTACGGCTCCAAGTGTG	GGGGAA	GCTT	LIDSPTTNDRLE
					AGTTTTGTTATGTGCAAACCAAAGGTACCG	GCTCTA	GAGT	AEFKAQLDRAE
					TGTGCTCTAAGCCTCGGACCGATGCGTGGTT	TGGCGG	GGCC	RSYMRRRLPRL
					GAAATGCGAGATTCTGTTCTTGCTTCATCAT	CGTTAG	AGCT	KNLSPDERMWI
					GCCTATACGGCTAATCAGAACAAGAGTATT	AAAGGT	CTGC	GTTVERLRLETV
					GAACTCGCCGAGAGTGCTTTTCGGCGAGCA	TGGACT	AGGC	SEPVCEQWRLA
					GGAATTACCCGACGTAGCAAGGCTACCATC	ACGGCA	AAAA	NAGLYAAIRSIA
					GCGAAAAGATGGTCCTTAATACAGCGAGGT	ACGCCA	GTCG	VMRPLDAAREA
					AAAGGCACCGACTATAAGGAATACTGGGAC	GGGAGA	AATT	HKTWLLNMKM
					GAGTACTTCGAAAAATTCAGATACGAGTGC	TGGGGA	TGGA	TERKLRQQIGW
					AACCCCACACCTATAGTACGGCGCAAACGA	AGGTTC	ACAG	VETTRRTKNEA
					AACAGGTTAGCCGCTGGGTTACAGTCGCCC	ATCAGG	CCGG	RTERQEIVYRKV
					AGCAGCGTGCCGAACGGATATGAATTTGAG	TGATAC	TAAT	AKLRRERFPEM
					CGGAAAAGGACTTGTGAAACGCCTCTTGAT	TAGTTC	GGAA	DLDSVSVHLKR
					ACCAAGGCTTCAAGTCTTCCACTAATCTGTA	GCTACT	GACC	KLELLKGRIQVR
					ACTTGCTAACAGGAATTGTCGGAGTTGAGA	GTCATT	TGCA	TAERLRRDTRE
					ATGTTGAGGAAAATATGAGTGTAGAATGCA	CGATGT	ACAA	AAGPYGKTALR
					CTGAGCCGAAAGAGCTAAGTGGAACTGCGA	ATCCGG	ACGT	GQGFAPNVKDA
					ACTCTTCTGTTCCGGGTTTGGCAGAAGGAGT	AACCAT	GGGG	TQYWSGLAQPS
					ATATGAAAGACGACATAACAATGTCAACGA	ACTCGC	TAGC	GQKCSENSAILS
					GCCAGCGGCTGGGTGTCCCCAAGATGTCCC	CAAGTT	AGGC	DWKELVECNLS

					AGTCGCAAACAACTTGATCGATTCCCCAAC	GTGAAC	AATA	SLPDQMEPLVV
					TACTAACGATCGATTAGAAGCTGAGTTTAA	TATGTG	TGTA	QGISRASPWKSP
					GGCACAGCTGGATAGAGCGGAACGGTCTTA	AAAGTC	ACTA	GPDGIFNYYWR
					TATGAGACGAAGGCTACCTAGGCTGAAGAA	TGGATC	TGAC	QDFIVDWLKQL
					CCTATCCCCTGACGAAAGGATGTGGATTGG	CAATCC	AGAC	MLDSLRTGHYP
					AACAACAGTCGAGCGCCTAAGACTCGAGAC	AAGACC	CAAA	WKLSSGRTVLL
					TGTGAGCGAACCTGTATGTGAACAGTGGCG	ACGGGG	ACTC	YKDGDPTKAEN
					GCTTGCCAACGCTGGGCTATATGCAGCTAT	CGCAAT	CGAA	YRPITCLNGCFK
					CCGGTCAATCGCGGTAATGCGCCCCCTGGA	TAAAAG	ACTC	MINSVVSEVILK
					CGCCGCCCGCGAAGCCCATAAAACTTGGCT	GTGTGA	TGGT	RVENTIALPIEQ
					TCTAAATATGAAAATGACGGAAAGGAAACT	AGCAGC	AATG	MALRRKVWAC
					AAGACAGCAGATTGGCTGGGTCGAGACGAC	TTGCTG	AGCC	VESQIWDQIKQR
					GCGCCGTACTAAGAACGAGGCACGTACTGA	GTGATC	CGTG	KLSDRTQKCKV
					AAGACAAGAAATTGTCTATAGAAAAGTGGC	ACCACG	CCCC	AWVDFSKAYDS
					CAAACTGCGAAGGGAAAGATTCCCCGAGAT	GTGGTA	CCAA	LNHDAIKFVIGV
					GGACCTGGACAGTGTGTCAGTCCACCTCAA	CCTACA	GCAT	LKLPTGINNYLL
					ACGCAAGCTTGAGTTACTTAAAGGGCGGAT	CCGGCG	GTGG	DSMQNWSTHLE
					TCAAGTGAGAACCGCAGAGAGGTTGCGAAG	GGGGAA	TCTC	LKSSGKVVRGPS
					GGATACCAGAGAGGCCGCTGGTCCATATGG	CATCTT	GTTC	YPIKRGVLQGDS
					TAAGACTGCTCTTCGGGGTCAAGGGTTTGC	TAATGC	GATG	LSPTLFVVVTSII
					CCCGAATGTAAAAGACGCGACTCAGTACTG	CGAGAT	TAGT	VRHIKTIESSDIQ
					GTCTGGATTGGCCCAGCCATCCGGACAAAA	GACCGC	TAGG	MYMDDIKLYGK
					ATGTTCTGAGAACTCGGCGATTCTCTCAGAT	GCGAAT	AACA	DQETLTRLIKEL

					TGGAAAGAACTCGTCGAGTGTAATTTAAGC	AATAGG	GTTCT	QTVSNKLGLCM
					AGCTTGCCGGATCAGATGGAGCCGCTGGTA	GAGCCG	CTTTA	NLKKCAILGDD
					GTCCAAGGCATAAGTCGGGCTAGTCCATGG	GCAAAA	ACCC	LPEEINGIEHLKE
					AAATCTCCTGGACCGGACGGTATTTTTAACT	GCCGAG	GTGA	SYKYLGVPQREI
					ACTATTGGCGACAAGACTTTATTGTGGACT	CGTAAG	TGAT	TQVRATMAALE
					GGTTAAAGCAGCTAATGCTGGATTCCTTGA	TGGAAA	TACG	KKILTEVDTSLG
					GGACCGGGCACTATCCGTGGAAACTTTCAA	GGATAC	CCCT	AAELSYRQRISR
					GTGGACGCACGGTGCTCCTGTACAAAGACG	AGAAAT	GTCTT	VNSKIAPLVRFV
					GGGATCCGACCAAAGCCGAGAACTACCGAC	TTGCTC	AAAT	VQSMLVTPRDV
					CTATCACCTGCCTAAACGGATGCTTCAAAA	AAAAGT	GGCA	LKVYNRLGGID
					TGATCAACTCGGTTGTTTCTGAAGTTATCCT	ATACCA	GGTG	VEIRRRLVKYEI
					GAAACGCGTCGAAAATACGATAGCTCTCCC	ACCCGC	CCAC	RYKKSNVARLY
					GATAGAGCAAATGGCTTTACGACGTAAGGT	CACTTA	CAAA	LDRKVGGIGFV
					ATGGGCATGTGTTGAGTCTCAGATATGGGA	TAAAAG	TACC	NLCRIMVEAVA
					TCAGATCAAACAGCGGAAGCTGTCTGATCG	ACCGAG	GAAC	ARAVYCRLAPS
					AACACAGAAATGCAAGGTAGCGTGGGTAG	AAAAGG	ACTC	FNEFQDFLAEQN
					ATTTCTCAAAGGCTTATGACAGTCTTAACCA	GTACCA	GTTG	TSPITAAQTILDK
					TGATGCTATAAAGTTCGTCATAGGAGTTCTA	CTAGAG	AGGT	CGINIELSTSTLG
					AAGCTTCCGACAGGTATTAACAACTACCTG	CTCATT	ATGG	DVKKIVRNHYH
					CTTGACAGTATGCAGAATTGGAGCACGCAC	ATAAAA	TGGT	ELWLTAWKNTG
					CTCGAGCTTAAAAGCTCTGGAAAGGTAGTA	CATCT	CCGA	LYKRWENDHV
					CGAGGCCCTAGTTACCCGATTAAAAGGGGC	(SEQ ID	ATGT	DIKRSSLWINRG
					GTTCTTCAAGGCGATAGTCTAAGCCCAACA	NO: 1252)	GAAG	NLSANNARIGIG

					CTGTTTGTCGTGGTGACCAGTATAATAGTAC		CTGG	IQDNSIFCRGFV
					GTCATATCAAGACGATCGAAAGTAGTGATA		GAGT	GNKCDTKYCRL
					TACAAATGTATATGGATGATATCAAGTTGT		ACAA	CGDGIESVSHIV
					ATGGCAAAGATCAGGAAACTCTGACACGGC		TTTG	TGCPTHRTNLYI
					TAATAAAAGAGCTGCAGACGGTATCGAACA		GTAC	ERHDCVARNVY
					AGTTAGGTCTATGTATGAACCTTAAGAAAT		GAGA	AYLAIRYGIPVP
					GCGCAATTCTCGGCGACGACCTCCCAGAGG		GCAC	HYTQRVKTIEK
					AAATCAACGGGATTGAGCACCTTAAGGAAT		CAGC	NGDQSVELYWN
					CATATAAGTACTTAGGTGTTCCCCAACGGG		GCCC	YKFPCTRALEAC
					AAATCACCCAAGTGCGTGCCACCATGGCCG		CCGA	RPDIVLIDKVSK
					CACTTGAAAAGAAAATACTAACTGAAGTAG		TCTA	RTHIIEVAVSWR
					ATACTAGTCTTGGAGCAGCCGAACTAAGTT		AGTG	GRLQEMVDRKV
					ACCGCCAACGGATATCGCGTGTTAACTCTA		ATGA	YKYTVNGEYEA
					AGATCGCGCCGTTGGTGAGGTTCGTTGTGC		CGCA	DGSSRGWNIVR
					AGAGTATGCTGGTCACGCCACGCGATGTGC		TGCG	ELNDQYGFPVE
					TTAAAGTGTACAATAGGCTTGGAGGCATTG		TCGG	VYTLVIGAGGEI
					ATGTCGAAATTCGAAGAAGACTCGTCAAGT		AACA	LPCTVKDVERLT
					ACGAGATCAGATACAAGAAAAGTAATGTAG		ATGA	GGAATDNLIER
					CCCGGCTCTACCTTGATAGAAAAGTTGGCG		AGAC	MERSAVLGSCRI
					GCATTGGTTTCGTAAACCTCTGCCGGATAAT		GGCT	IKRHLAL (SEQ
					GGTCGAGGCGGTTGCAGCCAGAGCTGTCTA		GGCA	ID NO: 1497)
					CTGCAGACTTGCTCCCTCGTTTAATGAGTTC		AACA	
					CAAGACTTCCTTGCAGAACAGAATACCAGC		TTCA	

					CCAATCACTGCTGCACAGACAATTTTAGAT AAGTGCGGCATTAATATTGAGCTGAGTACA TCTACTCTGGGAGATGTAAAGAAAATAGTG CGAAACCACTATCACGAGCTTTGGCTCACG GCATGGAAGAACACTGGGCTCTATAAACGT TGGGAGAACGACCATGTAGACATTAAAAGA TCTAGTTTATGGATCAATCGAGGTAACCTA AGTGCGAATAACGCTCGTATTGGAATTGGC ATTCAAGACAACAGTATCTTTTGTCGAGGCT TTGTTGGGAACAAGTGTGATACGAAATATT GCCGATTATGTGGGGATGGCATAGAATCCG TAAGCCATATTGTAACTGGTTGTCCGACTCA CCGTACAAACCTGTACATAGAGCGGCATGA CTGTGTTGCAAGAAATGTTTACGCATACCTC GCCATAAGGTATGGAATCCCTGTGCCACAC TATACGCAGCGAGTCAAGACAATAGAGAAA AACGGCGATCAGTCTGTAGAGCTATACTGG AATTACAAGTTCCCATGCACCCGTGCACTG GAAGCTTGCCGTCCGGATATTGTTCTCATAG ATAAGGTTAGTAAGCGAACTCATATAATTG AGGTGGCTGTCTCATGGCGGGGACGGTTAC AGGAAATGGTCGATAGGAAAGTGTACAAGT ATACGGTAAACGGAGAATACGAAGCCGATG		GGAG TCGC AAA (SEQ ID NO: 1375)	

					GCTCTTCCCGTGGGTGGAATATCGTCCGTGA ACTAAACGATCAGTACGGCTTTCCTGTCGA GGTATATACGCTAGTGATTGGTGCTGGAGG CGAAATCCTACCATGCACGGTTAAAGATGT AGAACGGCTCACCGGTGGAGCGGCTACTGA TAACCTCATTGAGCGTATGGAGCGGAGCGC AGTCCTTGGCTCATGTCGTATCATTAAAAGG CATCTGGCACTCTAAGAAAAGCATGAAATA ATAAGAAATCAGATAAGAATAACAAGAAT ACTAATAAGTATATCATGTAACTATGACAA AAAGAACGCACCAATAAGAACATGCTTGAG TGGCCAGCTCTGCAGGCAAAAGTCGAATTT GGAACAGCCGGTAATGGAAGACCTGCAACA AACGTGGGGTAGCAGGCAATATGTAACTAT GACAGACCAAAACTCCGAAACTCTGGTAAT GAGCCCGTGCCCCCCAAGCATGTGGTCTCG TTCGATGTAGTTAGGAACAGTTCTCTTTAAC CCGTGATGATTACGCCCTGTCTTAAATGGCA GGTGCCACCAAATACCGAACACTCGTTGAG GTATGGTGGTCCGAATGTGAAGCTGGGAGT ACAATTTGGTACGAGAGCACCAGCGCCCCC GATCTAAGTGATGACGCATGCGTCGGAACA			

					ATGAAGACGGCTGGCAAACATTCAGGAGTC GCAAA (SEQ ID NO: 1129)			
R4	R4-1_HG	.		Heterodera glycines	TGGCGATACTCGGAACCTCCGGGGAGCCTG GTAGGAGTTGGCCTACAGGTCGCGAAAGTC CCTAGGTGCTGCACGGGTTGCGCTAATCCG AGGGCGCTGGGTTACCTTCCCATCGGCCAA AAACGTCTGGGCCTTCTTAGCCGCGGGTCC AGTATTTCTGTTGAGCCTGACAGTTCTTCCC GGATATGGCGAAAGATTACAGGGCGGTATT TCGTGAAACCTAAAAAWGGTCGGGCCGAA TGGCACGGACAGACTCACTTCGGAGTGAGC TCGGGGGCATCCGTGTGTTACCCCGCTGCA CCACGCCGAAGCTGTCATAGCGAGCCCGAA GGGGAATGGCCATGGAGACTCCAGCCTCAC CCTGTAACTCGAACCTAAGTCCAGGCCCCTT TCTGGTGTTGGCTCGCACTGGTTAGGAACA CGACAGTCTCGTGTAATCCCACACGCACCG AAAGCCCTAGCCCTCGTAGGCGAAGGCTGC GTTGCTGGTTTCAGAACTGTGAGCWAGTGG GGTTCGGATGGCCGAGTGTACCAACCTGCT TGCTAACAGGTAGCATAGAGTAATATGCTA GTAAGCAGGGCACGAGAAAGGGCCGTAAA GGCTCCGGCTATGAAGGACCTTGCGACCAC	TGGCGA TACTCG GAACCT CCGGGG AGCCTG GTAGGA GTTGGC CTACAG GTCGCG AAAGTC CCTAGG TGCTGC CCTC ACGGGT TGCGCT AATCCG AGGGCG CTGGGT TACCTT CCCATC GGCCAA AAACGT	TGAG GACT CANA ATTG ACAA TACA CCTC AGA (SEQ ID NO: 1376)	MISCDLERETLT QMALFRARSDK TPTHAGIPAPDE VREGGCGQNRT NPAAPRGKAAA IQRQNGITIPIXA CAQSGLVRTQR VQQWSAVEESA LKDVVVRNTDD RGLINWAKGVL PEWQRLCQLNP TMYMARSSPSL SNKWASLRRTH VGPGCPSKEGSG PSQDLSDVKIQP ARLAHDTVAEL PQRTVPCGTDG HGVIDSDETETA LAEVSRSSPFGE REPLDLGATERI TRKRLRNAVRD

					GCGTGTGTCTCCCACGTGCGGATTCTTGAAG	CTGGGC		VVPPRKRRVPST
					CCAGAGTCTTGCACTGCGCGCAGGATGGAG	CTTCTT		PSRKEQDLVPEV
					CCTGTGCAACTCCTCCCTCGCTGATCGCAGG	AGCCGC		DGPAPTDVLTHP
					AGTGGAGGATCACCACTCTTTTTACCTTGCT	GGGTCC		PTESEPEPMLDP
					AGCTTGGGGTACCACCTTGAGCTGGGGCCG	AGTATT		LSLVQLVRPQLG
					GCCTTGCTAGCTTGGGGTACGACCCTTTGAG	TCTGTT		RAMGWAAEEM
					CTGGGGCTGGCCTTACTAGCTTGGCGCGCC	GAGCCT		ELGNVVMDVEL
					ACCCTTGGAGCTGGGGTGGCGCAAGATCAC	GACAGT		KREFNREVRRV
					TTGTATACGGTCTAACCAATACATTTGAAA	TCTTCC		GRTPPDQMYKR
					AGCGATCAAAGCGAAATGATATCCTGTGAT	CGGATA		GAGPPLPQKREP
					TTGGAAAGAGAAACCTTAACCCAAATGGCT	TGGCGA		ERVALLEQLIAA
					CTCTTTAGAGCCCGCAGTGACAAGACGCCG	AAGATT		RVERGINRGLD
					ACCCACGCCGGGATACCAGCCCCTGATGAG	ACAGGG		WFLELNVAVFA
					GTCCGAGAGGGAGGGTGCGGCCAAAACCG	CGGTAT		AARVLSRRERV
					AACCAATCCCGCGGCCCCCCGTGGTAAGGC	TTCGTG		ETLADRLHINDS
					GGCAGCGATTCAAAGGCAAAACGGCATAAC	AAACCT		ATLSEVSRRRAK
					TATTCCTATAAAWGCCTGCGCTCAATCCGG	AAAAA		AERKLRCAREQ
					CCTAGTCCGGACCCAGAGGGTTCAGCAGTG	WGGTCG		PWMSRRIRXLG
					GTCAGCCGTGGAGGAGTCCGCCTTAAAGGA	GGCCGA		VRVERLKQLAD
					TGTGGTGGTGCGAAACACAGACGACCGGGG	ATGGCA		LVRQRIAGRGN
					GCTCATTAATTGGGCCAAAGGTGTGCTCCCT	CGGACA		RSSYEGPRRRFR
					GAATGGCAAAGATTGTGCCAGCTCAACCCG	GACTCA		LRPSLRSVTEAP
					ACCATGTACATGGCTCGATCGTCTCCATCGC	CTTCGG		VNPPLNGNEVY

					TGTCCAATAAGTGGGCATCTCTGCGGCGGA	AGTGAG		TFWHSLWAQSL
					CCCATGTTGGTCCTGGATGCCCTTCGAAAG	CTCGGG		RANTDDCQLRE
					AAGGCTCCGGGCCCTCACAAGACCTTTCAG	GGCATC		FKNQLSAARHT
					ATGTAAAAATCCAACCAGCCAGGTTGGCCC	CGTGTG		DLTSVGTSSLVQ
					ATGATACGGTGGCCGAGCTACCTCAACGTA	TTACCC		MFSAALRKMKK
					CTGTCCCTTGCGGCACAGACGGTCATGGTG	CGCTGC		GKAPGPDGIRA
					TAATCGATTCCGATGAGACCGAGACTGCCT	ACCACG		AWWGVFRRIAP
					TAGCWGAGGTATCCCGGTCGAGTCCCTTTG	CCGAAG		YVATWVVRVIR
					GAGAACGCGAGCCGCTCGACTTAGGGGCGA	CTGTCA		GAEPVANWICN
					CAGAGAGGATCACGCGTAAGCGGCTGCGCA	TAGCGA		GLTVLLPKSSDN
					ACGCCGTTCGCGATGTGGTTCCACCTCGGA	GCCCGA		ADPSNYRPITCL
					AACGCAGGGTTCCATCCACCCCTTCACGTA	AGGGGA		NTCYKLFTAVIA
					AGGAACAAGATTTGGTTCCAGAGGTGGACG	ATGGCC		QITASYVDVLG
					GGCCTGCACCTACGGATGTGTTGACACACC	ATGGAG		GLPRQQVALRK
					CACCCACAGAATCCGAACCAGAACCGATGC	ACTCCA		GVWGTSVSLMI
					TTGACCCACTCTCACTTGTCCAACTGGTAAG	GCCTCA		DALTVADARRA
					GCCGCAATTAGGCCGGGCAATGGGATGGGC	CCCTGT		KRPLGVCWFDF
					AGCCGAGGAGATGGAACTCGGGAATGTCGT	AACTCG		KKAFDSVPHNLI
					GATGGACGTGGAGCTGAAGAGGGAGTTCAA	AACCTA		RWILRVIGLPPVI
					CCGGGAGGTGAGGCGTGTCGGAAGAACGCC	AGTCCA		LSVIVSVMDQW
					TCCCGACCAAATGTATAAGCGTGGCGCAGG	GGCCCC		ATRLKIGGKVM
					TCCACCACTCCCACAGAAACGAGAGCCCGA	TTTCTG		PKTIPVRTGVFQ
					ACGAGTAGCCCTACTCGAGCAACTCATTGC	GTGTTG		GDTLSPLLFCLS

					GGCGCGCGTGGAGCGTGGCATCAACAGGGG	GCTCGC		VWPISFALDQFP
					ACTTGATTGGTTCCTTGAACTGAATGTCGCG	ACTGGT		QYQFRCANHLQ
					GTCTTTGCAGCCGCTAGGGTTCTGAGCCGCC	TAGGAA		QGFSVGHVFYM
					GGGAGAGAGTGGAAACTCTTGCTGATCGGC	CACGAC		DDLKCYCPDRE
					TGCACATCAACGATTCTGCCACTCTGTCCGA	AGTCTC		VLTAVIQQVQK
					AGTATCCAGACGGCGAGCTAAAGCCGAAAG	GTGTAA		SASALGLTIHYK
					GAAACTCCGGTGTGCMCGGGAACAACCATG	TCCCAC		KSAWLDQDGG
					GATGAGCAGGCGGATCAGGAWACTTGGTG	ACGCAC		KSGKAVLGVPX
					TGAGGGTCGAGAGACTCAAACAACTAGCTG	CGAAAG		LVGTYKYLGMH
					ACCTGGTCCGGCAAAGAATAGCTGGGAGAG	CCCTAG		ERFMIVSKDSLE
					GTAATAGATCGTCTTACGAAGGCCCTAGGC	CCCTCG		SVRGKFMGRLK
					GGAGATTCCGACTAAGACCCTCTTTACGTTC	TAGGCG		TLWTSKLTFGQ
					TGTTACTGAAGCACCTGTCAATCCACCCCTT	AAGGCT		AMLGTKSXCMP
					AACGGGAATGAAGTTTACACGTTCTGGCAC	GCGTTG		VVRYVLQNLFL
					AGCCTTTGGGCGCAGAGCCTCCGTGCTAAC	CTGGTT		PKSEFNQTRLVL
					ACGGATGACTGTCAACTGAGGGAATTCAAA	TCAGAA		REWDRQIRDLL
					AACCAGCTTTCTGCTGCGCGACACACGGAT	CTGTGA		DECNIRQVFRSK
					CTGACGAGCGTCGGAACCAGCTCATTGGTT	GCWAGT		TELYVSREEGG
					CAGATGTTCAGCGCGGCCCTTCGGAAAATG	GGGGTT		WGLPSMEDALE
					AAGAAGGGCAAAGCTCCCGGTCCTGATGGT	CGGATG		EEVVTKLAILVA
					ATCCGTGCCGCGTGGTGGGGAGTGTTCCGG	GCCGAG		RQETEPLFRVCE
					CGAATTGCACCGTATGTGGCTACGTGGGTG	TGTACC		ALERKRCPTPLS
					GTCCGAGTAATACGWGGGGCAGAACCAGT	AACCTG		LGLQILKDWGV

					AGCTAACTGGATCTGCAACGGGCTCACTGT	CTTGCT		GVELQGRTLLL
					CCTGCTGCCCAAGAGTAGTGACAATGCCGA	AACAGG		NGNTVGPSQAT
					CCCTAGTAACTACCGGCCAATCACATGCCT	TAGCAT		RKLTGELVLRRE
					AAATACMTGCTACAAACTCTTCACTGCCGT	AGAGTA		AERLSRWRSKV
					TATCGCACAAATAACCGCAAGTTACGTTGA	ATATGC		KPGCGMTGGA
					TGTTCTTGGTGGCCTACCTAGACAACAAGT	TAGTAA		WRDVPGIDVHL
					AGCACTGCGAAAAGGCGTTTGGGGCACCTC	GCAGGG		SNRWLVKGALS
					TGTGTCCCTGATGATCGATGCGCTAACGGT	CACGAG		PTVVSNSLAIRA
					GGCTGACGCGAGAAGAGCTAAACGCCCACT	AAAGGG		NTVILRGSGGGY
					TGGAGTCTGCTGGTTCGACTTCAAGAAAGC	CCGTAA		TKGTLLRCRGC
					GTTTGACTCAGTACCGCACAACTTAATTCGC	AGGCTC		GNTGETRRHIVS
					TGGATCCTGAGAGTCATCGGACTCCCCCCG	CGGCTA		ACSLGRQKGAA
					GTCATCTTGTCTGTCATTGTCTCTGTGATGG	TGAAGG		SRRHDNVCRILV
					ATCAATGGGCAACGAGGCTMAAGATAGGT	ACCTTG		RAICHKLNIEPP
					GGAAAGGTCATGCCCAAGACCATCCCGGTG	CGACCA		NSANFPHVVVL
					CGAACCGGCGTATTCCAAGGGGACACCCTA	CGCGTG		EGSGAKMWIDF
					TCGCCCCTGCTKTTCTGCCTATCKGTKTGGC	TGTCTC		PFVVPHKIRHTR
					CAATMTCGTTCGCCCTAGACCAATTTCCTCA	CCACGT		PDIVVLFEWNG
					GTATCAGTTTCGTTGTGCAAACCATCTTCAG	GCGGAT		VRRLSVIEVAVS
					CAGGGCTTTAGTGTCGGACACGTGTTTTACA	TCTTGA		DVANMQTQHIR
					TGGACGACCTGAAATGCTATTGCCCAGACA	AGCCAG		KSHRYGTNSTEP
					GGGAAGTGTTGACAGCAGTTATTCAGCAGG	AGTCTT		FVAGVTPTYRN
					TGCAGAAGAGCGCTTCTGCCCTCGGTCTGA	GCACTG		DCLAAQLRAKF

					CGATCCACTACAAGAAATCTGCGTGGCTCG ACCAGGACGGCGGGAAAAGTGGCAAAGCC GTCTTGGGTGTACCAGWACTTGTCGGCACT TACAAGTACCTGGGAATGCACGAGAGATTC ATGATTGTAAGCAAAGACAGTCTTGAGTCA GTCAGGGGAAAGTTCATGGGGCGGCTTAAA ACCCTGTGGACCTCGAAACTGACATTCGGG CAGGCAATGCTTGGCACAAAGAGCMTCTGC ATGCCAGTCGTGCGCTACGTGCTGCAGAAT CTATTCCTACCAAAATCGGAATTCAACCAG ACCCGTCTGGTCCTGAGAGAGTGGGATCGT CAGATCAGAGACCTACTGGATGAATGTAAT ATCCGGCAGGTTTTCCGAAGCAAGACGGAG TTATATGTCAGCCGCGAAGAGGGTGGCTGG GGTCTTCCTAGCATGGAAGACGCCCTCGAG GAAGAGGTAGTCACAAAACTTGCCATACTT GTTGCAAGACAAGAAACTGAGCCGTTGTTC CGGGTGTGTGAAGCTCTTGAAAGGAAAAGG TGCCCCACCCCCCTGTCCCTAGGATTGCAAA TACTGAAGGACTGGGGAGTTGGGGTAGAGT TACAGGGCCGCACATTGCTGCTGAACGGGA ACACAGTGGGCCCATCTCAGGCAACCAGAA AACTCACTGGAGAACTGGTCCTGAGGAGGG	CGCGCA GGATGG AGCCTG TGCAAC TCCTCC CTCGCT GATCGC AGGAGT GGAGGA TCACCA CTCTTTT TACCTT GCTAGC TTGGGG TACCAC CTTGAG CTGGGG CCGGCC TTGCTA GCTTGG GGTACG ACCCTT TGAGCT		KAQQVDVIPIIV GTTGETLDGEFG RIRKGLPMLTKL QMPRLWSEIQR AVILGSYRILVE HLALPKGGA (SEQ ID NO: 1498)

					AAGCGGAACGGCTCAGTAGATGGAGAAGT AAAGTGAAACCTGGCTGTGGCATGACTGGA GGGGCATGGCGGGATGTGCCCGGAATCGAT GTCCATCTTAGCAATAGGTGGCTCGTCAAA GGTGCACTGTCTCCGACCGTTGTCTCAAACA GCCTTGCCATACGAGCAAACACGGTGATTC TGAGAGGGAGTGGCGGGGGCTACACCAAA GGTACCCTCCTCCGGTGCAGAGGATGTGGT AACACGGGAGAAACACGCAGACACATTGTG TCAGCATGCTCCCTGGGAAGGCAGAAGGGG GCTGCCTCCCGCAGGCACGACAACGTCTGC AGGATATTGGTGCGAGCCATCTGCCACAAA CTCAACATTGAACCGCCCAACTCAGCCAAC TTCCCCCATGTCGTTGTGCTGGAAGGCAGTG GTGCTAAGATGTGGATAGACTTTCCCTTCGT GGTGCCACATAAGATCCGCCACACACGACC TGACATCGTCGTCTTGTTTGAGTGGAACGG GGTCAGACGTCTTAGTGTCATCGAAGTGGC TGTTAGCGATGTGGCAAATATGCAAACACA ACACATCCGCAAATCCCACAGATATGGCAC CAACTCGACTGAACCCTTCGTCGCGGGGGT CACGCCTACGTATAGGAATGACTGCCTGGC AGCGCAGTTGCGAGCGAAATTTAAAGCCCA	GGGGCT GGCCTT ACTAGC TTGGCG CGCCAC CCTTGG AGCTGG GGTGGC GCAAGA TCACTT GTATAC GGTCTA ACCAAT ACATTT GAAAAG CGATCA AAGCGA A (SEQ ID NO: 1253)		

					GCAAGTTGATGTGATCCCGATCATTGTCGG GACCACTGGTGAAACATTGGATGGCGAGTT TGGGCGGATCCGAAAGGGTTTGCCTATGCT CACTAAACTGCAGATGCCAAGGCTGTGGTC AGAAATCCAGCGGGCGGTGATCCTCGGATC ATATCGAATTTTGGTTGAACATCTTGCGCTA CCAAAGGGGGGCGCTTGAGGACTCANAATT GACAATACACCTCAGA (SEQ ID NO: 1130)			
R4	R4-1_PH	.		Parhyale hawaiensis	CCGACCGCCAGCGGGATAACTGGCAAACCC TGTCTCGACCACCGGCCCGTGAATCCATCG GGGCGTATGAGTCTGACACAGGGGGGTGTT TAAGGTGACCCTGTTGCGAGGAAATGCGCA GCAAAAGCCGGATGAGCCTTAGAACATCGA AGGCCAACGACAACCTGCCGAAAGACTGGC GACTAATCCAGTCAACCTCCTGTAGGTCAC CGGCTGGTCATGCTGAATCTCAGCTTAACC AGGCGTACGAACTTACAGTTGGAGGGTCGA GCACCCCTGATGGCTGAAAAGGACCATCAA AGTCGAAGGTAGCCAACGAGAAGACCAAC GGCTGATTCAGGCGGAAGAGTCAACTCGTT GAATGCGTTCGACAGCTTGGGGTAGATGGA ACTCCTAAGCCCTGAAAGGCAGTCCATCTT CGCAGACGCTAAGGTGCCCCGCCGCCTGAG	CCGACC GCCAGC GGGATA ACTGGC AAACCC TGTCTC GACCAC CGGCCC GTGAAT CCATCG GGGCGT ATGAGT CTGACA CAGGGG GGTGTT	TGAG CCTT AGGT CGCG GGAT GTGA CCCG GCGC CAGA GTGT AGAG CTGA ACAT CGCT CAAC	MKMSHNRDTPS NGVKGTSVRLG TSLVRSPVGEAG AVRERGTHPSES VSQDSDASVNA TGEGSVREQAPL SPPGAEEATVPT QRRTRHKWSRE DRVVLWECFVA SKREGPGYLKR LKQLWDERGIP GNFPQASLSGQI RQICSKNLLSEE ERLQIAARMEA QVASPSADEPAR

					GGTTATCAGGCCCCGCCGCCTGAGGGTTAC	TAAGGT	CGAT	QVPTRPVTPPRS
					CAGGCAACCATGAAAATGAGTCATAATAGA	GACCCT	CCAA	PPVEPARRPSIPS
					GATACCCCTTCTAACGGCGTGAAGGGTACC	GTTGCG	TTTG	EETPDLGAVPSE
					TCAGTTAGACTGGGTACCTCGCTCGTCCGGT	AGGAAA	GGTC	IDSADPNRSPSR
					CTCCTGTGGGTGAGGCAGGGGCCGTTCGTG	TGCGCA	GTGA	GPRHLPAHNMS
					AACGCGGTACTCACCCGTCGGAATCGGTTA	GCAAAA	AATC	QSESEDDVTDPD
					GCCAGGATTCCGACGCTAGTGTGAATGCTA	GCCGGA	CCCT	VSDQQRSDSLEP
					CTGGAGAGGGGAGTGTGAGGGAGCAGGCT	TGAGCC	CGAT	RDLLRNSSVEST
					CCCCTTTCTCCGCCGGGGGCGGAGGAGGCT	TTAGAA	AATA	PGHPNQELSDTL
					ACTGTCCCGACGCAACGCCGCACGCGCCAC	CATCGA	ATAA	LSNYVPSEIDSD
					AAGTGGTCCAGGGAAGACCGAGTTGTGCTG	AGGCCA	TAAT	DPNQSPRRGPRH
					TGGGAGTGCTTCGTCGCCAGCAAGCGCGAG	ACGACA	AA	LPAHDMSLSDS
					GGGCCAGGCTATCTCAAGCGCCTCAAGCAG	ACCTGC	(SEQ	MDEETEPDLSD
					CTGTGGGACGAGAGAGGAATCCCGGGCAAT	CGAAAG	ID NO:	QQRSDLLELRDL
					TTCCCACAGGCTTCTCTTAGTGGCCAAATCC	ACTGGC	1377)	LRNSSVETTPKG
					GGCAAATCTGCAGCAAAAACCTCCTCAGCG	GACTAA		HPSLRHLPEPKI
					AAGAGGAACGGCTTCAAATCGCTGCGCGGA	TCCAGT		RAAAFRVNSVIG
					TGGAAGCTCAAGTTGCCTCTCCATCCGCTGA	CAACCT		KIHTNNITETNA
					CGAACCTGCACGCCAGGTTCCAACCCGACC	CCTGTA		LIKAGADLAVRI
					TGTGACCCCGCCTAGGTCCCCACCTGTGGA	GGTCAC		LEVQPRPQRTQR
					ACCTGCTAGGCGGCCCTCCATTCCATCGGA	CGGCTG		KKDPPWKHRLE
					GGAAACCCCCGACTTAGGCGCTGTCCCGTC	GTCATG		KNIAEIRKHLSW
					TGAGATCGACTCTGCCGACCCCAACCGGTC	CTGAAT		ISEWRRGNLHD

					GCCCTCACGTGGGCCTAGGCACTTACCGGC	CTCAGC		EEKKTLLESRYR
					CCACAATATGAGCCAGAGCGAGTCAGAAGA	TTAACC		CLEVGLTNLEDT
					CGATGTGACAGATCCCGACGTGTCTGACCA	AGGCGT		LKQRLSAKRSK
					GCAGCGGAGTGACTCGCTGGAGCCACGGGA	ACGAAC		VRRFEARVAGF
					CTTACTCCGAAACTCCAGCGTAGAATCAAC	TTACAG		HQNQLFNTNQK
					GCCAGGTCACCCTAACCAGGAACTTAGTGA	TTGGAG		RLYQTLRGEETS
					CACCCTACTGTCGAACTACGTCCCATCTGAG	GGTCGA		SDSPNAEESIRF
					ATCGACTCGGACGACCCCAACCAGTCGCCG	GCACCC		WSDIWSKEVRH
					CGACGTGGGCCTCGGCACTTACCGGCCCAC	CTGATG		NNTAEWLHDVK
					GATATGAGCCTGAGCGACTCGATGGACGAA	GCTGAA		EKNVAADPDLTI
					GAGACAGAACCTGACCTATCCGACCAGCAG	AAGGAC		TSQQLKKQLSK
					CGGAGTGACTTGCTGGAGCTGCGGGACTTA	CATCAA		TKNWKAPGPD
					CTCCGGAACTCCAGCGTAGAAACGACGCCA	AGTCGA		MVQGYWIKTFT
					AAAGGTCATCCTAGCCTACGACATCTACCG	AGGTAG		SLHSRIAAQLNH
					GAACCGAAGATCCGAGCAGCTGCTTTCCGC	CCAACG		CLQRGTVPTWM
					GTCAACTCAGTGATTGGTAAAATCCACACC	AGAAGA		TTGKTVLIQKDK
					AACAACATCACTGAGACAAACGCCCTCATC	CCAACG		AKGTEVSNYRPI
					AAAGCTGGAGCAGACCTCGCAGTGAGGATC	GCTGAT		TCLPLMWKVLT
					CTTGAGGTACAACCCCGACCGCAACGGACC	TCAGGC		GIIYERVYQHLD
					CAGCGGAAAAAGGATCCTCCTTGGAAGCAC	GGAAGA		SKKLLPDEQKG
					CGCCTCGAGAAGAACATTGCTGAGATCAGG	GTCAAC		CRRNTRGTKDH
					AAACATCTCAGCTGGATCTCCGAATGGCGC	TCGTTG		LLVDKLLTKDA
					AGAGGAAATCTCCACGACGAGGAAAAGAA	AATGCG		RSKKKNLSMAW

					GACGCTGCTCGAATCGAGGTACCGGTGCCT	TTCGAC		VDYKKAFDMVP
					CGAAGTCGGCCTAACCAACTTAGAAGATAC	AGCTTG		HSWILECLDIYG
					CTTAAAGCAGCGGCTATCTGCTAAAAGGAG	GGGTAG		IAGNIRNLIATT
					CAAAGTTAGACGCTTCGAGGCCAGAGTAGC	ATGGAA		MPNWKTQLTSA
					TGGATTCCATCAGAACCAGCTATTCAACAC	CTCCTA		NKHLGEVSIKRG
					CAACCAGAAACGCCTCTATCAAACCCTTAG	AGCCCT		IFQGDSLSPLLFV
					AGGAGAAGAGACGAGCTCTGACTCACCCAA	GAAAGG		LTMIPLSETLNK
					CGCAGAAGAGAGCATCCGATTCTGGAGTGA	CAGTCC		AGQGYNYSRTM
					CATCTGGAGCAAAGAGGTACGGCACAACAA	ATCTTC		KLNHLLYMDDL
					TACCGCAGAATGGCTCCACGACGTCAAGGA	GCAGAC		KLYAKSKDQVE
					GAAGAACGTCGCAGCAGATCCAGACCTTAC	GCTAAG		QLLNIVHQYSQ
					AATCACCAGTCAACAGCTCAAAAAGCAACT	GTGCCC		DIKMQFGVSKC
					GAGCAAGACCAAGAACTGGAAAGCACCCG	CGCCGC		GVLNIERGEVTA
					GACCAGATATGGTGCAAGGCTACTGGATCA	CTGAGG		SEGITIEEGTIKDI
					AAACTTTCACATCGCTGCACTCCAGAATTGC	GTTATC		EEAGYKYLGVM
					TGCACAGCTCAACCACTGCCTGCAACGCGG	AGGCCC		EYNTILHRTMK
					AACCGTGCCCACCTGGATGACCACAGGGAA	CGCCGC		DSIRKEYLTRLR
					GACGGTCCTAATCCAGAAGGACAAGGCAAA	CTGAGG		LILKSHLNGGNT
					GGGCACGGAGGTCAGCAATTACCGCCCAAT	GTTACC		IKAINTWAVPVV
					CACGTGCCTCCCCCTTATGTGGAAGGTGCTG	AGGCAA		RYSAGIINWTKK
					ACGGGCATAATCTACGAGAGAGTCTACCAA	CC (SEQ		DCTDMDIKTRK
					CATCTAGACTCAAAGAAGCTCCTACCCGAC	ID NO:		LMTIYRALHPRS
					GAACAGAAAGGATGCCGCAGGAACACCAG	1254)		CVDRLYINRREG

					GGGAACGAAGGACCATTTGCTCGTCGACAA			GRGLISVEDCVE
					GCTCCTCACGAAAGATGCCAGGAGCAAAAA			AEKRALSQHFR
					GAAAAACCTAAGCATGGCTTGGGTTGATTA			ESDDPWARCLV
					CAAAAAAGCCTTCGACATGGTCCCTCACTC			EAKLLKETETA
					CTGGATCCTGGAATGCCTGGACATCTACGG			DQFKERRRLDR
					CATTGCCGGCAACATCCGCAATCTGATCGC			TNKWKSMKMS
					GACGACCATGCCCAACTGGAAGACCCAGCT			GQYLEAVQDKI
					CACAAGCGCCAATAAACACCTCGGCGAGGT			VPDSWNWLLRG
					GAGCATCAAGCGGGGCATCTTCCAGGGCGA			ELKRETEGTILA
					CTCCCTATCTCCACTGCTCTTTGTCCTGACC			AQEQALRTRYIQ
					ATGATCCCCCTCTCCGAGACCCTGAACAAG			NKIDKRNVPSTC
					GCCGGCCAGGGCTACAACTACAGTCGTACC			RICRSSDETINH
					ATGAAGCTCAACCACCTGCTTTACATGGAC			VISECGVLAQKE
					GATCTGAAGCTCTACGCCAAATCAAAGGAC			YKRRHDKVARH
					CAAGTAGAGCAGCTCCTTAATATAGTCCAT			LHWTLLRIHNFP
					CAGTACTCCCAAGACATCAAGATGCAGTTC			VSERWYEHEPA
					GGTGTTAGCAAATGCGGAGTCCTTAACATC			PVVENEAVQIY
					GAAAGGGGAGAGGTTACTGCGTCCGAGGG			WDKRMETDRV
					AATCACCATCGAAGAAGGTACCATCAAGGA			LHANRPDIVVK
					TATCGAGGAGGCAGGTTACAAGTACCTGGG			DKQEKSAKLIDI
					TGTCATGGAGTATAACACCATTCTCCACAG			SIPFDSRIVDKEA
					AACGATGAAAGATTCCATTCGGAAGGAGTA			EKKEKYRDLAIE
					CCTCACCCGACTCCGGCTCATCCTGAAATCA			LQRLWQMKVD

					CACCTCAATGGGGGCAACACCATCAAGGCC ATCAACACCTGGGCCGTCCCCGTAGTCCGA TATTCTGCTGGCATTATTAACTGGACCAAGA AAGACTGCACCGACATGGACATCAAAACCA GGAAGCTCATGACCATCTATCGAGCACTGC ACCCAAGATCGTGCGTAGACAGACTCTACA TTAACAGAAGAGAGGGGGGCCGTGGCCTCA TCAGCGTCGAGGACTGCGTCGAAGCGGAAA AACGCGCTCTCTCCCAACACTTCAGAGAGA GCGACGACCCGTGGGCACGCTGCTTAGTCG AGGCGAAGTTGCTGAAAGAGACCGAAACG GCTGACCAATTCAAAGAGCGACGGCGCCTC GATCGGACTAACAAATGGAAGAGCATGAA GATGTCCGGCCAGTACCTCGAAGCTGTTCA GGACAAAATCGTACCAGACTCTTGGAACTG GCTTCTGAGGGGTGAGCTGAAACGGGAAAC AGAAGGCACCATCCTGGCAGCTCAAGAGCA AGCCCTCAGAACCAGGTACATCCAAAACAA GATCGACAAGAGAAACGTCCCTAGTACCTG TAGAATTTGCAGGAGCTCGGATGAAACCAT CAACCACGTGATAAGCGAATGCGGAGTGCT GGCGCAAAAGGAGTACAAGCGCAGGCACG ATAAAGTTGCCCGACACCTTCACTGGACCC			VVPVVIGALGA MSKNLKTALRE LKCGHLHPGTL QKSALLGTAHII RKVL (SEQ ID NO: 1499)

					TCCTGCGGATCCATAACTTCCCGGTCAGTGA GAGGTGGTACGAGCACGAACCCGCGCCCGT GGTTGAAAACGAAGCCGTGCAGATCTACTG GGACAAGAGGATGGAAACTGATCGAGTACT TCACGCCAACCGACCGGACATCGTCGTCAA AGACAAACAGGAGAAGTCGGCCAAACTGA TCGACATTTCAATCCCCTTCGACTCCAGAAT CGTGGACAAGGAAGCGGAGAAGAAGGAAA AGTACCGTGACCTTGCCATCGAACTCCAGA GACTGTGGCAGATGAAGGTGGACGTCGTTC CAGTGGTCATCGGAGCTCTGGGGGCAATGT CCAAGAATCTCAAGACTGCCCTGCGAGAGC TGAAGTGCGGCCATCTCCACCCAGGAACCC TCCAGAAGAGCGCTCTGCTTGGAACAGCGC ACATCATTCGCAAAGTGCTCTGAGCCTTAG GTCGCGGGATGTGACCCGGCGCCAGAGTGT AGAGCTGAACATCGCTCAACCGATCCAATT TGGGTCGTGAAATCCCCTCGATAATAATAA TAATAA (SEQ ID NO: 1131)			
R4	R4-2_BX	CADV01009048		Bursaphelenchus xylophilus	GAGGATCCTGGGTTCCTACTACCCTGCTCCA TCTCCTCGCGATGGATCCTTGGGGAAGTCTC CGGACTGAGCTAAGAGAGCGTTAAAGTAGA GGGTGGCGGCGTAGTGACTTCCAAGTTGCG	GAGGAT CCTGGG TTCCTA CTACCC	TAAC AAGT GTAA TAAA	MVFNNCKPKHL CPAIRPTGQQET NGGSEGTAEPT AGPSRPAVSEDA

					ATGGGGCGGAACATCTACTCTTCTGAGAGA	TGCTCC	AACC	AQPVPLFEEGEY
					GGGAAAGCCCTATGGCGGCGGTAGAAAGGT	ATCTCC	ACCC	IRAHRDKTCPYC
					TGGGCTACGGCAACACTTGCCATGATCAGA	TCGCGA	ATGC	EVLWIGARSSK
					TTCGATCAAAATTAGCCTCTGGGGCTGGCA	TGGATC	GTAG	ARSDSWPLCQIL
					ACCCTACAACGGATTGTAAACTGAACTATG	CTTGGG	TAAA	YLMKRNDDLRG
					CTATGCAAAATGAAAATAAAAAATGGGGGC	GAAGTC	CCGA	QRTRYPLLESSL
					TTTACAATCTAAGGTGTTGGCAGATCACGA	TCCGGA	TCAA	RAAGIARTKFAII
					AAACTGCCCGTCGATGAGGGTGAGATAACA	CTGAGC	TTATC	KCIRNVLRDRY
					TCCTACGGAACAGCCCCTGCTGGACCAAAC	TAAGAG	TAGC	VPNGPYSEHWK
					CAAATCATCCACAAATTGGAGGGTTTTCTTG	AGCGTT	AAAA	IYRANSGEVPQG
					GTAGTTTCCCTTGGCACGTCTTGTTTCATAA	AAAGTA	TCGC	ATITKGKRSARV
					GCCAGAATAAAAACGATACCATACAGACAT	GAGGGT	AGGT	AGLPSPSQSGHH
					GAGGCTGGGTACCTGCCAGCCGCGACACGG	GGCGGC	CAGA	TKRIQAGTGIET
					AAAACCGGTAGGTGCAACCGGAGACAGCTA	GTAGTG	AGAC	ETTVTETNTTPE
					GGGGAAAGAAAATAGTAAAGTGTCGAAAA	ACTTCC	CAAA	VSHEHRDPCGEP
					CAAAACAGGTAGCCCCTGGCTAGAGGGAAT	AAGTTG	GAAC	ETSAANVDKVT
					GGGACATTGTCCGATTAGGTTGCCTTGACCC	CGATGG	CGAC	ELTEDGSETRGT
					AATGAAAGCCAACCGGGTTTACATTATCGG	GGCGGA	CCAG	ANVANGGVSVS
					TCTACCAAGGGCAATGACCAGAGAAACTGC	ACATCT	AGGA	DPGRKRQSSSQ
					GACTATGCCGTAACCCTTCGCAGATTGCCG	ACTCTT	ATAG	NRGNIETTNPEL
					ATGAGAACCATCGATCGTAAGTCGAAGCCA	CTGAGA	GACC	VGMWEDMFGV
					AGCGGATTGACCAGTGGAGGGTTATCCCGA	GAGGGA	AGAG	QLDGAMRTTER
					CAACAGCATGGTATTCAATAATTGTAAACC	AAGCCC	CTGA	PRLPKLKHLSEP

					TAAACACCTGTGCCCTGCGATCCGTCCGAC	TATGGC	AACT	ERLWIRAKLEQ
					AGGGCAACAGGAAACGAACGGTGGTAGCG	GGCGGT	CTCA	AWLQCVSYDVE
					AGGGCACTGCTGAGCCTACCGCTGGTCCTT	AGAAAG	GATA	QQWLNANAVL
					CGAGACCGGCAGTGAGCGAGGATGCAGCCC	GTTGGG	CGCC	YAAIRSVAASRP
					AACCGGTGCCACTGTTCGAAGAGGGCGAAT	CTACGG	AACG	CKEAREAQKTW
					ATATAAGGGCTCACCGCGATAAGACTTGTC	CAACAC	GTCC	LDNKKKDEAKL
					CTTATTGCGAAGTCCTGTGGATAGGAGCGC	TTGCCA	TAAT	RRLIGRISSVHS
					GGTCTTCAAAGGCCCGCTCCGACAGCTGGC	TGATCA	AAAA	MPKGDRTPREK
					CATTGTGCCAAATCTTGTACTTAATGAAGCG	GATTCG	CGTC	KLVKNITKLKNT
					CAACGACGACCTCAGAGGTCAAAGGACGA	ATCAAA	GTTA	HYPDMDWGGL
					GGTATCCACTGCTCGAGAGTTCTTTGAGAG	ATTAGC	AGTA	LNHFKVKLSQL
					CAGCCGGAATTGCCCGTACTAAGTTTGCGA	CTCTGG	AAGC	KEKISVRVAEHK
					TTATCAAGTGTATTCGGAACGTGTTGAGAG	GGCTGG	ATCG	RKVNRNAAGQY
					ACCGGTATGTGCCCAACGGCCCATACTCTG	CAACCC	TTAA	GKSVAGSAGLA
					AGCATTGGAAAATTTACCGCGCGAATAGCG	TACAAC	GTAC	PDVVSATAYWS
					GGGAAGTTCCTCAGGGAGCGACCATAACAA	GGATTG	AAAA	GLAQPGPKKFK
					AAGGGAAGAGGTCCGCCAGAGTTGCGGGG	TAAACT	CAAA	ASSPIFQTWKDD
					CTTCCCTCCCCAAGCCAGAGCGGTCACCAC	GAACTA	AGCA	VAKNLNTEPVL
					ACGAAACGAATACAAGCTGGAACTGGAATT	TGCTAT	CTGT	LYPIIKECIRKPS
					GAAACGGAAACTACGGTGACTGAGACTAAC	GCAAAA	AAAC	PWKAPGPDGIY
					ACAACACCCGAGGTTTCGCATGAACACAGA	TGAAAA	GCGA	NYYWQQEFVA
					GATCCCTGCGGGGAACCAGAGACGAGTGCT	TAAAAA	GGCC	QWIQTLVKRTL
					GCTAATGTAGATAAGGTAACGGAATTGACG	ATGGGG	CCCT	DIGRFPTALMCG

					GAAGATGGGTCTGAAACCAGAGGAACTGCG	GCTTTA	CTTTG	RTVLLFKSGDKS
					AACGTCGCTAACGGTGGGGTTTCTGTTAGC	CAATCT	CCAA	MPQNYRPITCLN
					GATCCTGGTCGTAAAAGGCAGAGTTCAAGC	AAGGTG	ACTC	GCFKITNAVLTK
					CAAAATAGAGGAAATATCGAAACGACTAAT	TTGGCA	CGGT	VILQRVQDTCAL
					CCAGAATTAGTCGGTATGTGGGAAGACATG	GATCAC	AATC	PREQMALKPKV
					TTTGGAGTGCAGCTTGATGGAGCGATGCGC	GAAAAC	CTCG	WSCMEAQLRDQ
					ACAACAGAGCGACCAAGGCTACCCAAACTG	TGCCCG	TAGT	ALQSEIGDDCKT
					AAGCACCTGTCCGAACCCGAGAGGTTATGG	TCGATG	ACGG	AWIDFSKAYDS
					ATCAGGGCGAAGTTGGAACAGGCTTGGCTG	AGGGTG	TGCTT	LDHDALRFVIQT
					CAGTGCGTTAGCTATGACGTCGAACAGCAA	AGATAA	TTTCC	IALPAGMEEYLL
					TGGCTTAACGCCAACGCGGTTCTATACGCC	CATCCT	CGCT	KSLDSWRTQLV
					GCCATCAGGTCAGTCGCGGCTAGCAGACCT	ACGGAA	TCAC	LSDAGKVVSGK
					TGTAAAGAGGCCCGCGAAGCTCAAAAAACA	CAGCCC	GAAT	PYPIKRGVLQGD
					TGGCTGGATAATAAAAAGAAGGATGAGGC	CTGCTG	CAGA	SLSPALFVLTTSP
					GAAATTGAGGCGGCTGATCGGGAGAATCAG	GACCAA	ACGC	IVAHLQRTCPTG
					CTCTGTCCACAGTATGCCGAAAGGCGACCG	ACCAAA	TGCC	RIQLYMDDIKLY
					TACGCCACGAGAAAAGAAACTCGTGAAGA	TCATCC	AGAT	GKTESDLCMLIK
					ATATAACTAAGTTGAAAAACACCCACTATC	ACAAAT	CTTGT	ETQRVANKLGL
					CAGACATGGATTGGGGTGGACTTCTGAACC	TGGAGG	CCGA	NINLKKCALFGK
					ATTTTAAGGTAAAGTTGTCGCAACTTAAAG	GTTTTC	CATG	SIKQSIAGFDPLG
					AAAAGATTAGCGTGCGGGTGGCGGAGCATA	TTGGTA	GGCC	DRTYKYLGIPQR
					AGCGGAAAGTTAATCGTAACGCAGCAGGAC	GTTTCC	GTAG	DVADIKQAYDE
					AGTACGGCAAAAGTGTGGCCGGCAGCGCAG	CTTGGC	GGTT	LKAKTVQTIGET

					GGTTAGCGCCGGATGTCGTTTCCGCGACGG	ACGTCT	TGGG	MACDYLTTRQV
					CATACTGGTCAGGTCTTGCACAACCTGGGC	TGTTTC	AGTA	INRLNSKIPPVV
					CAAAGAAATTTAAAGCAAGCTCGCCGATCT	ATAAGC	CCAG	RFVTQSALCSAP
					TCCAGACTTGGAAAGACGACGTAGCGAAGA	CAGAAT	CGTG	MTRGLYNKITEL
					ACTTGAACACTGAGCCGGTACTGCTGTATC	AAAAAC	GGGC	DNVSRAELRKV
					CGATCATAAAAGAATGCATCAGAAAGCCAA	GATACC	GGAG	LIYKATNVSRFY
					GCCCCTGGAAAGCACCAGGCCCTGACGGGA	ATACAG	AGCG	LATKEGGFGYA
					TATACAACTATTACTGGCAGCAAGAGTTTG	ACATGA	TACC	SLQQVFVEAVV
					TGGCTCAGTGGATCCAAACCCTTGTGAAGA	GGCTGG	TGGG	SRAIYCLRAPSL
					GAACCCTGGATATAGGGCGGTTCCCGACTG	GTACCT	TACA	CDIREFILSKFDP
					CACTTATGTGTGGGAGGACGGTACTACTCTT	GCCAGC	CTGC	VKVARIALARS
					CAAAAGCGGTGACAAATCTATGCCTCAGAA	CGCGAC	ATAA	KIDMDIERMDM
					CTATCGACCCATTACATGCCTGAATGGATG	ACGGAA	TCGG	ASATRTIRQHYQ
					CTTTAAGATTACCAATGCGGTATTGACGAA	AACCGG	GTCT	AKWKTLFQQSK
					AGTAATCCTTCAACGGGTTCAAGACACTTG	TAGGTG	CAGA	LYQKWVQHKID
					CGCTCTTCCCAGAGAACAAATGGCTTTGAA	CAACCG	AAAC	IPNSSRWLQRGE
					ACCAAAGGTGTGGTCGTGCATGGAAGCGCA	GAGACA	CACC	ISPRNCRIAVAV
					GTTGCGCGACCAGGCACTTCAGTCTGAGAT	GCTAGG	TATG	QDNTLLCRGFV
					CGGTGACGATTGTAAGACTGCATGGATCGA	GGAAAG	GTTT	GSKDPNKQCRL
					TTTTTCCAAAGCATACGATAGTCTGGACCAC	AAAATA	ATTA	CNAGIETASHIV
					GACGCCTTAAGATTTGTAATCCAAACGATT	GTAAAG	TTCTG	TECSTHRVHMYI
					GCACTGCCGGCAGGGATGGAAGAGTACCTT	TGTCGA	TCTCC	ERHDSVARNIY
					CTGAAATCCCTGGATAGTTGGAGAACGCAA	AAACAA	CATC	AVLAKNCGFWI

					CTTGTGCTATCTGATGCAGGTAAGGTTGTCT	AACAGG	TGCA	PHYSQKIPTVKI
					CCGGAAAGCCCTACCCAATTAAAAGGGGAG	TAGCCC	GGGT	TKSYELYWNYK
					TGCTCCAAGGAGACAGTCTGAGTCCCGCGC	CTGGCT	AGCT	FPCTQALEACRP
					TATTCGTCTTAACGACGAGCCCAATTGTCGC	AGAGGG	TTTCG	DIVLIDRAKKRI
					TCACCTTCAGAGAACTTGTCCCACGGGTCGT	AATGGG	TTGG	LVVEVAVSYVT
					ATACAGTTGTATATGGATGATATTAAACTCT	ACATTG	GCCA	RLEQMTQRKLY
					ATGGCAAGACAGAGTCGGACCTGTGTATGT	TCCGAT	TAGG	KYGVNGEYQAD
					TAATAAAAGAAACTCAACGAGTCGCAAACA	TAGGTT	AGCC	GETRGWNICRE
					AATTGGGCTTAAACATTAATTTGAAAAAAT	GCCTTG	TAGG	LVQKYNMRIDL
					GTGCCCTCTTTGGTAAAAGTATAAAACAGA	ACCCAA	GGCA	CIVVIGACGEILP
					GTATTGCCGGGTTTGACCCGCTCGGCGACA	TGAAAG	AGAG	CMVKEIEKISKV
					GAACATACAAATATCTCGGCATTCCCCAAA	CCAACC	TGCA	SGRQLLERCQRS
					GGGATGTTGCAGATATCAAACAGGCATATG	GGGTTT	TGTA	AVLGTVRTVRR
					ACGAACTCAAAGCTAAAACGGTGCAGACCA	ACATTA	GTCTT	HLAN (SEQ ID
					TCGGCGAAACGATGGCGTGTGACTATCTAA	TCGGTC	CAAC	NO: 1500)
					CAACAAGGCAAGTTATCAACAGATTAAATT	TACCAA	GGCC	
					CGAAGATACCGCCTGTAGTCAGATTTGTCA	GGGCAA	ATGC	
					CACAATCTGCGCTGTGCAGCGCGCCTATGA	TGACCA	CAGG	
					CCAGAGGATTATATAATAAAATCACGGAAT	GAGAAA	GAAA	
					TGGATAACGTGAGTAGAGCAGAGTTAAGGA	CTGCGA	CTTGT	
					AAGTTCTTATCTATAAAGCGACCAATGTCTC	CTATGC	GAGA	
					TCGGTTCTACCTGGCCACGAAAGAAGGGGG	CGTAAC	GGTG	
					ATTCGGCTATGCAAGCCTGCAGCAGGTATT	CCTTCG	AGGG	

					TGTTGAAGCTGTTGTGAGTAGAGCGATCTA	CAGATT	ATAA	
					CTGCCTGAGAGCTCCCAGTCTGTGTGACAT	GCCGAT	CTAG	
					ACGGGAGTTCATATTGAGTAAGTTCGACCC	GAGAAC	CATC	
					TGTCAAGGTTGCTCGCATTGCGTTGGCAAG	CATCGA	AGAT	
					GAGTAAAATTGACATGGACATCGAACGGAT	TCGTAA	AATA	
					GGACATGGCCTCGGCCACGCGTACCATCAG	GTCGAA	TCAG	
					ACAACATTATCAGGCAAAGTGGAAAACTCT	GCCAAG	TCAT	
					GTTCCAACAATCGAAACTTTACCAAAAGTG	CGGATT	GAAA	
					GGTGCAGCATAAGATCGATATACCCAACTC	GACCAG	TTAG	
					GAGTAGGTGGTTGCAGAGGGGGGAAATCTC	TGGAGG	TAAC	
					GCCGCGAAACTGTCGAATCGCGGTAGCTGT	GTTATC	AACC	
					TCAAGATAACACTTTACTGTGCAGAGGATT	CCGACA	AACG	
					TGTGGGGAGTAAGGACCCAAACAAACAATG	ACAGC	TTCA	
					TCGGCTATGCAATGCTGGCATTGAGACCGC	(SEQ ID	CCGT	
					TAGTCACATCGTGACTGAATGTAGTACACA	NO: 1255)	CGTT	
					CCGGGTTCACATGTATATCGAACGTCACGA		GGCA	
					CTCTGTGGCGCGAAATATATATGCTGTCCTG		AAAC	
					GCAAAAAATTGCGGTTTTTGGATTCCTCATT		ACCG	
					ACTCACAGAAAATCCCAACAGTAAAGATTA		ACTA	
					CGAAGTCATATGAGTTGTATTGGAATTACA		ACGA	
					AGTTCCCGTGCACGCAAGCACTGGAGGCTT		TGCT	
					GTCGTCCCGATATCGTTCTCATTGACAGAGC		AGTT	
					GAAGAAAAGAATCCTCGTGGTGGAAGTGGC		AGAA	

					GGTCTCGTATGTCACACGGCTAGAGCAGAT GACGCAGAGGAAACTCTATAAATACGGGGT GAACGGCGAGTACCAGGCTGACGGGGAAA CCCGTGGCTGGAACATTTGTCGTGAACTTGT ACAAAAGTACAATATGAGAATTGATCTCTG CATAGTCGTAATTGGGGCATGTGGAGAAAT ACTCCCTTGCATGGTTAAAGAAATTGAAAA AATATCCAAGGTCTCTGGTCGGCAGCTCCT AGAACGGTGCCAGCGTAGCGCTGTACTTGG AACTGTGCGGACCGTAAGGAGACACTTGGC AAACTAACAAGTGTAATAAAAACCACCCAT GCGTAGTAAACCGATCAATTATCTAGCAAA ATCGCAGGTCAGAAGACCAAAGAACCGACC CAGAGGAATAGGACCAGAGCTGAAACTCTC AGATACGCCAACGGTCCTAATAAAACGTCG TTAAGTAAAGCATCGTTAAGTACAAAACAA AAGCACTGTAAACGCGAGGCCCCCTCTTTG CCAAACTCCGGTAATCCTCGTAGTACGGTG CTTTTTCCCGCTTCACGAATCAGAACGCTGC CAGATCTTGTCCGACATGGGCCGTAGGGTT TGGGAGTACCAGCGTGGGGCGGAGAGCGTA CCTGGGTACACTGCATAATCGGGTCTCAGA AAACCACCTATGGTTTATTATTCTGTCTCCC		AGAG TCGG GTCTT CCCA AAGT TAGG TGCTT GCAC CGAA GCCG ATCC GCTC TACC CACA GCTC TGCC CAGC GTT (SEQ ID NO: 1378)	

					ATCTGCAGGGTAGCTTTTCGTTGGGCCATAG GAGCCTAGGGGCAAGAGTGCATGTAGTCTT CAACGGCCATGCCAGGGAAACTTGTGAGAG GTGAGGGATAACTAGCATCAGATAATATCA GTCATGAAATTAGTAACAACCAACGTTCAC CGTCGTTGGCAAAACACCGACTAACGATGC TAGTTAGAAAGAGTCGGGTCTTCCCAAAGT TAGGTGCTTGCACCGAAGCCGATCCGCTCT ACCCACAGCTCTGCCCAGCGTT (SEQ ID NO: 1132)			
R4	R4-2_HG	ABLA01000389		Heterodera glycines	TGTGGCGATACTCGGAACCTCCGGGGAGCC TGGTAGGAGTTGGCCTACAGGTCGCGAAAG TCCCTAGGTGCTGCACGGGTTGCGCTAATCC GAGCCCCTTCCGGTGTTGGCTAGCACTGGA GGGGGGGGGGTCCAAAGAATACGTGAAGC GGACCCTGTGTTTTCTCTGACCTCATACGGG TATATCCCGATGATATCTATACAGTGTTCAT TTTCATTTCTTAACCTGCTTTTTCCTCAAGA GAAATATGACCACATTCGTTTGCCATTTCGA AGGTGCGACTTTTGAGGGAGTTCTCAGGGA CATCAAGCTGTTCATGGACGGCTTGACGGC TCGTGGCGGCACGCAGCCGCGCGGGCCAGT TAGGGCTGGTGCCAATGAGGCAGCGAAGAA	TGTGGC GATACT CGGAAC CTCCGG GGAGCC TGGTAG GAGTTG GCCTAC AGGTCG CGAAAG TCCCTA GGTGCT GCACGG	TAGT CGTA GCCC AGAT GTCG ACAG CCCC TACA CCGG TAGA TGGA AGGA GGTA	MRKSFLQHIPEL SSHIAMSVPARN YPKMCSLAQGS SGTLSHNGKGV AMHRCPSDDCA GKDPPQRGSQK GNLRSVRWTPS EEKAVFEYWSR LEQHAMLNGSE ARGTCAITRSQF LIHWDGERESRS LSDGVPEYPMR TERAYYERVRL

					GTCTCTGAAACGCCAAAAGCAGCGGGAGAA	GTTGCG	TGTA	LRQRGWQWDC
					GAGAGAGAAAGCGGGGACGGGTTAGAGAT	CTAATC	TCTA	ANECLVIGQCAE
					AATAGCATGATTGACGGCGAAGCAATACTT	CGAGCC	AGCC	PCRKPNVVAIKA
					TGACTTCCAGGTACGGAGGACGGCGGCAAA	CCTTCC	CACG	DKGMKRSLVKG
					AAGCAAAAAGGGGACGAAACCGATGGCAT	GGTGTT	GCAA	KLLSLPHVMGEI
					TGCCAAAGGGGCCGAAACCATTGAGCTGGA	GGCTAG	GCCA	NQVSVQVEVPL
					CGCTGATGGGCCGTCTGCCGCGCAAAAGGA	CACTGG	CCAG	PSVPASVPQVEG
					GGGGGGAGCCCCTCCGTCTGGGAAGAAGGT	AGGGGG	TGGA	VESKGFTETEPS
					GCGGCGCCATTCGTCCAACAAACCGGAATA	GGGGGT	AACG	NKPSLEGNPAEE
					GATCTTGATGATCTCAGCAATTGATCATATT	CCAAAG	GTGA	GLRKPERVNVP
					GATCTCTTTTTCCAGTTGTTGCGATAAATTA	AATACG	CCAC	VHGIISDSERKD
					TCGTGCATTATTTCTTCGATTTCTCCAAAGC	TGAAGC	CTCT	LKDRFWSAYKT
					TTAACTCCTTTCCTAATACCTACTCATGTAT	GGACCC	GTTC	AKRSVGFRPAL
					ACGTTAGTAGGCATGTTTTATGCAGGTAAA	TGTGTT	GCGG	KIEPNRVNRAQ
					GATAGACCTGGTGCCCCCGGGCGACTTGGG	TTCTCT	AAAT	WEVLDSCVVEV
					ATGTGGATGATGGTTGGCGGAGAGTTCTGA	GACCTC	GCCC	LKKRETSNGYR
					TGACGCCGTAGTTCCGGAGGAACAGTTCCT	ATACGG	CGTA	GCVLRHLNVAV
					ACTAGTGCCAGCCTGGGCTAGTGGAGCTGG	GTATAT	GAGT	YAAGYVLAEGN
					TCTGGCGAGTGCTTTGCTCGTCAGATAAGG	CCCGAT	ATAT	KERRQVIRRQSA
					GGTGGTTGGGGTGGTACCTCTGGAATGATG	GATATC	GAAG	EWLLRQKSEINN
					TGTCGGATAAGCACTCTGCCTTATAAAGCT	TATACA	GTCG	IRRHIGWITDEL
					GTCGTTCTGCCCAGGTTTCTCCTAACCAGGT	GTGTTC	GAGC	TRRRTGKNPTSR
					TGAACTTGTAATGAGCCTTTGGGTATCTGGT	ATTTTC	ATTG	QLSNFAWLQRR

					CGGGGTCTGCGCGGAGATAGCTCGTGAGGA	ATTTCT	AAGC	YQVLGKPVRET
					GCTGTTATTACATTATTGGTTGAACGCTTTT	TAACCT	ACAA	RDLEVQRERLV
					GGGTCGTACTGCACAGATAAATATGAGAAA	GCTTTT	GTGA	SRLRLAQDRINS
					AAGTTTTTTACAACATATACCTGAGCTTTCA	TCCTCA	GAAC	SMDREERVRKR
					AGTCATATTGCTATGTCTGTGCCTGCGAGAA	AGAGAA	CCTG	MLPLRRKLEEPL
					ATTATCCCAAAATGTGCTCTTTAGCACAGG	ATATGA	GAAG	GDSKLDTKQAR
					GTTCTTCCGGCACATTAAGTCATAATGGGA	CCACAT	TATG	TFWASLIGERKE
					AGGGTGTCGCAATGCACCGGTGCCCCAGCG	TCGTTT	GTGG	FGKIPELENWAE
					ATGATTGTGCTGGTAAAGATCCCCCTCAGC	GCCATT	TGAG	EVRSKVTDGQG
					GGGGGTCTCAAAAAGGGAATTTGAGGAGTG	TCGAAG	CTGG	FASDHVDQTVW
					TGCGATGGACTCCCAGCGAGGAGAAAGCTG	GTGCGA	TAAC	KKILGKARPLKA
					TCTTCGAGTATTGGTCACGGTTGGAACAGC	CTTTTG	CTCG	PGPDGIPNLLWK
					ATGCCATGTTAAATGGTAGCGAAGCCCGGG	AGGGAG	AATT	RLPSANQALFK
					GGACCTGTGCCATAACTAGAAGTCAGTTTC	TTCTCA	CCTC	WLMGIKRKQLS
					TCATTCATTGGGATGGCGAGAGAGAGTCAA	GGGACA	CTAT	VPSWLTKGRVV
					GATCCTTGTCGGATGGTGTGCCCGAGTACC	TCAAGC	GGGT	LLPKGGDPVDP
					CAATGAGGACGGAAAGGGCGTACTATGAA	TGTTCA	GCTT	ANYRPIACLNTQ
					AGGGTCCGACTCTTGCGACAACGCGGATGG	TGGACG	GCGC	YKLVTGMVTA
					CAATGGGATTGTGCTAATGAATGCTTAGTA	GCTTGA	CCGT	WVSEHLTTYSIL
					ATTGGCCAATGTGCAGAGCCATGCCGTAAG	CGGCTC	AGGT	PIEQRAMVSGT
					CCAAACGTGGTAGCGATTAAGGCTGATAAG	GTGGCG	CATTT	WGCTHAMVIDR
					GGTATGAAGAGATCGCTGGTTAAGGGAAAG	GCACGC	GTGT	AITSYAEATGLP
					CTTTTGTCCCTGCCCCATGTAATGGGCGAAA	AGCCGC	ATGT	LYVGFVDFAKA

					TAAACCAAGTGAGCGTGCAGGTTGAGGTGC	GCGGGC	AGGG	FDSVSQPWIRYA
					CTTTGCCAAGTGTACCCGCCAGCGTGCCCC	CAGTTA	ATGG	LKVAGVHKRIR
					AGGTGGAAGGAGTTGAAAGCAAAGGTTTTA	GGGCTG	AAGG	CLIGILMKCWSV
					CTGAAACGGAACCGTCGAATAAACCTTCCC	GTGCCA	AATC	RYEVFKSGRVL
					TGGAAGGGAATCCGGCGGAAGAAGGATTA	ATGAGG	TCGA	RSAPLAVKNGV
					AGGAAACCGGAGCGAGTAAACGTACCTGTT	CAGCGA	TAGC	LQGDTLSPLLFC
					CATGGGATAATCAGTGATAGTGAACGGAAA	AGAAGT	GCAA	LSVAVVSSAVG
					GACCTGAAAGACCGGTTTTGGAGCGCCTAT	CTCTGA	ATCG	SLFDFEVTIPGR
					AAAACGGCCAAGAGGTCTGTGGGTTTTAGA	AACGCC	GGAT	GVMQQQNHLFY
					CCGGCTCTAAAAATCGAACCGAATAGGGTT	AAAAGC	ACCG	MDDFKGFAPSE
					AACCGTGCCCAGTGGGAAGTACTGGATTCC	AGCGGG	CACT	ASLTRMLVTLE
					TGCGTTGTTGAGGTGCTCAAAAAGAGGGAA	AGAAGA	GGCG	RTASALGLKINK
					ACCTCCAATGGGTACCGCGGTTGCGTCCTCC	GAGAGA	ATCA	RKCALVHPRER
					GGCATTTGAACGTTGCAGTTTATGCGGCGG	AAGCGG	TGCT	ENEETGSDIPVL
					GATACGTGCTGGCGGAAGGCAACAAGGAA	GGACGG	GCCA	GLRDTYKYLGIE
					AGGAGGCAGGTGATCAGACGCCAGTCGGCG	GTTAGA	GTGG	ERFGIVFEDAW
					GAATGGTTGCTTCGCCAAAAGTCCGAAATC	GATAAT	GCCC	DRVRTKMFERM
					AACAACATCCGTAGACACATCGGGTGGATT	AGCATG	TGCT	RTLLCTEHTFGE
					ACCGATGAGTTAACACGGAGGAGGACGGG	ATTGAC	GAGT	LRAAFASTIAPV
					AAAGAACCCAACATCCAGACAGTTGAGTAA	GGCGAA	TGGG	ARYLFLNVIVGG
					TTTTGCATGGTTGCAAAGGCGATACCAGGT	GCAATA	CGGG	PSWSETLTKAK
					TTTGGGTAAACCCGTGAGGGAAACCCGGGA	CTTTGA	TTAG	DMDLRIRRLLW
					CTTGGAAGTCCAACGAGAAAGACTTGTAAG	CTTCCA	CAAC	ERRDNEPGWRF

					TCGTCTTCGATTGGCTCAGGACCGAATTAAC	GGTACG	CCGT	KHCSADRLYLR
					TCTTCCATGGATCGTGAAGAGCGAGTTCGT	GAGGAC	CCAG	VQYGGLGFVSV
					AAACGAATGTTACCCCTGCGGAGAAAGTTG	GGCGGC	GGAC	EDTLSESIIYCW
					GAAGAACCTTTGGGTGACTCTAAACTTGAC	AAAAAG	GCAG	AYVQCRPELEL
					ACAAAGCAGGCACGAACATTTTGGGCCAGC	CAAAAA	ATGC	ARELFGTLNRSA
					CTGATTGGAGAGCGAAAGGAATTTGGAAAA	GGGGAC	CAAA	RSGIKQSIAKGA
					ATCCCGGAGCTTGAAAACTGGGCCGAAGAA	GAAACC	AGCC	RKVFRSYALLSK
					GTGAGGAGCAAAGTCACCGATGGACAAGGT	GATGGC	TGGT	NSAQRVSDLDG
					TTCGCAAGTGACCATGTGGATCAAACTGTTT	ATTGCC	GACA	DASPGFRVGEMI
					GGAAAAAGATACTGGGTAAGGCCAGGCCA	AAAGGG	CTCT	FMEPTRGARAIV
					CTGAAGGCGCCCGGACCAGATGGAATCCCC	GCCGAA	AAAG	KILRKENDSRRL
					AATTTACTTTGGAAACGGTTACCATCCGCA	ACCATT	GGAG	AAWKGRPMGG
					AACCAGGCGTTATTTAAGTGGTTAATGGGG	GAGCTG	TCTA	RVVSLPELDQV
					ATCAAGCGGAAACAGCTCAGTGTCCCCAGT	GACGCT	CGGC	HSYHWLIRARIG
					TGGCTCACAAAAGGAAGGGTGGTCCTCCTA	GATGGG	GACG	RRSFRDCIAAQE
					CCCAAAGGAGGTGACCCGGTAGACCCGGCC	CCGTCT	AGAA	GQLKARELMCP
					AACTACCGACCAATCGCTTGTTTGAACACG	GCCGCG	CATC	HINAKAKWCRR
					CAGTACAAACTAGTGACTGGTATGGTGACG	CAAAAG	CCCT	CGDGRVETEQHI
					GCATGGGTATCGGAACACCTGACCACATAC	GAGGGG	AAGT	LSGCAWSRTGT
					TCTATTTTGCCGATTGAACAACGAGCGATG	GGAGCC	C (SEQ	MLDRHNGVVR
					GTGTCTGGCACTTGGGGGTGTACCCATGCT	CCTCCG	ID NO:	QVHTALCRKYG
					ATGGTGATTGACAGAGCAATTACGTCTTAT	TCTGGG	1379)	LPVSSHVVPLHA
					GCAGAGGCCACTGGACTACCGCTTTATGTA	AAGAAG		VIENEHAKILYD

					GGCTTTGTGGACTTTGCGAAGGCTTTCGACT	GTGCGG		VALHTSPAGVLP
					CAGTTAGTCAACCGTGGATCAGGTACGCTC	CGCCAT		REDGSTSYTGLR
					TAAAGGTCGCTGGAGTTCACAAGCGAATCC	TCGTCC		STRPDMVIFDKK
					GATGTCTTATTGGAATCCTTATGAAATGTTG	AACAAA		ARTILIVEISVPW
					GAGCGTAAGGTATGAAGTGTTCAAGAGCGG	CCGGAA		RENLVKQELIK
					GAGAGTGTTAAGGAGTGCGCCTTTGGCAGT	TAGATC		WRKYAINSMIEP
					TAAGAACGGAGTGTTACAAGGAGACACGTT	TTGATG		LELAEAEIPGPN
					GAGCCCGCTGCTCTTTTGTCTCTCAGTAGCC	ATCTCA		LKHALGLAYGT
					GTTGTTTCTTCTGCAGTAGGCAGCCTCTTTG	GCAATT		SFPTVKVVPIVV
					ATTTTGAGGTCACCATACCTGGACGAGGAG	GATCAT		GSCGEVLPNITK
					TGATGCAGCAGCAGAACCACTTGTTCTATA	ATTGAT		RLSELGIPKRGIP
					TGGATGACTTTAAAGGGTTTGCTCCTAGTGA	CTCTTTT		SLLESIQRAAIIG
					AGCGTCCTTGACTAGGATGTTAGTAACATT	TCCAGT		SGHVIRAHLSVP
					GGAGCGAACCGCTTCAGCCCTCGGGTTGAA	TGTTGC		RSESET (SEQ ID
					AATCAATAAGAGGAAATGCGCTCTGGTACA	GATAAA		NO: 1501)
					CCCCCGTGAAAGAGAGAACGAGGAAACTG	TTATCG		
					GAAGTGATATCCCGGTCCTTGGTCTACGAG	TGCATT		
					ACACTTATAAATACCTTGGAATCGAAGAGA	ATTTCT		
					GATTTGGTATCGTTTTCGAGGATGCGTGGG	TCGATT		
					ACCGAGTCCGTACCAAGATGTTTGAGAGAA	TCTCCA		
					TGCGCACCTTGCTATGCACGGAACATACCTT	AAGCTT		
					CGGGGAGTTGCGTGCGGCCTTCGCTAGTAC	AACTCC		
					GATCGCCCCAGTAGCGAGGTACCTCTTTCTC	TTTCCT		

					AATGTAATCGTTGGTGGGCCTAGTTGGAGT	AATACC		
					GAAACCCTGACGAAAGCTAAGGACATGGAC	TACTCA		
					CTGCGTATCCGACGCCTCCTGTGGGAAAGA	TGTATA		
					AGAGATAACGAGCCGGGATGGCGGTTTAAG	CGTTAG		
					CATTGTAGTGCTGACCGCTTGTACCTCCGCG	TAGGCA		
					TACAGTACGGTGGACTTGGCTTCGTGTCCGT	TGTTTT		
					AGAGGACACTTTGTCTGAGTCTATCATTTAC	ATGCAG		
					TGTTGGGCTTACGTGCAATGTAGGCCCGAG	GTAAAG		
					TTGGAGCTGGCTCGTGAACTTTTTGGAACCT	ATAGAC		
					TAAACAGGAGTGCTAGAAGCGGTATAAAAC	CTGGTG		
					AGAGCATTGCGAAAGGAGCGCGTAAAGTGT	CCCCCG		
					TCCGAAGTTATGCCCTTCTCTCGAAAAATTC	GGCGAC		
					CGCCCAGCGAGTATCTGACCTTGATGGGGA	TTGGGA		
					TGCCTCCCCAGGATTCCGAGTAGGGGAAAT	TGTGGA		
					GATATTCATGGAACCAACTCGGGGTGCCAG	TGATGG		
					AGCAATCGTAAAAATCCTCCGTAAGGAGAA	TTGGCG		
					TGATTCGCGACGCTTGGCTGCCTGGAAAGG	GAGAGT		
					TAGACCTATGGGAGGTCGGGTCGTGTCATT	TCTGAT		
					GCCAGAGCTCGACCAGGTTCATAGCTACCA	GACGCC		
					CTGGTTGATCCGTGCGAGGATTGGCAGGCG	GTAGTT		
					TTCGTTCCGAGACTGCATTGCAGCACAGGA	CCGGAG		
					GGGACAGTTGAAAGCGCGGGAGCTTATGTG	GAACAG		
					TCCACATATAAATGCGAAAGCAAAGTGGTG	TTCCTA		

					CCGACGATGTGGGGACGGTCGTGTTGAAAC	CTAGTG		
					CGAGCAGCATATCCTTAGTGGTTGTGCTTGG	CCAGCC		
					TCACGGACTGGAACTATGTTAGATCGACAT	TGGGCT		
					AACGGTGTTGTACGGCAAGTACACACGGCT	AGTGGA		
					TTATGCAGGAAATATGGACTGCCCGTTTCTT	GCTGGT		
					CTCATGTCGTTCCTTTGCATGCGGTAATCGA	CTGGCG		
					GAATGAACATGCCAAAATCCTGTACGATGT	AGTGCT		
					CGCGTTGCATACTTCGCCTGCTGGCGTCCTC	TTGCTC		
					CCTAGGGAGGATGGGAGTACGAGTTATACG	GTCAGA		
					GGTCTGCGTAGCACCAGGCCCGATATGGTA	TAAGGG		
					ATATTCGATAAAAAAGCACGTACGATCCTT	GTGGTT		
					ATTGTGGAAATTTCGGTGCCATGGAGAGAG	GGGGTG		
					AACCTGGTTAAACAGGAGCTCATAAAGTGG	GTACCT		
					CGGAAATATGCCATAAATTCTATGATCGAA	CTGGAA		
					CCATTGGAACTAGCGGAAGCTGAAATACCG	TGATGT		
					GGACCTAACCTGAAACATGCGTTAGGCCTT	GTCGGA		
					GCGTACGGTACGAGTTTCCCAACAGTGAAG	TAAGCA		
					GTGGTCCCGATAGTAGTTGGGTCGTGCGGC	CTCTGC		
					GAGGTATTACCAAATATAACGAAGAGGCTG	CTTATA		
					AGTGAACTTGGGATACCTAAACGAGGAATC	AAGCTG		
					CCGAGTTTGCTTGAGTCGATACAGCGTGCT	TCGTTC		
					GCGATAATTGGTTCTGGACATGTTATCCGG	TGCCCA		
					GCGCACTTGAGTGTACCTAGGAGCGAGTCC	GGTTTC		

					GAGACTTAGTCGTAGCCCAGATGTCGACAG CCCCTACACCGGTAGATGGAAGGAGGTATG TATCTAAGCCCACGGCAAGCCACCAGTGGA AACGGTGACCACCTCTGTTCGCGGAAATGC CCCGTAGAGTATATGAAGGTCGGAGCATTG AAGCACAAGTGAGAACCCTGGAAGTATGGT GGTGAGCTGGTAACCTCGAATTCCTCCTATG GGTGCTTGCGCCCGTAGGTCATTTGTGTATG TAGGGATGGAAGGAATCTCGATAGCGCAAA TCGGGATACCGCACTGGCGATCATGCTGCC AGTGGGCCCTGCTGAGTTGGGCGGGTTAGC AACCCGTCCAGGGACGCAGATGCCAAAAGC CTGGTGACACTCTAAAGGGAGTCTACGGCG ACGAGAACATCCCCTAAGTC (SEQ ID NO: 1133)	TCCTAA CCAGGT TGAACT TGTAAT GAGCCT TTGGGT ATCTGG TCGGGG TCTGCG CGGAGA TAGCTC GTGAGG AGCTGT TATTAC ATTATT GGTTGA ACGCTT TTGGGT CGTACT GCACAG ATAAAT (SEQ ID NO: 1256)		

R4	R4-2_SRa	CACX01002006		Strongyloides ratti	CGTCAGAAGAGCAGGTGTTTTTCAAAGCAA	CGTCAG	TAAT	MQKFSVPKDSS
					AGACTTATTCTACGAAGGGGAAAGATGATC	AAGAGC	TAGA	QIFLVDSILNKHI
					AAACATGCAGATTTGGCTGCAATGAAATAG	AGGTGT	TTTAT	CSTKNKVKDVI
					AATCAAACTACCATGTGGTGACATCTTGCA	TTTTCA	ATGT	KRRSIIKTLICAA
					AATACCATTCATACACAAGAAGACACGATA	AAGCAA	GCCA	GLTLRKLVCGK
					TGGTTGTGTACAATATATTAAAAAGACTAA	AGACTT	CTTTA	LGNNKYNSKIN
					AAGAAAAGTATAAGTTAGATGGTGATTTAC	ATTCTA	CCTC	QLWKKERKIINC
					AATATGGTAGATCAGTATTTAATGGGAAGA	CGAAGG	AAAG	IEDLKHLIETNK
					AGGAAAAAATTAGGATCAGATCTGGCCATA	GGAAAG	ATAT	RRHNFGKRLRN
					AGTTTTTGACTGAACAACCGTTAGTAAACA	ATGATC	ATTA	AKVSPSEMLKD
					ATAAACCGGATATTGTTATAACGATGAAAG	AAACAT	ATAC	YYNKLRYIKNEI
					AAGGGAAAAAAAAGATAACGTACATACTA	GCAGAT	TAAC	TACLDEHKKAIL
					GAGATGTCGATACCACATATCTAGAACTTG	TTGGCT	TTAA	RTKFKLTPSIKII
					AAGATGCAAGAAAGAATAAAGTATAAAAA	GCAATG	TTAT	SNIQNHNDEEAE
					ATATTGTGTGAGCTCCATGGTAAAAATAAC	AAATAG	AATT	LPKEEEFVKYY
					GAATGATAATGTTGATTCGATAGCTCGTGA	AATCAA	TATT	KELFTNKDGDD
					TTTTAATCTACTAAATTTAATGGAAAGAAA	ACTACC	ATGA	KETPHLDNWLK
					AGAAAGGTGTAAGATAAAGTTTGGATCATT	ATGTGG	ATTA	KFSKTLIVDWTI
					TGTATTTGGCTGTTATGGAGAATATGTCTTG	TGACAT	TAAT	NDKEILEALKYC
					ACTGAATCTGCTCTAAGAACACAGAAAATC	CTTGCA	ATAA	GNFKAPGSDMV
					CTGATTGAATTGGGATTCTCCAAAAAAGAG	AATACC	TAAG	MKVCYKWFKS
					ATGGATAGCTTGATCAAAGAGTGTTCTTAC	ATTCAT	TTTTA	AQNYLIRWIKST
					AGTTAAATGAATGAGACAGCAAGAATTATT	ACACAA	AAAT	WYGEYTINKKD

					ATGAAACATCTAGAAGGCGAAAAGAATCA	GAAGAC	AAAA	TNAVTFMIWKR
					AGATGAAATCAGCCATTAATAACTTTTATTT	ACGATA	(SEQ	DGKPKNDVKSY
					ACCCAATTATTTATTGCGTTAACTTGATTAT	TGGTTG	ID NO:	RPISCLNCDFKL
					ATGATATTTATTGTATATTTATTGTTTAATTT	TGTACA	1380)	LNKLIANKIYESI
					AGTTTCATAGTACAAATAAAAGAATATACT	ATATAT		EKILPINQMAVI
					ACGATTCTTTACTTTTAAGTTCTACGAGAAC	TAAAAA		KNKHGTCEALL
					GTTGTTTTAGAATATTTTAATAATATTTCTA	GACTAA		LYKSLVQSMKF
					CTACAATTAAGTAATTAGAAGAAATCAACG	AAGAAA		RRTKDVKEIWC
					AAAGCAGCTAAACTTACTCGCAAAATTCGT	AGTATA		SWIDFSKCYDSI
					TGATCGAGGCTGGAATGGCACCAACTAATA	AGTTAG		SHKCLKKMIQSI
					TACTAACCAACAATAGAAAAAAAAAGAAG	ATGGTG		KAPPIIHKLILDG
					AACTTCAGTGGACTTCAAATAAAATAAAAT	ATTTAC		IDSWNISICNGK
					TATTGGTAGAATTGTATGATAAAACTGAGA	AATATG		NISKTKIPVKSGI
					AAATTTGACTAAAACAAAAAGATTAGAGCA	GTAGAT		LQGKVASSLYF
					AATATGTCATCATTTTCCTAATCATACAATT	CAGTAT		VLLTGEISYALN
					AAAGCGATGATGACGAAATTGAGAGAATTT	TTAATG		KEEQVPIETITPS
					GAAAGAGAAAAAAGGGAAAATTGTGAAAT	GGAAGA		NTLKINHISFIDD
					TAAAATGAGGGAAAATGAGGAAAAACCTG	AGGAAA		YQLYATSQKKV
					AGGAAAAATTAAATTTTGACAACTATGAAG	AAATTA		EKLTIKLREIAEE
					AAGCAAAATTGAAAAGAGGAATTAAGTGTA	GGATCA		MNLKLNPQKCG
					AAAAAGAAGTAAAACCAATTGTTATTGAAA	GATCTG		IYGTDDLGKRL
					ACAAGGATTAGGAATTTTTGAAAACGGAAA	GCCATA		MLKESSLNFPYT
					AATTATTTCTCAAGTTTGTCAACACAATCAA	AGTTTT		SEYKYLGLVEN

					TTACCAAAGAAAGAAGGGTAAATCAACCAG	TGACTG		SLDLKDINIQLF
					AATGCAAAAATTTAGTGTTCCAAAGGACTC	AACAAC		KDKILSKYSTIFE
					TTCTCAAATATTCTTAGTGGATAGTATCCTA	CGTTAG		SRLTTHQKRKV
					AACAAACATATTTGCTCAACCAAAAACAAA	TAAACA		FNSTISPCAAYY
					GTTAAGGATGTCATTAAGAGGAGAAGTATT	ATAAAC		LGNLITNKCSIQ
					ATTAAGACATTAATATGTGCAGCTGGTTTA	CGGATA		ELLNECKKFDQ
					ACTCTTAGAAAGTTAGTTTGTGGGAAACTT	TTGTTA		MVRNQLVNQNI
					GGAAATAATAAATATAATTCCAAGATAAAT	TAACGA		KKLQVSNSRIYL
					CAACTTTGGAAAAAAGAAAGGAAAATTATT	TGAAAG		PKEYNSLGLNEI
					AATTGTATAGAAGATTTAAAGCATTTGATA	AAGGGA		EIEVAANIIRKA
					GAAACGAATAAACGTAGGCACAATTTTGGG	AAAAAA		CYIKKRETLRGV
					AAAAGATTAAGGAACGCCAAAGTGAGCCCT	AGATAA		DKLYIAMSKNG
					TCTGAGATGTTGAAGGATTACTACAACAAG	CGTACA		HRNTLSDALYIT
					TTAAGATATATTAAAAATGAAATTACGGCA	TACTAG		KKYSNFQINWNI
					TGCTTGGATGAACACAAAAAAGCAATTTTA	AGATGT		MGMVKDQNNIL
					AGAACAAAATTTAAACTAACACCAAGCATT	CGATAC		LDAKKIIENIKE
					AAGATCATCTCAAATATTCAAAATCATAAT	CACATA		KRRNLWLEHW
					GATGAAGAGGCTGAACTCCCAAAGGAAGA	TCTAGA		KKGNMTYANE
					AGAATTCGTCAAATACTATAAAGAACTATT	ACTTGA		AIKKEFHLPDLN
					TACAAACAAAGATGGGGATGATAAAGAAA	AGATGC		IDSKYLMLCYA
					CACCACATCTTGACAATTGGTTGAAGAAGT	AAGAAA		GSEEQIIYNGHV
					TTAGTAAGACATTAATTGTAGACTGGACTA	GAATAA		SLVNQSSPSSRL
					TAAATGACAAGGAGATTCTTGAAGCTTTAA	AGTATA		CRKCNKLEETS

					AATATTGTGGTAATTTCAAAGCACCGGGAT	AAAAAT		YHVASVCEFHK
					CTGACATGGTAATGAAAGTTTGTTACAAAT	ATTGTG		KNLHLMRHNSA
					GGTTCAAATCTGCACAAAATTATCTAATCA	TGAGCT		VYHIITELCRIM
					GATGGATCAAATCAACTTGGTATGGTGAAT	CCATGG		KVKCTLRYPEA
					ACACAATAAATAAGAAAGATACAAATGCA	TAAAAA		SGIIKSGNMKIA
					GTCACATTCATGATTTGGAAAAGAGATGGT	TAACGA		AGVKYTFGTAK
					AAACCAAAAAATGATGTCAAATCATATAGA	ATGATA		IYHNKPDLVWY
					CCGATTTCTTGCTTGAACTGTGATTTCAAGT	ATGTTG		TPEVIYVIEVSIS
					TACTAAACAAACTTATTGCAAACAAGATTT	ATTCGA		SLKNAKSQMKM
					ATGAATCAATAGAGAAAATTTTGCCTATAA	TAGCTC		KTARYAVNSTK
					ATCAAATGGCAGTAATTAAGAATAAACATG	GTGATT		KLENFAALNNL
					GAACATGTGAAGCCCTACTCCTATACAAAT	TTAATC		KKGENFVEILSH
					CATTGGTTCAATCAATGAAATTTAGAAGAA	TACTAA		KANFKRVHFMP
					CGAAAGATGTGAAAGAAATATGGTGTTCAT	ATTTAA		LVFCTFGEIPKE
					GGATTGACTTCAGCAAATGTTATGACTCAA	TGGAAA		TMKYLEKLNFS
					TAAGTCATAAGTGTTTGAAAAAAATGATAC	GAAAAG		NEKIKTIASPIAR
					AAAGTATTAAGGCTCCACCAATTATTCACA	AAAGGT		YTGRTLKAHFT
					AACTCATTTTAGATGGAATTGATAGCTGGA	GTAAGA		N (SEQ ID NO:
					ACATATCAATTTGTAATGGGAAAAATATTT	TAAAGT		1502)
					CCAAAACTAAAATACCTGTTAAATCTGGTA	TTGGAT		
					TTCTTCAAGGTAAAGTGGCTAGCTCTTTATA	CATTTG		
					TTTTGTATTATTGACTGGTGAAATATCTTAT	TATTTG		
					GCTTTGAACAAGGAAGAACAAGTACCAATT	GCTGTT		

					GAAACGATTACACCAAGCAACACTTTGAAA	ATGGAG		
					ATCAACCATATCAGTTTTATAGATGATTATC	AATATG		
					AACTGTATGCTACATCACAAAAGAAAGTTG	TCTTGA		
					AAAAACTAACTATTAAATTGAGGGAAATAG	CTGAAT		
					CAGAAGAGATGAACCTGAAGCTGAATCCAC	CTGCTC		
					AAAAATGTGGCATATATGGCACAGATGATC	TAAGAA		
					TGGGAAAGAGACTAATGCTGAAAGAAAGTT	CACAGA		
					CATTAAACTTTCCATATACTTCCGAGTACAA	AAATCC		
					GTATCTTGGTCTGGTAGAAAATTCATTGGAC	TGATTG		
					TTAAAGGATATCAATATCCAACTGTTTAAA	AATTGG		
					GATAAAATACTATCTAAATACTCAACTATTT	GATTCT		
					TTGAATCAAGACTCACTACACACCAGAAAA	CCAAAA		
					GAAAAGTATTTAATAGTACAATATCTCCAT	AAGAGA		
					GCGCAGCATATTACCTAGGTAATTTAATAA	TGGATA		
					CCAATAAATGTTCAATTCAAGAACTATTGA	GCTTGA		
					ATGAATGTAAGAAATTCGATCAAATGGTCA	TCAAAG		
					GAAATCAGTTAGTGAACCAAAACATCAAAA	AGTGTT		
					AGCTTCAAGTATCAAATAGTAGAATCTACT	CTTACA		
					TACCTAAAGAATACAACTCCCTTGGATTAA	GTTAAA		
					ACGAAATAGAAATTGAAGTAGCAGCCAATA	TGAATG		
					TAATCAGGAAAGCTTGTTACATTAAAAAAA	AGACAG		
					GAGAAACTCTTAGAGGAGTAGACAAACTAT	CAAGAA		
					ATATTGCAATGTCAAAAAATGGTCACAGAA	TTATTA		

					ACACTCTTAGTGATGCTTTATATATAACTAA	TGAAAC		
					AAAATACTCTAACTTTCAAATAAATTGGAA	ATCTAG		
					TATTATGGGTATGGTGAAAGATCAAAACAA	AAGGCG		
					TATTCTTCTGGATGCGAAGAAAATTATTGA	AAAAGA		
					AAATATCAAAGAGAAAAGGAGGAACCTAT	ATCAAG		
					GGCTGGAACATTGGAAAAAAGGAAACATG	ATGAAA		
					ACATATGCAAATGAAGCAATCAAAAAAGA	TCAGCC		
					ATTCCATCTCCCGGACCTAAACATTGACTCC	ATTAAT		
					AAATACCTCATGTTGTGCTATGCTGGTTCAG	AACTTT		
					AGGAACAAATCATTTATAATGGACACGTAA	TATTTA		
					GTTTGGTTAATCAATCCTCACCGTCTAGTAG	CCCAAT		
					ATTATGTAGAAAGTGTAACAAATTGGAAGA	TATTTA		
					AACTTCATACCATGTGGCTAGTGTTTGTGAA	TTGCGT		
					TTCCATAAAAAGAATCTACATCTAATGCGT	TAACTT		
					CACAATAGTGCAGTTTACCACATAATAACT	GATTAT		
					GAACTCTGTAGAATCATGAAAGTCAAATGT	ATGATA		
					ACGCTAAGATATCCAGAAGCATCAGGAATA	TTTATT		
					ATAAAAAGTGGAAACATGAAGATTGCTGCT	GTATAT		
					GGAGTAAAATACACTTTTGGAACAGCAAAA	TTATTG		
					ATATACCATAATAAGCCTGACTTGGTATGG	TTTAAT		
					TATACTCCAGAGGTAATTTATGTTATAGAA	TTAGTT		
					GTAAGTATTTCATCACTGAAGAACGCCAAA	TCATAG		
					TCTCAAATGAAAATGAAGACTGCTAGATAT	TACAAA		

					GCAGTAAACTCAACAAAGAAATTGGAAAAT TTTGCTGCTCTGAATAATTTAAAGAAGGGA GAAAACTTTGTTGAAATATTAAGTCATAAA GCAAATTTCAAAAGAGTCCATTTCATGCCC CTAGTTTTCTGCACTTTTGGTGAAATACCTA AAGAAACAATGAAATACCTGGAAAAACTCA ACTTTTCCAATGAAAAGATTAAAACAATTG CTTCACCAATCGCACGTTACACTGGAAGAA CTCTCAAAGCACATTTTACTAATTAATTAGA TTTATATGTGCCACTTTACCTCAAAGATATA TTAATACTAACTTAATTATAATTTATTATGA ATTATAATATAATAAGTTTTAAAATAAAA (SEQ ID NO: 1134)	TAAAAG AATATA CTACGA TTCTTT ACTTTT AAGTTC TACGAG AACGTT GTTTTA GAATAT TTTAAT AATATT TCTACT ACAATT AAGTAA TTAGAA GAAATC AACGAA AGCAGC TAAACT TACTCG CAAAAT TCGTTG		

						ATCGAG GCTGGA ATGGCA CCAACT AATATA CTAACC AACAAT AGAAAA AAAAAG AAGAAC TTCAGT GGACTT CAAATA AAATAA AATTAT TGGTAG AATTGT ATGATA AAACTG AGAAAA TTTGAC TAAAAC AAAAAG		

						ATTAGA GCAAAT ATGTCA TCATTT TCCTAA TCATAC AATTAA AGCGAT GATGAC GAAATT GAGAGA ATTTGA AAGAGA AAAAAG GGAAAA TTGTGA AATTAA AATGAG GGAAAA TGAGGA AAAACC TGAGGA AAAATT		

						AAATTT TGACAA CTATGA AGAAGC AAAATT GAAAAG AGGAAT TAAGTG TAAAAA AGAAGT AAAACC AATTGT TATTGA AAACAA GGATTA GGAATT TTTGAA AACGGA AAAATT ATTTCT CAAGTT TGTCAA CACAAT		

						CAATTA CCAAAG AAAGAA GGGTAA ATCAAC CAGA (SEQ ID NO: 1257)		
R4	R4-3_BX	CADV01008832		Bursaphelenchus xylophilus	GGGTCCTCGGTTCTTACTACCGTGCTCCACC TCCTCGCGATGGACCCTGGGGTAGGCCTCC GGGCTGAGCTAAGCAGAGCATTAAAGTAAA GAGTGACGGCGCAGTTGCTTCCAAGTTGCG GTGGGGCGGAACATCTACTCTTCTGAGAGA GGGGAAGCCCTATGGCGGCGTTAGAAAGGT TGGACTGCGGCAACACTAGCCATGATCAGA TTCGATCAAAATAGCCTCTGGGGCTGGCGA CCCTATAACGGATTGTAAACTGAACTATGC TAACCTGTCAGTAAAGACAGAACGGGGGCT TTGCAATCTAAGGTGTTGGCAGACCACTAA AACTGCCCTTTGATGAGGGTGAGATAACAT CCTACGGAACAGCCCCTGCTGATCCAAACA AAATCATCCACAAATCGGAGGGTTTTCTTG GTAGTTTCCCTTGGCACGTCTTGTTTCAAAA	GGGTCC TCGGTT CTTACT ACCGTG CTCCAC CTCCTC GCGATG GACCCT GGGGTA GGCCTC CGGGCT AGCTA AGCAGA GCATTA AAGTAA	TGAG ACCA CCCA TGCG CAGA GTAT CCGA ATCA GTGA AAGT CCAA GTTTC AGGA CAGA AACG	MISRSQADRPVE GQPVTAMSFHN LEPNNLYPENLR PTGSQDANRGV ADIAEEVTGPSG LVTNEEAARAPP LFVEGEYKRAH CGGGKCHYCRV LWIGARSSKART DSWNLCEILFLI NKCMELGNVRR IYSPLESSLKEA GINRTRHAIVKC RLAVMRDRFVD NAPYSEHWRLY

					GCCAGAATATAAAACAATACCGAAGTACAG	AGAGTG	TCAG	NACAENRAVVV
					AATGGCTGGAACCTAACAGCCATGGCACAG	ACGGCG	ATAA	PMDSATTVKKR
					AAAACCGGTGGGTGCACACCGGAGACAGC	CAGTTG	GTCC	TARQAGLESPSQ
					AGGGGAAAGATTGCAAAGTGCCGTAAATAA	CTTCCA	AAGA	IGVAGKRVHEA
					AGATGGTAGCTCTCTGACCTGAGGGAATGG	AGTTGC	GAAA	ETGTDRINAVIE
					GACACTGTCCGATTAATGCCTTATCCCGAA	GGTGGG	CGAG	TNTTPLEDIDLSP
					GACGGTCAAGGTTCTATCTTATCGGTCTCCA	GCGGAA	AAAA	ETPEGLAELPST
					AAGGGCTCTGGCCATAGAAGCTGCGAGGAA	CATCTA	CAAG	VEIMELTEDGSR
					GCCGTAACCCTACGCAGATTGCCGCCGTGC	CTCTTC	TTCA	SRGTANDADGG
					ATGATAAGTCGTAGCCAAGCGGATCGACCA	TGAGAG	AGTA	VSISDPLRNRPSS
					GTGGAGGGCCAACCCGTAACAGCCATGAGT	AGGGGA	TGCA	SQESRNVPEQV
					TTCCATAATTTAGAACCAAACAACCTGTAC	AGCCCT	AGAG	DPDGELVWESL
					CCAGAAAATCTTCGTCCTACTGGGTCACAG	ATGGCG	TTAA	YGAQLRGAMRT
					GACGCGAATCGCGGAGTTGCGGACATTGCT	GCGTTA	TCAA	TDRPRLPKLTKF
					GAGGAGGTGACTGGTCCTTCGGGGCTGGTT	GAAAGG	TAAG	SAAEQLWIKSK
					ACTAACGAGGAGGCAGCGCGAGCACCGCC	TTGGAC	AGAG	VEKARLECVSY
					GCTTTTCGTAGAGGGCGAATATAAAAGGGC	TGCGGC	TACC	GIEQQWLRASA
					ACACTGCGGTGGGGGCAAATGTCACTACTG	AACACT	GTAA	VLYATIKTVAA
					CAGGGTCCTGTGGATAGGTGCGAGGTCTTC	AGCCAT	ATGT	CRPYNKAREAH
					AAAGGCCCGTACCGACAGTTGGAATCTCTG	GATCAG	ATGA	KVWLENKRAEE
					TGAGATTCTATTTTTGATTAACAAATGCATG	ATTCGA	CCCC	KRVRRIIGRIETV
					GAACTTGGAAACGTGCGTCGGATTTACAGT	TCAAAA	CCCC	RTMPKGKRTDK
					CCTCTCGAGAGTTCTTTGAAAGAGGCTGGA	TAGCCT	TTTGC	QIRLARKINRLK

					ATCAACCGAACACGTCATGCAATAGTTAAG	CTGGGG	CAAG	RVSFPEMDWHG
					TGTCGACTAGCTGTCATGCGTGATCGATTTG	CTGGCG	TCGA	FLNHFKAKLDL
					TGGATAACGCCCCATATTCTGAGCATTGGC	ACCCTA	CAAC	LKKLISVRVAEH
					GTCTGTATAACGCGTGTGCAGAAAATAGGG	TAACGG	TGTC	ERKISRKIAGTY
					CGGTGGTCGTCCCGATGGACTCGGCCACAA	ATTGTA	ATGC	GKSVSGQSGFTP
					CGGTGAAGAAAAGAACTGCTAGACAAGCG	AACTGA	AGGT	DVVAATTFWSG
					GGCTTGGAATCACCTTCGCAAATCGGGGTC	ACTATG	GTCT	LAQPGPKKFKK
					GCTGGCAAGCGAGTCCACGAAGCTGAAACT	CTAACC	CTCTT	NSLIFQTWKDSV
					GGAACTGATCGGATTAACGCGGTGATTGAA	TGTCAG	TTCA	VENMNTEPVLL
					ACGAACACAACTCCGCTAGAGGATATTGAC	TAAAGA	CCCG	HPLIIECMNKPSP
					TTGAGTCCAGAGACCCCAGAAGGGCTTGCA	CAGAAC	CCAT	FKATGPDGIFNS
					GAGTTGCCTAGTACTGTAGAGATAATGGAA	GGGGGC	ATGG	YWRQGFIANWV
					TTGACGGAAGACGGGTCTAGATCCAGAGGG	TTTGCA	ACCA	KSLIQRTIQTGEF
					ACTGCGAACGACGCTGATGGCGGGGTTTCC	ATCTAA	AACG	PASLMCGRTVL
					ATCAGCGATCCACTACGAAATCGCCCAAGC	GGTGTT	CTAT	LYKNGDTAKPE
					TCAAGTCAGGAAAGCAGAAATGTCCCGGAG	GGCAGA	CCAG	NYRPITCLNGCF
					CAAGTAGACCCGGATGGGGAACTTGTATGG	CCACTA	CCTC	KMTNAVITKVIV
					GAGAGTCTATACGGAGCACAGTTGAGGGGA	AAACTG	GCTC	QRVQDTCALPG
					GCGATGCGCACAACTGACAGGCCGAGGTTG	CCCTTT	AGAA	EQMALKPKVW
					CCAAAACTTACGAAATTCTCTGCAGCCGAA	GATGAG	GAGC	ACMEAQLRDQA
					CAGCTGTGGATTAAATCGAAGGTTGAGAAA	GGTGAG	CTTA	LQSEIGNDCKTA
					GCTCGGCTGGAGTGCGTTAGCTACGGCATT	ATAACA	GGGC	WIDFSKAYDSL
					GAACAGCAATGGCTCCGTGCCAGCGCAGTT	TCCTAC	TGGG	DHDAIRFVIETL

					CTATATGCCACGATCAAGACTGTAGCGGCT	GGAACA	GAGT	ALPDGMEKYLL
					TGTCGTCCGTACAACAAGGCCCGCGAAGCT	GCCCCT	ACCA	KSLESWKTKLV
					CACAAGGTATGGCTCGAGAACAAACGTGCT	GCTGAT	CATG	LSNRGKVATGR
					GAAGAAAAGAGAGTACGAAGGATTATTGG	CCAAAC	TGGC	PYKIKRGVLQG
					GCGAATTGAGACTGTACGCACCATGCCGAA	AAAATC	GGAA	DSLSPALFVIATS
					AGGCAAACGAACTGACAAGCAGATCCGCCT	ATCCAC	ACTG	PIVSHLKRVCPS
					TGCGCGGAAGATTAACAGGTTGAAACGTGT	AAATCG	AATC	GRIQLYMDDIKL
					CAGCTTCCCAGAGATGGACTGGCACGGATT	GAGGGT	TGGA	YGKSETELRMLI
					CCTGAATCACTTTAAAGCTAAGTTAGATCTT	TTTCTT	TGCG	KEVQKVANKLG
					CTTAAAAAGTTGATCTCAGTTCGGGTGGCA	GGTAGT	ATGC	LQMNLKKCSTY
					GAACATGAGCGGAAGATTAGTCGGAAGATA	TTCCCT	ATAC	GAGLTESIAGFD
					GCAGGTACATACGGCAAAAGTGTATCCGGT	TGGCAC	CGGG	PLGDRAYKYLG
					CAGTCAGGGTTTACGCCGGATGTCGTTGCC	GTCTTG	AGCG	VPQRSVADTNL
					GCAACGACTTTCTGGTCAGGCCTTGCACAG	TTTCAA	CAGC	AFGELEGKVIRS
					CCTGGACCAAAGAAATTTAAAAAGAACTCG	AAGCCA	GAAA	IEETMACEYLT
					CTGATCTTCCAGACTTGGAAAGACAGTGTA	GAATAT	TCAC	MRQVVTRLNSV
					GTTGAGAATATGAACACTGAGCCAGTACTG	AAAACA	TTAA	IGPLVRFVAQSV
					CTACACCCGCTGATCATAGAATGTATGAAT	ATACCG	CGCT	LTSQAKVSWIY
					AAGCCCAGTCCGTTCAAGGCCACTGGCCCT	AAGTAC	GGTG	NKISDLDSKIRA
					GACGGGATTTTCAACTCCTATTGGAGGCAA	AGAATG	CACC	KLAQTGLRYKK
					GGATTCATTGCGAATTGGGTTAAGTCGCTA	GCTGGA	TCCT	SNVARLYLSKS
					ATCCAGCGAACTATTCAAACAGGAGAGTTC	ACCTAA	GCTA	KNGIGLVNVQQ
					CCGGCCTCACTTATGTGTGGTCGTACGGTGC	CAGCCA	TCGT	VLVEALVSRAIY

