(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 September 2021 (10.09.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/178709 A1

(51) International Patent Classification:
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2021/020933

(22) International Filing Date:
04 March 2021 (04.03.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/985,291	04 March 2020 (04.03.2020)	US
63/035,638	05 June 2020 (05.06.2020)	US

(71) Applicant: **FLAGSHIP PIONEERING INNOVATIONS VI, LLC** [US/US]; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US).

(72) Inventors: **STEINBERG, Barrett, Ethan**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **BOTHMER, Anne, Helen**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **SALOMON, William, Edward**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **SHCHERBAKOVA, Inna**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **COTTA-RAMUSINO, Cecilia, Giovanna, Silvia**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **RUBENS, Jacob, Rosenblum**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **CITORIK, Robert, James**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **WANG, Zi, Jun**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US).

(74) Agent: **KOYFMAN, Hannah, R.** et al.; Lando & Anastasi, LLP, 60 State Street, 23rd Floor, Boston, MA 02109 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: METHODS AND COMPOSITIONS FOR MODULATING A GENOME

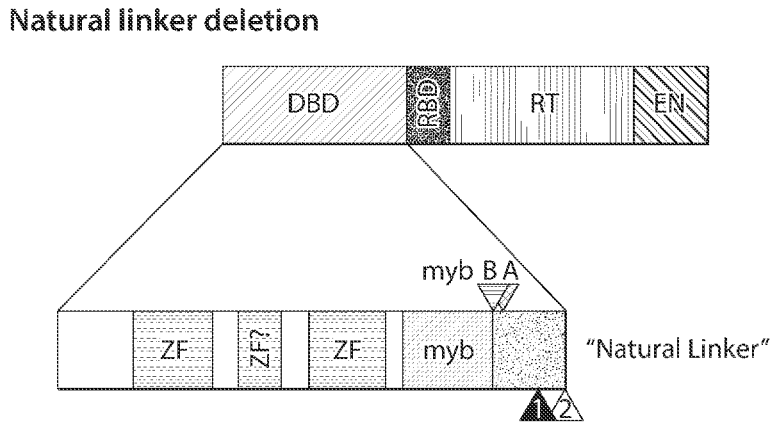


FIG. 1A

(57) Abstract: Methods and compositions for modulating a target genome are disclosed.

[Continued on next page]

WO 2021/178709 A1

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS AND COMPOSITIONS FOR MODULATING A GENOME

RELATED APPLICATIONS

This application claims priority to U.S. Serial No.: 62/985,291 filed Mar 4 2020 and U.S. Serial No.: 63/035638 filed Jun 5 2020, the entire contents of each of which is incorporated herein by reference.

BACKGROUND

Integration of a nucleic acid of interest into a genome occurs at low frequency and with little site specificity, in the absence of a specialized protein to promote the insertion event. Some existing approaches, like CRISPR/Cas9, are more suited for small edits and are less effective at integrating longer sequences. Other existing approaches, like Cre/loxP, require a first step of inserting a loxP site into the genome and then a second step of inserting a sequence of interest into the loxP site. There is a need in the art for improved proteins for inserting sequences of interest into a genome.

SUMMARY OF THE INVENTION

This disclosure relates to novel compositions, systems and methods for altering a genome at one or more locations in a host cell, tissue or subject, *in vivo* or *in vitro*. In particular, the invention features compositions, systems and methods for the introduction of exogenous genetic elements into a host genome.

Features of the compositions or methods can include one or more of the following enumerated embodiments.

1. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the polypeptide comprises a mutation inactivating and/or deleting a nucleolar localization signal.

2. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a first target DNA binding domain, e.g., comprising a first Zn finger domain, (ii) a reverse transcriptase domain, (iii) an endonuclease domain, and (iv) a second target DNA binding domain, e.g., comprising a second Zn finger domain, heterologous to the first target DNA binding domain; and

optionally (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein (a) binds to a smaller number of target DNA sequences in a target cell than a similar polypeptide that comprises only the first target DNA binding domain, e.g., wherein the presence of the second target DNA binding domain in the polypeptide with the first DNA binding domain refines the target sequence specificity of the polypeptide relative to the polypeptide target sequence specificity of the polypeptide comprising only the first target DNA binding domain.

3. The system of embodiment 2, wherein (iii) comprises (iv).

4. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

optionally, (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the system is capable of cutting the first strand of the target DNA at least twice (e.g., twice), and

optionally wherein the cuts are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or 200 nucleotides away one another (and optionally no more than 500, 400, 300, 200, or 100 nucleotides away from one another).

5. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

optionally, (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the system is capable of cutting the first strand and the second strand of the target DNA, and

wherein the distance between the cuts is the same as the distance between cuts made by the endonuclease domain, e.g., the endonuclease domain of a naturally occurring retrotransposase.

6. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein (a), (b), or (a) and (b) further comprises a 5' UTR and/or 3' UTR operably linked to the sequence encoding the polypeptide, the heterologous object sequence (e.g., a coding sequence contained in the heterologous object sequence), or both.

7. The system of embodiment 6, wherein the 5' UTR and/or 3' UTR increase expression of the operably linked sequence(s) by at least 10%, 20%, 30%, 40%, 50%, 70%, 70%, 80%, 90%, or 100% relative to an otherwise similar nucleic acid comprising the endogenous UTR(s) associated with the heterologous object sequence or a minimal 5' UTR and a minimal 3' UTR.

8. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a second strand of a site in a target

genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain;

wherein:

(i) the polypeptide comprises a heterologous targeting domain (e.g., in the DBD or the endonuclease domain) that binds specifically to a sequence comprised in the target site; and/or

(ii) the template RNA comprises a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target site.

9. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide, (ii) a heterologous object sequence, and (iii) a ribozyme that is heterologous to (a)(i), (a)(ii), (b)(i), or a combination thereof.

10. The system of embodiment 9, wherein the ribozyme is heterologous to (b)(i).

11. The system of embodiment 9 or 10, wherein the template RNA comprises (iv) a second ribozyme, e.g., that is endogenous to (a)(i), (a)(ii), (b)(i), or a combination thereof, e.g., wherein the second ribozyme is endogenous to (b)(i).

12. The system of embodiment 9 or 10, wherein the heterologous ribozyme replaced a ribozyme endogenous to (a)(i), (a)(ii), (b)(i), or a combination thereof, e.g., wherein the second ribozyme is endogenous to (b)(i).

13. A system for modifying DNA comprising:

optionally (a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide, (ii) a heterologous object sequence, (iii) a 5' UTR capable of being

cleaved into a fragment and a cleaved template RNA, wherein the 5' UTR is optionally the sequence that binds the polypeptide,

wherein the 5' UTR comprises one or more mutations (e.g., relative to a wildtype 5' UTR, e.g., described herein) which increase the affinity of the fragment for the cleaved template RNA, e.g., such that the fragment hybridizes to the cleaved template RNA (e.g., the 5' UTR of the cleaved template RNA), e.g., under stringent conditions, e.g., wherein the stringent conditions comprise hybridization in 4x sodium chloride/sodium citrate (SSC), at about 65°C, followed by a wash in 1xSSC, at about 65°C.

14. The system of embodiment 13, wherein the template RNA, e.g., the 5' UTR, comprises a ribozyme which cleaves the template RNA (e.g., in the 5' UTR).

15. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein (a), (b), or (a) and (b) comprise an intron that increases the expression of the polypeptide, the heterologous object sequence (e.g., a coding sequence situated in the heterologous object sequence), or both.

16. A method of modifying a target DNA strand in a cell, tissue or subject, comprising administering a system to a cell, wherein the system comprises:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the system reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand, and

wherein the cell has decreased Rad51 repair pathway activity, decreased expression of Rad51 or a component of the Rad51 repair pathway, or does not comprise a functional Rad51

repair pathway, e.g., does not comprise a functional Rad51 gene, e.g., comprises a mutation (e.g., deletion) inactivating one or both copies of the Rad51 gene or another gene in the Rad51 repair pathway.

17. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the heterologous object sequence comprises a sequence, e.g., a gene or fragment thereof, of any of Tables 10A-10D or 11A-11G.

18. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein the polypeptide is modified for enhanced activity or altered specificity; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

19. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence, wherein the template RNA comprises one or more chemical modification selected from dihydrouridine, inosine, 7-methylguanosine, 5-methylcytidine (5mC), 5′ Phosphate ribothymidine, 2′-O-methyl ribothymidine, 2′-O-ethyl ribothymidine, 2′-fluoro ribothymidine, C-5 propynyl-deoxycytidine (pdC), C-5 propynyl-deoxyuridine (pdU), C-5 propynyl-cytidine (pC), C-5 propynyl-uridine (pU), 5-methyl cytidine, 5-methyl uridine, 5-methyl deoxycytidine, 5-methyl deoxyuridine methoxy, 2,6-diaminopurine, 5′-Dimethoxytrityl-N4-ethyl-2′-deoxycytidine, C-5 propynyl-f-cytidine (pfC), C-5 propynyl-f-uridine (pfU), 5-methyl f-cytidine, 5-methyl f-uridine, C-5

propynyl-m-cytidine (pmC), C-5 propynyl-f-uridine (pmU), 5-methyl m-cytidine, 5-methyl m-uridine, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), or 5-methoxyuridine (5-MO-U).

20. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a target DNA binding domain, (ii) a reverse transcriptase domain, optionally (iii) an endonuclease domain, wherein the polypeptide comprises a heterologous linker replacing a portion of (i), (ii), or (iii), or replacing an endogenous linker connecting two of (i), (ii), or (iii); and

optionally (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

21. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide, (ii) a heterologous object sequence, (iii) a first homology domain having at least 5 or at least 10 bases of 100% identity to a target DNA strand, at the 5' end of the template RNA, and (iv) a second homology domain having at least 5 or at least 10 bases of 100% identity to a target DNA strand, at the 3' end of the template RNA.

22. The system of any preceding embodiments, wherein the polypeptide comprises a mutation inactivating and/or deleting a nucleolar localization signal.

23. The system of embodiment 22, wherein activity of the nucleolar localization signal is reduced by at least 50%, 60%, 70%, 80%, 90%, 95%, or 99%.

24. The system of either of embodiments 22 or 23, wherein the polypeptide comprises a nuclear localization signal (NLS), e.g., an endogenous NLS or an exogenous NLS.

25. The system of any preceding embodiments, wherein the polypeptide of (a) comprises a target DNA binding domain (e.g., the endonuclease domain comprises a target DNA binding domain), e.g., a first target DNA binding domain, or (a) further comprises a target DNA binding domain, e.g., a first target binding domain.

26. The system of embodiment 25, wherein:

the polypeptide of (a) further comprises a second target DNA binding domain, e.g., a Zn finger domain, that is heterologous, e.g., to the first target DNA binding domain or to the endonuclease domain.

27. The system of embodiment 26, wherein the endonuclease domain comprises the second target DNA binding domain.

28. The system of embodiment 26 or 27, wherein the second target DNA binding domain affects the endonuclease activity of the polypeptide.

29. The system of any preceding embodiments, wherein the second target DNA binding domain affects DNA nicking activity of the polypeptide.

30. The system of any preceding embodiments, wherein the second target DNA binding domain binds a locus provided in Table E3.

31. The system of any preceding embodiments, wherein the locus in Table E3 has a genomic score of at least 6.

32. The system of any preceding embodiments, wherein the polypeptide of (a) binds to a smaller number of target DNA sequences than a similar polypeptide that comprises only the first target DNA binding domain or the second target DNA binding domain, e.g., wherein the presence of the second target DNA binding domain in the polypeptide with the first target DNA binding domain refines the target sequence specificity of the polypeptide relative to the

polypeptide target sequence specificity of the polypeptide comprising only the first target DNA binding domain.

33. The system of any preceding embodiments, wherein the second target DNA binding domain binds to a genomic DNA sequence that is less than 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 nucleotides away from a genomic sequence to which the first target DNA binding domain binds.

34. The system of any preceding embodiments, wherein the second target DNA binding domain binds to a genomic DNA sequence that is 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, or 90-100 nucleotides away from a genomic sequence to which the first target DNA binding domain binds.

35. The system of any preceding embodiments, wherein the first or second target DNA binding domain comprises a CRISPR/Cas protein, a TAL Effector domain, a Zn finger domain, or a meganuclease domain.

36. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a CRISPR/Cas protein and the second target DNA binding domain comprises a TAL effector domain.

37. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a CRISPR/Cas protein and the second target DNA binding domain comprises a Zn finger domain.

38. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a CRISPR/Cas protein and the second target DNA binding domain comprises a CRISPR/Cas protein.

39. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a CRISPR/Cas protein and the second target DNA binding domain comprises a meganuclease domain.

40. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a TAL effector domain and the second target DNA binding domain comprises a Zn finger domain.

41. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a TAL effector domain and the second target DNA binding domain comprises a TAL effector domain.

42. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a TAL effector domain and the second target DNA binding domain comprises a meganuclease domain.

43. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a Zn finger domain and the second target DNA binding domain comprises a Zn finger domain.

44. The system of any preceding embodiments, wherein the first target DNA binding domain comprises a Zn finger domain and the second target DNA binding domain comprises a meganuclease domain.

45. The system of any preceding embodiments, wherein the second DNA binding domain binds to a sequence in a genomic safe harbor (GSH) site or a genomic Natural Harbor™ site.

46. The system of any preceding embodiments, wherein the system is capable of cutting the first strand of the target DNA and the second strand of the target DNA, e.g., wherein the cuts are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or 200 nucleotides away from one another (and optionally no more than 500, 400, 300, 200, or 100 nucleotides away from one another).

47. The system of any preceding embodiments, wherein the system is capable of cutting the first strand of the target DNA at least twice (e.g., twice), e.g., wherein the cuts are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or 200 nucleotides away from one another (and optionally no more than 500, 400, 300, 200, or 100 nucleotides away from one another).

48. The system of any preceding embodiments, wherein the cuts are 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, or 400-500 nucleotides away from one another.

49. The system of any preceding embodiments, wherein the distance between the cuts is the same as the distance between cuts made by the endonuclease domain, e.g., the endonuclease domain of a naturally occurring retrotransposase.

50. The system of any preceding embodiments, wherein the two cuts are both made by the same endonuclease domain (e.g., a CRISPR/Cas protein, e.g., directed by a plurality of gRNAs, e.g., disposed in the template RNA).

51. The system of any preceding embodiments, wherein the polypeptide further comprises a second endonuclease domain.

52. The system of any preceding embodiments, wherein:

i) the first endonuclease domain (e.g., nickase) cuts the to-be-edited strand of the target DNA and the second endonuclease domain (e.g., nickase) cuts the non-edited strand of the target DNA, or

ii) the first endonuclease domain (e.g., nickase) makes one of the two cuts to the to-be-edited strand of the target DNA and the second endonuclease domain (e.g., nickase) makes the other cut to the to-be-edited strand of the target DNA.

53. The system of any preceding embodiments, wherein (a), (b), or (a) and (b) further comprises a 5' UTR and/or 3' UTR operably linked to the sequence encoding the polypeptide, the heterologous object sequence (e.g., a coding sequence contained in the heterologous object sequence), or both, wherein the 5' UTR and/or 3' UTR increase expression of the operably linked sequence(s).

54. The system of preceding embodiment, wherein the 5' UTR and/or 3' UTR:

increase the stability, e.g., half-life, of the template RNA, an RNA transcribed from (a), or both; and/or

increases the efficiency of translation of the heterologous object sequence, the polypeptide, or both.

55. The system of preceding embodiment, wherein the 5' UTR comprises a 5' UTR from complement factor 3 (C3) or a functional fragment or variant thereof.

56. The system of any preceding embodiments, wherein the 3' UTR comprises a 3' UTR from orosomucoid 1 (ORM1) or a functional fragment or variant thereof.

57. The system of any preceding embodiments, wherein

i) the 5' UTR increases the rate of translation, e.g., relative to an otherwise similar nucleic acid comprising the endogenous UTR(s) associated with the heterologous object sequence or a minimal 5' UTR and a minimal 3' UTR,

ii) the 3' UTR increases nucleic acid half-life, e.g., relative to an otherwise similar nucleic acid comprising the endogenous UTR(s) associated with the heterologous object sequence or a minimal 5' UTR and a minimal 3' UTR, or

iii) both i) and ii).

58. The system of any preceding embodiments, wherein the template RNA comprises a ribozyme that is heterologous to (a)(i), (a)(ii), (b)(i), or a combination thereof.

59. The system of any preceding embodiments, wherein the heterologous ribozyme replaced a ribozyme endogenous to (a)(i), (a)(ii), (b)(i), or a combination thereof.

60. The system of any preceding embodiments, wherein the template RNA comprises a second ribozyme, e.g., that is endogenous to (a)(i), (a)(ii), (b)(i), or a combination thereof.

61. The system of any preceding embodiments, wherein the heterologous ribozyme is situated in a 5' UTR or 3' UTR of the template RNA.

62. The system of any preceding embodiments, wherein the heterologous ribozyme is 5' of the heterologous object sequence or 3' of the heterologous object sequence.

63. The system of any preceding embodiments, wherein the heterologous ribozyme is capable of cleaving RNA comprising the ribozyme, e.g., 5' of the ribozyme, 3' of the ribozyme, or within the ribozyme.

64. The system of any preceding embodiments, wherein the heterologous ribozyme is 5' of the heterologous object sequence and cleaves 3' of the heterologous ribozyme, e.g., wherein the heterologous ribozyme is a synthetic or naturally occurring hammerhead ribozyme.

65. The system of any preceding embodiments, wherein the heterologous ribozyme is 3' of the heterologous object sequence and cleaves 5' of the heterologous ribozyme, e.g., wherein the heterologous ribozyme is chosen from an HDV family ribozyme or a hatchet ribozyme.

66. The system of any preceding embodiments, wherein the template RNA further comprises a ribozyme-hybridizing region, e.g., a template with altered targeting, such as through a homology arm, comprises a modified 5' UTR comprising the ribozyme-hybridizing region.

67. The system of any preceding embodiments, wherein a portion of the ribozyme hybridizes (e.g. via Watson-crick basepairing) to sequence 5' or 3' of the ribozyme.

68. The system of any preceding embodiments, wherein the ribozyme sequence is altered from its natural sequence by at least 1, 2, 3, 4, 5, 6, 8, 9, 10, 15, 20, 25 or more basepairs.

69. The system of any preceding embodiments, wherein the ribozyme sequence is altered from its natural sequence in order to hybridize to a homology arm that is 5' or 3' of the target ribozyme

70. The system of any preceding embodiments, wherein the system integrates a heterologous object sequence into a target genome with a greater efficiency than an otherwise similar system lacking the heterologous ribozyme, e.g., wherein at least 10%, 20%, 30%, 405, 50%, 60%, 70%, 80%, 90%, or 100% more cells show integration in the presence of the system comprising the heterologous ribozyme compared to the system lacking the heterologous ribozyme.

71. The system of any preceding embodiments, wherein the template RNA comprises a 5' UTR capable of being cleaved into a fragment and a cleaved template RNA.

72. The system of any preceding embodiments, wherein the template RNA comprises a ribozyme which cleaves the template RNA, e.g., in the 5' UTR.

73. The system of any preceding embodiments, wherein the 5' UTR comprises one or more mutations (e.g., relative to a wildtype 5' UTR described herein, e.g., in Tables 1 or 3, or from a protein domain listed in Table 2).

74. The system of any preceding embodiments, wherein the one or more mutations increase the affinity of the fragment for the cleaved template RNA, e.g., such that the fragment hybridizes to the cleaved template RNA (e.g., the 5' UTR of the cleaved template RNA) under stringent conditions, e.g., wherein the stringent conditions for hybridization includes hybridization in 4x sodium chloride/sodium citrate (SSC), at about 65°C, followed by a wash in 1xSSC, at about 65°C.

76. The system of any preceding embodiments, wherein (a), (b), or (a) and (b) comprise an intron that increases the expression of the polypeptide, the heterologous object sequence (e.g., a coding sequence situated in the heterologous object sequence), or both.

77. The system of any preceding embodiments, wherein the intron is operably linked (e.g., to be recognized by cellular splicing proteins) to the sequence encoding the polypeptide, the heterologous object sequence (e.g., a coding sequence situated in the heterologous object sequence), or both.

78. The system of any preceding embodiments, wherein the intron is situated in a 5' UTR (e.g., 5' of the heterologous object sequence).

79. The system of any preceding embodiments, wherein the intron is situated in a coding sequence of the heterologous object sequence.

80. The system of any preceding embodiments, wherein the intron is situated in the forward direction in relation to the coding sequence of the heterologous object sequence.

81. The system of any preceding embodiments, wherein the intron is situated in the reverse direction in relation to the coding sequence of the heterologous object sequence.

82. The system of any preceding embodiment, wherein the intron is spliced after transcription of the template RNA and before target primed reverse transcription into target, e.g., genomic, DNA.

83. The system of any preceding embodiments, wherein the intron is spliced after transcription of the heterologous object sequence after the heterologous object sequence is integrated in the target, e.g., genomic, DNA.

84. The system of any preceding embodiments, wherein the intron comprises a microRNA binding site.

85. The system of any of the preceding embodiments, wherein the enonuclease domain (e.g., an endonuclease domain of R2Tg or R2-1_ZA) recognizes a motif (e.g., GG or AAGG, TAAGGT, or TTAAGGTAGC), and the heterologous DNA binding domain recognizes a genomic DNA sequence, wherein the motif and the genomic DNA sequence are within 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-100, 100-150, 150-200, or 200-250 nucleotides of each other, optionally wherein the motif recognized by the endonuclease domain comprises 4, 5, 6, 7, 8, 9, or 10 consecutive nucleotides of TTAAGGTAGC, AAGGTAGCCAAA, or TAAGGTAGCCAAA, or wherein the motif recognized by the endonuclease domain comprises 2 or 3 consecutive nucleotides of AAGG.

86. The system of any preceding embodiments, wherein the motif is upstream of the genomic DNA sequence, e.g., the motif is about 30-80, 40-70, 50-60, or 55 nt upstream of the genomic DNA sequence.

87. The system of any preceding embodiments, wherein the motif is downstream of the genomic DNA sequence, e.g., the motif is about 10-30, 15-25, or 20 nt downtream of the genomic DNA sequence.

88. The system of any preceding embodiments, wherein the motif is in the same orientation as the genomic DNA sequence or in the reverse complement orientation as the genomic DNA sequence.

89. The system of any preceding embodiments, wherein the heterologous DNA binding domain (e.g., a zinc finger domain) is N-terminal or C-terminal of the endonuclease domain.

90. The system of any preceding embodiments, wherein a linker (e.g., a linker of Table 38) is disposed between the heterologous DNA binding domain and the endonuclease domain.

91. The system any of the preceding embodiments, wherein the system comprises one or more circular RNA molecules (circRNAs).

92. The system of any preceding embodiments, wherein the circRNA encodes the Gene Writer polypeptide.

93. The system of any preceding embodiments, wherein the circRNA comprises a template RNA.

94. The system of any preceding embodiments, wherein circRNA is delivered to a host cell.

95. The system of any of the preceding embodiments, wherein the circRNA is capable of being linearized, e.g., in a host cell, e.g., in the nucleus of the host cell.

95. The system of any of the preceding embodiments, wherein the circRNA comprises a cleavage site.

97. The system of any preceding embodiments, wherein the circRNA further comprises a second cleavage site.

98. The system of any preceding embodiments, wherein the cleavage site can be cleaved by a ribozyme, e.g., a ribozyme comprised in the circRNA (e.g., by autocleavage).

99. The system of any of the preceding embodiments, wherein the circRNA comprises a ribozyme sequence.

100. The system of any preceding embodiments, wherein the ribozyme sequence is capable of autocleavage, e.g., in a host cell, e.g., in the nucleus of the host cell.

101. The system of any preceding embodiments, wherein the ribozyme is an inducible ribozyme.

102. The system of any preceding embodiments, wherein the ribozyme is a protein-responsive ribozyme, e.g., a ribozyme responsive to a nuclear protein, e.g., a genome-interacting protein, e.g., an epigenetic modifier, e.g., EZH2.

103. The system of any preceding embodiments, wherein the ribozyme is a nucleic acid-responsive ribozyme.

104. The system of any preceding embodiments, wherein the catalytic activity (e.g., autocatalytic activity) of the ribozyme is activated in the presence of a target nucleic acid molecule (e.g., an RNA molecule, e.g., an mRNA, miRNA, ncRNA, lncRNA, tRNA, snRNA, or mtRNA).

105. The system of any preceding embodiments, wherein the ribozyme is responsive to a target protein (e.g., an MS2 coat protein).

106. The system of any preceding embodiments, wherein the target protein localized to the cytoplasm or localized to the nucleus (e.g., an epigenetic modifier or a transcription factor).

107. The system of any preceding embodiments, wherein the ribozyme comprises the ribozyme sequence of a B2 or ALU retrotransposon, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

108. The system of any preceding embodiments, wherein the ribozyme comprises the sequence of a tobacco ringspot virus hammerhead ribozyme, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

109. The system of any preceding embodiments, wherein the ribozyme comprises the sequence of a hepatitis delta virus (HDV) ribozyme, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

110. The system of any preceding embodiments, wherein the ribozyme is activated by a moiety expressed in a target cell or target tissue.

111. The system of any preceding embodiments, wherein the ribozyme is activated by a moiety expressed in a target subcellular compartment (e.g., a nucleus, nucleolus, cytoplasm, or mitochondria).

112. The system of any of the preceding embodiments, wherein the ribozyme is comprised in a circular RNA or a linear RNA.

113. A system comprising a first circular RNA encoding the polypeptide of a Gene Writing system; and

a second circular RNA comprising the template RNA of a Gene Writing system.

114. The system of any of the preceding embodiments, wherein the template RNA, e.g., the 5' UTR, comprises a ribozyme which cleaves the template RNA (e.g., in the 5' UTR).

115. The system of any of the preceding embodiments, wherein the template RNA comprises a ribozyme that is heterologous to (a)(i) (the a reverse transcriptase domain), (a)(ii) (the endonuclease domain), (b)(i) (a sequence of the template RNA that binds the polypeptide), or a combination thereof.

116. The system of any of the preceding embodiments, wherein the heterologous ribozyme is capable of cleaving RNA comprising the ribozyme, e.g., 5' of the ribozyme, 3' of the ribozyme, or within the ribozyme.

117. A lipid nanoparticle (LNP) comprising the system, polypeptide (or RNA encoding the same), nucleic acid molecule, or DNA encoding the system or polypeptide, of any preceding embodiment.

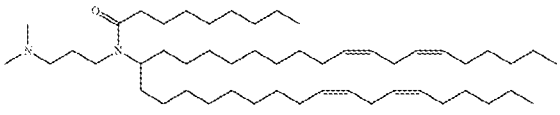
118. A system comprising a first lipid nanoparticle comprising the polypeptide (or DNA or RNA encoding the same) of a Gene Writing system (e.g., as described herein); and

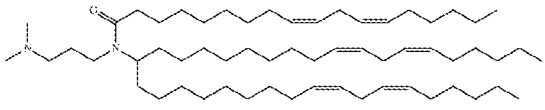
a second lipid nanoparticle comprising a nucleic acid molecule of a Gene Writing System (e.g., as described herein).

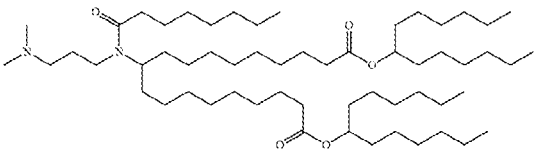
119. The system, kit, polypeptide, or reaction mixture of any preceding embodiments, wherein the system, nucleic acid molecule, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

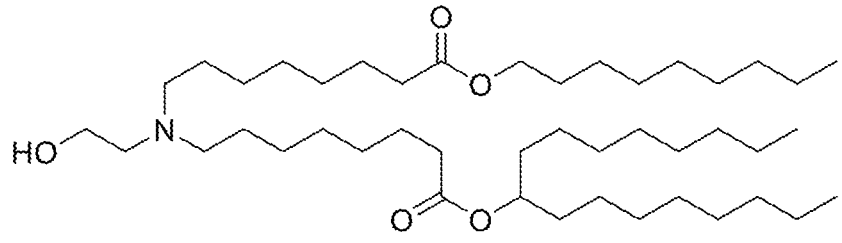
120. The LNP of any preceding embodiments, comprising a cationic lipid.

121. The LNP of any preceding embodiments wherein the cationic lipid having a following structure:

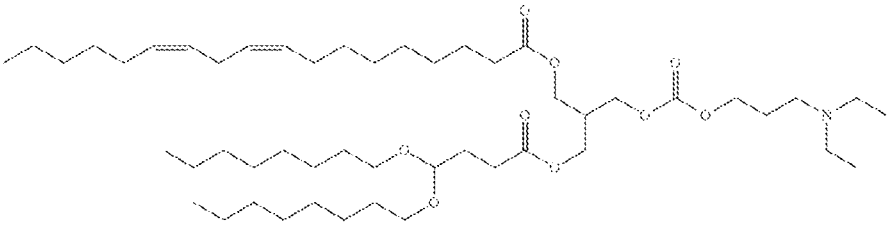

(i),


(ii),


(iii),



(vii), or



(ix).

122. The LNP of any any preceding embodiments, further comprising one or more neutral lipid, e.g., DSPC, DPPC, DMPC, DOPC, POPC, DOPE, SM, a steroid, e.g., cholesterol, and/or one or more polymer conjugated lipid, e.g., a pegylated lipid, e.g., PEG-DAG, PEG-PE, PEG-S-DAG, PEG-cer or a PEG dialkyoxypropylcarbamate.

123. The system, kit, or polypeptide, of any of the preceding embodiments, wherein the system, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

124. The system, kit, or polypeptide of embodiment M1, wherein the lipid nanoparticle (or a formulation comprising a plurality of the lipid nanoparticles) lacks reactive impurities (e.g., aldehydes), or comprises less than a preselected level of reactive impurities (e.g., aldehydes).

125. The system, kit, or polypeptide of embodiment M1, wherein the lipid nanoparticle (or a formulation comprising a plurality of the lipid nanoparticles) lacks aldehydes, or comprises less than a preselected level of aldehydes.

126. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle is comprised in a formulation comprising a plurality of the lipid nanoparticles.

127. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

128. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 3% total reactive impurity (e.g., aldehyde) content.

128. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

129. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagent comprising less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

130. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

131. The system, kit, or polypeptide of any any preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

132. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 3% total reactive impurity (e.g., aldehyde) content.

133. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

134. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

135. The system, kit, or polypeptide of any preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

136. The system, kit, or polypeptide of any preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

137. The system, kit, or polypeptide of any preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 3% total reactive impurity (e.g., aldehyde) content.

138. The system, kit, or polypeptide of any preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

139. The system, kit, or polypeptide of any preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

140. The system, kit, or polypeptide of any preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

141. The system, kit, or polypeptide of any preceding embodiments, wherein the total aldehyde content and/or quantity of any single reactive impurity (e.g., aldehyde) species is determined by liquid chromatography (LC), e.g., coupled with tandem mass spectrometry (MS/MS), e.g., according to the method described in Example 26.

142. The system, kit, or polypeptide of any preceding embodiments, wherein the total aldehyde content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleic acid molecule (e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents.

143. The system, kit, or polypeptide of any preceding embodiments, wherein the total aldehyde content and/or quantity of aldehyde species is determined by detecting one or more chemical modifications of a nucleotide or nucleoside (e.g., a ribonucleotide or ribonucleoside, e.g., comprised in or isolated from a nucleic acid molecule, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents, e.g., as described in Example 27.

144. The system, kit, or polypeptide of embodiment M21, wherein the chemical modifications of a nucleic acid molecule, nucleotide, or nucleoside are detected by determining the presence of one or more modified nucleotides or nucleosides, e.g., using LC-MS/MS analysis, e.g., as described in Example 27.

145. A method of modifying a target DNA strand in a cell, tissue or subject, comprising administering any preceding numbered system to the cell, tissue or subject, wherein the system reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand, and wherein the cell has decreased Rad51 repair pathway activity,

decreased expression of Rad51 or a component of the Rad51 repair pathway, or does not comprise a functional Rad51 repair pathway, e.g., does not comprise a functional Rad51 gene, e.g., comprises a mutation (e.g., deletion) inactivating one or both copies of the Rad51 gene or another gene in the Rad51 repair pathway.

146. A host cell (e.g., a mammalian cell, e.g., a human cell) comprising any preceding numbered system, wherein the host cell has decreased Rad51 repair pathway activity, decreased expression of Rad51 or a component of the Rad51 repair pathway, or does not comprise a functional Rad51 repair pathway, e.g., does not comprise a functional Rad51 gene, e.g., comprises a mutation (e.g., deletion) inactivating one or both copies of the Rad51 gene or another gene in the Rad51 repair pathway.

147. The system of any preceding embodiments, wherein the polypeptide binds a promoter region, a 5' UTR region, an exon, an intron, or a 3' UTR region of a sequence, e.g., a gene or fragment thereof, of any of Tables 10A-10D or 11A-11G .

148. The system of any preceding embodiments, wherein the polypeptide further comprises a heterologous linker replacing a portion of (i) a target DNA binding domain, (ii) a reverse transcriptase domain, optionally (iii) an endonuclease domain, or replacing an endogenous linker connecting two of (i), (ii), or (iii), wherein optionally the linker is a linker of Table 38.

149. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i).

150. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (ii).

151. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (iii).

152. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i) and (ii).

153. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i) and (iii).

154. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (ii) and (iii).

155. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (i) and (ii).

156. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (i) and (iii).

157. The system of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (ii) and (iii).

158. The system of any preceding embodiments, wherein the heterologous linker comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

159. The system of any preceding embodiments, wherein the heterologous linker comprises at least 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, or 500 amino acids.

160. The method of any of the preceding embodiments, wherein the tissue is liver, lung, skin, muscle tissue (e.g., skeletal muscle), eye or ocular tissue, blood, blood cells, immune cells, or central nervous system.

161, The method of any of the preceding embodiments, wherein the cell is a hematopoietic stem cell (HSC), a T-cell, a B cell, or a Natural Killer (NK) cell.

162. The method of any of the preceding embodiments, wherein the cell is a fibroblast.

163. The method of any of the preceding embodiments, wherein the cell is a primary cell.

164. The method of any of the preceding embodiments, where in the cell is not immortalized.

165. The system of any of the preceding embodiments, wherein (a) comprises RNA and (b) comprises RNA.

166. The system of any of the preceding embodiments, wherein (a) and (b) are part of the same nucleic acid.

167. The system of any preceding embodiments, wherein (a) and (b) are separate nucleic acids.

168. The system of any of the preceding embodiments, which comprises only RNA, or which comprises more RNA than DNA by an RNA:DNA ratio of at least 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, or 100:1.

169. The system of any preceding embodiments, wherein the heterologous object sequence comprises an open reading frame in a 5' to 3' orientation on the template RNA.

170. The system of any preceding embodiments, wherein the heterologous object sequence comprises an open reading frame in a 3' to 5' orientation on the template RNA.

171. The system of any of the preceding embodiments, wherein the sequence that binds the polypeptide is a 3' untranslated sequence.

172. The system of any preceding embodiments, wherein the template RNA further comprises a 5' untranslated sequence.

173. The system of any of the preceding embodiments, wherein the template RNA further comprises a promoter operably linked to the heterologous object sequence, e.g., the heterologous object sequence can, in some embodiment, comprise a promoter operably linked to a sequence, such as a protein coding sequence.

174. The system of any preceding embodiments, wherein the promoter is disposed between the 5' untranslated sequence and the heterologous object sequence.

175. The system of any preceding embodiments, wherein the promoter is disposed between the 3' untranslated sequence that binds the polypeptide and the heterologous object sequence.

176. The system of any any preceding embodiments, wherein the 5' untranslated sequence is a sequence of column 5 of Table 3, or a sequence having at least 80% identity thereto.

177. The system of any any preceding embodiments, wherein the 3' untranslated sequence is a sequence of column 6 of Table 3, or a sequence having at least 80% identity thereto.

178. The system of any of the preceding embodiments, wherein the heterologous object sequence comprises an enzyme, a membrane protein, a blood factor, an intracellular protein, an extracellular protein, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, a storage protein, an immune receptor protein, (e.g. a synthetic immune receptor protein such as a chimeric antigen receptor protein (CAR), a T cell receptor, a B cell receptor), or an antibody.

179. The system of any of the preceding embodiments, wherein the template RNA comprises at least 5 based or at least 10 bases of 100% identity to a target DNA strand, at the 5' end of the template RNA.

180. The system of any of the preceding embodiments, wherein the template RNA comprises at least 5 bases or at least 10 bases of 100% identity to a target DNA strand, at the 3' end of the template RNA.

181. A method of modifying a target DNA strand in a cell, tissue, or subject, comprising administering the system of any preceding embodiments to the cell, tissue, or subject, thereby modifying the target DNA strand.

182. The method of any preceding embodiments, which results in the addition of at least 5 base pairs of exogenous DNA sequence to the genome of the cell.

183. The method of any preceding embodiments, which results in the addition of at least 100 base pairs of exogenous DNA sequence to the genome of the cell.

184. The method of any preceding embodiments, which results in insertion of the heterologous object sequence into the target DNA at an average copy number of at least 0.01, 0.05, or 0.5 copies per genome.

185. The method of any preceding embodiments, which results in about 50-100% of insertions of the heterologous object sequence into the target DNA being non-truncated.

186. The method of any preceding embodiments, wherein the nucleic acid of (a) is not integrated into the genome of the cell.

187. The method of any preceding embodiments, wherein the template RNA comprises at least 5 or at least 10 bases of 100% identity to the target DNA strand, at the 5' end of the template RNA.

188. The method of any of any preceding embodiments, wherein the template RNA comprises at least 5 or at least 10 bases of 100% identity to the target DNA strand, at the 3' end of the template RNA.

189. The system or method of any preceding embodiments, wherein the heterologous object sequence encodes a therapeutic polypeptide or that encodes a mammalian (e.g., human) polypeptide, or a fragment or variant thereof.

190. The system or method of any preceding embodiments, wherein one or more of:

i. the heterologous object sequence encodes a protein, e.g. an enzyme (e.g., a lysosomal enzyme) or a blood factor (e.g., Factor I, II, V, VII, X, XI, XII or XIII);

ii. the heterologous object sequence comprises a tissue specific promoter or enhancer;

iii. the heterologous object sequence encodes a polypeptide of greater than 250, 300, 400, 500, or 1,000 amino acids, and optionally up to 7,500 amino acids;

iv. the heterologous object sequence encodes a fragment of a mammalian gene but does not encode the full mammalian gene, e.g., encodes one or more exons but does not encode a full-length protein;

v. the heterologous object sequence encodes one or more introns;

vi. the heterologous object sequence is other than a GFP, e.g., is other than a fluorescent protein or is other than a reporter protein; or

vii. the heterologous object sequence is other than a T cell chimeric antigen receptor.

191. The system or method of any preceding embodiments, wherein one or both of the reverse transcriptase domain or endonuclease domain are derived from an avian retrotransposase, e.g., have a sequence of Table 1 or 3 or at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

192. The system or method of any preceding embodiments, wherein the polypeptide has an activity at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

193. The system or method of any preceding embodiments, wherein the polypeptide is derived from an avian retrotransposase, e.g., an avian retrotransposase of column 8 of Table 3, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

194. The system or method of any preceding embodiments, wherein the avian retrotransposase is a retrotransposase from *Taeniopygia guttata*, *Geospiza fortis*, *Zonotrichia albicollis*, or *Tinamus guttatus*, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

195. The system or method of any preceding embodiments, wherein the polypeptide is derived from a retrotransposase of column 8 of Table 3, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

196. The system of any of the preceding embodiments, wherein the template RNA comprises a sequence of Table 3 (e.g., one or both of a 5' untranslated region of column 6 of Table 3 and a 3' untranslated region of column 7 of Table 3), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

197. The system or method of any preceding embodiments, wherein one or more of:

i. the nucleic acid encoding the polypeptide and the template RNA or a nucleic acid encoding the template RNA are separate nucleic acids;

ii. the template RNA does not encode an active reverse transcriptase, e.g., comprises an inactivated mutant reverse transcriptase, e.g., as described in Examples 1-2, or does not comprise a reverse transcriptase sequence; or

iii. the template RNA does not encode an active endonuclease, e.g., comprises an inactivated endonuclease or does not comprise an endonuclease; or

iv. the template RNA comprises one or more chemical modifications.

198. The system or method of any preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) further comprises a promoter operably linked to the heterologous object sequence,

wherein the promoter is disposed between the 5' untranslated sequence that binds the polypeptide and the heterologous sequence, or

wherein the promoter is disposed between the 3' untranslated sequence that binds the polypeptide and the heterologous sequence.

199. The system or method of any preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) further comprises a 5' untranslated sequence that binds the polypeptide and a 3' untranslated sequence that binds the polypeptide, and

wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 5' to 3' orientation on the template RNA; or

wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 3' to 5' orientation on the template RNA.

200. The system or method of any preceding embodiments, wherein at least one of the reverse transcriptase domain, endonuclease domain, or target DNA binding domain are heterologous.

201. The system or method of any preceding embodiments, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a purinic/apyrimidinic endonuclease (APE)-type non-LTR retrotransposon and a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of an APE-type non-LTR retrotransposon.

202. The system or method of any preceding embodiments, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a restriction enzyme-like endonuclease (RLE)-type non-LTR retrotransposon and a sequence at least 80% identical (e.g., at least 85%, 90%, 95%,

97%, 98%, 99%, 100% identical) to an endonuclease domain of a RLE-type non-LTR retrotransposon.

203. The system or method of any preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, wherein the RT domain further comprises a number of substitutions relative to the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

204. The system or method of any preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

205. The system or method of any preceding embodiments, wherein the template RNA comprises a promoter operably linked to the heterologous object sequence.

206. The system or method of any of the preceding embodiments, wherein the polypeptide further comprises (iii) a DNA-binding domain.

207. The system or method of any of embodiments 140-144, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to the sequence of SEQ ID NO: 1016.

208. The system or method of any of the preceding embodiments, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a sequence in column 8 of Table 3.

209. The system or method of any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide and the template RNA or the nucleic acid encoding the template RNA are covalently linked, e.g., are part of a fusion nucleic acid.

210. The system or method of any preceding embodiments, wherein the fusion nucleic acid comprises RNA.

211. The system or method of any preceding embodiments, wherein the fusion nucleic acid comprises DNA.

212. The system or method of any of the preceding embodiments, wherein (b) comprises template RNA.

213. The system or method of any preceding embodiments, wherein the template RNA further comprises a nuclear localization signal.

214. The system or method of any preceding embodiments, wherein the RNA of (a) does not comprise a nuclear localization signal.

215. The system or method of any of the preceding embodiments, wherein the polypeptide further comprises a nuclear localization signal and/or a nucleolar localization signal.

216. The system or method of any of the preceding embodiments, wherein (a) comprises an RNA that encodes: (i) the polypeptide and (ii) a nuclear localization signal and/or a nucleolar localization signal.

217. The system or method of any of the preceding embodiments, wherein the RNA comprises a pseudoknot sequence, e.g., 5' of the heterologous object sequence.

218. The system or method of any preceding embodiments, wherein the RNA comprises a stem-loop sequence or a helix, 5' of the pseudoknot sequence.

219. The system or method of any preceding embodiments, wherein the RNA comprises one or more (e.g., 2, 3, or more) stem-loop sequences or helices 3' of the pseudoknot sequence, e.g. 3' of the pseudoknot sequence and 5' of the heterologous object sequence.

220. The system or method of any preceding embodiments, wherein the template RNA comprising the pseudoknot has catalytic activity, e.g., RNA-cleaving activity, e.g, cis-RNA-cleaving activity.

221. The system or method of any of the preceding embodiments, wherein the RNA comprises at least one stem-loop sequence or helix, e.g., 3' of the heterologous object sequence, e.g. 1, 2, 3, 4, 5 or more stem-loop sequences, hairpins or helices sequences.

222. Any above-numbered system or method, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a sequence of a polypeptide listed in Table 1-3, or a reverse transcriptase domain or endonuclease domain thereof.

223. Any above-numbered system or method, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a sequence of a polypeptide listed in any of Tables 1-3 or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

224. Any above-numbered system or method, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to the amino acid sequence of column 8 of Table 3, or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

225. Any above-numbered system or method, wherein the template RNA comprises a sequence of Table 3 (e.g., one or both of a 5' untranslated region of column 6 of Table 3 and a 3' untranslated region of column 7 of Table 3), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

226. The system or method of any preceding embodiments, wherein the template RNA comprises a sequence of about 100-125 bp from a 3' untranslated region of column 7 of Table 3, e.g., wherein the sequence comprises nucleotides 1-100, 101-200, or 201-325 of the 3' untranslated region of column 7 of Table 3, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

227. Any above-numbered system or method, wherein (a) comprises RNA and (b) comprises RNA.

228. Any above-numbered system or method, wherein (a), (b), or (a) and (b) do not comprise DNA, or do not comprise more than 10%, 5%, 4%, 3%, 2%, or 1% DNA by mass or by molar amount.

229. Any above-numbered system, which is capable of modifying DNA by insertion of the heterologous object sequence without an intervening DNA-dependent RNA polymerization of (b).

230. Any above-numbered system, which is capable of modifying DNA by target primed reverse transcription.

231. Any above-numbered system, which is capable of modifying DNA by insertion of a heterologous object sequence in the presence of an inhibitor of a DNA repair pathway (e.g., SCR7, a PARP inhibitor), or in a cell line deficient for a DNA repair pathway (e.g., a cell line deficient for the nucleotide excision repair pathway or the homology-directed repair pathway).

232. Any above-numbered system, which does not cause formation of a detectable level of double stranded breaks in a target cell.

233. Any above-numbered system, which is capable of modifying DNA using reverse transcriptase activity, and optionally in the absence of homologous recombination activity.

234. Any above-numbered system, wherein the template RNA has been treated to reduce secondary structure, e.g., was heated, e.g., to a temperature that reduces secondary structure, e.g., to at least 70, 75, 80, 85, 90, or 95°C.

235. The system of any preceding embodiments, wherein the template RNA was subsequently cooled, e.g., to a temperature that allows for secondary structure, e.g, to less than or equal to 30, 25, or 20°C.

236. A host cell (e.g., a mammalian cell, e.g., a human cell) comprising any preceding numbered system.

237. The method of any preceding embodiments, wherein the cell, tissue or subject is a mammalian (e.g., human) cell, tissue or subject.

238. The method of any of the preceding embodiments, wherein the cell is a fibroblast.

239. The method of any of the preceding embodiments, wherein the cell is a primary cell.

240. The method of any of the preceding embodiments, where in the cell is not immortalized.

241. A method of modifying the genome of a mammalian cell, comprising contacting the cell with the system of any preceding embodiments.

242. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with the system of any preceding embodiments.

243. A method of adding at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA to the genome of a mammalian cell, without delivery of DNA to the cell, comprising contacting the cell with a system of any preceding embodiments.

244. A method of adding at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA to the genome of a mammalian cell, comprising contacting the cell with a system of any preceding embodiments,

wherein the method does not comprise contacting the mammalian cell with DNA, or

wherein the method comprises contacting the mammalian cell with a composition comprising less than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

245. A method of adding at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA to the genome of a mammalian cell, comprising contacting the cell with a system of any preceding embodiments, wherein the method delivers only RNA to the mammalian cell.

246. A method of adding at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA to the genome of a mammalian cell, comprising contacting the cell with a system of any preceding embodiments, wherein the method delivers RNA and protein to the mammalian cell.

247. The method of any preceding embodiments, wherein the template RNA serves as the template for insertion of the exogenous DNA.

248. The method of any preceding embodiments, which does not comprise DNA-dependent RNA polymerization of exogenous DNA.

249. The method of any preceding embodiments, which results in the addition of at least 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA to the genome of the cell, e.g., the mammalian cell.

250. A method of modifying the genome of a human cell, comprising contacting the cell with a system of any preceding embodiments,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq, e.g., as described in Example 14 of PCT/US2019/048607, incorporated herein by reference.

251. A method of adding an exogenous coding region to the genome of a cell (e.g., a mammalian cell), comprising contacting the cell with a system of any preceding embodiments, wherein the template RNA comprises the non-coding strand of the exogenous coding region, wherein optionally the template RNA does not comprise a coding strand of the exogenous coding region, wherein optionally the delivery comprises non-viral delivery.

252. A method of expressing a polypeptide in a cell (e.g., a mammalian cell), comprising contacting the cell with a system of any preceding embodiments, wherein the template RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide, wherein optionally the template RNA does not comprise a coding strand encoding the polypeptide, wherein optionally the delivery comprises non-viral delivery.

253. The method of any preceding embodiments, wherein the sequence that is inserted into the mammalian genome is a sequence that is exogenous to the mammalian genome.

254. The method of any preceding embodiments, wherein the system operates independently of a DNA template.

255. The method of any preceding embodiments, wherein the cell is part of a tissue.

256. The method of any preceding embodiments, wherein the mammalian cell is euploid, is not immortalized, is part of an organism, is a primary cell, is non-dividing, is a hepatocyte, or is from a subject having a genetic disease.

257. The method of any preceding embodiments, wherein the contacting comprises contacting the cell with a plasmid, virus, viral-like particle, virosome, liposome, vesicle, exosome, fusosome, or lipid nanoparticle.

258. The method of any preceding embodiments, wherein the contacting comprises using non-viral delivery.

259. The method of any preceding embodiments, which comprises comprising contacting the cell with the template RNA (or DNA encoding the template RNA), wherein the template RNA comprises the non-coding strand of an exogenous coding region, wherein optionally the template RNA does not comprise a coding strand of the exogenous coding region, wherein optionally the delivery comprises non-viral delivery, thereby adding the exogenous coding region to the genome of the cell.

260. The method of any preceding embodiments, which comprises contacting the cell with the template RNA (or DNA encoding the template RNA), wherein the template RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide, wherein optionally the template RNA does not comprise a coding strand encoding the polypeptide, wherein optionally the delivery comprises non-viral delivery, thereby expressing the polypeptide in the cell.

261. The method of any preceding embodiments, wherein the contacting comprises administering (a) and (b) to a subject, e.g., intravenously.

262. The method of any preceding embodiments , wherein the contacting comprises administering a dose of (a) and (b) to a subject at least twice.

263. The method of any preceding embodiments, wherein the polypeptide reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand.

264. The method of any preceding embodiments, wherein (a) and (b) are administered separately.

265. The method of any preceding embodiments, wherein (a) and (b) are administered together.

266. The method of any any preceding embodiments, wherein the nucleic acid of (a) is not integrated into the genome of the host cell.

267. Any preceding numbered method, wherein the sequence that binds the polypeptide has one or more of the following characteristics:

(a) is at the 3' end of the template RNA;

(b) is at the 5' end of the template RNA;

(b) is a non-coding sequence;

(c) is a structured RNA; or

(d) forms at least 1 hairpin loop structures.

268. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 20 nucleotides of at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a target DNA strand.

269. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 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 nucleotides of at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a target DNA strand.

270. Any preceding numbered method, wherein the sequence comprising at least 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 nucleotides, or about: 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100, or 2-100 nucleotides, of at least 80% identity to a target DNA strand is at the 3' end of the template RNA.

271. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 100 nucleotides of at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a target DNA strand, e.g., at the 3' end of the template RNA.

272. The method of any preceding embodiments, wherein the site in the target DNA strand to which the sequence comprises at least 80% identity is proximal to (e.g., within about: 0-10, 10-20, 20-30, 30-50, or 50-100 nucleotides of) a target site on the target DNA strand that is recognized (e.g., bound and/or cleaved) by the polypeptide comprising the endonuclease.

273. Any preceding numbered method, wherein the sequence comprising at least 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 nucleotides, or about: 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100, or 2-100 nucleotides, of at least 80% identity to a target DNA strand is at the 3' end of the template RNA;

optionally wherein the site in the target DNA strand to which the sequence comprises at least 80% identity is proximal to (e.g., within about: 0-10, 10-20, or 20-30 nucleotides of) a target site on the target DNA strand that is recognized (e.g., bound and/or cleaved) by the polypeptide comprising the endonuclease.

274. The method of any preceding embodiments, wherein the target site is the site in the human genome that has the closest identity to a native target site of the polypeptide comprising the endonuclease, e.g., wherein the target site in the human genome has at least about: 16, 17, 18, 19, or 20 nucleotides identical to the native target site.

275. Any preceding numbered method, wherein the template RNA has at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand.

276. Any preceding numbered method, wherein the at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand are at the 3' end of the template RNA.

277. Any preceding numbered method, wherein the at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand are at the 5' end of the template RNA.

278. Any preceding numbered method, wherein the template RNA comprises at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand at the 5' end of the template RNA and at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand at the 3' end of the template RNA.

279. Any preceding numbered method, wherein the heterologous object sequence is between 50-50,000 base pairs (e.g., between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp).

280. Any preceding numbered method, wherein the heterologous object sequence is at least 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, or 700 bp.

281. Any preceding numbered method, wherein the heterologous object sequence is at least 715, 750, 800, 950, 1,000, 2,000, 3,000, or 4,000 bp.

282. Any preceding numbered method, wherein the heterologous object sequence is less than 5,000, 10,000, 15,000, 20,000, 30,000, or 40,000 bp.

283. Any preceding numbered method, wherein the heterologous object sequence is less than 700, 600, 500, 400, 300, 200, 150, or 100 bp.

284. Any preceding numbered method, wherein the heterologous object sequence comprises:

(a) an open reading frame, e.g., a sequence encoding a polypeptide, e.g., an enzyme (e.g., a lysosomal enzyme), a membrane protein, a blood factor, an exon, an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein), an extracellular protein, a structural protein, a signaling protein, a regulatory

protein, a transport protein, a sensory protein, a motor protein, a defense protein, or a storage protein;

(b) a non-coding and/or regulatory sequence, e.g., a sequence that binds a transcriptional modulator, e.g., a promoter, an enhancer, an insulator;

(c) a splice acceptor site;

(d) a polyA site;

(e) an epigenetic modification site; or

(f) a gene expression unit.

285. Any preceding numbered method, wherein the target DNA is a genomic safe harbor (GSH) site.

286. Any preceding numbered method, wherein the target DNA is a genomic Natural Harbor™ site.

287. Any preceding numbered method, which results in insertion of the heterologous object sequence into the a target site in the genome at an average copy number of at least 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, or 5 copies per genome.

288. Any preceding numbered method, which results in about 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90%, of integrants into a target site in the genome being non-truncated, as measured by an assay described herein, e.g., an assay of Example 6.

289. Any preceding numbered method, which results in insertion of the heterologous object sequence only at one target site in the genome of the cell.

290. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target site in a cell, wherein the insertered heterologous sequence comprises less than 10%, 5%, 2%, 1%, 0.5%, 0.2%, or 0.1% mutations (e.g., SNPs or one or more deletions, e.g., truncations or internal deletions) relative to the heterologous sequence prior to insertion,

e.g., as measured by an assay of Example 12 of PCT/US2019/048607, incorporated herein by reference.

291. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target site in a plurality of cells, wherein less than 10%, 5%, 2%, or 1% of copies of the inserted heterologous sequence comprise a mutation (e.g., a SNP or a deletion, e.g., a truncation or an internal deletion), e.g., as measured by an assay of Example 12 of PCT/US2019/048607, incorporated herein by reference.

292. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target cell genome, and wherein the target cell does not show upregulation of p53, or shows upregulation of p53 by less than 10%, 5%, 2%, or 1%, e.g., wherein upregulation of p53 is measured by p53 protein level, e.g., according to the method described in Example 30 of PCT/US2019/048607, incorporated herein by reference, or by the level of p53 phosphorylated at Ser15 and Ser20.

293. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target cell genome, and wherein the target cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq, e.g., as described in Example 14 of PCT/US2019/048607, incorporated herein by reference.

294. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site (e.g., at a copy number of 1 insertion or more than one insertion) in about 1-80% of cells in a population of cells contacted with the system, e.g., about: 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, or 70-80% of cells, e.g., as measured using single cell ddPCR, e.g., as described in Example 17 of PCT/US2019/048607, incorporated herein by reference.

295. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site at a copy number of 1 insertion in about 1-80% of cells in a population of cells contacted with the system, e.g., about: 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, or 70-80% of cells, e.g., as measured using colony isolation and ddPCR, e.g., as described in Example 18 of PCT/US2019/048607, incorporated herein by reference.

296. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site (on-target insertions) at a higher rate that insertion into a non-target site (off-target insertions) in a population of cells, wherein the ratio of on-target insertions to off-target insertions is greater than 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1. 90:1, 100:1, 200:1, 500:1, or 1,000:1, e.g., using an assay of Example 11 of PCT/US2019/048607, incorporated herein by reference.

297. Any above-numbered method, results in insertion of a heterologous object sequence in the presence of an inhibitor of a DNA repair pathway (e.g., SCR7, a PARP inhibitor), or in a cell line deficient for a DNA repair pathway (e.g., a cell line deficient for the nucleotide excision repair pathway or the homology-directed repair pathway).

298. Any preceding numbered system, formulated as a pharmaceutical composition.

299. Any preceding numbered system, disposed in a pharmaceutically acceptable carrier (e.g., a vesicle, a liposome, a natural or synthetic lipid bilayer, a lipid nanoparticle, an exosome).

300. A method of making a system for modifying DNA (e.g., as described herein), the method comprising:

(a) providing a template nucleic acid (e.g., a template RNA or DNA) comprising a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target DNA molecule, and/or

(b) providing a polypeptide of the system (e.g., comprising a DNA-binding domain (DBD) and/or an endonuclease domain) comprising a heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule.

301. The method of any preceding embodiments, wherein:

(a) comprises introducing into the template nucleic acid (e.g., a template RNA or DNA) a heterologous homology sequence having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to the sequence comprised in a target DNA molecule, and/or

(b) comprises introducing into the polypeptide of the system (e.g., comprising a DNA-binding domain (DBD) and/or an endonuclease domain) the heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule.

302. The method of any preceding embodiments, wherein the introducing of (a) comprises inserting the homology sequence into the template nucleic acid.

303. The method of any preceding embodiments, wherein the introducing of (a) comprises replacing a segment of the template nucleic acid with the homology sequence.

304. The method of any preceding embodiments, wherein the introducing of (a) comprises mutating one or more nucleotides (e.g., at least 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, or 100 nucleotides) of the template nucleic acid, thereby producing a segment of the template nucleic acid having the sequence of the homology sequence.

305. The method of any preceding embodiments, wherein the introducing of (b) comprises inserting the amino acid sequence of the targeting domain into the amino acid sequence of the polypeptide.

306. The method of any preceding embodiments, wherein the introducing of (b) comprises inserting a nucleic acid sequence encoding the targeting domain into a coding sequence of the polypeptide comprised in a nucleic acid molecule.

307. The method of any preceding embodiments, wherein the introducing of (b) comprises replacing at least a portion of the polypeptide with the targeting domain.

308. The method of any preceding embodiments, wherein the introducing of (a) comprises mutating one or more amino acids (e.g., at least 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, or more amino acids) of the polypeptide.

309. The method of any preceding embodiments, wherein the motif recognized by the endonuclease domain (e.g., at least 2, 4, 6, 8, 10, 20, 30, 40, or at least 50 nt, or no more than 50, 40, 30, 20, 10, 8, 6, 4, or 2) or less than 3 less than Gene Write polypeptide, is used as a seed for retargeting the Gene Writing system, wherein the DNA binding domain is modified such that the binding of the Gene Writer polypeptide to the new target site results in the proper positioning of the endonuclease domain to the core motif to enable endonuclease activity , optionally wherein the motif recognized by the endonuclease domain comprises 4, 5, 6, 7, 8, 9, or 10 consecutive nucleotides of TTAAGGTAGC, AAGGTAGCCAAA, or TAAGGTAGCCAAA, or wherein the motif recognized by the endonuclease domain comprises 2 or 3, or 4 consecutive nucleotides of AAGG.

310. The method of any preceding embodiments, wherein AAGG sequence in the genome is used as a seed for retargeting the Gene Writing system, wherein the DNA binding domain is modified such that the binding of the Gene Writer polypeptide to the new target site results in the proper positioning of the endonuclease domain to the AAGG motif to enable endonuclease activity.

311. A method for modifying a target site in genomic DNA in a cell, the method comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the target site (e.g., a second strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain,

wherein:

(i) the polypeptide comprises a heterologous targeting domain (e.g., in the DBD or the endonuclease domain) that binds specifically to a sequence comprised in or adjacent to the target site of the genomic DNA; and/or

(ii) the template RNA comprises a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in or adjacent to the target site of the genomic DNA;

thereby modifying the target site in genomic DNA in a cell.

312. A method of making a system for modifying the genome of a mammalian cell, comprising:

a) providing a template RNA as described in any of the preceding embodiments, e.g., wherein the template RNA comprises (i) a sequence that binds a polypeptide comprising a reverse transcriptase domain and an endonuclease domain, and (ii) a heterologous object sequence; and

b) treating the template RNA to reduce secondary structure, e.g., heating the template RNA, e.g., to at least 70, 75, 80, 85, 90, or 95°C, and

c) subsequently cooling the template RNA, e.g., to a temperature that allows for secondary structure, e.g, to less than or equal to 30, 25, or 20°C.

313. The method of any preceding embodiments, which further comprises contacting the template RNA with a polypeptide that comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, or with a nucleic acid (e.g., RNA) encoding the polypeptide.

314. The method of any preceding embodiments, which further comprises contacting the template RNA with a cell.

315. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a therapeutic polypeptide.

316. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a mammalian (e.g., human) polypeptide, or a fragment or variant thereof.

317. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes an enzyme (e.g., a lysosomal enzyme), a blood factor (e.g., Factor I, II, V, VII, X, XI, XII or XIII), a membrane protein, an exon, an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein), an extracellular protein, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, or a storage protein.

318. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises a tissue specific promoter or enhancer.

319. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a polypeptide of greater than 250, 300, 400, 500, or 1,000 amino acids, and optionally up to 1300 amino acids.

320. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a fragment of a mammalian gene but does not encode the full mammalian gene, e.g., encodes one or more exons but does not encode a full-length protein.

321. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes one or more introns.

322. The system or method of any of the preceding embodiments, wherein the heterologous object sequence is other than a GFP, e.g., is other than a fluorescent protein or is other than a reporter protein.

323. The system or method of any of the preceding embodiments, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) are derived from an avian retrotransposase, e.g., have a sequence of Table 1 or 3, or of a protein domain listed in Table 2, or at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

324. The system or method of any preceding embodiments, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) are derived from an avian retrotransposase, and wherein one or both of (i) or (ii) further comprises a number of substitutions relative to the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

325. The system or method of any of the preceding embodiments, wherein the polypeptide has an activity at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

326. The system or method of any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide and the template RNA or a nucleic acid encoding the template RNA are separate nucleic acids.

327. The system or method of any of the preceding embodiments, wherein the template RNA does not encode an active reverse transcriptase, e.g., comprises an inactivated mutant reverse transcriptase, e.g., as described in Example 1 or 2 of PCT/US2019/048607, incorporated herein by reference, or does not comprise a reverse transcriptase sequence.

328. The system or method of any of the preceding embodiments, wherein the template RNA comprises one or more chemical modifications.

329. The system or method of any of the preceding embodiments, wherein the heterologous object sequence is disposed between the promoter and the sequence that binds the polypeptide.

330. The system or method of any of the preceding embodiments, wherein the promoter is disposed between the heterologous object sequence and the sequence that binds the polypeptide.

331. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 5' to 3' orientation on the template RNA.

332. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 3' to 5' orientation on the template RNA.

333. The system or method of any of the preceding embodiments, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein at least one of (a) or (b) is heterologous.

334. The system or method of any of the preceding embodiments, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one of (a), (b) or (c) is heterologous.

335. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain; wherein the DBD and/or the endonuclease domain comprise a heterologous targeting domain that binds specifically to a sequence comprised in a target DNA molecule (e.g., a genomic DNA).

336. A polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a first target DNA binding domain, e.g., comprising a first Zn finger domain, (ii) a reverse transcriptase domain, (iii) an endonuclease domain, and (iv) a second target DNA binding domain, e.g., comprising a second Zn finger domain, heterologous to the first target DNA binding domain.

337. The polypeptide or nucleic acid encoding the polypeptide of any preceding embodiments, wherein (iii) comprises (iv).

338. A polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a target DNA binding domain, (ii) a reverse transcriptase domain, optionally (iii) an endonuclease domain, wherein the polypeptide comprises a heterologous linker replacing a portion of (i), (ii), or (iii), or replacing an endogenous linker connecting two of (i), (ii), or (iii).

339. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i).

340. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (ii).

341. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (iii).

342. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i) and (ii).

343. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (i) and (iii).

344. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, a portion of (ii) and (iii).

345. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (i) and (ii).

346. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (i) and (iii).

347. The polypeptide of any preceding embodiments, wherein the heterologous linker replaces, e.g., deletes, the endogenous linker connecting (ii) and (iii).

348. The polypeptide of any preceding embodiments, wherein the heterologous linker comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

349. The polypeptide of any preceding embodiments, wherein the heterologous linker comprises at least 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, or 500 amino acids.

350. A nucleic acid encoding the polypeptide of any preceding numbered embodiment.

351. A vector comprising the nucleic acid of any preceding embodiments.

352. A host cell comprising the nucleic acid of any preceding embodiments.

353. A host cell comprising the polypeptide of any preceding numbered embodiment.

354. A host cell comprising the vector of any preceding embodiments.

355. A pharmaceutical composition, comprising any preceding numbered system, nucleic acid, polypeptide, or vector; and a pharmaceutically acceptable excipient or carrier.

356. The pharmaceutical composition of Any preceding embodiments, wherein the pharmaceutically acceptable excipient or carrier is selected from a vector (e.g., a viral or plasmid vector), a vesicle (e.g., a liposome, an exosome, a natural or synthetic lipid bilayer), a lipid nanoparticle.

357. A polypeptide of any of the preceding embodiments, wherein the polypeptide further comprises a nuclear localization sequence.

358. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

359. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

360. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

361. Any preceding numbered embodiment, wherein the polypeptide, reverse transcriptase domain, or retrotransposase comprises a linker comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

362. Any preceding numbered embodiment, wherein the polypeptide comprises a DNA binding doman covalently attached to the remainder of the polypeptide by a linker, e.g., a linker comprising at least 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, or 500 amino acids.

363. Any preceding embodiments, wherein the linker is attached to the remainder of the polypeptide at a position in the DNA binding domain, RNA binding domain, reverse transcriptase domain, or endonuclease domain.

364. Any preceding embodiments, wherein the linker is attached to the remainder of the polypeptide at a position in the N-terminal side of an alpha helical region of the polypeptide, e.g., at a position corresponding to version v1 as described in Example 26 of PCT/US2019/048607, incorporated herein by reference.

365. Any preceding embodiments, wherein the linker is attached to the remainder of the polypeptide at a position in the C-terminal side of an alpha helical region of the polypeptide, e.g., preceding an RNA binding motif (e.g., a -1 RNA binding motif), e.g., at a position corresponding to version v2 as described in Example 26 of PCT/US2019/048607, incorporated herein by reference.

366. Any preceding embodiments, wherein the linker is attached to the remainder of the polypeptide at a position in the C-terminal side of a random coil region of the polypeptide, e.g., N-terminal relative to a DNA binding motif (e.g., a c-myb DNA binding motif), e.g., at a position corresponding to version v3 as described in Example 26 of PCT/US2019/048607, incorporated herein by reference.

367. Any preceding embodiments, wherein the linker comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

368. Any preceding numbered embodiment, wherein a polynucleotide sequence comprising at least about 500, 1000, 2000, 3000, 3500, 3600, 3700, 3800, 3900, or 4000 contiguous nucleotides from the 5' end of the template RNA sequence are integrated into a target cell genome.

369. Any preceding numbered embodiment, wherein a polynucleotide sequence comprising at least about 500, 1000, 2000, 2500, 2600, 2700, 2800, 2900, or 3000 contiguous nucleotides from the 3' end of the template RNA sequence are integrated into a target cell genome.

370. Any preceding numbered embodiment, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.21, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0 integrants/genome.

371. Any preceding numbered embodiment, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.085, 0.09, 0.1, 0.15, or 0.2 integrants/genome.

372. Any preceding numbered embodiment, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.036, 0.04, 0.05, 0.06, 0.07, or 0.08 integrants/genome.

373. Any preceding numbered embodiment, wherein the polypeptide comprises a functional endonuclease domain (e.g., wherein the endonuclease domain does not comprise a mutation that abolishes endonuclease activity, e.g., as described herein).

374. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof.

375. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence of the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof, and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

376. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof.

377. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence of the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

378. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof.

379. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence of the R2 polypeptide from a medium ground finch, e.g., *Geospiza fortis* (e.g., as described herein), or a functional fragment thereof and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

380. Any preceding embodiments, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000,

2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.21 integrants/genome.

381. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R4 polypeptide from a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof.

382. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence of the R4 polypeptide from a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof, and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

383. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R4 polypeptidefrom a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof.

384. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence of the R4 polypeptidefrom a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

385. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the R4 polypeptide from a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof.

386. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence of the R4 polypeptide from a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein), or a functional fragment thereof and further comprises a number of substitutions relative to the the sequence the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

387. Any preceding embodiments, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.085 integrants/genome.

388. Any preceding numbered embodiment, wherein introduction of the system into a target cell does not result in alteration (e.g., upregulation) of p53 and/or p21 protein levels, H2AX phosphorylation (e.g., gamma H2AX), ATM phosphorylation, ATR phosphorylation, Chk1 phosphorylation, Chk2 phosphorylation, and/or p53 phosphorylation.

389. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p53 protein level in the target cell to a level that is less than about 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%, or 90% of the p53 protein level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

390. Any preceding embodiments, wherein the p53 protein level is determined according to the method described in Example 30 of PCT/US2019/048607, incorporated herein by reference.

391. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p53 phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the p53

phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

392. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p21 protein level in the target cell to a level that is less than about 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%, or 90% of the p53 protein level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

393. Any preceding embodiments, wherein the p21 protein level is determined according to the method described in Example 30 of PCT/US2019/048607, incorporated herein by reference.

394. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of H2AX phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the H2AX phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

395. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of ATM phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the ATM phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

396. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of ATR phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the ATR

phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

397. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of Chk1 phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the Chk1 phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

398. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of Chk2 phosphorylation level in the target cell to a level that is less than about 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%, or 90% of the Chk2 phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

399. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and
(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence (e.g., a CRISPR spacer) that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

400. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the RT domain has a sequence of Table 1 or 3, or of a protein domain listed in Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

411. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (etRNA) (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides.

412. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the heterologous object sequence is at least 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, or 1,000 nts in length.

413. The system of any any preceding embodiments, wherein one or more of: the RT domain is heterologous to the DBD; the DBD is heterologous to the endonuclease domain; or the RT domain is heterologous to the endonuclease domain.

414. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and
(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,
wherein the system is capable of producing a deletion into the target site of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides.

415. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and
(b) a template (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,
wherein (a)(ii) and/or (a)(iii) comprises a TALE molecule; a zinc finger molecule; or a CRISPR/Cas molecule; or a functional variant (e.g., mutant) thereof.

416. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence (e.g., a CRISPR spacer) that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the endonuclease domain, e.g., nickase domain, cuts both strands of the target site DNA, and wherein the cuts are separated from one another by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 30 nucleotides.

417. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) a sequence that specifically binds the RT domain, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

418. The system of any preceding embodiments, wherein the template RNA further comprises a sequence that binds (a)(ii) and/or (a)(iii).

419. A system for modifying DNA comprising:

(a) a first polypeptide or a nucleic acid encoding the first polypeptide, wherein the first polypeptide comprises (i) a reverse transcriptase (RT) domain and (ii) optionally a DNA-binding domain,

(b) a second polypeptide or a nucleic acid encoding the second polypeptide, wherein the second polypeptide comprises (i) a DNA-binding domain (DBD); (ii) an endonuclease domain, e.g., a nickase domain; and

(c) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the second polypeptide (e.g., that binds (b)(i) and/or (b)(ii)), (ii) optionally a sequence that binds the first polypeptide (e.g., that specifically binds the RT domain), (iii) a heterologous object sequence, and (iv) a 3' homology domain.

420. A system for modifying DNA comprising:
(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, and (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain;
(b) a first template RNA (or DNA encoding the RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds the polypeptide (e.g., that binds (a)(ii) and/or (a)(iii)) and (ii) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (e.g., wherein the first RNA comprises a gRNA);
(c) a second template RNA (or DNA encoding the RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the polypeptide (e.g., that specifically binds the RT domain), (ii) a heterologous object sequence, and (iii) a 3' homology domain.

421 The system of any preceding embodiments, wherein the second template RNA comprises (i).

422 The system of any preceding embodiments, wherein the first template RNA comprises a first conjugating domain and the second template RNA comprises a second conjugating domain.

423 The system of any preceding embodiments, wherein the first and second conjugating domains are capable of hybridizing to one another, e.g., under stringent conditions.

424 The system of any preceding embodiments, wherein association of the first conjugating domain and the second conjugating domain colocalizes the first template RNA and the second template RNA.

425. The system of any previous embodiment, wherein the template RNA comprises (i).

426. The system of any previous embodiment, wherein the template RNA comprises (ii).

427. The system of any previous embodiment, wherein the template RNA comprises (i) and (ii).

428. A template RNA (or DNA encoding the template RNA) comprising a targeting domain (e.g., a heterologous targeting domain) that binds specifically to a sequence comprised in the target DNA molecule (e.g., a genomic DNA), a sequence that specifically binds an RT domain of a polypeptide, and a heterologous object sequence.

429. The system, method, or template RNA of any of the preceding embodiments, wherein the polypeptide comprises a heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule (e.g., a genomic DNA).

430. The system, method, or template RNA of any preceding embodiments, wherein the heterologous targeting domain binds to a different nucleic acid sequence than the unmodified polypeptide.

431. The system, method, or template RNA of any preceding embodiments, wherein the polypeptide does not comprise a functional endogenous targeting domain (e.g., wherein the polypeptide does not comprise an endogenous targeting domain).

432. The system, method, or template RNA of any preceding embodiments, wherein the heterologous targeting domain comprises a zinc finger (e.g., a zinc finger that binds specifically to the sequence comprised in the target DNA molecule).

433. The system, method, or template RNA of any preceding embodiments, wherein the heterologous targeting domain comprises a Cas domain (e.g., a Cas9 domain, or a mutant or variant thereof, e.g., a Cas9 domain that binds specifically to the sequence comprised in the target DNA molecule).

434. The system, method, or template RNA of any preceding embodiments, wherein the Cas domain is associated with a guide RNA (gRNA)..

435. The system, method, or template RNA of any preceding embodiments, wherein the heterologous targeting domain comprises an endonuclease domain (e.g., a heterologous endonuclease domain).

436. The system, method, or template RNA of any preceding embodiments, wherein the endonuclease domain comprises a Cas domain (e.g., a Cas9 or a mutant or variant thereof).

437. The system, method, or template RNA of any preceding embodiments, wherein the Cas domain is associated with a guide RNA (gRNA).

438. The system, method, or template RNA of any preceding embodiments, wherein the endonuclease domain comprises a Fok1 domain.

439. The system, method, or template RNA of any any preceding embodiments, wherein the template nucleic acid molecule comprises at least one (e.g., one or two) heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target DNA molecule (e.g., a genomic DNA).

440. The system, method, or template RNA of any preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at or within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of the 5' end of the template nucleic acid molecule.

441. The system, method, or template RNA of any preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at or within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of the 3' end of the template nucleic acid molecule.

442. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick

site (e.g., produced by a nickase, e.g., an endonuclease domain, e.g., as described herein) in the target DNA molecule.

443. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence has less than 50%, 40%, 30%, 20%, 10%, 5%, 4%, 3%, 2%, or 1% sequence identity with a nucleic acid sequence complementary to an endogenous homology sequence of an unmodified form of the template RNA.

444. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence has having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence of the target DNA molecule that is different the sequence bound by an endogenous homology sequence (e.g., replaced by the heterologous homology sequence).

445. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 3' end) having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence positioned 5' to a nick site of the target DNA molecule (e.g., a site nicked by a nickase, e.g., an endonuclease domain as described herein).

446. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 5' end) suitable for priming target-primed reverse transcription (TPRT) initiation.

447. The system, method, or template RNA of any preceding embodiments, wherein the heterologous homology sequence has at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence positioned within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 3' relative to) a target insertion site, e.g., for a heterologous object sequence (e.g., as described herein), in the target DNA molecule.

448. The system, method, or template RNA of any preceding embodiments, wherein the template nucleic acid molecule comprises a guide RNA (gRNA), e.g., as described herein.

449. The system, method, or template RNA of any preceding embodiments, wherein the template nucleic acid molecule comprises a gRNA spacer sequence (e.g., at or within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides of its 5' end).

450. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), (ii) a sequence that specifically binds an RT domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

451. The template RNA of any preceding embodiments, further comprising (v) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide (e.g., the same polypeptide comprising the RT domain).

452. The template RNA of any preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2 or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

453. The template RNA of any preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, wherein the RT domain further comprises a number of substitutions relative to the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

454. The template RNA of any preceding embodiments, wherein the sequence of (ii) specifically binds the RT domain.

455. The template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is a sequence, e.g., a UTR sequence, that binds the RT domain in a wild-type context, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

456. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

457. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (iii) a heterologous object sequence, (iv) a 3' target homology domain, (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), and (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide.

458. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

459. The template RNA of any preceding embodiments, wherein the template RNA comprises (i).

460. The template RNA of any preceding embodiments, wherein the template RNA comprises (ii).

461. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) a sequence that specifically binds an RT domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

462. The template RNA of any preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3, or of a protein domain listed in Table 2or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

463. The template RNA of any preceding embodiments, further comprising (v) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide (e.g., the same polypeptide comprising the RT domain).

464. The template RNA of any preceding embodiments, wherein the sequence of (ii) specifically binds an RT domain of Table 1 or 3, or listed in Table 2, or an RT domain sequence that has at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

465. The template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is a sequence of Table 1 or 3, or of a protein domain listed in Table 2, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

466. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (iii) a heterologous object sequence, and (iv) a 3' homology domain.

467. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (iii) a heterologous object sequence, (iv) a 3' homology domain, (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), and (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide,.

468. The system or template RNA of any preceding embodiments, wherein the template RNA, first template RNA, or second template RNA comprises a sequence that specifically binds the RT domain.

469. The system or template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (i) and (ii).

470. The system or template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (ii) and (iii).

471. The system or template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (iii) and (iv).

472. The system or template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (iv) and (i).

473. The system or template RNA of any preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (i) and (iii).

474. A system for modifying DNA, comprising:
(a) a first template RNA (or DNA encoding the first template RNA) comprising (i) sequence that binds an endonuclease domain, e.g., a nickase domain, and/or a DNA-binding domain (DBD) of a polypeptide, and (ii) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (e.g., wherein the first RNA comprises a gRNA);
(b) a second template RNA (or DNA encoding the second template RNA) comprising (i) a sequence that specifically binds a reverse transcriptase (RT) domain of a polypeptide (e.g., the polypeptide of (a)), (ii) a target site binding sequence (TSBS), and (iii) an RT template sequence.

475. The system of any preceding embodiments wherein the nucleic acid encoding the first template RNA and the nucleic acid encoding the second template RNA are two separate nucleic acids.

476. The system of any preceding embodiments, wherein the nucleic acid encoding the first template RNA and the nucleic acid encoding the second template RNA are part of the same nucleic acid molecule, e.g., are present on the same vector.

477. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain, wherein the RT domain has a sequence of Table

1 or 3, or of a protein domain listed in Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

478. A system for modifying DNA, comprising:

(a) a first polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises a reverse transcriptase (RT) domain, wherein the RT domain has a sequence of Table 1 or 3, or of a protein domain listed in Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and optionally a DNA-binding domain (DBD) (e.g., a first DBD); and

(b) a second polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a DBD (e.g., a second DBD); and (ii) an endonuclease domain, e.g., a nickase domain.

479. The system of any preceding embodiments, wherein the nucleic acid encoding the first polypeptide and the nucleic acid encoding the second polypeptide are two separate nucleic acids.

480. The system of any preceding embodiments, wherein the nucleic acid encoding the first polypeptide and the nucleic acid encoding the second polypeptide are part of the same nucleic acid molecule, e.g., are present on the same vector.

481. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein an RNA of the system (e.g., template RNA, the RNA encoding the polypeptide of (a), or an RNA expressed from a heterologous object sequence integrated into a target DNA) comprises a microRNA binding site, e.g., in a 3' UTR.

482. The system, method, kit, template RNA, or reaction mixture of embodiment 481, wherein the microRNA binding site is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type.

483. The system, method, kit, template RNA, or reaction mixture of embodiment 481 or 482, wherein the miRNA is miR-142, and/or wherein the non-target cell is a Kupffer cell or a blood cell, e.g., an immune cell.

484. The system, method, kit, template RNA, or reaction mixture of embodiment 481 or 482, wherein the miRNA is miR-182 or miR-183, and/or wherein the non-target cell is a dorsal root ganglion neuron.

485. The system, method, kit, template RNA, or reaction mixture of any of embodiments 481-484, wherein the system comprises a first miRNA binding site that is recognized by a first miRNA (e.g., miR-142) and the system further comprises a second miRNA binding site that is recognized by a second miRNA (e.g., miR-182 or miR-183), wherein the first miRNA binding site and the second miRNA binding site are situated on the same RNA or on different RNAs of the system.

486. The system, method, kit, template RNA, or reaction mixture of any of embodiments 481-485, wherein the template RNA comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

487. The system, method, kit, template RNA, or reaction mixture of any of embodiments 481-486, wherein the RNA encoding the polypeptide of (a) comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

488. The system, method, kit, template RNA, or reaction mixture of any of embodiments 481-487, wherein the RNA expressed from a heterologous object sequence integrated into a target DNA comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

Definitions

Domain: The term "domain" as used herein refers to a structure of a biomolecule that contributes to a specified function of the biomolecule. A domain may comprise a contiguous region (e.g., a contiguous sequence) or distinct, non-contiguous regions (e.g., non-contiguous sequences) of a biomolecule. Examples of protein domains include, but are not limited to, an endonuclease domain, a DNA binding domain, a reverse transcription domain; an example of a domain of a nucleic acid is a regulatory domain, such as a transcription factor binding domain.

Exogenous: As used herein, the term exogenous, when used with reference to a biomolecule (such as a nucleic acid sequence or polypeptide) means that the biomolecule was introduced into a host genome, cell or organism by the hand of man. For example, a nucleic acid that is as added into an existing genome, cell, tissue or subject using recombinant DNA techniques or other methods is exogenous to the existing nucleic acid sequence, cell, tissue or subject.

First/Second Strand: As used herein, first strand and second strand, as used to describe the individual DNA strands of target DNA, distinguish the two DNA strands based upon which strand the reverse transcriptase domain initiates polymerization, e.g., based upon where target primed synthesis initiates. The first strand refers to the strand of the target DNA upon which the reverse transcriptase domain initiates polymerization, e.g., where target primed synthesis initiates. The second strand refers to the other strand of the target DNA. First and second strand designations do not describe the target site DNA strands in other respects; for example, in some embodiments the first and second strands are nicked by a polypeptide described herein, but the designations 'first' and 'second' strand have no bearing on the order in which such nicks occur.

Genomic safe harbor site (GSH site): A genomic safe harbor site is a site in a host genome that is able to accommodate the integration of new genetic material, e.g., such that the inserted genetic element does not cause significant alterations of the host genome posing a risk to the host cell or organism. A GSH site generally meets 1, 2, 3, 4, 5, 6, 7, 8 or 9 of the following criteria: (i) is located >300kb from a cancer-related gene; (ii) is >300kb from a miRNA/other functional small RNA; (iii) is >50kb from a 5' gene end; (iv) is >50kb from a replication origin; (v) is >50kb away from any ultraconservered element; (vi) has low transcriptional activity (i.e.

no mRNA +/- 25 kb); (vii) is not in copy number variable region; (viii) is in open chromatin; and/or (ix) is unique, with 1 copy in the human genome. Examples of GSH sites in the human genome that meet some or all of these criteria include (i) the adeno-associated virus site 1 (AAVS1), a naturally occurring site of integration of AAV virus on chromosome 19; (ii) the chemokine (C-C motif) receptor 5 (CCR5) gene, a chemokine receptor gene known as an HIV-1 coreceptor; (iii) the human ortholog of the mouse Rosa26 locus; (iv) the rDNA locus (v) the albumin locus, e.g., for liver cell applications; (vi) the T-cell receptor alpha constant (TRAC) locus, e.g., for T-cell applications. Additional GSH sites are known and described, e.g., in Pellenz et al. epub August 20, 2018 (https://doi.org/10.1101/396390).

Heterologous: The term heterologous, when used to describe a first element in reference to a second element means that the first element and second element do not exist in nature disposed as described. For example, a heterologous polypeptide, nucleic acid molecule, construct or sequence refers to (a) a polypeptide, nucleic acid molecule or portion of a polypeptide or nucleic acid molecule sequence that is not native to a cell in which it is expressed, (b) a polypeptide or nucleic acid molecule or portion of a polypeptide or nucleic acid molecule that has been altered or mutated relative to its native state, or (c) a polypeptide or nucleic acid molecule with an altered expression as compared to the native expression levels under similar conditions. For example, a heterologous regulatory sequence (e.g., promoter, enhancer) may be used to regulate expression of a gene or a nucleic acid molecule in a way that is different than the gene or a nucleic acid molecule is normally expressed in nature. In another example, a heterologous domain of a polypeptide or nucleic acid sequence (e.g., a DNA binding domain of a polypeptide or nucleic acid encoding a DNA binding domain of a polypeptide) may be disposed relative to other domains or may be a different sequence or from a different source, relative to other domains or portions of a polypeptide or its encoding nucleic acid. In certain embodiments, a heterologous nucleic acid molecule may exist in a native host cell genome, but may have an altered expression level or have a different sequence or both. In other embodiments, heterologous nucleic acid molecules may not be endogenous to a host cell or host genome but instead may have been introduced into a host cell by transformation (e.g., transfection, electroporation), wherein the added molecule may integrate into the host genome or can exist as extra-chromosomal genetic material either transiently (e.g., mRNA) or semi-stably for more than one generation (e.g., episomal viral vector, plasmid or other self-replicating vector).

Mutation or Mutated: The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference (e.g., native) nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art.

Nucleic acid molecule: Nucleic acid molecule refers to both RNA and DNA molecules including, without limitation, cDNA, genomic DNA and mRNA, and also includes synthetic nucleic acid molecules, such as those that are chemically synthesized or recombinantly produced, such as RNA templates, as described herein. The nucleic acid molecule can be double-stranded or single-stranded, circular or linear. If single-stranded, the nucleic acid molecule can be the sense strand or the antisense strand. Unless otherwise indicated, and as an example for all sequences described herein under the general format "SEQ. ID NO:," "nucleic acid comprising SEQ. ID NO:1" refers to a nucleic acid, at least a portion which has either (i) the sequence of SEQ. ID NO:1, or (ii) a sequence complimentary to SEQ. ID NO:1. The choice between the two is dictated by the context in which SEQ. ID NO:1 is used. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complimentary to the desired target. Nucleic acid sequences of the present disclosure may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more naturally occurring nucleotides with an analog, inter-nucleotide modifications such as uncharged linkages (for example, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (for example, phosphorothioates, phosphorodithioates, etc.), pendant moieties, (for example, polypeptides), intercalators (for example, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (for example, alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of a molecule. Other modifications can include, for example, analogs in which the

ribose ring contains a bridging moiety or other structure such as modifications found in "locked" nucleic acids.

Gene expression unit: a gene expression unit is a nucleic acid sequence comprising at least one regulatory nucleic acid sequence operably linked to at least one effector sequence. A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if the promoter or enhancer affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be contiguous or non-contiguous. Where necessary to join two protein-coding regions, operably linked sequences may be in the same reading frame.

Host: The terms host genome or host cell, as used herein, refer to a cell and/or its genome into which protein and/or genetic material has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell and/or genome, but to the progeny of such a cell and/or the genome of the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A host genome or host cell may be an isolated cell or cell line grown in culture, or genomic material isolated from such a cell or cell line, or may be a host cell or host genome which composing living tissue or an organism. In some instances, a host cell may be an animal cell or a plant cell, e.g., as described herein. In certain instances, a host cell may be a bovine cell, horse cell, pig cell, goat cell, sheep cell, chicken cell, or turkey cell. In certain instances, a host cell may be a corn cell, soy cell, wheat cell, or rice cell.

Operative association: As used herein, "operative association" describes a functional relationship between two nucleic acid sequences, such as a 1) promoter and 2) a heterologous object sequence, and means, in such example, the promoter and heterologous object sequence (e.g., a gene of interest) are oriented such that, under suitable conditions, the promoter drives expression of the heterologous object sequence. For instance, the template nucleic acid may be single-stranded, e.g., either the (+) or (-) orientation but an operative association between promoter and heterologous object sequence means whether or not the template nucleic acid will transcribe in a particular state, when it is in the suitable state (e.g., is in the (+) orientation, in the

presence of required catalytic factors, and NTPs, etc.), it does accurately transcribe. Operative association applies analogously to other pairs of nucleic acids, including other tissue-specific expression control sequences (such as enhancers, repressors and microRNA recognition sequences), IR/DR, ITRs, UTRs, or homology regions and heterologous object sequences or sequences encoding a transposase.

Pseudoknot: A "pseudoknot sequence" sequence, as used herein, refers to a nucleic acid (e.g., RNA) having a sequence with suitable self-complementarity to form a pseudoknot structure, e.g., having: a first segment, a second segment between the first segment and a third segment, wherein the third segment is complementary to the first segment, and a fourth segment, wherein the fourth segment is complementary to the second segment. The pseudoknot may optionally have additional secondary structure, e.g., a stem loop disposed in the second segment, a stem-loop disposed between the second segment and third segment, sequence before the first segment, or sequence after the fourth segment. The pseudoknot may have additional sequence between the first and second segments, between the second and third segments, or between the third and fourth segments. In some embodiments, the segments are arranged, from 5' to 3': first, second, third, and fourth. In some embodiments, the first and third segments comprise five base pairs of perfect complementarity. In some embodiments, the second and fourth segments comprise 10 base pairs, optionally with one or more (e.g., two) bulges. In some embodiments, the second segment comprises one or more unpaired nucleotides, e.g., forming a loop. In some embodiments, the third segment comprises one or more unpaired nucleotides, e.g., forming a loop.

Stem-loop sequence: As used herein, a "stem-loop sequence" refers to a nucleic acid sequence (e.g., RNA sequence) with sufficient self-complementarity to form a stem-loop, e.g., having a stem comprising at least two (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) base pairs, and a loop with at least three (e.g., four) base pairs. The stem may comprise mismatches or bulges.

Tissue-specific expression-control sequence(s): As used herein, a "tissue-specific expression-control sequence" means nucleic acid elements that increase or decrease the level of a transcript comprising the heterologous object sequence in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). In some embodiments, a tissue-specific expression-control sequence preferentially drives or represses transcription, activity, or the half-life of a transcript comprising the heterologous object sequence

in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). Exemplary tissue-specific expression-control sequences include tissue-specific promoters, repressors, enhancers, or combinations thereof, as well as tissue-specific microRNA recognition sequences. Tissue specificity refers to on-target (tissue(s) where expression or activity of the template nucleic acid is desired or tolerable) and off-target (tissue(s) where expression or activity of the template nucleic acid is not desired or is not tolerable). For example, a tissue-specific promoter (such as a promoter in a template nucleic acid or controlling expression of a transposase) drives expression preferentially in on-target tissues, relative to off-target tissues. In contrast, a micro-RNA that binds the tissue-specific microRNA recognition sequences (either on a nucleic acid encoding the transposase or on the template nucleic acid, or both) is preferentially expressed in off-target tissues, relative to on-target tissues, thereby reducing expression of a template nucleic acid (or transposase) in off-target tissues. Accordingly, a promoter and a microRNA recognition sequence that are specific for the same tissue, such as the target tissue, have contrasting functions (promote and repress, respectively, with concordant expression levels, i.e., high levels of the microRNA in off-target tissues and low levels in on-target tissues, while promoters drive high expression in on-target tissues and low expression in off-target tissues) with regard to the transcription, activity, or half-life of an associated sequence in that tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Figure 1. The linker region at the C-terminus of the DNA-binding domain of R2Tg can be truncated and modified. Deletions in the Natural Linker from the myb domain at A or B to positions 1 or 2 along with replacement by 3GS or XTEN synthetic linkers were constructed (A). Integration efficiency was measured in HEK293T cells by ddPCR (B).

Figure 2. Landing pads designed for testing target site mutations of R2Tg Gene Writer.

Figure 3a. ddPCR assay measuring percentage of integrations from all lentiviral integrated landing pads per cell

Figure 3b. Amplicon-sequencing and NGS analysis of indels present at landing pads sites.

Figure 4. AAVS1 ZFP replacement of DNA binding domain of a Retrotransposase Gene Writer.

Figure 5. Cas9 or Cas9 nickase replacement of DNA binding domain of Retrotransposase GeneWriters with or without active EN domain (*= mutant)

Figure 6. AAVS1 ZFP fusion to a Retrotransposase Gene Writer with or without functional DNA binding domain.

Figure 7. Schematic of second strand nicking. (A) A Cas9 nickase is fused to a Gene Writer protein. The Gene Writer protein introdces a nick in a DNA strand through its EN domain (shown as *), and the fused Cas9 nickase introduces a nicks on either top or bottom DNA strands (shown as X). (B) A Gene Writer is targeted to DNA through its DNA biding domain and introduces a DNA nick with its EN domain (*). A Cas9 nickase is then used the generate a second nick (X) at the top or bottom strand, upstream or downstream of the EN introduced nick.

Figure 8. Schematic of nickaseCas9-GeneWriter fusions. (A) Schematic of nickaseCas9 fused to Gene Writer protein. (B) Schematic of 3' extended gRNA.

Figure 9. Schematic of nickaseCas9-GeneWriter fusions. (A) Schematic of nickaseCas9 fused to Gene Writer protein. (B) Schematic of donor transgene flanked by UTRs and homology to the cut site.

Figure 10. Schematic of constructs. (A) Schematic of Gene Writer protein. (B) Schematic of donor transgene flanked by UTRs and homology to the cut site. (C) Schematic of Cas9 constructs used.

Figure 11. The schematics for mRNA encoding Gene Writer (A). The native untranslated regions (UTRs) were replaced by 5' and 3' UTRs optimized for the protein expression (shown as 5' UTRexp and 3' UTRexp). The Gene Writer protein expression was assayed by HiBit assay by probing HiBit tag expression (B).

Figure 12. Genome integration induced by Gene Writer protein with its native UTRs and UTRs optimized for the protein expression. The Gene Writing activity with non-native UTRs is stimulated by the presence of the RNA template bearing the retrotransposon native UTRs.

Figure 13. Delivery of Gene Writer system using mRNA encoding the polypeptide and plasmid DNA encoding the RNA template for retrotransposition.

Figure 14. Diagrams of example 5'UTR engineering strategies. HA = homology arm; K = Kozak sequence; pA = poly A signal; AMa = *A. maritima*; Rx = other species of retrotransposon.

Figure 15. Possible location of an intron (or introns) within the RNA template. Introns are shown by curved lines. 5'HA: 5' homology arm; 3' HA: 3' homology arm; 5' UTR: Retrotransposon-specific 5'UTR; 3' UTR: Retrotransposon-specific 3' UTR; GOI: gene of interest. Orange blocks correspond to the sequence designed to be expressed from the genomic location harboring its own cell specific promoter, poly(A) signal and UTRs for the protein expression (5' and 3' UTR_{exp}). The sequence can be oriented in the sense (shown above) or the antisense orientation related to retrotransposon UTRs and homology arms. The intron can be located within GOI, or within UTR_{exp}.

Figure 16. Genome integration in HEK293T cells as reported by 3' ddPCR assay. The Gene Writer mRNA at 0.5 µg/well was co-transfected with the RNA templates with or without enzymatically added cap 1 and the poly(A) tail. The Gene Writer mRNA to RNA transgene ratio was 1:1.

Figure 17. Genome integration detected by 3' ddPCR induced by expression of Gene Writer mRNA produced with either unmodified (G0) or modified nucleotides (pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U) or 5-methylcytidine (5mC)). 1 ug of Gene Writer mRNA per well was used. The non-modified RNA template was used. The Gene Writer RNA to the RNA template were co-transfected in 1:8 molar ratio.

Figure 18. The modules comprising a typical Gene Writer RNA template, where individual modules can be combined, re-arranged, and/or left out to produce a Gene Writer template. A = 5' homology arm; B = Ribozyme; C = 5' UTR; D = heterologous object sequence; E = 3' UTR; F = 3' homology arm.

Figure 19. The modules comprising a typical Gene Writer RNA template, where individual modules can be combined, re-arranged, and/or left out to produce a Gene Writer template. A = 5' homology arm; B = Ribozyme; C = 5' UTR; D = heterologous object sequence; E = 3' UTR; F = 3' homology arm

Figure 20. Construct diagram of driver and transgene plasmids. Homology arms (HA) and stuffer sequences are variable in this set of experiments

Figure 21. Integration efficiency at 3' or 5' end of transgene across constructs tested as measured via digital droplet PCR. Each point represents a replicate experiment. Bars represent

mean of two replicate experiments. (A,B) Integration efficiency as measured across the 3' junction between transgene and host rDNA. (C,D) Integration efficiency as measured across the 5' junction.

Figure 22. Example illustration of homology shift design tested for +/-3bp. Red indicates homology to 5' of the wildtype (WT) nick site, and blue indicates homology 3' to the nick. 3' shifted constructs (+) begin 3' homology farther downstream from the nick. 5' shifted constructs (-) incorporate homology from the 5' of the nick into the 3' homology arm.

Figure 23. 3' integration results from shifting the 3' homology arm of the transgene. Each data point represents a replicate, while the bar represents the mean of two replicates.

Figure 24. (A) Timeline of experiment. (B) Schematic of R2Tg and transgene construct configurations. (C) Western Blot against Rad51 shows loss of Rad51 protein expression at day 3.

Figure 25. U2OS cells were treated with a non targeting control siRNA (ctrl) or siRNA against Rad51, along with R2Tg Wt or control RT and EN mutants. ddPCR at the 3' (A) or 5' (B) junction was used to assess integration efficiency on day 3.

Figure 26. (A) Sequence map of Ribozyme of R2 element from *Taeniopygia guttata* (R2Tg) in context of modules of Gene Writer transgene molecule RNA. The Ribozyme features are denoted as: P, based paired region; P′, based pair region complement strand; L, loop at end of P region; J, nucleotides joining base paired regions. This Figure discloses SEQ ID NO: 1592. (B) Prediction of ribozyme secondary structure of R2Tg. Shaded box indicates a predicted catalytic position that could be used to inactivate the ribozyme. This Figure discloses SEQ ID NO: 1592.

Figure 27. Sequence map of Ribozyme of R2 element from *Taeniopygia guttata* (R2Tg) in context of modules of Gene Writer transgene molecule RNA. The Ribozyme features are denoted as: P, based paired region; P′, based pair region complement strand; L, loop at end of P region; J, nucleotides joining base paired regions. This Figure discloses SEQ ID NO: 1592.

Figure 28. Prediction of ribozyme secondary structure of R2 element from Taeniopygia guttata. This Figure discloses SEQ ID NO: 1592.

Figures 29A and 29B are a series of diagrams showing examples of configurations of Gene Writers using domains derived from a variety of sources. Gene Writers as described herein may or may not comprise all domains depicted. For example, a GeneWriter may, in some instances, lack an RNA-binding domain, or may have single domains that fulfill the functions of multiple domains, e.g., a Cas9 domain for DNA binding and endonuclease activity. Exemplary

domains that can be included in a GeneWriter polypeptide include DNA binding domains (e.g., comprising a DNA binding domain of an element of a sequence listed in any of Tables 1 or 3, or a domain listed in Table 2; a zinc finger; a TAL domain; Cas9; dCas9; nickase Cas9; a transcription factor, or a meganuclease), RNA binding domains (e.g., comprising an RNA binding domain of B-box protein, MS2 coat protein, dCas, or an element of a sequence listed in any of Tables 1 or 3, or a domain listed in Table 2), reverse transcriptase domains (e.g., comprising a reverse transcriptase domain of an element of a sequence listed in any of Tables 1 or 3, or a domain listed in Table 2), and/or an endonuclease domain (e.g., comprising an endonuclease domain of an element of a sequence listed in any of Tables 1 or 3, or a domain listed in Table 2; Cas9; nickase Cas9; a restriction enzyme (e.g., a type II restriction enzyme, e.g., FokI); a meganuclease; a Holliday junction resolvase; an RLE retrotranspase; an APE retrotransposase; or a GIY-YIG retrotransposase). Exemplary GeneWriter polypeptides comprising exemplary combinations of such domains are shown in the bottom panel.

Figures 30A and B illustrates mutations to the DNA binding motifs in a Gene Writer polypeptide that inhibit native site integration. Figure 30A discloses a general domain structure of a R2Tg retrotransposase (top), comprising a DNA-binding domain containing multiple predicted DNA-binding elements (bottom). The two zinc finger motifs and c-myb motif indicated in the protein were mutated as according to Example 30. Figure 30B illustrates that integration activity for the mutants of the ZF1, ZF2, and c-myb domains was assessed in HEK293T cells by analyzing native rDNA site integration frequency using ddPCR. Each individual mutant, as well as the triple mutant, was compared to wild-type (positive control) and an endonuclease-inactivated enzyme (negative control). Data indicate averages of two replicates. Figures legends: ZF=zinc finger; myb=c-myb-like DNA binding motif; RBD=RNA-binding domain; RT=reverse transcriptase domain; EN=endonuclease domain; *=mutated domain; CNV/Genome=average copies of integrated DNA per genome copy.

Figures 31A and 31B illustrates that the endonuclease cleavage site of a retrotransposase can be detected by indel signature. Figures 31A shows the predicted binding and cleavage locations in the target site of the R2Tg retrotransposase. Figure 31B shows the cleavage site of the R2Tg retrotransposase was validated by analysis of genome alterations resulting from endonuclease activity. Plasmid DNA encoding the R2Tg retrotransposase was nucleofected into U2OS cells and genomic DNA was harvested after three days. Target site amplicons were

generated using site-specific primers and sequenced to determine the location of genome alterations indicative of endonuclease activity. Shown here is a graph depicting the frequence of insertions (circles) and deletions (triangles) per nucleotide of sequence (x-axis). The peak of insertion signal (horizontal line under figure) was localized to the predicted GG dinucleotide. Figure legend: ZF=zinc finger; myb=c-myb-like DNA binding motif

Figures 32A and B shows determination of sequence determinants for endonuclease activity of a retrotransposase by schematic representation of Landing pad screen. Figure 33A shows a lentiviral expression vector was used to clone landing pads containing a native R2 retrotransposase target site or sites comprising mutations relative to the native site. Lentiviral constructs were packaged and used to transduce U2OS cells for generating cell lines with the landing pads integrated into the genome. The landing pad additionally comprised a green fluorescent protein (GFP) reporter cassette for titer determinations. Figure 33B shows Landing pad sequences comprising wild-type or mutational variants of the R2 site. A native rDNA sequence landing pad containing the unmodified rDNA sequence (WT_R2Tg) was used as a positive control. A series of 16 landing pads are shown with mutated regions indicated in dark gray and the GG cleavage site in light gray (left). The graph (right) was used to visualize the magnitude of each target site change on endonuclease activity of the enzyme. Mutation to the AA dinucleotide adjacent to the GG dinucleotide cleavage site was found to severely impair endonuclease activity, thus the motif AAGG is important for R2Tg endonuclease activity.

Figure 33 shows the overview of landing pad screen for retargeting a Gene Writer polypeptide. Schematic of the landing pad library built to analyze the sequences recognized in R2Tg retargeting. The AAVS1-ZF binding site (dark gray and labeled AAVS1) was used as a DNA binding motif for retargeting, and all landing pads were built in the context of the human AAVS1 genomic sequence. rDNA sequence (black) was added to the AAVS1 sequence in various ways: (Category 1) different length of rDNA sequence, (Category 2) different distances between the AAVS1 ZF binding site and the rDNA sequence, (Category 3) different orientations of the rDNA sequence relative to the AAVS1 site. Categories 1, 2, and 3 were explored combinatorially, resulting in lading pads of various rDNA sequence lengths and various distances and orientations relative to the AAVS ZF binding site. The AAGG minimum sequence for R2Tg cleavage was maintained in all landing pads (black box with white fill). Each landing

pad was designed with a unique barcode at the 3' end of the sequence to enable computational extraction and analysis of landing pad sequences from the pool.

Figure 34 represents sequencing-based determination of landing pad representation in U2OS pool. The landing pad pool of U2OS cells was sequenced and analyzed to determine barcode representation. Approximately 94% of landing pads were represented by at least 10,000 reads (horizontal black bar). The x-axis indicates landing pad identity and the y-axis shows the total reads for that barcode.

Figures 35 A and B discloses generation of indel signatures in a landing pad library enables screening of chimeric Gene Writer polypeptides. Figure 35A shows a landing pad library comprising various compositions of AAVS1 and R2 rDNA target sequences was treated with a full-length R2Tg retrotransposase fused to a zinc finger for AAVS1 sequence recognition. Amplicon sequencing was performed and insertion frequencies at the GG target site (y-axis) are plotted for each landing pad (x-axis). A representative number of 230 landing pads is shown on the x-axis. Positive controls containing 200 nt of rDNA sequence are indicated and showed the expected insertion signatures at the GG cleavage site. The negative control lacking any rDNA sequence did not harbor any insertions. The lengths of the rDNA sequence comprised in landing pads where insertion signatures were found indicated and corresponded to 44, 64, and 84 nt. Figure 35B is an illustrative representation of landing pad configurations found to contain signatures of endonuclease activity.

Figures 36 A and B discloses generation of indel signatures in a landing pad library enables screening of chimeric Gene Writer polypeptides. Figure 36A shows a landing pad library comprising various compositions of AAVS1 and R2 rDNA target sequences was treated with a full-length R2Tg retrotransposase fused to a zinc finger for AAVS1 sequence recognition. Amplicon sequencing was performed and insertion frequencies at the GG target site (y-axis) are plotted for each landing pad (x-axis). A representative number of 230 landing pads is shown on the x-axis. The negative control lacking any rDNA sequence did not harbor any insertions. The lengths of the rDNA sequence comprised in landing pads where insertion signatures were found indicated and corresponded to 44, 64, and 84 nt. Figure 36B is an illustrative representation of landing pad configurations found to contain signatures of endonuclease activity.

Figure 37 Aand B describes luciferase activity assay for primary cells. LNPs formulated as according to Example 38 were analyzed for delivery of cargo to primary human (A) and

mouse (B) hepatocytes, as according to Example 39. The luciferase assay revealed dose-responsive luciferase activity from cell lysates, indicating successful delivery of RNA to the cells and expression of Firefly luciferase from the mRNA cargo.

Figure 38 shows LNP-mediated delivery of RNA cargo to the murine liver. Firefly luciferase mRNA-containing LNPs were formulated and delivered to mice by iv, and liver samples were harvested and assayed for luciferase activity at 6, 24, and 48 hours post administration. Reporter activity by the various formulations followed the ranking LIPIDV005>LIPIDV004>LIPIDV003. RNA expression was transient and enzyme levels returned near vehicle background by 48 hours, post-administration.

Figure 39. Shows improving expression of Cas-RT fusions through choice of linker sequence. To assess how linkers can alter the expression of novel Gene Writer polypeptides in human cells, U2OS cells were transfected with Cas-RT expression plasmids harboring various linkers from Table 42 fusing the Cas9(N863A) nickase to the RT domain of an RNA-binding domain mutated R2Bm retrotransposase. Cell lysates were collected and analyzed by Western blot using a primary antibody against Cas9. A primary antibody against vinculin (left) or GADPH (right) was included as a loading control. Cas9 controls on the left represent titration of a Cas9 expression plasmid. Empty arrows indicate the original linker tested, while the filled arrow represents a linker (Linker **10;** SEQ ID NO: 468)) found to substantially improve expression of the fusion polypeptide. Sample numbers correspond to linker sequence identifiers in Table 42.

DETAILED DESCRIPTION

This disclosure relates to compositions, systems and methods for targeting, editing, modifying or manipulating a DNA sequence (e.g., inserting a heterologous object DNA sequence into a target site of a mammalian genome) at one or more locations in a DNA sequence in a cell, tissue or subject, e.g., *in vivo* or *in vitro*. The object DNA sequence may include, e.g., a coding sequence, a regulatory sequence, a gene expression unit.

More specifically, the disclosure provides retrotransposon-based systems for inserting a sequence of interest into the genome. This disclosure is based, in part, on a bioinformatic

analysis to identify retrotransposase sequences and the associated 5' UTR and 3' UTR from a variety of organisms (see Table 3).

Gene-writer™ genome editors

Non-long terminal repeat (LTR) retrotransposons are a type of mobile genetic elements that are widespread in eukaryotic genomes. They include two classes: the apurinic/apyrimidinic endonuclease (APE)-type and the restriction enzyme-like endonuclease (RLE)-type. The APE class retrotransposons are comprised of two functional domains: an endonuclease/DNA binding domain, and a reverse transcriptase domain. The RLE class are comprised of three functional domains: a DNA binding domain, a reverse transcription domain, and an endonuclease domain. The reverse transcriptase domain of non-LTR retrotransposon functions by binding an RNA sequence template and reverse transcribing it into the host genome's target DNA. The RNA sequence template has a 3' untranslated region which is specifically bound to the transposase, and a variable 5' region generally having Open Reading Frame(s) ("ORF") encoding transposase proteins. The RNA sequence template may also comprise a 5' untranslated region which specifically binds the retrotransposase.

Reverse transcription by non-LTR retrotransposons occurs via a unique process described as target-primed reverse transcription (Luan et al. Cell 72, 595-605 (1993)). To initiate the integration, a first single-stranded nick is generated by an endonuclease domain of the retrotransposase, releasing a free 3'-OH. The retrotransposon RNA, bound by the retrotransposase using structural features at the 3' end, is then primed by the target site with polymerization at the free 3'-OH and used as a template for reverse transcription. In some systems, a second nick is targeted to the second DNA strand and the new free 3'-OH is used to initiate second strand synthesis. Some non-LTR retrotransposons, e.g., R2, are believed to additionally require interaction with a second retrotransposase unit at the 5' end of the retrotransposon RNA for this second nick, which is activated upon the release of the 5' end (Craig, Mobile DNA III, ASM, ed. 3 (2105)).

As described herein, the elements of such non-LTR retrotransposons can be functionally modularized and/or modified to target, edit, modify or manipulate a target DNA sequence, e.g., to insert an object (e.g., heterologous) nucleic acid sequence into a target genome, e.g., a mammalian genome, by reverse transcription. Such modularized and modified nucleic acids,

polypeptide compositions and systems are described herein and are referred to as Gene Writer™ gene editors. A Gene Writer™ gene editor system comprises: (A) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain, and either (x) an endonuclease domain that contains DNA binding functionality or (y) an endonuclease domain and separate DNA binding domain; and (B) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous insert sequence. For example, the Gene Writer genome editor protein may comprise a DNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. In other embodiments, the Gene Writer genome editor protein may comprise a reverse transcriptase domain and an endonuclease domain. In certain embodiments, the elements of the Gene Writer™ gene editor polypeptide can be derived from sequences of non-LTR retrotransposons, e.g., APE-type or RLE-type retrotransposons or portions or domains thereof. In some embodiments the RLE-type non-LTR retrotransposon is from the R2, NeSL, HERO, R4, or CRE clade. In some embodiments the Gene Writer genome editor is derived from R4 element X4_Line, which is found in the human genome. In some embodiments the APE-type non-LTR retrotransposon is from the R1, or Tx1 clade. In some embodiments the Gene Writer genome editor is derived from Tx1 element Mare6, which is found in the human genome. The RNA template element of a Gene Writer™ gene editor system is typically heterologous to the polypeptide element and provides an object sequence to be inserted (reverse transcribed) into the host genome. In some embodiments the Gene Writer genome editor protein is capable of target primed reverse transcription. In some embodiments, the Gene Writer genome editor protein is capable of second strand synthesis.

In some embodiments the Gene Writer genome editor is combined with a second polypeptide. In some embodiments the second polypeptide is derived from an APE-type non-LTR retrotransposon. In some embodiments the second polypeptide has a zinc knuckle-like motif. In some embodiments the second polypeptide is a homolog of Gag proteins. In some embodiments, the second polypeptide possesses specific binding activity for the RNA template. In some embodiments, the second polypeptide aids in localization of the RNA template to the nucleus.

In embodiments, the disclosure provides a nucleic acid molecule or a system for retargeting, e.g., of a Gene Writer polypeptide or nucleic acid molecule, or of a system as described herein. Retargeting (e.g., of a Gene Writer polypeptide or nucleic acid molecule, or of

a system as described herein) generally comprises : (i) directing the polypeptide to bind and cleave at the target site; and/or (ii) designing the template RNA to have complementarity to the target sequence. In some embodiments, the template RNA has complementarity to the target sequence 5' of the first-strand nick, e.g., such that the 3' end of the template RNA anneals and the 5' end of the target site serves as the primer, e.g., for target-primed reverse transcription (TPRT). In some embodiments, the endonuclease domain of the polypeptide and the 5' end of the RNA template are also modified as described.

Polypeptide component of Gene Writer gene editor system

RT domain:

In certain aspects of the present invention, the reverse transcriptase domain of the Gene Writer system is based on a reverse transcriptase domain of an APE-type or RLE-type non-LTR retrotransposon. A wild-type reverse transcriptase domain of an APE-type or RLE-type non-LTR retrotransposon can be used in a Gene Writer system or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) to alter the reverse transcriptase activity for target DNA sequences. In some embodiments the reverse transcriptase is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the reverse transcriptase domain is a heterologous reverse transcriptase from a different retrovirus, LTR-retrotransposon, or non-LTR retrotransposon. In certain embodiments, a Gene Writer system includes a polypeptide that comprises a reverse transcriptase domain of an RLE-type non-LTR retrotransposon from the R2, NeSL, HERO, R4, or CRE clade, or of an APE-type non-LTR retrotransposon from the R1, or Tx1 clade. In certain embodiments, a Gene Writer™ system includes a polypeptide that comprises a reverse transcriptase domain of a non-LTR retrotransposon, LTR retrotransposon, group II intron, diversity-generating element, retron, telomerase, retroplasmid, retrovirus, or an engineered polymerase listed in Table 1 or Table 3. In some embodiments, a Gene Writer™ system includes a polypeptide that comprises a reverse transcriptase domain listed in Table 2. In embodiments, the amino acid sequence of the reverse transcriptase domain of a Gene Writer system is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of a reverse transcriptase domain of a non-LTR retrotransposon, LTR retrotransposon, group II intron, diversity-generating element, retron, telomerase, retroplasmid,

retrovirus, or an engineered polymerase whose sequence is referenced in Table 1 or Table 3, or to a peptide comprising a reverse transcriptase domain listed in Table 2. In some embodiments, the RT domain has a sequence selected from Table 1 or 3, or a sequence of a peptide comprising an RT domain selected from Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the RT domain is derived from the RT of a retrovirus, e.g., HIV-1 RT, Moloney Murine Leukemia Virus (MMLV) RT, avian myeloblastosis virus (AMV) RT, Rous Sarcoma Virus (RSV) RT. In some embodiments, the RT domain is derived from the RT of a Group II intron, e.g., the group II intron maturase RT from *Eubacterium rectale* (MarathonRT) (Zhao et al. *RNA* 24:2 2018), the RT domain from LtrA, the RT TGIRT (or trt). In some embodiments, the RT domain is derived from the RT of a retron, e.g., the reverse transcriptase from Ec86 (RT86). In some embodiments, the RT domain is derived from a diversity-generating retroelement, e.g., from the RT of Brt. In some embodiments, the RT domain is derived from the RT of a retroplasmid, e.g., the RT from the Mauriceville plasmid. In some embodiments, the RT domain is derived from a non-LTR retrotransposon, e.g., the RT from R2Bm, the RT from R2Tg, the RT from LINE-1, the RT from Penelope or a Penelope-like element (PLE). In some embodiments, the RT domain is derived from an LTR retrotransposon, e.g., the reverse transcriptase from Ty1. In some embodiments, the RT domain is derived from a telomerase, e.g., TERT. A person having ordinary skill in the art is capable of identifying reverse transcription domains based upon homology to other known reverse transcription domains using routine tools as Basic Local Alignment Search Tool (BLAST). In some embodiments, the reverse transcriptase contains the InterPro domain IPR000477. In some embodiments, the reverse transcriptase contains the pfam domain PF00078. In some embodiments, the reverse transcriptase contains the InterPro domain IPR013103. In some embodiments, the RT contains the pfam domain PF07727. In some embodiments, the reverse transcriptase contains a conserved protein domain of the cd00304 RT_like family, e.g., cd01644 (RT_pepA17), cd01645 (RT_Rtv), cd01646 (RT_Bac_retron_I), cd01647 (RT_LTR), cd01648 (TERT), cd01650 (RT_nLTR_like), cd01651 (RT_G2_intron), cd01699 (RNA_dep_RNAP), cd01709 (RT_like_1), cd03487 (RT_Bac_retron_II), cd03714 (RT_DIRS1), cd03715 (RT_ZFREV_like). Proteins containing these domains can additionally be found by searching the domains on protein databases, such as InterPro (Mitchell et al. Nucleic Acids Res 47, D351-360 (2019)), UniProt (The UniProt Consortium Nucleic Acids Res 47, D506-515

(2019)), or the conserved domain database (Lu et al. Nucleic Acids Res 48, D265-268 (2020)), or by scanning open reading frames for reverse transcriptase domains using prediction tools, for example InterProScan. The diversity of reverse transcriptases (e.g., comprising RT domains) has been described in, but not limited to, those used by prokaryotes (Zimmerly et al. *Microbiol Spectr* 3(2):MDNA3-0058-2014 (2015); Lampson B.C. (2007) Prokaryotic Reverse Transcriptases. In: Polaina J., MacCabe A.P. (eds) Industrial Enzymes. Springer, Dordrecht), viruses (Herschhorn et al. *Cell Mol Life Sci* 67(16):2717-2747 (2010); Menéndez-Arias et al. *Virus Res* 234:153-176 (2017)), and mobile elements (Eickbush et al. *Virus Res* 134(1-2):221-234 (2008); Craig et al. *Mobile DNA III* 3rd Ed. DOI:10.1128/9781555819217 (2015)), each of which is incorporated herein by reference.

In some embodiments, the RT domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety).

In some embodiments, the RT domain forms a dimer (e.g., a heterodimer or homodimer). In some embodiments, the RT domain is monomeric. In some embodiments, an RT domain, e.g., a retroviral RT domain, naturally functions as a monomer or as a dimer (e.g., heterodimer or homodimer). In some embodiments, an RT domain naturally functions as a monomer, e.g., is derived from a virus wherein it functions as a monomer. Exemplary monomeric RT domains, their viral sources, and the RT signatures associated with them can be found in Table 30 with

descriptions of domain signatures in Table 32. In some embodiments, the RT domain of a system described herein comprises an amino acid sequence of Table 30, or a functional fragment or variant thereof, or a sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto. In embodiments, the RT domain is selected from an RT domain from murine leukemia virus (MLV; sometimes referred to as MoMLV) (e.g., P03355), porcine endogenous retrovirus (PERV) (e.g., UniProt Q4VFZ2), mouse mammary tumor virus (MMTV) (e.g., UniProt P03365), Mason-Pfizer monkey virus (MPMV) (e.g., UniProt P07572), bovine leukemia virus (BLV) (e.g., UniProt P03361), human T-cell leukemia virus-1 (HTLV-1) (e.g., UniProt P03362), human foamy virus (HFV) (e.g., UniProt P14350), simian foamy virus (SFV) (e.g., UniProt P23074), or bovine foamy/syncytial virus (BFV/BSV) (e.g., UniProt O41894), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). In some embodiments, an RT domain is dimeric in its natural functioning. Exemplary dimeric RT domains, their viral sources, and the RT signatures associated with them can be found in Table 31 with descriptions of domain signatures in Table 32. In some embodiments, the RT domain of a system described herein comprises an amino acid sequence of Table 31, or a functional fragment or variant thereof, or a sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain is derived from a virus wherein it functions as a dimer. In embodiments, the RT domain is selected from an RT domain from avian sarcoma/leukemia virus (ASLV) (e.g., UniProt A0A142BKH1), Rous sarcoma virus (RSV) (e.g., UniProt P03354), avian myeloblastosis virus (AMV) (e.g., UniProt Q83133), human immunodeficiency virus type I (HIV-1) (e.g., UniProt P03369), human immunodeficiency virus type II (HIV-2) (e.g., UniProt P15833), simian immunodeficiency virus (SIV) (e.g., UniProt P05896), bovine immunodeficiency virus (BIV) (e.g., UniProt P19560), equine infectious anemia virus (EIAV) (e.g., UniProt P03371), or feline immunodeficiency virus (FIV) (e.g., UniProt P16088) (Herschhorn and Hizi Cell Mol Life Sci 67(16):2717-2747 (2010)), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). Naturally heterodimeric RT domains may, in some embodiments, also be functional as homodimers. In some embodiments, dimeric RT domains are expressed as fusion proteins, e.g., as homodimeric fusion proteins or heterodimeric fusion proteins. In some embodiments, the RT function of the system is fulfilled by multiple RT

domains (e.g., as described herein). In further embodiments, the multiple RT domains are fused or separate, e.g., may be on the same polypeptide or on different polypeptides.

In some embodiment, a GeneWriter described herein comprises an integrase domain, e.g., wherein the integrase domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an integrase domain. In some embodiments, an RT domain (e.g., as described herein) lacks an integrase domain, or comprises an integrase domain that has been inactivated by mutation or deleted. In some embodiment, a GeneWriter described herein comprises an RNase H domain, e.g., wherein the RNase H domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain, e.g., an endogenous RNAse H domain or a heterologous RNase H domain. In some embodiments, an RT domain (e.g., as described herein) lacks an RNase H domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain that has been added, deleted, mutated, or swapped for a heterologous RNase H domain. In some embodiments, mutation of an RNase H domain yields a polypeptide exhibiting lower RNase activity, e.g., as determined by the methods described in Kotewicz et al. Nucleic Acids Res 16(1):265-277 (1988) (incorporated herein by reference in its entirety), e.g., lower by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to an otherwise similar domain without the mutation. In some embodiments, RNase H activity is abolished.

In some embodiments, an RT domain is mutated to increase fidelity compared to to an otherwise similar domain without the mutation. For instance, in some embodiments, a YADD (SEQ ID NO: 1547) or YMDD (SEQ ID NO: 1548) motif in an RT domain (e.g., in a reverse transcriptase) is replaced with YVDD (SEQ ID NO: 1549). In embodiments, replacement of the YADD (SEQ ID NO: 1547) or YMDD (SEQ ID NO: 1548) or YVDD (SEQ ID NO: 1549) results in higher fidelity in retroviral reverse transcriptase activity (e.g., as described in Jamburuthugoda and Eickbush J Mol Biol 2011; incorporated herein by reference in its entirety).

In some embodiments, reverse transcriptase domains are modified, for example by site-specific mutation. In some embodiments, reverse transcriptase domains comprise a number of amino acid substitutions relative to the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions. In embodiments, the reverse transcriptase domain is engineered to bind a heterologous template RNA.

Table 1: Exemplary reverse transcriptase domains from different types of sources.

Sources include Group II intron, non-LTR retrotransposon, retrovirus, LTR retrotransposon, diversity-generating retroelement, retron, telomerase, retroplasmid, and evolved DNA polymerase. Also included are the associated RT signatures from the InterPro, pfam, and cd databases. Although the evolved polymerase RTX can perform RNA-dependent DNA polymerization, no RT signatures were identified by InterProScan, so polymerase signatures are included instead.

Protein	Type	Accession	UniProt	Sequence	RT signatures
Marathon RT	Group II intron	CBK92290.1	D4JMT6	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC (SEQ ID NO: 1550)	IPR000477, PF00078, cd01651
TGIRT, trt	Group II intron	AAT72329.1	Q6DKY2	MALLERILADRNLITALKRVEANQGAPGIGDVSTDQLRDIYRAHWSTIRAQLLAGTYRPAPVRRVGIPKGPGGTRQLGITPVVDRLIQQIALQELTPIFDPDFSPSSFGFRPGRNAHDAVRQAQGYIQEYGRYVVDMDLKEFFDRVNHDLIMSRVARKVDKKRVLKLIRYALQAGVMIEGVKVQTEEGTQPGGPLSPLLANILLDDLDKELEKRGLKFCYRADDCNIYVSKLRAGQRVKQSIQRFLEKTLKLKVNEEKSVADRPWKRAFGLFSFTPERKARIRLAPRSIQRLKQRIRQLTNPNWSISMPREIHRVNQYVGMWIGYFRLVTEPSVLQTIEGWI	IPR000477, PF00078, cd01651

				RRRLRLCWQLQWKRVRTRIREL RALGLKETAVMEIANRTKGAWR TTKPQTLHQALGKYTWTAQGLK TSLQRYFELRQG (SEQ ID NO: 1551)	
LtrA	Group II intron	AAB06503.1	P0A3U0	MKPTMAILERISKNSQENIDEVFT RLYRYLLRPDIYYVAYQNLYSN KGASTKGILDDTADGFSEEKIKKI IQSLKDGTYYPQPVRRMYIAKKN SKKMRPLGIPTFTDKLIQEAVRIIL ESIYEPVFEDVSHGFRPQRSCHTA LKTIKREFGGARWFVEGDIKGCF DNIDHVTLIGLINLKIKDMKMSQ LIYKFLKAGYLENWQYHKTYSG TPQGGILSPLLANIYLHELDKFVL QLKMKFDRESPERITPEYRELHN EIKRISHRLKKLEGEEKAKVLLE YQEKRKRLPTLPCTSQTNKVLKY VRYADDFIISVKGSKEDCQWIKE QLKLFIHNKLKMELSEEKTLITHS SQPARFLGYDIRVRRSGTIKRSG KVKKRTLNGSVELLIPLQDKIRQ FIFDKKIAIQKKDSSWFPVHRKYL IRSTDLEIITIYNSELRGICNYYGL ASNFNQLNYFAYLMEYSCLKTIA SKHKGTLSKTISMFKDGSGSWGI PYEIKQGKQRRYFANFSECKSPY QFTDEISQAPVLYGYARNTLENR LKAKCCELCGTSDENTSYEIHHV NKVKNLKGKEKWEMAMIAKQR KTLVVCFHCHRHVIHKHK (SEQ ID NO: 1552)	IPR000477, PF00078, cd01651
R2Bm	non-LTR retrotransposon	AAB59214.1	V9H052	MMASTALSLMGRCNPDGCTRGK HVTAAPMDGPRGPSSLAGTFGW GLAIPAGEPCGRVCSPATVGFFP VAKKSNKENRPEASGLPLESERT GDNPTVRGSAGADPVGQDAPG WTCQFCERTFSTNRGLGVHKRR AHPVETNTDAAPMMVKRRWHG EEIDLLARTEARLLAERGQCSGG DLFGALPGFGRTLEAIKGQRRRE PYRALVQAHLARFGSQPGPSSGG CSAEPDFRRASGAEEAGEERCAE DAAAYDPSAVGQMSPDAARVLS ELLEGAGRRRACRAMRPKTAGR RNDLHDDRTASAHKTSRQKRRA	IPR000477, PF00078, cd01650

				EYARVQELYKKCRSRAAAEVID GACGGVGHSLEEMETYWRPILE RVSDAPGPTPEALHALGRAEWH GGNRDYTQLWKPISVEEIKASRF DWRTSPGPDGIRSGQWRAVPVH LKAEMFNAWMARGEIPEILRQC RTVFVPKVERPGGPGEYRPISIAS IPLRHFHSILARRLLACCPPDARQ RGFICADGTLENSAVLDAVLGDS RKKLRECHVAVLDFAKAFDTVS HEALVELLRLRGMPEQFCGYIAH LYDTASTTLAVNNEMSSPVKVG RGVRQGDPLSPILFNVVMDLILA SLPERVGYRLEMELVSALAYAD DLVLLAGSKVGMQESISAVDCV GRQMGLRLNCRKSAVLSMIPDG HRKKHHYLTERTFNIGGKPLRQV SCVERWRYLGVDFEASGCVTLE HSISSALNNISRAPLKPQQRLEILR AHLIPRFQHGFVLGNISDDRLRM LDVQIRKAVGQWLRLPADVPKA YYHAAVQDGGLAIPSVRATIPDL IVRRFGGLDSSPWSVARAAAKSD KIRKKLRWAWKQLRRFSRVDST TQRPSVRLFWREHLHASVDGRE LRESTRTPTSTKWIRERCAQITGR DFVQFVHTHINALPSRIRGSRGR RGGGESSLTCRAGCKVRETTAHI LQQCHRTHGGRILRHNKIVSFVA KAMEENKWTVELEPRLRTSVGL RKPDIIASRDGVGVIVDVQVVSG QRSLDELHREKRNKYGNHGELV ELVAGRLGLPKAECVRATSCTIS WRGVWSLTSYKELRSIIGLREPT LQIVPILALRGSHMNWTRFNQMT SVMGGGVG (SEQ ID NO: 1553)	
LINE-1	non-LTR retrotransposon	AAC51271.1	O00370	MTGSNSHITILTLNVNGLNSPIKR HRLASWIKSQDPSVCCIQETHLT CRDTHRLKIKGWRKIYQANGKQ KKAGVAILVSDKTDFKPTKIKRD KEGHYIMVKGSIQQEELTILNIYA PNTGAPRFIKQVLSDLQRDLDSH TLIMGDFNTPLSILDRSTRQKVN KDTQELNSALHQTDLIDIYRTLH PKSTEYTFFSAPHHTYSKIDHIVG	IPR000477, PF00078, cd01650

				SKALLSKCKRTEIITNYLSDHSAI KLELRIKNLTQSRSTTWKLNNLL LNDYWVHNEMKAEIKMFFETNE NKDTTYQNLWDAFKAVCRGKFI ALNAYKRKQERSKIDTLTSQLKE LEKQEQTHSKASRRQEITKIRAEL KEIETQKTLQKINESRSWFFERIN KIDRPLARLIKKKREKNQIDTIKN DKGDITTDPTEIQTTIREYYKHLY ANKLENLEEMDTFLDTYTLPRLN QEEVESLNRPITGSEIVAIINSLPT KKSPGPDGFTAEFYQRYKEELVP FLLKLFQSIEKEGILPNSFYEASIIL IPKPGRDTTKKENFRPISLMNIDA KILNKILANRIQQHIKKLIHHDQV GFIPGMQGWFNIRKSINVIQHINR AKDKNHVIISIDAEKAFDKIQQPF MLKTLNKLGIDGMYLKIIRAIYD KPTANIILNGQKLEAFPLKTGTRQ GCPLSPLLFNIVLEVLARAIRQEK EIKGIQLGKEEVKLSLFADDMIV YLENPIVSAQNLLKLISNFSKVSG YKINVQKSQAFLYNNNRQTESQI MGELPFTIASKRIKYLGIQLTRDV KDLFKENYKPLLKEIKEDTNKW KNIPCSWVGRINIVKMAILPKVIY RFNAIPIKLPMTFFTELEKTTLKFI WNQKRARIAKSILSQKNKAGGIT LPDFKLYYKATVTKTAWYWYQ NRDIDQWNRTEPSEIMPHIYNYLI FDKPEKNKQWGKDSLLNKWCW ENWLAICRKLKLDPFLTPYTKINS RWIKDLNVKPKTIKTLEENLGITI QDIGVGKDFMSKTPKAMATKDK IDKWDLIKLKSFCTAKETTIRVNR QPTTWEKIFATYSSDKGLISRIYN ELKQIYKKKTNNPIKKWAKDMN RHFSKEDIYAAKKHMKKCSSSLA IREMQIKTTMRYHLTPVRMAIIK KSGNNRCWRGCGEIGTLVHCW WDCKLVQPLWKSVWRFLRDLEL EIPFDPAIPLLGIYPKDYKSCCYK DTCTRMFIAALFTIAKTWNQPNC PTMIDWIKKMWHIYTMEYYAAI KNDEFISFVGTWMKLETIILSKLS	

				QEQKTKHRIFSLIGGN (SEQ ID NO: 1554)	
Penelope	non-LTR retrotransposon	AAL14979.1	Q95VB5	MERSPEPSININGRHAVCTATNMSYAKIKTKYKDSKRTINKFQLTLVKLTKLKSSLKFLLKCRKSNLIPNFIKNLTQHLTILTTDNKTHPDITRTLTRHTHFYHTKILNLLIKHKHNLLQEQTKHMQKAKTNIEQLMTTDDAKAFFESERNIENKITTTLKKRQETKHDKLRDQRNLALADNNTQREWFVNKTKIEFPPNVVALLAKGPKFALPISKRDFPLLKYIADGEELVQTIKEKETQESARTKFSLLVKEHKTKNNQNSRDRAILDTVEQTRKLLKENINIKILSSDKGNKTVAMDEDEYKNKMTNILDDLCAYRTLRLDPTSRLQTKNNTFVAQLFKMGLISKDERNKMTTTTAVPPRIYGLPKIHKEGTPLRPICSSIGSPSYGLCKYIIQILKNLTMDSRYNIKNAVDFKDRVNNSQIREEETLVSFDVVSLFPSIPIELALDTIRQKWTKLEEHTNIPKQLFMDIVRFCIEENRYFKYEDKIYTQLKGMPMGSPASPVIADILMEELLDKITDKLKIKPRLLTKYVDDLFAITNKIDVENILKELNSFHKQIKFTMELEKDGKLPFLDSIVSRMDNTLKIKWYRKPIASGRILNFNSNHPKSMIINTALGCMNRMMKISDTIYHKEIEHEIKELLTKNDFPPNIIKTLLKRRQIERKKPTEPAKIYKSLIYVPRLSERLTNSDCYNKQDIKVAHKPTNTLQKFFNKIKSKIPMIEKSNVVYQIPCGGDNNNKCNSVYIGTTKSKLKTRISQHKSDFKLRHQNNIQKTALMTHCIRSNHTPNFDETTILQQEQHYNKRHTLEMLHIINTPTYKRLNYKTDTENCAHLYRHLLNSQTTSVTISTSKSADV (SEQ ID NO: 1555)	IPR000477, PF00078, cd00304
M-MLV RT	Retrovirus	ADS42990.1	P03355[660-1330]	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG	IPR000477, PF00078, cd03715

				LPPSHQWYTVLDLKDAFFCLRLH PTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDL ADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLK EGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAE MAAPLYPLTKTGTLFNWGPDQQ KAYQEIKQALLTAPALGLPDLTK PFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQ PLVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTT ETEVIWAKALPAGTSAQRAELIA LTQALKMAEGKKLNVYTDSRYA FATAHIHGEIYRRRGLLTSEGKEI KNKDEILALLKALFLPKRLSIIHC PGHQKGHSAEARGNRMADQAA RKAAITETPDTSTLL (SEQ ID NO: 1556)	
RSV RT	Retro virus	AAC82561.1	P03354[709-1567]	TVALHLAIPLKWKPDHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFAFTLPSVNNQAP ARRFQWKVLPQGMTCSPTICQL VVGQVLEPLRLKHPSLCMLHYM DDLLLAASSHDGLEAAGEEVIST LERAGFTISPDKVQREPGVQYLG YKLGSTYVAPVGLVAEPRIATLW DVQKLVGSLQWLRPALGIPPRL MGPFYEQLRGSDPNEAREWNLD MKMAWREIVRLSTTAALERWDP ALPLEGAVARCEQGAIGVLGQG LSTHPRPCLWLFSTQPTKAFTAW LEVLTLLITKLRASAVRTFGKEV DILLLPACFREDLPLPEGILLALK GFAGKIRSSDTPSIFDIARPLHVSL KVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGA	IPR000477, PF00078, cd01645

				SVQQLEARAVAMALLLWPTTPT NVVTDSAFVAKMLLKMGQEGV PSTAAAFILEDALSQRSAMAAVL HVRSHSEVPGFFTEGNDVADSQA TFQAYPLREAKDLHTALHIGPRA LSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQT DFTLEPRMAPRSWLAVTVDTASS AIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTR EWLARWGIAHTTGIPGNSQGQA MVERANRLLKDRIRVLAEGDGF MKRIPTSKQGELLAKAMYALNH FERGENTKTPIQKHWRPTVLTEG PPVKIRIETGEWEKGWNVLVWG RGYAAVKNRDTDKVIWVPSRKV KPDITQKDEVTKKDEASPLFAG (SEQ ID NO: 1557)	
AMV RT	Retro virus	HW60668 0.1	-	TVALHLAIPLKWKPNHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFAFTLPSVNNQAP ARRFQWKVLPQGMTCSPTICQLI VGQILEPLRLKHPSLRMLHYMD DLLLAASSHDGLEAAGEEVISTL ERAGFTISPDKVQREPGVQYLGY KLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPPRLM GPFYEQLRGSDPNEAREWNLDM KMAWREIVQLSTTAALERWDPA LPLEGAVARCEQGAIGVLGQGLS THPRPCLWLFSTQPTKAFTAWLE VLTLLITKLRASAVRTFGKEVDIL LLPACFREDLPLPEGILLALRGFA GKIRSSDTPSIFDIARPLHVSLKV RVTDHPVPGPTVFTDASSSTHKG VVVWREGPRWEIKEIADLGASV QQLEARAVAMALLLWPTTPTNV VTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHV RSHSEVPGFFTEGNDVADSQATF QAY (SEQ ID NO: 1558)	IPR000477 , PF00078, cd01645

HIV RT	Retrovirus	AAB50259.1	P04585[588-1147]	PISPIETVPVKLKPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKI SKIGPENPYNTPVFAIKKKDSTK WRKLVDFRELNKRTQDFWEVQL GIPHPAGLKKKKSVTVLDVGDA YFSVPLDEDFRKYTAFTIPSINNE TPGIRYQYNVLPQGWKGSPAIFQ SSMTKILEPFRKQNPDIVIYQYM DDLYVGSDLEIGQHRTKIEELRQ HLLRWGLTTPDKKHQKEPPFLW MGYELHPDKWTVQPIVLPEKDS WTVNDIQKLVGKLNWASQIYPGI KVRQLCKLLRGTKALTEVIPLTE EAELELAENREILKEPVHGVYYD PSKDLIAEIQKQGQGQWTYQIYQ EPFKNLKTGKYARMRGAHTNDV KQLTEAVQKITTESIVIWGKTPKF KLPIQKETWETWWTEYWQATWI PEWEFVNTPPLVKLWYQLEKEPI VGAETFYVDGAANRETKLGKAG YVTNRGRQKVVTLTDTTNQKTE LQAIYLALQDSGLEVNIVTDSQY ALGIIQAQPDQSESELVNQIIEQLI KKEKVYLAWVPAHKGIGGNEQV DKLVSAGIRKVL (SEQ ID NO: 1559)	IPR000477, PF00078, cd01645
Ty1	LTR retrotransposon	AAA66938.1	Q07163-1[1218-1755]	AVKAVKSIKPIRTTLRYDEAITYN KDIKEKEKYIEAYHKEVNQLLK MKTWDTDEYYDRKEIDPKRVIN SMFIFNKKRDGTHKARFVARGDI QHPDTYDSGMQSNTVHHYALM TSLSLALDNNYYITQLDISSAYLY ADIKEELYIRPPPHLGMNDKLIRL KKSLYGLKQSGANWYETIKSYLI QQCGMEEVRGWSCVFKNSQVTI CLFVDDMVLFSKNLNSNKRIIEK LKMQYDTKIINLGESDEEIQYDIL GLEIKYQRGKYMKLGMENSLTE KIPKLNVPLNPKGRKLSAPGQPG LYIDQDELEIDEDEYKEKVHEMQ KLIGLASYVGYKFRFDLLYYINT LAQHILFPSRQVLDMTYELIQFM WDTRDKQLIWHKNKPTEPDNKL VAISDASYGNQPYYKSQIGNIYL LNGKVIGGKSTKASLTCTSTTEA EIHAISESVPLLNNLSYLIQELNK	IPR013103, PF07727

				KPIIKGLLTDSRSTISIIKSTNEEKF RNRFFGTKAMRLRDEVSGNNLY VYYIETKKNIADVMTKPLPIKTF KLLTNKWIH (SEQ ID NO: 1560)	
Brt	Diversity-generating retroelement	NP_958675.1	Q775D8	MGKRHRNLIDQITTWENLLDAY RKTSHGKRRTWGYLEFKEYDLA NLLALQAELKAGNYERGPYREF LVYEPKPRLISALEFKDRLVQHA LCNIVAPIFEAGLLPYTYACRPDK GTHAGVCHVQAELRRTRATHFL KSDFSKFFPSIDRAALYAMIDKKI HCAATRRLLRVVLPDEGVGIPIG SLTSQLFANVYGGAVDRLLHDE LKQRHWARYMDDIVVLGDDPEE LRAVFYRLRDFASERLGLKISHW QVAPVSRGINFLGYRIWPTHKLL RKSSVKRAKRKVANFIKHGEDES LQRFLASWSGHAQWADTHNLFT WMEEQYGIACH (SEQ ID NO: 1561)	IPR000477, PF00078, cd01646
RT86	Retron	AAA61471.1	P23070	MKSAEYLNTFRLRNLGLPVMNN LHDMSKATRISVETLRLLIYTADF RYRIYTVEKKGPEKRMRTIYQPS RELKALQGWVLRNILDKLSSSPF SIGFEKHQSILNNATPHIGANFILN IDLEDFFPSLTANKVFGVFHSLGY NRLISSVLTKICCYKNLLPQGAPS SPKLANLICSKLDYRIQGYAGSR GLIYTRYADDLTLSAQSMKKVV KARDFLFSIIPSEGLVINSKKTCIS GPRSQRKVTGLVISQEKVGIGRE KYKEIRAKIHHIFCGKSSEIEHVR GWLSFILSVDSKSHRRLITYISKL EKKYGKNPLNKAKT (SEQ ID NO: 1562)	IPR000477, PF00078, cd03487
TERT	Telomerase	AAG23289.1	O14746	MPRAPRCRAVRSLLRSHYREVLP LATFVRRLGPQGWRLVQRGDPA AFRALVAQCLVCVPWDARPPPA APSFRQVSCLKELVARVLQRLCE RGAKNVLAFGFALLDGARGGPP EAFTTSVRSYLPNTVTDALRGSG AWGLLLRRVGDDVLVHLLARCA LFVLVAPSCAYQVCGPPLYQLGA ATQARPPPHASGPRRRLGCERA WNHSVREAGVPLGLPAPGARRR GGSASRSLPLPKRPRRGAAPEPE	IPR000477, PF00078, cd01648

				RTPVGQGSWAHPGRTRGPSDRG FCVVSPARPAEEATSLEGALSGT RHSHPSVGRQHHAGPPSTSRPPR PWDTPCPPVYAETKHFLYSSGDK EQLRPSFLLSSLRPSLTGARRLVE TIFLGSRPWMPGTPRRLPRLPQR YWQMRPLFLELLGNHAQCPYGV LLKTHCPLRAAVTPAAGVCARE KPQGSVAAPEEEDTDPRRLVQLL RQHSSPWQVYGFVRACLRRLVP PGLWGSRHNERRFLRNTKKFISL GKHAKLSLQELTWKMSVRDCA WLRRSPGVGCVPAAEHRLREEIL AKFLHWLMSVYVVELLRSFFYV TETTFQKNRLFFYRKSVWSKLQS IGIRQHLKRVQLRELSEAEVRQH REARPALLTSRLRFIPKPDGLRPI VNMDYVVGARTFRREKRAERLT SRVKALFSVLNYERARRPGLLGA SVLGLDDIHRAWRTFVLRVRAQ DPPPELYFVKVDVTGAYDTIPQD RLTEVIASIIKPQNTYCVRRYAVV QKAAHGHVRKAFKSHVSTLTDL QPYMRQFVAHLQETSPLRDAVVI EQSSSLNEASSGLFDVFLRFMCH HAVRIRGKSYVQCQGIPQGSILST LLCSLCYGDMENKLFAGIRRDGL LLRLVDDFLLVTPHLTHAKTFLR TLVRGVPEYGCVVNLRKTVVNF PVEDEALGGTAFVQMPAHGLFP WCGLLLDTRTLEVQSDYSSYAR TSIRASLTFNRGFKAGRNMRRKL FGVLRLKCHSLFLDLQVNSLQTV CTNIYKILLLQAYRFHACVLQLPF HQQVWKNPTFFLRVISDTASLCY SILKAKNAGMSLGAKGAAGPLPS EAVQWLCHQAFLLKLTRHRVTY VPLLGSLRTAQTQLSRKLPGTTL TALEAAANPALPSDFKTILD (SEQ ID NO: 1563)	
Mauric eville RT	Retro plasm id	NC_0015 70.1	Q36578	MPNHRLPNCVSYLGENHELSWL HGMFGLLKRSNPQTGGILGWLN TGPNGFVKYMMNLMGHARDKG DAKEYWRLGRSLMKNEAFQVQ AFNHVCKHWYLDYKPHKIAKLL KEVREMVEIQPVCIDYKRVYIPK	cd00304

				ANGKQRPLGVPTVPWRVYLHM WNVLLVWYRIPEQDNQHAYFPK RGVFTAWRALWPKLDSQNIYEF DLKNFFPSVDLAYLKDKLMESGI PQDISEYLTVLNRSLVVLTSEDKI PEPHRDVIFNSDGTPNPNLPKDV QGRILKDPDFVEILRRRGFTDIAT NGVPQGASTSCGLATYNVKELF KRYDELIMYADDGILCRQDPSTP DFSVEEAGVVQEPAKSGWIKQN GEFKKSVKFLGLEFIPANIPPLGE GEVKDYPRLRGATRNGSKMELS TELQFLCYLSYKLRIKVLRDLYIQ VLGYLPSVPLLRYRSLAEAINELS PKRITIGQFITSSFEEFTAWSPLKR MGFFFSSPAGPTILSSIFNNSTNLQ EPSDSRLLYRKGSWVNIRFAAYL YSKLSEEKHGLVPKFLEKLREINF ALDKVDVTEIDSKLSRLMKFSVS AAYDEVGTLALKSLFKFRNSERE SIKASFKQLRENGKIAEFSEARRL WFEILKLIRLDLFNASSLACDDLL SHLQDRRSIKKWGSSDVLYLKSQ RLMRTNKKQLQLDFEKKKNSLK KKLIKRRAKELRDTFKGKENKEA (SEQ ID NO: 1564)	
RTX	Engineered polymerase	QFN49000.1	-	MILDTDYITEDGKPVIRIFKKENG EFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHGTVVTVKRVEK VQKKFLGRPVEVWKLYFTHPQD VPAIMDKIREHPAVIDIYEYDIPF AIRYLIDKGLVPMEGDEELKLLA FDIETLYHEGEEFAEGPILMISYA DEEGARVITWKNVDLPYVDVVS TEREMIKRFLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINF ALGRDGSEPKIQRMGDRFAVEV KGRIHFDLYPVIRRTINLPTYTLE AVYEAVFGQPKEKVYAEEITTA WETGENLERVARYSMEDAKVTY ELGKEFLPMEAQLSRLIGQSLWD VSRSSTGNLVEWFLLRKAYERNE LAPNKPDEKELARRHQSHEGGYI KEPERGLWENIVYLDFRSLYPSIII THNVSPDTLNREGCKEYDVAPQ VGHRFCKDFPGFIPSLLGDLLEER	IPR006134, PF00136, cd05536

				QKIKKRMKATIDPIERKLLDYRQ RAIKILANSLYGYYGYARARWY CKECAESVIAWGREYLTMTIKEI EEKYGFKVIYSDTDGFFATIPGA DAETVKKKAMEFLKYINAKLPG ALELEYEGFYKRGLFVTKKKYA VIDEEGKITTRGLEIVRRDWSEIA KETQARVLEALLKDGDVEKAVR IVKEVTEKLSKYEVPPEKLVIHK QITRDLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKGSGRI VDRAIPFDEFDPTKHKYDAEYYI EKQVLPAVERILRAFGYRKEDLR YQKTRQVGLSARLKPKGTLEGSS HHHHHH (SEQ ID NO: 1565)	

Table 2: InterPro descriptions of signatures present in reverse transcriptases in Table 1.

Signature	Database	Short Name	Description
cd00304	CDD	*RT_like*	RT_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)_like family. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]
cd01645	CDD	*RT_Rtv*	RT_Rtv: Reverse transcriptases (RTs) from retroviruses (Rtvs). RTs catalyze the conversion of single-stranded RNA into double-stranded viral DNA for integration into host chromosomes. Proteins in this subfamily contain long terminal repeats (LTRs) and are multifunctional enzymes with RNA-directed DNA polymerase, DNA directed DNA polymerase, and ribonuclease hybrid (RNase H) activities. The viral RNA genome enters the cytoplasm as part of a nucleoprotein complex, and the process of

			reverse transcription generates in the cytoplasm forming a linear DNA duplex via an intricate series of steps. This duplex DNA is colinear with its RNA template, but contains terminal duplications known as LTRs that are not present in viral RNA. It has been proposed that two specialized template switches, known as strand-transfer reactions or "jumps", are required to generate the LTRs. [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]
cd01646	CDD	*RT_Bac_retron_I*	RT_Bac_retron_I: Reverse transcriptases (RTs) in bacterial retrotransposons or retrons. The polymerase reaction of this enzyme leads to the production of a unique RNA-DNA complex called msDNA (multicopy single-stranded (ss)DNA) in which a small ssDNA branches out from a small ssRNA molecule via a 2'-5'phosphodiester linkage. Bacterial retron RTs produce cDNA corresponding to only a small portion of the retron genome. [PMID: 1698615, PMID: 16093702, PMID: 8828137]
cd01648	CDD	*TERT*	TERT: Telomerase reverse transcriptase (TERT). Telomerase is a ribonucleoprotein (RNP) that synthesizes telomeric DNA repeats. The telomerase RNA subunit provides the template for synthesis of these repeats. The catalytic subunit of RNP is known as telomerase reverse transcriptase (TERT). The reverse transcriptase (RT) domain is located in the C-terminal region of the TERT polypeptide. Single amino acid substitutions in this region lead to telomere shortening and senescence. Telomerase is an enzyme that, in certain cells, maintains the physical ends of chromosomes (telomeres) during replication. In somatic cells, replication of the lagging strand requires the continual presence of an RNA primer approximately 200 nucleotides upstream, which is complementary to the template strand. Since there is a region of DNA less than 200 base pairs from the end of the chromosome where this is not possible, the chromosome is continually shortened. However, a surplus of repetitive DNA at the chromosome ends protects against the erosion of gene-encoding DNA. Telomerase is not normally expressed in somatic cells. It has been suggested that exogenous TERT may extend the lifespan of, or even immortalize, the cell. However, recent studies have shown that telomerase activity can be induced by a number of oncogenes. Conversely, the oncogene c-myc can be activated in human TERT immortalized cells. Sequence comparisons place the

			telomerase proteins in the RT family but reveal hallmarks that distinguish them from retroviral and retrotransposon relatives. [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_like*	RT_nLTR: Non-LTR (long terminal repeat) retrotransposon and non-LTR retrovirus reverse transcriptase (RT). This subfamily contains both non-LTR retrotransposons and non-LTR retrovirus RTs. RTs catalyze the conversion of single-stranded RNA into double-stranded DNA for integration into host chromosomes. RT is a multifunctional enzyme with RNA-directed DNA polymerase, DNA directed DNA polymerase and ribonuclease hybrid (RNase H) activities. [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_G2_intron*	RT_G2_intron: Reverse transcriptases (RTs) with group II intron origin. RT transcribes DNA using RNA as template. Proteins in this subfamily are found in bacterial and mitochondrial group II introns. Their most probable ancestor was a retrotransposable element with both gag-like and pol-like genes. This subfamily of proteins appears to have captured the RT sequences from transposable elements, which lack long terminal repeats (LTRs). [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_Bac_retron_II*	RT_Bac_retron_II: Reverse transcriptases (RTs) in bacterial retrotransposons or retrons. The polymerase reaction of this enzyme leads to the production of a unique RNA-DNA complex called msDNA (multicopy single-stranded (ss)DNA) in which a small ssDNA branches out from a small ssRNA molecule via a 2'-5'phosphodiester linkage. Bacterial retron RTs produce cDNA corresponding to only a small portion of the retron

			genome. [PMID: 1698615, PMID: 8828137, PMID: 11292805, PMID: 9281493, PMID: 2465092, PMID: 1722556, PMID: 1701261, PMID: 1689062]
cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
cd05536	CDD	*POLBc_B3*	DNA polymerase type-B B3 subfamily catalytic domain. Archaeal proteins that are involved in DNA replication are similar to those from eukaryotes. Some members of the archaea also possess multiple family B DNA polymerases (B1, B2 and B3). So far there is no specific function(s) has been assigned for different members of the archaea type B DNA polymerases. Phylogenetic analyses of eubacterial, archaeal, and eukaryotic family B DNA polymerases are support independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. Structural comparison of the thermostable DNA polymerase type B to its mesostable homolog suggests several adaptations to high temperature such as shorter loops, disulfide bridges, and increasing electrostatic interaction at subdomain interfaces. [PMID: 10997874, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321]
cd05780	CDD	*DNA_polB_Kod1_like_exo*	The 3'-5' exonuclease domain of archaeal family-B DNA polymerases with similarity to Pyrococcus kodakaraensis Kod1, including polymerases from Desulfurococcus (D. Tok Pol) and Thermococcus gorgonarius (Tgo Pol). Kod1, D. Tok Pol, and Tgo Pol are thermostable enzymes that exhibit both polymerase and 3'-5' exonuclease activities. They are family-B DNA polymerases. Their

			amino termini harbor a DEDDy-type DnaQ-like 3'-5' exonuclease domain that contains three sequence motifs termed ExoI, ExoII and ExoIII, with a specific YX(3)D pattern at ExoIII. These motifs are clustered around the active site and are involved in metal binding and catalysis. The exonuclease domain of family B polymerases contains a beta hairpin structure that plays an important role in active site switching in the event of nucleotide misincorporation. Members of this subfamily show similarity to eukaryotic DNA polymerases involved in DNA replication. Some archaea possess multiple family-B DNA polymerases. Phylogenetic analyses of eubacterial, archaeal, and eukaryotic family-B DNA polymerases support independent gene duplications during the evolution of archaeal and eukaryotic family-B DNA polymerases. [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*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. [PMID: 1698615]
PF00136	Pfam	*DNA_pol_B*	This region of DNA polymerase B appears to consist of more than one structural domain, possibly including elongation, DNA-binding and dNTP binding activities. [PMID: 9757117, PMID: 8679562]
PF07727	Pfam	*RVT_2*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. This Pfam entry includes reverse transcriptases not recognised by the Pfam:PF00078 model. [PMID: 1698615]
IPR000477	InterPro	*RT_dom*	The use of an RNA template to produce DNA, for integration into the host genome and exploitation of a host cell, is a strategy employed in the replication of retroid elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse trancriptase (RT) (2.7.7.49). Reverse transcriptase occurs in a variety of mobile elements, including retrotransposons, retroviruses,

			group II introns [PMID: 12758069], bacterial msDNAs, hepadnaviruses, and caulimoviruses. Retroviral reverse transcriptase is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. The discovery of retroelements in the prokaryotes raises intriguing questions concerning their roles in bacteria and the origin and evolution of reverse transcriptases and whether the bacterial reverse transcriptases are older than eukaryotic reverse transcriptases [PMID: 8828137]. Several crystal structures of the reverse transcriptase (RT) domain have been determined [PMID: 1377403].
IPR006134	InterPro	DNA-dir_DNA_pol_B_multi_dom	DNA is the biological information that instructs cells how to exist in an ordered fashion: accurate replication is thus one of the most important events in the life cycle of a cell. This function is performed by DNA- directed DNA-polymerases 2.7.7.7) by adding nucleotide triphosphate (dNTP) residues to the 5' end of the growing chain of DNA, using a complementary DNA chain as a template. Small RNA molecules are generally used as primers for chain elongation, although terminal proteins may also be used for the de novo synthesis of a DNA chain. Even though there are 2 different methods of priming, these are mediated by 2 very similar polymerases classes, A and B, with similar methods of chain elongation. A number of DNA polymerases have been grouped under the designation of DNA polymerase family B. Six regions of similarity (numbered from I to VI) are found in all or a subset of the B family polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. It has been suggested that it may be involved in binding a magnesium ion. All sequences in the B family contain a characteristic DTDS motif, (SEQ ID NO: 1566) and possess many functional domains, including a 5'-3' elongation domain, a 3'-5' exonuclease domain [PMID: 8679562], a DNA binding domain, and binding domains for both dNTP's and pyrophosphate [PMID: 9757117]. This domain of DNA polymerase B appears to consist of more than one activities, possibly including elongation, DNA-binding and dNTP binding [PMID: 9757117].
IPR013103	InterPro	*RVT_2*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group

			II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. This entry includes reverse transcriptases not recognised by IPR000477 [PMID: 1698615].

Table 30: Exemplary monomeric retroviral reverse transcriptases and their RT domain signatures

Name	Accession	Organism	Sequence	RT Signatures
Q4VFZ2_9GAMR - residues only	Q4VFZ2	Porcine endogenous retrovirus	MGATGQQQYPWTTRRTVDLGVGRVT HSFLVIPECPAPLLGRDLLTKMGAQISF EQGKPEVSANNKPITVLTLQLDDEYRL YSPLVKPDQNIQFWLEQFPQAWAETA GMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQ SPWNTPLLPVRKPGTNDYRPVQDLRE VNKRVQDIHPTVPNPYNLLCALPPQRS WYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPGTGRTGQLTWTRLPQGFKNSPTI FDEALHRDLANFRIQHPQVTLLQYVDD LLLAGATKQDCLEGTKALLLELSDLGY RASAKKAQICRREVTYLGYSLRDGQR WLTEARKKTVVQIPAPTTAKQVREFLG TAGFCRLWIPGFATLAAPLYPLTKEKG EFSWAPEHQKAFDAIKKALLSAPALAL PDVTKPFTLYVDERKGVARGVLTQTL GPWRRPVAYLSKKLDPVASGWPVCLK AIAAVAILVKDADKLTLGQNITVIAPH ALENIVRQPPDRWMTNARMTHYQSLL LTERVTFAPPAALNPATLLPEETDEPVT HDCHQLLIEETGVRKDLTDIPLTGEVLT WFTDGSSYVVEGKRMAGAAVVDGTR TIWASSLPEGTSAQKAELMALTQALRL AEGKSINIYTDSRYAFATAHVHGAIYK QRGLLTSAGREIKNKEEILSLLEALHLP KRLAIIHCPGHQKAKDPISRGNQMADR VAKQAAQGVNLLPMIETPKAPEPGRQ YTLEDWQEIKKIDQFSETPEGTCYTSD GKEILPHKEGLEYVQQIHRLTHLGTKH LQQLVRTSPYHVLRLPGVADSVVKHC VPCQLVNANPSRIPPGKRLRGSHPGAH WEVDFTEVKPAKYGNKYLLVFVDTFS GWVEAYPTKKETSTVVAKKILEEIFPR FGIPKVIGSDNGPAFVAQVSQGLAKILG IDWKLHCAYRPQSSGQVERMNRTIKET	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			LTKLTAETGVNDWIALLPFVLFRVRNT PGQFGLTPYELLYGGPPPLVEIASVHSA DVLLSQPLFSRLKALEWVRQRAWRQL REAYSGGGDLQIPHRFQVGDSVYVRR HRAGNLETRWKGPYHVLLTTPTAVKV EGISTWIHASHVKPAPPPDSGWKAEKT ENPLKLRLHRVVPYSVNNFSS (SEQ ID NO: 1567)	

POL_SFV1 - residues only	P23074	Simian foamy virus type 1	MDPLQLLQPLEAEIKGTKLKAHWDSG ATITCVPEAFLEDERPIQTMLIKTIHGEK QQDVYYLTFKVQGRKVEAEVLASPYD YILLNPSDVPWLMKKPLQLTVLVPLHE YQERLLQQTALPKEQKELLQKLFLKY DALWQHWENQVGHRRIKPHNIATGTL APRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSI YRGKYKTTLDLTNGFWAHPITPESYW LTAFTWQGKQYCWTRLPQGFLNSPAL FTADVVDLLKEIPNVQAYVDDIYISHD DPQEHLEQLEKIFSILLNAGYVVSLKKS EIAQREVEFLGFNITKEGRGLTDTFKQK LLNITPPKDLKQLQSILGLLNFARNFIPN YSELVKPLYTIVANANGKFISWTEDNS NQLQHIISVLNQADNLEERNPETRLIIK VNSSPSAGYIRYYNEGSKRPIMYVNYIF SKAEAKFTQTEKLLTTMHKGLIKAMD LAMGQEILVYSPIVSMTKIQRTPLPERK ALPVRWITWMTYLEDPRIQFHYDKSLP ELQQIPNVTEDVIAKTKHPSEFAMVFY TDGSAIKHPDVNKSHSAGMGIAQVQFI PEYKIVHQWSIPLGDHTAQLAEIAAVE FACKKALKISGPVLIVTDSFYVAESAN KELPYWKSNGFLNNKKKPLRHVSKW KSIAECLQLKPDIIIMHEKGHQQPMTTL HTEGNNLADKLATQGSYVVHCNTTPS LDAELDQLLQGHYPPGYPKQYKYTLE ENKLIVERPNGIRIVPPKADREKIISTAH NIAHTGRDATFLKVSSKYWWPNLRKD VVKSIRQCKQCLVTNATNLTSPPILRPV KPLKPFDKFYIDYIGPLPPSNGYLHVLV VVDSMTGFVWLYPTKAPSTSATVKAL NMLTSIAIPKVLHSDQGAAFTSSTFAD WAKEKGIQLEFSTPYHPQSSGKVERKN SDIKRLLTKLLIGRPAKWYDLLPVVQL ALNNSYSPSSKYTPHQLLFGVDSNTPF ANSDTLDLSREEELSLLQEIRSSLHQPT SPPASSRSWSPSVGQLVQERVARPASL RPRWHKPTAILEVVNPRTVIILDHLGN RRTVSVDNLKLTAYQDNGTSNDSGTM ALMEEDESSTSST (SEQ ID NO: 1568)	IPR043502, SSF56672, IPR000477, PF00078

POL_MPMV - residues only	P07572	Mason-Pfizer monkey virus	MGQELSQHERYVEQLKQALKTRGVK VKYADLLKFFDFVKDTCPWFPQEGTID IKRWRRVGDCFQDYYNTFGPEKVPVT AFSYWNLIKELIDKKEVNPQVMAAVA QTEEILKSNSQTDLTKTSQNPDLDLISL DSDDEGAKSSSLQDKGLSSTKKPKRFP VLLTAQTSKDPEDPNPSEVDWDGLED EAAKYHNPDWPPFLTRPPPYNKATPSA PTVMAVVNPKEELKEKIAQLEEQIKLE ELHQALISKLQKLKTGNETVTHPDTAG GLSRTPHWPGQHIPKGKCCASREKEEQ IPKDIFPVTETVDGQGQAWRHHNGFDF AVIKELKTAASQYGATAPYTLAIVESV ADNWLTPTDWNTLVRAVLSGGDHLL WKSEFFENCRDTAKRNQQAGNGWDF DMLTGSGNYSSTDAQMQYDPGLFAQI QAAATKAWRKLPVKGDPGASLTGVK QGPDEPFADFVHRLITTAGRIFGSAEAG VDYVKQLAYENANPACQAAIRPYRKK TDLTGYIRLCSDIGPSYQQGLAMAAAF SGQTVKDFLNNKNKEKGGCCFKCGKK GHFAKNCHEHAHNNAEPKVPGLCPRC KRGKHWANECKSKTDNQGNPIPPHQG NRVEGPAPGPETSLWGSQLCSSQQKQP ISKLTRATPGSAGLDLCSTSHTVLTPEM GPQALSTGIYGPLPPNTFGLILGRSSITM KGLQVYPGVIDNDYTGEIKIMAKAVN NIVTVSQGNRIAQLILLPLIETDNKVQQ PYRGQGSFGSSDIYWVQPITCQKPSLTL WLDDKMFTGLIDTGADVTIIKLEDWPP NWPITDTLTNLRGIGQSNNPKQSSKYL TWRDKENNSGLIKPFVIPNLPVNLWGR DLLSQMKIMMCSPNDIVTAQMLAQGY SPGKGLGKKENGILHPIPNQGQSNKKG FGNFLTAAIDILAPQQCAEPITWKSDEP VWVDQWPLTNDKLAAAQQLVQEQLE AGHITESSSPWNTPIFVIKKKSGKWRLL QDLRAVNATMVLMGALQPGLPSPVAI PQGYLKIIIDLKDCFFSIPLHPSDQKRFA FSLPSTNFKEPMQRFQWKVLPQGMAN SPTLCQKYVATAIHKVRHAWKQMYII HYMDDILIAGKDGQQVLQCFDQLKQE LTAAGLHIAPEKVQLQDPYTYLGFELN GPKITNQKAVIRKDKLQTLNDFQKLLG DINWLRPYLKLTTGDLKPLFDTLKGDS DPNSHRSLSKEALASLEKVETAIAEQF	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			VTHINYSLPLIFLIFNTALTPTGLFWQD NPIMWIHLPASPKKVLLPYYDAIADLII LGRDHSKKYFGIEPSTIIQPYSKSQIDW LMQNTEMWPIACASFVGILDNHYPPN KLIQFCKLHTFVFPQIISKTPLNNALLVF TDGSSTGMAAYTLTDTTIKFQTNLNSA QLVELQALIAVLSAFPNQPLNIYTDSAY LAHSIPLLETVAQIKHISETAKLFLQCQ QLIYNRSIPFYIGHVRAHSGLPGPIAQG NQRADLATKIVASNINTNLESAQNAHT LHHLNAQTLRLMFNIPREQARQIVKQC PICVTYLPVPHLGVNPRGLFPNMIWQM DVTHYSEFGNLKYIHVSIDTFSGFLLAT LQTGETTKHVITHLLHCFSIIGLPKQIKT DNGPGYTSKNFQEFCSTLQIKHITGIPY NPQGQGIVERAHLSLKTTIEKIKKGEW YPRKGTPRNILNHALFILNFLNLDDQN KSAADRFWHNNPKKQFAMVKWKDPL DNTWHGPDPVLIWGRGSVCVYSQTYD AARWLPERLVRQVSNNNQSRE (SEQ ID NO: 1569)	

POL_MMTVB - residues only	P03365	Mouse mammary tumor virus	MGVSGSKGQKLFVSVLQRLLSERGLH VKESSAIEFYQFLIKVSPWFPEEGGLNL QDWKRVGREMKRYAAEHGTDSIPKQ AYPIWLQLREILTEQSDLVLLSAEAKSV TEEELEEGLTGLLSTSSQEKTYGTRGT AYAEIDTEVDKLSEHIYDEPYEEKEKA DKNEEKDHVRKIKKVVQRKENSEGKR KEKDSKAFLATDWNDDDLSPEDWDD LEEQAAHYHDDDELILPVKRKVVKKK PQALRRKPLPPVGFAGAMAEAREKGD LTFTFPVVFMGESDEDDTPVWEPLPLK TLKELQSAVRTMGPSAPYTLQVVDMV ASQWLTPSDWHQTARATLSPGDYVL WRTEYEEKSKEMVQKAAGKRKGKVS LDMLLGTGQFLSPSSQIKLSKDVLKDV TTNAVLAWRAIPPPGVKKTVLAGLKQ GNEESYETFISRLEEAVYRMMPRGEGS DILIKQLAWENANSLCQDLIRPIRKTGT IQDYIRACLDASPAVVQGMAYAAAMR GQKYSTFVKQTYGGGKGGQGAEGPV CFSCGKTGHIRKDCKDEKGSKRAPPGL CPRCKKGYHWKSECKSKFDKDGNPLP PLETNAENSKNLVKGQSPSPAQKGDG VKGSGLNPEAPPFTIHDLPRGTPGSAGL DLSSQKDLILSLEDGVSLVPTLVKGTLP EGTTGLIIGRSSNYKKGLEVLPGVIDSD FQGEIKVMVKAAKNAVIIHKGERIAQL LLLPYLKLPNPVIKEERGSEGFGSTSHV HWVQEISDSRPMLHIYLNGRRFLGLLD TGADKTCIAGRDWPANWPIHQTESSLQ GLGMACGVARSSQPLRWQHEDKSGII HPFVIPTLPFTLWGRDIMKDIKVRLMT DSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKLQALQQLVTE QLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLP SPVAVPKGWEIIIIDLQDCFFNIKLHPED CKRFAFSVPSPNFKRPYQRFQWKVLPQ GMKNSPTLCQKFVDKAILTVRDKYQD SYIVHYMDDILLAHPSRSIVDEILTSMI QALNKHGLVVSTEKIQKYDNLKYLGT HIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNG DSNPISTRKLTPEACKALQLMNERLST ARVKRLDLSQPWSLCILKTEYTPTACL WQDGVVEWIHLPHISPKVITPYDIFCTQ	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			LIIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLP KDPLLTFTLQTAIIFPHMTSTTPLEKGIV IFTDGSANGRSVTYIQGREPIIKENTQN TAQQAEIVAVITAFEEVSQPFNLYTDSK YVTGLFPEIETATLSPRTKIYTELKHLQ RLIHKRQEKFYIGHIRGHTGLPGPLAQG NAYADSLTRILTALESAQESHALHHQN AAALRFQFHITREQAREIVKLCPNCPD WGHAPQLGVNPRGLKPRVLWQMDVT HVSEFGKLKYVHVTVDTYSHFTFATA RTGEATKDVLQHLAQSFAYMGIPQKIK TDNAPAYVSRSIQEFLARWKISHVTGIP YNPQGQAIVERTHQNIKAQLNKLQKA GKYYTPHHLLAHALFVLNHVNMDNQ GHTAAERHWGPISADPKPMVMWKDL LTGSWKGPDVLITAGRGYACVFPQDA ETPIWVPDRFIRPFTERKEATPTPGTAE KTPPRDEKDQQESPKNESSPHQREDGL ATSAGVDLRSGGGP (SEQ ID NO: 1570)	

POL_MLVMS - residues only	P03355	Moloney murine leukemia virus	MGQTVTTPLSLTLGHWKDVERIAHNQ SVDVKKRRWVTFCSAEWPTFNVGWP RDGTFNRDLITQVKIKVFSPGPHGHPD QVPYIVTWEALAFDPPPWVKPFVHPKP PPPLPPSAPSLPLEPPRSTPPRSSLYPALT PSLGAKPKPQVLSDSGGPLIDLLTEDPP PYRDPRPPPSDRDGNGGEATPAGEAPD PSPMASRLRGRREPPVADSTTSQAFPL RAGGNGQLQYWPFSSSDLYNWKNNN PSFSEDPGKLTALIESVLITHQPTWDDC QQLLGTLLTGEEKQRVLLEARKAVRG DDGRPTQLPNEVDAAFPLERPDWDYT TQAGRNHLVHYRQLLLAGLQNAGRSP TNLAKVKGITQGPNESPSAFLERLKEA YRRYTPYDPEDPGQETNVSMSFIWQSA PDIGRKLERLEDLKNKTLGDLVREAEK IFNKRETPEEREERIRRETEEKEERRRTE DEQKEKERDRRRHREMSKLLATVVSG QKQDRQGGERRRSQLDRDQCAYCKE KGHWAKDCPKKPRGPRGPRPQTSLLT LDDQGGQGQEPPPEPRITLKVGGQPVT FLVDTGAQHSVLTQNPGPLSDKSAWV QGATGGKRYRWTTDRKVHLATGKVT HSFLHVPDCPYPLLGRDLLTKLKAQIH FEGSGAQVMGPMGQPLQVLTLNIEDE HRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKATSTPVSI KQYPMSQEARLGIKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYV DDLLLAATSELDCQQGTRALLQTLGN LGYRASAKKAQICQKQVKYLGYLLKE GQRWLTEARKETVMGQPTPKTPRQLR EFLGTAGFCRLWIPGFAEMAAPLYPLT KTGTLFNWGPDQQKAYQEIKQALLTA PALGLPDLTKPFELFVDEKQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDRVQFGPVVALNPATLLPL PEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIAL	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			TQALKMAEGKKLNVYTDSRYAFATA HIHGEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEAR GNRMADQAARKAAITETPDTSTLLIEN SSPYTSEHFHYTVTDIKDLTKLGAIYDK TKKYWVYQGKPVMPDQFTFELLDFLH QLTHLSFSKMKALLERSHSPYYMLNR DRTLKNITETCKACAQVNASKSAVKQ GTRVRGHRPGTHWEIDFTEIKPGLYGY KYLLVFIDTFSGWIEAFPTKKETAKVV TKKLLEEIFPRFGMPQVLGTDNGPAFV SKVSQTVADLLGIDWKLHCAYRPQSS GQVERMNRTIKETLTKLTLATGSRDW VLLLPLALYRARNTPGPHGLTPYEILY GAPPPLVNFPDPDMTRVTNSPSLQAHL QALYLVQHEVWRPLAAAYQEQLDRP VVPHPYRVGDTVWVRRHQTKNLEPR WKGPYTVLLTTPTALKVDGIAAWIHA AHVKAADPGGGPSSRLTWRVQRSQNP LKIRLTREAP (SEQ ID NO: 1571)	

POL_HTL1A - residues only	P03362	Human T-cell leukemia virus 1	MGQIFSRSASPIPRPPRGLAAHHWLNFL QAAYRLEPGPSSYDFHQLKKFLKIALE TPARICPINYSLLASLLPKGYPGRVNEIL HILIQTQAQIPSRPAPPPPSSPTHDPPDS DPQIPPPYVEPTAPQVLPVMHPHGAPP NHRPWQMKDLQAIKQEVSQAAPGSPQ FMQTIRLAVQQFDPTAKDLQDLLQYL CSSLVASLHHQQLDSLISEAETRGITGY NPLAGPLRVQANNPQQQGLRREYQQL WLAAFAALPGSAKDPSWASILQGLEEP YHAFVERLNIALDNGLPEGTPKDPILRS LAYSNANKECQKLLQARGHTNSPLGD MLRACQTWTPKDKTKVLVVQPKKPPP NQPCFRCGKAGHWSRDCTQPRPPPGP CPLCQDPTHWKRDCPRLKPTIPEPEPEE DALLLDLPADIPHPKNLHRGGGLTSPP TLQQVLPNQDPASILPVIPLDPARRPVI KAQVDTQTSHPKTIEALLDTGADMTV LPIALFSSNTPLKNTSVLGAGGQTQDH FKLTSLPVLIRLPFRTTPIVLTSCLVDTK NNWAIIGRDALQQCQGVLYLPEAKRPP VILPIQAPAVLGLEHLPRPPQISQFPLNP ERLQALQHLVRKALEAGHIEPYTGPGN NPVFPVKKANGTWRFIHDLRATNSLTI DLSSSSPGPPDLSSLPTTLAHLQTIDLRD AFFQIPLPKQFQPYFAFTVPQQCNYGP GTRYAWKVLPQGFKNSPTLFEMQLAH ILQPIRQAFPQCTILQYMDDILLASPSHE DLLLLSEATMASLISHGLPVSENKTQQ TPGTIKFLGQIISPNHLTYDAVPTVPIRS RWALPELQALLGEIQWVSKGTPTLRQP LHSLYCALQRHTDPRDQIYLNPSQVQS LVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKEQWPLVWLHAPL PHTSQCPWGQLLASAVLLLDKYTLQS YGLLCQTIHHNISTQTFNQFIQTSDHPS VPILLHHSHRFKNLGAQTGELWNTFLK TAAPLAPVKALMPVFTLSPVIINTAPCL FSDGSTSRAAYILWDKQILSQRSFPLPP PHKSAQRAELLGLLHGLSSARSWRCL NIFLDSKYLYHYLRTLALGTFQGRSSQ APFQALLPRLLSRKVVYLHHVRSHTNL PDPISRLNALTDALLITPVLQLSPAELHS FTHCGQTALTLQGATTTEASNILRSCH ACRGGNPQHQMPRGHIRRGLLPNHIW QGDITHFKYKNTLYRLHVWVDTFSGAI	IPR043502, SSF56672, IPR000477, PF00078

			SATQKRKETSSEAISSLLQAIAHLGKPS YINTDNGPAYISQDFLNMCTSLAIRHTT HVPYNPTSSGLVERSNGILKTLLYKYF TDKPDLPMDNALSIALWTINHLNVLTN CHKTRWQLHHSPRLQPIPETRSLSNKQ THWYYFKLPGLNSRQWKGPQEALQEA AGAALIPVSASSAQWIPWRLLKRAACP RPVGGPADPKEKDLQHHG (SEQ ID NO: 1572)	

POL_FOAMV - residues only	P14350	Human spumaretrovirus	MNPLQLLQPLPAEIKGTKLLAHWDSG ATITCIPESFLEDEQPIKKTLIKTIHGEK QQNVYYVTFKVKGRKVEAEVIASPYE YILLSPTDVPWLTQQPLQLTILVPLQEY QEKILSKTALPEDQKQQLKTLFVKYDN LWQHWENQVGHRKIRPHNIATGDYPP RPQKQYPINPKAKPSIQIVIDDLLKQGV LTPQNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIV RQKYKTTLDLANGFWAHPITPESYWL TAFTWQGKQYCWTRLPQGFLNSPALF TADVVDLLKEIPNVQVYVDDIYLSHDD PKEHVQQLEKVFQILLQAGYVVSLKKS EIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGLLNFARNFIPN FAELVQPLYNLIASAKGKYIEWSEENT KQLNMVIEALNTASNLEERLPEQRLVI KVNTSPSAGYVRYYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKA MDLAMGQEILVYSPIVSMTKIQKTPLP ERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGV FYTDGSAIKSPDPTKSNNAGMGIVHAT YKPEYQVLNQWSIPLGNHTAQMAEIA AVEFACKKALKIPGPVLVITDSFYVAES ANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVF ILKGNALADKLATQGSYVVNCNTKKP NLDAELDQLLQGHYIKGYPKQYTYFL EDGKVKVSRPEGVKIIPPQSDRQKIVLQ AHNLAHTGREATLLKIANLYWWPNM RKDVVKQLGRCQQCLITNASNKASGPI LRPDRPQKPFDKFFIDYIGPLPPSQGYL YVLVVVDGMTGFTWLYPTKAPSTSAT VKSLNVLTSIAIPKVIHSDQGAAFTSST FAEWAKERGIHLEFSTPYHPQSGSKVE RKNSDIKRLLTKLLVGRPTKWYDLLPV VQLALNNTYSPVLKYTPHQLLFGIDSN TPFANQDTLDLTREEELSLLQEIRTSLY HPSTPPASSRSWSPVVGQLVQERVARP ASLRPRWHKPSTVLKVLNPRTVVILDH LGNNRTVSIDNLKPTSHQNGTTNDTAT MDHLEKNE (SEQ ID NO: 1573)	IPR043502, SSF56672, IPR000477, PF00078

POL_BLVJ - residues only	P03361	Bovine leukemia virus	MGNSPSYNPPAGISPSDWLNLLQSAQR LNPRPSPSDFTDLKNYIHWFHKTQKKP WTFTSGGPTSCPPGRFGRVPLVLATLN EVLSNEGGAPGASAPEEQPPPYDPPAIL PIISEGNRNRHRAWALRELQDIKKEIEN KAPGSQVWIQTLRLAILQADPTPADLE QLCQYIASPVDQTAHMTSLTAAIAAAE AANTLQGFNPKTGTLTQQSAQPNAGD LRSQYQNLWLQAGKNLPTRPSAPWSTI VQGPAESSVEFVNRLQISLADNLPDGV PKEPIIDSLSYANANRECQQILQGRGPV AAVGQKLQACAQWAPKNKQPALLVH TPGPKMPGPRQPAPKRPPPGPCYRCLK EGHWARDCPTKATGPPPGPCPICKDPS HWKRDCPTLKSKNKLIEGGLSAPQTIT PITDSLSEAELECLLSIPLARSRPSVAVY LSGPWLQPSQNQALMLVDTGAENTVL PQNWLVRDYPRIPAAVLGAGGVSRNR YNWLQGPLTLALKPEGPFITIPKILVDT SDKWQILGRDVPSRLQASISIPEEVRPP VVGVLDTPPSHIGLEHLPPPPEVPQFPL NLERLQALQDLVHRSLEAGYISPWDGP GNNPVFPVRKPNGAWRFVHDLRATNA LTKPIPALSPGPPDLTAIPTHPPHIICLDL KDAFFQIPVEDRFRFYLSFTLPSPGGLQ PHRRFAWRVLPQGFINSPALFERALQE PLRQVSAAFSQSLLVSYMDDILYASPT EEQRSQCYQALAARLRDLGFQVASEK TSQTPSPVPFLGQMVHEQIVTYQSLPTL QISSPISLHQLQAVLGDLQWVSRGTPTT RRPLQLLYSSLKRHHDPRAIIQLSPEQL QGIAELRQALSHNARSRYNEQEPLLAY VHLTRAGSTLVLFQKGAQFPLAYFQTP LTDNQASPWGLLLLLGCQYLQTQALS SYAKPILKYYHNLPKTSLDNWIQSSED PRVQELLQLWPQISSQGIQPPGPWKTLI TRAEVFLTPQFSPDPIPAALCLFSDGAT GRGAYCLWKDHLLDFQAVPAPESAQK GELAGLLAGLAAAPPEPVNIWVDSKY LYSLLRTLVLGAWLQPDPVPSYALLYK SLLRHPAIVVGHVRSHSSASHPIASLNN YVDQLLPLETPEQWHKLTHCNSRALS RWPNPRISAWDPRSPATLCETCQKLNP TGGGKMRTIQRGWAPNHIWQADITHY KYKQFTYALHVFVDTYSGATHASAKR GLTTQTTIEGLLEAIVHLGRPKKLNTD	IPR043502, SSF56672, IPR000477, PF00078

			QGANYTSKTFVRFCQQFGVSLSHHVP YNPTSSGLDERTNGLLKLLLSKYHLDE PHLPMTQALSRALWTHNQINLLPILKT RWELHHSPPLAVISEGGETPKGSDKLF LYLLPGQNNRRWLGPLPALVEASGGA LLATDPPVWVPWRLLKAFKCLKNDGP EDAHNRSSDG (SEQ ID NO: 1574)	

O41894_9RETR - residues only	O41894	Bovine foamy virus	MPALRPLQVEIKGNHLKGYWDSGAEI TCVPAIYIIEEQPVGKKLITTIHNEKEHD VYYVEMKIEKRKVQCEVIATALDYVL VAPVDIPWYKPGPLELTIKIDVESQKHT LITESTLSPQGQMRLKKLLDQYQALW QCWENQVGHRRIEPHKIATGALKPRPQ KQYHINPRAKADIQIVIDDLLRQGVLR QQNSEMNTPVYPVPKADGRWRMVLD YREVNKVTPLVATQNCHSASILNTLYR GPYKSTLDLANGFWAHPIKPEDYWITA FTWGGKTYCWTVLPQGFLNSPALFTA DVVDILKDIPNVQVYVDDVYVSSATE QEHLDILETIFNRLSTAGYIVSLKKSKL AKETVEFLGFSISQNGRGLTDSYKQKL MDLQPPTTLRQLQSILGLINFARNFLPN FAELVAPLYQLIPKAKGQCIPWTMDHT TQLKTIIQALNSTENLEERRPDVDLIMK VHISNTAGYIRFYNHGGQKPIAYNNAL FTSTELKFTPTEKIMATIHKGLLKALDL SLGKEIHVYSAIASMTKLQKTPLSERK ALSIRWLKWQTYFEDPRIKFHHDATLP DLQNLPVPQQDTGKEMTILPLLHYEAI FYTDGSAIRSPKPNKTHSAGMGIIQAKF EPDFRIVHLWSFPLGDHTAQYAEIAAF EFAIRRATGIRGPVLIVTDSNYVAKSYN EELPYWESNGFVNNKKKTLKHISKWK AIAECKNLKADIHVIHEPGHQPAEASP HAQGNALADKQAVSGSYKVFSNELKP SLDAELEQVLSTGRPNPQGYPNKYEYK LVNGLCYVDRRGEEGLKIIPPKADRVK LCQLAHDGPGSAHLGRSALLLKLQQK YWWPRMHIDASRIVLNCTVCAQTNST NQKPRPPLVIPHDTKPFQVWYMDYIGP LPPSNGYQHALVIVDAGTGFTWIYPTK AQTANATVKALTHLTGTAVPKVLHSD QGPAFTSSILADWAKDRGIQLEHSAPY HPQSSGKVERKNSEIKRLLTKLLAGRP TKWYPLIPIVQLALNNTPNTRQKYTPH QLMYGADCNLPFENLDTLDLTREEQL AVLKEVRDGLLDLYPSPSQTTARSWTP SPGLLVQERVARPAQLRPKWRKPTPIK KVLNERTVIIDHLGQDKVVSIDNLKPA AHQKLAQTPDSAEICPSATPCPPNTSL WYDLDTGTWTCQRCGYQCPDKYHQP QCTWSCEDRCGHRWKECGNCIPQDGS SDDASAVAAVEI (SEQ ID NO: 1575)	IPR043502, SSF56672, IPR000477, PF00078

Table 31: Exemplary dimeric retroviral reverse transcriptases and their RT domain signatures

Name	Accession	Organism	Sequence	RT Signatures
Q83133_AVIMA	Q83133	Avian myeloblastosis-associated virus type 1	RATVLTVALHLAIPLKWKPNHTPVWID QWPLPEGKLVALTQLVEKELQLGHIEP SLSCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRGWP LMVLDLKDCFFSIPLAEQDREAFAFTLP SVNNQAPARRFQWKVLPQGMTCSPTI CQLIVGQILEPLRLKHPSLRMLHYMDD LLLAASSHDGLEAAGEEVISTLERAGF TISPDKVQREPGVQYLGYKLGSTYVAP VGLVAEPRIATLWDVQKLVGSLQSVR PALGIPPRLMGPFYEQLRGSDPNEARE WNLDMKMAWREIVQLSTTAALERWD PALPLEGAVARCEQGAIGVLGQGLSTH PRPCLWLFSTQPTKAFTAWLEVLTLLIT KLRASAVRTFGKEVDILLLPACFREDL PLPEGILLALRGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDASS STHKGVVVWREGPRWEIKEIADLGAS VQQLEARAVAMALLLWPTTPTNVVTD SAFVAKMLLKMGQEGVPSTAAAFILE DALSQRSAMAAVLHVRSHSEVPGFFTE GNDVADSQATFQAYPLREAKDLHTAL HIGPRALSKACNISMQQAREVVQTCPH CNSAPALEAGVNPRGLGPLQIWQTDFT LEPRMAPRSWLAVTVDTASSAIVVTQ HGRVTSVAAQHHWATAIAVLGRPKAI KTDNGSCFTSKSTREWLARWGIAHTT GIPGNSQGQAMVERANRLLKDKIRVL AEGDGFMKRIPTSKQGELLAKAMYAL NHFERGENTKTPIQKHWRPTVLTEGPP VKIRIETGEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPDITQKDE VTKKDEASPLFAGISDWAPWEGEQEG LQEETASNKQERPGEDTPAANES (SEQ ID NO: 1576)	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

POL_SIVM1	P05896	Simian immunodeficiency virus	MGARNSVLSGKKADELEKIRLRPGGK KKYMLKHVVWAANELDRFGLAESLL ENKEGCQKILSVLAPLVPTGSENLKSL YNTVCVIWCIHAEEKVKHTEEAKQIVQ RHLVMETGTAETMPKTSRPTAPFSGRG GNYPVQQIGGNYTHLPLSPRTLNAWV KLIEEKKFGAEVVSGFQALSEGCLPYDI NQMLNCVGDHQAAMQIIRDIINEEAAD WDLQHPQQAPQQGQLREPSGSDIAGT TSTVEEQIQWMYRQQNPIPVGNIYRRW IQLGLQKCVRMYNPTNILDVKQGPKEP FQSYVDRFYKSLRAEQTDPAVKNWMT QTLLIQNANPDCKLVLKGLGTNPTLEE MLTACQGVGGPGQKARLMAEALKEA LAPAPIPFAAAQQKGPRKPIKCWNCGK EGHSARQCRAPRRQGCWKCGKMDHV MAKCPNRQAGFFRPWPLGKEAPQFPH GSSASGADANCSPRRTSCGSAKELHAL GQAAERKQREALQGGDRGFAAPQFSL WRRPVVTAHIEGQPVEVLLDTGADDSI VTGIELGPHYTPKIVGGIGGFINTKEYK NVEIEVLGKRIKGTIMTGDTPINIFGRN LLTALGMSLNLPIAKVEPVKSPLKPGK DGPKLKQWPLSKEKIVALREICEKMEK DGQLEEAPPTNPYNTPTFAIKKKDKNK WRMLIDFRELNRVTQDFTEVQLGIPHP AGLAKRKRITVLDIGDAYFSIPLDEEFR QYTAFTLPSVNNAEPGKRYIYKVLPQG WKGSPAIFQYTMRHVLEPFRKANPDV TLVQYMDDILIASDRTDLEHDRVVLQL KELLNSIGFSSPEEKFQKDPPFQWMGY ELWPTKWKLQKIELPQRETWTVNDIQ KLVGVLNWAAQIYPGIKTKHLCRLIRG KMTLTEEVQWTEMAEAEYEENKIILSQ EQEGCYYQESKPLEATVIKSQDNQWS YKIHQEDKILKVGKFAKIKNTHTNGVR LLAHVIQKIGKEAIVIWGQVPKFHLPVE KDVWEQWWTDYWQVTWIPEWDFIST PPLVRLVFNLVKDPIEGEETYYVDGSC SKQSKEGKAGYITDRGKDKVKVLEQT TNQQAELEAFLMALTDSGPKANIIVDS QYVMGIITGCPTESESRLVNQIIEEMIK KTEIYVAWVPAHKGIGGNQEIDHLVSQ GIRQVLFLEKIEPAQEEHSKYHSNIKEL VFKFGLPRLVAKQIVDTCDKCHQKGE AIHGQVNSDLGTWQMDCTHLEGKIVI	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			VAVHVASGFIEAEVIPQETGRQTALFLL KLASRWPITHLHTDNGANFASQEVKM VAWWAGIEHTFGVPYNPQSQGVVEA MNHHLKNQIDRIREQANSVETIVLMAV HCMNFKRRGGIGDMTPAERLINMITTE QEIQFQQSKNSKFKNFRVYYREGRDQL WKGPGELLWKGEGAVILKVGTDIKVV PRRKAKIIKDYGGGKEMDSSSHMEDT GEAREVA (SEQ ID NO: 1577)	

POL_ RSVP	P03354	Rous sarcoma virus	MEAVIKVISSACKTYCGKTSPSKKEIGA MLSLLQKEGLLMSPSDLYSPGSWDPIT AALSQRAMILGKSGELKTWGLVLGAL KAAREEQVTSEQAKFWLGLGGGRVSP PGPECIEKPATERRIDKGEEVGETTVQR DAKMAPEETATPKTVGTSCYHCGTAI GCNCATASAPPPPYVGSGLYPSLAGVG EQQGQGGDTPPGAEQSRAEPGHAGQA PGPALTDWARVREELASTGPPVVAMP VVIKTEGPAWTPLEPKLITRLADTVRT KGLRSPITMAEVEALMSSPLLPHDVTN LMRVILGPAPYALWMDAWGVQLQTVI AAATRDPRHPANGQGRGERTNLNRLK GLADGMVGNPQGQAALLRPGELVAIT ASALQAFREVARLAEPAGPWADIMQG PSESFVDFANRLIKAVEGSDLPPSARAP VIIDCFRQKSQPDIQQLIRTAPSTLTTPG EIIKYVLDRQKTAPLTDQGIAAAMSSAI QPLIMAVVNRERDGQTGSGGRARGLC YTCGSPGHYQAQCPKKRKSGNSRERC QLCNGMGHNAKQCRKRDGNQGQRPG KGLSSGPWPGPEPPAVSLAMTMEHKD RPLVRVILTNTGSHPVKQRSVYITALLD SGADITIISEEDWPTDWPVMEAANPQI HGIGGGIPMRKSRDMIELGVINRDGSL ERPLLLFPAVAMVRGSILGRDCLQGLG LRLTNLIGRATVLTVALHLAIPLKWKP DHTPVWIDQWPLPEGKLVALTQLVEK ELQLGHIEPSLSCWNTPVFVIRKASGSY RLLHDLRAVNAKLVPFGAVQQGAPVL SALPRGWPLMVLDLKDCFFSIPLAEQD REAFAFTLPSVNNQAPARRFQWKVLP QGMTCSPTICQLVVGQVLEPLRLKHPS LCMLHYMDDLLLAASSHDGLEAAGEE VISTLERAGFTISPDKVQREPGVQYLGY KLGSTYVAPVGLVAEPRIATLWDVQK LVGSLQWLRPALGIPPRLMGPFYEQLR GSDPNEAREWNLDMKMAWREIVRLST TAALERWDPALPLEGAVARCEQGAIG VLGQGLSTHPRPCLWLFSTQPTKAFTA WLEVLTLLITKLRASAVRTFGKEVDIL LLPACFREDLPLPEGILLALKGFAGKIR SSDTPSIFDIARPLHVSLKVRVTDHPVP GPTVFTDASSSTHKGVVVWREGPRWE IKEIADLGASVQQLEARAVAMALLLW PTTPTNVVTDSAFVAKMLLKMGQEGV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			PSTAAAFILEDALSQRSAMAAVLHVRS HSEVPGFFTEGNDVADSQATFQAYPLR EAKDLHTALHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAGVNPRGLG PLQIWQTDFTLEPRMAPRSWLAVTVD TASSAIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTREWL ARWGIAHTTGIPGNSQGQAMVERANR LLKDRIRVLAEGDGFMKRIPTSKQGEL LAKAMYALNHFERGENTKTPIQKHWR PTVLTEGPPVKIRIETGEWEKGWNVLV WGRGYAAVKNRDTDKVIWVPSRKVK PDITQKDEVTKKDEASPLFAGISDWIP WEDEQEGLQGETASNKQERPGEDTLA ANES (SEQ ID NO: 1578)	

POL_HV2D2	P15833	Human immunodeficiency virus type 2	MGARGSVLSGKKTDELEKVRLRPGGK KKYMLKHVVWAVNELDRFGLAESLL ESKEGCQKILKVLAPLVPTGSENLKSLF NIVCVIFCLHAEEKVKDTEEAKKIAQR HLAADTEKMPATNKPTAPPSGGNYPV QQLAGNYVHLPLSPRTLNAWVKLVEE KKFGAEVVPGFQALSEGCTPYDINQML NCVGEHQAAMQIIREIINEEAADWDQQ HPSPGPMPAGQLRDPRGSDIAGTTSTV EEQIQWMYRAQNPVPVGNIYRRWIQL GLQKCVRMYNPTNILDIKQGPKEPFQS YVDRFYKSLRAEQTDPAVKNWMTQT LLIQNANPDCKLVLKGLGMNPTLEEM LTACQGIGGPGQKARLMAEALKEALT PAPIPFAAVQQKAGKRGTVTCWNCGK QGHTARQCRAPRRQGCWKCGKTGHI MSKCPERQAGFLRVRTLGKEASQLPH DPSASGSDTICTPDEPSRGHDTSGGDTI CAPCRSSSGDAEKLHADGETTEREPRE TLQGGDRGFAAPQFSLWRRPVVKACIE GQSVEVLLDTGVDDSIVAGIELGSNYT PKIVGGIGGFINTKEYKDVEIEVVGKRV RATIMTGDTPINIFGRNILNTLGMTLNF PVAKVEPVKVELKPGKDGPKIRQWPLS REKILALKEICEKMEKEGQLEEAPPTNP YNTPTFAIKKKDKNKWRMLIDFRELN KVTQDFTEVNWVFPTRQVAEKRRITVI DVGDAYFSIPLDPNFRQYTAFTLPSVN NAEPGKRYIYKVLPQGWKGSQSICQYS MRKVLDPFRKANSDVIIIQYMDDILIAS DRSDLEHDRVVSQLKELLNDMGFSTPE EKFQKDPPFKWMGYELWPKKWKLQK IQLPEKEVWTVNAIQKLVGVLNWAAQ LFPGIKTRHICKLIRGKMTLTEEVQWTE LAEAELQENKIILEQEQEGSYYKERVPL EATVQKNLANQWTYKIHQGNKVLKV GKYAKVKNTHTNGVRLLAHVVQKIG KEALVIWGEIPVFHLPVERETWDQWW TDYWQVTWIPEWDFVSTPPLIRLAYNL VKDPLEGRETYYTDGSCNRTSKEGKA GYVTDRGKDKVKVLEQTTNQQAELEA FALALTDSEPQVNIIVDSQYVMGIIAAQ PTETESPIVAKIIEEMIKKEAVYVGWVP AHKGLGGNQEVDHLVSQGIRQVLFLE KIEPAQEEHEKYHGNVKELVHKFGIPQ LVAKQIVNSCDKCQQKGEAIHGQVNA	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			DLGTWQMDCTHLEGKIIIVAVHVASGF IEAEVIPQETGRQTALFLLKLASRWPIT HLHTDNGANFTSPSVKMVAWWVGIE QTFGVPYNPQSQGVVEAMNHHLKNQI DRLRDQAVSIETVVLMATHCMNFKRR GGIGDMTPAERLVNMITTEQEIQFFQA KNLKFQNFQVYYREGRDQLWKGPGEL LWKGEGAVIIKVGTEIKVVPRRKAKIIR HYGGGKGLDCSADMEDTRQAREMAQ SD (SEQ ID NO: 1579)	

POL_HV1A2	P03369	Human immunodeficiency virus type 1	MGARASVLSGGELDKWEKIRLRPGGK KKYKLKHIVWASRELERFAVNPGLLET SEGCRQILGQLQPSLQTGSEELRSLYNT VATLYCVHQRIDVKDTKEALEKIEEEQ NKSKKKAQQAAAAAGTGNSSQVSQN YPIVQNLQGQMVHQAISPRTLNAWVK VVEEKAFSPEVIPMFSALSEGATPQDL NTMLNTVGGHQAAMQMLKETINEEA AEWDRVHPVHAGPIAPGQMREPRGSD IAGTTSTLQEQIGWMTNNPPIPVGEIYK RWIILGLNKIVRMYSPTSILDIRQGPKEP FRDYVDRFYKTLRAEQASQDVKNWM TETLLVQNANPDCKTILKALGPAATLE EMMTACQGVGGPGHKARVLAEAMSQ VTNPANIMMQRGNFRNQRKTVKCFNC GKEGHIAKNCRAPRKKGCWRCGREGH QMKDCTERQANFLREDLAFLQGKARE FSSEQTRANSPTRRELQVWGGENNSLS EAGADRQGTVSFNFPQITLWQRPLVTI RIGGQLKEALLDTGADDTVLEEMNLP GKWKPKMIGGIGGFIKVRQYDQIPVEI CGHKAIGTVLVGPTPVNIIGRNLLTQIG CTLNFPISPIETVPVKLKPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKISKIG PENPYNTPVFAIKKKDSTKWRKLVDFR ELNKRTQDFWEVQLGIPHPAGLKKKK SVTVLDVGDAYFSVPLDKDFRKYTAF TIPSINNETPGIRYQYNVLPQGWKGSPA IFQSSMTKILEPFRKQNPDIVIYQYMDD LYVGSDLEIGQHRTKIEELRQHLLRWG FTTPDKKHQKEPPFLWMGYELHPDKW TVQPIMLPEKDSWTVNDIQKLVGKLN WASQIYAGIKVKQLCKLLRGTKALTE VIPLTEEAELELAENREILKEPVHEVYY DPSKDLVAEIQKQGQGQWTYQIYQEPF KNLKTGKYARMRGAHTNDVKQLTEA VQKVSTESIVIWGKIPKFKLPIQKETWE AWWMEYWQATWIPEWEFVNTPPLVK LWYQLEKEPIVGAETFYVDGAANRET KLGKAGYVTDRGRQKVVSIADTTNQK TELQAIHLALQDSGLEVNIVTDSQYAL GIIQAQPDKSESELVSQIIEQLIKKEKVY LAWVPAHKGIGGNEQVDKLVSAGIRK VLFLNGIDKAQEEHEKYHSNWRAMAS DFNLPPVVAKEIVASCDKCQLKGEAM HGQVDCSPGIWQLDCTHLEGKIILVAV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659

			HVASGYIEAEVIPAETGQETAYFLLKL AGRWPVKTIHTDNGSNFTSTTVKAAC WWAGIKQEFGIPYNPQSQGVVESMNN ELKKIIGQVRDQAEHLKTAVQMAVFIH NFKRKGGIGGYSAGERIVDIIATDIQTK ELQKQITKIQNFRVYYRDNKDPLWKG PAKLLWKGEGAVVIQDNSDIKVVPRR KAKIIRDYGKQMAGDDCVASRQDED (SEQ ID NO: 1580)	

POL_FIVPE	P16088	Feline immunodeficiency virus	KEFGKLEGGASCSPSESNAASSNAICTS NGGETIGFVNYNKVGTTTTLEKRPEILI FVNGYPIKFLLDTGADITILNRRDFQVK NSIENGRQNMIGVGGGKRGTNYINVH LEIRDENYKTQCIFGNVCVLEDNSLIQP LLGRDNMIKFNIRLVMAQISDKIPVVK VKMKDPNKGPQIKQWPLTNEKIEALTE IVERLEKEGKVKRADSNNPWNTPVFAI KKKSGKWRMLIDFRELNKLTEKGAEV QLGLPHPAGLQIKKQVTVLDIGDAYFT IPLDPDYAPYTAFTLPRKNNAGPGRRF VWCSLPQGWILSPLIYQSTLDNIIQPFIR QNPQLDIYQYMDDIYIGSNLSKKEHKE KVEELRKLLLWWGFETPEDKLQEEPP YTWMGYELHPLTWTIQQKQLDIPEQPT LNELQKLAGKINWASQAIPDLSIKALT NMMRGNQNLNSTRQWTKEARLEVQK AKKAIEEQVQLGYYDPSKELYAKLSLV GPHQISYQVYQKDPEKILWYGKMSRQ KKKAENTCDIALRACYKIREESIIRIGK EPRYEIPTSREAWESNLINSPYLKAPPP EVEYIHAALNIKRALSMIKDAPIPGAET WYIDGGRKLGKAAKAAYWTDTGKW RVMDLEGSNQKAEIQALLLALKAGSE EMNIITDSQYVINIILQQPDMMEGIWQE VLEELEKKTAIFIDWVPGHKGIPGNEE VDKLCQTMMIIEGDGILDKRSEDAGYD LLAAKEIHLLPGEVKVIPTGVKLMLPK GYWGLIIGKSSIGSKGLDVLGGVIDEG YRGEIGVIMINVSRKSITLMERQKIAQL IILPCKHEVLEQGKVVMDSERGDNGY GSTGVFSSWVDRIEEAEINHEKFHSDP QYLRTEFNLPKMVAEEIRRKCPVCRIIG EQVGGQLKIGPGIWQMDCTHFDGKIIL VGIHVESGYIWAQIISQETADCTVKAV LQLLSAHNVTELQTDNGPNFKNQKME GVLNYMGVKHKFGIPGNPQSQALVEN VNHTLKVWIQKFLPETTSLDNALSLAV HSLNFKRRGRIGGMAPYELLAQQESLR IQDYFSAIPQKLQAQWIYYKDQKDKK WKGPMRVEYWGQGSVLLKDEEKGYF LIPRRHIRRVPEPCALPEGDE (SEQ ID NO: 1581)	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

POL_EIAVY	P03371	Equine infectious anemia virus	TAWTFLKAMQKCSKKREARGSREAPE TNFPDTTEESAQQICCTRDSSDSKSVPR SERNKKGIQCQGEGSSRGSQPGQFVGV TYNLEKRPTTIVLINDTPLNVLLDTGAD TSVLTTAHYNRLKYRGRKYQGTGIIGV GGNVETFSTPVTIKKKGRHIKTRMLVA DIPVTILGRDILQDLGAKLVLAQLSKEI KFRKIELKEGTMGPKIPQWPLTKEKLE GAKETVQRLLSEGKISEASDNNPYNSPI FVIKKRSGKWRLLQDLRELNKTVQVG TEISRGLPHPGGLIKCKHMTVLDIGDA YFTIPLDPEFRPYTAFTIPSINHQEPDKR YVWKCLPQGFVLSPYIYQKTLQEILQP FRERYPEVQLYQYMDDLFVGSNGSKK QHKELIIELRAILQKGFETPDDKLQEVP PYSWLGYQLCPENWKVQKMQLDMVK NPTLNDVQKLMGNITWMSSGVPGLTV KHIAATTKGCLELNQKVIWTEEAQKEL EENNEKIKNAQGLQYYNPEEEMLCEV EITKNYEATYVIKQSQGILWAGKKIMK ANKGWSTVKNLMLLLQHVATESITRV GKCPTFKVPFTKEQVMWEMQKGWYY SWLPEIVYTHQVVHDDWRMKLVEEPT SGITIYTDGGKQNGEGIAAYVTSNGRT KQKRLGPVTHQVAERMAIQMALEDTR DKQVNIVTDSYYCWKNITEGLGLEGP QNPWWPIIQNIREKEIVYFAWVPGHKG IYGNQLADEAAKIKEEIMLAYQGTQIK EKRDEDAGFDLCVPYDIMIPVSDTKIIP TDVKIQVPPNSFGWVTGKSSMAKQGL LINGGIIDEGYTGEIQVICTNIGKSNIKLI EGQKFAQLIILQHHSNSRQPWDENKIS QRGDKGFGSTGVFWVENIQEAQDEHE NWHTSPKILARNYKIPLTVAKQITQECP HCTKQGSGPAGCVMRSPNHWQADCT HLDNKIILHFVESNSGYIHATLLSKENA LCTSLAILEWARLFSPKSLHTDNGTNF VAEPVVNLLKFLKIAHTTGIPYHPESQG IVERANRTLKEKIQSHRDNTQTLEAAL QLALITCNKGRESMGGQTPWEVFITNQ AQVIHEKLLLQQAQSSKKFCFYKIPGE HDWKGPTRVLWKGDGAVVVNDEGK GIIAVPLTRTKLLIKPN (SEQ ID NO: 1582)	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

POL_ BIV29	P19560	Bovine immunodeficiency virus	MKRRELEKKLRKVRVTPQQDKYYTIG NLQWAIRMINLMGIKCVCDEECSAAE VALIITQFSALDLENSPIRGKEEVAIKNT LKVFWSLLAGYKPESTETALGYWEAF TYREREARADKEGEIKSIYPSLTQNTQ NKKQTSNQTNTQSLPAITTQDGTPRFD PDLMKQLKIWSDATERNGVDLHAVNI LGVITANLVQEEIKLLLNSTPKWRLDV QLIESKVREKENAHRTWKQHHPEAPK TDEIIGKGLSSAEQATLISVECRETFRQ WVLQAAMEVAQAKHATPGPINIHQGP KEPYTDFINRLVAALEGMAAPETTKEY LLQHLSIDHANEDCQSILRPLGPNTPME KKLEACRVVGSQKSKMQFLVAAMKE MGIQSPIPAVLPHTPEAYASQTSGPEDG RRCYGCGKTGHLKRNCKQQKCYHCG KPGHQARNCRSKNREVLLCPLWAEEP TTEQFSPEQHEFCDPICTPSYIRLDKQPF IKVFIGGRWVKGLVDTGADEVVLKNI HWDRIKGYPGTPIKQIGVNGVNVAKR KTHVEWRFKDKTGIIDVLFSDTPVNLF GRSLLRSIVTCFTLLVHTEKIEPLPVKV RGPGPKVPQWPLTKEKYQALKEIVKD LLAEGKISEAAWDNPYNTPVFVIKKKG TGRWRMLMDFRELNKITVKGQEFSTG LPYPPGIKECEHLTAIDIKDAYFTIPLHE DFRPFTAFSVVPVNREGPIERFQWNVL PQGWVCSPAIYQTTTQKIIENIKKSHPD VMLYQYMDDLLIGSNRDDHKQIVQEI RDKLGSYGFKTPDEKVQEERVKWIGF ELTPKKWRFQPRQLKIKNPLTVNELQQ LVGNCVWVQPEVKIPLYPLTDLLRDKT NLQEKIQLTPEAIKCVEEFNLKLKDPE WKDRIREGAELVIKIQMVPRGIVFDLL QDGNPIWGGVKGLNYDHSNKIKKILRT MNELNRTVVIMTGREASFLLPGSSEDW EAALQKEESLTQIFPVKFYRHSCRWTSI CGPVRENLTTYYTDGGKKGKTAAAVY WCEGRTKSKVFPGTNQQAELKAICMA LLDGPPKMNIITDSRYAYEGMREEPET WAREGIWLEIAKILPFKQYVGVGWVP AHKGIGGNTEADEGVKKALEQMAPCS PPEAILLKPGEKQNLETGIYMQGLRPQS FLPRADLPVAITGTMVDSELQLQLLNI GTEHIRIQKDEVFMTCFLENIPSATEDH ERWHTSPDILVRQFHLPKRIAKEIVARC	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661

			QECKRTTTSPVRGTNPRGRFLWQMDN THWNKTIIWVAVETNSGLVEAQVIPEE TALQVALCILQLIQRYTVLHLHSDNGP CFTAHRIENLCKYLGITKTTGIPYNPQS QGVVERAHRDLKDRLAAYQGDCETV EAALSLALVSLNKKRGGIGGHTPYEIY LESEHTKYQDQLEQQFSKQKIEKWCY VRNRRKEWKGPYKVLWDGDGAAVIE EEGKTALYPHRHMRFIPPPDSDIQDGSS (SEQ ID NO: 1583)	

A0A142BKH1_ALV	A0A142BKH1	Avian leukosis and sarcoma virus	TVALHLAIPLKWKPDHTPVWIDQWPL PEGKLVALTQLVEKELQLGHIEPSLSC WNTPVFVIRKASGSYRLLHDLRAVNA KLVPFGAVQQGAPVLSALPRGWPLMV LDLKDCFFSIPLAEQDREAFAFTLPSVN NQAPARRFQWKVLPQGMTCSPTICQL VVGQVLEPLRLKHPSLRMLHYMDDLL LAASSHDGLEAAGEEVISTLERAGFTIS PDKIQREPGVQYLGYKLGSTYVAPVGL VAEPRIATLWDVQKLVGSLQWLRPAL GIPPRLMGPFYEQLRGSDPNEAREWNL DMKMAWREIVQLSTTAALERWDPALP LEGAVARCEQGAIGVLGQGLSTHPRPC LWLFSTQPTKAFTAWLEVLTLLITKLR ASAVRTFGKEVDVLLLPACFREDLPLP EGILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGASVQQ LEARAVAMALLLWPTTPTNVVTDSAF VAKMLLKMGQEGVPSTAAAFILEDAL SQRSAMAAVLHVRSHSEVPGFFTEGN DVADSQATFQAYPLREAKDLHTALHI GPRALSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQTDFTL EPRMAPRSWLAVTVATASSAIVVTQH GRVTSVAARHHWATAIAVLGRPKAIK TDNGSCFTSKSTREWLARWGIAHTTGI PGNSQGQAMVERANRLLKDKIRVLAE GDGFMKRIPTGKQGELLAKAMYALNH FERGENTKTPIQKHWRPTVLTEGPPVKI RIETGEWEKGWNVLVWGRGYAAVKN RDTDKIIWVPSRKVKPDITQKDELTKK DEASPLFAGISDWAPWKGEQEGL (SEQ ID NO: 1584)	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

Table 32: InterPro descriptions of signatures present in reverse transcriptases in Table 30 (monomeric viral RTs) and Table 31 (dimeric viral RTs).

Signature	Database	Short Name	Description

cd01645	CDD	*RT_Rtv*	RT_Rtv: Reverse transcriptases (RTs) from retroviruses (Rtvs). RTs catalyze the conversion of single-stranded RNA into double-stranded viral DNA for integration into host chromosomes. Proteins in this subfamily contain long terminal repeats (LTRs) and are multifunctional enzymes with RNA-directed DNA polymerase, DNA directed DNA polymerase, and ribonuclease hybrid (RNase H) activities. The viral RNA genome enters the cytoplasm as part of a nucleoprotein complex, and the process of reverse transcription generates in the cytoplasm forming a linear DNA duplex via an intricate series of steps. This duplex DNA is colinear with its RNA template, but contains terminal duplications known as LTRs that are not present in viral RNA. It has been proposed that two specialized template switches, known as strand-transfer reactions or "jumps", are required to generate the LTRs. [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]

cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
PF00078	Pfam	*RVT_1*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. [PMID: 1698615]

IPR000477	InterPro	*RT_dom*	The use of an RNA template to produce DNA, for integration into the host genome and exploitation of a host cell, is a strategy employed in the replication of retroid elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse trancriptase (RT) (2.7.7.49). Reverse transcriptase occurs in a variety of mobile elements, including retrotransposons, retroviruses, group II introns [PMID: 12758069], bacterial msDNAs, hepadnaviruses, and caulimoviruses. Retroviral reverse transcriptase is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. The discovery of retroelements in the prokaryotes raises intriguing questions concerning their roles in bacteria and the origin and evolution of reverse transcriptases and whether the bacterial reverse transcriptases are older than eukaryotic reverse transcriptases [PMID: 8828137]. Several crystal structures of the reverse transcriptase (RT) domain have been determined [PMID: 1377403].
IPR043502	InterPro	*DNA/RNA polymerase superfamily*	This entry represents the DNA/RNA polymerase superfamily, which includes DNA polymerase I, reverse transcriptase, T7 RNA polymerase, lesion bypass DNA polymerase (Y-family), RNA-dependent RNA-polymerase and dsRNA phage RNA-dependent RNA-polymerase. These enzymes share a similar protein fold at their active site, which resembles the palm subdomain of the right-hand-shaped polymerases. [PMID: 26931141]
SSF56672	Superfamily	*DNA/RNA polymerases*	This superfamily comprises DNA polymerases and RNA polymerases
PF06817	Pfam	*RVT_thumb*	This domain is known as the thumb domain. It is composed of a four helix bundle [PMID:1377403].

IPR010661	InterPro	*RVT_thumb*	This domain is known as the thumb domain. It is composed of a four helix bundle. Reverse transcriptase converts the viral RNA genome into double-stranded viral DNA. Reverse transcriptase often occurs in a polyprotein; with integrase, ribonuclease H and/or protease, which is cleaved before the enzyme takes action. The impact of antiretroviral treatment on the first 400 amino acids of HIV reverse transcriptase is good. Little is known, however, of the antiretroviral drug impact on the C-terminal domains of Pol, which includes the thumb, connection and RNase H. Evidence suggests that these might be well conserved domains. [PMID:1377403, PMID:18335052]
PF06815	Pfam	*RVT_connect*	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
IPR010659	InterPro	*RVT_connect*	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]

Endonuclease domain:

In certain embodiments, the endonuclease/DNA binding domain of an APE-type retrotransposon or the endonuclease domain of an RLE-type retrotransposon can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a Gene Writer system described herein. In some embodiments the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the endonuclease element is a heterologous endonuclease element, such as Fok1 nuclease, a type-II restriction l-like endonuclease (RLE-type nuclease), or another RLE-type endonuclease (also known as REL). In some embodiments the heterologous endonuclease activity has nickase activity and does not form double stranded breaks. In some embodiments, the heterologous endonuclease is a CRISPR-associated nuclease, e.g., Cas9, or a CRISPR-associated nuclease with nickase activity, e.g., a Cas9 nickase. The amino acid sequence of an endonuclease domain of a Gene Writer system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain of a retrotransposon whose DNA sequence is referenced in Table 1, 2, or 3. A person having ordinary skill in the art is capable of identifying endounclease domains based upon homology to other known endonuclease domains using tools as Basic Local Alignment Search Tool (BLAST). In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). For example, a Gene Writer polypeptide described herein may comprise a reverse transcriptase domain from an APE- or RLE-type retrotransposon and an endonuclease domain that comprises Fok1 or a functional fragment thereof. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease

activity. In still other embodiments, endonuclease domains are modified to remove any latent DNA-sequence specificity.

In addition to the target-site nick that is needed to initiate target-primed reverse transcription, supplemental endonuclease activity may be beneficial for improving the resolution of the integration event (Anzalone et al., Nature 576, 149-157 (2019)). In some embodiments, the endonuclease element of the polypeptide provides the nick for initiating target-primed reverse transcription and an additional heterologous domain of the polypeptide provides additional endonuclease activity. In some embodiments, the additional endonuclease activity is provided by a nickase. In some embodiments, the additional endonuclease activity may be provided by a heterologous DNA-binding element that also possesses endonuclease activity, e.g., a Cas9 nickase. In some embodiments, the additional endonuclease activity may be contained within the first Gene Writer polypeptide. In some embodiments, the additional endonuclease activity may be provided by a separate polypeptide.

In some embodiments, a Gene Writer polypeptide described herein comprises an endonuclease domain that cleaves at a predefined location in a target DNA sequence, e.g.. as measured using an assay of Example 32 herein. In some embodiments, the endonuclease domain cleaves at a GG site in a target DNA sequence. In some embodiments, the endonuclease domain cleaves at an AAGG site in a target DNA sequence. In some embodiments, a target DNA sequence described herein comprises a GG or AAGG motif, e.g., a naturally occurring motif in the human genome.

DNA binding domain:

In certain aspects, the DNA-binding domain of a Gene Writer polypeptide described herein is selected, designed, or constructed for binding to a desired host DNA target sequence. In certain embodiments, the DNA-binding domain of the engineered RLE is a heterologous DNA-binding protein or domain relative to a native retrotransposon sequence. In some embodiments the heterologous DNA binding element is a zinc-finger element or a TAL effector element, e.g., a zinc-finger or TAL polypeptide or functional fragment thereof. In some embodiments the heterologous DNA binding element is a sequence-guided DNA binding element, such as Cas9, Cpf1, or other CRISPR-related protein that has been altered to have no endonuclease activity. In

some embodiments the heterologous DNA binding element retains endonuclease activity. In some embodiments, the heterologous DNA binding element retains only single-stranded DNA cleavage activity, e.g., is a DNA nickase, e.g., is a Cas9 nickase. In some embodiments the heterologous DNA binding element with endonuclease activity replaces the endonuclease element of the polypeptide. In some embodiments, the heterologous DNA binding element with endonuclease activity supplements the endonuclease element of the polypeptide, e.g., causes an additional nick at the target site. In specific embodiments, the heterologous DNA-binding domain can be any one or more of Cas9, TAL domain, ZF domain, Myb domain, combinations thereof, or multiples thereof. In certain embodiments, the heterologous DNA-binding domain is a DNA binding domain of a retrotransposon described in a table herein. A person having ordinary skill in the art is capable of identifying DNA binding domains based upon homology to other known DNA binding domains using tools as Basic Local Alignment Search Tool (BLAST). In still other embodiments, DNA-binding domains are modified, for example by site-specific mutation, increasing or decreasing DNA-binding elements (for example, number and/or specificity of zinc fingers), etc., to alter DNA-binding specificity and affinity. In some embodiments the DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells.

In some embodiments, a polypeptide described herein comprises a mutation in a DNA binding domain. In some embodiments, the mutation reduces or abrogates DNA-binding activity of the DNA binding domain, e.g., to less than 50%, 40%, 30%, 20%, 10%, 5%, 2%, or 1% of the corresponding wild-type sequence, e.g., in an assay of Example 30. The mutation may be, e.g., in a ZF1 domain, a ZF2 domain, or a c-myb domain. The mutation may be a point mutation. The mutation may be in a C residue (e.g., C to S), for instance in a C residue in a ZF1 or ZF2 domain; in an R residue (e.g., R to A), for instance in an R residue in a c-myb domain; or in a W residue (e.g., W to A), for instance in a W residue in a c-myb domain; or any combination thereof. In some embodiments, the polypeptide ecomprising a mutation in a DNA binding domain further comprises a heterologous DNA binding domain.

In some embodiments, a naturally occurring AAGG sequence in the genome is used as a seed for retargeting an R2 retrotransposase-based Gene Writing system, wherein the DNA binding domain is mutated or replaced with a heterologous DNA binding domain such that the binding of the Gene Writer polypeptide to the new target site results in the proper positioning of

the endonuclease domain to the AAGG motif to enable endonuclease activity. In some embodiments, a target DNA sequence described herein comprises a motif recognized by an endonuclease domain (e.g., a GG or AAGG motif), e.g., a naturally occurring motif in the human genome. In some embodiments, a GeneWriter comprises a DNA binding domain (e.g., a heterologous DNA binding domain) that binds near the motif recognized by the endonuclease domain, e.g., in such a way that the endonuclease domain of the GeneWriter is positioned to cleave the motif. In some embodiments, the DNA binding domain binds a site that is within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 nucleotides of the motif recognized by an endonuclease domain (e.g., the GG or AAGG motif). The DNA binding domain may bind a site that is upstream or downstream of the GG or AAGG motif. The DNA binding domain may bind a site that is in the same orientation or the reverse complement orientation compared ot the motif recognized by an endonuclease domain (e.g., the GG or AAGG motif). In some embodiments, a retargeted GeneWriter polypeptide comprises (i) an endonuclease domain that recognizes a motif, and (ii) a heterologous DNA binding domain that recognizes a genomic DNA sequence. In some embodiments, the motif is about 30-80, 40-70, 50-60, or 55 nt upstream of the genomic DNA sequence, wherein optionally the motif and the genomic DNA sequence are in the same orientation. In some embodiments, the motif is about 10-30, 15-25, or 20 nt downtream of the genomic DNA sequence, wherein optionally the motif is in the reverse orientation to the genomic DNA sequence.In some embodiments, the DNA binding domain comprises a meganuclease domain (e.g., as described herein, e.g., in the endonuclease domain section), or a functional fragment thereof. In some embodiments, the meganuclease domain possesses endonuclease activity, e.g., double-strand cleavage and/or nickase activity. In other embodiments, the meganuclease domain has reduced activity, e.g., lacks endonuclease activity, e.g., the meganuclease is catalytically inactive. In some embodiments, a catalytically inactive meganuclease is used as a DNA binding domain, e.g., as described in Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012), incorporated herein by reference in its entirety. In embodiments, the DNA binding domain comprises one or more modifications relative to a wild-type DNA binding domain, e.g., a modification via directed evolution, e.g., phage-assisted continuous evolution (PACE).

In certain aspects of the present invention, the host DNA-binding site integrated into by the Gene Writer system can be in a gene, in an intron, in an exon, an ORF, outside of a coding

region of any gene, in a regulatory region of a gene, or outside of a regulatory region of a gene. In other aspects, the engineered RLE may bind to one or more than one host DNA sequence.

In some embodiments, a Gene Writing system is used to edit a target locus in multiple alleles. In some embodiments, a Gene Writing system is designed to edit a specific allele. For example, a Gene Writing polypeptide may be directed to a specific sequence that is only present on one allele, e.g., comprises a template RNA with homology to a target allele, e.g., a gRNA or annealing domain, but not to a second cognate allele. In some embodiments, a Gene Writing system can alter a haplotype-specific allele. In some embodiments, a Gene Writing system that targets a specific allele preferentially targets that allele, e.g., has at least a 2, 4, 6, 8, or 10-fold preference for a target allele.

In certain embodiments, a Gene Writer™ gene editor system RNA further comprises an intracellular localization sequence, e.g., a nuclear localization sequence. The nuclear localization sequence may be an RNA sequence that promotes the import of the RNA into the nucleus. In certain embodiments the nuclear localization signal is located on the template RNA. In certain embodiments, the retrotransposase polypeptide is encoded on a first RNA, and the template RNA is a second, separate, RNA, and the nuclear localization signal is located on the template RNA and not on an RNA encoding the retrotransposase polypeptide. While not wishing to be bound by theory, in some embodiments, the RNA encoding the retrotransposase is targeted primarily to the cytoplasm to promote its translation, while the template RNA is targeted primarily to the nucleus to promote its retrotransposition into the genome. In some embodiments the nuclear localization signal is at the 3' end, 5' end, or in an internal region of the template RNA. In some embodiments the nuclear localization signal is 3' of the heterologous sequence (e.g., is directly 3' of the heterologous sequence) or is 5' of the heterologous sequence (e.g., is directly 5' of the heterologous sequence). In some embodiments the nuclear localization signal is placed outside of the 5' UTR or outside of the 3' UTR of the template RNA. In some embodiments the nuclear localization signal is placed between the 5' UTR and the 3' UTR, wherein optionally the nuclear localization signal is not transcribed with the transgene (e.g., the nuclear localization signal is an anti-sense orientation or is downstream of a transcriptional termination signal or polyadenylation signal). In some embodiments the nuclear localization sequence is situated inside of an intron. In some embodiments a plurality of the same or different nuclear localization signals are in the RNA, e.g., in the template RNA. In some embodiments the nuclear localization signal is less

than 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or 1000 bp in legnth. Various RNA nuclear localization sequences can be used. For example, Lubelsky and Ulitsky, *Nature* 555 (107-111), 2018 describe RNA sequences which drive RNA localization into the nucleus. In some embodiments, the nuclear localization signal is a SINE-derived nuclear RNA localization (SIRLOIN) signal. In some embodiments the nuclear localization signal binds a nuclear-enriched protein. In some embodiments the nuclear localization signal binds the HNRNPK protein. In some embodiments the nuclear localization signal is rich in pyrimidines, e.g., is a C/T rich, C/U rich, C rich, T rich, or U rich region. In some embodiments the nuclear localization signal is derived from a long non-coding RNA. In some embodiments the nuclear localization signal is derived from MALAT1 long non-coding RNA or is the 600 nucleotide M region of MALAT1 (described in Miyagawa et al., *RNA* 18, (738-751), 2012). In some embodiments the nuclear localization signal is derived from BORG long non-coding RNA or is a AGCCC motif (described in Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014). In some embodiments the nuclear localization sequence is described in Shukla et al., *The EMBO Journal* e98452 (2018). In some embodiments the nuclear localization signal is derived from a non-LTR retrotransposon, an LTR retrotransposon, retrovirus, or an endogenous retrovirus.

In some embodiments, a polypeptide described herein comprises one or more (e.g., 2, 3, 4, 5) nuclear targeting sequences, for example, a nuclear localization sequence (NLS), e.g., as described above. In some embodiments, the NLS is a bipartite NLS. In some embodiments, an NLS facilitates the import of a protein comprising an NLS into the cell nucleus. In some embodiments, the NLS is fused to the N-terminus of a Gene Writer described herein. In some embodiments, the NLS is fused to the C-terminus of the Gene Writer. In some embodiments, the NLS is fused to the N-terminus or the C-terminus of a Cas domain. In some embodiments, a linker sequence is disposed between the NLS and theneighboring domain of the Gene Writer.

In some embodiments, an NLS comprises the amino acid sequence 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), or KRGINDRNFWRGENGRKTR(SEQ ID NO: 1590), KRPAATKKAGQAKKKK(SEQ ID NO: 1591), or a functional fragment or variant thereof. Exemplary NLS sequences are also described

in PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises an amino acid sequence as disclosed in Table 39. An NLS of this table may be utilized with one or more copies in a polypeptide in one or more locations in a polypeptide, e.g., 1, 2, 3 or more copies of an NLS in an N-terminal domain, within peptide domains, between peptide domains, in a C-terminal domain, or in a combination of locations, in order to improve subcellular localization to the nucleus. Multiple unique sequences may be used within a single polypeptide. Sequences may be naturally monopartite or bipartite, e.g., having one or two stretches of basic amino acids, or may be used as chimeric bipartite sequences. Sequence references correspond to UniProt accession numbers, except where indicated as SeqNLS for sequences mined using a subcellular localization prediction algorithm (Lin et al BMC Bioinformat 13:157 (2012), incorporated herein by reference in its entirety).

Table 39: Exemplary nuclear localization signals for use in Gene Writing systems

Sequence	Sequence References	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

IINGRKLKLKKSRR RSSQTSNNSFTSRR S	SeqNLS	1834
KAEQERRK	Q8LH59	1835
KEKRKRREELFIEQ KKRK	SeqNLS	1836
KKGKDEWFSRGKK P	P30999	1837
KKGPSVQKRKKT	Q6ZN17	1838
KKKTVINDLLHYK KEK	SeqNLS, P32354	1839
KKNGGKGKNKPSA KIKK	SeqNLS	1840
KKPKWDDFKKKK K	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
KKSTALSRELGKIM RRR	SeqNLS, P32354	1847
KKSYQDPEIIAHSRP RK	Q9U7C9	1848
KKTGKNRKLKSKR VKTR	Q9Z301, O54943, Q8K3T2	1849
KKVSIAGQSGKLW RWKR	Q6YUL8	1850
KKYENVVIKRSPRK RGRPRK	SeqNLS	1851
KNKKRK	SeqNLS	1852
KPKKKR	SeqNLS	1853
KRAMKDDSHGNST SPKRRK	Q0E671	1854
KRANSNLVAAYEK AKKK	P23508	1855
KRASEDTTSGSPPK KSSAGPKR	Q9BZZ5, Q5R644	1856
KRFKRRWMVRKM KTKK	SeqNLS	1857
KRGLNSSFETSPKK VK	Q8IV63	1858
KRGNSSIGPNDLSK RKQRKK	SeqNLS	1859
KRIHSVSLSQSQIDP SKKVKRAK	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
RKRSPKDKKEKDL DGAGKRRKT	Q7RTP6	1914
RKRTPRVDGQTGE NDMNKRRRK	O94851	1915
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270	1916
RLRFRKPKSK	P69469	1917
RQQRKR	Q14980	1918
RRDLNSSFETSPKK VK	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
RRKRSKSEDMDSV ESKRRR	Q7TT18	1924
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8	1925
RRPKGKTLQKRKP K	Q6ZN17	1926
RRRGFERFGPDNM GRKRK	Q63014, Q9DBR0	1927
RRRGKNKVAAQNC RK	SeqNLS	1928
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9	1929
RRRQKQKGGASRR R	SeqNLS	1930
RRRREGPRARRRR	P08313, P10231	1931

RRTIRLKLVYDKCD RSCKIQKKNRNKC QYCRFHKCLSVGM SHNAIRFGRMPRSE KAKLKAE	SeqNLS	1932
RRVPQRKEVSRCR KCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5	1933
RVGGRRQAVECIE DLLNEPGQPLDLSC KRPRP	P03255	1934
RVVKLRIAP	P52639, Q8JMN0	1935
RVVRRR	P70278	1936
SKRKTKISRKTR	Q5RAY1, O00443	1937
SYVKTVPNRTRTYI KL	P21935	1938
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9	1939
TLSPASSPSSVSCPV IPASTDESPGSALNI	SeqNLS	1940
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9	1941
SPKKKRKVE		1942
KRTAD GSEFE SPKKKRKVE		1943
PAAKRVKLD		1944
PKKKRKV		1945
MDSLLMNRRKFLY QFKNVRWAKGRRE TYLC		1946
SPKKKRKVEAS		1947
MAPKKKRKVGIHR GVP		1948

In some embodiments, the NLS is a bipartite NLS. A bipartite NLS typically comprises two basic amino acid clusters separated by a spacer sequence (which may be, e.g., about 10 amino acids in length). A monopartite NLS typically lacks a spacer. An example of a bipartite NLS is the nucleoplasmin NLS, having the sequence KR[PAATKKAGQA]KKKK (SEQ ID NO: 1591), wherein the spacer is bracketed. Another exemplary bipartite NLS has the sequence PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1593). Exemplary NLSs are described in International Application WO2020051561, which is herein incorporated by reference in its entirety, including for its disclosures regarding nuclear localization sequences.

In certain embodiments, a Gene Writer™ gene editor system polypeptide further comprises an intracellular localization sequence, e.g., a nuclear localization sequence and/or a

nucleolar localization sequence. The nuclear localization sequence and/or nucleolar localization sequence may be amino acid sequences that promote the import of the protein into the nucleus and/or nucleolus, where it can promote integration of heterologous sequyence into the genome. In certain embodiments, a Gene Writer gene editor system polypeptide (e.g., a retrotransposase, e.g., a polypeptide according to any of Tables 1, 2, or 3 herein) further comprises a nucleolar localization sequence. In certain embodiments, the retrotransposase polypeptide is encoded on a first RNA, and the template RNA is a second, separate, RNA, and the nucleolar localization signal is encoded on the RNA encoding the retrotransposase polypeptide and not on the template RNA. In some embodiments, the nucleolar localization signal is located at the N-terminus, C-terminus, or in an internal region of the polypeptide. In some embodiments, a plurality of the same or different nucleolar localization signals are used. In some embodiments, the nuclear localization signal is less than 5, 10, 25, 50, 75, or 100 amino acids in length. Various polypeptide nucleolar localization signals can be used. For example, Yang et al., *Journal of Biomedical Science* 22, 33 (2015), describe a nuclear localization signal that also functions as a nucleolar localization signal. In some embodiments, the nucleolar localization signal may also be a nuclear localization signal. In some embodiments, the nucleolar localization signal may overlap with a nuclear localization signal. In some embodiments, the nucleolar localization signal may comprise a stretch of basic residues. In some embodiments, the nucleolar localization signal may be rich in arginine and lysine residues. In some embodiments, the nucleolar localization signal may be derived from a protein that is enriched in the nucleolus. In some embodiments, the nucleolar localization signal may be derived from a protein enriched at ribosomal RNA loci. In some embodiments, the nucleolar localization signal may be derived from a protein that binds rRNA. In some embodiments, the nucleolar localization signal may be derived from MSP58. In some embodiments, the nucleolar localization signal may be a monopartite motif. In some embodiments, the nucleolar localization signal may be a bipartite motif. In some embodiments, the nucleolar localization signal may consist of a multiple monopartite or bipartite motifs. In some embodiments, the nucleolar localization signal may consist of a mix of monopartite and bipartite motifs. In some embodiments, the nucleolar localization signal may be a dual bipartite motif. In some embodiments, the nucleolar localization motif may be a KRASSQALGTIPKRRSSSRFIKRKK (SEQ ID NO: 1530). In some embodiments, the nucleolar localization signal may be derived from nuclear factor-κB-inducing kinase. In some

embodiments, the nucleolar localization signal may be an RKKRKKK motif (SEQ ID NO: 1531) (described in Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004).

Since an endogenous nucleolar localization signal may help drive the Gene Writer polypeptide to the nucleolus for those polypeptides derived from retrotransposons naturally targeting the rDNA, e.g., R1, R2, R4, R8, R9, it may be beneficial to inactivate this signal when retargeting to a site outside of the rDNA. An endogenous nucleolar localization signal (NoLS) can be computationally predicted using a published algorithm trained on validated proteins that localize to the nucleolus (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–7399 (2010)). The predicted NoLS sequence is based on both amino acid sequence, amino acid sequence context, and predicted secondary structure of the retrotransposase. The identified sequence is typically rich with basic amino acids (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–7399 (2010)) and mutating these residues to simple side-chain, non-basic, amino acids or removing them from the polypeptide chain can prevent localization to the nucleolus (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)). In some embodiments, the NoLS sequence is located in the amino acid region of a retrotransposase that is between the reverse transcriptase domain and the restriction-like endonuclease domain. In some embodiments, a predicted NoLS region contains lysine, arginine, histidine, and/or glutamine amino acids and nucleolar localization is inactivated by mutation of one or more of these residues to alanine and/or removal from the polypeptide.

In some embodiments, a nucleic acid described herein (e.g., an RNA encoding a GeneWriter polypeptide, or a DNA encoding the RNA) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a GeneWriter system. For instance, the microRNA binding site can be chosen on the basis that is is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the RNA encoding the GeneWriter polypeptide is present in a non-target cell, it would be bound by the miRNA, and when the RNA encoding the GeneWriter polypeptide is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the RNA encoding the GeneWriter polypeptide may reduce production of the GeneWriter polypeptide, e.g., by

degrading the mRNA encoding the polypeptide or by interfering with translation. Accordingly, the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells. A system having a microRNA binding site in the RNA encoding the GeneWriter polypeptide (or encoded in the DNA encoding the RNA) may also be used in combination with a template RNA that is regulated by a second microRNA binding site, e.g., as described herein in the section entitled "Template RNA component of Gene Writer™ gene editor system." In some embodiments, e.g., for liver indications, a miRNA is selected from Table 4 of WO2020014209, which is hereby incorporated by reference.

In some embodiments, the DNA encoding a Gene Writer polypeptide comprises a promoter sequence, e.g., a tissue specific promoter sequence. In some embodiments, the tissue-specific promoter is used to increase the target-cell specificity of a Gene Writer™ system. For instance, the promoter can be chosen on the basis that it is active in a target cell type but not active in (or active at a lower level in) a non-target cell type. A system having a tissue-specific promoter sequence in the DNA of the polypeptide may also be used in combination with a microRNA binding site, e.g., in the template RNA or a nucleic acid encoding a Gene Writer™ protein, e.g., as described herein. A system having a tissue-specific promoter sequence in the DNA encoding the Gene Writer polypeptide may also be used in combination with a DNA encoding the RNA template driven by a tissue-specific promoter, e.g., to achieve higher levels of RNA template in target cells than in non-target cells. In some embodiments, e.g., for liver indications, a tissue-specific promoter is selected from Table 3 of WO2020014209, which is hereby incorporated by reference.

A skilled artisan can, based on the Accession numbers provided in Tables 1-3 determine the nucleic acid and corresponding polypeptide sequences of each retrotransposon and domains thereof, e.g., by using routine sequence analysis tools as Basic Local Alignment Search Tool (BLAST) or CD-Search for conserved domain analysis. Other sequence analysis tools are known and can be found, e.g., at https://molbiol-tools.ca, for example, at https://molbiol-tools.ca/Motifs.htm. SEQ ID NOs 1-112 align with each row in Table 1, and SEQ ID NOs 113-1015 align with the first 903 rows of Table 2.

Tables 1-3 herein provide the sequences of exemplary transposons, including the amino acid sequence of the retrotransposase, and sequences of 5' and 3' untranslated regions to allow

the retrotransposase to bind the template RNA, and the full transposon nucleic acid sequence. In some embodiments, a 5' UTR of any of Tables 1-3 allows the retrotransposase to bind the template RNA. In some embodiments, a 3' UTR of any of Tables 1-3 allows the retrotransposase to bind the template RNA. Thus, in some embodiments, a polypeptide for use in any of the systems described herein can be a polypeptide of any of Tables 1-3 herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the system further comprises one or both of a 5' or 3' untranslated region of any of Tables 1-3 herein (or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), e.g., from the same transposon as the polypeptide referred to in the preceding sentence, as indicated in the same row of the same table. In some embodiments, the system comprises one or both of a 5' or 3' untranslated region of any of Tables 1-3 herein, e.g., a segment of the full transposon sequence that encodes an RNA that is capable of binding a retrotransposase, and/or the sub-sequence provided in the column entitled Predicted 5' UTR or Predicted 3' UTR.

In some embodiments, a polypeptide for use in any of the systems described herein can be a molecular reconstruction or ancestral reconstruction based upon the aligned polypeptide sequence of multiple retrotransposons. In some embodiments, a 5' or 3' untranslated region for use in any of the systems described herein can be a molecular reconstruction based upon the aligned 5' or 3' untranslated region of multiple retrotransposons. A skilled artisan can, based on the Accession numbers provided herein, align polypeptides or nucleic acid sequences, e.g., by using routine sequence analysis tools as Basic Local Alignment Search Tool (BLAST) or CD-Search for conserved domain analysis. Molecular reconstructions can be created based upon sequence consensus, e.g. using approaches described in *Ivics et al., Cell 1997, 501 – 510 ; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99.* In some embodiments, the retrotransposon from which the 5' or 3' untranslated region or polypeptide is derived is a young or a recently active mobile element, as assessed via phylogenetic methods such as those described in Boissinot et al., Molecular Biology and Evolution 2000, 915-928.

Table 3 (below) shows exemplary Gene Writer proteins and associated sequences from a variety of retrotransposases, identified using data mining. Column 1 indicates the family to which the retrotransposon belongs. Column 2 lists the element name. Column 3 indicates an

accession number, if any. Column 4 lists an organism in which the retrotransposase is found. Column 5 lists the DNA sequence of the retrotransposon. Column 6 lists the predicted 5' untranslated region, and column 7 lists the predicted 3' untranslated region; both are segments of the sequence of column 5 that are predicted to allow the template RNA to bind the retrotransposase of column 8. (It is understood that columns 5-7 show the DNA sequence, and that an RNA sequence according to any of columns 5-7 would typically include uracil rather than thymidine.) Column 8 lists the predicted retrotransposase sequence encoded in the retrotransposon of column 5.