					TCCTGTATAAGAATGGTGACACTGCTAAGC	TGGCAC	AGTA	CLRAPSLVEIRE
					CGGAGAATTATCGACCTATCACATGCCTGA	AGAAAA	CTTCC	HILTAEFDPVGA
					ATGGTTGCTTCAAGATGACTAATGCTGTTAT	CCGGTG	TAGA	ARTVLRRSRIQL
					CACGAAAGTAATCGTTCAACGGGTTCAAGA	GGTGCA	TAGA	EIERVEMASAIS
					CACCTGTGCTCTCCCGGGAGAACAGATGGC	CACCGG	TGAG	AIKTNYQARWM
					CCTGAAGCCGAAGGTGTGGGCGTGCATGGA	AGACAG	TAGG	TKFTQSKLYQK
					AGCGCAGTTGCGCGACCAGGCTCTACAGTC	CAGGGG	GTGG	WVHHDIDLANS
					TGAGATCGGTAACGATTGTAAGACTGCGTG	AAAGAT	GCTA	NLWLERGEISPQ
					GATCGATTTTTCAAAAGCTTATGACAGTCTG	TGCAAA	AAGG	NARIAVAAQDN
					GACCACGACGCGATCCGATTCGTGATAGAA	GTGCCG	TAGT	TLLCRGFVGNR
					ACACTTGCTTTACCGGACGGAATGGAGAAA	TAAATA	CGTC	ESEKQCRMCNM
					TATCTCCTGAAGTCGCTAGAGAGTTGGAAG	AAGATG	TTCA	GIETCSHILTECS
					ACAAAACTGGTGCTTAGTAATCGAGGTAAA	GTAGCT	ATGG	YHRAHMYIERH
					GTTGCCACTGGGCGACCTTACAAGATTAAG	CTCTGA	CAAT	DSVARNIYAVL
					AGAGGTGTTCTGCAAGGTGACAGTCTTAGT	CCTGAG	ACCG	AKDHGLWIPHY
					CCGGCCCTATTTGTCATTGCGACAAGTCCTA	GGAATG	AGGG	SQPVSSVTKTPT
					TCGTCTCCCATCTAAAGCGTGTTTGCCCATC	GGACAC	GTCT	CELYWNYKFPC
					CGGCCGTATACAATTGTATATGGATGACAT	TGTCCG	CGTG	TRALEACRPDIV
					TAAACTGTATGGCAAATCAGAGACCGAATT	ATTAAT	ACAG	LIDRAKRTILIVE
					ACGTATGCTCATTAAAGAAGTTCAAAAGGT	GCCTTA	GTGA	VAVSYVTRLKQ
					AGCCAACAAATTGGGGCTCCAGATGAATCT	TCCCGA	GGGA	MVSRKVYKYG
					AAAAAAATGTAGTACCTATGGTGCTGGCCT	AGACGG	TAAC	VNGEKGADGES
					AACAGAGAGTATCGCCGGGTTTGACCCACT	TCAAGG	TAGT	RGWNMIRELSE

					CGGCGACAGAGCATACAAGTATCTCGGAGT TCCCCAGCGGAGCGTCGCAGATACTAATTT GGCCTTTGGTGAGCTTGAAGGCAAGGTAAT ACGGTCCATCGAGGAAACGATGGCGTGCGA ATACTTGACAATGCGGCAGGTGGTCACAAG ATTGAACTCTGTGATCGGTCCTCTTGTCCGG TTTGTAGCACAGTCCGTACTAACCAGTCAG GCGAAAGTAAGCTGGATATATAACAAAATA TCGGATCTGGATAGCAAGATAAGAGCGAAA TTAGCTCAAACAGGCCTAAGGTACAAGAAA AGCAACGTAGCAAGACTCTACCTAAGCAAG AGTAAAAATGGCATCGGTTTGGTGAATGTC CAGCAGGTATTAGTTGAAGCATTGGTGAGC CGAGCCATCTATTGTCTGAGAGCCCCGAGT CTCGTTGAGATTCGGGAGCATATACTAACA GCAGAGTTCGACCCTGTAGGAGCTGCTCGT ACCGTGCTGCGGAGGAGTAGGATCCAACTG GAGATCGAGCGCGTCGAAATGGCTTCGGCT ATATCGGCGATTAAAACAAATTACCAGGCA AGATGGATGACGAAGTTCACCCAATCGAAA CTATACCAGAAGTGGGTTCACCATGACATC GATCTAGCCAACTCAAACTTATGGTTAGAG AGGGGCGAAATCTCTCCGCAAAACGCAAGG	TTCTAT CTTATC GGTCTC CAAAGG GCTCTG GCCATA GAAGCT GCGAGG AAGCCG TAACCC TACGCA GATTGC CGCCGT GC (SEQ ID NO: 1258)	ATCA GATA ATAT CAGT CATG AAAT TAGC AACA ACCA ACGT CACC GTCG TTGG CAAA ACAC CGAC TAAC GATG CTAG TTAG AAAG AGTC GGGT	VYNMKVNLCA VVIGASGEVLPC TVKAIQSISSKTS SRQLLERCQRSA VLGSTRVVKRH LAEFH (SEQ ID NO: 1503)

					ATTGCGGTAGCGGCTCAAGATAATACGTTA CTGTGTCGCGGGTTTGTAGGTAATCGGGAG TCAGAAAAACAGTGCCGCATGTGTAACATG GGCATTGAGACCTGCAGTCATATCCTGACT GAATGTAGTTATCACCGGGCGCATATGTAT ATAGAACGTCACGATTCTGTGGCAAGGAAT ATCTATGCTGTCCTGGCAAAAGACCACGGT CTCTGGATTCCGCATTACTCGCAGCCCGTGA GTTCGGTAACAAAGACGCCGACGTGTGAAC TTTACTGGAATTACAAGTTTCCATGTACCCG TGCATTGGAAGCTTGCCGTCCAGATATCGTT CTCATCGATCGCGCGAAAAGAACTATCCTC ATAGTGGAAGTTGCGGTGTCGTATGTCACA CGGCTAAAGCAGATGGTGTCACGAAAAGTG TATAAATATGGAGTAAACGGTGAAAAGGGA GCCGACGGCGAAAGCCGCGGCTGGAATATG ATCCGTGAACTATCTGAAGTTTATAATATGA AAGTGAACCTTTGCGCAGTCGTGATTGGGG CGAGTGGGGAAGTTCTCCCTTGTACGGTCA AAGCCATTCAGAGCATATCTAGCAAGACCT CCAGCCGACAGCTCCTAGAGCGGTGCCAGC GTAGCGCTGTCCTAGGGTCGACGAGGGTTG TCAAACGACATCTTGCAGAATTCCATTGAG		CTTCC CAAA GTTA GGTG CTTG CACC GAAG CCGA TCCG CTCT ACCC ACAG CTCT GCCC AGCG TT (SEQ ID NO: 1381)	

					ACCACCCATGCGCAGAGTATCCGAATCAGT GAAAGTCCAAGTTTCAGGACAGAAACGTCA GATAAGTCCAAGAGAAACGAGAAAACAAG TTCAAGTATGCAAGAGTTAATCAATAAGAG AGTACCGTAAATGTATGACCCCCCCCTTTGC CAAGTCGACAACTGTCATGCAGGTGTCTCT CTTTTCACCCGCCATATGGACCAAACGCTAT CCAGCCTCGCTCAGAAGAGCCTTAGGGCTG GGGAGTACCACATGTGGCGGAAACTGAATC TGGATGCGATGCATACCGGGAGCGCAGCGA AATCACTTAACGCTGGTGCACCTCCTGCTAT CGTAGTACTTCCTAGATAGATGAGTAGGGT GGGCTAAAGGTAGTCGTCTTCAATGGCAAT ACCGAGGGGTCTCGTGACAGGTGAGGGATA ACTAGTATCAGATAATATCAGTCATGAAAT TAGCAACAACCAACGTCACCGTCGTTGGCA AAACACCGACTAACGATGCTAGTTAGAAAG AGTCGGGTCTTCCCAAAGTTAGGTGCTTGC ACCGAAGCCGATCCGCTCTACCCACAGCTC TGCCCAGCGTT (SEQ ID NO: 1135)			
R4	R4-4_BX	.		Bursaphelenchus xylophilus	TGCCAGCGGTGTTGATTAGGTCCAAGTTCTT TGGCCAAAGATCCGCCCTCGGTTTAGCAGT ACCGAACGAGTATACCTTCAAGTGGTGGCA	TGCCAG CGGTGT TGATTA	TAAT TAGG CAGT	MEILWEDLRLKI EDRYGVTLPQR SASSLKNQYPK

					CTGAATTAGACTGAATACTCTGAACTGTAG	GGTCCA	GCTC	VILRGLPDSGLP
					ACTTTTGTGCAACTGTGTATGGTGTGGAAG	AGTTCT	CTAG	WAGVQVNDTG
					ACTTGTTTGTACCACTATCAGCTTTATTGGG	TTGGCC	CGGA	QVVVVDHAEAA
					GCTCGTTACTGTTTCATACAGGTAGATGTCC	AAAGAT	GTGC	TLRGSSPAAVD
					CCTTTAGAGATTTCCCTGCAGTTTGCGCTCC	CCGCCC	CGTG	GEAEEPVVPPLP
					AAGTCGCTAGCCTCTTGCGTGTAGTCAAAG	TCGGTT	AAGT	AAEVVESAADA
					GAATACATTCGCCGTCGGTGACAGGGCTAT	TAGCAG	GGTG	AVPDPQSEIVAD
					ACCCGGCGACTACGGACTTGTTTATTACGTA	TACCGA	TCAG	QGVETRPVENPP
					GTGCAGCCTCGTTTAAGACGAATGTGAAAA	ACGAGT	TACC	ANSRETETEPVE
					GAAGGTGTGATTACTAAGCGTTATGAGTCG	ATACCT	CGTC	VEPYLEGQYKFF
					GGTTATCTGGAAACTCCGCCCCGCCGCAAT	TCAAGT	GTGT	VSKILGKSMWR
					GGCTTCAAGGGTCAAACTGCTAACATTTTA	GGTGGC	GGAA	KPIKYPRRVPET
					AACCAATATTTGGCAGCAGCAGACTATGAT	ACTGAA	AGCC	LWRQANELIERS
					GTCAGAGTGCCGGGGGTCTATCGATATAAG	TTAGAC	CAGG	IRQGEVSIQSLN
					ACTGAAAGCGATAGACGTGGAGTGAAAGG	TGAATA	AGGG	CMVYAAGCAV
					ATTCCCTTTGTTAAGAGAATCGGTAGAATTC	CTCTGA	TTAG	KSSLDKKDQEA
					ACTTTTTACTTATTCAAGAACTTAACAGCAA	ACTGTA	TACC	KRRESEWYACR
					CAAGCACTCGCGAGGATTACCGCCCCAGAT	GACTTT	GACA	KAEIKALERYLN
					TCGGTCGGCGGTACTTCACCTGCTTTCTTCC	TGTGCA	GTGG	FIDLELKRRSAS
					ACTTTCGGAATCTGGCATCCTGGCTTTCAGT	ACTGTG	GAAA	RPLTSRQRQNLG
					GTGGTGATGGCCGGCTTGAGTTTTCTTGAGT	TATGGT	CCCG	VLITKYGRARV
					CGTGCGAGTGCCTCATCTGGGACGTCCGGA	GTGGAA	CTGC	RSGVRLSELQA
					CCGATTGATGGAGTCTGCAGTGGACGAGGA	GACTTG	ACGC	MLRDALVGIRK

					CTTGATGGACCGTAACCATAGTATATCCCTC	TTTGTA	AACC	CMAKRSADKKR
					ATGCGTCTTCTCGACTCGAGGGGGTAGCTT	CCACTA	TAAA	KQGKFVPIQRYL
					GCACTACCCACCCTTCTCTTCTCCGATTGGG	TCAGCT	GACA	EPSSAEPRLSPD
					ATTTAGACCTAGCCCTCTGGTGTGTCTCGAC	TTATTG	TTTGC	TVRAYWNDIVG
					CCGCGATATCAGATTCCTGAATCGACTGTG	GGGCTC	CCTTC	SSQQSTSDSTIQ
					AGAAATGTCTACGCGCAAAGATCGACCCAT	GTTACT	GGGG	DWSSNLSVPSQE
					TGCCACCCGGCTATGTGGATCGGGCTCTTG	GTTTCA	GAGA	LNASKIMGWWR
					ACTGCTTATCTCCGGCTTTAATCGCTTGAGG	TACAGG	GGTA	AAVSKSKPNKA
					AAAGGGGGGTGTCGCCCGAAAGGGTTACGC	TAGATG	TGAG	AGPDGIPGVLW
					GATCATCCGATTCCTACCGTAAAAACGTAG	TCCCCT	ACCA	KRFRSASEWVC
					TTGAAGTAGGATTGAACCTGAGTACCAAGT	TTAGAG	CACT	TWLYRLLQKRR
					GAAAAGAGTGGCTATAATGCTCCATGGTAT	ATTTCC	AGTC	IITPRWLSVGRV
					ACCCTAGGGATAACCGTGTCGAGCGCACTG	CTGCAG	ATGG	VLLPKKGPLEDP
					CTCAATACCCTGATTTGTTAGTGTAGGTATG	TTTGCG	TTGCT	ANYRPIACLNTV
					TCTAGTGGCTGCCTAGGCAGATTCGCCTATT	CTCCAA	TGCG	YKLITSVVEMA
					CACTACGTAAAATCTGGGTGAGTACCATAA	GTCGCT	CAAG	VREQIQACPGLV
					GAACCTCCTGTAGGGCCAGGAGTCAAACTA	AGCCTC	CATG	PYEQIANRKGV
					GTCCAGGTTCGGTATGCTTCGGTATACTCTC	TTGCGT	ACTC	WGCTHASIVDR
					TCACGGGCTAGTCACCTAAGAGTTAAGAAC	GTAGTC	ATTC	MITGASREGKG
					CGCCTTTTTCTGCACTGTGAATAGAAAAAG	AAAGGA	ATTG	GGFPDLRVLFY
					AAGGGCGGGAGAATATACGGGCGAGGTAA	ATACAT	TAAG	DCKKAFDSVNR
					GCACGTCGGAACGGGGTGTGCCCAACTCGT	TCGCCG	TTCGT	DHMFAVLRVAN
					ACCGTACGCCCGGACAAGGACCATCTTTCG	TCGGTG	ATTA	VNVKVVHLLHT

					TACCCCGTTCACCCGGCACAAGTCCGATTGT	ACAGGG	TGAG	LSQQWCVRYEL
					CTCTCCCGAGAGGCGTCGGCGGAGGTTGGC	CTATAC	GCCT	RRNNRVERSSPL
					TAACGCCGTCTCTCCAACCATCGGTTTGGGT	CCGGCG	CGCG	RVKRGLLQGDT
					TTAAGTAACGCCCCAGTGGCCGGTAGCCAA	ACTACG	GAAT	LSPTWFCLCMA
					ACTGATGGAGGCAGAAATGCCGGTCCGTTG	GACTTG	GCGA	PISASIKTLNPGP
					CTAAATGCGGGAACGGACCAAAACGCCGGT	TTTATT	AAGT	TLRPNMGNGRN
					GTGGTGGATACAGGTGGAAGAGTCTTTTGG	ACGTAG	CCAA	RGQVAIQVSHV
					TCTACGCAAGAAAAGACCAGGCTCAAGTAC	TGCAGC	GAGC	FYMDDLKVYCP
					GAATACGACCGTCTTCGTGAACGACGCCGT	CTCGTT	GCCA	RVADQRRMEQN
					CGAGACAAAGTTCGAAATCCTACGGTCGAT	TAAGAC	ACTT	IPQLFGEIGLSIN
					AAGGCCATGGAAATCCTTTGGGAAGATCTC	GAATGT	GTCT	ASKSAAAAAVG
					CGGTTGAAAATTGAAGATAGGTACGGTGTC	GAAAAG	CTTTG	RYVESELPVLGT
					ACATTGCCGCAAAGAAGCGCAAGCTCTTTG	AAGGTG	GGAG	KDEYKYLGIESG
					AAAAACCAGTACCCCAAAGTTATTCTGCGG	TGATTA	GTTC	FVVNEVAALDR
					GGGCTCCCAGATTCGGGCCTGCCTTGGGCG	CTAAGC	CTCC	MQAVLLNRVEA
					GGGGTTCAGGTAAATGATACTGGACAAGTC	GTTATG	GGGT	ILSVKEHTVGQR
					GTAGTCGTAGACCATGCTGAGGCGGCTACC	AGTCGG	TCCTT	RDAIRAKAIPGG
					CTTCGGGGGTCCAGCCCAGCAGCGGTCGAC	GTTATC	AGGG	AYILGHIILSDLD
					GGGGAGGCCGAAGAACCGGTAGTACCCCCG	TGGAAA	TTGG	PRGAAERMRRL
					TTACCCGCGGCAGAGGTCGTTGAATCGGCT	CTCCGC	CGTA	DIEIRRLVKSAGI
					GCTGACGCTGCGGTGCCCGATCCGCAAAGC	CCCGCC	CCAT	LHDKCSTARIHL
					GAGATCGTGGCTGACCAGGGTGTGGAGACA	GCAATG	GCTT	SCEQGGLAWPS
					CGCCCTGTCGAGAACCCCCCCGCTAATAGC	GCTTCA	CGTG	MERAYYVAVA

					CGGGAGACTGAAACGGAACCCGTAGAGGT	AGGGTC	TGAG	YSASYLLTSQDE
					AGAGCCTTACCTCGAAGGGCAGTATAAATT	AAACTG	ACCT	TISRARDYFVSG
					CTTCGTTTCTAAGATCCTTGGGAAATCGATG	CTAACA	AGGC	RLSNKFTVYKH
					TGGCGGAAGCCTATCAAATACCCGAGGAGA	TTTTAA	CGCT	LTSIVDSLGLSV
					GTCCCGGAAACCCTGTGGCGTCAAGCCAAT	ACCAAT	TGGA	ELPDPNGLPTGQ
					GAGTTGATTGAGAGATCGATTCGACAGGGG	ATTTGG	CCTA	PSVLARTIARAI
					GAAGTGTCGATTCAGAGCCTGAACTGCATG	CAGCAG	GAGT	DAKLEAQWKET
					GTTTATGCTGCTGGGTGCGCTGTTAAGTCTT	CAGACT	TAAG	LLTYQRAGRVE
					CGCTGGACAAGAAGGATCAGGAAGCTAAG	ATGATG	CTGTT	RADPTVVDHAN
					CGGCGTGAATCCGAGTGGTATGCGTGTAGG	TCAGAG	GCAT	SYHWLRKAWIN
					AAGGCCGAGATTAAGGCACTCGAGAGGTAT	TGCCGG	GTTT	EKAYQHAVSVM
					CTCAATTTCATTGACCTGGAGTTAAAGCGTC	GGGTCT	GAGG	EGTLLEGVNPH
					GCTCTGCGTCGCGGCCGTTGACATCAAGGC	ATCGAT	ATGC	GVLTMCRACKA
					AAAGGCAGAACTTGGGTGTGCTCATCACGA	ATAAGA	CTTA	PSASIAHIITGCA
					AGTATGGGCGTGCTCGGGTTAGGTCTGGTG	CTGAAA	GGCG	ELRKSHMKVRH
					TCCGGCTCTCGGAATTGCAGGCAATGTTAC	GCGATA	GACG	DGVTRWLYNAL
					GAGACGCCTTAGTGGGAATCCGAAAGTGTA	GACGTG	CCAT	TEVDGSLPKFHY
					TGGCGAAGCGCTCGGCAGACAAGAAACGG	GAGTGA	G (SEQ	TQQIPAEMRGER
					AAACAGGGCAAATTTGTTCCTATTCAACGTT	AAGGAT	ID NO:	LTVRYDSDIVTP
					ACTTAGAGCCTAGCAGTGCGGAGCCTCGAC	TCCCTT	1382)	NKPRHNRPDLV
					TCTCTCCAGACACGGTTCGTGCATATTGGAA	TGTTAA		VFDSTRKVIYIV
					TGACATTGTGGGATCTTCCCAACAGTCAAC	GAGAAT		EVSVTWLSVLQ
					GTCAGATTCGACGATTCAAGACTGGTCCTC	CGGTAG		KQYDNKLNRYA

					CAATTTAAGCGTTCCGAGTCAAGAGTTGAA	AATTCA		VNSNHEFSESIP
					CGCAAGTAAGATCATGGGGTGGTGGAGAGC	CTTTTT		YPPGVNLANEIR
					CGCCGTTTCGAAGAGCAAGCCCAACAAAGC	ACTTAT		VLYPQFTGGVK
					TGCTGGTCCTGATGGTATTCCCGGGGTACTG	TCAAGA		VFPMIISPTGEV
					TGGAAACGCTTCAGGTCGGCTAGTGAGTGG	ACTTAA		HMQFVPHLAEL
					GTTTGTACCTGGTTATATCGCTTATTACAGA	CAGCAA		LENPNIPRILEKI
					AGCGAAGGATAATCACGCCAAGATGGTTGT	CAAGCA		QRSVVLGTDYII
					CTGTTGGTAGAGTCGTATTGTTACCAAAGA	CTCGCG		RSYFAM (SEQ
					AAGGACCTCTCGAGGATCCCGCCAACTATA	AGGATT		ID NO: 1504)
					GACCCATCGCGTGCCTGAACACTGTATATA	ACCGCC		
					AGTTGATAACCTCGGTAGTAGAAATGGCAG	CCAGAT		
					TTCGAGAGCAGATCCAGGCTTGTCCTGGAC	TCGGTC		
					TAGTCCCATACGAGCAGATAGCTAACCGTA	GGCGGT		
					AAGGCGTATGGGGCTGTACGCATGCGAGTA	ACTTCA		
					TCGTAGACCGTATGATAACCGGTGCAAGCC	CCTGCT		
					GAGAAGGTAAGGGCGGTGGGTTTCCGGACC	TTCTTC		
					TACGCGTCCTATTTTACGATTGTAAAAAGGC	CACTTT		
					CTTCGACAGCGTAAACCGAGATCATATGTT	CGGAAT		
					CGCTGTTTTGAGAGTAGCAAACGTGAACGT	CTGGCA		
					TAAGGTCGTGCATCTACTCCACACACTATCT	TCCTGG		
					CAGCAATGGTGTGTCCGATACGAGCTCCGT	CTTTCA		
					CGAAACAACCGTGTGGAGAGAAGTAGCCCG	GTGTGG		
					CTGAGGGTGAAAAGAGGACTGCTGCAAGGC	TGATGG		

					GACACTCTCTCTCCTACATGGTTTTGCCTAT	CCGGCT		
					GCATGGCTCCAATATCAGCCTCTATTAAAA	TGAGTT		
					CCCTTAACCCTGGCCCAACCCTTCGGCCGA	TTCTTG		
					ACATGGGGAATGGGCGAAATAGAGGGCAG	AGTCGT		
					GTGGCGATCCAAGTTAGTCATGTGTTCTACA	GCGAGT		
					TGGATGACTTGAAAGTGTATTGCCCCCGCG	GCCTCA		
					TAGCCGATCAACGCAGAATGGAACAGAATA	TCTGGG		
					TTCCACAGTTGTTCGGTGAAATTGGACTAA	ACGTCC		
					GCATAAATGCGAGTAAGAGCGCAGCAGCCG	GGACCG		
					CAGCCGTGGGAAGATACGTTGAATCCGAGT	ATTGAT		
					TGCCAGTTCTTGGGACGAAGGATGAGTACA	GGAGTC		
					AGTACCTGGGAATAGAAAGCGGTTTTGTTG	TGCAGT		
					TAAATGAGGTGGCCGCCCTTGACAGAATGC	GGACGA		
					AGGCCGTTCTGCTAAACCGTGTAGAAGCGA	GGACTT		
					TTCTATCGGTTAAAGAACATACTGTTGGTCA	GATGGA		
					GAGGCGCGACGCTATACGAGCGAAAGCGAT	CCGTAA		
					CCCTGGGGGGGCCTATATCCTGGGTCACAT	CCATAG		
					CATTTTGAGTGACTTGGACCCCAGGGGTGC	TATATC		
					TGCTGAGCGCATGCGTCGGTTGGATATCGA	CCTCAT		
					GATTCGGCGTTTAGTCAAATCAGCTGGTATT	GCGTCT		
					CTTCACGACAAGTGTAGTACAGCGCGTATT	TCTCGA		
					CACCTTAGCTGTGAACAGGGAGGTCTAGCT	CTCGAG		
					TGGCCGTCGATGGAGAGAGCTTATTATGTA	GGGGTA		

					GCAGTTGCATATTCAGCCTCCTATCTTCTGA	GCTTGC		
					CAAGTCAAGACGAAACTATCAGTCGTGCCA	ACTACC		
					GAGACTACTTCGTCTCGGGTCGATTGAGTA	CACCCT		
					ACAAGTTTACTGTTTACAAGCACCTTACGA	TCTCTT		
					GCATCGTAGACTCCCTTGGTTTATCGGTAGA	CTCCGA		
					ACTGCCGGATCCGAATGGACTTCCTACCGG	TTGGGA		
					TCAACCGTCCGTCCTGGCTAGAACTATTGCG	TTTAGA		
					AGAGCGATAGATGCCAAGTTGGAAGCGCAG	CCTAGC		
					TGGAAGGAAACTTTGCTAACTTACCAGCGC	CCTCTG		
					GCAGGGCGAGTCGAAAGGGCAGATCCGAC	GTGTGT		
					GGTCGTGGACCATGCAAACTCATACCATTG	CTCGAC		
					GTTGAGAAAAGCATGGATAAACGAGAAAG	CCGCGA		
					CGTACCAACATGCCGTGAGTGTTATGGAGG	TATCAG		
					GAACTCTTTTGGAGGGAGTGAACCCTCACG	ATTCCT		
					GTGTGTTGACCATGTGTAGAGCGTGTAAAG	GAATCG		
					CTCCAAGTGCAAGCATTGCTCACATCATTAC	ACTGTG		
					TGGCTGCGCCGAATTGCGCAAAAGCCATAT	AGAAAT		
					GAAGGTGAGGCATGATGGAGTTACTCGCTG	GTCTAC		
					GCTATACAATGCTCTGACTGAGGTAGATGG	GCGCAA		
					TAGCCTACCGAAGTTCCACTATACGCAACA	AGATCG		
					GATCCCGGCAGAGATGAGGGGTGAGCGACT	ACCCAT		
					CACTGTCCGTTACGACTCGGACATAGTTACT	TGCCAC		
					CCCAATAAACCTCGTCACAACAGACCGGAT	CCGGCT		

					CTTGTTGTCTTCGACTCTACTAGAAAGGTGA	ATGTGG		
					TTTATATAGTAGAGGTCAGTGTAACGTGGTT	ATCGGG		
					GAGTGTGTTGCAAAAGCAATACGATAACAA	CTCTTG		
					GCTCAACAGGTATGCAGTGAACAGTAATCA	ACTGCT		
					CGAGTTCAGTGAGAGCATACCATATCCGCC	TATCTC		
					TGGTGTGAACCTTGCTAACGAGATAAGGGT	CGGCTT		
					TTTGTATCCTCAGTTTACTGGGGGTGTGAAG	TAATCG		
					GTATTCCCTATGATTATCTCGCCAACAGGGG	CTTGAG		
					AAGTACACATGCAATTTGTGCCCCACCTTGC	GAAAGG		
					AGAGTTACTCGAAAATCCGAACATTCCTCG	GGGGTG		
					GATTCTAGAGAAAATTCAACGCTCGGTTGT	TCGCCC		
					GCTTGGTACAGATTACATTATCCGGTCATAC	GAAAGG		
					TTCGCCATGTAATTAGGCAGTGCTCCTAGCG	GTTACG		
					GAGTGCCGTGAAGTGGTGTCAGTACCCGTC	CGATCA		
					GTGTGGAAAGCCCAGGAGGGTTAGTACCGA	TCCGAT		
					CAGTGGGAAACCCGCTGCACGCAACCTAAA	TCCTAC		
					GACATTTGCCCTTCGGGGGAGAGGTATGAG	CGTAAA		
					ACCACACTAGTCATGGTTGCTTGCGCAAGC	AACGTA		
					ATGACTCATTCATTGTAAGTTCGTATTATGA	GTTGAA		
					GGCCTCGCGGAATGCGAAAGTCCAAGAGCG	GTAGGA		
					CCAACTTGTCTCTTTGGGAGGTTCCTCCGGG	TTGAAC		
					TTCCTTAGGGTTGGCGTACCATGCTTCGTGT	CTGAGT		
					GAGACCTAGGCCGCTTGGACCTAGAGTTAA	ACCAAG		

					GCTGTTGCATGTTTGAGGATGCCTTAGGCG GACGCCATG (SEQ ID NO: 1136)	TGAAAA GAGTGG CTATAA TGCTCC ATGGTA TACCCT AGGGAT AACCGT GTCGAG CGCACT GCTCAA TACCCT GATTTG TTAGTG TAGGTA TGTCTA GTGGCT GCCTAG GCAGAT TCGCCT ATTCAC TACGTA AAATCT		

						GGGTGA GTACCA TAAGAA CCTCCT GTAGGG CCAGGA GTCAAA CTAGTC CAGGTT CGGTAT GCTTCG GTATAC TCTCTC ACGGGC TAGTCA CCTAAG AGTTAA GAACCG CCTTTTT CTGCAC TGTGAA TAGAAA AAGAAG		

						GGCGGG AGAATA TACGGG CGAGGT AAGCAC GTCGGA ACGGGG TGTGCC CAACTC GTACCG TACGCC CGGACA AGGACC ATCTTT CGTACC CCGTTC ACCCGG CACAAG TCCGAT TGTCTC TCCCGA GAGGCG TCGGCG		

						GAGGTT GGCTAA CGCCGT CTCTCC AACCAT CGGTTT GGGTTT AAGTAA CGCCCC AGTGGC CGGTAG CCAAAC TGATGG AGGCAG AAATGC CGGTCC GTTGCT AAATGC GGGAAC GGACCA AAACGC CGGTGT GGTGGA		

						TACAGG TGGAAG AGTCTT TTGGTC TACGCA AGAAAA GACCAG GCTCAA GTACGA ATACGA CCGTCT TCGTGA ACGACG CCGTCG AGACAA AGTTCG AAATCC TACGGT CGATAA GGCC (SEQ ID NO: 1259)		

					GGGGCCGGTGGGTTTACTCACTTCTGACCC	GGGGCC	TAGT	MPCSTNSFFERG
					ACCACCAACGGAACGAGGGAAAGCAGAGC	GGTGGG	CGCT	TPEPHREPISGTD
					TGGGGCCCTCTTCCGATTGGCATGGAACCG	TTTACT	AAGG	SSESLGMGTHRS
					ACCTCCACGTGGTGGCCCTGGGCAACGGAA	CACTTC	GGTC	PRLNDDEVINGP
					TTCAAGAGAGGATTTAATCCTCTCTATCATT	TGACCC	CGGA	KGHESDPVHVV
					TGCAAGATGGATGAGATCGAGGTATCCGGC	ACCACC	AATG	RAPRTLHPRRLE
					AAACAGGTTCCAAGTGAGCACCTTTCCCAT	AACGGA	GTCC	LPIGVNNLGEAS
					AGCTGGGAATATGGGTTAGGCGTCCTCTGA	ACGAGG	GGTC	QLRQDSAIAEEA
					CATATAAGAGGAATCAGACTCGTTCGCGCC	GAAAGC	CTGC	QLESTENHDGR
					CGGTCATTAACATCGATCAGCGGGAGGGCC	AGAGCT	GCTA	RPPLRGGRKLW
				Ascaris	GGACTGAAGTAAATTTCCTGTTGGCCCGAG	GGGGCC	CCCG	SEKEIATLRRLC
R4	R4_AL	U29445		lumbricoide	TGCAGGTGGAGCTCGGACCCGAAAAACGAT	CTCTTC	GTTCT	EAYGNRQVCW
				s	CCCTAAGAGGACCACAACCCGAAGGGATGG	CGATTG	GGTA	KEVQRKFADFH
					ACGCAGTCGCCCCGGCACGTTTGGTGTTGG	GCATGG	GCAC	EERTVAALATK
					TTATCCTGGAGTGTTGTGGGACGAATAGCT	AACCGA	GTTC	WGALKRPRAPM
					ATGCCTTGTTCAACTAACTCTTTTTTTGAAC	CCTCCA	AAGC	VGAPPTPDHDPE
					GAGGGACCCCTGAGCCCCATCGTGAGCCAA	CGTGGT	GCTC	RGPAGEGDGGT
					TCTCAGGGACTGACTCTTCGGAGTCCTTAGG	GGCCCT	AATC	TSQENVPTDDPI
					TATGGGTACCCACCGTTCACCTCGGCTAAAT	GGGCAA	GCCT	PANGPTEGKESD
					GATGACGAGGTGATAAACGGCCCCAAGGGT	CGGAAT	GCCT	VRPAVACRCTE
					CACGAAAGTGACCCGGTTCATGTAGTGCGT	TCAAGA	TGTA	PEEQLMESDVRP
					GCGCCTCGAACGCTACACCCGAGGCGATTG	GAGGAT	GGCA	PAVVRLADPEQ
					GAACTCCCAATCGGAGTAAATAATCTCGGA	TTAATC	GTCC	HTMKSGVKPVA

					GAAGCCTCTCAACTACGGCAAGATTCAGCA	CTCTCT	ATCT	LDGSADLEERPK
					ATAGCTGAGGAAGCCCAATTGGAGAGCACC	ATCATT	GTGG	EKDIEQMGVDF
					GAGAACCATGATGGACGTCGCCCTCCTTTG	TGCAAG	AAGT	EGEPRFRAFRKA
					AGAGGCGGCAGGAAGCTCTGGAGTGAAAA	ATGGAT	CGCG	FYGYFRWAVNS
					GGAGATCGCCACATTAAGGAGATTGTGTGA	GAGATC	CTCTT	FDREPVKRVRR
					GGCCTATGGAAATAGGCAGGTATGCTGGAA	GAGGTA	GATA	DCPKVFYAYAD
					AGAGGTTCAGAGGAAGTTTGCGGACTTCCA	TCCGGC	CAGA	YLIATGSSKALG
					CGAGGAAAGAACAGTGGCTGCGTTAGCCAC	AAACAG	TGTG	PNQSRIGRLNGL
					GAAGTGGGGGGCTCTTAAACGCCCGAGAGC	GTTCCA	GACG	VYAAARTIHQF
					TCCGATGGTTGGTGCGCCACCGACTCCTGAT	AGTGAG	GATG	WREEVGHRQQG
					CATGATCCAGAGCGTGGACCAGCGGGCGAG	CACCTT	GAAG	EKGWYTKTKAT
					GGAGATGGAGGTACAACCTCGCAAGAGAAT	TCCCAT	CAGA	REDLQMLISMM
					GTGCCTACCGATGATCCAATACCCGCGAAC	AGCTGG	TGAT	ESELARRKEKR
					GGCCCGACAGAGGGCAAGGAATCGGATGT	GAATAT	AGAG	KPGAKELENIHK
					GAGACCAGCCGTGGCCTGTAGGTGCACGGA	GGGTTA	CCGG	LVARLGTRSTSG
					ACCGGAGGAGCAACTGATGGAGTCGGATGT	GGCGTC	TGAC	IVRRLEMTRQRL
					GAGACCGCCTGCAGTCGTAAGACTTGCAGA	CTCTGA	GGCC	KLLEDRISLHEQ
					TCCGGAGCAGCATACGATGAAGTCGGGTGT	CATATA	CTAC	EKRRKRLRKQF
					GAAACCGGTTGCACTTGACGGGTCAGCGGA	AGAGGA	TAGC	AETPSLKLLTKG
					TCTGGAGGAGCGTCCGAAAGAGAAGGATAT	ATCAGA	CAAA	AKDRGDTMVT
					CGAGCAGATGGGTGTTGACTTTGAAGGGGA	CTCGTT	CGC	MKSVMDFWRPI
					ACCGAGGTTTCGTGCCTTTCGGAAAGCCTTT	CGCGCC	(SEQ	IGRRVTSNPDQL
					TACGGGTATTTCAGATGGGCTGTCAACTCGT	CGGTCA		QVLRDWRDEQK

					TTGATAGGGAACCTGTCAAGCGAGTTAGGC	TTAACA	ID NO:	KAYPADLDLEK
					GGGACTGTCCCAAGGTATTTTACGCTTATGC	TCGATC	1383)	ADLEEKYEGAIR
					GGATTACCTTATCGCAACCGGAAGTTCTAA	AGCGGG		RIQPWKAPGPD
					GGCGTTGGGACCGAATCAATCCAGGATTGG	AGGGCC		GLHAHWWKAL
					TCGCTTGAACGGATTGGTCTATGCGGCAGC	GGACTG		PSAKRLLGELVV
					TAGAACAATCCACCAATTCTGGAGAGAAGA	AAGTAA		DWLTTGKVTTG
					GGTAGGCCATCGTCAGCAAGGCGAGAAGG	ATTTCC		WMCRGRTILIPK
					GATGGTACACGAAAACCAAGGCAACCCGTG	TGTTGG		KGDRGDPSNYR
					AAGACCTTCAGATGCTCATCTCTATGATGG	CCCGAG		PITCLNTCYKVL
					AATCTGAACTTGCAAGAAGAAAGGAGAAG	TGCAGG		TSVMNSVILSHL
					AGGAAGCCCGGCGCGAAAGAGCTGGAGAA	TGGAGC		SRGEALPMNQR
					TATCCATAAGCTAGTTGCAAGACTTGGAAC	TCGGAC		AMRKREWGCT
					GCGCAGCACATCGGGCATCGTCAGAAGGCT	CCGAAA		HAMVLDRAMV
					GGAGATGACAAGACAGAGGCTCAAACTGTT	AACGAT		MDAMAQKKHS
					GGAAGACAGAATAAGTTTGCATGAGCAAGA	CCCTAA		LSVAWLDYRKA
					GAAGAGGCGGAAACGATTGCGCAAGCAATT	GAGGAC		YDSVSHEYIRW
					TGCGGAAACCCCTTCTCTAAAATTACTCACA	CACAAC		AINSVNIPRSVQ
					AAAGGAGCCAAGGATAGGGGCGATACGAT	CCGAAG		LTLKRLMSDWE
					GGTAACCATGAAATCTGTAATGGACTTCTG	GGATGG		TRFESTQCRPKL
					GAGACCAATAATTGGCAGACGAGTTACCTC	ACGCAG		RSDKMKVLNGI
					CAATCCGGACCAATTGCAAGTCTTGAGAGA	TCGCCC		FQGDSLSPTLFV
					CTGGAGAGATGAGCAGAAGAAGGCTTATCC	CGGCAC		LCIAPISYALNK
					GGCAGACCTGGATTTAGAAAAGGCCGATCT	GTTTGG		GVGQCQSSSGW

					TGAGGAGAAATATGAGGGAGCAATCAGGA	TGTTGG		SAGYGFEIGHQF
					GAATCCAACCGTGGAAAGCTCCCGGTCCCG	TTATCC		YMDDLKLYART
					ACGGATTACATGCCCACTGGTGGAAAGCTT	TGGAGT		PAMLDSQIQVVS
					TACCGTCGGCCAAGAGGCTACTGGGTGAAC	GTTGTG		EVSEAMGLHLN
					TGGTGGTTGATTGGCTGACAACAGGTAAGG	GGACGA		LSKCAKAHYAP
					TTACCACTGGCTGGATGTGCCGAGGGAGGA	ATAGCT		HGAGGAQEAVE
					CAATCCTGATCCCTAAGAAGGGTGATAGGG	(SEQ ID		GAEGSRKGEIPI
					GCGACCCTTCTAATTACCGACCTATAACATG	NO: 1260)		LGLRSTYKYLG
					TTTAAATACATGTTATAAGGTGCTAACATCG			VEQRLLPMEVA
					GTAATGAATTCAGTCATTCTGAGTCACCTGA			LKEFEDKFMDR
					GCAGAGGCGAAGCTTTACCAATGAACCAGC			AETIFASELTWG
					GAGCAATGCGGAAACGCGAGTGGGGTTGCA			QMATAYNTIAIA
					CCCACGCTATGGTCCTTGACAGGGCCATGG			GLRYVYSNTNG
					TAATGGATGCAATGGCTCAAAAGAAACACT			ASPKLLEALKRA
					CATTAAGTGTGGCCTGGCTTGACTATCGTAA			ATLDTRIRDLLR
					AGCATACGATAGTGTGTCGCATGAATATAT			RHKCRFRNSFVE
					TCGCTGGGCGATTAACTCCGTGAATATACC			RLYIPRECGGYG
					CCGGAGTGTTCAGCTGACGCTCAAGAGGCT			LKSVEDTLRESI
					CATGAGTGACTGGGAGACACGCTTTGAGTC			LATWSYIATNPH
					GACGCAATGCCGGCCGAAGTTAAGGTCTGA			LAGQQYFFERL
					CAAAATGAAAGTGCTGAATGGCATCTTTCA			AARGKRTPMAD
					AGGTGACTCGTTATCACCAACCCTTTTTGTA			GVKILLDLGVEP
					TTATGCATAGCACCTATCAGCTACGCACTCA			QVDLKRRTVTV

					ATAAGGGTGTCGGCCAGTGTCAATCCTCAT			DGIVFEDPTKLH
					CCGGCTGGAGTGCAGGTTACGGTTTTGAGA			RYLVGKLLKAR
					TTGGACATCAGTTCTATATGGACGATCTTAA			TEARIRRWKEAS
					ACTGTACGCTAGGACGCCTGCGATGCTAGA			LAGRLVNDTSID
					CTCCCAAATCCAGGTGGTGTCTGAGGTGTC			MRLSCLWMKK
					GGAAGCAATGGGACTCCATTTGAATTTGAG			GFVSARNLRDA
					TAAATGTGCGAAAGCACATTATGCTCCGCA			LAVQEGSLLTR
					TGGGGCGGGCGGAGCGCAAGAAGCTGTGG			ACPALKGKGGQ
					AAGGTGCAGAAGGATCGAGGAAGGGAGAA			EVCRCCHAAPE
					ATCCCGATACTCGGGCTTCGAAGCACCTAT			TAEHITSACRY
					AAATATCTTGGAGTGGAACAACGACTCCTT			WLPSLYVERHD
					CCGATGGAAGTAGCCCTCAAAGAGTTCGAG			SVARNLYYVICC
					GATAAGTTTATGGATCGAGCAGAAACAATC			RYGITPVHYSNR
					TTCGCTAGCGAACTCACATGGGGGCAGATG			VSPLSENSQCRV
					GCCACAGCGTATAATACTATAGCTATCGCT			LWNMDMQTRT
					GGTCTACGATACGTCTATAGTAATACAAAT			PMKHRKPDIVV
					GGAGCATCACCAAAGCTTCTAGAAGCCCTG			FDLKREKILMFE
					AAAAGGGCGGCCACCTTAGACACGCGGATA			VSIAHASGLLKQ
					AGAGATCTTTTGAGGCGACATAAATGTCGG			REIKINRYTVNS
					TTTCGAAATAGCTTTGTCGAGAGGCTGTATA			EELPDETITPYPP
					TCCCTAGAGAATGCGGTGGATACGGGTTAA			GPNLAADLAAT
					AATCAGTGGAGGATACGCTGCGAGAGAGTA			YGWQVEFAPVV
					TCCTCGCTACGTGGAGTTACATCGCCACGA			VGTCGEHVPAV

					ACCCGCATTTGGCTGGACAACAGTATTTCTT CGAGAGGCTTGCAGCAAGAGGCAAGCGCA CCCCGATGGCGGACGGGGTAAAGATATTGT TGGATCTGGGAGTGGAACCCCAGGTGGACT TGAAGCGAAGGACGGTGACCGTCGACGGTA TAGTCTTCGAAGACCCGACCAAGCTTCATC GATACCTGGTGGGAAAGCTCTTAAAAGCAA GAACTGAGGCGAGGATTCGAAGATGGAAG GAAGCCAGCTTAGCTGGACGGTTGGTAAAT GACACAAGTATTGATATGCGACTATCATGC TTGTGGATGAAAAAAGGTTTTGTGAGTGCG AGGAACCTCAGAGATGCGCTTGCTGTGCAA GAGGGGAGTTTGCTTACTAGAGCATGCCCT GCTCTAAAGGGTAAAGGCGGCCAAGAAGTT TGCCGGTGTTGCCATGCAGCGCCGGAAACT GCAGAGCACATAACATCAGCCTGCCGCTAT TGGCTTCCAAGTCTCTACGTTGAGAGACAT GACTCGGTAGCAAGGAACCTCTATTACGTC ATATGCTGCCGCTACGGCATAACACCGGTG CATTACTCAAATAGGGTATCACCGCTCTCG GAGAATAGCCAATGCCGCGTTCTTTGGAAC ATGGATATGCAGACTCGGACGCCAATGAAG CATCGAAAGCCTGATATAGTCGTCTTTGATC			KEDLQRTLDLK PHQVEALLERIS RSAVIGTARVVR AHLACS (SEQ ID NO: 1505)

					TCAAGAGGGAAAAGATCCTCATGTTCGAAG TTTCGATAGCCCATGCCAGCGGGTTATTGA AACAGCGGGAAATCAAGATCAATCGGTATA CGGTGAACTCCGAAGAGTTGCCTGATGAGA CCATAACACCGTATCCGCCTGGGCCGAATT TGGCCGCTGACCTCGCTGCCACCTATGGTTG GCAGGTTGAATTTGCCCCAGTGGTGGTTGG CACGTGTGGTGAGCACGTACCAGCCGTCAA AGAGGACCTGCAAAGAACGTTGGATCTAAA ACCTCATCAAGTCGAAGCCCTTCTTGAAAG GATATCCCGATCGGCGGTGATCGGAACGGC TAGAGTAGTCCGAGCACACCTCGCCTGCTC CTAGTCGCTAAGGGGTCCGGAAATGGTCCG GTCCTGCGCTACCCGGTTCTGGTAGCACGTT CAAGCGCTCAATCGCCTGCCTTGTAGGCAG TCCATCTGTGGAAGTCGCGCTCTTGATACAG ATGTGGACGGATGGAAGCAGATGATAGAGC CGGTGACGGCCCTACTAGCCAAACGC (SEQ ID NO: 1137)			
R4	R4_Hmel	.		Heliconius melpomene	ATAAATAATAATAATAATAATAATAATAAG CCCCCTAAAATCCAACCATACGTCCGAGTC GAACATCTGATTCTCGTGGGGGGCGGACAC GTGAAAATAAATTACTTACACGGCTAACAT	ATAAAT AATAAT AATAAT AATAAT	TAAC TTATT GTCA GAAT	ITYTANMALVT LFMENMENKRY NLRPLPGGRRG ASGANAGCHSM

					GGCTCTGGTAACCCTTTTTATGGAGAATATG	AATAAG	TCCTT	RTVGDGGLSRR
					GAAAATAAAAGATATAATTTGCGGCCGCTA	CCCCCT	ACTA	VPLEKNVTAEQ
					CCCGGGGGTCGCCGGGGCGCATCTGGAGCT	AAAATC	GTAA	SSSPLTSSSSHSP
					AACGCTGGATGCCACAGCATGCGAACCGTT	CAACCA	TAAT	VSSIPSPSSTRTL
					GGCGACGGGGGACTGAGCAGGCGGGTTCCC	TACGTC	AATT	LNSPNSSPTSSHS
					CTCGAAAAAAATGTTACAGCCGAACAATCA	CGAGTC	ATCG	SLVIRSADVVQE
					TCCTCACCACTCACTTCTTCTTCCTCTCATTC	GAACAT	CTGA	ALANYPAPTAG
					ACCGGTATCAAGCATACCCTCACCATCCTC	CTGATT	AAAT	SIRARKKWTDI
					AACTCGCACCCTATTAAATTCCCCTAATTCA	CTCGTG	CTCC	MNRYIWRTYLII
					TCACCAACATCATCACATTCATCACTGGTTA	GGGGGC	ACCC	TKCETTLLNNYL
					TAAGGTCAGCAGATGTTGTGCAGGAGGCAC	GGACAC	AAAT	EPLHQEFSSKFP
					TTGCCAATTACCCTGCACCCACTGCGGGCTC	GTGAAA	ATTG	EMQVTRQRIGD
					TATTCGTGCGCGAAAAAAATGGACCGACAT	ATAA	CTTG	QRRAIIRNKLLS
					AATGAACCGTTATATTTGGCGCACATATCTT	(SEQ ID	GCTA	DDTLAQILIEVK
					ATTATAACAAAATGCGAAACCACGTTATTG	NO: 1261)	TATG	ELLQIGDQPLTQ
					AATAATTATTTAGAACCCCTACATCAAGAA		CTCG	NNIHSTQLSHSN
					TTTAGTAGTAAATTTCCCGAAATGCAGGTA		CAAT	TRIKWSNELNEE
					ACTAGACAACGTATAGGAGATCAACGTAGA		TTTTG	IVKCYFEVTLLE
					GCAATAATAAGAAATAAACTACTATCAGAT		GTTA	VNKTSYRKNLY
					GATACCTTAGCTCAAATTTTAATTGAAGTAA		ACGT	SLFISRNPHLSHL
					AAGAACTATTACAAATCGGTGATCAACCAC		ACCC	TEQRIADQRRLI
					TTACACAAAATAACATTCATTCAACTCAACT		CAAT	FMNKSVHNDRII
					TTCACACTCAAACACTAGAATAAAATGGTC		GATT	ELKREVEIKLAN

					AAACGAACTTAATGAAGAAATTGTCAAATG		TGGG	SNSLTKNITESN
					TTACTTTGAAGTCACCCTACTTGAAGTTAAT		AGAA	SPSSQTNEINDS
					AAGACTTCTTATCGTAAAAACCTGTACAGTT		CAAA	AYVQSNLQPVE
					TATTTATATCTAGAAATCCACACCTATCACA		AATG	PLDQHCINRHNL
					CCTTACAGAACAAAGAATAGCTGACCAAAG		GTAA	IEKHYVEQEFNN
					ACGACTAATATTTATGAACAAAAGTGTACA		AACT	ALIQFNNTNPET
					CAACGATAGAATTATTGAATTAAAAAGAGA		ATAA	RPYIPRQKSSRK
					AGTAGAAATTAAATTAGCTAACAGTAATTC		TAAT	FSQIVSFLNSEVL
					TTTAACAAAAAATATTACAGAGTCAAATTC		AATA	PKHLNNELDFN
					TCCTAGCAGTCAAACAAATGAAATAAATGA		ATTA	ALHNIIYTAXYT
					TTCTGCATATGTACAAAGTAATTTACAACCC		TATT	ASLCNGTKFSFI
					GTTGAACCACTAGACCAACATTGTATAAAT		AATA	DNYRPRNSKPS
					AGACACAATTTAATAGAAAAACATTATGTT		T (SEQ	WQRRLESRIDK
					GAACAAGAATTTAATAATGCTTTAATTCAA		ID NO:	YRLQIGRLTQYI
					TTTAACAATACAAATCCTGAAACCAGACCC		1384)	SGNRNRKILKTV
					TATATACCAAGACAAAAGTCATCGCGTAAA			EEIKTQYKIHSH
					TTTTCCCAAATTGTCAGTTTCTTAAATAGTG			HEEPNTELPHFL
					AAGTACTTCCAAAACATTTAAATAACGAAT			DTLKQKLNATS
					TAGATTTCAATGCTCTTCATAACATTATATA			NRLRRYLTCTK
					TACMGCTGYCTACACTGCATCTCTGTGTAA			RKQQNNTFVNN
					TGGAACAAAATTCAGTTTTATTGATAATTAC			EKHFYRTLSSTN
					CGACCTCGTAACAGCAAGCCATCGTGGCAA			QNTTTQLXEHPT
					AGAAGACTTGAAAGTAGAATAGACAAATAT			ENNLQQYWANI

					AGATTGCAAATAGGTAGACTTACACAATAT			WETSIEHNADA
					ATATCAGGAAATAGGAACCGTAAAATATTA			EWLNKIPDXEIN
					AAAACTGTAGAAGAAATAAAAACACAATAT			XMKFKDISIETF
					AAAATTCATTCACATCACGAAGAACCTAAT			NQIIQRTHNWK
					ACAGAATTACCACATTTCTTAGACACACTA			APGTDNIHNYW
					AAACAAAAATTAAATGCAACTTCGAACAGA			YKKLTCTHSLLL
					CTTAGAAGATATTTGACATGTACAAAACGG			KHINQFIQSPCTL
					AAGCAACAAAATAATACATTCGTTAATAAC			PLFITNGITYMLP
					GAAAAACATTTTTATAGAACTCTATCTAGC			KGLDPTNPANY
					ACAAACCAAAATACCACAACACAATTACW			RPITCLQTIYKIIT
					AGAACATCCAACGGAAAATAATTTGCAACA			ACITDIIYKHIDQ
					ATACTGGGCCAATATTTGGGAAACTTCAAT			NNILAEQQKGC
					AGAACAYAATGCAGACGCAGAATGGTTAA			RKNSQGCKEQL
					ATAAAATACCTGACKTAGAAATAAATKTAA			TIDAIVMKQAH
					TGAAATTCAAAGATATATCCATAGAAACAT			NKNXNTMYIDY
					TCAATCAAATCATCCAGCGAACACACAATT			RKAFDSVPHSW
					GGAAGGCACCGGGAACCGATAATATACACA			LLYILKKYKIHPI
					ACTAYTGGTACAAAAAGCTAACATGCACTC			LITFLSSVMLSW
					ATTCACTTCTTCTTAAACACATAAACCAGTT			KTRLKLINNNET
					CATACAATCACCTTGCACACTACCTTTATTT			LITDWIKIQRGIF
					ATAACAAACGGAATAACGTATATGTTACCG			QGDALSPLWFC
					AAAGGCTTAGATCCCACAAACCCAGCTAAC			LALNPLSELLNN
					TATAGGCCAATAACATGCCTCCAAACGATA			TNTGFKLKHNN

					TATAAAATTATCACTGCGTGTATTACTGACA			TYHIISHLMYMD
					TAATTTATAAACATATCGATCAAAATAACA			DIKLYASNNKEL
					TATTAGCTGAACAACAGAAAGGCTGTAGGA			KILADLTQSFST
					AAAATAGTCAAGGGTGTAAAGAGCAGCTTA			DIRMEFGIEKCK
					CTATAGACGCAATAGTTATGAAACAGGCAC			VHSIKRGKSQQ
					ATAATAAAAATAHTAATACRATGTATATAG			NTYILNTGEQIE
					ACTACAGAAAAGCATTTGATTCAGTTCCAC			SMDENSTYKYL
					ACAGTTGGTTACTATATATTTTAAAAAAGTA			GFQQAKQIQQK
					CAAGATTCACCCCATTCTAATAACATTTTTA			QTKIELTNKFKF
					AGCTCCGTYATGCTCTCATGGAAAACTAGA			RLNQILRSQLNS
					TTAAAATTGATTAACAATAATGAAACACTG			RNIIKAINTYAIPI
					ATTACTGATTGGATTAAAATACAACGAGGC			LTYSFAIINWSQ
					ATCTTTCAAGGTGATGCGCTGAGTCCATTAT			TDLSNLQRIINT
					GGTTCTGTCTAGCGTTAAACCCTCTCTCTGA			HMTTHRKHHPK
					ATTGCTTAATAATACCAATACAGGATTTAA			SCIQRLTISRLDG
					ACTAAAACATAATAACACATATCATATAAT			GRGLIDIRNLHN
					TTCACATTTAATGTACATGGACGATATCAA			NLVTKFRNYFY
					ACTATACGCCAGTAATAATAAAGAACTAAA			AKAEISELHKFI
					AATACTAGCCGACCTCACACAATCCTTCTCT			VNIDNKYTPLNL
					ACAGACATACGCATGGAATTTGGAATTGAA			NDRNIQLNQTLI
					AAATGCAAGGTACATTCAATTAAACGTGGC			TKQQKIEAWSL
					AAATCACAACAAAATACATATATACTTAAC			KSLHGRHLADL
					ACTGGTGAACAAATTGAATCAATGGACGAA			SQTHVDKVASN

					AATAGCACGTATAAATACCTTGGATTTCAA CAAGCTAAGCAAATACAACAAAAACAAAC CAAGATAGAACTAACCAATAAATTTAAATT TCGTTTGAATCAAATTTTAAGATCACAATTA AACTCACGTAATATTATAAAAGCCATCAAC ACATACGCAATACCGATCCTTACTTACTCTT TTGCTATAATAAACTGGTCTCAAACTGATTT ATCGAATCTGCAACGTATAATAAATACACA CATGACTACACATAGGAAACATCACCCTAA ATCATGTATCCAGCGTCTAACTATATCACGT CTGGATGGAGGACGAGGTCTCATTGATATT CGTAACCTACATAATAATCTAGTGACTAAA TTCCGAAATTACTTTTACGCTAAGGCTGAAA TTAGTGAACTACACAAATTCATAGTTAATAT AGATAATAAATATACACCACTTAATTTAAA CGATAGAAATATACAACTAAATCAAACTTT GATAACTAAGCAGCAAAAAATAGAAGCAT GGTCTTTAAAGTCACTACACGGTAGACATC TAGCCGATCTTAGCCAAACTCACGTCGACA AAGTTGCGTCGAACGAATGGCTACGACGGG GAGATCTCTTCCCAGAGACGGAAGCCTTCA TGATGGCTATTCAGGACCAAGTAATAGACA CTAGGAACTATCAAAAACACATCATAAAAA			EWLRRGDLFPE TEAFMMAIQDQ VIDTRNYQKHII KRPNMVNDLCR RCYSSPETIQHIT GACKTIVQTDY KHRHDQVAAIIH QHLAFKHSLITQ AQKTPYYKYSP QAILESTNFKLY WDRTIITDKTVH YNRPDILLHDKV KXSVYLIDIAIPN THNLASTFSNKI DKYTDLTIELKS QWKVQSVTTVP IVLSTTGVVPHT LHTSLETLGIHR LSYILLQKAAIL NTCRIVRKFLSS NN (SEQ ID NO: 1506)

					GACCAAATATGGTAAACGATTTATGCAGAA GATGCTATTCTAGTCCTGAAACAATCCAAC ACATTACCGGAGCCTGTAAAACCATAGTAC AAACCGATTACAAACATAGACATGATCAGG TCGCTGCAATTATTCATCAACACCTAGCATT TAAACATTCTCTCATCACACAAGCACAAAA AACTCCCTATTACAAATATTCACCTCAGGCT ATTTTAGAATCTACAAATTTTAAGCTGTATT GGGATCGTACAATCATCACAGATAAAACCG TACACTACAATAGACCGGATATTTTACTTCA TGATAAAGTWAAAAWTTCTGTATACTTGAT AGACATAGCCATTCCAAATACACACAATCT TGCATCCACTTTTTCAAACAAAATAGATAA GTATACAGATTTAACCATAGAGCTCAAAAG TCAATGGAAAGTCCAATCCGTCACTACAGT ACCAATAGTTCTTTCAACAACCGGAGTAGT ACCTCATACGCTACACACGAGTCTGGAAAC TTTGGGTATTCACCGGTTAAGTTATATTCTT TTACAGAAAGCCGCAATCTTAAATACATGC CGTATAGTTAGGAAGTTTTTATCATCAAATA ATTAACTTATTGTCAGAATTCCTTACTAGTA ATAATAATTATCGCTGAAAATCTCCACCCA AATATTGCTTGGCTATATGCTCGCAATTTTT			

					GGTTAACGTACCCCAATGATTTGGGAGAAC AAAAATGGTAAAACTATAATAATAATAATT ATATTAATAT (SEQ ID NO: 1138)			
R4	Rex6	.		Takifugu rubripes	TTCTATGCGCCTTATGCGACTGGATAGGCCA	TTCTAT	TAGA	MSGTXTDRVIPA
					GTGGTTTACGCCGCTGACTTTGGTGCGGAA	GCGCCT	GGAC	RTSPGSTRSASG
					GGTTGTCGGTTCGAATCCAGGCGAGCCCTT	TATGCG	CCGA	VGEPGPPDVKL
					AGGCAAGGCTCCTTACGCATATATGCCTAC	ACTGGA	GTCT	ATGTRHSWSRA
					ACCTCGGTATGAGCGGAACMWTAACTGAT	TAGGCC	GAAG	ENVVLMECYYG
					AGAGTCATACCGGCTCGGACGTCGCCCGGG	AGTGGT	GAAG	SNPSERGYMQR
					TCAACAAGGTCCGCGTCAGGGGTCGGGGAA	TTACGC	GAGG	MWEKWVLRNP
					CCAGGGCCACCTGATGTCAAATTGGCTACT	CGCTGA	CACC	TSSLTKKQLLAQ
					GGAACAAGGCATTCATGGAGCAGGGCAGA	CTTTGG	GCCC	CSNIRNKKLLSQ
					GAACGTGGTACTGATGGAGTGTTACTACGG	TGCGGA	AGGA	LEIDEARRCASP
					GAGTAACCCCAGCGAGAGAGGGTACATGCA	AGGTTG	GGGC	TVQICYGKGEPG
					GAGGATGTGGGAGAAATGGGTGCTTCGAAA	TCGGTT	GAGG	RQVSXGVISSSP
					CCCCACATCCTCACTGACTAAGAAGCAGCT	CGAATC	AAGA	PNIEIGYKAPMT
					CTTAGCTCAGTGTTCGAATATCCGCAATAA	CAGGCG	GATT	DGLGTRAADLR
					GAAGCTGCTATCACAGCTAGAGATTGACGA	AGCCCT	TTTTT	ERIMKSWGNST
					GGCACGGCGNTGTGCCTCACCCACGGTACA	TAGGCA	TTAT	TSLPRLTHKVPD
					AATATGCTACGGCAAGGGGGAGCCAGGAC	AGGCTC	ATAT	QSLLEDMNTAL
					GCCAGGTCAGCNTGGGGGTGATTTCATCAT	CTTACG	ATAT	STIPTTTITETNQ
					CACCCCCCAATATCGAGATTGGGTACAAAG	CATATA	ATAT	LMYAAATVILQ
					CCCCAATGACAGACGGACTCGGCACAAGGG	TGCCTA	ATAT	MLGYKMKSMN

					CAGCTGACCTGAGAGAGAGAATCATGAAGT CCTGGGGGAACTCAACTACCTCCCTGCCAA GGCTAACACACAAAGTACCAGACCAATCTC TACTGGAGGACATGAACACGGCACTGTCAA CCATCCCTACCACTACCATCACTGAGACCA ATCAGCTGATGTATGCAGCGGCAACAGTAA TCCTACAGATGCTTGGCTATAAGATGAAGA GCATGAACAGCCAGAAGGAGCAAATGGCC CCATGGAGGAGAAGGCTAGAGGCAAAAAT CATGGCGACACGGAGAGAAGTCAGCCTTCT AACAGAGCTGAGTAGAGGCGTGAATCTAAG GACAGAGCNGCCCAAGAAATATAACAAACT GTCCACAACTGAGGCACTGGAGACTGCTAA GCAAAGGCTCACAGCCCTGGCTACCCGACT AAAGAGGTACACAAGAGAAGTAGAGGCAA GGAGAATAAACAAGGTGTTCTCCACCAATC CAGCNAAGGTCTACTCTCAATGGCAGGGCA ACAAGATGACAACAGACCCCCCCAGGGCTG AGACTGAGCAATACTGGAAGAGTATCTGGG AGAAAGAGGCAACGCACAACACTANTGCCC AATGGCTGCAAGACCTACAGACAGAGCACA GCCAACTTCCAGAACAAGACCCAGTAGTCA TCACCTTAGCAGACATCCAAACAAGAGTGT	CACCTC GGT (SEQ ID NO: 1262)	ATA (SEQ ID NO: 1385)	SQKEQMAPWRR RLEAKIMATRRE VSLLTELSRGVN LRTEXPKKYNK LSTTEALETAKQ RLTALATRLKR YTREVEARRINK VFSTNPAKVYS QWQGNKMTTD PPRAETEQYWK SIWEKEATHNT XAQWLQDLQTE HSQLPEQDPVVI TLADIQTRVSK MKSWTAPGPDK IHAYWLKKLTA LHERLAAQMNQ LLTSGNHPEWL TQGRTVLIMKD PQKGTIPSNYRPI TCLSTTWKLLSG IIAAKISRHMDQ YMSRAQKGIGN

					CCAAAATGAAGAGCTGGACAGCACCAGGG			NTRGAKHQLLV
					CCCGATAAGATCCACGCCTACTGGCTTAAG			DRAIAQDCRTR
					AAGCTGACTGCACTCCATGAACGCCTGGCA			HTNLCTAWIDY
					GCACAGATGAACCAGCTGCTAACATCAGGG			KKAYDSMPHT
					AACCACCCAGAGTGGCTAACCCAGGGCCGG			WILECLKLYNIN
					ACAGTCCTCATAATGAAGGACCCCCAGAAG			RTLREFIQNSMK
					GGCACAATACCATCCAACTACCGGCCCATA			LWNTTLEANSK
					ACCTGCCTCAGCACCACATGGAAGCTCCTA			PIARVSIRCGIYQ
					TCAGGCATCATAGCGGCTAAGATCAGTAGG			GDALSPLLFCIG
					CACATGGATCAATACATGAGCAGAGCACAG			LNPLSQIITKSGY
					AAAGGCATAGGCAACAACACCAGAGGTGC			GYQFRSGTTVS
					CAAGCACCAGCTACTGGTCGACAGGGCAAT			HLLYMDDIKLY
					CGCCCAGGACTGTAGGACGCGGCACACCAA			AKNERDIDSLIH
					CCTGTGCACTGCCTGGATTGATTACAAGAA			LTRIYSKDIGMS
					AGCCTATGACTCAATGCCGCACACATGGAT			FGLDKCGRMISR
					ACTGGAGTGCCTAAAGCTGTATAACATCAA			RGKVIATDGVE
					CAGGACACTAAGAGAGTTCATCCAGAACTC			LPEGNITDVQDS
					CATGAAGCTGTGGAACACGACTCTGGAGGC			YKYLGIPQANG
					CAACTCAAAGCCAATTGCGCGGGTGAGCAT			NHEEAARRSAT
					CAGATGCGGCATATATCAAGGAGATGCCCT			AKYLQRLRQVL
					GTCCCCCCTGCTGTTCTGCATAGGCCTAAAC			KSQLNGKNKIQ
					CCCCTCAGCCAGATCATCACAAAGAGTGGC			AINTYALPVIRY
					TATGGGTACCAGTTCCGAAGTGGAACAACC			PAGIIPWPLEEIQ

					GTCAGCCACCTCCTCTACATGGATGACATC			ATDIKTRKLNG
					AAGCTGTATGCCAAGAACGAGCGTGACATC			KHKIQAINTYAL
					GACTCCCTGATTCACCTCACTAGGATCTACA			PVIRYPAGIIPWP
					GCAAAGACATCGGGATGTCATTCGGACTAG			LEEIQATDIKTR
					ATAAATGTGGGCGGATGATATCTAGAAGAG			KLLTMHGGFHP
					GAAAGGTGATCGCAACTGATGGGGTTGAAC			KSSVLRLYTKR
					TACCTGAAGGGAACATCACAGATGTGCAGG			KEGGRGLVSVR
					ACAGTTACAAATACCTGGGGATCCCGCAGG			TTVQEETTSLRE
					CAAATGGTAACCATGAGGAGGCAGCTAGGA			YIKKMAPTDRL
					GGTCAGCCACAGCCAAATACCTACAGAGGT			LSECLRQQKPTK
					TAAGGCAGGTCCTGAAAAGTCAGCTGAATG			EEEPEGLSWKD
					GTAAGAATAAGATCCAGGCCATAAACACCT			KPLHGMYHRQI
					ACGCCCTGCCAGTAATCAGATACCCTGCTG			EEVADIEKTYQ
					GCATAATACCCTGGCCACTGGAAGAGATAC			WLEKAGLKDST
					AAGCCACTGACATCAAGACAAGGAAGCTGA			EALLMAAQEQA
					ATGGTAAGCATAAGATCCAGGCCATAAACA			LSTRAIEARVYH
					CCTACGCCCTGCCAGTAATCAGATACCCTG			TRQDPRCRLCG
					CTGGCATAATACCCTGGCCACTGGAAGAGA			DAPETVQHITAG
					TACAAGCCACTGACATCAAGACAAGGAAGC			CKMLAGKAYM
					TCCTCACCATGCACGGAGGGTTTCACCCTA			ERHNQVAGIVY
					AGTCCAGCGTCCTGAGGCTGTACACGAAGC			RNICTEYGLEVP
					GAAAGGAAGGGGGCCGAGGACTAGTAAGT			GSRWETPPKVL
					GTCCGAACTACTGTCCAGGAGGAAACAACA			ENKQAKILWDF

					AGCCTCCGAGAATACATCAAGAAGATGGCC CCCACTGACCGACTGCTGAGTGAATGCCTC AGGCAACAAAAGCCCACCAAGGAGGAGGA GCCTGAGGGGCTATCATGGAAGGACAAGCC CCTGCATGGNATGTACCACCGACAAATTGA GGAAGTGGCTGATATCGAGAAAACATACCA GTGGCTGGAAAAGGCCGGACTGAAAGACA GCACAGAGGCACTACTCATGGCTGCACAAG AACAGGCCCTGAGCACCAGAGCAATAGAG GCCAGGGTCTACCATACCAGACAAGACCCC AGGTGCAGGCTGTGTGGAGACGCCCCTGAG ACAGTCCAGCACATCACAGCAGGGTGCAAG ATGTTGGCAGGCAAGGCATACATGGAGCGG CATAACCAGGTGGCTGGCATAGTGTACAGG AACATCTGCACTGAGTATGGACTGGAGGTC CCAGGGTCCAGGTGGGAGACACCCCCGAAA GTGCTGGAGAACAAGCAGGCCAAGATCCTG TGGGACTTCCAGATCCAGACTGACAAGATG GTGGTGGCCAACCAGCCTGACATAGTGGTG GTGGATAAACACCAGAAGACAGTGGTGGTG ATAGATGTAGCAATCCCAAGTGATAGCAAC ATCAGGAAGAAGGAACACGAGAAGCTGGA GAAGTACCAAGGGCTGAAGGAGGAGATGG			QIQTDKMVVAN QPDIVVVDKHQ KTVVVIDVAIPS DSNIRKKEHEKL EKYQGLKEEME RMWGMKATVV PVVIGTLGAVTP KLSRWLQQIPGT TSEISVQKSAVL GTAKILRRTLRL PGLW (SEQ ID NO: 1507)

					AGAGAATGTGGGGCATGAAGGCAACAGTG GTCCCAGTAGTGATTGGGACACTAGGGGCA GTAACACCCAAGCTGAGTAGATGGCTCCAA CAGATACCAGGAACCACATCAGAGATCTCT GTCCAGAAGAGCGCAGTCCTAGGAACAGCT AAGATCCTGCGCAGAACCCTCAGACTCCCA GGCCTCTGGTAGAGGACCCGAGTCTGAAGG AAGGAGGCACCGCCCAGGAGGGCGAGGAA GAGATTTTTTTTTATATATATATATATATATAT ATA (SEQ ID NO: 1139)			

Иллюстративные варианты осуществления цис-действующих систем Gene Writing

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.

Домен, связывающий нуклеиновую кислоту-матрицу

Полипептид Gene Writer, как правило, содержит области, способные ассоциировать с нуклеиновой кислотой-матрицей Gene Writer (например, РНК-матрицей). В некоторых вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу, представляет собой РНК-связывающий домен. В некоторых вариантах осуществления РНК-связывающий домен представляет собой модульный домен, который может ассоциировать с молекулами РНК, содержащими специфические сигнатуры, например, структурные мотивы, например, вторичные структуры, присутствующие в 3'-UTR в ретротранспозонах, не содержащих LTR. В других вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу (например, РНК-связывающий домен), содержится в домене обратной транскрипции, например, компонент, полученный из обратной транскриптазы, имеет известную сигнатуру для предпочтения РНК, например, вторичные структуры, присутствующие в 3'-UTR в ретротранспозонах, не содержащих LTR. В других вариантах осуществления домен, связывающий нуклеиновую кислоту-матрицу (например, РНК-связывающий домен), содержится в ДНК-связывающем домене. Например, в некоторых вариантах осуществления ДНК-связывающий домен представляет собой белок, ассоциированный с CRISPR, который распознает структуру нуклеиновой кислоты-матрицы (например, РНК-матрицы), содержащей gRNA. В некоторых вариантах осуществления gRNA представляет собой короткую синтетическую РНК, состоящую из каркасной последовательности, которая участвует в связывании белка, ассоциированного с CRISPR, и определяемой пользователем нацеливающей последовательности из ~ 20 нуклеотидов для геномной мишени. Структура полной gRNA была описана в Nishimasu et al. Cell 156, P935-949 (2014). gRNA (также называемая sgRNA или единичная направляющая RNA) состоит из последовательностей, полученных из crRNA и tracrRNA, соединенных искусственной тетрапетлей. Последовательность crRNA можно разделить на направляющую (20 нуклеотидов) и повторяющуюся (12 нуклеотидов) области, тогда как последовательность tracrRNA можно разделить на антиповтор (14 нуклеотидов) и три структуры "стебель-петля" tracrRNA (Nishimasu et al. Cell 156, P935-949 (2014)). На практике последовательности направляющих РНК, как правило, разрабатывают таким образом, чтобы они имели длину 17-24 нуклеотида (например, 19, 20 или 21 нуклеотид) и были комплементарными по отношению к последовательности нуклеиновой кислоты-мишени. Специализированные генераторы и алгоритмы для создания gRNA являются коммерчески доступными для применения в разработке эффективных направляющих РНК. В некоторых вариантах осуществления gRNA содержит два компонента РНК из нативной

системы CRISPR, например, crRNA и tracrRNA. Как хорошо известно из уровня техники, gRNA может также содержать химерную единичную направляющую РНК (sgRNA), содержащую последовательность как из tracrRNA (для связывания нуклеазы), так и из по меньшей мере одной crRNA (для направления нуклеазы к последовательности, предназначенной для редактирования/связывания). Также было продемонстрировано, что химически модифицированные sgRNA являются эффективными при использования с белками, ассоциированными с CRISPR; см., например, Hendel et al. (2015) Nature Biotechnol. 985-991. В некоторых вариантах осуществления gRNA содержит последовательность нуклеиновой кислоты, которая является комплементарной последовательности ДНК, ассоциированной с геном-мишенью. В некоторых вариантах осуществления полипептид содержит ДНК-связывающий домен, содержащий белок, ассоциированный с CRISPR, который ассоциирует с gRNA, что позволяет ДНК-связывающему домену связываться с геномной последовательностью ДНК-мишени. В некоторых вариантах осуществления gRNA содержится в нуклеиновой кислоте-матрице (например, РНК-матрице), поэтому ДНК-связывающий домен также является доменом, связывающим нуклеиновую кислоту-матрицу. В некоторых вариантах осуществления полипептид обладает функцией связывания РНК в нескольких доменах, например, может связывать структуру gRNA в ДНК-связывающем домене, ассоциированном с CRISPR, и структуру 3'-UTR в домене обратной транскрипции, полученном из ретротранспозона, не содержащего LTR.

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 Fok1, эндонуклеаза, подобная ферменту рестрикции 1 II типа (нуклеаза RLE-типа), или другая эндонуклеаза RLE-типа

(также известная как REL). В некоторых вариантах осуществления гетерологичная эндонуклеаза обладает никазной активностью и не образует двухнитевые разрывы. Аминокислотная последовательность эндонуклеазного домена системы Gene Writer, описанной в данном документе, может являться на по меньшей мере приблизительно 50%, на по меньшей мере приблизительно 60%, на по меньшей мере приблизительно 70%, на по меньшей мере приблизительно 80%, на по меньшей мере приблизительно 85%, на по меньшей мере приблизительно 90%, на по меньшей мере приблизительно 95%, на по меньшей мере приблизительно 96%, на по меньшей мере приблизительно 97%, на по меньшей мере приблизительно 98%, на по меньшей мере приблизительно 99% идентичной аминокислотной последовательности эндонуклеазного домена ретротранспозона, последовательность ДНК которого указана в таблице 1, 2 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 Fok1 или ее функциональный фрагмент. В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой резольвазу для структур Холлидея или ее гомолог, такой как фермент, разрешающий структуры Холлидея, из Sulfolobus solfataricus -- Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). В определенных вариантах осуществления гетерологичная эндонуклеаза представляет собой эндонуклеазу из большого фрагмента белка сплайсосомы, такого как Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). В определенных вариантах осуществления гетерологичная эндонуклеаза получена из белка, ассоциированного с CRISPR, например, Cas9. В определенных вариантах осуществления гетерологичная эндонуклеаза сконструирована таким образом, что она обладает только активностью расщепления однонитевой ДНК, например, только никазной активностью, например, является никазой Cas9. Например, полипептид Gene Writer, описанный в данном документе, может содержать домен обратной транскриптазы из ретротранспозона APE- или RLE-типа и эндонуклеазный домен, который содержит Fok1 или его функциональный фрагмент. В еще нескольких других вариантах осуществления 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 никазной активностью и не образует двухнитевые разрывы. В некоторых вариантах осуществления эндонуклеазный домен образует однонитевые разрывы с большей частотой, чем двухнитевые разрывы, например, по меньшей мере 90%, 95%, 96%, 97%, 98% или 99% разрывов являются однонитевыми разрывами, или менее 10%, 5%, 4%, 3%, 2% или 1% разрывов являются двухнитевыми разрывами. В некоторых вариантах осуществления

эндонуклеаза практически не образует двухнитевых разрывов. В некоторых вариантах осуществления эндонуклеаза не образует двухнитевые разрывы на выявляемых уровнях.

В некоторых вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию; например, в некоторых вариантах осуществления эндонуклеазный домен разрезает сайт-мишень геномной ДНК рядом с сайтом изменения в нити, которая будет удлинена записывающим доменом. В некоторых вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию, и не обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, не подлежащей редактированию. Например, в случае, если полипептид содержит эндонуклеазный домен, ассоциированный с CRISPR, обладающий никазной активностью, и не образующий двухнитевых разрывов, то в некоторых вариантах осуществления указанный эндонуклеазный домен, ассоциированный с CRISPR, вносит однонитевые разрывы в ДНК сайта-мишени в нити, содержащей PAM-сайт (например, и не вносит однонитевые разрывы в ДНК сайта-мишени в нити, не содержащей PAM-сайт).

В некоторых других вариантах осуществления эндонуклеазный домен обладает никазной активностью, которая обеспечивает внесение однонитевых разрывов в ДНК сайта-мишени в нити, подлежащей редактированию, и нити, не подлежащей редактированию. Не желая ограничиваться какой-либо теорией, полагают, что после полимеризации (например, обратной транскрипции) записывающего домена (например, домена RT) полипептида, описанного в данном документе, из последовательности гетерологичного объекта нуклеиновой кислоты-матрицы (например, РНК-матрицы), механизм репарации клеточной ДНК должен исправить однонитевой разрыв в нити ДНК, подлежащей редактированию. ДНК сайта-мишени теперь содержит две разные последовательности для нити ДНК, подлежащей редактированию: одна соответствует исходной геномной ДНК, и вторая соответствует той, которая полимеризована из последовательности гетерологичного объекта. Считается, что две разные последовательности уравновешиваются друг с другом, где сначала первая из них гибридизируется с нередактируемой нитью, а затем другая, и то, какую из них механизм репарации клеточной ДНК включает в свой исправленный сайт-мишень, считается случайным. Не желая ограничиваться какой-либо теорией, полагают, что введение дополнительного однонитевого разрыва в нередактируемую нить может привести к тому, что механизм репарации клеточной ДНК будет чаще использовать последовательность на основе последовательности гетерологичного объекта, чем исходную геномную последовательность. В некоторых вариантах осуществления дополнительный однонитевой разрыв расположен на расстоянии по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145 или 150 нуклеотидов с 5'- или 3'-стороны от модификации сайта-мишени (например, вставки, делеции или замены) или однонитевого разрыва в нити, подлежащей редактированию.

В качестве альтернативы или дополнительно, не желая ограничиваться какой-либо теорией, полагают, что дополнительный однонитевой разрыв в нередактируемой нити может способствовать синтезу второй нити. В некоторых вариантах осуществления, где полипептид Gene Writer вставил или заменил часть редактируемой нити, необходим синтез новой последовательности, соответствующей вставке/замене в нередактируемой нити.

В некоторых вариантах осуществления полипептид содержит один домен, обладающий эндонуклеазной активностью (например, один эндонуклеазный домен), и указанный домен вносит однонитевой разрыв как в нити, подлежащей редактированию, так и в нити, не подлежащей редактированию. Например, в таком варианте осуществления эндонуклеазный домен может являться эндонуклеазным доменом, ассоциированным с CRISPR, и нуклеиновая кислота-матрица (например, РНК-матрица) содержит gRNA, которая направляет внесение однонитевого разрыва в нити, подлежащей редактированию, и дополнительную gRNA, которая направляет внесение однонитевого разрыва в нити, не подлежащей редактированию. В некоторых вариантах осуществления полипептид содержит множество доменов, обладающих эндонуклеазной активностью, и первый эндонуклеазный домен вносит однонитевой разрыв в нить, подлежащую редактированию, и второй эндонуклеазный домен вносит однонитевой разрыв в нить, не подлежащую редактированию (необязательно первый эндонуклеазный домен не вносит (например, не может вносить) однонитевой разрыв в нить, не подлежащую редактированию, а второй эндонуклеазный домен не вносит (например, не может вносить) однонитевой разрыв в нить, подлежащую редактированию).

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в первую нить и вторую нить. В некоторых вариантах осуществления однонитевые разрывы в первой и второй нитях происходят в одном и том же положении в сайте-мишени, но в противоположных нитях. В некоторых вариантах осуществления однонитевой разрыв второй нити происходит в смещенном местоположении, например, выше или ниже первого однонитевого разрыва. В некоторых вариантах осуществления эндонуклеазный домен образует делецию сайта-мишени, если однонитевой разрыв второй нити находится выше однонитевого разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен образует дупликацию сайта-мишени, если однонитевой разрыв второй нити находится ниже однонитевого разрыва первой нити. В некоторых вариантах осуществления эндонуклеазный домен не образует дупликации и/или делеции, если однонитевые разрывы первой и второй нитей находятся в одном и том же положении в сайте-мишени (например, как описано в Gladyshev and Arkhipova Gene 2009, включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления эндонуклеазный домен обладает измененной активностью, зависящей от конформации белка или статуса связывания с РНК, например, что способствует внесению однонитевого разрыва в первую или вторую нить (например, как описано в Christensen et al. PNAS 2006; включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен содержит хоминг-эндонуклеазу или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен содержит мегануклеазу из семейств LAGLIDADG (SEQ ID NO: 1594), 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: 1594) с одной копией мотива LAGLIDADG (SEQ ID NO: 1594) обычно образуют гомодимеры, тогда как представители с двумя копиями мотива LAGLIDADG (SEQ ID NO: 1594) 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 или ДНК-связывающий домен (например, описанный в данном документе) содержит белок Cas, например, белок Cas9 *Streptococcus pyogenes* (SpCas9), или его функциональный фрагмент или вариант. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит модифицированный SpCas9. В вариантах осуществления модифицированный SpCas9 содержит модификацию, которая изменяет специфичность мотива, прилегающего к протоспейсеру (PAM). В вариантах осуществления PAM характеризуется специфичностью в отношении последовательности нуклеиновой кислоты 5'-NGT-3'. В вариантах осуществления модифицированный SpCas9 содержит одну или несколько аминокислотных замен, например, в одном или нескольких из положений L1111, D1135, G1218, E1219, A1322, R1335, например, выбранных из L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. В вариантах осуществления модифицированный SpCas9 содержит аминокислотную замену T1337R и одну или несколько дополнительных аминокислотных замен, например, выбранных из L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q и T1337M, или соответствующие им аминокислотные замены. В вариантах осуществления модифицированный SpCas9 содержит: (i) одну или несколько аминокислотных замен, выбранных из D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q и T1337; и (ii) одну или несколько аминокислотных замен, выбранных из L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337R, T1333K, T1337N, T1337N, T1337Q и T1337M, или соответствующие им аминокислотные замены.

В некоторых вариантах осуществления Gene Writer может содержать белок Cas, который указан в таблице 40. Предсказанные или валидированные мутации никазы для установления никазной активности в белке Cas, который показан в таблице 40,

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 40A. В некоторых вариантах осуществления белок Gene Writer из таблицы 3 слит с белком Cas из таблицы 40A.

Таблица 40A. Белки CRISPR/Cas, виды и мутации

Вариант	**Исходный хозяин**	**Последовательность белка**	**Мутация никазы**	**SEQ ID NO:**
Nme2 Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDE EENPIRLIDLGVRVFERAEVPKTGDSLAMAR RLARSVRRLTRRRAHRLLRARRLLKREGVL QAADFDENGLIKSLPNTPWQLRAAALDRKL TPLEWSAVLLHLIKHRGYLSQRKNEGETAD KELGALLKGVANNAHALQTGDFRTPAELAL NKFEKESGHIRNQRGDYSHTFSRKDLQAELI LLFEKQKEFGNPHVSGGLKEGIETLLMTQRP ALSGDAVQKMLGHCTFEPAEPKAAKNTYTA ERFIWLTKLNNLRILEQGSERPLTDTERATL MDEPYRKSKLTYAQARKLLGLEDTAFFKGL RYGKDNAEASTLMEMKAYHAISRALEKEGL KDKKSPLNLSSELQDEIGTAFSLFKTDEDITG RLKDRVQPEILEALLKHISFDKFVQISLKALR RIVPLMEQGKRYDEACAEIYGDHYGKKNTE EKIYLPPIPADEIRNPVVLRALSQARKVINGV VRRYGSPARIHIETAREVGKSFKDRKEIEKR QEENRKDREKAAAKFREYFPNFVGEPKSKDI LKLRLYEQQHGKCLYSGKEINLVRLNEKGY VEIDHALPFSRTWDDSFNNKVLVLGSENQN KGNQTPYEYFNGKDNSREWQEFKARVETSR FPRSKKQRILLQKFDEDGFKECNLNDTRYVN RFLCQFVADHILLTGKGKRRVFASNGQITNL LRGFWGLRKVRAENDRHHALDAVVVACST VAMQQKITRFVRYKEMNAFDGKTIDKETGK VLHQKTHFPQPWEFFAQEVMIRVFGKPDGK PEFEEADTPEKLRTLLAEKLSSRPEAVHEYV	N611A	1949

		TPLFVSRAPNRKMSGAHKDTLRSAKRFVKH NEKISVKRVWLTEIKLADLENMVNYKNGRE IELYEALKARLEAYGGNAKQAFDPKDNPFY KKGGQLVKAVRVEKTQESGVLLNKKNAYTI ADNGDMVRVDVFCKVDKKGKNQYFIVPIY AWQVAENILPDIDCKGYRIDDSYTFCFSLHK YDLIAFQKDEKSKVEFAYYINCDSSNGRFYL AWHDKGSKEQQFRISTQNLVLIQKYQVNEL GKEIRPCRLKKRPPVR		
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEESSP IRLIDVGVRTFERAEVAKTGESLALSRRLARS SRRLIKRRAERLKKAKRLLKAEKILHSIDEKL PINVWQLRVKGLKEKLERQEWAAVLLHLSK HRGYLSQRKNEGKSDNKELGALLSGIASNH QMLQSSEYRTPAEIAVKKFQVEEGHIRNQRG SYTHTFSRLDLLAEMELLFQRQAELGNSYTS TTLLENLTALLMWQKPALAGDAILKMLGKC TFEPSEYKAAKNSYSAERFVWLTKLNNLRIL ENGTERALNDNERFALLEQPYEKSKLTYAQ VRAMLALSDNAIFKGVRYLGEDKKTVESKT TLIEMKFYHQIRKTLGSAELKKEWNELKGNS DLLDEIGTAFSLYKTDDDICRYLEGKLPERV LNALLENLNFDKFIQLSLKALHQILPLMLQG QRYDEAVSAIYGDHYGKKSTETTRLLPTIPA DEIRNPVVLRTLTQARKVINAVVRLYGSPAR IHIETAREVGKSYQDRKKLEKQQEDNRKQR ESAVKKFKEMFPHFVGEPKGKDILKMRLYE LQQAKCLYSGKSLELHRLLEKGYVEVDHAL PFSRTWDDSFNNKVLVLANENQNKGNLTPY EWLDGKNNSERWQHFVVRVQTSGFSYAKK QRILNHKLDEKGFIERNLNDTRYVARFLCNF IADNMLLVGKGKRNVFASNGQITALLRHRW GLQKVREQNDRHHALDAVVVACSTVAMQQ KITRFVRYNEGNVFSGERIDRETGEIIPLHFPS PWAFFKENVEIRIFSENPKLELENRLPDYPQY	N605A	1950

		NHEWVQPLFVSRMPTRKMTGQGHMETVKS AKRLNEGLSVLKVPLTQLKLSDLERMVNRD REIALYESLKARLEQFGNDPAKAFAEPFYKK GGALVKAVRLEQTQKSGVLVRDGNGVADN ASMVRVDVFTKGGKYFLVPIYTWQVAKGIL PNRAATQGKDENDWDIMDEMATFQFSLCQ NDLIKLVTKKKTIFGYFNGLNRATSNINIKEH DLDKSKGKLGIYLEVGVKLAISLEKYQVDEL GKNIRPCRPTKRQHVR		
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL LEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRK WKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRELINDTLYSTRKDDKGNTLIVNNLNGL	N580A	1951

		YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYNNDLIKINGELYRVIGVNNDLLN RIEVNMIDITYREYLENMNDKRPPRIIKTIAS KTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL LEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRK WKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRKLINDTLYSTRKDDKGNTLIVNNLNGL YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ	N580A	1952

		KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNNDLLN RIEVNMIDITYREYLENMNDKRPPHIIKTIAS KTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
Sauri Cas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTR EVIDAGVRLFPEADSENNSNRRSKRGARRLK RRRIHRLNRVKDLLADYQMIDLNNVPKSTD PYTIRVKGLREPLTKEEFAIALLHIAKRRGLH NISVSMGDEEQDNELSTKQQLQKNAQQLQD KYVCELQLERLTNINKVRGEKNRFKTEDFV KEVKQLCETQRQYHNIDDQFIQQYIDLVSTR REYFEGPGNGSPYGWDGDLLKWYEKLMGR CTYFPEELRSVKYAYSADLFNALNDLNNLV VTRDDNPKLEYYEKYHIIENVFKQKKNPTLK QIAKEIGVQDYDIRGYRITKSGKPQFTSFKLY HDLKNIFEQAKYLEDVEMLDEIAKILTIYQD EISIKKALDQLPELLTESEKSQIAQLTGYTGT HRLSLKCIHIVIDELWESPENQMEIFTRLNLK PKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQS IKVINAVINRFGLPEDIIIELAREKNSKDRRKF INKLQKQNEATRKKIEQLLAKYGNTNAKYM IEKIKLHDMQEGKCLYSLEAIPLEDLLSNPTH YEVDHIIPRSVSFDNSLNNKVLVKQSENSKK GNRTPYQYLSSNESKISYNQFKQHILNLSKA KDRISKKKRDMLLEERDINKFEVQKEFINRN LVDTRYATRELSNLLKTYFSTHDYAVKVKTI NGGFTNHLRKVWDFKKHRNHGYKHHAED ALVIANADFLFKTHKALRRTDKILEQPGLEV NDTTVKVDTEEKYQELFETPKQVKNIKQFR DFKYSHRVDKKPNRQLINDTLYSTREIDGET YVVQTLKDLYAKDNEKVKKLFTERPQKILM YQHDPKTFEKLMTILNQYAEAKNPLAAYYE	N588A	1953

		DKGEYVTKYAKKGNGPAIHKIKYIDKKLGS YLDVSNKYPETQNKLVKLSLKSFRFDIYKCE QGYKMVSIGYLDVLKKDNYYYIPKDKYEAE KQKKKIKESDLFVGSFYYNDLIMYEDELFRV IGVNSDINNLVELNMVDITYKDFCEVNNVTG EKRIKKTIGKRVVLIEKYTTDILGNLYKTPLP KKPQLIFKRGEL		
Sauri Cas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTR EVIDAGVRLFPEADSENNSNRRSKRGARRLK RRRIHRLNRVKDLLADYQMIDLNNVPKSTD PYTIRVKGLREPLTKEEFAIALLHIAKRRGLH NISVSMGDEEQDNELSTKQQLQKNAQQLQD KYVCELQLERLTNINKVRGEKNRFKTEDFV KEVKQLCETQRQYHNIDDQFIQQYIDLVSTR REYFEGPGNGSPYGWDGDLLKWYEKLMGR CTYFPEELRSVKYAYSADLFNALNDLNNLV VTRDDNPKLEYYEKYHIIENVFKQKKNPTLK QIAKEIGVQDYDIRGYRITKSGKPQFTSFKLY HDLKNIFEQAKYLEDVEMLDEIAKILTIYQD EISIKKALDQLPELLTESEKSQIAQLTGYTGT HRLSLKCIHIVIDELWESPENQMEIFTRLNLK PKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQS IKVINAVINRFGLPEDIIIELAREKNSKDRRKF INKLQKQNEATRKKIEQLLAKYGNTNAKYM IEKIKLHDMQEGKCLYSLEAIPLEDLLSNPTH YEVDHIIPRSVSFDNSLNNKVLVKQSENSKK GNRTPYQYLSSNESKISYNQFKQHILNLSKA KDRISKKKRDMLLEERDINKFEVQKEFINRN LVDTRYATRELSNLLKTYFSTHDYAVKVKTI NGGFTNHLRKVWDFKKHRNHGYKHHAED ALVIANADFLFKTHKALRRTDKILEQPGLEV NDTTVKVDTEEKYQELFETPKQVKNIKQFR DFKYSHRVDKKPNRKLINDTLYSTREIDGET YVVQTLKDLYAKDNEKVKKLFTERPQKILM YQHDPKTFEKLMTILNQYAEAKNPLAAYYE	N588A	1954

		DKGEYVTKYAKKGNGPAIHKIKYIDKKLGS YLDVSNKYPETQNKLVKLSLKSFRFDIYKCE QGYKMVSIGYLDVLKKDNYYYIPKDKYEAE KQKKKIKESDLFVGSFYKNDLIMYEDELFRV IGVNSDINNLVELNMVDITYKDFCEVNNVTG EKHIKKTIGKRVVLIEKYTTDILGNLYKTPLP KKPQLIFKRGEL		
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSK KFKVLGNTNRKSIKKNLMGALLFDSGETAE ATRLKRTARRRYTRRKNRIRYLQEIFANEMA KLDDSFFQRLEESFLVEEDKKNERHPIFGNL ADEVAYHRNYPTIYHLRKKLADSPEKADLR LIYLALAHIIKFRGHFLIEGKLNAENSDVAKL FYQLIQTYNQLFEESPLDEIEVDAKGILSARL SKSKRLEKLIAVFPNEKKNGLFGNIIALALGL TPNFKSNFDLTEDAKLQLSKDTYDDDLDELL GQIGDQYADLFSAAKNLSDAILLSDILRSNSE VTKAPLSASMVKRYDEHHQDLALLKTLVRQ QFPEKYAEIFKDDTKNGYAGYVGADKKLRK RSGKLATEEEFYKFIKPILEKMDGAEELLAK LNRDDLLRKQRTFDNGSIPHQIHLKELHAILR RQEEFYPFLKENREKIEKILTFRIPYYVGPLA RGNSRFAWLTRKSEEAITPWNFEEVVDKGA SAQSFIERMTNFDEQLPNKKVLPKHSLLYEY FTVYNELTKVKYVTERMRKPEFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EIIGVEDRFNASLGTYHDLLKIIKDKDFLDNE ENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRHYTGWGRLSRKMINGIR DKQSGKTILDFLKSDGFSNRNFMQLIHDDSL TFKEEIEKAQVSGQGDSLHEQIADLAGSPAIK KGILQTVKIVDELVKVMGHKPENIVIEMARE NQTTTKGLQQSRERKKRIEEGIKELESQILKE NPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTR	N872A	1955

		SVENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIKR QLVETRQITKHVARILDSRMNTKRDKNDKPI REVKVITLKSKLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYG DYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANGEIRKRPLIETNGETGEVV WNKEKDFATVRKVLAMPQVNIVKKTEVQT GGFSKESILSKRESAKLIPRKKGWDTRKYGG FGSPTVAYSILVVAKVEKGKAKKLKSVKVL VGITIMEKGSYEKDPIGFLEAKGYKDIKKELI FKLPKYSLFELENGRRRMLASAKELQKANE LVLPQHLVRLLYYTQNISATTGSNNLGYIEQ HREEFKEIFEKIIDFSEKYILKNKVNSNLKSSF DEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITG LYETRTDLSQLGGD		
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR	N863A	1956

		KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT	N863A	1957

		PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI RPKRNSDKLIARKKDWDPKKYGGFVSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASARFLQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPRAFKYF		

		DTTIDRKVYRSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAER TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY	N863A	1958

		DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI RPKRNSDKLIARKKDWDPKKYGGFLWPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAKQLQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTRLGAPRAFK YFDTTIDPKQYRSTKEVLDATLIHQSITGLYE TRIDLSQLGGD		
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNS RIFPAAQAENNLVRRTNRQGRRLARRKKHR RVRLNRLFEESGLITDFTKISINLNPYQLRVK GLTDELSNEELFIALKNMVKHRGISYLDDAS DDGNSSVGDYAQIVKENSKQLETKTPGQIQL ERYQTYGQLRGDFTVEKDGKKHRLINVFPT SAYRSEALRILQTQQEFNPQITDEFINRYLEIL TGKRKYYHGPGNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLL NDLNNLTVPTETKKLSKEQKNQIINYVKNEK AMGPAKLFKYIAKLLSCDVADIKGYRIDKSG KAEIHTFEAYRKMKTLETLDIEQMDRETLDK LAYVLTLNTEREGIQEALEHEFADGSFSQKQ VDELVQFRKANSSIFGKGWHNFSVKLMMEL IPELYETSEEQMTILTRLGKQKTTSSSNKTKY IDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKE YGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGH KQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATA NQEKGQRTPYQALDSMDDAWSFRELKAFV RESKTLSNKKKEYLLTEEDISKFDVRKKFIER NLVDTRYASRVVLNALQEHFRAHKIDTKVS	N622A	1959

		VVRGQFTSQLRRHWGIEKTRDTYHHHAVD ALIIAASSQLNLWKKQKNTLVSYSEDQLLDI ETGELISDDEYKESVFKAPYQHFVDTLKSKE FEDSILFSYQVDSKFNRKISDATIYATRQAKV GKDKADETYVLGKIKDIYTQDGYDAFMKIY KKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKG NGPEIKSLKYYDSKLGNHIDITPKDSNNKVV LQSVSPWRADVYFNKTTGKYEILGLKYADL QFEKGTGTYKISQEKYNDIKKKEGVDSDSEF KFTLYKNDLLLVKDTETKEQQLFRFLSRTMP KQKHYVELKPYDKQKFEGGEALIKVLGNVA NSGQCKKGLGKSNISIYKVRTDVLGNQHIIK NEGDKPKLDF		
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGV RTFEKAENPKNGEALAVPRREARSSRRRLRR KKHRIERLKHMFVRNGLAVDIQHLEQTLRS QNEIDVWQLRVDGLDRMLTQKEWLRVLIHL AQRRGFQSNRKTDGSSEDGQVLVNVTENDR LMEEKDYRTVAEMMVKDEKFSDHKRNKNG NYHGVVSRSSLLVEIHTLFETQRQHHNSLAS KDFELEYVNIWSAQRPVATKDQIEKMIGTCT FLPKEKRAPKASWHFQYFMLLQTINHIRITN VQGTRSLNKEEIEQVVNMALTKSKVSYHDT RKILDLSEEYQFVGLDYGKEDEKKKVESKET IIKLDDYHKLNKIFNEVELAKGETWEADDY DTVAYALTFFKDDEDIRDYLQNKYKDSKNR LVKNLANKEYTNELIGKVSTLSFRKVGHLSL KALRKIIPFLEQGMTYDKACQAAGFDFQGIS KKKRSVVLPVIDQISNPVVNRALTQTRKVIN ALIKKYGSPETIHIETARELSKTFDERKNITK DYKENRDKNEHAKKHLSELGIINPTGLDIVK YKLWCEQQGRCMYSNQPISFERLKESGYTE VDHIIPYSRSMNDSYNNRVLVMTRENREKG NQTPFEYMGNDTQRWYEFEQRVTTNPQIKK	N607A	1960

		EKRQNLLLKGFTNRRELEMLERNLNDTRYIT KYLSHFISTNLEFSPSDKKKKVVNTSGRITSH LRSRWGLEKNRGQNDLHHAMDAIVIAVTSD SFIQQVTNYYKRKERRELNGDDKFPLPWKFF REEVIARLSPNPKEQIEALPNHFYSEDELADL QPIFVSRMPKRSITGEAHQAQFRRVVGKTKE GKNITAKKTALVDISYDKNGDFNMYGRETD PATYEAIKERYLEFGGNVKKAFSTDLHKPKK DGTKGPLIKSVRIMENKTLVHPVNKGKGVV YNSSIVRTDVFQRKEKYYLLPVYVTDVTKG KLPNKVIVAKKGYHDWIEVDDSFTFLFSLYP NDLIFIRQNPKKKISLKKRIESHSISDSKEVQE IHAYYKGVDSSTAAIEFIIHDGSYYAKGVGV QNLDCFEKYQVDILGNYFKVKGEKRLELET SDSNHKGKDVNSIKSTSR		
cCas9 -v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL	N580A	1961

		LEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRK WKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRKLINDTLYSTRKDDKGNTLIVNNLNGL YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNSDKN NLIEVNMIDITYREYLENMNDKRPPHIIKTIA SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
cCas9 -v17	Staphylococ cus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL LEERDINRFSVQKDFINRNLVDTRYATRGLM	N580A	1962

		NLLRSYFRVNNLDVKVKSINGGFTSFLRRK WKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRKLINDTLYSTRKDDKGNTLIVNNLNGL YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNNSTRN IVELNMIDITYREYLENMNDKRPPHIIKTIAS KTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL LEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRK	N580A	1963

		WKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRKLINDTLYSTRKDDKGNTLIVNNLNGL YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNSDDRN IIELNMIDITYREYLENMNDKRPPHIIKTIASK TQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNE VEEDTGNELSTKEQISRNSKALEEKYVAELQ LERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY GLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQE GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSS SDSKISYETFKKHILNLAKGKGRISKTKKEYL LEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRK WKFKKERNKGYKHHAEDALIIANADFIFKE	N580A	1964

		WKKLDKAKKVMENQMFEEKQAESMPEIET EQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRKLINDTLYSTRKDDKGNTLIVNNLNGL YDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTV KNLDVIKKENYYEVNSKCYEEAKKLKKISN QAEFIASFYKNDLIKINGELYRVIGVNNNRL NKIELNMIDITYREYLENMNDKRPPHIIKTIA SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKT LSLVSHIHDSGLDPDEIKSAVTRLASSGIARR TRRLYRRKRRRLQQLDKFIQRQGWPVIELED YSDPLYPWKVRAELAASYIADEKERGEKLS VALRHIARHRGWRNPYAKVSSLYLPDGPSD AFKAIREEIKRASGQPVPETATVGQMVTLCE LGTLKLRGEGGVLSARLQQSDYAREIQEICR MQEIGQELYRKIIDVVFAAESPKGSASSRVG KDPLQPGKNRALKASDAFQRYRIAALIGNLR VRVDGEKRILSVEEKNLVFDHLVNLTPKKEP EWVTIAEILGIDRGQLIGTATMTDDGERAGA RPPTHDTNRSIVNSRIAPLVDWWKTASALEQ HAMVKALSNAEVDDFDSPEGAKVQAFFAD LDDDVHAKLDSLHLPVGRAAYSEDTLVRLT RRMLSDGVDLYTARLQEFGIEPSWTPPTPRI GEPVGNPAVDRVLKTVSRWLESATKTWGA PERVIIEHVREGFVTEKRAREMDGDMRRRA ARNAKLFQEMQEKLNVQGKPSRADLWRYQ SVQRQNCQCAYCGSPITFSNSEMDHIVPRAG QGSTNTRENLVAVCHRCNQSKGNTPFAIWA KNTSIEGVSVKEAVERTRHWVTDTGMRSTD FKKFTKAVVERFQRATMDEEIDARSMESVA WMANELRSRVAQHFASHGTTVRVYRGSLT AEARRASGISGKLKFFDGVGKSRLDRRHHAI	H573A (альтернативная)	1965

		DAAVIAFTSDYVAETLAVRSNLKQSQAHRQ EAPQWREFTGKDAEHRAAWRVWCQKMEK LSALLTEDLRDDRVVVMSNVRLRLGNGSAH KETIGKLSKVKLSSQLSVSDIDKASSEALWC ALTREPGFDPKEGLPANPERHIRVNGTHVYA GDNIGLFPVSAGSIALRGGYAELGSSFHHAR VYKITSGKKPAFAMLRVYTIDLLPYRNQDLF SVELKPQTMSMRQAEKKLRDALATGNAEY LGWLVVDDELVVDTSKIATDQVKAVEAELG TIRRWRVDGFFSPSKLRLRPLQMSKEGIKKE SAPELSKIIDRPGWLPAVNKLFSDGNVTVVR RDSLGRVRLESTAHLPVTWKVQ		
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIF TKVENPKTGESLALPRRLARSARKRLARRK ARLNHLKHLIANEFKLNYEDYQSFDESLAK AYKGSLISPYELRFRALNELLSKQDFARVILH IAKRRGYDDIKNSDDKEKGAILKAIKQNEEK LANYQSVGEYLYKEYFQKFKENSKEFTNVR NKKESYERCIAQSFLKDELKLIFKKQREFGFS FSKKFEEEVLSVAFYKRALKDFSHLVGNCSF FTDEKRAPKNSPLAFMFVALTRIINLLNNLK NTEGILYTKDDLNALLNEVLKNGTLTYKQT KKLLGLSDDYEFKGEKGTYFIEFKKYKEFIK ALGEHNLSQDDLNEIAKDITLIKDEIKLKKAL AKYDLNQNQIDSLSKLEFKDHLNISFKALKL VTPLMLEGKKYDEACNELNLKVAINEDKKD FLPAFNETYYKDEVTNPVVLRAIKEYRKVLN ALLKKYGKVHKINIELAREVGKNHSQRAKIE KEQNENYKAKKDAELECEKLGLKINSKNIL KLRLFKEQKEFCAYSGEKIKISDLQDEKMLEI DHIYPYSRSFDDSYMNKVLVFTKQNQEKLN QTPFEAFGNDSAKWQKIEVLAKNLPTKKQK RILDKNYKDKEQKNFKDRNLNDTRYIARLV LNYTKDYLDFLPLSDDENTKLNDTQKGSKV HVEAKSGMLTSALRHTWGFSAKDRNNHLH	N582A	1966

		HAIDAVIIAYANNSIVKAFSDFKKEQESNSAE LYAKKISELDYKNKRKFFEPFSGFRQKVLDK IDEIFVSKPERKKPSGALHEETFRKEEEFYQS YGGKEGVLKALELGKIRKVNGKIVKNGDMF RVDIFKHKKTNKFYAVPIYTMDFALKVLPN KAVARSKKGEIKDWILMDENYEFCFSLYKD SLILIQTKDMQEPEFVYYNAFTSSTVSLIVSK HDNKFETLSKNQKILFKNANEKEVIAKSIGIQ NLKVFEKYIVSALGEVTKAEFRQREDFKK		
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGV RIFDRAENPQTGESLALPRRLARSARRRLRR RKHRLERIRRLVIREGILTKEELDKLFEEKHEI DVWQLRVEALDRKLNNDELARVLLHLAKR RGFKSNRKSERSNKENSTMLKHIEENRAILSS YRTVGEMIVKDPKFALHKRNKGENYTNTIA RDDLEREIRLIFSKQREFGNMSCTEEFENEYI TIWASQRPVASKDDIEKKVGFCTFEPKEKRA PKATYTFQSFIAWEHINKLRLISPSGARGLTD EERRLLYEQAFQKNKITYHDIRTLLHLPDDT YFKGIVYDRGESRKQNENIRFLELDAYHQIR KAVDKVYGKGKSSSFLPIDFDTFGYALTLFK DDADIHSYLRNEYEQNGKRMPNLANKVYD NELIEELLNLSFTKFGHLSLKALRSILPYMEQ GEVYSSACERAGYTFTGPKKKQKTMLLPNIP PIANPVVMRALTQARKVVNAIIKKYGSPVSI HIELARDLSQTFDERRKTKKEQDENRKKNET AIRQLMEYGLTLNPTGHDIVKFKLWSEQNG RCAYSLQPIEIERLLEPGYVEVDHVIPYSRSL DDSYTNKVLVLTRENREKGNRIPAEYLGVG TERWQQFETFVLTNKQFSKKKRDRLLRLHY DENEETEFKNRNLNDTRYISRFFANFIREHLK FAESDDKQKVYTVNGRVTAHLRSRWEFNK NREESDLHHAVDAVIVACTTPSDIAKVTAFY QRREQNKELAKKTEPHFPQPWPHFADELRA RLSKHPKESIKALNLGNYDDQKLESLQPVFV	N605A	1967

		SRMPKRSVTGAAHQETLRRYVGIDERSGKIQ TVVKTKLSEIKLDASGHFPMYGKESDPRTYE AIRQRLLEHNNDPKKAFQEPLYKPKKNGEP GPVIRTVKIIDTKNQVIPLNDGKTVAYNSNIV RVDVFEKDGKYYCVPVYTMDIMKGILPNKA IEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIE LPREKTVKTAAGEEINVKDVFVYYKTIDSAN GGLELISHDHRFSLRGVGSRTLKRFEKYQVD VLGNIYKVRGEKRVGLASSAHSKPGKTIRPL QSTRD		
iSpy MacCas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRKLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDNL LAQIGDQYADLFLAAKNLSDAILLSDILRVN TEITKAPLSASMIKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLKREDLLRKQ RTFDNGSIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDI VLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQK	N863A	1968

		GQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVET RQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEIQTVGQNGG LFDDNPKSPLEVTPSKLVPLKKELNPKKYGG YQKPTTAYPVLLITDTKQLIPISVMNKKQFE QNPVKFLRDRGYQQVGKNDFIKLPKYTLVD IGDGIKRLWASSKEIHKGNQLVVSKKSQILL YHAHHLDSDLSNDYLQNHNQQFDVLFNEIIS FSKKCKLGKEHIQKIENVYSNKKNSASIEEL AESFIKLLGFTQLGATSPFNFLGVKLNQKQY KGKKDYILPCTEGTLIRQSITGLYETRVDLSK IGEDSGGSGGSKRTADGSEFES		
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDE EENPIRLIDLGVRVFERAEVPKTGDSLAMAR RLARSVRRLTRRRAHRLLRTRRLLKREGVL QAANFDENGLIKSLPNTPWQLRAAALDRKL TPLEWSAVLLHLIKHRGYLSQRKNEGETAD KELGALLKGVAGNAHALQTGDFRTPAELAL NKFEKESGHIRNQRSDYSHTFSRKDLQAELI LLFEKQKEFGNPHVSGGLKEGIETLLMTQRP ALSGDAVQKMLGHCTFEPAEPKAAKNTYTA ERFIWLTKLNNLRILEQGSERPLTDTERATL MDEPYRKSKLTYAQARKLLGLEDTAFFKGL RYGKDNAEASTLMEMKAYHAISRALEKEGL KDKKSPLNLSPELQDEIGTAFSLFKTDEDITG RLKDRIQPEILEALLKHISFDKFVQISLKALRR IVPLMEQGKRYDEACAEIYGDHYGKKNTEE	N611A	1969

		KIYLPPIPADEIRNPVVLRALSQARKVINGVV RRYGSPARIHIETAREVGKSFKDRKEIEKRQE ENRKDREKAAAKFREYFPNFVGEPKSKDILK LRLYEQQHGKCLYSGKEINLGRLNEKGYVEI DHALPFSRTWDDSFNNKVLVLGSENQNKGN QTPYEYFNGKDNSREWQEFKARVETSRFPR SKKQRILLQKFDEDGFKERNLNDTRYVNRFL CQFVADRMRLTGKGKKRVFASNGQITNLLR GFWGLRKVRAENDRHHALDAVVVACSTVA MQQKITRFVRYKEMNAFDGKTIDKETGEVL HQKTHFPQPWEFFAQEVMIRVFGKPDGKPE FEEADTLEKLRTLLAEKLSSRPEAVHEYVTP LFVSRAPNRKMSGQGHMETVKSAKRLDEG VSVLRVPLTQLKLKDLEKMVNREREPKLYE ALKARLEAHKDDPAKAFAEPFYKYDKAGN RTQQVKAVRVEQVQKTGVWVRNHNGIADN ATMVRVDVFEKGDKYYLVPIYSWQVAKGIL PDRAVVQGKDEEDWQLIDDSFNFKFSLHPN DLVEVITKKARMFGYFASCHRGTGNINIRIH DLDHKIGKNGILEGIGVKTALSFQKYQIDEL GKEIRPCRLKKRPPVR		
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSK KFKVLGNTNRKSIKKNLMGALLFDSGETAE ATRLKRTARRRYTRRKNRIRYLQEIFANEMA KLDDSFFQRLEESFLVEEDKKNERHPIFGNL ADEVAYHRNYPTIYHLRKKLADSPEKADLR LIYLALAHIIKFRGHFLIEGKLNAENSDVAKL FYQLIQTYNQLFEESPLDEIEVDAKGILSARL SKSKRLEKLIAVFPNEKKNGLFGNIIALALGL TPNFKSNFDLTEDAKLQLSKDTYDDDLDELL GQIGDQYADLFSAAKNLSDAILLSDILRSNSE VTKAPLSASMVKRYDEHHQDLALLKTLVRQ QFPEKYAEIFKDDTKNGYAGYVGIGIKHRKR TTKLATQEEFYKFIKPILEKMDGAEELLAKL NRDDLLRKQRTFDNGSIPHQIHLKELHAILR	N872A	1970

		RQEEFYPFLKENREKIEKILTFRIPYYVGPLA RGNSRFAWLTRKSEEAITPWNFEEVVDKGA SAQSFIERMTNFDEQLPNKKVLPKHSLLYEY FTVYNELTKVKYVTERMRKPEFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EIIGVEDRFNASLGTYHDLLKIIKDKDFLDNE ENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRHYTGWGRLSRKMINGIR DKQSGKTILDFLKSDGFSNRNFMQLIHDDSL TFKEEIEKAQVSGQGDSLHEQIADLAGSPAIK KGILQTVKIVDELVKVMGHKPENIVIEMARE NQTTTKGLQQSRERKKRIEEGIKELESQILKE NPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTR SVENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIKR QLVETRQITKHVARILDSRMNTKRDKNDKPI REVKVITLKSKLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYG DYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANGEIRKRPLIETNGETGEVV WNKEKDFATVRKVLAMPQVNIVKKTEVQT GGFSKESILSKRESAKLIPRKKGWDTRKYGG FGSPTVAYSILVVAKVEKGKAKKLKSVKVL VGITIMEKGSYEKDPIGFLEAKGYKDIKKELI FKLPKYSLFELENGRRRMLASATELQKANEL VLPQHLVRLLYYTQNISATTGSNNLGYIEQH REEFKEIFEKIIDFSEKYILKNKVNSNLKSSFD EQFAVSDSILLSNSFVSLLKYTSFGASGGFTF LDLDVKQGRLRYQTVTEVLDATLIYQSITGL YETRTDLSQLGGD		
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSK KFKVLGNTNRKSIKKNLMGALLFDSGETAE ATRLKRTARRRYTRRKNRIRYLQEIFANEMA KLDDSFFQRLEESFLVEEDKKNERHPIFGNL	N872A	1971

		ADEVAYHRNYPTIYHLRKKLADSPEKADLR LIYLALAHIIKFRGHFLIEGKLNAENSDVAKL FYQLIQTYNQLFEESPLDEIEVDAKGILSARL SKSKRLEKLIAVFPNEKKNGLFGNIIALALGL TPNFKSNFDLTEDAKLQLSKDTYDDDLDELL GQIGDQYADLFSAAKNLSDAILLSDILRSNSE VTKAPLSASMVKRYDEHHQDLALLKTLVRQ QFPEKYAEIFKDDTKNGYAGYVGADKKLRK RSGKLATEEEFYKFIKPILEKMDGAEELLAK LNRDDLLRKQRTFDNGSIPHQIHLKELHAILR RQEEFYPFLKENREKIEKILTFRIPYYVGPLA RGNSRFAWLTRKSEEAITPWNFEEVVDKGA SAQSFIERMTNFDEQLPNKKVLPKHSLLYEY FTVYNELTKVKYVTERMRKPEFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EIIGVEDRFNASLGTYHDLLKIIKDKDFLDNE ENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRHYTGWGRLSRKMINGIR DKQSGKTILDFLKSDGFSNANFMQLIHDDSL TFKEEIEKAQVSGQGDSLHEQIADLAGSPAIK KGILQTVKIVDELVKVMGHKPENIVIEMARE NQTTTKGLQQSRERKKRIEEGIKELESQILKE NPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTR SVENRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSEADKAGFIKR QLVETRQITKHVARILDSRMNTKRDKNDKPI REVKVITLKSKLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYG DYKVYDVRKMIAKSEQEIGKATAKRFFYSNI MNFFKTEVKLANGEIRKRPLIETNGETGEVV WNKEKDFATVRKVLAMPQVNIVKKTEVQT GGFSKESILSKRESAKLIPRKKGWDTRKYGG FGSPTVAYSILVVAKVEKGKAKKLKSVKVL VGITIMEKGSYEKDPIGFLEAKGYKDIKKELI		

		FKLPKYSLFELENGRRRMLASAKELQKANE LVLPQHLVRLLYYTQNISATTGSNNLGYIEQ HREEFKEIFEKIIDFSEKYILKNKVNSNLKSSF DEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITG LYETRTDLSQLGGD		
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMVKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGIIPHQIHLGELHAILRRQGDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDI VLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRLRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLIT	N863A	1972

		QRKFDNLTKAERGGLSELDKAGFIKRQLVET RQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKGNSDKLIARKKDWDPKKYGGFNSPTA AYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAGVLHKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGVPAAF KYFDTTIDKKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGD		
SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMVKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGIIPHQIHLGELHAILRRQGDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRF	N863A	1973

		NASLGTYHDLLKIIKDKDFLDNEENEDILEDI VLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRLRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVET RQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKGNSDKLIARKKDWDPKKYGGFNSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASASVLHKGNELALPSK YVNFLYLASHYEKLKGSSEDNKQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGASAAF KYFDTTIGRKLYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGD		
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL	N863A	1974

		AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMVKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGIIPHQIHLGELHAILRRQGDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDI VLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRLRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVET RQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKGNSDKLIARKKDWDPKKYGGFNSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAGVLQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTINRKQYNTTKEVLDATLIRQSITGL		

		YETRIDLSQLGGD		
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK	N863A	1975

		TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV	N863A	1976

		VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASARELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADAQLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTFKQKQYRSTKEVLDATLIHQSITGLYET RIDLSQLGGD		
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD	N863A	1977

		NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFLWPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAKQLQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKQYRSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD	N863A	1978

		EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFVSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL		

		FELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKQYRSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLITQ	N863A	1979

		RKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVY DVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFVSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASARELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHK HYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKEYRSTKEVLDATLIHQSITGLYETR IDLSQLGGD		
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDTKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKLYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGIIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEKVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGDQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN	N863A	1980

		ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFIQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVIT LKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAY SVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGVLQKGNELALPSKYVN FLYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFD TTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD		
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFI QLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDTKLQLSKDTYDDDLDNLL	N863A	1981

		AQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKLYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGIIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEKVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGDQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFIQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVIT LKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESIR PKRNSDKLIARKKDWDPKKYGGFVSPTVAY SVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASARFLQKGNELALPSKYVN FLYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYNK HRDKPIREQAENIIHLFTLTNLGAPRAFKYFD TTIDRKVYRSTKEVLDATLIHQSITGLYETRI		

		DLSQLGGD		
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTL	N622A	1982

		PKQKHYVELKPYDKQKFEGGEALIKVLGNV ANGGQCIKGLAKSNISIYKVRTDVLGNQHII KNEGDKPKLDF		
St1Cas9-LMG1831	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNS RIFPAAQAENNLVRRTNRQGRRLARRKKHR RVRLNRLFEESGLITDFTKISINLNPYQLRVK GLTDELSNEELFIALKNMVKHRGISYLDDAS DDGNSSVGDYAQIVKENSKQLETKTPGQIQL ERYQTYGQLRGDFTVEKDGKKHRLINVFPT SAYRSEALRILQTQQEFNPQITDEFINRYLEIL TGKRKYYHGPGNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLL NDLNNLTVPTETKKLSKEQKNQIINYVKNEK AMGPAKLFKYIAKLLSCDVADIKGYRIDKSG KAEIHTFEAYRKMKTLETLDIEQMDRETLDK LAYVLTLNTEREGIQEALEHEFADGSFSQKQ VDELVQFRKANSSIFGKGWHNFSVKLMMEL IPELYETSEEQMTILTRLGKQKTTSSSNKTKY IDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKE YGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGH KQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATA NQEKGQRTPYQALDSMDDAWSFRELKAFV RESKTLSNKKKEYLLTEEDISKFDVRKKFIER NLVDTRYASRVVLNALQEHFRAHKIDTKVS VVRGQFTSQLRRHWGIEKTRDTYHHHAVD ALIIAASSQLNLWKKQKNTLVSYSEEQLLDI ETGELISDDEYKESVFKAPYQHFVDTLKSKE FEDSILFSYQVDSKFNRKISDATIYATRQAKV GKDKKDETYVLGKIKDIYTQDGYDAFMKIY KKDKSKFLMYRHDPQTFEKVIEPILENYPNK QMNEKGKEVPCNPFLKYKEEHGYIRKYSKK GNGPEIKSLKYYDSKLLGNPIDITPENSKNKV VLQSLKPWRTDVYFNKNTGKYEILGLKYAD	N622A	1983

		LQFEKKTGTYKISQEKYNGIMKEEGVDSDSE FKFTLYKNDLLLVKDTETKEQQLFRFLSRTM PNVKYYVELKPYSKDKFEKNESLIEILGSAD KSGRCIKGLGKSNISIYKVRTDVLGNQHIIKN EGDKPKLDF		
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNS RIFPAAQAENNLVRRTNRQGRRLARRKKHR RVRLNRLFEESGLITDFTKISINLNPYQLRVK GLTDELSNEELFIALKNMVKHRGISYLDDAS DDGNSSVGDYAQIVKENSKQLETKTPGQIQL ERYQTYGQLRGDFTVEKDGKKHRLINVFPT SAYRSEALRILQTQQEFNPQITDEFINRYLEIL TGKRKYYHGPGNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLL NDLNNLTVPTETKKLSKEQKNQIINYVKNEK AMGPAKLFKYIAKLLSCDVADIKGYRIDKSG KAEIHTFEAYRKMKTLETLDIEQMDRETLDK LAYVLTLNTEREGIQEALEHEFADGSFSQKQ VDELVQFRKANSSIFGKGWHNFSVKLMMEL IPELYETSEEQMTILTRLGKQKTTSSSNKTKY IDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKE YGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGH KQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATA NQEKGQRTPYQALDSMDDAWSFRELKAFV RESKTLSNKKKEYLLTEEDISKFDVRKKFIER NLVDTRYASRVVLNALQEHFRAHKIDTKVS VVRGQFTSQLRRHWGIEKTRDTYHHHAVD ALIIAASSQLNLWKKQKNTLVSYSEDQLLDI ETGELISDDEYKESVFKAPYQHFVDTLKSKE FEDSILFSYQVDSKFNRKISDATIYATRQAKV GKDKADETYVLGKIKDIYTQDGYDAFMKIY KKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKG	N622A	1984

		NGPEIKSLKYYDSKLGNHIDITPKDSNNKVV LQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEF KFTLYKNDLLLLKDSENGEQILLRFTSRNDT SKHYVELKPYNRQKFEGSEYLIKSLGTVAKG GQCIKGLGKSNISIYKVRTDVLGNQHIIKNEG DKPKLDF		
St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNS RIFPAAQAENNLVRRTNRQGRRLARRKKHR RVRLNRLFEESGLITDFTKISINLNPYQLRVK GLTDELSNEELFIALKNMVKHRGISYLDDAS DDGNSSVGDYAQIVKENSKQLETKTPGQIQL ERYQTYGQLRGDFTVEKDGKKHRLINVFPT SAYRSEALRILQTQQEFNPQITDEFINRYLEIL TGKRKYYHGPGNEKSRTDYGRYRTSGETLD NIFGILIGKCTFYPDEFRAAKASYTAQEFNLL NDLNNLTVPTETKKLSKEQKNQIINYVKNEK AMGPAKLFKYIAKLLSCDVADIKGYRIDKSG KAEIHTFEAYRKMKTLETLDIEQMDRETLDK LAYVLTLNTEREGIQEALEHEFADGSFSQKQ VDELVQFRKANSSIFGKGWHNFSVKLMMEL IPELYETSEEQMTILTRLGKQKTTSSSNKTKY IDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKE YGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGH KQLATKIRLWHQQGERCLYTGKTISIHDLIN NSNQFEVDHILPLSITFDDSLANKVLVYATA NQEKGQRTPYQALDSMDDAWSFRELKAFV RESKTLSNKKKEYLLTEEDISKFDVRKKFIER NLVDTRYASRVVLNALQEHFRAHKIDTKVS VVRGQFTSQLRRHWGIEKTRDTYHHHAVD ALIIAASSQLNLWKKQKNTLVSYSEDQLLDI ETGELISDDEYKESVFKAPYQHFVDTLKSKE FEDSILFSYQVDSKFNRKISDATIYATRQAKV GKDKADETYVLGKIKDIYTQDGYDAFMKIY	N622A	1985

		KKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKG NGPEIKSLKYYDSKLGNHIDITPKDSNNKVV LQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEF KFTLYKNDLLLLKDSENGEQILLRFTSRNDT SKHYVELKPYNRQKFEGSEYLIKSLGTVVKG GRCIKGLGKSNISIYKVRTDVLGNQHIIKNEG DKPKLDF		

В таблице 40B представлены параметры для определения необходимых компонентов для конструирования gRNA и/или РНК-матриц в применении вариантов Cas, перечисленных в таблице 3A, для Gene Writing. Ярус указывает на предпочтительные варианты Cas, если они доступны для использования в данном локусе. Место разреза указывает на валидированные или предсказанные требования для мотива, прилегающего к протоспейсеру (PAM), валидированное или предсказанное местоположение сайта разреза (относительно наиболее вышерасположенного основания сайта PAM). gRNA для данного фермента может быть собрана посредством соединения последовательностей crRNA, тетрапетли и tracrRNA и дополнительного добавления спейсера на 5'-конце с длиной в пределах от спейсера (min) до спейсера (max), которая соответствует протоспейсеру в сайте-мишени. Дополнительно, предсказанное местоположение однонитевого разреза ssDNA в мишени является важным для разработки 3'-концевой области РНК-матрицы, которую необходимо отжечь с последовательностью непосредственно с 5'-стороны от однонитевого разреза с целью инициации обратной транскрипции с праймированием мишенью.