Table 3. Exemplary Gene Writer Proteins and Associated Sequences and Information

1. Family	2. Element	3. Accession	4. Organism	5. DNA Sequence	6. Predicted 5'UTR	7. Predicted 3'UTR	8. Predicted Amino Acid Sequence
R2	R2-1_TG	.	Taeniopygia guttata	GTCTAGTTACAACTGGGCATCGCTGCAGAGATCGCACCTCCTCGTGGTC	GTCTAGTT	TTCAGG	MASCPKPGPPVSAG
				CCGCTGGTAGCCCTTCGAAGGGTGACTAAGTCGATCTCTGCCCCAGGTA	ACAACTGG	TTATTTA	AMSLESGLTTHSVLAI
				CGGAGCCGTTGGGACTCACCAGTCCAACGTAACTCCTGCCTAAATTCGG	GCATCGCT	GATGCT	ERGPNSLANSGSDFG
				TGAAACAAATTCCTCGGTAAAAAGCCCCATGGCTTCTTGCCCGAAACCT	GCAGAGAT	TAGTTTT	GGGLGLPLRLLRVSV
				GGCCCCCCGGTTTCAGCAGGGGCAATGAGTTTGGAAAGTGGACTGACC	CGCACCTC	TGTACCT	GTQTSRSDWVDLVS
				ACCCACTCCGTTCTCGCCATCGAACGTGGTCCCAATTCGTTGGCAAATTC	CTCGTGGT	TTCTTGT	WSHPGPTSKSQQVD
				CGGATCAGACTTTGGGGGGGGGGGTCTGGGGCTACCGTTACGCCTATT	CCCGCTGG	TTTGTTT	LVSLFPKHRVDLLSKN
				GAGGGTATCGGTCGGCACTCAGACCTCCCGCTCCGACTGGGTAGACCTG	TAGCCCTT	AGGATT	DQVDLVAQFLPSKFP
				GTGTCCTGGAGCCACCCAGGACCCACGTCTAAGTCCCAGCAGGTTGACC	CGAAGGG	TTGATA	PNLAENDLALLVNLE
				TGGTGTCTTTATTTCCTAAACACCGGGTTGACCTGTTATCCAAAAACGAC	TGACTAAG	GTGTTA	FYRSDLHVYECVHFA
				CAGGTAGACCTGGTGGCTCAATTTTTACCATCTAAATTTCCCCCCAATTT	TCGATCTC	GTATTTT	AHWEGLSGLPEVYE
				GGCAGAAAATGATTTGGCTTTGCTGGTGAACTTAGAGTTCTACAGATCG	TGCCCCAG	TATATTT	QLAPQPCVGETLHSS
				GATTTGCATGTGTATGAGTGTGTTCATTTTGCTGCACATTGGGAGGGAT	GTACGGA	TTGTAC	LPRDSELFVPEEGSSE
				TAAGTGGTTTGCCTGAGGTGTATGAACAACTTGCACCACAACCGTGTGT	GCCGTTGG	GATTGC	KESEDAPKTSPPTPG
				GGGAGAAACTTTACATTCTAGCCTCCCACGAGACAGTGAACTGTTTGTG	GACTCACC	ATAATG	KHGLEQTGEEKVMV
				CCTGAAGAGGGGAGCAGCGAGAAGGAGAGCGAGGACGCGCCAAAAAC	AGTCCAAC	TTCTTTT	TVPDKNPPCPCCGTR
				ATCTCCTCCGACGCCTGGGAAACATGGTTTGGAACAGACTGGGGAGGA	GTAACTCC	TTATACA	VNSVLNLIEHLKVSH
				AAAAGTGATGGTGACTGTTCCTGACAAAAATCCACCTTGTCCTTGCTGTG	TGCCTAAA	GTTCTGT	GKRGVCFRCAKCGKE
				GTACCCGGGTAAACTCTGTGTTGAATCTGATTGAACATCTGAAAGTGTC	TTCGGTGA	TTTAATA	NSNYHSVVCHFPKCR
				ACACGGGAAAAGGGGGGTTTGTTTTCGGTGTGCAAAATGTGGAAAGGA	AACAAATT	AAATAG	GPETEKAPAGEWICE
				AAATAGTAACTATCACAGTGTTGTTTGTCATTTTCCAAAATGCAGGGGTC	CCTCGGTA	ACGATA	VCNRDFTTKIGLGQH
				CAGAGACGGAGAAAGCCCCAGCTGGGGAGTGGATTTGTGAGGTATGC	AAAAGCCC	GCTAGA	KRLAHPAVRNQERIV
				AACAGAGATTTTACAACCAAAATTGGCCTGGGACAACACAAGAGATTG	C (SEQ ID	GACGTT	ASQPKETSNRGAHKR
				GCACACCCAGCAGTGAGAAATCAGGAAAGGATCGTTGCTTCCCAACCG	NO: 1140)	AGGGCA	CWTKEEEELLIRLEAQ
				AAAGAAACATCAAATAGAGGTGCTCACAAAAGGTGCTGGACAAAGGAG		GCCACA	FEGNKNINKLIAEHIT
				GAGGAAGAATTACTAATAAGACTGGAGGCTCAGTTCGAGGGAAACAAA		AGCCAG	TKTAKQISDKRRLLSR
				AATATTAATAAGCTTATTGCAGAACACATAACCACCAAAACAGCTAAGC		TTAGGT	KPAEEPREEPGTCHH
				AGATCAGTGACAAAAGGCGATTGCTGTCCAGAAAGCCAGCAGAGGAGC		AGCGGA	TRRAAASLRTEPEMS
				CACGTGAGGAGCCTGGAACGTGTCATCACACCAGGAGAGCAGCTGCGA		TAGTAG	HHAQAEDRDNGPG
				GCCTGAGAACGGAGCCTGAGATGAGTCATCACGCCCAGGCAGAGGACA		GTAGGA	RRPLPGRAAAGGRT
				GAGATAATGGACCTGGGAGACGCCCTCTGCCAGGCAGGGCAGCTGCCG		ACAGAC	MDEIRRHPDKGNGQ

		GAGGGAGAACAATGGACGAGATAAGACGCCACCCTGATAAGGGCAAC		TTTTACT	QRPTKQKSEEQLQAY
		GGACAGCAGAGACCCACCAAGCAAAAATCAGAAGAACAGCTGCAGGCT		ATTTCAT	YKKTLEERLSAGALNT
		TACTATAAAAAGACACTAGAGGAACGACTTTCAGCTGGGGCACTTAACA		AACGCG	FPRAFKQVMEGRDIK
		CCTTCCCCCGAGCATTCAAGCAGGTAATGGAAGGCCGGGATATAAAGCT		TCAATTA	LVINQTAQDCFGCLE
		AGTAATCAATCAGACAGCGCAGGACTGCTTCGGATGCCTGGAATCCATA		CCACCT	SISQIRTATRDKKDTV
		AGCCAAATAAGAACGGCAACCCGAGATAAAAAGGACACGGTGACCCGG		GATTTG	TREKHPKKPFQKWMI
		GAGAAACACCCAAAGAAACCTTTTCAGAAGTGGATGAAGGACAGAGCA		GACCAA	KDRAIKKGNYLRFQR
		ATCAAAAAAGGTAATTATCTTCGGTTCCAGCGTTTATTTTATCTTGATAG		TTCACG	LFYLDRGKLAKIILDDI
		AGGGAAACTGGCTAAAATCATTTTAGATGATATTGAATGCTTGTCTTGT		GGATTT	ECLSCDIPLSEIYSVFK
		GACATACCACTCAGTGAAATTTATTCGGTTTTTAAAACAAGATGGGAAA		GTCCAA	TRWETTGSFKSLGDF
		CAACTGGTAGCTTTAAAAGCCTTGGGGACTTTAAAACTTACGGGAAGGC		GGTGGA	KTYGKADNTAFRELIT
		TGACAACACTGCCTTCAGAGAATTAATTACGGCTAAAGAAATTGAGAAA		CGGGCC	AKEIEKNVQEMSKGS
		AATGTGCAGGAAATGAGCAAAGGCTCCGGCTCCCGGTCCAGACGGGATT		ACCTTTA	APGPDGITLGDVVK
		ACTCTTGGGGACGTCGTAAAGATGGATCCCGAGTTTTCCCGGACCATGG		CTTAACC	MDPEFSRTMEIFNL
		AGATTTTCAATTTATGGTTAACAACTGGTAAAATCCCGGACATGGTGAG		CGGAAA	WLTTGKIPDMVRGC
		GGGGTGCAGAACCGTTTTGATTCCAAAATCATCAAAGCCGGATCGTTTG		AGGAAC	RTVLIPKSSKPDRLKDI
		AAAGACATTAATAACTGGAGACCTATCACGATCGGTTCCATCTTGCTGA		ATATATA	NNWRPITIGSILLRLF
		GACTGTTCTCCAGGATTGTAACAGCTAGGCTGAGCAAAGCGTGCCCCCT		ATTTATG	SRIVTARLSKACPLNP
		GAACCCAAGGCAAAGAGGCTTTATCAGAGCGGCGGGATGCTCTGAAAA		TGTGTTC	RQRGFIRAAGCSENL
		CTTAAAACTCCTGCAAACTATAATTTGGTCGGCCAAAAGAGAACACAGA		GATAAA	KLLQTIIWSAKREHRP
		CCACTGGGTGTTGTATTCGTGGACATCGCCAAGGCTTTTGACACCGTAA		(SEQ ID	LGVVFVDIAKAFDTV
		GCCACCAGCACATCATTCATGCTTTGCAGCAAAGAGAGGTGGATCCCCA		NO:	SHQHIIHALQQREVD
		CATCGTCGGTCTGGTGAGCAATATGTACGAGAGAACATCAGTACGTATATC		1263)	PHIVGLVSNMYENIS
		ACCACAAAGAGGAACACACACACAGACAAAATCCAGATCCGGGTTGGA			TYITTKRNTHTDKIQI
		GTAAAGCAGGGTGACCCGATGTCGCCCCTTTTATTTAACCTGGCAATGG			RVGVKQGDPMSPLL
		ACCCTCTATTATGCAAGCTGGAAGAGAGTGGCAAAGGATACCACCGAG			FNLAMDPLLCKLEES
		GACAGAGCAGCATCACAGCGATGGCATTTGCAGACGATCTGGTTTTGCT			GKGYHRGQSSITAM
		GAGCGACTCCTGGGAAAATATGAATACAAATATTAGCATACTGGAGACC			AFADDLVLLSDSWEN
		TTCTGCAATCTGACCGGTCTCAAAACACAGGGGCAAAAGTGCCACGGCT			MNTNISILETFCNLTG
		TTTACATCAAGCCGACAAAGGACTCTTACACCATCAATGACTGCGCTGCC			LKTQGQKCHGFYIKP
		TGGACTATCAACGGCACACCCCTGAACATGATCGACCCCGGCGAATCTG			TKDSYTINDCAAWTI
		AGAAATACCTCGGCCTGCAGTTTGACCCGTGGATTGGAATAGCAAGGTC			NGTPLNMIDPGESEK
		CGGTCTCTCCACAAAACTAGATTTTTGGCTTCAGCGGATCGATCAAGCAC			YLGLQFDPWIGIARS
		CACTTAAACCTCTGCAGAAAACTGATATTCTCAAAACATACACCATCCCT			GLSTKLDFWLQRIDQ
		CGGCTGATCTACATAGCTGACCACTCAGAAGTGAAAACTGCACTACTCG			APLKPLQKTDILKTYTI
		AAACCCTTGACCAGAAGATCCGGACAGCGGTCAAGGAATGGCTTCACCT			PRLIYIADHSEVKTALL
		ACCTCCGTGCACCTGCGATGCCATCCTGTACTCGAGCACGAGAGACGGC			ETLDQKIRTAVKEWL
		GGTTTGGGCATCACCAAATTGGCAGGACTGATCCCCAGCGTGCAGGCCC			HLPPCTCDAILYSSTR

			GTAGACTGCATCGGATCGCACAGTCATCTGACGATACGATGAAATGCTT				DGGLGITKLAGLIPSV
			CATGGAAAAAAGAGAAAATGGAACAGCTGCATAAGAAATTGTGGATTCA				QARRLHRIAQSSDDT
			AGCTGGAGGGGACAGAGAGAACATACCCTCGATTTGGGAAGCACCACC				MKCFMEKEKMEQL
			GTCGAGTGAACCACCAAACAACGTGAGCACAAATTCGGAATGGGAAGC				HKKLWIQAGGDRENI
			ACCGACCCAGAAAGATAAATTTCCAAAGCCTTGCAATTGGAGGAAAAAC				PSIWEAPPSSEPPNN
			GAATTCAAAAAATGGACCAAATTGGCATCCCAAGGCCGCGGAATTGTAA				VSTNSEWEAPTQKD
			ATTTTGAAAGAGACAAAATTAGTAACCATTGGATCCAATACTACAGACG				KFPKPCNWRKNEFKK
			CATACCTCACAGGAAACTCCTCACTGCACTACAACTCAGGGCCAACGTTT				WTKLASQGRGIVNFE
			ACCCCACGAGAGAATTTCTAGCCAGGGGTAGACAAGACCAATACATCAA				RDKISNHWIQYYRRI
			GGCGTGTAGGCACTGCGATGCGGACATTGAATCCTGCGCCCACATCATC				PHRKLLTALQLRANV
			GGCAACTGCCCAGTGACACAGGACGCCCGAATCAAGAGGCACAATTAC				YPTREFLARGRQDQY
			ATCTGCGAACTGCTTCTCGAGGAGGCGAAGAAGAAGGACTGGGTAGTG				IKACRHCDADIESCA
			TTCAAGGAACCGCACATAAGGGATTCCAACAAGGAACTGTACAAACCTG				HIIGNCPVTQDARIKR
			ACCTGATATTTGTGAAGGATGCCCGTGCACTTGTCGTGGATGTGACAGT				HNYICELLLEEAKKKD
			ACGGTATGAAGCAGCCAAATCATCGCTGGAGGAAGCCGCTGCAGAGAA				WVVFKEPHIRDSNKE
			AGTGAGAAAGTACAAACACCTGGAAACGGAAGTAAGACATCTCACGAA				LYKPDLIFVKDARALV
			TGCAAAGGACGTTACTTTTGTGGGCTTTCCCCTAGGAGCGCGCGGGGAA				VDVTVRYEAAKSSLE
			ATGGCACCAAGATAACTTTAAACTTTTGACTGAGCTTGGCCTCTCCAAAT				EAAAEKVRKYKHLET
			CGAGGCAAGTGAAAATGGCAGAGACTTTTTCCACAGTAGCGCTCTTTTC				EVRHLTNAKDVTFVG
			ATCTGTGGACATTGTACATATGTTTGCCAGTAGGGCCAGAAAATCTATG				FPLGARGKWHQDNF
			GTTATGTAATTCAGGTTATTTAGATGCTTAGTTTTTGTACCTTTCTTGTTT				KLLTELGLSKSRQVK
			TGTTTAGGATTTTGATAGTGTTAGTATTTTTATATTTTTGTACGATTGCAT				MAETFSTVALFSSVDI
			AATGTTCTTTTTTATACAGTTCTGTTTTAATAAAATAGACGATAGCTAGA				VHMFASRARKSMV
			GACGTTAGGGCAGCCACAAGCCAGTTAGGTAGCGGATAGTAGGTAGGGA				M (SEQ ID NO:
			ACAGACTTTTACTATTTCATAACGCGTCAATTACCACCTGATTTGGACCA				1016)
			ATTCACGGGATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGA				
			AAAGGAACATATATAATTTATGTGTGTTCGATAAA (SEQ ID NO: 1539)				
R2	R2-1_Gfo	Geospiza fortis	AGACTTAAGTGAGTTTGGTTACAACTGGGCATAGCTGCAGAGACCGCG	AGACTTAA	GGTAGA		VGLCPSPGVDGTHQ
			CCTCCTCGCGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAAGTCGGT	GTGAGTTT	TAATCTT		PNDSFQNFGETNFSV
			CTCTGCCCCAGTCCGGGAGTCGATGGGACTCACCAGCCCAACGATTCCT	GGTTACAA	TGTATA		QVARLVTRNLAPRSV
			TCCAAAATTTCGGTGAAACAAATTTCTCGGTGCAAGTCGCAAGGCTTGT	CTGGGCAT	GTGGGG		RGNGFGSGMATHPV
			CACCCGAAACCTAGCCCCCCGGTCGGTCAGGGGCAACGGGTTCGGAAG	AGCTGCAG	GGGGAT		PADESGHESDPFLVG
			TGGGATGGCCACCCACCCCGTTCCCGCAGACGAATCCGGCCATGAATCT	AGACCGCG	CTCATGT		RSCGQPARLTRQSVG
			GATCCATTCCTTGTAGGGAGGAGCTGCGGACAACCGGCACGCCTTACTA	CCTCCTCG	ACCGGG		TQTSRDDILPSKTTKL
			GGCAATCGGTTGGCACCCAGACCTCCCGAGATGATATTTTACCATCTAA	CGGCCCCG	TTTCTTT		TENELDLLVNFSLELY
			AACCACCAAATTGACAGAGAATGAATTGGACTTGCTGGTGAACTTTTCT	CTGGTAAG	TATTTGA		RSDLQGFVQEGIHFS
			TTAGAATTGTATAGGTCAGATCTGCAGGGATTTGTGCAGGAGGGGATTC	CCCTTAAC	TTTTCAA		VNREVLEGFPEVYEQ
			ATTTTTCTGTGAATAGGGAGGTGTTAGAGGGGTTTCCTGAGGTGTATGA	AGGGTGA	TAAAAC		PAPQPAVGDDLNTSL

		ACAACCTGCACCACAACCGGCAGTAGGGGACGATTTAAACACCAGTCTC	CTAA (SEQ ID NO: 1141)	AGACGG		PPDNNICVLEKGSSE
		CCACCGGACAATAATATATGCGTACTTGAGAAGGGTAGCAGTGAAGCA		TAGCTA		AVEDGTPEVAHPVPE
		GTGGAGGATGGCACACCGGAGGTAGCGCACCCCGTGCCTGAAACCCAG		GGTTCG		TQGKESPNNIVMVTL
		GGCAAAGAGTCACCGAATAACATCGTGATGGTAACTCTTCCCAACAAAA		CAAGGC		PNKNPPCPCCRVRLH
		ATCCACCATGTCCTTGCTGTAGGGTCAGACTGCATTCAGTACTGGCTCTG		AGCCAC		SVLALIEHLKGSHGKK
		ATTGAACATCTTAAGGGGTCGCATGGGAAGAAGAGGGCATGCTTTAGG		AAGCCA		RACFRCVKCGRENFN
		TGTGTCAAGTGTGGGAGGGAGAACTTTAACTATCATAGTACTGTTTGTC		AAGATA		YHSTVCHIAKCKGPK
		ACATCGCAAAATGCAAGGGACCAAAAGTTGAGAAGGCCCCAGTGGGAG		GGTAGG		VEKAPVGEWICEVCG
		AGTGGATCTGTGAGGTATGTGGTAGGGACTTTACAACCAAAATCGGCCT		GTGCTC		RDFTTKIGLGQHKRL
		GGGACAACATAAAAGATTGGCACATCCCTTGGTTAGAAACCAAGAAAG		ATAGTG		AHPLVRNQERIDASQ
		GATCGATGCTTCCCAACCGAAGGAGACATCAAACAGAGGAGCCCACAA		AGTAGG		PKETSNRGAHKRCW
		GAGATGTTGGACAAAAGAGGAGGAGGAGATGCTGATAAAGTTGGAGG		GACAGT		TKEEEEMLIKLEVQFE
		TACAGTTCGAGGGACACAGAAACATCAATAAGCTTATCGCGGAACACTT		GCCTTTT		GHRNINKLIAEHLTTK
		AACAACTAAAACATCCAAACAGATTAGTGATAAAAGGAGACTATTACCC		GATTCA		TSKQISDKRRLLPRKQ
		AGAAAACAATTAACAGATCTAAGTAAGGGAGTGGCTGGACAGAAGGGTG		CAACGC		LTDLSKGVAGQKVLD
		CTGGACCCAGGACTGAGTCATCAACCCCAGCTGGGGGTAGTTGACAAT		GTCAAT		PGLSHQPQLGVVDN
		GGACTTGGTGGGGGTCATCTGCCAGGGGGGCCAGCTGCTGAAGGAAG		ACCATCT		GLGGGHLPGGPAAE
		AACAATAGAGCCATTAGGACACCACCTTGATAAGGATAACGGTCACCGG		GACACG		GRTIEPLGHHLDKDN
		GAAATCGCTGACCAGCACAAGGCAGGGAGGCTGCAGGCCCATTACCGA		GATACC		GHREIADQHKAGRL
		AAGAAGATAAGGAAGCGCCTTTCAGAAGGGATGATTAGCAACTTCCCC		CTTACCG		QAHYRKKIRKRLSEG
		GAAGTATTTGAACAACTACTGGACTGCCAGGAAGCACAACCATTGATCA		GACTTG		MISNFPEVFEQLLDC
		ATCAAGCAGCGCAGGATTGCTTTGGATGCCTGGATTCAGCAAGCCAGAT		TCATGAT		QEAQPLINQAAQDC
		AAGGAAGGCGCTCCGAAAACAGAACACACAGAAAGACCAGGGGGATC		CTCCCA		FGCLDSASQIRKALRK
		AACCCAAAAGACCAGCTCAGAAGTGGATGAAAAAAAAGAGCAGTTAAGA		GACTTG		QNTQKDQGGDQPKR
		GGGGTCACTTCCTCCGCTTTCAGAAATTATTTCATCTTGACAGGGGGGAA		TCCAAG		PAQKWMKKRAVKR
		ATTGGCAAAGATTATTTTGGACGACGTAGAGTGTTTGTCCTGTGATATA		GTGGAC		GHFLRFQKLFHLDRG
		CCACCCAGTGAAATTTATTCGGTATTCAAAGCCCGATGGGAAACACCTG		GGGCCA		KLAKIILDDVECLSCDI
		GACAGTTTGCTGGCCTTGGGGATTTCGAAATTAATAGGAAGGCGAACA		CCTTTAC		PPSEIYSVFKARWETP
		ATAAAGCCTTCAGGGACTTAATTACGGCCAAAGAAATTCTCAAAAATGT		TTAACCC		GQFAGLGDFEINRKA
		GCGGGAGATGACCAAGGGCTCGGCCCCAGGTCCAGATGGGATCGCGCT		GGAAAA		NNKAFRDLITAKEILK
		TGGGGACATCAGGAAGATGGACCCTGAGTACACCCGGACCGCCGAACT		GGAACA		NVREMTKGSAPGPD
		CTTCAACTTATGGTTAACATCTGGTGAGATCCCGGACATGGTGAGGGGG		TATATTA		GIALGDIRKMDPEYT
		TGCAGAACTGTGTTAATCCCCAAATCGTCAAAACCGGAACGCCTGAAGG		ATTATAT		RTAELFNLWLTSGEIP
		ACATCAATAACTGGAGACCCATCACGATTGGATCCATCTTGCTGAGACTT		GTGTTC		DMVRGCRTVLIPKSS
		TTCTCCAGGATCATAACAGCGAGGTTAACAAAGGCGTGCCCCCTCAACC		GGAAAA		KPERLKDINNWRPITI
		CTAGGCAAAGAAGCTTCATCAGTGCGGCAGGATGCTCCGAGAACTTGA		(SEQ ID		GSILLRLFSRIITARLTK
		AGCTCCTGCAAACCATAATTCGGACTGCTAAAAATGAACACAGACCACT		NO:		ACPLNPRQRSFISAA
		GGGTGTTGTATTCGTGGACATCGCCAAGGCCTTTGACACCGTGAGCCAC		1264)		GCSENLKLLQTIIRTA

			CAACACATCATACATGTATTGCAAAGGAGGAGAGTGGACCCCCACATCA			KNEHRPLGVVFVDIA
			TTGGATTGGTGAAAAATATGTACAAAGACATCAGTACGGTTATCACCAC			KAFDTVSHQHIIHVL
			AAAGAAGAACACATACACGGACAAAATCCAGATCCAGGTTGGAGTGAA			QRRRVDPHIIGLVKN
			GCAAGGTGATCCGCTTTCGCCCCTTCTATTCAACCTGGCGATGGACCCCC			MYKDISTVITTKKNTY
			TGTTGTGCAAGCTGGAAGAACACGGCAAAGGATTCCACCGAGGACAGA			TDKIQIQVGVKQGDP
			GCAAGATAACAGCGATGGCATTCGCTGATGACCTGGTCCTGTTGAGCGA			LSPLLFNLAMDPLLCK
			TTCCTGGGAAGACATGAATGCGAACATCAAGATACTGGAGACCTTCTGC			LEEHGKGFHRGQSKI
			GACCTCACCGGTCTCAAAACACAGGGTCAAAAGTGCCACGGCTTCTACA			TAMAFADDLVLLSDS
			TCAAGCCTACAAAGGACTCTTACACCGTCAACAACTGCGCTGCGTGGAC			WEDMNANIKILETFC
			CATCAATGGCACACCCCTGAACATGATCAACCCCGGGGAATCAGAGAAA			DLTGLKTQGQKCHG
			TACCTCGGCCTGCAGTTTGACCCCTGGGTGGGAATTGCAAAGACCAGCC			FYIKPTKDSYTVNNCA
			TCCCCGAAAAACTGGACTTCTGGCTCGAACGCATTGATCGAGCTCCACT			AWTINGTPLNMINP
			CAAACCATTTCAGAAACTGGACATTCTTAAGACATACACCATACCTCGAC			GESEKYLGLQFDPWV
			TGACCTACGTAGCTGACCACACTCAGAGATGAAAGCGGGGGGCCCTTGAAG			GIAKTSLPEKLDFWLE
			CCCTTGACCGGACAATTCGATCGGCGGTCAAGGACTGGCTGCACCTACC			RIDRAPLKPFQKLDIL
			TTCGAGCACCTGTGATGCCATCTTGTACACGAGCATGAAGGACGGTGGT			KTYTIPRLTYVADHSE
			TTGGGAGTGACCAAATTGGTGGGACTGATTCCGAGTGTACAAGCCCGG			MKAGALEALDRTIRS
			AGGCTGCACAGGATTGCGCAGTCACCGGAGGAGACGATGAAAGACTTC			AVKDWLHLPSSTCDA
			CTGGAAAAGGCCCAGATGGAGAAGATGTACGAGAAATTGTGGGTCCAA			ILYTSMKDGGLGVTK
			GCTGGAGGGAAAAGAAAGAGGATGCCGTCAATTTGGGAAGCGCTCCC			LVGLIPSVQARRLHRI
			GGAGGTTGTACCATCCATAGACACAGCCACAACTTCGGAGTGGGAAGC			AQSPEETMKDFLEKA
			ACCGAACCCTAAAAGTAAGTACCCTAGACCTTGTAATTGGCGCAGAAAA			QMEKMYEKLWVQA
			GAATTTAAAAAGTGGACTAAATTAATAGCCCAGGGCTGGGGAATTAGG			GGKRKRMPSIWEAL
			TGTTTTAAGGGGGACAAAATTAGTAACAATTGGATTCGACATTATAGAT			PEVVPSIDTATTSEW
			ACATACCTCACAGGAAACTTCTCACTGCCATACAGCTCCGGGCCAGTGT			EAPNPKSKYPRPCN
			GTACCCCACAAGGGAATTTCTCGCGCGGGGGAGGGAAGATAACTGTGT			WRRKEFKKWTKLIA
			TAAGTCTTGTAGGCACTGTGAGGCGGCAGAGGAGTCCTGTGCCCACATC			QGWGIRCFKGDKISN
			ATCGGCATGTGTCCAGTCGTGAGGGATGCCCGAATCAAGAGGCACAAT			NWIRHYRYIPHRKLLT
			CGCATTTGCGAGAGGCTGATGGAGGAGGCGGGGAAGAGGGACTGGAC			AIQLRASVYPTREFLA
			GGTGTTTCAGGAGCCGCACATAAGGGACGTCACCAAGGAACTGTACAA			RGREDNCVKSCRHCE
			ACCGGACTTGATATTCGTGAAAGAAGGCCTTGCACTTGTTGTGGATGTT			AAEESCAHIIGMCPV
			ACAATACGGTTCGAGTCAACCAAGACAACGTTGGAGGAGGCTGCTGCA			VRDARIKRHNRICERL
			GAGAAGGTGAACAAGTACAAACATCTGGAGACCGAAGTACGGAACCTC			MEEAGKRDWTVFQE
			ACCAACGCTAAGGACGTTATCTTTATGGGGTTTCCCCTTGGAGCGCGGG			PHIRDVTKELYKPDLI
			GACAATGGTACAATAAGAACTTTGAACTTTTGGACACTCTTGGCCTCCCC			FVKEGLALVVDVTIRF
			AGATCGAGGCAGGACATTATTGCAAAGACTTTATCCACGGACGCGCTCA			ESTKTTLEEAAAEKVN
			TTTCATCTGTGGACATTATACATATGTTTGCCAGTAGAGGCAGAAGACA			KYKHLETEVRNLTNA
			GCATGCTTAGGGTAGATAATCTTTGTATAGTGGGGGGGGGATCTCATGTA			KDVIFMGFPLGARG

			CCGGGTTTCTTTTATTTGATTTTCAATAAAACAGACGGTAGCTAGGTTCG CAAGGCAGCCACAAGCCAAAGATAGGGTAGGGTGCTCATAGTGAGTAGG GACAGTGCCTTTTGATTCACAACGCGTCAATACCATCTGACACGGATACC CTTACCGGACTTGTCATGATCTCCCAGACTTGTCCAAGGTGGACGGGCC ACCTTTACTTAACCCGGAAAAGGAACATATATTAATTATATGTGTTCGGA AAA (SEQ ID NO: 1540)				QWYNKNFELLDTLGL PRSRQDIIAKTLSTDA LISSVDIIHMFASRGR RQHA (SEQ ID NO: 1386)
R2	R2-1_ZA	.	Zonotrichia albicollis	CGACTTGAGAAGGTCTGGTTACAACTGGGCATAGCTGCAGAGATCGCG CCTCCTCGTGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAAGTCGAT CTCTGCCCCAGTCCAGGAGCCGCTGGGTTTCACCAGCCCAGCGATTCCTT CCAAATTCGGTGAAACAAATTCCTCGGTAAAAGCCGCGTGGCTTATTGC CTGAAACCTGGCCCCCCGGTTTCAGACAGGGGCAAAGAGTTCGGAAGT GGACTGACCACCCACCCCGAACCCGAGAGCGAATCTGGTCATGACCCAA CTGTCCCAAATCCTGGTCCGTCTCTTGGAGCGGGGGAAGGTGCACAGCC ACTACCCTTACTCAGGGTATCGGTGGGCACCCAAACCTGTGAAGAGGAC TTTATAACATCTAGACCAACCAAATTACCCGGAATTGAATCAGAATTAGG CCCGCTGGTGAAGTTTTCTTTAGAGGTTTACAGGTCAGATCTTAAGGGG GATGTGCAATTTGAGGGGGATTCATTTTCCAGATAATTGGGGGGTACTGG AGGGGTTTCCTGAGGTGTACGAACAACTGGCACCACAGCCAAACGGGG GAGACGAGTTAAATCATAGTCTCCCAGGGGACAGGGAGGGGGATGTAC TTGAGAAGGATAGCAGCGAAAAGGAGAAGGCTGCACCAGAGGCA TTGCCCTCAGTGCAAAGGGCCCGCAGTGAACAGTTGCCAGATAACATCG TAAAGGTGACTGTTCCCGACAAAAAATCCACCATGTCCCTGCTGTGGTGT CCGCTTAAACTCAGTGTTAGCTCTGATTGAACATCTGAAGGGCTCACAC GGGAGGAGGAGGGTGTGCTTTAGGTGTGCCAAATGTGGGAGGGAGAA TTTTAACCACCATAGTACTGTTTGTCATTACGCAAAGTGCAAAGGTCCAC AGATTGAAAGGCCACCAGTGGGAGAGTGGATCTGTGAGGTATGCGGA AGGGACTTCACGACCAAAATTGGCCTGGGACAACACAAAAGACATATG CATGCAATGGTGAGAAACCAGGAAAGGATCGATGCTTCCCAACCGAAA GAGACATCAAATCGAGGAGCCCACACAAGAGGTGCTGGACGAAGGAGGA GGAAGAACTGCTCATGAAGTTGGAGGTACAGTTTGAGAATCACAAAAA CATCAATAAGCTTATCGCAGAGCAATTAACAACTAAAACAGCTAAACAA ATTAGTGATAAAAGGAGAATGCTGCTCAAAAAAGGTAGGGGGACAACT GGTAATTTGGAAACAGAGCCTGGGATGAGTCATCAATCGCAGGCAAAA GTTAAGGACAATGGACTGGGTGGGGACCATCTGCCGGGAGGACCAGTT GTCGATAAGGGAACAATAGGGAAGCCAGGACAACATCTTGACACAGAT AACAGCCATCAAATAACTGCTGGCAAGAAGAAAGGGGGAGGGCTGCA GGCTCGTTATAGAAGGAGAATAATGAAACGATTAGCGGCCGGGACAAT TAACATCTTCCCCAAAGTGTTTAAAGAACTGATTAACGACCAAGAGGCG	CGACTTGA GAAGGTCT GGTTACAA CTGGGCAT AGCTGCAG AGATCGCG CCTCCTCG TGGCCCCG CTGGTAAG CCCTTAAC AGGGTGA CTAAGTCG ATCTCTGC CCCAGTCC GCCAGCGAT TCCTTCCA AATTCGGT GA (SEQ ID NO: 1142)	GTAGTC ACATTG CACTTTC TGTAACT TGCACT GGGTGT GGGATG TGGGCC TGGGGT GTGGGT TATGGG GTATAT ATGTGG GATATTC TGGTGG GAATGT CCATTCA CTGTAT GCCTATC TTTTTAA TAAAAA GACGGT AGCTAG GTTCGC GAAGCA GCCACA AGCCAA TAGCCA GTTAGG TAGCTC ATAGTG GGTAGG	NKFLGKSRVAYCLKP GPPVSDRGKEFGSGL TTHPEPESESGHDPT VPNPGPSLGAGEGA QPLPLLRVSVGTQTC EEDFITSRPTKLPGIES ELGPLVKFSLEVYRSD LKGDVQFEGIHFPDN WGVLEGFPEVYEQL APQPNGGDELNHSL PGDREGDVLEKDSSE KEKEAAPEALPSVQR ARSEQLPDNIVKVTV PDKNPPCPCCGVRLN SVLALIEHLKGSHGRR RVCFRCAKCGRENFN HHSTVCHYAKCKGP QIERPPVGEWICEVC GRDFTTKIGLGQHKR HMHAMVRNQERID ASQPKETSNRGAHKR CWTKEEEELLMKLEV QFENHKNINKLIAEQ LTTKTAKQISDKRRM LLKKGRGTTGNLETE PGMSHQSQAKVKD NGLGGDHLPGGPVV DKGTIGKPGQHLDTD NSHQITAGKKKGGGL QARYRRRIMKRLAAG TINIFPKVFKELINDQE ARPLINQTTEDCFGLL

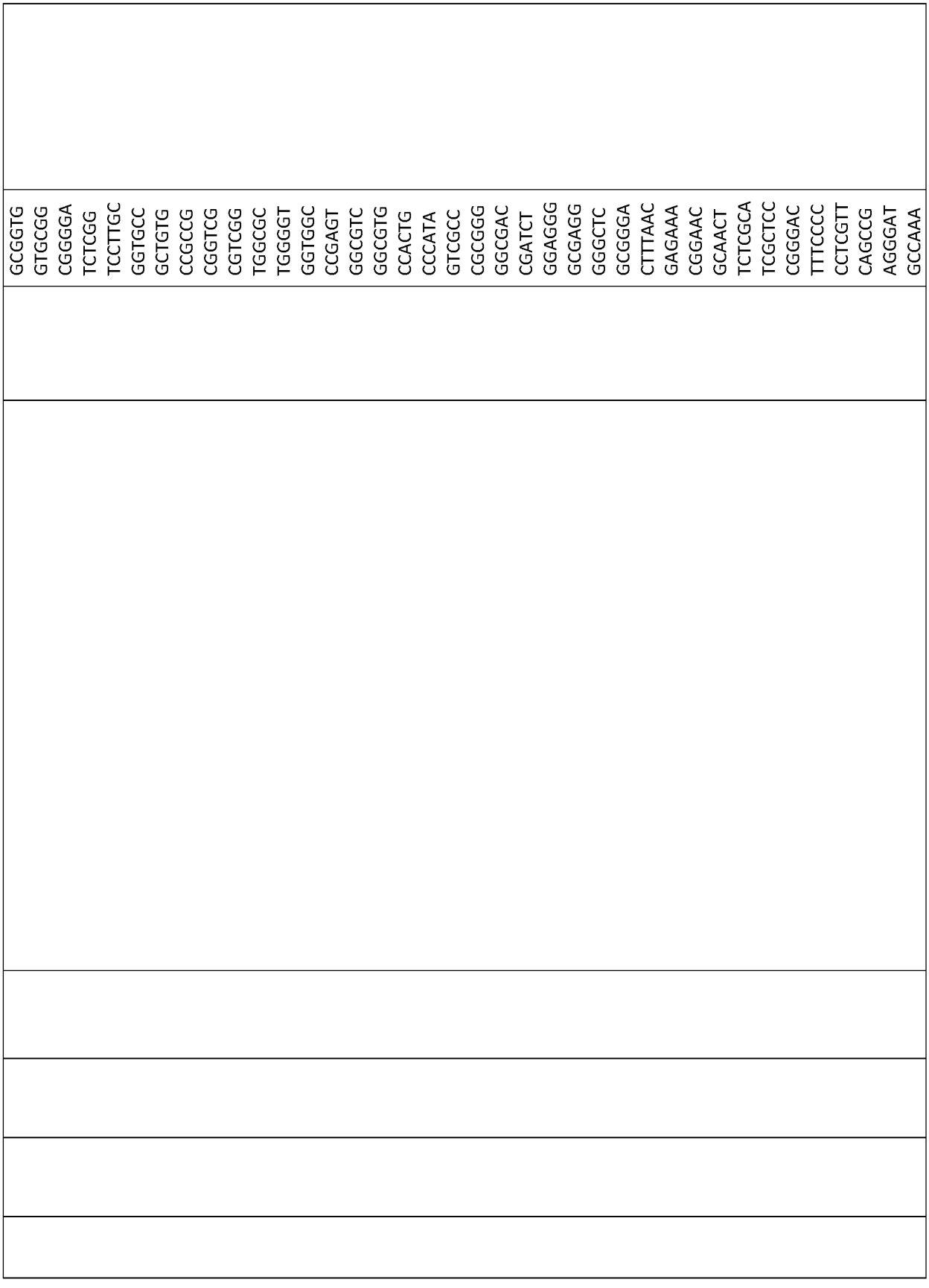
				TGACAG	DSACQIRTALREKGK
				GAACCT	SQEERPRKQYQKW
				TTGACTC	MKKRAIKRGDYLRFQ
				AGAACG	RLFHLDRGKLARIILD
				CGTCCAT	NTESLSCDISPSEIYSV
				TAACATC	FKARWETPGHFNGL
				TAGAAC	GDFEIKGKANNKAFR
				GGACCA	DFITAKEIEKNVREMS
				AACTTC	KGSAPGPDGIALGDI
				GGACAT	KKMDPGYSRTAELFN
				GCACCG	LWLTAGDIPDMVRG
				ATTAACC	CRTVLIPKSTTPERLK
				GGATTT	DINNWRPITIGSILLRL
				GTCCAA	FSRIITARMTKACPLN
				GGTGGA	PRQRGFISAPGCSEN
				CGGGCC	LKLLQSIIRTAKNEHK
				ACCTTTA	PLGVIFVDIAKAFDTV
				CTTAACC	SHQHIIHVLQQRRVD
				CGGAAA	PHIVGLVNNMYKDIS
				GGGAAC	TYVTTKKNTHTDKIQI
				ATATATA	RVGVKQGDPLSPLLF
				GTTATAT	NLAMDPLLCKLEESG
				GTGTTC	KGFHRGQSSITAMAF
				GTAATA	ADDLVLLSDSWENM
				(SEQ ID	KENIKILETFCNLTGLK
				NO:	TQGQKCHGFYIKPTK
				1265)	DSYTINNCPAWTING
					TPLNMINPGESEKYL
					GLQIDPWTGVAKYD
					LSTKLKIWLESIDRAPL
					KPLQKLDILKTYTIPRL
					TYLADHSEMKAGALE
					ALDQQIRTAVKDWL
					HLPSCTCDAILYVSTR
					DGGLGVTKLAGLIPS
					VQARRLHRIAQSPDE
					TMKDFLEKAQMEK
					MYEKLWVQAGGKK

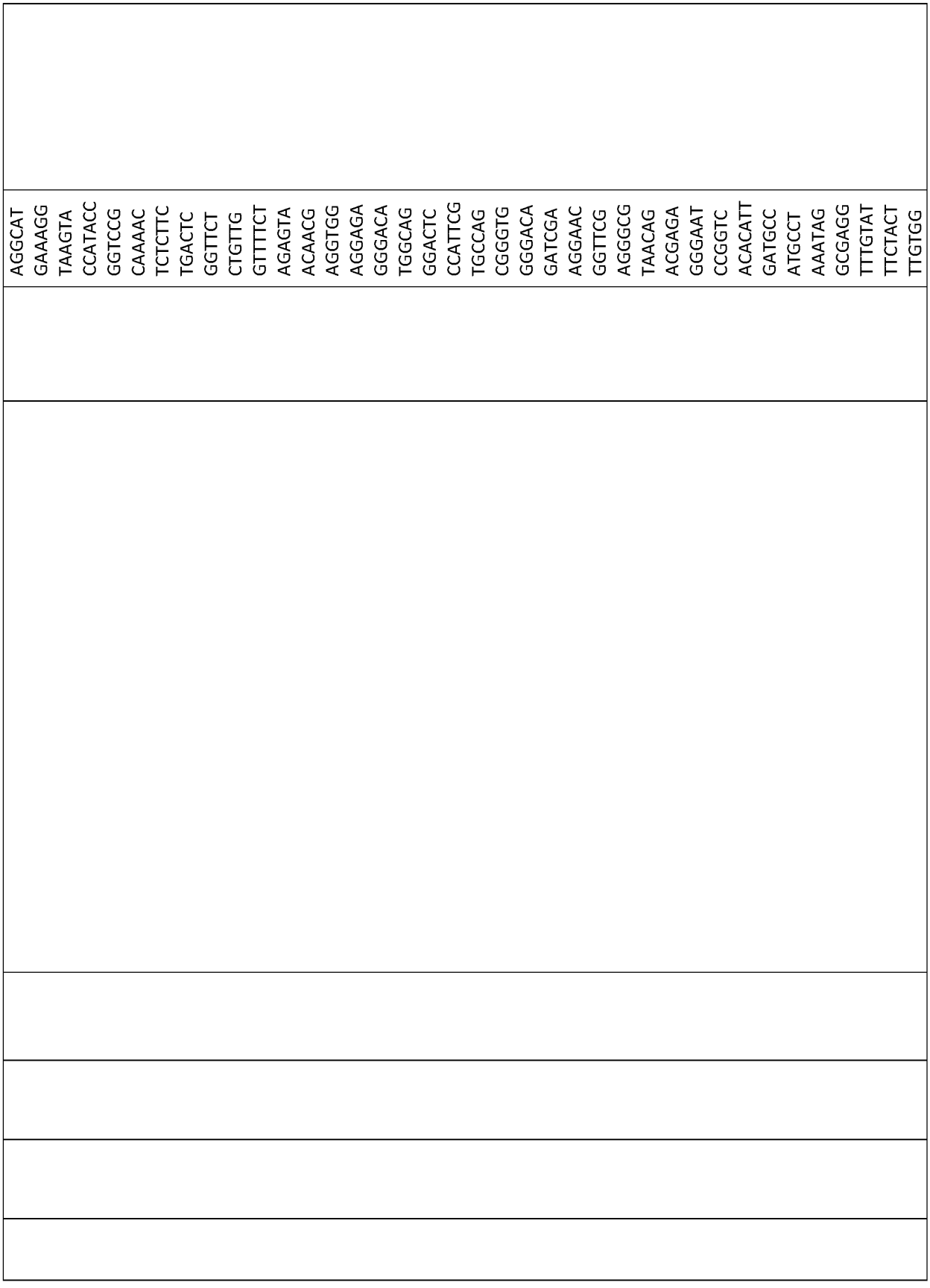
			TGGGAAGCACCGAACCCCAAAGTAAGTACCCAAAACCTTGTGATTGGA				KGMPSIWEALPMTV
			GAAGGAAAGAGCTTAAAAAGTGGACAAAATTGGAGTCCCAAGGTCGTG				PPTNTGNLSEWEAP
			GAGTCAAAAATTTTAGGAATGATACAATTAGTAACGATTGGATCCAATA				NPKSKYPKPCDWRR
			TTATAGACGCATACCTCACAGGAAACTCCTCACTGCCATACAACTCAGG				KELKKWTKLESQGRG
			GCCAATGTATACCCCACAAGGGAATTTCTCGCGCGGGGGAGGGGTGAT				VKNFRNDTISNDWIQ
			AACTATGTTAAGTTTTGTAGGCACTGTGAAGCGGACCTTGAAACCTGTG				YYRRIPHRKLLTAIQL
			GCCATATCATCGGCTTTTGCCCAGTAACGAAGGACGCCCGAATCAAGAG				RANVYPTREFLARGR
			GCACAATCGCATATGCGACAGGCTTTGCGAGGAGGCAGCTAAGAGGGA				GDNYVKFCRHCEADL
			ATGGGTGGTCTTCAAGGAGCCGCACTTGAGGGATGCCACCACGGAACT				ETCGHIIGFCPVTKDA
			GTTTAAACCGGATGTGATATTCGTGAAAGAGGACCGTGCACTGGTTGTG				RIKRHNRICDRLCEEA
			GATGTGACAGTACGATATGAATCAGCCAAGACAACGCTGGAGGCAGCT				AKREWVVFKEPHLR
			GCTATGGAGAAAGTGGACAAGTACAAACATCTGGAGGCAGAAGTGAA				DATTELFKPDVIFVKE
			GGAACTCACCAACGCAAAGGACGTTGTTTTATGGGGTTCCCCCTTGGA				DRALVVDVTVRYESA
			GCGCGAGGGAAATTCTACAAAGGGAACTTTAACTTGCTAGAGACTCTTG				KTTLEAAAMEKVDKY
			GCCTCCCAAAAACGAGGCAATTGAGTGTGGCAAAGACTCTATCCACGTA				KHLEAEVKELTNAKD
			CGCGCTCATGTCATCTGTGGACATTGTGCATATGTTTGCCAGTAGATCTA				VVFMGFPLGARGKF
			GGAAACCAAATGTCTAGGTAGTCACATTGCACTTTCTGTAACTTGCACTG				YKGNFNLLETLGLPKT
			GGTGTGGGATGTGGGCCTGGGGTGTGGGTTATGGGGTATATATGTGG				RQLSVAKTLSTYALM
			GATATTCTGGTGGGAATGTCCATTCACTGTATGCCTATCTTTTTAATAAA				SSVDIVHMFASRSRK
			AAGACGGTAGCTAGGTTCGCGAAGCAGCCACAAGCCAATAGCCAGTTA				PNV (SEQ ID NO:
			GGTAGCTCATAGTGGGTAGGTGACAGGAACCTTTGACTCAGAACGCGT				1387)
			CCATTAACATCTAGAACGGACCAAACTTCGGACATGCACCGATTAACCG				
			GATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGAAAGGGAAC				
			ATATATAGTTATATGTGTTCGTAATA (SEQ ID NO: 1541)				
R2	R2Dr	Danio rerio	AATCCCCCCTACCCAATCCCCCCGTCGTGACCTCCAGGCCAGGAATCACG	AATCCCCC	AAATCC		MESTAKGKSYWMA
			AGCGTACGACAGTGGCCATCCGGCAATGACAATAGCGTGACTAACGAC	CTACCCAA	CAGCGG		RRPVEGATEGSLGRV
			AATGAGTCAGATCCATGACCCTTGGAGTGGGTTAACCTCCGCCTCTTTAA	TCCCCCCG	GATACA		PFVTRDPKRKPEAKR
			AAACATGGAAAGTACAGCAAAAGGAAAGTCATACTGGATGGCCCGTCG	TCGTGACC	GCAAGA		TLTHGLGLRECSVVLT
			CCCAGTAGAAGGTGCCACGGAGGGATCTTTGGGTCGGGTCCCTTTCGTA	TCCAGGCC	AGGTAT		RLIEGRRGRDHTPSG
			ACGCGAGATCCTAAGCGCAAACCAGAGGCTAAACGAACACTTACGCAT	AGGAATCA	CGGATC		WNAQRGMPNDESS
			GGCTTAGGACTACGAGAATGCTCGGTTGTCTTGACACGCCTCATCGAGG	CGAGCGTA	TAATAA		VEEPNGPIPSNPIPTG
			GGCGTCGAGGTCGCGATCACACACCATCAGGATGGAACGCACAGCGCG	CGACAGTG	GGTTGA		TQALPEPMADGEQG
			GCATGCCAAACGACGAAAGCTCGGTCGAGGAGCCCAATGGGCCGATAC	GCCATCCG	GCGAGG		EHPGVVTLPLRDLN
			CATCTAACCCCATACCAACGGGCACCCAAGCCCTGCCTGAACCTATGGC	GCAATGAC	AGAGGG		CPLCGGSASTAVKVQ
			GGACGGGGAGCAGGGGGAGCACCCGGGAGTGGTGGTGACCCTGCCGC	AATAGCGT	TGGAGA		RHLAFRHGTVPVRFS
			TCAGGGACTTAAACTGCCCCCTATGTGGCGGGTCGGCGAGCACCGCGG	GACTAACG	TCCTTTG		CESCGKTSPGCHSVL
			TGAAAGTGCAAAGACACTTGGCATTTCGCCACGGAACAGTGCCGGTTA	ACAATGAG	GGGGGG		CHIPKCRGPTGEPPE
			GATTCAGCTGTGAATCATGTGGAAAAAACTTCTCCGGGTTGCCATTCCGTC	TCAGATCC	GTCGGG		KVVKCEGCSRTFGTR

	CTCTGTCACATTCCGAAATGTCGCGGACCGACAGGCGAGCCGCCTGAGA	ATGACCCT	CTAAGTT	RACSIHEMHVHSEIR
	AAGTGGTTAAGTGCGAGGGATGCAGTAGGACGTTTGGCACAAGGAGA	TGGAGTG	CCCCTCT	NRKRIAQDRQEKGTS
	GCGTGTAGTATACATGAGATGCACGTTCACTCAGAAATCCGCAATAGGA	GGTTAACC	CGGGTC	TDGEGRAGVERADA
	AAAGAATTGCTCAAGACAGGCAAGAAAAAGGGACCTCGACAGATGGA	TCCGCCTC	CTCCCAC	GEGPSGEGIPPKRPR
	GAGGGGAGAGCTGGAGTCGAAAGGGCTGACGCTGGGGAAAGGTCCCTC	TTTAAAAA	GGTGAC	RARTPREPSEPPANP
	TGGGGAAGGGATCCCCCCTAAACGTCCCAGACGTGCGAGAACGCCCAG	C (SEQ ID	GCTCTAC	PILSPQPDLPPGGLRD
	AGAACCGTCTGAGCCCCCCCGCGAATCCGCCGATTCTCTCGCCACAACCC	NO: 1143)	CCCTCCC	LLREVASGWVRAAR
	GATCTGCCCCCAGGAGGCCTCCGGGACCTACTCCGGGAGGTGGCCAGT		TCCTCGC	DGGTVIDSVLAAWL
	GGGTGGGTAAGGGCAGCGAGAGACGGAGGTACGGTGATTGACAGCGT		TCGTAG	DGNDRLPELVDAAT
	GCTCGCAGCATGGTTGGATGGCAACGATCGGCTCCCTGAGCTGGTTGAC		AACCCA	QRTLQGLPAGRLARR
	GCGGCGACGCAAAGGACACTGCAGGGCTTACCTGCAGGGAGGTTGGCC		ACGGTG	PATFVAPNRRRGRW
	CGAAGACCCGCAACTTTTGTTGCGCCTAACCGGAGGAGAGGCAGGTGG		AACACG	GRRLKLLAKRRAYHD
	GGGCGCCGGCTCAAACTGCTCGCTAAGCGCCGCGCCTACCACGATTGCC		GTTGGC	CQIRFRKDPARLAANI
	AAATTCGGTTCCGAAAAGACCCAGCCCGCCTAGCCGCGCGAACATCCTAGA		AGGATG	LDGKSETSCPINEQAI
	CGGCAAAAGCGAAACAAGTTGCCCAATCAATGAGCAAGCGATTCATGA		AAGTGA	HEHFRNKWANPSPF
	GCACTTTCGAAACAAATGGGCAAATCCAAGTCCATTTGGTGGGCTGGGA		CGTGAG	GGLGRFGTENRANN
	CGATTTGGGACGGAAAACAGGGCCAACAACGCCCACCTCCTCGGGCCA		GGGTAA	AHLLGPISKSEVQTSL
	ATCTCCAAAAGCGAGGTCCAAACTAGCCTCCGAAATGCATCGAACGCCT		GACATG	RNASNASTPGPDGV
	CCACACCAGGCCCAGACGGCGTTGGGAAAAGGGACATTTCCAACTGGG		CGTACG	GKRDISNWDPECETL
	ATCCTGAGTGTGAGACCCTCACTCAGCTGTTTAACATGTGGTGGTTCACA		TGAGCG	TQLFNMWWFTGVIP
	GGTGTCATCCCCTCTCGCTTGAAGAAAAGTCGTACGGTGCTTCTGCCCA		CGCATTT	SRLKKSRTVLLPKSSD
	AGTCCTCAGACCCAGGAGCGGAGATGGAGATCGGCAACTGGAGACCAA		TTGCTGT	PGAEMEIGNWRPITI
	TCACCATCGGGTCGATGGTCTTGCGGCTTTTCACAAGGGTGATCAATAC		TCTCTG	GSMVLRLFTRVINTR
	GAGATTAACGGAAGCCTGTCCGTTGCACCCAAGACAGAGAGGGTTTCG		GACTGG	LTEACPLHPRQRGFR
	ACGAAGCCCCGGGTGTTCGGAGAACCTGGAAGTACTCGAATGTCTCCTC		GTTTCGT	RSPGCSENLEVLECLL
	CGACACTCCAAAGAAAAGCGCAGCCAACTGGCAGTGGTATTCGTCGATT		CCCCCTC	RHSKEKRSQLAVVFV
	TTGCACAAGCGTTTGACACCGTCTCTCATGAACACATGCTGTCAGTCCTT		ACAACC	DFAQAFDTVSHEHM
	GAGCAGATGAACGTGGATCCCCACATGGTAAATCTGATCCGGGAGATTT		ATCACTT	LSVLEQMNVDPHMV
	ACACAAACAGCTGCACAAGTGTCGAGCTAGGCCGGAAAGAGGGACCAG		ACACTAT	NLIREIYTNSCTSVELG
	ACATCCCAGTGAGGGTTGGTGTTAAGCAAGGGGATCCTCTGTCCCCGCT		AGGGGC	RKEGPDIPVRVGVKQ
	GCTTTTCAACCTGGCTTTGGATCCTCTCATCCAAAGTCTCGAACGCACAG		ACAGCG	GDPLSPLLFNLALDPL
	GCAAAGGGTGTGAGGCCGAAGGTCACAAAGTGACAGCTTTAGCGTTCG		GCTCCTA	IQSLERTGKGCEAEG
	CGGATGACCTGGCACTGGTTGCGGGCTCGTGGGAGGGAATGGCACACA		CCTCCCT	HKVTALAFADDLALV
	ACCTTGCGCTTGTAGACGAATTCTGCCTAACCACCGGCCTCACAGTCCAA		CCCTATG	AGSWEGMAHNLAL
	CCCAAAAAGTGCCACACAGTTTCATGGTCAGGCCCTGCAGAGGTGCCTTCA		ACCCCCC	VDEFCLTTGLTVQPK
	CAGTGAACGACTGCCCCCCCATGGGTTCTGGGGGGCAAGGCCCTGCAGC		CTTCCCA	KCHSFMVRPCRGAF
	TAACAAACATCGAAAACTCCATCAAATATCTGGGAGTAAAAGTCAATCC		TACCGA	TVNDCPPWVLGGKA
	TTGGGCGGGGATTGAAAAGCCTGACCTTACAGTGGCACTAGACCGATG		TCCATG	LQLTNIENSIKYLGVK

		GTGCAAGCGCATTGGGAAGTCACTGCTCAAACCCTCACAGAAGGTATAC		GCTGTT	VNPWAGIEKPDLTVA
		ATTCTCAATCAGTTTGCCATCCCGCGACTCTTCTACCTGGCTGATCACGG		CTAGTCT	LDRWCKRIGKSLLKPS
		TGGGGCCGGCGACGTCATGCTCCAGAACCTGGATGGGACAATCAGGAA		GGACCG	QKVYILNQFAIPRLFY
		GGCGGTGAAGAAATGGCTGCATCTTCCACCGTCAACCTGCAACGGGCT		AGGGTC	LADHGGAGDVMLQ
		GTTGTATGCCAGGAACTGTAATGGTGGCCTCGGTATATGCAAGCTCACT		GGACGG	NLDGTIRKAVKKWLH
		CGGCACATCCCATCAATGCAGGCGAGACGAATGTTCCGCTTGGCCAACT		GGCATT	LPPSTCNGLLYARNC
		CATCGGACCCGTTGATGAAGGCCATGATGCGCGGCTCCCGAGTCGAAC		TGAAGG	NGGLGICKLTRHIPS
		AGAAATTCAAAAAGGCCTGGATGCGGGCCGGGGGAGAGGAGAGTGCG		TAGCTG	MQARRMFRLANSSD
		CTCCCACGGGTGTTCGGGGCGAATCAGTACCAGGAAGGGGAGGAGGT		GAATCC	PLMKAMMRGSRVE
		CGCTAACGATCTGGTACCTCGCTGCCCAATGCCGAGCGATTGGAGACTG		TCCGCT	QKFKKAWMRAGGE
		GAAGAATTCCAACACTGGATGGGCCTGCCGATCCAGGGTGTGGGTATA		GCTGCG	ESALPRVFGANQYQE
		GCCGGCTTCTTCAGAAACAGGGTGGCTAACGGATGGCTCAGGAAGCCG		AGCCTG	GEEVANDLVPRCPM
		GCAGGGTTCAAAGAGCGGCACTACATCGCCGCTCTACAACTGCGAGCAT		AGGTCG	PSDWRLEEFQHWM
		GTGTATACCCCACCCTCGAATTCCAGCAAAGGGGCAGGAGCAAAGCGG		ATGGTT	GLPIQGVGIAGFFRN
		GTGCGGCCTGCAGGCGGTGCTCATCCCGGTTGGAATCCAGCTCTCACAT		AGAGGT	RVANGWLRKPAGFK
		CCTCGGCAAATGTCCGGCGGTGCAGGGAGCCAGAATCAGGCGTCATAA		GAAATA	ERHYIAALQLRACVYP
		CAAAATATGCGACCTCCTGAAGGCCGAAGCCGAAACCCGGGGTTGGGA		CTTGGG	TLEFQQRGRSKAGAA
		GGTACGCCGGGAATGGGCCTTCAGAACTCCGGCTGGGGAACTGAGAAG		AGGAGA	CRRCSSRLESSSHILG
		GCTCGACCTGGTACTCATCCTCGGGGATGAGGCATTGGTCATTGACGTC		CACAGC	KCPAVQGARIRRHNK
		ACAGTAAGGTACGAGTTCGCTCCGGATACCCTCCAGAAATGCCGGAAAG		CTCCGG	ICDLLKAEAETRGWE
		GACAAGGTCAGCTACTACGGCCCCGCACAAAGAAGCGATCGCTCGGGAG		AGAGCC	VRREWAFRTPAGELR
		CTGGGCGTAAGAAGGGTCGACATACATCATGGGTTTCCGTTGGGTGCACGC		CCTCCCG	RLDLVLILGDEALVID
		GGACTTTGGCTCGCCAGCAACTCCAAAGTGCTGGAACTGATGGGATTGA		GGTGGT	VTVRYEFAPDTLQNA
		GCAGGGAAAGAGTGAAGGTCTTCTCCAGACTCTTGAGTCGGAGAGTGC		CATCAT	GKDKVSYYGPHKEAI
		TCCTGTACTCTATCGACATCATGAGGACATTTTACGCAACCCTGCAATGA		GGCAAC	ARELGVRRVDIHGFP
		AAATCCCAGCGGGATACAGCAAGAAGGTATCGGATCTAATAAGGTTGA		CGGGTG	LGARGLWLASNSKVL
		GCGAGGAGAGGGTGGAGATCCTTTGGGGGGGGTCGGGCTAAGTTCCC		AAACCTT	ELMGLSRERVKVFSR
		CTCTCGGGTCCTCCCACGGTGACGCTCTACCCCTCCCTCCTCGCTCGTAG		ACGGTT	LLSRRVLLYSIDIMRTF
		AACCCAACGGTGAACACGGTTGGCAGGATGAAGTGACGTGAGGGGTA		TCACTTA	YATLQ (SEQ ID NO: 1388)
		AGACATGCGTACGTGAGCGCGCATTTTTGCTGTTCTCTGGACTGGGTTTC		CGAAAC	
		GTCCCCCTCACAACCATCACTTACACTATAGGGGCACAGCGGCTCCTACC		AGCACC	
		TCCCTCCCTATGACCCCCCCTTCCCATACCGATCCATGGCTGTTCTAGTCT		ATAACA	
		GGACCGAGGGTCGGACGGGGCATTTGAAGGTAGCTGGAATCCTCCGCT		GCGCCG	
		GCTGCGAGCCTGAGGTCGATGGTTAGAGGTGAAATACTTGGGAGGAGA		TAATAG	
		CACAGCCTCCGGAGAGCCCCTCCCGGGTGGTCATCATGGCAACCGGGT		CGCACC	
		GAAACCTTACGGTTTCACTTACGAAACAGCACCATAACAGCGCCGTAAT		GGTGTG	
		AGCGCACCGGTGTGACTACTGTCCAGTGCTGATATTCTCATCTGGAGAA		ACTACT	
		TACAACACGGGTAATGGCAGAGTATTCAAAACCCAAATGTTTACGATCG		GTCCAG	

			ACCAACGGAGTCGTTCCCTTGCATCTAGGCCGGACCCGAAACTGCCGTA ATTGCCCGTCCCCAAGGTAGCCTCTTAGAAAACCGAAGCCCGGTCGGGG CGGTGGTTGCGGCGGCGCTGCGGGGGCCTGCTGCTCGGGCGGCGTCG GTGTGCCGCGGTGGTTGCGGTGGTGCGGCGGGGATCTCGGTCCTTGCG GTGCCGCTGTGCCGCCGCGGTCGCGTCGGTGGCGCTGGGGTGGTGGCC CGAGTGGCGTCGGCGTGCCACTGCCCATAGTCGCCCGCGGGGGCGACC GATCTGGAGGGGCGAGGGGGCTCGCGGGACTTTAACGAGAAACGGAA CGCAACTTCTCGCATCGCTCCCGGGACTTTCCCCCCTCGTTCAGCCGAGG GATGCCAAAAGGCATGAAAGGTAAGTACCATACCGGTCCGCAAAACTCT CTTCTGACTCGGTTCTCTGTTGGTTTTCTAGAGTAACAACGAGGTGGAG GAGAGGGACATGGCAGGGACTCCCATTCGTGCCAGCGGGTGGGGACA GATCGAAGGAACGGTTCGAGGGCGTAACAGACGAGAGGGAATCCGGT CACACATTGATGCCATGCCTAAATAGGCGAGGTTTGTATTTCTACTTTGT GGGTTCAGTATAGTCGGAGCATATGGTCGGTTGTCCCGTTGTTTTCACG GCGGGCAAGCGACTATCATGATAAAGTAGAATGGGAGACGGGCTCCCT GACAAACCCGGAAAGGCGCCCCCCCGTGGTTCGTAGCAGCTGACGGAT CACGCTCGAAGAAAAATGAGTGAGAGGGGACGCCGCAACCAC (SEQ ID NO: 1542)		TGCTGA TATTCTC ATCTGG AGAATA CAACAC GGGTAA TGGCAG AGTATTC AAAACC CAAATG TTTACGA TCGACC AACGGA GTCGTT CCCTTGC ATCTAG GCCGGA CCCGAA ACTGCC GTAATT GCCCGT CCCCAA GGTAGC CTCTTAG AAAACC GAAGCC CGGTCG GGGCGG TGGTTG CGGCGG CGCTGC GGGGGC CTGCTG CTCGGG CGGCGT CGGTGT GCCGCG GTGGTT


GCGGTG
GTGCGG
CGGGGA
TCTCGG
TCCTTGC
GGTGCC
GCTGTG
CCGCCG
CGGTCG
CGTCGG
TGGCGC
TGGGGT
GGTGGC
CCGAGT
GGCGTC
GGCGTG
CCACTG
CCCATA
GTCGCC
CGCGGG
GGCGAC
CGATCT
GGAGGG
GCGAGG
GGGCTC
GCGGGA
CTTTAAC
GAGAAA
CGGAAC
GCAACT
TCTCGCA
TCGCTCC
CGGGAC
TTTCCCC
CCTCGTT
CAGCCG
AGGGAT
GCCAAA


AGGCAT
GAAAGG
TAAGTA
CCATACC
GGTCCG
CAAAAC
TCTCTTC
TGACTC
GGTTCT
CTGTTG
GTTTTCT
AGAGTA
ACAACG
AGGTGG
AGGAGA
GGGACA
TGGCAG
GGACTC
CCATTCG
TGCCAG
CGGGTG
GGGACA
GATCGA
AGGAAC
GGTTCG
AGGGCG
TAACAG
ACGAGA
GGGAAT
CCGGTC
ACACATT
GATGCC
ATGCCT
AAATAG
GCGAGG
TTTGTAT
TTCTACT
TTGTGG

						GTTCAG TATAGTC GGAGCA TATGGT CGGTTG TCCCGTT GTTTTCA CGGCGG GCAAGC GACTAT CATGAT AAAGTA GAATGG GAGACG GGCTCC CTGACA AACCCG GAAAGG CGCCCC CCCGTG GTTCGT AGCAGC TGACGG ATCACG CTCGAA GAAAAA TGAGTG AGAGGG GACGCC GCAACC AC (SEQ ID NO: 1266)	
R2	R2-1_GA	.	Gasterosteus aculeatus	CATATTGGGGTCTCAGGAGGAGACACAGGGTCTGTTGCGGCTCCGGTA AACGGTACCGGAGTCGGTTAAGCATCGTTTGGGCCCGCCTCCACGTGGT GGTCCGCGGTAACACCAATAGGGTGGCTAAGAGGCCCAGTAATTTCCCC GAATTGTCTTCCCCCCCGCGCGGGGGGGACCCCCCTTTAGTGTCGGAGC GGTCGCGCCTCCGCGTTTGGGGTGTCGCAGGCGTGAGCCTTCGTCCCCT	CATATTGG GGTCTCAG GAGGAGA CACAGGGT CTGTTGCG	GGAGGG GAGTAG GTCTCTA CTCTGAC CCGAAG	MLRGGVGTPPAGGA GAVGPGMASPGGCS VRFSPGGRRLLGHRT GGLSPSVSWRLKRLS VSLRRWSGPGLLGA

		TAAGTTCAGACGGTCCCGGCTTCTTGCCGGGCCAACCCCCGGTGCAGCG	GCTCCGGT	GGCCCC	DGAGGGAAVASPRG
		TTCTCCCATGTTGGATCGGCACCCAGCCCCGGGTGCCATGCGAGTTCAG	AAACGGTA	CCCGTTT	TQVLGSGAGRRWLG
		ACATTTTGTTTATGTATCGTCTGCGTGGTTGACTTGCTAAGCTCATTTCCT	CCGGAGTC	CAGACC	HGSRGSSPSAARGLR
		CCTCTCACTGCGTCCCCCCAGGTGCTGATCGGTTGAAGAGGATTCGTCG	GGTTAAGC	TGATTCT	RLTVRLKRLSGGLLSP
		TTGACCTCGGCGGTGAATTTGGGATTGTATTATACAGGTAGGTATAGAG	ATCGTTTG	AGGCTA	KACRDAEEGSSSSPG
		GGCGTGCGGATGTTGCGTGGCGGTGTTGGTACTCCCCCGGCTGGGGGA	GGCCCGCC	CCTGTG	FRNPKGLGGRGLTPL
		GCGGGTGCGGTGGGGCCAGGCATGGCCTCGCCGGGTGGTTGCAGTGTC	TCCACGTG	CCTAATT	GSRRFCRLTVSLNRW
		CGGTTCAGTCCCGGAGGGAGGCGACTGCTTGGCCACAGGACTGGAGG	GTGGTCCG	GGGGGG	RGSLVKLNASSRASG
		GTTGAGTCCCTCCGTGTCCTGGAGGCTCAAGCGACTGTCTGTCTCTCTGA	CGGTAACA	GTCCCA	RRTPVKPACDSRAGR
		GGCGCTGGAGCGGGCCTGGGCTGCTAGGTGCGGATGGTGCGGGGGGGA	CCAATAGG	AAGAGA	GSEHAEGGGVSAAP
		GGCGCTGCGGTGGCCTCCCCCAGGGGTACGCAGGTCCTGGGAAGTGGG	GTGGCTAA	TGTTGTC	MVLRSRRKLTFSVDG
		GCCGGGCGTCGGTGGCTTGGGCACGGGTCGCGAGGGTCTTCTCCTTCT	GAGGCCCA	TGTTGTA	DSNSGDRARSGSVSA
		GCGGCCCGGGGGGCTAAGGCGCGCTGACGGTACGGTTGAAGCGACTCAG	GTAATTTC	GAAGGG	ARPGHLLVDGESASS
		CGGTGGCCTGTTGTCCCCTAAGGCGTGTCGGGATGCGGAAGAAGGAAG	CCCGAATT	TTTGCG	RSGPAGDARLAGPST
		CTCCAGCAGCCCAGGGTTCCGGAATCCAAAGGTCTCGGGGGAAGGGG	GTCTTCCC	CCACTG	RSRRKGCLPPVDFEN
		GTTGACGCCTCTCGGATCCCGTAGATTTTGTCGGCTGACCGTCTCCCTGA	CCCCGCGC	ACTGCA	PKKRTRLMAKMTNG
		ATCGCTGGAGGGGGCAGTCTGGTGAAGTTGAACGCTAGTAGCAGGGCCT	GGGGGGG	CGGAAG	NPTSHVPCPAPCSNG
		CCGGCCGGAGGACCCCTGTGAAACCCGCTTGTGACTCTAGAGCCGGAC	ACCCCCCT	GGTGGG	HEGGGRVAVIEGRLP
		GGGGCTCGGAGCATGCGGAGGGAGGTGGAGTGAGCGCTGCACCTATG	TTAGTGTC	CCTCGA	ELSGSRISGIQPALPV
		GTGTTGCGCAGTCGGCGTAAGCTCACCTTCTCTGTGGATGGCGACTCTA	GGAGCGG	CAGGTA	ETSFVGQSTGRGAD
		ACTCCGGGGATAGGGCCCGGAGCGGGTCCGTCTCTGCAGCCCGTCCTG	TCGCGCCT	GGGGTT	GDANANSSPPSPNL
		GCCACTTGTTGGTGGATGGTGAGAGTGCGTCCTCAAGATCTGGCCCCGC	CCGCGTTT	ACATGA	GGSVGMVPAVRDGT
		GGGGGATGCCAGGTTGGCGGGGCCTTCTACGCGGAGTAGGAGGAAGG	GGGGTGT	CTCCGT	PPLGRPGEDHSRECA
		GTTGCCTTCCCCCGGTCGACTTTGAAAACCCGAAGAAGCGCACACGGTT	CGCAGGC	GCTGCT	GGNTPLWMLEDSFR
		GATGGCTAAGATGACGAATGGTAATCCTACCTCGCACGTCCCTTGCCCT	GTGAGCCT	CAGCAG	CDYCPREFGTRAGRS
		GCCCCGTGCTCAAATGGGCATGAAGGAGGTGGGCGAGTTGCGGTGATC	TCGTCCCC	ACCCGC	LHMRRAHLAEYDGA
		GAGGGGCGGCTGCCGGAGTTAAGCGGTAGTAGGATCTCTGGAATACAG	TTAAGTTC	GCCTCT	GFCWGERLSEFAATR
		CCAGCCCTGCCTGTTGAAACCAGCTTTGTCGGCCAATCGACTGGCCGGG	AGACGGTC	GAGACC	LWSTEETKKLAVFCE
		GCGCGGACGGCGATGCGAATGCGAATAGTAGCCCGCCTTCTCCTAATCT	CCGGCTTC	GGGTAG	RGVPSPSECRAIAASL
		GGGCGGCTCGGTTGGGATGGTGCCTGCCGTGCGTGATGGTACCCCGCC	TTGCCGGG	GGCTAC	GAGKTHHQVRSKCR
		GCTTGGGCGTCCAGGAGAGGATCACTCGCGGGAGTGTGCAGGGGGAA	CCAACCCC	TTGAAC	LVFEAIRRRELLEVAA
		ATACTCCCCTCTGGATGCTGGAGGACAGTTTCCGGTGTGACTACTGTCCT	CGGTGCAG	AAGCGA	ATERLEKSARRKQPA
		AGGGAATTCGGCACAAGAGCGGGGCGCTCGTTGCACATGCCGCAGGGCT	CGTTCTCC	CGCCCT	VPPAPVHGVRGVLR
		CACCTGGCCGAGTACGACGGGGCAGGTTTCTGTTGGGGTGAACGTCTC	CATGTTGG	GGTGTA	GLLGKRVPREGGTTG
		AGTGAATTCGCCGCTACGCGCCTCTGGTCGACGGAGGAAACCAAAAAG	ATCGGCAC	TGTCCG	STSARIVRRDDCRQG
		CTGGCCGTGTTTTGTGAGAGGGGTGTGCCCTCACCGTCGGAATGCAGA	CCAGCCCC	TATCCTA	AVASASLNLIRRLGRK
		GCCATTGCAGCCTCTCTGGGCGCAGGAAAAACACATCATCAGGTTAGAT	GGGTGCCA	ACCTGG	ATGRSGRRRVLGRPP
		CGAAGTGTCGACTGGTGTTCGAGGCCATTCGGCGGCGTGAATTGCTTGA	TGCGAGTT	TTTGGG	RMDVRRSVRMRRM

	GGTGGCTGCTGCCACGGAGCGTTTGGAGAAAAGCGCTAGGCGGAAGC	CAGACATT	AAAGCC	RRFLYRLARLGWAKL
	AGCCCGCCGTACCACCGGCACCCGTACACGGAGTGAGAGGGGTCCTGC	TTGTTTAT	GATACC	AMFVLDGQMGASC
	GGGGCCTACTAGGGAAGCGGGTGCCGAGAGAGGGTGGTACCACAGGC	GTATCGTC	GGCAAT	PVPLVEVSAVFRERW
	AGCACCTCAGCAAGGATCGTCAGGAGAGACGACTGCCGTCAGGGGGCA	TGCGTGGT	GCCCGC	SIVRAFLGLGQFGGF
	GTTGCGTCGGCTTCTCTCAATCTGATCAGAAGGCTGGGTCGAAAGGCAA	TGACTTGC	CACAGG	GTADNAGFGKLIDPA
	CGGGCCGCTCCGGCAGGAGACGGGTCCTTGGACGCCCACCCAGGATGG	TAAGCTCA	TGTCGC	EVRAHLQSIKNRSSP
	ATGTAAGGCGTAGCGTGAGGATGAGGAGGATGCGCAGGTTCCTCTATC	TTTCCTCCT	GCACCC	GPDGITKVALSKWDP
	GGTTGGCCCGGCTGGGCCAAGTTGGCTATGTTTGTCCTGGACG	CTCACTGC	CACGGG	EGIKLAHMYSTWLVS
	GACAGATGGGGGCGAGCTGCCCCGTTCCACTCGTCGAAGTGTCGGCGG	GTCCCCCC	ATGACG	AGIPKVFKKCRTTLIP
	TCTTCCGGGAGAGGTGGAGCATAGTCAGAGCCTTCCTGGGTCTGGGTC	AGGTGCTG	TATGGG	KTGDVSLHGDVGQ
	AGTTCGGGGGCTTCGGGACTGCCGACAACGCAGGATTTGGGAAGCTGA	ATCGGTTG	CCCCGG	WRPITIASLVLRLYSRI
	TCGATCCGGCTGAAGTCAGGGCCCATCTCCAGTCCATCAAGAACCGGTC	AAGAGGA	GGGACC	LTERMTVACPSHPRQ
	TTCCCCGGGCCCGGATGGCATCACCAAGGTGGCGCTGTCCAAATGGGA	TTCGTCGT	TCATGG	RGFIASPGCSENLML
	CCCCGAAGGGATTAAATTGGCGCACATGTACTCAACATGGTTGGTATCG	TGACCTCG	ATACTCC	LEGCMSLSKAGNGSL
	GCAGGCATCCCTAAGGTCTTCAAGAAGTGCAGGACGACACTTATCCCAA	GCGGTGA	ACTGGA	AVVFVDFAKAFDTVS
	AGACCGGGGACGTTAGTCTACATGGTGACGTGGGGCAATGGAGGCCCA	ATTTGGGA	CTTGCAC	HEHLLSVLVQKGLDQ
	TAACCATTGCGTCCCTGGTCCTGAGACTCTATTCGCGGATCCTGACGGA	TTGTATTA	AATCCT	HMVELIKDSYENSVT
	AAGGATGACAGTGGCCTGTCCTAGCCACCCGCGCCAGAGGGGCTTCATT	TACAGGTA	GGTGTA	KVHCQEGCSTDIAM
	GCCTCCCCGGGCTGTTCGGAAAACCTCATGCTGTTGGAAGGTTGCATGA	GGTATAGA	CTGGAT	KVGVKQGDSMSPLL
	GTCTCAGCAAGGCAGGAAATGGCTCCCTCGCGGTTGTGTTCGTCGACTT	GGGCGTG	GCAGCG	FNLALDPLIQQLEREG
	TGCGAAGGCCTTCGATACCGTCTCCCACGAGCACCTCCTGAGTGTTCTG	CGG (SEQ	ACGTTG	RGFPVNGKSITAMAF
	GTGCAGAAAGGCTTGGACCAACACATGGTGGAGTTGATCAAGGACTCC	ID NO:	GTGACA	ADDLAIVSDSWEGM
	TACGAGAACAGCGTGACCAAGGTGCACTGTCAGGAGGGTTGTTCCACT	1144)	TAAGCA	RANLDILVDFCELTG
	GACATCGCCATGAAGGTGGGAGTGAAGCAGGGTGACTCCATGTCCCCT		ATCGCT	MRTQPSKCHGFLIEK
	CTCCTCTTTAACCTGGCGCTGGATCCGCTTATCCAGCAACTTGAACGCGA		AAGTCG	SGSRSYKVNRCEPWL
	GGGCCGGGGCTTCCCAGTAAATGGGAAGTCCATTACTGCGATGGCATTT		GGGTAG	LNDTALHMVGPKESI
	GCGGATGACTTGGCCATAGTGAGTGACTCTTGGGAAGGCATGAGAGCC		GGGAGG	KYLGVQVNPWTGIF
	AACCTTGATATCCTGGTGGACTTCTGCGAGCTTACTGGAATGCGAACCC		TGGGGA	AEDTVAKLRQWVVA
	AGCCCAGTAAGTGCCACGGGTTCCTGATTGAGAAGAGTGGCAGCAGGT		CCTCGG	ISKTPLRPLDKVSLLC
	CGTACAAAGTGAACAGGTGCGAACCGTGGCTGCTGAACGACACAGCTC		CACGGC	QFAVPRVIFVADHC
	TTCACATGGTCGGGCCTAAGGAATCAATCAAGTACCTGGGCGTCCAGGT		TGTAGG	MLSAKALTEMDRSIR
	GAACCCGTGGACAGGGATCTTCGCTGAGGATACGGTTGCCAAACTACG		AACGGG	QAVKRWLHLARCTT
	ACAGTGGGTAGTTGCAATCTCCAAGACGCCTCTACGTCCGCTTGACAAG		TGTATG	NGLLYSRKSSGGLGIP
	GTGTCCCTGTTGTGCCAGTTTGCCGTACCGAGGGTCATCTTCGTGGCTG		GGCTCC	KLSMIVPAMQARRLL
	ATCACTGCATGCTATCTGCGAAGGCCCTGACAGAAATGGATAGGAGCAT		GGCAGC	GLSRSKDETVRWMF
	AAGACAAGCAGTGAAGAGGTGGTTGCACCTGGCCAGGTGTACCACGAA		CGTCGT	LETTDHVAFERAWLR
	CGGCCTCCTCTACTCAAGGAAATCCAGCGGTGGTCTGGGTATCCCAAAA		CACTCCC	AGGSPDEVPELGPDL
	TTGTCGATGATTGTTCCGGCCATGCAGGCCAGGAGACTCCTGGGCCTGT		ATACAA	VEGSPAEGNADPVST

		CCCGTTCTAAGGACGAGACGGTCAGGTGGATGTTTCTGGAGACAACTG		CACAGG	VRPRKRIVPCDWRQ
		ATCACGTGGCGTTTGAGAGGGCATGGCTGAGGGCTGGAGGGTCGCCA		GGCTGC	VEFDRWAGQLVQGK
		GATGAGGTACCGGAGCTGGGTCCGGATCTGGTGGAGGGCTCCCCTGCG		ATCCTG	GIRTFEADKISNCWLY
		GAGGGGAACGCTGACCCTGTCAGCACGGTGAGGCCAAGGAAGCGCAT		GTGGCC	DYPPNKLKPGDFTAA
		AGTCCCGTGTGACTGGCGTCAAGTCGAGTTCGACAGATGGGCCGGTCA		GGTGCT	VQLRANVYPTRELAG
		ATTGGTGCAGGGAAAAGGGATTCGGACGTTCGAAGCGGACAAGATCA		AGTTGG	RGRTDTIDVCCRHCG
		GCAACTGCTGGTTGTACGACTACCCGCCAAACAAGCTGAAGCCTGGGG		TTCTGG	EAPETCWHILALCPK
		ATTTTACGGCGGCTGTCCAGCTTAGAGCGAACGTTTACCCGACCCGGGA		AAGCCC	VKRCRIQRHHKVCQV
		GCTAGCGGGTCGCGGAAGGACCGATACGATAGATGTCTGTTGTCGACA		GCCCGG	LVAEAERHGWEVER
		CTGTGGGGAGGCCCCAGAGACTTGCTGGCACATCCTTGCGCTCTGCCCG		GCTGGT	EKRWMLPSGECVAP
		AAGGTTAAGCGGTGCCGTATTCAGAGGCACCACACAAGGTGTGCCAGGTC		TCGCAG	DLICWLDELALIVDVT
		CTCGTCGCGGAGGCTGAGCGCCATGGATGGGAAGTGGAAAGGGAAAA		AAGCAG	VRYEFDEESLERARIE
		GCGCTGGATGCTGCCCTCCGGGGAGTGTGTCGCGCCGGACCTGATCTG		GGTGCG	KECKYRPLIPVIRASR
		CTGGTTGGATGAGCTGGCGCTCATTGTCGATGTGACGGTGAGGTACGA		CCCAGG	VQTKKVTVYGFPLGA
		GTTCGATGAGGAGTCGCTAGAACGCGCGCGAATCGAGAAGGAATGCAA		GTAGGT	RGKWPAKNELLLADL
		GTACCGCCCTCTCATTCCAGTGATCAGGGCGAGCAGAGTTCAGACGAAG		TTGGTAT	GLSKARTRSFAKLLSR
		AAGGTGACGGTCTATGGCTTCCCTCTGGGAGCCAGGGGAAAGTGGCCT		ATCTGG	RVLLHSLDVMRTFM
		GCTAAGAACGAGCTGCTGCTCGCCGACCTCGGCCTGAGCAAGGCTCGG		GTCCGG	R (SEQ ID NO: 1389)
		ACTCGGAGTTTTGCTAAACTCCTGAGCCGCAGAGTTCTCTTACATTCTCT		TGCGAT	
		GGATGTTATGAGGACGTTTATGCGTTAAGGAGGGGAGTAGGTCTCTAC		ACCTATC	
		TCTGACCCGAAGGGCCCCCCCGTTTCAGACCTGATTCTAGGCTACCTGTG		GATGGG	
		CCTAATTGGGGGGGGTCCCAAAGAGATGTTGTCTGTTGTAGAAGGGTTTG		CAGCGA	
		CGCCACTGACTGCACGGAAGGGTGGGCCTCGACAGGTAGGGGTTACAT		GGGCCG	
		GACTCCGTGCTGCTCAGCAGACCCGCGCCTCTGAGACCGGGTAGGGCT		CCTCGT	
		ACTTGAACAAGCGACGCCCTGGTGTATGTCCGTATCCTAACCTGGTTTG		GACGCG	
		GGAAAGCCGATACCGGCAATGCCCGCCACAGGTGTCGCGCACCCCACG		CTGTGT	
		GGATGACGTATGGGCCCCGGGGGACCTCATGGATACTCCACTGGACTT		GGAGCT	
		GCACAATCCTGGTGTACTGGATGCAGCGACGTTGGTGACATAAGCAATC		GGAGCC	
		GCTAAGTCGGGGTAGGGGAGGTGGGGACCTCGGCACGGCTGTAGGAA		GGCCTG	
		CGGGTGTATGGGCTCCGGCAGCCGTCGTCACTCCCATACACAACAGGG		GGTATG	
		GCTGCATCCTGGTGGCCGGTGCTAGTTGGTTCTGGAAGCCCGCCCGGGC		AACAGT	
		TGGTTCGCAGAAGCAGGGTGCGCCCAGGGTAGGTTTGGTATATCTGGG		TCTTGC	
		TCCGGTGCGATACCTATCGATGGGCAGCGAGGGCCGCCTCGTGACGCG		GGATGT	
		CTGTGTGGAGCTGGAGCCGGCCTGGGTATGAACAGTTCTTGCGGATGT		GGCGTA	
		GGCGTAGCTAGATAGTACCCGTGGTTGTGGGCGTGGTGTCGACCAAAT		GCTAGA	
		GTTGTCCTGTGTGCACATAGGCCAAGGGTTACGTGGGTGGCAGTCAGA		TAGTAC	
		AGCACCCGCACCTGGAAGTGATTGCCCCGGGATCCCGGCTCTCTGTGAA		CCGTGG	
		GAGCTACCTTGAGGAAAGGTGTTCCGCTGGAACTCAAGACCCTACAGTA		TTGTGG	

	GGGGATATCAACTGGCTTTGAGGTGCTGTGATTCCGGAACCAGGGCGA GGGCGAGTACTTAGAGCATGTCCAAAAGCCCGGGGAACGTTCCGGGGGG CCTGCTTGGGTCGTTGGACCCACATCCGTAAAACGATGGATCTCGCGTC GGCGCTCGGGAGAACTTCCCGCATGAACGCTGATTGCATGTGAGAACG CCCCCACGGCGGCGGGGCAGGCGCTCCCCCTGGGTGTAAGGCTCGGGG GGGTCACGGCTCCGCTCTAAAAG (SEQ ID NO: 1543)					
						GCGTGG TGTCGA CCAAAT GTTGTC CTGTGT GCACAT AGGCCA AGGGTT ACGTGG GTGGCA GTCAGA AGCACC CGCACC TGGAAG TGATTG CCCCGG GATCCC GGCTCT CTGTGA AGAGCT ACCTTG AGGAAA GGTGTT CCGCTG GAACTC AAGACC CTACAG TAGGGG ATATCA ACTGGC TTTGAG GTGCTG TGATTCC GGAACC AGGGCG AGGGCG AGTACTT AGAGCA

R2	R2_BM	AB076841	Bombyx mori	GGGCGATACGCATAATTTTAATTTCCCGATTGAAATCCAGTCGTCTTAATCTGGTGACCAGTGGCGCGGTCACCAGTATAGTGCACAGGACGTGAATGGCTCCGAGGCTGGCGGAGTCACTCACTATAAGTGTGAGAGACGATGTC	GGGCGAT ACGCATAA TTTTAATTT	GCCTTG CACAGT AGTCCA	TGTCCA AAAGCC CGGGGA ACGTTCC GGGGGC CTGCTT GGGTCG TTGGAC CCACATC CGTAAA ACGATG GATCTC GCGTCG GCGCTC GGGAGA ACTTCCC GCATGA ACGCTG ATTGCAT GTGAGA ACGCCC CCACGG CGGCGG GGCAGG CGCTCCC CCTGGG TGTAAG GCTCGG GGGGGT CACGGC TCCGCTC TAAAAG (SEQ ID NO: 1267)	MMASTALSLMGRC NPDGCTRGKHVTAA PMDGPRGPSSLAGT

		CTGTGCCAAGTATACGTCCAACCCTAACGGGTTAAGTGAAATTAGTTGC	CCCGATTG	GCGGTA	FGWGLAIPAGEPCG
		TCATAACAGGGACGGTGTACCTGTTTGCTCGTGGCTGGCTATCGAATGG	AAATCCAG	AGGGTG	RVCSPATVGFFPVAK
		ACGGGACCAATACACCCCCCTGTTAGTAATGGGGTAAGAGAGAGCGGT	TCGTCTTA	TAGATC	KSNKENRPEASGLPL
		CTGAAACTATGGCCGAAATCACGACGCCCCACTCCTACCCATAACCTGCA	ATCTGGTG	AGGCCC	ESERTGDNPTVRGSA
		CGTGGTACCGCCGCACATTGACCGATACGGGAGGAGGGGCAGCACTTG	ACCAGTGG	GTCTGTT	GADPVGQDAPGWT
		AATCACGTAGTCTTGGTGTAGCCATTGCGGGACTACAGCCCTCGTAAGT	CGCGGTCA	TCTTCCC	CQFCERTFSTNRGLG
		GCCGCCTTAGAACGCAACGGGGCAATAGGTGGGCCGGGGCGCTAGCG	CCAGTATA	CGGAGC	VHKRRAHPVETNTD
		GGGGGGAGTAATCTCCCCTGTTGGCGTGCACCGCACTGCTCCCACTGGG	GTGCACAG	TCGCTCC	AAPMMVKRRWHGE
		GGCAGTGTCATCCGGAAACAGGTGGGCCGGGGCGCCACCAGGGGGGGA	GACGTGAA	CTTGGC	EIDLLARTEARLLAER
		GCAATCCCTCCTGATGATGGCGAGCACCGCACTGTCCCTTATGGGACGG	TGGCTCCG	TTCCCTT	GQCSGGDLFGALPGF
		TGTAACCCGGATGGCTGTACACGTGGTAAACACGTGACAGCAGCCCCCG	AGGCTGGC	ATATTTA	GRTLEAIKGQRRREP
		ATGGACGGACCGCGAGGACCGTCAAGCCTAGCAGGTACCTTCGGGTGG	GGAGTCAC	ACATCA	YRALVQAHLARFGSQ
		GGCCTTGCGATACCTGCGGGCGAACCCTGTGGTCGGGTTTGCAGCCCG	TCACTATA	GAAACA	PGPSSGGCSAEPDFR
		GCCACAGTGGGTTTTTTTCCTGTTGCAAAAAAGTCAAATAAAGAAAATA	AGTGTGAG	GACATT	RASGAEEAVEERCAE
		GACCTGAAGCCTCTGGCCTCCCGCTGGAGTCAGAGAGGACAGGCGATA	AGACGATG	AAACAT	DAAAYDPSAVGQMS
		ACCCGACTGTGCGGGGTTCCGCCGGCGCAGATCCTGTGGGTCAGGATG	TCCTGTGC	CTACTG	PDAARVLSELLEGAG
		CGCCTGGTTGGACCTGCCAGTTCTGCGAACGAACCTTTTCGACCAACAG	CAAGTATA	ATCCAAT	RRRACRAMRPKTAG
		GGGTTTGGGTGTCCACAAGCGTAGAGCCCACCCTGTTGAGACCAATACG	CGTCCAAC	TTCGCC	RRNDLHDDRTASAH
		GATGCCGCTCCGATGATGGTGAAGCGGCGGTGGCATGGCGAGGAAATC	CCTAACGG	GGCGTA	KTSRQKRRAEYARVQ
		GACCTCCTCGCTCGCACCGAGGCCAGGTTGCTCGCTGAGCGGGGTCAGT	GTTAAGTG	CGGCCA	ELYKKCRSRAAAEVID
		GCTCGGGTGGAGACCTCTTTGGCGCGCTTCCAGGGTTTGGAAGAACTCT	AAATTAGT	CGATCG	GACGGVGHSLEEME
		GGAAGCGATTAAGGGACAACGGCGGAGGGAGCCTTATCGGGCATTGG	TGCTCATA	GGAGGG	TYWRPILERVSDAPG
		TGCAAGCGCACCTTGCCCGATTTGGTTCCCAGCCGGGTCCCTCGTCGGG	ACAGGGA	TGGGAA	PTPEALHALGRAEW
		GGGGTGCTCGGCCGAGCCTGACTTCCGGCGGGCTTCTGGAGCTGAGGA	CGGTGTAC	TCTCGG	HGGNRDYTQLWKPI
		AGCGGTCGAGGAACGATGCGCCGAAGACGCCGCTGCCTATGATCCATC	CTGTTTGC	GGATCT	SVEEIKASRFDWRTS
		CGCAGTCGGTCAGATGTCGCCCGATGCCGCTCGGGTTCTCTCCGAACTC	TCGTGGCT	TCCGATC	PGPDGIRSGQWRAV
		CTTGAGGGTGCGGGGAGAAGACGAGCGTGCAGGGCTATGAGACCCAA	GGCTATCG	CTAATCC	PVHLKAEMFNAWM
		GACTGCAGGGCGGCGAAACGATTTGCACGATGATCGGACACGCTAGTGC	AATGGACG	ATGATG	ARGEIPEILRQCRTVF
		CCACAAAACCAGTAGACAAAAGCGCAGGGCAGAGTACGCGCGTGTGCA	GGACCAAT	ATTACG	VPKVERPGGPGEYRP
		GGAACTGTACAAGAAGTGTCGCAGCAGAGCAGCAGCTGAGGTGATCGA	ACACCCCC	ACCTGA	ILIASIPLRHFHSILARR
		TGGCGCGTGTGGGGGGTGTCGGACACTCGCTCGAGGAGATGGAGACCTA	CTGTTAGT	GTCACT	LLACCPPDARQRGFI
		TTGGCGACCTATCCTCGAGAGAGTGTCCGATGCACCTGGGCCTACACCG	AATGGGGT	AAAGAC	CADGTLENSAVLDAV
		GAAGCTCTTCACGCCCTAGGGCGTGCGGAGTGGCACGGGGGCAATCGC	AAGAGAG	GATGGC	LGDSRKKLRECHVAV
		GACTACACCCAGCTGTGGAAGCCGATCTCGGTGGAAGAGATCAAGGCC	AGCGGTCT	ATGATG	LDFAKAFDTVSHEAL
		TCCCGCTTTGACTGGCGAACTTCGCCGGGCCCGGACGGTATACGTTCGG	GAAACTAT	ATCCGG	VELLRLRGMPEQFCG
		GTCAGTGGCGTGCGGTTCCTGTGCACTTGAAGGCGGAAATGTTCAATGC	GGCCGAA	CGATGA	YIAHLYDTASTTLAVN
		ATGGATGGCACGAGGCGAAATACCCGAAATTCTACGGCAGTGCCGAAC	ATCACGAC	AAA	NEMSSPVKVGRGVR
		CGTCTTTGTACCTAAGGTGGAGAGACCAGGTGGACCGGGGGGAATATCG	GCCCCACT	(SEQ ID	QGDPLSPILFNVVMD

		ACCGATCTTGATCGCGTCGATTCCCCTGAGACACTTTCACTCCATCTTGG	CCTACCCA	NO: 1268)	LILASLPERVGYRLEM
		CCCGGAGGCTGTTGGCTTGCTGCCCCCCTGATGCACGACACGCGCGGATT	TAACCTGC		ELVSALAYADDLVLLA
		TATCTGCGCCGACGGTACGCTGGAGAATTCCGCAGTACTGGACGCGGT	ACGTGGTA		GSKVGMQESISAVDC
		GCTTGGGGATAGCAGGAAGAAGCTGCGGGAATGTCACGTGGCGGTGC	CCGCCGCA		VGKQMGLRLNCRKS
		TAGACTTCGCCAAGGCATTTGACACAGTGTCTCACGAGGCACTTGTCGA	CATTGACC		AVLSMIPDGHRKKH
		ATTGCTGAGGTTGAGGGGCATGCCCGAACAGTTCTGCGGCTACATTGCT	GATACGG		HYLTERTFNIGGKPLR
		CACTTATACGATACGGCGTCCACCACCTTAGCCGTGAACAATGAAATGA	GAGGAGG		QVSCVERWRYLGVD
		GCAGCCCTGTGAAAGTGGGACGAGGGGTTCGTCAAGGGGACCCTCTGT	GGCAGCAC		FEASGCVTLEHSISSA
		CGCCGATACTCTTCAACGTGGTGATGGACCTCATCCTAGCTTCCCTGCCG	TTGAATCA		LNNISRAPLKPQQRL
		GAGAGGGTCGGGTATAGGTTGGAGATGGAACTTGTGTCCGCTCTGGCC	CGTAGTCT		EILRAHLIPRFQHGFV
		TATGCTGACGACCTAGTCCTGCTTGCGGGGTCGAAGGTAGGGATGCAG	TGGTGTAG		LGNISDDRLRMLDVQ
		GAGTCCATCTCTGCTGTGGACTGTGTTGGTAAGCAGATGGGCCTACGCC	CCATTGCG		IRKAVGQWLRLPAD
		TGAATTGCAGGAAGAGCGCGGTTCTGTCTATGATACCGGATGGCCACC	GGACTACA		VPKAYYHAAVQDGG
		GCAAGAAGCATCACTACCTGACTGAGCGAACCTTCAATATTGGAGGTAA	GCCCTCGT		LAIPSVRATIPDLIVRR
		GCCGCTCAGGCAGGTGAGTTGTGTTGAGCGGTGGCGATATCTTGGTGT	AAGTGCCG		FGGLDSSPWSVARA
		CGATTTTGAGGCCTCTGGATGCGTGACATTAGAGCATAGTATCAGTAGT	CCTTAGAA		AAKSDKIRKKLRWA
		GCTCTGAATAACATCTCAAGGGCACCTCTCAAACCCCAACAGAGGTTGG	CGCAACGG		WKQLRRFSRVDSTT
		AGATTTTGAGAGCTCATCTGATTCCGAGATTCCAGCACGGTTTTGTGCTT	GGCAATAG		QRPSVRLFWREHLH
		GGAAACATCTCGGATGACCGATTGAGAATGCTCGATGTCCAAATCCGGA	GTGGGCC		ASVDGRELRESTRTP
		AAGCAGTCGGACAGTGGCTAAGGCTACCGGCGGATGTGCCCAAGGCAT	GGGGCGC		TSTKWIRERCAQITG
		ACTATCACGCCGCAGTTCAGGACGGCGGCTTAGCGATCCCATCGGTGCG	TAGCGGG		RDFVQFVHTHINALP
		AGCGACCATCCCGGACCTCATTGTGAGGCGTTTCGGGGGGCTCGACTCG	GGGGAGT		SRIRGSRGRRGGGES
		TCACCATGGTCAGTGGCAAGAGCCGCCGCCAAATCTGATAAGATTCGTA	AATCTCCC		SLTCRAGCKVRETTA
		AGAAACTGCGGTGGGCCTGGAAACAGCTCCGCAGGTTCAGCCGTGTTG	CTGTTGGC		HILQQCHRTHGGRIL
		ACTCCACAACGCAACGACCATCTGTGCGCTTGTTTTGGCGAGAACATCT	GTGCACCG		RHNKIVSFVAKAMEE
		GCACGCATCTGTTGATGGACGCGAACTTCGCGAATCCACACGCACCCCG	CACTGCTC		NKWTVELEPRLRTSV
		ACATCCACAAAGTGGATTAGGGAGCGATGCGCGCAGATAACCGGACGG	CCACTGGG		GLRKPDIIASRDGVG
		GACTTCGTGCAGTTCGTGCACACTCATATCAACGCCCTCCCATCCCGCAT	GGCAGTGT		VIVDVQVVSGQRSLD
		TCGCGGATCGAGAGGGCGTAGAGGTGGGGGTGAGTCTTCGTTGACCTG	CATCCGGA		ELHREKRNKYGNHG
		CCGTGCTGGTTGCAAGGTTAGGGAGACGACGGCTCACATCCTACAACA	AACAGGTG		ELVELVAGRLGLPKAE
		GTGTCACAGAACACACACGGCGGCCGGATTCTACGACACAACAAGATTGTA	GGCCGGG		CVRATSCTISWRGV
		TCTTTCGTGGCGAAAGCCATGGAAGAGAACAAGTGGACGGTTGAGCTG	GCGCCACC		WSLTSYKELRSIIGLRE
		GAGCCGAGGCTACGAACATCGGTTGGTCTCCGTAAGCCGGATATTATCG	AGGGGGG		PTLQIVPILALRGSHM
		CCTCCAGGGATGGTGTCGGAGTGATCGTGGACGTGCAGGTGGTCTCGG	AGCAATCC		NWTRFNQMTSVMG
		GCCAGCGATCGCTTGACGAGCTTCACCGTGAGAAACGTAATAAATACGG	CTCCTG		GGVG (SEQ ID NO: 1390)
		GAATCACGGGGAGCTGGTTGAGTTGGTCGCAGGTAGACTAGGACTTCC	(SEQ ID		
		GAAAGCTGAGTGCGTGCGAGCCACTTCGTGCACGATATCTTGGAGGGG	NO: 1145)		
		AGTATGGAGCCTGACTTCTTATAAGGAGTTAAGGTCCATAATCGGGCTT			

			CGGGAACCGACACTACAAATCGTTCCGATACTGGCGTTGAGAGGTTCAC ACATGAACTGGACCAGGTTCAATCAGATGACGTCCGTCATGGGGGGCG GCGTTGGTTGAGCCTTGCACAGTAGTCCAGCGGTAAGGGTGTAGATCA GGCCCGTCTGTTTCTTCCCCGGAGCTCGCTCCCTTGGCTTCCCTTATATTT AACATCAGAAACAGACATTAAACATCTACTGATCCAATTTCGCCGGCGT ACGGCCACGATCGGGAGGGTGGGAATCTCGGGGATCTTCCGATCCTAA TCCATGATGATTACGACCTGAGTCACTAAAGACGATGGCATGATGATCC GGCGATGAAAA (SEQ ID NO: 1544)			
R2	R8Hm-A	.	Hydra vulgaris	TTCAAGTGGATGAAGCTGGGAAGGTAATCTGTAGTTGGTTGAGTTGGTT GCAGATTACTGCTGTCGATTTTGCTTTCTATTGAAAGCCTGTCTCTACGG GTCCTGAAGCTTGAATTTTGGTAGCTATAGTTTTGTGGGAGGAAAGTGG AATTTTGTACCATCTTTTGTCTCTCGTATCTACTATAGTAAATCCGGTCAT GCAGCCTCTACGCGGCGCAACTAGAAACTTGGATCAGTGATCAAGGCTA ATGCATGCCGGGTCTCCTCAGATTAGGAGTATAATACAAATCTGACTTC ATCACTAAGAGGCTATGGGGCTAACGATCCTATAGTCTCGATGAACCTA TTGATTGTTACTAGTAGCATAAAAGAAAGTGACGTACCCTCTAGTGGAA AGGGGGGTGTAGCAGTCAATAACATAACAGCAGGAGCTAGTGGAAAA GATACGTGCGTGATCATACACCCAGGTACCGATGGTATTTGGTGCTGTA CTGAGTGTGTAGAGATACATAACAGCGGTAAGGATCTGAAACGACATCT TGCAAAACGTCACCCGAGTGTAACGATAAGCGGTTACAAATGCAATCTG TGTCCATTTGTTAGTGAACGCCAACTAAGTGTGGGGACACATCTGAGGT ACTGCAGAGGCGTAAAAGAAGTGGTTAAAAGAGAGTTTGCATGCGCGA GCTGCTCTTTTTCTTCGGATACGTTCTCAGGACTTCAGGTGCATATGCAA AGAAAGCATATAGCAGAATGGAACGACCAGCTGAAGGAGAAAACGGA GTTTGCTTGGACAGACCGAGAATTGAGGGAGCTTGCTGAGAAGGAACT TACCACTCCTTCCTTCAGGTACAACAAAATTTTCTATGCTGCGCTAGGTA CCTCCCGGACCTACGACGCTGTGAGGAAAATTCGCTATAATGACAGATA CAAATCTGCCATTGCTGAAATGCGATCACAGATAGCAGATGCGGCTGCC GCTGCACAAGAGAGGGATGTAGAGCGGGGTTTAGTTTCAGCACACTCA GACAGAGGAAAAGAAATGCTCCCTGTTGTTGAAACCAAAAGTGATATCC AAGTAAACAACGATATCAAAAAGGATATTGAATTAACACCGAATTCAAG ACAGAAACAAACTAATCTAGCGCTGGCAAGGCCAGCTGTAATTGAGGT GGAGGAAGACTTGGGTAGGCAGGATGTGAAACAATATCTCGCATCCCT GCGCCAAGACGACTACACACAAGTCCGGCCGAGCGGTCAATCTTTGCATAC TGCAGGGAGGAAACCAATTGGTCTGCGACAAAAAGACAGGTATTAAAG ATATCGAGAACTACCAGAGGTTTAAGACAACCTAAGAAGGTTCGTCCAT TTGAGTTTCCGGAAGGGTTCAAACCTAACAGAAATATGAGAAAGTGGA GAAAGTATAGATTCCTTCAGGAATGCTATAGGGAAAAGAGAGCTGAGA	TTCAAGTG GATGAAGC TGGGAAG GTAATCTG TAGTTGGT TGAGTTGG TTGCAGAT TACTGCTG TCGATTTT GCTTTCTA TTGAAAGC CTGTCTCT ACGGGTCC TGAAGCTT GAATTTTG GTAGCTAT AGTTTTGT GGGAGGA AAGTGGA ATTTTGTA CCATCTTTT GTCTCTCG TATCTACT ATAGTAAA TCCGGTCA TGCAGCCT CTACGCGG CGCAACTA GAAACTTG GATCAGTG	TAAATG CCAAAA GTTGCTT GGGCTA AATGAT ACGTAC GCTAGA AAAAGC GACTTG CTGCAC GGATGA CGGTTC ATCAGA GCCCGA TATGTG CATGTC AAGGCG GCAGGG AGAATC ACTAGT GTAGCT GTTCTTT CCATTAC GACTTA CGCGGT TAACGT GGCACG ATAGAT TTACACC AGGAAA	MNLLIVTSSIKESDVP SSGKGGVAVNNITAG ASGKDTCVIIHPGTD GIWCCTECVEIHNSG KDLKRHLAKRHPSVTI SGYKCNLCPFVSERQ LSVGTHLRYCRGVKE VVKREFACASCSFSSD TFSGLQVHMQRKHI AEWNDQLKEKTEFA WTDRELRELAEKELT TPSFRYNKIFYAALGT SRTYDAVRKIRYNDR YKSAIAEMRSQIADA AAAAQERDVERGLV SAHSDRGKEMLPVV ETKSDIQVNNDIKKDI ELTPNSRQKQTNLAL ARPAVIEVEEDLGRQ DVKQYLASLRQDDYT SPAERSIFAYCREETN WSATKRQVLKISRTT RGLRQPKKVRPFEFP EGFKPNRNMRKWR KYRFLQECYREKRAET VSKILDGTFIDEPEEEI RPELEEVQRMYIDRL EKRTQLDTTKIVQTD EVFCLQSYGRITIGEV RDALGASKKDSASGP

		CTGTTAGCAAGATCCTGGACGGGACTTTTATCGATGAACCGGAGGAAG	ATCAAGGC	TAATAC	DGLLLQDVRRLGPLLL
		AGATTAGACCAGAGTTAGAGGAAGTACAACGTATGTACATTGACCGGCT	TAATGCAT	GTGAAG	CNIFNMWYLHGIPVE
		GGAGAAAAGAACTCAGCTGGATACCACGAAGATTGTGCAAACAGACGA	GCCGGGTC	GGTTCC	ENRCRTILLYKSGDRH
		GGTGTTTTGTCTGCAAAGCTACGGTCGCATTACGATCGGGGAAGTAAGA	TCCTCAGA	ACCATAT	LASNYRPVTIGNMLN
		GATGCACTCGGTGCAAGCAAGAAGGACTCGGCCTCGGGTCCTGACGGC	TTAGGAGT	ACTGGA	RLYAKIWDKRIRKNV
		CTGCTTCTACAGGATGTGAGGAGGCTGGGACCACTATTATTGTGTAACA	ATAATACA	GTTTAG	RLHVRQKAFIPVDGC
		TCTTTAACATGTGGTACTTACATGGGATCCCTGTGGAAGAAAACAGGTG	AATCTGAC	ATCTATG	FENVKTIQCVLQSYR
		TCGAACAATACTCTTATACAAGAGTGGCGATAGACATCTGGCATCAAAC	TTCATCAC	AGGGAA	KRKLEHNVVFIDLAK
		TATAGACCTGTGACAATCGGCAACATGCTGAACAGGCTTTACGCCAAAA	TAAGAGGC	ACATTTG	AFDTVLHDSIRKALW
		TCTGGGACAAACGGATCCGGAAGAACGTGCGTCTTCATGTGAGGCAAA	TATGGGGC	TAATAA	RKGVPSGVVKVVDSL
		AAGCATTTATCCCGGTGGATGGGTGCTTTGAGAACGTAAAAACGATCCA	TAACGATC	GTCAGT	YAGAVTSISVGKTKTR
		ATGCGTTCTCCAGTCTTACAGAAAGCGTAAGTTGGAACACAACGTCGTA	CTATAGTC	CTGGTA	SICINSGVKQGCPLSP
		TTTATTGATCTTGCCAAGGCCTTTGACACGGTCTTGCATGACTCGATAAG	TCG (SEQ	ACCTGG	LLFNLILDELAERIEAT
		GAAAGCGTTGTGGCGGAAAGGTGTTCCGTCTGGGGTTGTTAAAGTGGT	ID NO:	CGCCGC	GCGLDLDGHVLSSM
		AGACAGCTTATATGCGGGAGCTGTCACAAGCATAAGTGTTGGAAAAAC	1146)	TGTTGA	AFADDYVLLAKDSVE
		GAAAACTCGTTCTATATGTATAAACTCTGGAGTCAAGCAGGGTTGTCCT		GTCAAA	MNELIRVCSTFFKEK
		CTGTCACCTCTTCTATTCAACCTAATACTGGATGAACTAGCGGAGAGGAT		TTAACTA	GLSVNPGKCQSLRVL
		AGAGGCAACCGGCTGCGGGTTAGATCTTGATGGTCACGTTCTATCATCT		TGTCAAT	PVKEKKRSMKVLVRP
		ATGGCCTTTGCTGACGACTACGTGTTGCTAGCGAAGGACTCCGTGGAGA		ACTCATT	HRWWRIKDQDVDIP
		TGAACGAGTTGATAAGAGTGTGTAGTACATTCTTCAAAGAGAAAGGCTT		AAGTTA	SMTYDSLGKYLGVSI
		ATCTGTAAACCCAGGTAAATGTCAATCGCTAAGAGTTCTTCCCGTAAAG		TCGACTT	DPTGKIALPIEEWKN
		GAGAAGAAACGGTCAATGAAGGTCCTTGTTAGACCTCATAGATGGTGG		TGATAT	WMTKLKECKLKPEQ
		AGGATAAAAGACCAGGATGTTGACATCCCATCTATGACATATGACAGCT		GGCATG	KVKILKEVVCSRVNYV
		TAGGAAAATACCTTGGTGTTTCGATTGACCCAACTGGTAAGATAGCGCT		GGGTGA	LRMSECGISELRSWT
		TCCGATTGAGGAGTGGAAGAATTGGATGACCAAGCTAAAAGAGTGTAA		TTCCGC	RFVRNWAKNIIHLPT
		GCTCAAGCCCGAGCAGAAAGTTAAAATTCTGAAAGAAGTGGTTTGCTCT		GTTATAT	WCSSDWIHSIKGLGI
		CGGGTAAACTACGTTTTGCGGATGTCAGAGTGTGGCATCAGCGAACTTC		CAAAGT	PDVSKGIVIQRMRAS
		GGAGTTGGACACGATTTGTAAGGAATTGGGCGAAAAACATCATTCACTT		CAAACA	EKMSTSEDGIVRVVG
		ACCCACATGGTGCAGTAGTGACTGGATACACTCGATCAAAGGGTTAGG		TGATGA	ARLVQKNRVLWEKA
		CATTCCCGACGTTTCGAAGGGAATTGTCATACAACGTATGAGGGCTTCG		TTGCAAT	GFEGIELKAARRHCE
		GAGAAAATGTCTACGTCTGAAGACGGTATAGTCCGCGTGGTTGGTGCA		GAGAAA	VERLNNIGNITNGVA
		CGACTTGTTCAGAAGAACAGAGTCTTGTGGGAAAAGGCCGGTTTCGAA		CTACCAC	LKTIAAVSSVNRYW
		GGTATCGAACTGAAGGCAGCCAGGAGGCACTGCGAAGTGGAGAGACT		GCTTGG	MIEDNLKSGNKILVW
		CAACAACATTGGTAACATTACCAACGGCGTTGCACTCAAAACTATCGCA		TCACGTT	KAMAGAIPTKINLSR
		GCAGTCTCCTCGGTAAATCGGTACTGGATGATTGAAGACAACTTGAAAT		TGTGAG	GVADQTLKKCRRCGL
		CCGGGAACAAGATTCTCGTTTGGAAAGCAATGGCGGGTGCCATTCCAAC		GAGAAC	TAETDGHILAGCHTS
		AAAGATTAACCTTTCGCGGGGCGTAGCAGACCAGACCCTCAAAAAATGT		ATCTCAT	SDAYSKRHNMLCDKL
		CGTCGATGCGGTTTAACAGCGGAAACGGATGGACACATCTTGGCTGGA		TCAAGC	AKELKLNGGPNRRV

			TGCCATACTAGCAGCGACGCGTACTCAAAACGTCACAACATGCTCTGTG			CTCCCG	WRERTCFTSTGRRYR
			ATAAACTCGCCAAAGAGCTCAAACTCAATGGTGGACCAAACAGACGTGT			GATGTC	PDIIVKDDSKITVIDM
			GTGGCGCGAGAGGACGTGCTTCACTAGTACAGGCAGGCGATATAGACC			GGCACC	TCPYEKSEGHLIQCES
			TGACATTATCGTTAAAGATGACAGTAAAATCACAGTCATCGATATGACTT			CGCTGA	AKVTKYEPLKLDKYW
			GTCCGTATGAGAAATCAGAAGGACACCTGATCCAATGTGAAAGTGCGA			CATCTTC	TRELEGANGIVAEKV
			AAGTAACTAAATACGAGCCACTCAAGCTAGATAAGTATTGGACTCGAGA			TGGCTT	ELMGLAIGAIGTIMR
			ACTCGAGGGAGCAAATGGTATTGTTGCTGAAAAGGTAGAGCTGATGGG			ATGAAA	STLRKLCELKSGRIVR
			ATTGGCAATAGGGGCGATCGGCACAATCATGCGTAGTACCCTTCGGAA			ATTTTCA	RLQMIACNNSAQIIK
			ACTCTGTGAGTTAAAGTCGGGCAGGATCGTAAGACGTCTACAAATGATT			TTAATTT	GHLSRATRRNLR
			GCTTGTAATAATAGCGCCCAAATTATAAAGGGTCACCTGTCAAGGGCGA			TTGTAA	(SEQ ID NO: 1391)
			CTCGGAGGAATTTGCGGTGATAAATGCCAAAAGTTGCTTGGGCTAAATG			GTCATG	
			ATACGTACGCTAGAAAAAGCGACTTGCTGCACGGATGACGGTTCATCAG			GGCGGC	
			AGCCCGATATGTGCATGTCAAGGCGGCAGGGAGAATCACTAGTGTAGC			TTGAAA	
			TGTTCTTTCCATTACGACTTACGCGGTTAACGTGGCACGATAGATTTACA			GC (SEQ	
			CCAGGAAATAATACGTGAAGGGTTCCACCATATACTGGAGTTTAGATCT			ID NO:	
			ATGAGGGAAACATTTGTAATAAGTCAGTCTGGTAACCTGGCGCCGCTGT			1269)	
			TGAGTCAAATTAACTATGTCAATACTCATTAAGTTATCGACTTTGATATG				
			GCATGGGGTGATTCCGCGTTATATCAAAGTCAAACATGATGATTGCAAT				
			GAGAAACTACCACGCTTGGTCACGTTTGTGAGGAGAACATCTCATTCAA				
			GCCTCCCGGATGTCGGCACCCGCTGACATCTTCTGGCTTATGAAAATTTT				
			CATTAATTTTTGTAAGTCATGGGCGGCTTGAAAGC (SEQ ID NO: 1545)				
R2	R8Hm-B	Hydra vulgaris	CTTGGGGTCACTGACACATTTTTCGGTAGCCATAGTTTTTTGAGAGGAA	CTTGGGGT	ATGCCC		MSNRITIGDVPSVGK
			GAGTGGAAGTTTTTCCATGAGTCGTCTCTCGTATAAAACTGTGGTAAATCC	CACTGACA	GAGGTA		GGLTVNKQTAGADG
			GGCCATCCAGCCTCTACGCGGCGCAACTAGAAACTTGGATCAGTGATCA	CATTTTTC	GTTGGG		AEACVVIHPGAKGIW
			AGGCTAATGGATGACGGGACTCCATGGATAAGGAGATATAAAGATCTT	GGTAGCCA	ATAATG		SSPACLRKFTIGKELR
			ATTTGAACGCATCTTAAGGGGTTATGGGGCTAACACCCCCTTAATTCTG	TAGTTTTTT	ATGCAC		AHLAQIHKLAPSAVR
			GTGCACATTTATTGACCGTTATGAGCAATAGAATCACGATAGGTGATGT	GAGAGGA	AAGCTC		YRCNKCPYEGDVQLS
			ACCCTCGGTAGGAAAGGGGGGGTTTAACTGTCAATAAACAAACAGCAGG	AGAGTGG	GTAAGG		VGTHLRYCKGIAGVV
			AGCTGATGGTGCTGAAGCGTGTGTAGTCATACACCCAGGTGCCAAGGG	AAGTTTTT	CGACTT		EEKKQFACAICNFSSD
			TATTTGGTCCTCTCCTGCGTGTTTAAGAAAGTTTACGATCGGAAAAGAAC	CCATGAGT	GCTGCA		TFSGLQVHKQRKHV
			TAAGGGCACATTTGGCTCAAATTCATAAACTTGCACCGAGTGCAGTTCG	CGTCTCTC	CGTATG		VEWNEQLKEKTEFA
			GTACAGGTGTAATAAGTGTCCGTATGAGGGGTGATGTCCAACTCAGTGTG	GTATAAAC	CCGCTA		WTDRELRELAVKEVT
			GGAACACATCTGAGGTACTGTAAGGGTATTGCGGGAGTGGTGGAGGA	TGTGGTAA	AACGCT		IPFSVVNTETFAVLDI
			GAAAAAGCAATTCGCTTGCGCGATTTGTAATTTCTCTTCGGATACCTTTT	ATCCGGCC	TAGCTC		TTRTKDAVRKIRYTDR
			CAGGACTTCAGGTGCATAAGCAAAGAAAGCATGTAGTTGAATGGAACG	ATCCAGCC	GATGAG		YKSILAEVRAQVNAV
			AGCAGCTGAAAGAGAAAACGGAGTTTGCTTGGACAGACAGGGAACTGC	TCTACGCG	TGCATG		AEEAPQASDESQITLL
			GGGAGCTGGCGGTTAAGGAAGTAACGATTCCTTTCTCTGTGGTGAATAC	GCGCAACT	TCAAGA		VNTGRGAELQPAVIN
			GGAGACCTTTGCTGTGCTAGATATTACGACGCGGACTAAGGATGCTGTG	AGAAACTT	CGGTCG		ITDSIELVTDVNEVEM

		AGGAAAATTCGCTACACGGATAGATACAAATCTATCCTGGCTGAAGTAC	GGATCAGT	GGAGTA	VTSNSTNEEQPINAP
		GCGCACAAGTTAACGCTGTGGCGGAGGAAGCGCCGCAAGCTAGTGATG	GATCAAGG	TGATCA	VEPAVIEADLGRQDA
		AGAGTCAAATAACGCTCTTAGTTAACACAGGCAGGGGAGCAGAATTAC	CTAATGGA	GTGGAG	KLYLASLRQSDCTNA
		AACCTGCTGTGATTAATATAACTGATTCAATTGAATTAGTTACTGATGTC	TGACGGG	CTGACTT	SDRWTLAYCRGEVD
		AATGAGGTTGAAATGGTAACATCGAATTCAACCAATGAAGAACAGCCTA	ACTCCATG	TCCAGA	WCKTKSRLFKVSRHA
		TCAACGCGCCGGTGGAACCGGCTGTAATTGAGGCGGACTTGGGAAGAC	GATAAGG	CAACTC	RGLRQPQRVENWEF
		AGGATGCGAAACTATATCTCGCATCGCTGCGTCAAAGCGATTGCACAAA	AGATATAA	ACGCGG	PEGFRPNRNLRKWR
		CGCATCTGATCGATGGACCCTTGCGTATTGCAGGGGAGAAGTTGATTGG	AGATCTTA	ATTCGC	KYSFLQSCYRTKKKET
		TGTAAGACGAAAAGCAGGCTTTTCAAAGTATCAAGACATGCCCGGGGTT	TTTGAACG	GTGCGG	VSKILDGTFKDTPEEEI
		TAAGACAACCTCAAAGGGTGGAGAATTGGGAGTTTCCAGAGGGATTCA	CATCTTAA	TGGATA	RPELEEVQRVYVDRL
		GACCTAACAGGAACCTTCGTAAATGGAGGAAGTATTCATTCTTGCAAAG	GGGGTTAT	CAACAC	EVRTQLDTTRTVHID
		TTGCTATAGAACGAAGAAGGAAACTGTTAGTAAGATTCTTGATGGT	GGGGCTA	CTGGTA	ERFDLVSYGRITIREV
		ACTTTCAAGGACACACCTGAGGAAGAGATTAGGCCAGAGTTGGAGGAA	ACACCCCC	TAACAT	QDAISASKKDASGGP
		GTACAACGTGTGTACGTTGACCGGCTAGAGGTAAGAACTCAGCTGGAT	TTAATTCT	ATGAAG	DGLLLQDVKKASPRQ
		ACCACTAGGACAGTGCATATAGACGAAAGATTCGATTTAGTAAGCTATG	GGTGCACA	GGTTCC	LCIIFNMWYLHGIPV
		GTCGCATTACGATCAGGGAGGTACAAGACGCAATCAGCGCAAGCAAGA	TTTATTGA	ATCTAGT	VENRCRTILLHKGGE
		AGGATGCCTCAGGGGGTCCCGACGGCTTGCTCCTACAGGACGTGAAAA	CCGTT	ACAGGG	KHLTSNYRPVTIGNM
		AGGCGAGCCCACGCCAATTGTGTATCATCTTTAATATGTGGTACTTGCAT	(SEQ ID	ATAACG	LNRVYAKIWDRRIRK
		GGAATCCCTGTAGTGGAAAATAGGTGCCGAACAATACTCTTGCATAAGG	NO: 1147)	ATCCAT	NLQLHVRQKAFVPLD
		GTGGCGAGAAGCATCTAACGTCGAACTACCGACCTGTGACGATCGGCA		GGGAGC	GCFENVKTIQCILQSY
		ATATGCTGAATAGGGTATACGCTAAGATCTGGGACAGACGGATCAGAA		AAACTA	RRSRREHNVVFVDLA
		AAAACCTGCAACTTCATGTGAGACAGAAAGCATTCGTCCCGCTGGATGG		ATTAGTT	KAFDTILHDSIEKALLR
		GTGCTTTGAGAATGTAAAAACCATCCAATGCATTCTCCAGTCTTACAGAA		GGAGGT	KGIPRSVIKVVDSLYA
		GGAGCAGGCGGGAACACAATGTCGTATTTGTCGATCTTGCAAAAGCGTT		AATCCA	GAVTSITVGKTKTRPI
		TGATACGATTTTGCATGATTCGATAGAGAAAGCATTGCTGAGGAAAGGC		ACGCCG	CINSGVKQGCPLSPLL
		ATACCGCGAAGTGTGATAAAAGTGGTAGACAGCTTATATGCGGGAGCT		CTGTTG	FNLVIDELAERLEATG
		GTCACGAGCATTACGGTTGGGAAAACAAAGACTCGACCTATATGTATAA		AGTCAG	CGLDLEGHVISSMAF
		ATTCAGGGGTGAAGCAGGGTTGTCCTCTATCTCCTTTGCTGTTCAATCTA		TTTTTAA	ADDYVLLAKDSVEM
		GTAATAGATGAACTAGCGGAGAGGCTGGAGGCAACTGGCTGCGGTCTT		CCGCCA	NVLMNVCNTFFEEK
		GATCTGGAAGGTCACGTCATTTCTTCCATGGCTTTTGCTGATGACTACGT		GTCAAC	GLAVNPAKCQSLRVL
		GTTGTTGGCGAAAGACTCGGTTGAAATGAACGTGCTAATGAACGTGTG		TCTTGTA	PVKGKRSMKVLTRTH
		CAATACGTTCTTTGAGGAGAAGGGTTTAGCTGTAAATCCAGCAAAATGT		GGTTAT	RWWKINNQDVEIPS
		CAGTCGTTACGCGTTTTGCCTGTAAAAGGCAAACGGTCCATGAAAGTCC		CGGTCT	MTYESVGKYLGVMI
		TTACGAGGACGCATAGATGGTGGAAAATTAATAACCAGGATGTTGAAA		TCGGCA	DPAGKIALPIEEWKL
		TCCCATCTATGACATACGAAAGTGTTGGAAAATATCTTGGGGTAATGAT		GACCTT	WLTRLRECKLKPDQK
		TGACCCAGCTGGTAAGATTGCTCTTCCGATTGAGGAATGGAAGCTTTGG		GGACCG	VKVLKEVVCARANYV
		CTAACTAGGTTAAGGGAGTGTAAGCTCAAACCTGATCAAAAAGTGAAG		CCTAGC	LRMSGCGICELRKWS
		GTGCTGAAAGAGGTAGTTTGTGCCCGAGCAAACTATGTTCTCCGGATGT		GCCGGC	RFVRGWVKSIIHFPA

			CCGGGTGCGGAATCTGTGAGCTCCGTAAGTGGTCACGATTTGTGAGGG		CAACAG		WCNSEWMHSSKGL
			GATGGGTGAAATCCATCATTCACTTCCCCGCATGGTGCAATAGCGAATG		TTTGTCG		GIPDVVSGIVIQRMR
			GATGCATTCGAGCAAAGGCTTAGGCATTCCTGATGTAGTGTCAGGAATT		TCGACT		AAEKMAKSTDGVVR
			GTCATCCAACGAATGAGAGCTGCGGAAAAAATGGCTAAGTCAACAGAC		AACATG		VVGARIVQTNRVLW
			GGAGTAGTCCGAGTTGTCGGGGCCCGCATTGTGCAGACAAATAGAGTT		ATGATTT		KRAGLAGIELDAARK
			TTGTGGAAAAGGGCCGGATTAGCAGGCATAGAACTGGATGCCGCCAGG		GCGAGA		FCEVKRVNKIGNQTN
			AAGTTCTGTGAGGTTAAGAGGGTGAACAAAATTGGCAATCAAACCAAT		GAAACC		GGALKTIAESSVSRH
			GGAGGCGCCCTCAAGACTATAGCAGAGTCCTCGGTGAGCCGGCACTGG		CACGCTT		WLLEKNIRPGNKILV
			TTATTGGAAAAGAATATAAGACCTGGAAACAAAATTCTAGTTTGGAAGG		TGTCACT		WKAMAGVIPTKINLS
			CAATGGCAGGAGTGATTCCAACAAAGATCAATCTGTCTAGAGGCGTAGC		TATGTG		RGVADQTLKKCRCC
			CGACCAGACTCTCAAAAAATGTCGGTGTTGTGGTTTAACAGCAGAAACT		AGGATA		GLTAETDCHILAGCPT
			GATTGTCACATCTTGGCCGGATGTCCTACCAGTCGGGATGCGTACTCGA		AAATCTC		SRDAYSKRHNLLCDK
			AACGTCATAACTTGCTTTGTGATAAACTCGCCAAAGAGCTAAGACTCAAT		TTGTCCA		LAKELRLNGGPSRRV
			GGTGGGCCAAGCAGACGGGTGTGGCGCGAGAGGATGTGTCTCTCTGG		TATGATC		WRERMCLSGNGRRY
			GAATGGCAGGCGTTATAAGCCCGATATTGTTGTGAAAGATGATGGTGT		CTTTGAA		KPDIVVKDDGVITVID
			AATTACTGTCATCGATATGGCATGTCCGTACGAGAAATCGGAAAGACAC		GGGAAC		MACPYEKSERHLSQC
			CTAAGTCAATGCGAAGATGCAAAAGTTGCTAAGTACGAGCCACTAAGG		AGCGCT		EDAKVAKYEPLRLDR
			CTTGATAGGAGTTGGACTCAAGAACTTGAGGGGAATAACGGCAGAAGT		TTGAGC		SWTQELEGNNGRSA
			GCTAATGAAATATCAGTTGTAGGGATTGCAGTAGGGGCGATTGGAACA		TTGCTC		NEISVVGIAVGAIGTI
			ATTACGCGTAAAACCCAGCGGATACTTAGCAAGTTGAAACTGGCCAAGG		GGCGTT		TRKTQRILSKLKLAKV
			TCGGAAGACCGTTACAAATAATTGCATGTAATGAAAGCGCCCAAATTAT		GGCACC		GRPLQIIACNESAQIIR
			AAGACGACATCTTTCGGGATCGAGACTTAGAAATTTGCGGTGAATGCCC		TTTAGTC		RHLSGSRLRNLR
			GAGGTAGTTGGGATAATGATGCACAAGCTCGTAAGGCGACTTGCTGCA		TGTAAT		(SEQ ID NO: 1392)
			CGTATGCCGCTAAACGCTTAGCTCGATGAGTGCATGTCAAGACGGTCGG		ATTTTCT		
			GAGTATGATCAGTGGAGCTGACTTTCCAGACAACTCACGCGGATTCGCG		TGATATT		
			TGCGGTGGATACAACACCTGGTATAACATATGAAGGGTTCCATCTAGTA		ATGGAC		
			CAGGGATAACGATCCATGGGAGCAAACTAATTAGTTGGAGGTAATCCA		GAAAAA		
			ACGCCGCTGTTGAGTCAGTTTTTAACCGCCAGTCAACTCTTGTAGGTTAT		GGTAGT		
			CGGTCTTCGGCAGACCTTGGACCGCCTAGCGCCGGCCAACAGTTTGTCG		ATGGTT		
			TCGACTAACATGATGATTTGCGAGAGAAACCCACGCTTTGTCACTTATGT		GCA		
			GAGGATAAAATCTCTTGTCCATATGATCCTTTGAAGGGAACAGCGCTTT		(SEQ ID		
			GAGCTTGCTCGGCGTTGGCACCTTTAGTCTGTAATATTTTCTTGATATTA		NO:		
			TGGACGAAAAAGGTAGTATGGTTGCA (SEQ ID NO: 1546)		1270)		
R2	R9Av	Adineta vaga	GAAATAGTTTGCAATGGTAGGTGTATGGCGCCTCTGTGTCTCTCTTTCGC	GAAATAGT	ACTAGT		MNLPIREHAVSVHNI
		GQ398057	TGGATATAGTTTGACGATTTTGTACCAGGTATCTGTTTCTTGTGAGTTCA	TTGCAATG	CTCCTTC		NKFNYLCQLCSKSYD
			GCACCAGTTTGAACAGGCTTAGCGATAGACCTTCGAACTTGAAACACTG	GTAGGTGT	TTCTATT		TINSVKAHYVACRRQ
			TTGTGAAGCTGGCTGGGCCCCTGCAGATTTTCTCGATTAGAACGTGAGT	ATGGCGCC	AGTCAG		KNASSTTAVPTNVIN
			GTTACGTCCAGAATGACCCACCAGTGGTTAGTTCTACGTTGCCCTGGAA	TCTGTGTC	TCTAATT		NNQLAINTNQVISRN

		AGGAGAAAAGTTGAGCTAAAATCGCACGGCCTAGTTGTTTATCAAATAG	TCTCTTTC	AATTTTT		PLQCVECLMKQVDF
		GCACGGTGAGGAACTCTTCTATGTACCCTGACTAAAGTACTCACTTGTGC	GCTGGATA	CTTACAT		YAKDTKALVTHMRTK
		GCTGGGTTTGCTCCCCCTCGCATTGACTTATCTGATCGCACTACCCACCA	TAGTTTGA	TCTACAT		HAAAYEESKKVATRR
		AACGAAACATAAACTTAGCTCGTGGTATCAGTCCACACAGCGTGTGCAGTC	CGATTTTG	CTAGTTC		VAWSPDEDQILAELE
		GGATTCAGGGGAGCGTGTTAGTGACAAGCAGGATAATATTAACATAGT	TACCAGGT	CATTATT		VKLKKIQKGQLLSRLV
		TAATGTTAAGGCGTTCAACATTCCTTATCCAATTGGAAGAGTTGACTGTG	ATCTGTTT	AAATTG		VEYNKCADKSKAPSR
		AAGTTTGTCATGAAGACATTGGACAAATGAATTTGCCGATTCGAGAGCA	CTTGTGAG	GTATGA		SKDAIRTRRQQHDYK
		TGCCGTATCTGTACACAATATAAACAAATTTAATTATTTATGCCAGCTAT	TTCAGCAC	TCAGTG		LLLRSLQSQQPPVGS
		GTTCTAAGTCTTATGATACTATTAATAGTGTTAAAGCTCACTATGTTGCA	CAGTTTGA	CTATCTC		EDSDSDISSSNNNPLT
		TGCAGAAGACAGAAGAATGCCTCATCCACAACAGCTGTTCCAACCAATG	ACAGGCTT	TGCTAC		TTHNVTPTPDSSNVV
		TCATCAACAACAACCAACTTGCTATAAATACTAATCAAGTAATATCAAGA	AGCGATAG	ACTCAAT		LLIQKIRESVDSIVKIT
		AATCCACTTCAGTGCGTTGAGTGTCTAATGAAACAAGTTGATTTCTATGC	ACCTTCGA	GCTTAAT		NLKLNTNMLNAASA
		TAAAGATACAAAGGCACTAGTCACGCACATGCGTACTAAACATGCTGCT	ACTTGAAA	CGTATG		FINQNNNMDPLELS
		GCCTACGAGGAATCAAAGAAAGTCGCAACAAGAAGAGTTGCCTGGAGC	CACTGTTG	TTATTGA		MRGIEEDVKAIRDKE
		CCTGATGAGGATCAAATTCTTGCTGAACTAGAAGTCAAATTGAAAAAGA	TGAAGCTG	CAGTCT		LQKPTRNVPSSTTSR
		TACAAAAAGGTCAATTACTTAGTCGTCTTGTCGTTGAATATAATAAATGT	GCTGGGCC	GACACT		KPTRNAKRLEKSKKY
		GCTGATAAATCGAAAGCTCCTTCCAGGTCCAAGGATGCTATTCGTACAA	CCTGCAGA	TGATTAC		GYYQHLYYNNKKKLV
		GACGCCAACAACATGATTACAAACTATTGCTTCGCTCACTCCAATCTCAA	TTTTCTCG	TCTTACG		AEILDGETSGAKPPP
		CAACCGCCAGTTGGTAGCGAAGACAGTGACATATCTTCTAGTA	ATTAGAAC	ACATAT		MNLVEDYYRNIWSR
		ATAACAATCCTTTAACAACACATAATGTCACTCCAACGCCAGATTCA	GTGAGTGT	GCACTG		STIDDSPVNNIKTVNS
		TCCAACGTTGTACTACTAATACAAAAGATCCGTGAATCTGTAGATTCCAT	TACGTCCA	TTTGCTT		DSIFAPISRDEIKLALS
		TGTAAAAATAACGAACCTCAAATTGAACACGAATATGCTGAACGCAGCA	GAATGACC	CAGAGA		NTKKDSAAGPDAVTI
		AGTGCGTTCATTAATCAAAATAACAACATGGATCCACTTGAACTATCTAT	CACCAGTG	AACCAC		KEAKAIIDNLYVAYNI
		GCGTGGTATCGAAGAGGATGTGAAGGCAATTCGAGACAAAGAACTTCA	GTTAGTTC	TGTTCAT		WLGVQGIPEQLKLNK
		GAAACCAACCAGGAACGTTCCTTCTTCAACAACTTCGAGAAAGCCAACT	TACGTTGC	ATAGTG		TILIPKGNSDLSLLKN
		CGAAATGCCAAAAGGCTTGAGAAATCAAAAAAATATGGCTATTATCAAC	CCTGGAAA	AAGTTC		WRPITISSIILRVYNRL
		ATCTGTACTATAATAACAAGAAAAAATTAGTAGCGGAAATCCTCGATGG	GGAGAAA	CTCAGTT		LAYRMNKIFKTNDKQ
		CGAAACAAGTGGTGCTAAGCCACCTCCAATGAACCTGGTTGAAGATTAT	AGTTGAGC	TTCTGTT		VGFKPVNGCGINISW
		TATAGAAATATTTGGTCACGTTCTACTATTGATGATTCGCCTGTTAACAA	TAAAATCG	GATATA		LHSLLKHARLNKNSIY
		TATTAAAACCGTTAATAGTGACTCTATATTTGCTCCAATTTCGCGTGATG	CACGGCCT	TTCTTCT		ACLVDVSKAFDSVSH
		AAATCAAATTAGCATTATCAAATACGAAAAAGGATTCAGCAGCTGGACC	AGTTGTTT	TTCATTC		QSIVRALTMNGAPSL
		TGACGCTGTAACAATAAAAGAAGCAAAAGCTATTATTGACAATCTTTAT	ATCAAATA	TCGCTTC		LVKLIMDQYTNVNTV
		GTTGCATATAATATATGGCTAGGTGTTCAAGGAATTCCTGAACAACTGA	GGCACGGT	TCCTTTT		ITCSGSISNKINISSGV
		AATTGAATAAAACTATCTTAATTCCAAAAGGAAATTCCGATCTTAGTCTA	GAGGAACT	CTACTGT		KQGDPLSSLLFNLVID
		CTGAAAAACTGGCGACCTATTACAATCTCGTCTATTATCCTAAGAGTATA	CTTCTATG	GTTCTTT		ELFDVIKDQYGYTIDN
		CAACAGATTATTAGCATACAGAATGAACAAGATCTTTAAAACTAATGAT	TACCCTGA	TTATCAG		IGTTNARCFADDLTLI
		AAACAAGTTGGATTCAAACCTGTTAATGGTTGTGGTATTAATATATCTTG	CTAAAGTA	TTTTTTG		SSSRMGMNKLLELTT
		GCTTCACTCTCTCTTGAAGCATGCACGCTTAAACAAAAATTCAATATATG	CTCACTTG	TGGAAA		KFFKERGLNVNPSKC

	CTTGTCTTGTCGATGTGTCTAAAGCCTTTGATTCTGTGTCACATCAATCAA	TGCGCTGG	AATTGA	MSIGMSKGYKGKKS
	TAGTAAGAGCTCTCACAATGAATGGTGCACCATCCTTGCTAGTGAAATT	GTTTGCTC	GAATAA	KIESEPLFSITDAQIPM
	AATAATGGATCAATATACGAATGTAAATACTGTCATCACATGTTCTGGTT	CCCCTCGC	ATAAAG	LGYIDKTTRYLGVNFT
	CTATATCAAACAAGATAAATATCTCCAGTGGTGTCAAGCAAGGTGACCC	ATTGACTT	T (SEQ ID	SIGAIDAKRIKKDLQD
	ACTATCTAGCTTGTTGTTCAATCTGGTTATAGATGAACTGTTCGATGTAA	ATCTGATC	NO:	TLDKLEHLKLKAQCK
	TAAAGGACCAATATGGTTATACAATTGATAACATTGGCACCACCAATGC	GCACTACC	1271)	MDLLRTYMIPRFMF
	ACGATGCTTCGCCGATGATTTAACACTAATATCATCATCTAGAATGGGTA	CACCAAAC		QLIHTELYPKLLIKMDI
	TGAATAAATTGCTTGAGCTCACCACGAAATTCTTCAAAGAACGTGGACT	GAAACATA		LIRKLAKRILHLPISTSS
	AAATGTAAACCCATCAAAGTGCATGTCTATTGGCATGTCCAAAGGTTAT	AACTTAGC		EFFYLPFKEGGLQLTS
	AAAGGAAAGAAGAGTAAAATCGAATCTGAACCACTCTTCTCTATCACCG	TCGTGGTA		LKEAVGLAKIKLHKKI
	ATGCTCAGATACCGATGTTGGGCTATATTGATAAGACAACTCGATATCTC	TCAGTCCA		MSSNDPMLCYLIESQ
	GGTGTAAATTTCACATCTATTGGTGCCATTGATGCAAAAAGAATCAAAA	CAGCGTGT		RSRIVEHFMKDLKLG
	AAGACCTTCAGGACACACTCGATAAGCTTGAACATCTTAAACTCAAAGC	GCAGTCGG		DSLTLNEMNNIKECF
	TCAGTGCAAAATGGATCTCTTACGAACTTATATGATACCAAGATTCATGT	ATTCAGGG		MKEKRISFAQKIHGV
	TTCAATTAATTCATACTGAGTTATATCCGAAATTGCTTATTAAAATGGAC	GAGCGTGT		GFEVFSSSPLTNQWI
	ATCTTAATTAGGAAATTAGCTAAACGAATCCTACATCTGCCCATATCAAC	TAGTGACA		NGEIKTMTTKTYINSI
	GAGTAGTGAATTCTTTTACTTACCCTTCAAAGAAGGAGGTCTTCAACTAA	AGCAGGAT		KLRTNTLETRVTTSRG
	CCTCACTTAAAGAAGCAGTTGGTTTAGCCAAAATAAAATTACACAAGAA	AATATTAA		LNIIKTCRRCHVADES
	GATAATGTCCAGTAATGATCCAATGTTATGCTACTTGATTGAGAGCCAG	CATAGTTA		LMHVLQCCSSTKGLR
	AGGAGCCGTATTGTCGAACATTTTATGAAAGACCTTAAACTTGGAGATT	ATGTTAAG		YSRHHKICAKVANKL
	CTTTAACATTAAACGAAATGAATAACATCAAAGAGTGCTTCATGAAAGA	GCGTTCAA		VMNGYGVFREKSYP
	AAAAAGAATCTCATTTGCTCAAAAAATTCACGGTGTCGGCTTCGAAGTA	CATTCCTT		DPNNSGSYLRPDIIAV
	TTCTCATCAAGTCCTTTGACGAACCAATGGATTAATGGCGAAATTAAGA	ATCCAATT		KNGHVIVLDVTVVYE
	CAATGACAACTAAAACATACATTAACTCAATTAAACTTAGAACAAATACT	GGAAGAG		VTGATFINAYQTKINK
	CTAGAAACTCGGGTAACAACATCTCGGGGACTGAACATCATAAAAACAT	TTGACTGT		YNAIMVQIEQMFNC
	GTAGAAGATGCCACGTAGCTGACGAAAGTCTCATGCATGTGCTCCAATG	GAAGTTTG		VNGELHGLVIGSRGSI
	TTGCTCTTCTACCAAAGGTTTACGATACTCTCGTCATCACAAAATATGTG	TCATGAAG		HHSQLHIWHQMGFS
	CCAAAGTAGCAAATAAATTGGTAATGAATGGTTATGGTGTATTTCGTGA	ACATTGGA		SIELKYVAIGCMEDSL
	GAAGAGTTATCCAGATCCAAACAACTCAGGTTCATACCTTCGACCGGAT	CAA (SEQ		RIMSTFSKAIT (SEQ
	ATAATTGCAGTAAAAAATGGTCATGTTATTGTTCTTGATGTAACGGTTGT	ID NO:		ID NO: 1393)
	GTACGAAGTAACTGGTGCTACGTTTATTAATGCCTACCAAACAAAAATA	1148)		
	AATAAATATAATGCGATTATGGTACAAATCGAGCAAATGTTCAATTGTG			
	TTAATGGTGAATTGCATGGTCTAGTAATTGGATCACGTGGTTCAATTCAT			
	CACAGTCAACTCCACATCTGGCATCAAATGGGATTCTCTTCCATAGAACT			
	TAAATATGTGGCTATAGGATGCATGGAGGATTCGCTCAGAATCATGTCC			
	ACATTCTCAAAAGCTATCACATGAACTAGTCTCCTTCTTCTATTAGTCAGT			
	CTAATTAATTTTTCTTACATTCTACATCTAGTTCCATTATTAAATTGGTATG			
	ATCAGTGCTATCTCTGCTACACTCAATGCTTAATCGTATGTTATTGACAG			

			TCTGACACTTGATTACTCTTACGACATATGCACTGTTTGCTTCAGAGAAA				
			CCACTGTTCATATAGTGAAGTTCCTCAGTTTTCTGTTGATATATTCTTCTT				
			TCATTCTCGCTTCTCCTTTCTACTGTGTTCTTTTATCAGTTTTTTGTGGA				
			AAAATTGAGAATAAATAAAGT (SEQ ID NO: 1025)				
R2	R2Ol	LC349444	Oryzias latipes	CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGG	CGCACAGG	GGGGGA	MGTDTVYVGQDYPS
				GAAACGGGGGGGTGACTATCCCCTGGGGTCCGGCGAGAGCGCTGGTCT	GGACACA	CAGCTG	GLSKRVPARLVAGP
				ACGGACCAGGGGGTGGCTGTGGGCAGGCTGCTCCTCAGGCCAGTTGATT	GAGCCTGC	GGAGTC	MLRERSCHAHVFRA
				AGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTAACGACTTGT	CCAAGTAC	TCGGCA	GHMWNWRTSLPSG
				CGGCTAAATCAGCCCGCCCACCATCTGGGATATGGTTGACCGTCTAACC	CGCTCCCG	TGATTAC	RWDQPALEKSRVLTR
				CCAGTACTCAGGTCACAAACAAAATGGGAACAGATACAGTGTATGTCG	AGGGAGC	AAATCTT	SVATATDPEITSYPGK
				GCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACGGTTAGT	GGGAAAC	GCGCTG	SVSTSTQVQEEDWC
				GGCGGGACCGATGCTGCGAGAGCGAAGCTGTCACGCCCATGTGTTTAG	GGGGGGG	CACTCG	SRESGWISPGLAPEE
				GGCTGGACACATGTGGAACTGGCGAACCAGCCTTCCGAGCGGGCGCTG	TGACTATC	GATGTC	PSVVSEITASMVATM
				GGACCAGCCCGCTTTGGAGAAGTCTCGGGTCCTAACCCGGTCGGTGGC	CCCTGGGG	GTCCCC	RVATEEVVLEPQPEQ
				GACGGCCACCGACCCCGAAATTACCTCTTACCCAGGAAAGTCCGTATCG	TCCGGCGA	GTGACG	VVTILPEHGRNVPPG
				ACAAGTACGCAGGTTCAGGAGGAGGACTGGTGTAGCCGGGAGAGCGG	GAGCGCTG	GACACA	LAEQDTASPIEVSVLL
				GTGGATCTCGCCAGGACTTGCTCCTGAAGAACCCTCGGTGGTGTCCGAA	GTCTACGG	TTAATCC	PDLAENCPLCGVPSG
				ATTACAGCCTCCATGGTAGCGACAATGAGGGTAGCAACCGAGGAGGTC	ACCAGGG	GGAAAG	GLRLLGKHFAVRHAG
				GTGCTGGAACCACAGCCTGAACAGGTCGTCACAATACTGCCGGAGCAT	GTGGCTGT	CGAGTG	VPVTYECRKCAWRSP
				GGTCGAAACGTTCCTCCGGGGCTGGCAGAACAGGACACCGCCAGCCCC	GGGCAGG	GTGACT	NSHSISCHVPKCRGR
				ATAGAAGTCTCGGTGCTCCTCCCAGACCTCGCTGAGAACTGCCCATTGT	CTGCTCCT	CGCCTC	ARMPSGDPGIACDLC
				GTGGCGTGCCGAGCGGGGGGCCTACGCTTGCTCGGGAAGCATTTTGCTG	CAGGCCAG	AAG	EARFATEVGVAQHK
				TCCGACATGCGGGGGGTGCCTGTAACGTATGAGTGCCGTAAGTGTGCGT	TTGATTAG	(SEQ ID	RHVHPVEWNKVRLE
				GGCGGAGCCCCAACAGCCACTCAATCTCGTGTCACGTCCCCAAATGCCG	TTACGCAT	NO:	RRGARGGGIKATKL
				GGGGCGTGCGCGGATGCCCAGTGGCGATCCAGGGATCGCCTGCGATCT	GGGCTGTA	1272)	WSVAEVETLIRLIREH
				CTGTGAAGCCCGGTTTGCCACGGAGGTTGGGGTCGCCCAACACAAGCG	CCTCCACG		GDSGATYQLIADELG
				GCACGTTCATCCGGTGGAGTGGAACAAGGTGAGGCTGGAAAGGAGAG	TGGTCCCG		RGKTAEQVRSKKRLL
				GTGCGCGCGGAGGGGGGAATTAAGGCGACGAAGCTCTGGAGTGTAGCG	CTGGTAAC		RIDTASNSPDDAEVE
				GAGGTAGAGACGCTAATCCGGCTCATCCGTGAGCACGGAGATTCAGGT	GACTTGTC		EERLESLAVRSSSRSP
				GCCACTTACCAGCTCATTGCCGATGAGCTGGGAAGGGGCAAGACGGCC	GGCTAAAT		PSLVATRVREAVARG
				GAACAGGTGAGGAGTAAAAAGAGGCTCCTGCGCATAGATACGGCAAGC	CAGCCCGC		ESEGGEEIRAIAALIRD
				AATAGCCCAGATGATGCAGAGGTTGAGGAGGAGAGGTTGGAATCTCTG	CCACCATC		VDQNPCLIETSASDIIS
				GCGGTTCGGTCCTCGTCACGGTCACCCCCGAGCCTGGTGGCGACCAGG	TGGGATAT		KLGRRVDGPKRPRPV
				GTCAGGGAGGCAGTTGCCAGGGGTGAATCAGAAGGTGGCGAGGAGAT	GGTTGACC		VREQTQEKGWVRRL
				CAGGGCTATTGCTGCTCTCATTAGGGACGTAGATCAGAATCCTTGTCTG	GTCTAACC		ARRKREYREAQYLYS
				ATTGAAACCTCGGCGTCGGACATCATCTCGAAGCTGGGAAGGAGGGTG	CCAGTACT		RDQARLAAQILDGAA
				GATGGGCCCAAGAGACCCAGGCCCGTTGTCAGAGAACAGACCCAAGAG	CAGGTCAC		SQECALPVDQVYGAF
				AAGGGATGGGTAAGGCGGCTTGCCCGGCGGAAAAGGGAGTACAGAGA	AAACAAA		REKWETVGQFHGLG

		AGCGCAGTACCTGTACTCAAGGGATCAAGCAAGGCTGGCGGCCCAGAT	(SEQ ID NO: 1149)		EFRTGARADNWEFY
		CCTCGATGGTGCCGCCAGCCAGGAATGCGCCCTCCCGGTGGACCAGGT			SPILAAEVKENLMRM
		CTACGGAGCGTTCCGTGAGAAATGGGAAACCGTAGGGCAGTTCCACGG			ANGTAPGPDRISKKA
		ACTTGGTGAGTTCCGGACGGGTGCACGCGCAGACAACTGGGAGTTCTA			LLDWDPRGEQLARLY
		CTCTCCAATTCTGGCGGCTGAGGTGAAAGAAAACCTAATGAGAATGGCT			TTWLIGGVIPRVFKEC
		AACGGCACGGCCCCGGGACCAGACAGGATAAGCAAAAAGGCTCTGCTT			RTKLLPKSSDPVELQD
		GACTGGGACCCCCGGGGTGAGCAACTGGCACGGCTGTACACGACGTGG			IGGWRPVTIGSMVT
		CTGATCGGTGGGGTCATACCAAGGGTCTTCAAGGAGTGCAGGACTAAG			RLFSRILTMRLTRACP
		CTGCTACCGAAATCCAGCGACCCGGTCGAGTTGCAGGACATCGGTGGA			INPRQRGFLASSSGC
		TGGAGGCCGGTGACGATTGGGTCGATGGTGACTAGGCTGTTCAGTCGG			AENLLIFDEIVRRSRR
		ATTCTAACGATGAGGCTAACCCGAGCCTGTCCGATCAATCCGAGGCAGC			DGGPLAVVFVDFAR
		GCGGTTTCTTGGCCTCCTCGAGTGGATGCGCGGAAAACCTGTTGATCTT			AFDSISHEHILCVLEE
		TGACGAGATCGTCAGGCGCTCGAGGCGGGACGGGGGGCCGCTGGCAG			GGLDRHVIGLIRNSYV
		TGGTGTTTGTGGACTTTGCGAGGGCCTTTGACTCCATCTCACATGAACAT			DCVTRVGCVEGMTP
		ATCCTGTGTGTTCTCGAAGAAGGCGGGGCTTGACAGGCACGTTATCGGGT			PIQMKVGVKQGDP
		TGATCCGAAACTCGTACGTGGATTGCCGTGACCAGGGTGGGTTGTGTCG			MSPLLFNLAMDPLIH
		AGGGCATGACACCACCAATACAAATGAAGGTTGGAGTGAAGCAGGGA			KLETAGTGLKWGDLS
		GACCCCATGTCCCCCTTGCTCTTCAACCTGGCTATGGATCCCCTCATCCAT			IATLAFADDLVLVSDS
		AAACTCGAGACGGCCGGAACTGGACTGAAATGGGGCGATCTTTCAATC			EEGMGRSLGILEKFC
		GCCACGCTGGCCTTTGCCGACGATCTGGTGCTGGTGAGTGACTCTGAGG			QLTGLRVQPRKCHGF
		AAGGCATGGGGAGGAGTCTCGGGATTTTGGAGAAGTTTTGCCAACTGA			FMDKGVVNGCGTW
		CTGGGCTGAGGGTTCAGCCCAGGAAGTGTCACGGTTTCTTTATGGACAA			EICGSPIHMIPPGESV
		GGGCGTGGTGAACGGCTGTGGAACCTGGGAAATCTGTGGGTCACCGAT			RYLGVQVGPGRGVM
		CCACATGATTCCCCCGGGGGAATCAGTTCGTTATTTGGGAGTCCAGGTA			EPDLIPTVHTWIERIS
		GGCCCGGGGCGCGGCGTGATGGAACCGGATCTTATCCCTACGGTCCAC			EAPLKPSQRMRVLNS
		ACGTGGATCGAAAGGATCTCGGAGGCTCCTCTCAAAGCCCTCACAACGCA			FALPRIIYQADLGKVT
		TGAGGGTTTTGAACTCATTCGCTCTCCCCCGGATAATTTACCAGGCCGAT			VTKLAQIDGIVRKAVK
		CTAGGGAAGGTTACGGTAACCAAATTGGCCCAGATAGATGGGATTGTC			KWLHLSPSTCNGLLY
		CGGAAGGCTGTGAAGAAGTGGCTCCATTTGTCACCATCCACGTGCAATG			SRNRDGGLGLLKLER
		GACTGCTGTATTCACGGAACCGCGACGGTGGTTTGGGCCTCCTAAAGCT			LIPSVRTKRIYRMSRS
		GGAAAGACTAATCCCATCCGTGCGCACGAAGCGTATCTATCGGATGTCC			PDIWTRRMTSHSVSK
		AGGTCTCCGGATATCTGGACACGGCGAATGACCAGCCATTCTGTGTCAA			SDWEMLWVQAGGE
		AATCTGACTGGGAGATGTTGTGGGTCCAAGCGGGAGGTGAGAGGGGC			RGSAPVMGAVEAAP
		AGTGCACCTGTAATGGGTGCCGTGGAGGCTGCCCCGACCGATGTGGAG			TDVERSPDYPDWRR
		AGATCGCCAGACTACCCAGACTGGCGGCGTGAGGAAAAACCTGGCATGG			EENLAWSALRVQGV
		TCGGCCCTGCGGGTGCAGGGTGTGGGTGCAGACCAGTTTCGAGGCGAC			GADQFRGDRTSSSW
		AGGACCAGCAGCTCTTGGATCGCCGAGCCCGCTTCGGTTGGGTTCGCGC			IAEPASVGFAQRHWL
		AGCGCCACTGGTTGGCTGCCCTGGCGCTGAGGGCTGGGGTGTATCCGA			AALALRAGVYPTREF

			CTCGGGAGTTTCTGGCTCGGGGTAAGGAAAAGTCAGGAGCAGCTTGCA				LARGKEKSGAACRRC
			GACGCTGCCCGGCCAGGTTGGAATCATGTTCACACATACTTGGGCAATG				PARLESCSHILGQCPF
			TCCGTTCGTTCAGGCGAACAGAATTGCGAGGCACAACAAGGTGTGTGT				VQANRIARHNKVCVL
			GCTCTTGGCCACGGAGGCGGAGAGGTTCGGCTGGACGGTAATAAGGG				LATEAERFGWTVIRE
			AGTTCCGTCTTGAGGACGCCGCTGGCGGTCTCAAGATACCCGACCTGGT				FRLEDAAGGLKIPDLV
			TTGCAAGAAGGCCGACACAGTTCTCATTGTCGACGTGACCGTCCGGTAC				CKKADTVLIVDVTVR
			GAGATGGATGGAGAGACGCTAAAAAGGGCCGCATCGGAGAAGGTGAA				YEMDGETLKRAASEK
			ACACTATCTCCCAGTAGGGCAACAGATAACGGACAAGGTCGGAGGGCG				VKHYLPVGQQITDKV
			TTGCTTTAAAGTCATGGGGTTCCCTGTAGGTGCTAGGGGAAAGTGGCCG				GGRCFKVMGFPVGA
			GCGAGCAACAACACAGTTTTGGCTGAGTTAGGCGTCCCTGCAGGTCGG				RGKWPASNNTVLAE
			ATGAGGACCTTTGCCAGGCTGGTGAGCCGGAGGACTCTTCTTTATTCTTT				LGVPAGRMRTFARL
			GGATATATTGAGGGACTTCATGCGTGAGCCGGCCGGCAGGGGAACTCG				VSRRTLLYSLDILRDF
			GGTTGCTCTCATCCCTGCGGCAACGGGTGCCGCGAATTGAGGGGGACA				MREPAGRGTRVALIP
			GCTGGGAGTCTCGGCATGATTACAAATCTTGCGCTGCACTCGGATGTCG				AATGAAN (SEQ ID
			TCCCCGTGACGGACACATTAATCCGGAAAGCGAGTGGTGACTCGCCTCA				NO: 1394)
			AG (SEQ ID NO: 1026)				
R2	R2_LP	AF015814	TGGGAGGAGACCCAAACTATCCTAGGATGGGGCGGAACCGACCATATG	TGGGAGG	ATTTTGT		GIDGYMFGYARASG
		Limulus polyphemus	AGCCATATTAACATTGCCCACACTATCCTCTGGAGGTACCTCCTCGTGGT	AGACCCAA	CTCTTTC		STSVSIQSSSMTEGET
			ACGGCTGGATATAGGTAAATCCTGTAACCAAATCCTCCAACCCGTGAAG	ACTATCCT	CCCAAT		NERATPRASDSSSVSI
			GAGAACACTAAAACCCATATAGTGGCCTCGCCAACCACTATATGTCCAA	AGGATGG	GATGTC		QSSCVTEGECLPPTD
			CGGCAGGAGAAGCTATCTCCCGGATGGGAAGGAAAACCCTAAACCGTG	GGCGGAA	TACTAG		NCNPSVENQLPCVTE
			ATGGGAACTTACCGGCCCCATCAGCTATTGGGTACCCGGTAGGGACTTG	CCGACCAT	CACGCT		GRFERVGSLVTVRLP
			CAACCCTACCCTGTATTTGCATTTTATAGGGAACCGGTCGGCCCTATATC	ATGAGCCA	GCCGAA		FRKVACDLCSKEFLTY
			AGAGTAGACCGTTTATTAAATATGGGTGAAAATATTAACAGTAAAAGCT	TATTAACA	GCTAGA		SKFAVHQANFHNSET
			ATGGTTTGGCGTCCGTGTGGTGCCAGGGCGGCGGCCAAACCCGAGCTA	TTGCCCAC	TAGATT		QACCTYCGKSDGNH
			CTTGGCACCAACTGGGGATGGTAGCTTCCGAGCGATTCCCTGGCGACGT	ACTATCCT	GAGGAA		HSIACHVPKCPWRRT
			GGGACCGATCGACGATGGAGTCCAAACATCCGGAATAGAGGAATTGAG	CTGGAGGT	TCTGCG		VTFAANLSNFLCDLC
			AAATACCTATTCCACCACCGGCTCACATACCCAAGGTGAACCCGGTGCA	ACCTCCTC	TAATCTG		NDSFKTKSGLSQHKR
			ACTAGAGTACAACCTATCTGTGGCGGTAGGTGCCGAACCACTCAGGTGA	GTGGTACG	TAATGA		HKHPCSRNAERILSLG
			CGGGCTTGTTTATTGATGTCTCCCTACGAGACACGAATTGTGACAAATCC	GCTGGATA	TTACGCC		VRTPSARPRQVVWS
			ACTCCGGTGGACAATTACCCGATCTATGAACCTGTTACCGATATTAGACA	TAGGTAAA	TCATGG		EEETRTLREVEVVYSG
			AGAAAATAAAGAACTGACAACGCCTAGAGCTTCAGGCAGCATGTCTGTA	TCCTGTAA	GCATCT		QKNINVLCAGHLPGK
			AGTATCCAGTCATCGAGCGTGACTGAGGGCGAAATTGATAATAACTCTG	CCAAATCC	ATCGGT		TSKQVSDKRRDLHRI
			AAACTGAGGAATTGACGGATATATGTTTGGCTACGCTAGAGCTTCAGGC	TCCAACCC	AGCGTC		RSSNVHGTPTTQSRG
			AGCACGTCTGTAAGCATCCAGTCATCGAGCATGACTGAGGGCGAAACTA	GTGAAGG	GACCCT		DPVEQVEEYEELDWE
			ACGAAAGGGCCACGCCTAGAGCTTCAGACAGCTCGTCTGTAAGCATCCA	AGAACACT	GACGTT		GMHPFPDPDSKFCSY
			GTCATCGTGCCGTGACTGAGGGTGAATGTCTACCTCCTACAGACAACTGC	AAAACCCA	AAATTG		LDQLRDQKGLTEPV
			AACCCGTCTGTAGAGAACCAGTTACCGTGCGTAACTGAGGGTAGGTTTG	TATAGTGG	GGTAAT		WQEIEIVAQEWVEN

	AACGGGTAGGCTCACTGGTGACGGTGCGTCTGCCCTTCAGAAAGGTGG	CCTCGCCA	AAGAAA	LAHVQSSWNHERTT
	CATGTGACTTGTGTTCTAAAGAGTTCTTGACATATTCGAAGTTTGCAGTC	ACCACTAT	TATCGA	KQVPENNTPARRPFK
	CACCAGGCAAACTTCCACAATTCAGAAACTCAGGCATGCTGCACATATT	ATGTCCAA	(SEQ ID	RRLHRVERYKRFQR
	GCGGTAAAAGTGATGGCAATCATCACTCTATAGCCTGTCACGTTCCGAA	CGGCAGG	NO:	MYDLQRKRLAEEILD
	ATGTCCCTGGCGGCGCGAACTGTTACGTTTGCTGCGAACTTAAGCAATTTCT	AGAAGCTA	1273)	GREAVTCNLKKEEIK
	TGTGTGATCTTTGCAATGATAGTTTTAAGACCAAATCAGGGCTTTCGCAA	TCTCCCGG		DHYDQVYGVSNDRV
	CATAAGCGTCATAAGCATCCTTGTTCAAGGAATGCTGAACGCATCCTTTC	ATGGGAA		SLDDCPRPPGANNT
	TCTTGGAGTCAGGACGCCGTCGGCCCGCCCTCGCCAGGTAGTGTGGTCC	GGAAAACC		DLLKPFTPTEVMDSL
	GAAGAAGAAACACGAACCCTCCGGGAAGTGGAAGTAGTGTATTCGGGC	CTAAACCG		QGMKNGAPGPDKIT
	CAAAAGAACATTAATGTCCTCTGTGCGGGGCATCTACCTGGTAAGACTT	TGATGGGA		LPFLQKRLKNGIHVSL
	CCAAACAGGTCTCGGACAAGCGCCGAGACTTGCACAGGATACGGTCTTC	ACTTACCG		ANVFNLWQFSGRIPE
	TAACGTACATGGTACACCCACCACTCAGAGTCGTGGAGATCCTGTTGAA	GCCCCATC		CMKSNRSVLIPKGKS
	CAGGTCGAGGAGTACGAGGAGTTGGACTGGGAAGGAATGCATCCTTTT	AGCTATTG		NLRDVRNWRPITISSI
	CCCGACCCTGACTCTAAGTTTTGCTCGTACCTTGATCAGCTGAGAGATCA	GGTACCCG		VLRLYTRILARRLERA
	GAAGGGACTCACTGAACCGGTATGGCAGGAGATCGAAATCGTGGCACA	GTAGGGA		VQINPRQRGFVPQA
	AGAATGGGTAGAAAACCTTGCCCATGTTCAATCGTCTTGGAATCATGAG	CTTGCAAC		GCRDNIFLLQSAMRR
	AGAACAACCAAGCAGGTGCCAGAAAACAATACACCTGCACGAAGACCA	CCTACCCT		AKRKGTLALGLLDLSK
	TTTAAAAGGCGTCTCCATCGTGTGGAACGTTATAAGCGGTTTCAGAGAA	GTATTTGC		AFDTVGHKHLLTSLE
	TGTACGACCTCCAGCGAAAGCGCCTGGCTGAGGAAATACTAGACGGCC	ATTTTATA		RFAVHPHFVRIVEDM
	GGGAAGCCGTCACATGTAACCTCAAAAAGGAGGAGATCAAAGACCACT	GGGAACC		YSGCSTSFRVGSQST
	ATGATCAGGTCTACGGTGTGTCAAATGATAGAGTTTCTCTAGATGACTG	GGTCGGCC		RPIVLMRGVKQGDP
	CCCCAGGCCACCAGGGGCCAATAACACCGACCTCCTGAAACCGTTTACG	CTATATCA		MSPILFNIALDPLLRQ
	CCAACCGAAGTGATGGACTCACTTCAGGGTATGAAGAACGGGGCGCCT	GAGTAGAC		LEEESRGFMFREGQA
	GGCCCTGATAAGATTACCCTACCGTTCCTCCAAAAACGTCTTAAAAATGG	CGTTTATT		PVSSLAYADDMALLA
	CATCCATGTTTCCTTGGCAAATGTGTTTAACCTTTGGCAATTCTCGGGTC	AAATATGG		KDHASLQSMLGTVD
	GCATCCCCGAATGCATGAAGTCAAATAGGTCAGTCCTCATCCCGAAAGG	GTGAAAAT		KFCSGNGLGLNIAKS
	GAAGAGCAATCTGCGGGATGTCAGAAACTGGCGGCCAATCACAATCTC	ATTAACAG		AGLLIRGANKTFTVN
	CTCGATTGTGTTGCGGCTATACACCAGGATCTTGGCACGCCGTCTCGAG	TAAAAGCT		DCPSWLVNGETLPM
	CGGGCGGTGCAGATTAATCCCCGACAGCGAGGCTTCGTCCCTCAGGCTG	ATGGTTTG		IGPEQTYRYLGASICP
	GGTGTAGGGATAATATATTCCTGCTTCAGTCTGCTATGAGGAGGGCTAA	GCGTCCGT		WTGINSGPVKPTLEK
	GCGAAAGGGAACTCTGGCTCTGGGGCTTCTTGACTTGTCGAAGGCATTT	GTGGTGCC		WIANITESPLKPHQR
	GACACAGTTGGTCACAAACATCTTCTGACCAGCCTAGAAAGGTTCGCTG	AGGGCGG		VDILCKYALPRLFYQL
	TCCACCCGCATTTCGTCCGAATTGTGGAGGACATGTACAGTGGTTGTTC	CGGCCAAA		ELGTLNFKELKELDS
	GACGTCCTTTCGAGTAGGCAGCCAGTCTACTCGCCCCATCGTTCTGATGA	CCCGAGCT		MVKQAVKRWCHLP
	GAGGCGTCAAACAAGGGGACCCCATGTCTCCTATATTGTTCAACATCGC	ACTTGGCA		ACTADGLLYSRHRDG
	TCTGGACCCTCTTCTTCGTCAACTGGAAGAGGAAAGCCGAGGCTTTATG	CCAACTGG		GLAVVKLESLVPCLKI
	TTTAGGGAGGGGCAGGCCCCTGTCTCATCTCTAGCATATGCCGATGATA	GGATGGTA		KTNLRLVHSTDPVISS
	TGGCACTACTGGCTAAAGATCACGCCAGTCTTCAGTCGATGTTGGGCAC	GCTTCCGA		LAESDGLVGAIEGIAQ

		TGTGGATAAATTTTGTTCAGGGAACGGACTTGGCCTTAACATCGCCAAA	GCGATTCC		KAGLPIPTPDQRSGT
		AGTGCCGGACTTCTGATTAGGGGAGCGAATAAGACCTTCACTGTCAATG	CTGGCGAC		YHSNWRDMERRSW
		ACTGCCCTTCCTGGCTAGTAAATGGTGAAACGCTCCCGATGATCGGTCC	GTGGGACC		ERLALHGQGVELFKG
		CGAACAAACTTACCGTTATCTTGGGGCAAGCATCTGTCCGTGGACTGGG	GATCGACG		SRSANHWLPRPVGM
		ATAAACAGCGGGCCTGTTAAACCCACCCTGGAGAAATGGATAGCCAATA	ATGGAGTC		KPHHWVKCLAMRA
		TCACAGAGTCTCCCCTCAAGCCACATCAGAGGGTCGACATACTCTGTAA	CAAACATC		NVYPTKRGLSRGNLS
		GTACGCTTTACCCCGGCTGTTTTACCAACTTGAGCTGGGCACTCTGAATT	CGGAATAG		KNKDSAKCRGCTSM
		TCAAAGAACTGAAGGAACTAGACAGCATGGTCAAACAAGCTGTCAAAC	AGGAATTG		RETLCHLSGQCPKLKS
		GTTGGTGCCATCTACCTGCCTGTACGGCTGACGGCCTGCTATACTCCCGT	AGAAATAC		MRIRRHNKICEHLIAE
		CATCGTGATGGGGGTTTAGCTGTAGTAAAATTAGAGTCTCTTGTCCCTTG	CTATTCCA		ASFKGWKVLQEPTLV
		TCTAAAGATCAAGACAAATCTCAGACTAGTGCATTCGACCGACCCCGTC	CCACCGGC		TDNGERRRPDLIFHR
		ATATCATCTTTGGCGGAATCCGATGGTTTAGTGGGTGCCATCGAGGGTA	TCACATAC		DDKAVVVDVTVRYEI
		TTGCTCAAAAGGCTGGGCTTCCGATCCCTACGCCTGACCAGCGATCTGG	CCAAGGTG		SKDTLREAYASKVRR
		AACATATCATTCTAATTGGAGAGATATGGAAAGGAGAAGCTGGGAAAG	AACCCGGT		YGCLTEQIKDLTGATS
		GTTGGCCCTGCACGGGCAAGGTGTGGAGCTCTTCAAAGGCTCAAGATCT	GCAACTAG		VVFHGFPMGARGA
		GCCAACCACTGGTTGCCTAGGCCAGTTGGTATGAAGCCACACCACTGGG	AGTACAAC		WFPESSDVMADLNI
		TGAAGTGTCTGGCAATGAGAGCTAATGTATACCCTACAAAAAGAGGCCT	CTATCTGT		RSKYFEEFLCRRTILYT
		CAGTAGAGGGAATCTATCTAAGAACAAAGATTCCGCCAAGTGTCGGGG	GGCGGTA		LDLLWKSNNEQYLER
		ATGCACATCAATGAGGGAGACCCTATGTCATCTAAGTGGTCAATGCCCG	GGTGCCGA		LAP (SEQ ID NO:
		AAATTGAAGTCGATGAGAATAAGGCGCCACAATAAGATCTGTGAGCAC	ACCACTCA		1395)
		TTGATCGCCGAGGCCAGCTTTAAAGGCTGGAAGGTTCTGCAAGAGCCTA	GGTGACG		
		CCTTGGTTACAGACAATGGTGAACGTCGGCGACCTGATCTGATCTTCCA	GGCTTGTT		
		TCGTGATGATAAAGCGGTGGTTGTTGACGTGACGGTTCGCTACGAAATT	TATTGATG		
		TCGAAAGACACGTTGAGAGAAGCTTATGCTTCTAAAGTTCGAAGGTATG	TCTCCCTA		
		GATGTTTGACCGAACAAATTAAAGACCTTACAGGGGCTACCTCCGTTGT	CGAGACAC		
		TTTTCATGGATTTCCAATGGGTGCCCGCGGTGCCTGGTTTCCTGAAAGCT	GAATTGTG		
		CGGACGTGATGGCCGACCTGAACATTCGGTCAAAATATTTTGAAGAGTT	ACAAATCC		
		CTTGTGTAGACGCACCATCCTATATACACTGGACTTATTATGGAAATCGA	ACTCCGGT		
		ATAACGAACAATATTTAGAAAGGCTTGCACCATAAATTTTGTCTCTTTCC	GGACAATT		
		CCAATGATGTCTACTAGCACGCTGCCGAAGCTAGATAGATTGAGGAATC	ACCCGATC		
		TGCGTAATCTGTAATGATTACGCCTCATGGGCATCTATCGGTAGCGTCG	TATGAACC		
		ACCCTGACGTTAAATTGGGTAATAAGAAATATCGA (SEQ ID NO: 1027)	TGTTACCG		
			ATATTAGA		
			CAAGAAAA		
			TAAAGAAC		
			TGACAACG		
			CCTAGAGC		
			TTCAGGCA		

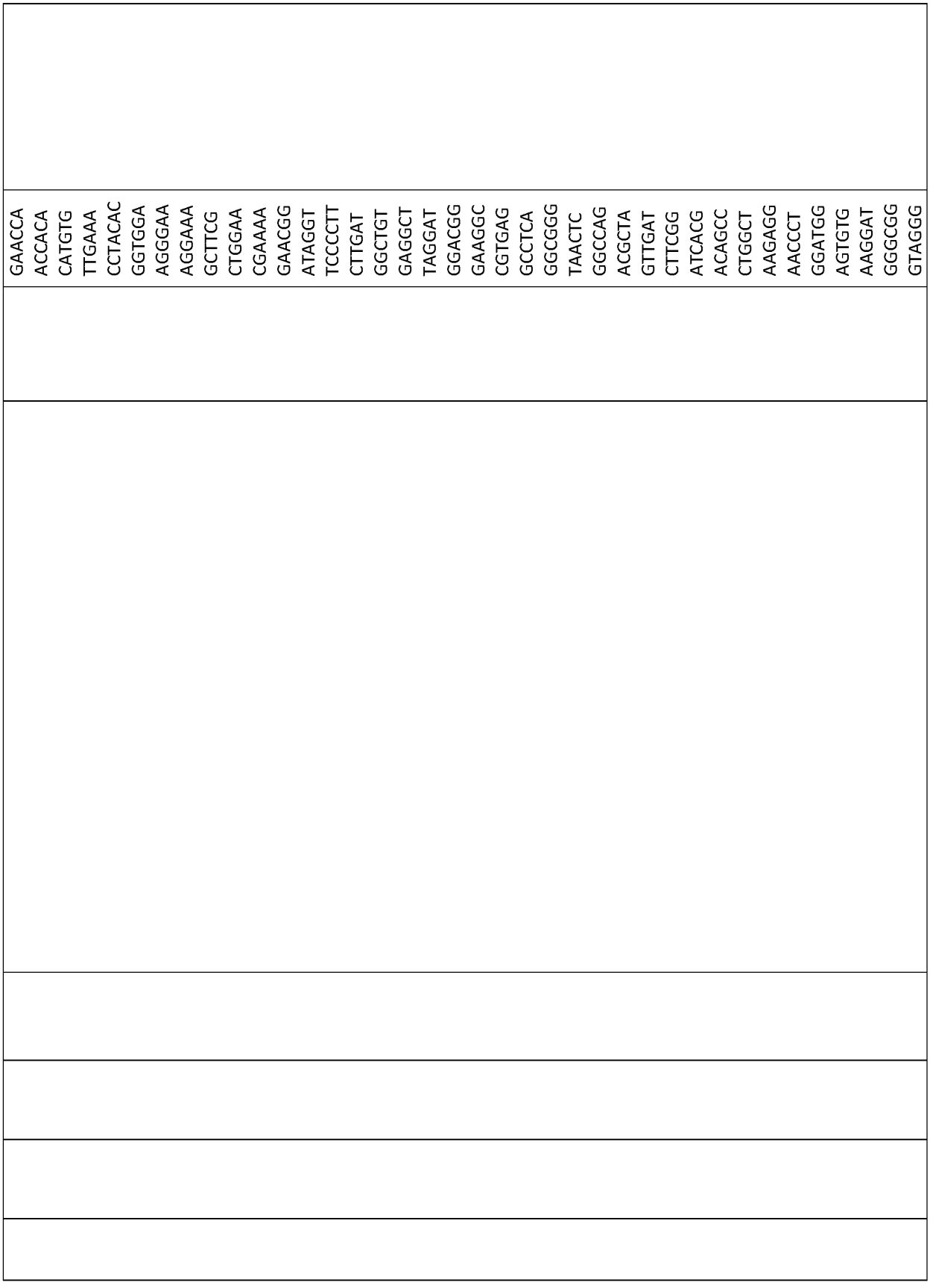
				GCATGTCT GTAAGTAT CCAGTCAT CGAGCGTG ACTGAGG GCGAAATT GATAATAA CTCTGAAA CTGA (SEQ ID NO: 1150)			
NeSL	NeSL-1	Z82058	Caenorhabditis elegans	GCTCACTTTCTATCGTGTTAACCGTACGTTTACACTCCCAGTGAGTGTAA	GCTCACTT	CCTCCA	MLRRKGRHRMVMV
				TAAAGGTTATTCGATAGAGGGTGTCTCCCTCTTTCTTGGGTAATTCTTCG	TCTATCGT	GGGCAC	NSVKWQPSAHAEAI
				GCGGTCCGGGGTCTCTCCCTCGTCTTTTTTTTAAACTTTTCTTTCTCATCC	GTTAACCG	GCCGCA	GTGKSWAPQRSQAS
				ACTCTTTTGCTCCTTTTTACTAACTCTTGTACTCTATAGTCTTTTCTCATCC	TACGTTTA	CGCCAA	EHGWQSNAMFDPP
				CCCATCCGCCGTTGGGCAAAGTTTATTTACTTTGTTAAATCCATATTTTAT	CACTCCCA	AAGTCC	NRILFARDSWSLNQS
				CTCTCTCACCCGTACAGAAAGCGTCTCCTTCTCAAACGCTTTTCTGTACTT	GTGAGTGT	TGGCAT	THLQNQRSGSGLGIR
				TTTCTTATATTTTCATTAACATATTTTTCCTGTTTATACTAACCTAACCTCC	AATAAAGG	AACTCT	PGQVRNNMVGGGP
				ATTGTCAATTACTAACTAACTTGTACAACGGATTTCGATGTTGCGCCGAA	TTATTCGA	GCAAAT	HRAGDPKRRVELVSI
				AAGGACGTCACCGAATGGTTATGGTCAATTCTGTCAAATGGCAACCCAG	TAGAGGGT	AACATC	QGSEVTVRTIYPSDEI
				TGCACATGCTGAAGCAATTGGAACAGGAAAGTCCTGGGCACCACAGCG	GTCTCCCT	AAACGT	FSCYSKSCDIKTKAGY
				GTCCCAGGCATCCGAACACGGCTGGCAATCAAATGCAATGTTTGATCCC	CTTTCTTG	CAATCA	GPEDLKHLTRHIKNE
				CCCAACAGGATTCTCTTCGCCAGAGACTCATGGTCGCTCAACCAATCAAC	GGTAATTC	ACTCCAC	HGLKARWAYQCGLC
				GCATCTTCAAAATCAAAGGAGCGGATCAGGATTGGGTATAAGACCTGG	TTCGGCGG	AAACTCT	NEKSDPSVSEGHKW
				TCAGGTAAGGAACAATATGGTGGGGGGTGGGCCTCACAGAGCAGGGG	TCCGGGGT	CCACTCT	MEAHMVAVHQSSA
				ACCCAAAGCGTCGTGTCGAGCTGGTCAGCATACAAGGAAGCGAAGTGA	CTCTCCCT	CTTCAA	EKRIKSYQKCTGARV
				CCGTCAGAACAATCTACCCGTCGGATGAAATATTCAGTTGTTACTCCAAA	CGTCTTTTT	GTCTTCT	AEQLQAAAPSLTVPG
				TCATGTGATATCAAAACAAAAGCTGGCTATGGCCCTGAGGACCTAAAGC	TTTAAACT	CGGTGC	KHKSGSRDAAKDSM
				ACCTGACTCGTCATATCAAGAACGAGCATGGTCTCAAAGCTCGCTGGGC	TTTCTTTCT	TTCCAAC	TPTKDDDPKTRIYQT
				ATATCAATGTGGATTGTGCAATGAGAAGTCGGACCCAAGTGTATCGGA	CATCCACT	ACCACA	RSVVKKSTQKTAEPT
				AGGCCACAAATGGATGGAGGCACACATGGTCGCCGTTCACCAAAGCTC	CTTTTGCT	ATGGTG	DEGSRGPKYASIFQKS
				TGCGGAAAAAAGGATAAAGTCCTATCAGAAATGCACGGGTGCAAGAGT	CCTTTTTAC	AAAGCT	VKARKSLALLCELSSP
				TGCAGAACAGCTACAAGCTGCTGCTCCATCGCTTACTGTGCCGGGGAAG	TAACTCTT	CCTTCAC	KPMNPLPTNELTLKE
				CACAAATCAGGCTCTAGAGACGCTGCCAAAGATTCGATGACACCAACAA	GTACTCTA	CTTTTCC	GNSRELAKEEAPSEGI
				AGGATGATGACCCGAAAACCAGGATCTATCAGACACGAAGCGTAGTTA	TAGTCTTT	CTCCAA	DDIVIIDLDESEESPPR
				AAAAGTCGACTCAGAAAACAGCAGAGCCAACAGATGAAGGGTCTAGAG	TCTCATCC	AATTCTT	RKRFNTWCLDHESSR
				GCCCAAAGTACGCATCCATTTTTCAGAAATCCGTCAAAGCAAGGAAGAG	CCCATCCG	CCCATGT	EAWLDDTAIFWYISY
				CTTGGCGCTTCTCTGTGAATTAAGCAGCCCTAAGCCTATGAACCCCCTTC	CCGTTGGG	GGGGAA	LCRGSTKYSALDPCL

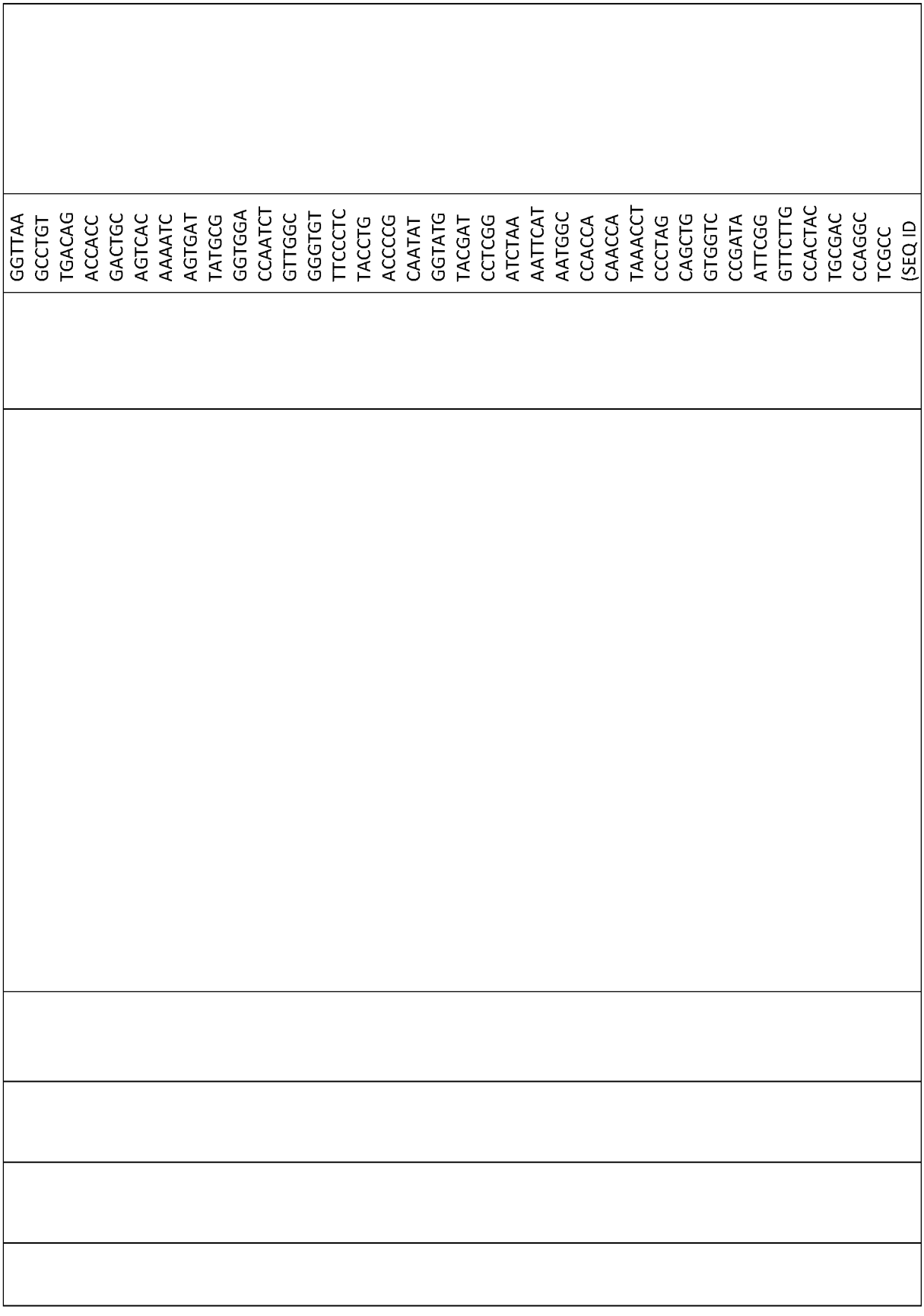
		CTACAAATGAGCTAACTCTGAAAGAAGGGAATTCAAGAGAGCTCGCCA	CAAAGTTT	GTCCTG		WSMYKVKGSRYILDR
		AAGAGGAAGCACCATCTGAAGGTATAGACGACATCGTCATCATCGATCT	ATTTACTTT	TTCTTGT		LESSITYFFPICEEDH
		GGACGAATCGGAGGAGTCGCCACCCAGAAGGAAACGATTCAACACCTG	GTTAAATC	AAGCTC		WTLLVLKDNSYYYAN
		GTGTCTGGATCATGAGTCAAGCCGTGAAGCATGGCTGGATGACACAGC	CATATTTT	TCCGGA		SLHQEPRGPVRDFIN
		AATCTTCTGGTACATCTCCTATCTCTGCAGAGGAAGTACAAAGTACTCAG	ATCTCTCT	GGCTGC		DSKRARKEFKVQVPL
		CTTTGGACCCATGCCTCTGGAGTATGTACAAAGTCAAAGGCTCAAGATA	CACCCGTA	AAGAGC		QRDSFNCGVHICLM
		CATTCTTGACCGCTTGGAAAGCTCCATCACATATTTTTTCCCGATATGCG	CAGAAAGC	AGAAGA		TNSIMAGGKWHSEE
		AGGAGGACCATTGGACACTGTTGGTATTGAAAGACAATTCATACTATTA	GTCTCCTT	AATTCTT		DVRNFRKRLKKTLQE
		TGCAAACAGTCTGCACCAAGAGCCACGTGGCCCGGTCAGGGACTTCATC	CTCAAACG	CTTTCTG		EGYELYSVNSLGIPFQ
		AACGACTCAAAACGGGCTCGGAAGGAGTTTAAGGTGCAAGTACCTCTTC	CTTTTCTGT	ACAAGG		APTTEQMDYKETRCK
		AAAGAGACTCCTTTAACTGTGGAGTGCACATCTGTCTAATGACCAACTC	ACTTTTTCT	TCAGAA		RSYASVLTQISPPAKR
		GATTATGGCAGGAGGCAAATGGCACTCTGAAGAAGACGTCAGAAACTT	TATATTTTC	GGAAGT		PDCKPDNNIFVPTKD
		CAGAAAAAGACTGAAGAAGACACTCCAGGAAGAAGGCTATGAGCTTTA	ATTAACAT	CCTGTTC		CAAEGNPQEKGRNE
		CTCGGTCAATAGTCTGGGTATACCATTCCAAGCCCCAACGACTGAGCAA	ATTTTTCCT	TTGAGG		SPEEINTEHIVVAGKP
		ATGGACTACAAAGAAACAAGATGCAAAAGAAGCTATGCCAGTGTTCTTA	GTTTATAC	CGTCCAT		ANNISPRCRSTSEML
		CTCAAATAAGCCCGCCGGCCAAAAGGCCGGACTGCAAACCTGACACAA	TAACCTAA	CCCGGG		FEMVKATTSSGRSSL
		ACATATTCGTACCAACCAAGGATTGTGCTGCCGAAGGTAACCCGCAGGA	CCTCCATT	CGTCAT		GTMTQDEFIRTSTIAE
		AAAAGGCCGAAATGAATCTCCTGAAGAGATCAATACGGAACATATCGTC	GTCAATTA	AGGAGA		AVPLMSIKLPPMELP
		GTCGCAGGAAAACCTGCAAACAACATCAGTCCAAGGTGTCGGAGCACC	CTAACTAA	GATCAG		RKILPPIPPRKPTQTN
		TCGGAAATGCTGTTTGAGATGGTGAAAGCCACAACCAGCAGTGGAAGA	CTTGTACA	ATGCAC		GGQKGKQQRVPTGK
		AGCAGCTTGGGCACCATGACGCAGGATGAGTTCATCCGAACCAGCACA	ACGGATTT	CTTCTAG		PDTLNAKVRNWFNN
		ATCGCCGAGGCAGTTCCCCTAATGAGCATAAAACTCCCACCAATGGAGT	CG (SEQ ID	CAGGAG		QLESYAMEGRSFQRL
		TGCCAAGGAAAATTCTGCCACCAATTCCCCCCAGAAAACCAACCCAAAC	NO: 1151)	CTAGAA		EWLTEVLTASIQKAA
		CAATGGAGGTCAAAAGGGAAAGCAACAGAGGGTGCCTACAGGAAAAC		GGGCTG		AGDEGIVDIICKRNPP
		CAGACACCCTAAATGCTAAAGTCCGGAACTGGTTCAACAACCAACTTGA		CCCTGTC		LEVAKGEMCTQTEN
		GTCGTATGCGATGGAGGGTCGCAGCTTCCAACGACTGGAATGGCTGAC		TTGAGA		KRKTTNNAARIADPI
		GGAAGTACTCACTGCGTCGATACAAAAAGCAGCAGGTGATGAAGG		TCCCCAC		QSSKGAGDVKASYW
		AATAGTTGATATTATTTGCAAACGGAACCCGCCACTTGAAGTTGCGAAG		GGGGGT		KERARTYNRIIGSKEE
		GGTGAAATGTGCACCCAGACCGAAAACAAAAGGAAAACGACCAACAAT		CAATAG		LCKIPIDQLEDFFKKST
		GCAGCAAGAATTGCGGACCCAATCCAGAGCAGCAAGGGAGCTGGTGAT		ACGGGA		SRTNVQESIMKEKSS
		GTGAAGGCATCGTACTGGAAAGAAAGGGCTCGCACTTACAACAGGATT		GGGGCT		KIPALKIGNWMEKKF
		ATTGGTAGCAAGGAGGAACTCTGCAAAATTCCCATCGATCAACTGGAGG		GCTGGC		IGKEVAFALRKTKDTA
		ATTTCTTCAAGAAATCCACGTCCCGCACCAACGTGCAGGAGTCGATCAT		TTTCTCT		QGADGLRYHHLQWF
		GAAGGAGAAAAGCTCCAAAATTCCTGCTCTCAAGATAGGTAACTGGATG		TTTTAAG		DPSGELLAKVYNECQ
		GAGAAGAAGTTTATCGGAAAGGAGGTGGCGTTCGCTCTGCGGAAAACA		AGGAAG		RHRKIPKHWKEAETIL
		AAAGACACCGCGCAGGGTGCAGACGGACTGCGATACCACCACCTTCAA		CACCAA		LFKNGDQSKPENWR
		TGGTTTGATCCCAGTGGTGAGTTATTGGCGAAGGTATATAACGAGTGCC		TCCGGA		PISLMPVIYKLYSSLW
		AACGACACAGGAAGATCCCAAAACACTGGAAGGAGGCCGAGACCATCT		GATCCTT		NRRIRAVPNVLSKCQ

		TGCTGTTCAAAAATGGAGATCAGTCAAAACCAGAAAACTGGCGCCCAAT		AGGGGT		RGFQEREGCNESLAIL
		TAGCCTGATGCCTGTGATCTACAAACTTTACTCCAGTCTGTGGAACCGGA		CAAAGG		RTAIDVAKGKRRNLA
		GAATTAGAGCTGTACCAAATGTGTTGAGCAAATGTCAGCGAGGGTTCCA		ATTAAA		VAWLDLTNAFGSIPH
		GGAGCGCGAAGGTTGCAATGAGAGTCTAGCAATACTCAGAACAGCAAT		AGGCAG		ELIEYALTAYGFPQM
		CGACGTGGCCAAAGGAAAACGAAGAAACCTGGCGGTGGCATGGCTGG		CAGGTC		VVDVVKDMYQGAS
		ATCTGACGAACGCGTTTGGATCCATCCCGCACGAATTGATTGAGTACGC		CAATTCT		MRVKNATEKSDRIPI
		GCTGACAGCGTATGGATTTCCGCAAATGGTCGTCGATGTGGTCAAAGAT		CCTCACT		MSGVKQGDPISPTLF
		ATGTACCAGGGAGCATCAATGAGGGTTAAGAACGCGACGGAAAAAAGC		GACTTC		NICLETVIRRHLESAN
		GATCGAATCCCAATAATGTCTGGGGTGAAACAAGGCGATCCCATTTCAC		GGTCAG		GHQCLKTRIKVLAFA
		CAACACTTTTCAATATATGCCTGGAAACTGTGATTAGAAGACACCTGGA		AGAGGA		DDMAILTDSPDQLQ
		GTCTGCAAATGGTCACCAGTGCCTCAAAACAAGAATTAAGGTACTGGCG		GTCCCG		RELSKLDNDCTPLNLI
		TTCGCCGACGACATGGCGATTTTAACGGATTCCCCCGACCAGCTCCAGC		CCTTGG		FKPAKCASLVIQKGV
		GAGAACTGTCAAAGCTAGACAATGATTGCACGCCCCTGAATCTTATTTTC		AGACCT		VRSASIKLKGNAIRCL
		AAGCCAGCAAAATGTGCATCACTTGTGATCCAAAAAGGAGTTGTGCGG		CCCCGG		DENTTYKYLGVQTGS
		AGCGCATCAATTAAGCTTAAAGGAAACGCCATTCGATGCCTTGACGAGA		GGAGGT		AARISAMDLLEKVTK
		ACACCACTTACAAATATTTGGGAGTTCAGACGGGTTCGGCAGCAAGAAT		TGCTGA		ELECVVKSDLTPPQKL
		TTCAGCAATGGATCTACTGGAGAAAGTCACGAAGGAACTTGAATGCGT		AGAGGC		DCLKTFTLSKLTYMYG
		GGTCAAAAGTGACCTGACGCCCGCCGCAAAAGCTGGACTGTCTTAAAACA		GGAAGC		NSIPLITEIKMFANIVI
		TTCACGCTGTCCAAACTGACATACATGTATGGAAATTCCATACCACTGAT		TCCTTCT		RGVKVMHRIPVRGS
		CACGGAGATAAAAATGTTTGCAAATATCGTCATTCGAGGAGTCAAAGTG		AGCAAG		PLEYIHLPVKDGGLG
		ATGCATAGAATCCCAGTCCGAGGGTCACCACTGGAGTACATCCATCTTC		AGCTAG		VACPKTTCMITFLVST
		CAGTGAAGGATGGAGGGCTTGGTGTAGCATGTCCCAAGACAACCTGCA		AGGGAG		LKKLWSDDEYIKTLFT
		TGATTACGTTCCTTGTCTCTACTCTTAAAAAACTCTGGTCAGATGATGAA		TTCCCAG		SLAEEVVKKESKKSTV
		TACATCAAAACATTATTCACATCACTGGCGGAAGAAGTAGTAAAGAAAG		TCCTGA		TMDDIADYLNVEERI
		AGTCAAAGAAGAGCACAGTCACTATGGATGATATAGCCGACTATCTCAA		AACCCTT		NRSEFGYNSITRLRDV
		CGTTGAGGAGAGGATCAATAGGAGCGAATTTGGGTACAATTCCATTAC		GCGGTT		MRNLAITGDSPLYRL
		GAGACTGCGGGATGTGATGAGGAACTTGGCCATCACTGGCGACTCCCC		GATGAT		KMVVKNGKIALLVQ
		ACTTTACAGGCTGAAAATGGTAGTAAGAACGGGAAAATCGCTTTGCTC		GGAATG		ATSESMERIYTEEDAK
		GTCCAAGCCACAAGCGAAAGCATGGAAAGGATCTACACGGAAGAAGAT		GAAGAG		KLQRSLKDQVNKALK
		GCGAAAAAGCTGCAGCGCTCACTGAAGGATCAAGTGAACAAAGCACTC		TACTTCG		HRFNTTKVVKSKVVR
		AAACATCGATTCAACACCACCAAAGTAGTGAAAAGCAAAGTCGTCCGAG		GTACTG		VVQQHPASNRFVTK
		TCGTGCAACAGCACCCAGCAAGCAACAGGTTTGTCACAAAAGGTGGCA		CTCGTT		GGNLSLACHRFVHKA
		ACCTGAGCCTTGCATGTCACCGCTTTGTGCATAAAGCACGTCTGAATCTA		GCTCTCT		RLNLLACNYNNYDKS
		CTGGCCTGCAACTACAACAACTACGACAAATCCAAATCAAAAGTCTGTA		CTGCGT		KSKVCRRCGKDLETQ
		GGCGTTGTGGGAAGGATCTGGAGACGCAGTGGCACATACTGCAAAACT		TTTACTG		WHILQNCPFGFSKKI
		GTCCGTTTGGTTCTCAAAGAAGATCACTGAGAGGCATGATGCCGTCTT		CCGAGG		TERHDAVLHKVKTLIE
		GCACAAGGTCAAAACTCTCATTGAAAGCGGTGGAAAAAAGAATTGGAC		GCCGGA		SGGKKNWTMKIDEE
		AATGAAGATTGATGAAGAACTTCCAGGATTCAGCAGACTCCGTCCAGAT		TTTGCTC		LPGFSRLRPDICLKSP

		ATCTGCCTCAAAAGCCCTGATGAAAAACAAATCATCTTGGCAGATGTCG	GAATCG	DEKQIILADVACPYEH	
		CATGCCCATATGAGCATGGAGTAGAAGCGATGGAAAGGAGCTGGCAG	CGAAAG	GVEAMERSWQAKID	
		GCAAAAATCGACAAATACGAGACGGGATTCGCCCACCTGCGGAAATCG	GTCTCA	KYETGFAHLRKSGTKL	
		GGAACCAAGCTGACCGTCCTTCCGATTATAATCGGGTCACTTGGATCAT	ATCGAC	TVLPIIIGSLGSWWKP	
		GGTGGAAACCGACAGGTGACAGTCTCAAGGAATTGGGAATCAAGGGA	CATTCAA	TGDSLKELGIKGSVIN	
		AGCGTGATCAACAGTGCCATTCCAGAACTCTGTGCTACTGTTCTCGAACA	GATGAC	SAIPELCATVLEHSKN	
		CAGTAAGAATACGTACTGGAATCACATCTTCGGTGAAGCGTACATACCA	GGCTTA	TYWNHIFGEAYIPNP	
		AATCCAATGCGAAACGGACACGCAAAACCTGCTGGAAATGGATGGAAA	TCTAAG	MRNGHAKPAGNGW	
		AAGGAAAGATTGCAGAAGGCCCCTGTGAGGCCTACCAACTAGCCTCCA	GTCCGA	KKERLQKAPVRPTN	
		GGGCACGCCGCACGCCAAAAGTCCTGGCATAACTCTGCAAATAACATCA	AAGCAG	(SEQ ID NO: 1396)	
		AACGTCAATCAACTCCACAAACTCTCCACTCTCTTCAAGTCTTCTCGGTGC	TTGGGA		
		TTCCAACACCACAATGGTGAAAGCTCCTTCACCTTTTCCCTCCAAAATTCT	GAGTAA		
		TCCCATGTGGGGAAGTCCTGTTCTTGTAAGCTCTCCGGAGGCTGCAAGA	CGTGTT		
		GCAGAAGAAATTCTTCTTTCTGACAAGGTCAGAAGGAAGTCCTGTTCTT	CTCCTAC		
		GAGGCGTCCATCCCGGGCGTCATAGGAGAGATCAGATGCACCTTCTAG	CTTTCAA		
		CAGGAGCTAGAAGGGCTGCCCTGTCTTGAGATCCCCACGGGGGTCAAT	GTTGAA		
		AGACGGGAGGGGCTGCTGGCTTTCTCTTTTTAAGAGGAAGCACCAATCC	TGGTCG		
		GGAGATCCTTAGGGGTCAAAGGATTAAAAGGCAGCAGGTCCAATTCTC	TTTTACT		
		CTCACTGACTTCGGTCAGAGAGGAGTCCCGCCTTGGAGACCTCCCCGGG	GTTTGG		
		GAGGTTGCTGAAGAGGCGGAAGCTCCTTCTAGCAAGAGCTAGAGGGA	GATAGC		
		GTTCCCAGTCCTGAAACCCTTGCGGTTGATGATGGAATGGAAGAGTACT	TGACTT		
		TCGGTACTGCTCGTTGCTCTCTCTGCGTTTTACTGCCGAGGGCCGGATTT	GATGCT		
		GCTCGAATCGCGAAAGGTCTCAATCGACCATTCAAGATGACGGCTTATC	AGTACG		
		TAAGGTCCGAAAGCAGTTGGGAGAGTAACGTGTTCTCCTACCTTTCAAG	CTTCATC		
		TTGAATGGTCGTTTTACTGTTTGGGATAGCTGACTTGATGCTAGTACGCT	TGTGGA		
		TCATCTGTGGATGACGCTCCCCAAGCAGTCAAGTAGACTTGAAAGGTGC	TGACGC		
		CCTCGCCCTAGTTAGCTCTTAGACCTTATGGGTCGCCATGGTTGTGGACG	TCCCCAA		
		GGTATGCTTGCCGGAGCCGAGTCGTGTTTCTTAGAACCAACCTCGACGA	GCAGTC		
		GGCGAAAGCTTGCACAAGTTAGCACAATTGTGGTAGGGCCGACTAGAA	AAGTAG		
		AATGAGTCCCTTAGGGGGTTACGCCTTGGCGAAAGTGAGGACAATTGG	ACTTGA		
		CATTGACGGGTGCTTCGGCACTAGGCAAAGGCGCCACCACACACTGTCCAA	AAGGTG		
		TCTCTAAAAAGTTCACATTCATCGAAGAACTACCGGAACCAACCACACACAT	CCCTCGC		
		GTGTTGAAACCTACACGGTGGAAGGGAAAGGAAAGCTTCGCTGGAACG	CCTAGTT		
		AAAAGAACGGATAGGTTCCCCTTCTTGATGGCTGTGAGGCTTAGGATGG	AGCTCTT		
		ACGGGAAGGCCGTGAGGCCTCAGGCGGGTAACTCGGCCAGACGCTAGT	AGACCT		
		TGATCTTCGGATCACGACACGCCCTGGCTAAGAGGAACCCTGGATGGAG	TATGGG		
		TGTGAAGGATGGGCGGGTAGGGGGTTAAGCCTGTTGACAGACCACCGA	TCGCCAT		
		CTGCAGTCACACAAATCAGTGATTATGCGGGTGGACCAATCTGTTGGCGG	GGTTGT		

	GTGTTTCCCTCTACCTGACCCCGCAATATGGTATGTACGATCCTCGGATC TAAAATTCATAATGGCCCACCACAACCATAAACCTCCCTAGCAGCTGGTG GTCCCGATAATTCGGGTTCTTGCCACTACTGCGACCCAGGCTCGCC (SEQ ID NO: 1028)				GGACGG GTATGC TTGCCG GAGCCG AGTCGT GTTTCTT AGAACC AACCTC GACGAG GCGAAA GCTTGC ACAAGT TAGCAC AATTGT GGTAGG GCCGAC TAGAAA ATGAGT CCCTTAG GGGGTT ACGCCTT GGCGAA AGTGAG GACAAT TGGCAT TGACGG GTGCTT CGGCAC TAGGCA AAGGCG CCACCA CACTGTC CAATCTC TAAAAA GTTCAC ATTCATC GAAGAA CTACCG


GAACCA
ACCACA
CATGTG
TTGAAA
CCTACAC
GGTGGA
AGGGAA
AGGAAA
GCTTCG
CTGGAA
CGAAAA
GAACGG
ATAGGT
TCCCCTT
CTTGAT
GGCTGT
GAGGCT
TAGGAT
GGACGG
GAAGGC
CGTGAG
GCCTCA
GGCGGG
TAACTC
GGCCAG
ACGCTA
GTTGAT
CTTCGG
ATCACG
ACAGCC
CTGGCT
AAGAGG
AACCCT
GGATGG
AGTGTG
AAGGAT
GGGCGG
GTAGGG


GGTTAA
GCCTGT
TGACAG
ACCACC
GACTGC
AGTCAC
AAAATC
AGTGAT
TATGCG
GGTGGA
CCAATCT
GTTGGC
GGGTGT
TTCCCTC
TACCTG
ACCCCG
CAATAT
GGTATG
TACGAT
CCTCGG
ATCTAA
AATTCAT
AATGGC
CCACCA
CAACCA
TAAACCT
CCCTAG
CAGCTG
GTGGTC
CCGATA
ATTCGG
GTTCTTG
CCACTAC
TGCGAC
CCAGGC
TCGCC
(SEQ ID

CRE	Cnl1	.	Cryptococcus neoformans	CCCTCTTAATACCCCATAACACATAACAACCCCCTAATCAACGTTCTCTGC	CCCTCTTA	TGAGGA	MSLQRAKNARGDPG
				ACCTTAAACACCACCACCAACATGTCCCTGCAGAGGGCCAAAAACGCCCGTG	ATACCCCA	AGAGGA	RCNLCSADYRDLKDH
				GAGATCCTGGTCGGTGCAACCTATGCTCTGCCGACTATAGGGACCTCAA	TAACACAT	GGTTGG	LNKQHSTHFFVPSDL
				AGATCATCTCAATAAACAACATTCCACCCATTTCTTCGTCCCCTCCGACCT	AACAACCC	ATTATTT	RGSSLVACPRCGTPC
				CCGTGGCTCTTCCCTAGTCGCTTGCCCTCGCTGCGGCACCCCCTGCTCAG	CCTAATCA	TTTCTTT	SAGTGLSRHQSRYCG
				CTGGCACTGGTTTATCTCGTCACCAGAGCCGGTATTGCGGTCTCACCGCT	ACGTTCTC	TCTTTAA	LTAPRIRRNRVGNST
				CCTCGAATCCGCCGAAATCGCGTGGGAAACTCAACAAACACATCTCGCT	TGCACCTT	TAAGTT	NTSRCPPSNTAASPIV
				GCCCTCCCTCCAATACTGCAGCTTCACCCATCGTTCCTTCGCCTTCCCCAG	AAACACCA	GTTTATT	PSPSPERPSPPQPAE
				AACGCCCAAGCCCCCCTCAGCCTGCTGAAGTTGTTGCCAGTCTCGAACC	CCAAC (SEQ ID NO: 1152)	TAAGTA	VVASLEPLSEAEEVLE
				ATTGTCTGAAGCCGAGGAGGTGCTGGAGGTCGCCCAGGTTGATGCCGA		GTTTCTT	VAQVDAETVDTLEGT
				GACTGTTGACACGCTGGAAGGGACCCGGAGAGCTCCGGAATCCGTTCC		TCATTCG	RRAPESVPRSAEEGS
				GAGATCTGCCGAGGAAGGTAGCACGCGAGTTAGGGAGCTAAACATGAC		GGCAAC	TRVRELNMTAPEEEH
				AGCGCCGGAGGAGGAGCATCGTGGGGAGGAGAGTAGTCATACCA		CCACAC	RGEEESSHTNPTAPA
				ACCCAACTGCCCCAGCAGGGCTCGAGAACGCGGTGAGCTCAACGCTGG		GACAAC	GLENAVSSTLGPSPG
				GGCCTTCCCCTGGGACGTTGCCTTCCTTACTTCCGTCCCAAGAGTGTGCT		CCAATA	TLPSLLPSQECANERF
				AACGAAAGATTCCTGTACCTTGCGCACCTGCCTGTTCGGAGCAAGCCTC		AATTAA	LYLAHLPVRSKPLPN
				TGCCCAACAACCTAGTTACCGACTTCATGGACGCCGCTGAGCGTTGTGC		ACAACG	NLVTDFMDAAERCA
				TCTTGCCTACATTGCACACCCTCGGACTCTACACTGCTGGCATTTCTCG		AAAAAT	LAYIAQPSDSTLLAFL
				CCCTTCCAAAGGTCGGCCTCACCCAGGCGCTCGCTCCAGAACAGCCCCT		GCAACC	ALPKVGLTQALAPEQ
				CAGGCCGTCAACCTTCCTTAAGCAGTTCCCGCATATCCCCTGGCCAGAAC		TCTATAA	PLRPSTFLKQFPHIPW
				AGCCACCCGCTCGTCGTCCTCCCAGCAATATTCGTCCAGACACCACCAAA		CCC (SEQ	PEQPPARRPPSNIRP
				CAAGTCATCAAACTCGTTGAGAATGGGCGCCTAGGTGCGGCAGAGAGG		ID NO:	DTTKQVIKLVENGRL
				GTGTTGGAGGAGGATGCTTCAGTAGCCGAACTCGATCAAGGGGTCATC		1275)	GAAERVLEEDASVAE
				GACCAGCTCATCACCAAGCACCCCAAAGGGCCGTCTTGTCCATTCGGCA			LDQGVIDQLITKHPK
				ATGCAGTGGGTCCAACTCCTGGTAAAGCTCCCGACATCGACACCATCCA			GPSCPFGNAVGPTP
				AAAGGCCCTCGACTCCTTCAAGCCCGACACAGCACCCGGCGTTAGTGGC			GKAPDIDTIQKALDSF
				TGGTCAGTCCCTCTCTTGAAGACGGCTGCCAAGAGGGAGCCGGTCAAG			KPDTAPGVSGWSVP
				CAGTTTCTCCAACTCCTCTGCGCCGCCATCGCCAACAACACCGCCCCTGG			LLKTAAKREPVKQFL
				TCGCTCTATGCTCCGCACTTCTCGTCTCATCCCCTTGAAGAAGGACGATG			QLLCAAIANNTAPGR
				GCTCTATCCGACCTATCGCTGTTGGTGAACTTATCTATCGGCTGTGTGCG			SMLRTSRLIPLKKDDG
				AAAGCTCTCATCATCTCGCATTTCCAACCCGACTTCCTCCTCCCGTTCCAG			SIRPIAVGELIYRLCAK
				CTCGGGGTCAAGTCAATCGGTGGTGTAGAGCCGATCGTGAGGCTGACA			ALIISHFQPDFLLPFQL
				GAGAGAGTCTTGGAGGGTTCTGCCGGCGCTGAGTTCTCCTTTTTAGCCT			GVKSIGGVEPIVRLTE
				CGCTCGATGCTTCTAACGCTTTCAACCGTGTAGATAGGGCCGAGATGGC			RVLEGSAGAEFSFLAS
				AGCAGCGGTCAAGACCCATGCGCCGACGCTTTGGAGGACATGCAAATG			LDASNAFNRVDRAE
				GGCCTATGGCGACTCGTCCGACCTTGTGTGGTGACAAAATCCTTCAA			MAAAVKTHAPTLWR

							TCKWAYGDSSDLVC
				TCCTCTCAAGGTGTTCGACAGGGTGACCCCTTTGGCCCTCTCTTCTTCTC			GDKILQSSQGVRQG
				GATCACCCTCCGACCAACCTTGAATGCCCTCAGTCAATCGCTAGGTCCGT			DPFGPLFFSITLRPTL
				CTACGCAAGCACTCGCTTACCTCGATGACATCTACCTCTTCTCAAACGAC			NALSQSLGPSTQALA
				TCGCAAGTCCTCAGCAAAACTACCCAATTCCTCGCCGACAAGCAGCACA			YLDDIYLFSNDSQVLS
				TCATCAAGCTCAATGAAAAGAAATGCAAGTTAATCAGCTTCGATGAGAT			KTTQFLADKQHIIKLN
				CAGGCAGGAGGGCTTCAAGATGCTAGGGACGATGGTAGGAGGTAAGG			EKKCKLISFDEIRQEG
				AGAAGCGAGCGGAGTTTCTGGAAGGCAGGATTCGGAAGGAAATGGCA			FKMLGTMVGGKEKR
				AAGGTGGGCAAGCTCAAGGATCTTCCACATCAACACGCGCTCCTTCTAT			AEFLEGRIRKEMAKV
				TACGTTTCTGCATTCAGCAAAATCTACGACACCTGCAGAGAAGTCTGCG			GKLKDLPHQHALLLL
				CTCGGACGACCTTGTAGACCTATGGGAGAGGCTGGACACGATGCTATG			RFCIQQNLRHLQRSL
				GGAGGAGGTGAAAAGGATGAGGATGAGGCAGCGGGAGGATACGGCG			RSDDLVDLWERLDT
				GAAGAGGAGGCTCTAGGGAGATCGTTGACGAAGCTACCAGCGCGACTG			MLWEEVKRMRMRQ
				GGCGGACTAGGTCTACTTTCCTTCAAAGATGTAGCCCCCCTTGCTTACCG			REDTAEEEALGRSLTK
				CTCGGCAGCCGAGGCCTCCGACACTCTCCTCGATAACCTAGGTCTCCTTT			LPARLGGLGLLLSFKDV
				CTTCGCCAGAGGAACCTCCAACTCCGATCCCCCAACGAACTCGATGCCGC			APLAYRSAAEASDTLL
				AGAACTCTGGGAATCGCAACAGGAAGCCATCCTACATAACCTCGGCGAC			DNLGLLSSPEEPPTPI
				ACTGAACGCAAGCGACTCACCGAGAATGCCTCCAGACTCGGCCGAAGTT			PQRTRCAELWESQQ
				GGTTATCAGTTATCCCTTACCTTCAACCCCTGCGCCTTTCCAATGTCGAG			EAILHNLGDTERKRLT
				ATTGCCTCTGGTCTCCATGACCGCACCCTGGTCGGCTCCTCGATCCCTGT			ENASRLGRSWLSVIP
				CTGTCGCTTCTGTGGGTCGGACTCACCTTTGGGTCACGACGAGCTTTGCC			YLQPLRLSNVEIASGL
				GCGCCCGCAACCCCTGGACCCAGCGCCGGCACAATGCCATCAACCGCGT			HDRTLVGSSIPVCRFC
				CATTTATCAACACCTCAAACAAATTCAAGGTGCCACGGTTGAGATTGAG			GSDSPLGHDELCRAR
				CCCCACACGCTGTCGGGACAAAGGAGAAACGACCTTCGGGTCAGAGGT			NPWTQRRHNAINRV
				TCCAGCGCTCTGGCCTTCACTGACTACGACCTGAAGGTTTACTCCCTCGG			IYQHLKQIQGATVEIE
				GGACCGAGACGCGAGAAGCACCGTCACACCCTGCGCCCCCAACGGCAA			PHTLSGQRRNDLRVR
				GCTGGCCGACTTCTGCTTGGACCGGTGCGTGAACTGGCTCGACAAGGT			GSSALAFTDYDLKVYS
				GGGTCAGGTCGTCTCTAAGAACGCTCCGAAGGTCACTGGTGGGGGTCTTT			LGDRDARSTVTPCAP
				AAACCAATCATCCTTTCCACTGGTGGCTTGATGAGCAGGAGCACAGCAG			NGKLADFCLDRCVN
				ACGAATGGAAGGACTGGAGGGACGCGATGCCGGTGGGGGGGTTCGAG			WLDKVGQVVSKNAP
				AAAATGGAGAAACGGATTGGTGTCGAGTTAGTAAAGGCAAGGGCGAG			KVTGGVFKPIILSTGG
				GACGCTGGTCTTATGAGGAAGAGGAGGTTGGATTATTTTTTCTTTTCTTT			LMSRSTADEWKDW
				AATAAGTTGTTTATTTAAGTAGTTTCTTTCATTCGGGCAACCCACACGAC			RDAMPVGGFEKMEK
				AACCCAATAAATTAAACAACGAAAAATGCAACCTCTATAACCC (SEQ ID NO: 1029)			RIGVELVKARARTLVL (SEQ ID NO: 1397)
CRE	CRE-12_CCri	.	Chondrus crispus	ACGCCCCCTATCCATTTCTGCCAGCCTCCCATCGGCTCGCCGTCTCCGCA	ACGCCCCC	TAAGTC	MSQPNISSAETPLSQ
				ACCCCTCTTCCTCGGCTGTACCAGTTCCGCTCCCACAACCTCCCTCGCCAC	TATCCATT	CTTGAC	LPTPVPTPPSPNPSL
				AATGAGTCAACCTAATATTTCGTCCGCTGAGACCCCCTTGTCTCAGCTGC	TCTGCCAG	GCCTGC	SLPTVRDLLLCPIRSSH

	CCACGCCTGTTCCCACCCCGCCTTCTCCCTCCAATCCCTCTCTCTCTCCC	CCTCCCAT	CCCCGTG			VYSSIPSSCLHSFTML
	TACTGTGCGTGACCTCCTCCTCTGCCCCATACGCTCCTCCCATGTTTACTC	CGGCTCGC	ATACAG			LIKTVRAASATMTPT
	ATCCATCCCTTCCTCGTGCTTACACAGTTTCACGATGCTCCTCATCAAGAC	CGTCTCCG	CATCGG			ESHRAFIHLHILPIAVL
	TGTCCGCGCTGCGTCGGCCACAATGACCCCAACTGAATCMCATCGCGCA	CAACCCCT	TACCCCT			RRSFRGETGWRSRT
	TTCATTCATCTACACATTCTTCCCATCGCTGTCTTGCGACGCTCGTTCCGT	CTTCCTCG	AGCATTT			GQHHALRQRIRRASS
	GGAGAAACCGGATGGCGTTCCCGCACAGGGCAACATCATGCCCTCCGC	GCTGTACC	GAATAA			GRHWAALWHEALA
	CAACGGATTCGCCGCGCCTCCTCGGGTCGGCATTGGGCTGCCTTGTGGC	AGTTCCGC	AAAA			AHQVDLDYRTRHSR
	ACGAAGCCCTTGCTGCACATCAGGTCGACCTTGACTACCGCACGCGTCA	TCCCACAA	(SEQ ID			RYQASATSRHRIGRA
	CAGCCGTCGTTACCAAGCCTCCGCTACATCGCGCCACCGCATCGGCCGT	CCTCCCTC	NO:			MRLAADAQYGRAM
	GCCATGCGTCTGGCCGCCGATGCCCAATATGGACGTGCAATGTCGGCCC	GCCACA	1276)			SALKAKPLPDLHAAA
	TCAAGGCCAAACCGCTGCCCGATCTACATGCTGCCGCCACCCGCGACAC	(SEQ ID				TRDTLTALHPPPASP
	ACTCACCGCGCTTCACCCTCCTCCTGCCAGCCCGGTTCAGCCTCTCTCACC	NO: 1153)				VQPLSPTDLPPVPEIT
	GACTGACCTCCCCCCCGGTCCCTGAAATTACGGAAGGTCAAGTCCTCCGA					EGQVLRAARALNPTS
	GCGGCACGCGCCCTTAATCCCACATCCGCTGCAGGACCGGACCATCTCT					AAGPDHLSPRILQLLA
	CTCCCCGTATCCTGCAGCTCTTAGCCCGCACCACTATCAGCCCGGAAGCT					RTTISPEAGVTGLSAL
	GGGGTTACGGGGTTGTCCGCATTGACGAACCTGGTTCGACGTCTCGCCC					TNLVRRLARGDIPDR
	GAGGTGACATTCCGGATAGAACTGCGCCTCTCCTTGCTGCTGCCACTCT					TAPLLAAATLIPLQPR
	GATCCCCCTCCAACCCCGCCCTCACAAAATACGGCCGATTGCTGTAGGG					PHKIRPIAVGQALRRL
	CAGGCTTTGCGCCGTCTGGTCACGAAGGTCCTTCTGCCCCCCGCCATCCA					VTKVLLPPAIQDTRD
	GGACACCCGCGATCACCTTCTCCCAGAACAGCTCGCCAACTCGGTTGCC					HLLPEQLANSVASG
	TCGGGCATGGACGCAATCGTCCATGACACGCGCATGCTTATGCATCGTC					MDAIVHDTRMLMH
	ACGGTCGAAACCCAGACTACATCATGGTCTCCGTAGACGCGCGGAATGC					RHGRNPDYIMVSVD
	CTTTAACACCTTCTCACGTCAGTCCCTGCTGGATCGTCTCCCTCTGCAGAC					ARNAFNTFSRQSLLD
	TCCTTCCCTCGCCCGTTTTCTCAATCTAATCTATGGCCGCACCGTTCCTGA					RLPLQTPSLARFLNLIY
	TCTCGTGCTGCCCTCTTCTCCGCGGTTTCTGATGAAAAGTCAGGAGGGC					GRTVPDLVLPSSPRFL
	ACCCAACAGGGGGACCCGGCAAGTATGCTCTTATTTTCGCTGGCAATCC					MKSQEGTQQGDPAS
	AGCCGCTCCTGCGTCGTCTCACCCGCGAGTGCCGTCTCGACCTGAACCG					MLLFSLAIQPLLRRLT
	CTGGTACGCGGATGACGGCACTCTGGTCGGGCCAATCTCGGAGGTCAT					RECRLDLNRWYADD
	CAAGGCACTCCGAATTCTTCGTGATGACGGCCCGCAGTCCGGATTTCAC					GTLVGPISEVIKALRIL
	GTCAATATCAACAAGTGCCGGGCATACTGGCCGACCGTAATGCCAGAAA					RDDGPQSGFHVNIN
	AGTTGTCCGAATTGCTCCGTATCTTCCCCCTTCACGTCGAGTGCGGCGAA					KCRAYWPTVMPEKL
	GGCGGTGTCGCCTTGCTGGGTGCCCCGCTCGGCACAGATGCCTTTGTGC					SELLRIFPLHVECGEG
	GCCGACATCTCATGAACAAGGTTCAATCATGCCATGCCTCCCTCAGCCTC					GVALLGAPLGTDAFV
	CTTGATGAAATTCCCGACGCGCGTACGCGGTTTCACCTTCACCGTGTAAC					RRHLMNKVQSCHAS
	AGGCTCGGTATGCAAAGTCGAGCATGTTTTTTCGCCTTACACCTCCCCACC					LSLLDEIPDARTRFHL
	TCTCCCTCCCAGCAGCTACAAAATTTGATGAACAACAAATCGCTGCTTCTAT					HRVTGSVCKVEHVFR
	TCTCGGTTGAATGATGTGGCCGTTTCCACATCCATGGCTACACAAATAG					LTPPHLSLPAATKFDE
	GCCTTCCGTTTCGCCTCGGTGGACACGGCTTCACCCCACTGTCACCATTC					QQIAAYSRLNDVAVS

			ATCCATGCGTCCTACGCTGCCAGTTTAATTGAGGCGGGCACCTGTTCGTGT				TSMATQIGLPFRLGG
			GAAGGGCCCACATAACCCCTCCGAGTCGTTTTATCGCCGCATGGCCCGT				HGFTPLSPFIHASYAA
			CGTCATATCGTCCACGTACTAGGGGCCTTGAACCCTGAGGTCCGCACCC				SLIEAAPVRVKGPHN
			GAGGCATTCTTGGGACCCATTCTCCCCTCGGACCATTTGAACCAGAGGC				PSESFYRRMARRHIV
			CCTTTTGTCTAGACCTGAACGCGTACACCACACATTAATTCAGGCTATGC				HVLGALNPEVRTRGI
			AGGGGGCCACTTCTCGACTCTACTGGGAACACACCGCGTGGGACCTTGA				LGTHSPLGPFEPEALL
			CCCTCTCCCTCGCAACCACAGTGCCGCCTCTGTCCGCCGACGTGCCCGGT				SRPERVHHTLIQAM
			ACAATTCCCTCCGTGCTCCGGGCGCCGCGTCGTTTCTATGCAGCCACCCC				QGATSRLYWEHTAW
			TCACTCACTTCTCGAGTCCCTTCTGCGGTGTGGTCCTGTATGCTACGCCG				DLDPLPRNHSAASVR
			GCATCTGGATACACCCGTCTACTGTGACTCTATTCGGCCTCTCATATGTT				RRARYNSLRAPGAAS
			CGCATTGCTGTAAGCCAATGGACGCTCGCGGTGATCATGCTGCAATATG				FLCSHPSLTSRVPSAV
			CCGTCATGGCTTCGGCGTCGTTCACCGTCATAACACTGTACGCAACCTAC				WSCMLRRHLDTPVY
			TCGCCCGTCACGCGTTCCGCGCCGCCGGTCTCTGCTGCGACCTTGAGGT				CDSIRPLICSHCCKPM
			CCCTTCTCTCTTGCCGAATACCGCGAACCGCCCCGCCGATATTCTCGTCC				DARGDHAAICRHGF
			AGCCCGCCCCGCCTCCTTCGGGCGCTCTCCCGGACCGCCCCACTGCGTA				GVVHRHNTVRNLLA
			CGACGTAACCGTTCGTTCCCCCTACTGTCGCTCTACAATGTCTCTCGCTG				RHAFRAAGLCCDLEV
			CGAAAGGCCTCGCGGGGTGCAGCGGAAGCTGCTGATTTGGACAAGCTTC				PSLLPNTANRPADILV
			GCGTCCATTCCCGTACAGTGCGTGACGCATTTCACCTCCAGCCTGACTCC				QPAPPPSGALPDRPT
			CCACTCCCTCTACTCGACTGGCACTTTGTCCCGCTCGCATTTGATACTCTC				AYDVTVRSPYCRSTM
			GGCGCGACCAGCTCTCGCACGATGGCAGTCCTTGAGTACCTCGCTCACC				SLAAKGLAGAAEAAD
			GCATTGCCAACCGGACATATTCATCTTACGGGACCGCCAAGATACGTCT				LDKLRVHSRTVRDAF
			ACTACAACGCATCAGTTTCGCTGTTTGGTCCAGTTTGGCCTCTGCCACCC				HLQPDSPLPLLDWHF
			TTTCCCGTATGCCCTATCACGGCGCGGCCCTATCGAGCCCCGCCCAAGT				VPLAFDTLGATSSRT
			GTAAGTCCTTGACGCCTGCCCCGTGATACAGCATCGGTACCCCTAGCATT				MAVLEYLAHRIANRT
			TGAATAAAAAA (SEQ ID NO: 1030)				YSSYGTAKIRLLQRISF
							AVWSSLASATLSRMP
							YHGAALSSPAQV
							(SEQ ID NO: 1398)
CRE	CRE-13_CCri	.	Chondrus crispus	CNCCAGCCAMCGATCCCGCCGCCACTCGCMGCCCGGCCGTCTCGACCG	CNCCAGCC	TAATTCA	MAXXPXISPPGAPPA
				CCACCTCCCCGASGCCCCAGCCCATCATGGCCTSTWMGCCCCWGATWT	AMCGATC	CCTTCAT	PLRYRMLQCPPPLPK
				CTCCCCCGGGGGCCCCCCCCTGCTCCGCTGCGGTACCGGATGTTACAATG	CCGCCGCC	ATCTGCT	XXXXPVPHPMSSPIR
				TCCACCSCCGCTACCSAAGCSMCASTMGTTKCCGGTCCCCCCACCCGATGT	ACTCGCM	AGTGTC	XRLPHRXMRGPPSXT
				CGTCACCCATACGCCKCCGCCTCCCMCACCGGMCGATGAGGGGACCCCC	GCCCGGCC	TCTGTAA	PPRDMHRPHGTPGP
				CTTCCCSAACCCCTCCCCGAGACATGCACAGACCSCATGGCACCCCCGG	GTCTCGAC	GCGCAC	HSHRXCGRPPXHCTH
				WCCGCACTCTCACCGATKCTGTGGCCGCCCTCCCMACCACTGCACCCAT	CGCCACCT	CCCTCAT	ASXQPRXAXHXLQXP
				GCCTCCSCACAACCSCGSMSAGCCCMGCACSCTCTCCAAWGGCCGAAA	CCCCGASG	GCATTG	KLRSPPPHPHVSPLIL
				CTSCGCAGCCCACCTCCACATCCGCATGTTTCACCCCTCATCCTCTGCAST	CCCCAGCC	ATAAAA	CXGPLPTPMTQPRM
				GGCCCCCTCCCAACGCCCATGACCCAACCACGCATGAAACGAGCGCTCT	CATC (SEQ	TTACCCC	KRALSXSAKAPPTKRP

		CCASAAGCGCCAAGGCGCCCCCTACCAAGCGCCCCTCTGCCTCTCAGGG	ID NO: 1154)	CCA (SEQ ID NO: 1277)	SASQGPAASSHDXPR
		CCCAGCCGCGTCTTCCCATGACKAATGACCSCGGACGCCCCCCACCTCMA			TPPPXPPRPPPPYRFPP
		CCCCCGCGCCCGCCTCCCTACAGGTTCCCTCCCCCWACTCTCGACCAGCA			PTLDQHXFALSXAYP
		CTTMTTCGCCCTCTCWTMAGCCTACCCCCACCCGMCTCCCMGGCGCCC			HPXPRRPPSPXRXLR
		ACCCTCMCCTTKCCGTCSGTTGAGGCACTCCTTTCCTCCCCGATTCGGCC			HSFPPRFGXQTFSSIP
		STCAGACATTTTCTTCCATTCCGGGACCGCGCCTTCATAGCACTGTATTAC			GPRLHSTVLLLIRLVR
		TTCTCATCCGCCTCGTCCGCGCCGCTACAGCCGCCAACACTCCCGAAACC			AATAANTPETTTLXS
		ACCACACTGMATTCTTGCACCTTCACCTGCTCCCGACTGCCSTTCTTCGA			CTFTCSRLPFFERPSX
		GAGGCCTTCCGKGGCGAGSCTGGCTGGAGGTCCTCGCGCGGTCAACTTC			AXLAGGPRAVNFMIL
		ATGCTCTCCGCTTGCKGATACGGAGAGCGTGTACKGGACGAGAGTGGG			SACXYGERVXDESGX
		GWCTCTTATGKAAWGAAGCMCTAGATGCCCACAGCKCCAGGACAGAA			SYXXKXHCITSHPPRP
		TGGCAGCACACGCATGCCCGGCGCCCCTCGCCACCCGTTTCCCCATCGG			RRYSRQHSRNHHTXF
		CACGTGCCGCCCGCGCTATGCGCCTTGCCTCCCAAGCTCAATACGGCCG			LHLHLLPTAXLREAFR
		CGCCATGCGCACATTTACCAACCCCCCTCTAGCTGACCTCAACGACCCGG			GEXGWRSSRGQLHA
		CCACGATGGAGCGGCTCCAAGCCCTTCACCCCACTCCTACCGTGCCCGTC			LRLXIRRACTGREWG
		GTGCCCCTGCCACCCTCCGCACAGCCTCGACCACCCGAAGTCACCGMGG			LLXXEALDAHSXRTE
		AGGCGGTCWTGCGTGCGGTTCGTCGCCTCAATCCGAACTCGGCGGCCG			WQHTHARRPSPPVS
		GCCCTGATCGCATGTCCCCGAAATTGCTTCACCTCCTGGCTCACACTCCC			PSARAARAMRLASQ
		ATAAGCCCAGAAGCGGGCGTCACCGGTCTCTCKGCGCTAACCAACCTCG			AQYGRAMRTFTNPP
		TCAGCCGCCTGGCTCGCGGCTCCCTCCCACCCTGTACGATCCCACTGGCC			LADLNDPATMERLQ
		AGTGCGGCGACACTTCTKCCGTTGCAGCCCCGACCGGGAAAAATACGCC			ALHPTPTVPVVPLPPS
		CGATCGCTATTGGGCAAGCCCTWCGCCGGCTTGTCACAAAAKTMCTTCT			AQPRPPEVTXEAVXR
		TCCTGCCGCCATCGACGACTGTCGGGACCACCTTGCTCCCGAACAAMTG			AVRRLNPNSAAGPD
		GCMAACGGCATACCMAACGGCATTGACGCTATCGTACACGACGCACGC			RMSPKLLHLLAHTPIS
		ATGCTAGTACGACGCCACGGTAACGACCCACACTACKTAATGGTGTCTA			PEAGVTGLSALTNLV
		TTGACGCTTCCAATGCGTTCAATAATTTCTCACGSCAACAAGTCCTCGAC			SRLARGSLPPCTIPLA
		CAGCTGCCCACTCGAGCACCATCGCTCTCACGATATTTGGATATGGTGTA			SAATLLPLQPRPGKIR
		CGCACGCGCCCCCTCCCCCCTCGTCTTGCCTTCATSCCCGCCTACCATACT			PIAIGQALRRLVTKXL
		CCACAGCCGGGAMGGATCACACAACAAGGGGACCCTGCAAGCATGCTCCT			LPAAIDDCRDHLAPE
		TTTCTCGCTTGCCCTCCAGCCGCTCACGCGCCTCATTTCACGTGAGTGTG			QXANGIPNGIDAIVH
		AMCTWKTAATGAACCGCTGGTATGCGGACGACGGAACTATCATTGGAC			DARMLVRRHGNDP
		GGATTGACGAAGTTKCCAAAGCCCTTGATATCATCACTAAAGAGGGGCC			HYXMVSIDASNAFN
		CAGGTTCCAATTCTTCCTCAACCCTTCGAAGACACGCGTCTTCTGGCCAA			NFSRQQVLDQLPTRA
		GCAGGCAGCMAGACCTCCTCAGCCCGCTCATGACAGTGGGTCCTCTGC			PSLSRYLDMVYARAP
		GMGTCATCGATGAAGGCGGTGTGGMTCTGCTCGGCGCCCCCCATMGGG			SPLVLPSXPPTILHSRX
		TCACCAAGCTWTATGGCACAGTACATTCGGGAAAAWTTGAACACTTGC			GSQQGDPASMLLFSL
		AAAACCGCMCTCGCCCATCTCGACCATATCCCCGAGGCCCGCATGCGCT			ALQPLTRLISRECXLX
		TTCACCTGCATCGGGTGTCTGCTTCTGCATGCCGCTTGCAGCACCTCTTC			MNRWYADDGTIIGRI

		CGGTTGGTCCCCCCCGGATTTCGCGWTGCCGTTTGCACACAACAATTCGACC			DEVXKALDIITKEGPR
		GTGACCAACTCGMAGCCTATGMGCGCTTTAATAGTGTGACTATGTCGC			FQFFLNPSKTRVFWP
		CAAGAATCGTGCCCAAATACGGCTGCSTTTTTCMCACGGWGGCCACGG			SRQXDLLSPLMTVGP
		CCTCACCTCATTGGCATCTACCATACACGCCTCWTACGCTGCTAGCCTCA			LRVIDEGGVXLLGAPI
		TCGATACCGCTCCAGCACGGCTACAAGGTCCCCACTTTCCCGCCGTCTCT			GSPSXMAQYIREXLN
		CAGTATCAGCGTTTTGCACGAGGCCCGTTGCGGGTCGTTCTTCGAAATTT			TCKTALAHLDHIPEAR
		ACCTTCATTCGTKCAACCCGCACACTTCTCGATGACGGAAWCGGACCTC			MRFHLHRVSASACRL
		GGMTGCCTTGAACCAKCTGCGCTACTGGCGCGACCTGAACGCATACAC			QHLFRLVPPDFAXPF
		ACCTTTCTACTTCAGGCGCAATACAGTGCAGCAGCWAGCTCGTACTGGC			AQQFDRDQLXAYXR
		AAAWACCCCTCTGGGAGTCCTTCCCCAACCCTGGTGATCACAGCGCAGC			FNSVTMSPRIVPKYG
		CTCGCTACGCAAACGAGTACGCTACAACTCCCTGCTTGCCCCWGGGGCC			CXFXTXATASPHWHL
		ACCAGTTTTCTCACTGCACACCCSGCCGCCACCTCTCGGGTCCACAACGC			PYTPXTLLASSIPLQH
		AACTTGGTCCACGATGTTACGTCGGCACCTCGACGCCCCCGTGACCAAC			GYKVPTFPPSLSISVL
		GATTCCATATCGCCGTTGCGATGTKCTCACTGCTCCAAGCCTATGGATGC			HEARCGSFFEIYLHSX
		CCGCGGCGACCACGCGWCCATTKGCAGCCACGGGTTTGGTACGTTGCA			NPHTSRCDYVAKNR
		CCGGCATAACACCGTCAGGAACGTCCTCGCCMGGCAGTTATTCCGMGT			AQIRLXFSHGGHGLT
		CGCTGGCCTCGCCTACTCGCTCGAAGTACCCTTTCTGATTCCCAACACCG			SLASTIHASYAASLIDT
		CCGCCCGTCCCGCAGATATTCTCGTCCAACCACCTCCTCCAGCCCCCTGGC			APARLQGPHFPAVS
		CTACCTCCTGACMAACCCACAGCCTATGATGTCACGATTTGTAGCCCTTT			QYQRFARGPLRVVLR
		TCGCCGCGGAATGTTATACCATGCCGCCCGTCACCGCGGCGGAGCCGCC			NLPSFVQPAHFSMTE
		GACGCCGCATCTGTAAGGAAGMGCAAAGCCCTCGAGCGCACTATCCGC			XDLGCLEPXALLARPE
		MACGCTCTCCTTATCGAGGACGACAATCSTCCGCCGCCTCTTGACTGGC			RIHTFLLQAQYSAAAS
		ACTTTCAACCGCTTTCCTTCGACGCWCTGGGMGCCCCCTCTCAGTCTAC			SYWQXPLWESFPNP
		TGTACACGTTATCGAAGATCACGCTAAGCTCATGGCCCTCCGCAACTCGT			GDHSAASLRKRVRYN
		GCACMATTGCAACTGCCAAATCACGCATCCAACAACGCCTCAGCTTTGC			SLLAPGATSFLTAHPA
		TATATGGTCCAGTGCTGCCGCCGCTATCCTCTCTCGCCTACCGACACACG			ATSRVHNATWSTML
		CCGCGGACATCTCATACCCGATAGAAGTATAATTCACCTTCATATCTGCT			RRHLDAPVTNDSISPL
		AGTGTCTCTGTAAGCGCACCCCTCATGCATTGATAAAATTACCCCCCA			RCXHCSKPMDARGD
		(SEQ ID NO: 1031)			HAXIXSHGFGTLHRH
					NTVRNVLARQLFRVA
					GLAYSLEVPFLIPNTA
					ARPADILVQPPPPAP
					GLPPDXPTAYDVTICS
					PFRRGMLYHAARHR
					GGAADAASVRKXKA
					LERTIRXALLIEDDNX
					PPPLDWHFQPLSFDA
					LGAPSQSTVHVIEDH

CRE	CRE-1_ACas	.	Acanthamoeba castellanii	TAACCCTAACCCTCTCCCTCGGCCCCTCTACCCTAAAGCGCCCTAATCGA	TAACCCTA	TAAGCC	AKLMALRNSCTIATA
				CCGGCGACGCCCTAATCGCTACCCTCTACGCCCTAATCGACTTTGGCGCC	ACCCTCTC	GCGCGA	KSRIQQRLSFAIWSSA
				AAAGCGACTTTCCCCGGCCGATTTTCTTCCTGCCTTTTTCTTTCTCTCCA	CCTCGGCC	CGAGGA	AAAILSRLPTHAADIS
				AGCGACGCGCCTTTTACTTTGCCGCCGTTCTGTTTTTTCTTTTCTCTTTGC	CCTCTACC	CGGCCA	YPIEV (SEQ ID NO: 1399)
				ACTTCGCTTCTACTTCACACCTCCTCCTCCTCCTTCTCGACCCGCGCGGCC	CTAAAGCG	GGACGA	MATTTISRSPSSSSSS
				TCGAGCGACTTGCTGCAGCGGCTCCCGGCCTCCCCCACGCGGCCTGCTA	CCCTAATC	CCAGGA	SSARSRASASTSASVA
				CTCCCGCTTTCTAGACGCCCCCGGTCTTGCTCTCAGTCTCCCGCATCGAA	GACCGGC	CGACGG	SIPRLFRDGRFHCPLA
				GCGGTAGTCGGGGTACGTGCTCAAGTGACTCAAGCCTCTTTTCAGCCTC	GACGCCCT	CGACGG	HCQTRTSTWQDLSA
				GGCGCTCTCTCAATCCGCCTCAGTCTTAGCCTTTCAAGTTGCTCGATTAC	AATCGCTA	CGACCA	HLTRMHDGDVPRDV
				GCTCTCGAATCGCTCTCTCTCTCAGTCTCAGTCTCAATCTCGATC	CCCTCTAC	CCTAGC	AAACGIVQCLHEGCR
				TTGCCTTCGCCTTCGTCTCGACGCCTTGCTCTCGWAATCGCTGCCACTAC	GCCCTAAT	ACCGCA	KWFRGAAGLASHRG
				GTGCCAGCTTTTTCGTGCCTTGTCTTCGTGTCGACCCGGACCGTTTGCAA	CGACTTTG	CGCGCC	KARHAPPPAPRAALA
				GCCCTCGCCTTCGTACCCCGCTCTCGTAGCCGTTCTCATCGCTGAAGCGT	GCGCCAAA	ACGACA	VAAVPRADSRGRTP
				TCTACGCGCTGGCAGCAAGCCTCGGCCCTAGCTTGTAGCGCCGCCGGTG	GCGACTTT	TATTGTC	APTPSVAPPXAGPPP
				GCCGCTCGCCAACATGGCTACGACGACCATCTCACGATCCCCTTCGTCTT	CCCCGGCC	GCGCGC	RAAPRAAPSPLPCPP
				CTTCTTCTTCTTCCGCTCGCTCGGGGCGTCAGCTTCCACGTCCGCCT	GATTTTCT	TGTACA	ALPHPPPSASPPTSSV
				CAGTTGCGTCGATACCCCGCCTCTTCCGCGATGGCCGCTTCCACTGCCCT	TCCTGCCT	GGCGGC	TSPCSPPTTPPSQPSP
				CTCGCCCACTGTCAGACCCGCACGTCCACATGGCAGGACCTCTCCGCGC	TTTTCTTTT	TAGGTC	DLFSGFANAPTTPSP
				ACCTCACACGCATGCACGACGGTGACGTGCCCCGTGACGTCGCCGCCGC	CTCTCCAA	GAGCCC	PSTPXSSPAGSPIPAA
				CTGCGGCATCGTGCAGTGCCTACACGAGGGCTGCCGCAAGTGGTTTCGC	GCGACGC	AGCCGA	RRFVLPVATPYPAPA
				GGAGCTGCAGGACTTGCCTCTCATAGGGGCAAGGCCCGTCACGCCCCG	GCCTTTTA	CCGTTCT	PRANRPKLSPVARPF
				CCACCAGCCCCCCGCGCCGCCTCGCCGTCGCCGTGCCCCGCGCGG	CTTTGCCG	GAGCCT	VPKARAGAIPEASSP
				ATTCTCGCGGTAGGACCCCAGCCCCGACCCCCTCGGTAGCCCCGCCCWA	CCGTTCTG	CAGTCG	VTPQDRAVSRREDA
				CGCCGGTCCTCCGCCGCGAGCTGCGCCACGCGCCGCCCCCAGCCCACTG	TTTTTTCTT	GCTTGA	AAAPSSAPGLGLADE
				CCGTGCCCGCCTGCGCTCCCGCACCCGCCCCCCTCTGCCTCTCCTCCCAC	TTCTCTTTG	GCCCCC	HEDDDTYGGDTIALT
				CTCCAGCGTGACGTCCCCGTGCTCCCCGCCCACGACCCCGCCGTCGCAG	CACTTCGC	GGCTTC	APHAPRETRAPFEFE
				CCTTCGCCCGACCTGTTCTCGGGTTTCGCGAACGCGCCTACCACGCCATC	TTCTACTTC	CCAAGG	ACFLEEEAPATAGDL
				GCCGCCCTCCACGCCGWCWTCGTCGCCAGCAGGCTCGCCCATCCCGGC	ACACCTCC	CCTACC	TQGYNAHLRLFIELPA
				TGCCAGACGCTTCGTCCTGTGGCCACGCCCTACCCGGCCCCCGCG	TCCTCCTC	GGGGCG	RGLAVPTNCRGAART
				CCGCGTGCTAACAGGCCCAAGCTGTCGCCGGTCGCGCGCCCCTTCGTCC	CTTCTCGA	GCTCTTT	KLQRERLLDIAAGRIP
				CTAAGGCGCGAGCCGGAGCGATACCTGAGGCGTCCTCACCTGTGACGC	CCCGCGCG	TTCGCCC	AIPDPPCDAPGADDA
				CTCAGGACCGCGCCGTCTCACGCCGCGAAGACGCCGCCGCCGCCCCTTC	GCCTCGAG	TGGTTTT	LRGFPVSGTTAGDVS
				GTCCGCGCCGGGCCTCGGCCTAGCAGACGAACACGAGGACGATGACAC	CGACTTGC	TGCCGG	NDDDSGGVHDRPAA
							TASARQAKRLVEQGL

		GTACGGCGGTGACACAATCGCGCTCACTGCCCCCCACGCGCCCCGTGAG	TGCAGCGGG	CCTGTTT	SSRALRALERGEPAV
		ACCCGCGCCCCCTTCGAGTTCGAGGCGTGTTTCCTCGAGGAGGAAGCCC	CTCCCGGC	TTTCTCT	ASADTLGRLEALHPP
		CAGCCACCGCCGGCGACCTCCCGCCTTACGCGCGCGCCTTCCTCGCTTGC	CTCCCCCA	CCCCCTT	NPTDRGLWPGAPKA
		CCGTCAGCCCGTCTCCAGGAGATCCCGCGCCGCCTCAAGTCCGCGTGGC	CGCGGCCT	TTCCCCC	AIPRVTAKHLAQVAK
		AGGCCGCCGCCAAGACCATCGCGGAAGCCGCGCTGGATTGCCACACCG	GCTACTCC	CTTTTCC	ELPRGSAPGPSGWTF
		CGGGCGACACGCAGGGCTACAACGCCCACTTGCGGCTCTTCATCGAGCT	CGCTTTCT	ATTTGTA	ELVQAAIDRQPTGTV
		GCCGGCCCCGCGGGCTGGCAGTGCCCACCAACTGTCGAGGCGCCGCCCG	AGACGCCC	CTTAGTT	AAFLIDMAQRALRGT
		CACCAAGCTTCAACGAGAGCGCCTGCTCGACATCGCCGCCGCCGGAAGGAT	CCGGTCTT	TTTCCTT	LHWRGLLTASRLVAL
		CCCGGCCATCCCGGACCCGCCGTGCGACGCCCCGGGCGCCGACGACGC	GCTCTCAG	CGGCCG	KKPDGGVRPIAVGEA
		CCTACGCGGTTTCCCCGTATCGGGGACGACAGCCGGCGACGTCAGCAA	TCTCCCGC	CGGCAG	LYRVIGRLVLKADRV
		CGACGACGACAGCGGAGGTGTGCACGATCGGCCCGCGGCCACCGCCAG	ATCGAAGC	CTTGTTG	MSSADATQYVGRHQ
		CGCCCGGCAAGCCAAACGGCTAGTGGAGCAAGGGCTCTCCTCCCGAGC	GGTAGTCG	CCCGGC	YGVAYPGGVEAPVH
		CCTCCGTGCCCTTGAACGGGGCGAGCCCCGCGGTCGCCTCGGCAGACAC	GGGTACGT	ATAGTG	AVRELHDSGQLRAVV
		CCTCGGGCGTCTCGAGGCGCTCCACCCGCCTAACCCCACCGACACAGAGGA	GCTCAAGT	TTAATAT	SLDWRNAFNSLDRV
		CTATGGCCCGGCGCCCCGAAGGCCGCGATCCCGCGGTCACCGCCAAG	GACTCAAG	GTTTAA	HTALLIADRAPALARL
		CACCTGGCCCAAGTGGCCAAAGAGCTCCCGCGCGGTAGTGCGCCGGGT	CCTCTTTTC	AAAACG	YEWSYREDSVLVLPR
		CCCTCGGGCTGGACGTTCGAGCTCGTGCAGGCGGCCATCGACCGCCAA	AGCCTCGG	TGTAAA	AFEKAGLPASLLSQA
		CCCACGGGCACGGTTGCCGCGTTCCTCATCGACATGGCGCAGAGAGCCC	CGCTCTCT	TAAATA	GVRQGDVLGPLFFAI
		TCCGGGGCACCCTACACTGGCGGGGATTGCTCACCGCCAGCCGCCTTGT	CAATCCGC	ACTGTTT	GAAPVLDEIDAIPYVT
		CGCGCTGAAGAAGCCCGACGGCGGTGTACGACCCATCGCCGTAGGCGA	CTCAGTCT	AACCCT	PRAYLDDIFVTIPHGV
		GGCCCTCTATCGCGTCATCGGCCGCCTTGTTCTCAAGGCCGACAGGGTG	TAGCCTTT	AACCCT	TDAATKAAVAATFAT
		ATGTCGAGCGCCGACGCCACGCAATATGTCGGGCGGCACCAGTATGGC	CAAGTTGC	AACCCT	AEREGAAAGLRLNRC
		GTGGCCTACCCCGGTGGGGTTGAGGCCCCGGTCCACGCCGTCCGCGAA	TCGATTAC	AA (SEQ	KSAVWAADAEALLP
		CTGCACGACAGCGGCCAGCTCCGAGCGGTCGTCTCGCTCGACTGGCGTA	GCTCTCGA	ID NO:	PHAAGAREDVESCA
		ACGCGTTCAACTCGCTCGACCGCGTGCACACGGCCCTGCTCATCGCCGA	ATCGCTCT	1278)	PVREGLKILGAPVGSP
		CCGCGCACCCGCTCTCGCGCGACTCTACGAGTGGTCCTACCGTGAGGAC	CTCTCTCA		AFVAKSLDGIIKRAIG
		TCAGTCCTCGTGCTGCCGCGCGTTCGAAAAGGCGGGGGCTGCCGGCC	GTCTCAGT		TLDLVADAELPLQHK
		TCCCTGCTCTCCCAGGCCGGCGTGCGCCAGGGCGACGTCCTGGGACCCC	CTCAGTCT		LVLLRQCVAQIPTFW
		TCTTCTTCGCCATCGGCGCTGCCCCGGTCCTCGACGAGATCGACGCCATA	CAATCTCG		ARAVPDAGPALAVW
		CCGTACGTGACGCCGCGAGCGTACCTCGACGACATCTTCGTCACGATAC	ATCTTGCC		DTALLRRTGALVGLD
		CCCACGGTGTCACGGACCGCGACCAAGGCCGCCGTCGCTGCCACCTT	TTCGCCTT		VRDGSLQADIARLPV
		CGCTACGGCGGAACGCGAAGGCGCGGCCGCTGGCTTGCGGCTCAACCG	CGTCTCGA		RLGGLGLRSMKDTAP
		CTGCAAGTCGGCGGTGTGGGCCGCGGACGCAGAAGCCCTCCTTCCCCCC	CGCCTTGC		RAFVASILFAAALANT
		CACGCCGCTGGCGCGCGGAGGACGTCGAGAGCTGCGCACCAGTGCG	TCTCGWA		RRSELTCSASTARRLR
		CGAGGGCCTCAAAATCCTCGGCGCGCCCGTGGGCTCGCCCGCCTTCGTC	ATCGCTGC		AALPELARTDACNDE
		GCCAAGTCGCTCGACGGCATCATCAAGCGCCATCGGCACACTCGACC	CACTACGT		AAWRRSIARGVFPD
		TCGTCGCTGACGCCGAGCTACCGCTGCAGCACAAGCTGGTGCTGCTACG	GCCAGCTT		VDKLGTTQLQRVLQ
		GCAGTGCGTGGCCCAGATACCCACGTTCTGGGCCCGCGCCGTGCCCGAC	TTTCGTGC		GMADSKSAHRTRRQ

		GCAGGCCCGGCCCTCGCCGTCTGGGATACAGCGCTCCTCAGGCGCACG	CTTGTCTT		VPFLFAAVFEDAATP
		GGCGCGCTGGTCGGACTTGACGTGCGGGACGGGTCCCTGCAAGCCGAC	CGTGTCGA		GSGAWLAAIPSDPTL
		ATCGCGCGCCTGCCTGTCCGCCTGGGCGGTCTCGGCCTCCGTTCAATGA	CCCGGACC		VLPDAELAEAVRIKLL
		AGGACACGGCGCCCCGGGCCTTCGTGGCCTCGATCCTGTTCGCCGCCGC	GTTTGCAA		TTTANAAGVCPACH
		GCTCGCCAACACGCGCCGATCCGAGCTCACGTGCAGCGCCAGTACGGC	GCCCTCGC		KTGIDPSHAYTCVSLS
		CCGACGCCTCAGAGCCGCCCTGCCCGAGCTGGCACGCACCGACGCGTG	CTTCGTAC		HLRTARHDVVVRRVE
		CAACGACGAAGCCGCCTGGCGGCGTCCATCGCCAGGGGAGTCTTCCC	CCCGCTCT		LACKTEKPVREHVLAI
		CGACGTGGACAAGCTGGGCACCACACAACTGCAGCGCGTCCTCCAGGG	CGTAGCCG		PPVAPTDNNNNGDE
		GATGGCGGACTCCAAGTCCGCCCATCGGACCCGCCGCCAAGTGCCCTTC	TTCTCATC		DGSPVTTADDNADG
		CTCTTCGCCGCCGTGTTCGAGGACGCCGCCACGCCGGGATCCGGTGCCT	GCTGAAGC		HAVATKRRPETRASA
		GGCTGGCCGCCATACCCTCCGACCCGACCCTCGTCTTGCCAGACGCCGA	GTTCTACG		RAAAAAATAAAAAAI
		ACTGGCCGAAGCCGTGCCGCATTAAGTTGCTCACGACGACGGCCAATGC	CGCTGGCA		INDNSLLSDDDDDDD
		GGCCGGCGTCTGTCCGGCATGCCACAAGGACCGGCATCGACCCGTCCCAC	GCAAGCCT		HDDNCHGEERGEGE
		GCGTACACGTGCGTTTCGCTATCCCATTTGCGCACAGCACGCCCACGACG	CGGCCCTA		RNVTCPGHYTATPFA
		TGGTCGTTCGCCGAGTCGAGCTCGCCTGCAAGACCGAGAAGCCGGTCC	GCTTGTAG		ADDTLDNSDEDNED
		GCGAACACGTGCTCGCCATCCCCCCCGTCGCGCCCACCGACAACAACAA	CGCCGCCG		NAHEDDDEDGKDD
		CAACGGCGACGAGGACGGCAGCCCAGTCACCACCGCCGACGACAACGC	GTGGCCGC		NDDDVYNNCNSSSS
		AGACGGCCACGCGGTCGCGACCAAACGTCGCCCCGAGACCCGCGCTTC	TCGCCAAC		DGDEGGDDLDYEYS
		AGCCAGAGCCGCCGCCGCCACCGCCGCCGCCGCCGCCGCCAAT	(SEQ ID		DQSVTRSVDAATGES
		CATCAACGACAACAGCCTCCTGAGCGACGACGACGACGACGACGACCA	NO: 1155)		PNPERPTTPTRALLRA
		TGACGACAACTGCCACGGAGAGGAAAGAGGAGAGGGAGAAAGGAAC			DLWLPATSTAVDVM
		GTCACGTGCCCCGGCCACTACACCGCCACACCCTTCGCCGCGGACGACA			VAAACRRSRAKAFDR
		CGCTCGACAACAGCGACGAGGACAACGAGGACAACGCTCACGAAGACG			AVSRKAAKYGPAVA
		ATGACGAAGACGGTAAAGATGACAACGACGACGACGTCTACAACAACT			DGSIAKVVPFVVSPF
		GCAACAGCAGCAGCGACGGGGATGAAGGCGGTGACGACCTGGAC			GVLSRPAKAFLKRAM
		TACGAGTACAGTGACCAGAGCGTCACTCGAAGCGTCGACGCCGCGACG			GDTTAAKQAKARLRL
		GGAGAGAGCCCCAACCCCGAGCGCCCTACCAGCGCCCACCCGCGCACTAC			AVAAVRGTARLSYA
		TACGCGCGGACCTGTGGCTACCCGCCACCTCCACCGCGGTGGACGTGAT			WGACAALIVGGN
		GGTCGCGGCCGCCTGCCGTCGGTCACGCGCCAAGGCCTTCGACCGAGC			(SEQ ID NO: 1400)
		CGTCAGCCGCAAGGCCGCGAAATACGGCCCTGCGGTAGCCGACGGCTC			
		GATCGCCAAGGTGGTGCCGTTCGTCGTGTCGCCCTTTGGCGTACTCTCG			
		AGGCCGGCCAAGGCCTTCCTCAAGCGCGCCATGGGCGACACGACGGCG			
		GCCAAACAGGCCAAGGCGCGTCTGCGCCTCGCCGTGGCCGCCGTCCGA			
		GGCACGGCCCGCCTCTCCTACGCCTGGGGCGCCTGCGCCGCCCTCATCG			
		TCGGCGGCAACTAAGCCGCGCGACGAGGACGGCCAGGACGACCAGGA			
		CGACGGCGACGGCGACCACCTAGCACCGCACGCCACGACATATTGT			
		CGCGCGCTGTACAGGCGGCTAGGTCGAGCCCAGCCGACCGTTCTGAGC			
		CTCAGTCGGCTTGAGCCCCCGGCTTCCCAAGGCCTACCGGGGCGGCTCT			

			TTTTCGCCCTGGTTTTTGCCGGCCTGTTTTTTCTCTCCCCCTTTCCCCCCT				
			TTTCCATTTGTACTTAGTTTTTCCTTCGGCCGCGGCAGCTTGTTGCCCGGC				
			ATAGTGTTAATATGTTTAAAAAACGTGTAAATAAATAACTGTTTAACCCT				
			AACCCTAACCCTAA (SEQ ID NO: 1032)				
CRE	Cre-1_FCy	.	Fragilariopsis cylindrus	ATCAATCTAATACTGAAGGCAATACCAAACTCAACCCGAAATCAAAATC	ATCAATCT	TAGCAC	MAPLPWNAATSSPP
				GTTAGAATCAATATACGACCCCCGCTGCTGTACATGTCCAGCCGGATCTC	AATACTGA	CACCATC	SPVPLTNDKKKDSTLP
				GTTGTAAAGAAGATTGCAGCTGTAAAAAAAGTTGGACTTCTTTGTTCTTCC	AGGCAATA	TATTCAT	TATSKNLSKNNNNK
				TGTGAAGAAGTTGATTGTGGCTGTTCAAATTCTTTTCATAATAAAGAATT	CCAAACTC	ATCCAC	NNNTNRINNIKNND
				AATGGCTCCACTTCCTTGGAATGCTGCGACATCTTCACCACCGTCACCTG	AACCCGAA	ACACTG	NTNDGSNKINLKLPP
				TCCCATTAACTAACGATAAGAAAAAAGGATTCTACTCTACCTACTGCAACA	ATCAAAAT	ACCACCT	AAVKITNPYKNKKKN
				TCAAAAAATTTATCCAAAAATAATAATAATAAAAATAATAACTAATAG	CGTTAGAA	CCACCTT	KKKNNAGKSNPKTN
				GATTAATAATATTAAAAATAATGATAATACAAATGATGGTTCTAATAAG	TCAATATA	CACAAC	QNPNSSPLSDNDDD
				ATAAATTTGAAACTGCCCCCGGCAGCTGTTAAAATCACAAATCCTTATAA	CGACCCCC	TCCACTC	DTDSSNITINRRLKFG
				GAACAAAAAGAAGAACAAGAAGAATAACGCTGGAAAGTCGAACC	GCTGCTGT	TCAATTC	TDDLAPPNPPSNTNT
				CCAAAACAAACCAAAATCCAAATTCAAGTCCACTTTCGGATAATGATGAT	ACATGTCC	CCCTGA	IGTATAATAATATTT
				GATGATACTGATAGCTCAAACATCACCATTAATAGACGACTAAAATTTG	AGCCGGAT	CTACTAA	ATAATATTATNTTTT
				GTACTGATGATTTAGCACCTCCAAACCCACCGTCAAATACTAATACTATT	CTCGTTGT	GAAATA	TTTTNNTTGDNLASN
				GGTACTGCTACTGCTGCCACCGCTGCCACTGCTACAACTACGGCAACCG	AAAGAAG	TTTCATG	INNNNNNNNSGSNN
				CAGCGACTGCTACTACTGCTACCAACACTACTACTACTACTACTACT	ATTGCAGC	GTGGTT	SNTNNINNTDGNGS
				AATAACACTACTGGAGACAATCTTGCAAGTAATATTAATAATAATAATAA	TGTAAAAA	ACATTG	NNRPPPRVYTVDPRS
				TAATAATAATAGTGGCAGTAACAATAGCAACACCAATAATATTAACAAT	AGTTGGAC	GAGGTA	DLPGAEISAANKMLD
				ACCGATGGCAATGGTAGTAATAATCGCCCACCTCCCCGTGTTTACACAG	TTCTTTGTT	TCCCAAC	EVYGDHVHDNPGSH
				TCGATCCACGAAGCGACCTTCCCGGTGCAGAAATCTCTGCTGCAAATAA	CTTCCTGT	ACCAAG	LSGLISSSQDQLWQG
				AATGTTAGATGAAGTATATGGAGACCATGTCCATGACAACCCCGGCTCC	GAAGAAG	CACAAA	YFRRLIPHNQSLYDCP
				CATCTCAGCGGACTTATTAGCAGCTCTCAAGATCAACTCTGGCAGGGTT	TTGATTGT	ATGAAC	KGKLGKDITNEYSNLF
				ACTTTCGTCGCCTTATACCTCACACCAATCTCTCTATGACTGCCCTAAGG	GGCTGTTC	CCACTA	EAIMNGKCNMEKLL
				GAAAACTGGGTAAGGATATAACGAATGAATTCAAACTTATTTGAGGC	AAATTCTT	ACCCTCT	VFPVVVLQRRHGVTK
				AATCATGAATGGCAAGTGTAATATGGAGAAGCTACTCGTGTTTCCAGTA	TTCATAAT	CATCCTA	NADVKRRLLSRLTAW
				GTGGTTCTACAACGGAGACACGGAGTGACTAAGAATGCCGACGTTAAA	AAAGAATT	TCCACG	KEGKFKYLVEDTHRD
				CGCCGTCTTCTTAGCCGACTCACCGCTTGGAAGGAAGGCAAATTCAAGT	A (SEQ ID	GGGACC	LIAKQSKARGDTTPA
				ATCTTGTTGAAGACACACATCGAGATCTTATTGCCAAACAATCCAAAGCA	NO: 1156)	ACCTTTG	HRAKVYSSKLMRGHL
				AGAGGAGATACAACCCCCGCGCACAGAGCTAAAGTTTACTCAAGTAAAC		AGCAGA	QSAVNYITDREGGGI
				TCATGCGTGGACATCTCCAATCAGCCGTCAACTACATCACTGACCGCGA		ACACCA	LYPYDVDEKSGHTVS
				AGGAGGGGGCATCCTTTATCCTTATGACGTCGATGAGAAATCAGGCCAT		TTCATAT	RVLQDKHPSMRDPG
				ACTGTATCAAGAGTGCTACAGGATAAGCATCCCAGCATGCGTGATCCTG		TACAAC	PTAMPAYESVPELPT
				GTCCCACAGCCATGCCTGCCTACGAGTCCGTCCCGGAACTTCCAACACTT		CTTTAGC	LEITADTVEIVAGKLS
				GAAATTACAGCTGATACAGTTGAGATAGTCGCTGGAAAGCTCAGTGGT		TAGATT	GGAGLSGVDSIQLKH
				GGTGCAGGTCTGAGTGGAGTTGATTCAATACAACTCAAGCACCTCCTCC		AAGATA	LLLHHGQASQRLRNV

		TCCATCACGGTCAAGCAAGCCAACGACTGCGCAATGTTTGTGCAAAATT		ATTATTT AGTACA TATTTTA TACTATT AAAAAA AAAAAA AAAA (SEQ ID NO: 1279)	CAKFGRWLANEHPP WASYRAMLANRLIAL DKMPGIRPVGIGDT WRRFFAKLVLAVSM SYATDCCGSDQLCAG LRAGVDGAIHGLSA MWREMESEENTGF VLIDADNAFNEVSRI NMLWTIRHEWPAG ARFAFNCYRHHSLLV VRNPGGKPFTFFSKE GVTQGDPFAMIAYG VALLPLIRKLKELNVLL VQSWYADDASAAGK FDEILRLFQDLLRMG PDFGYFPNASKSILIT HPDNVVAAHHFFNE THGLGFKISTGSRFLG GFIGDTTSRDEYVSTK IADWIHGTKELAAVA RLKYPHAAYTGITKCL QHKWSFTQRVIPGID DLFQPLEDELTNNLLP ALFGDPPSTMDDKLR LLTALPVKHAGLALP NPVTSSATNYKNSTL MSSHLLLAVQGKINF SLQDHRDTCQSSLSA SRELRQTENDSSLTNL LAALPPAAAGQPSTT RAIKRAGETGLWLTTI PNHINGNILGCDEFID AIRLRYQKVPHNLPA KCDGCGSAFDVGHA LQCKSGGLIIRRHDEL NLELASLAKMALRES AIRAEPEINPSASIMD SPTTITAIDTNGDRG
		TGGTAGATGGCTTGCCAACGAGCACCCCCCCTGGGCCTCGTACCGTGCC			
		ATGCTAGCAAATAGGCTTATTGCGCTAGACAAAATGCCCGGAATTCGAC			
		CAGTCGGTATAGGTGATACATGGCGTCGTTTCTTCGCCAAACTTGTTCTA			
		GCAGTCTCTATGTCTTATGCTACTGACTGTTGTGGGTCAGACCAGCTCTG			
		TGCCGGACTAAGAGCCGGAGTTGATGGTGCCATACATGGACTATCGGC			
		TATGTGGAGGGAGATGGAATCTGAGGAAAACACAGGTTTCGTACTTATT			
		GACGCAGACAATGCATTCAATGAGGTCTCACGCATCAACATGTTATGGA			
		CGATCCGCCACGAATGGCCTGCTGGAGCTCGATTCGCCTTCAACTGCTA			
		TCGGCACCACAGCCTACTAGTGGTACGGAATCCAGGCGGGAAACCCTTC			
		ACTTTCTTTTCTAAAGAAGGTGTCACACAGGGCGACCCATTTGCGATGAT			
		AGCATATGGTGTCGCTCTCCTACCACTCATCCGCAAACTGAAAGAATTAA			
		ATGTATTATTAGTTCAATCTTGGTATGCAGATGATGCTAGCGCAGCTGG			
		CAAATTTGATGAAATACTACGCCTTTTTCAAGATTTATTACGAATGGGAC			
		CTGATTTTGGGTACTTTCCTAATGCATCTAAGAGTATCCTCATCACCCATC			
		CCGACAATGTGGTTGCAGCTCACCACTTCTTCAACGAGACCCATGGCCT			
		AGGTTTCAAGATCAGCACACAGGAAGTCGTTTCCTGGGTGGTTTCATTGGA			
		GATACCACAAGTCGAGATGAATACGTATCAACAAAAATCGCCGACTGGA			
		TCCACGGCACCAAGGAGCTAGCAGCAGTAGCAAGATTGAAGTATCCAC			
		ACGCAGCTTACACAGGCATTACCAAGTGTTTGCAGCACAAGTGGAGTTT			
		TACTCAACGTGTTATTCCTGGCATTGATGACCTCTTCCAACCACTGGAGG			
		ATGAACTCACCAATAATTTGCTCCCCGCCCTATTTGGAGACCCCCCATCC			
		ACTATGGATGACAAGCTCAGACTTTTGACCGCTCTGCCAGTCAAACATG			
		CTGGGCTTGCTCTCCCGAATCCAGTTACCTCCTCCGCAACCAACTACAAG			
		AATAGCACTCTTATGAGTTCTCATCTTCTGTTGGCTGTTCAAGGCAAGAT			
		CAACTTCAGTTTACAGGACCACACAGAGATACCTGTCAATCCTCTCTCTCCG			
		CGTCCCGAGAGCTCCGACAAACCGAAAATGATTCTTCATTGACCAACCTC			
		CTTGCAGCTCTCCCTCCAGCTGCTGCAGGTCAACCAAGCACACAACAAGAG			
		CAATCAAGCGTGCTGGGGAAACCGGTCTTTGGCTTACTACTATCCCTAAT			
		CACATCAACGGTAACATTCTCGGATGTGACGAATTTATTGATGCTATTCG			
		ATTGAGATACCAAAAAGTGCCACACAATCTCCCTGCCAAATGTGATGGC			
		TGTGGCTCTGCATTTGATGTAGGGCACGCGCTCCAATGCAAATCCGGGG			
		GCCTAATCATTAGACGTCATGATGAACTCAATCTTGAGCTTGCATCTTTA			
		GCAAAGATGGCCTTGAGAGAATCTGCAATACGTGCTGAACCTGAAATCA			
		ACCCCAGCGCCTCTATTATGGATTCTCCCACCACCATCACAGCCATCGAC			
		ACAAACGGAGACCGAGGAGATTTGTTGATCAAGGGCTTTTGGGACAAT			
		GGAATGGACGCTATCATCGATGTCAGAATAACAGACACAGATGCCAAAT			
		CCTATCGAACAAGAGACCCAAAAAAAGTCCTACAGTCACAAGAGAAGG			

			AGAAAAAGAAGAAATACCTCGATCAATGTCTACTCCAACGTCGAGCCTT				DLLIKGFWDNGMDA
			TACCCCTTTTGTTGTCTCTGTGGACGGCCTGATTGGTTACGAGGCCAGCA				IIDVRITDTDAKSYRT
			ATGTGCTAAAGCAATTATCAAAACGTTTAGCAGATAAATGGAATAAGCC				RDPKKVLQSQEKEKK
			TTATTCAGTTACATGTGGAATAGTCCGCTCACGTATCAGCATTGCATGTG				KKYLDQCLLQRRAFT
			CGCGAGCTTCCAATCAATGTCTGAGAGGTTCTCGAATACCATTCAAAAC				PFVVSVDGLIGYEAS
			AATGAGCAGACAAATTCAATGGGAGGACGGTGCAGGCCGCCGGCCTCTA				NVLKQLSKRLADKW
			TAGAATTGTCCGCTAGCACCACCATCTATTCATATCCACACACTGACCAC				NKPYSVTCGIVRSRISI
			CTCCACCTTCACAACTCCACTCTCAATTCCCCTGACTACTAAGAAATATTT				ACARASNQCLRGSRI
			CATGGTGGTTACATTGGAGGTATCCCAACACCAAGCACAAAATGAACCC				PFKTMSRQIQWEDG
			ACTAACCCTCTCATCCTATCCACGGGGACCACCTTTGAGCAGAACACCAT				AGAGLYRIVR (SEQ
			TCATATTACAACCTTTAGCTAGATTAAGATAATTATTTAGTACATATTTTA				ID NO: 1401)
			TACTATTAAAAAAAAAAAAAAAA (SEQ ID NO: 1033)				
CRE	Cre-1_HM	.	Hydra vulgaris	TTTCTAATGTTACGTGATATGATATGGTTAGTTCATGGTTAGTTTATGTTT	TTTCTAAT	TAACTTG	MNMVSICKRCDRSF
				ATGCTTAGTTTATGGAAAATCGTTTATTTATGGCACAATATTGTTTGCTG	GTTACGTG	TATTTTT	TTLKGLNIHKGQCKIF
				TTTTTAAATTTATGTAACGTGTGCATTTGATGTATATTCTTGAACTTTTTA	ATATGATA	AAATTG	VSNTNKQINNVVNN
				ATCTGAATTTTTACTTGGTTTAATACGTTTATTATATTCTTCGATTGAGCA	TGGTTAGT	TTTTATT	ELTTPNNKVEINTIL
				ATTTTATCCTATCAAAGCAATTTATCCTTCGATTCGAGCAATTTATCCTTCG	TCATGGTT	AGTTT	NCDEISVEHYSTNTPY
				ATTCGAGCAATTTATCCTTCGATTGAGCAATTTATCCTATCAAAATTAGC	AGTTTATG	(SEQ ID	LPKINICESIIDPNDYL
				ATATATACTGCAATTTTCAAATAATCTACGAAATAAGTTCACTTACTGAA	TTTATGCT	NO:	WGHMPFSFLLNHVN
				AATCATTAAGTAAAAGAAGAAAGGAAGAAAAATAAAAATAAAAAGTA	TAGTTTAT	1280)	TIYDEIVFYHKNLFKV
				GTAAATCCTTTCATAACAATAATCATTCTATTATTAAATTTAAAGGAATAT	GGAAAATC		PSGKGGKMFIEELTF
				TTTGGTTTTGTACTAAATCATGCGTTCATATTTCACCGAAGAAGGGGGCT	GTTTATTT		WLKQFNNRTKLNGI
				GCTATATTTTTGTTTGAAGTTGTTTATCTTAAAACTTTAAACTTGTGTTCA	ATGGCACA		AMKCFMIVPSLMLQ
				ACCAACCGTAAACATTAGTTCGCTGTTCGCTCAAATTATCTACAATATAA	ATATTGTT		KPSIRSKAKEHAECLV
				AATTTATCAATCTTTTTTCGTTACGGTAAACAATAAACAATAAAATAACT	TGCTGTTT		RRITLWRNGNFSELM
				ATAGTTATTTTATTGTTTACCGCATATTGTTTAACTATAGTTAAACAAAGT	TTAAATTT		REIRYIQSKINTSKKKR
				ATTTGTTTATGGAACATTACCAGTATCTCTTGTTAAGGTAAACAACAAAA	ATGTAACG		TFEDISRIFAKLMME
				CATAGACGGCATCTCTTTTTAAGGTAATTAAGTATACGGCTAATAATAAA	TGTGCATT		GKVAAALKVLDRESS
				AATATACAGCTAATAATAAAATCTTCAATGAACATGGTTTCTATATGCAA	TGATGTAT		GILQCSESVLKELKSK
				AAGATGTGATCGTAGCTTTACTACCCTTAAGGGACTAAATATTCATAAA	ATTCTTGA		HPDETPVQDNCLLYG
				GGTCAATGTAAGATCTTTGTTTCCAATACAAATAAACAAATAAACAATGT	ACTTTTTA		PLQNTPECLFDSIDEI
				AGTTAACAATGAATTAACAACACCGAATAAAAACAAGGTGGAAATTAAT	ATCTGAAT		SIFNSALQTKGSAGPS
				ACGATATTAAACTGCGATGAGATATCTGTAGAACACTATTCAACCAACA	TTTTACTT		GMDADLYRRVLCSK
				CACCTTACTTACCCAAAATAAATATTTGTGAATCTATTATAGATCCCAAC	GGTTTAAT		CFGPSCKTLREEIATF
				GACTATCTATGGGGTCATATGCCGTTTAGCTTCCTTCTCAACCATGTCAA	ACGTTTAT		TKNIATKSYQPDIVQP
				CACAATATACGATGAAATAGTATTTTACCATAAAAACCTTTTTAAAGTGC	TATATTCTT		YIACRLIPLDKNPGIRP
				CATCAGGAAAAGGTGGTAAAATGTTTATAGAAGAACTGACCTTTTGGCT	CGATTGAG		IGIGEVLRRIVGKTISH
				AAAACAGTTTAATAATCGAACCAAATTGAATGGAATAGCCATGAAATGT	CAATTTAT		HCQKEIKEAAGPLQT

		TTCATGATAGTCCCTTCCCTAATGTTACAGAAGCCCTCAATACGGTCCAA	CCTATCAA		CAGHGAGAEAAIHA
		AGCCAAAGAACATGCAGAATGTTTAGTAAGACGAATTACATTATGGAGA	AGCAATTT		MQKIFHQEDTDGVL
		AACGGGAACTTTAGTGAATTGATGCGGGAAATTAGATATATTCAGAGCA	ATCCTTCG		LIDARNAFNCLNRSV
		AAATTAACACCTCAAAAAAGAAAAGGACATTTGAGGATATCTCAAGGAT	ATTCGAGC		ALHNIQITCPILAMYL
		ATTCGCAAAACTAATGATGGAAGGTAAAGTTGCTGCCGCACTGAAGGTT	AATTTATC		VNTYRKPAKLFIYGGE
		TTAGATAGAGAGTCATCTGGCATCTTGCAATGCTCGGAAAGTGTATTGA	CTTCGATT		TIFSKEGTTQGDPLA
		AAGAATTGAAAAGTAAACACCCAGACGAAACTCCTGTACAAGATAATTG	CGAGCAAT		MPWYSLSTVTIINTLK
		TTTACTATACGGCCCGTTACAAAACACTCCAGAATGTTTATTCGATTCAA	TTATCCTTC		LVIPDVKQVWLADD
		TTGATGAGATAAGTATATTTAACTCAGCTTTACAGACTAAAGGATCTGCA	GATTGAGC		ATAAGKLQSLKKWYK
		GGTCCTTCTGGAATGGATGCAGATCTTTACCGTCGAGTCCTATGCTCAAA	AATTTATC		CLEDVGGLYGYYVNQ
		ATGTTTTGGACCCTCTTGTAAGACTCTACGAGAAGAAATAGCAACATTTA	CTATCAAA		SKCWLIVKSDNQAEE
		CAAAAAATATTGCAACAAAATCCTACCAACCGGATATAGTTCAACCCTAC	ATTAGCAT		AKLIFGNSINITTQGK
		ATTGCATGTCGACTAATTCCCTTAGACAAAAAATCCCGGGATTCGCCCCAT	ATATACTG		RHLGAALGSEAYKKV
		AGGAATTGGGGAAGTGTTACGTAGGATTGTAGGTAAAACCATTAGCCA	CAATTTTC		YCEDLVSKWSKELNN
		CCATTGTCAAAAAGAAATCAAAGAGGCAGCTGGACCACTACAAACTTGC	AAATAATC		LCEIATTQPQAAYSAF
		GCAGGACACGGTGCAGGAGCAGAAGCTGCAATACATGCTATGCAAAAG	TACGAAAT		IKGYRSKFTYFLRTIEA
		ATATTTCATCAGGAAGATACAGATGGTGTTTTGTTAATCGATGCTAGGA	AAGTTCAC		FENFVTPVEKILSEKLL
		ACGCGTTTAACTGCCTAAACCGTTCTGTTGCACTACATAATATACAGATA	TTACTGAA		PVLFGTDCSIIKENRD
		ACTTGCCCAATCTTAGCTATGTATTTAGTCAACACTTACCGTAAACCGGC	AATCATTA		LLALNPSEGGLGICNL
		AAAATTATTCATCTACGGTGGAGAAACTATTTTTTCGAAAGAAGGCACA	AGTAAAAG		ITEAKEQHTASKKITN
		ACGCAGGGCGATCCCCTCGCCATGCCATGGTACTCACTTAGCACTGTGA	AAGAAAG		LHIKSILDQSDVMKEK
		CAATCATAAATACATTGAAACTAGTAATTCCTGATGTAAAACAAGTATG	GAAGAAA		DDFGKTFSEIKTKTN
		GTTAGCCGATGATGCTACCGCTGCAGGAAAATTACAGTCTTTAAAAAG	AAATAAAA		MDKSKKKKEEVKKIH
		TGGTATAAATGCCTAGAGGATGTCGGTGGTTTGTATGGTTATTATGTAA	ATAAAAAG		AGLPENLKLLVEQAC
		ATCAGTCAAAATGCTGGCTAATAGTAAAATCTGATAACCAAGCTGAAGA	TAGTAAAT		DKGASSWLNTLPIKE
		AGCTAAACTTATATTTGGCAACTCCATAAATATAACTACTCAGGGAAAAA	CCTTTCAT		QHLDLNKEEFKDALR
		GGCACTTAGGAGCTGCACTTGGTTCGGAAGCATACAAAAAAAGTGTATTG	AACAATAA		LRYNVPLANLPSYCA
		CGAGGATTTAGTAAGTAAATGGTCTAAAGAACTTAACAATCTCTGCGAA	TCATTCTA		CGEKFDELHAMSCKK
		ATCGCCACCACGCAACCACAAGCTGCTTATTCAGCTTTTATTAAAGGGTA	TTATTAAA		GGFVCNRHDNIRDLL
		CAGATCTAAATTCACTTACTTCTTACGCACAATTGAAGCTTTTGAAAATTT	TTTAAAGG		TVCLNKVCTDVQAEP
		CGTAACACCAGTGGAAAAAATTTTATCAGAAAAATTATTACCTGTATTGT	AATATTTT		HLIPLTNEKFNFKTAN
		TTGGAACTGATTGTTCTATAATCAAAGAAAATAGGGATTTATTGGCGCT	GGTTTTGT		TNDEARLDIKAKGFW
		AAATCCATCGGAAGGAGGACTTGGAATTTGTAACTTAATAACTGAGGCC	ACTAAATC		RKGETAFFDVRVTHV
		AAGGAACAGCATACTGCCTCTAAGAAAATAACTAACTTGCACATAAAAT	ATGCGTTC		NSKSSKKQPTKHIFRR
		CAATACTCGATCAGTCAGATGTTATGAAAGAAAAAGATGATTTCGGGAA	ATATTTCA		HEDAKKREYLERVLE
		AACATTTTCAGAAATAAAAACAAAAACAAATATGGATAAATCTAAAAAA	CCGAAGAA		VEHGTFTPLIFGTNG
		AAAAAAGAAGAGGTTAAAAAAATACATGCAGGACTTCCAGAAAACCTT	GGGGGCT		GFGDECKRFTALLAQ
		AAACTTCTGGTTGAACAGGCCTGTGACAAAGGTGCCAGCAGCTGGTTAA	GCTATATT		KLSLKMGERYGAVIN

		ACACCTTACCAATTAAAGAACAACATCTAGATCTGAATAAGGAAGAGTT	TTTGTTTG		WLRTRLSMEITRASLL
		TAAGGACGCACTTAGATTGAGATATAATAATGTGCCACTTGCCAATTTACCAT	AAGTTGTT		CLRGSRTPFRHYNTD
		CCTACTGTGCTTGTGGAGAAAAATTTGACGAGCTACACGCAATGTCATG	TATCTTAA		DVGLENVQCGLI
		CAAAAAAGGTGGCTTTGTTTGTAACAGACATGATAACATCAGAGATTTA	AACTTTAA		(SEQ ID NO: 1402)
		TTAACTGTTTGCCTAAATAAAGTTTGTACTGATGTTCAAGCGGAGCCGCA	ACTTGTGT		
		TTTAATTCCATTGACAAATGAAAAATTTAATTTCAAAACTGCCAATACCA	TCAACCAA		
		ACGACGAAGCTAGATTGGATATAAAAGCAAAAGGGTTTTGGAGAAAAG	CCGTAAAC		
		GAGAAACTGCATTTTTTGATGTTAGAGTAACGCACGTAAACTCCAAATC	ATTAGTTC		
		CTCCAAAAAACAACCAACAAAACACATATTCCGTAGGCATGAAGATGCA	GCTGTTCG		
		AAAAAACGTGAGTATTTAGAACGAGTTCTAGAGGTTGAACACGGGACA	CTCAAATT		
		TTTACCCCATTAATTTTTGGTACGAATGGTGGGTTTGGAGACGAATGCA	ATCTACAA		
		AACGCTTCACGGCACTACTCGCACAAAAACTGTCCTTAAAAATGGGTGA	TATAAAAT		
		GCGGTACGGAGCTGTTATAAATTGGCTAAGGACACGTCTTTCCATGGAG	TTATCAAT		
		ATTACTAGAGCCTCCCTACTCTGCTTAAGAGGGTCACGAACCCCATTTAG	CTTTTTCG		
		GCATTATAACACTGACGATGTTGGCCTGGAAAATGTGCAATGTGGACTT	TTACGGTA		
		ATTTAACTTGTATTTTTAAATTGTTTTATTAGTTT (SEQ ID NO: 1034)	AACAATAA		
			ACAATAAA		
			ATAACTAT		
			AGTTATTT		
			TATTGTTT		
			ACCGCATA		
			TTGTTTAA		
			CTATAGTT		
			AAACAAAG		
			TATTTGTTT		
			ATGGAACA		
			TTACCAGT		
			ATCTCTTG		
			TTAAGGTA		
			AACAACAA		
			AACATAGA		
			CGGCATCT		
			CTTTTTAA		
			GGTAATTA		
			AGTATACG		
			GCTAATAA		
			TAAAAATA		
			TACAGCTA		

				ATAATAAA ATCTTCA (SEQ ID NO: 1157)		
CRE	CRE-1_LSa	.	Lactuca sativa	ACATTAAATTAGAGAGGTTGATGTTTCAATGGAAGAAGATGAAATTCCA	ACATTAAA	TGAACT	MASSSTSSSDICLCPF
				AGAAGCTATTTTTGTTGCCCACCAAGTGTTTGATAAAATGTCCAAACTAA	TTAGAGAG	ATATTTT	RSFHCCPNGEVGSKG
				TTTTTCTCTTGTTGCAGCTTTATTGTTCAAGATAATGTAGTTTGCTTAGTT	GTTGATGT	ATATATT	IXRMISHIKRHHLLTE
				TGAGCGTTCCTTGTGCACACCAACAGTGTGTTGGTGTGCCATTTCCTTTC	TTCAATGG	AAAAAA	DRKCVLREALSSDVG
				CTTCCTTTTAACTATTGCTTCATAGCTTAAGCTTCATCTCGAGGCTTGTT	AAGAAGAT	A (SEQ	LFMAVEETLKAFGQ
				CTCTTGTATGGCTTCTTCTTCTACAAGTTCGAGTGATATTTGTCTGTGCCC	GAAATTCC	ID NO:	WMCGKCMTLHALS
				GTTCAGAAGCTTCCATTGTTGCCCAAATGGTGAAGTGGGAAGTAAGGG	AAGAAGCT	1281)	RYCHHPDGRVXFVT
				GATTGKCCGTATGATTTCACACATCAAAAGGCATCATCTACTTACTGAAG	ATTTTTGTT		GADGSSRYIVGILKPS
				ATCGTAAATGTGTTTTACGTGAAGCTCTTTCTAGTGATGTTGGTTTATTT	GCCCACCA		TKESVTNALGGLVFD
				ATGGCGGTGGAAGAAACTTTGAAGGCCTTTGGTCAATGGATGTGTGGG	AGTGTTTG		VGLLDRVFKEPITTVK
				AAGTGTATGACTTTGCATGCTCTTAGCCGTTATTGTCATCACCCGGATGG	ATAAAATG		SIPHSCRLAFSQALKT
				TCGTGTGAGKTTTGTTACAGGGGCTGACGGCTCGAGTCGTTACATTGTC	TCCAAACT		ALYKVIAQPGSVDA
				GGTATTCTAAAGCCGTCTACTAAAGAGTCGGTGACAAATGCTCTTGGAG	AATTTTTCT		WICLLLLPRCTLQVFR
				GTTTGGTTTTTGATGTTGGGCTCCTTGATCGTGTGTTTTTAAAGAGCCTATC	CTTGTTGC		PKNRQECRSGNRKSL
				ACTACTGTCAAGAGTATCCCCCATAGTTGTCGCCTTGCTTTCTCTCAGGC	AGCTTTAT		QQSSILKSLDTWGKE
				TTTGAAAACTGCTCTTTACAAGGTGATTGCCCAACCTGGCTCGGTTGATG	TGTTCAAG		DGIRKLVQNMLDNP
				CATGGATTTGTTTGTTACTTCTTCCTCGCTGCACACTGCAGGTGTTTAGG	ATAATGTA		EVGAMGQGGGILQK
				CCCAAAAATAGACAAGAATGTAGGTCTGGGAATAGAAAATCCTTACAAC	GTTTGCTT		ESTSSNTNIRQCLRKV
				AAAGCTCCATCCTGAAGTCCTTGGATACATGGGGGAAAGAGGATGGTA	AGTTTGAG		ADGHFTAAVKVLCSS
				TCAGGAAGTTAGTTCAAAATATGTTAGACAATCCCGAGGTTGGGGCCAT	CGTTCCTT		GVAPYNGDTIKALED
				GGGACAGGGTGGAGGCATCCTTCAGAAGGAGTCTACATCAAGTAACAC	GTGCACAC		KHPFRPPPSMPSPIIS
				CAACATCAGGCAGTGTCTCCGTAAGGTTGCAGATGGKCATTTTACCGCA	CAACAGTG		EPPLVADFDCVFGCIK
				GCAGTGAAAGTGTTATGCTCATCGGGTGTTGCGCCATATAATGGTGATA	TGTTGGTG		SFPKGTSCGRDGLRA
				CTATTAAAGCTTTGGAGGACAAAACACCCTTTCAGGCCACCCCCATCCATG	TGCCATTT		QHXLDALCGEGSAIA
				CCGAGCCCCATAATTTCTGAACCTCCCCTTGTAGCAGACTTTGACTGTGT	CCTTTCCTT		TDLIRAITSVVNLWLA
				ATTTGGTTGCATCAAATCCTTCCCTAAAGGAACTTCWTGCGGGAGAGAT	CCTTTTTA		GRCPTILAEFVASAPL
				GGCTTGAGGGCTCAACACWTACTAGATGCCCTTTGTGGAGAAGGGTCT	ACTATTGC		TPLIKPDNGIRPIAVG
				GCTATAGCCACAGATCTCATACGTGCTATCACTTCAGTGGTTAATTTATG	TTCATAGC		TIWRRLVSKVAMKG
				GTTAGCGGGAAGATGTCCGACCATTTTGGCAGAGTTTGTTGCATCCGCT	TTAAGCTT		VGKEMAKYLNDFQF
				CCTCTCACGCCTCTGATTAAACCTGACAACGGGATCCGTCCAATTGCAGT	CATCTCGA		GVGVSGGAEVVLHS
				AGGCACTATATGGAGACGTCTGGTTTCCAAGGTTGCCATGAAAGGTGTG	GGCTTGTT		ANRVLSEHHADGSLA
				GGTAAAGAAATGGCCAAGTACCTTAATGATTTTCAGTTCGGGGTTGGTG	CTCTTGT		MLTVDFSNAFNLVD
				TGTCCGGGGGTGCTGAGGTTGTGTTACACAGTGCCAATAGGGTGTTGA	(SEQ ID		RSALLHEVKRMCPSIS
				GTGAACACCACGCTGATGGGTCTCTTGCAATGCTGACAGTGGATTTCTC	NO: 1158)		LWVNFLYGQAARLYI

		GAATGCCTTTAACCTGGTGGATAGATCAGCCTTGCTCCACGAGGTTAAG			GDQHIWSATGVQQ
		AGGATGTGCCCTTCTATTTCTTTGTGGGTGAATTTCTTGTACGGGCAAGC			GDPLGPLLFALVLHPL
		AGCGAGACTTTATATAGGAGACCAACATATATGGTCTGCCACTGGGGTG			VHKIRDNCKLLLHAW
		CAGCAAGGCGACCCCTTGGGCCCTCTTCTTTTTGCCCTCGTTTTGCACCC			YLDDGTVIGDSEEVA
		GCTTGTGCACAAGATTAGAGACAATTGTAAGCTCCTTCTCCATGCTTGGT			RVLNIIRVNGPGLGLE
		ATCTAGATGATGGGACTGTCATTGGGGATTCAGAGGAGGTGGCTAGAG			LNIKKTEIFWPSCDGR
		TGTTGAACATTATTCGGGTGAATGGTCCAGGCTTGGGTCTTGAGTTGAA			KLRADLFPTDIGRPSL
		CATCAAGAAAACGGAGATTTTTTGGCCCTCCTGTGATGGTAGGAAGCTT			GVKLLGGAVSRDAGF
		CGTGCCGATTTATTCCCAACGGATATAGGGAGACCTTCTTTGGGGGTGA			ISGLAMKRAVNAVDL
		AGCTCCTTGGGGGGGCTGTTAGCAGAGACGCAGGGTTTATTAGCGGGC			MGLLPQLCDPQSELL
		TGGCCATGAAGAGAGCGGTCAATGCTGTTGATTTGATGGGTCTTCTTCC			LLRSCMGIAKLFFGLR
		ACAACTATGTGACCCGCAGAGTGAGCTCCTTTTGCTTCGATCATGTATGG			TCQPVHIEEAALFFDK
		GCATTGCAAAACTTTTCTTTGGTTTAAGGACATGCCAGCCGGTGCACATA			GLRRSIEDMVVCGGP
		GAAGAGGCAGCTTTGTTCTTTGACAAAGGATTGCGCAGGTCTATCGAGG			FFGDIQWRLASLPIRF
		ATATGGTGGTATGTGGAGGCCCCTTCTTTGGAGACATCCAGTGGCGTCT			GGLGLYSAYEVSSYAF
		GGCTTCCTTACCTATTCGTTTCGGTGGTTTGGGTTTGTACTCGGCATACG			VASRAQSWALQDHI
		AGGTTTCCTCCTACGCATTTGTAGCCTCGAGGGCCCAATCTTGGGCATTA			LRDSGICGMDSDYLC
		CAAGACCACATCTTACGTGACAGTGGCATATGTGGTATGGACTCTGATT			AMTRLRDTIPGFDCS
		ACCTATGTGCTATGACTCGTCTTCGCGATACGATTCCGGGATTCGACTGT			GFTNKDTPPKSQKAL
		AGCGGTTTCACTAATAAGGACACCCCCCCTAAATCCCAAAAAGCATTGG			ACALFSKIVKDMEVD
		CGTGTGCCCTTTTTAGCAAAATCGTCAAAGATATGGAAGTCGACTTCGA			FDMTVRQKAVFECL
		CATGACTGTTAGACAGAAAGCAGTTTTTGAGTGTCTTCGGGCACCTCAT			RAPHAQDFLLTIPIDG
		GCTCAGGATTTTTCTGCTAACTATCCCTATTGATGGCCTTGGCCAGCATAT			LGQHMSPVEYRTILR
		GTCTCCTGTGGAGTACCGAACTATCCTTCGTTACCGCCTCATGATTCCTCT			YRLMIPLFPIDEICPVC
		ATTCCCAATTGACGAGATATGCCCAGTTTGCCGCAAGGCATGTTTGGAT			RKACLDTFGEHAVHC
		ACCTTTGGGGAACATGCGGTTCATTGTAGAGAGCTCCCTGGTTTCAAGT			RELPGFKYRHDVVRD
		ACAGACATGATGTGGTTAGGGATGTTCTCTTTGATGCTTGTCGGCGTGC			VLFDACRRAGISAKKE
		TGGTATTTCTGCGAAGAAAGAAGCGCCAGTGAACTTTTTGACGGACCCG			APVNFLTDPQDGRST
		CAGGATGGAAGATCCACACTTAGACCGGCTGACATTTTGGTCTTTGGAT			LRPADILVFGWVGGK
		GGGTAGGAGGGAAGCACGCGTGTGTGGATCTTACTGGGGTCTCTCCTC			HACVDLTGVSPLVGL
		TCGTCGGTTTGAGGAGCGGGGGTTTCACAGCAGGGCATGCCGCTTTGA			RSGGFTAGHAALKA
		AAGCCGCTGCGTGCAAAGTGGCAAAGCACGAGAATGCATGTATAGAAA			AACKVAKHENACIEN
		ATCAACATGTGTTTGTACCTTTTGCATTTGATACATTTGGTTTTCTCGCAC			QHVFVPFAFDTFGFL
		CAGAGGCGGTGGAGCTCCTCAACAGAGTCCAACGGGTCATGCATTCTA			APEAVELLNRVQRV
		ATGTCATATCTCCTAGATCCACAGATGTTGTTTTCAAAAGAATTAGTTTT			MHSNVISPRSTDVVF
		GCCATCCAGAAAGGGCTAGCGGCGCAGCTTGTTGCCCGTTTGCCTTCCA			KRISFAIQKGLAAQLV
		TCGATATGTATTGAACTATATTTTATATATTAAAAAAA (SEQ ID NO: 1035)			ARLPSIDMY (SEQ ID NO: 1403)

CRE	Cre-1_MB	.	Monosiga brevicollis	CATCTTGGCGTGAACCACGTTGTCAGACAAAATCTGCAACCCCGCTCTTT	CATCTTGG	TAGGTA	MATESGGEDSWTQ
				GCGGCCCGCGTTTTGGCGGCGCCCTCGCTCCCACCGTGTCCGCTCGCTT	CGTGAACC	GGCACC	VRGAKRPSAESPPSN
				GCTCGCTTGCTTGCCCCGCGGACATGGCCACTGAGTCCGGCGGCGGAGG	ACGTTGTC	GTCTCG	TTTSPSQTHRSAKHT
				ATTCTTGGACCCAGGTCCGCGGTGCTAAACGCCCGAGTGCCGAATCACC	AGACAAAA	GGGGTC	KHGSARHDRNHVFP
				TCCAAGCAACACCACCACCTCGCCTTCCCAAACTCATCGTTCTGCAAAAC	TCTGCAAC	CCTCTGT	DPMTTPLRPHARHS
				ACACAAAACATGGCAGCGCTCGCCACGACCGTAACCATGTTTTCCCTGA	CCCGCTCT	GGGGAT	VPTARASSHVPSTSP
				CCCCATGACCACCCCGCTTCGCCCTCACGCCCGCCACTCTGTCCCTACCG	TTGCGGCC	CCCTGT	AAGATESSARAVVPA
				CCCGTGCCTCGTCTCATGTGCCCTCGACGTCCCCCGCTGCCGGTGCGACC	CGCGTTTT	GTGCAC	AEPVTRTSNGGGEQ
				GAGTCTTCGGCACGTGCCGTCGTGCCCCGCGGCCGAGCCAGTGACCAGG	GGCGGCG	CTGTCG	HPIIGNTSNASPRTPR
				ACGTCAAACGGCGGCGGGGAGCAACATCCCATCATCGGAAACACTTCC	CCCTCGCT	CTCCCTA	TPSSPRSFAQVAAA
				AATGCTTCTCCCCGCACCCCTCGCACGCCATCCAGCCCTCGCTCCTTTGCT	CCCACCGT	GGTGGT	MPAAATATSSAPMT
				CAAGTTGCTGCGGCAATGCCTGCCGCCGCCACTGCCACATCTTCGGCCC	GTCCGCTC	TCCTCGT	EDLSASVPSEPNGSG
				CTATGACCGAGGATTTGTCAGCATCGGTGCCCTCTGAGCCAAATGGCAG	GCTTGCTC	TGTGTCT	EQQPSPESTGQTHHS
				CGGGGAGCAACAACCCTCGCCCGAGTCCACAGGGCAGACACATCATTC	GCTTGCTT	TTTGATG	IPNTPSDFLTMSSDES
				GATTCCTAACACACCATCGGATTTTTTGACCATGTCTTCGGATGAAAGCG	GCCCCGCG	GCTTGA	DSPPRSTALRAPTPIA
				ACTCCCCTCCTCGCTCCACCGCACTCCGCGCCCACCCCTATCGCCCCCTC	GAC (SEQ	CTTGTAT	PPAHDGDGDTNGSA
				CCGCGCATGATGGTGACGGTGACACAAACGGCAGTGCCACGCCTGAGC	ID NO:	TTTTGTT	TPEPLVQSPTPAQM
				CATTGGTGCAATCACCTACACCCGCTCAAATGGTGCTGCCATATCCATCG	1159)	TTAATTT	VLPYPSGTQQTHSDP
				GGTACACAACAAACCCATTCCGATCCCTCTCCGCCCTCTGCTTCACCCCT		TGCTTTA	SPPSASPPATTILPAAI
				GCCACTACCATTTTGCCCGCTGCCATTTCACATCCTGTCGAACACAGTGA		ATTTTTG	SHPVEHSEHANSAPL
				GCATGCAAACTCAGCCCCACTTGGCGAAGTCAGTGAGAGTGAAACACA		CTGTATT	GEVSESETHNTAGEH
				CAATACAGCGGGCGAACACAGTGAGAGTGAGCAAGATGTTCTTCTCAG		TGTGTG	SESEQDVLLSDPAPPI
				CGATCCCGCTCCGCCCATCGCTGCCAACGTGCTGGATGCCCAGCGCAAG		GTATTTT	AANVLDAQRKVLLKT
				GTCCTGCTGAAGACATCTGGCCACAGGCAACTCCTCGCCTGCCCATTTG		TGCTGA	SGHRQLLACPFGLCK
				GGCTTTGCAAATGCAAGGGGCCCCGCCTTGACCGCAAAGCCTGGGTCA		ATTTTTG	CKGPRLDRKAWVNH
				ATCATGTACTACGCGAGCACCCCTACGACGAGCAAGCCACTGATCTGGT		TAAGGT	VLREHPYDEQATDLV
				CAAGCAGGTGATGGAGGCCAAATTGGTCGCCCAGTGCAACAAGTGTCA		CCTTTGT	KQVMEAKLVAQCNK
				CCTCTTCTTCGAAGCTGCTGGTATCAGTCAACACCGCTCCCGATGTGGTG		ATGATG	CHLFFEAAGISQHRS
				CCAATCTGAAGCGAGCGACCGAGGCGTTGTTTCATGCGGCTGGACACG		TCTTTGT	RCGANLKRATEALFH
				ACCTGCTTGAGATTATGCGTGGCGCTTGGCCCCAACAGTGTGTAGGGTC		CTTCTGT	AAGHDLLEIMRGAW
				TCGCATCAGTGTCTGCGAGCTGCTCAAGCTCGCCCGGCATCCACTGATG		GGTCGG	PQQCVGSRISVCELLK
				CAGCGCAGCCGCTACCCATCCAACGCCACCGAGACCAAGCTGATGGCTG		TTGTTTT	LARHPLMQRSRYPSN
				CCACCCTGAGCCAGCTGTATTGGTCTGCCGTCCACTCGGACTATACCGCT		CCTCAAT	ATETKLMAATLSQLY
				GAAGAGCGAGAGATGTGCTGGGCTTTGATTTTGGCCTTGCCTAGCATGT		CCGACG	WSAVHSDYTAEERE
				TGTTGTCTGCTCCCTCGACCGCACTGTCTACGATTGACCTGCGCAATATG		TTGTGTC	MCWALILALPSMLLS
				TTTCACGATCGTCTCCGTTGGCTTGTGACGGGGCAACTAGGTCGGGTCG		TCGTTG	APSTALSTIDLRNMF
				TGGACGCCATGCGCAAGGCAGTCGCACGCAAGCAGAGCCGTCGAGGAC		GATGTG	HDRLRWLVTGQLGR
				AGCTGAACGCCGGCGCGGGCGCCCACCCGAACGACGCAGTCGACCAGA		AGCGTG	VVDAMRKAVARKQS

				GCCTCAGGTCGCTCGTCCGCGACCCGGACCTGGCGGACGAAGCCTGGG		CCGTGG	RRGQLNAGAGAHPN
				CAAACCACGTCACGAACCGTCTGAACCGAGGTCAGATTGCCAAAGCATT		TGTTCTT	DAVDQSLRSLVRDPD
				TGATGCCGACAAGGCTCGTGCCGTGATTGGTAATTCTGAGGTTCAGGCC		TGTGTTT	LADEAWANHVTNRL
				GTGCGCGACCTCTTGGTACCGCCCGGGCTGACCCCGTACATTGCTTCGA		GTGCTG	NRGQIAKAFDADKA
				CACCCGCCTCCACGTCTACACTGGCACCAGCCACGGCTGTGAGCTCCCC		TGATGG	RAVIGNSEVQAVRDL
				AACCGTGTCCTTCACCAAGGGTGAGCTCCCCAAGGCGTTGGCGGCCACC		CTTGTA	LVPPGLTPYIASTPAS
				AAGGGTGTCACCGACCCCTATGGTTGGTCTGGTGAGCTTCTTGCCTCCAT		GTTGTG	TSTLAPATAVSSPTVS
				CTACCGCATCAAGGAGCACTTCAGTCAAGTCTTGGGCCCACGCCAGGGT		ATGTGT	FTKGELPKALAATKG
				TCTACCAGCGACCCGACTGCTCCTTCTGATGGAGACGCGCCTCAGGGCC		GACTGC	VTDPYGWSGELLASI
				CCACCACCGCCACTGGAGGTCCTCAGGTTGCCTTGAACAAGATCTTTCAC		CTTTTTG	YRIKEHFSQVLGPRQ
				CACATTGCCAACAACACCGTGCCCGAGTCGATTCGACATGCCCTTTGCTC		GGTGTC	GSTSDPTAPSDGDAP
				CATCAACTACACTATCCTGGAGAAGGCCAATGGCAAGTTTCGACCCGTG		TTGTGTT	QGPTTATGGPQVAL
				GGCACGGATTCCATCTTCAACAAGGTTGTCAACCGCGCTCTGCTCGAGC		TGAAAT	NKIFHHIANNTVPESI
				AACAGCAGCCCCATATTGCCCACTTGCTACAGGCCAGTCCAGAGCTGGC		GGCCGT	RHALCSINYTILEKAN
				CGTCGGAGTCAAGGACGGCATTTCAGCAGCGGTTGGCATGGCCTTTGG		ATCTCTG	GKFRPVGTDSIFNKV
				TGAGCTTCAAGCCTGTGAGTCTACCCCGGGCTGGACCATGCTCTCCCTC		GTTATAC	VNRALLEQQQPHIAH
				GATTTCAAGAGTGCCTTCAACTACACCGACCGAGCACGGCTGCACGAGA		TTGGTC	LLQASPELAVGVKDG
				TTGTGGCCGACAAGGTCCCTGGCCTCTTGCGCGCCTTTGAACGACACTA		GTTTTGT	ISAAVGMAFGELQAC
				TGCTCGACCCACTACCCACTGCATTGTCGACAAGTTCTTCAAGGTGATTG		ACGATTT	ESTPGWTMLSLDFKS
				ACATTGATGTTGGCCAAGGCATTGTGCAGGGCAACGAGCTATCGCCCTT		TTGTTTC	AFNYTDRARLHEIVA
				CTTCTTTGCCCTGTACTCCTGTGAGGTCCTGGGTCTCCTCGACGCCACCA		TATGTG	DKVPGLLRAFERHYA
				CTGACTACCGCTGCAAGGTCATCAAGTACCTCGACGACATTGTACTGAT		CGTGAT	RPTTHCIVDKFFKVIDI
				GGGTCCCGCGGAGGACGTGGCGGCCGACGTGGAGATTGTCAAGGCTC		TCTTCGC	DVGQGIVQGNELSPF
				GTGCAGAGTCTGCTGGCCTTCATCTGCAGCCCAGTAAGAGCCGCTTCTA		GCTTGT	FFALYSCEVLGLLDAT
				CATGCCTCGCCACCATTCGGCTTCCATCACTGCTATCAAGTCTGTATTGC		ACTTCTT	TDYRCKVIKYLDDIVL
				CAGATGCCGTGCGCGAGACGGCCAACACGGGCATGACGGTCTTGGGAA		GGCATG	MGPAEDVAADVEIV
				CGCCGATTGGCCGTCGCGAGTGGATGAAGAAGCAGCTGAACGACAAG		ATAGAA	KARAESAGLHLQPSK
				GCAAAGCACATTGCTGGCAAGCTCAATGACATGCTGACGACCGGTGTCT		GCCAAT	SRFYMPRHHSASITAI
				CGCTTCAGGCCCTCCTCACGGCCATGCAGTACGTGCCTAGCCTCATCAAC		GAATGT	KSVLPDAVRETANTG
				CACCTCTACACGCTGCCCCCAAGTCTCACGTCGGGCTTGTCCGAGCTCTT		GTCTTGT	MTVLGTPIGRREWM
				GAACCGTGCTTGCAAGGACACCTTTGTCAAAGCCTTTTTTGCCAAGGTA		TCTCTTG	KKQLNDKAKHIAGKL
				AACCTGTCTGCACCGGCTGGAGCTGAAGGTCATGACGTTACGCTGGAAC		TGTTGTT	NDMLTTGVSLQALLT
				AGCTCCTTGAGGCTCGCCTCTTCACACGGGCCAACACCGGGGGCTTTGG		TTGCGT	AMQYVPSLINHLYTL
				CCTGCACGACTTGGTTGAGCGCGGTCCGGTGGCTTATGTCTGCAACATG		GCCGTC	PPSLTSGLSELLNRAC
				GCCAAGCTGGCCACTCGCTACCCTCGGGTCTACGATCGACTTTTGGAGG		GTGATTT	KDTFVKAFFAKVNLS
				ATGCATCGAGGGCTGCCGACTTTGAGGCCCACGTGCAGCGAGCTGGCT		TGATGT	APAGAEGHDVTLEQ
				TCCAGATGGCCACGGTCAAGGACGCGGCGACCCAGCGACCAGCTGAGA		CGGGGT	LLEARLFTRANTGGF
				TCATTGCCCTCCGCTCCAAGGCGGCACTGGACGACCTGATGGCCAAGTG		TGCACA	GLHDLVERGPVAYVC

			CGCGCTGGACCTGCAGCAGGCATATCTGGCCTCACGCGAGTGGGGCGT CAGCACTGTCTTGACCATGCGGGGTCGGGACAAGTTGCGTCGCTTGAGC GACACGACCTTTGCCATTGCGGTCGTGTCCATGATGGGTTTTGGCCTCCA TGAACTCATCAACGTCAAGCCGACGGACAAGTGCCCGCTCTGCAGCAGC AAGACACCTCAGCCGCGACTGACCCGCGAGCACCTGCTGACCTGCCGTC CCATCAAGCGTCACAACGCCCTTCGCGACGAGATGGGCCGCCTGCTCAG GTACGCCACCCTCTCCCATGTCTGGGTGGAAAAGTCTGGCTACAACGCC AACGGTCAGAGCTGCCGCATCGACCTGCACTGCCGCAACCCCTTTCCCG GCGGTGCTCTGGGCCCAGCTCTGCCCGACCTGGGCATTGACGTGACTGT GCGCACAGCCCAACCCCCGACCACCTCGCAAGCCTGCATCAAGGTGGGC GCTGCCCTTCGCCGAGCCGAAAAGGAGAAGCGCGACTACTACACCGGT TTCAACCATGGAAAAACTCTGATCGTCCCTGCGGCGATGACGACAACCG GTGGGTTCGCCTCCTCCTCTTTGTGGATCTGCTTGGTCAGCTCGCCCGCTGC GCCGAGGCCCGTGGTGTGTACCAGCCGGGGCTGGATGAGGCCTTTGTT CCTCGGTGGAAGGGTCGCTTTGCGGCGCTGGTCCATCAGATGAACGCT GACCACATCCAGCGCCACTTTGGCGGTGTCTGCCTGCGCTCGTCGTAGG TAGGCACCGTCTCGGGGGTCCCTCTGTGGGGATCCCTGTGTGCACCTGT CGCTCCCTAGGTGGTTCCTCGTTGTGTCTTTGATGGCTTGACTTGTATTT TTGTTTTAATTTTGCTTTAATTTTTGCTGTATTTGTGTGGTATTTTTGCTGA ATTTTTGTAAGGTCCTTTGTATGATGTCTTTGTCTTCTGTGGTCGGTTGTT TTCCTCAATCCGACGTTGTGTCTCGTTGGATGTGAGCGTGCCGTGGTGTT CTTTGTGTTTGTGCTGTGATGGCTTGTAGTTGTGATGTGTGACTGCCTTT TTGGGTGTCTTGTGTTTGAAATGGCCGTATCTCTGGTTATACTTGGTCGT TTTGTACGATTTTTTGTTTCTATGTGCGTGATTCTTCGCGCTTGTACTTCTT GGCATGATAGAAGCCAATGAATGTGTCTTGTTCTCTTGTGTTGTTTTGCG TGCCGTCGTGATTTTGATGTCGGGGTTGCACAGCTTTGCTTTCAGCTCTG AGGTTCAAACACCTAATTTA (SEQ ID NO: 1036)		GCTTTG CTTTCAG CTCTGA GGTTCA AACACC TAATTTA (SEQ ID NO: 1282)	NMAKLATRYPRVYD RLLEDASRAADFEAH VQRAGFQMATVKD AATQRPAEIIALRSKA ALDDLMAKCALDLQ QAYLASREWGVSTVL TMRGRDKLRRLSDTT FAIAVVSMMGFGLH ELINVKPTDKCPLCSS KTPQPRLTREHLLTCR PIKRHNALRDEMGRL LRYATLSHVWVEKSG YNANGQSCRIDLHCR NPFPGGALGPALPDL GIDVTVRTAQPPTTS QACIKVGAALRRAEK EKRDYYTGFNHGKTL IVPAAMTTTGGFASS FVDLLGQLARCAEAR GVYQPGLDEAFVPR WKGRFAALVHQMN ADHIQRHFGGVCLRS S (SEQ ID NO: 1404)
CRE	CRE-2_HMa	Hydra vulgaris	AATTTAAAAAAAAAAAAATCGTTTATTTATGGCATAAATACTGTTTGTAATT TTTGAAAATTCGTGCAACAACTGCAGTTAAATTGAAGAGCTGAAATTTA AGATCTGAGCTTTTCAATCAGAGTTTTTTACCCTAAAACATTAAATTTTAT CATAACAAAAATCGTTCTAATATTATTAAACTTAAAGAAATTCGTTCTTAT ATCAAATCTTATTTCAGTGTTTCACAGACGAAGGGTTTTACTAGATTTTT ATTTTTTCAACTTTTGAATTTGTTTATTATAAAACTGTAAACTAGTGTGCA ACCAACCGTAAAAAWTAGTTAGCTGTTCACCCAAAATATTATTCAGTAT GAAAATATTTAATCTTCTTTATTTCGCAGTAAACAATAAAATATCTAGTTA AACAAAATATTTCTTAATAATAAAAAACAAAAACTTTTTCTTAACAAGTA CAATGAGCAGCTGCAAAGTTACTATACCTCATGTGTGCCCTTATTGTAAA GTAGAACTTAAAACAATATGCGGAATAAACCGTCACATTTTAAAATGCA	AATTTAAA AAAAAAAA ATCGTTTA TTTATGGC ATAATACT GTTTGTAA TTTTTGAA AATTCGTG CAACAACT GCAGTTAA ATTGAAGA	TGAGCT CTTATAA ATTTATA TTATAGC ATTTTGT TTTA (SEQ ID NO: 1283)	MSSCKVTIPHVCPYC KVELKTICGINRHILKC KKNPLQIPSLQKTNTS LTLEPNTKVIPSITKQ NDIIIASTSSNNLAFN QKKDYTLTPTYSRKTT PVSILSSMKMTPISITS HIVRRKLPELPSQTTN HLFNENFINVPFLPEI MNHLPVPNNNVM WGVYSYQQQFKLFVD

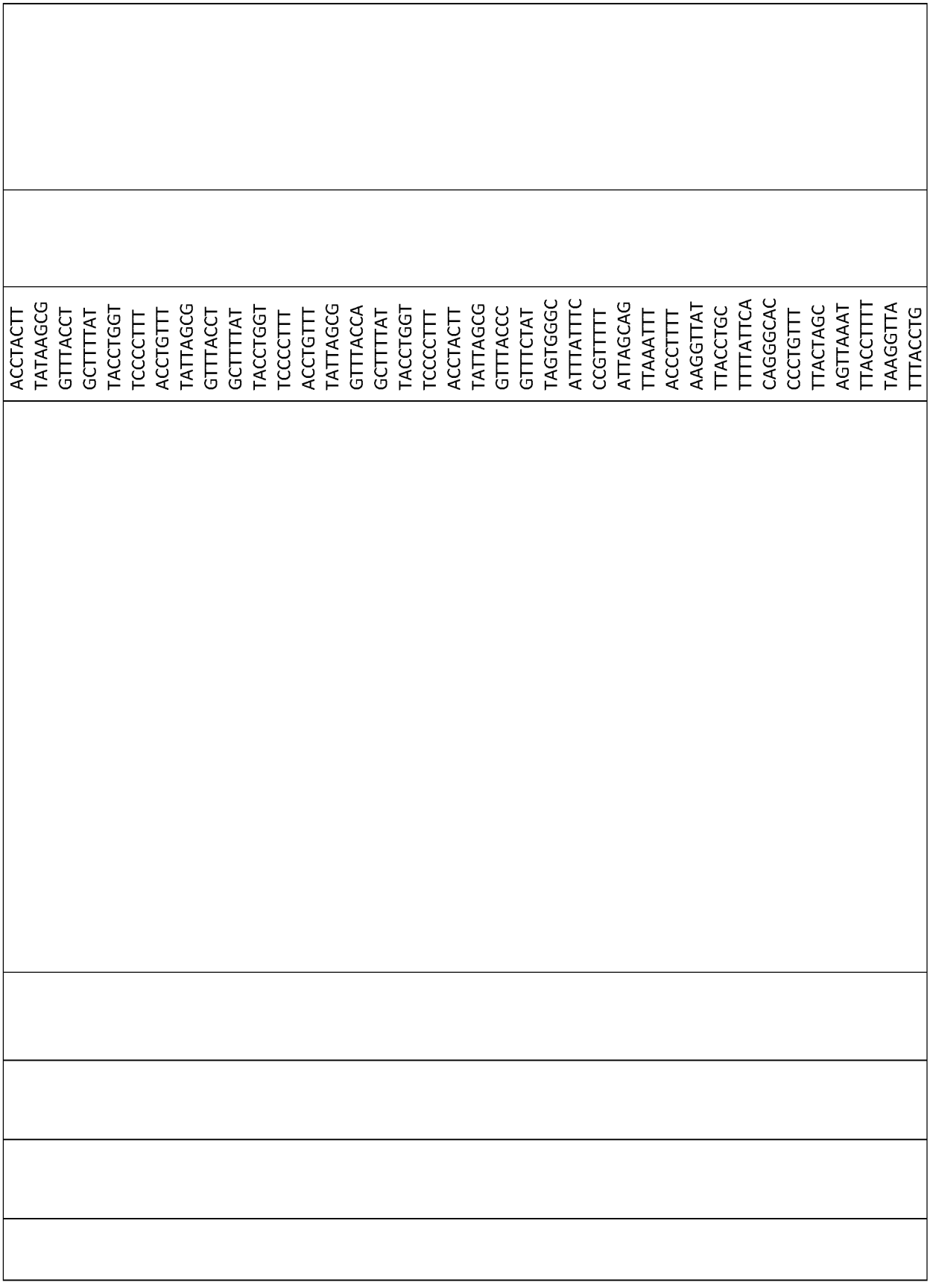
		AAAAGAATCCTTTACAAATACCCAGTCTACAAAAAACTAATACCTCTTTA	GCTGAAAT		STYDEIVNYRRNIFNI
		ACACTCGAACCAAATACTAAAGTAATACCCTCAATTACAAAACAAAACG	TTAAGATC		PSGKAGKEFIEELTFW
		ATATTATAATAGCATCCACTTCGTCTAACAACTTAGCGTTTAATCAAAAA	TGAGCTTT		LRKFNSTSSLNSIALK
		AAGGACTACACATTAACACCTACATATTCTAGAAAAACGACACCCGTAA	TCAATCAG		VTMILPNLLLQKPSAK
		GCATACTGTCTTCTATGAAAATGACACCCATAAGTATAACATCACATATA	AGTTTTTT		SKSKEHTLCLTRRIDL
		GTTCGCAGAAAACTACCTGAGCTTCCTTCTCAAACAACAAATCATTTATT	ACCCTAAA		WKKGDTSLLLKEVRN
		TAATGAGAATTTTATAAATGTTCCCTTCTTGCCTGAAATAATGAACCATC	ACATTAAA		IQKKFVNSKXKRSMD
		TACCAGTTCCAAATAACAATGTCATGTGGGGAGTATACTCATATCAACA	TTTTATCAT		DISRIFAKLIMEGKITA
		ATTTAAATTGTTTGTGGATTCTACCTATGATGAGATCGTAAATTACCGAA	AACAAAAA		ALKFLEKEASSGILPLS
		GAAATATTTTCAACATTCCATCTGGAAAGGCAGGTAAAGAATTTATAGA	TCGTTCTA		DNTLKDLKSKHPEPS
		GGAGCTAACCTTTTGGTTAAGAAAGTTTAATTCCACTTCTAGTTTAAATT	ATATTATT		RVEDYSLLFGPIDLIPK
		CAATCGCGCTGAAAGTTACAATGATTTTGCCGAATCTTCTTTTGCAAAAA	AAACTTAA		CFFDCIDEQLVMKAA
		CCCTCCGCCAAATCAAAGTCCAAAGAACATACATTATGTTTAACTCGTAG	AGAAATTC		FATKGSAGPSGMDA
		GATTGACCTTTGGAAAAAAAGGAGATACTAGTTTACTGTTAAAAGAAGTT	GTTCTTAT		DIYRRILCSKNFIKEGK
		CGAAATATACAAAAAAAAATTTGTAAATTCCAAAAAKAAAAGATCTATGG	ATCAAATC		ELRKEIAKMTQNLLTE
		ACGATATATCTAGAATATTTGCCAAATTAATTATGGAAGGCAAAATCACT	TTATTTCA		TYEPTFLEAFTACRLIP
		GCAGCGCTGAAATTTTTAGAAAAAGAGGCATCATCCGGCATACTACCAC	GTGTTTCA		LDKNPGIRPIGVGEVL
		TATCAGACAACACATTAAAAGACCTTAAAAGCAAACACCCTGAACCCTC	CAGACGAA		RRIIGKVISWSFNSEIK
		CCGAGTAGAAGATTATAGCTTACTGTTTGGTCCGATTGATTTAATCCCAA	GGGTTTTA		EAAGPLQTCAGHGA
		AATGTTTCTTCGATTGTATTGATGAGCAACTAGTTATGAAAGCAGCATTT	CTAGATTT		GAEAAVHAMKEIFD
		GCAACTAAAGGATCTGCTGGACCATCAGGAATGGATGCCGATATTTATC	TTATTTTTT		NVQTDAILLIDAKNA
		GCCGCATCTTATGTTCTAAAAACTTCATCAAAGAAGGTAAAGAACTCCG	CAACTTTT		FNCMNRQVALHNIQ
		AAAAGAAATTGCTAAAATGACACAAAACTTACTAACAGAAACATATGAA	GAATTTGT		IICPLISIYLINTYRNPS
		CCAACATTTCTAGAAGCTTTCACTGCTTGTCGATTAATTCCTCTAGATAAA	TTATTATA		RLFVAGGKEISSQEGT
		AATCCAGGTATTAGGCCAATTGGAGTAGGAGAAGTATTAAGGCGTATC	AAACTGTA		TQGDPLAMPWYSC
		ATAGGTAAAGTAATTAGCTGGAGTTTCAACAGTGAGATAAAAGAGGCA	AACTAGTG		NTTIIIEHLLVNYPQV
		GCCGGGCCATTACAAACATGTGCTGGACATGGGGCAGGAGCCGAAGCG	TGCAACCA		KQVWLADDAAASGS
		GCTGTACATGCCATGAAGGAAATATTCGACAATGTGCAAACAGATGCAA	ACCGTAAA		IANLHSWYQHLIDEG
		TACTTTTGATTGACGCAAAGAACGCTTTTAATTGTATGAATCGACAAGTC	AAWTAGT		CKHGYYVNQSKCWLI
		GCCTTACACAACATCCAGATCATTTGTCCATTAATTTCAATTTACTTAATC	TAGCTGTT		VKSPSLAENAGIVFGK
		AATACTTATCGAAATCCATCGAGGCTCTTTGTGGCAGGGGGGTAAAGAAA	CACCCAAA		SVNITTEGQRHLGSVI
		TATCATCCCAAGAAGGCACAACTCAAGGTGATCCCCTTGCTATGCCATG	ATATTATT		GSQNFKNKYCTEKVA
		GTACTCTTGTAACACCACGATTATTATAGAACACTTACTTGTAAATTACC	CAGTATGA		KWLTELKQLCKVAET
		CACAAGTTAAGCAGGTGTGGTTAGCAGACGATGCTGCAGCTAGTGGAA	AAATATTT		QPQAAFIAFTKGFRS
		GCATTGCAAACTTACATAGCTGGTATCAACACCTTATTGATGAAGGATG	AATCTTCT		KFTYFLRTIPKFEQYLA
		TAAACATGGCTACTATGTAAACCAATCTAAATGCTGGTTAATTGTAAAAT	TTATTCGC		PVDEILSHLLLPTLFGK
		CCCCCTCGTTAGCAGAGAATGCAGGCATAGTATTTGGTTCAGTAATAGGTTCGCAA	AGTAAACA		DTPFEDHIRKLFTLTP
		CATAACTACAGAGGGTCAACGACATTTGGGTTCAGTAATAGGTTCGCAA	ATAAAATA		RDGGLGIPILVEEAPH

			AATTTTAAGAACAAATATTGCACTGAGAAAGTAGCAAAATGGTTAACCG	TCTAGTTA		QFLSSVKLTKNLVQQI
			AGTTAAAACAACTTTGTAAAGTAGCAGAGACGCAACCACAGGCCGCTTT	AACAAAAT		IDQDKILKTKNSSGNV
			TATTGCGTTTACAAAAGGATTTCGTTCAAAATTTACATATTTCCTAAGAA	ATTTCTTA		LEDLEKILTTDRLKHR
			CTATTCCAAAATTTGAACAATATCTAGCGCCCGTAGACGAAATACTTAGT	ATAATAAA		KEKIIAVDSMQPDSM
			CATTTGTTGTTGCCAACTCTTTTTGGAAAAGATACGCCCTTTGAGGATCA	AAACAAAA		LRNIQQTRSECASTW
			CATTAGAAAACTTTTTTACATTAACTCCTCGAGATGGAGGATTGGGTATAC	ACTTTTTCT		LNALPLENQGFVLNK
			CTATACTAGTTGAAGAAGCGCCTCACCAGTTTTTATCATCTGTTAAATTA	TAACAAGT		EEFRDALCLRYNFDLK
			ACTAAAAATCTTGTACAGCAAATTATAGATCAAGATAAAATTTTAAAAAC	ACA (SEQ		NIPRICECGEPFNVTH
			AAAAAACTCTTCCGGGAATGTTCTAGAAGATCTTGAAAAAATATTAACT	ID NO:		ALSCKKGGFISSRHD
			ACTGACAGACTTAAGCATCGCAAAGAGAAAAATAATTGCAGTTGATTCAA	1160)		NIRNLFTTLLKRVCIN
			TGCAACCAGATTCAATGTTAAGAAACATACAGCAAACAAGAAGCGAAT			VQSEPHLIPLDNENFY
			GTGCTAGCACCTGGTTAAACGCCTTACCACTAGAAAACCAAGGTTTTGTT			FHTANKSNQARLDIK
			TTAAATAAAGAAGAATTTCGAGATGCACTTTGCTTACGTTATAATTTTGA			ANGFWRNGQTAFF
			CTTGAAAAATATCCCTCGTATTTGTGAATGCGGAGAACCTTTCAATGTAA			DVRVTHVNSMSNKN
			CTCATGCATTGTCGTGTAAGAAAGGTGGCTTCATTTCAAGTCGTCATGAT			LDIAAIFRKHEKEKKR
			AACATAAGAAATTTGTTCACCACATTGCTTAAAAGAGTGTGCATAAATGT			EYGERVREVEHGSLT
			TCAATCCGAACCACACCTCATACCCCTTGATAATGAGAATTTTTATTTCCA			PLVFGTNGGMGKEC
			TACAGCAAATAAAAGTAACCAAGCTCGTTTAGATATTAAAGCAAATGGT			HRFVRRLAEKLAEKQ
			TTTTGGCGAAATGGACAGACAGCATTTTTTGACGTAAGAGTTACGCATG			NEKYSVVMTWLRTK
			TCAATTCCATGAGCAACAAAAATCTAGATATAGCTGCTATATTCAGAAAA			LSFEILRSTILCLRGSRT
			CACGAAAAGGAAAAAAAAAGAGAGTATGGCGAACGAGTTCGTGAAGT			PWTKKNDFEIGVDFK
			CGAACATGGCAGCTTGACACCACTAGTGTTTGGTACAAATGGAGGTATG			MDALEARI (SEQ ID
			GGTAAAGAATGTCATCGGTTTGTTAGAAGGCTAGCGGAGAAACTAGCG			NO: 1405)
			GAAAAACAAAATGAAAAATACTCAGTTGTTATGACATGGTTAAGAACAA			
			AGTTATCTTTTGAGATACTTCGCTCAACTATCCTTTGTTTAAGAGGCTCAA			
			GAACACCCTGGACTAAAAAAAAACGATTTTGAAATCGGTGTTGATTTTAA			
			GATGGATGCTCTAGAAGCAAGAATCTGAGCTCTTATAAATTTATATTATA			
			GCATTTTGTTTTA (SEQ ID NO: 1037)			
CRE	MoTeR1	JQ747487	CCCGAACCCGAACCCAAACCCAAACCCAAACCCAAACCCAAACCCAAAC	CCCGAACC	TAATAG	MVCPTCNGVYADYN
		Magnaporthe oryzae	CCAAACCCAAACCCGGAGGGTTCCCAAGTCGCCTAAACCCGAAGGGTTT	CGAACCCA	GTAACG	DHIRKKHPDERYTAL
			AGGATATTATTTCGTTTATTAGAATTGGATAAATTATTTACCCCTGTTGGA	AACCCAAA	TCCCTAT	QLQPLGLTPCPICKTA
			CAGGGGGGTTGCAGGGGTTAAATTAAGGTTTTTTATTATTTATGCGCCG	CCCAAACC	TTTTGTC	CKNDLGVKTHLSKIH
			TTTATTTGTTTACCCCCCCCAAATATTATAAAAGCGCGTTCCATCCTCTTAG	CAAACCCA	TTTGGTT	KISGASKISTQPRIRTE
			GAAAAGCGAAGCTTTTCCTTGTAAAAGTCGCTAGACTTTTACTATAAAAA	AACCCAAA	TTGTTTT	NTDNTNSVPTSSFNP
			GTCGCTAGACTTTTATACCAATCTTTTAACAAAAAGCGTAGCTTTTTGTT	CCCAAACC	TATCTTT	VLPEIQTLTPGLNNSR
			GCCAATCTATTAAAAAAAGCGGAGCTTTTTTTAACTTTTTCTTTTTTTTTT	CAAACCCG	GTTTTTG	WADNPRKRRADTPS
			TTTTTCTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTATATAT	GAGGGTTC	TTTTTGT	PTRGRNTRPRRFSYT
			TATTATTATTATTATTAGCGGTGGGGCTATTTATGCGCTTTAATTTGTGC	CCAAGTCG	TTTCGTT	DIDLTNDEPADNPRA

		GGGGCTATTTATGCGCTTTAATTTGTGCGGGGCTATTAATGCGCTTTAAC	CCTAAACC	TTTGTTT		NNPRVNNPRVNNEP
		TTTACAAATTTTATTTATGCGCTTTAATTGCTGCGGGCCTGTTAATGCGCT	CGAAGGG	TTGTTTT		PSSPNSLPSISEFHTP
		TTAATTTACAAATTTCATTAATGCGCTTTAACTTTTATATTTACTAATGCG	TTTAGGAT	CGTTTTT		GTLPLTNSNISLKDQH
		TTATTTATATAATTGCTATTATTATCGTTGCTATTATTATTATTGCTATTAT	ATTATTTC	GTTTTTT		DKITGPILQKPLIQKLI
		TATCGTTATTATTATTGCAATTTTATTATATAAACCCTCGTTTGTCCCTCG	GTTTATTA	TTTTTGT		EYSKIPIPEHHLHARQ
		ATTTATCCCGTTTCTTTTCCATCCCATCGCGCGTTTTCGTAAGCTTTGGTT	GAATTGGA	TTTTGTT		AKIFADAANRIAKNFI
		TTCGTAGGATTTGCTTTCGTAGGCTTTGCTTTCGTAGGCTTTCGTCAGCTT	TAATTATTT	TTTGTTT		QSPTEKTLFNLLILPRI
		TTACCTGCTTTTATTTTTTCTTTTTCTTTTTATTCCCCCCCCCTTTTTTTACCT	ACCCCTGT	TTGCCTT		FGIGLINGKVTKIMQ
		GGTTTATTAGCGGTTTACCTGCTTTTATTACCTGGTTCCCCTTTACCTGTT	TGGACAG	TGTTTTT		NFPSQIPPIPKIDFPSE
		TTATTAGCGGTTTACCTGCTTTTATTACCTGGTTCCCCTTTACCTACTTTAT	GGGGGTT	GTTTTTA		KTDSDPVLNAKKLLE
		AAGCGGTTTACCTGCTTTTATTACCTGGTTCCCCTTTACCTGTTTTATTAG	GCAGGGG	TCTTTAT		KGYIGRAAKAIIDPTP
		CGGTTTACCTGCTTTTATTACCTGGTTCCCCTTTACCTGTTTTATTAGCGG	TTAAATTA	TTTTGTT		VAPETPESLNILREKH
		TTTACCAGCTTTTATTACCTGGTTCCCCTTTACCTACTTTATTAGCGGTTTA	AGGTTTTT	TTTGTTT		PIGQNNPFNTKSQPI
		CCCGTTTCTATTAGTGGGCATTTATTTCCCGTTTTTTATTAGCAGTTAAATT	TATTATTTA	TTACTTT		SGRQITEKAILLAISSI
		TACCCTTTTAAGGTTATTTACCTGCTTTTATTCACAGGGCACCCCTGTTTTT	TGCGCCGT	GTTTTAT		GREKAPGLSGWTRSL
		TACTAGCAGTTAAATTTACCTTTTTAAGGTTATTTACCTGCTTTTATTCAC	TTATTTGTT	TTGTTTT		LDAAIKIPTQNDVIPA
		AGGGCACCCCTGTTTTTACCAGCAGTTAAATTTACCTTTTTTAAGGTTATTT	TACCCCCC	ATATTTA		LRLLTDMIRQGTAPG
		ACCTGCTTTTATTAACAACCCTTTATTTTTTCCTATTAACGGGTATTTATTT	CAAATATT	CCTTTTG		RELLCASRLIGLSKPD
		ACCTGTTTTATTGGAATTCACCCGTTGGACGGCATGGTTTGCCCAACCTG	ATAAAAGC	ATTTTTT		GGVRPIAVGDLLYKIA
		TAACGGCGTTTACGCCGATTACAACGACCATATCCGGAAAAAACACCCG	GCGTTCCA	CTATTTT		FKAILNTLWSPNCLLP
		GACGAACGTTATACCGCCCTCCAACTCCAACCATTGGGTTTAACCCCCTG	TCCTCTTA	TCCCACC		YQLGVNSIGGVEPAIF
		CCCTATATGCAAAACCGCTTGCAAAAACGATTTGGGCGTTAAAACCCAC	GGAAAAG	CTTATTA		TLEEAIMGPNINGIKS
		CTATCCAAAATCCACACAAAATATCCGGTGCATCGAAAATTTCAACCCAACC	CGAAGCTT	TTATAAC		ITSLDLKNAFNSVSRA
		GCGTATACGAACGGAAAATACGGATAATACCAATTCGGTCCCCACGTCG	TTCCTTGT	CCCAAC		AIASSVAKYAPTFYRS
		TCGTTTAACCCTGTCCTTCCCGAAATCCAAACGTTAACCCCGGGGTTAAA	AAAAGTCG	CTACTAA		TCWAYNQPSILITEN
		TAACAGCCGTTGGGCCGATAACCCCAGAAAACGACGGGCCGATACCCC	CTAGACTT	TATTTTT		GSVLASAQGIRQGDP
		CTCCCCAACACGGGGTCGGAATACACGCCCACGTCGATTTTCATATACG	TTACTATA	TCTTTTT		LGPLLFSLAFRPTLETI
		GATATCGATTTAACAAACGACGAACCGGCGGATAACCCCAGGGCTAATA	AAAGTCGC	TCTTTTT		QKSLPYTYIAAYLDDV
		ACCCCAGGGTTAATAACCCCAGGGTTAATAACGAACCCCCCTCCAGCCC	TAGACTTT	TCTTTTT		YILSKTPVKDKIAKIIEK
		AAATTCGTTACCTTCGATTTCCGAATTTCACACCCCTGGGACCCTACCCCT	TATACCAA	ACGGTT		SPFTLNSAKTTETDID
		AACCAATTCGAATATATCGTTAAAAGACCAGCACGACAAAATTACCGGC	TCTTTTAA	TTATTTT		TLKTNGLKTLGSFIGP
		CCTATATTGCAAAAACCGTTAATCCAAAAATTAATCGAATATTCGAAAAT	CAAAAAGC	CCCGTTT		TELRKEFLQNKIQNFE
		CCCAATCCCAGAACACCACCTCCACGCCAGGCAGGCTAAAATTTTTGCTG	GTAGCTTT	GTTTTTT		SSINALKKLPKQYGLLI
		ACGCCGCAAATCGAATCGCCAAAAATTTTATACAAAGCCCAACGGAGAAA	TTGTTGCC	CTATTTT		LRKSTQLLLRHLLRTL
		AACATTATTTAATTTACTTATATTACCCCGCATATTCGGTATCGGGTTAAT	AATCTATT	ATTTGTA		NSQDLWELWEKTDK
		AAACGGAAAAGTAACTAAAATAATGCAAAACTTCCCATCCCAAATACCC	AAAAAAA	CGACAA		LIADFVINLTVTKRKK
		CCTATTCCAAAAATTGATTTTCCATCCGAAAAAAACCGATTCCGACCCGGT	GCGGAGCT	AACCCTT		RPITDFVTPLITLPIKD
		TTTAAACGCCAAAAAAATTATTGGAAAAAGGGTATATTGGCCGTGCGGCA	TTTTTTAAC	AGCAAA		GGFGLLRHNGIAQDI

	AAGGCTATTATCGATCCAACCCCCGTTGCCCCAGAAACCCCGGAATCGT	TTTTTCTTT	TAAGCTT	YFAAKDLTTEIRHKIQ
	TAAATATTTTACGGGAAAAACACCCTATTGGCCAAAATAACCCGTTTAAT	TTTTTTTTT	AGAATA	RISNDFPQNQSPTAT
	ACAAAATCCCAACCAATATCAGGCAGGCAAATTACCGAAAAAGCTATTT	TTTTCTTTT	TAATAA	EILHLLHNGVLADCK
	TATTAGCTATTTCGTCTATTGGCCGGGAAAAAGCTCCGGGCCTTAGCGG	TTTTTTTTT	AGCGCG	NGLTNAQLNALTEN
	GTGGACGAGATCGTTATTAGATGCAGCCATTAAAATACCTACCCAAAAC	TTTTTCTTT	AATTAA	ASYLGRKWLNILPIQK
	GACGTAATTCCGGCTTTACGACTCTTAACGGATATGATTCGCCAGGGTA	TTTTTTTTT	AA (SEQ	SNRLTDWEMAEAVR
	CCGCACCGGGTAGGGAATTATTATGCGCTTCGCGTTTAATAGGGCTATC	TTTTTTTTT	ID NO:	LRLLAPVKPLTHPCN
	CAAACCCGACGGCGGCGTACGCCCAATAGCCGTTGGGGACCTATTATAT	TATATATA	1284)	HCGNRTNINHEDVC
	AAAATAGCCTTTAAAGCTATTTTAAATACCCTATGGTCCCCAAACTGTTT	TTATTATTA		KGAVRKYTARHDQIN
	ATTACCTTACCAATTAGGTGTAAATAGTATAGGTGGCGTCGAACCCGCT	TTATTATTA		RSFVNSLKSRPEIDVEI
	ATTTTTACCCTCGAAGAGGCTATAATGGGCCCTAATATTAACGGTATAAA	GCGGTGG		EPDLNNENNVNNAN
	ATCGATTACCTCCCTCGATTTAAAAAACGCGTTTAATAGCGTATCCAGGG	GGCTATTT		TTTENPTPSPNGQN
	CTGCAATAGCCTCGTCGGTAGCTAAATACGCACCAACTTTCTACCGTTCT	ATGCGCTT		DTGCLFTTPIRSGTRN
	ACCTGTTGGGCCTATAACCAACCTTCGATTTTAATAACGGAAAACGGTTC	TAATTTGT		GQNGLRADFAVING
	CGTCCTGGCTAGTGCACAAGGTATACGCCAAGGCGATCCGTTAGGCCCG	GCGGGGC		VSKYYYDVQIVAINKD
	TTGTTATTCAGCCTTGCTTTTCGACCTACGTTGGAAACGATCCAAAAATC	TATTTATG		SGNTNPLNTLADAA
	GCTTCCATATACGTATATAGCGGCTTATTTGGACGACGTTTATATTTTATC	CGCTTTAA		NNKRRKYQFLDPFFH
	CAAAACGCCCGTTAAAGATAAAATAGCCAAAATAATCGAAAAAAGCCC	TTTGTGCG		PIIISAGGLMEKDTAQ
	GTTTACCCTAAATTCCGCCAAAACGACAGAAACGGATATCGATACGTTA	GGGCTATT		AYKQIQKLIGPVAAH
	AAAACCAATGGTTTAAAAAACGCTCGGCTCGTTTATTGGACCAACGGAAT	AATGCGCT		WLDTSISLILLRSRTTA
	TACGGAAGGAATTTTTGCAAAATAAAATTCAAAATTTCGAATCGTCCATT	TTAACTTT		AISIAKNRPRA (SEQ
	AACGCCCTGAAAAAAACTCCCTAAACAATACGGATTGCTAATCTTGCGTA	ACAAATTT		ID NO: 1406)
	AAAGTACACAATTACTTTTACGCCATTTGCTCCGTACTTTAAATTCCCAG	TATTTATG		
	GACCTGTGGGAATTATGGGAAAAAAACAGATAAATTAATAGCGGATTTC	CGCTTTAA		
	GTTATAAATTTAACTGTTACAAAACGGAAAAAACGGCCAATTACGGATT	TTGCTGCG		
	TCGTTACGCCGTTAATTACGTTACCTATAAAGGACGGAGGTTTTGGATTA	GGCCTGTT		
	TTACGGCATAACGGAATAGCCCAAGATATTTATTTTGCGGCCAAGGATT	AATGCGCT		
	TAACAACCGAAATTCGGCACAAAATCCAACGTATATCCAACGATTTTCCA	TTAATTTA		
	CAAAATCAAAGCCCTACCGCCACCGAGATTTTTGCATTTGTTGCATAACG	CAAATTTC		
	GGGTTTTAGCAGATTGCAAAAAACGGGTTAACAAACGCCCAATTAAACGC	ATTAATGC		
	TTTAACCGAAAACGCTAGTTATTTAGGTCGAAAATGGCTTAACATTTTAC	GCTTTAAC		
	CTATCCAAAAATCAAATCGATTAACGGATTGGGAAATGGCTGAAGCCGT	TTTTATATT		
	TCGATTAAGATTATTAGCCCCGGTTAAACCGTTAACCCACCCCTGCAACC	TACTAATG		
	ATTGCGGAAATCGGACCAATATAAACCACGAGGACGTTTGCAAAGGTG	CGTTATTT		
	CCGTACGCAAATATACGGCCCGTCACGACCAAATAAACAGAAGTTTCGT	ATATAATT		
	CAATTCGTTAAAAAGTCGACCAGAAATCGACGTCGAAATCGAACCCGAT	GCTATTAT		
	TTAAATAACGAAAATAACGTAAATAACGCCAATACAACCACCGAAAATC	TATCGTTG		
	CCACCCCTAGCCCCCAACGGCCAAAACGATACCGGATGCCTTTTTACAACC	CTATTATT		

			CCTATTCGCTCCGGGACCCGTAACGGCCAAAACGGCCTTAGGGCGGATT	ATTATTGC	
			TTGCCGTTATTAACGGCGTATCCAAATATTATTACGACGTGCAAATCGTT	TATTATTAT	
			GCAATTAATAAGGATTCCGGTAATACAAATCCGTTAAATACGTTAGCAG	CGTTATTA	
			ACGCAGCAAATAACAAACGACGTAAATACCAATTTTTGGATCCATTTTTC	TTATTGCA	
			CATCCAATTATAATAAGCGCCGGAGGCCTTATGGAAAAGGATACAGCAC	ATTTTATTA	
			AGGCGTACAAACAAATCCAAAAATTAATAGGCCCCGTTGCGGCCCATTG	TATAAACC	
			GTTGGATACGTCGATTTCGTTAATTTTGTTACGGTCCAGAACGACGGCA	CTCGTTTG	
			GCAATTTCTATTGCTAAAAACCGCCCTCGTGCGTAATAGGTAACGTCCCT	TCCCTCGA	
			ATTTTTGTCTTTGGTTTTGTTTTTATCTTTGTTTTTTGTTTTTCGTTTTT	TTTATCCC	
			TGTTTTTGTTTTCGTTTTTTGTTTTTTTTTTTTGTTTTTTGTTTTTTGCCTT	GTTTCTTTT	
			TGTTTTTGTTTTTATCTTTATTTTTTGTTTTTTGTTTTTACTTTGTTTTTATTTGTT	CCATCCCA	
			TTATATTTACCTTTTGATTTTTTTCTATTTTTTCCCACCCTTATTATTATAACCC	TCGCGCGT	
			CAACCTACTAATATTTTTTTCTTTTTTTCTTTTTTACGGTTTTTATTTTTC	TTTCGTAA	
			CCGTTTGTTTTTTTCTATTTTATTTGTACGACAAAACCCTTAGCAAATAAGC	GCTTTGGT	
			TTAGAATATAATAAAGCGCGAATTAAAA (SEQ ID NO: 1038)	TTTCGTAG	
				GATTTGCT	
				TTCGTAGG	
				CTTTGCTTT	
				CGTAGGCT	
				TTCGTCAG	
				CTTTTACCT	
				GCTTTTAT	
				TTTTTCTTT	
				TTCTTTTTA	
				TTCCCCCC	
				CCTTTTTTT	
				TACCTGGT	
				TTATTAGC	
				GGTTTACC	
				TGCTTTTA	
				TTACCTGG	
				TTCCCCTTT	
				ACCTGTTT	
				TATTAGCG	
				GTTTACCT	
				GCTTTTAT	
				TACCTGGT	
				TCCCCTTT	


ACCTACTT
TATAAGCG
GTTTACCT
GCTTTTAT
TACCTGGT
TCCCCTTT
ACCTGTTT
TATTAGCG
GTTTACCT
GCTTTTAT
TACCTGGT
TCCCCTTT
ACCTGTTT
TATTAGCG
GTTTACCA
GCTTTTAT
TACCTGGT
TCCCCTTT
ACCTACTT
TATTAGCG
GTTTACCC
GTTTCTAT
TAGTGGGC
ATTTATTTC
CCGTTTTT
ATTAGCAG
TTAAATTT
ACCCTTTT
AAGGTTAT
TTACCTGC
TTTTATTCA
CAGGGCAC
CCCTGTTT
TTACTAGC
AGTTAAAT
TTACCTTTT
TAAGGTTA
TTTACCTG

HERO	HERO-2_BF	.	Branchiostoma floridae	TTTTCAGTCTGGCTCAGCCAGTGACCGCCGGGAAAGTCCGGCTGACTAC CACGAATAGGGTGGTGACAGCTGGATAGACAGACGACAGCTCGGAAA GACGGCATTGGGGCAGTATGGGTTGGCACCCCTAACTGCATCTCCCCTA GGAGAGCATCCCGCAACACGCTACAAAGAACCACAAAGAGCAATACCC CCAGGGATGCCCGAGAGAGGGGGGAGGATGAGCATCCCATTCGGACGG TCCAATCGGTATTGACCCCAGCAAACGGAGAATCGACAATGAATGCAGT CTGTGTGTGTGGCAAGGTATGTAAGAACCAGAGAGGTTTGAGAATCCA CCAAACAAAGATGGCCTGCTTAAGGAGGGGTGCAGGCGGAGCACCGCTC AGGGGCTGTGGCAACCACTGTAGAACCAGTGTTGTCAGCATCAGCCCCT GGTCAGACGGAGGAGGATCAGGGCCCGGAAGCTCCCCACAGTGCCCG GAACCTCCGCGCAACGCCTGCCCCTCCACAAGGCAGGAAGTCAGATCAT CACCGAGTGAAGTGGCCAGCCGCAAACTCCAAGGAGTGGTCGCAGTTT GACGAGGACGTTGACATGATCTTGGAGTCGGTGTCAAGAGGTAGTACA GACCAAAAGCTTCAGTCCATGTGCACAGTGATTATGTCCATGGGGGCAG AACGATTTGGCACGATTGGGCAGAGGAAACCGACAGACACAATGAAGC CAAATCGCCGGGAAGTAAAGATCCGTCAACTGAGGCAGGAGCTAAAGT CGTTGAGGCGGAGCTTTAAGGCGAGTACGTCGGGAGAAGCTTAGGACCCTCA	TTTTCAGT CTGGCTCA GCCAGTGA CCGCCGGG AAAGTCCG GCTGACTA CCACGAAT AGGGTGG TGACAGCT GGATAGAC AGACGACA GCTCGGAA AGACGGC ATTGGGGC AGTATGGG TTGGCACC CCTAACTG CATCTCCC	TGATTA AAGACC CGAAAC ACCCAA TGACCC CGGGTT CATCACT GATGAT GTGTCC CTGTTC GCACTA CCAGAG TGTATTC TAGAG (SEQ ID NO: 1285)	MNAVCVCGKVCKN QRGLRIHQTKMACLR RVQAEHRSGAVATT VEPVLSASAPGQTEE DQGPEAPHSARNLR ATPAPPQGRKSDHH RVKWPAANSKEWS QFDEDVDMILESVSR GSTDQKLQSMCTVI MSMGAERFGTIGQR KPTDTMKPNRREVKI RQLRQELKSLRRSFKA STSGEERAALAELTH HLREKLRTLRRAEWH KKKGKERARKRSAFIT NPFGFTKRLLGQKRS GNLTCPVEEINLHLSN TFSDASRDVDLGPCP

		GAAGGGCAGAGTGGCACAAGAAGAAGGGTAAAGAAAGAGCCCGGAA	CTAGGAGA		LLVTSPEPEVHFDISE
		GCGCAGTGCTTTCATCACCAACCCTTTCGGCTTCACCAAGCGACTCCTAG	GCATCCCG		PTLKEVRETVKAARSS
		GGCAGAAGAGGAGTGGGAACCTGACCTGCCCAGTCGAGGAGATCAACC	CAACACGC		SAPGPSGVVYKVYKH
		TCCACCTCAGCAATACCTTCAGTGATGCCTCGAGAGATGTGGATCTTGG	TACAAAGA		CPRLVVRLWRILKVV
		TCCTTGCCCTTTGCTGGTGACTTCACCTGAGCCGGAAGTGCACTTTGACA	ACCACAAA		WRRGKVAADWRQA
		TCTCTGAACCAACTCTGAAGGAGGTCAGAGAGACAGTCAAGGCGGCGA	GAGCAATA		EGVWIPKEEESSKVD
		GGTCCAGTTCGGCGCCAGGTCCCAGTGGCGTGGTATACAAGGTCTACA	CCCCCAGG		QFRLISLLSVEGKIFFKI
		AACATTGCCCACGGCTTGTGGTGCGCCTCTGGAGGATCCTAAAGGTGGT	GATGCCCG		VAQRLIKYLLDNQYID
		CTGGCGCAGAGGTAAAGTGGCGGCTGATTGGAGGCAAGCCGAGGGGG	AGAGGGG		TSVQKGGVPGVPGC
		TTTGGATCCCAAAGGAAGAGGAGTCAAGTAAGGTAGACCAGTTCCGCT	GGGAGGA		LEHTGVVTQLIREAKE
		TAATTTCTCTGCTCAGTGTTGAGGGAAAGATCTTCTTCAAGATTGTGGCC	TGAGCATC		NRGDLAVLWLDLAN
		CAGCGTCTAATAAAGTACCTTCTGGACAACCAGTATATTGACACATCTGT	CCATTCGG		AYGSIPHKLVETALTR
		GCAGAAGGGGGGAGTTCCTGGTGTCCCAGGATGTCTTGAACACACGGG	ACGGTCCA		HHVPESIQNLILDYYS
		CGTAGTGACCCAGCTCATCCGGGAGGCTAAGGAGAACAGAGGGGACTT	ATCGGTAT		NFWLRAGSSTATSA
		GGCAGTCTTGTGGCTGGATCTCGCGAATGCGTATGGTTCGATCCCCCAC	TGACCCCA		WQRLEKGIITGCTISV
		AAGCTTGTGGAAACAGCACTGACCAGACACCATGTTCCAGAGTCAATTC	GCAAACG		PLFALAMNMIVKGA
		AGAACCTCATCTTAGATTACTACAGCAACTTCTGGCTAAGAGCTGGCTCC	GAGAATCG		EAGCRGPVSRSGTRQ
		AGTACAGCAACTTCAGCATGGCAACGGTTAGAGAAGGGCATCATTACTG	ACA (SEQ		PPIRAFMDDLTVMT
		GATGTACGATTTCAGTGCCCCTCTTTGCACTAGCGATGAACATGATTGTT	ID NO:		ATVPVCRWLLQGLER
		AAAGGAGCGGAAGCAGGATGTAGGGGTCCCGTGTCTAGGTCTGGAACC	1162)		LITWARMSFKPAKSR
		AGGCAGCCGCCGATTCGAGCCTTCATGGACGATCTGACGGTGATGACT			SLVLKKGKVAERFRFT
		GCAACAGTCCCGGTGTGTAGATGGCTCCTACAGGGATTAGAGCGTCTCA			LGGTQIPTVSEKPVKS
		TTACATGGGCACGGATGAGTTTCAAGCCGGCCAAGTCAAGATCTCTTGT			LGKVFNSSLKDTASV
		CCTGAAGAAGGGGAAGGTGGCTGAAAGGTTCCGTTTCACCCTGGGAGG			QQTRSDLTTWLEGID
		CACTCAGATTCCCACAGTGTCAGAGAAACCAGTCAAGAGTCTGGGCAAG			KTGLPGSFKAWMFQ
		GTGTTCAACAGCTCTCTGAAGGACACCGCTTCAGTTCAGCAGACTAGGA			HGVLPRVLWPLLVYE
		GTGACCTGACAACGTGGCTCGAGGGAATTGACAAGACAGGGCTACCTG			VPMTMVEQLERTISR
		GTAGCTTCAAGGCCTGGATGTTCCAGCATGGAGTCTTGCCAAGGGTACT			FLRKWLGLPRSLSNIA
		CTGGCCTCTTCTTGTGTACGAGGTGCCGATGACCATGGTGGAGCAACTG			LYGRSTKLQLPLSGLT
		GAGAGAACCATCAGCAGGTTCCTTCGCAAATGGTTGGGGCTCCCGAGG			EEFKVTRAREVLMYR
		TCCTTAAGCAACATTGCCCTGTACGGTAGATCCACCAAGCTGCAGCTTCC			DSSDSKVSSAGIHVR
		CTTGAGTGGCCTGACTGAAGAGTTCAAGGTTACCCGTGCAAGAGAAGT			TGRKWKAQEAVDQ
		GTTGATGTACCGGGACTCCTCAGACTCCAAGGTCTCTTCAGCCGGCATC			AEARLRHSVLVGSVA
		CATGTCAGGACTGGAAGAAATGGAAGGCACAGGAAGCAGTGGATCA			VGRAGLGSCPKPRYD
		GGCAGAGGCAAGGTTGAGACACAGTGTCCTCGTGGGGTCCGTGGCAGT			KVSGKEKRLLIQDEIR
		AGGACGGGCAGGACTGGGCAGCTGCCCAAAGCCTCGGTACGACAAAGT			AGEEEDRRCRMVG
		CAGCGGGAAGGAGAAGCGTCTACTGATCCAGGATGAGATAAGGGCTG			MRKQGAWTRWEH
		GGGAAGAGGAGGATCGGCGATGCAGGATGGTAGGCATGCGCAAGCAA			ADSRKVTWPELCRAE

			GGTGCGTGGACTAGGTGGGAACATGCTGACTCCCGCAAGGTCACATGG				PSRIKFLISSVYDVLPS
			CCAGAGTTGTGCAGAGCTGAGCCTTCTCGGATCAAATTTCTCATCTCTTC				PANLHVWGLAETPS
			AGTGTACGACGTGCTTCCAAGTCCAGCTAACTTGCATGTCTGGGGCTTG				CQLCQRRGTLEHILSC
			GCAGAGACCCCCTCATGCCAACTCTGTCAGAGGAGAGGTACCCTTGAAC				CPKALGEGRYRWRH
			ACATTCTCAGTTGTTGTCCGAAAGCACTAGGGGAAGGGAGGTACCGCT				DQVLRVLADTVSNAI
			GGCGGCATGACCAGGTTCTTAGGGTGTTGGCAGACACAGTTAGCAACG				QSSRSQQPPKKSIVF
			CCATCCAGAGTAGCAGGAGTCAGCAACCCCCCAAGAAGTCAATTGTCTT				VRAGEKTRQQPTSA
			TGTCAGGGCCGGAGAGAAAACCCGACACAACAACCCACTTCCGCAGGTGG				GGLLSTARDWQLLV
			GCTTCTCTCCACTGCTAGAGATTGGCAGCTTCTAGTCGACCTTGGGAGA				DLGRQLKFPEHIVATS
			CAGCTCAAGTTTCCAGAACACATTGTAGCCACGTCACTTCGCCCTGACAT				LRPDMVLVSESTRQV
			GGTACTCGTGTCAGAATCCACCAGACAAGTGGTTCTGCTGGAGCTAACT				VLLELTVPWEERISEA
			GTTCCCTGGGAGGAGCGGATAAGCGAAGCCAACGAGCCGGAAGAGGGC				NERKRAKYAELVVQS
			GAAGTATGCCGAACTGGTAGTACAAAGCCAGAGTAATGGGTGGAGAGC				QSNGWRARCVPVEV
			CCGGTGTGTACCAGTGGAGGTTGGTTGCCGGGGTTTCGCAGGGCAGTC				GCRGFAGQSLAYVLK
			TTTGGCTTATGTGTTAAAACTCCTTGGAGTAAGAGGTTTCCGTCTTCGGA				LLGVRGFRLRKSIRDIL
			AATCCATCAGGGATATTCTAGAGGCTGCGGAGAAAGCCTCACGTTGGTT				EAAEKASRWLWFRR
			GTGGTTCCGTAGGGGGGAACCGTGGAAGCCACACGGACACAGGTCGG				GEPWKPHGHRSGN
			GGAATGATCAACCTCGGCTGGGTCGCCCGGGCGAGGGTGTATGGTGAT				DQPRLGRPGEGVW
			TAAAGACCCGAAACACCCAATGACCCCGGGTTCATCACTGATGATGTGT				(SEQ ID NO: 1407)
			CCCTGTTCGCACTACCAGAGTGTATTCTAGAG (SEQ ID NO: 1039)				
HERO	HERO-2_DR	.	TTCAAGCCTGGCGCAGCCAGTGACTCCTAGGAATAGACTAGGTGGCAA	TTCAAGCC	TGATCA		MTHANEQTTNKIYVT
		Danio rerio	CCAAGAATAGTTTGGTCGACTACTGGAGAGACAGTTGACGGCACGGAA	TGGCGCAG	ACCCCG		CICGKLCKNHWGLKI
			AGACGGCACTTGGGACAGTATGGGTTAGCACCCCAGCCTGTGTCTTTCG	CCAGTGAC	GCTGGG		HQARMKCLEQESKV
			TGAGAGAGAACCCAAACAAGCTACGGAAAGCCCCACAGAGATATACCC	TCCTAGGA	TCACCTG		QRTGPEPGETQEEPG
			CCAGGAGATCCCGAGAGGGGGGGAGGATGAGATCTCCAATCGGACGG	ATAGACTA	GGTGAG		PEATHRAKSLHVPEP
			ATCAAAGGTTAATGACCCATGCAAACGAACAGACGACGAACAAAATAT	GGTGGCA	AGTGTA		QTPSEVVQQRIKWP
			ATGTGACATGCATTTGCGGAAAGCTGTGTAAGAACCATTGGGGCCTAAA	ACCAAGAA	TGATGTT		PASKGSEWLQFDED
			AATCCATCAGGCCAGAATGAAATGTTTGGAGCAGGAGAGTAAGGTGCA	TAGTTTGG	GAGAGA		VSNIIQAIAKGDADSR
			ACGCACAGGTCCTGAACCTGGTGAGACGCAGGAGGAGCCCGGCCCGG	TCGACTAC	CCCGAA		LKTMTTIIFSYALERFG
			AGGCAACCCACAGAGCCAAGTCCCTCCATGTACCAGAGCCTCAAACTCC	TGGAGAG	ACACTC		CIEKGKTKPTTPYTM
			AAGCGAAGTAGTTCAACAGCGGATTAAATGGCCCCCAGCCAGCAAAGG	ACAGTTGA	AATGAT		NRRATQIHHLRQELR
			AAGTGAGTGGCTGCAGTTCGATGAAGATGTGTCCAACATCATTCAAGCC	CGGCACG	CCCAGG		SLKKLYKKATDEEKQP
			ATAGCCAAAGGAGATGCAGATAGCCGACTCAAAACGATGACTACCATC	GAAAGAC	ATACATC		LAELKNILRKKLMILR
			ATCTTCAGCTATGCTCTAGAAAGATTCGGTTGCATAGAGAAAGGAAAGA	GGCACTTG	ACTGAT		RAEWHRRRGRERAR
			CCAAGCCCACCACCCCCTACACTATGAACCGTAGGGCTACCCAGATACA	GGACAGTA	GATGTG		KRAAFITNPFGFTKQL
			TCACCTGCGTCAGGAGCTTCGCTCCCTCAAGAAACTGTATAAGAAAGCT	TGGGTTAG	TCCCAA		LGDKRSGRLECSIEEV
			ACGGATGAGGAGAAGCAACCATTAGCGGAGTTGAAAAACATTTTGCGG	CACCCCAG	ATGCAT		NRFIEETVSDPLREQE
			AAGAAGCTGATGATCCTACGCAGGGCAGAGTGGCATCGGAGACGAGG	CCTGTGTC	CCATGA		LEPNKALISPTPPARE

		GCGAGAGAGAGCCAGGAAGCGAGCTGCCTTCATCACCAATCCCTTTGG	TTTCGTGA	GATGTTT	FSLRGPSLKEVKEIIKA
		CTTCACAAAACAGCTGCTCGGGGACAAGCGGAGCGGTCGACTTGAATG	GAGAGAA	CTTGCAT	SRSASTPGPSGIPYLV
		CTCAATAGAGGAAGTGAATCGCTTCATTGAGGAAACAGTGAGTGATCCA	CCCAAACA	AA (SEQ	YKRCPGLLLHLWKILK
		CTGAGAGAGCAGGAGCTGGAGCCCAACAAAGCTCTTATCAGCCCCACCC	AGCTACGG	ID NO:	VIWQRGRVAEQWR
		CTCCAGCAAGAGAGTTCAGTTTGAGGGGGCCAAGTCTGAAGGAGGTCA	AAAGCCCC	1286)	CAEGVWIPKEENSKN
		AGGAAATCATTAAGGCATCTCGCTCAGCATCTACTCCAGGCCCTAGTGG	ACAGAGAT		INQFRIISLLSVEGKVF
		CATACCTTACCTTGTCTATAAGCGCTGCCCAGGGCTTCTCCTGCATCTGT	ATACCCCC		FSIVSRRLTEFLLENNY
		GGAAGATCTTGAAGGTGATTTGGCAACGAGGAAGAGTTGCTGAGCAGT	AGGAGATC		IDPSVQKGGIPGAPG
		GGAGGTGTGCCGAGGGAGTGTGGATTCCTAAAGAGGAAAACTCGAAA	CCGAGAG		CLEHTGVVTQLIREA
		AACATCAACCAGTTTCGAATCATCTCTCTATTGAGTGTTGAAGGGAAGG	GGGGGGA		HENRGDLVVLWLDL
		TGTTTTTCAGCATCGTCTCACGAAGACTGACAGAGTTCCTCCTCGAGAAC	GGATGAG		ANAYGSIPHKLVELAL
		AATTATATTGACCCTTCAGTGCAGAAGGGAGGGATTCCTGGAGCTCCCG	ATCTCCAA		HRHHVPSKIKDLILDY
		GCTGCTTGGAACACACACTGGAGTAGTTACACAACTCATCAGAGAGGCCCA	TCGGACGG		YNNFKMRVTSGSETS
		TGAGAACAGAGGGGACTTGGTTGTCTTGTGGTTGGACTTGGCAAATGC	ATCAAAGG		SWHRIGKGIITGCTIS
		CTATGGGTCCATACCCCACAAGCTGGTTGAGCTCGCTCTACACCGCCACC	TTA (SEQ		VILFALAMNMVVKS
		ACGTTCCTAGTAAGATTAAGGACCTAATTCTGGATTACTACAATAATTTC	ID NO:		AEVECRGPLTKSGVR
		AAGATGCGGGTCACATCTGGGTCAGAAACATCAAGCTGGCATCGCATC	1163)		QPPIRAYMDDLTITTT
		GGGAAAGGAATAATAACAGGCTGCACCATCTCAGTTATTCTTTTCGCTCT			TVPGSRWILQGLERLI
		CGCCATGAACATGGTGGTCAAGTCAGCCGAAGTGGAATGCAGAGGGCC			AWARMSFKPSKSRS
		CTTAACTAAGTCAGGTGTGCGACAGCCCCCTATTAGAGCATATATGGAT			MVLKKGKVVDKFHF
		GACCTTACCATCACACAACAACACGGTCCCAGGGAGCAGGTGGATCTTAC			SISGSVIPTITEQPVKS
		AAGGACTTGAGAGACTCATCGCCTGGGCTAGAATGAGTTTTAAGCCCTC			LGKLFDSSLKDSAAIQ
		CAAGTCTAGGTCCATGGTGCTGAAGAAGGGGAAAGTGGTTGACAAGTT			KSKKELGAWLAKVDK
		CCATTTTTCCATCTCAGGAAGTGTCATCCCAACCATCACGGAGCAACCTG			SGLPGRFKAWIYQHS
		TCAAGAGTTTGGGGAAGCTCTTTGACTCCAGCCTAAAAGACTCTGCAGC			ILPRVLWPLLIYAVPM
		CATCCAGAAGTCCAAAAAAGAACTTGGAGCTTGGCTGGCGAAGGTTGA			STVESLERKISGFLRK
		CAAATCCGGCCTGCCTGGTAGATTCAAAGCCTGGATCTATCAGCATTCA			WLGLPRSLTSAALYG
		ATTCTGCCCCGAGTTTTGTGGCCTCTGCTGATCTATGCAGTCCCAATGTC			TSNTLQLPFSGLTEEF
		AACAGTTGAGTCCCTAGAAAGGAAGATCAGTGGCTTTCTTCGAAAATGG			MVVRTREALQYRDS
		TTGGGCCTCCCACGCAGTCTTACCAGTGCTGCACTATACGGGACAAGTA			RDGKVSSACIEVRTG
		ACACCTTGCAGCTACCATTCAGTGGCCTCACAGAGGAATTCATGGTTGT			RKWNAGKAVEVAES
		ACGCACCAGAGAAGCCCTACAGTACAGGGACTCTAGAGATGGCAAGGT			RLQQKALVGTVATG
		GTCATCAGCCTGCATCGAGGTGAGGACAGGCAGGAAATGGAATGCAG			RAGLGYFPKTLVSQV
		GGAAAGCAGTGGAGGTGGCAGAGTCACGCCTGCAACAAAAGGCTCTG			KGKERHHLLQGEVRA
		GTGGGCACTGTAGCGACAGGCAGAGCGGGCTTGGGCTATTTTCCAAAG			SVEEERVSRVVGLRQ
		ACCTTAGTAAGCCAGGTCAAAGGCAAGGAAAGACACCACCTACTCCAG			QGAWTRWNTLQRRI
		GGAGAGGTTCGAGCAAGTGTGGAGGAAGAGAGTCAGTAGGGTGGT			TWANILQADFQRVR
		AGGACTCCGGCAGCAGGGAGCATGGACTAGGTGGAATACACTGCAACG			FLVQAVYDVLPSPSN

			TAGGATCACCTGGGCGAACATCTTGCAGGCGGATTTCCAACGTGTCCGT TTCCTAGTACAAGCTGTCTACGATGTACTGCCAAGCCCATCAAACCTCCA CGTTTGGGGAAAGAATGAGACACCTTCCTGCCTTCTTTGCTCTGGAAGA GGCTCTCTAGAACATCTCCTCAGCAGTTGCCCCAAGGCTCTGGCTGATG GTCGCTATCGTTGGCGCCATGACCAGGTGCTTAAGGCAATTGCTGCGAG CTTAGCTTCAGCCATTAACACGAGCAAGAACCATCGTGCTCCAAGGAAG GCAGTCCACTTCATCAAAGCTGGAGAAAAACCCCGGGCCCTCCCACAAT TAACAACAGGCCTCCTTCACAAAGCCTCGGACTGGCAGCTGGAGGTCGA CCTGGGAAAACAGCTGAGGTTTCCTCATCACATCGCTGCAACACGTCTC CGTCCAGACATTATAGCTATCTCAGAAGCTTCAAGACAGCTAATTATTCT GGAGCTTACAGTGCCGTGGGAAGAGCGTATTGAAGAAGCAAATGAGA GGAAGCGCGCTAAGTACCAGGAATTAGTGGAGGAGTGCAGGGAGAGA GGCTGGAGAACTTACTATGAGCCCATAGAAATTGGATGCAGAGGCTTT GCAGGGCGTTCACTTTGCAAAGTCCTCAGTCGTTTGGGCATTACAGGCG TGGCGAAGAAAAGGGCCATTCGATCCGCAAGCGAAGCCGCAGAGAAG GCCACAAGGTGGCTGTGGATTAAGAGGGCAGATCCGTGGACTGCTGTT GGGACACAAGTCGGGACTTGATCAACCCCGGCTGGGTCACCTGGGTGA GAGTGTATGATGTTGAGAGACCCGAAACACTCAATGATCCCAGGATACA TCACTGATGATGTGTCCCAAATGCATCCATGAGATGTTTCTTGCATAA (SEQ ID NO: 1040)	CTGACCAG CAGACGG GAAGCCCG CGACCAAC TAGTCTCC GCAAATAT TGCACACA GGGCGAC CCTATGGA GCTGATTC AGTCAAAT TTCCTCTG AGATATAC CGATAACT ATCTACAG AAACTGCA CAGTTAGT TTGGAAAG	TAGAAA CCCACA AGGCTG AGAAAT GTAGAG CATCTGT ATGGAC AATATT GATGAT TGAAAT GTTGTG ATTTTAG ATCAAA TTTAGA AATATG AAAACC GAACTA AACTAA	LHVWGKNETPSCLLC SGRGSLEHLLSSCPKA LADGRYRWRHDQVL KAIAASLASAINTSKN HRAPRKAVHFIKAGE KPRALPQLTTGLLHK ASDWQLEVDLGKQL RFPHHIAATRLRPDII AISEASRQLIILELTVP WEERIEEANERKRAK YQELVEECRERGWRT YYEPIEIGCRGFAGRS LCKVLSRLGITGVAKK RAIRSASEAAEKATR WLWIKRADPWTAV GTQVGT (SEQ ID NO: 1408)
HERO	HERO-3_BF	.	Branchiostoma floridae	CTGACCAGCAGACGGGAAGCCCGCGACCAACTAGTCTCCGCAAATATTG CACACAGGGCGACCCTATGGAGCTGATTCAGTCAAATTTCCTCTGAGAT ATACCGATAACTATCTACAGAAACTGCACAGTTAGTTTGGAAAGAGCTT TTCTACTGAAAGACAGCAAAATCCGCCACTTTAGACGAGCGTCAAGACT GCCCTCCCCATAACCAATATGGCGCTACCTGCTGTACGTTCTGGACCAGC CAGCACCTGGACACTGTTAATCACGCTGGTCATCGTCGCTGCTAAAGGT ACAGATGGTTTTATGTCTTTTAAACTGCCACTGCTGTCTACTGATACCTG GTCTGGGTATAACAATGATGTGAAAACCCTGCTAGGCCCGCTCCACCAC GAACTGGCCACAAATGAAATGTCCCCCAAACTAGCTGGGGAGGGATTC AGTGACATCATGTGCGACTTTATGGCCAGTAAACCAGAGTTCAGCCACA CTACCGAAGAAAGTCACTCAGAAGGCTATATAAGCCACGAACCACAGTC TCTCGCACAAGTAAAAACGCCTGAAAAACAAGCTACGTAAGAAGGCATTC AGAGCTGACGCAACACCTGAGGATCGAAAGGCTTTCAGAGATGCAATT AAAACATACTCCTTCATGAAGCGACAACAGAAACGAAAGGAAACTACA AAATCGGCAGCACACCAAGAGAAAGAATATCATAAGAACTTTTGGAAG TTTGCCGGAAAATGTGCAAAAGGACAGCTCGATATCCCCCCAGTAAAAC CGGCATTCTCTGTTTATTATGCAAATGAGTACTACAAAAAACAAATACTCA CACCCAACCCGTGTTGACTTCAACAAACTGCTCTGGTTTCCTCATTTGCC			

		GGTGGAGGAACAACTACCTGCGAACTCTTTGACATGTCACCTGTCAGG	AGCTTTTC	ATATAAT	ACHLFLSTLFSKLLESG
		CCGAAAGACATTAAGGCAGTCTTATCCAAACGATGCGCTACATCTGCAC	TACTGAAA	GTTTTTT	DPPTSWSSGNVSLIH
		CTGGCCCGGACGGGATCATGTATGGCCACCTCAAGCACCTGCCAGCTTG	GACAGCAA	TTAAAG	KDGSPEAAENFRMIC
		TCACCTGTTCCTTAGTACACTGTTCTCCAAACTGCTTGAGTCCGGAGACC	AATCCGCC	TAATGA	LTSCVSKIFHQILSER
		CACCGACATCATGGTCATCTGGCAACGTGTCACTTATACACAAGGATGG	ACTTTAGA	TAAGCA	WAKYMTCNDLIDPE
		TAGTCCAGAAGCTGCCGAAAACTTTCGAATGATCTGCCTTACTTCCTGCG	CGAGCGTC	ATACCC	TQKAFLTGINGCVEH
		TCTCCAAGATTTTCCACCAAATACTCTCGGAACGATGGGCAAAGTACAT	AAGACTGC	ACATTGT	VQVMREILAHAKKN
		GACTTGCAATGATCTGATAGACCCAGAAACACAAAGGCATTCCTGACC	CCTCCCCA	GCAATA	RRTVHITWFDLADAF
		GGAATCAACGGCTGTGTGGAGCATGTCCAAGTTATGCGGGAGATCTTA	TAACCAAT	CTATCTA	GSVEHELIYYQMERN
		GCACATGCCAAGAAAAACCGCCGAACAGTCCACATTACATGGTTTGACC	(SEQ ID	TGTTATG	GFPPIITTYIKNLYSRL
		TCGCGGATGCCTTTGGTTCTGTAGAACACGAACTGATCTACTACCAGAT	NO: 1164)	TCCTTTG	KGKVKGPGWESDPF
		GGAGAGAAACGGCTTCCCGCCAATTATCACCACGTACATTAAAAACCTG		TCCCCCC	PFGRGVFQGDNLSPII
		TATTCTCGCCTGAAAGGGAAAGTGAAGGGTCCAGGCTGGGAAAGTGAT		TGCATG	FLTVFQPILQHLKGVE
		CCGTTCCCGTTCGGAAGAGGAGTGTTCCAAGGAGACAAACTTGTCACCCA		TTTGGTC	QQHGYNLNDKHYVT
		TCATCTTCCTAACGGTGTTCCAGCCTATTCTACAGCATCTCAAGGGAGTA		AATAAT	LPFADDFCLITTNKRQ
		GAGCAGCAACATGGCTACAACCTCAATGACAAGCATTATGTTACACTGC		GACCAT	HQKLITQISSNTKSM
		CTTTCGCAGACGACTTTTGTCTCATAACCACAAACAAACGACAGCATCAG		CGTGTC	NLKLKPRKCKSMSIVS
		AAACTAATTACTCAAATTTCTTCCAACACAAAGTCAATGAACCTAAAGCT		CTGGGC	GKPSDISFTIDGDPVK
		AAAACCACGCAAGTGTAAGTCTATGTCTATAGTGAGCGGAAAGCCATCG		TCCGTG	TTKDAPEKFLGGYITF
		GACATCAGCTTCACAATAGATGGGGACCCTGTCAAAACGACCAAAGATG		TACCTTT	LSKTKETYDILAKTIET
		CACCGGAGAAATTCCTAGGTGGCTACATCACCTTCCTGAGTAAAACAAA		CTTTACT	TVENINKSAIRNEYKL
		AGAGACCTATGACATCCTAGCAAAGACAATAGAAACGACTGTTGAAAAC		ATGAAT	RVYMEYAFPSWRYM
		ATAAACAAATCAGCGATAAGGAACGAATACAAACTCAGGGTTTACATG		AAAGAA	LMVHDLTDTQLQKL
		GAGTACGCCTTCCCATCTTGGAGGTACATGCTGATGGTACACGACCTGA		TGATTTT	DSIHTKAIKTWLRMQ
		CAGACACCCAGCTACAAAAACTCGATTCCATCCACACAAAGGCGATCAA		ACTAC	PSATNAILYNTRGLNF
		AACATGGCTCAGAATGCAACCTAGTGCAACAAATGCAATTCTGTACAAC		(SEQ ID	KSISDLYLEAHALAYS
		ACAAGGGGTCTCAACTTCAAAAGCATCTCAGACTTGTACCTAGAAGCCC		NO:	RSVLKADEKVKHALQ
		ACGCTCTGGCCTACAGTAGGTCAGTCCTCAAAGCAGATGAGAAGGTAA		1287)	AKLDRESQWTRKMQ
		AACACGCTTTACAAGCCAAACTGGACCGCGAATCGCAATGGACTAGGA			KWGIGKCHTIHQQAI
		AAATGCAGAAATGGGGTATTGGAAAGTGTCACACCATCCACCAGCAAG			HVAKDSEWTSVRKH
		CCATCCATGTAGCAAAGGACTCAGAATGGACATCAGTACGCAAACATGT			VKQQVTDMRHDVW
		CAAACAACAAGTCACAGATATGCGTCATGACGTCTGGACTAAACATCAG			TKHQENLLQQGQML
		GAAAACCTTCTACAGCAAGGGCAGATGCTACAACTGCTTGAGGAAGAA			QLLEEEKCDLTWRSA
		AAATGCGACCTGACATGGCGGTCCGCTATGTACAACCTGCCGAGGGGC			MYNLPRGILSFAVRA
		ATCCTCAGTTTCGCTGTGCGTGCCTCCATCGACGCCCTCCCCACACTCTG			SIDALPTLCNLTTWG
		TAACCTGACCACCTGGGGAAAACGTAACACTGACAAATGTAAACTGTGT			KRNTDKCKLCGNRET
		GGCAACCGGGAAACACTCCACCACGTTCTGAACCACTGCGGTGTCGCTC			LHHVLNHCGVALQQ
		TCCAACAAGGACGGTACACATTCCGACACAACTCGGTATTGAAGCACAT			GRYTFRHNSVLKHIT

			AACGGACACCATCATAGAGTCCATTGACACCTCTCGGATCAACGCCACC ATCTATGCGGACATACAAGGTTACACAACTAACGGAGGTACCATCCCGG TCCATACAATACCCACTACCCAGAAACCAGACCTGATCATATATTTACCA GAACAGAAGACCCTCCACATCCATGAACTGACTGTACCCTTTGAAAAGA ACATCAAAACAAGTCATGACCGAAAGGTCAACAAATACAGCACCCTAGC GGCAGATTTAGAAACTGCTGGCATTTCCGCTACACTAACCTGCTTTGAA GTCGGATCAAGGGGACTCGTCACGCCAGAGAACAAGACCAGGCTTAGA ACACTGTTCAAAATAGTTAAAGCCAAACCACCGAAGACTCTGTTTACTGA TATAAGCCGCATTGCGATGTTATCGTCATATGCTATTTGGAACTCACGCC ACGAACCGTATTGGGAGTCAGAAACGCTATTGTAGAAACCCACAAGGCT GAGAAATGTAGAGCATCTGTATGGACAATATTGATGATTGAAATGTTGT GATTTTAGATCAAATTTAGAAATATGAAAACCGAACTAAACTAAATATA ATGTTTTTTTTAAAGTAATGATAAGCAATACCCACATTGTGCAATACTAT CTATGTTATGTCCTTTGTCCCCCCTGCATGTTTGGTCAATAATGACCATCG TGTCCTGGGCTCCGTGTACCTTTCTTTACTATGAATAAAGAATGATTTTA CTAC (SEQ ID NO: 1041)	AAAGCAGT AGAG (SEQ ID NO: 1165)	TAGCAT GCCACTT GGACAC AGGCCG GGGTCT GATCAG CCTCGG TCGGGT CGCCTG GAGGAG GGTGTC TGTTGC AAGACC CGAAAC ACCCTGT GAGCCC AGGAAA CAACAC TGATGA TGTGTC CAAGGT TGTGCA	DTIIESIDTSRINATIYA DIQGYTTNGGTIPVH TIPTTQKPDLIIYLPEQ KTLHIHELTVPFEKNIK TSHDRKVNKYSTLAA DLETAGISATLTCFEV GSRGLVTPENKTRLR TLFKIVKAKPPKTLFT DISRIAMLSSYAIWNS RHEPYWESETLL (SEQ ID NO: 1409)
HERO	HERO Dr	.	Danio rerio	AAAGCAGTAGAGATGACGACACATCGCGCGCAGAAGTTACAACTTCTGGT AAGACGCAGGAGGAGCCAGGCCCGGAGGCAACCCACAGTGCCCAGAG CCTCCTAGTGTCGCCAACACCTGCTGCCGGCCGCTCGCCTGCTACTCAAA GCTGCCCTCAAGTGACAGCAGCTCATAACAGTCCACACAAAGCCCCCAAAG TCAGCAAGTGGCAGTTACAAGATCTGACTGTGTTCCCTTGGCACAGCCA AGAATCCAGTGGCCCCAATCCTCCAAAGAAAGCTGAGTGGCTCCAGTTCG ACAAGGACGTGAATCAGATCCTGGAAGTGACAGGCAAGGGGGGTGTG GACCAGCGACTGTCAACAATGACCACGCTCATAGTGAACATTGCAGCTG AGCGATTCGGAACTGTGACACCCAAACCCACTCCATCGACATATACTCCA AGCCACAGAGTAAAGGAAATCAACGTCTCAGGAAAGAACTTAAGCTA CTAAAGAGGCAGTACAAGGCAGCAGGGGAAGTAGAAAGAGCGGGCCT AGAAGATCTGAGAGGAATCCTGAGGAAACAGCTCGTGAACCTATGTAG GGCAGAGTATCACAGGAAGAGGCGGAGAGAGAGCAAGGAAAAGG GCAGCATTTTTGGCCAACCCTTTCAAGTTGACCAAGCAGCTCCTTGGCCA AAAGAGGACTGGCAAACTCACCTGCTCCAAGGAGGCTATCAACAATCAC CTCAAGGCCACTTATTCTGACCCGAATAGAGAACAACCCCTGGGGCCTT GCGGTGCACTGCTGACACCACCTGAGCCCACATCAGAGTTCAACATGAA GGAACCCTGCCGGAGTGAAGTAGAGGAAGTGGTGAGGAGAGCAAGGT CAAGCTCAGCACCAGGCCCAAGCGGAGTGCCTTACAAGGTATATAAGA ACTGCCCAAAGCTTCTACACAGGCTCTGGAAGGCCCTGAAAGTCATATG GAGAAGAGGGAAGATTGCCCAGCCATGGAGGTATGCGGAGGGAGTGT ACATCCCAAAAGAGGAGAAGTCGGAGAACATCGACCAGTTTCGAGTCA			DTIIESIDTSRINATIYA DIQGYTTNGGTIPVH TIPTTQKPDLIIYLPEQ KTLHIHELTVPFEKNIK TSHDRKVNKYSTLAA DLETAGISATLTCFEV GSRGLVTPENKTRLR TLFKIVKAKPPKTLFT DISRIAMLSSYAIWNS RHEPYWESETLL (SEQ ID NO: 1409)
						MTTHRAEVTTSGKT QEEPGPEATHSAQSL LVSPTPAAGRSPATQ SCPQVTAAHNSPQSP QSQQVAVTRSDCVP LAQPRIQWPQSSKKA EWLQFDKDVNQILEV TGKGGVDQRLSTMT TLIVNIAAERFGTVTP KPTPSTYTPSHRVKEI KRLRKELKLLKRQYKA AGEVERAGLEDLRGI LRKQLVNLCRAEYHR KRRRERARKRAAFLA NPFKLTKQLLGQKRT GKLTCSKEAINNHLK ATYSDPNREQPLGPC GALLTPPEPTSEFNM KEPCRSEVEEVVRRA RSSSAPGPSGVPYKV YKNCPKLLHRLWKAL KVIWRRGKIAQPWR

				TCAGGA GATGTTT CTGTAA C (SEQ ID NO: 1288)	YAEGVYIPKEEKSENI
					DQFRVISLLSVESKIFF
					SIVAKRLSNFLLSNKYI
					DTSMQKGGIPGVPG
					CLEHTGVVTQLIREA
					REGRGDLAVLWLDLT
					NAYGSIPHKLVEVALE
					KHHVPQKVKDLIIDYY
					SKFSLRVSSGQLTSD
					WHQLEVGIITGCTISV
					TLFALAMNMMVKA
					AETECRGPLSKSGVR
					QPPIRAFMDDLTVTT
					TSVPGARWILQGLER
					LVAWARMSFKPAKS
					RSLVLRKGKVRDEFR
					FRLGQHQIPSVTERP
					VKSLGKAFNCSLNDR
					DSIRETSTAMEAWLK
					AVDKSGLPGRFKAW
					VYQHGILPRLLWPLLI
					YEVPMTVVEGFEQK
					VSSYLRRWLGLPRSLS
					NIALYGNTNKLKLPFG
					SVREEFIVARTREHLQ
					YSGSRDAKVSGAGIVI
					RTGRKWRAAEAVEQ
					AETRLKHKAILGAVA
					QGRAGLGSLAATRYD
					SASGRERQRLVQEEV
					RASVEEERTSRAVAM
					RQQGAWMKWEQA
					MERNVTWKDIWTW
					NPLRIRFLIQGVYDVL
					PSPSNLYIWGRVETP
					ACPLCSKPGTLEHILS
					SCSKALGEGRYRWR
					HDQVLKSIAEAISKGI

			GACATAATCTTGGTCTCTGAGGCCACAAAGCAATTAATCCTGCTGGAGC TGACGGTGCCCTGGGAGGAGAGGATGGAGGAGGCTCAGGAGAGAAA GAGGGGAAAATATCAGGAGCTAGTGGAGCAATGTAGGGCGAATGGAT GGAGGACCAGGTGCATGCCAGTGGAAGTGGGCAGTAGGGGATTTGCC AGCTACACCCTGAGCAAGGCCTATGGTACACTGGGAATAACAGGCACA AACCGAAGAAGAGCCCTAAGCAACAACGTGGAAGCAGCGGAAAAAGC ATCCAGATGGCTCTGGTTGAAGAGGGGGGAACAGTGGGGGCAGTAGC ATGCCACTTGGACACAGGCCGGGGTCTGATCAGCCTCGGTCGGGTCGC CTGGAGGAGGGTGTCTGTTGCAAGACCCGAAACACCCTGTGAGCCCAG GAAACAACACTGATGATGTGTCCAAGGTTGTGCATCAGGAGATGTTTCT GTAAC (SEQ ID NO: 1042)			KDSRYRQATAKVIQFI KEGQRPERTAKNCSA GLLSTARDWVMTVD LERQLKIPPHITQSTLR PDIILVSEATKQLILLEL TVPWEERMEEAQER KRGKYQELVEQCRA NGWRTRCMPVEVG SRGFASYTLSKAYGTL GITGTNRRRALSNNV EAAEKASRWLWLKR GEQWGQ (SEQ ID NO: 1410)
HERO	HEROFor	.	Takifugu rubripes	AGACTAGGTGACAACCAAGAACAGTTWGGTCGACTACTGGAAAGACA GTTGGCAGCTCGGAAAGACGGCACCCGGGACAGTATGGGTTAGCACCC CAGCCTGTATCTTTCGCGAGAAGGAACCCAAACAAGCTACGGAAAGCCC TACAGAGAAACACCCCCAGGAGATCCCGAGAGGGGGGGAGGATGAGA TCTCCAATCGGACGGACCTAACGTTAATGACCCCTGCAATGGAAATGAC TACGACAGTAACATGTATCTGCAGCAAGCTGTGCAAGAACCAGCGTGGC TTAAAGATCCATCAGGCCAGAATGAAATGTCTGGAGCGGGAGGTTGAG GTGCAACGCACAGGTCCTGGACCTGGTGAGACGCAGGAGGAGCCCGG ACAGGAGGCAACCCACAGATCCCAGTCCCTCCACGTACCGGAGCCTCCC AACCCTAACAGAGTAGTTCAACAGCAGCGGATTAAGTGGCCCCCAGCAA ATAGACGGAGTGAGTGGCTGCAGTTTGATGAGGATGTGTCCAACATCA TCCAAGCCACAGCCAAAGGAGATGTCGACAGCAGACTCCAGGCGATAA GTACCATCATCGTCAGCTATGGCTCAGAAAGATTTGGACGGATCGAGAA GGGCAACACTGAGACCACCTCTTACACCATGAACCGCAGGTCCTTTAAG ATACACCAACTGCGCAAGGAGCTGCGAACCCTCAAGAAACAGTTCAAG AGAGCTKCTGATGGGGACAAGCAAGCTTTAAAAGAGCTGTATAACATCC TGCGGAAGAAGTTGAAAACTCTCCGCAGAGCAGAGTGGCACAGGAGGC GCGGGAGAGAGAGCAAGGAAGCGAGCAGCCTTCATTGCCAATCCCT TCCGGTTTTCTAAACAGCTGCTCGGGGACAAGCGGAGTGGCCGACTTGA GTGCTCAAGGGAGGAAGTGAATCGCTTCCTCCAAAACACCATGAGCGA CCCACTGAGGGGTCAAGACCTAGGACCCAACAGAGCGCTCATCAGCCCT GCCCCACCATCGGCAGAGTTCAAGCTGGCAGAGCCTAGTTTGAAGGAG GTTGAAGAAGTCATCAAGGCAGCCCGTTCTGCATCTTCCCCGGGCCCCA GTGGTGTACCTTACCTCGTCTACAAGCGCTGTCCAGAAATTCTCCGGCAT CTGTGGAAGGCCTTGAAAGTGATCTGGCGAAGGGGGAGAGTAGCCGA	AGACTAGG TGACAACC AAGAACA GTTWGGT CGACTACT GGAAAGA CAGTTGGC AGCTCGGA AAGACGG CACCCGGG ACAGTATG GGTTAGCA CCCCAGCC TGTATCTT TCGCGAGA AGGAACCC AAACAAGC TACGGAAA GCCCTACA GAGAAAC ACCCCCAG GAGATCCC GAGAGGG ATGAGATC	TGATCA CCCCGG CTGGGT CGCCTG GGCGAG GGTGTA TGATGT CGTGAG ACCCGA AACACC CTATGA ACCCAG GATACA TCCTGAC GATGTG TCCCAGT GCATCC AGGAGA TGTAKCT TTAAGT (SEQ ID NO: 1289)	MTPAMEMTTTVTCI CSKLCKNQRGLKIHQ ARMKCLEREVEVQR TGPGPGETQEEPGQ EATHRSQSLHVPEPP NPNRVVQQQRIKWP PANRRSEWLQFDED VSNIIQATAKGDVDS RLQAISTIIVSYGSERF GRIEKGNTETTSYTM NRRSFKIHQLRKELRT LKKQFKRAXDGDKQ ALKELYNILRKKLKTLR RAEWHRRRGRERAR KRAAFIANPFRFSKQL LGDKRSGRLECSREE VNRFLQNTMSDPLR GQDLGPNRALISPAP PSAEFKLAEPSLKEVE EVIKAARSASSPGPSG VPYLVYKRCPEILRHL WKALKVIWRRGRVA DQWRCAEGLWIPKE EDSKNINQFRTISLLS VEGKVFFSIVSRRLTE

		CCAGTGGAGGTGTGCTGAGGGACTTTGGATACCCAAGGAGGAGGACTC	TCCAATCG GACGGACC TAACGTTA (SEQ ID NO: 1166)		FLLKNNYIDTSVQKG
		GAAAAACATCAACCAGTTTCGGACTATCTCACTACTGAGTGTGGAAGGG			GIPGVPGCLEHNGVV
		AAGGTGTTTTTTAGCATCGTCTCCCGAAGACTGACCGAGTTTCTCCTCAA			TQLIREAHESKGELAV
		GAACAACTACATCGACACTTCAGTGCAGAAGGGTGGGATCCCTGGAGT			LWLDLTNAYGSIPHK
		CCCCGGCTGTCTAGAGCACAATGGTGTAGTCACACAGCTCATCAGAGAG			LVELALHLHHVPSKIK
		GCCCATGAGAGCAAAGGAGAACTAGCGGTTTTGTGGTTGGACCTGACT			DLILDYYNNFRLRVTS
		AACGCCTACGGGTCCATCCCACACAAGCTAGTTGAGCTTGCGCTACACC			GSVTSDWHRLEKGII
		TACACCATGTTCCCAGTAAGATCAAGGACCTGATTCTGGATTACTATAAT			TGCTISVVLFVLAMN
		AACTTCAGGCTCAGGGTCACTTCAGGGTCAGTAACCTCAGACTGGCATC			MVVKAAEVECRGPL
		GCCTTGAGAAAGGAATAATAACAGGCTGTACCATCTCCGTCGTTCTCTTC			SRSGVRQPPIRAYMD
		GTACTGGCGATGAATATGGTGGTAAAGGCGGCTGAGGTGGAGTGCAG			DLTVTTTSVPGCRWI
		AGGGCCTCTATCCAGATCAGGTGTTCGACAGCCCCCCATAAGAGCCTAC			LQGLERLILWARMSF
		ATGGACGACCTTACCGTCACAACAACATCAGTCCCAGGGTGTAGGTGGA			KPTKSRSMVLKKGKV
		TCTTGCAGGGTTTGGAGAGACTCATCCTATGGGCTAGGATGAGTTTTAA			VDKFRFSISGTVIPSIT
		GCCCACCAAGTCAAGGTCCATGGTACTGAAGAAGGGGAAAGTGGTGGA			EQPVKSLGKLFDSSLK
		CAAATTCCGATTCTCAATCTCAGGAACCGTAATTCCATCGATCACGGAGC			DTAAIQKSTEELGGW
		AACCAGTCAAGAGCCTGGGAAAGCTCTTTGACTCCAGCCTGAAGGACAC			LTKVDKSGLPGRFKA
		TGCTGCTATCCAGAAGTCTACGGAAGAGCTTGGAGGGTGGCTCACTAA			WIYQYSILPRVLWPLL
		GGTGGACAAGTCTGGCCTGCCTGGTAGATTTAAAGCCTGGATCTACCAG			VYAVPVTTVESFERKI
		TACTCCATCCTTCCCAGAGTCCTGTGGCCTCTCCTCGTGTATGCAGTCCC			SSFLRRWLGLPRSLNS
		AGTAACAACAGTGGAATCCTTTGAAAGGAAGATCAGCAGCTTTCTGCGC			AALYGTSNTLQLPFS
		AGATGGCTGGGTCTTCCTCGCAGCCTCAACAGCGCTGCACTGTACGGGA			GLTEEFKVARTREAL
		CAAGTAACACCCTGCAGCTACCCTTCAGTGGGCTCACTGAAGAATTTAA			QYRDSRDCKVSSAGI
		GGTGGCACGCACAAGAGAAGCCCTACAGTACAGAGACTCCAGGGACTG			EVKTGRKWKAEKAV
		CAAGGTGTCATCAGCCGGGATTGAGGTGAAGACAGGAAGGAAGTGGA			XVAESRLRQKALVGA
		AGGCAGAAAAGGCAGTGGAKGTGGCTGAGTCACGCCTAAGGCAAAAG			VATGRTGLGYFPKTQ
		GCACTAGTTGGGGCCGTGGCAACAGGAAGAACAGGCTTGGGCTACTTC			VSHARGKERNHLLQE
		CCAAAGACCCAAGTCAGCCATGCCCGGGGCAAAGAGAGAAACCACCTA			EVRAGVEEERVGRAV
		CTTCAGGAGGAGGTCCGAGCAGGCGTGGAGGAAGAGCGAGTGGGTAG			GLRQQGAWTRWES
		GGCAGTGGGACTCCGGCAGCAGGGGGCATGGACAAGGTGGGAGAGC			ALQRKVTWSNIMQA
		GCGTTACAGCGCAAAGTTACCTGGTCAAACATCATGCAGGCAGACTTCC			DFHRVRFLVAAVYDA
		ACCGCGTCCGGTTCCTTGTGGCGGCAGTCTACGATGCCCTCCCCAGCCC			LPSPANLHAWGKSET
		AGCAAACCTCCATGCGTGGGGAAAGAGTGAGACACCCACCTGTTCCCTT			PTCSLCSGRGSLEHLL
		TGCTCCGGAAGAGGCTCCCTGGAACATCTCCTTAGCAGCTGCCCAAAGT			SSCPKSLADGRYRWR
		CCCTGGCTGATGGTCGCTATCGCTGGCGCCACGACCAGGTACTCAAAGC			HDQVLKAVAESIALAI
		AGTGGCTGAGAGCATAGCCTTGGCCATTAGCACCASCAAACACCATCAT			STXKHHHAPKKAISFI
		GCTCCGAAGAAGGCAATCTCCTTCATAAAAGCTGGAGAGACCTCGTG			KAGERPRAGPQITTG
		CAGGCCCACAGATAACAACGGGACTCCTCCACACAGCTMCTGATTGGC			LLHTAXDWQLHVDL

			AACTGCACGTTGACCTGGGAAAACAACTGATATTCCCCCAGCACATCGC AACAACGTCTCTACGGCCAGACATGATCATCATCTCAGAGGCTTCGAAA CACCTGATCATGCTGGAGCTTACAGTGCCCTGGGAAGAGCGGATTGAG GAAGCCAACGAAAGGAAACGTGCCAAGTATCAGGAGCTGGTGGAGGA GTGCAGGGGCAGGGGCTGGAGGACCTTCTACGAGCCCATAGAAGTTGG CTGTAGAGGCTTTGCAGGACGCTCCCTCTGCAAAGCCTTTGGCCGACTG GGAGTCACAGGGACAGCCAAAAAGAGGGCCATTAAAKCCGCGAGTGA AGCTGCAGAGAGAGCCACGAGGTGGSTGTGGCTKAAAAGGGCAGATCC GTGGGTTGCTACTGGGACACAAGCCGGGTCTTGATCACCCCGGCTGGG TCGCCTGGGCGAGGGTGTATGATGTCGTGAGACCCGAAACACCCTATG AACCCAGGATACATCCTGACGATGTGTCCCAGTGCATCCAGGAGATGTA KCTTTAAGT (SEQ ID NO: 1043)				GKQLIFPQHIATTSLR PDMIIISEASKHLIML ELTVPWEERIEEANE RKRAKYQELVEECRG RGWRTFYEPIEVGCR GFAGRSLCKAFGRLG VTGTAKKRAIKXASE AAERATRWXWLKRA DPWVATGTQAGS (SEQ ID NO: 1411)
HERO	HEROTn	.	Tetraodon nigroviridis	AGATTGGTCTGGCTAAGCCAGTGACGTCCAGGAACAGACTGGCTGACG ACCACGAATAGAGTGGTGACAGCTTGGATAGACAGCTGACAGCAGGGA AAGACGGCAACCGGGGCAGGAAGGGCTAGCAACCCAGCCTGCATCTTC CGTGAGGAAGAACCCAAAACTTGCTACGAAGAGCCCGAAGCAAAGATA CCCCCAGGGGAGCCCGAGAGGGGGGGAGAATGAGCTCCCCAAACGGA CGGATAACATGGCAACGACCCAGGCTAGCGTTAAACCGACAGCGGTTG CCACATGTGTATGTGGCAAAATCTGCAAAAACCCACGAGGTCTGAAGAT CCACCAGACCAAGATGGGGGTGCTTGGCAAGTGTGCAACCAGAGCAGCG CGCAAGGTTCAGCCTCAGCGAGTCGCGGGAGGTGCCAGCCAGGGCCGA GCCCTATGGCCCTCAGCAACCGCATTCTCCTGAGGCCCTTGGTGAGACG CAGGAGGAGCGGGGCCAGGAGTCACCCCACAGTGCCCAGAACCTCCGT GCTCAGGTAGCACAAGCGCCAGACAACCCACAACACCACCGGCGGGTT AAGTGGCCCCCAGCCAGCAAAGTGAGCGAGTGGCAGCAGCTTGATGAG GATTTGGAAGGTATTCTGGAGTCCACCGCAAAAGGTGGAGTAGACAGA AAACTCCAAACAATGACCACGCTGGTCATCAGCTTTGCCACCGAGAGAT ATGGTACAATGGAGAAAACGCGCTGCTCCAGAGAAGTACACCAAAAACC GCAGGGCAGAAAAGATCTCCCAACTGCGGCAGGAACTTCGGGTCCTGA AAAAGCAGTTCAAGGGCGCCAGCGAGGATCAGAAGCCAGGATTGGCA GAGCTTCGTTGCACCCTTAGGAAAAAACTGCTTACCCTTCGCCGAGCAG AGTGGCACCGGAGACGGGCCAAGGAAAGAGCCAAGAAACGCGCTGCA TTTTTAGCCAACCCTTTTGGGTTCACTAAACAACTTTTAGGCCAGAAGCG TAGCGCCCACTTGGAATGTGCAAAAGAGGAGGTTGATTCCTACCTCCAC GACACATTCAGTGACGCAGAACGGGAGAACAGCCTAGGCGAATGTAGA GTGCTGATCAGTCCACCTGAGCCAGCCTGCAGTTTCAACACCAAGGCTC CAACTTGGAAAGAAATCCAAACTGTGGTCAGGGCTGCAAGAAACAACT CAGCTCCTGGACCCAATGGAGTCCCATATCTGGTGTACAAAAGATGCCC	AGATTGGT CTGGCTAA GCCAGTGA CGTCCAGG AACAGACT GGCTGACG ACCACGAA TAGAGTGG TGACAGCT TGGATAGA CAGCTGAC AGCAGGG AAAGACG GCAACCGG GGCAGGA AGGGCTA GCAACCCA GCCTGCAT CTTCCGTG AGGAAGA ACCCAAAA CTTGCTAC GAAGAGC CCGAAGCA AAGATACC CCCAGGG	TGATCA CTCCCA GTCGGG TCGCCT GGGTGA GGGGGT CTGATG TTGAAA GACCCG AAACCC CCGATG ACCCCA GGTACT ATCACT GACGAT GTGTCC AAGACA TGCATC AATAGG TGTATTT AGAAAT C (SEQ ID NO: 1290)	MATTQASVKPTAVA TCVCGKICKNPRGLKI HQTKMGCLASVQPE QRARFSLSESREVPA RAEPYGPQQPHSPEA LGETQEERGQESPHS AQNLRAQVAQAPDN PQHHRRVKWPPASK VSEWQQLDEDLEGIL ESTAKGGVDRKLQT MTTLVISFATERYGT MEKRAAPEKYTKNR RAEKISQLRQELRVLK KQFKGASEDQKPGLA ELRCTLRKKLLTLRRA EWHRRRAKERAKKR AAFLANPFGFTKQLL GQKRSAHLECAKEEV DSYLHDTFSDAEREN SLGECRVLISPPEPAC SFNTKAPTWKEIQTV VRAARNNSAPGPNG VPYLVYKRCPKLLARL WKILRVIWRRGKVA HQWRWAEGVWVP KEEKSTLIEQFRTISLL

		CAAACTCCTAGCCCGGCTCTGGAAGATCCTAAGGGTGATCTGGAGAAG	GAGCCCGA		NVEGKIFFSILSHRLSD
		GGGGAAGGTCGCCCATCAATGGAGATGGGCGGAAGGGGGTGTGGGTTC	GAGGGGG		FLLKNQYIDSSVQKG
		CGAAGGAGGAGAAGTCAACCTTGATAGAGCAGTTTAGGACCATCTCACT	GGAGAAT		GIPGVPGCLEHCGVV
		GCTCAATGTCGAGGGGGAAGATATTCTTTAGTATCCTCTCCCATCGTCTAT	GAGCTCCC		TQLIREAREGRGSLA
		CAGACTTCCTCCTTAAGAACCAGTACATCGACTCCTCGGTGCAAAAGGG	CAAACGGA		VLWLDLANAYGSIPH
		GGGGATCCCTGGGGTACCAGGGTGTTTAGAACACTGTGGCGTGGTGAC	CGGATAAC		KLVEMALARHHVPG
		ACAACTAATTAGGGAGGCGCGCGAAGGGAGAGGTAGCCTGGCCGTACT	(SEQ ID		PIKTLIMDYYDSFHLR
		TTGGCTGGACTTAGCTAACGCTTATGGCTCCATACCCCACAAGCTGGTG	NO: 1167)		VTSGSVTSEWHRLEK
		GAAATGGCATTAGCGAGGCACCATGTCCCAGGCCCGATCAAGACTCTG			GIITGCTISVIIFALAM
		ATCATGGACTACTATGATAGCTTCCACCTGAGAGTCACGTCAGGCAGTG			NMLAKSAEPECRGPI
		TCACATCTGAATGGCACCGACTAGAGAAAGGGATCATCACTGGATGCAC			TKSGIRQPPIRAFMD
		CATCTCAGTGATAATATTCGCCCTGGCCATGAATATGCTGGCCAAGTCG			DLTVTTTSVPGCRWI
		GCTGAGCCAGAGTGCAGAGGACCCATAACCAAGTCAGGCATTCGCCAG			LQGLERLMTWARM
		CCCCCCATCAGAGCATTCATGGATGATCTGACAGTAACAACAACGTCAG			RFKPGKSRSLVLKAG
		TTCCAGGGTGCCGTTGGATCCTCCAGGGCCTGGAGAGGCTTATGACTTG			KVTDRFRFYLGGTQIP
		GGCCCGTATGCGCTTTAAACCTGGAAAATCTAGGTCCTTAGTCCTGAAG			SVSEKPVKSLGKMFD
		GCAGGGAAGGTGACCGACCGCTTCCGCTTCTACCTGGGAGGGCACCCAG			GSLKDAASIRETNDQ
		ATTCCATCAGTCTCTGAGAAACCGGTGAAAAGCCTAGGTAAAATGTTCG			LGHWLTLVDKSGLPG
		ACGGCTCCTTAAAGGATGCCGCTTCCATCAGGGAAACCAATGATCAGCT			KFKAWVYQHGILPRI
		GGGGCACTGGCTGACGTTGGTCGATAAGTCAGGTCTTCCGGGGAAATT			LWPLLVYEFPISTVEG
		CAAGGCATGGGTATACCAGCATGGTATCCTACCTAGGATACTGTGGCCA			LERRVSSCLRRWLGL
		CTGCTGGTGTATGAATTTCCAATTTCCACCGTGGAAGGGCTTGAGAGGA			PRSLSSNALYGNNNK
		GGGTCAGCAGCTGCCTCAGGCGTTGGCTGGGACTACCTAGGAGTCTGA			LTLPFSSLAEEFMVTR
		GCAGCAATGCCCTCTACGGTAACAACAACAAGCTGACACTCCCCTTCAG			AREVLQYRESKDPKV
		CAGCCTGGCAGAGGAATTCATGGTTACCAGAGCTAGGGAAGTTCTCCA			ALAGIEVRTGRRWRA
		GTACAGGGAGTCCAAGGATCCCAAGGTAGCTCTTGCCGGCATTGAGGT			QEAVDQAESRLHHK
		GCGGACTGGCAGAAGGTGGAGGGCTCAGGAGGCAGTGGACCAGGCAG			ELVGAVATGRAGLGT
		AATCTCGGCTGCACCACAAAGAGCTTGTGGGAGCCGTGGCGACTGGCC			TPTTHLSRLKGKERR
		GTGCAGGCCTGGGAACAACACCGACCACCCACCTCAGCAGGCTCAAGG			DQVQLEVRASIEEQR
		GCAAGGAAAGGCGGGATCAGGTCCAACTAGAAGTGAGGGCCAGTATT			ASQWVGLRQQGAW
		GAGGAACAGCGAGCTAGTCAGTGGGTGGGGCTGAGGCAGCAAGGCGC			TRWEEAMARKISWP
		TTGGACTAGGTGGGAAGAGGCCATGGCCAGAAAGATCTCATGGCCTGA			ELWRAEPLRIRFLIQS
		GCTGTGGAGGGCTGAGCCCTTGCGCATCCGCTTCCTTATTCAGTCAGTTT			VYDVLPSPSNLFLWG
		ATGACGTCTTGCCCAGCCCATCAAACCTCTTCCTGTGGGGCAAGGTGGA			KVESPSCPLCQGRGT
		ATCCCCATCATGTCCCTTGTGCCAGGGAAGGGGCACCTTGGAGCACATC			LEHILSSCPKALGEGR
		CTCAGCAGCTGTCCCAAAGCACTTGGAGAGGGTCGCTATCGCTGGCGTC			YRWRHDQVLKAIAES
		ACGACCAGGTGCTGAAGGCAATCGCTGAGTCTATCAGCTCCGCCATGGA			ISSAMEYSKRLPLPGR
		GTACAGCAAGCGCCTACCCTTACCGGGACGCGGAGTTAGGTTTGTCAG			GVRFVRAGEQPPPQ

			GGCCGGTGAACAACCTCCTCCCCAACCAAGGGCCCAACCAGGCCTCCTT GCAACAGCTAGGGACTGGCAACTAAGGGTTGACCTGGGGAAACAATTA AAGTTCCCGGAAAACATCGTAGAAACCAACCTGAGGCCAGACATTGTTC TGCACTCACAGTCGTCCAAGCAAGTTATTTTGCTGGAGCTGACTGTGCCC TGGGAGGAGAGAATGGAGGAAGCGTATGAAAGGAAGGCAGGGAAGT ACGCTGAGCTGGTGGAGGATTGCCGCAGAGCAGGGTGGCGCAGTAGA TGCCTGCCTATAGAGGTTGGGGGTAGGGGCTTTGCAGGGAAGTCACTC TGCAAGGCCTTTAGCCTCCTGGGCATCACAGGCATGCGCAGGAGGAAA GCCATCTGCGCGGCCTCAGAGGCTGCAGAGAGGGCGTCCAGATGGCTG TGGATCCAGCGGGACAAGCCGTGGACGAGCGCTTCTTGGACACAGGCC GGGAACTGATCACTCCCAGTCGGGTCGCCTGGGTGAGGGGGTCTGATG TTGAAAGACCCGAAACCCCCGATGACCCCAGGTACTATCACTGACGATG TGTCCAAGACATGCATCAATAGGTGTATTTAGAAATC (SEQ ID NO: 1044)				PRAQPGLLATARDW QLRVDLGKQLKFPEN IVETNLRPDIVLHSQS SKQVILLELTVPWEER MEEAYERKAGKYAEL VEDCRRAGWRSRCL PIEVGGRGFAGKSLC KAFSLLGITGMRRRK AICAASEAAERASRW LWIQRDKPWTSASW TQAGN (SEQ ID NO: 1412)
NeSL	LIN9_SM	.	Schmidtea mediterranea	AAACGACATCATGAACGCTTGGCCGCAACAATCCAGTTATCCCTGCGGT AACATTGTGGAACTCATAAGACAAGTACTAAAAGAAGAATTAGAAAAAT TAGAAGAAAAAATTGAAAATAATTATTTATAAAATTTAAAATTTAAAT AAATTTAAAAATTTAAATTTAAATTTAAATGAAGATAAAAATTTATTTAA TCCAATAAATAATCAAGAAAATCAAGAAAATGATGGATTCAAGACAATT AAATACTCCAAAAATAAGAAAATATCAGAACCCAAAAATGACAAACGAC ATCATGAAAAGCTACAACTACGCGGTTTTGAGCGATGTCACGCCTCAAG AAACCACTCAAACAACAACCCACTTAAATGTCGATATAGACAATGAAAC CACCCAACCAAAACAGCCACTTACGAAGTCTGGCAAACCAAAATCTAAA CCAATTGCGGTATCATACAAATTTAAAGATGCCACCTTCATCTGGGACAC TACCCCACAAACAAATCCACCAAGAGATTGCACCAAACTTATTGATAAA ACAAGACCAAGAAAGACCATCTTCAAAAAATCAGCATTTCAAAGCTACC TCAAAAAGAACTGTCCAATGAGACATTTGTGGAAGTAAAAACCTTCCT CATGGCAACTCACAAATATCGTTTTAAAGACGAAAACTCAAGACTCTTG GCATACCGAATAATTAATCGCTATGTCATGGAGACAGCAAATGAATTCA AAGAAACCGAATTTGACATGGCTCGCTTTGCCAAATTCTTCACAATCCCA GAGAATTGGTTAAAACATCTAAAACCATACTCTACAGCTACCGAAACAT CACCGGCTGATAGAATAAAAGTACAAAAATTAGTGGATCTCACATGCAG ATACCCATTCAAAACTCAAGAAGAGCAAACAAGTGTAGCAAACTTCCTA CACTTCTTCACCCAAAGATCAATAATTGGAATCTCAAGAGATTATAAATT CCAAAAATTTATACCATTTATGGCAAGAAAAAACACCAGGCCGGAGACA ACCTCCACTATGGTTACGACTTCTCCAACAGAACAAAACAGACTACCAAT GGTAATAATCACACCACTTGAAGAACCAAAAAGTGAACATCGTAGACCA GAGAAAAGAGGCGCAAGCAATGACACAATTGTGCTTAGCGACGAAGAG	AAACGACA TCATGAAC GCTTGGCC GCAACAAT CCAGTTAT CCCTGCGG TAACATTG TGGAACTC ATAAGACA AGTACTAA AAGAAGA ATTAGAAA AATTAGAA GAAAAAAT TGAAAATA ATTTATTTA TAAAATTT AAAAATTT AAATAAAT TTAAAAAT TTAAATTT AAATTTAA ATGAAGAT AAAAATTT	TAAAAT GGCAAA AAGATA TTTCAAG ATGAAT TGTGGA CTCATCT AAAAAA TGACCA CCTTGA GTCCAA ATATGC CTAGCT ATCATG GTTGCT GATGGA AACAGT AAGGCA CCTGAT AGCTAA CTTTTCA CTGTGA ATATCTT CAGATA	MMDSRQLNTPKIRK YQNPKMTNDIMKSY NYAVLSDVTPQETTQ TTTHLNVDIDNETTQ PKQPLTKSGKPKSKPI AVSYKFKDATFIWDT TPQTNPPRDCTKLID KTRPRKTIFKKSAFQS YLKKELSNETFVEVKT FLMATHKYRFKDENS RLLAYRIINRYVMETA NEFKETEFDMARFAK FFTIPENWLKHLKPYS TATETSPADRIKVQKL VDLTCRYPFKTQEEQ TSVANFLHFFTQRSII GISRDYKFQKFIPFMA RKNTRPETTSTMVTT SPTEQNRLPMVIITPL EEPKSEHRRPEKRGA SNDTIVLSDEEFPLLK RRTLPTRKSKNPTGA GNVPTETECTDEVKFI LNNEYQIECKECGKV

		TTCCCACTACTTAAAAGGAGAACTCTTCCAACCAGAAAATCCAAAAATCC	ATTTAATC CAATAAAT AATCAAGA AAATCAAG AAA (SEQ ID NO: 1168)	TTCACA GTGACA CGAAAG GACACC ACTAGT AAAAAC CACTAG TTTTTTC TGACAC CTCTTGC TACAAA CTCTGTA AAAATC AAAAGG ATCGAT AGGCCG CGCTTTC GTACTG AAAATC AAGGAA GCTTTTC CCCTTTT AGTCAA CACCAG GTTTCTG TCCTAGT TGAGCT TCCCTTG GGACAT CTGCGT TACCATT TGACAG ATGTAC CGCCCC AGTCAA	WENVRNGLNHLRQK HDFPNRTDVMVSCV RCEVPIKGAECVNHIK NHKKDDKEESEAGSL VANTQDIPNESSLSQ AAIEVYLRNILKMKEN QERNIQYLEPSTANFL INRNLRAFYQNVKIEK LIGWEQVIWLIHWN KCHWIVYLANCDSKT SVILDSDNQMTLQQ RCNIKAKFDKFLEGTF EEKTVLGTLERKVPQ QPNNFDCGIYVIQYIS DFLKDPQRIDYHTPD SKRIRKEIGELILEEMIK NPASKIKNPNKEIQSL LQKFRLLQINVNDVF HWFAAEYQKSLPKIR TKRDGKLNKLSCSYQI QRLFGLAPKRAVKEIY FQETSTADLETRVLN EHFKKDESTMKECKI KNGNHYQDWITKAQ IDNKEILEALKNSTDS APGEDNIPLRQWIIW NNDGVLFDMFNYIK RTHDIPDMWKNYTT TLLIKPGKSQESNIPA NWRPISILPTSYRIFM KVLNKRVLEWANRG ELISKWQKAVDKAN GCDEHSYVIQALIEKA NRSYYKNEQCHLAFL DLADAFGSIPFQVIW HTLKNMGMDEETIN LLKEIYKDCSTKYKCG KNESEKIKITKGVRQG
		TACTGGTGCAGGAAATGTACCWACAGAAACCGAATGCACTGATGAAGT				
		TAAATTCATCCTCAACAATGAATACCAAATAGAATGTAAAGAGTGTGGA				
		AAAGTGTGGGAAAACGTACGAAATGGATTAAACCACCTTCGTCAAAAAC				
		ACGATTTCCCAAACCGAACAGATGTTATGGTATCTTGCGTAAGATGTGA				
		AGTACCGATCAAAGGAGCAGAATGTGTAAATCACATTAAAAATCACAAA				
		AAAGATGACAAAGAAGAAAGTGAAGCSGGGAGTCTTGTGGCTAACACT				
		CAAGACATCCCAAATGAAAGTAGCTGACTGTCACAAGCCGCAATCGAAG				
		TATATCTGAGGAATATTCTGAAAATGAAAGAAACCAGGAAAGGAATA				
		TTCAATATCTTGAACCTAGTACTGCGAATTTCCTCATAAATAGGAACCTC				
		AGAGCATTTTATCAAAACGTCAAAATCGAAAAGCTTATCGGATGGGAAC				
		AAGTCATCTGGCTTATACATTGGAACAAATGTCATTGGATTGTATACCTA				
		GCTAATTGCGACTCAAAAACCTCTGTTATCTTGGACTCTGACAACCAAAT				
		GACATTACAGCAAAGATGTAACATAAAAGCCAAATTTGACAAATTCCTA				
		GAAGGTACCTTTGAAGAAAAAACAGTGCTTGGAACCCTAGAAAGAAAA				
		GTTCCTCAGCAACCAAACAACTTCGATTGCGGTATATATGTGATACAATA				
		CATCAGCGACTTTCTTAAAGACCCACAAAGAATAGATTATCATACACCCG				
		ACTCCAAAAGAATTAGAAAAGAAATAGGAGAATTAATATTAGAAGAAA				
		TGAAAAACCCTGCCTCAAAAATCAAAAATCCAAACAAAGAAATACAATC				
		TTTACTCCAAAAATTCAGACTACTGCAAATCAATGTGAATGATGTATTCC				
		ATTGGTTTGCGGCTGAATACCAAAAATCTCTACCGAAGATACGTACCAA				
		AAGAGATGGAAAACTGAATAAACTAAGCTGCTCCTATCAAATCCAAAGA				
		TTATTTGGTCTAGCTCCTAAAAGAGCAGTCAAAGAAATATATTTCCAAGA				
		AACCTCTACAGCAGACTTGGAAACAAGAGTTCTAAATGAACATTTCAAA				
		AAGGATGAATCAACGATGAAAGAATGTAAAATAAAAAATGGAAACCAT				
		TACCAAGACTGGATAACAAAGGCCCAAATTGATAATAAAGAAATATTGG				
		AAGCCCTAAAAAACAGTACAGATTCTGCCCCCGGAGAAGATAACATTCC				
		TCTGAGGCAATGGATAATCTGGAACAACGACGGTGTCCTCTTTGATATG				
		TTTAACTACATCAAAAGGACACACACGATATCCCAGATATGTGGAAAAACT				
		ACACCACAACACTACTTATAAAACCCGGAAAAAGCCAAGAAAGCAACAT				
		CCCCGCTAATTGGAGGCCAATATCGATATTGCCAACAAGCTATCGTATAT				
		TTATGAAAGTCCTAAATAAAAGAGTACTAGAATGGGCTAATAGAGGAG				
		AACTGATATCAAAATGGCAGAAAGCCGTAGACAAAGCTAATGGATGTG				
		ATGAGCACAGCTATGTCATACAAGCGCTTATCGAAAAAGCAAACAGAA				
		GCTACTACAAAAACGAGCAATGTCACCTCGCCTTCTTGGATTTGGCAGA				
		TGCTTTTGGAAGCATCCCATTCCAAGTAATATGGCATACCCTAAAAAATA				
		TGGGTATGGATGAGGAAACCATCAACTTGCTCAAAGAAATCTACAAAGA				
		TTGCTCCACAAAATATAAATGTGGAAAGAATGAGTCAGAAAAGATCAAA				

		ATTACGAAAGGAGTCCGACAGGGATGCCCATTGTCGATGACCCTCTTCA	ACTCCCC ACCTGA CACTGTC CTCAAA ACAGTT CAATTG CATCCG AAGATC GCAATTT TTTCACT AAAATA AATTAA CAAAAG TTAATTA TACTGCT TCATTGA GTAAGT AGAAAA ACAATC (SEQ ID NO: 1291)	CPLSMTLFSLCIQYLI QGIAEKKKGATIAGQ EVCILAYADDLVIVAN TAKDMQMLLTTIENL AKQADLIFKPAKCGY YRDPRDKKSMMKIY GKEISIVDEKNVYTYL GVRIGDTKKKDLNVR FEEVKKKTTAIFKSKLR SDQKLEAYNIFCQSKF VYILQGEDIAKTKIETY DEEIKKMIKEDILKLQ DKSPFTDFVIYSPREK GGLGITKIIDEQTIQTI NRTAKLLNSSHRAIR AIIYEELIQVANLRGEK EINTIEEALKWLEGTN KYKKNSNAKTTWITR VREAFQTLEKKHKIKV RFVPKENCIGYKIKCD TQEKIVELDNSKELSK SLHWMIKEAYYKEW KALKCQGYIISLKTSEF MEWKMPRGLPDPD WRFLTKVKANMLDV NMKQANQGGRLGS TKCRKCEDKESASHVI NHCASGNWSRVEKH NQVQNELAKELTKR NISFEKDSIPKETKESL RPDLVIRLKDKIMIVD IKCPFDEESAIESARN KNIDKYRELAKEIQAK TGLQTTVSTFVVCSL GTWDKRNNELLRQ MGIRYEESKEMRIN MIQKAIHGSRKTYDH
		TCTAGGTGTAAGAATCGGTGACACAAAGAAAAAAGACCTAAATGTCAGA		
		TTCGAAGAGGTCAAAAAGAAAACGACAGCAATCTTCAAATCGAAATTGC		
		GAAGTGACCAAAAACTAGAGGCATACAACATCTTTTGCCAATCAAAATT		
		TGTGTACATCCTACAAGGCGAAGATATCGCAAAAACCAAAATTGAAACT		
		TACGACGAAGAAATCAAGAAAATGATAAAAGAAGATATATTAAAATTAC		
		AAGACAAAAGTCCGTTCACAGACTTCGTTATCTACTCCCCAAGAGAAAA		
		AGGGGGGTTAGGAATAACAAAGATAATAGATGAACAAACAATTCAAAC		
		TATTAATAGAACGGCAAAACTCCTAAATAGTAGCCATAGAGCAATCCGG		
		GCTATTATTTATGAAGAGCTAATACAAGTAGCTAACCTAAGAGGAGAAA		
		AAGAAATCAACACCATTGAAGAAGCACTAAAATGGTTGGAAGGTACCA		
		ACAAATACAAAAAGAACTCCAACGCCAAGACCACCTGGATAACAAGGG		
		TTCGGGAGGCCTTTCAAACTCTAGAAAAGAAACACAAAATCAAGGTTAG		
		ATTTGTGCCCAAAGAAAACTGCATTGGATATAAAATCAAATGCGACACC		
		CAAGAAAAGATAGTGGAGCTTGATAACTCAAAAGAGTTATCAAAAAGC		
		TTACACTGGATGATAAAAGAGGCATATTATAAAGAATGGAAAGCCCTAA		
		AATGCCAAGGATATATTATAAGCCTAAAAACCTCCGAATTTATGGAGTG		
		GAAAATGCCCAGAGGCCTTCCGGACCCTGATTGGAGATTCCTAACAAAA		
		GTAAAGGCAAATATGTTGGACGTAAACATGAAACAAGCCAACCAGGGA		
		GGAAGGTTGGGAAGCACAAAATGCCGAAAATGTGAAGATAAAGAATC		
		GGCAAGCCATGTTATAAACCACTGTGCCTCAGGTAACTGGAGTAGAGTG		
		GAAAAGCACAACCAGGTGCAAAATGAGCTAGCAAAAGAACTGACAAAG		
		CGGAATATCAGCTTCGAAAAGGACAGCATCCCAAAAGAAACAAAAGAG		
		AGCCTAAGACCAGATTTGGTTATAAGACTCAAAGACAAGATAATGATAG		
		TGGACATCAAATGCCCATTTGATGAGGAATCTGCTATCGAGAGTGCCAG		
		AAACAAGAACATAGACAAATATCGAGAACTGGCCAAAGAGATCCAAGC		
		AAAAACTGGGTTACAAACAACAGTCTCAACTTTCGTTGTCTGTTCTTTGG		
		GAACCTGGGATAAGAGGAACAACGAGCTCCTACGGCAGATGGGAATAA		
		GATATGAAGAATCCAAAGAGATGAGGATCAATATGATCCAAAAAGCCA		
		TCCACGGGTCTAGAAAAACCTACGACCACCACAGAAATTTTAACAATGG		
		TTAAAATGGCAAAAAGATATTTCAAGATGAATTGTGGACTCATCTAAAA		
		AATGACCACCTTGAGTCCAAATATGCCTAGCTATCATGGTTGCTGATGG		

			AAACAGTAAGGCACCTGATAGCTAACTTTTCACTGTGAATATCTTCAGAT ATTCACAGTGACACGAAAGGACACCACTAGTAAAAACCACTAGTTTTTTC TGACACCTCTTGCTACAAACTCTGTAAAAATCAAAAGGATCGATAGGCC GCGCTTTCACGGTCTGTATTCGTACTGAAAATCAAGATCAAGGAAGCTT TTCCCCTTTTAGTCAACACCAGGTTTCTGTCCTAGTTGAGCTTCCCTTGGG ACATCTGCGTTACCATTTGACAGATGTACCGCCCCAGTCAAACTCCCCAC CTGACACTGTCCTCAAAACAGTTCAATTGCATCCGAAGATCGCAATTTTT TCACTAAAATAAATTAACAAAAGTTAATTATACTGCTTCATTGAGTAAGT AGAAAAACAATC (SEQ ID NO: 1045)			HRNFNNG (SEQ ID NO: 1413)
NeSL	NeSL-1_CBre	.	Caenorhabditis brenneri	AAGGACGCTGGTTTAAGGCCGAATTCGTTCGTTCTTTTTCTGGCGGTCTT GCTTTGAGCTTGGTTTCCGATCCTATGCCCTTGWGCATCAGCGATTGCG TTCATCTGGTCTCGGCCGAAGGAGATACGATGAACGGGAGGTCCACTT GTGGGCCATTGTCTCGTTCATCCTCTGTCGTAAGTAGGTCGAGGTCTTCC CCTTCCCCCAGTGTTCCCCCCCACCCTTCCCCCAGTATCGGTCCAGATACA GGATTGTCGGCTGGAATCATCGGCACATCGAGAGGATGTAGCCTTTGGT TGCCAGAGGTGGACAATGCCTTATCACAGTGGCTGAGAAAAGGGTTGG AACGAGACCATGAAGTTCTGGTTTGTGGATTTGAGGCAGCAAAGCCACT GTCACTTTCCAAAGCTAGACTTCTAAGAAAGACCCCAAGGAACACTGGT GTGGTTAGGCACATATTAGAATTTGACGGAAGGTTAGTTCATACTAACT GTAACGAGACCGAGTGTGTTCTTTCTACTTTGTKCAGTGAMGWGGCTG TCGAAGTAGTCAGGATATCTCTCAAATGTGAACCCCGTGAACCCTGTGA ACCCAAATGTGTTCTTTCTATTTTATGCAGTGATAAGATAGTCWGGATAT CATTTGAATGTGAAACWCGTGAACCTTTTCCTTTCTTCMCGGATCGGAA ATTCAGAGAACCTATTCCTTTCGTTTTCGAGAGGATGTATGACCCAAGA GACCCTATTCCTTCATTTATTTGTTGGATGTATGACCTGAGACAAAGGAT GACCCCTGGCACSTTGCCAAGWAATCCCCTTTCTCMAGAGAACAAAGA CAGCTGGGGACGCCCAGCTGTCATAAAGAATGAGATAAGATCTATGAG ATCTTATCTCGAAGAGAATGTGAAGGAAAACCGCCTGAACCTTTTGAGA AGGTTAAGAGGTGGTGGTGAAGGAAAGAAGATGATCAGAAAGTTGGT TGCAGAAAAGAAAAGCGACACAGAGGCTGTCTGCAGGATACTGTACCC ACTTGATGATCGTTATGAGTGTTTTGTTGATGGTTGTGAGACAACATCAA CGATGGGATACGGGTCTAGTGACCTGAAATACATGACCACACACATAAA GAAAGAGCATGGTGTGAAAGTCCAATGGACATATGAGTGCTCCCTGTG TAATAAGCAAGCTCCTTTCATGGGGGGAGCTGCGTCCAAGTGGGTTACA GCGCACATGGCAACAAAGCATACCGAAACGGTGAAGTTGAAGCTCAAA CCAAGCATCTCGACTACTGCCAAGGTTGCTGCGAAGCTAGATGAGATCG CCGTGTCGCTACCCAAACCGAGACAAGTACGTGTATTGAGAGACCCAGA TGAAGTGAAAGAGAAGGTTGCAAAACCAACACTTGCTTCCACGAGAGA	AAGGACG CTGGTTTA AGGCCGA ATTCGTTC GTTCTTTTT CTGGCGGT CTTGCTTT GAGCTTGG TTTCCGAT CCT (SEQ ID NO: 1169)	TAAACC CACACG AGAMCT ACGACG CCATAA GATCAG GCATGT ACGGAT GTGAAT GAGACT GATGAA CGGAAT GAGCAC GTGCCC ATAAGA TCGGGT ATKAAA GAWCA GAGACG ATCCCTA MCATCG GGAAAA CACGAG TTATACT GCTTCAC TGAMCT CGCTAA GCTCTCA TAATGA	MPLXISDCVHLVSAE GDTMNGRSTCGPLS RSSSVVSRSRSSPSPS VPPHPSPSIGPDTGLS AGIIGTSRGCSLWLPE VDNALSQWLRKGLE RDHEVLVCGFEAAKP LSLSKARLLRKTPRNT GVVRHILEFDGRLVH TNCNETECVLSTLXSX XAVEVVRISLKCEPRE PCEPKCVLSILCSDKIV XISFECETREPEPFFXD RKFREPIPFVFERMY DPRDPIPSFICWMYD LRQRMTPGTLPXNPL SXENKDSWGRPAVIK NEIRSMRSYLEENVK ENRLNLLRRLRGGGE GKKMIRKLVAEKKSD TEAVCRILYPLDDRYE CFVDGCETTSTMGY GSSDLKYMITTHIKKE HGVKVQWTYECSLC NKQAPFMGGAASK WVTAHMATKHTETV KLKLKPSISTTAKVAA KLDEIAVSLPKPRQVR VLRDPDEVKEKVAKP

		AGAAGTGAAGAGAAATGCNTTGCGAAACATGGCCCCACTAGTCGAACT	CCGAAC	TLASTREEVKRNALR	
		GAGTTCTCAGAATCAGTTGACWGGAGCCGAAAGACCTGAGGAAACTA	TTGTTCG	NMAPLVELSSQNQL	
		GTGAAGCTATGCGACTCGAGGAGTGTAGGACTCCAGAGAAGATTGCTG	CAACTG	TGAERPEETSEAMRL	
		AACTAGAAGGAAAGATACAGACCCGAACAGTGACTAAAAAGCTTAGTG	CCTCCTA	EECRTPEKIAELEGKI	
		CACTGAAAGAGTCAATGGAAAAGAGAACGAGAGAGGAAGGTTGGG	ACCGGG	QTRTVTKKLSALKES	
		AAACCATCACTTGCTCCAATTCATGAAGAAGTGAAAAAAGACTGCAAGAC	CGGGTG	MEKRTREEKVGKPSL	
		GGAGCTTGGCACCTCTAGTTGAACCGAGTACGTTCACTCATTTGACTGG	TGAGAA	APIHEEVKKTARRSLA	
		GGCGTCAAGACTTCAGGCTGTTCGTGACGCGTTCTCGAAAGCCAACAAA	GGGAGG	PLVEPSTFTHLTGASR	
		GACGCTGCGGCGAAAAGAAGGTCTAGCCTGGCGAAACCAGCTAGATTA	TCGCCTT	LQAVRDAFSKANKD	
		TCAGAGATTATGAATACCACCTTCACGAAGGAGACGGTAAATGAGACG	GAGGCG	AAAKRRSSLAKPARL	
		AAAGAACCTGTGAATGATACTGACGAGAGTATCGCAACAATCCAGCCAC	GACGCA	SEIMNTTFTKETVNE	
		AAGTACGTGTCTACCGGTTTAATACATGGTGTCTCGATCATGAAACCAC	ATGAGG	TKEPVNDTDESIATIQ	
		GAGAGAAGCCTGGTTAACCGGAGAAGTTGTGGATTGGTTCATGGGAAA	GATGTG	PQVRVYRFNTWCLD	
		AGTGACTGAGAAGAAAGACCAGTACAGAGTGTTTGACTCACTTGTATG	TGCAGG	HETTREAWLTGEVV	
		GTCAATGTACAAGTTCCATGGTGTAGGGTATGTATTGGATCTGATGAGG	TTCCCCC	DWFMGKVTEKKDQ	
		GATCCTCTAACATACTTCTTACCAATATGTGAACACGATCACTGGGTTTT	TCTTGA	YRVFDSLVWSMYKF	
		GCTAGTGATTGATGAGAAAGGAATTTGGTACGGTGACTCGAAAGGTGC	GATCCG	HGVGYVLDLMRDPL	
		AGAACCGTGTAGAGAAATCGCCAAATTCATAGAAGAGACGAAAAGAGA	AAAGTC	TYFLPICEHDHWVLL	
		AAGACGAATGTTCCCAGTCCCGTACCTCTTCAAAGAGACGGAGTGAAC	TAAAAG	VIDEKGIWYGDSKGA	
		TGTGGTGTACATATATGTCTAATGGTTAAATCCATCGTGAATGGCGAAC	TACTAG	EPCREIAKFIEETKRER	
		CATGGTACACCGAAGAAGAAGTGAAAGTGTTCAGAAGAAATGTGAAAA	ACCGAA	RMFPVPVPLQRDGV	
		GAGGTCTGAAAGAATTTGGTTTTGAACTTTATTCTGAAAGGATCGTCTAT	AGATCG	NCGVHICLMVKSIVN	
		GTCGGAGATGACAGCATAAAAGTGAATGATGAGCATGATGATGACGTG	AGGACG	GEPWYTEEEVKVFRR	
		GTATTCCTCTCGGAGGAGACGAATAACACTACGTTCACGATCGAGCAAG	GACGGG	NVKRGLKEFGFELYSE	
		CAGAAGATCCGGCTGAAGAGGATGCCCAGCATCTGGAGAGTCCGGTGA	ATGGCC	RIVYVGDDSIKVNDE	
		AACCTGTAAAGCTCATGGAGTTGAAAATTCCAAAGATTGAGATAAAGAA	GCGAGG	HDDDVVFLSEETNNT	
		GAAAGAGATTCGGAGAAAACCGAAACAACAAATCGAAAAGAAAAGAA	CACACG	TFTIEQAEDPAEEDA	
		AGGTGCCAACAGGGAAACCAGATGAACTGTTGGTCAGAGTGCGATTAT	GCGGGT	QHLESPVKPVKLMEL	
		GGTTGGAAAGAGAAGTCCAATCATACTACTTCGACTCTGGAAAGAGATTCCA	AACACA	KIPKIEIKKKEIRRKPK	
		AAGACTGGAGTGGATATTAGATGTCCTCACGGCTGCGATTCACAAGGCT	GCCAGA	QQIEKKRKVPTGKPD	
		ACCGCCGGTGATGAGCAAGCAATTGAAAGAATTGAGAAGAGATCACCC	TAACCTA	ELLVRVRLWLEREVQ	
		CCTTTGGAAGTGGAAGAGGGTGAAATGTCTACACAGACAGAACCAAAG	GTAGAT	SYFDSGKRFQRLEWI	
		AAAAGAGAAAGAAAAGAAAAGGAGTCAGGTTGTGAAATGAAAGCTTCT	CTTCGG	LDVLTAAIHKATAGD	
		CACAAGGAGATGTACTTCAAAAACCGCTCCAAAGCGTTCAATGTGATAA	ATCTCGT	EQAIERIEKRSPPLEVE	
		TTGGAAAAGACTCAAAGCAATGCGAGATTCCAATTGAGACCCTGCAAAA	CGGCCT	EGEMSTQTEPKKRER	
		GTTCTTTGAGGGAACAACTGCAGAAACGAATGTGCCAGCAGAAGTGCT	GGAGAT	KEKESGCEMKASHKE	
		GAAAGAGATGGGTTCACGTCTGCCAAAGTTGGAGGCGTTGGACTGGAT	ATGTGG	MYFKNRSKAFNVIIG	
		GGAAGCTAATTTCATTGAAAGTGAAGTGTCAGATGCGATGAAAAAGAC	AACCCT	KDSKQCEIPIETLQKF	

				GGGAAA	FEGTTAETNVPAEVL
				GGAGAA	KEMGSRLPKLEALD
				AGTTGTT	WMEANFIESEVSDA
				TGTTGG	MKKTKDTAPGVDGL
				GCTGGC	RYHHLKWFDPEYKM
				AAGAGT	LTLLYNECKNHRKIPS
				GAAGTT	HWKEAETILLYKGGD
				TGAATG	ETRPDNWRPISLMPT
				TGAACC	IYKLYSSLWNRRIRSV
				ACCGTC	GGVMSKCQRGFQER
				ATGCAA	EGCNESIGILRTAIDV
				CCACTA	AKGKRRNLSVAWLD
				AACCAG	LTNAFGSVPHELIKST
				TGGCGA	LESYGFPEMVTEIVM
				TGCGGG	DMYRGASIRIKSKNE
				TGGAGT	KSEQIVIKSGVKQGD
				CATCAC	PISPTLFNMCLENVIR
				AGGAAA	RHLDSASGHRCIKTK
				ATGTTTC	VKVLAFADDMAILAE
				TGTTGCT	NRDQLQTELNKLDKE
				TGACTTA	CESLNLIFKPVKCASLI
				TCAGTG	IERGMVNKNAEVVL
				TTTGATA	RGKPIRNLDENGSYK
				TCGCCCT	YLGVHTGIATRVSTM
				CAGGCA	QLLESVTKEMDLVN
				CAAGTA	QSGMAPFQKLDCLK
				TGAAGG	TFVLPKLTYMYANAIP
				CCCCCAC	KLTELKVFANLTMRM
				CCACAT	VKEIHEIPKGSPLEYV
				AAACTC	QLPPSQGGLGVACPK
				CCTAGC	ITALITFLVNVMKKL
				AACTGG	WSSDSYIRKLYRDYLD
				TAGTCC	EVAETETGMEEMTK
				AGCAAG	EDIAKYLSGDVPIDKK
				CGCTGG	AFGYNTFTRVRDVCN
				TWCTTG	SLTXIXGAPLHKLKIVE
				CTACTAT	RDGDFAILVQATKEG
				TGCGCC	MEKIFTCAQEKKLQQ

			GWCATGCGAGTACGGGTTACCAAAGCTAATCAACGAAAGACATGATGC GGTGTTACATGTGGTGAGAAACCTCATCCGCAAAGGCTCAAAGAAAGA CTGGAAGCTAAAGATAGATGAAACTGTGTCAAGTTGTAATCAACTTCGT CCAGACATCTATATGTGTAGCCCAGATGGGAAAGAGGTCATAATGGCA GATGTAACCTGTCCTTATGAATCAGGAATGCAAGCTATGCAAGAGAGTT GGAACCGAAAGGTCACSAAATACGAAGGAGGCTTTAGCCACTTCCAWA AGATGGGAAAGAAATTCACAGTGTTGCCAATAGTGGTTGGATCACTGG GAACGTGGTGGAAACCCACAACGAACAGCTTAGTTCAACTAGGCATAG AGAAAGASACGATAAGAAGAGTGATCCCCGAGCTGTGCTCAATGACCA TGGAATACAGTAAGGATGTCTACTGGAACCATATATTCGGGGACACCTT CAGGAAGCCACCAATGAGATTTGGTGTAGAGAAGCCAAAGGGTAATAG TTGGAAGAAGGAAGGCAGCGAGCCGAAAGGTGCTGCTTCCTCCGACTA AACCCACACGAGAMCTACGACGCCATAAGATCAGGCATGTACGGATGT GAATGAGACTGATGAACGGAATGAGCACGTGCCCATAAGATCGGGTAT KAAAGAWCAGAGACGATCCCTAMCATCGGGAAAACACGAGTTATACT GCTTCACTGAMCTCGCTAAGCTCTCATAATGACCGAACTTGTTCGCAAC TGCCTCCTAACCGGGCGGGTGTGAGAAGGGAGGTCGCCTTGAGGCGGA CGCAATGAGGGATGTGTGCAGGTTCCCCCTCTTGAGATCCGAAAGTCTA AAAGTACTAGACCGAAAGATCGAGGACGGACGGGATGGCCGCGAGGC ACACGGCGGGTAACACAGCCAGATAACCTAGTAGATCTTCGGATCTCGT CGGCCTGGAGATATGTGGAACCCTGGGAAAGGAGAAAGTTGTTTGTTG GGCTGGCAAGAGTGAAGTTTGAATGTGAACCACCGTCATGCAACCACTA AACCAGTGGCGATGCGGGTGGAGTCATCACAGGAAAATGTTTCTGTTG CTTGACTTATCAGTGTTTGATATCGCCCTCAGGCACAAGTATGAAGGCCC CCACCCACATAAACTCCCTAGCAACTGGTAGTCCAGCAAGCGCTGGTWC TTGCTACTATTGCGCCCCAGGCTCGCCC (SEQ ID NO: 1046)		CCAGGC TCGCCC (SEQ ID NO: 1292)	LLKAEVNTALAHRFFT EKPVKSAVMSVMRQ YPQSNAFVKNGKNV SIAVHSWIHKARLNA LHCNFNTYGENKSKV CRRCGKDVETQLHIL QXCEYGLPKLINERH DAVLHVVRNLIRKGS KKDWKLKIDETVSSC NQLRPDIYMCSPDGK EVIMADVTCPYESG MQAMQESWNRKVT KYEGGFSHFXKMGK KFTVLPIVVGSLGTW WKPTTNSLVQLGIEK XTIRRVIPELCSMTME YSKDVYWNHIFGDTF RKPPMRFGVEKPKG NSWKKEGSEPKGAA SSD (SEQ ID NO: 1414)
NeS L	NeSL- 1_Clap	. Caenor habditis japonic a	GCGCCCCGGTTACATTGTCGGGGCCACCTTTCTCTTGGAGTAGAGTAC AGTCTACTAATTTTTTGATAAGCTAGTCGGGTCCGAACCACTAGAGTTTG CTTGAAAATGCGTCAAACCAGCATTTTAGAACTCGCCCAAAAGTTCGGC CCCGACCCCCAAACAAATGGGACCTTCTTGACGATTTTCCCTGAAAATCG GAGGATGGAATGGTCCCCTATTCTTGTAAATAGKACTGTGCAATACCCC TTCGTCATCTGTGGGGAACAGATGACACGTGACGTCATCCGTGTAGACG TCACGTTTTCCCGTGCCTGCGGGAGCCCCCAATCGAGCAATTTTTGCTCT TTTGAGTGTCTGGAACGCTTGAAACCCCAGACAAATCAGGCCCAGTCGT CGGAAAATTTCTTTTTGAAATTTTTTGGCGCCTGCGAAAAAAATTTTTTA ACCGCCACAAACCCCCGGGAGGCGCGGWTAGGGATATCGATGTCATCG ACTCGTCGGTGATCTTTGATTTTCTCTCTGCGTCTCCTATTTTGGAACAGT CTCGACCAAAAAACCGGGCCTGGCAACCCACCGAATCCGGATGTCGGA	GCGCCCCG GGTTACAT TGTCGGG GCCACCTT TCTCTTGG AGTAGAGT ACAGTCTA CTAATTTTT TGATAAGC TAGTCGGG TCCGAACC ACTAGAGT	TAAAAG CCAAAA GCCACG GAGCAT CGGGAA AGAAAA ATGGAA AAGGAC TGAAAA CGAGAC TGAAAA ATCCCA	WRRPAPKQTKNSSL HHLGHEVKRIARLKP GIFEFHAKPKNSSLHH LGHGVKRXARLKPGI FEFHAKXKNSSLHHL GHEVKRIARLKPGIFE FHAKPKNSSLHHLGH EVRRNSRLKPGIFGFY QKSKNSSLHHLGHEV RRIARLKPGILEFHAK NRIKSGLKVTFLSDLX AHAGALACSRFLAST

		GGGATTTGGCAAGAAATGTTGGAAATAACGAAATTTCGTTATTTTCAGC	TTGCTTGA	AACAAA		LKTEHCRQKSFKPVG
		ACAATTGTCAAACCGGCAAGAAAACTGGATGGACAAGACACACAATTTA	AAATGCGT	ACAAAT		FLLHFLKNSSINEVAS
		CCGGAAATTGTGCTTGTTACGTCGAATTTCCCAATTTTGAAAAAATTCCT	CAAACCAG	CCAAAA		LRNVKKXFLEFFSGKP
		CGTTCCACTGGTCGGGACGCGAGGTCAGACGATCTGCACGTCTGAAACC	CATTTTAG	CAAACT		IGGMASFSRTKITFFK
		CAAAATCTTCGGATTTTATGCAGTAGTGGCGCCGCCCGGCTCCCAAACA	AACTCGCC	GAAAAA		LCLKNFVLSAENPPIIR
		AACAAAAAATTCCTCGTTGCACCATTTGGGGCACGAGGTCAAACGAATT	CAAAAGTT	AAAAAA		QKTNQNKASXVQJA
		GCACGTTTGAAACCTGGAATTTTTGAATTCCATGCAAAACCAAAAAATTC	CGGCCCCG	AAAACA		RGGHLSDCLPSQKM
		CTCGTTGCACCATTTGGGGCACGGGGTCAAACGAWTTGCACGTTTGAA	ACCCCCAA	AAACAA		AGVLGRLFLSVQSTLS
		ACCCGGAATCTTCGAATTCCATGCAAAACMAAAAAATTCCTCGTTGCAC	ACAAATGG	AAACTG		HRPFDTLLRSDDDKR
		CATTTGGGGCACGAGGTCAAACGAATTGCACGTTTGAAACCCGGAATTT	GACCTTCT	GACAGA		GRKTIKLQFFIKENLV
		TCGAATTCCATGCAAAGCCAAAAAATTCCTCGTTGCACCATCTGGGGCA	TGACGATT	CACTGG		TPXVARDVKILXKQT
		CGAGGTCAGACGAAATTCACGTCTGAAACCCGGAATCTTCGGATTCTAT	TTCCCTGA	AAACAG		KNNSGNSDSNSETK
		CAAAAATCAAAAAATTCCTCGTTGCACCATTTGGGGCACGAGGTCAGAC	AAATCGGA	TGTCAG		NFSKNKVSRQNGPLI
		GAATTGCACGTTTGAAACCCGGAATTCTCGAATTCCATGCAAAAAACCG	GGATGGA	GCAAAG		GGGNHKKIGENQITR
		GATAAAATCCGGTCTAAAAGTGACGTTTTTGTCAGATCTTWCTGCTCAC	ATGGTCCC	TCGCCG		TLEIESKSDDNKVLVL
		GCTGGTGCGTTGGCGTGTAGCAGATTTCTGGCATCGACACTTAAAACGG	CTATTCTT	ATTATAC		RILYPTNDWYKCYSQ
		AGCACTGCCGACAAAAAAAGCTTCAAGCCAGTCGGTTTTCTGCTTCATTTC	GTAAATAG	TGTTCCA		WCQHKSLVGYGAHD
		CTAAAAAATTCCTCGATCAACGAGGTGGCGTCTCTCCGCAACGTCAAAA	KACTGTGC	CGCCTTA		LKYLTDHIKSTHSKKV
		AAAWKTTTCTTGAATTTTTCTCAGGAAAACCTATCGGTGGGATGGCTTC	AATACCCC	AAAGTC		EWSYQCSICDAKAEG
		TTTCAGTAGAACCAAAATAACTTTTTTCAAACTTTGTTTGAAGAATTTTGT	TTCGTCAT	CCGAAA		TGTKAARWITAHMP
		TTTGTCTGCTGAAAATCCCCCCGATAATCCGCCAAAAAACAAACCAAAAC	CTGTGGG	TGGCGC		KVHGIEATHRIKQNS
		AAAGCGAGCCKTGTCCAAATCGCAAGAGGAGGTCATCTGTCAGACTGTC	GAACAGAT	AAAACA		EKTTNVKTANSLQE
		TKCCGTCCCAAAAGATGGCAGGAGTTCTCGGACGACTATTTCTCTCGGTT	GACACGTG	ACCTGA		MALSLQKPKNGPKK
		CAGAGCACTCTCTCGCACCGCCCTTTCGACACGTTATTGCGGAGCGATG	ACGTCATC	ATCTATC		VVMATSTTPEKKISEL
		ACGACAAAAGAGGGAGGAAAACGATCAAACTCCAGTTTTTTATTAAAGA	CGTGTAGA	TGAAAG		ESKIQTREVAKQLSAL
		AAATCTGGTCACACCTKTGGTTGCTAGGGACGTGAAAATTTTAAAMAAA	CGTCACGT	TGCTCCA		KESAQKNQQGNKTK
		CAAACAAAAAACAATTCTGGGAACTCTGATAGCAACAGTGAAACAAAA	TTTCCCGT	AACCAC		NVKSSLKTIAENTNET
		AACTTCTCTAAAAATAAAGTTTCCAGACAAAATGGCCCATTGATTGGGG	GCCTGCGG	GCACAA		KKISARKSLINYLKPED
		GCGGTAACCACAAAAAAATCGGAGAAAACCAAATCACACGCACTTTGG	GAGCCCCC	CTCGGA		VLNHIPKEPKPASAKX
		AAATTGAATCCAAAAGCGATGACAACAAAGTTTTGGTCCTCCGAATACT	AATCGAGC	GAAAAT		GLQELTGAQRLQETR
		GTACCCAACTAATGATTGGTACAAGTGTTACTCCCAATGGTGCCAACAC	AATTTTTG	CAGGGA		RRFMAGNRRDSIAR
		AAATCCCTTGTTGGATATGGCGCTCACGATTTAAAATACTTGACAGACCA	CTCTTTTG	CAAGTT		RESLSLGKISNSFKIEL
		CATAAAGTCCACTCATTCTAAAAAGGTTGAGTGGTCTTATCAGTGTAGTA	AGTGTCTG	GCTTCAC		KNAPEKTTLKKPAVT
		TTTGTGACGCAAAAGCCGAAGGTACCGGTACAAAAGCWGCTAGATGG	GAACGCTT	GCAACG		QKQNTSQNVSSSTV
		ATTACAGCCCACATGCCAAAAGTACACGGTATTGAAGCAACACACAGAA	GAAACCCC	GGCTGG		VKENKTGNDVITIDD
		TTAAACAAAATTCTGAAAAAACACAAATGTTAAAACTGCGAACAGTCT	AGACAAAT	GACAGG		TETVKRKINTWCLDH
		CCAGGAAATGGCGCTGTCGCTCCAAAAACCAAAAAATGGTCCGAAAAA	CAGGCCCA	TACCCCC		ESTENAWMADDIIF
		AGTTGTAATGGCAACTAGTACGACCCCAGAAAAGAAAATCTCTGAACTG	GTCGTCGG	TCCTGA		WYIQKQIEISLDNKKF

		GAATCAAAAATCCAAACCAGAGAAGTGGCCAAACAATTGAGCGCTCTG	AAAATTTC	AACCGC	KVIDPLIWTTYRIYGV	
		AAGGAGTCAGCTCAAAAAAATCAGCAAGGAAACAAAACAAAAAATGTT	TTTTTGAA	GAGGTT	ECVQDELVGFEKYFF	
		AAATCAAGCTTAAAAACAATTGCTGAAAACACAAATGAAACMAAAAAG	ATTTTTTG	GAGGAT	PICENGHWVLLIIDDK	
		ATWAGCGCTCGAAAGAGCCTTATAAACTATCTGAAACCTGAGGATGTG	GCGCCTGC	GGACGG	RVWYSDSLADKPIEVI	
		CTCAATCACATTCCAAAGGAGCCAAAACCAGCTTCTGCGAAAAKTGGCC	GAAAAAA	GAAGGC	EDLINKLNRTQGKFN	
		TKCAAGAACTGACTGGTGCTCAAAGACTGCAGGAAACMAGAAGAAGG	ATTTTTAA	CGCGAG	QTVPKQKDGFNCGV	
		TTTATGGCWGGAAACAGAAGAGATTCAATTGCAAGAAGAGAAAGTCTG	CCGCCACA	GCTTAT	HVCLVAKSVITENFW	
		TCTCTCGGCAAAATCTCAAACTCATTTAAAATTGAGCTGAAAAATGCTCC	AACCCCCG	GGCGGG	YTEKDVNDFRKTVKL	
		GGAAAAAACAACTCTTAAAAAACCGGCTGTCACTCAGAAACAAAACACG	GGAGGCG	TAACTC	WLFSEGFELYSEPYK	
		AGTCAGAATGTATCTAGTTCTACGGTTGTAAAAGAGAAGAACAAAACAGGAA	CGGWTAG	GGTTGG	QIQNKNISVNSEKNQ	
		ATGACGTGATCACAATTGATGACACAGAAACTGTTAAAAGAAAAAATAAA	GGATATCG	TGTGCT	ISDNEKNWGDKTQT	
		CACTTGGTGTCTCGACCACGAATCCACAGAAAATGCGTGGATGGCTGAC	ATGTCATC	AGTAGA	VNESTLKERDEDIFLL	
		GACATCATATTCTGGTACATCCAGAAACAGATTGAAATCAGTTTGGACA	GACTCGTC	TGATTTA	RPHISVGVALKTEDEK	
		ATAAAAAGTTCAAAGTGATTGATCCACTCATCTGGACCACATATCGAATT	GGTGATCT	TATCCG	NQKAENLKAPQKLK	
		TATGGTGTCGAATGTGTCCAAGATGAACTAGTTGGATTTGAAAAATACT	TTGATTTT	ACAGCC	AIRRLKILKTCLKKLTA	
		TTTTTCCAATCTGTGAAAATGGTCATTGGGTTTTGCTGATTATCGATGAC	CTCTCTGC	CCAACT	VKGKPEETERAAIPNL	
		AAAAGAGTCTGGTACAGTGATTCCCTGGCCGATAAACCAATTGAGGTTA	GTCTCCTA	AAGAGG	MAIKLKTPPKVEPVR	
		TTGAGGACCTCATAAACAAACTAAATCGAACCCAAGGTAAATTTAACCA	TTTTGGAA	AATCCT	RNPEKGENYXKSQP	
		AACGGTTCCAAAACAAAAAGACGGCTTTAATTGTGGAGTTCATGTATGT	CAGTCTCG	GGGAAA	NKKRQIPTGKPDELV	
		CTGGTGGCCAAATCCGTTATCACTGAGAACTTTTGGTACACAGAAAAAG	ACCAAAAA	GGAAAA	KKVREWFEIQFQAYF	
		ACGTTAATGACTTCAGAAAAACTGTCAAGCTTTGGCTTTTCAGTGAAGG	ACCGGGCC	CTTGAA	EDGKSFQRLEWXTGL	
		GTTTGAACTCTATTCAGAGCCGTACAAACAAATCCAAAACAAAAACATTT	TGGCAACC	AAAGTT	LTAAIHKASAGDEQA	
		CCGTTAATTCGGAAAAAAAATCAAATCAGTGATAATGAAAAAAATTGGGG	CACCGAAT	TTTACAG	VGKIIKRCPPLEIEEGE	
		TGATAAAACTCAAACTGTGAATGAGAGTACTCTGAAAGAAAGAGATGA	CCGGATGT	GGCTGG	MATQTETKQKPKNQ	
		AGACATCTTTTTGCTCAGACCACACATCAGTGTTGGAGTTGCTCTCAAGA	CGGAGGG	TAATAG	KSTKGANSSSSIREAY	
		CAGAAGACGAGAAAAATCAAAAAAGCTGAAAACTTGAAAGCCCCACAAA	ATTTGGCA	TTCAGC	AENRARTFNKIIGKD	
		AACTCTGAAACACGGAAGAATTCCAAGTGGACAAAAACGAGAAAACCAG	AGAAATGT	ACAATT	DKCEIPIEKIEKFFENT	
		AATCTCCAAATGCCCAGGAAACTCCAAAAAACGAGCCAAAAATGGTTCC	TGGAAATA	GTAGTC	TSNTNVPTETLARITS	
		AAGTCCGAAGAATTCGGAAAAAGAAATTTCTACTGAGCTTCTGGATGCT	ACGAAATT	TACTGTC	DLPKLEIGSWIEEEFR	
		CAAGAAGCCGGAGAAGAGCTGAAAAGCGATCCGAAGGCTGAAAATCCT	TCGTTATT	TTGCAA	EKEVAEALKKTKDTA	
		GAAAACCTGTCTCAAAAAGCTGACGGCAGTGAAGGGAAAACCGGAAGA	TTCAGCAC	CCACAA	PGVDGLRYHHLSWF	
		GACGGAAAGAGCCGCCATTCCAAACCTCATGGCAATCAAGCTCAAGAC	AATTGTCA	CAAACC	DPKXKLLTKLYNECRE	
		GCCTCCAAAAGTTGAACCTGTAAGAAGAAACCCTGAAAAGGGTGAAAA	AACCGGCA	AGTGGT	HKKIPGHWKEAETVL	
		TTACMAAAAAAGTCAGCCAAACAAAAAGAGACAAATACCAACCGGAAA	AGAAAACT	TCTGCG	LYKGGDETQAENWR	
		ACCGGATGAATTGGTTAAAAAAAGTCCGAGAATGGTTTGAAATTCAATTT	GGATGGA	GGTAGA	PISLMPTICKLYSSLW	
		CAAGCATATTTTGAGGACGGAAAATCCTTCCAGAGGTTAGAGTGGWTG	CAAGACAC	TCAAACT	NKRIKSVTGVLSKCQ	
		ACAGGTTTGCTCACGGCTGCAATTCACAAAGCTTCGGCTGGAGATGAGC	ACAATTTA	ATAATTT	RGFQEREGCNESIAIL	
		AAGCTGTGGGMAAAATCATCAAACGTTGTCCACCTCTGGAAATTGAAG	CCGGAAAT	GTGTGT	RTAIEAAKGTKKSLSI	

	AAGGGGAAATGGCTACCCAAACTGAAACAAAACAAAACCAAAAAACC	TGTGCTTG	TTTCTTT	AWLDLTNAFGSVPH
	AAAAGAGCACAAAAGGAGCAAATAGTTCCAGCTCAATTCGGGAAGCCT	TTACGTCG	TACTTGA	ESIEATLIAYGFPGMV
	ATGCTGAAAACCGAGCGAGAACCTTCAACAAAATTATTGGAAAAGACG	AATTTCCC	CCCGGG	TEVIKDMYNGASIRV
	ACAAATAGTGTGAAATTCCAATTGAAAAAATTGAAAAGTTCTTCGAGAA	AATTTTGA	CAACAC	KTKNEKSKQILIKSGV
	CACAACTTCAAATACCAATGTTCCAACAGAAACACTAGCGAGGATCACT	AAAAATTC	ATTATAC	KQGDPISPTLFNICLE
	TCTGATCTTCCAAAACTCGAGATTGGTAGTTGGATTGAAGAAGAGTTCA	CTCGTTCC	CACGTC	SVIXRHLKSADGHKCI
	GGGAGAAAGAAGTAGCCGAAGCTCTTAAAAAAAACAAAGGATACTGCCC	ACTGGTCG	CACAAG	XSNIKLLAFADDMAIL
	CAGGTGTAGATGGATTACGGTACCATCATCTGAGCTGGTTTGATCCAAA	GGACGCG	GACGAA	SDSKTKLQQELQKM
	AAKGAAACTGCTCACAAAACTGTACAACGAATGCAGGGAGCACAAGAA	AGGTCAGA	TTCATAA	DDDCTPLNLIFKPAKC
	AATCCCAGGTCACTGGAAAGAGGCAGAAACTGTACTCCTCTACAAAGG	CGATCTGC	TGGCCC	ASLIIEWGKVQKDQK
	GGGGGACGAGACGCAGGCCGAGAATTGGCGACCAATCAGTCTCATGCC	ACGTCTGA	CTCCCTA	IKLKGQFIRSLAEQDT
	AACCATCTGCAAGCTATACTCTAGCCTGTGGAACAAAAGAATAAAATCC	AACCCAAA	AATAAA	YKYLGVQTGIETRVSA
	GTGACAGGTGTTCTGAGCAAATGCCAAAGGGGTTTTCAAGAAAGAGAG	ATCTTCGG	CTCCCTA	MQLMKKTVSELDKI
	GGTTGTAATGAAAGCATTGCAATTCTCAGAACCGCTATTGAAGCGGCAA	ATTTTATG	GCAACT	NCSALAXWQKLDAV
	AAGGAACAAAAAAGAGCCTGTCAATTGCTTGGTTGGACCTTACCAATGC	CAGTAG	GGTGGT	KTFVLPKMTYMYAN
	ATTTGGCTCAGTTCCACACGAATCGATCGAGGCCACACTAATTGCTTACG	(SEQ ID	CCGGCG	TVPKLSELKEFANITM
	GTTTTCCGGGGAATGGTAACCGAGGTAATAAAAGACATGTATAATGGCG	NO: 1170)	AAGCCG	RAIKVMQNIPVKGSP
	CATCGATTCGTGTAAAAACAAAAAAACGAAAAGAGTAAACAAATCCTGAT		GTTCTTG	LEYVQLPIGKGGLGV
	TAAATCGGGTGTAAAACAGGGTGATCCAATCTCACCTACTCTTTTCAACA		CCACTAT	ACPKTTALITYLVSTM
	TTTGCCTTGAAAGTGTCATTMGTCGCCACCTAAAAAGCGCGGATGGTCA		TGCGCC	KKLWSTDDYIRKLHT
	CAAATGCATCGAMTCAAACATCAAATTATTGGCGTTTGCCGATGACATG		CCAGGC	DYLKMVAIKETKTKE
	GCAATCCTGTCAGATTCCAAACAAAACTCCAACAAGAGTTACAAAAAA		TCGCCC	VTLEDLASYLSDDKTV
	TGGATGATGACTGTACACCGCTCAACCTTATCTTCAAACCCGCCAAATGT		(SEQ ID	CKKAVGYNSFTRVREI
	GCAAGTCTGATAATTGAGTGGGGAAAAGTACAAAAAAGATCAAAAAATA		NO:	CKTLSKNKGALLSQLK
	AAACTAAAAGGTCAATTCATCAGAAGTTTGGCCGAACAAGACACCTACA		1293)	IIAKDGKLAILVQAXK
	AATATCTTGGGGTGCAAACTGGCATCGAAACGCGCGTTTCTGCAATGCA			DGKTKIFTHDHVKTL
	ACTGATGAAAAAAAACKGTCAGCGAGCTTGACAAAATAAATTGCTCTGCA			QKXLKKEINEALLHRF
	CTGGCTCMWTGGCAAAAACTGGACGCAGTAAAAACTTTTGTGCTCCCA			TTEKRVKSEVVRVVQ
	AAAATGACGTACATGTATGCAAATACTGTACCGAAACTCTCCGAGCTCA			EYPQCNSFVRDGGK
	AAGAGTTCGCAAATATTACAATGAGAGCAATAAAGTAATGCAAAACAT			VSIGAHRFVHKARLN
	TCCAGTAAAAGGTTCACCATTGGAGTATGTACAGTTACCCATTGGAAAA			LLACNYNTWQDAAT
	GGTGGACTAGGAGTGGCATGTCCAAAAACAACTGCGTTGATAACCTATC			KQCRRCGYEKETQW
	TGGTTTCAACAATGAAAAAATTGTGGTCCACTGATGACTATATCAGGAA			HILSSCPKSMGGKITE
	ACTACACACAGACTACCTGAAAATGGTGGCCATAAAAGAAACGAAAAC			RHDSVLKTVKEMIQT
	AAAAGAGGTCACACTAGAGGACCTTGCCTCCTACCTAAGTGATGATAAA			GSLKNWKLKLDHELP
	ACCGTCTGCAAAAAAGCGGTTGGTTATAATTCATTCACAAGGGTACGAG			GSTRLRPDIYLRSPNG
	AAATCTGCAAAACGCTATCAAAAAACAAGGAGCACTGTTAAGCCAACT			SEIILGDVTIPYEHGIE
	AAAAATCATTGCAAAAGATGGAAAGTTGGCTATTCTGGTACAGGCTSTG			AMQTAWQKKIEKYE

			AAAGATGGCAAAACAAAGATTTTCACGCATGACCACGTGAAAACCTTGC AAAAASTTCTTAAAAAAGAAATAATGAAGCCCTTCTGCACAGATTCACA ACTGAAAAAAGAGTGAAAAGCGAAGTGGTGCGAGTGGTCCAAGAGTA CCCCCAGTGCAACTCCTTTGTCAGAGATGGAGGAAAAGTTAGCATTGGA GCGCATCGCTTTGTGCACAAAGCCAGGTTGAACCTGCTCGCGTGTAATT ACAACACGTGGCAGGATGCAGCCACAAAACAATGCAGAAGGTGTGGAT ATGAAAAAGAAACCCAATGGCACATCCTCTCATCTTGCCCAAAAAGTAT GGGAGGAAAAATAACTGAAAGACACGATTCTGTGTTAAAAAACAGTAAA AGAGATGATTCAAACTGGATCTCTCAAAAACTGGAAACTAAAACTTGAT CATGAATTGCCAGGATCAACCAGACTTCGCCCGGATATCTATTTGAGAA GCCCAAATGGATCCGAAATAATTCTTGGCGATGTCACAATCCCGTATGA ACACGGAATTGAAGCTATGCAAACAGCATGGCAGAAAAAAAATTGAAAA ATATGAAGAGGGCTTCAAATACCTTCGTTCTACCGGCAAAAAACTCACA ATTGTGCCAATTGTGGTCGGAGCACTAGGAAGTTGGTGGAAGCCCACA ACAGACAGTCTTGTCAGTCTGGGAATCGACAAAAATACTGTAAAAAGAG CTATTCCAGAAATTTGCTCTACAGTACTCGAATACAGTAAAAACATTTAC TGGAACCATATATTCGGGGATTCCTACCAAAAAGTACCCATGTTTTTCGG CGGTGAAAAACCAAAGGGGCAAAGTTGGAAGAAAGTGAAGCCTCCTGA AGGCAAAACTGCTTCTAACCATGAGCCTCCAGGTTAAAAGCCAAAAGCC ACGGAGCATCGGGAAAGAAAAATGGAAAAGGACTGAAAACGAGACTG AAAAATCCCAAACAAAACAAATCCAAAACAAACTGAAAAAAAAAAAAAA AACAAAACAAAAACTGGACAGACACTGGAAACAGTGTCAGGCAAAGTC GCCGATTATACTGTTCCACGCCTTAAAAGTCCCGAAATGGCGCAAAACA ACCTGAATCTATCTGAAAGTGCTCCAAACCACGCACAACTCGGAGAAAA TCAGGGACAAGTTGCTTCACGCAACGGGCTGGGACAGGTACCCCCTCCT GAAACCGCGAGGTTGAGGATGGACGGGAAGGCCGCGAGGCTTATGGC GGGTAACTCGGTTGGTGTGCTAGTAGATGATTTATATCCGACAGCCCCA ACTAAGAGGAATCCTGGGAAAGGAAAACTTGAAAAAGTTTTTACAGGG CTGGTAATAGTTCAGCACAATTGTAGTCTACTGTCTTGCAACCACAACAA ACCAGTGGTTCTGCGGGTAGATCAAACTATAATTTGTGTGTTTTCTTTTA CTTGACCCGGGCAACACATTATACCACGTCCACAAGGACGAATTCATAA TGGCCCCTCCCTAAATAAACTCCCTAGCAACTGGTGGTCCGGCGAAGCC GGTTCTTGCCACTATTGCGCCCCAGGCTCGCCC (SEQ ID NO: 1047)	CGCGAACC AGTCAT (SEQ ID NO: 1171)	TAGCCG ATCGTA AAAGAA ACCGAG CCGTAA	EGFKYLRSTGKKLTIV PIVVGALGSWWKPT TDSLVSLGIDKNTVKR AIPEICSTVLEYSKNIY WNHIFGDSYQKVPM FFGGEKPKGQSWKK VKPPEGKTASNHEPP G (SEQ ID NO: 1415)
NeS L	NeSL- 1_CRe m	. 	Caenor habditis remane i	CGCGAACCAGTCATATGACAGTCTTTATTGATCGCGGTATAGGCGAGCG AGGCCAGATGGCCGTATGTAGCCTCCACCGTTATTTTTCGTTTTCACCTTT TTCCCCCATCCCCCCGTATGTAAATAATGGATCGTTCGGCGAAAATGGCT GTGGCACAGACAAATCACTGTTGCCCGTCATAGAAGTTGTTGTTCGTGA AGTTAAGATAAATTGGTCTGAGAATATTTTGGTAGTAGAGTGTCTGATA	CGCGAACC AGTCAT (SEQ ID NO: 1171)	TAGCCG ATCGTA AAAGAA ACCGAG CCGTAA	MTVFIDRGIGERGQ MAVCSLHRYFSFPF SPIPPYVNNGSFGEN GCGTDKSLLPVIEVVV REVKINWSENILVVE

		ATGGTAAAGAGCGGAGAAAGAGTCGTTGTAAAGAGACAAAATCTGGAA		CAACAA		CLIMVKSGERVVVKR
		AAAGTTATTCAGAATTTGGCAAGAATCAACTCAACTCTATTTTCCAATCT		GCAAAG		QNLEKVIQNLARINST
		AGGAAATCAGATATTTTGCGTAGTACCCAGAATAAAAGACAGTACCAAT		TAAACA		LFSNLGNQIFCVVPRI
		AAAGAGCAGGGATACAGGAAAGAGAAGCAAWTGAAATTCCATGTATC		AAAGAA		KDSTNKEQGYRKEKQ
		ATTCCGAAGTATAAAATCCCAAGTTCCACCATATTTGAGAGGTGGGGGA		AAATCA		XKFHVSFRSIKSQVPP
		GATGTAATGGAAGATACAGAGATAAGAGGTATCAGAAAGTTGGAGCCA		ATAAAA		YLRGGGDVMEDTEIR
		GAGGCTCAGTTAGACAGCTCAAAACCGCTGATCTGCAGAGTTCTCTACC		AGGAAG		GIRKLEPEAQLDSSKP
		CAACGCAAGGTTATATGTATAAATGTTTTTATCCAAAGTGTAAAGGACA		GTTGAC		LICRVLYPTQGYMYK
		TAGTAATGGATCAACAGATCTGAGAAGTCTGAAGAAACACATGGTGGA		CTCAGA		CFYPKCKGHSNGSTD
		TAAGCATTTCACGAATATTGAATTTGCATATAAAATGTGCTACGTGTATGT		CCCCGA		LRSLKKHMVDKHFTN
		TTTTAACGACTGGGAAATCGGCCACAGCGTTAAAATCAATAAAGGCACA		GGAGGG		IEFAYKCATCMFLTTG
		TATGGCAAGTCACCACAAGGTAACGATGGAACCCGGTAAAAAGAGTCT		AAGAGA		KSATALKSIKAHMAS
		CGTGCAAAAGTTGAATGCCAGACTCGAAGAAGCTGCTCCATCACTTCCA		GACACC		HHKVTMEPGKKSLV
		ATGCCGAGAAATCGATCAAAGGTCATACAGTTGACCCCCGAGAAATCGA		SAGAAA		QKLNARLEEAAPSLP
		TATCGGAATTGGAGAAAAAGAAGCAAACTCGTTCTGTGGCAAAACAGC		AAGAGA		MPRNRSKVIQLTPEK
		TTAGCACACTGAAAGAGTCGGCACAGAAAAAGGAAGAGGAGGTGAAG		GACGCA		SISELEKKKQTRSVAK
		ATAGCGGAGGTCAAAAAGAGAGAACCCCGTCTATCAATAATCCCAGAG		GAGAAA		QLSTLKESAQKKEEEV
		TCGAATGTCAGGCGAAGTCTGGCGGCAGGACTCGAACAATGTATAAAC		AGGAGA		KIAEVKKREPRLSIIPE
		CCTGAGCAATCGGTAGCTCAGAGGATAAGAGAAAAAAGAGAAGAATAC		GACACC		SNVRRSLAAGLEQCI
		GCCAAAGCTTCTAGGGAGGCAGCGGCAAAAAGAAGATCGAGTTTGGCA		TCTCATA		NPEQSVAQRIREKRE
		ATGAAGCCAGCTAGATTACCAGACAAAGAAAACGAGATTACACTCCAG		AGGAGA		EYAKASREAAAKRRS
		GAAACGAAAAAGATCGATGATCCAATCGTTATAGACCTGGAAAAAGAA		GGTAGG		SLAMKPARLPDKENE
		TGTATTCTCACTACAGTACTTCAAGTCCCAAGAAACCAGTTCAACTCGTG		TCAATCC		ITLQETKKIDDPIVIDL
		GTGTCTAGAGCATGAGACAACGATTGACGCTTGGTTAACGGATGAGGT		AAATGT		EKECILTTVLQVPRNQ
		AATACATATGTACATGTGCACAATAACCGAGAATCGAAAATATTTTATG		AAACAG		FNSWCLEHETTIDA
		GCAATCGATCCGGTTCTGTGGCCAGTCTATGTGAGAAATGGAGCAGAG		AAAAAA		WLTDEVIHMYMCTI
		GATCTACTGAGGCGTACTAGTTGCCCAGGAACATTCTTCTTTCCAATTTG		CCAGTG		TENRKYFMAIDPVL
		TGAAAGTAATCATTGGGTTCTATTAGTGATAGAACACGATGTGTATTGG		GGGAGG		WPVYVRNGAEDLLR
		TATCTGGATCCGAAAGGCGAGGAACCAAAAGGAAATGTAGAGATTCTT		AAAGAA		RTSCPGTFFFPICESN
		TTAGAGTCCATGAAAAGGAAAAGGCAGTACTATGAATTCCCACCACCCT		AGACTG		HWVLLVIEHDVYWY
		CACAGAGAGATAATGTGAATTGTGGAGTGCATGTCTGTCTTATGGCAAA		ATTTCAC		LDPKGEEPKGNVEILL
		ATCAATAGTAGATGAATGTGGTTATAATTGGTATTCTGAAGAGGACGTA		CCACTA		ESMKRKRQYYEFPPP
		AGGTCATTCAGAACCAATATGAAGGACATTCTGAAAAGTAAGGGATAT		AAATGA		SQRDNVNCGVHVCL
		GAGTTATGTCCTGAGCCTTATAATAGGCAAAATTTATTAAAAACAGAAA		ATTTGG		MAKSIVDECGYNWY
		AACAAAAGGAAGTTATTCTGGAAGAAATGATCGATTCATTCGTTGTAGA		AAACAG		SEEDVRSFRTNMKDI
		AGACGATATGACGTTCACAGTGCATCGGGATTCTGATCATGGTGATGAT		AATTTG		LKSKGYELCPEPYNR
		GAAGTTGAACATCTGAAGACCATTGAGCAGGAACCTGAAAATGAAATA		GAAGAG		QNLLKTEKQKEVILEE
		AGTGAAATTGAGAATGTAGAGGGATCTGTAGACTCAGTCATTCCAAAGT		AAAAGA		MIDSFVVEDDMTFT

					GAAAGG	VHRDSDHGDDEVEH
					GAAACC	LKTIEQEPENEISEIEN
					TAAAGA	VEGSVDSVIPKLMEM
					AAATAG	RVQTPPVINEKRGKK
					TTCTCTT	RVSAKEKPRKQKEKE
					GCCAAA	QKVPTGKPDELVKRV
					ATTCTGT	RVWFEKEFKSYVEDG
					AGAGGA	KSFQRLEWXTDVLTA
					ATACTTT	AIQKASAGDEKAVELI
					GTCAAA	EKRCPPLEXEEGEMC
					ACATGA	TQTEKKKKPKSGKGN
					TAGAAA	GGQESMKSLMASYS
					CCAGTA	ENRAKTYNRIIGKHSK
					ATCTGG	QCEIPIAKVQKFFEGT
					TACGAA	TAETNVPKETLKEMC
					AGACAA	SRLPKVEVGTWIEGE
					GTAAGA	FSESEVTEALKKTKDT
					CCTGAA	APGVDGLRYHHLKW
					CTGACA	FDPELKMLSQIYNEC
					AGAAGG	REHRKIPKHWKEAET
					AAGTCA	ILLYKGGDESKXDNW
					GAAAGA	RPISLMPTIYKLYSSL
					AATACC	WNRRIRAVKGVMSK
					GCTCAC	CQRGFQEREGCNESI
					AAAGCC	GILRTAIDVAKGKKR
					TGTGAT	NIAVAWLDLTNAFGS
					CGATTCT	VPHELIKETLESYGFP
					CTTACCT	EIVVDVVEDMYRDAS
					ACTGAA	IRVTTRTEKSDQIMIK
					CTTGTTC	SGVKQGDPISPTLFN
					TCTTGG	MCLESVIRRHLDRSV
					CCTCGTA	GHRCLKTKIKVLAFAD
					ACCGGC	DMAVLAESSEQLQKE
					TAAAGG	LTAMDADCSALNLLF
					GAGAAG	KPAKCASLILEKGIVN
					GAATGT	RLNEVVLRGKPIRNL
					TAATTG	MENETYKYLGVQTG
					GAGATA	TETRVSIMDHITEVSR
		TGATGGAAATGAGAGTGCAGACACCTCCAGTGATCAATGAAAAAAGAG				
		GTAAAAAGCGAGTATCGGCCAAAGAGAAACCGAGAAAGCAAAAGGAA				
		AAAGAGCAAAAAGTGCCAACAGGAAAACCAGATGAGCTGGTTAAAAGA				
		GTAAGAGTATGGTTTGAGAAAGAATTCAAATCGTATGTGGAAGATGGA				
		AAAAGTTTCCAAAGGTTGGAATGGMTAACAGATGTTCTCACTGCAGCA				
		ATTCAGAAGGCGTCAGCCGGAGATGAGAAAGCAGTAGAACTGATTGAG				
		AAAAGATGTCCACCTTTGGAAWKCGAGGAGGGTGAAATGTGTACCCAG				
		ACTGAAAAGAAAAAGAAACCAAAAAGTGGTAAAGGGAATGGCGGTCA				
		AGAAAGTATGAAGTCCTTGATGGCCTCATACAGTGAGAACCGAGCCAA				
		AACCTACAATAGAATAATTGGTAAGCATTCAAAGCAGTGTGAGATCCCA				
		ATAGCCAAAGTACAAAAGTTCTTTGAAGGGACCACTGCCGAGACAAATG				
		TGCCAAAGGAAACACTTAAGGAAATGTGTTCACGCCTCCCGAAAGTTGA				
		AGTGGGAACGTGGATTGAAGGTGAATTCAGTGAAAGTGAAGTGACTGA				
		AGCATTGAAGAAGACAAAGGACACAGCACCAGGGGTAGATGGATTAA				
		GGTACCATCACCTGAAATGGTTTGATCCCGAGTTGAAAATGCTGTCACA				
		GATCTATAATGAGTGTAGAGAACACAGAAAAATTCCAAAGCATTGGAA				
		AGAGGCAGAGACAATTCTTCTCTATAAGGGAGGAGATGAGTCMAAAM				
		CGGATAATTGGAGGCCTATCAGTCTGATGCCAACCATCTATAAACTGTA				
		TTCTAGTCTCTGGAACAGAAGGATTAGAGCAGTGAAAGGGGTGATGAG				
		CAAGTGTCAGAGAGGTTTCCAAGAAAGAGAAGGATGTAATGAAAGTAT				
		CGGAATATTGAGAACAGCCATAGATGTGGCCAAGGGCAAAAAGAGAAA				
		CATAGCCGTAGCATGGTTAGATCTCACGAATGCCTTTGGATCAGTACCA				
		CATGAGCTGATAAAAGAAACTCTGGAATCTTACGGATTTCCAGAAATAG				
		TAGTAGACGTCGTAGAAGACATGTATCGAGATGCATCGATCCGTGTGAC				
		GACGCGAACGGAGAAAAGTGATCAGATTATGATCAAGTCAGGAGTGAA				
		GCAGGGAGATCCAATCTCGCCTACTCTCTTCAACATGTGTCTCGAGAGT				
		GTCATCAGAAGGCATCTCGACAGATCAGTCGGCCATCGGTGCCTGAAAA				
		CAAAAATAAAAGTATTAGCCTTTGCAGACGATATGGCAGTATTAGCAGA				
		AAGTAGTGAACAGTTGCAAAAGGAGTTGACAGCTATGGATGCTGACTG				
		CTCAGCACTGAATTTGCTATTCAAACCGGCTAAATGTGCAAGTCTGATAT				
		TGGAAAAAGGAATAGTAAACAGGTTAAATGAGGTAGTTTTGAGAGGGA				
		AACCGATCAGAAACCTCATGGAAAATGAGACCTACAAGTACTTAGGTGT				
		TCAGACAGGTACGGAAACAAGGGTTTCCATAATGGATCATATAACGGA				
		AGTGTCAAGGGAGATAGATCTAGTGAATATGAGTCAACTGGCAATGCA				
		CCAGAAACTAGATATACTCAAAGCCTTCATACTTCCAAAGATGACCTATA				
		TGTATCAGAACACGACACCTAAACTGTCAGAACTGAAAGTGTTTGCCAA				
		TTTGGTAATGAGGTCAGTGAAGGAATTCCACACAACATTCCCCTAAAAGGG				
		TCACCGTTGGAGTATGTCCAACTTCCCGTAGGAAAAGGAGGATTAGGA				

		GTGGCATGTCCAAAGAACACAGCCTTATTAACATTCTTGGTAACCATTAT		GACATA	EIDLVNMSQLAMHQ
		GAAAAAGTTATGGTCATCTGATAGCTATATCAGAAAGTTGTATACAGAC		AAGATA	KLDILKAFILPKMTYM
		TACCTAGAGGAGGTGGCAAAAGTGGAAATTGGAAAGTTCGAGGTCAAC		GGTGGA	YQNTTPKLSELKVFA
		TTGAACGATCTAGCAGAATTCCTAAGTGACGAAAGAGCAGTCGACAGC		GTGAAG	NLVMRSVKEFHNIPL
		AAGTTGTTCGGCTTCAATGCGTTCACGAGGGTGAGAGAAGTGGTGAGG		GTCCTG	KGSPLEYVQLPVGKG
		AGTCTCTGTAAGAATAAAGATTCTCCACTACATAGTCTGAAAATAATTGA		TTCTTGA	GLGVACPKNTALLTF
		AAGAGAAGGGAAACTTGCCATAAGTGTGCAAGCAACCGAAGAAAGTAT		AACTAG	LVTIMKKLWSSDSYIR
		TGAGAAAATCTTCACTGAAGACCAGGAAAAGAAGTTAATGTACCTACTG		GAGGAA	KLYTDYLEEVAKVEIG
		AAAGGGAGCTAAATACAGCTCTCCAGCACAGGTTCTTTACTCAAAAGG		TGTGGA	KFEVNLNDLAEFLSDE
		TATTCAAAAGTGAAGTAATGAGAGTGGTTCAACAGCATCCACAAAGTAA		AAGAGC	RAVDSKLFGFNAFTR
		CAGTTTTGTCAGAAATGGTGGAAAAATGAGTTTTTCGGCTCAAAGATTT		AGAAGG	VREVVRSLCKNKDSP
		GTCCACCCAGGAAGACTGAACCAGTTGCCATGTAACTACAACACTTGGG		CCGCGA	LHSLKIIEREGKLAISV
		CAAAAGGCCGTACGAAGTTGTGTAGAAGGTGTGCAAAGAATGAAAATG		GGCTTT	QATEESIEKIFTEDQE
		AGACACAGTCGCATATACTGCAAGTGTGTGACTACTCAATAGGAAATAT		AGACGG	KKLMYLLKGELNTAL
		CATAAAGGAAAGACACGATGCAGTTCTTTATAAGTTTAGAGAACTCATT		GTAACT	QHRFFTQKVFKSEV
		AAAAGAGGGTCAAAAGGTCATTGGTTAGAGAGAACTGACCGGACAGTA		CAGTCA	MRVVQQHPQSNSFV
		CCAAATACTGGATCACAGCTGAAGCCAGATCTCTATCTGGAAAGCCCAG		GTTGCT	RNGGKMSFSAQRFV
		ACGGGAAGCATGTGATACTAGCCGATGTGACAGTTCCATATGAAAGAG		AGTGGT	HPGRLNQLPCNYNT
		GCATCGAAGGAATGCAAAAGGCATGGAATGAGAAAATCAACAAGTATA		CTTCGG	WAKGRTKLCRRCAK
		CTGATGGATATAAAGAAATATTCAGAAGACAAGGAAAATCCCTAGTAGT		ATCCAA	NENETQSHILQVCDY
		GTTACCATTAGTAGTTGGTTCACTGGGAACGTGGTGGAAGCCCACGGA		CGGCTT	SIGNIIKERHDAVLYK
		GGAAAGTCTGATCAAACTAGGTGTTGAGAAGACTACAGTAAGAAGGAT		CGGACA	FRELIKRGSKGHWLE
		AATACCTGAGACGTGTGGAATGGTGGCTGAATACAGTAAGAACTGCTA		TAGTGA	RTDRTVPNTGSQLKP
		TTGGAGACACATCTACGGTGAAAAGTATGTTCAAACTCCAATGATAAAT		GGAACC	DLYLESPDGKHVILAD
		GGAGGAAAAAAGCCTGAAGGAAATGATTGGAAAAAGTGTGAAAAAGG		CTGGGT	VTVPYERGIEGMQKA
		AATAGAAGTTCCTAAAGTTACTAATTAGCCGATCGTAAAAGAAACCGAG		ACGGAG	WNEKINKYTDGYKEI
		CCGTAACAACAAGCAAAGTAAACAAAAGAAAAATCAATAAAAAGGAAG		AAGAAA	FRRQGKSLVVLPLVV
		GTTGACCTCAGACCCCGAGGAGGGAAGAGAGACACCSAGAAAAAGAG		TGGAAA	GSLGTWWKPTEESLI
		AGACGCAGAGAAAAGGAGAGACACCTCTCATAAGGAGAGGTAGGTCA		AGAGAT	KLGVEKTTVRRIIPET
		ATCCAAATGTAAACAGAAAAAAACCAGTGGGGAGGAAAGAAAGACTGAT		AGGGCG	CGMVAEYSKNCYWR
		TTCACCCACTAAAATGAATTTGGAAACAGAATTTGGAAGAGAAAAGAG		GGCAAA	HIYGEKYVQTPMING
		AAAGGGAAACCTAAAGAAAATAGTTCTCTTGCCAAAATTCTGTAGAGGA		GGCTAA	GKKPEGNDWKKCEK
		ATACTTTGTCAAAACATGATAGAAACCAGTAATCTGGTACGAAAGACAA		GTTCATA	GIEVPKVTN (SEQ ID
		GTAAGACCTGAACTGACAAGAAGGAAGTCAGAAAGAAATACCGCTCAC		CACTGTC	NO: 1416)
		AAAGCCTGTGATCGATTCTCTTACCTACTGAACTTGTTCTCTTGGCCTCGT		ATGCAA	
		AACCGGCTAAAGGGAGAAGGAATGTTAATTGGAGATAGACATAAAGAT		CCACTA	
		AGGTGGAGTGAAGGTCCTGTTCTTGAAACTAGGAGGAATGTGGAAAGA		AACCAG	
		GCAGAAGGCCGCGAGGCTTTAGACGGGTAACTCAGTCAGTTGCTAGTG		TGGGAT	

			GTCTTCGGATCCAACGGCTTCGGACATAGTGAGGAACCCTGGGTACGG AGAAGAAATGGAAAAGAGATAGGGCGGGCAAAGGCTAAGTTCATACA CTGTCATGCAACCACTAAACCAGTGGGATCTGCGGGTGAATCACTTTCG AAAAGAAGTGAATGGACGTGCTGATGTCTGACTTTAAAGAAGTCTGAA ATTAAAAAAACAGATATAAAGGCCCCTCACTATAAACTCCACAGCAACA GGTGGTCCGGCGAGGCCGGTTCTTGCCACCATTGCACCCCAGGCTCGTC (SEQ ID NO: 1048)			CTGCGG GTGAAT CACTTTC GAAAAG AAGTGA ATGGAC GTGCTG ATGTCT GACTTTA AAGAAG TCTGAA ATTAAA AAAACA GATATA AAGGCC CCTCACT ATAAAC TCCACA GCAACA GGTGGT CCGGCG AGGCCG GTTCTTG CCACCAT TGCACC CCAGGC TCGTC (SEQ ID NO: 1294)	
NeSL	NeSL-1_TV	.	Trichomonas vaginalis	GGGTGAGTAGTCTAGTGGTATGATTCCTGTTTTGGGTACAGGAGGTCCC GAGAAGCTTCCACTGCAATCGTACGTGTACTGTGGCAACACAGCTATAA CAGACAGTTTCACGCCAACCGCGAAAACGATTTTGAAGCCTGAGGAACA AAATTTAGATATCGTTTTGAAAAATATTGCAGCGTTGAATCCAGAAAATT ACTCCGACTTAATCAGGAGCCTATCGAAGATGGAGTTCAGATTAGATTA CCCGAAAGAAATAGAGAATTACTGGATTTCGGAAAAATTATTTAGCCAA TCCATCGCATCATTGCCCATCAGTTTGTTAGTCGCATCCATGTTCTCACCT GAAGACCGTGACTTGAGTACAGAACCGTTCCACTGTAACGCTGATGGCT	GGGTGAG TAGTCTAG TGGT (SEQ ID NO: 1172)	TAAGAA GAGATA AGACGA GTGAGA AGAACA GAAGCA TAGTAG GATTGG	MIPVLGTGGPEKLPL QSYVYCGNTAITDSF TPTAKTILKPEEQNLD IVLKNIAALNPENYSD LIRSLSKMEFRLDYPK EIENYWISEKLFSQSIA SLPISLLVASMFSPED RDLSTEPFHCNADGC