Таблица 40B. Параметры для определения необходимых компонентов для конструирования gRNA и/или РНК-матриц в применении вариантов Cas, перечисленных в таблице 40A, для Gene Writing

Вариант	PAM	Разрез	Ярус	Спейсер (min)	Спейсер (max)	crRNA	Тетрапетля	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGCTCCCTTTCTCATTTCG (SEQ ID NO: 386)	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCC

								CTTAAAGCT TCTGCTTTA AGGGGCAT CGTTTA (SEQ ID NO: 387)
PpnCas9	NNNNRTT		1	21	24	GTTGTAG CTCCCTT TTTCATT TCGC (SEQ ID NO: 388)	GAAA	GCGAAATG AAAAACGT TGTTACAAT AAGAGATG AATTTCTCG CAAAGCTCT GCCTCTTGA AATTTCGGT TTCAAGAG GCATCTTTT T (SEQ ID NO: 389)
SauCas9	NNGRR;NNGRRT	-3	1	21	23	GTTTTAG TACTCTG (SEQ ID NO: 390)	GAAA	CAGAATCTA CTAAAACA AGGCAAAA TGCCGTGTT TATCTCGTC AACTTGTTG GCGAGA (SEQ ID NO: 391)
SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAG TACTCTG TAAT (SEQ ID NO: 392)	GAAA	ATTACAGA ATCTACTAA AACAAGGC AAAATGCC GTGTTTATC TCGTCAACT TGTTGGCGA

								GA (SEQ ID NO: 393)
Sauri Cas9	NNGG	-3	1	21	21	GTTTTAG TACTCTG (SEQ ID NO: 394)	GAAA	CAGAATCTA CTAAAACA AGGCAAAA TGCCGTGTT TATCTCGTC AACTTGTTG GCGAGATTT TT (SEQ ID NO: 395)
Sauri Cas9-KKH (SEQ ID NO: 401)	NNRG	-3	1	21	21	GTTTTAG TACTCTG (SEQ ID NO: 396)	GAAA	CAGAATCTA CTAAAACA AGGCAAAA TGCCGTGTT TATCTCGTC AACTTGTTG GCGAGATTT TT (SEQ ID NO: 397)
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAG AGCTA (SEQ ID NO: 398)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 399)
SpyCas9	NGG	-3	1	20	20	GTTTTAG AGCTA (SEQ ID NO: 400)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG

								AGTCGGTGC (SEQ ID NO: 401)
SpyCas9-NG	NG (NGG=NGA=NGT>NGC)	-3	1	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 402)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 403)
SpyCas9-SpRY	NRN>NYN	-3	1	20	20	GTTTTAGAGCTA (SEQ ID NO: 404)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 405)
St1Cas9	NNAGAAW>NNAGGAW=NNGGAAW	-3	1	20	20	GTCTTTGTACTCTG (SEQ ID NO: 407)	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT (SEQ ID NO: 406)
BlatCas9	NNNNCNAA>NNNNCNDD>NNNNC	-3	1	19	23	GCTATAGTTCCTTACT (SEQ	GAAA	GGTAAGTTGCTATAGTAAGGGCAACAGACCCGAG

						ID NO: 408)		GCGTTGGG GATCGCCTA GCCCGTGTT TACGGGCTC TCCCCATAT TCAAAATA ATGACAGA CGAGCACCT TGGAGCATT TATCTCCGA GGTGCT (SEQ ID NO: 409)
cCas9 -v16	NNVACT; NNVATG M;NNVAT T;NNVGC T;NNVGT G;NNVGT T	-3	2	21	21	GUCUUAG UACUCUG (SEQ ID NO: 410)	GAAA	CAGAAUCU ACUAAGAC AAGGCAAA AUGCCGUG UUUAUCUC GUCAACUU GUUGGCGA GAUUUUUU U (SEQ ID NO: 411)
cCas9 -v17	NNVRRN	-3	2	21	21	GUCUUAG UACUCUG (SEQ ID NO: 412)	GAAA	CAGAAUCU ACUAAGAC AAGGCAAA AUGCCGUG UUUAUCUC GUCAACUU GUUGGCGA GAUUUUUU U (SEQ ID NO: 413)
cCas9 -v21	NNVACT; NNVATG	-3	2	21	21	GUCUUAG UACUCUG	GAAA	CAGAAUCU ACUAAGAC

	M;NNVATT;NNVGCT;NNVGTG;NNVGTT					(SEQ ID NO: 414)		AAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU (SEQ ID NO: 415)
cCas9-v42	NNVRRN	-3	2	21	21	GUCUUAGUACUCUG (SEQ ID NO: 416)	GAAA	CAGAAUCUACUAAGACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU (SEQ ID NO: 417)
CdiCas9	NNRHHHY;NNRAAAY		2	22	22	ACUGGGGUUCAG (SEQ ID NO: 418)	GAAA	CUGAACCUCAGUAAGCAUUGGCUCGUUUCCAAUGUUGAUUGCUCCGCCGGUGCUCCUUAUUUUUAAGGGCGCCGGC (SEQ ID NO: 419)
CjeCas9	NNNNRYAC	-3	2	21	23	GTTTTAGTCCCT (SEQ ID NO: 420)	GAAA	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAA

								AACCGCTTTTTT (SEQ ID NO: 421)
GeoCas9	NNNNCRAA		2	21	23	GUCAUAGUUCCCCUGA (SEQ ID NO: 421)	GAAA	UCAGGGUUACUAUGAUAAGGGCUUUCUGCCUAAGGCAGACUGACCCGCGGCGUUGGGGAUCGCCUGUCGCCCGCUUUUGGCGGGCAUUCCCCAUCCUU(SEQ ID NO: 422)
iSpyMacCas9	NAAN	-3	2	19	21	GTTTTAGAGCTA (SEQ ID NO: 423)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 44)
NmeCas9	NNNNGAYT;NNNNGYTT;NNNNGAYA;NNNNGTCT	-3	2	20	24	GTTGTAGCTCCCTTTCTCATTTCG (SEQ ID NO: 425)	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTA

								AGGGGCAT CGTTTA (SEQ ID NO: 426)
ScaCas9	NNG	-3	2	20	20	GTTTTAG AGCTA (SEQ ID NO: 427)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 428)
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAG AGCTA (SEQ ID NO: 429)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 430)
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAG AGCTATG CTG (SEQ ID NO: 431)	GAAA	CAGCATAG CAAGTTTAA ATAAGGCT AGTCCGTTA TCAACTTGA AAAAGTGG CACCGAGTC GGTGC (SEQ ID NO: 432)
SpyCas9-3var-	NRTH	-3	2	20	20	GTTTAAG AGCTATG CTG (SEQ	GAAA	CAGCATAG CAAGTTTAA ATAAGGCT AGTCCGTTA

NRTH						ID NO: 433)		TCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 434)
SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 435)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 436)
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 437)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 438)
SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 439)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 440)
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCG

						(SEQ ID NO: 441)		TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 442)
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAG AGCTA (SEQ ID NO: 443)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 444)
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAG AGCTA (SEQ ID NO: 445)	GAAA	TAGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTGC (SEQ ID NO: 446)
SpyCas9-xCas	NG;GAA; GAT	-3	2	20	20	GTTTAAG AGCTATG CTG (SEQ ID NO: 447)	GAAA	CAGCATAG CAAGTTTAA ATAAGGCT AGTCCGTTA TCAACTTGA AAAAGTGG CACCGAGTC GGTGC (SEQ ID NO: 448)
SpyCas9-	NG	-3	2	20	20	GTTTAAG AGCTATG	GAAA	CAGCATAG CAAGTTTAA

xCas-NG						CTG (SEQ ID NO: 449)		ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 450)
St1Cas9-CNRZ1066	NNACAA	-3	2	20	20	GTCTTTGTACTCTG (SEQ ID NO: 451)	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT (SEQ ID NO: 452)
St1Cas9-LMG1831	NNGCAA	-3	2	20	20	GTCTTTGTACTCTG (SEQ ID NO: 453)	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT (SEQ ID NO: 454)
St1Cas9-MTH17CL396	NNAAAA	-3	2	20	20	GTCTTTGTACTCTG (SEQ ID NO: 455)	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT

								(SEQ ID NO: 456)
St1Cas9-TH1477	NNGAAA	-3	2	20	20	GTCTTTGTACTCTG (SEQ ID NO: 457)	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT (SEQ ID NO: 458)

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит домен Cas, например, домен Cas9. В вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит нуклеазоактивный домен Cas, Cas-никазный домен (nCas) или нуклеазонеактивный домен Cas (dCas). В вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит нуклеазоактивный домен Cas9, Cas9-никазный домен (nCas9) или нуклеазонеактивный домен Cas9 (dCas9). В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит домен Cas9 из Cas9 (например, dCas9 и nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas9 (например, dCas9 и nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas9 из S. pyogenes или S. thermophilus или ее функциональный фрагмент. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит последовательность Cas9, например, описанную в Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; включенном в данный документ посредством ссылки. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит HNH-нуклеазный субдомен и/или субдомен RuvC1 из Cas, например, Cas9, например, описанной в данном документе, или ее вариант. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h или Cas12i. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит полипептид Cas (например, фермент) или его функциональный фрагмент. В вариантах осуществления полипептид Cas (например, фермент) выбран из Cas1, Cas1B, Cas2, Cas3,

Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (например, Csn1 или Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, эффекторных белков Cas II типа, эффекторных белков Cas V типа, эффекторных белков Cas VI типа, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, сверхточного варианта Cas9 (HypaCas9), их гомологов, их модифицированных или сконструированных вариантов и/или их функциональных фрагментов. В вариантах осуществления Cas9 содержит одну или несколько замен, например, выбранных из H840A, D10A, P475A, W476A, N477A, D1125A, W1126A и D1127A. В вариантах осуществления Cas9 содержит одну или несколько мутаций в положениях, выбранных из D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987, например, одну или несколько замен, выбранных из D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A и/или D986A. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит последовательность Cas (например, Cas9) из Corynebacterium ulcerans, Corynebacterium diphtheria, Spiroplasma syrphidicola, Prevotella intermedia, Spiroplasma taiwanense, Streptococcus iniae, Belliella baltica, Psychroflexus torquis, Streptococcus thermophilus, Listeria innocua, Campylobacter jejuni, Neisseria meningitidis, Streptococcus pyogenes или Staphylococcus aureus или ее фрагмент или вариант.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит домен Cpf1, например, содержащий одну или несколько замен, например, в положениях D917, E1006A, D1255 или любой их комбинации, например, выбранных из D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A и D917A/E1006A/D1255A.

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит spCas9, spCas9-VRQR (SEQ ID NO: 1696), spCas9-VRER (SEQ ID NO: 1697), xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER (SEQ ID NO: 1698), spCas9-LRKIQK (SEQ ID NO: 1699) или spCas9-LRVSQL (SEQ ID NO: 1700).

В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен (например, описанный в данном документе) содержит аминокислотную последовательность, указанную в таблице 37 ниже, или аминокислотную последовательность, характеризующуюся по меньшей мере 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью последовательности с ней. В некоторых вариантах осуществления эндонуклеазный домен или ДНК-связывающий домен содержит аминокислотную последовательность, которая 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: 1595)
Иллюстративный линкерный мотив	$(GGGS)_n$ (SEQ ID NO: 1596)
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 Streptococcus iniae	Эталонная последовательность в NCBI: NC_021314.1
Cas9 из Belliella baltica	Эталонная последовательность в NCBI: NC_018010.1
Cas9 из Psychroflexus torquis	Эталонная последовательность в NCBI: NC_018721.1
Cas9 из Streptococcus thermophilus	Эталонная последовательность в NCBI: YP_820832.1
Cas9 из Listeria innocua	Эталонная последовательность в NCBI: NP_472073.1
Cas9 из Campylobacter jejuni	Эталонная последовательность в NCBI: YP_002344900.1
Cas9 из Neisseria meningitidis	Эталонная последовательность в NCBI: YP_002342100.1
dCas9 (D10A и H840A)	
Каталитически неактивная Cas9 (dCas9)	
Никаза Cas9 (nCas9)	
Каталитически активная Cas9	
CasY	((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1 белок CasY, ассоциированный с CRISPR [некультивируемая бактерия из группы Parcubacteria])
CasX	uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53
CasX	>tr|F0NH53|F0NH53_SULIR белок, ассоциированный с CRISPR, Casx OS=Sulfolobus islandicus (штамм REY15A) GN=SiRe_0771 PE=4 SV=1
CasX Deltaproteobacteria	
Cas12b/C2c1	((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG эндонуклеаза C2c1, ассоциированная с CRISPR OS=Alicyclobacillus acidoterrestris (штамм ATCC 49025/DSM 3922/ CIP 106132/NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1)
BhCas12b (Bacillus hisashii)	Эталонная последовательность в NCBI: WP_095142515
BvCas12b (Bacillus sp. V3-13)	Эталонная последовательность в NCBI: WP_101661451.1

Cpf1 Francisella novicida дикого типа	
Cpf1 D917A Francisella novicida	
Cpf1 E1006A Francisella novicida	
Cpf1 D1255A Francisella novicida	
Cpf1 D917A/E1006A Francisella novicida	
Cpf1 D917A/D1255A Francisella novicida	
Cpf1 E1006A/D1255A Francisella novicida	
Cpf1 D917A/E1006A Francisella novicida	
SaCas9	
SaCas9n	
PAM-связывающий SpCas9	
PAM-связывающий SpCas9n	
PAM-связывающий SpCas9-EQR	
PAM-связывающий SpCas9-VQR	
PAM-связывающий SpCas9-VRER	
PAM-связывающий SpCas9-VRQR	
SpyMacCas9	

В некоторых вариантах осуществления полипептид Gene Writing имеет эндонуклеазный домен, содержащий никазу Cas9, например, Cas9 H840A. В вариантах осуществления Cas9 H840A имеет следующую аминокислотную последовательность.

Никаза Cas9 (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR
KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII
HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 1597)

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, например, RT M-MLV дикого типа, например, содержащий следующую последовательность.

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS
IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI
SGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQ
GTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK
TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTA
PALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR
MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD
RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSS
LLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSR

YAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1598)

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, например, RT M-MLV, например, содержащий следующую последовательность:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL
DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG
QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQ
ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWP
PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL
LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT
DGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVY
TDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH
SAEARGNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 1556)

В некоторых вариантах осуществления полипептид Gene Writing содержит домен RT из ретровирусной обратной транскриптазы, содержащий последовательность из аминокислот 659-1329 NP_057933. В вариантах осуществления полипептид Gene Writing дополнительно содержит одну дополнительную аминокислоту на N-конце последовательности из аминокислот 659-1329 NP_057933, например, как показано ниже:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQD**
LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF
EWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDL
LLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLT
EARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPD
QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN
ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP
LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK
MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLS
IIHCPGHQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 1599)

Сердцевинная RT (жирный шрифт), аннотирована согласно вышеуказанному

РНКаза H (подчеркнуто), аннотирована согласно вышеуказанному

В вариантах осуществления полипептид Gene Writing дополнительно содержит одну дополнительную аминокислоту на C-конце последовательности из аминокислот 659-1329 NP_057933. В вариантах осуществления полипептид Gene Writing содержит домен РНКазы H1 (например, аминокислоты 1178-1318 NP_057933).

В некоторых вариантах осуществления домен ретровирусной обратной транскриптазы , например, RT M-MLV, может содержать одну или несколько мутаций относительно последовательности дикого типа, которые могут улучшать характеристики RT, например, термостабильность, процессивность и/или связывание с матрицей. В некоторых вариантах осуществления домен RT M-MLV содержит, по сравнению с приведенной выше последовательностью M-MLV (WT), одну или несколько мутаций, например, выбранных из D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, например, комбинацию мутаций, таких как D200N, L603W и T330P, необязательно дополнительно включая T306K и W313F. В некоторых вариантах осуществления RT M-MLV, используемая в данном документе, содержит мутации D200N, L603W, T330P, T306K и W313F. В вариантах осуществления мутантная RT M-MLV содержит следующую аминокислотную последовательность.

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS
IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI
SGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQ
GTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK
TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTA
PALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR
MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD
RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSS
LLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSR
YAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1600)

В некоторых вариантах осуществления полипептид Gene Writer содержит гибкий линкер между эндонуклеазой и доменом RT, например, линкер, содержащий аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1601). В некоторых вариантах осуществления домен RT полипептида Gene Writer может быть расположен с C-концевой стороны от эндонуклеазного домена. В некоторых вариантах осуществления домен RT полипептида Gene Writer может быть расположен с N-концевой стороны от эндонуклеазного домена.

В некоторых вариантах осуществления полипептид Gene Writer содержит последовательность dCas9, содержащую мутацию D10A и/или H840A, например, следующую последовательность:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK

NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG
MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES
EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD (SEQ ID NO: 1602)

В некоторых вариантах осуществления молекула РНК-матрицы для применения в системе содержит в направлении от 5' к 3' (1) спейсер gRNA; (2) каркас gRNA; (3) последовательность гетерологичного объекта; (4) 3'-концевой домен гомологии. В некоторых вариантах осуществления:

(1) Представляет собой спейсер Cas9 размером ~ 18-22 нуклеотида, например, 20 нуклеотидов

(2) Представляет собой каркас gRNA, содержащий одну или несколько шпилечных петель, например, 1, 2 или 3 петли для ассоциации матрицы с никазным доменом Cas9. В некоторых вариантах осуществления каркас gRNA в направлении от 5' к 3' имеет последовательность
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
AAGTGGGACCGAGTCGGTCC (SEQ ID NO: 1603).

(3) В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину, например, 7-74, например, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70 или 70-80 нуклеотидов или 80-90 нуклеотидов. В некоторых вариантах осуществления первое (наиболее близкое к 5'-концу) основание последовательности не представляет собой С.

(4) В некоторых вариантах осуществления 3'-концевой домен гомологии, который связывает праймирующую последовательность-мишень после того, как происходит однонитевой разрез, составляет, например, 3-20 нуклеотидов, например, 7-15 нуклеотидов,

например, 12-14 нуклеотидов. В некоторых вариантах осуществления 3'-концевой домен гомологии характеризуется 40-60% содержанием GC.

Вторая gRNA, ассоциированная с системой, может способствовать обеспечению полной интеграции. В некоторых вариантах осуществления вторая gRNA может нацеливаться на местоположение, которое находится на расстоянии 0-200 нуклеотидов от однонитевого разрыва первой нити, например, 0-50, 50-100, 100-200 нуклеотидов от однонитевого разрыва первой нити. В некоторых вариантах осуществления вторая gRNA может связываться со своей последовательностью-мишенью только после осуществления редактирования, например, gRNA связывает последовательность, присутствующую в последовательности гетерологичного объекта, но не в исходной последовательности-мишени.

В некоторых вариантах осуществления описанная в данном документе система Gene Writing применяется для осуществления редактирования в клетках HEK293, K562, U2OS или HeLa. В некоторых вариантах осуществления система Gene Writing применяется для осуществления редактирования в первичных клетках, например, первичных кортикальных нейронах мышей E18.5.

В некоторых вариантах осуществления домен обратной транскриптазы или RT (например, описанный в данном документе) содержит последовательность RT MoMLV или ее вариант. В вариантах осуществления последовательность RT MoMLV содержит одну или несколько мутаций, выбранных из D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S и K103L. В вариантах осуществления последовательность RT MoMLV содержит комбинацию мутаций, таких как D200N, L603W и T330P, необязательно дополнительно включая T306K и/или W313F.

В некоторых вариантах осуществления эндонуклеазный домен (например, описанный в данном документе) содержит nCAS9, например, содержащую мутацию H840A.

В некоторых вариантах осуществления последовательность гетерологичного объекта (например, системы, описанной в данном документе) имеет длину приблизительно 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000 или больше нуклеотидов.

В некоторых вариантах осуществления домен RT и эндонуклеазный домен соединены гибким линкером, например, содержащим аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1601).

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.

Компонент системы редактирования генома Gene Writer, представляющий собой нуклеиновую кислоту-матрицу (например, РНК-матрицу), описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) имеет 3'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 3'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer. Область связывания может ассоциировать с нуклеиновой кислотой-матрицей (например, РНК-матрицей) с помощью любого из полипептидных модулей. В некоторых вариантах осуществления область связывания нуклеиновой кислоты-матрицы (например, РНК-матрицы) может ассоциировать с РНК-связывающим доменом в полипептиде. В некоторых вариантах осуществления область связывания нуклеиновой кислоты-матрицы (например, РНК-матрицы) может ассоциировать с доменом обратной транскрипции полипептида (например, специфично связываться с доменом RT). Например, в случае, когда домен обратной транскрипции получен из ретротранспозона, не содержащего LTR, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать область связывания, полученную из ретротранспозона, не содержащего LTR, например, 3'-UTR из ретротранспозона, не содержащего LTR. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может ассоциировать с ДНК-связывающим доменом полипептида, например, в случае с gRNA, ассоциирующей с ДНК-связывающим доменом, полученным из Cas9. В некоторых вариантах осуществления область связывания может также обеспечивать распознавание ДНК-мишени, например, в случае с gRNA, гибридизирующейся с последовательностью ДНК-мишени и связывающей полипептид, например, домен Cas9. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может ассоциировать с несколькими компонентами полипептида, например, с ДНК-связывающим доменом и доменом обратной транскрипции. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать область gRNA, которая ассоциирует с ДНК-связывающим доменом, полученным

из Cas9, и 3'-UTR из ретротранспозона, не содержащего LTR, которая ассоциирует с доменом обратной транскрипции, полученным из ретротранспозона, не содержащего LTR.

В некоторых вариантах осуществления в системе или способе, описанных в данном документе, содержится одна нуклеиновая кислота-матрица (например, РНК-матрица). В некоторых вариантах осуществления в системе или способе, описанных в данном документе, содержится множество нуклеиновых кислот-матриц (например, РНК-матриц). Например, в системе, описанной в данном документе, содержится первая РНК, содержащая (например, в направлении от 5' к 3') последовательность, которая связывает полипептид Gene Writer (например, ДНК-связывающий домен и/или эндонуклеазный домен, например, gRNA) и последовательность, которая связывает сайт-мишень (например, нередактируемую нить сайта в геноме-мишени), и вторая РНК (например, РНК-матрица), содержащая (например, в направлении от 5' к 3') необязательно последовательность, которая связывает полипептид Gene Writer (например, которая специфично связывает домен RT), последовательность гетерологичного объекта и 3'-концевой домен гомологии. В некоторых вариантах осуществления, когда в системе содержится множество нуклеиновых кислот, каждая нуклеиновая кислота содержит конъюгирующий домен. В некоторых вариантах осуществления конъюгирующий домен обеспечивает ассоциацию молекул нуклеиновых кислот, например, посредством гибридизации комплементарных последовательностей.

В некоторых вариантах осуществления молекула нуклеиновой кислоты-матрицы, описанная в данном документе, 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% идентичностью последовательности с последовательностью нуклеиновой кислоты, содержащейся в молекуле нуклеиновой кислоты-мишени. В вариантах осуществления последовательность нуклеиновой кислоты в молекуле нуклеиновой кислоты-мишени расположена в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 или 100 нуклеотидов от сайта-мишени для вставки (например, с 5'-стороны относительно него), например, для последовательности гетерологичного объекта, например, содержащейся в молекуле нуклеиновой кислоты-матрицы.

В некоторых вариантах осуществления 3'-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 однонитевого 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, может содержать gRNA (например, pegRNA). В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, может связываться с полипептидом Gene Writer™ посредством взаимодействия gRNA-части нуклеиновой кислоты-матрицы с доменом, связывающим нуклеиновую кислоту-матрицу, например, с РНК-связывающим доменом (например, с гетерологичным РНК-связывающим доменом). В некоторых вариантах осуществления гетерологичный РНК-связывающий домен представляет собой белок CRISPR/Cas, например, Cas9.

В некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, характеризуется частично раскрученной

лентовидной структурой gRNA, связанной с ДНК-мишенью (например, как описано в Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484). Не желая ограничиваться какой-либо теорией, полагают, что эта неканоническая структура облегчается вращением каждого шестого нуклеотида из гибрида РНК-ДНК. Таким образом, в некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, может допускать увеличенное количество несовпадений с сайтом-мишенью с некоторым интервалом, например, по каждому шестому основанию. В некоторых вариантах осуществления область нуклеиновой кислоты-матрицы, например, РНК-матрицы, содержащая gRNA, характеризующуюся гомологией с сайтом-мишенью, может иметь неоднозначные положения с регулярными интервалами, например, по каждому шестому основанию, которые не требуют образования пары оснований с сайтом-мишенью.

gRNA с индуцируемой активностью

В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, содержит направляющую РНК (gRNA) с индуцируемой активностью. Индуцируемая активность может быть достигнута нуклеиновой кислотой-матрицей, например, РНК-матрицей, дополнительно содержащей (в дополнение к gRNA) блокирующий домен, где последовательность части блокирующего домена или его всего является по меньшей мере частично комплементарной части gRNA или ей всей. Таким образом, блокирующий домен способен гибридизироваться или по сути гибридизироваться с частью gRNA или с ней всей. В некоторых вариантах осуществления блокирующий домен и gRNA с индуцируемой активностью расположены в нуклеиновой кислоте-матрице, например, РНК-матрице, так что gRNA может принимать первую конформацию, в которой блокирующий домен гибридизирован или по сути гибридизирован с gRNA, и вторую конформацию, в которой блокирующий домен не гибридизирован или по сути не гибридизирован с gRNA. В некоторых вариантах осуществления в первой конформации gRNA не способна связываться с полипептидом Gene Writer (например, с доменом, связывающим нуклеиновую кислоту-матрицу, ДНК-связывающим доменом или эндонуклеазным доменом (например, белком CRISPR/Cas)) или связывается со значительно сниженной аффинностью по сравнению с в остальном аналогичной РНК-матрицей без блокирующего домена. В некоторых вариантах осуществления во второй конформации gRNA способна связываться с полипептидом Gene Writer (например, с доменом, связывающим нуклеиновую кислоту-матрицу, ДНК-связывающим доменом или эндонуклеазным доменом (например, белком CRISPR/Cas)). В некоторых вариантах осуществления то, находится ли gRNA в первой или во второй конформации, может влиять на активность связывания ДНК или эндонуклеазную активность полипептида Gene Writer (например, белка CRISPR/Cas, который содержит полипептид Gene Writer). В некоторых вариантах осуществления гибридизация gRNA с блокирующим доменом может быть нарушена с помощью открывающей молекулы. В некоторых вариантах осуществления открывающая молекула содержит средство, которое связывается с частью gRNA или с ней

всей частью или с блокирующим доменом и ингибирует гибридизацию gRNA с блокирующим доменом. В некоторых вариантах осуществления открывающая молекула содержит нуклеиновую кислоту, например, содержащую последовательность, частично или полностью комплементарную gRNA, блокирующему домену или им обоим. С помощью выбора или разработки подходящей открывающей молекулы, предоставление открывающей молекулы может способствовать изменению конформации gRNA таким образом, чтобы она могла ассоциировать с белком CRISPR/Cas и обеспечивать функции, ассоциированные с белком CRISPR/Cas (например, активность связывания ДНК и/или эндонуклеазную активность). Не желая ограничиваться какой-либо теорией, полагают, что предоставление открывающей молекулы в выбранное время и/или в выбранном местоположении может обеспечить пространственный и временной контроль активности gRNA, белка CRISPR/Cas или системы Gene Writer, содержащей их. В некоторых вариантах осуществления открывающая молекула является экзогенной по отношению к клетке, содержащей полипептид Gene Writer и/или нуклеиновую кислоту-матрицу. В некоторых вариантах осуществления открывающая молекула содержит эндогенное средство (например, эндогенное по отношению к клетке, содержащей полипептид Gene Writer и/или нуклеиновую кислоту-матрицу, содержащую gRNA и блокирующий домен). Например, индуцируемая gRNA, блокирующий домен и открывающая молекула могут быть выбраны таким образом, чтобы открывающая молекула представляла собой эндогенное средство, экспрессируемое в клетке или ткани, являющихся мишенями, например, за счет чего обеспечивается активность системы Gene Writer в клетке или ткани, являющихся мишенями. В качестве дополнительного примера, индуцируемая gRNA, блокирующий домен и открывающая молекула могут быть выбраны таким образом, чтобы открывающая молекула отсутствовала или практически не экспрессировалась в одной или нескольких клетках или тканях, не являющихся мишенями, например, за счет чего обеспечивается то, что активность системы Gene Writer не проявляется или практически не проявляется в одной или нескольких клетках или тканях, не являющихся мишенями, или проявляется на сниженном уровне по сравнению с клеткой или тканью, являющимися мишенями. Иллюстративные блокирующие домены, открывающие молекулы и их пути применения описаны в публикации заявки согласно PCT WO2020044039A1, которая включена в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления нуклеиновая кислота-матрица, например, РНК-матрица, может содержать одну или несколько UTR (например, из ретротранспозона типа R2) и gRNA. В некоторых вариантах осуществления UTR облегчает взаимодействие нуклеиновой кислоты-матрицы (например, РНК-матрицы) с записывающим доменом, например, доменом обратной транскриптазы, полипептида Gene Writer. В некоторых вариантах осуществления gRNA облегчает взаимодействие с доменом, связывающим нуклеиновую кислоту-матрицу (например, РНК-связывающим доменом) полипептида. В некоторых вариантах осуществления gRNA направляет полипептид к соответствующей последовательности-мишени, например, в геноме клетки-мишени. В некоторых вариантах осуществления

нуклеиновая кислота-матрица может содержать только мотив связывания обратной транскриптазы (например, 3'-UTR из R2), и gRNA может быть предоставлена в качестве второй молекулы нуклеиновой кислоты (например, второй молекулы РНК) для распознавания сайта-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица, содержащая RT-связывающий мотив, может существовать в той же молекуле, что и gRNA, но процессироваться с образованием двух молекул РНК благодаря активности расщепления (например, рибозима).

В некоторых вариантах осуществления РНК-матрица может быть специально разработана для коррекции заданной мутации в геномной ДНК клетки-мишени (например, ex vivo или in vivo, например, в ткани или органе, являющимися мишенями, например, у субъекта). Например, мутация может представлять собой мутацию, ассоциированную с заболеванием, по сравнению с последовательностью дикого типа. Не желая ограничиваться какой-либо теорией, полагают, что наборы эмпирических параметров помогают обеспечить оптимальные исходные конструкции РНК-матриц или их частей in silico. В качестве неограничивающего иллюстративного примера для выбранной мутации можно использовать следующие параметры разработки. В некоторых вариантах осуществления разработка начинается с получения примерно 500 п. о. (например, вплоть до 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 или 700 п. о. и необязательно по меньшей мере 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 или 650 п. о.) фланкирующей последовательности с каждой стороны от мутации, которые служат в качестве области-мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица содержит gRNA. Методика разработки gRNA известна специалистам в данной области. В некоторых вариантах осуществления gRNA содержит последовательность (например, спейсер CRISPR), которая связывает сайт-мишень. В некоторых вариантах осуществления последовательность (например, спейсер CRISPR), которая связывает сайт-мишень, для применения при нацеливании нуклеиновой кислоты-матрицы на область-мишень выбирают посредством рассмотрения конкретного используемого полипептида Gene Writer (например, эндонуклеазного домена или записывающего домена, например, содержащего домен CRISPR/Cas) (например, для Cas9, мотив, прилегающий к протоспейсеру (PAM), NGG непосредственно с 3'-стороны от gRNA-связывающей области размером в 20 нуклеотидов). В некоторых вариантах осуществления спейсер CRISPR выбирают посредством ранжирования сперва в зависимости от того, будет ли разрушен PAM в результате редактирования, индуцированного Gene Writing. В некоторых вариантах осуществления разрушение PAM может привести к увеличению эффективности редактирования. В некоторых вариантах осуществления PAM можно разрушить также посредством введения (например, в качестве части или в качестве дополнения другой модификации сайта-мишени в геномной ДНК) молчащей мутации (например, мутации, при которой не изменяется аминокислотный остаток, кодируемый последовательностью нуклеиновой кислоты-мишени, если таковой имеется) в сайт-мишень во время использования Gene Writing. В некоторых вариантах осуществления спейсер CRISPR

выбирают посредством ранжирования последовательностей по близости их соответствующего геномного сайта к желаемому месту редактирования. В некоторых вариантах осуществления gRNA содержит каркас gRNA. В некоторых вариантах осуществления используемый каркас gRNA может представлять собой стандартный каркас (например, для Cas9, 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC-3' (SEQ ID NO: 1603)) или может содержать одну или несколько нуклеотидных замен. В некоторых вариантах осуществления последовательность гетерологичного объекта характеризуется по меньшей мере 90% идентичностью, например, по меньшей мере 90%, 95%, 98%, 99% или 100% идентичностью, или содержит не более 1, 2, 3, 4 или 5 положений неидентичности по отношению к сайту-мишени с 3'-стороны от однонитевого разрыва в первой нити (например, непосредственно с 3'-стороны от однонитевого разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от однонитевого разрыва в первой нити), за исключением любой вставки, замены или делеции, которые могут быть записаны в сайт-мишень с помощью Gene Writer. В некоторых вариантах осуществления 3'-концевой домен гомологии с мишенью характеризуется по меньшей мере 90% идентичностью, например, по меньшей мере 90%, 95%, 98%, 99% или 100% идентичностью, или содержит не более 1, 2, 3, 4 или 5 положений неидентичности по отношению к сайту-мишени с 5'-стороны от однонитевого разрыва в первой нити (например, непосредственно с 5'-стороны от однонитевого разрыва в первой нити или на расстоянии до 1, 2, 3, 4 или 5 нуклеотидов с 3'-стороны от однонитевого разрыва в первой нити).

В некоторых вариантах осуществления 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 нагрузки без какой-либо дополнительной последовательности (например, единицы экспрессии гена без UTR, используемой для связывания белка Gene Writer).

Нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана таким образом, чтобы приводить к вставкам, мутациям или делециям в локусе ДНК-

мишени. В некоторых вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана таким образом, чтобы вызывать вставку в ДНК-мишень. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может содержать гетерологичную последовательность, где обратная транскрипция приведет к вставке гетерологичной последовательности в ДНК-мишень. В других вариантах осуществления РНК-матрица может быть разработана для записи делеции в ДНК-мишень. Например, нуклеиновая кислота-матрица (например, РНК-матрица) может соответствовать ДНК-мишени выше и ниже желаемой делеции, где обратная транскрипция будет приводить к копированию вышерасположенной и нижерасположенной последовательностей нуклеиновой кислоты-матрицы (например, РНК-матрицы) без вставочной последовательности, например, вызывая делецию вставочной последовательности. В других вариантах осуществления нуклеиновая кислота-матрица (например, РНК-матрица) может быть разработана для записи редактирования в ДНК-мишень. Например, РНК-матрица может соответствовать последовательности ДНК-мишени, за исключением одного или нескольких нуклеотидов, где обратная транскрипция приведет к копированию этих редактирований в ДНК-мишени, например, что приведет к мутациям, например, мутациям по типу транзиции или трансверсии.

В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов (и необязательно не более 500, 400, 300, 200 или 100 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить вставку в сайт-мишень по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 или 200 нуклеотидов (и необязательно не более 500, 400, 300 или 200 нуклеотидов). В некоторых вариантах осуществления система Gene Writer способна производить делецию по меньшей мере 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 или 10 тысяч оснований (и необязательно не более 1, 5, 10 или 20 тысяч оснований). 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.

Способы и композиции для модифицированной РНК (например, gRNA или РНК-матрицы)

В некоторых вариантах осуществления РНК-компонент системы (например, РНК-матрица или gRNA, например, описанные в данном документе) содержит одну или несколько модификаций нуклеотидов. В некоторых вариантах осуществления паттерн модификаций gRNA может значительно влиять на активность *in vivo* по сравнению с немодифицированными направляющими или направляющими с модифицированными концами (например, как показано на фигуре 1D в Finn et al. *Cell Rep* 22(9):2227-2235 (2018); включенном в данный документ посредством ссылки во всей своей полноте). Не желая ограничиваться какой-либо теорией, полагают, что данный процесс может быть обусловлен по меньшей мере частично стабилизацией РНК, обеспечиваемой модификациями. Неограничивающие примеры таких модификаций могут включать 2'-O-метил (2'-O-Me), 2'-O-(2-метоксиэтил) (2'-O-MOE), 2'-фтор (2'-F), фосфотиоатную (PS) связь между нуклеотидами, замены G-C и перевернутые связи, лишенные оснований, между нуклеотидами и их эквивалентами.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывает сайт-мишень) или направляющая РНК содержит 5'-концевую область. В некоторых вариантах осуществления РНК-матрица или направляющая РНК не содержит 5'-концевую область. В некоторых вариантах осуществления 5'-концевая область содержит спейсерную область CRISPR, например, как описано в отношении sgRNA в Briner AE et al, Molecular Cell 56: 333-339 (2014) (включенном в данный документ посредством ссылки во всей своей полноте; применимо в данном документе, например, ко всем направляющим РНК). В некоторых вариантах осуществления 5'-концевая область содержит модификацию 5'-конца. В некоторых вариантах осуществления 5'-концевая область со спейсерной областью или без нее может быть ассоциирована с crRNA, trRNA, sgRNA и/или dgRNA. Спейсерная область CRISPR может в некоторых случаях содержать направляющую область, направляющий домен или нацеливающий домен. В некоторых вариантах осуществления домен-мишень или последовательность-мишень могут содержать последовательность нуклеиновой кислоты, к которой направляющая область/домен направляет нуклеазу для расщепления. В некоторых вариантах осуществления белок spyCas9 может быть направлен направляющей областью/доменом к последовательности-

мишени молекулы нуклеиновой кислоты-мишени с помощью нуклеотидов, присутствующих в спейсерной области CRISPR.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывает сайт-мишень) или направляющая РНК, например, описанная в данном документе, содержит любую из последовательностей, показанных в таблице 4 документа WO2018107028A1, включенного в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления, где последовательность содержит направляющую и/или спейсерную область, композиция может содержать или не содержать данную область. В некоторых вариантах осуществления направляющая РНК содержит одну или несколько модификаций любой из последовательностей, показанных в таблице 4 документа WO2018107028A1, например, идентифицированных в нем с помощью SEQ ID NO. В вариантах осуществления нуклеотиды могут быть одинаковыми или разными, и/или показанный паттерн модификаций может быть таким же, как паттерн модификаций направляющей последовательности, показанной в таблице 4 документа WO2018107028A1, или сходным с ним. В некоторых вариантах осуществления паттерн модификаций включает относительное положение и отличительные особенности модификаций gRNA или области gRNA (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области). В некоторых вариантах осуществления паттерн модификаций содержит по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% модификаций в любой из последовательностей, показанных в столбце "Последовательности" таблицы 4 в документе WO2018107028A1, и/или в одной или нескольких областях последовательности. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным паттерну модификаций любой из последовательностей, показанных в столбце "Последовательности" таблицы 4 в документе WO2018107028A1. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в одной или нескольких областях последовательности, показанной в таблице 4 в документе WO2018107028A1, например, в 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2 и/или 3'-концевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным паттерну модификаций последовательности в 5'-концевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в нижней стеблевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в области выпетливания. В некоторых вариантах осуществления паттерн

модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в верхней стеблевой области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в связующей области. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в шпилечной области 1. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в шпилечной области 2. В некоторых вариантах осуществления паттерн модификаций является на по меньшей мере 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичным в 3'-концевой области. В некоторых вариантах осуществления паттерн модификаций отличается от паттерна модификаций последовательности из таблицы 4 документа WO 2018107028A1 или области (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области) такой последовательности, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам. В некоторых вариантах осуществления gRNA содержит модификации, которые отличаются от модификаций последовательности из таблицы 4 документа WO 2018107028A1, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам. В некоторых вариантах осуществления паттерн модификаций отличается от паттерна модификаций последовательности из таблицы 4 документа WO 2018107028A1 или области (например, 5'-концевой области, нижней стеблевой области, области выпетливания, верхней стеблевой области, связующей области, шпилечной области 1, шпилечной области 2, 3'-концевой области) такой последовательности, например, по 0, 1, 2, 3, 4, 5, 6 или больше нуклеотидам.

В некоторых вариантах осуществления РНК-матрица (например, та ее часть, которая связывается с сайтом-мишенью) или gRNA содержит 2'-O-метил- (2'-O-Me-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит 2'-O-(2-метоксиэтил-) (2'-O-MOE-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит 2'-фтор- (2'-F-) модифицированный нуклеотид. В некоторых вариантах осуществления gRNA содержит фосфотиоатную (PS) связь между нуклеотидами. В некоторых вариантах осуществления gRNA содержит модификацию 5'-конца, модификацию 3'-конца или модификации 5'- и 3'-концов. В некоторых вариантах осуществления модификация 5'-конца включает в себя фосфотиоатную (PS) связь между нуклеотидами. В некоторых вариантах осуществления модификация 5'-конца включает в себя 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модифицированный нуклеотид. В некоторых вариантах осуществления модификация 5'-конца включает в себя по меньшей мере одну фосфотиоатную (PS) связь и один или несколько 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модифицированных нуклеотидов. Модификация конца может включать в себя

фосфотиоатную (PS), 2'-O-метил- (2'-O-Me-), 2'-O-(2-метоксиэтил-) (2'-O-MOE-) и/или 2'-фтор- (2'-F-) модификацию. Эквивалентные модификации конца также охватываются вариантами осуществления, описанными в данном документе. В некоторых вариантах осуществления РНК-матрица или gRNA содержит модификацию конца в комбинации с модификацией одной или нескольких областей РНК-матрицы или gRNA. Дополнительные иллюстративные модификации и способы для защиты РНК, например, gRNA, и их формулы описаны в документе WO2018126176A1, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления для введения модификаций (например, модификаций 2'-OMe-РНК, 2'-F-РНК и PS) в РНК-матрицу или направляющую РНК используются структурно-ориентированные и систематические подходы, например, описанные в Mir et al. *Nat Commun* 9:2641 (2018) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления включение 2'-F-РНК приводит к повышению термической и нуклеазной стабильности дуплексов РНК:РНК или РНК:ДНК, например, минимально препятствуя излому плоскости сахарного кольца в конформации C3'-эндо. В некоторых вариантах осуществления 2'-F может лучше переноситься, чем 2'-OMe, в положениях, где 2'-OH важен для стабильности дуплекса РНК:ДНК. В некоторых вариантах осуществления crRNA содержит одну или несколько модификаций, которые не приводят к снижению активности Cas9, например, C10, C20 или C21 (полностью модифицированные), например, как описано в дополнительной таблице 1 в Mir et al. *Nat Commun* 9:2641 (2018), включенном в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления tracrRNA содержит одну или несколько модификаций, которые не приводят к снижению активности Cas9, например, T2, T6, T7 или T8 (полностью модифицированные) из дополнительной таблицы 1 в Mir et al. *Nat Commun* 9:2641 (2018). В некоторых вариантах осуществления crRNA, содержащая одну или несколько модификаций (например, описанных в данном документе), может быть спарена с tracrRNA, содержащей одну или несколько модификаций, например, C20 и T2. В некоторых вариантах осуществления gRNA содержит химерное соединение, например, из crRNA и tracrRNA (например, в Jinek et al. *Science* 337(6096):816-821 (2012)). В вариантах осуществления модификации из crRNA и tracrRNA картируются на химерное соединение с одной направляющей молекулой, например, для получения модифицированной gRNA с повышенной стабильностью.

В некоторых вариантах осуществления молекулы gRNA можно модифицировать посредством добавления или удаления встречающихся в природе структурных компонентов, например, шпилек. В некоторых вариантах осуществления gRNA может содержать gRNA с одним или несколькими удаленными 3'-шпилечными элементами, например, как описано в документе WO2018106727, включенном в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления gRNA может содержать дополнительную шпилечную структуру, например, дополнительную шпилечную структуру в спейсерной области, которая, как было показано, приводит к

повышению специфичности системы CRISPR-Cas согласно идеям Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019). Дополнительные модификации, в том числе примеры укороченной gRNA и специфичных модификаций, приводящих к улучшению активности in vivo, можно найти в документе US20190316121, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления структурно-ориентированный и систематический подходы (например, описанные в Mir et al. *Nat Commun* 9:2641 (2018); включенном в данный документ посредством ссылки во всей своей полноте) используются для поиска модификаций для РНК-матрицы. В вариантах осуществления модификации идентифицируют по включению или исключению направляющей области РНК-матрицы. В некоторых вариантах осуществления структуру полипептида, связанного с РНК-матрицей, используют для определения не контактирующих с белком нуклеотидов в РНК, которые затем могут быть выбраны для модификаций, например, с более низким риском разрушения ассоциации РНК с полипептидом. Вторичные структуры в РНК-матрице также можно предсказать in silico с помощью программных инструментов, например, инструмента RNAstructure, доступного по адресу rna.urmc.rochester.edu/RNAstructureWeb (Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); включенный в данный документ посредством ссылки во всей своей полноте), например, для определения вторичных структур для выбора модификаций, например, шпилек, стеблей и/или выпетливаний.

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 T7-полимеразой, из линеаризованной ДНК-матрицы, где UTP необязательно замещен 1-метилпсевдо-UTP. В некоторых вариантах осуществления транскрипт содержит 5'- и 3'-UTR, например, GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1604) и UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1605), или их функциональные фрагменты или варианты и необязательно содержит поли(A)-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 T7 начинается с мотива GGG. В некоторых вариантах осуществления транскрипт от промотора T7 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 T7 лэддера поли(G)-продуктов в результате проскальзывания транскрипта по трем остаткам C 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 T7 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 T4, может использоваться для соединения двух или более сегментов РНК друг с другом. При использовании реагента, такого как РНК-лигаза, 5'-

конец, как правило, связывают с 3'-концом. В некоторых вариантах осуществления, если два сегмента соединены, то могут быть образованы две возможные линейные конструкции (т. е. (1) 5'-сегмент 1-сегмент 2-3' и (2) 5'-сегмент 2-сегмент 1-3′). В некоторых вариантах осуществления также может иметь место внутримолекулярная циркуляризация. Обе эти проблемы можно решить, например, блокируя один 5'-конец или один 3'-конец, чтобы РНК-лигаза не могла лигировать этот конец с другим концом. В вариантах осуществления, если конструкция 5'-сегмент 1-сегмент 2-3' является желаемой, то размещение блокирующей группы либо на 5'-конце сегмента 1, либо на 3'-конце сегмента 2 может привести к образованию только правильного линейного продукта лигирования и/или предотвратить внутримолекулярную циркуляризацию. Композиции и способы для ковалентного соединения двух сегментов нуклеиновой кислоты (например, РНК) раскрыты, например, в US20160102322A1 (включенном в данный документ посредством ссылки во всей своей полноте), наряду со способами, включающими применение РНК-лигазы для направленного лигирования двух однонитевых сегментов РНК друг с другом.

Одним из примеров блокатора концов, который можно применять в сочетании, например, с РНК-лигазой Т4, является средство для дидезокси-терминации. РНК-лигаза Т4, как правило, катализирует АТР-зависимое лигирование фосфодиэфирных связей между 5'-фосфатным и 3'-гидроксильным концами. В некоторых вариантах осуществления при применении РНК-лигазы Т4 на концах, подвергаемых лигированию, должны присутствовать подходящие концы. Один из способов блокирования РНК-лигазы Т4 на конце предусматривает отсутствие правильного формата конца. Обычно концы сегментов РНК с 5-гидроксилом или 3'-фосфатом не будут выступать в качестве субстратов для РНК-лигазы Т4.

Дополнительные иллюстративные способы, которые можно применять для соединения сегментов РНК, включают клик-химию (например, описанную в патентах США №№ 7375234 и 7070941 и публикации заявки на патент США № 2013/0046084, полное раскрытие которых включено в данный документ посредством ссылки). Например, одна иллюстративная реакция клик-химии происходит между алкиновой группой и азидной группой (см. фиг. 11 документа US20160102322A1, который включен в данный документ посредством ссылки во всей своей полноте). Любая клик-реакция потенциально может быть использована для связывания сегментов РНК (например, Cu-катализируемое азид-алкиновое циклоприсоединение, промотируемое напряжением азид-алкиновое циклоприсоединение, лигирование по Штаудингеру, лигирование тетразином, фотоиндуцированная тетразол-алкеновая реакция, тиол-еновая реакция, использование сложных NHS-эфиров, эпоксидов, изоцианатов и реакция альдегид-аминоокси). В некоторых вариантах осуществления лигирование молекул РНК с использованием реакции клик-химии является преимущественным, поскольку реакции клик-химии являются быстрыми, модульными, эффективными, часто не приводят к получению токсичных побочных продуктов, могут проводиться с использованием воды в качестве растворителя и/или могут быть настроены так, чтобы они были стереоспецифическими.

В некоторых вариантах осуществления сегменты РНК могут быть соединены с использованием азид-алкинового циклоприсоединения Хьюсгена - реакции, которая, как правило, представляет собой 1,3-диполярное циклоприсоединение между азидом и концевым или внутренним алкином с получением 1,2,3-триазола для лигирования сегментов РНК. Не желая ограничиваться какой-либо теорией, полагают, что одним из преимуществ данного способа лигирования может быть то, что эта реакция может быть инициирована посредством добавления необходимых ионов Cu(I). Другие иллюстративные механизмы, с помощью которых сегменты РНК могут быть соединены, включают без ограничения использование реакций присоединения галогенов (F-, Br-, I-) и алкинов, карбонилов/сульфгидрилов/малеимида и связей между карбоксильными и аминогруппами. Например, одна молекула РНК может быть модифицирована с помощью тиола в положении 3' (с использованием дисульфидамидита и универсальной подложки или подложки, модифицированной дисульфидом), и другая молекула РНК может быть модифицирована акридитом в положении 5' (с использованием акрилового фосфорамидита), тогда две РНК молекулы могут быть соединены с помощью реакции присоединения Михаэля. Такая стратегия также может быть применена для поэтапного соединения нескольких молекул РНК. Также предусмотрены способы связывания более чем двух (например, трех, четырех, пяти, шести и т. д.) молекул РНК друг с другом. Не желая ограничиваться какой-либо теорией, полагают, что это может быть полезно в случае, если желаемая молекула РНК длиннее, чем приблизительно 40 нуклеотидов, например, так что снижается эффективность химического синтеза, например, как указано в документе US20160102322A1 (включенном в данный документ посредством ссылки во всей своей полноте).

В качестве иллюстрации, длина tracrRNA, как правило, составляет около 80 нуклеотидов. Такие молекулы РНК могут быть получены, например, с помощью таких процессов, как транскрипция in vitro или химический синтез. В некоторых вариантах осуществления, когда для получения таких молекул РНК используется химический синтез, они могут быть получены в качестве единого продукта синтеза или посредством связывания двух или более синтезированных сегментов РНК друг с другом. В вариантах осуществления, когда три или более сегмента РНК соединены друг с другом, можно применять разные способы для связывания отдельных сегментов вместе. Также сегменты РНК могут быть соединены друг с другом в одном сосуде (например, контейнере, колбе, лунке, пробирке, планшете или другом резервуаре) все одновременно, или в одном сосуде в разное время, или в разных сосудах в разное время. В неограничивающем примере для сборки сегментов 1, 2 и 3 РНК в порядке номеров сегменты 1 и 2 РНК могут быть сначала соединены друг с другом 5'-конец к 3'-концу. Затем продукт реакции можно очистить от компонентов реакционной смеси (например, с помощью хроматографии), затем поместить во второй сосуд для соединения 3'-конца с 5'-концом сегмента 3 РНК. Затем конечный продукт реакции можно соединить с 5'-концом сегмента 3 РНК.

В другом неограничивающем примере сегмент 1 РНК (приблизительно 30 нуклеотидов) представляет собой последовательность распознавания локуса-мишени

crRNA и часть шпилечной области 1. Сегмент 2 РНК (приблизительно 35 нуклеотидов) содержит оставшуюся часть шпилечной области 1 и часть линейной tracrRNA между шпилечной областью 1 и шпилечной областью 2. Сегмент 3 РНК (приблизительно 35 нуклеотидов) содержит оставшуюся часть линейной tracrRNA между шпилечной областью 1 и шпилечной областью 2 и всю шпилечную область 2. В данном примере сегменты 2 и 3 РНК связаны 5'-конец к 3'-концу с помощью клик-химии. Дополнительно, оба из 5'- и 3'-концов продукта реакции являются фосфорилированными. Затем продукт реакции приводят в контакт с сегментом 1 РНК, имеющим 3'-концевую гидроксильную группу, и с РНК-лигазой Т4 для получения молекулы направляющей РНК.

Для соединения сегментов РНК в соответствии со способом по настоящему изобретению можно использовать ряд дополнительных связывающих химических соединений. Некоторые из этих химических соединений представлены в таблице 6 документа US20160102322A1, который включен в данный документ посредством ссылки во всей своей полноте.

Системы Gene Writer, например, термостабильные системы Gene Writer

Не желая ограничиваться какой-либо теорией, полагают, что в некоторых вариантах осуществления ретротранспозазы, эволюционировавшие в холодных условиях, не могут функционировать так же хорошо при температуре тела человека. В настоящей заявке представлен ряд термостабильных систем Gene Writer, включая белки, полученные из ретротранспозаз птиц. 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, 2 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 клетке человека, культивируемой при 37°C, например, при использовании анализа из примера 6 или примера 7 в данном документе.

В некоторых вариантах осуществления анализ включает стадии: (1) введения клеток HEK293T в одну или несколько лунок диаметром 6,4 мм из расчета 10000 клеток/лунка, (2) инкубирования клеток при 37°C в течение 24 часов, (3) получения смеси для трансфекции, содержащей 0,5 мкл реагента для трансфекции FuGENE® HD и 80 нг ДНК (где ДНК представляет собой плазмиду, содержащую в указанном порядке: (a) промотор CMV, (b) 100 п. о. последовательности, гомологичные 100 п. о. выше сайта-мишени, (c) последовательность, кодирующую 5'-нетранслируемую область, которая связывается с белком Gene Writer, (d) последовательность, кодирующую белок Gene Writer, (e) последовательность, кодирующую 3'-нетранслируемую область, которая связывается с белком Gene Writer, (f) 100 п. о. последовательности, гомологичные 100 п. о. ниже сайта-мишени, и (g) последовательность полиаденилирования BGH) в 10 мкл Opti-MEM, и инкубирования в течение 15 минут при комнатной температуре, (4) добавления смеси для трансфекции к клеткам, (5) инкубирования клеток в течение 3 дней и (6) анализа интеграции экзогенной последовательности в локус-мишень (например, рДНК) в геноме клетки, например, где одна или несколько из предыдущих стадий выполняются, как описано в примере 6 в данном документе.

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. 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. Вставку можно измерить, например, используя пороговое значение выше 1%, 1,5%, 2%, 2,5%, 3%, 3,5%, 4%, 4,5%, 5%, например, как описано в примере 8. 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.

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. 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 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.

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 (см. фигуру 17). 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. В некоторых вариантах осуществления полипептид Gene Writer может содержать линкер, например, пептидный линкер, например, линкер, описанный в таблице 38. В таблице 38 представлены линкерные последовательности для увеличения экспрессии, стабильности и функции полипептидов Gene Writer, содержащих несколько функциональных доменов.

Таблица 38. Иллюстративные линкерные последовательности

Аминокислотная последовательность	**SEQ ID NO**
GGS	1701

GGSGGS	1702
GGSGGSGGS	1703
GGSGGSGGSGGS	1704
GGSGGSGGSGGSGGS	1705
GGSGGSGGSGGSGGSGGS	1706
GGGGS	1707
GGGGSGGGGS	1708
GGGGSGGGGSGGGGS	1709
GGGGSGGGGSGGGGSGGGGS	1710
GGGGSGGGGSGGGGSGGGGSGGGGS	1711
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	1712
GGG	1713
GGGG	1714
GGGGG	1715
GGGGGG	1716
GGGGGGG	1717
GGGGGGGG	1718
GSS	1719
GSSGSS	1720
GSSGSSGSS	1721
GSSGSSGSSGSS	1722
GSSGSSGSSGSSGSS	1723
GSSGSSGSSGSSGSSGSS	1724
EAAAK	1725
EAAAKEAAAK	1726
EAAAKEAAAKEAAAK	1727
EAAAKEAAAKEAAAKEAAAK	1728
EAAAKEAAAKEAAAKEAAAKEAAAK	1729
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	1730
PAP	1731
PAPAP	1732
PAPAPAP	1733
PAPAPAPAP	1734
PAPAPAPAPAP	1735

PAPAPAPAPAPAP	1736
GGSGGG	1737
GGGGGS	1738
GGSGSS	1739
GSSGGS	1740
GGSEAAAK	1741
EAAAKGGS	1742
GGSPAP	1743
PAPGGS	1744
GGGGSS	1745
GSSGGG	1746
GGGEAAAK	1747
EAAAKGGG	1748
GGGPAP	1749
PAPGGG	1750
GSSEAAAK	1751
EAAAKGSS	1752
GSSPAP	1753
PAPGSS	1754
EAAAKPAP	1755
PAPEAAAK	1756
GGSGGGGSS	1757
GGSGSSGGG	1758
GGGGGSGSS	1759
GGGGSSGGS	1760
GSSGGSGGG	1761
GSSGGGGGS	1762
GGSGGGEAAAK	1763
GGSEAAAKGGG	1764
GGGGGSEAAAK	1765
GGGEAAAKGGS	1766
EAAAKGGSGGG	1767
EAAAKGGGGGS	1768
GGSGGGPAP	1769

GGSPAPGGG	1770
GGGGGSPAP	1771
GGGPAPGGS	1772
PAPGGSGGG	1773
PAPGGGGGS	1774
GGSGSSEAAAK	1775
GGSEAAAKGSS	1776
GSSGGSEAAAK	1777
GSSEAAAKGGS	1778
EAAAKGGSGSS	1779
EAAAKGSSGGS	1780
GGSGSSPAP	1781
GGSPAPGSS	1782
GSSGGSPAP	1783
GSSPAPGGS	1784
PAPGGSGSS	1785
PAPGSSGGS	1786
GGSEAAAKPAP	1787
GGSPAPEAAAK	1788
EAAAKGGSPAP	1789
EAAAKPAPGGS	1790
PAPGGSEAAAK	1791
PAPEAAAKGGS	1792
GGGGSSEAAAK	1793
GGGEAAAKGSS	1794
GSSGGGEAAAK	1795
GSSEAAAKGGG	1796
EAAAKGGGGSS	1797
EAAAKGSSGGG	1798
GGGGSSPAP	1799
GGGPAPGSS	1800
GSSGGGPAP	1801
GSSPAPGGG	1802
PAPGGGGSS	1803

PAPGSSGGG	1804
GGGEAAAKPAP	1805
GGGPAPEAAAK	1806
EAAAKGGGPAP	1807
EAAAKPAPGGG	1808
PAPGGGEAAAK	1809
PAPEAAAKGGG	1810
GSSEAAAKPAP	1811
GSSPAPEAAAK	1812
EAAAKGSSPAP	1813
EAAAKPAPGSS	1814
PAPGSSEAAAK	1815
PAPEAAAKGSS	1816
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	1817
GGGGSEAAAKGGGGS	1818
EAAAKGGGGSEAAAK	1819
SGSETPGTSESATPES	1820
GSAGSAAGSGEF	1821
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	1822

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 GeneWriter, описанный в данном документе, имеет ковалентно димеризованную конфигурацию, например, показанную на любой из фиг. 17A-17F документа PCT/US2019/048607, включенного в данный документ посредством ссылки. Изображенные белки представляют собой следующие. Фиг. 17A - полноразмерный фермент дикого типа. Фиг. 17B - два полноразмерных фермента (каждый из которых содержит ДНК-связывающий домен, РНК-связывающий домен, домен обратной транскриптазы и эндонуклеазный домен), соединенные посредством линкера. Фиг. 17C - ДНК-связывающий домен и РНК-связывающий домен, соединенные посредством линкера с полноразмерным ферментом. Фиг. 17D - ДНК-связывающий домен и РНК-связывающий домен, соединенные посредством линкера с РНК-связывающим

доменом, доменом обратной транскриптазы и эндонуклеазным доменом. Фиг. 17E - ДНК-связывающий домен, соединенный посредством первого линкера с РНК-связывающим доменом, который соединен посредством второго линкера со вторым РНК-связывающим доменом, доменом обратной транскриптазы и эндонуклеазным доменом. Фиг. 17F - ДНК-связывающий домен, соединенный посредством первого линкера с РНК-связывающим доменом, который соединен посредством второго линкера со множеством РНК-связывающих доменов (на данной фигуре молекула содержит три РНК-связывающих домена), которые соединены посредством линкера с доменом обратной транскриптазы и эндонуклеазным доменом. В некоторых вариантах осуществления каждый R2 связан с UTR в РНК-матрице. В некоторых вариантах осуществления по меньшей мере один модуль содержит домен обратной транскриптазы и эндонуклеазный домен. В некоторых вариантах осуществления белок содержит множество РНК-связывающих доменов. В некоторых вариантах осуществления модульная система является разделенной и является активной только при ее связывании с ДНК, при этом система использует два разных ДНК-связывающих модуля, например, первый белок, содержащий первый ДНК-связывающий модуль, слитый с РНК-связывающим модулем, который привлекает РНК-матрицу для осуществления обратной транскрипции с праймированием мишенью, и второй белок, который содержит второй ДНК-связывающий модуль, который связывается в сайте интеграции и слит с модулем обратной транскрипции и эндонуклеазным модулем. В некоторых вариантах осуществления нуклеиновая кислота, кодирующая Gene Writer, содержит интеин, вследствие чего белок Gene Writer экспрессируется из двух отдельных генов и после трансляции подвергается слиянию в ходе белкового сплайсинга. В некоторых вариантах осуществления Gene Writer получен из белка, кодируемого не содержащим LTR транспозоном, например, белка R2.

В некоторых вариантах осуществления одна субъединица может содержать только эндонуклеазный домен. В некоторых вариантах осуществления две субъединицы, составляющие один полипептид, могут выполнять взаимодополняющие функции.

В некоторых вариантах осуществления одна субъединица может не обладать функциональными свойствами эндонуклеазы. В некоторых вариантах осуществления в одной субъединице может отсутствовать эндонуклеазный домен. В некоторых вариантах осуществления одна субъединица может обладать только активностью обратной транскриптазы. В некоторых вариантах осуществления одна субъединица может содержать только домен обратной транскриптазы. В некоторых вариантах осуществления одна субъединица может обладать только функциональным свойством ДНК-зависимого синтеза ДНК.

Линкеры

В некоторых вариантах осуществления домены композиций и систем, описанных в данном документе (например, эндонуклеазный домен и домен обратной транскриптазы полипептида или ДНК-связывающий домен и домены обратной транскриптазы полипептида), могут быть соединены линкером. Описанная в данном документе

композиция, содержащая линкерный элемент, имеет общую форму S1-L-S2, где S1 и S2 могут быть одинаковыми или разными и представляют собой два доменных компонента (например, каждый из которых является доменом полипептида или нуклеиновой кислоты), связанных друг с другом с помощью линкера. В некоторых вариантах осуществления линкер может соединять два полипептида. В некоторых вариантах осуществления линкер может соединять две молекулы нуклеиновой кислоты. В некоторых вариантах осуществления линкер может соединять полипептид и молекулу нуклеиновой кислоты. Линкер может представлять собой химическую связь, например, одну или несколько ковалентных связей или нековалентных связей. Линкер может быть гибким, жестким и/или расщепляемым. В определенных вариантах осуществления линкер представляет собой пептидный линкер. Как правило, пептидный линкер имеет длину по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше аминокислот, например, имеет длину 2-50 аминокислот, имеет длину 2-30 аминокислот.

Наиболее часто используемые гибкие линкеры имеют последовательности, состоящие главным образом из отрезков из остатков Gly и Ser (линкеры "GS"). Гибкие линкеры могут быть применимы для соединения доменов, которые требуют определенной степени подвижности или взаимодействия, и могут содержать небольшие неполярные (например, Gly) или полярные (например, Ser или Thr) аминокислоты. Включение Ser или Thr может также обеспечивать поддержание стабильности линкера в водных растворах благодаря образованию водородных связей с молекулами воды и, следовательно, уменьшение неблагоприятных взаимодействий между линкером и другими компонентами. Примеры таких линкеров включают линкеры, имеющие структуру $[GGS]^{\geq 1}$ или $[GGGS]^{\geq 1}$ (SEQ ID NO: 1536). Жесткие линкеры применимы для сохранения фиксированного расстояния между доменами и поддержания их независимых функций. Жесткие линкеры также могут быть применимы, когда пространственное разделение доменов является критически важным для сохранения стабильности или биологической активности одного или нескольких компонентов в средстве. Жесткие линкеры могут иметь альфа-спиральную структуру или Pro-богатую последовательность (XP)n, при этом X обозначает любую аминокислоту, предпочтительно Ala, Lys или Glu. Расщепляемые линкеры могут высвобождать свободные функциональные домены in vivo. В некоторых вариантах осуществления линкеры могут расщепляться в специфических условиях, таких как присутствие восстанавливающих реагентов или протеаз. Расщепляемые in vivo линкеры могут использовать обратимую природу дисульфидной связи. Один пример включает в себя последовательность, чувствительную к тромбину (например, PRS), между двумя остатками Cys. Обработка тромбином in vitro CPRSC (SEQ ID NO: 1537) приводит к расщеплению последовательности, чувствительной к тромбину, тогда как обратимая дисульфидная связь остается интактной. Такие линкеры известны и описаны, например, в Chen et al., 2013. Fusion Protein Linkers: Property, Design and Functionality. Adv Drug Deliv Rev. 65(10): 1357-1369. Расщепление линкеров in vivo в композициях, описанных в данном документе, может также осуществляться протеазами, которые экспрессируются in vivo при

патологических состояниях (например, раке или воспалении), в определенных клетках или тканях или ограничены определенными клеточными компартментами. Специфичность многих протеаз обеспечивает более медленное расщепление линкера в ограниченных компартментах.

В некоторых вариантах осуществления аминокислотные линкеры представляют собой эндогенные аминокислоты (или гомологичны им), которые существуют между такими доменами в нативном полипептиде. В некоторых вариантах осуществления эндогенные аминокислоты, которые существуют между такими доменами, заменены, но длина не изменена по сравнению с естественной длиной. В некоторых вариантах осуществления к существующим в естественных условиях аминокислотным остаткам между доменами добавлены дополнительные аминокислотные остатки.

В некоторых вариантах осуществления аминокислотные линкеры разработаны вычислительным путем или подвергнуты скринингу для максимального увеличения функции белка (Anad et al., FEBS Letters, 587:19, 2013).

В дополнение к тому, что полипептид полностью кодируется одним транскриптом, он может быть получен посредством отдельной экспрессии двух или более полипептидных фрагментов, из которых восстанавливается полный фермент. В некоторых вариантах осуществления полипептид Gene Writer получают путем экспрессии в виде отдельных субъединиц, из которых повторно собирается полный фермент посредством сконструированных белок-белковых взаимодействий. В некоторых вариантах осуществления восстановление полного фермента не предполагает ковалентное связывание между субъединицами. Пептиды также могут сливаться вместе посредством *транс*-сплайсинга интеинов (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). В некоторых вариантах осуществления полный фермент Gene Writer экспрессируется в виде отдельных субъединиц, которые предназначены для создания слитого белка за счет присутствия расщепленных интеинов в субъединицах. В некоторых вариантах осуществления полный фермент Gene Writer восстанавливается посредством образования ковалентных связей между субъединицами. В некоторых вариантах осуществления белковые субъединицы собираются повторно с помощью сконструированных партнеров по белок-белковому связыванию, например, SpyTag и SpyCatcher (Zakeri et al. PNAS 109, E690-E697 (2012)). В некоторых вариантах осуществления дополнительный домен, описанный в данном документе, например, никаза Cas9, экспрессируется в виде отдельного полипептида, который ассоциирует с полипептидом Gene Writer посредством ковалентных или нековалентных взаимодействий, описанных выше. В некоторых вариантах осуществления разделение полипептида Gene Writer на субъединицы может способствовать доставке белка за счет поддержания нуклеиновой кислоты, кодирующей каждую часть, в оптимальных пределах упаковки вирусного вектора для доставки, например, AAV (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). В некоторых вариантах осуществления полипептид Gene Writer предназначен для димеризации посредством использования ковалентных или нековалентных взаимодействий, описанных выше.

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 (см., например, пример 22). 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.

Известное явление заключается в том, что некоторые мобильные элементы способны перемещать чужеродные элементы, например, ретротранспозаза L1 облегчает перемещение неавтономных элементов Alu и SVA в человеческом геноме (Craig, Mobile DNA III, ASM, ed. 3 (2105)). В недавних исследованиях были картированы различные подвижные элементы, присутствующие в человеческом геноме, включая ретротранспозоны, не содержащие LTR (Kojima Mobile DNA 9 (2018)). С учетом того, что активная транспозиция в человеческом геноме, как было установлено, связана с заболеваниями, например, известна роль ретротранспозиции LINE-1 в онкогенезе (Rodriguez-Martin et al. Nat Genet (2020)), то является желательным, чтобы Gene Writer не распознавал и не мобилизовал подвижные элементы или псевдоэлементы. В некоторых вариантах осуществления полипептид Gene Writer не приводит к мобилизации какой-либо эндогенной ДНК человека. В некоторых вариантах осуществления Gene Writer получен из ретротранспозазы, которая отсутствует в человеческом геноме. В некоторых вариантах осуществления Gene Writer, полученный из ретротранспозазы, присутствующей в человеческом геноме (см., например, Kojima Mobile DNA 9 (2018)) сконструирован таким образом, что он распознает гетерологичные последовательности в РНК-матрице и больше не распознает природные UTR исходного ретротранспозона, например, имеет гетерологичный РНК-связывающий домен, который не ассоциирует с 3'-UTR, присутствующей в человеческом геноме. В некоторых вариантах осуществления Gene Writer содержит РНК-связывающий домен, который не распознает какие-либо последовательности, присутствующие в человеческом геноме.

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 PCT/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 нагрузкой. В одном аспекте 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 под SEQ ID NO: 3 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; DHFR человека (hDHFR), содержащего аминокислотную последовательность под SEQ ID NO: 2 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; DHFR *E. coli*, содержащего аминокислотную последовательность под SEQ ID NO: 1 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; PDE5, содержащего аминокислотную последовательность под SEQ ID NO: 4 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; PPAR-гамма, содержащего аминокислотную последовательность под SEQ ID NO: 5 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; CA2, содержащего аминокислотную последовательность под SEQ ID NO: 6 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте; или NQ02, содержащего аминокислотную последовательность под SEQ ID NO: 7 из документа PCT/US2018/020704, включенного в данный документ посредством ссылки во всей своей полноте. 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 нагрузки, интегрированной в геном. В определенных вариантах осуществления нагрузка, содержащая DD, может представлять собой терапевтический белок, например, функциональную копию подвергнутого эндогенной мутации гена. В определенных вариантах осуществления нагрузка, содержащая DD, может представлять собой гетерологичный белок, например, CAR.

При применении в системах и способах, представленных в данном документе, Gene Writer™ могут быть представлены либо в виде полипептидов, либо в виде кодирующих их нуклеиновых кислот.

Признаки нуклеиновых кислот

Элементы систем, предусмотренных настоящим изобретением, могут быть представлены в виде нуклеиновых кислот, например, нуклеиновой кислоты-матрицы (например, РНК-матрицы), как описано, среди прочего, в формуле изобретения и пронумерованных вариантах осуществления, а также, в определенных вариантах осуществления, нуклеиновой кислоты, кодирующей полипептид Gene Writer™ (например, ретротранспозазу). 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, 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 однонитевую, двухнитевую, кольцевую, линейную, линейную с незамкнутыми концами, линейную с замкнутыми концами и их конкретные варианты, такие как ДНК Doggybone (dbDNA), ДНК с замкнутыми концами (ceDNA).

В определенных вариантах осуществления тканеспецифическая(тканеспецифические) последовательность(последовательности), контролирующая(контролирующие) экспрессию, относится(относятся) к одной или нескольким последовательностям в: таблице 3 документа WO2020014209, включенного в данный документ посредством ссылки, исключая ее последний столбец (эталонный SEQ ID NO); или таблице 4 документа WO2020014209, включенного в данный документ посредством ссылки, исключая ее последний столбец (эталонный SEQ ID NO).

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, нуклеиновая кислота-матрица или матрица, кодирующая ретротранспозазу), содержит промоторную последовательность, например, тканеспецифический промотор. 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 нуклеиновая кислота, кодирующая

полипептид, была доставлена в клетку, не являющуюся мишенью, она не будет управлять экспрессией (или будет управлять экспрессией только на низком уровне) ретротранспозазы, что ограничивает интеграцию РНК-матрицы. Система, содержащая последовательность тканеспецифического промотора в ДНК ретротранспозазы, также может применяться в комбинации с сайтом связывания с микроРНК, например, закодированным в ДНК ретротранспозазы, например, как описано в данном документе. Система, содержащая последовательность тканеспецифического промотора в ДНК ретротранспозазы, также может применяться в комбинации с РНК-матрицей, содержащей последовательность гетерологичного объекта под управлением тканеспецифического промотора, например, для достижения более высоких уровней интеграции и экспрессии последовательности гетерологичного объекта в клетках-мишенях, чем в клетках, не являющихся мишенями.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК, кодирующая полипептид Gene Writer™, или ДНК, кодирующая РНК, или нуклеиновая кислота-матрица), 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, вставка последовательности гетерологичного объекта в геном клеток-мишеней будет происходить более эффективно, чем в геном клеток, не являющихся мишенями. Система, имеющая сайт связывания с микроРНК в РНК, кодирующей полипептид Gene Writer™ (или кодируемый в ДНК, кодирующей РНК), может также применяться в комбинации с РНК-матрицей, регулируемой вторым сайтом связывания с микроРНК, например, как описано в данном документе в разделе под названием "Компонент системы редактирования генов Gene Writer™, представляющий собой матрицу".

В некоторых вариантах осуществления компонент, представляющий собой нуклеиновую кислоту, в системе, предусмотренной по настоящему изобретению, представляет собой последовательность (например, ретротранспозазу или последовательность гетерологичного объекта), фланкированную нетранслируемыми областями (UTR), которые модифицируют уровни экспрессии белка (иногда называемыми

UTR_{exp}) (фигуры 11 и 15, пример 6). Эффекты различных 5'- и 3'-UTR в отношении экспрессии белка известны из уровня техники. Например, в некоторых вариантах осуществления кодирующей последовательности может предшествовать 5'-UTR, которая модифицирует стабильность РНК или трансляцию белка. В некоторых вариантах осуществления за последовательностью может следовать 3'-UTR, которая модифицирует стабильность РНК или трансляцию белка. В некоторых вариантах осуществления последовательности может предшествовать 5'-UTR, и за ней может следовать 3'-UTR, которые модифицируют стабильность РНК или трансляцию. В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны из 5'- и 3'-UTR фактора комплемента 3 (C3) (cactcctccccatcctctccctctgtccctctgtccctctgaccctgcactgtcccagcacc (SEQ ID NO: 1606)) или орозомукоида 1 (ORM1) (caggacacagccttggatcaggacagagacttgggggccatcctgcccctccaacccgacatgtgtacctcagctttttccctcacttgcat caataaagcttctgtgtttggaacagctaa (SEQ ID NO: 1607)) (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: 1608), и/или 3'-UTR содержит UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1609), например, как описано в Richner et al. *Cell* 168(6): P1114-1125 (2017), последовательности из которого включены в данный документ посредством ссылки.

В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны для усиления экспрессии белка. В некоторых вариантах осуществления 5'- и/или 3'-UTR могут быть выбраны для модификации экспрессии белка таким образом, чтобы минимизировать ингибирование избыточного продуцирования. В некоторых вариантах осуществления UTR расположены вокруг кодирующей последовательности, например, вне кодирующей последовательности, и в других вариантах осуществления находятся проксимальнее кодирующей последовательности. В некоторых вариантах осуществления в UTR включены дополнительные регуляторные элементы (например, сайты связывания с miRNA, цис-регуляторные сайты).

В некоторых вариантах осуществления открытая рамка считывания (ORF) системы Gene Writer, например, ORF мРНК (или ДНК, кодирующей мРНК), кодирующая полипептид Gene Writer, или одна или несколько ORF мРНК (или ДНК, кодирующей мРНК) в последовательности гетерологичного объекта фланкированы 5'- и/или 3'-нетранслируемой областью (UTR), которая усиливает их экспрессию. В некоторых вариантах осуществления 5'-UTR мРНК-компонента (или транскрипта, полученного из

ДНК-компонента) системы содержит последовательность 5'-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3' (SEQ ID NO: 1610). В некоторых вариантах осуществления 3'-UTR мРНК-компонента (или транскрипта, полученного из ДНК-компонента) системы содержит последовательность 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3' (SEQ ID NO: 1611). Данная комбинация 5'-UTR и 3'-UTR приводит к желаемой экспрессии функционально связанной ORF, как было показано в Richner et al. *Cell* 168(6): P1114-1125 (2017), идеи и последовательности из которого включены в данный документ посредством ссылки. В некоторых вариантах осуществления система, описанная в данном документе, содержит ДНК, кодирующую транскрипт, где ДНК содержит соответствующие последовательности 5'-UTR и 3'-UTR, где T заменяет U в вышеперечисленной последовательности). В некоторых вариантах осуществления ДНК-вектор, применяемый для получения РНК-компонента системы, дополнительно содержит промотор для инициации транскрипции in vitro выше 5'-UTR, например, промотор T7, T3 или SP6. Вышеуказанная 5'-UTR начинается с GGG, что является подходящим сайтом начала для оптимизации транскрипции с применением РНК-полимеразы T7. Для настройки уровней транскрипции и изменения нуклеотидов в сайте начала транскрипции таким образом, чтобы они соответствовали альтернативным 5'-UTR, в идеях Davidson et al. *Pac Symp Biocomput* 433-443 (2010) описаны варианты промотора T7 и способы их обнаружения, которые соответствуют обоим этим признакам.

Кольцевые РНК в системах 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, линеаризуется (например, в клетке-хозяине) перед трансляцией.

В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид Gene Writer, или РНК-матрицу, или и то, и другое) представлена в виде circRNA. В некоторых вариантах осуществления circRNA содержит одну или несколько последовательностей рибозима. В некоторых вариантах осуществления последовательность рибозима активируется для саморасщепления, например, в клетке-хозяине, например, что таким образом приводит к линеаризации circRNA. В некоторых вариантах осуществления рибозим активируется, когда концентрация магния достигает уровня, достаточного для расщепления, например, в клетке-хозяине. В некоторых вариантах осуществления circRNA поддерживается в среде с низким содержанием магния перед доставкой в клетку-хозяина. В некоторых вариантах осуществления рибозим представляет собой рибозим, отвечающий на белки. В некоторых вариантах осуществления рибозим представляет собой рибозим, отвечающий на нуклеиновые кислоты.

В некоторых вариантах осуществления circRNA линеаризуется в ядре клетки-мишени. В некоторых вариантах осуществления линеаризация circRNA в ядре клетки включает компоненты, присутствующие в ядре клетки, например, для активации события расщепления. Например, ретротранспозоны B2 и ALU содержат саморасщепляющиеся рибозимы, активность которых усиливается при взаимодействии с белком Polycomb EZH2 (Hernandez et al. PNAS 117(1):415-425 (2020)). Таким образом, в некоторых вариантах осуществления рибозим, например, рибозим из элемента B2 или ALU, который отвечает на ядерный элемент, например, ядерный белок, например, белок, взаимодействующий с геномом, например, эпигенетический модификатор, например, EZH2, включен в circRNA, например, системы Gene Writing. В некоторых вариантах осуществления ядерная локализация circRNA приводит к увеличению автокаталитической активности рибозима и линеаризации circRNA.

В некоторых вариантах осуществления 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, тРНК, snRNA или mtRNA. В некоторых вариантах осуществления нуклеиновая кислота, которая запускает линеаризацию, присутствует в целевых клетках на более высоких уровнях, чем в нецелевых клетках.

В некоторых вариантах осуществления любого из аспектов в данном документе система Gene Writing включает один или несколько рибозимов с индуцируемой специфичностью в отношении ткани-мишени или клетке-мишени, представляющим интерес, например, рибозим, который активируется лигандом или нуклеиновой кислотой, присутствующими на более высоких уровнях в ткани-мишени или клетке-мишени, представляющих интерес. В некоторых вариантах осуществления система Gene Writing включает рибозим с индуцируемой специфичностью в отношении субклеточного компартмента, например, ядра, ядрышка, цитоплазмы или митохондрий. В некоторых вариантах осуществления рибозим, который активируется лигандом или нуклеиновой

кислотой, присутствует на более высоких уровнях в целевом субклеточном компартменте. В некоторых вариантах осуществления РНК-компонент системы Gene Writing представлен в виде circRNA, например, которая активируется путем линеаризации. В некоторых вариантах осуществления линеаризация circRNA, кодирующей полипептид Gene Writing, активирует молекулу для трансляции. В некоторых вариантах осуществления сигнал, который активирует circRNA-компонент системы Gene Writing, присутствует на более высоких уровнях в целевых клетках или тканях, например, так что система специфично активируется в данных клетках.

В некоторых вариантах осуществления РНК-компонент системы Gene Writing представлен в виде circRNA, которая инактивируется путем линеаризации. В некоторых вариантах осуществления circRNA, кодирующая полипептид Gene Writer, инактивируется путем расщепления и разрушения. В некоторых вариантах осуществления circRNA, кодирующая полипептид Gene Writing, инактивируется путем расщепления, при котором сигнал трансляции отделяется от последовательности, кодирующей полипептид. В некоторых вариантах осуществления сигнал, который инактивирует circRNA-компонент системы Gene Writing, присутствует на более высоких уровнях в нецелевых клетках или тканях, так что система специфично инактивируется в данных клетках.

В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид Gene Writer или кодирующая РНК-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 нуклеиновая кислота (например, кодирующая полипептид Gene Writer или кодирующая РНК-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 кодирующую РНК-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 Gene Writer или кодирующая РНК-матрицу или и то, и другое), доставляемая в клетки, представляет собой линейный ДНК-дуплекс с замкнутыми концами (CELiD-ДНК или ceDNA). В некоторых вариантах осуществления ceDNA получена из репликативной формы генома AAV (Li et al. PLoS One 2013). В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид Gene Writer или кодирующая РНК-матрицу или и то, и другое) фланкирована ITR, например, ITR AAV, где по меньшей мере одна из ITR содержит сайт концевого разрешения и сайт связывания белка репликации (иногда называемый сайтом связывания репликативного белка). В некоторых вариантах осуществления ITR получены из аденоассоциированного вируса, например, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12 или их комбинации. В некоторых вариантах осуществления ITR являются симметричными. В некоторых вариантах осуществления ITR являются асимметричными. В некоторых вариантах осуществления предусмотрен по меньшей мере один белок Rep, обеспечивающий репликацию конструкции. В некоторых вариантах осуществления по меньшей мере один белок Rep получен из аденоассоциированного вируса, например, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12 или их комбинации. В некоторых вариантах осуществления ceDNA получают посредством создания клетки-продуцента с (i) ДНК, фланкированной ITR, например, ITR AAV, и (ii) компонентами, необходимыми для ITR-зависимой репликации, например, белками AAV Rep78 и Rep52 (или нуклеиновой кислотой, кодирующей данные белки). В некоторых вариантах осуществления ceDNA не содержит какого-либо капсидного белка, например, не упакована в инфекционную частицу AAV. В некоторых вариантах осуществления ceDNA составляют в LNP (см., например, WO2019051289A1).

В некоторых вариантах осуществления ceDNA-вектор состоит из двух самокомплементарных последовательностей, например, асимметричных, или симметричных, или по сути симметричных ITR, определенных в данном документе, фланкирующих указанную кассету экспрессии, где ceDNA-вектор не ассоциирован с капсидным белком. В некоторых вариантах осуществления ceDNA-вектор содержит две самокомплементарные последовательности, обнаруженные в геноме AAV, где по меньшей мере одна ITR содержит действующий Rep-связывающий элемент (RBE) (также иногда называемый в данном документе "RBS") и сайт концевого разрешения (trs) AAV или функциональный вариант RBE. См., например, WO2019113310.

В некоторых вариантах осуществления нуклеиновая кислота (например, кодирующая полипептид Gene Writer или кодирующая РНК-матрицу или и то, и другое),

доставляемая в клетки, разработана в виде миниколец, где последовательности остова плазмиды, не относящиеся к Gene Writer™, удаляются перед введением в клетки. 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, кодирующий матрицу Gene Writer™, 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, кассеты экспрессии полипептида Gene Writer или кассеты экспрессии РНК-матрицы, из каркасной последовательности 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, rAAV, 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™ или кодирующей РНК-матрицу или и то, и другое.

Вирусные векторы и их компоненты

Вирусы являются полезным источником средств доставки для систем, описанных в данном документе, в дополнение к источнику соответствующих ферментов или доменов, описанных в данном документе, например, в качестве источников полимераз и функций полимеразы, используемых в данном документе, например, ДНК-зависимой ДНК-полимеразы, РНК-зависимой РНК-полимеразы, РНК-зависимой ДНК-полимеразы, ДНК-зависимой РНК-полимеразы, обратной транскриптазы. Некоторые ферменты, например, обратные транскриптазы, могут характеризоваться несколькими видами активности, например, быть способными как к РНК-зависимой полимеризации ДНК, так и к ДНК-зависимой полимеризации ДНК, например, к синтезу первой и второй нитей. В некоторых вариантах осуществления вирус, используемый в качестве системы доставки Gene Writer

или источника ее компонентов, может быть выбран из группы, описанной в Baltimore *Bacteriol Rev* 35(3):235-241 (1971).

В некоторых вариантах осуществления вирус выбран из вируса группы I, например, представляет собой ДНК-вирус и упаковывает dsDNA в вирионы. В некоторых вариантах осуществления вирус группы I выбран, например, из аденовирусов, герпесвирусов, поксвирусов.

В некоторых вариантах осуществления вирус выбран из вируса группы II, например, представляет собой ДНК-вирус и упаковывает ssDNA в вирионы. В некоторых вариантах осуществления вирус группы II выбран, например, из парвовирусов. В некоторых вариантах осуществления парвовирус представляет собой депендопарвовирус, например, аденоассоциированный вирус (AAV).

В некоторых вариантах осуществления вирус выбран из вируса группы III, например, представляет собой РНК-вирус и упаковывает dsRNA в вирионы. В некоторых вариантах осуществления вирус группы III выбран, например, из реовирусов. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, представляют собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, могут непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты.

В некоторых вариантах осуществления вирус выбран из вируса группы IV, например, представляет собой РНК-вирус и упаковывает ssRNA(+) в вирионы. В некоторых вариантах осуществления вирус группы IV выбран, например, из коронавирусов, пикорнавирусов, тогавирусов. В некоторых вариантах осуществления ssRNA(+), содержащаяся в таких вирионах, представляет собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, может непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты.

В некоторых вариантах осуществления вирус выбран из вируса группы V, например, представляет собой РНК-вирус и упаковывает ssRNA(-) в вирионы. В некоторых вариантах осуществления вирус группы V выбран, например, из ортомиксовирусов, рабдовирусов. В некоторых вариантах осуществления РНК-вирус с ssRNA(-)-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую РНК-полимеразу, способную копировать ssRNA(-) в ssRNA(+), которая может транслироваться непосредственно хозяином.

В некоторых вариантах осуществления вирус выбран из вируса группы VI, например, представляет собой ретровирус и упаковывает ssRNA(+) в вирионы. В некоторых вариантах осуществления вирус группы VI выбран, например, из ретровирусов. В некоторых вариантах осуществления ретровирус представляет собой лентивирус, например, HIV-1, HIV-2, SIV, BIV. В некоторых вариантах осуществления ретровирус

представляет собой спумавирус, например, вирус пенистости, например, HFV, SFV, BFV. В некоторых вариантах осуществления ssRNA(+), содержащаяся в таких вирионах, представляет собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, может непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты. В некоторых вариантах осуществления ssRNA(+) сначала подвергается обратной транскрипции и копированию с образованием промежуточного dsDNA-генома, из которого мРНК может транскрибироваться в клетке-хозяине. В некоторых вариантах осуществления РНК-вирус с ssRNA(-)-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую ДНК-полимеразу, способную копировать ssRNA(+) в dsDNA, которая может транскрибироваться в мРНК и транслироваться хозяином. В некоторых вариантах осуществления обратная транскриптаза из ретровируса группы VI включена в качестве домена обратной транскриптазы полипептида Gene Writer.

В некоторых вариантах осуществления вирус выбран из вируса группы VII, например, представляет собой ретровирус и упаковывает dsRNA в вирионы. В некоторых вариантах осуществления вирус группы VII выбран, например, из гепаднавирусов. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, представляют собой кодирующую молекулу, способную непосредственно служить в качестве мРНК при трансдукции клетки-хозяина, например, могут непосредственно транслироваться в белок при трансдукции клетки-хозяина, не требуя каких-либо промежуточных стадий репликации или полимеризации нуклеиновой кислоты. В некоторых вариантах осуществления одна или обе нити dsRNA, содержащиеся в таких вирионах, сначала подвергаются обратной транскрипции и копированию с образованием промежуточного dsDNA-генома, из которого мРНК может транскрибироваться в клетке-хозяине. В некоторых вариантах осуществления РНК-вирус с dsRNA-геномом также несет фермент внутри вириона, который вводится путем трансдукции в клетки-хозяева с вирусным геномом, например, РНК-зависимую ДНК-полимеразу, способную копировать dsRNA в dsDNA, которая может транскрибироваться в мРНК и транслироваться хозяином. В некоторых вариантах осуществления обратная транскриптаза из ретровируса группы VII включена в качестве домена обратной транскриптазы полипептида Gene Writer.

В некоторых вариантах осуществления вирионы, используемые для доставки нуклеиновой кислоты в настоящем изобретении, могут также нести ферменты, участвующие в процессе Gene Writing. Например, ретровирусный вирион может содержать домен обратной транскриптазы, который доставляется в клетку-хозяина вместе с нуклеиновой кислотой. В некоторых вариантах осуществления РНК-матрица может быть ассоциирована с полипептидом Gene Writer внутри вириона, так что оба они совместно доставляются в клетку-мишень при трансдукции нуклеиновой кислоты из вирусной частицы. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может

содержать ДНК, например, линейную ssDNA, линейную dsDNA, кольцевую ssDNA, кольцевую dsDNA, миникольцевую DNA, dbDNA, ceDNA. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может содержать РНК, например, линейную ssRNA, линейную dsRNA, кольцевую ssRNA, кольцевую dsRNA. В некоторых вариантах осуществления вирусный геном может циркуляризоваться при трансдукции клетки-хозяина, например, линейная молекула ssRNA может участвовать в ковалентной связи с образованием кольцевой ssRNA, линейная молекула dsRNA может участвовать в ковалентной связи с образованием кольцевой dsRNA или одной или нескольких кольцевых ssRNA. В некоторых вариантах осуществления вирусный геном может реплицироваться путем репликации по типу катящегося кольца в клетке-хозяине. В некоторых вариантах осуществления вирусный геном может содержать одну молекулу нуклеиновой кислоты, например, содержать несегментированный геном. В некоторых вариантах осуществления вирусный геном может содержать две или более молекулы нуклеиновых кислот, например, содержать сегментированный геном. В некоторых вариантах осуществления нуклеиновая кислота в вирионе может быть ассоциирована с одним или несколькими белками. В некоторых вариантах осуществления один или несколько белков вириона могут быть доставлены в клетку-хозяина при трансдукции. В некоторых вариантах осуществления природный вирус может быть адаптирован для доставки нуклеиновой кислоты посредством добавления сигналов упаковки вириона к нуклеиновой кислоте-мишени, где клетка-хозяин используется для упаковки нуклеиновой кислоты-мишени, содержащей сигналы упаковки.

В некоторых вариантах осуществления вирион, используемый в качестве средства доставки, может содержать комменсальный вирус человека. В некоторых вариантах осуществления вирион, используемый в качестве средства доставки, может содержать анелловирус, применение которого описано в WO2018232017A1, которая включена в данный документ посредством ссылки во всей своей полноте.

Аденоассоциированные вирусы

В некоторых вариантах осуществления вирус представляет собой аденоассоциированный вирус (AAV). В некоторых вариантах осуществления геном AAV содержит два гена, которые кодируют четыре белка репликации и три капсидных белка соответственно. В некоторых вариантах осуществления гены фланкированы с обеих сторон инвертированными концевыми повторами (ITR) размером 145 п. о. В некоторых вариантах осуществления вирион содержит до трех капсидных белков (Vp1, Vp2 и/или Vp3), например, продуцируемых в соотношении 1:1:10. В некоторых вариантах осуществления капсидные белки получены из одной и той же открытой рамки считывания и/или из сайта дифференциального сплайсинга (Vp1) и альтернативных сайтов начала трансляции (Vp2 и Vp3 соответственно). В целом Vp3 является наиболее распространенной субъединицей вириона и принимает участие в распознавании рецепторов на клеточной поверхности, определяя тропизм вируса. В некоторых вариантах осуществления Vp1 содержит фосфолипазный домен, например, который функционирует для обеспечения инфекционности вируса, на N-конце Vp1.

В некоторых вариантах осуществления упаковочная способность вирусных векторов ограничивает размер редактора оснований, который может быть упакован в вектор. Например, упаковочная способность AAV может составлять приблизительно 4,5 т. о. (например, приблизительно 3,0, 3,5, 4,0, 4,5, 5,0, 5,5 или 6,0 т. о.), например, включая один или два инвертированных концевых повтора (ITR), например, ITR размером 145 оснований.

В некоторых вариантах осуществления рекомбинантный AAV (rAAV) содержит цис-действующие ITR размером 145 п. о., фланкирующие кассеты трансгена в векторе, например, обеспечивая до 4,5 т. о. для упаковки чужеродной ДНК. После инфицирования rAAV в некоторых случаях может экспрессировать слитый белок по настоящему изобретению и персистировать без интеграции в геном хозяина, существуя эписомально в виде кольцевых конкатемеров "голова к хвосту". rAAV можно использовать, например, in vitro и in vivo. В некоторых вариантах осуществления опосредованная AAV доставка гена требует, чтобы длина кодирующей последовательности гена была равна геному AAV дикого типа или превышала его по размеру.

Доставка с помощью AAV генов, которые превышают этот размер, и/или применение 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) (1989) (включенных в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления Gene Writer, описанный в данном документе (например, с одной или несколькими направляющими нуклеиновыми кислотами или без них), может быть доставлен с применением AAV, лентивируса, аденовируса или других типов плазмидных или вирусных векторов, в частности, с использованием составов и доз из, например, патента США № 8454972 (составы, дозы для аденовируса), патента США № 8404658 (составы, дозы для AAV) и патента США № 5846946 (составы, дозы для ДНК-плазмид), а также из клинических испытаний и публикаций, касающихся клинических испытаний с участием лентивирусов, AAV и аденовирусов. Например, для AAV путь введения, состав и доза могут быть такими, как описано в патенте США № 8454972 и в клинических испытаниях с использованием AAV. Для аденовируса путь введения, состав и доза могут быть такими, как описано в патенте США № 8404658 и в клинических испытаниях с использованием аденовируса. Для доставки плазмиды путь введения, состав и доза могут быть такими, как описано в патенте США № 5846946 и в клинических испытаниях с использованием плазмид. Дозы могут быть рассчитаны на основе средней массы индивидуума 70 кг (например, взрослого человека мужского пола) или экстраполированы на нее и могут быть скорректированы для пациентов, субъектов, млекопитающих разной массы и разных видов. Частота введения регламентируется практикующим врачом или ветеринаром (например, терапевтом, ветеринаром) в зависимости от обычных факторов, включая возраст, пол, общее состояние здоровья, другие состояния пациента или субъекта и конкретное состояние или симптомы, на которые направлено лечение. В некоторых вариантах осуществления вирусные векторы можно инъецировать в ткань, представляющую интерес. Для Gene Writing, специфичной для типа клеток, экспрессия Gene Writer и необязательной направляющей нуклеиновой кислоты в некоторых вариантах осуществления может находиться под управлением промотора, специфичного для типа клеток.

В некоторых вариантах осуществления AAV обеспечивает низкую токсичность, например, благодаря способу очистки, не требующему ультрацентрифугирования клеточных частиц, которые могут активировать иммунный ответ. В некоторых вариантах осуществления AAV обеспечивает низкую вероятность приведения к инсерционному мутагенезу, например, потому, что он по сути не интегрируется в геном хозяина.

В некоторых вариантах осуществления AAV имеет предел упаковки, составляющий приблизительно 4,4, 4,5, 4,6, 4,7 или 4,75 т. о. В некоторых вариантах осуществления Gene Writer, промотор и терминатор транскрипции могут входить в один вирусный вектор. SpCas9 (4,1 т. о.) в некоторых случаях может быть сложно упаковать в AAV. Следовательно, в некоторых вариантах осуществления применяется Gene Writer, который короче по длине, чем другие Gene Writer или редакторы оснований. В некоторых вариантах осуществления Gene Writer составляет менее чем приблизительно 4,5 т. о., 4,4 т. о., 4,3 т. о., 4,2 т. о., 4,1 т. о., 4 т. о., 3,9 т. о., 3,8 т. о., 3,7 т. о., 3,6 т. о., 3,5 т. о., 3,4 т. о., 3,3 т. о., 3,2 т. о., 3,1 т. о., 3 т. о., 2,9 т. о., 2,8 т. о., 2,7 т. о., 2,6 т. о., 2,5 т. о., 2 т. о. или 1,5 т. о.