		GTAATTTCCATTGTGACAATTGTGAAAGAATGGTTGAACACATCAGAGA		CAGAGC	NFHCDNCERMVEHI
		GCACCATAACACTGACCCCATGATCAATACATTTGAAACAACAGAAGAC		TTAAGC	REHHNTDPMINTFET
		ACATTTAGAAGAATAACGGCCATCAAAATAGACAAGACAGGCATCGAA		GATGTC	TEDTFRRITAIKIDKTG
		GAACTTAACCCTCTAAAATACAGATGCTCGTATTGCGACGAGTTATTCAC		ACTCGG	IEELNPLKYRCSYCDE
		CGAAGCAGAAGATCATGCCATCCATATGATTTCACATCTCACAGAAAAA		TACGAA	LFTEAEDHAIHMISHL
		TTATCACCAGATATATCTTTCTTTTTCAACGACATTTTACGCCTTTACAAA		ACGTGT	TEKLSPDISFFFNDILR
		ACTATCGACAAACCAACAGTACAAAAATTTATTTCCAGAAACACAAGTCG		ACCAAA	LYKTIDKPTVQNLFPE
		CAATTTTTGACACACTTGAAGAAACAAACAGATTCAGACTTATCGTAGG		CACCGG	TQVAIFDTLEETNRFR
		AAGAGAAGCCATAGAAACAATTGAAGAAGCATTCCCTCCAAGTCCACCA		ATTCCGT	LIVGREAIETIEEAFPP
		GGAACAGATCGGAAACCATCCATAATCATCACAGACACCTGTCAACTCA		GCTAGG	SPPGTDRKPSIIITDTC
		GGTTTGTACCATGCATGGATGAACCACCAAAAGGAGATCTCGGAATTCT		AATCAC	QLRFVPCMDEPPKG
		GACTCTACTTTTAAGAGATTTCAGCGCACACAATATCCCGATTAAATCAC		AAGCCA	DLGILTLLLRDFSAHNI
		TGAACAATAAGGAACTAATTGCTGATAAAGACATCGATTACAGCCCAGA		AAATAA	PIKSLNNKELIADKDI
		TTTTGTCGAAGGAGCTCTAGCCAACGCAGAAGAACATGATACAACGAAC		AAGAGA	DYSPDFVEGALANAE
		AGCCAGAACAACAATGGAAGATACATTAACTCAGCCGAAAAACTTACAG		CACCAC	EHDTTNSQNNNGRY
		AATTTTTAATACAATGTGAAGACTACTTAACGAACATCAAAACACTTGAA		GAAAAT	INSAEKLTEFLIQCEDY
		GACTTAGAACGTTTCTACACAACGATTAAAGACTACAGAGTCAACAAAG		TACTCAC	LTNIKTLEDLERFYTTI
		AGGTTATCGCCGAAGATACACCAATCTTTGTATATTTCCTAGTAGAAGAA		CCTCCCT	KDYRVNKEVIAEDTPI
		GGGAAATTACCAAAACCAGGTCTTAGATGCCCACTTGAATCATACGAAG		CAAACA	FVYFLVEEGKLPKPGL
		GACACGAAGACAAGGCATTCGAATCACTGAGAAAACTTTGCGACCACTT		GATAAT	RCPLESYEGHEDKAF
		CAAAGGAGAAATCGCGAAAACGAGCTTTGACCCAAAGGTTCACACCAT		AATATTA	ESLRKLCDHFKGEIAK
		AGACATCTGGGTTGAATTTTTGGCCCAAGCCTATGGCACAGGCACGTTT		ACCTCCC	TSFDPKVHTIDIWVEF
		GTCTACAAAGATGAAAACGGAAACATCGACCTTGATACGCACGTATTCA		ATCCATC	LAQAYGTGTFVYKDE
		AATGCCCTTATGCAGACTGCTCATACACGAACAACGACAGATCAAAACT		AGTCCG	NGNIDLDTHVFKCPY
		CATGGACCACATGAAAACGAAGAAACACGCCAAGAACGTATACATCGA		TATGGT	ADCSYTNNDRSKLM
		GAGATACGGCTTCTTTTGGGGTATTGTCATAGAAGGAGTCAACCGACCA		CTGATA	DHMKTKKHAKNVYI
		AAAGGAATCGTCTACCCGACACTCAAAGACATCAAAGAACACGCTTGTC		ACAGAC	ERYGFFWGIVIEGVN
		GCAAATGTCCAGAAGCAGGATGCAACACATATGTAACAGAATTGAGCG		TAGCAC	RPKGIVYPTLKDIKEH
		ACATCAAAGAACATCTAAAGAAGAAACATAAGTCTACACAACAGCAGGAG		CACATCC	ACRKCPEAGCNTYVT
		TAGACGGAGAAATCGCGCACACTGATGCTACATACTGCTGGATTACCAA		ATGATA	ELSDIKEHLKKKHKST
		AGAAGAACTCGACGCATTACATGCCGAGAGAGCAAGAGAAAGAGCAG		CACTCAT	TAGVDGEIAHTDATY
		AGCAAGTAGACAACACTCCAGTACAACAGATAATTAATGCTGACAACAA		TGGAGT	CWITKEELDALHAER
		TGAAGAGAACAACGAGAACCAAGAAGACAACGGAAACAACGAAGAAG		GAAAAC	ARERAEQVDNTPVQ
		CAGATGCCCTCGACCCGCCAAATAACACAACAGAGACAGAAGATGAAG		CACCAA	QIINADNNEENNEN
		CGGTTCATGCCGTCATCAATCCACCAGCAACAGAAGAGGAAGAGGT		CAACAA	QEDNGNNEEADALD
		AGCCATCATCGCCGAGGCAAGAAGAAACATTCCAGAACTCCAACACAAGC		ATCCACC	PPNNTTETEDEAVHA
		AGAAGAGAGAGGCTGCGTTACACCGAAAATGACATCACTCGTCCGATTA		TAGACC	VIINPPATEEEEVAIIA
		AAACTATTGAAAGGAGGAGGAGAACTTTTCAACAAGAAACTCACTCCAT		AAATCCT	EARRNIPELQQAEER

		TAGCCACAAGATACGCAGCTACAGGAAATACAGAAGCAGACAAAATCA		GCCCCA		GCVTPKMTSLVRLKL
		AGGTAGATTACTTGACACTAAAATGCAATGCCGCCTTGAGAGAAATGAT		CCTCCAC		LKGGGELFNKKLTPL
		CTACACCAATAACCACAGCGAATCAAAGTTTATGACAGCAGAAAATGGA		CCAAGT		ATRYAATGNTEADKI
		GAAGACACAGCACCACCGCCAAGGATATCGGAAGACACAAGAGATCGC		AGCTCG		KVDYLTLKCNAALRE
		ATTCAAAAAGCAGCCAATGAAATAAAAGGAACTCTCATCAAAGTAGTCA		CTTCGCT		MIYTNNHSESKFMT
		AACACATAAGTCACGCGAGATGCCTCAAAGACAGCACGAGAGACGATG		CGCTCA		AENGEDTAPPPRISE
		AACACAATAAATTCGTCGAAATGATTGCAAAAATCAAAAACGATCTCAG		CCTAAA		DTRDRIQKAANEIKG
		AGATAACAAATTCGAACAATATAACATTGAAGAAATATTTCAAGGACCG		ACTTTGC		TLIKVVKHISHARCLK
		ATCTCCGACCAGAGTATTCTCAACATCGTCAACACGGAGGACAACAACG		TCGCTC		DSTRDDEHNKFVEMI
		AATTCATCAAGAAAATGGATTACATTAACCGAATTCTCGGAACACCACA		GCTTCG		AKIKNDLRDNKFEQY
		GGATGCATCACCATATGCAAGGAAGAAGTTACAAGCATGTTTCGCCGAT		CTCGCTC		NIEEIFQGPISDQSILN
		AACCCAACAAAGACTCTCAGAAACATAATCTTAGCCGACAAAGTTCCAC		GTCTTAA		IVNTEDNNEFIKKMD
		AACAATCATTGAAGCCAAGCGAATACCTTGATTACTACGGACCTCAATG		CCCTTTC		YINRILGTPQDASPYA
		GGCAAACGAAGCTGAAGGCTACGAAAAACTTCCTGCATCATGACTACGCG		CGAATA		RKKLQACFADNPTKT
		TTACCGGAGAGATATGGCCAAGTTTTCGCAAACGACTTCCTCGACTTCAT		AACACTT		LRNIILADKVPQQSLK
		GACAAACGAATCGAAGATCATCGAAGTAATCCGCAACAAGAATCATTTA		ACAATTC		PSEYLDYYGPQWAN
		TCGGCACACGGCCTCGATGGAATTCCGAACTCAGTTTACATGCTATTCCC		CCGGCT		EAEGYENFLHHDYAL
		AGTCAGCGCCGCAAAATTCCTCAGTATATTATTCAGATCAATCATCATAT		CGCCCC		PERYGQVFANDFLDF
		CAGGTCACATCCCAGACTGCTGGAAGCTCTCCAAGACAGTGATGCTTTT		ATTTTTT		MTNESKIIEVIRNKNH
		TAAGAAGGACGACCCATCGTTAGCAAAGAACTGGAGACCAATCGGCAT		(SEQ ID		LSAHGLDGIPNSVYM
		CACGTCATGCACTTACAGAATCTTCATGACTTTAGTCAACAAAGCGTTAC		NO:		LFPVSAAKFLSILFRSII
		AGATGATCCCAATGTTCCACGCAATGCAAAAAGGTTTCGTTCGCGGAGC		1295)		ISGHIPDCWKLSKTV
		AACACTGAGTGAGCACATTGCAGTCGCGAACGAAGTCCTTTGCCAATCA				MLFKKDDPSLAKNW
		ACCAGAACACAGTCTGAAATGTTCCAAACAGCAATCGATTTCACGAACG				RPIGITSCTYRIFMTLV
		CTTTCGGCACAGTTCCTCATCAATTGATCTTTGATTCTCTCGAAGCGAAG				NKALQMIPMFHAM
		AAAGTTCCCGATTCGATCATCAATCTGCTCAAGGACCTCTACAAAGGAG				QKGFVRGATLSEHIA
		CAAGAACGGCTATCTATACAAGACATGCACACTCCGAGATAGTTCCGGT				VANEVLCQSTRTQSE
		TCGCAGAGGTGTCATCCAAGGCTGTCCACTCAGTCCAATCCTCTTCAACT				MFQTAIDFTNAFGTV
		GCTGCTTAGATCCTTTATTATATATGCAGTCCAGAGGAGACACTTTGAGGA				PHQLIFDSLEAKKVPD
		CGGTTACAGATTCCAAGACAAAGCAGGACAGTATTCAATTGCCATTCAA				SIINLLKDLYKGARTAI
		GCTTACGCTGACGACGTTCTAGTCATCTCTCCAACACATGAAGGAATGC				YTRHAHSEIVPVRRG
		AAAGAATCTTAAACACAGTAGATGAATTCCAGAAAATTGCGAAACTCAA				VIQGCPLSPILFNCCL
		AGTTGCACCACAGAAATGCGTCACACTTGCCAAACATCCCACTGCAATC				DPLLYAVQRRHFEDG
		CAACCTTTCCGCATTGGTCCAGACGAAATCCCAATCAAGACGAGCATGG				YRFQDKAGQYSIAIQ
		ACAACATCACATATCTTGGAATACCAATCTCTGGAACAAAGACATCAAG				AYADDVLVISPTHEG
		ATTTGCAGCTGCAACTGGCATTCTGGAAAAGGTCAAAGCACAGATCAGA				MQRILNTVDEFQKIA
		GTCGTCTTCGCGTCACATCTCGCTCTCTCTCAGAAGATTATCGCTCTCAG				KLKVAPQKCVTLAKT
		AGTCTTCATCTTGCCACAACTTGACTTTTACATGTTCCACAACGTATTCAG				STAIQPFRIGPDEIPIK

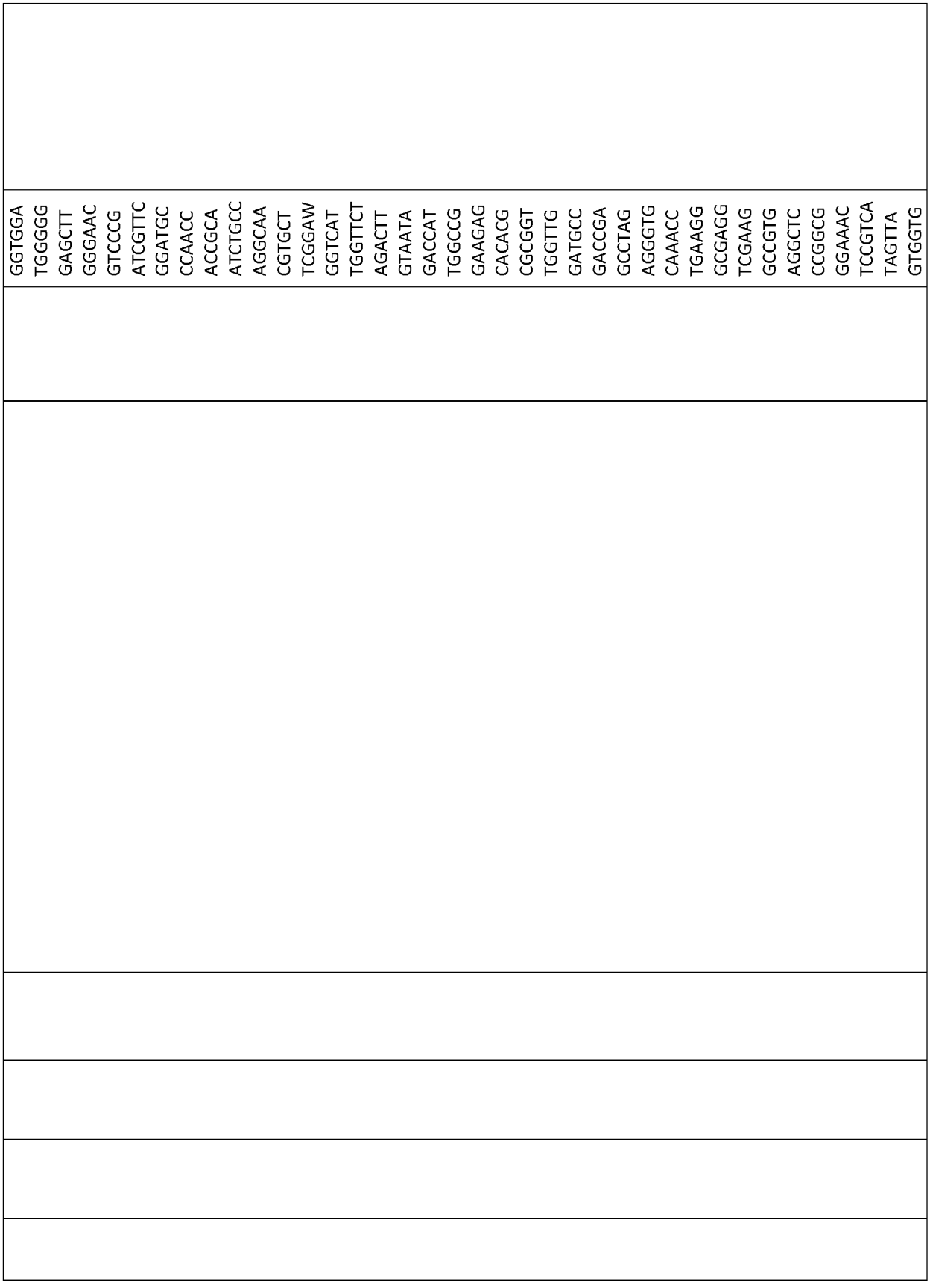
		AGTCAATGACTTGAAAGCGACAGATCAGATGATCCGAGGCCTGATCGA			TSMDNITYLGIPISGT
		CAAAGAAGCGCCGACGTCAAACATTCCGGTTTCATTTTTCTACATGCCGA			KTSRFAAATGILEKVK
		AGAACAAAGGCGGCTTTGGACTCGTTAAATTGGAACTTCGCCAGCCTCA			AQIRVVFASHLALSQ
		GCTCGTTCTCACTAAATTTGCGAGGTTATGGTTAAGTCAACAAGCAGAA			KIIALRVFILPQLDFYM
		ACCAAAGCCTTCTTTCACACAATGGCTCAAGAAGAGAAGTCATTCCGCA			FHNVFRVNDLKATD
		AGGTCGTCGAAGACCAAGAAAATGGTTTCTTAGGCATCAAGATGGAAA			QMIRGLIDKEAPTSNI
		ACGGCAAAATTGTCCAGAAGAACGAAAGATCCAAACGCACACAAATTGTTT			PVSFFYMPKNKGGF
		CATCACACAGGCGGCTAAAGCAGCAGACAAACTGGAAGTCAGATTCAA			GLVKLELRQPQLVLTK
		AGAATGGGACAAAGGAGGCATACAAGTCAGAGGTGTAGGAGAAAATG			FARLWLSQQAETKAF
		CAACAGACTGGTACCGCTCGAAACACATCGGCCAAATCTCACCCTTAATC			FHTMAQEEKSFRKVV
		GGTCGCGTCATCCAACAGAGGCAGTACGAGGAGTTCAAGAAAGACGAA			EDQENGFLGIKMEN
		ACACACTCACACACTTTCTGCGAACCAGCAGCGCTAGCGGAGTCACACG			GKIVQKNERSKRTNC
		ACATCATGAAGAGACCACACAAGCTGTTCCAAACAACCTCTACTCAGCGGC			FITQAAKAADKLEVR
		TATTGCTCTCCGTACAAACACAGCTCCAACCCCAGCAAACATGCACTTCC			FKEWDKGGIQVRGV
		ACAACCCAGAAGTTTTGGCTAATTGTCCATTGTGCGGATGCCAATCCTGC			GENATDWYRSKHIG
		ACTCTCTTCCACACATTGAACATGTGCAGAAACCGTTTCAGTCTATACAA			QISPLIGRVIQQRQYE
		ATGGCGCCACAATATCATATGCGATGACATTTACCAATTCATTCACGATC			EFKKDETHSHTFCEP
		ACTATCCAGGAGTAACCATCAAATGCTCGGCGAGAATTACAAGTGACGG			AALAESHDIMKRPQA
		CTACCAAACAACAGGCCCAGAGCTCGACGACACAGTTAAAGATCTCCTC			VPNNLYSAAIALRTN
		CCAGACCTTGTTGTCTACGATGAAGCGAACAAGATGATCAAGATCATTG			TAPTPANMHFHNPE
		AAGTCACATGCCCTTACGGCACGGACAACAATGTTGGCAACTCTCTTGA			VLANCPLCGCQSCTL
		CGCGGCATACGACAAAAAGGTTAACAAGTATAAGAGCCTTGCTGAACA			FHTLNMCRNRFSLYK
		AACAGAGAGATTATTTAACTGGACCACGACGCTCTCAATTATCGTAGTCT			WRHNIICDDIYQFIH
		CATCACTAGGAGTCATCCCTCTCCGTACAAAACTCGACGCATTGAGAATC			DHYPGVTIKCSARITS
		TCACCTGCAGATCACATACAGCTACTCAAGAGACTTTCGATGCACGCGA			DGYQTTGPELDDTVK
		TAGCTGCGAGTGCTTGCATTGTTTTTGAAAAAGTGCCAGAATTCTTCGGT			DLLPDLVVYDEANK
		ATGCGCTGCCGTCCCCTCCCAGGACGAGTCACAGCTCCCAATGCAGCGA			MIKIIEVTCPYGTDNN
		TCCCACCAAACAACAATGAAAACAATAACGACACAGATCATGGTCAGGA			VGNSLDAAYDKKVN
		GAACCAACAGGCAACCTCTGAAGAGCAACCAACCAACAATGGAAATGC			KYKSLAEQTERLFNW
		TCAAGAAGACAATGGCCAAGGCGAACAAATAAATAATTCAACCGAACA			TTTLSIIVVSSLGVIPLR
		AACTATCTCTGTTGATCAAATCATCGAAGAAGATGCTGAGAACAACGCG			TKLDALRISPADHIQL
		ATAGAACAAGCCTTAGACCAACCCGATGAGGACGAATTCCTTAACTAAG			LKRLSMHAIAASACIV
		AAGAGATAAGACGAGTGAGAAGAACAGAAGCATAGTAGGATTGGCAG			FEKVPEFFGMRCRPL
		AGCTTAAGCGATGTCACTCGGTACGAAACGTGTACCAAACACCGGATTC			PGRVTAPNAAIPPNN
		CGTGCTAGGAATCACAAGCCAAAATAAAAGAGACACCACGAAAATTACT			NENNNDTDHGQEN
		CACCCTCCCTCAAACAGATAATAATATTAACCTCCCATCCATCAGTCCGT			QQATSEEQPTNNGN
		ATGGTCTGATAACAGACTAGCACCACATCCATGATACACTCATTGGAGT			AQEDNGQGEQINNS
		GAAAACCACCAACAACAAATCCACCTAGACCAAATCCTGCCCCCACCTCCA			TEQTISVDQIIEEDAE

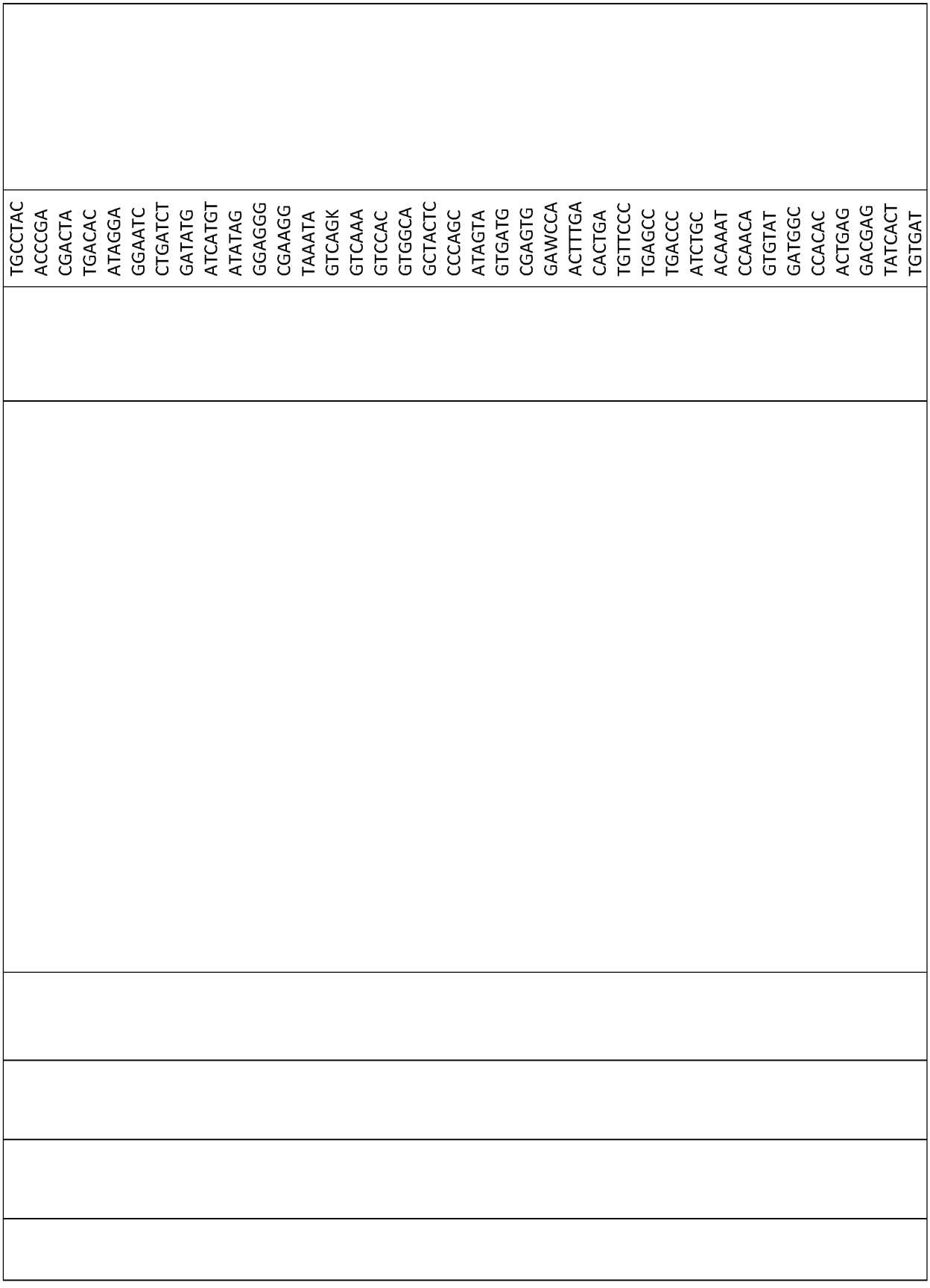
			CCCAAGTAGCTCGCTTCGCTCGCTCACCTAAAACTTTGCTCGCTCGCTTC GCTCGCTCGTCTTAACCCTTTCCGAATAAACACTTACAATTCCCGGCTCG CCCCATTTTTT (SEQ ID NO: 1049)				NNAIEQALDQPDEDE FLN (SEQ ID NO: 1417)
NeSL	NeSL-2_CBre	.	Caenorhabditis brenneri	GACTCGCCTTGGGGAAGGTWTTTCAGGGGKSAATTGCCGMAGGCAAG GCAGCCCCCSMMTAGCTTACAAAGTAAGTACMCATTTTCATTTCTTGTG AATTCTTTAAACATATTTTTTCTTGTTTTTTGATTTCTTTTTTCTCTACCTTCC CCCAATTCTTCCCCTCATCTTGTGTATACATCCCCCCTCCTCCAACCAATCA ATACATTGACCTCTCTCTTCTGTCAAAAAATCAATACTAGTATATTGTCCC TTGTATAGTATTATTTGACGTCGTCTTTGTATTAGGAGTAGGTAACAAW CTGTGTATGGCTTCAAAAAGCATGCACACAAACWCCTGTCAAAAAGTAWT TCCCATCMTGTGAATAGCTCAACGACWKGAAGMCCAATGATATGAGA TATCATMGCTCCAACMACCCAGCASCCCGCACCTCAGACAATCAMTGG AGATCAATCCMAAAGGACGTCCGTCGCCCAGATCCGTCAACTATCGAG GAGAAAAGCAGGTATAACAGGTCCATAGGTATTCCAGATTCGCTCAAAG AWCGGAGCAGTGCAGTCCGCAGCAKGAGCAGCCMACCTCCGTCAGGT CCACAGGACGTCCGTCTCWCCAATTCGCCATCTCTCGATGATAGGAGAA GGTTAGTWGATTGTGAAACAACACTAGGGTCATACCGCGAATGGACSG ATAAACCAATGATGGGAAAGATGACGTATGCGGCAGTGACAAAAAGAG CGCCCCCAAGACCGCAAACGGGAGGAGCCCGGTTGAGCACCAATCTCC TAGCAGATGAAATGGAGATAAAGTATCGAGACACCAATGACATCCGCCT TGTCATAGACCTTCCCAATCCCCACCTCATCAAGTGTCCGCTCTGTAAAA GCTGCATAAGTGCGCGGGGAAGAGGTGCTAATGCGCTGAAGTACATGA AAAGGCACATAGCCGACGCCCACCACCTCAACGCCGACTTCGTCTACAA ATGTAGCAGGTGTCAAGAGCATGAACCAGAAAATGTATGCGGCGCGAA GTGGATTGTGAATCATCTCAAAAGAGTACATGGCTATACTCTAGAAGAT GCCGTATCCACAGCAAAACCCTCTACAAGGCAGCAGATTGCAAACGCCT TCAACGACTCTGCTCCATTCATAGACGCCCGGAAAACATCCGATGTGCC AGAGAAGAAGAGCAGAGAGGCAGGACTTGAGAAGTTCCTGGCCCCTAC AAAGTCCGAGGACACAAGGGAAAAAACCCCGCCCTCCACCAGAAAGTC CTCTGAAAGTTCAGAGGCATCAATCCAATCGACCATCCAAGAGACTCTTT CGGAGTCGTCAGACACATTGACCGTCCAAGAAATAATCAATATCAGCAG TGAAGATGAAATGGACGAGGAGCCACCGAAACGGCGTGTCAATGTCTG GGCCTTGATCCATGAGAATGGCAAGGACGCCTGGATAGACTCAGACTT GATGGTCATATTCCTGGAATCAAGAGCAAGAGGATATGAATCATGCAGC ATCATAGACCCTCTGAACTTCATTTGCACTGACATGTCCTATCTGACCAC AATAGTCAGAAGGCGCATGGAAGAAGGCTACAAGAAAATCATATTTCC ATTATGTGCAAATGACCACTGGACACTGGTCACGATAACAGGTAGCACG GCCACCTTTTACGATCCAATGGGAAATGAGCCAACTGAGACTGTCAAGA	GACTCGCC TTGGGGAA GGTWTTTC AGGGGKSA ATTGCCG MAGGCAA GGCAGCCC CCSMMTA GCTTACAA AGTAAGTA CMCATTTT CATTTCTT GTGAATTC TTTAAACA TATTTTCT TGTTTTTT GATTTCTT TTTTCTCTA CCTTCCCC CAATTCTT CCCCTCAT CTTGTGTA TACATCCC AACCAATC GACCTCTC TCTTCTGT CAAAAAAT CAATACTA GTATATTG GAGGAC ATTCTCC GTGGGC GGATGG GAGGAG TAGGGT AACGAC CCGTCAT TCTGGA TGCCTA AACCAC CACAAT CTGTCA AGGCAA AGTGCC CCAAAA GCACAC GCGTGG ATCGGT TTGGAT GCCGAC TGAGCC AGAGGG NDHWTLVTITGSTAT FYDPMGNEPTETVKK MIDELDLEMQLAPS NSPRQRDSWNCGVF VMKMAEAYIKDTQ WDLTDVDTDVKTFR RSLLTELKAKFNIFAE DIQTYRPPSRKALTRN

	AGATGATCGATGAGCTCGACCTTGAAATGCAATTAGCCCCGTCAAACTC	GTATTAGG	AGTCAG	SQSPVVVCHKCSRPA
	TCCTAGACAGAGAGACTCGTGGAACTGCGGCGTTTTCGTCATGAAAATG	AGTAGGTA	TGGTGC	TPIQDVSRMEVEEAP
	GCGGAAGCGTACATCAAGGATACGCAATGGGATCTCACGGACGTAGAC	ACAAWCT	GCCTAC	VLVPTPEEPPQEWTF
	ACGGACGTCAAAACGTTCAGAAGGAGCCTCCTAACAGAGCTCAAAGCA	GTGTATGG	ACCCAA	VGKNRKRGVTSRTP
	AAGTTCAACATCTTTGCCGAGGATATCCAGACCTACCGGCCACCCTCAA	CTTCAAAA	CTGCTAT	NTSPEAKRPAFPPVP
	GGAAAGCCTTAACGAGGAACAGCCAATCGCCTGTCGTTGTTTGTCACAA	AGCATGCA	GACACA	LKPSANRWHFPEEET
	GTGCTCTCGGCCAGCCACACCGATCCAGGATGTGAGCAGAATGGAAGT	CAAACWC	CAAGGA	EKMEVSSADEVKNST
	GGAAGAAGCGCCAGTGCTGGTACCGACTCCTGAAGAGCCTCCACAGGA	CTGTCAAA	CAACCC	PPKPPKIPNLLAMKIA
	ATGGACCTTCGTCGGAAAAAACAGAAAGCGTGGTGTGACAAGCCGAAC	AAGTAWT	AAAATA	SPVPLKRGNPSKKHG
	CCCGAACACGTCGCCGGAAGCCAAGCGACCGGCTTTCCCACCAGTACCC	TCCCATCM	AATAAG	KGHMMNTARKGPT
	CTCAAACCATCAGCCAACAGATGGCACTTTCCAGAAGAGGAAACTGAAA	TGTGAATA	CCAAGG	KKEMPKGEPANLIVKI
	AGATGGAGGTCTCAAGTGCCGACGAGGTGAAGAACTCTACCCCTCCAA	GCTCAACG	CGGCGT	RSWFDEQLKMYKDE
	AACCACCCAAGATACCTAATCTTCTCGCGATGAAAATCGCCAGTCCCGTA	ACWKGAA	TAGCTTC	GSNLQRLTWLSDSLT
	CCTCTGAAGAGAGGAAACCCGTCAAAGAAGCACGGTAAAGGACACATG	GMCCAAT	GAGCTA	AAIGKAFNGNKYIVD
	ATGAATACAGCGAGAAAGGGTCCGACACAAAGAAAGAGATGCCCAAAGG	GAT (SEQ	ACAAGC	QIIKRNPPPLVEKGA
	GGAACCAGCGAACCTTATAGTTAAAATCAGAAGCTGGTTTGATGAGCAA	ID NO:	TCCCCG	MSTQTSRKRDEFKPR
	CTGAAGATGTACAAGGATGAAGGGTCCAATCTACAAAGACTCACATGGT	1173)	AGAGGA	ERMAQEPNEPLRIQY
	TATCGGACTCTCTGACCGCCGCCATCGGAAAAGCATTCAATGGCAACAA		TGGTTG	AKNRQKTFFKIIGKQS
	ATACATAGTGGACCAAATCATCAAGAGAAACCCACCACCACTTGTTGAA		CCACAG	EQCTINIETVEQHFRK
	AAGGGCGCAATGTCAACACAGACAAGCCGAAAGAGAGACGAGTTCAA		GGCACC	TLKAPVVSENAIKTVC
	GCCAAGGGAGAGAATGGCCCAAGAGCCCAATGAGCCGCTTCGTATTCA		ATCCTG	GSIKKVLMPKTIEDPI
	ATATGCCAAGAATAGGCAAAAAACGTTCTTCAAGATCATTGGGAAACAG		GGGAAC	SSVEVKSILTKVKDTS
	TCTGAACAGTGCACCATCAACATTGAGACTGTCGAACAGCACTTCCGAA		GACCCG	PGTDGVKYSNLRWF
	AAACACTCAAGGCTCCTGTAGTCTCAGAGAATGCAATTAAAACGGTCTG		ATCTTTC	DPEGERLAKLFEECRK
	CGGAAGCATCAAGAAAGTATTGATGCCAAAGACCATAGAAGACCCAAT		GGATGC	HREIPSHWKEAETILL
	CTCCTCCGTAGAAGTCAAATCCATCTTGACGAAAGTGAAAGACACGTCA		CCAACC	PKDCSDEEKKKPEN
	CCAGGAACAGATGGTGTCAAGTATAGCAATCTACGCTGGTTCGACCCAG		ACCGCC	WRPIALMATIYKLYS
	AAGGGGAACGCCTCGCCAAACTGTTCGAAGAATGTCGCAAGCATAGAG		AATCTGT	AVWSRRISGVQGVIS
	AAATACCCAGCCATTGGAAGGAGGCGGAGACGATCTTATTGCCAAAGG		CAGGCA	PCQRGFQSLDGCNES
	ACTGTTCAGATGAAGAGAAGAAAAAGCCGGAAAATTGGCGTCCCATCG		ACGTGC	IGILRMCIDTASVLNR
	CCCTCATGGCAACAATCTATAAGTTGTATTCAGCAGTGTGGAGCAGAAG		CCCAAA	NLSCSWLDLTNAFGS
	AATCTCCGGTGTTCAAGGGGTAATTAGCCCGTGCCAAAGAGGCTTTCAG		AGCACA	VPHELIRRSLESFGYP
	TCCCTCGACGGATGCAATGAGTCGATCGGAATATTGCGCATGTGTATTG		CGTGCG	QSVIQIVTDMYKGAT
	ACACCGCTTCCGTACTCAATAGGAATCTCTCTTGCTCATGGCTTGATCTC		GAGCGG	MKVKTADQKTQSIKI
	ACCAACGCCTTCGGGAGCGTTCCTCACGAGCTGATAAGGAGATCCCTAG		TTGGAT	EAGVKQGDPISPTLF
	AATCATTCGGATATCCACAATCAGTTATCCAAATCGTGACTGACATGTAC		GCCGAC	NICLEGIIRMHQMRE
	AAGGGAGCAACGATGAAAGTCAAAACGGCAGATCAAAAGACGCAAAG		TGAGCC	KGYDCVGHKVRCLAF
	CATCAAAATAGAAGCGGGGGTGAAACAAGGAGACCCCATTTCTCCAAC		AGAGGG	ADDLAILTNNKDEM

					CAAAGT	QEVIDKLDADCRSVS
					CGTAGG	LIFKPRKCASLTIVRGA
					CCGGTA	VDKYAKIRINGDAIRT
					GGCTCC	MADRDTYRYLGVKT
					CGGCGG	GVGGRASETEALIQV
					GTTCTCC	VKELQKVHETDLAPH
					GTCGTA	QKLDILKTFLLPRLQH
					GTCAGT	LYRNATPKLSELREFE
					GGTGTG	NVVMKSVKRYHNIPI
					CCTACAC	KGSPVEYVQIPVKKG
					CTAACT	GLGVLSPRLTCLITFLT
					GCTATG	STLCKLWSDDPFISSI
					ACAAGC	HKDALSRITVKAMGL
					GTATAG	TTQSATIKETCEYLNT
					GAGGCC	RKAVTKGGYSLFCRM
					CGGAAA	NESLRTLSVIQGAPLK
					AACAAG	SMEFIPVNNEIGIAV
					CCAAGG	QATKDSEIKVFTKADS
					CGGCGT	LKLMSKLKDLVRSAM
					TAGCTG	LKRFLEEKSVKSRVTQ
					AGAGCT	VLQHHPQSNRFVRD
					AACAAG	GRNCSIAAQRFVHPA
					CTTCTCG	RLNLLSCNANTYDVN
					TGGATG	HPKGCRRCQADFES
					GGTGCC	QQHILQNCHYSLAG
					AGAGGG	GITQRHDRVMNRIL
					CACCATC	QEIGNGRKAHYKIMV
					CTGGTG	DMETGATRERPDIIM
					GGTGGA	EERDGPEVLLADVTV
					TGGGGG	PYENGVQAVERAWD
					GAGCTT	KKIEKYKHFLDYYRKI
					GGGAAC	GKKATILPLVVGSLGT
					GTCCCG	YWPDTSHSLKMLGLS
					ATCCTTC	DGQIRNVIPEICQIAL
					GGATGC	ESSKNIYWKHILGDSY
					CCAGAC	KTVEGLFCQRNNKEV
					CACCGC	RFEGKGEKHHVSQRF
					AATCTGT	QPLKCEKVRTMKSTK

				CAAGGC	EEGRSRSNAKKGPN
				ACCGTG	WRRSKSESDGRSVSK
				CTCCAA	GRYWRDPSNKPPHS
				AAGCAC	KMTQSALAKR (SEQ
				ACGCGC	ID NO: 1418)
				GGGTTG	
				GTTTGG	
				ATGCCG	
				ACTGAG	
				CCAGAG	
				GGCAAA	
				GTCGTA	
				GGCCGG	
				TAGGCT	
				CCCGGC	
				GGGCTC	
				TCCGTCA	
				TAGTCA	
				GTGGTG	
				TGCCTTC	
				ACCCAA	
				CTGCTAT	
				GACATG	
				CGTACA	
				GGAGGC	
				CCGGAA	
				AAATAA	
				GCCAAG	
				GCGGCG	
				TTAGCAT	
				AGGGCT	
				AACAAG	
				CTTCTCG	
				TGGATG	
				GGTGCC	
				AGAGGG	
				CACCATC	
				CTGGTG	


GGTGGA
TGGGGG
GAGCTT
GGGAAC
GTCCCG
ATCGTTC
GGATGC
CCAACC
ACCGCA
ATCTGCC
AGGCAA
CGTGCT
TCGGAW
GGTCAT
TGGTTCT
AGACTT
GTAATA
GACCAT
TGGCCG
GAAGAG
CACACG
CGCGGT
TGGTTG
GATGCC
GACCGA
GCCTAG
AGGGTG
CAAACC
TGAAGG
GCGAGG
TCGAAG
GCCGTG
AGGCTC
CCGGCG
GGAAAC
TCCGTCA
TAGTTA
GTGGTG


TGCCTAC
ACCCGA
CGACTA
TGACAC
ATAGGA
GGAATC
CTGATCT
GATATG
ATCATGT
ATATAG
GGAGGG
CGAAGG
TAAATA
GTCAGK
GTCAAA
GTCCAC
GTGGCA
GCTACTC
CCCAGC
ATAGTA
GTGATG
CGAGTG
GAWCCA
ACTTTGA
CACTGA
TGTTCCC
TGAGCC
TGACCC
ATCTGC
ACAAAT
CCAACA
GTGTAT
GATGGC
CCACAC
ACTGAG
GACGAG
TATCACT
TGTGAT

							ACTCAG AGGTGT CCCCCAT GATCAA CCAATAT CACAGC TAGCGG ACCTACC GTGAGG TAGACC CCCGCC GCTGTA GCAGGC TCGCCTC (SEQ ID NO: 1296)	MTNVYLKPVNDNQT NKTGDNSRNTMSNS QCEMTWKPVARTYA QAASTNPADDKTVT VLGCKYNLLKLGNTP QTSKRSPPKPSRGGA RISSVYTLTDELEITHR EEGKITFAIDLPNKNN ILCPLCRECTQTRGRG SSFTKHMKLHVKEKH QLDATFIYKCSMCNE YEPEKKCGTKWIQTH LQKVHNYKYDESAIV VPVPPNTRQQIANEL NNAAPFVDIRKPKAA AVEEKKTENGALLKFL TKSNKDDQVKSPSXD IPDAESPEKETQALTI DPKGNNSPSKSSIRSS QSSASSVCQEIQEIITL SEDEDPKGARPKPGI
						CCAACTCT CATCGTAT TAACCTAC GGTATTCA CTCCTAGT GAGTGTAA TAAAGGTT AATTACGT TTTCTCTTG CMAGAGA AAAAGAA AATTCGAA TCCTTTTTG TGTAACTC ACAAACTG ACAGAGAC CTATCGAA TTTCCTTTG ATAGGAAT AGTCACTC	TGAATA CCGTCA GATAAG CCCCCA ACATAA AAATAA AAGTCG GCGTTA GCTAAC CACTAA ACCGGC TCCTCAT TGGGGG AGAGTA TCATTCC GGTGCT CTCCGTT TGGGCG GTAGGG AGGAGT TGGGTA	AAGGTTAATTACGTTTTCTCTTGCMAGAGAAAAAGAAAATTCGAATCCT TTTTGTGTAACTCACAAACTGACAGAGACCTATCGAATTTCCTTTGTTTC GTATATAGGAATAGTCACTCTGGACCACGAAGTGGACAGTTGTCGGCG GACTTCCAGAGTGGAGAGAAAAGGTGTGAAGAGAGGAGGTCTAGAAA CACTTCGGCTGTCTAGGACCAGTTCCTGAGTGGAAAGAGGAAGGTCTA GAAACACTTCGGCTGTCTAGGACCAGTTCGTGAGATCTCTCGTGGAGAG TTGAAAACAGTCAGCTGAGGCTACTGTATTTCTTGATAGCCCCGCCCCCA ATCCCCCTCCCCCCCCCCTCGACAGATTTTTCTGTTTGACCTCCTGGAATT TGCGAGGAGTGCGCGAGAATTTTCGAATTCTTCGCGCGTTTTCTCGAAA TTTTCCAGAAGATTCGAGCGGAGAATCTTCGAGAAAGTGAGCTGAATTT CGCGCGAATTTTCCGCGATTTTCAAATTATCGATTTTTTGTCGGAAAATTT ATTTTCTGGCAAAAATTTGATTGAGTTCACGCGGGAGAGAAGGAATTGTT GGAAAAGGGTATTGATTTTTGTGGCGGAGGAAACTCCCACTGAATCAAT AACTCTCAAAGGAGAACTCATCGAACAACCTCGGGTGACCTGAATCTTG GGCGAAATTTTCGCATTGACACAAGATAAMACAAATTACTGTKGAAAAT AAATCAGAACAAACTGTCAAAAAGAGAGACAAAAAGTATTGATTAACA ACATCATGACAAATGTATATCTTAAGCCTGTGAATGATAACCAGACTAAC AAAACCGGTGATAATTCTAGAAATACTATGTCAAATAGTCAATGTGAAA TGACGTGGAAACCTGTAGCCAGAACATATGCTCAGGCAGCCAGTACTAA CCCGGCCGACGACAAAACGGTGACTGTCCTTGGGTGCAAATACAATCTG	264
NeSL	NeSL-2_CRem	.	Caenorhabditis remanei					

		CTAAAACTGGGAAATACTCCTCAGACGTCGAAAAGGTCGCCTCCAAAAC	TGGACCAC	GCGACC		NVWSLINETGKDAYI
		CATCGAGAGGAGGAGCTCGAATCAGCAGTGTGTATACTCTGACTGATG	GAAGTGG	CGGAAG		DTDIMMAFLKMRVE
		AGCTGGAGATTACGCACAGAGAAGAAGGTAAGATCACATTCGCGATAG	ACAGTTGT	TATGGA		NCDSVNIIDPLNYQF
		ACCTTCCAAACAAGAATAACATCTTGTGCCCGCTGTGTCGGGAGTGCAC	CGGCGGA	TGCCCA		PARVDLVPLIQRNLE
		CCAAACCCGTGGGAGAGGGTCCAGTTTTACCAAGCATATGAAACTCCAC	CTTCCAGA	ACCACC		DGKKRVVFPICADEH
		GTGAAAGAGAAGCACCAACTTGATGCCACGTTCATCTACAAGTGTAGTA	GTGGAGA	GCAATC		WTLLTISNGIAAFYDP
		TGTGCAACGAGTACGAACCGGAAAAAAAATGCGGTACGAAGTGGATCC	GAAAAGG	TGATCT		TGSRMSSYIEELVNEL
		AGACCCACCTTCAAAAAGTGCACAACTACAAGTATGACGAGTCTGCAAT	TGTGAAGA	GGCATT		GLIIPKEQDEQPRQR
		AGTTGTCCCAGTACCACCCAACACAAGACAGCAAATAGCTAATGAGTTG	GAGGAGG	GTGTTTC		DSYNCGVFVMKMAE
		AACAATGCTGCCCCCATTCGTTGACATCAGAAAACCGAAAAGCTGCTGCTG	TCTAGAAA	GGATGG		AFIQDTEWEMEEVE
		TTGAGGAGAAGAAGACTGAAAATGGTGCTCTGTTAAAATTCCTGACCAA	CACTTCGG	TCTCTGT		EDVKNFRRNLLEELK
		GTCCAATAAGGACGATCAGGTAAAATCCCCATCGGAWGATATTCCAGA	CTGTCTAG	CTCTAG		PNYEIFAEKIKYYNSP
		TGCGGAAAGCCCTGAAAAAAGAAACTCAGGCGCTCACTATCGATCCGAA	GACCAGTT	ATCTGA		GKSFAQSRPTSRSSQ
		AGGGAACAACTCACCATCAAAAAGCTCAATAAGATCGAGCCAGTCCTCA	CCTGAGTG	AATAGA		CAVCPTCSRSATPM
		GCTTCCTCCGTTTGTCAAGAAATCCAGGAAATCATCACGTTGAGTGAGG	GAAAGAG	GCTCTG		MDVGNMEVDPVPQ
		ATGAAGACCCAAAAGGGGCTCGTCCAAAACCAGGAATCAACGTGTGGA	GAAGGTCT	GCCTGA		QQETPKSREPEQDEG
		GCTTGATAAATGAAACGGGAAAGGATGCATACATTGATACAGATATCAT	AGAAACAC	AGAACA		WKVVGKARKRGVVT
		GATGGCGTTCTTGAAGATGAGAGTGGAAAACTGTGACTCCGTGAACAT	TTCGGCTG	CACGCG		ERSPNISPEAKRQFTG
		AATTGATCCACTCAATTACCAGTTTCCCGCGAGAGTGGACCTAGTCCCAC	TCTAGGAC	CGGACC		PEIKVVSPGKFHPLVG
		TTATCCAGAGGAATCTGGAAGACGGAAAGAAAAGAGTCGTGTTTCCGA	CAGTTCGT	GGTTGG		ETEEMEVTCDSPPTK
		TCTGTGCAGACGAACACTGGACGCTCTTGACCATCTCGAATGGAATTGC	GAGATCTC	ATGCCG		EPTTEPKVTPSLPAM
		TGCATTCTATGATCCGACTGGATCGCGAATGAGTAGTTATATTGAAGAG	TCGTGGAG	ACTCGA		KIASPEVTKKQTSKKK
		TTGGTGAACGAACTTGGACTGATTATCCCAAAGGAACAGGATGAACAG	AGTTGAAA	TCTGGA		GKYGKKKQXTKKAQ
		CCAAGACAAAGAGACAGCTACAACTGTGGGGTATTTGTGATGAAAATG	ACAGTCAG	GGGTGC		PPKGEPTKKAQPKGE
		GCGGAAGCCTTCATCCAAGATACCGAATGGGAAATGGAGGAAGTAGAG	CTGAGGCT	AAACCT		PAKLIEQVRTWFDKQ
		GAAGACGTGAAAAACTTCCGAAGAAATCTCCTTGAAGAACTGAAACCCA	ACTGTATT	GAAAGG		MKSYQEQGSNIQTLT
		ACTACGAGATATTTGCTGAAAAAATCAAATATTATAACTCTCCGGGAAA	TCTTGATA	GAAAGT		WIADSLTAAIFKANS
		AAGTTTCGCCCAAAGTCGACCCACAAGTCGAAGCAGCCAGTGTGCCGTC	GCCCCGCC	TGAAGG		GNKYLVDKITARCPP
		TGTCCGACGTGCTCTCGTTCAGCTACACCGATGATGGATGTAGGAAACA	CCCAATCC	CCGTGA		PLLNEGEMATQTSRR
		TGGAAGTGGATCCCGTTCCACAGCAACAAGAGACACCGAAGAGTCGCG	CCCTCCCC	GGCTCC		TEAVKPKDRFVKESN
		AGCCAGAACAAGATGAAGGCTGGAAAGTGGTGGGAAAGGCTAGAAAG	CCCCCCTC	TGGCGG		EPLRIQYAKNRAKTF
		AGGGGAGTGGTAACTGAACGATCGCCAAACATCTCTCCAGAAGCCAAG	GACAGATT	GAAACT		NVIIGKHSARCEIDIN
		AGACAATTCACTGGTCCAGAGATCAAAGTCGTCTCACCTGGGAAGTTTC	TTTCTGTTT	CCGTCAT		VVENHFRQTLKAQP
		ACCCACTTGTGGGCGAAACTGAGGAGATGGAGGTGACGTGTGACAGCC	GACCTCCT	AGTCAG		VTEEALNTVCSGIKKA
		CACCAACGAAAGAGCCCACTACGGAACCGAAAGTGACTCCAAGCCTGC	GGAATTTG	TGGTGT		KVDPSIEGPISSGEVK
		CAGCAATGAAAATTGCTAGCCCAGAAGTGACGAAAAAGCAAACGTCAA	CGAGGAG	GCCAAC		AILAKIKDTSPGTDGV
		AGAAGAAGGGAAAGTATGGCAAAAAGAAACAGSAGACAAAGAAAGCT	TGCGCGAG	ACCCGA		KYSDLKWFDPEGERL
		CAGCCGCCGAAAGGGGAGCCAACAAGAAAGCTCAGCCAAAAGGAGA	AATTTTCG	CGACTA		ALLFDECRQHGKIPS

		ACCGGCAAAGCTCATTGAGCAAGTGAGAACTTGGTTTGATAAACAGATG	AATTCTTC	TGACAT		HWKEAETVLLPKDCT
		AAATCGTACCAAGAGCAAGGTTCTAACATCCAGACACTGACCTGGATTG	GCGCGTTT	AGTTGG		EEERKKPENWRPISL
		CCGACTCACTCACTGCCGCCATCTTCAAGGCAAACAGTGGAAACAAGTA	TCTCGAAA	AGGAAT		MATVYKLYSSVWNR
		TCTGGTAGATAAGATAACTGCAAGATGCCCACCACCATTGCTGAATGAA	TTTTCCAG	CCTGATC		RISSVKGVISDCQRGF
		GGTGAGATGGCGACGCAGACGAGCAGAAGGACAGAAGCGGTGAAACC	AAGATTCG	TGATAA		QAIDGCNESIGILRM
		AAAAGATCGATTTGTAAAAGAATCTAACGAGCCGCTCAGAATCCAGTAT	AGCGGAG	TAATCAT		CIDTATVLNRNLSCS
		GCAAAGAACCGAGCAAAGACCTTCAATGTGATAATTGGGAAACACTCC	AATCTTCG	TGTTCAT		WLDLTNAFGSVPHEL
		GCACGATGTGAAATTGATATTAACGTCGTGGAAAACCACTTCAGGCAAA	AGAAAGT	ATAAGG		IRRSLAAFGYPESVIN
		CCCTGAAAGCACAACCAGTAACAGAAGAAGCATTGAATACTGTGTGCA	GAGCTGAA	GAGGGG		ISDMYNGSSMRVKT
		GTGGAATCAAAAAGGCGAAAGTTGATCCAAGCATTGAAGGTCCGATCT	TTTCGCGC	GATGGT		AEQKTQNIMIEAGVK
		CGTCAGGAGAAGTGAAAGCGATTCTTGCAAAGATCAAAGATACCTCTCC	GAATTTTC	AAATAC		QGDPISPTLFNICLEGI
		CGGAACTGATGGAGTGAAGTACAGTGATCTGAAATGGTTCGATCCGGA	CGCGATTT	CCAGGG		IRRHQTRKTGYNCVG
		AGGTGAACGTTTGGCGTTGTTCGATGAATGTCGACAGCACGGGAA	TCAAATTA	TCCGAA		NDVRCLAFADDLAIL
		GATTCCGAGCCACTGGAAAGAAGCAGAAACTGTTCTGCTACCAAAAGAT	TCGATTTT	ACCATC		TNNQDEMQDVLNQ
		TGCACTGAAGAGGAAAGAAAGAAGCCAGAGAATTGGAGACCCATCTCT	TGTCGGAA	AAAGCA		LDKDCRSVALIFKPKK
		CTAATGGCTACTGTATACAAACTCTACTCCTCAGTCTGGAACAGGAGAA	AATTTATTT	GCTACT		CASLTIKKGSVDQYA
		TCTCCTCAGTKAAAGGAGTCATCAGTGATTGCCAAAGAGGCTTCCAGGC	TCTGGCAA	GACCAG		RIKIHGMPIRTMSDG
		GATCGATGGATGCAATGAGTCAATCGGAATTCTGCGGATGTGCATAGA	AATTTGAT	CATAGT		DTYKYLGVQTGNGG
		CACAGCCACAGTTCTCAACCGAAACCTGTCGTGTTCATGGTTAGACTTGA	TGAGTTCA	AGTGAT		RASESESLTQIAAELQ
		CGAACGCTTTTGGAAGCGTGCCCCACGAGTTGATCAGAAGATCACTAGC	CGCGGGA	GAACAC		MVHDTDLAPNQKLD
		CGCATTCGGGTATCCTGAATCAGTCATCAATATATAATCAGTGACATGTATA	GAGAAGG	ATAGAC		VLKAFILPRLQHMYR
		ATGGATCGTCAATGAGAGTCAAGACAGCGGAGCAGAAAACTCAGAACA	AATTGTTG	CCTGGG		NATPKLTELKEFENTV
		TCATGATTGAAGCTGGAGTTAAGCAAGGTGATCCCATCTCGCCAACTCT	GAAAAGG	GTTCCCT		MKSVKMYHNIPIKGS
		ATTCAACATCTGTCTTGAAGGCATAATCCGAAGGCATCAGACGAGGAAG	GTATTGAT	GAACTC		PLEYVQIPVKNGGLG
		ACAGGTTACAACTGCGTTGGAAACGACGTACGTTGCCTGGCATTTGCTG	TTTTGTGG	GACCCA		VMSPRFTCLITFLAST
		ACGATCTTGCTATCCTTACCAACAACCAGGATGAGATGCAAGATGTGCT	CGGAGGA	TCTGCAC		LFKLWSDDEYISSIHK
		CAATCAGCTGGACAAGGACTGTCGTAGTGTTGCCCTGATATTTAAGCCA	AACTCCCA	AAACCC		KALSRITAKVMGLKT
		AAGAAGTGTGCTTCACTGACGATCAAAAAAGGAAGTGTTGATCAGTATG	CTGAATCA	ACTTTGT		QKATLQEQCEYLNTK
		CAAGAATCAAGATTCATGGAATGCCCATTCGGACTATGTCGGATGGGG	ATAACTCT	ACAAAT		KAITKGGYSLFSRMN
		ATACCTACAAGTATCTCGGAGTCCAAACCGGAAACGGTGGTAGAGCCTC	CAAAGGA	GAACCA		EAIRTLSVNLGAPLKS
		GGAATCAGAATCCCTGACTCAGATTGCCGCGGAACTCCAAATGGTCCAT	GAACTCAT	AACTGA		MQFIPENGEIALEVQ
		GACACAGACCTGGCGCCGAACCAGAAACTTGATGTGCTGAAGGCATTC	CGAACAAC	TGAAGA		ASENSQIKVFSKADS
		ATCCTGCCGAGACTGCAACATATGTACAGAAACGCCACTCCAAAGCTGA	CTCGGGTG	GTTTAAT		MKLVTKLKDLVKSA
		CGGAGTTAAAGGAGTTTGAGAACACAGTCATGAAAAGTGTGAAGATGT	ACCTGAAT	GATTTCT		MLKNFLENKKVKSKV
		ATCACAACATCCCGATTAAAGGATCACCACTCGAATATGTCCAAATTCCA	CTTGGGCG	TACATCA		VQVLQHHPQSNKFV
		GTAAAGAATGGAGGACTCGGAGTTATGTCTCCCCGATTCACGTGTCTCA	AAATTTTC	CAGCTA		NDGKNXSISSQKFVH
		TAACGTTCCTGGCGTCCACACTGTTCAAACTGTGGTCAGACGACGAATA	GCATTGAC	GCGGAC		PARLSQLVCNGNSYS
		CATCTCGTCCATCCACACAAAAAGGCGTTGAGTAGAATCACGGCAAAGGTG	ACAAGATA	CTACCGT		KDLPKNCRWCGYEC

		ATGGGACTGAAGACCCAAAAAGCCACGCTCCAAGAGCAGTGCGAGTAC	AMACAAA	GAGGTA	ESQAHILQHCTYSLSS
		CTGAACACCAAGAAAGCAATCACGAAAGGAGGTTACAGCCTCTTCTCGC	TTACTGTK	GACTCC	GITQRHDRVLNRILXE
		GAATGAACGAAGCTATTCGAACGCTCAGTGTCAACCTTGGAGCACCGCT	GAAAATAA	CGCCGC	VIKGRKNNDYYDIMV
		CAAATCAATGCAATTCATTCCGGAAAAATGGCGAAATTGCTTTAGAAGTG	ATCAGAAC	TGTAGC	DTEPGPTRERPDIIMI
		CAAGCATCAGAAAACTCACAGATCAAAGTATTCTCGAAAGCTGACAGTA	AAACTGTC	AGGCTC	QKDGPEVLLADVTVP
		TGAAACTGGTGACAAAGCTGAAAGATCTGGTGAAATCGGCGATGCTCA	AAAAAGA	GCCATT	YENGVVAIEAAWDW
		AGAACTTCTTGGAAAACAAGAAGGTCAAAAGCAAGGTTGTGCAGGTGC	GAGACAA	G (SEQ	KMEKYSHFIDYFARL
		TTCAACACCACCCACAATCAAACAAATTCGTCAATGATGGAAAGAACWK	AAAGTATT	ID NO:	GKRAVILPLVVGSLGT
		CAGCATWTCCTCCCAAAAGTTCGTACACCCAGCACGGCTGAGCCAGCTG	GATTAACA	1297)	YWPDTSNSLRMLGL
		GTCTGCAACGGGAACAGCTACAGTAAAGACCTTCCGAAAAACTGCAGAT	ACATC		SDGQIRNLIPDISMIA
		GGTGCGGCTACGAATGCGAGTCTCAGGCTCACATCCTCCAGCATTGCAC	(SEQ ID		LESSKQIYWRHIFGDS
		ATACAGCCTTTCATCTGGAATCACCCAGAGGCATGACCGTGTCCTGAAC	NO: 1174)		YRIVSDLYCRKDQQEI
		AGGATCTTGCASGAGGTGATAAAAGGCAGAAAAAACAACGACTACTAT			RFGDEPMENVQVSD
		GACATAATGGTGGATACGGAGCCCGGACCAACCAGAGAGCGTCCAGAT			RFQPFKTREREKKSEE
		ATCATCATGATACAGAAAGATGGTCCGGAAGTCCTACTGGCGGATGTTA			EKKRRSKSKKGKTWR
		CGGTACCATACGAGAATGGAGTTGTTGCGATCGAAGCCGCGTGGGATT			GSKKQTDSRQSGKSN
		GGAAGATGGAGAAGTACAGTCACTTTATTGATTACTTCGCAAGACTGGG			QNQGFQRSVGQGVS
		AAAGAGAGCAGTAATCCTTCCACTAGTGGTTGGAAGTCTTGGGACCTAC			R (SEQ ID NO: 1419)
		TGGCCAGACACCTCAAACTCTCTGAGAATGCTGGGTCTCTCTGATGGGC			
		AAATCAGAAACCTGATCCCAGACATCTCCATGATTGCTCTAGAGTCTTCC			
		AAACAAATCTACTGGAGGCATATCTTCGGAGATAGCTACAGAATTGTGA			
		GTGATCTATACTGCAGAAAAGACCAGCAGGAGATCAGATTCGGAGATG			
		AACCCATGGAAAATGTTCAAGTCTCAGATCGATTCCAGCCTTTTAAAACA			
		AGAGAGCGTGAGAAGAAATCCGAGGAAGAAAAAGAGAAGATCAAA			
		GTCCAAAAAAGGCAAAACTTGGCGAGGATCCAAAAAACAAACTGATTC			
		CCGGCAATCCGGCAAAAGCAATCAGAATCAGGGCTTCCAAAGAAGCGT			
		TGGACAAGGCGTATCACGGTGAATACCGTCAGATAAGCCCCCAACATAA			
		AAATAAAAGTCGGCGTTAGCTAACCACTAAACCGGCTCCTCATTGGGGG			
		AGAGTATCATTCCGGTGCTCTCCGTTTGGGCGGTAGGGAGGAGTTGGG			
		TAGCGACCCGGAAGTATGGATGCCCAACCACCGCAATCTGATCTGGCAT			
		TGTGTTTCGGATGGTCTCTGTCTCTAGATCTGAAATAGAGCTCTGGCCTG			
		AAGAACACACGCGCGGACCGGTTGGATGCCGACTCGATCTGGAGGGTG			
		CAAACCTGAAAGGGAAAGTTGAAGGCCGTGAGGCTCCTGGCGGGAAAC			
		TCCGTCATAGTCAGTGGTGTGCCAACACCCGACGACTATGACATAGTTG			
		GAGGAATCCTGATCTGATAATAATCATTGTTCATATAAGGGAGGGGGAT			
		GGTAAATACCCAGGGTCCGAAACCATCAAAGCAGCTACTGACCAGCATA			
		GTAGTGATGAACACATAGACCCTGGGGTTCCCTGAACTCGACCCATCTG			
		CACAAACCCACTTTGTACAAATGAACCAAACTGATGAAGAGTTTAATGA			

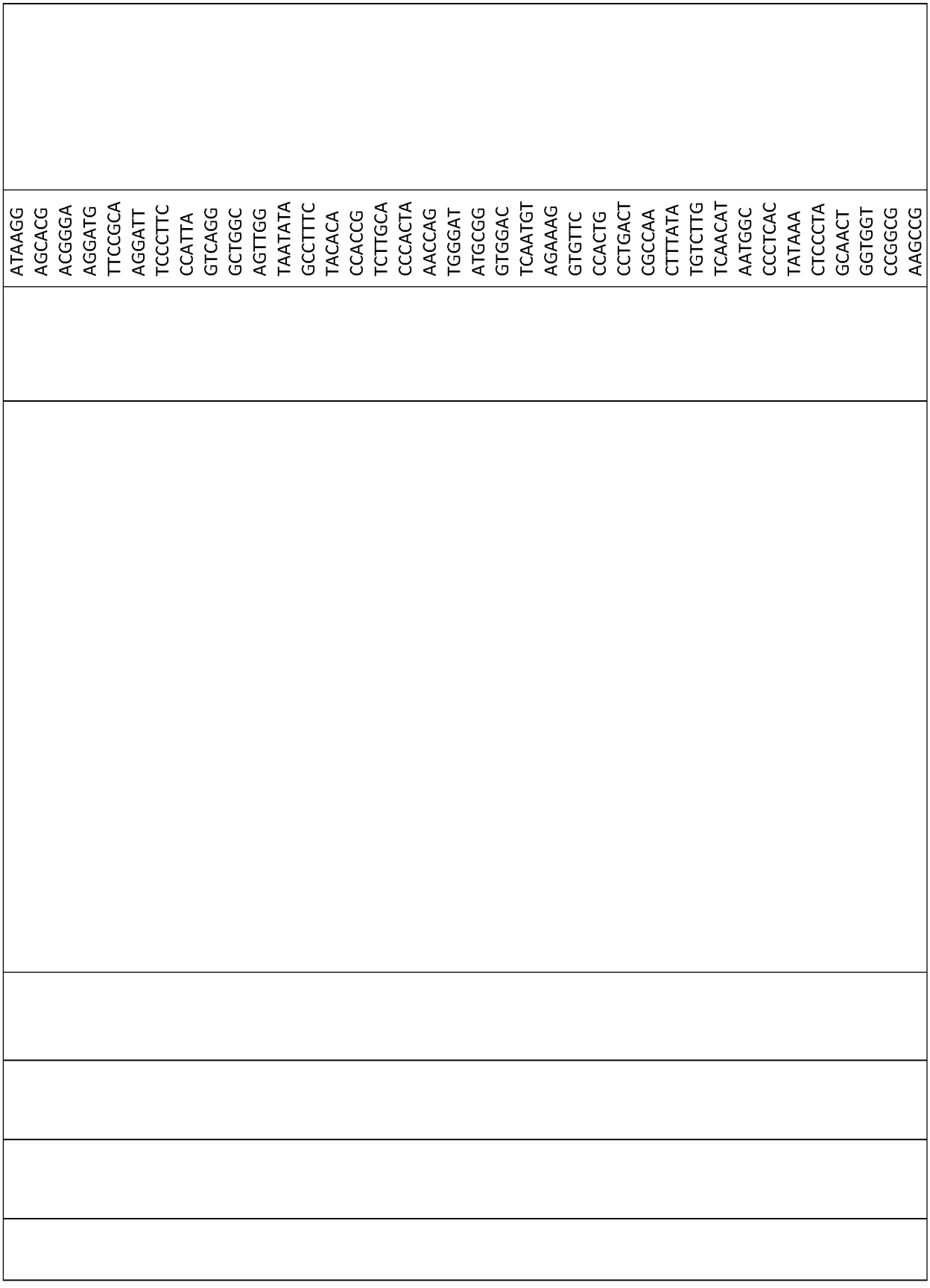
				TTTCTTACATCACAGCTAGCGGACCTACCGTGAGGTAGACTCCCGCCGCT GTAGCAGGCTCGCCATTG (SEQ ID NO: 1051)			
NeSL	NeSL-4_CRem	chrUn	Caenorhabditis remanei	CCCTTTTCTATCGTATTAACTACGATAACCGCTCATTTGAGTGTAAAAAA	CCCTTTTCT	TAACAT	MTKTEWSWRHRSRS
				GGTTCCCCCCTCCTCGCCTGCCTTACCCACGCATCTCTGCCTCTGGGAAG	ATCGTATT	GCCTTG	RSVGIVVKIDTSDYAN
				GCGGAGGGTCAACTTGCGGGTCTGTGGATTTCCTTTCCTATCCACCGCCC	AACTACGA	GAAGGC	VRVHVAADLSNEDG
				ATATTCTCTGTCGAAAGCCTACCTAGATCAGCCGGGAGTTTTTCCTATCC	TAACCGCT	ACCACG	HTSHNNGIILPIPMKP
				CATTCAGGCGATCGCTCAAGGCTGTTTTATCGACACTCCTTCTTGACAAG	CATTTGAG	CCAAAA	SVDRFCQIQYPPRGY
				TATTTATTTCTTGACAAATTCTATTTTTCCTTTTATCGATTTTTCTCTTATTTA	TGTAAAAA	GTCCTG	YVPHPQSQKGHDAK
				TCGATTCTTGTGAAAATATGACCAAGACCGAATGGTCCTGGCGTCATCG	AGGTTCCC	GCAACT	PSRHWNEEAQPPYY
				ATCTCGCTCCCGCTCTGTTGGAATCGTTGTGAAAATCGATACAAGCGACT	CCCTCCTC	GATTTG	HNNNHGRRGRSAKP
				ATGCTAACGTCCGAGTGCATGTCGCGGCGGACCTTTCCAATGAGGATGG	GCCTGCCT	AATAAT	SGRRPPRKPILQEESL
				CCACACGAGCCACACAACGGCATCATTCTCCCCATCCCAATGAAGCCC	TACCCACG	GTATAA	AAHPQIPGDTASAVP
				AGCGTCGATCGATTCTGTCAAATTCAATACCCTCCAAGAGGGTACTATGT	CATCTCTG	AAGTAA	LYSDVVNNENKSQG
				TCCGCATCCTCAAGTCAGAAAGGCCATGATGCAAAGCCCTCGCGTCAT	CCTCTGGG	CTGGAA	KPPQGSHRRSGRPGT
				TGGAATGAAGAGGCACAACCTCCCTACTACCACACAACAACAATCATGGGA	AAGGCGG	CCAAAT	KPSVPVGEAEQETNS
				GAAGGGGGCGTTCGGCAAAACCAAGTGGACGCCGACCCCCACGAAAGC	AGGGTCAA	GCCCGA	RPIAPEPIVKFKHDKH
				CCATACTTCAGGAAGAGTCCCTGGCAGCGCACCCCCAAATACCCGGGGA	CTTGCGGG	TAGGTA	GWTTVQGSHSSGRP
				TACTGCGTCAGCGGTCCCACTGTACTCCGACGTCGTCAACAATGAAAAC	TCTGTGGA	GGGCGG	VPKPSVPVVSEANRF
				AAGAGTCAGGGGAAACCACCGCAAGGGTCCCACAGAAGAAGTGGAAG	TTTCCTTTC	GAGAAA	QLLQEGDFPPLTTSES
				ACCAGGAACAAAGCCCTCTGTTCCGGTTGGTGAGGCAGAGCAAGAAAC	CTATCCAC	ATGACC	SQEEIKVPNYQRIVSP
				GAATTCCCGTCCAATTGCTCCAGAACCCATCGTGAAATTCAAACACGATA	CGCCCATA	TAGAAA	IPLPSEEDSKLPTKSNY
				AACACGGGTGGACTACTGTCCAAGGGTCCCACAGTAGTGGAAGGCCGG	TTCTCTGT	ACACAA	RAPKGRKSRNYKKPQ
				TACCAAAGCCCTCGGTACCGGTGGTTTCAGAGGCAAATCGGTTCCAGTT	CGAAAGCC	AGTCCC	QQNPKKYQQRLPYQ
				ACTCCAGGAAGGGGATTTTCCACCCCTTACAACATCCGAATCTTCGCAA	TACCTAGA	AAGCCC	PKVNNAPTDRMAPE
				GAAGAGATTAAAGTACCGAACTACCAACGAATAGTGTCACCGATTCCTC	TCAGCCGG	CCGGAT	QLKGGGGKTAHNDI
				TCCCCTCTGAAGAGGATAGTAAGTTGCCGACTAAATCAAATTACAGAGC	GAGTTTTT	TCGAAA	EEMEIEEDTDEKIIQV
				GCCCAAGGGACGAAAGAGTCGCAACTACAAGAAGCCACACAACAAAAA	CCTATCCC	GACCTA	KRIKIVNKLTPHHFVC
				TCCGAAGAAATATCAGCAGAGGTTACCCTATCAACCCAAGGTCAACAAT	ATTCAGGC	TAGGAA	MMTYPTDNIYRCFV
				GCTCCGACGGATCGCATGGCCCCAGAACAACTCAAAGGAGGAGGAGGA	GATCGCTC	GTCAGT	KGCTATSQGGWGAE
				AAAACCGCCCACAATGACATTGAAGAGATGGAAATTGAGGAAGACACT	AAGGCTGT	GAATAG	DLKYLTVHIRQEHKIK
				GACGAGAAGATTATCCAAGTGAAACGAATCAAAATCGTCAATAAGCTAA	TTTATCGA	AGAGAA	VEWTYECGICGDLSG
				CTCCGCATCACTTTGTTTGCATGATGACGTATCCAACCGACAACATCTAT	CACTCCTT	ATATCA	GAGKHISKWIKPHM
				AGATGCTTCGTCAAAGGTTGCACAGCAACATCACAAGGTGGTTGGGGA	CTTGACAA	AACAAA	RKKHNRDAPTNFKM
				GCAGAGGACCTTAAGTACCTGACTGTCCATATCAGACACAAGAACACAAAA	GTATTTAT	TCTCACC	GSRSSGKPKITELLEES
				TTAAGGTCGAATGGACGTACGAATGCGGGATATGCGGTGACCTATCGG	TTCTTGAC	CATTCAC	APSCSNPRRKTLNQK
				GAGGTGCTGGCAAACATATCAGTAAATGGATCAAACCCCATATGAGGA	AAATTCTA	AAGGAC	KTAIITQVTPEKLKTG
				AGAAACACAATAGAGATGCCCCAACCAATTTCAAGATGGGTTCAAGAA	TTTTTCCTT	TTACTG	YQTRSVTKALSVLKES
				GTTCAGGTAAACCCAAGATTACTGAACTACTGGAGGAGAGCGCCCCGTC	TTATCGAT	GTCGAG	RQKELEVLREEEKAN

		TTGCTCGAATCCAAGAAGGAAAACCCTCAACCAGAAGAAGACTGCTATA	TTTCTCTTA	TAGAAA	AKQKSKLHPFFTKAP
		ATCACGCAAGTCACTCCGGAGAAATTGAAAACGGGCTATCAAACGAGA	TTTATCGA	ACAAGC	HIDGVKPTVRRELSK
		AGTGTCACGAAGGCTCTCAGCGTCCTGAAAGAGTCACGACAAAAAAGAG	TTCTTGTG	CAAAAC	MITPGGEHKGTKIPM
		CTGGAAGTGTTGAGAGAAGAAGAAAAGGCTAACGCTAAACAAAAGTCT	AAAAT	ATCAAG	VHTKRGLIQKINRKAK
		AAACTTCATCCTTTCTTCACCAAAGCCCCCTCATATAGATGGTGTGAAACC	(SEQ ID	CACGAC	KAKPMHLDESTIIEAS
		AACAGTACGGAGAGAAGAACTATCAAAAATGATTACTCCCGGAGGAGAACA	NO: 1175)	GCAAAA	QLDVITIDDDDEDDN
		TAAGGGAACAAAGATACCAATGGTCCACACCAAGAGCGCGGTCTCATCCAA		AGGGGT	MTPMRRRFNTWCL
		AAGATAAACAGAAAAAGCTAAAAAGGCTAAACCAATGCATCTTGACGAA		AACTTTG	DHETTQEAWLTDDVI
		AGTACCATCATAGAAGCGTCACAGCTCGACGTCATCACTATTGACGACG		GGCAAC	NWYLKDLCFGNEQY
		ACGACGAAGACGACAACATGACACCAATGCGAAGAAGATTCAACACTT		TAATTAA	MLVDPLVWLIYKMIG
		GGTGTCTTGACCACGAGACGACTCAAGAAGCATGGTTAACTGACGACG		CGGATA	GMAGVEQRFKSKKT
		TAATCAATTGGTACTTGAAAGACCTATGCTTTGGTAACGAACAATACAT		CCTCCGT	CLFPICEADHWILLVF
		GCTCGTAGACCCACTAGTATGGCTGATATACAAGATGGGAGGAATGGC		GTATCA	DETNLCYANSLGSQP
		AGGCGTCGAACAAAGGTTCAAAGCAAGAAGACGTGCCTATTCCCAATC		GGCAAA	NGQVKNFIQQLNRKL
		TGCGAAGCTGACCACTGGATTCTTCTTGTATTCGATGAGACCAACTTGTG		GCCGCC	CSFEKEVPLQKDSVN
		CTACGCGAATAGTCTTGGATCCCAACCAAACGGACAAGTTAAGAACTTC		ACCAAC	CGVHVCLIAKSIVNG
		ATTCAACAACTCAACCGAAAGCTCTGCAGCTTTGAGAAAGAAGTTCCAC		AGCAAA	QFWYDDSDVRTFRT
		TTCAGAAAGATAGTGTAAACTGCGGAGTACATGTCTGCCTGATAGCAAA		TTACTGC	NAKAALKAQGYELFS
		GTCAATAGTCAATGGACAATTTTGGTACGATGATTCAGACGTTCGAACG		CCGATA	EAPKQIENPDSSHRE
		TTTAGAACCAACGCCAAGGCGGCTCTGAAAGCCCAGGGCTACGAGCTCT		GGTAGG	DIKENSMEMCSESL
		TCTCGGAAGCACCAAAACAAATCGAAAACCCAGACTCCAGCCACAGAG		GCGTGA	MIVATPQRSEAPMEL
		AAGACATCAAGGAGAACAGTATGGAAATGTGTTCGGAATCTTTGATGAT		GAAAAT	VDTEPSDLESPKSDR
		CGTTGCGACTCCACAGAGGAGTGAAGCACCTATGGAACTAGTCGACACT		GACCTA	VVYEDCITALSDVSEP
		GAGCCTAGTGATCTGGAATCGCCAAAGTCAGACAGAGTAGTCTACGAA		CAACCTC	RMTPEKSETPEVPVV
		GACTGCATCACAGCTCTATCTGATGTTTCGGAGCCAAGAATGACTCCAG		CAAGAC	EERDLDWPKLESPKS
		AAAAGAGCGAAACTCCAGAGGTGCCAGTGGTGGAAGAAAGAGATCTG		CCGAGC	DRVVYEDCITDLSDVS
		GATTGGCCAAAACTGGAATCGCCAAAGTCAGACAGAGTAGTCTATGAA		CCACGG	EQRMTPEKCETPEAP
		GACTGCATCACAGATCTGTCTGATGTTTCGGAGCAAAGAAATGACTCCAG		AATCGA	LVVECVELERLPKDLP
		AAAAGTGCGAAACCCCAGAAGCGCCATTGGTTGTAGAATGTGTTGAGTT		AAGACC	VTDRSTVVAIPEAVKL
		GGAAAGGCTACCTAAGGATCTGCCAGTCACAGACAGGTCAACTGTCGT		TATAGG	EEKSEVVIPRLMELSY
		GGCAATCCCTGAAGCAGTAAAACTGGAGGAAAAGTCAGAAGTGGTAAT		AAGTCA	TVPPEPSPVVEYTQP
		TCCACGGCTCATGGAGTTATCATACACCGTCCCTCCAGAACCCTCTCCAG		GTGAAT	YTHTHTKPKVKATCQ
		TGGTTGAATACACCCAACCATACACTCACACTCACACTAAACCAAAGGTC		TGATGG	MGKKRKVPTGKPDE
		AAAGCTACATGCCAGATGGGAAAGAAAAGGAAGGTACCAACTGGGAA		AAATAC	LIQIVRQWFEKEFND
		ACCAGACGAACTGATTCAGATTGTGAGACAATGGTTTGAGAAAGAATTC		AAAACC	YVTEGRNFQRLEWLT
		AACGATTATGTTACGGAAGGACGAAACTTTCAACGACTGGAATGGCTTA		AAATTTC	NLLTAAIQKASAGDE
		CGAACTTACTCACCGCCGCAATACAGAAGGCATCAGCTGGTGATGAGG		TTCCATT	ETIEKIRKRCPPPEVRE
		AAACAATCGAAAAGATTCGAAAGAGATGCCCACCTCCAGAAGTTAGAG		CACAAG	NEMSTQTSQRQKPT

			AAAACGAAATGTCCACTCAGACATCTCAACGTCAAAAGCCTACCACAAC		GACTTA	TTNQKKRSRNTTQSD
			GAATCAGAAGAAACGCTCTAGAAACACTACTCAATCGGATACACAAGCC		CTGGTC	TQANTYWRNRAKTY
			AACACATACTGGCGAAATCGAGCCAAGACATATAATCAAATCATAGGTC		GAGTAG	NQIIGQDFKQCDIPIA
			AAGATTTCAAACAGTGTGACATACCGATCGCGATACTAGAAGAAATTCTA		AGCACA	ILEEFYKKTTSVTNVP
			TAAAAAGACTACCTCAGTGACCAATGTCCCTCAGGAAACCCTTGTGAAA		AGCCAA	QETLVKVTSRLPRLDI
			GTCACCTCAAGACTACCAAGGTTAGACATTGGAAAGTGGATCGAGGATC		AATATC	GKWIEDPFTEQEVFG
			CGTTCACGGAACAAGAGGTATTTGGTGCCCTCAAAAAGACAAAAGACA		AAGTAT	ALKKTKDTAPGTDGL
			CTGCGCCAGGAACAGATGGGCTCAGATACTATCACCTCCAATGGTTTGA		GACGCA	RYYHLQWFDPDCKM
			TCCCGACTGTAAAATGTTGAGTAGCATTTACAATGAATGCCAGCACCAT		AAAATG	LSSIYNECQHHLKIPA
			CTGAAAATTCCTGCCCAATGGAAAGAAGCTGAAACAATTCTCCTCTTCAA		GGTAAC	QWKEAETILLFKSGD
			AAGTGGCGACGAATCCAAACCAGACAACTGGCGGCCTATAAGTCTCATG		CTTGGG	ESKPDNWRPISLMPT
			CCCACCATCTACAAGCTATACTCAAGTCTCTGGAATAGGAGAATACGGA		CATCCA	IYKLYSSLWNRRIRTV
			CGGTGAAGGGGGATTATGAGCAAGTGCCAACGAGGATTCCAAGAGAGA		ATCAAC	KGIMSKCQRGFQERE
			GAAGGTTGCAATGAGAGTATCGGAATACTGCGGAGTGCTATTGATGTG		GGATAC	GCNESIGILRSAIDVA
			GCTAAAGGGAAAAGATCCCACCTGTCCGTTGCATGGCTGGACCTCACCA		CTCTGC	KGKRSHLSVAWLDLT
			ATGCCTTCGGTTCAGTACCTCACGAGCTGATTGAAAGCACGTTAAGTGC		GTATCA	NAFGSVPHELIESTLS
			ATACGGCTTTCCGGAGATGGTTGTACACATTGTCAAGGACATGTATAAA		GGCAAA	AYGFPEMVVHIVKD
			GACGCTTCCATAAGAGTCAAGAATAGAACGGAGAAAAGTGAGCAGATT		GTCGCC	MYKDASIRVKNRTEK
			ATGATAAAATCTGGGGTAAAACAAGGCGACCCTATCTCACCAACACTAT		ACCAAA	SEQIMIKSGVKQGDP
			TCAACATGTGCCTCGAAACGGTGATTAGACGACATCTGAAAGAATCATC		CTGTACT	ISPTLFNMCLETVIRR
			AGGTCACAAATGCATTGACACCAGAATCAAGCTTCTTGCATTTGCAGAT		ACTCCG	HLKESSGHKCIDTRIK
			GATATGGCCGTTCTAGCAGAATCAAAGAGCAGCTACAAAAGGAGCTT		AAAAAA	LLAFADDMAVLAESK
			ACAGAAATGGATGAAGACTGTACACCTCTCAACCTAATTTTCAAGCCGG		CCAAGA	EQLQKELTEMDEDCT
			CGAAGTGTGCAAGTCTCATCATAGAGTTCGGGAAAGTGAGGACCCATG		AACATG	PLNLIFKPAKCASLIIEF
			AGCAGATCATGTTGAAGCGAGAGCCGATCCGAAACCTCAATGATGACG		ATTTTCC	GKVRTHEQIMLKREP
			GAACATACAAGTATCTGGGAGTGCATACGGGAGCAGATGCAAGGACAT		CACTCC	IRNLNDDGTYKYLGV
			CAGAAGAGGAGCTGATCATTTCTGTAACAAAAGAGGTAGACCTTGTCAA		GTTAAA	HTGADARTSEEELIIS
			TCGCTCGGCGCTTACGCCACCCCAGAAACTGGACTGTCTTAAGACGTTC		GCATCTC	VTKEVDLVNRSALTP
			ACACTCCCAAAGATGACCTACATGTATGCCAACGCCATACCAAAACTTAC		AACCAA	PQKLDCLKTFTLPKM
			CGAACTTTCAGCGTTCGCTAACATGGTCATGCGAGGAGTCAAGATAATC		GCTAAA	TYMYANAIPKLTELSA
			CACTATATCCCAGTTAGAGGATCTCCTCTTGAATATATTCAAATTCCGAC		GCGGTA	FANMVMRGVKIIHYI
			CGGCAAAGGAGGACTTGGAGTTCCATGCCCTAGAATCACGGCATTGATT		AGGTTA	PVRGSPLEYIQIPTGK
			ACCTTCCTTGTCTCAACCATGAAGAAACTGTGGTCTGATGATGAATACAT		TCATGTC	GGLGVPCPRITALITF
			TCGTAAGCTCTACAACTCTTATCTGAAGAAGGTTGTGGAGGCGGAAACG		AAAAGG	LVSTMKKLWSDDEYI
			GGAATAGTGGAGGTCTCCACAAAGGATCTAGCAGAGTACCTCAGCAAC		TGTAGC	RKLYNSYLKKVVEAET
			AAGGTACCATCCAGAAAGCACGAATTCGGGTATAACTGCTACTCGAGGA		TACAGC	GIVEVSTKDLAEYLSN
			TTCGCGAAGTTTGTAATGGGCTAGCTCTCAACCAAGCTGCCCCTCTCTAC		AACCTA	KVPSRKHEFGYNCYS
			AAACTTGAATTCATCGAACAAGACAATGAGTTAGCAGTTGTTGTCCAGC		AAGCCC	RIREVCNGLALNQAA

		CGACTGAGGAGAGCAAGGAAAGGATTTTCACTAAAGATCATGTGAAAA	GAAAGG	PLYKLEFIEQDNELAV	
		AGCTCCAGTCGCTACTGAAAGCCAGCGTGAATGACGCACTGCTACACAG	TAGGGC	VVQPTEESKERIFTKD	
		ATTCTTGACAACAAAACCCGTCAAAAGTGAAGTGGTACAAGTTCTCCAG	CGTATA	HVKKLQSLLKASVND	
		CAGCACCCTCAAAGCAACAGCTTCGTCCGAATGGGAGGTAAAGTAAGT	AAAAGA	ALLHRFLTTKPVKSEV	
		ATATCGGTACATGTATGGATCCACAGGTCACGGTTAAACCAACTAACGT	CCTACAC	VQVLQQHPQSNSFV	
		GCAATTATAACATCTTTGATCCAAAGCAACCGAAAAACTGCCGGAGGTG	CCTCCAA	RMGGKVSISVHVWI	
		TGGTTATAAGAACGAGACTCAATGGCACATCCTGCAAGACTGCACATAT	GACCTA	HRSRLNQLTCNYNIF	
		GGCTGGGCTAAACTTATACGAGAAAGACACGATGCCGTACATCACAAG	AACCCA	DPKQPKNCRRCGYK	
		GTAGTCACAATGATTTGCGCTGGGGCAAAGAAGAACTGGGGCCGGAAA	CGAACT	NETQWHILQDCTYG	
		ATCGACCAAGAACTGCCCGGTTTCACTTCACTCCGTCCAGACATTTGTCT	CGAACG	WAKLIRERHDAVHH	
		GACGAGTCCGGATGGCAAAGAGGTTATCTTTGCGGATGTTTGTGTCCCT	ACCTAC	KVVTMICAGAKKNW	
		TACTCAAGGACAAGGAACATCGAATTCGCGTGGAAAGAGAAAATCCGA	AGGAAG	GRKIDQELPGFTSLRP	
		AAGTATACAGAAGGATACAGTCATCTTGTTGCACAAGGAATCAAAGTGA	TCCGTG	DICLTSPDGKEVIFAD	
		CAGTCCTTCCGATAGCCATAGGATCACTCGGAACTTGGTGGACGCCAAC	AATGGA	VCVPYSRTRNIEFAW	
		CAACGAAAGTCTCTATCAACTGGGTATCAGCAAGAGCGATATTCGCAGT	GAGAAA	KEKIRKYTEGYSHLVA	
		GCCATTCCATTACTATGCTCTACTGTGATGGAGTATAGTAAGAACGCCTA	TATCTCA	QGIKVTVLPIAIGSLG	
		CTGGAATCACATATACGGAAACTCATATACCTCGGTCCCACTGAGATAC	CCAAAT	TWWTPTNESLYQLGI	
		GGACACCAGAAGCCCGATGGAGACGATTGGAAGAAAGAACTGAGTTGC	CTCTTCC	SKSDIRSAIPLLCSTV	
		GAACCAGTTCTAGCTCTCCAACAATAACATGCCTTGGAAGGCACCACGC	ATTCACA	MEYSKNAYWNHIYG	
		CAAAAGTCCTGGCAACTGATTTGAATAATGTATAAAAGTAACTGGAACC	AAGGCT	NSYTSVPLRYGHQKP	
		AAATGCCCGATAGGTAGGGCGGGAGAAAATGACCTAGAAAACACAAA	AACTGG	DGDDWKKELSCEPV	
		GTCCCAAGCCCCCGGATTCGAAAGACCTATAGGAAGTCAGTGAATAGA	TCAAGT	LALQQ (SEQ ID NO:	
		GAGAAATATCAAACAAATCTCACCCATTCACAAGGACTTACTGGTCGAG	AGAGCA	1420)	
		TAGAAAACAAGCCAAAACATCAAGCACGACGCAAAAAGGGGTAACTTT	CAAGCT		
		GGGCAACTAATTAACGGATACCTCCGTGTATCAGGCAAAGCCGCCACCA	AAGCCT		
		ACAGCAAATTACTGCCCGATAGGTAGGGCGTGAGAAAATGACCTACAA	CCAAGC		
		CCTCCAAGACCCGAGCCCACGGAATCGAAAGACCTATAGGAAGTCAGT	ACGAAG		
		GAATTGATGGAAATACAAAACCAAATTTCTTCCATTCACAAGGACTTACT	TGATAT		
		GGTCGAGTAGAGCACAAGCCAAAATATCAAGTATGACGCAAAAAATGGG	GGGTAA		
		TAACCTTGGGCATCCAATCAACGGATACCTCTGCGTATCAGGCAAAGTC	TTTAGG		
		GCCACCAAACTGTACTACTCCGAAAAAACCAAGAAACATGATTTTCCCAC	CAACCA		
		TCCGTTAAAGCATCTCAACCAAGCTAAAGCGGTAAGGTTATCATGTCAA	ATCAAC		
		AAGGTGTAGCTACAGCAACCTAAAGCCCGAAAGGTAGGGCCGTATAAA	GGATAC		
		AAGACCTACACCCTCCAAGACCTAAACCCACGAACTCGAACGACCTACA	CTCCGT		
		GGAAGTCCGTGAATGGAGAGAAATATCTCACCAAATCTCTTCCATTCAC	GTATCA		
		AAAGGCTAACTGGTCAAGTAGAGCACAAGCTAAGCCTCCAAGCACGAA	GGCAAA		
		GTGATATGGGTAATTTAGGCAACCAATCAACGGATACCTCCGTGTATCA	GTCGCC		
		GGCAAAGTCGCCACAAACACTGTACTACTCCGTTACTCCCAAACACATG	ACAAAC		

			GATCTCCTTCTCTCACCAAAAAGCTTTATAACCAAGCTAACGGTGGAAAG GACATCATGTCACGAGGAGTAGCTACAGTAACCTCTCTCTTGAGACTGC AAAGTCGAGGATGGATTGGGAAGGCCGCGAGGCAAAAGGCGGGTAAC TCGGCCAGACGCTAGTGATCTTCGGATCCGACAGCCCTGGCCTTAGAGG AACCCTGGGATAAGGAGCACGACGGGAAGGATGTTCCGCAAGGATTTC CCTTCCCATTAGTCAGGGCTGGCAGTTGGTAATATAGCCTTTCTACACAC CACCGTCTTGCACCCACTAAACCAGTGGGATATGCGGGTGGACTCAATG TAGAAAGGTGTTCCCACTGCCTGACTCGCCAACTTTATATGTCTTGTCAA CATAATGGCCCCTCACTATAAACTCCCTAGCAACTGGTGGTCCGGCGAA GCCGGTTCTTGCCACTATTGCGCCCCAGGCTCGCC (SEQ ID NO: 1052)		ACTGTA CTACTCC GTTACTC CCAAAC ACATGG ATCTCCT TCTCTCA CCAAAA AGCTTTA TAACCA AGCTAA CGGTGG AAAGGA CATCAT GTCACG AGGAGT AGCTAC AGTAAC CTCTCTC TTGAGA CTGCAA AGTCGA GGATGG ATTGGG AAGGCC GCGAGG CAAAAG GCGGGT AACTCG GCCAGA CGCTAG TGATCTT CGGATC CGACAG CCCTGG CCTTAG AGGAAC CCTGGG	


ATAAGG
AGCACG
ACGGGA
AGGATG
TTCCGCA
AGGATT
TCCCTTC
CCATTA
GTCAGG
GCTGGC
AGTTGG
TAATATA
GCCTTTC
TACACA
CCACCG
TCTTGCA
CCCACTA
AACCAG
TGGGAT
ATGCGG
GTGGAC
TCAATGT
AGAAAG
GTGTTC
CCACTG
CCTGACT
CGCCAA
CTTTATA
TGTCTTG
TCAACAT
AATGGC
CCCTCAC
TATAAA
CTCCCTA
GCAACT
GGTGGT
CCGGCG
AAGCCG

NeSL	NeSL-4_SM	.	Schmidtea mediterranea	TTAAATCATTTTTAAATGTGTTTGAATATCTTAAATTATCAAATCATATTA	TTAAATCA	TGAGTG	MNVDLDATIKSIGM
				ATATCAATGCTAAAAAAAAAAATCGTGCKCATCAGGCGCACGAAAATAATG	TTTTTAAAT	TGCTAC	NTKETTYPNSQLRVE
				GACACAACTCGTCGACCTGCTGTCGACTCACAGAGAACCTCAATTTGGA	GTGTTTGA	GAGGCA	TTPCTSTTIMHASCN
				AGAATGGGAAGCCTATAATGCTACAATTCCGCCAACCCCTATTTGAATG	ATATCTTA	GCGCTG	TTSTISYSPLPSAVSLP
				ACAGATAGTCAAATATCAAAAAATATACAAACTGCTGTCAAGCGTGACT	AATTATCA	GTAATT	ESPASSITITTTDDNC
				CACTTCCTTCCAATCGAAAAATAGGAAKATGTAAGAAACATGAAAGTCA	AATCATAT	GCATCG	DIIETPYPLPQTNGDL
				AGCTGAAAAACCAATAATATGTCCTAAAATAAAACAATTTGAAAAATATG	TAATATCA	GCGTTG	SEILKDIEANKDTTMS
				CAAAAAATACCTATAAAATCACAGCCGAATAAATTCCCATCCGTTCTAAG	ATGCTAAA	CAGATTT	NKVLDCDSDSGDDR
				CAGAAACCGCTACGAACTACTGCAAGAATCGGATCAAGTATATTAATTT	AAAAAATC	GTGTAC	DMIIENDRESDMDLF
				CCCCCCMGGGGGAAATTAATATACTTGTTAKAAAATTAATTTTTTAATAA	GTGCKCAT	GATAGA	SQSLLNTNQSDERRE
				AAATAAATAAATCGAATAAATATAAAATAAAAATCAAATTAAACTT	CAGGCGCA	TAAAAA	KNLTENAPTEITTEKS
				TTATTAACAATAAAATCGCAGTAAGTAAATTTCCACTGTTATTAAATTTA	CGAAAATA	CCAATA	YFDIISKASDNTTSKKL
				AAACAAAATTCCTTTAAAAATGCCTCTCTTTTTCAGTAATAACACCTTTTC	ATGGACAC	GTAATA	LNVKNELTAGLPPMP
				TTGCTTTTATTACTATTTCTTGTGTACTGTACAAATCGAGCACAGTTATTG	AACTCGTC	AATGCT	PVTNTAKFIRNVRPE
				CAAATAGGACATAGAAATTCCTTTTTAAGTAAATTTAAATCCATGAGAAA	GACCTGCT	GAGCCT	DIADPTLYRLDSRGKL
				TAAAATAAAATCCTTTTGATTCAAAGTTTCTATGTTGCTTTCTAATAGAAT	GTCGACTC	AGCTCG	GCRTQYKKPGCGDIA
				GGTGTAAGCATTAATGGGTCTTGATTTTTATAAATTAAATATATTTAATC	ACAGAGA	CATATCT	VYDYEAIVEHAAFIHT
				TATTAAATTAATATGTTTTTATTAATTATTTTTATAGTGGGGGGGAAA	ACCTCAAT	AAGCCG	IPFNEQNNVDCQPC
				TTAATATACTTGATCCCAAGAATCAACTGATGATGAAGAATATGTTATTT	TTGGAAGA	AAAGGC	HPKKGKDVHTIVLIKY
				CAAAATACATACAAGAAGCTGGAAAAAACAAATCAATCGCTACAATGAA	ATGGGAA	AGCATA	ADIFNHIEAHSHVVQ
				TGTGGATCTCGATGCAACAATTAAAAGTATTGGAATGAACACAAAAGAG	GCCTATAA	TATATG	TAITDNMKTYLRLTKE
				ACGACCTATCCAAATTCACAACTGCGAGTTGAGACGACTCCCTGTACCTC	TGCTACAA	AGACAA	NXFYCSYRNNKKKNK
				AACGACTATTATGCATGCATCTTGCAACACACAACCAGCACTATATCTTACT	TTCCGCCA	TTTAAAA	CKKAFNLESNMMDI
				CTCCATTACCATCGGCTGTGTCACTTCCCGAAAGCCCTGCCTCGTCAATC	ACCCCTAT	AAAAA	TEHMKTHTGYSFDX
				ACAATAACCACACAACAGACGATAATTGCGATATTATAGAGACCCCTTACC	TTGAATGA	(SEQ ID	NLNILCYCGIWKPFTE
				CATTACCTCAAACACAAATGGTGACTTGAGTGAAATATTAAAGGATATAGA	CAGATAGT	NO:	LIAHIKTEHLQEYINSI
				AGCTAATAAGGACACCACCACCATGTCGAATAAAGTATTGGACTGTGACTCT	CAAATATC	1299)	PNKENIHNTTTIVSPL
				GACAGCGGCGATGATCGGGACATGATGATAATAGAAAATGACCGAGAATCT	AAAAAATA		NFAGILASGETQNIP
				GACATGGACCTGTTTTCGCAATCTTTATTGAACACTAATCAATCTGATGA	TACAAACT		DEEIIKPRDLPENLAF
				GAGGAGGGAGAAAAACTTAACAGAAAATGCTCCAACAGAGATTACTAC	GCTGTCAA		NRNIENELSWSQHLV

		TGAGAAGAGCTACTTTGATATCATCAGTAAAGCATCTGATAATACAACCT	GCGTGACT		KAYIFSYAVKTSTIFIN
		CTAAGAAACTGCTTAATGTAAAAAACGAATTGACTGCTGGACTACCTCCT	CACTTCCT		PYTCNALIQCNYKTFF
		ATGCCTCCAGTGACCAATACTGCAAAATTCATTCGAAATGTTCGACCTGA	TCCAATCG		ETFPFKDFAKWNEIV
		GGATATTGCAGATCCTACCCTATATCGACTTGACAGCAGGGGAAAGCTT	AAAAATAG		LPIHNNTSSWSFFFL
		GGATGCAGAACWCAATACAAAAAACCCGGATGCGGGGACATAGCAGT	GAAKATGT		NKKKRVAMIIDPTAD
		ATATGACTATGAGGCGATAGTTGAACATGCCGCATTTATCCACACAATC	AAGAAACA		DSHTLHFELATDILRTI
		CCATTTAATGAACAAAATAATGTGGATTGTCAACCATGCCACCCTAAAAA	TGAAAGTC		LNVQNIFEDLNFPLTE
		AGGAAAAGATGTCCATACAATAGTTCTGATAAAATATGCAGATATCTTT	AAGCTGAA		VEYPVCHEANLSAFX
		AACCATATTGAAGCCCATAGCCACGTTGTGCAAACCGCGATTACAGATA	AAACCAAT		VCHFLKCLMSDLPIDI
		ACATGAAAACCTATCTACGTTTAACAAAGGAAAATTTKTTCTACTGCTCA	AATATGTC		PDIDHMKETMRPIIR
		TATCGTAACAACAAAAAAAAGAATAAATGCAAAAAGGCTTTTAACCTTG	CTAAAATA		KYNCAKFPESDVRNY
		AATCAAACATGATGGACATAACAGAGCACATGAAAACTCATACCGGATA	AAACAATT		RVLIEDLIYQLNLDTIT
		CAGTTTCGACMAAAACTTAAACATTCTATGCTATTGTGGTATCTGGAAG	TGAAAATA		CEEILCEIERINGRLNP
		CCGTTCACAGAGCTCATTGCCCACATCAAGACTGAGCATTTGCAAGAAT	TGCAAAAA		KRYFKESKPKTDIIHL
		ATATTAACTCAATACCAAACAAAGAAAAATATCCATAAATACTACTACCATA	ATACCTAT		QKKKSAELLCVKRLKF
		GTTTCCCCTCTAAACTTTGCTGGGATACTTGCATCTGGCGAAACTCAAAA	AAAATCAC		QISQKTEIGKIWEND
		TATCCCCGATGAAGAAATAATTAAACCCAGAGATCTGCCAGAAAATCTT	AGCCGAAT		DVDHRPPMARFLKT
		GCCTTCAACCGAAACATCGAAAATGAATTAAGTTGATGGTCGCAGCACT	AAATTCCC		FASQDCPVSNTSSINL
		TGGTCAAAGCATATATCTTCTCTTATGCTGTTAAAACATCAACCATCTTCA	ATCCGTTC		PYYMDTDTDXCTDC
		TCAATCCTTATACTTGCAATGCTTTGATCCAGTGCAACTACAAAACTTTCT	TAAGCAGA		ENLSHIMKNLDSSAP
		TTGAAACCTTCCCTTTCAAAGACTTTGCCAAGTGGAACGAGATAGTCCTG	AACCGCTA		GMDLITGGDWKKISP
		CCAATTCACAACAACACTTCTTCTTCTTGGTCCTTCTTCTTCTTGAACAAGAAA	CGAACTAC		KHELITAICNCILRNKV
		AAACGAGTTGCGATGATTATCGATCCAACTGCAGATGACAGTCATACCC	TGCAAGAA		CPEKWKLFRTVLILKP
		TGCACTTTGAATTGGCTACAGATATCCTAAGGACTATACTTAACGTCCAG	TCGGATCA		GKMSESFRANSWRP
		AATATATTTGAGGACTTAAATTTCCCTCTTACTGAGGTCGAATACCCCGT	AGTATATT		LAIMDTAYRIFTTLLN
		GTGTCATGAGGCAAACCTTTCCGCATTTTMTGTATGCCACTTTCTTAAAT	AATTTCCC		NRLLQWIRNGNLISP
		GTTTAATGTCGGACTTGCCAATTGATATTCCGGATATCGATCACATGAAA	CCCMGGG		NQKAIGIPDGCAEHN
		GAGACWATGAGACCAATTATTAGAAAATATAACTGCGCAAAGTTTCCG	GGAAATTA		ATLHFAIDRAKRCKTE
		GAGAGTGATGTTAGGAATTACCGCGTACTTATCGAGGACCTGATATACC	ATATACTT		LHIVWLDIADXFGSLP
		AATTGAACCTTGACACAATTACTTGTGAGGAAATACTGTGCGAAATCGA	GTTAKAAA		HDLIWYTLANMGLK
		AAGAATAAATGGAAGGTTAAATCCCAAAAGATATTTTAAAGAGAGTAA	ATTAATTTT		NETLTLIKELYKDVKTI
		ACCAAAGACGGATATAATACATCTGCAAAAGAAAAAGTCGGCGGAACT	TTAATAAA		FDCQGTLSEPVPITKG
		CCTCTGTGTTAAAAGATTGAAATTCCAAATCAGTCAGAAAACAGAAATC	AATAAATA		VKQGCPLSMTLFCLSI
		GGAAAGATATGGGAAAACGACGATGTGGATCACAGACCGCCTATGGCC	AATCGAAT		DYILKSILTNYPFLLHD
		AGATTCTTGAAGACTTTCGCGAGTCAAGACTGCCCCGTTTCGAATACGTC	AAATATAA		LNISILAYADDLVLLSD
		ATCCATAAACCTACCTTACTACATGGATACTGATACAGATAMGTGTACT	AATAAAAA		SYLEIKKSLESTVELAA
		GATTGTGAAAATTTGTCGCACATCATGAAGAACTTGGATAGCTCGGCAC	TAAATCAA		FANLKFKPSKSGYLSI
		CTGGAATGGACCTCATTACAGGTGGAGACTGGAAAAAGATCTCCCCGA	ATTAAACT		NNVNSDILKLHLYNE

		AGCATGAACTGATAACAGCAATCTGCAATTGTATACTACGAAATAAGGT	TTTATTAA		EIPTISENNKYRYLGV
		CTGCCCAGAGAAATGGAAGCTGTTTAGAACAGTTTTAATCCTAAAACCA	CAATAAAA		DFSYKRNQDVDGRL
		GGAAAAATGTCCGAGAGTTTCAGAGCTAACTCATGGAGACCTCTTGCAA	TCGCAGTA		GSALALTRSLFKSYLH
		TCATGGACACAGCCTATAGAATCTTTACGACTCTGCTGAATAACCGCCTG	AGTAAATT		PAQKLNAYKTFIHSKL
		CTGCAATGGATCAGGAATGGCAACCTCATAAGCCCGAACCAAAAAGCG	TCCACTGT		IFSLRNCVIGHRILDC
		ATTGGTATACCGGATGGATGTGCTGAGCATAATGCTACTCTACACTTCG	TATTAAAT		DRNRVTQGREKQLG
		CAATTGACCGAGCTAAACGATGTAAAACTGAACTACACATTGTTTGGCT	TTAAAACA		FDQEIKALLKTMIGD
		CGATATCGCCGATKCATTTGGTTCGCTGCCTCATGACCTGATCTGGTATA	AAATTCCT		KFQAXNNYFPYTHCK
		CACTGGCTAATATGGGTCTGAAGAATGAAACACTAACCTTGATTAAGGA	TTAAAAAT		LGGLGITSAIDEYLIQS
		ACTATATAAGGATGTGAAGACTATCTTCGACTGTCAGGGAACCTTGTCC	GCCTCTCT		ITGITRLFHSSNLSFRK
		GAACCTGTCCCAATTACTAAAGGAGTTAAACAAGGTTGCCCATTATCAAT	TTTTCAGT		MLITELAHSRGGKNF
		GACACTCTTCTGCCTGTCTATTGACTACATTCTAAAGTCAATACTGACTA	AATAACAC		EAGLKWLNCEVNKA
		ATTATCCCTTCCTTCTTCATGATCTGAACATCAGTATTTTGGCATATGCTG	CTTTTCTTG		FPNTSFFVKFQKSALA
		ATGACTTGGTTCTTCTTTCTGACTCTTATCTAGAAATCAAAAAATCTTTAG	CTTTTATTA		LKRKFCICVNLKFVED
		AGAGTACTGTGGAATTGGCAGCMTTTGCCAACCTTAAGTTTAAACCTTC	CTATTTCTT		NFSLEMTYKKRTSYV
		GAAGTCTGGATACTTGTCCATCAACAATGTTAACTCCGATATCCTTAAAT	GTGTACTG		NHQNLSTLSKELHDF
		TACATCTCTATAATGAGGAGATACCAACGATATCCGAGAATAACAAATA	TACAAATC		VGLYYAEQXCQMRV
		CAGATATCTTGGAGTTGACTTCTCTTACAAAAGAAATCAGGATGTTGAT	GAGCACA		QGHIATAIGDSITAKY
		GGACGACTTGGGTCTGCACTTGCACTCACCAGATCTCTATTTAAATCATA	GTTATTGC		LIASDILNDAQYYFLV
		CTTGCATCCGGCGCAAAAGCTGAATGCTTACAAAACCTTCATCCACTCCA	AAATAGGA		RARNNLLNLNYNAYR
		AGCTTATCTTCTCCTTGCGTAATTGCGTGATAGGTCATAGAATCCTCGAC	CATAGAAA		LKYNIGTKCRLCHLDE
		TGTGATCGGAATAGAGTTACACAAGGTCGGGAAAAACAGCTGGGCTTT	TTCCTTTTT		ETQAHXFNHCRAKP
		GATCAGGAAATCAAGGCACTWCTGAAAACCATGATTGGAGACAAATTT	AAGTAAAT		NARRVKHENVLVSIV
		CAGGCAKTAAATAACTACTTTCCTTACACTCACTGCAAGCTGGGGGGAC	TTAAATCC		AFLEKIGFEIDVEKSPK
		TTGGTATAACCTCAGCTATTGATGAATATTTGATCCAAAGCATTACCGGA	ATGAGAAA		YISIPTKLKPDMVIRSK
		ATAACAAGATTATTTCACTCATCCAACCTCAGCTTCAGAAAAATGCTAAT	TAAAATAA		RNKDIHVLDLKVPYD
		CACAGAACTCGCTCATTCTAGAGGAGGGAAAAACTTTGAAGCGGGGCT	AATCCTTT		SGEGFEKAREDNYVK
		AAAATGGCTTAACTGTGAAGTTAACAAGGCATTCCCCAACACCTCTTTCT	TGATTCAA		YKDLAIEIGKAFNQKA
		TTGTAAAATTCCAAAAATCGGCACTTGCTCTTAAAAGAAAGTTCTGTATA	AGTTTCTA		TISAVVIGCLGTWDK
		TGCGTTAACCTTAAATTTGTAGAGGACAATTTCTCACTTGAGATGACCTA	TGTTGCTT		KNNAALSKIGLTKTEII
		CAAAAAGCGCACTTCTTATGTAAACCATCAAAACCTCAGCACACTTTCCA	TCTAATAG		SLARIACPNAVIACYH
		AAGAACTCCACGACTTCGTGGGCCTTTACTATGCWGAGCAATGWTGTC	AATGGTGT		IYREHVSFTKSAMAL
		AAATGAGAGTACAAGGACACATTGCGACTGCGATCGGGGATAGCATAA	AAGCATTA		PFSLA (SEQ ID NO: 1421)
		CAGCTAAATACCTAATAGCTAGTGACATCCTTAACGACGCACAGTACTAC	ATGGGTCT		
		TTCTTGGTACGTGCGAGAAATAATCTTCTGAATCTTAACTACAATGCGTA	TGATTTTT		
		TCGACTCAAGTATAATATTGGCACAAAGTGCAGACTTTGCCACCTTGAT	ATAAATTA		
		GAAGAAACTCAGGCCCATSTGTTCAATCACTGCCGTGCCAAACCAAACG	AATATATT		
		CTAGAAGAGTGAAACACGAAAATGTGCTAGTAAGCATAGTTGCCTTCCT	TAATCTAT		

			AGAGAAAATTGGATTTGAGATTGATGTGGAAAAATCACCCAAATATATC TCAATACCAACAAAGCTGAAACCTGACATGGTAATTAGGTCTAAGAGGA ATAAAGATATACATGTCCTAGACCTAAAAGTGCCATATGACTCAGGAGA AGGCTTTGAAAAAGCGCGGGAAGACAACTATGTTAAATACAAAGATCT AGCCATTGAAATTGGAAAGGCATTTAATCAAAAAGCGACTATATCTGCT GTGGTGATTGGATGCCTGGGCACATGGGACAAGAAGAACAATGCCGCT CTTTCCAAAATCGGGTTGACTAAGACCGAGATCATATCTCTTGCCAGGAT AGCATGCCCAAATGCGGTAATCGCATGCTATCACATATACCGTGAGCAC GTCTCATTTACAAAGAGTGCCATGGCCCTCCCCTTCAGCCTTGCATGAGT GTGCTACGAGGCAGCGCTGGTAATTGCATCGGCGTTGCAGATTTGTGTA CGATAGATAAAAACCAATAGTAATAAATGCTGAGCCTAGCTCGCATATC TAAGCCGAAAGGCAGCATATATATGAGACAATTTAAAAAAAAAA (SEQ ID NO: 1053)		TAAATTAA TATGTTTTT ATTAATTA TTAATTTTT ATAGTGGG GGGAAATT AATATACT TGATCCCA AGAATCAA CTGATGAT GAAGAATA TGTTATTT CAAAATAC ATACAAGA AGCTGGAA AAAACAAA TCAATCGC TACA (SEQ ID NO: 1176)	TGATCC GTGTGT TTGTGTC GTATGA TTGTTTC CGTGTG TGTCTAT ATTTTTC TTTTTTA TACTTTC AATTACC TCGTTGT AATGTT ATAACTT CATATG GAATAT ATGTAA TTTAGTT	
NeS L	R5	AY216 701	Girardia tigrina	GTAGGTAACTATGACTGCAAAATAATAATTCTACACCTATTGTTGATAAC TCATCTCGTGCGCAAACGGAGCATGTTATTTCTAATCATTTCGTCACACA GGATTCTTCTAATTCTGATAGTAATATTATAGATAGAGATAGGAACCTTG TTGATTTAGATGCGTCAATAACTTCTCCTACTATTATACAGCCAGAGGAT AGTAAGATATCTGAGGATGAGGACTTCATCTTAGTCAATAGGAAAAAGA GCAAAAATAAGAAAAAATCTAAGAAAACAACTGAAAATAAAAATGAAA TTCCTATTCAAAAGAGTAAAGATAAGAAAAAGAAGTCTAAAATTAATAC CGAAAAACTAACTGAAAATATTACTACTTCTGAAATACCACTTGAAATTG CTCCTTCCATACCTTTACCTTCAGCAAGTACCTCGGGTTCTCAACAACCG GCCAATCCTCCAGAAGACGCTACTCTAAGTGATACGGATCTCTTCCTTAC ACAGGATGATCCCGATAGTCTTATTCTTTCTGGAAGTACTCAACCAACCT TTGTTGACCTCAACCCTTCACAGCAATCGGAACTTCCTTCAAATACTGAC AGCCAAAGATTTGAGGCGGGTGAAACACCCAAAATCATAACTTCTTACA GGGATGACCTTTTCTACTCTACAGTCCTTCACTACAACTCAGATACAGGT TACGGTATAAGTGTTGACAATGGGGAGCAGAGGTTTCGAATTCTTGCTA GGAATCTTGTCAGGAAAACCAAGGATAAGTTCCCCTCTTTATATGCTGG ACAAGTAATTAGACACACAGTCTTCTTCAATCACTTCAACCAGGCATACT ACGCCAATAATATAACTGATAGTAAAGGTAATCTAATTGAGTTTTCTGAT	GTAGGTAA CTATGACT GCAAAATA ATAATTCT ACACCTAT TGTTGATA ACTCATCT CGTGCGCA AACGGAG CATGTTAT TTCTAATC ATTTCGTC ACACAGGA TTCTTCTA ATTCTGAT AGTAATAT TATAGATA GAGATAG		

	GATAAGCCTTTTCAAAGTATACCGACTGACCCAAAAACTGAACTAGAGC	GAACCTTG	TAGTTTA	DDFEFRIPGRDQIAKY
	AAATTAGGAGAGAGAGACAACATCTAGTTGATAGAGCTCTTAGACATAA	TTGATTTA	GTTAGTT	IKFASKSAAGLDWITY
	TCAGTTACGGGAAACTTATATTTTAAATAAACTTAATAATAATGGGG	GATGCGTC	TAGTTTA	EDIKLGDPSGEILQPIF
	GGGGTGGCGAACATTTGAAAAGGAAAAAGATCAAAGTCAATACGGATG	AATAACTT	GTTTAGT	EYIVQNNICPSEGKAS
	ATGTCTCCAGCAATGATGGAGACAGAAAACATAGACGACAGGAAGAAA	CTCCTACT	TTAGTTT	RTIMIPKPGKSDYSDP
	CCTCGGACAATGGAGCTGCTATTCCCGCTCAATCCAACAATCAAATTACT	ATTATACA	AGTTAG	SSWRPITITSAVYRLL
	CCTTTAAACTGAGTTCGACTGAAGTGGGAGAGCTCGTTGAGCAATCTCC	GCCAGAG	TTTAGTT	MKYLTWELYNWILL
	AGCTCCCCTTCAGTCGCCTCAGTTCTCTAACAATTATAACAATCTAAATAT	GATAGTAA	AGTTTA	NQMLSRSQKSLGKFE
	CAACAACTTATACTATAGTCTCAATACTTTTAATCAGTCTAATAACCT	GATATCTG	GTTAGT	GCHDHNAMLNMLI
	TTGCTGTCTTGTTAATATTGAGTTTTTCCCAACTCAACACCTTCTTGGTGA	AGGATGA	(SEQ ID	QDVRRQTNPSNPIN
	TATAGTTAACTCGGGGTGCATAAACTATATGAATAATTATAATAACTTTG	GGACTTCA	NO:	KNKRLYIVFLDFTNAF
	ATAATATTAATTTATATATTAATAGTAATGTATTATCTTACAATAATTATA	TCTTAGTC	1300)	GSVPLDTLMYVPQRF
	ATCACAGTTTTCTCGCTTCCCCATATACTACTCAACATCACAGAACATGCA	AATAGGAA		GLGTSALTLIKNLYLD
	GACATAAACATGCACGTGCAAGAAGTTAACATGCAGCAAGATAACAAT	AAAGAGC		NYTNVTCGESKIENV
	ACACAACATGCTATAACACAACAAGTCTCTCTACAAGCAACATCTCTGCA	AAAAATAA		KLNKGVKQGCPLSM
	ACACACGTTGGACGAAATGATAGTCCAGTTTAACACTGCTGTCAGGTTA	GAAAAAAT		LLFNIFINIIIRAIEAMP
	AAGAAAAAGCACACAAAGTTGCAAAAATCTTTAGGGGACATAATCATCGTA	CTAAGAAA		DVHGYPLGDMDIRIL
	AAGACCTTCCAACATTGCCTGCTAGGGAACAGTATATAAAACTAAACCGAA	ACAACTGA		AYADDIALISDSHKDL
	ACTTGCAATTAGAGAGGTACTTCATCGAAAAACAACAGCTACGTCTTCCC	AAATAAAA		QEMVYKAEYIGRILGL
	CTTCTGAAAATGCAATTAAGGCTTTTTTCTCCTCCTACAGCCGTCCAGCT	ATGAAATT		LFNPSKCALMDIPHD
	GAACTTTTCACTGGTCAGGAACTTCTTGAATCATCTTGGTTCCCAGTACA	CCTATTCA		KKRTPPILVNGEMIKC
	CCCGGAAGATGACTTTGAGTTTAGAATTCCGGGTAGAGACCAAATAGC	AAAGAGTA		VGKADPYKYLGTFRS
	GAAATACATCAAGTTTGCTAGTAAATCAGCTGCTGGTCTTGACTGGATC	AAGATAAG		WFRKLDIKELLQMM
	ACGTACGAGGATATTAAGTTAGGCGATCCGTCCGGGGAAATTCTCCAAC	AAAAAGA		MDETKLITESNLHPH
	CCATTTTTGAATATATAGTACAAAATAACATATGCCCATCCGAGGGGAA	AGTCTAAA		QKIHAYETFIHSQLPF
	GGCTAGTAGGACCATTATGATTCCCAAACCGGGAAAAAGTGACTATTCA	ATTAATAC		HLRHSRIPFSDFITNR
	GATCCTTCTTCTTGGCGGCCCATTACAATTACCAGCGCAGTATACAGACT	CGAAAAAC		KTNKTTNNSNDSEKS
	TCTCATGAAATATCTTACATGGGAGCTGTATAACTGGATTCTTCTTAATC	TAACTGAA		IQKAYDPESGQLFLN
	AGATGCTGTCCAGGAGCCAAAAGAGTTTAGGGAAGTTTGAGGGATGTC	AATATTAC		TFALPSGCAKDFFYIT
	ATGATCACAACGCAATGTTGAACATGCTCATCCAAGACGTTAGGAGACA	TACTTCTG		KDAGGPQLTSGLDEY
	GACCAACCCGTCTAATCCAATCAATAAGAATAAGAGGCTATACATAGTC	AAATACCA		LIQSIMYIFRLLGSEDP
	TTCCTAGACTTTACGAATGCTTTCGGGTCGGTTCCGTTAGATACTCTCAT	CTTGAAAT		TLNSAIKHDLISHLNL
	GTATGTCCCTCAACGCTTTGGCTTAGGCACCTCTGCTTTAACGCTGATTA	TGCTCCTT		KGFVNINFSQAISIFN
	AAAACTTATATCTAGATAACTACACAAATGTAACATGTGGGGAAAGCAA	CCATACCT		SNFTDRTDHFSHLSR
	AATAGAAAACGTAAAATTAAATAAAGGGGTTAAGCAAGGCTGCCCTCTA	TTACCTTC		TEWARLQLARKKLKS
	TCTATGCTGCTTTTCAACATTTTTATCAATATTATAATTAGGGCAATAGAA	AGCAAGTA		TLAIQTNVCLINGHLV
	GCTATGCCAGATGTCCATGGATACCCACTTGGAGATATGGACATCCGGA	CCTCGGGT		LTLSLENNVLLIDSKEK
	TACTGGCATATGCTGATGATATTGCTCTAATATCTGACTCCCACAAAGAC	TCTCAACA		GDVKKIHASLMGFLR

		CTGCAGGAAATGGTCTACAAGGCGGAATATATCGGTCGGATTCTTGGAC	ACCGGCCA		LAHLIRLQKHGWSKL
		TACTCTTCAACCCGTCAAAATGTGCACTTATGGACATTCCGCACGACAAG	ATCCTCCA		LFSATTHHEILNKRIL
		AAGAGGACGCCGCCTATCCTCGTCAACGGTGAGATGATCAAGTGTGTTG	GAAGACG		NGHVPYKIWYFIHRA
		GAAAGGCCGACCCATACAAATATCTTGGAACCTTTAGATCCTGGTTCCG	CTACTCTA		RLGLLPTKLFSVSNLC
		GAAGCTGGATATAAAGGAGCTCCTCCAGATGATGGATGAGACTAA	AGTGATAC		RKCGGKKETMSHAL
		ACTCATCACCGAGTCAAATCTACATCCTCACCAAAAAATCCACGCGTATG	GGATCTCT		VNCPMMQTLINERH
		AGACCTTCATTCACAGCCAGCTCCCATTTCACCTTAGACACAGCCGAATT	TCCTTACA		DALEISLVQILSSKFQ
		CCGTTCTCAGACTTCATAACAAACAGAAAAACAAACAAAACAAACA	CAGGATGA		GTVIRQKTYVNELRP
		ATTCAAACGACTCAGAAAAATCTATACAGAAAGCCTACGATCCGGAATC	TCCCGATA		DITMESDTQYYLVEV
		AGGACAATTATTCCTCAACACCTTCGCCCTTCCAAGTGGATGTGCTAAGG	GTCTTATT		KCPFDTKMSFELRTQ
		ATTTCTTTTACATTACAAAAGATGCAGGTGGACCTCAACTCACAAGCGG	CTTTCTGG		QTTDKYNIIIEILEDVH
		ACTGGATGAGTACTTAATCCAATCAATTATGTACATCTTCCGACTATTGG	AAGTACTC		PGKEVRLVTFIVGTLG
		GCAGTGAGGACCCCACCTTAAACTCTGCAATAAAACATGATCTCATTTCC	AACCAACC		SWGPQNSDFLRDLG
		CACTTAAATTTAAAGGGTTTTGTAAATATTAATTTTTCTCAAGCCATTTCA	TTTGTTGA		FSKDEIDQVKTRLML
		ATCTTTAATTCAAATTTTACCGACCGAACCGATCACTTTTCACATCTTAGC	CCTCAACC		QNINSSCEQWKRFV
		CGCACTGAATGGGCAAGACTTCAATTAGCTCGGAAAAAATTGAAGTCAA	CTTCACAG		QYAPTITPGPIPDAES
		CCTTAGCCATCCAAACTAATGTCTGTCTGATAAATGGGCATCTTGTCTTA	CAATCGGA		EDDQGTSDNGPTAA
		ACTCTTTCGCTAGAAAACAACGTTCTGTTAATTGATAGTAAAGAAAAGG	ACTTCCTT		TVQGPVIGDEEEELQI
		GGGATGTCAAGAAGATCCATGCATCCCTCATGGGGTTTCTTAGGTTAGC	CAAATACT		YDSGLDESSDDEPDP
		TCACCTTATCAGACTGCAAAAACATGGATGGTCAAAACTGCTCTTCAGT	GACAGCCA		DDAELLFTIDIEQYLN
		GCGACCACTCATCACGAAATACTAAATAAGCGTATCTTGAATGGTCACG	AAGATTTG		SVITD (SEQ ID NO:
		TCCCTTATAAGATTTGGTACTTTATTCATAGGGCGCGGCTGGGGTTGTTG	AGGCGGG		1422)
		CCTACTAAACTCTTTAGTGTTAGTAACCTTTGTAGGAAGTGCGGGGGGA	TGAAACAC		
		AGAAAGAGACCATGTCGCATGCTTTGGTCAACTGTCCAATGATGCAGAC	CCAAAATC		
		CCTCATTAATGAGAGACATGATGCTCTTGAAATCTCCCTTGTACAAATTC	ATAACTTC		
		TTTCTTCTAAATTTCAGGGTACGGTTATAAGGCAAAAGACCTATGTCAAC	TTACAGGG		
		GAGTTAAGACCCGATATCACAATGGAATCGGATACCCAATATTATCTTG	ATGACCTT		
		TTGAGGTAAAATGCCCCTTTGACACGAAGATGAGTTTTGAATTGAGAAC	TTCTACTCT		
		ACAACAAACTACTGATAAATACAACATTATTATTGAAATATTAGAAGATG	ACAGTCCT		
		TACACCCTGGGAAGGAGGTGCGCCTTGTTACGTTTATTGTAGGCACCTT	TCACTACA		
		AGGCTCATGGGGCCCGCAGAACTCGGACTTTTTGAGAGATCTGGGATTC	ACTCAGAT		
		TCCAAAGACGAAATCGACCAGGTGAAGACGCGGCTGATGCTTCAGAAT	ACAGGTTA		
		ATCAATTCCTCCTGCGAGCAGTGGGAAAAGATTTGTGCAATATGCACCCA	CGGTATAA		
		CAATTACACCTGGGCCGATTCCAGACGCGGAGAGCGAGGACGATCAGG	GTGTTGAC		
		GGACGAGCGACAATGGGCCAACAGCTGCTACAGTGCAAGGACCGGTGA	AATGGGG		
		TTGGCGATGAGGAGGAGGAACTTCAAATCTACGATTCCGGCCTTGACG	AGCAGAG		
		AGTCCAGCGATGATGAACCGACCCAGATGATGCTGAATTACTTTTCAC	GTTTCGAA		
		AATTGACATAGAACAATATTTGAATTCTGTGATAACAGACTGATCCGTGT	TTCTTGCT		

	GTTTGTGTCGTATGATTGTTTCCGTGTGTCTATATTTTTCTTTTTTATAC TTTCAATTACCTCGTTGTAATGTTATAACTTCATATGGAATATATGTAATT TAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAG TTAGTTTAGTTAGT (SEQ ID NO: 1054)			AGGAATCT TGTCAGGA AAACCAAG GATAAGTT CCCCTCTTT ATATGCTG GACAAGTA ATTAGACA CACAGTCT TCTTCAAT CACTTCAA CCAGGCAT ACTACGCC AATAATAT AACTGATA GTAAAGGT AATCTAAT TGAGTTTT CTGATGAT AAGCCTTT TCAAAGTA TACCGACT GACCCAAA AACTGAAC TAGAGCAA ATTAGGAG AGAGAGA CAACATCT AGTTGATA GAGCTCTT AGACATAA TCAGTTAC GGGAAACT TATATTTTA AATAAACT TAATAATA ATAATGGG GGGGGTG		

				GCGAACAT TTGAAAAG GAAAAAG ATCAAAGT CAATACGG ATGATGTC TCCAGCAA TGATGGAG ACAGAAAA CATAG (SEQ ID NO: 1177)		
NeSL	Utopia_1B_CPB	Chrysemys picta bellii	GTTTAATTCCTTCTGATGGACATCTGCAACACCCGTCCTGAAGATGGAGT	GTTTAATT	TAACCG	MESPAXIFEKIDAALX
			CTCCTGCWKCCATTTTTGAAAAAATTGATGCTGCCTTGMAGATATACTC	CCTTCTGA	AGACCG	IYSAAAXLXXNSLSLSP
			CGCTGCTGCTGMTTTGGAWGAMAATTCTCTCTCTCTCACCTWCAGM	TGGACATC	CCGACC	XXAXXSXXAAPASST
			TGCASTCWCGTCAMCTSCTGCTGCTCCTCGTCCTCTACTCCCCAGAAA	TGCAACAC	AGGGAA	PQKTQXKPIPXTTLG
			ACTCAGWGGAAGCCTATCCCGAAKACCACTCTTGGTGCCTCACGGAAG	CCGTCCTG	ATAACC	ASRKXRTTXKDEXIXX
			AMCMGGACCACCASCAAGGATGAAAAMATCAGSASCTGGCKGAAGAA	AAG (SEQ	CACTTCC	WXKKAPVDTSXGRX
			AGCCCCTGTGGATACCTCTKCAGGGAGAMCTAGCACCAGAAGGACAGC	ID NO:	TTCCCTG	STRRTALRDLTSRSXN
			TCTTCGGGACCTCACATCCAGGAGCAGKAATATCWCAMCAGCTCTTCA	1178)	ACGAAC	IXXALQEEDPRRTPPX
			GGAGGAGGACCCCCGGAGAACCCCTCCCWCTTCCCGGGACCAGGATGC		CAAGGG	SRDQDAERRPAAPEK
			TGAGCGCCGCCCTGCTGCTCCTGAGAAGGCTGCTACCAGAGGAGCCCCC		ACGCAC	AATRGAPPTIQDQD
			CCGACGATCCAGGACCAGGATGCTGATCGCTGCCCTGCTGGGAGGGAT		CCCACCC	ADRCPAGRDATGGA
			GCCACCGGAGGAGCCCCCCGACGACCCAGGACCAGGATGCTGANCGCT		ATGTACT	PRRPRTRMLXAAPLG
			GCCCCGCTGGGAAGGATGCCGCCGGAGGAGCCCCCCCCGACGACCCGG		TATTCGC	RMPPEEPPPTTRDQ
			GACCAGGATGCTGACCGCCGCCCCGCTGCTCCAGAGAGGGGATGCTCCM		TACACC	DADRRPAAPERDAP
			GAAGGAACCACCTCCTCAACCCCGGACCCCGAAACCACTTACCACCCGC		GATACT	EGTTSSTPDPETTYHP
			CTGTCCGGAGGAGGGCCGCTCCAAGGGGAACGCACTCCGSAGCCCWG		GACTTG	PVRRRAAPRGTHSXA
			GATCTCGATGCCGCACGCTGTCCTTCCGGGCAAAGAGACATCGTGGCCA		GACTCCT	XDLDAARCPSGQRDI
			GTGAGTCCAGCACCCCCCCAGGAGCGACTTCACCTCCTCAAGCTTCTCTG		TATACAT	VASESSTPPGATSPP
			CCAGACCSAGAGGAATCACCTGCCGAGTCKGCAGGCACAACAGAGGTC		TCCATG	QASLPDXEESPAESA
			CGCCCCACAGAGGGTGAGGCAGGGGAAGACGACTGCATCTACCTCCAG		GGTGGC	GTTEVRPTEGEAGED
			TACCCGMTCCCTACAGGCCTCCTCCTCTGCCCCTTCTGCCTCCCCKTCCAT		GTACCC	DCIYLQYPXPTGLLLC
			GGAGTCCAGACCCTCGGGGCCCTCAGCAAACACGTTCGKAAGGCCCAC		GAGCCC	PFCLPXHGVQTLGAL
			AACAAACGGATTGCCTTCCGGTGTAGCCGCTGCGATGCACCMTTCGAG		ACTTATC	SKHVRKAHNKRIAFR
			ACTCAAAAGAAATGCAAGTMCCATCAMGCCACATGCAAGGGACCCCTC		CACTGA	CSRCDAPFETQKKCK
			ACAACCGCGAAGGTSAACCCCACTGACACCCTGCGGGTTCCAACCCCGA		CACTTTA	XHXATCKGPLTTAKV
			CCCCCACCGACGGTCCAGCTTCAGCACCCCAGCCAGCATCCCCAGAGCC		AAAACT	NPTDTLRVPTPTPTD

	ACAGCANGTAAGGGGGGGACCAACCGCCAACCGAGGGAAGCGCAACCC		CTTGCAC		GPASAPQPASPEPQX
	CCGCCTCGAGGACTGACGATGCCACCAAAAGGACCAGCCCCGCCTCCAG		CCCAATC		VRGDQPPTEGSATPA
	AATCCCCACGCTGGACCCTGCCGTGAGGGGGATCACCGCCACCTCSCAG		TGGGTC		SRTDDATKRTSPASRI
	GTCAGCGACCTCACCAGATGCCTCAGCGACCTCATAAAAACCATCCGGC		TATGCC		PTLDPAVRGITATSQ
	ACAACACGGATACGAGACGGKGCAGCGCTCCCCCACAGGTMACCTCAT		GGTTAT		VSDLTRCLSDLIKTIRH
	GCCGCCCTGCCGTAGGAGCAACTAGCACTGCCCCCCAGGCCGCACGGC		GCGATA		NTDTRRXSAPPQVTS
	GAGACCCAGCCAACGGAGGAGCCTCCCGCAGCCCCCAGATCCCACGGC		TGTATGT		CRPAVGATSTAPQA
	CGGACCCCGCCCCCGGGAGACCCAACACCTCCTCCAAGGTTACCCAACG		ATCTCTT		ARRDPANGGASRSP
	AGACTCCGACCGCCAAAAACCCCATGCCCCCACCGAGGACCCCCCCAGCCG		CATCCTT		QIPRPDPAPGRPNTS
	GATACCACCCGCAGAAGAACCAGAACCATCCCCAGCGCTTCCAAACACG		GCAACC		SKVTQRDSDRQKPH
	ACCGCGCCCCCACAAAGCCCCAACACCGGTGTTTCCAGAACCCCACTCCCT		GATACC		APPRTPQPDTTRRRT
	CCCGGAAGATCCAGCGCTGCCTCGGAGACACCGAGAGCTGCCCCCCCCCC		TGTAATC		RTIPSASKHDRAPTKP
	CACCGCACCAGGACCCCCGCCTCAAGACCCACCTGAACACCGCTCCCCA		CCTCATA		NTGVSRTPLPPGRSS
	GTCCGAGGGACAACAAGGCCACAGACTGTCTCCGCAGCACCTGAACCC		ACCCAA		AASETPRAAPPPPHQ
	GCGGAGACAACGCAGCAGGAGGAACGACGGCCGCGAGCAGAGGGTCG		GCCTGA		DPRLKTHLNTAPQSE
	CCACGCCGTGGCAATCCGCCTGGATGGAGGAGCTGGCAAAGGCTGAGG		CCCCAG		GQQGHRLSPQHLNP
	ACTTCGAGACCTTCGACACCCTGATGGACAGACTGACCGCAGAACTGTC		ATGTAC		RRQRSRRNDGREQR
	TGCGGAAATTACAGCCAGAAGGAGGGAACCCCAGGAGGCCTCACGGG		AGTACC		VATPWQSAWMEEL
	CCACGCGCAGATTCCCCGCGCCGACCCGTAACAACACTGCCAGAGAAG		TTCCCTC		AKAEDFETFDTLMDR
	GCAGGAGAGGGGACGTCGGCCGCCGCTACGATCCGGCAGCTGCATCCC		TTAACTC		LTAELSAEITARRREP
	GTATTCAGAAACTATACAGGACGAACCGGACGAAAGCCATGAGGGAGA		GTGTAT		QEASRATRRFPAPTR
	TCCTCGACGGGACCTCCTCCTACTGTGCCATCCAGCCCGAGAGACTCTAC		ATTTAAT		NNTAREGRRGDVGR
	TCCTACTTCAAGGATGTGTTCGACCACGAGGCCCAGACCAACTTGCAAC		TTTAAAC		RYDPAAASRIQKLYRT
	GCCCAGAGTGCCTTCTCCCGCTACCCCGGATCAACCTCACGGAGGACCT		ATTAACT		NRTKAMREILDGTSS
	GGAGCGAGATTTTTCCCCGCAGGAGGTGCAGGCGAGGCTGATGAGGAC		TTAATAA		YCAIQPERLYSYFKDV
	CAAAAACACTGCCCCTGGAAAAGATGGCATCCGCTACCACCTGCTGAAG		AATTTTT		FDHEAQTNLQRPECL
	AAGCGAGACCCCGGCTGCCTGGTGCTTGCTGCCATCTTCACCAAATGCA		AAA		LPLPRINLTEDLERDF
	AGCAGTTTCATCGCGTTCCCCGCTCCTGGAAAAAGTCCATGACCGTGCTC		(SEQ ID		SPQEVQARLMRTKN
	ATCCACAAAAAAGGCGAGCGAGACGACCCCGGCAACTGGAGGCCCATC		NO:		TAPGKDGIRYHLLKK
	TCCCTCTGCTCCACCATCTACAAGCTGTATGCCAGCTGCCTCGCGGCAAG		1301)		RDPGCLVLAAIFTKCK
	GATCACGGACTGGTCAGTGTGCGGGGGCGCCGTCAGCTCAGTACAGAA				QFHRVPRSWKKSMT
	GGGTTTCATGTCCTGCGAGGGATGCTACGAGCACACAACTTCCTCCTTCAG				VLIHKKGERDDPGN
	ACGGCCATCCAGGAAGCCAGGAGGTCCAAGAGGCAGTGCGCCGTAGCA				WRPISLCSTIYKLYAS
	TGGCTTGACCTGACCAACGCCTTCGGGTCCATACCCCACCATCACATCTT				CLAARITDWSVCGG
	TGCCACCCTGGGAGAGTTCGGGATGCCAGAAACCTTCATCCAGATCCTC				AVSSVQKGFMSCEG
	CGGGACCTCTACAAGGACTGCACCACCCATCCGCCGGCGCCACGGACGGA				CYEHNFLLQTAIQEA
	GAGACGGACGCCATCCCCATCCGCCGGCGTGAAACAAGGATGCCCC				RRSKRQCAVAWLDL
	CTCAGCCCCATCATCTTCAACCTGGCCATGGCAACCGCTCATCCGAGCCAT				TNAFGSIPHHHIFATL

		CTCCAGCGGCCCGACCGGCTTCGACCTGCACGGCAAGAAAATCAGCATT				GEFGMPETFIQILRDL
		CTGGCCTACGCGGACGATCTGGCCCTGGTCGCCGACAGCTCGGAAAGC				YKDCTTTIRATDGET
		CTCCAGCAAATGCTCGACGTCACCAGCCAAGCCGCCGAGTGGATGGGC				DAIPIRRGVKQGCPLS
		CTCCGCTTCAACCCCAAAAAGTGTGCCTCCCTCCACGTCGACGGTGGCG				PIIFNLAMEPLIRAISS
		CCAGGGCGCTGGTCCGGCCATCACGATTCCTGATCCAGGGCGAGCCCAT				GPTGFDLHGKKISILA
		GGCCTCCCTCGAAGAGGGAGAGGTATACCAACACCTCGGCACACCCAC				YADDLALVADSSESL
		AGGAGTCCGCGTCCGACAGACCCCCGAAGACACCATCGCGGAGATCCT				QQMLDVTSQAAEW
		GCGAGACGCGGCCCAAATCGACTCCTCCCTGCTCGCCCCCTGGCAAAAG				MGLRFNPKKCASLHV
		ATCAACGCCCTCAATACCTTCCTGATCCCCCGCATCTCCTTTGTCCTCAGG				DGGARALVRPSRFLI
		GGATCGGCCGTAGCCAAGGTGCCCCTGAACAAGGCCGACAGCACCATC				QGEPMASLEEGEVY
		AGGCAGCTGGTGAAGAAGTGGCTCTACCTTCCCCAGAGGGCCAGCACG				QHLGTPTGVRVRQT
		GACATCATCTACATTTCCCACAGGCAGGGCGGCGCCAACGTACCTCGGA				PEDTIAEILRDAAQID
		TGGGTGACCTGTGCGACGTGGCGGTGATGACCCACGCCTTCCGCCTCCT				SSLLAPWQKINALNT
		GACGTGCCCGGACCCGACGGTGAGGAGCATCGCGCAGGAAGCCGTAC				FLIPRISFVLRGSAVAK
		GGGACGTGGTCAGGAAACGCATCGCCAGGGCCCCCTCCGAGCAGGACA				VPLNKADSTIRQLVK
		TCGCCACTTACCTCAGCGGCTCCCTGGAGGCTGAGTTCGGGAGAGAGG				KWLYLPQRASTDIIYI
		GGGGAGACCTGTCCTCTCTCTGGTCCCGCGCCCGCAACGCCTCGAGACG				SHRQGGANVPRMG
		CCTGGGTAAGAGGATCGGCTGCTGCTGGAAGTGGTGCGAGGAGCGCC				DLCDVAVMTHAFRLL
		GGGAGCTGGGAATACTGGTGCCACGCATAAAGACCCCGGACCACACCA				TCPDPTVRSIAQEAV
		TCGTCACCCCGACCGCCAGAGCTATGCTGGAAAGGACCCTGAAAGACG				RDVVRKRIARAPSEQ
		CCATCCGCTGCCACTATGCCGAGAACCTCAAGCGGAAGCCGGACCAGG				DIATYLSGSLEAEFGR
		GCAAGGTGTTCGAGGTGTCCAGCAAGTGGGACGCCAGCAACCACTTCC				EGGDLSSLWSRARN
		TCCCCGGGGGCAGCTTCACCAGGTTCGCCGACTGGCGGTTCGTCCACAG				ASRRLGKRIGCCWK
		GGCCCGACTCAACTGCGTTCCCCTCAACGGAGCCATCCGCCACGGCAAC				WCEERRELGILVPRIK
		CGGGACAAGCGCTGCAGGAAGTGCGGCTACGCAAACGAGACCCTGCCC				TPDHTIVTPTARAML
		CACGTCCTGTGTGGATGCAAACAGCACTCCGGAGCCTGGCGGCACCGC				ERTLKDAIRCHYAENL
		CACAACGCCATCCAGAACCGGCTGGTGAAAGCCATCCCGCCGTCCCTGG				KRKPDQGKVFEVSSK
		GGAAGATCACCCTCGACTCCGCCATCCCCGGGACAGACAGCAGACTGC				WDASNHFLPGGSFT
		GACCCGACATCGTCGTGACGGACGCAGAAAAGAAGAAGGTCCTCATGG				RFADWRFVHRARLN
		TAGACGTCACGGTGCCTTTTGAAAACAGGTCACCGGCCTTCCACGAGGC				CVPLNGAIRHGNRDK
		CCGAGCACGGAAGGCGTTGAAGTACACCCCGCTGGCCGAGACCCTGAG				RCRKCGYANETLPHV
		AGCCCAGGGCTACGAGGTCCAGATACACGCCCTGATCGTGGGAGCCCT				LCGCKQHSGAWRHR
		GGGCTCGTGGGACCCCCACAACGAGCCGGTTCTGAGAGCGTGCGGAGT				HNAIQNRLVKAIPPSL
		CGGTCGACGCTACGCCCGGCTCATGAGACAGCTCATGGTGTCCGACACC				GKITLDSAIPGTDSRL
		ATCAGGTGGTCCAGAGACATTTATACGGAACACATCACAGGACACCGTC				RPDIVVTDAEKKKVL
		AATACCACACTGAGTAACCGAGACCGCCGACCAGGGAAATAACCCACTT				MVDVTVPFENRSPA
		CCTTCCCTGACGAACCAAGGGACGCACCCCACCCATGTACTTATTCGCTA				FHEARARKALKYTPL
		CACCGATACTGACTTGGACTCCTTATACATTCCATGGGTGGCGTACCCGA				AETLRAQGYEVQIHA

			GCCCACTTATCCACTGACACTTTAAAAACTCTTGCACCCCAATCTGGGTC TATGCCGGTTATGCGATATGTATGTATCTCTTCATCCTTGCAACCGATAC CTGTAATCCCTCATAACCCAAGCCTGACCCCAGATGTACAGTACCTTCCC TCTTAACTCGTGTATATTTAATTTTAAACATTAACTTTAATAAAATTTTTA AA (SEQ ID NO: 1055)				LIVGALGSWDPHNEP VLRACGVGRRYARL MRQLMVSDTIRWSR DIYTEHITGHRQYHTE (SEQ ID NO: 1423)
NeSL	Utopia-1_ACa	.	Acanthamoeba castellanii	CCCGTCAAGGGTGCTCCACGAGATCCCTGTCGCTAGCCGACCGGTTTTA CCACCCCACCCCGCCCGGACAACCACGGACCCTGCTCCGCAGCAGGACC CCACGCACGATGGCCGCTAAATCCGTCGCCTGCCCTCACGATGGATGCG CCAACAAGTACGCGTCGGAAGCCTCCCTCCGAAGACACATTAAGAACAA ACACGCTACAGATGAGGAAGGAGATGAGACCTCACACTCCTGTCCCCAC TGCCACCGACCTTTCTCCACCGCCCGCGGGCTCAGCGTCCACATTGGCA AATCGCACCGTCAGGCCCCCCCTGAGCCGACGCGCCCCCCCCGGCCCC GGCCCCTGCCGATCCCGGCCTCGATCCCGACCCCGGCCCCACCGTGACG CCCCCCAGCCGTGATGACGAAGACCGCGAGGAACCCGACGACGACCCC GTGGAGATCGCGGACCTAAGCTGCCCTCACTGCGCCCAGGCCCTCCCGT CGGCCCACGGCCTCGCCAACCACCTTCGCGCCTGCAAGGACCACAGGGT CCCCGCCCCTGGAGCACCCCGCTCGGGTCCGCCCAGCTCCAGGTACTGG ACTGCTGTCGAGCACCACCGCTATGTGGAGGCCATGGCGCGCTTCGCG GATCACCCCGACCTACTTGCGCGCGCGGCTGCCCACATCGGGACCCGCA CGTACAAACAGGTTGACTCCCACCGCACCAAGGTGATCGCGGCGGAGC GCGAGGGCCGCCCTGTCCGCACGCTCGACCCCACGATGGACTGGCGCA TGCGGCCCTACTGCGCCAGCACCACGGCCCGGTGGCTGGCTGAGCAGG GGCGTAGCCCAGTAGCGCCCCGCTCGCCCTGCCCCGAGCCCCACGCCCC GCCGCCTGCAGCCGCGCTGCTGTACATCCCGGCCACGCCCCCCGCGCCA ACGCCCCGTGCCCCAGTGGCGCCTCCCAAGCTTGCGCCTCCCGCCGAGA GCACCGTGCCCGCCACGCCCGATGGGAATCCGGAGGCGCCAGCACCCC CGTTTAGCGCCCCCGGACCTCCCACCCCCAAGGCATTGCCGCCCCCGCCC CCGTCCCGCCGCAACCTGCGCCCTCACCTCGTGCCCAAGGATGCTTGGC AGGGGGTCGCCGATGCCGTCGCCCCTGCCGCCTCGCGCCTCCTGCGCAC GCCCCTTGCGCACCTCTCCACCGAGCAGTGGGCCACGTTCGAAGCCGCC CTCGCCGGCCTCGAGGCTACGCTCCACCATGCCGCCCGCAGTGCAGAGG CGGTGCCCACACGCTGCGCTAGCCGAGCAAGGGAAGACGCCGAGCGCC AACTCCGTGAAGCCCGAAAGACGCGTGAGATCTTTGGCAAGGCCGCTG CCCTCTACGCAGCCGGCAAGGACCCCACTGCCACCATCGAGCGCATCCC CCCAGAAGTCCGCCTACACCTGCCAACCCCTGGCTCGGCTGAATGGCCC GCCAGGGCGGCCGCCCGCCCGCAGGGTGATCCGCCGTGCAGTCGCGCGA GCGGACCGGTTGCGCAAGCGCATGGGCATCCTCGATAGCGACCGCGAC CTCCAACGCCTCTTCAACGCTAACCAGAAGAAGGCAGTTCGGCAGATCC	CCCGTCAA GGGTGCTC CACGAGAT CCCTGTCG CTAGCCGA CCGGTTTT ACCACCCC ACCCCGCC CGGACAAC CACGGACC CTGCTCCG CAGCAGG ACCCCACG CACG (SEQ ID NO: 1179)	TAACAA CCATGT ATGGTG AACCAC ACCTCTC TCGATCT TGTATTC TGTGATT GGACAT CAGAGT TCCTGC GAAGGG ATACACT CTGCCA ATCTCGT GGGTTG TAATAA ATCCAC ACCTTCA ACA (SEQ ID NO: 1302)	MAAKSVACPHDGCA NKYASEASLRRHIKNK HATDEEGDETSHSCP HCHRPFSTARGLSVH IGKSHRQAPPEPTRP PPAPAPADPGLDPDP GPTVTPPSRDDEDRE EPDDDPVEIADLSCP HCAQALPSAHGLAN HLRACKDHRVPAPG APRSGPPSSRYWTAV EHHRYVEAMARFAD HPDLLARAAAHIGTR TYKQVDSHRTKVIAA EREGRPVRTLDPTM DWRMRPYCASTTAR WLAEQGRSPVAPRS PCPEPHAPPPAAALL YIPATPPAPTPRAPVA PPKLAPPAESTVPATP DGNPEAPAPPFSAPG PPTPKALPPPPPSRR NLRPHLVPKDAWQG VADAVAPAASRLLRT PLAHLSTEQWATFEA ALAGLEATLHHAARS AEAVPTRCASRARED AERQLREARKTREIFG KAAALYAAGKDPTAT IERIPPEVRLHLPTPGS AEWPARAAAARRVI RRAVARADRLRKRM GILDSDRDLQRLFNA

			TCGCCCCGTCCACCAAGGCGCCGCGGTGCCAGCTAGACCCAGCCGCCGT			NQKKAVRQILAPSTK
			CGAGGAGGCCTACATCCAGACCCTCGCCAAGCCGCCGCCGATCGACCCC			APRCQLDPAAVEEAY
			AGCCCCCCGTGGAAGAACTCCGTCCAGTGGCCCCGCCCGCCCACTGCCG			IQTLAKPPPIDPSPPW
			CCGATGACGGAGGCAGCCCCTTCAGCGTCGCCGAGGTCCGGGCCCAGC			KNSVQWPRPPTAAD
			TCCGCCGACTGCCCAACGGGTCCGCCCCAGGGATCGATGGCATACCGTA			DGGSPFSVAEVRAQL
			CGAGGCCTACAAGCGTACGAAACTGGACGCCACGCTCGCCCATGTCTTC			RRLPNGSAPGIDGIPY
			GAGGTCGTGCGGCTGAATGCGCGCCTGCCAGCTCGATGGGATGTGGCG			EAYKRTKLDATLAHV
			CGCACGGTCCTGCTCTACAAGAAAGGCGACCCTAACGACACCGGCAACT			FEVVRLNARLPARW
			GGCGACCGATAAGCCTCCAGGTCACCATCTATAAGATCTTCACGGCCGC			DVARTVLLYKKGDPN
			CCTGTCGAAGCGGCTCATCTCCTGGGCTGGCAAGCACAACACTTTCTCC			DTGNWRPISLQVTIY
			GCATCGCAGAAGGGATTCCTACCGGCCGAAGGCTGCCACGAGCACGCG			KIFTAALSKRLISWAG
			TTTGTCTTGCGAAGCGTGCTTGACGACGCCCGTCGGCACAAGCAGAACG			KHNTFSASQKGFLPA
			TGTACCTTGCCTGGTACGATCTGCGCAACGCCTTCGGATCGGTGTCGCA			EGCHEHAFVLRSVLD
			CGACCTCATCGCCTGGTGCGCTGCCATGTTGGGCCTGCCCCGCTACCTCC			DARRHKQNVYLAWY
			GGGATGCCATCGGCGCAATCTATCGGCACTCAGCGCTCTTCGTCCAAGT			DLRNAFGSVSHDLIA
			TGGGGATCAGGAGACCACCGGCGTCATTCCTATGCGCTGCGGCGTCAA			WCAAMLGLPRYLRD
			GCAGGGCTGCCCTCTCAGCCCCCCTCCTCTTCAACCTGTGCGTCGAGCCG			AIGAIYRHSALFVQV
			GCCCTTCGCTGCCTACGCCGCACCACCGGGTACAAGTTCTACGGCACGT			GDQETTGVIPMRCG
			CGATCACCGTCGAGGGCCAGGCCTACGCCGACGACCTGCTCACTGCCGC			VKQGCPLSPLLFNLC
			GCCCTCCGCCTACCATGCGGCCCGGCAGGTGGCCACGATCGAGGAATG			VEPALRCLRRTTGYKF
			GGCCAACTGGGCGGGAGTCTCCTTCGTCGTCCAAGCCCTCTCCCTGGAT			YGTSITVEGQAYADD
			GCGCCGGCCGGCAAGTGTGCCGCCCTCGCGATCAACTTCGAAGGTGGT			LLTAAPSAYHAARQV
			CTAATGCACTCTATCGACCCTGCCCTCAAGGTCCAAGGCGCAGCCATCCC			ATIEEWANWAGVSF
			GGCCATGTCAAGAAACAACGTGTACCGCTACCTCGGAGTACATGTCGGT			VVQALSLDAPAGKCA
			CTCACAGATGCGCTCGGCCAAGCGAACGAGCTCCTCGAGAAGGCCTCA			ALAINFEGGLMHSID
			CGCGATGCACGCACGATCTGTGCCTCTGGCCTCGAACCCTGGCAGAAGG			PALKVQGAAIPAMSR
			TGGTCGCAATCAAGACCTTCATCCTCTCCCGGCTCCCCTTCTTCTTCACA			NNVYRYLGVHVGLT
			ACGGGAAGATCCAGAGGGGCCGATGCCAGCAATTCGACCGCGAGCTTC			DALGQANELLEKASR
			GAGAAAAACCTGCGGGCCGCCCTCCGACTCCCCGTCTGCACCACGAACGC			DARTICASGLEPWQK
			CTTCTTCCATTCCCGCGTGGCCTCAGGCGGCCTTGGCATCCTGCCCATCG			VVAIKTFILSRLPFFFH
			CGGAAGAACAACAAGTCTACCTGGCAGCCCACGTGTTCAAGCTCCTGAC			NGKIQRGRCQQFDR
			TTCGCCAGATCTGTCGATCCGCGCCATCGCCCGACACCAACTTGCCGAG			ELRENLRAALRLPVCT
			GTCACCCACGCCGCGACACACCACGCCAGTCCAGGACGGCGAAGCGTCA			TNAFFHSRVASGGLG
			CCCTTCTTCGGATGGCTCATGCGGGGGGCAGGAGGTCGCATCAACTACCC			ILPIAEEQQVYLAAHV
			CCTCGGGTGACGTCAGTTCAATCTGGTTCGCAGCTGCAGGCGCCTACTC			FKLLTSPDLSIRAIARH
			GAGGATGGGATGGTCAGTCCGCGATGCACTCCACCCGACGCTGACAGT			QLAEVTHARHTTPV
			TGGTCCGGGCGTCCAATTCGAGGGCCGATTCCAACGTGCCAACGTCATC			QDGEASPFFGWLMR
			CCAGCTCTCCGGGCTAGCGCCTTTTCCCGCCATGCTGTGGAATGGAGTG			GQEVASTTPSGDVSS

			CCCTCCGCACCCAGGGTCGAGCAGCAGCCTACCAACATGCCGTCCACCC				IWFAAAGAYSRMG
			TGCAACGCACCACTGGGTCCACAACAGCGCTGGCCTGACGACCAAGGA				WSVRDALHPTLTVG
			GTACCGATTCGCGATCAAGTGTCGATTGGGTCTCCTGCCGACGCGAGCA				PGVQFEGRFQRANVI
			GCTCCACACCACCGCAATGGGCCAACAGCGTGCAGGGCGTGCTCCTAC				PALRASAFSRHAVEW
			GCCCGCGAGACGGCCAACCATGTTCTCGGACACTGCCCGGCGACCAAG				SALRTQGRAAAYQH
			GCCGAAGTCATCGCGCGCCACAACAGGATATGCCGAGCTCTGGCCCAG				AVHPATHHWVHNS
			GCGGCTGAAGCCTCATGGACGTCTGTCCTTGAAGACGTCCCGATCCCGG				AGLTTKEYRFAIKCRL
			GGGTGGACTCCCCCCTACGACCCGACATCTACTGCTCTCGGCCGGGCCA				GLLPTRAAPHHRNGP
			GTGTGCCATCATCGAGGTCGCGGTCTCCTACGAGGACGCCTTCAACGCT				TACRACSYARETANH
			TCGATGGAGGGCCGGGCGAAGCAGAAGACCGACAAGTACGCTGGCCT				VLGHCPATKAEVIAR
			GGCTGCTACCGTCGAGGAGCAGCTGCGGCTCCAAACCCGGCACGCGGC				HNRICRALAQAAEAS
			TTTCGTGGTGGGCTTCTCTGGCGTCGTGCTCCCAGCCTCGGTAACCGCTA				WTSVLEDVPIPGVDS
			CGGCCACCTCCCTTGATCTCCCCCCCCAAAACTTGGAATGTGCTTCTTAAA				PLRPDIYCSRPGQCAI
			CGTTGTGTTGCTGCCTCAATCAAAGGCAGTTACACAGCGTGGAGAAGAT				IEVAVSYEDAFNASM
			TCCGGCGCTCTACTCCATAACAACCATGTATGGTGAACCACACACCTCTCTC				EGRAKQKTDKYAGLA
			GATCTTGTATTCTGTGATTGGACATCAGAGTTCCTGCGAAGGGATACAC				ATVEEQLRLQTRHAA
			TCTGCCAATCTCGTGGGTTGTAATAAATCCACACCTTCAACA (SEQ ID NO: 1056)				FVVGFSGVVLPASVT
							ATATSLDLPPKTWNV
							LLKRCVAASIKGSYTA
							WRRFRRSTP (SEQ ID NO: 1424)
NeSL	Utopia-1_AEc	.	Acromyrmex echinatior	GGTGCACAACGGATGCATCATACGTGTACCGGAGCATACGGGCTGTCA	GGTGCACA	TAAATTA	VCSVRGCRREDSRRF
				CGGCGGCTGCATGCGCGATCTAGCTCGGAGATTTTATTTATTTATTTATT	ACGGATGC	TTTTGTC	YKFKFPLNFVKVPKTI
				AATTTATTTATTTATTCATCGAGTGTGAGTGTTCGCGTTTTGCCGAGAAG	ATCATACG	TTTGTCT	VIGSAFQKSSVSARS
				CGATTTTCGTTAAGTGATACGCGCGCCGCGTTCATAGGTTAGGTGTGCAGT	TGTACCGG	TGGCCC	QNHSRSTRVPKTRQP
				GTACGTGGCTGTCGTCGCGAAGATTCTCGAAGATTCTACAAGTTCAAAT	AGCATACG	CCCCTTT	RTSNTIGRYTAASAN
				TTCCGTTAAATTTCGTTAAAGTTCCTAAAACCATCGTGATCGGGAGTGCG	GGCTGTCA	TTAAACC	NYLTVIITGNYTVFAQ
				TTCCAAAAATCATCAGTTTCGGCTCGTTCGCAAAACCACTCCCGGTCGAC	CGGCGGCT	AAGCAG	WICYRECTWLLSKFV
				CCGCGTTCCGAAAACCAGACAGCCACGTACAAGCAATACGATTGGTCGG	GCATGCGC	GAGAGA	NFFLTIIGYFFQLRLVV
				TACACTGCCGCGAGTGCGAATAACTATTTAACTGTTATTATTACTGGAAA	GATCTAGC	GTGGCC	IYEGPVILDTFSNCGS
				TTACACTGTATTCGCCCAGTGGATCTGTTATCGCGAGTGCACTTGGCTTT	TCGGAGAT	CAATGC	SLFMRGQXSKALLVR
				TAAGTAAATTTGTTAATTTTTTCTTAACTATAATTGGATACTTTTTCCAATT	TTTATTTAT	CCAACT	LNRSALAMADPQVH
				GCGGCTCGTCGTTATTTATGAGGGGCCAGTKATCTAAGGCCCTTTTAGTC	TTATTTATT	ATTATAT	YIDYPLPPRVKCVKCF
				AGACTAAACCGTAGTGCACTCGCCATGGCGGATCCACAAGTGCACTATA	AATTTATTT	ATTAACT	GAEGAGKVKGEYSD
				TAGACTACCCGCTGCCCCCTAGAGTCAAATGCGTAAAATGTTTCGGTGC	ATTTATTC	ATTTACT	PPHLAKHLKKCHPGD
				TGAGGGGGCAGGCAAAGTAAAGGGCGAATACAGCGACCCGCCGCATTT	ATCGAGTG	GTGATA	TLNYKCSICDLRGTGK
				AGCAAAACATCTGAAAAAGTGCCACCCGGGAGACACATTAAATTATAAA	TGAGTGTT	TTTATTA	YPLRDVKAHYAECHV
				TGCTCAATTTGTGATCTAAGGGGGACCGGTAAATACCCCCTTAGAGATG	CGCGTTTT	TTTGACT	SPAVDAAGPSTRGSL

	TTAAGGCACATTACGCCGAGTGCCATGTGTCTCCCGCAGTGGATGCGGC	GCCGAGA	GTTGGG		GECSGAGQPTASRA
	GGGTCCAAGCACTCGCGGCAGCCTCGGCGAGTGCAGTGGTGCGGGTCA	AGCGATTT	CGGGCC		AKATTRLAETVGGTD
	ACCGACAGCCAGCCGCGCGGCTAAAGCGACCACGCGATTGGCGGAGAC	TCGTTAAG	CCTCTCT		KRRAATSGSRQLTLP
	GGTTGGGGGTACGGATAAGCGCCGTGCCGCGACATCGGGATCGCGGC	TGATACGC	GCTGGT		FAATPSPSTAAGEAR
	AACTCACGCTGCCGTTCGCAGCCACCCCATCGCCATCCACAGCAGCCGG	GCCGCGTT	TTTATTT		APRSXSTTPTSRSPSY
	TGAGGCAAGGGCCCCAAGAAGCGMGTCAACGACACCGACGAGCAGGT	CATAGGTT	ATATATA		AAVTAGPPSMRSTTT
	CCCCCTCATATGCGGCAGTCACTGCGGGCCCGCCATCGATGAGGAGCAC	AG (SEQ ID	TTTTTTA		STTARSKTVAKGAAP
	GACAACTTCCACCACAGCCCGCAGTAAGACTGTCGCGAAAGGCGCCGC	NO: 1180)	CTCGCG		NTTTTTTARRSGEAA
	GCCCAACACMACGACGACAACGGCCAGGAGATCCGGCGAGGCCG		TACTTTT		ATRKPPTTATVSKPR
	CCGCAACGAGGAAGCCGCCTACGACCGCCACGGTGAGTAAACCGCGTG		TGTACTA		VLSVETVRLPVDDIQ
	TGTTGTCGGTAGAAACTGTTAGGTTGCCTGTCGACGACATCCAGCGAGC		CTCTATT		RAGVQNAAKPARAP
	AGGCGTGCAAAACGCGGCCAAACCGGCGCGCTCCCTCTCGCCCCCC		TTTCTTT		SRPPQRTSPEAGGPR
	GCAGAGAACATCACCGGAGGCGGGGGGTCCAAGAACAACGGGCGCAA		TTATTTT		TTGAKEKCGEGAYKK
	AGGAGAAATGCGGAGAGGGAGCATACAAGAAGTTGCCCGCAAACAGC		AGCTAT		LPANSGNPISTRTRR
	GGCAATCCAATTTCAACCAGGACGAGGCGGGCAACTAGCGTGCCGGTT		GCTATTT		ATSVPVEKSEGTARR
	GAGAAGAGCGAAGGCACGGCAAGACGGGAGCGCGTCTCCCCACACCCT		TTATCTC		ERVSPHPPPKGIDIILS
	CCTCCCAAAGGAATTGATATCATCCTATCGTCGACATCGGAGGAGGAGG		TTTCTTT		STSEEEGTPYQPGGV
	GCACGCCATACCAGCCCGGCGGCGTGGGGAGACTAAGACTAAGGAGG		GTCTCTA		GRLRLRRKKVTGPPP
	AAAAAGGTGACCGGACCACCCCCAAAGATGACACCCAGAGAGGGGGT		TTTTCTT		KMTPREGVVTRARR
	GGTCACAAGAGCCAGGCGGTCCACCAGCGCTCCCGTCGAGAAGAGTGC		TCTTTTT		STSAPVEKSALDARLT
	CTTGGATGCACGCCTGACGGCTCTGGACCGGACATCGTCCAGAGCGAC		TTCTTTC		ALDRTSSRATGNPTS
	AGGCAACCCGACGTCGCAAATCGCAGGGGGGCCTTTACACCAGTAGAGG		CTTTTCT		QIAGGLYTSRGQPER
	CCAACCGGAGAGGACGCCCCCTGCGAGGCTCCCCAGCCTGTCTCCGACC		TTTCTTT		TPPARLPSLSPTTRGS
	ACCAGAGGCAGTCCATCGGGGAGCCTAGGCGAGATACGGACACCCATC		CTTTTAT		PSGSLGEIRTPISPATS
	TCGCCTGCGACGTCGCTACCGGCAACGCTCACCACTTGCACGGTGACGA		TCTTCTT		LPATLTTCTVTTTTCG
	CGACCACCTGTGGAAGCCCCATAACATCCACGGGCTTCACAGGTGGCGT		TTATTTA		SPITSTGFTGGVGRLI
	GGGGAGGCTGATAACACCWCCGAGCCTCCCCCAAACGAACATCCTCCC		TCTTTTT		TPPSLPQTNILPTIGEE
	GACCATCGGGGAGGAAGGAACGTCACCGTGCGTGGCGGTCGTCACCAC		TTCTTTT		GTSPCVAVVTTHPRP
	CCATCCTTAGGCGGCAGGAGCACGTCTCCGCTCATCCTACCAAGGCCAA		CTGTTGT		TGEDAPCEAPQPVSD
	CGACACCGGAGCCTGAGCGGGGACAAGAGGAGCGGCGGCTAGAAGGC		GGGGCC		HQRQSIGEPRRDTDT
	GCGGCGCAGCCACCTACCACACCCGTCGTCGAGGGGGACAACCAGTGG		CTGACC		HLACDVATGNAHHL
	GATGGCCAGTGGACGGTGAGCGTGAGGAGAAGAGCGAGGAGGCAACA		GTCCGA		HGDDDHLWKPHNIH
	ACTGAACGATACATCCCCCTCCAACTCCGAGTCCCCGCCAACCGCTGGA		GTGTGA		GLHRWRGEADNTXE
	CCATCGCGTTCGCCGCGCATAGCCCCACTATCTGCGCTGATAGCGGCGT		ATGCCG		PPPNEHPPDHRGGR
	CGACGAGCCGCCATGAGACCAGCTTAAATCTCAATTGCACGAACGGCAA		CGAAAA		NVTVRGGRHHPSLG
	TATTTGCATGGACCGAACTCCGCCCCGTAACATTTTGCCGGTARGGGCG		ACAATA		GRSTSPLILPRPTTPEP
	GAGCGCCGTCGCGAGACATCGCCACAGGATCGCGTGGAGGGAGACATC		TTATGTT		ERGQEERRLEGAAQP
	GGTTATGGTGCTGGAAAGGTAAGTGCCGAACACCCGAGTGCTCCCGTA		TTATACG		PTTPVVEGDNQWD

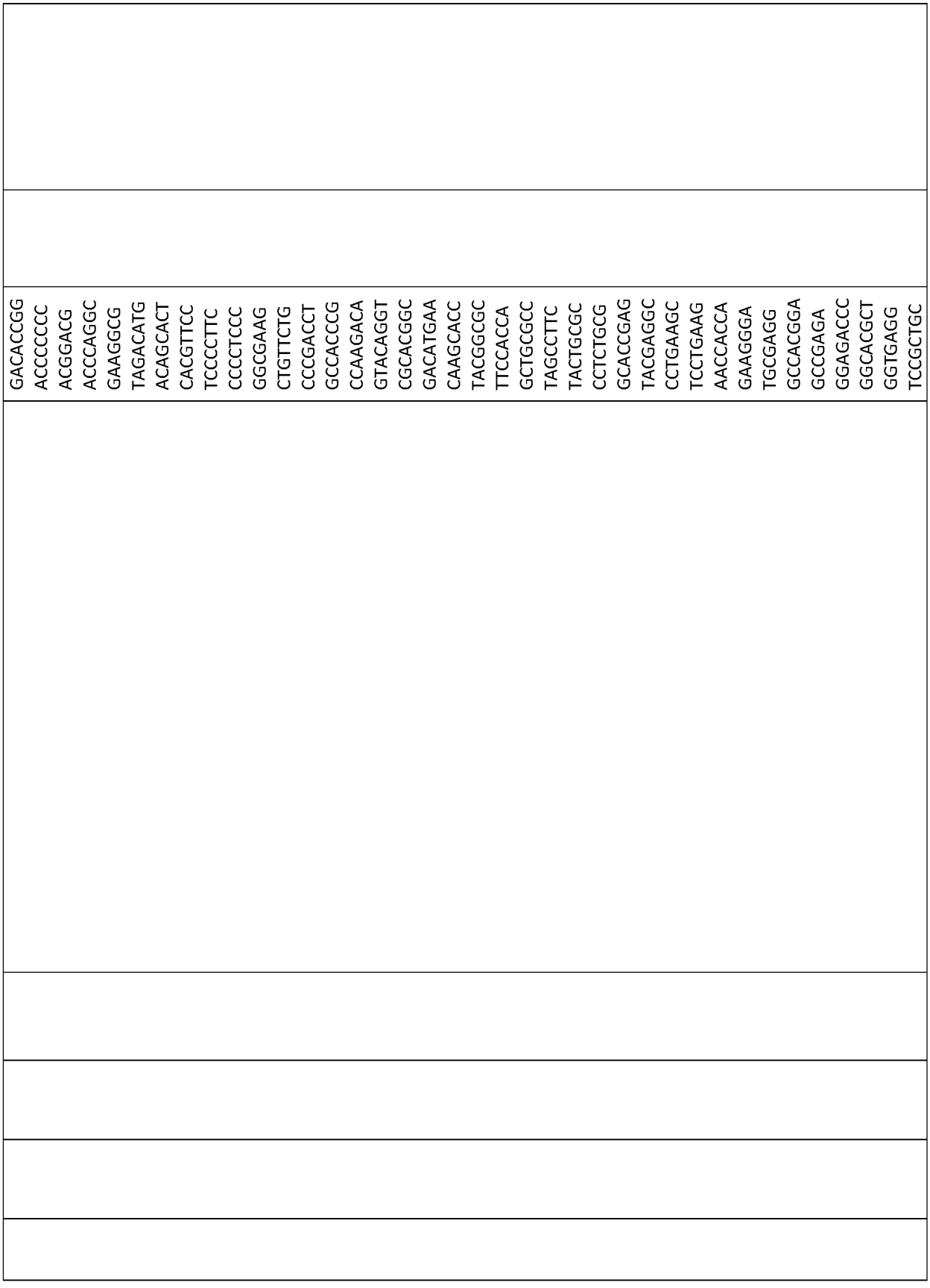
	AATGTCCGTGGTGTGATGTCTCGAGGGAGAGCAACCGCGTCATCCATCG	AGTGTG	GQWTVSVRRRARRQ
	TGCCACCGCGAGCCAACCGTGGGGAGGGCGGTCGGCAGCATCATAGTC	CATGTG	QLNDTSPSNSESPPT
	GGCGGCGTCCGGACGCTCCTGTCGGTCAGCCGTCGCGGGATCACCCGG	CGTGAT	AGPSRSPRIAPLSALI
	CGCCTGCGACTGTCGCGAGGCAGCGTAGGCGTGAGCGGGTGGCCGCCC	ATATTTA	AASTSRHETSLNLNC
	GCGACGCGCTGCTCGATCGGGCTAAGGACGTCGCTACGATTGCGGATC	TCTATTT	TNGNICMDRTPPRNI
	TGGAGGCGTTCGCGGCTTCGGTCGCGGCGTTCTTCGGGGAGGATGCAT	TATTTTA	LPVXAERRRETSPQD
	CGGCCACTGGTGCTGCAGCCCGCGCTCGCGATCGTTCGGTACGCTCACG	TTTATTT	RVEGDIGYGAGKVSA
	GGAGGCGGGTGCGTCGGGGGGTGAAGGGAGGTGAGCGTCCGGAG	ATTATAA	EHPSAPVNVRGVMS
	AGAGAGGGCGCCGGTAGGCCGGGTCAGCGCCGGCTGACCCCGGAGC	TTTATTG	RGRATASSIVPPRAN
	GTCGGGAGAAGCACGCGGGGACTGGGTGCGCGAGGCCAAACGCTTGC	CCGCGC	RGEGGRQHHSRRRP
	AGGCGCTGTACAGGGCGAACCGCCGCAAGGCAGTGCGAGAGGTGCTC	GCGCTC	DAPVGQPSRDHPAP
	CAGGGACCTGCCGATCAGTGCCAGGTGCCTAAACGTCAGGTCCAGGAG	CTCCGG	ATVARQRRRERVAA
	TACTTCGAGCGGCTGTACAGCGGCGGGGAAGACCTGGCTGGCGCCGGC	GACTTTT	RDALLDRAKDVATIA
	GTGGAAGCCGAACGCCCTGACCCCTCGAGTCCGCGTGAGGTATCTGCG	ATTCGTT	DLEAFAASVAAFFGE
	GTCCTGGGTCCGCTCGCGGAGCGAGAGGTGGACCGTCGGCTCCGGCGT	GACAAT	DASATGAAARARDR
	ATGAATAACTCTGCGCCCGGGTCCCGACGGTGTATCCTATCGTGACCTCC	ACTGTG	SVRSREAGARRGVKG
	GTGGGGCGGACCGGGGAGCGCGGCTCCTCACGGCGCTCTACAACATCT	ATATTTT	GERPEREGAGRPGSA
	GCCTGCGGCTCGAGGCAGTCCCCGCGTCCTGGAAGACCTCCAACACTGT	TCTGCKC	PADPGASGEARGDW
	GTTGATACACAAGAAAGGAGACCGGGGCATGTTGGAGAACTGGCGCCC	AGGCTG	VREAKRLQALYRANR
	TCTCGCTCTGGGGGACACCGTCCCCAAACTCTTCGCCGCGCTCTTGGCCG	GGGGGG	RKAVREVLQGPADQ
	ACCGATTGACCGACTGGGCGGTCACCCGCGGGAAGCTCTGCTCCGCGC	CTTGCCC	CQVPKRQVQEYFERL
	AGAAGGGCTTCCTGCGGGACGAGGGGTGCTACGAGCACAACTTCGTCC	CCCAGC	YSGGEDLAGAGVEAE
	TGCAGGAGGTCCTGACGCACGCCAAGCGCTCTAAGCGCCAGGCGGTCG	CCCTTAG	RPDPSSPREVSAVLG
	TCGCGTGGCTGGACCTGTCCAACGCGTTTGGATCGATCCCGCACGCGAC	TTTTAAT	PLAEREVDRRLRRM
	GATCCGCCGCGCGCTTATAAGATCCGCGGTGCCACGGGGTCTCATAGCG	TGCCTAT	NNSAPGPDGVSYRD
	ATATGGGACTCCATGTACGATGGTTGCACGACGAGGGTGCGAACCGCC	GCGGGG	LRGADRGARLLTALY
	GAGGGTCACACAGCACCCATCCCCATCCGGTCGGGCGTCCGTCAGGGTT	GGGGCT	NICLRLEAVPASWKT
	GTCCGCTAAGCCCTATTATCTTCAACCTGGCCATCGACTCGGTCGTCCGT	TTTGTCC	SNTVLIHKKGDRGML
	GTGGCGGCCGAGWCGAATGACGGGTATTCCCTCCACGGAAATACCTGG	CCCGCA	ENWRPLALGDTVPKL
	TCGGCATTGGCTTACGCGGACGACATCGCACTACTGGCCCAGACGCCCG	AATGTA	FAALLADRLTDWAVT
	AGGGGATGGAGAGGATGCTAGCCTCTGTGGAGGCGGAGGCAGCGTCG	TATATAT	RGKLCSAQKGFLRDE
	GTGGGGCTGCGGTTCAACCCTGCAAAGTGTGCCACCCTGCACGTCGGTG	ATATATT	GCYEHNFVLQEVLTH
	CGGGGAATGGCGGCAGGGTCCTACCGACGTCATTCCAAATCCAGGGGG	TAGCGC	AKRSKRQAVVAWLD
	AGACGATCAACCCCCTGGCTCAGGGTGAGTCGTACACCCACCTTGGCGT	GCGGCT	LSNAFGSIPHATIRRA
	TCCAACGGGGTTCTCCGTGGACCAGACGCCCTACGCCGCCGTCGGGGA	TAGCCG	LIRSAVPRGLIAIWDS
	CATCGTCTCGGACCTGCCGCTGTCGACCGCTCACTCCTTGCCCCGTGGC	CTTTTGT	MYDGCTTRVRTAEG
	AGAAGATAGAAATGCTGGGGACCTTCATCCTATCCAGGCTTGACTTTCT	TTGTATT	HTAPIPIRSGVRQGC
	GCTCCGGGGGGCCAGAGTGTTCAAGGGTCCCCTCACGGCCGTGGACCT	ACCCCA	PLSPIIFNLAIDSVVRV

		TAACATCCGWAGGCATGTTAAATCCTGGCTTAACCTCCCTCAGCGAGCA	GAGGGG		AAEXNDGYSLHGNT
		AGCGCGGAGGGGAGTCTACATGCCGCCCCGTTGGGGGGGATGTGGACTC	AATTGTC		WSALAYADDIALLAQ
		CTGCCGCTCTCTGACCTCGCCGACGTCCTCACGGTTGCCCACGCGTACCG	CCTCTG		TPEGMERMLASVEA
		TATGTTAACGGTGCGCGCATGGCGCCCGTGAGGGAGTTGGCGTGGGAATC	GGGAAA		EAASVGLRFNPAKCA
		GCTGAGGGGAGTGGTTGGGCGCAGGATCGGCCACGCCCCTAGTTGCGA	AAAAAT		TLHVGAGNGGRVLP
		GGATATCGCCTCCTTCCTATCCGGCTCGCTGGATGGAAGGATGAGGGGC	GATTGG		TSFQIQGETINPLAQ
		GGAGGGGAGGCTTCGCTCTGGTCGAGTGCGCGGAACGCTGCGCTCAGA	AAAAAT		GESYTHLGVPTGFSV
		CAGTCCGAGAGGTTGTCCCTGCGTTGGCGGTGGGTCGAGGCCACGGAG	AAAGTG		DQTPYAAVGDIVSDL
		GAGATGACGTTGGAGTGTCGAGGGCCCAGGGGGGCAGCGATTAAGAT	AGCTAA		RAVDRSLLAPWQKIE
		TCCGCCTGAAGCGCGCGGTCAGGTAGTGAATCGGTTGCGCTCAGCTGTA	(SEQ ID		MLGTFILSRLDFLLRG
		GCAGAGCACTACGCAAGTAGGTTGCTTAGCAAGCCTGATCAGGGTAAG	NO:		ARVFKGPLTAVDLNI
		GTCTTCGAGGTGTCGTCGCGGAGCCGAGTGAGCAATCACTCACTTTATCCGCG	1303)		RRHVKSWLNLPQRA
		GCGGCAGCTTCACTCGCTTCGCCGACTGGCGCTTTATCCATAAGGCCCG			SAEGVYMPPRWGG
		GTTAGATGTTCTTCCTCTCAACGGCGCACGACGTTGGGAGGCCAACGAC			CGLLPLSDLADVLTVA
		AAGCGCTGTCGGCGATGCGGTGAGGTATCGGAGACATTACCCCATGTG			HAYRMLTVRDGAVR
		CTCTGTCACTGCGGCATCCACTCCGCCGCGATACAGCTGAGGCACGACG			ELAWESLRGVVGRRI
		CTGTCCTGCACCGCCTTTGGAAGGCCACTCGCCTTCCAGGGGTAGTGCG			GHAPSCEDIASFLSGS
		GGTTAACCAGCGGGTGGAGGGCGTCAGCGACGAATTGGGGGCGCTAC			LDGRMRGGGEASL
		GACCTGATCTCGTGGTCAGGCACGAGCCCTCCAAAAGTGTCGTCATCTG			WSSARNAALRQSERL
		CGACGTCACGGTGCCATTCGAAAACCGCTGGACCGCTTTCGAGGACGCC			SLRWRWVEATEEMT
		AGGGCGAGGAAAATCGCCAAATACTCGCCTCTGGCAGAGGAGCTACAG			LECRGPRGAAIKIPPE
		CGGCGAGGGTACCGTGTCGTCGTGACGGCCTTCGTCGGCGCCCTCG			ARGQVVNRLRSAVA
		GCTCGTGGGATCCGAGGAATGAGGCGGTGTTGAGACTGCTGCGGGTTG			EHYASRLLSKPDQGK
		GCAACCAGTATGCAGCTATGATGCGGCGCCTCATTGTCTCGGATACCAT			VFEVSSRSRVSNHFIR
		TCGCTGGTCACGCGACATATATGTGGAGCATGTGTCCGGCACCCGCCAG			GGSFTRFADWRFIHK
		TACCTGGCTCCTTCCCGTCCCTCTGGGGATCTCGCGACGCCGCCGAGAG			ARLDVLPLNGARRW
		CGGTTCGTCGACGCTGGCTCGCCGAGGAGAAGCGCACAGGACGCG			EANDKRCRRCGEVSE
		GCGCGTCGCGGTTCGGATAGTGTGAGTGTCGCGTAAATTATTTTGTCTT			TLPHVLCHCGIHSAAI
		TGTCTTGGCCCCCCCCTTTTTAAACCAAGCAGGAGAGAGTGGCCCAATGC			QLRHDAVLHRLWKA
		CCAACTATTATATATTAACTATTTACTGTGATATTTATTATTTGACTGTTG			TRLPGVVRVNQRVE
		GGCGGGCCCCTCTCTGCTGGTTTTATTTATATATATTTTTTACTCGCGTAC			GVSDELGALRPDLVV
		TTTTTGTACTACTCTATTTTTCTTTTTTATTTTTAGCTATGCTATTTTTTATCTCT			RHEPSKSVVICDVTVP
		TTCTTTGTCTCTATTTTCTTTCTTTTTTTCTTTCCTTTTCTTTTCTTTCTTTTAT			FENRWTAFEDARAR
		TCTTCTTTTATTTATCTTTTTTTTCTTTTCTGTTGTGGGGCCCTGACCGTCCG			KIAKYSPLAEELQRRG
		AGTGTGAATGCCGCGAAAAACAATATTATGTTTTATACGAGTGTGCATG			YRVVVTAFVVGALGS
		TGCGTGATATATTTATCTATTTTATTTTATTTATTATAATTTATTGCC			WDPRNEAVLRLLRV
		GCGCGCGCTCCTCCGGGACTTTTATTCGTTGACAATACTGTGATATTTTT			GNQYAAMMRRLIVS
		CTGCKCAGGCTGGGGGGGGCTTGCCCCCCAGCCCCTTAGTTTTAATTGCCT			DTIRWSRDIYVEHVS

			ATGCGGGGGGGGCTTTTGTCCCCCGCAAATGTATATATATATATATTTA GCGCGCGGCTTAGCCGCTTTTGTTTGTATTACCCCAGAGGGGAATTGTC CCTCTGGGGAAAAAAAAATGATTGGAAAAATAAAGTGAGCTAA (SEQ ID NO: 1057)				GTRQYLAPSRPSGDL ATPPRAVRRRWLAEE RSAQDAARRGSDSV SVA (SEQ ID NO: 1425)
NeSL	Utopia-1_AMi	.	Alligator mississippiensis	TGCTGGAAAGACGGAGAACCGCTTCCTTTTCCCTGCGCCTGGCCTGGT	TGCTGGAA	TGAACC	CHHAGLRPGTPNRT
				ATTGCAGTACCTCCAGGATTAGCGCCAACTAGTCCGGCAGACTGTCGGA	AGACGGA	CCCCCTC	RRPDQTAPLPDPRG
				ATACAGCAATAGAAAGWGAGCTGACTAGCAGCTTGCTTTCCTTCCTCCG	GAACCGCT	TGCACC	HPMPPNRRGSRSRP
				GTGCAGCATGGGTTCTCGTCAGTCMTGACGGGCTAGGGAAGGCGGTG	TCCTTTTTC	AGATGG	EEPSRREPPXPRACQ
				CTGCCAGTACGTCCGAAAGAGTGCCGGTTGCGCAAGCGACCGCGCCAC	CCTGCGCC	ACCTTCA	GLRVWSPPQQRMPT
				TCAGGTGAGTAGCCAAGGGTCTTACAGTTCACCGGACCCGAWAACGCG	TGGCCTGG	CTTCGA	PWQTLWLEELSRATT
				AAAACCCCAACTCGGGCTAGTAGCCGAAGACCTGGGTCCCCCCCCMGGT	TATTGCAG	GAGGAT	FKAFEASVARLTEELS
				CAGAGTAGGCGAACGCCWGKGCTCAGAGGACGGAACGCGGAAAACAC	TACCTCCA	TCTTCAG	AAARPGQPRGGNNR
				CCCCAGGTCCCAAGGACGCCCTGATCCACTGACAAGAACGCTCGAGGCA	GGATTAGC	CAATGG	PATRRDHRLQPQRR
				CGCCAGGAGACCCCCAGCTAGGGTGGACCGCCGACTGCAGGTCCGGAG	GCCAACTA	ACGACC	PRRQRYDPAAASRIQ
				GACCCTCCCAGGAGGGTGGACCAGCGAACCCAAGTTGGCGACGAACCC	GTCCGGCA	CCGCTCC	KLYRANRPKAVREILE
				TGACGCACCCCCCACGATGTCAGGACCCCGACAGGCGGCGGTGGACCA	GACTGTCG	ACCCGA	GPSAFCQVPRETLFN
				CTGACCATCGACCGACCCCCAGAGGCAGAGAGACTCTCAGAGCCCGGA	GAATACAG	AGAGGA	YFSRVFNPPAEAAAP
				ACCCCGGCTGACGAGAGCCGCCTCCCGGCGGAGGACCCCGGAGCCTGA	CAATAGAA	CCCCCG	RPATVEALTPVPPAE
				GGATGCCCCCCGGATGACGGCGGAGCGCCCCGAGCGACAGCGGACCCC	AGWGAGC	CGATGA	GFEDAFTPQEVEARL
				TCCGGACCCCCACGGCCCCTCGGTGACGATGGCGGGCCCCGAACGACG	TGACTAGC	GACTCT	KRTRDTAPGRDGIRY
				ACGACCCCCGGACCCCGGCGGTCCCGAGGACGCCCCCCCCGAGGGTCT	AGCTTGCT	ATATGG	SLLKKRDPGCLVLSVL
				CCCCACGCTGGTGGAGGAGCCCCGGACCCCCCCGACACCGGACCCCCCC	TTCCTTCCT	ACTGAG	FNRCREFRRTPTTWK
				ACGGACGACCCAGGCGAAGGCGTAGACATGACAGCACTCACGTTCCTC	CCGGTGCA	ACACTTT	RAMTVLIHKKGDPTD
				CCCTTCCCCCTCCCGGCGAAGCTGTTCTGCCCGACCTGCCACCCGCCAAG	GCATGGGT	TTCTTCG	PGNWRPIALCSTVAK
				ACAGTACAGGTCGCACGGCGACATGAACAAGCACCTACGGCGCTTCCAC	TCTCGTCA	AACCAC	LYASCLAARITDWAV
				CAGCTGCGCCTAGCCTTCTACTGCGCCCTCTGCGGCACCGAGTACGAGG	GTCMTGA	TTCCTCC	TGGAVSRSQKGFMS
				CCCTGAAGCTCCTGAAGAACCACCAGAAGGGATGCGAGGGCCACGGAG	CGGGCTAG	ACCATT	TEGCYEHNFTLQMAL
				CCGAGAGGAGACCCGGCACGCTGGTGAGGTCCGCTGCCCCGGCCCGCC	GGAAGGC	GCGGAC	DNARRTRKQCAVA
				GGACCCAGGCCGCGGTGCGAAGGCCCGCCAGACTGGCCACCCCGCCGA	GGTGCTGC	CATTGTA	WLDISNAFGSVPHRH
				CAACCCCACCGGACCAGACCTCCAGGGACCACCCGACGGAGAGACCTG	CAGTACGT	ACGGGT	IFGTLRELGLPDGVID
				CCCCAGTGATGCCACCACGCAGGCCTCCGCCCAGGGACCCCCAACCGGA	CCGAAAGA	TTGTGT	LVRELYHGCTTTVRA
				CACGACGCCCCGACCAGACAGCCCCCCTCCCAGACCCCCGGGGCCACCC	GTGCCGGT	GTATCTA	TDGETAEIPIRSGVRQ
				GATGCCCCCGAACCGCCGGGGATCCCGGAGCCGCCCGGAGGAACCGA	TGCGCAAG	TCTTCTT	GCPLSPIIFNLAMEPL
				GCCGCCGGGAGCCCCCCCGKCCCCCGAGCGTGCCAGGGTCTCCGGGTGT	CGACCGCG	TCTCTCT	LRAVAGGPGGLDLY
				GGAGCCCTCCCCAGCAGAGGATGCCCACCCCATGGCAAACCCTCTGGCT	CCACTCAG	CAGCGT	GQKLSVLAYADDLVL
				GGAGGAGCTCTCCCGGGCCACCACCTTCAAGGCCTTCGAGGCCTCGGTG	GTGAGTAG	CGCGAA	LAPDATQLQQMLDV
				GCCCGGCTCACAGAGGAGCTCTCGGCGGCCGCCCGGCCCGGCCAGCCC	CCAAGGGT	CCCCCTC	TSEAARWMGLRFNV

	CGGGGGGGCAACAACAGACCGGCGACGCGACGGGACCACAGACTGCA	CTTACAGT	CCTCCCC		AKCASLHIDGRQKSR
	GCCGCAGAGGCGACCCAGGCGCCAGCGCTACGACCCGGCGGCAGCCTC	TCACCGGA	TTCCCCT		VLDSTLTIQGQAMRH
	CCGGATCCAGAAGCTGTACCGGGCCAACCGTCCCAAGGCGGTGAGAGA	CCCGAWA	CCCCCTC		LRDGEAYCHLGTPTG
	GATCCTGGAGGGACCCTCGGCCTTCTGCCAGGTCCCCCGGGAGACTCTG	ACGCGAAA	CCCCCCA		HRAKQTPEETINGIV
	TTCAACTACTTCAGCAGGGTCTTCAACCCCCCCAGCAGAAGCCGCCGCCC	ACCCCAAC	CCCCCG		QDAHKLDSSLLAPW
	CACGCCCTGCGACCGTCGAAGCGCTGACCCCCGTCCCCCCGGCGGAGG	TCGGGCTA	GGCTTA		QKIDAVNTFLIPRVAF
	GGTTCGAGGATGCCTTCACGCCGCAGGAAGTGGAAGCCCGCCTCAAGA	GTAGCCGA	GTTGGC		VLRGSAVPKTPLKKA
	GGACCAGGGACACCGCCCCCGGCAGGGACGGCATCAGGTACAGTCTCC	AGACCTGG	TAACATT		DAEIRRLLKKWLHLPL
	TCAAAAAGCGTGACCCGGGCTGCCTTGTTCTTTCTGTTCTCTTCAACAGG	GTCCCCCC	GTATCTC		RASNEVLHIPYRQGG
	TGCAGAGAGTTCCGGCGCACGCCCACACCTGGAAGAGGGCCATGACG	CMGGTCA	CTGTAA		ANVPRMGDLCDIAV
	GTCCTCATCCACAAGAAGGGAGACCCGACCGGGCAACTGGAGA	GAGTAGG	CCTAGTT		VTHAFRLLTCPDXTVS
	CCCATCGCCCTGTGCTCCACCGTCGCCAAGCTGTACGCCAGCTGCCTGG	CGAACGCC	GCGTTC		IIAASALEETARKRIGR
	CGGCCCGCATCACCGACTGGGCGGTGACCGGCGGGGCCGTCAGCCGGA	WGKGCTC	CCCTCCT		QPTRRDLATFLSGSLE
	GCCAGAAGGGCTTCATGTCGACGGAGGGCTGCTACGAACACAACTTCA	AGAGGAC	CACCCCC		GEFSRDGGDFASLW
	CCCTCCAGATGGCCCTGGACAATGCCCGGAGGACCAGGAAGCAGTGCG	GGAACGC	ATCCCTC		SRARNATRRLGKRIG
	CGGTGGCGTGGCTGGACATCTCCAATGCCTTCGGCTCCGTGCCCCACCG	GGAAAAC	TATTGTT		CAWTWTEERRELGV
	CCACATCTTCGGCACCCTCCGCGAGCTGGGCCTACCGGACGGCGTCATC	ACCCCCAG	AGTCCCT		SLQPAPHADRVTVTP
	GACCTGGTGCGAGAGCTCTACCACGGCTGCACCACGACCGTCCGCGCCA	GTCCCAAG	CGCTCG		RTRTFLERFLKDAVR
	CCGACGGAGAGACCGCGGAGATCCCCATCCGGTCGGGGGTGAGACAG	GACGCCCT	GGCGCT		NKYAGDLRAKPDQG
	GGCTGCCCCCTCAGCCCCATCATCTTCAACCTGGCCATGGAACCGCTCCT	GATCCACT	CTGTATT		KVFDVTSKWDSSNH
	TCGAGCCGTGGCAGGCGGCCCCGGCGGCTCGACCTGTACGGCCAGAA	GACAAGA	TCCCTAC		FMPSGSFTRFADWR
	GTTGAGCGTCCTGGCCTACGCCGACGACCTCGTTCTCCTCGCCCCCGAC	ACGCTCGA	CGGCTT		FLHRARLNCLPLNGA
	GCCACCCAGCTGCAGCAGATGCTGGACGTGACGTCCGAGGCGGCCAGG	GGCACGCC	TGTCATC		VRFGHRDKRCRRCG
	TGGATGGGCCTGCGCTTCAACGTCGCCAAGTGCGCCTCCCTGCACATCG	AGGAGAC	TTTTTTG		YVAETLPHVLCSCKP
	ACGGSAGGCAGAAGAGCCGCGTCCTGGACTCCACCCTCACGATCCAGG	CCCCAGCT	GATTCA		HARAWQLCHNAVQ
	GCCAGGCGATGAGGCACCTGCGCGACGGCGAGGCCTACTGCCACCTGG	AGGGTGG	CAATCCT		DRLVRAIPAAAGEISV
	GGACGCCCACCGGCCACCGGGCCAAGCAGACGCCGGAGGAGACCATCA	ACCGCCGA	AAACAT		NRTVPGCESQMRPD
	ACGGGATCGTGCAGGACGCCCACAAGCTGGACTCGTCCCTGCTGGCCCC	CTGCAGGT	CTACTAA		IVITNEEAKKVVIVDV
	CTGGCAGAAGATAGACGCGGTGAACACCTTCCTCATCCCCCGCGTCGCG	CCGGAGG	TAAAAG		TIPFENRRQAFTDAR
	TTCGTCCTGAGAGGCTCGGCGGTCCCCAAGACCCCCCTCAAGAAGGCG	ACCCTCCC	TCAATC		ARKREKYAPLADILRG
	GACGCCGAGATCCGGCGGCTGCTCAAGAAGTGGCTGCACCTGCCGCTG	AGGAGGG	(SEQ ID		RGYDVTVDALIVGTL
	AGGGCCAGCAATGAGGTCCTGCACATCCCCTACCGGCAGGGAGGTGCC	TGGACCAG	NO:		GAWDPSNESVLHAC
	AACGTCCCCCGCATGGGAGACCTCTGCGACATCGCGGTGGTCACCCACG	CGAACCCA	1304)		RVSRRYAKLMRCLM
	CCTTCCGCCTCCTGACCTGCCCGGACSCGACGGTAAGTATCATCGCCGCC	AGTTGGCG			VSDTIRWSRDIYVEHI
	AGCGCCCTCGAGGAGACCGCCCGCAAGAGGATCGGGAGGCAGCCCACC	ACGAACCC			TGHRQYTDPTRRTAA
	AGACGTGACTTGGCCACCTTCCTCAGCGGCTCGCTGGAGGGCGAGTTCA	TGACGCAC			GPDPEGTA (SEQ ID
	GCAGAGACGGCGGGGACTTTGCCTCGCTGTGGAGCCGAGCCCGCAACG	CCCCCACG			NO: 1426)
	CCACGCGCCGCCTCGGGAAGCGCATCGGCTGCGCCTGGACCTGGACCG	ATGTCAGG			

		AGGAGCGCCGGGAGCTGGGAGTCTCCCTGCAACCAGCCCCGCACGCCG	ACCCCGAC		
		ACCGCGTCACCGTGACGCCCCCGCACGAGGACCTTCCTGGAGAGGTTCCT	AGGCGGC		
		GAAGGACGCCGTCCGAAACAAGTACGCCGGCGACCTGAGGGCCAAACC	GGTGGACC		
		CGACCAGGGCAAGGTCTTCGACGTCACCTCGAAGTGGGACTCCAGCAA	ACTGACCA		
		CCACTTCATGCCCAGCGGGAGCTTCACGCGCTTCGCGGACTGGCGCTTC	TCGACCGA		
		CTCCACCGCGCCCGCCTCAACTGCCTGCCTCTGAACGGGGCCGTCCGCTT	CCCCCAGA		
		CGGCCACCGGGACAAGAGGTGCCGACGGTGCGGCTACGTGGCAGAGA	GGCAGAG		
		CCCTCCCCCACGTGCTGTGCAGCTGCAAGCCGCACGCCAGAGCCTGGCA	AGACTCTC		
		GCTCTGCCACAACGCTGTCCAGGACCGCCTGGTGAGGGCCATCCCGGCC	AGAGCCCG		
		GCAGCGGGGGAGATCTCCGTGAACCGCACCGTCCCGGGCTGCGAGAGC	GAACCCCG		
		CAGATGCGCCCCGACATCGTCATCACCAACGAGGAGGCCAAGAAGGTC	GCTGACGA		
		GTGATCGTGGACGTGACCATCCCCTTCGAGAACCGGCGCCCAAGCCTTCA	GAGCCGCC		
		CCGACGCCCGGGCTCGCAAGCGGGAGAAGTACGCCCCGCTGGCCGACA	TCCCGGCG		
		TCCTGAGGGGCCGCGGCTACGACGTGACGGTCGACGCGCTCATCGTGG	GAGGACCC		
		GAACGCTCGGAGCCTGGGACCCCAGCAACGAGAGCGTCCTGCATGCCT	CGGAGCCT		
		GCCGCGTCTCCCGCCGCTACGCCAAGCTGATGCGCTGCCTCATGGTGTC	GAGGATG		
		CGACACCATCCGTTGGTCCCGTGACATCTACGTGGAGCACATCACGGGC	CCCCCCGG		
		CACCGCCAGTACACCGACCCCACCAGACGAACCGCCGCCGGACCGGAC	ATGACGGC		
		CCAGAGGGGACCGCCTGAACCCCCCCTCTGCACCAGATGGACCTTCACT	GGAGCGC		
		TCGAGAGGATTCTTCAGCAATGGGACGACCCCGCTCCACCCGAAGAGGA	CCCGAGCG		
		CCCCCGCGATGAGACTCTATATGGACTGAGACACTTTTTCTTCGAACCAC	ACAGCGG		
		TTCCTCCACCATTGCGGACCATTGTAACGGGTTTGTGTGTATCTATCTTCT	ACCCCTCC		
		TTCTCTCTCAGCGTCGCGAACCCCCTCCCTCCCCTTCCCCTCCCCTCCCC	GGACCCCC		
		CCCACCCCCGGGCTTAGTTGGCTAACATTGTATCTCCTGTAACCTAGTTG	ACGGCCCC		
		CGTTCCCCTCCTCACCCCCATCCCTCTATTGTTAGTCCCTCGCTCGGGCGC	TCGGTGAC		
		TCTGTATTTCCCTACCGGCTTTGTCATCTTTTTTGGATTCACAATCCTAAA	GATGGCG		
		CATCTACTAATAAAAGTCAATC (SEQ ID NO: 1058)	GGCCCCGA		
			ACGACGAC		
			GACCCCCG		
			GACCCCGG		
			CGGTCCCG		
			AGGACGCC		
			CCCCCCGA		
			GGGTCTCC		
			CCACGCTG		
			GTGGAGG		
			AGCCCCGG		
			ACCCCCCC		


GACACCGG
ACCCCCCC
ACGGACG
ACCCAGGC
GAAGGCG
TAGACATG
ACAGCACT
CACGTTCC
TCCCCTTC
CCCCTCCC
GGCGAAG
CTGTTCTG
CCCGACCT
GCCACCCG
CCAAGACA
GTACAGGT
CGCACGGC
GACATGAA
CAAGCACC
TACGGCGC
TTCCACCA
GCTGCGCC
TAGCCTTC
TACTGCGC
CCTCTGCG
GCACCGAG
TACGAGGC
CCTGAAGC
TCCTGAAG
AACCACCA
GAAGGGA
TGCGAGG
GCCACGGA
GCCGAGA
GGAGACCC
GGCACGCT
GGTGAGG
TCCGCTGC

					CCCGGCCC GCCGGACC CAGGCCGC GGTGCGA AGGCCCGC CAGACTGG CCACCCCG CCGACAAC CCCACCGG ACCAGACC TCCAGGGA CCACCCGA CGGAGAG ACCTGCCC CAGTGA (SEQ ID NO: 1181)		
NeSL	Utopia_1_CMy	.	Chelonia mydas	CTCTTCTTATGAATACTTGCAACACCTGCACTGAAGATGGATTCTCCGGC	CTCTTCTTA	TGAGCC	MTTKKVLGASTTLQT
				TGCTATTTTTGAAAAAACTGATGCTGCTTTGAAGGTGTATTCTGCTGCTGC	TGAATACT	GGTACG	SSTKGKNSGCSKDPL
				TACCTTGGAAGGAAATTCTCTCTCTGCTCCTGAGACATCCCCAGCTGCAC	TGCAACAC	ACATCG	RDAVPGRSWILRPAC
				CGTGTACCACCACCACTGCTGCTCCACAGAAGGTTTCTCGGAC	CTGCACTG	TGCATC	RDITTRRNIPPAPQQ
				AATGACTACAAAGAAGGTCCTAGGTGCCTCCACAACATTACAGACCAGC	AAGATGG	AACTAT	QQPPMESPPTLQLQ
				AGCACGAAGGGGGAAGAACAGTGGCTGCTCAAAGGACCCCCTCCGAGAT	ATTCTCCG	GAGAAA	DALRRPSPTPAAAQV
				GCTGTTCCAGGAAGATCCTGGATTCTGAGGCCGGCCTGTCGGGACATCA	GCTGCTAT	GGGACT	ADAGGALAALHTIKR
				CAACCAGAAGGAACATCCCCCCCGCCCCCCAGCAGCAGCCGCCAAT	TTTTGAAA	GAGAGA	GISVDWTSISPKXXQ
				GGAGAGCCCCCCCCACTCTGCAGCTGCAAGATGCTCTCAGGCGACCATCT	AACTGATG	CTTTTTC	RXTSASPDACPASET
				CCCACCCCCGCAGCTGCCCAGGTCGCTGACGCTGGTGGTGCTCTCGCTG	CTGCTTTG	CATTGG	TQRDXRXLLDARPAG
				CTCTACACACCATCAAGAGAGGAATCTCCGTAGACTGGACCAGCATCTC	AAGGTGTA	ACCATAT	PLDPTRPHQDEPASD
				TCCAAAGASCMCCCAGAGGSTCACCAGCGCCTCGCCGGATGCCTGCCCA	TTCTGCTG	GAACTG	TADAAGTPLLQGNE
				GCCTCAGAAACCACCCAGAGGGACCSCAGGWSCCTGCTGGACGCCCGC	CTGCTACC	GAACCA	DTIYLQYPLAADMLIC
				CCAGCCGGACCTCTCGACCCTACCCGCCCACACCAGGATGAACCAGCCA	TTGGAAGG	TAAACTC	PICSPPQSFHLLGVVT
				GCGATACCGCTGATGCTGCTGGAACCCCCCCTGCTGCAAGGTAATGAGG	AAATTCTC	ACTGAA	RHLKRCHSKRVAFSC
				ACACCATCTACCTGCAGTATCCCCTCGCTGCGGACATGCTCATCTGCCCC	TCTCTGCT	CATTAA	ALCSLPFETQKQCKM
				ATCTGCTCTCCGCCCCCAAAGCTTCCACCTCCTCGGTGTCGTCACCAGGCA	CCTGAGAC	ATCTCAC	HQVACRKCLKGTTQS
				CCTGAAGAGATGCCACACAGCAAGCGGGTTGCCTTCAGCTGTGCCCTCTGC	ATCCCCAG	CAAATG	PAPAPSPPAARRPAA
				AGCCTGCCCTTCGAGACGCAGAAGCAATGCAAGATGCACCAAGTCGCCT	CTGCACCG	AGGGTA	PEPQRRKXTSQAAVK
				GCAGGAAATGCCTCAAGGGAACMACACAGTCTCCTGCCCCGGCTCCCA	TGTACCAC	AATCCAT	KPAPVARPAERDAAI
				GCCCTCCTGCTGCACGCCGGCCCGCTGCTCCTGAGCCTCAACGAAGAAA	CACCACCA	CCTCATC	EKVPAASGNITQVLA

		GSCGACCTCGCAAGCTGCCGTCAAGAAGCCTGCCCCCGTCGCCAGGCCA	CTGCTGCT	ATCGTAT		SRRPVSPSHVAKXIS
		GCGGAACGGGATGCTGCGATCGAGAAGGTACCTGCTGCCTCGGGGAAC	GCTCCACA	CCACTCA		MLRRLSAASPPVQH
		ATCACCCAGGTCCTCGCCAGCAGGAGGCCCGTCTCACCCTCTCATGTCG	GAAGGTTT	TTATACT		VPVPRRISAPPRIAAR
		CCAAGMAGATCTCCATGCTGAGACGACTCAGTGCTGCCTCGCCACCTGT	CTCGGACA	CCACAC		DPVAGRASAAPQTA
		CCAGCACGTCCCCGTCCCCAGAAGGATCAGCGCCCCACCGCGCATAGCT	(SEQ ID	CTGAAC		LRTPAAGGASTTPQT
		GCTCGAGATCCTGTCGCCGGAAGAGCCAGCGCCGCCCCTCAGACCGCCC	NO: 1182)	ATAGCC		ALRTPTAGGASAMP
		TGCGAACTCCAGCCGCCGGAGGAGCCAGCACCACGCCTCAGACCGCCC		ATTATAT		QTTLPXPRRPDWRN
		TGCGAACTCCAACCGCCGGAGGAGCCAGTGCCATGCCTCAGACCACCCT		GAACAA		QPRSHSKAPGLHRQT
		GCCAGMCCCCAGACGTCCAGACTGGAGGAACCAGCCCCGCAGCCACAG		CATACCC		DQHGPQVHSAGHCL
		CAAAGCACCGGGCCTTCATCGCCAGACGGACCAGCACGGCCCCCAAGTC		CCATATC		REISRSSSNRLGSSHS
		CATTCTGCGGGACACTGCCTACGGGAGATCTCACGCTCCAGCAGCAACC		TCAATGT		AAATHRRTGGVPAT
		GCCTAGGCAGCAGCCACTCATAGAAGGACCGGCGGTGTCCCAGCAACC		CTGTACT		PEPDRVSPTTSNAXIP
		CCCGAGCCGGACCGCGTCTCTCCGACCACCAGCAACGCCASCATCCCGC		TTGACCC		PEIPPQHPTEGNPDP
		CAGAGATCCCGCCCCAGCACCCAACCGAAGGGAATCCTGACCCACGAG		GTTAAC		RDRRQADHTAGSEP
		ATAGACGGCAGGCCGACCATACAGCAGGCTCWGAGCCTGCACCAGAC		CTTTTAC		APDEVEDXEGQRPM
		GAGGTCGAGGACCMTGAGGGCCAGCGGCCGATGGTGAGGGCTGCCAC		CCCCAAT		VRAATPWQTAWTEE
		NCCGTGGCAGACTGCCTGGACCGAGGAGCTACAAGCGGCAGCTTCCTT		CGGGGA		LQAAASFDDFDLLVD
		TGACGACTTCGACCTCCTCGTAGACAGGCTCACCCGAGAACTGTCTGCG		TATTGCA		RLTRELSAEIAPRRSS
		GAAATCGCTCCCAGGAGGAGTTCGAACCAGGAGAAGCGCCCCGCCTGCC		GATTAT		NQENAPPAHRTPAP
		CACAGAACGCCTGCTCCGAACCACACAACCACCACCAGGGGAGCCAGA		GTATTCC		NHNTTTRGARSRDA
		AGTAGAGACGCCAGCCGCCGCTACGATCCAGCAGCGGCTTCAAGGATC		TTACGCC		SRRYDPAAASRIQKLY
		CAAAAGCTGTACCGGGCAAACCGCTCCAAGGCCATGAGGGAGATCCTA		ACCCGA		RANRSKAMREILDGP
		GACGGGCCCTCGCCCTACTGCACGATCCCATCTGAGCGTCTCTACAGCT		TCCTAAA		SPYCTIPSERLYSYFKD
		ACTTCAAGGATGTATTCGACCGCATAGCCCGGAATGACGCGCAGCGCCC		CCGAAT		VFDRIARNDAQRPEC
		AGAGTGCCTCCGCCCCCTGCCCCGTGTCGACGAAGCAGGTGTCCTGGAA		TTCGCAC		LRPLPRVDEAGVLET
		ACTGACTWTACGCCCAAGGAAGTGATGGCCAGACTCTCAAAAACAAAA		CCCTTGA		DXTPKEVMARLSKTK
		AACACAGCTCCTGGGAAAGACGGCATCCCCTACAGCCTCCTGAAAAAGC		TAATCTG		NTAPGKDGIPYSLLKK
		GAGATCCCGGCTGCCTGGTCCTCGCCACGCTCTTCAACCAGTGCAAGCG		TACCTTA		RDPGCLVLATLFNQC
		ATTCTGCCGGACTCCCAGCTCCTGGAAGAAGGCCATGACGGTACTGGTG		TTCCCTG		KRFCRTPSSWKKAM
		TACAAGAAGGGCGAGCGGGATGACCCCAGCAACTGGAGGCCCATCTCC		ATAACC		TVLVYKKGERDDPSN
		CTCTGCTCCACGATGTACAAGCTCTATGCCAGCTGCCTGGCGTCGAGGA		AGAAAC		WRPISLCSTMYKLYA
		TCACGGAGTGGTCGGTGAGCGGGGGGAGCCATCAGCTCCATCCAGAAAG		TTCTATG		SCLASRITEWSVSGG
		GCTTCATGTCCTGCGAGGGCTGCTACGAACACAACTTCGTCCTCCAAACC		CTTAAAC		AISSIQKGFMSCEGCY
		ACCATCGAAACGGCCAGAAGGGCGCGGAGGCAGTGCGCGGTAGCGTG		TCTGTAC		EHNFVLQTTIETARR
		GCTCGACCTGGCTAACGCCTTTGGGTCCATGCCCCACCACCACACATCTTTG		CGTTTTT		ARRQCAVAWLDLAN
		CCACGCTCCAGGAGTTTGGGATGCCAGAGAACTTCCTTCGTGTGATCCG		TTTTATT		AFGSMPHHHIIFATLQ
		AGAGGTGTACGAGGGATGCAGCACCACCATTCGCTCGGTCGAAGGGGA		TCAACAT		EFGMPENFLRVIREV
		GACCGCCGAGATCCCGATCCGGAGCGGAGTTAAGCAGGGCTGTCCCCT		CATCTTA		YEGCSTTIRSVEGETA

			CAGCCCCATCATCTTTAACCTCGCCATGGAGCCGTTGCTGCGAGCGATCT	ATAAAA TTATTAA A (SEQ ID NO: 1305)	EIPIRSGVKQGCPLSPI
			CCAATGGCACAGATGGCTTCAACCTCCACGGTGAGAGGGTGAGCGTCC		IFNLAMEPLLRAISNG
			TGGCTTACGCGGATGACCTGGTCCTGACCGCGGATGACCCAGAGAGCC		TDGFNLHGERVSVLA
			TCCAAGGTATGCTAGATGCCACCAGTCGAGCTGCCGACTGGATGGGGC		YADDLVLTADDPESL
			TCCGCTTCAATGCAAAGAAGTGCGCAACTCTCCACATCGACGGCAGCAA		QGMLDATSRAADW
			AAGGGACTCGGTGCAGACGACGGGGTTCCAGATCCAGGGCGAGCCCGT		MGLRFNAKKCATLHI
			CATCCCCCTGGCAGAGGGGCAGGCGTACCAGCACCTCGGCACGCCGAC		DGSKRDSVQTTGFQI
			GGGTTTCCGTGTCCGGCAGACACCCGAGGACACCATCCAGGAGATCTTG		QGEPVIPLAEGQAYQ
			CAGGATGCCGCCAAGATCGACGCCTCCCTGCTAGCACCGTGGCAGAAG		HLGTPTGFRVRQTPE
			ATAAACGCCCTGAACACCTTCCTGATCCCCCGCATCTCGTTCGTCCTAAG		DTIQEILQDAAKIDAS
			GGGATCCGCCGTGGCGAAGGTACCCCTCAACAAGGCAGACAAGATCGT		LLAPWQKINALNTFLI
			CCGGCAGCTGGTGAAGAAGTGGCTGTTCCTTCCCCAGAGAGCCAGCAA		PRISFVLRGSAVAKVP
			CGAGCTGGTCTACATCGCCCACAGGCATGGCGGTGCCAACGTCCCCCGC		LNKADKIVRQLVKKW
			ATGGGCGACCTGTGTGACATCGCGGTGATCACCCACGCCTTCCGCCTGC		LFLPQRASNELVYIAH
			TGACGTGTCCCGACGCCATGGTAAGGAACATCGCGGCAAACGCCCTCCA		RHGGANVPRMGDL
			TGACGCGACAAAGAAGCGGATCGGCAGAGCCCCCTCCAACCAAGACAT		CDIAVITHAFRLLTCP
			CGCCACCTTCCTGAGCGGTTCCCTGGATGGCGAATTCGGACGGGACGG		DAMVRNIAANALHD
			GCGCGACATCGCTTCACTGTGGTCCCGCGCTCGCAACGCCACGCGTCGC		ATKKRIGRAPSNQDI
			CTGGGGAAGCGCATCGGCTGCCGCTGGGAGTGGTGCGAGGAGCGCCA		ATFLSGSLDGEFGRD
			GGAGCTGGGAGTCCTGGTGCCGCAGATCAGGTCCAACGACACAACCACCAT		GRDIASLWSRARNAT
			CGTCACCCCGAGCGCCAGGGGCATGCTGGAGAGGACCCTGAAGGCAGC		RRLGKRIGCRWEWC
			CATCCACTCACTGTACGTGGAAACCCTGAAGCGTAAACCGGACCAGGGT		EERQELGVLVPQIRS
			AAAGCCTTCGAACTGACCAGCAAGTGGGACGCCAGCCAACCACTTCCTC		NDNTIVTPSARGMLE
			GCCGGGGGCGGCTTCACCCGTTTCGCCGACTGGCGGTTCATCCACCGTG		RTLKAAIHSLYVETLK
			CCCGGCTCAACTGCGTCCCGCTCAACGGAGCCGTCCGCCACGGGAACCG		RKPDQGKAFELTSK
			AGACAAGCGTTGCAGGAAGTGCGGCTACTCCAACGAGACCCTGCCCCA		WDASQPLPRRGRLH
			CGTCCTGTGCAGCTGCAAGCCCCACTCCAGAGCCTGGCAGCTGCGCCAC		PFRRLAVHPPCPAQL
			AATGCCATCCAGAACCGCCTGGTGAAAGCCATCGCACCGCGCCTGGGG		RPAQRSRPPREPRQA
			GAGGTCGCCGTGAACTGCGCCATCCCCGGTACTGACAGCCAGTTGCGAC		LQEVRLLQRDPAPRP
			CTGACGTGGTAGTCACCGACGAGGCCCAGAAAAAGATCATCCTCGTCG		VQLQAPLQSLAAAP
			ACGTCACGGTCTCCTTTGAGAACAGGACCCCGGCCTTCCGCGAAGCCCG		QCHPEPPGESHRTAP
			AGCTCGTAAGCTGGAAAAATACGCCCCCCTGGCCGACACCCTGAGAGC		GGGRRELRHPRYPAS
			GAAGGCTACGAGGTGCAGATGGATGCCCTGATCGTCGGAGCCCTGGG		GTPANHFLAGGGFT
			CGCTTGGGACCCCTGCAATGAGCGTGTGCTGCGGACCTGTGGGATCGG		RFADWRFIHRARLNC
			TCGACGCTACGCACGGCTCATGCGGCGCCTCATGGTCTCGGACACCATC		VPLNGAVRHGNRDK
			CGATGGTCCAGGGACATCTACATCGAACACATCACCGGCCACCGACAGT		RCRKCGYSNETLPHV
			ACCAGGAGGTGTGAGCCGGTACGACATCGTGCATCAACTATGAGAAAG		LCSCKPHSRAWQLR
			GGACTGAGAGACTTTTTCCATTGGACCATATGAACTGGAACCATAAACT		HNAIQNRLVKAIAPR

			CACTGAACATTAAATCTCACCAAATGAGGGTAAATCCATCCTCATCATCG TATCCACTCATTATACTCCACACCTGAACATAGCCATTATATGAACAACA TACCCCCATATCTCAATGTCTGTACTTTGACCCGTTAACCTTTTACCCCCA ATCGGGGATATTGCAGATTATGTATTCCTTACGCCACCCGATCCTAAACC GAATTTCGCACCCCTTGATAATCTGTACCTTATTCCCTGATAACCAGAAA CTTCTATGCTTAAACTCTGTACCGTTTTTTTTTTATTTCAACATCATCTTAAT AAAATTATTAAA (SEQ ID NO: 1059)			
NeSL	Utopia_CPB	.	Chrysemys picta bellii	TTTTTTCTGATGCTTGACTGCAAACACCCATCCAGAAGATGGAATCTCCT GCAGCCATTTTTGAAAAAAATTGATGCTGCTTTAAAGATATACTCCATTCT CCTWKTTTGKAAGAAAACTCTTTTTCAGCTTCAGCTATTCTGTCATCGGC TGCTGCTGTTCCTGCTTCCCAGAAAGCTCAGCMAAAACCTATCCTGAAG ACCWCCCTTGGTGCCTCACGGAAGACCCGGASCACCTGCAAGAACCAA AACATTAGGAGCTGGCTGAAGAAACCCCCCGTGGATACCTCWGCAGGG AGACCTGGSTCCAGMAGGACAKCTCTTCGGGACCTCMCATCSAGGAGC AAGAATATCTCAACAGCTCTTCAGGAGGGGGACCCCCGGAGAACCCTG CCCGCTTCCCAGAACCAGGATGCTGATCGCCGCCCCACCGGGAAGGAT GCCACCGCAGGAGCCCCCCCCAATGACCCAGGACCAGGATGCTGATTGCT GCCCCGCCGGGAAGGATGCCACCAGAGGAGCCCCCCCGATGACCCAGG ACCAGGATGCTGATCGCTGCCCCGCTGCTCCAGAGAGGGATGCTCCGG AAGGAACCACCTCCTCAACCCCAGACCCCAAAACTACTTACCACCCGGCT GTCCGGAGGAGGGCCGCTCGAAGGGGAATGCACCTCAGAGCCCAAGA TCTCGATGCCGCACGCTGCCCTTCCGGGCAAAGAGACAACGTGGCCAGT GAGTCCAGCGCCCCCCCAAGAGCGACTTCACCTCCCCAAGCTTCTCTACC AGACCCAGAGGAATCACCTGGCGAGTCTGCAGGCACAACAGAGATCCG CCCCACGGAGGGTGAGGCAGGGGAAGAAGACCGCATCTACCTCCAGTA CCCGCTCCCTACAGGTCTCCTCCTCTGCCCCTTCTGTCTCCCCGTCCATGG AGTCCAGACCCTCGCGGCCCTCAGCAAACACGTTCGTAAGACCTACAAC AAACGGATTGCTTTCCGGTGTAGCCGCTGCGATCTCCCCTTCGAGACCC AAAAGAAATGTAAGTTTCATCAAGCCACGTGCAGGGGACCCCCCCACGAC CGCGAAAGTGAATCCCACTGACATCCTCCGGGTTCCAACCCTGACCCCC ACCGATGATCTGGCTTCAGCACCCCAGCCAGCATCCCCAGAGTCACAGC AGATAAGGGGGGGACCAACCGCCAACTGAGGGAAGCGTAACCCCCGCCT CGAGGACTGACGATGCCACCAAAAGGACCAGTCCCGTCTCCAGAATCCC CACGCTGGACCCTGCTGTGAGGGGGACCACCGCCACCTCTCAGGTCAAC AACCTCACCAGACGCCTCAGCGACCTCATAAAAACCATCCGGCACAACA	TTTTTTCTG ATGCTTGA CTGCAAAC ACCCATCC AGAAGAT GGAATCTC CTGCAGCC ATTTTTGA AAAAATTG ATGCTGCT TTAAAGAT ATACTCCA TTCTCCTW KTTTGKAA GAAAACTC TTTTTCAG CTTCAGCT ATTCTGTC ATCGGCTG CTGCTGTT CCTGCTTC CCAGAAAG CTCAGCM AAAACCTA TCCTGAAG ACCWCCCT TGGTGCCT CACGGAA	TGAGCC AGAGTG ACATCG TTCTCCC ACTACG AGAAAG GGACCA AGTGAC CTTCTCC GTTGGA TCATATG AACTGG AACCAT AAACTC CCTGAA CATTAA ATCTCAC CAAATG AGGGTC AATCCAT CCTCATC ATCATAT CCACTCA TTATAM TCCACAC CCGAAC ACAGCC ACTCTAT	LGEVAVNCAIPGTDS QLRPDVVVTDEAQK KIILVDVTVSFENRTP AFREARARKLEKYAPL ADTLRAKGYEVQMD ALIVGALGAWDPCN ERVLRTCGIGRRYARL MRRLMVSDTIRWSR DIYIEHITGHRQYQEV (SEQ ID NO: 1427) MTQDQDADCCPAG KDATRGAPPMTQDQ DADRCPAAPERDAP EGTTSSTPDPKTTYHP AVRRRAARRGMHLR AQDLDAARCPSGQR DNVASESSAPPRATS PPQASLPDPEESPGE SAGTTEIRPTEGEAGE EDRIYLQYPLPTGLLL CPFCLPVHGVQTLAA LSKHVRKTYNKRIAFR CSRCDLPFETQKKCKF HQATCRGPPTTAKV NPTDILRVPTLTPTDD LASAPQPASPESQQI RGDQPPTEGSVTPAS RTDDATKRTSPVSRIP TLDPAVRGTTATSQV NNLTRRLSDLIKTIRH NTDTRRCSAPPQVTS CRPAVGATSIVPQAA RRDPANGGASRSPQI PQPDPAPGRPNTSSK VTQRASDRQKPHAP PRTHQPDAARRRTRT IPSASKHDRAPTKPST GASRTPLPPGRSSAA

	CGGACACGAGACGCTGCAGCGCTCCCCCACAGGTAACCTCATGCCGCCC	GACCCGGA	GAACTT	SETPRAALPTTPGPPP
	TGCCGTAGGAGCAACTAGCATCGTCCCCCAGGCTGCACGGCGAGATCC	SCACCTGC	CATACCC	QDPPEHRSTVRGTTR
	AGCCAACGGAGGAGCCTCCCGTAGCCCCCAGATCCCACAGCCAGACCCC	AAGAACCA	TCATATC	PQTVPAAPEPAETTQ
	GCCCCCGGGAGACCCAACACCTCCTCCAAGGTTACCCAAAGAGCCTCTG	AAACATTA	TCAATGT	QEERRPRARVATPW
	ACCGCCAAAAAACCCCATGCCCCCACCGAGGACCCACCAGCCGGATGCCGC	GGAGCTG	CTGTACT	QSAWMEELAKAEDF
	CCGCAGAAGAACCAGAACCATCCCCAGCGCTTCCAAACACGACCGCGCC	GCTGAAGA	TTGACCC	ENFDTLMDRLTAELS
	CCGACAAAGCCCAGCACCGGTGCTTCCAGAACCCCACTCCCTCCCGGAA	AACCCCCC	ATCAAC	AEITARRREPQEAAR
	GATCCAGTGCTGCCTCGGAGACACCGAGAGCTGCCCTCCCCACCACACC	GTGGATAC	CTTTTAC	ATRRFPAPSRNNTAR
	AGGACCCCCGCCTCAAGACCCACCTGAACACCGCTCCACAGTCCGAGGG	CTCWGCA	CCCCAAT	EGRRGDVGRRYDPA
	ACAACAAGGCCGCAAACCGTCCCCGCAGCACCTGAACCTGCAGAGACA	GGGAGAC	CGGGGA	AASRIQKLYRMNRTK
	ACGCAGCAGGAGGAGCGACGGCCACGAGCGAGGGTCGCCACGCCGTG	CTGGSTCC	TATTGCA	AMREILDGTSSYCAI
	GCAATCCGCCTGGATGGAGGAGCTGGCAAAGGCTGAGGACTTTGAGAA	AGMAGGA	GATTAT	QPERLYSYFKDVFDH
	CTTCGACACCCTGATGGACAGACTGACTGCAGAACTGTCTGCGGAAATT	CAKCTCTT	GTATTCC	EAQTNLRRPECLSPLP
	ACGGCCAGAAGGAGGGAACCCCAGGAAGCCGCACGGGCCACTCGCAG	CGGGACCT	TCATGCC	RIDLTEDLERDFSPQE
	ATTCCCTGCGCCGAGCCGTAACAACACCGCCAGAGAAGGCAGGAGAGG	CMCATCSA	ACCTGA	VQARLSRTKNTAPGK
	GGACGTCGGCCGCCTACGATCCGGCGGCTGCATCCCGTATTCAGAAA	GGAGCAA	TCTTAAA	DGIRYPLLKKRDPGCL
	CTATACAGGATGAACCGGACGAAAGCCATGAGGGAGATCCTCGACGGG	GAATATCT	CCAAAC	VLAAIFNKCKQFHRV
	ACCTCCTCCTACTGTGCCATCCAGCCCGAGAGGCTCTACTCCTACTTCAA	CAACAGCT	TTTGCAC	PRSWKKSMTVLIHKK
	GGATGTGTTTGATCACGAGGCCCAGACCAACTTGCGACGCCCAGAGTG	CTTCAGGA	CCTCGAT	GXRDDPGNWRPISL
	CCTTTCCCCGCTACCCCGGATCGACCTCACGGAGGACTTGGAGCGAGAT	GGGGGAC	AATCTGT	CSTIYKLYASCLAARIT
	TTTTCCCCGCAGGAGGTGCAGGCGAGGCTGTCGAGGACCAAAAAACACC	CCCCGGAG	ATGTTAT	DWSVCGGAVSSVQK
	GCCCCTGGAAAAGATGGCATCCGCTACCCCCCTGCTGAAGAAGCGAGAC	AACCCTGC	TCCCTGA	GFMSCEGCYEHNFLL
	CCCGGCTGCTTGGTGCTCGCTGCCATCTTCAACAAATGCAAGCAGTTCCA	CCGCTTCC	TAACCA	QTAIQEARRSKRQCA
	TCGCGTTCCCCGCTCCTGGAAAAAGTCCATGACCGTGCTCATCCACAAA	CAGAACCA	GAAACT	VAWLDLTNAFGSIPH
	AAAGGCGAMCGAGACGACCCCGGCAACTGGAGGCCCATCTCCCTCTGC	GGATGCTG	TCTATGC	HHIFATLGEFGMPET
	TCCACCATCTACAAGCTGTATGCCAGCTGCCTCGCGGCAAGGATCACAG	ATCGCCGC	TCAAACT	FIQILRDLYKDCTTTIR
	ACTGGTCAGTGTGCGGGGGCGCCGTCAGCTCAGTGCAGAAGGGTTTCA	CCCACCGG	CTGTTCA	ATDGETDAIPIRRGV
	TGTCCTGCGAGGGATGCTACGAGCACAACTTCCTCCTTCAGACGGCCAT	GAAGGAT	CTATTTT	KQGCPLSPIIFNLAME
	CCAGGAGGCCAGGAGGTCCAAGAGGCAGTGCGCAGTAGCATGGCTTG	GCCACCGC	TTTTAAC	PLIRAISSGPTGFDLH
	ACCTGACCAACGCCTTTGGGTCCATACCCCACCATCACATCTTTGCCACC	AGGAGCCC	ATCATCT	GKKLSILAYADDLVLT
	CTGGGAGAGTTCGGGATGCCAGAAACCTTCATCCAGATCCTCCGGGACC	CCCCA	TAATAA	ADDPESLQGMLDAT
	TCTACAAGGACTGCACCACCACCATCCGCGCCACGGACGGAGAGACGG	(SEQ ID	AATTTTT	SRATDWMGLRFNAK
	ACGCCATCCCCATCCGCCCGCGGCGTGAAACAAGGATGCCCCCTTAGCCC	NO: 1183)	AAATCT	KCATLHIDGSKRDSV
	CATCATCTTCAACCTGGCCATGGAACCGCTCATCCGAGCCATCTCCAGCG		GTT (SEQ	QTTGFQIQGEPVIPL
	GCCCGACCGGCTTCGACCTGCACGGCAAGAAACTCAGCATTCTGGCCTA		ID NO:	AEGQAYQHLGTPTG
	CGCGGACGATCTGGTCCTGACCGCGGATGACCCAGAGAGCCTCCAAGG		1306)	FRVRQTPEDTIQEILQ
	TATGCTAGATGCCACCAGCCGAGCTACTGACTGGATGGGGCTCCGCTTC			DAAKIDASLLAPWQK
	AATGCGAAGAAGTGCGCAACTCTGCACATTGACGGCAGCAAAAGGGAC			INALNTFLIPRISFTLR

			TCGGTGCAGACAACGGGGTTCCAGATCCAGGGTGAGCCCGTCATCCCCC			GSAVAKVPLNKADKII
			TGGCAGAGGGGCAGGCATACCAGCACCTGGGCACGCCAACAGGGTTCC			RKLVKKWLFLPQRAS
			GTGTCCGGCAGACACCCGAGGACACCATCCAGGAGATCTTGCAGGACG			NELVYIAHRHGGANV
			CCGCCAAGATTGATGCCTCCCTGCTGGCACCGTGGCAGAAGATAAACGC			PRMGDLCDVAVITH
			CCTGAACACCTTCCTGATCCCACGCATCTCGTTCACCCTAAGGGGATCCG			AFRLLTCPDATVRNIA
			CCGTGGCGAAGGTGCCCCTCAACAAGGCAGACAAGATCATCCGGAAGC			ANALRDATEKRIGRA
			TGGTGAAGAAGTGGCTGTTCCTTCCCCAGAGAGCCAGCAACGAGCTGG			PSNQDIATFLSGSLD
			TCTACATCGCCCACAGGCACGGCGGCGCCAACGTCCCCCGCATGGGTGA			GEFGRDGRDIASLWS
			CCTGTGCGACGTCGCGGTGATCACCCACGCCTTCCGCCTGCTGACATGT			RTRNATRRLGKRIGC
			CCCGACGCCACGGTGAGGAACATTGCGGCGAACGCCCTGCGTGATGCG			RWEWCEERQELGIR
			ACAGAGAAGCGGATCGGCAGAGCCCCCTCGAACCAAGACATCGCCACC			VPQIRSDDNTIVTPTA
			TTCCTGAGCGGCTCCCTGGATGGGGAATTCGGACGGGACGGGCGCGAC			RGLLERTLKAAIRSLY
			ATCGCTTCACTGTGGTCCCGCACTCGCAACGCCACGCGTCGCCTGGGGA			VETLKRKPDQGKAFE
			AGCGCATCGGCTGCCGCTGGGAGTGGTGCGAGGAGCGCCAGGAGCTG			LTSKWDASNHFLDG
			GGAATCCGGGTGCCGCAGATCAGGTCCGACGACAACACCATCGTCACC			GGFTRFADWRFIHRA
			CCGACGGCCAGGGGCTTGCTGGAGAGGACTCTGAAGGCCGCCATCCGC			RLNCVPLNGAVRHG
			TCGCTGTACGTGGAAACCCTGAAGCGTAAACCGGACCAGGGTAAAGCC			NRDKRCRKCGYPNET
			TTTGAGTTGACCAGCAAGTGGGACGCCAGCAACCACTTCCTCGACGGG			LPHVLCSCKPHSRAW
			GGCGGCTTCACCCGTTTCGCCGACTGGCGGTTCATCCACCGTGCCCGGC			QLRHNAIQNRLVKAI
			TCAACTGCGTCCCGCTCAACGGAGCCGTCCGCCACGGGAACCGAGACA			APRLGEISVNCTIAGT
			AGCGTTGCAGGAAGTGCGGCTACCCCAACGAGACCCTGCCCCACGTCCT			DSQLRPDVVVTDEA
			GTGCAGCTGCAAACCCCACTCCAGAGCCTGGCAGCTGCGCCACAACGCC			QKKIILVDVTVSFENR
			ATCCAGAACCGCCTGGTGAAAGCCATCGCGCCACGCCTGGGGGAGATC			TPAFREARARKLEKY
			TCCGTGAACTGCACCATCGCCGGTACCGACAGCCAGCTACGACCTGACG			APLADTLRAKGYEVQ
			TGGTCGTCACCGACGAGGCCCAGAAAAAGATCATCCTCGTCGACGTCAC			MDALIVGALGAWDP
			GGTCTCCTTTGAGAACAGGACCCCGGCATTTCGCGAAGCCCGAGCTCGT			CNERVLRTCGIGRRY
			AAGCTGGAAAAGTACGCCCCCCTGGCTGACACCCTGAGAGCGAAGGGC			ARLMRRLMVSDAIR
			TATGAGGTGCAGATGGACGCCCTGATTGTCGGAGCCCTGGGCGCCTGG			WSRDIYIEHITGHRQ
			GACCCCTGCAACGAGCGTGTGCTGCGGACCTGCGGGATCGGTCGACGC			YQEA (SEQ ID NO: 1428)
			GTGACCTTCTCCGTTGGATCATATGAACTGGAACCATAAACTCCCTGAAC			
			TACGCACGTCTCATGCGGCGCCTCATGGTCTCAGACGCCATCCGATGGT			
			ATTAAATCTCACCAAATGAGGGTCAATCCATCCTCATCATCATATCCACT			
			CCAGGGACATCTACATCGAGCACATCACCGGCCACCGACAGTACCAGGA			
			CATTATAMTCCACACCCGAACACAGCCACTCTATGAACTTCATACCCTCA			
			TATCTCAATGTCTGTACTTTGACCCATCAACCTTTTACCCCCAATCGGGG			
			ATATTGCAGATTATGTATTCCTCATGCCACCTGATCTTAAACCAAACTTTG			
			CACCCTCGATAATCTGTATGTTATTCCCTGATAACCAGAAACTTCTATGCT			

			CAAACTCTGTTCACTATTTTTTTTTAACATCATCTTAATAAAATTTTTAAATC TGTT (SEQ ID NO: 1060)				
NeSL	Utopia-1_DYak	.	Drosophila yakuba	AAAGTGTAGTTCTTTTCTGTTTTAGTGTAGTGGGAAGTCTGTTTCTTTTTA	AAAGTGTA	TAAAAA	YAPGYEAAQSPCGRE
				TTATGTTTTTTACGAAAAAAGTCCTGGTCTTTGAAATTCATTGTCTAAATTT	GTTCTTTTC	ATTAAA	PPRDHHRRPRDACG
				TAAATAAAATTATAAAATTTAAAAAGAAAATTAATTAAAGAAGCGATGA	TGTTTTAG	ATGCCTT	SSHSPEPCLTTPRLLP
				AATATCTCTGAAATTCAATCAATTAATCATGGCGTCTCAGCGAGTG	TGTAGTGG	AAAAAT	ETVSAEPCDDESQRT
				CACGTATTTGCCTACCCCTTCGTGGGACCATTCCGGTGCTCCGTATGCAT	GAAGTCTG	AAATAA	RYASPHKQARTLHDA
				GGATGCGTCCGGGATGCATCCCACTAGGTCGCTGGGCGAATACGGCAC	TTTCTTTTT	ATATATC	EPRDASREHAPSCAE
				ATACGCTGCGGCATATCAGCACATAACCGGCGCCACCCACAAGTGGTT	ATTATGTT	AAAATTT	PRCHRCQWTHWKD
				ATTACATACCGTTGCCGGGTCTGTGGCGCTGATATGCCCCGGGGTATGA	TTTTACGA	AAAAAA	CCPHSTNTTDGPEGT
				AGCAGCTCAAAGCCCATGTGGCCGCGAGCCACCCCGAGACCACCACAG	AAAAGTCC	AAAAAC	DRCADTITSPATAAC
				ACGCCCACGGGATGCTTGTGGAAGCAGCCACAGCCCCGAACCCTGCCTT	TGGTCTTT	GAGGAA	PQRSPCPLGSSNGCD
				ACCACTCCCCGCCTTCTCCCCGAGACAGTCAGCGCTGAGCCCTGTGACG	GAAATTCA	CAAATA	ETAPEKRQPAADLVH
				ACGAGAGCCAGCGAACGCGGTACGCCTCTCCCCACAAGCAGGCGCGTA	TTGTCTAA	AACACA	TAPFAVLVRAGPFAD
				CTCTGCACGACGCCGAGCCGCGCGATGCTTCCCGGGAACATGCCCCATC	ATTTTAAA	AATTCTG	LVRAGPFADHHQDD
				CTGCGCTGAGCCCCGTTGTCACAGGTGCCAGTGGACGCACTGGAAGGA	TAAAATTA	AAAGAT	DPLPHRSGSLGPLCSK
				CTGTTGCCCCCATTCCACCAACACGACAGATGGCCCCGAGGGCACCGAC	TAAAATTT	TTATATA	QKDPRKTHQHRHSG
				CGCTGCGCAGACACCATCACCAGCCCCGCAACGGCCGCCTGCCCGCAAC	AAAAAGA	ATTTAAA	QAGNQTHTDIPRAA
				GTTCCCCCTGCCCCCTTGGGTCAAGTAACGGGTGTGACGAGACGGCTCC	AAATTAAT	AWATAA	PSRRAAICLMANAAA
				TGAGAAGCGGCAACCAGCCGCCGATCTCGTCCATACCGCCCCCGTTCGCC	TAAAGAAG	ATCGAA	TREDLLRAATSLSEM
				GTCCTCGTCCGTGCCGGCCCGTTCGCCGATCTCGTCCGCGCCGGCCCGT	CGATGAAA	AATAAA	AAANQPTRSPTGGG
				TCGCCGACCACCACCAGGACGACGACCCCCTCCCGCACCGGTCTGGGAG	TATCTCTG	TGTTGA	EPTSQGRRGPQALA
				TCTCGGCCCCCTCTGCTCCAAGCAGAAGGACCCCCGAAAGACCCACCAG	AAATTCAA	AAACAA	DAAKRIQQIYRTNIPR
				CACCGCCACAGCGGGCAGGCCGGGAACCAAACCCACACGGACATACCC	TCAATCAA	AAAAAA	AMRKVLRTLLTAVFS
				AGAGCCGCCCCCAGCAGAAGGGCCGCAATCTGCCTGATGGCCAATGCC	TTAATCAT	AATAAT	ACLRTGHVPDLCKKS
				GCGGCCACCAGGGAGGACCTGCTGAGGGCCGCCACCAGCCTTTCCGAA	GGCGTCTC	AATAAT	RTVLIHKKGDRTDLS
				ATGGCGGCCGCGAACCAGCCTACCCGCTCGCCCACTGGAGGTGGCGAG	AGCGAGT	AATAAA	NWRPLSMGDTIPKLF
				CCCACCTCACAGGGTAGGCGCGGACCGCAAGCACTGGCAGACGCAGCG	GCACGTAT	AACACA	AAVMADRLTAFLTN
				AAAAGGATCCAACAAATATACAGGACCAACATACCTCGCGCCATGAGAA	TTGCCTAC	ATAACA	GGRLSEEQKGFLQHE
				AAGTCCTGAGAACACTGCTCACGGCAGTGTTTAGCGCCTGCCTGAGGAC	CCCTTCGT	CTCACCC	GCHEHNFVLGQVLEE
				AGGTCATGTCCCCGATCTGTGTAAAAAGTCCAGAACGGTCTTAATCCAC	GGGACCAT	GGCCTG	SRRQGKDLVMGWL
				AAGAAAGGSGACAGAACTGACCTGTCAAATTGGAGGCCTCTTTCCATGG	TCCGGTGC	CCCCAG	DLSNAFGSIPHATIM
				GTGACACCATCCCCAAATTGTTCGCAGCCGTCATGGCGGACAGGCTGAC	TCCGTATG	AGGCAG	DAVAGMGIPSRIRTII
				GGCGTTCCTCACTAACGGAGGAAGGCTCAGCGAGGAGCAGAAGGGCTT	CATGGATG	GTAAAC	HQLATGAATTAKTID
				CCTCCAGCACGAAGGCTGCCATGAACACAATTTTGTTCTTGGCCAAGTG	CGTCCGGG	ATTTACT	GMSEEIPIEAGVRQG
				CTGGAGGAGAGCAGACGCCAAGGCAAGGACCTCGTCATGGGCTGGCT	ATGCATCC	GGCCAT	CPASPILFNIAIERVLR
				GGACCTGTCCAACGCGTTCGGGTCGATTCCGCATGCCACCATCATGGAC	CACTAGGT	ATGGCT	KIKTVNAGYLLYGSRI
				GCGGTCGCCGGTATGGGGATCCCTTCGAGGATCCGGACCATAATCCACC	CGCTGGGC	TTTTTTT	SPLAYADDLVLIASSP