AAV может представлять собой AAV1, AAV2, AAV5 или любую их комбинацию. В некоторых вариантах осуществления тип AAV выбирают с учетом клеток, подлежащих нацеливанию; например, серотипы 1, 2, 5 AAV или гибридный капсид AAV1, AAV2, AAV5, или любая их комбинация могут быть выбраны для нацеливания на клетки головного мозга или нервные клетки; или AAV4 может быть выбран для нацеливания на сердечную ткань. В некоторых вариантах осуществления AAV8 выбирают для доставки в печень. Иллюстративные серотипы AAV, используемые в отношении данных клеток, описаны, например, в Grimm, D. et al, J. Virol. 82: 5887-5911 (2008) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления AAV относится ко всем серотипам, подтипам и встречающимся в природе AAV, а также к рекомбинантным AAV. AAV может применяться для обозначения самого вируса или его производного. В некоторых вариантах осуществления AAV включает AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64R1, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrh10, AAVLK03, AV10, AAV11, AAV 12, rh10 и их гибриды, AAV птиц, AAV крупного рогатого скота, AAV собак, AAV лошадей, AAV приматов, AAV животных, отличных от приматов, и AAV овец. Геномные последовательности различных серотипов AAV, а также последовательности нативных концевых повторов (TR), белков Rep и субъединиц капсида известны из уровня техники. Такие последовательности можно найти в литературе или в общедоступных базах данных, таких как GenBank. Дополнительные иллюстративные серотипы AAV указаны в таблице 36.

Таблица 36. Иллюстративные серотипы AAV

Ткань-мишень	Средство доставки	Литературный источник
Печень	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2, NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4] Аденовирус (Ad5, HC-AdV[6])	1. Wang *et al.*, *Mol. Ther.* **18**, 118-25 (2010) 2. Ginn *et al.*, *JHEP Reports*, 100065 (2019) 3. Paulk *et al.*, *Mol. Ther.* **26**, 289-303 (2018).

		4. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014). 5. L. Wang *et al.*, *Mol. Ther.* **23**, 1877-87 (2015). 6. Hausl *Mol Ther* (2010)
Легкое	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) Аденовирус (Ad5, Ad3, Ad21, Ad14)[3]	1. Duncan *et al.*, *Mol Ther Methods Clin Dev* (2018) 2. Cooney *et al.*, Am J Respir Cell Mol Biol (2019) 3. Li *et al.*, *Mol Ther Methods Clin Dev* (2019)
Кожа	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010) 2. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014).
HSC	Аденовирус (HDAd5/35^{++})	Wang et al. *Blood Adv* (2019)

В некоторых вариантах осуществления фармацевтическая композиция (например, содержащая AAV, описанный в данном документе) содержит менее 10% пустых капсидов, менее 8% пустых капсидов, менее 7% пустых капсидов, менее 5% пустых капсидов, менее 3% пустых капсидов или менее 1% пустых капсидов. В некоторых вариантах осуществления фармацевтическая композиция содержит менее чем приблизительно 5% пустых капсидов. В некоторых вариантах осуществления количество пустых капсидов находится ниже предела выявления. В некоторых вариантах осуществления преимущественно, чтобы фармацевтическая композиция содержала небольшое количество пустых капсидов, например, поскольку пустые капсиды могут вызывать нежелательный ответ (например, иммунный ответ, воспалительный ответ, печеночный ответ и/или сердечный ответ), например, при незначительном или отсутствующем терапевтическом эффекте.

В некоторых вариантах осуществления содержание остаточного белка клеток-хозяев (rHCP) в фармацевтической композиции меньше или равняется 100 нг/мл rHCP на 1×10^{13} vg/мл, например, меньше или равняется 40 нг/мл rHCP на 1×10^{13} vg/мл или 1-50 нг/мл rHCP на 1×10^{13} vg/мл. В некоторых вариантах осуществления фармацевтическая композиция содержит менее 10 нг rHCP на $1{,}0\times10^{13}$ vg или менее 5 нг rHCP на $1{,}0\times10^{13}$ vg, менее 4 нг rHCP на $1{,}0\times10^{13}$ vg или менее 3 нг rHCP на $1{,}0\times10^{13}$ vg или любое промежуточное значение концентрации. В некоторых вариантах осуществления содержание остаточной ДНК клеток-хозяев (hcDNA) в фармацевтической композиции меньше или равняется 5×10^{6} пг/мл hcDNA на 1×10^{13} vg/мл, меньше или равняется $1{,}2\times10^{6}$ пг/мл hcDNA на 1×10^{13} vg/мл или 1×10^{5} пг/мл hcDNA на 1×10^{13} vg/мл. В некоторых вариантах осуществления содержание остаточной ДНК клеток-хозяев в указанной фармацевтической композиции составляет

менее $5,0\times10^5$ пг на 1×10^{13} vg, менее $2,0\times10^5$ пг на $1,0\times10^{13}$ vg, менее $1,1\times10^5$ пг на $1,0\times10^{13}$ vg, менее $1,0\times10^5$ пг hcDNA на $1,0\times10^{13}$ vg, менее $0,9\times10^5$ пг hcDNA на $1,0\times10^{13}$ vg, менее $0,8\times10^5$ пг hcDNA на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации.

В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции меньше или равняется $1,7\times10^5$ пг/мл на $1,0\times10^{13}$ vg/мл, или 1×10^5 пг/мл на $1,0\times10^{13}$ vg/мл, или $1,7\times10^6$ пг/мл на $1,0\times10^{13}$ vg/мл. В некоторых вариантах осуществления содержание остаточной плазмидной ДНК в фармацевтической композиции составляет менее $10,0\times10^5$ пг на $1,0\times10^{13}$ vg, менее $8,0\times10^5$ пг на $1,0\times10^{13}$ vg или менее $6,8\times10^5$ пг на 1,0 x 10^{13} vg. В вариантах осуществления фармацевтическая композиция содержит менее 0,5 нг на $1,0\times10^{13}$ vg, менее 0,3 нг на $1,0\times10^{13}$ vg, менее 0,22 нг на $1,0\times10^{13}$ vg или менее 0,2 нг на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации бычьего сывороточного альбумина (BSA). В вариантах осуществления содержание бензоназы в фармацевтической композиции составляет менее 0,2 нг на $1,0\times10^{13}$ vg, менее 0,1 нг на $1,0\times10^{13}$ vg, менее 0,09 нг на $1,0\times10^{13}$ vg, менее 0,08 нг на $1,0\times10^{13}$ vg или любое промежуточное значение концентрации. В вариантах осуществления содержание полоксамера 188 в фармацевтической композиции составляет приблизительно 10-150 ppm, приблизительно 15-100 ppm или приблизительно 20-80 ppm. В вариантах осуществления содержание цезия в фармацевтической композиции составляет менее 50 пг/г (ppm), менее 30 пг/г (ppm) или менее 20 пг/г (ppm) или любое промежуточное значение концентрации.

В вариантах осуществления фармацевтическая композиция содержит общее количество примесей, например, как определено с помощью SDS-PAGE, составляющее менее 10%, менее 8%, менее 7%, менее 6%, менее 5%, менее 4%, менее 3%, менее 2% или любую промежуточную процентную долю. В вариантах осуществления общее количество примесей, например, как определено с помощью SDS-PAGE, составляет более 90%, более 92%, более 93%, более 94%, более 95%, более 96%, более 97%, более 98% или любую промежуточную процентную долю. В вариантах осуществления содержание никакой отдельной неназванной родственной примеси, например, как измерено с помощью SDS-PAGE, не составляет более 5%, более 4%, более 3% или более 2% или любой промежуточной процентной доли. В вариантах осуществления фармацевтическая композиция содержит процентную долю заполненных капсидов по отношению к общему количеству капсидов (например, пик 1+пик 2 согласно измерению с помощью аналитического ультрацентрифугирования), составляющую более 85%, более 86%, более 87%, более 88%, более 89%, более 90%, более 91%, более 91,9%, более 92%, более 93% или любую промежуточную процентную долю. В вариантах осуществления фармацевтической композиции процентная доля заполненных капсидов, измеренная в пике 1 с помощью аналитического ультрацентрифугирования, составляет 20-80%, 25-75%, 30-75%, 35-75% или 37,4-70,3%. В вариантах осуществления фармацевтической композиции процентная доля заполненных капсидов, измеренная в пике 2 с помощью аналитического ультрацентрифугирования, составляет 20-80%, 20-70%, 22-65%, 24-62% или 24,9-60,1%.

В одном варианте осуществления фармацевтическая композиция содержит титр геномов, составляющий от 1,0 до $5,0\times10^{13}$ vg/мл, от 1,2 до $3,0\times10^{13}$ vg/мл или от 1,7 до $2,3\times10^{13}$ vg/мл. В одном варианте осуществления фармацевтическая композиция демонстрирует биологическую нагрузку, составляющую менее 5 КОЕ/мл, менее 4 КОЕ/мл, менее 3 КОЕ/мл, менее 2 КОЕ/мл или менее 1 КОЕ/мл или любое промежуточное значение концентрации. В вариантах осуществления количество эндотоксина в соответствии с USP, например, USP <85> (включенной посредством ссылки во всей своей полноте), составляет менее 1,0 EU/мл, менее 0,8 EU/мл или менее 0,75 EU/мл. В вариантах осуществления осмолярность фармацевтической композиции в соответствии с USP, например, USP <785> (включенной посредством ссылки во всей своей полноте), составляет от 350 до 450 мОсм/кг, от 370 до 440 мОсм/кг или от 390 до 430 мОсм/кг. В вариантах осуществления фармацевтическая композиция содержит менее 1200 частиц размером более 25 мкм на контейнер, менее 1000 частиц размером более 25 мкм на контейнер, менее 500 частиц размером более 25 мкм на контейнер или любое промежуточное значение. В вариантах осуществления фармацевтическая композиция содержит менее 10000 частиц размером более 10 мкм на контейнер, менее 8000 частиц размером более 10 мкм на контейнер или менее 600 частиц размером более 10 пм на контейнер.

В одном варианте осуществления фармацевтическая композиция имеет титр геномов, составляющий от 0,5 до $5,0\times10^{13}$ vg/мл, от 1,0 до $4,0\times10^{13}$ vg/мл, от 1,5 до $3,0\times10^{13}$ vg/мл или от 1,7 до $2,3\times10^{13}$ vg/мл. В одном варианте осуществления фармацевтическая композиция, описанная в данном документе, характеризуется одним или несколькими из следующего: менее чем приблизительно 0,09 нг бензоназы на $1,0\times10^{13}$ vg, менее чем приблизительно 30 пг/г (ppm) цезия, приблизительно 20-80 ppm полоксамера 188, менее чем приблизительно 0,22 нг BSA на $1,0\times10^{13}$ vg, менее чем приблизительно $6,8\times10^{5}$ пг остаточной плазмидной ДНК на $1,0\times10^{13}$ vg, менее чем приблизительно $1,1\times10^{5}$ пг остаточной hcDNA на $1,0\times10^{13}$ vg, менее чем приблизительно 4 нг rHCP на $1,0\times10^{13}$ vg, pH от 7,7 до 8,3, приблизительно 390-430 мОсм/кг, менее чем приблизительно 600 частиц размером > 25 мкм на контейнер, менее чем приблизительно 6000 частиц размером > 10 мкм на контейнер, титр геномов от приблизительно $1,7\times10^{13}$ до $2,3\times10^{13}$ vg/мл, инфекционный титр от приблизительно $3,9\times10^{8}$ до $8,4\times10^{10}$ МЕ на $1,0\times10^{13}$ vg, содержание общего белка приблизительно 100-300 пг на $1,0\times10^{13}$ vg, средняя выживаемость > 24 дней у мышей A7SMA с дозой вирусного вектора приблизительно $7,5\times10^{13}$ vg/кг, относительная активность приблизительно 70-130% на основании клеточного анализа in vitro и/или менее чем приблизительно 5% пустых капсидов. В различных вариантах осуществления фармацевтические композиции, описанные в данном документе, содержат любую из вирусных частиц, обсуждаемых в данном документе, сохраняют активность в пределах ± 20%, ± 15%, ± 10% или ± 5% от эталонного стандарта. В некоторых вариантах осуществления активность измеряют с использованием подходящего клеточного анализа in vitro или животной модели in vivo.

Дополнительные способы получения, определения характеристик и введения доз частиц AAV описаны в WO 2019094253, которая включена в данный документ посредством ссылки во всей своей полноте.

Дополнительные конструкции rAAV, которые можно использовать согласно с настоящим изобретением, включают конструкции, описанные в Wang et al 2019, доступные по адресу: //doi.org/10.1038/s41573-019-0012-9, включая таблицу 1, которая включена посредством ссылки во всей своей полноте.

Интеины

В некоторых вариантах осуществления, которые более подробно описаны ниже, интеин-N может быть слит с N-концевой частью первого домена, описанного в данном документе, и интеин-C может быть слит с C-концевой частью второго домена, описанного в данном документе, для присоединения N-концевой части к C-концевой части с соединением таким образом первого и второго доменов. В некоторых вариантах осуществления каждый из первого и второго доменов независимо выбраны из ДНК-связывающего домена, РНК-связывающего домена, домена RT и эндонуклеазного домена.

Используемый в данном документе термин "интеин" относится к самовырезающемуся интрону белка (например, пептиду), например, который лигирует фланкирующие N-концевые и C-концевые экстеины (например, фрагменты, подлежащие соединению). В некоторых случаях интеин может содержать фрагмент белка, который способен вырезать себя и соединять оставшиеся фрагменты (экстеины) пептидной связью в процессе, известном как сплайсинг белка. Интеины также называют "интронами белков". Процесс самовырезания интеина и соединения оставшихся частей белка в данном документе называется "сплайсингом белка" или "интеин-опосредованным сплайсингом белка". В некоторых вариантах осуществления интеин белка-предшественника (белка, содержащего интеин, до интеин-опосредованного сплайсинга белка) происходит из двух генов. Такой интеин упоминается в данном документе как расщепленный интеин (например, расщепленный интеин-N и расщепленный интеин-C). Например, у цианобактерий DnaE, каталитическая a-субъединица ДНК-полимеразы III, кодируется двумя отдельными генами - dnaE-n и dnaE-c. Интеин, кодируемый геном dnaE-n, может называться в данном документе "интеином-N". Интеин, кодируемый геном dnaE-c, может называться в данном документе "интеином-C".

Применение интеинов для соединения гетерологичных фрагментов белка описано, например, в Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (включенном в данный документ посредством ссылки во всей своей полноте). Например, при слиянии с отдельными фрагментами белка интеины IntN и IntC могут распознавать друг друга, самовырезаться и/или одновременно лигировать фланкирующие N- и C-концевые экстеины фрагментов белка, с которыми они были слиты, тем самым восстанавливая полноразмерный белок из двух фрагментов белка.

В некоторых вариантах осуществления применяется синтетический интеин на основе интеина dnaE - пара интеинов Cfa-N (например, расщепленный интеин-N) и Cfa-C

(например, расщепленный интеин-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 Prp8 (например, описанные в патенте США № 8394604, включенном в данный документ посредством ссылки).

В некоторых вариантах осуществления интеин-N и интеин-С могут быть слиты с N-концевой частью расщепленного Cas9 и С-концевой частью расщепленного Cas9 соответственно для соединения N-концевой части расщепленного Cas9 и С-концевой части расщепленного Cas9. Например, в некоторых вариантах осуществления интеин-N слит с С-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-С слит с N-концом С-концевой части расщепленного Cas9, т. е. с образованием структуры N-[интеин-С]~[С-концевая часть расщепленного Cas9]-С. Механизм интеин-опосредованного сплайсинга белка для соединения белков, с которыми слиты интеины (например, расщепленный Cas9), описан в Shah et al., Chem Sci. 2014; 5(l):446-461, включенном в данный документ посредством ссылки. Способы разработки и применения интеинов известны из уровня техники и описаны, например, в WO2020051561, W02014004336, WO2017132580, US20150344549 и US20180127780, каждый из которых включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления расщепление относится к разделению на два или более фрагмента. В некоторых вариантах осуществления расщепленный белок Cas9 или расщепленный Cas9 содержит белок Cas9, который представлен в виде N-концевого фрагмента и С-концевого фрагмента, кодируемых двумя отдельными нуклеотидными последовательностями. Полипептиды, соответствующие N-концевой части и С-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 С, Т, А или S, например, в области SpCas9 между аминокислотами A292-G364, F445-K483 или E565-T637 или в соответствующих положениях любого другого Cas9, варианта Cas9 (например, nCas9, dCas9) или другого napDNAbp. В некоторых вариантах осуществления белок разделен на два фрагмента по Т310, Т313, А456, S469 или С574 SpCas9. В некоторых вариантах

осуществления процесс разделения белка на два фрагмента называется расщеплением белка.

В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 2-1000 аминокислот (например, 2-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900 или 900-1000 аминокислот). В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 5-500 аминокислот (например, 5-10, 10-50, 50-100, 100-200, 200-300, 300-400 или 400-500 аминокислот). В некоторых вариантах осуществления длина фрагмента белка находится в диапазоне приблизительно 20-200 аминокислот (например, 20-30, 30-40, 40-50, 50-100 или 100-200 аминокислот).

В некоторых вариантах осуществления часть или фрагмент Gene Writer (например, Cas9-R2Tg) слиты с интеином. Нуклеаза может быть слита с N-концом или C-концом интеина. В некоторых вариантах осуществления часть или фрагмент слитого белка слиты с интеином и слиты с капсидным белком AAV. Интеин, нуклеаза и капсидный белок могут быть слиты друг с другом в любой конфигурации (например, нуклеаза-интеин-капсид, интеин-нуклеаза-капсид, капсид-интеин-нуклеаза и т. д.). В некоторых вариантах осуществления N-конец интеина слит с C-концом слитого белка, и C-конец интеина слит с N-концом капсидного белка AAV.

В некоторых вариантах осуществления эндонуклеазный домен (например, никазный домен Cas9) слит с интеином-N, и полипептид, содержащий домен RT, слит с интеином-C.

Иллюстративные нуклеотидные и аминокислотные последовательности интеинов представлены ниже.

ДНК интеина-N DnaE:

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAA
TCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAAT
GGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGT
ATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAAT
TTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGG
ACCTCATGCGAGTTGACAACCTTCCTAAT (SEQ ID NO: 1612)

Белок интеин-N DnaE:

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEV
FEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1613)

ДНК интеина-C DnaE:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATA
TTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT
(SEQ ID NO: 1614)

Интеин-C:

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1615)

ДНК Cfa-N:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTAT TGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATG GTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTAT TTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTC ATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGA TCTCAAACAAGTGGATGGATTG CCA (SEQ ID NO: 1616)

Белок Cfa-N:

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVF EYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 1617)

ДНК Cfa-C:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGG AAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGG AGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ ID NO: 1618)

Белок Cfa-C:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 1619)

Липидные наночастицы

В способах и системах, предусмотренных в настоящем изобретении, может использоваться любой пригодный носитель или способ доставки, в том числе, в некоторых вариантах осуществления, липидные наночастицы (LNP). Липидные наночастицы в некоторых вариантах осуществления содержат один или несколько ионных липидов, таких как некатионные липиды (например, нейтральные, или анионные, или цвиттерионные липиды); один или несколько конъюгированных липидов (таких как липиды, конъюгированные с PEG, или липиды, конъюгированные с полимерами, описанные в таблице 5 документа WO 2019217941; включенного в данный документ посредством ссылки во всей своей полноте); один или несколько стеринов (например, холестерин) и необязательно одну или несколько нацеливающих молекул (например, конъюгированные рецепторы, лиганды рецепторов, антитела) или комбинации вышеперечисленного.

Липиды, которые можно применять для образования наночастиц (например, липидных наночастиц), включают, например, липиды, описанные в таблице 4 документа WO2019217941, который включен посредством ссылки, например, липидосодержащая наночастица может содержать один или несколько липидов из таблицы 4 документа WO2019217941. Липидные наночастицы могут содержать дополнительные элементы, такие как полимеры, такие как полимеры, описанные в таблице 5 документа WO2019217941, включенного посредством ссылки.

В некоторых вариантах осуществления конъюгированные липиды в случае их присутствия могут включать одно или несколько из PEG-диацилглицерина (DAG) (такого как 1-(монометоксиполиэтиленгликоль)-2,3-димиристоилглицерин (PEG-DMG)), PEG-диалкилоксипропила (DAA), PEG-фосфолипида, PEG-церамида (Cer), пегилированного

фосфатидилэтаноламина (PEG-PE), PEG-сукцинатдиацилглицерина (PEG-DAG) (такого как 4-0-(2',3'-ди(тетрадеканоилокси)пропил-1-0-(w-метокси(полиэтокси)этил)бутандиоат (PEG-S-DMG)), PEG-диалкоксипропилкарбама, натриевой соли N-(карбонилметоксиполиэтиленгликоль-2000)-1,2-дистеароил-sn-глицеро-3-фосфоэтаноламина и соединений, описанных в таблице 2 документа WO 2019051289 (включенного посредством ссылки), а также комбинаций вышеперечисленного.

В некоторых вариантах осуществления стерины, которые могут быть включены в липидные наночастицы, включают одно или несколько из холестерина или производных холестерина, например, таких как в WO2009/127060 или US2010/0130588, которые включены посредством ссылки. Дополнительные иллюстративные стерины включают фитостерины, в том числе описанные в Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386, включенном в данный документ посредством ссылки.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид, некатионный липид, конъюгированный липид, который ингибирует агрегацию частиц, и стерин. Количества данных компонентов могут варьироваться независимо и для достижения желаемых свойств. Например, в некоторых вариантах осуществления липидная наночастица содержит ионизируемый липид в количестве от приблизительно 20 мол. % до приблизительно 90 мол. % от общего количества липидов (в других вариантах осуществления это может составлять 20-70% (мол.), 30-60% (мол.) или 40-50% (мол.); от приблизительно 50 мол. % до приблизительно 90 мол. % от общего количества липидов, присутствующих в липидной наночастице), некатионный липид в количестве от приблизительно 5 мол. % до приблизительно 30 мол. % от общего количества липидов, конъюгированный липид в количестве от приблизительно 0,5 мол. % до приблизительно 20 мол. % от общего количества липидов и стерин в количестве от приблизительно 20 мол. % до приблизительно 50 мол. % от общего количества липидов. Соотношение общего количества липидов и нуклеиновой кислоты (например, кодирующей Gene Writer или нуклеиновой кислоты-матрицы) можно варьировать по желанию. Например, соотношение общего количества липидов и нуклеиновой кислоты (по массе или весу) может составлять от приблизительно 10: 1 до приблизительно 30: 1.

В некоторых вариантах осуществления ионизируемый липид может быть катионным липидом, ионизируемым катионным липидом, например, катионным липидом, который может существовать в положительно заряженной или нейтральной форме в зависимости от pH, или аминосодержащим 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 аминогрупп, которые несут положительный заряд. В некоторых вариантах осуществления липидная частица содержит катионный липид в составе с одним или несколькими нейтральными липидами, ионизируемыми аминосодержащими липидами, биоразлагаемыми алкиновыми

липидами, стероидами, фосфолипидами, содержащими полиненасыщенные липиды, структурными липидами (например, стеринами), PEG, холестерином и липидами, конъюгированными с полимером. В некоторых вариантах осуществления катионный липид может представлять собой ионизируемый катионный липид. Иллюстративный катионный липид, раскрытый в данном документе, может характеризоваться эффективным значением pKa, составляющим более 6,0. В вариантах осуществления липидная наночастица может содержать второй катионный липид, характеризующийся другим эффективным значением pKa (например, более высоким, чем первое эффективное значение pKa) по сравнению с первым катионным липидом. Липидная наночастица может содержать от 40 до 60 молярных процентов катионного липида, нейтрального липида, стероида, липида, конъюгированного с полимером, и терапевтического средства, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), инкапсулированной внутри липидной наночастицы или ассоциированной с ней. В некоторых вариантах осуществления нуклеиновая кислота составлена совместно с катионным липидом. Нуклеиновая кислота может быть адсорбирована на поверхности LNP, например, LNP, содержащей катионный липид. В некоторых вариантах осуществления нуклеиновая кислота может быть инкапсулирована в LNP, например, в LNP, содержащую катионный липид. В некоторых вариантах осуществления липидная наночастица может содержать нацеливающий фрагмент, например, покрытый нацеливающим средством. В вариантах осуществления состав на основе LNP является биоразлагаемым. В некоторых вариантах осуществления липидная наночастица, содержащая один или несколько липидов, описанных в данном документе, например, формулы (i), (ii), (ii), (vii) и/или (ix), инкапсулирует по меньшей мере 1%, по меньшей мере 5%, по меньшей мере 10%, по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 92%, по меньшей мере 95%, по меньшей мере 97%, по меньшей мере 98% или 100% молекулы РНК, например, РНК-матрицы и/или мРНК, кодирующей полипептид Gene Writer.

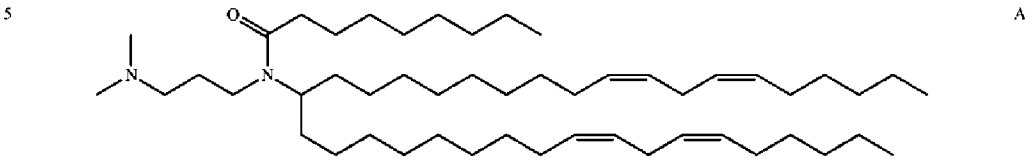
В некоторых вариантах осуществления соотношение липидов и нуклеиновой кислоты (соотношение масса/масса; соотношение вес/вес) может находиться в диапазоне от приблизительно 1: 1 до приблизительно 25: 1, от приблизительно 10: 1 до приблизительно 14: 1, от приблизительно 3: 1 до приблизительно 15: 1, от приблизительно 4: 1 до приблизительно 10: 1, от приблизительно 5: 1 до приблизительно 9: 1 или от приблизительно 6: 1 до приблизительно 9: 1. Количества липидов и нуклеиновой кислоты можно регулировать для обеспечения желаемого соотношения N/P, например, соотношения N/P, составляющего 3, 4, 5, 6, 7, 8, 9, 10 или выше. В целом, общее содержание липидов в составе на основе липидных наночастиц может находиться в диапазоне от приблизительно 5 мг/мл до приблизительно 30 мг/мл.

Иллюстративные ионизируемые липиды, которые можно применять в составах на основе липидных наночастиц, включают без ограничения липиды, перечисленные в таблице 1 документа WO2019051289, включенного в данный документ посредством ссылки. Дополнительные иллюстративные липиды включают без ограничения одну или несколько из следующих формул: X из US2016/0311759; I из US20150376115 или из US2016/0376224; I, II или III из US20160151284; I, IA, II или IIA из US20170210967; I-c из US20150140070; A из US2013/0178541; I из US2013/0303587 или US2013/0123338; I из US2015/0141678; II, III, IV или V из US2015/0239926; I из US2017/0119904; I или II из WO2017/117528; A из US2012/0149894; A из US2015/0057373; A из WO2013/116126; A из US2013/0090372; A из US2013/0274523; A из US2013/0274504; A из US2013/0053572; A из W02013/016058; A из W02012/162210; I из US2008/042973; I, II, III или IV из US2012/01287670; I или II из US2014/0200257; I, II или III из US2015/0203446; I или III из US2015/0005363; I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID или III-XXIV из US2014/0308304; из US2013/0338210; I, II, III или IV из W02009/132131; A из US2012/01011478; I или XXXV из US2012/0027796; XIV или XVII из US2012/0058144; из US2013/0323269; I из US2011/0117125; I, II или III из US2011/0256175; I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII из US2012/0202871; I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV или XVI из US2011/0076335; I или II из US2006/008378; I из US2013/0123338; I или X-A-Y-Z из US2015/0064242; XVI, XVII или XVIII из US2013/0022649; I, II или III из US2013/0116307; I, II или III из US2013/0116307; I или II из US2010/0062967; I-X из US2013/0189351; I из US2014/0039032; V из US2018/0028664; I из US2016/0317458; I из US2013/0195920; 5, 6 или 10 из US10,221,127; III-3 из WO2018/081480; I-5 или I-8 из WO2020/081938; 18 или 25 из US9867888; A из US2019/0136231; II из WO2020/219876; 1 из US2012/0027803; OF-02 из US2019/0240349; 23 из US10,086,013; cKK-E12/A6 из Miao et al (2020); C12-200 из WO2010/053572; 7C1 из Dahlman et al (2017); 304-O13 или 503-O13 из Whitehead et al; TS-P4C2 из US9708628; I из WO2020/106946; I из WO2020/106946.

В некоторых вариантах осуществления ионизируемый липид представляет собой MC3 (6Z,9Z,28Z,31Z)-гептатриаконта-6,9,28,31-тетраен-19-ил-4-(диметиламино)бутаноат (DLin-MC3-DMA или MC3), например, описанный в примере 9 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой липид ATX-002, например, описанный в примере 10 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой (13Z,16Z)-A, A-диметил-3-нонилдокоза-13-16-диен-1-амин (соединение 32), например, описанное в примере 11 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой соединение 6 или соединение 22, например, описанные в примере 12 документа WO2019051289A9 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый

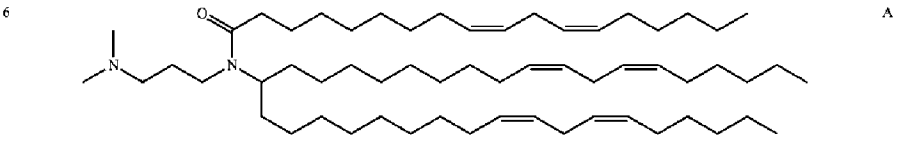
липид представляет собой гептадекан-9-ил-8-((2-гидроксиэтил)(6-оксо-6-(ундецилокси)гексил)амино)октаноат (SM-102); например, описанный в примере 1 документа US9867888 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой (9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат (LP01), например, синтезированный в примере 13 документа WO 2015/095340 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой ди((Z)-нон-2-ен-1-ил)-9-(((4-диметиламино)бутаноил)окси)гептадекандиоат (L319), например, синтезированный в примерах 7, 8 или 9 документа US2012/0027803 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой 1,1'-((2-(4-(2-((2-(бис(2-гидроксидодецил)амино)этил)(2-гидроксидодецил)амино)этил)пиперазин-1-ил)этил)азандиил)бис(додекан-2-ол) (C12-200), например, синтезированный в примерах 14 и 16 документа WO2010/053572 (включенного в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ионизируемый липид представляет собой липид на основе сложного эфира холестерина, содержащего имидазол (ICE) - (3S,10R,13R,17R)-10,13-диметил-17-((R)-6-метилгептан-2-ил)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-тетрадекагидро-1H-циклопента[a]фенантрен-3-ил-3-(1H-имидазол-4-ил)пропаноат, например, структуру (I) из документа WO2020/106946 (включенного в данный документ посредством ссылки во всей своей полноте).

Некоторые неограничивающие примеры липидных соединений, которые можно применять (например, в комбинации с другими липидными компонентами) для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), включают



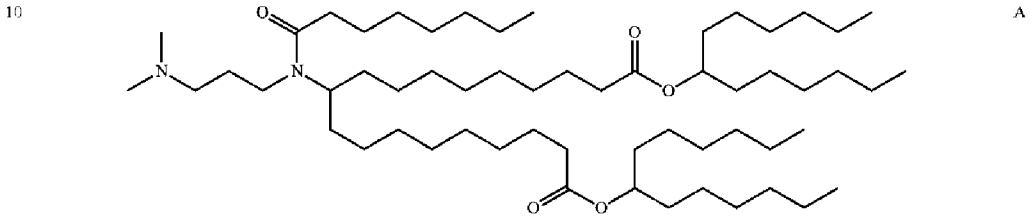
(i)

В некоторых вариантах осуществления LNP, содержащую формулу (i), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



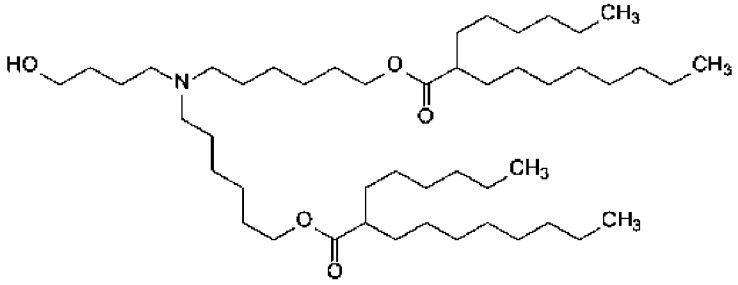
(ii)

В некоторых вариантах осуществления LNP, содержащую формулу (ii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

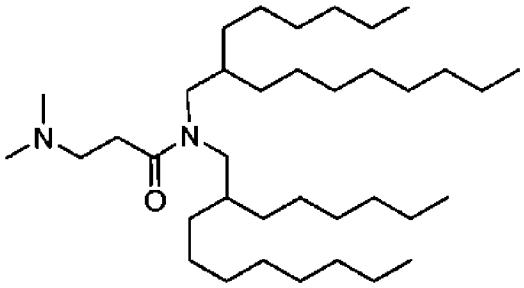


(iii)

В некоторых вариантах осуществления LNP, содержащую формулу (iii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

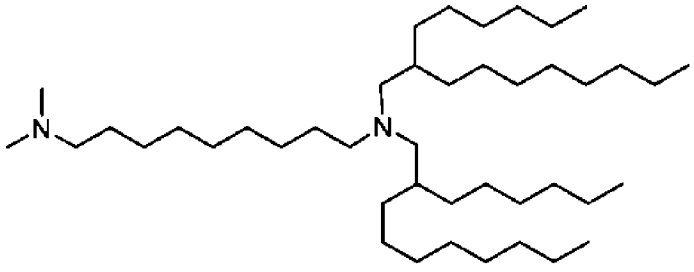


(iv)



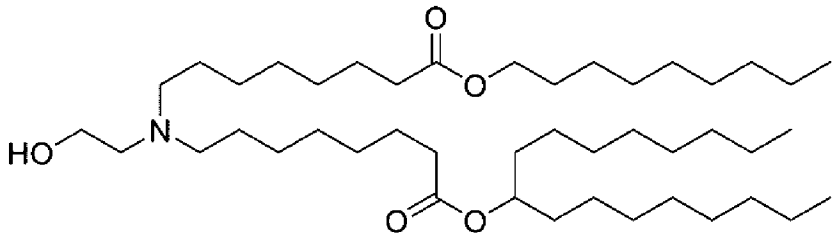
(v)

В некоторых вариантах осуществления LNP, содержащую формулу (v), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

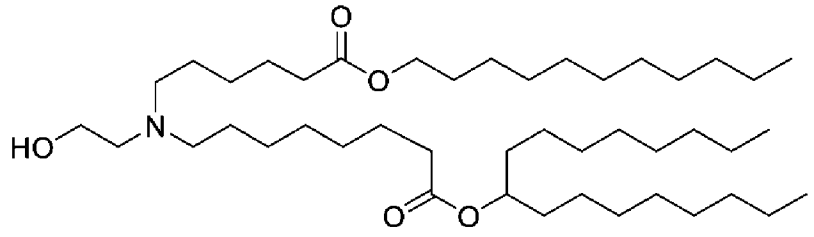


(vi)

В некоторых вариантах осуществления LNP, содержащую формулу (vi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

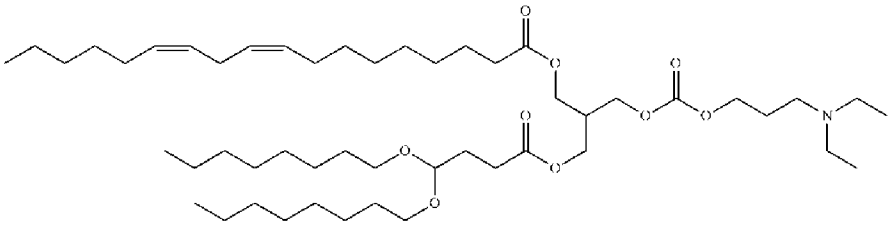


(vii)



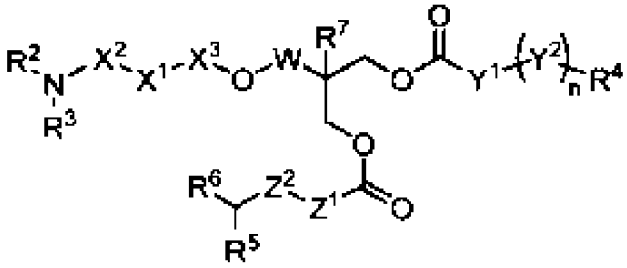
(viii)

В некоторых вариантах осуществления LNP, содержащую формулу (viii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



(ix)

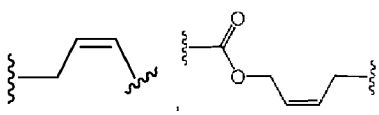
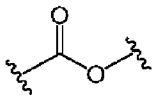
В некоторых вариантах осуществления LNP, содержащую формулу (ix), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.



(x),

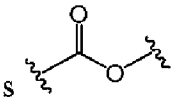
где

X^1 представляет собой O, NR^1 или прямую связь, X^2 представляет собой C2-5-алкилен, X^3 представляет собой C(=0) или прямую связь, R^1 представляет собой H или Me, R^3 представляет собой C-3-алкил, R^2 представляет собой C-3-алкил, или R^2, взятый вместе с атомом азота, к которому он присоединен, и 1-3 атома углерода из X^2 образуют 4-, 5- или 6-членное кольцо, или X^1 представляет собой NR^1, R^1 и R^2, взятые вместе с атомами азота, к которому они присоединены, образуют 5- или 6-членное кольцо, или R^2, взятый вместе с R^3 и атомом азота, к которому они присоединены, образует 5-, 6- или 7-членное кольцо, Y^1 представляет собой C2-12-алкилен, Y^2 выбран из

 orientation), and  (in either orientation),

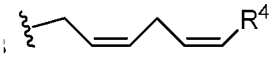
(в любой ориентации), (в любой ориентации), (в любой ориентации),

n равняется 0-3, R^4 представляет собой C1-15-алкил, Z^1 представляет собой C1-6-алкилен или прямую связь,

Z^2 is 

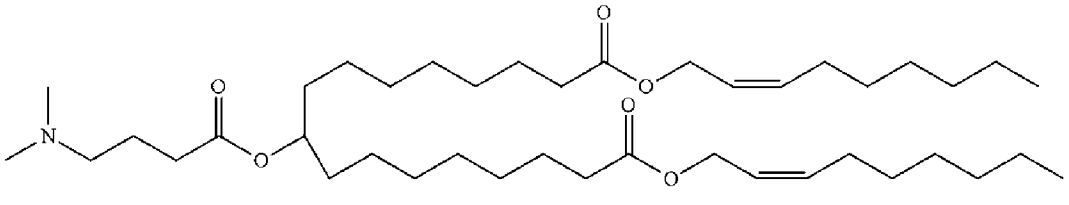
(в любой ориентации) или отсутствует, при условии, что если Z^1 представляет собой прямую связь, то Z^2 отсутствует;

R^5 представляет собой C5-9-алкил или C6-10-алкокси, R^6 представляет собой C5-9-алкил или C6-10-алкокси, W представляет собой метилен или прямую связь, и R^7 представляет собой H или Me или их соль, при условии, что если R^3 и R^2 представляют собой C2-алкилы, X^1 представляет собой O, X^2 представляет собой линейный C3-алкилен, X^3 представляет собой C(=0), Y^1 представляет собой линейный C1-алкилен, $(Y^2)n\text{-}R^4$ представляет собой



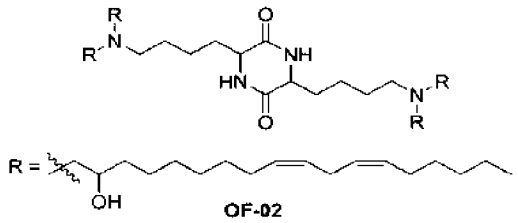
, R^4 представляет собой линейный C5-алкил, Z^1 представляет собой C2-алкилен, Z^2 отсутствует, W представляет собой метилен, и R^7 представляет собой H, то R^5 и R^6 не представляют собой Cx-алкокси.

В некоторых вариантах осуществления LNP, содержащую формулу (xii), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

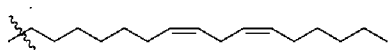


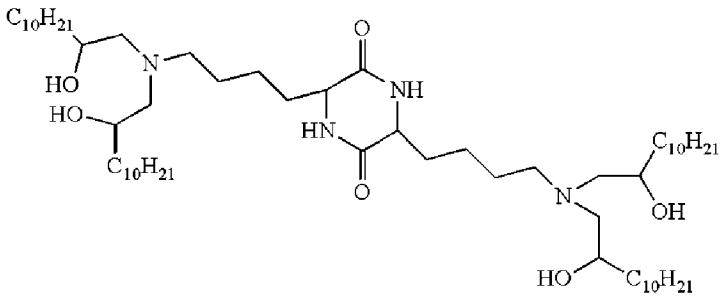
(xi)

В некоторых вариантах осуществления LNP, содержащую формулу (xi), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

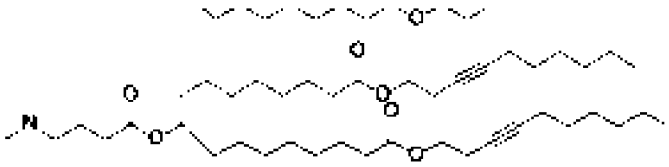


OF-02

где R=  (xii)

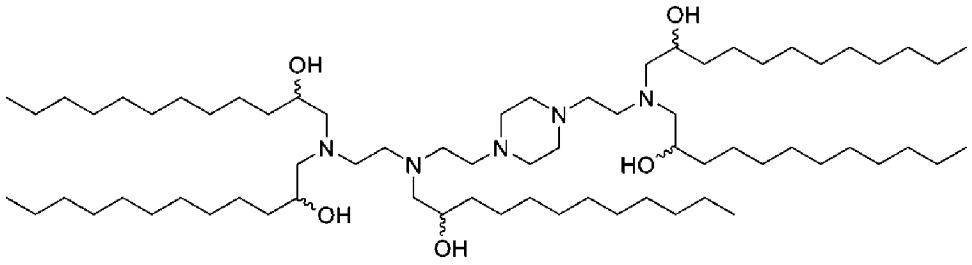


(xiii)



(xiv)

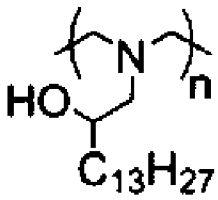
В некоторых вариантах осуществления LNP содержит соединение формулы (xiii) и соединение формулы (xiv).



(xv)

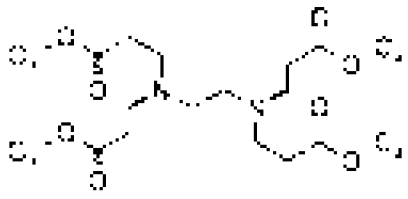
В некоторых вариантах осуществления LNP, содержащую формулу (xv), применяют для доставки композиции Gene Writer, описанной в данном документе, в клетки печени и/или гепатоциты.

Сердцевинный PEI_{600}

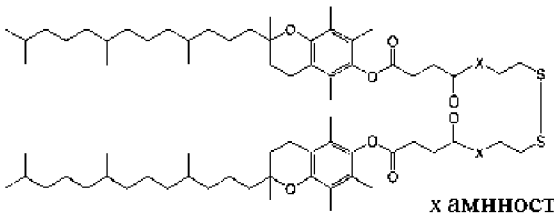


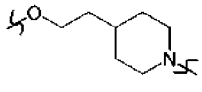
(xvi)

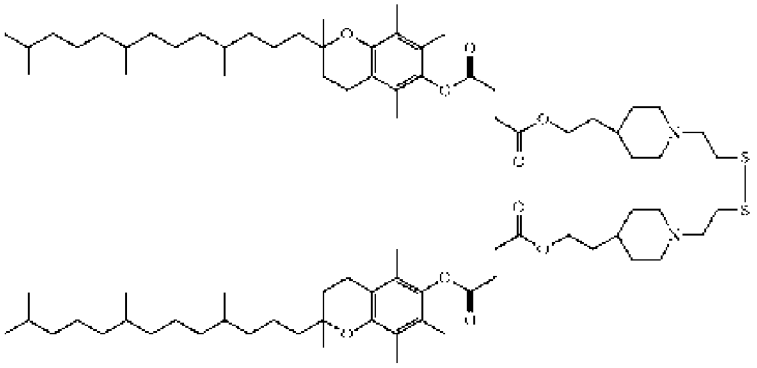
В некоторых вариантах осуществления LNP, содержащую состав формулы (xvi), применяют для доставки композиции Gene Writer, описанной в данном документе, в эндотелиальные клетки легкого.



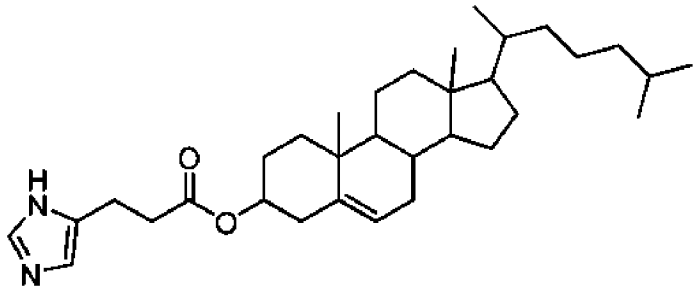
(xvii)



x аминоструктура где X=  (xviii) (a)

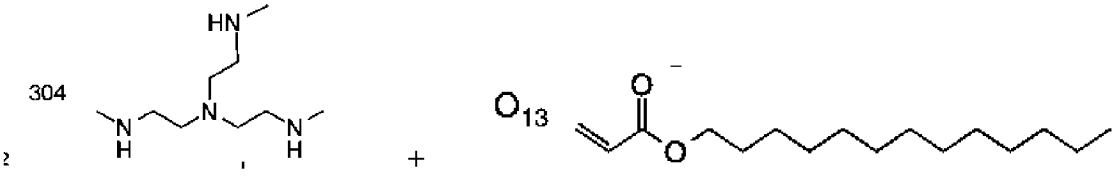


(xviii) (b)

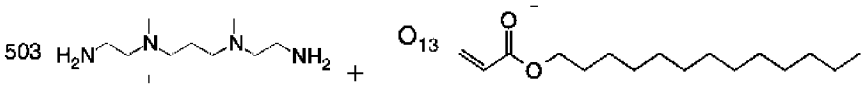


(xix)

В некоторых вариантах осуществления липидное соединение, применяемое для образования липидных наночастиц для доставки композиций, описанных в данном документе, например, нуклеиновой кислоты (например, РНК), описанной в данном документе (например, нуклеиновой кислоты-матрицы или нуклеиновой кислоты, кодирующей Gene Writer), получают с помощью одной из следующих реакций:



(xx) (a)

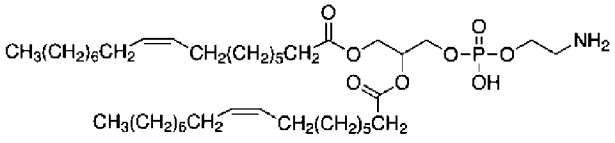


(xx)(b)

Иллюстративные некатионные липиды включают без ограничения дистеароил-sn-глицерофосфоэтаноламин, дистеароилфосфатидилхолин (DSPC), диолеоилфосфатидилхолин (DOPC), дипальмитоилфосфатидилхолин (DPPC), диолеоилфосфатидилглицерин (DOPG), дипальмитоилфосфатидилглицерин (DPPG), диолеоилфосфатидилэтаноламин (DPPG), 1,2-диолеоил-sn-глицеро-3-фосфоэтаноламин (DOPE), пальмитоилолеоилфосфатидилхолин (POPC), пальмитоилолеоилфосфатидилэтаноламин (POPE), диолеоилфосфатидилэтаноламин-4-(N-

малеимидометил)-циклогексан-1-карбоксилат (DOPE-Mal), дипальмитоилфосфатидилэтаноламин (DPPE), димиристоилфосфоэтаноламин (DMPE), дистеароилфосфатидилэтаноламин (DSPE), монометилфосфатидилэтаноламин (такой как 16-O-монометил-PE), диметилфосфатидилэтаноламин (такой как 16-O-диметил-PE), 18-1-транс-PE, 1-стеароил-2-олеоилфосфатидилэтаноламин (SOPE), гидрогенизированный соевый фосфатидилхолин (HSPC), яичный фосфатидилхолин (EPC), диолеоилфосфатидилсерин (DOPS), сфингомиелин (SM), димиристоилфосфатидилхолин (DMPC), димиристоилфосфатидилглицерин (DMPG), дистеароилфосфатидилглицерин (DSPG), диэрукоилфосфатидилхолин (DEPC), пальмитоилолеоилфосфатидилглицерин (POPG), диэлаидоилфосфатидилэтаноламин (DEPE), лецитин, фосфатидилэтаноламин, лизолецитин, лизофосфатидилэтаноламин, фосфатидилсерин, фосфатидилинозитол, сфингомиелин, яичный сфингомиелин (ESM), цефалин, кардиолипин, фосфатидную кислоту, цереброзиды, дицетилфосфат, лизофосфатидилхолин, дилинолеоилфосфатидилхолин или их смеси. Понятно, что можно также применять другие диацилфосфатидилхолиновые и диацилфосфатидилэтаноламиновые фосфолипиды. Ацильные группы в данных липидах предпочтительно представляют собой ацильные группы, полученные из жирных кислот, имеющих углеродные цепи C10-C24, например, лауроил, миристоил, пальмитоил, стеароил или олеоил. Дополнительные иллюстративные липиды в некоторых вариантах осуществления включают без ограничения липиды, описанные в Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386, включенном в данный документ посредством ссылки. Такие липиды включают в некоторых вариантах осуществления растительные липиды, которые, как было обнаружено, улучшают трансфекцию мРНК в печени (например, DGTS).

В некоторых вариантах осуществления некатионный липид может иметь следующую структуру:



(xxi)

Другие примеры некатионных липидов, пригодных для применения в липидных наночастицах, включают без ограничения липиды, не содержащие фосфор, такие как, например, стеариламин, додециламин, гексадециламин, ацетилпальмитат, глицеринрицинолеат, гексадецилстеарат, изопропилмиристат, амфотерные акриловые полимеры, лаурилсульфат триэтаноламина, алкиларилсульфат, полиэтилоксилированные амиды жирных кислот, бромид диоктадецилдиметиламмония, церамид, сфингомиелин и т. п. Другие некатионные липиды описаны в WO2017/099823 или в публикации заявки на патент США US2018/0028664, содержание которых включено в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления некатионный липид представляет собой олеиновую кислоту или соединение формулы I, II или IV из US2018/0028664, включенного

в данный документ посредством ссылки во всей своей полноте. Некатионный липид может составлять, например, 0-30% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления содержание некатионных липидов составляет 5-20% (мол.) или 10-15% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В вариантах осуществления молярное соотношение ионизируемых липидов и нейтральных липидов находится в диапазоне от приблизительно 2:1 до приблизительно 8:1 (например, составляет приблизительно 2:1, 3:1, 4:1, 5:1, 6:1, 7:1 или 8:1).

В некоторых вариантах осуществления липидные наночастицы не содержат каких-либо фосфолипидов.

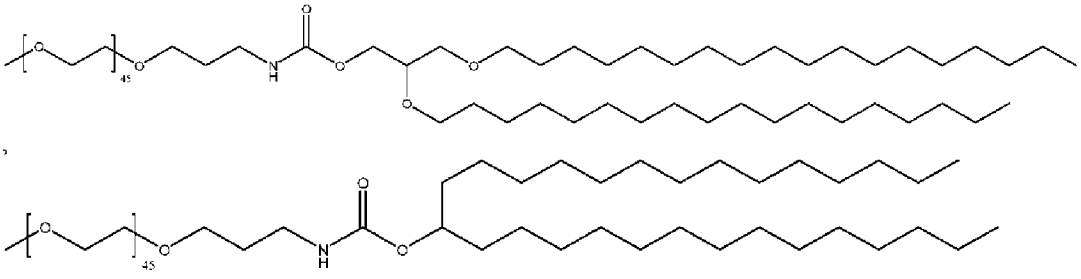
В некоторых аспектах липидная наночастица может дополнительно содержать компонент, такой как стерин, для обеспечения целостности мембраны. Одним иллюстративным стерином, который можно применять в липидной наночастице, является холестерин и его производные. Неограничивающие примеры производных холестерина включают полярные аналоги, такие как 5α-холестанол, 53-копростанол, холестерил-(2'-гидрокси)-этиловый эфир, холестерил-(4'-гидрокси)-бутиловый эфир и 6-кетохолестанол; неполярные аналоги, такие как 5а-холестан, холестенон, 5а-холестанон, 5р-холестанон и холестерилдеканоат; и их смеси. В некоторых вариантах осуществления производное холестерина представляет собой полярный аналог, например, холестерил-(4'-гидрокси)бутиловый эфир. Иллюстративные производные холестерина описаны в публикации согласно PCT WO2009/127060 и публикации заявки на патент США US2010/0130588, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления компонент, обеспечивающий целостность мембраны, такой как стерин, может составлять 0-50% (мол.) (например, 0-10%, 10-20%, 20-30%, 30-40% или 40-50%) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления такой компонент составляет 20-50% (мол.) или 30-40% (мол.) от общего количества липидов в липидной наночастице.

В некоторых вариантах осуществления липидная наночастица может содержать полиэтиленгликоль (PEG) или конъюгированную молекулу липида. Обычно они применяются для ингибирования агрегации липидных наночастиц и/или обеспечения стерической стабилизации. Иллюстративные конъюгированные липиды включают без ограничения конъюгаты PEG-липид, конъюгаты полиоксазолин (POZ)-липид, конъюгаты полиамид-липид (такие как конъюгаты ATTA-липид), конъюгаты катионный полимер-липида (CPL) и их смеси. В некоторых вариантах осуществления молекула конъюгированного липида представляет собой конъюгат PEG-липид, например, липид, конъюгированный с (метоксиполиэтиленгликолем).

Иллюстративные конъюгаты PEG-липид включают без ограничения PEG-диацилглицерин (DAG) (такой как 1-(монометоксиполиэтиленгликоль)-2,3-димиристоилглицерин (PEG-DMG)), PEG-диалкилоксипропил (DAA), PEG-фосфолипид,

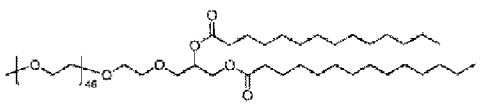
PEG-церамид (Cer), пегилированный фосфатидилэтаноламин (PEG-PE), PEG-сукцинатдиацилглицерин (PEG-DAG) (например, 4-0-(2',3'-ди(тетрадеканоилокси)пропил-1-0-(w-метокси(полиэтокси)этил)бутандиоат (PEG-S-DMG)), PEG-диалкоксипропилкарбам, натриевую соль N-(карбонилметоксиполиэтиленгликоль 2000)-1,2-дистеароил-sn-глицеро-3-фосфоэтаноламина, 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:



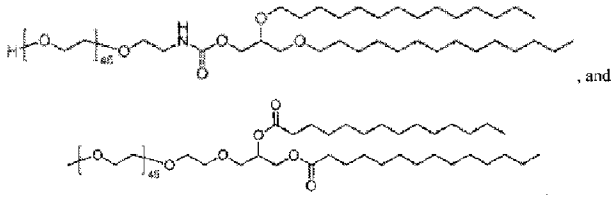
(xxii),



, and

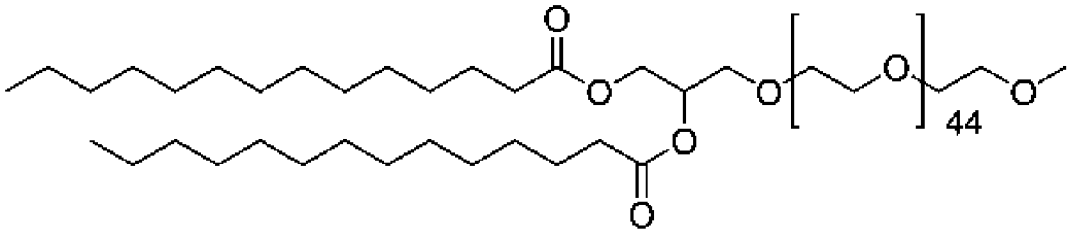


(xxiii),



, and

(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 ATTA-липид) и конъюгаты катионный полимер-липид (GPL).

Иллюстративные конъюгированные липиды, т. е. конъюгаты PEG-липид, конъюгаты (POZ)-липид, конъюгаты ATTA-липид и катионный полимер-липид, описаны в патентных заявках согласно PCT и LIS, указанных в таблице 2 документа WO2019051289A9 и в документе WO2020106946A1, содержание каждого из которых включено в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления LNP включает соединение формулы (xix), соединение формулы (xxi) и соединение формулы (xxv). В некоторых вариантах осуществления LNP, содержащую состав формулы (xix), формулы (xxi) и формулы (xxv), применяют для доставки композиции Gene Writer, описанной в данном документе, в легкое или клетки легкого.

В некоторых вариантах осуществления PEG или конъюгированный липид могут составлять 0-20% (мол.) от общего количества липидов, присутствующих в липидной наночастице. В некоторых вариантах осуществления содержание PEG или конъюгированного липида составляет 0,5-10% или 2-5% (мол.) от общего количества липидов, присутствующих в липидной наночастице. Молярные соотношения ионизируемого липида, некатионного липида, стерина и PEG-конъюгированного липида можно изменять по мере необходимости. Например, липидная частица может содержать 30-70% ионизируемого липида по молям или от общего веса композиции, 0-60% холестерина по молям или от общего веса композиции, 0-30% некатионного липида по молям или от общего веса композиции и 1-10% конъюгированного липида по молям или от общего веса композиции. Предпочтительно композиция содержит 30-40% ионизируемого липида по молям или от общего веса композиции, 40-50% холестерина по молям или от общего веса композиции и 10-20% некатионного липида по молям или от общего веса композиции. В некоторых других вариантах осуществления композиция содержит 50-75% ионизируемого липида по молям или от общего веса композиции, 20-40% холестерина по

молям или от общего веса композиции, 5-10% некатионного липида по молям или от общего веса композиции и 1-10% конъюгированного липида по молям или от общего веса композиции. Композиция может содержать 60-70% ионизируемого липида по молям или от общего веса композиции, 25-35% холестерина по молям или от общего веса композиции и 5-10% некатионного липида по молям или от общего веса композиции. Композиция может также содержать до 90% ионизируемого липида по молям или от общего веса композиции и 2-15% некатионного липида по молям или от общего веса композиции. Состав также может представлять собой состав на основе липидных наночастиц, например, содержащий 8-30% ионизируемого липида по молям или от общего веса композиции, 5-30% некатионного липида по молям или от общего веса композиции и 0-20% холестерина по молям или от общего веса композиции; 4-25% ионизируемого липида по молям или от общего веса композиции, 4-25% некатионного липида по молям или от общего веса композиции, 2-25% холестерина по молям или от общего веса композиции, 10-35% конъюгированного липида по молям или от общего веса композиции и 5% холестерина по молям или от общего веса композиции; или 2-30% ионизируемого липида по молям или от общего веса композиции, 2-30% некатионного липида по молям или от общего веса композиции, 1-15% холестерина по молям или от общего веса композиции, 2-35% конъюгированного липида по молям или от общего веса композиции и 1-20% холестерина по молям или от общего веса композиции; или даже до 90% ионизируемого липида по молям или от общего веса композиции и 2-10% некатионных липидов по молям или от общего веса композиции, или даже 100% катионного липида по молям или от общего веса композиции. В некоторых вариантах осуществления состав на основе липидных частиц содержит ионизируемый липид, фосфолипид, холестерин и пегилированный липид в молярном соотношении 50: 10:38,5: 1,5. В некоторых других вариантах осуществления состав на основе липидных частиц содержит ионизируемый липид, холестерин и пегилированный липид в молярном соотношении 60:38,5: 1,5.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид, некатионный липид (например, фосфолипид), стерин (например, холестерин) и пегилированный липид, где молярное соотношение липидов находится в диапазоне от 20 до 70 молярных процентов для ионизируемого липида при целевом значении 40-60, молярный процент для некатионного липида находится в диапазоне от 0 до 30 при целевом значении от 0 до 15, молярный процент для стерина находится в диапазоне от 20 до 70 при целевом значении от 30 до 50, и молярный процент для пегилированного липида находится в диапазоне от 1 до 6 при целевом значении от 2 до 5.

В некоторых вариантах осуществления липидная частица содержит ионизируемый липид/некатионный липид/стерин/конъюгированный липид в молярном соотношении 50: 10:38,5: 1,5.

В одном аспекте в настоящем изобретении предусмотрен состав на основе липидных наночастиц, содержащий фосфолипиды, лецитин, фосфатидилхолин и фосфатидилэтаноламин.

В некоторых вариантах осуществления также могут быть включены одно или несколько дополнительных соединений. Данные соединения можно вводить отдельно, или дополнительные соединения могут быть включены в липидные наночастицы по настоящему изобретению. Другими словами, липидные наночастицы могут содержать другие соединения в дополнение к нуклеиновой кислоте или по меньшей мере вторую нуклеиновую кислоту, отличающуюся от первой. Другие дополнительные соединения могут без ограничений быть выбраны из группы, состоящей из малых или больших органических или неорганических молекул, моносахаридов, дисахаридов, трисахаридов, олигосахаридов, полисахаридов, пептидов, белков, аналогов пептидов и их производных, пептидомиметиков, нуклеиновых кислот, аналогов нуклеиновых кислот и производных, экстрактов, полученных из биологических материалов, или любых их комбинаций.

В некоторых вариантах осуществления липидная наночастица (или состав, содержащий липидные наночастицы) не содержит реакционноспособных примесей (например, альдегидов или кетонов) или содержит реакционноспособные примеси (например, альдегиды или кетоны) в количестве, меньшем, чем предварительно выбранный уровень. Без ограничения какой-либо теорией, в некоторых вариантах осуществления липидный реагент применяется для получения состава на основе липидных наночастиц, и липидный реагент может содержать загрязняющую реакционноспособную примесь (например, альдегид или кетон). Липидный реагент может быть выбран для изготовления на основании наличия реакционноспособных примесей (например, альдегидов или кетонов) в количестве, меньшем, чем предварительно выбранный уровень. Без ограничения какой-либо теорией, в некоторых вариантах осуществления альдегиды могут вызывать модификацию и повреждение РНК, например, поперечную сшивку оснований и/или ковалентное конъюгирование липида с РНК (например, образование аддуктов липид-РНК). В некоторых случаях это может привести к неудачной реакции с участием обратной транскриптазы и/или включению неподходящих оснований, например, в месте(местах) повреждения(повреждений), например, мутации во вновь синтезированной ДНК-мишени.

В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента, характеризующегося общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента, содержащего менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием липидного реагента: (i) характеризующегося общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержащего менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности

реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц получают с использованием множества липидных реагентов, и каждый липидный реагент из множества независимо соответствует одному или нескольким критериям, описанным в данном абзаце. В некоторых вариантах осуществления каждый липидный реагент из множества соответствует одному и тому же критерию, например, критерию из данного абзаца.

В некоторых вариантах осуществления состав на основе липидных наночастиц характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления состав на основе липидных наночастиц содержит менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления состав на основе липидных наночастиц: (i) характеризуется общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержит менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

В некоторых вариантах осуществления один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%. В некоторых вариантах осуществления один или несколько или необязательно все из липидных реагентов, используемых для получения липидной наночастицы, описанной в данном документе, или состава на ее основе, содержат менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов). В некоторых вариантах осуществления один или несколько или необязательно все липидные реагенты, используемые для получения липидной наночастицы, описанной в данном документе, или состава на ее основе: (i) характеризуются общим содержанием реакционноспособных примесей (например, альдегидов), составляющим менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1%; и (ii) содержат менее 5%, 4%, 3%, 2%, 1%, 0,9%, 0,8%, 0,7%, 0,6%, 0,5%, 0,4%, 0,3%, 0,2% или 0,1% какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов).

В некоторых вариантах осуществления общее содержание альдегидов и/или количество какой-либо отдельной разновидности реакционноспособных примесей (например, альдегидов) определяется с помощью жидкостной хроматографии (LC), например, в сочетании с тандемной масс-спектрометрией (MS/MS), например, в соответствии со способом, описанным в примере 34. В некоторых вариантах осуществления

содержание реакционноспособных примесей (например, альдегидов) и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяют посредством выявления одной или нескольких химических модификаций молекулы нуклеиновой кислоты (например, молекулы РНК, например, описанной в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах. В некоторых вариантах осуществления содержание реакционноспособных примесей (например, альдегидов) и/или количество разновидностей реакционноспособных примесей (например, альдегидов) определяют посредством выявления одной или нескольких химических модификаций нуклеотида или нуклеозида (например, рибонуклеотида или рибонуклеозида, например, содержащихся в молекуле нуклеиновой кислоты-матрицы или выделенных из нее, например, как описано в данном документе), ассоциированных с наличием реакционноспособных примесей (например, альдегидов), например, в липидных реагентах, например, как описано в примере 35. В вариантах осуществления химические модификации молекулы нуклеиновой кислоты, нуклеотида или нуклеозида выявляют посредством определения присутствия одного или нескольких модифицированных нуклеотидов или нуклеозидов, например, с применением анализа методом LC-MS/MS, например, как описано в примере 35.

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, зависимых от ASGPR, для наблюдаемого эффекта доставки груза с помощью LNP (см., например, фигуру 6). Другие составы на основе LNP, на которых представлен лиганд, например, в которые включены фолат, трансферрин или антитела, обсуждаются в WO2017223135, который включен в данный документ посредством ссылки во всей своей полноте, в дополнение к литературным источникам, используемым в данном документе, а именно Kolhatkar et al., Curr Drug Discov Technol. 2011 8:197-206; Musacchio and Torchilin, Front Biosci. 2011 16:1388-1412; Yu et al., Mol Membr Biol. 2010 27:286-298; Patil et al., Crit Rev Ther Drug Carrier Syst. 2008 25:1-61 ; Benoit et al., Biomacromolecules. 2011 12:2708-2714; Zhao et al., Expert Opin Drug Deliv. 2008 5:309-319; Akinc et al., Mol Ther. 2010 18:1357-1364; Srinivasan et al., Methods Mol Biol. 2012 820:105-116; Ben-Arie et al., Methods Mol Biol. 2012 757:497-507; Peer 2010 J Control Release. 20:63-68; Peer et al., Proc Natl Acad Sci U S A. 2007 104:4095-4100; Kim et al., Methods Mol Biol. 2011 721:339-353; Subramanya et al., Mol Ther. 2010 18:2028-2037; Song et al., Nat Biotechnol. 2005 23:709-717; Peer et al., Science. 2008 319:627-630; и Peer and Lieberman, Gene Ther. 2011 18:1127-1133.

В некоторых вариантах осуществления LNP выбирают по тканеспецифической активности посредством добавления молекулы селективного нацеливания на орган (SORT) в состав, содержащий традиционные компоненты, такие как ионизируемые катионные липиды, амфипатические фосфолипиды, холестерин и конъюгированные с поли(этиленгликолем) (PEG) липиды. Идеи Cheng et al. Nat Nanotechnol 15(4):313-320 (2020) демонстрируют, что добавление дополнительного компонента "SORT" точно изменяет профиль доставки РНК in vivo и опосредует тканеспецифическую (например, в легкие, печень, селезенку) доставку и редактирование генов в зависимости от процентного содержания и биофизических свойств молекулы SORT.

В некоторых вариантах осуществления LNP содержат биоразлагаемые ионизируемые липиды. В некоторых вариантах осуществления LNP содержит (9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат, также называемый 3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропил(9Z,12Z)-октадека-9,12-диеноат, или другой ионизируемый липид. См., например, липиды из WO2019/067992, WO/2017/173054, WO2015/095340 и WO2014/136086, а также приведенные в них литературные источники. В некоторых вариантах осуществления термины "катионный" и "ионизируемый" применительно к липидам LNP являются взаимозаменяемыми, например, где ионизируемые липиды являются катионными в зависимости от значения pH.

В некоторых вариантах осуществления несколько компонентов системы Gene Writer могут быть получены в виде одного состава на основе LNP, например, состав на основе LNP содержит мРНК, кодирующую полипептид Gene Writer, и РНК-матрицу. Соотношения компонентов, представляющих собой нуклеиновые кислоты, можно изменять с целью усиления до максимума свойств терапевтического средства. В некоторых вариантах осуществления соотношение РНК-матрицы и мРНК, кодирующей полипептид Gene Writer,

составляет от приблизительно 1:1 до 100:1, например, от приблизительно 1:1 до 20:1, от приблизительно 20:1 до 40:1, от приблизительно 40:1 до 60:1, от приблизительно 60:1 до 80:1 или от приблизительно 80:1 до 100:1 в молярном соотношении. В других вариантах осуществления система из нескольких нуклеиновых кислот может быть получена при использовании отдельных составов, например, одного состава на основе LNP, содержащего РНК-матрицу, и второго состава на основе LNP, содержащего мРНК, кодирующую полипептид Gene Writer. В некоторых вариантах осуществления система может содержать более двух компонентов, представляющих собой нуклеиновые кислоты, составленных в виде LNP. В некоторых вариантах осуществления система может содержать белок, например, полипептид Gene Writer, и РНК-матрицу, составленные в виде по меньшей мере одного состава на основе LNP.

В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от нескольких 10 до нескольких 100 нм, например, при измерении с помощью динамического светорассеяния (DLS). В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от приблизительно 40 нм до приблизительно 150 нм, например, приблизительно 40 нм, 45 нм, 50 нм, 55 нм, 60 нм, 65 нм, 70 нм, 75 нм, 80 нм, 85 нм, 90 нм, 95 нм, 100 нм, 105 нм, 110 нм, 115 нм, 120 нм, 125 нм, 130 нм, 135 нм, 140 нм, 145 нм или 150 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от приблизительно 50 нм до приблизительно 100 нм, от приблизительно 50 нм до приблизительно 90 нм, от приблизительно 50 нм до приблизительно 80 нм, от приблизительно 50 нм до приблизительно 70 нм, от приблизительно 50 нм до приблизительно 60 нм, от приблизительно 60 нм до приблизительно 100 нм, от приблизительно 60 нм до приблизительно 90 нм, от приблизительно 60 нм до приблизительно 80 нм, от приблизительно 60 нм до приблизительно 70 нм, от приблизительно 70 нм до приблизительно 100 нм, от приблизительно 70 нм до приблизительно 90 нм, от приблизительно 70 нм до приблизительно 80 нм, от приблизительно 80 до приблизительно 100 нм, от приблизительно 80 до приблизительно 90 нм или от приблизительно 90 нм до приблизительно 100 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять от приблизительно 70 нм до приблизительно 100 нм. В конкретном варианте осуществления средний диаметр LNP в составе на основе LNP может составлять приблизительно 80 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP может составлять приблизительно 100 нм. В некоторых вариантах осуществления средний диаметр LNP в составе на основе LNP находится в диапазоне от приблизительно 1 мм до приблизительно 500 мм, от приблизительно 5 мм до приблизительно 200 мм, от приблизительно 10 мм до приблизительно 100 мм, от приблизительно 20 мм до приблизительно 80 мм, от приблизительно 25 мм до приблизительно 60 мм, от приблизительно 30 мм до приблизительно 55 мм, от приблизительно 35 мм до приблизительно 50 мм или от приблизительно 38 мм до приблизительно 42 мм.

В некоторых случаях LNP может являться относительно однородной. Индекс полидисперсности может применяться для указания однородности LNP, например, распределения липидных наночастиц по размерам. Небольшой (например, менее 0,3) индекс полидисперсности обычно указывает на узкое распределение частиц по размерам. LNP может характеризоваться индексом полидисперсности, составляющим от приблизительно 0 до приблизительно 0,25, например, 0,01, 0,02, 0,03, 0,04, 0,05, 0,06, 0,07, 0,08, 0,09, 0,10, 0,11, 0,12, 0,13, 0,14, 0,15, 0,16, 0,17, 0,18, 0,19, 0,20, 0,21, 0,22, 0,23, 0,24 или 0,25. В некоторых вариантах осуществления индекс полидисперсности LNP может составлять от приблизительно 0,10 до приблизительно 0,20.

Дзета-потенциал LNP можно использовать для обозначения электрокинетического потенциала композиции. В некоторых вариантах осуществления дзета-потенциал может описывать поверхностный заряд LNP. Липидные наночастицы с относительно низким зарядом, положительным или отрицательным, обычно являются желательными, поскольку более сильно заряженные соединения могут взаимодействовать с клетками, тканями и другими элементами в организме нежелательным образом. В некоторых вариантах осуществления дзета-потенциал LNP может составлять от приблизительно -10 мВ до приблизительно +20 мВ, от приблизительно -10 мВ до приблизительно +15 мВ, от приблизительно -10 мВ до приблизительно +10 мВ, от приблизительно -10 мВ до приблизительно +5 мВ, от приблизительно -10 мВ до приблизительно 0 мВ, от приблизительно -10 мВ до приблизительно -5 мВ, от приблизительно -5 мВ до приблизительно +20 мВ, от приблизительно -5 мВ до приблизительно +15 мВ, от приблизительно -5 мВ до приблизительно +10 мВ, от приблизительно -5 мВ до приблизительно +5 мВ, от приблизительно -5 мВ до приблизительно 0 мВ, от приблизительно 0 мВ до приблизительно +20 мВ, от приблизительно 0 мВ до приблизительно +15 мВ, от приблизительно 0 мВ до приблизительно +10 мВ, от приблизительно 0 мВ до приблизительно +5 мВ, от приблизительно +5 мВ до приблизительно +20 мВ, от приблизительно +5 мВ до приблизительно +15 мВ или от приблизительно +5 мВ до приблизительно +10 мВ.

Эффективность инкапсуляции белка и/или нуклеиновой кислоты, например, полипептида Gene Writer или мРНК, кодирующей полипептид, описывает количество белка и/или нуклеиновой кислоты, которое инкапсулировано или иным образом ассоциировано с LNP после получения, относительно предоставленного исходного количества. Желательно, чтобы эффективность инкапсуляции была высокой (например, близкой к 100%). Эффективность инкапсуляции можно измерить, например, путем сравнения количества белка или нуклеиновой кислоты в растворе, содержащем липидную наночастицу, до и после разрушения липидной наночастицы с помощью одного или нескольких органических растворителей или детергентов. Анионообменную смолу можно применять для измерения количества свободного белка или нуклеиновой кислоты (например, РНК) в растворе. Флуоресценцию можно применять для измерения количества свободного белка и/или нуклеиновой кислоты (например, РНК) в растворе. Для описанных в данном документе

липидных наночастиц эффективность инкапсуляции белка и/или нуклеиновой кислоты может составлять по меньшей мере 50%, например, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 80%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 90%. В некоторых вариантах осуществления эффективность инкапсуляции может составлять по меньшей мере 95%.

LNP необязательно может содержать одно или несколько покрытий. В некоторых вариантах осуществления LNP может быть составлена в виде капсулы, пленки или таблетки с покрытием. Капсула, пленка или таблетка, содержащие композицию, описанную в данном документе, могут иметь любые пригодные размер, прочность на разрыв, твердость или плотность.

Дополнительные иллюстративные липиды, составы, способы и определение характеристик LNP описаны в WO2020061457, который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления липофекцию клеток in vitro или ex vivo осуществляют с использованием Lipofectamine MessengerMax (Thermo Fisher) или реагента для трансфекции TransIT-mRNA (Mirus Bio). В определенных вариантах осуществления LNP составлены с использованием смеси ионизируемых липидов GenVoy-ILM (Precision NanoSystems). В некоторых вариантах осуществления LNP составлены с использованием 2,2-дилинолеил-4-диметиламиноэтил-[1,3]-диоксолана (DLin-KC2-DMA) или дилинолеилметил-4-диметиламинобутирата (DLin-MC3-DMA или MC3), состав и применение которого in vivo описаны в Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012), включенном в данный документ посредством ссылки во всей своей полноте.

Составы на основе LNP, оптимизированные для доставки систем CRISPR-Cas, например, RNP Cas9-gRNA, gRNA, мРНК Cas9, описаны в WO2019067992 и WO2019067910, оба из которых включены посредством ссылки.

Дополнительные конкретные составы на основе LNP, пригодные для доставки нуклеиновых кислот, описаны в US8158601 и US8168775, оба из которых включены посредством ссылки, и они включают составы, применяемые в патисиране, продаваемом под наименованием ONPATTRO.

Иллюстративная дозировка LNP с Gene Writer может составлять приблизительно 0,1, 0,25, 0,3, 0,5, 1, 2, 3, 4, 5, 6, 8, 10 или 100 мг/кг (РНК). Иллюстративная дозировка AAV, содержащего нуклеиновую кислоту, кодирующую один или несколько компонентов системы, может включать MOI, составляющую приблизительно 10^{11}, 10^{12}, 10^{13} и 10^{14} vg/кг.

Компонент системы редактирования генов Gene Writer™, представляющий собой РНК-матрицу

Системы Gene Writer, описанные в данном документе, могут осуществлять транскрипцию последовательности РНК-матрицы в сайты-мишени ДНК хозяина с помощью обратной транскрипции с праймированием мишенью. Путем записывания

последовательности(последовательностей) ДНК посредством обратной транскрипции последовательности РНК-матрицы непосредственно в геном хозяина система Gene Writer может осуществлять вставку последовательности объекта в геном-мишень без необходимости введения экзогенных последовательностей ДНК в клетку-хозяина (в отличие от, например, систем CRISPR), а также исключая стадию вставки экзогенной ДНК. Таким образом, система Gene Writer представляет собой платформу для использования индивидуально адаптированных последовательностей РНК-матриц, содержащих последовательности объекта, например, последовательности, предусматривающие кодирование гетерологичных генов и/или информацию о функции.

В целях описания компонента системы Gene Writer, представляющего собой РНК-матрицу, РНК-матрицу можно рассматривать как содержащую модульные части (фигуры 18-19). Матрица Gene Writer может содержать все или некоторые из проиллюстрированных частей, и модули можно комбинировать, переупорядочивать и/или исключать. Проиллюстрированные модули не предполагают ограничение потенциальных элементов, включенных в матрицу, и можно легко предусмотреть дополнительные компоненты, например, 5'- и 3'-концевые домены, для увеличения стабильности матрицы.

Таблица 4. Модули, содержащиеся в типичной РНК-матрице Gene Writer

A=5'-плечо гомологии; B=рибозим; C=5'-UTR; D=последовательность гетерологичного объекта; E=3'-UTR; F=3'-плечо гомологии.

Модуль	**Функция**
A: 5ʿ-плечо гомологии	Модуль 5ʿ-плечо гомологии является комплементарным последовательности ДНК с 5ʿ-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени
B: рибозим	Модуль рибозим является компонентом последовательности 5ʿ-UTR эндогенного ретротранспозона на 5'-конце 5'-UTR (Ruminski, D. J., et al, Journal of Biological Chemistry, 286(48), 41286-41295, 2011).
C: 5ʿ-UTR	Модуль 5ʿ-UTR представляет собой последовательность РНК или взаимодействующий компонент, с которым взаимодействует полипептид Gene Writer и который может быть вовлечен в процесс 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 3ʿ-UTR представляет собой последовательность РНК или любой другой взаимодействующий компонент, с которым взаимодействует полипептид Gene Writer для связывания с молекулой РНК-матрицы и который может быть вовлечен в процесс ретротранспозиции.
F: 3ʿ-плечо гомологии	Модуль 3ʿ-плечо гомологии является комплементарным последовательности ДНК с 3ʿ-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени

В некоторых вариантах осуществления РНК-матрица кодирует белок Gene Writer в *цис*-положении по отношению к последовательности гетерологичного объекта. Различные цис-конструкции были описаны, например, в Kuroki-Kami et al (2019) *Mobile DNA* 10:23 (включена в данный документ посредством ссылки во всей своей полноте) и могут применяться в комбинации с любым из вариантов осуществления, описанных в данном документе. Например, в некоторых вариантах осуществления РНК-матрица содержит последовательность гетерологичного объекта, последовательность, кодирующую белок Gene Writer (например, белок, содержащий (i) домен обратной транскриптазы и (ii) эндонуклеазный домен, например, как описано в данном документе), 5'-нетранслируемую область и 3'-нетранслируемую область. Компоненты могут быть включены в различном порядке. В некоторых вариантах осуществления белок Gene Writer и последовательность гетерологичного объекта кодируются в разных направлениях (смысловом и антисмысловом), например, с использованием схемы расположения, показанной на фигуре 3A в Kuroki-Kami et al, там же. В некоторых вариантах осуществления белок Gene Writer и последовательность гетерологичного объекта кодируются в одном направлении. В некоторых вариантах осуществления нуклеиновая кислота, кодирующая полипептид, и РНК-матрица или нуклеиновая кислота, кодирующая РНК-матрицу, связаны ковалентной связью, например, являются частью слитой нуклеиновой кислоты и/или являются частью одного и того же транскрипта. В некоторых вариантах осуществления слитая нуклеиновая кислота содержит РНК или ДНК.

Нуклеиновая кислота, кодирующая полипептид Gene Writer, может в некоторых случаях находиться с 5'-стороны от последовательности гетерологичного объекта. Например, в некоторых вариантах осуществления РНК-матрица содержит в направлении от 5' к 3' 5'-нетранслируемую область, полипептид Gene Writer, кодируемый в смысловом направлении, последовательность гетерологичного объекта, кодируемую в смысловом направлении, и 3'-нетранслируемую область. В некоторых вариантах осуществления РНК-матрица содержит в направлении от 5' к 3' 5'-нетранслируемую область, полипептид Gene Writer, кодируемый в смысловом направлении, последовательность гетерологичного объекта, кодируемую в антисмысловом направлении, и 3'-нетранслируемую область.

В некоторых вариантах осуществления РНК дополнительно предусматривает гомологию в отношении сайта-мишени ДНК.

Понятно, что, если РНК-матрица описывается как содержащая открытую рамку считывания или последовательность, обратно комплементарную ей, в некоторых вариантах осуществления РНК-матрица должна быть преобразована в двухнитевую ДНК (например, посредством обратной транскрипции) до того, как открытая рамка считывания может быть транскрибирована и транслирована.

В определенных вариантах осуществления индивидуально адаптированную последовательность РНК-матрицы можно идентифицировать, разработать, создать и сконструировать таким образом, чтобы она содержала последовательности, изменяющие или определяющие функционирование генома хозяина, например, путем введения гетерологичной кодирующей области в геном; осуществления или обеспечения сплайсинга экзонной структуры/альтернативного сплайсинга; обеспечения нарушения функции эндогенного гена; обеспечения активации транскрипции эндогенного гена; обеспечения эпигенетической регуляции эндогенной ДНК; обеспечения повышения или снижения экспрессии функционально связанных генов и т. д. В определенных вариантах осуществления индивидуально адаптированную последовательность РНК-матрицы можно сконструировать таким образом, чтобы она содержала последовательности, кодирующие экзоны и/или трансгены, обеспечивающие сайты связывания с активаторами, репрессорами, энхансерами факторов транскрипции и т. д., а также их комбинации. В других вариантах осуществления кодирующую последовательность можно дополнительно индивидуально адаптировать с помощью акцепторных сайтов сплайсинга, поли(А)-хвостов. В определенных вариантах осуществления последовательность РНК может содержать последовательности, кодирующие последовательность РНК-матрицы, гомологичной транспозазе RLE, являться сконструированной таким образом, чтобы она содержала гетерологичные кодирующие последовательности, или представлять собой комбинации вышеперечисленного.

РНК-матрица может характеризоваться некоторой степенью гомологии с ДНК-мишенью. В некоторых вариантах осуществления РНК-матрица имеет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 или больше оснований с точной гомологией с ДНК-мишенью на 3'-конце РНК. В некоторых вариантах осуществления РНК-матрица имеет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 175, 180 или 200 или 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 РНК-матрицы. В некоторых вариантах осуществления РНК-матрица имеет 3'-нетранслируемую область, полученную из не содержащего LTR ретротранспозона, например, не содержащих LTR ретротранспозонов, описанных в данном документе. В некоторых вариантах осуществления РНК-матрица имеет 3'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 или больше оснований с по

меньшей мере 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% или 100% гомологией с 3'-последовательностью не содержащего LTR ретротранспозона, например, не содержащего LTR ретротранспозона, описанного в данном документе, например, не содержащего LTR ретротранспозона из таблицы 1, 2 или 3. В некоторых вариантах осуществления РНК-матрица имеет 5'-нетранслируемую область, полученную из не содержащего LTR ретротранспозона, например, не содержащих LTR ретротранспозонов, описанных в данном документе. В некоторых вариантах осуществления РНК-матрица имеет 5'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180 или 200 или больше оснований с по меньшей мере 40%, 50%, 60%, 70%, 80%, 90%, 95% или большей гомологией с 5'-последовательностью не содержащего LTR ретротранспозона, например, не содержащего LTR ретротранспозона, описанного в данном документе, например, не содержащего LTR ретротранспозона, описанного в таблице 2 или 3.

Компонент системы редактирования генома Gene Writer, представляющий собой РНК-матрицу, описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления РНК-матрица имеет 3'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 3'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer.

Компонент системы редактирования генома Gene Writer, представляющий собой РНК-матрицу, описанный в данном документе, как правило, способен связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления РНК-матрица имеет 5'-область, которая способна связывать белок Gene Writer, редактирующий геном. Область связывания, например, 5'-область, может представлять собой структурированную область РНК, например, имеющую по меньшей мере 1, 2 или 3 шпилечные петли, способные связывать белок системы редактирования генома Gene Writer. В некоторых вариантах осуществления 5'-нетранслируемая область содержит псевдоузел, например, псевдоузел, который способен связываться с белком Gene Writer.

В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит последовательность структуры "стебель-петля". В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит шпильку. В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит спираль. В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 5'-нетранслируемая область) содержит псевдоузел. В некоторых вариантах осуществления РНК-матрица содержит рибозим. В некоторых вариантах осуществления рибозим сходен с рибозимом вируса гепатита дельта (HDV), например, имеет вторичную структуру, сходную с таковой у рибозима HDV, и/или обладает одной или несколькими формами активности

рибозима HDV, например, активностью саморасщепления. См., например, Eickbush et al., Molecular and Cellular Biology, 2010, 3142-3150.

В некоторых вариантах осуществления РНК-матрица (например, нетранслируемая область шпилечной РНК, например, 3'-нетранслируемая область) содержит одну или несколько структур "стебель-петля" или спиралей. Иллюстративные структуры 3'-UTR R2 показаны, например, в Ruschak et al. "Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs", RNA. 2004 Jun; 10(6): 978-987, например, на фигуре 3 в ней, и в Eikbush and Eikbush, "R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements", Mobile DNA (2012) 3:10; например, на фигуре 3 в ней, при этом содержание данных статей настоящим включено посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления РНК-матрица, описанная в данном документе, содержит последовательность, которая способна связываться с белком Gene Writer, описанным в данном документе. Например, в некоторых вариантах осуществления РНК-матрица содержит последовательность РНК MS2, способную связываться с последовательностью белка оболочки MS2 в белке Gene Writer. В некоторых вариантах осуществления РНК-матрица содержит последовательность РНК, способную связываться с последовательностью B-бокса. В некоторых вариантах осуществления РНК-матрица содержит последовательность РНК (например, последовательность crRNA и/или последовательность tracrRNA), способную связываться с последовательностью dCas в белке Gene Writer. В некоторых вариантах осуществления в дополнение к UTR или вместо нее РНК-матрица связана (например, ковалентной связью) с объектом, отличным от UTR РНК, например, белком или малой молекулой.

В некоторых вариантах осуществления РНК-матрица имеет поли(A)-хвост на 3'-конце. В некоторых вариантах осуществления РНК-матрица не имеет поли(A)-хвост на 3'-конце.

В некоторых вариантах осуществления РНК-матрица имеет 5'-область из по меньшей мере 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 или больше оснований с по меньшей мере 40%, 50%, 60%, 70%, 80%, 90%, 95% или большей гомологией с 5'-последовательностью не содержащего LTR ретротранспозона, например, не содержащего LTR ретротранспозона, описанного в данном документе.

РНК-матрица системы обычно содержит последовательность объекта для вставки в ДНК-мишень. Последовательность объекта может быть кодирующей или некодирующей.

В некоторых вариантах осуществления в системе или способе, описанных в данном документе, предусмотрена одна РНК-матрица. В некоторых вариантах осуществления в системе или способе, описанных в данном документе, предусмотрено множество РНК-матриц. В некоторых вариантах осуществления ДНК, кодирующая матрицу, циркуляризуется под действием ферментов, таких как рекомбиназы, для увеличения активности, как описано в Yant et al., *Nature Biotechnology* 20: 990-1005, 2002.

В некоторых вариантах осуществления последовательность гетерологичного объекта может содержать открытую рамку считывания. В некоторых вариантах осуществления РНК-матрица имеет последовательность Козак. В некоторых вариантах осуществления РНК-матрица имеет участок внутренней посадки рибосомы. В некоторых вариантах осуществления РНК-матрица имеет сайт саморасщепляющегося пептида, такого как T2A или P2A. В некоторых вариантах осуществления РНК-матрица имеет старт-кодон. В некоторых вариантах осуществления РНК-матрица имеет акцепторный сайт сплайсинга. В некоторых вариантах осуществления РНК-матрица имеет донорный сайт сплайсинга. Иллюстративные акцепторные и донорные сайты сплайсинга описаны в WO2016044416, включенной в данный документ посредством ссылки во всей своей полноте. Иллюстративные последовательности акцепторных сайтов сплайсинга известны специалистам в данной области и включают, только в качестве примера, CTGACCCTTCTCTCTCTCCCCCAGAG (SEQ ID NO: 1620) (из гена HBB человека) и TTTCTCTCCCACAAG (SEQ ID NO: 1621) (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 в РНК-матрице последовательность гетерологичного объекта кодирует полипептид и кодируется в антисмысловом направлении по отношению к 5'- и 3'-UTR. В некоторых вариантах осуществления в РНК-матрице последовательность гетерологичного объекта кодирует полипептид и кодируется в смысловом направлении по отношению к 5'- и 3'-UTR.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК-матрица или ДНК, кодирующая РНК-матрицу), содержит сайт связывания с микроРНК. В некоторых вариантах осуществления сайт связывания с микроРНК используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, сайт связывания с микроРНК может быть выбран на основании того, что он распознается miRNA, которая присутствует в типе клеток, не являющихся мишенями, но которая отсутствует (или присутствует на сниженном уровне по сравнению с клетками, не являющимися мишенями) в типе клеток, являющихся мишенями. Таким образом, если РНК-матрица присутствует в клетке, не являющейся мишенью, то она будет

связываться с miRNA, а если РНК-матрица присутствует в клетке-мишени, то она не будет связываться с miRNA (или будет связываться, но на сниженных уровнях по сравнению с клеткой, не являющейся мишенью). Не желая ограничиваться какой-либо теорией, полагают, что связывание miRNA с РНК-матрицей может препятствовать вставке последовательности гетерологичного объекта в геном. Соответственно, вставка последовательности гетерологичного объекта в геном клеток-мишеней будет происходить более эффективно, чем в геном клеток, не являющихся мишенями. Система, имеющая сайт связывания с микроРНК в РНК-матрице (или кодирующей ее ДНК), также может применяться в комбинации с нуклеиновой кислотой, кодирующей полипептид Gene Writer, при этом экспрессия полипептида Gene Writer регулируется вторым сайтом связывания с микроРНК, например, как описано в данном документе, например, в разделе под названием "Полипептидный компонент системы редактирования генов GeneWriter". В некоторых вариантах осуществления, например, для печеночных показаний, miRNA выбрана из таблицы 4 WO2020014209.

В некоторых вариантах осуществления последовательность объекта может содержать некодирующую последовательность. 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 РНК-матрица содержит тканеспецифический промотор или энхансер, каждый из которых может быть однонаправленным или двунаправленным. В некоторых вариантах осуществления промотор представляет собой промотор РНК-полимеразы I, промотор РНК-полимеразы II или промотор РНК-полимеразы III. В некоторых вариантах осуществления промотор содержит элемент TATA. В некоторых вариантах осуществления промотор содержит элемент распознавания B. В некоторых вариантах осуществления промотор имеет один или несколько сайтов связывания для факторов транскрипции. В некоторых вариантах осуществления некодирующая последовательность транскрибируется в антисмысловом направлении по отношению к 5'- и 3'-UTR. В некоторых вариантах осуществления некодирующая последовательность транскрибируется в смысловом направлении по отношению к 5'- и 3'-UTR.

В некоторых вариантах осуществления нуклеиновая кислота, описанная в данном документе (например, РНК-матрица или ДНК, кодирующая РНК-матрицу), содержит последовательность промотора, например, последовательность тканеспецифического промотора. В некоторых вариантах осуществления тканеспецифический промотор используется для увеличения специфичности системы Gene Writer в отношении клеток-мишеней. Например, промотор может быть выбран на основании того, что он является активным в типе клеток, являющихся мишенями, но не является активным (или является активным на более низком уровне) в типе клеток, не являющихся мишенями. Таким

образом, даже если промотор интегрирован в геном клетки, не являющейся мишенью, он не будет управлять экспрессией (или будет управлять экспрессией только на низком уровне) интегрированного гена. Система, содержащая последовательность тканеспецифического промотора в РНК-матрице, также может применяться в комбинации с сайтом связывания с микроРНК, например, в РНК-матрице или нуклеиновой кислоте, кодирующей белок Gene Writer, например, как описано в данном документе. Система, содержащая последовательность тканеспецифического промотора в РНК-матрице, также может применяться в комбинации с ДНК, кодирующей полипептид Gene Writer, под управлением тканеспецифического промотора, например, для достижения более высоких уровней белка Gene Writer в клетках-мишенях, чем в клетках, не являющихся мишенями. В некоторых вариантах осуществления, например, для печеночных показаний, тканеспецифический промотор выбран из таблицы 3 WO2020014209, которая включена 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, NK-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 NK-клетках, промоторы, которые активны как в Т-клетках, так и в NK-клетках.

В некоторых вариантах осуществления РНК-матрица содержит последовательность микроРНК, последовательность siRNA, последовательность направляющей РНК, последовательность РНК, взаимодействующей с Piwi.

В некоторых вариантах осуществления РНК-матрица содержит сайт, который координирует эпигенетическую модификацию. В некоторых вариантах осуществления РНК-матрица содержит элемент, который ингибирует, например, предотвращает, эпигенетический сайленсинг. В некоторых вариантах осуществления РНК-матрица

содержит инсулятор хроматина. Например, РНК-матрица содержит сайт для CTCF или сайт, являющийся мишенью метилирования ДНК.

Для способствования более высокому уровню или более стабильной экспрессии гена РНК-матрица может содержать признаки, которые обеспечивают предотвращение или ингибирование сайленсинга генов. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование метилирования ДНК. В некоторых вариантах осуществления эти признаки способствуют деметилированию ДНК. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование деацетилирования гистонов. В некоторых вариантах осуществления эти признаки обеспечивают предотвращение или ингибирование метилирования гистонов. В некоторых вариантах осуществления эти признаки способствуют ацетилированию гистонов. В некоторых вариантах осуществления эти признаки способствуют деметилированию гистонов. В некоторых вариантах осуществления в РНК-матрицу могут быть включены несколько признаков для способствования одной или нескольким таким модификациям. CpG-динуклеотиды подвергаются метилированию метилтрансферазами хозяина. В некоторых вариантах осуществления РНК-матрица лишена CpG-динуклеотидов, например, не содержит CpG-нуклеотиды или содержит сниженное количество CpG-динуклеотидов по сравнению с соответствующей неизмененной последовательностью. В некоторых вариантах осуществления промотор, управляющий экспрессией трансгена из интегрированной ДНК, лишен CpG-динуклеотидов.

В некоторых вариантах осуществления РНК-матрица содержит единицу экспрессии гена, состоящую из по меньшей мере одной регуляторной области, функционально связанной с эффекторной последовательностью. Эффекторная последовательность может представлять собой последовательность, которая транскрибируется в РНК (например, кодирующую последовательность или некодирующую последовательность, такую как последовательность, кодирующая микроРНК).

В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень в эндогенном интроне. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень и таким образом выступает в качестве нового экзона. В некоторых вариантах осуществления вставка последовательности объекта в геном-мишень приводит к замене природного экзона или пропуску природного экзона.

В некоторых вариантах осуществления последовательность объекта из РНК-матрицы вставляется в геном-мишень в геномном сайте "безопасная гавань", таком как AAVS1, CCR5, ROSA26 или локус альбумина. В некоторых вариантах осуществления Gene Writer применяют для интеграции CAR в локус константной области α-цепи Т-клеточного рецептора (*TRAC*) (Eyquem 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 РНК-матрицы добавлена в геном в межгенной или внутригенной области. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы добавлена в геном с 5'- или 3'-стороны в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от эндогенного активного гена. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы добавляется в геном с 5'- или 3'-стороны в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от эндогенного промотора или энхансера. В некоторых вариантах осуществления последовательность объекта из РНК-матрицы может составлять, например, 50-50000 пар оснований (например, 50-40000 п. о., 500-30000 п. о., 500-20000 п. о., 100-15000 п. о., 500-10000 п. о., 50-10000 п. о., 50-5000 п. о.). В некоторых вариантах осуществления последовательность гетерологичного объекта имеет длину, составляющую менее 1000, 1300, 1500, 2000, 3000, 4000, 5000 или 7500 нуклеотидов.

В некоторых вариантах осуществления геномный сайт "безопасная гавань" представляет собой сайт Natural Harbor™. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой рибосомную ДНК (рДНК). В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой 5S рДНК, 18S рДНК, 5,8S рДНК или 28S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт Mutsu в 5S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт R2, сайт R5, сайт R6, сайт R4, сайт R1, сайт R9 или сайт RT в 28S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт R8 или сайт R7 в 18S рДНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую транспортную РНК (тРНК). В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую тРНК-Asp или тРНК-Glu. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую сплайсосомную РНК. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой ДНК, кодирующую малую ядерную РНК (snRNA), такую как snRNA U2.

Таким образом, в некоторых аспектах в настоящем изобретении предусмотрен способ обеспечения вставки последовательности гетерологичного объекта в сайт Natural Harbor™. В некоторых вариантах осуществления способ включает применение системы Gene Writer, описанной в данном документе, например, применение полипептида из любой из таблиц 1-3 или полипептида, характеризующегося сходством последовательности с ним, например, по меньшей мере 80%, 85%, 90% или 95% идентичностью с ним. В некоторых вариантах осуществления способ включает применение фермента, например, ретротранспозазы, для вставки последовательности гетерологичного объекта в сайт Natural Harbor™. В некоторых аспектах в настоящем изобретении предусмотрена клетка-хозяин, являющаяся клеткой человека, содержащая последовательность гетерологичного объекта

(например, последовательность, кодирующую терапевтический полипептид), находящуюся в сайте Natural Harbor™ в геноме клетки. В некоторых вариантах осуществления сайт Natural Harbor™ представляет собой сайт, описанный в таблице 5 ниже. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований от последовательности, показанной в таблице 5. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от последовательности, показанной в таблице 5. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в сайт, характеризующийся по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью, показанной в таблице 5. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований или в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от сайта, характеризующегося по меньшей мере 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичностью с последовательностью, показанной в таблице 5. В некоторых вариантах осуществления последовательность гетерологичного объекта вставлена в ген, указанный в столбце 5 таблицы 5, или в пределах 20, 50, 100, 150, 200, 250, 500 или 1000 пар оснований или в пределах 0,1 т. о., 0,25 т. о., 0,5 т. о., 0,75 т. о., 1 т. о., 2 т. о., 3 т. о., 4 т. о., 5 т. о., 7,5 т. о., 10 т. о., 15 т. о., 20 т. о., 25 т. о., 50, 75 т. о. или 100 т. о. от данного гена.

Таблица 5. Сайты Natural Harbor™. В столбце 1 указан ретротранспозон, который осуществляет вставку в сайт Natural Harbor™. В столбце 2 указан ген в сайте Natural Harbor™. В столбцах 3 и 4 показаны иллюстративные последовательности человеческого генома с 5'- и 3'-стороны от сайта вставки (например, длиной 250 п. о.). В столбцах 5 и 6 приведены иллюстративный символ гена и соответствующий ID гена.

Таблица 5. Иллюстративные сайты Natural Harbor™.

Сайт-мишень	**Ген-мишень**	**5'-фланкирующая последовательность**	**3'-фланкирующая последовательность**	**Иллюстративный символ гена**	**Иллюстративный ID гена**
R2	28S рДНК	CCGGTCCCCCCCGCC GGGTCCGCCCCCGG GGCCGCGGTTCCGCG CGGCGCCTCGCCTCG GCCGGCGCCTAGCA	GTAGCCAAATGCCTC GTCATCTAATTAGTGA CGCGCATGAATGGAT GAACGAGATTCCCAC TGTCCCTACCTACTAT	RNA28SN1	106632264

		GCCGACTTAGAACTG GTGCGGACCAGGGG AATCCGACTGTTTAA TTAAAACAAAGCAT CGCGAAGGCCCGCG GCGGGTGTTGACGC GATGTGATTTCTGCC CAGTGCTCTGAATGT CAAAGTGAAGAAAT TCAATGAAGCGCGG GTAAACGGCGGGAG TAACTATGACTCTCT TAAG (SEQ ID NO: 1508)	CCAGCGAAACCACAG CCAAGGGAACGGGCT TGGCGGAATCAGCGG GGAAAGAAGACCCTG TTGAGCTTGACTCTAG TCTGGCACGGTGAAG AGACATGAGAGGTGT AGAATAAGTGGGAGG CCCCCGGCGCCCCCCC GGTGTCCCCGCGAGG GGCCCGGGGCGGGGT CCGCCG (SEQ ID NO: 1513)		
R4	28S рДНК	GCGGTTCCGCGCGGC GCCTCGCCTCGGCCG GCGCCTAGCAGCCG ACTTAGAACTGGTGC GGACCAGGGGAATC CGACTGTTTAATTAA AACAAAGCATCGCG AAGGCCCGCGGCGG GTGTTGACGCGATGT GATTTCTGCCCAGTG CTCTGAATGTCAAAG TGAAGAAATTCAAT GAAGCGCGGGTAAA CGGCGGGAGTAACT ATGACTCTCTTAAGG TAGCCAAATGCCTCG TCATCTAATTAGTGA CG (SEQ ID NO: 1509)	CGCATGAATGGATGA ACGAGATTCCCACTGT CCCTACCTACTATCCA GCGAAACCACAGCCA AGGGAACGGGCTTGG CGGAATCAGCGGGGA AAGAAGACCCTGTTG AGCTTGACTCTAGTCT GGCACGGTGAAGAGA CATGAGAGGTGTAGA ATAAGTGGGAGGCCC CCGGCGCCCCCCCGGT GTCCCCGCGAGGGGC CCGGGGCGGGGTCCG CCGGCCCTGCGGGCC GCCGGTGAAATACCA CTACTC (SEQ ID NO: 1514)	RNA28SN1	106632264
R5	28S рДНК	TCCCCCCCGCCGGGT CCGCCCCCGGGGCC GCGGTTCCGCGCGGC	CCAAATGCCTCGTCAT CTAATTAGTGACGCGC ATGAATGGATGAACG	RNA28SN1	106632264

		GCCTCGCCTCGGCCG GCGCCTAGCAGCCG ACTTAGAACTGGTGC GGACCAGGGGAATC CGACTGTTTAATTAA AACAAAGCATCGCG AAGGCCCGCGGCGG GTGTTGACGCGATGT GATTTCTGCCCAGTG CTCTGAATGTCAAAG TGAAGAAATTCAAT GAAGCGCGGGTAAA CGGCGGGAGTAACT ATGACTCTCTTAAGG TAG (SEQ ID NO: 1510)	AGATTCCCACTGTCCC TACCTACTATCCAGCG AAACCACAGCCAAGG GAACGGGCTTGGCGG AATCAGCGGGGAAAG AAGACCCTGTTGAGCT TGACTCTAGTCTGGCA CGGTGAAGAGACATG AGAGGTGTAGAATAA GTGGGAGGCCCCCGG CGCCCCCCCGGTGTCC CCGCGAGGGGCCCGG GGCGGGGTCCGCCGG CCC (SEQ ID NO: 1515)		
R9	28S рДНК	CGGCGCGCTCGCCG GCCGAGGTGGGATC CCGAGGCCTCTCCAG TCCGCCGAGGGCGC ACCACCGGCCCGTCT CGCCCGCCGCGCCG GGGAGGTGGAGCAC GAGCGCACGTGTTA GGACCCGAAAGATG GTGAACTATGCCTGG GCAGGGCGAAGCCA GAGGAAACTCTGGT GGAGGTCCGTAGCG GTCCTGACGTGCAAA TCGGTCGTCCGACCT GGGTATAGGGGCGA AAGACTAATCGAAC CATCTAG (SEQ ID NO: 1511)	TAGCTGGTTCCCTCCG AAGTTTCCCTCAGGAT AGCTGGCGCTCTCGCA GACCCGACGCACCCC CGCCACGCAGTTTTAT CCGGTAAAGCGAATG ATTAGAGGTCTTGGG GCCGAAACGATCTCA ACCTATTCTCAAACTT TAAATGGGTAAGAAG CCCGGCTCGCTGGCGT GGAGCCGGGCGTGGA ATGCGAGTGCCTAGT GGGCCACTTTTGGTAA GCAGAACTGGCGCTG CGGGATGAACCGAAC GCC (SEQ ID NO: 1516)	RNA28SN1	106632264

R8	18S рДНК	GCATTCGTATTGCGCCGCTAGAGGTGAAATTCTTGGACCGGCGCAAGACGGACCAGAGCGAAAGCATTTGCCAAGAATGTTTTCATTAATCAAGAACGAAAGTCGGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCCGACCATAAACGATGCCGACCGGCGATGCGGCGGCGTTATTCCCATGACCCGCCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTTGCAAAGC (SEQ ID NO: 1512)	TGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCCGGCCCGGACACGGACAGGATTGACAGATTGATAGCTCTTTCTCGATTCCGTGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTGGCATGCTAACTAGTTACGCGACCCCCGAGCGGTCGGCGTCCC (SEQ ID NO: 1517)	RNA18SN1	106631781
R4-2_SRa	тРНК-Asp			TRD-GTC1-1	100189207
LIN25_SM	тРНК-Glu			TRE-CTC1-1	100189384
R1	28S рДНК	TAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGACTGTTTAATTAAAACAAAGCATCGCGAAGGCCCGCGGCGGGTGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCG	ACCTACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCCCCCGGCGCCCCCCCGGTGTCCCCGCGAGGGGCCCGGG	RNA28SN1	106632264

		GGAGTAACTATGACT CTCTTAAGGTAGCCA AATGCCTCGTCATCT AATTAGTGACGCGC ATGAATGGATGAAC GAGATTCCCACTGTC CCT (SEQ ID NO: 1518)	GCGGGGTCCGCCGGC CCTGCGGGCCGCCGG TGAAATACCACTACTC TGATCGTTTTTTCACT GACCCGGTGAGGCGG GGGG (SEQ ID NO: 1524)		
R6	28S рДНК	CCCCCCGCCGGGTCC GCCCCCGGGGCCGC GGTTCCGCGCGGCGC CTCGCCTCGGCCGGC GCCTAGCAGCCGACT TAGAACTGGTGCGG ACCAGGGGAATCCG ACTGTTTAATTAAAA CAAAGCATCGCGAA GGCCCGCGGCGGGT GTTGACGCGATGTGA TTTCTGCCCAGTGCT CTGAATGTCAAAGTG AAGAAATTCAATGA AGCGCGGGTAAACG GCGGGAGTAACTAT GACTCTCTTAAGGTA GCC (SEQ ID NO: 1519)	AAATGCCTCGTCATCT AATTAGTGACGCGCA TGAATGGATGAACGA GATTCCCACTGTCCCT ACCTACTATCCAGCGA AACCACAGCCAAGGG AACGGGCTTGGCGGA ATCAGCGGGGAAAGA AGACCCTGTTGAGCTT GACTCTAGTCTGGCAC GGTGAAGAGACATGA GAGGTGTAGAATAAG TGGGAGGCCCCCGGC GCCCCCCCGGTGTCCC CGCGAGGGGCCCGGG GCGGGGTCCGCCGGC CCTG (SEQ ID NO: 1525)	RNA28SN1	10663 2264
R7	18S рДНК	GCGCAAGACGGACC AGAGCGAAAGCATT TGCCAAGAATGTTTT CATTAATCAAGAAC GAAAGTCGGAGGTT CGAAGACGATCAGA TACCGTCGTAGTTCC GACCATAAACGATG CCGACCGGCGATGC	GGAGCCTGCGGCTTA ATTTGACTCAACACGG GAAACCTCACCCGGC CCGGACACGGACAGG ATTGACAGATTGATA GCTCTTTCTCGATTCC GTGGGTGGTGGTGCA TGGCCGTTCTTAGTTG GTGGAGCGATTTGTCT	RNA18SN1	10663 1781

		GGCGGCGTTATTCCC ATGACCCGCCGGGC AGCTTCCGGGAAAC CAAAGTCTTTGGGTT CCGGGGGGAGTATG GTTGCAAAGCTGAA ACTTAAAGGAATTG ACGGAAGGGCACCA CCAGGAGT (SEQ ID NO: 1520)	GGTTAATTCCGATAAC GAACGAGACTCTGGC ATGCTAACTAGTTACG CGACCCCCGAGCGGT CGGCGTCCCCCAACTT CTTAGAGGGACAAGT GGCGTTCAGCCACCC GAG (SEQ ID NO: 1526)		
RT	28S рДНК	GGCCGGGCGCGACC CGCTCCGGGGACAG TGCCAGGTGGGGAG TTTGACTGGGGCGGT ACACCTGTCAAACG GTAACGCAGGTGTCC TAAGGCGAGCTCAG GGAGGACAGAAACC TCCCGTGGAGCAGA AGGGCAAAAGCTCG CTTGATCTTGATTTT CAGTACGAATACAG ACCGTGAAAGCGGG GCCTCACGATCCTTC TGACCTTTTGGGTTT TAAGCAGGAGGTGT CAGAAAAGTTACCA CAGGGAT (SEQ ID NO: 1521)	AACTGGCTTGTGGCG GCCAAGCGTTCATAG CGACGTCGCTTTTTGA TCCTTCGATGTCGGCT CTTCCTATCATTGTGA AGCAGAATTCACCAA GCGTTGGATTGTTCAC CCACTAATAGGGAAC GTGAGCTGGGTTTAG ACCGTCGTGAGACAG GTTAGTTTTACCCTAC TGATGATGTGTTGTTG CCATGGTAATCCTGCT CAGTACGAGAGGAAC CGCAGGTTCAGACATT TGGTGTATGTGCTTGG C (SEQ ID NO: 1527)	RNA28SN1	10663 2264
Mutsu	5S рДНК	GTCTACGGCCATACC ACCC (SEQ ID NO: 1522)	TGAACGCGCCCGATCT CGTCTGATCTCGGAAG CTAAGCAGGGTCGGG CCTGGTTAGTACTTGG ATGGGAGACCGCCTG GGAATACCGGGTGCT	RNA5S1	10016 9751

			GTAGGCTTT (SEQ ID NO: 1528)		
Utopia/Keno	snRNA U2	ATCGCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTA (SEQ ID NO: 1523)	TCTGTTCTTATCAGTTTAATATCTGATACGTCCTCTATCCGAGGACAATATATTAAATGGATTTTTGGAGCAGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGCACCC (SEQ ID NO: 1529)	RNU2-1	6066

В некоторых вариантах осуществления система или способ, описанные в данном документе, приводят к вставке гетерологичной последовательности в сайт-мишень в человеческом геноме. В некоторых вариантах осуществления сайт-мишень в человеческом геноме характеризуется сходством последовательности с соответствующим сайтом-мишенью соответствующей ретротранспозазы дикого типа (например, ретротранспозазы, из которой был получен Gene Writer) в геноме организма, для которого он является нативным. Например, в некоторых вариантах осуществления идентичность между 40 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 40 нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%. В некоторых вариантах осуществления идентичность между 100 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 100 нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%. В некоторых вариантах осуществления идентичность между 500 нуклеотидами последовательности человеческого генома с центром в сайте вставки и 500 нуклеотидами последовательности генома нативного организма с центром в сайте вставки составляет менее 99,5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60% или 50% или составляет 50-60%, 60-70%, 70-80%, 80-90% или 90-100%.

Дополнительные признаки матрицы

В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые *in silico*. В вариантах осуществления предполагается, что РНК-матрица имеет структуры с минимальной энергией, составляющей от -280 до -480 ккал/моль (например, от -280 до -300, от -300 до -

350, от -350 до -400, от -400 до -450, или от -450 до -480 ккал/моль), например, при измерении структуры РНК, например, как описано в Turner and Mathews *Nucleic Acids Res* 38:D280-282 (2009) (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые *in vitro.* В вариантах осуществления РНК-матрица оптимизирована по последовательности, например, для ослабления вторичной структуры, определенной *in vitro*, например, с помощью SHAPE-MaP (например, как описано в Siegfried et al. Nat Methods 11:959-965 (2014); включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления матрица (например, РНК-матрица) содержит определенные структурные признаки, например, определяемые в клетках. В вариантах осуществления РНК-матрица оптимизирована по последовательности, например, для ослабления вторичной структуры, измеренной в клетках, например, с помощью DMS-MaPseq (например, как описано в Zubradt et al. Nat Methods 14:75-82 (2017); включенном в данный документ посредством ссылки во всей своей полноте).

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.

Полипептид Gene Writer

ДНК-связывающий домен

В некоторых вариантах осуществления ДНК-связывающий домен способен связываться с последовательностью-мишенью (например, последовательностью dsDNA-мишени) с большей аффинностью, чем эталонный ДНК-связывающий домен. В некоторых вариантах осуществления эталонный ДНК-связывающий домен представляет собой ДНК-связывающий домен из R2_BM B. mori. В некоторых вариантах осуществления ДНК-связывающий домен способен связываться с последовательностью-мишенью (например,

последовательностью dsDNA-мишени) с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ).

В некоторых вариантах осуществления аффинность ДНК-связывающего домена к его последовательности-мишени (например, последовательности dsDNA-мишени) измеряют in vitro, например, с помощью термофореза, например, описанного в Asmari et al. Methods 146:107-119 (2018) (включенном в данный документ посредством ссылки во всей своей полноте).

В вариантах осуществления ДНК-связывающий домен способен связываться со своей последовательностью-мишенью (например, последовательностью dsDNA-мишени), например, с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ), в присутствии молярного избытка конкурентной dsDNA со скремблированной последовательностью, например, приблизительно 100-кратного молярного избытка.

В некоторых вариантах осуществления ДНК-связывающий домен обнаруживается ассоциированным с последовательностью-мишенью (например, последовательностью dsDNA-мишени) чаще, чем с любой другой последовательностью в геноме клетки-мишени, например, клетки-мишени человека, например, при измерении с помощью ChIP-seq (например, в клетках HEK293T), например, как описано в He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления ДНК-связывающий домен обнаруживается ассоциированным с последовательностью-мишенью (например, последовательностью dsDNA-мишени) в по меньшей мере приблизительно 5 раз или 10 раз чаще, чем с любой другой последовательностью в геноме клетки-мишени, например, при измерении с помощью ChIP-seq (например, в клетках HEK293T), например, как описано в He and Pu (2010), выше.

В некоторых вариантах осуществления полипептид Gene Writer содержит модификацию ДНК-связывающего домена, например, по сравнению с полипептидом дикого типа. В некоторых вариантах осуществления ДНК-связывающий домен предусматривает добавление, делецию, замену или модификацию аминокислотной последовательности исходного ДНК-связывающего домена. В некоторых вариантах осуществления ДНК-связывающий домен модифицирован таким образом, чтобы он содержал гетерологичный функциональный домен, который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В некоторых вариантах осуществления функциональный домен заменяет по меньшей мере часть предшествующего ДНК-связывающего домена полипептида (например, его полностью). В некоторых вариантах осуществления функциональный домен содержит "цинковый палец" (например, "цинковый палец", который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В некоторых вариантах осуществления функциональный домен содержит домен Cas (например, домен Cas, который специфично связывается с

последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес. В вариантах осуществления домен Cas содержит Cas9 или его мутантную форму или вариант (например, как описано в данном документе). В вариантах осуществления домен Cas ассоциирован с направляющей РНК (gRNA), например, как описано в данном документе. В вариантах осуществления домен Cas направлен на последовательность нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес, посредством gRNA. В вариантах осуществления домен Cas кодируется в той же молекуле нуклеиновой кислоты (например, РНК), что и gRNA. В вариантах осуществления домен Cas кодируется в молекуле нуклеиновой кислоты (например, РНК), отличной от gRNA.

РНК-связывающий домен

В некоторых вариантах осуществления РНК-связывающий домен способен связываться с РНК-матрицей с большей аффинностью, чем эталонный РНК-связывающий домен. В некоторых вариантах осуществления эталонный РНК-связывающий домен представляет собой РНК-связывающий домен из R2_BM B. mori. В некоторых вариантах осуществления РНК-связывающий домен способен связываться с РНК-матрицей с аффинностью, составляющей от 100 пМ до 10 нМ (например, от 100 пМ до 1 нМ или от 1 нМ до 10 нМ). В некоторых вариантах осуществления аффинность РНК-связывающего домена к его РНК-матрице измеряют in vitro, например, с помощью термофореза, например, описанного в Asmari et al. Methods 146:107-119 (2018) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления аффинность РНК-связывающего домена к его РНК-матрице измеряют в клетках (например, с помощью FRET или CLIP-Seq).

В некоторых вариантах осуществления РНК-связывающий домен ассоциирован с РНК-матрицей *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной РНК. В некоторых вариантах осуществления частоту ассоциации между РНК-связывающим доменом и РНК-матрицей или скремблированной РНК измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления РНК-связывающий домен ассоциирован с РНК-матрицей в клетках (например, в клетках HEK293T) с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной РНК. В некоторых вариантах осуществления частоту ассоциации между РНК-связывающим доменом и РНК-матрицей или скремблированной РНК измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019), выше.

Эндонуклеазный домен

В некоторых вариантах осуществления эндонуклеазный домен ассоциирован с dsDNA-мишенью *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной dsDNA. В некоторых вариантах осуществления эндонуклеазный домен ассоциирован с dsDNA-мишенью *in vitro* с частотой в по меньшей мере приблизительно 5 раз или 10 раз выше, чем со скремблированной dsDNA, например,

в клетке (например, клетке HEK293T). В некоторых вариантах осуществления частоту ассоциации между эндонуклеазным доменом и ДНК-мишенью или скремблированной ДНК измеряют с помощью ChIP-seq, например, как описано в He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен может катализировать образование однонитевого разрыва в последовательности-мишени, например, с увеличением в по меньшей мере приблизительно 5 раз или 10 раз по сравнению с последовательностью, не являющейся мишенью (например, по сравнению с любой другой геномной последовательностью в геноме клетки-мишени). В некоторых вариантах осуществления уровень образования однонитевых разрывов определяют с использованием NickSeq, например, как описано в Elacqua et al. (2019) *BioRxiv* doi.org/10.1101/867937 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК in vitro. В вариантах осуществления однонитевой разрыв приводит к 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, выявленный с помощью анализа чувствительности к нуклеазе) увеличен на по меньшей мере 10%, 50% или больше по сравнению с эталонным эндонуклеазным доменом. В некоторых вариантах осуществления эталонный эндонуклеазный домен представляет собой эндонуклеазный домен из R2_BM B. mori.

В некоторых вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК в клетке. В вариантах осуществления эндонуклеазный домен способен вносить однонитевой разрыв в ДНК в клетке HEK293T. В вариантах осуществления неисправленный однонитевой разрыв, который подвергается репликации в отсутствие Rad51, приводит к увеличению частоты NHEJ в месте однонитевого разрыва, что может быть выявлено, например, с помощью анализа ингибирования Rad51, например, описанного в Bothmer et al. (2017) *Nat Commun* 8:13905 (включенном в данный документ посредством ссылки во всей своей полноте). В вариантах осуществления частота NHEJ увеличивается выше 0-5%. В вариантах осуществления частота NHEJ увеличивается до 20-70% (например, 30%-60% или 40-50%), например, при ингибировании Rad51.

В некоторых вариантах осуществления эндонуклеазный домен высвобождает мишень после расщепления. В некоторых вариантах осуществления высвобождение мишени определяется косвенно посредством оценки множественных оборотов фермента, например, как описано в Yourik al. RNA 25(1):35-44 (2019) (включенном в данный документ посредством ссылки во всей своей полноте) и показано на фигуре 2. В некоторых вариантах осуществления k_{exp} эндонуклеазного домена составляет 1×10^{-3}-1×10^{-5} мин-1 при измерении такими способами.

В некоторых вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью (k_{cat}/K_m), составляющей более чем приблизительно 1×10^8 c^{-1} M^{-1} in vitro. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью, составляющей более чем приблизительно 10^5, 1×10^6, 1×10^7 или 1×10^8 c^{-1} M^{-1} in vitro. В вариантах осуществления каталитическую эффективность определяют, как описано в Chen et al. (2018) *Science* 360(6387):436-439 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью (k_{cat}/K_m) в клетках, составляющей более чем приблизительно 1×10^8 c^{-1} M^{-1}. В вариантах осуществления эндонуклеазный домен характеризуется каталитической эффективностью в клетках, составляющей более чем приблизительно 10^5, 1×10^6, 1×10^7 или 1×10^8 c^{-1} M^{-1}.

В некоторых вариантах осуществления полипептид Gene Writer содержит модификацию эндонуклеазного домена, например, по сравнению с полипептидом дикого типа. В некоторых вариантах осуществления эндонуклеазный домен предусматривает добавление, делецию, замену или модификацию аминокислотной последовательности исходного эндонуклеазного домена. В некоторых вариантах осуществления эндонуклеазный домен модифицирован таким образом, чтобы он содержал гетерологичный функциональный домен, который специфично связывается с последовательностью нуклеиновой кислоты-мишени (например, ДНК), представляющей интерес, и/или индуцирует ее расщепление эндонуклеазой. В некоторых вариантах осуществления эндонуклеазный домен содержит "цинковый палец". В некоторых вариантах осуществления эндонуклеазный домен содержит домен Cas (например, Cas9 или его мутантную форму или вариант). В вариантах осуществления эндонуклеазный домен, содержащий домен Cas, ассоциирован с направляющей РНК (gRNA), например, как описано в данном документе. В некоторых вариантах осуществления эндонуклеазный домен модифицирован таким образом, чтобы он содержал функциональный домен, который не нацеливается на конкретную последовательность нуклеиновой кислоты-мишени (например, ДНК). В вариантах осуществления эндонуклеазный домен содержит домен Fok1.

Домен обратной транскриптазы

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется более низкой вероятностью частоты преждевременной терминации (P_{off}) *in vitro* по сравнению с эталонным доменом обратной транскриптазы. В некоторых вариантах осуществления эталонный домен обратной транскриптазы представляет собой домен обратной транскриптазы из R2_BM B. mori или вирусный домен обратной транскриптазы, например, домен RT из M-MLV.

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется более низкой вероятностью частоты преждевременной терминации (P_{off}) *in vitro*, составляющей менее чем приблизительно 5×10^{-3}/нуклеотид, 5×10^{-4}/нуклеотид или 5×10^{-6}/нуклеотид, например, при измерении в РНК длиной 1094 нуклеотида. В вариантах

осуществления частоту преждевременной терминации *in vitro* определяют, как описано в Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы способен осуществлять по меньшей мере приблизительно 30% или 50% полных интеграций в клетках. Процент полных интеграций может быть измерен путем деления количества событий по сути полноразмерной интеграции (например, геномных сайтов, которые составляют по меньшей мере 98% ожидаемой интегрированной последовательности) на общее количество событий интеграции (включая по сути полноразмерную и частичную) в популяции клеток. В вариантах осуществления интеграцию в клетках определяют (например, по сайту интеграции) с использованием секвенирования ампликонов с длинными ридами, например, как описано в Karst et al. (2020) *BioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте).

В вариантах осуществления количественная оценка интеграций в клетках включает подсчет доли интеграций, которые содержат по меньшей мере приблизительно 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% последовательности ДНК, соответствующей РНК-матрице (например, РНК-матрице длиной по меньшей мере 0,05, 0,1, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 3, 4 или 5 т. о., например, длиной 0,5-0,6, 0,6-0,7, 0,7-0,8, 0,8-0,9, 1,0-1,2, 1,2-1,4, 1,4-1,6, 1,6-1,8, 1,8-2,0, 2-3, 3-4 или 4-5 т. о.).

В некоторых вариантах осуществления домен обратной транскриптазы способен полимеризовать dNTP *in vitro*. В некоторых вариантах осуществления домен обратной транскриптазы способен полимеризовать dNTP *in vitro* со скоростью от 0,1 до 50 нуклеотидов/с (например, 0,1-1, 1-10 или 10-50 нуклеотидов/с). В вариантах осуществления полимеризацию dNTP с помощью домена обратной транскриптазы измеряют с помощью анализа отдельных молекул, например, описанного в Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы характеризуется частотой ошибок *in vitro* (например, неправильным включением нуклеотидов), составляющей от 1×10^{-3} до 1×10^{-4} или от 1×10^{-4} до 1×10^{-5} замен/нуклеотид, например, как описано в Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления домен обратной транскриптазы характеризуется частотой ошибок (например, неправильным включением нуклеотидов) в клетках (например, клетках HEK293T), составляющей от 1×10^{-3} до 1×10^{-4} или от 1×10^{-4} до 1×10^{-5} замен/нуклеотид, например, при секвенировании ампликонов с длинными ридами, например, как описано в Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы способен осуществлять обратную транскрипцию РНК-мишени in vitro. В некоторых вариантах осуществления обратной транскриптазе требуется праймер длиной по меньшей мере 3 нуклеотида для инициации обратной транскрипции матрицы. В некоторых вариантах осуществления обратную транскрипцию РНК-мишени определяют посредством выявления кДНК из РНК-мишени (например, при наличии праймера для ssDNA, например, который отжигается с мишенью с по меньшей мере 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидами на 3'-конце), например, как описано в Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы выполняет обратную транскрипцию в по меньшей мере 5 или 10 раз более эффективно (например, по продуцированию кДНК), например, при преобразовании своей РНК-матрицы в кДНК, например, по сравнению с РНК-матрицей, не имеющей белок-связывающего мотива (например, 3'-UTR). В вариантах осуществления эффективность обратной транскрипции измеряют, как описано в Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (включенном в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления домен обратной транскриптазы специфично связывает специфическую РНК-матрицу с большей частотой (например, в приблизительно 5 или 10 раз большей частотой), чем любую эндогенную клеточную РНК, например, при экспрессии в клетках (например, клетках HEK293T). В вариантах осуществления частоту специфичного связывания между доменом обратной транскриптазы и РНК-матрицей измеряют с помощью CLIP-seq, например, как описано в Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (включенном в данный документ посредством ссылки во всей своей полноте).

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 однонитевого разрыва в молекуле нуклеиновой кислоты-мишени. В вариантах осуществления связывание сайта связывания праймера с молекулой нуклеиновой кислоты-мишени приводит к праймированию TPRT.

В некоторых вариантах осуществления домен обратной транскриптазы (RT) демонстрирует повышенную строгость инициации обратной транскрипции с праймированием мишенью (TPRT), например, по сравнению с эндогенным доменом RT. В некоторых вариантах осуществления домен RT инициирует TPRT, когда 3 нуклеотида в сайте-мишени непосредственно перед однонитевым разрывом первой нити, например, геномная ДНК, праймирующая РНК-матрицу, характеризуются по меньшей мере 66% или 100% комплементарностью по отношению к 3 гомологичным нуклеотидам в РНК-матрице. В некоторых вариантах осуществления домен RT инициирует TPRT при наличии несовпадения менее 5 нуклеотидов (например, несовпадения менее 1, 2, 3, 4 или 5 нуклеотидов) между гомологичной РНК-матрицей и ДНК-мишенью, праймирующей обратную транскрипцию. В некоторых вариантах осуществления домен RT модифицирован таким образом, что строгость в отношении несовпадений при праймировании реакции TPRT увеличивается, например, где домен RT не допускает каких-либо несовпадений или допускает меньшее количество несовпадений в области праймирования по сравнению с доменом RT дикого типа (например, немодифицированным). В некоторых вариантах осуществления домен RT содержит домен RT HIV-1. В вариантах осуществления домен RT HIV-1 инициирует более низкие уровни синтеза даже при несовпадении трех нуклеотидов по сравнению с альтернативным доменом RT (например, как описано в Jamburuthugoda и Eickbush J Mol Biol 407(5):661-672 (2011); включенном в данный документ посредством ссылки во всей своей полноте).

Сайт-мишень

В некоторых вариантах осуществления после Gene Writing сайт-мишень, окружающий интегрированную последовательность, содержит ограниченное количество вставок или делеций, например, в менее чем приблизительно 50% или 10% событий интеграции, например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления в сайте-мишени не обнаруживается несколько событий вставки, например, дупликаций по типу "голова к хвосту" или "голова к голове", например, как определено с помощью секвенирования ампликонов с длинными ридами сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903

(2020) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице. В некоторых вариантах осуществления сайт-мишень не содержит вставок, полученных из эндогенной РНК, в более чем приблизительно 1% или 10% событий, например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице.

В некоторых вариантах осуществления сайт-мишень содержит интегрированную последовательность, соответствующую РНК-матрице. В вариантах осуществления сайт-мишень не содержит последовательности за пределами матрицы для RT (например, каркаса gRNA, остова вектора и/или ITR), например, как определено с помощью секвенирования ампликонов с длинными ридами для сайта-мишени, например, как описано в Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (включенном в данный документ посредством ссылки во всей своей полноте).

Варианты Gene Writer, полученные методом направленной эволюции

В некоторых вариантах осуществления в настоящем изобретении предусмотрены варианты Gene Writer, полученные методом направленной эволюции. В некоторых вариантах осуществления варианты, полученные методом направленной эволюции, могут быть получены посредством мутагенеза эталонного Gene Writer или одного из содержащихся в нем фрагментов или доменов. В некоторых вариантах осуществления метод направленной эволюции применяют в отношении одного или нескольких доменов (например, домена обратной транскриптазы, ДНК-связывающего домена (в том числе, например, направляемых последовательностью ДНК-связывающих элементов), РНК-связывающего домена или эндонуклеазного домена). В некоторых вариантах осуществления один или несколько таких вариантов доменов, полученных методом направленной эволюции, могут быть получены в результате применения метода направленной эволюции к ним отдельно или совместно с другими доменами. Вариант домена или домены, полученные методом направленной эволюции, могут в некоторых вариантах осуществления быть объединены с когнатным(когнатными) компонентом(компонентами), к которым не был применен метод направленной эволюции, или вариантами когнатного(когнатных) компонента(компонентов), полученными методом направленной эволюции, например, при получении которых методом направленной эволюции мог применяться либо параллельный, либо последовательный подход.

В некоторых вариантах осуществления процесс мутагенеза эталонного Gene Writer или его фрагмента или домена предусматривает мутагенез эталонного Gene Writer или его фрагмента или домена. В вариантах осуществления мутагенез предусматривает способ непрерывной эволюции (например, PACE) или способ прерывистой эволюции (например, PANCE), например, как описано в данном документе. В некоторых вариантах

осуществления Gene Writer, полученный методом направленной эволюции, или его фрагмент или домен содержат одну или несколько аминокислотных вариаций, введенных в их аминокислотную последовательность, по сравнению с аминокислотной последовательностью эталонного Gene Writer или его фрагмента или домена. В вариантах осуществления вариации аминокислотной последовательности могут предусматривать один или несколько мутантных остатков (например, консервативные замены, неконсервативные замены или их комбинацию) в аминокислотной последовательности эталонного Gene Writer, полученные, например, в результате изменения в последовательности нуклеотидов, кодирующей Gene Writer, которое приводит, например, к изменению кодона в любом конкретном положении в кодирующей последовательности, делеции одной или нескольких аминокислот (например, в случае усеченного белка), вставке одной или нескольких аминокислот или любой комбинации изложенного выше. Вариант Gene Writer, полученный методом направленной эволюции, может предусматривать наличие вариантов в одном или нескольких компонентах или доменах Gene Writer (например, варианты, введенные в домен обратной транскриптазы, эндонуклеазный домен, ДНК-связывающий домен, РНК-связывающий домен или их комбинации).

В некоторых аспектах в настоящем изобретении предусмотрены Gene Writer, системы, наборы и способы, в которых используется или содержится вариант Gene Writer, полученный методом направленной эволюции, например, в которых используется вариант Gene Writer, полученный методом направленной эволюции, или Gene Writer, который получен или который может быть получен с помощью PACE или PANCE. В вариантах осуществления эталонный Gene Writer, к которому не был применен метод направленной эволюции, представляет собой Gene Writer, раскрытый в данном документе.

Термин "непрерывная эволюция, осуществляемая с использованием фагов (PACE)", используемый в данном документе, обычно относится к методу непрерывной направленной эволюции, в котором используются фаги в качестве вирусных векторов. Примеры технологии PACE были описаны, например, в международной заявке согласно PCT № PCT/US 2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной заявке согласно PCT PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; патенте США № 9023594, выданном 5 мая 2015 г.; патенте США № 9771574, выданном 26 сентября 2017 г.; патенте США № 9394537, выданном 19 июля 2016 г.; международной заявке согласно PCT PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г.; патенте США № 10179911, выданном 15 января 2019 г.; и международной заявке согласно PCT PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., содержание каждого из которых включено в данный документ посредством ссылки во всей своей полноте.

Термин "прерывистая эволюция, осуществляемая с использованием фагов (PANCE)", используемый в данном документе, обычно относится к методу прерывистой направленной эволюции, в котором используются фаги в качестве вирусных векторов.

Примеры технологии PANCE были описаны, например, в Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017), включенном в данный документ посредством ссылки во всей своей полноте. Вкратце, PANCE представляет собой методику быстрой направленной эволюции in vivo с использованием серийных переносов в колбах фага для селекции (SP), подвергающегося действию направленной эволюции, который содержит ген, представляющий интерес, который необходимо подвергнуть действию направленной эволюции, в свежих клетках-хозяевах (например, клетках E. coli). Гены внутри клетки-хозяина могут поддерживаться в неизменном состоянии, в то время как гены, содержащиеся в SP - непрерывно эволюционировать. После обеспечения роста фага аликвоту инфицированных клеток можно использовать для трансфекции следующей колбы, содержащей E. coli в качестве хозяина. Этот процесс можно повторять и/или продолжать до тех пор, пока в результате направленной эволюции не будет получен необходимый фенотип, например, столько переносов, сколько необходимо.

Способы применения PACE и PANCE к Gene Writer могут быть легко оценены специалистом в данной области посредством ссылки, помимо прочего, на предшествующие литературные источники. Дополнительные иллюстративные способы управления непрерывной направленной эволюцией белков или систем, модифицирующих геном, например, в популяции клеток-хозяев, например, с использованием фаговых частиц, могут применяться для создания вариантов Gene Writer, полученных методом направленной эволюции, или их фрагментов или субдоменов. Неограничивающие примеры таких способов описаны в международной заявке согласно PCT PCT/US2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной заявке согласно PCT PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; патенте США № 9023594, выданном 5 мая 2015 г.; патенте США № 9771574, выданном 26 сентября 2017 г.; патенте США № 9394537, выданном 19 июля 2016 г.; международной заявке согласно PCT PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г.; патенте США № 10179911, выданном 15 января 2019 г.; международной заявке № PCT/US2019/37216, поданной 14 июня 2019 г., международной патентной публикации WO 2019/023680, опубликованной 31 января 2019 г., международной заявке согласно PCT PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., и международной патентной публикации № PCT/US2019/47996, поданной 23 августа 2019 г., каждый из которых включен в данный документ посредством ссылки во всей своей полноте.

В некоторых неограничивающих иллюстративных вариантах осуществления способ направленной эволюции в случае варианта Gene Writer, полученного методом направленной эволюции, его фрагмента или домена включает: (a) приведение популяции клеток-хозяев в контакт с популяцией вирусных векторов, содержащих ген, представляющий интерес (исходный Gene Writer или его фрагмент или домен), где (1)

клетка-хозяин поддается инфицированию вирусным вектором; (2) в клетке-хозяине экспрессируются вирусные гены, необходимые для образования вирусных частиц; (3) экспрессия по меньшей мере одного вирусного гена, необходимого для продуцирования инфекционной вирусной частицы, зависит от функции гена, представляющего интерес; и/или (4) вирусный вектор обеспечивает экспрессию белка в клетке-хозяине и может реплицироваться и упаковываться в вирусную частицу клеткой-хозяином; В некоторых вариантах осуществления способ включает (b) приведение клеток-хозяев в контакт с мутагеном, использование клеток-хозяев с мутациями, которые повышают скорость мутаций (например, за счет наличия мутационной плазмиды или некоторой модификации генома, например, ДНК-полимеразы с нарушенной функцией исправления ошибок, генов SOS, таких как UmuC, UmuD' и/или RecA, мутации которых, если они связаны с плазмидой, могут находиться под контролем индуцируемого промотора), или их комбинации. В некоторых вариантах осуществления способ включает (c) инкубирование популяции клеток-хозяев в условиях, обеспечивающих репликацию вируса и продуцирование вирусных частиц, где клетки-хозяева удаляют из популяции клеток-хозяев, а свежие, неинфицированные клетки-хозяева вводят в популяцию клеток-хозяев, за счет чего пополняется популяция клеток-хозяев и создается поток клеток-хозяев. В некоторых вариантах осуществления клетки инкубируют в условиях, позволяющих гену, представляющему интерес, приобрести мутацию. В некоторых вариантах осуществления способ дополнительно включает (d) выделение мутантного варианта вирусного вектора, кодирующего продукт гена, полученный методом направленной эволюции (например, вариант Gene Writer, полученный методом направленной эволюции, или его фрагмент или домен), из популяции клеток-хозяев.

Специалисту в данной области будет известно множество признаков, которые можно использовать в рамках вышеописанной структуры. Например, в некоторых вариантах осуществления вирусный вектор или фаг представляет собой нитевидный фаг, например, фаг M13, например, фаг M13 для селекции. В некоторых вариантах осуществления ген, необходимый для продуцирования инфекционных вирусных частиц, представляет собой ген III M13 (gIII). В вариантах осуществления фаг может не содержать функционального gIII, но при этом содержать gI, gII, gIV, gV, gVI, gVII, gVIII, gIX и gX. В некоторых вариантах осуществления образование инфекционных частиц VSV предусматривает наличие белка оболочки VSV-G. В различных вариантах осуществления могут использоваться разные ретровирусные векторы, например, векторы на основе вируса лейкоза мышей или лентивирусные векторы. В вариантах осуществления ретровирусные векторы могут быть эффективно упакованы с использованием белка оболочки VSV-G, например, в качестве замены нативного белка оболочки вируса.

В некоторых вариантах осуществления клетки-хозяева инкубируют в соответствии с подходящим количеством жизненных циклов вируса, например, по меньшей мере 10, по меньшей мере 20, по меньшей мере 30, по меньшей мере 40, по меньшей мере 50, по меньшей мере 100, по меньшей мере 200, по меньшей мере 300, по меньшей мере 400, по

меньшей мере 500, по меньшей мере 600, по меньшей мере 700, по меньшей мере 800, по меньшей мере 900, по меньшей мере 1000, по меньшей мере 1250, по меньшей мере 1500, по меньшей мере 1750, по меньшей мере 2000, по меньшей мере 2500, по меньшей мере 3000, по меньшей мере 4000, по меньшей мере 5000, по меньшей мере 7500, по меньшей мере 10000 или больше последовательных жизненных циклов вируса, который в иллюстративных и не ограничивающих примерах с фагом M13 составляет 10-20 минут на жизненный цикл вируса. Аналогичным образом, условия можно модулировать для регулирования времени, в течение которого клетка-хозяин остается в популяции клеток-хозяев, например, приблизительно 10, приблизительно 11, приблизительно 12, приблизительно 13, приблизительно 14, приблизительно 15, приблизительно 16, приблизительно 17, приблизительно 18, приблизительно 19, приблизительно 20, приблизительно 21, приблизительно 22, приблизительно 23, приблизительно 24, приблизительно 25, приблизительно 30, приблизительно 35, приблизительно 40, приблизительно 45, приблизительно 50, приблизительно 55, приблизительно 60, приблизительно 70, приблизительно 80, приблизительно 90, приблизительно 100, приблизительно 120, приблизительно 150 или приблизительно 180 минут. Популяции клеток-хозяев можно частично регулировать по плотности клеток-хозяев или в некоторых вариантах осуществления по плотности клеток-хозяев в поступающей среде, например, 10^3 клеток/мл, приблизительно 10^4 клеток/мл, приблизительно 10^5 клеток/мл, приблизительно 5-10^5 клеток/мл, приблизительно 10^6 клеток/мл, приблизительно 5-10^6 клеток/мл, приблизительно 10^7 клеток/мл, приблизительно 5-10^7 клеток/мл, приблизительно 10^8 клеток/мл, приблизительно 5-10^8 клеток/мл, приблизительно 10^9 клеток/мл, приблизительно 5-10^9 клеток/мл, приблизительно 10^{10} клеток/мл или приблизительно 5-10^{10} клеток/мл.

Промоторы

В некоторых вариантах осуществления один или несколько промоторных или энхансерных элементов функционально связаны с нуклеиновой кислотой, кодирующей белок Gene Writer, или нуклеиновой кислотой-матрицей, например, которая контролирует экспрессию последовательности гетерологичного объекта. В некоторых вариантах осуществления один или несколько промоторных или энхансерных элементов содержат специфичные для типа клеток или тканеспецифические элементы. В некоторых вариантах осуществления промотор или энхансер является таким же, как промотор или энхансер, который в естественных условиях контролирует экспрессию последовательности гетерологичного объекта, или получен из него. Например, промотор и энхансер гена орнитинтранскарбамилазы можно применять для контроля экспрессии гена орнитинтранскарбамилазы в системе или способе, предусмотренных в настоящем изобретении, для коррекции дефицита орнитинтранскарбамилазы. В некоторых вариантах осуществления промотор для применения в настоящем изобретении предназначен для гена, описанного в любой из таблиц **9-22**, например, который можно применять с аллелем эталонного гена или, в других вариантах осуществления, с гетерологичным геном. В

некоторых вариантах осуществления промотор представляет собой промотор из таблицы 33 или его функциональный фрагмент или вариант.

Иллюстративные тканеспецифические промоторы, которые являются коммерчески доступными, можно найти, например, по унифицированному указателю ресурса (например, www.invivogen.com/tissue-specific-promoters). В некоторых вариантах осуществления промотор представляет собой **нативный промотор** или минимальный промотор, например, состоящий из одного фрагмента из 5'-области данного гена. В некоторых вариантах осуществления нативный промотор содержит коровый промотор и его природную 5'-UTR. В некоторых вариантах осуществления 5'-UTR содержит интрон. В других вариантах осуществления они включают **составные промоторы, в которых** объединены промоторные элементы различного происхождения или которые были получены посредством сборки дистального энхансера с минимальным промотором того же происхождения.

Иллюстративные клеточно- или тканеспецифические промоторы представлены в таблицах ниже, и иллюстративные последовательности нуклеиновых кислот, кодирующие их, известны из уровня техники и могут быть легко доступны с использованием различных ресурсов, таких как база данных NCBI, в том числе RefSeq, а также Eukaryotic Promoter Database (//epd.epfl.ch//index.php).

Таблица 33. Иллюстративные клеточно- или тканеспецифические промоторы

Промотор	**Клетки-мишени**
Промотор B29	B-клетки
Промотор CD14	Моноцитарные клетки
Промотор CD43	Лейкоциты и тромбоциты
Промотор CD45	Гемопоэтические клетки
Промотор CD68	Макрофаги
Промотор гена десмина	Мышечные клетки
Промотор гена эластазы-1	Ацинарные клетки поджелудочной железы
Промотор гена эндоглина	Эндотелиальные клетки
Промотор гена фибронектина	Дифференцирующиеся клетки, рубцовая ткань
Промотор Flt-1	Эндотелиальные клетки
Промотор GFAP	Астроциты
Промотор GPIIB	Мегакариоциты
Промотор ICAM-2	Эндотелиальные клетки
Промотор INF-бета	Гемопоэтические клетки
Промотор Mb	Мышечные клетки
Промотор Nphs1	Подоциты
Промотор OG-2	Остеобласты, одонтобласты

Промотор SP-B	Легкое
Промотор Syn1	Нейроны
Промотор WASP	Гемопоэтические клетки
Промотор SV40/bAlb	Печень
Промотор SV40/bAlb	Печень
Промотор SV40/Cd3	Лейкоциты и тромбоциты
Промотор SV40/CD45	Гемопоэтические клетки
Промотор NSE/RU5'	Зрелые нейроны

Таблица 34. Дополнительные иллюстративные клеточно- или тканеспецифические промоторы

Промотор	**Описание гена**	**Специфичность гена**
APOA2	Аполипопротеин A-II	Гепатоциты (из предшественников гепатоцитов)
SERPINA1 (hAAT)	Ингибитор пептидаз SERPIN, клада A (альфа-1 антипротеиназа, антитрипсин), представитель 1 (также называемый альфа-1-антитрипсином)	Гепатоциты (из стадии дефинитивной энтодермы)
CYP3A	Цитохром P450, семейство 3, подсемейство A, полипептид	Зрелые гепатоциты
MIR122	МикроРНК 122	Гепатоциты (из клеток эмбриональной печени на ранней стадии) и энтодерма
Промоторы, специфичные для поджелудочной железы		

Промотор	**Описание гена**	**Специфичность гена**
INS	Инсулин	Бета-клетки поджелудочной железы (из стадии дефинитивной энтодермы)
IRS2	Субстрат 2 рецептора инсулина	Бета-клетки поджелудочной железы
Pdx1	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 (aMHC)	Миозин, тяжелая цепь 6, сердечная мышца, альфа	Маркер поздней дифференцировки клеток сердечной мышцы (предсердная специфичность)
MYL2 (MLC-2v)	Миозин, легкая цепь 2, регуляторный, сердечный, медленный	Маркер поздней дифференцировки клеток сердечной мышцы (желудочковая специфичность)

ITNNI3 (cTnl)	Тропонин I 3 типа (сердечный)	Кардиомиоциты (из незрелого состояния)
ITNNI3 (cTnl)	Тропонин I 3 типа (сердечный)	Кардиомиоциты (из незрелого состояния)
NPPA (ANF)	Предшественник натрийуретического пептида A (также называемый предсердным натрийуретическим фактором)	Предсердная специфичность во взрослых клетках
Slc8a1 (Ncx1)	Семейство 8 переносчиков растворенных веществ (натрий/кальциевый обменник), представитель 1	Кардиомиоциты из ранних стадий развития
Промоторы, специфичные для ЦНС		
Промотор	**Описание гена**	**Специфичность гена**
SYN1 (hSyn)	Синапсин I	Нейроны
GFAP	Глиальный фибриллярный кислый белок	Астроциты
INA	Белок промежуточных филаментов нейронов альфа-интернексин (a-интернексин)	Нейропредшественники
NES	Нестин	Нейропредшественники и эктодерма
MOBP	Основной белок олигодендроцитов, ассоциированный с миелином	Олигодендроциты
MBP	Основной белок миелина	Олигодендроциты
THT	Тирозингидроксилаза	Дофаминергические нейроны

FOXA2 (HNF3-бета)	Бокс Forkhead A2	Дофаминергические нейроны (также используется в качестве маркера энтодермы)
Промоторы, специфичные для кожи		
Промотор	**Описание гена**	**Специфичность гена**
FLG	Филаггрин	Кератиноциты из зернистого слоя
K14	Кератин 14	Кератиноциты из зернистого и базального слоев
TGM3	Трансглютаминаза 3	Кератиноциты из зернистого слоя
Промоторы, специфичные для иммунных клеток		
Промотор	**Описание гена**	**Специфичность гена**
ITGAM (CD11B)	Интегрин, альфа-M (субъединица рецептора 3 компонента 3 системы комплемента)	Моноциты, макрофаги, гранулоциты, естественные клетки-киллеры
Промоторы, специфичные для клеток мочеполовой системы		
Промотор	**Описание гена**	**Специфичность гена**
Pbsn	Пробасин	Эпителий предстательной железы
Upk2	Уроплакин 2	Мочевой пузырь
Sbp	Спермин-связывающий белок	Предстательная железа
Fer1L4	Fer-1-подобный белок 4	Мочевой пузырь
Промоторы, специфичные для эндотелиальных клеток		
Промотор	**Описание гена**	**Специфичность гена**
ENG	Эндоглин	Эндотелиальные клетки

Промоторы, специфичные для плюрипотентных и эмбриональных клеток		
Промотор	**Описание гена**	**Специфичность гена**
Oct4 (POU5F1)	Гомеобокс 1 POU класса 5	Плюрипотентные клетки (зародышевые клетки, ES-клетки, iPS-клетки)
NANOG	Гомеобокс Nanog	Плюрипотентные клетки (ES-клетки, iPS-клетки)
Синтетический Oct4	Синтетический промотор на основе корового энхансерного элемента Oct-4	Плюрипотентные клетки (ES-клетки, iPS-клетки)
Brachyury T	Brachyury	Мезодерма
NES	Нестин	Нейропредшественники и эктодерма
SOX17	SRY (область Y, определяющая пол)-бокс 17	Энтодерма
FOXA2 (HNFJ бета)	Бокс Forkhead A2	Энтодерма (также используется в качестве маркера дофаминергических нейронов)
MIR122	МикроРНК 122	Энтодерма и гепатоциты (из клеток эмбриональной печени на ранней стадии)

В зависимости от используемой системы хозяин/вектор в векторе экспрессии можно использовать любой из ряда пригодных элементов, контролирующих транскрипцию и трансляцию, включая конститутивные и индуцируемые промоторы, энхансерные элементы транскрипции, терминаторы транскрипции и т. д. (см., например, Bitter et al. (1987) *Methods in Enzymology*, 153:516-544; включенный в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления нуклеиновая кислота, кодирующая Gene Writer или нуклеиновую кислоту-матрицу, функционально связана с контролирующим

элементом, например, элементом, контролирующим транскрипцию, таким как промотор. В некоторых вариантах осуществления элемент, контролирующий транскрипцию, может быть функциональным в эукариотической клетке, например, в клетке млекопитающего; либо в прокариотической клетке (например, бактериальной клетке или клетке археи). В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая полипептид, функционально связана с несколькими контролирующими элементами, например, которые обеспечивают экспрессию нуклеотидной последовательности, кодирующей полипептид, как в прокариотических, так и в эукариотических клетках.

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 (CamKIIα) (см., например, Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250 и Casanova et al. (2001) Genesis 31:37); энхансер CMV/промотор гена тромбоцитарного фактора роста-β (см., например, Liu et al. (2004) Gene Therapy 11:52-60) и т. п.

Специфичные для адипоцитов пространственно ограниченные промоторы включают без ограничения промотор/энхансер гена aP2, например, область от -5,4 т. о. до +21 п. о. гена aP2 человека (см., например, Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590 и Pavjani et al. (2005) Nat. Med. 11:797); промотор гена транспортера глюкозы 4 типа (GLUT4) (см., например, Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); промотор гена транслоказы жирных кислот (FAT/CD36) (см., например, Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476 и Sato et al. (2002) J. Biol. Chem. 277:15703); промотор гена стеароил-CoA-десатуразы-1 (SCD1) (Tabor et al. (1999) J. Biol. Chem. 274:20603); промотор гена лептина (см., например, Mason et al. (1998) Endocrinol. 139:1013 и Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); промотор

гена адипонектина (см., например, Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484 и Chakrabarti (2010) Endocrinol. 151:2408); промотор гена адипсина (см., например, Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); промотор гена резистина (см., например, Seo et al. (2003) Molec. Endocrinol. 17:1522) и т. п.

Специфичные для кардиомиоцитов пространственно ограниченные промоторы включают без ограничения контролирующие последовательности, полученные из следующих генов: легкой цепи-2 миозина, тяжелой цепи α-миозина, AE3, сердечного тропонина C, сердечного актина и т. п. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617 и Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Специфичные для гладких мышц пространственно ограниченные промоторы включают без ограничения промотор SM22α (см., например, Akyürek et al. (2000) Mol. Med. 6:983 и патент США № 7169874); промотор гена смутелина (см., например, WO 2001/018048); промотор гена α-гладкомышечного актина и т. п. Например, было показано, что область промотора SM22α длиной 0,4 т. о., внутри которой находятся два элемента CArG, опосредует экспрессию, специфичную для гладкомышечных клеток сосудов (см., например, Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; и Moessler, et al. (1996) Development 122, 2415-2425).

Специфичные для фоторецепторов пространственно ограниченные промоторы включают без ограничения промотор гена родопсина; промотор гена родопсинкиназы (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); промотор гена бета-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 фосфоглицераткиназы (PKG), CAG (составную молекулу из энхансера CMV, промотора гена бета-актина курицы (CBA) и интрона бета-глобина кролика), NSE (нейронспецифической енолазы), промоторы генов синапсина или NeuN, ранний промотор SV40, промотор LTR вируса опухоли молочной железы мышей; главный поздний промотор аденовируса (MLP Ad); промотор вируса простого герпеса (HSV), промотор цитомегаловируса (CMV), такой как немедленно-ранняя промоторная область CMV (CMVIE), промотор SFFV, промотор вируса саркомы Рауса (RSV), синтетические промоторы, гибридные промоторы и т. п. Другие промоторы могут происходить от человека или от других видов, в том числе от мышей. Обычные промоторы включают, например, промотор гена немедленно-раннего ответа цитомегаловируса человека (CMV), ранний промотор SV40, длинный концевой повтор вируса саркомы Рауса, промотор гена бета-актина, промотор гена инсулина крыс, промотор гена фосфоглицераткиназы, промотор гена альфа-1-антитрипсина человека (hAAT), промотор гена транстиретина, промотор TBG и другие промоторы, специфичные для печени, промотор гена десмина и аналогичные промоторы, специфичные для мышц, промотор EF1-альфа, промотор CAG и другие конститутивные промоторы, гибридные промоторы со специфичностью для нескольких тканей, промоторы, специфичные для нейронов, такие как промотор гена синапсина и глицеральдегид-3-фосфатдегидрогеназы, все из которых являются промоторами, хорошо известными и легко доступными для специалистов в данной области, и могут использоваться для достижения экспрессии кодирующей последовательности, представляющей интерес, на высоком уровне. Кроме того, в данном документе также найдут применение последовательности, полученные из невирусных генов, таких как ген металлотионеина мыши. Такие промоторные последовательности коммерчески доступны, например, от Stratagene (Сан-Диего, Калифорния). Дополнительные иллюстративные последовательности промоторов описаны, например, в WO 2018213786A1 (включенной в данный документ посредством ссылки во всей своей полноте).

В некоторых вариантах осуществления используется энхансер гена аполипопротеина E (ApoE) или его функциональный фрагмент, например, для управления экспрессией в печени. В некоторых вариантах осуществления используются две копии энхансера ApoE или его функционального фрагмента. В некоторых вариантах осуществления энхансер ApoE или его функциональный фрагмент используют в комбинации с промотором, например, промотором гена альфа-1-антитрипсина человека (hAAT).

В некоторых вариантах осуществления регуляторные последовательности придают возможности экспрессии тканеспецифических генов. В некоторых случаях тканеспецифические регуляторные последовательности связывают тканеспецифические факторы транскрипции, которые индуцируют транскрипцию тканеспецифическим

образом. Различные тканеспецифические регуляторные последовательности (например, промоторы, энхансеры и т. д.) известны из уровня техники. Иллюстративные тканеспецифические регуляторные последовательности включают без ограничения следующие тканеспецифические промоторы: специфичный для печени промотор гена тироксинсвязывающего глобулина (TBG), промотор гена инсулина, промотор гена глюкагона, промотор гена соматостатина, промотор гена панкреатического полипептида (PPY), промотор гена синапсина-1 (Syn), промотор гена креатинкиназы (MCK), промотор гена десмина млекопитающих (DES), промотор гена тяжелой цепи α-миозина (a-MHC) или промотор гена сердечного тропонина T (cTnT). Другие примеры промоторов включают промотор гена бета-актина, коровый промотор вируса гепатита B, Sandig et al., Gene Ther., 3:1002-9 (1996); промотор гена альфа-фетопротеина (AFP), Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), промотор гена костного остеокальцина (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); промотор гена костного сиалопротеина (Chen et al., J. Bone Miner. промотор гена тяжелой цепи иммуноглобулина; промотор гена α-цепи T-клеточного рецептора, нейрональные промоторы, такие как промотор нейронспецифической енолазы (NSE) (Andersen et al. Cell. Mol. Neurobiol., 13:503-15 (1993)); промотор гена легкой цепи нейрофиламента (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)) и промотор нейронспецифического гена vgf (Piccioli et al., Neuron, 15:373-84 (1995)) и другие. Дополнительные иллюстративные промоторные последовательности описаны, например, в патенте США № 10300146 (включенном в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления тканеспецифический регуляторный элемент, например, тканеспецифический промотор, выбран из элемента, о котором известно, что он функционально связан с геном, который экспрессируется на высоком уровне в данной ткани, например, как измерено с помощью секвенирования РНК или данных об экспрессии белка или их комбинации. Способы анализа тканевой специфичности по экспрессии описаны в Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014), который включен в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления вектор, описанный в данном документе, представляет собой полицистронную экспрессионную конструкцию. Полицистронные экспрессионные конструкции включают, например, конструкции, содержащие первую кассету экспрессии, например, содержащую первый промотор и первую кодирующую последовательность нуклеиновой кислоты, и вторую кассету экспрессии, например, содержащую второй промотор и вторую кодирующую последовательность нуклеиновой кислоты. Такие полицистронные экспрессионные конструкции могут в некоторых случаях быть особенно полезными для доставки нетранслируемых продуктов генов, таких как шпилечные РНК, вместе с полипептидом, например, Gene Writer и матрицы Gene Writer. В некоторых вариантах осуществления полицистронные экспрессионные конструкции могут демонстрировать сниженные уровни экспрессии одного или нескольких включенных трансгенов, например, из-за интерференции промоторов или присутствия несовместимых элементов, представляющих собой нуклеиновую кислоту, в непосредственной близости.

Если полицистронная экспрессионная конструкция является частью вирусного вектора, присутствие самокомплементарной последовательности нуклеиновой кислоты может в некоторых случаях препятствовать образованию структур, необходимых для репродукции или упаковки вируса.

В некоторых вариантах осуществления последовательность кодирует РНК со шпилькой. В некоторых вариантах осуществления шпилечная РНК представляет собой направляющую РНК, РНК-матрицу, shRNA или микроРНК. В некоторых вариантах осуществления первый промотор представляет собой промотор РНК-полимеразы I. В некоторых вариантах осуществления первый промотор представляет собой промотор РНК-полимеразы II. В некоторых вариантах осуществления второй промотор представляет собой промотор РНК-полимеразы III. В некоторых вариантах осуществления второй промотор представляет собой промотор U6 или H1. В некоторых вариантах осуществления конструкция на основе нуклеиновой кислоты содержит структуру конструкции B1 или B2 AAV.

Не желая ограничиваться какой-либо теорией, полагают, что полицистронные экспрессионные конструкции могут не достигать оптимальных уровней экспрессии по сравнению с системами экспрессии, содержащими только один цистрон. Одной из предполагаемых причин более низких уровней экспрессии, достигаемых при использовании полицистронных экспрессионных конструкций, содержащих два или более промоторных элемента, является явление интерференции промоторов (см., например, Curtin J A, Dane A P, Swanson A, Alexander I E, Ginn S L. *Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct*. Gene Ther. 2008 March; 15(5):384-90 и Martin-Duque P, Jezzard S, Kaftansis L, Vassaux G. *Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes*. Hum Gene Ther. 2004 October; 15(10):995-1002; оба литературных источника включены в данный документ посредством ссылки для раскрытия явления интерференции промоторов). В некоторых вариантах осуществления проблему интерференции промоторов можно решить, например, путем получения полицистронных экспрессионных конструкций, содержащих только один промотор, управляющий транскрипцией нескольких кодирующих последовательностей нуклеиновых кислот, разделенных участками внутренней посадки рибосомы, или путем разделения цистронов, содержащих их собственный промотор, транскрипционными инсуляторными элементами. В некоторых вариантах осуществления экспрессия нескольких цистронов, управляемая одним промотором, может приводить к неравномерным уровням экспрессии цистронов. В некоторых вариантах осуществления промотор не может быть эффективно выделен, и элементы для выделения могут быть несовместимы с некоторыми векторами для переноса генов, например, некоторыми ретровирусными векторами.

МикроРНК

miRNA и другие малые интерферирующие нуклеиновые кислоты обычно регулируют экспрессию генов посредством расщепления/разрушения РНК-транскрипта-

мишени или репрессии трансляции матричной РНК-мишени (мРНК). miRNA могут в некоторых случаях экспрессироваться нативно, обычно в виде конечных нетранслируемых РНК-продуктов длиной 19-25. miRNA обычно проявляют свою активность посредством специфичных в отношении последовательности взаимодействий с 3′-нетранслируемыми областями (UTR) мРНК-мишеней. Эти эндогенно экспрессируемые miRNA могут образовывать шпилечные предшественники, которые впоследствии процессируются в miRNA-дуплекс, а затем в зрелую однонитевую молекулу miRNA. Эта зрелая miRNA обычно направляет многобелковый комплекс miRISC, который идентифицирует 3′-UTR-области-мишени мРНК-мишеней на основе их комплементарности по отношению к зрелой miRNA. Применимые трансгенные продукты могут включать, например, miRNA или сайты связывания miRNA, которые регулируют экспрессию связанного полипептида. Неограничивающий перечень генов miRNA; продукты этих генов и их гомологов являются применимыми в качестве трансгенов или в качестве мишеней для малых интерферирующих нуклеиновых кислот (например, miRNA-губок, антисмысловых олигонуклеотидов), например, в способах, таких как способы, приведенные в US 10300146, 22:25-25:48, включенном посредством ссылки. В некоторых вариантах осуществления один или несколько сайтов связывания для одной или нескольких вышеуказанных miRNA включены в трансген, например, трансген, доставляемый вектором на основе rAAV, например, для ингибирования экспрессии трансгена в одной или нескольких тканях животного, несущего трансген. В некоторых вариантах осуществления сайт связывания может быть выбран таким образом, чтобы обеспечивать контроль экспрессии трансгена тканеспецифическим образом. Например, сайты связывания для специфичной в отношении печени miR-122 могут быть включены в трансген для ингибирования экспрессии этого трансгена в печени. Дополнительные иллюстративные последовательности miRNA описаны, например, в патенте США № 10300146 (включенном в данный документ посредством ссылки во всей своей полноте).

Однако для специфичного в отношении печени Gene Writing можно использовать сверхэкспрессию miR-122 вместо использования сайтов связывания для осуществления miR-122-специфичного разрушения. Эта miRNA положительно ассоциирована с дифференцировкой и созреванием клеток печени, а также с повышенной экспрессией генов, специфичных для печени. Таким образом, в некоторых вариантах осуществления кодирующая последовательность miR-122 может быть добавлена к компоненту системы Gene Writing для усиления терапии, направленной на печень.

Ингибитор miR или ингибитор miRNA обычно представляет собой средство, которое блокирует экспрессию и/или процессинг miRNA. Примеры таких средств включают без ограничения антагонисты микроРНК, специфичные в отношении микроРНК антисмысловые последовательности, микроРНК-губки и специфичные в отношении микроРНК олигонуклеотиды (двухнитевые, шпилечные, короткие олигонуклеотиды), которые ингибируют взаимодействие miRNA с комплексом Drosha. Ингибиторы микроРНК, например, miRNA-губки, могут экспрессироваться в клетках из трансгенов

(например, как описано в работе Ebert, M. S. Nature Methods, Epub Aug. 12, 2007; включенной в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления микроРНК-губки или другие ингибиторы miR используются с AAV. МикроРНК-губки обычно специфично ингибируют miRNA посредством комплементарной гептамерной затравочной последовательности. В некоторых вариантах осуществления все семейство miRNA может быть подвергнуто сайленсингу с помощью одной последовательности губки. Другие способы сайленсинга функции miRNA (подавление мишеней miRNA) в клетках будут очевидны специалисту средней квалификации в данной области.

В некоторых вариантах осуществления miRNA, описанная в данном документе, содержит последовательность, приведенную в таблице 4 публикации согласно PCT № WO2020014209, включенной в данный документ посредством ссылки. Также в данный документ посредством ссылки включен перечень иллюстративных последовательностей miRNA из WO2020014209.

В некоторых вариантах осуществления преимущественно подвергать сайленсингу один или несколько компонентов системы Gene Writing (например, мРНК, кодирующую полипептид Gene Writer, РНК-матрицу Gene Writer или последовательность гетерологичного объекта, экспрессируемую из генома после успешного осуществления способа Gene Writing) в части клеток. В некоторых вариантах осуществления преимущественно ограничивать экспрессию компонента системы Gene Writing выбранными типами клеток в ткани, представляющей интерес.

Например, известно, что в заданной ткани, например, в печени, макрофаги и иммунные клетки, например, клетки Купфера в печени, могут участвовать в поглощении средства доставки одного или нескольких компонентов системы Gene Writing. В некоторых вариантах осуществления по меньшей мере один сайт связывания для по меньшей мере одной miRNA, экспрессируемой на высоком уровне в макрофагах и иммунных клетках, например, клетках Купфера, включен в по меньшей мере один компонент системы Gene Writing, например, нуклеиновую кислоту, кодирующую полипептид Gene Writing, или трансген. В некоторых вариантах осуществления miRNA, которая нацелена на один или несколько сайтов связывания, приведена в таблице, упоминаемой данном документе, например, miR-142, например, зрелая miRNA hsa-miR-142-5p или hsa-miR-142-3p.

В некоторых вариантах осуществления может быть предпочтительно снижение уровней Gene Writer и/или активности Gene Writer в клетках, в которых экспрессия Gene Writer или сверхэкспрессия трансгена может оказывать токсический эффект. Например, было показано, что доставка кассеты сверхэкспрессии трансгена в ганглиозные нейроны дорсального корешка может привести в результате к токсичности генной терапии (см. Hordeaux et al. Sci Transl Med 12(569):eaba9188 (2020), включенный в данный документ посредством ссылки во всей своей полноте). В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA может быть включен в компонент системы Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии

данного компонента системы в нейроне, например, в ганглиозном нейроне дорсального корешка. В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA, включенный в компонент системы Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии данного компонента системы в нейроне представляет собой сайт связывания miR-182, например, зрелой miRNA hsa-miR-182-5p или hsa-miR-182-3p. В некоторых вариантах осуществления по меньшей мере один сайт связывания miRNA, включенный в компонент системы Gene Writing, представляющий собой нуклеиновую кислоту, для снижения экспрессии данного компонента системы в нейроне представляет собой сайт связывания miR-183, например, зрелой miRNA hsa-miR-183-5p или hsa-miR-183-3p. В некоторых вариантах осуществления для усиления ограничения экспрессии одного или нескольких компонентов системы Gene Writing тканью или типом клеток, представляющими интерес, можно использовать комбинации сайтов связывания miRNA.

В таблице A5 ниже приведены иллюстративные miRNA и соответствующие экспрессирующие клетки, например, miRNA, для которых в некоторых вариантах осуществления можно включать сайты связывания (комплементарные последовательности) в трансген или нуклеиновую кислоту, кодирующие полипептид, например, для снижения экспрессии в этой нецелевой клетке.

Таблица A5. Иллюстративные miRNA из нецелевых клеток и тканей

Тип клеток, подвергнутых сайленсингу	**Название miRNA**	**Зрелая miRNA**	**Последовательность miRNA**	**SEQ ID NO**
Клетки Купфера	miR-142	hsa-miR-142-5p	cauaaaguagaaag cacuacu	481
Клетки Купфера	miR-142	hsa-miR-142-3p	uguaguguuuccua cuuuaugga	482
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-5p	uuuggcaaugguag aacucacacu	483
Ганглиозные нейроны дорсального корешка	miR-182	hsa-miR-182-3p	ugguucuagacuug ccaacua	484
Ганглиозные нейроны дорсального корешка	miR-183	hsa-miR-183-5p	uauggcacugguag aauucacu	485
Ганглиозные нейроны дорсального корешка	miR-183	hsa-miR-183-3p	gugaauuaccgaag ggccauaa	486
Гепатоциты	miR-122	hsa-miR-122-5p	uggagugugacaau gguguuug	
Гепатоциты	miR-122	hsa-miR-122-3p	aacgccauuaucaca cuaaaua	

Системы, противодействующие CRISPR, для регуляции активности GeneWriter

Различные подходы для модулирования активности молекулы Cas могут быть применены в сочетании с системами и способами, описанными в данном документе. Например, в некоторых вариантах осуществления полипептид, описанный в данном документе (например, молекулу Cas или Gene Writer, содержащий домен Cas), можно регулировать с помощью средства, противодействующего CRISPR (например, белка, противодействующего CRISPR, или малой молекулы, противодействующей CRISPR). В некоторых вариантах осуществления молекула Cas или домен Cas содержит отвечающий интеин, такой как, например, 4-гидрокситамоксифен (4-HT)-отвечающий интеин, молекулу iCas (например, iCas9); 4-HT-отвечающий Cas (например, аллостерически регулируемый Cas9 (arC9) или мертвый Cas9 (dC9)). В системах и способах, описанных в данном документе, могут также использовать химически индуцированную систему димеризации расщепленных белковых фрагментов (например, опосредованную рапамицином димеризацию FK506-связывающего белка 12 (FKBP) и FKBP-рапамицин-связывающего домена (FRB), доменов гетеродимеризации ABI-PYL1, индуцируемых абсцизовой кислотой, и доменов гетеродимеризации GID1-GAI, индуцируемых гиббереллином); димер пептида BCL-xL и пептидов BH3, малую молекулу A385358 (A3), систему дегрона (например, дестабилизированную систему FKBP-Cas9, систему дегрона, индуцируемую ауксином (AID), или систему дегрона DHFR E. coli), аптамер или аптазим, слитый с gRNA (например, тетрациклин- и теофиллин-отвечающие биопереключатели), белки AcrIIA2 и AcrIIA4, а также BRD0539.

В некоторых вариантах осуществления интеин, отвечающий на малые молекулы (например, 4-гидрокситамоксифен (4-HT)-отвечающий интеин), встроен в определенные сайты молекулы Cas (например, Cas9). В некоторых вариантах осуществления встраивание 4HT-отвечающего интеина нарушает ферментативную активность Cas9. В некоторых вариантах осуществления молекула Cas (например, iCas9) слита с гормонсвязывающим доменом эстрогенового рецептора (ERT2). В некоторых вариантах осуществления лигандсвязывающий домен эстрогенового рецептора-α человека может быть вставлен в молекулу Cas (например, Cas9 или мертвого Cas9 (dC9)), например, в положении 231, с получением 4HT-отвечающего средства, противодействующего CRISPR-Cas9 (например, arC9 или dC9). В некоторых вариантах осуществления dCas9 может обеспечивать репрессию функции Cas9, зависимую от дозы 4-HT. В некоторых вариантах осуществления arC9 может обеспечивать контроль функции Cas9, зависимый от дозы 4-HT. В некоторых вариантах осуществления молекула Cas (например, Cas9) слита с расщепленными белковыми фрагментами. В некоторых вариантах осуществления химически индуцированная димеризация расщепленных белковых фрагментов (например, опосредованная рапамицином димеризация FK506-связывающего белка 12 (FKBP) и FKBP-рапамицин-связывающего домена (FRB)) может индуцировать активность молекулы Cas9 на низких уровнях. В некоторых вариантах осуществления химически индуцированная система димеризации (например, домены гетеродимеризации ABI-PYL1, индуцируемые

абсцизовой кислотой, и домены гетеродимеризации GID1-GAI, индуцируемые гиббереллином) может индуцировать дозозависимую и обратимую активацию/репрессию транскрипции Cas9. В некоторых вариантах осуществления индуцируемая система Cas9 (ciCas9) включает замену домена REC2 молекулы Cas (например, Cas9) пептидом BCL-xl и присоединение пептида BH3 к N- и C-концам модифицированного Cas9.BCL. В некоторых вариантах осуществления взаимодействие между пептидами BCL-xL и BH3 может удерживать Cas9 в неактивном состоянии. В некоторых вариантах осуществления малая молекула (например, A-385358 (A3)) может нарушать взаимодействие между пептидами BCL-xl и BH3 для активации Cas9. В некоторых вариантах осуществления индуцируемая система Cas9 может демонстрировать дозозависимый контроль нуклеазной активности. В некоторых вариантах осуществления система дегрона может индуцировать разрушение молекулы Cas (например, Cas9) при активации или деактивации внешним фактором (например, низкомолекулярным лигандом, светом, температурой или белком). В некоторых вариантах осуществления малая молекула BRD0539 обратимо ингибирует молекулу Cas (например, Cas9). Дополнительную информацию о белках, противодействующих CRISPR, или малых молекулах, противодействующих CRISPR, можно найти, например, в Gangopadhyay, S.A. et al. Precision control of CRISPR-Cas9 using small molecules and light, Biochemistry, 2019, Maji, B. et al. A high-throughput platform to identify small molecule inhibitors of CRISPR-Cas9, и Pawluk Anti-CRISPR: discovery, mechanism and function, Nature Reviews Microbiology volume 16, pages 12-17(2018), каждый из которых включен посредством ссылки во всей своей полноте.

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 п сравнению с геномным сайтом-мишенью. В некоторых вариантах осуществления сайт-мишень модифицирован с помощью обратной транскрипции с праймированием мишенью с использованием РНК-матрицы. В некоторых вариантах осуществления сайт-мишень имеет однонитевой разрыв. Сайт-мишень может быть включен в энхансер, промотор, нетранслируемую область, экзон, интрон, открытую рамку считывания или спейсерную последовательность.

В некоторых вариантах осуществления при инактивации экспрессия снижается на 25%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99,9% или является еще более низкой по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. В некоторых вариантах осуществления система Gene Writer, содержащая самоинактивирующийся модуль, характеризуется 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% или более высоким уровнем интеграции в сайтах-мишенях, чем в нецелевых сайтах, по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. Система Gene Writer, содержащая самоинактивирующийся модуль, характеризуется 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% или более высокой эффективностью модификации сайта-мишени по сравнению с системой Gene Writing, которая не содержит самоинактивирующийся модуль. В некоторых вариантах осуществления самоинактивирующийся модуль включен, если полипептид Gene Writer доставляется в виде ДНК, например, с помощью вирусного вектора.

Для нуклеаз были описаны самоинактивирующиеся модули. См., например, в Li et al A Self-Deleting AAV-CRISPR System for In Vivo Genome Editing, Mol Ther Methods Clin Dev. 2019 Mar 15; 12: 111-122, P. Singhal, Self-Inactivating Cas9: a method for reducing exposure while maintaining efficacy in virally delivered Cas9 applications (доступно на www.editasmedicine.com/wp-content/uploads/2019/10/aef_asgct_poster_2017_final_-_present_5-11-17_515pm1_1494537387_1494558495_1497467403.pdf), и Epstein and Schaffer Engineering a Self-Inactivating CRISPR System for AAV Vectors Targeted Genome Editing I|Volume 24, SUPPLEMENT 1, S50, May 01, 2016, *и* WO2018106693A1.

Малые молекулы

В некоторых вариантах осуществления полипептид, описанный в данный документе (например, полипептид Gene Writer), можно контролировать с помощью малой молекулы. В некоторых вариантах осуществления полипептид димеризуется посредством малой молекулы.

В некоторых вариантах осуществления полипептид можно контролировать с помощью химически индуцированной димеризации (CID) посредством малых молекул. CID обычно применяется для создания переключателей функции белка, чтобы изменить физиологические параметры клеток. Иллюстративным высокоспецифичным эффективным димеризатором является римидуцид (AP1903), который имеет две идентичные связывающие белок поверхности, расположенные по типу "хвост к хвосту", каждая из

которых обладает высокой аффинностью и специфичностью в отношении мутантной формы FKBP12: FKBP12(F36V) (FKBP12v36, F_{V36} или F_v). Присоединение одного или нескольких доменов F_V к одной или нескольким сигнальным молекулам клетки, которые в обычных условиях зависят от гомодимеризации, может привести к преобразованию данного белка в контрольный римидуцид. Гомодимеризация с римидуцидом применяется в случае с индуцируемым переключателем безопасности каспазы. В основе данного молекулярного переключателя, который контролируется отдельным лигандом димеризации, лежит гетеродимеризующаяся малая молекула - рапамицин или аналоги рапамицина ("рапалоги"). Рапамицин связывается с FKBP12 и его вариантами и может индуцировать гетеродимеризацию сигнальных доменов, слитых с FKBP12, посредством связывания как с FKBP12, так и с полипептидами, содержащими FKBP-рапамицин-связывающий (FRB) домен mTOR. В некоторых вариантах осуществления настоящей заявки предусмотрены молекулярные переключатели, которые значительно расширяют возможности применения рапамицина, рапалогов и римидуцида в качестве средств для терапевтического применения.

В некоторых вариантах осуществления технологии двойного переключателя гомодимеризатор, такой как AP1903 (римидуцид), непосредственно индуцирует димеризацию или мультимеризацию полипептидов, содержащих область мультимеризации FKBP12. В других вариантах осуществления полипептид, содержащий мультимеризованный FKBP12, мультимеризуется или агрегируется посредством связывания с гетеродимеризатором, таким как рапамицин или рапалог, который также связывается с областью мультимеризации FRB или варианта FRB в химерном полипептиде, также экспрессируемом в модифицированной клетке, таком как, например, химерный антигенный рецептор. Рапамицин представляет собой природный макролид, который с высокой аффинностью (< 1 нМ) связывается с FKBP12 и совместно с ним инициирует высокоаффинное ингибирующее взаимодействие с FKBP-рапамицин-связывающим (FRB) доменом mTOR. FRB имеет небольшой размер (89 аминокислот) и поэтому может применяться в качестве белковой "метки" или "указателя" при добавлении ко многим белкам. Коэкспрессия белка, слитого с FRB, с белком, слитым с FKBP12, делает их сближение индуцируемым рапамицином (12-16). Это может служить основой для переключателя клеточной безопасности, регулируемого перорально доступным лигандом - рапамицином или производными рапамицина (рапалогами), которые не ингибируют mTOR при низкой терапевтической дозе, но вместо этого связываются с выбранными мутантными доменами FRB, слитыми с каспазой-9. (См. Sabatini D M, et al., Cell. 1994; 78(1):35-43; Brown E J, et al., Nature. 1994; 369(6483):756-8; Chen J, et al., Proc Natl Acad Sci USA. 1995; 92(11):4947-51; и Choi J, Science. 1996; 273(5272):239-42).

В некоторых вариантах осуществления в терапевтических клетках предусмотрено два уровня контроля. В вариантах осуществления первый уровень контроля может быть настраиваемым, т. е. уровень удаления терапевтических клеток можно регулировать таким образом, чтобы он приводил к частичному удалению терапевтических клеток. В некоторых

вариантах осуществления химерный антигенный полипептид содержит сайт связывания рапамицина или аналога рапамицина. В вариантах осуществления в терапевтической клетке также присутствует "суицидальный" ген, такой как, например, ген, кодирующий полипептид каспазу. При использовании данного контролируемого первого уровня в некоторых вариантах осуществления потребность в продолжении терапии может быть уравновешена необходимостью устранения или снижения уровня возникновения отрицательных побочных эффектов. В некоторых вариантах осуществления пациенту вводят аналог рапамицина рапалог, который затем связывается как с полипептидом каспазой, так и с химерным антигенным рецептором, привлекая, таким образом, полипептид каспазу в местоположение CAR и обеспечивая агрегацию полипептида каспазы. При агрегации полипептид каспаза индуцирует апоптоз. Количество рапамицина или аналога рапамицина, вводимого пациенту, может варьироваться; если желательно удаление более низкого уровня клеток посредством апоптоза для уменьшения побочных эффектов и продолжения CAR-терапии, то пациенту можно вводить более низкий уровень рапамицина или рапамицина. В некоторых вариантах осуществления второй уровень контроля может быть предназначен для достижения максимального уровня устранения клеток. Данный второй уровень может быть основан, например, на применении римидуцида или AP1903. Если есть необходимость быстро устранить до 100% терапевтических клеток, пациенту можно вводить AP1903. Мультимерный AP1903 связывается с полипептидом каспазой, что приводит к мультимеризации полипептида каспазы и апоптозу. В определенных примерах второй уровень также можно настраивать или контролировать с помощью уровня AP1903, вводимого субъекту.

В некоторых вариантах осуществления малые молекулы могут применяться для контроля генов, как описано, например, в US10584351 в 47:53-56:47 (включенном в данный документ посредством ссылки во всей своей полноте), вместе с пригодными лигандами для контрольных признаков, например, в US10584351 в 56:48 и след., а также в U10046049 в 43:27-52:20, включенных посредством ссылки, а также в описании лигандов для таких систем контроля в 52:21 и след.

Химически модифицированные нуклеиновые кислоты и концевые признаки нуклеиновых кислот

Описанная в данном документе нуклеиновая кислота (например, нуклеиновая кислота-матрица, например, РНК-матрица; или нуклеиновая кислота (например, мРНК), кодирующая Gene Writer) может содержать немодифицированные или модифицированные азотистые основания. Встречающиеся в природе РНК синтезируются из четырех основных рибонуклеотидов: ATP, CTP, UTP и GTP, но могут содержать нуклеотиды с посттранскрипционными модификациями. Дополнительно в РНК было идентифицировано примерно сто различных модификаций нуклеозидов (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). РНК также может содержать полностью синтетические нуклеотиды, не встречающиеся в природе.

В некоторых вариантах осуществления химическая модификация представлена в PCT/US2016/032454, публикации патента США № 20090286852, международной заявке № WO/2012/019168, WO/2012/045075, WO/2012/135805, WO/2012/158736, WO/2013/039857, WO/2013/039861, WO/2013/052523, WO/2013/090648, WO/2013/096709, WO/2013/101690, WO/2013/106496, WO/2013/130161, WO/2013/151669, WO/2013/151736, WO/2013/151672, WO/2013/151664, WO/2013/151665, WO/2013/151668, WO/2013/151671, WO/2013/151667, WO/2013/151670, WO/2013/151666, WO/2013/151663, WO/2014/028429, WO/2014/081507, WO/2014/093924, WO/2014/093574, WO/2014/113089, WO/2014/144711, WO/2014/144767, WO/2014/144039, WO/2014/152540, WO/2014/152030, WO/2014/152031, WO/2014/152027, WO/2014/152211, WO/2014/158795, WO/2014/159813, WO/2014/164253, WO/2015/006747, WO/2015/034928, WO/2015/034925, WO/2015/038892, WO/2015/048744, WO/2015/051214, WO/2015/051173, WO/2015/051169, WO/2015/058069, WO/2015/085318, WO/2015/089511, WO/2015/105926, WO/2015/164674, WO/2015/196130, WO/2015/196128, WO/2015/196118, WO/2016/011226, WO/2016/011222, WO/2016/011306, WO/2016/014846, WO/2016/022914, WO/2016/036902, WO/2016/077125 или WO/2016/077123, каждая из которых включена в данный документ посредством ссылки во всей своей полноте. Понятно, что включение химически модифицированного нуклеотида в полинуклеотид может привести к встраиванию модификации в азотистое основание, остов или и в то, и в другое в зависимости от местоположения модификации в нуклеотиде. В некоторых вариантах осуществления модификация остова представлена в EP 2813570, который включен в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления модифицированный кэп представлен в публикации патента США № 20050287539, которая включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота (например, РНК, например, мРНК) содержит один или несколько из ARCA: антиинвертированного аналога кэпа (m27.3'-OGP3G), GP3G (неметилированного аналога кэпа), m7GP3G (монометилированного аналога кэпа), m32.2.7GP3G (триметилированного аналога кэпа), m5CTP (5'-метилцитидинтрифосфата), m6ATP (N6-метиладенозин-5'-трифосфата), s2UTP (2-тиоуридинтрифосфата) и Ψ (псевдоуридинтрифосфата).

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота содержит 5'-кэп, например: 7-метилгуанозиновый кэп (например, кэп O-Me-m7G); гиперметилированный аналог кэпа; аналог кэпа, полученный из NAD+ (например, как описано в Kiledjian, Trends in Cell Biology 28, 454-464 (2018)); или модифицированный, например, биотинилированный, аналог кэпа (например, как описано в Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)).

В некоторых вариантах осуществления химически модифицированная нуклеиновая кислота содержит 3'-признак, выбранный из одного или нескольких из поли(A)-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'-O-метилированных NTP или фосфотиоат-NTP; химической модификации одного нуклеотида (например, окисление 3'-концевой рибозы до реакционноспособного альдегида с последующей конъюгацией альдегид-реактивного модифицированного нуклеотида) или химического лигирования с другой молекулой нуклеиновой кислоты.

В некоторых вариантах осуществления нуклеиновая кислота (например, нуклеиновая кислота-матрица) содержит один или несколько модифицированных нуклеотидов, например, выбранных из дигидроуридина, инозина, 7-метилгуанозина, 5-метилцитидина (5mC), 5'-риботимидинфосфата, 2'-O-метилриботимидина, 2'-O-этилриботимидина, 2'-фторриботимидина, C5-пропинилдезоксицитидина (pdC), C5-пропинилдезоксиуридина (pdU), C5-пропинилцитидина (pC), C5-пропинилуридина (pU), 5-метилцитидина, 5-метилуридина, 5-метилдезоксицитидина, 5-метоксиметилдезоксиуридина, 2,6-диаминопурина, 5'-диметокситритил-N4-этил-2'-дезоксицитидина, 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 8. 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.

Таблица 6. Модифицированные нуклеотиды

5-азауридин	N2-метил-6-тиогуанозин
2-тио-5-азауридин	N2,N2-диметил-6-тиогуанозин
2-тиоуридин	пиридин-4-оновый рибонуклеозид
4-тиопсевдоуридин	2-тио-5-азауридин
2-тиопсевдоуридин	2-тиоуридин
5-гидроксиуридин	4-тиопсевдоуридин
3-метилуридин	2-тиопсевдоуридин
5-карбоксиметилуридин	3-метилуридин
1-карбоксиметилпсевдоуридин	1-пропинилпсевдоуридин
5-пропинилуридин	1-метил-1-деазапсевдоуридин
1-пропинилпсевдоуридин	2-тио-1-метил-1-деазапсевдоуридин
5-тауринометилуридин	4-метоксипсевдоуридин
1-тауринометилпсевдоуридин	5'-O-(1-тиофосфат)-аденозин
5-тауринометил-2-тиоуридин	5'-O-(1-тиофосфат)-цитидин
1-тауринометил-4-тиоуридин	5'-O-(1-тиофосфат)-гуанозин
5-метилуридин	5'-O-(1-тиофосфат)-уридин
1-метилпсевдоуридин	5'-O-(1-тиофосфат)-псевдоуридин
4-тио-1-метилпсевдоуридин	2'-O-метиладенозин
2-тио-1-метилпсевдоуридин	2'-O-метилцитидин
1-метил-1-деазапсевдоуридин	2'-O-метилгуанозин
2-тио-1-метил-1-деазапсевдоуридин	2'-O-метилуридин

дигидроуридин	2'-O-метилпсевдоуридин
дигидропсевдоуридин	2'-O-метилинозин
2-тиодигидроуридин	2-метиладенозин
2-тиодигидропсевдоуридин	2-метилтио-N6-метиладенозин
2-метоксиуридин	2-метилтио-N6 изопентениладенозин
2-метокси-4-тиоуридин	2-метилтио-N6-(цис-гидроксиизопентенил)аденозин
4-метоксипсевдоуридин	N6-метил-N6-треонилкарбамоиладенозин
4-метокси-2-тиопсевдоуридин	N6-гидроксинорвалилкарбамоиладенозин
5-азацитидин	2-метилтио-N6-гидроксинорвалилкарбамоиладенозин
псевдоизоцитидин	2'-O-рибозиладенозин (фосфат)
3-метилцитидин	1,2'-O-диметилинозин
N4-ацетилцитидин	5,2'-O-диметилцитидин
5-формилцитидин	N4-ацетил-2'-O-метилцитидин
N4-метилцитидин	Лизидин
5-гидроксиметилцитидин	7-метилгуанозин
1-метилпсевдоизоцитидин	N2,2'-O-диметилгуанозин
пирролоцитидин	N2,N2,2'-O-триметилгуанозин
пирролопсевдоизоцитидин	2'-O-рибозилгуанозин (фосфат)
2-тиоцитидин	Вибутозин
2-тио-5-метилцитидин	Пероксивибутозин
4-тиопсевдоизоцитидин	Гидроксивибутозин
4-тио-1-метилпсевдоизоцитидин	недомодифицированный гидроксивибутозин
4-тио-1-метил-1-деазапсевдоизоцитидин	метилвиозин
1-метил-1-деазапсевдоизоцитидин	квеуозин
зебуларин	эпоксиквеуозин
5-азазебуларин	галактозилквеуозин
5-метилзебуларин	маннозилквеуозин
5-аза-2-тиозебуларин	7-циано-7-деазагуанозин
2-тиозебуларин	7-аминометил-7-деазагуанозин
2-метоксицитидин	археозин
2-метокси-5-метилцитидин	5,2'-O-диметилуридин
4-метоксипсевдоизоцитидин	
4-метокси-1-метилпсевдоизоцитидин	
2-аминопурин	
2,6-диаминопурин	

7-деазааденин	4-тиоуридин
7-деаза-8-азааденин	5-метил-2-тиоуридин
7-деаза-2-аминопурин	2-тио-2'-О-метилуридин
7-деаза-8-аза-2-аминопурин	3-(3-амино-3-карбоксипропил)уридин
7-деаза-2,6-диаминопурин	5-метоксиуридин
7-деаза-8-аза-2,6-диаминопурин	уридин-5-оксиуксусная кислота
1-метиладенозин	Сложный метиловый эфир уридин-5-оксиуксусной кислоты
N6-изопентениладенозин	5-(карбоксигидроксиметил)уридин)
N6-(цис-гидроксиизопентенил)аденозин	Сложный метиловый эфир 5-(карбоксигидроксиметил)уридина
2-метилтио-N6-(цис-гидроксиизопентенил)аденозин	5-метоксикарбонилметилуридин
N6-глицинилкарбамоиладенозин	5-метоксикарбонилметил-2'-О-метилуридин
N6-треонилкарбамоиладенозин	5-метоксикарбонилметил-2-тиоуридин
2-метилтио-N6-треонилкарбамоиладенозин	5-аминометил-2-тиоуридин
N6,N6-диметиладенозин	5-метиламинометилуридин
7-метиладенин	5-метиламинометил-2-тиоуридин
2-метилтиоаденин	5-метиламинометил-2-селеноуридин
2-метоксиаденин	5-карбамоилметилуридин
инозин	5-карбамоилметил-2'-О-метилуридин
1-метилинозин	5-карбоксиметиламинометилуридин
виозин	5-карбоксиметиламинометил-2'-О-метилуридин
вибутозин	5-карбоксиметиламинометил-2-тиоуридин
7-деазагуанозин	N4,2'-О-диметилцитидин
7-деаза-8-азагуанозин	5-карбоксиметилуридин
6-тиогуанозин	N6,2'-О-диметиладенозин
6-тио-7-деазагуанозин	N, N6,O-2'-триметиладенозин
6-тио-7-деаза-8-азагуанозин	N2,7-диметилгуанозин
7-метилгуанозин	N2,N2,7-триметилгуанозин
6-тио-7-метилгуанозин	3,2'-О-диметилуридин
7-метилинозин	5-метилдигидроуридин
6-метоксигуанозин	5-формил-2'-О-метилцитидин
1-метилгуанозин	1,2'-О-диметилгуанозин
N2-метилгуанозин	
N2,N2-диметилгуанозин	

8-оксогуанозин	4-деметилвиозин
7-метил-8-оксогуанозин	Изовиозин
1-метил-6-тиогуанозин	N6-ацетиладенозин

Таблица 7. Модификации остова

2'-O-метилмодифицированный остов
Остов на основе пептидной нуклеиновой кислоты (PNA)
Фосфотиоатный остов
Морфолиновый остов
Карбаматный остов
Силоксановый остов
Сульфидный остов
Сульфоксидный остов
Сульфоновый остов
Формацетильный остов
Тиоформацетильный остов
Метиленформацетильный остов
Рибоацетильный остов
Алкенсодержащий остов
Сульфаматный остов
Сульфонатный остов
Сульфаниламидный остов
Метилениминовый остов
Метиленгидразиновый остов
Амидный остов

Таблица 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
m7GpppA
m7GpppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

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 селектируемый маркер, например, маркер устойчивости к антибиотикам. В некоторых вариантах осуществления маркер устойчивости к антибиотикам представляет собой маркер устойчивости к канамицину. В некоторых вариантах осуществления маркер устойчивости к антибиотикам не придает устойчивости к бета-лактамным антибиотикам. В некоторых вариантах осуществления вектор не содержит маркер устойчивости к ампициллину. В некоторых вариантах осуществления вектор содержит маркер устойчивости к канамицину и не содержит маркер устойчивости к ампициллину. В некоторых вариантах осуществления вектор, кодирующий полипептид Gene Writer, интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления вектор, кодирующий полипептид Gene Writer, не интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления вектор, кодирующий нуклеиновую

кислоту-матрицу (например, РНК-матрицу), не интегрируется в геном клетки-мишени (например, при введении в клетку-, ткань-, орган-мишень или субъекту-мишени). В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то селектируемый маркер не интегрируется в геном. В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то гены или последовательности, участвующие в поддержании вектора (например, гены поддержания плазмиды), не интегрируются в геном. В некоторых вариантах осуществления, если вектор интегрируется в сайт-мишень в геноме клетки-мишени, то последовательности, регулирующие перенос (например, инвертированные концевые повторы, например, из AAV), не интегрируются в геном. В некоторых вариантах осуществления введение вектора (например, кодирующего полипептид Gene Writer, описанный в данном документе, нуклеиновую кислоту-матрицу, описанную в данном документе, или и то, и другое) в клетку-, ткань-, орган-мишень или субъекту-мишени приводит к интеграции части вектора в один или несколько сайтов-мишеней в геноме(геномах) указанных клетки-, ткани-, органа- или субъекта-мишени. В некоторых вариантах осуществления менее 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2 или 1% сайтов-мишеней (например, ни один из сайтов-мишеней), содержащих интегрированный материал, содержат селектируемый маркер (например, ген устойчивости к антибиотикам), последовательность, регулирующую перенос (например, инвертированный концевой повтор, например, из AAV), или и то, и другое из вектора.

Иллюстративные способы получения терапевтического фармацевтического белка или полипептида, описанного в данном документе, включают экспрессию в клетках млекопитающих, хотя рекомбинантные белки также могут быть получены с использованием клеток насекомых, дрожжей, бактерий или других клеток под контролем соответствующих промоторов. Экспрессионные векторы для клеток млекопитающих могут содержать нетранскрибируемые элементы, такие как точка начала репликации, подходящий промотор и другие 5'- или 3'- фланкирующие нетранскрибируемые последовательности и 5'- или 3'- нетранслируемые последовательности, такие как необходимые сайты связывания рибосомы, сайт полиаденилирования, донорные и акцепторные сайты сплайсинга и последовательности терминации. Последовательности ДНК, полученные из генома вируса SV40, например, точку начала репликации, ранний промотор, сайты сплайсинга и полиаденилирования SV40, можно использовать для получения других генетических элементов, требуемых для экспрессии гетерологичной последовательности ДНК. Подходящие клонирующие и экспрессионные векторы для применения с клетками-хозяевами, являющимися клетками бактерий, грибов, дрожжей и млекопитающих, описаны в Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

Различные системы культивирования клеток млекопитающих можно использовать для экспрессии и получения рекомбинантного белка. Примеры систем экспрессии на основе клеток млекопитающих включают линии клеток CHO, COS, HEK293, HeLa и BHK.

Способы культивирования клеток-хозяев для получения белковых терапевтических средств описаны в Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology)*, Springer (2014). Композиции, описанные в данном документе, могут содержать вектор, такой как вирусный вектор, например, лентивирусный вектор, кодирующий рекомбинантный белок. В некоторых вариантах осуществления вектор, например, вирусный вектор, может содержать нуклеиновую кислоту, кодирующую рекомбинантный белок.

Очистка белковых терапевтических средств описана в Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); и в Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

В некоторых вариантах осуществления стандарты качества включают без ограничения:

(i) длину мРНК, кодирующей полипептид Gene Writer, например, имеет ли мРНК длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК длину более 3000, 4000 или 5000 нуклеотидов;

(ii) присутствие, отсутствие и/или длину поли(A)-хвоста в мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК поли(A)-хвост (например, поли(A)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствие, отсутствие и/или тип 5'-кэпа на мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК 5'-кэп, например, является ли этот кэп 7-метилгуанозиновым кэпом, например, кэпом O-Me-m7G;

(iv) присутствие, отсутствие и/или тип одного или нескольких модифицированных нуклеотидов (например, выбранных из псевдоуридина, дигидроуридина, инозина, 7-метилгуанозина, 1-N-метилпсевдоуридина (1-Me-Ψ), 5-метоксиуридина (5-MO-U), 5-метилцитидина (5mC) или запертого нуклеотида) в мРНК, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей мРНК один или несколько модифицированных нуклеотидов;

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

(vi) активность мРНК в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая мРНК, анализируется на активность.

Наборы, изделия и фармацевтические композиции

В одном из аспектов настоящего изобретения предусмотрен набор, содержащий Gene Writer или систему Gene Writer, например, описанную в данном документе. В некоторых вариантах осуществления набор содержит полипептид Gene Writer (или

нуклеиновую кислоту, кодирующую полипептид) и РНК-матрицу (или ДНК, кодирующую РНК-матрицу). В некоторых вариантах осуществления набор дополнительно содержит реагент для введения системы в клетку, например, реагент для трансфекции, LNP и т. п. В некоторых вариантах осуществления набор подходит для любого из описанных в данном документе способов. В некоторых вариантах осуществления набор содержит один или несколько элементов, композиций (например, фармацевтических композиций), Gene Writer и/или систем Gene Writer или их функциональных фрагментов или компонентов, например, размещенных в изделии. В некоторых вариантах осуществления набор содержит инструкции по его применению.

В одном из аспектов настоящего изобретения предусмотрено изделие, например, в котором находится описанный в данном документе набор или его компонент.

В одном из аспектов настоящего изобретения предусмотрена фармацевтическая композиция, содержащая Gene Writer или систему Gene Writer, например, описанную в данном документе. В некоторых вариантах осуществления фармацевтическая композиция дополнительно содержит фармацевтически приемлемый носитель или наполнитель. В некоторых вариантах осуществления фармацевтическая композиция содержит РНК-матрицу и/или РНК, кодирующую полипептид. В вариантах осуществления фармацевтическая композиция обладает одной или несколькими (например, 1, 2, 3 или 4) следующими характеристиками:

(a) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) ДНК-матрицы относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(b) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) некэпированной РНК относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(c) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) РНК частичной длины относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(d) по сути отсутствуют непрореагировавшие динуклеотиды кэпа.

Физико-химические испытания, меры контроля качества и единообразия при производстве (CMC)

Очистка белковых терапевтических средств описана, например, в Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); и в Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

В некоторых вариантах осуществления система Gene Writer™, полипептид и/или нуклеиновая кислота-матрица (например, РНК-матрица) соответствуют определенным стандартам качества. В некоторых вариантах осуществления система Gene Writer™, полипептид и/или нуклеиновая кислота-матрица (например, РНК-матрица), полученные с помощью описанного в данном документе способа, соответствуют определенным стандартам качества. Соответственно, настоящее изобретение в некоторых аспектах

относится к способам получения системы Gene Writer™, полипептида и/или нуклеиновой кислоты-матрицы (например, РНК-матрицы), которые соответствуют определенным стандартам качества, например, в которых указанные стандарты качества являются проанализированными. В некоторых аспектах настоящее изобретение также относится к способам анализа указанных стандартов качества в системе Gene Writer™, полипептиде и/или нуклеиновой кислоте-матрице (например, РНК-матрице). В некоторых вариантах осуществления стандарты качества включают без ограничения одно или несколько (например, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 или 12) из следующего:

(i) длину РНК-матрицы, например, имеет ли РНК-матрица длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы длину более 100, 125, 150, 175 или 200 нуклеотидов;

(ii) присутствие, отсутствие и/или длину поли(A)-хвоста в РНК-матрице, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы поли(A)-хвост (например, поли(A)-хвост длиной по меньшей мере 5, 10, 20, 30, 50, 70, 100 нуклеотидов);

(iii) присутствие, отсутствие и/или тип 5'-кэпа в РНК-матрице, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы 5'-кэп, например, является ли этот кэп 7-метилгуанозиновым кэпом, например, кэпом O-Me-m7G;

(iv) присутствие, отсутствие и/или тип одного или нескольких модифицированных нуклеотидов (например, выбранных из псевдоуридина, дигидроуридина, инозина, 7-метилгуанозина, 1-N-метилпсевдоуридина (1-Me-Ψ), 5-метоксиуридина (5-MO-U), 5-метилцитидина (5mC) или запертого нуклеотида) в РНК-матрице, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующей РНК-матрицы один или несколько модифицированных нуклеотидов;

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

(vi) активность РНК-матрицы в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая РНК-матрицу, анализируется на активность;

(vii) длину полипептида, первого полипептида или второго полипептида, например, имеет ли полипептид, первый полипептид или второй полипептид, длину, которая превышает эталонную длину или находится в пределах диапазона эталонных длин, например, имеет ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, длину более 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600,

1700, 1800, 1900 или 2000 аминокислот (и необязательно длину не более 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700 или 600 аминокислот);

(viii) присутствие, отсутствие и/или тип посттрансляционной модификации полипептида, первого полипептида или второго полипептида, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% полипептида, первого полипептида или второго полипептида, фосфорилирование, метилирование, ацетилирование, миристоилирование, пальмитоилирование, изопренилирование, гликозилирование или липоилирование или любую их комбинацию;

(ix) присутствие, отсутствие и/или тип одной или нескольких искусственных, синтетических или неканонических аминокислот (например, выбранных из орнитина, β-аланина, GABA, δ-аминолевулиновой кислоты, PABA, D-аминокислоты (например, D-аланина или D-глутамата), аминоизомасляной кислоты, дегидроаланина, цистатионина, лантионина, дженколовой кислоты, диаминопимелиновой кислоты, гомоаланина, норвалина, норлейцина, гомонорлейцина, гомосерина, O-метилгомосерина и O-этилгомосерина, этионина, селеноцистеина, селеногомоцистеина, селенометионина, селеноэтионина, теллуроцистеина или теллурометионина) в полипептиде, первом полипептиде или втором полипептиде, например, содержит ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, одну или несколько искусственных, синтетических или неканонических аминокислот;

(x) стабильность полипептида, первого полипептида или второго полипептида (например, с течением времени и/или в заранее выбранных условиях), например, остается ли по меньшей мере 80, 85, 90, 95, 96, 97, 98 или 99% присутствующего полипептида, первого полипептида или второго полипептида, интактным (например, длиной более 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900 или 2000 аминокислот (и необязательно длиной не более 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700 или 600 аминокислот)) после теста на стабильность;

(xi) активность полипептида, первого полипептида или второго полипептида, в системе для модификации ДНК, например, модифицируется ли по меньшей мере 1% сайтов-мишеней после того, как система, содержащая полипептид, первый полипептид или второй полипептид, анализируется на активность; или

(xii) присутствие, отсутствие и/или уровень одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов или белков клетки-хозяина, например, свободна ли или по сути свободна ли система от пирогенов, вирусов, грибов, бактериальных патогенов или контаминации белками клетки-хозяина.

В некоторых вариантах осуществления система или фармацевтическая композиция, описанные в данном документе, не содержит эндотоксинов.

В некоторых вариантах осуществления определяют присутствие, отсутствие и/или уровень одного или нескольких пирогенов, вирусов, грибов, бактериальных патогенов

и/или белков клетки-хозяина. В вариантах осуществления определяют, свободна ли или по сути свободна ли система от пирогена, вируса, гриба, бактериального патогена и/или контаминации белками клетки-хозяина.

В некоторых вариантах осуществления фармацевтическая композиция или система, описанная в данном документе, обладает одной или несколькими (например, 1, 2, 3 или 4) следующими характеристиками:

(a) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) ДНК-матрицы относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(b) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) некэпированной РНК относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(c) менее 1% (например, менее 0,5%, 0,4%, 0,3%, 0,2% или 0,1%) РНК частичной длины относительно РНК-матрицы и/или РНК, кодирующей полипептид, например, в пересчете на моли;

(d) по сути отсутствуют непрореагировавшие динуклеотиды кэпа.

Пути применения

При применении систем, описанных в данном документе, необязательно при применении любого из способов доставки, описанных в данном документе (включая способы вирусной доставки, такие как AAV), в настоящем изобретении также предусмотрены пути применения (способы) для модификации молекулы ДНК, такой как ядерная ДНК, т. е. в геноме клетки, будь то in vitro, ex vivo, in situ или in vivo, например, в ткани, в организме, таком как субъект, в том числе у субъектов-млекопитающих, таких как человек. 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 определенных вариантах осуществления последовательность нуклеиновой кислоты-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, например, для терапевтического вмешательства.

В определенных аспектах в настоящем изобретении предусмотрены способы модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающие введение

системы, описанной в данном документе (необязательно с помощью способа, описанного в данном документе), в клетку, ткань или субъекту, где система встраивает последовательность гетерологичного объекта в нить ДНК-мишени, тем самым модифицируя нить ДНК-мишени. Таким образом, в определенных вариантах осуществления последовательность гетерологичного объекта экспрессируется в клетке, ткани или у субъекта. В некоторых вариантах осуществления клетка, ткань или субъект являются клеткой, тканью млекопитающего (например, человека) или субъектом-млекопитающим (например, человеком). Иллюстративные клетки, модифицированные таким образом, включают гепатоцит, клетку легочного эпителия, ионоцит. Такая клетка может быть первичной клеткой или иным образом не иммортализованной. В связанных аспектах в настоящем изобретении также предусмотрены способы обработки ткани млекопитающего, включающие введение млекопитающему системы, описанной в данном документе, за счет чего обеспечивается обработка ткани, где в ткани наблюдается дефицит последовательности гетерологичного объекта. В определенных вариантах осуществления любого из предшествующих аспектов и вариантов осуществления полипептид Gene Writer представлен в качестве нуклеиновой кислоты, которая присутствует транзиентно.

В некоторых вариантах осуществления система по настоящему изобретению способна производить вставку, замену, делецию или их комбинацию в ДНК-мишени. 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 вариантах осуществления система редактирования генов Gene Writer™ может предоставлять терапевтические трансгены, экспрессирующие, например, заместительные факторы свертывания крови или заместительные ферменты, например, лизосомальные ферменты. Например, композиции, системы и способы, описанные в данном документе, применимы для экспрессии в геноме-мишени, являющимся человеческим геномом, агалсидазы альфа или бета для лечения болезни Фабри; имиглюцеразы, талиглюцеразы альфа, велаглюцеразы альфа или алглюцеразы для лечения болезни Гоше; себелипазы альфа для лечения дефицита лизосомной кислой липазы (болезни Вольмана/CESD); ларонидазы, идурсульфазы, элосульфазы альфа или галсульфазы для лечения мукополисахаридозов; алглюкозидазы альфа для лечения болезни Помпе. Например, композиции, системы и способы, описанные в данном документе, применимы для экспрессии в геноме-мишени, являющимся человеческим геномом, фактора I, II, V, VII, X, XI, XII или XIII при дефицитах факторов свертывания крови.

В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует внутриклеточный белок (например, цитоплазматический белок, ядерный белок, белок органеллы, такой как митохондриальный белок или лизосомальный белок, или мембранный белок). В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует мембранный белок, например, мембранный белок, отличный от CAR, и/или эндогенный мембранный белок человека. В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует внеклеточный белок. В некоторых вариантах осуществления последовательность гетерологичного объекта кодирует фермент, структурный белок, сигнальный белок, регуляторный белок, транспортный белок, сенсорный белок, двигательный белок, защитный белок или запасной белок. Другие 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 T-, NK- или NKT-клеток. В некоторых вариантах осуществления клетки, модифицированные по меньшей мере одним CAR (например, CAR-T-клетки, CAR-NK-клетки, CAR-NKT-клетки), или комбинация клеток, модифицированных по меньшей мере одним CAR (например, смесь CAR-NK/T-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 a, GD2 (ганглиозида G2), GD3 (ганглиозида G3), GPC3 (глипикана-3), GPI00, hTERT (теломеразной обратной транскриптазы человека), ICAM-1, интегрина B7, рецептора интерлейкина 6, IL13Ra2 (рецептора интерлейкина-13 субъединицы альфа-2), каппа-легкой цепи, KDR (рецептора домена вставки киназы), LeY (антигена Y системы Льюис), L1CAM (молекулы клеточной адгезии LI), LILRB2 (рецептора B2, подобного лейкоцитарному иммуноглобулину), MARTI, MAGE-A1 (ассоциированного с меланомой антигена Al), MAGE-A3, MSLN (мезотелина), MUC16 (муцина 16), MUCI (муцина I), лигандов KG2D, NY-ESO-1 (антигена рака/яичка), PRI (протеиназы 3), TRBCI, TRBC2, TFM-3, TACI, 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 B, гепатита C, CMV (цитомегаловируса), EBV (вируса Эпштейна-Барр), HPV (вируса папилломы человека).

В некоторых вариантах осуществления иммунные клетки, например, T-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, T-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, паращитовидной железы, гипофиза, яичек, щитовидной железы, молочной железы, селезенки, миндалин, тимуса, лимфатического узла, костного мозга, легкого, сердечной мышцы, пищевода, желудка, тонкой кишки, толстой кишки, печени, слюнной железы, почки, предстательной железы, крови или в другой тип клеток или тканей. В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, применяется для лечения заболевания, такого как рак, воспалительное заболевание, инфекционное заболевание, генетический дефект или другое заболевание. Рак может являться раком головного мозга, мозжечка, надпочечников, яичников, поджелудочной железы, паращитовидной железы, гипофиза, яичек, щитовидной железы, молочной железы, селезенки, миндалин, тимуса, лимфатического узла, костного мозга, легкого, сердечной мышцы, пищевода, желудка, тонкой кишки, толстой кишки, печени, слюнной железы, почки, предстательной железы, крови или из другого типа клеток или тканей или может включать несколько видов рака.

В некоторых вариантах осуществления описанная в данном документе система Gene Writer™ вводится посредством энтерального введения (например, перорального, ректального, желудочно-кишечного, сублингвального, сублабиального или трансбуккального введения). В некоторых вариантах осуществления систему Gene Writer™, описанную в данном документе, вводят посредством парентерального введения (например, внутривенного, внутримышечного, подкожного, внутрикожного, эпидурального, внутримозгового, интрацеребровентрикулярного, накожного, назального, внутриартериального, внутрисуставного, внутрикавернозного, внутриглазного, внутрикостной инфузии, внутрибрюшинного, интратекального, внутриматочного, интравагинального, внутривезикального, периваскулярного или чресслизистого введения). В некоторых вариантах осуществления система Gene Writer™, описанная в данном документе, вводится посредством местного введения (например, трансдермального введения).

В некоторых вариантах осуществления система Gene Writing может применяться для осуществления множественных модификаций 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 нагрузка предназначены для настраиваемого контроля, например, посредством применения индуцируемого промотора. Например, промотор, например, Tet, управляющий геном, представляющим интерес, может являться "молчащим" при интеграции, но в некоторых случаях может активироваться при воздействии низкомолекулярного индуктора, например, доксициклина. В некоторых вариантах осуществления настраиваемая экспрессия позволяет контролировать ген (например, терапевтический ген) после обработки, например, обеспечивая эффект дозирования, зависящий от малых молекул. В вариантах осуществления эффект дозирования, зависящий от малых молекул, включает изменение уровней продукта гена во времени и/или в пространстве, например, посредством местного введения. В некоторых вариантах осуществления промотор, используемый в системе, описанной в данном документе, может быть индуцируемым, например, отвечающим на эндогенную молекулу хозяина и/или экзогенную малую молекулу, вводимую ему.

В некоторых вариантах осуществления система Gene Writing применяется для внесения изменений в некодирующие области и/или регуляторные контрольные области, например, для настройки экспрессии эндогенных генов. В некоторых вариантах осуществления система Gene Writing применяется для индукции повышения или понижения экспрессии генов. В некоторых вариантах осуществления регуляторная контрольная область содержит один или несколько из промотора, энхансера, UTR, сайта CTCF и/или области, контролирующей экспрессию гена.

В некоторых вариантах осуществления система Gene Writing применяется для лечения или предупреждения болезни экспансии повторов (например, заболевания из таблицы 26) или для уменьшения его тяжести или симптома. В некоторых вариантах осуществления болезнь экспансии повторов включает экспансию тринуклеотидных повторов. В некоторых вариантах осуществления у субъекта имеется по меньшей мере 10, 20, 30, 40 или 50 копий повтора. В вариантах осуществления болезнь экспансии повторов является наследственным заболеванием. Неограничивающие примеры болезней экспансии повторов включают болезнь Хантингтона (HD) и миотоническую дистрофию. Например, здоровые люди могут иметь от 10 до 35 тандемных копий тринуклеотидного повтора CAG, тогда как пациенты с болезнью Хантингтона часто имеют > 40 копий, что может привести, например, к образованию удлиненного и дисфункционального белка гентингтина. В некоторых вариантах осуществления Gene Writer корректирует экспансию повтора, например, распознавая ДНК на конце области повторов и внося однонитевой разрез в одну нить (фигура 30). В некоторых вариантах осуществления компонент Gene Writer, представляющий собой РНК-матрицу, содержит область с числом повторов, характерным для здорового субъекта, например, приблизительно 20 повторов (например, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35 или 35-40 повторов). В некоторых вариантах осуществления компонент Gene Writer, представляющий собой РНК-матрицу, копируется с помощью TPRT в сайт-мишень. В некоторых вариантах осуществления однонитевой разрыв во второй нити и синтез второй нити затем приводят к интеграции новой скопированной ДНК, содержащей правильное количество повторов (например, как описано в данном документе). В некоторых вариантах осуществления система распознает ДНК на конце области повторов, и при этом матрица несет информацию о новом количестве повторов. В вариантах осуществления таким образом можно применять Gene Writer независимо от количества повторов, присутствующих у индивидуума и/или в отдельной клетке. Из-за наличия множественных повторов альтернативное терапевтическое средство, отличное от Gene Writer (например, терапевтическое средство для гомологичной рекомбинации на основе CRISPR), в некоторых вариантах осуществления может привести к непредсказуемому поведению репарации. Дополнительные неограничивающие примеры болезней экспансии повторов и вызывающих их повторов можно найти, например, в La Spada and Taylor Nat Rev Genet 11(4):247-258 (2010), которая включена в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления систему Gene Writing можно применять для лечения здорового человека, например, в качестве профилактической терапии. В некоторых вариантах осуществления системы Gene Writing могут быть нацелены на создание мутаций, например, которые, как было показано, защищают от заболевания, представляющего интерес. Иллюстративный перечень таких заболеваний и мишеней защитных мутаций можно найти в таблице 22.

Систему Gene Writing™ можно применять для лечения заболеваний печени. В некоторых вариантах осуществления заболевания печени, предпочтительные для

терапевтического применения Gene Writing™, включают без ограничения заболевания, выбранные из любой из таблиц 10A-10D или 11A-11G или таблицы 5 документа WO2020014209, который включен в данный документ посредством ссылки. В иллюстративных вариантах осуществления дефицит OTC устраняют посредством доставки всего гена OTC или его фрагмента, например, всего гена OTC или его фрагмента, содержащегося в таблице 5 документа WO2020014209. В некоторых вариантах осуществления дефицит OTC устраняют посредством доставки полной кассеты экспрессии гена OTC в геном, которая дополняет функцию мутантного гена. В некоторых вариантах осуществления используется фрагмент гена OTC, который заменяет патогенную мутацию в ее эндогенном локусе. В других вариантах осуществления система Gene Writing™ применяется для лечения состояния, выбранного из столбца 6 таблицы 4 документа WO2020014209, абзац "Заболевания легких" ниже, или любой из таблиц 10A-10D или 11A-11G посредством доставки всей кассеты экспрессии гена, кодирующей соответствующий ген, указанный в столбце 1 таблицы 4 документа WO2020014209, абзац "Заболевания легких" ниже, или в любой из таблиц 10A-10D или 11A-11G, или ее фрагмента. В некоторых вариантах осуществления вся указанная кассета экспрессии гена или ее фрагмент доставляются в эндогенный локус патогенной мутации. В некоторых вариантах осуществления вся указанная кассета экспрессии гена или ее фрагмент интегрированы в отдельный локус в геноме и дополняют функцию мутантного гена.

В определенных вариантах осуществления система Gene Writer™ предоставляет последовательность гетерологичного объекта, содержащую ген, указанный в таблице 4 документа WO2020014209 или в абзаце "Заболевания легких" ниже.

Таблица 10A. Показания и генетические мишени, например, в печени

Заболевание	**Пораженный ген**
Острая интермиттирующая порфирия	HMBS
Дефицит альфа-1-антитрипсина (AAT)	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-CoA-дегидрогеназы)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
Болезнь накопления гликогена IV типа	GBE1
Гемофилия A	F8
Гемофилия B	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-CoA-мутазы)	MMUT
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS
MPS3 (синдром Санфилиппо)	Тип IIIa: SGSH Тип IIIb: NAGLU Тип IIIc: HGSNAT Тип IIId: 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

Таблица 10B. 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 A и B	SMPD1
Болезнь Ниманна-Пика C	NPC1
Болезнь Помпе	GAA
Дефицит пируваткиназы (PKD)	PKLR
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB

Таблица 10C. 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 окуломоторной апраксией)	SETX
Болезнь Тея-Сакса	HEXA
Х-сцепленная адренолейкодистрофия	ABCD1

Таблица 10D. 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" ретинальная 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

В некоторых вариантах осуществления система GeneWriter, описанная в данном документе, применяется для лечения показаний, указанных в любой из таблиц 11A-11G. Например, в некоторых вариантах осуществления система GeneWriter модифицирует сайт-мишень в геномной ДНК в клетке, где клетка содержит мутацию в гене, указанном в любой из таблиц 11A-11G, например, у субъекта, имеющего соответствующие показания, указанные в любой из таблиц 11A-11G. В некоторых вариантах осуществления сайт-мишень находится в случайном месте генома. В некоторых вариантах осуществления сайт-мишень находится в последовательности GSH. В некоторых вариантах осуществления клетка представляет собой клетку печени, которая содержит мутацию в гене из таблицы

11A, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11A. В некоторых вариантах осуществления клетка представляет собой HSC, которая содержит мутацию в гене из таблицы 11B, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11B. В некоторых вариантах осуществления клетка представляет собой клетку ЦНС, содержащую мутацию в гене из таблицы 11C, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11C. В некоторых вариантах осуществления клетка представляет собой клетку глаза, которая содержит мутацию в гене из таблицы 11D, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11D. В некоторых вариантах осуществления клетка представляет собой клетку легкого, которая содержит мутацию в гене из таблицы 11E, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11E. В некоторых вариантах осуществления клетка представляет собой мышечную клетку (например, клетку скелетных мышц), которая содержит мутацию в гене из таблицы 11F, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11F. В некоторых вариантах осуществления клетка представляет собой клетку кожи, которая содержит мутацию в гене из таблицы 11G, например, у субъекта, имеющего соответствующие показания, указанные в таблице 11G.