		AGCTGGCCACCGGCGCCGCGACCACCGCCAAAACCATTGATGGCATGTC	GAATACGG	TAA	EEMRSLLRAADDAAI
		GGAAGAGATCCCGATCGAAGCGGGGGTCAGACAGGGCTGCCCAGCCA	CACATACG	(SEQ ID	EAGLHFNPKKCATLH
		GCCCAATCCTCTTTAACATCGCAATAGAGCGGGTACTTCGCAAAATCAA	CTGCGGCA	NO:	LTGKKSSRRAVQTGF
		AACCGTCAACGCGGGGGTACCTGCTCTATGGGAGCCGCATTAGCCCGCTG	TATCAGCA	1307)	LVRGTPIPAMTEGDA
		GCGTACGCCGATGACCTGGTGCTAATTGCGAGCTCCCCAGAGGAGATG	CATAACCC		YEYLGIPLGLKKNQTP
		AGGTCCTTGCTGCGTGCTGCGGACGACGCCGCAATAGAAGCCGGTCTG	GGCGCCAC		RAAMEAIVGDIAKID
		CACTTCAACCCCAAGAAGTGCGCGACCCTACACCTCACGGGGAAGAAAT	CCACAAGT		DSLLAPWQKIDAART
		CCTCGCGGAGGGCAGTGCAGACCGGCTTCCTCGTCCGTGGCACGCCAAT	GGTTATTA		FVAPKLDFVLRSGATL
		ACCGGCCATGACAGAGGGGGATGCCTACGAATACCTCGGCATCCCCCT	CATACCGT		RAPLRHLDTVIKKHIK
		GGGTTTAAAAAAAAACCAAACACCCAGGGCAGCGATGGAAGCGATAGT	TGCCGGGT		KWLYLPQRASAEVVY
		TGGGGACATAGCCAAGATAGATGACTCGCTGCTCGCCCCGTGGCAAAA	CTGTGGCG		TPLKKGGAGILPSSILA
		GATCGACGCGGCCCGCACCTTCGTGGCACCGAAGCTTGACTTCGTGCTA	CTGA (SEQ		DVLTIAQAHRMVSCP
		CGAAGTGGCGCCACCTTGCGGGCCCCGCTGCGTCATCTGGATACAGTCA	ID NO:		GEVVSRIASEGLREAV
		TTAAAAAACACATTAAAAAAATGGCTGTATCTGCCGCAGAGGGCGAGCG	1184)		KRKINREPSGDEMAH
		CGGAGGTAGTATACACCCCGCTGAAGAAAGGTGGAGCGGCGCATACTAC			FLSGSTLSGETASFGD
		CTTCATCTATATTGGCTGATGTCCTAACTATCGCCCAGGCTCACCGCATG			AGFWSRVRMATKR
		GTGTCCTGCCCTGGGGAGGTCGTCTCCCGGATTGCAAGTGAGGGCCTG			QAVHLGVRWAWRG
		AGAGAAGCGGTAAAGCGAAAAATAAACCGGGAGCCATCCGGCGACGA			GELLVESRGQRNRPV
		GATGGCCCACTTTCTCTCAGGCTCCACTCTATCCGGGGAGACAGCCAGC			ATDSNSRSQLIQRLR
		TTTGGCGACGCCGGATTCTGGTCGAGGGTGAGGATGGCCACCAAAAGG			CAAQDEFLTILINKPD
		CAAGCTGTGCATCTGGGGGTGCGTTGGGCCTGGAGAGGAGGTGAGCTA			QGKVAKLSTLTPVSN
		CTGGTCGAGAGTAGAGGACAAAGAAACCGACCAGTGGCCACCGACTCG			AFIRDGSFTRFADWR
		AACTCCAGGTCCCAACTCATCCAACGTCTCAGGTGCGCAGCTCAGGATG			FIHRARLGVLPLNGAI
		AGTTCCTGACCATCCTCATAAATAAACCCGACCAGGGGAAGGTGGCGAA			RWGSGDKRCRVCGY
		GCTCTCCACGCTAACCCCAGTCAGCAACGCGTTCATACGCGACGGTAGC			QLESVPHVLCHCMH
		TTTACCAGGTTTGCTGACTGGCGGTTTATCCACAGAGCCCGACTGGGAG			HSNAMQQRHNAV
		TCCTCCCACTCAACGGAGCGATCCGATGGGGCAGCGGCGACAAGCGCT			MDRLAKAGSRLGTP
		GCCGGGTCTGTGGATATCAGCTGGAGAGCGTTCCACACGTGTTGTGCCA			RVNCRVEGVAEDMA
		CTGCATGCACCACTCAAACGCAATGCAGCAGAGGCACAACGCGGTGAT			ALRPDLVWRDERSRK
		GGATCGCCTCGCCAAGGCTGGCTCACGGCTGGGGACCCCCAGGGTGAA			IVIVDVTVPFENGAEA
		CTGCCGCGTGGAAGGGGTCGCCGAGGACATGGCGGCCCTCAGGCCGG			FDNARGEKEEKYRPL
		ACCTGGTATGGCGCGACGAACGGAGCAGAAAAATCGTCATAGTTGACG			AEALRAMGYQVKLE
		TGACTGTTCCGTTCGAGAACGGGGCTGAAGCGTTTGATAACGCGAGGG			AFIVGALGSWDPKNE
		GCGAGAAAGAAGAAAAATACCGCCCCCCTAGCTGAAGCCCTCGCGCGCCA			RVLKTLGVSRFYAGL
		TGGGATACCAGGTAAAACTGGAGGCATTCATTGTCGGAGCCTTGGGCTC			MRRLMVADTIRWSR
		GTGGGACCCTAAAAACGAAAGGGTCCTTAAGACTTTGGGTGTCTCCAG			DIYVEHVSGIRQFTLP
		GTTTTATGCTGGCCTGATGCGCAGACTGATGGTGGCCGACACCATCAGG			SGAPSN (SEQ ID
		TGGTCCCGGGACATTTATGTGGAGCATGTATCCGGGATCAGGCAGTTCA			NO: 1429)

			CCCTGCCAAGTGGAGCTCCCTCCAACTAAAAAATTAAAATGCCTTAAAA ATAAATAAATATATCAAAATTTAAAAAAAAAAAAACGAGGAACAAATAAAC ACAAATTCTGAAAGATTTATATAATTTAAAAWATAAATCGAAAATAAAT GTTGAAAACAAAAAAAAAAAATAATAATAATAAAAAACACAATAACACT CACCCGGCCTGCCCCAGAGGCAGGTAAACATTTACTGGCCATATGGCTT TTTTTTTAA (SEQ ID NO: 1061)	CGCTGGAA AGACGGA	TGAACC GCCCCC	MSGPRQAAADPRPS TDPRRQRDSQSPEPR
NeSL	Utopia-1_Gav	.	Gavialis gangeticus	CGCTGGAAAGACGGAGAACCGCTTCTTTTTCCTGCGCCCGGCCTGGTAT TGCACTTCCTCCAGGACCAGCGCCAACCTAGTCCGGCAGACTGCCGGAA	GCCTGGAA AGACGGA	TGAACC GCCCCC	MSGPRQAAADPRPS TDPRRQRDSQSPEPR
				TAATAGCCTCAGAAAGAGAGCTGGCTAGCAGCCCTCTTTTCTTTCCTCCG GTGCAGCGTGGGTTCTTGTCAGTCCTGATGGGCTAGGGAAGGCGGTGC	GAACCGCT TCTTTTTCC	CCTCCGC GCCAGA	LTRAASRRRTPDPED APRTTAEHPERRRTP
				CGCCAGTACGTCCGAAAGAGCGCCGGTTGCGCGAGCGACCGCGCCGCT CAGGCGAGTAGCCCAAGGGTCTTACGGTTCGCCGGACCCGATAACGCG	TGCGCCCG GCCTGGTA	CGGACC TTCACTT	PDPRGPSATTAGPER RRPPDPGGPEDDPPE
				AAAGCCCCGACTCGGGCCAGTAGCCGAAGACCNTGGGCCTCCCTCCCCA GGTCGGAGTAGGCGAACGCCCGTGCTCGGAGGACGGAACGTGGACAA	TTGCACTT CCTCCAGG	CACTCC GAGAGG	GLPTLVEEPRTPPTPD PPDGRPRRGCRRGS
				AACACCCCCAGGTCCCAATGACGCCCTGATCCACTGACAAGAACGCTCG AGGCACNCCAGGAGACCCCCAGCTAGGGCAGACCGCCGACCACGGGTC	ACCAGCGC CAACCTAG	ATTCTTC GACCAC	AHVPPLPPPCEAAVP DLPPAKAVQVAQRH
				GCGGAGGACCCTCCCAGGAGGGTGGACCAGCGAACCCGAGTCGGCGA CGAACCCCGACGCACCCCCCCCGCGATGTCGGGACCCCGACAGGCGGC	TCCGGCAG ACTGCCGG	GGACGA CCCCGCT	EQTPTALPPAAPSVLL LPLRHRVRGPEAPEE
				GGCGGACCCCCCGGCCATCGACCGACCCCCGGAGGCAGAGAGACTCTCA GAGCCCGGAACCCCGGCTGACGAGAGCCGCCTCCCGGCGGAGGACCCC	AATAATAG CCTCAGAA	CCACCC GAAGAG	PPQGMPGPRGREET RHAGEVRRPTTRAAA
				GGACCCCGAGGACGCCCCCCGGACGACGGCGGAGCACCCCGAGCGAC GGCGGACTCCTCCGGACCCCCGCGGNCCCTCGGCGACGACGGCGGGCCC	AGAGAGCT GGCTAGCA	GACCCC CGCGAT	RRPARPAAPPATPPD QTSGDRPTERPAPAT
				CCGAGCGGCGACGNCCCCCGGACCCCGGCGGTCCCGAGGACGACCCCC CCGAGGGCCTCCCCACNCTGGTGGAGGAGCCCCGAACCCCCCCGACAC	GCCCTCTT TTCTTTCCT	GAGACT CTATAC	PPRRSAPRDPRPDVT PRPDGPPPGPPGPP
				CGGACCCCCCCCGACGGACGACCCAGGCGAGGGTGCAGACGCGGCAGC GCTCACGTTCCTCCCCTTCCCCCTCCCTGCGAAGCTGCTGTGCCCGACCT	CCGGTGCA GCGTGGG	GGACTG AGGCAC	DAPDPPRIPEPPGEP EPPGALQLPSVPGSP
				GCCACCCGCCAAGGCAGTACAGGTCGCACAACGACATGAACAAACACC TACGGCGCTTCCACCAGCTGCGCCTAGCGTTCTACTGCTCCCTCTGCGGC	TTCTTGTC AGTCCTGA	TTCCTTC GAACCA	GAETSAQQRMPTPR QALWLEELSRATAFE
				ACCGAGTACGAGGCCCTGAAGCTCCTGAAGAACCACCACACAAGGTATGC CAGGGCCCCGGGGCCGAGAGGAGACCCGGCACGCTGGTGAGGTCCGC	TGGGCTAG GGAAGGC	CTTCCTC CACCATT	AFEASVARLTEELSAA ARPGQPRRGADNGP
				CGCCCCACGACCCGGGCCGCGGCGCGAAGGCCCGCCAGACCGGCCGCC CCGCCGGCGACCCCACCGGACCAGACCTCCGGGGACCGCCCCGACGGAG	GGTGCCGC CAGTACGT	GCGGAC CATTGTA	TTRRDHRPQPQRRP RRQRYDPAAASRIQK
				AGACCCGCCCCGGCGACGCCACCACGCAGGTCTGCACCCAGGGACCCC CGACCGGACGTGACGCCCCGACCGGACGGCCCCCCTCCCGGACCCCCG	CCGAAAGA GCGCCGGT	ACGGGT TTGTGT	LYRANRPKAAREILEG PSAFCQVPRETLFNYF
				GGGCCGCCCGACGCCCCCGACCCGCCGAGGATCCCGGAGCCGCCCGGN GAGCCCGAGCCGCCGGGAGCCCTCCAGCTCCCGAGCGTGCCGGGGTCT	TGCGCGAG CGACCGCG	GTATCTA TCTCCTT	SRVFNPPAEAAAPRP ATVEALTPVPPAEGF
				CCGGGTGCGGAGACCTCCGCACAGCAGAGGATGCCCACCCCGCGGCAA GCCCTCTGGCTGGAGGAGCTCTCCCGGGCCACCGCCTTCGAGGCCTTCG	CCGCTCAG GCGAGTA	TCTCTCT CAGCGT	EEAFTPREVEARLKRT RDTAPGRDGIRYGLL

	AGGCCTCGGTGGCCCGGCTCACGGAGGAGCTCTCGGCGGCCGCCCGGC	GCCCAAGG	CGCGAA		KKRDPGCLVLSVLFN
	CCGGCCAGCCCCGGAGGGGCGCCGACAACGGACCGACGCGACGA	GTCTTACG	CCCCCTC		RCREFRRTPAAWKR
	GACCACAGACCGCAGAGGCGACCCAGGCGCCAGCGCTACGAC	GTTCGCCG	CCCCACC		AMTVLIHKKGDPTDP
	CCGGCGGCAGCCTCCCGGATCCAGAAGCTGTACCGGGCCAACCGCCCC	GACCCGAT	CCCCACC		GNWRPIALCSTVAKL
	AAGGCGGCGAGAGAGATCCTGGAGGGACCCTCGGCTTTCTGCCAGGTC	AACGCGAA	CCCGGG		YASCLAARITDWAVT
	CCCCGGGAGACTCTGTTCAACTATTTCAGCAGGGTCTTCAACCCCCCGGC	AGCCCCGA	CTTAGTT		GGAVSRSQKGFMST
	AGAAGCCGCCGCCCCACGCCCCGCGACCGTCGAAGCGCTGACCCCCGTC	CTCGGGCC	GGCTAA		EGCYEHNFTLQMAL
	CCCCCGGCAGAGGGGTTCGAGGAGGCCTTCACGCCGCGGGAAGTGGA	AGTAGCCG	CATTGTA		DNARRTRKQCAVA
	AGCCCGCCTGAAGAGGACCAGGGACACCGCCCCCGGCAGGGACGGCAT	AAGACCNT	TCTCCTG		WLDISNAFGSVPHRR
	CAGGTACGGTCTCCTNAAGAAACGTGACCCGGGCTGCCTCGTTCTTTCT	GGGCCTCC	TAACCTA		IFGTLRELGLPDGVID
	GTTCTCTTCAACAGGTGCAGAGAGTTCCGGCGCACGCCCGCCGCCTGGA	CTCCCCAG	GTCGCG		LVRELYHGCTTTVRA
	AGAGGGCCATGACGGTCCTCATCCACACAAGAAGGAGACCCGACCGACC	GTCGGAGT	TTCCCCT		TDGETAEIPIRSGVRQ
	CGGGCAACTGGAGACCCATCGCCCTGTGCTCCACCGTGGCCAAGCTGTA	AGGCGAA	CCTCACC		GCPLSPIIFNLAMEPL
	CGCCAGCTGCCTGGCGGCCCGCATCACCGACTGGGCGGTGACCGGCGG	CGCCCGTG	CCCATCC		LRAVAGGPGGLDLY
	GGCCGTCAGCCGGAGCCAGAAGGGCTTCATGTCGACGGAGGGCTGCTA	CTCGGAGG	CTCTATT		GQKLSVLAYADDLVL
	CGAACACAACTTCACCCTCCAGATGGCCCTGGACAACGCCCGGAGGACC	ACGGAAC	GTTAGT		LAPDATQLQQMLDV
	AGGAAGCAGTGCGCGGTGGCGTGGCTGGACATCTCCAACGCCTTCGGC	GTGGACAA	CCCTCGC		TSEAARWMGLRFNV
	TCCGTGCCCCACCGCCGCATCTTCGGCACCCTCCGCGAGCTGGGCCTAC	AACACCCC	TCGGGC		AKCASLHIDGRQKSR
	CGGACGGCGTCATCGACCTGGTGCGAGAGCTCTACCACGGNTGCACCA	CAGGTCCC	GATCTG		VLDSTLTIQGQAMRH
	CGACCGTCCGCGCCACCGACGGAGAGACCGCGGAGATCCCCATCCGGT	AATGACGC	TATTTCC		LRDGEAYCHLGTPTG
	CGGGGGTGAGGCAGGGCTGCCCCCTCAGCCCCATCATCTTCAACCTGGC	CCTGATCC	CTATCG		HRAKQTPEETINGIV
	CATGGAACCGCTCCTTCGAGCCGTGGCGGGCGCCCGGCGCGGCTCGA	ACTGACAA	GCTTTGT		QDAHKLDSSLLAPW
	CCTGTACGGCCAGAAGTTGAGCGTCCTGGCCTACGCCGACGACCTCGTC	GAACGCTC	CATCTTT		QKIDAANTFLIPRVAF
	CTCCTCGCCCCCGACGCCACCCAGCTGCAGCAGATGCTGGACGTGACGT	GAGGCAC	TTTCTGG		VLRGSAVPKTPLKKA
	CCGAGGCGGCCAGGTGGATGGGCCTGCGCTTCAACGTCGCCAAGTGCG	NCCAGGA	ATTCCCG		DAEIRRLLKKWLHLPL
	CCTCCCTGCACATCGACGGCAGGCAGAAGAGCCGCGTCCTGGACTCCAC	GACCCCCA	ATCCTAA		RASNEVLHIPYRQGG
	CCTCACGATCCAGGGCCAGGCGATGAGGCACCTGCGCGACGGCGAGGC	GCTAGGGC	ACATTTA		ANVPRMGDLCDIAV
	CTACTGCCACCTGGGGACGCCCACCGGCCACCGGGCCAAGCAGACGCC	AGACCGCC	CTAATA		VTHAFRLLTCPDATV
	GGAGGAGACCATCAACGGGATCGTGCAGGACGCCCACAAGCTGGACTC	GACCACGG	AAAGTC		SIIAASALEETARKRIA
	GTCCCTGCTGGCCCCCTGGCAGAAGATAGACGCGGCGAACACCTTCCTC	GTCGCGGA	AATCTGT		RQPTGRDLATFLSGS
	ATCCCCCGCGTTCGTCCTGAGAGGCTCGGCGGTCCCCAAGACCC	GGACCCTC	TCTTT		LEGEFGRDGGDFASL
	CCCTCAAGAAGGCGGACGCCGAGATCCGGCGGCTGCTCAAGAAGTGGC	CCAGGAG	(SEQ ID		WSRARNATRRLGKRI
	TGCACCTGCCGCTGAGGGCCAGCAACGAGGTCCTGCACATCCCCTACCG	GGTGGACC	NO:		GCAWTWTEECRELG
	GCAGGGAGGCGCCAACGTCCCCCGCATGGGAGACCTCTGCGACATCGC	AGCGAACC	1308)		VSLQPAPHADRVTVT
	GGTGGTCACCCACGCCTTCCGCCTCCTGACCTGCCCGGACGCGACGGTA	CGAGTCGG			PRTRTFLERFLKDAVR
	AGTATCATCGCCGCCAGCGCCCTCGAGGAGACCGCCCGCAAGAGGATC	CGACGAAC			NKYAGDLRAKPDQG
	GCGAGGCAGCCGACCGGACGNGACTTGGCCACCTTCCTCAGCGGCTCG	CCCGACGC			KVFDVTSKWDASNH
	CTGGAGGGCGAGTTCGGCCGAGACGGCGGGGACTTTGCCTCGCTGTGG	ACCCCCCC			FMPSGSFTRFADWR

			AGCCGAGCCCGCAACGCCACGCGCCGCCTCGGGAAGCGCATCGGCTGC GCCTGGACCTGGACCGAGGAGTGCCGGGAGCTGGGAGTCTCCCTGCAA CCAGCCCCGCACGCCGACCGCGTCACCGTGACGCCCCGCACGAGGACCT TCCTGGAGAGGTTCCTGAAGGACGCCGTCCGAAACAAGTACGCCGGCG ACCTGAGGGCCAAACCCGACCAGGGCAAGGTCTTCGACGTCACCTCGA AGTGGGACGCTAGCAACCACTTCATGCCCAGCGGGAGCTTCACGCGCTT CGCGGACTGGCGCTTCCTCCACCGCGCCCGCCTCAACTGCCTGCCTCTGA ACGGGGCCGTGCGCTTCGGCCACCGGGACAAGAGGTGCCGACGGTGC GGCTACGCGGCAGAGACCCTCCCCCACGTGCTGTGCAGCTGCAAGCCG CACGCCAGAGCCTGGCAGCTCCGCCACAACGCTGTCCAGGACCGCCTG GTGAGGGCCATCCCGGCCGCGGCGGGGGAGATCTCCGTGAACCGCACC GTCCCGGGCTGCGAGAGCCAGATGCGACCCGACATAGTCATCACCAAC GAAGAGGCCAAGAAGGTCGTGATCGTGGACGTCACCATCCCCTTCGAG AACCGGCGCCAAGCCTTCACCGACGCCCGGGCTCGCAAGCGGGAGAAG TACGCCCCGCTGGCCGACACCCTGAGGGGCCGCGGCTACGACGTGACG GTCGACGCGCTCATCGTGGGAACGCTCGGAGCCTGGGACCCCAGCAAC GAGAGCGTCCTGCGTGCCTGCCGCGTCTCCCGCCGCTACGCCAAGCTGA TGCGCTGCCTCATGGTGTCCGACACCATCCGTTGGTCCCGCGACATCTAC GTGGAACACATCACGGGCCACCGCCAGTACTCCGACCCCACCAGACGA GCCGCCGCCGGACCGGACCCGGAGGGGACCGCCTGAACCGCCCCCCCT CCGCGCCAGACGGACCTTCACTTCACTCCGAGAGGATTCTTCGACCACG GACGACCCCGCTCCACCCGAAGAGGACCCCCGCGATGAGACTCTATACG GACTGAGGCACTTCCTTCGAACCACTTCCTCCACCATTGCGGACCATTGT AACGGGTTTGTGTGTATCTATCTCCTTTCTCTCTCAGCGTCGCGAACCCC CTCCCCCACCCCCCACCCCCGGGCTTAGTTGGCTAACATTGTATCTCCTG TAACCTAGTCGCGTTCCCCTCCTCACCCCCATCCCTCTATTGTTAGTCCCT CGCTCGGGCGATCTGTATTTCCCTATCGGCTTTGTCATCTTTTTTCTGGAT TCCCGATCCTAAACATTTACTAATAAAAGTCAATCTGTTCTTT (SEQ ID NO: 1062)	CGCG (SEQ ID NO: 1185)	ATCTACTATC TC (SEQ ID NO: 1186)	TGAATA GCATTTA TATTGTG TTCCAAA CAACAT ACTCATT ATTATAT CTAAAC ATTTTTT	FLHRARLNCLPLNGA VRFGHRDKRCRRCG YAAETLPHVLCSCKP HARAWQLRHNAVQ DRLVRAIPAAAGEISV NRTVPGCESQMRPD IVITNEEAKKVVIVDV TIPFENRRQAFTDAR ARKREKYAPLADTLR GRGYDVTVDALIVGT LGAWDPSNESVLRA CRVSRRYAKLMRCL MVSDTIRWSRDIYVE HITGHRQYSDPTRRA AAGPDPEGTA (SEQ ID NO: 1430)
NeSL	Utopia-1_LV	AGCV013581 06	Lytechinus variegatus	ATCTACTATCATGTCTTGTCCAAGAGAGGGAAGCGATCACCTCGGTCCT GATCCTGAGACACCCGCCCTCCATCAGGGTTCTGACATCCGGGTTACCA GTTCTCGCCTTCGAGGTTCCCGCGGAAAGAGTTCTCGCCAACCAAGCTC CCGACACCAAGTTCCTGCCAGCGAGGCTTCCGCCACCGCCCAGCAGACT GCCGCCAACGAGTGCCAGGTGTGTGGATCTTCCTTCGCCACCTCCAGTG GACTCCGCCGCCACATGGCCAGGCTTCATCGAGCTGCCTCTGCGGATCC TGAGGGTGCTGCGCCGGCTTCCATCACAGAGATTTTCGACTACCCCTTG CCTTCCCGGTGGAAATGCTCGGCATGCTCGGAGAACTTTTTCAACCAGC AGACCCTCAAGCGACACCAGACCAGGCATCATCCAGCTACCACCTTCGC			

		GTATGCCTTTCGGTGTTCGTCATGCCGGTCCGAGTTCGACTCAGCACGG		TTTCTGT		QTRHHPATTFAYAFR
		AGGGCTGCGAACCATTGGCAGGTCCACAAGAAGGAGCGATCTCAACTC		TCCTGAC		CSSCRSEFDSARRAA
		TCTGGCACCGAGCCCCAGGCCTCTTCCCAAGCCAGAGTTAGCATGGCTC		AATCTAC		NHWQVHKKERSQLS
		ATTCTCCTCCACCTCTGCCCAACACTTCTTGGGCGGAGCTCGCCTCGAAT		GTAAAG		GTEPQASSQARVSM
		CCTGCCGAGATACCTTCCTTCGTCTGGGAGTCTCCTCCCAAGAACCGCCC		TCTGCTA		AHSPPPLPNTSWAEL
		CTCGGTTGAGGAGTTCGGTTCGTCTCTGCCAACTGATGTTACGATGATG		ACCAAC		ASNPAEIPSFVWESP
		TCTCAAAGTCCTCCACCGCAGGTACAGTCGTCTCCTGTCCCTGCTCTGAC		TGGCAT		PKNRPSVEEFGSSLPT
		TCCTCTTTCACCCGCTGCCACTGCCTCCAGTTCTCCTCCAGGGGCTGCAA		GATGAA		DVTMMSQSPPPQV
		GGCAGCTGACCCCTCCTACACAGACTAACACCCCAGTCACCCAGAGGGC		ATAAGA		QSSPVPALTPLSPAAT
		TCGCCTGCAACCTGAAGCAGACGTCGTACCTGAACTCCCTCCTTCAGTCA		TAAAAT		ASSSPPGAARQLTPP
		CCGAGCACCCTGTGTCTGACGCTCAACACTGGGTTGATGCTGTATCCTCT		CCCCTTA		TQTNTPVTQRARLQ
		GCATCAGATTGGTCTGAGTTTGAAGCAGTATGTGATCAATTTGTCATCCA		CACATTA		PEADVVPELPPSVTE
		CGCTGTTGCTGTTTCCCGTCCCAATCTTGCTCGACCCCAGCAGCAAGATA		ATTTCTT		HPVSDAQHWVDAV
		GGCAGAGATCTGGTGACCACCCTCCTAGACAGCAAAGAGGTCAGCATC		GTCACA		SSASDWSEFEAVCD
		GACCAACCTTCGATGTCCGTGAGGCAAGTAGAATCCAGAAGCTCTATCG		TCATAAT		QFVIHAVAVSRPNLA
		TACCAGCAAGAAAAGAGCCATCAGACACATACTGAAAGAGAAATCACC		GCTTTGT		RPQQQDRQRSGDH
		TTCCTTCTCTGGTTCCGAGTCAGACGTCTTAGACTTCTTCCGCGAGGTGT		CAAAGC		PPRQQRGQHRPTFD
		ATTCTGCTAAAGAAGTTGACGAGGAAGCAGTTGGTAAACTAGCATCCTC		AATGTC		VREASRIQKLYRTSKK
		GCTCTTCGATGTCCCTCAAGGTGATGACTCTGCGACATCTCTGTCTCTGC		CTACATA		RAIRHILKEKSPSFSGS
		CCACGTCAGCGAAGGAGATCGGAGCAAGGCTGTCAAGGATGACAAACT		ATATCTC		ESDVLDFFREVYSAKE
		CTGCCCCCGGGAAGGATCGCTTGGAGTACAGACACATTCGACGTGCGG		GATGTC		VDEEAVGKLASSLFD
		ACGGGTCCTTCAGCATCTCTGAGGCCATCTTTAACAAATGCCTGGCTGA		ACCCCA		VPQGDDSATSLSLPT
		AGGTCGGATCCCAGCTCCTTGGAAGACAGCATCTACCATCCTACTTCACA		ATTAATT		SAKEIGARLSRMTNS
		AGGCTGGCCCCACGGATGATCCCGCCAACTTCCGCCCAATCGCCTTACA		TTACATC		APGKDRLEYRHIRRA
		GTCATGTCTCTACAAGCTTTTTATGGCTGTACTTGCGGACCGGCTGACCA		CTTCGG		DGSFSISEAIFNKCLA
		AGTGGGCCTGTGAGAACCAGTACCTCAGCCCCGAGCAGAAGTCCGCTC		TAACCTT		EGRIPAPWKTASTILL
		GCCCCTGCGAGGGGTGCTTCGAGCACTCCTTCCTCTCTCAGCTGCCCTG		TATACC		HKAGPTDDPANFRPI
		AAGGACTGCAGGAGAAACCAGAAGACCATCTGCATCGGTTGGTTGGAC		GTTGGA		ALQSCLYKLFMAVLA
		CTTAGGAATGCATTTGGAAGCATTCCTCATCCTGTCATCAAGATCGTCCT		TCAACAT		DRLTKWACENQYLSP
		GTCCAGTCTGGGTGTCCCTGATTCGCTTGTTACCCTCCTCATGGATGCCT		ATATGA		EQKSARPCEGCFEHS
		ACAATGGTGCGTCAACCTCGTTCACGCTGACCGGGGGCCAGACCGACA		TTTGTAA		FLLSAALKDCRRNQK
		CCGTACCCATCAGATCAGGGGTGAAGCAAGGCTGCCCGATGTCCCCAAT		AACTGTT		TICIGWLDLRNAFGSI
		CCTCTTCAACCTGGCCATCGAACTTATCATCAGGGCAGTCAAGAAGAAT		ATTTCTG		PHPVIKIVLSSLGVPD
		GCATCAGACAACCATCTCGGAGTGACTGTCCAGGGCAAGAACCTCTCCA		AGTTTTT		SLVTLLMDAYNGAST
		TCCTGGCCTATGCTGATGACCTAGTGCTGCTCAGCCGAGACACTGAAGG		TCTATGC		SFTLTGGQTDTVPIRS
		CCTCCAATCCCTCCTTCAAGTTGCTGGCTCTTCTGCATCTACCCTTCAGAT		TAATAA		GVKQGCPMSPILFNL
		GCAGTTTAAGCCCCAGAAGTGTGCAACACTCACCCTTGACTGCAAGCGT		A (SEQ		AIELIIRAVKKNASDN
		GGTACCAATGTTAGGCAGTCTGCTCACCATATCCAAGGGGCTGCCATCC				HLGVTVQGKNLSILA

		CATCGCTTACTGAAGAGGAGCGTTACCGCTACCTTGGGGTCCCCATTGG	ID NO: 1309)	YADDLVLLSRDTEGL
		ACTCCCAAGGCTCACAAGTCTACAGGAGAGCTCTAGGAAGCTGTCGTCA		QSLLQVAGSSASTLQ
		GACATCGAGACCATTTCATCCTCCCTTCTAGCGCCTTGGCAGAAACTGGA		MQFKPQKCATLTLDC
		TGCAATCAAGACCTTTGTGATACCCATCCTTCAGTATACCTTGCGGGCCA		KRGTNVRQSAHHIQ
		CAGAATACCTCAAGTCCGACTTGAAACCCCTACGAGCAGCCATTATCAA		GAAIPSLTEEERYRYL
		GCACGTGAAGAAGATCTGCCATCTTCCGGTAAGATCATCCAACGCCTTC		GVPIGLPRLTSLQESS
		GTCTTCGCGTCCCGACCCTCCGGTGGTTTGGCTTTTGTTGACCCAGGAGT		RKLSSDIETISSSLLAP
		GGATGCCGACATCCTTGTCGTAACCCAAGCCGTCCGTACCCTGGCATCG		WQKLDAIKTFVIPILQ
		GATGATGACACGGTGAGAGCCGTTGCTCTAGGACAGCTGACCTCAGTG		YTLRATEYLKSDLKPL
		GTCCACAGGACGGTCCATTCTGCACCATCAGATGACTGCATCGACAAGT		RAAIIKHVKKICHLPV
		TCTTGTCCGGAAGCAGTGAGGGCCCTCTGGCAAACTCAGGGAACAGTG		RSSNAFVFASRPSGG
		GCCAGGCATCATCTCTTTGGTCACGTACAAGGGCTGCATCCCGTCGCCT		LAFVDPGVDADILVV
		GAAGATAAGGATCGTCGGAGCCTCGAGCGGCGACATCAAGGTCGAGA		TQAVRTLASDDDTVR
		GTGGAGGCCGTGCCATCCCCTCCAAGAAAGTGACTGCAGGTCTCCGCTC		AVALGQLTSVVHRTV
		CGACCATCACAACGAGATGTCAGAGAAGCTGCGGTCCCTTCCAGACCAG		HSAPSDDCIDKFLSGS
		GGCAAGGTAGCCAGGGCACTCTCCTTGGATAGTTTCGCCAACGCAACAT		SEGPLANSGNSGQAS
		CATGGCTGACATCCGGGAGTTTCATCCGATTTTGCGACTGGAGGTTCAT		SLWSRTRAASRRLKIR
		CCACCGTGCTCGCCTAAACTGTCTACCTACCAACGCTGCTGTCCGGCGGT		IVGASSGDIKVESGGR
		GGAAACAGAACGCCAACACCAAGTGCCGAAGGTGTGATCACCAGTTGG		AIPSKKVTAGLRSDH
		AGACGTTGCCCCACATTATTAACAACTGCCGTCCCAACATGGTACCGATC		HNEMSEKLRSLPDQ
		CGCCGGAGGCACAACTCCATCCAAGAACGCCTGGTGAAGGCGATTCAC		GKVARALSLDSFANA
		TACGGTGACATCTACCAGGACCAGCATGTCCCCGGCGACCCGAACCCAA		TSWLTSGSFIRFCDW
		GGGAGCGACCCGACATCACAGTGGTTGAGGGTAACAAGGTGACCATTG		RFIHRARLNCLPTNA
		TCGACATCACCATCCCATTCGACAATGGTCCGGATGCCCTCTCCACGGCT		AVRRWKQNANTKCR
		GCCAACGCCAAAGTCATGAAATATGACACACTGCGACAAGAGCTGGCC		RCDHQLETLPHIINNC
		AGCCGTGGTATGGATGTCGAGGTCCACGCTTTTGTAATCGGCTCCCTTG		RPNMVPIRRRHNSIQ
		GGTCGTGGCACGGGGACAACGAGCGTGTTCTGGGACGTCTAGGTATCT		ERLVKAIHYGDIYQD
		CTCGCCGATACAGAACACTGATGAGGCGCCTGTGCTGCATAGATGCGAT		QHVPGDPNPRERPDI
		CAAAGGCAGCCGGGATATCTATATCGAACATGTTACCGGTCACCGGCAG		TVVEGNKVTIVDITIP
		TATTGAATAGCATTTATATATTGTGTTCCAAACAACATACTACTCATTATTATATC		FDNGPDALSTAANA
		TAAACATTTTTTTTTTCTGTTCCTGACAATCTACGTAAAGTCTGCTAACCAA		KVMKYDTLRQELASR
		CTGGCATGATGAAATAAGATAAAATCCCCTTACACATTAATTTCTTGTCA		GMDVEVHAFVIGSL
		CATCATAATGCTTTGTCAAAGCAATGTCCTACATAATATCTCGATGTCAC		GSWHGDNERVLGRL
		CCCAATTAATTTTACATCCTTCGGTAACCTTTATACCGTTGGATCAACATA		GISRRYRTLMRRLCCI
		TATGATTTGTAAAACTGTTATTTCTGAGTTTTTTCTATGCTAATAAA (SEQ ID NO: 1063)		DAIKGSRDIYIEHVTG HRQY (SEQ ID NO: 1431)

NeSL	Utopia - 1_NVit	.	Nasonia vitripennis	CCATTCCTTCGTACGGGTTTTCGTGCCGGCATAGCCGGGTGGGAGACTC GCGCGGGGGAGGTCATATCTCACCACCATCCTCGAGCTTTTTGCTGCAC CTGATCCGGGMTAACTGGCCGAGAAGTTAAATGTATTACTGTTAATGTA CTCATGGAGCAACAACCACACACACCAAAGCGATCCGGGAGGGGGGACTTC ATCGTGATACTCCTCCCGCAAAGTGACGACGAGACACTTTGCTGTCCACT CTGCGTGGGACGCGGACGGTACTCCGGCAAGACGCGGGTTGAGTGTCT CAATAGACATGTCAAAGAGGTGCACCCAGACCTAACCACCACCTTTAGG TGCTGGGGGTGCGGGTTCGCGGCTCCCGGGGATAAGAAATATCCCCGG AAGATCGTGACCCAGCACTGCGCCACCTGCGTCCCCGAGGTTTCGTCCG CGCCCTCGGGCAGAGTGGACGGAGAGAGGAGAGTTAACACCCGCCGTC GGTTGGGTATTGCCGCGGCGGCAACCGAGGCCAGTCCCGTGAGACGTACTC GTCGTAACGGCCTGGCCTCACCTCCAGTCGAGCAGAACATCAGCCAAAG CGCGGCTCCTCCCGGAACCCGCAAGGGTGCCACAACATCCGGAGATAGT TGCGCTGGGAGAGTCAGCGGATGACGAAGTATTTCGATCTCCGGTGAA CTCTCCCCCAAGAGACTGGAGGGCTGCGGCGCCCCAACAGGCGGCTTC CTCCTCTCCTCSGGCGGTGCCCGGTATCACTGCTGCAACCCCGTCCAACA CCACCAGGACAGGGAATGGCTCGGCANACTCCATCCTAGCCGAGCACC CGATACCTGCACCTCCGCCAACCAACACCACGGAGGCGAATGGTCGTGC CGACATACCCCGTTCTGGGCGGGCTCCGCCGCCCGGTCNACAAGCCGCC CGTAGAAGGGCTCCCACGACAGAACAACGTCGCATTGTGGGACTCCTC GAGGCGGCCACTGGTCGGGAGCAGCTTGAGGAGGCAACGACTCAAGC TATGCTATTCCTGGCCAGATTGACCGGGCGGCGGCCTGAACCCCGGAAC GCCATTCGGCCGGGAMCGCGGCAACGTCATCCAGCTCAGGGGGATGTC CAAGCGCAGGCGCCTGATCGGATTCASGAGGCCAAGAAACTCCAGCGT CTCTACCGAACATCCAAGAAGCGGGCAGTMCAGAAGATTTTGGCAGGT CCTNATATGAACTGCCAAATCGACAAGAATACCATCACAGCCCACTTTGT GGAGCTGGCTGCCCGCAGAGACGGTGGAGAGGATTGGCCCGATGTCTT CGATCGGGAAGAGCCCACGGCGGCGAGCGGCGAGGCCCTTTGTACCCC TATCACAAGGGAGGAAGTGTTTCGACGGCTCAAGGGGAGGAACAACAC CTCACCTGGGCCGGACGGAATCACCTACAGGGACCTCGCCAAGGCCSAT CCAGGGGCCCATGTGCTCGCCGCCCTCTATAACASCATCTGGCGCATTG AGGCGACGCCTGCCTTGTGGGGGGTGTCCAACACMACCCTKATATACA AGAAAGGCGACGCCATGGACATCTCCAACTGGAGACCGATCTCCTTGG GAGATACGGTCCCCAAGCTATTTGCCGCGATCCTGGCCGATAGAATCAA AAGGTGGGCCGTGGCCAACGGGCGCTACAGTGCCTCCCAGAAGGGCTT CTTGGAGTTCGAGGGATGCTATGAACATAACTTCGTGCTCCAGGAGGCC ATTCGGGAGGCAAAAGGCGGGAGAAAGGAGTTGGTGGTGGCCTGGCT GGACTTAGCCAGTGCCTTTACATCCGTGCCCCACTCCTCTATCCTGCAGG	CCATTCCT TCGTACGG GTTTTCGT GCCGGCAT AGCCGGGT GGGAGAC TCGCGCGG GGGAGGT CATATCTC ACCACCAT CCTCGAGC TTTTTGCT GCACCTGA (SEQ ID NO: 1187)	TAGTGG GGCCAT AACACC TAGGCC CCACAG TGTGGC GATGTC CATGTG TGTCGT CCTTACT TATTTAT TTATTTG TCCTGTC GAGCTG TTTAATC TATTCAT ATGTAT GTGTTG TGTGTG TCGTCCC TCGTAG CAGTTTT ATTCCGT CCAACC AGAGGT CGACCA ATATTAA AATAAG CATGGC TTGAAG CAGGCC AAGCGC CGTGTT CTAACCC CGTTTTA GGGGAA GTTACTT AACCTA	SGXTGREVKCITVNV LMEQQPHTKAIREG DFIVILLPQSDDETLC CPLCVGRGRYSGKTR VECLNRHVKEVHPDL TTTFRCWGCGFAAP GDKKYPRKIVTQHCA TCVPEVSSAPSGRVD GERRVNTRRRLGIAA ATEASPVRRTRRNGL ASPPVEQNISQSAAP PEPARVPQHPEIVAL GESADDEVFRSPVNS PPRDWRAAAPQQA ASSSPXAVPGITAATP SNTTRTGNGSAXSILA EHPIPAPPPTNTTEA NGRADIPRSGRAPPP GXQAARRRAPTTEQ RRIVGLLEAATGREQL EEATTQAMLFLARLT GRRPEPRNAIRPGXR QRHPAQGDVQAQA PDRIXEAKKLQRLYRT SKKRAVQKILAGPXM NCQIDKNTITAHFVEL AARRDGGEDWPDVF DREEPTAASGEALCT PITREEVFRRLKGRNN TSPGPDGITYRDLAK AXPGAHVLAALYNXI WRIEATPALWGVSN TTLIYKKGDAMDISN WRPISLGDTVPKLFA AILADRIKRWAVANG RYSASQKGFLEFEGC YEHNFVLQEAIREAK GGRKELVVAWLDLA

		CGCTCGAGGGACATGGCCTACCAAGCAAAGCGAGAAACATTATCTCCTC	AAAATA CAACTTT TCC (SEQ ID NO: 1310)	SAFTSVPHSSILQALE
		GCTCTACACAGGCATGACCACCCGCTTCCACACGGCGGAGGGACCCACC		GHGLPSKARNIISSLY
		GACCCTATTCTCATCCAGTCGGGGGTCAGACAGGGTTGCCCATTAAGCC		TGMTTRFHTAEGPT
		CCGACGTGTTCAACCTTACCCTGGAAGTTGTCCTACGGGAAATACAACG		DPILIQSGVRQGCPLS
		GACGGGGGAGGGCTACACGATCGAGGGCAGGCGTATTTCCCATCTCGC		PDVFNLTLEVVLREIQ
		CTACGCCGATGATGTCGCGATCCTGGCTGACTCCCCGGCTGGAATGAGG		RTGEGYTIEGRRISHL
		AGGCTGCTTTTTGCAGCAGAACGGGGAGCTCGGGCGGTCGGCCTCACC		AYADDVAILADSPAG
		TTCAACCCAGCGAAGTGCGCTACACTGCACATAGCGGGACGAGGKGAG		MRRLLFAAERGARA
		GAGGCAGTTAGACCGACTGAGTTCTCCGTTCAAGGGACACCTGTCCGTG		VGLTFNPAKCATLHI
		CCCTTGCTTCCGGGGAGGCCTACGAACATCTTGGCATCCCCACAGGCTA		AGRGEEAVRPTEFSV
		CCAAGTCAGGCAGACCCCAATCAACACCTTGCGGGATCTCCTTGCTGAC		QGTPVRALASGEAYE
		ATCGGCTCCATCGACCGGTCCTTACTGGCGSCCTGGCAGAAGCTGGATG		HLGIPTGYQVRQTPI
		CAGTGGGAACCTTCCTACTGCCCCGGCTTGATTTCACCATGCAGGGCGC		NTLRDLLADIGSIDRS
		CCATATCGACAAGGGCTTCTTGACCGAGGCGGACAAGATCATCAAAAAA		LLAXWQKLDAVGTFL
		GCTGCAAAATCTTGGCTAAGCCTTCCTCAGCGAGCGAGCGCGGAGCTCG		LPRLDFTMQGAHIDK
		TGTTTCTACCTCCCTCGCAAGGGGGAGGGGGTTTGCTCACGGTCGCTCA		GFLTEADKIIKKAAKS
		CAGCTATAAAATGCTGTACTCGAGCGATGTCACCGTGAGCACCATAGCG		WLSLPQRASAELVFL
		GGATCCACACTGAGGAGGACCGTTTCGGAGAGGCTTAAAAAGCGGGCG		PPSQGGGGLLTVAHS
		TCTAACATCGATATCGCCCGCTTCCTCTCCGGAGATCTTGATCTGCCCCG		YKMLYSSDVTVSTIA
		GTCGACGAGCCCGTCGACATTCTGGACCAAGGTTCGGAGCGCGGCCCT		GSTLRRTVSERLKKRA
		GAGAATTAAGACCAAACTCGGGCTGCGCTGGAGCTGGTGTCAGGGGG		SNIDIARFLSGDLDLP
		GGGAAGTACTCCTGATGGCCTGTGGGGATCCACGTGCCCCCGGAACAA		RSTSPSTFWTKVRSA
		GAGTCTCGCCGCAGACAAAACATCTTGTCACCACGAGTCTGCGGCGCTG		ALRIKTKLGLRWSWC
		CCTGAATCGGCACTATGCGGAGAGTCTCCTTGCCAAAAAAGATCAAGGC		QGGEVLLMACGDPR
		AAGGTCTTTGAGGTTACCCGGAGGAGCGGACAATCCAACCACTTCCTCC		APGTRVSPQTKHLVT
		GGAGTGGCAGCTTCACTCGCTTCTGCGACTGGCGCTTTATCCACCGCGC		TSLRRCLNRHYAESLL
		CCGTCTCGACGTTCTCCCCCTCAATGCTGCCAAACGGTGGCAACGAGGG		AKKDQGKVFEVTRRS
		ATGGACAAGAGATGTCGCAGATGCGGGTCAGACCTGGAGACTCTCCCG		GQSNHFLRSGSFTRF
		CATGTGCTGAGTCACTGCGGCCCGCACTCGGCGGCGAGACAAAAGAGG		CDWRFIHRARLDVLP
		CACAACAACATCCAGGACAGGTTGGTGAAGGCAGCCTCGAGATGTCCG		LNAAKRWQRGMDK
		GGGACCATTAGTGTCAACCAGACCGTTGTGGGCGTCCGCGGCCCCGAC		RCRRCGSDLETLPHV
		GCGGCATTAAGACCKGACATTGTCGTTCGCGACGATGTCAACCGCCGCG		LSHCGPHSAARQKR
		TCACCATCGTCGATGTGGCCGTGCCGTTCGAAAACCGCCTCGAGGCGTT		HNNIQDRLVKAASRC
		CGACGGCGTTCGKGAGGCCAAGATTGCCAAATATACACCTTTGGCCAGG		PGTISVNQTVVGVRG
		CAACTCACGGACAGCGGTTACACCGTAACGGTGGAGGCCTTTGTAGTG		PDAALRPDIVVRDDV
		GGCGCCCTCGGCGCGTGGGACCCGCGAAATGAACGGGTCCTGAGCCTC		NRRVTIVDVAVPFEN
		CTCAGCATTAGTCGTTATTATGCGATACTGATGAGGAGGCTCATGGTTA		RLEAFDGVREAKIAKY
		GTGACACAATTCGCTGGTCCCGGGACATCTACGTGGAACATGTGTCGGG		TPLARQLTDSGYTVT

			CATTCGGCAGTACCGGGAATAGTGGGGCCATAACACCTAGGCCCCACA GTGTGGCGATGTCCATGTGTGTCGTCCTTACTTATTTATTTATTTGTCCTG TCGAGCTGTTTAATCTATTCATATGTATGTTGTGTGTGTCGTCCCTCGT AGCAGTTTTATTCCGTCCAACCAGAGGTCGACCAATATTAAAATAAGCAT GGCTTGAAGCAGGCCAAGCGCCGTGTTCTAACCCCGTTTTAGGGGAAGT TACTTAACCTAAAAATACAACTTTTCC (SEQ ID NO: 1064)				VEAFVVGALGAWDP RNERVLSLLSISRYYAI LMRRLMVSDTIRWS RDIYVEHVSGIRQYRE (SEQ ID NO: 1432)
NeSL	Utopia-1_PCa	.	Phytophthora capsici	GCCCGCCGGTGGAGTAGCCATGTTGGCCACCACCGCCCAAGTCTCCGCC GCAGCTGCGACTGCTGCTCATCGAGTCGCAGTAGTCGCAGCTGCAC AAGCACCACCTCCTGCCGGACGCGCCGCCGTGGAGCACCACGCGCGCG CTGAGCCGTACCAAGACCAAGGNTTCCAGGCTCGCGCGCGCGTGGMG CTCCCAGCCAGCGACGCGTTCAGCAGCAGGGTCGGCTGCGGCAGCTCG ACGCTGCACGGGGACGGCAGCGACGGCACCAAGACCACGACGA CCAAGCCGTCGCTGCCGTCCCCATGGACGTCGACCAAGGTGCTCGGTGG CCGACAGCGGATCAMCACCCGCTGTCGCCGCCACCGGAGCTCCGAGTC GGCGGCAAGCGCCGCCGCCTGAACGACGGCGACGACGAAGACATCCGC GAGCTGGCCGAGCTTCTGCTGCCCGACGAGGAGGAGGCCGACGACCAC AAACCAGCGCCCAGGTTACCCGCGACCAGCGCTCATCCGGCCTCCGTCC TCGCTGTGTACGCGCACAACGCGCAGCGCTTCAACTGCACGTTGTGCGT GTACACGGCTGCCAGCTTCGCTGCTCTTACGCGACACACGGGACTCTCGG CACCGGCGCGTGACCTTCCTGGACAGGTTCTCGGCGGGTTGCGCGTGC GGCAAACCTTTTGCCTCGAGGCTGGCCGCAGCAAGACACGCACAAACG TGCGCCAGCCTCAGCACCACACTGGTCGCGGTTTCGACGACAGGAGGA GCATCAAGCCACACTGTCGTCGGAGCCAACACCACCGTCGCCACGGCG GTCACCGCCGAACCCCCCCTGCTCCACCATCAAGCCTCGGAACTCACTGT GCCCCCCCCCACGTGTGAGTTCCCCCGACGTCGATGTGCAGCTGCACAG TCCGCCACAGGAAGATCAGCACGAGGACGCCACCCANCACCCGGAGAG CACGCAACACCAACCTCCTGAGGCAACCCGCTGGGGTTCGCCGCTCGCG CCCACGCTCGTTGCCTCCAGGATTGCTCAGCGACTCGGCGAGCTGGAAC CTCCACGCTGGGGCCCACCATTACCCCGCGCGATGGTTGTTTCGCGAAT TACAGCACGGCTTGAAGCTACACCTGCTCCTCGATGGGATCCGCCACTC CCACGCCGCGTCATCGCCTCCCGCATTGCCGACCGGCTTGTACCTGCGA CGGCACCGTGTCGCTCGGCCCTGAACGCTGCATTCCCTTCTCCATCGCGA GACACGGTGACGGAATCGTTCACCCAGGAGGACAGGCAACTGGAACCC CTGACGCGGCACGTGGACGAAGAGACGAAGGACAGCGAGCTGCCGGG CAGGGCGCCGACTGTACTCGACGAGGAGTCCAAGGACAATGACGCCAC TGCCGGTGAGTGGCTCCTCCGGTTCGACGGAGCTTGTCGAGCGAACCCA GGCCCAGGCGGAGCAGGTGCCGTGCTATTCAACCCTAGCGGTGCAGCG TCATGGACCTGCTCGCACTTTATGCCGGGCGCCACGGAGACAAACAACA	GCCCGCCG GTGGAGTA GCCATGTT GGCCACCA CCGCCCAA GTCTCCGC CGCAGCTG CGACTGCT GCTGCTCA TCGAGTCG CAGTAGTC GCAGCTGC ACAAGCAC CACCTCCT GCCGGAC GCGCCGCC GTGGAGC ACCACGCG CGCGCTGA GCCGTACC AAGACCAA GGNTTCCA GGCTCGCG CGCGCGTG GMGCTCC CAGCCAGC GACGCGTT CAGCAGCA GGGTCGG CTGCGGCA GCTCGACG CTGCACGG	TAGACG GCACAG TTCTGG CCCACG TAGGCC GAAAGG GCCCCA CCCATGT AGGGAA CCGCCCT CGGGAA ATCCATT CCGGTG TTCGACT GAAGAG ATGCTCC TTCGCCT TGACGG AGGTAC ATCTCG ACAGTC GAACTT CAACTC GCAACA TATCCG ATACAG TTACAA ACCACA GTTAGA TATCAG ATAGGA ACCTTCC	MVVSRITARLEATPA PRWDPPLPRRVIASRI ADRLVPATAPCRSAL NAAFPSPSRDTVTES FTQEDRQLEPLTRHV DEETKDSELPGRAPT VLDEESKDNDATAGE WLLRFDGACRANPG PGGAGAVLFNPSGA ASWTCSHFMPGATE TNNTAEYTAFLLGAR AAADHGATLLRVQG DSQLVLRQVKGIYGA KSTRLRRLRDAVRAEL ARVGQFSLHHIDRQD NAHADRLANRALDM KSTLVECATHPGRNA CTTTLTTSAAAESPAS PPPVGARDTPMADA GEERLADVDDGEVY AAMRLGPGEVPERR PRLRLRQLSDEELEAA SEMVERLGAALSAKI TDAEDWASAEGYITA LPYMLYDKLQSYSQA PRGPQQPVLTRSPRG DDRPASSEPNASSTT GGVASEHQPRRRRR RGRRKGRRQRRNPR RSGREGATGGHQQH KKHKPRPPRETQHHR EHRLDEALDELHALE

		CCGCCGAGTACACGGCATTCCTCCTAGGCGCCAGAGCTGCTGCAGATCA	GGACGGC	TTTAGG	RTDPHNRPAIAKARR
		CGGTGCGACGCTACTTCGCGTCCAGGGGGACAGCCAGCTCGTACTCAG	AGCGACG	AAGCTA	RVGRIRSAINQQLLR
		GCAGGTCAAGGGCATTTACGGTGCGAAGAGCACCCGTCTACGGAGACT	GCAGGCAC	ACGGGT	HKFDTDEKACVDGIL
		ACGCGACGCTGTTCGTGCCGAACTGGCTCGGGTGGGGCAATTCTCCCTC	CAAGACCA	ACACTG	STARAERAARAATPS
		CACCACATCGACAGGCAGGACAATGCGCACGCAGACCGCCTTGCGAAC	CGACGACC	GATGGT	PPASGAPTTTVSAPG
		AGAGCGCTCGACATGAAGAGCACACTAGTGGAATGCGCAACGCACCCC	AAGCCGTC	AAATAC	AIVTNDDGTCPIPSDK
		GGCCGGAACGCCTGCACCACGACACTGACCACCAGTGCTGCCGCTGAAT	GCTGCCGT	ACATAC	LWRHFDAVNTPRLD
		CACCGGCATCACCCCCGCCGGTGGGAGCTCGAGACACGCCGATGGCCG	CCCCATGG	ATTTC	FDAEAPGSAAFRAA
		ATGCCGGCGAAGAGCGCCTCGCAGACGTTGACGATGGTGAAGTTTACG	ACGTCGAC	(SEQ ID	MDHLPAATRLLDLLK
		CGGCGATGCGGCTTGGCCCAGGCGAAGTTCCTGAACGACGGCCACGAC	CAAGGTGC	NO:	EAPSTDEIETQLQHV
		TCCGCCTCCGCCAGCTGAGCGACGAGGAACTGGAAGCAGCGAGCGAGA	TCGGTGGC	1311)	KASSSPGLDGVGYDV
		TGGTGGAGCGACTGGGCGCCGCCTTATCGGCCAAGATCACAGACGCGG	CGACAGCG		YKRFTIQLLPVLRAAF
		AAGACTGGGCGAGTGCGGAGGGCTACATCACGGCCCTACCGTACATGC	GATCAMC		RCCWLYKKVPQSWK
		TGTACGACAAGCTTCAGTCGTACTCACAGGCACCGCGTGGACCCCAACA	ACCCGCTG		LGVVRLLHKKGPRED
		ACCGGTCTTGACCCGGTCCCCGCGCGGCGACGATCGCCCAGCTTCCTCG	TCGCCGCC		PANWRPICLQQAIYK
		GAGCCCAACGCATCATCCACGACTGGAGGGGTGGCATCGGAGCACCAA	ACCGGAGC		LYTGILARRLTRWMD
		CCGCGGCGCCGCCGACGACGAGGACGACGCAAGGGCCGACGACAGCG	TCCGAGTC		ANDRHAPGQKGFRA
		CCGCAACCCGCCGCCGATCGGGACGCGAGGGAGCTACAGGTGGGCATCA	GGCGGCA		VNGCGEHNFLAATLI
		ACAACACAAGAAACACAAACCACACGCCCACCCCGCGAGACGCAGCATCAT	AGCGCCGC		DNARRKHRPLYEVW
		CGGGAACACCGTCTCGACGAGGCTCTCGATGAGCTCCACGCCCTCGAGC	CGCCTGAA		YDFRNAFGSVPFALL
		GTACCGATCCTCACAACCGGCCAGCAATTGCCAAGGCCCGCCGCCGCGT	CGACGGC		WDSLQRLGVPPDYV
		CGGTCGCATCCGGTCTGCCATCAACCAGCAGCTACTCCGCCACACAAGTTC	GACGACG		DMCKGLYNQASFVV
		GACACGGACGAGAAGGCCTGTGTCGACGGCATCCTGTCTACAGCCCGC	AAGACATC		GNAVDGSTAPVEQR
		GCCGAACGCGCTGCACGAGCAGCAACGCCGTCGCCACCAGCATCAGGA	CGCGAGCT		VGVFQGCPLSPQLFN
		GCGCCCACCACGACAGTCTCGGCGCCCGGTGCCATCGTGACCAACGATG	GGCCGAG		AAISPLLYALRRLPDT
		ATGGCACGTGCCCCATCCCGAGTGACAAGCTTTGGCGACACTTCGACGC	CTTCTGCT		GVQLSSVDRPGASAY
		CGTGAACACACCGCGCCTTGACTTCGATGCGGAGGCACCCGGGAGTGC	GCCCGACG		ADDLKIFSGTKAGITQ
		GGCCTTCCGAGCGGCGATGGACCACCTGCCCGCGGCGACACGGTTGCT	AGGAGGA		QHELVATFLRWTGM
		GGACCTGCTCAAAGAGGCGCCGAGCACGGACGAGATCGAGACGCAACT	GGCCGAC		QANPAKCRSMGVRR
		CCAGCACGTGAAGGCCTCCTCCAGCCCCCGGCCTCGACGGCGTCGGCTAC	GACCACAA		NTNGAVEADNVHLE
		GACGTGTACAAGCGCTTCACGATACAGTTGCTGCCTGTGCTGCGGGCGG	ACCAGCGC		LDDTPIPSMTHMQS
		CGTTCCGGTGCTGCTGGCTGTATAAGAAGGTTCCTCAGAGCTGGAAGCT	CCAGGTTA		YTYLGIGDGFDHVRR
		CGGCGTGGTGCGCTTGCTGCACAAGAAGGGCCCGCGTGAGGACCCTGC	CCCGCGAC		RVELAPKLKTLKHDVT
		GAACTGGAGACCCATCTGTCTCCAGCAAGCCATCTACAAACTATACACG	CAGCGCTC		ALVESGLAPWQVVK
		GGCATCCTCGCGCGGCGACTAACCCGGTGGATGGACGCGAATGACCGC	ATCCGGCC		AVKVYLYPRVEYALR
		CACGCACCCGGTCAGAAGGGATTCCGAGCGGTGAACGGCTGCGGGGA	TCCGTCCT		HLRPDDQHLESFDLH
		ACACAACTTCCTCGCAGCTACGCTCATCGACAACGCCCGCCGCAAGCAT	CGCTGTGT		LRRGLRHLLRLPKNAT
		CGGCCACTGTACGAAGTCTGGTACGACTTCCGTAACGCGTTCGGCAGCG	ACGCGCAC		NEFFYAPVSRGGLGL

	TGCCGTTCGCGTTGCTGTGGGATTCGCTACAGCGCCTCGGTGTCCCGCC	AACGCGCA		LPLVELHAALQIAHG
	CGACTACGTGGACATGTGCAAGGGACTCTACAACCAAGCCTCGTTCGTC	GCGCTTCA		WQTLHSPDPAIRRVA
	GTCGGGAACGCAGTGGATGGCAGCACAGCGCCGGTAGAGCAGCGCGT	ACTGCACG		REQLYQIADARHRLD
	CGGCGTTTTTCAAGGCTGTCCGCTCAGCCCCCAGCTCTTTAACGCCGCCA	TTGTGCGT		KDHWPHRREELCELL
	TCAGCCCACTGCTCTATGCGCTGCGCCGGCTACCAGATACCGGCGTGCA	GTACACGG		LNGELGTSAHAPPKR
	GCTCTCCAGTGTTGACCGCCCTGGTGCCTCAGCGTACGCCGACGACCTT	CTGCCAGC		RNGDIGSLWVDVRK
	AAGATTTTCAGCGGCACCAAGGCGGGCATCACGCAGCAACACGAGCTG	TTCGCTGC		NLKTFGLKVATAPAN
	GTGGCGACGTTCCTGCGGTGGACCGGGATGCAGGCCAACCCTGCCAAG	TCTTACGC		QETGVPAQPLQLRVP
	TGCAGGTCCATGGGCGTGCGGCGCAACACCAACGGMGCCGTCGAGGC	GACACAG		HHAEWLDHGNVLRH
	AGACAACGTCCACCTCGAGCTCGACGATACGCCGATCCCCTCCATGACA	GGACTCTC		VKLHIKNLHWQTWC
	CACATGCAATCCTACACGTACCTGGGGATCGGCGACGGCTTCGATCACG	GGCACCG		ALSDQGKTARVHGG
	TGCGCCGCCGCGTCGAGCTCGCTCCCAAGCTGAAGACTCTCAAGCACGA	GCGCGTGA		VGSAFLTRPRGMWE
	CGTGACGGCGCTCGTGGAGTCCGGCCTTGCGCCATGGCAGGTGGTGAA	CCTTCCTG		SDYRFAVAARLNVVD
	GGCCGTGAAGGTGTACCTCTACCCTCGGGTAGAGTACGCCCTCCGGCAC	GACAGGTT		TVNTLSRRRLRAHDR
	CTCCGACCCGATGATCAGCACCTGGAGAGCTTCGACCTCCACCTTCGCC	CTCGGCGG		CRYPACRWKETLAH
	GCGGCCTCCGCCACCTGCTTCGCCTTCCCAAGAACGCCACGAACGAATT	GTTGCGCG		VLNHCPGTMDAVRG
	CTTCTACGCGCCGGTGTCGCGCGGTGGTTTGGGGCTGCTGCCGCTGGTG	TGCGGCAA		RHDDALKEIEHTLRAS
	GAACTGCACGCAGCACTGCAAATCGCGCACGGATGGCAGACGCTCCAC	ACCTTTTG		SGDRRELRVNQTVP
	TCACCCGACCCAGCCATCCGTCGCGTCGCGCGAGCAGCTCTACCAGA	CCTCGAGG		GLPGPPLRPDIQVYN
	TTGCCGACGCGCGGCACCGGCTCGACAAGGACCACTGGCCACACCGGA	CTGGCCGC		HDKRTVAVVDLAVAF
	GAGAGGAACTTTGCGAGCTCCTCCTCAACGGGGAGCTCGGGACGTCGG	AGCAAGAC		DEQPSEDPESSGLAK
	CCCACGCACCGCCCAAGCGACGCAACGGTGACATCGGCTCTCTGTGGGT	ACGCACAA		AVQIKKAKYAGIKEHL
	CGACGTCCGCAAGAACCTCAAGACGTTTGGACTGAAGGTGGCCACAGC	ACGTGCGC		ENQGWKVHLSAIVY
	GCCGGCCAACCAGGAGACCGGTGTACCAGCCCAGCCGCTGCAGCTCCG	CAGCCTCA		GSLGSVAASNHKVYT
	CGTGCCTCACCACGCCGAGTGGCTGGACCACGCAACGTGCTGAGGCA	GCACCACA		EHLGLLKRDAKRLDR
	CGTGAAGCTACACATCAAGAACCTCCACTGGCAAACCTGGTGCGCCCTA	CTGGTCGC		QLSSACIQSSRRIWN
	AGTGACCAAGGGAAGACCGCACGGGTGCACGGCGGTGTTGGCAGCGC	GGTTTCGA		FHCAKHRARQHEHQ
	ATTCCTGACCCGACCTCGCGGCATGTGGGAAAGCGACTACCGCTTCGCG	CGACAGG		APPSQATRGRRVTET
	GTGGCCGCTCGCCTCAACGTGGTCGACACCGTCAACACACACTGTCGCGCC	AGGAGCAT		GGNPSRTDRR (SEQ
	GGCGACTCCGCGCCCACGACCGGTGTCGGTACCCGGCCTGCAGGTGGA	CAAGCCAC		ID NO: 1433)
	AGGAGACTCTGGCACACGTCCTCAACCACTGCCCGGGTACCATGGACGC	ACTGTCGT		
	GGTACGTGGCCGGCACGATGACGCGCTCAAGGAAATCGAACACACGCT	CGGAGCCA		
	CCGCGCGTCGTCGGGCGACCGCCGCGAGCTGCGTGTTAACCAAACGGT	ACACCACC		
	GCCTGGCCTGCCCGGGCCACCCCTGCGCCCCGACATCCAGGTCTACAAC	GTCGCCAC		
	CATGACAAACGCACGGTCGCAGTTGTAGACCTGGCTGTGGCGTTCGATG	GGCGGTCA		
	AGCAGCCGAGTGAGGACCCAGAGAGCTCGGGGTTGGCCAAGGCAGTC	CCGCCGAA		
	CAGATCAAGAAAGCCAAGTACGCGGGCATCAAGGAGCATCTCGAGAAC	CCCCCCCT		
	CAGGGGTGGAAGGTTCATCTCTCGGCCATCGTGTACGGCTCGCTTGGCT	GCTCCACC		

			CGGTGGCGGCCAGCAACCACAAGGTGTACACTGAACACCTTGGGTTGC TCAAGCGCGATGCAAAGCGGCTGGACCGGCAGCTCTCGTCGGCATGCA TCCAGTCCAGCCGCCGCATCTGGAATTTCCACTGCGCCAAGCACCGAGC ACGCCAGCACGAGCACCAAGCACCACCAAGCCAAGCGACACGGGGCAG GAGGGTGACGGAAACCGGGGGGGAATCCGTCGCGCACCGACCGCCGCT AGACGGCACAGTTCTGGCCCACGTAGGCCGAAAGGGCCCCACCCATGT AGGGAACCGCCCTCGGGAAATCCATTCCGGTGTTCGACTGAAGAGATG CTCCTTCGCCTTGACGGAGGTACATCTCGACAGTCGAACTTCAACTCGCA ACATATCCGATACAGTTACAAACCACAGTTAGATATCAGATAGGAACCT TCCTTTAGGAAGCTAACGGGTACACTGGATGGTAAATACACATACATTT C (SEQ ID NO: 1065)	ATCAAGCC TCGGAACT CACTGTGC CCCCCCCC ACGTGTGA GTTCCCCC GACGTCGA TGTGCAGC TGCACAGT CCGCCACA GGAAGATC AGCACGA GGACGCCA CCCANCAC CCGGAGA GCACGCAA CACCAACC TCCTGAGG CAACCCGC TGGGGTTC GCCGCTCG CGCCCACG CTCGTTGC CTCCAGGA TTGCTCAG CGACTCGG CGAGCTGG AACCTCCA CGCTGGG GCCCACCA TTACCCCG CGCG (SEQ ID NO: 1188)	TAAGCT GGTCAT CATGAC CGACAG	
NeSL	Utopia-1_PI	.	Phytophthora infestans	AGCTCGGCCTCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTG GCGCGGCCCACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTT CGCAGACGCAGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCC CCCCTACGATGTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGA	AGCTCGGC CTCGCGGC TGCCTTCC CAGGCGCC	MLADPAALAAGLAR APPPPSAPQDPSPAF PAGPAGQNPRAAAP ARVEVHTVVAPPGR

		CTCGGTATCGCCGGGTGCTACACTCTCAGCCGCCACAGCTCGGGCCTTG	GCCGACTT	GGCACT	AGGMLPDPGLVEEPI
		GCGGTCCGCCATTGGCCCTTGGAGCTCGACAGCGACAGCAGCGACGAC	CGCGCTCT	ACCCAG	QATYAHDAAQFECA
		GAGGACGCTCAAGACCCCCACGCCGCCGCCCCAGAACCCCCAGAAGAC	GGCGCGG	GTAGGG	LCPYVAESMAVLVQ
		GTCGCGAGTGTGCTTGCCCCACCCGGCAGGGCAGGCAGCATGCTCGCC	CCCACACG	AACCGC	HRRSAHRGTRFKDIF
		GACCCAGCAGCGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCCGCCCCCAT	CCGCCGCC	CCTTAAA	TSGCQCSLVFYARIVA
		CGGCACCCCAGGACCCAAGCCCCGCCTTCCCTGCTGGCCCCGCCGGCCA	GAGCCTCC	AAACCC	ASHAVACARRNQRA
		GAACCCTCGCGCCCGCCGCCCCAGCCCCGTGTTGAAGTCCACACAGTCGTG	AAGCGCGC	AGGAAG	VPPAPTPVAPTRPEA
		GCACCACCCGGCAGGGCGGGCGGCATGCTGCCCGACCCGGGGTTGGTT	CCGTTGGC	ACACAA	TPQPTGYLAAAMTA
		GAAGAACCTATCCAAGCGACGTACGCGCACGACGCTGCCCAGTTCGAG	TTTCGCAG	ACACCCT	AAAAASSDTVVAAAT
		TGTGCGTTGTGCCCGTACGTCGCTGAGAGCATGGCAGTGCTCGTTCAGC	ACGCAGG	CCACTTA	NMQSAVPAAAKTTG
		ACCGACGGTCGGCTCACCGCGGCACCCGCTTCAAGGACATCTTCACCAG	GCTCGCGG	GTGACA	LQLVPPELEPALPQR
		CGGGTGCCAGTGCAGCCTAGTGTTTTACGCGCGCATTGTTGCCGCCAGC	CGACGGCC	TACATAT	ASCHAGKRRRLNAD
		CATGCAGTGGCCTGTGCCAGGCGCAACCAGCGAGCCGTGCCACCAGCT	CTCGCCAG	TTTAGCC	EAVTPCTPTARVSPQ
		CCGACCCCTGTTGCCCCCACGCGGCCCGAGGCTACCCCACAACCTACTG	CCCCCAAG	TAGATTT	TEVAMAPHDAPQD
		GGTACCTCGCCGCGGCGATGACTGCCGCTGCGGCAGCCGCATCAAGCG	ACCCCCCC	CAGTTA	DTVLQREAAEPQPDP
		ACACCGTAGTGGCAGCCGCCACCAACATGCAGAGCGCCGTGCCAGCCG	TACGATGT	CGGAGA	AATPGAQVQRVEDT
		CCGCCAAGACCACTGGGCTTCAGCTCGTCCCCCCCGAGCTGGAGCCTGC	GGCACCAC	GGTTAC	TAAQDDTVQQDHD
		CCTTCCGCAGCGTGCTTCTTGCCATGCAGGCAAACGCCGTCGGCTCAAT	CCGGCAG	TAACTG	ADTAQVSPPRRTPTR
		GCCGACGAGGCGGTGACACCGTGTACGCCTACGGCGCGGGTGTCGCCG	GGCGGCC	GTAAAT	WGPRPSSTQEPSPM
		CAAACAGAAGTGGCGATGGCGCCTCACGACGCACCCCAGGACGACACC	GGCAGGCT	ACGAAC	TGEPAATLAARRPLT
		GTCTTACAGCGTGAGGCCGCCGAGCCACAGCCAGACCCAGCAGCAACG	GCCCGACT	ACATATT	PAATGTRATRWGPC
		CCGGGTGCCCAAGTGCAGCGAGTCGAAGACACAACTGCGGCGCAGGAT	CGGTATCG	CTGTTCT	HRAIGAAAIARLVTGL
		GACACTGTGCAGCAGGACCACGACGCCGACACTGCCCAGGTCTCGCCG	CCGGGTGC	AATCAG	STEPAQPQRRQPPPP
		CCCCGGCGCACACCCACACGCTGGGGACCAAGACCAAGCTCAACTCAA	TACACTCT	TGTGAA	QEPPSQPEPLAAAAT
		GAGCCTAGCCCCATGACAGGCGAGCCGGCTGCCACCCTGGCAGCGCGA	CAGCCGCC	AACTGG	AAADIAATVAADIAA
		AGACCACTGACTCCTGCCGCAACGGGCACTCGTGCCACGCGGTGGGGG	ACAGCTCG	TTTTCGC	AAANAAMDVDGGP
		CCATGTCACCGCGCGATCGGGGCCCGCAGCCATCGCACGACTTGTCACGG	GGCCTTGG	CTTTTGG	AADETWLLRFDGAC
		GCCTTTCAACGGAGCCAGCGCAGCCTCAGCGACGGCAGCCACCGCCAC	CGGTCCGC	CGGACT	RRNPGPGGAGAALF
		CACAAGAACCTCCATCGCAACCAGAGCCGCTAGCAGCAGCAGCAACAG	CATTGGCC	TTTTCAC	APSGAVVWTCSHYM
		CCGCAGCAGACATAGCAGCAACAGTTGCAGCAGACATTGCGGCAGCGG	CTTGGAGC	TCGCATT	PSRSETNNTAEYTALL
		CAGCGAACGCAGCAATGGACGTGGATGGTGGACCGGCAGCCGACGAG	TCGACAGC	TTTGGG	LGVQSAVHHGASHL
		ACATGGCTGTTGCGCTTTGACGGCGCATGCCGACGGAACCCAGGACCT	GACAGCA	CAATCG	EVEGDSSLVIAQVKG
		GGTGGAGCAGGTGCCGCACTCTTTGCGCCCAGCGGTGCCGTGGTGTGG	GCGACGAC	TCTGCG	TFACRNARLRQLRNR
		ACGTGCTCGCACTACATGCCAAGCCGCTCCGAAACGAACAACACCGCCG	GAGGACG	GCTAGC	VRHALRSVDTHKLRH
		AGTACACGGCGTTACTGCTCGGCGTACAGAGCGCAGTCCACCACGGTG	CTCAAGAC	TTGCTA	IDRQANAHADRLAN
		CCTCGCACCTGGAAGTCGAGGGCGACAGCAGCCTGGTAATCGCTCAGG	CCCCACGC	GCGGCG	RALDQRRTSSECGTH
		TAAAGGGCACCTTCGCATGCAGGAATGCGCGGCTCCGCCAACTCCGCAA	CGCCGCCC	GACGAG	GSCMDSCLAVPTALA
		CCGAGTGCGTCACGCGCTCCGCTCCGTGGACACGCACAAACTCCGGCAC	CAGAACCC	CGGTCT	AQETPPAAPPSTSAT

	ATCGACCGCCAGGCCAACGCGCATGCAGATCGCCTTGCCAACCGCGCGC	CCAGAAGA	CCGGGG	PAEGNAMDDIAAEIA
	TAGACCAGCGGAGGACCAGCAGCGAGTGCGGGACGCATGGCTCGTGC	CGTCGCGA	GCGTTC	ARDEGETFPVLPIGP
	ATGGACTCCTGCCTCGCCGTACCCACGGCCCTGGCGGCGCAAGAAACGC	GTGTGCTT	ACCTTTC	GSAPERQPRLRLRQL
	CACCGGCAGCACCACCCAGCACCAGTGCCACTCCAGCCGAAGGTAATGC	GCCCCACC	CCCCGC	SDEERDAAADALQEL
	CATGGACGACATCGCAGCCGAGATTGCTGCGCGTGACGAGGGCGAGAC	CGGCAGG	GAGGCC	ADTMASKIEDADSW
	CTTCCCAGTCCTGCCAATTGGCCCAGGCTCAGCCCCAGAACGGCAACCC	GCAGGCA	AACTAC	TSGEGYISSIPERIREV
	CGACTCCGTCTGCGCCAACTTAGCGATGAAGAGCGCGACGCAGCGGCA	GC (SEQ ID	ACCGAT	LQPYATAPPQPGRSR
	GATGCCCTGCAGGAGCTCGCCGATACCATGGCGAGTAAGATCGAGGAT	NO: 1189)	CTTCTCT	RQQRRRPPRVTRNQ
	GCTGACAGCTGGACCTCCGGCGAAGGATACATCAGCTCCATCCCCGAGC		ACACTTT	REHRLDEALDDMAA
	GTATCCGAGAGGTACTGCAGCCATACGCTACGGCCCCGCCGCAGCCGG		TCTAATT	TQQATPRDQRAVRR
	GACGCTCTCGCCGGCAACAACGCCGTCGACCACCACGGGTGACGCGCA		CGCCTCC	ARRRVGRVRASMAQ
	ACCAGCGCGAACATCGACTCGACGAAGCTTTGGACGACATGGCAGCCA		GTCTTC	QELRHEFAKDESKCV
	CACAGCAAGCGACCCCGCGAGACCAGCGGGCTGTGCGCCGTGCGCGGC		GGTCTT	AKILKTASTETAAEDE
	GACGTGTGGGCCGGGTGCGAGCCTCAATGGCACAACAGGAGCTGCGCC		CGGTTG	HPETCPIDAATLHAH
	ACGAGTTTGCCAAGGACGAGTCTAAGTGTGTGGCCAAGATCCTGAAGA		TCGGGC	FTGVNAPRTDFDYDA
	CAGCGTCGACAGAGACGGCAGCGGAGGACGAGCACCCGGAGACGTGC		TTTTTC	TSGREFRAAMSDLPP
	CCCATTGATGCAGCTACGCTCCACGCCCATTTCACGGGAGTCAACGCGC		TTTTTGA	ATVEIDAFDAELTIDE
	CAAGGACGGACTTCGACTACGATGCCACCAGCGGCCGGGAGTTTCGAG		CCAATC	VEDQLTRAAKTSSPG
	CGGCGATGAGCGACCTTCCGCCAGCCACCGTGGAGATTGATGCGTTCG		AGAGCG	HDGIDYGIYSRFAAQ
	ACGCGGAGCTGACGATCGACGAGGTGGAGGACCAGCTCACACGCGCG		CGCCAT	LVPLLHAVFQFCWRH
	GCCAAGACCTCAAGCCCCGGCCACGACGGCATCGACTACGGCATCTACA		GCGCCT	RRVPRLWKVGIVRLI
	GCCGCTTCGCCGCGCAGCTCGTTCCGCTCCTCCACGCGGTGTTCCAGTTT		CTTCTG	HKKGDPRQPTNWRP
	TGCTGGAGACACCGTCGTGTGCCAAGACTGTGGAAGGTGGGCATCGTA		GCCAAT	ICLQPTIYKLYSGLLAH
	CGCCTCATCCACAAGAAGGGCGACCCAAGGCAGCCGACTAACTGGAGG		CAGAGA	RLSRWLEGNDRLPM
	CCGATCTGCCTTCAGCCCACGATCTACAAGCTCTACAGCGGGCTGCTGG		CCGGGC	AQKGFRAFNGCHEH
	CGCATCGGCTCTCACGCTGGCTGGAAGGTAACGACCGCCTGCCGATGG		CCTGTCC	NFMATTLLDQTRRQ
	CCCAGAAGGGGTTCAGGGCCTTCAACGGGTGCCACGAGCACACAATTTCAT		TCGGAC	HRKLYQVWYDLRNA
	GGCTACCACGTTGCTGGACCAGACCCGCCGGCAGCACCGAAAGCTGTA		AGCGAG	FGSLPQQLMWRVLR
	CCAAGTGTGGTACGACCTTCGCAACGCCTTCGGCTCGCTTCCTCAGCAG		GCCTCC	HLGVDSGFIDRCRDIY
	CTCATGTGGCGCGTGCTTCGCCACCTTGGCGTCGACTCAGGCTTCATCG		ACGGCC	RDSAFVVANAADGA
	ACCGCTGCAGGGATATCTACAGGGACTCGGCCTTCGTGGTGGCGAACG		AGCCAA	TDPVRQEVGVYQGC
	CGGCGGACGGTGCGACGGACCCGGTGCGCCAAGAGGTGGGCGTATAC		TCGAGT	PLSPLLFVAALVPLVR
	CAAGGCTGCCCACTCAGCCCGCTCCTGTTCGTCGCGGCGCTGGTGCCTC		CTCGGC	RLEKLDGVGVPLADG
	TCGTGCGCCGTCTCGAGAAACTCGACGGCGTTGGTGTGCCACTGGCGG		AGCGAC	VRPCTTAYADDLKVF
	ACGGTGTGCGGCCCTGCACCACGGCCTACGCCGATGACCTCAAGGTGTT		GCGTCT	SDSAAGIRKCHDTVA
	CAGCGACAGCGCCGCCGGCATCCGCAAGTGCCACGACACAGTGGCTGG		TTCTATA	GFLAWTGLRANPGK
	CTTCCTCGCGTGGACAGGGCTGCGCGCTAACCCTGGCAAGTGTGCCAGC		GCGCAG	CASLAVTTNARGNPT
	CTCGCAGTGACGACGAACGCCCGGGGCAACCCGACTAGAGACAGCAGC		CTGACG	RDSSMRLEVHDAAIT

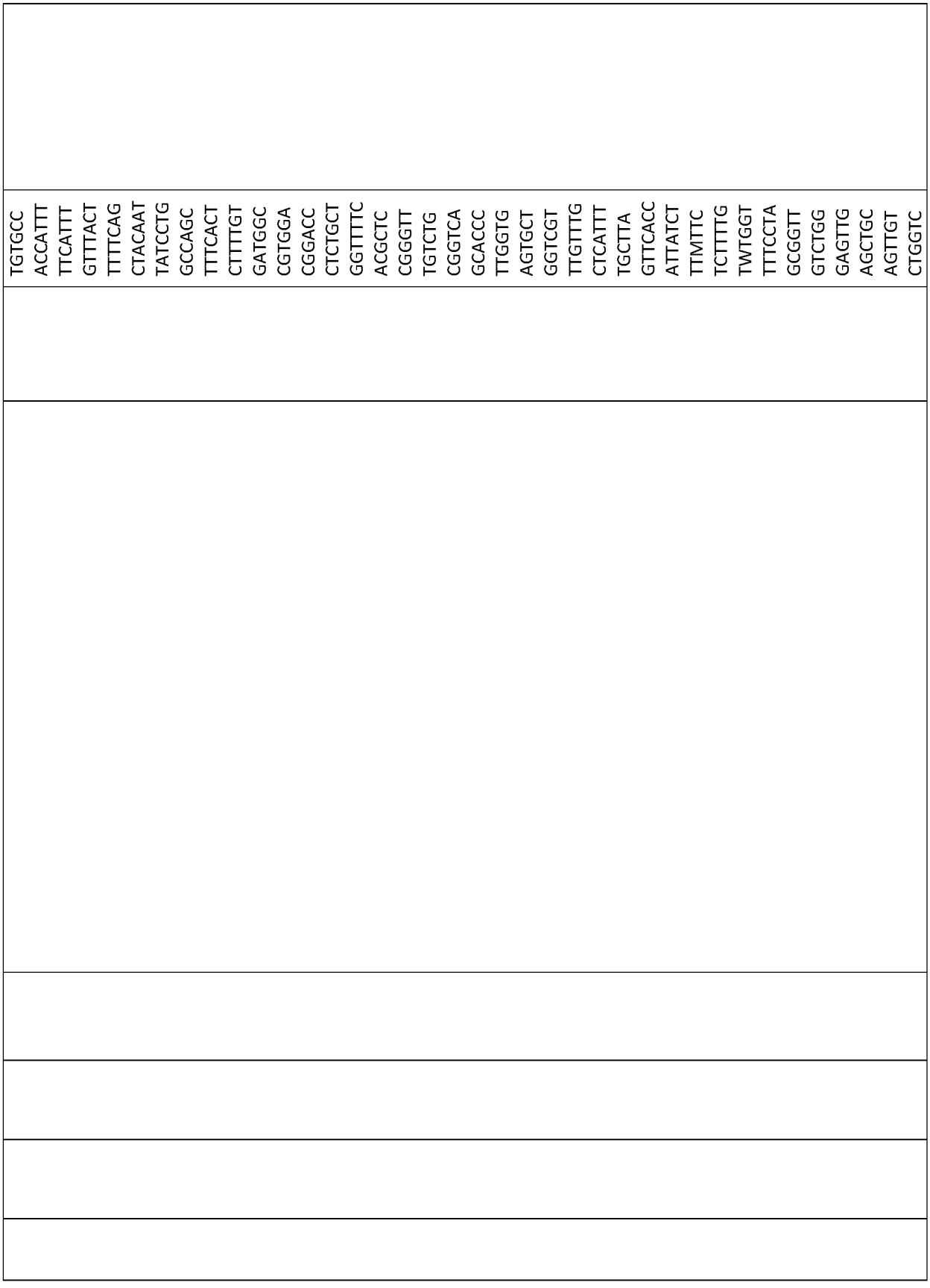
	ATGCGTCTCGAGGTCCACGACGCCGCCATCACCACGCTCAGCCTCCACG	AGGCCG	TLSLHESYRYLGVGD
	AAAGCTACCGCTACCTGGGCGTGGGTGATGGATATGACCACGTGCGCC	ATCTGG	GYDHVRHRLQLEPKL
	ATCGTCTCCAGCTCGAGCCAAAACTCAAGCAGCTCAAGCGGGAGGCGG	CGGCCC	KQLKREAVALLTSGL
	TGGCGCTGCTGACGTCGGGTTTGGCACCGTGGCAGGTGGTGCGCGCGC	CCGATT	APWQVVRALKVYVY
	TCAAGGTGTACGTCTACCCCAAGGTGGAGTACGCGCTGCGCCACCTGCG	GGTCCG	PKVEYALRHLRPLQS
	CCCGCTGCAGTCGCAACTGCAGGCGTTCGACCGTGTCGTCTCCAAGGGC	ACTTTCG	QLQAFDRVVSKGLRH
	TTGCGACACCTGCTGAGCCTGCCTCGCTCAGCCACCAGCGAGGTGCTGT	GCCAAT	LLSLPRSATSEVLYAP
	ACGCCCCGACGTCCAGCGGAGGCCTGGGCCTGCAGCCGCTCGTGGAGC	CAGCGA	TSSGGLGLQPLVELH
	TGCACCGCGCCCTACAGCTCGCCCACGCGTGGCAGATGCTGCATTCCAA	CGACGA	RALQLAHAWQMLH
	GGACCCAGCCATCCAAGCAGTGGCAAGAGCGCAGGCTTGCCAGGTGGT	GGGGGC	SKDPAIQAVARAQAC
	GCGCAAGCGCTACCGCCTGCAAGAAGACCATTGGCGCGGGCGCGATGA	AGGGGT	QVVRKRYRLQEDHW
	CGAGCTCGTGCGCAGCTTCCTGAACTCCGAGCTCGCCGCCTCGCCACAC	TTACACT	RGRDDELVRSFLNSE
	GCTGAGGTCCTCCGTCGCAACGGTGACATTGCCTCGCTGTGGTCCGACG	TTTGCCC	LAASPHAEVLRRNGD
	TGCAGCGGTGGCTGCGCATCTACCACCTGCGGTTCGAGCACTGCGACGA	CCGTTTC	IASLWSDVQRWLRIY
	GACAGAGGCCCATGGCCCGCTCAGCTTCCGCGTGCCCGCACCACAACAA	GACTTC	HLRFEHCDETEAHGP
	GTGGCTTACGCACAAGACTGTGCTGCGGCACGTGAAGCTACACCTGAA	AACTTCA	LSFRVPHHNKWLTH
	GATCCGACACCAGACCCGATGGAAGGGCATGGTGGACCAGGGCAAGA	GGCCAA	KTVLRHVKLHLKIRH
	CGGTGCGCGTGCACGGTGGGGTGGGCGCCAAGTTCATGACGACCGGA	AATGGC	QTRWKGMVDQGKT
	GCGGGCCTGTCGGACGACGACTACCGCTTCGGGGTGAAGGGCCGCCTC	GATTTG	VRVHGGVGAKFMTT
	AACCAAGTCGACACGAACTCGGTGCTCAAGCGCAAACGACTGCGGGCA	GACCCT	GAGLSDDDYRFGVK
	CACACGACCTGCAGGGACCCAACCTGCTCCTCGGCTGAGACGCTGGCAC	CCACGC	GRLNQVDTNSVLKRK
	ATGTACTCAACCACTGCGAGTCCAACATGGACGCCATTCGCCAGCGTCA	GCCGTG	RLRAHTTCRDPTCSS
	CGACGACGCCCTGGAGCAGATTGGGTCTAAGATCCGCGGAGCGCTCGA	CCACTG	AETLAHVLNHCESN
	CCGCGCCAAGTCTCCCACAGAGCTGCGCCTGAACCAGACTGTACCCGAG	CTCGGC	MDAIRQRHDDALEQ
	TACACAGGCGCAGCTCTACGCCCAGACATTGTGCTTAGAAACGTGGCCG	ACCGGC	IGSKIRGALDRAKSPT
	CCAAGACGATGGTGATCGCCGACTTGGCGGTCACGTTCGAAGACCAAG	GGCGAT	ELRLNQTVPEYTGAA
	CGGCTCGCGCTCGTCACTCCTCGCTGCAGCTCAGCCACGACCACAAGAC	TCAGCG	LRPDIVLRNVAAKTM
	TCTCAAGTACCAACCAATCGTGGCGGAGCTGCAGCACAAGGGATGGCG	GGTGCA	VIADLAVTFEDQAAR
	AGTGCAGACTGCTGCCATAGTGTACGGCACGCTGGGCTCGGTGCAGCC	ACTTCG	ARHSSLQLSHDHKTL
	CAGCAACTTCAAGGCCTACACTGAGAAGTTCAAGCTTCACAAGCGCGAG	GGCACG	KYQPIVAELQHKGW
	GCTAGGCAGCTGGATCTCCAGCTGTCTAGCCACTGCATTCGCGCCAGCC	TGTGCA	RVQTAAIVYGTLGSV
	ACCGCATCTGGGGCTGGCACTGCCGACAGCACCGCGATCGGCAACGAA	ACACAT	QPSNFKAYTEKFKLH
	GCGGCACTGCCTCGCGAGCGTCGCGTGGGTCTGGGGGGGCCCCGCGGC	GCAGCG	KREARQLDLQLSSHCl
	GCACATCGCAGGCACCGGCACGGCGATAAGCTGGTCATCATGACCGAC	CCCATTG	RASHRIWGWHCRQ
	AGGGCACTACCCAGGTAGGGAACCGCCCTTAAAAAACCCAGGAAGACA	CACGCC	HRDRQRSGTASRASR
	CAAACACCCTCCACTTAGTGACATACATATTTTAGCCTAGATTTCAGTTA	AAGCGG	GSGGAPRRTSQAPA
	CGGAGAGGTTACTAACTGGTAAATACGAACACATATTCTGTTCTAATCA	CATCGC	RR (SEQ ID NO:
	GTGTGAAAACTGGTTTTCGCCTTTTGGCGGACTTTTTCACTCGCATTTTT	GGGACG	1434)

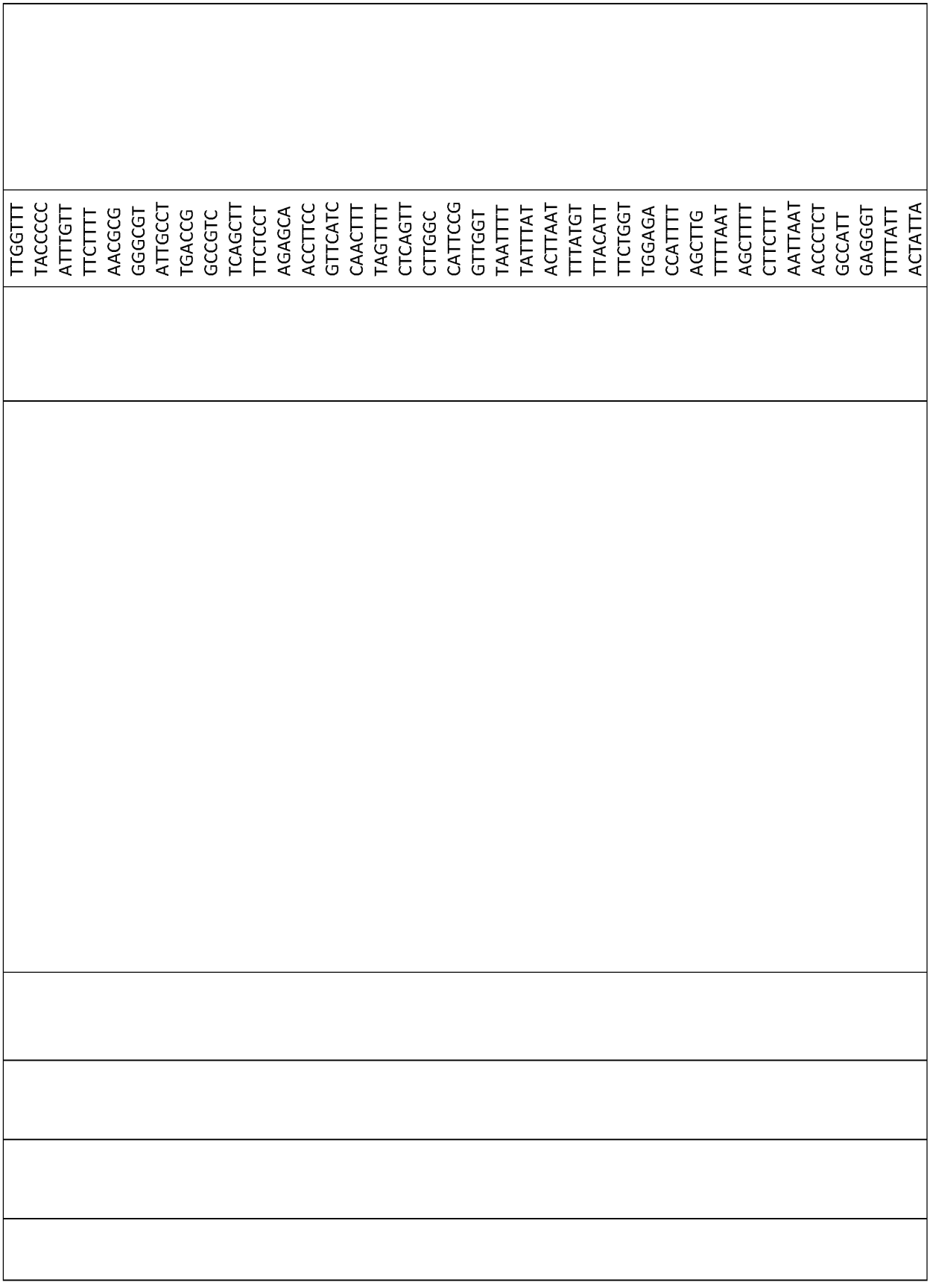
			GGGCAATCGTCTGCGGCTAGCTTGCTAGCGGCGGACGAGCGGTCTCCG GGGGCGTTCACCTTTCCCCCGCGAGGCCAACTACACCGATCTTCTCTACA CTTTTCTAATTCGCCTCCGTCTTCGGTCTTCGGTTGTCGGGCTTTTTTTCTTT TTGACCAATCAGAGCGCGCCATGCGCCTCTTCTGGCCAATCAGAGACCG GGCCCTGTCCTCGGACAGCGAGGCCTCCACGGCCAGCCAATCGAGTCTC GGCAGCGACGCGTCTTTCTATAGCGCAGCTGACGAGGCCGATCTGGCG GCCCCCGATTGGTCCGACTTTCGGCCAATCAGCGACGACGAGGGGGCA GGGGTTTACACTTTTGCCCCCGTTTCGACTTCAACTTCAGGCCAAAATGG CGATTTGGACCCTCCACGGCCGTGCCACTGCTCGGCACCGGCGGCGAT TCAGCGGGTGCAACTTCGGGCACGTGTGCAACACATGCAGCGCCCATTG CACGCCAAGCGGCATCGCGGGACGACGCCTCGGCCGCCCAAGCGCAGC CCCGCCCTTCCAGCACGACCTCGCGCCGTTTGGCGGATCGCCATCAAGA CGTGCGAGAGCCAGGCGGGGTCGGGCAAAATATACTTACTCTAAGTAT GCCCGAATCCCTGCCCTCTCAGGCTGAACGCGGCCCCATACTTGATCTAA GTATGGGAGGATCCCTGGCCTCTCAGGCTGTACGCGAGACCC (SEQ ID NO: 1066)		ACGCCT CGGCCG CCCAAG CGCAGC CCCGCC CTTCCAG CACGAC CTCGCG CCGTTT GGCGGA TCGCCAT CAAGAC GTGCGA GAGCCA GGCGGG GTCGGG CAAAAT ATACTTA CTCTAA GTATGC CCGAAT CCCTGCC CTCTCAG GCTGAA CGCGGC CCCATAC TTGATCT AAGTAT GGGAGG ATCCCTG GCCTCTC AGGCTG TACGCG AGACCC (SEQ ID NO: 1312)	

NeSL	Utopia_PMi	.	Patiria miniata	CTGATGTGGATACCTTGGAATTACTCAACCGTGTCGGAGTCTTTGTCTT TTGCGCCCAACACCTCATGGATACCATGCTTGTCGCTGGAGCGACGTTA CAAGCGTGAGGGCGCCCTCCATGCCGGACAGCTGGTCTGTGCCATGTG CTTGAAGTCGTTCTCCTCGACTTCGGGCTTGCGGCGGCACATGGCCCGG CTGCACCGGCAGCCTTCTCCCGACGCCTCCACGCCCTCGACTATGACGG AAGTGTTCCCGTACCCCCTTCCGAAGGTGTGGCCGTGCGTGGTGTGCCG CGAGAACTTCTATCACAATCAAACCTTGAAGCGTCACCAGAAGAATTTTC ACCCCAATGTTGACCTCACCACCGTCTACCAGTGCTCAGTGTGCGGGCA GGAGTTTGTCACCGGCAGGAAAGCATCCTTCCATTTCAAAGTTCATCGG CGGATGTCAGCCAGCGCCATCCCTACCCCGTCTGCAATGCCCTCCTCACC GATGGATTTGATCCGAGGCCTCGTCGGGGAACCCCTGCCTCCTAGCCCT GCCCGTACACCGCCACCACTTGCCCGATACATCTCCCCAGCGCCCCGGTC TTCTTTTCACCACCCTGGAATCCCTCCCCACCGCCCCGGTCGCCCACCCC ACTGCCACGGCCACTCACTCCACCGCCACGATCACCGAGCCCGCCGCCC CGGTCGCCAACACCACCTCCCCCGGTCACCCTCACCACCGCCCCGGTCAC CGAGCCCGCCGTCCCGGTCGCCCTCACCACCGCCCAGGTCACCGAGCCC TCCGCCCCGGCCGTCCACACCGCCGCCCCGGTCACCCTCTCCAACGCCCC GGTCACCGAGCCCGCCACCCCGGCCACCGATCCTGCCACCCCGGTCACC AGACTCCACTCACCGGTCACTCACATCTCATGCTCGATCTCCTTCACGGC CAGCCACGCCCCCTATTCCTGTGCAGCCCCGACATCTCCGTCCGTCTACG CCTGCTCACCCAGGCGTAGACAATGCAGTTCCACCATCGCAGCAGTCTG CAATTGACGTGTGGCTGGCCGAGCTGTCACGCTCGGCTGACTTCGAGTC TTTTGAGGATGTGTGTGATCGGTTTGTGGAGTTTGCCGCCGCTGAGGGT CGAAACAACGGACGCCCAGCCCGCCACCAACCGCCCCGTGAC CGCGGGAACCAGGGACCTCGCCCCCAACGTCCCCCTCGTCCCCACCGCC CCGGGCTTGGCCCGGAGTTTGATGCACAGGAGGCCAGTCGTCTGCAAA AGCTGTATCGGACCAGCAAGAAGCGCGCCATCAGGACTATCTTGACCG GCTCTGATGTCCGATACAGCGGCTACCGAGGTCCTCACTAGTGGATTTC TTTCACTTGAAGTATTTTCCGCTCGTGAGGTCGACACCGACACCATTGCC ACGGATACCAGCCTCCTGTTTCCGAACTCCGCCCAAGCCCGTGAGTCCG GCCAAGACCTGCTTCGCCCAGTCACCCAGCGTGAAGTCTCCCTCCGTCTT GGCCGGATGTCCAACTCCGCACCAGGGAAGGACCGGTTGGAGTACCGT CACATCCGCCAGGTTGATGGCGCATTCCGCGTCACCCTGGAAATCTTCA ACCGGTGCCTGCGTGAAAGCCGGGTCCCATCCTCCTGGAAGACAGCGA CCACCGTGCTGATCCACAAGAAGGGGGGATGCCACCGACCCGGCCAACT TCCGCCCCATCGCCCTCCAGTCCTGCCTCTACAAGCTGTTGATGGCGATC TTGTCCGACCGTGTCACCACATGGGCCCTGGACAACGATCTCATCAGCA GCTCGCAGAAGAGCGCCCGCCCCGGTGAAGGTTGCTACGAACACACAT	CTGATGTG GATACCTT GGAATTAC TCAACCGT GTCGGAGT CTTTTGTCT TTTGCGCC CAACACCT CATGGATA CCATGCTT GTCGCTGG AGCGACGT TACAAGCG TGAGGGC GCCCTCCA TGCCGGAC AGCTGGTC TGTGCC (SEQ ID NO: 1190)	GATTAG CGAACA CTAATAT CCTGCC MTAGAC GTGATT GCTAAT CCGCAA ACCAAC CGGATC TACAGC CTGAAC ACTGAA CTTTAAT CTTCACC CATGTC ACATCT GGACAC TAGGTTT TTGCTCT GTTTGT GTTTTCC TGCCTTT WCMTT GGAWCT TTCGCM CTGGAA TTTATTT GTCGCT TGGATT ATTTTTT TTCTCAC AATTTG GATCTAT TTTCGTT TGTTCAC TTCGAA CTCTAGC	MCLKSFSSTSGLRRH MARLHRQPSPDAST PSTMTEVFPYPLPKV WPCVVCRENFYHNQ TLKRHQKNFHPNVDL TTVYQCSVCGQEFVT GRKASFHFKVHRRM SASAIPTPSAMPSSP MDLIRGLVGEPLPPS PARTPPPLARYISPAP RSSFSPPWNPSPPPR SPTPLPRPLTPPPRSP SPPPRSPTPPPPVTLT TAPVTEPAVPVALTT AQVTEPSAPAVHTAA PVTLSNAPVTEPATP ATDPATPVTRLHSPV THISCSISFTASHAPYS CAAPTSPSVYACSPR RRQCSSTIAAVCNSE ASSGNPCLLALPVHR HHLPDTSPQRPGLLF HHPGIPPHRPGRPPH CHGHSLHRHDHRAR RPGRQHHLPRSPSPP PRSPSPPSRSPSPPPR SPSPPPRPSTPPPRSP SPTPRSPSPPPRPPILP PRSPDSTHRSLTSHA RSPSRPATPPIPVQPR HLRPSTPAHPGVDN AVPPSQQSAIDVWL AELSRSADFESFEDVC DRFVEFAAAEGRNN GRPARPAHQPPRDR GNQGPRPQRPPRPH RPGLGPEFDAQEASR LQKLYRTSKKRAIRTIL

		TTCTTCTCTCCACCGTCGTGAAGGATGCCCGGCGTAACCAGAAAAACAT	TGCCCTT	TGSDVRYSGYRGPHP		
		GTACGCCGCCTGGCTGGATTTACAAAATGCATTCGGCAGCATCCCCCAC	TCTTCG	ALMSDTAATEVLTSG		
		GACGCAATGTTCACCGTGTTAACCTCCATCGGGGCCCCGGAGGGCCTGG	GACACT	FLSLEVFSAREVDTDT		
		TCTCCCTGGTGCGCGATGTGTACACTGACGCCTCCACCGACTTCGTCACA	GAACTTT	IATDTSLLFPNSAQAR		
		CCCACCGGCCGCACTGCTGCAGTCCCGATTCACTCAGGCGTGAAGCAGG	AATCTTC	ESGQDLLRPVTQREV		
		GCTGCCCGATCTCTCCAGTGCTTTTCAACCTGACGCTGGAACTTATTATC	GCCATG	SLRLGRMSNSAPGK		
		CGAGCAGTGAATGCTTCAGCGACTCGGGGACCAGCAGTGCGCCGGTGGTC	GCTGTC	DRLEYRHIRQVDGAF		
		CATGGCCAGGCCGTCCCCATACTTGCGTATGCTGACGACCTGGTCATAC	AGTCGC	RVTLEIFNRCLRESRV		
		TCAGCCGATCCAGTGACGGCCTGCAGAGCCTCCTGACAACAGCATCCAT	CGGTTC	PSSWKTATTVLIHKK		
		CATGGCCACCAAGATCCAGCTGAAGTTCAAGCCCGCCAAGTGCGCTTCC	ACTTGCT	GDATDPANFRPIALQ		
		CTCTCCCTCGAGTGCCGACGGGGGACCAAGGTACGCCCTCTTGAGTTCA	GCGGTG	SCLYKLLMAILSDRVT		
		ACGTGCAGGACAAGATCATCCCCGCCCTTACAGAGGAGCAGCACTACCG	GGATCC	TWALDNDLISSSQKS		
		GTACCTTGGTGTCCCCATCGGCCTCTACCGCACCGACGACAGCCTGGAG	TGTTGT	ARPGEGCYEHTFLLST		
		ACCTTGGTGGCCAAGATGACTGATGACATCCAGCGGATTGACTCCTCAC	GATAAT	VVKDARRNQKNMY		
		TCCTCGCTCCCTGGCAGAAACTGGATGCCATCCGGACCTTCGTACAGCC	CCCCGT	AAWLDLQNAFGSIP		
		CTGCCTAGCCTACACCCTGCGGGCAGGGGACTGTGCGAAGAAACACCT	GCATTG	HDAMFTVLTSIGAPE		
		CAAGAGACTGCGGGGGACAGCTGGTGAAGACAGCCCGCAAAGTATGCAA	CCCATG	GLVSLVRDVYTDAST		
		CCTGCCGACTCGGGCCACCACAAACTACATCTTTGCTGATCGTAGGGCG	GATTTAT	DFVTPTGRTAAVPIH		
		GGTGGCTTGGGTTTCATCGACCCAAACGTCGATGCCGATATTCAAATCA	TTCCGCC	SGVKQGCPISPVLFNL		
		TCACGCAGGCCGTACGCATGCTGTCTTCACCCGACGACATCACCAGAGC	TTAGTTG	TLELIIRAVNASATRD		
		CATCGCCACCGGCCAGCTGTCTTCCGTGGTCCACCGCACCATCCACCGA	TTCCTAA	RSAPVVHGQAVPILA		
		GCACCGACACAGGAGGAGACGGACGAGTTCTTGTCAGCATCAATGGAG	CCTTGG	YADDLVILSRSSDGLQ		
		GGGGACTTTGCCAACAGTGGGAACAGCGGCCAGGCCTCTTCATTGTGG	ATTTATT	SLITTASIMATKIQLK		
		TCCAGAGCACGTGCGGCGGCCCGGCGACTGAAAGTTACCATCAGCGGC	GCTGTG	FKPAKCASLSLECRRG		
		AGCCTTTCCGGCTCAGTGATCACCAAGTCAACAGAGAACCGGGAAATG	GGTGAT	TKVRPLEFNVQDKIIP		
		GCTGCCAAGTCCATCACTACCGCCCTGCGTGCTCAGTCCCGCGCCCACTA	GCCCGG	ALTEEQHYRYLGVPIG		
		CACCCACCGGCTGCTGTCCCTGCCCGACCAGGGGAAAGTTGGCCAGTCC	GTTTTGT	LYRTDDSLETLVAKM		
		CTCAACCAAGACCAATACATGAACTCGTCCTCGTGGATGTCCTCCGGTA	TTACATC	TDDIQRIDSSLLAPW		
		GCTACATCCTGTTCTGTGACTGGAGATTCATCCATCGGGCCCGTCTCAAC	GGGATC	QKLDAIRTFVQPCLA		
		ACCCTTCCGACCAACGCAACGGCGCAGCGATGGAAACCCAACACCAGTC	CCGCTG	YTLRAGDCAKKHLKR		
		CCGCCTGTCGCCGTTGCCAGCACCCGCAAGAGACCCTGCCCCACATCCT	CGGCTC	LRGQLVKTARKVCNL		
		CAACCACTGCCCGCCCAACATGGTACCCATCCGCCGCAGACACAACCTC	GGTGTT	PTRATTNYIFADRRA		
		GTCCAGCAACGCATCGTGTCTGCGGTACGCCACGGGCGGGTGTTCGTC	CCATGC	GGLGFIDPNVDADIQ		
		GACCAGCACGTCCCTGAGGATCCCAACCCCCGGGAGCGCCCGGACATC	GACTGG	IITQAVRMLSSPDDIT		
		ACGGTCGTTGAGGGCGACAAAGTCACCATCATCGATGTATGTTGCCCCT	CAGCCC	RAIATGQLSSVVHRTI		
		TCGACAACGGCCGTGATGCCCTGATGACCGCCGCAGCCGCCAAGGAGA	CTTTGTT	HRAPTQEETDEFLSA		
		CCAAGTATGCCGACCTGAAGCAGGCGCTGGTGGCTGCGGGGAAAGAC	TACTCTS	SMEGDFANSGNSGQ		
		GTCGAGGTGTTCGGATTCGCCGTTGGATCGCTAGGCTCCTGGTTACCTA	GACTCT	ASSLWSRARAAARRL		

		GCAACGAGCGGGGCACTTCGACGTCTCGGCATCGCCAAGCGCTTCCGGA		ATTCATT	KVTISGSLSGSVITKST
		CTCTGATGCGCAAACTCCTGCGCATCGACGCCATCAAGGGATCCCGGGA		GTTGTAT	ENREMAAKSITTALR
		CGTGTACATCGAGCACATGTGTGGCCATCGCCAATACACCGATTAGCGA		TTCTTCC	AQSRAHYTHRLLSLP
		ACACTAATATCCTGCCMTAGACGTGATTGCTAATCCGCAAACCAACCGG		ACACTG	DQGKVGQSLNQDQY
		ATCTACAGCCTGAACACTGAACTTTAATCTTCACCCATGTCACATCTGGA		GCCAGT	MNSSSWMSSGSYILF
		CACTAGGTTTTTGCTCTGTTTGTGTTTTCCTGCCTTTWCMTTGGAWCTTT		GATCAC	CDWRFIHRARLNTLP
		CGCMCTGGAATTTATTTGTCGCTTGGATTATTTTTTTTCTCACAATTTGG		ACTGCT	TNATAQRWKPNTSP
		ATCTATTTTCGTTTGTTCACTTCGAACTCTAGCTGCCCTTTCTTCGGACAC		GTGTTTC	ACRRCQHPQETLPHI
		TGAACTTTAATCTTCGCCATGGCTGTCAGTCGCCGGTTCACTTGCTGCGG		CCGGGA	LNHCPPNMVPIRRR
		TGGGATCCTGTTGTGATAATCCCCGTGCATTGCCCATGGATTTATTTCCG		AGATAT	HNLVQQRIVSAVRH
		CCTTAGTTGTTCCTAACCTTGGATTTATTGCTGTGGGTGATGCCCGGGTT		CCTCTGC	GRVFVDQHVPEDPN
		TTGTTTACATCGGGATCCCGCTGCGGCTCGGTGTTCCATGCGACTGGCA		GGTTTTC	PRERPDITVVEGDKV
		GCCCCTTTGTTTACTCTSGACTCTATTCATTGTTGTATTTCTTCCACACTGG		ACGCTCT	TIIDVCCPFDNGRDAL
		CCAGTGATCACACTGCTGTGTTTCCCGGGAAGATATCCTCTGCGGTTTTC		GGGTGT	MTAAAAKETKYADLK
		ACGCTCTGGGTGTCTCCCCGGGCAACGCACTGGTTGCTTGCTGCGCCAT		CTCCCCG	QALVAAGKDVEVFG
		CACCCTTTTCGTTTATATTCATTTTCAGTCTGCCGTTATCTTGGCCAGCGC		GGCAAC	FAVGSLGSWLPSNER
		TCATTCTTTTGTGATGGCCGTGGACTGACCCTCTGCGGTTTTCTGCGGTC		GCACTG	ALRRLGIAKRFRTLM
		ASCACTCTCGGTGAATGCTGTGCCACCATTTTTCATTTGTTTACTTTTTCA		GTTGCTT	RKLLRIDAIKGSRDVYI
		GCTACAATTATCCTGGCCAGCTTTCACTCTTTTGTGATGGCCGTGGACGG		GCTGCG	EHMCGHRQYT (SEQ
		ACCCTCTGCTGGTTTTCACGCTCCGGGTTTGTCTGCGGTCAGCACCCTTG		CCATCAC	ID NO: 1435)
		GTGAGTGCTGGTCGTTTGTTTGCTCATTTTGCTTAGTTCACCATTATCTTT		CCTTTTC	
		MTTCTCTTTTGTWTGGTTTTCCTAGCGGTTGTCTGGGAGTTGAGCTGCA		GTTTATA	
		GTTGTCTGGTCTTGGTTTTACCCCCATTTGTTTTCTTTTAACGCGGGGCGT		TTCATTT	
		ATTGCCTTGACCGGCCGTCTCAGCTTTTCTCCTAGAGCAACCTTCCGTTC		TCAGTCT	
		ATCCAACTTTTAGTTTTCTCAGTTCTTGGCCATTCCGGTTGGTTAATTTTT		GCCGTT	
		ATTTATACTTAATTTTATGTTTACATTTTCTGGTTGGAGACCATTTTAGCT		ATCTTG	
		TGTTTTAATAGCTTTTCTTCTTTAATTAATACCCTCTGCCATTGAGGGTTT		GCCAGC	
		TTATTACTATTAATTTTGTTTACTCTTTGTAACTTGTTTGATTGAATATTTT		GCTCATT	
		AATAAACCAC (SEQ ID NO: 1067)		CTTTTGT	
				GATGGC	
				CGTGGA	
				CTGACC	
				CTCTGC	
				GGTTTTC	
				TGCGGT	
				CASCACT	
				CTCGGT	
				GAATGC	


TGTGCC
ACCATTT
TTCATTT
GTTTACT
TTTTCAG
CTACAAT
TATCCTG
GCCAGC
TTTCACT
CTTTTGT
GATGGC
CGTGGA
CGGACC
CTCTGCT
GGTTTTC
ACGCTC
CGGGTT
TGTCTG
CGGTCA
GCACCC
TTGGTG
AGTGCT
GGTCGT
TTGTTTG
CTCATTT
TGCTTA
GTTCACC
ATTATCT
TTMTTC
TCTTTTG
TWTGGT
TTTCCTA
GCGGTT
GTCTGG
GAGTTG
AGCTGC
AGTTGT
CTGGTC


TTGGTTT
TACCCCC
ATTTGTT
TTCTTTT
AACGCG
GGGCGT
ATTGCCT
TGACCG
GCCGTC
TCAGCTT
TTCTCCT
AGAGCA
ACCTTCC
GTTCATC
CAACTTT
TAGTTTT
CTCAGTT
CTTGGC
CATTCCG
GTTGGT
TAATTTT
TATTTAT
ACTTAAT
TTTATGT
TTACATT
TTCTGGT
TGGAGA
CCATTTT
AGCTTG
TTTTAAT
AGCTTTT
CTTCTTT
AATTAAT
ACCCTCT
GCCATT
GAGGGT
TTTTATT
ACTATTA

						ATTTTGT TTACTCT TTGTAAC TTGTTTG ATTGAA TATTTTA ATAAAC CAC (SEQ ID NO: 1313)		
NeS L	Utopia -1_PS	.	Phytop hthora sojae	AGACGAGCAACGCGCTGGGGCCCAAGACCTGGCACGAACGACACTGCC	AGACGAG	TGAAGC	MDVDGGPAMPEPW
				CAGGCTGTGAACGACGAGCACGCTGCTAACCCGGCAGCGCACCGGCCT	CAACGCGC	TGCACA	VLRFDGACRRNPGP
				CTGGGCTCCGCTGCACCGGTTACCGGTGCAACACGGTGGGGTCCACGTC	TGGGGCCC	AGCGCG	GGAGAALFKPCGTV
				ACGGCGCGATCGGGGCAGCGGCTGTAGCTCGCCTGCTCACAGGCCTAC	AAGACCTG	AGGCTC	AWTCSHYMPNSSET
				CCACGGCACCAGCACCAGCTACGCGCCGGCCTGCTTCGGCTCGGCGCTG	GCACGAAC	GACAGC	NNTAEYTALLLGVQS
				CCCAGACCCGCCCCGCTCCCCCCGCGGCCACGACGACAGCCCCCGATGCG	GACACTGC	TGGACC	AVHHGASHLEIEGDS
				ATGGACGTCGACGGAGGGCCGGCTATGCCCGAGCCGTGGGTGCTCCGC	CCAGGCTG	TTCAGCT	HLVVAQVKGTFACR
				TTCGACGGCGCCTGCCGCCGGAACCCGGGACCTGGTGGAGCTGGTGCC	TGAACGAC	GTCGAG	NPRLRQLRNRVRHAL
				GCCTTGTTCAAGCCGTGCGGGACCGTGGCGTGGACGTGCTCGCACTACA	GAGCACGC	CCACTG	RAVTSLTLKHIDRKAN
				TGCCAAACAGCAGCGAGACCAACACACCGCTGAATACACGGCGCTGC	TGCTAACC	CATCCG	AHADRLANRALDLKR
				TGCTTGGCGTGCAGAGTGCCGTCCACCACGGCGCCTCGCACCTGGAGAT	CGGCAGC	CGCCAG	SLAECGEHQGAMES
				CGAAGGTGACAGCCACCTCGTCGTAGCGCAGGTCAAGGGCACCTTCGC	GCACCGGC	CCACCG	CLHMNPAAQRQRE
				CTGCAGGAACCCGCGCCTCCGCCAGCTGCGCAACCGCGTCCGCCACGCC	CTCTGGGC	CACCTG	QPAPPARPACAPTRA
				CTCCGCGCCGTGACGTCCTTAACGCTCAAGCACATCGACCGGAAGGCGA	TCCGCTGC	GGGCTG	ESASDHDEDIDAEIAA
				ACGCGCACGCGGACAGACTGGCAAACCGGGCGCTGGACCTGAAGCGTT	ACCGGTTA	GTACTG	RDGGEAFPTLPIGPG
				CGCTGGCTGAGTGCGGCGAGCACCAAGGGGCGATGGAGAGCTGCCTCC	CCGGTGCA	CCGGCG	TAPARQPRLRLRQLT
				ACATGAATCCTGCTGCACAACGCCAGCGCGAGCAACCCGCACCTCCTGC	ACACGGTG	CCACCG	EDEQEAAASALQAM
				CCGCCCCGCGTGCGCACCCACGCGCGCCGAGTCTGCCAGCGATCATGAC	GGGTCCAC	CGAAGG	AEELACKIEDADSWT
				GAGGACATCGACGCAGAGATTGCCGCGCGACGGCGGCGAGGCGTT	GTCACGGC	ACAACG	SGDGYISAIPSRIRQLL
				CCCCACGCTGCCTATCGGGCCCGGGACGGCTCCAGCGCGGCAGCCGCG	GCGATCGG	GAGCGG	QPFTAAQPHPRPPLQ
				CCTTCGACTGCGGCAGCTGACCGAAGACGAGGCAGGAAGCGGCGGCGTC	GGCAGCG	CAACGC	QQRQRPPRVTRTQR
				GGCCCTGCAGGCCATGGCGGAGGAGCTAGCGTGCAAGATCGAAGACG	GCTGTAGC	CTCGCG	EHRLDEALDEMAAV
				CAGACAGCTGGACCTCGGGCGATGGATACATCAGCGCCATCCCCAGTC	TCGCCTGC	AGCGCC	QQERPTSRSAVRRAR
				GCATCCGGCAGTTGCTGCAGCCCTTCACAGCAGCCCAGCCCCACCCTCG	TCACAGGC	GCGTGG	RRVGRIRASMRQQQ
				GCCTCCACTGCAACAGCAGCGACAGCGACCGCCCCGGGTGACGCGCAC	CTACCCAC	GTCTGG	LRHDFARNESKCVED
				TCAGCGCGAGCATCGACTCGACGAGGCCCTCGACGAGATGGCGGCGGGT	GGCACCAG	GGGGGG	ILRAASAETAAEEHPE
				GCAGCAGGAGCGACCCACCAGCAGATCGGCCGTCCGCCGTGCGCGGCG	CACCAGCT	CCCGCG	TCPIDSGTLHEHFTAV

	CCGCGTCGGCCGGATCCGCGCCTCGATGCGCCAACAGCAGCTCCGCCAC	ACGCGCCG	GCGCAC	NSPRINFLPDEACGA
	GACTTCGCTCGCAACGAGAGCAAGTGCGTCGAGGACATTCTGCGTGCT	GCCTGCTT	ATCGCA	LFREAMADVGTPQE
	GCGTCGGCGGAGACCGCAGCGGAGGAGCACCCGGAGACGTGCCCCAT	CGGCTCGG	GGCTCG	RRSALTDELTMDEVE
	CGACTCGGGCACGCTCCACGAGCACTTCACGGCGGTGAACTCGCCCCGC	CGCTGCCC	GGCACG	DQLMQAATNSSPGH
	ATCAACTTCCTACCAGATGAAGCTTGCGGGGCGTTGTTCCGCGAGGCGA	AGACCCGC	GCGGTA	DGVGYDIYKKFAAQL
	TGGCCGACGTCGGGACACCGCAGGAGCGGCGGAGCGCGCTCACGGAC	CCGCTCCC	AGCTGG	VPLLHAAFQSCWRH
	GAGCTGACGATGGACGAGGTGGAGGACCAGCTCATGCAAGCGGCCAC	CCCGCGGC	TCATTTG	HRVPALWKVGFVRLI
	GAACTCCAGCCCCGGCCACGACGGCGTGGGCTACGACATCTACAAGAA	CACGACGA	ACCAAC	HKKGDPNDPANWR
	GTTCGCGGCGCAGCTGGTCCCCCTCCTCCACGCAGCCTTCCAGTCCTGTT	CAGCCCCC	AGGGCA	PICLQTAIYKLYSGLLA
	GGCGCCACCACCGGGTGCCTGCCCTGTGGAAGGTGGGGTTCGTCCGCC	GATGCG	CTACCCA	RRLSAYLEANGLLLM
	TCATCCACAAGAAGGGCGACCCCAACGACCCAGCGAACTGGCGGCCGA	(SEQ ID	GGTAGG	AQKGFRAYNGCHEH
	TCTGCCTGCAGACCGCGATCTACAAGCTCTACAGCGGTCTGCTGGCGCG	NO: 1191)	GAACCG	NFVATTLLDQTRRM
	TCGGCTGTCGGCGTACCTGGAGGCGAACGGGCTGCTGCTCATGGCGCA		CCCTTCA	RRRLYQVWYDLRNA
	GAAGGGGTTCAGGGCGTACAATGGTTGCCACGAGCACAACTTCGTGGC		AAAACC	FGSVHQDMLWYVLR
	AACCACGCTCCTGGACCAAACCCGACGCATGCGCCGCAGGCTGTACCAG		CAGGAA	LLGVERAFVERCDDIY
	GTCTGGTACGACCTGCGGAACGCGTTCGGGTCCGTTCACCAGGACATGC		GACACA	EDSYFVVGNAADGA
	TGTGGTATGTGCTCCGCCTGCTCGGCGTCGAGCGTGCGTTCGTCGAGCG		AACACC	TEPVRQEVGVYQGC
	CTGCGACGACATCTACGAGGACTCCTACTTCGTCGTCGGCAACGCGGCC		CTCCCTT	PLSPLLFIAALVPLLRA
	GACGGAGCGACGGAACCAGTGCGGCAGGAGGTGGGCGTGTACCAAGG		TAGTGA	LEKLDGVGVALADGV
	CTGTCCGCTCAGCCCCCTCCTGTTCATCGCCGCGCTGGTGCCGTTGCTCC		CATGCA	RPCTTAYADDLKVFS
	GAGCATTGGAGAAGCTCGACGGAGTGGGCGTCGCGTTGGCGGACGGG		TATTTTA	DSAAGITRCHAVVEK
	GTGCGGCCATGCACCACCGCCTACGCCGACGACCTCAAGGTGTTCAGCG		GCCTAC	FLEWTVLQANPGKC
	ACAGCGCGGCCGGCATCACGCGCTGCCACGCTGTCGTTGAGAAGTTCCT		ATTTCAG	AFLAVTRNARGNPA
	GGAGTGGACGGTGCTGCAGGCCAATCCAGGCAAGTGCGCCTTCCTGGC		TTACGG	HDKDMKLSLHDEEVS
	CGTGACACGCAACGCGCGCGGCAACCCCGCCCACGACAAGGACATGAA		AGAGGT	SIKLHDSYRYLGVGD
	GCTGTCGCTCCACGACGAGGAGGTCTCAAGCATCAAGCTGCACGACAG		TACTAAC	GFDHVRHRLQLEPKL
	CTACCGCTACCTCGGCGTGGGCGACGGCTTCGATCACGTGCGCCACCGT		TGGTAA	QQIKREAVALMQSG
	CTCCAGCTAGAGCCCAAGCTACAGCAGATCAAGCGGGAGGCGGTGGCG		ACACGA	LAPWQVVKALKTYVY
	CTGATGCAGTCGGGGTTGGCGCCGTGGCAGGTGGTGAAGGCGCTCAAG		ACACAC	PKVEYALRHLRPLQS
	ACCTACGTCTACCCCAAGGTCGAGTACGCGCTCCGCCACCTCCGACCGC		AT (SEQ	QLQGFDRVVAKGLR
	TCCAGTCCCAACTGCAGGGCTTCGACCGTGTGGTAGCCAAGGGCCTGCG		ID NO:	HLLRLPRSATNEVLYA
	GCACCTGCTACGCCTGCCCCGCTCGGCAACCAACGAGGTGCTCTACGCA		1314)	PTSSGGLGLQPLVEM
	CCAACGTCGAGCGGGGGCTTGGGGCTCCAGCCGCTGGTGGAAATGCAC			HRALQIAHAWQML
	CGCGCGCTGCAGATCGCCCATGCGTGGCAGATGCTGCACTCGAAGGAC			HSKDPAIREVARAQV
	CCAGCGATCCGGGAAGTGGCGCGCGCACAGGTGTGGCAGGTGGCCCG			WQVARKRHRLREEH
	CAAGCGTCACCGCCTCCGCGAAGAACACTGGCGGGAGCGTGACGACGA			WRERDDELVRCFLNS
	GCTCGTGCGGTGCTTCCTCAACTCCGAGCTCGCCGCCTCGCCCCACGCG			ELAASPHAEALRRHG
	GAGGCGCTGCGCCGCCACGGCGACATCGGGTCACTGTGGTCCGACGTA			DIGSLWSDVQRWLRI

			CAGCGGTGGCTGCGCATCTACCACCTCAGCCTGGTGGTGCAGGACGAC					YHLSLVVQDDRNGL
			CGCAACGGCCTGGACCCGCTCGGCCTCCGCGTGCCGCACCACGCCAAGT					DPLGLRVPHHAKWL
			GGCTCGACCACAAGAGCGTGCTGCGGCACGTCAAGTTGCACCTCAAGA					DHKSVLRHVKLHLKIR
			TCCGCCACCAGACCCGATGGAAGGGTCTGGCGGACCAGGGCAAGACG					HQTRWKGLADQGKT
			GTGCGGGCCCACGGCGGTGTCGGTGCCAAGTTCATGTCGACATGGGCG					VRAHGGVGAKFMST
			GGCCTGTCGGACGACGACTACCGCTTCGGCGTGAAGGCCCGCCTCAACC					WAGLSDDDYRFGVK
			AGATCGACACGAACGCCGTGCTCAAGCGCAAGCGGCTGCGGTCGCACA					ARLNQIDTNAVLKRK
			AGACCTGCAGGGACCCGACCTGCTCCTCGGCCGAGACGCTGGCCCATGT					RLRSHKTCRDPTCSS
			CCTCAACCATTGCGAGTCCAACATGGACGCAATCCGCCAGCGTCATGAC					AETLAHVLNHCESN
			GACGCCTTGGAGCAGATCGGGTCCAAGATCCGCAACGCGCTCAAGCGC					MDAIRQRHDDALEQ
			GGCAAGTCGACCGCGGAGCTGCGCCTGAACCAGACGGTACCCGAGTAC					IGSKIRNALKRGKSTA
			ACGGGCGCAGCTCTGCGGCCAGACATCGTACTGCGGATCGTGGCCGCC					ELRLNQTVPEYTGAA
			AAGAAGATGGTGATCGCCGACTTGGCGGTCACGTTCGAAGAGCACGCG					LRPDIVLRIVAAKKM
			GCCGGCGCCCGCCACTCCTCGCTGCAGCTCAGCCACGACCACAAGACAC					VIADLAVTFEEHAAG
			TCAAGTACCAACCGATCGTGGCGGAGCTGCAGCTCAAGGGGTGGCAGG					ARHSSLQLSHDHKTL
			TGCAGACCGCGGCGATCGTGTACGGGTCGCTGGGCTCGGTGCAGCCCA					KYQPIVAELQLKGW
			GCAACTCAACACCTACACGGAAAAGTTGAAGCTGCACAAGCGCGAGGC					QVQTAAIVYGSLGSV
			TCGACAGCTGGACCTTCAGCTGTCGAGCCACTGCATCCGCGCCAGCCAC					QPSNSTPTRKS (SEQ
			CGCACCTGGGGCTGGTACTGCCGGCGCCACCGCGAAGGACAACGGAGC					ID NO: 1436)
			GGCAACGCCTCGCGAGCGCCGCGTGGGTCTGGGGGGGGCCCGCGGCG					
			CACATCGCAGGCTCGGGCACGGCGGTAAGCTGGTCATTTGACCAACAG					
			GGCACTACCCAGGTAGGGAACCGCCCTTCAAAAACCCAGGAAGACACA					
			AACACCCTCCCTTTAGTGACATGCATATTTTAGCCTACATTTCAGTTACG					
			GAGAGGTTACTAACTGGTAAACACGAACACACAT (SEQ ID NO: 1068)					
NeSL	Utopia-1_PU	ADOS010013 21	Pythium ultimum	GCGGTGTACGCGCACAACGCCCGCGCTCTTCGAGTGCACGTTGTGCGCGC	GCGGTGTA	TGATGC		MGTQSARERGAPSA
				ACACCGCGCGGGATCTCGCCGCGCTCCAGCAGCATCGGCGCTCCGCGC	CGCGCACA	GGGTCA		PHSHTLGPRTPPRPP
				ACCGCAGCGTCCGCTTTGTGGATCACTTCCACAGCGGATGCGCGTGCGG	ACGCCGCG	TATTGAC		ACSKHGELESAAGGR
				CGTGAGCTTCCACTCGCGTGCGGCGGCAACCAAGCACGCGCGCGAATG	CTCTTCGA	CGAAAG		DGQCSDGAERERDA
				TCCAGAGAGCGCGTTCTCGGTCGCCGCCGCGCGCACTGCAGCGGC	GTGCACGT	GGCACC		ERDIRANERDCNGD
				CAACACCGCAGGTATGTCTCTCGGCGCCGACGAACGCGACCACCTCGCG	TGTGCGCG	ATCCAC		GDGDDADSDSDDRN
				TCCGTCGGCGCCTTGCATGATGTTGCATCCCCCGCTTTTGGCAACATTTT	CACACCGC	GTAGGA		DARRRSRRPRATATT
				GCCGGTTGCGTTTGCGACCGCGGCAGACGCATCAAGCGCCACCGTGAT	GCGGGATC	CACCGC		TTSAPTTTTTTTTSAT
				CGCAGACGCAGCCATGCAGCACAGTGCTGTGCCCTCTGCTGCCGCCCAA	TCGCCGCG	CCTCAA		TSATTPATDSSPWVL
				TCCCCTCGGCGTGCCGCACGTCCCCCCCCGTGCCGCGCGCCACCACGACAC	CTCCAGCA	AAACCC		RFDGACRRNPGPGG
				CATCCGCGCTGCGGATTGGTGGCAAACGCCGCCGCCTGAACGACGACG	GCATCGGC	AGTTCA		AGAALFEPGGAVVW
				GCGACAACGAAAACAGCGACGGCCGCGACGCCGACATCGAGATGCGC	GCTCCGCG	GTTTATT		TVSHYLPGSETNNTA
				GCCGACGACACCGACCACCAGCGCCGACAACCCCGCGACCAGTGCG	CACCGCAG	GACACC		EYSAMLLGVRSAIHH
				GCTGCAACGCCCGCGCGCACAGCACCAGCACCAACGGATGCCGCGACC	CGTCCGCT	CTCCACT		GATRLRVEGDSHLAL

	GCGCGCCGCCGCCACACGCGATGGGGACCCAGAGTGCTCGTGAACGTG	TTGTGGAT	TAGTGA		SQVRGTFACTNRRLR
	GCGCGCCTTCCGCGCCGCACTCACACACGCTGGGGCCCCGCACTCCGCC	CACTTCCA	CATGCA		KLRNRVQAALRELGD
	GCGACCACCAGCATGCAGCAAACATGGAGAACTCGAATCAGCTGCTGG	CAGCGGAT	TATTCAA		YRLVHIDRQANAHA
	TGGCCGCGACGGACAATGCAGCGATGGCGCGGAACGCGAGCGCGACG	GCGCGTGC	ACCGAT		DRLANRALDLRKTKV
	CCGAGCGCGACATCCGCGCCAACGAGCGCGACTGCAACGGCGACGGCG	GGCGTGA	ACATATT		DCGPHATTTDACVQ
	ACGGCGACGACGCCGACAGCGACGACCGCAACGACGCGCGA	GCTTCCAC	CGTTAG		PAEILAPTARLSSSSSS
	CGACGCAGCAGACGACCGCGAGCGACGGCGACAACAACGACGAGCGC	TCGCGTGC	GAGAGG		SSSSSSDEPMPGLEEP
	GCCGACCACGACCACGACCACGACAACGAGCGCAACGACGAGCGCGAC	GGCGGCA	TTACTAA		AADDETDADAEADIA
	GACGCCGGCGACAGACAGCTCCCCCTGGGTCTTGCGCTTTGACGGCGCA	ACCAAGCA	CTGGTA		MRDGGEIFPTLQIGP
	TGTCGCCGCAACCCAGGCCCCGGCGGCGCCGGTGCTGCGCTCTTCGAGC	CGCGCGCG	ATATATC		GSAPAQQPRLRLRQL
	CAGGTGGCGCCGTGGTGTGGACGGTGTCTCACTACCTCCCGGGCAGTG	AATGTCCA	ACCATTT		SDDESEAAARTLEHF
	AGACCAACAACACCGCCGAGTACTCGGCGATGCTGCTCGGGGTGCGCA	GAGAGCG	C (SEQ ID		ANDMASKIADADD
	GCGCCATCCACCACGGCGCGACCCGACTCCGCGTCGAGGGCGACAGCC	CGTTCTCG	NO:		WRSGEGYISAIPVRLR
	ACCTCGCGCTCTCGCAGGTGCGCGGCACCTTCGCGTGCACGAACCGGCG	GTCGCCGC	1315)		ELLAPYAVPIRSPPRN
	TCTGCGCAAGCTGCGCAACCGAGTGCAAGCGGCACTGCGCGAGCTCGG	CGCCGCGC			ASSRPPPRPQSRPPPRP
	CGACTACCGGCTCGTGCACATCGACCGGCAAGCGAACGCGCATGCCGA	GCACTGCA			PRVTRHQREHRLDEA
	CCGCCTCGCGAACCGCGCCCTCGACCTCCGCAAGACCAAGGTCGACTGC	GCGGCCAA			LDDLAAAQRSTSTDQ
	GGGCCCCACGCAACGACCACCGACGCGTGCGTGCAGCCGGCGGAAATC	CACCGCAG			RSIRNARRRVGRIRTA
	CTGGCGCCAACAGCCCGCCTGTCCAGCTCCAGCTCCAGCTCCAGCTCCA	GTATGTCT			QAQSDLRSQFATNE
	GCTCCAGCGACGAGCCGATGCCGGGACTCGAGGAGCCAGCTGCGGAC	CTCGGCGC			RACVESILRAAKPDG
	GACGAGACCGACGCTGACGCCGAAGCCGACATCGCAATGCGCGATGGA	CGACGAAC			TEPQASAGTCPIDRA
	GGCGAGATTTTTCCGACGCTCCAGATCGGCCCCGGCTCGGCGCCAGCAC	GCGACCAC			TLHAHFAGVNTPRER
	AACAACCGCGGCTCCGCCTCCGCCAGCTCTCGGACGACGAGTCCGAGG	CTCGCGTC			FDFDDALGADFRAAL
	CAGCCGCGCGCACGCTCGAGCACTTCGCCAACGACATGGCGAGCAAGA	CGTCGGCG			DVLPPPDQAADAFA
	TTGCGGACGCTGACGACTGGCGTTCCGGGGAGGGCTACATCAGTGCCA	CCTTGCAT			DELSLGEVEDQLDRV
	TCCCTGTGCGCCTCCGTGAACTCCTCGCACCATATGCGGTGCCAATCAG	GATGTTGC			VASSSPGLDGVGYDV
	GAGCCCTCCACGGAACGCCAGCTCCCGTCCCCCGCGCCCGCAATCTCGT	ATCCCCCG			FKRFRLQLLPLLHAAY
	CCGCCTCGCCCGCCGCGAGTGACTCGACACCAGCGCGAACACCGACTG	CTTTTGGC			QCCWRHRRVPATW
	GATGAGGCGCTCGACGACCTGGCCGCTGCGCAGCGCTCGACATCGACC	AACATTTT			KVGLVRLLHKKGDRA
	GACCAGCGCTCCATCCGCAACGCCCGGCGTCGAGTCGGCCGCATCCGCA	GCCGGTTG			EPNNWRPICLQQAIY
	CGGCACAAGCCCAATCCGACCTCCGCAGCCAGTTCGCCACGAACGAGCG	CGTTTGCG			KVYSGLLARRLSRWL
	CGCCTGCGTCGAGAGCATCCTCCGCGCCGCCAAACCCGACGGCACCGA	ACCGCGGC			EANERFTTAQKGFRE
	GCCTCAAGCAAGCGCGGGCACGTGCCCCATCGACCGTGCGACGCTCCA	AGACGCAT			FNGCHEHNFVASSLL
	CGCGCATTTCGCGGGCGTCAACACGCCGCGCGAGCGCTTCGACTTTGAC	CAAGCGCC			DQTRRLHRKLYAVW
	GACGCGCTGGGAGCCGACTTTCGCGCCGCGCTCGACGTTCTGCCACCGC	ACCGTGAT			YDLRNAFGSMPQPL
	CGGACCAAGCCGCCGATGCGTTCGCCGACGAGCTCTCGCTGGGCGAGG	CGCAGACG			MWRVLARLGVDTAF
	TCGAGGACCAGCTCGACCGCGTCGTCGCCTCCTCCAGTCCCGGCCTGGA	CAGCCATG			LQRCEDIYADSFFVV
	CGGAGTGGGCTACGACGTGTTCAAGCGCTTTCGGCTGCAGCTCCTGCCG	CAGCACAG			GNAADGATDPVRQE

		CTCCTCCACGCTGCATACCAGTGCTGCTGGCGCCACCGCCGCGTGCCTG	TGCTGTGC		VGVYQGCPLSPLLFIS
		CAACGTGGAAGGTGGGATTGGTGCGCCTGCTGCACAAGAAAGGCGATC	CCTCTGCT		ALIPLLRALQRLPGVG
		GCGCTGAGCCAAACAACTGGCGGCCGATCTGCTTGCAGCAGGCCATCTA	GCCGCCCA		VPLADGVRPCTTAYA
		CAAGGTCTACTCTGGGCTGCTGGCGCGCCGACTGTCGCGGTGGCTCGA	ATCCCCTC		DDLKVFSDSAAGIQQ
		GGCGAACGAACGCTTCACCACCGCCCAGAAAGGTTTCCGGGAGTTCAA	GGCGTGC		CHGTVARFLRWTGL
		CGGGTGCCACGAGCACAATTTCGTGGCATCCTCGCTCCTCGACCAGACG	GCACGTCC		RANASKCALLPVTTT
		CGTCGGCTCCATCGGAAGCTGTACGCTGTCTGGTACGACTTGCGGAACG	CCCCCGTG		ARGNPAIDDTLQLEL
		CCTTCGGTTCAATGCCGCAGCCCCTGATGTGGCGCGTCCTCGCTCGCCTC	CCGCGCGC		HGDAIARLTLQSSYAY
		GGAGTCGACACTGCGTTCCTCCAGCGGTGCGAGGACATCTACGCGGAC	CACCACGA		LGVGDGFDHVQHRV
		TCGTTTTTCGTCGTCGGCAACGCAGCCGACGGTGCCACGGACCCCGTCC	CACCATCC		QLAPKLAELKRDAVA
		GACAGGAGGTCGGCGTCTACCAAGGATGCCCGCTGAGCCCGCTGCTGT	GCGCTGCG		LLRSGLAPWQVLKAI
		TCATCTCAGCATTGATTCCTCTCCTGCGGCTCTGCAGCGGCTACCAGGC	GATTGGTG		KVYLYPRIEYALRHLR
		GTCGGCGTTCCCCTGGCGGACGGAGTCCGTCCGTGTACGACCGCGTAC	GCAAACGC		PLQSQLEGFDRAVAK
		GCGGACGACCTCAAGGTGTTCAGCGACAGTGCCGCCGGCATCCAGCAA	CGCCGCCT		GFRHLLRLPANATNE
		TGCCACGGCACTGTCGCTCGATTCCTCCGATGGACGGGGCTGCGCGCGA	GAACGAC		LLYAPVSSGGLGLLPL
		ACGCGAGCAAGTGTGCGCTGCTGCCAGTGACAACCACTGCGCGCGGCA	GACGGCG		VELHKALQIAHGWQ
		ACCCCGCCATCGACGACACGCTGCAGCTCGAGCTGCACGGCGACGCCAT	ACAACGAA		MLHSKDAAVQAIAR
		CGCCCGCCTCACCCTCCAATCGAGCTATGCCTACCTTGGGGTGGGTGAT	AACAGCGA		AQVRQVVQKRYTLD
		GGCTTCGACCACGTGCAGCACCGCGTGCAGCTCGCGCCCAAGCTCGCA	CGGCCGCG		ADHWQGRDDELVQ
		GAGCTGAAGCGCGACGCTGTGGCACTGCTGCGGTCGGGGTTGGCGCCT	ACGCCGAC		LFLNSELAASPHATIK
		TGGCAGGTACTGAAGGCCATCAAGGTCTATCTGTACCCGCGCATCGAGT	ATCGAGAT		RRNGDIGSLWSDVQ
		ACGCGCTGCGCCACCTCCGACCGCTGCAGTCGCAGCTCGAGGGGTTTGA	GCGCGCCG		RHLKTLQLRLETREPT
		TCGCGCGGTAGCGAAGGGGTTCCGACACCTCCTCCGTCTACCCGCGAAC	ACGACACC		ADAPDSPNGLLHLRV
		GCCACCAACGAGCTGCTCTACGCGCCTGTGTCGAGCGGCGGCCTGGGG	GACGCACC		PHHRKWLSHKTVLR
		CTTCTGCCACTCGTGGAGCTCCACAAAGCTCTGCAGATTGCCCACGGGT	AGCGCCGA		HMKLHIRLCHKHKW
		GGCAGATGCTCCACTCGAAAGACGCCGCCGTGCAGGCGATCGCGCGCG	CCAACCCC		QSMSDQGRTVRAH
		CACAGGTCCGCCAAGTGGTCCAAAAGCGCTACACGCTGGACGCGGACC	GCGACCAG		GQAGSHFVSRGVGL
		ACTGGCAGGGACGAGACGACGAGCTCGTGCAGCTCTTTCTCAACAGTG	TGCGGCTG		WDADYRFALQARLN
		AGCTCGCCGCCTCGCCCCACGCGACCATCAAGCGGCGCAACGGAGACA	CAACGCCC		QLDTNSTLKRRRQRT
		TTGGTTCGCTCTGGAGCGATGTCCAGCGCCACCTGAAGACGCTGCAGCT	GCGCGCAC		NATCRAPNCSRTETL
		CCGGCTCGAGACGCGCGAGCCGACTGCCGACGCGCCTGACTCCCCCAA	AGCACCAG		AHVLNHCETNMDVI
		CGGCCTCCTCCACCTCCGTGTGCCGCACCACCGCAAATGGCTCAGCCAC	CACCAACG		RQRHDGALEQIGAAI
		AAAACGGTGTTGCGGCACATGAAGCTCCACATCCGCTTGTGCCACAAGC	GATGCCGC		NAAIKGRRTDTEVRL
		ACAAGTGGCAGAGCATGAGCGACCAGGGGCGCACGGTGCGAGCGCAC	GACCGCGC		NQTVPEFNGPAWRP
		GGACAGGCTGGCAGCCACTTTGTGTCGCGCGGGGTGGGCCTATGGGAC	GCCGCCGC		DIQVRDARSKTMVIA
		GCGGACTACCGCTTCGCGCTCCAGGCCCGCCTCAACCAACTCGACACCA	CACACGCG		DLAITFEDQPNDQSA
		ACTCGACACTCAAGCGTCGCCGCCAGCGCACCAATGCCACCTGCCGCGC	(SEQ ID		SSSLQHSREHKIAKYQ
		GCCGAACTGCTCGCGCACCGAGACACTGGCGCACGTCCTCAATCATTGC	NO: 1192)		PIAAALERQGWRVH

			GAAACCAACATGGACGTGATCCGCCAGCGTCACGACGGCGCGCTCGAG CAGATTGGTGCGGCGATCAACGCTGCCATCAAGGGCCGCCGCACCGAC ACCGAGGTCCGACTCAACCAGACAGTTCCAGAGTTCAACGGACCGGCG TGGCGGCCCGACATCCAGGTGCGCGACGCGCGCTCCAAGACCATGGTG ATCGCGGACCTCGCGATCACCTTCGAGGACCAGCCAAACGACCAGTCCG CCAGCTCGTCGCTCCAACACAGCCGCGAACACAAGATCGCGAAATACCA GCCAATCGCGGCGGCACTGGAGCGGCAAGGGTGGCGGGTACACACAA GTGCCATTGTGTACGGGTCGCTACGCTCGGTCCACCCCAGCAACTTCAC CGTGTACACGGAGTTGCTGGGACTGCTCAAGCGCGACGCCCGTCGCCT GAACACAACGCTGTCGTGCCACTGCATCCGCAGCAGCCGACGCGTGTG GAACTGGCACTGCGGCCAGCACAGAGCACGACAGCATCAGCGATGCCA AGAAGGCCGCGCGCACGGGTCCGGGGGGAACCAGCGCGCGGAAGGG GGCACGGCAACAACATGATGCGGGTCATATTGACCGAAAGGGCACCAT CCACGTAGGACACCGCCCTCAAAAACCCAGTTCAGTTTATTGACACCCTC CACTTAGTGACATGCATATTCAAACCGATACACATATTCGTTAGGAGAGGT TACTAACTGGTAATATATCACCATTTC (SEQ ID NO: 1069)				TSAIVYGSLRSVHPSN FTVYTELLGLLKRDAR RLNTTLSCHCIRSSRR VWNWHCGQHRAR QHQRCQEGRAHGS GGNQRAEGGTATT (SEQ ID NO: 1437)
NeSL	Utopia-1_SM	.	Strigamia maritima	GGAGTGTTCTTTTCGGAGACGCCGCCTACTTTAGAGGAGAGAATCCCCA CGGGCATCCTCATTTGATCTGATCCATCGAGTATCTGCGAATAGTCGGC GCACTCCTTTTGCCATGATCCCGGGGGGTCTCATGGTAAAAAGGTTTGTG GCACGGCTTAGTTGACGCCCCTCTTCCACGTCACTCGGCCTGCATCGATC GACTGCTCTCCTCCTTCCTCCCTCCTCTTCACGCTCTCTTCACGTGA CCCCTTCCCATCCCGCCCCTCGGCTTTTGGCAAAGATCTGTGTGTCCTCC AAAGCACCCATCTACCATTTGCTCGAGTTGCGATTGGTCGAAGCTGCCA CGCCACTCGTCTACTCTGCCTCTCTGACCCCTTCCCCCCTCTCCCTCTGC CGCCCTCGCCTCTCGGGTCCCTTCCCATCTCCCTCCTCCCATCACGTGCT CCCTCCTCTCGTCACGTGATCGTTGCGGCTCGACATGGCCACGGTCAGA CTCAAGTACCCCTATCCTCCCGAGGGTATCTTGTGTGGACCGTGTGCCGC CAACACGAACGCCCCTCAGACGCGGCCATACTCCGATAAGTCGGGGTTG GCCAAGCATCTAAAGCTCTATCACACAAAGCCACTCTAGTGGTGGAGTGTC GTCACTGTGGTCACGAGGAGTCTGATCTGAGGAAGATGAAGAAGCACA TCTCGACGAACCACACCCCGTCGCTGCGGCCGCGCCTACCGTACCCCC TCGCCTGGGTCCCACCGCCCCTCCTCCGCCCCGGGTTATCCTCCGCCCTC GCTTCATCCCACGGCCTCGGACTCCCTCTCCTTCCTCCTCGTCATCGGG CGTCGTCTCCGGCCTCTTCCAGAAGGTCTGTATCCCTTCCCCCTGCGAGT CCCCCCGTCTCCTCCGCTTCCTCCCCGGCCGCTCGCTCGGGTCGGAATTC GCCGGACTCGCAGGGGACGGCTCCGGTCACTCCGATCGGTACCGTACG GAACTCGCCGGCCGGTAGCCCTGCCCTCAGCTACTCGACGGCCTCCCCG ATCGCCTCTACCATCACGACGCCACGCCATCTCTCGCCTGCGTCCCCCGC	GGAGTGTT CTTTTCGG AGACGCCG CCTACTTT AGAGGAG AGAATCCC CACGGGCA TCCTCATTT GATCTGAT CCATCGAG TATCTGCG AATAGTCG GCGCACTC CTTTTGCC ATGATCCC GGGGGTC TCATGGTA AAAAGGTT TGTGGCAC GGCTTAGT TGACGCCC CTCTTCCA	TGATGG GAGAGT GAGGAA TCTTCTC CACTGT GCAAAA CCATAC AGTCAG AAGATG CTAACTA CTAGTTT GATACC CTGTGC CCCCTGC AATGTC CCGCGT GTCGTA CCCAAG CCCGGC TGGCAT TGAGAC ACATTA	MATVRLKYPYPPEGIL CGPCAANTNAPQTR PYSDKSGLAKHLKLY HKATLVVECRHCGHE ESDLRKMKKHISTNH PVAAAAAPTVPPRLG PTAPPPPRVILRPRFIP RPRTPSPSSSSSASS PASSRRSVSLPPASPP VSSASSPAARSGRNS PDSQGTAPVTPIGTV RNSPAGSPALSYSTAS PIASTITTPRHLSPASP ALSAGPGSLGASPPV SPTAATVPPAPPATV PAVMAATVPFVAAT TVPSVGSSTVPQRPA GPRRPPPFPIDDWIG RIARVSSLPELDAVSR LLEDEVVKRRPPDPN ARPASLHPTRRPPPP STRPRPCHGTGGTSV

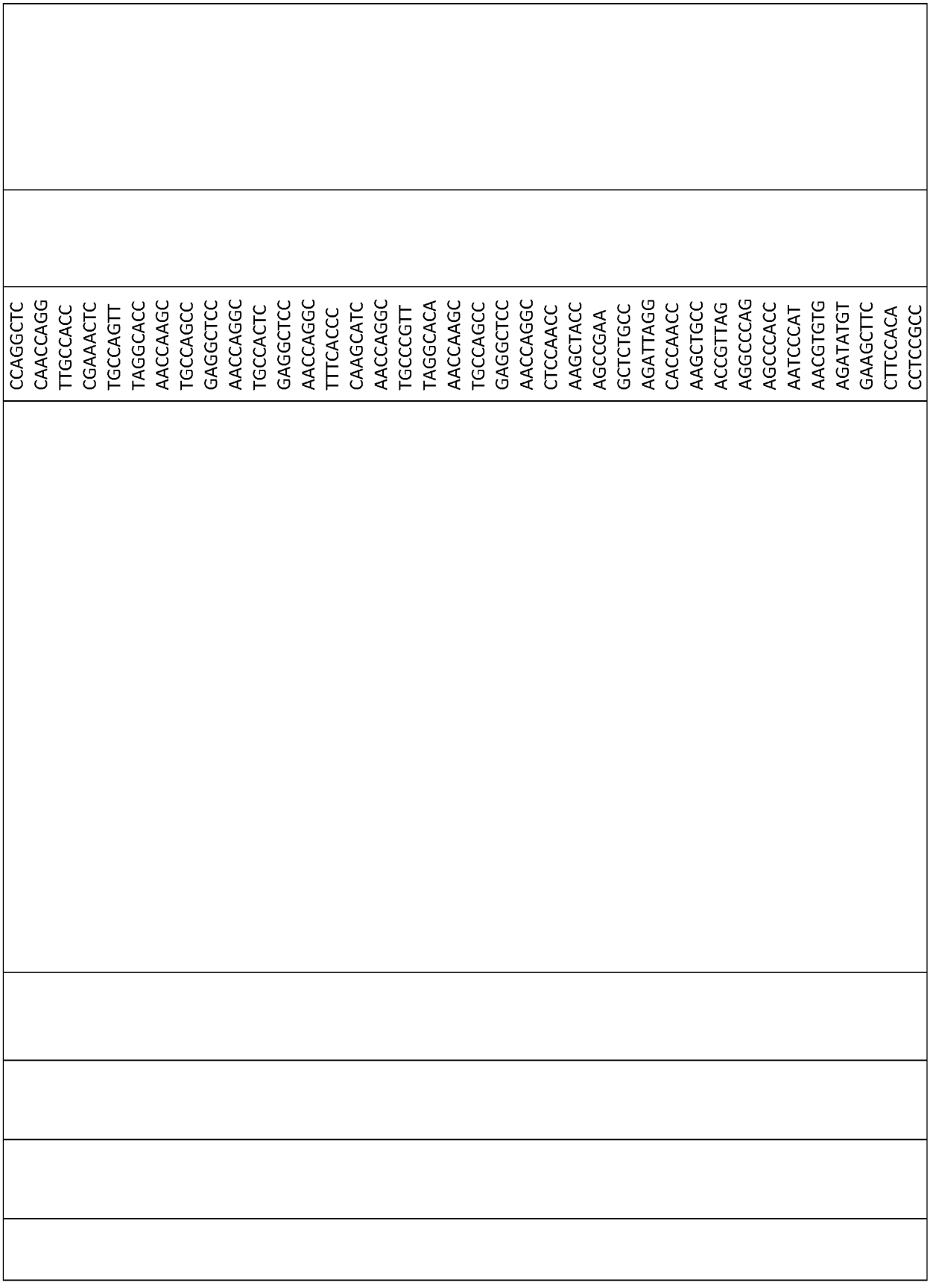
	TCTCTCTGCCGGGCCTGGCTCTCTCGGCGCGTCTCCGCCCGTCTCTCCCA	CGTCACTC	GGCTCT		SLAALTSSCIREDHRG
	CGGCCGCTACAGTACCCCCCCGCTCCTCCAGCTACCGTACCCGCAGTGAT	GGCCTGCA	CGCTCCC		SLPLLCTSGWDSPWP
	GGCGGCTACCGTACCCTTCGTCGCCGCGACTACCGTACCCTCCGTCGGC	TCGATCGA	CCGTAT		LSPSPSSHCPSSNPCP
	TCTAGCACGGTACCCCAGCGTCCCGCCGGCCCTCGACGACCCCCTCCCTT	CTGCTCTC	ACTCTCA		SSSTPPSSLLGPPRHS
	CCCTATCGACGATTGGATCGGAAGGATCGCCCGAGTGTCATCTCTGCCG	TCCTCTTCC	TAATTTC		HLTWRGSGSTTPSHR
	GAATTGGACGCCGTCTCAAGGTTGCTCGAGGATGAAGTGGTCAAGAGA	TCCCTCCC	GTGTAC		HCSRARYHHLLWRLP
	CGCCCCCCTGACCCCAACGCCAGACCCGCCTCACTCCATCCGACACGCC	TCCTCTTC	GCTAAT		YPSSPRLPYPPSALAR
	GCCCTCCCCCTCCATCTACTCGCCCTCGTCCCTGCCATGGGACCGGCGGA	ACGCTCTC	CCTACCC		YPSVPPALDDPLPSLS
	ACTTCAGTCTCTCTGGCGGCGCTCACGTCCTCGTGCATTCGCGAAGATCA	TTCACGTG	TACCCCT		TIGSEGSPECHLCRN
	CCGAGGGAGCCTCCCCCTTTTGTGCACTTCCGGTTGGGACAGTCCATGG	ACCCCTTC	CCCTTTG		WTPSQGCSRMKWS
	CCACTTTCTCCAAGTCCATCAAGCCACTGCCCATCTTCCAACCCCTGTCCC	CCATCCCG	ACCACTC		RDAPLTPTPDPPHSIR
	TCTTCCTCCACTCCGCCCTCCTCGCTCCTCGGACCCCCTCGTCACTCCCAT	CCCCTCGG	ACCCAA		HAALPLHLLALVPAM
	CTCACCTGGAGAGGTTCTGGATCGACTACGCCGAGCCACCGACACTGCT	CTTTTGGC	CCATGT		GPAELQSLWRRSRPR
	CCCGGGCCAGATACCATCATCTACTCTGAGTGGAGGGCCATCGACCCGA	AAAGATCT	GTATAG		AFAKITEGASPFCALP
	CTGGACGACTCCTCTCCTCTCTTCCAGAAGGTCCAGACCTTCGGGGCG	GTGTGTCC	CTGTGC		VGTVHGHFLQVHQA
	CCAACCCGATGGAAGGAGAGTACGACCACACTCATCCACAAGGGTGGT	TCCAAAGC	TGGTGA		TAHLPTPVPLPPLRPP
	GACCACACAGCGATGTCCTCTTGGCGTCCAATCGCCCTCCTCTCCACAGT	ACCCATCT	TCCCGG		RSSDPLVTPISPGEVL
	GGCCAAGATCTATGGGAGCATCCTGTCTCATCGGCTGACGACCTGGGCC	ACCATTTG	GGCGGT		DRLRRATDTAPGPDT
	GTCCAGAATGGTCGTCTCTCGCTCTCGCAGAAGGGATTCCTGCCATTTCG	CTCGAGTT	TATTCAC		IIYSEWRAIDPTGRLL
	AGGATGCCTCGACCAGAACTACCTTGTTCAGTCGTGCCTCCAAGACGCC	GCGATTGG	TGGTTAT		SSLFQKVQTFGAPTR
	CGCCGCAACAAGAAGACACTCTCACTCGCCTTCCTCGATCTGAAGAATG	TCGAAGCT	CATATCA		WKESTTTLIHKGGDH
	CCTTTGGGAGCATCCCCCACCTGACCATTCGTCACTCCCTCGAGTGGCTC	GCCACGCC	TTCTAAA		TAMSSWRPIALLSTV
	GGACTGGCCCCTTCATCGATCGACATTCTTGAAGCGTCATTCCTGGGATC	ACTCGTCT	ATGATCT		AKIYGSILSHRLTTWA
	GTCGACTCGCGTTCGAACCGAGACTGGACTCACTCCCCCCATCAGCCTC	ACTCTGCC	TTGATCT		VQNGRLSLSQKGFLP
	GACACTGGAGTGGTCCAAGGAGCTCCCCTCTCCCCATCCTCTTCAATCT	TCTCTGAC	CTCAATT		FRGCLDQNYLVQSCL
	AGCCATCGAACCCCTCCTCCGCACGGTCCCCTCTGCCCATTCTGGATTTT	CCCTTCCC	AACAAC		QDARRNKKTLSLAFL
	CTCTCCATGGCCATGTCGTCTCTGTGGTCGCCTACGCGGATGACCTCGCC	CCCTCTCT	TAACTTA		DLKNAFGSIPHLTIRH
	ATCCTCGCCCCATCTACTCCCGCTCTACAATCCCAATTGGATGCTATCTCT	CCCTCTGC	TTTCTGT		SLEWLGLAPSSIDILE
	GGCATGGCAGATTGGGCGGGCCTCTCATTCAATCCGGCCAAGTGCGCG	CGCCCTCG	TTGTTTC		ASFLGSSTRVRTETGL
	ACCGTCACTCTCACTGGGAAGGATAACTCTCGTGATACGCTGTCCCTGC	CCTCTCGG	ATTGTTT		TPPISLDTGVVQGAP
	AAGGCTCCCCCGTCCCTTCCATCTCTGACGGCGATGCATACAAGCACCTT	GTCCCTTC	CACCTC		LSPILFNLAIEPLLRTV
	GGAGTCCCCACTGGCACTACTACCTTCCCCAGTGGCACCGATGCAATCA	CCATCTCC	GTAAGA		PSAHSGFSLHGHVVS
	AGAAGATGACGACGGATCTACAGGCCATAGACCACTCCGACCTCGCACC	CTCCTCCC	GGAAGT		VVAYADDLAILAPSTP
	GGCCCAGAAGCTAGATGCCCTCCGCACTTTCATCATGCCCCAGCTCTCCT	ATCCACGT	TCATTGT		ALQSQLDAISGMAD
	TCCATCTCTCTCACGGCTCAGTCCCCAAAGCCCCCCTCACTCAACTGGAC	GCTCCCTC	GCGATA		WAGLSFNPAKCATV
	AAGAAGATCAAGCGGGCTGCCAAGCACTGGCTGTTCCTTCCTCAACGAG	CTCTCGTC	AATCAA		TLTGKDNSRDTLSLQ
	CCTCGAACGAAATACTGTACATGTCCCACTTGCACGGAGAGGGCAATCCCT	ACGTGATC	(SEQ ID		GSPVPSISDGDAYKH
	ACTCCCCCTCTCGGTGCTAGCGGACATCGGACAAGTGACTCACGCCGTC	GTTGCGGC			LGVPTGTTTFPSGTD

			GCTCTTCTTCAATCAAGGGATCCAGCCGTGGCTGACCTCGCTCTCCGGAC	AAGGCTCA	TGACGA	AIKKMTTDLQAIDHS
			TTGCCGAGAAGTCGCATCGAAACGGGCAAAGAAGACGGTCAACGGCCC	AACCAGGC	AATGTTC	DLAPAQKLDALRTFI
			CGAGTTGGCCCAGTACCTCTCCGGGTCCACTGACGGAATCTACTGCACC	TGCCAACC	AATATAT	MPQLSFHLSHGSVPK
			CCGACTTCTGACATCCCGAGTCTCTGGACGACTGCCCGAGCTGCAACAC	AAGCTGCC	GACATT	APLTQLDKKIKRAAK
			GCCGACTCTCCTCCACTCTGCCACTCACGTGGACATCCCCCCCTTCCCTCTG	AGCTTAGG	CTAATGT	HWLFLPQRASNEILY
			GCGTCCCCTTCCTTTCTATCAATGGGTCCCCCCTCTCCCCATTCCGAGTAC	CACCAACC	TCATGTA	MSHLHGGQSLLPLSV
			AGTCCACCCTCACCAATGCCATTAGACATAACCACCTCTCCACTCTCATT	AAGCTGCC	TTTTTTG	LADIGQVTHAVALLQ
			GCCAAACGGGACCAGGGAAATTCCTACCGTACCTCACACGACCCCGACC			SRDPAVADLALRTCR
			CCTCTAACTACTGGGTCAAGGGAGGAGACTTCCTCCGATTCTGTGACTG			EVASKRAKKTVNGPE
			GCGATTCATACACAGAGCCCGACTGAACCTCCTCCCAGTGAACGGGGCG			LAQYLSGSTDGIYCTP
			CGCCGATGGGACGCAAACAGCATCAAGACGTGCCGGCGCTGTGGTGCG			TSDIPSLWTTARAAT
			CCCAATGAGACACTCGCCCACGTGCTCAACGTCTGCCCTGTCGGACTGC			RRLSSTLPLTWTSPLP
			CGGAGATGAAGAAGAGGCACGATGCAATTCATGCCCGCATCAAGAAAG			SGVPFLSINGSPLSPF
			CCCTCCGACCCTCTCCCCATACGGTAGTACACCACGACCGGACAGTTCCC			RVQSTLTNAIRHNHL
			GGATGCGGCCCCCTGAGACCGGACATTCTGAGAATCAGCGAGAGAGAC			STLIAKRDQGNSYRTS
			AAATCTGTTGCCATTGTAGATATCCACGTTCCCTTCGACAATGGCACTGA			HDPDPSNYWVVKGG
			TGCGGTTGAGCGAGCTCACGAGACAAAGCGGGCGAAGTACGAACTGAT			DFLRFCDWRFIHRAR
			CCGGCGGCACTACGAACACCAAGGGTATCGAGTCACGTTTGACTCACTA			LNLLPVNGARRWDA
			GTAGTCACAGCGCTTGGTAGACTGTGGCGGGGGAGTGAGGCAGCGCTA			NSIKTCRRCGAPNETL
			CAAGCGCTCCAGATCTCGTCACAGTATAGTAAGCTCCTCCGTAAGTTACT			AHVLNVCPVGLPEM
			CGTGGCTGATGCCATCCACGGCTCGAGAAACGTGTACGCCCATCATATG			KKRHDAIHARIKKALR
			ACCGGTATGGTTATGTGATGGGAGAGTGAGGAATCTTCTCCACTGTGCA			PSPHTVVHHDRTVP
			AAACCATACAGTCAGAAGATGCTAACTACTAGTTTGATACCCTGTGCCCC			GCGPLRPDILRISERD
			CTGCAATGTCCCGCGTGTCGTACCCAAGCCCGGCTGGCATTGAGACACA			KSVAIVDIHVPFDNG
			TTAGGCTCTCGCTCCCCCGTATACTCTCATAATTTCGTGTACGCTAATCCT			TDAVERAHETKRAKY
			ACCCTACCCCTCCCTTTGACCACTCACCCAACCATGTGTATAGCTGTGCT			ELIRRHYEHQGYRVT
			GGTGATCCCGGGGCGGTTATTCACTGGTTATCATATCATTCTAAAATGAT			FDSLVVTALGRLWRG
			CTTTGATCTCTCAATTAACAACTAACTTATTTCTGTTTGTTTCATTGTTTCA			SEAALQALQISSQYSK
			CCTCGTAAGAGGAAGTTCATTGTGCGATAAATCAA (SEQ ID NO: 1070)			LLRKLLVADAIHGSRN
						VYAHHMTGMVM
						(SEQ ID NO: 1438)
NeSL	Utopia-1_SP	AAGJ021405137	Strongylocentrotus purpuratus	AAGGCTCAAACCAGGCTGCCAACCAAGCTGCCAGCTTAGGCACCAACCA	AAGGCTCA	TGACGA	MSHSITEVFDYPLPSR
				AGCTGCCAGTCCAGGCTCCAACCAGGCTGCCCACCAATCTGCCAGCTTA	AACCAGGC	AATGTTC	WKCTVCLENFFNQQ
				GGCTCCAACCAAGCTGCCCACCTAGCTGCCAGCTTAGGCACCAACCAAG	TGCCAACC	AATATAT	TLKRHQARHHQTTSF
				CTGCCAGCCGAGGATCCAACCAGGCTGCCCACCTAGCTGCCAGCTTAGG	AAGCTGCC	GACATT	LYVFRCSACQAEFDS
				CTCCAACCAAGCCGCCAGCCCAGGCTCCAACCAGGTTGCCACCCGAAAG	AGCTTAGG	CTAATGT	ARKASNHWQSHKRK
				TCTGCCAGTTTAGGCACCAACCAAGCTGCCAGCCGAGGCTCCAACCAGG	CACCAACC	TCATGTA	PILSQPAVNEIPSSGL
				CTGCCACTCGAGGCTCCAACCAGGCTTTCACCCGAAGCATTACCCAGGC	AAGCTGCC	TTTTTTG	DPSPPRSRPPVEVIGS

		TGCCCGTTTAGGCACAAACCAGGCTGCCAGCCGAGGCTCCAACCAAGCA	AGTCCAGG	TTTGCTT	SFPDDVSMLSEPSTP
		ACCAGCCCAGGTTCCAATCAAGCTGCCAGCCGAGGCTCCAAGTAAGCTG	CTCCAACC	GACAAA	STSLQMDPEVVHPPS
		CCAGCCGAGGCTCCAACCAGGCTGCCACTCGAGGCTCCAACCAGGCTTT	AGGCTGCC	GCTAGA	RSISFSPMHLSPTQPA
		CACCCGAAGCATTACCCAGGCTGCCCGTTTAGGCACAAACCAAGCTGCC	CACCAATC	TGAATA	SPIQIGVEVSFNSSSSL
		AGCCGAGGCTCCAACCAAGCAACCAGCCCAGGTTCCAATCAAGCTGCCA	TGCCAGCT	TATTCCC	QMDPEVVQPPSPSIS
		GCCGAGGCTCCAACCAAGCTGCCAGCCGAGGCTCCAACCAGGCTGCCC	TAGGCTCC	TGTCCTA	FSPMHLSPTQPASPI
		ACCAAGCTGCCAACTTAGGCTCCAACCAAGCTGCCAGCCCAGGCTCCAA	AACCAAGC	CTACATA	QIGVEVSFNSSGSLQ
		CCAGGTTGCCACCCGAATCTCTGCCAGTTTAGGCACCAACCAAGCTGCC	TGCCCACC	CATCTCG	MDPEVVQPPSPSISF
		AGCCTAGGCTCCAACCAGGCTGCCACTCGAGGCTCCAACCAGGCTTTCA	TAGCTGCC	ATGTGC	SPMHLSPTQPVSPIQI
		CCCGAAGCATCACCCAGGCTGCCCTTTTAGGCACAAATCAAGCTGCCAG	AGCTTAGG	CCTTGCT	GIEVSFTSSSPLQMD
		CCGAGGCTCCAACCAAGCAACCAGCCCAGGTTCCAATGAAGCTACCAGC	CACCAACC	AGACGA	PEVVQPPSPSMSYSP
		CGAGGCTCCAACCAAGCTGTCAGCCGAGGCTCCAACCAGGCTGCCCACC	AAGCTGCC	TCCATTG	MHLSLTQPDSPIPVDI
		AAGCTGCCAGCTTAGGTACCAACCAAGCTGCCAGCCCAGGCTCCAACCA	AGCCGAG	GTAACC	DVIPAAEVPLPDIEIPP
		GGTTGCCACCCGAAACTCTGCCAGTTTAGGCACCAACCAAGCTGCCAGC	GATCCAAC	ATGATA	SPDRHPAAEVPLPDI
		CGAGGCTCCAACCAGGCTGCCACTCGAGGCTCCAACCAGGCTTTCACCC	CAGGCTGC	TCAATG	EIPPSPDRHPVAEVPL
		CAAGCATCAACCAGGCTGCCCGTTTAGGCACAAACCAAGCTGCCAGCCG	CCACCTAG	GATTAA	PDIEIPPSPDRHPQSP
		AGGCTCCAACCAGGCCTCCAACCAAGCTACCAGCCGAAGCTCTGCCAGA	CTGCCAGC	CACATG	PRPVMMEQPVHTPP
		TTAGGCACCAACCAAGCTGCCACCGTTAGAGGCCCAGAGCCCACCAATC	TTAGGCTC	ATCTGTA	PADTQQANGPQHW
		CCATAACGTGTGAGATATGTGAAGCTTCCTTCCACACCTCCGCCGGTCTC	CAACCAAG	AACAAT	VTVLANATNWEDFG
		CGTCGCCACACGGCCAGGCTTCATCGCACCACTGCTGAACACACTGACG	CCGCCAGC	ATATGA	RVCVEFANHAVEAA
		ACAGCATGAGCCATTCCATCACCGAGGTATTCGACTATCCGTTGCCCTCC	CCAGGCTC	TTTACAC	RSRQDAPQVRPAAQ
		CGTTGGAAGTGCACAGTGTGCTTAGAGAACTTTTTCAATCAGCAGACCC	CAACCAGG	AATGTA	RQPRRPTPRPQPTFD
		TCAAGCGCCACCAGGCCCGGCACCACCAAACGACCTCCTTCCTCTACGTC	TTGCCACC	TTTATTG	VREASRLQKLYKRSKK
		TTTCGATGCTCAGCCTGCCAGGCTGAGTTTGATTCTGCTCGTAAGGCTTC	CGAAAGTC	TTTCAAT	RAVRHILRDDAPSFS
		TAATCACTGGCAGAGCCACAAGAGGAAACCAATCCTCTCACAACCAGCA	TGCCAGTT	AAATCT	GSNEQLLDYFKEIYAP
		GTGAATGAAATCCCTTCTTCGGGTCTAGATCCATCCCCTCCAAGGTCTCG	TAGGCACC	GTTCTTT	PEIDENRAQQLAESL
		TCCACCGGTCGAGGTGATTGGTTCTTCTTTTCCAGATGATGTGTCTATGC	AACCAAGC	TCACTTT	FTDLEEAKESAAALM
		TGTCAGAACCATCTACCCCCTCTACATCACTACAAATGGACCCAGAGGTG	TGCCAGCC	AATACA	SPISQQEISTRLSRMS
		GTGCATCCTCCGTCACGATCCATATCCTTTTCGCCCATGCATCTTAGCCCC	GAGGCTCC	TGAAAC	NSAPGKDRLEYRHIR
		ACTCAGCCAGCCAGCCCGATTCAAATAGGCGTTGAGGTTTCCTTTAACTC	AACCAGGC	ATGACT	QADGACRVTHIMFN
		TTCCAGCTCACTTCAGATGGACCCAGAGGTCGTGCAGCCCCCGTCACCG	TGCCACTC	GTGCCT	RCLQEHRIPSAWKEA
		TCCATTTCCTTCTCACCCATGCATCTTAGCCCCACTCAGCCCGCCAGCCCG	GAGGCTCC	TCTCCAA	TTILIHKSGTTDDPAN
		ATTCAAATAGGCGTTGAGGTTTCCTTTAACTCTTCCGGCTCACTTCAGAT	AACCAGGC	CTGGAG	FRPIALQSCLYKLFMG
		GGACCCAGAGGTGGTACAGCCCCCGTCACCGTCCATCTCCTTCTCACCCA	TTTCACCC	ACCTAC	ILSDRMTQWACNHN
		TGCATCTTAGCCCCCACCCAACCAGTCAGCCCGATTCAAATAGGCATTGA	GAAGCATT	ATAATTT	LLSPEQKSARPCEGC
		GGTTTCCTTTACCTCTTCCAGCCCACTTCAGATGGACCCCGAGGTAGTGC	ACCCAGGC	GTTAAA	HEHTFLLSSVIKDTKR
		AGCCCCCGTCACCCTCCATGTCCTACTCACCCATGCATCTTAGCCTCACTC	TGCCCGTT	ATGATA	NQKTANIAWLDLRN
		AGCCAGACAGCCCCATCCCAGTAGACATTGATGTCATACCAGCGGCTGA	TAGGCACA	TAAATAT	AFGSIPHQAIHAVLTT

	GGTCCCCTTGCCTGATATTGAGATTCCTCCTTCTCCGGATAGACATCCAG	AACCAGGC	TTCGAA		IGAPVSLVMLLKDTYT
	CGGCTGAGGTCCCCTTGCCTGATATTGAGATTCCTCCTTCTCCGGATAGA	TGCCAGCC	GATGAA		GASTSFLSTSGETDPI
	CATCCAGTGGCTGAGGTCCCCTTGCCTGATATTGAGATTCCTCCTTCTCC	GAGGCTCC	ATTATTA		QIQSGVKQGCPMSAI
	GGATAGACATCCACAATCGCCTCCACGACCAGTGATGATGGAGCAGCCA	AACCAAGC	TTAATAA		LFNLTIELIIRAVKKKA
	GTCCATACACCACCGCCGGCCGATACACAGCAGGCTAATGGTCCACAGC	AACCAGCC	(SEQ ID		TDDGLGLVVHGQRL
	ATTGGGTGACTGTTCTAGCCAACGCCACCAATTGGGAAGACTTCGGGAG	CAGGTTCC	NO:		SIMAYADDLVLMSKT
	AGTCTGTGTAGAGTTTGCTAATCATGCTGTAGAAGCAGCCCGGTCTCGG	AATCAAGC	1317)		PEGLDAILSVASEQAE
	CAAGATGCACCTCAAGTAAGGCCTGCTGCACAGAGACAGCCTCGCAGA	TGCCAGCC			TLRLAFKPTKCASLSLS
	CCGACTAGACCTCGTCAACCCACGTTTGATGTCAGGGAGGCAAGCCGTT	GAGGCTCC			CRHGTSVLPREYTVQ
	TGCAGAAGCTCTACAAACGCAGCAAGAAAAGAGCGGTTCGGCACATTC	AAGTAAGC			GHLMPALDEEEQYR
	TAAGAGATGACGCCCCCCTCCTTCTCTGGCAGCAATGAGCAACTGCTTGA	TGCCAGCC			YLGVPFGLPRFTNLK
	CTACTTCAAAGAAATCTATGCTCCTCCCGAGATTGATGAAAATAGAGCA	GAGGCTCC			DLIGKLKGNIETIASSL
	CAACAGTTGGCAGAGTCTCTCTTTACCGATCTTGAGGAAGCGAAGGAGT	AACCAGGC			LAPWQKLDAIKTFVQ
	CTGCGGCCGCCCTAATGTCTCCTATATCTCAGCAAGAGATCAGTACACG	TGCCACTC			PGLSFVLRAADYLKSD
	CCTTTCTCGAATGAGCAATTCTGCCCCCTGGAAAGGACCGCCTCGAGTAT	GAGGCTCC			LRSLKSAITTNVKKIC
	AGACATATTCGACAAGCTGATGGTGCATGCCGTGTTACCCACATCATGT	AACCAGGC			QLPLRAANAYIFAAK
	TCAACAGGTGCCTACAGGAACACCGTATACCATCAGCATGGAAGGAAG	TTTCACCC			ESGGLAFIDPNVDAD
	CTACAACCATCCTCATTCACAAGTCCGGTACAACCGATGACCCAGCAAAC	GAAGCATT			IQVITQAVRVLSSDDE
	TTCCGCCCGATCGCACTCCAGTCATGCTTGTACAAACTCTTCATGGGCAT	ACCCAGGC			VVQTIATSQLKSVVH
	CCTCTCAGACAGAATGACACAGTGGGCGTGCAATCACAACCTTCTCAGC	TGCCCGTT			RTIHAVPTEEDIDNYL
	CCAGAGCAGAAGAGCGCTCGCCCATGTGAGGGATGTCATGAACACACC	TAGGCACA			SGSNEGLLANSGNSG
	TTTCTTCTGTCGTCTGTCATCAAAGATACCAAGCGAAACCAGAAGACTGC	AACCAAGC			QASSLWSRTRSAARR
	CAACATCGCCTGGCTAGATCTTAGGAACGCCTTCGGAAGCATCCCCCAC	TGCCAGCC			LHLTLRATTSGTVVV
	CAGGCTATCCATGCAGTCCTGACAACGATAGGAGCACCTGTCAGCCTGG	GAGGCTCC			NQQADIDHTRDILPA
	TTATGCTCCTGAAGGACACATACACGGGAGCATCCACATCCTTTCTTAGC	AACCAAGC			SITRGLRLIQRTTNAE
	ACGAGTGGAGAGACTGACCCAATACAGATCCAGTCCGGTGTTAAACAA	AACCAGCC			KLKSLPDQGKVARSL
	GGATGCCCGATGTCTGCCATCCTCTTCAACCTAACCATTGAACTTATCAT	CAGGTTCC			SNDPFANGSSWHAT
	CAGAGCCGTCAAGAAGAAGGCAACGGATGATGGCCTCGGGCTTGTGGT	AATCAAGC			GKFIRFCDWRFIHRA
	GCATGGACAACGCCTATCCATCATGGCCTACGCTGACGACCTCGTCCTG	TGCCAGCC			RLNCLPTNVATKRW
	ATGAGCAAGACACCAGAAGGCCTGGATGCCATTTTGTCTGTGGCATCCG	GAGGCTCC			KANANGKNGHQQE
	AGCAGGCAGAAACCCTACGTCTGGCATTCAAGCCTACCAAGTGTGCCTC	AACCAAGC			TLPHVLNHCLPNMV
	CCTCTCACTCTCATGTAGACACGGTACATCTGTCCTTCCACGCGAATACA	TGCCAGCC			PIRRRHDNIQQRLVT
	CCGTCCAAGGTCACCTGATGCCAGCTCTCGACGAGGAAGAGCAGTACC	GAGGCTCC			AIRHGDVFVNQHVP
	GCTACCTCGGAGTTCCCTTCGGCTTACCTCGTTTCACAAATCTCAAGGAC	AACCAGGC			GDPNPRERPDITVIE
	CTGATCGGCAAGCTGAAGGGAAACATCGAGACCATTGCATCATCATTGC	TGCCCACC			GNKVTVIDISVPFDN
	TGGCCCCATGGCAAAAGTTGGATGCAATCAAAACCTTCGTCCAACCAGG	AAGCTGCC			GPNACTTAAQAKVE
	TCTCAGTTTCGTACTGCGGGCAGCGGACTACTTGAAGTCGGACTTGCGT	AACTTAGG			KYSALRQALRDMGR
	TCTCTGAAATCTGCCATCACCACCAACGTCAAAAAGATATGCCAGCTTCC	CTCCAACC			DVEVHGFIVGALGT

		CCTTCGGGCGGCCAACGCATACATCTTCGCGGCTAAGGAATCGGGTGGT	AAGCTGCC		WHQGNERALGRLGV
		CTAGCCTTCATAGACCCCAACGTAGATGCGGACATCCAAGTCATCACTCA	AGCCCAGG		SRWYRTLMRKLCCID
		AGCCGTTCGTGTCCTGTCGTCTGACGATGAAGTCGTTCAGACAATAGCA	CTCCAACC		AIQASRDIWVEHVTG
		ACATCCCAACTCAAGTCGGTTGTTCATCGCACAATCCACGCAGTCCCTAC	AGGTTGCC		HRQYE (SEQ ID NO:
		TGAAGAGGACATCGACAACTACTTGTCTGGCAGCAACGAGAGGGCCTGCTT	ACCCGAAT		1439)
		GCCAATTCGGGTAATAGCGGCCAAGCTTCTTCTCTGTGGTCTCGAACCA	CTCTGCCA		
		GGTCAGCCGCAAGAAGACTTCATCTTACCCTCAGAGCCACCACCAGTGG	GTTTAGGC		
		TACAGTTGTTGTCAACCAGCAGGCAGACATCGACCACACACAAGGGACATT	ACCAACCA		
		CTACCAGCATCAATAACAAGAGGACTGAGGCTCATCCAACGAACCACAA	AGCTGCCA		
		ATGCTGAGAAACTCAAGTCCCTACCAGACCAAGGGAAAGTTGCCAGATC	GCCTAGGC		
		CCTCAGCAACGATCCATTCGCTAACGGCTCTTCTTGGCATGCAACTGGAA	TCCAACCA		
		AGTTCATCCGCTTCTGCGATTGGCGGTTCATTCACCGAGCCAGACTGAA	GGCTGCCA		
		CTGCCTCCCCACGAATGTAGCCACGAAGAGGTGGAAGGCAAATGCGAA	CTCGAGGC		
		TGGCAAGAATGGCCATCAACAAGAGACACTGCCTCACGTACTGAATCAT	TCCAACCA		
		TGCCTACCCAACATGGTCCCCATCAGACGCCGCCATGACAACATCCAAC	GGCTTTCA		
		AACGACTGGTAACAGCCATCAGACATGGCGATGTCTTTGTCAATCAACA	CCCGAAGC		
		TGTACCGGGAGACCCCAACCCAAGAGAACGCCCAGATATCACTGTTATC	ATCACCCA		
		GAGGGAAACAAAGTCACAGTCATCGATATTTCTGTCCCCTTCGACAATG	GGCTGCCC		
		GTCCCAACGCATGCACCACAGCTGCACAAGCCAAAGTCGAAAAGTACTC	TTTTAGGC		
		TGCCCTCCGTCAAGCCTTGAGAGATATGGGAAGAGACGTCGAAGTCCAT	ACAAATCA		
		GGCTTCATAGTTGGTGCCTTAGGAACCTGGCACCAGGGCAACGAACGT	AGCTGCCA		
		GCTCTCGGACGCCTGGGTGTCAGTCGCTGGTACAGAACACTGATGAGA	GCCGAGG		
		AAGCTGTGCTGCATTGATGCTATCCAAGCTTCCAGAGACATTTGGGTTG	CTCCAACC		
		AACATGTCACTGGACATCGACAGTACGAGTGACGAAATGTTCAATATAT	AAGCAACC		
		GACATTCTAATGTTCATGTATTTTTTGTTTGCTTGACAAAGCTAGATGAA	AGCCCAGG		
		TATATTCCCTGTCCTACTACATACATCTCGATGTGCCCTTGCTAGACGATC	TTCCAATG		
		CATTGGTAACCATGATATCAATGGATTAACACATGATCTGTAAACAATAT	AAGCTACC		
		ATGATTTACACAATGTATTTATTGTTTCAATAAATCTGTTCTTTTCACTTTA	AGCCGAG		
		ATACATGAAACATGACTGTGCCTTCTCCAACTGGAGACCTACATAAATTTG	GCTCCAAC		
		TTAAAATGATATAAATATTTCGAAGATGAAATTATTATTAATAA (SEQ ID	CAAGCTGT		
		NO: 1071)	CAGCCGAG		
			GCTCCAAC		
			CAGGCTGC		
			CCACCAAG		
			CTGCCAGC		
			TTAGGTAC		
			CAACCAAG		
			CTGCCAGC		


CCAGGCTC
CAACCAGG
TTGCCACC
CGAAACTC
TGCCAGTT
TAGGCACC
AACCAAGC
TGCCAGCC
GAGGCTCC
AACCAGGC
TGCCACTC
GAGGCTCC
AACCAGGC
TTTCACCC
CAAGCATC
AACCAGGC
TGCCCGTT
TAGGCACA
AACCAAGC
TGCCAGCC
GAGGCTCC
AACCAGGC
CTCCAACC
AAGCTACC
AGCCGAA
GCTCTGCC
AGATTAGG
CACCAACC
AAGCTGCC
ACCGTTAG
AGGCCCAG
AGCCCACC
AATCCCAT
AACGTGTG
AGATATGT
GAAGCTTC
CTTCCACA
CCTCCGCC

NeSL	Utopia-1_TSP	.	Trichinella spiralis	TTTCTGGTATGAATCCCAAGCGGATTCGTTACGAAATTTGCATAAGTTTT	GGTCTCCG TCGCCACA CGGCCAG GCTTCATC GCACCACT GCTGAACA CACTGACG ACAGC (SEQ ID NO: 1194)	TTTCTGGT	TGATCC	MCSAKTPALKSGRRR
				TGAAAAAATAGGCATTTGGTCGAGTGCTCGCACCACCATTTGTCGCGGG		ATGAATCC	GTCCCG	GKEVNYEGQIVRVER
				TCGTCCTGATATTGCACTACATTCAGGAACGGCCTATTCCCTTCGGGGAA		CAAGCGG	AACCAA	RRGSRSSTSATDLGT
				TTGTGTTTTAGGAATTGGAATCGGTTTGGTTACGATCGGTCGAGTGGTT		ATTCGTTA	CGGAAC	RMVTRGRKKLMEAS
				CGTGAGATCGAGTGACAGCCGGGTGGCAGCGACAATGTGTAGTGCAAA		CGAAATTT	CACATT	VREAGHHGGESAST
				GACACCAGCCTTGAAGTCGGGTCGCCGTCGCGGAAAAGAGGTCAACTA		GCATAAGT	GCCGCA	VDVDVVESKKITGKT
				TGAAGGCCAAATTGTACGGGTAGAAAGACGGCGTGGATCTCGCAGTAG		TTTTGAAA	TGACTTC	ARRNRRAPSGDGKR
				TACGAGCGCAACGGACCTTGGTACGAGGATGGTTACGCGTGGGCGAAA		AAATAGGC	GATTTC	RESCGAECGQAVCG
				GAAGCTGATGGAGGCGTCGGTTCGTGAAGCGGGTCACCACGGAGGTG		ATTTGGTC	GCTTTTG	NAVADRSEASSPRTP
				AATCGGCGAGTACTGTGGACGTCGACGTGGTGGAATCGAAGAAAATTA		GAGTGCTC	CTCTTTT	NVSKSGRDKCGQPTI
				CGGGTAAAACGGCGCGCCGTAATAGAAGAGCGCCAAGCGGTGATGGA		GCACCACC	TGTATTT	KASTPSPPKRKPTTSS
				AAAAGACGAGAGAGCTGTGGTGCGGAATGTGGCCAAGCCGTATGCGG		ATTTGTCG	AATTTTG	SPRTPCLSKRGARSKI
				TAACGCTGTTGCGGACCGTAGTGAGGCCAGTTCGCCCCGCACTCCCAAC		CGGGTCGT	CTATTAA	PSTPDTPSTSGGSGK
				GTTAGTAAATCTGGTCGTGACAAGTGTGGTCAACCCACTATAAAAGCCT		CCTGATAT	CAATTCA	QRVLVSPLLRTEKLPD
				CCACACCATCACCACCGAAAAGAAAACCGACCACTTCGAGTAGCCCACG		TGCACTAC	GTTTGTT	LEVLQRTEEQVTVRA
				TACGCCATGCCTCTCCAAGCGCGGAGCGAGATCCAAAATTCCGTCCACG		ATTCAGGA	AACTGTT	TFPIAQAVVCPLGCE
				CCAGATACACCATCGACTTCTGGCGGTTCAGGTAAGCAACGTGTTTTGG		ACGGCCTA	TTGTATT	KPYTAVRPDGQFAH
				TGTCGCCGTTGCTTAGGACGGAAAAATTGCCCGACCTCGAAGTTCTACA		TTCCCTTC	CATTTGA	QTLTRHFMRVHNCH
				AAGAACCGAGGAACAGGTGACTGTACGGGCAACTTTCCCAATCGCCCA		GGGGAATT	AGATCC	SVQWHYRCRNCNTD
				GGCGGTGGTCTGCCCACTGGGATGCGAGAAACCGTACACGGCCGTGAG		GTGTTTTA	AAATAA	FLPADHRYPLRVVNT
				ACCAGACGGTCAATTTGCGCACCAAACGCTGACTCGTCATTTCATGCGA		GGAATTGG	AAC	HVRSCVSRWEITRKL
				GTCCACAATTGTCACAGCGTCCAGTGGCATTACCGGTGTCGCAATTGTA		AATCGGTT	(SEQ ID	GESEDLHGVRCDLCD
				ATACGGATTTCTTGCCGGCAGATCACCGGTATCCCCTCCGAGTTGTGAA		TGGTTACG	NO:	YVGVSKRAVGLHRRR
				CACCCACGTTCGCAGCTGTGTTAGCCGGTGGGAGATAACCCGGAAGTT		ATCGGTCG	1318)	HANENIMQNTGTAA
				GGGAGAGAGCGAAGATTTGCATGGAGTTCGTTGCGATCTTTGCGATTAC		AGTGGTTC		QIEALSKQVGEIRVA
				GTAGGCGTGTCGAAAAGAGCCGTCGGATTGCATCGCCGCCGGCACGCC		GTGAGATC		GDYSQFKFGKRVRQY
				AACGAGAACATTATGCAAAATACGGGTACGGCCGCTCAAATCGAGGCC		GAGTGACA		VAPTQRRDGLDEEEV
				TTGAGCAAACAAGTCGGCGAAATACGTGTAGCAGGCGACTATAGTCAG		GCCGGGT		HEAEEEEVPAESRTIL

		TTTAAGTTTGGTAAACGGGTTCGGCAATACGTGGCACCAACGCAGAGG	GGCAGCG ACA (SEQ ID NO: 1195)		GEPSTATAIGAEEISA
		CGAGACGGGTTGGACGAAGAGGAGGTTCATGAAGCCGAAGAAGAGGA			TGPVRADTAAQQMI
		GGTGCCAGCCGAATCACGTACCATCCTAGGTGAACCCAGCACCGCAACT			CRIGQWCVWPQDY
		GCTATAGGAGCTGAGGAGATCTCTGCAACGGGTCCAGTTCGAGCAGAC			HSIPAPQCWTDTLM
		ACAGCTGCTCAGCAGATGATCTGCCGGATCGGGCAATGGTGTGTCTGG			DLMIEQIVLQRYPDG
		CCGCAGGACTACCATTCGATCCCGGCACCTCAGTGCTGGACGGACACGT			AGVSVMSCSAVSAAI
		TGATGGACCTCATGATTGAACAAATCGTTCTGCAGCGCTATCCCGATGG			HHEISAEFAAQVMSS
		TGCGGGTGTGAGCGTGATGAGCTGCTCAGCAGTAAGCGCAGCCATACA			HDASLYCIIPVNVRN
		CCACGAGATATCTGCAGAGTTTGCCGCTCAGGTCATGTCTTCACATGAT			HWQMIVLDVAERVV
		GCCAGTTTGTATTGCATCATTCCTGTGAATGTACGTAATCATTGGCAGAT			HYYCSLREHNTVVLSS
		GATTGTGCTGGATGTGGCTGAACGAGTAGTCCATTACTACTGTTCTCTCC			LLSLVELSGKHTGCTS
		GCGAGCATAACACTGTTGTTTTATCGTCTCTGTTATCGCTAGTGGAGCTG			WKIETHDGAPVQTN
		AGTGGTAAGCACACTGGCTGCACCTCGTGGAAGATCGAGACTCACGAT			AFDCGPFSCLFLKHLL
		GGCGCGCCGGTGCAAACGAATGCGTTCGATTGTGGTCCGTTTTCGTGCT			HGIDMNFGDRESAA
		TGTTTTTGAAGCACCTGCTGCACGGCATCGACATGAATTTCGGTGATCG			LRTDLKFMIDAVSTP
		GGAGTCGGCGGCTTTGCGCACCGACCTGAAGTTTATGATCGATGCGGTA			VVPATDKLKKKPDGS
		TCCACTCCCGTCGTTCCGGCGACGGATAAGCTAAAGAAAAAACCTGACG			ATQLTQFQQKFLSAS
		GCTCAGCGACTCAGTTGACCCAGTTCCAGCAGAAGTTCCTGAGCGCTTC			DNWQSPDVDLQAV
		TGACAACTGGCAATCTCCTGATGTCGACTTGCAAGCGGTGTATGACGAA			YDEVVESIVSGHNEP
		GTGGTGGAGAGTATTGTGTCCGGACACAATGAACCGAGCTCTCGTAAC			SSRNTSRLQKKSPGK
		ACCAGCCGTTTGCAAAAGAAAAGTCCGGGTAAAGGAGGTAAAGGACAG			GGKGQVRRRSAVTR
		GTTCGCCGACGCAGTGCGGTAACGCGTGACCCGGCATGGTTGAAATCT			DPAWLKSASAVQKA
		GCGTCAGCGGTTCAAAAGGCGTTCAATTCGGCACCGGCCCGTACTGTGA			FNSAPARTVNAILRR
		ATGCAATCCTCCGGCGTCCGAATGCCTGTCCTTCGTTTACAGCCACGCAA			PNACPSFTATQVAD
		GTGGCAGATCATTATTTCAATCTCAGACCGGCGGTTACCTCGCTTGCACC			HYFNLRPAVTSLAPE
		CGAGGTTATTGACATCTTACCTCCGCCCGCCACTGACCACTCGATGCTTG			VIDILPPPATDHSMLV
		TTGCGGAGTTGAGTCGGAGGTCTGGGAAAAGATGCAGAAAGCA			AELSESEVWEKMQK
		CCCAACTCTGCGCCTGGTGCGGACAGGATCACCATTCGTATGGTGCGGA			APNSAPGADRITIRM
		TGGCAGATCCAGGTGCGATGATTCTCACCAGGTTCTACCGTGCCTGTCT			VRMADPGAMILTRF
		ACTAAGAAAATGGGTGCCCCTGCAATGGAAGCAATCCGTGTGCAAGCTC			YRACLLRKWVPLQW
		CTTTATAAGGACGGGGATAAGGAGCGCCTCGCCAACTGGCGGCCTATC			KQSVCKLLYKDGDKE
		GCCCTGGAGCCCGTGTTACAGCGGGTGCTCAGTGCGGTAGTCGCCAGC			RLANWRPIALEPVLQ
		AGAGTGACGAACTGGGCGAGGGCAAACGGGTTGATTTCCTTGGAAGCC			RVLSAVVASRVTNW
		CAGAAAGGCTTCCAGCCAGCTGATGGTACCTCCGAGCACAATTTCGTCA			ARANGLISLEAQKGF
		TGGAGGTTGCAATTCAAGAGGCAAGACGCACTAATGCACAGCTTGCCAT			QPADGTSEHNFVME
		CAGCTGGTTAGACATCTCCAACGCCTTTGGCACAGTGTCGCACCAATTGT			VAIQEARRTNAQLAI
		TGTTTTCGTTGTTGGAGCGCTATGGGCTGGACCCGACTTTTACCAGTTTT			SWLDISNAFGTVSHQ
		ATCCAGAATTTGTATAAAGACGCGACCATTGTAGTCAAAGGCGCCAATG			LLFSLLERYGLDPTFTS

			GAACCCACGTTACCGCGCGCTGGTCAGTCGGAGTACGCCAAGGTGACC		FIQNLYKDATIVVKGA
			CTTGCAGTGGTATCTTGTTTTGTTTGTTTGTTGAGCCACTCTTGCGGTCA		NGTHVTARWSVGVR
			GTCTTACCATCTTTACCGTGTGAGGCGGAAACGACTGCGGTTAACGTGT		QGDPCSGILFCLFVEP
			TAGGACAGCCGATCACAGCGCTGGCTTACGCTGATGACATCGCGCTGTT		LLRSVLPSLPCEAETT
			CGCCCCATCCATCGGGGTTATGCAACAACAGCTGTGCAAAATCCAGGGC		AVNVLGQPITALAYA
			ATGGCCAGCGCCATGGGCTTTCGGTTCAACCCGAAGAAGTGCGCGTCCC		DDIALFAPSIGVMQQ
			TGTACTTGAATCGTGCAGTAGTGAATGCCGCAACATTCACGATAAGCGG		QLCKIQGMASAMGF
			GGAGGAAATACCTGCTCTCGTACACGGGGACACTTTTCGCTATCTTGGA		RFNPKKCASLYLNRA
			GTAGCCGCTGGGCTGGGTAAACCCCAGACTCCTTTCTCATTACTGCGTG		VVNAATFTISGEEIPA
			AGAACCTCCGAGAAGCGGAACTTATATTCCGGTCAAAGCTGGCTCCGTG		LVHGDTFRYLGVAAG
			GCAGAAGATGGATGCGTACCGGACTTACGTACTGCCACGTTTGACATTT		LGKPQTPFSLLRENLR
			CAGCTTATGATTGCTAAATTTAACAACATCAAGCAATCGGCCGGGCAGT		EAELIFRSKLAPWQK
			ATGACCGCGCAATATTGAGGCTCGTGAAGCGTTGCTTCCAGCTTCCGGT		MDAYRTYVLPRLTFQ
			GGAGACGTCCACCGACTTCATTAGAGCCCCCAGACAGTGTGGTGGTCTG		LMIAKFNNIKQSAGQ
			GGCGTGCCATCATTGCGTGAGTTGTACGCAACCGCAAAGGTTTCGCGG		YDRAILRLVKRCFQLP
			GCACTCAAGATGCTCTGGTCTCCCTGTCGGGTAGTCTCGAGCCTGGCAG		VETSTDFIRAPRQCG
			CCAGTCAGCTTCAGCGAGTTGCGTCCGCGTACTTTGCGAAGAGGTTGAG		GLGVPSLRELYATAK
			AGATGTTGAGGCGGCTGACCTCTCTACCTTCATGAACGCCGCGCGATCT		VSRALKMLWSPCRV
			ACGCCGCTAGATCGGTCAGGGTATCCGACTTGCTTGTGGATGGACGCAC		VSSLAASQLQRVASA
			GCAAGCAGATGTCGTACCTGACAAAGGTAGCCGGCGTTGACTGTTACTT		YFAKRLRDVEAADLS
			CCTCGTTGGCGAGGCTGGGACCTCTTTCTTCATCCGGAATGGCCTCGGT		TFMNAARSTPLDRSG
			CAGACGGTCTCAGTTCTAAGCCCCTTGAGGAAGAACAAAGTTATGAGCG		YPTCLWMDARKQM
			TTTTAGGGGGGGCGATCCAAACTCGCCATCTGGACGCCTGGCTCCAATG		SYLTKVAGVDCYFLV
			CAAGCGCCAAGGTAAGACTGCGAGCTGCATTGTGCTCGACCGGTCATCC		GEAGTSFFIRNGLGQ
			TCCCGTTTTATAACGACCGGACGGTACACCAGCTTTGCAGCTATGCGCTT		TVSVLSPLRKNKVMS
			TGCGTTGCCCGCTCGCCTGGATTTGTTACCTTGCCGGGCGAGATCCTCTA		VLGGAIQTRHLDAW
			TGCGTTCGTACCAGAACTGTCGGCGGTGCGGTTACGACCGAGAGACCCT		LQCKRQGKTASCIVL
			GCCGCACATCCTTCAGCACTGTCGGCAGTTCAGCGCCCCCGCTTACCAG		DRSSSRFITTGRYTSF
			GCCCGGCATGACGCAGTTCAGGGACGCCTGGAAACGGTCATGCGTAGG		AAMRFALPARLDLLP
			CGGTTCCCCAACCTCCGCGTAAACCGGGCTCTACCGGAGATTGGGTCGA		CRARSSMRSYQNCR
			ACAAGCGGCCTGATCTAGTGGTCGTTGACGAGGAAAAGCGGTTAGTGA		RCGYDRETLPHILQH
			TTCTGCTTGATATCGCAATCGTGTTTGAAAACACTGCCGCTGCCTTCGTC		CRQFSAPAYQARHD
			GATGCCCGGACCAGGAAGTGGGCGCACTATGAAAAGGAGATCCTCGCC		AVQGRLETVMRRRF
			TACCGGCTACGAGGATACTCCGTGACGTACGACGCCATCGTCGTCGGTG		PNLRVNRALPEIGSN
			CTCTGGGGACGTGGGACCCGAAGAACGATGCCATCCTGAAAAGGATCG		KRPDLVVVDEEKRLVI
			GTGTCGTCTCCCAAAGGTACCTACGGCTCATGAAGGTTCTCGTGGTGTC		LLDIAIVFENTAAAFV
			CGAGATGTTGGAACATTCTTCACGGATCTACAGAAAACACCTCGGCCTG		DARTRKWAHYEKEIL
			CGAGACCTTCTCCCGGACACCGGCACGAAACGGCGACCGGTTGGCACG		AYRLRGYSVTYDAIVV

			ACGGAGACAGACCCACCCGGTGGTGATCTCCGTCAGAAGAAGCGCAAT ACCATCAGCGCTAGGGCATCAGGTGGGAAATGCCTGGAACGGCGGTTC ACGTCACCCGTCGGGACGCCCTCGCAGCGTGGTGAGCTGCAGTGCCAA CCGTGCCCTGGCCCTAGAAGACCCGCCCTCGCTGGAATAGCGCCCAACC CGCCTAGTCTCCAGCCCAGAAAACCACCCCCGCGGCAACACCAGAAGCC GGTTACGAAGTCTACAGCGCATTGATCCGTCCCGAACCAACGGAACCAC ATTGCCGCATGACTTCGATTTCGCTTTTGCTCTTTTTGTATTTAATTTTGCT ATTAACAATTCAGTTTGTTAACTGTTTTGTATTCATTTGAAGATCCAAATA AAAC (SEQ ID NO: 1072)				GALGTWDPKNDAIL KRIGVVSQRYLRLMK VLVVSEMLEHSSRIYR KHLGLRDLLLPDTGTK RRPVGTTETDPPGGD LRQKKRNTISARASG GKCLERRFTSPVGTPS QRGELQCQPCPGPR RPALAGIAPNPPSLQ PRKPPPRQHQKPVTK STAH (SEQ ID NO: 1440)
NeSL	Utopia_2_CMy	.	Chelonia mydas	GGCAGAAACTGCACSTTCTAGAAGACTCACTGCCTATCCTGAGGAAGAC TACCGCTTTGGAGATGGATTCTACTGCCGCTGCTTCAAAGAAAATCTTC ATGCTGCTTCGGAGGCTCCAGGACAGASTGAGAAGATGATCGCTCCCTC GCCGATTCCACAGAAACCTTCAACTGCCGCTCGGACCGCTGCTGCTCCA CGGAGCGCCCATCGGGGAACAGCCTCTAAGAGACCCTCAGCGCTTCCA GGACATCGCAGATGAGCAGCATCGATTGGAAAAGCAGCGCCTCCCTGA AGAAAACCCGTGGAGATGCTGCTGCAGGGAWATCTTGCACCTCGAGGA CGGCCTCCCAGGACATCGCAGCCAGGACAAACATCATCTCGCCTGCTCT TCAGGASAAGGATGCCMSAAGAACTTMTCCCACCTCCTCMACTGCCCA GGATCCTSATGCTGGTCGTCGTCCTGCTGTGCTGGAAGMKACCCCAACT GGCGAGACCTCAGAAACCACCCAGMWGGACCKCAACATCTCACTAGAT GCCTGCCCAGCCGAACCTCTCCATGCWACTCTCCCAGAGCAACSAGAAC CATCCAGMGAATCCGCTGATATGACTGAAGCCMATCCAACAGAGGGA GAAGGAAAGGAGAATGACTGCATCTATCTCCAGTATCCCCTCCCTACGG ACACGCTCCTCTKCCCCTTCCGCTATCCGAGGGTTCCAGTACATTGGCAG TCTCAGCAMACACCTCAAGAGAATCCATASCAAGCGGATCACCTTCCGG TGTGCCCTCTSCGACCTGCCTTTCGAGACGCAGATGAAATGTAAGTCTCA TCAAGTCACCTGCAAAGGACATCTCGAACTGGAAGAGTCCAACTTTACC AGTCTATGTTGCCGCCACACCCCATCTCTGCTCCGAAAGCAGAAACACCACA TGCTCCAGCTCCGGCTTCCGACTCCCCAAACACTCCGGCTACTGCATCCCC AGCCAACTCCCCCAGAGCCACAGCACGAAGAGGTGGTCCAACATCTATG AGGAAAGGGCCCGGGCGCCCAAGGATACTACCATTGAAAGAGTAAGTG CTGCCTCAAAAATCCCTAAGCTGGACCCTGCCAAGAGGAGGATCGGCG CTCCCTTGCAGCTTATGCAAGGCAACTCCATCTCCAGACAGCTTAGTGCT TCCTCGCAGTATGTCCAGCACAACGCCGTCGTCAGAAGGGTCAGCGCCC CTCCACACACGGGTAACTTCACGGGATCCTGTCGTCGGAAGCTAGCACT	GGCAGAA ACTGCACS TTCTAGAA GACTCACT GCCTATCC TGAGGAA GACTACCG CTTTGGAG ATGGATTC TACTGCCG CTGCTTCA AAAGAAA ATCTTCAT GCTGCTTC GGAGGCTC CAGGACA GASTGAGA AGATGATC GCTCCCTC GCCGATTC CACAGAAA CCTTCAAC TGCCGCTC GGACCGCT GCTGCTCC ACGGAGC	TGAGCT GGAGTG CCGATG AGAAGC GCAGTC GGGAAA GTAACT GAAATA CTTTCCT CATGGA TTGTATT TTCTAAA TGGACA ACCTACC TAATTCT CAATTAC TGAGGG ACAATCT CCACTCA TTGATAT ATTTTGC TTTCCAC AACCAA ATCTCTG TACAACT TTTCATG	MLQLRLPTPQTLRLL HPSQLPQSHSTKRW SNIYEERARAPKDTTI ERVSAASKIPKLDPAK RRIGAPLQLMQGNSI SRQLSASSQYVQHN AVVRRVSAPPHTGNF TGSCRRKLALPHRPP SETQPAESYLHCWTT QRDPTLADQHGLQD HPTGLPPEGSHCIKD PRQNSSTEGQRRSG NSQARRIPKRASTAA SKTVLPKRTSAALKSV REDTALVLEDPAKWS SQHREGXRQQANPT AVFQPEPAEIEQQPX VRAATPWQAAWME ELARTASFXDFDLLVD RLTKDLSAEIVSGRKG TQENTPTAHRQNQN NMREARRRNISRCYD PAAASRIQKLYRSNR PKAMREILDGPSSYC AIPSERLFLYFKGVFD RVAQNDMQRPECLX

	ACCTCACAGGCCACCCAGCGAGACCCAGCCAGCGGAGAGCTATCTCCAC	GCCCATCG	AGTGAT		PXPRVDYAEDXEQDF
	TGCTGGACCACGCAACGGGACCCCACCCTAGCGGACCAGCACGGCCTC	GGGAACA	GTACCC		TSWEVEARLTKTKNT
	CAAGACCACCCAACGGGATTGCCCCCAGAAGGCTCACACTGCATCAAAG	GCCTCTAA	GAGTAC		APGKDGIRYNFLKKR
	ACCCCAGACAGAACTCATCCACAGAAGGGCAGCGTCGCTCCGGAAACTC	GAGACCCT	TTGGATT		DPGCLVLTAIFNKCK
	CCAAGCAAGACGCATCCCTAAAAGGGCCAGTACCGCTGCCTCTAAAACT	CAGCGCTT	CTAATAT		QFRRTPSSWKKSMM
	GTCCTTCCTAAAAGGACCAGTGCTGCCCTAAAATCCGTAAGGGAAGACA	CCAGGACA	CTAAACT		VLVYKKGKQDNPNT
	CTGCCCTCGTTCTGGAAGATCCGGCCAAATGGAGTTCTCAACACCGGGA	TCGCAGAT	GTATTGT		RRPISLCSTMYKLYAS
	GGGAAAWAGGCAACAGGCTAATCCCACAGCAGTTTTCCAGCCTGAGCC	GAGCAGC	TAAATCT		CLAARITDWSVNGG
	TGCGGAGATTGAGCAGCAGCCCWCGGTGAGGGCCGCCACACCATGGC	ATCGATTG	ATTCACC		AISSIQKGFMSCKGC
	AAGCCGCTTGGATGGAGGAGTTGGCGAGGACAGCCTCCTTCGAKGACT	GAAAAGC	TAAATTT		YEHNFVLQTAIHMAR
	TTGACCTCCTTGTAGACAGACTCACTAAAGACCTGTCTGCAGAAATTGTA	AGCGCCTC	GGGTTA		RAWRQCAIAWLDLA
	TCTGGGAGGAAGGGAACTCAGGAGAACACACCAACTGCKCACAGACAG	CCTGAAGA	TTGCTG		NAFGSMPHQHIFDM
	AACCAAAACAACATGAGGGAAGCCAGGAGAAGGAACATCAGTCGCTGC	AAACCCGT	ATTATGT		LREFGMPENFLQLVR
	TATGACCCAGCAGCAGCATCCAGGATCCAGAAGCTCTACCGGTCAAACC	GGAGATG	ACTCTAT		ELYEGCTTTICSMEGE
	GCCCTAAGGCTATGAGAGAGATCCTCGATGGGCCCTCATCCTACTGCGC	CTGCTGCA	GTATCAT		TPEIPIRSGVKQGCPL
	AATCCCGTCCGAGAGACTATTCTTGTACTTCAAGGGGGTGTTTGACCGT	GGGAWAT	ATGACTT		SPIVFNLAMEPLIRAIS
	GTAGCTCAGAACGACATGCAGCGTCCAGAGTGCCTTCAKCCTTWACCCC	CTTGCACC	TTAAAA		SGLGGFDLYDNRVNI
	GGGTTGACTATGCAGAGGACCWGGAACAAGACTTTACATCATGGGAA	TCGAGGAC	ACAAAC		LAYADDLVLIADNPES
	GTAGAGGCCAGACTTACAAAGACCAAAAACACAGCCCCAGGAAAAGAT	GGCCTCCC	TTTGTAT		LQQMLDITSQAANW
	GGCATTCGCTACAACTTCCTGAAAAAACGAGACCCCGGTTGCCTGGTAC	AGGACATC	TTGTGG		MGLRFNARKCASLHI
	TAACTGCCATTTTCAACAAATGCAAACAGTTCCGCCGTACTCCCAGCTCC	GCAGCCAG	ATAATCT		DGSRRDSVQATSFQI
	TGGAAGAAATCTATGATGGTGCTCGTCTACAAGAAAGGCAAGCAAGAC	GACAAACA	AAGCAC		QGEPMIFLEDGQAY
	AACCCCAACACTCGGAGACCCATCTCTCTCTGTTCCACCATGTACAAACT	TCATCTCG	TATACCC		QHLGTPTGFRVQQT
	GTATGCCAGCTGCCTTGCGGCTAGGATCACAGACTGGTCGGTGAACGG	CCTGCTCT	AGATGT		PEDTIAEILRDVARIDS
	AGGAGCCATCAGCTCCATCCAGAAAGGCTTCATGTCATGCAAAGGCTGC	TCAGGASA	ACAGAC		SLLAPWQKINALNTF
	TACGAACACAACTTTGTCCTTCAGACTGCCATCCACATGGCCAGGAGGG	AGGATGCC	ACTCTTT		LIPRISFVLRGSAMVK
	CATGGAGGCAATGTGCCATAGCATGGCTCGACCTGGCTAATGCTTTTGG	MSAAGAA	TCCCAAC		VPLNKADNTIRQLVK
	ATCGATGCCCCACCAGCACACATTTTTGACATGCTGCGAGAGTTCGGGATG	CTTMTCCC	CTATGTA		KWMFLPQRASNELV
	CCTGAAAACTTTCTCCAACTGGTCCGGGAACTGTATGAAGGCTGCACCA	ACCTCCTC	TTATATT		YISHRQGGANVPRM
	CCACCATCTGCTCCATGGAAGGAGAGACACCCGAAATTCCTATCCGTAG	MACTGCC	TTTTTAA		GDLCDVAVITHAFRL
	CGGCGTGAAGCAAGGCTGCCCCCTGAGCCCCATCGTTTTCAACTTAGCC	CAGGATCC	CATTAG		LTCPDAMVRNIAESA
	ATGGAGCCGCTCATTCGAGCGATCTCCAGCGGTCTAGGCGGTTTCGATC	TSATGCTG	CTTTAAT		LQDAVKKRIARTPSN
	TGTATGACAACAGAGTGAATATCCTGGCCTACGCAGACGATCTGGTCCT	GTCGTCGT	AAAATTT		QDVATYLSGSLEGEF
	GATCGCAGACAACCCCGAGAGTCTCCAACAAATGCTTGACATCACCAGC	CCTGCTGT	TTAAA		GRDGGDFASLWTRA
	CAGGCTGCCAACTGGATGGGACTCCGCTTCAATGCCAGGAAGTGTGCAT	GCTGGAA	(SEQ ID		RNATRRLEKRIGCHW
	CCCTGCATATCGATGGCAGTAGAAGGGATTCGGTCCAGGCGACGTCTTT	GMKACCC	NO:		TWCEERQELGVLVP
	TCAGATCCAGGGTGAACCCATGATCTTCCTCGAGGACGGGCAAGCGTAC	CAACTGGC	1319)		QVKNTDHTIITPRART
	CAACATCTCGGCACGCCGACGGGTTTCCGCGTCCAGCAGACACCTGAGG	GAGACCTC			MLERTLKDAIRCQYV

		ATACCATCGCGGAGATCCTACGAGATGTGGCCAGGATCGACTCCTCCCT	AGAAACCA	ENLKRKPDQGKAFEV
		ACTGGCACCATGGCAGAAGATCAACGCCTTGAACACCTTCCTGATCCCC	CCCAGMW	TCKWDASNHFLPGG
		CGTATCTCATTCGTCCTGAGGGGATCTGCCATGGTGAAGGTACCTCTGA	GGACCKCA	SFTRFADWRFIHRAR
		ACAAGGCAGACAACACCATCCGGCAGCTGGTGAAGAAGTGGATGTTTC	ACATCTCA	LNCVPLNGAVRHGN
		TTCCCCAGAGGGCCAGCAATGAACTGGTGTACATCTCGCACAGGCAGG	CTAGATGC	RDKRCRKCGYANETL
		GCGGTGCCAACGTCCCTCGAATGGGTGATCTGTGCGACGTTGCCGTGAT	CTGCCCAG	PHVLCSCKPHSRAW
		CACTCATGCCTTCCGCCTTCTGACGTGCCCGGATGCCATGGTGAGGAAC	CCGAACCT	QLRHNAIQDRLARAI
		ATCGCGGAGAGTGCTCTGCAGGATGCCGTCAAGAAGCGAATCGCCAGG	CTCCATGC	PPPVGKVAVNSAIPG
		ACCCCCTCCAACCAAGATGTTGCCACCTACCTGAGCGGCTCGCTGGAAG	WACTCTCC	TDSQLRPDIVITNEDR
		GCGAATTTGGAAGAGACGGGGGAGACTTTGCTTCGCTCTGGACTCGTG	CAGAGCAA	KKIIMVDVTVPFENR
		CCCGCAATGCTACGCGACGACTGGAGAAGCGTATCGGCTGCCACTGGA	CSAGAACC	TPAFHDARARKVEKY
		CGTGGTGCGAGGAACGCCAGGAGCTGGGAGTCCTGGTGCCACAGGTG	ATCCAGM	APLAETLRAKGYQVQ
		AAGAACACAGATCACACTATCATCACTCCGAGAGCTAGAACCATGCTGG	GAATCCGC	THALIVGALGAWDPS
		AGAGGACCCTGAAGGATGCCATCCGCTGCCAATACGTGGAAAACCTGA	TGATATGA	NERVLRECGIGQRYA
		AACGGAAGCCGGACCAGGGCAAGGCATTCGAGGTGACGTGCAAGTGG	CTGAAGCC	RLMRQLMVSDAIRW
		GATGCCAGCAACCACTTCCTCCCCGGGGGCAGCTTCACCCGATTTGCCG	MATCCAA	SRDIYIEHITGHRQYQ
		ACTGGAGGTTCATCCACAGGGCCCGGCTCAACTGCGTCCCGCTGAACGG	CAGAGGG	EG (SEQ ID NO:
		AGCCGTCCGCCACGGGAATCGGGACAAGCGATGCAGGAAGTGCGGCT	AGAAGGA	1441)
		ATGCCAACGAGACGCTACCCCACGTCCTGTGTAGCTGCAAGCCCCATTC	AAGGAGA	
		CAGAGCCTGGCAGCTGCGACACAACGCCATCCAAGATCGCCTGGCCAG	ATGACTGC	
		AGCCATCCCGCCACCCGTGGGGAAGGTTGCCGTAAACTCCGCCATCCCC	ATCTATCT	
		GGAACCGACAGCCAACTGCGACCGGACATCGTCATCACCAACGAGGAC	CCAGTATC	
		CGGAAGAAGATCATCATGGTGGACGTCACAGTGCCCTTCGAGAACAGG	CCCTCCCT	
		ACCCCAGCCTTCCACGATGCCCGAGCTCGAAAGGTGGAGAAATATGCCC	ACGGACAC	
		CTCTGGCCGAAACCTTGAGAGCTAAGGGTTACCAGGTCCAGACACACGC	GCTCCTCT	
		GCTGATCGTCGGAGCCTTGGGCGCATGGGACCCCAGTAACGAGCGAGT	KCCCCTTC	
		GCTGAGAGAATGCGGAATCGGTCAACGCTACGCTCGGCTGATGCCGGCA	CGCTATCC	
		ACTCATGGTGTCAGATGCCATCAGGTGGTCGAGGGACATCTACATAGAA	GAGGGTTC	
		CACATCACCGGACATCGGCAATACCAGGAGGGATGAGCTGGAGTGCCG	CAGTACAT	
		ATGAGAAGCGCAGTCGGGAAAGTAACTGAAATACTTTCCTCATGGATTG	TGGCAGTC	
		TATTTTCTAAATGGACAACCTACCTAATTCTCAATTACTGAGGGACAATC	TCAGCAM	
		TCCACTCATTGATATATATTTTGCTTTCCACAACCAAATCTCTGTACAACTTT	ACACCTCA	
		TCATGAGTGATGTACCCGAGTACTTGGATTCTAATATCTAAACTGTATTG	AGAGAATC	
		TTAAATCTATTCACCTAAATTTGGGTTATTGCTGATTATGTACTCTATGTA	CATASCAA	
		TCATATGACTTTTAAAAACAAACTTTGTATTTGTGGATAAATCTAAGCACT	GCGGATCA	
		ATACCCAGATGTACAGACACTCTTTTCCCAACCTATGTATTATATTTTTTT	CCTTCCGG	
		AACATTAGCTTTAATAAAATTTTTAAA (SEQ ID NO: 1073)	TGTGCCCT	
			CTSCGACC	

				TGCCTTTC GAGACGC AGATGAAA TGTAAGTC TCATCAAG TCACCTGC AAAGGAC ATCTCGAA CTGGAAGA GTCCAACT TTACCAGT CTATGTTG CCGCCACC CCATCTCT GCTCCGAA AGCAGAA ACACCAC (SEQ ID NO: 1196)		
NeSL	Utopia-2_PCa	.	Phytophthora capsici	ACCGCCCAAGTCTCCACCGCAGCTGCGACTGCTGCTGCTCATCGAGCCG CAGTAGTCGCAGCTGTACAAGCACCACCTCCTACCGGACGCGCCGTCGT GGAGCACCACGCGCGCTGAGCCGCNTCAAGACCAAGAATTCCAGGC TCGCGCGCGCGTGGCGCTTCAAGATGGCGACGCGACCAGCAGGAGTGG CCGCTCTACTGACGGCGGAGACGAGGCCAAGACCGAGGACAGCGACGC GGACGCAGCCAAGCGCCAAGCCGCTCAAGCAGCACCGAGACCCCATAC CCMGGACACCGCGCCGTACGACGGCCACAGGGACTCGGTGCTGGCTGT GTACGCACACAACGCACCTGCATTCACCTGCGCGTTGTGTGTGTACACA GCACGCAACTTCGCCGAGCTGACCAAGCATCGCCATGCAGCGCATCGTC ACACCCGCTTCGTGGATCACTTCCACAGCGGGTGCACGTGTGGCATCGG CTTCCAGTCGCGCGCGGCAGCTACGCGACACGCTCAAGCCTGTGCAGAC AGCACACACGCCACCGTAGCTGCCTCGCGCGACCCGGCMNCCASTGCSS CCSGNGCCGGMAGCGASGAGGAGGACCNCGCACCCCCCGGTCCTCTSS TCGCGGCAGCATCTGCAGCTGCAGCAGCCGCATCAAGCGCCACTCCAGC TGCAGATACTGCCACCACACGCAGAGCGCCGTGCCCATCGCTACTCCTGGG CCCCAGTACGCCCCCCCACGTGCTGGAGCCACCTCCAGAGCTCCGAGTTT CCGGCAAACGCCGMCGCCTCAACACGCCGATCGACCTGCAGCCGCTGG ACGTGGACGCGCTGATGCAGGACATGGAGGAGGAGCTGCTCCTGGAC GTGGAGATGGAGACGGAGACGACGGAACCGCAGACCTCGACGGCCAM	ACCGCCCA AGTCTCCA CCGCAGCT GCGACTGC TGCTGCTC ATCGAGCC GCAGTAGT CGCAGCTG TACAAGCA CCACCTCC TACCGGAC GCGCCGTC GTGGAGC ACCACGCG CGCGCTGA GCCGCNTC AAGACCAA GAATTCCA GGCTCGCG	TAAGCT GGTCAT TTGACC GACAGG GCACTA CCCAGG TAGGGA ACCGCC CTTCAG AAACCC AGGAAG ACACAA ACACCCT CCCTTTA GTGACA TACATAT TTTAGG CTACATT TCAGTTA	MQDMEEELLLDVEM ETETTEPQTSTAXDA TTTTDRPTRWGPHP RAVAAAAIAQLVTGE XAXPALPSRQDRRPA PRSHAPTRSRWGPR HQAVGAAAIASLATG LPASAAPVSRATKHG EGRRRLQTRWGPRV SIPRAARRPGSRWGP PRAAGASGQLPASAS GATGQLPEHVEAITT TPRVASDADEGPTPP DPWILRFDGACRRN PGPGGAGAALFKPS GAVVWTCSHYMPSS NETNNTAEYTALLLG VQSAVHHGATRLDIE GDSSLVIAQVKGTFA

		GGACGCCACCACCACCACGGACAGGCCAACACGATGGGGCCCCCACCC	CGCGCGTG	CGGAGA	CRNAKLRQLRNRVR
		CAGAGCAGTGGCGGCTGCAGCCATMGCGCAGCTGGTGACCGGCGAGS	GCGCTTCA	GGTTAC	HALRSVEKYTLRHIDR
		CAGCAGMTCCAGCTCTGCCTTCGAGACAGGACCGCCGCCCAGCACCAC	AGATGGC	TAACTG	KANAHADRLANRAL
		GATCCCACGCACCCACCCGGTCTCGATGGGGACCGCGCCATCAAGCTGT	GACGCGAC	GTAAAT	DRRSSSSECEPHGSC
		CGGTGCCGCAGCCATCGCGTCGCTAGCTACAGGACTACCGGCCTCTGCG	CAGCAGG	AAAAAG	MERCCGTDTTPAVQ
		GCACCCGTGTCTCGCGCAACCAAGCACGGGGAAGGACGACGCCGTCTC	AGTGGCCG	CACTTT	GPTPQAAAAVPVQV
		CAGACACGATGGGGACCTCGGGTGAGCATCCCACGAGCCCGCACGCCGT	CTCTACTG	(SEQ ID	WPQWQRQTMVAW
		CCAGGATCGAGGTGGGGACCTCCACGCGCGGCAGGGGCCTCTGGCCAA	ACGGCGG	NO:	TTSHGGRCRDCSTRC
		CTTCCTGCCAGCGCCTCCGGGGCCACTGGCCAACTTCCGGAGCATGTGG	AGACGAG	1320)	GRSLPSLAHRPRLSPR
		AGGCCATCACCACGACACCACGTGTGGCATCGGACGCTGATGAAGGAC	GCCAAGAC		RQPRLRLRQLSDEER
		CGACACCGCCGGATCCGTGGATACTCCGCTTCGACGGCGCATGCAGGC	CGAGGAC		DXAADALQELSDVM
		GTAACCCCGGACCGGGCGGTGCAGGTGCTGCATTGTTTAAGCCGAGTG	AGCGACGC		ASKIVDADSWDTGE
		GCGCCGTGGTGTGGACGTGCTCCCACTACATGCCAAGCAGCAACGAAA	GGACGCA		GYISSIPERIREVLQPY
		CCAACAACACCGCCGAGTACACGGCACTGCTGCTGGGCGTACAGAGCG	GCCAAGCG		TTRPPRPGHQQQQR
		CGGTCCACCACGGTGCCACTCGCCTAGACATCGAGGGCGACAGCAGCC	CCAAGCCG		RRPPRVTRNQREHRL
		TGGTGATTGCCCAGGTCAAGGGCACATTCGCGTGCAGGAACGCGAAAC	CTCAAGCA		DEALDDMQATQQA
		TCCGGCAGCTCCGCAACCGCGTCCGGCATGCGCTCCGCTCCGTGGAGAA	GCACCGAG		APRDQRAIHRARRRV
		GTACACCCTCCGCCACATCGACCGCAAGGCCAACGCCCACGCAGACCGT	ACCCCATA		GRVRASMAKQELRQ
		CTGGCCAACCGCGCGCTGGACCGGCGGTCGAGCAGCAGCGAGTGCGA	CCCMGGA		AFAKDESKCVSKILAG
		GCCACACGGCTCGTGCATGGAGCGATGCTGTGGCACCGACACGACACC	CACCGCGC		ASAETAAEEHVDECP
		AGCGGTGCAAGGACCGACACCGCAAGCAGCAGCAGCGTGCCAGTGCA	CGTACGAC		IDAATLHAHFTGTNA
		GGTGTGGCCCCAGTGGCAGCGGCAGACGATGGTAGCGTGGACGACGA	GGCCACAG		PRTDFDYDAACGQEF
		GCCATGGAGGTCGATGCCGAGATTGCAGCACGAGATGCGGGCGAAGC	GGACTCGG		RGALDSMQPPTVAT
		CTTCCCAGTCTTGCCCATCGGCCCAGGCTCAGCCCCAGGCGGCAACCTC	TGCTGGCT		DAFEEELTIDEVEDQL
		GGCTTCGACTGCGCCAACTCAGCGACGAGGAACGCGACRMAGCAGCG	GTGTACGC		TRAAKTSSPGHDGIG
		GACGCCCTCCAGGAGCTTAGCGACGTCATGGCCAGCAAGATCGTTGAC	ACACAACG		YDIYSRFAAQLVPLLH
		GCAGACAGTTGGGACACCGGCGAAGGCTACATCAGCTCCATACCGGAG	CACCTGCA		AAYQFCWLHRRVPA
		CGCATCCGAGAGGTGCTGCAGCCGTACACCACGCGACCACCACGGCCT	TTCACCTG		LWKLGIVRLIHKKGD
		GGGCACCAGCAGCAGCAACGGCGACGGCCACCACGTGTGACACGCAAC	CGCGTTGT		PMQPTNWRPICLQP
		CAGCGCGAACACCGGCTAGACGAAGCGCTGGACGACATGCAAGCCACG	GTGTGTAC		AIYKIYSGLLARRLSR
		CAGCAAGCAGCACCGCGGGACCAACGGGCGATACATCGAGCGCGGCGT	ACAGCACG		WMEQNQRLPMAQ
		CGCGTCGGGCGTGTCCGGGCATCGATGGCCAAGCAAGAGCTACGGCAG	CAACTTCG		KGFRAFNGCHEHNF
		GCGTTCGCCAAGGACGAGTCCAAGTGTGTCCAAGATATTGGCCGGT	CCGAGCTG		VATTLLDQTRRSHRR
		GCCTCGGCGGAGACAGCGGCGGAAGAGCACGTCGACGAGTGCCCGAT	ACCAAGCA		LYQVWYDLRNAFGS
		TGACGCCGCCACGCTGCATGCCCACTTCACAGGCACCAATGCGCCACGG	TCGCCATG		LPQQLMWSVLRHLG
		ACGGACTTCGACTACGACGCTGCATGCGGTCAAGAATTCCGAGGAGCG	CAGCGCAT		VDASFIARCKNIYQDS
		CTGGACAGCATGCAGCCGCCGACCGTTGCGACCGACGCGTTCGAGGAA	CGTCACAC		AFVVANAVDGATDP
		GAGTTGACGATCGACGAGGTGGAGGACCAGCTTACACGCGCGGCCAAG	CCGCTTCG		VRQEVGVYQGCPLSP

		ACCTCCAGCCCGGGACATGACGGCATCGGCTACGACATCTACAGTCGCT	TGGATCAC		LLFISALVPLIRRLEKL
		TCGCCGCGCAGCTCGTGCCACTGCTCCACGCCGCGTACCAGTTCTGCTG	TTCCACAG		DGVGVPLAEGVRPC
		GCTGCATCGGCGCGTGCCAGCCCTGTGGAAGTTGGGCATCGTCCGCCTC	CGGGTGCA		ATAYADDIKVFSDSA
		ATCCACAAGAAGGGCGACCCGATGCAGCCAACGAACTGGCGACCGATC	CGTGTGGC		AGIRKCHDAVTRFLE
		TGTCTGCAGCCCGCGATTTACAAGATCTACAGCGGACTGCTGGCGCGTC	ATCGGCTT		WTGLRANPGKCASL
		GGCTGTCGCGCTGGATGGAGCAGAACCAGCGACTCCCCATGGCACAGA	CCAGTCGC		AVTTNARGNPVRDD
		AGGGGTTCCGAGCGTTCAATGGTTGCCACGAACACAATTTCGTGGCAAC	GCGCGGC		GVHLELQGEVIAPLSL
		CACGCTCCTGGATCAGACGCGCCGGTCGCACCGCAGGCTGTACCAAGTC	AGCTACGC		HDSYRYLGVGDGFD
		TGGTACGACCTGCGGAACGCGTTCGGGTCACTGCCGCAGCAGCTCATGT	GACACGCT		HVRHRLQLEPKLQQI
		GGAGCGTGCTCCGCCACCTCGGCGTCGACGCAAGCTTCATCGCCCGGTG	CAAGCCTG		KREAVALMQSGLAG
		CAAGAACATCTACCAAGACTCGGCTTTCGTGGTGGCGAACGCGGTGGA	TGCAGACA		WQVVKALKTFVYPK
		CGGCGCGACGGATCCTGTCCGCCAGGAGGTGGGCGTGTACCAAGGTTG	GCACACAC		VEYALRHLRPLQSQL
		TCCCCTGAGCCCCCTCCTCTTCATCTCGGCGCTTGTGCCGCTCATTCGCCG	GCCACCGT		QGFDRAVVRGLRHLL
		CCTCGAGAAGCTCGACGGCGTAGGGGTGCCGTTGGCCGAGGGTGTACG	AGCTGCCT		RLPQSATTEFFYTPTS
		ACCGTGTGCCACCGCCTACGCCGACGACATCAAGGTGTTCAGTGACAGC	CGCGCGAC		GGGLGLQSLVEMHQ
		GCCGCTGGCATCCGTAAATGCCACGACGCTGTCACTCGTTTCCTCGAGT	CCGGCMN		ALQVAHAWQMLHS
		GGACAGGCCTGCGGGCAAACCCCGGCAAGTGTGCCAGCCTAGCGGTGA	CCASTGCS		KDAAVVAVAKEQVC
		CGACAAACGCCCGCGGCAACCCGGTCCGCGACGACGGCGTGCACCTAG	SCCSGNGC		QVARKRYRLQEEHW
		AGCTGCAAGGCGAAGTCATCGCCCCCCTCAGTCTCCACGATAGCTATCG	CGGMAGC		RGRGDELVRLFLNSE
		CTACCTGGGTGTGGGCGATGGCTTCGACCATGTGCGCCACCGCCTCCAG	GASGAGG		LAASPFADCLRRNGD
		CTTGAACCCAAGCTCCAACAGATCAAGCGAGAGGCGGTGGCGCTGATG	AGGACCN		IGSLWTDVQRTLRLH
		CAGTCCGGACTGGCGGGGTGGCAGGTGGTGAAGGCGCTCAAGACCTTC	CGCACCCC		HLSLTAQDDRDGQD
		GTGTACCCCAAGGTCGAGTACGCGCTGCGCCACCTACGACCACTCCAGT	CCGGTCCT		PLALRVPHHTKWLD
		CCCAGCTCCAAGGCTTCGACCGCGCCGTCGTCCGTGGCCTGCGCCACCT	CTSSTCGC		HKTVLRHVKLHMKIR
		ACTGCGCCTCCCCCAGTCCGCCACCACCGAGTTCTTCTACACACCAACGT	GGCAGCAT		HQTRWKGLVDQGKT
		CTGGAGGTGGGCTGGGGCTGCAGTCCCTCGTGGAGATGCATCAGGCGC	CTGCAGCT		VRVHGGLGAKFVST
		TGCAGGTGGCGCATGCCTGGCAGATGCTCCATTCCAAGGACGCGGCCG	GCAGCAGC		GAGLSDDAYRFGVK
		TCGTGGCCGTTGCGAAGGAGCAGGTCTGCCAGGTAGCACGCAAGCGCT	CGCATCAA		ARLNQVDTNAVLKR
		ACCGCCTGCAGGAAGAACATTGGCGCGGGCGTGGCGACGAGCTCGTGA	GCGCCACT		KRLRSSKTCRDPTCSS
		GGCTCTTCCTCAACTCCGAGCTCGCCGCCTCGCCCTTCGCTGACTGCCTC	CCAGCTGC		AETLAHALNHCASN
		CGCCGCAACGGCGACATCGGATCCCTCTGGACGGACGTGCAGCGCACG	AGATACTG		MDAIRQRHDDALEQ
		CTCCGCCTCCATCACCTCAGCCTGACGGCTCAAGACGACAGAGATGGCC	CCACCACG		IGSKIRGALERAKSTT
		AAGACCCGCTTGCCCTTCGCGTGCCGCACCACACCAAGTGGCTCGACCA	CAGAGCGC		ELRLNQTVPEYTGAA
		CAAGACTGTGCTGCGGCACGTGAAGTTGCACATGAAGATCCGCCACCA	CGTGCCCA		LRPDIVLRNVAAKKM
		GACCCGATGGAAGGGGCTAGTGGACCAGGGCAAGACGGTACGCGTGC	TCGCTACT		VIADLAVTFEDHAAG
		ACGGTGGCCTCGGTGCGAAGTTCGTGTCGACCGGGGGCGGGGCTGTCGG	CCTGGGCC		ARHSSLQLSHDHKTL
		ACGACGCATACCGGTTCGGCGTGAAGGCCCGCCTCAACCAGGTCGACA	CCAGTACG		KYQPIVAELRVQGW
		CGAACGCCGTGCTCAAGCGCAAGCGGCTGCGGTCCAGCAAGACCTGCA	CCCCCCAC		QVQTAAIVYGSLGSV

			GAGACCCGACCTGCTCCTCGGCCGAGACGCTGGCTCATGCTCTCAATCA	GTGCTGGA		QPSNFKTYTEKLKLHK
			CTGCGCGTCCAACATGGACGCCATCCGCCAGCGTCATGACGACGCCTTA	GCCACCTC		REARQLDLQLSSHCI
			GAGCAGATCGGGTCCAAGATCCGCGGCGCGCTTGAGCGAGCCAAGTCG	CAGAGCTC		QASHRIWGWHCRR
			ACCACAGAGCTGCGCCTGAACCAGACGGTGCCTGAGTATACAGGCGCA	CGAGTTTC		HREGQRSGNTSRAS
			GCTCTGCGACCCGACATCGTCCTGCGGAACGTGGCCGCCAAGAAGATG	CGGCAAAC		RGSGGTPRRTSQVRA
			GTGATCGCCGACTTGGCAGTCACGTTCGAGGACCACGCGGCCGGCGCT	GCCGMCG		RR (SEQ ID NO:
			CGTCACTCCTCGCTGCAGCTCAGCCACGATCACAAGACACTCAAGTACC	CCTCAACA		1442)
			AGCCAATCGTGGCAGAGCTGCGGGTGCAGGGGTGGCAGGTGCAGACC	CGCCGATC		
			GCGGCGATTGTGTACGGGTCGCTGGGCTCGGTGCAGCCCAGCAACTTC	GACCTGCA		
			AAGACGTACACCGAGAAGTTGAAGTTGCACAAGCGCGAGGCTCGACAG	GCCGCTGG		
			CTGGATCTTCAACTGTCGAGCCACTGCATTCAAGCCAGCCACCGCATCTG	ACGTGGAC		
			GGGCTGGCATTGCCGGCGCCATCGCGAAGGACAACGAAGCGGCAATAC	GCGCTG		
			CTCGCGAGCATCGCGTGGGTCTGGGGGGACCCCGCGGCGCACATCGCA	(SEQ ID		
			GGTAAGGGCACGGCGATAAGCTGGTCATTTGACCGACAGGGCACTACC	NO: 1197)		
			CAGGTAGGGAACCGCCCTTCAGAAACCCAGGAAGACACAAACACCCTC			
			CCTTTAGTGACATACATATTTTAGGCTACATTTCAGTTACGGAGAGGTTA			
			CTAACTGGTAAATAAAAAGCACTTT (SEQ ID NO: 1074)			
NeSL	Utopia-2_PI	.	GCTCGGCCTCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTGG	GCTCGGCC	TAGGCG	MLADPAALAAGLAR
			CGCGGCCCACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTCC	TCGCGGCT	GAAACC	APPPPSAPQDPSPAF
			GCAGACGCAGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCCC	GCCTTCCC	AGGCCC	PAGPAGQNPRAAAP
			CCCTACGATGTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGAC	AGGCGCC	AAGACG	ARVEVHTVVAPPGR
			TCGGTATCGCCGGGTGCTACACTCTCAGCCGCCACAGCTCGGGCCTTGG	GCCGACTT	GCCGAC	AGGMLPDPGLVDSS
			CGGTCCGCCATTGGCCCTTGGAGCTCGACAGCGACAGCAGCGACGACG	CGCGCTCT	AGGGCC	PAAATAATPAPVAAT
			AGGACGCTCAAGACCCCCACGCCCGCCCCCAGGACCCCCCCGCAGACGT	GGCGCGG	CCACCC	ATTARAAARVAVEH
			CGCGAGTGTGCTTGCCCCACCCCGGCAAGGCAGGCAGCATGCTCGCCGA	CCCACACG	AGGTAG	HAHAEPNQEHLPMA
			CCCAGCAGCGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCGCCCCCATCG	CCGCCGCC	GGAACC	RVLVEPMQVDECSSC
			GCACCCCAGGACCCAAGCCCCGCCTTCCCTGCTGGCCCCGCCGGCCAGA	GAGCCTCC	GCCCTA	DRSTLTADDGSGDD
			ACCCTCGCGCCGCCGCCCCAGCCCCGTGTTGAAGTCCACACAGTCGTGGC	AAGCGCGC	GAAACC	VAAPSSMLSNDVAA
			ACCACCCGGCAGGGCGGGCGGCATGCTGCCCGACCCGGGGTTGGTTGA	CCGTTGGC	CATTTCG	PMDVDSGTSCPPTL
			CTCAAGCCCAGCAGCGGCTACCGCAGCTACTCCAGCTCCGGTAGCTGCT	TTCCGCAG	GTGGTC	QQPLQRPRALHVGS
			ACTGCTACAACTGCTCGCGCTGCTGCTCGCGTCGCCGTGGAGCACCACG	ACGCAGG	GACTCG	KRRRLDADDGEEAH
			CGCACGCTGAGCCGAACCAAGAACATCTACCGATGGCTCGCGTGCTCGT	GCTCGCGG	AAGGCC	QLQEEEEAGIHAPAL
			GGAGCCCATGCAGGTCGACGAGTGCAGCAGCTGCGACCGCAGCACTTT	CGACGGCC	TTACCTA	RLSAASAQPASVLAV
			GACAGCCGACGACGGGAGCGGTGACGACGTTGCAGCCCCAAGCTCCAT	CTCGCCAG	TTTTTTC	YTHNASRFDCTLCAY
			GCTCAGCAACGACGTCGCCGCCCCCCATGGACGTTGACTCCGGCACCAGT	CCCCCAAG	CTTAGA	TAGSFASLLTHRNSR
			TGCCCTCCAACGTTGCAGCAGCCGCTCCAGCGACCGCGTGCGCTGCACG	ACCCCCCC	CATTCAA	HRRTAFLDRFSAGCA
			TAGGCAGCAAGCGCCGCCGACTCGACGCAGATGACGGTGAAGAAGCTC	TACGATGT	TTAGGT	CGVPFASRLAAARHA
			ACCAGCTGCAAGAAGAGGAGGAGGCTGGCATTCATGCCCCAGCGCTCA	GGCACCAC	AGCGAC	QACASLSSAPSAEASS

	GGCTATCCGCGGCCAGCGCTCAACCGGCCTCCGTCCTCGCCGTGTACAC	CCGGCAG	CAAATT	AAGTSSPTADGADST
	GCACAACGCGTCGCGGTTCGACTGCACGCTGTGCGCGTACACGGCCGG	GGCGGCC	ACAAAT	VSAVAHAEPGLPHH
	CAGCTTCGCGTCACTTCTGACGCACAGGAACTCTCGGCACCGCCGCACC	GGCAGGCT	TTGGTA	NDTELTASPPLVSSSD
	GCATTCCTCGACCGATTCTCGGCGGGTTGCGCGTGTGGCGTGCCATTCG	GCCCGACT	ACGAGT	VEVQATKTEATDNR
	CCTCGAGGCTGGCTGCAGCAAGACACGCGCAAGCGTGCGCCAGCCTCT	CGGTATCG	AAGCCA	WGAPLPRVLVASRIA
	CGAGCGCACCATCGGCCGAGGCTTCGTCGGCAGCAGGAACATCAAGCC	CCGGGTGC	AATGGT	GRLAQVPPPRWGPP
	CCACTGCCGACGGAGCCGACTCCACCGTCTCAGCCGTGGCTCACGCTGA	TACACTCT	AATACA	LPRTTIAARIATRLAA
	GCCCGGCCTGCCCCACCACAACGACACGGAGCTCACAGCGTCCCCCCCG	CAGCCGCC	CAAAAC	TPAPRWDPPLPRSLV
	CTGGTGAGCTCCTCCGATGTTGAAGTGCAGGCAACAAAGACCGAAGCC	ACAGCTCG	TTTTCTG	VSRIAARLLPALPDAP
	ACCGACAACCGCTGGGGCGCCCCCGCTTCCGCGCGTGCTGGTCGCGTCTC	GGCCTTGG	TTCTAAT	ACEEEAKDSDTMDW
	GAATCGCCGGGCGCCTCGCTCAGGTGCCACCCCCACGCTGGGGCCCCCC	CGGTCCGC	CAGTGT	APTWTNEETKESEPH
	GCTGCCGCGCACCACCATCGCAGCCCGCATTGCGACGAGGCTCGCAGCC	CATTGGCC	GAAAAC	DEAPGQVDEETIDDA
	ACGCCTGCCCCCCGCTGGGACCCACCGCTGCCACGAAGCCTGGTTGTGT	CTTGGAGC	TGGTTTT	DGEWLLRFDGACRA
	CACGCATCGCCGCCCCGCCTGCTACCGCGCGCTACCAGACGCGCCTGCATG	TCGACAGC	CGCCTTT	NPGPGGAGAALFKP
	CGAGGAAGAGGCCAAGGACAGCGACACGATGGACTGGGCGCCCAACTT	GACAGCA	TGGCGG	SGPVVWTCSHYDPS
	GGACAAACGAGGAGACCAAGGAAAGCGAGCCACATGACGAAGCGCCA	GCGACGAC	ACTTTTT	TTATNNTAEYTALLL
	GGGCAGGTCGACGAGGAGACCATCGACGACGCCGACGGCGAGTGGCT	GAGGACG	CACTCG	GARAAADHGVTKLRI
	TTTGCGCTTCGACGGTGCTTGCCGTGCAAACCCAGGCCCAGGCGGCGCC	CTCAAGAC	CATTTTT	EGDSTLVIQQVRGIF
	GGCGCAGCTCTCTTCAAGCCCAGCGGCCCCGTAGTGTGGACTTGTTCCC	CCCCACGC	GGGCAA	ATRSTRLRALRNKVKL
	ACTACGACCCGAGCACCACGGCAACGAACAACACAGCAGAGTACACTG	CGCCGCCC	TCGTCT	ELARVGSFSLHHIDR
	CGTTGCTGCTCGGGGCCAGGGCCGCCGCCGACCACGGCGTCACGAAGC	CAGGACCC	GCGGCT	QANGHADRLANAGL
	TGCGCATCGAGGGCGACAGCACGCTGGTCATCCAGCAAGTGCGGGGCA	CCCGCAGA	AGCTTG	DRRRTKLECSVHPDG
	TCTTCGCGACGAGAAGCACGCGGCTCCGAGCACTGCGCAACAAGGTCA	CGTCGCGA	CTAGCG	RGCTNTSVATAAPTA
	AGCTCGAGCTCGCGCGTGGGGAGCTTCTCGCTCCACCACATCGACCG	GTGTGCTT	GCGGAC	PAAPLPSARPPASTA
	GCAGGCGAACGGTCACGCCGACCGCCTTGCCAACGCCGGCCTCGACCG	GCCCCACC	GAGCGG	APSPDDDHSDQGDI
	ACGCCGCACCAAGCTCGAGTGCAGTGTGCACCCCGATGGCAGGGGCTG	CGGCAAG	TCTCCG	DDGEVYAAMCISPD
	CACCAACACCTCCGTGGCGACCGCAGCGCCTACTGCTCCTGCAGCGCCC	GCAGGCA	GGGGCG	AVPHRRPRLRLRRLT
	TTGCCCTCAGCCCGGCCTCCTGCATCTACGGCCGCTCCGTCTCCTGATGA	GC (SEQ ID	TTCACCT	DEESEEAGNVVERLA
	CGACCACAGCGACCAAGGCGACATCGACGACGGCGAGGTGTACGCCGCC	NO: 1198)	TTCCCCC	ASLAAKIADAPDWET
	AATGTGCATCAGCCCCGATGCTGTGCCGCACCGCCGACCGCGCCTCCGC		GCGAGG	AEGYITALPYALYDKL
	CTACGGCGCTTGACGGACGAAGAGTCCGAGGAGGCTGGCAACGTCGTG		CCAACT	QPYSQSQHQPPRQQ
	GAGAGGTTGGCAGCGTCGCTGGCGGCGAAGATTGCAGATGCGCCAGA		ACACCG	QQQQRQRPRQQQQ
	CTGGGAGACGGCGGAGGGCTACATCACGGCCCTACCGTACGCCCTCTA		ATCTTCT	TRQRRQRRCKRGGG
	CGACAAGCTGCAGCCGTACTCCCAGTCCCAGCACCAACCTCCACGTCAA		CTACACT	SQHRQRKTRRRRPPR
	CAGCAGCAGCAGCAGCGGCAACGCCCACGGCAGCAGCAGCAAACACG		TTTCTAA	VTRHHREHRIDEALD
	GCAGCGTCGACAGCGGAGATGCAAGCGCGGCGGTGGCAGCCAACACA		TTCGCCT	DLHALESRRPQDRTA
	GGCAGCGGAAGACGCGCCGCCGACGGCCCCCACGGGTGACACGCCACC		CCGTCTT	ISKARRVGRIRSALD
	ACCGCGAGCACCGGATCGATGAAGCCCTCGACGACCTCCACGCGCTCG		CGGTCT	QHQLRHRFDTDEKA