Таблица 11A. Показания и генетические мишени для печени

Заболевание	**Пораженный ген**
Дефицит 3-гидрокси-3-метилглутарил-CoA-лиазы	HMGCL
Дефицит 3-метилкротонил-CoA-карбоксилазы (3MCC)	MCCC1
Острая печеночная порфирия	ALAD
Острая интермиттирующая порфирия	HMBS
Синдром Алажилля	JAG1
Дефицит альфа-1-антитрипсина (AAT)	SERPINA1
Синдром Альстрема	ALMS1
Дефицит аргиназы	ARG1
Дефицит аргининосукцинатлиазы	ASL
Дефицит карбамоилфосфатсинтетазы I	CPS1
Болезнь накопления сложных эфиров холестерина	LIPA
Дефицит цитрина	SLC25A13
Цитруллинемия I типа	ASS1
Синдром Криглера-Найяра (гипербилирубинемия)	UGT1A1
D-2-гидроксиглутаровая ацидурия I типа	D2HGDH
Болезнь Фабри	GLA
Синдром семейной хиломикронемии	LPL

Семейная гиперхолестеринемия 4 типа (гомозиготная семейная холестеринемия)	LDLRAP1
Болезнь Фарбера	ASAH1
Болезнь Гоше	GBA
Глутаровая ацидурия I типа	GCDH
Глутаровая ацидурия II типа (множественный дефицит ацил-CoA-дегидрогеназы)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
Глутаровая ацидурия III типа	C7orf10
Болезнь накопления гликогена I типа	GSD1A: G6PC GSD1B: SLC37A4 GSD1C: SLC37A4
Болезнь накопления гликогена IV типа	GBE1
Гемофилия A	F8
Гемофилия B	F9
Наследственный амилоидоз (hTTR)	TTR
Наследственная непереносимость фруктозы	ALDOB
Наследственный гемохроматоз	HFE
Гомоцистинурия	CBS
Синдром гипераммониемии-гиперорнитинемии-гомоцитруллинурии (HHH)	SLC25A15
Гипероксалурия 1 типа (оксалоз)	HP1: AGXT HP2: GRHPR
Изовалериановая ацидемия	IVD
Дефицит липопротеинлипазы	LPL
Болезнь мочи с запахом кленового сиропа (MSUD)	Тип Ia: BCKDHA Тип Ib: BCKDHB Тип II: DBT
Метилмалоновая ацидемия (дефицит метилмалонил-CoA-мутазы)	MMUT
Синдром митохондриальной нейрогастроинтестинальной энцефалопатии	TYMP
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS

MPS3 (синдром Санфилиппо)	Тип IIIa: SGSH Тип IIIb: NAGLU Тип IIIc: HGSNAT Тип IIId: GNS
MPS4	Тип IVA: GALNS Тип IVB: GLB1
MPS6	ARSB
MPS7	GUSB
MPS9	HYAL1
Дефицит N-ацетилглутаматсинтазы	NAGS
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
Прогрессирующий семейный внутрипеченочный холестаз 4 типа	TJP2
Пропионовая ацидемия	PCCB; PCCA
Дефицит пируваткарбоксилазы	PC
Тирозинемия I типа	FAH
Болезнь Вильсона	ATP7B

Таблица 11B. Показания и генетические мишени для HSC

Заболевание	**Пораженный ген**
ADA-SCID	ADA
JAK-3 SCID	JAK3
Дефицит пируваткиназы (PKD)	PKLR
Дефицит RAG 1/2 (SCID с гранулемами)	RAG
Синдром Вискотта-Олдрича	WAS
Х-сцепленная агаммаглобулинемия	BTK

X-сцепленный SCID	IL2RG
Адренолейкодистрофия (CALD)	ABCD1
Альфа-маннозидоз	MAN2B1
Аспартилглюкозаминурия	AGA
Анемия Даймонда-Блекфена	
Хроническая гранулематозная болезнь	CYBB; CYBA; NCF1; NCF2; NCF4
Врожденная амегакариоцитарная тромбоцитопения	MPL
Врожденная дизэритропоэтическая анемия	CDAN1
Анемия Фанкони	FANC
Фукозидоз	FUCA1
Болезнь Гоше	GBA
Тромбастения Гланцмана	ITGA2B, ITGB3
Глобоидно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Болезнь Гюнтера (врожденная эритропоэтическая порфирия)	UROS
Гемофагоцитарный лимфогистиоцитоз	PPRF1; STX11; STXBP2; UNC13D
Гемофилия A	F8
Гемофилия B	F9
Синдром IPEX	FOXP3
Синдром Костманна (тяжелая врожденная нейтропения)	HAX1
Недостаточность адгезии лейкоцитов	ITGB2, FERMT3
Злокачественный детский остеопетроз - аутосомно-рецессивный остеопетроз	Вовлечены многие гены
Метахроматическая лейкодистрофия	PSAP
Дефицит кофактора молибдена	MOCS1
MPS 1S (синдром Шейе)	IDUA
MPS2	IDS
MPS6	ARSB
MPS7	GUSB
Муколипидоз II типа	GNPTAB

Болезнь Ниманна-Пика A и B	SMPD1
Болезнь Ниманна-Пика C	NPC1
Болезнь Помпе	GAA
Дефицит пуриннуклеозидфосфорилазы	PNP
Синдром Швахмана-Даймонда	SBDS
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB
Синдром Вольмана	LIPA
X-сцепленный синдром гипер-IgM	TNFSF5
X-сцепленное лимфопролиферативное заболевание	SH2D1A
X-сцепленная сидеробластная анемия	ALAS2
Дефицит 17-альфа-гидроксилазы	CYP17A1
Вариабельный неклассифицируемый иммунодефицит	Вовлечены многие гены
IL-7R SCID	IL7R
Синдром Костманна (тяжелая врожденная нейтропения)	ELANE

Таблица 11C. Показания и генетические мишени для ЦНС

Заболевание	**Пораженный ген**
Синдром Корнелии де Ланге (SMC1A)	SMC1A
Нейрофиброматоз 1 типа	NF1
Нейрофиброматоз 2 типа	NF2
Синдром Ретта (MECP2)	MECP2

Таблица 11D. Показания и генетические мишени для глаза

Заболевание	**Пораженный ген**
Палочко-колбочковая дистрофия (CRX)	CRX
Палочко-колбочковая дистрофия (GUCY2D)	GUCY2D
Решетчатая дистрофия роговицы I типа	TGFBI
Пигментный ретинит (AD)	RHO
Вителлиформная макулярная дистрофия	BEST1, PRPH2

Таблица 11E. Показания и генетические мишени для легких

Заболевание	**Пораженный ген**
Дефицит альфа-1-антитрипсина	SERPINA1

Кистозный фиброз	CFTR
Последовательность деформации акинезии плода 1	MUSK
Последовательность деформации акинезии плода 2	RAPSN
Последовательность деформации акинезии плода 3	DOK7
Последовательность деформации акинезии плода 4	NUP88
Синдромальная микрофтальмия 9	STRA6
Первичная цилиарная дискинезия	DNAI1; DNAH5 (~30%) многие другие
Легочный альвеолярный микролитиаз	SLC34A2
Легочная веноокклюзионная болезнь 2	EIF2AK4
Дефицит сурфактантного белка B (SP-B) (нарушение метаболизма легочного сурфактанта 1)	SFTPB
Дисфункция метаболизма сурфактанта, легочная, 5	CSF2RB
Дисфункция метаболизма сурфактанта, легочная, 4	CSF2RA
Дисфункция метаболизма сурфактанта, легочная, 3	ABCA3

Таблица 11F. Показания и генетические мишени для скелетных мышц

Заболевание	**Пораженный ген**
Мышечная дистрофия Беккера	DMD
CMD типа Фукуяма	FKTN
Заболевания мышц, глаз и головного мозга	POMGNT1
Синдром ригидного позвоночника	SELENON
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса, AR	LMNA
Поясно-конечностная мышечная дистрофия 2A	CAPN3
Поясно-конечностная мышечная дистрофия 2B	DYSF
Поясно-конечностная мышечная дистрофия типа 2E	SGCB
Поясно-конечностная мышечная дистрофия типа 2F	SGCD
Поясно-конечностная мышечная дистрофия типа 2H	TRIM32
Поясно-конечностная мышечная дистрофия типа 2C	SGCG

Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Центронуклеарная миопатия 2	BIN1
Центронуклеарная миопатия, X-сцепленная	MTM1
Центронуклеарная миопатия 5	SPEG
Центронуклеарная миопатия 6	ZAK
Врожденные миопатии с диспропорцией типов мышечных волокон	SELENON; TPM3; ACTA1
Болезнь центральных волокон	RYR1
Множественная миниядерная миопатия	RYR1
Миопатия с накоплением миозина, AR	MYH7
Немалиновые миопатии	NEB ACTA1
Миопатия GNE/миопатия Нонака/наследственная миопатия с тельцами включения	GNE
Дистальная миопатия с поздним началом типа Марксбери-Григгса	TTN
Мышечная дистрофия Миеси 1	DYSF
Мышечная дистрофия Миеси 3	ANO5
Миозит с тельцами включения 2	GNE
Дефицит карнитина	SLC22A5
Дефицит карнитинпальмитоилтрансферазы 1A	CPT1A
Дефицит миоаденилатдезаминазы	AMPD1
Дефицит фосфофруктокиназы (болезнь Таруи) (GSD VII)	PFKM
Дефицит фосфоглицераткиназы	PGK1
Дефицит фосфоглицератмутазы (GSD X)	PGAM2
Дефицит фосфорилазы (болезнь Мак-Ардля) (GSD V)	PYGM
Миофибриллярная миопатия 1	DES
Миофибриллярная миопатия 7	KY
Миофибриллярная миопатия 8	PYROXD1
Врожденный миастенический синдром	CHRNE (> 50% случаев); RAPSN, CHAT, COLQ, DOK7; другие

Болезнь Шарко-Мари-Тута 2	MFN2 (20% случаев)
Болезнь Шарко-Мари-Тута 4	A: GDAP1 F: PRX K: SURF1
Гигантская аксональная нейропатия 1	GAN
Миотония Беккера	CLCN1
Синдром истощения митохондриальной ДНК 2	TK2

Таблица 11G. Показания и генетические мишени для кожи

Заболевание	**Пораженный ген**
Энтеропатический акродерматит	SLC39A4
Алкаптонурия	HGD
Атаксия-телеангиэктазия	ATM
Синдром Берардинелли-Сейпа	BSCL2; AGPAT2; CAVIN1; CAV1
Дефицит биотинидазы	BTD
Синдром Блума	RECQL3
Церебротендинозный ксантоматоз	CYP27A1
Болезнь Чедиака-Хигаси	LYST
Хондродисплазия точечная	ARSE
Хроническая гранулематозная болезнь	CYBB; CYBA; NCF1; NCF2; NCF4
Синдром Коккейна	ERCC8, ERCC6
Врожденная эритропоэтическая порфирия	UROS
Синдром дряблой кожи	ATP6VOA2; FBLN5; ALDH18A1; PYCR1; LTBP4
Синдром де Барси	ALHD18A1, PYCR1
Дерматоспараксис (синдром Элерса-Данлоса VII типа)	ADAMTS2
Врожденный дискератоз	TERC; TERT; RTEL1; ACD; DKC1; NOLA3
Синдром Элерса-Данлоса VI типа	PLOD1, FKBP14
Бородавчатая эпидермодисплазия	TMC6
Претибиальная форма дистрофического буллезного эпидермолиза	COL7A1
Рецессивный дистрофический буллезный эпидермолиз (подтип Аллопо-Сименса)	COL7A1, MMP1

Буллезный узловой эпидермолиз	LAMA3; LAMC2; LAMB3; COL17A1; ITGB4
Эритропоэтическая протопорфирия	FECH
Болезнь Фабри	GLA
Семейная дизавтономия	IKBKAP
Семейный кожно-слизистый кандидоз	CARD9; TRAF3IP2; IL17F; CLEC7A; IL17RC
Анемия Фанкони	FANC
Очаговая лицевая эктодермальная дисплазия	EDA
Синдром GAPO	ANTXR1
Синдром Гришелли 1, 2 и 3 типов	MYO5A; RAB27A; MLPH
Ихтиоз Арлекина	ABCA12
Болезнь Хартнупа	SLC6A19
Гемохроматоз	HFE; HJV; TFR2
Синдром Германски-Пудлака	HPS1; HPS6; AP3B1; HPS4; DTNBP1; HPS5; HPS3; AP3D1; BLOC1S3; BLOC1S6
Синдром гиалинового фиброматоза (системный гиалиноз)	ANTXR2
Гипогидротическая эктодермальная дисплазия	EDA; EDAR; EDARADD
Фолликулярный декальвирующий шиповидный кератоз	MBTPS2
Синдром Киндлера	FERMT1
Пластинчатое отшелушивание у новорожденного	TGM1
Ламеллярный ихтиоз/небуллезная врожденная ихтиозиформная эритродермия (ARCI)	ALOX12B
Лепреконизм	INSR
Липогранулематоз	ASAH1
Липоидный протеиноз	ECM1
Болезнь острова Меледа	SLURP1
Болезнь Менкеса	ATP7A
Вариант синдрома множественных птеригиумов Эскобара	CHRNG
Синдром Нетертона	SPINK5

Синдром Ноя-Лаксовой	PHGDH; PSAT1
Болезнь накопления нейтральных липидов с ихтиозом	ABHD5
Тирозиназо-отрицательная форма глазо-кожного альбинизма	TYR
Тирозиназо-положительная форма глазо-кожного альбинизма	OCA2, MC1R
Синдром Папийона-Лефевра	CTSC
Синдром шелушения кожи	CDSN; TGM5; CHST8; CSTA; SERPINB8; FLG2
Стекловидные волосы (синдром нерасчесываемых волос)	PADI3
Дефицит пролидазы	PEPD
Эластическая псевдоксантома (PXE)	ABCC6
Болезнь Рефсума	PHYH
Рестриктивная дермопатия	ZYMPSTE24; LMNA
Синдром Рихнера-Ханхарта (тирозинемия II типа)	TAT
Синдром Ротмунда-Томсона	RECQL4, ANAPC1
Синдром Шегрена-Ларссона	ALDH3A2
Транзиторный буллезный дермолиз новорожденного	COL7A1
Трихотиодистрофия	ERCC2; ERCC3; MPLKIP; GTF2H5; GTF2E2
Опухолевый кальциноз	GALNT3; SAMD9; FRF23;
Витилиго	Различные
Синдром Вернера	RECQL2
Синдром Вискотта-Олдрича	WAS
Болезнь курчавых волос	KRT25, LPAR6
Пигментная ксеродерма	ERCC2; ERCC3; ERCC4; ERCC5; XPA; XPC; DDB2; POLH
Х-сцепленный рецессивный ихтиоз	STS
Альбинизм, желтый мутантный тип	TYR

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, BH4-дефицит в следствие частичного дефицита PTS, BH4-дефицит, D и несвязанный с ФКН; гиперинсулинемическая семейная гипогликемия 3, 4 и 5; фолликулярный кератоз; орально-лицевой-пальцевой синдром; синдром SeSAME; глухота, несиндромальная сенсоневральная, митохондриальная; протеинурия; синдром секреторной диареи при инсулинзависимом сахарном диабете; болезнь моямоя 5; анемия Даймонда-Блэкфана 1, 5, 8 и 10; синдром псевдоахондропластической спондилоэпифизарной дисплазии; синдром ломкой роговицы 2; метилмалоновая ацидемия с гомоцистинурией; синдром Адамса-Оливера 5 и 6; аутосомно-рецессивная агаммаглобулинемия 2; пороки развития затылочной части коры головного мозга; фебрильные судороги, семейные, 11; мукополисахаридоз VI типа, типа VI (тяжелый) и VII типа; синдром Мардена Уокера; псевдонеонатальная адренолейкодистрофия; миопатия сфероидного тела; ключично-черепной дизостоз; множественные кожные и слизистые венозные мальформации; детская острая печеночная недостаточность; неонатальный внутрипеченочный холестаз, вызванный дефицитом цитрина; дефект межжелудочковой перегородки 1; окуло-денто-дигитальная дисплазия; опухоль Вильмса 1; синдром Вейля-Марчезани; почечная адисплазия; катаракта 1, 4, аутосомно-доминантная, аутосомно-доминантная, множественные, с микрокорнеа, по типу Coppock, ювенильная, с микрокорнеа и глюкозурией, ядерная диффузная непрогрессирующая; одонто-гипофосфатазия; цереброокулофациоскелетный синдром; шизофрения 15; церебральная амилоидная ангиопатия, связанная с APP; гемофагоцитарный лимфогистиоцитоз, семейный, 3; недостаточность порфобилиногенсинтазы; эпизодическая атаксия 2 типа; трихо-рино-фаланговый синдром 3 типа; прогрессирующая семейная сердечная блокада типа IB; восприимчивость к глиоме 1; синдром Лихтенштейна-Кнорра; гипогидротическая X-сцепленная эктодермальная дисплазия; синдром Барттера 3, 3 с гипокальциурией и 4 ; недостаточность карбоангидразы VA, в следствие гипераммониемии; кардиомиопатия; пойкилодермия, наследственная фиброзирующая, с контрактурами сухожилий, миопатией и легочным фиброзом; комбинированная d-2- и 1-2-гидроксиглутаровая ацидурия; недостаточность аргиназы; палочко-колбочковая дистрофия 2 и 6; синдром Смита-Лемли-Опица; муколипидоз III гамма; синдром Блау; синдром Вемера; менингиома; дефект йодтирозильного сочетания; синдром Дубина-Джонсона; недостаточность 3-оксо-5-альфа-стероид-дельта-4-дегидрогеназы; синдром Буше-Нейхаузера; накопление железа в

головном мозге; умственная отсталость, сцепленная с X-хромосомой 102 и синдромальная 13; семейная предрасположенность к аденоме гипофиза; гипоплазия мозолистого тела; гиперальфалипопротеинемия 2; недостаточность ферроксидазы; нечувствительность к гормону роста при иммунодефиците; синдром Маринеско- Sj\xc3\xb6gren; синдром Марцольфа; семейный горизонтальный паралич взгляда с прогрессирующим сколиозом; синдром Митчелла-Райли; Гипокальциурическая гиперкальциемия, семейная, 1 и 3 типы; синдром Рубинштейна-Тайби; синдром Эпштейна; ювенильный ретиношизис; мышечная дистрофия Беккера; синдром Лойса-Дитца 1, 2, 3; врожденный мышечно-гипертрофический церебральный синдром; семейная ювенильная подагра; сперматогенная недостаточность 11, 3 и 8; орофациальные расщелины 11 и 7, синдром расщелины губы/неба и эктодермальной дисплазии; умственная отсталость, X-сцепленная, неспецифическая, синдромальная, типа Хедера и синдромальная, типа wu; комбинированные недостаточности окислительного фосфорилирования 1, 3, 4, 12, 15 и 25; лобно-височная деменция; дисплазия Кнеста; семейная кардиомиопатия; доброкачественная семейная гематурия; феохромоцитома; глухота, вызванная аминогликозидами; недостаточность трансаминазы гамма-аминомасляной кислоты; глазокожный альбинизм типа IB, типа 3 и типа 4; синдром почечной колобомы; гипомиелинизация ЦНС; синдром лимфангиэктазии-лимфедемы Хеннекама 2; базилярная мигрень, семейная; дистальная спинальная мышечная атрофия, сцепленная с X-хромосомой 3; X-сцепленная перивентрикулярная гетеротопия; микроцефалия; мукополисахаридоз, MPS-IH/S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; инфантильный паркинсонизм-дистония; лобно-височная деменция с включениями TDP43, связанная с TARDBP; наследственный диффузный рак желудка; сиалидоз типа I и II; синдром микроцефалии-капиллярной мальформации; синдром наследственного рака молочной железы и яичников; поражение мелких сосудов головного мозга с кровоизлиянием; некетотическая гиперглицинемия; нейрогепатопатия Навахо; аурикулокондилярный синдром 2; спастическая параплегия 15, 2, 3, 35, 39, 4, аутосомно-доминантная, 55, аутосомно-рецессивная и 5A; аутосомно-рецессивный эластолизис типа IA и IB; гемолитическая анемия несфероцитарная, обусловленная недостаточностью глюкозофосфатизомеразы; синдром Хатчинсона-Гилфорда; семейная амилоидная нефропатия с крапивницей и глухотой; надклапанный аортальный стеноз; диффузная ладонно-подошвенная кератодермия Ботнического типа; синдром Холта-Орама; синдром Коффина-Сириса/умственная отсталость; пороки развития лево-правой оси; синдром Рападилино; нанофтальм 2; краниосиностоз и зубочелюстные аномалии; параганглиомы 1; синдром Снайдера-Робинсона; фибрилляция желудочков; синдром активированного PI3K-дельта; синдром Хауэла-Эванса; синдром Ларсена, доминантный тип; синдром Ван Мальдергема 2; MYH-ассоциированный полипоз; недостаточность 6- pymvoyl-тетрагидроптеринсинтазы; синдромы Алажиля 1 и 2; лимфангиомиоматоз; мышечно-глазное заболевание головного мозга; расстройства, связанные с WFSl; первичная гипертрофическая остеоартропатия, аутосомно-рецессивная 2; бесплодие; синдром

прогерии Нестора-Гильермо; недостаточность митохондриального трифункционального белка; синдром гипоплазии левых отделов сердца 2; Первичная дилатационная кардиомиопатия; пигментный ретинит; болезнь Гиршпрунга 3; синдром Апшоу-Шульмана; дисплазия Дебукуа 2; диарея 3 (секреторная натриевая, врожденная, синдромальная) и 5 (при тафтинговой энтеропатии, врожденная); врожденная пахионихия 4 и 2 типа; церебральная аутосомно-доминантная и рецессивная артериопатия с подкорковыми инфарктами и лейкоэнцефалопатией; дистрофия формы Vi tel 1i; тип II, тип IV, IV (сочетанный печеночный и миопатический), тип V и тип VI; атипичный синдром Ретта; дефект атриовентрикулярной перегородки 4; синдром Папийона- Lef\xc3\xa8vre; амавроз Лебера; X-сцепленная наследственная моторная и сенсорная невропатия; прогрессирующая склерозирующая полиодистрофия; синдром Гольдмана-Фавра; почечно-печеночно-панкреатическая дисплазия; синдром Паллистера-Холла; амилоидогенный транстиретиновый амилоидоз; синдром Мельника-Нидлса; синдром гипериммуноглобулина E; атаксия задних столбов спинного мозга с пигментным ретинитом; точечная хондродисплазия 1, X-сцепленная рецессивная и 2 X-сцепленная доминантная; эктопия хрусталика, изолированная аутосомно-рецессивная и доминантная; семейная холодовая крапивница; семейный аденоматозный полипоз 1 и 3; порокератоз 8, диссеминированный поверхностный актинический тип; спектр синдромов избыточного роста, связанных с PIK3CA; церебральные кавернозные мальформации 2; экссудативная витреоретинопатия 6; врожденная мраморная кожа, телеангиэктатическая мегалэнцефалия; TARP синдром; сахарный диабет, перманентный неонатальный, с неврологическими характеристиками; короткореберная грудная дисплазия 11 или 3 с полидактилией или без нее; гипертрихотическая остеохондродисплазия; бета-талассемия; болезнь Ниманна-Пика типа C1, C2, типа A и типа C1, взрослая форма; болезнь Шарко-Мари-Тута типы IB, 2B2, 2C, 2F, 21, 2U (аксональный), 1C (демиелинизирующий), доминантный промежуточный C, рецессивный промежуточный A, 2A2, 4C, 4D, 4H, IF, IVF и X; тирозинемия I типа; пароксизмальная фибрилляция предсердий; УФ-чувствительный синдром; агенезия зубов, селективная, 3 и 4; врожденная мышечная дистрофия с недостаточностью мерозина; недостаточность длинноцепочечной 3-гидроксиацил-КоА-дегидрогеназы; врожденная аниридия; гипертрабекулярность левого желудочка 5; дефицит декарбоксилазы ароматических L-аминокислот; ишемическая болезнь сердца; тотальная лейконихия; дистальный артрогрипоз типа 2B; пигментный ретинит 10, 11, 12, 14, 15, 17 и 19; синдром Робинова-Сорауфа; синдром Тенорио; пролактинома; нейрофиброматоз, тип 2; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями головного мозга и глаз, типы A2, A7, A8, A11 и A14; гетеротаксия, висцеральная, 2, 4 и 6, аутосомная; синдром Янковича Риверы; семейная парциальная липодистрофия 2 и 3 типа; заболевание гемоглобина H, без делеций; мультицентрический остеолиз, нодулез и артропатия; агенезия щитовидной железы; недостаточность семейства ацил-КоА-дегидрогеназ, члена 9; болезнь Александра; болезнь накопления фитановой кислоты; рак молочной железы и яичников, семейный 1, 2 и 4; недостаточность пролиндегидрогеназы; детская гипофосфатазия; агенезия

поджелудочной железы и врожденные пороки сердца; витамин D-зависимый рахит, тип 2; иридогониодизогенез доминантный тип и тип 1; синдром аутосомно-рецессивной гипогидротической эктодермальной дисплазии; умственная отсталость, сцепленная с X-хромосомой, 3, 21, 30 и 72; наследственные геморрагические телеангиэктазии 2 типа; блефарофимоз, птоз и обратный эпикантус; недостаточность аденинфосфорибозилтрансферазы; судороги, доброкачественные семейные инфантильные, 2; акродизостоз 2, с гормонорезистентностью или без нее; тетрада Фалло; пигментный ретинит 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; недостаточность лизосомальной кислотной липазы; врожденная мышечная дистрофия типа Эйхсфельда; врожденная мышечная дистрофия Уокера-Варбурга; синдром периодической лихорадки, ассоциированной с рецептором TNF (TRAPS); прогрессирующая миоклонус-эпилепсия с атаксией; эпилепсия, детские абсансы 2, 12 (идиопатическая генерализованная, предрасположенность) 5 (лобная ночная эпилепсия), лобная ночная эпилепсия типа 1, парциальная, с различными очагами, прогрессирующая миоклония 3 и X-сцепленная, с различными нарушениями обучаемости и поведенчискими расстройствами; синдром удлиненного интервала QT; дикарбоновая аминоацидурия; брахидактилия типов A1 и A2; заболевание, подобное эластической псевдоксантоме, с множественной недостаточностью факторов свертывания крови; синдром мультисистемной дисфункции гладкой мускулатуры; синдактилия Сенани типа Ленца; синдром Жубера 1, 6, 7, 9/15 (дигенный), 14, 16 и 17 и орофациодигитальный синдром xiv; дигитореноцеребральный синдром; ретинобластома; семейная дискинезия с миокимией лица; наследственная сенсорная и вегетативная нейропатия типа IIB и IIA; семейный гиперинсулинизм; мегалэнцефальная лейкоэнцефалопатия с подкорковыми кистами тип 2a; синдром Ааза; синдром Видемана-Штайнера; эксфолиативный ихтиоз; врожденная миотония; гранулематозное заболевание, хронический, сцепленный с X-хромосомой вариант; недостаточность 2-метилбутирил-КоА-дегидрогеназы; саркоидоз с ранним началом; глаукома, врожденная и глаукома врожденная, колобома; рак молочной железы, предрасположенность к нему; цероидный липофусциноз нейронов 2, 6, 7 и 10; врожденная генерализованная липодистрофия 2 типа; недостаточность фруктозо-бифосфатазы; врожденная контрактурная арахнодактилия; синдром Линча I и II; недостаточность фосфоглицератдегидрогеназы; синдром Берна-Маккеуна; инфаркт миокарда 1; ахроматопсия 2 и 7; пигментный ретинит 73; дефект протана; полимикрогирия, асимметричная, двусторонняя лобно-теменная; спинальная мышечная атрофия, дистальная, аутосомно-рецессивная, 5; метилмалоновая ацидурия из-за недостаточности метилмалонил-КоА-мутазы; семейная порэнцефалия; синдром Гурлера; ухо-небно-пальцевый синдром I и II типов; синдром Сотоса 1 или 2; кардиоэнцефаломиопатия, инфантильная с летальным исходом, обусловленная недостаточностью цитохром-с-оксидазы; парастремматическая карликовость; резистентность к тиреотропин-рилизинг-гормону, генерализованная; сахарный диабет 2 типа и инсулинозависимый - 20; аневризмы грудной аорты и расслоение аорты; резистентность к эстрогенам; заболевание мочи “кленового сиропа” типа 1A и типа 3;

гипоспадия 1 и 2, сцепленная с Х-хромосомой; метахроматическая лейкодистрофия ювенильный, поздний инфантильный и взрослый типы; ранний Т-клеточный предшественник острого лимфобластного лейкоза; нейропатия, наследственная сенсорная, тип IC; умственная отсталость, аутосомно-доминантная 31; пигментный ретинит 39; рак молочной железы с ранним началом; аномалия Мэя-Хегглина; болезнь Гоше 1 типа и подострая нейропатическая; синдром Темтами; спинальная мышечная атрофия с преимущественным поражением нижних конечностей 2 типа, аутосомно-доминантный; анемия Фанкони, группы комплемента E, I, N и O; алкаптонурия; болезнь Гиршпрунга; комбинированная малоновая и метилмалоновая ацидурия; аритмогенная правожелудочковая кардиомиопатия 5, 8 и 10 типов; врожденное липоматозное разрастание, сосудистые мальформации и эпидермальные невусы; синдром Тимоти; недостаточность гуанидиноацетатметилтрансферазы; миоклоническая дистония; болезнь Канзаки; дефект транспорта нейтральной 1 аминокислоты; нейрогипофизарный несахарный диабет; метаболизм гормонов щитовидной железы, патологический; доброкачественная лопаточно-перонеальная мышечная дистрофия с кардиомиопатией; гипогликемия с дефицитом гликогенсинтетазы в печени; гипертрофическая кардиомиопатия; миастенический синдром, врожденный, 11, связанный с дефицитом ацетилхолиновых рецепторов; синдромная Х-сцепленная умственная отсталость 5; синдром Сторморкена; апластическая анемия; умственная отсталость; нормокалиемический периодический паралич, калий-чувствительный; болезнь Данона; нефронофтиз 13, 15 и 4; тиреотоксический периодический паралич и тиреотоксический периодический паралич 2; бесплодие, связанное с многохвостыми сперматозоидами и дополнительными ДНК; глаукома, первичная открытоугольная, с ранним началом; афибриногенемия и врожденная афибриногенемия; поликистоз почек 2, взрослый и детский типы; поздняя семейная кожная порфирия; церебелло-окуло-ренальный синдром (нефронофтиз, окуломоторная апраксия и мозжечковые нарушения); лобно-височная деменция, сцепленная с хромосомой 3 и убиквитин-позитивная лобно-височная деменция; метатрофическая дисплазия; синдром иммунодефицита, нестабильности центромерного участка и лицевых аномалий 2; анемия несфероцитарная гемолитическая, обусловленная дефицитом G6PD; бронхоэктазы с повышенным содержанием хлоридов в поте или без 3; врожденная миопатия с диспропорцией мышечных волокон; комплекс Карнея, тип 1; крипторхизм односторонний или двусторонний; буллезный ихтиоз Сименса; изолированный дефицит лютропина; DFNA 2 несиндромальная потеря слуха; синдром Кляйна-Ваарденберга; синдром серых тромбоцитов; дефект синтеза желчных кислот, врожденный, 2; 46, нарушение формирования пола XY, типы 1, 3 и 5; острая интермитирующая порфирия; синдромы Корнелии де Фанге 1 и 5; гиперглицинурия; колбочко-палочковая дистрофия 3; дисфибриногенемия; синдром Карака; врожденная мышечная дистрофия-дистрогликанопатия без умственной отсталости, тип B5; инфантильный нистагм, сцепленный с Х-хромосомой; врожденный дискератоз, аутосомно-рецессивный, 1, 3, 4 и 5; микроцефалия с хориоретинопатией, лимфедемой или умственной

отсталостью или без; гиперлизинемия; синдромы Барде-Бидля 1, 11, 16 и 19; аутосомно-рецессивная центронуклеарная миопатия; синдром Фрейзера; синдром каудальной регрессии; врожденный фиброз экстраокулярных мышц, 1, 2, 3a (с или без экстраокулярного поражения), 3б; синдром Прадера-Вилли; злокачественная меланома; синдром Блума; заболевание Дарье, сегментарное; многоочаговая остеолизная нефропатия; гемохроматоз 1, 2B и 3 типов; детская мозжечковая атаксия с прогрессирующей наружной офтальмоплегией и мозжечковая атаксия, умственная отсталость и синдром нарушения равновесия 2; синдром гипоплазии левых отделов сердца; эпилепсия, потеря слуха и синдром умственной отсталости; локус 2 количественного признака уровня трансферрина в сыворотке крови; глазной альбинизм I типа; синдром Марфана; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями головного мозга и глаз, тип A14 и B14; гипераммониемия III типа; синдром криптофтальма; врожденная универсальная алопеция; гипофосфатазия взрослых; недостаточность манноз-связывающего белка; макулярная дистрофия «бычий глаз»; аутосомно-доминантная торсионная дистония 4; нефротический синдром, тип 3, тип 5, с глазными аномалиями или без них, тип 7 и тип 9; судороги, ранняя детская эпилептическая энцефалопатия 7;стойкая гиперинсулинемическая гипогликемия грудного возраста; тромбоцитопения, сцепленная с X-хромосомой; неонатальная гипотония; синдром Орставика-Линдемана-Сольберга; первичная легочная гипертензия, 1, с наследственными геморрагическими телеангиэктазиями; гипофизарно-зависимый гиперкортицизм; бородавчатая эпидермодисплазия; буллезный эпидермолиз, узловой, локальный вариант; дефицит цитохром с оксидазы i; синдром Киндлера; миосклероз, аутосомно-рецессивный; артериальный ствол; синдром Дуэйна 2 типа; ВЗРОСЛЫЙ синдром; спектр синдрома Цельвегера; лейкоэнцефалопатия с атаксией, с поражением ствола и спинного мозга и повышением лактата, с исчезновением белого вещества и прогрессирующая, с недостаточностью яичников; дефицит антитромбина III; голопрозэнцефалия 7; синдром фокомелии Робертса- SC; синдром истощения митохондриальной ДНК 3 и 7, гепатоцеребральные типы и 13 (энцефаломиопатический тип); порэнцефалия 2; микроцефалия, нормальный интеллект и иммунодефицит; гигантская аксональная нейропатия; синдром Стерджа-Вебера, капиллярные мальформации, врожденные, 1; заболевание Фабри и заболевание Фабри, кардиальный вариант; недостаточность глутаматформиминотрансферазы; синдром Фанкони-Биккеля; акромикрическая дисплазия; эпилепсия, идиопатическая генерализованная, предрасположенность к 12; кальцификация базальных ганглиев, идиопатическая, 4; миопатия с полиглюкозановыми тельцами 1 с иммунодефицитом или без; злокачественная опухоль простаты; врожденная эктодермальная дисплазия лица; врожденный порок сердца; возрастная макулодистрофия 3, 6, 11 и 12; врожденная миотония, аутосомно-доминантная и рецессивная формы; гипомагниемия 1, кишечная; недостаточность сульфитоксидазы, изолированная; заболевание Пика; дефицит плазминогена, тип I; синдактилия 3 типа; колбочко-палочковая дистрофия, несовершенный амелогенез; псевдопервичный гиперальдостеронизм; терминальная костная дисплазия; антенатальный синдром Барттера

2 типа; врожденная мышечная дистрофия-дистрогликанопатия с умственной отсталостью, типы B2, B3, B5 и B15; семейная инфантильная миастения; лимфопролиферативный синдром 1, 1 (сцепленный с 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; дисплазминогенемия; семейный колоректальный рак; спастическая атаксия 5, аутосомно-рецессивная, типа Шарлевуа-Сагене, 1, 10 или 11, аутосомно-рецессивная; лобно-метафизарная дисплазия 3; наследственное заболевание недостаточности факторов II, IX, VIII; спондилохейродисплазия, по типу синдрома Элерса-Данлоса, с иммунодефицитом, Аггреканского типа, с врожденными вывихами суставов, укороченно-кистевого типа, Седагатского типа, с колбочко-палочковой дистрофией, и Козловского типа; синдром ихтиоза недоношенных; синдром Стиклера 1 типа; очаговый сегментарный гломерулосклероз 5; дефицит 5-оксопролиназы; синдромальная микрофтальмия 5, 7 и 9; ювенильный полипоз/синдром наследственной геморрагической телеангиэктазии; недостаточность бутирил-КоА-дегидрогеназы; сахарный диабет взрослого возраста у молодых, тип 2; уственная отсталость, синдромальная, типа Класа-Йенсена, сцепленная с 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; гипериммуноглобулинеимя D с периодической лихорадкой; предрасположенность к злокачественной гипертермии 1 типа; трихомегалия с умственной отсталостью, карликовостью и пигментной дистрофией сетчатки; аденокарцинома молочной железы; недостаточность фактора комплемента B; врожденная мышечная дистрофия Ульриха; гипертрабекулярность левого желудочка; заболевание рыбьего глаза; финский врожденный нефротический синдром; поясно-конечностная мышечная дистрофия, тип IB, 2A, 2B, 2D, Cl, C5, C9, C14; идиопатический фиброзирующий альвеолит, хроническая форма; первичная семейная гипертрофическая кардиомиопатия; доброкачественное увеличение в сыворотке ангиотензин i-превращающего фермента; недостаточность Cd8, семейная; синдром Протея; дефект транспорта глюкозо-6-фосфата; синдром Борджесона-Форсмана-Лемана; синдром Зельвегера; спинальная мышечная

атрофия II типа; рак предстательной железы, наследственный, 2; тромбоцитопения, дисфункция тромбоцитов, гемолиз и нарушение синтеза глобина; врожденные нарушения гликозилирования типов IB, ID, 1G, 1H, 1 J, IK, IN, IP, 2C, 2J, 2K, Ilm; пограничный буллезный эпидермолиз Герлица; генерализованная эпилепсия с фебрильными приступами плюс 3, тип 1, тип 2; шизофрения 4; ишемическая болезнь сердца, аутосомно-доминантная 2; врожденный дискератоз, аутосомно-доминантный, 2 и 5; подкорковая ламинарная гетеротопия, сцепленная с Х-хромосомой; недостаточность аденилаткиназы; Х-сцепленный тяжелый комбинированный иммунодефицит; копропорфирия; амилоидная кардиомиопатия, связанная с транстиретином; гипокальциемия, аутосомно-доминантная 1; синдром Бругада; врожденный миастенический синдром, чувствительный к ацетазоламиду; первичная гипомагниемия; склеростеоз; лобно-височная деменция и/или боковой амиотрофический склероз 3 и 4; мевалоновая ацидурия; шванноматоз 2; наследственная моторная и сенсорная нейропатия с атрофией зрительного нерва; поздняя кожная порфирия; рассекающий остеохондрит; судороги, доброкачественные семейные неонатальные, 1 и/или миокимия; синдром удлиненного интервала QT, подтип LQT1; умственная отсталость, передняя протрузия верхней челюсти и косоглазие; идиопатическая гиперкальциемия грудного возраста; гипогонадотропный гипогонадизм 11 с аносмией или без; поликистозная липомембранозная остеодисплазия со склерозирующей лейкоэнцефалопатией; первичная аутосомно-рецессивная микроцефалия 10, 2, 3 и 5; прерванная дуга аорты; врожденная амегакариоцитарная тромбоцитопения; синдром Германского-Пудлака 1, 3, 4 и 6; синдром удлиненного интервала QT 1, 2, 2/9, 2/5, (дигенный), 3, 5 и 5, приобретенный, предрасположенность к; синдром Андермана; дистрофия колбочек сетчатки 3Б; эритропоэтическая протопорфирия; недостаточность сепиаптеринредуктазы; недостаточность ацил-КоА-дегидрогеназы с очень длинной углеродной цепью; гиперферритинемический катарактальный синдром; синдром спастической параплегии Сильвера; болезнь Шарко-Мари-Тута; дефект межпредсердной перегородки 2; синдром Карнавала; врожденная нечувствительность к боли с ангидрозом; катехоламинергическая полиморфная желудочковая тахикардия; гипокалиемический периодический паралич 1 и 2 степени; синдром внезапной детской смерти; гипохромная микроцитарная анемия с перегрузкой железом; синдром дефицита GLUT1 2; лейкодистрофия, гипомиелинизирующая, 11 и 6; колбочковый монохроматизм; остеопетроз аутосомно-доминантный 1 и 2 типа, рецессивный 4, рецессивный 1, рецессивный 6; тяжелая врожденная нейтропения 3, аутосомно-рецессивная или доминантная; недостаточность метионин-аденозилтрансферазы, аутосомно-доминантная; пароксизмальная семейная фибрилляция желудочков; недосаточность пируваткиназы эритроцитов; дисплазия Шнекенбекена; пируэтная желудочковая тахикардия; дистальная миопатия типа Марксбери-Григгса; недостаточность УДФ-глюкозо-гексозо-1-фосфат уридилилтрансферазы; внезапная сердечная смерть; синдром Ней-Лаксова 1; атрансферринемия; гиперпаратиреоз 1 и 2; злокачественная меланома кожи 1; симфалангизм, проксимальный, lb; прогрессирующая псевдоревматоидная дисплазия;

болезнь Верднига-Гоффмана; ахондрогенез 2 типа; голопрозэнцефалия 2, 3, 7 и 9; хаболевание Шиндлера, тип 1; цереброретинальная микроангиопатия с кальцинатами и кистами; гетеротаксия, висцеральная, сцепленная с Х-хромосомой; синдром туберозного склероза; синдром Картагенера; резистентность к гормонам щитовидной железы, генерализованная, аутосомно-доминантная; бестрофинопатия, аутосомно-рецессивная; поражение ногтей, несиндромальное врожденное, 8; синдром Мора-Транебьерга; колбочко-палочковая дистрофия 12; нарушение слуха; овариолейкодистрофия; почечный канальцевый ацидоз, проксимальный, с глазными аномалиями и умственной отсталостью; недостаточность дигидроптеридинредуктазы; фокальная эпилепсия с нарушением речи с задержкой психического развития или без; синдром атаксии-телеангиэктазии; синдром Брауна-Виалетто-Ван Лаэра и синдром Брауна-Виалетто-Ван Лаэра 2; кардиомиопатия; периферическая демиелинизирующая нейропатия, центральная дисмиелинизация; эндотелиальная дистрофия роговицы Фукса 4; синдром Каудена 3; дистония 2 (торсионная, аутосомно-рецессивная), 3 (торсионная, Х-сцепленная), 5 (допа-чувствительный тип), 10, 12, 16, 25, 26 (миоклоническая); эпифизарная дисплазия, множественная, с миопией и кондуктивной глухотой; неспецифическое нарушение проводимости сердца; бранхиотические синдромы 2 и 3; нарушение биогенеза пероксисом 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 Х-хромосомой; синдром истощения митохондриальной ДНК 11, 12 (кардиомиопатический тип), 2, 4В (тип MNGIE), 8В (тип MNGIE); брахидактилия при гипертонии; плоская роговица 2; синдром Аарского; множественная эпифизарная дисплазия 5 или доминантная; эндотелиальная дистрофия роговицы 2 типа; недостаточность аминоацилазы 1; задержка речевого и языкового развития; синдром Николаидеса-Барайцера; недостаточность энтерокиназы; эктродактилия, эктодермальная

дисплазия и синдром расщелины губы/неба 3; множественный врожденный артрогрипоз, дистальный, сцепленный с Х-хромосомой; синдром Перро 4; синдром Джервелла и Ланге-Нильсена 2; наследственные неполипозные колоректальные новообразования; синдром Робинова, аутосомно-рецессивный, аутосомно-рецессивный, с брахи-син-полидактилией; нейрофибросаркома; недостаточность цитохром-с-оксидазы; пузырно-мочеточниковый рефлюкс 8; недостаточность дофамин-бета-гидроксилазы; синдром углеводдефицитных гликопротеинов I и II типов; прогрессирующий семейный внутрипеченочный холестаз 3; доброкачественные семейные неонатально-инфантильные судороги; панкреатит, хронический, восприимчивость к; ризомелическая точечная хондродисплазия 2 и 3 типа; нарушение стероидогенеза из-за недостаточности цитохрома p450 оксидоредуктазы; глухота с лабиринтной аплазией, микротией и микродонтией (FAMM); синдром Ротмунда-Томсона; комплексная кортикальная дисплазия с другими пороками развития головного мозга 5 и 6; миастения, семейная инфантильная, 1; трихоринофаланговая дисплазия I типа; заболевание Worth; гипоплазия селезенки; недостаточность кофактора молибдена, группа комплементации A; синдром Себастьяна; прогрессирующий семейный внутрипеченочный холестаз 2 и 3; синдром Вейля-Маркезани 1 и 3; микроцефалическая остеодиспластическая примордиальная карликовость 2 типа; нарушение обмена сурфактанта, легочное, 2 и 3; тяжелая Х-сцепленная миотубулярная миопатия; рак поджелудочной железы 3; кровотечения тромбоцитарного типа 15 и 8; тирозиназо-положительный глазо-кожный альбинизм; синдром Борроне-ди-Рокко-Кровато; синдром ATR-X; недостаточность сахаразы-изомальтазы; частичная недостаточность компонента комплемента 4 в следствии дисфункционального ингибитора cl; врожденная центральная гиповентиляция; инфантильная гипофосфатазия; недостаточность ингибитора активатора плазминогена 1 типа; злокачественная лимфома, неходжкинская; синдром гиперорнитинемии-гипераммониемии-гомоцитруллинурии; синдром Шварца-Джампеля 1 типа; локус 1 количественного признака фетального гемоглобина; миопатия дистальная с началом в передней большеберцовой кости; синдром Нунана 1 и 4, синдром LEOPARD 1; глаукома 1, открытоугольная, e, F и G; синдром Кенни-Кэффи 2 типа; PTEN-ассоциированный синдром гамартомных опухолей; мышечная дистрофия Дюшенна; инсулинорезистентный сахарный диабет и черный акантоз; микрофтальм изолированный 3, 5, 6, 8 и при колобоме 6; синдром Рейна; преждевременная недостаточность яичников 4, 5, 7 и 9; синдром Аллана-Хемдона-Дадли; цитруллинемия I типа; болезнь Альцгеймера, семейная, 3, со спастическими парапарезами и апраксией; семейная гемиплегическая мигрень 1 и 2 типов; вентрикуломегалия при кистозной болезни почек; эластическая псевдоксантома; гомоцистеинемия в следствии недостаточности MTHFR, недостаточности CBS и гомоцистинурия, реагирующая на пиридоксин; дилатационная кардиомиопатия 1A, 1AA, 1C, 1G, IBB, 1DD, IFF, 1HH, II, IKK, IN, IS, 1Y и 3B; недостаточность гуаниноксидазы АМФ в мышцах; семейный рак молочной железы; наследственная сидеробластная анемия; миоглобинурия, острая рецидивирующая, аутосомно-рецессивная; нейроферритинопатия; аритмия сердца; синдром дефицита переносчика глюкозы 1 типа; последовательность,

приводящая к голопрозэнцефалии; ангиопатия наследственная с нефропатией, аневризмами и мышечными спазмами; недостаточность изовалерил-КоА-дегидрогеназы; синдром Каллмана 1, 2 и 6; постоянный неонатальный сахарный диабет; акрокаллезный синдром по типу Шинцеля; синдром Гордона; расстройства, связанные с MYH9; синдром Доннея Барроу; тяжелая врожденная нейтропения и 6, аутосомно-рецессивная; заболевание Шарко-Мари-Тута, типы ID и IVF; синдром Коффина-Лоури; недостаточность митохондриальной 3-гидрокси-3-метилглутарил-КоА-синтазы; гипомагниемия, судороги и умственная отсталость; ишиопателлярная дисплазия; множественные врожденные аномалии - гипотония - судорожный синдром 3; спастическая параплегия 50, аутосомно-рецессивная; низкорослость с неспецифическими аномалиями скелета; тяжелая миоклоническая эпилепсия в младенчестве; пропионовая ацидемия; подростковый нефронофтиз; макроцефалия, макросомия, синдром лицевого дисморфизма; заболевание Штаргардта 4; синдром Элерса-Данлоса 7-го типа (аутосомно-рецессивный), классический тип, 2-й тип (прогероидный), гидроксилизин-дефицитный, 4-й тип, вариант 4-го типа и обусловлен дефицитом тенасцина-X; миопия 6; плоский таз; семейный холодовой аутовоспалительный синдром 2; пороки развития сердца и крупных сосудов; болезнь фон Виллебранда типа 2M и типа 3; недостаточность галактокиназы; синдром Бругада 1; X-сцепленный ихтиоз с недостаточностью стерилсульфатазы; врожденная колобома глаза; гистиоцитоз-лимфаденопатия плюс синдром; аниридия, мозжечковая атаксия и умственная отсталость; гипертрабекулярность левого желудочка 3; боковой амиотрофический склероз типов 1, 6, 15 (с лобно-височной деменцией или без), 22 (с лобно-височной деменцией или без) и 10; несовершенный остеогенез тип 12, тип 5, тип 7, тип 8, тип I, тип III, с нормальными склерой, доминантная форма, рецессивный перинатальный летальный; гематологическое новообразование; фавизм, подверженность; легочный фиброз и/или недостаточность костного мозга, связанные с теломерами, 1 и 3; доминантная наследственная атрофия зрительного нерва; доминантный дистрофический буллезный эпидермолиз с отсутствием кожи; мышечная дистрофия врожденная, мегакониального типа; множественные атрезии желудочно-кишечного тракта; синдром МакКьюна-Олбрайта; синдром ногтя-надколенника; синдром нейроакантоцитоза Мак-Леода; системный вариабельный иммунодефицит 9; частичная недостаточность гипоксантин-гуанинфосфорибозилтрансферазы; псевдогипоальдостеронизм 1 типа аутосомно-доминантный и рецессивный и 2 тип; недостаточность уроканатгидратазы; гетеротопия; синдром Меккеля 7 типа; Ch\xc3\xa9diak-Higashi синдром, синдром Чедиака-Хигаси, взрослый тип; тяжелый комбинированный иммунодефицит в следствии недостаточности ADA, с микроцефалией, задержкой роста и чувствительностью к ионизирующему излучению, атипичный, аутосомно-рецессивный, Т-клеточный отрицательный, В-клеточный положительный, NK-клеточный отрицательный или NK-положительный; резистентность к инсулину; недостаточность стероид-11-бета-монооксигеназы; синдром подколенного птеригиума; легочная артериальная гипертензия, связанная с наследственными геморрагическими телеангиэктазиями; глухота, аутосомно-рецессивная

1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86 и 89; Первичная гипероксалурия, тип I, тип и тип III; врожденная парамиотония фон Эйленбурга; синдром Дебюкуа; недостаточность карнитинпальмитоилтрансферазы I, II, II (позднее начало) и II (инфантильный); вторичный гипотиреоз; нижнечелюстно-лицевой дизостоз, по типу Treacher Collins, аутосомно-рецессивный; синдром Каудена 1; синдром Ли-Фраумени 1; недостаточность аспарагинсинтетазы; «сотовидная» ретинальная дистрофия Дойна; атрофия зрительного нерва 9; инфантильные судороги и пароксизмальный хореоатетоз, семейные; атаксия при дефиците витамина Е; гиперплазия островковых клеток; мышечная дистрофия Миоши 1; тромбофилия, наследственная, обусловленная дефицитом протеина С, аутосомно-доминантная и рецессивная; синдром Фехтнера; недостаточность пропердина, сцепленная с Х-хромосомой; умственная отсталость, стереотипные движения, эпилепсия и/или церебральные пороки развития; недостаточность креатина, Х-сцепленная; пиломатриксома; цианоз, преходящий неонатальный и атипичный нефропатический; взрослая атаксия с окуломоторной апраксией; гемангиома, капиллярная инфантильная; PC-K6a; генерализованный доминантный дистрофический буллезный эпидермолиз; болезнь Пелизеуса-Мерцбахера; центроядерная миопатия 1, врожденная с избытком мышечных веретен, дистальная 1, лактоацидоз и сидеробластная анемия 1, митохондриальная прогрессирующая с врожденной катарактой, тугоухостью, задержкой развития и тубулярным агрегатом 2; доброкачественные семейные неонатальные судороги 1 и 2; первичная легочная гипертензия; лимфедема первичная с миелодисплазией; врожденный синдром удлиненного интервала QT; семейная экссудативная витреоретинопатия, сцепленная с Х-хромосомой; аутосомно-доминантная гипогидротическая эктодермальная дисплазия; примордиальная карликовость; семейный легочный капиллярный гемангиоматоз; недостаточность карнитин-ацилкамитин-транслоказы; висцеральная миопатия; семейная средиземноморская лихорадка и семейная средиземноморская лихорадка, наследуемая по аутосомно-доминантному типу; комбинированная частичная и полная недостаточность 17-альфа-гидроксилазы/17, 20-лиазы; ухо-небо-пальцевый синдром I типа; нефролитиаз/остеопороз, гипофосфатемия, 2; Семейная гиперлипопротеинемия 1 и 3 типа; фенотипы; CHARGE-синдром; синдром Фурманна; синдром гипотрихоз-лимфедемы-телеангиэктазии; хондродисплазия типа Бломстранда; акроэритрокератодермия; замедленная скорость нервной проводимости, аутосомно-доминантная; наследственный синдром, предрасполагающий к раку; краниодиафизарная дисплазия, аутосомно-доминантная; спиноцеребеллярная атаксия аутосомно-рецессивная 1 и 16; недостаточность пропротеинконвертазы 1/3; D-2-гидроксиглутаровая ацидурия 2; гиперэкплексия 2 и наследственная гиперэкплексия; миопатия центрального стержня; синдром Опица G/BBB; муковисцидоз; дистрофия боуменовой мембраны II типа Тиля-Бенке; недостаточность бисфосфоглицератмутазы; недостаточность митохондриальной короткоцепочечной эноил-КоА-гидратазы 1; синдром ломкости кожи при эктодермальной дисплазии; вольфрамоподобный синдром, аутосомно-доминантный; микроцитарная анемия; недостаточность пируваткарбоксилазы; недостаточность адгезии лейкоцитов I и III

типов; множественная эндокринная неоплазия, тип 4; Транзиторный буллезный дермолиз новорожденных; синдром Примулы; немелкоклеточный рак легкого; врожденная мышечная дистрофия; комбинированная недостаточность липазы; СИНДРОМ КОУЛА-КАРПЕНТЕРА 2; дефект атриовентрикулярной перегородки и общий атриовентрикулярный переход; недостаточность ксантиноксидазы; синдром Ваарденбурга типа 1, 4C и 2E (с неврологическим поражением); синдром Стиклера, типы I (несиндромальный глазной) и 4; синдром ломкой роговицы, голубых склер и гипермобильность суставов; микросферофакия; синдром Чадли-Маккаллоу; эпидермолизная буллезная симплексная и поясно-конечностная мышечная дистрофия, симплексная с пятнистой пигментацией, симплексная с атрезией привратника, симплексная, аутосомно-рецессивная и с атрезией привратника; расстройство Ретта; нарушение миграции нейронов; недостаточность гормона роста с аномалиями гипофиза; болезнь Ли; ладонно-подошвенный полосатый кератоз 1; синдром Вейсенбахера-Цвеймюллера; недостаточность ацил-кофермента А дегидрогеназы со средней длиной цепи; недостаточность УДФ-глюкозо-4-эпимеразы; предрасположенность к аутизму, сцепленному с Х-хромосомой 3; регматогенная отслойка сетчатки, аутосомно-доминантная; семейные фебрильные судороги 8; отсутствие локтевой и малоберцовой кости при тяжелом дефиците конечностей; гипертрабекулярность левого желудочка 6; центромерная нестабильность хромосом 1,9 и 16 и иммунодефицит; наследственная диффузная лейкоэнцефалопатия со сфероидами; синдром Кушинга; сниженная плотность допаминовых рецепторов d2 в головном мозге; 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 Х-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 Х-хромосомой), 3, 6 (с микроцефалией), сцепленная с Х-хромосомой; соматотрофная аденома; синдром Гамсторпа-Вольфарта; липидный протеиноз; миопатия с тельцами включения 2 и 3; синдром увеличенного вестибулярного водопровода; остеопороз с псевдоглиомой; приобретенный синдром удлиненного интервала QT; фенилкетонурия; CHOPS-синдром; глобальная задержка развития; кристаллическая корнеоретинальная дистрофия Биетти; расстройство, подобное синдрому Нунана, с ювенильным миеломоноцитарным лейкозом или без; врожденная эритропоэтическая порфирия; наследственная бульбовидная атрофия; параганглиомы 3; синдром Ван дер Вуда; недостаточность ароматазы; синдром дисморфизма умственной отсталости Бирка Барела; боковой амиотрофический склероз 5 типа; метгемоглобинемия I 1 и 2 типов; врожденная стационарная куриная слепота типов 1A, IB, 1C, IE, IF и 2A; судороги; рак щитовидной железы, фолликулярный; синдром летальной врожденной контрактуры 6; дистальная наследственная моторная нейронопатия типа 2B; стромальная опухоль полового тяжа; эпилептическая энцефалопатия с началом в детском возрасте, ранняя инфантильная, 1, 19, 23, 25, 30 и 32; миофибриллярная миопатия 1 и связанная с ZASP; детская мозжечковая атаксия с прогрессирующей наружной офтальмоплегией; недостаточность пуриннуклеозидфосфорилазы; дефекты переднего мозга; эпилептическая энцефалопатия типа Леннокса-Гасто; ожирение; 4, гипертрабекулярность левого желудочка 10; синдром Верхейя; синдром Мовата-Уилсона; одонтотрихомелический синдром; узорчатая дистрофия пигментного эпителия сетчатки; синдром Lig4; синдром Бараката; недостаточность IRAK4; соматотрофная аденома; недостаточность киназы дегидрогеназы кетокислот с разветвленной цепью; цистинурия; семейная аплазия червоподобного отростка; недостаточность сукцинил-КоА-ацетоацетаттрансферазы; лопаточно-перонеальная спинальная мышечная атрофия; пигментная дистрофия сетчатки; тромбастения Гланцмана; первичная открытоугольная глаукома с началом в юношеском возрасте 1; синдромы Айкарди-Гутьера 1, 4 и 5; почечная дисплазия; задержка внутриутробного развития, метафизарная дисплазия, врожденная гипоплазия надпочечников и генитальные аномалии; монилетрикс; низкий рост, ониходисплазия, лицевой дисморфизм и гипотрихоз; метахроматическая лейкодистрофия; недостаточность лизосомной кислой липазы; синдром 3 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 apob32; синдром Коккейна типа A; тяжелый неонатальный гиперпаратиреоз; расстройство, подобное атаксии-телеангиэктазии; синдром Пендреда; I система групп крови; семейная доброкачественная пузырчатка; висцеральная гетеротаксия 5, аутосомная; нефрогенный несахарный диабет, нефрогенный несахарный диабет, X-сцепленный; миопатия с мини-ядрами и наружной офтальмоплегией; синдром Перри; гипогидротический/волосистый/зубной тип, аутосомно-рецессивный; наследственный панкреатит; умственная отсталость и микроцефалия с гипоплазией моста и мозжечка; заболевание накопления гликогена 0 (мышечная), II (взрослая форма), IXa2, IXc, тип 1A; полосчатая остеопатия с черепным склерозом; недостаточность глутатионсинтетазы; синдром Бругада и синдром Бругада 4; эндометриальная карцинома; гипогидротическая эктодермальная дисплазия с иммунодефицитом; внутрипеченочный холестаз беременных 3; синдром Бемара-Сулье, типы A1 и A2 (аутосомно-доминантный); болезнь Салла; недостаточность орнитинаминотрансферазы; синдром опухоли гамартомы PTEN; дистихиаз-синдром лимфедемы; недостаточность кортикостероидсвязывающего глобулина; нейрональный цероидный липофусциноз взрослых; болезнь Дежерина-Сотта; тетраамелия, аутосомно-рецессивная; синдром Сениора-Локена 4 и 5, ; глутаровая ацидемия IIA и IIB; аневризма аорты, семейная грудная 4, 6 и 9; гиперфосфатазия с синдромом умственной отсталости 2, 3 и 4; врожденный дискератоз, сцепленный с X-хромосомой; артрогрипоз, почечная дисфункция и холестаз 2; синдром Баннаяна-Райли-Рувалькабы; 3- метилглутаконовая ацидурия; изолированная недостаточность 17,20-лиазы; синдром Горлина; HFGS; заболевание Тея-Сакса, вариант B1, Gm2-ганглиозидоз (у взрослых), Gm2-ганглиозидоз (начало у взрослых); заболевание Даулинга-Дего 4; заболевание Паркинсона 14, 15, 19 (юношеское начало), 2, 20 (раннее начало), 6, (аутосомно-рецессивное раннее начало и 9; сенсорная, аутосомно-доминантная атаксия; врожденная микроворсинчатая атрофия; миоклонико-атоническая эпилепсия; заболевание Танжера; 2-метил-3-гидроксимасляная ацидурия; семейная почечная гипоурикемия; шизэнцефалия; синдром истощения митохондриальной ДНК 4B, тип MNGIE; синдром Фейнгольда 1; дефект почечного транспорта карнитина; семейная гиперхолестеринемия; Таунса-Брокса-бранхио-ото-реноподобный синдром; синдром Гришелли тип 3; синдром Меккеля-Грубера; буллезная ихтиозиформная эритродермия; синдром нейтрофильного иммунодефицита; миастенический синдром, врожденный, 17, 2A (медленный канал), 4B (быстрый канал) и без канальцевых агрегатов; микрососудистые осложнения сахарного диабета 7; синдром МакКьюсика Кауфмана; хроническое гранулематозное заболевание, аутосомно-рецессивное, цитохром b-положительное, типы 1 и 2; недостаточность аргининосукцинатлиазы; недостаточность переносчиков митохондриального фосфата и пирувата; решетчатая l дистрофия роговицы III типа; синдром эктодермальной дисплазии-синдактилии 1; гипомиелинизирующая лейкодистрофия 7; умственная отсталость, аутосомно-доминантная 12, 13, 15, 24, 3, 30, 4, 5, 6 и 9; генерализованная эпилепсия с

фебрильными судорогами плюс 1 и 2 типы; восприимчивость к псориазу 2; синдром Франка Тер Хаара; аневризмы грудной аорты и расслоение аорты; синдром Крузона; гранулезоклеточная опухоль яичника; эпидермолитическая ладонно-подошвенная кератодермия; дисхондростеоз Лери Вейля; недостаточность 3-бета-гидроксистероиддегидрогеназы; семейная рестриктивная кардиомиопатия 1; аутосомно-доминантная прогрессирующая наружная офтальмоплегия с делециями 1 и 3 митохондриальной ДНК; синдром Антли-Бикслера с аномалиями половых органов и нарушением стероидогенеза; синдром Хайду-Чейни; пигментно-узелковое заболевание коры надпочечников, первичное, 1; эпизодический болевой синдром, семейный, 3; синдром Дежерина-Сотта, аутосомно-доминантный; FG синдром и FG синдром 4; недостаточность дендритных клеток, моноцитов, B-лимфоцитов и природных лимфоцитов-киллеров; гипотиреоз врожденный, незобный, 1; синдром Миллера; немалиновая миопатия 3 и 9; олигодонтия - синдром колоректального рака; синдром холодового потоотделения 1; болезнь Ван Бухема 2 типа; глаукома 3, первичная врожденная, d; цитруллинемия I и II типа; миопатия Нонака; врожденная мышечная дистрофия вследствие частичного дефицита LAMA2; синдром мионевральной желудочно-кишечной энцефалопатии; синдром Ли из-за недостаточности митохондриального комплекса I; медуллобластома; недостаточность пируватдегидрогеназы El-альфа; рак толстой кишки; синдром Нанса-Хорана; заболевание Сандхоффа, взрослый и детский типы; артрогрипоз, почечная дисфункция, синдром холестаза; аутосомно-рецессивная гипофосфатемическая болезнь костей; «Сотовидная» ретинальная дистрофия Дойна; спиноцеребеллярная атаксия 14, 21, 35, 40 и 6; деменция с тельцами Леви; митохондриальное заболевание, связанное с RRM2B; Броуди миопатия; синдром мегалэнцефалии-полимикрогирии-полидактилии-гидроцефалии 2; синдром Ушера, типы 1, IB, ID, 1G, 2A, 2C и 2D; тип гипокальцификации и тип гипоматурации, IIA1 несовершенный амелогенез; недостаточность гормонов гипофиза, сочетание 1, 2, 3 и 4; Симфалангизм Кушинга; почечный тубулярный ацидоз, дистальный, аутосомно-рецессивный, с поздней сенсоневральной тугоухостью или с гемолитической анемией; детский нефронофтиз; синдром ювенильного полипоза; сенсорная атаксическая невропатия, дизартрия и офтальмопарез; недостаточность 3-гидроксиацил-КоА-дегидрогеназы; карцинома паращитовидной железы; 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 птеригиумов, синдром Эскобара; гомоцистинурия-мегалобластная анемия вследствие дефекта метаболизма кобаламина, комплементарный тип cblE; холецистит; сфероцитоз 4 и 5 типов; множественные врожденные аномалии; пигментная ксеродерма, комплементарная группа b, группа D, группа E и группа G; болезнь Лейнера; комеальная дистрофия Гроеноу I типа;

недостаточность коэнзима Q10, первичные 1, 4 и 7; врожденная непрогрессирующая дистальная спинальная мышечная атрофия; микросиндром Варбурга 2 и 4; дефект синтеза желчных кислот, врожденный, 3; АКТГ-независимая макронодулярная гиперплазия надпочечников 2; акро-капито-феморальная дисплазия; заболевание костей Педжета, семейное; тяжелая неонатальная энцефалопатия с микроцефалией; синдром Циммерманна-Лабанда и синдром Циммерманна-Лабанда 2; синдром Райфенштейна; семейная гипокалиемия-гипомагниемия; светочувствительная трихотиодистрофия; узловой буллезный эпидермолиз у взрослых; рак легких; синдром Фримена-Шелдона; синдром гиперинсулинизма-гипераммониемии; задняя полярная катаракта 2 типа; Склерокорнеа, аутосомно-рецессивный; ювенильный GM >1< ганглиозидоз; синдром Коэна; наследственные синдромы параганглиомы-феохромоцитомы; неонатальный инсулинозависимый сахарный диабет; гипохондрогенез; синдром плавучей гавани; эластолизис с остеодистрофией и тяжелыми легочными, желудочно-кишечными и мочевыми нарушениями; врожденные контрактуры конечностей и лица, гипотония и задержка развития; врожденный дискератоз аутосомно-доминантный и аутосомно-доминантный, 3; гистиоцитарный медуллярный ретикулез; синдром Костелло; иммунодефицит 15, 16, 19, 30, 31C, 38, 40, 8, обусловленный дефектом cd3-дзета, с гипериммуноглобулинемией M 1 и 2 и X-сцепленным наследованием, с дефектом магния, вирусом Эпштейна-Барра и неоплазией ; дефекты межпредсердной перегородки 2, 4 и 7 (с дефектами атриовентрикулярной проводимости или без них); недостаточность GTP-циклогидролазы I; эквиноварусная деформация стопы; недостаточность фосфоглицераткиназы 1; туберозный склероз 1 и 2; аутосомно-рецессивный врожденный ихтиоз 1, 2, 3, 4A и 4B; и семейная гипертрофическая кардиомиопатия 1, 2, 3, 4, 7, 10, 23 и 24.

Показания по ткани

Дополнительные подходящие заболевания и нарушения, которые можно лечить с помощью систем и способов, представленных в данном документе, включают без ограничения заболевания центральной нервной системы (ЦНС) (см. иллюстративные заболевания и пораженные гены в **таблице 13**), заболевания глаз (см. иллюстративные заболевания и пораженные гены в **таблице 14**), заболевания сердца (см. иллюстративные заболевания и пораженные гены в **таблице 15**), заболевания гемопоэтических стволовых клеток (HSC) (см. иллюстративные заболевания и пораженные гены в **таблице 16**), заболевания почек (см. иллюстративные заболевания и пораженные гены в **таблице 17**), заболевания печени (см. иллюстративные заболевания и пораженные гены в **таблице 18**), заболевания легких (см. иллюстративные заболевания и пораженные гены в **таблице 19**), заболевания скелетных мышц (см. иллюстративные заболевания и пораженные гены в **таблице 20**) и заболевания кожи (см. иллюстративные заболевания и пораженные гены в **таблице 21**). В **таблице 22** представлены иллюстративные защитные мутации, которые снижают риск указанных заболеваний. В некоторых вариантах осуществления система Gene Writer, описанная в данном документе, применяется для лечения показаний,

указанных в любой из таблиц 13-21. В некоторых вариантах осуществления система Gene Writer модифицирует сайт-мишень в геномной ДНК в клетке, где сайт-мишень находится в гене, указанном в любой из таблиц 13-21, например, у субъекта, имеющего соответствующие показания, указанные в любой из таблиц 13-21. В некоторых вариантах осуществления Gene Writer исправляет мутацию в гене. В некоторых вариантах осуществления Gene Writer вставляет последовательность, которая была удалена из гена (например, вследствие мутации, вызывающей заболевание). В некоторых вариантах осуществления Gene Writer удаляет последовательность, которая была дуплицирована в гене (например, вследствие мутации, вызывающей заболевание). В некоторых вариантах осуществления Gene Writer заменяет мутацию (например, мутацию, вызывающую заболевание) соответствующей последовательностью дикого типа. В некоторых вариантах осуществления мутация представляет собой замену, вставку, делецию или инверсию.

Таблица 13. Заболевания ЦНС и пораженные гены

Заболевание	**Пораженный ген**
Альфа-маннозидоз	MAN2B1
Атаксия-телеангиэктазия	ATM
CADASIL	NOTCH3
Болезнь Канавана	ASPA
Дефицит карбамоилфосфатсинтетазы 1	CPS1
Заболевание CLN1	PPT1
Заболевание CLN2	TPP1
Заболевание CLN3 (ювенильный нейрональный цероидный липофусциноз, болезнь Баттена)	CLN3
Синдром Коффина-Лоури	RPS6KA3
Врожденный миастенический синдром 5	COLQ
Синдром Корнелии де Ланге (NIPBL)	NIPBL
Синдром Корнелии де Ланге (SMC1A)	SMC1A
Синдром Драве (SCN1A)	SCN1A
Глициновая энцефалопатия (GLDC)	GLDC
GM1-ганглиозидоз	GLB1
Болезнь Хантингтона	HTT
Гидроцефалия со стенозом сильвиева водопровода	L1CAM
Синдром Ли	SURF1

Метахроматическая лейкодистрофия (ARSA)	ARSA
MPS 2 типа	IDS
MPS 3 типа	Тип 3a: SGSH Тип 3b: NAGLU
Муколипидоз IV типа	MCOLN1
Нейрофиброматоз 1 типа	NF1
Нейрофиброматоз 2 типа	NF2
Нейродегенерация, ассоциированная с пантотенаткиназой	PANK2
Пиридоксин-зависимая эпилепсия	ALDH7A1
Синдром Ретта (MECP2)	MECP2
Болезнь Сандхоффа	HEXB
Семантическая деменция (лобно-височная деменция)	MAPT
Спиноцеребеллярная атаксия с аксональной нейропатией (атаксия с окуломоторной апраксией)	SETX
Болезнь Тея-Сакса	HEXA
Х-сцепленная адренолейкодистрофия	ABCD1

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

Заболевание	Пораженный ген
Пигментный ретинит (AD)	RHO
Болезнь Штаргардта	ABCA4
Вителлиформная макулярная дистрофия	BEST1, PRPH2

Таблица 15. Заболевания сердца и пораженные гены

Заболевание	Пораженный ген
Аритмогенная кардиомиопатия правого желудочка (ARVC)	PKP2
Синдром Барта	TAZ
Мышечная дистрофия Беккера	DMD
Синдром Бругада	SCN5A
Катехоламинергическая полиморфная желудочковая тахикардия (RYR2)	RYR2
Дилатационная кардиомиопатия (LMNA)	LMNA
Дилатационная кардиомиопатия (TTN)	TTN
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса I типа	EMD
Семейная гипертрофическая кардиомиопатия	MYH7
Семейная гипертрофическая кардиомиопатия	MYBPC3
Синдром Джервелла-Ланге-Нильсена	KCNQ1
Дефицит LCHAD	HADHA
Поясно-конечностная мышечная дистрофия типа 1B (Эмери-Дрейфуса EDMD2)	LMNA
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Синдром удлиненного интервала QT 1 типа (Романо-Уорда)	KCNQ1
Заболевание	**Пораженный ген**
Аритмогенная кардиомиопатия правого желудочка (ARVC)	PKP2
Синдром Барта	TAZ
Мышечная дистрофия Беккера	DMD
Синдром Бругада	SCN5A

Катехоламинергическая полиморфная желудочковая тахикардия (RYR2)	RYR2
Дилатационная кардиомиопатия (LMNA)	LMNA
Дилатационная кардиомиопатия (TTN)	TTN
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса I типа	EMD
Семейная гипертрофическая кардиомиопатия	MYH7
Семейная гипертрофическая кардиомиопатия	MYBPC3
Синдром Джервелла-Ланге-Нильсена	KCNQ1
Дефицит LCHAD	HADHA
Поясно-конечностная мышечная дистрофия типа 1B (Эмери-Дрейфуса EDMD2)	LMNA
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Синдром удлиненного интервала QT 1 типа (Романо-Уорда)	KCNQ1

Таблица 16. Заболевания HSC и пораженные гены

Заболевание	**Пораженный ген**
ADA-SCID	ADA
Адренолейкодистрофия (CALD)	ABCD1
Альфа-маннозидоз	MAN2B1
Хроническая гранулематозная болезнь	CYBB; CYBA; NCF1; NCF2; NCF4
Вариабельный неклассифицируемый иммунодефицит	TNFRSF13B
Анемия Фанкони	FANCA; FANCC; FANCG
Болезнь Гоше	GBA
Глобоидно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Гемофагоцитарный лимфогистиоцитоз	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 A и B	SMPD1
Болезнь Ниманна-Пика C	NPC1
Пароксизмальная ночная гемоглобинурия	PIGA
Болезнь Помпе	GAA
Дефицит пируваткиназы (PKD)	PKLR
Дефицит RAG 1/2 (SCID с гранулемами)	RAG1/RAG2
Тяжелая наследственная нейтропения	ELANE; HAX1
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB
Синдром Вискотта-Олдрича	WAS
X-сцепленная агаммаглобулинемия	BTK
X-сцепленный SCID	IL2RG
Заболевание	**Пораженный ген**
ADA-SCID	ADA
Адренолейкодистрофия (CALD)	ABCD1
Альфа-маннозидоз	MAN2B1
Хроническая гранулематозная болезнь	CYBB; CYBA; NCF1; NCF2; NCF4
Вариабельный неклассифицируемый иммунодефицит	TNFRSF13B
Анемия Фанкони	FANCA; FANCC; FANCG
Болезнь Гоше	GBA
Глобоидно-клеточная лейкодистрофия (болезнь Краббе)	GALC
Гемофагоцитарный лимфогистиоцитоз	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 A и B	SMPD1
Болезнь Ниманна-Пика C	NPC1
Пароксизмальная ночная гемоглобинурия	PIGA
Болезнь Помпе	GAA
Дефицит пируваткиназы (PKD)	PKLR
Дефицит RAG 1/2 (SCID с гранулемами)	RAG1/RAG2
Тяжелая наследственная нейтропения	ELANE; HAX1
Серповидно-клеточная анемия	HBB
Болезнь Тея-Сакса	HEXA
Талассемия	HBB
Синдром Вискотта-Олдрича	WAS
X-сцепленная агаммаглобулинемия	BTK
X-сцепленный SCID	IL2RG

Таблица 17. Заболевания почек и пораженные гены

Заболевание	**Пораженный ген**
Синдром Альпорта	COL4A5
Аутосомно-доминантный поликистоз почек (PKD1)	PKD1
Аутосомно-доминантный поликистоз почек (PKD2)	PDK2
Аутосомно-доминантное тубулоинтерстициальное заболевание почек (MUC1)	MUC1

Аутосомно-доминантное тубулоинтерстициальное заболевание почек (UMOD)	UMOD
Аутосомно-рецессивный поликистоз почек	PKHD1
Врожденный нефротический синдром	NPHS2
Цистиноз	CTNS

Таблица 18. Заболевания печени и пораженные гены

Заболевание	**Пораженный ген**
Острая интермиттирующая порфирия	HMBS
Синдром Алажилля	JAG1
Дефицит альфа-1-антитрипсина	SERPINA1
Дефицит карбамоилфосфатсинтетазы I	CPS1
Цитруллинемия I типа	ASS1
Синдром Криглера-Найяра	UGT1A1
Болезнь Фабри	LPL
Синдром семейной хиломикронемии	GLA
Болезнь Гоше	GBE1
GSD IV	GBA
Гем A	F8
Гем B	F9
Наследственный амилоидоз (hTTR)	TTR
Наследственный ангионевротический отек	SERPING1 (KLKB1 для CRISPR)
HoFH	LDLRAP1
Гиперхолестеринемия	PCSK9
Метилмалоновая ацидемия	MMUT
MPS II	IDS
MPS III	Тип IIIa: SGSH Тип IIIb: NAGLU Тип IIIc: HGSNAT Тип IIId: GNS
MPS IV	Тип IVA: GALNS

	Тип IVB: GLB1
MPS VI	ARSB
MSUD	Тип Ia: BCKDHA Тип Ib: BCKDHB Тип II: DBT
Дефицит OTC	OTC
Поликистозная болезнь печени	PRKCSH
Болезнь Помпе	GAA
Первичная гипероксалурия 1 типа	AGXT (HAO1 или LDHA для CRISPR)
Прогрессирующий семейный внутрипеченочный холестаз 1 типа	ATP8B1
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11
Прогрессирующий семейный внутрипеченочный холестаз 3 типа	ABCB4
Пропионовая ацидемия	PCCB; PCCA
Болезнь Вильсона Болезнь накопления гликогена типа 1a Болезнь накопления гликогена типа IIIb Изовалериановая ацидемия Болезнь Вольмана	ATP7B G6PC AGL IVD LIPA

Таблица 19. Заболевания легких и пораженные гены

Заболевание	**Пораженный ген**
Дефицит альфа-1-антитрипсина	SERPINA1
Кистозный фиброз	CFTR
Первичная цилиарная дискинезия	DNAI1
Первичная цилиарная дискинезия	DNAH5
Первичная легочная гипертензия I типа	BMPR2
Дефицит сурфактантного белка B (SP-B) (нарушение метаболизма легочного сурфактанта 1)	SFTPB

Таблица 20. Заболевания скелетных мышц и пораженные гены

Заболевание	**Пораженный ген**
Мышечная дистрофия Беккера	DMD
Миотония Беккера	CLCN1
Бетлемская миопатия	COL6A2
Центронуклеарная миопатия, Х-сцепленная (миотубулярная)	MTM1
Врожденный миастенический синдром	CHRNE
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса, AD	LMNA
Лице-лопаточно-плечевая мышечная дистрофия	Хромосомная область DUX4 - D4Z4
Гиперкалиемический периодический паралич	SCN4A
Гипокалиемический периодический паралич	CACNA1S
Поясно-конечностная мышечная дистрофия 2A	CAPN3
Поясно-конечностная мышечная дистрофия 2B	DYSF
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Мышечная дистрофия Миеси 1	DYSF
Врожденная парамиотония	SCN4A
Миотония Томсена	CLCN1
Миопатия VCP (IBMPFD) 1	VCP
Заболевание	**Пораженный ген**
Мышечная дистрофия Беккера	DMD
Миотония Беккера	CLCN1
Бетлемская миопатия	COL6A2
Центронуклеарная миопатия, Х-сцепленная (миотубулярная)	MTM1
Врожденный миастенический синдром	CHRNE
Мышечная дистрофия Дюшенна	DMD
Мышечная дистрофия Эмери-Дрейфуса, AD	LMNA

Лице-лопаточно-плечевая мышечная дистрофия	Хромосомная область DUX4 - D4Z4
Гиперкалиемический периодический паралич	SCN4A
Гипокалиемический периодический паралич	CACNA1S
Поясно-конечностная мышечная дистрофия 2A	CAPN3
Поясно-конечностная мышечная дистрофия 2B	DYSF
Поясно-конечностная мышечная дистрофия типа 2D	SGCA
Мышечная дистрофия Миеси 1	DYSF
Врожденная парамиотония	SCN4A
Миотония Томсена	CLCN1
Миопатия VCP (IBMPFD) 1	VCP

Таблица 21. Заболевания кожи и пораженные гены

Заболевание	**Пораженный ген**
Доминантная форма дистрофического буллезного эпидермолиза	COL7A1
Рецессивный дистрофический буллезный эпидермолиз (подтип Аллопо-Сименса)	COL7A1
Буллезный узловой эпидермолиз	LAMB3
Простой буллезный эпидермолиз	KRT5; KRT14
Эпидермолитический ихтиоз	KRT1; KRT10
Болезнь Хейли-Хейли	ATP2C1
Ламеллярный ихтиоз/небуллезная врожденная ихтиозиформная эритродермия (ARCI)	TGM1
Синдром Нетертона	SPINK5

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: 1622) 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

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).

Таблица 23. Показания, гены и причинные патогенные мутации

Заболевание	**Ген**	**Патогенная мутация**[#]
Ахроматопсия	CNGB3	1148delC
Дефицит альфа-1-антитрипсина	SERPINAI	E342K
Дефицит альфа-1-антитрипсина	SERPINAI	E342K
Дефицит альфа-1-антитрипсина	SERPINAI	R48C (R79C)
Врожденный амавроз (LCA10)	CEP290	2991+1655A>G
Синдром Андерсена-Тавила	KCNJ2	R218W
Аритмогенная кардиомиопатия правого желудочка (ARVC)	PKP2	c.235C>T
ассоциированное с врожденным дефицитом фактора XI	F11	E117*
ассоциированное с врожденным дефицитом фактора XI	F11	F283L
ATTR-амилоидоз	TTR	V50M/N30M
Аутосомно-доминантная глухота	COCH	G88E
Аутосомно-доминантная глухота	TECTA	Y1870C
Аутосомно-доминантная болезнь Паркинсона	SNCA	A53T
Аутосомно-доминантная болезнь Паркинсона	SNCA	A30P
Аутосомно-доминантный рахит	FGF23	R176Q
Аутосомно-рецессивная глухота	CX30	T5M
Аутосомно-рецессивная глухота	DFNB59	R183W
Аутосомно-рецессивная глухота	TMC1	Y182C
Аутосомно-рецессивная гиперхолестеринемия	ARH	Q136*
Анемия Даймонда-Блекфена	RPS19	R62Q
Монохроматизм синего конуса	OPN1LW	C203R
Синдром Бругада	SCN5A	E1784K
Синдром CADASIL	Ген NOTCH3	R90C
Синдром CADASIL	Ген NOTCH3	R141C
Болезнь Канавана	ASPA	E285A

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	QI00R
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	E200K
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	M129V
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	P102L
Болезнь Крейтцфельда-Якоба (CJD)	PRNP	D178N
Кистозный фиброз	CFTR	G551D
Кистозный фиброз	CFTR	W1282*
Кистозный фиброз	CFTR	R553*
Кистозный фиброз	CFTR	R117H
Кистозный фиброз	CFTR	дельта-F508
Цистиноз	CTNS	W138*
Болезнь Дарье	ATP2A2	N767S
Болезнь Дарье	ATP2A2	N767S

Заболевание	**Ген**	**Патогенная мутация**[#]
Болезнь Дарье	ATP2A2	N767S
Буллезный узловой эпидермолиз	LAMB3	R42X
Буллезный узловой эпидермолиз	LAMB3	R635X
Семейный боковой амиотрофический склероз (ALS)	SOD1	A4V
Семейный боковой амиотрофический склероз (ALS)	SOD1	H46R
Семейный боковой амиотрофический склероз (ALS)	SOD1	G37R
Болезнь Гоше	GBA	N370S
Болезнь Гоше	GBA	N370S
Болезнь Гоше	GBA	L444P
Болезнь Гоше	GBA	L444P
Болезнь Гоше	GBA	L483P
Дефицит глутарил-CoA-дегидрогеназы	GCDH	R138G
Дефицит глутарил-CoA-дегидрогеназы	GCDH	M263V
Дефицит глутарил-CoA-дегидрогеназы	GCDH	R402W
Глициновая энцефалопатия	GLDC	A389V
Глициновая энцефалопатия	GLDC	G771R
Глициновая энцефалопатия	GLDC	T269M
Гемофилия A	F8	R2178C
Гемофилия A	F8	R550C
Гемофилия A	F8	R2169H
Гемофилия A	F8	R1985Q
Гемофилия B	F9	T342M
Гемофилия B	F9	R294Q
Гемофилия B	F9	R43Q
Гемофилия B	F9	R191H
Гемофилия B	F9	G106S
Гемофилия B	F9	A279T
Гемофилия B	F9	R75*
Гемофилия B	F9	R294*

Заболевание	**Ген**	**Патогенная мутация**[#]
Гемофилия B	F9	R379Q
Врожденный дефицит антитромбина I типа	SERPINCI	R48C (R79C)
Врожденный хронический панкреатит	PRSS1	R122H
Синдром Хантера	IDS	R88C
Синдром Хантера	IDS	G374G
Синдром Гурлер (MPS1)	IDUA	Q70*
Синдром Гурлер (MPS1)	IDUA	W402*
Гиперкалиемический периодический паралич	SCN4A	T704M
Гиперкалиемический периодический паралич	SCN4A	M1592V
Гиперкалиемический периодический паралич	CACNA1S	p.Arg528X
Гиперкалиемический периодический паралич	CACNA1S	p.Arg1239
Интермиттирующая порфирия	HMBS	R173W
Изолированная агаммаглобулинемия	E47	E555K
Решетчатая дистрофия роговицы I типа	TGFBI	Arg124Cys
Дефицит LCHAD	HADHA	Glu474Gln
Врожденный амавроз Лебера 2 типа	RPE65	R44*
Врожденный амавроз Лебера 2 типа	RPE65	IVS1
Врожденный амавроз Лебера 2 типа	RPE65	G-A, +5
Синдром Леша-Нихена	HPRTI	R51*
Синдром Леша-Нихена	HPRTI	R170*
Поясно-конечностная мышечная дистрофия типа 2D	SGCA	Arg77Cys
Синдром Марото-Лами (MSPVI)	ARSB	Y210C
Средиземноморский вариант дефицита G6PD	G6PD	S188D
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Дефицит ацил-CoA-дегидрогеназы со средней длиной цепи	ACADM	K329E
Эпителиальная дистрофия роговицы Месманна	KRT12	L132P

Заболевание	**Ген**	**Патогенная мутация**[#]
Метахроматическая лейкодистрофия	ARSA	P426L
Метахроматическая лейкодистрофия	ARSA	c.459+1G>A
Синдром Моркио (MPSIVA)	GALNS	R386C
Муколипидоз IV типа	MCOLN1	406-2A>G
Муколипидоз IV типа	MCOLN1	511_6943del
Болезнь Ниманна-Пика типа A	SMPDI	L302P
Нейрональный цероидный липофусциноз (NCL)	CLN2	R208*
Нейрональный цероидный липофусциноз 1 типа	PPT1	R151*
Болезнь Паркинсона	LRRK2	G2019S
Синдром Пендреда	PDS	T461P
Синдром Пендреда	PDS	L236P
Синдром Пендреда	PDS	c.1001+1G>A
Синдром Пендреда	PDS	IVS8, +1 G>A,
Фенилкетонурия	PAH	R408W
Фенилкетонурия	PAH	I65T
Фенилкетонурия	PAH	R261Q
Фенилкетонурия	PAH	IVS10-11G>A
Фенилкетонурия	PCDH15	R245*
Фенилкетонурия	PCDH15	R245*
Болезнь Помпе	GAA	c.-32-13T>G
Первичная цилиарная дискинезия	DNAI1	IVS1+2_3insT
Первичная цилиарная дискинезия	DNAH5	10815delT
Первичная гипоксалурия	AGXT	G170R
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11	D482G (c.1445A>G)
Прогрессирующий семейный внутрипеченочный холестаз 2 типа	ABCB11	E297G
Пропионовая ацидемия	PCCB; PCCA	c.1218_1231del14ins12
Эластическая псевдоксантома	ABCC6	R1141*
Дефицит пируваткиназы (PKD)	PKLR	c.1456c->T

Заболевание	**Ген**	**Патогенная мутация**[#]
Пигментный ретинит	USH2a	C759F
Пигментный ретинит	IMPDHI	D226N
Пигментный ретинит	PDE6A	V685M
Пигментный ретинит	PDE6A	D670G
Пигментный ретинит	PRPF3	T494M
Пигментный ретинит	PRPF8	H2309R
Пигментный ретинит	RHO	P23H
Пигментный ретинит	RHO	P347L
Пигментный ретинит	RHO	P347L
Пигментный ретинит	RHO	D190N
Пигментный ретинит	RPI	R667*
Пигментный ретинит/синдром Ашера типа 1C	USH1C	V72V
Синдром Ретта	MECP2	R106W
Синдром Ретта	MECP2	R133C
Синдром Ретта	MECP2	R306C
Синдром Ретта	MECP2	R168*
Синдром Ретта	MECP2	R255*
Синдром Санфилиппо A (MPSIIIA)	SGSH	R74C
Синдром Санфилиппо A (MPSIIIA)	SGSH	R245H
Синдром Санфилиппо B (MPSIIIB)	NAGLU	R297*
Синдром Санфилиппо B (MPSIIIB)	NAGLU	Y140C
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	G216R
Тяжелый комбинированный иммунодефицит	ADA	Q3*
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E6V
Серповидно-клеточная анемия	HBB	E26K
Серповидно-клеточная анемия	HBB	E26K
Серповидно-клеточная анемия	HBB	E7K
Серповидно-клеточная анемия	HBB	c.-138C>T

Заболевание	**Ген**	**Патогенная мутация[#]**
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
Х-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.

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.

Таблица 24. Показания, гены, компенсаторные редактирования и иллюстративные признаки конструкций

Заболевание	**Ген**	**Изменение нуклеотида#**
Дефицит альфа-1-антитрипсина	SERPINAI	F51L
Дефицит альфа-1-антитрипсина	SERPINAI	M374I
Дефицит альфа-1-антитрипсина	SERPINAI	A348V/A347V
Дефицит альфа-1-антитрипсина	SERPINAI	K387R
Дефицит альфа-1-антитрипсина	SERPINAI	T59A
Дефицит альфа-1-антитрипсина	SERPINAI	T68A
ATTR-амилоидоз	TTR	A108V
ATTR-амилоидоз	TTR	R104H
ATTR-амилоидоз	TTR	T119M
Кистозный фиброз	CFTR	R555K
Кистозный фиброз	CFTR	F409L
Кистозный фиброз	CFTR	F433L
Кистозный фиброз	CFTR	H667R
Кистозный фиброз	CFTR	R1070W
Кистозный фиброз	CFTR	R29K
Кистозный фиброз	CFTR	R553Q
Кистозный фиброз	CFTR	1539T
Кистозный фиброз	CFTR	G550E
Кистозный фиброз	CFTR	F429S

Заболевание	**Ген**	**Изменение нуклеотида[#]**
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.

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).

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

Заболевание	**Ген**	**Изменение нуклеотида#**
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
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

Заболевание	**Ген**	**Изменение нуклеотида#**
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
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция
Серповидно-клеточная анемия	HBG1/2	C. -114~ -102 делеция

Заболевание	**Ген**	**Изменение нуклеотида[#]**
Серповидно-клеточная анемия	HBG1/2	c. -90 связывание с BCLIIA
Серповидно-клеточная анемия	HBG1/2	c. -90 связывание с BCLIIA
Серповидно-клеточная анемия	HBG1/2	C. -202 C>T, -201 C>T, -198 T>C, -197 C>T, -196 C>T, -195 C>G
Серповидно-клеточная анемия	HBG1/2	C. -197 C>T,-196 C>T, -195 C>G

[#]: См. J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, включенный в данный документ посредством ссылки во всей своей полноте, в отношении подробностей номенклатуры генных мутаций.

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).

Таблица 26. Иллюстративные болезни экспансии повторов, гены, причинные повторы и местоположения повторов

Заболевание	**Ген**	**Причинный повтор**	**Местоположение повтора**
миотоническая дистрофия 1 типа	DMPK/DM1	CTG	3'-UTR
миотоническая дистрофия 2 типа	ZNF9/CNBP	CCTG	Интрон 1
Дентаторубрально-паллидолуйзийская атрофия	ATN1	CAG	Кодирующая область

Заболевание	**Ген**	**Причинный повтор**	**Местоположение повтора**
Синдром ломкой Х-хромосомы с умственной отсталостью	FMR1	CGG	5'-UTR
Синдром ломкой Х-хромосомы с умственной отсталостью	FMR2	GCC	5'-UTR
Атаксия Фридрейха	FXN	GAA	Интрон
Синдром ломкой Х-хромосомы с тремором/атаксией	FMR1	CGG	5'-UTR
Болезнь Хантингтона	HTT	CAG	Кодирующая область
Болезнь 2, подобная болезни Хантингтона	JPH3	CTG	3'-UTR, кодирующая область
Миоклоническая эпилепсия Унферрихта-Лундборга	CSTB	CCCCGCCCCGCG (SEQ ID NO: 1623)	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 Х-сцепленная умственная отсталость	ARX	GCG	Кодирующая область

Иллюстративные матрицы

В некоторых вариантах осуществления в системах или способах, предусмотренных в данном документе, используются последовательности матрицы, указанные в **таблице 27**. В таблице 27 представлены иллюстративные последовательности РНК-матрицы (столбец 5) и необязательные последовательности gRNA для второго однонитевого разрыва (столбец 6), предназначенные для спаривания с полипептидом Gene Writing для коррекции указанных патогенных мутаций (столбец 4). Все матрицы в таблице 27 предназначены для иллюстрации общей последовательности из: (1) gRNA, нацеливающейся на однонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по однонитевому разрыву первой нити.

Таблица 27. Иллюстративные заболевания, ткани, гены, патогенные мутации, последовательности РНК-матрицы и последовательности gRNA для второго однонитевого разрыва

Заболевание	Ткань	Ген	Мутация	РНК-матрица	gRNA для второго однонитевого разрыва
Альфа-1-антитрипсин	Печень	SERPINA1	PiZ	TCCCCTCCAGGCCGTGCATAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTcGTCGATGGTCAGCAC	TTTGTTGAACTTGACCTCGG (SEQ ID

				AGCCTTATGCACGGCCTGGA (SEQ ID NO: 1624)	NO: 1625)
Кистозный фиброз	Легкое	CFTR	Дельта-F508	ACCATTAAAGAAAATATCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACCAaagATGATATTTTCTTTA (SEQ ID NO: 1626)	AaagATGATATTTCTTAA (SEQ ID NO: 1627)
Серповидно-клеточная анемия	HSC	HBB	HbS	GTAACGGCAGACTTCTCCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGACTCCTGaGGAGAAGTCTGCC (SEQ ID NO: 1628)	TGGTGAGGCCCTGGGCAGGT (SEQ ID NO: 1629)
Болезнь Вильсона	Печень	ATP7B	H1069Q	TTGGTGACTGCCACGCCCAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAcCCCTTGGGCGTGGCAGTC (SEQ ID NO: 1630)	GGCCAGCAGTGAACAcCCCT (SEQ ID NO: 1631)
ARVC	Сердце	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTTCACtGAACCAGCAGTGTTCC (SEQ ID NO: 1632)	TTGGTTGAAAATGATTTTGT (SEQ ID NO: 1633)
Синдром удлиненного интервала QT 1 типа	Сердце	KCNQ1	P343S	CCAGGGAAAACGCACCCACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCcCAGCGGTAGGTGCC	CTCCTTCTTTGCGCTCcCAG (SEQ ID

				CCGTGGGTGCGTTTTC (SEQ ID NO: 1634)	NO: 1635)
Муколипидоз IV типа	ЦНС	MCOLN1	406-2A>G	GCCCTCCCCTTCTCTGCCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGGTACtGTGGGCAGAGAAGGGG (SEQ ID NO: 1636)	TCAGGCAACGCCAGGTACtG (SEQ ID NO: 1637)

В некоторых вариантах осуществления в системах или способах, предусмотренных в данном документе, используются последовательности матрицы, указанные в **таблице 35**. В таблице 35 представлены иллюстративные последовательности РНК-матрицы (столбец 5) и необязательные последовательности gRNA для второго однонитевого разрыва (столбец 6), предназначенные для спаривания с полипептидом Gene Writing для коррекции указанных патогенных мутаций (столбец 4). Все матрицы в таблице 35 предназначены для иллюстрации общей последовательности из: (1) gRNA, нацеливающейся на однонитевой разрыв в первой нити, (2) домена связывания полипептида, (3) последовательности гетерологичного объекта и (4) домена гомологии с мишенью для проведения TPRT по однонитевому разрыву первой нити.

Таблица 35. Иллюстративные матрицы для Gene Writing и последовательностей gRNA для второго однонитевого разрыва для коррекции иллюстративных болезней экспансии повторов. Область матрицы, охватывающая повтор(повторы), указана строчными буквами

Заболевание	**Ген**	**Ссылочный номер доступа**	**Повтор**	**Местоположение**	**РНК-матрица**	**gRNA для второго однонитевого разрыва**

миотоническая дистрофия 1 типа	DMPK	NC_000019.10 (45769709..45782490, комплементарная нить)	CTG	3'-UTR	CTCGAAGGGTCCTTGTAGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGTGATCCCCCcagcagcag

					cctgcctgcctgTCTGTCTCACTTTGTCCCCTAGGCTGGAGTGCA (SEQ ID NO: 1640)	
Синдром ломкой X-хромосомы с умственной отсталостью	FMR1	NC_000023.11 (147911919..147951127)	CGG	5'-UTR	GGGGGCGTGCGGCAGCGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTGCGGGCGCTCGAGGCCCAGccgccgccgccgccgccgccgccgccgcctccgccgccgccgccgccgccgccgccgccgCGCTGCCGCACG (SEQ ID NO: 1642)	GCTCAGAGGCGGCCCTCCAC (SEQ ID NO: 1643)

Атаксия Фридрейха	FXN	NC_000009.12 (69035752..69079076)	GAA	Интрон	CAGGCGCGCG ACACCACGCC GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG GACCGAGTCG GTCCAACCCA GTATCTACTAA AAAATACAAA AAAAAAAAAA AAAgaagaagaaga agaagaaAATAAA GAAAAGTTAG CCGGGCGTGG TGTCGCGC(SEQ ID NO: 1644)	CGCTT GAGCC CGGGA GGCAG (SEQ ID NO: 1645)
Болезнь Хантингтона	HTT	NC_000004.12 (3074681..3243960)	CAG	Кодирующая область	GGCGGCTGAG GAAGCTGAGG GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG GACCGAGTCG GTCCAGTCCCT CAAGTCCTTCc agcagcagcagcagca gcagcagcagcagcag cagcagcagcagcagc agcagcaacagcc gccaccgccgccgccg	CGCTG CACCG ACCGT GAGTT (SEQ ID NO: 1647)

					ccgccgccgcctcctC AGCTTCCTCAG (SEQ ID NO: 1646)	
Спиноцереб еллярная атаксия 1	ATX N1	NC_000006.12 (16299112..16 761490, комплементар ная нить)	CAG	Кодирую щая область	TGAGCCCCGG AGCCCTGCTG GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG GACCGAGTCG GTCCACAAGG CTGAGcagcagca gcagcagcagcagcag cagcagcagcatc agcatcagcagcagca gcagcagcagcagcag cagcagcagcagcag CACCTCAGCA GGGCTCCGGG (SEQ ID NO: 1648)	TCCAG TTCTCC GCAGA ACAC (SEQ ID NO: 1649)

Иллюстративные последовательности гетерологичного объекта

В некоторых вариантах осуществления системы или способы, предусмотренные в данном документе, содержат последовательность гетерологичного объекта, где последовательность гетерологичного объекта или последовательность, обратно комплементарная ей, кодирует белок (например, антитело) или пептид. В некоторых вариантах осуществления лечение одобрено регулирующим органом, таким как FDA.

В некоторых вариантах осуществления белок или пептид представляет собой белок или пептид из базы данных THPdb (Usmani et al. PLoS One 12(7):e0181748 (2017), включенной в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления белок или пептид представляет собой белок или пептид, раскрытый в **таблице 28**. В некоторых вариантах осуществления системы или способы, раскрытые в данном документе, например, предусматривающие Gene Writer, могут

применяться для интеграции кассеты экспрессии белка или пептида из таблицы 28 в клетку-хозяина для обеспечения экспрессии белка или пептида у хозяина. В некоторых вариантах осуществления последовательности белка или пептида в первом столбце таблицы 28 можно найти в патентах или заявках, представленных в третьем столбце таблицы 28, которые включены в данный документ посредством ссылки во всей своей полноте.

В некоторых вариантах осуществления белок или пептид представляет собой антитело, раскрытое в таблице 1 из Lu et al. J Biomed Sci 27(1):1 (2020), включенного в данный документ посредством ссылки во всей своей полноте. В некоторых вариантах осуществления белок или пептид представляет собой антитело, раскрытое в **таблице 29**. В некоторых вариантах осуществления системы или способы, раскрытые в данном документе, например, предусматривающие Gene Writer, могут применяться для интеграции кассеты экспрессии антитела из таблицы 29 в клетку-хозяина для обеспечения экспрессии антитела в хозяине. В некоторых вариантах осуществления система или способ, описанные в данном документе, применяются для экспрессии средства, которое связывает мишень из столбца 2 таблицы 29 (например, моноклональное антитело из столбца 1 таблицы 29), у субъекта, имеющего показания, указанные в столбце 3 таблицы 29.

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

Терапевтический пептид	Категория	Номер патента
Эпоэтин альфа	Гематиники	CA1339047
Лососевый кальцитонин	Средства для сохранения плотности костей	US6440392
Интерферон альфа-n3	Иммунодепрессивные средства	
Пегфилграстим	Иммунодепрессивные средства	CA1341537
Сарграмостим	Иммунодепрессивные средства	CA1341150
Секретин	Средства диагностики	
Пегилированный интерферон альфа-2b	Иммунодепрессивные средства	CA1341567
Аспарагиназа	Антинеопластические средства	
Тиротропин альфа	Средства диагностики	US5840566
Антигемофильный фактор	Коагулянты и тромботические средства	CA2124690
Анакинра	Противоревматические средства	CA2141953
Грамицидин D	Антибактериальные средства	
Внутривенный иммуноглобулин	Иммунологические факторы	
Анистреплаза	Фибринолитические средства	
Инсулин регуляр	Противодиабетические средства	
Тенектеплаза	Фибринолитические средства	CA2129660
Менотропины	Средства для улучшения фертильности	
Интерферон гамма-1b	Иммунодепрессивные средства	US6936695
Интерферон альфа-2a, рекомбинантный		CA2172664
Коагуляционный фактор VIIa	Коагулянты	
Опрелвекин	Антинеопластические средства	
Палифермин	Средства против мукозита	

Терапевтический пептид	Категория	Номер патента
Глюкагон рекомбинантный	Гипогликемические средства	
Альдеслейкин	Антинеопластические средства	
Ботулотоксин типа B	Антидистонические средства	
Омализумаб	Противоаллергические средства	CA2113813
Лютропин альфа	Средства для улучшения фертильности	US5767251
Инсулин лиспро	Гипогликемические средства	US5474978
Инсулин гларгин	Гипогликемические средства	US7476652
Коллагеназа		
Расбуриказа	Супрессоры подагры	CA2175971
Адалимумаб	Противоревматические средства	CA2243459
Имиглюцераза	Ферментативные заместительные средства	US5549892
Абциксимаб	Антикоагулянты	CA1341357
Ингибитор альфа-1-протеиназы	Ингибиторы сериновых протеиназ	
Пегаспаргаза	Антинеопластические средства	
Интерферон бета-1a	Антинеопластические средства	CA1341604
Пегадемаза крупного рогатого скота	Ферментативные заместительные средства	
Сывороточный альбумин человека	Заменители сыворотки	US6723303
Эптифибатид	Ингибиторы агрегации тромбоцитов.	US6706681
Иодированный сывороточный альбумин	Средства диагностики	
инфликсимаб	Антиревматические средства, противовоспалительные средства, нестероидные, дерматологические	CA2106299

Терапевтический пептид	Категория	Номер патента
	средства, желудочно-кишечные средства и иммунодепрессивные средства	
Фоллитропин бета	Средства для улучшения фертильности	US7741268
Вазопрессин	Антидиуретические средства	
Интерферон бета-1b	Адъюванты, иммунологические и иммуносупрессивные средства	CA1340861
Интерферон альфакон-1	Противовирусные средства и иммунодепрессивные средства	CA1341567
Гиалуронидаза	Адъюванты, анестетики и пермеабилизирующие средства	
Инсулин свиной	Гипогликемические средства	
Трастузумаб	Антинеопластические средства	CA2103059
Ритуксимаб	Антинеопластические средства, иммунологические факторы и антиревматические средства	CA2149329
Базиликсимаб	Иммунодепрессивные средства	CA2038279
Muromo b	Иммунологические факторы и иммунодепрессивные средства	
Digoxin Immune Fab (овец)	Антидоты	
Ибритумомаб		CA2149329
Даптомицин		US6468967
Тозитумомаб		
Пегвисомант	Гормонозаместительные средства	US5849535
Ботулотоксин типа A	Нервно-мышечные блокирующие средства, средства против морщин и антидистонические средства	CA2280565
Панкрелипаза	Желудочно-кишечные средства и ферментативные заместительные средства	
Стрептокиназа	Фибринолитические средства и тромболитические средства	

Терапевтический пептид	Категория	Номер патента
Алемтузумаб		CA1339198
Альглюцераза	Ферментативные заместительные средства	
Капромаб	Индикаторы, реагенты и диагностические средства	
Ларонидаза	Ферментативные заместительные средства	
Урофоллитропин	Средства для улучшения фертильности	US5767067
Эфализумаб	Иммунодепрессивные средства	
Сывороточный альбумин	Заменители сыворотки	US6723303
Хориогонадотропин альфа	Средства фертильности и гонадотропины	US6706681
Антитимоцитарный глобулин	Иммунологические факторы и иммунодепрессивные средства	
Филграстим	Иммунодепрессивные средства, антинейтропенические средства и гемопоэтические средства	CA1341537
Фактор свертывания ix	Коагулянты и тромботические средства	
Бекаплермин	Средства, индуцирующие ангиогенез	CA1340846
Агалсидаза бета	Ферментативные заместительные средства	CA2265464
Интерферон альфа-2b	Иммунодепрессивные средства	CA1341567
Окситоцин	Утеротоники, антитоколитические средства и средства для индукции родов	
Энфувиртид	Ингибиторы слияния HIV	US6475491
Паливизумаб	Противовирусные средства	CA2197684
Даклизумаб	Иммунодепрессивные средства	
Бевацизумаб	Ингибиторы ангиогенеза	CA2286330
Арцитумомаб	Средства диагностики	US8420081

Терапевтический пептид	Категория	Номер патента
Арцитумомаб	Средства диагностики	US7790142
Экулизумаб		CA2189015
Панитумумаб		
Ранибизумаб	Офтальмологические средства	CA2286330
Идурсульфаза	Ферментативные заместительные средства	
Альглюкозидаза альфа	Ферментативные заместительные средства	CA2416492
Эксенатид	Гипогликемические средства	US6872700
Мекасермин		US5681814
Прамлинтид		US5686411
Галсульфаза	Ферментативные заместительные средства	
Абатацепт	Антиревматические средства и иммунодепрессивные средства	CA2110518
Косинтропин	Гормоны и диагностические агенты	
Кортикотропин		
Инсулин Аспарт	Гипогликемические средства и антидиабетические средства	US5866538
Инсулин детемир	Противодиабетические средства	US5750497
Инсулин глулизин	Противодиабетические средства	US6960561
Пегаптаниб	Предназначен для предотвращения респираторного дистресс-синдрома (RDS) у недоношенных детей с высоким риском развития RDS.	
Несиритид		
Тималальфазин		
Дефибротид	Антитромбины	
Натуральный интерферон альфа ИЛИ мультиферон		
Глатирамера ацетат		

Терапевтический пептид	Категория	Номер патента
Преотакт		
Тейкопланин	Антибактериальные средства	
Каникумаб	Противовоспалительные средства и моноклональные антитела	
Ипилимумаб	Антинеопластические средства и моноклональные антитела	CA2381770
Сулодексид	Антитромбины и фибринолитические средства, и гипогликемические средства, и антикоагулянты, и гиполипидемические средства	
Тоцилизумаб		CA2201781
Терипаратид	Средства для сохранения плотности костей	US6977077
Пертузумаб	Моноклональные антитела	CA2376596
Рилонацепт	Иммунодепрессивные средства	US5844099
Деносумаб	Средства для сохранения плотности костей и моноклональные антитела	CA2257247
Лираглутид		US6268343
Голимумаб	Антипсориатические средства и моноклональные антитела, и ингибитор TNF	
Белатацепт	Антиревматические средства и иммунодепрессивные средства	
Бусерелин		
Велаглюцераза альфа	Ферменты	US7138262
Тесаморелин		US5861379
Брентуксимаб ведотин		
Талиглюцераза альфа	Ферменты	
Белимумаб	Моноклональные антитела	

Терапевтический пептид	Категория	Номер патента
Афлиберцепт	Антинеопластические средства и офтальмологические средства	US7306799
Аспарагиназа erwinia chrysanthemi	Ферменты	
Окриплазмин	Офтальмологические средства	
Глюкарпидаза	Ферменты	
Тедуглутид		US5789379
Раксибакумаб	Антиинфекционные средства и моноклональные антитела	
Цертолизумаб пегол	Ингибитор TNF	CA2380298
Инсулин-изофан	Гипогликемические средства и антидиабетические средства	
Эпоэтин дзета		
Обинутузумаб	Антинеопластические средства	
Фибринолизин, он же плазмин		US3234106
Фоллитропин альфа		
Ромиплостим	Колониестимулирующие факторы и тромбопоэтические средства	
Люцинактант	Поверхностно-активные вещества легких	US5407914
Тализумаб	Иммунодепрессивные средства	
Алискирен	Ингибитор ренина	
Экстракт пыльцы амброзии		
Секукинумаб	Ингибитор	US20130202610
Соматотропин рекомбинантный	Гормонозаместительные средства	CA1326439
Дротрекогин альфа	Антисептические средства	CA2036894

Терапевтический пептид	Категория	Номер патента
Алефацепт	Дерматологические и иммунодепрессивные средства	
Липопротеин OspA	Вакцины	
Урокназа		US4258030
Абареликс	Антитестостероновые средства	US5968895
Серморелин	Гормонозаместительные средства	
Апротинин		US5198534
Гемтузумаб озогамицин	Антинеопластические средства и иммунотоксины	US5585089
Сатумомаб Пендетид	Средства диагностики	
Альбиглутид	Лекарственные средства, применяемые при сахарном диабете; пищеварительный тракт и обмен веществ; лекарственные средства, снижающие уровень глюкозы в крови, помимо инсулина.	
Алирокумаб		
Анцестим		
Антитромбин альфа		
Антитромбин III человеческий		
Асфотаза альфа	Ферменты пищеварительного тракта и обмена веществ	
Атезолизумаб		
Культивированные аутологичные хондроциты		
Беректант		
Блинатумомаб	Антинеопластические средства, иммунодепрессивные средства, моноклональные антитела,	US20120328618

Терапевтический пептид	Категория	Номер патента
	антинеопластические средства и иммуномодулирующие средства	
Ингибитор эстеразы C1 (человека)		
A-субъединица фактора свертывания крови XIII (рекомбинантная)		
Конестат альфа		
Даратумумаб	Антинеопластические средства	
Дезирудин		
Дулаглутид	Гипогликемические средства; лекарственные средства, применяемые при сахарном диабете; пищеварительный тракт и обмен веществ; лекарственные средства, снижающие уровень глюкозы в крови, помимо инсулина	
Элосульфаза альфа	Ферменты пищеварительного тракта и обмена веществ	
Элотузумаб		US2014055370
Эволокумаб	Средства, модифицирующие липиды, простые; Сердечно-сосудистая система	
Концентрат фибриногена (человека)		
Филграстим-sndz		
Внутренний фактор желудка		
Иммуноглобулин к гепатиту B		

Терапевтический пептид	Категория	Номер патента
Кальцитонин человека		
Человеческий противостолбнячный иммуноглобулин		
Человеческий иммуноглобулин к вирусу бешенства		
Человеческий иммуноглобулин Rho(D)		
Гиалуронидаза (человеческая рекомбинантная)		US7767429
Идаруцизумаб	Антикоагулянт	
Человеческий иммуноглобулин	Иммунологические факторы; иммунодепрессивные средства; Противоинфекционные средства	
Ведолизумаб	Иммунодепрессивное средство; антинеопластическое средство	US2012151248
Устекинумаб	Дерматологическое средство, иммунодепрессивное средство; антинеопластическое средство	
Туроктоког альфа		
Очищенное туберкулиновое производное белка		
Симоктоког альфа	Антигеморрагические средства: фактор свертывания крови VIII	

Терапевтический пептид	Категория	Номер патента
Силтуксимаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства	US7612182
Себелипаза альфа	Ферменты	
Сакрозидаза	Ферменты	
Рамуцирумаб	Антинеопластические средства и иммуномодулирующие средства	US2013067098
Концентрат протромбинового комплекса		
Поракtан альфа	Поверхностно-активные вещества легких	
Пембролизумаб	Антинеопластические средства и иммуномодулирующие средства	US2012135408
Пегинтерферон бета-1а		
Офатумумаб	Антинеопластические средства и иммуномодулирующие средства	US8337847
Обилтоксаксимаб		
Ниволумаб	Антинеопластические средства и иммуномодулирующие средства	US2013173223
Нецитумумаб		
Метрелептин		US20070099836
Метокси полиэтиленгликоль -эпоэтин бета		
Меполизумаб	Антинеопластические и иммуномодулирующие средства, иммунодепрессивные средства, ингибиторы интерлейкина	US2008134721
Иксекизумаб		
Инсулин свиной	Гипогликемические средства, антидиабетические средства	

Терапевтический пептид	Категория	Номер патента
Инсулин Деглудек		
Инсулин из говядины		
Тиреоглобулин	Гормональная терапия	US5099001
Человеческий противосибиреязвенный 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 С1 (рекомбинантный)	Кровь и органы кроветворения	
Каникумаб	Антинеопластические средства и иммуномодулирующие средства	
Хорионический гонадотропин (человека)	Гормоны	US6706681
Хорионический гонадотропин (Рекомбинантный)	Гормоны	US5767251
Человеческий фактор	Факторы свертывания крови	

Терапевтический пептид	Категория	Номер патента
свертывания крови X		
Динутуксимаб	Антитело, иммунодепрессивное средство, антинеопластическое средство	US20140170155
Эфмороктоког альфа	Антигемофильный фактор	
Комплекс фактора IX (человека)	Антигемофильное средство	
Вакцина к гепатиту A	Вакцина	
Человеческий иммуноглобулин против вируса ветряной оспы	Антитело	
Ибритумомаб тиуксетан	Антитело, иммунодепрессивные средства	CA2149329
Ленограстим	Антинеопластические средства и иммуномодулирующие средства	
Пеглотиказа	Ферменты	
Протамина сульфат	Антагонисты гепарина, гематологические средства	
Белок S человека	Антикоагулянтный белок плазмы	
Сипулеуцель-Т	Антинеопластические средства и иммуномодулирующие средства	US8153120
Соматропин рекомбинантный	Гормоны, гормонозаместители и антагонисты гормонов	CA1326439, CA2252535, US5288703, US5849700, US5849704, US5898030, US6004297, US6152897,

Терапевтический пептид	Категория	Номер патента
		US6235004, US6899699
Сусоктоког альфа	Факторы свертывания крови, Антигеморрагические средства	
Тромбомодулин альфа	Антикоагулянтное средство, антиагрегантое средство	

Таблица 29. Иллюстративные средства терапии на основе моноклональных антител

mAb	Мишень	Показание
Муромонаб-CD3	CD3	Отторжение почечного трансплантата
Абциксимаб	GPIIb/IIIa	Предупреждение образования тромбов при ангиопластике
Ритуксимаб	CD20	Неходжкинская лимфома
Паливизумаб	RSV	Предупреждение респираторно-синцитиальной вирусной инфекции
Инфликсимаб	TNFα	Болезнь Крона
Трастузумаб	HER2	Рак молочной железы
Алемтузумаб	CD52	Хронический миелоидный лейкоз
Адалимумаб	TNFα	Ревматоидный артрит
Ибритумомаб тиуксетан	CD20	Неходжкинская лимфома
Омализумаб	IgE	Астма
Цетуксимаб	EGFR	Колоректальный рак
Бевацизумаб	VEGF-A	Колоректальный рак
Натализумаб	ITGA4	Рассеянный склероз
Панитумумаб	EGFR	Колоректальный рак
Ранибизумаб	VEGF-A	Макулярная дегенерация
Экулизумаб	C5	Пароксизмальная ночная гемоглобинурия
Цертолизумаб пегол	TNFα	Болезнь Крона
Устекинумаб	IL-12/23	Псориаз

Канакинумаб	IL-1β	Синдром Макла-Уэльса
Голимумаб	TNFα	Ревматоидный и псориатический артрит, анкилозирующий спондилоартрит
Офатумумаб	CD20	Хронический лимфоцитарный лейкоз
Тоцилизумаб	IL-6R	Ревматоидный артрит
Деносумаб	RANKL	Потеря костной массы
Белимумаб	BLyS	Системная красная волчанка
Ипилимумаб	CTLA-4	Метастатическая меланома
Брентуксимаб ведотин	CD30	Лимфома Ходжкина, системная анапластическая крупноклеточная лимфома
Пертузумаб	HER2	Рак молочной железы
Трастузумаб эмтанзин	HER2	Рак молочной железы
Раксибакумаб	PA *B. anthracis*	Сибиреязвенная инфекция
Обинутузумаб	CD20	Хронический лимфоцитарный лейкоз
Силтуксимаб	IL-6	Болезнь Кастлемана
Рамуцирумаб	VEGFR2	Рак желудка
Ведолизумаб	Интегрин α4β7	Неспецифический язвенный колит, болезнь Крона
Блинатумомаб	CD19, CD3	Острый лимфобластный лейкоз
Ниволумаб	PD-1	Меланома, немелкоклеточный рак легкого
Пембролизумаб	PD-1	Меланома
Идаруцизумаб	Дабигатран	Устранение дабигатран-индуцированной антикоагуляции
Нецитумумаб	EGFR	Немелкоклеточный рак легкого
Динутуксимаб	GD2	Нейробластома
Секукинумаб	IL-17α	Псориаз
Меполизумаб	IL-5	Тяжелая эозинофильная астма
Алирокумаб	PCSK9	Высокое содержание холестерина
Эволокумаб	PCSK9	Высокое содержание холестерина
Даратумумаб	CD38	Множественная миелома
Элотузумаб	SLAMF7	Множественная миелома
Иксекизумаб	IL-17α	Псориаз
Реслизумаб	IL-5	Астма

Оларатумаб	PDGFRα	Саркома мягких тканей
Безлотоксумаб	Энтеротоксин B *Clostridium difficile*	Предупреждение рецидива инфекции *Clostridium difficile*
Атезолизумаб	PD-L1	Рак мочевого пузыря
Обилтоксаксимаб	PA *B. anthracis*	Предупреждение ингаляционной формы сибирской язвы
Инотузумаб озогамицин	CD22	Острый лимфобластный лейкоз
Бродалумаб	IL-17R	Бляшечный псориаз
Гуселькумаб	IL-23, p19	Бляшечный псориаз
Дупилумаб	IL-4Rα	Атопический дерматит
Сарилумаб	IL-6R	Ревматоидный артрит
Авелумаб	PD-L1	Карцинома из клеток Меркеля
Окрелизумаб	CD20	Рассеянный склероз
Эмицизумаб	Фактор IXa, X	Гемофилия A
Бенрализумаб	IL-5Rα	Астма
Гемтузумаб озогамицин	CD33	Острый миелоидный лейкоз
Дурвалумаб	PD-L1	Рак мочевого пузыря
Буросумаб	FGF23	X-сцепленная гипофосфатемия
Ланаделумаб	Плазменный калликреин	Приступы наследственного ангионевротического отека
Могамулизумаб	CCR4	Грибовидный микоз или синдром Сезари
Эренумаб	CGRPR	Предупреждение мигрени
Галканезумаб	CGRP	Предупреждение мигрени
Тилдракизумаб	IL-23, p19	Бляшечный псориаз
Цемиплимаб	PD-1	Плоскоклеточный рак кожи
Эмапалумаб	IFNγ	Первичный гемофагоцитарный лимфогистиоцитоз
Фреманезумаб	CGRP	Предупреждение мигрени
Ибализумаб	CD4	HIV-инфекция
Моксетумомаб пасудодокс	CD22	Волосатоклеточный лейкоз
Равулизумаб	C5	Пароксизмальная ночная гемоглобинурия

Каплацизумаб	фактор фон Виллебранда	Приобретенная тромботическая тромбоцитопеническая пурпура
Ромосозумаб	Склеростин	Остеопороз у женщин в постменопаузе с повышенным риском переломов
Рисанкизумаб	IL-23, p19	Бляшечный псориаз
Полатузумаб ведотин	CD79β	Диффузная крупноклеточная В-клеточная лимфома
Бролуцизумаб	VEGF-A	Макулярная дегенерация
Кризанлизумаб	P-селектин	Серповидно-клеточная анемия

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.

A. 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 подвида *japonica*) или в часть растения (например, в каллюс растения) посредством агробактерий. В некоторых вариантах осуществления системы и способы, описанные в данном документе, могут применяться в растениях путем замены гена растения (например, гена гигромицинфосфотрансферазы (HPT)) нулевым аллелем (например, содержащим замену основания в старт-кодоне). Системы и способы модификации генома растения описаны в Xu et. al. *Development of plant prime-editing systems for precise genome editing*, 2020, Plant Communications.

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, имеющихся до или после сбора урожая, которые считаются необходимыми для сельского хозяйства или садоводства (например, вкуса, внешнего вида, срока хранения), или улучшению признаков, которые иным образом приносят пользу людям (например, понижение продуцирования аллергенов). Улучшение урожайности растения относится к повышению выхода продукта (например, определяемого по выходу биомассы, зерна, семян или плодов растения, содержанию белка, углеводов или масел или площади листьев) растения на измеримую величину по сравнению с выходом того же продукта растения, полученного в тех же условиях, но без применения композиций по настоящему изобретению, или по сравнению с применением обычных средств для модификации растений. Например, урожайность может быть повышена на по меньшей мере приблизительно 0,5%, приблизительно 1%, приблизительно 2%, приблизительно 3%, приблизительно 4%, приблизительно 5%, приблизительно 10%, приблизительно 20%, приблизительно 30%, приблизительно 40%, приблизительно 50%, приблизительно 60%, приблизительно 70%, приблизительно 80%, приблизительно 90%, приблизительно 100% или более 100%. В некоторых случаях способ является эффективным для повышения урожайности в приблизительно 2, 5, 10, 25, 50, 75, 100 или более 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 (например, средств для модификации растений, доставляемых без использования PMP).

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.

B. 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 так желатин, при этом данное покрытие растворяется или разрушается в среде его использования, что делает тогда композицию на основе системы Gene Writer для модификации растений доступной, или за счет распределения средства в растворимой или разрушаемой матрице. Таки средства/устройства с непрерывным высвобождением и/или распределением могут преимущественно использоваться для устойчивого поддержания эффективной концентрации одной или нескольких композиций для модификации растений, описанных в данном документе.

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.

С. 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, ликофиты, мохообразные и водоросли (например, многоклеточные или одноклеточные водоросли). Растения, которые можно обрабатывать в соответствии со способами по настоящему изобретению, дополнительно включают любое сосудистое растение, например, однодольные, или двудольные, или голосеменные, в том числе без ограничения люцерну, яблоню, растение рода Arabidopsis, банан, ячмень, канолу, клещевину, цикорий, хризантему, клевер, какао, кофе, хлопчатник, семена хлопчатника, кукурузу, крамбе, клюкву, огурец, дендробиум, диоскорею, эвкалипт, овсяницу, лен, гладиолус, растения семейства Лилейные, семя льна, просо, дыню, горчицу, овес, масличную пальму, масличный рапс, папайю, арахис, ананас, декоративные растения, фасоль, картофель, рапс, рис, рожь, райграс, сафлор, кунжут, сорго, сою, сахарную свеклу, сахарный тростник, подсолнечник, клубнику, табак, томат, газонную траву, пшеницу и овощные культуры, такие как салат, сельдерей, брокколи, цветная капуста, тыквенные культуры; фруктовые и ореховые деревья, такие как яблоня, груша, персик, апельсин, грейпфрут, лимон, лайм, миндаль, пекан, грецкий орех, лещина; вьющиеся растения, такие как виноград (например, виноградник), киви, хмель; плодовые кустарники и колючие кустарники, такие как малина, ежевика, крыжовник; лесные деревья, такие как ясень, сосна, пихта, клен, дуб, каштан, тополь; такое как люцерна, канола, клещевина, кукуруза, хлопчатник, крамбе, лен, семя льна, горчицу, масличную пальму, масличный рапс, арахис, картофель, рис, сафлор, кунжут, сою, сахарную свеклу, подсолнечник, табак, томат или пшеницу. Растения, которые можно обрабатывать в соответствии со способами по настоящему изобретению, включают любое культурное растение, например, кормовую культуру, масличную культуру, зерновую культуру, плодовую культуру, овощную культуру, волокнистую культуру, пряную культуру, ореховую культуру, газонную культуру, сахарную культуру, культуру для производства напитков и лесную культуру. В определенных случаях культурное растение, которое обрабатывают в этом способе, представляет собой растение сои. В других определенных случаях культурное растение представляет собой пшеницу. В определенных случаях культурное растение представляет собой кукурузу. В определенных случаях культурное растение представляет собой хлопчатник. В определенных случаях культурное растение представляет собой люцерну. В определенных случаях культурное растение представляет собой сахарную свеклу. В определенных случаях культурное растение представляет собой рис. В определенных случаях культурное растение представляет собой томат. В определенных случаях культурное растение представляет собой картофель.

В определенных случаях растение представляет собой сельскохозяйственную культуру. Примеры таких культурных растений включают без ограничения однодольные и

двудольные растения, в том числе без ограничения кормовые растения или кормовые бобовые растения, декоративные растения, продовольственные культуры, деревья или кустарники, выбранные из Acer spp., Allium spp., Amaranthus spp., Ananas comosus, Apium graveolens, Arachis spp, Asparagus officinalis, Beta vulgaris, Brassica spp. (например, Brassica napus, Brassica rapa ssp. (канола, масличный рапс, репа масличная), Camellia sinensis, Canna indica, Cannabis saliva, Capsicum spp., Castanea spp., Cichorium endivia, Citrullus lanatus, Citrus spp., Cocos spp., Coffea spp., Coriandrum sativum, Corylus spp., Crataegus spp., Cucurbita spp., Cucumis spp., Daucus carota, Fagus spp., Ficus carica, Fragaria spp., Ginkgo biloba, Glycine spp. (например, Glycine max, Soja hispida или Soja max), Gossypium hirsutum, Helianthus spp. (например, Helianthus annuus), Hibiscus spp., Hordeum spp. (например, Hordeum vulgare), Ipomoea batatas, Juglans spp., Lactuca sativa, Linum usitatissimum, Litchi chinensis, Lotus spp., Luffa acutangula, Lupinus spp., Lycopersicon spp. (например, Lycopersicon esculenturn, Lycopersicon lycopersicum, Lycopersicon pyriforme), Malus spp., Medicago sativa, Mentha spp., Miscanthus sinensis, Morus nigra, Musa spp., Nicotiana spp., Olea spp., Oryza spp. (например, Oryza sativa, Oryza latifolia), Panicum miliaceum, Panicum virgatum, Passiflora edulis, Petroselinum crispum, Phaseolus spp., Pinus spp., Pistacia vera, Pisum spp., Poa spp., Populus spp., Prunus spp., Pyrus communis, Quercus spp., Raphanus sativus, Rheum rhabarbarum, Ribes spp., Ricinus communis, Rubus spp., Saccharum spp., Salix sp., Sambucus spp., Secale cereale, Sesamum spp., Sinapis spp., Solanum spp. (например, Solanum tuberosum, Solanum integrifolium или Solanum lycopersicum), Sorghum bicolor, Sorghum halepense, Spinacia spp., Tamarindus indica, Theobroma cacao, Trifolium spp., Triticosecale rimpaui, Triticum spp. (например, Triticum aestivum, Triticum durum, Triticum turgidum, Triticum hybernum, Triticum macha, Triticum sativum или Triticum vulgare), Vaccinium spp., Vicia spp., Vigna spp., Viola odorata, Vitis spp. и Zea mays. В определенных вариантах осуществления культурное растение представляет собой рис, масличный рапс, канолу, сою, кукурузу (маис), хлопчатник, сахарный тростник, люцерну, сорго или пшеницу.

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.

Способы введения и доставки

Элементы систем, представляющие собой нуклеиновую кислоту, предусмотренные в настоящем изобретении, применяемые в способах, предусмотренных в настоящем изобретении, могут быть доставлены посредством различных способов. В вариантах осуществления, где система содержит две отдельные молекулы нуклеиновой кислоты (например, ретротранспозаза и нуклеиновая кислота-матрица представляют собой отдельные молекулы), две молекулы могут доставляться одним и тем же способом, тогда как в других вариантах осуществления две молекулы доставляются разными способами. 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, ex vivo или in vivo. В некоторых вариантах осуществления клетки представляют собой эукариотические клетки, например, клетки многоклеточного организма, например, животного, например, млекопитающего (например, человека, свиньи, крупного рогатого скота), птицы (например, домашней птицы, такой как курица, индейка или утка) или рыбы. В некоторых вариантах осуществления клетки представляют собой клетки животного, отличного от человека (например, лабораторного животного, сельскохозяйственного животного или животного-компаньона). В некоторых вариантах осуществления клетка представляет собой стволовую клетку (например, гемопоэтическую стволовую клетку), фибробласт или Т-клетку. В некоторых вариантах осуществления клетка представляет собой неделящуюся клетку, например, неделящийся фибробласт или неделящуюся Т-клетку. 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.

Липидные наночастицы являются другим примером носителя, который представляет собой биосовместимую и биоразлагаемую систему доставки для фармацевтических композиций, описанных в данном документе. Наноструктурированные липидные носители (NLC) представляют собой модифицированные твердые липидные наночастицы (SLN), которые сохраняют характеристики SLN, улучшают стабильность и емкость загрузки лекарственного средства и предотвращают утечку лекарственного средства. Полимерные наночастицы (PNP) являются важным компонентом для доставки лекарственных средств. Эти наночастицы могут эффективно направлять доставку лекарственного средства к конкретным мишеням и улучшать стабильность лекарственного средства и контролируемое высвобождение лекарственного средства. Также можно использовать липидно-полимерные наночастицы (PLN) - новый тип носителя, который является комбинацией липосом и полимеров. Эти наночастицы обладают взаимодополняющими преимуществами PNP и липосом. PLN состоит из структуры сердцевина-оболочка; полимерная сердцевина обеспечивает стабильную структуру, а фосфолипидная оболочка обеспечивает хорошую биосовместимость. Таким образом, эти два компонента увеличивают показатель эффективности инкапсуляции лекарственного средства, способствуют модификации поверхности и предотвращают утечку водорастворимых лекарственных средств. Для обзора см., например, Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Экзосомы также можно применять в качестве носителей для доставки лекарственных средств для композиций и систем, описанных в данном документе. Для

обзора см. Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; https://doi.org/10.1016/j.apsb.2016.02.001.

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).

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.

ПРИМЕРЫ

Настоящее изобретение дополнительно проиллюстрировано с помощью следующих примеров. Примеры представлены только для иллюстративных целей, и их не следует истолковывать как ограничивающие объем или содержание настоящего изобретения каким-либо образом.

Пример 1. 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 c-myb в ДНК-связывающем домене (DBD) и замену его гибким линкером (фигура 1a). 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 A (предсказанный статистический клубок, следующий за мотивом c-myb), либо в местоположении B (конец предсказанной альфа-спирали, которая содержит часть мотива c-myb), и удаленная область заканчивается либо в местоположении v1 (альфа-спиральная область R2Tg, которая предшествовала предсказанному -1 РНК-связывающему мотиву), либо в местоположении v2 (C-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 A, XTEN: SGSETPGTSESATPES (SEQ ID NO: 1023), и линкера B, 3GS: GGGS (SEQ ID NO: 1024)). В каждой из этих мутантных ретротранспозаз, содержащих разные удаленные области (местоположение A - v1, местоположение A - v2, местоположение B - v1 или местоположение B - v2), они были заменены либо линкером A, либо линкером B с помощью ПЦР на ДНК-плазмиду, которая экспрессировала R2Tg, в результате чего были получены следующие последовательности: c-mybA - v1 заменен линкером 3GS, c-mybA - v2 заменен линкером 3GS, c-mybA - v1 заменен линкером XTEN, c-mybA - v2 заменен линкером XTEN, c-mybA - v2 заменен линкером XTEN, c-mybA - v2 заменен линкером 3GS, mybB - v1 заменен линкером 3GS, c-mybB - v2 заменен линкером 3GS, c-mybB - v1 заменен линкером XTEN, c-mybB - v2 заменен линкером XTEN, как показано в таблице E1 ниже. Вставку линкеров проверяли посредством секвенирования по Сэнгеру и ДНК-плазмиды очищали для трансфекции.

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 c-mybA - v1" заменен линкером 3GS	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSR KPAEEPREEPGTCHHTRRAA***GGGS***CFGCLESISQIRTATRDK KDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDV VKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKP DRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFI RAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVS*	

	HQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDK IQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQS SITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQG QKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKY LGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILK TYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCT CDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTM KCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPN NVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREF LARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRH NYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKD ARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHL TNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVK MAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg с делецией природного линкера, участок "местоположение c-mybA - v2" заменен линкером 3GS	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF* *GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV* *DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL* *VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE* *TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG* *EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF* *RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR* *DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW* *TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSR* *KPAEEPREEPGTCHHTRRAA***GGGS**TATRDKKDTVTREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIEC LSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT MEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQR EVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGD PMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDL VLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIAR	

	SGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHS EVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGG LGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQL HKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHW IQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC RHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKK DWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYE AAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPL GARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSV DIVHMFASRARKSMVM	
R2Tg с делецией природного линкера, участок "местоположение c-mybA - v1" заменен линкером XTEN	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF* *GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV* *DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL* *VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE* *TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG* *EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF* *RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR* *DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW* *TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSR* *KPAEEPREEPGTCHHTRRAA***SGSETPGTSESATPES**CFGCLE SISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYL RFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWET TGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGS APGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRG CRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSK ACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVF VDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTY ITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLE ESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILET FCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPL NMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLH RIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSI	

	WEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFK KWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQ LRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCP VTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKEL YKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKY KHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTEL GLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg с делецией природного линкера, участок "местоположение c-mybA - v2" заменен линкером XTEN	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF* *GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV* *DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL* *VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE* *TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG* *EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF* *RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR* *DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW* *TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSR* *KPAEEPREEPGTCHHTRRAA***SGSETPGTSESATPES**TATRDK KDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDV VKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKP DRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFI RAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVS HQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDK IQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQS SITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQG QKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKY LGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILK TYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCT CDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTM KCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPN NVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREF LARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRH NYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKD	

	ARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHL TNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVK MAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg с делецией природного линкера, участок "местоположение c-mybB - v1" заменен линкером 3GS	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL**GG GS**CFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDR AIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTG KIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFS RIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKR EHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLA MDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDF WLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLET LDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLI PSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAG GDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCN WRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHR KLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESC AHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHI RDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAA AEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRA RKSMVM*	

R2Tg с делецией природного линкера, участок "местоположение c-mybB - v2" заменен линкером 3GS	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***GG GS***TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTG SFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAP GPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCR TVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKAC PLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVD IAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITT KRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCN LTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMI DPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKP LQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRI AQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWE APPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRA NVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQ DARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKP DLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHL ETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLS KSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM*	
R2Tg с делецией природного линкера, участок	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG*	

"местоположение c-mybB - v1" заменен линкером XTEN	*EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***SG*** **SETPGTSESATPES**CFGCLESISQIRTATRDKKDTVTREKHP KKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIE CLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAF RELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT MEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQR EVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGD PMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDL VLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTK DSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIAR SGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHS EVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGG LGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQL HKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHW IQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC RHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKK DWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYE AAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPL GARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSV DIVHMFASRARKSMVM*	
R2Tg с делецией природного линкера, участок "местоположение c-mybB - v2" заменен	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTG EEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCF RCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNR DFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCW TKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***SG***	

линкером XTEN	**SETPGTSESATPES**TATRDKKDTVTREKHPKKPFQKWMKD RAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTG KIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFS RIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKR EHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLA MDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDF WLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLET LDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLI PSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAG GDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCN WRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHR KLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESC AHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHI RDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAA AEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRA RKSMVM	

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 Fugene 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 с помощью измерения числа ампликонов с 3ʻ соединениями (фигура 1b).

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 (С-конец по отношению к альфа-спирали, предшествующей -1 РНК-связывающему мотиву). При делеции, начинающейся в местоположении А 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 В (c-mybB), данные конфигурации демонстрируют более заметное снижение активности интеграции по сравнению с диким типом или делециями в местоположении А (c-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 В с местоположениями v1 или v2. Даже несмотря на то, что N-концевое местоположение делеции природного линкера mybB является субоптимальным, С-концевой конец делеции характеризовался наилучшей переносимостью в v2 с любым из линкеров 3GS или XTEN, и, по-видимому, это является предпочтительным местоположением для наличия полипептида, предшествующего области RBD-1.

Пример 2. Определение специфичности эндонуклеазного домена Gene Writer в отношении мишени

В данном примере описывается 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 данном примере были созданы линии клеток, содержащие "посадочные площадками" или стабильные интеграции, которые имитируют область рДНК, содержащую положение R2, которое используется ретротранспозазами R2 в качестве мишени для ретротранспозиции (см. фиг. 2). 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 Е2). 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 пакующую линию клеток LentiX-293T (Takara Bio). Трансфицированные клетки инкубировали при 37°С, 10% CO_2 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 матричной молекулы Gene Writer были разработаны таким образом, чтобы они характеризовались 100% гомологией с последовательностью длиной 100 п. о., расположенной слева (модуль A 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 матричной молекулы 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 матричные молекулы 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% CO_2 в течение 3 дней перед лизисом клеток и экстракцией геномной ДНК. С помощью ddPCR экстрагированную gDNA оценивали на предмет интеграции матричной молекулы Gene Writer в сайт посадочной площадки. Активность внесения однонитевых 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 матричной молекулы 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 3a). Более того, интеграция не обнаруживается при использовании матричных молекул 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 однонитевых разрывов в ДНК измеряли с помощью NGS-анализа посадочной площадки. Активность внесения однонитевых разрывов не зависит от матричной молекулы Gene Writer, поскольку управляющий элемент R2Tg WT Gene Writer характеризовался сопоставимыми показателями вставок/делеций на посадочной площадке WT при использовании матричной молекулы Gene Writer WT, с мутантной последовательностью длиной 12 п. о. или мутантной последовательностью длиной 75 п. о. (фигура 3b). В случае посадочных площадок с 12 п. о. и 75 п. о., независимо от матричной молекулы 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.

Таблица E2. Информация о посадочной площадке

Название последовательности посадочной площадки	Последовательность 5′-> 3‘ (рДНК, подчеркнуто; **область расщепления**, выделено полужирным шрифтом; ***подвергнутая мутации последовательность***, выделено полужирным курсивом
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**AAGGTAGCCAAA**TGCCTCGTCATCTAATTAG TGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTAC CTACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCC GTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACC CAACTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1658)
Мутантная форма 12 п. о.	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT

	GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT***TCCAATATGATT***TGCCTCGTCATCTAATTAGTG ACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACCT ACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCGT CGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCA ACTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1659)
Мутантная форма 75 п. о.	Gctcacacaggaaacagctatgaccatgattacgccaagctgttgacgcgatgtgatttctgcccagt gctctgaatgtcaaagtgaagaaattcaatgaagcgcgggtaaacggcgggagtaactatgactctctt t***ccaatatgattccacccatggcaaattccatggcaccgtcaaggctgagaacgggaagcttgtca tcaatggaa***actatccagcgaaaccacagccaagggaaattcactggccgtcgttttacaacgtcgtg actgggaaaaccctggcgttacccaacttaatcgccttgcagcacatcc (SEQ ID NO: 1660)

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 однонитевого разрыва. В конкретных вариантах осуществления последовательность из 13 нуклеотидов (TAAGGTAGCCAAA) (SEQ ID NO: 1661) на основе сайта однонитевого разрыва элемента R2, например, R2Tg, применяют для поиска в геноме человека подходящих местоположений для перенацеливания Gene Writer, в котором гетерологичный ДНК-связывающий домен предназначен для обеспечения перемещения 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 3). 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 однонитевого разрыва в новой мишени.

Таблица E3. Сайты в геноме человека, соответствующие отрезку длиной 13 нуклеотидов вблизи сайта однонитевого 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

Пример 3. 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 активности эндонуклеазы, специфической в отношении двухнитевой ДНК (называемой каталитически мертвым 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 никазой Cas9) (см. фиг. 5).

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 4 и 6. 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).

Таблица E4. 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 местоположении	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***<u>GGGS</u>**TATRDKKD TVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKII LDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNT AFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT

v2 DBD R2Tg с помощью линкера 3GS	MEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIG SILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWS AKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPL LCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISIL ETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLN MIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPL QKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLP PCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTM KCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVS TNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFER DKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYI KACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKK DWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAK SSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKW HQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRA RKSMVM (SEQ ID NO: 1662)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный в местоположении v2 DBD R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***<u>SGSETPGTSESAT PES</u>***TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRL FYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGD FKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDV VKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSE NLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQR EVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMS PLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQ RIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHR IAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP SSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQ GRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFL

	ARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICE LLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDV TVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGF PLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDI VHMFASRARKSMVM (SEQ ID NO: 1663)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный в местоположении v1 DBD R2Tg с помощью линкера 3GS	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC* *DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR* *THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***GGGS**CFGCLESIS QIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRL FYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGD FKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDV VKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSE NLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQR EVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMS PLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCA AWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQ RIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHR IAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP SSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQ GRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICE LLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDV TVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGF PLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDI VHMFASRARKSMVM (SEQ ID NO: 1664)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный в местоположении v1 DBD R2Tg с	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC* *DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR* *THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT** **PES**CFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAI KKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTR WETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGS APGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRT

помощью линкера XTEN	VLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNP RQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDT VSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKI QIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITA MAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGI ARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSE VKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITK LAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQ AGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNW RKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLT ALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCP VTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKP DLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETE VRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVK MAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1665)
Специфичный в отношении AAVS1 правый ZFP, прикрепленный в местоположении v2 DBD R2Tg с помощью линкера 3GS	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***GGGS***TATRDKKDT VTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIIL DDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAF RELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEI FNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSIL LRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAK REHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSN MYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLL CKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILE TFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMI DPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQK TDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPC TCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKC FMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNS EWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKI SNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKA CRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDW*

	VVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSL EEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQ DNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARK SMVM (SEQ ID NO: 1666)
Специфичный в отношении AAVS1 правый ZFP, прикрепленный в местоположении v2 DBD R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***<u>SGSETPGTSESATP ES</u>***TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLF YLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDF KTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVV KMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKD INNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSEN LKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQRE VDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSP LLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSW ENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAA WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQR IDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTA VKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRI AQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPS SEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQG RGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLA RGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICEL LLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1667)*
Специфичный в отношении AAVS1 правый ZFP, прикрепленный в местоположении v1 DBD R2Tg с	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***<u>GGGS</u>***CFGCLESISQI RTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFY LDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFK TYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVK MDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENL*

помощью линкера 3GS	KLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPL LFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWE NMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAW TINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1668)
Специфичный в отношении AAVS1 правый ZFP, прикрепленный в местоположении v1 DBD R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATP ES***CFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIK KGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRW ETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAP GPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLI PKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQR GFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSH QHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRV GVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAF ADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPT KDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG LSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLI PSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDR ENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEF KKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLR ANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQD ARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFV

	KDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLT NAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETF STVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1669)
Специфичные в отношении AAVS1 левый и правый ZFP (разделенные линкером XTEN), прикрепленные в местоположении v2 DBD R2Tg с помощью линкера 3GS	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT PES****GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFAC DICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIR THTGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***GGGS*TATRDKK DTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAK IILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNT AFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT MEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIG SILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWS AKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPL LCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISIL ETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLN MIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPL QKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLP PCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTM KCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVS TNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFER DKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYI KACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKK DWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAK SSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKW HQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRA RKSMVM (SEQ ID NO: 1670)*
Специфичные в отношении AAVS1 левый и правый ZFP (разделенные линкером XTEN),	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT PES****GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFAC DICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIR THTGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESA*

прикрепленные в местоположении v2 DBD R2Tg с помощью линкера XTEN	**TPES**TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQR LFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLG DFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGD VVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRL KDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCS ENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQ REVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPM SPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSD SWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDC AAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWL QRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIR TAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLH RIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAP PSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLAS QGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREF LARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYIC ELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVG FPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVD IVHMFASRARKSMVM (SEQ ID NO: 1671)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный к N-концу R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT PES***ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSD FGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLV NLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETL HSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKV MVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCFRCAKC GKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNRDFTTKIG LGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEEL LIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPR EEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEER LSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRT

	ATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTY GKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINN WRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKL LQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDP HIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLF NLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWT INGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1672)
Специфичный в отношении AAVS1 правый ZFP, прикрепленный к N-концу R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI* *CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH* *TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATP** **ES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLV NLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETL HSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKV MVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCFRCAKC GKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNRDFTTKIG LGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEEL LIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPR EEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEER LSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRT

	ATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTY GKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINN WRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKL LQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDP HIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLF NLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWEN MNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWT INGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1673)
Специфичные в отношении AAVS1 левый и правый ZFP, прикрепленные к N-концу R2Tg с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC* *DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR* *THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT** **PES***GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFAC* *DICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIR* *THTGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESA** **TPES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGS DFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQ VDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEE KVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVSHGKRGVCFRC AKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCNRDFTT KIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEE ELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE

	PREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPG RAAAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLE ERLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQI RTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFY LDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFK TYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVK MDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENL KLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPL LFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWE NMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAW TINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1674)
Специфичный в отношении AAVS1 левый ZFP, прикрепленный к N-концу R2Tg, содержащего мутации инактивации DBD, с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC* *DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR* *THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT** **PES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSD FGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLV NLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETL HSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKV MVTVPDKNPPSPSSGTRVNSVLNLIEHLKVSHGKRGVCFRCAKC GKENSNYHSVVCHFPKCRGPETEKAPAGEWISEVSNRDFTTKIGL GQHKRLAHPAVRNQERIVASQPKETSNRGAHKACATKEEEELLI RLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREE

	PGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAA AGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERL SAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTA TRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDR GKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYG KADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDP EFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNW RPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLL QTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPH IVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFN LAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENM NTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQ APLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVK EWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQS SDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEP PNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGI VNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLL EEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTV RYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPL GARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVH MFASRARKSMVM (SEQ ID NO: 1675)
Специфичный в отношении AAVS1 правый ZFP, прикрепленный к N-концу R2Tg, содержащего мутации инактивации DBD, с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDI* *CGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTH* *TGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATP** **ES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDF GGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQV DLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLV NLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETL HSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKV MVTVPDKNPPSPSSGTRVNSVLNLIEHLKVSHGKRGVCFRCAKC GKENSNYHSVVCHFPKCRGPETEKAPAGEWISEVSNRDFTTKIGL GQHKRLAHPAVRNQERIVASQPKETSNRGAHKACATKEEEELLI RLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREE

	PGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAA AGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERL SAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTA TRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDR GKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYG KADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDP EFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNW RPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLL QTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPH IVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFN LAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENM NTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQ APLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVK EWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQS SDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEP PNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGI VNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLL EEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTV RYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPL GARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVH MFASRARKSMVM (SEQ ID NO: 1676)
Специфичные в отношении AAVS1 левый и правый ZFP, прикрепленные к N-концу R2Tg, содержащего мутации инактивации DBD, с помощью линкера XTEN	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFAC* *DICGRKFARSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIR* *THTGEKPFACDICGRKFAQNSTRIGHTKIHLRGS***SGSETPGTSESAT** **PES***GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFAC* *DICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIR* *THTGEKPFACDICGRKFAQGYNLAGHTKIHLRGS***SGSETPGTSESA** **TPES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGS DFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQ VDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALL VNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGE TLHSSLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEE KVMVTVPDKNPPSPSSGTRVNSVLNLIEHLKVSHGKRGVCFRCA

	KCGKENSNYHSVVCHFPKCRGPETEKAPAGEWISEVSNRDFTTK IGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKACATKEEEE LLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEP REEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGR AAAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEE RLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIR TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYL DRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKM DPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDIN NWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVD PHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLL FNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWE NMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAW TINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRID QAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSS EPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGR GIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIV HMFASRARKSMVM (SEQ ID NO: 1677)
Cas9 *S. pyogenes*, прикрепленная в местоположении v2 DBD R2Tg с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH*

	*QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV

	NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 1678)
Cas9 *S. pyogenes*, содержащая каталитические мутации (dCas9), прикрепленная в местоположении v2 DBD R2Tg с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI*

	*DLSQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 1679)
Мутантная форма никазы Cas9 D10A *S. pyogenes*, прикрепленная в местоположении v2 DBD R2Tg с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

	LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI *HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV* *DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE* *LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD* *VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW* *RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH* *VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI* *NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS* *EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW* *DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA* *RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI* *MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS* *AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH* *KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH* *LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI* *DLSQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA

	RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 16780)
Мутантная форма никазы Cas9 N863A *S. pyogenes*, прикрепленная в местоположении v2 DBD R2Tg с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE*

	FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 1681)
Мутантная форма никазы Cas9 D10A *S. pyogenes*, прикрепленная в местоположении v2 DBD R2Tg, содержащей мутацию EN, с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKK* *FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI* *CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV* *AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL* *NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL* *ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD* *TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS* *ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG* *GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH* *QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV* *LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT* *NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD* *KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ* *LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI* *HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV* *DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE* *LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD*

	*VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 1682)*
Мутантная форма никазы Cas9 N863A *S.*	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV*

pyogenes, прикрепленная в местоположении v2 DBD R2Tg, содержащей мутацию EN, с помощью линкера XTEN33aa	*AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN*

	TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM (SEQ ID NO: 1683)
Cas9 *S. pyogenes*, прикрепленная в местоположении v2 DBD R2Tg, содержащего мутацию EN, с помощью линкера XTEN33aa	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKK* *FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI* *CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV* *AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL* *NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL* *ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD* *TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS* *ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG* *GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH* *QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV* *LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT* *NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD* *KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ* *LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI* *HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV* *DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE* *LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD* *VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW* *RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH* *VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI* *NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS* *EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW*

	*DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TAT RDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRG KLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGK ADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPE FSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWR PITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHI VGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMN TNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTIN GTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQA PLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKE WLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIV NFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGR QDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEE AKKKDWVVFKEPHIRDSNKELYKPALIFVKDARALVVDVTVRY EAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGA RGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMF ASRARKSMVM* (SEQ ID NO: 1684)

Пример 4. 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) (APTQKDKFPKPCNWRKNEFKKWTKLAS (SEQ ID NO: 1685)) подвергнут мутации по 1, 2, 3, 4, 5, 6 или 7 остаткам с образованием инактивированного NoLS (APTQ**A**D**A**FP**A**PCNWR**A**NEF**AA**WT**A**LAS (SEQ ID NO: 1686)).

Пример 5. Применение внесения однонитевого разрыва во вторую нить в системе Gene Writer

В данном примере описана система Gene Writer, в которой ретротранспозиция сочетается с целенаправленной активностью внесения однонитевого разрыва во вторую нить для увеличения эффективности событий интеграции. Однонитевой разрыв второй нити может быть получен с помощью (1) никазы Cas9, слитой с системой Gene Writer, в которой Gene Writer вводит один однонитевой разрыв с помощью своего эндонуклеазного домена (EN), а слитая никаза Cas9 размещает другой однонитевой разрыв либо на верхней, либо на нижней нити ДНК (фиг. 7A), или (2) системы Gene Writer, в которой активный домен EN вводит однонитевой разрыв, а никаза Cas9 вводит второй однонитевой разрыв либо на верхней, либо на нижней нити ДНК, выше или ниже от индуцированного Gene Writer однонитевого 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 однонитевой разрыв в ДНК с помощью своего домена EN, а дополнительный однонитевой разрыв создается за счет никазной активности Cas9. Такой дополнительный однонитевой разрыв может быть нацелен на верхнюю или нижнюю нити ДНК, окружающие

однонитевой разрыв, введенный Gene Writer (фиг. 1A). Конструкции, разработанные и испытанные, включают (см. схематическую фиг. 8A):

- 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 неактивен, применяют gRNA, удлиненную на своем 3'-конце с обеспечением включения донорной последовательности для модификации генома (фиг. 8B). 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 однонитевой разрыв в ДНК с помощью своего домена EN, а дополнительный однонитевой разрыв создается за счет никазной активности Cas9. Такой дополнительный однонитевой 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. 9A), и матрица, применяемая для модификации генома, представляет собой трансген, фланкированный UTR (фиг. 9B). Конструкции включают (см. схематическую фиг. 9A):

- 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/делеций. Сайт однонитевого разрыва, осуществляемого доменом EN, идентифицируют по вставкам/делециям, которые домен EN осуществляет в сайте AAVS1. Плечи гомологии длиной 100 п. о., фланкирующие сайт однонитевого разрыва, осуществляемого EN, разрабатывают и включают в трансген (см. фиг. 9 и 10, где показано местоположение плеч гомологии в трансгене). 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 данном примере осуществляют нуклеофекцию плазмиды Gene Writer, нацеливающаейся на сайт AAVS1 (фиг. 10A), и фланкированного UTR трансгена с гомологией сайту AAVS1 (фиг. 10B), в 200 тыс. клеток U2OS с применением импульсного кода DN100. Трансфекцию следующих конструкций на основе Cas9 осуществляли вместе с плазмидами Gene Writer (фиг. 10C):

- 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 однонитевого 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.

Пример 6. Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR

В данном примере описано применение гетерологичных UTR для усиления внутриклеточной экспрессии полипептида Gene Writer.

В данном примере полипептид Gene Writer экспрессировали из мРНК (фигура 11). В плазмидной матрице для получения мРНК нативные UTR ретротранспозона заменяли с помощью UTR, оптимизированными 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: 1687). Плазмиду линеаризовали 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'-O-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-матрицу транскрибировали in vitro с плазмиды, как описано в примере 8 ("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. Обнаружили, что экспрессия белка значительно улучшается при применении 5'- и 3'-UTR_{exp} из C3-ORM по сравнению с применением нативных UTR из R2Tg (фигура 11). Интеграцию в геном анализировали через 3 дня после трансфекции с применением 3'-ddPCR (фигура 12).

Пример 7. 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 данном примере полипептидный компонент системы Gene Writer™ экспрессируется из мРНК, описанной в примере 6 ("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/лунка 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 высеянными 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. Плазмида, 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 13. 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.

Пример 8. 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 данном примере полипептидный компонент системы Gene Writer™ экспрессируется из мРНК, описанной в примере 6 ("Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR"). Плазмидная матрица для получения РНК-матрицы содержала промотор T7, 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 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 РНК-полимеразы T7 (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.

Пример 9. 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 15), и в зависимости от местоположения 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 сплайсосомой 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 подвергать 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 местоположении, показанном на фигуре 15, будет обеспечивать рекрутирование 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 данном примере как РНК-матрицы, несущие интрон или интроны, так и полипептид Gene Writer доставляются в клетки в виде транскрибированных *in vitro* кэпированных РНК, описанных в примере 8 ("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.

Пример 10. 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 активность саморасщепляющего рибозима, наличие которой было показано в определенных элементах и предсказано в дополнительных ретротранспозонах (см. модули B и C на фигурах 26-27) (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 28.

С целью оценки конструирования 5'UTR были разработаны конструкции для усиления или ослабления данных активностей (фигура 14). В случае (A) природный 5'UTR R2Tg применяют для интеграции в транс-положении, как и в предыдущих экспериментах. Случай (B) иллюстрирует делецию 5'UTR. (C) 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 (C) 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 14.B, 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.

Пример 11. 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 данном примере полипептидный компонент системы Gene Writer™ экспрессируется из мРНК, описанной в примере 6 ("Улучшение экспрессии полипептида Gene Writer с помощью гетерологичных UTR"). Плазмидная матрица для получения РНК-матрицы содержала промотор T7, 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 T7 (NEB). После стадии IVT РНК обрабатывали ДНКазой I (NEB) и либо ферментативно полиаденилировали поли(А)-полимеразой (NEB), либо нет. После стадии замены буфера реакцию ферментативного кэпирования, приводящую к получению структуры кэпа 1, проводили, как описано в примере 8 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"), или не проводили. РНК-матрицу концентрировали и заменяли буфер. 50000 клеток HEK293T 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 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 через 3'-конец мест соединения интегрированного элемента. Интеграцию в геном улучшали при применении матрицы с ферментативным кэпированием и поли(А) хвостом (фигура 16).

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 кэпирование Gene Writer и РНК-матрицы будут проводить, как описано в примере 8 ("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 (Whifield 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 включат dNTP, 2'O-метилированные NTP или фосфоротиоат-NTP в 3'-конец РНК трансгенов для увеличения периода полужизни данных молекул за счет защиты 3'-конца РНК от экзорибонуклеаз. Авторы настоящего изобретения включат 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 Т4 или РНК-лигазы Т4 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 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"), будут исследоваться другие защитные группы 5'-конца. В частности, авторы настоящего изобретения будут применять гиперметилированные (Wurth et al. Nucleic Acid Res 42, 8663-8677 (2014)), фосфоротиоатные (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 в геном будут тестировать, как описано в примере 8 ("Улучшенные компоненты Gene Writer для доставки на основе РНК").

Пример 12. 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 rNTP на один из модифицированных rNTP: псевдоуридин (Ψ), 1-N-метилпсевдоуридин (1-Me-Ψ), 5-метоксиуридин (5-MO-U) или 5-метилцитидин (5mC). В противном случае получение, очистку РНК и трансфекции ими клеток проводили, как описано в примере 8 ("Улучшенные компоненты Gene Writer для доставки на основе РНК"). Способность интеграции генов модифицированных мРНК сравнивали со способностью немодифицированной мРНК (G0), применяя ddPCR, при этом все мРНК полипептида рассматривали в паре с немодифицированной РНК-матрицей (фигура 17). Интеграцию обнаруживали, в случае когда полипептид кодировали с применением каждого модифицированного rNTP, 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 матричной молекулы Gene Writer, где она состоит из всех или подмножества иллюстративных модулей, перечисленных на фигуре 19 и проиллюстрированных на фигуре 18. Отдельные модули могут быть получены посредством химического синтеза или синтеза in vitro в виде непрерывной молекулы нуклеиновой кислоты или в виде раздельных фрагментов, которые впоследствии объединяются вместе. Отдельные модули матричной молекулы 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 матричной молекулы Gene Writer.

В некоторых вариантах осуществления матричную молекулу 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'-O-метилом, 2'-O-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'-дезоксицитидином, 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 матричная молекула Gene Writer обычно полностью или частично совместима с активностью обратной транскриптазы полипептидной последовательности Gene Writer; для модулей или частей модулей матричной молекулы 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 матричными молекулами, содержащими модифицированные rNTP, тестируют, как описано выше и в примере 8 ("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'-O-метил, 2'-O-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'-дезоксицитидин, 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 матричной молекулы Gene Writer.

Примером модифицированной матричной молекулы 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) фосфоротиоатные связи между первыми 3 нуклеотидами на 5‘-конце и модуль F содержит (3) фосфоротиоатные 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.

Пример 13. Матрицы Gene Writer, которые не содержат UTR

В данном примере описана конфигурация матричной молекулы 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 матричной молекулы Gene Writer (фигуры 18 и 19), приводящие к тому, что управляющий элемент Gene Writer не включает модули UTR в геном в качестве части ретротранспозиции. В некоторых вариантах осуществления модули матричной молекулы Gene Writer для 5‘- и 3‘-UTR (модули B+C и E матричной молекулы 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 B и/или C пропущены в матричной молекуле Gene Writer, при этом модуль E следует за модулем F.

Дополнительными примерами невключения UTR в геном являются удаление модулей B и C из матричной молекулы Gene Writer, изменение положения модуля F (3'-плечо гомологии) таким образом, чтобы он следовал за модулем D (последовательностью гетерологичного объекта), и замена модуля E на связывающий лиганд, такой как биотин. Данная матричная молекула Gene Writer в данном случае будет состоять из модуля A (5'-плеча гомологии) - модуля D (последовательности гетерологичного объекта) - модуля F (3'-плеча гомологии) - модуля E, 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, матричной молекулы Gene Writer с полипептидной последовательностью управляющего элемента Gene Writer.

Пример 14. Длина плеча гомологии влияет на эффективность ретротранспозиции

В данном примере описывают модулирование плеч гомологии (HA), фланкирующих трансген, для увеличения частоты ассоциированных событий ретротранспозиции.

Считается, что ретротранспозиция опосредована событиями праймирования на 3'-конце интегрированного трансгена. Праймирование трансгенной РНК с помощью полученной в результате однонитевого разрыва геномной ДНК хозяина требует наличия гомологии между 3'-концом трансгенной РНК и геномной ДНК, находящейся в направлении 3'-конца от однонитевого разрыва в ДНК хозяина. Хотя способ 5'-концевого разрешения ретротранспозиции неизвестен, обеспечение гомологии посредством путей репарации, опосредованных хозяином, может также оказывать благоприятное влияние на данное разрешение. Кроме того, процессинг 5'-конца трансгенной РНК может влиять на ретротранспозицию, т. е. посредством рибозимного расщепления выше 5'-UTR. Следовательно, последовательность гомологии, фланкирующую трансгенную нагрузку, модулировали для оптимизации ретротранспозиции.

Трансфекции плазмид выполняли для тестирования эффектов последовательности гомологии трансгена в отношении эффективности интеграции в *транс*-положении. Плазмиду, экспрессирующую R2Tg, или контрольную R2Tg с мутацией, обеспечивающей инактивацию эндонуклеазы, доставляли совместно с плазмидами с трансгеном, содержащими последовательности гомологии с мишенью различной длины, и необязательную случайную спейсерную последовательность ДНК (таблица E5 ниже, фигура 20). Спейсерную последовательность применяли, чтобы контролировать эффект длины транскрипта в отношении ретротранспозиции. Всего 500 нг плазмид при молярном соотношении плазмида с управляющим элементом:плазмида с трансгеном, составляющим 1:4, подвергали нуклеофекции в 2e5 клеток 293T посредством челночного вектора Lonza 4D с помощью программы DS-150. После 3 дней выделяли геномную ДНК клеток. Цифровую капельную ПЦР (ddPCR) проводили для каждого места соединения с применением праймера и зонда Taqman, специфичных для трансгена, в комбинации с праймером в ожидаемом локусе рДНК. Установленное число копий представляет собой эффективность интеграции на 3'- или 5'-конце трансгена.

Результаты показаны на фигуре 21. Протестированная конструкция с наиболее длинной последовательностью гомологии (100 п. о. с каждой стороны) продемонстрировала наибольшую эффективность интеграции.

Таблица E5. Протестированные конструкции матриц с различной длиной последовательности гомологии на 5'- и 3'-концах, включая спейсерную последовательность для поддержания общей длины фланкирующей последовательности на уровне 100 п. о. на каждом конце

Длина плеча гомологии на 5'-конце (п. о.)	Длина плеча гомологии на 3'-конце (п. о.)	Наличие случайного спейсера (до 100 п. о.)
100	100	n/a
0	0	N
0	0	Y
5	5	N
5	5	Y
10	10	N
10	10	Y
20	20	N
20	20	Y
50	50	N
50	50	Y
90	90	N
90	90	Y

Пример 15. Переносимость и специфичность плеча гомологии влияют на эффективность ретротранспозиции

В данном примере описана необходимость точного выравнивания конструкции плеча гомологии с местоположением однонитевого разрыва во время ретротранспозиции.

Данные из приведенного выше примера 14 подчеркивают важность плеча гомологии для эффективной ретротранспозиции. Во многих случаях местоположение разрыва данного ретротранспозона может быть плохо охарактеризовано. В таких случаях разработанные плечи гомологии могут не инициироваться рядом с сайтом однонитевого разрыва в геномной ДНК. Для оценки зависимости положения плеча гомологии от эффективности ретротранспозиции, авторы настоящего изобретения разработали конструкции с плечом гомологии на 3'-конце, сдвинутым на несколько оснований относительно известного сайта разреза R2Tg (таблица E6 ниже, фигура 22). В тех случаях, когда 3'-плечо гомологии сдвигали в направлении 3'-конца от сайта разреза (+), плечо гомологии успешно укорачивалось. В случаях, когда 3'-плечо гомологии сдвигали в направлении 5'-конца от сайта разреза, гомологичные основания добавлялись из 5'-плеча.

Клетки 293T котрансфицировали плазмидами с управляющим элементом и плазмидами с трансгеном, как в вышеприведенном примере. В день 3 экстрагировали геномную ДНК. Частоту интеграции измеряли, как указано выше, с помощью ddPCR. Значительное снижение частоты интеграции отмечали при смещении 3'-плеча гомологии в любое положение относительно сайта однонитевого разрыва WT (фигура 23).

Таблица E6. Протестированные конструкции матрицы с различными положениями последовательности гомологии на 3'-конце, включая относительный сдвиг положения последовательности на 3'-конце

Длина плеча гомологии на 5'-конце (п. о.)	Длина плеча гомологии на 3'-конце (п. о.)	Сдвиг правого плеча (относительно ожидаемого сайта однонитевого разрыва)
100	100	0
100	99	+1
100	97	+3
100	95	+5
100	90	+10
100	101	-1
100	103	-3
100	105	-5
100	110	-10

Пример 16. 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 A, B). 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 в день 3, показали отсутствие Rad51 для условия обработки с помощью siRad51 (фиг. 24C). 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 25. Отсутствие Rad51 приводит к снижению на ~20% интеграции трансгена, опосредованной R2Tg, в локусе рДНК как в 3'-, так и в 5'-концевых местах соединения (фиг. 25), указывая на то, что интеграция трансгена, опосредованная R2TG, 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 G_0 или G_1, 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.

Пример 17. Gene Writer могут интегрировать генетический груз независимо от пути репарации посредством однонитевой матрицы

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации посредством однонитевой матрицы (SSTR) является ингибированным.

В таком примере путь SSTR будут ингибировать с помощью siRNA к основным компонентам пути: FANCA, FANCD2, FANCE, USP1. Также будут включены контрольные siРНК ненацеленного контроля. Нуклеофекцию 200 тысяч клеток U2OS будут осуществлять с использованием 30 пмоль (1,5 мкМ) siRNA, а также с использованием плазмиды с управляющим элементом R2Tg и плазмиды с трансгеном (транс-конфигурация). В частности, 250 нг плазмид, экспрессирующих R2Tg, контрольный R2Tg с мутацией в домене RT или контрольный R2Tg с мутацией, обеспечивающей инактивацию эндонуклеазы, применяют в сочетании с трансгеном в молярном соотношении 1:4 (управляющий элемент к трансгену). Трансфекции клеток U2OS проводят в буфере SE с применением программы DN100. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят ddPCR для оценки интеграции в сайте рДНК. Интеграцию трансгена в рДНК обнаруживают по отсутствию основных компонентов пути SSTR.

Пример 18. 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 плазмида, 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 (uguaguguuuccuacuuuaugga (SEQ ID NO: 1688)), 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.

(HEK293T по сравнению с HEK293T с miRNA?)

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.

Пример 19. Применение системы Gene Writer™ для доставки терапевтического гена в печень в мышиной модели химерной печени человека.

В данном примере описана система редактирования генома Gene Writer™, доставляемая в печень in vivo для интеграции и стабильной экспрессии генетической нагрузки. Промотор и последовательность распознавания miRNA для гена контроля экспрессии и терапевтического гена предназначены для иллюстрации подхода и выбраны из таблицы 2 документа WO2020014209, включенного в данный документ посредством ссылки, и таблиц 3 и 4 документа WO2020014209 соответственно.

В данном примере гепатоциты человека, полученные от пациентов с дефицитом OTC, прививают мышиной модели (Ginn et al JHEP Reports 2019), и при этом систему Gene Writer™ применяют для доставки кассеты экспрессии OTC для интеграции в клетки печени. Полипептидный компонент Gene Writer™ предусмтаривает кассету экспрессии для ретротранспозазы R2Tg (таблица 3), и матричный компонент предусматривает кассету экспрессии для гена OTC человека (таблица 5 документа WO2020014209), фланкированного последовательностями UTR, необходимыми для связывания и ретротранспозиции с помощью R2Tg (таблица 3), и дополнительно фланкированного последовательностями гомологии с сайтом-мишенью длиной 100 нуклеотидов в рибосомной ДНК. В данном примере кассеты экспрессии как транспозазы, так и матрицы, дополнительно содержат промотор hAAT (таблица 3 документа WO2020014209) для специфичной для гепатоцитов экспрессии и последовательность распознавания miRNA, комплементарную затравочной последовательности miR-142 (таблица 4 документа WO2020014209), для подавления экспрессии в гемопоэтических клетках.

1. Полипептидный компонент Gene Writer™: rAAV2/NP59.hAAT.R2Tg

2. Полипептид Gene Writer™ с эндонуклеазой, подвергнутой мутации: rAAV2/NP59.hAAT.R2TgEN*

3. Матричный компонент Gene Writer™: rAAV2/NP59.hAAT.OTC

4. Матричный компонент Gene Writer™ с репортером: rAAV2/NP59.hAAT.GFP

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™, упакованным в NP59, капсид AAV с высокой гепатотропностью у человека. Следующие векторы вводят с помощью внутрибрюшинной инъекции:

- Активный Gene Writing™ терапевтического средства: (1) и (3);
- Активный Gene Writing™ репортера: (1) и (4);
- Интеграционно-инактивированный контроль терапевтического средства: (2) и (3);
- Интеграционно-инактивированный контроль репортера: (2) и (4).

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 OTC впоследствии анализируют посредством проведения RT-qPCR на выделенных образцах РНК с применением праймеров, специфических в отношении последовательности. Для подтверждения интеграции конструкции и анализа местоположений в геноме выполняют однонаправленное секвенирование на образцах геномной ДНК с применением специфических праймеров, отжигающихся с вставленным геном, для считывания наружу в окружающую геномную последовательность на MiSeq.

Пример 20. Применение системы Gene Writer™ для доставки терапевтического гена в печень в модели заболевания на новорожденных или взрослых мышах.

В данном примере описана система редактирования генома Gene Writer™, доставляемая в печень in vivo для интеграции и стабильной экспрессии генетической нагрузки. Промотор и последовательность распознавания miRNA для гена контроля экспрессии и терапевтического гена предназначены для иллюстрации подхода и выбраны из таблиц 2, 3 и 4 документа WO2020014209 соответственно.

В данном примере мышиную модель с дефицитом OTC применяют для оценки системы Gene Writer ™, предназначенной для доставки кассеты экспрессии OTC для интеграции в клетки печени. Полипептидный компонент Gene Writer™ предусмтаривает кассету экспрессии для ретротранспозазы R2Tg (таблица 3), и матричный компонент предусматривает кассету экспрессии для гена OTC человека (таблица 5 документа WO2020014209), фланкированного последовательностями UTR, необходимыми для связывания и ретротранспозиции с помощью R2Tg (таблица 3), и дополнительно фланкированного последовательностями гомологии с сайтом-мишенью длиной 100 нуклеотидов в рибосомной ДНК. В данном примере кассеты экспрессии как транспозазы, так и матрицы, дополнительно содержат промотор hAAT (таблица 3 документа WO2020014209) для специфичной для гепатоцитов экспрессии и последовательность распознавания miRNA, комплементарную затравочной последовательности miR-142 (таблица 4 документа WO2020014209), для подавления экспрессии в гемопоэтических клетках.

1. Полипептидный компонент Gene Writer™: rAAV2/8.hAAT.R2Tg

2. Полипептид Gene Writer™ с эндонуклеазой, подвергнутой мутации: rAAV2/8.hAAT.R2TgEN*

3. Матричный компонент Gene Writer™: rAAV2/8.hAAT.OTC

4. Матричный компонент Gene Writer™ с репортером: rAAV2/8.hAAT.GFP

Самкам мышей *Spf*^{ash} с дефицитом Otc в возрасте либо одного-двух дней, либо восьми-12 недель (фоновый генотип C57BL/6/C3H-F1) вводили Gene Writer™, упакованные в AAV8, гепатотропный капсид AAV. Следующие векторы вводят с помощью внутрибрюшинной инъекции:

- Активный Gene Writing™ терапевтического средства: (1) и (3);
- Активный Gene Writing™ репортера: (1) и (4);
- Интеграционно-инактивированный контроль терапевтического средства: (2) и (3);
- Интеграционно-инактивированный контроль репортера: (2) и (4).

Впоследствии, через 5 недель, ДНК и РНК экстрагируют из лизатов печени стандартными способами. Экспрессию OTC впоследствии анализируют посредством проведения RT-qPCR на выделенных образцах РНК с применением праймеров, специфических в отношении последовательности. Для подтверждения интеграции конструкции и анализа местоположений в геноме выполняют однонаправленное секвенирование на образцах геномной ДНК с применением специфических праймеров, отжигающихся с вставленным геном, для считывания наружу в окружающую геномную последовательность на MiSeq.

Пример 21. Совместимость рибозима и последовательности плеча гомологии в сайтах, в отношении которых осуществляли перенацеливание

В данном примере описана матричная молекула Gene Writer, которую применяют в сочетании с мутантной полипептидной последовательностью Gene Writer, нацеленной на местоположение в геноме, находящееся за пределами его природного сайта-мишени в геноме. Эндогенные последовательности РНК ретротранспозона содержат рибозим на 5'-конце своей РНК, который активен только в том случае, если РНК образует правильную вторичную структуру, или в случае правильного сворачивания РНК (Eickbush, D. G., et. al, Molecular and Cellular Biology, 30 (13), 3142-3150, 2010; Eickbush, D. G., et. al., PLoS ONE, 8(9), 1-16, 2013; Ruminski, D. J., et. al., Journal of Biological Chemistry, 286(48), 41286-41295, 2011). Для образования правильной структуры активной формы рибозима ретротранспозона РНК некоторых ретротранспозонов должна содержать часть рибосомной РНК 28S для обеспечения образования правильной вторичной структуры стебля P1 рибозима (Eickbush, D. G., et. al., PLoS ONE, 8(9), 1-16, 2013).

Часть РНК эндогенного ретротранспозона, представляющая собой рибосомную РНК 28S, которая взаимодействует с 5'-UTR РНК ретротранспозона, является аналогичной матричной молекуле Gene Writer (модульность на фигуре 18), где часть последовательности в модуле A взаимодействует с частью последовательности модуля B матричной молекулы Gene Writer. С целью того, чтобы модуль B был активным рибозимом, он должен свернуться в правильную вторичную структуру, часть P1' рибозима, находящаяся в модуле

В, взаимодействует с некоторой комплементарностью с последовательностью P1 рибозима, находящейся в модуле А. В некоторых вариантах осуществления, где 5'-плечи гомологии матричной молекулы Gene Writer (модуль А) необходимы для интеграционной активности наряду с наличием активного рибозима (модуль В), последовательность последовательности P1' рибозима, находящуюся в модуле В, изменяют для обеспечения некоторой комплементарности последовательности, находящейся в модуле А, где находится последовательность P1. Длины комплементарности, в нуклеотидах, между последовательностью P1, находящейся в модуле А, и последовательностью P1' модуля В могут варьировать в пределах от 0 до 100 нуклеотидов. В некоторых вариантах осуществления, если перенацеленная полипептидная последовательность Gene Writer основана на элементе R2 из *Taeniopygia guttata*, то нуклеотиды 49-54 модуля В матричной молекулы Gene Writer взаимодействуют посредством полной или частичной комплементарности с последними 28 нуклеотидами модуля А, где модуль А представляет собой плечо гомологии, которое комплементарно местоположению в геноме, которое совместимо с мутантной полипептидной последовательностью Gene Writer, нацеленной на геномную область, в отношении которой модуль А характеризуется комплементарностью (фигура 28, предсказанный рибозим).

Пример 22. Система Gene Writing интегрирует специфическую последовательности матрицы

В данном примере описан анализ вставок, выполненных системой Gene Writing, с целью анализа в отношении непреднамеренного включения РНК, не являющейся матрицей, например, клеточной эндогенной РНК, в сайт-мишень.

В данном примере систему Gene Writing применяют, как описано в предыдущих примерах, с целью интеграции РНК-матрицы в сайт-мишень в клетках HEK293. Клетки HEK293 трансфицировали следующими реагентами:

1. мРНК, кодирующей полипептид Gene Writer;

2. мРНК, кодирующей инактивированный полипептид Gene Writer (например, мутантная форма обратной транскриптазы R2Tg);

3. РНК-матрицей Gene Writer (например, содержащей в направлении от 5'- к 3'-концу: 5'HA - 5'UTR - кассета GFP - 3'UTR - 3'HA);

4. РНК-матрицей Gene Writer без мотива связывания (например, содержащей в направлении от 5'- к 3'-концу: 5'HA - кассета GFP - 3'HA).

После 3 дней инкубации геномную ДНК экстрагируют и анализируют в отношении частоты вставок с помощью ddPCR, как описано в другом месте в данном документе. В некоторых вариантах осуществления комбинация (1) и (2) будет приводить к интеграции матрицы. В некоторых вариантах осуществления комбинация (1) и (4) не будет приводить к обнаруживаемой интеграции матрицы, поскольку матрица (4) не содержит мотива связывания полипептида, например, 3'UTR из ретротранспозона R2Tg. В некоторых вариантах осуществления комбинация (1) и (4) будет приводить к меньшей частоте

интеграции, чем частота интеграции при использовании комбинации (1) и (2), например, меньше 10%, 5%, 4%, 3%, 2%, 1%, 0,05% или 0,01%.

Для дополнительного анализа всех включенных последовательностей выполняется секвенирование ампликона в клетках, полученных в результате трансфекции (1) и (2), с применением ПЦР для амплификации последовательности, проходящей через место соединения сайта-мишени. Необязательно может быть выполнен отрицательный отбор в отношении неотредактированных сайтов-мишеней посредством расщепления конкретно по месту соединения неотредактированного сайта-мишени (сайт однонитевого разрыва Gene Writer) для улучшения сигнала. Ампликоны обрабатывают для секвенирования следующего поколения на приборе Illumina MiSeq, как описано ранее. Непреднамеренные включения обнаруживают посредством поиска ридов, которые содержат вставленную последовательность, например, по меньшей мере 100 нт, новой ДНК, которая не картируется на РНК-матрице. Необязательно вставку сравнивают с транскриптомом человека для определения источника любого транскрипта, непреднамеренно включенного в сайт-мишень. В некоторых вариантах осуществления система Gene Writing не будет включать в сайт-мишень какие-либо матрицы, которые не представляют собой РНК-матрицу Gene Writing. В некоторых вариантах осуществления система Gene Writing не будет включать в сайт-мишень какие-либо матрицы, которые не представляют собой РНК-матрицу Gene Writing, на уровне, превышающем 1% от общего числа вставок.

Пример 23. 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 недостаточностью альфа-1-антитрипсина. Полипептид Gene Writer™ использует никазу Cas9 как 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 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: 1689), где числа применяются для обозначения модулей матрицы в порядке (5'-3'): (1) спейсер gRNA, (2) каркас gRNA, (3) последовательность гетерологичного объекта, (4) 3'-гомологичный домен праймирования, и строчная буква "c" указывает положение в матрице, несущее нуклеотидную замену, которая должна быть записана в сайт-мишень для исправления мутации E342K. Иллюстративная gRNA для обеспечения второго однонитевого разрыва, как описано в вариантах осуществления данной системы, содержит спейсерную последовательность TTTGTTGAACTTGACCTCGG (SEQ ID NO: 1625) и направляет никазу Cas9 для введения однонитевого разрыва во вторую нить сайта-мишени в пределах гомологичной области. В некоторых вариантах осуществления данный второй однонитевой разрыв обеспечивает улучшение эффективности редактирования.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и конверсии 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.

Пример 24. Коррекция недостаточности 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 недостаточность альфа-1-антитрипсина.

Составление LNP (система LNP-INT01), несущих Cas9 и gRNA, и обработка ими мышиных 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 линеаризованной 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 NanoAssemblr в соответствии с протоколом производителя. После смешивания LNP собирают и разбавляют в PBS (приблизительно 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 мм. Фильтрат хранят при температуре 2-8°C. Составы с несколькими дозами можно изготовить с применением 25 мМ цитрата, буфера для груза с 100 мМ NaCl (pH 5) и осуществить замену буфера посредством TFF на Tris-солевой сахарозный буфер (TSS) (5% сахарозы, 45 мМ NaCl и 50 мМ 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.

Пример 25. 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 гентингтина (*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. Могут быть разработаны многочисленные примеры таких РНК-матриц, при этом иллюстративная РНК-матрица, кодирующаяся в ДНК, содержит последовательность
(1)GGCGGCTGAGGAAGCTGAGG(2)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)AGTCCCTCAAG TCCTTCcagcagcagcagcagcagcagcagcagccgccaccgccgccgccgccgccgcctcct(4)CAGCTTCC TCAG(SEQ ID NO: 1690), где числа применяются для обозначения модулей матрицы в порядке (5'-3'): (1) спейсер gRNA, (2) каркас gRNA, (3) последовательность гетерологичного объекта, (4) 3'-гомологичный домен праймирования, 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 однонитевого разрыва, как описано в вариантах осуществления данной системы, содержит спейсерную последовательность CGCTGCACCGACCGTGAGTT (SEQ ID NO: 1647) и направляет никазу Cas9 для введения однонитевого разрыва во вторую нить сайта-мишени в пределах гомологичной области. В некоторых вариантах осуществления данный второй однонитевой 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 никазы Cas9, слитый с интеином-N расщепленной пары интеинов (DnaE Интеин-N: CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN(SEQ ID NO: 1613)), и второй кодирует домен RT, слитый с интеином-C расщепленной пары интеинов (DnaE интеин-C, MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN(SEQ ID NO: 1615)), 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 однонитевого 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 rAAV2/1, 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 rAAV2/1-Cas9 и вируса rAAV-MMLV_RT/hU6templateRNA в соотношении 1:1. Для инъекций rAAV мышей анестезируют изофлураном и 5 мкл смеси rAAV 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.

Пример 26. 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 мыши с помощью комбинации липидных наночастиц (мРНК, кодирующую полипептид) и AAV (ДНК, кодирующую РНК-матрицу) для коррекции мутации SERPINA1 PiZ, обуславливающей недостаточность 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-матрицу 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 однонитевого разрыва gRNA, управляемой промотором U6 (последовательности РНК, ранее описанные в примере 1).

Мыши 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 ~10^{11} 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.

Пример 27. Применение системы Gene Writer™ для доставки терапевтического гена в печень в мышиной модели химерной печени человека.

В данном примере описана система редактирования генома Gene Writer™, доставляемая в печень in vivo для интеграции и стабильной экспрессии генетической нагрузки. В частности, LNP применяют для доставки системы Gene Writing, способной интегрировать полную кассету экспрессии OTC для лечения недостаточности OTC в гуманизированной мышиной модели.

В данном примере систему Gene Writing применяют для лечения недостаточности OTC в гуманизированной мышиной модели, в которой гепатоциты человека, полученные от пациентов с недостаточностью OTC, 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) область гомологии с мишенью на 3'-конце для отжига с геномной ДНК непосредственно перед однонитевым разрывом первой нити, чтобы праймировать TPRT последовательности гетерологичного объекта. Иллюстративной последовательностью для (1) является GGGGCCACTAGGGACAGGAT(SEQ ID NO: 1691). Область (2) несет каркас gRNA, описанную в данной заявке, обычно являющуюся последовательностью GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGGACCGAGTCGGTCC (SEQ ID NO: 1603). В данном примере (3) содержит полную кассету экспрессии OTC, где последовательность, кодон-оптимизированная для печени, кодирующая OTC человека (UniProt P00480), функционально ассоциирована с

промоторной системой ApoE.hAAT, описанной в примере 25. Иллюстративной последовательностью для (4) является CTGTCCCTAGTG (SEQ ID NO: 1692). Иллюстративной последовательностью дополнительного спейсера gRNA для создания однонитевого разрыва во второй нити с целью увеличения эффективности интеграции является AGAGAGATGGCTCCAGGAAA(SEQ ID NO: 1693).

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 OTC впоследствии анализируют посредством проведения RT-qPCR на выделенных образцах РНК с применением праймеров, специфических в отношении последовательности. Уровни OTC человека также измеряли в плазме крови на протяжении всего эксперимента с применением набора OTC ELISA для человека (например, набора Aviva Systems Biology OTC 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 нормализуется 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 OTC.

Пример 28. Gene Writer для интеграции CAR в Т-клетки ex vivo

В данном примере описывают доставку системы редактирования генома Gene Writer™ в Т-клетки ex vivo для интеграции и стабильной экспрессии генетической нагрузки. В частности, LNP применяют для доставки системы Gene Writing, способной интегрировать химерный антигенный рецептор (CAR) в локус *TRAC* для создания CAR-T-клеток для лечения В-клеточной лимфомы.

В данном примере система Gene Writing содержит полипептид Gene Writing, например, домен никазы Cas9 и домен обратной транскриптазы R2Tg, описанные в данном документе, gRNA для направления активности никазы 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 однонитевого 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 однонитевого 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 данном примере направляют к 5'-концу первого экзона *TRAC* с помощью нацеленной gRNA, например, TCAGGGTTCTGGATATCTGT(SEQ ID NO: 1694), чтобы поместить груз под эндогенный контроль экспрессии из того локуса, одновременно разрушая эндогенный TCR, как указано в Eyquem et al. *Nature* 543:113-117 (2017). Все данные три компонента (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 siРНК, придающие полученным гидрофобно модифицированным молекулам siРНК 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 NK-клетками иммунных ответов и контроля поглощения макрофагами, например, PD-L1, HLA-G, CD47 (Han et al. PNAS 116(21):10441-10446 (2019)). Нокин может быть осуществлен 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 применением gRNA GGGGCCACTAGGGACAGGAT для нацеливания полипептида 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-T-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-T-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-T-клеток ex vivo с желаемой цитотоксической активностью.

Пример 29. 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-T-клеток в мышиной модели.

В данном примере система Gene Writing содержит полипептид Gene Writing, например, домен никазы Cas9 и домен обратной транскриптазы R2Tg, описанные в данном документе, gRNA для направления активности никазы 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 однонитевого 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 однонитевого 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 данном примере направляют к локусу Rosa26 мыши с применением gRNA, например, ACTCCAGTCTTTCTAGAAGA(SEQ ID NO: 1695) (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: 1604) и 3'-UTR (например, UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA(SEQ ID NO: 1605)), 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.

Пример 30. Мутация ДНК-связывающих мотивов ретротранспозазы предотвращает интеграцию в их нативный сайт

В данном примере внутренние свойства ДНК-связывания ДНК-связывающего домена полипептида Gene Writer на основе ретротранспозона анализируют по мутагенезу или укорочению области. Создание таких мутантных форм и обнаружение сниженной

активности полезны для понимания детерминант нацеливания ДНК, а также для создания мутантных производных с уменьшенной или исключенной функцией, которые затем можно применять в качестве каркасов для продуктов слияния, обеспечиваемых гетерологичными ДНК-связывающими доменами. В данном документе показано, что мутации в С-концевых доменах цинковых пальцев, а также в домене c-myb оказывают сильное влияние на ферментативную активность ретротранспозазы R2Tg в ее нативной последовательности распознавания.

Чтобы проверить, приводят ли мутантные формы R2Tg с измененной ДНК-связывающей способностью к устранению ДНК-связывающей активности, были сконструированы иллюстративные мутантные формы и оценены в отношении активности интеграции в сайт-мишень нативной рДНК в виде считывания активности связывания с ДНК-мишенью. В данном примере исходные и мутантные (*) домены были названы и сконструированы следующим образом:

- ZF1: CPCCGTRVNSVLNLIEHLKVSH
- ZF1*: SPSSGTRVNSVLNLIEHLKVSH
- ZF2: CEVCNRDFTTKIGLGQHKRLAH
- ZF2*: SEVSNRDFTTKIGLGQHKRLAH
- c-myb: RCWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISD
- c-myb*: ACATKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISD

Общая структура данных доменов показана на фигуре 30A. В данном документе мутантные полипептиды Gene Writer кодировались в матрице Gene Writer, что означает, что последовательность, кодирующая полипептид, была дополнительно фланкирована полученными из R2Tg 5'-UTR и 3'-UTR и последовательностями гомологии с нативным сайтом-мишенью длиной 100 нуклеотидов, как описано в настоящем изобретении.

Чтобы проверить, влияют ли мутации в ДНК-связывающих доменах R2Tg на функцию R2Tg Gene Writer, 250 нг плазмиды, содержащей исходный полипептид R2Tg, мутантную форму с инактивированной эндонуклеазой (отрицательный контроль активности Gene Writer), каждую отдельную мутацию DBD или все три мутации DBD подвергали нуклеофекции в клетки HEK293T с применением системы Lonza Amaxa Nucleofector 96 Well Shuttle System с использованием программы DS150 в соответствии с протоколами производителя. После нуклеофекции клетки выращивали при 37°C, 5% CO_2 в течение 3 дней перед лизисом клеток и экстракцией геномной ДНК. С помощью ddPCR оценивали экстрагированную gDNA на предмет интеграции матрицы Gene Writer в нативный сайт рДНК. В данном документе заметили, что мутации в доменах ZF1, ZF2 или c-myb ингибируют интеграцию, при этом величина эффекта составляет ZF2>c-myb>ZF1 (фигура 30B).

Пример 31. Определение сайта расщепления эндонуклеазного домена Gene Writer по сигнатуре вставки/делеции

В данном примере описывают клеточный анализ для определения сайта активности расщепления полипептида Gene Writer, содержащего эндонуклеазный домен (EN) в

геномной ДНК-мишени. В частности, показано, что активность ретротранспозазы может вызывать низкий уровень модификации сайта-мишени, вероятно, в результате репарации ДНК хозяина однонитевых разрывов мишени. Применяя чрезвычайно чувствительный и направленный анализ секвенирования ампликона, данную сигнатуру можно оценить для определения местоположения сайта расщепления. Способность быстро оценить раннюю стадию в соответствующих системах Gene Writing позволяет анализировать и конструировать ДНК-специфичность фермента, например, специфичное для последовательности связывание ДНК или специфичную для последовательности эндонуклеазную активность, не завися от полной интеграции, которая может зависит от других свойств системы.

С целью создания профиль сайта расщепления, создали анализ для анализа модификаций геномной последовательности в предполагаемом сайте расщепления ретротранспозазы R2Tg. Схема последовательности-мишени R2Tg, изображающая предполагаемые области связывания ДНК и ожидаемый сайт расщепления эндонуклеазой, показана на фигуре 31 А. Чтобы определить, можно ли обнаружить эндонуклеазную активность R2Tg вокруг предполагаемого сайта-мишени, 73 нг плазмиды экспрессии R2Tg были нуклеофицированы в 200000 клеток U2OS с применением программы DN100 и буфера SE нуклеофектора Amaxa Lonza в соответствии с инструкциями производителя. Клетки культивировали в DMEM с 10% FBS в течение трех дней после нуклеофекции перед выделением геномной ДНК. Ампликоны создавали с применением праймеров, фланкирующих сайт-мишень рДНК (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTaggggaatccgactgtttaatta-3' и 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcacctctcatgtctcttcaccg-3'). Секвенирование ампликона выполняли с применением Illumina MiSeq и результаты анализировали с применением конвейера CRISPResso2 (Clement et al Nat Biotechnol 37(3):224-226 (2019)) для определения мутационных сигнатур в сайте-мишени. Вставки и делеции обнаруживали в предполагаемом сайте расщепления GG и вокруг него, при этом пик сигнатуры вставки приходится непосредственно на предполагаемый сайт расщепления R2Tg (фигура 31B). Такие данные подтверждают применение анализа сигнатуры вставки/делеции для обнаружения и определения местоположения эндонуклеазной активности в системе Gene Writing.

Пример 32. Уточнение специфичности последовательности эндонуклеазного домена ретротранспозазы

В данном примере описан эксперимент по выяснению специфичности последовательности эндонуклеазной активности, ассоциированной с системой Gene Writer. В частности, информация о последовательности-мишени ДНК, важная для эндонуклеазной активности, выясняется посредством направленного мутагенеза областей различных размеров в сайте расщепления или вокруг него. В данном документе профилируют специфичность последовательности нативного эндонуклеазного домена ретротранспозазы R2Tg посредством применения анализа сигнатур вставок/делеций (см. пример 31) на

клетках, содержащих геномные "посадочные площадки" с нативной или измененной последовательностью-мишенью R2Tg, для определения активности расщепления этой библиотеки мишеней.

В данном примере создали клеточные линии со стабильной интеграцией "посадочных площадок", содержащих полученную из рДНК последовательность, соответствующую нативному сайту-мишени, где системы ретротранспозонов класса R2 облегчают ретротранспозицию. Данные "посадочные площадки" конструировали так, чтобы они имели либо (1) последовательность-мишень дикого типа, содержащую последовательность, полученную от рДНК, из области R2 рДНК; (2) "посадочную площадку" из (1) с мутацией последовательности длиной 12 п. о. в сайте расщепления R2 и вокруг него; или (3) серию мутаций в последовательности длиной 12 п. о. из (2) для дополнительного определения минимальных требований к последовательности в контексте 12 п. о., описанном в примере 2. Полный список "посадочных площадок" можно найти в таблице "Посадочные площадки" и он показан на фигуре 32 В. Чтобы создать такие клеточные линии, ДНК для различных "посадочных площадок" синтезировали и клонировали в лентивирусный вектор экспрессии гена ниже репортера GFP (фигура 32 А). Лентивирусные векторы "посадочной площадки" верифицировали секвенированием "посадочных площадок" по Сэнгеру. Чтобы получить лентивирусные частицы для трансдукции, 9 мкг плазмид с подтвержденной последовательностью и 9 мкг смеси для упаковки лентивирусов (Biosettia) трансфицировали в линию клеток для упаковки лентивирусов LentiX-293T (Takara Bio). Трансфицированные клетки инкубировали при 37°C, 5% CO2 в течение 48 часов (включая одну смену среды через 24 часа) и собирали среду, содержащую вирусные частицы. Собранную среду фильтровали через фильтр с размером пор 0,2 мкм для удаления клеточного дебриса и готовили для трансдукции клеток U2OS. Вируссодержащий фильтрат разводили в DMEM и смешивали с полибреном для приготовления серии разведений для трансдукции клеток с конечной концентрацией полибрена, составляющей 8 мкг/мл. Реципиентные клетки U2OS выращивали в среде, содержащей вирус, в течение 48 часов, и затем разделяли свежей средой и выращивали до слияния. Эффективность трансдукции различных разведений вируса измеряли по экспрессии GFP с помощью проточной цитометрии и ddPCR для определения среднего числа копий интегрированных лентивирусных "посадочных площадок".

С применением профилирования вставок/делеций анализировали способность ретротранспозазы R2Tg проявлять эндонуклеазную активность в отношении сайта-мишени с различными мутациями в положении расщепления R2 и вокруг него (пример 31). В частности, 73 нг вектора экспрессии полипептида Gene Writer, содержащего ретротранспозазу R2Tg, подвергали нуклеофекции в различных клеточных линиях "посадочной площадки" U2OS с применением системы Lonza Amaxa Nucleofector 96 Well Shuttle с программой нуклеофекции DN100 в соответствии с инструкциями производителя. После нуклеофекции клетки выращивали при 37°C, 5% CO2 в течение 3 дней перед лизисом клеток и выделением геномной ДНК. Праймеры, специфичные к последовательностям

"посадочной площадки" (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctcacacaggaaacagctatg-3' и 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTggatgtgctgcaaggcgatt-3') применяли для амплификации области, и затем очищенные ампликоны секвенировали на Illumina MiSeq. Сигнатуры эндонуклеазной активности в сайте-мишени анализировали посредством обнаружения вставок и делеций в "посадочной площадке" с применением конвейера CRISPResso2 (Clement et al Nat Biotechnol 37(3):224-226 (2019)). Как показано в примере 31, эндонуклеазная активность R2Tg проявляется в сайте расщепления GG (фигуры 31 A и B). Применяя данную серию "посадочных площадок", было обнаружено, что минимальная последовательность, важная для эндонуклеазной активности для тестируемого полипептида, включает данный динуклеотид GG и дополнительный динуклеотид AA, расположенные непосредственно выше, тем самым определяя мотив 5'-AAGG-3', важный для эндонуклеазной активности в пределах нативной последовательности-мишени (фигура 32 B). В некоторых вариантах осуществления может быть желательно найти нативный эндонуклеазо-специфический мотив в участке генома человека для перенацеливания системы Gene Writer на основе ретротранспозазы. В некоторых вариантах осуществления природная последовательность AAGG в геноме используется в качестве затравки для перенацеливания системы Gene Writing на основе ретротранспозазы R2, где ДНК-связывающий домен подвергнут мутации или заменен гетерологичным ДНК-связывающим доменом таким образом, что связывание полипептида Gene Writer с новым сайтом-мишенью приводит к правильному расположению эндонуклеазного домена в мотиве AAGG для обеспечения эндонуклеазной активности.

Таблица 41. "Посадочные площадки", содержащие мутантные последовательности-мишени или последовательности-мишени дикого типа, которые описаны в примере 32. Последовательности включали в геном с применением лентивирусной векторной системы для стабильной интеграции. Последовательности в таблице представлены в ориентации от 5' до 3', где подчеркнутый текст указывает на нативную последовательность рДНК, и полужирный текст указывает на мутации из нативной последовательности рДНК.

Таблица 41. "Посадочные площадки", предусматривающие мутантные последовательности-мишени или последовательности-мишени дикого типа

"Посадочная площадка"	**Соответствующая последовательность-мишень**	**SEQ ID NO.**
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**AAGGTAGCCAAA**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC	1986

	TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	
Мутантная форма 12 п. о.	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TCCAATATGATT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1987
10 п. о. (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TCCAATATGA**AATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1988
6 п. о. (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TCCAAT**GCCAAATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1989
4 п. о. (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TCCA**TAGCCAAATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA	1990

	ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	
2 п. о. (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TC**GGTAGCCAAATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1991
12 п. о. (GG)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TC**GG**ATATGATT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1992
10 п. о. (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TC**GG**ATATGA**AATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1993
8 п. о. (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTT**TC**GG**ATAT**CAAATGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1994

6 п. о. (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC T<u>GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA</u> <u>AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA</u> <u>GTAACTATGACTCTCTT**TC**GG**AT**GCCAAATGCCTCGTCATC</u> <u>TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC</u> <u>TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA</u> ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1995
12 п. о. (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC T<u>GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA</u> <u>AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA</u> <u>GTAACTATGACTCTCTT**TAGGTTATGATT**TGCCTCGTCATC</u> <u>TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC</u> <u>TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA</u> ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1996
10 п. о. (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC T<u>GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA</u> <u>AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA</u> <u>GTAACTATGACTCTCTT**TAGGTTATGA**AATGCCTCGTCATC</u> <u>TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC</u> <u>TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA</u> ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1997
8 п. о. (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC T<u>GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA</u> <u>AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA</u> <u>GTAACTATGACTCTCTT**TAGGTTAT**CAAATGCCTCGTCATC</u> <u>TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC</u> <u>TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA</u> ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1998

10 п. о. (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAA**CAATATGATT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	1999
8 п. о. (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAAGG**ATATGATT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	2000
6 п. о. (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAAGGTA**ATGATT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	2001
2 п. о. (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC TGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAAGGTAGCCA**TT**TGCCTCGTCATC TAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC TGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC	2002

Пример 33. Определение специфичности последовательности для перенаправления полипептидов Gene Writer

В данном примере описано перенаправление полипептидного компонента системы Gene Writer из его нативной последовательности распознавания в новое местоположение в геноме человека. Как описано в настоящем изобретении, система на основе ретротранспозазы может быть направлена на распознавание новой последовательности ДНК посредством добавления гетерологичного ДНК-связывающего домена отдельно или в сочетании с мутагенезом эндогенного ДНК-связывающего домена. В данном документе цинковый палец, способный нацеливаться на нативный сайт AAVS1 человека, сливают с полипептидом Gene Writer, чтобы обеспечить распознавание последовательности AAVS1.

Чтобы направить полипептид Gene Writer, содержащий ретротранспозазу R2Tg, на последовательность ДНК, отличную от его последовательности-мишени нативной рДНК, домен цинковых пальцев (ZF), нацеленный на сайт AAVS1 человека, сливали с N-концом полипептида. Применяя подход, аналогичный примеру 32, создавали клетки, содержащие различные композиции геномных "посадочных площадок", представляющие различные комбинации последовательности AAVS1, распознаваемой ZF, и последовательности-мишени нативной рДНК, распознаваемой ретротранспозазами R2. В целом библиотека состояла из 460 различных "посадочных площадок", содержащих последовательность AAVS1 вместе с последовательностями рДНК различной длины, которые дополнительно диверсифицировались за счет изменения расстояния и ориентации между двумя последовательностями (фигура 33). В данном примере все "посадочные площадки" конструировались для включения геномной последовательности AAVS1 человека, содержащей сайт связывания ZF AAVS1. Кроме того, к последовательности AAVS1 добавляли последовательность рДНК, содержащую минимальную последовательность AAGG, важную для эндонуклеазной активности R2Tg (см. пример 32), в соответствии со следующими параметрами: (1) последовательность рДНК включала последовательность рДНК длиной 12, 22, 32, 42, 52, 62 или 72 нуклеотидов, расположенную непосредственно на 3'-конце тетрануклеотида AAGG; или (2) последовательность рДНК включала 12, 22, 32 или 42 нуклеотида непосредственно на 3'- и 5'-концах сайта расщепления AAGG, в результате чего общая длина рДНК составляла 24, 44, 64 или 84 нуклеотидов. Различные композиции последовательностей рДНК дополнительно размещали на различных расстояниях от сайта связывания ZF AAVS1, в том числе на расстояние 5, 10, 15, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80 или 85 нуклеотидов. Были учтены конструктивные соображения, касающиеся сохранения общей длины "посадочной площадки" между постоянными праймерами для анализа, с целью предотвращения смещения при ПЦР-амплификации, например, более длинные последовательности рДНК нельзя было размещать на том же расстоянии от сайта AAVS1, что и более короткие последовательности рДНК. В качестве окончательного варианта, ориентация между двумя сайтами была изменена таким образом, что композиции последовательности рДНК, описанные выше, располагались либо выше, либо ниже сайта AAVS1, либо в прямой, либо в обратной ориентации. Применяя такие параметры, была разработана библиотека из 454 гибридных "посадочных площадок" AAVS1-рДНК, с контролями, включающими различные

комбинации полноразмерной последовательности рДНК и последовательности AAVS1 (положительные), или без последовательности рДНК (отрицательные). Иллюстративное представление стратегии проектирования "посадочных площадок" показано на фигуре 33, и краткий список примерных последовательностей "посадочных площадок" для демонстрации конкретных композиций приведен в таблице 42.

Библиотека лентивирусных конструкций, описанная выше, была синтезирована с 3'-штрих-кодами для анализа последовательности и клонирована в лентивирусный вектор экспрессии гена. Затем лентивирусную систему применяли для создания клеточных линий U2OS со встроенными "посадочными площадками", описанными в примере 32. Чтобы подтвердить успешное создание библиотеки, пул клеточных линий U2OS с "посадочной площадкой" анализировали на предмет представления "посадочной площадки". Праймеры, специфичные для консервативных последовательностей "посадочных площадок" (см. пример 32), применяли для амплификации в области-мишени, включая специфический для конструкции штрих-код. Штрих-коды, присутствующие в каждой "посадочной площадке", демультиплексировали и сортировали с помощью вычислений, при этом примерно 94% "посадочных площадок" были представлены по меньшей мере 10000 ридов (фигура 34).

После одобрения библиотеки клетки U2OS с "посадочной площадкой" применяли для определения минимальных детерминант последовательности для перенацеливания полипептида Gene Writer на основе ретротранспозазы R2Tg. Две конструкции перенацеливания создавали посредством слияния кодирующей последовательности ZF-AAVS1 либо с полноразмерным R2Tg (SEQ ID NO: 1672) (ZF-R2Tg, фигура 35 A) или с укороченным ДНК-связывающим доменом R2Tg под SEQ ID NO: 1663) (ZF-R2Tg(noDBD), фигура 36A), и осуществляли электропорацию объединенных клеток U2OS с "посадочной площадкой" с использованием соответствующих плазмид экспрессии. В частности, 400 нг либо ZF-R2Tg, либо ZF-R2Tg (без DBD) доставляли в клетки с применением 96-луночной челночной системы Lonza Amaxa Nucleofector с программой нуклеофекции DN100 в соответствии с инструкциями производителя. После нуклеофекции клетки выращивали при 37°C, 5% CO_2 в течение 3 дней перед лизисом клеток и выделением геномной ДНК. Как и при одобрении библиотеки выше, праймеры, специфичные для последовательностей "посадочных площадок", применяли для амплификации области-мишени для секвенирования ампликона с помощью Illumina MiSeq. Прочтения, получаемые в результате секвенирования вариантов "посадочных площадок", сначала демультиплексировали с применением соответствующих штрих-кодов, затем сигнатуры эндонуклеазной активности в сайте-мишени анализировали посредством обнаружения вставок и делеций с применением конвейера CRISPResso2 (Clement et al Nat Biotechnol 37(3):224-226 (2019)).

Учитывая, что конструкция ZF-R2Tg содержит полноразмерный белок R2Tg, было предсказано, что данный полипептид по-прежнему будет сохранять способность распознавать свою нативную последовательность-мишень. Частоту вставок в сайте-мишени GG рассчитывали и наносили на график для каждой "посадочной площадки"

(фигура 35 A). Обнаружили, что "посадочные площадки" положительного контроля, содержащие 200 нуклеотидов последовательности рДНК, содержат сигнатуры вставки в сайте расщепления GG, как в примере 31. Обнаружили, что "посадочная площадка" отрицательного контроля, лишенная последовательности рДНК, не содержит каких-либо вставок. В данном документе сигнатуры вставки, возникающие в результате активности эндонуклеазы ZF-R2Tg в сайте-мишени расщепления, обнаружили в "посадочных площадках", содержащих 44, 64 и 84 нуклеотидов последовательности-мишени нативной рДНК, но не в тех, которые содержат только 24 нуклеотидов последовательности-мишени нативной рДНК (фигура 35 A и B). Было показано, что такие длины последовательности рДНК, которые дали положительный результат, оказались положительными на "посадочных площадках" с разным расстоянием от последовательности AAVS1.

Затем аналогичным образом оценивали конструкцию ZF-R2Tg (без DBD), не имеющую эндогенного ДНК-связывающего домена и предположительно зависящую от ZF в отношении связывания с мишенью и активности. Частоту вставок в сайте-мишени GG рассчитывали и наносили на график для каждой "посадочной площадки" (фигура 36 A). Обнаружили, что "посадочная площадка" отрицательного контроля, лишенная последовательности рДНК, не содержит каких-либо вставок. Учитывая, что белок R2Tg в такой конструкции содержал значительную делецию, не было конфигурации положительного контроля "посадочной площадки" для проверки активности. В данном эксперименте две разные конфигурации "посадочных площадок" показали сигнатуры вставок/делеций на сайте-мишени GG. Оба попадания содержали одну и ту же последовательность рДНК длиной 44 нуклеотида, но были по-разному расположены относительно сайта AAVS1, где одна рДНК-мишень находилась на расстоянии 55 нуклеотидов выше сайта AAVS1, и другая находилась на расстоянии 20 нуклеотидов ниже сайта AAVS1 и в обратной комплементарной ориентации (фигуры 36 A и B). Несмотря на различный состав, данные два попадания показали восстановление активности мутантного ДНК-связывающего домена ретротранспозазы R2 путем компенсации удаленного эндогенного домена с помощью слитого гетерологичного ДНК-связывающего домена, который, как известно, нацелен на нативный локус в геноме человека. Кроме того, в данном примере предусматривается способ дополнительного уточнения требований для перенацеливания полипептида Gene Writer на измененную последовательность в геноме человека.

В таблице 42 представлен выбор примерных последовательностей-мишеней "посадочных площадок", предназначенных для тестирования полипептидов Gene Writer, содержащих продукты слияния цинковых пальцев AAVS1 с ретротранспозазой R2Tg, как показано в данном документе, например, в данном примере. Включены: 1) положительный контроль, содержащий последовательность распознавания цинковых пальцев AAVS1 и полноразмерную область из 200 нуклеотидов из рДНК, сосредоточенную вокруг сайта расщепления R2; 2) отрицательный контроль без последовательности рДНК; 3) последовательность рДНК тз 44 нуклеотидов, расположенная выше последовательности

AAVS1; и 4) последовательность рДНК из 44 нуклеотида, расположенная ниже последовательности AAVS1 и в обратной ориентации. Было показано, что две экспериментальные "посадочные площадки", включенные в данный документ, обеспечивают расщепление основной последовательности AAGG с применением цинкового пальца, слитого с мутантным R2Tg, лишенным его эндогенного N-концевого ДНК-связывающего домена, например, см. данный пример. Последовательность рибосомной ДНК обозначена подчеркнутым текстом, сердцевина AAGG обозначена скобками, и последовательность связывания цинкового пальца AAVS1 обозначена строчными буквами.

Таблица 42. Выбор примерных последовательностей-мишеней "посадочных площадок", предназначенных для тестирования полипептидов Gene Writer, содержащих продукты слияния цинковых пальцев AAVS1 с ретротранспозазой R2Tg

Описание	**Последовательность**	**SEQ ID NO**
Положительный контроль R2(200)_левый(прямой)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATG TCAAAGTGAAGAAATTCAATGAAGCGCGGGTAAAC GGCGGGAGTAACTATGACTCTCTT[AAGG]TAGCCAA ATGCCTCGTCATCTAATTAGTGACGCGCATGAATGG ATGAACGAGATTCCCACTGTCCCTACCTACTATCCA GCGAAACCACAGCCAAGGGAAGAGTGGCCccactgtggg gtGGAGGGGACGACTGGGAAAAGTTAGGATCCCCTG GCGTTACCCAACTTAATCGCCTTGCAGCACATCC	2003
Отрицательный контроль AAVS1_без R2	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCTGCCTAACAGGAGGTGGGGGTTAGACCCA ATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGAT GGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCCc cactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CAGAGCCACATTAACCGGCCCTGGGAATATAAGGTG GTCCCAGCTCGGGGACACAGGATCCCTGGAGGCAGC AAACATGCTGACTGGGAAAATCTGGCTTCCCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC	2004

AAVS1_R2(44 нуклеотида) при -55(прямой)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCGGGAGTAACTATGACTCTCTT[AAGG]TAG CCAAATGCCTCGTCAGGAAGGAGGAGGCCTAAGGA TGGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCC ccactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CAGAGCCACATTAACCGGCCCTGGGAATATAAGGTG GTCCCAGCTCGGGGACACAGGATCCCTGGAGGCAGC AAACATGCTGACTGGGAAAACTTAGAGAACCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC	2005
AAVS1_R2(44 нуклеотида) при +20(rc)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCTGCCTAACAGGAGGTGGGGGTTAGACCCA ATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGAT GGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCCc cactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CTGACGAGGCATTTGGCTA[CCTT]AAGAGAGTCATA GTTACTCCCGCGGGGACACAGGATCCCTGGAGGCAG CAAACATGCTGACTGGGAAAACCTTCTCGGCCCTGG CGTTACCCAACTTAATCGCCTTGCAGCACATCC	2006

Пример 34. Выбор липидных реагентов со сниженным содержанием альдегидов

В этом примере липиды выбирают для дальнейшего использования в составах на основе липидных наночастиц, содержащих компонент Gene Writing, представляющий собой нуклеиновую(нуклеиновые) кислоту(кислоты), и липиды выбирают, по меньшей мере частично, на основании отсутствия или низкого уровня контаминирующих альдегидов. Реакционноспособные альдегидные группы в липидных реагентах могут обуславливать химические модификации компонента, представляющего собой нуклеиновую(нуклеиновые) кислоту(кислоты), например РНК, например РНК-матрицу, во время составления LNP. Следовательно, в некоторых вариантах осуществления содержание альдегидов в липидных реагентах сводят к минимуму.

Жидкостную хроматографию (LC) в сочетании с тандемной масс-спектрометрией (MC/MC) можно использовать для разделения, определения характеристик и количественной оценки содержания альдегидов в реагентах, например, как описано в работе Zurek et al. The Analyst 124(9):1291-1295 (1999), включенной в данный документ посредством ссылки. В данном документе каждый липидный реагент подвергают анализу методом LC-MS/MS. Посредством метода LC/MS-MS сначала отделяют липид и одну или несколько примесей на колонке HPLC C8, а затем следует 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.

Пример 35. 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 34).

Модификации РНК можно обнаружить 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 36. Gene Writer™ позволяет осуществлять вставку крупного участка в геномную ДНК

В этом примере описано использование системы редактирования генов Gene Writer™ для изменения геномной последовательности путем вставки крупной последовательности нуклеотидов.

В этом примере полипептид Gene Writer™, gRNA и матрица для записи представлены в виде ДНК, трансфицированной в клетки HEK293T. Полипептид Gene Writer™ использует никазу Cas9 как 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 ник в сайте-мишени, а затем выполняет TPRT на матрице, инициируя реакцию с использованием праймирующих областей на матрице, которые комплементарны последовательности, непосредственно примыкающей к сайту ника, и копируя полезную нагрузку GFP в геномную ДНК.

После трансфекции клетки инкубируют в течение трех дней для обеспечения экспрессии системы Gene Writing™ и конверсии 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.

Пример 37. Gene Writer™ позволяет осуществлять вставку крупного участка в геномную ДНК

В данном примере описано применение системы Gene Writer в клетке человека, где путь репарации посредством однонитевой матрицы (SSTR) является ингибированным.

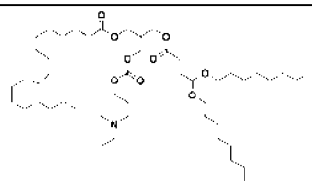
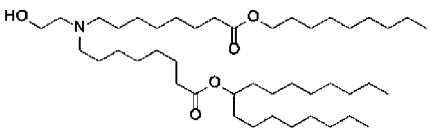
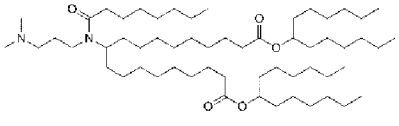
В таком примере путь SSTR будут ингибировать с помощью siRNA к основным компонентам пути: FANCA, FANCD2, FANCE, USP1. Также будут включены контрольные siРНК ненацеленного контроля. Нуклеофекцию 200 тысяч клеток U2OS будут осуществлять с использованием 30 пмоль (1,5 мкМ) siRNA, а также с использованием плазмиды с управляющим элементом R2Tg и плазмиды с трансгеном (транс-конфигурация). В частности, 250 нг плазмид, экспрессирующих R2Tg, контрольный R2Tg с мутацией в домене RT или контрольный R2Tg с мутацией, обеспечивающей инактивацию эндонуклеазы, применяют в сочетании с трансгеном в молярном соотношении 1:4 (управляющий элемент к трансгену). Трансфекции клеток U2OS проводят в буфере SE с применением программы DN100. После нуклеофекции клетки выращивают в полной среде в течение 3 дней. gDNA собирают в день 3 и проводят ddPCR для оценки интеграции в сайте рДНК. Интеграцию трансгена в рДНК обнаруживают по отсутствию основных компонентов пути SSTR.

Пример 38. Состав на основе липидных наночастиц, инкапсулирующих мРНК люциферазы светлячка

В этом примере мРНК репортерного гена, кодирующую люциферазу светлячка, составляли с включением в липидные наночастицы, содержащие различные ионизируемые липиды. Компоненты липидных наночастиц (LNP) (ионизируемый липид, вспомогательный липид, стерин, PEG) растворяли в 100% растворе этанола с липидным компонентом. Затем их получали при молярных соотношениях 50:10:38,5:1,5 с использованием ионизируемого липида LIPIDV004 или LIPIDV005 (таблица A1), DSPC, холестерина и DMG-PEG 2000 соответственно. мРНК люциферазы светлячка-LNP, содержащие ионизируемый липид LIPIDV003 (таблица A1), получали при молярном соотношении 45:9:44:2 с использованием LIPIDV003, DSPC, холестерина и DMG-PEG 2000 соответственно. мРНК люциферазы светлячка, используемую в этих составах, получали путем транскрипции in vitro, и она кодировала белок люциферазы светлячка, дополнительно содержащий 5′-кэп, 5′- и 3′-UTR и поли(A)-хвост. мРНК синтезировали в стандартных условиях для транскрипции РНК-полимеразой T7 in vitro с кэпированием во время транскрипции, однако с 100% заменой нуклеотидтрифосфата UTP на N1-метил-псевдоуридинтрифосфат в реакции. Очищенную мРНК растворяли в 25 мМ цитрате натрия, pH 4, до концентрации 0,1 мг/мл.

мРНК люциферазы светлячка составляли с включением в LNP с молярным соотношением амина липида и фосфата РНК (N:P), составляющим 6. LNP составляли посредством микрофлюидного смешивания растворов липидов и РНК с использованием настольного прибора Precision Nanosystems NanoAssemblr™ с использованием параметров, рекомендованных производителем. Соотношение водного и органического растворителей 3:1 поддерживали во время смешивания с использованием различных скоростей потока. После смешивания LNP собирали и подвергали диализу в 15 мМ Tris, 5% буфере сахарозы при 4°C в течение ночи. Сосав 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 А1. Ионизируемые липиды, используемые в примере 37

ID ЛИПИДА	**Химическое название**	**Молекулярная масса**	**Структура**
LIPIDV003	(9Z,12Z)-3-((4,4-бис(октилокси)бутаноил)окси)-2-((((3-(диэтиламино)пропокси)карбонил)окси)метил)пропилоктадека-9,12-диеноат	852,29	
LIPIDV004	Гептадекан-9-ил 8-((2-гидроксиэтил)(8-(нонилокси)-8-оксооктил)амино)октаноат	710,18	
LIPIDV005		919,56	

Полученные LNP анализировали в отношении размера, однородности и % инкапсуляции РНК. Измерения размера и однородности проводили с помощью динамического светорассеяния с использованием прибора для DLS Malvern Zetasizer (Malvern Panalytical). LNP разводили в PBS перед измерением с помощью DLS с определением среднего размера частиц (нанометры, нм) и индекса полидисперсности (pdi). Размеры частиц мРНК люциферазы светлячка-LNP показаны в **таблице 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 A3.

Таблица A3. Инкапсуляция РНК после составления LNP

ID LNP	Ионизируемый липид	% Инкапсуляции мРНК
LNPV019-002	LIPIDV005	98
LNPV006-006	LIPIDV004	92
LNPV011-003	LIPIDV003	97

Пример 39. Тестирование активности мРНК-LNP in vitro в первичных гепатоцитах

В этом примере LNP, содержащие мРНК репортерного гена люциферазы, использовали для доставки полезной нагрузки РНК в клетки в культуре. Первичные гепатоциты мыши или первичные гепатоциты человека размораживали и высевали в покрытые коллагеном 96-луночные планшеты для тканевых культур с плотностью 30000 или 50000 клеток на лунку соответственно. Клетки высевали в 1x среду E Уильяма без фенолового красного и инкубировали при 37°C с 5% CO_2. Через 4 часа среду заменяли поддерживающей средой (1x среда E Уильяма без фенола, содержащая Hepatocyte Maintenance Supplement Pack (ThermoFisher Scientific)) и клетки выращивали в течение ночи при 37°C с 5% CO_2. мРНК люциферазы светлячка-LNP размораживали при 4°C и осторожно перемешивали. LNP разводили до соответствующей концентрации в поддерживающей среде, содержащей 7,5% эмбриональной бычьей сыворотки. LNP инкубировали при 37°C в течение 5 минут перед добавлением к высеянным первичным гепатоцитам. Для оценки

доставки полезной нагрузки РНК 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 15000*g* при 4°C в течение 3 минут с удалением клеточного дебриса. Концентрацию белка измеряли для каждого образца с использованием набора для анализа белка Pierce™ 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. 37, и указывают на успешную LNP-опосредованную доставку РНК в первичные клетки в культуре. Как показано на фиг. 37A, LNP, составленные согласно примеру 38, анализировали в отношении доставки полезной нагрузки к первичным гепатоцитам человека (A) и мыши (B) в соответствии с примером 39. В результате анализа с использованием люциферазы было выявлено дозозависимую активность люциферазы в лизатах клеток, что указывает на успешную доставку РНК в клетки и экспрессию люциферазы светлячка из полезной нагрузки мРНК.

Пример 40. 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 60, и тестировали in vitro до (пример 39) 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 показана успешная доставка мРНК при всех липидных составах, при этом активность репортерного гена соответствовала ранжированию LIPIDV005>LIPIDV004>LIPIDV003 (ФИГ. 38). Как показано на фигуре 38, 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 временной, и уровни ферментов возвращались к уровню, близкому к фоновому уровню в среде-носителе, 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, предусматривающая второй однонитевой разрыв, описаны для редактирования генома in vitro путем нуклеофекции, с использованием модифицированных нуклеотидов, путем липофекции и для редактирования первичных Т-клеток. Как описано в настоящей заявке, эти системы, состоящие полностью из РНК, характеризуются многими уникальными преимуществами в отношении клеточной иммуногенности и токсичности, что важно при работе с более чувствительными первичными клетками, особенно иммунными клетками, например Т-клетками, в отличие от линий клеток на основе иммортализованной культуры клеток. Кроме того, предполагается, что все эти системы на основе РНК могут быть нацелены на альтернативные ткани и типы клеток с использованием новых систем доставки липидов, как указано в данном документе, например, для доставки в печень, легкие, мышцы, иммунные клетки и другие, учитывая, что функцию систем Gene Writing проверяли в данном документе in vitro на нескольких

типах клеток, и функция других систем на основе РНК, доставляемых с помощью нацеленных LNP, известна в данной области техники. Доставка in vivo систем Gene Writing может оказать большое влияние на многие области медицины, например, исправление патогенных мутаций, внедрение защитных вариантов и улучшение клеток, эндогенных для организма, например Т-клеток. Предполагается, что при надлежащем составлении Gene Writing, состоящая полностью из РНК, сделает возможным изготовление средств клеточной терапии in situ в организме пациента.

Пример 41. 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. 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 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 слияния Cas-RT варьируется с применением библиотеки последовательностей линкера. Более конкретно, линкеры из таблицы 38B применяли для создания новых вариантов слитой конструкции Cas9-RT, ранее демонстрировавшей низкую экспрессию белка, и доставляли в клетки человека для скрининга на улучшенную экспрессию белка Cas-RT.

Набор из 22 пептидных линкеров (таблица 38B) с различной степенью длины, гибкости, гидрофобности и вторичной структуры впервые применяли для создания вариантов слитого белка Cas-RT посредством замены исходного линкера (SQ ID NO: 480). 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 на 2-й день после трансфекции и анализировали с помощью вестерн-блоттинга с применением первичных антител к Cas9. Линкер 10 (SEQ ID NO: 468), перечисленный в таблице 38 B, значительно улучшил экспрессию слияния Cas-RT (Фигура 39), демонстрируя потенциально сильное влияние последовательности пептидного линкера на экспрессию Cas-RT.

Таблица 38B. Пептидные последовательности, применяемые в качестве линкеров между доменами Cas и RT в полипептидах Gene Writer, содержащих продукты слияния Cas-RT

№	**Последовательность линкера**	**Примечания**	**Seq Id. No.**

1	GGS	Короткий	**459**
2	GGGGS	Гибкий, короткий	**460**
3	$(GGGGS)_2$	Гибкий	**461**
4	$(GGGGS)_3$	Гибкий, длинный	**462**
5	$(GGGGS)_4$	Гибкий, очень длинный	**463**
6	$(G)_6$	Гибкий	**464**
7	$(G)_8$	Гибкий	**465**
8	GSAGSAAGSGEF	Гибкий	**466**
9	(GSSGSS)	Средний	**467**
10	$(GSSGSS)_2$	Средний, гибкий	**468**
11	$(GSSGSS)_3$	Средний	**469**
12	SGSETPGTSESATPES	XTEN	**470**
13	(EAAAK)	Жесткая спираль, короткий	**471**
14	$(EAAAK)_2$	Жесткая спираль, средний	**472**
15	$(EAAAK)_3$	Жесткая спираль, длинный	**473**
16	PAP	Жесткий, короткий	**474**
17	PAPAP	Жесткий, короткий	**475**
18	PAPAPAPAP	Жесткий, средний	**476**
19	$A(EAAAK)_4ALEA(EAAAK)_4A$	Жесткий, очень длинный со спиралями	**477**

20	GGGGS(EAAAK)GGGGS	Гибкий - спираль - изгиб	**478**
21	(EAAAK)GGGGS(EAAAK)	Спираль - изгиб - спираль	**479**
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	Гибкий - XTEN - гибкий	**480**

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, предусматривающая второй однонитевой разрыв, описаны для редактирования генома in vitro путем нуклеофекции, с использованием модифицированных нуклеотидов, путем липофекции и для редактирования первичных Т-клеток. Как описано в настоящей заявке, эти системы, состоящие полностью из РНК, характеризуются многими уникальными преимуществами в отношении клеточной иммуногенности и токсичности, что важно при работе с более чувствительными первичными клетками, особенно иммунными клетками, например Т-клетками, в отличие от линий клеток на основе иммортализованной культуры клеток. Кроме того, предполагается, что все эти системы на основе РНК могут быть нацелены на альтернативные ткани и типы клеток с использованием новых систем доставки липидов, как указано в данном документе, например, для доставки в печень, легкие, мышцы, иммунные клетки и другие, учитывая, что функцию систем Gene Writing проверяли в данном документе in vitro на нескольких типах клеток, и функция других систем на основе РНК, доставляемых с помощью нацеленных LNP, известна в данной области техники. Доставка in vivo систем Gene Writing может оказать большое влияние на многие области медицины, например, исправление патогенных мутаций, внедрение защитных вариантов и улучшение клеток, эндогенных для организма, например Т-клеток. Предполагается, что при надлежащем составлении Gene Writing, состоящая полностью из РНК, сделает возможным изготовление средств клеточной терапии in situ в организме пациента.

Следует понимать, что для всех числовых границ, описывающих какой-либо параметр в настоящей заявке, таких как "приблизительно", "по меньшей мере", "меньше чем" и "больше чем", описание также обязательно охватывает любой диапазон, ограниченный указанными значениями. Соответственно, например, описание "по меньшей мере 1, 2, 3, 4 или 5" также описывает в числе прочего, диапазоны 1-2, 1-3, 1-4, 1-5, 2-3, 2-4, 2-5, 3-4, 3-5 и 4-5 и так далее..

Следует понимать, что все патенты, заявки или другие ссылки, цитируемые в данном документе, такие как непатентная литература и информация об эталонных последовательностях, включены посредством ссылки во всей своей полноте для всех целей,

а также для цитируемого предложения. Если существует какое-либо противоречие между документом, включенным посредством ссылки, и настоящей заявкой, настоящая заявка будет иметь преимущественную силу. Вся информация, связанная с эталонными последовательностями генов, раскрытыми в настоящей заявке, такая как GeneID или номера доступа (обычно ссылки на номера доступа в NCBI), в том числе, например, геномные локусы, геномные последовательности, функциональные аннотации, аллельные варианты и эталонные мРНК (в том числе, например, границы экзонов или элементы отклика) и последовательности белков (такие как структуры консервативных доменов), а также ссылки, касающиеся химических соединений (например, соединения в PubChem, вещества в PubChem или записи в PubChem Bioassay, в том числе содержащиеся в них аннотации, такие как структуры и анализы, и т.д.), настоящим включены посредством ссылки во всей своей полноте.

Заголовки, используемые в настоящей заявке, предназначены только для удобства и не влияют на интерпретацию настоящей заявки.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где полипептид содержит мутацию, приводящую к инактивации и/или делеции сигнала ядрышковой локализации.