		AGAGCAGGCGCCCTCAGGACAGGACGGCCATCAGCAAGGCCAGGCGTC	TCGGCT			CVDGILAAARDKDRA
		GAGTCGGTCGCATCCGCTCCGCCCTTGACCAGCACCAATTGCGCCACCG	GTCGGA			ASVTTTAQTAAPPHS
		CTTCGACACCGACGAGAAGGCCTGCGTGGACGGCATCCTGGCCGCCGC	TTTTTTT			APASAPSSAVDDGIC
		GCGCGACAAGGACAGGGCGGCATCAGTGACGACAACAGCGCAAACAG	CTTTTTG			PIPGDLLHAFFTDVN
		CGGCTCCGCCTCATTCTGCACCAGCTTCAGCTCCTTCGTCGGCGGTGGAC	ACCAAT			TPRTEFDADSPIGARF
		GACGGAATCTGCCCGATTCCGGGAGACTTGCTCCATGCCTTCTTCACAG	CAGAGC			REALAQLPAAIAATEL
		ACGTGAACACGCCGCGCACGGAGTTCGACGCCGACTCGCCCATCGGAG	GCGCCA			LMEPPSPDEVEDQL
		CCCGCTTTCGGGAGGCGCTTGCTCAGCTGCCCGCTGCCATCGCTGCCAC	TGCGAC			QRVRGTSSPGLDGV
		GGAGCTACTCATGGAGCCGCCGTCGCCGGACGAGGTTGAGGACCAGTT	TCTTCTG			GYDVYKTFTQQLLPA
		GCAACGTGTACGGGGTACGTCCAGCCCCGGCTTGGACGGAGTGGGGTA	GCCAAT			LHAAFSRCWTDQRV
		CGACGTCTACAAGACCTTCACACAGCAACTCCTACCGGCCCTGCACGCC	CAGAGA			PQSWKLGVVRLLFKK
		GCTTTCTCCAGGTGTTGGACTGATCAGCGCGTCCCACAGAGCTGGAAGC	CCGGGC			GDRQDPANWRPICL
		TCGGAGTCGTGCGGCTCCTGTTCAAGAAAGGAGATCGGCAGGACCCCG	CCTGTCC			QQAVYKLYAGILAHR
		CCAACTGGAGGCCGATCTGCCTCCAGCAAGCCGTCTACAAGCTGTACGC	TCGGAC			FTRWLDANTRHADA
		CGGTATCCTGGCCCACCGGTTCACACGCTGGCTCGACGCCCAACACCCGG	AGCGAG			QKGFRAVNGCGEHN
		CACGCTGACGCACAGAAGGGATTCCGCGCGGTTAACGGCTGCGGGGAA	GCCTCC			FLAATLTDNARRRRR
		CACAATTTCCTCGCAGCCACGCTGACGGACAACGCCCGCCGCCGACGCC	ACGGCC			ELHVVWYDIKNAFGS
		GCGAGCTGCACGTGGTCTGGTACGACATCAAGAACGCGTTCGGCAGTG	AGCCAA			VPHELLWEVLRRMG
		TGCCCCACGAACTGCTGTGGGAGGTGCTGAGGCGCATGGGAGTGCCGG	TCAAGT			VPAQFIACCQGIYDA
		CGCAGTTCATCGCGTGCTGCCAAGGGATCTACGACGCGGCAGCTTTCAC	CTCGGC			AAFTVGNAADGTTA
		GGTCGGGAACGCAGACGGCACGACGGCCCCAATTCAGCTGCGCTT	AGCGAC			PIQLRLGVFQGCPLSP
		GGGCGTGTTTCAGGGATGTCCCTTGAGCCCCCACCTCTTCACGGCGGTG	GCGTCT			HLFTAVISPLLHALKR
		ATCAGCCCACTCCTCCACGCACTCAAGCGGCTCCCTGGCACGGGGGTAC	TTCTATA			LPGTGVQLSAVDRPG
		AGCTCTCAGCCGTCGACCGGCCAGGTGCGTCAGCATATGCCGACGACCT	GCGCAG			ASAYADDLKVFSDTK
		CAAGGTCTTCAGTGACACCAAGGATGGCATCACGCGGCAGCACCAGCT	CTGACG			DGITRQHQLVTDFLR
		CGTCACCGACTTCCTGCGCTGGACAGGCATGGTGGCAAACCCGTCCAAG	AGGCCG			WTGMVANPSKCST
		TGCAGCACCATGTCTGTCCAGAGGGACAACCGCGGCGTTTTGAAGACC	ATCTGG			MSVQRDNRGVLKTA
		GCCAACCTCACGCTCCAGCTGGACGGCGCCCAAATCCCAGCACTCGGGA	CGGCCC			NLTLQLDGAQIPALG
		TGACCGAGGCGTACGCGTACCTGGGCATAGGCGACGGATTCGACCACG	CCGATT			MTEAYAYLGIGDGFD
		TACGCCGCCGCGTCGAGTTGGCCCCCAAGCTCCGCGAACTGAAGGCAG	GGTCCG			HVRRRVELAPKLREL
		ACACGACGGCGCTGATGCAGTCGGGGCTGGCGCCGTGGCAAGTTGTGA	ACTTTCG			KADTTALMQSGLAP
		AGGCCCTCAAGGTTTACATCTACCCACGCGTGGAATATGCCCTCCGCCA	GCCAAT			WQVVKALKVYIYPRV
		CCTCCGACCGTTTCAGCAGCAGCTCCAAGGCTTTGACCGACACCTCGCG	CAGCGA			EYALRHLRPFQQQLQ
		CGCGGGCTGCGTCACCTTCTCCGCCTGCCAACATCAGCCACCACCGAGT	CGACGA			GFDRHLARGLRHLLR
		TCCTGTATGCACCAACGTCGCGAGGAGGTTTGGGGCTGCTACCGCTTAC	GGGGGC			LPTSATTEFLYAPTSR
		AGAGGTTCACGGTGCCCTTCAGATTGCGCACGCCTGGCAGACCCTGCAC	AGGGGT			GGLGLLPLTEVHGAL
		TCGCCGGACCCCGCCATCCGCCGCATTGCCCGCGTGCAGCTTCGCCAAG	TTACACT			QIAHAWQTLHSPDP
		TCGCCGACGCCCGGCACCGCTTGGACGCCGAGCACTGGAAGGAGCGGG	TTTGCCC			AIRRIARVQLRQVAD

			GGGAGGAGCTGTGCGAGCGGCTGCTAAACTCGCAGCTCGGGACATCGG		CCGTTTC	ARHRLDAEHWKERG
			CCCACGCGCCTCCCAAACGACGCAATTGCGACATCGGCTCGCTGTGGGT		GACTTC	EELCERLLNSQLGTSA
			CGATGTCCAGCGCCACCTCCGCTCTCTCGGCCTACAGCTACAGACAGCC		AACTTCA	HAPPKRRNCDIGSLW
			CCCGCGGACACGCACACTGGCGCACCAGCTCAGCCTCTCCAGCTTCGCG		GGCCAA	VDVQRHLRSLGLQLQ
			TGCCGCACCACGACACAAGTGGCTGACGCACAAGGACGTGCTGCGGCACG		AATGGC	TAPADTHTGAPAQPL
			TGAAGCTGCACATCAAGAATAACCACTGGCACCGCTGGACATCCATGCG		GATTTC	QLRVPHHDKWLTHK
			GGACCAAGGGAAGACGGCCCGCGCCCACGGCGGTGAAGGCAGCGGCT		GACCCT	DVLRHVKLHIKNNH
			TCCTTACGCAGCCGCGAGGCATGTGGGAAGCCGACTACCGGTTCGCGG		CCACGC	WHRWTSMRDQGKT
			TGGCGGGTCGTCTTAACCAAGTGGACACGTACAGCGTACTGAAGCGCC		GCCGTG	ARAHGGEGSGFLTQ
			GGCGCCTGCGCTCGCACGACAGATGCCGCCAGCCAGGATGCCATCGTG		CCACTG	PRGMWEADYRFAV
			CCGAAACACTGGCTCACGTGCTAAACCACTGTCCCGGCACGATGGACGC		CTCGGC	AGRLNQVDTYSVLKR
			GGTCCGAGGGCGGCACGACGGAGCGCTCAAGCGCATCGAGCGTGAGC		ACCGGC	RRLRSHDRCRQPGC
			TCCACGCGTCGGCCACGGACCGCCGGGACCGCGTCGAGCTCCGAGTGA		GGCGAT	HRAETLAHVLNHCPG
			ACCAGACCGTGCCGTCACTCGCCGGCCCCGCGTTACGGCCCGATCTCCA		TCAGCG	TMDAVRGRHDGALK
			GCTCTACAACCACACCAAGAAGACGGTGGCGGTGGTCGACCTGGCCGT		GGTGCA	RIERELHASATDRRD
			GGCGTTTGAGGAGCAGGCGAGTGACGACGCGAGCAGCTCGGCGCTGT		ACTTCG	RVELRVNQTVPSLAG
			CCCTTATCGCCAGCCACAAGCGAGCCAAGTACGACCGCATCAAGCGACA		GGCACG	PALRPDLQLYNHTKK
			CCTCGAGCGCCAAGGATGGAAGGTACACCTCTCGGCGCTCGTGTACGG		TGTGCA	TVAVVDLAVAFEEQA
			GTCGCTTGGGGCGGTCGCTAGTGGCAACTACCAGGTGTACACCACACAC		ACACAT	SDDASSSALSLIASHK
			CTGGGGCTACTCAAGCGCGATGCAAAGCGGCTGGACCGGCAGCTGTCT		GCAGCG	RAKYDRIKRHLERQG
			GTCGAATGCATCCAGTCCAGCCGCCGCATCTGGAATCTACACTGCAGCC		CCCATTG	WKVHLSALVYGSLGA
			AGCACCGGACTCGCCAACACCAGGCGAGGCCCAGCCAAGGACCAAGAG		CACGCC	VASGNYQVYTTHLGL
			GCAGCCGGGCGACGGAGACCGGGGGGGACTCCGTCCCAGACAAGCCGC		AAGCGG	LKRDAKRLDRQLSVE
			CGCTAGGCGGAAACCAGGCCCAAGACGGCCGACAGGGCCCCACCCAGG		CATCGC	CIQSSRRIWNLHCSQ
			TAGGGAACCGCCCTAGAAACCCATTTCGGTGGTCGACTCGAAGGCCTTA		GGGACG	HRTRQHQARPSQGP
			CCTATTTTTCCTTAGACATTCAATTAGGTAGCGACCAAATTACAAATTTG		ACGCCT	RGSRATETGGTPSQT
			GTAACGAGTAAGCCAAATGGTAATACACAAAACTTTTCTGTTCTAATCAG		CGGCCG	SRR (SEQ ID NO:
			TGTGAAAACTGGTTTTCGCCTTTTGGCGGACTTTTTCACTCGCATTTTTG		CTCAAG	1443)
			GGCAATCGTCTGCGGCTAGCTTGCTAGCGGCGGACGAGCGGTCTCCGG		CGCAGC	
			GGGCGTTCACCTTTCCCCCGCGAGGCCAACTACACCGATCTTCTCTACAC		CCCGCC	
			TTTTCTAATTCGCCTCCGTCTTCGGTCTTCGGCTGTCGGATTTTTTTCTTTT		CTTCCAG	
			TGACCAATCAGAGCGCGCCATGCGACTCTTCTGGCCAATCAGAGACCGG		CACGAC	
			GCCCTGTCCTCGGACAGCGAGGCCTCCACGGCCAGCCAATCAAGTCTCG		CTCGCG	
			GCAGCGACGCGTCTTTCTATAGCGCAGCTGACGAGGCCGATCTGGCGG		CCGTTT	
			CCCCCGATTGGTCCGACTTTCGGCCAATCAGCGACGACGAGGGGGGCAG		GGCGGA	
			GGGTTTACACTTTTGCCCCCGTTTCGACTTCAACTTCAGGCCAAAATGGC		TCGCCAT	
			GATTTCGACCCTCCACGCGCCGTGCCACTGCTCGGCACCGGCGGCGATT		CAAGAC	
			CAGCGGGTGCAACTTCGGGCACGTGTGCAACACATGCAGCGCCCATTG		GTGCGA	

			CACGCCAAGCGGCATCGCGGGACGACGCCTCGGCCGCTCAAGCGCAGC CCCGCCCTTCCAGCACGACCTCGCGCCGTTTGGCGGATCGCCATCAAGA CGTGCGAGAGCCAGGCGGGGTCGGGCAAAATATACTTACTCTAAGTAT GCCCGAATCCCTGCCCTCTCAGGCTGAACGCGGCCCCATACTTGATCTAA GTATGGGAGGGATCCCTGGCCTCTCAGGCTGTACGCGAGACCCGTACG GCCGAATCCCCTGGCCTCTCAGCCTGTACGCGGGGC (SEQ ID NO: 1075)			GAGCCA GGCGGG GTCGGG CAAAAT ATACTTA CTCTAA GTATGC CCGAAT CCCTGCC CTCTCAG GCTGAA CGCGGC CCCATAC TTGATCT AAGTAT GGGAGG GATCCCT GGCCTC TCAGGC TGTACG CGAGAC CCGTAC GGCCGA ATCCCCT GGCCTC TCAGCCT GTACGC GGGGC (SEQ ID NO: 1321)	
NeSL	Utopia-2_PR	.	Phytophthora ramorum	TCAAGCCCCGCCGCCAAGCCCAGCTGCGGCTGTTGCCGCCCCTCCAGCA GCAGCAGTCGCAGCTGAACCTACAGCCCCTCCTGCTGATCGCGCCGCCG TGGAGCCCCCGCGCGCCGCCGAGCCGCCCCAAGAACAAGCACCCCCAG CTAGCGCGCGCGTGGAGCCCATGTCGGGCGACGTGGTCAGCAGCGACG GCAGCAGCAGGACGACGGACGCGAGCGGCGACGGCGACGACGGCGCA GGCAGCAGCGATGCAGCAGGCGACGTCGGCGTCGTCGCCATGGACGTT GACCAAGGTGCCCGCCGACAGCAGCCGCCATGGCAGCGCGTCGGCGGC	TCAAGCCC CGCCGCCA AGCCCAGC TGCGGCTG TTGCCGCC CCTCCAGC AGCAGCA	TGAGCA CCTTGG GTTGCT CAAGCG TGATGC GAAGCG GCTGGA	MSGDVVSSDGSSRT TDASGDGDDGAGSS DAAGDVGVVAMDV DQGARRQQPPWQR VGGKRRRLNDVDDE DTRELAELLLEEEDEA GDHAPAPRLSAASAR

		AAGCGTCGACGTCTCAACGACGTCGACGACGAAGACACCCGCGAGCTG	GTCGCAGC	CCGGCA	PASVLSVYAHNAQRF
		GCCGAGCTCCTGCTCGAGGAGGAGGACGAGGCTGGGGACCACGCTCCA	TGAACCTA	GCTCTC	QCTLCTYTAASFASLK
		GCGCCTAGGCTATCCGCGGCCAGCGCTCGCCCGGCCTCCGTCCTCTCCG	CAGCCCCT	GGTGGC	RHRDSRHRRTAFLDR
		TGTACGCGCACAATGCGCAGCGCTTCCAGTGCACGCTGTGCACGTACAC	CCTGCTGA	GTGCAT	FSAGCACGAPFASRL
		GGCCGCCAGCTTCGCGTCGCTCAAGCGACACCGGGACTCCCGGCACCG	TCGCGCCG	CCAGTC	AAANHAHACASLNR
		CCGCACCGCCTTCCTCGACCGGTTCTCGGCGGGATGTGCGTGCGGTGCG	CCGTGGAG	CAGCCG	TLSVAATPAAGELSPT
		CCCTTCGCGTCGAGGCTGGCTGCTGCGAACCACGCACACGCGTGCGCCA	CCCCGCGC	CCGCAT	AGAANATVKAATVT
		GCCTCAACCGCACCTTATCGGTAGCGGCTACGCCGGCAGCAGGGGAAC	GCGCGCCG	CTGGAA	PDSPRQDPPELAASP
		TGAGCCCCACTGCTGGTGCAGCCAACGCCACCGTCAAGGCGGCCACCGT	AGCCGCCC	CCTGCA	PLASSPDVAVQAAD
		CACGCCCGACTCCCCCCCGCCAAGACCCCCCGGAGCTCGCAGCGTCCCCC	CAAGAACA	CTGCAG	MQAPTRWDPPLPRT
		CCGCTAGCGAGCTCCCCCCGACGTTGCTGTGCAGGCCGCCGACATGCAG	AGCACCCC	CCAGCA	LVATRVASRLTDLTPP
		GCCCCGACTCGTTGGGACCCGCCCCTCCCCCCGCACGCTGGTGGCGACAC	CAGCTAGC	CCGCGC	RWGPPLPRATVVSRI
		GGGTCGCCTCCCCGCCTCACCGACCTGACACCGCCGCGGTGGGGACCGC	GCGCGCGT	GCGCCA	AARLEAAPTPRWGP
		CGCTGCCTCGAGCCACGGTGGTGTCTCGCATCGCGGCGAGACTCGAGG	GGAGCCC	GCACCA	PLPRVVVASRIAERLA
		CCGCACCAACGCCCCCGTTGGGGCCCTCCACTTCCACGCGTCGTGGTTGC	(SEQ ID	AGCACC	PPELAADDETKDGEE
		GTCCMGGATCGCCGAACGCTTAGCTCCGCCCGAGCTGGCAGCTGACGA	NO: 1199)	AGGGGG	DQSFTEPVAAARSXG
		CGAGACCAAGGACGGCGAGGAAGACCAGAGCTTTACGGAACCTGTGGC		AAGTCG	GEDANGEWLLRFDG
		TGCCGCCCGGTCMGAKGGCGGCGAGGACGCGAACGGCGAGTGGCTGC		GGCGGC	ACRANPGPGGAGAA
		TTCGCTTCGACGGCGCCTGCCGAGCCAACCCAGGCCCCGGCGGTGCGG		GGAGAC	LFKPSGPVVWTCSHY
		GCGCAGCGCTGTTCAAGCCCAGCGGCCCGGTGGTGTGGACCTGCTCCC		CGGGGG	MPSSSETNNTAEYTA
		ACTATATGCCGAGCAGCAGCGAGACCAACAACACGGCCGAGTACACGG		GACTCC	LLLGMRAAADHGAT
		CTCTGCTGCTCGGCATGCGGGCTGCGGCGGACCATGGCGCTACGCGCG		GCCGCG	RVHVEGDSTLVIQQV
		TGCACGTCGAAGGTGACAGCACCCTGGTGATCCAGCAAGTGCGGGGCA		CACCGG	RGIFATRSTRLRGLRK
		TCTTCGCTACCCGTAGCACCCGGCTCCGTGGACTGCGCAAGAGCGTGAA		CCGCCG	SVKAEMARMEHVTL
		GGCGGAGATGGCGCGGATGGAGCATGTCACGCTCCACCACATCGACCG		CTAGAC	HHIDRQANGHADRL
		CCAAGCCAACGGCCATGCCGACCGCCTCGCCAACGCCGCCCTCGACCGC		GGCACA	ANAALDRRKTKLECG
		CGCAAGACCAAGCTGGAGTGCGGACTCCACCCTGACGGCCAAGGCTGC		CAGGCC	LHPDGQGCSSTAATT
		TCCAGCACCGCCGCGACGGCGGTCCCCTCTGTGGTGCCGGACCGG		CACAGC	AVPSVVPDRPPSSTA
		CCTCCCTCGTCAACCGCTGCCGCGCCAACGCCAAGCGCCGAACCCGACG		GGCCGA	AAPTPSAEPDETEQG
		AGACCGAGCAGGGCGACATCGACGACGGCGAGGTCTACGCCGCCATGT		CAGGGC	DIDDGEVYAAMCIGP
		GCATTGGCCCGGACTCCATACCGGAGCGCCGACCACGCCTCCGCCTACG		CACACC	DSIPERRPRLRLRQLS
		GCAGCTGTCCGAGACCGAGGAGGAGGAGGCGGGTGCCATTGTGGAGA		CAGGTA	ETEEEEAGAIVERLAA
		GGCTGGCGGCAACGTTGGCGGGCAAGATCACGGACGCGAGTGACTGG		GGGAAC	TLAGKITDASDWATA
		GCGACGGCGGAGGGCTACATCACTGCCCTCCCCTACACTCTATACGACA		CGCCCTC	EGYITALPYTLYDKLQ
		AACTGCAGCCGTTCGCCCAGCACCGCCACCAGCCTCGCCCACAGCACCG		AAACCC	PFAQHRHQPRPQHR
		GCAGCAACCCCAGCGCGACCCACCCCTCGGCACCCACGACGGGGACCA		CGCCGG	QQPQRDPPLGTHDG
		CGGCCAGCCGAGCACCAGCCGCAGCCGTCGCCGTCGCCGCCACCACCGTGAG		TACATTA	DHGQPSTSRSRRRRR
		GGATCGCTTACGCCGACGCCCACCGCGTATTACCCGCCACCACCGTGAG		TGGTCC	RAKDRLRRRPPRITR

		CACCGCCTCGACGAGGCTCTGGACGACCTGCGGGCGGTGGAGCACGCC		GACACC	HHREHRLDEALDDLR
		AGTCCCCACGACCGGCCGGCCGTCGCTCGGGCCCGGCGGCGCGTTGGC		TATGAG	AVEHASPHDRPAVA
		CGCGTCAACTCCGCCATTGCCCAGCAACAGCTGCGACACAAGTTTGACA		GTGCAA	RARRRVGRVNSAIAQ
		AGGACGAGAAGGCGTGCGTGGATGGAATCTTGGCCGCGGCCCGAGCTT		CCTGTAC	QQLRHKFDKDEKAC
		CTCGCGGGTTAGCAACCCCATCGGCCTCAGCCTCGAGGCATCCACCGCC		ACAAGT	VDGILAAARASRGLA
		GGTACCCTCCACCGCTGCTGATGACGGCTCCTGCCCCATCCCCAGCGAT		TACACA	TPSASASRHPPPVPST
		GAGCTCCACGCCTTCTTCACGGCCGTGAACACGCCGGCGGGCACCTTCG		CCACAT	AADDGSCPIPSDELH
		AGCCCATGGCACCAGTGGGTGCCCCGTTCCGCTCCGCCGTCGCCCACCT		AGCGAC	AFFTAVNTPAGTFEP
		CCCCGCCGCGACATCGCAGCCTGAGTTGCTCTCGGACGCGCCGACCACG		TACCAG	MAPVGAPFRSAVAH
		GACGACATCGAGGACCAACTGCAGCGTGCCCGTGGCAGCTCCAGCCCT		GTATTTA	LPAATSQPELLSDAPT
		GGCCTGGACGGCGTGGGGTACGACATTTACAAGGCGTTCGCGGCCCAG		CTACCTG	TDDIEDQLQRARGSS
		CTCCTGCCGGCCCTGCACGCAGCATTCGCGTGCTGCTGGCGCCACAAGC		GAAGCC	SPGLDGVGYDIYKAF
		AAGTCCCGCAGAGCTGGAAGGTGGGGGTGGTGCCGCCTCCTGTTCAAGA		AAGGAT	AAQLLPALHAAFACC
		AGGGGGAGCGAACCGAGCCGGCGAACTGGCGCCCGATCTGCCTGCAG		TAACCG	WRHKQVPQSWKVG
		CAGGCAATCTACAAGCTGTACGCCGGGGTGCTAGCTCGCCGACTCACCC		GTCGGT	VVRLLFKKGERTEPA
		GCTGGCTGGACGCCAACGGCCGCCACGCAGACACGCAGAAGGGATTCC		AATACA	NWRPICLQQAIYKLY
		GGGCGATGAACGGCTGCGGGGGAACACAATTTCCTCGCAGCCACGCTGG		CATAACT	AGVLARRLTRWLDA
		TCGACCAGGCCCGCCGCAAGCGACGAGAGCTCCACGTCGTGTGGTACG		TT (SEQ	NGRHADTQKGFRA
		ACTTCGCCAACGCCTTCGGCAGCGTGCCACATGACCTGCTGTGGGAGGC		ID NO:	MNGCGEHNFLAATL
		GCTCGAGCGGCAGGGCGTCCCCAGCCCCTTCATTGCTTGCTGCCGGGGC		1322)	VDQARRKRRELHVV
		CTCTACGCGGATGCCGTCTTCACAGTCGGCAACGCAGCGGACGGCACCA			WYDFANAFGSVPHD
		CGGCACCGATCGCGCTACGGGTGGGGGTCTTCCAGGGATGTCCGCTGA			LLWEALERQGVPSPF
		GCCCCCACCTCTTCACGGCCGCGATCGCCCCGCTGCTCCACGCGCTCAA			IACCRGLYADAVFTV
		GCGCTTACCAGACACGGGCGTACAGCTGTCTCGTGTCGACTGCCTCGGT			GNAADGTTAPIALRV
		GCCTCCGCGTACGCGGACGACCTCAAGATATTCAGCGGCACCGAGGGC			GVFQGCPLSPHLFTA
		GGGACCAAGAGACAGCACGCGCTCGTGGCGGACTTCCTGCGCTGGACC			AIAPLLHALKRLPDTG
		GGCATGCCGCGCGAATCCGGCGAAGTGCTGCACCATGTCGGTCCAGCGC			VQLSRVDCLGASAYA
		GACACACGTGGAGTGCTCAAGGCCTGCAACCTCGGACTGCAGCTGGAT			DDLKIFSGTEGGTKR
		GGCGCCCCGATCCCCGCGCTGACCATGAGCGCCTCGTACGCCTACCTGG			QHALVADFLRWTG
		GGATCGGGGATGGCTTCGACCATGTGCGCCGCCGCATCGAGCTCGCGC			MRANPAKCCTMSV
		CCAAACTCCAGGAGCTCAAGCACGATGCGACGGCGCTGCTGCAGTCCG			QRDTRGVLKACNLGL
		GACTGGCCCCATGGCAGGTAGTGAAGGCGGTGAAGGTGTACCTCTACC			QLDGAPIPALTMSAS
		CTCGGGTGGAGTACGCCCTTCGCCACCTTCGCCCCTTCCACCAGCAGCTG			YAYLGIGDGFDHVRR
		GAGGGCTTCGACAGACACCTCGTGCGCGGCCTCCGCCACCTCCTCCGCC			RIELAPKLQELKHDAT
		TGCCCGCCAACGCCACGACGGCATTCTTCTACGCGCCGGTGTCGCGCGG			ALLQSGLAPWQVVK
		CGGGCTGGGCCTGCTGCCCCTGACGGAGCTGCACGCGGCGGCTACAAGT			AVKVYLYPRVEYALR
		CGCGCACGGATGGCAGATGCTGAACTCGAAGGACCCTGCCATCCGACG			HLRPFHQQLEGFDR
		CATCGCGCGCCGTGCAGCTCCGGCAAATCGCCGACGCGCGCCACCGGAT			HLVRGLRHLLRLPAN

			CGACGCTCAAGCGTGGCAAGACCGGGAGGAAGAGCTGGCGCAGCTCCT				ATTAFFYAPVSRGGL
			CCTGAACTCGCAGCTCGGGGCGTCCACCGGCGCACCCCCGAAGCGCCG				GLLPLTELHAALQVA
			CAACGGCGACATCGGCTCGCTCTGGGTGGACGTCCAGCGCCACCTCCGC				HGWQMLNSKDPAIR
			CATCTCAGCCTCAAGCTGGAGACCGCACCGGCGTGTGCGGAGACCGGC				RIARVQLRQIADARH
			ACCGCCGCGGCCATGCTGCAGCTTCGCGTGCCGCACCACGACACAAGTGG				RIDAQAWQDREEEL
			CTAGACCACAAGACTGTGCTGCGGCACGTGAAGCTGCACTACACAAGAAC				AQLLLNSQLGASTGA
			AAGCACTGGGCCAGGTGGGCGGCCATGMGSGACCAAGGCAAGACGGC				PPKRRNGDIGSLWV
			ACGCACGCATGGGGGCGCGGGGAGCGGCTTCCTCACCCGGCCACGAG				DVQRHLRHLSLKLET
			GCATGTGGGAAGCCGACTACCGCTTCGCGGTGGCCCGCCCGCCTGAACC				APACAETGTAAAML
			AGCTCGACACGCACAGCGTSCTGAAGCGCCGGCGCCTCCGGSMSCACG				QLRVPHHDKWLDHK
			ACCGCTGTCGACAGCCGGGCTGCACCCAGGGSGGAGACGCTGGCGCAC				TVLRHVKLHYKNKH
			GTGCTCAACCACTGCGCCGGCACCATGGACGCGGTCCGCGGCCGCCAC				WARWAAAMXDQGK
			GACGACGCCCTCAAGCACATCGAGCGCGCGCTGCACGCGTCGTCGCCG				TARTHGGAGSGFLTR
			GGCGGCCAGGACCGCGTCGAGCTCCGGGTCAACCAGACGACGGTGCCGTCG				PRGMWEADYRFAV
			CTCGCCGGCCCCGCGCTACGGCCCGACCTCCAGCTGTACAACCACACCA				AARLNQLDTHSVLKR
			AGAAGACGGTGGCGGTGGTCGACCTGGCCGTGGCGTTCGAGGAGCAG				RRLRXHDRCRQPGCT
			GCGAGCGACGACCCGGAGAGCTCGGCGCTGGCACGCATCGCCGCACAC				QGGDAGARAQPLRR
			AAGCGCGCGAAGTATGCCGGCGTCAAGCGACACCTCGAGCGCCAAGGG				HHGRGPRPPRRRPQ
			TGGAAGGTCCACCTCTCGGCGCTCGTGTACGGCTCGCTCGGCGCGGTGC				AHRARAARVVAGRP
			CGGCCGGCAACCACAAGGTGCTTACTGAGCACCTTGGGTTGCTCAAGCG				GPRRAPGQPDGAVA
			TGATGCGAAGCGGCTGGACCGGCAGCTCTCGGTGGCGTGCATCCAGTC				RRPRATARPPAVQP
			CAGCCGCCGCATCTGGAACCTGCACTGCAGCCAGCACCGCGCGCCA				HQEDGGGGRPGRG
			GCACCAAGCACCAGGGGGAAGTCGGGCGGCGGAGACCGGGGGGGACTC				VRGAGERRPGELGA
			CGCCGCGCACCGGCCGCCGCTAGACGGCACACAGGCCCACAGCGGCCG				GTHRRTQAREVCRR
			ACAGGGCCACACCCAGGTAGGGAACCGCCCTCAAACCCCGCCGGTACA				QATPRAPRVEGPPLG
			TTATGGTCCGACACCTATGAGGTGCAACCTGTACACAAGTTACACACCA				ARVRLARRGAGRQP
			CATAGCGACTACCAGGTATTTACTACCTGGAAGCCAAGGATTAACCGGT				QGAY (SEQ ID NO: 1444)
NeSL	Utopia-2_PU	.	CGGTAATACACATAACTTT (SEQ ID NO: 1076)				
		Pythium ultimum	TGACTGGTGTTTGATCACGATCAATGAGGTGATTAACATGAGCCGGAGC	TGACTGGT	TAAGCG		MDYDDSEFFDAICIP
			AGGCCCCTTACACGCTGGTGCTGTAATGGTTCAGGAATGCTCTTATGAG	GTTTGATC	GGGGGT		DEDADVLDDGDEGD
			TAACTCCACAGTATAATTTTTGTTAGCGGGGGTGAGCGCGTGCCGCCGCC	ACGATCAA	CCAGAC		EGGNDDESSEPLPLAI
			CCCACCTTTCTCTTTTCGTTTGTATTTGGTTCCTCTGGAGCGACTTCCGTG	TGAGGTGA	CCCACA		TNAPSAPLHATMLC
			GCTTTTGTCGCGCGCTATCGCGCCCGGCATGAGGATCGTCGCTCCTGAC	TTAACATG	AGAGAG		GTVTQPWLLRFDGA
			GCCATGTGCGCCAGTATATCGGAGCCCCGTGCTCCGCTTGGGTGTGTGT	AGCCGGA	AAGCAG		CRRNPGPAGAGATL
			TCCGCCACGACGCCACGCGATTTGCGTGCACGGCATGCGCTGGCGACTT	GCAGGCCC	GAATCA		TRPNGIILWTHYRYIP
			CCGCTCAGTGCGCGAACTCGCGGTGCATCGTCGACGTGCGCACCGGTCT	CTTACACG	TGGTCC		DKTATNNVAEYEALL
			CTGGCCTTCCAGGACTGGTATGAGTCGTCGTGCGCGTGCAGAGCGCTCT	CTGGTGCT	GCATGG		DGLRCAAHHGVKHL
			TCACCGCCCGTCTCGACGCGTTCGTTCATTCGAGCCGTTGCGACCACAAT	GTAATGGT	ACCAGT		RIEGDSNLVIEQVKGI

	GCGAACCAATCAGCGAACGCCGTCCCCTCGTCCGCCGCCTCGACAACGA	TCAGGAAT	AGGGCA	FACSTSLRPRRDQVR
	GAGCGGACGCGGATGCGACGGCCCCCCCTCCCGCTATCTCTCTTCCGCG	GCTCTTAT	CGCTCC	EILRHFETYSFRHIDR
	GACGGGCTTCCCCCGCCCCTGTCCTAGTTGCGCCGCTCTCCTTCTGGATG	GAGTAACT	ACAAAG	ALNRQADRLANQAL
	CGGCAACTTGGACGACCCACGTGCACGCGTGCCGCCAGTGCGA	CCACAGTA	GTTATC	DLLKTVSVCALSQTR
	CGGCGTTGATCCAGCGTCATCTACACGCCGTTCCTCCCGTATCGACAGCA	TAATTTTT	GCCCTC	VQDDTGAAHGCWH
	ACCGATCTCAACGGCGATGATAGCCAACCTCCAGACCCGCGTGCCAACA	GTTAGCGG	AAACCC	WTPPDASPTDDASTS
	TCCAACAGCTTCCAGCCATCCATTCGACAACGACAACGACAACGACAAC	GGGTGAG	ATCACA	ILTQDVPVPMDIDDY
	GACGACGACCCGCTGCTCGACGACGTGCTCATGGACTACGACGACTCG	CGCGTGCG	CGAAGG	DPDEPMNAADDPVS
	GAGTTTTTCGACGCCATCTGCATTCCAGACGAAGACGCTGACGTGCTCG	CCGCCCCC	ATCTAA	INAEREGGTVYPVLRL
	ACGATGGTGACGAAGGTGACGAAGGTGGTAACGACGACGAGTCGTCC	ACCTTTCT	AAAGAA	GPNVVPERQKRLQIP
	GAGCCACTGCCTCTAGCGATCACGAACGCTCCAAGCGCTCCCTTGCACG	CTTTTCGTT	AGCAAC	WLPPREMQKLEKKIE
	CGACCATGCTCTGCGGCACAGTTACACAGCCATGGCTCCTGCGCTTCGA	TGTATTTG	AATCGA	VLGETFASRIRDAPD
	CGGCGCGTGCCGACGCAACCCAGGACCCGCCGGCGCCGGCGCCACGCT	GTTCCTCT	AATAGT	WFSAEGYITALTSELA
	CACGCGGCCCAATGGTATCATCCTATGGACGCACTACAGGTATATCCCG	GGAGCGA	AAATAG	TLIRQSTAATTGPNA
	GACAAGACGGCGACCAACAACGTTGCCGAGTACGAAGCGCTGCTTGAC	CTTCCGTG	TAAATA	ARPCERTISKEKRRAR
	GGGCTGCGCTGCGGCGCACCACGGCGTCAAACATCTGCGTATTGAA	GCTTTTGT	GCTTAG	RTTPLQRALAEAKHE
	GGTGACTCCAACCTCGTCATCGAGCAAGTCAAAGGCATCTTCGCGTGCT	CGCGCGCT	AAAGGT	LQIIQPDASRKSVRKA
	CCACATCGCTGCGCCCGCGCCGCGACCAAGTCCGCGAGATCCTGCGCCA	ATCGCGCC	CAACAT	ARRVKRISQAQQRH
	CTTCGAGACCTACAGCTTCCGCCACATCGACCGCGCACTTAATCGACAA	CGGCATGA	GCGAAA	DLRRLFSTNERRCVEK
	GCCGACCGCCTGGCCAACCAGGCACTCGATCTCCTCAAGACGGTCTCCG	GGATCGTC	TGCATG	ILRDPPVGPSSTSSSL
	TGTGCGCGCTCTCTCAAACTCGCGTCCAAGACGATACCGGTGCCGCACA	GCTCCTGA	AGGAAC	PATDDDRCTIDPADL
	TGGGTGCTGGCACTGGACGCCGCCAGATGCGTCGCCAACCGACGATGC	CGCCATGT	GTAAGA	FAYFQTQATAPTNFD
	ATCCACGTCGATACTCACGCAGGACGTTCCAGTCCCCATGGATATCGAT	GCGCCAGT	TGGTAA	FDDEGGELFRSVLDE
	GATTACGACCCCGACGAGCCCATGAATGCGGCCGATGATCCAGTCAGC	ATATCGGA	TACATTT	LPRADQEVHLLEDEIT
	ATTAATGCTGAACGAGAAGGTGGTACTGTCTACCCAGTGCTTCGCCTCG	GCCCCGTG	TTATA	RDEIEDQLSRISKSTA
	GTCCCAACGTCGTCCCCGAACGACACAAAAGAGGTTGCAAATCCCGTGGCT	CTCCGCTT	(SEQ ID	PGLDGITNAVYVRFK
	CCCGCCGCGTGAGATGCAAAAACTCGAAAAGAAGATCGAAGTGCTGGG	GGGTGTGT	NO:	LQLLDALQAAFNAC
	GGAGACTTTTGCGTCGCGCATTCGTGACGCGCCCGATTGGTTCTCTGCA	GTTCCGCC	1323)	WRYNRVPSMWKAA
	GAAGGCTACATCACGGCTCTAACGTCCGAGCTTGCAACGCTCATTCGCC	ACGACGCC		FVRLIYKKGNRAVPS
	AATCAACGGCGGCGACGACGGGACCGAATGCAGCACGACCCTGCGAGC	ACGCGATT		NWRPICLQQTVYKLY
	GCACCATCAGCAAGGAAAAGCGGCGCGCGAAGAACGACACCGTTAC	TGCGTGCA		TAILASRLQRWMDA
	AACGCGCGCTCGCGGAGGCAAAGCACGAGCTCCAGATCATTCAGCCTG	CGGCATGC		NARFTMSQKGFRAF
	ATGCCTCGCGCAAGAGCGTGAGGAAGGCTGCACGACGAGTCAAACGGA	GCTGGCGA		NGCHEHNFVATCLH
	TCTCCCAAGCGCAGCAGCGCCACGACCTGCGTCGCCTCTTCTCGACCAA	CTTCCGCT		DQTRRLRKKLAIVWY
	CGAGCGTCGTTGTGTCGAGAAAATCCTTCGTGATCCCCCAGTGGGGCCA	CAGTGCGC		DLRNAFGSLPHEYLW
	TCGAGTACCAGCTCCAGCTTGCCAGCCACCGACGATGACCGCTGTACCA	GAACTCGC		RVLARLGMPPQFVA
	TTGACCCAGCCGACCTGTTCGCCTACTTCCAGACGCAAGCCACGGCCCC	GGTGCATC		RVRQLYADASFTVES
	AACGAACTTCGACTTTGACGACGAAGGCGGCGAACTGTTCCGCTCGGTG	GTCGACGT		RDGTTDPVQLERGVY

		CTCGACGAACTGCCCGCGCGCCGACCAAGAGGTTCACCTTCTTGAGGACG	GCGCACCG		QGCPLSPYLFIAALIPL
		AGATCACTCGGGACGAGATCGAAGACCAGCTCTCGCGCATCTCGAAATC	GTCTCTGG		VRALHKLKDQHGIVL
		GACGGCACCGGGACTCGACGGCATCACGAACGCTGTGTACGTGCCGCTT	CCTTCCAG		APGVTDCVSAYADDI
		CAAGCTACAGCTGCTTGATGCGCTACAAGCTGCATTCAATGCATGTTGG	GACTGGTA		KIFARSGTGAKALHEI
		CGATACAATCGTGTTCCCAGCATGTGGAAGGCGGCGTTTGTGCGTCTGA	TGAGTCGT		VVRFLSWTNMAANP
		TCTACAAAAAGGGCAATCGCGCTGTGCCAAGCAATTGGCGGCCGATCT	CGTGCGCG		AKCALMVTDGARGG
		GCCTACAGCAGACAGTGTACAAACTATACACAGCCATCCTCGCCTCGCG	TGCAGAGC		DDTDASMTLSIEGETI
		ACTGCAGAGATGGATGGACGCGAACGCGCGGTTCACGATGTCGCAAAA	GCTCTTCA		PRLTGKEGYVYLGVE
		AGGATTTAGAGCGTTTAACGGTTGCCACGAACACAATTTTGTGGCTACC	CCGCCCGT		DGLAHERRATCLRDS
		TGTCTACACGACCAAACACGACGCCTTCGAAAGAAGCTTGCGATCGTGT	CTCGACGC		LKAASADVVRLLRSD
		GGTATGATCTCCGGAACGCATTCGGCTCGCTCCCGCATGAGTATTTGTG	GTTCGTTC		LAPWQIVRAIKSHVL
		GCGAGTACTGGCGCGCCTTGGCATGCCACCGCAGTTTGTCGCAAGAGT	ATTCGAGC		SRFDYVLRHLRPFLSL
		GAGGCAGCTGTACGCGGATGCAAGCTTCACCGTGGAATCCCGTGACGG	CGTTGCGA		FDGFDKMLVRGIKRL
		GACGACAGACCCCGTCCAGCTGGAGCGTGGCGTATATCAAGGGTGTCC	CCACAATG		CQLPQTATSEFLFSPT
		TCTGAGTCCATACTTGTTTATCGCCGCACTCATTCCGCTGGTGCGCGCAC	CGAACCAA		SAGGLGFLPLKELFAA
		TCCACAAGCTCAAAGACCAGCACGGCATCGTGCTGGCTCCCGGCGTCAC	TCAGCGAA		LQIVHALQMLHSKDA
		CGACTGCGTCTCGGCTTATGCCGACGACATCAAGATTTTCGCTCGATCG	CGCCGTCC		NVRAIARHQALQVV
		GGAACCGGCGCCAAAGCCCTCCACGAGATCGTGGTGCGATTTCTGTCCT	CCTCGTCC		RKRYALQSDHWSDR
		GGACGAACATGGCAGCGAACCCGGCCAAGTGCGCGCTGATGGTGACCG	GCCGCCTC		EEELLEEFFNGTLERS
		ATGGCGCTCGTGGCGGTGACGACACCGATGCGAGCATGACGCTATCTA	GACAACGA		PFALAKKVSGDIASL
		TCGAAGGTGAGACGATCCCACGCCTCACCGGCAAAGAAGGCTACGTGT	GAGCGGA		WTDVRVNLTKYGLKF
		ATCTCGGAGTGGAAGATGGGCTCGCGCATGAGCGCAGAGCAACGTGCC	CGCGGATG		GEAHGRRLQPLVSHT
		TGCGCGACTCGCTCAAGGCAGCGAGCGCGGATGTCGTCCGCCTGCTTCG	CGACGGCC		DKQLAPQQWASAIK
		CAGCGACCTCGCACCTTGGCAAATCGTGCGTGCGATCAAGTCGCACGTG	CCCCCTCC		THMRLRHLKRWTTL
		CTCTCGCGTTTCGATTATGTACTCCGACACACTTGCGCCCCGTTCCTTAGTCTC	CGCTATCT		VDQGKTARMHERIG
		TTTGACGGCTTCGACAAGATGCTGGTGCGCGGAATCAAGAGGCTGTGC	CTCTTCCG		SAFLTRPSGVYDASY
		CAGCTGCCGCAAACAGCCACGTCCGAGTTTTTGTTCTCGCCCACCTCGGC	CGGACGG		AFAVRARLNQVDTRS
		GGGTGGCCTCGGCTTCCTGCCACTGAAGGAGCTTTTTGCGGCGCTGCAG	GCTTCCCC		ALKRKRIVNNSRCRV
		ATTGTGCACGCACTTCAAATGCTGCACTCGAAGGACGCCAACGTCCGTG	CGCCCCTG		SGCSELETLAHVLNH
		CCATTGCACGGCACCAAGCGCTGCAAGTGGTGCGTAAGCGCTACGCGC	TCCTAGTT		CRFGSDSIRARHAETL
		TGCAGTCTGACCATTGGAGCGACCGCGAAGAGGAACTTCTCGAAGAGT	GCGCCGCT		LLIKTTMERELTRPGR
		TCTTTAACGGCACGCTGGAGCGCTCGCCCTTCGCCCTCGCAAAGAAAGT	CTCCTTCT		QHQRLLVDATVPEAR
		GTCAGGGGACATCGCGTCACTCTGGACGGATGTACGAGTGAATCTCACC	GGATGCG		DPVPSNDAAESNIGA
		AAGTACGGACTCAAATTCGGCGAGGCGCACGGACGGCGACTGCAGCCG	GCAACTTG		MAPSISHLRPDIQLYD
		CTCGTCAGTCACACTGACAAGCAAGCAGCTTGCACCTCAACAATGGGCTAGTG	GACGACCC		NKTMEAVIIDLAVAF
		CAATCAAGACGCACATGCGCCTGCGCCACCTCAAGCGATGGACCACCCT	ACGTGCAC		EDQSTDDAASSSSFA
		GGTCGACCAGGGTAAGACAGCCCGGATGCATGAGCGCATCGGAAGTGC	GCGTGCGC		RVKGVKTKKYEVIKQ
		GTTTCTGACAAGGCCGTCTGGAGTGTACGACGCCAGCTATGCGTTCGCG	TGCCGCCA		FLEYKGYTVHVAALV

			GTTCGAGCAAGACTAAACCAAGTCGATACGCGCAGCGCATTGAAGCGC	GTGCGACG		YGSLGSVDTGNFAVY
			AAGCGCATCGTGAACAATAGCCGCTGCCGTGTGAGTGGATGCAGCGAG	GCGTTGAT		TERLGLRKGAVRRLE
			CTGGAAACGCTCGCCCACGTACTGAATCACTGCCGATTCGGGTCAGACT	CCAGCGTC		CSLSARHINFAHRMI
			CGATACGAGCACGACACGCAGAGACACTCTTGCTGATCAAGACGACCAT	ATCTACAC		WRRHAIAHTTGLRLI
			GGAACGTGAGCTCACGCGGCCAGGGCGGCAGCACCAGCGCCTCCTGGT	GCCGTTCC		GTNSVQQQGVQRA
			CGACGCGACAGTACCCGAAGCTCGTGATCCCGTACCAAGTAATGACGCT	TCCCGTAT		PAEKQQHQRPVQRP
			GCTGAGAGCAACATTGGCGCAATGGCACCCAGCATCTCCCACTTGCGCC	CGACAGCA		SRAQAPRDQPSQQQ
			CCGACATTCAACTCTACGACACAACAAGACAATGGAAGCCGTGATCATCGA	ACCGATCT		QSQQSFQQQSQQS
			CCTCGCGGTTGCGTTCGAAGACCAGTCAACGGACGACGCCGCATCCAGT	CAACGGCG		QQSQQSQQSQQSR
			TCCAGTTTCGCGCGCGTCAAGGGTGTGAAGACAAAGAAATATGAGGTA	ATGATAGC		QRHAPTPTPVPVPVP
			ATCAAGCAATTTCTGGAATACAAAGGCTACACTGTGCATGTAGCAGCGC	CAACCTCC		VPVSTPTLTPTPTPTR
			TAGTCTATGGTTCACTAGGCTCCGTTGACACCGGCAACTTTGCCGTGTAT	AGACCCGC		RPKPAPSSTQPQQG
			ACGGAACGTCTCGGTCTCCGCAAGGGCGCAGTGCGGCGACTGGAGTGC	GTGCCAAC		APAQRRQQREQKKQ
			TCGCTCTCGGCGCGCCACACATCAACTTCGCGCACCGTATGTGGCGACGCC	ATCCAACA		PACRRRHATAVRETP
			ACGCGATCGCCCACACGACCGGTTTGCGCCTTATCGGCACGAACAGCGT	GCTTCCAG		PAAPTAARTATPTAR
			GCAACAGCAGGGTGTGCAGCGCGCCCCGGCAGAGAAGCAACAACATCA	CCATCCAT		PTATTRSTSTTRSTAT
			GCGACCAGTGCAGCGACCCAGCCGCGCGCAAGCTCCACGTGATCAGCC	TCGACAAC		ATSTTRSSAPTSRPSA
			ATCGCAGCAACAGCAGTCCCAGCAGTCGTTCCAGCAGCAGTCCCAGCAG	GACAACGA		PRPRSSAPRPRSSAPT
			TCCCAGCAGTCCCAGCAGTCCCAGCAGTCCCAGCAGTCGCGTCAACGCC	CAACGACA		TRSAAPTTRSAAPTAI
			ACGCCCCAACTCCTACCCCAGTCCCAGTCCCAGTTCCAGTATCG	ACGACGAC		ATTSVYKSRRTTNAIS
			ACACCGACACTAACACCGACACCGACACGCCGGCCCAAGCCG	GACCCGCT		GATTRRSASSKRAPM
			GCTCCGTCCTCGACTCAACCCCAGCAAGGAGCTCCAGCACAGCGACGAC	GCTCGACG		QPRTALTPTQQQQR
			AACAACGAGAGCAGAAAAAGCAACCTGCGTGCCGCCGTCGACACGCAA	ACGTGCTC		Q (SEQ ID NO: 1445)
			CGGCAGTGAGAGAGACACCGCCAGCAGCTCCGACAGCAGCTCGGACAG	(SEQ ID		
			CAACTCCGACAGCGAGACCGACAGCGACAACGAGATCAACATCGACAA	NO: 1200)		
			CGAGATCAACAGCGACAGCGACATCGACAACGAGATCATCAGCACCGA			
			CATCGAGACCATCAGCACCGAGACCGAGATCATCAGCACCGAGACCGA			
			GATCATCAGCACCGACAACGAGATCAGCAGCTCCGACAACGAGATCAG			
			CTGCACCGACAGCGATAGCGACAACCAGCGTTTACAAAAGTCGACGCAC			
			GACCAACGCCATCTCTGGCGCAACGACGCGACGCAGTGCGTCATCCAA			
			GCGTGCGCCAATGCAGCCCCGAACTGCATTGACGCCAACACAACAACAA			
			CAACGACAGTAAGCGGGGGGGTCCAGACCCCACAAGAGAGAAGCAGGA			
			ATCATGGTCCGCATGGACCAGTAGGGCACGCTCCACAAAGGTTATCGCC			
			CTCAAACCCATCACACGAAGGATCTAAAAAGAAAGCAACAATCGAAATA			
			GTAAATAGTAAATAGCTTAGAAAGGTCAACATGCGAAATGCATGAGGA			
			ACGTAAGATGGTAATACATTTTTATA (SEQ ID NO: 1077)			
NeS L	Utopia -3_PI	.	GTACGGCCGAATCCCTGGCCTCTCAGCCTGTACGCGGGGCTATACTTGG	GTACGGCC	TAGACG	MPRSLASEPVHSSAS
		Phytop hthora	TCTAAGTATGCCCCGGTCGCTGGCCTCTGAGCCCGTACACAGCTCGGCC	GAATCCCT	GCAAAG	RLPSQAPPTSRSGAA

		infestans	TCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTGGCGCGGCCC	GGCCTCTC	TTCTGG	HTPPPSLQARPLASA
			ACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTCCGCAGACGC	AGCCTGTA	CCCGCG	DAGLAATALASPQDP
			AGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCCCCCCTACGAT	CGCGGGG	TAGGCC	PYDVAPPGRAAGRLP
			GTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGACTCGGTATCG	CTATACTT	GAAAGG	DSVSPGATLSAATAR
			CCGGGTGCTACACTCTCAGCCGCCACAGCTCGGGCCTTGGCGGTCCGCC	GGTCTAAG	GCCCCG	ALAVRHWPLELDSDS
			ATTGGCCCTTGGAGCTCGACAGCGACAGCGACGAGGACGCTC	T (SEQ ID	CCCAGG	SDDEDAQDPHAAAP
			AAGACCCCCACGCCGCCCGCCCAGGACCCCCCGCAGACGTCGCGAGTG	NO: 1201)	TAGGGT	GPPADVASVLAPPGR
			TGCTTGCCCCACCCGGCAGGGCAGCATGCTCGCCGACCCAGCAG		AACGCC	AGSMLADPAALAAG
			CGCTGGCAGCCGGTCTGGCTCGCGCTCCGCGCCCCCATCGGCACCCCA		CTCGGG	LARAPPPPSAPQDPS
			GGACCCAAGCCCCGCCCTCCCTGCTGGCCCCGCCGGCCAGAACCCTCGC		AAAACC	PALPAGPAGQNPRA
			GCCGCCGCCCCAGCCCGTGTTGAAGTCCACACAGTCGTGGCACCACCCG		ATTCTG	AAPARVEVHTVVAPP
			GCAGGGCGGGCGGCATGCTGCCCGACCCGGGGTTGGTTGACTCAAGCC		GTGTTT	GRAGGMLPDPGLVD
			CAGCAGCGGCTACCGCAGCTACTCCAGCTCCGGTAGCTGCTACTGCTAC		GGCTGT	SSPAAATAATPAPVA
			AACTGCTCGCGTCGCCGTGGAGCACCACGCGCACGCTGAGCCGAACCA		TTTTCAA	ATATTARVAVEHHA
			AGAACATCTACCGATGGCTCGCGTGCTCGTGGAGCCCATGCAGGTCGAC		CAGTCG	HAEPNQEHLPMARV
			GAGTGCAGCAGCTGCGACCGCAGCACTTTGACAGCCGACGACGGGAGC		AACTTCA	LVEPMQVDECSSCD
			GGTGACGACGTTGCAGCCCCAAGCTCCATGCTCAGCAACGACGTCGCCG		ACTCGG	RSTLTADDGSGDDV
			CCCCCATGGACGTTGACTCCGGCACCAGTTGCCCTCCAACGTTGCAGCA		AGCATA	AAPSSMLSNDVAAP
			GCCGCTCCAGCGACCGCGTGCGCTGCACGTAGGCAGCAAGCGCCGCCG		TCAGAT	MDVDSGTSCPPTLQ
			ACTCAACGCAGATGACGGTGAAGAAGCTCACCAGCTGCAAGAAGAGGA		ACACTTA	QPLQRPRALHVGSKR
			GGAGGCTGGCATTCATGCCCCAGCGCTCAGGCTATCCGCGGCCAGCGCT		CTCACAC	RRLNADDGEEAHQL
			CAACCGGCCTCCGTCCTTGCCGTGTACACGCACAACGCGTCGCGGTTCG		ATTTAG	QEEEEAGIHAPALRLS
			ACTGCACGCTGTGCGCGTACACGGCCGGCAGCTTCGCATCACTGAAGAC		ATATCA	AASAQPASVLAVYTH
			GCACAGGAACTCTCGGCACCGCCGCACACCGCATTCCTCGACCGATTCTCG		GATAGG	NASRFDCTLCAYTAG
			GCGGGTTGCGCGTGTGGCGTGCCCTTCGCCTCGAGGCTGGCTGCAGCA		GAACCT	SFASLKTHRNSRHRR
			AGACACGCGCAAGCGTGCGCCAGCCTCTCGAGCGCACCATTGGCCGAG		TTATTAG	TAFLDRFSAGCACGV
			GCTTCGTCGGCAGCAGGAGCATCAAGCCACACTGTCGACGGAGCCGAC		GGAGAT	PFASRLAAARHAQAC
			TCCACCGTCTCAGCCGCGGGTCACGCTGAGCCCGACCTGCCCCGCCACA		AACGGG	ASLSSAPLAEASSAAG
			ACGCCACGGAGCTCACAGCGTCCCCCCCGCTGGTGAGCTCCACCGATGT		TACACC	ASSHTVDGADSTVSA
			TGAAGTGCAGGCAACAGAGACCGAGGCCACCGAGAACCGCTGGGGCA		GGATGG	AGHAEPDLPRHNAT
			CTCCACTTCCGCGCGTGCTGGTCGCGTCTAGGATCGCCGGGCGCCTCGC		TAAATAT	ELTASPPLVSSTDVEV
			TCAGGTGCCACCCCCACGCTGGGGCCCCCCGCTACCGCGCACCACCATC		ACAAAA	QATETEATENRWGT
			GCGGGCCGCATTGCGACGAGGCTCGCCGCCACACCTGCCCCCCGCTGG		CCTTCTC	PLPRVLVASRIAGRLA
			AGCCCACCGCTTCCACGAAGCCTGGTTGCGTCACGCATCGCCGGCGCCGCC		TGTTCTA	QVPPPRWGPPLPRT
			TGCTACCCGCTCTGCCAGACGCGCCTGCATGCGAGGACGAGGCCAAGG		ATCAGT	TIAGRIATRLAATPAP
			ACAGCGACGAAATGGACTGGGAGGCGAGCCGCACGTCGAAGCG		GTGAAA	RWSPPLPRSLVASRIA
			CCAGGGCCGGTCGACGAGGAGACCATCGACGCCGACGGCGAGTG		ACTGGT	GRLLPALPDAPACED
			GCTCCTGCGCTTTGACGGTGCCTGCCGTGCCAACCCAGGCCCAGGCGGC		TTTCGCC	EAKDSDEMDWEASE

				TTTTGGC	PHVEAPGPVDEETID
				GGACTT	DADGEWLLRFDGAC
				TTTCACT	RANPGPGGAGAALF
				CGCATTT	KPSGPVVWTCSHYD
				TTGGGC	PSTTATNNTAEYTALL
				AATCGT	LGARAAADHGVTKL
				CTGCGG	RVEGDSTLVIQQVRG
				CTAGCTT	IFATRSTRLRALRNKV
				GCTAGC	KLELARVGSFSLHHID
				GGCGGA	RQANGHADRLANAG
				CGAGCG	LDRRRTQLECSVHPD
				GTCTCC	GRGCTNTSVATAAPT
				GGGGGC	ASAAPSTPTRPPATT
				GTTCACC	AAPFHSDQGHIDED
				TTTCCCC	DERRADIDDGEIYAP
				CGCGAG	MTLGPDEVPARRPRL
				GCCAAC	RLRQLSDEELEAAGAI
				TACACC	VERLSASLSAKITDAE
				GATCTTC	DWGTAEGYITALPHL
				TCTACAC	LYDKLLPYSRTAPRH
				TTTTCTA	QRPPPRPSRNQQDHP
				ATTCGCC	QPRRDQPQRNVDEQ
				TCCGTCT	QHAESQQGEDQRQ
				TCGGTC	QQPPTRRRRRRGKR
				TTCGGTT	RGRRQRRHPRQPGQ
				GTCGGG	STRASQQSRQRRPRP
				CTTTTTT	PRVTRHHRREHRIDEA
				CTTTTTG	LDELHTLERARPQDR
				ACCAAT	SAIDKARRRVGRVRG
				CAGAGC	AINQHLLRHRFDTDE
				GCGCCA	KACVADILEKAHAAR
				TGCGAC	AARTAQAAGAATST
				TCTTCTG	GGAATAPTQQAATS
				GCCAAT	ALGDADDGTCPILAD
				CAGAGA	ELWQYFTGTNTPRW
				CCGGGC	EFNPATPVGEAFRTA
				CCTGTCC	MARLPPATRLRELLT
				TCGGAC	EAPTADEIETQLQHV

		GGACGTGAACGACAGGCACGCACCAGGACAGAAGGGGTTCAGGGCAG	AGCGAG		RGSSSPGLDGIGYDV
		TGAACGGCTGCGGGGAGCACAACTTCCTCGCAGCCACGCTCATCGACCA	GCCTCC		YQRFAQQLLPVLTAS
		AGCACGTCGCAAGCGCCGATCCCTATACGAGGTATGGTACGACTTCAGG	ACGGCC		FKRCWTAKMVPQS
		AACGCCTTCGGCAGCGTGCCATTCCAGCTGCTGTGGGACTCGCTGCAGC	AGCCAA		WQVGVVRLLYKKGA
		GCCTCGGCGCTCCCGCAGACTTCATCGACATGTGCAAGGGGCTGTACCA	TCAAGT		HDDPANWRPICLQQ
		CCAGGCCGCCTTCGTCATCGGGAACGCCGCCGACGGACCGACAGCAGC	CTCGGC		AIYKLYTGVLARRLVR
		AATCCGGCAGCAGGTTGGCGTCTTCCAAGGGTGCCCACTGAGCCCCCAG	AGCGAC		WLDVNDRHAPGQK
		CTCTTCAACGTCGCGATCAGCCCGCTGCTCTTCGCGCTCCGCAGGCTGCC	GCGTCT		GFRAVNGCGEHNFL
		GGAGACCGGGGTGCAGCTGTCCGGCGACGATCGTGTTGGAGTTTCGGC	TTCTATA		AATLIDQARRKRRSLY
		TTATGCGGACGACCTCAAGACCTTCAGCAGCACCAAGGCGGGCGCCCAC	GCGCAG		EVWYDFRNAFGSVP
		CAAGCAACACGAGCTCGTCGCCGCGTTCCTGGCTTGGACTGGCATGAAA	CTGACG		FQLLWDSLQRLGAP
		GCCAACGCAGCCAAGTGCAGCAGCATGGGCGTACGGCGCAACAGCAAC	AGGCCG		ADFIDMCKGLYHQA
		GGCGCGACAGAGGCGGACAATCTGGACCTCGCGCTCGACGGCACTCCG	ATCTGG		AFVIGNAADGPTAAI
		ATCCCCAGCATGACTCACATGCAATCGTACACCTACCTGGGGATCGGAG	CGGCCC		RQQVGVFQGCPLSP
		ACGGTTTCGACCATGTGCACCGCCGCATCGAGCTCGCGCCCCAAGCTGAA	CCGATT		QLFNVAISPLLFALRR
		GACGCTCAAGCAGGACACGACGGCGCTACTCGAGTCCGGCCTCGCGCC	GGTCCG		LPETGVQLSGDDRVG
		TTGGCAGGTGGTGAAGGCCGTGAAGGTTTACCTTTACCCTCGTGTCGAG	ACTTTCG		VSAYADDLKTFSSTKA
		TACGCGCTTCGGCATCTGAGACCCGAAGACCAGCTGCTGGAGAGCTTCG	GCCAAT		GATKQHELVAAFLA
		ACCTTCACCTCCGCGCCGGCCTTCGCCACCTGCTCCGCCTGCCCAAGAAC	CAGCGA		WTGMKANAAKCSS
		GCCAACAACGACTTCTTCTACTCGCCGGTCTCGCGCGGTGGCTTGGGAC	CGACGA		MGVRRNSNGATEAD
		TCCTGCCGCTGGTGGAGCTGCACGCGGCGCTACAAATCGCGCACGGGT	GGGGGC		NLDLALDGTPIPSMT
		GGCAGATGCTTAACTCGACGGACCCAGCCACCCGACGCATCGCGCGCG	AGGGGT		HMQSYTYLGIGDGF
		AGCAGCTCCACCAGATTGCGGATGCAAGGCACCGCCTTGACAAGGCAC	TTACACT		DHVHRRIELAPKLKTL
		ACTGGAAAGAGCGTGGCGATGAGCTCTGCCAGCTCTTCCTTAACTTGGA	TTTGCCC		KQDTTALLESGLAPW
		CCTCGGGACGTCGGCCCACGCACCGCCCAAACGCCGCAACTGCGACATC	CCGTTTC		QVVKAVKVYLYPRVE
		GGCTCGCTGTGGGTCGACGTCCGCAAGAACCTGCAGGCCTTTGGGCTCA	GGCTTC		YALRHLRPEDQLLESF
		AGCTCGAGACCGCCCCGGCCGACGCTGAGTCCGGCACACCCGCCCTGCC	AACTTCA		DLHLRAGLRHLLRLPK
		GCTGCAGCTTCGCGTGCCACACCACGAGAAGTGGTTGACGCACCGCGA	GGCCAA		NANNDFFYSPVSRG
		CGTGCTGCGGCACGTGAAGCAGCACCTGAAGAACAAACATTGGCGGGC	AATGGC		GLGLLPLVELHAALQI
		GTGGTGTGCCTTCCAAGACCAAGGCAGGACGGCACGGGCACACGGTGG	GATTTG		AHGWQMLNSTDPA
		GGTCGGCAGCAGCTTCATCACCCGCCCCCGTGGCATGTGGGAGAGCGA	GACCCT		TRRIAREQLHQIADA
		CTACCGCTTTGCGGTGGCCGCTCGCCTCAATATGGTGGACACGAGTGCC	CCACGC		RHRLDKAHWKERGD
		ACTCTCGCGCGCCGACGCCTCCGCGCCCATGACCGCTGCCGCTACCCGG	GCCGTG		ELCQLFLNLDLGTSA
		GCTGCCGCTGGAAGGAGTCACTGGAGCACGTGCTCAACCACTGTCCGG	CCACTG		HAPPKRRNCDIGSLW
		GCACGATGGACGCCGTGCGTGGCCGCCACGACGGTGTGCTCCGAGAGA	CTCGGC		VDVRKNLQAFGLKLE
		TTGAGCACGCGCTCCGTGCACCATCGGGAGCACGCAGAGAGCTGCGCG	ACCGGC		TAPADAESGTPALPL
		TGAACCAGACGGTTCCGGGCCTACCCGGCCCCGCACTACGCCCCGACAT	GGCGAT		QLRVPHHEKWLTHR
		CCAGGTCTACAACCACGACCAACGCACGGTCGCAGTGGTGGACCTGGC	TCAGCG		DVLRHVKQHLKNKH

		GGTGGCCTTCGACAGGCAGGACCGCGACGACCCAGAGACCTCGGGCCT		GGTGCA	WRAWCAFQDQGRT
		GGCCAAAGCCGCGGCGGAGAAGAAGGCCAAGTACACTGGCATCCAGC		ACTTCG	ARAHGGVGSSFITRP
		GACACCTCGAGCGCCAAGGGTGGAAGGTCCACCTTTCCGCGCTCGTCTA		GGCACG	RGMWESDYRFAVAA
		CGGGTCACTGGGCTCGGTGGCCCCCAACAACTACAAAGTCTACACGGA		TGTGCA	RLNMVDTSATLARRR
		GCACCTGGGGCTCCTCAAGCGTGACGCCAAGCGGCTGGACCGGACGCT		ACACAT	LRAHDRCRYPGCRW
		GTCGGTCGCGTGCATCCAGTCCAGCCGCCGCATCTGGAACCTGCACTGC		GCAGCG	KESLEHVLNHCPGTM
		GCCAAGCATCGGGCGCGGCAGCACCAGACGCCATCCCAGTCAAGGGGC		CCCATTG	DAVRGRHDGVLREIE
		AGGCGGGTGACGGAGACCGGGGGGGCTCCGTCGCGCACCGATCGCCG		CACGCC	HALRAPSGARRELRV
		CTAGACGGCAAAGTTCTGGCCCGCGTAGGCCGAAAGGGCCCCCGCCCAG		AAGCGG	NQTVPGLPGPALRPD
		GTAGGGTAACGCCCTCGGGAAAACCATTCTGGTGTTTGGCTGTTTTTCA		CATCGC	IQVYNHDQRTVAVV
		ACAGTCGAACTTCAACTCGGAGCATATCAGATACACTTACTCACACATTT		GGGACG	DLAVAFDRQDRDDP
		AGATATCAGATAGGGAACCTTTATTAGGGAGATAACGGGTACACCGGA		ACGCCT	ETSGLAKAAAEKKAK
		TGGTAAATATACAAAACCTTCTCTGTTCTAATCAGTGTGAAAACTGGTTT		CGGCCG	YTGIQRHLERQGWK
		TCGCCTTTTGGCGGACTTTTTCACTCGCATTTTTGGGCAATCGTCTGCGG		CTCAAG	VHLSALVYGSLGSVA
		CTAGCTTGCTAGCGGCGGACGAGCGGTCTCCGGGGGGCGTTCACCTTTCC		CGCAGC	PNNYKVYTEHLGLLK
		CCCGCGAGGCCAACTACACCGATCTTCTCTACACTTTTCTAATTCGCCTCC		CCCGCC	RDAKRLDRTLSVACI
		GTCTTCGGTCTTCGGTTGTCGGGCTTTTTTCTTTTTGACCAATCAGAGCG		CTTCCAG	QSSRRIWNLHCAKH
		CGCCATGCGACTCTTCTGGCCAATCAGAGACCGGGCCCTGTCCTCGGAC		CACGAC	RARQHQTPSQSRGR
		AGCGAGGCCTCCACGGCCAGCCAATCAAGTCTCGGCAGCGACGCGTCTT		CTCGCG	RVTETGGAPSRTDRR
		TCTATAGCGCAGCTGACGAGGCCGATCTGGCGGCCCCCGATTGGTCCGA		CCGTTT	(SEQ ID NO: 1446)
		CTTTCGGCCAATCAGCGACGAGGGGGCAGGGGTTTACACTTTTGCC		GGCGGA	
		CCCGTTTCGGCTTCAACTTCAGGCCAAAATGGCGATTTGGACCCTCCACG		TCGCCAT	
		CGCCGTGCCACTGCTCGGCACCGGCGGCGATTCAGCGGGTGCAACTTC		CAAGAC	
		GGGCACGTGTGCAACACATGCAGCGCCCATTGCACGCCAAGCGGCATC		GTGCGA	
		GCGGGACGACGCCTCGGCCGCTCAAGCGCAGCCCCGCCCTTCCAGCAC		GAGCCA	
		GACCTCGCGCCGTTTGGCGGATCGCCATCAAGACGTGCGAGAGCCAGG		GGCGGG	
		CGGGGTCGGGCAAAATATACTTACTCTAAGTATGCCCGAATCCCTGCCC		GTCGGG	
		TCTCAGGCTGTACGCGGCCCCATACTTGACCTAAGTATGGGAGGATCCC		CAAAAT	
		TGGCCTCTCAGGCTGTACGCGRGGACCAAGTACAGCCGAATCCCTGGCC		ATACTTA	
		TCTCAGCCTGTACGCGGGGCTATACTTGGTCTAAGTATGCCCCGGTCGC		CTCTAA	
		TGGCCTCTGAGCCCGTACGC (SEQ ID NO: 1078)		GTATGC	
				CCGAAT	
				CCCTGCC	
				CTCTCAG	
				GCTGTA	
				CGCGGC	
				CCCATAC	
				TTGACCT	

NeSL	Utopia-3_PR	.	Phytophthora ramorum	TGCGCGGCAGACCAAGACGCGCAGCCAACAACAGACCGTGCAGCGTGC	TGCGCGGC	TAGACG	MQDQVDAEQQARN
				GGGTGGAAAGCGCCGCCGCCTGAACGCTGGTGACGATGAAGACCAGC	AGACCAAG	GCAACA	RWGPPLPRPLVASRV
				GAGAGCTGGCCGAGCTCCTGCTCGTGGACGAGGACAAGGCTGGCGCCG	ACGCGCAG	CTCTGG	AARLGEVPPPRWGP
				AACACCCCGCGCTCAGGCTGCCCACGGCCAGCGCTCATCCGGCCTCCGT	CCAACAAC	CCCGCG	PLPRGVVVSRIAARLE
				CCTCTCCGTGTACGCGCACGCTGCAACTCGCTTCGACTGCACGCTGTGCA	AGACCGTG	TCGGCC	AVPVPRGGPPLPRSF
				CGTACACGGCTGCCAGCCTCGCTTCGCTCAAGCGCCACCGCTCGTCTCG	CAGCGTGC	GAAAGG	VATRIADRLAPPSPDL
				GCACCGACGCACGGCCTTCCTCGACAAGTTCTTGGCGGGCTGCGCGTGC	GGGTGGA	GCCCCA	SLLDEEMKESEPPDP
				GGCACGCCCTTCGCATCGAGGTTGGCCGCAGCCAGACACGCGCAAGCG	AAGCGCCG	CCCACG	THHSADEDSTDAETA
				TGCGCCAACCTCTGCACCACCTCGGCGACGACTTCGACGGCAGCAAAGG	CCGCCTGA	TAGGGA	DAVMEPAFVSDPPT
				CATCAAGCCCCACTGCTGCCGGAGGCAGACCCACCGTCCGTGCAGTGGT	ACGCTGGT	ACCGCC	ATPREWRLQFDGAC
				CACCGCCGCGCCCGACCTGCCCCGCCAGTATCCCTCGGAGCTCGTTGCG	GACGATGA	CTCGGG	RGGPNPGGAGALLY
				TCCCCCCCGCAGCCGAGCTCCACCAACGTTGCAATGCAGGACCAGGTCG	AGACCAGC	AAACCC	NPEGAVVWTGSHY
				ACGCAGAGCAGCAGGCGCGAAATCGCTGGGGCCCACCGCTCCCTCGAC	GAGAGCT	AGTCTG	MPGAKETNNSAEYT

	CGCTGGTGGCCTCCCGGGTCGCAGCCCGACTCGGCGAGGTACCGCCGC	GGCCGAG	GTGTTC		ALLIGARAAADHGAR
	CGCGATGGGGCCCACCACTTCCGCGAGGGGTGGTGGTGTCCCGTATTG	CTCCTGCT	GGCCAA		QLRIEGDSLLVIRQVK
	CGGCCCGGCTGGAAGCCGTCCCAGTCCCACGCGGGGGACCCCCGCTCC	CGTGGACG	GAAATG		GLYATKSTRLRQLRN
	CACGCAGCTTCGTCGCGACCAGGATCGCAGACCGCCTCGCTCCCCCCTC	AGGACAA	CACCAC		AVRHELARVGQHSL
	CCCGGACCTGTCGCTGCTGGACGAGGAGATGAAGGAGAGCGAGCCGC	GGCTGGC	CACCAC		HHIDRQGNAFADRL
	CTGACCCGACGCACCACTCAGCGGACGAGGATTCCACGGACGCAGAGA	GCCGAACA	GGCGGA		ANRALDLKSDKVECK
	CGGCGGACGCAGTGATGGAGCCCGCGTTCGTCAGCGACCCACCCACGG	CCCCGCGC	GGTGCA		EHPVAGACTTCMGS
	CGACACCCCGGGAGTGGCGCCTCCAATTCGATGGAGCTTGCCGTGGCG	TCAGGCTG	TTTCGAC		PSAGPPATPPPTTADI
	GGCCCAACCCCGGTGGTGCTGGCGCACTGCTCTACAACCCCGAGGGTG	CCCACGGC	AGTCGA		EMADAGSDDELRAD
	CTGTCGTGTGGACGGGTTCGCACTACATGCCGGGCGCCAAGGAGACGA	CAGCGCTC	ACTTCAA		IDDGEVYAPMRLEPG
	ACAACTCGGCCGAGTACACGGCCCTCCTGATCGGCGCCCGCGCAGCAG	ATCCGGCC	CCCGCC		VIPTRRSRLRLRQLTD
	CAGACCACGGTGCACGACAGCTGCGCATCGAGGGCGACAGCCTGCTGG	TCCGTCCT	ACATATC		DEMEAAGEVVERLS
	TCATTCGTCAAGTCAAGGGATTGTACGCCACCAAGAGCACTCGGCTGCG	CTCCGTGT	GGATAT		AGLSAKIADADDWE
	ACAGCTCCGGAACGCCGTTCGCCACGAGCTGGCTCGGGTGGGGCAGCA	ACGCGCAC	AGTTAC		TAEGYITALPYMLYD
	CTCGCTCCACCACATCGACCGGCAAGGCAACGCGTTCGCGGACCGCCTC	GCTGCAAC	AGCTCT		KLQQYTQVRHGTAR
	GCCAACCGGGCGCTCGACCTGAAGAGCGCGACAAAGTCGAGTGCAAGGAA	TCGCTTCG	AGTTAG		SPAPHPQRRDVQGQ
	CACCCCGTGGCCGGCGCGTGCACGACATGCATGGGTTCGCCCTCGGCG	ACTGCACG	ACATCG		VETHREPRHETIGQP
	GGCCCGCCAGCGACACCGCCCCCTACCACTGCCGACATCGAGATGGCG	CTGTGCAC	GATAGG		DQPGEPSPTRRRRRG
	GACGCGGGGAGCGACGACGAGCTCCGAGCAGACATCGACGACGGTGA	GTACACGG	AACTTCT		KRKGRRQRRHPRRT
	AGTGTATGCTCCGATGCGCTTAGAGCCCGGCGTGATCCCGACCCGCCGC	CTGCCAGC	TAGAAA		NCGGGGRQQRKQR
	TCGCGCCTCCGCCTTCGCCAGCTGACCGACGATGAGATGGAGGCAGCT	CTCGCTTC	ATTAAC		HPRPPRGTRHHREH
	GGTGAGGTAGTCGAGCGCTTGAGTGCGGGCCTGTCGGCCAAGATCGCG	GCTCAAGC	GGGTAT		RIDEAIDELHALERAR
	GACGCCGACGACTGGGAGACGGCGGAAGGCTACATCACAGCCCTCCCG	GCCACCGC	ACCGGA		PQARPAIAKARRRVG
	TACATGCTCTACGACAAGCTGCAGCAGTACACTCAGGTGCGGCATGGCA	TCGTCTCG	TGGTAA		RIRSAIDQQLLRHRFD
	CGGCTCGTTCACCAGCGCCCCACCCACAGCGCCGAGACGTCCAAGGCCA	GCACCGAC	ATAAAA		TAEKECVDGILAAAR
	GGTGGAAACCCACCGGGAGCCGCGGCACGAGACCATTGGTCAGCCAGA	GCACGGCC	TAAAAA		TARDARTTVRAAAAT
	CCAGCCAGGTGAGCCGTCACCGACGCGCCGCCGCCGAGGAAAGCG	TTCCTCGA	CTTC		GTTATPETAVTSGTE
	TAAAGGCCGCCGCCAGCGACGTCATCCCCGACGGACCAATTGTGGTGG	CAAGTTCT	(SEQ ID		QQDDNGTCPIPSEVL
	TGGTGGCCGTCAGCAGCGGAAGCAACGGCATCCACGCCCACCGCGCGG	TGGCGGG	NO:		WRHFDSVNTPQRDF
	GACACGCCACCACAGGGAGCACCGCATCGACGAAGCCATCGACGAACT	CTGCGCGT	1325)		DPEAPEGAAFRSAM
	GCACGCGCTGGAGCGTGCCCGTCCCCAAGCTCGCCCAGCCATTGCCAAA	GCGGCAC			ARLPAATRFMELLKE
	GCTCGCCGCCGAGTTGGCCGGATCCGATCGGCGATCGATCAGCAGCTTC	GCCCTTCG			EPSTDGIEVQLQHAS
	TGCGCCACCGCTTCGACACCGCCGAGAAGGAGTGCGTGGATGGCATCT	CATCGAGG			STSSPGLDGVGYDVY
	TGGCAGCAGCTCGCACGGCACGCGACGCACGGACAACAGTGAGGGCG	TTGGCCGC			KRFASQLLPVLKAAFK
	GCCGCGGCAACAGGGACGACAGCGACTCCAGAGACGGCAGTGACGAG	AGCCAGAC			CCWTHKQVPQSWK
	CGGTACGGAGCAGCAGGACGATAACGGCACCTGCCCCATCCCGAGCGA	ACGCGCAA			LGVVRLLYKKGDRED
	GGTGCTGTGGAGACACTTCGACAGCGTCAATACGCCCCAGAGGGACTT	GCGTGCGC			PANWRPICLQQAIYK
	CGACCCGGAGGCTCCCGAGGGTGCTGCGTTCCGTTCTGCGATGGCCCG	CAACCTCT			IYTGVLARRLTRWQD

		GCTGCCAGCAGCCACGCGATTCATGGAGCTGCTGAAGGAAGAGCCGTC	GCACCACC		ANDRHAPGQKGFRP
		CACGGACGGCATCGAGGTCCAGCTGCAGCACGCGAGCAGCACCTCCAG	TCGGCGAC		VNGCGEHNFLAAML
		CCCGGGCCTCGACGGCGTCGGCTACGACGTGTACAAGCGCTTCGCGTC	GACTTCGA		IDHARRKHRPLYEVW
		GCAGCTGCTGCCGGTGCTGAAGGCGGCTTTCAAGTGCTGCTGGACGCA	CGGCAGCA		YDFRNAFGSVPLGLL
		CAAGCAGGTGCCCCAGAGCTCGGCGTGGTGCGCCTCCTGTA	AAGGCATC		WDALERTGVPAEYIA
		CAAGAAGGGCGACCGTGAAGACCCGGCGAACTGGAGGCCCATCTGCCT	AAGCCCCA		AVQGLYDHAAFMVG
		CCAGCAAGCCATCTACAAGATCTACACGGGGGGTGCTCGCGCGGCGCCTC	CTGCTGCC		NAVDGSTAPILQRVG
		ACCCGCTGGCAGGACGCCAACGACCGCCACGCCCCGGGCCAGAAGGGC	GGAGGCA		VFQGCPLSPPLFSAAI
		TTCCGGCCGGTGAATGGCTGCGGGGAACACAACTTCCTCGCGGCCATGC	GACCCACC		SPLLHALQLLPSSGVQ
		TCATCGACCACGCGCGCCGCAAGCACCGGCCACTGTACGAGGTGTGGT	GTCCGTGC		LSGDDRPGVSAYAD
		ACGACTTCCGGAACGCTTTCGGAAGCGTACCCCTCGGGCTGCTGTGGGA	AGTGGTCA		DLKTFSGTKAGVTEQ
		CGCCCTGGAGCGGACAGGCGTTCCCGCCGAGTACATCGCTGCCGTCCA	CCGCCGCG		HELVAMFLRWTGM
		GGGCTTGTACGACCACGCAGCGTTCATGGTGGGCAACGCGGTGGACGG	CCCGACCT		AANPAKCRSMGVRR
		CAGCACGGCGCCGATCCTGCAGCGTGTCGGAGTGTTCCAGGGGTGCCC	GCCCCGCC		NGNGAIEADHLELAL
		GCTGAGCCCACCGCTGTTCTCGGCGGCGATCAGCCCCCTCCTCCACGCG	AGTATCCC		DDTPIPTLTHLQSYTY
		TTGCAGCTACTGCCCAGCTCAGGAGTACAGCTCTCGGGTGACGACCGCC	TCGGAGCT		LGIGDGFDHVRRRVA
		CGGGTGTCTCCGCGTACGCCGACGACCTCAAGACCTTCAGCGGCACCAA	CGTTGCGT		LAPKLKLLKQDATAL
		GGCGGGCGTCACCGAGCAGCACGAGCTAGTGGCCATGTTCCTGCGATG	CCCCCCCG		MESGLAPWQVVKA
		GACCGGCATGGCGCCGAACCCGGCCAAGTGCAGGTCGATGGGCGTGC	CAGCCGAG		VKGYLYPRVEYALRH
		GCCGCAACGGCAACGGCGCCATCGAGGCGGACCACCTGGAGCTGGCGC	CTCCACCA		LRPDDQLLESFDLHLR
		TGGACGACACGCCGATCCCTACGTTGACGCACCTGCAGTCGTACACGTA	ACGTTGCA		RGLRHLLRLPKSANN
		CCTGGGGATCGGCGACGGCTTCGACCACGTGCGCCGCCGCGTCGCGCT	(SEQ ID		DFVYAPVSRGGLGFL
		CGCGCCCAAGCTCAAGTTGCTCAAGCAGGACGCGACGGCGCTGATGGA	NO: 1202)		PLVELHAALQIAHGW
		GTCCGGCCTCGCGCCGTGGCAGGTGGTGAAGGCCGTGAAGGGGTACCT			QMINSPDPAIRRIAR
		CTACCCTCGGGTGGAGTACGCCCTCCGGCACCTGCGACCCGACGACCAG			EQLHQVADARHRLD
		CTCCTGGAGAGCTTCGACCTCCACCTCCGCCGCGGCCTCCGCCACCTGCT			KDHWKQRGDELCEL
		CCGCCTCCCCAAGAGCGCCAACAACGACTTCGTGTACGCGCCGGTGTCG			LLNGELGTSAHAPPK
		CGCGGGGGCCTGGGCTTCCTGCCGCTGGTGGAGCTGCACGCGGCGCTG			RRNGDIGSLWVDVR
		CAGATTGCGCACGGGTGGCAGATGATGATCAACTCGCCCGACCCAGCCATCC			KNLKAFGLKLATAPA
		GGCGCATCGCGCGCGAGCAGCTTCACCAGGTTGCCGACGCGCGGCACC			DPESGAPAKPLQLCV
		GGCTCGACAAGGACCACTGGAAGCAGCGCGGTGACGAGCTGTGCGAG			PHHAEWLDHRNVLR
		CTCCTCCTCAACGGGGGAGCTCGGGACGTCGGCCCACGCGCCACCCAAAC			HVKQHMKNKRWRA
		GGCGCAACGGCGACATTGGCTCGCTGTGGGTCGACGTCCGCAAGAATC			WCSHVDQGRTARA
		TCAAGGCCTTCGGGCTGAAGCTCGCAACCGCGCCCGCCGACCCGGAGT			HGGVGSGFLTRPRG
		CCGGCGCACCTGCCAAGCCGCTGCAGCTTTGCGTGCCGCACCACGCTGA			MWESDYRFAVAARL
		GTGGCTGGACCACCGCAACGTGCTGCGGCACGTGAAGCAGCACATGAA			NMLDTVNVLARRRL
		GAACAAGCGCTGGCGGGCCTGGTGCAGCCACGTGGACCAGGGGCGGA			RAHDRCRHPGCRWK
		CGGCGCGCGCGCACGGAGGCGTCGGCAGCGGTTTCCTGACGAGACCGC			ETLAHVLNHCPGTM

			GGGGCATGTGGGAGAGCGACTACCGCTTCGCGGTGGCCGCTCGCCTCA				DSIRGRHDDALKEIER
			ACATGCTTGACACGGTGAACGTGCTCGCGCGCAGACGCCTCCGCGCCCA				TLHASSGDRQGRTEL
			CGACCGTTGCCGTCACCCGGGCTGTCGGTGGAAGGAGACGTTGGCACA				RTNQTVPGLAGPALR
			CGTCCTCAACCACTGCCCGGGTACGATGGACAGCATCCGCGGCCGCCAC				PDLQVYNHDQRTVA
			GACGACGCGCTCAAGGAAATCGAGCGCACGCTCCACGCGTCGTCGGGC				VVDLAIAFDEQPRDD
			GACCGCCAGGGCCGCACCGAGCTCCGCACGAACCAGACCGTGCCGGGC				PESSGLAKAAAEKKA
			CTGGCCGGGCCAGCTCTGCGCCCCGACCTCCAGGTCTACAACCACGACC				KYAGIKRHLERQGW
			AACGCACGGTCGCGGTGGTGGACCTGGCCATCGCGTTCGACGAGCAGC				KVHLSALVYGSLGSV
			CGCGCGACGACCCAGAGAGCTCCGGGCTGGCGAAGGCGGCGGCCGAG				APSNYKVYTEHLGLLK
			AAGAAGGCCAAGTATGCGGGCATCAAGCGACACCTCGAGCGCCAAGGG				RDAKRLDRQLSVACI
			TGGAAGGTCCACCTCTCGGCGCTCGTGTACGGCTCGCTGGGCTCGGTGG				QSSRRIWNLHCAQH
			CACCCAGCAACTACAAGGTGTACACCGAGCACCTGGGGCTGCTCAAGC				RARQHQDQPAPRGR
			GGGACGCCAAGCGGCTGGACCGGCAGCTGTCGGTGGCGTGCATCCAGT				RVTETGGTPSRTDRR
			CCAGCCGCCGCATCTGGAACCTGCACTGCGCCCAGCACCGCGCCAGGCA				(SEQ ID NO: 1447)
			GCACCAAGACCAACCAGCCCCAAGGGGCAGGCGGGTGACGGAGACCG				
			GGGGGACTCCGTCGCGCACCGATCGCCGCTAGACGGCAACACTCTGGC				
			CCGCGTCGGCCGAAAGGGCCCCACCCACGTAGGGAACCGCCCTCGGGA				
			AACCCAGTCTGGTGTTCGGCCAAGAAATGCACCACCACCACGGCGGAG				
			GTGCATTTCGACAGTCGAACTTCAACCCGCCACATATCGGATATAGTTAC				
			AGCTCTAGTTAGACATCGGATAGGAACTTCTTAGAAAATTAACGGGTAT				
			ACCGGATGGTAAATAAAATAAAAACTTC (SEQ ID NO: 1079)				
NeSL	Utopia-4_PI	AATU0 10012 81.1	Phytophthora infestans	AGCTCGGCCTCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTG	AGCTCGGC	TAAACG	MLADPAALAAGLAR
				GCGCGGCCCACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTT	CTCGCGGC	GGTCAC	APPPPSAPQDPSPAF
				CGCAGACGCAGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCC	TGCCTTCC	TTGACC	PAGPAGQNPRAAAP
				CCCCTACGATGTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGA	CAGGCGCC	GACAGG	ARVEVHTVVAPPGR
				CTCGGTATCGCCGGGTGCTACACTCTCAGCCGCTACAGCTCGGGCCTTG	GCCGACTT	GCACCA	AGGMLPDPGLVEEPI
				GCGGTCCGCTATTGGCCCTTGGAGCTCGAGGCGACAGCAGCGACGAC	CGCGCTCT	CCCAGG	QATYAHDAAQFECA
				GAGGACGCTCAAGACCCCCACGCCGCCGCCCCAGAACCCCCAGAAGAC	GGCGCGG	TAGGGA	LCPYVAESMAVLVQ
				GTCGCGAGTGTGCTTGCCCCACCCGGCAGGGCAGCATGCTCGCC	CCCACACG	ACCGCC	HRRSAHRGTRFKDIF
				GACCCAGCAGCGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCGCCCCCAT	CCGCCGCC	CTTTAAA	TSGCQCSLVFYARIVA
				CGGCACCCCAGGACCCAAGCCCCCGCCTTCCCTGCTGGTCCCGCCGGCCA	GAGCCTCC	ACCCAG	ASHAVACARRNQRA
				GAACCCTCGCGCCGCCGCCCCAGCCCGTGTTGAAGTCCACACAGTCGTG	AAGCGCGC	GAAGAC	VPPAPTPVAPTRPEA
				GCACCACCCGGCAGGGCGGGCGGCATGCTGCCCGACCCGGGGTTGGTT	CCGTTGGC	ACAAAC	TPQPTGYLAAAMTA
				GAAGAACCTATCCAAGCGACGTACGCGCACGACGCTGCCCAGTTCGAG	TTTCGCAG	ACCCTCC	AAAAASSDTVVAAAT
				TGTGCGTTGTGCCCGTACGTCGCTGAGAGCATGGCAGTGCTCGTTCAGC	ACGCAGG	ACATAG	NMQSAVPAAAKTTG
				ACCGACGGTCGGCTCACCGCGGCACCCGCTTCAAGGACATCTTCACCAG	GCTCGCGG	TGACAT	LQLVPPELEPALPQR
				CGGGTGCCAGTGCAGCCTAGTGTTTTACGCGCGCATTGTTGCCGCCAGC	CGACGGCC	ACATATT	ASCHAGKRRRLNAD
				CATGCAGTGGCCTGTGCCAGGCGCAACCAGCGAGCCGTGCCACCAGCT	CTCGCCAG	TTAGCCT	EAVTPCTPTARVSPQ

	CCGACCCCTGTTGCCCCCACGCGGCCCGAGGCTACCCCACAACCTACTG	CCCCCAAG	AGATTTC	TEVAMAPHDAPQD
	GGTACCTCGCCGCGGCGATGACTGCCGCTGCGGCAGCCGCATCAAGCG	ACCCCCCC	AGTTAC	DTVLQREAAEPQPDP
	ACACCGTAGTGGCAGCCGCCACCAACATGCAGAGCGCCGTGCCAGCCG	TACGATGT	GGAGAG	AATQGAQVQRVEDT
	CCGCCAAGACTACTGGGCTTCAGCTCGTCCCCCCCGAGCTGGAGCCTGC	GGCACCAC	GTTACTA	TAAQDDTVQQDHD
	CCTTCCGCAGCGTGCTTCTTGCCATGCAGGCAAACGCCGTCGGCTCAAT	CCGGCAG	ACTGGT	ADTAQVSPPRRTPTR
	GCCGACGAGGCAGTGACACCGTGTACGCCTACGGCGCGGGTGTCGCCG	GGCGGCC	ACATAA	WGPRPSSTQEPSPM
	CAAACAGAAGTGGCGATGGCGCCTCACGACGCACCCCAGGACGACACC	GGCAGGCT	AATTAC	TGEPAATLAARRPLT
	GTCCTACAGCGTGAGGCCGCCGAGCCACAGCCAGACCCAGCAGCAACG	GCCCGACT	ACATTCT	PAATGTRATRWGPC
	CAGGGTGCCCAAGTGCAGCGAGTCGAAGACACAACTGCGGCGCGAGGAT	CGGTATCG	GTTCTAA	HRAIGAAAIARLVTGL
	GACACTGTGCAGCAGGACCACGACGCCGACACTGCCCAGGTCTCGCCG	CCGGGTGC	TCAGTG	PTEPAQPQRRQPPPP
	CCCCGGCGCACACCGACACGCTGGGGACCAAGACCAAGCTCAACTCAA	TACACTCT	TGAAAA	QEPPLQPEPQAAAAT
	GAGCCTAGCCCCATGACAGGCGAGCCGGCTGCCACCCTGGCAGCGCGA	CAGCCGCT	CTGGTTT	VAADIAATVAADIAA
	AGACCACTGACTCCTGCCGCAACGGGCACTCGTGCCACGCGGTGGGGG	ACAGCTCG	TCGCCTT	AAANAAMDVDGGP
	CCATGTCACCGCGCGATCGGGGCCGCAGCCATCGCACGACTTGTCACGG	GGCCTTGG	TTGGCG	AADETWLLRFDGAC
	GCCTTCCAACGGAGCCAGCGCAGCCTCAGCGACGGCAGCCACCGCCAC	CGGTCCGC	GACTTTT	RRNPGPGGAGAALF
	CACAAGAACCTCCTTTGCAGCCAGAGCCACAAGCAGCAGCAGCAACAGT	TATTGGCC	TCACTCG	APSGAVVWTCSHFM
	CGCAGCAGACATCGCAGCAACAGTAGCAGCAGACATTGCGGCAGCAGC	CTTGGAGC	CATTTTT	PSRSETNNTAEYTALL
	AGCGAACGCAGCAATGGACGTGGATGGTGGACCGGCAGCCGACGAGA	TCGACGGC	GGGCAA	LGAQSAVHHGAKRL
	CATGGCTGTTGCGCTTTGACGGCGCATGCCGACGGAACCCAGGACCTG	GACAGCA	TCGTCT	NIEGDSHLILSQVRGA
	GTGGAGCAGGTGCCGCACTCTTTGCGCCCAGCGGTGCCGTAGTATGGA	GCGACGAC	GCGGCT	FACNNKRLRSLRNRV
	CGTGCTCGCACTTCATGCCGAGCCGCTCCGAAACGAACAACACCGCCGA	GAGGACG	AGCTTG	QASLRQLDWYRLQHI
	GTACACGGCGCTACTGCTCGGCGCACAGAGCGCAGTCCACCACGGCGC	CTCAAGAC	CTAGCG	DRKANQHADRLANR
	CAAACGCTTGAATATCGAGGGCGATAGCCACCTGATCCTCTCGCAAGTG	CCCCACGC	GCGGAC	ALDLRRTVTECGPHA
	CGCGGTGCCTTCGCGTGCAACAACAAGCGGTTGCGGTCTCTCCGAAACC	CGCCGCCC	GAGCGG	ETRNRCFQTPQPLVE
	GAGTCCAGGCCTCTCTTCGGCAGCTCGACTGGTACCGCTTGCAGCACAT	CAGAACCC	TCTCCG	PGETHCVPGSDEVLA
	CGACCGCAAGGCCAACCAGCATGCCGACCGCCTCGCGAACCGCGCCCT	CCAGAAGA	GGGGCG	ANTAMEDATAVPTE
	GGACCTGCGCCGCACGGTGACCGAGTGCGGCGCCCACACGCGGAGACTCG	CGTCGCGA	TACACCT	DDEAEVAARDGGEV
	CAATCGGTGCTTTCAGACACCGCAGCCCCTCGTCGAGCCCGGCGAGACG	GTGTGCTT	TTCCCCC	FPTIAIGPDSAPAKQP
	CATTGTGTACCGGGGGTCTGATGAGGTCCTCGCTGCTAATACCGCGATGG	GCCCCACC	GCGAGG	RLRLKKLDEDDFDAA
	AGGATGCGACCGCTGTGCCTACCGAAGACGACGAGGCTGAGGTTGCAG	CGGCAGG	CCAACT	AAAVTRVSEELASKIV
	CGCGAGACGGGGGTGAGGTGTTTCCAACCATCGCCATCGGGCCTGACT	GCAGGCA	ACACCG	DAGDWTSGEGYISAI
	CCGCCCCTGCCAAGCAACCACGGCTTCGCCTCAAGAAGCTGGACGAAGA	GC (SEQ ID	ATCTTCT	PERLRAALRPFALPTQ
	CGATTTCGATGCGGCTGCCGCCGCCGTGACGCGTGTGTCCGAGGAGCT	NO: 1203)	CTACACT	PARPQPREPRMQQP
	GGCAAGCAAGATCGTGGACGCTGGCGACTGGACATCAGGTGAAGGCTA		TTTCTAA	PRRPPRVTRDHLEHR
	CATCAGCGCAATTCCTGAACGACTTCGAGCAGCGCTCCGGCCATTTGCC		TTCGCCT	LDEALDTMENVQRS
	CTTCCGACGCAGCCAGCACCACAGCCCCGGGAGCCTCGTATGCAGC		TCGTCTT	TPQNQKAVRRARRR
	AGCCACCCCGCCGGCCACCCCGCGTCACACGCGACCACCTGGAGCACAG		CGGTCT	VGRLRSAMDRTRLRK
	GCTTGACGAGGCGCTCGACACCATGGAGAATGTGCAGCGCTCCACGCC		TCGGCT	KFATHERECVAEILRR

		GCAAAACCAGAAGGCAGTGCGCAGGGCCCGGCGCCGTGTCGGGCGGC	GTCGGA	ASTEEAANPSQEKCPI	
		TTCGTTCGGCGATGGACCGGACCAGACTGCGGAAGAAGTTCGCCACAC	TTTTTTT	DRATLHEYFTATSTQ	
		ACGAACGGGAGTGCGTGGCCGAGATCCTGCGACGTGCTTCAACGGAGG	CTTTTTG	RTPFDYDSAKGTEFR	
		AGGCGGCGAACCCCAGTCAGGAGAAATGTCCTATCGACCGCGCCACCC	ACCAAT	TFLEVMSTPSHETSAL	
		TCCACGAGTACTTTACCGCCACCAGCACCCAGCGGACACCATTCGACTAC	CAGAGT	TAEPTLDEIEDQLAH	
		GACAGCGCCAAAGGCACAGAGTTCCGCACGTTCCTGGAGGTGATGTCG	GCGCCA	VKAGSSPGHDGVGY	
		ACGCCTAGCCACGAGACAAGCGCGTTGACCGCAGAGCCTACGCTGGAC	TGCGAC	DVYRRFQVQLLPLLH	
		GAGATCGAAGACCAGCTCGCCCACGTCAAAGCCGGCTCAAGTCCAGGG	TCTTCTG	AAFRFCWRHRRVPA	
		CATGATGGGGTCGGGTACGACGTGTACCGCCGCTTCCAGGTTCAGCTGC	GCCAAT	LWKVGFVRLLHKKG	
		TGCCCCTCCTGCATGCAGCCTTCCGGTTTTGCTGGCGCCACCGCCGCGTG	CAGAGA	DPQQPNNWRPICLQ	
		CCTGCCCTGTGGAAGGTCGGCTTCGTTCGCCTGCTGCACAAGAAAGGTG	CCGGGC	TAIYKLYSGLLARRLSK	
		ACCCGCAGCAGCCAAACAACTGGCGACCAATCTGCCTACAGACCGCCAT	CCTGTCC	FLEANELLPMAQKGF	
		CTACAAGCTCTACAGCGGCCTCCTCGCCCGTCGACTGTCCAAGTTCCTGG	TCGGAC	RAFNGCHEHNFVAT	
		AAGCCAACGAGCTGTTGCCCATGGCACAGAAGGGCTTCCGAGCGTTCA	AGCGAG	TLLDQTRRMHRRLY	
		ACGGTTGCCACGAGCACAACTTCGTGGCAACCACGCTCCTGGACCAGAC	GCCTCC	QVWYDLRNAFGSLP	
		GCGCCGCATGCATCGCAGGCTGTACCAGGTCTGGTACGACCTGCGCAAT	ACGGCC	QQLMWGVLRQLGV	
		GCTTTCGGGTCGCTGCCGCAGCAGCTCATGTGGGGCGTCTTACGCCAGC	AGCCAA	TEEFVARCSGIYEDSY	
		TTGGGGTGACAGAGGAGTTTGTGGCTCGCTGCAGTGGCATCTACGAGG	TCAAGT	FVVGNASDGATEPV	
		ACTCATACTTCGTGGTCGGCAACGCGAGCGACGGAGCGACGGAGCCAG	CTCGGC	RQEVGVYQGCPLSPL	
		TGCGCCAGGAGGTGGGAGTATACCAGGGCTGCCCGCTCAGCCCCCTCCT	AGCGAC	LFITALVPLLRALENQ	
		CTTTATCACTGCGCTCGTGCCACTCCTCCGGGCGCTAGAGAACCAGGAT	GCGTCT	DGVGVPLADGVRPC	
		GGCGTTGGCGTCCCGCTCGCAGACGGAGTACGCCCGTGCGCCACCGCG	TTCTATA	ATAYADDIKVFCDSA	
		TACGCGGACGACATCAAGGTCTTCTGTGACAGCGCGACGGGCATCCAG	GCGCAG	TGIQRCHALVTRFLE	
		CGCTGCCACGCCCTCGTGACGCGCTTCCTGGAGTGGACTGGGCTGCAG	CTGACG	WTGLQANPAKCAFL	
		GCCAACCCGGCCAAGTGCGCGTTCCTGCCGGTGACACGCAGCCAGCAC	AGGCCG	PVTRSQHSNPTRDRD	
		AGCAACCCCACACGCGACCGAGACATCGAGCTTCGCATCCACGGCGAG	ATCTGG	IELRIHGEAIATLGLQE	
		GCCATCGCCACCCTCGGACTTCAAGAAAGCTACCGGTACCTGGGGGGTTG	CGGCCC	SYRYLGVGDGFDHVR	
		GCGACGGCTTCGATCACGTGCGACACCGCCTCCAGCTCGAACCCAAGCT	CCGATT	HRLQLEPKLKQIKREA	
		GAAGCAGATCAAGCGAGAGGCGGTGGCACTGCTGCACTCGGAGCTGGT	GGTCCG	VALLHSELVPWQILK	
		GCCGTGGCAGATTCTCAAGGCTCTGAAGGTGTACATCTACCCGAAGGTA	ACTTTCG	ALKVYIYPKVEYALRH	
		GAGTACGCCCTCAGACACCTGCGGCCGCTCAAGTCACAGCTTCAGGGCT	GCCAAT	LRPLKSQLQGFDSAIV	
		TCGACAGCGCCATCGTGCGCGGCCTGCGCCACCTGCTCCGCCTGCCGGA	CAGCGA	RGLRHLLRLPENSHD	
		GAACAGCCACGACGGCCTCTTCTTCTCACCCACCTCGGCTGGGGGACTG	CGACGA	GLFFSPTSAGGLGLLS	
		GGACTGCTGTCGCTCGTCGAGCTGCACGAAGCGCTGCAGGTGGCACAC	GGGGGC	LVELHEALQVAHAW	
		GCGTGGCAAATGCTCCACTCCAAGGACCCGGCCATCCGTGCCATCGCGC	AGGAGT	QMLHSKDPAIRAIAR	
		GCACGCAGGTGGGACAAGTGGCGCGCAAGCGGTTCAAGCTGGTCGAG	TTACACT	TQVGQVARKRFKLVE	
		GAGCACTGGCGCGGACGGGAAGACGACCTCGCACAGCGCTTCCTCAAC	TTTGCCC	EHWRGREDDLAQRF	
		ACGGAGCTCGCCGCCTCGCCCCACGCCACGGAGACACGCCGCAACGGC	CCGTTTC	LNTELAASPHATETR	

				GACATC	RNGDIGSLWNDVRD
				AACTTCA	TLQTLGLKFAAGDEE
				GGCCAA	EAPGLLQLRVPHHTK
				AATGGC	WLSHSTVLRHVKLH
				GATTTC	MKLRRMDTWKSKVS
				GACCCT	QGTTVREHGGVGSR
				CCACGC	FITAGAGLSDAEYRFA
				GCCGTG	IAPRAHLIDTNSTLKR
				CCACTG	RRLRANDTCRAPGCS
				CTCGGC	YTEPPAHILNKCSPN
				ACCGGC	MDAIRKRHDDALERI
				GGCGAT	ADALRRKVEKSGGRL
				TCAGCG	EVAINKTVPEYDGAA
				GGTGCA	LRPDIVLRNTETKRAII
				ACTTCG	ADLAITHENQPTDAT
				GGCACG	TSSALQQSRDNKITKY
				TGTGCA	QTVAAAMMRAGW
				ACACAT	RVRVTGIVYGSLGSVL
				GCAGCG	PSNFKVYTELLALLKR
				CCCATTG	DARRLNRQLSSHCIR
				CACGCC	ASARIWSAHCRRHRE
				AAGCGG	RQRSGNASRASRGS
				CATCGC	GGAPRRTSQASARR
				GGGACG	(SEQ ID NO: 1448)
				ACGCCT	
				CGGCCG	
				CTCAAG	
				CGCAGC	
				CCCGCC	
				CTTCCAG	
				CACGAC	
				CTCGCG	
				CCGTTT	
				GGCGGA	
				TCGCCAT	
				CAAGAC	
				GTGCGA	
				GAGCCA	

			GGCAAAATATACTTACTCTAAGTATGCCCGAATCCCTGCCCTCTCAGGCT GTACGCGGCCCCATACTTGATCTAAGTATGGGAGGATCCCTGGCCTCTC AGGCTGTACGCGAGACCCGTACGGCCGAATCCCTGGCTTCTCAGCCTGT ACGCGGGGCTATACTTGGTCTAAGTATGCCCCGGTCGCTGGCCTCTGAG CCCGTACACA (SEQ ID NO: 1080)		GGCGGGG GTCGGG CAAAAT ATACTTA CTCTAA GTATGC CCGAAT CCCTGCC CTCTCAG GCTGTA CGCGGC CCCATAC TTGATCT AAGTAT GGGAGG ATCCCTG GCCTCTC AGGCTG TACGCG AGACCC GTACGG CCGAAT CCCTGG CTTCTCA GCCTGT ACGCGG GGCTAT ACTTGG TCTAAGT ATGCCC CGGTCG CTGGCC TCTGAG CCCGTA CACA (SEQ ID NO: 1326)

NeSL	Utopia-4_PR	.	Phytophthora ramorum	CCAGGAATCACCCCCGCCGCCCCCAAGCGCTCCGCCGCCGAGCCCAGCT	CCAGGAAT	TAGACG	MSGDVVSSDGSSRT
				GCGGCTGCTGCCGCCCCACTGGTGACGGCAGTCGCAGCTGCCCCTGCCA	CACCCCCG	GCACAC	TDASGDGDDGAGSS
				CCTCTCCTGTTGGCCGCCGCCGTGGAGCCCCCGCCGCGCCGCCGAGCC	CCGCCCCC	AGGCCC	DAAGDVGVVAMDV
				GCCCCAAGAACAAGCACCCCCAGCTAGCGCGCGCGTGGAGCCCCATGTC	AAGCGCTC	ACAGCG	DQGARRQQPPWQR
				GGGCGACGTGGTCAGCAGCGACGGCAGCAGGACGACGGACGCGA	CGCCGCCG	GCCGAC	VGGKRRRLNDVDDE
				GCGGCGACGGCGACGACGGCGCAGGCAGCGATGCAGCAGGCGAT	AGCCCAGC	AGGGCC	DTRELAELLLEEEDEA
				GTCGGCGTCGTCGCCATGGATGTTGACCAAGGTGCCCCGCCGACAGCAG	TGCGGCTG	ACACCC	GDHAPAPRLSAASAR
				CCGCCATGGCAGCGCGTCGGCGGCAAGCGTCGACGTCTCAACGACGTC	CTGCCGCC	AGGTAG	PASVLSVYAHNAQRF
				GACGACGAAGACACCCGCGAGCTGGCCGAGCTCCTGCTCGAGGAGGA	CCACTGGT	GGAACC	QCTLCTYTAASFASLK
				GGACGAGGCTGGGGACCACGCTCCAGCGCCTAGGCTATCCGCGGCCAG	GACGGCA	GCCTTCA	RHRDSRHRRTAFLDR
				CGCTCGCCCGGCCTCCGTCCTCTCCGTGTACGCGCACAATGCGCAGCGC	GTCGCAGC	AACCCG	FSAGCACGVPFASRL
				TTCCAGTGCACGCTGTGCACGTACACGGCCGCCAGCTTCGCGTCGCTCA	TGCCCCTG	GCCGGT	AAANHAHACDSLNR
				AGCGACACCGGGACTCCCGGCACCGCCGCACCGCCTTCCTCGACCGGTT	CCACCTCT	ACATTAT	TFSVAAAPAAGELSP
				CTCGGCGGGATGTGCGTGCGGCGTGCCCTTCGCGTCGAGGCTGGCTGC	CCTGTTGG	GGTCCG	TAGAANATVKAATV
				TGCGAACCACGCACACGCGTGCGACACGCCTCAACCGCCACCTTCTCGGTA	CCGCGCCG	ACACCT	TPDSPRQDPPKLAAT
				GCGGCTGCGCCGGCAGCAGGGGAATTGAGCCCCACTGCTGGTGCAGCC	CCGTGGAG	ATGAGG	PPLASSALVVDPDHA
				AACGCCACTGTCAAGGCGGCCACTGTCACGCCCGACTCCCCCCGCCAAG	CCCCGCGC	TGCAAC	EQQARERWGPPLPR
				ACCCCCCGAAGCTCGCAGCGACCCCCCCCGCTAGCGAGCTCCGCTCTTGT	GCGCGCCG	CGGTAC	TLVAGRVAARLSEVP
				CGTCGACCCGGACCACGCCGAGCAACAAGCCCGCGAGCGCTGGGGCCC	AGCCGCCC	ACAAGT	APRWGPPLPRGVVA
				GCCCCTACCCAGGACGCTGGTGGCAGGCCGTGTTGCAGCCCGCCTCAGC	CAAGAACA	TACACA	FRIGHRVLPPEMTSD
				GAGGTTCCAGCTCCCCGCTGGGGACCCCCACTGCCACGCGGCGTGGTG	AGCACCCC	CCACAT	EETKDDSSVQDGDR
				GCATTCCGGATCGGCCACCGCGTGCTCCCGCCGGAGATGACGAGCGAC	CAGCTAGC	AGCGAC	QDYPVAAMDVDSG
				GAGGAGACGAAGGACGACTCGAGTGTCCAGGACGGCGACCGCCAGGA	GCGCGCGT	TACCAG	MSGEWLLRFDGACR
				CTACCCGGTGGCCGCGATGGACGTGGACAGCGGCATGAGCGGCGAGT	GGAGCCC	GTATTTA	ANPGPGGAGAALSQ
				GGCTGCTGCGCTTCGACGGCGCCTGCCGCGCCAACCCTGGCCCAGGTG	(SEQ ID	CTACCTG	PDGSVVWTCSHYMP
				GCGCCGGCGCAGCGCTGTCCCAGCCCGACGGCTCCGTTGTATGGACTTG	NO: 1204)	GAAGCC	SSSETNNTAEYTALLL
				CTCCCACTACATGCCGAGCAGCAGCGAAACCAACAACACCGCCGAATAC		AAGGAT	GTRAAADHGTTTLRV
				ACGGCGCTGCTGCTCGGTACGCGAGCCGCCGCCGACCACGGAACCACG		TAACCG	EGDSTLVIQQVRGIF
				ACCCTGCGAGTGGAGGGCGACACGCACCTTGGTGATCCAGCAGGTGCGT		GTCGGT	ATRSVTLRHLRDQVK
				GGCATCTTTGCCACGCGGAGCGTCACTCTCCGCCACCTCCGCGACCAAG		AATACA	LELARVGRFSLHHIDR
				TCAAGCTGGAACTCGCCCGGGTGGGGAGATTTTCGCTCCACCACATTGA		CATAACT	QANAHADRLANRAL
				TCGCCAAGCCAATGCCCACGCCGACCGCCTCGCCAACCGAGCCCTCGAC		TT (SEQ	DLRRTVSECGVHPDG
				CTCCGGCGGACAGTGTCGGAGTGCGGCGTCCACCCCGACGGGAACGGC		ID NO:	NGCTPTAIDDRPLAP
				TGCACGCCCACGGCCATCGACGACCGACCCTAGCGCCGACTCAGCAGC		1327)	TQQPPDAPPPPPAA
				CACCTGACGCCCCGCCACCGCCGCCCGCCGCCGACATTGAGATGGAGG			DIEMEDPDDEDLADI
				ACCCGGACGACGAGGACCTCGCCGACATCGACGACGGCGAGGTCTACG			DDGEVYAAMRVGP
				CGGCCATGCGCGTTGGCCCGAATGCCACTCCGCAGCGACGACGCCGTG			NATPQRRRRGRSGT
				GACGCTCAGGCACCGCGCGAAGAAGCACCGCCGCCAGCGCCCACCCCGGG			AKKHRRQRPPRVTR

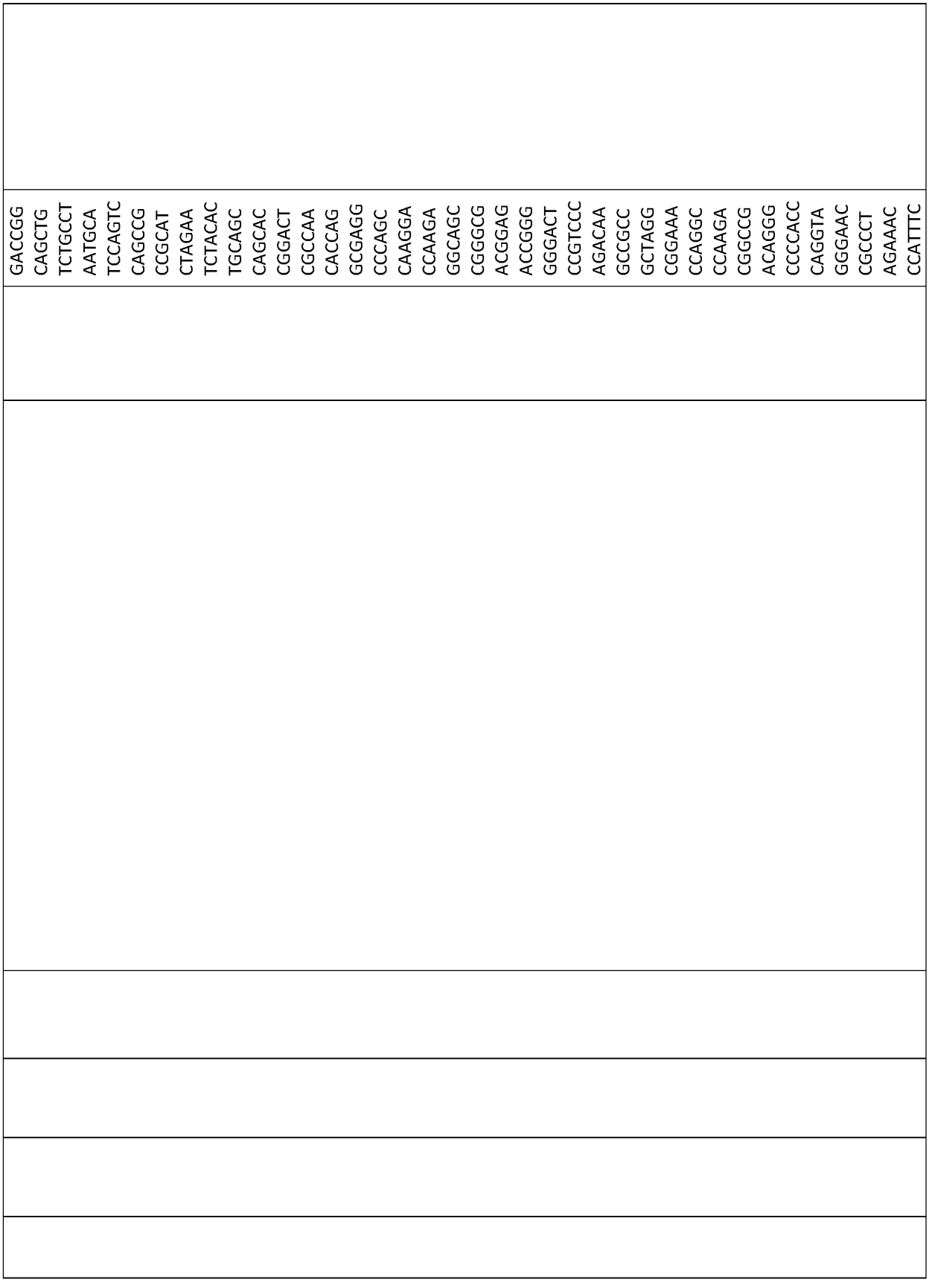
			TGACCCGACACCACCGCGAACACCGCCTCGACGAGGCACTCGATGACCT			HHREHRLDEALDDLH
			CCATGCAGTGGAGCGCTCCACCCCCAGCGACCGCACCACCGTCCGCCGG			AVERSTPSDRTTVRR
			GCACGACCCGCGTTGGACGCGTCAATTCGGCCATCGAACAGCAGCGC			ARRRVGRVNSAIEQQ
			CTCCGCCACCGCTTCGACACGGACGAGAAGGCATGTGTCACCGACATCC			RLRHRFDTDEKACVT
			TGGCGAAGGCGTGCGCGACGCGAGAGGCAGCGAGGACCACGGCCTCC			DILAKACATREAARTT
			GGCGGTGATCCACCTGCAGGCCCGGCGACGCCAGCGGCAGGCAGCGC			ASGGDPPAGPATPA
			AGACGACGGCACCTGCCCCATCTTGGGAGAAGAACTCTGGCGGTTTTTC			AGSADDGTCPILGEE
			GACAGCGTCAACACGCCGCGCCAAGAATTCGCCCCGGACGCCCCGGTG			LWRFFDSVNTPRQEF
			GGTGCCGCCTTCCGCTCGGCGTTGGCACGCCTACCGGCCGCGACCAGTT			APDAPVGAAFRSALA
			GCAAGGAGCTGCTGACTGCGGCGCCGTCGGCCGGCGAAGTGGAAGAC			RLPAATSCKELLTAAP
			CAGCTACAACACGTCCGTGGCGCCTCCAGCCCTGGACTCGACGGCGTCG			SAGEVEDQLQHVRG
			GGTACGACGTGTACCAGCACTTCGCGGCGCAGCTCCTGCCGGCACTGAC			ASSPGLDGVGYDVY
			CGCGGCATTCAAGGCATGCTGGACGGCGAAGAGAGTGCCTCAGAGCTG			QHFAAQLLPALTAAF
			GAAGCTCGGGGTGGTGCGGCTGCTCCACAAGAAGGGCGCGCGCGAGG			KACWTAKRVPQSWK
			ACCCGGCAAACTGGAGGCCGATCTGCCTCCAGCAAGCCATTTACAAGCT			LGVVRLLHKKGARED
			CTACACGGGACTCCTCGCGCGGCGGCTGGTGCGTTGGCTAGACGCCAA			PANWRPICLQQAIYK
			CGACCGCCACGCTCCAGGCCAGAAGGGCTTCCGAGCGGTGAACGGCTG			LYTGLLARRLVRWLD
			CGGGGAGCACAACTTCCTCGCAGCCACGCTCGTCGACCAAGCCCGCCGC			ANDRHAPGQKGFRA
			AAGCGCCGCACCCTCTTCGAAGTCTGGTACGATTTTCGCAACGCGTTCG			VNGCGEHNFLAATLV
			GCAGCGTGCCATTCGCGCTGCTGTGGGACGCGCTGGCGCGGCTCGGCG			DQARRKRRTLFEVW
			TGCCCGACGACTACGTTACCATGTGCAAGGGGCTCTACGAGAGCGCGG			YDFRNAFGSVPFALL
			CGTTTGTCGTCGGGAACGCTATCGACGGCACGACGGACCCAATTGCGCT			WDALARLGVPDDYV
			GCGGGTGGGCGTCTTCCAAGGGTGCCCACTGAGCCCCCAACTCTTCAAC			TMCKGLYESAAFVVG
			GCTGCCATTAGCCCCGCTGCTCTTCGCGTTGCAGCGGCTGCCGGCCACGG			NAIDGTTDPIALRVG
			GCGTGCAGCTGTCGGGCGATGACTGCCCTGGTGCCTCGGCTTATGCCGA			VFQGCPLSPQLFNAA
			CGACCTCAAGATTTTCAGTGGCACCGAGGACGGCATCAAGCGACAGCA			ISPLLFALQRLPATGV
			CGCGCTCGTGGCCGACTTCTTGCGCTGGACCGGCATGGCGGCGAACCCC			QLSGDDCPGASAYA
			AACAAGTGCTGCACCATGTCGGTCCAGCGCGACGGACGCGGCGTGCTC			DDLKIFSGTEDGIKRQ
			AAGACTGACGACCTCCAGCTAGACCTGGCTGGCACGCCGATCCCCGCGC			HALVADFLRWTGMA
			TGTCCATGAGCGCTTCGTACACCTACCTGGGGATCGGCGATGGATTCGA			ANPNKCCTMSVQRD
			TCACGTGCGCCGCCGCGTCGAGCTCGCTCCAGCACTCAAGCAGCTTAAG			GRGVLKTDDLQLDLA
			GACGACGCGACGACGCTGCTGCAGTCGGGCCTCGCCCCGTGGCAGGTG			GTPIPALSMSASYTYL
			GTGAAGGCGGTGAAGACGTACCTCTACCCCCGGGTGGAGTACGCTCTC			GIGDGFDHVRRRVEL
			CGCCACCTTCGCCCCTTCCAACAGCAGCTTGAGGGCTTCGACCGGCACC			APALKQLKDDATTLL
			TCGCGCGCGGCCTCCGCCACCTCCTGCGACTGCCGGGCAACGCCACGGC			QSGLAPWQVVKAVK
			GGAGTGCTTCTACGCTCCGGTGTCGCGCGGCGGGCTGGGCCTGCTGCC			TYLYPRVEYALRHLRP
			ACTGACGGAGCTGCACGCGGCGCTCCAAGTCGCGCACGGATGGCAGCT			FQQQLEGFDRHLAR
			GCTGAATTCGAAGGACCCTGCCATCCGGCGCATCGCCGCGCGTGCAGCTC			GLRHLLRLPGNATAE

NeSL	Utopia-5_PI	.	Phytophthora infestans	CTCAAGCCTAGCAGCGGCTACCGCAGCTACTCCAGCTCCGGTAGCTGCT	CTCAAGCC	TGACGC	CFYAPVSRGGLGLLPL
				ACTGCTACAACTGCTCCGCGCTGCTGCTCGCCGTGGAGCACCACG	TAGCAGCG	ACCGTG	TELHAALQVAHGWQ
				CGCACGCTGAACCGAACCAAGAACATCTACCGATGGCTCGCGTGCTCGT	GCTACCGC	ACATAG	LLNSKDPAIRRIARVQ
				GGAGCCCATGCAGGTCGACGAGTGCAGCAGCTGCGACCGCAGCACTTT	AGCTACTC	TGCGGC	LRQIADARHRIDSRA
				GACAGCCGACGACGGGAGCGGTGACGACGTTGCAGCCCCAAGCTCCTT	CAGCTCCG	ACGTGA	WEGRDEELCELLLNS
				GAACAGCAACGACGTCGCCGCCCCCATGGACGTTGACTCCGGCACTGAT	GTAGCTGC	AGATGC	QLGTSPDAPPKRRN
				TGCCCTCCAGCGTTGCAACAGCCGCCCCAGCGACCGCGTGCGCTGCACG	TACTGCTA	ACATGA	GDIGSLWVDVQRHL
				CGGCAGATCGCCGACGCGCGCCACCGTATCGACAGCCGCGCATGGGAG			RTLGLKFATAPACAD
				GGGCGCGACGAGGAGTTGTGCGAGCTGCTGCTCAACTCGCAGCTCGGG			AGSAATTLQLRVPHH
				ACGTCGCCCGACGCGCCACCGAAGCGTCGGAACGGCGACATCGGTTCT			DKWLDHRTVLRHVK
				CTTTGGGTCGACGTCCAGCGCCACCTCCGCACGCTCGGCCTCAAGTTCG			LHVKHRHWSKWAA
				CGACCGCGCCGGCCTGCGCGGACGCCGGCTCTGCGGCGACCACGTTGC			MRDQGKTARAHGG
				AGCTCCGCGTGCCGCACCACGATAAGTGGTTGGACCATCGCACTGTCCT			AGSGFLTRPRGMWE
				GCGGCACGTGAAACTGCACGTGAAGCACCGGCACTGGTCCAAGTGGGC			ADYRFAVAARLNQL
				GGCCATGCGCGACCAAGGCAAGACGGCCCGCCCATGGCGGCGCGG			DTHSVLKRRRLRAHD
				GGAGCGGCTTCCTCACCCGGCCACGAGGCATGTGGGAAGCCGACTACC			HCRQPGCSRAETLAH
				GCTTCGCGGTGGCCGCCCGCCTGAACCAGCTCGACACGCACAGCGTGCT			VLNHCAGTMDAVRG
				GAAGCGCCGGCGTCTCCGAGCCCACGACCACTGCAGGCAGCCGGGCTG			RHDDALKHIERALHA
				CTCCAGGGCGGAGACGCTGGCGCACGTGCTCAACCACTGCGCCGGCAC			SSPGGQDRVELRVN
				CATGGACGCGGTCCGCGGCCGCCACGACGACCCTCAAGCACACATCGA			QTVPSLAGPALRPDL
				GCGCGCGCTGCACGCGTCGTCGCCGGGCGGCCAGGACCGCGTCGAGCT			QLYNHTKKMVAVVD
				CCGGGTGAACCAGACGGTGCCGTCGCTCGCCGGCCCCGCGCTACGGCC			LAVAFEEQASDDPES
				CGACCTCCAGCTGTACAACCACACCACAAGAAGATGGTGGCGGTGGTCGA			SALARIAAHKRAKYA
				CCTGGCCGTGGCGTTCGAGGAGCAGGCGAGCGACGACCCGGAGAGCT			GVKRHLERQGWKVH
				CGGCGCTGGCACGCATCGCCGCACACAAGCGCGCGAAGTATGCCGGCG			LSALVYGSLGAVPAG
				TCAAGCGACACCTCGAGCGCCAAGGGTGGAAGGTCCACCTCTCGGCGC			NHKVLTEHLGLLKRD
				TCGTGTACGGCTCGCTCGGCGCGGTGCCGGCGGCAACCACAAGGTGC			AKRLDRQLSVACIQS
				TTACTGAGCACCTTGGGTTGCTCAAGCGTGATGCGAAGCGGCTGGACC			SRRIWNLHCSQHRA
				GGCAGCTCTCGGTGGCGTCATCCAGTCCAGCCGCCGCATCTGGAACCT			RQHQAPGGSRAAET
				GCACTGCAGCCAGCACCGCGCGCGCCAGCACCAAGCACCAGGGGGAAG			GGTPPRTGRR (SEQ
				TCGGGCGGCGGAGACCGGGGGGACTCCGCCGCGCACCGGCCGCCGCT			ID NO: 1449)
				AGACGGCACACAGGCCCACAGCGGCCGACAGGGCCACACCCAGGTAG			
				GGAACCGCCTTCAAACCCGGCCGGTACATTATGGTCCGACACCTATGAG			
				GTGCAACCGGTACACAAGTTACACACCACATAGCGACTACCAGGTATTT			
				ACTACCTGGAAGCCAAGGATTAACCGGTCGGTAATACACATAACTTT			
				(SEQ ID NO: 1081)			
							MARVLVEPMQVDEC
							SSCDRSTLTADDGSG
							DDVAAPSSLNSNDV
							AAPMDVDSGTDCPP
							ALQQPPQRPRALHV
							GSKRRRLDADDEEEA
							RQLQEEEEAGIHAPA

	TAGGCAGCAAGCGCCGCCGACTCGACGCAGATGATGAGGAAGAAGCTC	CAACTGCT	AGCTCC	LRLSAASAQPASVLA
	GCCAGCTGCAAGAAGAGGAGGAGGCTGGCATCCATGCCCCAGCGCTCA	CGCGCTGC	GACACT	VYTHNASRFDCTLYA
	GGCTATCCGCGGCCAGCGCTCAACCGGCCTCCGTCCTCGCCGTGTACAC	TGCTCGCG	GGGCCA	YTAGSFASLKTHRNS
	GCACAACGCGTCGCGGTTCGACTGCACGCTGTACGCGTACACGGCCGG	TCGCCGTG	AGTGGG	RHRRTAFLDRFSAGC
	CAGCTTCGCATCACTGAAAACGCACAGGAACTCTCGGCACCGCCGCACC	GAGCACCA	CGGCCA	ACGVPFASRLAAARH
	GCATTCCTCGACCGATTCTCGGCGGGTTGCGCGTGTGGCGTGCCCTTCG	CGCGCACG	TGCGCG	AQACASLSSAPLAEA
	CCTCGAGGCTGGCTGCAGCAAGACACGCGCAAGCGTGCGCCAGCCTCT	CTGAACCG	ACCAAG	SSAAGASSHTVDEAD
	CGAGCGCACCATTGGCCGAGGCTTCGTCGGCAGCAGGAGCATCAAGCC	AACCAAGA	GCAAGA	STVSAAGHTEPDLPR
	ACACTGTCGACGAAGCCGACTCCACCGTCTCAGCCGCGGGTCACACCGA	ACATCTAC	CAGCTC	HNATELTASPPLVSSP
	GCCCGACCTGCCCCCGCCACAACGCCACGGAGCTCACAGCGTCCCCCCG	CG (SEQ ID	GTGCAC	DVEVQAPETEATENR
	CTGGTGAGCTCCCCCGATGTTGAAGTGCAGGCACCAGAGACAGAGGCC	NO: 1205)	ATGGTG	WGTPLPRVLVASRIA
	ACCGAGAACCGCTGGGGCACTCCACTTCCGCGCGTGCTGGTCGCGTCTA		GGGTTG	GRLAQVPPPRWGPP
	GGATCGCCGGGCGCCTCGCTCAGGTGCCACCCCCACGCTGGGGCCCCCC		GTAGTG	LPRTTIAGRIATRLAA
	GCTACCGCGCACTACCATCGCGGGCCGCATTGCGACGAGGCTCGCCGCC		GCTTCCT	TPAPRWDPPLPRSLV
	ACACCTGCCCCCCGCTGGGACCCACCGCTGCCACGAAGCCTGGTTGTGT		CACACG	VSRIAARLLPALPDAP
	CACGCATCGCCGCCCCGCCTGCTACCCGCGCTACCAGACGCGCCTGCATG		GCCGCG	ACEEEAKDSDTMDW
	CGAGGAAGAGGCCAAGGACACGACACGATGGACTGGGCGCCAACTT		AGGCCT	APTWTNEETKDSEPH
	GGACAAACGAGGAGACCAAGGACAGCGAGCCGCATGACGAAGCGCCA		GTGGGA	DEAPGQVDEETIDDA
	GGGCAGGTCGACGAGGAGACCATCGACGACGCCGACGGCGAGTGGCT		AGCCGA	DGEWLLRFDGACRA
	TTTGCGCTTCGACGGTGCTTGCCGTGCAAACCCAGGCCCAGGCGGCGCC		CTACCG	NPGPGGAGAALFKP
	GGCGCAGCTCTCTTCAAGCCCAGCGGCCCCGTAGTGTGGACTTGTTCCC		GTTCGC	SGPVVWTCSHYDPS
	ACTACGACCCGAGCACCACGGCAACGAACAACACAGCAGAGTACACTG		GGTGGC	TTATNNTAEYTALLL
	CGTTGCTGCTCGGGGCCAGGGCTGCCGCCGACCACGGCGTCACGAAGC		CGGCCG	GARAAADHGVTKLR
	TGCGCGTCGAGGGCGACAGCACGCTGGTCATCCAGCAAGTGCGGGGCA		CTTAAAC	VEGDSTLVIQQVRGIF
	TCTTCGCGACGAGAAGCACGCGGCTCCGAGCACTGCGCAACAAGGTCA		CAGGTA	ATRSTRLRALRNKVKL
	AGCTCGAGCTCGCGCGCGTGGGGAGCTTCTCGCTCCACCACATCGACCG		GACACG	ELARVGSFSLHHIDR
	GCAGGCGAACGGCCACGCCGACCGCCTCGCCAACGCCGGCCTCGACCG		CACAGT	QANGHADRLANAGL
	ACGCCGCACCAAGCTCGAGTGCAGTGTGCACCCCGATGGCAGGGGCTG		GTCCTCA	DRRRTKLECSVHPDG
	CACCAACACCTCTGTGGCGACCGCAGCGCCTACTGCTCCTGCAGCGCCC		AGCGCC	RGCTNTSVATAAPTA
	TTGCCCCCAGCCCGGCCTCCTGCAACTACAGCCGCTCCGTCTCATGATGA		GGCGCC	PAAPLPPARPPATTA
	CGACCACAGCGTTCAAGGCGACATCGACGATGGCGAAGTATACGCTGC		TCCGAG	APSHDDDHSVQGDI
	AATGTGCATCGGCCCCGATGCGGTGCCGCACCGCCGACCGCGCCTCCGC		CACATG	DDGEVYAAMCIGPD
	CTACGGCACTTGACGGACGAGGAGTCCGAGGAGGCTGGCGACGTCGTG		ACAGGT	AVPHRRPRLRLRHLT
	GAGAGGCTGGCAGCGTCGCTGGCGGCGAAGATTGCAGACGCGCCCGA		GCAGAC	DEESEEAGDVVERLA
	CTGGGAGACGGCGGAGGGCTACATCACAGCCCTACCGTACGCTCTCTAC		ACCCAG	ASLAAKIADAPDWET
	GACAAGCTGCAGCCGTACTCCCAGGCCCAGCCTCAGCCTCCAAGCCAAC		GATGCA	AEGYITALPYALYDKL
	AGCAGCAGCAGCGGAGAGGCAAGCGCGGCGGTGGCAGCCAACGCAG		CGCGCT	QPYSQAQPQPPSQQ
	CAGCGTCGACAGCGGAGAGGCAAGCGCGGCGGTGGCAGCCAACGCAG		CCGAGA	QQQQQQRPRQQQ

		GCAGCGGAAGACGCGCCGCCGACGGCCCCCGCGGGTGACACGCCACCA		CGCTGG	QTRQRRQRRGKRGG
		CCGCGAGCACCGGATCGATGAAGCCCTCGACGACCTCCACGCGATCGA		CGCATG	GSQRRQRKTRRRRRPP
		GAGCAGGCGCCCTCAAGACAGGACGGCCATCAGCAAGGCCAGGCGTCG		TGCTTAA	RVTRHHREHRIDEAL
		AGTCGGCCGCATCCGCTCCGCCCTTGACCAGCACCAATTGCGCCACCGC		CCACTG	DDLHAIESRRPQDRT
		TTCGACACCGATGAGAAGGTGTGCGTGGACGCCATCCTGGCTGGGGCG		CGACGG	AISKARRRVGRIRSAL
		CGAGCGAGCCAAGGAGCGACAACGGCACCACCGTCGGCAACCACGGAC		AACCAT	DQHQLRHRFDTDEK
		CCCCCAGCTCCGATGGACGACAGCAGGTGCCCCATTCCCGGCGACGATC		GGACGC	VCVDAILAGARASQG
		TATGGAGGTTTTTCGACAGCGTTAACACGCCACGGCGGTCGTTCGACGC		AGTCCG	ATTAPPSATTDPPAP
		AGAGGCACCAGACGGAGCTGCCTTCCGCGAAGCGATGGCGTGCTTGCC		TGGCCG	MDDSRCPIPGDDLW
		CGCCGCCACGAGAGCCCAGGAACTGCTGACGGAGGCGCCTACGGTGGA		GCCATG	RFFDSVNTPRRSFDA
		CGAGGTTGAAGACCAGATTCAGCACGCCCGCGCCAGCTCCAGCCCCGG		ACGCCG	EAPDGAAFREAMAC
		CCTCGACGGCGTTGGTTACGACATCTACAAGCAGTTCGCCGCCCACCTC		CACTCA	LPAATRAQELLTEAPT
		CTGCCAGCACTGCACACAGCGTTCGTGTGCTGGAACCACAAGCGG		AGATTA	VDEVEDQIQHARASS
		GTGCCTCAGAGCTGGAAGCTCGGCGTGGTGCGCCTTCTCCACAAGAAA		TTGAGC	SPGLDGVGYDIYKQF
		GGCGATCGACAGGATCCTGCCAACTGGCGGCCGATCTGCTTGCAGCAG		GTGCGC	AAHLLPALHTAFVCC
		ACCATCTACAAGCTCTATGCCGGCATCCTGTCACGCCGGTTCGTGCGCTG		TCCTCGC	WNHKRVPQSWKLG
		GCTGGACGCAAACGCTCGTCACGCAGAGGCGCAGAAAGGATTCCGGGC		ATCGTC	VVRLLHKKGDRQDP
		GATGAACGGCTGCGGGGAGCACAACTTCCTCGCGGCCACGCTCGTCGA		GGCCGA	ANWRPICLQQTIYKL
		CCACGCCCGCCGCAAGGAGCTGCACGTCGTGTGGTACGACCT		CCAGCA	YAGILSRRFVRWLDA
		TGCCAACGCCTTCGGCAGTGTGCCACACGACTTGCTATGGGAAACGCTG		GGACCG	NARHAEAQKGFRAM
		GCGCGCCAAGGCGTACCGCCCACGTTCGTGGACTGCTGCCGCGGGATC		TGCTGA	NGCGEHNFLAATLV
		TACAGTGATGCAGCCTTCACCATCGGGAACGCCGCCGACGGCACCACA		GCTCCG	DHARRKRKELHVVW
		GCGCCCATCAGGCTTCGCGTGGGAGTATTCCAGGGCTGCCCGCTCAGCC		CGTGAA	YDLANAFGSVPHDLL
		CCCACCTCTTCACCGCCGCAATCGCCCCGCTGCTCCATGCGCTGAAGAG		CCAGAC	WETLARQGVPPTFV
		GCTGCCGGTCACTGGAGTGCAACTCACCGGCGTGGACCGTCCCGGAGC		CGTGCC	DCCRGIYSDAAFTIG
		TGCTGCCTACGCGGACGACCTCAAGACGTTCAGTAGCTCCGTGGACGGA		GTCACTC	NAADGTTAPIRLRVG
		ATCAAGAGGCAGCACGAGCTCGTGGCGACCTTCCTACGATGGACCGGC		GCCGGC	VFQGCPLSPHLFTAAI
		ATGGCGGCGAACCTTTCCAAGTGCAGCGCCATGTCGGTCCAGCGCGAC		CCCGCG	APLLHALKRLPVTGV
		AGCCGGGGCGTGCTCAAGACAGGTGACCTATGCCTCAAGCTCAACGAC		CTACGG	QLTGVDRPGAAAYA
		GCAGAGATCCCAGCGCTGAGCATGACGGCCTCCTACGCGTACCTCGGCA		CCCGAC	DDLKTFSSSVDGIKR
		TCGGGGACGGCTTCGACCACGTGCGCCGCCTCGAGCTCGCGCCAAT		CTTCAGC	QHELVATFLRWTGM
		GATGAAGCAGCTCAAGCACGACGACGGCGCTGATGCAGTCGGGGCT		TCTACAA	AANLSKCSAMSVQR
		GGCCTCATGGCAAGTGGTGAAGGCGGTGAAGGTATACCTCTACCCCCG		CCACAC	DSRGVLKTGDLCLKL
		CATAGAGTACGCCCTTCGCCACCTTCGCCCGTTCAAACAGCAATTAGAA		CAAGAA	NDAEIPALSMTASYA
		GCTTTCGACGAGCACCTCCGCCGCGGGCTCCGCCACCTGCTGCGACTAC		GACGGT	YLGIGDGFDHVRRRL
		CGACGAACGCCACCAGCGCCTTCTTCTCAGCGCCGACATCGCGCGGCGG		GGCGGT	ELAPMMKQLKHDAT
		GCTGGGCCTTCTGCCGCTGACGGAGCTGCACGCGGCGCTTCAGATCGCC		GGTCGA	ALMQSGLASWQVVK
		CATGGGTGGCAGATACTCAACTCCCCAGACGGCGCCACCCAGAGGATT		CCTGGC	AVKVYLYPRIEYALRH

		GCCCGCGAGCAGCTCCGGGAGATTCCAGACGCGCGCCACCGCCTGGAC		CGTGGC	LRPFKQQLEAFDEHL
		ACTGCCCACTGGAGGAACCGCGACGCAGAGCTGTGCGAGCTGCTGCTC		GTTGAG	RRGLRHLLRLPTNATS
		AACTCGCAGCTCGGACAGTCCAGCCACGCCCCACCCAAGCGGCGCAACT		GAGCAG	AFFSAPTSRGGLGLLP
		GCGACATCGGATCCCTTTGGATCGACATTCGGCGCCAGCTCGGCACGCT		GCGAGT	LTELHAALQIAHGW
		CGGGCTCAAATTCGAGACCGCGCCTGGCCGACGCAGTCACCAACCAGC		GACGAC	QILNSPDGATQRIAR
		CCGCCCAGCCATTGCAGCTTTCGCGTGCCGCACCACGACAAGTGGCTGA		GCGAGT	EQLREIPDARHRLDT
		CGCACCGTGACATAGTGCGGCACGTGAAGATGCACATGAAGCTCCGAC		AGCTCG	AHWRNRDAELCELLL
		ACTGGGCCAAGTGGGCGGCCATGCGCGACCAAGGCAAGACAGCTCGTG		GCACTG	NSQLGQSSHAPPKRR
		CACATGGTGGGGTTGGTAGTGGCTTCCTCACACGGCCGCGAGGCCTGT		TCCCGG	NCDIGSLWIDIRRQL
		GGGAAGCCGACTACCGGTTCGCGGTGGCCGGCGCCTTAAACCAGGTAG		ATCGCC	GTLGLKFETAPGRRS
		ACACGCACAGTGTCCTCAAGCGCCGGCGCCTCCGAGCACATGACAGGT		AACCAC	HQPARPAIAAFACRT
		GCAGACACCCAGGATGCACGCGCTCCGAGACGCTGGCGCATGTGCTTA		AAGCGA	TTSG (SEQ ID NO:
		ACCACTGCGACGGAACCATGGACGCAGTCCGTGGCCGGCCATGACGCC		GCCAAG	1450)
		GCACTCAAGATTATTGAGCGTGCGCTCCTCGCATCGTCGGCCGACCAGC		TACGAC	
		AGGACCGTGCTGAGCTCCGCGTGAACCAGACCGTGCCGTCACTCGCCG		CACATC	
		GCCCCGCGCTACGGCCCGACCTTCAGCTCTACAACCACACCAAGAAGAC		AAGCTA	
		GGTGGCGGTGGTCGACCTGGCCGTGGCGTTGAGGAGCAGGCGAGTGA		CACCTC	
		CGACGCGAGTAGCTCGGCACTGTCCCGGATCGCCAACCACAAGCGAGC		GAGCGC	
		CAAGTACGACCACATCAAGCTACACCTCGAGCGCCAAGGATGGAAGGT		CAAGGA	
		ACACCTCTCGGCACTCGTGTACGGGTCGCTTGGGGCGGTCGCTAGTGGC		TGGAAG	
		AACTACCAGGTGTACACCACACACCTGGGGCTACTCAAGCGCGATGCAA		GTACAC	
		AGCGGCTGGACCGGCAGCTGTCTGCCTAATGCATCCAGTCCAGCCGCCG		CTCTCG	
		CATCTAGAATCTACACTGCAGCCAGCACCGGACTCGCCAACACCAGGCG		GCACTC	
		AGGCCCAGCCAAGGACCAAGAGGCAGCCGGGCGACGGAGACCGGGGGG		GTGTAC	
		GACTCCGTCCCAGACAAGCCGCCGCTAGGCGGAAACCAGGCCCAAGAC		GGGTCG	
		GGCCGACAGGGCCCCACCCAGGTAGGGAACCGCCCTAGAAACCCATTT		CTTGGG	
		CGGTGGTCGACTCGCAAGCCTTACCTATATTTTAGACGTAGCGACCAAA		GCGGTC	
		TTACAAATTTGGTAACGAGTAAGCCAAATGGTAATACACAAAACTTT		GCTAGT	
		(SEQ ID NO: 1082)		GGCAAC	
				TACCAG	
				GTGTAC	
				ACCACA	
				CACCTG	
				GGGCTA	
				CTCAAG	
				CGCGAT	
				GCAAAG	
				CGGCTG	


GACCGG
CAGCTG
TCTGCCT
AATGCA
TCCAGTC
CAGCCG
CCGCAT
CTAGAA
TCTACAC
TGCAGC
CAGCAC
CGGACT
CGCCAA
CACCAG
GCGAGG
CCCAGC
CAAGGA
CCAAGA
GGCAGC
CGGGCG
ACGGAG
ACCGGG
GGGACT
CCGTCCC
AGACAA
GCCGCC
GCTAGG
CGGAAA
CCAGGC
CCAAGA
CGGCCG
ACAGGG
CCCCACC
CAGGTA
GGGAAC
CGCCCT
AGAAAC
CCATTTC

						GGTGGT CGACTC GCAAGC CTTACCT ATATTTT AGACGT AGCGAC CAAATT ACAAAT TTGGTA ACGAGT AAGCCA AATGGT AATACA CAAAAC TTT (SEQ ID NO: 1328)	
NeS L	Utopia -5_PR	.	Phytop hthora ramoru m	CCAGGCATCACCCCCGCCGCCCCCAAGCGCTCCGCCGCCGAGCGCAGCT GCGGCTGCTGCCGCCCCACTGGTGACGGCAGTCGCAGCTGCCCCTGCCA CCTCTCCTGTTGGCCGCGCCGCCGTGGAGCCCCGCGCGCGCGCCGAGCC GCCCCAAGAACAAGCACCCCCAGCTAGCGCGCGCGCGGAGCCCATGTC GGGCGACGTGGTCAGCAGCGACGGCAGCAGCAGGACGACGGACGTGA GCGGCGACGGCGACGACGGCGCCGACGGCGCAGGCAGCAGCGATGCA GCAGGCGATGTCGGCGTCGTCGCCATGGATGTTGACCAAGGTGCCCGC CGACAGCGGCCGCCATGGCAGCGCGTCGGCGGCAAGCGTCGACGTCTC AACGACGTCGACGACGAAGACACCCGCGAGCTGGCAGAGCTCCTGCTC GAGGAGGAGGACGAGGTTGGGGCCCAAGCACCAGCGCTCAGGCTATT CGCGGCCAGCGCTCACCCGGCCTCCGTCCTCTCCGTGTACGCACACAAT GCGCAGCGCTTCGTGTGCACGCTGTGCGCGTACACGGCCGCCAGCTTCG CGTCGCTCAAGCGACACCGGGACTCACGACACCGGCGCGTGTCCTTCGT GGACAAGTTCTCGGCGGGCTGCGCGTGCGGCACGCCCTTCGGCTCGAG GCTGGCAGCGGCGAGACATGCGCAAGCGTGCGCCAGCCTCTCGATCCC ACGCACGGTAACAGCTCCGGCGGCAGCAGGAGATTTGAGCCCCACTGC CACGGGAGCCAACGCCACCGCCTCGGCGGCCGCAACTTCGCCCGATTTG CCCCGCCCAGCTTCCCCGGAGCTCGCTGCGTCCCCCCCGCAGACGAGCC CCTTTGACGTTGCCATACAGGCAGACGCTGCAGAGCAGACAGCGTGGA CGCGCTGGGACCCTCCGCTCACCCGCGCTGCGGTGGCAGCCCGTGTCGC	CCAGGCAT CACCCCCG CCGCCCCC AAGCGCTC CGCCGCCG AGCGCAGC TGCGGCTG CTGCCGCC CCACTGGT GACGGCA GTCGCAGC TGCCCCTG CCACCTCT CCTGTTGG CCGCGCCG CCGTGGAG CCCCGCGC GCGCGCCG AGCCGCCC CAAGAACA	TGATGC CCGCCA ACTACA AGGTGC TTACTGA GCACCTT GGGCTG CTCAAG CGTGAT GCGAAG CGGCTG GACCGG CAGCTG TCGGTG GCGTGC ATCCAG TCCAGC CGCCGC ATCTGG AACCTG	MSGDVVSSDGSSRT TDVSGDGDDGADGA GSSDAAGDVGVVA MDVDQGARRQRPP WQRVGGKRRRLND VDDEDTRELAELLLEE EDEVGAQAPALRLFA ASAHPASVLSVYAHN AQRFVCTLCAYTAAS FASLKRHRDSRHRRV SFVDKFSAGCACGTP FGSRLAAARHAQAC ASLSIPRTVTAPAAAG DLSPTATGANATASA AATSPDLPRPASPELA ASPPQTSPFDVAIQA DAAEQTAWTRWDP PLTRAAVAARVASRL AVVPAPRWGPPLSR TLVASRIAARLDAQTS

		CTCCCGCCTCGCCGTGGTTCCCGCGCCGCGATGGGGGCCGCCGCTGTCA	AGCACCCCC	CACTGC	RWGPPLPPAMVASR
		AGGACGCTGGTGGCCTCCAGGATTGCAGCACGCCTGGACGCGCAGACG	CAGCTAGC	GCGCAG	IASRLAAMPAPRWG
		TCACGATGGGGACCGCCCCTGCCACCGGCGATGGTGGCGTCCCGGATT	GCGCGCGC	CATCGA	PPLLRTVIASRIADRLL
		GCGTCGAGACTCGCCGCCATGCCGGCACCCCGCTGGGGCCCGCCGCTG	GGAGCCC	GCACGG	PPELAADEETKDDDV
		CTGCGGACCGTGATCGCATCCCGCATCGCCGACCGACTGCTCCCACCGG	(SEQ ID NO: 1206)	CAGCAC	HMDNAASVDVDEES
		AGTTGGCGGCCGACGAGGAGACGAAGGATGACGACGTCCACATGGAC		CAGGGC	EVADVVMTDHDGE
		AACGCGGCCAGCGTGGACGTCGACGAGGAGTCCGAGGTCGCCGACGT		CAAGCG	WLLRFDGACRANPG
		GGTCATGACCGACCACGACGGCGAATGGCTGTTGCGCTTCGACGGCGC		CCAAGG	PGGAGAALFKPSGPV
		CTGTCGTGCCAACCCTGGCCCAGGCGGCGCCGGAGCCGCTCTGTTCAAG		GGCAGT	VWACSRYMPSSSAT
		CCCAGCGGCCCCCGTAGTGTGGGCATGCTCGCGCTACATGCCGAGCAGC		CGGGCG	NNTAEYTALLLGARA
		AGCGCGACCAACAACACGGCCGAGTACACGGCGCTGCTGCTCGGCGCT		GCGGAG	AADHGATHLRVEGD
		AGGGCCGCTGCCGACCATGGAGCCACGCACCTGCGAGTGGAAGGCGAC		ACCGGG	STLVIQQVRGIFAARS
		AGTACGCTTGTGATCCAGCAGGTGCGTGGCATCTTCGCAGCGCGCAGCA		GGGACT	TRMRALRNQVQSEL
		CCCGGATGAGGGCGCTGCGCAACCAGGTGCAGTCCGAGCTCGCGCGGG		CCGCCA	ARVGSFSLHHIDRQD
		TGGGCAGCTTCTCGCTCCACCACATCGACAGGCAAGACAACGCACACGC		CAGAGC	NAHADRLANRALDL
		GGACCGCCTTGCCAATCGCGCGCTGGACCTGCGCCGGACCGTGATAGA		GGCCGC	RRTVIECGIHCDGVG
		GTGCGGCATCCACTGTGATGGTGTTGGCTGCACCGCCACCACGACGGA		CGCTAA	CTATTTEVQSSSAPEI
		GGTACAGTCATCATCAGCGCCCGAGATACCGACCCGGCCGGTTGCTGAC		GCGGAC	PTRPVADDHDEHEV
		GACCATGATGAACATGAAGTTGTTGATGTTGTTGATGTTTGTGGAGTTT		ATCGGG	VDVVDVCGVCGVCG
		GTGGAGTTTGTGGCGACCGTGGAACTTGTGGAGTCTGTGACGTTAGTG		CCCGTA	DRGTCGVCDVSGDI
		GAGATATCGATGATGGCGAGGTCTACGCAGCCATGCGTACGGGACCTG		GCGGCC	DDGEVYAAMRTGPD
		ACGCTGTTCCTGCGCGCCGACCACGCCTCCGCTTGCGGAAACTCACTGA		GACAGG	AVPARRPRLRLRKLT
		CGAAGAGCAAGAAGAGGCAGGCACGCTGGCGGAACGGCTTGGAGCGA		GCCACA	DEEQEEAGTLAERLG
		CGCTGGCGGCCAAGATTGCAGACGCACGCGACTGGGAGTCGGCGGAG		CCCATGT	ATLAAKIADARDWES
		GGATACATCACAGCCCTCCCTTATCTACTCTACGACAAATTGTTGCCGTA		AGGGAA	AEGYITALPYLLYDKLL
		CTCCCAGGGCCCAGCTCGCTCCTTGCCGGTCCGCCAGCACCAACGGCAG		CCGCCCT	PYSQGPARSLPVRQH
		CAGCAGCAGCCGGACGGACAATTCCAGCGACCCACCCAGTCCCGATCTG		CTAAAC	QRQQQQPDGQFQR
		CGGCTCGTCGACAGCGACGCCAGCGCCACCGGGCACGTCGCCGACCGC		CCGCCC	PTQSRSAARRQRRQ
		CACGCGTGACTCGTCACCACCGTGAACACCGACTAGACGAAGCGCTTGA		GGTACA	RHRARRRPPRVTRH
		CGACCTCCACGCCGGTGGAGAGAGCTACGCCCAGCGACCGCCGCTCCATT		TTATGGT	HREHRLDEALDDLHA
		CGCCGAGCCCGCCGCCGGGTCGGCCGCGTTAATTCGGCCGTCGAACAA		CCGACA	VERATPSDRRSIRRAR
		CAACGGCTCCGGCACCACTTCGATACGAACGAGAAGGGCTGCGTGGAG		CCTATG	RRVGRVNSAVEQQR
		ATCCTACTGGCCAAGGCGCGTGCACAGAGGTCGACGACGGTGGCCCGG		AGGTGC	LRHHFDTNEKGCVEI
		ACGGCAGTTGGAGAGCCGAACAGCGGTGCGGCCGAGGACGACGGCAC		AACCGG	LLAKARAQRSTTVAR
		GTGCCCCATCCCGAGCGAGCGCCTTCACCGCCATTTCACCGAGGTGAAC		TACACA	TAVGEPNSGAAEDD
		ACGCCCGGGTCCAGCTTTGATGCGATGGCCCCCGTGGGTGCACCGTTCC		AGTTAC	GTCPIPSERLHRHFTE
		GTGCGGCACTTGCGCACCTCCCCGCGGCCACAGAGGCCTCGGAGCTGCT		ACACCA	VNTPGSSFDAMAPV
		GACAGAGGCGCCGACGCCCGATGAGATCGAGGATCAGCTGCAGCGAG		CATAGC	GAPFRAALAHLPAAT

	CCAAGGGAACGACCAGCCCTGGCCTGGACGGAGTGGGCTACGACGTCT	GACTAC		EASELLTEAPTPDEIE
	ACAAGGCCTTCTCCACCCAGTTACTGCCGGTGCTGCACGCGGCATTCCA	CAGGTA		DQLQRAKGTTSPGLD
	GTGCTGCTGGCAGCACCACCGTGTGCCGCAGAGCAGGAAGCAAGGCAT	TTTACTA		GVGYDVYKAFSTQLL
	CGTGAGGCTCCTGTACAAAAAGGGCCCGCGCGAAGACCCCGCGAACTG	CCTGGA		PVLHAAFQCCWQHH
	GAGACCTATCTGCCTCCAGCAAGTGATCTACAAGACCTACGCCGGGGTC	AGCCAA		RVPQSRKQGIVRLLY
	CTCGCGCGGAGGTTTACGCGCTGGTTGGCTGCAAACGGCCGCCATGCC	GGATTA		KKGPREDPANWRPI
	GACGCCCAGAAGGGGTTCCGCACTGTAAATGGCTGCGGTGAGCACAAT	ACCGGT		CLQQVIYKTYAGVLA
	TTTCTCGCGTCGACGCTGATCGATCATGCGCGCCGCAGCCGCAGAGAGC	CGGTAA		RRFTRWLAANGRHA
	TCCACATGGTCTGGTACGACCTGAAGAACGCCTTCGGCAGCGTGCCCCA	TACACAT		DAQKGFRTVNGCGE
	GGAGTTGCTGTGGGAGGTGCTGCAGCGCATGGGTGTCCCGCCAGCCTT	AACTTT		HNFLASTLIDHARRSR
	TGTGGAGGTCTGCCAGGGGCTGTACCAGGATGCTGCCTTCACCGTCGGC	(SEQ ID		RELHMVWYDLKNAF
	AACGCTGCTGACGGCCCGACCGATCTCGTGCGGCAGCTCGTGGGCGTG	NO:		GSVPQELLWEVLQR
	TTTCAAGGATGCCCGCTGAGCCCCCACCTGTTCACTGCGGCGATCAGCC	1329)		MGVPPAFVEVCQGL
	CGCTCCTCCACGCGCTGGACCGACTCAAGGACACGGGCGTGCGGCTGT			YQDAAFTVGNAADG
	CAGCGGACGACCGTCCCGGTGCCTCGGCCTACGCCGACGACCTCAAGAT			PTDLVRQLVGVFQG
	CTTCAGTGGCACCGCGGACGGCGTTAAGCGACACAGCACGCGCTCGTGGC			CPLSPHLFTAAISPLLH
	CGACTTCCTGCGCTGGACCGGCATGGTGGCGAACCCCAACAAGTGCTGC			ALDRLKDTGVRLSAD
	ACCATGTCGGTCCAGCGCGACGGACGGCGTGCTGAAGGCCTGCGAC			DRPGASAYADDLKIF
	CTGGAGCTGCAGCTGGACGGAGCCCGGATCCCCTCGCTGATCATGAAC			SGTADGVKRQHALV
	GCGTCGTACGCGTACCTGGGGACAGGCGACGGCTTTGACCACGTGCGC			ADFLRWTGMVANP
	CGCCGCATCGAGCTCGTCCCGGCGCTCATGCAGCTCAAGGACGATGCG			NKCCTMSVQRDGRG
	ACGGCACTGCTCCAGTCGGGGCTGGCGCCGTGTCAGGTGGTGAAGGCG			VLKACDLELQLDGARI
	GTGAAGACTTACCTCTTCCCGCGGGTGGAGTACGCACTCCGCCACCTCC			PSLIMNASYAYLGTG
	GCCCCTTTCAGCAGCAACTGGAAGGCTTTAATCGCCACCTCGTTTGCGG			DGFDHVRRRIELVPA
	CCTTCGCCACCTGCTCCGACTTCCGGTCAGCGCCACGACCAGCTTCTTCT			LMQLKDDATALLQS
	TCGTGCCGGTGTCGCGCGGCGGGCTAGGCCTGTTGCCACTGACGGAGC			GLAPCQVVKAVKTYL
	TGCACGCAGCTCCGGCTGATCGCCGACCGGCGCCACCGCCTAGACCCCG			FPRVEYALRHLRPFQ
	ACCACTGGAAGGAGCGGGAGGAGAAAATATGCGAGCTGCGGATCAAC			QQLEGFNRHLVCGLR
	TCGCAGCTCGGGACGTCGGCCCACGCCCCACCCAAGCGCCGCAACGGC			HLLRLPVSATTSFFFV
	GACATCGGCTCGTTGTGGGTCGACGTCCAGCGCCACCTCCGAGGACTCG			PVSRGGLGLLPLTELH
	GCCTCAAGCTAGAGACCGCCGGCGTGCGCGGAGACCGGCACCGCG			AAPADRRPAPPPRPR
	GAGGCGACGCTGCAGCTTCGCGTGCCGCACCACGAGAAGTGGCTGGAC			PLEGAGGENMRAAD
	CACCGCACGGTCCTGCGGCAGGTGAAGCTGCACATGAAGAACAAGCAT			QLAARDVGPRPTQA
	TGGCAGCGGTGGGCCGGGATGAAGGACCAAGGCAAGACCGCCCGCAC			PQRRHRLVVGRRPA
	CCATGGTGGTGCAGGGAGCGGCTTCCTCACGCGGCCTCGCGGCCTATG			PPPRTRPQARDRAG
	GGAGACCGACTACCGCTTTGCAGTGGCGGGTCGGCTCAACCAGCTGGA			VRGDRHRGGDAAAS
	CACGCACAGCGTGCTGAAGCGCCGGCGCCTCCGCGCGCACGACCAGTG			RAAPREVAGPPHGP
	TCGACAGCCGGGGCTGCTCCCGTGCCGAGACACTGGCACACGTGCTGAAT			AAGEAAHEEQALAA

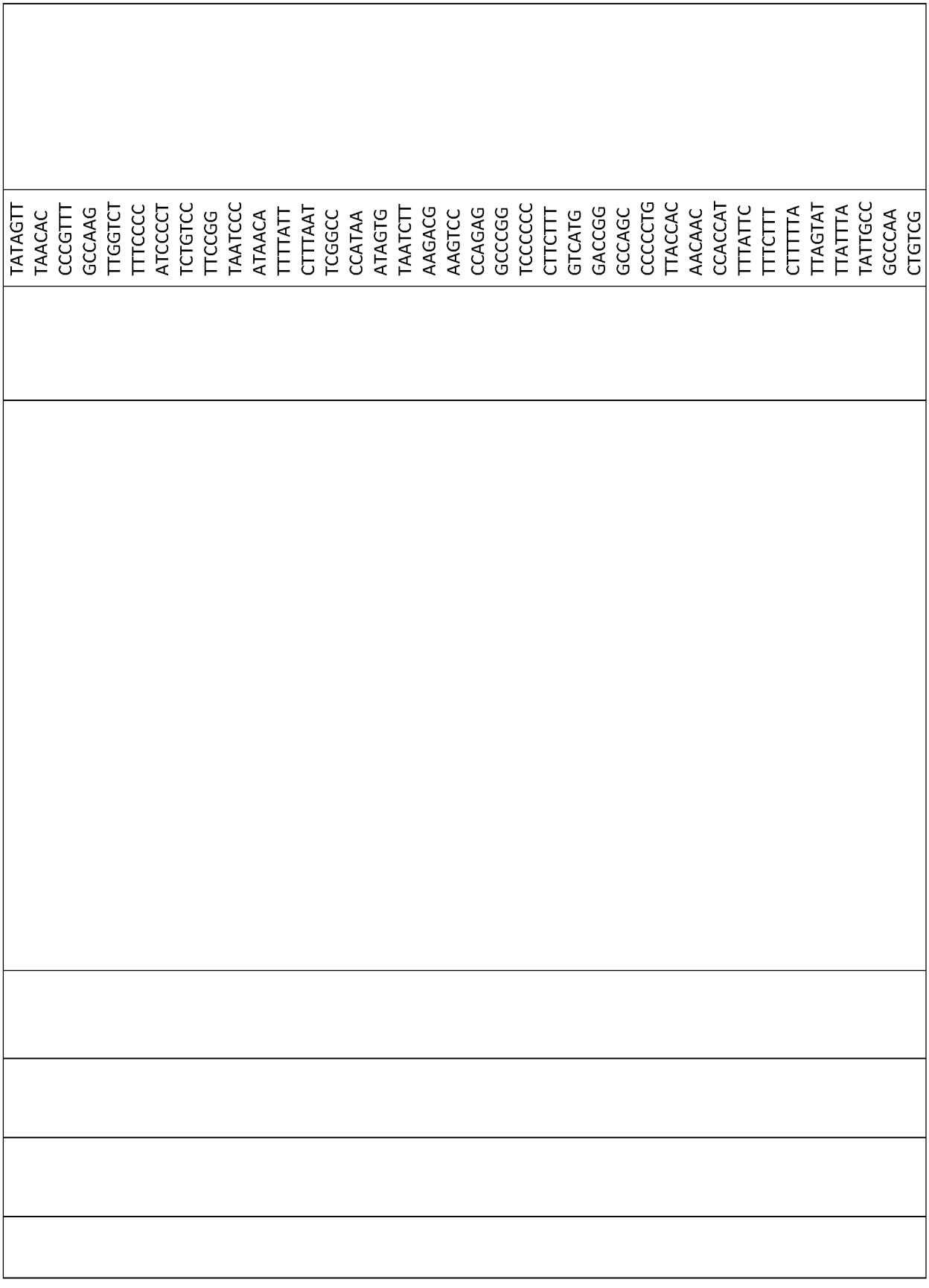
			CACTGCGCCGGCACCATGGACGCGGTCCGCGGCCGCCACGACGATGCG				VGRDEGPRQDRPHP
			CTCAAGACCATCGAGCGCGCGCTCATCGCGTCGTCCGGGCGGCCAGGA				WWCRERLPHAASRP
			CCGCGCCGAGCTCCGGGTCAACCAGACCGTGCCCTCGCTCGCCGGCCCC				MGDRLPLCSGGSAQ
			GCGCTGCGGCCCGACCTCCAGCTCTTCAACCACACCAAGAAGACGGTGG				PAGHAQRAEAPAPP
			CGGTGGTCGACCTGGCCGTGGCGTTCGAGGAGCAGGCGAGCGACGAC				RARPVSTAGLLPCRD
			ACCGAGAGCTCTGCGCTGGCGCGCATCGCCGCGCACAAGCGAGCCAAG				TGTRAESLRRHHGRG
			TACGCCGGCATCAAGCGGCACCTCGAGCGCCAAGGGTGGAAGGTCCAC				PRPPRRCAQDHRAR
			CTCTCGGCGCTCGTGTACGGCTCGTTGGGCGCGGTGATGCCCGCCAACT				AHRVVRAARTAPSSG
			ACAAGGTGCTTACTGAGCACCTTGGGCTGCTCAAGCGTGATGCGAAGC				STRPCPRSPAPRCGP
			GGCTGGACCGGCAGCTGTCGGTGGCGTGCATCCAGTCCAGCCGCCGCA				TSSSSTTPRRRWRWS
			TCTGGAACCTGCACTGCGCGCAGCATCGAGCACGGCAGCACCAGGGCC				TWPWRSRSRRATTP
			AAGCGCCAAGGGGCAGTCGGGCGGCGGAGACCGGGGGGGACTCCGCCA				RALRWRASPRTSEPS
			CAGAGCGGCCGCCGCTAAGCGGACATCGGGCCCGTAGCGGCCGACAG				TPASSGTSSAKGGRS
			GGCCACACCCATGTAGGGAACCGCCCTCTAAACCCGCCCGGTACATTAT				TSRRSCTARWAR
			GGTCCGACACCTATGAGGTGCAACCGGTACACAAGTTACACACCACATA				(SEQ ID NO: 1451)
			GCGACTACCAGGTATTTACTACCTGGAAGCCAAGGATTAACCGGTCGGT				
			AATACACATAACTTT (SEQ ID NO: 1083)				
NeSL	YURECi	.	Ciona intestinalis	ATCAACCC	TAGCCA		MATSSSSVSSGNVQT
			ATCAACCACTACTACACCCTCTACAAAGAACCCACTACAGCAAGATCTT	ACTACTAC	AAAGAT		EVRCVYHGKGDLLLE
			GGCGACCAACTACACCTGCAGCCTCCCGTCGACTCTTCACCACCTGCGCA	ACCCTCTA	TTGGTG		CPVAHCPSIHPTVATI
			CCTCCCTCCGACTCCAAGCGGCAGACAGTGGCACCACCAGCACTCCAAG	CAAAGAAC	TTCTTGG		TKHLKKHHTPQFEQI
			GGTCCACAGCCAGAAGAGTGCACCCCCCACTGCTAGGAAGAGCTAGGC	CCACTACA	CGAGCT		TTKNLTITYTCSQCSF
			CATTGCCCGCTGTGGGCCTTGTGCTGGCTGTCTGCCGCTGATTGGCACG	GCAAGATC	GCAGTC		STTGLTQHHISKHYKT
			TTAGGGAGTGTGCCGCTGAGCACGATAGGGAGGCTGCTGAAGCCGAAG	TTGGCGAC	CCGGCA		CKGVGAVQEGNKGR
			AACCCGCGCCGGGGACTTTAGGACTTTAACTAAAATCTGTAACACTAAA	CAACTACA	GAGTAG		FCCPACGTRWALLCK
			ATTGGAATACTGGATTACTATTGGATTACTACTGGACAATAACCTGATCT	CCTGCAGC	GCGAGA		ARHHFNNVHFEYDT
			ACAACAACCAATTGGATTACAGAACCAAATTACAATCTCAACTACAAAC	CTCCCGTC	TCGTGA		PPIAAFSGTPYKLKKR
			AAAAAGTGAGTCCCCGGCCGGGTTTCTATATTACCCAATTACTGTTTTTA	GACTCTTC	GCTCTAC		KFTIINKALTYSCPLPL
			CAATTTTAAAATTTTATAGTATTTTACTAATTTTTCCGCCCCGCTAGCACT	ACCACCTG	ACCGCTT		NQLLCPLWSCSLTILN
			TAAATTGGCACCCCCCCCCCCCSTCCAAAAAAAAATAGATAACCCTCCAC	CGCACCTC	CACGAT		KPLSSVQQETAHGD
			CACCTAAACCCCGGTCACCCCCTCAGAACTAATTGGAACCCGGATCTTTT	CCTCCGAC	TTGACG		GSQGQSYVPTQLRQ
			CCCTCTATCAAATTTCCGGTTCAGTTTGGCTTAATTACCTAAAACCCGTCT	TCCAAGCG	AGTTCTT		VLRARCHCGNPPIGK
			TTATTTTATCCCTTTATTGCCCCTTAAACCAAATTATTCAAATTGAAATCC	GCAGACA	CTTGAG		GHWASCQGKRPLSS
			GGCCAAATTGCCTTGCTGAACCACCATTTTTGTTTTTCTTATAAAAGTAA	GTGGCACC	TCGATCC		PKGGRSSPTPPANLT
			ATTTTTCAACAACCCAAAATCATAAAATTAAAACTGTTATCTGATTAGTCT	ACCAGCAC	CCAACG		LHFLNYLPFQLPSQSS
			AAATTCTGTTATTGACAAAAAACCTCCATACCAACCTTAATCTTATCATTAT	TCCAAGGG	GCAACG		SPQSSTLDPTACKAR
			CTTCACCCAGACTGCTCAATCTAGTTCTTTTTTCAATCAATAAGGAGGTG	TCCACAGC	CACTCCA		VPIPSFLRGDCEVTFFI
			TTTTCTTTGGTCTGTTGTTTTATAGTTTACCGTTATAATAGTCTGTTGTCCT	CAGAAGA	AATTTAA		IPSVNFYRPYLSYPLR
			ATAGTTTTCCGTGGTGTATTTCTGGTCTACTATAGTTGGTGCCTCTAAAA				

	CATTATTATACCGAAGAGTGCTAGCGATTTCTATTATTAACTCACCCACC	GTGCACCC	CGACCG	MFWRNRTSSGHCSL
	AGTACTGCTGTGCATCCACGCAGCACCACGCCTGCCATCACGTACAGTG	CCCACTGC	CATGGC	HRVVRGFVRERLVPH
	CGCTTGCTGCCTTGTGTTTTGTTTTGCAGACACCACTGGACGATAGATGG	TAGGAAG	GACGCC	PRSKSPARTPLEFLCE
	CGACTTCGTCTTCCTCCGTTTCATCGGGGAAATGTTCAGACCGAGGTGCGT	AGCTAGGC	ACTGAT	FRLAGVFPDPGKVAS
	TGCGTCTACCACGGAAAAGGAGACTTGTTGCTGGAATGCCCAGTTGCCC	CATTGCCC	CTCCCA	LRPVPAPLTLCLSPPV
	ATTGCCCTAGCATACACCCCACAGTTGCCACAATCACAAAACACTTAAAA	GCTGTGG	GTCATCC	AGPMISCEDHSAPPS
	AAACATCACACACCTCAATTCGAACAGATAACCACTACTAAAAATCTCACAAT	GCCTTGTG	GTGCTC	VRSSSPIPNSPASVSS
	AACATATACATGCAGTCAATGTAGTTTCTCCACCACCGGTCTCACACAAC	CTGGCTGT	CGAGGG	VEAHLSDLLDKVSSG
	ACCACATATCCAAACACTACAAAACATGCAAAGGCGTGGGCGCTGTTCA	CTGCCGCT	CACTTAA	ELRPLSPTLPSSGFFG
	GGAAGGCAACAAGGGTCGATTCTGCTGCCCTGCATGCGGTACCAGGTG	GATTGGCA	GGGGAA	PLLPPTPPPRPTPSAE
	GGCTCTTTTGTGTAAGGCACGGCACCATTTTAATAATGTCCACTTCGAGT	CGTTAGGG	CCGATTC	KASPSGLSYLPCREVK
	ATGACACCCCTCCAATTGCCGCTTTCTCTGGCACACCCTATAAACTAAAA	AGTGTGCG	ACTGAT	IASIARPSPASQRVGC
	AAACGGAAATTTACAATTATTAATAAGGCCCTAACATATTCATGCCCTCT	CTGAGCAC	GCTGCA	DADRTGPSLNPNYQ
	CCCCCTCAACCAACTACTTTGCCCTTTGTGGTCGTGCTCCCTCACGATCTT	GATAGGG	AGTTGG	QTSPPSTPSFSPIVRP
	AAATAAACCTCTTTCAAGCGTGCAGCAGGAAACGGCACATGGCGACGG	AGGCTGCT	CGGGGG	PKFPRSGAKVNSKSK
	CTCACAAGGCCAGAGTTATGTTCCAACTCAGTTGCGGCAGGTGCTCAGA	GAAGCCG	CGCACT	PPGVRPRRAKPIEPG
	GCCCGTTGCCACTGCGGAAACCCTCCAATTGGAAAGGGCCACTGGGCAT	AAGAACCC	CACCAA	TESASPVDVDTISSSV
	CTTGCCAAGGGAAGCGACCACTCTCTTCACCTAAGGGTGGGCGCAGTTC	GCGCCGG	TGGAAG	QEPCTPENRTPEFFYE
	GCCCACTCCACCTGCTAATCTAACCCTCCACTTCCTTAACTATCTCCCCTT	GGACTTTA	ATGAGG	RKWLVSILNIHEREGS
	CCAACTTCCCTCACAAAGCTCCTCCCCCCAAAGTAGCACCCTCGACCCTA	GGACTTTA	GAGCCC	NFFQFNRDLEYWTQ
	CGGCCTGTAAAGCTCGAGTCCCCATCCCCTCTTTCCTACGCGGCGACTGT	ACTAAAAT	AACGAT	LLSGSQKGGRAKRAS
	GAAGTCACCTTTTTCATCATCCCCTCCGTTAATTTTTATAGACCCTACCTC	CTGTAACA	GCCCCG	YNRGAANQAMKNR
	TCCTACCCCCTCAGGATGTTCTGGCGTAACCGAACATCCTCCGGTCACTG	CTAAAATT	AACTAC	DSGRKDFDPRPVAG
	TTCCCTACACCGGGTCGTTCGAGGCTTTGTCCGAGAGCGACTCGTCCCCC	GGAATACT	CGCACC	HSSGGGTELGSRPRY
	ATCCTCGGTCGAAATCGCCGGCGAGGACTCCCCTTGAATTCCTCTGCGA	GGATTACT	GTCCCC	PKGARLRADFWRD
	GTTCCGGCTCGCGGGGGTCTTCCCCGATCCTGGGAAGGTCGCGTCCCTT	ATTGGATT	AGCACC	MKGTVRKLLDGSNG
	CGACCCGTACCAGCTCCCCTTACGCTTTGCCTTTCTCCACCAGTCGCTGG	ACTACTGG	GTGTCC	ERRCGIPLDIIERKFRQ
	TCCAATGATTTCCTGTGAGGACCACAGCGCCCCCCCTTCCGTTCGCTCGA	ACAATAAC	TTGTGA	VSMPGWIDHRRYAA
	GCTCCCCCATTCCTAACTCCCCAGCCTCCGTAAGCTCCGTGGAGGCGCAC	CTGATCTA	GTAATT	GASPSLVTQAETDVA
	TTGTCTGACTTGTTAGACAAAGTGTCTTCCGGAGAACTGCGGCCGCTGT	CAACAACC	ATTTTCA	ITSEEVEAVLSGLNVQ
	CCCCCACCCTGCCTTCTTCAGGCTTTTTTGGCCCTCTGTTGCCACCCACTC	AATTGGAT	CTAATCC	SAPGSDGLSYRFWK
	CCCCTCCCCCGACCTACCCCCTCGGCCGAAAAGGCATCTCCTTCAGGCCTT	TACAGAAC	AGGGAC	GLDPSGRLLSCLFEIV
	TCCTACCTACCCTGCAGAGAAGTAAAAATCGCGTCCATTGCACGGCCAT	CAAATTAC	GGGGCC	RRHGRIPGAWPTCS
	CCCCTGCCTCACAGAGGGTCGGCTGCGACGCGGACCGAACAGGGCCTT	AATCTCAA	ATGAAC	VILLCKDAQGDVQDV
	CGCTGAACCCTAATTATCAGCAAACCTCCCCCCCGTCTACTCCGAGTTTC	CTACAAAC	TGAACT	GNWRPITICRTLYKLY
	TCCCCCATAGTTAGACCCCCTAAATTCCCCCGGTCGGGCGCCAAAGTAA	AAAAAGTG	ATGTCT	AAVIARRIQTWAKQ
	ATTCCAAGTCCAAACCCCCCGGCGTCCGACCACGCCGTGCCAAGCCAAT	AGTCCCCG	GCACGC	GGVLSRLQKGFMPV
	TGAGCCAGGGACCGAGTCGGCATCACCCGTTGATGTCGACACTATTAGT	GCCGGGTT	TGCCCC	EGVFEHVFMLDTVLS

		AGTTCGGTTCAGGAACCTTGTACTCCAGAAAACAGAACCCCGGAATTTT	TCTATATT	GCCTAG	DAKLRRKNLLAVFLD
		TTTATGAGCGAAAGTGGCTCGTGAGTATTTTAAATATTCACGAGCGGGA	ACCCAATT	CCTCGG	VRNAFGSVRHECLLK
		GGGATCCAATTTTTTTCAATTTAACCGGGATCTGGAGTATTGGACGCAG	ACTGTTTT	CCACAA	VLRHFDAPHYLVELV
		TTGCTTTCAGGCTCCCAGAAGGGAGGACGCGCGAAGCGTGCGTCCTAC	TACAATTT	ATAAAT	RDIYTGATCRVRSSV
		AATCGTGGTGCGGCGAACCAAGCGATGAAAAACCGCGATTCCGGCCGT	TAAAATTT	AGTCAA	GETGDIPCDRGVRQ
		AAGGATTTCGATCCCCGGCCGGTAGCAGGCCACAGTAGTGGAGGTGGC	TATAGTAT	GCCGGG	GYPLSGILFNLVTEVLI
		ACCGAGCTCGGTTCGCGGCCACGATACCCTAAGGGCGCACGTCTTCGTG	TTTACTAA	GCGTAC	PGLSAGNDGYRMAC
		CCGACTTCTGGAGGGACATGAAAGGGACGGTCCGTAAGTTACTTGATG	TTTTTCCG	TACCAAT	LGGKLTQVLAYADDL
		GTTCGAACGGTGAGCGTCGCTGCGGGATTCCCCTCGATATCATAGAGCG	CCCCGCTA	TGATTC	VVVTENRDQMLRQL
		GAAATTCCGGCAAGTATCCATGCCCGGCTGGATAGACCACAGACGGTA	GCACTTAA	GGATTT	GVCEEFGRWAGLAF
		CGCGGCCGGCGCGTCCCCGTCCCTAGTAACCCAAGCTGAAACGGATGTT	ATTGGCAC	GGCGGG	NQRKCGLIGWRTLR
		GCCATCACCAGTGAGGAAGTTGAGGCGGTCCTCAGTGGCCTAAACGTTC	CCCCCCCC	GCCCAT	GGRRVALEDPLLLNG
		AGTCCGCTCCGGGCTCGGATGGCCTCTCCTATCGATTTTGGAAGGGCCT	CCCSTCCA	CACGCG	VEIPLLRPGEHYKYLG
		CGATCCTAGTGGTCGATTGTTGTCATGCCTTTTCGAAATCGTTAGGAGGC	AAAAAAAA	CCTCCGC	AMTGVMSVPRTGS
		ATGGTCGTATTCCGGGCGCGTGGCCGACTTGTAGCGTCATCCTGCTCTG	TAGATAAC	CCCCAC	QLIKDFRARLQRLFTS
		TAAAGATGCTCAGGGTGACGTTCAGGACGTGGGAAACTGGAGACCGAT	CCTCCACC	ACACAA	FLTPHQKLIALKRFLLP
		CACCATCTGCCGCACTCTGTACAAATTGTACGCGGCTGTGATCGCCAGA	ACCTAAAC	AAACAC	SLSFHLRVRPIARSELI
		CGTATTCAGACTTGGGCGAAACAGGGCGGGGTGCTCTCGCGGCTCCAG	CCCGGTCA	TTTTAAA	ALDRRVRECLRVAFR
		AAAGGGTTCATGCCAGTTGAGGGTGTGTTTGAGCACGTCTTCATGTTGG	CCCCCTCA	CTCTTCG	LTKPSCQAVFHTPTD
		ATACCGTCCTTTCCGACGCCAAGCTGAGGCGCAAGAATCTTCTTGCCGTC	GAACTAAT	GTTCCCC	MDGLGVPSVCSESSI
		TTCCTTGACGTTCGGAATGCGTTCGGAAGCGTCCGGCATGAATGCTTGC	TGGAACCC	AACCAC	LTIAQGFKVLTSPDGT
		TCAAAGTTCTTCGGCACTTCGATGCTCCCCATTACCTGGTTGAACTCGTG	GGATCTTT	TACAAC	VSATASARVKLYAAK
		AGGGATATTTACACCGGTGCGACCTGCCGTGTTAGATCTTCGGTTGGTG	TCCCTCTA	AAAGCG	FGGLTEAGPSDWAR
		AGACCGGTGATATACCCTGTGATCGAGGTGTCAGGCAGGGCTATCCCCT	TCAAATTT	AGCGGC	YLSGDDVNGNSTRKP
		GTCTGGCATTCTTTTCAACCTGGTAACGGAGAGGTGCTAATACCGGGGCTC	CCGGTTCA	CCCTCTT	GANLPSGLWTRVRC
		AGTGCCGGGAATGATGGCTACCGGATGGCTTGCTTAGGTGGTAAGCTC	GTTTGGCT	AGATCC	ASRQLGAVWRVCPE
		ACTCAAGTCCTGGCCTACGCTGACGATCTGGTGGTAGTCACGGAAAACC	TAATTACC	AATTTTA	NGITVRVRNSVITSRD
		GTGATCAGATGCTACGTCAGCTGGGTGTGTGTGAGGAATTCGGCCGTT	TAAAACCC	AAATTTT	RRKLIRSFHDCSNQQ
		GGGCAGGCCTTGCATTCAATCAACGGAAGTGTGGCCTGATCGGCTGGA	GTCTTTAT	AAATCA	WKEQWMQHPNQE
		GGACGTTGCGGGGAGGGAGGAGGGTCGCACTGGAGGATCCTCTGCTC	TTTATCCCT	GTGCAC	KTAAAHMAYADANR
		CTTAACGGAGTAGAGATCCCCCTGTTGCGGCCCGGTGAACACTATAAAT	TTATTGCC	TCAACTT	WVKQPSVMEPHTYF
		ACCTTGGCGCCATGACTGGCGTCATGAGTGTTCCCCGCACCGGATCCCA	CCTTAAAC	TTTTACG	FALRARLNLLPTRVSR
		GCTGATCAAGGACTTCCGTGCAAGGCTCCAACGACTGTTTACATCGTTCC	CAAATTAT	TGTTGT	AIYSRDQHPDILCRRC
		TCACACCACACCAGAAGCTGATCGCTCTGAAACGGTTTCTGCTGCCAAG	TCAAATTG	GTTTTTT	GASVESLPHVLNHCP
		TCTCAGCTTTCATCTGAGGGTTCGTCCGATTGCGCGGTCTGAGCTGATCG	AAATCCGG	GTATTTT	PNMSILGRHNLVLQ
		CGCTCGACCGTCGCGTGCGCGAGTGCCTTAGGGTGGCGTTTCGCTTGAC	CCAAATTG	TTCCCAC	EVLNAVDKTQFKEIS
		CAAGCCCTCGTGTCAGGCTGTGTTTCATACACCGACTGACATGGATGGT	CCTTGCTG	AGATTTT	VDRTVPEHMSETGE
		CTGGGTGTCCCGTCGGTATGCAGTGAGAGCAGCATACTGACGATCGCCC	AACCACCA	GTATTTT	ALRPDIVARRNDGSV

	AGGGGTTCAAGGTGTTAACGTCACCCGACGGCACAGTGTCCGCGACGG	TTTTTGTTT	TATATTA	VVVDVACPFDQKAN		
	CTTCGGCGCGCGTGAAGCTCTACGCCGCTAAATTTGGCGGCCTCACTGA	TTCTTATA	TATTTTA	FDEAAKRKLLKYDKLC		
	AGCAGGTCCTTCGGACTGGGCGCGTTATCTATCTGGTGATGACGTCAAC	AAAGTAAA	TATACAC	CNIAASTGKPVECHSI		
	GGGAATAGCACTCGCAAACCCGGCGCTAACCTGCCTTCTGGTCTGTGGA	TTTTTCAA	AACACT	VVGSLGSLAEGLSTSL		
	CTCGCGTCCGCTGTGCCAGCAGGCAGTTAGGTGCTGTGTGGCGCGTTTG	CAACCCAA	ATTTTTA	RALGITDFARSKLVAC		
	CCCGGAAAACGGCATCACTGTCCGGGTGCGTAATTCTGTGATTACCTCC	AATCATAA	TACACTA	HQG (SEQ ID NO:		
	CGAGACCGACGGAAGCTCATCAGATCCTTCCACGATTGCTCTAATCAGC	AATTAAAA	CCTTGCA	1452)		
	AGTGGAAGGAGCAGTGGATGCAGCACCCCAACCAGGAGAAGACCGCT	CTGTTATC	CTGTCCC			
	GCAGCGCATATGGCTTATGCCGATGCGAACCGCTGGGTGAAGCAGCCG	TGATTAGT	ACTTTTT			
	TCGGTTATGGAACCTCATACATACTTTTTTGCATTGAGGGCGCGCTTGAA	CTAAATTC	GTAATT			
	CCTGTTACCCACTAGGGTTTCACGAGCTATTTATAGCAGGGACCAACAC	TGTTATTG	ATCACCT			
	CCGGACATTTTATGCCGACGTTGTGGTGCGTCAGTTGAGTCCCTCCCGC	ACAAAAAC	TTTACCT			
	ATGTACTAAACCATTGTCCTCCCAATATGAGCATAATACTCGGGAGGCA	CTCCATAC	TTTATGC			
	TAATTTGGTACTGCAGGAGGTTCTGAACGCGGTAGACAAAACCCAGTTC	CAACCTTA	CGCTCG			
	AAAGAGATATCCGTTGACCGAACGGTCCCCGAGCACATGTCGGAGACA	ATCTTATC	CTGAGC			
	GGTGAGGCCTTGCGGCCTGACATAGTGGCACGGCGGAATGACGGTTCC	ATTATCTT	TCCTTGC			
	GTCGTGGTCGTCGACGTGGCTTGCCCGTTCGATCAAAAGGCGAACTTTG	CACCCAGA	ACGGAT			
	ATGAGGCGGCAAAACGGAAGCTCCTAAAGTACGACAAGCTTTGTTGCA	CTGCTCAA	GACCAG			
	ATATCGCTGCCTCCACAGGTAAACCTGTGGAGTGTCATTCGATAGTCGT	TCTAGTTC	ACAAAC			
	GGGCTCCTTGGGGAGCCTCGCGGAGGGTCTGTCCACCTCCCTCCGGGCT	TTTTTTCAA	TTTTATA			
	TTGGGAATAACCGATTTTGCTCGCTCTAAGCTTGTCGCTTGCCATCAAGG	TCAATAAG	AAATTAT			
	GTAGCCAAAAGATTTGGTGTTCTTGGCGAGCTGCAGTCCCGGCAGAGTA	GAGGTGTT	AACATT			
	GGCGAGATCGTGAGCTCTACACCGCTTCACGATTTGACGAGTTCTTCTTG	TTCTTTGG	GTTTTAA			
	AGTCGATCCCCAACGGCAACGCACTCCAAATTTAACGACCGCATGGCGA	TCTGTTGT	TTGTCG			
	CGCCACTGATCTCCCAGTCATCCGTGCTCCGAGGGCACTTAAGGGGAAC	TTTATAGT	CGGGGT			
	CGATTCACTGATGCTGCAAGTTGGCGGGGGCGCACTCACCAATGGAAG	TTACCGTT	ATCAGT			
	ATGAGGGAGCCCAACGATGCCCCGAACTACCGCACCGTCCCCAGCACCG	ATAATAGT	GGCGCC			
	TGTCCTTGTGAGTAATTATTTTCACTAATCCAGGGACGGGGCCATGAACT	CTGTTGTC	CCCTTGC			
	GAACTATGTCTGCACGCTGCCCCGCCTAGCCTCGGCCACACAAATAAATAG	CTATAGTT	GGCCGT			
	TCAAGCCGGGGCGTACTACCAATTGATTCGGATTTGGCGGGGGCCCATCA	TTCCGTGG	GAAAGC			
	CGCGCCTCCGCCCCCACACACAAAAACACTTTTAAACTCTTCGGTTCCCC	TGTATTTC	CCTTTTC			
	AACCACTACAACAAAGCGAGCGGCCCCTCTTAGATCCAATTTTAAAATTT	TGGTCTAC	ACCTTAA			
	TAAATCAGTGCACTCAACTTTTTTACGTGTTGTGTTTTTTGTATTTTTTCCC	TATAGTTG	TTCGCCC			
	ACAGATTTTGTATTTTTATATTATATTTTATATACACAACACTATTTTTATA	GTGCCTCT	CTTAGTC			
	CACTACCTTGCACTGTCCCACTTTTTTGTAATTATCACCTTTTACCTTTTATG	AAAACATT	CAAATTT			
	CCGCTCGCTGAGCTCCTTGCACGGATGACCAGACAAACTTTTATAAAATT	ATTATACC	TTCCCCA			
	ATAACATTGTTTTAATTGTCGCGGGGTATCAGTGGCGCCCCCTTGCGGC	GAAGAGT	TTGCCT			
	CGTGAAAGCCCTTTTCACCTTAATTCGCCCCTTAGTCCAAATTTTTCCCCA	GCTAGCGA	GTAAAA			

				TTGCCTGTAAAAGTGCGTTGCCGTCGACTCGAACTCGTCCTTTTTACCGCCTCTCTGTTACTGAACTATGACCAGCTTGGCTGTTGAAGTCGGCTTTAATTCGCCGGCTTCACATTATATTTTTTTGTTGGGTAATCTGTTTTTTTCATATATAAAACCATTCGGCTTAAAATCACCAATCCCCATCATCCACCCTGGTCACCCATTGAAACATCTCTTTTTAGTCAATTATTTTTTCAGATTACCCCTCCCAAATTGTCATATAGTTTAACACCCCGTTTGCCAAGTTGGTCTTTTCCCCATCCCCTTCTGTCCTTCCGGTAATCCCATAACATTTTATTCTTTAATTCGGCCCCATAAATAGTGTAATCTTAAGACGAAGTCCCCAGAGGCCCGGTCCCCCCCTTCTTTGTCATGGACCGGGCCAGCCCCCCTGTTACCACAACAACCCACCATTTTATTCTTTCTTTCTTTTTATTAGTATTTATTTATATTGCCGCCCAACTGTCGTGTCGTGGCGCAGGGGGGGTTCCCTCTGTCGGCCCAGTGGACGACCGTCTAAAAAACAGGCACAGGCAGCAGGCACTTATCACCCGGTACCTCCGGGTACCGGAGGCTGGTTTTCAGCGCACTGTACGTGATGGCCAATTTTATTCATTGCATTTTATCCGCGTCGTGGTGTTTGCGTGGATGCATTAATAAAAGATGAAATTCC (SEQ ID NO: 1084)	TTTCTATTATTAACTCACCCACCAGTACTGCTGTGCATCCACGCAGCACCACGCCTGCCATCACGTACAGTGCGCTTGCTGCCTTGTGTTTTGTTTTGCAGACACCACTGGACGATAG (SEQ ID NO: 1207)	GTGCGTTGCCGTCGACTCGAACTCGTCCTTTTTACCGCCTCTCTGTTACTGAACTATGACCAGCTTGGCTGTTGAAGTCGGCTTTAATTCGCCGGCTTCACATTATATTTTTTTGTTGGGTAATCTGTTTTTTTCATATATAAAACCATTCGGCTTAAAATCACCAATCCCCATCATCCACCCTGGTCACCCATTGAAACATCTCTTTTTAGTCAATTATTTTTTCAGATTACCCCTCCCAAATTGTCA	


TATAGTT
TAACAC
CCCGTTT
GCCAAG
TTGGTCT
TTTCCCC
ATCCCCT
TCTGTCC
TTCCGG
TAATCCC
ATAACA
TTTTATT
CTTTAAT
TCGGCC
CCATAA
ATAGTG
TAATCTT
AAGACG
AAGTCC
CCAGAG
GCCCGG
TCCCCCC
CTTCTTT
GTCATG
GACCGG
GCCAGC
CCCCCTG
TTACCAC
AACAAC
CCACCAT
TTTATTC
TTTCTTT
CTTTTTA
TTAGTAT
TTATTTA
TATTGCC
GCCCAA
CTGTCG

R2	PERER E-9	BN000 800	Schistos oma	ATCTCACGTTTTAATTTATTTTTGAACTACTGCAGTCTGAGTGCTTCTAAC GACCCGAAGGCTCAGAAACTACCCACTTCTTGAACTGCTACTTTTTGCTG	ATCTCACG TTTTAATTT	TGTCGT GGCGCA GGGGGG TTCCCTC TGTCGG CCCAGT GGACGA CCGTCTA AAAAAC AGGCAC AGGCAG CAGGCA CTTATCA CCCGGT ACCTCC GGGTAC CGGAGG CTGGTTT TCAGCG CACTGT ACGTGA TGGCCA ATTTTAT TCATTGC ATTTTAT CCGCGT CGTGGT GTTTGC GTGGAT GCATTA ATAAAA GATGAA ATTCC (SEQ ID NO: 1330)	TAACGG CTGAAC	MPVSTGAETDITSSLP IPASSIVSPNYTLPDSS

	manson i	TTTATCCACAACAACAGTTGTGATTCTATTCTCCANATATTCCTTGTGCTT	ATTTTTGA	GAATAG		STCLICFAIFPTHNILLS
		TTGTCAACATTATTCTATACCAACTGTACCACCTACTTCTTCATCTCACGT	ACTACTGC	CCCCCTT		HATAIHHISCPPTPVQ
		TTTAATTCTGGTCTTATTTTCTCATCATTAGTCACGGAGAGGGCCTATGA	AGTCTGAG	CACTCTT		DGSQQMSCVLCAAA
		ACGGTCCGTGACGCGAAATTCTATCCGCGATTTCGACCTCTCCTGCTAGT	TGCTTCTA	AGACAT		FSSNRGLTQHIRHRHI
		GGTCCCCGAAGTACGGTTCCTCTGGCCTGTCAGTTGTGTTAAACTATAT	ACGACCCG	TCCCCCA		SEYNELIRQRIAVQPT
		AATAACGATGCCGGTCTCAACCGGCGCAGAAACTGACATAACCTCTTCT	AAGGCTCA	CTGTTGT		SRIWSPFDDASLLSIA
		TTGCCTATTCCTGCATCCTCAATCGTCTCGCCAAACTACACACTCCCTGAT	GAAACTAC	TGCTTAT		NHEAHRFPTKNDLC
		TCCTCTTCAACCTGCCTTATATGTTTCGCTATCTTCCCCACCCACAACATA	CCACTTCT	CTTCATG		QHISTILTRRTAEAVK
		CTCCTCTCCCATGCCACTGCAATCCACCATATTTCTTGTCCTCCTACTCCA	TGAACTGC	CTCTTGT		RRLLHLQWSRSPTAI
		GTGCAAGACGGTTCTCAGCAGATGTCTTGTGTTCTTTGCGCCGCCGCTTT	TACTTTTT	GTTAATT		TTSSNNHTITDIPNTE
		TTCATCTAACAGGGGACTAACACACAACACATTCGCCACCGGCACATCTCC	GCTGTTTA	GACTGC		ARYIFPVDLDEHPPLS
		GAATATAACGAACTAATCAGACAACGAATTGCAGTGCAGCCGACGTCTC	TCCACAAC	TCTCTTC		DATTPNASTHPLPELL
		GCATATGGTCACCATTCGATGATGCTTCTTTACTATCCATCGCTAACCAT	AACAGTTG	TGGGTT		VILTPLPSPTRLQNISE
		GAAGCCCATAGATTCCCCACGAAGAATGACCTATGCCAACATATCAGCA	TGATTCTA	GACGTC		SQTSHESNKNSMHT
		CCATACTAACACGCAGGACGGCAGAAGCCGTCAAACGCCGACTCCTCCA	TTCTCCAN	TGATTGT		PPTYACDPDETLGAT
		CCTACAGTGGTCGAGATCACCAACAGCGATTACTACCTCTTCGAATAATC	ATATTCCT	CTCTCTC		PSSTIPSCFHSYQDPL
		ACACAATCACAGACATCCCCAATACCGAGGCCCGATATATTTTTCCGGTA	TGTGCTTT	TCTTTCC		AEQRGKLLRASASLL
		GACCTAGACGAACATCCACCATTGTCTGATGCCACACAACCCCCAACGCATC	TGTCAACA	ATATTGC		QSSCTRIRSSSLLAFL
		GACACATCCACTCCCAGAACTCCTTGTCATCTTGACACCGCTTCCATCCCC	TTATTCTAT	TTGCTCT		QNESTLMDEEHVSTF
		GACTAGACTACAAAACATATCCGAATCACAGACCTCCCATGAATCTAATA	ACCAACTG	GCCCGC		LNSHAEFVFPRTWTP
		AGAACTCAATGCATACACCGCCAACGTATGCCTGCGATCCGGATGAGAC	TACCACCT	TTACTTC		SRPKHPSHAPANVSR
		ACTAGGGGCTACTCCCTCATCAACTATTCCCTCATGCTTCCACAGTTATCA	ACTTCTTC	CAATAG		KKRRKIEYAHIQRLFH
		GGACCCCCTAGCTGAACAAAGAGGCAAACTCCTGAGGGCATCCGCCAG	ATCTCACG	TTGTCAT		HRPKDASNTVLDGR
		CCTACTACAAAGCAGTTGTACTCGCATACGGTCCTCCAGCCTGCTCGCCT	TTTTAATTC	ATTATGT		WRNPYVANHSMIPD
		TCCTCCAAAACGAATCCACATTAATGGACGAGGAACACGTGTCCACCTT	TGGTCTTA	CTTTGTT		FDCFWTTVFTKTNSP
		CCTCAATAGTCATGCAGAATTCGTCTTCCCTAGAACATGGACCCCATCCC	TTTTCTCAT	TACTTGC		DSREITPIIPMTPSLID
		GACCCAAACACCCCTCCCACGCCCCAGCTAATGTTTCTAGGAAGAAAAG	CATTAGTC	CATGTCT		PILPSDVTWALKEMH
		GAGGAAAATAGAGTACGCACACATCCAGAGACTCTTCCACCACCGTCCC	ACGGAGA	AACGAC		GTAGGIDRLTSYDLM
		AAAGATGCCTCCAACACCGTTCTAGACGGTCGGTGGAGAAACCCCTATG	GGGCCTAT	AATTACT		RFGKNGLAGYLNML
		TCGCAAACCATTCAATGATTCCAGACTTCGACTGCTTCTGGACAACAGTC	GAACGGTC	TTATCTA		LALAYLPTNLSTARVT
		TTTACTAAAACAAATTCCCCAGACACAGCCGGGAGATTACTCCAATCATCCC	CGTGACGC	CCTTAGT		FVPKSSSPVSPEDFRP
		TATGACTCCCTCTCTCATTGACCCGATCCTCCCCTCTGACGTCACATGGG	GAAATTCT	TTGTCCT		ISVAPVATRCLHKILA
		CGCTGAAAGAAATGCATGGCACGGCCGGTGGGATTGATCGTCTGACAT	ATCCGCGA	CTTGGTT		KRWMPLFPQERLQF
		CGTACGATCTGATGAGATTCGGGAAGAATGGTCTTGCTGGATATCTCAA	TTTCGACC	TCGATT		AFLNRDGCFEAVNLL
		CATGCTACTCGCTCTTGCATACCTTCCCACTAATCTTTCAACAGCACGGG	TCTCCTGC	GCCTTCA		HSVIRHVHTRHTGAS
		TAACTTTCGTCCCCAAGTCATCAAGTCCTGTGTCACCTGAGGACTTCCGT	TAGTGGTC	TATGTTC		FALLDISRAFDTVSHD
		CCCATCAGTGTCGCTCCAGTAGCCACTAGGTGCCTGCACAAAATTCTAG	CCCGAAGT	ATGGCG		SIIRAAKRYGAPELLC
		CTAAGAGATGGATGCCGCTCTTTCCACAGGAACGACTTCAGTTCGCTTTC	ACGGTTCC	GAATCT		RYLNNYYRRSTSCVN

	CTAAACCGAGATGGATGCTTTGAAGCAGTTAATCTTCTGCACTCGGTCAT	TCTGGCCT	GATGTTT	RTELHPTCGVKQGDP
	ACGGCACGTCCACACCCGCCATACAGGAGCATCCTTCGCCCTGCTCGAC	GTCAGTTG	ATAATG	LSPLLFIMVLDEVLEG
	ATATCACGGGCCTTTGACACTGTATCACATGACTCCATCATCAGAGCGG	TGTTAAAA	ACTATTC	LDPMTHLTVDGESLN
	CGAAAAGATATGGGGCACCTGAACTGTTATGCCGCTACCTCAATAACTA	CTATATAA	CTATTAC	YIAYADDLVVFAPNA
	TTACCGACGTTCAACCAGCTGCGTCAACCGCACTGAATTGCATCCTACGT	TAACG	CACCACT	ELLQRKLDRISILLHEA
	GTGGGGTGAAGCAAGGAGACCCCCTGTCGCCACTCCTCTTCATCATGGT	(SEQ ID	ACAACT	GWSVNPEKSRTLDLI
	TCTCGACGAAGTACTGGAAGGTCTAGATCCAATGACCCACCTAACAGTT	NO: 1208)	ACTATTA	SGGHSKITALSQTEFT
	GATGGAGAGAGCTTGAACTACATAGCTTATGCTGACGATCTCGTAGTTT		TTATTTT	IAGMRIPPLSAADTF
	TCGCTCCAAATGCGGAACTCCTTCAACGGAAACTCGATCGGATCTCCAT		CATTACT	DYLGIKFNFKGRCPV
	ACTTCTACACGAGGCTGGATGGTCGGTTAACCCTGAAAAAAGCCGGACC		ATTAAC	AHIDLLNNYLTEISCA
	CTGGACCTAATCTCTGGTGGCCATTCCAAAATCACAGCGCTCTCTCAGAC		ATTATTA	PLKPQQRMKILKDNL
	AGAATTCACCATCGCGGGGATGCGTATACCACCGCTTTCTGCCGCCGAC		TAAACA	LPRLLYPLTLGIVHLKT
	ACCTTCGACTACCTGGGTATCAAATTCAACTTCAAGGGCCGATGCCCAG		TTATTAC	LKSMDRNIHTAIRKW
	TGGCCCATATTGACTTATTGAACAACTACCTCACGGAAATATCGTGCGCT		TATTATT	LRLPSDTPLAYFHSPV
	CCACTTAAGCCGCAGCAGCGCATGAAGATCTTGAAAGATAATCTACTCC		ATTATTA	AAGGLGILHLSSSVPF
	CTCGACTCCTATACCCCCTGACTCTAGGAATAGTACACCTGAAAACCCTG		CTATTAT	HRRKRLETLLSSPNRL
	AAGTCAATGGACCGAAATATCCACACGGCCATAAGGAAATGGTTGCGG		TACTTCT	LHKLPTSPTLASYSHL
	CTACCCTCCGACACCCCGCTAGCATATTTTCACTCACCCGTCGCTGCCGG		ACAATT	SQLPVRIGHETVTSRE
	AGGCCTAGGGATCCTCCATCTGTCCTCATCGGTTCCATTCCACCGTCGAA		AATATTA	EASNSWVRRLHSSCD
	AACGTCTAGAAACCCTCCTATCTTCACCGAACCGCCTACTGCACAAGTTG		TGGCTA	GKGLLLAPLSTESHA
	CCAACTTCCCCAACACTAGCTTCTTATTCACACCTTAGTCAACTGCCAGTT		CTCCTCT	WLRYPQSIFPSVYINA
	CGAATTGGGCACGAGACCGTAACGTCTAGAGAAGAGGCTTCCAACAGC		CAGCAC	VKLRGGLLSTKVRRSR
	TGGGTGAGACGATTACATTCGTCCTGCGACGGGGAAGGGACTACTCCTA		ACCAAT	GGRVTNGLNCRGGC
	GCACCACTAAGCACCGAGTCCCATGCATGGCTGCGCTACCCCCAGTCTA		AAAATA	AHHETIHHILQHCALT
	TTTTTCCATCTGTTTACATCAACGCCGTTAAATTACGAGGTGGCTTACTAT		TCAATCA	HDIRCKRHNELCNLV
	CCACCAAAGTCAGGAGATCTCGCGGAGGTAGAGTGACGAATGGCCTGA		AACATCT	AKKLRRQKIHFLQEPC
	ACTGTCGAGGCGGTTGCGCCCATCATGAAACGATCCACCACATTCTGCA		CAATTAT	IPLEKTYCKPDFIIIRDS
	ACATTGCGCGCTCACCCACGACATCAGATGCAAACGCCATAACGAACTA		ATCCACC	IAYVLDVTVSDDGNT
	TGCAACCTTGTGGCAAAGAAACTGCGTAGGCAAAAAATCCATTTCTTAC		TATTAAA	HASRLLKISKYGNERT
	AGGAGCCCTGCATTCCTCTAGAAAAAACCTACTGCAAACCTGATTTTATA		CTCTCTC	VASIKRFLTSSGYIITS
	ATTATACGTGACTCAATTGCTTATGTTCTAGACGTCACTGTATCGGACGA		TATTTCC	VRQTPVVLTFRGILDR
	CGGAAACACCCACGCCAGCCGCCTGTTAAAAATATCAAAATACGGCAAT		CCTGAG	ASSQSLRRLCFSSRDL
	GAGCGAACCGTCGCGTCGATCAAGCGTTTCCTCACATCCAGTGGATATA		TTATAAA	GDLCLSAIQGSIKIYN
	TCATTACCAGTGTTCGACAAACACCAGTCGTCCTTACATTCAGAGGTATT		CTTACAA	TYMRGT (SEQ ID
	CTGGATAGAGCAAGTTCACAATCCCTACGACGCCTATGTTTTTCATCCCG		TTCAGTC	NO: 1453)
	TGACCTCGGTGACCTTTGCCTGAGTGCGATTCAAGGCTCAATTAAAATAT		TAACCG	
	ATAATACCTATATGAGAGGAACCTAACGGCTGAACGAATAGCCCCCTTC		AATATCT	
	ACTCTTAGACATTCCCCCACTGTTGTTGCTTATCTTCATGCTCTTGTGTTA		CTCTTTT	

			ATTGACTGCTCTCTTCTGGGTTGACGTCTGATTGTCTCTCTCTCTTTCCAT ATTGCTTGCTCTGCCCGCTTACTTCCAATAGTTGTCATATTATGTCTTTGT TTACTTGCCATGTCTAACGACAATTACTTTATCTACCTTAGTTTGTCCTCTT GGTTTCGATTGCCTTCATATGTTCATGGCGGAATCTGATGTTTATAATGA CTATTCCTATTACCACCACTACAACTACTATTATTATTTTCATTACTATTAA CATTATTATAAACATTATTACTATTATTATTACTACTATTATTACTTCTACA ATTAATATTATGGCTACTCCTCTCAGCACACCAATAAAATATCAATCAAA CATCTCAATTATATCCACCTATTAAACTCTCTCTATTTCCCCTGAGTTATA AACTTACAATTCAGTCTAACCGAATATCTCTCTTTTACAAATCTTAAGTAT GTAATTTTGTGCCAAACCCATTTGGGTCTGTACAATTTGATACTTAAAAA TAAATGTTATTAGCC (SEQ ID NO: 1085)			ACAAAT CTTAAGT ATGTAA TTTTGTG CCAAAC CCATTTG GGTCTG TACAATT TGATACT TAAAAA TAAATG TTATTAG CC (SEQ ID NO: 1331)	
R2	R2-1_BTe	.	Bombus terrestris	TCAATAGTTACTCGGGGGCAGGCGGGATATTGGTCTTGCCTTGCCCAAG TCACACTCCTACCTCCTCGTGGTACCGCCGGTAACACGCGCACGTCCACA TCAGCGAGGGGCGTACTCCCCCGGATGTGGCGGCGCGTGGCTAAACGG AGTGTGGCGACGAAGGAGCGAAAGACTAACAACTATAACGGTCTTCCG TAACGGCTACTTGGAGCCGTGAATAATGGAGCCTATATTAAACCCTGGA ACTCGTTCCTTCGTTCTGTTGACGACTGGAACGGCAAAGGACATGATTT GGATAACAATTGGAAACTTAATATCGAATTCACAATTAAATTGCAAAATT CGATAATAATACGAATAGTGCCTCGGACGCAGCGCCTCTCAGTCCCGGA GGCGCTGTGGCGGATTTATCCGCTTCAGAGGGGACTACGGATAACGAC CAGGCTATGTCCCCGGCTATGAGTTTGATNACGGTGCCGCTCGTGGGCA ATCGGGTCGCCTGTCCAGNATGTGAGAAAAGGGAGGCTAACCTTTTCTT CTTAAATCTGTCGGACCTGGACAGGCACCTAACACAACACCACCCGGAT GCCCCGATCTNTTGGTCNTGTATCGACTGCGCGAAATGCTTCCCGAAGC TTCATGGGGCTAGATGCCACATACCTAAGTGTGGTGGCGCTAGCAGCCA GGCCAGGACAGGNGAGTTCCAGTGCGAGGCCTGTCCCATGAGCTTCGG GTCGCGCAGAGGGCTATCCACCCACGAACGGCATGCGCACCCTGCCGTC AGAAATATTAAAAGAAGGGGAGCGGACCCCCCAGAAGAAAATACAAAA TCTTGGAAAGTAGAANNGGTAGCACGCTTGAAGGGGCTATGGGAAATA TTCAAAGACCATAAATATCCCAACAAAGAAATCAGCAAATTCCTCACCAC NAAAACGGTNGATCANATNAAATATCAAAGAAAAAAGCTAAACTTAAT TGGTGNGGAAAGCCCCCAAGAGGCTACCTCACTGGCAACAGAGGGAG GGTGCGATCTCGTTAGTTCAGGCAATGCCAGTTTTGGCTCGCCTGTAGG CCGCAACGAGAACGAAGAAGAGCTCATCCACGAATGGAAGCTCTCGCT	TCAATAGT TACTCGGG GGCAGGC GGGATATT GGTCTTGC CTTGCCCA AGTCACAC TCCTACCT CCTCGTGG TACCGCCG GTAACACG CGCACGTC CACATCAG CGAGGGG CGTACTCC CCCGGATG TGGCGGC GCGTGGCT AAACGGA GTGTGGC GACGAAG GAGCGAA AGACTAAC	TGAAAA CACGAT AACGAT TATGAA TCAAAA TAAGAA AGTAAA CATCCCA GAAATT GTCTAC GTCTTAT TTGTTAT CTATTTA TTTGTTT A (SEQ ID NO: 1332)	IAKFDNNTNSASDAA PLSPGGAVADLSASE GTTDNDQAMSPAM SLXTVPLVGNRVACP XCEKREANLFFLNLSD LDRHLTQHHPDAPIX WSCIDCAKCFPKLHG ARCHIPKCGGASSQA RTGEFQCEACPMSF GSRRGLSTHERHAHP AVRNIKRRGADPPEE NTKSWKVEXVARLK GLWEIFKDHKYPNKE ISKFLTTKTVDXXKYQ RKKLNLIGXESPQEAT SLATEGGCDLVSSGN ASFGSPVGRNENEEE LIHEWKLSLKNEINKP TEVPPILKEVYNRLML IWEEHQDDRDSLTES LDHFIRTALYELINKIN KNQTDLKTKRAAKTK SPKNNRNSRKRFSYA

		AAAAAACGAAATAAATAAGCCAACCGAGGTGCCCCCTATTTTAAAGGAG	AACTATAA		RCQELFHECPRRLAD
		GTTTACAATCGGCTGATGTTAATCTGGGAAGAGCACCAAGATGATCGAG	CGGTCTTC		AVVNNDQAYLEPAR
		ACTCTCTAACGGAGAGTCTCGATCACTTTATACGCACAGCTCTATACGAG	CGTAACGG		QPPGSEEVRGLYEKL
		CTTATAAACAAAATAAATAAAAACCAAACGGATTTAAAAACTAAAAGAG	CTACTTGG		WGQVGSTYVPAPVT
		CAGCTAAAACAAAAAGCCCAAAAAATAATAGAAACTCCAGGAAGCGAT	AGCCGTGA		RVPKLSLSEIFPPIAAE
		TCTCATACGCTCGTTGCCAGGAGTTATTTCACGAATGCCCAAGGAGACT	ATAATGGA		DVGERIGKIRKKAAA
		GGCTGATGCCGTGGTCAATAACGATCAGGCATATCTCGAACCAGCCAG	GCCTATAT		GPDGLQRDHLTIPGL
		GCAACCTCCCGGATCAGAGGAAGTGAGGGGGCTCTACGAAAAGCTGTG	TAAACCCT		PIIMAKIYNILVYCSYF
		GGGACAAGTGGGTTCCACGTACGTCCCGGCTCCAGTTACGAGAGTCCCC	GGAACTCG		PSAWKENRTTLIPKIN
		AAACTATCCCTATCCGAGATCTTCCCGCCGATAGCGGCTGAGGATGTGG	TTCCTTCG		KPCSLVENWRPITISP
		GGGAGAGAATCGGCAAAATAAGAAAGAAAGCTGCGGCAGGACCGGAT	TTCTGTTG		ILGRIFSSIIDGRIRRGT
		GGATTACAAAGAGATCACCTAACTATCCCCGGCCTGCCTATCATAATGG	ACGACTGG		VLNMRQKGFTSENG
		CGAAAATCTATAATATACTCGTATATTGTTCTTATTTTCCCTCCGCATGGA	AACGGCAA		CKINIELLNSALNYSKR
		AGGAGAATAGAACAACTCTTATTCCAAAAATAAATAAGCCATGCAGCCT	AGGACATG		NSGGIFTIVDISKVFD
		GGTCGAGAACTGGCGACCAATTACTATTAGTCCGATTCTAGGCCGAATC	ATTTGGAT		TVPHAALKPCLAKKG
		TTCTCCTCCATTATCGACGGGAGGATAAGAAGAGGCACTGTATTAAATA	AACAATTG		VPALIVDLIDEMYKN
		TGAGGCAGAAGGGCTTTACATCTGAAAACGGATGTAAAATTAACATCGA	GAAACTTA		VKTTIKTKDGGVEIMI
		ATTATTAAATTCTGCCTTAAACTATAGTAAAAGAAATAGCGGCGGGATA	ATATCGAA		RRGVKQGDPLSPLLF
		TTCACTATCGTGGATATCTCGAAAGTTTTTGATACAGTGCCTCATGCAGC	TTCACAAT		NLCLEPLLEEIEEQAS
		CTTAAAACCTTGTCTGGCAAAAAAGGGTGTGCCCGCCCTTATCGTCGATT	TAA (SEQ		GINVSEHRKVSVLAF
		TGATCGACGAAATGTATAAAAACGTAAAAACTACTATAAAAACTAAAGA	ID NO:		ADDIVLLGADAREAQ
		TGGCGGGGTCGAGATTATGATCCGCCGAGGAGTTAAGCAAGGCGACCC	1209)		HQINVLTDYLQSLM
		CCTATCGCCCTTACTCTTCAATTTATGCTTGGAGCCACTGCTGGAAGAGA			MNLSIEKCQTFEVVA
		TCGAGGAGCAAGCCAGTGGAATAAATGTTAGCGAACATCGTAAAGTAT			KKDTWFIKEPGLKIG
		CAGTTCTGGCCTTCGCAGACGATATCGTTCTACTTGGAGCGGACGCGAG			NQIMPTVDPDEAFKY
		GGAAGCGCAACACCAAATAAACGTACTTACCGACTACCTGCAAAGCCTC			LGAKIGPWKGVHCG
		ATGATGAATCTATCAATTGAAAAATGCCAAACCTTCGAGGTTGTGGCCA			VIVPELLSVVKRVRKL
		AAAAGGATACCTGGTTCATTAAGGAACCTGGACTTAAAATCGGGAATCA			SLKPGQKVELLTKYIF
		AATAATGCCTACTGTAGATCCCGACGAGGCTTTCAAATACCTGGGCGCC			PRYIYHLLVSPPSDTV
		AAAATTGGTCCCTGGAAAGGCGTTCATTGTGGTGTAATTGTTCCAGAAC			LKLLDSEVRQEVKTIL
		TTCTGAGCGTGGTGAAAAGGGTGAGGAAACTCTCCCTCAAGCCGGGCC			HLVPSTATGFFYTPKA
		AGAAGGTGGAACTTCTAACTAAATACATCTTCCCCCGCTACATTTACCAT			CGGIGIPRFEHIIKLGT
		CTACTTGTAAGTCCGCCAAGCGATACCGTCCTTAAACTACTAGACAGCG			LKSAIKIANSIDPAVA
		AGGTCAGACAAGAAGTTAAAACTATCTTACATCTTGTGCCTTCCACTGCC			GLIDDAAIKKLKQTA
		ACAGGCTTCTTTTACACTCCAAAGGCCTGTGGAGGAATAGGAATACCGA			NSLRINWPASLEDIEK
		GGTTTGAACATATAATCAAACTCGGTACCCTTAAAAGTGCAATAAAAAT			ARKRLRKEHISQWAD
		CGCAAACTCGATCGATCCAGCGGTCGCTGGCCTAATCGATGATGCAGCC			LKCQGQGVPDFIKNK
		ATAAAAAAGTTAAAGCAAACGGCCAACTCCTTGCGGATCAATTGGCCAG			TGNLWLEDHSLLKPS

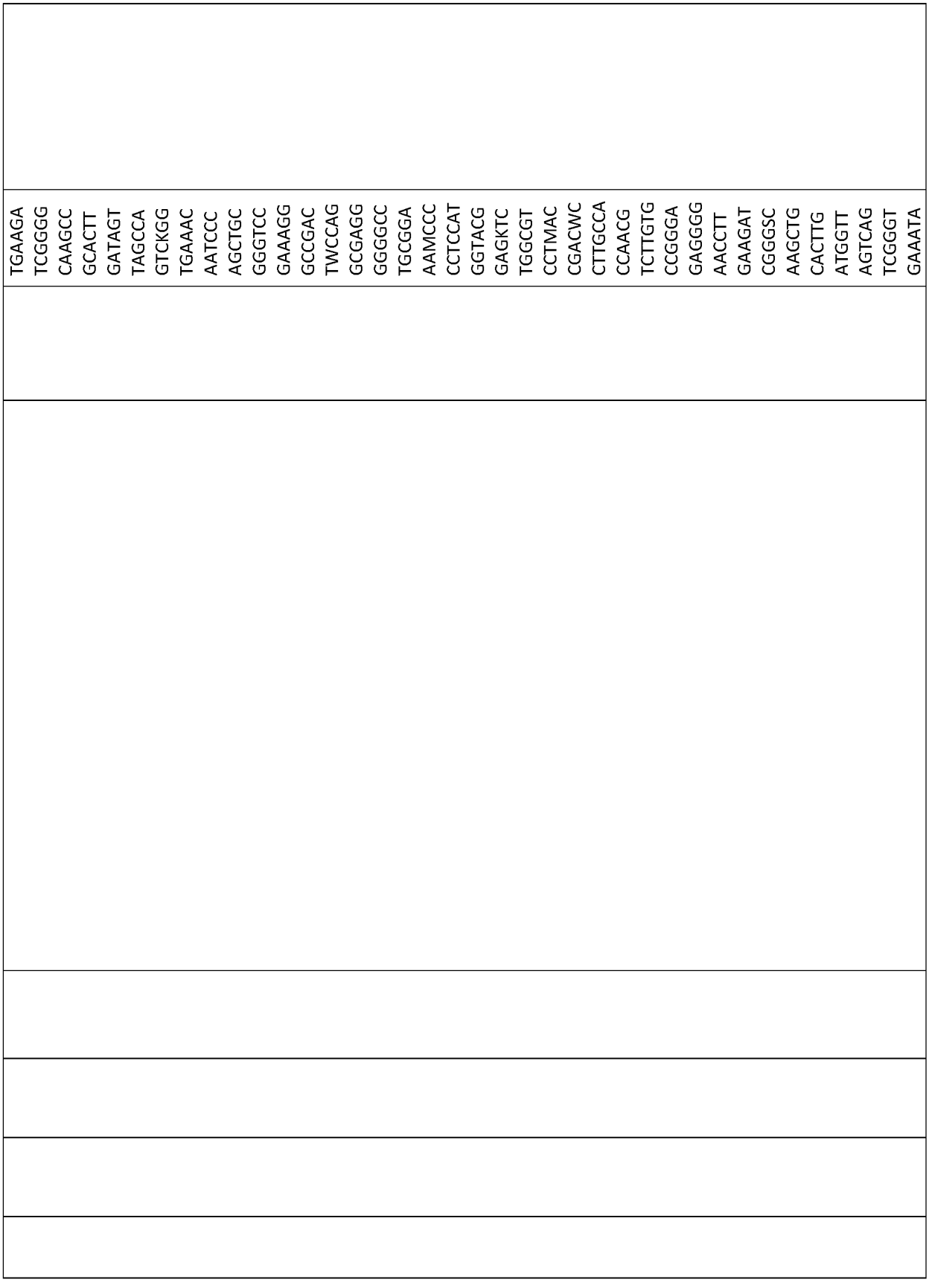
			CCTCCTTGGAGGATATTGAAAAAGCTCGTAAACGTCTAAGGAAAGAGCA				RLIDALRLRTNTFGTR
			TATCAGCCAATGGGCTGATTTAAAATGCCAGGGACAAGGCGTTCCTGAT				SVLARADKNIDVTCR
			TTTATCAAAAATAAAACTGGCAACCTGTGGCTTGAGGACCATAGTCTAC				RCRAQPETLGHILGLC
			TCAAGCCATCGAGACTCATCGATGCCCTTCGATTAAGAACAAATACCTTT				QHTKGLRIKRHDEVK
			GGTACAAGATCGGTGCTGGCACGGGCCGACAAAAATATTGATGTAACA				SLLEGRLKSKKNNEVF
			TGTCGAAGATGTCGTGCCCAGCCCGAGACCCTTGGACACATACTCGGGC				VEPTIKAGGSLFKPDL
			TATGTCAGCACACTAAAGGCTTAAGAATCAAAAGGCACGATGAGGTTAA				VIKNGERVLVVDVTV
			ATCCCTCCTCGAAGGAAGATTAAAAAGTAAAAAGAACAACGAAGTATTT				RYENKNYLALAEKEKI
			GTAGAGCCGACGATTAAGGCCGGAGGCAGTTTATTTAAACCAGACCTC				EKYRPCLRALKEIFNA
			GTAATTAAAAAACGGGGAAAGGGTTCTCGTGGTCGACGTAACTGTCCGCT				KGGEILPVVLGSRGTI
			ACGAGAACAAAAACTACCTGGCCTTAGCCGAAAAAGAAAAGATAGAGA				TPNTEKVLKRLGIAN
			AGTATCGGCCATGCCTAAGGGCATTAAAAGAAATATTTAACGCCAAAGG				NDIKTILLNVLRSSIEL
			AGGAGAGATTCTTCCGGTGGTCTTGGGTAGTAGAGGCACTATCACGCCT				CNIFIDD (SEQ ID
			AACACCGAGAAAGTCCTCAAGCGATTGGGAATCGCTAACAATGATATAA				NO: 1454)
			AGACAATACTCTTAAATGTATTAAGAAGCTCGATAGAGCTGTGTAATAT				
			ATTCATTGATGACTGAAAACACGATAACGATTATGAATCAAAATAAGAA				
			AGTAAACATCCCAGAAATTGTCTACGTCTTATTTGTTATCTATTTATTTGT				
			TTA (SEQ ID NO: 1086)				
R2	R2-1_Crp	.	Crocodylus porosus	AGGCGTCTCCTTTAAGGGCAACGGTCTGGTTACGCGGTCGCAGCAGKCT	AGGCGTCT	TGAATC	VPPGAEARGRYHHP
				TGCKCCAGGTACCTCCWCGTGGTTCCCGCCGGGTGCCGMAGMCCCAG	CCTTTAAG	CCACTCT	RXEXARQGEPPSXRV
				GSCTGTCGGTAGCTCGATCCTGGCACAGTAMGGCCAGGGGAGTKCTTC	GGCAACG	GGGGAC	FLVXLPDSNPPCPICG
				CTTGCTGCWGGTGCCCCACAAGCGTKTGGCAGSMCMCCTGCTTCWTC	GTCTGGTT	CCCCAA	DHVXXXSVLALHCVE
				GCAAMAATMASAGKGTSMTCAGTAGTCGGCCCCGCCGCTAGCCAAAA	ACGCGGTC	AAATTA	GHXWAXVQYQCTH
				CTGTTCGCCACSCAGTTACAGATGGTGCCTGTGCTGACCKGKKGCCCCGT	GCAGCAGK	GAAAAA	CGILCHIPRCQGRVXE
				GGGCTCGGWGGTGCAGGGGWGGCCGGCCCATGGCTGGGCCAGACST	CTTGCKCC	CCCAAA	XTGKDXXCPECPASF
				GGGCCKTGGAGCCCGCTCCCAMCCCAGAGTTCCCCCTCAAMGCCGCCA	AGGTACCT	ACAGTT	DEKAGLSQHKRHTVT
				GGKCAGMAMCAKCCGGGGAGKGMCACGGCCCCGGCCGTGAGTCCCC	CCWCGTG	GTGTTTA	XSXERVAGXLLRAXLR
				CCCGGTGCWGAGGCCCGGGGCCGATACCACACCACCCCCGGTGMGAGM	GTTCCCGC	AGTGTG	HGCWSVEEEETLTRL
				GCGCCCGCCAGGGTGAACCACCCTCCAMCAGGGTTTTCCTGGTGASGCT	CGGGTGCC	TTCTTGT	DAMFXGARNINQLIA
				GCCTGACTCAAACCCKCCGTGCCCGATTTGCGGGGACCACGTGSGTWA	GMAGMCC	TTGTCCC	AEXVSKMXKQISDK
				ACWCTCCGTGCTGGCCCTCCACTGCGTGGAGGGCCACGWGTGGGCGS	CAGGSCTG	TTTGGCT	WRXLXLXPEQTTXGG
				ATGTGCAGTACCAGTGCACCCATTGTGGCATCCTCTGCCACATCCCACGT	TCGGTAGC	TCACCTC	XAESASVVXXESMTP
				TGCCAGGGGAGGGTGCMKGAGCWGACCGGGAAGGACTKGGMTTGTC	TCGATCCT	CAAGTT	EMEAQSPAXPPGKIR
				CTGAGTGCCCTGCCTCCTTCGATGAGAAAGCWGGCCTATCGCAGCACA	GGCACAGT	GCGATC	KIFTGQDGHAGGXA
				AGCGGCACACGGTAACACAMAGTGWKGAACGGGTTGCAGGKWACCT	AMGGCCA	CCCCATC	WENQEDFHWTRWA
				GTTGAGGGCTKGTTTAAGGCACGGGTGCTGGTCGGTGGAAGAAGAGG	GGGGAGT	TCCCCTG	RRWLKRGQXLSDKV
				AAACTCTCACCCGCCTWGACGCGATGTTCCSTGGCGCCCGGAACATCAA	KCTTCCTT	CGCTGT	QEVLGXWVEGQPRI
				CCAGCTGATCGCTGCTGAGMTGGTWTCGAAGATGCSTAAGCAGATCAG	GCTGCWG	CTTTCTG	XAWVDXVSLDVLTLF

	CGACAAATGGAGGTWGCTCKSTCTGWGTCCTGAGCAGACCACATWGG	GTGCCCCA	AATGAC	LGVPPGPQRAPSKKG
	GTGGTGAKGCTGAGTCKGCCTCCGTGGTGSAGGWWGAGTCCATGACT	CAAGCGTK	CAGTGG	PXEGGKPTSWMNKR
	CCGGAGATGGAAGCCCAAAGCCCAGCTGAMCCGCCTGGGAAAATCAG	TGGCAGS	TGTTGA	AIKWGTFLRYQHLFG
	GAAGATTTTCACTGGACAAGATGGGCACGCCGGTGGCTGAAGAGGGG	MCMCCTG	GGCTGG	ANRKLLAAXILDGAX
	GCAGGKTCTGTCAGACAAGGTGCAAGAAGTCCTGGGCKCGTGGGTGGA	CTTCWTCG	TGTGAC	RNQXTLLLEEVXQXY
	GGGTCAACCCAGGATTCMTGCCTGGGTCGACKCMGTCTCCCTTGATGT	CAAMAAT	CTCGGT	XGKWEAEPPFEGLG
	TTTGACTTTGTTCTTGGGGGTGCCCCCGGGACCGCAGAGGGCTCCGAGC	MASAGKG	CACCTCC	RFGXXRDVDSFAFEA
	AAGAAGGGGCCCAMGGAGGGTGGCAAACCWACGTCCTGGATGAACA	TSMTCAGT	AAGGCC	LITXEEAVKHMMXM
	AACGTGCCATCAAATGGGGCACATTCCTTCGGTACCAGCACCTGTTTGG	AGTCGGCC	AAGTGC	AXXSAPGPDKLTLRD
	TGCCAACAGGAAGCTCCTGGCGGCCMTCATCCTGGATGGTGCTGASCG	CCGCCGCT	CCTGGC	LRRADPEGDALAELF
	TAATCAGTGSACCCTCCTGCTWGAGGAGGTCTTKCAGGMCTACCKAGG	AGCCAAAA	CCCGAG	SLWXITGVVPDRLKE
	AAAGTGGGAGGCCGAGCCGCCTTTCGAGGGTCTTGGACGGTTTGGGST	CTGTTCGC	TAGGAC	XQXVLIPKAVDFEKLR
	CCAMMGGGATGTGGATAGCTTCGCATTCGAGGCCCTGATCACTCKGGA	CACSCAGT	CAGGTG	QLGNWRPITIXSIVLQ
	GGAGGCTGTCAAGCACATGATGGMAATGGCCWMSAAMTCGGCTCCA	TACAGATG	GCCCAG	LXSRVLTARLTAACPI
	GGTCCGGATAAGCTCACCCTGAGAGATCTGCGCCCGCCGCKGACCCCGAG	GTGCCTGT	CTTGCT	XPHQQGFISAPXCAE
	GGAGATGCTCTTGCCGAACTGTTCAGCCTGTGGMTGATCACTGGCGTG	GCTGACCK	GGGCAC	NLKXPELIXRKVKXDR
	GTCCCGGACAGACTCAAGGAGTKTCAGTSTGTGCTGATACCAAAGGCAG	GKKGCCCC	CCGTCA	RPLGVAFVDXARAFD
	TGGACTTTGAGAAGCTGAGGCAGCTGGGCAACTGGCGCCCCATCACTAT	GTGGGCTC	CCGCCC	SVSHDXISWVLKAKG
	CGSGTCCATTGTKTTGCAGCTATTMTCACGGGTGCTGACCGCACGGCTC	GGWGGTG	AGGGCA	VDQHIVNLIEDSYQK
	ACCGCAGCWTGCCCCATAASCCCCCACCAGCAGGGKTTCATMTCGGCA	CAGGGGW	AATGGA	VTMRVQVFSGSTPPI
	CCKKGGTGTGCCGAAAACCTGAAGMTGCCTGAGCTCATCTTSCGSAAGG	GGCCGGCC	AGGGAT	SIKXGVKQGDPMSPL
	TGAAGCAMGACAGGCGGCCGCTGGGGGTGGCATTCGTGGATMTAGCS	CATGGCTG	CTATCCT	LFNIAMDPLIXKLKTV
	AGGGCSTTCGACTCGGTATCGCACGACMACATTTCWTGGGTCCTAAAG	GGCCAGAC	GACCAC	RQGVKVGSASLTTLA
	GCCAAAGGGGTGGACCAACACATCGTGAATCTCATWGAAGATTCATAC	STGGGCCK	TACCAG	FADGLXLLXDSWEG
	CAGAAGGTTACCATGAGAGTACAAGTGTTCAGTGGTTCCACCCCWCCC	TGGAGCCC	GCTAAG	MQHNITTSXTPGRAC
	ATCAGCATCAAAWCCGGGGTCAAGCAAGGCGACCCGATGTCMCCCCTS	GCTCCCA	TGTGGT	NTTSRHPRGLLQPHG
	TTGTTCAACATTGCGATGGACCCCCTGATAGSCAAGCTCAAGACAGTCA	MCCCAGA	GCTGCC	PTSATXKMXGVLLES
	GGCAGGGAGTGAAAGTCGGGAGTGCTTCCCTGACCACCCTGGCATTTG	GTTCCCCC	TAGCCT	XMRLLYGEQLRGLED
	CGGATGGCCTCGWMCTGCTGAKTGACTCCTGGGAGGGCATGCAACAC	TCAAMGC	GCCGTA	XRPXXHDAXARRAD
	AACATCACGACATCCTAGAGGACTTCTGCAACCTCACGGGCCTACGAGT	CGCCAGGK	AGGTCA	TISGLEXRSLGWDXQ
	GCAACCCAMAAAATGKCAGGGGTTCTTCTTGAATCCGAMATGCGACTC	CAGMAMC	AGCGCC	TRFGYATXLLAREDG
	CTWTACGGTGAACAACTGCGMGGCCTGGAAGATWGCCGGCCGTGMS	AKCCGGG	CTGCTG	DCXAQTNAEALCWX
	ATCACGATGCTSGGGCCAGGCGAGCCGACACGATATCTGGGCTTGAAK	GAGKGMC	CCGCTC	SXPFPGCAXRPXYAN
	GTCGGTCCCTGGGTTGGGATWGGCAAACCAGATTTGGGTACGCAACTA	ACGGCCCC	GGGTAK	XGWVASEALDSMSR
	AKCTCCTGGCTCGAGAGGATGGGGACTGCCWCGCTCAAACCAATGCAG	GGCCGTGA	CAGTCCT	RXVKEWFHLPACTDX
	AAGCTCTTTGCTGGTSCAGTATKCCATTCCCAGGCTGCGCTMTCAGGCC	(SEQ ID	CGTTCAC	LLXSRHRDGGLGLLRL
	GWTTTATGCGAACMCCGGTTGGGTGGCCTCGGAGGCTCTGGATTCCAT	NO: 1210)	TCGTCCC	ARXXLAAXVRRPIRV
	GAGCCGAAGAAAKGTGAAGGAGTGGTTCCACCTACCCGCCTGTACCGA		TCCTAGT	ATSSDEVTRKVSYAC

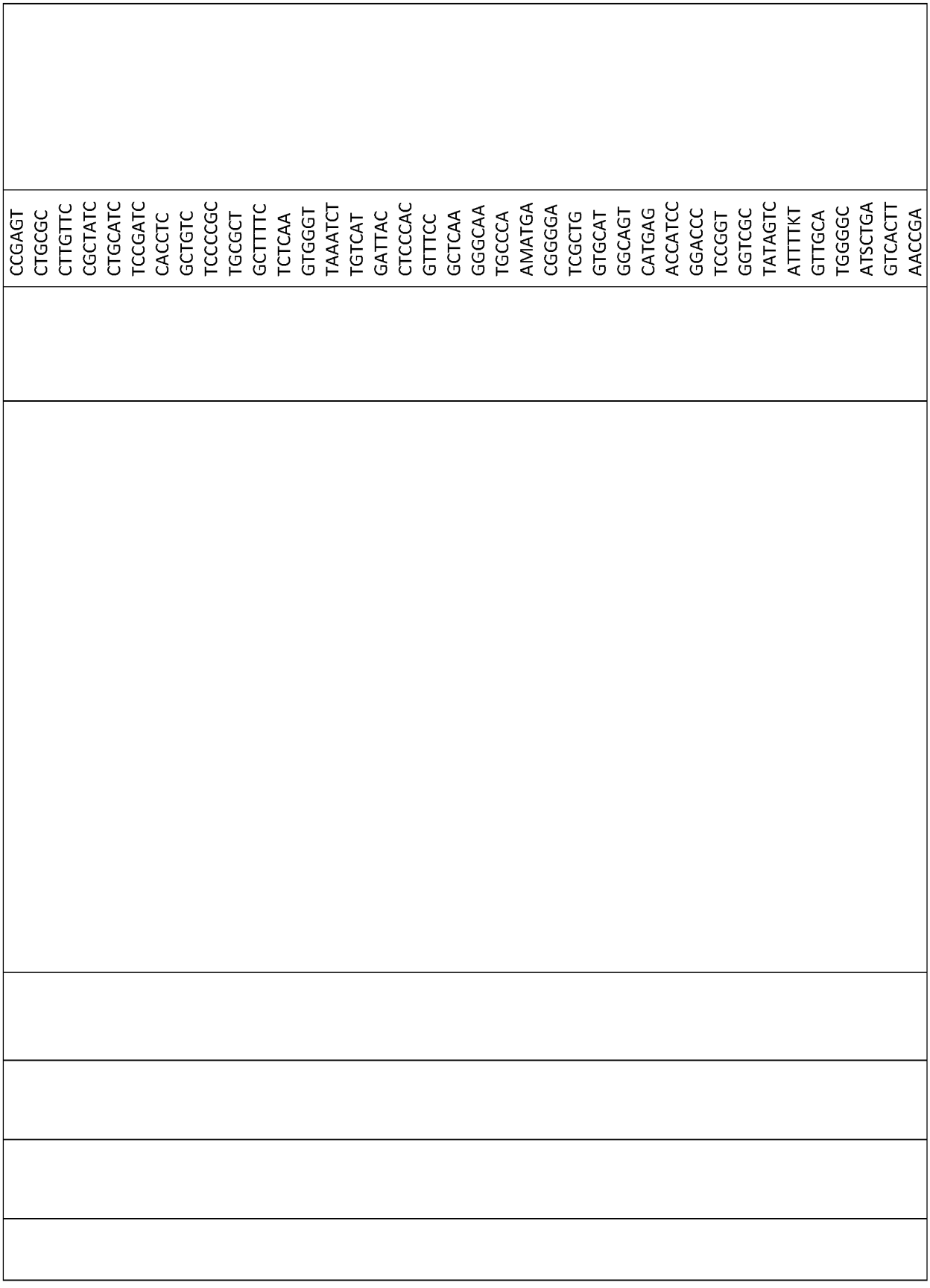
		CKGCCTCCTCCASTCCCGCCACCGGGATGGGGGCCTKGGGTTGCTGCGC	ACCCTCC	GISDEVERLXLAXGG		
		CTGGCGAGAGMCWTTCTGGCAGCGSAAGTGAGGAGGCCGATTCGTGT	GCMTCT	DXSNVPRFEDPXAPK		
		AGCCACCTCATCTGATGAAGTCACGCCGGAAAGTATCCTACGCATGTGGG	GCGCTT	SXXVQGPHEAAQET		
		ATAAGTGATGAAGTGGAGCGKCTCTGKTTGGCMAGWGGGGGAGACAT	CTGCTAT	PRVVRTQAIPWPSN		
		STCCAATGTCCCGAGGTTCGAGGATCCTGAWGCCCCGAAGTCTSCAKGG	CCACTGT	WRAEEHSKWAQLSC		
		GTGCAGGGCCCCCATGAGGCTGCCCAGGAGACACCACGTGTTGTCCGG	GGCCGG	QGERVELFCNDPVSN		
		ACACAGGCGATCCCCCTGGCCCTCCAACTGGAGAGCCGAGGAACACTCCA	WGATGC	GWINSRGQLAERLW		
		AATGGGCCCAACTCAGCTGTCAGGGAGAGAGGGTGGAGTTATTCTGCA	CGAGG	IMALKLRSNIYPTREF		
		ATGATCCAGTCAGCAATGGCTGGATCAACAGTCGGGGACAACTGGCGG	WTGSW	LGRGQAGTNIGCRH		
		AACGCCTCTGGATCATGGCCTTAAAACTGCGTTCGAACATTTATCCCACC	MCAWC	CTHPRETLGHILGICP		
		AGGGAGTTTCTTGGGAGAGGCCAGGCAGGTACCAACATAGGTTGCCGG	CTCGAC	AMQEARILRHNKLCK		
		CACTGTACACACCCTCGTGAAACATTAGGGCATATCCTGGGCATATGTC	ACCTCCA	ILAAEGKNCEWTVYY		
		CTGCCATGCAGGAGGCACGGATCCTCCGGCACACAACAAGCTGTGCAAGA	GGGCCA	EPHLHNAAGELRKPD		
		TCCTGGCAGCTGAGGGCAAAAACTGTGAGTGGACAGTGTATTATGAGC	AGCGCC	LIFVRDGTALVVDITV		
		CGCACCTACATAACGCAGCTGGGGAACTGCGTAAACCTGACCTCATTTT	TTGGCC	WYEGGPATLLSTTAE		
		TGTCCGAGACGGTACCGCTCTGGTTGTGGATATCACTGTGTGGTATGAA	MCAGGT	KATKYLDLNTQIQELT		
		GGGGGCCCCGCAACCCTCCTGTCCACCACTGCAGAAAAGGCCACACAAAT	AGGACC	GAEQVTYFGFPIGAR		
		ACCTGGATCTGAACACACAGATCCAGGAGCTCACAGGGGCTGAGCAGG	TGGCAC	GKWHADNWRVLSEL		
		TCACCTACTTTGGCTTCCCTATAGGAGCCAGGGGAAAGTGGCATGCTGA	CTGCCC	GLSNSRKERVTRLLS		
		CAATTGGCGAGTACTGTCTGAACTGGGATTGTCCAACTCCCGGAAGGAG	AGGGGG	WRALLGSVDMVNIF		
		CGGGTCACACGGCTCCTGTCGTGGCGAGCACTGCTCGGGTCAGTGGAC	CCAGAC	VSKHRQENLLDEHCT		
		ATGGTGAACATCTTTGTATCTAAACACAGGCAGGAAAACCTATTGGATG	ACTGCCT	PAEQVVSSYAS (SEQ		
		AGCACTGCACCCCAGCTGAGCAAGTTGTGAGTTCTTATGCGAGTTGAAT	GCGGCA	ID NO: 1455)		
		CCCACTCTGGGGACCCCCAAAAATTAGAAAAACCCAAAACAGTTGTGTT	AGGGAA			
		TAAGTGTGTTCTTGTTTGTCCCTTTGGCTTCACCTCCAAGTTGCGATCCCC	AGGGAG			
		CATCTCCCCTGCGCTGTCTTTCTGAATGACCAGTGGTGTTGAGGCTGGTG	CCGTCCC			
		TGACCTCGGTCACCTCCAAGGCCAAGTGCCCTGGCCCCGAGTAGGACCA	TGACCG			
		GGTGGCCCAGCTTGCTGGGCACCCGTCACCGCCCAGGGCAAATGGAAG	TTACCG			
		GGATCTATCCTGACCACTACCAGGCTAAGTGTGGTGCTGCCTAGCCTGC	GGCTTG			
		CGTAAGGTCAAGCGCCCTGCTGCCGCTCGGGTAKCAGTCCTCGTTCACT	ATGGTG			
		CGTCCCTCCTAGTACCCTCCGCMTCTGCGCTTCTGCTATCCACTGTGGCC	CTGGCT			
		GGWGATGCCGAGGWTGSWMCAWCCTCGACACCTCCAGGGCCAAGC	AGCCCG			
		GCCTTGGCCMCAGGTAGGACCTGGCACCTGCCCAGGGGGCCAGACACT	CCGTAT			
		GCCTGCGGCAAGGGAAAGGGAGCCGTCCCTGACCGTTACCGGGCTTGA	GTCAAG			
		TGGTGCTGGCTAGCCCGCCGTATGTCAAGCACTCCACACAGCTGCTCGAGT	CACTCCA			
		TCGCTGGCTTCACCTTCATCCCWCCTAGTGCCTTCCGCCTCTGCGCTATT	CAGCTG			
		TTCGTCCCGACTCGTACCTCCCCACCTCTGCGCTACMGCTATCCCTGMA	CTCGAG			
		AGGACCAAGTGGGCAGGGGGSGTTCGCCCCCCGCCCGCAGGAGGCTCGG	TTCGCT			

			CGTATCCGTGGCKTCWTGCCTCCACCGTCTTGTGCCGCTAGAGGGGTAC		GGCTTC
			CTCMGAGACCGGCGCAACACGACCTTGACGSTTAGACAGTAGGGTGM		ACCTTCA
			AACAKCCCTGCTGCAGGCSTGAAGGGCCAAACGGCTGTGCCATGAGAG		TCCCWC
			GGGAACCTTGAAGACCGGGSCAGTCAGCCAGTTAGTCAGTTGGGCGAA		CTAGTG
			ACAATCCCAGCTGCAGGCCCAMMAGGGCTGTCAGGTGAGGGGGTATC		CCTTCCG
			CCCAWCCACCCCCCGCCGCCGACTACGGAGGCAKGAAGTCCCTAGTGA		CCTCTGC
			CTTCKGACCCCCACGTCTTGTGCCGGGAGAGGGGAACCTTGAAGATCG		GCTATTT
			GGGCAAGCCGCACTTGATAGTTAGCCAGTCKGGTGAAACAATCCCAGCT		TCGTCCC
			GCGGGTCCGAAAGGGCCGACTWCCAGGCGAGGGGGGCCTGCGGAAA		GACTCG
			MCCCCCTCCATGGTACGGAGKTCTGGCGTCCTMACCGACWCCTTGCCA		TACCTCC
			CCAACGTCTTGTGCCGGGAGAGGGGAACCTTGAAGATCGGGSCAAGCT		CCACCTC
			GCACTTGATGGTTAGTCAGTCGGGTGAAATAATCCCAGCWGCCCCCGC		TGCGCT
			TGTGACTGCTAAGMCWGGTCCCCAAGGGGCATGAGGCATSTGCGCTG		ACMGCT
			AGCCGGSAGGGGGTGACACMCGGCGATCGGCGCAGCACACAKASTGAAGG		ATCCCTG
			GAGGCACTTGCTGAGACTGCTTCTGAGGCCCCAGACTTGGGGTGGTGC		MAAGG
			AGCCTTGTCTGGGGTATGGTACAGCACCCTACTGCTCCCTTTGGKCAGC		ACCAAG
			AGAATTCGTCCCGACCTCTTACCCACCCGAGTCTGCGCCTTGTTCCGCTA		TGGGCA
			TCCTGCATCTCCGATCCACCTCGCTGTCTCCCCGCTGCGCTGCTTTTCTCT		GGGGSG
			CAAGTGGGTTAAATCTTGTCATGATTACCTCCCACGTTTCCGCTCAAGGG		TTCGCCC
			CAATGCCCAAMATGACGGGGATCGCTGGTGCATGGCAGTCATGAGACC		CCCGCC
			ATCCGGACCCTCCGGTGGTCGCTATAGTCATTTTKTGTTGCATGGGGCAT		CGCAGG
			SCTGAGTCACTTAACCGAAAGACTCWAAATAACTCAAAAGAGGKAMCC		AGGCTC
			TCTGSGGTTCGGTAAA (SEQ ID NO: 1087)		GGCGTA
					TCCGTG
					GCKTCW
					TGCCTCC
					ACCGTCT
					TGTGCC
					GCTAGA
					GGGGTA
					CCTCMG
					AGACCG
					GCGCAA
					CACGAC
					CTTGAC
					GSTTAG
					ACAGTA
					GGGTG

MAACAK
CCCTGCT
GCAGGC
STGAAG
GGCCAA
ACGGCT
GTGCCA
TGAGAG
GGGAAC
CTTGAA
GACCGG
GSCAGT
CAGCCA
GTTAGT
CAGTTG
GGCGAA
ACAATC
CCAGCT
GCAGGC
CCAMM
AGGGCT
GTCAGG
TGAGGG
GGTATC
CCCAWC
CACCCCC
CGCCGC
CGACTA
CGGAGG
CAKGAA
GTCCCTA
GTGACT
TCKGAC
CCCCAC
GTCTTGT
GCCGGG
AGAGGG
GAACCT


TGAAGA
TCGGGG
CAAGCC
GCACTT
GATAGT
TAGCCA
GTCKGG
TGAAAC
AATCCC
AGCTGC
GGGTCC
GAAAGG
GCCGAC
TWCCAG
GCGAGG
GGGGCC
TGCGGA
AAMCCC
CCTCCAT
GGTACG
GAGKTC
TGGCGT
CCTMAC
CGACWC
CTTGCCA
CCAACG
TCTTGTG
CCGGGA
GAGGGG
AACCTT
GAAGAT
CGGGSC
AAGCTG
CACTTG
ATGGTT
AGTCAG
TCGGGT
GAAATA


ATCCCA
GCWGCC
CCCGCT
GTGACT
GCTAAG
MCWGG
TCCCCAA
GGGGCA
TGAGGC
ATSTGC
GCTGAG
CCGGSA
GGGGTG
ACACMC
GGCGAT
CGGCGC
AGCACA
KASTGA
AGGGAG
GCACTT
GCTGAG
ACTGCTT
CTGAGG
CCCCAG
ACTTGG
GGTGGT
GCAGCC
TTGTCTG
GGGTAT
GGTACA
GCACCC
TACTGCT
CCCTTTG
GKCAGC
AGAATT
CGTCCC
GACCTCT
TACCCAC


CCGAGT
CTGCGC
CTTGTTC
CGCTATC
CTGCATC
TCCGATC
CACCTC
GCTGTC
TCCCCGC
TGCGCT
GCTTTTC
TCTCAA
GTGGGT
TAAATCT
TGTCAT
GATTAC
CTCCCAC
GTTTCC
GCTCAA
GGGCAA
TGCCCA
AMATGA
CGGGGA
TCGCTG
GTGCAT
GGCAGT
CATGAG
ACCATCC
GGACCC
TCCGGT
GGTCGC
TATAGTC
ATTTTKT
GTTGCA
TGGGGC
ATSCTGA
GTCACTT
AACCGA

R2	R2-1_DWi	.	Drosophila willistoni	GAAGCTGGGTCGGATGAGCGCAGAAGGGGTGTTCTTTGGAACACTGTA	GAAGCTG	TAGATG	FERRSNSWGYRPLEP
				ATTCATAAGTCGTAAGTCTGATCAAGTCGACTCGAAACCTCCTCGTGGT	GGTCGGAT	TACTAAC	RSVGTESNNNSPRSN
				GTTTCCTGGGTGCTGTTGAGTTCCTAGTCTCTAGGTTCTTTTCAGTAGCT	GAGCGCA	CTCTAGC	ITITSATSRPGDQPRE
				AATTCGAGCGGCGAAGCAACTCTTGGGGTTACCGGCCCCTTGAGCCAA	GAAGGGG	TTTTCTT	AIAVVNLAGEIPCAVC
				GAAGCGTTGGTACAGAATCAAATAATAATAGTCCTCGGAGCAATATCAC	TGTTCTTT	ATACTTT	GRLFNTRRGLGVHM
				TATCACTTCAGCGACTTCACGTCCTGGAGACCAACCGAGAGAGGCTATA	GGAACACT	TGCCTG	SHQHKDELDTQRQR
				GCAGTGGTAAATCTCGCGGGAGAGATTCCCTGTGCAGTATGCGGGCGC	GTAATTCA	CTACCTT	EDVKLRWSEEEAWM
				CTCTTCAATACTAGAAGGGGGCTCGGTGTACACATGTCACATCAACACA	TAAGTCGT	GGCATT	MARKEVELEASGNLR
				AAGACGAACTAGATACGCAACGTCAGCGTGAAGATGTAAACTCCGAT	AAGTCTGA	ACATCTA	FPNKKLAEVFTHRSSE
				GGAGCGAGGAAGAAGCGTGGATGATGGCGAGAAAGGAGGTGGAGCT	TCAAGTCG	AAAAGG	AIKCFRKRGEYKAKLE
				CGAAGCAAGTGGTAATTTGAGATTTCCTAATAAGAAGCTAGCGGAAGTA	ACTCGAAA	TACAAA	QIRGQSTPTPEALDSI
				TTTACTCACCGTAGCTCCGAAGCAATTAAATGTTTTCGGAAGAGGGGTG	CCTCCTCG	CATCGC	TSQPRPSLLERNHQV
				AATATAAGGCAAAACTGGAGCAGATCAGAGGGCAATCTACTCCCACCCC	TGGTGTTT	ATTGTCA	SSSEAQPINPSEEQSN
				AGAAGCGTTGGACTCTATTACCTCACAGCCTCGCCCTAGTTTACTCGAGC	CCTGGGTG	TAAAGA	WEIMRILQGYRPVEC
				GAAACCACCAAGTATCATCGTCGGAAGCGCAACCAATCAATCCATCAGA	CTGTTGAG	GGTGGT	SPRWRAQVLQTIVD
				AGAACAGTCGAACTGGGAAATCATGCGGATACTACAGGGCTATCGCCC	TTCCTAGT	TTTAGTA	RAQAVGKETTLQCLS
				CGTAGAATGTAGTCCCCGGTGGAGAGCCCAGGTCTTGCAAACTATCGTA	CTCTAGGT	CGTAGG	NYLLEVFPLPNEPHTI
				GATAGGGCGCAGGCCGTAGGGAAGGAAACCACTCTCCAATGCTTATCC	TCTTTTCA	CGCTGT	GRSNLRRPRTRRQLR
				AACTATCTCCTGGAAGTATTTCCATTACCAAACGAACCACACACCATCGG	GTAGCTAA	GGGACT	QQEYAQVQRRWDK
				TCGGAGCAATTTGCGAAGACCTCGAACTAGGAGACAGTTAAGACAACA	(SEQ ID	TCATTGT	NTGRCIKSLLDGTDES
				AGAGTACGCACAGGTTCAGCGTCGTTGGGATAAGAATACTGGGAGATG	NO: 1211)	CCCGGT	VMPNQEIMEPYWK
				CATTAAATCCTTGCTTGATGGAACAGATGAGTCGGTTATGCCAAACCAA		GATGCA	QVMTNPSTCSCDNT
				GAGATAATGGAACCCTATTGGAAACAAGTAATGACGAATCCCAGCACAT		GTGAAT	RFRMEHSLETVWSAI
				GCTCTTGCGATAACACAAGATTCCGTATGGAACATTCGCTTGAGACGGT		CGTGCA	TPRDLRENKLKLSSAP
				TTGGTCAGCGATAACGCCACGCGACCTGAGGGAAAATAAGTTAAAGTT		TACGAG	GPDGITPRTARSVPL
				GTCAAGTGCTCCGGGTCCTGACGGTATCACTCCAAGAACAGCCAGGAGT		ATTGTCC	GIMLRIMNLILWCGK
				GTACCCTTAGGCATTATGCTACGCATAATGAACCTGATTCTCTGGTGCGG		AGTAGT	IPFSTRLARTIFIPKTVT

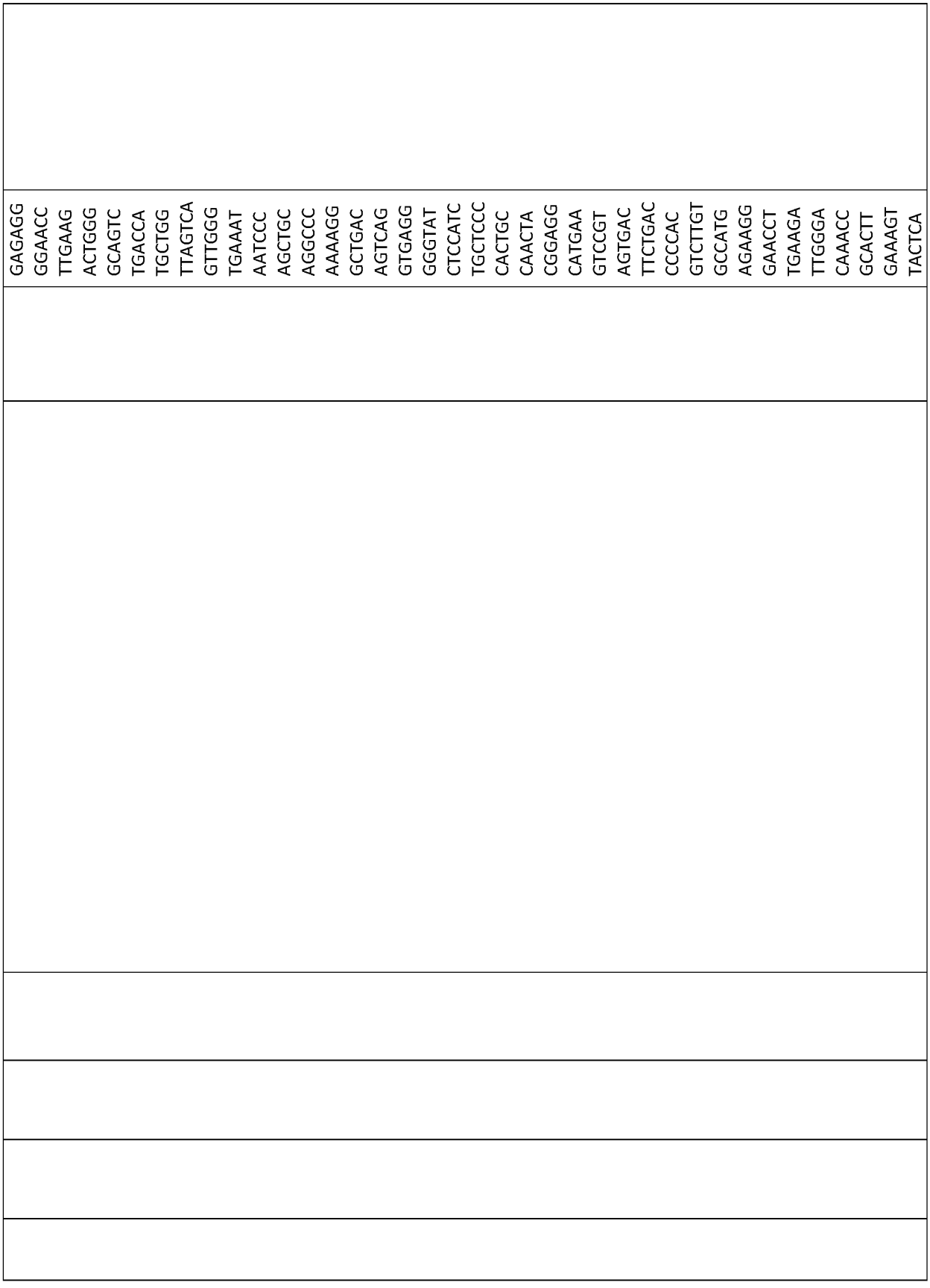
		CAAAATACCATTCTCTACCCGACTGGCCAGAACTATCTTCATTCCGAAGA		TGGTTG	ANRPQDFRPITVPSV
		CTGTGACGGCAAATCGACCGCAAGACTTTCGTCCAATAACAGTCCCCTC		CTCGTAT	LVRQLNAVLASRLAS
		GGTTTTGGTCAGGCAATTAAACGCTGTTCTGGCTTCTCGATTGGCTTCTA		CTTTAGA	KVNWDPRQRGFLPT
		AAGTCAACTGGGATCCAAGGCAGCGCGGTTTCCTACCTACCGATGGGTG		AGATTTC	DGCADNATLVDLILR
		TGCTGATAATGCGACGTTGGTTGATCTCATTTTGCGGGAGCACCATAAA		CTTCCTC	EHHKRWKSCYLATV
		CGGTGGAAGTCATGTTACCTTGCGACGGTGGATGTCAGCAAGGCTTTTG		GGCGAT	DVSKAFDSVSHQAIIK
		ACTCAGTATCACACCAGGCCATTATCAAGACTTTACAGGCCTATGGTGCT		CAAAAA	TLQAYGAPTNFVSFIE
		CCAACAAACTTTGTCAGCTTCATAGAAGAACAGTATAAGGGCGGCGGA		AAAAAA	EQYKGGGTSLNGAG
		ACCTCCCTCAATGGGGCAGGATGGAGTTCAGAGGTGTTTATACCCGCGC		AAAAAA	WSSEVFIPARGVKQG
		GGGGCGTTAAGCAAGGTGACCCTCTGTCTCCACTATTATTTAATCTTATC		AAAAAA	DPLSPLLFNLIIDRLLR
		ATTGATAGATTACTTAGGTCCTACCCCAGAGAGATTGGTGCCAAAGTCG		(SEQ ID	SYPREIGAKVGNTMT
		GAAATACCATGACAAGCGCGGCAGCGTTCGCGGATGATCTGGTGCTATT		NO:	SAAAFADDLVLFAET
		TGCGGAAACTCCGATGGGGCTTCAAACATTGTTGGATACCACGGTAGGC		1334)	PMGLQTLLDTTVGFL
		TTCCTAGCCTCCGTGGGACTCTCCCTTAATGCTGATAAGTGCTTCACTGT			ASVGLSLNADKCFTV
		CAGTATAAAGGGGCAAGCCAAGCAGAAGTGTACTGTCGTAGAACGACG			SIKGQAKQKCTVVER
		GAGCTTTTGTGTAGGTGAGCGCGAGTGTCCTTCATTGAAGCGTACTGAA			RSFCVGERECPSLKRT
		GAGTGGAAGTATTTAGGTATCCGGTTCACTGCGGATGGGCGGGGCTCGG			EEWKYLGIRFTADGR
		TATAGTCCAGCAGACGACCTCGGTCCGAAGCTGTTAAGATTAACAAGAG			ARYSPADDLGPKLLRL
		CCCCTCTGAAACCACACAACAGAAGTTATTTGCACTTAGGACTGTCCTTATC			TRAPLKPQQKLFALR
		CCACAACTCTATCACCAACTAACACTTGGGAGTGTGATGATAGGCGTCC			TVLIPQLYHQLTLGSV
		TAAGAAAATGTGACAGATTGGTACGGCAATTCGTAAGGAGATGGTTAG			MIGVLRKCDRLVRQF
		ATCTCCCACTGGATGTACCAGTTGCTTACTTTCACGCCCCCCACACTTGT			VRRWLDLPLDVPVAY
		GGGGGTCTCGGGATTCCGTCAATTAGATGGATAGCACCGATGCTGCGTC			FHAPHTCGGLGIPSIR
		TGAAGCGATTGAGCAATATTAAATGGCCCCACCTCGAACAATCCGAGGT			WIAPMLRLKRLSNIK
		AGCTAGCTCTTTCATTGACGACGAATTGCAAAGGGCTCGAGATAGATTA			WPHLEQSEVASSFID
		AAGGCGGAAAATGTGCAGCTGTGTTCGCGTCCAGAGATTGACTCGTATT			DELQRARDRLKAENV
		TCGCAAATAGATTGTACATGTCTGTTGATGGTTGCGGTCTCCGTGAAGC			QLCSRPEIDSYFANRL
		AGGTCATTATGGCCCCGCAACATGGATGGGTGAGTCAGCCCACGCGCTT			YMSVDGCGLREAGH
		GCTAACAGGAAAGGAATATTTGCACGGTGTCAAATTGCGGATAAATGCC			YGPQHGWVSQPTRL
		CTACCCTCGAAGTCTCGTACGACGAGGGGAAGGCACGAATTGGAGAGA			LTGKEYLHGVKLRINA
		CGGTGTCGTGCAGGATGTGATGCTCCCGAGACAACAACAAACCACATCTTGC			LPSKSRTTRGRHELER
		AAAAATGCTATCGTACGCATGGGAGGCGGGTAGCTAGACACACAACAGCG			RCRAGCDAPETTNHI
		TAGTAAATGCCGTCAAGCGGGGACTTGAACGGAAAGGCTGCGTTGTCC			LQKCYRTHGRRVAR
		ATGTCGAACCAAGTCTGCAATGCGACTCGGGCTTAAATAAACCGGACCT			HNSVVNAVKRGLER
		GGTGGGAATCCGACACAGAATCACACATTTATGTGATAGACGTTCAGGTTGTG			KGCVVHVEPSLQCDS
		ACAGACGGACATTCCTTAGACCAAGCGCACCAGCGCAAGGTCGAAAGG			GLNKPDLVGIRQNHI
		TACGACAGAGCTGACATAAGATCACAAATGCGGCGATTTTTCGGAGCGA			YVIDVQVVTDGHSLD
		CAGGTGAAATCGAGTTTCATTCCGTTACACTCAACTGGAGAGGAATCTG			QAHQRKVERYDRAD

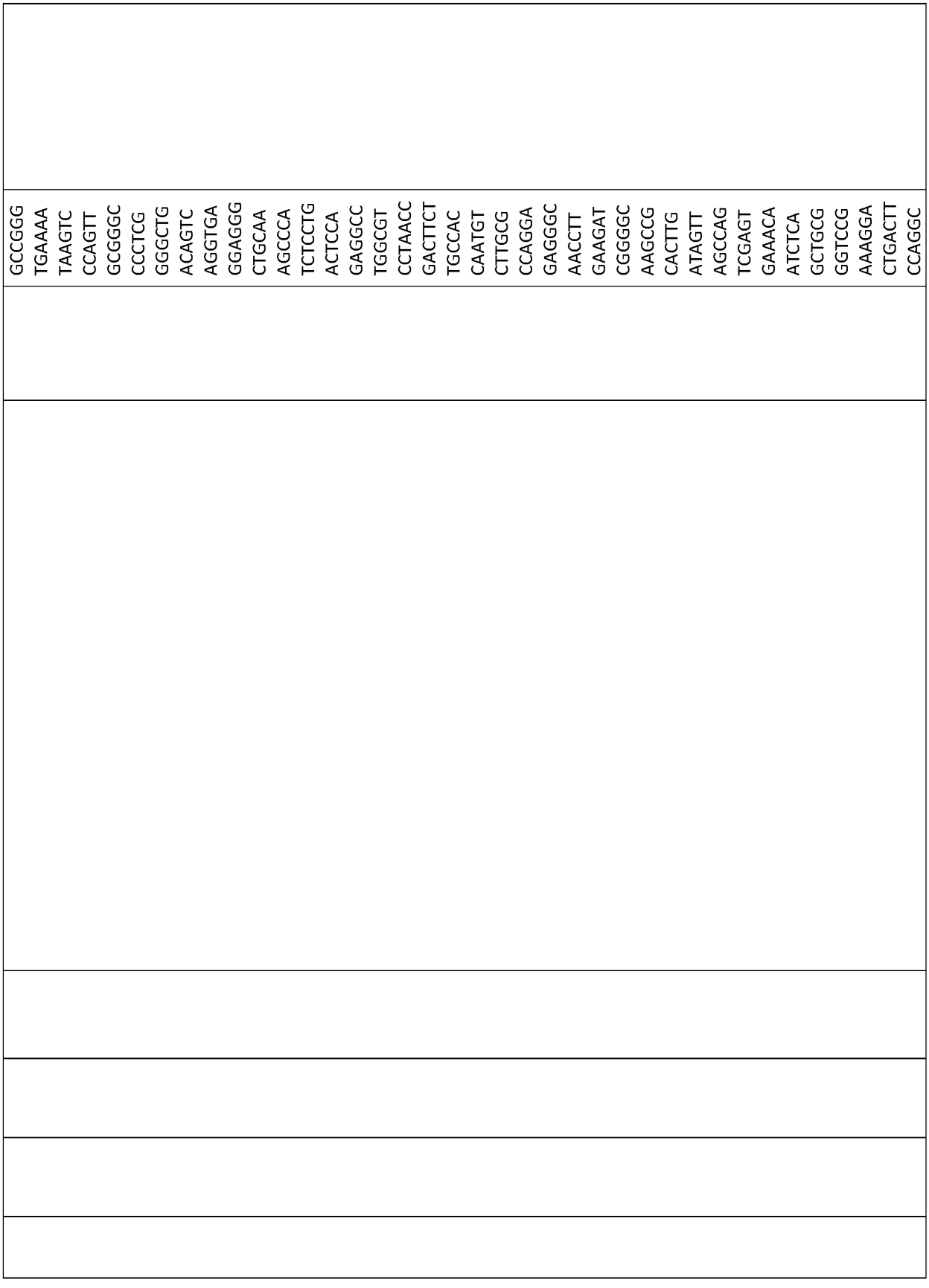
			GAGTGGTCAGTCGGTAAAACGATTGATTGCAAAGGATCTCCTCATCGCT GAAGATACCAAACTCATCAGCGTCAGAGCAGTAAATGGCGGAGTGACG TCCTTCAAATATTTCATGTATTGTGCTGGGTATACTCGAAGCTAGATGTA CTAACCTCTAGCTTTTCTTATACTTTTGCCTGCTACCTTGGCATTACATCTA AAAAGGTACAAACATCGCATTGTCATAAAGAGGTGGTTTTAGTACGTAG GCGCTGTGGGACTTCATTGTCCCGGTGATGCAGTGAATCGTGCATACGA GATTGTCCAGTAGTTGGTTGCTCGTATCTTTAGAAGATTTCCTTCCTCGG CGATCAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1088)				IRSQMRRFFGATGEI EFHSVTLNWRGIWS GQSVKRLIAKDLLIAE DTKLISVRAVNGGVT SFKYFMYCAGYTRS (SEQ ID NO: 1456)
R2	R2-1_Gav	.	Gavialis gangeticus	AGGCATCTCCTTKAAGGGTAATGGTCTGGTTACATGGTCATAGCAGGTT TGTGTCAGGTACCTCCCAGTGGTTCCCGCCGGGTGSCAMAGCCCCAGG GCTGTCGGTAGCTCGATCCTGGTACAGTACGGCCAAGGGAGTTCTTCCT TGCTGTCGGGTGCCTCGCAAGCACKTGGCAGCCCCAATCGCTTCATTGC GAAAAACACAAACGTCCTAAGGGGATGATCAGCTAGTTCAGTTCTGCCGC TAGCCAAAACTGTTTGCCACCCAGTTACAGATAGCGTCTGTGCTGACCA GCTGCCCCGCGGGCTTGGGGTGCAGTGGAGGCCGGCCCGTGGCCAGG CCGGACACGGGCCGTGGAGCCTGCTCCCAGTCCAGAGAGTTCCCCCTCG GAAGCTGCCAGGGCAGCACCAGCCGGGGAGGGCCACGGCCCCGGTCA TGAGTCCCCCTCGGTGCAGAGGCCTGAGGCCGATACCACTGCCCCTGGT GTGAGCGCGCCCACCAGGGAGGGTGAACCACCCTCCACCAGGGTTTTC CTGGTGAGGCTGCCTGATTCAAACCCGCCGTGCCCAATTTGCAGGGACC ATGTGGGTAAACCCTCCGCGCTGGCCCTCCACTGCGTGGAGAGCCATGC GTGGGCGGATGTGCAATACCAGTGCACCCATTGCAAAAAGGTCAGTGC TAACAAACACAGCATCCTCTGCCACATCCCATGTTGCCAGGGGAGGGTG CCCGAGTGGACCGGGAAGGACTGGGCTTGTCCTGAGTGCCCTGCCTCCT TCAATAAGAAAGTTGGCCTATCGCAGCACAAGCGGCACGTGCATCCGGT AACACGTAATGTGGAACGGGTTGCAGGGAGCCTATCGAGGGCTGGTTT AAGGCCCCAGACCAGACGCGGGTGCTGGTCGGTGGAAGAAGAGGAAA CTCTCACCTGCCTAGACGCGATGTTCCGTGGTGCCCGGAACATCAACCA GCTGATCGCTGCTGAAATGGTAACGAAGATGCCTAAGCAGATCAGTGA CAAACGGAGGCAGCTCGGTCTGTGTCCTGAGCAGACCACATTGGGTGG TGATGCTGAATCGACCTCCGTGGTGGAGGAAGAGTCCATGACTCCGGA GATGGAAACCCAAAGCCCAATTAACCCGCCTGGGAAAATCAGGAAGAT CCTGGCCCAGAGGGCACGCCGGTGGCTGAAGAAGGGGCAGGGTCTGT CGGACAAGGTGCGAGAAGTCCTGGGCGCATGGGTGGAGGGTCAACCC AGGATTCATGCCTGGGTCGACTCAGTCTCCCTTGATGTTTTGACTTTGTT CTTGGGGGTGCCCTCAGGACCGCAGAGGGCTCCGAACAAGAAGAGGCC CAAGGAGGGTGGCAAACCAACGTCCTGGATGAACAAATGTGCCGTCAA ATGGGGCACATTCCTTCGGTACCAGCACCTGTTTGGTGCCAACAGGAAG	AGGCATCT CCTTKAAG GGTAATGG TCTGGTTA CATGGTCA TAGCAGGT TTGTGTCA GGTACCTC CCAGTGGT TCCCGCCG GGTGSCA MAGCCCC AGGGCTGT CGGTAGCT CGATCCTG GTACAGTA CGGCCAAG GGAGTTCT TCCTTGCT GTCGGGT GCCTCGCA AGCACKTG GCAGCCCC AATCGCTT CATTGCGA AAAACACA AACGTCCT TGATCAGC TAGTCAGT	TGAGAA GTTGCG AGTTCTT ATGCAA GTTGAA TACCACT CTKGKG ACCCCA AAAAAW WAAACC CCAAAA CAGTTG TGTTTAA GTGTGT TCTTGTT CGTCCCT TTGGCTT CACCTCS AAGTTG CGATCC CCCCATC TCCCCTG CGCTGC CTTTCAG AACGGC CGGTGG CGCGAC CTCGGT	PAAPRAWGAVEAGP WPGRTRAVEPAPSP ESSPSEAARAAPAGE GHGPGHESPSVQRP EADTTAPGVSAPTRE GEPPSTRVFLVRLPDS NPPCPICRDHVGKPS ALALHCVESHAWAD VQYQCTHCKKVSAN KHSILCHIPCCQGRVP EWTGKDWACPECPA SFNKKVGLSQHKRHV HPVTRNVERVAGSLS RAGLRPQTRRGCWS VEEEETLTCLDAMFR GARNINQLIAAEMVT KMPKQISDKRRQLGL CPEQTTLGGDAESTS VVEEESMTPEMETQ SPINPPGKIRKILAQR ARRWLKKGQGLSDK VREVLGAWVEGQPR IHAWVDSVSLDVLTL FLGVPSGPQRAPNKK RPKEGGKPTSWMNK CAVKWGTFLRYQHLF GANRKLLVAIVLDGA DRNQCTLLLEEVFQA YREKWGLEEVLRAYR GKWEVESSFEGLGRF

	CTCCTGGTGGCSATCGTCCTGGACGGCGCTGACCGTAATCAGTGCACCC	TCTGCCGC	CACCTCC	GVRRDADNFAFKALI
	TCCTGCTAGAGGAGGTCTTCCAGGCCTACCGAGAAAAGTGGGGGCTAG	TAGCCAAA	AAGGCC	TPEEVVKHMMAMA
	AGGAAGTCCTTCGGGCCTACCGAGGAAAGTGGGAGGTAGAGTCATCTT	ACTGTTTG	AAGTGC	SKSAPGPDKLTLRDLR
	TTGAGGGTCTCGGACGGTTCGGGGTCCGCCGGGATGCGGATAACTTCG	CCACCCAG	CCTGGC	RADPEGDALAELFSL
	CATTCAAGGCCCTGATCACTCCTGAGGAGGTTGTCAAACACATGATGGC	TTACAGAT	CCCGAG	WLITGTVPDGLKECR
	AATGGCCTCGAAATCGGCTCCAGGTCCGGATAAGCTCACCCTGAGAGAT	AGCGTCTG	TAGGAC	SVLIPKTVDREKLGQL
	CTGCGCCGCGCTGACCCCGAGGGAGATGCTCTTGCCGAACTCTTCAGCC	TGCTGA (SEQ ID NO: 1212)	TGAGTG	GNWRPITIGSIVLRLF
	TGTGGCTGATTACCGGCACGGTCCCGGACGGACTCAAGGAGTGTCGGT		GCCCAG	SRVLTARLAAACPINP
	CTGTGTTGATACCCAAAACGGTGGACCGGGAGAAGTTGGGCCAGCTGG		CTCGCT	RQRGFIAAPGCAENL
	GCAACTGGCGCCCTATCACGATTGGGTCCATCGTTTTACGGCTATTCTCA		GGGCAC	KVLELLLRKRKRDRQP
	CGAGTGCTAACCGCACGGCTCGCCGCAGCATGTCCCATCAACCCCCGTC		CCGTCA	LGVVFVDLARAFDSV
	AGAGGGGTTTCATAGCGGCGCCGGGGTGCGCCGAGAACCTGAAGGTG		CCATCTG	SHDHISWVLKAKGV
	CTTGAGCTTCTCTTGCGGAAGAGGAAGCGAGACAGGCAGCCGTTGGGT		GGGCAA	DEHIVNLIEDSYQKVT
	GTGGTATTTGTGGATCTAGCGAGGGCGTTCGATTCGGTGTCACACGATC		ATGGAA	TRVQVFNGVTPPISIK
	ACATTTCTTGGGTCCTAAAGGCCAAAGGGGTGGACGAACACATCGTGA		GGGATC	TGVKQGDPMSPLLF
	ATCTCATCGAAGATTCTTACCAGAAGGTTACCACGAGAGTACAAGTGTT		TGTCCT	NIAMDPLIAKLETDG
	CAATGGCGTCACCCCTCCTATCAGCATCAAAACCGGGGTTAAGCAAGGC		GACCAC	QGVKVGSASLTTLAF
	GACCCGATGTCCCCCCTCTTGTTCAACATTGCGATGGACCCCCTGATAGC		TACCAG	ADDLVLLSDSWEGM
	GAAGCTCGAGACAGACGGACAGGGAGTAAAAGTCGGGAGTGCCTCCCT		GCTAAG	LKNISILEDFCNLTGLR
	GACCACCCTGGCCTTCGCGGATGATCTCGTCCTGCTTAGCGACTCTTGG		TGTGGT	VQPKKCQGFFLNPTC
	GAGGGCATGCTGAAGAACATCAGCATCCTAGAGGACTTCTGCAACCTCA		GCGGCC	DSFTVNNCEAWKIA
	CGGGCCTACGAGTGCAACCCAAAAAATGTCAGGGGTTCTTCTTGAATCC		TAGCCT	GREITMLGPGESTRY
	GACATGCGACTCCTTTACGGTGAACAACTGCGAGGCCTGGAAGATAGC		GCCGTA	LGLNVGPWVGIDKP
	CGGCCGTGAGATCACGATGCTCGGACCAGGCGAGTCGACACGATATCT		AGGTCA	DLGTQLSSWLERIGT
	GGGCTTGAATGTCGGTCCTTGGGTTGGGATCGACAAACCAGATTTGGGT		AGCGCC	APLKPMQKLSLLVQY
	ACGCAACTAAGCTCCTGGCTCGAGAGGATAGGGACTGCTCCACTCAAAC		CTGCTG	AIPRLNYQADYAGIG
	CGATGCAGAAGCTCTCTTTGCTGGTGCAGTATGCCATACCCAGGCTGAA		CCACTCA	RVALEALDSMNRRK
	CTATCAGGCCGATTACGCGGGCATCGGCAGGGTGGCCTTGGAGGCTCT		GGTATC	VKEWFHLPACTSDGL
	GGATTCTATGAACCGGAGAAAGGTAAAGGAGTGGTTCCATCTTCCCGCC		AGTCCTC	LHSRHRDGGLGLPRL
	TGTACCTCGGACGGTCTCCTCCACTCCCGTCACCGTGACGGGGGTCTTG		GTTCACT	AKAIPEAQVRRLIRVA
	GGTTACCGCGTCTGGCGAAAGCCATTCCGGAAGCGCAAGTGAGGAGGC		TGTCCCT	TSSDEVTRKVSYACGI
	TGATCCGCGTAGCCACTTCATCTGATGAAGTCACGCGGAAAGTATCCTA		CCTAGT	SDEVERLWLARGGD
	CGCGTGTGGGATAAGTGACGAAGTGGAGCGGCTCTGGTTGGCGAGGG		ACCCTCT	MSSVPRFEDPEAPRS
	GTGGGGACATGTCCAGTGTACCGAGGTTCGAGGATCCTGAGGCCCCGA		GCCTCT	PGVQGPCEAAQEIPS
	GGTCTCCGGGGGGTGCAGGGCCCCTGCGAGGCTGCCCAGGAGATTCCGA		GCTCTTT	VVRKLAIPRPSNWRS
	GCGTAGTCCGGAAGCTTGCGATCCCCCGGCCCTCCAACTGGAGATCCAA		TGCTATC	KKHSKWAQLSCQGE
	GAAACACTCCAAATGGGCCCAACTCAGCTGTCAGGGAGAGGGGATGGA		CACTAT	GMELFCNDPVSNG
	GTTATTCTGCAATGATCCAGTCAGCAATGGCTGGAACAACAGTCGGGGA		GGCCAG	WNNSRGQLAEHLQI

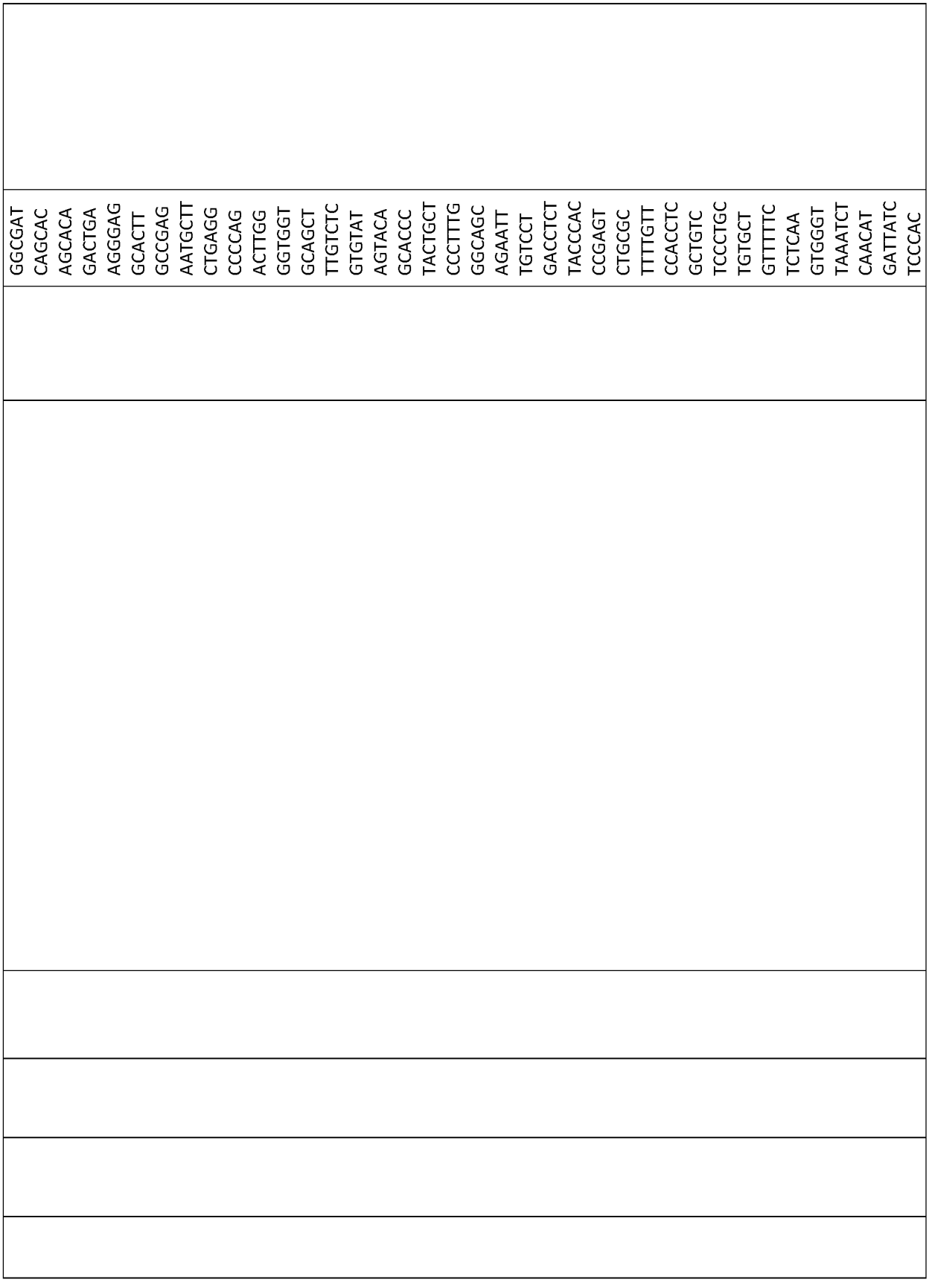
	CAACTGGCGGAACACCTCCAGATCGTGGCCTTAAAACTGCGTTCAAACA		TGATGTT	VALKLRSNIYPTREFL
	TTTATCCCACCAGGGAGTTTCTTGGGAGAAGCCAGGCAAGTACCAATGT		GAGGTT	GRSQASTNVGCRHC
	AGGTTGCCGGCATTGTACACACCCTCATGAAACACTAGGGCATATCTTG		GGTGCA	THPHETLGHILGICPA
	GGCATATGCCCTGCCGTGCAGGAGGCACGGATCATCCGGCATAACAAG		TCCTTG	VQEARIIRHNKLCKIL
	CTGTGCAAGATCCTAGCAGCTGAGGGCAAAAAGTGTGAGTGGACAGTG		GTCACCT	AAEGKKCEWTVYYEL
	TATTATGAACTGCAACTGCTTAACGCTGCAGGGGAACTGTGTAAACCTG		CCAGGG	QLLNAAGELCKPDLIF
	ACCTCATTTTTGTCCGAGACGGTACCGMTCTGGTTGTGAATGTCACTGT		CCAAGC	VRDGTXLVVNVTVG
	GGGGTACGAAGGGGGCCCCGCAWCCCTCCTATCCACCGCTGCAGAAAA		GCCTTG	YEGGPAXLLSTAAEK
	GGCCACAAAATACCTGGATMTGAACGCACAGATCCAGGAGCTCACAGG		GCCACA	ATKYLDXNAQIQELT
	GGCTGAGCAGGTCACCTACTTTGGCTTCCCTATWGGAGCCAGGGGAAA		GGTAGG	GAEQVTYFGFPIGAR
	GTGGCATGCTGACAACTGKCGAGTACTGTCTGAACTGGGATTGTCCAAC		ACCTGG	GKWHADNXRVLSEL
	TCCCGGAAGGAGCGGGTCGCACGGCTCCTGTMGTGGCGAGCACTGCTC		CACCTG	GLSNSRKERVARLLX
	GGGTCAGTGGACATGGTGAACATCTTTGCATCTAAGCACAGGCAGGAA		CCCAGG	WRALLGSVDMVNIF
	AACCTATCGGATGAMGCACTGAGCCCCAGCTGAGAAGTTGCGAGTTCT		GGGCCA	ASKHRQENLSDXALS
	TATGCAAGTTGAATACCACTCTKGKGACCCCAAAAAAWWAAACCCCAA		GACACT	PS (SEQ ID NO:
	AACAGTTGTGTTTAAGTGTGTTCTTGTTCGTCCCTTTGGCTTCACCTCSAA		GCCTGT	1457)
	GTTGCGATCCCCCCCATCTCCCCTGCGCTGCCTTTCAGAACGGCCGGTGGT		GGCAAG	
	GTCGAGGCTGGCGCGACCTCGGTCACCTCCAAGGCCAAGTGCCCTGGC		GGAAAG	
	CCCGAGTAGGACTGAGTGGCCCAGCTCGCTGGGCACCCGTCACCATCTG		GGAGCC	
	GGGCAAATGGAAGGGATCTGTCCTGACCACTACCAGGCTAAGTGTGGT		GTCCCT	
	GCGGCCTAGCCTGCCGTAAGGTCAAGCGCCCTGCTGCCACTCAGGTATC		GACCGT	
	AGTCCTCGTTCACTTGTCCCTCCTAGTACCCTCTGCCTCTGCTCTTTTGCT		TACCAG	
	ATCCACTATGGCCAGTGATGTTGAGGTTGGTGCATCCTTGGTCACCTCCA		GCTTGA	
	GGGCCAAGCGCCTTGGCCACAGGTAGGACCTGGCACCTGCCCAGGGGG		GATGGT	
	CCAGACACTGCCTGTGGCAAGGGAAAGGGAGCCGTCCCTGACCGTTAC		GCTGGC	
	CAGGCTTGAGATGGTGCTGGCTAGCCCACCATATGTCAAGCACTCCACA		TAGCCC	
	GCTGCTTGAGTTTGTTGGCTTCACCTTCATCCCACCTAGTGTCTTCTGCCT		ACCATAT	
	CTGCACTATTTTCATCCCAACTCGTACCTCCCCATCTCTGCGCTCCTGCTA		GTCAAG	
	TCCCTGCAAGGACCAAGTAGGCAGGGGGGTTCATCCCCCTACCTGCAG		CACTCCA	
	GAGACTCAGCATATCCATGACTTCTTGCCTCCACCGTCTTGTGGCGCTAG		CAGCTG	
	AGGGGTACCTCAGAGACCGGCACAACATGACCTTGACGGTTAGACAGT		CTTGAG	
	AGGGTCAAACAACCCTGCTGCAGGCCCAAAGGGCCAACAGCTGTGCCA		TTTGTTG	
	CGAGAGGGGAACCTTGAAGACTGGGGCAGTCTGACCATGCTGGTTAGT		GCTTCAC	
	CAGTTGGGTGAAATAATCCCAGCTGCAGGCCCAAAAGGGCTGACAGTC		CTTCATC	
	AGGTGAGGGGGTATCTCCATCTGCTCCCCACTGCCAACTACGGAGGCAT		CCACCTA	
	GAAGTCCGTAGTGACTTCTGACCCCCACGTCTTGTGCCATGAGAAGGGA		GTGTCTT	
	ACCTTGAAGATTGGGACAAACCGCACTTGAAAGTTACTCAGCCGGGTGA		CTGCCTC	
	AAATAAGTCCCAGTTGCGGGCCCCTCGGGGCTGACAGTCAGGTGAGGA		TGCACT	

			GGGCTGCAAAGCCCATCTCCTGACTCCAGAGGCCTGGCGTCCTAACCGA		ATTTTCA
			CTTCTTGCCACCAATGTCTTGCGCCAGGAGAGGGCAACCTTGAAGATCG		TCCCAAC
			GGGCAAGCCGCACTTGATAGTTAGCCAGTCGAGTGAAACAATCTCAGCT		TCGTACC
			GCGGGTCCGAAAGGACTGACTTCCAGGCGAGGGGGGGGCCTGCGGAA		TCCCCAT
			AACCCCCTCCATGGTACGGAGGTCTGGCATCCTAACCGACACCTTGCCA		CTCTGC
			CCAATGTCTTGTGCCAGGAGAGGGGAACCTTGAAGACTGGGGCAAGCC		GCTCCT
			GCAGTTGATGGTTAGTCAGTCGGGTGAAATAATCCCGGCTGCACCCTGC		GCTATCC
			TGTGACTGCTAAGCCCGGTCCCCAAGGGGCATGAGGCATGTGCGCTGA		CTGCAA
			GACGGGAGGGGTGACATCTGGCGATCAGCACAGCACAGACTGAAGGG		GGACCA
			AGGCACTTGCCGAGAAATGCTTCTGAGGCCCCAGACTTGGGGTGGTGCA		AGTAGG
			GCTTTGTCTCGTGTATAGTACAGCACCCTACTGCTCCCTTTGGGCAGCAG		CAGGGG
			AATTTGTCCTGACCTCTTACCCACCCGAGTCTGCGCTTTTGTTCCACCTCG		GGTTCA
			CTGTCTCCCTGCTGTGCTGTTTTTCTCTCAAGTGGGTTAAATCTCAACATG		TCCCCCT
			ATTATCTCCCACGTTTCCGCTCAAGGGCAATGCCCAACATGACGGAGAT		ACCTGC
			CGTTGGTGCATGGTAGTCACGAGACCATCCGGACCCTCCAGTGGTCGCT		AGGAGA
			ATAGTCATTTTGTGTTGCATGGGGCATGCTGAGTCACTTAACCGAAAGA		CTCAGC
			CTGTAAATAACTCAAAAGAGGGTACCCTCCGGGGTTCGGTAAA (SEQ ID NO: 1089)		ATATCCA
					TGACTTC
					TTGCCTC
					CACCGT
					CTTGTG
					GCGCTA
					GAGGGG
					TACCTCA
					GAGACC
					GGCACA
					ACATGA
					CCTTGAC
					GGTTAG
					ACAGTA
					GGGTCA
					AACAAC
					CCTGCT
					GCAGGC
					CCAAAG
					GGCCAA
					CAGCTG
					TGCCAC


GAGAGG
GGAACC
TTGAAG
ACTGGG
GCAGTC
TGACCA
TGCTGG
TTAGTCA
GTTGGG
TGAAAT
AATCCC
AGCTGC
AGGCCC
AAAAGG
GCTGAC
AGTCAG
GTGAGG
GGGTAT
CTCCATC
TGCTCCC
CACTGC
CAACTA
CGGAGG
CATGAA
GTCCGT
AGTGAC
TTCTGAC
CCCCAC
GTCTTGT
GCCATG
AGAAGG
GAACCT
TGAAGA
TTGGGA
CAAACC
GCACTT
GAAAGT
TACTCA


GCCGGG
TGAAAA
TAAGTC
CCAGTT
GCGGGC
CCCTCG
GGGCTG
ACAGTC
AGGTGA
GGAGGG
CTGCAA
AGCCCA
TCTCCTG
ACTCCA
GAGGCC
TGGCGT
CCTAACC
GACTTCT
TGCCAC
CAATGT
CTTGCG
CCAGGA
GAGGGC
AACCTT
GAAGAT
CGGGGC
AAGCCG
CACTTG
ATAGTT
AGCCAG
TCGAGT
GAAACA
ATCTCA
GCTGCG
GGTCCG
AAAGGA
CTGACTT
CCAGGC


GAGGGG
GGGGCC
TGCGGA
AAACCC
CCTCCAT
GGTACG
GAGGTC
TGGCAT
CCTAACC
GACACC
TTGCCAC
CAATGT
CTTGTG
CCAGGA
GAGGGG
AACCTT
GAAGAC
TGGGGC
AAGCCG
CAGTTG
ATGGTT
AGTCAG
TCGGGT
GAAATA
ATCCCG
GCTGCA
CCCTGCT
GTGACT
GCTAAG
CCCGGT
CCCCAA
GGGGCA
TGAGGC
ATGTGC
GCTGAG
ACGGGA
GGGGTG
ACATCT


GGCGAT
CAGCAC
AGCACA
GACTGA
AGGGAG
GCACTT
GCCGAG
AATGCTT
CTGAGG
CCCCAG
ACTTGG
GGTGGT
GCAGCT
TTGTCTC
GTGTAT
AGTACA
GCACCC
TACTGCT
CCCTTTG
GGCAGC
AGAATT
TGTCCT
GACCTCT
TACCCAC
CCGAGT
CTGCGC
TTTTGTT
CCACCTC
GCTGTC
TCCCTGC
TGTGCT
GTTTTTC
TCTCAA
GTGGGT
TAAATCT
CAACAT
GATTATC
TCCCAC

							GTTTCC GCTCAA GGGCAA TGCCCA ACATGA CGGAGA TCGTTG GTGCAT GGTAGT CACGAG ACCATCC GGACCC TCCAGT GGTCGC TATAGTC ATTTTGT GTTGCA TGGGGC ATGCTG AGTCAC TTAACC GAAAGA CTGTAA ATAACTC AAAAGA GGTACC CTCCGG GGTTCG GTAAA (SEQ ID NO: 1335)		MQCTSRLADAPRFA RVGVEGEGVGASGN GTDAQLWYGCTGCD EAFSSLRGLRIHAAQK KHGNQDGLLRLPAG RPRKRRVGKSTTAGA
								GTTCCAAA GGAAGGC ACTCCTTT GGTTCGTG ATGAGATG TTCATGGT	TAGTGT GACGGA GTCCTCA AGCCCC CACAAG TGCCTG	
R2	R2-1_IS	.	Ixodes scapularis	GTTCCAAAGGAAGGCACTCCTTTGGTTCGTGATGAGATGTTCATGGTGC TTGCCTAGCTGGAGAAATCCGACTCACACCTGCACGTGGTCCCTGCCGC CTGCCAGTATGCCGAGGAAACGGGTGCAACTTAATCCGTGGATACTGGT AGCAACGTGAGCAACGGTACGGTCCTTCGCGGACCACCCTGGGCGTTC GGGTTGCCAGCCCGTTCGCCCGAAATATCTTGGCCCTGAAACTAAAAGA AAAATGCAATGCACCAGCCGACTGGCTGATGCACCAAGATTTGCCCGAG				

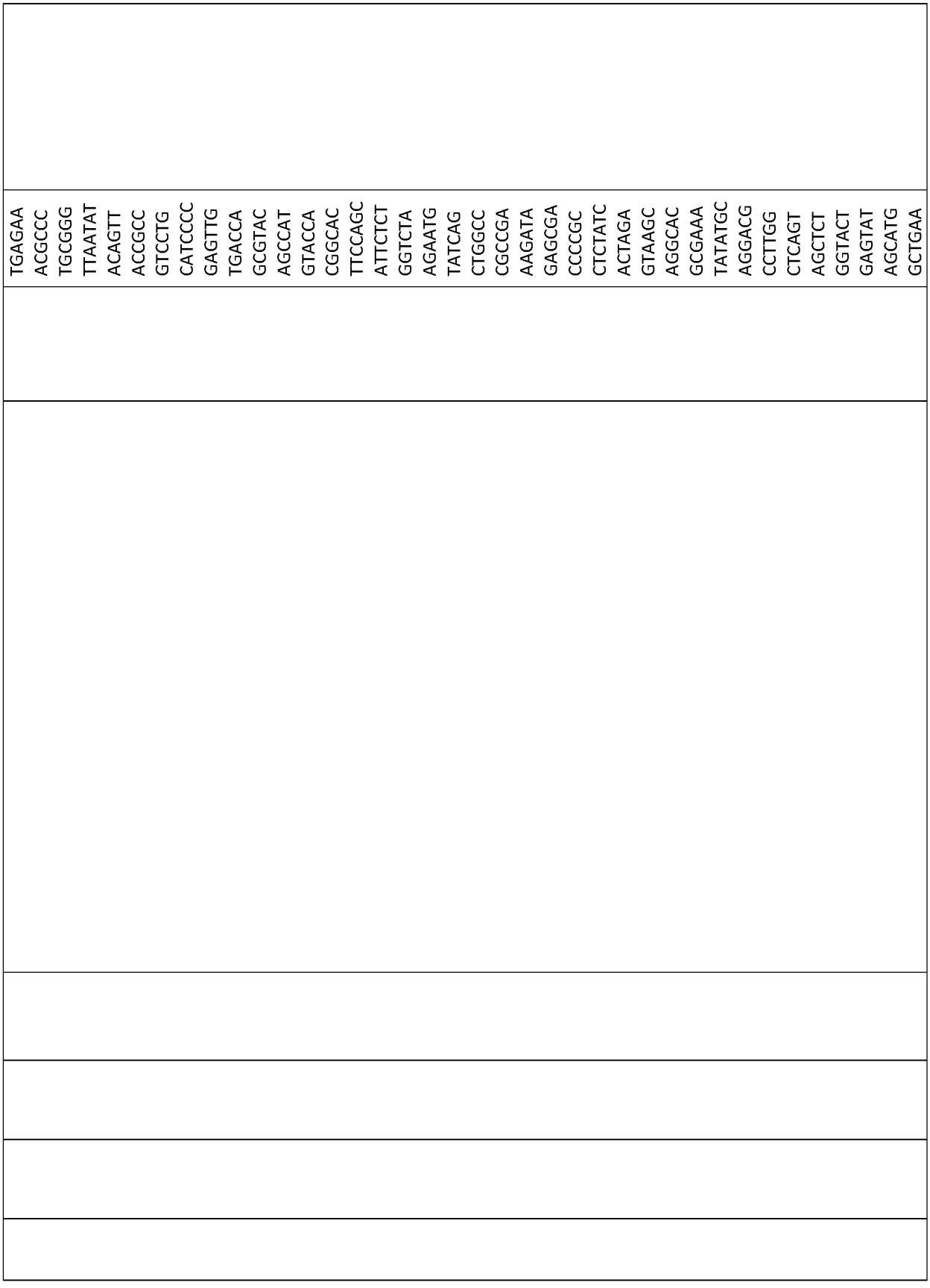
	TGGGCGTCGAGGGTGAAGGTGTCGGTGCGTCTGGTAACGGCACTGATG	GCTTGCCT	CCAGGT	SDRVTTDPVPAPVPE
	CGCAGTTATGGTATGGCTGCACGGGCTGTGACGAAGCCTTCTCGTCCCT	AGCTGGA	GGCAGG	SPGLLPGLPGPSLPGC
	GCGAGGACTGAGAATTCATGCGGCCCAAAAGAAACATGGAAACCAAGA	GAAATCCG	AAAGGG	SDLPPGVLPGGWSA
	TGGCCTTCTCCGCCTGCCGGCGGGACGGCCCCGAAAACGACGAGTGGG	ACTCACAC	CAACTA	SPGPLSWPPSLDAGP
	GAAGAGCACCACAGCGGGTGCTTCGGACCGGGTGACCACGGATCCAGT	CTGCACGT	CTGGTG	LPGPSRVSPGPSRPSP
	GCCTGCTCCAGTTCCTGAATCTCCTGGGCTGCTGCCTGGGCTACCTGGAC	GGTCCCTG	AGCGAC	GKPTGPPSLDAGPLP
	CATCGCTGCCTGGGTGCTCGGACCTGCCGCCTGGGGTGCTGCCTGGAG	CCGCCTGC	CCAAGC	GPSRVSPGPSRPSPG
	GGTGGTCTGCATCACCTGGGCCTCTCTCCTGGCCTCCTTCCCTGGATGCC	CAGTATGC	AAGGCG	KPPGTPEPLPGSPGG
	GGGCCTCTGCCTGGACCTTCAAGAGTATCACCTGGACCTTCAAGACCTTC	CGAGGAA	GAGCCA	RRGVSPGQPGSRTD
	GCCCGGGAAGCCGACCGGGCCTCCTTCCCTGGATGCCGGGCCTCTGCCT	ACGGGTGC	AGACCA	PSSSAGAGHFVCPQC
	GGACCTTCAAGAGTATCACCTGGACCTTCAAGACCTTCGCCCGGGAAGC	AACTTAAT	AGCTGG	SRAFSSKIGMSQHQK
	CGCCCGGGACGCCTGAGCCGCTGCCTGGATCTCCTGGCGGTCGGCGCG	CCGTGGAT	AGCCAA	HAHLEEYNAGINITRT
	GGGTGTCCCCGGGACAGCCCGGGTCACGGACCGACCCCTCAAGCTCTG	ACTGGTAG	GAGCAA	KARWDPEETYLLARL
	CTGGCGCCGGACACTTCGTATGCCCGCAGTGCAGCAGAGCCTTCTCAAG	CAACGTGA	CTCCAG	EATLNPDHKNINQTL
	CAAGATTGGCATGTCTCAACACCAAAAACATGCCCACCTCGAAGAATAC	GCAACGGT	GAGGCA	HAALPRGSCRTLESIK
	AACGCGGGCATTAACATCACCGTACCAAAGCCCCGGTGGGACCCCGAG	ACGGTCCT	GGGGTG	AHRKQAAYRDLVTSL
	GAGACCTATCTTCTGGCCCGCCTGGAGGCCACCCTCAACCCAGACCACA	TCGCGGAC	GATATC	RSARESSEAQHVPDR
	AGAACATCAACCAGACGCTGCACGCCGCGCTGCCCCGCGGTTCCTGTCG	CACCCTGG	AAGAGC	PLETPEPQTPANPQR
	AACCCTGGAGAGCATCAAGGCCCACCGAAAGCAGGCGGCTTACAGGGA	GCGTTCGG	AACCCC	DSKQAVIEALQSLIGR
	CCTGGTGACGAGCCTGCGGTCAGCCAGGGAGAGCAGCGAGGCGCAGC	GTTGCCAG	AAGGGA	APPGSFQGARLWDI
	ACGTTCCGGACCGGCCCCTGGAGACCCCGGAGCCCCAGACACCAGCGA	CCCGTTCG	CACAGA	ARQATRGTNILPLLN
	ACCCTCAAAGAGACTCGAAGCAGGCAGTCATCGAAGCGCTGCAATCCCT	CCCGAAAT	CCACGG	SYLRDVFTLPTKPTRK
	CATCGGCCGAGCACCACCACAGGCTCCTTCCAGGGAGCGCGTCTCTGGGAC	ATCTTGGC	GCAACT	KPAVRPARSRRKQKK
	ATCGCGAGGCAAGCCACAAGGGGGACGAACATCCTCCCACTCCTGAAC	CCTGAAAC	ACTGGT	QEYARTQDLFRKKQS
	AGCTACCTGAGGGATGTTTTCACCCTCCCCACAAAGCCAACAAGAAAGA	TAAAAGAA	GAGCGC	DCARAVLDGPTSSSV
	AACCTGCAGTGCGGCCCGCCCGGAGCCGCAGAAAACAAAAGAAACAAG	AA (SEQ ID	CCAAGA	PGTGAFLQTWREIM
	AGTATGCCAGAACACAAGATCTATTCAGGAAAAAGCAGTCCGACTGTGC	NO: 1213)	CAGGGG	TGPSPALEAPPLPTR
	CAGGGCGGTCCTGGACGGCCCCACGTCGTCATCGGTCCCTGGAACGGG		TGGATA	GEVDLFFPATAQEIQ
	CGCCTTCCTGCAAACCTGGCGAGAGATCATGACGGGGCCCAGCCCTGCA		TTAAGA	SAEIAVNSAAGPPDGF
	CTCGAGGCACCGCCTCTACCTACCCGGGGGGAAGTCGACCTGTTCTTCC		ACAGCC	SARLLKSVPALLLRVM
	CGGCGACGGCGCAAGAGATCCAGAGCGCTGAGATAGCCGTCAACTCGG		CCACAA	VNLLLLVRRVPAALR
	CTGCTGGACCCGACGGGTTCTCAGCCCGTCTCCTCAAGTCCGTCCCGGC		AGTGTT	DARTTFIPKVPDAVD
	CCTCCTCCTAAGGGTCATGGTTAACCTTCTGCTCCTCGTCCCGACGTGTCC		ACCTATA	PSQFRPITVASVLQRL
	CGGCGGCCCTCCGGGACGCGAGAACGACCTTCATCCCGAAGGTCCCCG		TTAACA	LHRILAKRALEAIPLNF
	ATGCAGTGGACCCCTCCCAATTTCGCCCAATAACGGTGGCCTCCGTTCTC		ATAAAG	RQRAFQPVDGCAEN
	CAGCGCCTACTACATCGCATCCTGGCCAAGAGGGCGCTGGAGGCCATTC		TTGAAG	IWLLSTALNEARTRR
	CCCTCAACTTTCGACAAAGAGCCTTTCAGCCGGTGGATGGCTGTGCCGA		CCTCAAC	RPLHMASVDLTKAFD
	GAATATATGGCTGCTGTCCACCGCGCTCAACGAGGCAAGAACCAGACG		CACGCA	RVTTDAILRGARRAG

				TTGCGG	LSGEFIGYLKELYTTSR
		GCGCCCGCTACACATGGCGAGCGTCGACCTAACCAAGGCATTCGACCG		GTTAGA	TLLQFQGESLLVEPTT
		GGTCACCACGGATGCCATCCTGAGGGGCGCAAGGCGCCGCCGGGCTGTC		TGGCGT	GVRQGDPLSPILFNL
		CGGGGAGTTCATCGGATACCTGAAGGAGCTCTACACAACATCCAGGAC		GGCTTG	VLDEYLSSLDPDISFVS
		CCTCCTGCAGTTCCAGGGAGAGAGCCTGCTTGTCGAACCCACGACCGGC		GCCCGC	GDLRLDAMAFADDLI
		GTGCGACAGGGCGACCCACTGTCGCCCATCCTCTTCAACCTGGTCCTGG		CGCCAT	VFASTPAGLQDRLDA
		ACGAGTACCTCTCCTCCCTGGACCCGGACATCTCCTTCGTCTCGGGCGAC		GATGAG	LVEFFDPRGLRVNVK
		TTGCGCCTCGATGCGATGGCATTCGCTGACGACTTGATCGTCTTTGCCTC		CTGGAA	KSFTLSLQPGRDKKV
		AACCCCAGCCGGCCTGCAGGATCGGCTCGATGCCCTCGTCGAGTTCTTC		CCCTCCA	KVVCDQIFTIGGTPLP
		GACCCAAGGGGGCTCAGGGTGAACGTGAAGAAGAGCTTCACGCTATCG		CCTGGT	ASKVATPWRYLGMT
		CTGCAGCCGGGACGAGACAAGAAGGTCAAGGTGGTGTGTGACCAGATC		GGGCCG	FTPQGSINKGTSEQL
		TTCACCATCGGAGGAACCCCACTCCCAGCCTCCAAGGTCGCAACCCCTT		CACGAG	DLLLTRTSKAPLKPQQ
		GGCGCTACCTGGGGATGACCTTCACCCCCCAGGGCTCAATCAACAAGG		ACCACC	RLVVLRNYLLPRLYHR
		GCACCAGCGAGCAGTTGGACCTACTGCTCACGAGAACCAGTAAGGCCC		GGCTCT	LVLGPWSAALLLKM
		CCCTCAAGCCACAACAGAGGCTGGTGGTCTTAAGAAACTACCTGCTCCC		TTCTACT	DTTIRGAIRRWMDLP
		GAGGCTATACCATCGCCTCGTGCTTGGACCTTGGTCGGCCGCCCTCCTAC		AAGGCC	HDTPLGFFHAPVTEG
		TGAAGATGGACACCACCATTCGAGGAGCCATTAGACGCTGGATGGATC		GGTCTC	GLGINSLRASIPAMVL
		TCCCGCACGACACACCGCTGGGTTTCTTCCACGCCCCAGTAACGGAGGG		CGTGAC	QRLDGLHFSTHPGAE
		AGGCCTAGGAATCAACTCCCTGCGAGCATCAATTCCAGCCATGGTGCTC		TGCGGT	VAIQLPFLTGLHRRAE
		CAACGGCTGGATGGACTTCACTTCAGCACGCATCCCGGAGCTGAGGTCG		TGGGAT	AAAQYQGQRLLSKA
		CCATCCAGCTGCCGTTCCTGACAGGACTCCATCGAAGAGCGGAAGCGG		AAACTC	DVHRMWSARLHGS
		CGGCCCAATACCAGGGACAGAGACTACTGTCCAAAGCGGACGTCCACC		CAAGCA	CDGRPLRESKRVPAA
		GGATGTGGAGCGCAAGACTCCACGGGAGCTGCGACGGAAGACCCCTTC		CTGAGC	HRWAAEGTRLLSGR
		GGGAGTCCAAGAGAGTGCCGGCTGCCCATCGTTGGGCCGCGGAAGGC		GGTAAA	DFISITKLKINALPTLE
		ACCAGACTACTCTCGGGAAGGGACTTCATCTCGATCACGAAACTCAAGA		AAAAAA	RTSRGQHKDIQCRA
		TAAACGCGCTACCTACACTCGAGCGCACCAGCCGGGGCCAGCACAAGG		AAAAAA	GCQAVESLGHVLQA
		ACATCCAGTGCAGAGCTGGCTGCCAGGCTGTTGAATCCCTGGGCCACGT		AAAAAA	CHRGHRGRIRRHDNI
		CCTACAAGCTTGCCATCGAGGACACCGTGGCCGAATCCGGCGGCATGAT		AAAAAA	ARYVCGRLTQIGWA
		AACATTGCCCGCTACGTCTGCGGCCGACTGACCCAGATTGGCTGGGCGG		(SEQ ID	VKWEPHYSVAGRTL
		TGAAGTGGGAGCCCCACTACTCTGTCGCTGGAAGGACCCTCAAACCTGA		NO:	KPDIVAHRGAETVVL
		CATCGTTGCCCATCGTGGAGCCGAGACTGTCGTGCTCGACGCCCAGGTC		1336)	DAQVVGTSMRLGFH
		GTCGGCACCAGCATGCGACTGGGCTTCCACCACGCTCAAAAGAAAGAA			HAQKKEKYSLPDLLH
		AAGTACTCTCTCCCAGACCTCCTCCACCAAGTCTGCGAGGGACGGAGAG			QVCEGRRDAARVSTI
		ACGCAGCCCCGGGTGTCAACAATCACCCTCAATTTTCGAGGTGTTTGGGC			TLNFRGVWAPESAQ
		ACCTGAGAGCGCCCAGGACCTGAAGTCCCTGGGCCTGACGGACAACGA			DLKSLGLTDNDLKLLT
		CCTAAAGCTTCTCACCGTCCGCTGCCTCCAGGGCGGCGCGCAGTGTTTC			VRCLQGGAQCFRLH
		CGGCTGCACCGCCGAATGACCACCGTGGTGAAGGCCACGGGCGATGAA			RRMTTVVKATGDEA
		GCCAACGCCCTCCCCGCCCATTCGGGCTTGCCGCCAACACAGCTTGGTG			NALPAHSGLPPTQLG
		GCCGAACCCTGGGTCCCTCTGCCCACAATCAGAGTGCAAGAACTACTTA			

			GTGTGACGGAGTCCTCAAGCCCCCACAAGTGCCTGCCAGGTGGCAGGA AAGGGCAACTACTGGTGAGCGACCCAAGCAAGGCGGAGCCAAGACCA AGCTGGAGCCAAGAGCAACTCCAGGAGGCAGGGGTGGATATCAAGAG CAACCCCAAGGGACACAGACCACGGGCAACTACTGGTGAGCGCCCAAG ACAGGGGTGGATATTAAGAACAGCCCCCACAAAGTGTTACCTATATTAAC AATAAAGTTGAAGCCTCAACCACGCATTGCGGGTTAGATGGCGTGGCTT GGCCCGCCGCCATGATGAGCTGGAACCCTCCACCTGGTGGGCCGCACG AGACCACCGGCTCTTTCTACTAAGGCCGGTCTCCGTGACTGCGGTTGGG ATAAACTCCAAGCACTGAGCGGTAAAAAAAAAAAAAAAAAAAAAAAAA AA (SEQ ID NO: 1090)			GRTLGPSAHNQSART T (SEQ ID NO: 1458)
R2	R2-1_MLe.	Mnemiopsis leidyi	TGGGGGCCCCTTGGACTTGCTCCCTGGGGCAGGACACCAGTGAAAGGA GATCCTCAAGACAGGACAGAGACAGGCACACCAACCCTTCGAACCT GAGGAGACACCCAGATCTGGTTACCCCATTCCTGTAACCATGGTTGCTCT CCGTGCGCTGTTAAGGAAACCCAGCCTAGTACCCTCGGGGAAAGGTTG CGTAGTACTTAGCAGTGTGCGGATCAAACCTCTACCGGTCTCTCTAGCG ATGAAAGTTTCTCCGACTGGAACTTGAGGGACTGGCTAGCCCAGCTAGC CTGTAGAGCAATGCGTATACGATGCCTTGGCGACAATGGCGACCGCTGC TTAGCAGACGGAGGTTAGTGAAAGGGCGACTTGCTGTTCATAGTCACGT GAGTCGTCTAGAAACTGCACCGATGCTGCCCTCTGTTCCAGGAGAAGGA CAGTGGAGGACTAAATCGTAGCGCGAGCGGTGTTATGTCGAACACTAG CCACTCGAAACTTAACCTCAAAATGGATAATAAACTCAAGACTAGTCTCG AGACTCCCTCCGGAGTCCGAGCAGATTCGATAATAACGCGGGTGAGGA CCAGCAGTAACCGTGGTGAACACAGCAATGGTGTCACGTATCCCCGGTG CGAACAGGGGGTTGCGCCTCTGGATACGCATGGGGGGATATGTGACGC TCCTCCACAGGTGACGGTTCCCGCCACTGAGACGGATAAACAGAAAAAA TGTGAATACTGCGAGTTCACATACCTGAAACCGAGACAGATAGGAACCC ACATGAGAAAACGGCATCCCCAGGAATGGAACGACATCAAACGGACCA AATTCCTGAGCGAAAAGAGACAAAAACGTTGGCTTGACGAAGACTTCG AGCTTCTCTGCATAGGCCAAGAGGAATACCTTGTGCTATCCTCGATTGG CAAACAGGGCAAAGGGATCAACCAGTATATTCAAACCAAGTACTTCCCA ACCCTTAGTACAGATGCTATTAAATCTCAAAGGAAATCCAGACGGTTTTC AGAGTACAGCGAAAAGAGATCACGTGAACTCCAACCTTGCAATACCAGC AGCGACCCCGAAGAACTTCCGAATGAAGCGGTCACCGAAAACAGTCCG CTCTCTTTTGATCCACTTGACCGAGATGTTGTAAAGAAAATAAGCTCCAA AGATCACGGCGATCAGATCCTCTTAGTACAGGAGCACCTCATCAACGGA CGGTATCAAGAAGCGAACACGTTGGCCAAAGCTATATTCGAGAAGCTCT CCGGTAAATTCCCGAACCTGAAGACCGGCGACCATCGGCCTGGTAAACA ACAAACAGCAAGGAAAGTCGGTAAGAAAAGAGTTAGAGGAAGCGGTA	TGGGGGC CCCTTGGA CTTGCTCC CTGGGGCA GGACACCA GTGAAAG GAGATCCT CAAGACAG GACAGAG AGACAGG CACACCAA CCCTTCGA ACCTGAGG AGACACCC AGATCTGG TTACCCCA TTCCTGTA ACCATGGT TGCTCTCC GTGCGCTG TTAAGGAA ACCCAGCC TAGTACCC TCGGGGA AAGGTTGC GTAGTACT TAGCAGTG TGCGGATC	TAGTGA CGAGAA AAGTCG CTTTATC CTTACAT AACAGT GTGATA GTCATCC TTATCAC AACTGT CCTGGC GAACAA AATGGC AGAGGG ATAGTA CTCGGT CCAACC AGAAGG AAGCCC ACACGA TGCCAA ACTTGCT GTAACC CACGTG AGCGGA AAACAT CCTCCG AGTATA	MSNTSHSKLNLKMD NKLKTSLETPSGVRA DSIITRVRTSSNRGEH SNGVTYPRCEQGVA PLDTHGGICDAPPQV TVPATETDKQKKCEY CEFTYLKPRQIGTHM RKRHPQEWNDIKRT KFLSEKRQKRWLDED FELLCIGQEEYLVLSSI GKQGKGINQYIQTKY FPTLSTDAIKSQRKSR RFSEYSEKRSRELQPC NTSSDPEELPNEAVT ENSPLSFDPLDRDVV KKISSKDHGDQILLVQ EHLINGRYQEANTLA KAIFEKLSGKFPNLKT GDHRPGKQQTARKV GKKRVRGSGKKLSPS KQNRRELYAIVQKQ WRTKKRSKVINQILT GNLNKEQSYTHTPD QLAQFWSTLFGRVSP RDDRPINHRRSVIPEL DKPLSVEEVEAALKG AKDAATGIDGVPISH LKHLGSAALTILYNGL

	AGAAGCTATCCCCCTCTAAACAGAATAGACGCGAGCTATACGCGATAGT	AAACCTCT	GTATGA		YVTGSIPDPWKRART
	ACAAAAGCAGTGGCGCACTAAGAAAAGATCTAAGGTGATCAACCAGAT	ACCGGTCT	TGGAAG		ILIPKSNPPASPGDYR
	ACTGACAGGTAATCTAAACAAAGAACAATCTTATACACATACACCCGAT	CTCTAGCG	GATACA		PISISSYFYRIYTSAISK
	CAGCTAGCGCAATTCTGGAGTACGCTGTTTGGCAGAGTCAGTCCGAGA	ATGAAAGT	GGATGT		RLASAVSLDDRQKGF
	GACGATCGACCGATTAACCACCGGCGGTCAGTTATCCCGGAGCTAGACA	TTCTCCGA	GGACCG		IKEDGIRDNLSLIDTLI
	AACCCCTCTCTGTTGAGGAGGTCGAAGCTGCGCTCAAAGGGGCCAAAG	CTGGAACT	CTCTAG		NETKAGSKSLFMTF
	ATGCTGCAACAGGCATTGACGGAGTCCCAATCTCCCACCTGAAGCACCT	TGAGGGA	GCGGCG		MDVKKAFDSVSHYAI
	AGGCAGTGCAGCTCTAACAATCTTGTACAACGGATTGTATGTGACCGGG	CTGGCTAG	GACGGT		ARSLEWAGVPDGM
	TCGATTCCTGACCCCTGGAAAAGAGCAAGAACCATACTGATCCCAAAGT	CCCAGCTA	GTAGAC		RSVIADLYQDCTTDIC
	CAAATCCACCAGCTTCACCTGGGGATTATAGACCTATCTCGATCAGCTCG	GCCTGTAG	GGCGAC		GRSVKVTRGVKQGD
	TATTTCTATAGAATCTACACGAGCGCGATCAGCAAAAGACTTGCTTCTGC	AGCAATGC	CTTACCA		PLSSTLFNLVIEMVM
	CGTCAGCCTTGACGACCGTCAGAAAGGATTCATCAAAGAAGATGGAAT	GTATACGA	TAGCTG		SNVPERLGIQFQGHR
	ACGAGACAACTTATCATTGATTGATACTCTCATCAATGAGACGAAAGCA	TGCCTTGG	CGACTG		LFYLAFADDLVLLTRG
	GGCTCGAAATCCCTATTCATGACCTTCATGGACGTCAAGAAAGCTTTCG	CGACAATG	CTGTGC		PTANQKLVSLVHEQL
	ACTCCGTGTCGCATTACGCAATTGCCAGATCTCTAGAATGGGCTGGCGT	GCGACCGC	GATCCG		ARVGLELHPGKCKSIA
	CCCCGATGGAATGAGAAGTGTCATAGCCGATCTATACCAGGACTGCACC	TGCTTAGC	GAACGG		IMADPKRKTTFVDQ
	ACTGATATCTGCGGCAGATCCGTCAAAGTGACCAGAGGCGTGAAGCAG	AGACGGA	GCCTTTC		GSSVLIGGEPVSSLGP
	GGAGACCCCCTAAGCTCCACTCTGTTCAACCTGGTAATAGAAATGGTGA	GGTTAGTG	TCACTTA		QEWYKYLGIKLGSGG
	TGTCAAACGTCCCCGAACGTTTGGGAATTCAATTCCAGGGACACAGACT	AAAGGGC	CCTCAA		MPQGIYRDQLADLLA
	GTTCTATCTGGCTTTTGCAGACGACCTAGTCCTCCTGACGAGAGGACCC	GACTTGCT	GCACCG		KTDSAPLKPQQRLYIL
	ACAGCTAACCAGAAGCTAGTATCCTTAGTCCATGAACAACTAGCAAGAG	GTTCATAG	TGTCGG		RSHILPKFNHRLMFE
	TTGGACTTGAGCTACATCCGGGAAAGTGCAAGTCAATCGCTATCATGGC	TCACGTGA	ATCGCG		RVTCQTLEGLDKLIRT
	TGACCCCAAGAGGAAAACCACTTTCGTAGATCAGGGATCCAGTGTTCTC	GTCGTCTA	GGGTGG		HVRKWLKLPKDTPG
	ATTGGAGGCGAACCCGTCTCTTCGCTCGGTCCGCAAGAATGGTACAAAT	GAAACTGC	GCGCGG		PAFYADKGSGGLGLI
	ACCTTGGGATTAAACTGGGATCAGGAGGCATGCCCCAGGGTATTTACCG	ACCGATGC	GATTCCT		TLRYRVPLLKLRRHKK
	AGACCAACTAGCAGACCTTCTAGCGAAGACAGACAGTGCCCCCCTGAAA	TGCCCTCT	GCTGGG		MADSPDPVIRLIPNA
	CCCCAACAAAGATTGTATATACTCAGATCGCATATACTACCGAAGTTCAA	GTTCCAGG	AAGTTG		EPTISLLARWTKMCS
	CCATAGACTAATGTTCGAGAGAGTAACGTGCCAAACACTAGAGGGTCTT	AGAAGGA	TCTGCAC		LYGKQYQHKSELSKII
	GACAAACTGATCAGGACCCACGTCAGAAAGTGGTTGAAATTACCCAAA	CAGTGGA	CCAAGT		RDKYWTMCDGKGLR
	GACACGCCTGGTCCAGCCTTCTACGCTGATAAAGGTTCAGGCGGACTCG	GGACTAAA	CGTCAG		TEVPPDTAKKTLSLLF
	GACTGATAACTCTCAGATACCGAGTTCCCCTCCTAAAACTTAGGAGACAC	TCGTAGCG	CTAACT		EDRTPLKPGQLIGAIG
	AAAAAGATGGCGGACTCACCTGATCCCGTTATCAGGTTGATACCAAACG	CGAGCGGT	GTCGGT		VRLNTLGTPARNNRA
	CGGAACCGACCATCTCCTTGTTAGCTAGATGGACCAAGATGTGCTCCCT	GTT (SEQ	ACTTATA		KGYSPEANICDKCPG
	CTATGGCAAACAGTACCAGCACAAGTCCGAACTCTCTAAGATAATTAGA	ID NO:	CGCACC		NRQATLGHISQTCPA
	GATAAATACTGGACAATGTGCGACGGGAAAGGCTTGAGGACAGAGGT	1214)	CAGAGC		THGRRVKRHDKIVNR
	GCCACCAGATACGGCTAAGAAAACACTAAGCCTGCTATTCGAGGACCG		CATGCTC		IAKALKERGSVKNILT
	GACTCCACTCAAACCAGGACAACTAATTGGAGCTATCGGGGTCCGCCTA		TTACCA		EPHLRHDKLPLRKPD
	AATACCCTTGGTACCCCCGCTCGTAACAATCGAGCTAAAGGCTACAGTC		GGCCAC		LIVHTEKSVEIIDVQV

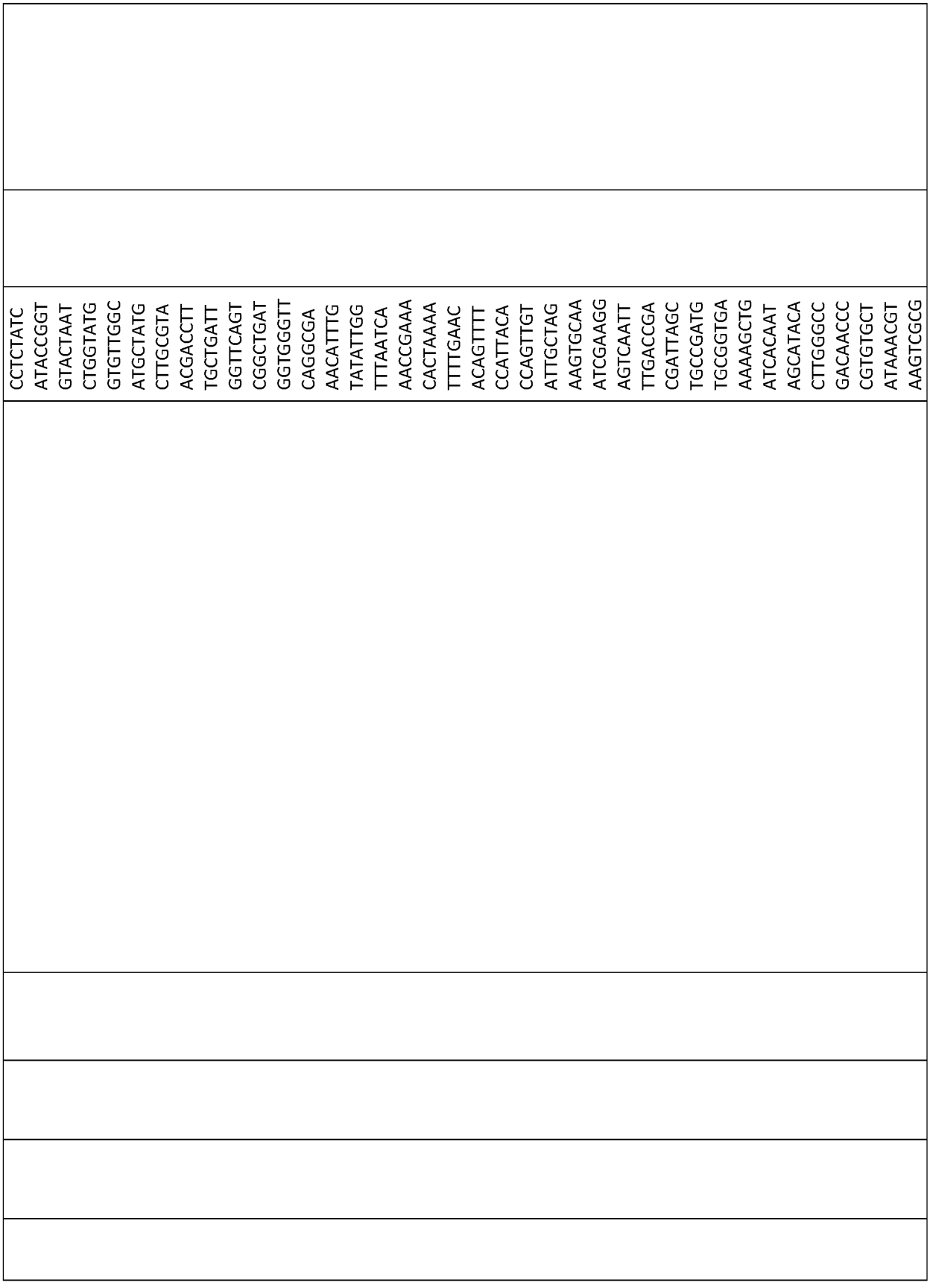
		CTGAGGCTAATATTTGTGACAAATGTCCAGGCAACAGGCAGGCCACTCT		TTGACG	VADQGISRHEDEDQ
		TGGCCACATCTCGCAAACTTGTCCGGCGACTCACGGGAGGAGAGTCAA		GCGCTA	QKKIVKYDVDGYKRA
		ACGACACGACAAGATTGTGAACAGAATCGCAAAAGCTCTAAAAGAACG		CACCGG	AYKMLGIDYGSIPCN
		AGGCTCCGTCAAGAACATACTTACCGAACCCCACCTCAGGCACGATAAA		TCTGTA	VSAFTITWRGNLAPH
		CTGCCCCTACGGAAACCCGACCTAATAGTACATACGGAAAAGTCGGTCG		GGAGGG	SLKLASRLQFSPVLKYI
		AAATCATTGATGTCCAGGTCGTTGCAGACCAGGGCATCTCGAGGCACGA		TATCTCA	VADSLVDTWGAFLI
		AGATGAAGACCAACAAAAGAAGATCGTCAAATATGACGTGGATGGATA		GCGACT	WGKTS (SEQ ID NO:
		CAAGAGAGCGGCCTACAAAATGCTAGGGATTGACTACGGGAGCATCCC		TGGACA	1459)
		CTGCAATGTAAGTGCGTTCACTATCACCTGGCGAGGAAACCTTGCCCCC		AAATGG	
		CATTCACTCAAACTGGCGAGTCGTCTCCAGTTTTCACCCGTGCTCAAATA		ATCTTGA	
		TATCGTGGCTGACTCCCTAGTTGACACATGGGGTGCGTTCCTGATATGG		CTGCCT	
		GGAAAGACGTCTTAGTGACGAGAAAAGTCGCTTTATCCTTACATAACAG		GAGGGC	
		TGTGATAGTCATCCTTATCACAACTGTCCTGGCGAACAAAATGGCAGAG		GCTTATC	
		GGATAGTACTCGGTCCAACCAGAAGGAAGCCCACACGATGCCAAACTT		GAGTGA	
		GCTGTAACCCACGTGAGCGGAAAACATCCTCCGAGTATAGTATGATGGA		CCCAGA	
		AGGATACAGGATGTGGACCGCTCTAGGCGGCGGACGGTGTAGACGGC		GTAAGC	
		GACCTTACCATAGCTGCGACTGCTGTGCGATCCGGAACGGGCCTTTCTC		TGGTAG	
		ACTTACCTCAAGCACCGTGTCGGATCGCGGGGTGGGCGCGGGATTCCT		GAAGAA	
		GCTGGGAAGTTGTCTGCACCCAAGTCGTCAGCTAACTGTCGGTACTTAT		TCTTGCA	
		ACGCACCCAGAGCCATGCTCTTACCAGGCCACTTGACGGCGCTACACCG		GTTGAG	
		GTCTGTAGGAGGGTATCTCAGCGACTTGGACAAAATGGATCTTGACTGC		AGGGCT	
		CTGAGGGCGCTTATCGAGTGACCCAGAGTAAGCTGGTAGGAAGAATCT		AGTAGG	
		TGCAGTTGAGAGGGCTAGTAGGGCCAGCCTCTGCTCGTCAACTGTCCTA		GCCAGC	
		GGCGTTAACGTGTGCCTCCCTATGAGGTAACCGTGAAGTATTACTTGTTT		CTCTGCT	
		CCCTGTAGAGACCATATAACTAGCTGAGAAACGCCCTGCGGGTTAATAT		CGTCAA	
		ACAGTTACCGCCGTCCTGCATCCCCGAGTTGTGACCAGCGTACAGCCAT		CTGTCCT	
		GTACCACGGCACTTCCAGCATTCTCTGGTCTAAGAATGTATCAGCTGGCC		AGGCGT	
		CGCCGAAAGATAGAGCGACCCCGCCTCTATCACTAGAGTAAGCAGGCA		TAACGT	
		CGCGAAATATATGCAGGACGCCTTGGCTCAGTAGCTCTGGTACTGAGTA		GTGCCT	
		TAGCATGGCTGAACGCCCCTTTAAGTTCGGTGGCCCAGAGACTCTCCGA		CCCTATG	
		GACTTCCCAGTCCTGGAGAGGGTACACCTTTACCAGAACTTTCTGTGGTC		AGGTAA	
		GGTTG (SEQ ID NO: 1091)		CCGTGA	
				AGTATT	
				ACTTGTT	
				TCCCTGT	
				AGAGAC	
				CATATA	
				ACTAGC	


TGAGAA
ACGCCC
TGCGGG
TTAATAT
ACAGTT
ACCGCC
GTCCTG
CATCCCC
GAGTTG
TGACCA
GCGTAC
AGCCAT
GTACCA
CGGCAC
TTCCAGC
ATTCTCT
GGTCTA
AGAATG
TATCAG
CTGGCC
CGCCGA
AAGATA
GAGCGA
CCCCGC
CTCTATC
ACTAGA
GTAAGC
AGGCAC
GCGAAA
TATATGC
AGGACG
CCTTGG
CTCAGT
AGCTCT
GGTACT
GAGTAT
AGCATG
GCTGAA

						CGCCCCT TTAAGTT CGGTGG CCCAGA GACTCTC CGAGAC TTCCCAG TCCTGG AGAGGG TACACCT TTACCA GAACTTT CTGTGG TCGGTT G (SEQ ID NO: 1337)	
R2	R2-1_PM	.	Petromyzon marinus	CTATTAATGGGATGAAGAAGGGGGACACGAGTTTGTGTGTGCATCCAG TTTCCATGGTGCATGCAGGAGTGGTGGTTTAAATGGCGAGACTCTACAG GGCTTCCATGGCTACACGGGATGCAAGGCATCAGACATTTTGGCACAG GCAATCCTTTTGGTCTCTACCGCAATCATGTCTTAGACCTCAGTAGCGAC CACTACAACCACAGTGGTGACTGCTGTTGAGTGAAGGACGACTGAGCG CTGGATAACAACTTTCTTGCGTGGCCCAACATCGAAGCAACCACTTCGG AGCTGGCACAAGGCAAGAGGGCAGCCCAAGGTGTGAATCATCTCAACT TCACTGCAGGAAGAAATGCTGTGCAAGGATGAGTGTGAACGACACCAA CGGGATTGTTGCTGACCAGGAGGTGCCAACCAAATTTGAATGGATTGAC TTTGGGCCTGGTTTCTCCTGCGTGTATTGCACGGAAAAACAAGTGGCTA CACGTGTGGCCGTCGTGTCCTGGGGTTTCGCAACACAACTCCACAAGAT CGACAACTATGAGGATGACAATGTACTTAAAGAACAAAGAGACTGACG CCAAAGGGGATTTAAACCGCCAAATCGTACATTGGGTCTCACTACAATT TTTTTACGTGTATTTATTTTCCTAAGTGTCTGTACTTGCCATTCTTCGCTGC TTTTTCTGCATTAATTGCATATCGTATGCAAATAAGCGAATTAACCACCA CCGTGCAACTATATGCAGATGTTACAGCTGAGCCCTCTATCATACCGGT GTACTAATCTGGTATGGTGTTGGCATGCTATGCTTGCGTAACGACCTTTG CTGATTGGTTCAGTCGGCTGATGGTGGGTTCAGGCGAAACATTTGTATA TTGGTTTAATCAAACCGAAACACTAAAATTTTGAACACAGTTTTCCATTA CACCAGTTGTATTGCTAGAAGTGCAAATCGAAGGAGTCAATTTTGACCG ACGATTAGCTGCCGATGTGCGGTGAAAAAGCTGATCACAATAGCATACA	CTATTAAT GGGATGA AGAAGGG GGACACG AGTTTGTG TGTGCATC CAGTTTCC ATGGTGCA TGCAGGA GTGGTGGT TTAAATGG CGAGACTC TACAGGGC TTCCATGG CTACACGG GATGCAAG GCATCAGA CATTTTGG CACAGGCA ATCCTTTT GGTCTCTA	TAATTTA AGGTAA AATCGT GGGATT GTTTTGA TGGCAA TCTGCCT AGTCGC GGCCTT CCATTTT GGGTAG GCAGCA GACCCA TCTATAT AACAAA CTACTTT GCCTTTC ATAGGG GTACCC GACCCT ACCAAC	MNERLTDELTTEFILS DMFLWDYPCTDQN KCYPCNLVFLDHRTW SSHMARVHPHANKT YKCRICNRTADSIHKI ASHYGRTCKSLIGKTN AITTTIDETLFSCLHCS RGFTTKTGLGVHTRR THPTEHEAILQQNTP GRKVRWGEEEVEIM AHKEAQQKDEDINM NQLIQNSVMPHRTLE AIKGKRRNIKYKELVR TLKETTYKVENQCLV NLVLPTTSEITTTPSE GDQPAIRAEKEQSPT AAEDLQVIINDLKSQ NFSHNQALLLLNSHV EKFLNRSKPIKRKDHV NQQEIDENRHRRQS KQTKYRRYQYLYHTN

	CTTGGGCCGACAACCCCGTGTGCTATAAACGTAAGTCGCGAATTATAAA	CCGCAATC	TTTCGG	KKALLDEITSDRSGPSI
	GAAAACAAACCGGACGGACTACTCGGTGACGAACTAACATCGCTCATG	ATGTCTTA	GGAAGT	YPTEESIRGTFVTLFES
	AATGAGCGATTAACAGACGAGCTGACTACGGAATTTATCCTTTCGGACA	GACCTCAG	AAAAGA	NSPPDNIPSKLKNDQ
	TGTTTTTATGGGACTACCCATGCACAGATCAGAACAAATGTTATCCATGC	TAGCGACC	AA (SEQ	SCIDIVKAITLDELIKTL
	AATCTTGTTTTTCCTAGACCACAGAACTTGGTCATCACATATGGCACGGGT	ACTACAAC	ID NO:	AIMKDKSPGQDNITL
	ACATCCACATGCAAACAAAACGTATAAATGTCGAATTTGTAATCGCACA	CACAGTGG	1338)	SDLRTLPIKYLLDILNII
	GCAGATAGCATACACAAGATAGCGTCACACTACGGAAGAACTTGCAAA	TGACTGCT		LYIQDIPQIWKQHRT
	AGTTTAATAGGTAAAACTAATGCTATAACCACCACAATTGATGAAACACT	GTTGAGTG		RLIPKTKEELEKPSNW
	ATTTAGTTGTTTACATTGCAGCAGAGGTTTTACTACGAAAACAGGTTTAG	AAGGACG		RPITISSIVIRLLHKILSY
	GGGTACATACTAGACGAACTCATCCGACAGAACATGAGGCTATACTACA	ACTGAGCG		RLGQQLKLNYRQKAF
	GCAAAACACACCAGGAAGGAAAGTTAGATGGGGAGAAGAAGAAGGTAG	CTGGATAA		LPVDGCFENSALLHFI
	AAATTATGGCCCATAAAGAAGCCCAACAGAAGGATGAGGACATAAACA	CAACTTTC		IHNARQKHENTQIVSI
	TGAATCAACTAATACAGAACTCAGTTATGCCACACAGAACGCTAGAAGC	TTGCGTGG		DLSKAFDSVSHESIIR
	GATTAAAGGGAAGCGGAGAAATATCAAGTATAAGGAATTGGTAAGGAC	CCCAACAT		ALNRFNLSKESITYLT
	TTTGAAAGAAACTACCTATAAGGTAGAAAATCAATGCCTTGTTAACTTAG	CGAAGCAA		NIYKCNLTDIVFGSTI
	TTTTACCGACAACATCGGAAATAACAACTACACCTTCGGAAGGAGATCA	CCACTTCG		MRNINLKRGVKQGD
	GCCAGCAATAAGGGCCGAAAAAGAACAATCACCGACACAGCAGCTGAGGA	GAGCTGGC		PLSPLLFNMIMDELL
	TCTTCAGGTCATAATTAACGATCTAAAGAGCCAGAATTTTTAGCCACAATC	ACAAGGCA		DNLPTYIGVNVGNQ
	AGGCGTTACTGCTACTCAATTCTCATGTAGAAAAGTTTTTAAATCGAAGT	AGAGGGC		KVNSMMFADDLILF
	AAACCAATTAAAAGGAAAGATCACGTAAACCAACAGGAGATAGATGAG	AGCCCAAG		AETECGMNKLLDITT
	AATAGGCATCGAAGACAATCAAAGCAAACTAAATACAGGAGATATCAA	GTGTGAAT		KFLDDRHLKININKCN
	TATTTATACCATACGAACAAGAAAGCTCTATTAGACGAGATTACTTCAGA	CATCTCAA		SLRFIKYGKQKTFSVA
	TAGATCGGGGCCAAGTATATACCCAACTGAGGAAAGCATACGGGGAAC	CTTCACTG		TTSSYFINNEPINPVS
	ATTCGTTACTTTATTCGAGTCAAACTCTCCTCCAGATAATATACCCTCTAA	CAGGAAG		YVKGFKYLGIEFDPRG
	ATTAAAAAACGACCAATCCTGCATCGATATCGTAAAAGCAATCACCTTA	AAATGCTG		KRSISCNLLAAMLNKL
	GATGAGTTGATTAAGACCCTAGCAATTATGAAGGATAAGTCACCTGGAC	TGCAAGGA		TRAPLKPEKKVYLINN
	AGGACAACATTACTCTTAGCGATCTTAGGACTTTACCAATAAAATATTTA	TGAGTGTG		NLIPRIIHQLVLGKVT
	CTAGATATCTTAAATATCATCCTTTACATACAGGATATACCACAAATATG	AACGACAC		KGLLMSLDSEIRKTVK
	GAAACAGCACAGAACAAGACTTATCCCGAAAACTAAAGAGGAATTAGA	CAACGGG		LLLRLPHDTPDSFFYT
	AAAACCCTCAAATTGGAGACCCATAACCATCTCATCAATTGTAATTAGGC	ATTGTTGC		SVSNGGMGIRNLCD
	TATTACATAAAATTTTAAGTTATCGTCTAGGACAGCAATTAAAGCTTAAT	TGACCAGG		SVALSIINRHNKLITSD
	TACAGGCAGAAAGCATTCCTCCCGGTAGACGGATGTTTCGAAAATAGTG	AGGTGCCA		DLVIRALSQQSYTIAT
	CATTACTACACTTCATCATACACAACGCCAGGCAAAAGCACGAAAACAC	ACCAAATT		LKQAHIIAGSKFPSKS
	GCAAATAGTGTCAATAGACCTCAGTAAGGCATTCGATTCTGTCAGCCAC	TGAATGGA		LNQNKWSNKLYQTT
	GAATCGATTATTAGAGCCTTAAACCGATTTAACTTATCAAAGGAATCCAT	TTGACTTT		DGRGLVYCQSQTEN
	AACGTACTTGACCAACATCTATAAGTGTAATCTAACTGATATTGTATTTG	GGGCCTG		NSWITGNHRTIKSYN
	GATCGACAATAATGCGTAACATAAATCTAAAAAGAGGCGTAAAGCAAG	GTTTCTCC		YIDMVKLRINALPTKS
	GAGATCCACTTTCACCGTTACTATTTAACATGATTATGGATGAATTATTA	TGCGTGTA		RCNRGTLETKQCRFK

		GATAACTTGCCGACATATATAGGAGTTAATGTAGGAAATCAGAAAGTAA	TTGCACGG		CRSINNQISEETLAHIL
		ATTCTATGATGTTTGCAGACGACCTTATCCTATTTGCAGAAACGGAATGT	AAAAACAA		QKCDRSHYSRIARHD
		GGCATGAATAAACTCTTAGATATAACTACTAAATTCCTCGATGACAGACA	GTGGCTAC		SLVQFLATAAQKLN
		CTTGAAAATAAATATAAACAAATGCAATTCGTTAAGATTTATCAAGTACG	ACGTGTGG		WEVIKEPTLPSDTNK
		GCAAACAGAAGACATTTAGTGTTGCAACGACATCATCGTACTTTATAAAT	CCGTCGTG		AKPDLILVRDSHVLIV
		AACGAACCCATTAATCCGGTATCATATGTAAAGGGATTCAAATATCTAG	TCCTGGGG		DVAVPWESRSLAHA
		GCATTGAATTTGACCCAAGGGGAAAACGATCTATAAGCTGTAACCTGCT	TTTCGCAA		YDFKVKKYATDKKM
		CGCAGCAATGTTAAACAAACTGACCAGAGCACCGTTAAAGCCAGAAAA	CACAACTC		QAYLKTIYPEKEIRTEA
		GAAAGTATATTTAATCAATAACAATTTAATACCTCGTATTATTCATCAATT	CACAAGAT		LIISARGGWCALNN
		GGTCCTCGGAAAAGTTACCAAGGGTTTATTGATGTCACTTGATTCTGAA	CGACAACT		MVTKKVGLSSAWVK
		ATTAGGAAAACAGTAAAGCTTCTGCTCAGGTTGCCACACGATACGCCCG	ATGAGGAT		LALIKVMEGSVKIWR
		ACAGTTTCTTTTATACATCAGTATCCAACGGAGGAATGGGTATAAGAAA	GACAATGT		SWSKG (SEQ ID NO:
		TTTATGCGACTCAGTTGCACTATCTATAATAAACAGACACAACAAATTGA	ACTTAAAG		1460)
		TAACTTCAGATGATCTAGTAATAAGAGCATTATCACAACAATCATACACT	AACAAAGA		
		ATTGCAACGTTAAAACAGGCCCATATCATTGCAGGCTCCAAATTTCCTTC	GACTGACG		
		GAAATCTTTAAATCAGAACAAATGGTCAAATAAACTATATCAAACAACA	CCAAAGG		
		GATGGTCGGGGGGTTGGTATACTGCCAATCTCAAACAGAAAACAATTCAT	GGATTTAA		
		GGATAACAGGGAATCATAGAACAATAAAATCGTATAATTACATAGACAT	ACCGCCAA		
		GGTTAAACTAAGGATTAATGCACTACCGACTAAATCGAGATGCAATCGA	ATCGTACA		
		GGGACGTTAGAGACCAAGCAATGTAGATTTAAATGTCGAAGTATTAACA	TTGGGTCT		
		ACCAAATTTCAGAGGAAACATTGGCACATATCTTGCAAAAGTGTGATCG	CACTACAA		
		CAGTCATTATTCAAGAATCGCAAGGCATGATTCTTTGGTGCAATTTCTGG	TTTTTTTAC		
		CAACGGCCGCACAAAAACTAAACTGGGAAGTGATCAAAGAACCCACTTT	GTGTATTT		
		ACCGAGCGATACAAATAAGGCAAAACCGGACTTAATTTTAGTAAGAGAC	ATTTTCCT		
		TCTCATGTCTTGATAGTAGATGTGGCAGTTCCGTGGGAGTCTCGATCATT	AAGTGTCT		
		GGCACATGCATACGATTTTAAGGTGAAAAAATACGCTACTGACAAAAAA	GTACTTGC		
		ATGCAAGCATATTTAAAAACTATATATCCGGAAAAAGAAATTAGAACGG	CATTCTTC		
		AGGCTTTAATCATATCTGCACGTGGGGGCTGGTGCGCTTTAAATAATAT	GCTGCTTT		
		GGTAACAAAAAAAGGTGGGATTGTCAAGTGCATGGGTAAAATTAGCATT	TTCTGCAT		
		GATCAAGGTCATGGAGGGTTCCGTAAAGATATGGCGCTCTTGGAGCAA	TAATTGCA		
		AGGATAATTTAAGGTAAAATCGTGGGATTGTTTTGATGGCAATCTGCCT	TATCGTAT		
		AGTCGCGGCCTTCCATTTTGGGTAGGCAGCAGACCCATCTATATAACAA	GCAAATAA		
		ACTACTTTGCCTTTCATAGGGGTACCCGACCCTACCAACTTTCGGGGAAG	GCGAATTA		
		TAAAAGAAA (SEQ ID NO: 1092)	ACCACCAC		
			CGTGCAAC		
			TATATGCA		
			GATGTTAC		
			AGCTGAGC		

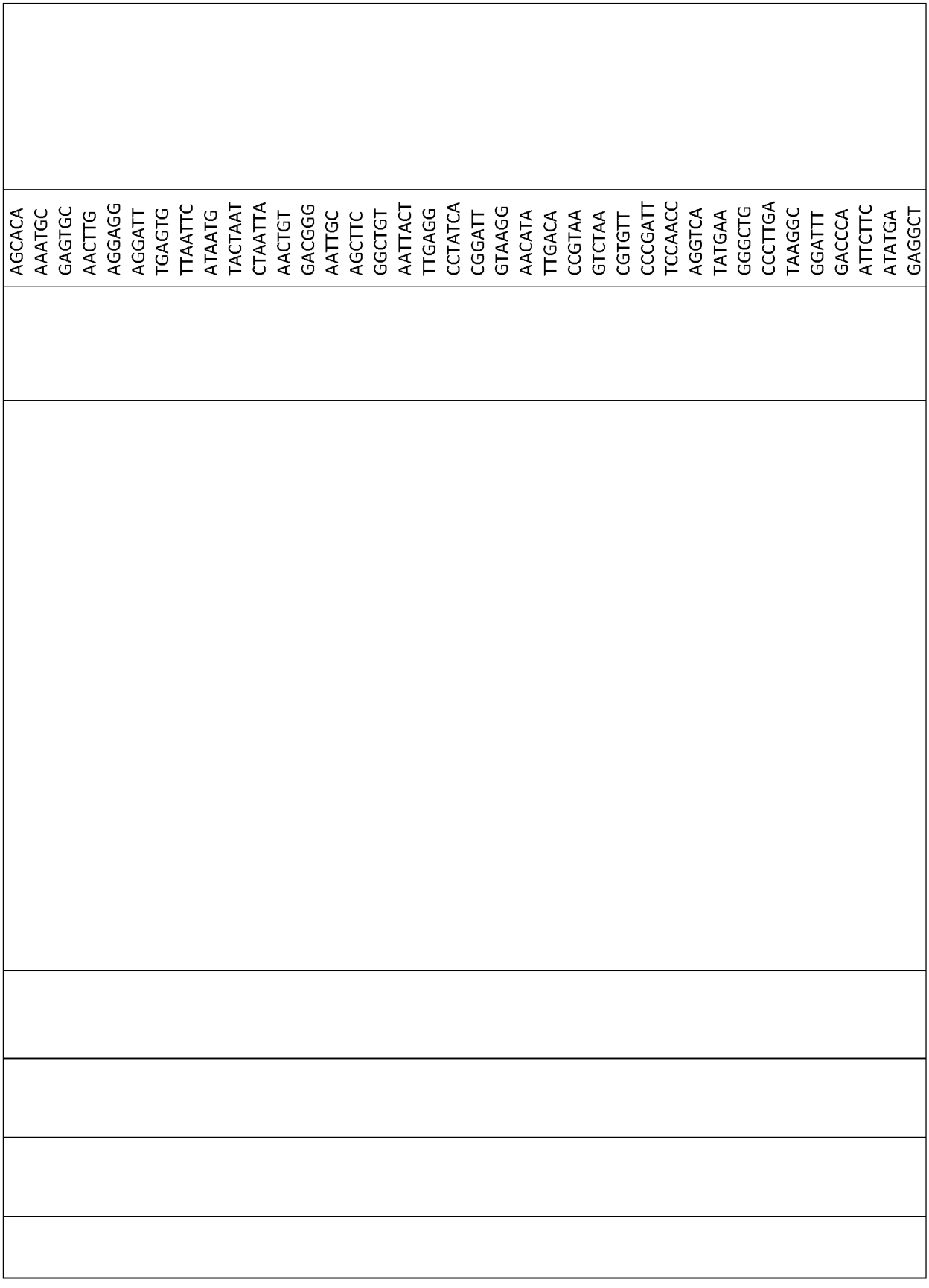

CCTCTATC
ATACCGGT
GTACTAAT
CTGGTATG
GTGTTGGC
ATGCTATG
CTTGCGTA
ACGACCTT
TGCTGATT
GGTTCAGT
CGGCTGAT
GGTGGGTT
CAGGCGA
AACATTTG
TATATTGG
TTTAATCA
AACCGAAA
CACTAAAA
TTTTGAAC
ACAGTTTT
CCATTACA
CCAGTTGT
ATTGCTAG
AAGTGCAA
ATCGAAGG
AGTCAATT
TTGACCGA
CGATTAGC
TGCCGATG
TGCGGTGA
AAAAGCTG
ATCACAAT
AGCATACA
CTTGGGCC
GACAACCC
CGTGTGCT
ATAAACGT
AAGTCGCG

				AATTATAA AGAAAACA AACCGGAC GGACTACT CGGTGACG AACTAACA TCGCTC (SEQ ID NO: 1215)			
R2	R2-1_SM	.	Schmidtea mediterranea	CAGTGCTATTCGAATGTCAATGTGAAGAAATTCAACTAAGCTCTGGTTA	CAGTGCTA	GGCCAC	MKKVLNNETEKLPGS
				ACGGCGGGAGTAACTATGACTCTCTTAAGGAATTAAGAATTTACCTGCC	TTCGAATG	GCGCGT	NLTFMCGFCDREFDT
				KTAKTAAAARTGAAATCMGTTGTTCATWGCAAGTGGTATTGTACACCTT	TCAATGTG	CGTCCTT	ARGRGVHESRGHLV
				CCCGCGGTGCTAGTCGTTTAAAACTAAGTTACAAACCACGAGGGGCGTC	AAGAAATT	GTTTGAT	ERDAAVQSRVKAVV
				CTGACGGACTGSAAAAGCATTGAGRGTCMTGAAGAGAGGCTCTTATTG	CAACTAAG	CACTAG	SKKYYSNEEDVALA
				TACGAATCTCTTCAACGATCGAAGTCTGGACCGATATGAGAACTAATAC	CTCTGGTT	TGGATC	KMQLXHADLAKSEXL
				ATTAGTTGACAGGTGAAAAATACTGTTGATTACTTAGTTCTCAGTCATGT	AACGGCG	AACCTTC	EAMYLALGKGRTREA
				GGTATATTGCCAGTCAATTACTACATWAATATTAGTGTGGCTCTCAAAG	GGAGTAAC	GACTCC	IEQHIRKSLRYKGVLE
				GAACACGATTGRTCGGCAGTCCAATGCGCGACTGGCGGGCTTGTTGTTT	TATGACTC	CCGGAA	EQRKLLETARGNVRQ
				GCATTTGTTACCGGCTACTTGAAAAGGTTATATATAGCAGACGCTTAAA	TCTTAAGG	CTGTGG	NNVGVPASNATKNL
				GCGCGACTGTAATTTACATYTCATTGCCCAGTATTTGTCTTTTGTCAGATT	AATTAAGA	GAGTGG	QRFLESLPLGTNRREE
				TAGCAAAATTTCATATTTTGTTAATTACCTTAACTGGTTAAACGATCCCAT	ATTTACCT	CGGAAG	RLDRIIRSNSIESQRLE
				AATTGCTTGCAATTATTATAAAGTAATTCAGGTAAAAATTACATATCTGG	GCCKTAKT	AAAGGC	LIHYCNDMCQDFVQ
				CTGATCCTGCCAGTAGTCATTTTACTTCCGCCGCGCTATAAAACAGTTTA	AAAARTGA	CAGAGG	LDCQXNPINAIRRRN
				AAAACTGAATAGGAATCAAAAAGAACATGGCAAGCGACTATATGTAAC	AATCMGTT	ATGTCCT	PKRLSKKQLKRAKFSA
				TGGGCATTCAACATTCCCTATTACATATGGTGGTGCCTGGGGTCTGTTTT	GTTCATW	GAAACC	LQRLWIRDRKAAAQL
				ATATAATGGGTACCCGGGAAGTGGATCTGTATCACCAGTCATGGTGCCA	GCAAGTG	ATATATT	VLKDKLDSLLSNKEDS
				TATCTTTKGATAAAGATACAGTTTAAAACTGCGATGATACTAATAGAGAT	GTATTGTA	TATTTAT	KDLGSYWQQVFERE
				CCTCTTAGACCTTCGTAAAGAAGTGGGGATTGATGACATTAGCATTGGA	CACCTTCC	AGAAGT	SELDRRPIPQVVENEE
				AGAATTAAATCTCCAAGGAAATGGAGTAACTTCAATGAAGTCCCACAAC	CGCGGTGC	TTTACTT	LNSPVLEKEVEWAVK
				CCCGTTGAAGGGCTGGGTTCGAGTATCGAGAGAAAACTCTAAATTCTCT	TAGTCGTT	CATCCTA	NIKKSTAAGPDGLTA
				TCGGTTMTGTCCAACGGAGGGGACATTACTGTAAAATATCCTCTAAAAA	TAAAACTA	TTTACGT	LALKKIPYSELVKLFNII
				CAACTATGAAAAAAAGTCTTAAATAATGAAACCGAAAAATTACCGGGAAG	AGTTACAA	ATTTCAG	LLVGFLPDVLKNSRTI
				TAATCTAACTTTCATGTGCGGGTTCTGTGATCGGGAATTCGACACGGCW	ACCACGAG	TATGAA	LIPEVDNPQGGGDYR
				AGGGGCAGAGGAGTGCACGAGAGTAGAGGTCATTTAGTTGAGCGGGA	GGGCGTCC	AATGAG	PISINSVLTRTLNKILA
				TGCGGCGGTTCAGAGCAGAGTGAAAGCCGTGGTGAGTAAAAAGTATTA	TGACGGAC	TAAAGT	KRVSEGDFGINGQKG
				TTATAGTAACGAAGAGGATGTGGCATTAGCGAAAATGCAGTTASAGCAT	TGSAAAAG	TCTCGAC	FKSVDGCLENLATVE
				GCAGATCTGGCCAAAAGCGAACWATTAGAAGCCATGTATCTTGCATTG	CATTGAGR	TCGATG	SILADARMKKNKLAV
				GGAAAGGGAAGAACTCGTGAAGCCATAGAGCAACACATAAGGAAATC	GTCMTGA	AGTTGG	VFLDMSKAFDSVNH

		GTTACGTTATAAGGGGGTCCTTGAAGAACAGCGAAAGCTCCTTGAGAC	AGAGAGG	GGGCAA	ESIVRAGEIKGYPKLL
		AGCAAGGGGAAATGTTCGGCAAAATAACGTGGGTGTGCCAGCTAGCAA	CTCTTATT	CCATTG	MTYVKECYNDATTN
		TGCCACTAAAAATCTGCAAAGATTCTTAGAATCGTTACCCTTGGGAACG	GTACGAAT	GGGGTC	VAGVTAKFNRGVKQ
		AATAGGCGCGAGGAACGATTGGATAGGATTATCCGATCTAACTCGATC	CTCTTCAA	CTGAAG	GDPLSPALFNNVIDL
		GAAAGCCAAAGACTTGAATTGATCCACTATTGTAACGATATGTGTCAAG	CGATCGAA	AGAGGC	AIERVSGTGIGYNMG
		ACTTTGTGCAACTAGACTGTCAAARGAACCCCATCAATGCTATAAGGCG	GTCTGGAC	TCTCACT	GKKYSVVAYADDLVL
		CAGAAATCCGAAAAAGACTATCGAAAAAGCAGTTGAAGAGAGCTAAATT	CGATATGA	GTAAAA	FGESREGLQIALTALL
		TAGCGCTCTTCAACGGCTTTGGATAAGGGATCGGAAAGCTGCTGCACAG	GAACTAAT	AATCTCT	EELKLNGLTPNPAKS
		TTAGTGTTGAAGGATAAGCTTGATAGTTTGCTCAGCAATAAAGAGGATT	ACATTAGT	TCGTGT	ASLTFERSGPHWFAS
		CCAAGGATTTGGGATCGTATTGGCAACAGGTCTTCGAACGTGAGTCCGA	TGACAGGT	CTGTTTA	TDTVTALGDQIPAM
		ATTAGACCGCAGACCCATACCACAAGTGGTGGAAAACGAAGAGTTAAA	GAAAAATA	TTCCTAG	GNIETYKYLGIKFNSC
		TTCCCCGGTATTAGAGAAGGAAGTAGAGTGGGCTGTCAAAAACATTAA	CTGTTGAT	GCACNT	GVVKGSLPGIYTKKLE
		GAAGTCCACTGCCGCAGGACCAGACGGGTTAACGGCACTTGCCTTGAA	TACTTAGT	GCTGCA	LISKAPLKPQQRLAM
		GAAAATACCGTATTCCGAGCTAGTCAAACTATTTAATATAATACTGTTGG	TCTCAGTC	TTATGA	LTDFLIPGVLHQAVF
		TGGGATTCCTACCTGATGTATTAAAAAATAGTAGAACTATCCTAATACCC	ATGTGGTA	AGCGG	GQTNAGDLRSLDKRT
		GAAGTGGATAATCCCCAAGGGGGYGGTGATTATAGACCGATTTCGATC	TATTGCCA	WGAAA	RRAVRSWCHLPSDTS
		AATTCAGTGCTCACTAGAACACTAAATAAGATCCTAGCGAAACGAGTCT	GTCAATTA	GTAAAG	TAFIHAKAKDGGMGI
		CGGAAGGTGATTTTGGTATCAATGGTCAAAAAGGATTCAAAAGTGTAG	CTACATW	TTAGAG	PSIRAEVQFGKLDRF
		ATGGTTGTCTAGAGAATCTAGCAACAGTTGAATCAATTTTGGCCGATGC	AATATTAG	CTGAGA	GKLPNVKDERSKVLA
		TAGAATGAAGAAAAATAAGCTTGCGGTAGTATTCTTGGATATGAGTAAA	TGTGGCTC	GATAGG	DNAHIKKKMLEKLGV
		GCTTTTGATTCTGTAAACCACGAGAGTATTGTTAGAGCTGGAGAAAATCA	TCAAAGGA	TACTTGC	GIPIKGVRCKNKLEFY
		AAGGTTATCCCAAACTATTAATGACGTATGTGAAGGAGTGTTATAACGA	ACACGATT	TGCATTA	NKMREELIKSNDGIG
		CGCTACTACGAACGTCGCAGGTGTAACAGCCAAGTTTAACCGGGGAGTT	GRTCGGCA	TGAAGC	LKEASLVPSANTWLK
		AAACAAGGTGATCCTCTTTCGCCGGCGTTATTTAATAACGTAATAGACCT	GTCCAATG	GGAGAA	LSDLHMSGRTFVGCL
		TGCAATTGAGCGAGTTTCCGGTACTGGAATTGGATATAATATGGGCGGT	CGCGACTG	AGGCCT	KTRGNLMATVTRTSR
		AAGAAGTATTCTGTGGTCGCGTATGCTGACGATCTTGTTCTATTCGGTGA	GCGGGCTT	CGAATA	GGQNPGIELNCKKGC
		GTCGAGAGAGGGGTTGCAAATAGCCCTGACGGCGCTACTGGAGGAATT	GTTGTTTG	AATAGG	QYQGSLNHIVQKCPV
		AAAGCTAAACGGTCTGACACCCAATCCAGCGAAAAGTGCTTCGTTGACA	CATTTGTT	GTGTTA	VKGLRIKRHDEVVKY
		TTTGAGAGATCAGGGCCACATTGGTTCGCAAGTACCGATACCGTTACAG	ACCGGCTA	GAGTTA	VEEITKKAGWSATME
		CACTAGGAGATCAAATACCGGCCATGGGTAACATCGAAACCTATAAATA	CTTGAAAA	TTGATG	PIIPFEGSHRKPDLVL
		CTTAGGAATCAAGTTTAATTCGTGCGGAGTGGTTAAAGGGAGTCTACCC	GGTTATAT	GAGAGT	VRGDLGKVVDIQIVS
		GGGATATACACCAAGAAATTGGAGTTAATCTCTAAGGCTCCTTTGAAGC	ATAGCAGA	ATACTA	DHCGLDEKNSCKIGK
		CGCAACAGCGGTTAGCTATGCTAACCGATTTCTTGATTCCTGGAGTATTA	CGCTTAAA	GTAAGC	YDNDIIRNYVRGLGP
		CACCAAGCCGTGTTTGGACAGACAAATGCAGGGGACCTGCGCAGTCTC	GCGCGACT	TTAAGCT	SRVEVAAITLNWRGV
		GATAAACGGACGAGAAGAGCGGTGAGATCTTGGTGTCACCTTCCCTCTG	GTAATTTA	GCGCCT	WSRDSFNLIKRLGMT
		ATACGTCAACGGCATTCATCCATGCTAAAGCTAAGGATGGAGGAATGG	CATYTCAT	CGCGCG	EMDAKIISMRVLAST
		GAATTCCATCTATAAGAGCCGAAGTCCAATTTGGAAAATTGGATAGATT	TGCCCAGT	GTGCCC	AKMFKTCKKVLEPVC
		TGGAAAGTTACCCAATGTCAAAGATGAACGATCKAAAGTTTTGGCCGAT	ATTTGTCT	AAAAAT	RTKTADCDGYGPEET

	AATGCTCACATTAAAAAGAAGATGTTAGAGAAATTGGGAGTGGGGATC	TTTGTCAG	ATACTTA	SARPCHELNLKESSGT * (SEQ ID NO: 1461)
	CCAATCAAAGGAGTGCGTTGTAAGAACAACTCGAGTTCTACAACAAGA	ATTTAGCA	ATGAGA	
	TGCGTGAAGAGTTGATTAAGTCGAATGATGGCATTGGCTTAAAAGAGG	AAATTTCA	GCAATA	
	CATCTTTGGTTCCCTCTGCGAACACTTGGTTAAAACTGAGTGACCTACAT	TATTTTGTT	ACTCAA	
	ATGAGCGGTCGCACATTTGTAGGATGTCTTAAAACACGGGGAAACCTTA	AATTACCT	GGNGAG	
	TGGCTACCGTAACAAGAACTAGCAGAGGAGGTCAGAACCCGGGTATAG	TAACTGGT	TTTAATT	
	AGTTGAACTGTAAAAAGGGATGCCAATATCAGGGAAGTCTGAACCATAT	TAAACGAT	CATATG	
	AGTCCAAAAGTGCCCAGTAGTGAAAGGGTTGAGGATAAAACGACATGA	CCCATAAT	CGCATG	
	TGAAGTCGTTAAGTATGTGGAAGAGATCACGAAAAAGGCTGGATGGTC	TGCTTGCA	CGGCAC	
	TGCAACAATGGAACCAATAATTCCGTTCGAGGGGTCACACCGAAAACCG	ATTATTAT	CAAGGT	
	GATTTAGTTCTGGTCCGGGGCGATCTCGGGAAAGTTGTGGATATTCAAA	AAAGTAAT	GCTGAA	
	TAGTTTCTGATCACTGCGGTCTGGATGAAAAAAATAGTTGTAAGATCGG	TCAGGTAA	TGGCAT	
	TAAGTATGATAATGACATCATACGAAATTATGTAAGGGGGCTAGGACCA	AAATTACA	CGATTA	
	TCGAGAGTGGAGGTAGCGGCGATAACCCTTAATTGGCGAGGCGTGTGG	TATCTGGC	AACCTCT	
	AGTAGAGACTCATTCAATCTCATTAAGAGATTAGGAATGACAGAAATGG	TGATCCTG	CCTGTT	
	ATGCGAAAATCATTTCTATGAGAGTGCTGGCGAGCACTGCGAAGATGTT	CCAGTAGT	GTAGAA	
	TAAGACGTGTAAAAAAGTGTTAGAGCCCGTTTGCAGAACCAAGACCGCT	CATTTTAC	GCAGGT	
	GATTGTGATGGTTACGGACCGGAGGAAACCTCCGCCCGACCATGTCATG	TTCCGCCG	CATAAA	
	AGCTTAACTTGAARGAGAGTTCGGGGACGTAAGGCCACGCGCCGTCGTC	CGCTATAA	TGGAGG	
	CTTGTTTGATCACTAGTGGATCAACCTTCGACTCCCCGGAACTGTGGGA	AACAGTTT	RGGGCA	
	GTGGCGGAAGAAAGGCCAGAGGATGTCCTGAAACCATATATTTATTTAT	AAAAACTG	ACCACT	
	AGAAGTTTTACTTCATCCTATTTACGTATTTCAGTATGAAAATGAGTAAA	AATAGGAA	GAAACT	
	GTTCTCGACTCGATGAGTTGGGGGCAACCATTGGGGGTCCTGAAGAGA	TCAAAAAG	TATGAG	
	GGCTCTCACTGTAAAAAATCTCTTCGTGTCTGTTTATTCCTAGGCACNTG	AACATGGC	CCAGAA	
	CTGCATTATGAAGCGGWGAAAGTAAAGTTAGAGCTGAGAGATAGGTA	AAGCGACT	GAAGCT	
	CTTGCTGCATTATGAAGCGGAGAAAGGCCTCGAATAAATAGGGTGTTA	ATATGTAA	TAACTAC	
	GAGTTATTGATGGAGAGTATACTAGTAAGCTTAAGCTGCGCCTCGCGCG	CTGGGCAT	AWAAGT	
	GTGCCCAAAAATATACTTAATGAGAGCAATAACTCAAGGNGAGTTTAAT	TCAACATT	TTTAGG	
	TCATATGCGCATGCGGCACCAAGGTGCTGAATGGCATCGATTAAACCTC	CCCTATTA	CAATTAC	
	TCCTGTTGTAGAAGCAGGTCATAAATGGAGGRGGGCAACCACTGAAAC	CATATGGT	TGAACG	
	TTATGAGCCAGAAGAAGCTTAACTACAWAAGTTTTAGGCAATTACTGAA	GGTGCCTG	GAGTTA	
	CGGAGTTAACTGTTAGTTAACACTACCATGTAGTTGTTTATAAAGCAAAT	GGGTCTGT	ACTGTTA	
	ATCAGGTTTCAGTCTATATACTAAAAGTATTTTTTGATACCGTGGTATAT	TTTATATA	GTTAAC	
	AGGCAACTAGTTAGGAAATAGTAAGGGATGACGCATTGTCTCTCTTTAT	ATGGGTAC	ACTACC	
	GGTACTGAGGAAACTTATCGACTCGCGAGGGATGAAACCCGTAAACCG	CCGGGAA	ATGTAG	
	ATCGATYTAGCCTATAAGTACCAGCGACAGTTAAACCATCTTACGCGAG	GTGGATCT	TTGTTTA	
	GGGTAAAACCTGAGGACCGATTATGGTATAACTTCTCAAGATTAGCACA	GTATCACC	TAAAGC	
	AAATGCGAGTGCAACTTGAGGAGGAGGATTTGAGTGTTAATTCATAAT	AGTCATGG	AAATAT	

	GTACTAATCTAATTAAACTGTGACGGGAATTGCAGCTTCGGCTGTAATT ACTTTGAGGCCTATCACGGATTGTAAGGAACATATTGACACCGTAAGTC TAACGTGTTCCCGATTTCCAACCAGGTCATATGAAGGGCTGCCCTTGATA AGGCGGATTTGACCCAATTCTTCATATGAGAGGCTTATTCCAGCCTTCCC GTAGTACCGTGAGGTTTTCCCGCCTCGAACGGAACAATGTTGCAGGGTA ATTAAGTACATCGGGCTATATMGCGATATTTAACGTTTTA (SEQ ID NO: 1093)	TGCCATAT CTTTKGAT AAAGATAC AGTTTAAA ACTGCGAT GATACTAA TAGAGATC CTCTTAGA CCTTCGTA AAGAAGT GGGGATT GATGACAT TAGCATTG GAAGAATT AAATCTCC AAGGAAAT GGAGTAAC TTCAATGA AGTCCCAC AACCCCGT TGAAGGG CTGGGTTC GAGTATCG AGAGAAA ACTCTAAA TTCTCTTC GGTTMTG TCCAACGG AGGGGAC ATTACTGT AAAATATC CTCTAAAA ACAACT (SEQ ID NO: 1216)	CAGGTT TCAGTCT ATATACT AAAAGT ATTTTTT GATACC GTGGTA TATAGG CAACTA GTTAGG AAATAG TAAGGG ATGACG CATTGTC TCTCTTT ATGGTA CTGAGG AAACTT ATCGAC TCGCGA GGGATG AAACCC GTAAAC CGATCG ATYTAGC CTATAA GTACCA GCGACA GTTAAA CCATCTT ACGCGA GGGGTA AAACCT GAGGAC CGATTAT GGTATA ACTTCTC AAGATT			


AGCACA
AAATGC
GAGTGC
AACTTG
AGGAGG
AGGATT
TGAGTG
TTAATTC
ATAATG
TACTAAT
CTAATTA
AACTGT
GACGGG
AATTGC
AGCTTC
GGCTGT
AATTACT
TTGAGG
CCTATCA
CGGATT
GTAAGG
AACATA
TTGACA
CCGTAA
GTCTAA
CGTGTT
CCCGATT
TCCAACC
AGGTCA
TATGAA
GGGCTG
CCCTTGA
TAAGGC
GGATTT
GACCCA
ATTCTTC
ATATGA
GAGGCT

R2	R2-1_SP	.	Strongylocentrotus purpuratus	TCTCGCGACGCGTTCTTCTGCCTGATGAAGTCACGTCAGGTAATAGACTT AGAAGGTTGATGAGCGTTCCTCTCCTGGACCGGGGGTGAGGATGTGTT CTACTGAATCATCGTTCCGGTTGTGAGGTCCGCTGCAAATAGGCCTTGG GGTGGTCTACCTCCGGGTCGCTACTCCTTTTGAGCTAGTCGATCCAGGT GAGAGTCGGGGAAGCCCACTTAGGTGGGCCAGCTAAGCAGATCACCCC CCCAGCACGACAGTGCGCTGTAATATAGCCTGTTGAGAGTGCACCCATT TATACTGTTATAAAAGTATAAATGGTTCTTAGCTAACCTATCCAATGATT TTTATTGTTATGTATAATACTGACCACTTTAGTCTAGTGTCATGTAAGGA TCCGAGACAATAGCCTATGTCTCTAGTCTAGCAATAACACAAGAGGGAT AATCCACCTCCCTACTCCAGCCACCACCTTTACCTTTCTTATCTTCCATATA GATGCCAACCAGACGGGTCGATGGGCAGTGAGAAACAAGAATGGGGA TAATTTGGAATTTGACATTCTTTCCTAAAGCAGACCTGAAGGGTTAAGA GTCACAAGCGGAGTCCGGACTGTCTTTATAAACAGACAATATTTTCTTAT TTTACCACCTATATGGGGATATTTCCCAACTCGTAATATAGGGCCCACTG TCAAGTGGCTGGTATAAGTTACCCTGTGGGTAACAAATAAATTCAAACA AAGAATGGAAAACTCCTTTGCGTGGGAGGGTACCAGTTCTGCTGAAGG CAGAACAACAGTTGAGGATTCACCCTCAAGCAGTGACGATTTTGTTAGC AATGTCGGATTCAAAGTAGCTAAGGCGGATCCCACTGTGTGGGAGGAA GCCAATATGTCAGAAGACAATACAATAATTGAGGACCCACCATCGAGTA	TCTCGCGA CGCGTTCT TCTGCCTG ATGAAGTC ACGTCAGG TAATAGAC TTAGAAGG TTGATGAG CGTTCCTC TCCTGGAC CGGGGGT GAGGATGT GTTCTACT GAATCATC GTTCCGGT TGTGAGGT CCGCTGCA AATAGGCC TTGGGGTG	TATTCCA GCCTTCC CGTAGT ACCGTG AGGTTTT CCCGCCT CGAACG GAACAA TGTTGC AGGGTA ATTAAG TACATC GGGCTA TATMGC GATATTT AACGTTT TA (SEQ ID NO: 1339)	TAAACCT TGCCTCC CCGGGC CCCCCTC AGTGAC TAAGAC AACTTTC ACCGTA ATAATC ATATATT TGTATAC CATGTAT TAATCTA GGTAAC AATTGA AAGTAA CATTGA ACCGTA TTCATAC	MENSFAWEGTSSAE GRTTVEDSPSSSDDF VSNVGFKVAKADPTV WEEANMSEDNTIIED PPSSSDDFANNVGFK VTRADPTAWEEASTS TETEDLPSSSNFIDNV ETQIDMAGPTAWED ADSNEDNIDEGTPN NINNNLAIVRGRADA YACSCCERNFISLKAI GTHLKETHNKKVVFE CAKCQHTFVKAHGL ACHVPKCKEDADTP MLNRLLHGCGECGL AFNTRRGLSQHERHR HPSACLSTRRRSRLD GIARKKSLRNRRDIW TNDEIRLLKQLMIQYE

		GCGACGACTTTGCTAACAATGTCGGCTTCAAAGTAACTAGGGCAGATCC	GTCTACCT	TTTAATG	HAKKINIKIAEHFNHK
		CACTGCGTGGGAGGAAGCCAGTACATCTACAGAGACTGAGGATTTACC	CCGGGTCG	AGTACA	NAKQVMIHKRRSLRE
		CTCGAGCAGTAATTTTATTGACAATGTCGAGACCCAAATAGATATGGCG	CTACTCCT	ATAGTG	KDMALGAPHDAPPP
		GGCCCCACTGCGTGGGAGGACGCCGATAGCAATGAAGATAACATAGAT	TTTGAGCT	AAGGGA	LAEEPIEVVEGAREEL
		GAGGGCACACCCAATAATATTAATAATTTAGCAATAGTCAGAGGGA	AGTCGATC	ACTGTAT	EQAPVDVILPDLEALT
		GGGCGGATGCCTATGCATGCTCGTGTTGCGAACGGAACTTCATTTCTCT	CAGGTGA	ATTACAT	VNDGRGGGSPVLTE
		CAAGGCAATCGGTACTCACCTCAAGGAAACCCATAACAAAAAGGTTGTC	GAGTCGG	ACCTCG	GGESTRDMEENGGT
		TTTGAATGTGCCAAATGTCAACACACACCTTCGTTAAAGCACACGGACTAG	GGAAGCCC	AGATAG	DTRSPSPREERAGST
		CGTGCCACGTCCCCAAATGCAAAGAGGACGCCGACACACCTATGCTTAA	ACTTAGGT	AGGTTTT	PWERGWQPRVDRG
		TAGACTGCTCCACGGCTGTGGAGAGTGCGGTCTTGCATTCAATACGAGA	GGGCCAG	TGTACCT	RGEYKGYGGERGDR
		AGAGGGCTGTCTCAACACGAGCGCCATAGGCACCCGTCGGCGTGTCTA	CTAAGCAG	TAAGGG	HTVSLSQRGESGVDP
		AGTACTAGGAGAAGGTCACGGCTAGATGGTATAGCTAGGAAAAAATCA	ATCACCCC	TTCGTG	LTPGGVVEDYDDSYL
		TTGAGGAATAGGAGGGATATTTGGACGAATGATGAGATCCGTCTACTC	CCCAGCAC	AGAATC	EDYFPGWDEDEHM
		AAACAACTCATGATACAGTACGAGCATGCTAAAAAGATAAACATTAAGA	GACAGTGC	CATACT	HIIGRLDLSDESEQGE
		TAGCGGAACATTTCAACCATAAGAATGCAAAACAAGTCATGCATAAACG	GCTGTAAT	GCATAA	AAVSPRLGSFGDLLEE
		TAGATCCTTAAGGGAAAAGGACATGGCCTTAGGTGCTCCACATGATGCA	ATAGCCTG	AGGGGT	VNAMEGKDNLSEAL
		CCTCCGCCATTGGCAGAAGAACCGATAATAGAAGTAGTTGAGGGTGCG	TTGAGAGT	TAACTTG	AETLGLVLHEGHRVE
		AGAGAAGAGTTAGAACAGGCTCCTGTAGATGTAATCCTTCCAGATCTTG	GCACCCAT	TAACTCA	YIKEKMNINVKQMA
		AAGCACTCACGGTTAATGATGGGAGAGGGGGTGGCAGCCCCGTGTTGA	TTATACTG	ACCCCA	TEILAHGANKGNPKR
		CCGAGGGAGGGGAGAGTACAAGGGATATGGAGGAGAACGGGGGGGAC	TTATAAAA	GGGAGA	RKEAVAAKRNPGTRL
		CGACACACGGTCTCCCTCTCCCAGAGAGGAGAGAGCGGGGTCGACCCC	GTATAAAT	CAAGCC	DRAQRDNQAKAKAK
		CTGACTCCGGGGGGGAGTGGTAGAAGATTATGATGACTCGTACTTAGAA	GGTTCTTA	GACAAA	EKKRIFSETQTQYKKN
		GACTACTTCCCTGGGTGGGATGAGGATGAGCACATGCATATTATTGGGA	GCTAACCT	TTCGGC	PHRLVEKLLDGKGDE
		GGCTCGACCTAAGTGACGAGTCGGAACAAGGGGAGGCTGCTGTCAGTC	ATCCAATG	ATTACTA	RCSVSLEVIQRTYMN
		CAAGACTAGGCTCCTTCGGAGACCTATTAGAAGAAGTTAATGCAATGGA	ATTTTTATT	TGTGAG	RFSRESKEVDIGAYVD
		AGGCAAGGATAACCTGTCTGAGGCACTGGCCGAAACCCTGGGCCTTGT	GTTATGTA	GGTCAT	PETVEDNQGIVDPISK
		GTTACACGAAGGTCATAGAGTAGAATACATTAAAGAGAAAATGAACAT	TAATACTG	AAGTGT	AEIERAISTTKKGSAP
		AAATGTTAAACAAATGGCAACAGAAATATTAGCCCACGGGGCAAATAA	ACCACTTT	TAAAGG	GPDGVTYDALKAYG
		GGGAAACCCCAAACGGAGAAAGGAGGCGGTCGCCGCCAAGCGGAATC	AGTCTAGT	ACCCATT	NCQLYLLIMYNTWLA
		CCGGTACGAGATTAGATCGAGCCCAGAGGGACAACCAAGCGAAGGCCA	GTCATGTA	GTATAT	MGKVPSEAKTYRSILI
		AGGCCAAGGAGAAAAAGAGAATCTTCTCCGAAACACAGACCCAATATA	AGGATCCG	AACTTGT	PKGQGDPMDINNFR
		AGAAGAATCCCCACAGACTCGTGGAAAAACTTCTTGATGGGAAGGGTG	AGACAATA	AAATGT	PLTLANVISRLYSKILT
		ACGAGCGGTGCTCTGTCAGCCTTGAAGTGATCCAAAGGACCTACATGAA	GCCTATGT	ACAGAG	RRLDGAVSVCPRQR
		CAGGTTCTCAAGAGAAAGCAAAGAGGTTGACATAGGAGCCTATGTTGA	CTCTAGTC	AATTCG	GFTHKASIEDNTLILR
		CCCCGAGACTGTGGAGGATAACCAAGGGATAGTGGACCCCATCAGCAA	TAGCAATA	AGTTGA	ELIMKSKRNKECLAV
		GGCTGAAATAGAAAGGGCCATCTCTACCACGAAAAAGGGATCCGCACC	ACACAAGA	GTTCGA	VLLDLAKAFDTVSHD
		TGGACCGGACGGGGTCACCTATGATGCCCTAAAGGCGTATGGAAACTG	GGGATAAT	AGAATA	LIIKALRRHRVHEHLIS
		TCAATTATACTTATTAATAATGTACAACACATGGCTGGCCATGGGGAAG	CCACCTCC	AACAAA	VIMDLYEGGTTSFTT

		GTCCCATCAGAGGCAAAAACCTACAGGTCAATTCTAATCCCAAAAGGTC	CTACTCCA	CAAAAA	DEGTTCPIAIRSEVKQ
		AAGGGGACCCCATGGACATCAACAATTTTCGTCCATTAACACTCGCCAAT	GCCACCAC	AGAAAA	GDSLSPVLFNLALDPL
		GTAATCAGCAGATTATATTCCAAAATCTTGACCAGGAGGCTGGACGGG	CTTTACCTT	CAAAGA	LATLEQRGKGVEIGG
		GCGGTAAGTGTCTGTCCGAGGCAAAGAGGCTTCACTCACAAAGCTAGT	TCTTATCTT	GAATTC	HTFVSLAYADDTALV
		ATTGAAGACAATACACTCATCCTACGGGAACTCATCATGAAGAGTAAGA	CCATATAG	GTTCGT	SSSHLDMTANLDITV
		GAAATAAAGAATGCCTTGCAGTGGTACTTCTTGACTTAGCCAAGGCTTT	ATGCCAAC	CCATAG	EYLNATGLSLNVRKC
		CGACACCGTTTCCCACGATCTCATCATCAAGGCCCTGAGGAGGCATCGG	CAGACGG	AAGCGG	QGFLLTPINKSFLVNE
		GTGCATGAACATCTCATTAGCGTCATAATGGACCTGTACGAAGGGGGA	GTCGATGG	AAAAAG	AESWVVEREAIPWV
		ACTACATCTTTCACGACAGATGAGGGCACCACCTGCCCCATCGCGATCC	GCAGTGA	CGCAAA	EPGDTAKYLGVQVG
		GGAGCGAAGTTAAGCAAGGAGACTCTTTGTCGCCGGTGTTATTCAACTT	GAAACAA	AAGGCG	PWSRPWPSIQPVIKR
		GGCTCTCGACCCGCTACTAGCTACACTCGAGCAAAGAGGGAAAGGTGT	GAATGGG	TCCTAAC	LTAYCESIDKAALKPR
		TGAGATCGGGGGACACACGTTCGTCAGCCTGGCCTATGCGGATGACAC	GATAATTT	CTCACG	QRIHILTTYIAPRIAFEI
		GGCACTCGTCTCGTCCAGCCACCTGGACATGACCGCAAACCTAGACATA	GGAATTTG	GTCTAA	AEGGYSTLVDCRGGI
		ACTGTAGAATACCTCAATGCAACGGGACTGAGTCTCAATGTGAGGAAAT	ACATTCTT	CCCAGA	QYTRIREVDMTIRNY
		GTCAAGGCTTCTTACTCACCCCCATCAATAAGTCCTTCTTAGTCAATGAA	TCCTAAAG	AACATTT	VRKWLFLPACLSNSF
		GCCGAGAGCTGGGTAGTGGAACGGGAAGCTATCCCGTGGGTTGAACCT	CAGACCTG	AAGGGG	LYTRRGEGGLGLVSF
		GGGGACACAGCTAAATATCTGGGGGTACAAGTGGGTCCGTGGTCACGT	AAGGGTTA	GGAAGG	YDYVPTERMRKLVRV
		CCGTGGCCATCCATTCAACCAGTAATTAAACGCCTCACAGCCTACTGTGA	AGAGTCAC	AGCTCA	CDSEDPVIAGAAASL
		GAGCATCGACAAAGCGGCGTTAAAGCCAAGACAGAGAATACACATTCT	AAGCGGA	TTACTCC	GLRERAAKISAQTGL
		AACGACCTATATTGCCCCTAGGATTGCCTTTGAGATTGCAGAGGGGGGA	GTCCGGAC	AATCCA	PVPVKPKGAHNAWR
		TACAGTACACTCGTATAAGGGAGGTAGACATGACAATCAGAAACTACGT	TGTCTTTA	AAACCA	KVQKKKWKAQPTQG
		CAGGAAATGGCTATTCTTGCCCGCCTGCCTTTCAAACTCCTTCCTTTACAC	TAAACAGA	CGTTCCC	KGVSCYQHRLGNKW
		AAGGAGAGGAGAGGGTGGTCTTGGGCTGGTAAGCTTTTACGACTACGT	CAATATTT	TAGCCC	LGAPSFLTENDYIWAI
		ACCGACTGAACGTATGAGGAAGCTAGTTAGGGTGTGTGATTCTGAAGA	TCTTATTTT	AAAAGC	KLRTNLVPTREAMGR
		TCCGGTTATAGCAGGGGCAGCAGCCTCGCTAGGGTTAAGGGAGAGAGC	ACCACCTA	GCGCTT	GIIGRNQVECRHCHT
		TGCGAAGATAAGCGCTCAGACAGGGTTGCCAGTTCCAGTTAAACCCAAA	TATGGGGA	CAATGA	TIETMGHISGYCQMV
		GGGGCACACAATGCATGGAGAAAAGTCCAGAAGAAAAAGTGGAAAGC	TATTTCCC	GGGAGC	LDIRLIRHNRICKALIK
		TCAACCCACCCAGGGGAAGGGAGTTAGTTGTTACCAACACAGGTTGGG	AACTCGTA	CGATATT	AATATGLRVTEEPRIV
		CAACAAATGGTTAGGTGCCCCCAGCTTCCTCACTGAAAATGACTACATTT	ATATAGGG	GGAGGA	GTDGKNYLPDLIFSA
		GGGCCATTAAACTGAGGACAAACCTCGTCCCTACTAGGGAAGCTATGG	CCCACTGT	CAGAGT	GAGEPCYVVDPTVV
		GGAGAGGTATCATAGGGAGGAATCAAGTGGAGTGCAGACACTGCCATA	CAAGTGGC	TCATCAA	WDDDPKNLREAWR
		CAACAATAGAAACAATGGGACACATTTCTGGATACTGCCAGATGGTCCT	TGGTATAA	AAGCCA	GKVRKYTPIIPAVEA
		GGATATTCGTTTGATTAGACACAACCGTATATGCAAAGCCCTGATCAAG	GTTACCCT	CCTGTA	MLHPSSVQIFGFVCG
		GCTGCTACGGCTACAGGACTCAGAGTTACGGAAGAACCACGAATAGTG	GTGGGTAA	AGTAGG	ARGTWCPMNDDIAK
		GGTACGGATGGGAAAAAACTATTTGCCAGATCTTATTTTCAGTGCGGGAG	CAAATAAA	CCCATTT	IVGLKNSGISRTLQIVL
		CTGGGGAGCCATGTTATGTAGTTGATCCAACAGTGGTGTGGGATGATG	TTCAAACA	CTGCCA	CDTIRMVKAFMAR
		ATCCCAAGAACCTGAGGGAGGCCTGGAGGGGCAAGGTGAGGAAATAC	AAGA (SEQ	AGGACA	(SEQ ID NO: 1462)
		ACACCAATCATTCCGGCGGTTGAAGCGATGCTTCACCCATCATCAGTTCA		TGCGCA	

		AATATTTGGATTTGTATGTGGCGCAAGGGGAACATGGTGCCCAATGAAT GATGACATAGCCAAGATTGTCGGTTTAAAGAACTCAGGAATTAGCAGG ACACTACAAATAGTCCTTTGCGACACCATAAGAATGGTCAAGGCCTTCAT GGCCAGATAAACCTTGCCTCCCCGGGCCCCCCTCAGTGACTAAGACAAC TTTCACCGTAATAATCATATATTTGTATACCATGTATTAATCTAGGTAACA ATTGAAAGTAACATTGAACCGTATTCATACTTTAATGAGTACAATAGTGA AGGGAACTGTATATTACATACCTCGAGATAGAGGTTTTTGTACCTTAAG GGTTCGTGAGAATCCATACTGCATAAAGGGGTTAACTTGTAACTCAACC CCAGGGAGACAAGCCGACAAATTCGGCATTACTATGTGAGGGTCATAA GTGTTAAAGGACCCATTGTATATAACTTGTAAATGTACAGAGAATTCGA GTTGAGTTCGAAGAATAAACACAAACAAAAAAGAAAACAAAGAGAATTCG TTCGTCCATAGAAGCGGAAAAAGCGCAAAAAGGCGTCCTAACCTCACG GTCTAACCCAGAAACATTTAAGGGGGGAAGGAGCTCATTACTCCAATCC AAAACCACGTTCCCTAGCCCAAAAGCGCGCTTCAATGAGGGAGCCGATA TTGGAGGACAGAGTTCATCAAAAGCCACCTGTAAGTAGGCCCATTTCTG CCAAGGACATGCGCAAAAGGAAGCAGATTATCAAACCAGTCAAACAAG CACAAACATTGGGGGGGATGGGATAACCCCGGAAAGAGAGGGATCTTTA GATAGTGGATGGAAGGGGTGGGACGTTGAATCCAACCATGCCGTTTTT ATGTTCCCGATAAAGAAGGATAAGGTCACTCCAGCCTGACACACAAAGT GGGGTAAAAGAACTCCGCTCGTACGGACTCCAAATAGAA (SEQ ID NO: 1094)	ID NO: 1217)	AAAGGA AGCAGA TTATCAA ACCAGT CAAACA AGCACA AACATT GGGGGG ATGGGA TAACCCC GGAAAG AGAGGG ATCTTTA GATAGT GGATGG AAGGGG TGGGAC GTTGAA TCCAACC ATGCCG TTTTTAT GTTCCC GATAAA GAAGGA TAAGGT CACTCCA GCCTGA CACACA AAGTGG GGTAAA AGAACT CCGCTC GTACGG ACTCCA AATAGA A (SEQ ID NO: 1340)

R2	R2-1_SSa	AGKD010724 55	Salmo salar	AATCTTTAAACCCCGGACTCTTGGGGTTCTTACGACTCTGTATGAGGAACA	AATCTTTA	TAGATC	MSGKRIVEMSGCDE
				GTCGAAGAGAGGGCGCTACCAATCCAAGTATATGTCCCAAGAGGGCTG	ACCCCGGA	CGTTTGT	KICQNKHCLKRRWA
				GGACAGGGTGGAAGAGTGCACCTCGCGATCTGGGGCAGGGAAGGAAT	CTCTTGGG	TATGATT	WISGPKGETSPPRKR
				GGAGAGAAGTCGAAGAAGGCTTGTAGAGAAGGGGCTCTCCTAGATCCT	GTTCTTAC	GGAGGG	GTCENVSFQDKSHAS
				AACCTGTATGACGCCCGTAAAACGGTGACCCCAGTAGCGAATAAAGGA	GACTCTGT	AGCCTG	DPDPLKAPEAREDAG
				GGCAGGTGACAATAGAGGGCAGGGCCGACTTCCCAGGTTTACATTGTT	ATGAGGA	CCGAGT	SVAPQWVGEIKTPSL
				GTACTTGTCAACATAAAGAGGGTGTCTCAATAGTTTGAATCAACAAGGGA	ACAGTCGA	GGTATG	TSRDGVSEVVLPPQP
				GAGGAATACCGACCTGCTCCCTTGGGGCGGGGGTACTGGTCTTAGCCC	AGAGAGG	AGCGCT	VHAEGVSPASDSKDK
				GGTTCCCCGCAAGTTTCCTTTGCCTGGGATGTGCCTGACTGGCTCCATCC	GCGCTACC	CCAACT	ATKITLLISLPVCDLRC
				CCTTTCCCCATTAGGCACGGCTAGATGACGCACCGATGGGCGGGTGTGT	AATCCAAG	ATTGAA	GRCERPLETVGKAVR
				AGGTCGCTACCGAAGGGGACTGGGGGTGTCCGGTGAACCAGGACTTCC	TATATGTC	CCCATAT	HFAVAHPTVSVVFKC
				CAAAATGGTCTCACATTTTTAAGCGGCTTGAGTATCGCCCAGTATCCTCG	CCAAGAG	GATTCCC	QKCEKSSKNSHSISCH
				CGCGGCACTGGGAACCCAGTCAACCGCTCTGTGCCCCGGCGCGACGGCGG	GGCTGGG	GAGGCC	IPKCKGMTETRTDVE
				GGGTTTAATGTCTCCCCGGCTTCACCGGCGCTTCGGCGACGACGCAGAG	ACAGGGT	TGGCCA	GDHGCDHCQEKFTT
				GAGCACCCGGAGGCCCCCATGAACTTAAACCAACCTATCTTGAAATATG	GGAAGAG	GACGCC	AMGLTQHKRHRHIV
				GCCTCTCGTTCGGGTGAAGGGCAGGTGGGAAGAGAGGGCTGCCTCACG	TGCACCTC	TAGATG	QYCKEKEGEMTARR
				ATAAACACCTAGTCAATAGCCAGTCGGGAAAAATGTGGAATGTTAGGA	GCGATCTG	CCTGCC	KGEVEAVKWSEWEE
				CAGGGAGGTAAGGGAGGCGGCTTTTTCGTAAGGCTCCTTCAACCCCTAC	GGGCAGG	ACAATT	SEVARLSDGLAGLKM
				CTGTAGTCCACCTATTGCAGGTGTTGACAACATGCAAGATGACCTGCCT	GAAGGAA	GAACGC	INRRIADSLGTGKTAE
				CGTTACGGGTCGCGTATCATTGCTACAGGTCGTGTGCCGCTTCTAAGAG	TGGAGAG	AGCCCT	QVRQKRRRMRPEKV
				GATAGTAAGGAGAGGTTATAGGGAGGTCCTGTTAGGGCTTCCTCAACC	AAGTCGAA	AGCTTG	RCDKPKEAKDKSNLIK
				CCTCTCTATGCGATTCCTTACAGGAGTGGATCGAGAAGTCCCGGACGTA	GAAGGCTT	CTAGGA	MLSIPSATPTPQTGLK
				ATACACCCTGGAGGTAAGGGAGTGGCCTTCTAGTAGGGCTGCTTCAACC	GTAGAGA	GATCCA	GFLLGELNGVATKGE
				CCTCTGATGGGAGTGTACCGGGAACCTCGACTTGTAAGCACAGGTTAGT	AGGGGCTC	TAGGAA	VQIGGVTLSLRGVEQ
				ATGGGAGCAGAAGGGGGAGCCGTAATGGGCTTCTCTTTCACCTGCTTAC	TCCTAGAT	CTGGCC	DSALLNTSALELQRLL
				ATAATACCTGTGGTGCATGTATCTAGGTCTTGGCGGGAGAGTACTGAGA	CCTAACCT	TATGGG	GGRAGSANPLSLQRE
				GACAAGGTTGAGACCCCAAGATTGGGTCTCCCTAGCCTCTATAGCTGCG	GTATGACG	GCGTCA	RETTLPSERRKTKQGE
				ACTCTTAGCGGGGATATGGAGTAACATGTACCAAGGGAGTGAATAATA	CCCGTAAA	TGACGG	YRRVQKMFRSNEKKI
				AAGGAATTGACGGGGTACAAGTGACTGTTGGCCCGAATCCTAGCCACTT	ACGGTGAC	TTGAAG	AKYILDGNGDGEAAS
				GATGACCAGGGATATATTACAGACTGAGAGCGAATCTAGAGACGTGAG	CCCAGTAG	TTCCTCC	PPLEIALAFKSRWEEV
				AATAAAGTGAGAATTGGTTGAGCGAATACAGAGGAAGATGAGCGGAA	CGAATAAA	ATAGCG	ETFHGLGQFYSRGEA
				AGAGAATTGTTGAGATGAGCGGTTGCGATGAGAAAATCTGCCAAAATA	GGAGGCA	TGCTTG	DGVVFRSLISMSEVC
				AGCATTGCTTGAAGCGCAGATGGGCGTGGATTTCCGGCCCAAAGGGGG	GGTGACAA	GGAGGG	ENLGAIKNNTAAGPD
				AGACATCTCCTCCTCGCAAGAGAGGAACTTGCGAAAACGTATCCTTCCA	TAGAGGG	GACGAC	GITKPALLEWDPTGA
				GGATAAATCTCATGCCTCGGACCCAGATCCTCTCAAAGCCCCGGAAGCG	CAGGGCC	AATGAC	KLAAIFSIWLTSGTLP
				AGAGAGGACGCGGGTTCGGTGGCGCCTCAGTGGGTTGGCGAAATAAA	GACTTCCC	GAGTCA	GPFKKCRTTLIPKTDD
				GACACCTAGCCTAACATCGCGGGATGGGGTTAGTGAAGTCGTGCTGCC	AGGTTTAC	TGACGT	PILLTQVAGWRPLTIG
				ACCACAGCCAGTTCATGCAGAGGGAGTATCCCCAGCCAGCGACTCAAA	ATTGTTGT	ACCGAG	SVVLRLYSRILTHRLER

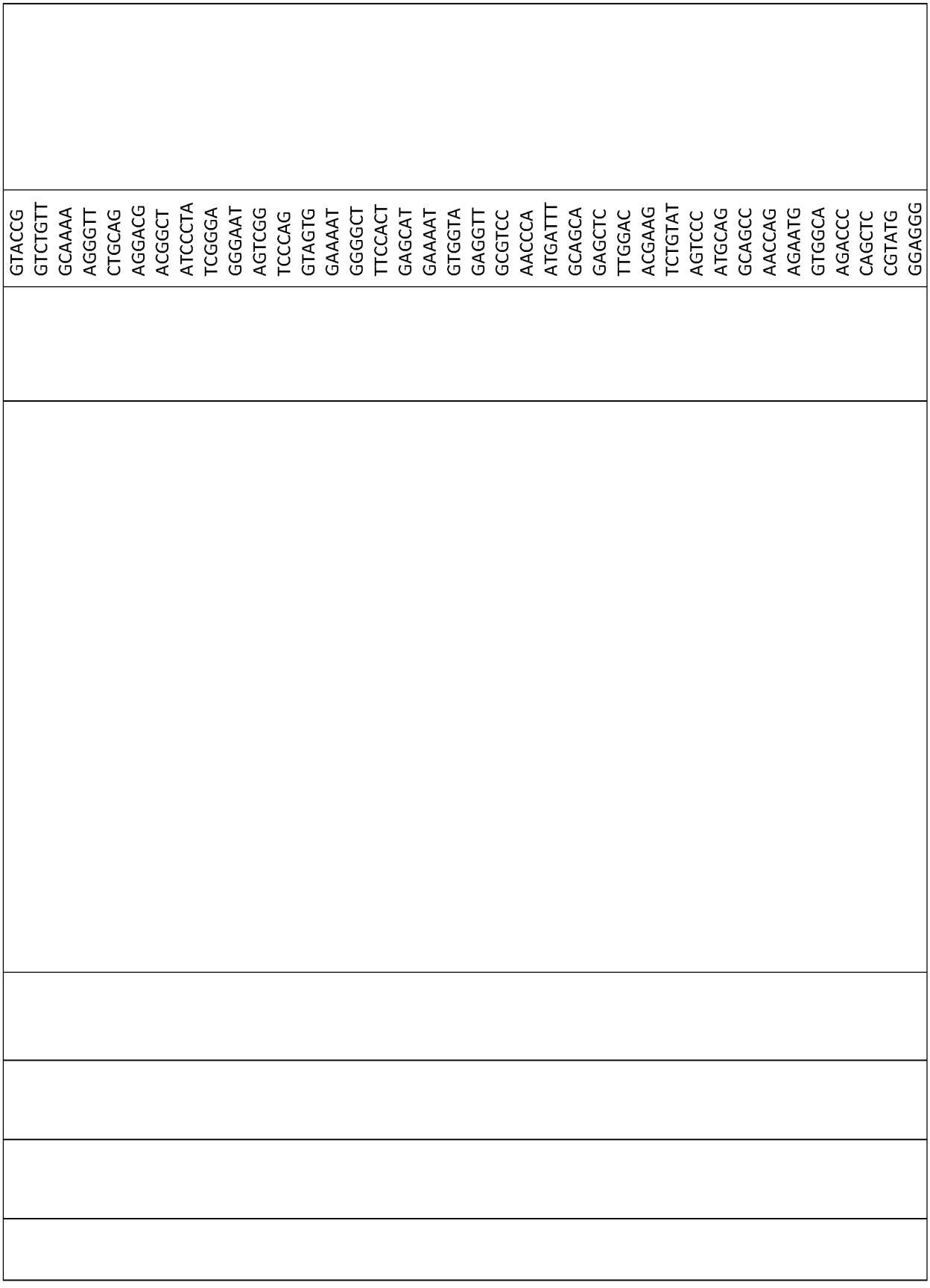
		GGACAAGGCTACGAAGATCACCCTGCTGATATCACTCCCCGTCTGTGAT	ACTTGTCA	AGAACC	ACPINPRQRGFISSPG
		TTGAGATGCGGACGATGTGAAAGACCATTAGAGACCGTTGGGAAGGCG	ACATAAAG	CCAACC	CSENLMILGGLIKRS
		GTGAGACACTTTGCTGTGGCTCATCCGACGGTGTCGGTGGTTTTTAAGT	AGGTGTCT	CAGGTT	WAKGERLAVVLVDF
		GTCAAAAGTGTGAGAAAGAGCAGCAAGAATAGCCACTCCATCTCCTGTCA	CAATAGTT	GGGGGA	ARAFDSVSHSHILEIL
		TATCCCTAAATGTAAGGGAATGACAGAGACCCGGACGGATGTGGAAGG	TGAATCAA	GAGAGC	RQRGLDEHIIGIVGDS
		TGATCACGGCTGTGATCATTGCCAGGAGAAGTTTACAACGGCTATGGG	CAAGGGA	CAGCAA	YTDVTTTITVSGEQSP
		GCTGACTCAACACAAGAGACACAGACACATCGTCCAGTATTGTAAGGAG	GAGGAAT	GAGCGG	PIDMRVGVKQGDP
		AAGGAGGGGGAGATGACAGCAAGAAGGAAGGGTGAAGTCGAAGCAG	ACCGACCT	AGATGC	MSPLLFNLALDPMID
		TCAAATGGAGCGAATGGGAAGAAAGTGAGGTGGCAAGGTTGAGCGAT	GCTCCCTT	TTGGTAT	TLERYGLGYRMGEQ
		GGACTGGCTGGGCTAAAAATGATCAACAGGCGAATCGCAGATAGCCTG	GGGGCGG	ACCAAG	QITALAFADDLVLVS
		GGGACTGGAAAAACTGCGGAACAGGTGAGGCAGAAAAGACGTAGAAT	GGGTACTG	CTAGCA	DSWEGMACNIRILEE
		GAGACCTGAGAAGGTACGGTGTGACAAACCTAAGGAGGCAAAAGATA	GTCTTAGC	GAGAGA	FCRLTGLRIQPRKCH
		AGAGCAATCTTATCAAAATGCTGTCCATACCGAGTGCAACACCAACACC	CCGGTTCC	GGGTTG	GFLIQKIQRARSVNLC
		CCAAACTGGGCTCAAAGGATTCCTCCTTGGAGAACTAAATGGGGTTGCT	CCGCAAGT	AAGAGG	KPWIVCGEELHMVG
		ACCAAAGGTGAAGTACAGATTGGGGGAGTTACGTTGTCCCTAAGGGGG	TTCCTTTG	ATGACT	PEESVSYLGMKVSP
		GTAGAGCAAGACTCAGCCTTGCTCAATACGTCGGCCCTTGAACTGCAAC	CCTGGGAT	ACTGGG	WHGIMEPDPVERLC
		GATTGTTGGGAGGGGAGAGCGGGGAGCGCGAATCCGTTAAGCTTACAG	GTGCCTGA	CTCAGA	NWISSIGRSPLKPSQK
		AGAGAACGTGAGACCACATTGCCAAGTGAAAGAAGGAAAACGAAACA	CTGGCTCC	GTCATCT	VRMLNVYAAPRMTY
		AGGAGAGTACCGGAGGGTACAAAAGATGTTTCGAAGCAATGAAAAGA	ATCCCCTT	CACCCTA	QADHGGLGPIVLNVL
		AAATTGCTAAATACATTCTAGATGGCAATGGAGATGGAGAAGCGGCCTC	TCCCCATT	AAAGGC	DGMIRKAVKVWLHL
		CCCGCCACTCGAGATCGCTTTGGCGTTCAAAAGTAGGTGGGAGGAAGT	AGGCACG	GGTGGG	PLCTCDGLLYSRCQD
		GGAAACATTTCATGGGCTTGGCCAGTTCTACTCGAGGGGGGAAGCGGA	GCTAGATG	GCATCG	GGLGIVKLACQIPSIQ
		TGGTGTGGTCTTCCGGTCACTTATCTCAATGAGCGAAGTATGTGAAAAC	ACGCACCG	GTTGAA	ARRVYRLWHSKEAIT
		CTGGGGGCAATTAAAAACAACACAGCAGCTGGGCCAGATGGGATAACA	ATGGGCG	CACCTAC	RVVTRRTVEAEEYRG
		AAACCGGCATTGCTTGAATGGGATCCCACTGGTGCGAAACTGGCCGCTA	GGTGTGTA	CCATACC	MWLRAGGSEAGLPP
		TCTTTTCCATATGGTTGACATCGGGCACCCTGCCTGGGCCTTTTAAGAAG	GGTCGCTA	GGGATG	LEDREEGAVQCTDTA
		TGCAGAACAACCTTAATACCCAAGACCGATGACCCGATTTTACTCACCCA	CCGAAGG	GGAGGT	GSVKPKNPVIPDWR
		GGTGGCTGGGTGGAGGCCTCTCACTATCGGGTCGGTGGTTTTGAGGCT	GGACTGG	GGTAGG	RAEFLKWQNLTAQG
		ATACTCTCGCATCCTGACACACAGGCTGGAGCGGGCGTGTCCCATTAAC	GGGTGTCC	CCGAAA	VGVQVFGGDKNSNH
		CCGCGCCAGAGAGGATTCATTTCCTCACCTGGGTGTTCGGAAAACTTGA	GGTGAACC	AAGAAC	WMANPETLGSKERH
		TGATCTTGGGAGGTCTAATCAAGAGAAGTTGGGCGAAGGGCGAGAGG	AGGACTTC	AGGAAG	YIAGLQLRANVYPTR
		CTGGCGGTAGTGTTGGTTGACTTTGCGCGTGCGTTTGACTCTGTGAGCC	CCAAAATG	ATGGTG	EALSRGRPDLPKVCR
		ACTCGCACATCTTGGAGATACTCAGACAGAGAGGCTTGATGAACATAT	GTCTCACA	GAGTAA	QCLAGTESCAHILGQ
		CATCGGAATCGTAGGTGACTCGTACACCGATGTAACGACCACAATTACA	TTTTTAAG	GTTGAG	CPAVKDSRIRRHHKL
		GTCAGTGGGGAGCAGTCCCCTCCCATTGACATGAGGGTAGGGGTTAAG	CGGCTTGA	AGCGGT	CDLLASEAESAGWTV
		CAGGGAGACCCGATGTCTCCGCTACTGTTTAATCTAGCCCTGGACCCAA	GTATCGCC	TGCTCG	IREMCCRTRAGALRR
		TGATCGACACCCTTGAACGCTACGGCTTGGGGTACAGGATGGGCGAGC	CAGTATCC	GGAAGT	PDLVFVKTGFALVVD
		AGCAGATCACGGCCCTAGCTTTTGCTGACGATCTGGTTCTGGTGAGCGA	TCGCGCGG	TATGTTG	VTVRYEMAYDTLMG

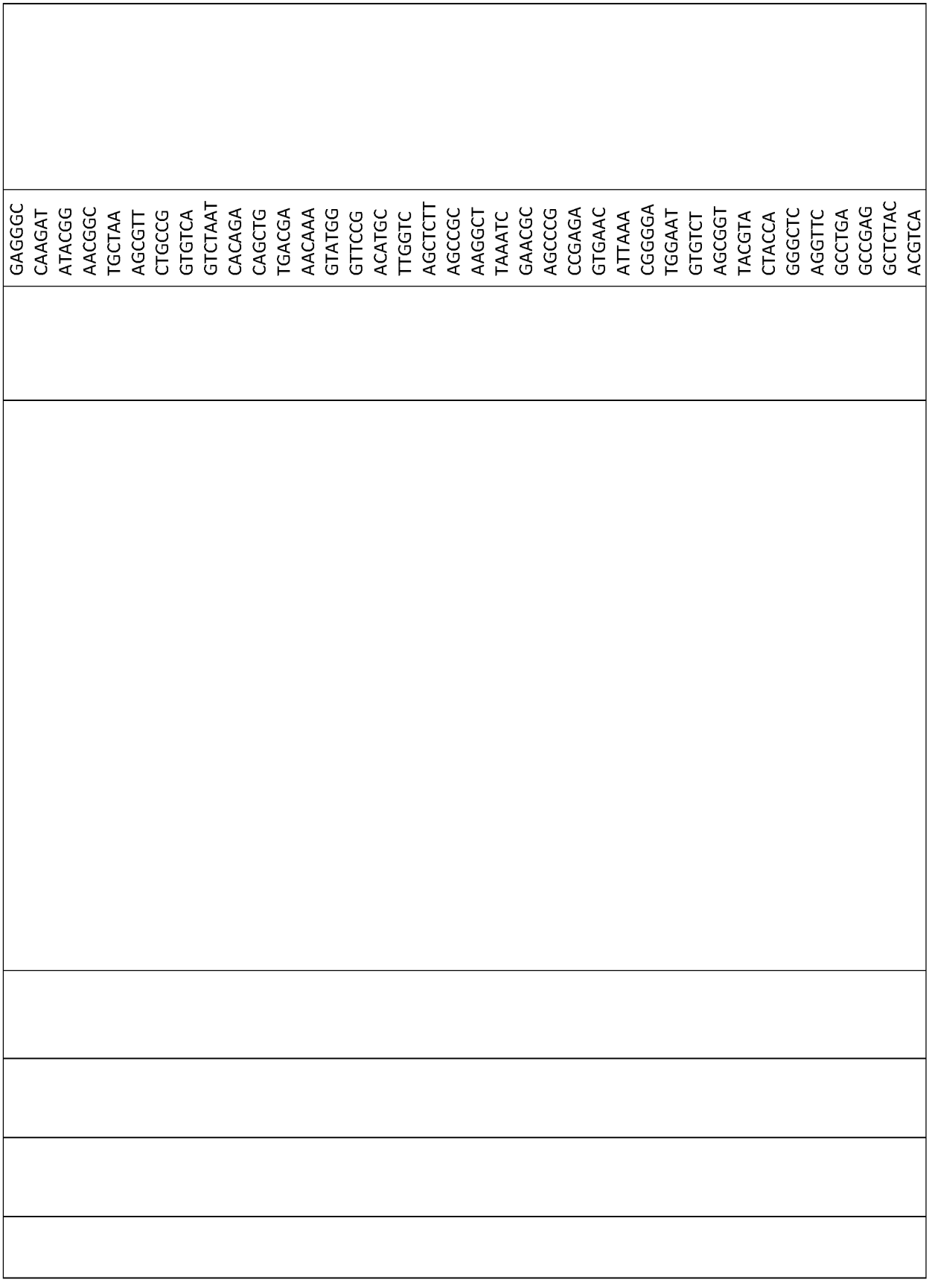
	CTCGTGGGAGGGCATGGCGTGCAATATCCGTATTCTGGAAGAATTTTGT	CACTGGGA	TGATAA	AAAEKVARYTPITPYV
	CGACTGACTGGGCTGAGGATTCAGCCTAGGAAGTGTCATGGTTTCCTCA	ACCCAGTC	CTCCATT	AMTLKARRVKVFGFP
	TACAGAAGATTCAGAGGGCGAGATCGGTAAACCTCTGTAAGCCCTGGA	AACCGCTC	AAGGCC	LGARGKWPGSNDRL
	TAGTGTGTGGTGAAGAACTACATATGGTCGGGCCGGAAGAGTCGGTTT	TGTGCCCC	GGTGGG	LKAMGVGGGRRKQL
	CCTACCTGGGTATGAAGGTGAGCCCATGGCATTATGGAGCCAG	GGCGCAG	CATGGT	AKLFSRRALLYSLDVL
	ATCCTGTCGAACGACTCTGTAACTGGATCAGTTCGATTGGGCGGTCACC	GCGGGGG	GCGGAT	RDFYRAEGETGDLDD
	GCTGAAGCCTTCTCAGAAGGTGAGGATGTTGAATGTTTATGCTGCCCCG	TTTAATGT	AATGGA	ESVDDHL (SEQ ID
	AGGATGACTTATCAGGCGGATCATGGCGGGCTGGGGCCAATTGTCCTG	CTCCCCGG	AACTAT	NO: 1463)
	AATGTACTCGATGGGATGATCAGGAAAGCAGTGAAGGTGTGGCTACAC	CTTCACCG	AAAAAC	
	CTTCCGCTGTGTACCTGTGATGGGCTACTTTACTCTAGATGCCAGGACG	GCGCTTCG	AATAAA	
	GTGGACTGGGCATAGTAAAAATTGGCTTGTCAAATCCCTTCTATCCAAGCT	GCGACGAC	AAGAAA	
	AGAAGGGTCTACCGCCTGTGGCATTCTAAGGAAGCCATAACGCGGGTA	GCAGAGG	GACCAA	
	GTCACCCGAAGGACGGTCGAAGCGGAAGAGTACCGTGGGATGTGGCT	AGCACCCG	AAAAAT	
	GAGAGCCGGTGGGAGTGAGGCAGGTTTGCCTCCCCTGGAAGATAGGG	GAGGCCCC	GTTCTGT	
	AAGAAGGTGCTGTACAGTGTACAGACACTGCCGGTTCGGTGAAGCCGA	CATGAACT	TATGAT	
	AAAACCCAGTCATTCCCGATTGGAGGCGAGCTGAGTTCCTCAAGTGGCA	TAAACCAA	GCCTTAC	
	AAACCTGACAGCGCAAGGGGTTGGAGTGCAGGTCTTTGGCGGTGACAA	CCTATCTT	ACATGT	
	AAACAGCAATCACTGGATGGCGAATCCGGAAACGTTGGGATCGAAAGA	GAAATATG	CTGGGA	
	GCGCCACTATATTGCAGGTCTACAGTTGAGGGCCAATGTATATCCAACT	GCCTCTCG	GACCCC	
	CGCGAGGCACTGTCCAGGGGCAGGCCGGACTTACCTAAAGTCTGCCGG	TTCGGGTG	ATAAGG	
	CAATGCCTAGCAGGAACTGAGTCTTGCGCGCATATTCTCGGGCAGTGTC	AAGGGCA	GTCTCCC	
	CTGCAGTGAAGGATAGCCGCATCAGGCGGCATCATAAACTGTGTGACCT	GGTGGGA	CTTATAC	
	GCTAGCAAGTGAAGCCGAAAGCGCCGGATGGACCGTCATCAGAGAAAT	AGAGAGG	TTCACTG	
	GTGTTGCAGAACTCGTGCCGGAGCTTTGCGGCGTCCAGACCTGGTGTTT	GCTGCCTC	GGAAAC	
	GTGAAAACCGGTTTTGCTTTGGTGGTGATGTTACTGTGCGGTACGAGA	ACGATAAA	CCCATA	
	TGGCCTATGATACGCTCATGGGTGCGGCTGCCGAGAAAGTTGCTCGGTA	CACCTAGT	AGGGTA	
	CACCCCAATTACTCCATATGTTGCGATGACCCTGAAGGCAAGGAGAGTC	CAATAGCC	TCCCCCT	
	AAGGTGTTTGGCTTTCCACTGGGAGCCCGAGGCAAATGGCCGGGAAGC	AGTCGGG	ATATTTA	
	AACGACCGGCTGCTGAAGGCTATGGGTGTTGGTGGCGGCAGGAGGAA	AAAAATGT	CTGGGA	
	ACAGCTGGCCAAGTTGTTTAGCAGGAGGGCGCTCCTGTACTCCTTAGAT	GGAATGTT	GACCCC	
	GTCCTTAGGGACTTCTACCGGGCGGAGGGAGAAACGGGGGACTTGGAT	AGGACAG	ATAAGG	
	GATGAGAGCGTGGATGATCATCTATAGATCCGTTTGTTATGATTGGAGG	GGAGGTA	GTCTCCC	
	GAGCCTGCCGAGTGGTATGAGCGCTCCAACTATTGAACCCATATGATTC	AGGGAGG	CCTATA	
	CCGAGGCCTGGCCAGACGCCTAGATGCCTGCCACAATTGAACGCAGCCC	CGGCTTTT	GATGTA	
	TAGCTTGCTAGGAGATCCATAGGAACTGGCCTATGGGGCGTCATGACG	TCGTAAGG	GAGCGT	
	GTTGAAGTTCCTCCATAGCGTGCTTGGGAGGGGACGACAATGACGAGT	CTCCTTCA	AAGGGG	
	CATGACGTACCGAGAGAACCCCAACCCAGGTTGGGGGAGAGAGGGTTGAA	ACCCCTAC	TCTCCAA	
	AAGAGCGGAGATGCTTGGTATACCAAGCTAGCAGAGAGAGGGTTGAA	CTGTAGTC	AGTACC	

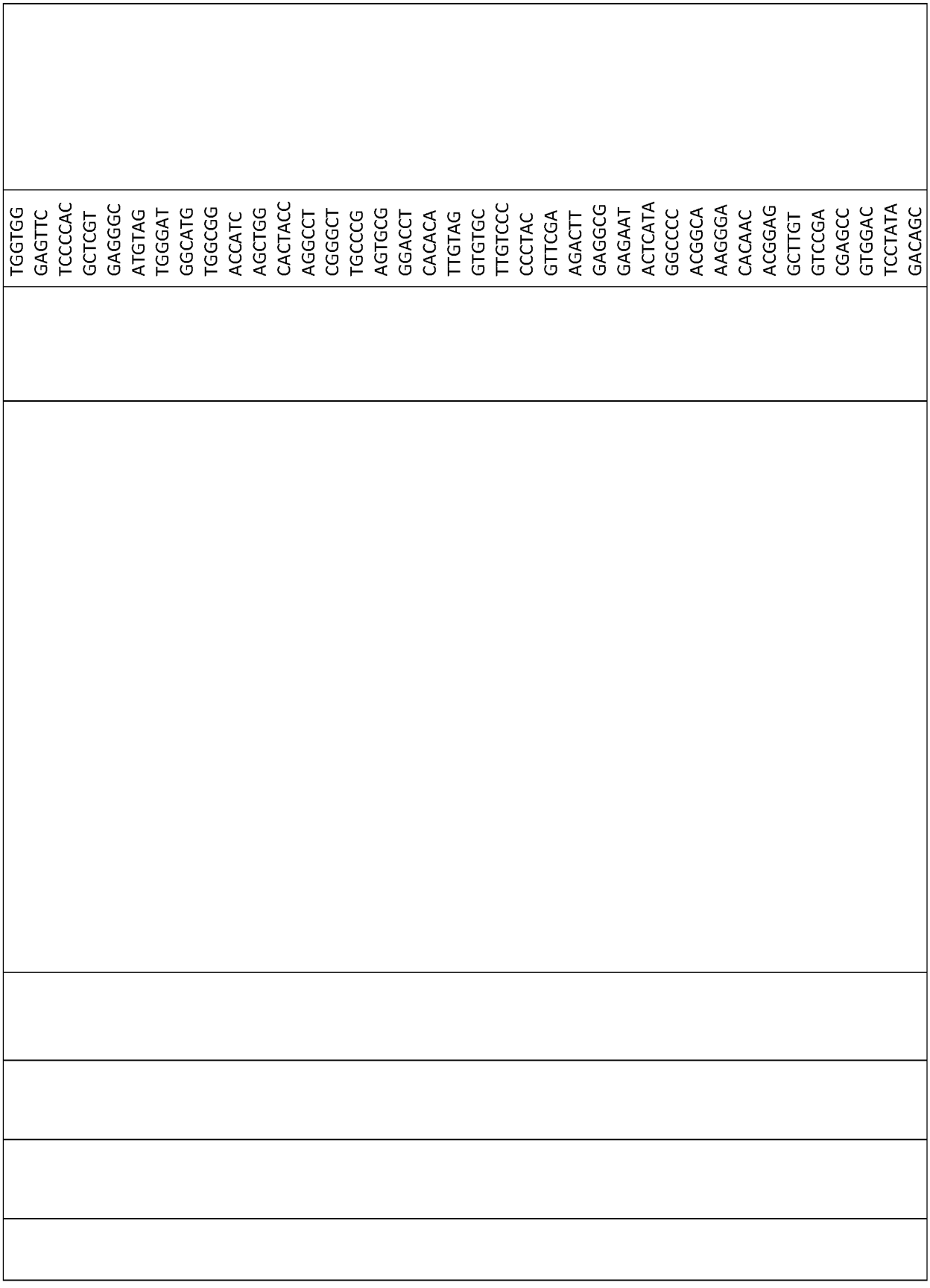
				GAGGATGACTACTGGGCTCAGAGTCATCTCACCCTAAAAGGCGGTGGG GCATCGGTTGAACACCTACCCATACCGGGATGGGAGGTGGTAGGCCGA AAAAGAACAGGAAGATGGTGGAGTAAGTTGAGAGCGGTTGCTCGGGA AGTTATGTTGTGATAACTCCATTAAGGCCGGTGGGCATGGTGCGGATAA TGGAAACTATAAAAACAATAAAAAGAAAGACCAAAAAAATGTTCTGTTA TGATGCCTTACACATGTCTGGGAGACCCCATAAGGGTCTCCCCTTATACT TCACTGGGAAACCCCATAAGGGTATCCCCCTATATTTACTGGGAGACCC CATAAGGGTCTCCCCCTATAGATGTAGAGCGTAAGGGGTCTCCAAAGTA CCGGCCGATATGGCCTTATGGCAAACTCTGGTGGTAGGGACAAGGAGG TAAGGGCAGTGCCAACCCCTACTTGATCGGGACCATCCAGGGAATGCCA TCCTCCCGCGAAGGTGATGTGGTGAGGTAAGGGGGGAGCCCGTCTTCG AGTTTCCCCAACCCCTACCCACAGGTGAGAGGAGGAGAAGAGGAATCT GTCCCCAACGGGAGGAGGGTGAGGTGTAAGGGGGAGACCTTCTAGTA GGGTCTTCTCAGTCGCCTGACGTCCTGACTGTGGGGTGGATCAGTACCC TACAGGTGAGACCGGTGAGGTAAGGGTGTGGCCCTCTTGAGGGCTGCG CCAACCCCTACTCGAGGTAACCTGAGGGAGTGGTGGAATGGCGGCATG TTAGTGCTGGGACTTGATTGCGAGGGTTTAATGAGAGTGGCCTGCTGA GAGCAACACTTGTGGTGCTTAAAGCGGGGCGGCCCATGACCACCGTGA GATAGGACACTGCACAGTGCAGCCATGAGGTTCCTGGAGGATGATGCG ATGAGGTGGGGGCCTCATCAGCCCCTCCTGGCAGGGCGTCGGCCAGGG AAACTAAATGTCTCTAGCATGTCAGTGCAGTGAGGTAAGGGGAGAGCA CTCTAGTAGGGCTCTTCCAACCCCTACCTGTAGGTCACCTGGTCCAGGTG TCGATGATGTGAAAACAAGAGCTACTTTGGTACCGGTCTGTTGCAAAAA GGGTTCTGCAGAGGACGACGGCTATCCCTATCGGGAGGGAATAGTCGG TCCCAGGTAGTGGAAAATGGGGCTTTCCACTGAGCATGAAAATGTGGT AGAGGTTGCGTCCAACCCAATGATTTGCAGCAGAGCTCTTGGACACGAA GTCTGTATAGTCCCATGCAGGCAGCCAACCAGAGAATGGTGGCAAGAC CCCAGCTCCGTATGGGAGGGGAGGGCCAAGATATACGGAACGGCTGCT AAAGCGTTCTGCCGGTGTCAGTCTAATCACAGACAGCTGTGACGAAACA AAGTATGGGTTCCGACATGCTTGGTCAGCTCTTAGCCGCAAGGCTTAAA TCGAACGCAGCCCGCCGAGAGTGAACATTAAACGGGGATGGAATGTGT CTAGCGGTTACGTACTACCAGGGCTCAGGTTCGCCTGAGCCGAGGCTCT ACACGTCATGGTGGGAGTTCTCCCCACGCTCGTGAGGGCATGTAGTGG GATGGCATGTGGCGGACCATCAGCTGGCACTACCAGGCCTCGGGCTTG CCCGAGTGCGGGACCTCACACATTGTAGGTGTGCTTGTCCCCCCTACGTT CGAAGACTTGAGGCGGAGAATACTCATAGGCCCCACGGCAAAGGGACA CAACACGGAGGCTTGTGTCCGACGAGCCGTGGACTCCTATAGACAGCCC GGGATATCACTGGGCACGCTCATACTGAAGAAATTCGATGAACCGGGC	CACCTATT GCAGGTGT TGACAACA TGCAAGAT GACCTGCC TCGTTACG GGTCGCGT ATCATTGC TACAGGTC GTGTGCCG CTTCTAAG AGGATAGT AAGGAGA GGTTATAG GGAGGTCC TGTTAGGG CTTCCTCA ACCCCTCT CTATGCGA TTCCTTAC AGGAGTG GATCGAGA AGTCCCGG ACGTAATA CACCCTGG AGGTAAG GGAGTGG CCTTCTAG TAGGGCTG CTTCAACC CCTCTGAT GGGAGTG TACCGGGA ACCTCGAC TTGTAAGC ACAGGTTA GTATGGGA GCAGAAG	GGCCGA TATGGC CTTATG GCAAAC TCTGGT GGTAGG GACAAG GAGGTA AGGGCA GTGCCA ACCCCTA CTTGATC GGGACC ATCCAG GGAATG CCATCCT CCCGCG AAGGTG ATGTGG TGAGGT AAGGGG GGAGCC CGTCTTC GAGTTT CCCCAA CCCCTAC CCACAG GTGAGA GGAGGA GAAGAG GAATCT GTCCCC AACGGG AGGAGG GTGAGG TGTAAG GGGGAG ACCTTCT	

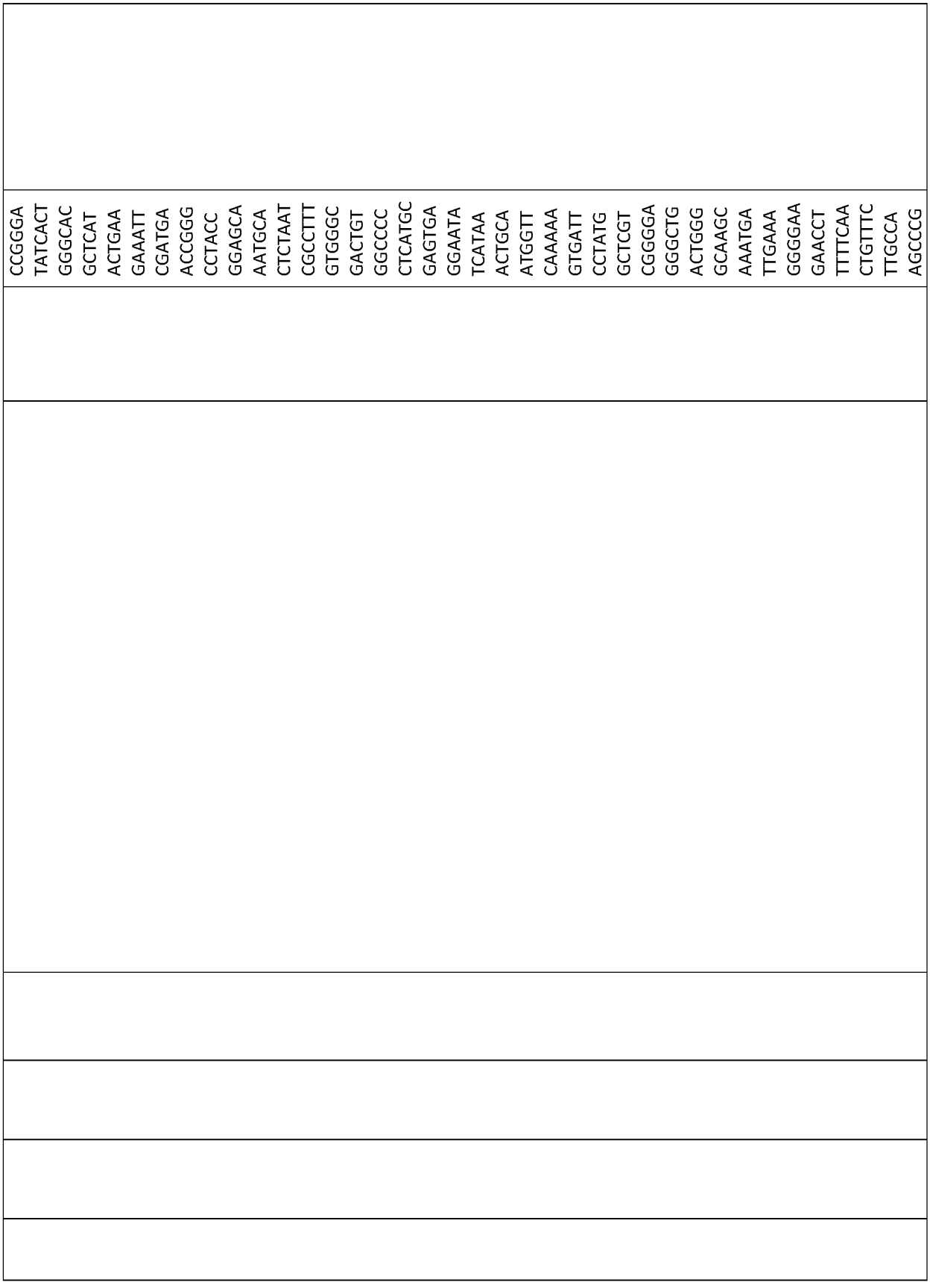
		CTACCGGAGCAAATGCACTCTAATCGCCTTTGTGGGCGACTGTGGCCCC (SEQ ID NO: 1095)	GGGGAGC	AGTAGG		
		CTCATGCGAGTGAGGAATATCATAAACTGCAATGGTTCAAAAAGTGATT (SEQ ID NO: 1096)	CGTAATGG	GTCTTCT		
		CCTATGGCTCGTCGGGAGGGCTGACTGGGGCAAGCAAATGATTGAAA (SEQ ID NO: 1097)	GCTTCTCT	CAGTCG		
		GGGGAAGAACCTTTTTCAACTGTTTCTTGCCAAGCCCGGTTGATGGTGG (SEQ ID NO: 1098)	TTCACCTG	CCTGAC		
		CGCTAGTAATTGCGACGGGGAAAATGCGGTTTAAGTCTCCGAAGTAGTGC (SEQ ID NO: 1099)	CTTACATA	GTCCTG		
		GTAGCACCGGATGTCGACCGGGTGTAAAAGCCCTTCGTAAAGTCCCTGG (SEQ ID NO: 1100)	ATACCTGT	ACTGTG		
		GGAGGTCAGTCCTGGGGCTACTGATGCGCAGTATGTAATTCGCAGAAT (SEQ ID NO: 1101)	GGTGCATG	GGGTGG		
		AGGGCCATCGATACCGCCTGCGTGACTCGACTGGGTTTCCACTTGAGGA (SEQ ID NO: 1102)	TATCTAGG	ATCAGT		
		TATCCGACCGTAGCGTGCACCCTCTTGTAGTTGCGCCGGAAACGGCTGT (SEQ ID NO: 1103)	TCTTGGCG	ACCCTAC		
		GTTCCCTCACGTATGTGAGGAAACTCAACAATGTGAGTGGGTAAACGGC (SEQ ID NO: 1104)	GGAGAGT	AGGTGA		
		GGGACGAACTATGGCTCTCGT (SEQ ID NO: 1095)	ACTGAGAG	GACCGG		
			ACAAGGTT	TGAGGT		
			GAGACCCC	AAGGGT		
			AAGATTGG	GTGGCC		
			GTCTCCCT	CTCTTGA		
			AGCCTCTA	GGGCTG		
			TAGCTGCG	CGCCAA		
			ACTCTTAG	CCCCTAC		
			CGGGGAT	TCGAGG		
			ATGGAGTA	TAACCT		
			ACATGTAC	GAGGGA		
			CAAGGGA	GTGGTG		
			GTGAATAA	GAATGG		
			TAAAGGAA	CGGCAT		
			TTGACGGG	GTTAGT		
			GTACAAGT	GCTGGG		
			GACTGTTG	ACTTGAT		
			GCCCGAAT	TGCGAG		
			CCTAGCCA	GGTTTA		
			CTTGATGA	ATGAGA		
			CCAGGGAT	GTGGCC		
			ATATTACA	TGCTGA		
			GACTGAGA	GAGCAA		
			GCGAATCT	CACTTGT		
			AGAGACGT	GGTGCT		
			GAGAATAA	TAAAGC		
			AGTGAGA	GGGGCG		
			ATTGGTTG	GCCCAT		

					AGCGAATA CAGAGGA AG (SEQ ID NO: 1218)	GACCAC CGTGAG ATAGGA CACTGC ACAGTG CAGCCA TGAGGT TCCTGG AGGATG ATGCGA TGAGGT GGGGGC CTCATCA GCCCCTC CTGGCA GGGCGT CGGCCA GGGAAA CTAAAT GTCTCTA GCATGT CAGTGC AGTGAG GTAAGG GGAGAG CACTCTA GTAGGG CTCTTCC AACCCCT ACCTGT AGGTCA CCTGGT CCAGGT GTCGAT GATGTG AAAACA AGAGCT ACTTTG	


GTACCG
GTCTGTT
GCAAAA
AGGGTT
CTGCAG
AGGACG
ACGGCT
ATCCCTA
TCGGGA
GGGAAT
AGTCGG
TCCCAG
GTAGTG
GAAAAT
GGGGCT
TTCCACT
GAGCAT
GAAAAT
GTGGTA
GAGGTT
GCGTCC
AACCCA
ATGATTT
GCAGCA
GAGCTC
TTGGAC
ACGAAG
TCTGTAT
AGTCCC
ATGCAG
GCAGCC
AACCAG
AGAATG
GTGGCA
AGACCC
CAGCTC
CGTATG
GGAGGG


GAGGGC
CAAGAT
ATACGG
AACGGC
TGCTAA
AGCGTT
CTGCCG
GTGTCA
GTCTAAT
CACAGA
CAGCTG
TGACGA
AACAAA
GTATGG
GTTCCG
ACATGC
TTGGTC
AGCTCTT
AGCCGC
AAGGCT
TAAATC
GAACGC
AGCCCG
CCGAGA
GTGAAC
ATTAAA
CGGGGA
TGGAAT
GTGTCT
AGCGGT
TACGTA
CTACCA
GGGCTC
AGGTTC
GCCTGA
GCCGAG
GCTCTAC
ACGTCA


TGGTGG
GAGTTC
TCCCCAC
GCTCGT
GAGGGC
ATGTAG
TGGGAT
GGCATG
TGGCGG
ACCATC
AGCTGG
CACTACC
AGGCCT
CGGGCT
TGCCCG
AGTGCG
GGACCT
CACACA
TTGTAG
GTGTGC
TTGTCCC
CCCTAC
GTTCGA
AGACTT
GAGGCG
GAGAAT
ACTCATA
GGCCCC
ACGGCA
AAGGGA
CACAAC
ACGGAG
GCTTGT
GTCCGA
CGAGCC
GTGGAC
TCCTATA
GACAGC


CCGGGA
TATCACT
GGGCAC
GCTCAT
ACTGAA
GAAATT
CGATGA
ACCGGG
CCTACC
GGAGCA
AATGCA
CTCTAAT
CGCCTTT
GTGGGC
GACTGT
GGCCCC
CTCATGC
GAGTGA
GGAATA
TCATAA
ACTGCA
ATGGTT
CAAAAA
GTGATT
CCTATG
GCTCGT
CGGGGA
GGGCTG
ACTGGG
GCAAGC
AAATGA
TTGAAA
GGGGAA
GAACCT
TTTTCAA
CTGTTTC
TTGCCA
AGCCCG


GTTGAT
GGTGGC
GCTAGT
AATTGC
GACGGG
AAAATG
CGGTTT
AAGTCT
CCGAAG
TAGTGC
GTAGCA
CCGGAT
GTCGAC
CGGGTG
TAAAAG
CCCTTCG
TAAAGT
CCCTGG
GGAGGT
CAGTCCT
GGGGCT
ACTGAT
GCGCAG
TATGTA
ATTCGC
AGAATA
GGGCCA
TCGATA
CCGCCT
GCGTGA
CTCGACT
GGGTTT
CCACTTG
AGGATA
TCCGAC
CGTAGC
GTGCAC
CCTCTTG

						TAGTTG CGCCGG AAACGG CTGTGTT CCCTCAC GTATGT GAGGAA ACTCAA CAATGT GAGTGG GTAAAC GGCGGG ACGAAC TATGGC TCTCGT (SEQ ID NO: 1341)	
R2	R2-1_TCas	.	Tribolium castaneum	AGTCATAGAGCCAGAACCTCCTCGTGGTCCCGCTGGGCACAGGGATTAA TTTTTCTGTGGCAAATTTGACTGGCTTCAGAGAGCGTTTTTCGAAGTGG WCTGTGTGACTGCGTTCCCCCCTTAGTTGCTATWTCCGCTKMGATTAAC ATCTCACCTCGACGTWTAAGATCATTATGTCGAGGCGCCCCGGGGAAGT CTAATGAGCCCCCGGTTCGTAGTAGGGCTATGGGTTTAACCACTTTATCT GGCACCAAGACATCAAATTCTGGTGCCCAGGGGCCCTCTACCAGTGCCC CCATGCAGAACATGGCTGGGGGTTTTGTTTGTGACTGCGGCCGGAGCTA TGCCTTGAAGACATCTCTGGCCCGACACAAGAAGGAGTGCGGAAAGAA CAATGCGGAGTGTCGATGGTGCGGGACAAGGTTTAACACCTTGGCTGG CACTCGTCAGCATGAGCGAAAGGCTCACTTTGTCCAGTATCAGTCTGAT CTGGCAAAGGCGCTCCCCCAACCTGAGTCGGAGCTTATGGAGAAAATC GCCATTGTGGAGGCCAGGTCCAKTAACGGCATCTTCTATAAGGAGATGA TGGCCTCCACGGGTCTGACCCACCAACAGGTCAGATCAAGGAGAGAGA AACCCGAATATAAGGGCTTCCTGGAGCGAGCACGCAGGAGCCTGGCCC AGACCAATATCAGGGCTGGCTCCATCTCACCAGCGAGTACGAWCGCAG GCTCTTTGGAGTCTGCGTCTCCGAAAGCGGGGGTGTAGTAGCAGTGCCTC GCCCGGACCGACCACCAGGTCCAGGGCACCTACCAAGGGTGTGCCCCW ACGAAGCTCTAATTCGGCGAGAATTGTGGTGGAGGCCCAAGTCCATACC CGGGCGCCACCCAATACCGGGGAAACGGAAGTGGCTCTTCGTGAGTCA CGCCGAACCGTGCCCCGGCTCGGGCMTAATCCATCTAGGCCCTGTGGT	AGTCATAG AGCCAGAA CCTCCTCG TGGTCCCG CTGGGCAC GTTTCAG GAGGATTA ATTTTTCT GTGGCAAA TTTGACTG GCTTCAGA GAGCGTTT GTGACTGC GTTCCACC CTTAGTTG CTATWTCC GCTKMGA TTAACATC TCACCTCG	TGAAAA GAGTGG CAGTTG TGGAGA CTTCCTT CGGACG TCGGGT CGGATT TTCGGA CGCCAG GGTACC TCCACC GCTGGG TTCACAA ACTAGG CGAACA TCTGCC GATACC CTCTTTA GGTCAT	MSRRPGKSNEPPVRS RAMGLTTLSGTKTSN SGAQGPSTSAPMQN MAGGFVCDCGRSYA LKTSLARHKKECGKN NAECRWCGTRFNTL AGTRQHERKAHFVQ YQSDLAKALPQPESE LMEKIAIVEARSXNGI FYKEMMASTGLTHQ QVRSRREKPEYKGFL ERARRSLAQTNIRAG SISPASTXAGSLESAS PKAGCSSSASPGPTT RSRAPTKGVPXRSSN SARIVVEAQVHTRAP PNTGETEVALRESRR TVPRLGXNPSRPCGIS PLMAIAIDEDSVLGG LRVQAGPSPTAVHSV

	ATCTCTCCTTTGATGGCGATTGCCATAGATGAAGACTCTGTGTTGGGCG	ACGTWTA AGATCATT (SEQ ID NO: 1219)	AGGACC ACATGT CTCTGCA CGAGAT TAACCC A (SEQ ID NO: 1342)		EAFPGTSSMTPMET
	GGCTACGAGTGCAGGCCGGGCCAAGTCCCACCGCGGTGCACTCCGTCG				DRVHNKSGIDPILEH
	AAGCTTTCCCCGGAACGTCCTCGATGACCCCGATGGAGACCGATCGGGT				NGTRQVRREESSTRE
	ACATAATAAGTCCGGTATAGACCCCATTCTCGAACATAATGGGACACGG				DPVEQWSPNYPKTP
	CAGGTTCGAAGGGAGGAGTCTTCCACACGGGAAGACCCCGTCGAGCAA				VTMPNITTTADAXXT
	TGGTCTCCCAACTATCCTAAAACGCCTGTCACTATGCCCAACATTACAAC				SYNRTPQTLPGNRRR
	TACTGCTGATGCTWCGTKCACCTCGTACAACCGAACACCTCAGACACTC				RSRSLPPVQRKSASD
	CCTGGAAACCGGAGGCGACGCAGTCGCAGTCTACCACCAGTGCAACGG				XLESVDSLGPWAVFL
	AAATCCGCTTCCGACGAWCTCGAGTCCGTTGATTCGCTGGGGCCCTGG				QDQVDAGSLSGNDS
	GCGGTCTTCTTGCAAGATCAAGTAGACGCCGGGTCACTAAGCGGCAAC				LADLVRVALTKSDRG
	GACTCGCTGGCGGATCTGGTGAGGGTGGCTCTCACTAAGTCCGATCGC				VLNDAVNRYLAQRA
	GGTGTCCTTAATGATGCCGTGAATCGGTATCTGGCCCAAAGGGCGGAG				ESLRIRKRGSKGKRKS
	AGCTTGAGGATACGTAAGCGAGGCAGTAAGGGCAAAAGAAAGAGTAA				KTGRHYGQTTSGSG
	GACTGGTCGCCACTATGGCCAAACAACATCTGGCAGTGGTCAAAGAGCT				QRAALFKKHQDLFLK
	GCGCTCTTTAAAAAGCACCAGGACCTCTTTCTGAAAAACAGGCGAGGAT				NRRGLAETILSGKEDF
	TGGCGGAGACTATCTTGTCGGGGAAGGAGGATTTTGGTCCTAGACCTG				GPRPEPPVTSVEEFY
	AACCTCCCGTAACAAGTGTGGAAGAGTTCTACGGCGGCATTTTCGAGTC				GGIFESPSPPDNEPFE
	GCCTTCTCCTCCCGATAACGAACCCTTCGAAGTCCGTGCTACTGGCGTCG				VRATGVEDPPPLTSP
	AAGACCCCCCCCCCACTTACATCACCATGGACGAAATAAAGGCGGCAAGA				WTKRTLRSPCYWRR
	GCCGGTTGGCAGATTTCAGCGCCTGGATCTGACCAAATACCGGTCGCGG				RPPPTYITMDEIKAAR
	CCGTCAAAACTATGTCCGAGCTGGAACTGGCGATACTGTTCAATATTAT				AGWQISAPGSDQIPV
	ATTGTTCAGAAACGTCCAGCCTTCAGCCTGGGGGGTACTTCGTACGACG				AAVKTMSELELAILFN
	CTGGTTCCCAAAGATGGGGATCTTAGAAATCCGGCCAACTGGCGGCCCA				IILFRNVQPSAWGVL
	TTACWATCAGTTCTGCCCTGCAGCGTCTTTTGCATAGAGTATTGGCGGC				RTTLVPKDGDLRNPA
	CCGTCTTTCGAAGTTGGTCTCGTTGAGTTCGTCTCAGCGGGGCTTTACCG				NWRPITISSALQRLLH
	AGATTGACGGTACTCTAGCGAATGCGCTGATACTGCATGAATATTTGCA				RVLAARLSKLVSLSSS
	GTATCGCAGGCAGACTGGCCGCACCTACCASGTCGTGTCTCTGGACGTC				QRGFTEIDGTLANALI
	AGAAAGGCGTTTGACACGGTTTCCCACTGCTCTGTTAGTAGAGCCCTGG				LHEYLQYRRQTGRTY
	GTCGTTTTGGTATCCCCTCTGTTATCAGGGAATACATCCTGGCCACATTT				XVVSLDVRKAFDTVS
	GGCGCACAGACGACCATAAAATGTGGGAGCGTTACGACAAGACCGATA				HCSVSRALGRFGIPSV
	AGGATGCTCCGTGGAGTGAGACAGGGTGACCCTCTTAGCCCGGTCCTAT				IREYILATFGAQTTIKC
	TCAATCTTGTGATGGATGAGCTTCTGGAAAAGGTCAATGAGAAATACGA				GSVTTRPIRMLRGVR
	GGGCGGCTCACTGCAAAGCGGCGAACGCTGTGCCATCATGGCGTTTGC				QGDPLSPVLFNLVM
	TGACGACCTCATCCTGATTGCTGACCGAGATCAGGATGTACCCGCAATG				DELLEKVNEKYEGGSL
	TTCGATGATGTTTCGACATTTCTCGAGCGGCGCGGGATGTCTGTCAACC				QSGERCAIMAFADDL
	CTGCCAAGTGCAGAGCGTTGATTGCGGGCGCGGTGAGCGGGAGAAGT				ILIADRDQDVPAMFD
	GTTGTGAGGACAGGCAGCTCTTATAAGATCCATAACACGCCGATTCCAA				DVSTFLERRGMSVNP
	ATGTCGACGCCCTAGATGCATTTAAGTATCTGGGGCTAGAATTTGGCCA				AKCRALIAGAVSGRS

			CAAAGGCGTAGAGCGCCCCACAATCCACAACCTCTCCGTCTGGCTTAAC				VVRTGSSYKIHNTPIP
			AACCTCAGACGAGCCCCCTCTCAAACCGGACCAAAAATGCTTATTCATCCG				NVDALDAFKYLGLEF
			GCAATATGTCATCCCTCGATTGCTCTACGGCATGCAAAATCCTCAAGTTA				GHKGVERPTIHNLSV
			CCTCTAGGGTACTTCGAGAAGCTGACCGTCTGATAAGGAGACACTTAAA				WLNNLRRAPLKPDQ
			GACATACTACCATCTTAATGTCCATACTCCAGATTCTCTTATTCACGCGA				KCLFIRQYVIPRLLYG
			GTGTGAGTGACGGAGGTCTGGGTATAATGGAATTGCGTAAAGCAATTC				MQNPQVTSRVLREA
			CAAGAATTTTCCTAGGACGGCTAGTTAAGCTTCTCAACAAAAACAAAGA				DRLIRRHLKTYYHLNV
			TTCGGTCTTAAGTTCGGTCCTGCAGTCTAACCGTGTGAGAACACTAATG				HTPDSLIHASVSDGG
			GGCAAATTATCAACCATGGCGGGTGAAGTACCGGAGTCGACGTTTTGG				LGIMELRKAIPRIFLG
			AGAAATCAGATTGCCTCGGGTCCCCTTTCGAAGGGTCTCGAACAGGCGG				RLVKLLNKNKDSVLSS
			CCGAGGACTCCGCAAGTCGACTCTGGATCAGCGAGAAACCTTCCGGTTG				VLQSNRVRTLMGKLS
			GTCGGGTAGAGATCATGTCCGTGCAGTCCAACTTCGCACGGGCAACCTC				TMAGEVPESTFWRN
			CCGACCAAAGCCATTCCATCGGTGCCGGTTGGTCAGAGACGCTGCAGAC				QIASGPLSKGLEQAA
			ACGGATGTGCCTGCGATGAATCGATATCGCATGTGCTGCAGATGTGCCC				EDSASRLWISEKPSG
			CCTGACGCACGCTGATCGGATTCGCAGGCACGATGAAGTGGTTAAGAA				WSGRDHVRAVQLRT
			AGTAGCCCGCCATTGCACCTCACGCGGTTGGACGGTGGAAGTCGAGCC				GNLPTKAIPSVPVGQ
			ACACATTCGCTCTAGGTGTGGTAGGCTTTTTAAGCCTGACCTGGCAGTCC				RRCRHGCACDESISH
			ATCAGCCAGGCGGCGCCATTGTGATAGCGGATGTCCAGGTATCATGGG				VLQMCPLTHADRIRR
			ATTCCGAAAGCTTGACGGTACCTTACGAACGCAAGCGAGCTAAGTACGA				HDEVVKKVARHCTSR
			CGTGCCCCAATTCCATCAGGCGGCACAACATGCATGGCCGGGTAAGGC				GWTVEVEPHIRSRCG
			ACTTACGTTTGCCCCTGTAATAGTGGGCGCCCGGGGCATATGGGCCCCGG				RLFKPDLAVHQPGGA
			ATAAACAACGACCGCTCTGCGGCGTTTGCAGATTCCTCCGGTCGTGCGAC				IVIADVQVSWDSESL
			GAGCTTGCGTAAATTCGGTTGTCAAGTGGGGGGTCGTCGATCCATGCTAC				TVPYERKRAKYDVPQ
			CTTTATGAGGTCAGTATGGGCAAATCGACTCAACCCCCGACCCTTAAGG				FHQAAQHAWPGKA
			GCTTGAAAAGAGTGGCAGTTGTGGAGACTTCCTTCGGACGTCGGGTCG				LTFAPVIVGARGIWP
			GATTTTCGGACGCCAGGGTACCTCCACCGCTGGGTTCACAAACTAGGCG				RINNDRSAALQIPPV
			AACATCTGCCGATACCCTCTTTAGGTCATAGGACCACATGTCTCTGCACG				VRRACVNSVVKWGS
			AGATTAACCCA (SEQ ID NO: 1096)				SIHATFMRSVVWANR
							LNPRPLRA (SEQ ID NO: 1464)
R2	R2-1_TGut	.	Tinamus guttatus	CTGGGGACCGTGGTTACAACCCGGGCTTAGCTGCAGAGACAGTACCTCC	CTGGGGAC	TAGGGG	MGSWIVNFVSVATQ
				CCGTGGTTCCCGCCGGACCCCGTAACATCGGGTGACTGAATCTGTCTCT	CGTGGTTA	GCTTGG	TGEFPVDTARRAPVP
				GCCCCGGGAGTAGTTCCTCCTTGCCCTATTGACCAGCGGTCGCCGGCTG	CAACCCGG	CATTTCT	VTSYPESECHXPLPLT
				CTCAATAGTATTCTAGGCGTGAAATATAGCGATAGTCCTAGTGGTTGTCT	GCTTAGCT	CATTGCC	FCNSDVTIWGGVRPE
				TACTGGGCCATAGCCCCTTGCTTCAGGGGTCATTCGCGAAGTCTCTCAG	GCAGAGA	TGCTCCT	PVDCLGDLPEXYDAL
				GAGAACTGGGGGTGGTGTTCTTCTGGGTATAGCTAAACCCCCTAGACTG	CAGTACCT	GAAAGG	PGVAGPREXVGGSP
				TGTCCGATCCATGGGGTCCTGGATCGTGAATTTCGTTTCGGTGGCGACT	CCCCGTGG	ATATGG	PGEGVRSPGIASPSG
				CAGACGGGAGAATTCCCTGTGGATACGGCCAGGAGGGCACCTGTGCCG	TTCCCGCC	GTCCTG	TAVQHDFGSPILVPG

		GTAACATCATACCCTGAGTCGGAATGCCACNTACCGTTGCCCCTGACATT	GGACCCCCG	CGTCGC	AEAAEVSTPVVKVPQ
		TTGTAACTCGGATGTGACTATTTGGGGAGGGGTTCGCCCTGAACCGGTG	TAACATCG	GTGGTA	DHPACPCCGTRVVK
		GACTGCTTGGGWGATCTTCCRGAGGYGTATGATGCACTCCCAGGGGTG	GGTGACTG	GGCAGA	VTALSEHLRRAHGRK
		GCTGGGCCTCGGGAAYCGGTGGGTGGGAGCCCGCCGGGRGAAGGGGT	AATCTGTC	CCCATTC	RVLFQCSRCGRMNE
		CAGGTCGCCAGGGATTGCGTCRCCCTCTGGTACTGCGGTCCAACATGAT	TCTGCCCC	GTCCGA	KHHSIACHFPKCRGP
		TTTGGGAGTCCCATCCTCGTACCGGGAGCCGAAGCCGCCGAGGTNTCTA	GGGAGTA	GTAGGG	PVEEGPLGAPEWCCE
		CCCCGGTAGTGAAGGTTCCNCAAGACCATCCAGCATGTCCNTGCTGTGG	GTTCCTCC	GGCTTG	ECGQKFNTKSGLSQH
		TACGAGGGTGGTGAAAGTAACGGCGTTGTCAGAACATCTTAGGAGAGGC	TTGCCCTA	GCAGTN	KRSVHPLTRNVERIEA
		CCACGGTCGGAAACGGGTCCTATTTCAGTGCTCCCGATGCGGGAGGAT	TTGACCAG	TCCATTG	ARPKGKGKRGAHKG
		GAATGAGAAACATCATAGCATCGCGTGCCATTTCCCGAAGTGCCGGGG	CGGTCGCC	CCTGTG	CWTEAEVAQLIELEG
		GCCCCCAGTTGAGGAGGGTCCCCTGGGTGCACCCGAGTGGTGCTGTGA	GGCTGCTC	CCCGAA	RFKNQRFINKLIAEHL
		GGAGTGCGGGCAGAAATTTAACACCAAAAGCGGCCTGTCTCAGCACAA	AATAGTAT	AGGACG	PSKSAKQISDKRRQL
		AAGATCTGTGCACCCACTTACGAGGAATGTGGAACGGATAGAGGCAGC	TCTAGGCG	TGGGTC	AAATKTSSPEKRVTSS
		TCGTCCGAAAGGAAAAGGGAAGCGTGGTGCCCACAAAGGCTGTTGGAC	TGAAATAT	ATCTGG	TSGESSPEVEKVEGIK
		CGAGGCGGAGGTGGCCCAGCTGATTGAACTGGAGGGGAGATTCAAGA	AGCGATAG	TCTGTCT	REYRRRVGEWLCAG
		ACCAGCGATTTATCAACAAGCTGATCGCGGAGCATTTGCCATCGAAATC	TCCTAGTG	GCCTAC	SLXDQTSFQKILEDVE
		GGCGAAGCAGATCAGCGATAAAGAGAAGGCAGCTGGCGGCAGCGACCA	GTTGTCTT	ACCTCTC	SGSEIVTGPLEELASF
		AGACATCGTCGCCCGAGAAGAGGGTAACGTCATCAACGAGTGGGGAGT	ACTGGGCC	TAGACTT	ARGKLAAARVRHHR
		CCTCCCCTGAAGTGGAAAAGGTGGAGGGTATCAAGAGAGAATATAGAA	ATAGCCCC	GTAACA	KHPAEAVPAREEQR
		GGCGTGTTGGAGAGTGGTTGTGCGCTGGGTCCCTGMAGGACCAGACN	TTGCTTCA	TCTAGTC	WMKRRVGRRGLYLR
		TCGTTCCAGAAGATCTTGGAAGATGTGGAGAGCGGCTCCGAGATTGTC	GGGGTCAT	TGTCAA	FQRLFALDRRKLAGII
		ACCGGTCCGCTGGAGGAACTGGCCTCCTTTGCGAGGGGGAAGCTCGCG	TCGCGAAG	CAAGAT	LDDVESIKCPLPMEE
		GCAGCTAGAGTGCGACATCATCGTAAGCACCCAGCTGAGGCCGTGCCT	TCTCTCAG	CAAAAT	VADVFRRRWEEVAP
		GCGCGAGAGGAGCAGAGGTGGATGAAGCGCAGGGTGGGTCGTCGGG	GAGAACTG	TCTTCAC	FTGSGSFRSLGKADN
		GCTTGTACCTCAGGTTCCAGCGGCTTTTTGCNTTGGATCGCAGRAAGCTT	GGGGTGG	ACAGAC	GAFKPMISAKEVMK
		GCTGGGATCATCCTCGACGATGTCGAGTCCATCAAGTGCCCCCTTCCGA	TGTTCTTCT	GACCGA	NVXEMSRRSAXGPD
		TGGAAGAAGTCGCTGACGTCTTCAGGAGAAGGTGGGAGGAGGTCGCC	GGGTATAG	GCTTGC	GLSLRDLMKIDPQGS
		CCCTTTACCGGCTCGGGCTCATTCCGAAGTTTGGGGAAGGCTGACAACG	CTAAACCC	TCAGTCT	RMAELFNLWLLAGR
		GTGCCTTCAAGCCCATGATCTCCGCYAAGGAGGTCATGAAGAACGTCRC	CCTAGACT	TCCTGTA	VPDQVKAGRTVLIPK
		GGAGATGTCTCGACGCTCCGCGCYGGGWCCCGAYGGCCTCTCCCTGCG	GTGTCCGA	CCCGCA	SADPGKIGNIDNWR
		GGATCTGATGAAGATTGATCCCCAGGGCAGCCGCATGGCTGAATTGTTC	TCC (SEQ	GAATTTT	PITIGSVXLRMFSRILS
		AACCTGTGGCTGTTGGCAGGACGGGTCCCGGACCAAGTGAAGGCGGGC	ID NO:	GCTCTT	ARLRRACPINRRQRG
		CGAACGGTCCTGATCCCNAAGTCGGCCGATCCCGGGAAGATCGGGAAC	1220)	GCTCTCC	FIAAPGCSENLKLLQA
		ATTGACAACTGGCGGCCCATCACCATCGGGTCCGTTATNCTCAGAATGT		TTTGGCT	LIKSAKRDHRTLGVVF
		TCTCTCGGATTTTGAGCGCTAGACTGCGGCGAGCATGCCCCATTAATAG		GTGTCC	VDLAKAFDSVNHQHI
		AAGGCAAAGGGGGTTTATAGCAGCCCCTGGCTGCTCGGAAAATCTGAA		TGGACG	FQVLVQKGVDGHIID
		GCTTCTCCAGGCGCTCATCAAGAGCGCGAAACGAGATCATAGGACCCTT		TGGGAC	ILRDLYTNAGTYLESG
		GGAGTCGTGTTTGTCGATTTGGCTAAGGCCTTCGACTCCGTGAACCACC		TATTCCA	SQRSGFIKILRGVKQG

	AGCATATCTTCCAAGTCCTGGTCCAGAAGGGTGTCGATGGGCATATTAT	TCTCGTC	DPLSPILFNLALDPLLC
	CGACATCCTAAGAGACCTGTATACCAACGCTGGAACGTATCTGGAGTCA	CCAAAT	RLEDRGLGYKYGDQ
	GGTTCCCAGCGATCGGGATTTATTAAGATCCTCAGGGGAGTGAAACAR	GCCGCG	QIXSLAFADDLALLSD
	GGGGACCCACTCTCTCCCATCCTGTTCAATCTTGCATTAGATCCTTTGCTG	TCCAATT	SWEGMQQSIRVVEE
	TGCCGCCTGGAAGATCGGGGCCTCGGTTATAAGTATGGAGACCAACAA	ATACCG	FCQRTGLRVQAPKC
	ATAWCATCGTTGGCATTTGCGGATGATCTCGCCCTGCTCAGCGACTCTT	GATTTG	HGFLIRPTKESYTIND
	GGGAGGGCATGCAGCAGAGTATTCGGGTGGTAGAGGAAATTTTGTCAAC	ACAAAG	CDPWTIADMQLDMI
	GGACCGGGCTGCGGGTTCAAGCGCCGAAATGCCACGGGTTTTTGATCA	CGGACG	DPGSSEKYLGLGIDP
	GGCCAACTAAAGAGTCATATACCATCAATGACTGTGACCCGTGGACGAT	GCCCGC	WIGLSRPELSEVLTR
	TGCAGATATGCAATTGGATATGATCGATCCGGGCAGTTCCGAGAAGTAT	TTTATAA	WVKNIGGAPLKPLQ
	CTTGGCCTAGGGATAGACCCATGGATTGGTCTATCGAGACCGGAACTGT	GCCGGA	KVDILRSYALPRLLFIA
	CCGAGGTGCTGACCCGCTGGGTGAAGAACATCGGGGGCGCCCCTTTGA	AAAGGT	DHAGLSATCLHSLDL
	AGCCACTCCAGAAGGTGGACATCTTGAGGAGCTACGCCCTNCCAAGGC	GCCTTG	SIRSAVKGWLHLPPS
	TGCTGTTCATTGCGGATCACGCAGGCCTGAGCGCCACCTGTTTGCATTCC	TAAAATT	TCDAIIYVSYKDGGLG
	CTGGACCTTTCGATAAGATCTGCCGTCAAGGGCTGGTTACATCTACCGC	GCAAGG	LPRLASLIPNVQARRL
	CTAGTACGTGTGACGCTATTATTTACGTCAGCTACAAGGACGGCGGGCT	TTCATTA	VRIAQSEDDVIRSVVL
	GGGTCTTCCCCGTCTGGCGAGCCTAATTCCAAATGTACAGGCTCGCAGG	AATAG	QEGIQEEIRKVWISA
	TTGGTGCGGATCGCCCAATCGGAGGATGATGTCATCAGGAGTGTGGTA	(SEQ ID	GGRPEKVPSVTGEFP
	CTCCAGGAGGGTATCCAGGAGGAGATCCGGAAGGTCTGGATCTCGGCT	NO:	VMEAQAADEALSEW
	GGGGGGCGACCGGAGAAGGTTCCATCTGTGACGGGGGGAGTTCCCAGT	1343)	ERRAPRTIYPIPCKWR
	GATGGAGGCTCAGGCGGCTGACGAGGCCCTATCCGAGTGGGAGAGGC		KREMENWTNLKSQG
	GAGCTCCACGAACCATCTATCCCATTCCCTGTAAGTGGAGGAAGAGAGA		HGIRNFENDRISND
	AATGGAGAATTGGACCAATCTAAAATCGCAAGGCCACGGGATTCGGAA		WLLHYGRIPHRKLITA
	TTTTGAAAATGACCGAATCAGTAATGATTGGCTCCTGCATTATGGCCGC		IQLRANVYPTREFLAR
	ATTCCCCACCGCAAACTAATAACAGCTATCCAGTTGCGGGCCAATGTCTA		GLGEGAPRGCRHCP
	TCCCACWCGGGAGTTTTTGGCCCGCGGCCTGGGCGAGGGCGCACCCAG		AEWESCSHIIGYCPA
	GGGATGTAGGCACTGTCCCGCGGAGTGGGAATCTTGTTCCCACATAATT		VQEARIKRHNDICGV
	GGCTACTGCCCGGCTGTCCAGGAGGCCAGGATCAAAAGGCATAATGAC		LAEEARKLGWVIFIEP
	ATCTGTGGTGTGCTGGCTGAAGAGGCNAGAAAGCTGGGATGGGTGATA		HLRDNTNELFKPDLV
	TTTATAGAGCCCCATCTCAGAGATAACACCAATGAGCTCTTCAAGCCAG		LVKGSCAKVVDVTIR
	ATTTGGTTTTGGTGAAGGGATCCTGTGCGAAGGTAGTGGATGTAACCAT		YESGLTTLSDAAAEKA
	CCGCTACGAGAGTGGGTTAACCACCTTGAGTGACGCCGCGGCAGAAAA		RKYQHLAGEVRALTS
	GGCTAGGAAGTATCAACATCTGGCAGGGGAGGTGCGGGCCCTAACATC		ATTVDFLGFPIGARG
	GGCCACTACTGTAGACTTCCTGGGTTTCCCTATTGGCGCTAGAGGGAAG		KWYVGNNGLLSDLG
	TGGTACGTTGGTAATAATGGACTCCTTTCCGACCTTGGGTTCTCCACTAG		FSTSRVVRIARALSKK
	CCGTGTAGTGCGGATAGCGAGGGCCCTCTCTAAAAAGGCTCTCCTATCG		ALLSSVDIIHIFASRAR
	TCCGTGGACATTATACATATTTTTGCGTCTCGCGCTAGACAGGCCCAAAC		QAQTSE (SEQ ID
	GTCCGAGTAGGGGGCTTGGCATTTCTCATTGCCTGCTCCTGAAAGGATA		NO: 1465)

			TGGGTCCTGCCGTCGCGTGGTAGGCAGACCCATTCGTCCGAGTAGGGGG				
			CTTGGCAGTNTCCATTGCCTGTGCCCGAAAGGACGTGGGTCATCTGGTC				
			TGTCTGCCTACACCTCTCTAGACTTGTAACATCTAGTCTGTCAACAAGAT				
			CAAAATTCTTCACACAGACGACCGAGCTTGCTCAGTCTTCCTGTACCCGC				
			AGAATTTTGCTCTTGCTCTCCTTTGGCTGTGTCCTGGACGTGGGACTATT				
			CCATCTCGTCCCAAATGCCGCGTCCAATTATACCGGATTTGACAAAGCG				
			GACGGCCCGCTTTATAAGCCGGAAAAGGTGCCTTGTAAAATTGCAAGGT				
			TCATTAAATAG (SEQ ID NO: 1097)				
R2	R2-1_TSP	.	Trichinella spiralis	CTCCTGACTAACCTGATTTCGTCCGTGCGGCGGCGTTTTCTTTTCGCTCTC	CTCCTGAC	TGAGGT	MSNRLANTAAAGGV
				CGCTCGTCGAAATTTGCTGTAGTTGATTCGCTTTTCTTTGCGTTTTTCTTCT	TAACCTGA	TTTTGTT	PEKTSGTLDIPGQPSS
				ACTTTCGCAGTTTTTTTCTGCATTGCCACGATGTCAAACCGCCTTGCCAAT	TTTCGTCC	TTCTTTT	SGEKRAISYPGPFGC
				ACTGCTGCGGCTGGTGGGGTTCCAGAGAAAACCTCGGGAACTTTAGAC	GTGCGGC	TTCCTTT	NSCSFTSTTWLSLELH
				ATTCCTGGCCAACCCTCTTCATCCGGTGAAAAGCGTGCGATCTCTTACCC	GGCGTTTT	TACCATT	FKSVHNIRDFVFLCSK
				TGGTCCATTCGGTTGCAATTCGTGTTCGTTTACGAGTACGACTTGGCTCT	CTTTTCGC	CTTGTTC	CKKSWPSINSVASHY
				CATTGGAATTGCATTTTAAAAGCGTCCATAATATTCGTGACTTCGTCTTC	TCTCCGCT	CATTGTT	PRCKGSVKAAVVPTS
				CTCTGCTCTAAATGTAAAAAAAGCTGGCCATCGATCAACTCCGTAGCTA	CGTCGAAA	GTTATTT	LANTCTTCGSSFGTFS
				GCCATTACCCTCGGTGCAAAGGTAGCGTCAAGGCTGCAGTTGTTCCTAC	TTTGCTGT	GCTTTAA	GLQLHRKRAHPDVF
				ATCTTTGGCGAATACGTGCACCACGTGCGGCTCAAGCTTCGGTACTTTCA	AGTTGATT	TCCTGTA	AASCSKKTKARWSN
				GTGGTCTTCAACTCCATCGGAAAAGAGCACATCCGGACGTTTTTGCTGC	CGCTTTTC	TTTTACC	DEFTLLARLEAGLDPA
				TTCTTGTAGCAAAAAAACGAAGGCGCGTTGGTCTAACGACGAATTTACC	TTTGCGTT	GCCGGC	CKNINQVLAERLMEY
				CTTCTGGCGAGACTCGAAGCAGGTCTGGATCCAGCCTGTAAAAACATTA	TTCTTCTAC	AATTCCA	NITRGVEMIKGQRRK
				ACCAAGTACTAGCGGAAAGGTTAATGGAGTATAACATCACCAGAGGCG	TTTCGCAG	TTGTTAT	DQYKALVRQLRSNSE
				TAGAAATGATAAAAGGCCAACGTAGAAAAGATCAGTACAAAGCGCTCG	TTTTTTCTG	TATTACT	TQQCVGLAGSMDSN
				TTCGTCAACTCCGGTCAAATTCTGAAACACAGCAATGTGTAGGTTTAGCC	CATTGCCA	GTTACT	VPANDTSSSVASEVSI
				GGAAGTATGGATTCGAACGTACCGGCCAACGATACATCGTCTTCCGTTG	CG (SEQ ID	GTTATTA	TYPEYGAVMSCDLIK
				CATCAGAGGTCAGCATTACGTACCCTGAGTACGGGGCCGTGATGTCGTG	NO: 1221)	TTGTTAC	EATGMAIVDINELQS
				CGACCTAATTAAAGAAGCGACTGGTATGGCCATAGTTGACATCAACGAG		TATTGTT	NLRKAFLSGRKLPMK
				TTGCAAAGCAACTTACGAAAAGCCTTCTTGTCCGGCCGCAAGCTTCCCAT		TTTACTT	FHGARETAQKKMAN
				GAAGTTCCATGGAGCGCGTGAAACCGCCCAGAAGAAAATGGCCAACCC		TTACTTA	PRVAKFKRFQRLFRS
				CCGTGTTGCGAAATTCAAGCGTTTCCAACGGTTGTTTCGAAGCAACAGG		CTACTGT	NRRKLASHIFDKASLE
				AGGAAACTGGCCAGCCACATCTTCGACAAAGCCTCACTGGAGCAATTCG		TATTATA	QFGGSIDEASDHLEK
				GTGGCAGCATCGATGAGGCATCTGACCATTTAGAAAAGTTCCTCTCCCG		CTTTAAT	FLSRPRLESDSYSVISG
				GCCAAGATTGGAGTCCGATTCTTATTCCGTGATAAGCGGTGATAAGTCA		TCGTTAA	DKSIGVAHPILAEEVE
				ATCGGAGTTGCACATCCAATTTTGGCCGAGGAGGTGGAATTGGAATTAA		CTTACGT	LELKASRPTAVGPDGI
				AAGCCTCCCGACCAACCGCTGTTGGTCCGGATGGAATTGCACTGGAAGA		TATTGTT	ALEDIKKLNTYDIASLF
				CATTAAAAAACTCAATACTTACGACATAGCCAGTCTTTTCAACCTCTGGC		ACCACT	NLWLKAGDLPASVK
				TAAAAGCTGGCGACCTACCCGCATCGGTGAAAGCCAGTAGAACCATCTT		ACTTACT	ASRTIFLPKSDGTTDIS
				TTTGCCCAAAAGCGACGGCACCACCGACATATCGAACTGTCGGCCAATC		TTGCTCT	NCRPITIASAMYRLFS

		ACAATCGCATCCGCCATGTATCGGCTGTTCAGCAGAATAATAACGCGAC		CTCGCA	RIITRRLAARLELNVR
		GTCTGGCAGCCAGGTTGGAATTGAACGTGCGGCAAAAAGCGTTCCGGC		AACGTT	QKAFRPEMNGVFEN
		CTGAAATGAACGGCGTATTCGAGAACTCCGCCATTTTATACGCCCTCATC		CGTTGTT	SAILYALIKDAKVRSRE
		AAGGATGCTAAGGTCAGGTCAAGGGAAATTTGCGTAACTACGCTCGAC		GTTTCTT	ICVTTLDLAKAFDTVP
		CTTGCCAAGGCCTTTGACACGGTGCCCCACTCACGCATTTTACGAGCCCT		TTGGAC	HSRILRALRKNNVDP
		GAGGAAAAATAATGTCGACCCGGAATCCGTCGACCTGATTTCGAAAATG		CAGGTT	ESVDLISKMLTGTTYA
		TTAACGGGTACGACTTATGCAGAAATAAAAGGGCTCCAGGGCAAACTTA		TAGAGA	EIKGLQGKLIPIRNGV
		TACCCATTCGCAATGGAGTCAGGCAAGGTGACCCCTTGTCGCCCCTATT		AATCGC	RQGDPLSPLLFSLFID
		ATTTAGTCTATTTATAGACGAGATAATAGGTCGCCTACAAGCCTGCGGC		ACGCAC	EIIGRLQACGPAYDFH
		CCTGCCTACGATTTCCATGGCGAAAAAATTTGCATCCTGGCTTTCGCCGA		AGCGGA	GEKICILAFADDLTLV
		TGATCTGACGCTGGTGGCTGACAGCGCAGCTGGTATGAAGATCCTTCTA		ACTGGA	ADSAAGMKILLKAAC
		AAAGCGGCTTGTGACTTCCTGGAGGAATCTGGAATGTCACTTAATGCAG		CCGCTTA	DFLEESGMSLNAEKC
		AGAAATGCCGCACTCTCTGTATTACAAGATCTCCCCGAAGCCGCAAGAC		AGCCAG	RTLCITRSPRSRKTFV
		TTTCGTCAACCCAGCTGCCAAATTCATCATCAGCGATTGGAAAACGGGT		AAATAG	NPAAKFIISDWKTGIS
		ATCAGCTCAGAAATCCCCTCCCTGTGTGCGACGGACACCTTTCGTTTCCT		TAAAGT	SEIPSLCATDTFRFLG
		GGGGCACACCTTCGATGGAGAAGGAAAGATCCACATCGATACGGAGGA		AACAA (SEQ ID NO: 1344)	HTFDGEGKIHIDTEEI
		AATTCGATCCATGCTCAAATCGGTGAAGTCAGCTCCACTGAAACCGGAA			RSMLKSVKSAPLKPE
		CAGAAGGTGGCTTTGATACGGTCACACCTTCTTCCCCGCCTTCAGTTCCT			QKVALIRSHLLPRLQF
		GTTTTCTACAGCTGAAGCTGACAGCCGGAAAGCCTGGTTGATCGATTCC			LFSTAEADSRKAWLI
		ATCATCAGGGGGTGTGTGAAGGAGATCTTGCACTCAGTGAAAGCTGGT			DSIIRGCVKEILHSVKA
		ATGTGCACTGATATCTTTTACATACCCTCTAGAGACGGTGGAATGGGAT			GMCTDIFYIPSRDGG
		TTACTTCCCTCGGGGAGTTTTCTCTTTTCAGCAGGCAGAAGGCACTCGCC			MGFTSLGEFSLFSRQ
		AAGATGGCTGGATCGTCGGACCCCCTCTCGAAACGGGTTGCTGAATTCT			KALAKMAGSSDPLSK
		TCATCGAAAGGTGGAACATCGCCCGTGACCCGAAAGTCATTGAAGCTGC			RVAEFFIERWNIARD
		TCGGCGCGTCTACCAGAAAAAACGGTACCAACGCTTTTTCCAGACGTAC			PKVIEAARRVYQKKR
		CAGAGCGGTGGATGGAATGAATTTTCGGGAAACACTATTGGGAACGCC			YQRFFQTYQSGGWN
		TGGTTGACAAACGGCCGTGCCCGCGGAAGAAATTTCATAATGGCTGTG			EFSGNTIGNAWLTN
		AAATTCCGTTCCAACACCGCAGCCACCCGGGCCGAAAACCTACGAGGCC			GRARGRNFIMAVKF
		GCCCCGGCACGAAAGAATGCCGGTTTTGCAAGAGTGCCACCGAAACTTT			RSNTAATRAENLRGR
		GGCACACATTTGCCAGAGGTGTCCGGCAAATCACGGCTTGGTTATCCAG			PGTKECRFCKSATETL
		CGCCATGACGCAGTCGTAACATTCCTGGGGGAAGTGGCGCGGAAGGAA			AHICQRCPANHGLVI
		GGTTACCAGGTCATGATAGAGCCTAAGGTGTCAACCCCGGTCGGCGCG			QRHDAVVTFLGEVA
		CTCAAGCCCGACCTCCTACTCATCAAAGCCGACACTGCATTCATTGTGGA			RKEGYQVMIEPKVST
		TGTAGGCATTGCATGGGAAGGTGGACGCCCACTAAAGCTGGTCAACAA			PVGALKPDLLLIKADT
		AATGAAATGTGACAAGTACAAAACTGCCATCCCGGCAATTTTGGAAACA			AFIVDVGIAWEGGRP
		TTTCACGTTGGCCATGCTGAGACGTACGGCGTTATTCTGGGCAGCCGCG			LKLVNKMKCDKYKTA
		GATGCTGGCTCAAGAGCAACGACAAGGCGTTGGCATCAATTGGGCTCA			IPAILETFHVGHAETY
		ATATCACACGGAAGATGAAAGAACACCTGAGCTGGTTGACGTTTGAAAT			GVILGSRGCWLKSND

			TATATTTATAACTCAAATAAGCCGGATTTATAACTCATTCATGAAAAAAT GAGGTTTTTGTTTTCTTTTTTCCTTTTACCATTCTTGTTCCATTGTTGTTATT TGCTTTAATCCTGTATTTTACCGCCGGCAATTCCATTGTTATTATTACTGT TACTGTTATTATTGTTACTATTGTTTTTACTTTTACTTACTACTGTTATTAT ACTTTAATTCGTTAACTTACGTTATTGTTACCACTACTTACTTTGCTCTCTC GCAAACGTTCGTTGTTGTTTCTTTTGGACCAGGTTTAGAGAAATCGCACG CACAGCGGAACTGGACCGCTTAAGCCAGAAATAGTAAAGTAACAA (SEQ ID NO: 1098)	Tetranychus urticae	CTCTCTTATTTTAACATATTCGATGTACTCGTACATTGAATATGCTTTTATT TTTTTTCAAAGTTTTTTTGGGTGCATACCCCTGGAAAATTCTGAGATGTAT AAATCTCCCATCAGCTTTGGCTGAAACGTTGGCTAAGTTTTGTAGGTTGT TTGCCCCCTACTACTTAGTCGCAAATGGTATTTGCTAACAGTTGTTAAAT TGTTACATTTACAAGTCCTATCCAGTGCCTCCTCGTGGCGCTACCCGGTA ACACTTAGAGTAATCTGAGTGGCTAAACTGGAAGGGCGGAAAATGCAA ACAGGCGGTTGGTAGATGCTTCGGCATTTTGCCAAAAATCCACGGCTTT TTAGCCCAACAACATCAGGGTGATGGACCCGCCAGCTTGTGGTCAGGAT CCCATCCATGAATAAAGCATGGCTCTGCTTCTGGTGCATCCTCAACGGG ATCGGCTTCGGCTGGATGTAAGTCTTGCGGAGGCATGTGCATCTTGGGT GGCCTAACAAGCCATTCAAGAGAAGGAGGGCTCTCGCGAGGGAGTTCT CAACTTAAGACCGTTAAGCCACAAACGAAGAAGATAATGGAACTACCC AGTTGAAGGCTGGGAGTGCGGACTCTTTCCCAAGGCCAAGTGGTGATTT GAACCCAGAGGAACCACTGAGCATTGATATCTGTCCAGTTTGTTTCAGA CAGATGAAGAGTTATTTAGGCGTTAGAGTGCATATGCAGAAGATGCACT TAGAAGAATATAATGCGTCTATTCCGGACCCAGTCGTTTCCCACACTAG GTGGTCCGATGAAGAGGCAGCACAATTAGCTTTTACTGAAGCTAAGATT GAGGTCGACAAGCTTTTACCTCGTGGCAAAGGCATAAATAAGTTCCTGT TAGAGCTTTTGCCTGGTCGAACCTTAGAGTCTATCAAGTCTCACAGAAA AAGACAATCGCATAAAGACTTGGTTCGCAAATATGTAAAGGAGTTCGTT GATACTTTAGCTGCAGATAACGATGACGATACCATCATCTGTCAAGATA ATGGTGACATTTTTAATGACCCCATCGTTGGAGCAACTGATTCTCAGTCT GAGACTGAAACTGTGGCTGATCCTGCTGAATTCAAGACTTTTATTGAGTT AGCAGACGATCCCACAAAACCAAAAGTCGTAGCAAAACTCAGAAACTTG ATAAAAGATAAACCAAAGAGCGAGATCCTTGGAAGTGACATTCTTGTAC GAATCCTGCGAAGAACTCTTCATGGCTTGCCTGTTGAGGACGAGTTGGA CCAGTACTTGGAGGTTTACTTCACTGGGAAGATAAAGCAGAGACGTTCT AAGACACAAACTGCCTTATCTAAGAAACAGATTAAGCAGAGAGATTATG GACGTCTTCAGGAGTTATACTCCCGGTCTAGAAAAAGGTGTGCAAATGA GATCTTGAACCCTACTTCTATGAGTGGTGGTTTTGGTCATCAGGAGCTAT	CTCTCTTAT TTTAACAT ATTCGATG TACTCGTA CATTGAAT ATGCTTTT ATTTTTTTT CAAAGTTT TTTGGGTG CATACCCC TGGAAAAT TCTGAGAT GTATAAAT CTCCCATC AGCTTTGG CTGAAACG TTGGCTAA GTTTTGTA GGTTGTTT GCCCCCTA CTACTTAG TCGCAAAT GGTATTTG CTAACAGT TGTTAAAT TGTTACAT TTACAAGT CCTATCCA GTGCCTCC TCGTGGCG	TGATGT ATCCCTT CAATAT ATTGTA ATCCTCA TTCGTCC CTATCCT TTCATTT GATAAA AGAAAC TTTGTTG CTCCTTT TAATAG TTGGTC CCTCCTG TCCCTTT TCTGGA ACCTGG TTGTATC GATTATT GAAAGT TGCAAT AAACGG ATTTAA (SEQ ID NO: 1345)	KALASIGLNITRKMKE HLSWLTFEIIFITQISRI YNSFMKK (SEQ ID NO: 1466)	MCILGGLTSHSREGG LSRGSSQLKTVKPQN EEDNGTTQLKAGSA DSFPRPSGDLNPEEP LSIDICPVCFRQMKSY LGVRVHMQKMHLEE YNASIPDPVVSHTRW SDEEAAQLAFTEAKIE VDKLLPRGKGINKFLL ELLPGRTLESIKSHRK RQSHKDLVRKYVKEF VDTLAADNDDDTIIC QDNGDIFNDPIVGAT DSQSETETVADPAEF KTFIELADDPTKPKVV AKLRNLIKDKPKSEIL GSDILVRILRRTLHGL PVEDELDQYLEVYFT GKIKQRRSKTQTALSK KQIKQRDYGRLQELY SRSRKRCANEILNPTS MSGGFGHQELSEFW TKTFGPDEQPTLGEV EIIPKENCWWDIFSPI SSDEIKASYPSIGKAA GPDNFSAYQLRKVPV WHLECLYNIFAFYKDI PSRLKDAKTILIPKKD NAESPGDFRPITLSSII TRHFHKILATRVNNF
R2	R2-1_TUr	scaffold_6							