2. Система для модификации ДНК, содержащая:

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

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

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

3. Система для модификации ДНК, содержащая:

(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

где система способна разрезать первую нить ДНК-мишени по меньшей мере дважды (например, дважды), и

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

4. Способ модификации нити ДНК-мишени в клетке, ткани или у субъекта, включающий введение системы в клетку, где система содержит:

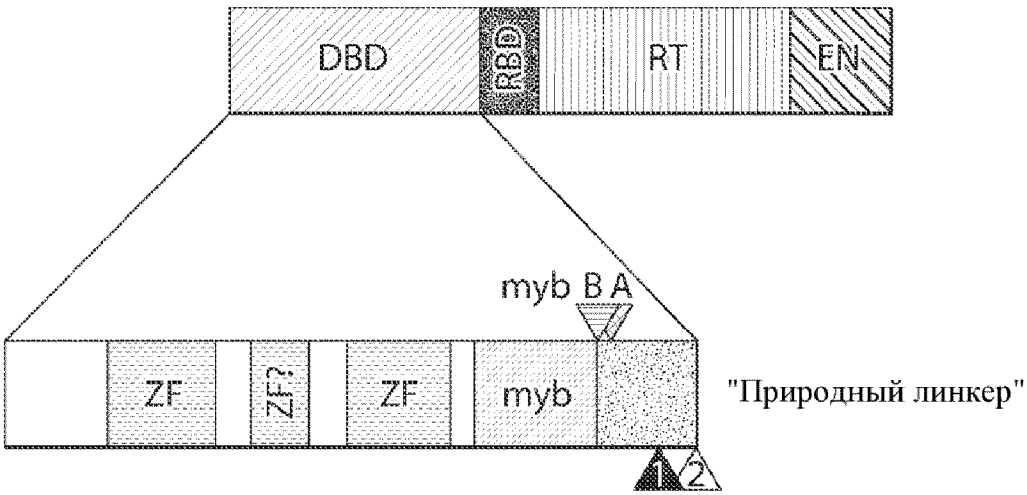
(a) полипептид или нуклеиновую кислоту, кодирующую полипептид, где полипептид содержит (i) домен обратной транскриптазы и (ii) эндонуклеазный домен; и

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

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

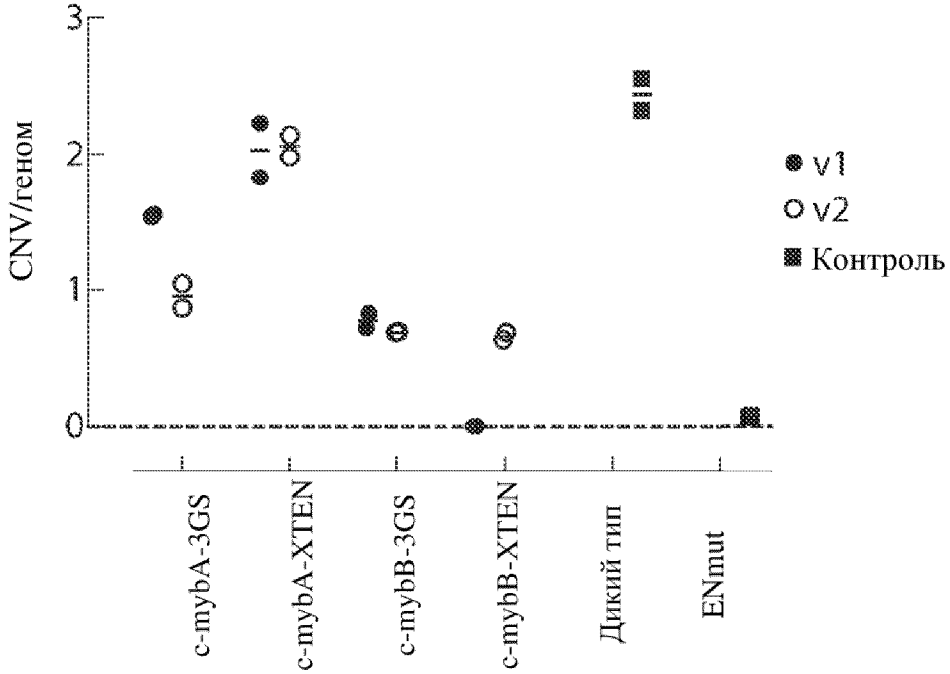
где клетка характеризуется уменьшенной активностью Rad51-опосредованного пути репарации, уменьшенной экспрессией Rad51 или компонента Rad51-опосредованного пути репарации или не содержит функциональный Rad51-опосредованный путь репарации, например, не содержит функциональный ген Rad51, например, содержит мутацию (например, делецию), приводящую к инактивации одной или обеих копий гена Rad51 или другого гена в Rad51-опосредованном пути репарации По доверенности

Делеция природного линкера



ФИГ. 1A

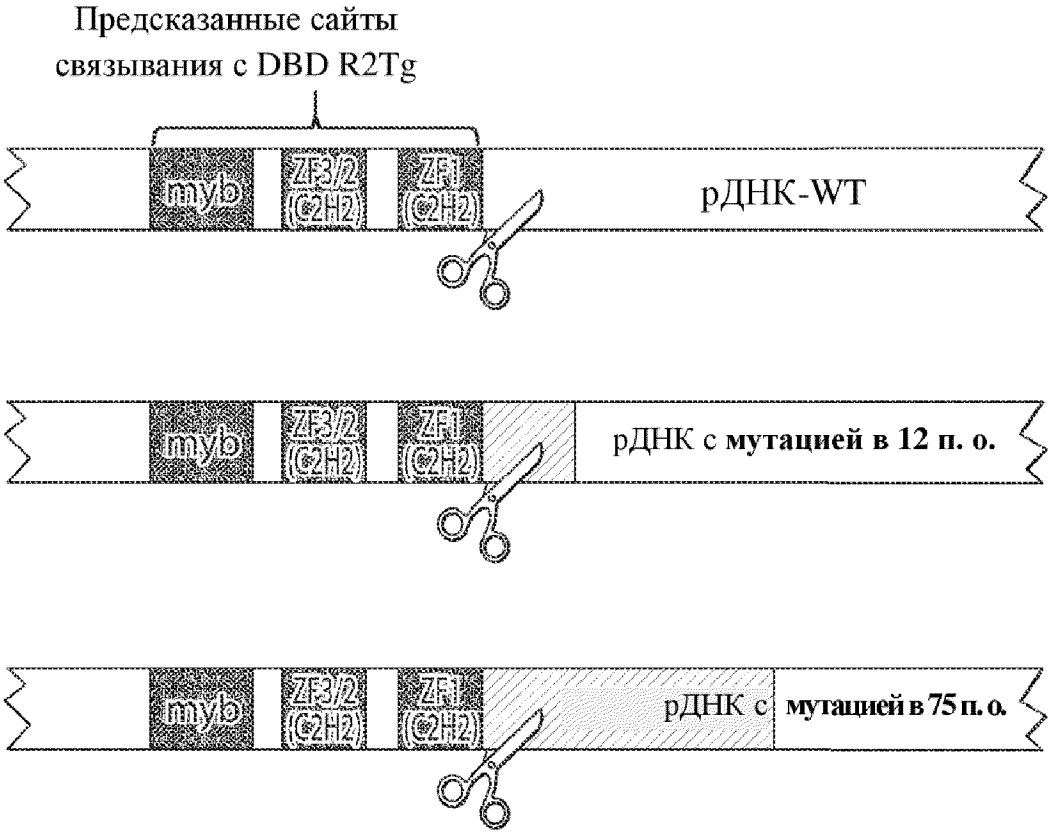
Делеция природного линкера – анализ на 3’ методом ddPCR



Делеция природного линкера

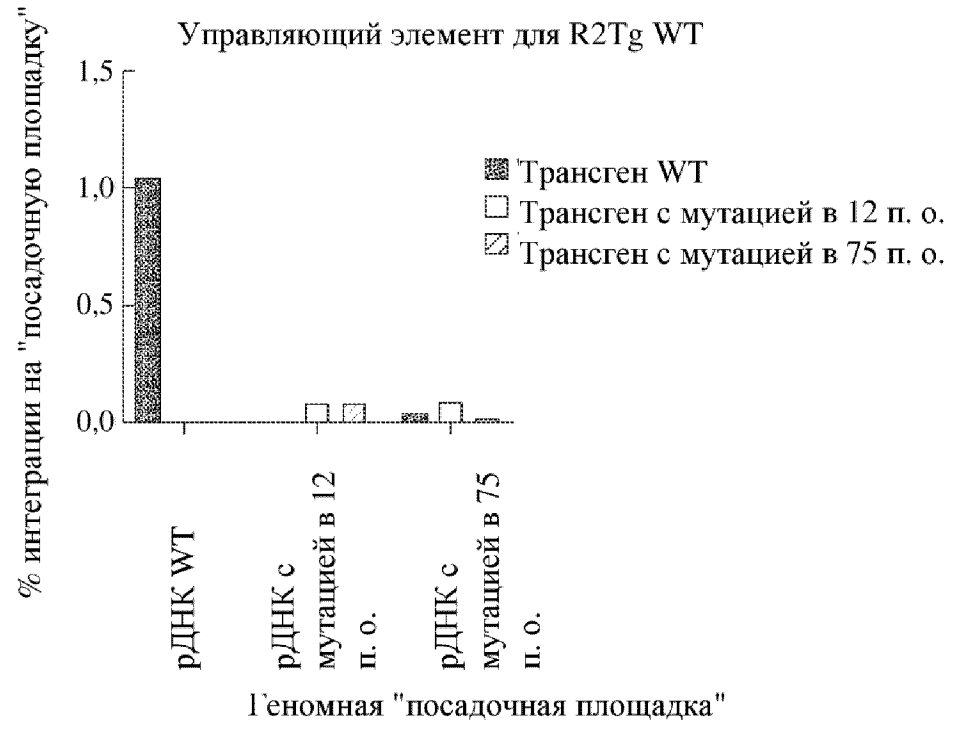
ФИГ. 1B

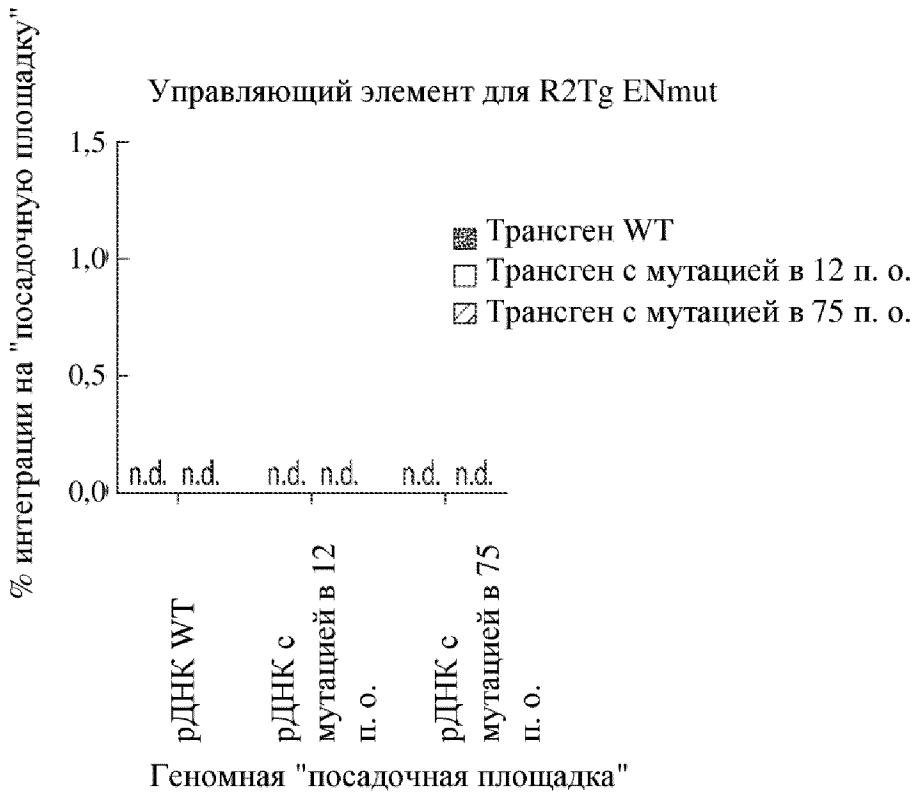
Фигуры, изображающие "посадочные площадки"



ФИГ. 2

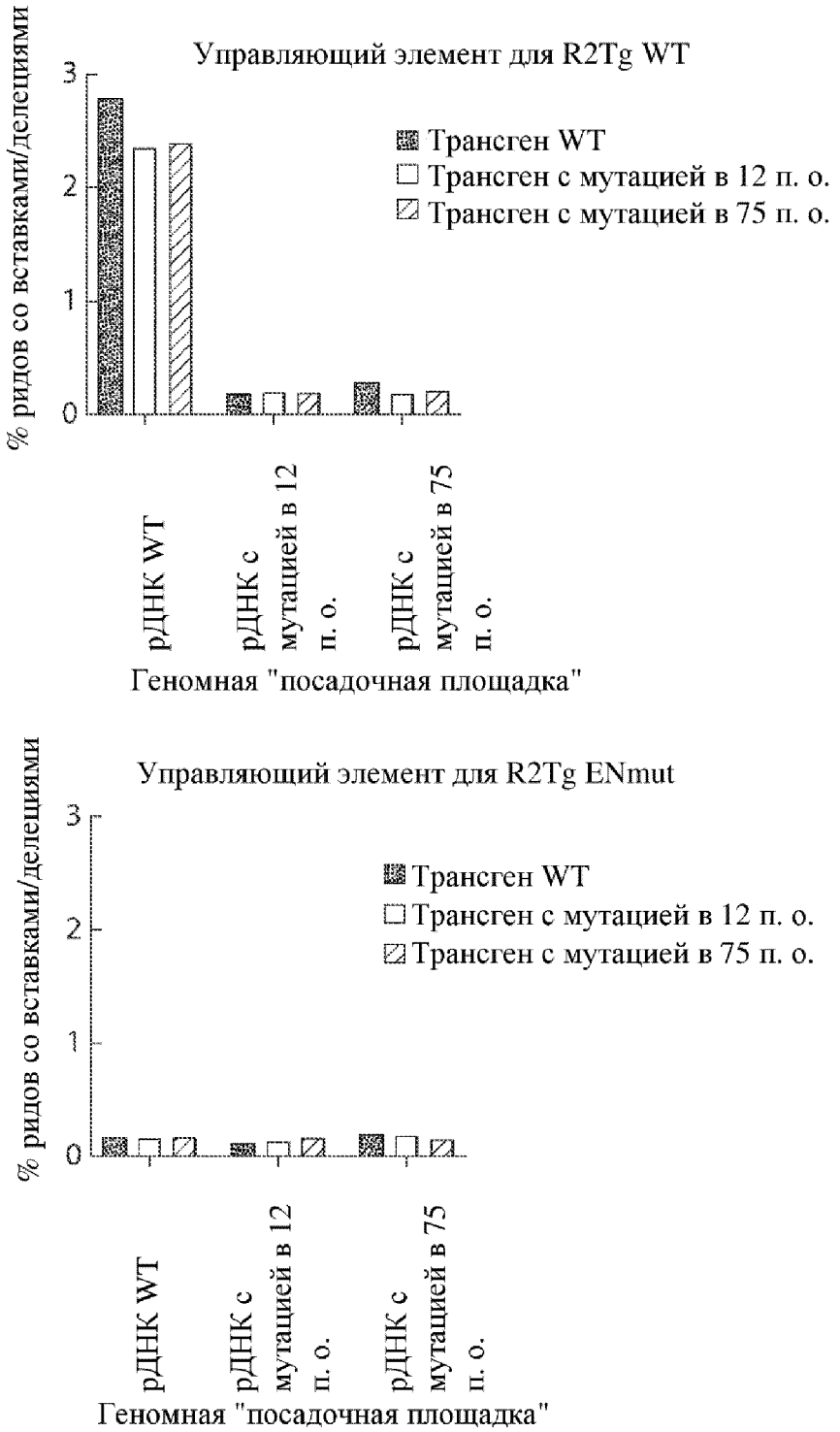
Интеграция в "посадочные площадки" с мутантным сайтом R2





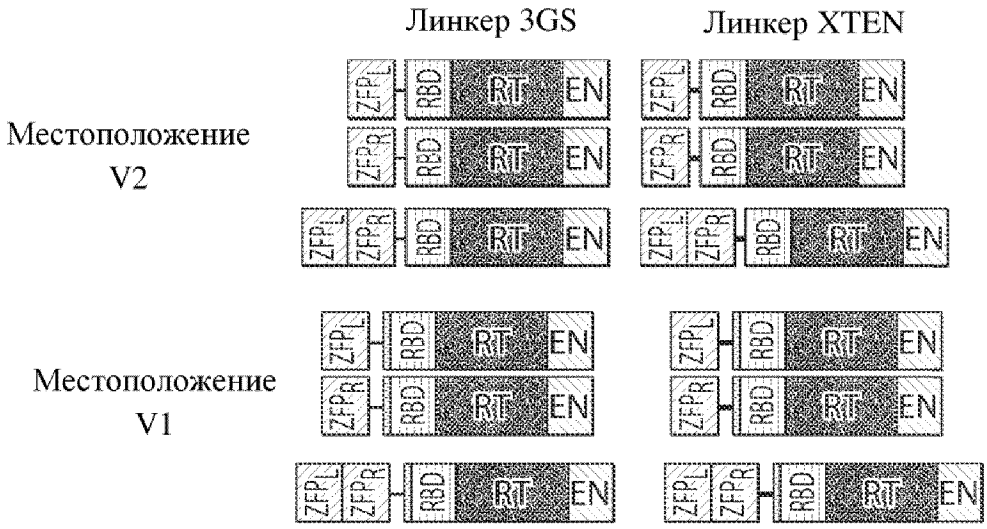
ФИГ. 3А

Вставки/делеции (внесение однонитевых разрывов)
в "посадочных площадках" с мутантным сайтом R2



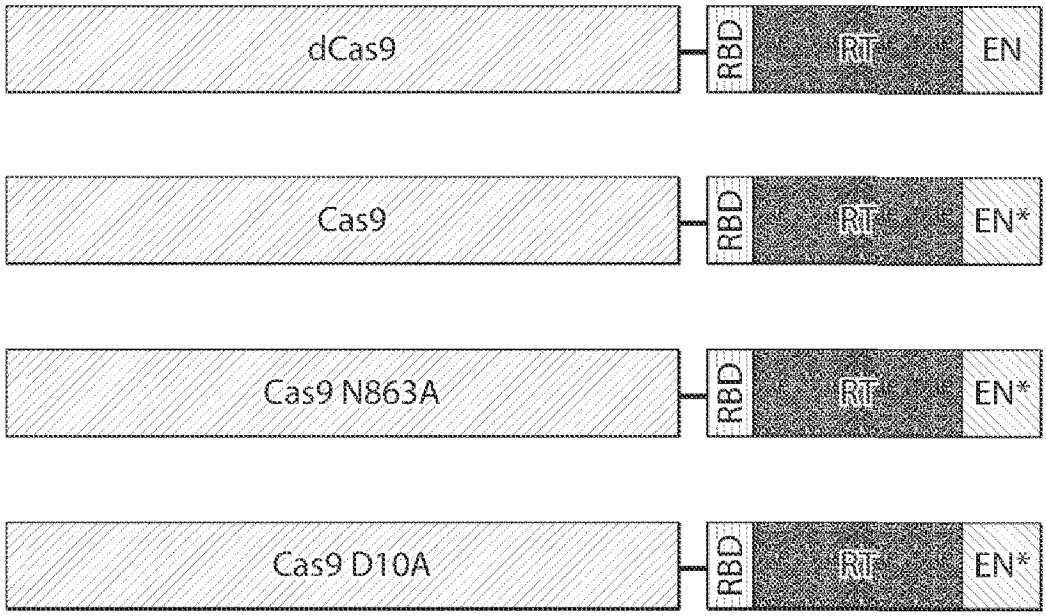
ФИГ. 3В

Примеры перенацеливания полипептида Gene Writer путем замены ДНК-связывающего домена



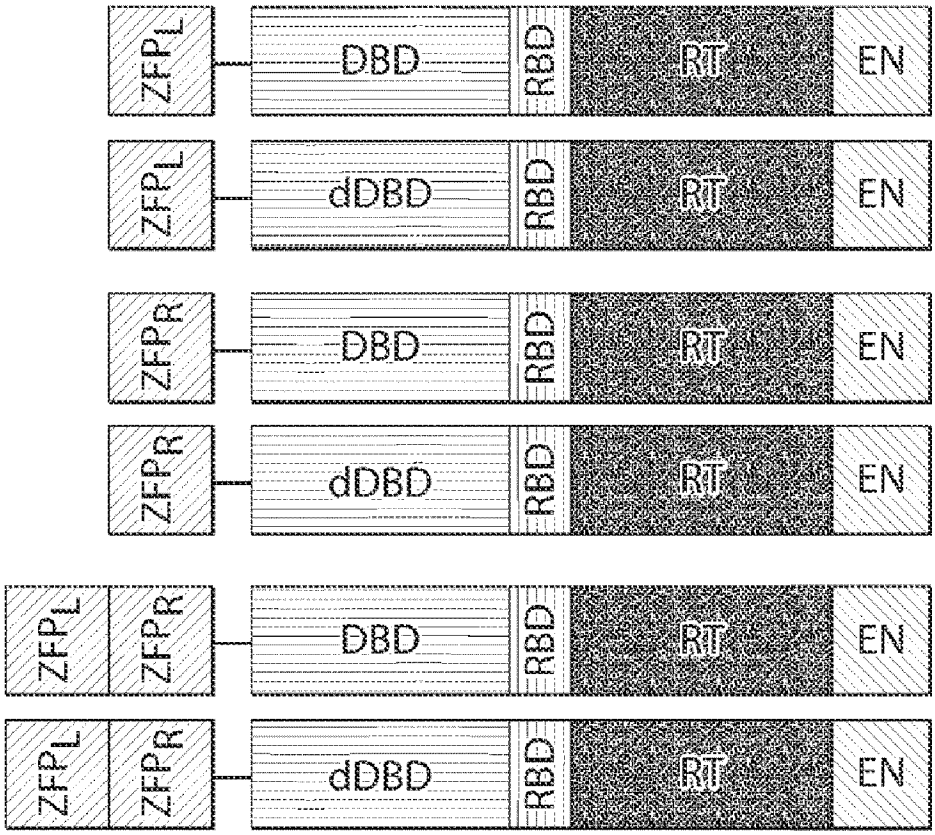
ФИГ. 4

Примеры перенацеливания полипептида Gene Writer путем замены ДНК-связывающего домена



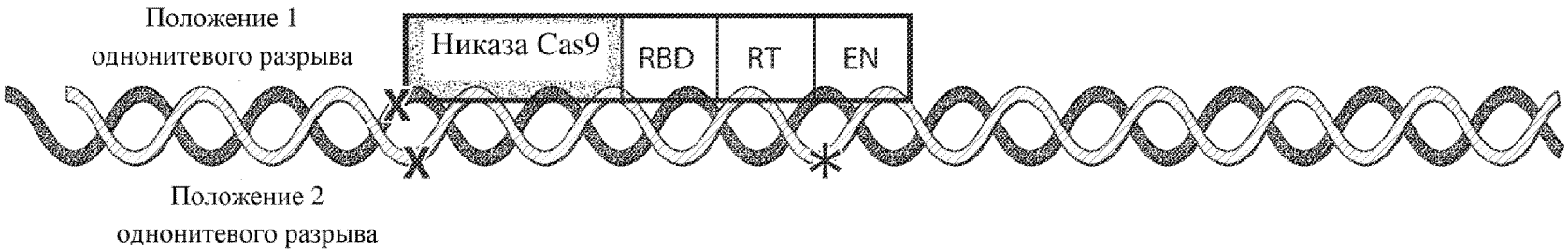
ФИГ. 5

Примеры перенацеливания полипептида Gene Writer путем слияния с концом

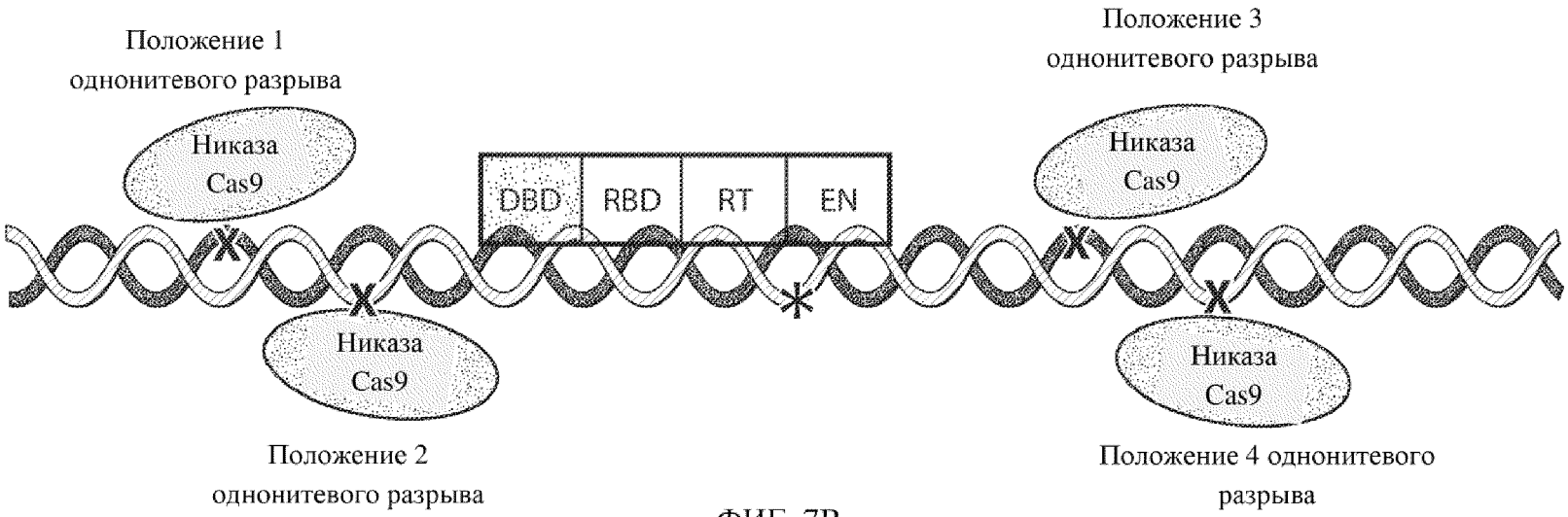


ФИГ. 6

Идеи внесения однонитевого разрыва во вторую нить



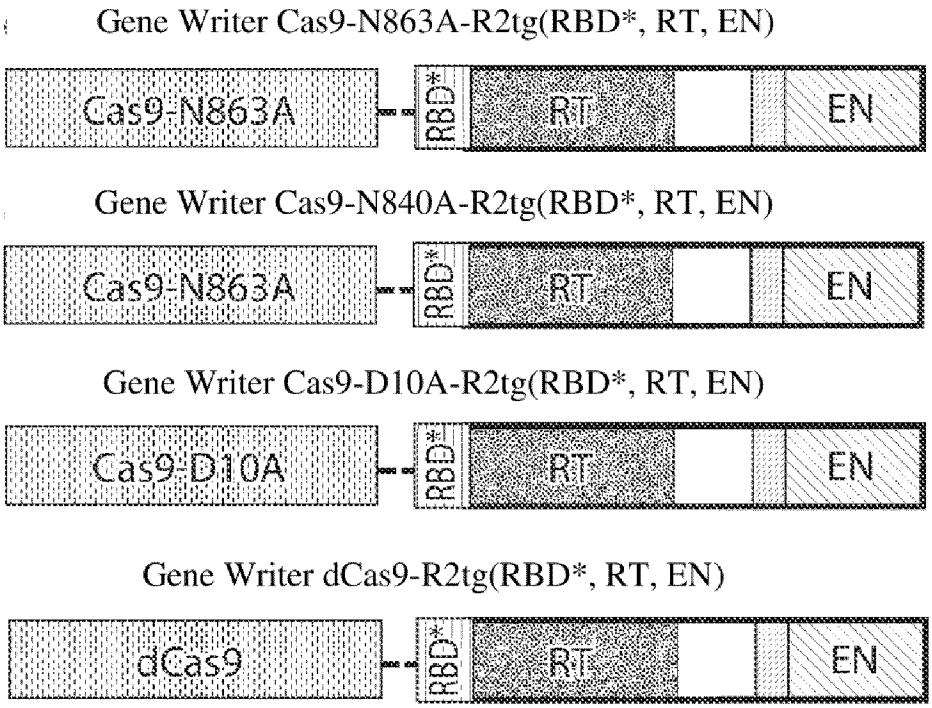
ФИГ. 7A



ФИГ. 7B

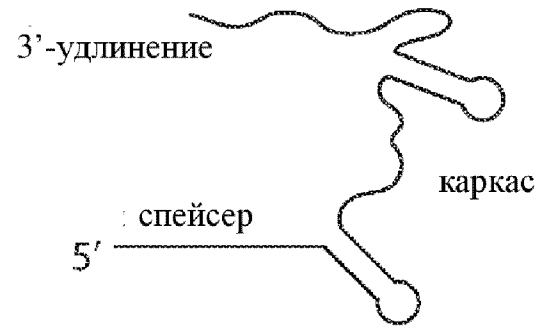
Внесение однонитевого разрыва во вторую нить: слияние Cas9-Gene Writer с удлиненной на 3’-конце донорной gRNA

Слитые конструкции R2Tg-Cas9



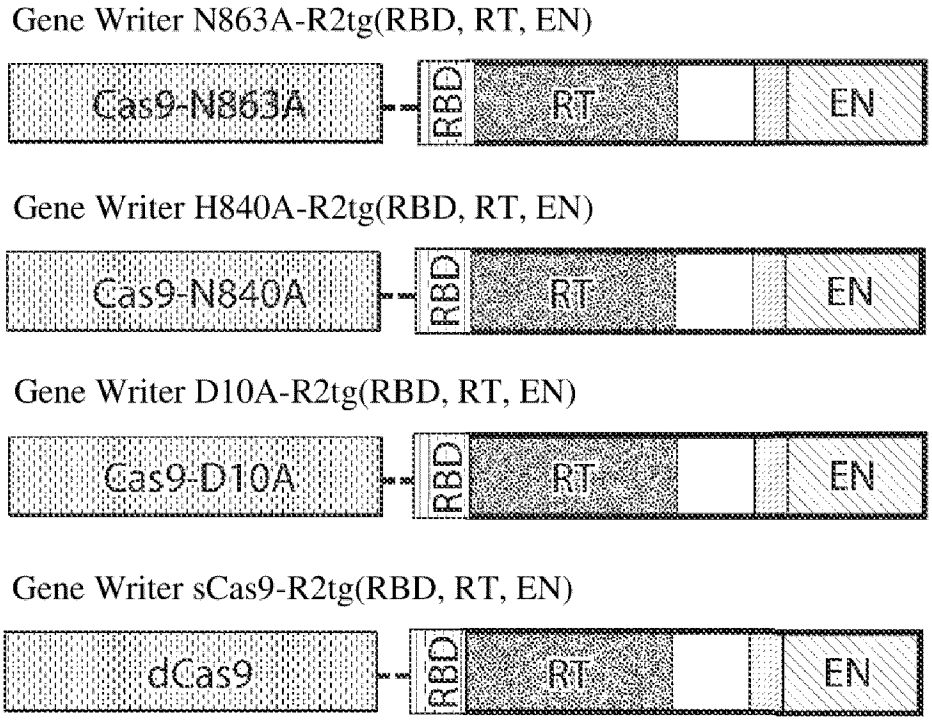
ФИГ. 8A

Донор: gRNA, удлиненная на 5’-конце



ФИГ. 8B

Внесение однонитевого разрыва во вторую нить: Слияния Cas9-Gene Writer с трансгеном, фланкированным UTR



ФИГ. 9A

Донор: UTR-трансген



ФИГ. 9B

Внесение однонитевого разрыва во вторую нить: никаза Cas9



ФИГ. 10A

Донор: UTR-трансген



ФИГ. 10B

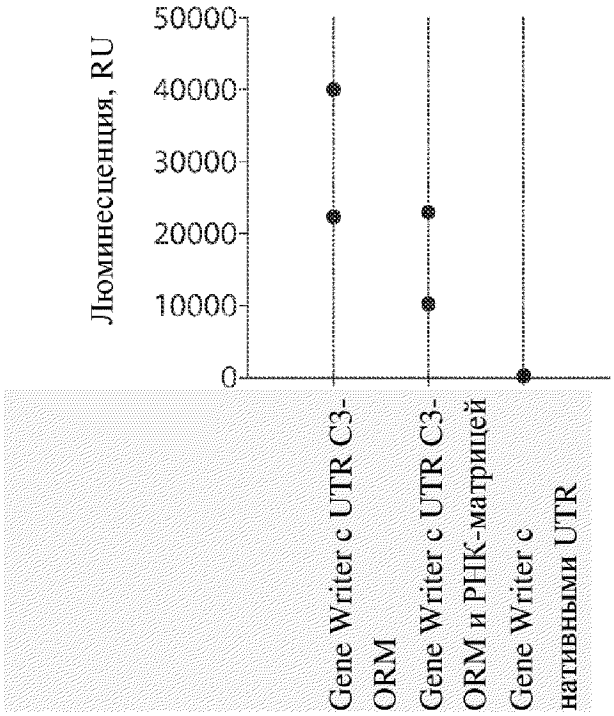


ФИГ. 10C

Улучшенная экспрессия полипептида Gene Writer

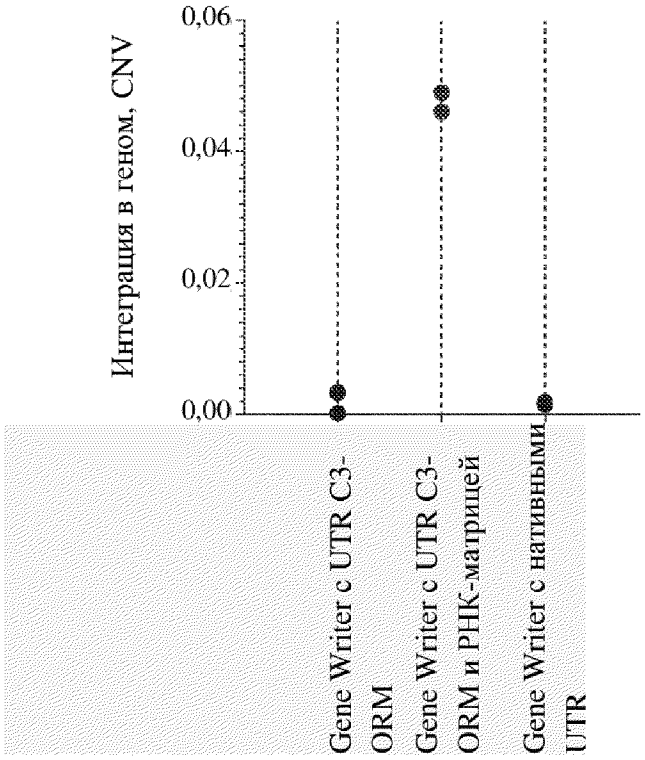


ФИГ. 11A



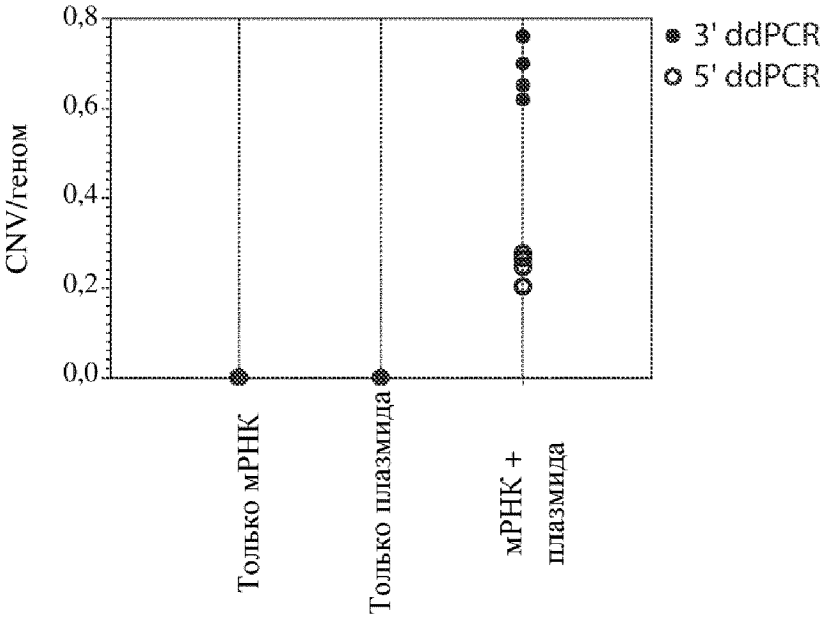
ФИГ. 11B

Улучшенная экспрессия полипептида Gene Writer



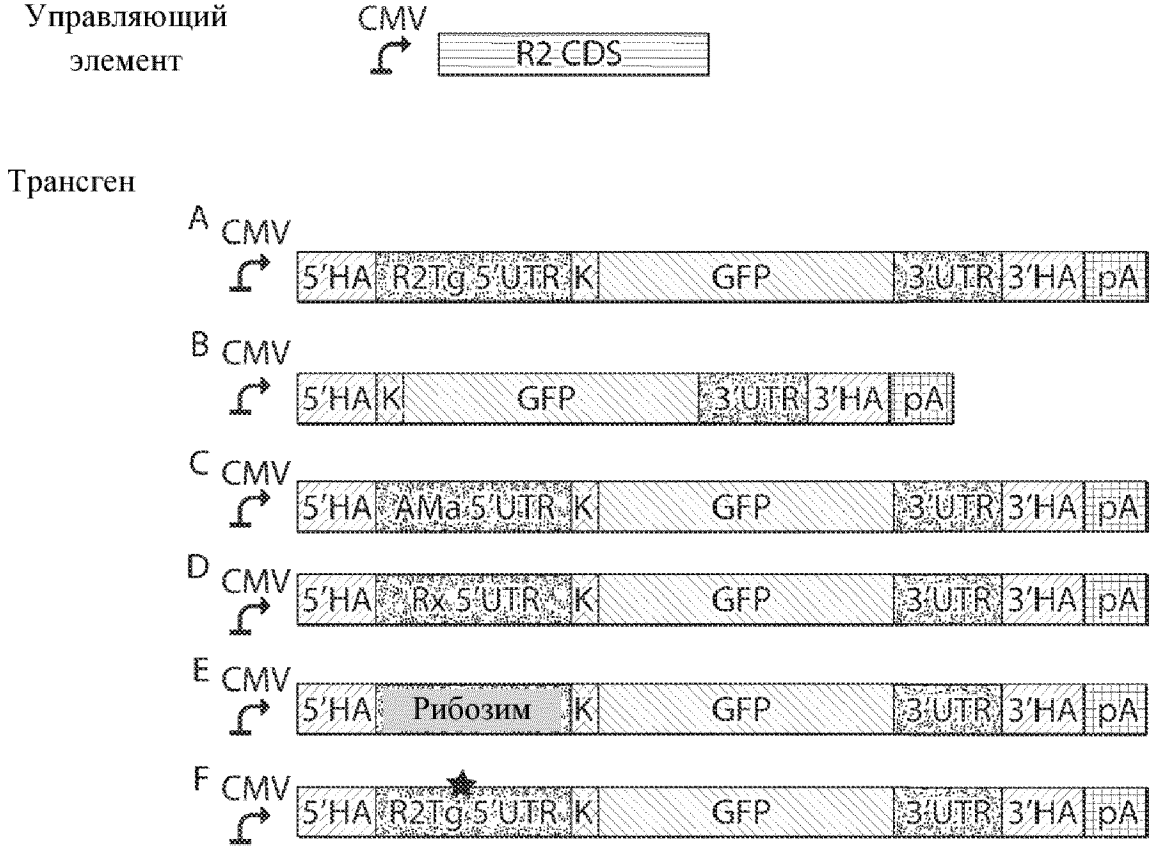
ФИГ. 12

Пример смешанной доставки с помощью РНК/ДНК



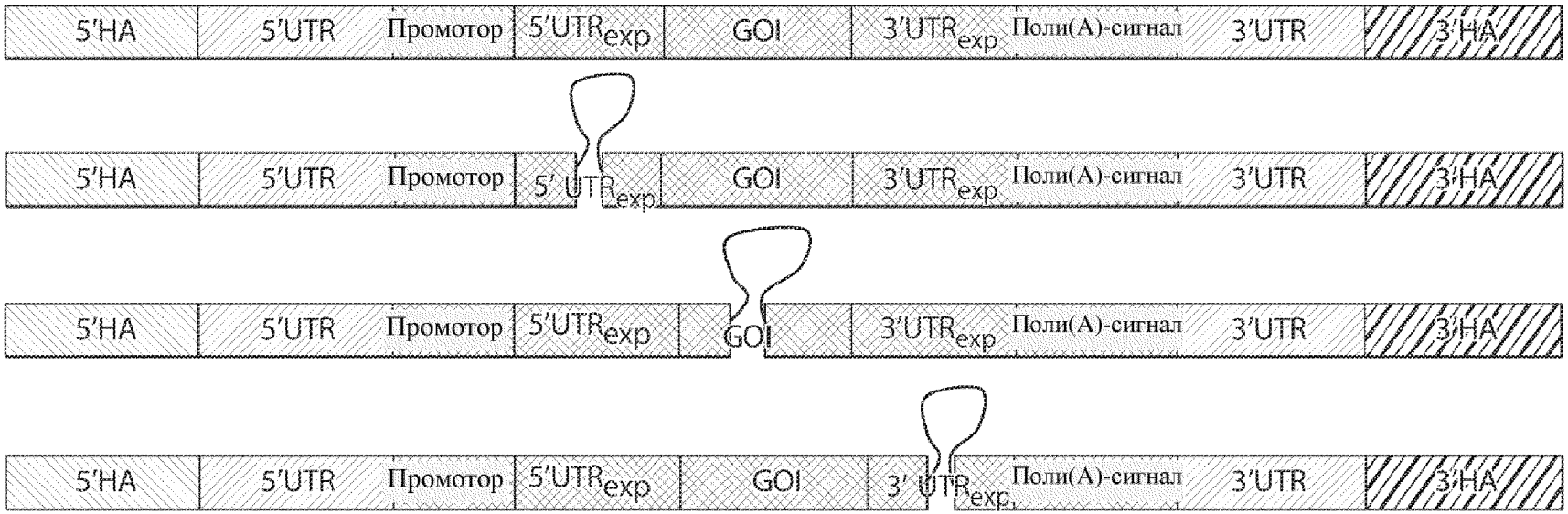
ФИГ. 13

Конструирование 5’-UTR

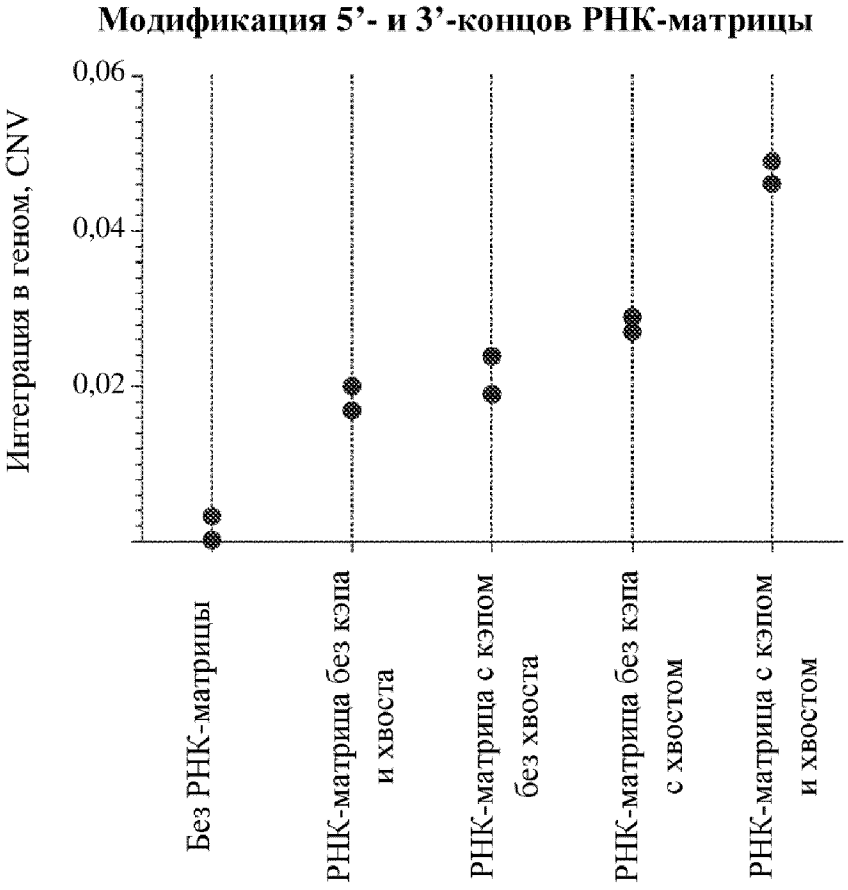


ФИГ. 14

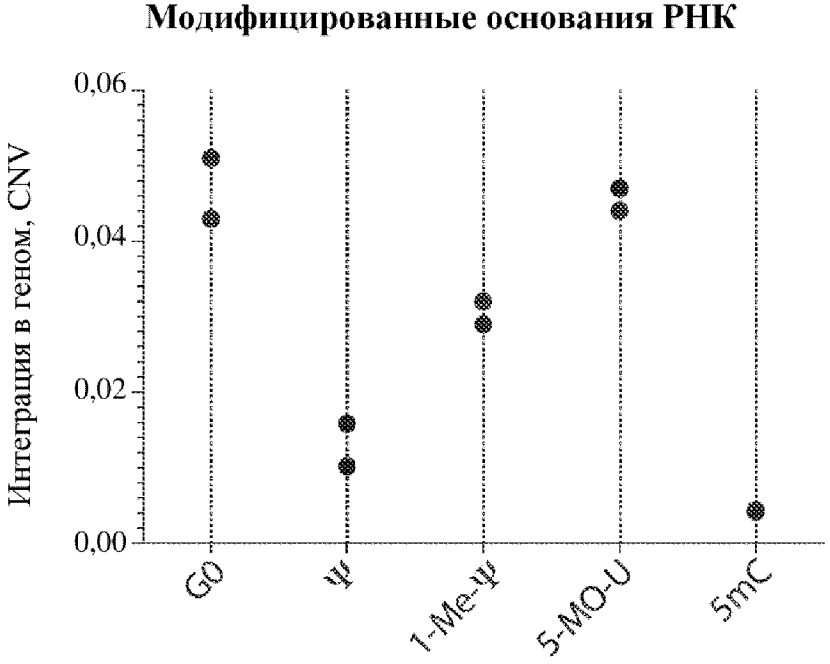
Интроны в РНК-матрице



ФИГ. 15

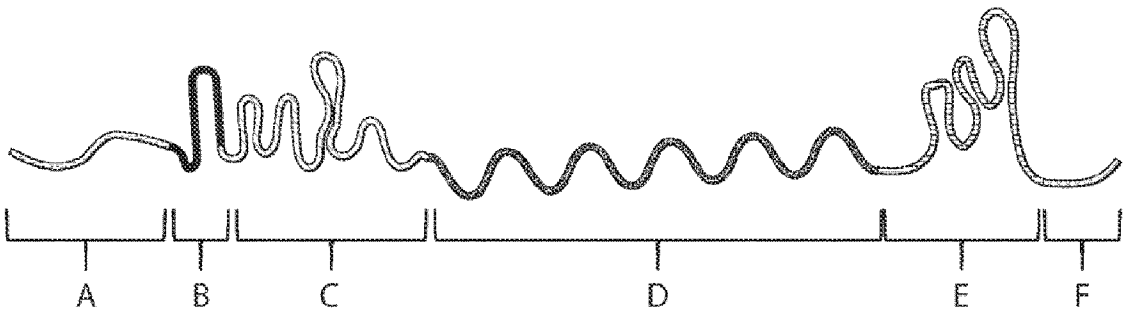

Модификация 5’- и 3’-концов РНК-матрицы
Интеграция в геном, CNV
0,06
0,04
0,02
Без РНК-матрицы
РНК-матрица без кэпа и хвоста
РНК-матрица с кэпом без хвоста
РНК-матрица без кэпа с хвостом
РНК-матрица с кэпом и хвостом

ФИГ. 16


Модифицированные основания РНК
Интеграция в геном, CNV
0,06
0,04
0,02
0,00
G0
Ψ
1-Me-Ψ
5-MO-U
5mC

ФИГ. 17

Модульность матрицы



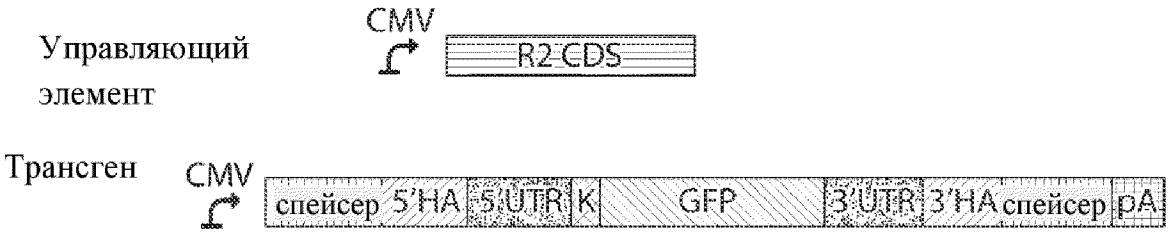
ФИГ. 18

Модульность матрицы

Модуль	Функция
A: 5'-плечо гомологии	Модуль 5'-плечо гомологии является комплементарным последовательности ДНК с 5'-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени
B: Рибозим	Модуль рибозим является подчастью последовательности 5'-UTR ретротранспозазы
C: 5′ UTR	Модуль 5'-UTR представляет собой последовательность РНК или взаимодействующий компонент, который использует управляющий элемент для ретротранспозазы в процессе ретротранспозиции
D: Последовательность гетерологичного объекта	Последовательность гетерологичного объекта представляет собой последовательность РНК, которая служит в качестве матрицы для полипептида Gene Writer для встраивания желаемой нагрузки в целевое местоположение генома
E: 3′ UTR	Модуль 3'-UTR представляет собой последовательность РНК или взаимодействующий компонент, с которым взаимодействует полипептид Gene Writer для связывания с молекулой РНК-матрицы, которая используется полностью и/или частично в качестве матрицы для ретротранспозиции
F: 3'-плечо гомологии	Модуль 3'-плечо гомологии является комплементарным последовательности ДНК с 3'-стороны от места, в котором система Gene Writer производит однонитевой разрыв в ДНК-мишени

ФИГ. 19

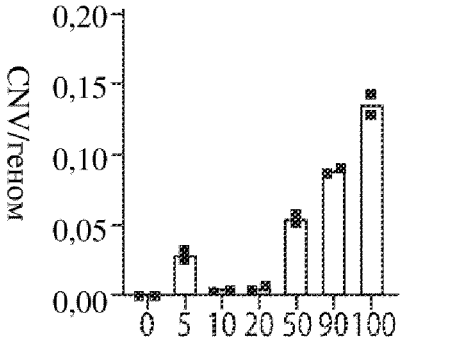
Управляющий элемент и трансген для экспериментов с HA



ФИГ. 20

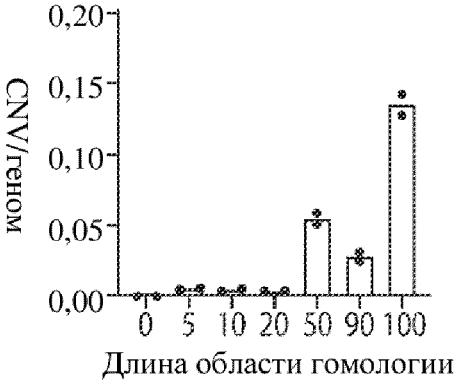
ddPCR для экспериментов с HA

Анализ на 3’ без спейсера



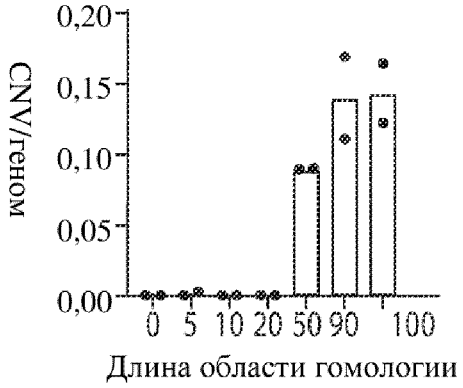
Фиг. 21A

Анализ на 3’ со спейсером



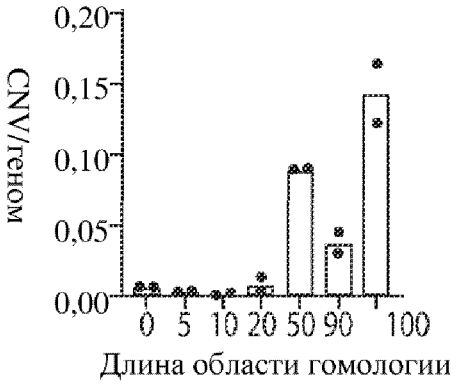
Фиг. 21B

Анализ на 5’ без спейсера



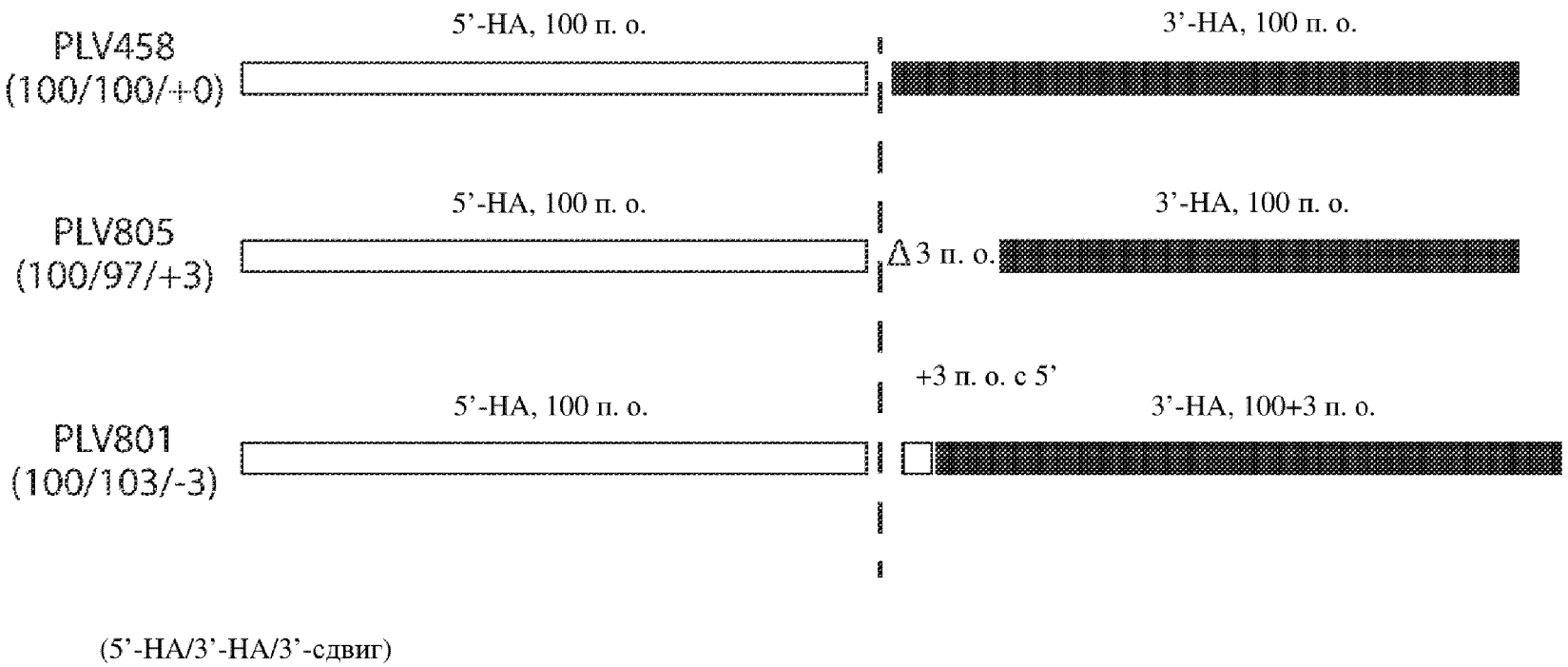
Фиг. 21C

Анализ на 5’ со спейсером



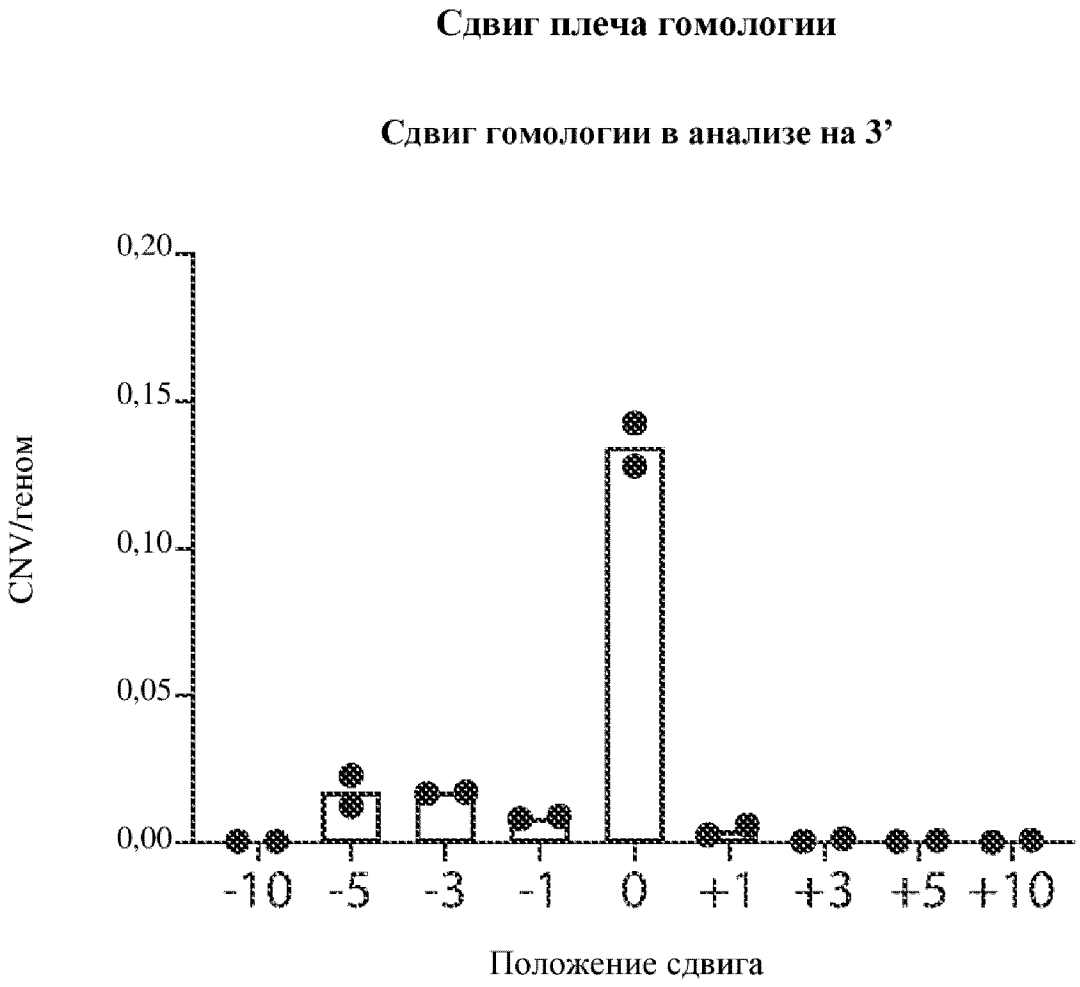
Фиг. 21D

Сдвиг плеча гомологии



(5'-HA/3'-HA/3'-сдвиг)

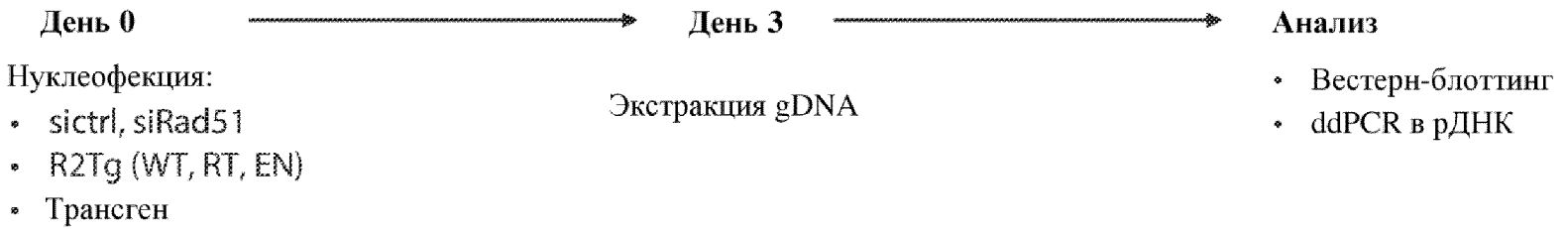
Фиг. 22


Сдвиг плеча гомологии
Сдвиг гомологии в анализе на 3’
CNV/геном
0,20
0,15
0,10
0,05
0,00
-10
-5
-3
-1
0
+1
+3
+5
+10
Положение сдвига

Фиг. 23

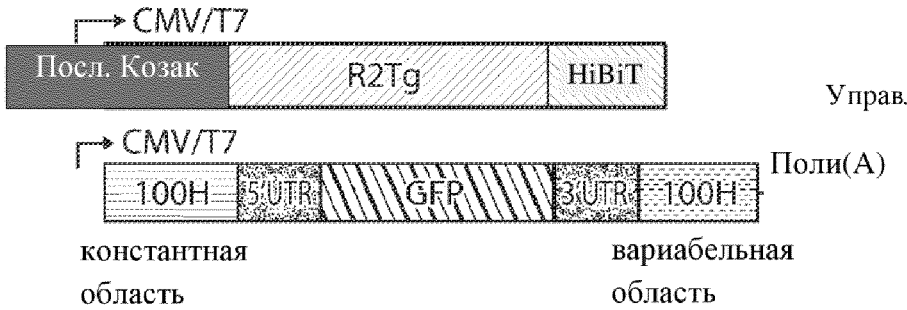
Gene Writing при отсутствии Rad51

Схема эксперимента



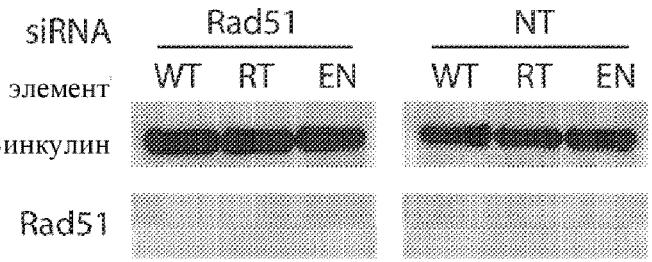
Фиг. 24A

Конфигурация R2Tg и трансгена



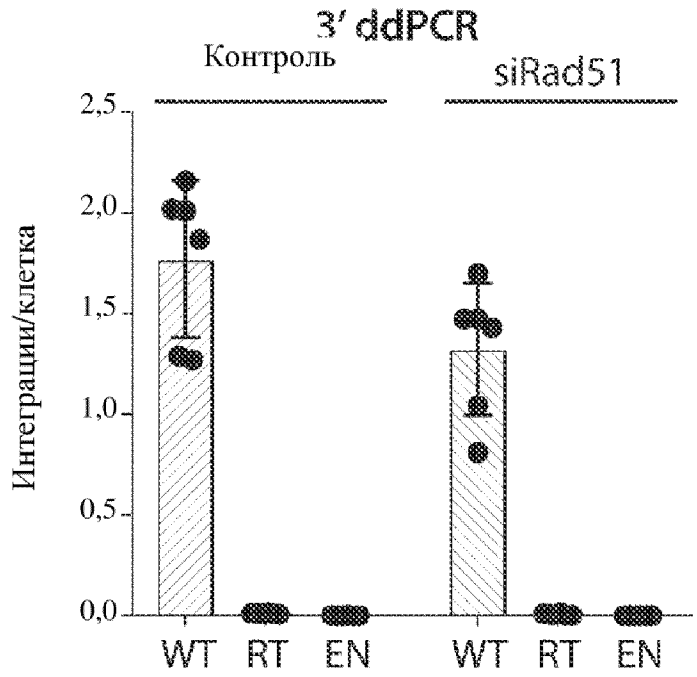
Фиг. 24B

Вестерн-блоттинг

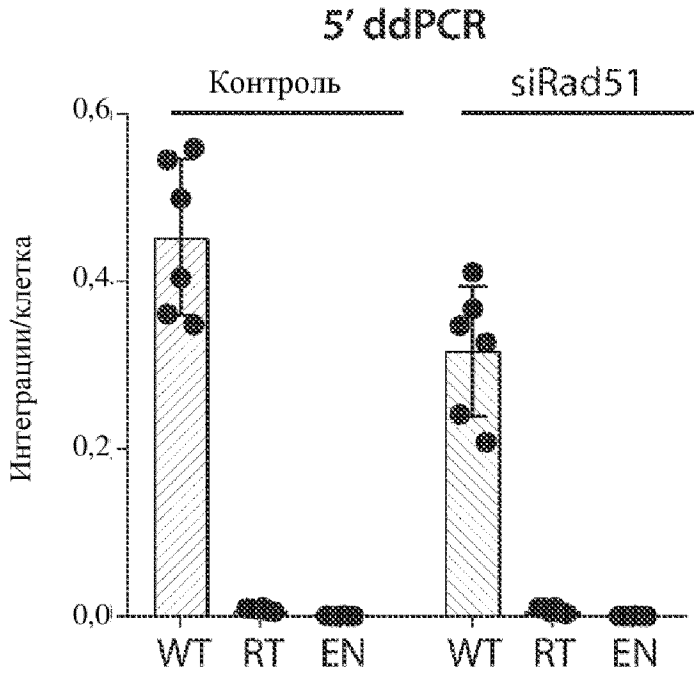


Фиг. 24C

Gene Writing при отсутствии Rad51

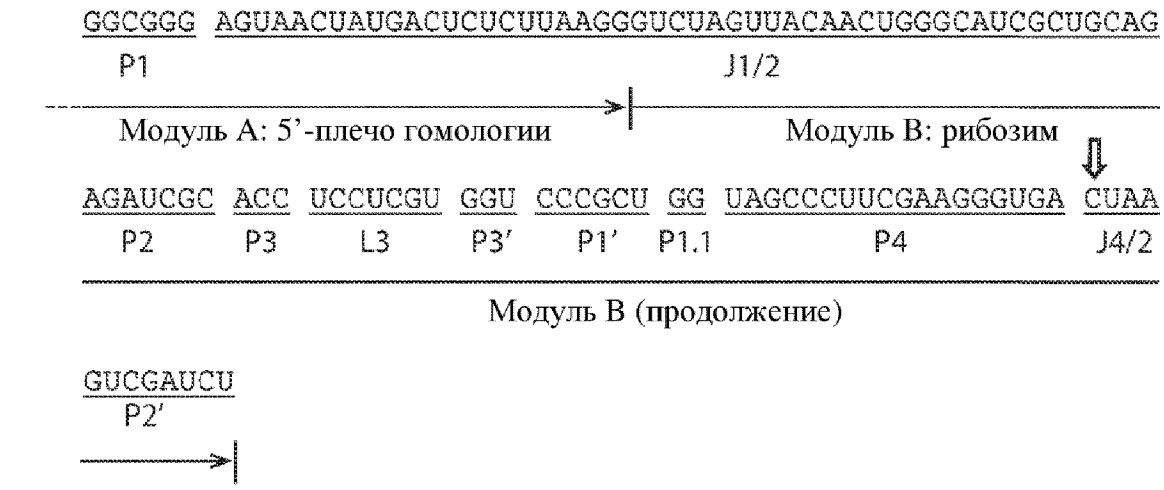


ФИГ. 25A

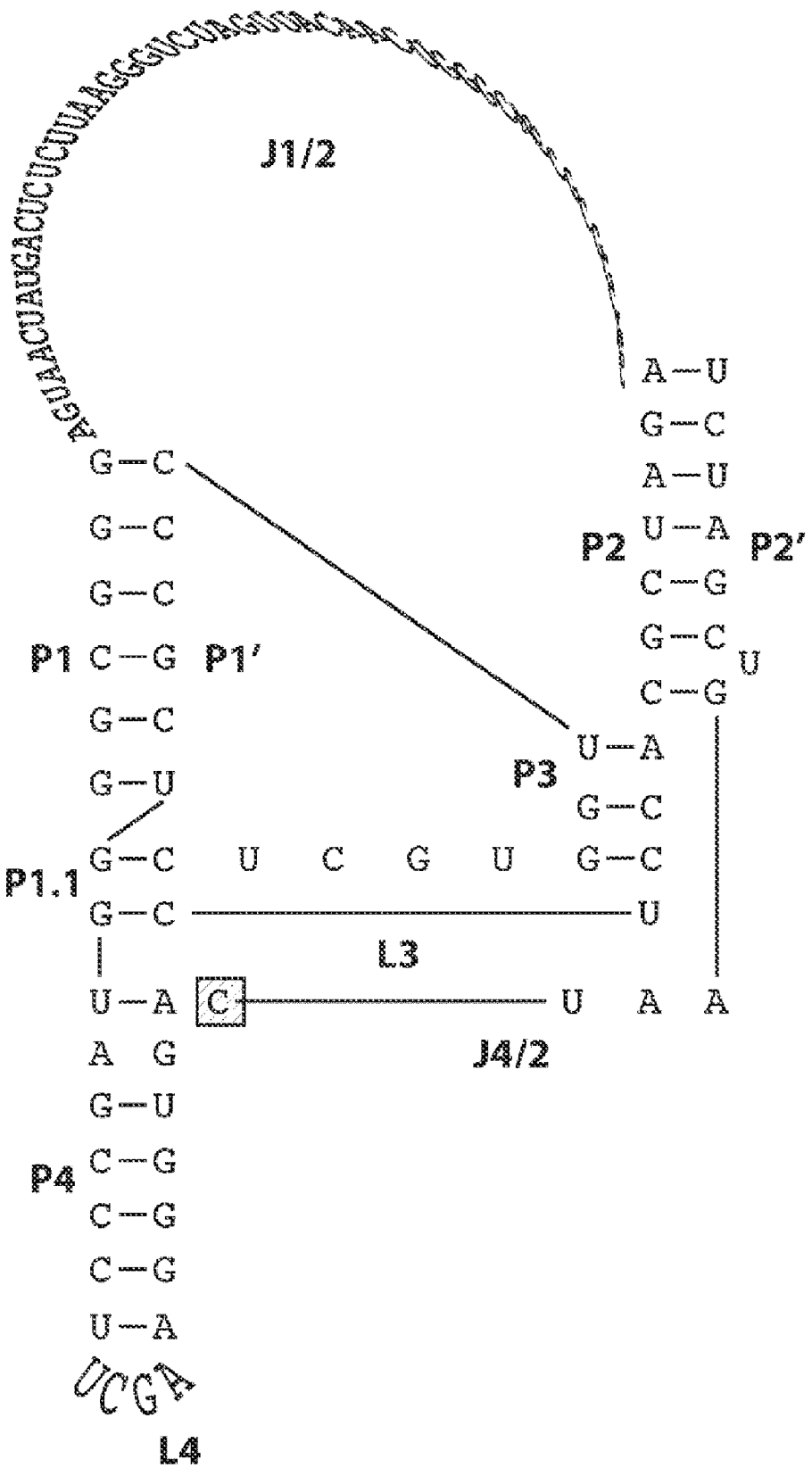


ФИГ. 25B

Рибозим R2Tg

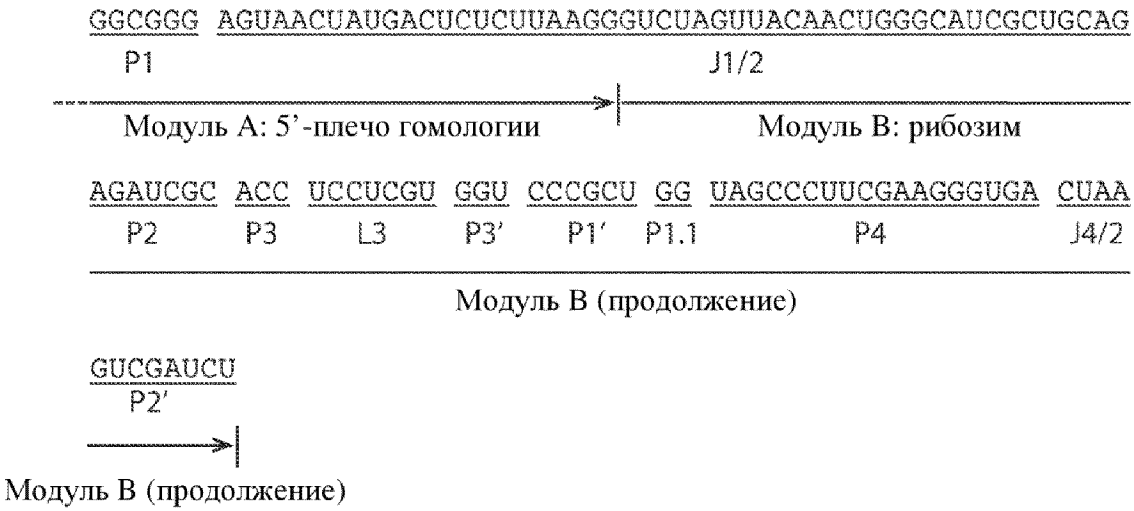


ФИГ. 26A



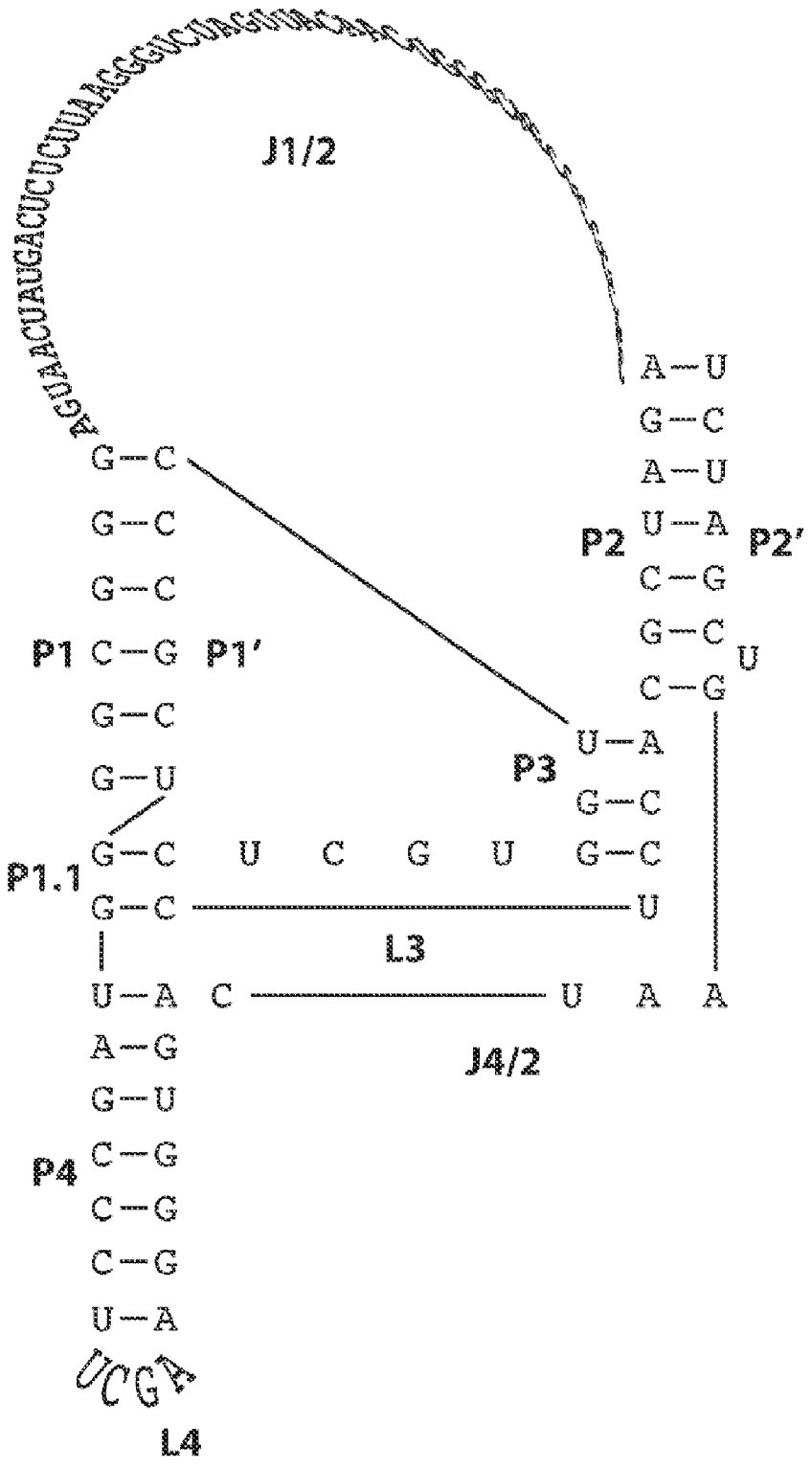
ФИГ. 26B

Последовательность рибозима R2Tg



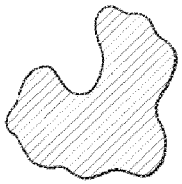
ФИГ. 27

Предсказанная вторичная структура рибозима R2Tg

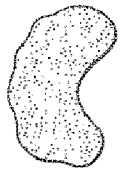


ФИГ. 28

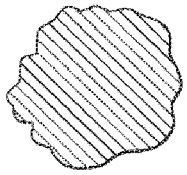
Модули Gene Writer



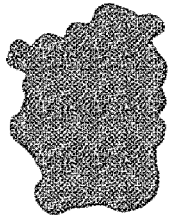
<u>ДНК-связывающий</u>
RTn
"Цинковый палец"
Домен TAL
Cas9, dCas9, никаза Cas9
Фактор транскрипции
Мегануклеаза



<u>РНК-связывающий</u>
Белок B-бокса
Белок оболочки MS2
Белок dCas
RTn



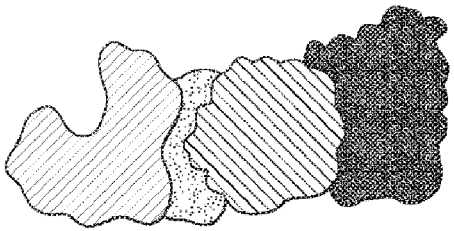
<u>Обратная транскриптаза</u>
RTn
Другая RT
Пептид, содержащий домен RT



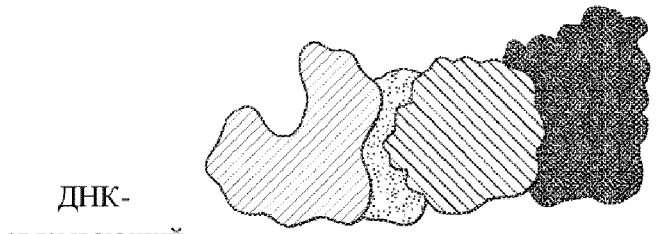
<u>Эндонуклеаза</u>
RTn
Cas9, никаза Cas 9
Рестрикционный фермент (например, типа II, например, FokI)
Мегануклеаза
Резольваза для структур Холлидея
RLE
APE
GIY-YIG

ФИГ. 29A

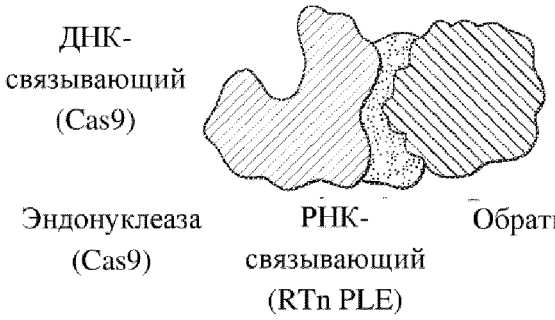
Gene Writer с ретротранспозазой дикого типа (например, R2)



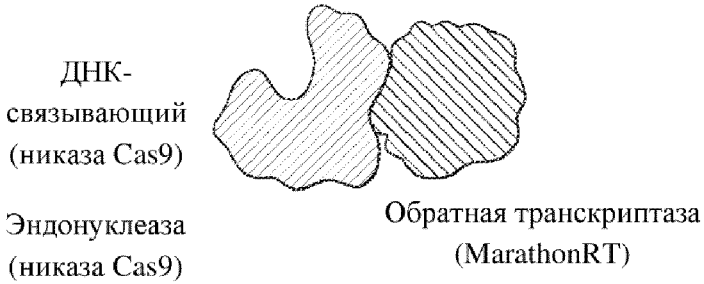
Gene Writer, все домены которого являются гетерологичными (например, ZF-R2-MarathonRT-FokI)



Gene Writer, перенацеленный с помощью Cas9 (например, Cas9-PLE)

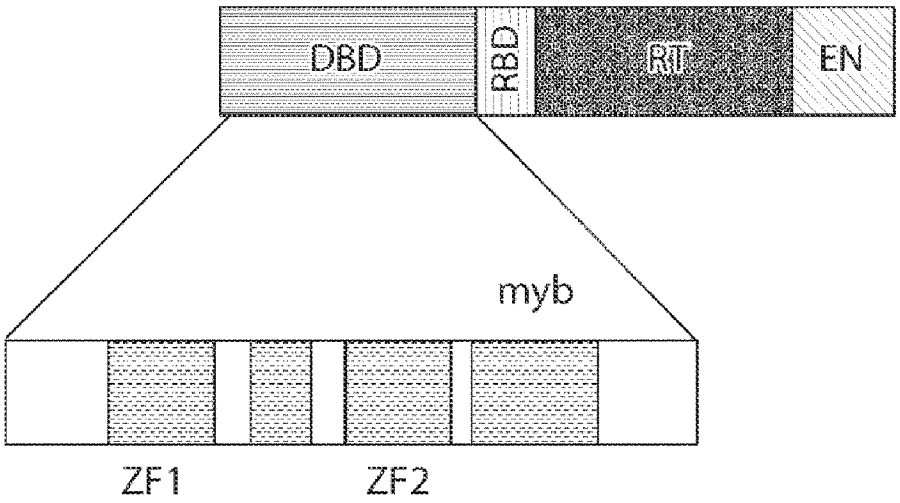


Gene Writer, перенацеленный с помощью Cas9 (например, Cas9-MarathonRT)

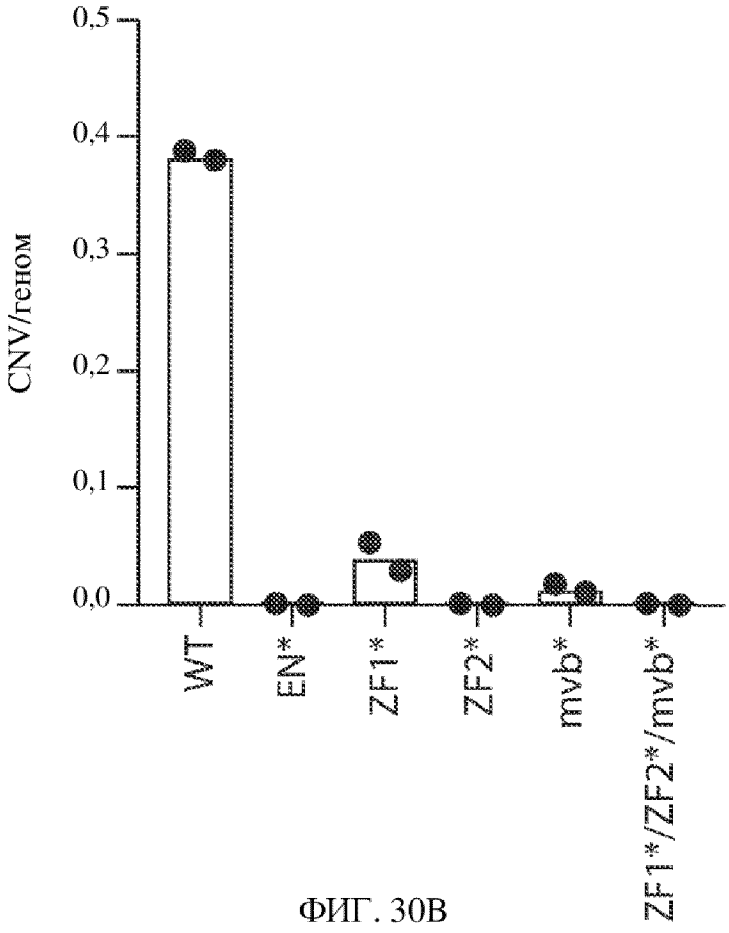


ФИГ. 29В

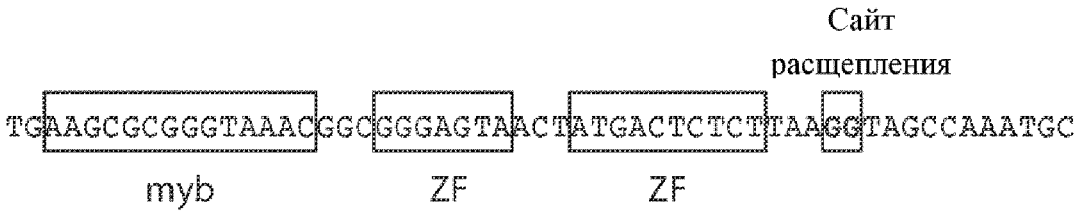
Делеция природного линкера



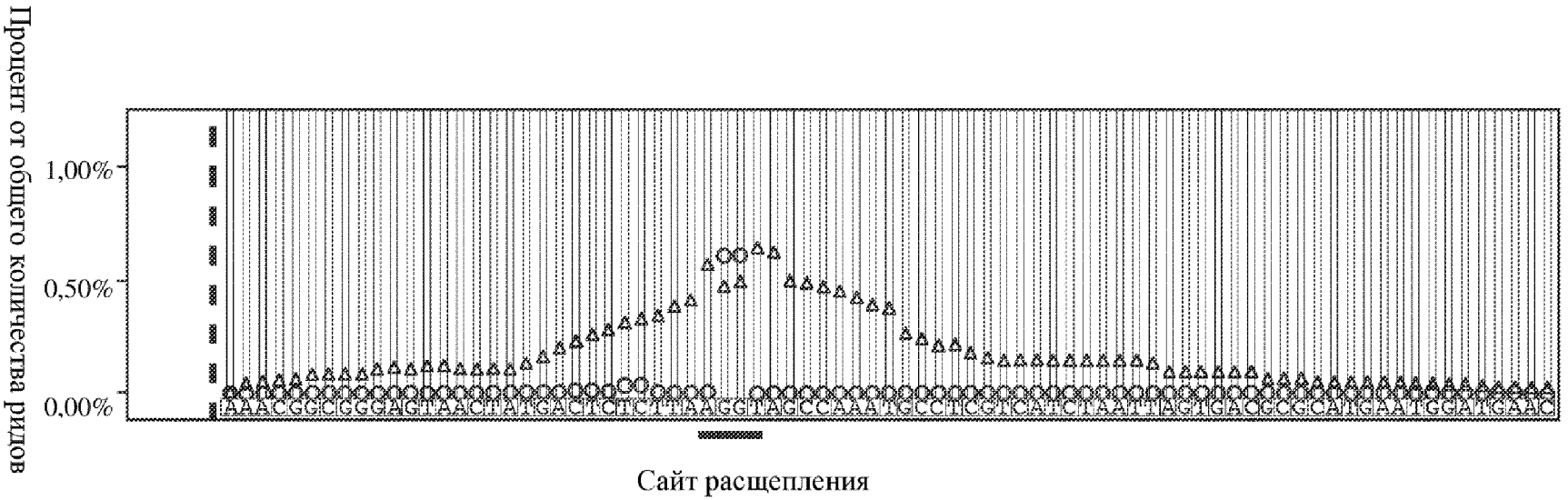
ФИГ. 30A



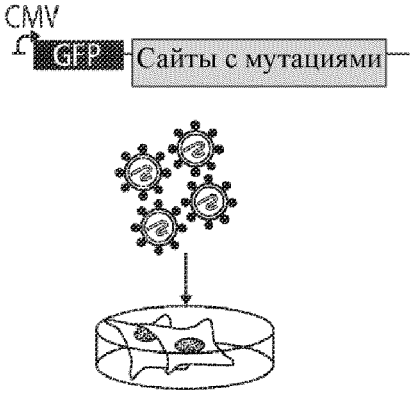
ФИГ. 30B



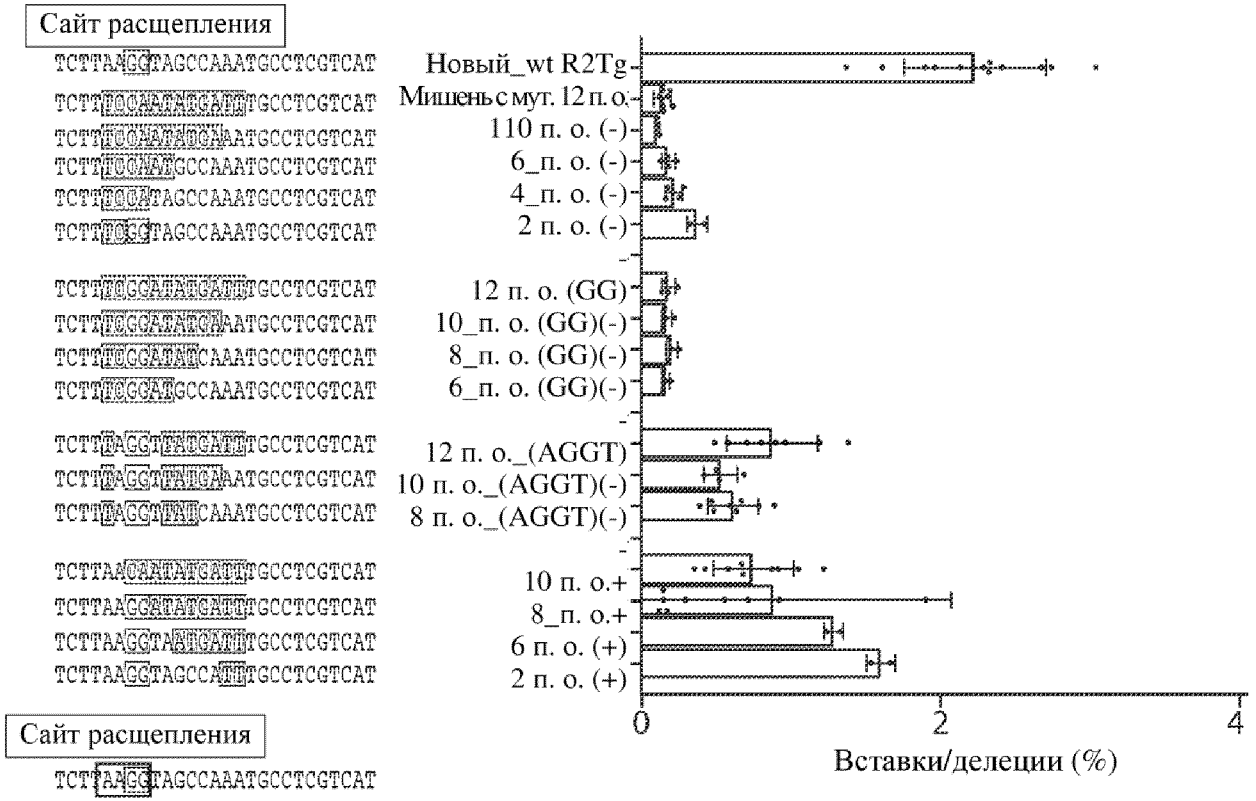
ФИГ. 31A



ФИГ. 31B

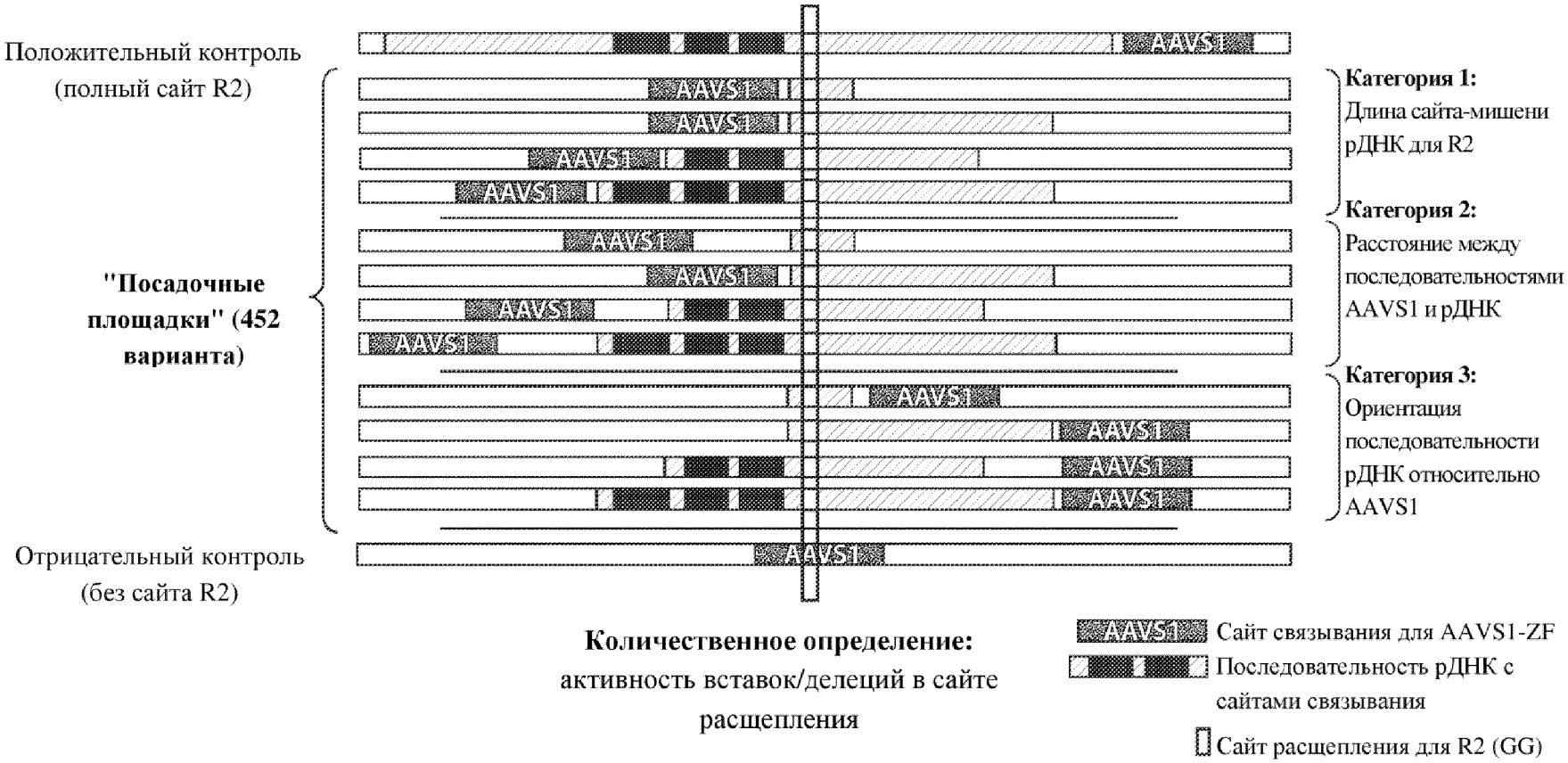


ФИГ. 32A



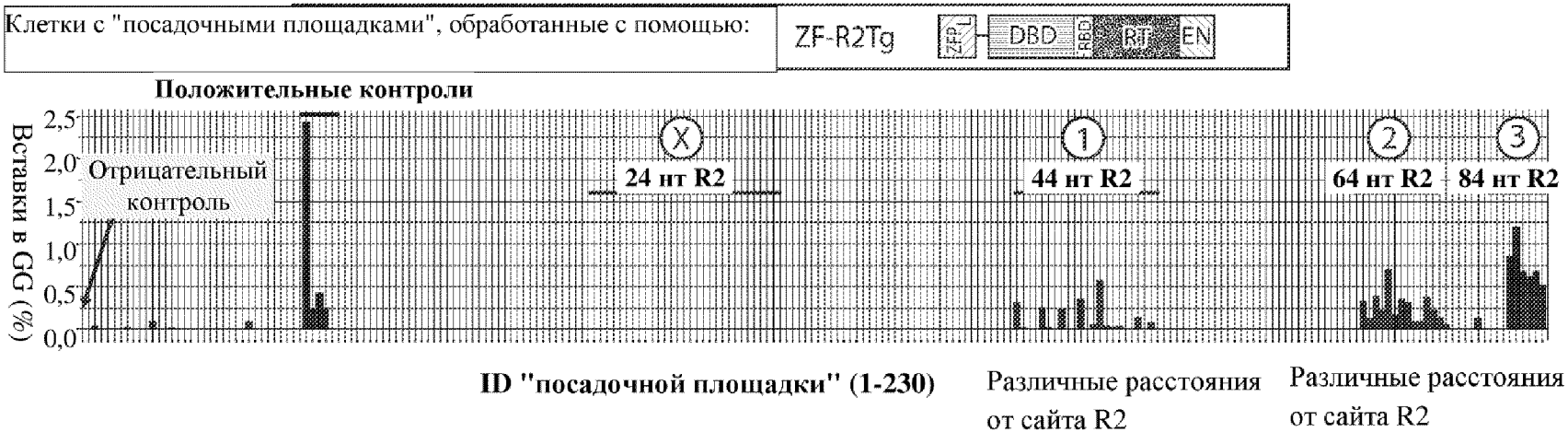
Мутация двух п. о., прилегающих к сайту расщепления,
приводит к значительной утрате расщепления

ФИГ. 32B

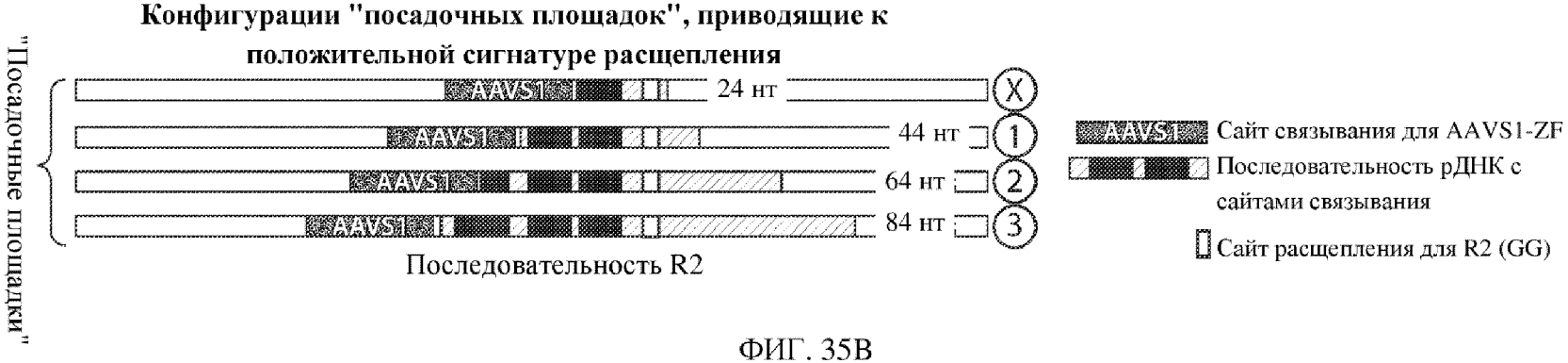

Положительный контроль
(полный сайт R2)
"Посадочные площадки" (452 варианта)
Отрицательный контроль
(без сайта R2)
AAVS1
Категория 1:
Длина сайта-мишени рДНК для R2
Категория 2:
Расстояние между последовательностями AAVS1 и рДНК
Категория 3:
Ориентация последовательности рДНК относительно AAVS1
Количественное определение:
активность вставок/делеций в сайте расщепления
Сайт связывания для AAVS1-ZF
Последовательность рДНК с сайтами связывания
Сайт расщепления для R2 (GG)

ФИГ. 33

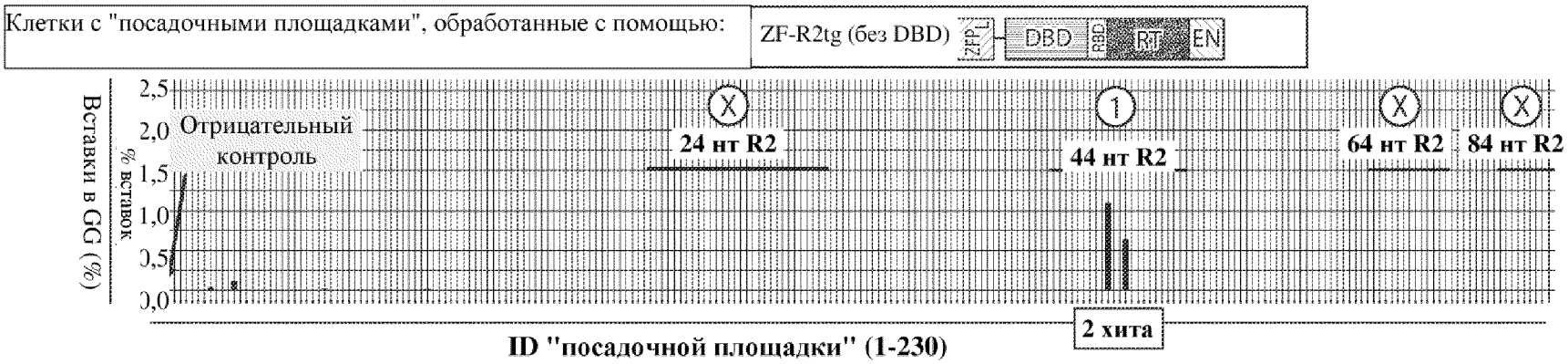
ФИГ. 34



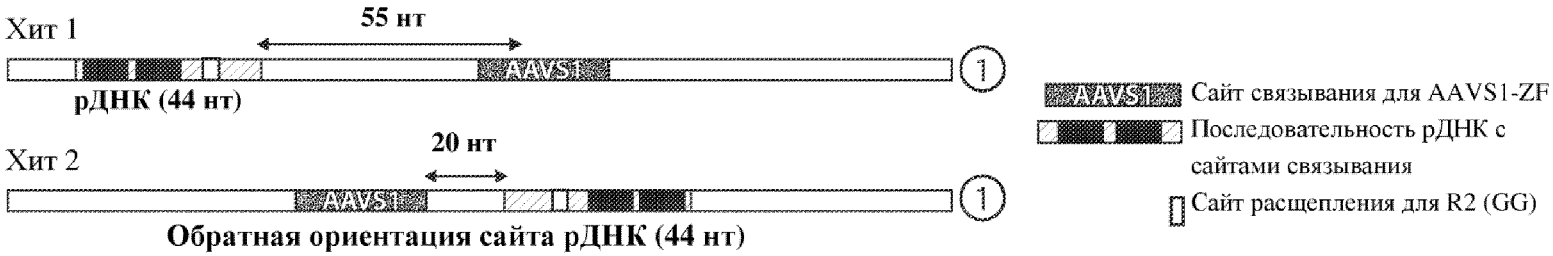
ФИГ. 35A



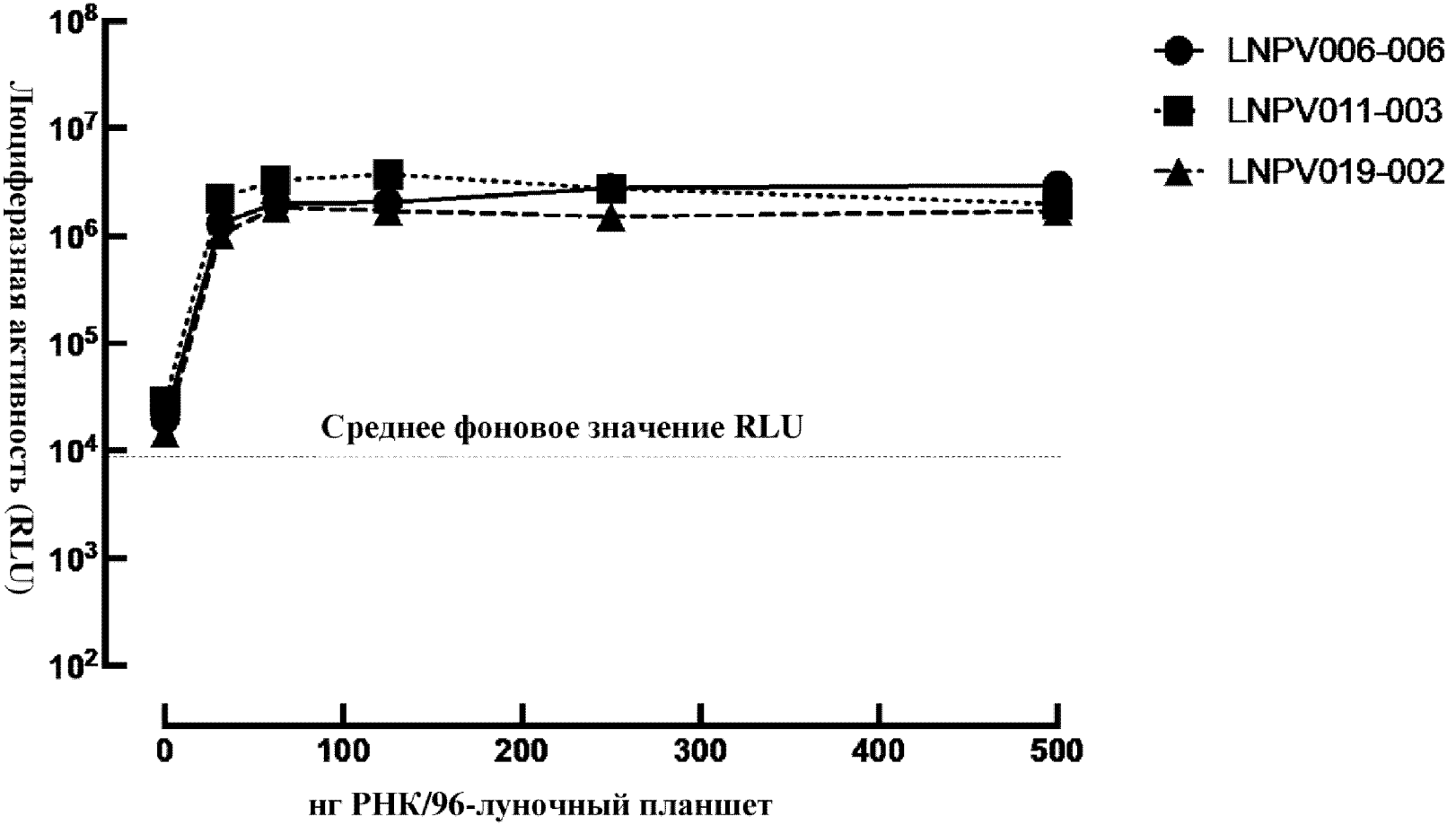
ФИГ. 35B



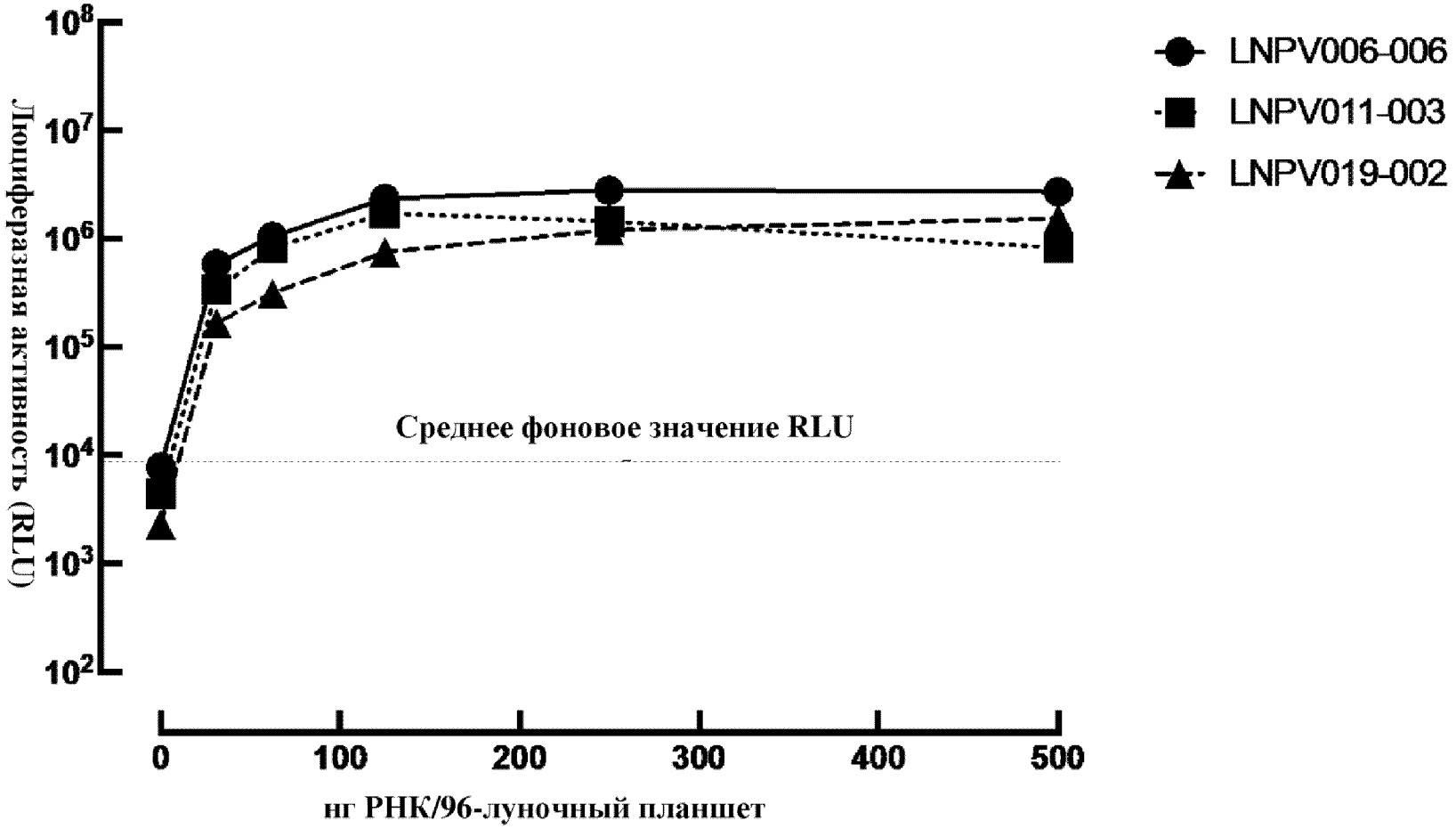
ФИГ. 36A



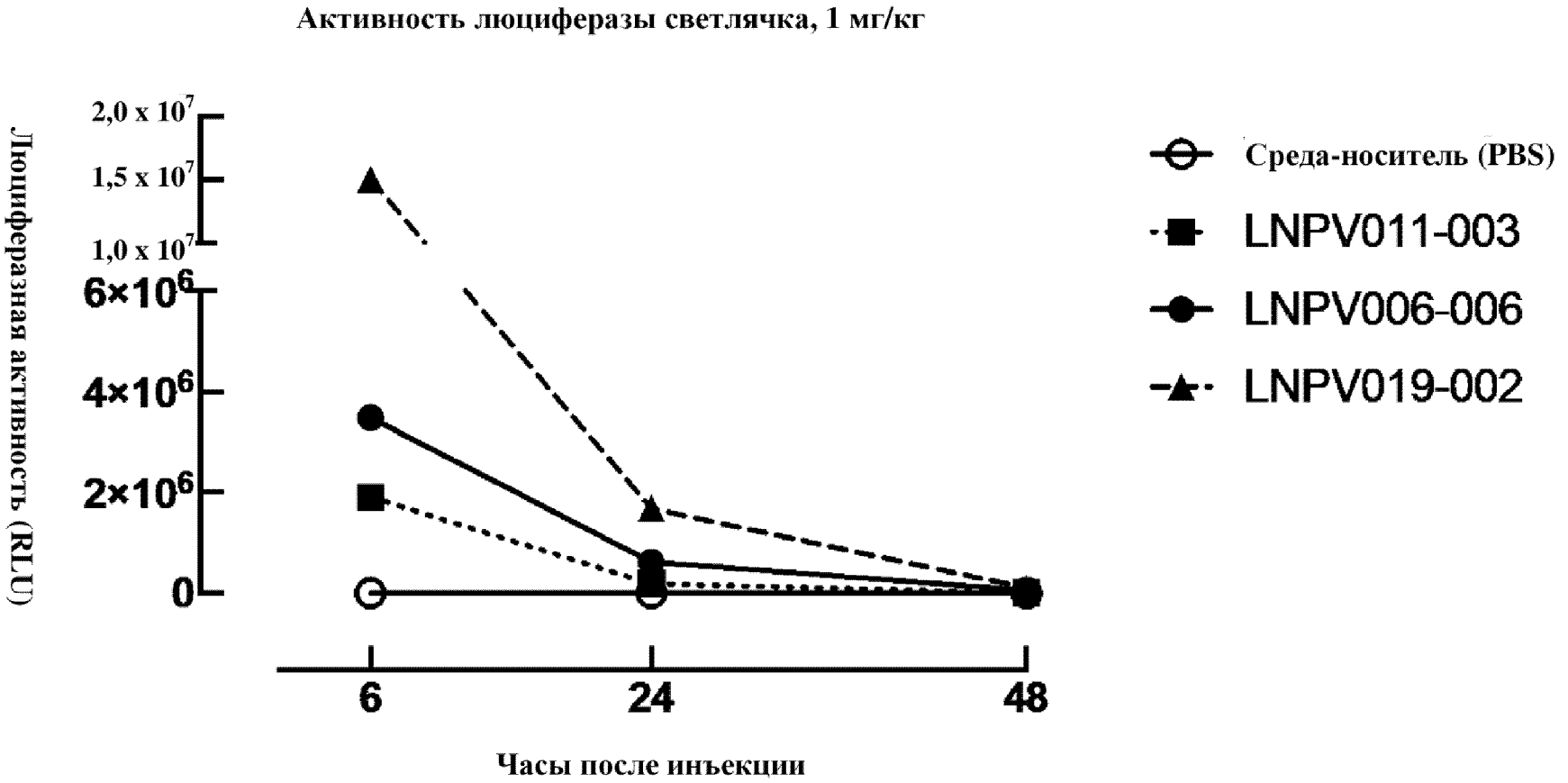
ФИГ. 36B


Дозозависимая активность люциферазы в первичных гепатоцитах человека (PHH)
Люциферазная активность (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Среднее фоновое значение RLU
0
100
200
300
400
500
нг РНК/96-луночный планшет
LNPV006-006
LNPV011-003
LNPV019-002

ФИГ. 37A


Дозозависимая активность люциферазы в первичных гепатоцитах мыши (PMH)
Люциферазная активность (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Среднее фоновое значение RLU
0
100
200
300
400
500
нг РНК/96-луночный планшет
LNPV006-006
LNPV011-003
LNPV019-002

ФИГ. 37B


Активность люциферазы светлячка, 1 мг/кг
Люциферазная активность (RLU)
2,0 x 10^7
1,5 x 10^7
1,0 x 10^7
6×10^6
4×10^6
2×10^6
0
6
24
48
Часы после инъекции
Среда-носитель (PBS)
LNPV011-003
LNPV006-006
LNPV019-002

ФИГ. 38

Cas9(N863A)-[линкер]-RT(R2BmRBDmut))
Cas9
10 нг
50 нг
200 нг
800 нг
Cas9-22-RT(MMLV)
Без Cas9/линкера
22 (исходный)
1
2
3
7
8
10
9
11
12
15
16
Без Cas9/линкера
22 (исходный)
4
5
6
13
14
17
18
19
20
21
α-Cas9
Контроль нагрузки

ФИГ. 39