	CCGAATTCTGGACAAAGACGTTCGGACCAGATGAGCAACCTACTCTGGG	CTACCCGG		VRFHPMQRGFIQSD
	CGAGGTGGAGATCATTCCGAAGGAGAACTGTTGGTGGGATATTTTCTCG	TAACACTT		GCLENTALIQTVIREA
	CCAATATCTTCAGACGAAATCAAAGCCTCATACCCCAGTATAGGGAAAG	AGAGTAAT		KVRRKQVHITFCDVR
	CTGCTGGTCCAGATAATTTCTCTGCTTATCAACTTAGAAAAGTTCCGGTT	CTGAGTGG		KAFDSVRYDSIIAAIA
	TGGCATCTGGAATGTCTCTATAATATCTTTGCTTTCTACAAGGACATACC	CTAAACTG		KKGAPGSFIMYLSNL
	GAGTAGACTTAAAGACGCGAAGACAATTTTGATACCAAAGAAGGATAA	GAAGGGC		YRGNKTTLLTAGGET
	CGCTGAAAGTCCAGGTGATTTCCGCCCTATCACCTTGAGCTCAATTATCA	GGAAAAT		RITPTRGVRQGDPLS
	CTCGACATTTTCACAAGATTCTAGCGACGAGAGTGAATAATTTTGTTCGA	GCAAACAG		PILFNCVMDQILTALP
	TTCCATCCTATGCAAAGAGGCTTCATTCAATCTGATGGTTGTTTGGAGAA	GCGGTTG		SRTGFTLSAGDESVN
	CACTGCGCTCATACAGACGGTCATTAGAGAAGCCAAAGTGCGTAGGAA	GTAGATGC		VNCLAFADDIILISKTK
	GCAAGTGCATATCACCTTCTGTGATGTGCGAAAAGCGTTTGACAGCGTC	TTCGGCAT		NGHQELLDVTQRILK
	CGGTATGACTCGATCATAGCAGCTATTGCAAAGAAAGGTGCACCAGGA	TTTGCCAA		ENGLDLNPDKCCSLS
	AGTTTCATCATGTACCTTAGTAACTTATATAGAGGAAATAAGACTACGCT	AAATCCAC		LIPHSKTKKIKVVRAD
	TTTGACTGCTGGCGGTGAAACAAGAATAACTCCAACACGGGGCGTTAG	GGCTTTTT		FVVNGVKVRSMSIG
	ACAAGGAGATCCGTTATCTCCGATACTTTTTAATTGCGTGATGGACCAG	AGCCCAAC		DSTCYLGVSINVTGQ
	ATTCTCACGGCTCTTCCTTCGAGAACTGGTTTTACGTTATCTGCTGGTGA	AACATCAG		VAPVKMYQALCEKL
	CGAATCAGTTAACGTTAACTGTTTGGCGTTTGCTGATGATATAATTTTAA	GGTGATG		DSAAIKPHQRLYILKH
	TCTCAAAGACCAAGAATGGACATCAAGAACTCCTTGATGTAACACAAAG	GACCCGCC		FVITKMFHPLILSTIAA
	AATTTTAAAGGAGAATGGCCTCGATCTCAATCCTGATAAGTGCTGTTCG	AGCTTGTG		HKIKNLDLISRRYVRK
	CTAAGCTTGATACCGCACAGCAAAACAAAGAAAATCAAAGTGGTTCGTG	GTCAGGAT		WLHLPHDCGSGMIH
	CTGACTTCGTGGTGAACGGAGTCAAAGTGCGATCAATGAGTATTGGTG	CCCATCCA		AKVSDGGLGVPLLFR
	ATTCAACTTGTTACCTAGGTGTCTCAATAAATGTAACTGGCCAAGTAGCA	TGAATAAA		TIADLKVRRKEKLQV
	CCAGTTAAAATGTATCAGGCTCTGTGTGAGAAACTTGATAGTGCTGCTA	GCATGGCT		HENPIFRILAKLSTVSK
	TTAAGCCTCATCAGAGACTGTACATATTGAAGCATTTCGTAATAACGAA	CTGCTTCT		ELENCKKIASKTTDIQ
	GATGTTTCACCCGCTGATTTTGTCAACTATTGCGGCACACAAGATAAAG	GGTGCATC		EKTFKEMLATYDGLS
	AATCTGGATTTGATTAGCAGACGTTATGTCAGGAAGTGGTTACACTTAC	CTCAACGG		LKEARAVPEVHKWV
	CTCATGATTGCGGATCTGGTATGATTCATGCCAAAGTAAGCGATGGAGG	GATCGGCT		DSYDKRYKFAGRDFV
	ATTGGGTGTTCCATTACTGTTCAGGACAATTGCTGATCTCAAAGTTCGGA	TCGGCTGG		QVIQARFNALPTRSR
	GAAAAGAAAAAGCTGCAGGTCCATGAAAACCCAATATTTAGAATATTGGC	ATGTAAGT		VWRGRGADEKSLRC
	AAAGCTATCTACAGTAAGCAAGGAGTTGGAAAATTGTAAAAAGATAGC	CTTGCGGA		RAGCNARETLNHVS
	TTCCAAGACGACAGACATTCAAGAAAAGACTTTCAAGGAAATGCTTGCG	GGC (SEQ		QSCFRTHRVRTARHD
	ACTTATGATGGTTTGTCATTGAAAGAAGCTCGAGCGGTACCTGAGGTGC	ID NO:		KILDFICERLDVVGVK
	ACAAATGGGTGGACAGCTATGATAAGCGTTATAAGTTTGCAGGCCGAG	1222)		YVREKPISFPGKKLIPD
	ATTTCGTTCAAGTTATACAGGCTCGTTTCAATGCGTTACCAACGAGATCT			LIVENTDQALVLDLQI
	CGAGTATGGCGGGGCAGAGGTGCTGATGAAAAATCGTTACGTTGCAGA			VGDNSELPLDERGKN
	GCAGGCTGCAACGCTAGGGAAACGTTGAATCACGTTTCACAATCTTGCT			KVIKYNCSEMQELYK
	TCAGGACGCATCGAGTACGAACAGCTAGACACGACAAAATCCTGGATTT			RKKKTLAVKALTLHYK
	TATTTGTGAAAGGCTGGACGTAGTCGGTGTTAAATACGTACGAGAAAA			GLMAPETSNILRSFG

			GCCAATCAGCTTTCCTGGCAAAAAGTTGATTCCTGATTTGATTGTGGAAA ACACGGACCAAGCTCTTGTGCTCGATCTCCAGATTGTTGGCGATAACAG TGAACTTCCACTTGATGAAAGAGGCAAGAACAAGGTTATCAAATACAAT TGTTCGGAGATGCAGGAATTGTACAAGAGGAAGAAAAAGACTTTGGCG GTTAAGGCACTCACGCTACATTATAAAGGGTTGATGGCACCGGAAACAA GCAATATCCTAAGGTCATTCGGATTCAAGTCTAAGGATTTGGAAAAGAT GGCGTATATGGCGTTGTTTGGTACGGTGGCTGCTTGGGGCATTTTCAAT CGCTCGACGGAGACAATGCGTTCAGTCGCTAACTGGCCTAGACCTGAG GAATTATGATGTATCCCTTCAATATATTGTAATCCTCATTCGTCCCTATCC TTTCATTTGATAAAAGAAACTTTGTTGCTCCTTTTAATAGTTGGTCCCTCC TGTCCCTTTTCTGGAACCTGGTTGTATCGATTATTGAAAGTTGCAATAAA CGGATTTAA (SEQ ID NO: 1099)				FKSKDLEKMAYMALF GTVAAWGIFNRSTET MRSVANWPRPEEL (SEQ ID NO: 1467)
R2	R2-2_DWi	.	Drosophila willistoni	GAAGCTGGGAAGCTGGGTCGGATGAGCGCAGAAGGGGGTGTTCTTCGG AGCACTGTAATTCATAAGTCGTAAGTCTGATCAAGTCGACTCGGAACCT CTTCGTGGTGTTTCCTGGGTGCTGTTGAGTTCCTAGTCTCTAGGTTCTCT CCAGTAGCTAATTCGAGCGGCGAAGCAACTCTTGGGGTTACCAGAACCT TGAGCCAAGTAACGTTGGTCAAGATATGAATACTGTGCCTCGGATAAAC AACACAACAACTCCAGCGACTTCCCGTCCTGGAGACCAACCGAGAG AGGCTATAGCAGTGGTAAATCTCGCGGGAGAGATTCCCTGTGCAGTAT GTGGGCGCCTCTTCAATACTAGAAGGGGGTTTGGGGTTCATATGTCACA TCAACACAAAGACGAACTAGATACGCAACGTCAGCGTGAAGATGTAAA ACTCCGATGGAGTGAGGAAGAAGCGTGGATGATGGCGAGAAAGGAGG TGGAGCTCGAAGCAAGTGGTAATTTGAGATTTCCTAATAAGAAGCTAGC GGAAGTATTTACTCACCGTAGCTCCGAAGCAATTAAATGTTTTCGGAAG AGGGGTGAATATAAGGCAAAACTGGAGCAGATCAGAGGGCAATCTACT CCCACCCCAGAAGCGTTGGACTCAATTACCTCACAGCCTCGCCCTAGTTT ACTCGAGCGAAACCACCAAGTATCATCGTCGGAAGCGCAACCAATCAAT CCATCAGAAGAACAGTCGAACTGGGAAATCATGCGGATACTACAGGGC TATCGCCCCGTAGAATGTAGTCCCCGGTGGAGAGCCCAGGTCTTGCAAA CTATCGTAGATAGGGCGCAGGCCGTAGGGAAGGAAACCACTCTCCAAT GCTTATCCAACTATCTCCTGGAAGTATTTCCATTACCAAACGAACCACAC ACCATCGGTCGGAGCAATTTGCGAAGACCTCGAACTAGGAGACAGTTA AGACAACAAGAGTACGCACAGGTTCAGCGTCGTTGGGATAAGAATACT GGGAGATGCATTAAATCCTTGCTTGATGGAACAGATGAGTCGGTTATGC CAAACCAAGAGATAATGGAACCCTATTGGAAACAAGTAATGACGAATCC CAGCACATGCTCTTGCGAAAACACAAGATTCCGTATGGAACATTCGCTT GAGACGGTTTGGTCAGCGATAACGCCACCGACCTGAGGGAAAATAAG TTAAAGTTGTCAAGTGCTCCGGGTCCTGACGGTATCACTCCAAGAACAG	GAAGCTG GGAAGCT GGGTCGG ATGAGCGC AGAAGGG GTGTTCTT CGGAGCAC TGTAATTC ATAAGTCG TAAGTCTG ATCAAGTC GACTCGGA ACCTCTTC GTGGTGTT TCCTGGGT GCTGTTGA GTTCCTAG TCTCTAGG TTCTCTCC AGTAGCTA A (SEQ ID NO: 1223)	TAGATG TACTAAC CTCTAGT TTCTCTA TACTTTT GCCTGC TACCTTG GCATTA CATCTAA AAAGGT ACAAAC ATCGCA TTGGCA AAAAGA GGTGGT TTTAGTA CATAGG CGCTGT GGGACT TCATTGT CCCGAT GATGCA GCGAAT CGTGCA TACGAG ATTGTCC	FERRSNSWGYQNLE PSNVGQDMNTVPRI NNTTTTPATSRPGDQ PREAIAVVNLAGEIPC AVCGRLFNTRRGFGV HMSHQHKDELDTQR QREDVKLRWSEEEA WMMARKEVELEAS GNLRFPNKKLAEVFT HRSSEAIKCFRKRGEY KAKLEQIRGQSTPTPE ALDSITSQPRPSLLER NHQVSSSEAQPINPS EEQSNWEIMRILQGY RPVECSPRWRAQVL QTIVDRAQAVGKETT LQCLSNYLLEVFPLPN EPHTIGRSNLRRPRTR RQLRQQEYAQVQRR WDKNTGRCIKSLLDG TDESVMPNQEIMEP YWKQVMTNPSTCSC ENTRFRMEHSLETV WSAITPRDLRENKLK LSSAPGPDGITPRTAR SVPLGIMLRIMNLIL

		CCAGGAGTGTACCCTTAGGCATTATGCTACGCATAATGAACCTGATTCTC		AGTAGT	WCGKIPFSTRLARTIFI
		TGGTGCGGCAAAATACCATTCTCTACCCGACTGGCCAGAACTATCTTCAT		TGGTTG	PKTVTANRPQDFRPI
		TCCGAAGACTGTGACGGCAAATCGACCGCAAGACTTTCGTCCAATAACA		CTCGTAT	TVPSVLVRQLNAVLA
		GTCCCCTCGGTTTTGGTCAGGCAATTAAACGCTGTTCTGGCTTCTCGATT		CTTTAGA	SRLASKVNWDPRQR
		GGCTTCTAAAGTCAACTGGGATCCAAGGCAGCGCGGTTTCCTACCTACC		AGATTTC	GFLPTDGCADNATLV
		GATGGGTGTGCTGATAATGCGACGTTGGTTGATCTCATTTTGCGGGAGC		CTTCCTC	DLILREHHKRWKSCY
		ACCATAAACGGTGGAAGTCATGTTACCTTGCGACGGTGGATGTCAGCAA		GGCGAT	LATVDVSKAFDLVSH
		GGCTTTTGACTTAGTATCACACCAGGCCATTATCAAGACTTTACAGGCCT		CAAAAN	QAIIKTLQAYGAPTNF
		ATGGTGCTCCAACAAACTTTGTTAGCTTTATAGAAGAACAGTATAAGGG		AAAAAA	VSFIEEQYKGGGTSL
		CGGCGGAACCTCCCTCAATGGGGCAGGATGGAGTTCAGAGGTGTTTAT		AAAAAA	NGAGWSSEVFIPAR
		ACCCGCGCGGGGCGTTAAGCAAGGTGACCCTCTGTCTCCACTATTATTT		AAAAAA	GVKQGDPLSPLLFNLI
		AATCTTATCATTGATAGATTACTTAGGTCCTACCCCAGAGAGATTGGTGC		(SEQ ID	IDRLLRSYPREIGAKV
		CAAAGTCGGAAATACCATGACAAGCGCGGCAGCGTTCGCGGATGATCT		NO:	GNTMTSAAAFADDL
		GGTGCTATTTGCGGAAACTCCGATGGGGCAAACATTGTTGGATACCACG		1346)	VLFAETPMGQTLLDT
		CTAGGCTTCCTAGCCTCCGTGGGACTCTCCCTTAATGCTGATAAGTGTTT			TLGFLASVGLSLNAD
		CACTGTCAGTATAAAGGGGCAAGCCAAGCAGAAGTGTACTGTCGTAGA			KCFTVSIKGQAKQKC
		ACGACGGAGCTTTTGTGTAGGGGAGCGCGAGTGTCCTTCATTGAAGCG			TVVERRSFCVGEREC
		TACTGAAGAGTGGAAGTATTTAGGTATCCGGTTCACTGCGGATGGGCG			PSLKRTEEWKYLGIRF
		GGCTCAATATAGTCCAGCAGACGACCTCGGTCCGAAGCTGTTAAGATTA			TADGRAQYSPADDL
		ACAAGAGCCCCTCTGAAACCACACAGAAGTTATTTGCACATAGGACTG			GPKLLRLTRAPLKPQ
		TCCTTATCCCACAACTCTATCACCAACTTAACACTTGGGAGTGTGATGATA			QKLFAHRTVLIPQLYH
		GGCGTCCTAGGAAAATGTGACAGATTGGTACGGCAATTCGTAAGGAGA			QLTLGSVMIGVLGKC
		TGGTTAGATCTCCCACTGGATGTACCAGTTGCGTACTTTCACGCCCCCCA			DRLVRQFVRRWLDL
		CACTTGTGGGGGTCTCGGGATTCCGTCAATTAGATGGATAGCACCGATG			PLDVPVAYFHAPHTC
		CTTCGTCTGAAGCGATTGAGCAATATTAAATGGCCCCACCTCGAACAAT			GGLGIPSIRWIAPML
		CCGAGGTAGCTAGCTCTTTCATTGACGACGAATTGCAAAGGGCTCGAGA			RLKRLSNIKWPHLEQ
		TAGATTAAAGGCGGAAAATGTGCAGCGGTGTTCGCGTCCAGAGATTGA			SEVASSFIDDELQRAR
		CTCGTATTTCGCAAATAGGTTGTACATGTCTGTTGATGGTTGCGGTCTCC			DRLKAENVQRCSRPE
		GTGAAGCAGGTCATTATGGCCCGCAACATGGATGGGTGAGTCAGCCCA			IDSYFANRLYMSVDG
		CGCGCTTGCTAACAGGAAAGGAATATTTGCACGGTGTCAAATTGCGGAT			CGLREAGHYGPQHG
		AAATGCCCTACCCTCGAAGTCTCGTACGACGAGGGGAAGGCACGAATT			WVSQPTRLLTGKEYL
		GGAGAGACGGTGTCGTGCAGGATGTGATGCTCCCGAGACAACAAACCA			HGVKLRINALPSKSRT
		CATCTTGCAAAAATGCTATCGTACGCATGGGAGGCGGGTAGCTAGACA			TRGRHELERRCRAGC
		CAACAGCGTAGTAAATGCCGTCAAGCGGGGACTTGAACGGAAAGGCTG			DAPETTNHILQKCYR
		CGTTGCCCATGTCGAACCAAGTCTGCAATGCGACTCGGGCTTAAATAAA			THGRRVARHNSVVN
		CCGGACCTGGTGGGAATCCGACAGAATCACATTTATGTGATAGACGTTC			AVKRGLERKGCVAH
		AGGTTGTGACAGACGGACATTCCTTAGACCAAGCGCACCAGCGCAAGG			VEPSLQCDSGLNKPD
		TCGAAAGGTACGACAGAGCTGACATAAGATCACAAATGCGGCGATTTTT			LVGIRQNHIYVIDVQ

			CGGAGTGACAGGTGAAATCGAGTTTCATTCCGTTACACTCAACTGGAGA GGAATCTGGAGTGGTCAGTCGGTAAAACGATTGATTGCAAAAGATCTCC TCATCGCTGAAGATACCAAACTCATCAGCGTCAGAGCAGTAAACGGCGG AGTGACGTCTTTTAAATATTTCATGTATTGTGCTGGGTATACTCGAAGCT AGATGTACTAACCTCTAGTTTCTCTATACTTTTGCCTGCTACCTTGGCATT ACATCTAAAAAGGTACAAACATCGCATTGGCAAAAAGAGGTGGTTTTAG TACATAGGCGCTGTGGGACTTCATTGTCCCGATGATGCAGCGAATCGTG CATACGAGATTGTCCAGTAGTTGGTTGCTCGTATCTTTAGAAGATTTCCT TCCTCGGCGATCAAAANAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1100)	CGGTGCGT TCCCTTGG GTAAGGA ACACGAGT CTTAGTGG CCTTGACC TCCACGTG GTCCCGCT GGTAACAT CATCTCTT GATGATGG CTAACAAG GCTAATGC ACCCATTC CATCTCCT ATCTCGCA TGGAGGCC GCTATGCG TGATTACT AG (SEQ ID NO: 1224)	TGAGAA TAATGA GGTGCT AACCTCC CTGGGC CTGACC AAACCC AGAACA CATCACT GGCCAA GATGAT TTCCCGC AGCACG TTGCTTT TCTCTCT CGATAC CCGAGA ATGTTCT GCGGAA CTACGC AGTCTAT GAACAG TCACAG ACAACC TCTGATC CAAG (SEQ ID NO: 1347)	VVTDGHSLDQAHQR KVERYDRADIRSQMR RFFGVTGEIEFHSVTL NWRGIWSGQSVKRL IAKDLLIAEDTKLISVR AVNGGVTSFKYFMY CAGYTRS (SEQ ID NO: 1468)
R2	R2-2_PM	.	Petromyzon marinus	RPQTKLMTDKLKFSS QLARGLAKQRAMDG ARVGDPPITVRPTET DLCNTEGSWGRRPM KLLFVSVSTQTQNED ALWASDVAKPMASR SALKMTSIPSMTFHN SSLEKEEEMNYDFYE QIKSLVESDDSSDDFT EDDEDVEESFLDISAE EPVLGKFPIDTKGTIT VVLPSLEYICVICKQH MGKASELVAHFNIKH RDIPLVFKCAKCDKT NSNHRSIACHAPKCG GIKLTEESLPMVCECC QARFATLSGLSQHKR HAHPVTRNEERIKDG IKGTSQRGVHRSCWS LKEVEQLALLELQFQ GKKNINKIIAEALGTK TNKQVSDKRRDLSKK TGAPMSDSLHFSSRP LETLSPPPNVTTGTSS ILAQAAERLTNENSG TLEKPAMEAIKAWLN GEGQHDALVETATAL MLCPMRLVKNKGKR SKPENDIIKPRILPTRS	CGGTGCGTTCCCTTGGGTAAGGAACACGAGTCTTAGTGGCCTTGACCTC CACGTGGTCCCGCTGGTAACATCATCTCTTGATGATGGCTAACAAGGCT AATGCACCCATTCCATCTCCTATCTCGCATGGAGGCCGCTATGCGTGATT ACTAGAGGCCACAAACAAAACTTATGACTGACAAACTTAAGTTCTCAAG CCAGTTGGCACGAGGCCTGGCAAAACAACGTGCTATGGATGGCGCTCG GGTTGGCGATCCACCCATTACAGTTAGACCCACAGAAACCGATCTGTGC AACACTGAGGGTTCATGGGGACGCCGTCCTATGAAACTCTTGTTTGTCT CGGTGTCAACCCAGACACACAGAATGAAGATGCCCTCTGGGCATCTGATGT TGCTAAACCTATGGCGTCTAGGTCGGCGCTAAAAATGACGAGTATACCT TCCATGACCTTCCATAACTCGTCCTTGGAAAAGGAAGAGGAGATGAACT ACGATTTTTACGAACAGATTAAAAGTCTAGTTGAGTCGGATGACTCTTCA GATGACTTTACAGAGGATGATGAGGATGTGGAGGAGTCCTTCCTCGAC ATATCGGCTGAGGAACCCGTGTTGGGAAAGTTTCCCATTGACACCAAGG GAACTATCACGGTTGTACTGCCTAGTCTGGAGTATATATGCGTTATCTGT AAACAACACATGGGCAAAGCATCCGAACTTGTTGCACATTTTAACATCA AACACAGAGACATTCCTCTGGTGTTTAAGTGCGCTAAATGTGACAAAAC CAACTCAAACCACCGATCGATTGCCTGTCATGCCCCCAAATGTGGGGGA ATAAAGTTAACTGAGGAAAGCTTACCAATGGTCTGTGAATGTTGCCAGG CACGCTTTGCGACTCTAAGTGGCCTTTCGCAGCACAAGAGGCATGCTCA CCCAGTCACCCGTAATGAGGAAAGGATTAAAGATGGTATAAAGGGTAC CTCGCAGAGAGGGGTACACCGTAGCTGCTGGTCTTTGAAGGAAGTAGA ACAGCTGGCCCTTCTAGAGTTGCAGTTTCAGGGGAAAAAGAATATCAAT AAGATCATTGCTGAAGCGCTTGGGACTAAGACCAACAAGCAAGTCTCTG ACAAGAGACGGGACCTAAGTAAAAAGACAGGGGCCCCCATGTCAGACA GCTTACATTTTTCTTCTAGGCCTCTTGAGACATTGTCTCCCCCACCAAATG TAACAACGGGGACTTCATCCATACTCGCTCAAGCAGCTGAGCGGCTTAC GAATGAGAATTCTGGGACCCTGGAAAAGCCTGCAATGGAGGCAATAAA GGCTTGGCTTAACGGCGAGGGCCAACATGATGCCCTCGTAGAAACTGC CACAGCATTGATGCTTTGTCCGATGAGATTGGTGAAAAACAAAGGCAAA

			CGTTCAAAACCCGAGAACGACATTATTAAACCCAGGATATTACCCACAC			WMKKRAEKRGSFM
			GATCTTGGATGAAGAAGAGAGCGGAAAAACGAGGAAGCTTCATGAAG			KHQKLFFKNRSLLASL
			CACCAGAAGCTCTTCTTTAAGAACCGCTCTCTTCTTGCGTCCTTAGTCCTG			VLDGTERHECRIPNA
			GATGGCACTGAACGTCATGAATGCCGAATCCCGAACGCAGATGTATATC			DVYRFYCEKWEKVLP
			GTTTTTACTGCGAAAAATGGGAGAAGGTGTTGCCATTCAATGGCCTGGG			FNGLGQFKSSGVAN
			CCAATTTAAGTCATCAGGTGTTGCAAATAACGAATACTTTGAGCCCCTAA			NEYFEPLISVEEVQTA
			TTTCGGTGGAGGAAGTTCAGACTGCCATACGGGCCATTAAACCAACGTC			IRAIKPTSAAGPDGLT
			AGCAGCTGGGCCAGATGGCCTAACAAGGGCTGCAATCTGTGCTGCCGA			RAAICAADPEGRTLT
			CCCCGAGGGTCGGACACTGACAGCCCTATTTAATGCATGGATGATTACA			ALFNAWMITGIIPKEL
			GGAATTATTCCCAAAGAGTTGAAAAAGAATAGGACGATTCTTATTCCTA			KKNRTILIPKVMDDE
			AGGTTATGGACGATGAAAAGCTGAAAGAATTGGGGAACTGGAGACCAA			KLKELGNWRPITIGS
			TAACGATTGGTTCAATGATTCTGAGATTATTTTCCAGAATAATGACTGCA			MILRLFSRIMTARLAR
			CGTCTTGCTCGTGCTTGTCCCTTAAACCCAAGGCAGCGTGGTTTTATAGC			ACPLNPRQRGFIAAS
			GGCATCTGGCTGCTCTGAAAATCTTAAGGTGCTACAGGACCTTATGAGA			GCSENLKVLQDLMR
			CACGCTAAGAAATTGCACAGGCCGTTGGCTGTCATGTTCATCGACATAG			HAKKLHRPLAVMFID
			CGAAAGCTTTTGACTCGGTTTCGCATGCTCATATTTTATGGGTGTTAAGG			IAKAFDSVSHAHILW
			CACAAGAAAGTAGATGAACACGTGGTGGGCATCATCCAGAACGCCTAC			VLRHKKVDEHVVGII
			GATCGGTGTACGACCTCGTTCAAAAGCAATGGCGAGTCGACTCGAGAA			QNAYDRCTTSFKSNG
			ATTAGCATACGTGTTGGTGTCAAACAGGGTGACCCCATGTCACCCCTGC			ESTREISIRVGVKQGD
			TCTTCAATCTTGCCATGGACCCTTTGATATGCACCCTAGAGTCACACGGA			PMSPLLFNLAMDPLI
			GTTGGGTACTCCATTGATACCGACCACGTGACAGCTCTTGCGTTTGCTGA			CTLESHGVGYSIDTD
			TGATTTGGTGTTGGTGAGCGAATCTTGGGTTGGTATGGCCGCCAATCTA			HVTALAFADDLVLVS
			GCGATCTTGGAATCATTTTGTGGGCTATCGGGATTGGAGGTTCAGGCCA			ESWVGMAANLAILE
			GAAAGTGCCAGGGCTTCATGATAAGCCCAACCAAAGATTCATATACGGT			SFCGILSGLEVQARKC
			GAACAACTGCGACCCATGGACTATCAAAAATAAAGATGTCCATATGATC			QGFMISPTKDSYTVN
			CAACCTGATGAATCAACGAAATACCTTGGTCTAAAAATTTGCCCTTGGAC			NCDPWTIKNKDVHM
			TGGCATTATACGGTCGGATCTACATGTTCAACTAAAGACACGGATCTCG			IQPDESTKYLGLKICP
			AAAATCGATGAGGCGCCTCTGAAACCGACTCAGAAGGTCGAACTCCTCA			WTGIIRSDLHVQLKT
			ATGCCTACGCCCTACCCAGATTATTGTACCCTGCTGACCACTCGGACTGC			RISKIDEAPLKPTQKV
			AAGCAATCAACTCTCCGTGTGTTGGACCAAGAAATAATAAAGGCGGTAA			ELLNAYALPRLLYPAD
			AAGGATGGCTCCATCTTCCCGCGTCAACCTGTGACGGGCTGTTGTACGC			HSDCKQSTLRVLDQE
			CAGAGCCCGAGACGGAGGCCTTGCCATCTTGAAACTGGAAAATGCAATT			IIKAVKGWLHLPASTC
			CCTTCGGTTCAAGTTAGAAGGCTGCAACGTATTGCAAACTCCTCTGACG			DGLLYARARDGGLAI
			CTATCGCTCGAAACATTGCGTCCTCGCAGGGTGTGGAGGAAGAGTACC			LKLENAIPSVQVRRLQ
			GAAGTCTGTGGGTACGGGCAGGGGGTGACAGCGAAGCAATCCCAACG			RIANSSDAIARNIASS
			TTCTTTCTCAGGGGTTCGGAATCAAAAGAGCCCGTGTATCCGAGACCCT			QGVEEEYRSLWVRA
			GCGATTGGAGGAAACGCGAATCTCGGAGACGGTGCGAAAAGCCGGTTC			GGDSEAIPTFFLRGSE
			AAGGAAGGGGCATTGTAAACTTTGCGCAAGATAGAATCAGTAATGCAT			SKEPVYPRPCDWRKR

			GGTTGGGGCCACGGTGCGGCTTTAAACAATGCTTCTTTATCGCAGCATT ACAATTAAGGGCAAATATTTACCCAACAAGAGAAAGCATAAACAGAGG CAGAGATGGTGCCTCACGGTCCTGCAGGAAATGCTCTGCCAGGCTGGA GTCTCTCTCGCACATTCTTGGTCAATGTCCCGCAGTACAAAAATTCAAGG ATTGCGCGACGCAATAAGATCAGCGACATTCTAGCTGACGAAGCGGCG AGACTGGGCTGGTGGGTGTACAAAGAGCCACGGTTCACATCTGAAGCC GGAGAGCTAAGGAAACCTGCCCTTGTGTTTGCCAAAGGTGAGGAAGCG CTTGTTATTGATGTCACCGTCCGGTTTGAGCTCTCGAGGAAAACCTCATC AGAGGCTGCCTCGCACCAAGTTGCGTACTACACCCCCCCTTGTGATCAA GTCAAAGTGCTGACGAAGGCAAGCAATGTCACATTCTTTGGATTCCAGG TTGGGGCAAGAGGGAAAGTGGCCCCTTGAGAATAATGAGGTGCTAACC TCCCTGGGCCTGACCAAACCCAGAACACATCACTGGCCAAGATGATTTC CCGCAGCACGTTGCTTTTCTCTCTCGATACCCGAGAATGTTCTGCGGAAC TACGCAGTCTATGAACAGTCACAGACAACCTCTGATCCAAG (SEQ ID NO: 1101)				ESRRRCEKPVQGRGI VNFAQDRISNAWLG PRCGFKQCFFIAALQL RANIYPTRESINRGRD GASRSCRKCSARLESL SHILGQCPAVQKFKD CATQTEAEMVPHGP AGNALPGWSLSRTFL VNVPQYKNSRIARRN KISDILADEAARLGW WVYKEPRFTSEAGEL RKPALVFAKGEEALVI DVTVRFELSRKTSSEA ASHQVAYYTPPCDQ VKVLTKASNVTFFGF QVGARGKVAP (SEQ ID NO: 1469)
R2	R2-2_SMed	.	Schmidtea mediterranea	AGTCATAGGGTGAACTGCAATTCTGACACGATGACCGAGCTGTGTCAGT TTGCAGCTAGTCGCTAAAGACTCGATCAGTCCGCCAAGTGAGGTGGCCG GGTATCTGCAGCACTAGAGCCACTGGTATCAAGAGCAGAGATACGCGA GTGGAAGTTGAGTACGACTACCTTCACGGGGTCCTCCTGATAACCACAG TGGACTGTGGGAACTAAATGTGTGCTCAGCGTTCCCTACTTTCTCGTAG GGTAAAGGGTATGATAACCCAGAGAATATCCCATGGGAGATATCCATG GAAAAAGCACCACGTTAGACAATCCGATGGTCTAACTCGGCTCCGAGG GGCTAACTATCCCAAAGGGCTTAAAGAAAAGAACTGGTTACAATCAAAA ATCTATTTGAAGAGTCAGGTGCCACCGCGCCTGCACCAGTTCCCCTAGA AGTTGCAGTTGAGGTACACCAGTCATCGAGTGTTCCCGAGATAACCGAC GAATCTACGACTACTCAGGAAGGAAGCTATTCAGAACCACCGATACACC GATGTGAGAACTGTGGAAGAGAATTCAGAACAAGAGCAGGAGTTCAAC AGCACCGGAGAAAAGCTCACACCAACGAGTTTATGGAGGAGAAAGAGA AGGCAGCTCCAACCAAGAAGCTTCGATGGACAGACGAAGAAAAAGAGA TCCTCATCGAAAGCGAGATAAAGATCATCAAGGAAGGATCCCTTAAAGA ACAACACGAAATAAACAAGATTCTAGCTTCAAGAATGCCTGGAAGATCA CAAGACGGAATAGCCAAAATTCGCCAGAAACAAGAGCACAAGGCTGAA ATACAGAGAAGGCTACATGGGACCGTCACTACCAACGAAACCAGAGGA AACAGAACCAGCGAAATTACCGAACCAATAAGAAGCTTACCAATAAACA CCAAAACCTGGAGCGAAGACGAAATGAAAAGAATGCTAGCCGAAGAA GTGAAGTTGAGAACCAAGAATGAAAAGGATATCAACAAGAAGCTAGCA	AGTCATAG GGTGAACT GCAATTCT GACACGAT GACCGAGC TGTGTCAG TTTGCAGC TAGTCGCT AAAGACTC GATCAGTC CGCCAAGT GAGGTGG CCGGGTAT CTGCAGCA CTAGAGCC ACTGGTAT CAAGAGCA GAGATACG CGAGTGG AAGTTGAG TACGACTA	TAGAAG GGAAAC AAAGGA AAAACG AAATGA CTGGAA ACTATG AAGGAT ATAGCT GAAAGC CGCAAG GAAGGC TAAGTC CTGAAA CCGATCT ACATCTT CGATCC CAAGAG GAACTG TGGGTT AAGCTT	

	GAAATATTCCCGAACAGAACAATGGGGTCCATAAAGAGCAAAAGGACG	CCTTCACG	GAGCCG	PGLKITNQTRKREKKI
	AAAGATAAGGACTACCAGGATTTGGTAAAGCTAACAATGCAAACAATCA	GGGTCCTC	ACGGAA	SKRETRRQEYAEIQKL
	GCGAGAACCCAGACAATGAAACAGACTTCAATACCAGCAACACGGAAA	CTGATAAC	AAAGCG	YKKNISSAAEKAINGK
	ACAACAGCACTGATGCAGAGAAAGAAGTGAAGAACTACCTCAATATGC	CACAGTGG	AATGCA	WSIKPEEEYHNNKDL
	TACTACTGACCATCAACGAGGAAGAATGGTTGACATCCACTCTGAAGGA	ACTGTGGG	TGTTAG	IKAWKPILEAPPFSDC
	AGCAGCAACACTAGCACTACAAGGAAAGAAAACTGAAGCATCTGAAAA	AACTAAAT	ACGACG	RPIENIKEMDYALMEI
	GCTTAACGAGTATGCAAGCAAAACGCTGTTCCCTGGATTGAAAATAACT	GTGTGCTC	AGGTAC	STAEIFLAIRAMGKTA
	AATCAGACGAGGAAACGAGAAAAGAAATATCAAAAAGAGAAACCAG	AGCGTTCC	AGTCAC	PGPDGIKYSKLKKNIQ
	AAGGCAAGAATACGCTGAAATACAGAAACTCTACAAAAAGAATATTTCA	CTACTTTCT	CTCCTCG	SMAILFNTCLLTSFLP
	AGCGCAGCCGAAAAAGCAATCAATGGGAAATGGTCCATAAAACCCGAA	CGTAGGGT	TGGTATT	LPLKIARTILIPKQENP
	GAAGAATACCACAACAATAAGGATCTCATTAAAGCATGGAAACCAATAC	AAAGGGT	TGGCGG	GILDYRPLTIASVVTR
	TAGAGGCACCTCCATTCAGTGACTGTAGGCCCATCGAAAACATCAAGGA	ATGATAAC	GCAATG	VFHSILAKKLDNNAQ
	AATGGACTACGCTTTAATGGAGATCAGCACGGCAGAAATCTTCCTCGCA	CCAGAGAA	CTCACTA	LSQRQKGFRKCDGV
	ATAAGAGCCATGGGGAAGACTGCACCAGGACCTGATGGCATTAAATAT	TATCCCAT	AATTAA	AENIVILETILTNSRSE
	TCAAAGCTCAAGAAAAATATCCAATCAATGGCAATATTATTCAACACATG	GGGAGAT	CTGTGA	KRPLCMAFVDLRKAF
	CCTACTAACGAGCTTCCTGCCACTCCCATTGAAGATAGCAAGGACAATCC	ATCCATGG	GTAGCT	DSVGHESIIRGAKRV
	TGATACCTAAACACAAGAGAATCCAGGTATCCTTGACTATCGACCACTAAC	AAAAAGCA	GAGAAC	GVPPMLLEYISSSYQ
	AATAGCCTCAGTGGTGACAAGAGTGTTCCACAGTATCCTTGCGAAGAAG	CCACGTTA	TGTATGT	NASTNLFGEILNSRR
	CTCGACAACAATGCCCAATTAAGTCAACGACAAAAAGGATTTCGAAAAT	GACAATCC	GTATCAT	GVRQGDPLSPILFNF
	GTGATGGAGTTGCGGAAAATATAGTAATACTCGAAACTATATTAACCAA	GATGGTCT	GAAAAA	VIDEALENLNRNIGYL
	CAGCAGAAGTGAAAAGAGACCGCTCTGTATGGCCTTCGTTGACTTAAGA	AACTCGGC	AAAA	LKEEKVSCLAFADDIV
	AAAGCATTCGATTCTGTGGGACATGAGTCTATCATCAGAGGAGCAAAAA	TCCGAGGG	(SEQ ID	LIAETKGGLENHIEKLL
	GAGTTGGAGTGCCACCAATGTTGCTCGAGTACATTTCKTCAAGTTACCA	GCTAACTA	NO:	EKLNGAGLELNASKC
	GAATGCGTCTACTAACTTGTTCGGCGAAATACTCAACTCGAGAAGAGGA	TCCCAAAG	1348)	ATLMVMKNGKEKST
	GTCAGGCAAGGCGACCCTCTGAGCCCTATTCTCTTCAATTTTGTTATCGA	GGCTTAA		YISTKAIKIKENDIPTM
	TGAAGCTCTAGAAAACCTCAACAGGAATATTGGATATCTACTGAAGGAA	(SEQ ID		KATETYKYLGLQMGF
	GAAAAAGTGAGTTGCCTAGCTTTCGCGGACGACATAGTCCTGATAGCTG	NO: 1225)		KAREQNANEVITEGL
	AGACAAAAGGAGGCCTAGAGAATCATATCGAGAAACTATTAGAGAAGC			ENITRAPLKPQQRIHI
	TGAATGGGGCCGGTCTCGAGTTGAACGCCTCGAAATGCGCAACACTGA			LRDFLIPRLIHKLVLGR
	TGGTGATGAAAAACGGAAAGGAAAAATCAACGTATATATCAACAAAAG			VAKKSLKRIDQNIRKK
	CAATCAAAATCAAAGAAAATGACATTCCAACAATGAAAGCCACAGAAAC			VRNWLHLPKDTTAA
	GTACAAATATCTCGGATTGCAAATGGGTTTCAAAGCTAGAGAACAGAAT			FIHADAGDGGLGVP
	GCTAATGAGGTTATTACAGAAGGACTGGAGAATATAACAAGAGCACCA			ALEHTIPLLKRERITNL
	CTGAAGCCCCAGCAGAGGATACATATCCTACGAGACTTCCTTATACCAA			RKSNDPVTKECLRME
	GATTAATACACACAAATTGGTATTAGGAAGAGTGGCCAAGAAGTCATTAAA			YTKQVLGKWVSRPTKI
	AAGAATCGACCAGAATATAAGAAAGAAAGTGAGGAATTGGCTACATCT			GETLATNKSQLKEAF
	CCCTAAAGACACGACAGCAGCATTCATACACGCTGATGCAGGAGATGG			RKQMLITLDGKGLKD
	AGGGCTTGGAGTACCAGCGTTAGAACACACACAATTCCTCTACTGAAAAGA			HHETPTIHKWIRRGE

			GAAAGAATAACTAATCTAAGAAAATCCAATGATCCAGTTACCAAAGAAT				NMTGKQFITAVKIRG
			GCCTGAGAATGGAGTACACCAAACAAGTACTGGGAAAATGGAGTAGAC				NLVATKSRNSRGRPE
			CAACTAAAATTGGAGAAACTCTGGCTACCAACAAAAGCCAACTAAAAGA				QEKLCEAQCGRPDSL
			AGCATTCAGAAAACAGATGTTAATAACGCTAGATGGAAAAGGGCTAAA				GHILQGCWRTHGM
			AGATCACCACGAAACGCCCACTATCCACAAATGGATCAGAAGAGGAGA				RVERHNNICRRIKAI
			GAACATGACCGGCAAACAGTTTATCACAGCAGTTAAAATAAGAGGAAA				MKGKESEVVEEPRLQ
			CCTTGTGGCAACTAAGTCAAGAAATAGCAGAGGGAGACCCGAACAAGA				TNEGLRKPDLLICHKG
			GAAACTATGTGAAGCTCAATGTGGACGACCAGATAGCCTGGGACACAT				KIIICDAQVVADSSNC
			ATTACAAGGTTGCTGGAGAACACATGGTATGCGAGTGGAAAGGCATAA				SLESENQRKIDYYKKD
			CAACATTTGTCGCAGAATTAAAGCAATAATGAAAGGAAAGGAAAGCGA				SVVSEARKLIGRVDE
			AGTAGTCGAAGAACCAAGACTACAAACGAATGAAGGTCTCAGAAAGCC				DIIIMAVTFNWRGAI
			TGACTTACTGATCTGTCACAAAGGTAAAATAATAATATGCGATGCACAA				SKTSIRDLDMLLDIKS
			GTAGTGGCAGATAGCTCGAACTGCAGTCTTGAAAGCGAAAACCAAAGA				KEVIKMSRKIIRDNSI
			AAGATAGATTACTACAAGAAAGATTCAGTAGTATCAGAAGCAAGAAAA				MVEMHRNRTEKRR
			CTTATCGGACGTGTCGACGAAGATATAATTATAATGGCAGTGACCTTTA				(SEQ ID NO: 1470)
			ACTGGAGAGGAGCCATCTCAAAAACATCAATAAGAGATTTGGACATGCT				
			TCTAGATATAAAATCAAAAGAAGTAATTAAAATGTCAAGGAAAATAATC				
			AGAGATAATAGCATCATGGTGGAGATGCACAGAAACCGGACTGAGAAA				
			AGGAGATAGAAGGGAAACAAAGGAAAAACGAAATGACTGGAAACTAT				
			GAAGGATATAGCTGAAAGCCGCAAGGAAGGCTAAGTCCTGAAACCGAT				
			CTACATCTTCGATCCCAAGAGGAACTGTGGGTTAAGCTTGAGCCGACGG				
			AAAAAGCGAATGCATGTTAGACGACGAGGTACAGTCACCTCCTCGTGGT				
			ATTTGGCGGGCAATGCTCACTAAATTAACTGTGAGTAGCTGAGAACTGT				
			ATGTGTATCATGAAAAAAAAAA (SEQ ID NO: 1102)	TGGAAGAC	TGATGC		MKSRSFRRIGDCAAG
R2	R2-2_TCas	.	Triboliu m castane um	GAAAAGTTATGCTGAGTACGTCGTGAGAGTCGGTAACTGTCCCAGGAT	CCCGCCCA	TCCTTTG		SSRRGVRLTGKAGRE
				GGTCTGGGATAGGCTAAACCTCAGCAGGGGAAAGTTGTAGGGGCCTGC	TGAGGCTT	GTTTTAC		GRFAASPHLSPRYLA
				CACCCCTACACTTTTATAGATATGGCATTCGATACCTCAAATAGAGCCTC	GGAGAGT	CATCTGT		GSVSGNVPSVPPGP
				GGACTTGGAGGAGCATGGTTCCCCTCCTCCTCGTACTAGACCTGGAACC	GTGATCCT	GGGGGC		GLGAGAPAFAAGRN
				AACGGTCTTGACAACCCCATTGGACCTACGGGAGCGGACCATGCTCATG	GATCACAC	ATCGGT		ADGGPAQNPCPYCA
				GACATGGATTCCGAAGACGAAGCGGGGGAACACGGACCCCCCCGCCGAT	TTGAAAAG	CCTCAC		RSFTTANGRGLHIRR
				AATGCTCACTTAACGTCAGGCGAACCCATCGAAATCATCTTGATGTTACC	TTATGCTG	GGTTTC		AHPDEANNAIDIERI
				CTTTCAAAGCAGGTCATGCGGCATATGTCTCAATGCCGGAAAGGGTGGC	AGTACGTC	CTCGGG		HARWSHEETAMMA
				TCCCGGGCGTAACGGACATCTATCGCTGATGGAASCTAGCTCGCTCTTG	GTGAGAGT	TTTCCTA		RLEAGAIQRGGVRF
				AAGAACGGCGGACGGGGACTTGGAATCGTGGTGTGGTTCTGATGTAAG	CGGTAACT	TTGTTTT		MNQFLVPRMPGRTL
				TCCTGAAATTATGGCGTGATGGCCCGCCCGCCCGACCGGAGGGACTTA	GTCCCAGG	TCCTAAA		EAVKSKRRDATYKAL
				GAACCCCCTTCCGCGAGGGTCCTGTCTGTAGGTCCWCCATCTCCGTAAA	ATGGTCTG	CCCGAC		VQRFLQAPQINLPEL
				ACGAGTTGGAGGAAACCGCAGACGGGGATGAAGAGTCGTTCGTTCCGT	GGATAGG	AAGGAG		RDGDAPRQPDPQQE

		CGCATCGGGGATTGCGCAGCTGGGTCAAGTCGGCGGGGGTGTTCGGCTA	CTAAACCT	CCCTTTG	NPPEPPSFDGAIRGA
		ACGGGCAAGGCTGGGCGTGAGGGGAGGTTTGCAGCTTCCCCTCACTTA	CAGCAGG	GCCCTCC	VADLVGGVDWQRL
		AGCCCACGTTATCTTGCAGGCTCGGTTTCAGGTAATGTTCCTTCAGTTCC	GGAAAGTT	TCCTTAA	GFQGDRLCNIARRAC
		CCCCGGCCCCGGCCTGGGGGGCCGGGGCTCCTGCATTTGCTGCTGGTCG	GTAGGGG	ACACCTC	DGGDVSGQLLGWLR
		GAATGCCGATGGGGGGCCCCGCCCAAAATCCGTGCCCGTATTGTGCACG	CCTGCCAC	TCCTTCA	DVFPVKRVSTRGDQS
		GTCGTTTACTACCGCCAATGGCCGCGGGTTGCACATACGGAGGGCGCAT	CCCTACAC	TCCTSTT	DLDVDGALVSRRTAR
		CCAGACGAGGCTAATAACGCCATAGACATCGAACGCATTCATGCCCGCT	TTTTATAG	AGTCCA	TREYARVQELYRKEP
		GGTCACACGAGGAGACAGCCATGATGGCCAGGCTTGAGGCCGGTGCCA	ATATGGCA	TTCCGG	KACLARILGDRREGA
		TCCAGCGAGGGGGTGTAAGGTTTATGAACCAGTTTCTGGTTCCGCGCAT	TTCGATAC	CTAAAA	NRAPNRDPAFIDFW
		GCCGGGCAGGACTCTGGAAGCGGTGAAAAGTAAGCGCAGAGATGCGA	CTCAAATA	TGATGA	RGVFSEASAEVEGW
		CTTACAAAGCACTGGTCCAGAGGTTTCTGCAGGCGCCGCAAATAAACCT	GAGCCTCG	AGACCG	AEEVSDHGELARRV
		CCCCGAACTGCGTGATGGTGATGCGCCTCGTCAGCCTGATCCTCAACAG	GACTTGGA	AGGAGT	WDPISVEEVGRSRVR
		GAAAACCCTCCCGAGCCCCCTTCTTTCGATGGGGCGATCAGGGGCGCTG	GGAGCAT	GTCACTC	NGAAPGPDGIAVSV
		TCGCCGACCTTGTCGGAGAGGGGTCGACTGGCAGCGCCTCGGGTTTCAGG	GGTTCCCC	TCTTGG	WNKLPPEAAALLFNV
		GAGACCGTCTCTGTAACATCGCGAGACGGGCTTGCGATGGGGGGGGATG	TCCTCCTC	CGGGGT	LLLGRCLPAELTRTRT
		TTTCGGGCCAGTTGCTGGCTGCGGGATGTCTTCCCCGTCAAGAG	GTACTAGA	TAACCC	VFIPKTDAPRTPADYR
		GGTTAGTACTAGAGGGGGATCAGAGCGACCTAGATGTCGACGGCGCATT	CCTGGAAC	GTCCAA	PISIASVVARHFHRVL
		GGTGAGCAGGCGCACGGCGAGGACAAGGGAGTACGCTCGGGTGCAGG	CAACGGTC	GTGTAA	SARVQRIPDLFTKYQ
		AGCTGTACCGGAAGGAACCTAAGGCGTGTTTGGCTCGAATCCTTGGTGA	TTGACAAC	ATGTGA	RGFLSGVDGIADNLS
		TCGTCGAGAGGGGGCCAATCGGGCTCCTAATCGCGATCCCGCATTCATC	CCCATTGG	CCTCGCC	VLDTMLTMSRRCCK
		GATTTTTGGCGAGGCGTCTTTTCGGAGGCGTCCGCTGAGGTCGAGGGA	ACCTACGG	ATTCGG	HLHLAALDVSKAFDT
		TGGGCGGAGGAGGTCTCCGACCATGGTGAGCTGGCACGCCGGGTTTGG	GAGCGGA	GCTCTG	VSHFAIVRACRSIFGS
		GACCCCATCTCTGTCGAGGAGGTGGGAAGGTCTCGGGTCCGGAATGGT	CCATGCTC	ATA	AETVLEEGGRRHFVQ
		GCAGCGCCGGGGCCCGACGGAATAGCGGTTTCCGTGTGGAACAAGCTT	ATGGACAT	(SEQ ID	VRXGVRQXDPLSPLL
		CCGCCGGAGGCCGCTGCCCTTCTTTTCAATGTGCTGCTGTTGGGGAGAT	GGATTCCG	NO:	FNLVLDRALKRLSTDV
		GTCTGCCTGCAGAACTCACCCGCACGCGAACGGTCTTTATTCCAAAGAC	AAGACGA	1349)	GFRLTDATKVTALAF
		CGACGCACCGCGAACACCTGCAGACTATCGGCCGATTTCGATCGCGTCG	AGCGGGG		ADDVVLCATTARGL
		GTAGTCGCACGGCACTTCCACCGTGTCTTATCCGCACGTGTTCAACGCAT	GAACACG		QTNLDVLEAELRLAG
		TCCCGACTTGTTCACAAAGTATCAAAGGGGCTTCTTAAGCGGCGTGGAC	GACCCCCC		LLLNPNKCQALSLVAS
		GGGATCGCAGATAATCTCTCGGTCCTTGACACAATGCTGACGATGTCCC	GCCGATAA		GRDHKVKLVTKPTFK
		GGAGATGCTGCAAGCATTTGCACTTGGCCGCGTTAGACGTGTCAAAAGC	TGCTCACT		VGQNTIHQVDASSI
		ATTTGACACTGTCTCTCACTTTGCCATAGTGCGTGCTTGCCGGTCCATATT	TAACGTCA		WKYLGIQFRGSGMC
		TGGGAGTGCGGAGACCGTGTTGGAGGAGGGCGGGAGGCACTTCG	GGCGAACC		GCGSEGVAAGLKRIT
		TGCAAGTGCGGYGTGGGGTGAGGCAGKGAGACCCACTCTCCTCTCCT	CATCGAAA		CAPLKPQQRMHLLR
		CTTTAACCTGGTCTTGGACCGGGCTCTGAAGAGGCTGTCCACCGATGTG	TCATCTTG		VFFLPKFYHAWTFGR
		GGCTTTCGTCTAACGGATGCGACCAAAGTTACGGCACTCGCTTTTGCGG	ATGTTACC		LNAGVLRRLDVVVRT
		ACGACGTCGTTCTGTGCGCGACCACCGCTAGAGGACTCCAGACCAATCT	CTTTCAAA		SVRTWLRLPHDIPVG
		GGATGTTCTCGAAGCGGAGCTTCGCCTAGCCGGTCTCTTGTTGAACCCC	GCAGGTCA		YFHAPTKSGGLGIPQ

			AACAAGTGTCAGGCCTTATCCTTGGTGGCTTCCGGACGGGACCACAAAG	TGCGGCAT		LSRFIPFLRLKRFDRLG
			TGAAGCTCGTGACAAAAACCCACCTTTAAGGTGGGACAAAACACGATTCA	ATGTCTCA		RSAVDYVRECAFTDI
			CCAAGTGGATGCGTCCAGTATTTGGAAGTACCTGGGAATCCAGTTTAGG	ATGCCGGA		ADRKIRWCRERLSGI
			GGTAGTGGGATGTGCGGATGCGGTTCGGAGGGTGTGGCTGCTGGGTT	AAGGGTG		VDQVAGGRDALDAY
			GAAGCGGATCACTTGTGCTCCGCTGAAGCCTCAGCAGAGGATGCACCT	GCTCCCGG		WTAQLHQSVDGRAL
			GCTGCGAGTCTTTTTCTTGCCCAAGTTTTATCATGCTTGGACATTTGGGA	GCGTAACG		RESASVASSTQWLRC
			GACTCAATGCGGGGTGTACTGCGCCGTCTGGACGTTGTGGTCCGAACTTC	GACATCTA		STRAIPASDWLHYTA
			GGTTCGGACCTGGCTACGCCTCCCCCACGATATACCGGTGGGATACTTT	TCGCTGAT		VHIGALPSRVRTSRG
			CACGCTCCTACGAAGTCCGGGGGGCTGGGGATCCCGCAGTTGTCACGA	GGAASCTA		RRGGQDVSCRGGCL
			TTCATCCCGTTCTTGCGATTAAAGCGGTTCGACCGGTTGGGGCGTTCCG	GCTCGCTC		LDETPAHCIQVCHRT
			CGGTTGACTATGTCCGAGAGTGTGCATTTACGGACATTGCGGATCGAAA	TTGAAGAA		HGGRVLRHDAIAKRI
			GATCCGATGGTGCCGTGAGCGCCTCTCGGGCATAGTGGACCAAGTCGC	CGGCGGA		SADLMELGWIVTREV
			GGGTGGGCGGGACGCGCTTGATGCATATTGGACAGCGCAGCTTCATCA	CGGGGACT		SFRTTAGVFRPDMV
			GTCGGTGGACGGAAGAGCACTAAGGGAGTCCGCTTCGGTGGCTTCCAG	TGGAATCG		AVKEGVTVILDVQIVS
			CACGCAGTGGCTCCGCTGTAGCACACGTGCAATTCCGGCCTCTGACTGG	TGGTGTGG		PAPTLDEAHRRKVAK
			CTGCATTACACGGCTGTCCATATTGGCGCCCTACCATCGCGTGTTCGCAC	TTCTGATG		YRDRADLARYLAEAA
			ATCGCGTGGGAGGCGGGGCGGCCAGGATGTCTCTTGTCGGGGTGGAT	TAAGTCCT		VARGRAPPANIRFAS
			GTCTTCTGGACGAAACTCCGGCACACTGCATCCAGGTTTGCCACCGCAC	GAAATTAT		ATISWRGVWSAESV
			CCACGGGGGGCGGGTGTTGCGACACGACGCCATCGCCAAGCGCATTTC	GGCGTGAT		GSLRELGLSARHFDR
			TGCCGACCTGATGGAACTTGGATGGATCGTAACGAGGGAGGTTTCCTTC	GGCCCGCC		YTTMALCGSWRNW
			CGGACGACTGCGGGTGTTTTCCGGCCGGACATGGTCGCGGTAAAGGAG	CGCCCGAC		VRFNASTASRMGRG
			GGTGTTACCGTGATCTTGGACGTGCAGATTGTTTCACCGGCCCCCGACAC	CGGAGGG		RGDASPRRHENQQ
			TGGATGAGGCCCATCGGCGAAAGGTTGCCAAGTATCGGGACAGGGCTG	ACTTAGAA		(SEQ ID NO: 1471)
			ACCTGGCCAGGTACTTGGCTGAGGCTGCGGTTGCTCGTGGGCGTGCGC	CCCCCTTC		
			CGCCTGCGAATATACGATTTGCCAGCGCGACCATCTCTTGGCGAGGCGT	CGCGAGG		
			TTGGAGTGCTGAGTCGGTGGGGTCCCTGCGCGAGCTGGGGCTCTCCGC	GTCCTGTC		
			TCGGCACTTCGATCGCTATACAACGATGGCACTCTGCGGTTCATGGAGA	TGTAGGTC		
			AATTGGGTGCGGTTCAATGCGTCCACGGCCTCTCGGATGGGTCGTGGAC	CWCCATCT		
			GCGGCGACGCGAGTCCTCGGAGGCACGAAAATCAGCAGTGATGCTCCT	CCGTAAAA		
			TTGGTTTTACCATCTGTGGGGGCATCGGTCCTCACGGTTTCCTCGGGTTT	CGAGTTGG		
			CCTATTGTTTTTCCTAAACCCGACAAGGAGCCCTTTGGCCCTCCTCCTTAA	AGGAAACC		
			ACACCTCTCCTTCATCCTSTTAGTCCATTCCGGCTAAAATGATGAAGACC	GCAGACG		
			GAGGAGTGTCACTCTCTTGGCGGGGTTAACCCGTCCAAGTGTAAATGTG	GGG (SEQ		
			ACCTCGCCATTCGGGCTCTGATA (SEQ ID NO: 1103)	ID NO:		
				1226)		
R2	R2-7_MR	.	Megachile	TCTAGTTAGCAAGCGGCCCCCTCTAATCGGGTCCCGCCACGAGTACCTTC	TCTAGTTA	TAAATG	SGPATSTFGETKSRLC
				GGGGAAACCAAGAGCCGCCTCTGCGAACCTACCTCCGCATTGGGTTGCC	GCAAGCG	TCGAAC	EPTSALGCRPGAVVI
				GGCCCGGTGCGGTTGTAATTCAGTGGGCCCAAATTCACAAGGAAAAGA	GCCCCCTC	CGAATTT	QWAQIHKEKRKRIVG

	rotunda ta	GAAAAAGAATTGTCGGTTGGCCCTTAGGGCATCTCGGATCACCGACCTC	TAA (SEQ ID NO: 1227)	TGGGTA		WPLGHLGSPTSLKLR
		ACTGAAACTTCGACACCCGCGCCTGCAAGCGAAGCGCATAGTTCCAGTG		ACGTGC		HPRLQAKRIVPVLAEL
		CTTGCGGAGTTGATGCAATGTTTGTGTGCTCGACATGTCAGCGGCCGTT		ACCCCA		MQCLCARHVSGRSP
		CGCCACAAAAATCGGCTTGGGCGTTCACGTTAGAAGAGCGCACGAGTG		CCATCCT		QKSAWAFTLEERTSA
		CTTGCGGAGTTGATGCAATGTTTGTGTGCTCGACATGTCAAAGGTCGTT		TAATCG		CGVDAMFVCSTCQR
		CGCCACAAAAATCGGCTTGGGCGTTCACGTTAGAAGAGCGCACGTGGA		GCAGCA		SFATKIGLGVHVRRA
		AGTCGCAAACGCTGCCATATCAGTGGAGCGCGTTAAGGATCGTTGGTC		CGCAAT		HVEVANAAISVERVK
		GGAGGAGGAGCGAAGAATAATGGCTGCCGTCGAGGTTCGAGGAGTAC		AAAGCC		DRWSEEERRIMAAV
		TCTCTGGGGCTCGTTTTATAAACGAGTACATTATGAGTCATCTCCAGACC		GTGGGC		EVRGVLSGARFINEYI
		AGCCGGACCCTCGAATCTGTCAAAGGAACTCGGAAAAAACCCCAAATACA		AGTGGT		MSHLQTSRTLESVKG
		AGGAACTTGTCGCAACCTTACTGGAAGAGGCACGAACCAGTGTGCGGG		TTTAGTG		TRKNPKYKELVATLLE
		AGGAGAGTCCTCGTAGTGCGGTCAACGATAGCGCCACCCAACCATCAG		GGTAGT		EARTSVREESPRSAV
		GCCCCTCTGATACACGAAGCCTGCGCACCGAACACACCTGTTCACTGAAAG		CATTAG		NDSATQPSGPSDTRS
		CACCGAGCCTTTCGAACATCGGATCCGAGAGTTGATCGGGGATCTTGAG		GAGTCC		LRTEHLFTESTEPFEH
		GGAGTGACCGACTTCAGAGCTGAGTTGCTGGTTTCCATTGCGGAACAGC		CACAGT		RIRELIGDLEGVTDFR
		AGCTTCAAGGCGATGAGGTGGCGGAATCCCTGACTCGGTGGCTTGGCG		ACCCAG		AELLVSIAEQQLQGD
		AGGTTTTTAAACCCGAGAACCAACAACAACAGGTTCAGCGTAAGCGTCG		CGAACA		EVAESLTRWLGEVFK
		GCGACAAAGAAAGGCCCCTGTCAGTGGGCAGCTGCCAAAGTGGCGCGA		TCTTAGT		PENQQQQVQRKRRR
		GAGGCGAAGGGATTATGCAGCGATGCAGACTCTCTTCCATCGGAACCCC		GGGTCT		QRKAPVSGQLPKWR
		AGTCTTGCTGCTGGTCGTGTGCTGGACGGCAAGAACGAGTCCCGGCCCC		GCGTAA		ERRRDYAAMQTLFH
		CGGATCTGCCAGAAATGACAGCGTTTTGGGAACCGATACTCACCGAGCA		ACGCAT		RNPSLAAGRVLDGK
		ATCTGCCGAACATCGGGCAGTTGGGCCGGCTTCCGAGAAATCGGAGTT		TTCCACT		NESRPPDLPEMTAF
		GTGCTCAGTATGGGGTCCAGTAGAGAAAGAAGAAGAACTTCTCTCCTCCGTG		GCCTATC		WEPILTEQSAEHRAV
		CCGCCGTTGGACACGGCTGTTGGCCCTGACGGTGTCACCGCCAGACAAT		CTCCGG		GPASEKSELCSVWGP
		GGAGAGCGGTTTTACCGGCCGTTAGAGCTCTCTTGTATAACATCATCCTC		GAAAAA		VEKEELLSSVPPLDTA
		AAGAGAGGCTCTTTCCCCGCCTCGATGCTGGAAAGCAGAACCGTGTTCC		AAAAAA		VGPDGVTARQWRA
		TGCCGAAGAAGCAACATAGTGTCAATCCTGCTGACTTTCGACCCATCAG		AAAAAA		VLPAVRALLYNIILKR
		CATAGCGTCGGTCGTCGTCCGTCAATTACACAAGATCTTGGCCATGCGC		AAAAAA		GSFPASMLESRTVFL
		CTTCGCCGTACGAACCTCGTTGACGAAAGGCAACGATGCATGGATGAC		AAAAAA		PKKQHSVNPADFRPI
		GGATGTGCCGAGAACATCACGGTGTTGGCCTCGCTGCTGGACGATGCG		AAAAAA		SIASVVVRQLHKILA
		AGGCATGGTCTGAAGGAGCTGCATCTGGTTTCGTTGGATTGTGCTAAGG		AAAAAA		MRLRRTNLVDERQR
		CGTTCGACAGCGTTAGTCACCATGCGATCGACGCAACACTAAAAGAATG		AAAA		CMDDGCAENITVLAS
		CGGTCTGCCCGCGGGATTTGTGCAATACATCTCCCGCACATATTCCGACA		(SEQ ID		LLDDARHGLKELHLV
		GTTCCACCAGGCTCGAGGTCGGAAGGAATCGGTCGGAGCCAATAAAGA		NO:		SLDCAKAFDSVSHHA
		CCAACCGGGGAGTGCGGCAGGGTGATCCCCTGTCAACCCTCATCTTCTG		1350)		IDATLKECGLPAGFV
		TCTGTGCTTCGATAGAGTCGCGCGTACCCTCTCACCACACATTGGGTATG				QYISRTYSDSSTRLEV
		ACCTTAATAACACCCGCATCAGTACTCTGCTTTACGCTGATGATGCTTTC				GRNRSEPIKTNRGVR
		CTTGTCTCGACGACTGCTCCGGGGATGAACATCCTGCTCCGGAGTGTGG				QGDPLSTLIFCLCFDR

			AGGAAAGTGCCGGTGAGGTGGGGCTTTCTTTTAACACCTCGAAATGCTC				VARTLSPHIGYDLNN
			AGCCCTCTCCTTAATTCCATCTGGAAAGGAAAAGAAATGAAGGTCGGA				TRISTLLYADDAFLVS
			ACGACTCCGACGTTCAAAACGTCGCAGGGTTTTATTACCCAGATAACGC				TTAPGMNILLRSVEE
			CCTCGCAAGAGTGGAGGTATCTAGGAGTCGACTTCCAATACTCCGGACC				SAGEVGLSFNTSKCS
			GAAAAAAGCAAGTCGGAGTCTCAAAATCGAGCTGGAACGCATTTCGAA				ALSLIPSGKEKKMKV
			GGCCCCCCTCAAGCCGCAACAGCGGCTTTTAATCCTCAGAGTCTACCTGC				GTTPTFKTSQGFITQI
			TGCCGCGATATTACCATCATCTCGTACTCTCCAGGACAACCCTTGGTCAC				TPSQEWRYLGVDFQ
			CTTAGGGGCCTGGACCTGCAAGTGCGCGCGGCTGTTCGTAGGTGGCTG				YSGPKKASRSLKIELE
			TCTCTGCCGCGCGACATCCCCATTGCCTACTTCCACACAACAGCGAAGG				RISKAPLKPQQRLLILR
			AAGGGGGTCTGGGTCTTCCGGCGTTTGAGACGTCCATTCCATGCCTGAT				VYLLPRYYHHLVLSRT
			GTTAGCCCGCTTGAGGTCGATGGAAACCTCGACTTGCAAGGCGGCTAG				TLGHLRGLDLQVRAA
			GGCGGCGGTTCAAGGCTTCTGGGTGCAGAAACGCATCCATTGGGCTAC				VRRWLSLPRDIPIAYF
			GGCCGCTCTCACTAAGAACGGAGAAGCCCTAACATGCAAGGCGGACGT				HTTAKEGGLGLPAFE
			CGATCGGTGGTGGGCCAGTAGGCTCCACAAGTCAGTGGATGGTCGCGA				TSIPCLMLARLRSME
			ACTGCGTGAGTGCTCAGGCGTTGGCTCGTCCAGCACCTGGGTAAATTCT				TSTCKAARAAVQGF
			GCGCTTAACATCACCGGCAGAGATTACGTGCAGTATCATCACGTCCGCA				WVQKRIHWATAALT
			TCAATTCTCTCCCGACTCGCATTAGGACATCTCGGGGTGTGCGAAGAGA				KNGEALTCKADVDR
			GGGGATGGAGGTCACGTGCCGCGCTGGGTGTCAGGTAACTGAGACCGC				WWASRLHKSVDGRE
			GGCGCATGTCATCCAGTCTTGCCACCGCACACATGGAGGTCGGATTCTT				LRECSGVGSSSTWVN
			CGACACAATGCCGTGTGCAAAGTGCTGGCGTCCGGCCTTCGCGACAAG				SALNITGRDYVQYHH
			GGATGGGAGGTGCGTGAAGAGCCAAAACTGCGAACGCGACAGGGTCT				VRINSLPTRIRTSRGV
			GAGGAAGCCGGACATAGTGGCGATAAAGGACGGCGTCGCAAGGGTGA				RREGMEVTCRAGCQ
			TAGACGCACAGGTCGTGAGTGGATCTGGACCGCTTGACGAGGCGCATG				VTETAAHVIQSCHRT
			AGACCAAGCGCAAGTATTATTCCGACAATGGCGACGTAACTGCTGCGAT				HGGRILRHNAVCKVL
			TGCTCGTGAGTGCAACATTGCTCCATCGAATGTTGCATACTCGTCATGCA				ASGLRDKGWEVREE
			CTATCTCTTGGAGGGGTGTGTGGTCGCCTCGGTCCGCTGCAGACCTACT				PKLRTRQGLRKPDIV
			ACAGGTGGGTCTGTCCAAAAAGCTCTTGGGTTTCATCACCCTAAGGGTC				AIKDGVARVIDAQVV
			TTGAGAGGCAGTCATCTAAATTGGACACGGTGGAACAAAATGACAACG				SGSGPLDEAHETKRK
			ATGCGCGTACACCATCAGCGAACAGGTATCGGGTAAATGTCGAACCGA				YYSDNGDVTAAIARE
			ATTTTGGGTAACGTGCACCCCACCATCCTTAATCGGCAGCACGCAATAA				CNIAPSNVAYSSCTIS
			AGCCGTGGGCAGTGGTTTTAGTGGGTAGTCATTAGGAGTCCCACAGTAC				WRGVWSPRSAADLL
			CCAGCGAACATCTTAGTGGGTCTGCGTAAACGCATTTCCACTGCCTATCC				QVGLSKKLLGFITLRV
			TCCGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA				LRGSHLNWTRWNK
			AAAA (SEQ ID NO: 1104)				MTTMRVHHQRTGIG (SEQ ID NO: 1472)
R2	R2Amel	.	Apis mellifera	TGGTAATCAAATGCCTCGCTATTTTAGTAGCGGTAGCGCTCCGCCCGCG	TGGTAATC	TGATCG	MSSNEEGASDTGAP
				CAGGAACCATTGACGCCGCCGTAGTGTGGGTGATTTTATATCCAACCAA	AAATGCCT	TTAAAA	GPGVPVADVSAADG
				TCACGTCAACTACGATCATTTGTAATCACCGACGGTACTTGGTAGGGGT	CGCTATTT	GTAAAA	RATYDDHGMSTDYE

		ACCACATGGGCATTCTTGCTCATTCCACAACGCCGCCTCCATCATGGCAA	TAGTAGCG	ATCTATT	KQTIELPLNGQIQCL
		CAATTTAAAATATATATAAATTCTTAAGGTTTGACCGTATTCATATATATA	GTAGCGCT	TATTTAT	WCHIEGRNQRFLQE
		TATATATATTTAATATTACAACCATAAATCTTATATCGAGCCTTCTATTTG	CCGCCCGC	TTTTATT	SQYLKHKDTQHPKGE
		GTCTCAAAAGCAATACGTTGTCAGATCTTGTAGAACATCAGGAGTGAGC	GCAGGAA	CCTATAT	IIWRCAACQKEFEKL
		GGTGCGCTGTGGTATCCGTGCTTTGTGCCGCGGCGACAAACCAATACGC	CCATTGAC	TATAAC	HGCRCHLPKCKGRKE
		TGCTGCCTGTCGCAAAGCAATACGCTGCTGATTCTCGGATGCGGGTGTC	GCCGCCGT	ACATTAT	AKGVAKFKCDSCEES
		GACGGTCACGCAAAGCGATACGCTGGTGGGGTTTCAAAACAATACGGC	AGTGTGG	TTATTTA	FLTQRGLSMHELHRH
		GCTGGTGCTAAAAAGCATTATGCCGCTAACGGCTGGATTGTCGATCGCC	GTGATTTT	TTTACTT	PAIRNLKRTQGTSRG
		GCTGCGGGGGCTAGTGGCGCACCCAGAGAGGTGCGACGCGCAAGCAT	ATATCCAA	ATTGTTT	NTRPINRASVVWSKEE
		TGGTTCTGTGCGAAGCGGAGTTCTTGAGAGTAATGGTTGCTGGGGGCA	CCAATCAC	TAAAGA	TDLLIKLNERYKHLKQ
		CAAAGCGCAACATATAGCCTCTTATGCCTCAAGTCGTAGTTCGTACCTCC	GTCAACTA	TGACGA	PNVALKEYFPDKTLK
		ACGTGGTCCCGCTGGAATGCCTATCGACTCCTCCCCGGAGGATCATAGA	CGATCATT	AGCCGC	QISDKRRLLPVQEPE
		GTTCGAAACCGGCTACGGCGAGGCAAGGGCGGTGAGGTGCACACCGA	TGTAATCA	AAGGCC	DVATTDETGPPPSDS
		TGGGGAGCAGCGACCCCACCTACCCTTAGCTAAGAGAGCAGGCGATCC	CCGACGGT	AATCCA	SEESIYESATEDEGGG
		GCCAACTGTCAGCACGAAATAAACTAATCATATGTATACGAGGGAGAAT	ACTTGGTA	AATTTAA	DMQQTAPNDSWKE
		TTACAACGGGTACCTTGTGCCCGAACCGCCTGTAGGTATCACCTACAGG	GGGGTACC	CAAAAG	PFIQSIRTNHLEEEDSL
		TGTTAAAATGAATCTGATAGCTGGCGGATCGTCGACCCTCTTTGATGGC	ACATGGGC	AACGAG	RKVEEAIERMAMNE
		TCTGCGCCAACGACTGGAAAGAATAGGAACGGAAGTCTAATGGAAGGA	ATTCTTGC	ACTACT	GVTEQEVGTLLEQFV
		AAGTGTCGGGAGCACTATAAATTCCCAAAGAAGAAAAGAAAAGAAAAA	TCATTCCA	GGTCGA	DSLTQSPTTERKGSR
		AAATAAAAAACCCAAATTAAATGTCGAGTAACGAAGAAGGAGCCTCGG	CAACGCCG	CATTAA	RKSQKTTKRKTTHNN
		ACACGGGGGCCCCCGGCCCTGGGGTCCCCGTGGCGGACGTGTCCGCTG	CCTCCATC	AAAGAC	RKKFLYAKHQELYKKS
		CGGATGGAAGGGCCACTTATGACGACCATGGCATGTCCACGGACTACG	ATGGCAAC	GAAGCA	PRRLLELALSGESSSG
		AGAAACAAACCATCGAACTGCCCCTGAACGGGCAGATACAATGCTTATG	AATTTAAA	GCTGCC	REVVNLPEADSVGPL
		GTGCCACATTGAAGGAAGGAACCAAAGATTCCTCCAGGAAAGCCAATA	ATATATAT	AGCTGA	YKSLWGQIGPEKTHR
		CTTAAAACACAAAGACACGCAACACCCTAAAGGCGAAATAATATGGAG	AAATTCTT	TAAACA	NQPMCNNIDMSEI
		GTGCGCTGCATGCCAGAAAGAATTCGAGAAACTCCATGGCTGCAGGTG	AAGGTTTG	ACAGAG	WTPIALESLVEKFKKI
		TCACTTACCCAAGTGTAAGGGACGCAAAGAAGCCAAGGGTGTTGCCAA	ACCGTATT	CCCGTCT	KSDTAAGADQIKKFH
		ATTTAAATGCGACTCCTGCGAGGAATCGTTCCTAACGCAAAGAGGATTG	CATATATA	CGGCCT	LRKKGALHVFAKLCN
		TCTATGCACGAGCTACATAGACATCCAGCGATTAGGAACTTGAAAAGAA	TATATATA	TTACACC	LLMLHRIYPAQWKT
		CACAAGGCACCAGTCGGGGAAATACCAGACCAATCAACAGAGCCTCGG	TATTTAAT	GAGCGG	NRTTLIPKPGKSAEEV
		TCTGGTCCAAAGAGGAAACGGACCTCTTGATAAAACTTAACGAGCGCTA	ATTACAAC	TGCAAG	ENWRPITIGSLLGRIY
		CAAACACTTAAAACAGCCGAACGTAGCGCTCAAAGAATATTTCCCCGAC	CATAAATC	TCCTGAC	SAMIDRKLRSKIKQHI
		AAGACACTAAAACAAATCAGCGACAAAAGAAGGCTCTTGCCCGTTCAG	TTATATCG	GTACTAT	RQKGFTQEDGCKNN
		GAACCCGAAGACGTGGCCACAACTGATGAAACGGGACCTCCTCCTTCCG	AGCCTTCT	TGTACG	IAILSSALTKMKEDSG
		ACTCATCGGAAGAGAGCATATACGAATCGGCCACGGAGGACGAAGGA	ATTTGGTC	TCTAGG	GIITIIDISKAFDTVPH
		GGAGGAGATATGCAACAAACGGCTCCAAACGATAGCTGGAAGGAGCC	TCAAAAGC	GCGCGG	GEISQSLMNKGVPSP
		GTTTATACAAAGTATAAGAACAAACCACCTCGAAGAGGAAGACTCCCTT	AATACGTT	GGCAGA	ICEYIQKMYIGCKTIIY
		CGAAAGGTGGAAGAAGCCATCGAAAGAATGGCTATGAATGAAGGGGT	GTCAGATC	TTCTACC	CRDKKTLPVDILRGVK

	AACTGAACAAGAGGTGGGCACCCTTCTTGAACAATTTGTCGACTCCCTA	TTGTAGAA	GTGTAG	QGDPLSPLLFNLIIDPI	
	ACTCAATCCCCAACAACGGAAAGAAAGGGGAGCCGACGTAAGAGTCAA	CATCAGGA	AATCTG	IGTLDETTEGIKLENE	
	AAGACTACAAAAAGAAAGACCACCCATAACAATAGAAAAAAAGTTCTTAT	GTGAGCG	GGGCGA	NISVLAFADDLVLLAK	
	ATGCCAAACACCAGGAGCTCTATAAAAAGAGCCCACGAAGGCTTCTGG	GTGCGCTG	CGCCTCC	DKETADKQNRLINEY	
	AGTTGGCGTTATCGGGTGAGTCTAGCAGTGGCAGAGAAGTGGTTAATC	TGGTATCC	GCGAGG	LDDLKMKVSAEKCTT	
	TCCCTGAGGCCGACTCAGTGGGTCCACTATATAAAAGTCTATGGGGCCA	GTGCTTTG	CACTCCC	FEIKRQNKTWFLGDP	
	AATAGGCCCGGAAAAAACTCACAGAAACCAACCTATGTGCAACAATATC	TGCCGCGG	TGGACA	QLTLGQQRIPYADPE	
	GATATGAGCGAAATTTGGACTCCAATCGCCCTGGAGAGCCTTGTCGAAA	CGACAAAC	ACGTAC	AAIKYLGTNFNPWRG	
	AATTCAAAAAGATAAAGTCCGACACCGCAGCCGGCGCGGACCAGATAA	CAATACGC	GCTAAA	LCKTSIKEIIDAARTVK	
	AGAAATTCCACCTGAGAAAGAAAGGGGCACTACACGTATTCGCCAAACT	TGCTGCCT	GCGTAC	QLKLKPHQKINLIRTY	
	GTGTAACCTCCTCATGCTGCACCGAATATACCCAGCACAGTGGAAAACC	GTCGCAAA	GGCTAA	LLPRYIHKLVANPPPL	
	AACCGAACCACGCTTATTCCCAAACCGGGGAAGAGCGCGGAAGAGGTT	GCAATACG	GTGCGC	GTLDLIDKELKTIIKEIL	
	GAGAACTGGAGACCAATCACCATCGGGTCTCTGCTGGGAAGAATATATT	CTGCTGAT	CTCCCG	HLHPSTTDGLIYTDKS	
	CGGCTATGATCGACCGTAAATTACGGTCGAAAATAAAGCAGCACATAAG	TCTCGGAT	AAAGGG	HGGLGIQRVANIVKL	
	ACAGAAGGGGTTTACACAGGAGGATGGCTGTAAAAATAATATAGCCAT	GCGGGTG	TCCCCGT	AKLKHSILMTRSEDN	
	TCTCAGTAGTGCCTTAACCAAAATGAAAGAGGACTCAGGTGGAATCATA	TCGACGGT	TCCTAAT	AVKIALNGQEGMVK	
	ACCATAATAGACATTTCCAAAGCCTTCGACACGGTTCCCCACGGCGAAA	CACGCAAA	TTTTCCG	RYATSIGLQWPCGIE	
	TAAGCCAAAGTCTGATGAACAAAGGAGTCCCATCGCCCATATGCGAATA	GCGATACG	AGCCCG	EIEETRKKLKRADTNK	
	CATTCAAAAAATGTACATAGGTTGTAAAACTATTATATATTGCAGAGACA	CTGGTGG	CGGGCA	WKTLISQGQGIKEFF	
	AGAAAACACTGCCAGTGGACATACTGAGAGGAGTCAAACAGGGAGACC	GGTTTCAA	GATCTC	GDKTGNAWLYNPE	
	CGCTATCGCCACTGCTTTTCAACTTGATAATAGATCCCATAATAGGGACA	AACAATAC	GTGGCA	MLRPSRYLDALKLRT	
	CTGGACGAGACGACGGAGGGCATTAAATTAGAAAACGAGAACATTTCA	GGCGCTG	GTGACG	NTYGTKAALHRAKRD	
	GTTCTCGCCTTCGCCGACGACCTTGTCCTTTTGGCGAAAGACAAAGAAA	GTGCTAAA	CTAGAA	IDINCRRCGVQVETL	
	CAGCCGATAAGCAAAATCGGCTCATCAATGAATATCTGGACGACCTGAA	AAGCATTA	AGTTAA	GHILGLCTHTKNKRIK	
	AATGAAAGTATCCGCCGAAAAATGCACACAACCTTCGAAATCAAACGGCAG	TGCCGCTA	GTCCGC	RHDEICDLIAKNVSKE	
	AACAAAACGTGGTTCCTAGGAGACCCACAGTTGACGTTGGGTCAGCAAC	ACGGCTGG	GGACAT	YVIFREPEVEVNGDR	
	GTATCCCGTATGCCGACCCAGAAGCAGCAATCAAATACCTAGGAACCAA	ATTGTCGA	ATAAAA	RKPDMVIKDHDKVY	
	CTTCAATCCATGGAGAGGGTTGTGCAAAACCTCGATAAAAGAAATCATC	TCGCCGCT	TTACAG	VVDVTVRYENNDSL	
	GATGCGGCTAGAACTGTCAAACAGCTGAAACTTAAGCCGCATCAAAAAA	GCGGGGG	CCTTAAA	NKAYKEKENKYKETA	
	TCAACCTTATAAGAACCTACCTCTTGCCAAGATACATACATAAATTGGTG	CTAGTGGC	TAATGA	EIMRRDLKAKESRVL	
	GCAAATCCTCCCCCTCTGGGGACTCTAGACCTAATCGATAAAGAGCTCA	GCACCCAG	ACCCCA	PVVIGSRGAVPRATIE	
	AAACTATAATAAAGGAAATATTGCACCTCCATCCGTCCACCACGGACGG	AGAGGTG	CGAAGG	NLKVLGLQTKHALTA	
	ACTAATATACACTGATAAGAGCCATGGCGGTCTAGGGATCCAGCGGGT	CGACGCGC	AGGTAT	SLIALRSSIEMANEFL	
	GGCGAACATAGTCAAGCTGGCCAAACTAAAACATAGTATACTAATGACA	AAGCATTG	CCTCGA	DYDHTT (SEQ ID	
	AGGTCAGAGGATAATGCCGTCAAGATAGCACTTAACGGGCAAGAGGGA	GTTCTGTG	AATTCC	NO: 1473)	
	ATGGTGAAAAGATACGCCACGTCCATAGGCCTACAATGGCCATGTGGA	CGAAGCG	GCCACG		
	ATAGAAGAAATCGAGGAAACGCGTAAAAAACTCAAGAGGGCGGATAC	GAGTTCTT	ATCCTTC		
	AAACAAATGGAAAACTTTAATTTCGCAAGGACAAGGCATAAAAGAATTT	GAGAGTA	TGATCG		

		TTCGGGGATAAAACCGGGAATGCCTGGTTATACAACCCCGAAATGCTGA	ATGGTTGC	TAGGCG	
		GACCGTCTCGATACCTGGACGCACTAAAACTGAGAACAAATACATATGG	TGGGGGC	CAAAAC	
		CACAAAAGCAGCACTCCACAGAGCGAAAAGAGACATAGACATAAACTG	ACAAAGCG	A (SEQ	
		TCGGAGATGCGGCGTCCAGGTGGAAACCCTAGGACATATATTGGGACT	CAACATAT	ID NO:	
		ATGCACCCATACAAAGAACAAAAGAATAAAAAGACACGACGAAATCTG	AGCCTCTT	1351)	
		CGATCTTATCGCAAAGAATGTCTCTAAAGAATACGTGATATTTAGGGAA	ATGCCTCA		
		CCAGAAGTAGAGGTAAACGGTGACAGACGTAAACCAGATATGGTCATA	AGTCGTAG		
		AAAGACCATGACAAGGTATACGTCGTGGACGTCACCGTAAGATACGAA	TTCGTACC		
		AACAATGATTCCCTAAACAAGGCCTACAAAGAAAAAGAAAACAAATACA	TCCACGTG		
		AAGAGACAGCGGAAATTATGAGAAGAGACTTGAAAGCAAAAGAGAGC	GTCCCGCT		
		AGAGTTCTGCCAGTGGTTATCGGGAGCAGAGGGGCAGTGCCCCGAGCC	GGAATGCC		
		ACTATAGAAAATCTAAAAGTCCTAGGGCTTCAAACAAAACATGCCCTGA	TATCGACT		
		CGGCTTCGCTCATAGCCCTCCGATCGTCGATCGAAATGGCAAATGAATT	CCTCCCCG		
		CCTGGACTACGATCACACTACGTGATCGTTAAAAGTAAAAATCTATTTAT	GAGGATCA		
		TTATTTTTATTCCTATATTATAACACATTATTTATTTACTTATTGTTTT	TAGAGTTC		
		AAAGATGACGAAGCCGCAAGGCCAATCCAAATTTAACAAAAGAACGAG	GAAACCG		
		ACTACTGGTCGACATTAAAAAGACGAAGCAGCTGCCAGCTGATAAACAA	GCTACGGC		
		CAGAGCCCGTCTCGGCCTTTACACCGAGCGGTGCAAGTCCTGACGTACT	GAGGCAA		
		ATTGTACGTCTAGGGCGCGGGGCAGATTCTACCGTGTAGAATCTGGGG	GGGCGGT		
		CGACGCCTCCGCGAGGCACTCCCTGGACAACGTACGCTAAAGCGTACG	GAGGTGC		
		GCTAAGTGCGCCTCCCGAAAGGGTCCCCGTTCCTAATTTTTCCGAGCCCG	ACACCGAT		
		CGGGCAGATCTCGTGGCAGTGACGCTAGAAAGTTAAGTCCGCGGACAT	GGGGAGC		
		ATAAAATTACAGCCTTAAATAATGAACCCCACGAAGGAGGTATCCTCGA	AGCGACCC		
		AATTCCGCCACGATCCTTCTGATCGTAGGCGCAAAACA (SEQ ID NO: 1105)	CACCTACC		
			CTTAGCTA		
			AGAGAGC		
			AGGCGATC		
			CGCCAACT		
			GTCAGCAC		
			GAAATAAA		
			CTAATCAT		
			ATGTATAC		
			GAGGGAG		
			AATTTACA		
			ACGGGTAC		
			CTTGTGCC		
			CGAACCGC		
			CTGTAGGT		

				ATCACCTA CAGGTGTT AAAATGAA TCTGATAG CTGGCGGA TCGTCGAC CCTCTTTG ATGGCTCT GCGCCAAC GACTGGAA AGAATAG GAACGGA AGTCTAAT GGAAGGA AAGTGTCG GGAGCACT ATAAATTC CCAAAGAA GAAAAGA AAAGAAA AAAAATAA AAAACCCA AATTAA (SEQ ID NO: 1228)		
R2	R2B_NVi	.	Nasonia vitripennis	GACTAGACTATGGGTTCAGTCAGTCCCAAATAGCCGATCCTGGCGCGTC CGGCAGTAATGCCACGTATGAGTCGGTTACCCATCTCTAAACGCGTAGA GGTGGGGAGCTAAAGGCCAGGCGGTTTACCCGACGTCGAATTTCTCCA GGTCTGTGTCAGTCGACGGAATAAAGGTACTACAACATCTACTATCTAT CGGGATCGGAAGACGCCTTACAGCGTTTTCCGATTTTTGCTCTTTGAGCA TTTTTCTTCAAATTGCGATAACCGACCCGATCACGCGGGGCTTTGACAAA GCAATGCGTGGTCGGTAAGATGGTTGCAATCTTTTCCACCTCGTTTCTTT TACGGAACGAAAGCAATGCGTGTGGGGAACGTTAAAAACTCCCTTCATG CATCCCAGGATTTATCCTGCTTACTGCAAAGCAATGCGTGTGGAGCGAC TTTACCACGAGTCGCTCCACCGCAAAGCAATGCGTATCGCGCAAAAGCA ATGCGTGTGGGGGACTTGTCAAAGATCCCCCGCCGCAAAGCAATGCGT GTCGGCACCACGTAGAGCAAAGCGTGTAGGCAGACTTTGTCAAAAGTA GTTCTGCCGCAAAGCAATGCGTGTGGAGATCTTCGCCGGTGAAAGCAAT	GACTAGAC TATGGGTT CAGTCAGT CCCAAATA GCCGATCC TGGCGCGT CCGGCAGT AATGCCAC GTATGAGT CGGTTACC CATCTCTA AACGCGTA GAGGTGG	TGACCT GAACAA AACGTG TTGTCTT GTCTTGT CTAAAA CTATTTA TTCGAA ATAAGG GGAGGC TAACTG CCTGCA AGTTGA	TFAPTHPMVRSGPC RKTKRPGSDYRESLI MDSGNNVASEPRGA VDVTSAAPIGAELNA EPCEGRNQRREAALS AQTRRRNXARRARN AQQADEPGDDEEIET HGPLTIRTXEPMEIVA IAKNPQACPKCLQGG TQLLCMGSWELSRHI NKEHPSVDVTWVCG ACQRRCTTLRSWSC HVLHCKGRQEPKDLP

		ACGTGTGGGCGAACTTAGACCTTCGCGCCCACACATCCCATGGTGAGGA	GGAGCTAA	ACGCGA	FKCEHCSLSFDSQIGL
		GCGGCCCATGTCGAAAGACTAAGCGCCCCGGTAGTGACTACCGTGAAA	AGGCCAG	AAGTTA	SQHERHVHPEVRND
		GTCTAATAATGGACAGCGGAAATAACGTTGCCTCGGAGCCGAGGGGAG	GCGGTTTA	GACCTTC	KRAAEANKPKGKSGR
		CTGTGGATGTGACCTCAGCAGCTCCAATCGGGGCGGAGTTAAACGCCG	CCCGACGT	CCACCTA	RPSIWSDEDLLLIRELE
		AACCTTGCGAAGGTCGCAACCAAAGGAGGGAGGCTGCCTTAAGTGCTC	CGAATTTC	AAGCCC	SEYHGARNINEKIAE
		AAACACGCCGGCGAAATNCGGCCCGCCGAGCTCGGAATGCCCAACAGG	TCCAGGTC	AAAAGT	HFPDRTGRQVSDAR
		CTGACGAGCCCCGGCGATGATGAGGAAATAGAGACACACGGGCCTCTAA	TGTGTCAG	GATCGG	RRKDYAALRGRGGP
		CTATCCGGACGNCGGAGCCGATGGAGATTGTCGCAATAGCGAAAAACC	TCGACGGA	GGAATG	QGPAEGVEAIEEVDE
		CACAGGCCTGTCCCAAATGCCTGCAGGGAGGTACCCAACTTCTCTGCAT	ATAAAGGT	AATCCG	GEIPEGEELVATDGA
		GGGCAGCTGGGAACTAAGCAGGCACATTAATAAAGAACATCCGTCAGT	ACTACAAC	CGGGTG	ALESGPPENGGSAPA
		CGACGTGACCTGGGTGTGCGGTGCTTGTCAAAGGCGCTGCACACAACGCT	ATCTACTA	ACCCCA	EQVNAPALESSSQQ
		CAGGTCGTGGAGCTGTCATGTTCTNCACTGTAAGGGGCGACAAGAACC	TCTATCGG	GAGTTG	DRECSPAVGSDEQIE
		AAAGGATCTGCCGTTCAAATGTGAGCATTGCAGCTTGTCGTTTGACTCG	GATCGGAA	GGTAAA	DSSDDDEFSDALGEIS
		CAAATCGGACTCTCTCAGCACGAGAGGCATGTCCATCCAGAGGTGCGA	GACGCCTT	CCCTTGA	LPEPLSVERTTISPPPR
		AACGATAAGCGCGCGGCAGAGGCCAATAAGCCAAAGGGCAAGAGTGG	ACAGCGTT	AACGTT	DDWKGPMRWEICN
		CCGTAGGCCTTCTATATGGTCCGACGAAGACTTGCTGCTCATCCGGGAA	TTCCGATT	GGAGAA	ASEEAGSYANWVTG
		TTAGAGAGCGAATACCACGGAGCTCGAAATATCAATGAAAAAATAGCT	TTTGCTCTT	GCGGAA	LQELVRNNALSEIGLD
		GAACATTTCCCGGATAGAACAGGCAGACAGGTGTCGGACGCCCGGAGG	TGAGCATT	GAGAGT	SLYDQLIQIMRHPSD
		CGTAAGGATTATGCCGCGCTCCGAGGGAGAGGAGGCCCGCAAGGCCCA	TTTCTTCA	CCCGCC	DNEQDRLQLNARGP
		GCAGAAGGAGTCGAGGCCATCGAAGAGGTAGACGAAGGCGAAATCCC	AATTGCGA	ACCGAG	PRRGHRKNRRRRRRLT
		TGAGGGGGAAGAGCTGGTCGCCACCGATGGCGCTGCGTTGGAAAGCG	TAACCGAC	CATCGA	AADRKRFAFARCQDL
		GCCCCCCGGAGAATGGAGGGAGCGCACCCGCAGAACAAGTCAATGCGC	CCGATCAC	GTGCTG	WNNNPKKLAELVIA
		CCGCGCTGGAAAGCAGTTCTCAGCAAGATCGAGAGTGCAGCCCGGCAG	GCGGGGC	CGGCGC	NDLSILQRRQAPGRT
		TGGGGTCCGATGAACAAATCGAGGACAGCAGTGACGACGACGAATTCA	TTTGACAA	CCGAAT	ETQTLYNELWGRVG
		GCGACGCATTAGGAGAAATATCACTCCCAGAACCTCTCTCGGTTGAACG	AGCAATGC	GAAACC	PNIEAPRRTEDPIPVS
		CACAACAATCTCACCACCTCCTCGAGATGACTGGAAAGGCCCAATGAGG	GTGGTCG	GATCGC	RIFTPITPQEIMGRIRR
		TGGGAGATTTGCAATGCGAGCGAAGAAGCCGGAAGTTACGCGAACTGG	GTAAGATG	GGATGG	IKNDSAAGPDGVTKD
		GTGACTGGACTGCAGGAGCTAGTCAGGAACAATGCGCTGAGTGAAATA	GTTGCAAT	TGCAAG	DLRGRGVSIALSKLFN
		GGACTAGACTCCCTGTATGACCAGCTCATCCAGATTATATGCGGCACCCTTC	CTTTTCCA	TCGTAG	SILLAGYYPKAWREN
		CGATGACAACGAACAGGATCGCCTTCAATTGAACGCTAGAGGCCCCCCA	CCTCGTTT	GACGGG	RTTLLPKPEKDPADV
		CGAAGGGGCCACCGCAAGAACCGACGGCGCCGTCGTCTCACGGCCGCT	CTTTTACG	GCACGA	KNWRPITISSMVSRV
		GATCGAAAGCGGTTTGCCTTTGCCAGGTGCCAAGATCTTTGGAACAACA	GAACGAA	CCTAAG	YSGLLDQRVRAVIKQ
		ACCCAAAGAAGCTAGCCGAGTTAGTGATTGCCAATGACCTGTCCATTCT	AGCAATGC	CCTCTGT	CDRQKGFTEENGCFS
		CCAAAGGCGCCAAGCGCCAGGTAGAACGGAGACACAGACTCTGTACAA	GTGTGGG	CACGGC	NIQLLDDAVSNAKKA
		CGAGCTGTGGGGGAGGGTCGGACCTAATATCGAAGCGCCAAGGCGCA	GAACGTTA	GGCGAA	GGVITILDVSKAFDTV
		CCGAAGACCCGATACCCGTATCGAGGATCTTCACTCCGATCACTCCCCAA	AAAACTCC	GCCAGG	PHAVIQGCLEKKGIPE
		GAGATAATGGGCAGAATCAGGCGAATCAAAAACGACTCGGCAGCGGGT	CTTCATGC	AATCAC	TVAAYISSMYRDCST
		CCTGACGGGGTAACGAAGGACGACCTGAGAGGAAGAGGAGTCAGCAT	ATCCCAGG	CATGCA	AIRTRSGDVKIGMKR

	AGCCCTCTCCAAGCTGTTCAACTCGATCCTGCTAGCGGGTTACTACCCAA	ATTTATCC	AAGGTG	GVKQGDPLSPLIFNL
	AGGCATGGAGAGAGAACAGAACAACCCTTCTGCCGAAGCCAGAAAAAG	TGCTTACT	TGAACT	VLEPLLERLQETSGVE
	ATCCTGCTGACGTTAAGAACTGGCGGCCCCATTACCATCAGCTCAATGGT	GCAAAGCA	GGGGCG	IEGMNLSCAAFADDI
	TAGTCGAGTCTACTCAGGCTTGCTTGACCAGCGAGTGAGGGCCGTCATT	ATGCGTGT	GATACC	VCFANTAPEAGRQLR
	AAGCAGTGTGATCGGCAGAAAGGATTCACGGAGGAAAATGGCTGTTTC	GGAGCGA	TCCACG	MVADYLGRLDMSLS
	AGCAACATACAGTTGTTGGATGACGCCGTATCGAACGCAAAGAAAGCG	CTTTACCA	GGGTTT	VSKCIAVEYVPHRKT
	GGCGGTGTCATTACTATCTTGGATGTTTCGAAAGCATTCGACACTGTCCC	CGAGTCGC	CCCTGG	WYTKNPGLEVNGNA
	GCATGCCGTGATCCAAGGGTGCTTGGAGAAAAAAGGAATCCCCGAAAC	TCCACCGC	GCATCG	VPSISPSETFKYLGAK
	CGTGGCCGCCTATATCTCGAGCATGTATCGCGACTGCTCCACTGCAATCC	AAAGCAAT	CGCGAG	VSPWKGLLEGFESDA
	GAACGAGGAGCGGGGGACGTAAAGATTGGAATGAAGAGAGGAGTCAAG	GCGTATCG	CGATGG	FREVISRVQRLPLKP
	CAGGGGGATCCCCTGTCACCTCTCATTTTCAATCTGGTTCTCGAACCTCT	CGCAAAAG	CCAAAG	MQKVDLLQMYIFPR
	ATTAGAACGATTGCAAGAGACAAGTGGAGTGGAAATCGAAGGCATGAA	CAATGCGT	TCCGCTT	YTYGLITSPPAKAVLK
	TCTCTCGTGCGCGGCTTTCGCAGACGACATAGTATGTTTTGCGAATACA	GTGGGGG	TCTCAGC	TIDRIIRTRIKEILHLPE
	GCCCCCGAGGCGGGGAAGGCAGCTACGGATGGTGGCGGATTATCTGGG	ACTTGTCA	TACAAA	SVSSSFLYTPRKQGGL
	CCGACTCGATATGAGTCTTTCAGTGTCAAAGTGTATAGCTGTAGAGTAT	AAGATCCC	ACAAAA	GLLEVEKMVLIAALR
	GTCCCCCACAGGAAGACCTGGTACACTAAAAACCCAGGCCTCGAGGTG	CCGCCGCA	ATGGTA	NGLRARQSHDPVTR
	AACGGTAATGCCGTTCCGAGCATCTCACCTAGTGAGACGTTCAAGTACC	AAGCAATG	TGAGAC	AAMNSNAADDRLKS
	TCGGGGCAAAGGTCTCTCCCTGGAAGGGGCTGCTCGAAGGCTTCGAAT	CGTGTCGG	TTCGTTA	YADALRLHWPLTTKE
	CTGACGCGTTCAGGGAAGTCATATCCCGCGTCCAAAGACTGCCGTTGAA	CACCACGT	ACACTA	LDTYKYQLRLSYAQK
	GCCCATGCAAAAGGTGGACCTTCTACAGATGTATATCTTTCCGAGGTAC	AGAGCAA	ATTTTTC	WAEQKWQGQGVEE
	ACCTATGGGTTGATAACATCGCCTCCGGCGAAGGCAGTCTTAAAGACTA	AGCGTGTA	CGAGCC	FAQDPVGNSWLQRY
	TCGACCGGATCATAAGAACGAGAATCAAGGAGATCCTCCACCTGCCAGA	GGCAGACT	TAGCAG	DLLPASRYIDAIKLRT
	ATCGGTAAGTAGCAGTTTTCTCTACACGCCGAGGAAGCAGGGTGGATT	TTGTCAAA	GCTCCCT	NTYPTRALMKIIDGR
	GGGGCTCCTTGAAGTGGAGAAGATGGTGCTGATAGCCGCTCTTCGGAA	AGTAGTTC	TGACAA	VDSSCRKCQGSSETL
	CGGCTTGAGAGCCCGTCAATCCCACGATCCGGTCACACGCGCGGCCATG	TGCCGCAA	CGCTTAT	GHILGRCRYTKDKRIS
	AACTCGAACGCAGCCGACGATCGGCTAAAATCTTACGCCGATGCTCTAA	AGCAATGC	GAATCT	RHNEIKDLLKARLAK
	GACTACACTGGCCACTAACAACCAAGGAGCTAGATACTTATAAGTATCA	GTGTGGA	GGAAAA	NHQVMDEPQITVRG
	GCTTCGCCTGAGCTATGCCCAGAAATGGGCTGAGCAAAAATGGCAAGG	GATCTTCG	GGACAC	QRFKPDLVVKTNEGR
	CCAGGGGGTCGAGGAGTTCGCACAAGATCCCGTCGGCAACTCATGGCT	CCGGTGAA	AAAGTG	VHVIDVTVRYEHRTY
	GCAGCGCTATGATCTTCTGCCCGCGTCAAGGTACATCGATGCTATCAAG	AGCAATAC	GAAAAA	LDEGRTEKIGKYRQIL
	CTACGAACAAACACGTACCCGACGCGAGCACTAATGAAGATCATAGATG	GTGTGGG	GCGCTG	STLRRDLHSNAEEVIP
	GACGTGTTGATAGCTCTTGCCGAAAGTGCCAAGGCAGTAGCGAGACCC	CGAACTTA	ATGGTG	IVIGSRGAIPRETRKAL
	TTGGTCATATACTTGGCAGATGCCGGTATACTAAGGATAAGCGAATAAG	G (SEQ ID	GACAAA	SKLGIGKSDWLTISLIA
	CCGGCACAATGAAATCAAAGACCTCCTCAAGGCTCGTCTAGCCAAAAAC	NO: 1229)	AGTCAG	LRSSLEIVNAFMDD
	CATCAGGTTATGGACGAGCCGCAGATAACGGTCCGAGGCCAGAGGTTT		TTGAGA	(SEQ ID NO: 1474)
	AAACCCGATCTCGTCGTGAAAACGAATGAGGGAAGGGTGCACGTAATC		CTTGATA	
	GACGTAACTGTCCGCTACGAGCACAGAACCTATCTGGATGAGGGCCGT		TCAGTT	
	ACTGAGAAAATTGGCAAATATCGCCAAATTCTCAGCACGCTTCGGCGAG		GTTTTGA	

			ATCTGCACTCGAACGCCGAGGAGGTCATTCCAATTGTAATCGGGTCGAG AGGTGCAATTCCAAGGGAAACGCGGAAAGCCCTCTCGAAACTCGGAAT TGGCAAGAGTGATTGGCTTACGATCTCACTAATAGCGCTGCGTAGCTCG CTAGAGATCGTCAACGCCTTCATGGATGACTGACCTGAACAAAACGTGT TGTCTTGTCTTGTCTAAAACTATTTATTCGAAATAAGGGGAGGCTAACTG CCTGCAAGTTGAACGCGAAAGTTAGACCTTCCCACCTAAAGCCCAAAAG TGATCGGGGAATGAATCCGCGGGTGACCCCAGAGTTGGGTAAACCCTT GAAACGTTGGAGAAGCGGAAGAGAGTCCCGCCACCGAGCATCGAGTGC TGCGGCGCCCGAATGAAACCGATCGCGGATGGTGCAAGTCGTAGGACG GGGCACGACCTAAGCCTCTGTCACGGCGGCGCGAAGCCAGGAATCACCAT GCAAAGGTGTGAACTGGGGCGGATACCTCCACGGGGTTTCCCTGGGCA TCGCGCGAGCGATGGCCAAAGTCCGCTTTCTCAGCTACAAAACAAAAAT GGTATGAGACTTCGTTAACACTAATTTTTCCGAGCCTAGCAGGCTCCCTT GACAACGCTTATGAATCTGGAAAAGGACACAAAGTGGAAAAAGCGCTG ATGGTGGACAAAAGTCAGTTGAGACTTGATATCAGTTGTTTTGACTAAG AATTTTATTATCGTTGACTTTTAAATATTTTATTATTGACTGTTAATATACT GACTTGGGACCAAGTCATCTCTGTTACCCGGTACCGGTTCCTGTCATCAA ACCGGAAAGTCCGTCCCACGTAATGTGGTAGACGCAGGAG (SEQ ID NO: 1106)			CTAAGA ATTTTAT TATCGTT GACTTTT AAATATT TTATTAT TGACTG TTAATAT ACTGAC TTGGGA CCAAGT CATCTCT GTTACCC GGTACC GGTTCC TGTCATC AAACCG GAAAGT CCGTCCC ACGTAA TGTGGT AGACGC AGGAG (SEQ ID NO: 1352)	
R2	R2Ci-B	AB097122	Ciona intestinalis	CGACGGTGAACCACCTTGTCGCGGTGTAAGAGCTTTAGTGTCTCGAACA AGAAATAGCTTGTGTGCTGTCCTTCTGGGCGGTGCACATACTTCTTAACC TCCCGAGGCCATGCCGGCGGGGGGCTTTAGCCCCCGGCAGGTTTTACCAT GCCGGACGGGTTCGAGAGGTAGAGGCCAAACTAAGAGTTCACCAGCAG ACTTCGCACGCGGCTGGCCACTGGCCGAAGTTTAAACAACAGGGCCGC ATCTTCCCAAACTCAATATATGGTGTTAAGTGAACCGTGCCGATGGGTG AATGGCCATGGGTAAGTTGGTCTCTAACGGTCCTTGTGGAGAAGTGGC GACCCTTCACAATACTTCAGCCTTACCCTATGCCCGGGCAACTTAGGGTG GACGTATATCTCCCGCGGAAAACATCTTATCTCATGGATAAAAATATATA CGAAAACACTACTTCTCCAGGAGGGGGACCGCTTTGCGGTGAGAAGAC GCATCGTTCCGATGTGATCATCCCTCCACCGGGGTTCGCCCCGTCGACG GACACTGCTTCCAATACATTAGGGGAAAATGTGGATGCATCTGCAACCA	CGACGGTG AACCACCT TGTCGCGG TGTAAGAG CTTTAGTG TCTCGAAC AAGAAATA GCTTGTGT GCTGTCCT TCTGGGCG GTGCACAT ACTTCTTA	TGACAG TAATAT GAAAAC ATCACAT CTGACC GGCACA GAATCA CCATGC CGTAAT GCACCC AACTAA GGATTC	MGEWPWVSWSLTV LVEKWRPFTILQPYP MPGQLRVDVYLPRK TSYLMDKNIYENTTS PGGGPLCGEKTHRSD VIIPPPGFAPSTDTAS NTLGENVDASATTSS ANPLSQEPGWCESCS KLFKSQRGLRVHQRS KHPELYHSQNQPLPR SKARWSDEEMVIFA REEIANRKIRFINQHL

	CTAGCAGTGCCAACCCATTAAGTCAGGAACCAGGATGGTGCGAATCTTG	ACCTCCCG	CAATGG	HKVFPHRTLESIKGLR
	TTCCAAGCTGTTCAAATCACAAAGAGGACTTAGAGTCCACCAAAGATCA	AGGCCATG	GTAAAA	GKNVRYARIMADLE
	AAACATCCAGAGTTGTACCATTCTCAGAATCAACCGCTTCCGCGATCTAA	CCGGCGG	AAAAAA	AEMTSQPEAATSLCT
	AGCACGATGGAGTGATGAGGAAATGGTTATCTTTGCTCGCGAAGAAAT	GGGCTTTA	AAAAAA	ETSENLASSNVLPQT
	TGCAAATAGGAAGATCCGCTTTATCAATCAGCACTTACACAAAGTGTTTC	GCCCCCGG	AAAAAA	RGWAENLVENIDTA
	CACACAGAACGCTAGAGTCAATTAAAGGGTTGAGGGGAAAAAACGTAA	CAGGTTTT	AAAAAA	HLANLGPLSQFEPGK
	GATACGCTCGCATCATGGCTGACCTTGAGGCGGAAATGACCTCTCAGCC	ACCATGCC	AAAAAA	PSSSTKEAINTEYND
	CGAAGCAGCCACAAGCTTGTGCACCGAAACATCAGAAAATCTCGCTTCT	GGACGGG	AAAAAA	WISKWLPSGAAHRE
	AGCAATGTTCTGCCCCAAACAAGAGGATGGGCTGAAAATCTCGTCGAAA	TTCGAGAG	A (SEQ	RRANPPSTKLNARAT
	ATATTGACACAGCACATCTGGCAAACCTTGGGCCGTTATCCCAGTTTGA	GTAGAGG	ID NO:	RRLQYSRIQNLYKLNR
	ACCTGGTAAACCAAGCAGCAGTACAAAAGAGGCTATCAACACCGAGTA	CCAAACTA	1353)	SACAQEVLSGAWKV
	CAACGACTGGATCAGCAAATGGCTTCCCTCTGGTGCTGCTCATAGAGAA	AGAGTTCA		QSGELNLKEVQPFW
	AGAAGAGCAAATCCACCCTCAACTAAGCTAAATGCTAGAGCAACAAGG	CCAGCAGA		EKMFRKESAKDRRKP
	AGGCTACAATACAGTAGGATCCAGAATCTCTACAAATTAAATCGTTCTGC	CTTCGCAC		KPTGEVLWGLMEPL
	CTGCGCCCAAGAAGTATTAAGCGGGGCTTGGAAGGTCCAGTCGGGCGA	GCGGCTG		TIAEVGSTLKSTTPSA
	ACTGAACTTGAAAGAGGTCCAACCGTTTTGGGAAAAAATGTTCCGGAAA	GCCACTGG		PGPDKLTLDGVKRIPI
	GAATCCGCCAAGGACAGGCGCAAGCCCAAGCCCACTGGTGAAGTACTC	CCGAAGTT		AELVSHYNLWLYAGY
	TGGGGACTGATGGAGCCTCTAACAATTGCCGAGGTGGGATCCACTCTG	TAAACAAC		QPEGLREGITTLIPKIK
	AAGTCAACGACACCATCAGCCCCAGGACCCGACAAGTTAACACTAGATG	AGGGCCG		GTRDPAKLRPITVSSF
	GCGTAAAGAGGATACCCATTGCTGAACTAGTATCCCATTACAACCTATG	CATCTTCC		ICRIFHRCLAQRMETS
	GTTGTACGCTGGCTACCAACCAGAGGGACTGCGCGAAGGAATAACCAC	CAAACTCA		LPLGERQKAFRKVDG
	CCTAATTCCGAAAATTAAGGGTACGCGTGACCCGGCTAAGCTACGCCCA	ATATATGG		ICHNIWSLRSLIHNSK
	ATTACAGTGTCTAGCTTTATTTGCCGCATCTTCCACCGATGTTTGGCACA	TGTTAAGT		DNLKELNITFLDVRKA
	ACGAATGGAAACCTCTCTTCCATTGGGCGAAAGACAGAAAGCTTTCCGC	GAACCGTG		FDSISHKSLGIAAARL
	AAGGTCGATGGCATCTGTCATAATATTTGGTCCCTAAGGTCTCTAATACA	CCG (SEQ		GLPPPLITYISNLYPNC
	CAATAGCAAAGACAACTTAAAAGAACTCAATATAACATTTTTGGACGTA	ID NO:		STKLKVNGKISKPIEV
	CGAAAAGCATTTGACTCCATCTCGCATAAATCACTGGGCATAGCTGCAG	1230)		RRGVRQGDPLSPLLF
	CCCGCCTTGGACTTCCCCCACCCTTGATAACATACATCTCCAACCTTTACC			NAVMDWALSELDPR
	CAAACTGTTCTACCAAATTAAAAGTAAACGGAAAAATATCAAAACCAAT			VGVQIGEQRINHLAF
	TGAAGTTCGTCGAGGAGTCCGACAAGGAGACCCACTCTCCCCGCTACTG			ADDIILVSSTKIGMVS
	TTTAATGCAGTAATGGATTGGGCCCTATCTGAACTGGACCCAAGAGTGG			SINTLSRHLAKSGLEIS
	GAGTACAAATTGGGGAACAACGAATAAACCATCTAGCATTCGCCGATGA			AGKEGKSASMAIVV
	TATTATTTTGGTCTCTTCAACAAAGATCGGAATGGTGTCCTCCATCAATA			DGKKKMWTVDPLPR
	CGTTGAGTCGCCATCTGGCAAAATCAGGCCTCGAAATTAGCGCTGGGGAA			FKVNSQKIPALSITQQ
	AGAAGGAAAATCAGCTAGTATGGCCATCGTTGTTGATGGAAAGAAAAA			YKYLGINIDAQGARN
	GATGTGGACAGTAGATCCCCTTCCTAGGTTCAAAGTAAACAGCCAGAAA			DAARILTEGLAELSRA
	ATTCCGGCACTGAGTATTACACAACAGTATAAGTATCTGGGGATAAACA			PLKPQQRLYLLRVHLL
	TAGATGCCCAAGGGGCAAGAAATGATGCAGCTCGAATCCTTACAGAGG			PKLQHGLVLSSCAKR

			GTCTAGCAGAATTATCTCGTGCCCCACTGAAGCCACAACAGCGACTATA				ALTYLDKSVRSAIRR
			TCTGCTTCGGGTGCATCTTCTTCCGAAACTGCAACACGGACTGGTTTTAA				WLTLPKDTPTAFYHA
			GTAGTTGTGCAAAAAGAGCGCTAACATACCTGGACAAAAGTGTAAGGT				KACDGGLGITRLEHTI
			CTGCAATAAGAAGATGGCTTACACTTCCGAAGGACACACCAACAGCCTT				PILKRNRMMKLTLSE
			TTACCATGCGAAAGCATGTGACGGTGGACTAGGCATTACACGTTTGGAA				DPVIMELVKLTYFTNL
			CACACAATACCAATACTCAAAAGAAATAGAATGATGAAGTTGACATTGT				LHKYSNVKLLNSWPV
			CAGAGGATCCAGTGATCATGGAACTTGTAAAATTGACATACTTTACAAA				TDKDSLARAEASMLH
			TTTACTACACAAATACTCAAATGTGAAGCTGCTTAACTCCTGGCCAGTAA				TSVDGRGLSNCSDVP
			CAGACAAAGACAGTCTGGCCCGTGCCGAAGCCTCCATGCTTCACACCTC				RQSDWVTNGASLLS
			TGTGGATGGGCGTGGACTAAGCAATTGCAGTGATGTCCCGCGCCAAAG				GRDFIGAIKVRGNLLP
			TGACTGGGTCACAAATGGCGCGTCGCTACTTAGTGGTCGTGACTTTATT				TKVSAARGRQREITC
			GGAGCAATAAAAGTGAGGGGCAACCTCTTGCCAACAAAGGTTAGTGCG				DCCRRPESLGHILQTC
			GCACGTGGACGGCAACGGGAGATAACATGCGACTGCTGCAGACGACCG				PRTWGPRISRHDSLL
			GAGTCACTGGGCCACATCCTGCAGACCTGCCCTAGAACATGGGGACCG				KRVRNQACLKNWTP
			AGAATCTCCAGGCATGACTCCCTGTTGAAGCGAGTTCGGAATCAGGCAT				IIEPSIPTNIGLRRPDL
			GCTTAAAAAACTGGACCCCGATAATAGAGCCATCGATACCAACCAACAT				VLAKGNIAFLVDATV
			AGGACTTCGTCGGCCAGACCTCGTTTTGGCCAAAGGTAACATCGCCTTT				VADNANMQLQHEA
			CTTGTAGATGCCACCGTGGTTGCAGACAATGCCAACATGCAGTTGCAAC				KVEKYNNSDIKEWIK
			ACGAAGCAAAAGTGGAAAAGTACAACAACTCAGATATAAAGGAATGGA				VHCPGVDEVRVTSLT
			TAAAAGTACACTGCCCAGGGGTTGATGAAGTGAGGGTAACGTCACTAA				ANWRGCLYGGSASF
			CTGCGAACTGGCGCGGTTGCTTGTATGGAGGTTCAGCATCCTTCCTAAC				LTEDLGLPKAELSLLS
			AGAAGACCTGGGACTGCCTAAAGCAGAGTTAAGTCTGCTCAGTGCAAA				AKINEKGYYLWCAHY
			GATCAATGAAAAAGGATACTACCTTTGGTGTGCTCACTACAGGGGAACA				RGTARLWNRPLRS
			GCAAGGTTGTGGAATCGACCGTTAAGGTCCTGACAGTAATATGAAAACA				(SEQ ID NO: 1475)
			TCACATCTGACCGGCACAGAATCACCATGCCGTAATGCACCCAACTAAG				
			GATTCCAATGGGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA				
			AAAAAA (SEQ ID NO: 1107)				
R2	R2C_NGi	.	Nasonia girault	CGGGTTCCCCCGACTTCGGCTTGCCGTGGTCTGGGGCTCACTGCTTTTTG	CGGGTTCC	TAGCGG	WVTSPRRPRYVGPQ
				TGGAGTCATGGTTACATGGTGACCCTGGTTCCTCGCACCCCCGCTGGAA	CCCGACTT	ACTGGA	KKKASDGNDGRAAA
				ACTATCTGGGGAGGCCATGATTGGGTAACGATAAAGGTCCTGGTCGTG	CGGCTTGC	CTGTCT	RAEPTNPGGPDRAD
				TCCTCCTGAGATAGGCTGAATGGGTCACTAAGTGGCACCTAATGGGTTA	CGTGGTCT	GGAGGA	DDEGDVKFWCEFPG
				CCAGCCCTAGGCGGCCGAGATACGTTGGTCCACAGAAAAAGAAAGCCT	GGGGCTCA	GTGTTTA	CDRFFMTRSGRGLH
				CGGATGGAAATGACGGACGAGCCGCTGCTCGTGCCGAACCAACGAATC	CTGCTTTTT	ACTCGG	HKKGHPDWNDQRN
				CGGGCGGACCAGACCGCGCTGACGACGACGAAGGGGATGTTAAATTCT	GTGGAGTC	GTTCTCA	LAGKQHRKEIWSEEE
				GGTGTGAATTTCCAGGTTGTGATCGCTTCTTTATGACCAGGAGCGGTAG	ATGGTTAC	TGGGAA	RLLLAKKEAELAISGA
				AGGCCTCCATCACAAGAAAGGCCACCCTGATTGGAATGATCAGAGAAA	ATGGTGAC	CCCGAC	RFINVELRDFTARSLD
				CCTGGCCGGAAAGCAACACCGAAAAAGAGATATGGTCGGAGGAAGAAC	CCTGGTTC	AACGTT	AIKGQRKRPDYKILVE
				GTCTCCTGCTTGCCAAAAAAGAGGCGGAGCTTGCCATCAGTGGAGCTA	CTCGCACC	GTTATCT	KFVRELRVRGIRQGV

		GGTTTATTAACGTAGAGTTGCGTGATTTTACAGCGCGCTCCCTAGACGCT	CCCGCTGG	TGTATG	ASRSQQARAMAVA
		ATCAAGGGCCAGCGAAAGAGACCCGACTATAAGATCTTAGTCGAGAAA	AAACTATC	ACAATTC	GAPAATSSGAPPVAT
		TTTGTCAGGGAGTTAAGGGTTAGAGGCATTCGTCAAGGAGTGGCCTCG	TGGGGAG	ATAAAA	QPPPSGRVLRSQVVE
		CGGAGTCAACAAGCTCGCGCGATGGCGGTGGCAGGAGCTCCTGCAGCG	GCCATGAT	AAAAAA	APAMEIPVAESEGDS
		ACGTCCTCGGGGGCACCACCTGTCGCGACTCAACCACCACCATCAGGTC	TGGGTAAC	AAAAAA	SGDELFEDVEPVRLS
		GCGTACTAAGATCTCAGGTCGTTGAAGCACCAGCGATGGAGATCCCCGT	GATAAAG	AAAAAA	DLPPDRFTIYFAGLEIP
		GGCAGAGTCGGAAGGTGACTCCTCGGGGGACGAGCTGTTTGAGGATGT	GTCCTGGT	AAAAAA	GTEDIYAHRLHTICL
		CGAGCCCGTGCGATTGTCCGACCTACCCCCTGACAGGTTTACGATATACT	CGTGTCCT	(SEQ ID	MTTWRTKEEVRLEL
		TTGCTGGGCTTGAAATACCCGGCACCGAAGATATATATGCCCACAGGCT	CCTGAGAT	NO:	GLFLKDLFPSKGSQER
		CCATACCATCTGCCTGATGACAACGTGGCGAACCAAAGAAGAGGTAAG	AGGCTGAA	1354)	PERTNLPDPRNRIER
		ATTAGAACTTGGCCTCTTTTTGAAAGATTTGTTCCCGAGTAAGGGCAGTC	TGGGTCAC		RRGEYKKCQDLWRR
		AAGAACGCCCGGAGAGAACCAACCTGCCGGACCCGAGAAATCGGATCG	TAAGTGGC		NKSTCVQRILKEDLSQ
		AGAGGCGGAGGGGGGGAGTACAAAAAAATGCCAGGATCTATGGCGACGA	ACCTAA		GECLPRELMEPFWN
		AATAAGTCAACCTGTGTTCAGCGGATCCTTAAAGAGGATCTTTCGCAGG	(SEQ ID		ATFTQNPGTAPVLPP
		GTGAATGTTTGCCTCGAGAGCTGATGGAGCCCTTTTGGAATGCGACTTT	NO: 1231)		PTEVYSSVWEPIRPE
		CACCCAGAATCCTGGCACGGCTCCGGTGCTCCCTCCTCCCACGGAGGTT			NIKGNYPPQNTAAGI
		TATTCTAGTGTTTGGGGAGCCCATTCGGCCCGAGAATATCAAGGGCAACT			DGLTVGDLKGVSRE
		ATCCGCCGCAGAACACCGCGGCAGGGATAGACGGACTGACAGTGGGT			MLARIFNLFMWCGK
		GACCTGAAAGGGGTGTCGCGGGAGATGCTGGCCAGAATTTTTAACTTAT			LPEHLCASRTILLPKKP
		TCATGTGGTGCGGCAAACTGCCAGAGCACCTTTGTGCCTCACGCACAAT			GAKVPGEFRPITVTSV
		TCTCCTGCCCAAGAAACCTGGGGCGAAAGTCCCCGGCGAATTCAGGCCT			LIRTFHKVLAERLKVV
		ATCACCGTGACATCCGTCCTCATCCGGACCTTTCACAAGGTTCTGGCCGA			PLDPRQRGFRESDGC
		AAGACTGAAGGTTGTCCCTCTTGACCCCCGCCAAAGAGGCTTCAGAGAG			AENVMLLDMTIRYH
		TCCGATGGATGTGCAGAGAACGTGATGCTACTGGACATGACCATCCGGT			HERRRKMFLALLDM
		ACCACCACGAGCGGCGCAGAAAGATGTTCTTGGCCCTGCTAGACATGG			AKAFDSVSFESMREV
		CTAAGGCATTTGACTCGGTCTCTTTCGAGTCCATGCGGGAGGTTTTGACT			LTTKGIPTPFIEYFMT
		ACTAAAGGCATACCAACGCCATTTATTGAGTATTTTATGACGCACTTGGA			HLEDSFTVLQHGNW
		GGATAGTTTTACTGTTCTTCAGCATGGTAACTGGCAATCGGGGAAAATC			QSGKIHPTCGVKQG
		CACCCAACATGTGGTGTGAAGCAAGGCGATCCACTGTCTCCGCCTATCT			DPLSPPIFNFIMDEM
		TCAACTTCATCATGGATGAAATGTTGAAGAGGTTGCCTAAGGAAATCGG			LKRLPKEIGVNLDGLF
		GGTTAACTTGGACGGGTTATTTGTTAATGCTATGGCATTTGCGGATGAC			VNAMAFADDLSLVA
		CTGAGCCTTGTTGCCAATACCGAACAAGGTCTGCAGATCCTCATAGATG			NTEQGLQLIDEATSF
		AAGCTACTTCCTTTCTGGGGCTCTGTGGACTCCGCGCCAATCCCAATAAG			LGLCGLRANPNKCVT
		TGCGTCACCCTAGCAATTAAGACCATCCCGAAGGAGAAAAAGACGGCC			LAIKTIPKEKKTAIDPS
		ATTGACCCCTCATCACATTTTAGGATAGGTAATGCGGTGATCCCCTCGTT			SHFRIGNAVIPSLKRT
		GAAGAGGACAGATGAGTGGGTGTACTTAGGGATCAAATTTAATTCAAA			DEWVYLGIKFNSNGR
		TGGTCGCCTTATCTCTGACGCAAAACCCAAGCTCATAAAAGATCTTGAG			LISDAKPKLIKDLELLT
		CTACTAACTAAGGCACCACTCAAACCACAGCAAAGACTGTGGGCGCTTA			KAPLKPQQRLWALK

			AGGTGATTGTCATTCCGGGCATCCTTTACAGAGGTACCCTGGGGAGCAG				VIVIPGILYRGTLGSST
			CACCGCAGGCTACCTACGCTCTCTTGACTGTGTAATAAGGGCCTATGTTC				AGYLRSLDCVIRAYVR
			GGCGATGGCTACGTCTCCCTGGAGATTGCCCGAATGGGTATTTTCATGC				RWLRLPGDCPNGYF
			AGCGGTTGCGGATGGAGGGCTGGGAGTTCACCCCATACGATACAAGGC				HAAVADGGLGVHPI
			GATGGTAGATCGCCTTGCCCGGCTCCGAAAATTAGAGAAATCCGCGTAC				RYKAMVDRLARLRKL
			ATCACGGGGCCTGAAGCCGCACGTTATCTTCAAAGACAAGTTTCTATCG				EKSAYITGPEAARYLQ
			CCGAAAATAGGCTCCGAGATGGGGCCAACCGCATTATGAGTGATGCGA				RQVSIAENRLRDGAN
			GTATGCTAAGGGAGTTCCTTCGGGAGCTTCTGTACAAGTCCTTTGATGG				RIMSDASMLREFLRE
			TCGTCCCCTGGAAAATTCCAGCAAAGTACCAGGTCAGCACCGCTGGGTC				LLYKSFDGRPLENSSK
			GAGGAGCCAACCCGTTTCCTATCCGGGGCGGACTATATGAATTGTATAC				VPGQHRWVEEPTRF
			GTGCGAGGATCGCAGCTCTTCCGACTGCAGCCAGGTGTGCTAGGGGAC				LSGADYMNCIRARIA
			GTCTCAAAGACAAGCATTGCCGGGCAGGTTGCGGAAATGTGGAGACGC				ALPTAARCARGRLKD
			TTAACCACGTCTTGCAATTCTGCCACCGTACCCATGGCACTCGCATTGGA				KHCRAGCGNVETLN
			CGCCATGATGCGGTTGTAAAGTATGTTGTAGGAGGACTCAAGAAGAGG				HVLQFCHRTHGTRIG
			GGGTACGCAGTGAAAGAAGAGCCGAAAATCGTCTTACAGGATGTGGTG				RHDAVVKYVVGGLK
			TACAAACCTGATATGGTTGCGACCAAGGAAGGGAAAACACTCATTCTGG				KRGYAVKEEPKIVLQ
			ACGCTCAGGTTCTAGGCGACCAGCGTGATATGAGACTGGCACATGAAG				DVVYKPDMVATKEG
			ATAAGCTCCGTAAGTATGGGGCCCCAGAATTTAAACGAAAGATCAGGA				KTLILDAQVLGDQRD
			GTGAGACGGGGTCGGCAACCATTAAGTCCTTGTCGGTTACATTGAGCTG				MRLAHEDKLRKYGA
			GCGAGGGTTGTGGGGACCTGACTCAGTAAAGGGGCTCCTCGAAGAGG				PEFKRKIRSETGSATIK
			GAGTGATTCTAAAGAAGGACCTTAAGATCTTGTCCACAAGAGTACTTAT				SLSVTLSWRGLWGP
			AGGAGCTTTGGCAGGCTGGAGAAGGTTTAATGAAAGGACGAGTATGGC				DSVKGLLEEGVILKKD
			AACATCTGGAAGAAGAAGAGGTTACAACAAGGATGGTGAGAAGGT				LKILSTRVLIGALAGW
			GGAAGAGAAGAGAGGGTCGGTGTTGGTTAGCGGACTGGACTGTCT				RRFNERTSMATSGRR
			GGAGGAGTGTTTAACTCGGGTTCTCATGGGAACCCGACACAACGTTGTTAT				EEVTTRMVRRWKRR
			CTTGTATGACAATTCATAAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1108)				ERVGVG (SEQ ID NO: 1476)
R2	R2La	JN937617	GGGGTAGCAATTGATCGATTCCCGCTTCCTCGTGGCGCTACCCTGGGTA	Lepidurus arcticus	GGGGTAG	TGATAA	MSGKSSKPRTVSSGS
			ATACTATGAGGAATTGATCACACCGTAGCGAACGTCATCAGTCACAGCT		CAATTGAT	TCGCTCC	SSQETPPSGSNACDI
			GCACGAATCCAGATAGAAATATAACAGACGAGTAATTCTTTTAAAAGCT		CGATTCCC	ATCCTGC	CGKCFMKPVGLSRV
			CGTCAGTAATCTTCCCAGATGTCTGGAAAGAGCAGCAAACCGAGAACTG		GCTTCCTC	AACTAA	HPSQYHARLEKNQPK
			TATCCAGTGGATCCAGTTCCCAAGAGACTCCTCCAAGTGGGTCCAACGC		GTGGCGCT	TTATGA	AKKFRWTDEDLYFLA
			ATGTGATATATGCGGGAAATGCTTCATGAAACCGGTCGGTTTGTCCAGA		ACCCTGGG	ATGCAA	KKEAELLHLGSIKFVN
			GTCCATCCTAGTCAATATCATGCACGCTTAGAAAAGAACCAACCTAAAG		TAATACTA	ATCTGTT	KELAEFFPEKSVDQIR
			CGAAGAAGTTCAGATGGACTGATGAGGATTTGTATTTTTTAGCGAAAAA		TGAGGAAT	AAGTGA	GQRRSETYKQQVLSI
			AGAAGCGGAACTTTTACACTTGGGCAGCATTAAATTTGTGAATAAAGAA		TGATCACA	CATTAGT	HSELLKLQTVADSPPP
			CTGGCAGAATTCTTTCCGGAGAAATCTGTCGACCAAATCAGAGGCCAAA		CCGTAGCG	GATACTT	SRIPAKEVSAWLDFFL
			GACGTTCTGAGACGTACAAACAGCAAGTTTTATCTATCCATTCTGAACTG		AACGTCAT	ACCTGA	ALPKTKNKFSEDKLD

	CTCAAGTTGCAGACTGTAGCCGATTCTCCTCCCCCATCGAGAATCCCGGC	CAGTCACA	TACTTAC	QLIRTAQDGTLILDDL	
	GAAAGAAGTCTCCGCTTGGCTTGACTTTTTTCTGGCCCTTCCCAAAACCA	GCTGCACG	CCTGGT	DLYLREVLVQPTSQG	
	AGAATAAATTTAGTGAAGACAAGTTAGACCAACTCATCCGTACCGCTCA	AATCCAGA	ATTTATT	EKQAKLLPPPKSSREK	
	AGATGGTACTCTCATCCTGGATGATCTAGATTTATATCTTAGAGAAGTTC	TAGAAATA	TGACCT	RDREYARAQNLYRK	
	TAGTTCAACCCACTAGTCAGGGAGAGAAACAAGCGAAGTTGCTCCCACC	TAACAGAC	ATACTTA	NKTACVNAILDGNKK	
	CCCCAAATCCAGTAGAGAGAAAAGAGACCGCGAATATGCCCGTGCCCA	GAGTAATT	CCCTGG	CENKIPDIDDFWKTIF	
	GAATTTATATCGGAAGAATAAGACCGCTTGTGTGAACGCCATCCTTGAT	CTTTTAAA	TATCTAC	ESHSPPDAEPVCYVV	
	GGTAACAAGAAGTGCGAAAATAAAATACCTGACATTGACGATTTCTGGA	AGCTCGTC	CTGACA	DEEPTNIWSWISFFE	
	AGACCATATTTGAGTCCCATTCTCCCCCAGATGCAGAACCGGTGTGTTAT	AGTAATCT	TATATTT	MNHNYPDSSTSPGP	
	GTCGTTGACGAGGAGCCGACAAATATTTGGAGTTGGATTTCATTTTTTG	TCCCAG	ATCTAAC	DGVTARMLRSIPARV	
	AGATGAATCATAAACTACCCCGATTCGTCTACCTCACCCGGTCCTGATGGC	(SEQ ID	CACCTAC	LNKLLNLLLFIEDLPAV	
	GTGACTGCTCGCATGCTCCGTTCCATCCCAGCACGTGTCCTAAATAAATT	NO: 1232)	CTATGAT	FKCHRTVLIPKIDNPT	
	GTTAAACCTTTTACTCTTTATTGAAGACTTACCTGCAGTGTTCAAGTGCC		GACTCC	SPGEFRPITISSIVVRQ	
	ATCGTACCGTGTTGATACCGAAAATCGACAATCCGACATCACCAGGCGA		CGCGGA	LNKIIAARVSEGVPIN	
	ATTTCGTCCCATCACGATTTCCTCCATAGTTGTTCGGCAGCTGAACAAAA		AACTCTC	PRQKAFRQIDGCAEN	
	TCATCGCTGCTCGCGTCTCAGAGGGCGTTCCCATAAATCCTCGCCAGAA		ACTTACC	VFLLDFILRDAKTKIKS	
	GGCCTTCCGGCAAATTGATGGATGCGCTGAGAATGTTTTCCTTTTGGATT		TTATTAC	LSLATVDIKKAFDSVS	
	TCATTCTTCGTGACGCAAAAACGAAAATTAAGTCCCTTTCGTTGGCTACT		CCACTTG	HHSIFRAIRGARCPEN	
	GTCGACATTAAGAAAGCATTTGATAGTGTGTCACATCATTCCATCTTCCG		GTCTTTT	LVNYIQNSYSGCTTQI	
	AGCGATTAGAGGAGCGAGATGCCCTGAGAATCTAGTGAATTATATCCA		ATTTTCT	SVGGSISASKIPMNR	
	GAACAGCTACTCGGGTTGTACGACCCAAATTTCTGTTGGCGGCTCTATAT		CGTTCCT	GVKQGDPLSPVLFNL	
	CAGCATCAAAAATCCCGATGAACCGTGGTGTCAAACAAGGGGACCCACT		TATTACT	VINEIIRKLPASIGYPIN	
	ATCTCCTGTCCTCTTTAATTTAGTGATAAACGAAATTATTCGCAAACTGCC		TTGTTCC	SELSINCIAYADDLILV	
	GGCCTCGATTGGATACCCAATCAATAGTGAGCTTTCGATAAACTGTATT		TTTGGT	TNTREGLKLLLGLLNE	
	GCTTATGCGGATGACCTGATACTAGTTACAAATACTAGAGAGGGCCTGA		GTAGGG	ELPKRGLELNASKCFG	
	AGCTACTACTCGGCCTATTGAATGAAGAGCTCCCTAAGCGAGGATTAGA		TTCTCTG	LSLTALGKLKKTHLCT	
	ATTGAATGCATCTAAGTGTTTCGGCTTATCCTTAACAGCACTCGGGAAAT		GTTTTTG	SDQLDLHGTLIKNLTA	
	TGAAGAAGACTCACCTGTGTACCAGTGACCAACTCGATTTACATGGGAC		GAACGG	EESWVYLGVPFSHIG	
	TCTAATTAAGAATCTTACCGCTGAGGAATCCTGGGTTTATTTGGGTGTTC		CTTCCTT	RSKSFSPDLEALLNKL	
	CTTTCTCACACATTGGGCGCTCTAAGAGCTTCAGTCCAGATCTTGAAGCT		AGCCGG	QKSPLKLQQKLFALR	
	CTTTTAAATAAACTCCAGAAATCGCCGCTTAAACTGCAACAAAAGCTTTT		AATTTTG	VYLIPRLLHGLVLSRV	
	TGCTTTGCGAGTTTATCTAATTCCTAGACTGCTCCACGGTTTAGTCCTATC		TCTGAT	AIGELKIMDKLILKHL	
	AAGAGTGGCCATTGGCGAATTAAAAATAATGGATAAGTTAATTTTGAAA		GTATCTT	RVWLRLPKDTPLGFF	
	CACTTACGCGTATGGCTTCGTTTACCGAAAGATACCCCTTTGGGATTTTT		GCTTGT	YSPVKLGGLGIKNLRT	
	CTACTCACCGGTGAAATTGGGCGGTTTGGGGATTAAAAACTTACGTACA		GTCCTT	NVLKCRKQRIERMLV	
	AACGTTTTGAAGTGCAGGAAACAACGCATTGAGCGAATGCTAGTCTCCC		GAAATA	SPDDVVRLVAESEIFL	
	CTGATGATGTGGTGCGCTTGGTTGCAGAGTCGGAGATTTTCCTGAAAGA		TACGAC	KETDKLKDLLTINGM	
	GACGGACAAGTTAAAGGACCTACTTACTATTAACGGAATGTGCCTAGAT		CCAGGC	CLDXRNVPRTGKNN	

			TANCGGAATGTGCCTAGAACGGGAAAAAATAACAAGTTTTGGAGTGAA AGGCTATATACGTCATTTGATGGTAAAACACTTGCCTATAGCGAGTACTT TACTCAAGGCGGCGGCTGGATCCGTGAAGATAAGATCCTTCAACCTGCC CATGTTTTCGCTGAATGCATCAAGTTACGTATTAATGCTTTACCCACTAA ATCTAGAGTTGCACATGGCCGTCCAACTAAAGATCGTTCCTGTCGTGCC GGATGTCTTGACGTTCAAAAAGTGCCAACTATCGAAACAATCAATCATA TAGCACAAGTTTGTCCTCGGACGCATGGTGCAAGAATAAAAAGACACG ATCGGCTCGTTCAATTCCTTTCTCTCAACCTTAGGAAAAATCCAAAAAGA AATGTTTTAGTGGAGTACAATTTCCGTACAGTCGCTGGTATCCGTAAACC GGATATCATTGTTATTGAAGATACCCGTGCAGTTATCCTTGACGTCCAAG TTGTTGGTGATTCGTCCAATTTGGAGATGGAATACCTGGAGAAATCCAG GAAATATTCTAATGACGCTAACTTTATCAATGCGCTTCAAAAACTTTATC CAACGGTCACTAACCTAACTTTTCATGCAGTTACCTTCAACAATAGAGGC CTGATTGCCAAATCCACAGTGGCGGCCCTCCGGATGCTCGGCGTCCCAC CTAGATGCATAATGATCCTTTGTGTGATTAGCCTTGAGAAGACCCTGGA AGTCTGGAGGATGTTTAACCAATCAACTGCTTCCGCTCGGAAATGATAA TCGCTCCATCCTGCAACTAATTATGAATGCAAATCTGTTAAGTGACATTA GTGATACTTACCTGATACTTACCCTGGTATTTATTTGACCTATACTTACCC TGGTATCTACCTGACATATATTTATCTAACCACCTACCTATGATGACTCCC GCGGAAACTCTCACTTACCTTATTACCCACTTGGTCTTTTATTTTCTCGTT CCTTATTACTTTGTTCCTTTGGTGTAGGGTTCTCTGGTTTTTGGAACGGCT TCCTTAGCCGGAATTTTTGTCTGATGTATCTTGCTTGTGTCCTTGAAATATA CGACCCAGGCTTGCGTCATTTAGGCTCTGGGAAA (SEQ ID NO: 1109)		TTGCGT CATTTAG GCTCTG GGAAA (SEQ ID NO: 1355)	KFWSERLYTSFDGKT LAYSEYFTQGGGWIR EDKILQPAHVFAECIK LRINALPTKSRVAHG RPTKDRSCRAGCLDV QKVPTIETINHIAQVC PRTHGARIKRHDRLV QFLSLNLRKNPKRNV LVEYNFRTVAGIRKP DIIVIEDTRAVILDVQ VVGDSSNLEMEYLEK SRKYSNDANFINALQ KLYPTVTNLTFHAVTF NNRGLIAKSTVAALR MLGVPPRCIMILCVIS LEKTLEVWRMFNQS TASARK (SEQ ID NO: 1477)
R2	R2LcA	Lepidurus couesii	TTTGGGGTAGCAATTGATCGATTCCCGCCTCCTCGTGGCGCTACCCCGG GATAGCCTCAAAGAAATTTGACGGTAAAGCAAAGAGGAATTGATCACC CAAGGCAGTACATCGGCCTTCCTGCAGGAGCTCTGATAAAGATATTAGT GAGTTATTCTGTTGAAGCTCGCTATTTCATTCCCCTGATGTCAGAGGAAA GCAGACCAAAGCAGACGGCCTCCAAGCGTGGAGCCGCCGTGGAGAAG ACTATGATGAGTGGGACCTACGTGTGTACTTTATGTGGTAGGAGTTTTG AGAAGTCGGTGGGTTTGTCACTCCACACGAACCGGATGCACCCTGAAGC TTACAATAAGCTGAAGGAGGCTAAGAAACCTGTTTTGAAGAAGGCGCG ATGGTCAGAGGAGGAGGTTTTCCTTCTAGCGCAGAAGGAAGCGGAGCT CTCATTCATTGGAGGAATTAAGTTCATGAACATCGAGCTGCATAAAATTT TTCCAGAGCGTGAGTTGGAGGGAATCAAAGGCCAAAGAAAGAACCCAA CGTACAAAGCGCAAGTCGTTAGCTTATTAGCTGAGATCCGAGAGTCTAA GGCCAATGATTCCTCTTCATCTTCCTCATCCTCCTCCTTGTGATTCGGC CTCTTTGGGCATTAGCAATTGGCTTGAATTTCTGCTAGCTCTCCCGAAGA CCTCCAACCAATTCCAGGAAGGTAGGCTAGATAGGCTAATCTCTGATGC	TTTGGGGT AGCAATTG ATCGATTC CCGCCTCC TCGTGGCG CTACCCCG GGATAGCC TCAAAGAA ATTTGACG GTAAAGCA AAGAGGA ATTGATCA CCCAAGGC AGTACATC GGCCTTCC	TAGTGC TTGAGT GATGCC TATCCTT TCTTTGA TTAACTC TTACCAT ATACTTA CCAGTTC TTACCCG TACTTAC CCTGTAT ACTTACC TGTGTG CGTACCT	MSEESRPKQTASKRG AAVEKTMMSGTYVC TLCGRSFEKSVGLSLH TNRMHPEAYNKLKE AKKPVLKKARWSEEE VFLLAQKEAELSFIGG IKFMNIELHKIFPEREL EGIKGQRKNPTYKAQ VVSLLAEIRESKANDS SSSSSSSSSCDSASLGI SNWLEFLLALPKTSN QFQEGRLDRLISDAL RGVDVLENLDAYLLE VFAKPMAQNPCPKP PPPAKNSRERRDREY

	CCTGCGGGGTGTTGATGTGCTAGAAAACCTTGATGCTTACCTTCTCGAG	TGCAGGA	GTGTAC		SRVQNFYKKNRSACI
	GTTTTCGCTAAGCCCATGGCGCAAAATCCTTGTCCAAAGCCCCCCACCCCC	GCTCTGAT	TTGTCCT		NSILDGNTRSQNVIP
	TGCGAAGAACAGTCGTGAGAGAAGGGATAGAGAATACTCTAGAGTTCA	AAAGATAT	TTAGCC		GLTKFWTETFEKNSP
	GAATTTTTATAAGAAGAAGAACCGTTCAGCGTGCATTAATTCCATTTTAGACG	TAGTGAGT	GCCTTG		PDDEAPDQFVADEP
	GGAACACCCGGTCTCAAAATGTGATACCTGGCTTGACGAAGTTTTGGAC	TATTCTGT	TGTTTTT		RDMYKWITFYEMSQ
	CGAGACGTTTGAGAAGAACTCGCCACCTGATGATGAAGCCCCTGACCA	TGAAGCTC	ACCATT		DYLDSSTAPGVDGFS
	GTTCGTTGCGGACGAGCCTAGAGACATGTATAAGTGGATAACTTTTTAT	GCTATTTC	GGTACT		AKQLRSMSPRVLNKI
	GAGATGTCTCAAGATTATCTTGATTCTTCCACTGCTCCTGGAGTGGATGG	ATTCCCCT	TACCTTG		LNLLLLSENLPNSFKM
	TTTCTCTGCCAAACAGCTCCGGTCCATGTCTCCCAGGGTGCTAAATAAGA	G (SEQ ID	TGTGGT		HKTVLIPKIDDPKSPG
	TATTGAACCTTTTACTTCTTAGTGAGAACTTACCTAATTCGTTTAAGATGC	NO: 1233)	TGCCCG		DFRPITISPVLARLLNK
	ATAAGACAGTTCTTATACCGAAGATTGATGATCCAAAGAGTCCTGGGGA		ATACTTA		ILAARLSKLVPISQRQ
	TTTTAGACCGATCACGATTTCCCCTGTCCTTGCTCGCCTCCTTAACAAGAT		CCTTGTA		KAFLPVDGCGENIFLL
	TCTCGCTGCTAGATTGAGCAAACTAGTCCCCATTAGTCAGCGTCAAAAG		TTTGCCT		DYILRSSKKSSKSVAM
	GCTTTTTTGCCTGTTGATGGCTGTGGGGAGAATATTTTCCTTTTGGATTA		TGTAATT		AVLDVKKAFDSVSHH
	TATTCTTCGTAGCTCTAAGAAAAGTAGTAAATCAGTGGCGATGGCAGTT		CTGCAT		SILRALNEAKCPINFIN
	TTGGATGTTAAAAAGGCATTTGATAGTGTATCGCATCACTCCATCCTTAG		GATATTT		FVRNSYDGCTTKLTC
	AGCCCTTAATGAGGCCAAGTGTCCTATTAACTTTATTAACTTCGTCAGGA		ATTGTGT		GGTSFPDSVRMNRG
	ATAGTTATGATGGATGCACTACCAAACTTACTTGTGGGGGTACTTCCTTT		AGGTTC		VKQGDPLSPVLFNLII
	CCGGACTCGGTACGCATGAACCGAGGAGTGAAACAGGGAGATCCCCTA		CTGATG		DSAIRKLPDSIGYVIR
	TCGCCTGTGCTCTTTAATTTGATCATTGATAGTGCAATTAGGAAGTTACC		CTTACCT		DGLKINCLAYADDLIL
	TGACTCTATCGGCTATGTGATAAGGGATGGCTTAAAGATTAATTGCCTG		GATTTGT		VASSRAGLKTLLNIVA
	GCATACGCTGATGACCTGATTTTGGTAGCTTCAAGTCGTGCTGGCCTTAA		CCCCCTC		EHLSLRGLDLNAAKC
	GACCTTACTGAATATTGTTGCTGAGCATCTCTCGCTACGCGGGTTGGATT		ATCATCT		HGLSIIASGKAKTTYV
	TGAATGCTGCTAAATGTCATGGTCTGTCAATTATCGCCAGTGGTAAGGC		TTAGTTT		SAADSLDLDGQPIKN
	CAAAACTACTTACGTTAGTGCTGCGGATTCCCTAGATCTTGACGGTCAAC		CGTTCTA		LGVLDTWTYLGIPFS
	CTATAAAGAATTTGGGAGTGCTGGACACGTGGACCTACCTCGGGATTCC		TTTCACT		HLGRAEKVSPDLTNL
	ATTTTCCCATCTGGGTCGAGCTGAGAAAGTTTCTCCTGATTTGACCAACT		CCATTAT		LNKLQKAPLKLQQKL
	TACTTAACAAACTTCAGAAAGCTCCCCTCAAACTCCAGCAGAAGTTATAT		GGAGTT		YAVRNFVIPRALHGLI
	GCTGTACGTAACTTTGTTATCCCTAGAGCACTCCATGGCCTCATTCTATC		CCGTTT		LSKTNLKELNTLDRAI
	GAAGACAAACCTAAAAGAATTAAACACGCTTGATAGAGCAATTAGAGTT		GTTTTTT		RVFLRTLLYLPKDTPL
	TTTCTACGTACCTTGCTTTACTTACCTAAAGATACGCCACTGGGATTTTTC		GGTGGA		GFFHSPIKSGGLGITC
	CATTCCCCTATCAAGTCCGGAGGACTTGGGATTACGTGTTTTAGAACATC		GGTACA		FRTSVLKCRLQRIAR
	TGTCCTTAAGTGCCGGCTACAGAGAATAGCCCGAATGCGTAGTTCCTGT		GCACCC		MRSSCDGVIQAVAES
	GATGGAGTTATACAGGCAGTGGCGGAATCTGATATCTTCGCTGATGAAT		TTTAAGC		DIFADEYAKLRDLIRIN
	ATGCCAAGTTGCGGGACCTTATTAGGATAAACGGCAATGTGCTCGACAC		TGGAAT		GNVLDTTESIKRYWA
	GACCGAGAGTATTAAGAGGTATTGGGCTCAACGTCTTCACTCTTCGGTA		TGAGTG		QRLHSSVDGKTLAY
	GATGGCAAGACTTTGGCCTACATGGATTATTTCCCTCAGGGTAATCTTTG		AGTTTAT		MDYFPQGNLWMSE
	GATGTCAGAGGATAAAGTGAGCCAGCGATCTTACGTCTTTGCTGACTGT		GTACTTT		DKVSQRSYVFADCVK

			GTCAAGTTACGGATTAATGCCATCCCAACTCGTGTCCGTGTTTCTCGTGG TCGACCTAATAAGGAAATGTGCTGTAGAGCGAAATGTTTTGATTCCCAG AGAATGCCAGCGTTTGAATCCTTGAACCACATAACGCAAGTGTGTCCGC GCACTCATGGCTCGCGAATCCAGCGGCATGATAAGATAGCCAAGTTCCT TTTTAAGAACCTGAACAATTGCCCATCTCGCTCGGTTCTGTATGAACCCC ATTTTGTGACGGTCGATGGCCTGCGGAAGCCTGACATTATCATTTACGA TGATAGCCACATGGTGGTTCTCGATGTCCAGGTGGTTTCCGACTCTGCTA ACCTGGAGAAGGAATTTGAGTGCAAGGCGAAGAAATATGCCAATGATG TTGCACTTCGATCTGCAATGCTAATTAAATACCCATTCATTAAGTCATTTT CTTTCGTTGCAGCTACCTACACAACAACCGTGGTCTGATAGCTAAATCCTCT GTCCAGGTCCTTCGCCAGCTGGGTCTCTCGCCGCGCTCCATAATGGTCTC CATCCTCATCTGCCTTGAGGGTACTTTGGAAACATGGCGCATCTTCAATC AGTCTACAATGAATGCCCATTAGTGCTTGAGTGATGCCTATCCTTTCTTT GATTAACTCTTACCATATACTTACCAGTTCTTACCCGTACTTACCCTGTAT ACTTACCTGTGTGCGTACCTGTGTACTTGTCCTTTAGCCGCCTTGTGTTTT TACCATTGGTACTTACCTTGTGTGGTTGCCCGATACTTACCTTGTATTTGC CTTGTAATTCTGCATGATATTTATTGTGTAGGTTCCTGATGCTTACCTGAT TTGTCCCCCTCATCATCTTTAGTTTCGTTCTATTTCACTCCATTATGGAGTT CCGTTTGTTTTTTGGTGGAGGTACAGCACCCTTTAAGCTGGAATTGAGT GAGTTTATGTACTTTGGATGGTTGTAATAAACTACCCGGAGGCAT (SEQ ID NO: 1110)	TTTTGGGG TAGCAATT GATCGATT CCCGCCTC CTCGTGGC GCTACCCT GGGATAAC CTCAAAGA AATTTGAC GGTAAAGC TAAGAGG AATTGATC ACACCGTG ACGAATAT CATCAGTC ACAGCTGC ACGAATCC	GGATGG TTGTAAT AAACTA CCCGGA GGCAT (SEQ ID NO: 1356)	LRINAIPTRVRVSRGR PNKEMCCRAKCFDS QRMPAFESLNHITQV CPRTHGSRIQRHDKI AKFLFKNLNNCPSRS VLYEPHFVTVDGLRK PDIIIYDDSHMVVLDV QVVSDSANLEKEFEC KAKKYANDVALRSA MLIKYPFIKSFSFVAA TYNNRGLIAKSSVQV LRQLGLSPRSIMVSILI CLEGTLETWRIFNQS TMNAH (SEQ ID NO: 1478)
R2	R2LcB	JN937619	Lepidurus couesii	TTTTGGGGTAGCAATTGATCGATTCCCGCCTCCTCGTGGCGCTACCCTGG GATAACCTCAAAGAAATTTGACGGTAAAGCTAAGAGGAATTGATCACAC CGTGACGAATATCATCAGTCACAGCTGCACGAATCCAGATAGATATATA ACAGGCGAGTAATTCTTTTCGAAGCTCGTCAGTAATCTTCCCAGATGTCT GGAAAGAGCAGCAAACCGAGAACTGTATCCAGCGGATCCAGTTCCCAA GAGACTCCTCCAAGTGGGTCCAACGCATGTGATATATGCGGGAAGTGCT TCATGAAACCAGTTGGTTTATCACTACACATGTCCAAAGTCCATCCTACT CAGTACCATGCACGCTTGGAAAAGAATCAACCGAAAGCCAAGAAATTTA GATGGACGGATGAAGATTTGTATTTTTTAGCGAAGAAGGAAGCTGAAC TTTTACTTTTGGGTGGCATTAAGTTCATGAACAAGGAACTGGCAGAATT CTTTCCGGAGAAATCTGTTGACCAAATCAAAGGCCAAAGACGTTCCGAG ACATACAAACAGCAAGTTGTATCAATTCATTCTGAACTGCTTAAGTTGCA GGCTGTAGCTGATTCTCCTCCACCATCGAGAATTCCAGCGAAAGAAGTC TCGGCCTGGCTTGATTTCCTTCTGGCCCTTCCCAAAACCAAGAATAAATT TAGTGAGGACAAGTTAGACCAACTTATACGTACCGCGCAAGAAGGCACT CCGGTCCTAAACGATCTAGATTTATATCTAAGAGAAGTTCTAGTTCAGCC CACAAGACAGGGAGAGAGACAGGCGAAGCCTCTCCCACCACCCAAATC			

		CAGTCGAGAGAAAAGAGACAGAGAATATGCTCGTGTCCAGAATTTTTAT	AGATAGAT	TATCTAC	DIDEFWKAIFESQSPP
		CGCAAGAATAAAACCGCATGTGTGAACGCCATCCTTGATGGAAATAAGA	ATATAACA	CTGACA	DAEPVSYVVDEEPKN
		AGTGCGAAAACAAAATACCTGACATTGACGAATTCTGGAAGGCCATATT	GGCGAGT	TATATTT	IWSWISFFEMNRNY
		TGAATCTCAATCTCCTCCTGATGCAGAACCAGTGAGTTACGTTGTAGAC	AATTCTTTT	ATCTAAC	PDTSTSPGPDGVTAR
		GAGGAACCGAAGAATATTTGGAGTTGGATTTCGTTCTTCGAAATGAATC	CGAAGCTC	CACCTAC	MLRSIPARVLNKLLNL
		GTAACTACCCTGATACGTCGACCTCACCCGGTCCAGACGGCGTGACGGC	GTCAGTAA	CTATGAT	LLFIEDLPAVFKCHRT
		TCGTATGCTCCGTTCCATCCCTGCGCGTGTCCTAAATAAATTGTTAAACC	TCTTCCCA	GACTCC	VLIPKVDNPALPGEFR
		TATTACTCTTTATTGAAGACTTACCTGCAGTGTTCAAGTGCCATCGTACC	G (SEQ ID	CGCGGA	PITISSIIVRQLNKIIAA
		GTTCTGATACCTAAAGTTGACAATCCAGCATTACCGGGAGAAATTTCGGC	NO: 1234)	AACTCTC	RVSEGVPINPPRQKAF
		CTATCACGATTTCTTCCATAATTGTTCGGCAGCTGAATAAAATCATCGCT		ACTTACC	RQIDGCAENVFLLDFI
		GCTCGCGTCTCAGAGGGCGTTCCCATAAATCCTCGTCAGAAGGCCTTTC		TTATTAC	LRDAKTKIKSLSLATV
		GGCAAATAGATGGATGTGCTGAGAATGTTTTCCTTTTAGATTTCATTCTT		CCACTTG	DIKKAFDSVSHHSIFR
		CGTGACGCAAAAACGAAAATTAAGTCCCTTTCGTTGGCTACTGTCGACA		GTCTTTT	AIRGARCPENLVNYI
		TTAAGAAAGCATTTGACAGTGTGTCACATCATTCCATTTTCCGAGCGATT		ATTTTCT	QNSYSGCTTQISVGG
		AGAGGAGCGAGATGCCCTGAGAATTTGGTGAATTATATCCAGAACAGC		CGTTCCT	SISTTKILMNRGVKQ
		TATTCGGGTTGTACGACCCAAATTTCCGTAGGCGGCTCTATATCTACAAC		TATTACT	GDPLSPVLFNLVINEII
		AAAAATCCTGATGAACCGTGGTGTAAAACAAGGGGACCCACTATCTCCT		TTGTTCC	RKLPASIGYPINSELSI
		GTCCTCTTTAACTTAGTGATAAACGAAATTATTCGCAAGCTGCCGGCCTC		TTTGGT	NCIAYADDLILVANTR
		GATTGGATATCCAATCAATAGTGAGCTCTCGATAAACTGTATTGCCTATG		GTAGGG	EGLKLLLNLLNEELPK
		CGGATGACCTGATACTAGTTGCAAATACTAGAGAGGGTCTGAAGCTACT		TTCTCTG	RGLELNASKCFGLSLT
		ACTCAACCTATTGAATGAAGAGCTCCCTAAGCGAGGATTAGAATTGAAT		GTTTTTG	ALGKLKKTHLCTSDQ
		GCATCTAAGTGTTTCGGCTTATCCTTAACAGCACTCGGGAAATTGAAGA		GAACGG	LDLHGTLIKNLTAEES
		AGACTCACCTGTGTACCAGTGACCAACTCGATTTACATGGGACTCTAATT		CTTCCTT	WVYLGVPFSHIGRSK
		AAGAATCTTACCGCTGAGGAATCCTGGGTTTATTTGGGTGTTCCTTTCTC		AGCCGG	SFSPDLEALLNKLQKS
		ACACATTGGGCGCTCTAAGAGCTTCAGTCCAGATCTTGAAGCTCTTTTAA		AATTTTG	PLKLQQKLFALRVYLI
		ATAAACTCCAGAAATCGCCGCTTAAACTGCAACAAAAGCTTTTTGCTTTG		TCTGAT	PRLLHGLVLSRVAIGE
		CGAGTTTATCTAATTCCTAGACTGCTCCACGGTTTAGTCCTATCAAGAGT		GTATCTT	LKIMDKLILKHLRVWL
		GGCCATTGGCGAATTAAAAATAATGGATAAGTTAATTTTGAAACACTTA		GCTTGT	RLPKDTPLGFFYSPVK
		CGCGTATGGCTTCGTTTACCGAAAGATACCCCTTTGGGATTTTTCTACTC		GTCCTT	LGGLGIKNLRTNVLKC
		ACCGGTGAAATTGGGCGGTTTGGGGATTAAAAACTTACGTACAAACGTT		GAAATA	RKQRIERMLVSPDDV
		TTGAAGTGCAGGAAACAACGCATTGAGCGAATGCTAGTCTCCCCTGATG		TACGAC	VRLVAESEIFLKETDKL
		ATGTGGTGCGCTTGGTTGCAGAGTCGGAGATTTTCCTGAAAGAGACGG		CCAGGC	KDLLTINGMCLDXRN
		ACAAGTTAAAGGACCTACTTACTATTAACGGAATGTGCCTAGATTANCG		TTGCGT	VPRTGKNNKFWSER
		GAATGTGCCTAGAACGGGAAAAAATAACAAGTTTTGGAGTGAAAGGCT		CATTTAG	LYTSFDGKTLAYSEYF
		ATATACGTCATTTGATGGTAAAACACTTGCCTATAGCGAGTACTTTACTC		GCTCTG	TQGGGWIREDKILQP
		AAGGCGGCGGCTGGATCCGTGAAGATAAGATCCTTCAACCTGCCCATGT		GGAAA	AHVFAECIKLRINALP
		TTTCGCTGAATGCATCAAGTTACGTATTAATGCTTTACCCACTAAATCTA		(SEQ ID	TKSRVAHGRPTKDRS
		GAGTTGCACATGGCCGTCCAACTAAAGATCGTTCCTGTCGTGCCGGATG			CRAGCLDVQKVPAIE

			TCTTGACGTTCAAAAAGTGCCAGCTATCGAAACAATCAATCATATAGCAC AAGTTTGTCCTCGGACGCATGGTGCAAGAATAAAAAGACACCGATCGGCT CGTTCAATTCCTTTCTCTCAACCTTAGGAAAAATCCAAAAAGAAATGTTT TAGTGGAGTACAATTTCCGTACAGTCGCTGGTATCCGTAAACCGGATAT CATTGTTATTGAAGATACCCGTGCAGCTATCCTTGACGTCCAAGTTGTTG GTGATTCGTCCAATTTGGAGATGGAATACCTGGAGAAATCCAGGAAATA TTCTAATGACGCTACTTTATCAATGCGTATCAATGCGCTTCAAAAACTTT ATCCAACGGTCACTAGCCTAACTTTTCATGCAGTTACCTTCAACAATAGA GGCCTGATTGCCAAATCCACAGTGGCGGCCCTCCGGATGCTCGGCGTCC CACCTAGATGCATAATGATCCTTTGTGTGATTAGCCTTGAGAAGACCCTG GAAGTCTGGAGGATGTTTAACCAATCAACTGCTTCCGCTCGGGAAATGAT AATCGCTCCATCCTGCAACTAATTATGAATGCAAATCTGTTAAGTGACAT TAGTGATACTTACCTGATACTTACCCTGGTATTTATTTGACCTATACTTAC CCTGGTATCTACCTGACATATATTTATCTAACCACCTACCTATGATGACTC CCGCGGAAACTCTCACTTACCTTATTACCCACTTGGTCTTTTATTTTCTCG TTCCTTATTACTTTGTTCCTTTGGTGTAGGGTTCTCTGGTTTTTGGAACGG CTTCCTTAGCCGGAATTTTTGTCTGATGTATCTTGCTTGTGTCCTTGAAATA TACGACCCAGGCTTGCGTCATTTAGGCTCTGGGAAA (SEQ ID NO: 1111)	GGTTGGGG GCCTTCTC GTGGCGG AGTCGTGA GTAAGGG GTATAGGG GTAAGGG ACACCACG GACCGAG AACGGTTA CCGCTCAA GGCGAGT GGTGGAA GGCATAAA ATCGTAAC GCCGCCCT CCGACCTG CTCCTGAA ACTAATGC	TGATGG TGGGTT ACTCGC CTCTGT GTAACA GGCAAA TGAAAG CTGCGC AAGCAG TCGATG AGCCAA AGCCGC ACAGCC CCCGAC TGGGGT ACAGGG CAGCCC TGGGCT ATGCCC	TINHIAQVCPRTHGA RIKRHDRLVQFLSLNL RKNPKRNVLVEYNFR TVAGIRKPDIIVIEDTR AAILDVQVVGDSSNL EMEYLEKSRKYSNDA TLSMRINALQKLYPT VTSLTFHAVTFNNRG LIAKSTVAALRMLGV PPRCIMILCVISLEKTL EVWRMFNQSTASAR K (SEQ ID NO: 1479)
R2	R2Nvec-A	Nematostella vectensis	GGTTGGGGCCTTCTCGTGGCGGAGTCGTGAGTAAGGGGTATAGGGGTA AGGGACACCACGGACCGAGAACGGTTACCGCTCAAGGCGAGTGGTGG AAGGCATAAAATCGTAACGCCGCCCTCCGACCTGCTCCTGAAACTAATG CCAACCAACTGACTGTGGGGCTAACCTCCCCAGAGTCAGGATGCTGAG GGGTACTGGGAATATGAATGACAAGAGGGACGGATCGGCCACTGCTGA TCCTACCTCAGCGCTTTTGGGGGCTGTCGGTGATGGCAGCCTGGTATGC AATCTGTGTGGCCTTGCTTGCAAGTCTCGTGGGGGGCTCTCGATTCATC GAAGGAGCAAGCATGCCACGGTGTACCATGCGGAAAGGCAGCCCGCTC CGCGGGCTAAGGCGAGATGGACGAATGATGAAATGATACTTGTGGCCA GAAAACAAATAGCTTCGGAAAAGAGTAGGTGTAGCGCAGTAGTTGAAG GGATGCGAGAAGCAGTTCCACATAGAACTTTTGACGCCGTCAAAAGTCT CAAAACCAAGAATAGGAACTACACACGGATTCTGGAGCAGATTCGCGCT GAATGCTCCGAGGAAGAGGTGATTGAATCAGGAGTCCTTAAGGATAGG ACGGAGAACGTGTGTGTGCAAACAACATCTAACGTTCCCGGTTCGGCAG GACGGGCAGCATCGGTGGAGCTAGAGGGTAATATCCAGGTGGGCCATC AACTCGCCCAAAAAAACCATGGCAGGCAATAATAGTAGGAAACAGCCTG CCAATCATACCAACTGGGCTGAGTTCAATATTGAGGAGGGAAATATTAC CCTCCGAAAAAGCAAAAGAAAGGCTAATGGTATGCCCGATGCGACCCA CCGCCCCGGCCCACCGACTGTCGACAGTCTTAAGCACCCAGTATGCTTG			

	TTGCAGGGTGCAGCCGACAAAAGGGACGAACCGCACACAGTGGAACAA	CAACCAAC	GAAGTC		EKLFDAIDSSLLSVEVE
	TTGTATTATAACATTGAGGAAGGAATGCCTCTTGCCGAAGAGCAACAGT	TGACTGTG	TTTTGAC		LGRIVPGCPDEETRQ
	GGTCAGAAAAGCTTTTCGATGCAATCGACTCATCTTTGTTGTCTGTCGAG	GGGCTAAC	AGGTCC		LIDREFLDFIHSYSREK
	GTCGAGCTGGGGAGAATCGTTCCTGGTTGCCCCGACGAAGAGACCCGA	CTCCCCAG	AAACTTC		PPQRGLAKSKPPPKG
	CAGCTCATTGATAGAGAATTCTTGGATTTTATCCATTCCTATTCGAGGGA	AGTCAGG	ACCCTTG		PKSLRRQQYRQLQRL
	AAAGCCGCCACAAAGAGGTCTCGCTAAGTCCAAACCACCACCCAAAGG	(SEQ ID	CCGCCA		WDKNRSAAAEQALT
	ACCAAAGAGCCTCAGACGCCAGCAATATCGGCAACTCCAGACCAGTGACTA	NO: 1235)	GTAGGG		GKWQEVRTAAGVPL
	GGACAAAAACCGATCCGCAGCAGCCGAACAGGCCCTAACAGGTAAATG		CACCAG		SLMEVPWREIFETPS
	GCAAGAAGTTCGGACTGCTGCTGGGGTCCCATTATCGTTGATGGAGGTC		GGCCAG		TTDVREPAPAGPVL
	CCGTGGCGGGAGATTTTCGAAACACCATCTACGACGGACGTCCGAGAG		GTGCAG		WQLLRPVTIAEVEDA
	CCGGCCCCAGCTGGGCCGGTTCTGTGGCAGCTTCTCAGACCAGTGACTA		GTGGGC		ISSKKSASGPDGVPC
	TCGCAGAAGTAGAGGACGCAATTTCCTCCAAGAAATCAGCTAGCGGTCC		GCTTTGT		AALQTMGAASLAAH
	AGATGGAGTGCCTTGTGCCGCCCTTCAAACCATGGGCGCCGCCTCCCTA		TCTATTT		FNLWLLAGTQPKRLT
	GCAGCTCACTTTAACTTGTGGCTGTTGGCCGGTACTCAGCCCAAACGCCT		GGTTTC		ECRTIFVPKEVNTHLP
	AACAGAATGCCGAACAATATTTGTGCCTAAGGAGGTAAATACTCACCTG		GCTTAAT		LHHRPITIGSVVVRLF
	CCCCTCCATCATAGGCCAATCACTATCGGCTCGGTGGTTGTAAGACTCTT		TTTTGTT		HQILGKPMEAVLPLG
	CCACCAGATTCTGGGAAAGCCAATGGAAGCAGTACTGCCATTGGGGTC		AAATTTT		SGQRGFRKGDGICQ
	GGGCCAGCGCGGTTTTCGAAAGGGTGACGGAATATGCCAGAACATCTG		TCCCGT		NIWLLHTLIRRSTDLL
	GCTCCTCCATACTCTAATCAGACGCAGCACTGATTTGCTGAGGCCACTCA		GCGCCC		RPLKLVFLDVKKAFDS
	AGCTGGTCTTCTTGGATGTGAAAAAGGCCTTCGATTCGGTTAGTCACGA		ACCCTTT		VSHESLLIAAKRLGVP
	ATCCCTTTTAATTGCGGCGAAGCGTCTCGGAGTGCCCGGGCCGCTCCTA		TTAACCC		GPLLTYINELYSRSETV
	ACGTACATCAATGAACTGTACTCTCGTTCGGAGACAGTCTCTTTGAAGTTG		TTTTAAC		FEVGGESSGSVKVSQ
	GGGGCGAGAGCAGTGGCTCTGTTAAAGTATCCCAAGGGGTAAAGCAAG		ACCAATT		GVKQGDPLSSTLFNC
	GGGATCCATTGAGTTCGACGCTCTTCAACTGCGTCATTGACTGGGCGGT		TTTTTTA		VIDWAVSDLDPHIGV
	CTCAGATCTCGACCCCCACATAGGCGTCTTACTAGGGGAGTCTCGAGTC		CAACCCT		LLGESRVSFLAYADDL
	TCCTTTCTAGCTTACGCTGATGACCTGGTTCTCCTATCGGAGACCGAGGC		CTACTCA		VLLSETEAALTSQLNS
	TGCACTGACAAGCCAGCTGAACTCAATTGAAAAGAGTCTAGCGCACTGT		ATAATCC		IEKSLAHCGLKLSTGD
	GGTCTCAAGCTCAGTACGGGGGACAGTGGTAAATCTGCTTCGCTGAACA		AAATGA		SGKSASLNIVIDGKAK
	TCGTGATTGATGGTAAAGCTAAGCGGTGGGTTGTAAACCCTACCCCGTT		ATAAAA		RWVVNPTPFLRASG
	CCTACGTGCGTCTGGTGGCGAAATTCGATCTCTCGTCGCAAATGAGACG		GCGGCA		GEIRSLVANETYKYLG
	TACAAGTACTTGGGCATTAACATCGGAGCCCAGGGCGTGAAAGCTGCT		TAACAG		INIGAQGVKAAEYNA
	GAATACAATGCATTCAAAGAAGCACTCGATAACTTGTCTCGAGCTCCACT		GTGAAC		FKEALDNLSRAPLKP
	TAAGCCGCAGCAGAGGTTATTCCTGCTGAAGACCTATCTGTTGCCGCAA		AC (SEQ		QQRLFLLKTYLLPQLH
	CTACACCACTCACTGGTATTGTCAAGGACAACTGGCAAACTATTGAACA		ID NO:		HSLVLSRTTGKLLNSL
	GCCTCGATGCTCTCGTAAGAAAAGCTGTACGGGGATGGCTTAAATTGCC		1358)		DALVRKAVRGWLKL
	ACACGACACTCATAGGGCCTTCTTTTATGCTCACCAGGCCGACGGGGGA				PHDTHRAFFYAHQA
	CTCAATGTCCCAAGTCTGTACCATCTAATACCTCTCCTAAGGAGGAGTAG				DGGLNVPSLYHLIPLL
	GTATGAACGCTTGACGAGGGTCGAAGACCCTGAGATAAGGGAAGTGTC				RRSRYERLTRVEDPEI

			CCGAACCGACTACTTCAAAAGAGTCTTGGGTGCCGCGGCGGCCAC				REVSRTDYFKRVLGA
			CACGGTTGCAGGGCATCGAATTGATTCCAAAGTAACGTTGCGCCTAGCT				AAAATTVAGHRIDSK
			TGGCGCGAGGCCCTTTACGCTTCAGCTGACGGGCGGCGGGTCTCTCCCAAT				VTLRLAWREALYASA
			GCCCGTTGGTACCCGAGGTTCACTCCTGGGTTACCGATGTTTCCGGTCTT				DGRGLSQCPLVPEVH
			CAGACCGGATCGCAGTACATCTCTGCTGTTCGCTTAAGGGGGGGCCTTGC				SWVTDVSGLQTGSQ
			TTCCCACGGCGGTAAGGAAATCGAGGGGTCGCGGGGGGTGTTAACTCGA				YISAVRLRGALLPTAV
			ATTGCGACTGTTGCGGTCGTGGACAACCAGAGTTTCTTGGGCACGTTCT				RKSRGRGGVNSNCD
			CCAGACGTGTCCTCGGACCTGGGGGGTCTCGCATCAGTCGTCACAACCAC				CCGRGQPEFLGHVL
			GTCCTCTCCCTTATCGCTAAAGCGTGCCGTTCGCGAAGATGGCAGGTGC				QTCPRTWGSRISRHN
			TGGAAGAGCCGATCATCCAGACTCCTGCGCAGTTGCTCAAGCCAGACTT				HVLSLIAKACRSRRW
			GGTGATCTGGAACCATCAAGCGGCGTACGTCGTTGACGTTAGCGTACCG				QVLEEPIIQTPAQLLK
			GGTGATAACACCCCGCTCAGTACGTGTCACAACAGGAAAGTCGCATATT				PDLVIWNHQAAYVV
			ACTCTGGTGAGAGCGTTCGTGAATGGGTGCGTTCCAAGACTGGGCATA				DVSVPGDNTPLSTCH
			ATCCTACAGTGTCCTCTGTAGTGATTAACTGGCGCGGTGCCATGGCAAA				NRKVAYYSGESVRE
			GGAGTCCTACCGCCTTTTGACGAAAGACCTTCGTCTAGCGAAGACCCTG				WVRSKTGHNPTVSS
			CCTCGTCTGCTGGTACTAAGAGTTCTGGAAGGCGGCCATGGAATCTGGC				VVINWRGAMAKESY
			TCAATTTCCATCGGTCGACCTTCGCCGTGGGAGTAACTTGATGGTGGGT				RLLTKDLRLAKTLPRL
			TACTCGCCTCTGTGTAACAGGCAAATGAAAGCTGCGCAAGCAGTCGATG				LVLRVLEGGHGIWLN
			AGCCAAAGCCGCACAGCCCCCGACTGGGGTACAGGGCAGCCCTGGGCT				FHRSTFAVGVT (SEQ
			ATGCCCGAAGTCTTTTGACAGGTCCAAACTTCACCCTTGCCGCCAGTAG				ID NO: 1480)
			GGCACCAGGGCCAGGTGCAGGTGGGCGCTTTGTTCTATTTGGTTTCGCT				
			TAATTTTTGTTAAATTTTTCCCGTGCGCCCACCCTTTTTAACCCTTTTAACA				
			CCAATTTTTTTTACAACCCTCTACTCAATAATCCAAATGAATAAAAGCGG				
			CATAACAGGTGAACAC (SEQ ID NO: 1112)				
R2	R2Sm-A	Schistosoma mansoni	ATGTTTTAATTTATTTTTGAACTACTACTGTCTGAGTGCTTCTTACAACCT	ATGTTTTA	TAGCCC		MPVSTGAETDITSSLP
			GAAGGCTCAGAAACTACCCACTTTTTGCTGTTTATCCACACAACACAGTTG	ATTTATTTT	CCTTCAC		IPASSIVSPNYTLPDSS
			TGAATCTATTCTCCAAATATTCCTTGTGCTTTTGTCAACATTATTCTATACC	TGAACTAC	TCTTAGA		STCLICFAIFPTHNILLS
			AACTGTACCACCTACTTCTTCATCTCACGTTTTAATTCTGGTCTAATTTTCT	TACTGTCT	CATTCCC		HATAIHHISCPPTPVQ
			CATCATTAGTCACGGAGAGGGCCTATGAACGGTCCGTGACGCGAAATTC	GAGTGCTT	CCACTGT		DGSQQMSCVLCAAA
			AATCCACGAATTCGTCCTCTTCTGCTAGTGGTCCCCGAAATACGGTTCCT	CTTACAAC	TGTTGCT		FSSNRGLTQHIRHRHI
			CTGGCCTGTCAGTTGTGTTAAAACTATATAATAACGATGCCGGTCTCAAC	CTGAAGGC	TATCTTC		SEYNELIRQRIAVQPT
			CGGCGCAGAAACTGACATAACCTCTTCTTTGCCTATTCCTGCATCCTCAA	TCAGAAAC	ATGTTTT		SRIWSPFDDASLLSIA
			TCGTCTCGCCAAACTACACACTCCCTGATTCCTCTTCAACCTGCCTTATAT	TACCCACT	TGTGTTA		NHEAHRFPTKNDLY
			GTTTCGCTATCTTCCCCACCCACAACATACTCCTCTCCCATGCCACTGCAA	TTTTGCTG	ATTGACT		QHISTVLTRRTAEAVK
			TCCACCATATTTCTTGTCCTCCTACTCCAGTGCAAGACGGTTCTCAGCAG	TTTATCCA	GCTCTCT		RRLLHLQWSRSPTAI
			ATGTCTTGTGTTCTTTGCGCCGCCGCTTTTTCATCTAACAGGGGACTAAC	CAACAACA	TCTGGG		TTSSNNHTTTDIPNTE
			ACAACACATTCGCCACCGGCATCATCTCCGAATATAACGAACTAATCAGA	GTTGTGAA	TTGATGT		ARYIFPVDLDEHPPLS
			CAACGAATTGCAGTGCAGCCGACGTCTCGCATATGGTCACCATTCGATG	TCTATTCTC	CTGATT		DATTPDASTHPLPELL

		ATGCTTCTCTACTATCAATCGCTAACCATGAAGCCCATAGATTCCCCACG	CAAATATT	GTCTCTC		VILTPLPSPTRLQNISE
		AAGAATGACTTATACCAACACATCAGCACTGTATTAACACGCAGGACGG	CCTTGTGC	TCTCTTT		SQTSHESNRNSMHT
		CAGAGGCCGTCAAACGCCGACTCCTCCACCTACAGTGGTCCAGATCACC	TTTTGTCA	CCATATT		PPTYACDSDESLGVT
		CACAGCGATTACTACCTCTTCGAATAATCACACAACCACAGACATCCCCA	ACATTATT	GCTTGC		PSSTIPSCFHSYRDPL
		ATACCGAGGCCCGATATATTTTTTCCGGTAGACCTAGACGAACATCCACC	CTATACCA	TCTCCCC		AEQRSKLLRASASLLQ
		ATTGTCTGATGCCACACAACCCCCGACGCATCGACACATCCACTCCCAGAAC	ACTGTACC	GCTTACT		SSCTRIRSSSLLAFLQN
		TCCTTGTCATCTTGACACCGCTTCCATCCCCGACTAGACTACAAAACATA	ACCTACTT	TCCAATA		ASTLMDEEHVSTFLN
		TCCGAATCACAGACCTCCCATGAATCCAATAGGAACTCAATGCATACACC	CTTCATCT	GTTGTC		SHGEFVFPRTWTPSR
		GCCAACGTATGCCTGCGATTCGGATGAGTCACTAGGGGTTACTCCCTCA	CACGTTTT	ATATTAT		PKHPSHAPANVSRKK
		TCAACTATCCCCTCATGCTTCCACAGTTATCGGGACCCCCTAGCTGAACA	AATTCTGG	GTCTTTG		RRKIEYAHIQTLFHHR
		AAGAAGCAAACTCCTGAGGGCATCCGCCAGCCTACTACAAAGCAGTTGT	TCTAATTTT	TTTACTT		PKDAANTVLDGRWR
		ACTCGCATACGGTCCTCCAGCCTGCTCGCCTTCCTCCAAAACGCATCCAC	CTCATCAT	GCCATG		NPYVANHSMIPDFD
		ATTAATGGACGAGGAACACGTGTCCACCTTCCTCAATAGTCATGGAGAA	TAGTCACG	TCTAAC		CFWTTVFTKTNSPDS
		TTCGTCTTCCCTAGAACATGGACCCCATCCCGACCCAAACACCCCTCCCA	GAGAGGG	GACAAT		REITPIIPMTPSLIDPIL
		CGCCCCAGCTAATGTTTCTAGGAAGAAAAGGAGGAAAATAGAGTACGC	CCTATGAA	TACTTTA		PSDVTWALKEMHGT
		ACACATCCAGACACTCTTCCACCACCGTCCCAAAGATGCCGCCAACACCG	CGGTCCGT	TCTACCT		AGGIDRLTSYDLMRF
		TTCTAGACGGTCGGTGGAGAAACCCCTATGTCGCAAACCATTCAATGAT	GACGCGA	TAGTTG		GKNGLAGYLNMLLA
		TCCAGACTTCGACTGCTTCTGGACAACAGTCTTTACTAAAACAAATTCCC	AATTCAAT	GTCCTCT		LAYLPTNLSTARVTFV
		CAGACAGCCGGGAGATTACTCCAATCATCCCTATGACTCCCTCTCTCATT	CCACGAAT	TGGTTT		PKSSSPVSPEDFRPIS
		GACCCGATCCTCCCCCTCTGACGTCACATGGGCGCTGAAAGAAATGCATG	TCGTCCTC	GGTTGC		VAPVATRCLHKILAKR
		GCACGGCCGGTGGGATTGATCGTCTAACATCGTACGATCTGATGAGATT	TTCTGCTA	CTTCATG		WMPLFPQERLQFAF
		CGGGAAGAATGGTCTTGCTGGATATCTCAACATGCTACTCGCTCTTGCAT	GTGGTCCC	TGTTCAT		LNRDGCFEAVNLLHS
		ACCTTCCCACTAATCTCTCAACAGCACGGGTAACTTTCGTCCCCAAGTCA	CGAAATAC	GGCGGA		VIRHVHTRHAGASFA
		TCAAGTCCTGTGTCACCTGAGGACTTCCGTCCCATCAGTGTCGCTCCAGT	GGTTCCTC	ATCTGAT		LLDISRAFDTVSHDSII
		AGCCACTAGGTGCCTGCACAAAATTCTAGCAAAGAGATGGATGCCGCTC	TGGCCTGT	GTTTATA		RAAKRYGAPELLCRYL
		TTTCCACAGGAACGACTTCAGTTCGCTTTCCTAAACCGAGATGGATGCTT	CAGTTGTG	ATGACT		NNYYRRSTSCVNRTE
		TGAAGCAGTTAATCTTCTGCACTCGGTCATACGGCACGTCCACACCCGCC	TTAAAACT	ATTCCTA		LHPTCGVKQGDPLSP
		ATGCAGGAGCATCCTTCGCCCTGCTCGACATATCACGGGCCTTTGACACT	ATATAATA	CTACCAC		LLFIMVLDELLEGLDP
		GTATCACATGACTCCATCATCAGAGCGGCGAAAAGATATGGGGCACCTG	ACG (SEQ	CATTACA		MTHLTVDGESLNYIA
		AACTGTTATGCCGCTACCTCAATAACTATTACCGACGTTCAACCAGCTGC	ID NO:	ACTATTA		YADDLVVFAPNAELL
		GTCAACCGCACTGAATTGCATCCTACGTGTGGGGTGAAGCAAGGAGAC	1236)	TTATTAT		QRKLDRISLLLHEAG
		CCCCTGTCGCCACTCCTCTTCATCATGGTTCTCGACGAATTACTGGAAGG		CACTATT		WSINPEKSRTLDLISG
		TCTAGATCCAATGACCCACCTAACAGTTGATGGAGAGAGCTTGAACTAC		ATTAAC		GHSKITALSQTEFTIA
		ATAGCTTATGCTGACGATCTCGTAGTTTTCGCTCCAAATGCAGAACTCCT		ATTATTA		GMRIPPLSAADTFDY
		TCAACGGAAACTCGATCGGATCTCCCTACTTCTACACGAGGCTGGATGG		TTACTTC		LGIKSNFKGRCPVAHI
		TCGATTAACCCTGAAAAAAGCCGGACCCTGGACCTAATCTCTGGTGGCC		TACAATT		DLLNNYLTEISCAPLK
		ATTCCAAAATCACAGCGCTCTCTCAGACAGAATTCACCATCGCGGGGAT		AGTATT		PQQRMKILKDNLLPR
		GCGTATACCACCGCTTTCCGCCGCCGACACCTTCGACTATCTGGGTATCA		ATGGCT		LLYPLTLGIVHLKTLKS

			AATCCAACTTCAAGGGCCGATGCCCAGTGGCCCATATTGACTTATTGAA		ACTCCTT		MDRNIHTAIRKWLRL
			CAACTACCTCACGGAAATATCGTGCGCTCCACTTAAGCCGCAGCAGCGC		TCAGCA		PSDTPLAYFHSPVAA
			ATGAAGATCTTGAAAGATAATCTACTCCCTCGACTCCTCTACCCCCTGAC		CACCAA		GGLGILHLSSSVPFHR
			TCTAGGAATAGTACACCTGAAAACCCTGAAGTCAATGGACCGAAATATC		TAAAAT		RKRLETLLSSPNRLLH
			CACACGGCCATAAGGAAATGGTTGCGGCTACCCTCCGACACCCCGCTAG		CTCAATC		KLPTSPTLASYSHLSQ
			CATATTTTCACTCACCCGTCGCTGCCGGAGGCCTAGGGATCCTCCATCTG		AAACAT		LPVRIGHETVTSREEA
			TCCTCATCGGTTCCATTCCACCGTCGAAAACGTCTAGAAACCCTCCTATC		CTCACTT		SNSWVRRLHSSCDG
			TTCACCGAACCGCCTACTGCACAAGTTGCCAACTTCCCCAACACTAGCTT		ATTAAA		KGLLLAPLSTESHAW
			CTTATTCACACCTTAGTCAACTGCCAGTTCGAATTGGGCACGAGACCGTA		CTCTCTA		LRYPQSIFPSVYINAV
			ACGTCTAGAGAAGAGGCTTCCAACAGCTGGGTGAGACGATTACATTCGT		TTTCCCC		KLRGGLLSTKVRRSR
			CCTGCGACGGGAAGGGACTACTCCTAGCACCACTAAGCACCGAGTCCCA		TTCGTTA		GGRVTNGLNCRGGC
			TGCATGGCTGCGCTACCCCCAGTCTATTTTTCCAAGTGTTTACATCAACG		TAAACTT		AHHETIHHILQHCALT
			CCGTTAAATTACGAGGTGGCTTACTATCCACCAAAGTCAGGAGATCTCG		ACAATTC		HDIRCKRHNELCNLV
			CGGAGGTAGAGTGACGAATGGCCTGAACTGTCGAGGCGGTTGCGCCCA		AGTTTA		AKKLRRQKIHFLQEPC
			TCATGAAACAATCCACCACATTCTGCAACATTGCGCCGCTCACCCATGACA		ACCGAA		IPLEKTYCKPDFIIIRDS
			TCAGATGCAAACGCCATAACGAACTATGCAACCTTGTGGCAAAGAAACT		TATCTCT		IAYVLDVTVSDDGNT
			GCGTAGGCAAAAAATCCATTTCTTACAGGAGCCCTGCATTCCTCTAGAA		CTTTTAC		HASRLLKISKYGNERT
			AAAACTTACTGCAAACCTGATTTTATAATTATACGTGATTCAATTGCTTAT		AAATCTT		VASIKRFLTSSGYIITS
			GTTCTAGACGTCACTGTATCGGACGGAAACACCCACGCCAGCCGCC		AAGTAT		VRQTPVLTFRGILERA
			TGTTAAAAATATCAAAATACGGCAATGAGCGAACCGTCGCATCGATCAA		GTAATTT		SSQSLRRLCFSSRDLG
			GCGATTCCTCACATCCAGTGGATATATCATTACCAGTGTTCGACAAACAC		TGTGCC		DLCLSAIQGSIKIYNTY
			CAGTCCTTACATTCAGAGGTATTCTGGAGAGAGCAAGTTCACAATCCCT		AAGCCC		MRGTQRLNE (SEQ
			ACGACGCCTATGTTTTTCGTCCCGTGACCTCGGTGACCTTTGCCTGAGTG		ATTTGG		ID NO: 1481)
			CGATTCAAGGCTCAATTAAAATATATAATACCTATATGAGAGGAACCCA		GTCTGT		
			ACGGCTGAACGAATAGCCCCCTTCACTCTTAGACATTCCCCCACTGTTGT		ACAATTT		
			TGCTTATCTTCATGTTTTTGTGTTAATTGACTGCTCTCTTCTGGGTTGATG		GATACTT		
			TCTGATTGTCTCTCTCTTTCCATATTGCTTGCTCTCCCCGCTTACTTCCA		AAAAAT		
			ATAGTTGTCATATTATGTCTTTGTTTACTTGCCATGTCTAACGACAATTAC		AAATGT		
			TTTATCTACCTTAGTTGGTCCTCTTGGTTTGGTTGCTTCATGTGTTCATG		TAT (SEQ		
			GCGGAATCTGATGTTTATAATGACTATTCCTACTACCACCATTACAACTA		ID NO:		
			TTATTATTATCACTATTATTAACATTATTATTACTTCTACAATTAGTATTAT		1359)		
			GGCTACTCCTTTCAGCACACCAATAAAAATCTCAATCAAACATCTCACTTAT				
			TAAACTCTCTATTTCCCCTTCGTTATAAACTTACAATTCAGTTTAACCGAA				
			TATCTCTCTTTTACAAATCTTAAGTATGTAATTTTGTGCCAAGCCCATTTG				
			GGTCTGTACAATTTGATACTTAAAAATAAATGTTAT (SEQ ID NO: 1113)		TTTTTGGT		MSQKRRPEKAVPDE
R2	R2Tc	EU854578	TTTTTGGTCTGGCATTTGATCGTTTCCGCCTCCTCGTGGCGCCAGACTGG		CTGGCATT		GATAHDVAQPDKSK
		Triops cancrifo rmis	GTAAGCTGATTTATCAGTGAGCTAAGAAAGAAACGATCACCGCAGGAGTCC		TGATCGTT		CSVCGETFKGPASVT
			ATCTACCTGCGCTGCACGTGATTTCATTGTGCACTTGGCGAGTTATCCCT				

	TGTGGAGCTCGCTCGTTAGCTTTGAGATGTCTCAAAAGCGCAGACCGG	TCCGCCTC	TTTGCTG	MHMVKKHPVEFNEL
	AGAAAGCAGTCCCCGACGAGGGAGCGACTGCTCATGATGTTGCGCAAC	CTCGTGGC	CCGAAG	KMAKKPVPKKVRWS
	CGGATAAGAGTAAGTGTTCTGTTTGTGGGGAGACTTTTAAGGGTCCTGC	GCCAGACT	AACTACT	EEEIFQLARTEAELTL
	TTCCGTGACAATGCACATGGTAAAGAAGCACCCAGTCGAATTTAATGAG	GGGTAAG	GAAGAC	QGVRFINVELQKIFPA
	CTTAAGATGGCCAAGAAGCCTGTTCCGAAGAAAGTACGATGGTCTGAG	CTGATTTA	TATTGA	REIEGIKGQRKLAKYK
	GAGGAGATCTTCCAACTTGCCAGGACAGAGGCCGAACTTACCCTTCAGG	TCAGTGAG	ACTAAC	ELVKDQLDEIGRAPN
	GAGTTCGGTTTATTAATGTCGAGCTGCAGAAGATCTTCCCAGCACGCGA	CTAAGAGA	CAACTG	PPEQEIGEDVPSPFKA
	GATCGAGGGCATCAAAGGCCAGAGGAAACTGGCTAAGTACAAAGAGCT	AACGATCA	TGTTAA	WLELLLALPKTPNDFL
	CGTTAAAGATCAGCTGGACGAAATCGGCCGTGCTCCTAACCCCCCTGAG	CCGCAGGA	GTAAGA	EHKLDNIIVQALKEDV
	CAAGAGATCGGAGAGGATGTGCCTAGCCCGTTCAAGGCATGGCTGGAA	GTCCATCT	ACTAAT	NSDQVFNDLNSYLKL
	CTTCTGCTTGCGCTCCCCAAGACCCCCAATGATTTTCTAGAACACAAACT	ACCTGCGC	GCCTCTT	ILEPSGRAKSVPGEIIH
	TGACAACATTATTGTTCAAGCGCTGAAGGAAGACGTGAACTCCGATCAG	TGCACGTG	TTTCCCT	GDPSGSAKTSVTKAP
	GTATTCAATGACCTTAACTCCTATTTAAAGCTGATTCTTGAGCCGAGTGG	ATTTCATT	GCATGT	KPATVSSSRKKRRDA
	TAGAGCCAAGTCAGTTCCCGGGGAAATTATACACGGTGATCCCAGTGG	GTGCACTT	ATCCCCT	EFARIQRLYRKNRTSC
	GTCTGCTAAGACCTCAGTAACCAAGGCTCCTAAGCCTGCTACAGTTTCTT	GGCGAGTT	GATCAG	INTILDGNTREHEAPK
	CTAGCAGGAAAAAACGCAGAGATGCCGAGTTTGCTAGGATCCAGCGTC	ATCCCTTG	TGACTTA	NMEGFWREIFERESP
	TCTATCGGAAAAATCGGACATCATGTATAAATACAATCCTTGATGGCAAT	TGGAGCTC	TTTTCTT	DDPDDPDIFLEEEAS
	ACAAGAGAGCATGAGGCCCCTAAGAACATGGAAGGATTTTGGAGGGAA	GCTCGTTA	TTCCTGT	DIWKYISFYEMCNLY
	ATATTCGAGAGGGAATCCCCGGATGACCCCGATGATCCCGACATTTTCC	GCTTTTGA	TGCGCC	PPPSTAPGPDGFSSK
	TTGAGGAAGAAGCCAGTGATATTTGGAAGTACATCTCTTTCTATGAGAT	G (SEQ ID	CTTTGTT	DLRRMTPRVLNKILN
	GTGCAACCTTTATCCGCCTCCCTCCACTGCTCCTGGCCCCGATGGGTTTT	NO: 1237)	TAGTTAT	LLLHLRDLPQILKSHR
	CCTCTAAGGACCTCCGGCGGATGACGCCAAGAGTCCTCAACAAAATTTT		TTCCTTT	TVLIPKTDLPTKPGDF
	AAACCTACTCCTTCATTTGAGAGATCTTCCTCAGATCCTGAAATCCCATC		AATTACT	RPITISNILVRHLNKIL
	GAACTGTACTAATACCAAAAACCGATTTGCCTACTAAACCCGGGGACTTT		AGAATT	ANRVSHLIPINERQK
	CGTCCCATAACGATCTCAAATATACTAGTCAGACATCTTAATAAGATTCT		ATCTTTT	AFLPIDGCAENIFTLD
	GGCTAATAGAGTGAGCCACCTTATACCAATCAATGAGAGGCAGAAAGC		CGTTCTC	FILHHARTKIKSLSMA
	TTTCCTACCAATTGATGGCTGCGCAGAAAATATTTTCACTTTGGACTTCA		CGTCTA	ILDISKAFDSVSHHSIF
	TACTCCATCATGCGCGGACCAAAATTAAATCGCTGTCAATGGCCATTCTA		ATTGCTT	RALREARCPIGFIKFIE
	GATATTAGTAAGGCATTTGATAGCGTCTCGCATCATTCCATCTTCCGCGC		TTTCGGT	NCYGGCFTKLFCGGV
	TTTGCGAGAGGCTCGATGCCCCATTGGTTTTATTAAATTCATTGAGAATT		GTAGGA	KYPSEVSMNRGVKQ
	GTTATGGTGGATGTTTTACCAAACTTTTCTGTGGTGGCGTGAAATACCCG		ACGGCT	GDPLSPVLFNLVIDGL
	AGTGAGGTTTCTATGAACCGAGGAGTAAAGCAGGGGGATCCGCTGTCA		ACCTAA	IRQIPSALGFNVSDQ
	CCAGTCCTATTTAACTTAGTCATTGATGGTCTGATCAGGCAAATTCCATC		AGCTGG	VKVSCIAYADDLILIAT
	CGCTCTTGGATTCAACGTAAGTGATCAGGTGAAAGTCAGCTGTATCGCC		AAGTGG	TRAGLKTLLDLTNSYL
	TACGCTGATGACCTGATCCTTATTGCGACCACACAAGGGCTGGCCTGAAGA		GAAGTG	AKRGLSLNPDKCSAL
	CCCTGCTTGACTTGACTAACTCCTACCTAGCAAAACGGGGACTTTCGCTC		TTTTCAA	SIVASGKQKLVYIASS
	AACCCGGACAAATGCTCGGCCCTATCGATTGTGGCATCCGGCAAGCAGA		TGTACTT	EHFDLAGQKMRNLN
	AGCTTGTATATATTGCGTCTTCCGAGCATTTTGATCTCGCAGGTCAGAAA		TGTGATT	VGDSWRYLGIQFSHL

			ATGCGCAATTTAAACGTCGGTGACTCTTGGAGATACCTGGGTATTCAAT		ATAGAA		GRAEKVTPDLTCLINR
			TCTCGCATCTCGGTAGAGCCGAAAAGGTCACTCCAGACCTAACATGTCT		ATATAT		LQKAPLKLQQKLYAL
			AATTAATAGACTTCAAAAGGCCCCTTTAAAGCTTCAACAGAAGCTTTATG		GACCCG		RIYLIPRLIHGLTLSKT
			CTCTCCGAATCTATCTTATACCCAGGCTTATCCATGGCCTCACCCTCTCTA		AGGTGC		NLGELKTLDKLIRKYIR
			AGACCAACCTCGGAGAGCTTAAAACGCTGGACAAGCTAATCCGCAAGT		ATTGTTT		AWLHLPDDTPMGYF
			ACATCCGCGCCTGGCTCCATCTCCCAGATGACACTCCGATGGGCTATTTC		GGCATT		YTPLKAGGLGLPSLRL
			TATACCCCCTTAAAAGCTGGTGGACTTGGCCTTCCTTCACTCCGTCTAGT		TCCTCGA		VILNNRLERILRMKAS
			GATCCTTAACAATCGACTTGAACGGATACTAAGGATGAAGGCTTCTCAA		GAAA		QDIIVRTIAESETLGVE
			GACATTATTGTCCGCACTATTGCCGAATCGGAGACGTTGGGGGTCGAGA		(SEQ ID		IRKLHDLLSIDGTILDT
			TTCGAAAGCTTCACGATCTGTTGTCCATTGATGGAACAATTCTGGACACG		NO:		SVKIHSFWAERLYSSY
			AGCGTTAAGATCCATAGCTTCTGGGCAGAAAGACTCTACTCGTCCTATG		1360)		DGKCLCNSANFPPG
			ATGGTAAATGTCTTTGCAATTCTGCAAATTTTCCCCCCGGAAACAAGTGG				NKWIGEDSLNQRSHI
			ATCGGTGAAGATTCTCTGAACCAACGCTCGCATATCTTCGCTGATTGCCT				FADCLKLRINALPTRS
			AAAGCTCCGGATAAATGCGCTCCCAACACGTTCCAGAACAGCACGTGGA				RTARGRPLKDKPCRA
			AGGCCACTAAAAGATAAACCATGTAGAGCTGGATGTCGGAACAGTGAT				GCRNSDGVKVIETLN
			GGTGTAAAGGTCATTGAGACACTGAATCACATTACACAGGTGTGTGAG				HITQVCERTHGARVK
			AGAACCCACGGAGCCCGCGTCAAGCGGCATGATCGCCTAGTGGACTTC				RHDRLVDFAVKGLQ
			GCGGTGAAAGGACTACAAAGACCCCATCGTGTGGTTCTAAAGGAGCCA				RPHRVVLKEPHYKTV
			CATTACAAGACAGTAAACGGTGTTCGGAAACCGGACATCGTCATTAAAA				NGVRKPDIVIKIPDHT
			TCCCCGACCACACATACATTTGTGATTTCCAAGTTGTATCTGATACCTCTT				YICDFQVVSDTSCLEL
			GTTTAGAGTTGGAGTTCCGGAAGAAAGCCCTCAAGTATGCGGAGGACA				EFRKKALKYAEDKGL
			AAGGCCTCTCGCGATCAACTTACCCGCGACCATCCTGGGGAGCTCTCATT				CDQLTRDHPGELSFT
			CACCGCAATTACTTTCAATACAAGAGGGCTTATAGCCAAGTCTTCAGTCA				AITFNTRGLIAKSSVT
			CCGCCCTCCGGAAATTGGGGATGCCACCCAGATCTATCATGACCCTCCA				ALRKLGMPPRSIMTL
			GAAAATATGTATGGAGGGTTCTCTGGAGATCTGGAGGATTTTTTAACCAG				QKICMEGSLEIWRIF
			ACTACTGCTATGGCCCGCAATTAGATGACTGCCCTACCCCTTTGCTGCCG				NQTTAMARN (SEQ
			AAGAACTACTGAAGACTATTGAACTAACCAACTGTGTTAAGTAAGAACT				ID NO: 1482)
			AATGCCTCTTTTTCCCTGCATGTATCCCCTGATCAGTGACTTATTTTCTTTT				
			CCTGTTGCGCCCTTTGTTTAGTTATTTCCTTTAATTACTAGAATTATCTTTT				
			CGTTCTCCGTCTAATTGCTTTTTCGGTGTAGGAACGGCTACCTAAAGCTG				
			GAAGTGGGAAGTGTTTTCAATGTACTTTGTGATTATAGAAATATATGAC				
			CCGAGGTGCATTGTTTGGCATTTCCTCGAGAAA (SEQ ID NO: 1114)		AGAATATG	TAGCCA	FERRKDPWGYRPPG
R2	R2_DAn	.	AGAATATGGATTTGATTGTGCAGAGGGGGTGCTATACCGTAACTCGTAA	Drosophila ananassae	GATTTGAT	ATGCAC	TLKQIGATENNEPRN
			GCCATGCAATCAGATCAAGTCGACTCAAAACCTCCTCGTGGTATTCTCTG		TGTGCAGA	GGGTTC	LNRFVRGESTASSLES
			GGTGCCAGTATTTACTGGTAGCTGATTTGAGCGGCGAAAGGATCCTTGG		GGGGGTG	CAGATT	TQFGTSAEVNLAGRV
			GGTTACCGGCCCCCTGGAACCTTAAAACAAATTGGTGCAACTGAAAATA		CTATACCG	AAGCTT	PCTICEMTFSSKRGLG
			ACGAGCCTCGGAACCTAAATCGTTTTGTAAGAGGAGAATCCACGGCTTC		TAACTCGT	GCTGCC	VHMSHRHKDDLDA
			CAGCCTGGAGAGCACACAATTTGGAACCAGTGCAGAGGTTAACCTTGCA				

	GGGAGGGTGCCCTGTACGATATGTGAAATGACGTTCAGCTCCAAGAGG	AAGCCATG	GAAGCA		QRLRVDKKARWSEE
	GGTTTGGGCGTTCACATGTCACATCGGCACAAAGACGATCTTGATGCAC	CAATCAGA	TACCATC		ETLMMARKEVELAA
	AACGTCTTCGTGTCGATAAAAAGGCAAGGTGGTCAGAGGAAGAAACCT	TCAAGTCG	AAAATC		SGVRFLNKKLAEIFTH
	TGATGATGGCGAGAAAGGAGGTTGAGCTTGCAGCAAGTGGTGTACGAT	ACTCAAAA	GGCATA		RSADAISSYRKRSEYK
	TTCTTAATAAGAAGCTAGCGGAGATTTTCACCCACCGCAGTGCCGATGC	CCTCCTCG	AAATTC		AKLEQIRGQSVPTPE
	GATATCTTCGTATCGGAAGAGGAGTGAGTACAAGGCAAAACTAGAGCA	TGGTATTC	GCTTAAT		AEEINTTQRRPSNSE
	GATAAGGGGGCAATCCGTTCCCACCCCAGAAGCAGAAGAAATCAACAC	TCTGGGTG	AAAGGA		QNRRVPRSEGGPIAP
	CACACAGCGCCGCCCTAGTAATAGCGAGCAAAACCGACGAGTACCAAG	CCAGTATT	GGTGGT		TEQTNNEILRVLQGL
	ATCAGAAGGGGGACCAATCGCACCAACCGAACAGACGAACAACGAAAT	TACTGGTA	TTTAGTA		APVVCLPRWRAEVL
	CCTTAGGGTACTACAGGGTCTAGCACCTGTAGTATGCTTACCCCGGTGG	GCTGA	CGTAGG		QNIVDNAQVSGQET
	AGAGCCGAGGTCCTGCAAAATATCGTAGATAATGCGCAGGTCTCGGGA	(SEQ ID	CGTCCC		TLQSLSSYLMEIFPPR
	CAGGAAACCACTCTCCAAAGTTTATCCAGTTATCTCATGGAAATTTTTCC	NO: 1238)	GGGACT		NEPHILTRPRTEPRN
	GCCACGGAATGAACCGCACATTCTGACGAGGCCCCGAACGGAGCCTCG		TGTCTC		MRQRRRQQYARVQ
	AAATATGAGACAACGTAGAAGGCAGCAGTACGCGAGGGTTCAGCGTAA		GGGATG		RNWDKHPGRCIKSLL
	CTGGGATAAACATCCGGGGCGATGCATAAAGTCCCTACTGGAGGAAGA		AATCGT		EEDDESVMPNQEV
	TGATGAGTCGGTGATGCCAAACCAGGAGGTCATGGAGCCATATTGGAG		GCATGC		MEPYWRRVMTQPS
	ACGGGTAATGACTCAGCCTAGCTCAAGCTCGATAAAACGCGACATGTTT		GTATAA		SSSIKRDMFNMEHSL
	AACATGGAGCATTCACTCGAGAGGGTATGGTCCGCTGTGAACCAGCGC		TTGGGA		ERVWSAVNQRDLRA
	GATCTTAGGGCCACAAAAGTCAAATTATCTAGTTCTCCAGGCCCGGACG		TCGATA		TKVKLSSSPGPDGITP
	GGATCACTCCAAAAACTGCCAGGAGTGTCCCCGAAGGCATTATGCTTCG		ACAAAT		KTARSVPEGIMLRIM
	CATAATGAACTTGATCCTCTGGTGCGGGAATTTGCCGTACTCTATCCGTC		ACCAAC		NLILWCGNLPYSIRLA
	TGGCCCGAACCATCTTCATTCCGAAGAAGGCGACGGCAAATCAACCGCA		TAAGTT		RTIFIPKKATANQPQ
	AGACTATCGTCCTATTTCAGTCCCCTCGGTTATAGTTAGGCAACTAAATG		ATTACTA		DYRPISVPSVIVRQLN
	CCATTTTGGCTTCCCGGTTGAGCGCAGCCATCAACTGGGACACGCGTCA		ATATATC		AILASRLSAAINWDT
	GCGAGGGTTCCTACCTACCGATGGGTGTGCTGATAATACGACGATTGTT		GAAATA		RQRGFLPTDGCADN
	GATTTAGTTTTGAGGGAACATCATAAGCGATTTAAATCGTGCTACATCG		CATAAA		TTIVDLVLREHHKRFK
	GGACCCTCGATGTTAGTAAGGCCTTTGATGCTGTAGCTCACGAAGCGGT		TATCCCG		SCYIGTLDVSKAFDAV
	CTACAACACATTGGCTTCATATGGTGCCCCGAAAGGCTTCATCAACTACT		TCCTTAC		AHEAVYNTLASYGAP
	TACGGAAGGCGTACGAGGGCGGCGGCACACAATGCTCGCTGGGAACGGG		GTATCTT		KGFINYLRKAYEGGG
	TGGGTTTCAGAGGCGTTCATTCCTGCCCGAGGAGTGAAGCAGGGTGAC		TGAAGA		TMLAGNGWVSEAFI
	CCTCTGTCTCCCATACTATTCAACTTGGTCATTGACCGGTTGCTTAGGTCC		TTTCCAT		PARGVKQGDPLSPIL
	TTACCCAGTGAGATTGGTGCCAAAGTCGGAAATGCCATGACAAACGCA		CCTCAG		FNLVIDRLLRSLPSEIG
	GCAGCATTCGCAGATGATATAGTCCTTTTTGCGGAAACTCCGATGGGGC		CGAACA		AKVGNAMTNAAAFA
	TTCAGAAATTGTTGGACACCACCGTTTGTTTCCTTTCCTCGGTGGGTCTC		AAAAAA		DDIVLFAETPMGLQK
	ACCCTTAATACTGATAAATGTTTCACGGTCAGTATTAAGGGGCAAGCCA		AAAAAA		LLDTTVCFLSSVGLTL
	AACAAAAGTGTACCGTCGTCGAACGGCGAAGCTTCTTGATTGGCGGGC		A (SEQ		NTDKCFTVSIKGQAK
	GCGAGTGTCCTTCATTGAAGCGTACTGACGAGTGGAAGTACTTAGGGAT		ID NO:		QKCTVVERRSFLIGG
	TAAATTCACTGCGGAGGGGGCGGGCCCGGTACGATCCAGCAGAGGACCT		1361)		RECPSLKRTDEWKYL

			CGGTCCAAAGCTGTTGAGATTGACTCGGGCCCCCCTGAAACCACAACAG				GIKFTAEGRARYDPA
			AAGTTATTTGCCCTTCGGACCGTCCTTATCCCACAACTCTATCACAAGCT				EDLGPKLLRLTRAPLK
			GACCCTTGGGAGTGTGACGATAGGCGTTCTGAAGAAATTTGACAAATTA				PQQKLFALRTVLIPQL
			GTTCGATATACCGCACGGAAGTGGTTGGGGCTTCCGGTGGACGTACCA				YHKLTLGSVTIGVLKK
			GTTTCTTTTTCCATGCCCCCCACAAGAGTGGGGGTCTCGGGTTACCATC				FDKLVRYTARKWLGL
			TCTAAGATGGACAGCACCAATGCTTCGACTAAAGCGATTGAGCAACATA				PVDVPVSFFHAPHKS
			AAATGGCCTCACCTCGAGCGATCCGAGGTAGCCAGCTCTTTCGTGGAGG				GGLGLPSLRWTAPM
			AAGAAATGCGGAGGGCCCGGGATAGGCTTCAGGCTGGAAGTGAAGAA				LRLKRLSNIKWPHLER
			CTGTTAACCCGTTCGCAGGTAGATTCGTACTTGGCAAATAGATTGCACAT				SEVASSFVEEEMRRA
			GTCTGTTGATGGTTGCGGGCTCCGTGAAGCAGAGCGTTTTGCTCCGCAA				RDRLQAGSEELLTRS
			CACGGGTGGGTTAGTCAGCCCACGCGTTTGCTAACAGGAAAGGAATAT				QVDSYLANRLHMSV
			ACTGATGGAATCAAACTGCGGATAAATGCCCTACCCTCAAGGTCTCGTA				DGCGLREAERFAPQ
			CTACGAGGGGAAGGCACGAATTGGAGAGACGGTGCCGTGCAGGATGT				HGWVSQPTRLLTGK
			GATGCTCCCGAAACAACAAATCACATCTTGCAGCAATGCTATAGAACTC				EYTDGIKLRINALPSR
			ACGGGAGGAGGATAGCTCGGCACAACGGCGTAGTAAATTTTCTCAAGC				SRTTRGRHELERRCR
			GGGGACTTGAGCGAAGAGGCTGCGTCGTTCATGTTGAACCAAGTCTGC				AGCDAPETTNHILQQ
			AGGGCGAAACCGGACTGAATAAACCTGACCTGGTGGCTATCCGACAAA				CYRTHGRRIARHNGV
			ATCGCATTTATGTGATTGACACTCAGATTGTGACTGACGGACATTCTCTC				VNFLKRGLERRGCVV
			GACCAAGCGCACCAGCGTAAGGTCGGGAAGTACGATACACCGGACATA				HVEPSLQGETGLNKP
			CGGACGAATTTGCGGAGATCTTTCGGTGCCTTTGACATTGAGTTCCATTC				DLVAIRQNRIYVIDTQ
			CGCCACTGTGAACTGGAGGGGAATATGGAGTGGTCAATCAGTAAAACG				IVTDGHSLDQAHQRK
			GTTGATCGCTTCAGACCTCCTCAGCTCTGGTGATAGCAATATCATCAGTG				VGKYDTPDIRTNLRR
			TCCGGGTAATCAGTGGTGGTCTCTGGAGCTGGCGGCAGTTCATGTATCT				SFGAFDIEFHSATVN
			GTCGGGGTACACTCGCGATTGGACTTAGCCAATGCACGGGTTCCAGATT				WRGIWSGQSVKRLI
			AAGCTTGCTGCCGAAGCATACCATCAAAATCGGCATAAAAATTCGCTTAA				ASDLLSSGDSNIISVR
			TAAAGGAGGTGGTTTTAGTACGTAGGCGTCCCGGGACTTGTCTCGGGAT				VISGGLWSWRQFMY
			GAATCGTGCATGCGTATAATTGGGATCGATAACAAATACCAACTAAGTT				LSGYTRDWT (SEQ ID
			ATTACTAATATATCGAAATACATAAATATCCCGTCCTTACGTATCTTTGAA				NO: 1483)
			GATTTCCATCCTCAGCGAACAAAAAAAAAAAAAAA (SEQ ID NO: 1115)				
R2	R2_DM	X51967	Drosophila melanogaster	TTGGGGATCATGGGTATTTGAGAGCAGAGGGGGAGTATTCTTCTGTA	TTGGGGAT	TAGCTA	MTTRPSVDIFPEDQY
				ATTCGTAAGTCATATCATATGATGTGCGGAAGGGGGAATTTTACTCTGTA	CATGGGGT	AATCGTT	EPNAAATLSRVPCTV
				ACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTG	ATTTGAGA	TGGTTC	CGRSFNSKRGLGVH
				GTCCCGGTAATGCTAAACTCGTTTAGCAGCTAATTTGAGCGGAAAAACT	GCAGAGG	AAAACA	MRSRHPDELDEERRR
				TTTCCGATGGGCTGGTTCCCCAGAGGAAATTTATTCATATTGGAACTACA	GGGAGTAT	TTTGCTT	VDIKARWSDEEKWM
				AGCACAAATAACGAGCCTCGGATACCTTTACACAATCTGATGACGACCC	TCTTCTGT	GCTGTC	MARKEVELTANGCK
				GACCCTCCGTGGATATCTTCCCGGAGGACCAATATGAACCAAACGCAGC	AATTCGTA	TTGGCA	HINKQLAVYFANRSV
				GGCTACTCTATCTAGGGTTCCCTGCACAGTATGTGGCCGGTCCTTTAACA	AGTCATAT	TAACATC	EAIKKLRQRGDYKEKI
				GCAAGAGAGGACTCGGTGTTCACATGCGATCTCGGCACCCAGACGAAC	CATATGAT	AATAAA	EQIRGQSALAPEVAN

	TTGATGAAGAACGTCGACGTGTCGATATAAAGGCAAGATGGAGTGATG	GTGCGGA	GGCATA		LTIRRRPSRSEQDHQ
	AAGAGAAGTGGATGATGGCGAGAAAGGAGGTTGAGCTCACAGCAAAT	AGGGGAA	AACATC		VTTSETTPITPFEQSN
	GGATGTAAACACATAAACAAGCAACTAGCGGTGTATTTTGCAAACCGCA	TTTTACTCT	GCAAAA		REILRTLRGYSPVECH
	GCGTCGAAGCCATCAAAAAGCTAAGACAGAGGGGCGATTATAAGGAGA	GTAACTCA	TAATGG		SKWRAQELQTIIDRA
	AAATAGAGCAGATAAGAGGGCAATCAGCTCTCGCCCCGGAAGTTGCTA	CAAGTCTC	TTATAAT		HLEGKETTLQCLSLYL
	ATCTAACCATAAGGCGCCGCCCTAGTAGAAGTGAGCAAGACCACCAAGT	TCCTTTACT	TAAATG		LGIFPAQGVRHTLTR
	AACAACATCGGAAACAACTCCAATCACTCCCTTCGAACAGTCGAACAGG	CAAGTCGA	GCTATG		PPRRPRNRRESRRQQ
	GAAATTTTGCGGACACTACGCGGGTATAGCCCCGTAGAATGCCATTCCA	CTCAAAAC	AGGATG		YAVVQRNWDKHKG
	AATGGAGAGCCCAAGAGTTGCAAACTATCATTGATAGGGCACATCTCGA	CTCCTCGT	GTTTTAG		RCIKSLLNGTDESVM
	GGGAAAGGAAACCACTCTCCAATGCTTATCGCTATATCTCCTGGGAATTT	GGTGGTCC	TACGTA		PSQEIMVPYWREVM
	TTCCGGCACAGGGTGTACGACACACACTGACGAGACCTCCTCGGAGACC	CGGTAATG	GGCGTT		TQPSPSSCSGEVIQM
	TCGGAACAGGAGAGAAAGCAGAAGGCAGCAGTACGCTGTCGTCCAGC	CTAAACTC	GCGGAA		DHSLERVWSAITEQD
	GTAACTGGGATAAGCATAAAGGAAGATGCATCAAGTCCTTGCTAAATG	GTTTAGCA	CTTCGG		LRASRVSLSSSPGPDG
	GAACTGATGAGTCGGTAATGCCAAGCCAAGAAATAATGGTTCCCTACTG	GCTAATTT	TTCATAT		ITPKSAREVPSGIMLR
	GAGAGAAGTAATGACTCAGCCTAGCCCAAGCTCTTGCAGTGGAGAAGT	GAGCGGA	AGAGCA		IMNLILWCGNLPHSI
	GATACAAATGGATCACTCGCTTGAGAGGGTATGGTCTGCTATTACAGAG	AAAACTTT	ATGAAT		RLARTVFIPKTVTAKR
	CAGGACCTTCGGGCGTCAAGAGTCTCATTATCCTCGTCTCCGGGGCCTG	TCCGATGG	CGTGCA		PQDFRPISVPSVLVR
	ACGGGATAACTCCAAAATCTGCCAGGGAGGTGCCGTCAGGTATTATGCT	GCTGGTTC	TGCTAG		QLNAILATRLNSSIN
	GCGCATAATGAACCTAATTCTATGGTGCGGTAATCTACCACACTCCATAC	CCCAGAGG	GAAAAC		WDPRQRGFLPTDGC
	GACTGGCCAGAACCGTCTTCATCCCGAAGACGGTGACGGCGAAGCGAC	AAATTTAT	TGACCA		ADNATIVDLVLRHSH
	CGCAAGACTTTCGTCCAATATCAGTGCCTTCAGTCCTGGTAAGACAGCTA	TCATATTG	CACACA		KHFRSCYIANLDVSKA
	AATGCAATATTGGCAACCCGGTTGAACTCATCAATCAATTGGGACCCGC	GAACTACA	GTGTTG		FDSLSHASIYDTLRAY
	GCCAGCGGGGCTTCTTACCAACCGACGGATGCGCCGATAATGCGACGA	AGCACAAA	GCAGAC		GAPKGFVDYVQNTY
	TAGTCGACTTAGTCTTGAGGCATAGCCATAAGCACTTTAGATCTTGCTAC	TAACGAGC	CTAGTAT		EGGGTSLNGDGWSS
	ATCGCAAATTTAGATGTAAGCAAGGCATTTGATTCTCTATCACATGCATC	CTCGGATA	CTTTCGA		EEFVPARGVKQGDPL
	TATATACGACACCTTACGTGCTTATGGTGCGCCAAAGGGCTTCGTTGACT	CCTTTACA	AGATTTC		SPILFNLVMDRLLRTL
	ACGTACAGAACACGTACGAGGGCGGTGGTACCAGTCTCAATGGGGACG	CAATCTG	CATACCT		PSEIGAKVGNAITNA
	GTTGGAGTTCAGAGGAATTCGTCCCTGCTAGAGGAGTGAAGCAGGGTG	(SEQ ID	CCGCGA		AAFADDLVLFAETRM
	ACCCTTTGTCTCCTATTCTATTTAACTTGGTAATGGACAGGTTACTTAGAA	NO: 1239)	TCAAAA		GLQVLLDKTLDFLSIV
	CCTTACCCAGCGAAATTGGTGCCAAAGTCGGAAATGCCATTACTAACGC		AAAAAA		GLKLNADKCFTVGIK
	GGCCGCGTTTGCAGATGATTTGGTACTATTTGCGGAAACTCGGATGGGG		AAAAAA		GQPKQKCTVLEAQSF
	CTTCAAGTATTGTTGGACAAGACGTTGGATTTTCTATCTATCGTCGGCCT		AAAAAA		YVGSSEIPSLKRTDEW
	CAAACTTAATGCCGACAAATGTTTTACCGTTGGCATTAAGGGCCAGCCG		(SEQ ID		KYLGINFTATGRVRC
	AAACAGAAGTGTACCGTGTTAGAGGCACAGAGCTTCTACGTAGGCTCG		NO:		NPAEDIGPKLQRLTK
	AGTGAGATTCCATCACTGAAGCGCACGGACGAGTGGAAGTACTTAGGC		1362)		APLKPQQRLFALRTV
	ATCAACTTCACTGCAACCGGGAGGGTTCGATGCAATCCGGCCGAGGAC				LIPQLYHKLALGSVAI
	ATTGGTCCAAAGCTACAAAGATTGACAAAGGCCCCCCTCAAACCACAAC				GVLRKTDKLIRYYVRR
	AGAGGTTGTTCGCCCTTCGGACTGTCCTTATCCCACAGCTCTACCACAAG				WLNLPLDVPIAFVHA

			TTAGCCCTTGGGAGTGTGGCGATAGGCGTCCTAAGAAAAACTGATAAA CTTATAAGATATTATGTGCGAAGATGGCTAAATCTTCCGCTGGATGTGC CGATAGCATTTGTTCATGCACCCCCAAAAAGTGGAGGTCTCGGAATTCC ATCACTAAGATGGGTAGCTCCAATGTTAAGGCTAAGACGCTTGAGTAAC ATTAAATGGCCTCACCTCACGCAAAACGAGGTAGCCAGCTCTTTCCTCG AAGCCGAAAAACAACGGGCCCGAGATAGATTATTAGCTGAACAAAATG AACTGTTATCGCGTCCGGCAATAGAAAAATATTGGGCGAACAAGTTGTA CCTCTCAGTTGATGGTAGCGGACTCCGTGAAGGCGGCCATTATGGCCCG CAACACGGGTGGGTTAGTCAACCCACGCGTTTATTAACAGGAAAGGAAT ATATGGACGGTATTCGTCTGCGGATAAATGCCCTACCCACAAAGTCTCG TACTACAAGGGGAAGGCACGAATTGGAACGACAGTGTCGTGCAGGATG TGATGCTCCCGAAACAACAAACCACATAATGCAAAAATGCTACCGCTCG CATGGGAGGCGGGTAGCTAGACACAACTGCGTAGTAAATCGAATCAAG CGGGGACTTGAGGAGAGAGGCTGCGTGGTCATTGTTGAACCAAGTCTG CAGTGCGAATCCGGCCTTAATAAACCAGACCTGGTGGCACTACGACAAA ATCACATTGATGTGATCGACACACAAATTGTGACAGACGGACACTCTAT GGATGATGCGCACCAGCGCAAAATCAATAGATACGACAGACCGGACAT ACGAACTGAATTGCGTCGCAGATTCGAAGCCGCAGGTGACATTGAATTC CATTCTGCCACCCTGAACTGGAGGGGGATCTGGAGTGGTCAATCCGTTA AAAGATTGATAGCAAAGGGTCTCCTCAGCAAATATGATAGTCATATCAT TAGCGTCCAGGTTATGAGAGGCAGTCTCGGTTGTTTTAAACAGTTCATG TACCTGAGCGGGTTTTCCCGAGATTGGACTTAGCTAAATCGTTTGGTTCA AAACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACATCG CAAAATAATGGTTATAATTAAATTAAATGGCTATGAGGATGGTTTTAGTACGTA GGCGTTGCGGAACTTCGGTTCATATAGAGCAATGAATCGTGCATGCTAG GAAAACTGACCACACACAGTGTTGGCAGACCTAGTATCTTTCGAAGATT TCCATACCTCCGCGATCAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1116)			PPKSGGLGIPSLRWV APMLRLRRLSNIKWP HLTQNEVASSFLEAE KQRARDRLLAEQNEL LSRPAIEKYWANKLYL SVDGSGLREGGHYG PQHGWVSQPTRLLT GKEYMDGIRLRINAL PTKSRTTRGRHELER QCRAGCDAPETTNHI MQKCYRSHGRRVAR HNCVVNRIKRGLEER GCVVIVEPSLQCESGL NKPDLVALRQNHIDV IDTQIVTDGHSMDD AHQRKINRYDRPDIR TELRRRFEAAGDIEFH SATLNWRGIWSGQS VKRLIAKGLLSKYDSHI ISVQVMRGSLGCFKQ FMYLSGFSRDWT (SEQ ID NO: 1484)
R2	R2_DPe	Drosophila persimilis	AAGATATGGATCTGAATAATAGCGTAGAAGGGGAGTCATTCCGTAATTC GTAAATCGTAAAAATCAGATCAAGTTGATTCAAGACCTCCTCGTGGTAT CTTCTGGATGCTATTAGACTGAAGTTCTTTTGGTCTAATAGTAACTAACT TGAACAGCGAAACAGTCCTATGGGGTTGCCAGCCCCTTGGACAGTTCAG TTTGATTGGCACTAATATGCAAAATACAACGCCTCGGATAATAAACACTA ATTCGTTGACGAACCAAATCCCTACGGTCTCTAGCCTAGGGGCCCAATCT GAACATAGTGCACAGGTTAACCCAAACAGTGGTTACCAATGCACGATAT GTGAATCGTCTTTCCGTAGCAAAAGCGGACTAGGCGTTCACATGTCACG TCGGCACAAGGACGAGTTTGATCAACTTCGTCTGCGTACCGACCGTAAG GCACAATGGAGTGAGGAAGAGTTGAGTATGATGGCAAGAAAAGAGAT	AAGATATG GATCTGAA TAATAGCG TAGAAGG GGAGTCAT TCCGTAAT TCGTAAAT CGTAAAAA TCAGATCA AGTTGATT	TAGCCT ATACACT ATGTTG GAGAGA AGACGC TTGCTAC CTAGGC AAAATG TGAAAT TAGGTA	SSFGLIVTNLNSETVL WGCQPLGQFSLIGT NMQNTTPRIINTNSL TNQIPTVSSLGAQSE HSAQVNPNSGYQCTI CESSFRSKSGLGVHM SRRHKDEFDQLRLRT DRKAQWSEEELSM MARKEIELAANGERY LNKKLAEVFTNRSVD

	CGAGCTCGCAGCAAATGGAGAAAGATATCTAAATAAGAAGCTAGCGGA	CAAGACCT	TAAACA	AIKKCRQRERYKTKIE
	AGTATTTACGAACCGTAGCGTCGACGCTATCAAGAAATGTCGACAGAGG	CCTCGTGG	TCGTGG	QLKGQAVPLPEALES
	GAGAGATATAAGACCAAAATCGAACAGCTAAAGGGTCAAGCTGTTCCTC	TATCTTCT	TTGTAA	ETIQRRPSIRERDLLV
	TCCCAGAAGCATTAGAATCTGAAACCATACAGCGCCGCCCTAGTATACG	GGATGCTA	AACTTG	TPPNTLGTTPTELSNS
	CGAGCGAGATCTCCTAGTAACGCCACCTAACACTCTCGGAACCACTCCA	TTAGACTG	AGGTGG	EILAVLQGYPPVVCN
	ACCGAACTGTCGAACAGTGAAATCCTGGCAGTACTACAGGGGTACCCAC	A (SEQ ID	GTTTTTA	DQWRVEVLQSIVDG
	CTGTAGTATGCAATGACCAATGGAGAGTTGAGGTTTTGCAATCCATCGT	NO: 1240)	GTACGT	AQASGKEITLQRLSTY
	AGATGGTGCGCAGGCCTCGGGTAAGGAAATTACTCTTCAGCGCTTGTCT		ATGCGT	LMEVFPSQNDRPIQT
	ACTTACCTTATGGAAGTATTTCCCTCACAGAATGACCGCCCCCATTCAAAC		GATTACT	RPPRRPRNRRQGRR
	GAGACCTCCACGGAGACCTCGTAATAGGAGACAAGGTAGGAGACAGCA		TCGTAAT	QQYALTQRNWDKH
	GTACGCCTTAACACAGCGTAACTGGGACAAGCACACAAAGGTCGTTGTATA		CATGAA	KGRCIKAILDGTEGTA
	AAAGCCATTTTGGATGGAACTGAGGGGACAGCAACTATGCCAAGTCAA		TCGTGC	TMPSQGIMGSYWR
	GGTATCATGGGGTCCTATTGGAGACAAGTCATGACACAAACAAGCCCAA		ATGCTA	QVMTQTSPTYSGTN
	CATATAGTGGTACGAACACCACGTTCCGGACGGAACACCCACTTGAAGG		GTGGGG	TTFRTEHPLEGVWSP
	GGTTTGGTCCCCGATAACACTAGGGGACCTAAGGGTACACAGAGTGTC		TTTGGC	ITLGDLRVHRVSLTKS
	ATTGACGAAATCTCCAGGACCTGATGGAATTACTCCAAGAACTGTCAGG		CTCCACT	PGPDGITPRTVRSIPS
	AGTATTCCGTCAGGAGTTATGCTTCGCATAATGAACCTGATACTTTGGTG		AGTATCT	GVMLRIMNLILWCG
	CGGAAAGTTGCCTGTCTCCATCCGACAGGCACGAACCATCTTCATTCCG		TTGAAG	KLPVSIRQARTIFIPKV
	AAGGTGGGGAATGCTTCTCGACCGCAAGACTTTCGTCCAATTACGGTAC		ATTTTCC	GNASRPQDFRPITVQ
	AATCTGTTATGGTAAGGATTTTAAATGCCATTTTGGCTTCCCGGTTGACC		TTCCTCA	SVMVRILNAILASRLT
	TCATCAGTCGACTGGGATCCGCGTCAGCGAGGTTTCCTTCCAACCGACG		GCGATC	SSVDWDPRQRGFLP
	GATGTGCCGATAATACGACGATAGTCGACTTAATCTTAAGGGATCACCA		AAAAAA	TDGCADNTTIVDLILR
	TAAACGTTGTAAATCACTTTATATCGCAACTTTAGATATAAGCAAAGCAT		AA (SEQ	DHHKRCKSLYIATLDI
	TTGACTCGGTGTCTCATGCAGCAGTTAGCGCCACTCTAACTGCATATGGT		ID NO:	SKAFDSVSHAAVSAT
	GCCCCTAAAGAATTCGTTGACTACGTACAAAATTCGTACGAGGTCTGTG		1363)	LTAYGAPKEFVDYVQ
	GCACAACGCTCAATGGGGACGGATGGAGATCAGAGGAATTCATACCTG			NSYEVCGTTLNGDG
	CTCGAGGTGTCAGACAGGGTGACCCGCTATCTCCCATAATATTCAACTT			WRSEEFIPARGVRQG
	GATCATCGATCAGTTGCTTAGGTCCTACCCCAATGAGATTGGTGCCACA			DPLSPIIFNLIIIDQLLRS
	ATCGGTGATCACACAACAAACGCGGCCGCGTTCGCAGATGATATTGTCT			YPNEIGATIGDHTTN
	TATTTGCGGAAACTCGTTTAGGCCTTCAAACAATGCTAGACACGACTGTC			AAAFADDIVLFAETRL
	GATTTTCTATCTTCAGTCGGGCTTACCCTTAACTCGGATAAAATGTTTTACA			GLQTMLDTTVDFLSS
	GTTGGAATAAAGGGGCAACCGAAACAGAAGTGTACTGTGGTCATCCCA			VGLTLNSDKCFTVGIK
	GAGACCTTCCGTATCGGTTCGCGCTCGTGTCCTGCATTGAAGCGCACAG			GQPKQKCTVVIPETF
	ACGAGTGGAAGTATTTAGGCATAACATTCACTGCACAAGGGAGGACCA			RIGSRSCPALKRTDE
	GGTACAGTCCAGCCGACGACCTAGGTCCGAAGCTGTTGAGGCTGACAA			WKYLGITFTAQGRTR
	GGTCCCCCCTAAAACCACAACAGAAATTGTTCGCACTCAGAACAGTTCTT			YSPADDLGPKLLRLTR
	ATCCCACAACTTTACCATAAGCTGACCCTAGGTAGTGTGATGATAGGTG			SPLKPQQKLFALRTVL
	TTCTGAGAAAGTGTGACATACTGGTACGTTCGACCGTAAGGAAGTGGTT			IPQLYHKLTLGSVMIG

			AGGGCTTCCTCTGGACGTGTCAACTGCATTCTTCCATGCTCCTCATACTT				VLRKCDILVRSTVRK
			ATGGGGGCCTCGGAATCCCTTCAGTTCGTTGGGTAGCGCCAATGCTACG				WLGLPLDVSTAFFHA
			TATGAAAAGATTGAGCAATATTAAGTGGGCCCACCTCGCGCAATCCGAG				PHTYGGLGIPSVRWV
			GCGGCCAGCTCATTTCTTACCGACGAATTGAATAAGGCCCGGGGTAGAA				APMLRMKRLSNIKW
			CTCTGGCTGGACTGAATGAGTTGACATCGCGTACAGAGATCGAAACGTA				AHLAQSEAASSFLTD
			CTGGGCGAACAGGTTGTATATGTCTGTTGATGGTCGCGGCTTACGTGAA				ELNKARGRTLAGLNE
			GCGGGACTTTTTCGTCCCCAACACGGCTGGGTGTGTCAGCCCACGCGTT				LTSRTEIETYWANRLY
			TGCTAACAGGTCAAGATTACCGAAACAGTATCAAGCTGCGAATAAATGC				MSVDGRGLREAGLF
			CCTACCATCGAGGTCTCGTACCACGAGGGGCAGAAATGAATTGGAACG				RPQHGWVCQPTRLL
			GCAATGTCGTGCAGGTTGTGATGCTCCCGAAACAACAAACCACATCCTG				TGQDYRNSIKLRINAL
			CAGAATTGTTACCGTACGCATGGGAGGCGGGTAGCAAGACATAACTGT				PSRSRTTRGRNELER
			GTAGTCAATAACCTTAAGAGGATTCTTGAGGAGAAGGGCCACACAGTA				QCRAGCDAPETTNHI
			CACGTCGAACCAAGTTTGCAGCTGGAAACCTCGGTAAGTAAACCAGACC				LQNCYRTHGRRVAR
			TGGTGTGTATCCGTGACAATCACGCTTGCGTGATTGATGCGCAGATTAT				HNCVVNNLKRILEEK
			AACGGATGGACTGTTTCTCGACGATGTGCACCATCGCAAAGTTGAGAAA				GHTVHVEPSLQLETS
			TATAAAAGACCGGAAGTTATATCTGCACTGCGGAGAGAATTCGGAGTG				VSKPDLVCIRDNHAC
			TCGGGCAACGTCGAAGTCCTATCCGCGACGTTAAACTGGCGTGGGATCT				VIDAQIITDGLFLDDV
			GGAGCAATCAATCCGTTAGAAGATTGATAGCAAAGGGTCTCATCTCATC				HHRKVEKYKRPEVIS
			CGGTGACAGCAATGTCATTAGCGCCAGAGTGGTAACAGGCGGACTATA				ALRREFGVSGNVEVL
			TTGCTTCAGACAGTTCATGTATCTCGCAGGTTACACTCGAGATTGGACTT				SATLNWRGIWSNQS
			AGCCTATACACTATGTTGGAGAGAAGACGCTTGCTACCTAGGCAAAATG				VRRLIAKGLISSGDSN
			TGAAATTAGGTATAAACATCGTGGTTGTAAAACTTGAGGTGGGTTTTTA				VISARVVTGGLYCFR
			GTACGTATGCGTGATTACTTCGTAATCATGAATCGTGCATGCTAGTGGG				QFMYLAGYTRDWT
			GTTTGGCCTCCACTAGTATCTTTGAAGATTTTCCTTCCTCAGCGATCAAA				(SEQ ID NO: 1485)
			AAAAA (SEQ ID NO: 1117)				
R2	R2_DPs	Drosophila pseudoobscura	CAATTGGAAAGATATGGGTCTGAATAATAGCGTAGAAGGGGAGTCATT	CAATTGGA	TAGCCT		SSFGLIVTNLNSETVL
			CCGTAATTCGTAAATCGTAAAAATCAGATCAAGTTGATTCAAGACCTCCT	AAGATATG	ATACACT		WGCCQPLGQFSLIGT
			CGTGGTATCTTCTGGATGCTATTAGACTGAAGTTCTTTTGGTCTAATAGT	GGTCTGAA	ATGTTG		NMQNTTPRIINTNSL
			AACTAACTTGAACAGCGAAACAGTCCTATGGGGCTGCCAGCCCCTTGGA	TAATAGCG	GAGAGA		TNQIPTVSSLGAQSE
			CAGTTCAGTTTGATTGGCACTAATATGCAAAATACAACGCCTCGGATAA	TAGAAGG	AGACGC		HSAQVNPNSGYQCTI
			TAAACACTAATTCGTTGACGAACCAAATCCCTACGGTCTCTAGCCTAGG	GGAGTCAT	TTGCTAC		CESSFRSKSGLGVHM
			GGCCCAATCTGAACATAGTGCACAGGTTAACCCAAACAGTGGTTACCAA	TCCGTAAT	CTAGGC		SRRHKDEFDQLRLRT
			TGCACGATATGTGAATCGTCTTTCCGTAGCAAAAGCGGACTAGGCGTTC	TCGTAAAT	ATAATG		DRKAQWSEEELSM
			ACATGTCACGTCGGCACAAGGACGAGTTTGATCAACTTCGTCTGCGTAC	CGTAAAAA	TGAAAT		MARKEIELAANGERY
			CGACCGTAAGGCACAATGGAGTGAGGAAGAGTTGAGTATGATGGCAA	TCAGATCA	TAGGTA		LNKKLAEVFTNRSVD
			GAAAAGAGATCGAGCTCGCAGCAAATGGAGAAAGATATCTAAATAAGA	AGTTGATT	TAAACA		AIKKCRQRERYKTKIE
			AGCTAGCGGAAGTATTTACGAACCGTAGCGTCGACGCTATCAAGAAATG	CAAGACCT	TCGTGG		QLKGQAVPLPEALES
			TCGACAGAGGGAGAGATATAAGACCAAAATCGAACAGCTAAAGGGTCA	CCTCGTGG	TTGTAA		ETIQRRPSIRERDLLV

	AGCTGTTCCTCTCCCAGAAGCATTAGAATCTGAAACCATACAGCGCCGC	TATCTTCT	AACTTG	TPPNTLGTTPTELSNR
	CCTAGTATACGCGAGCGAGATCTCCTAGTAACGCCACCTAACACTCTCG	GGATGCTA	AGGTGG	EILAVLQGYPPVVCN
	GAACCACTCCAACCGAACTGTCGAACAGGGAAATCCTGGCAGTACTACA	TTAGACTG	GTTTTTA	DQWRVEVLQSIVDG
	GGGGTACCCACCTGTAGTATGCAATGACCAATGGAGAGTTGAGGTTTTG	A (SEQ ID	GTACGT	AQASGKEITLQRLSTY
	CAATCCATCGTAGATGGTGCGCAGGCCTCGGGTAAGGAAATTACTCTTC	NO: 1241)	ATGCGT	LMEVFPSQNDRPIQT
	AGCGCTTGTCTACTTACCTTATGGAAGTATTTCCCTCACAGAATGACCGC		GATTACT	RPPRRPRNRRQGRR
	CCCATTCAAACGAGACCTCCACGGAGACCTCGTAATAGGAGACAAGGTA		TCGTAAT	QQYALTQRNWDKH
	GGAGACAGCAGTACGCCTTAACACAGCGTAACTGGGACAAGCACAAAG		CATGAA	KGRCIKAILDGTEGTA
	GTCGTTGTATAAAAGCCATTTTGGATGGAACTGAGGGGACAGCAACTAT		TCGTGC	TMPSQGIMGSYWR
	GCCAAGTCAAGGTATCATGGGGTCCTATTGGAGACAAGTCATGACACAA		ATGCTA	QVMTQTSPTYSGTN
	ACAAGCCCAACATATAGTGGTACGAACACCACGTTCCGGACGGAACACC		GTGGGG	TTFRTEHPLEGVWSP
	CACTTGAAGGGGTTTGGTCCCCGATAACACTAGGGGACCTAAGGGTAC		TTTGGC	ITLGDLRVHRVSLTKS
	ACAGAGTGTCATTGACGAAATCTCCAGGACCTGATGGGATTACTCCAAG		CTCCACT	PGPDGITPRTVRSIPS
	AACTGTCAGGAGTATTCCGTCAGGAGTTATGCTTCGCATAATGAACCTG		AGTATCT	GVMLRIMNLILWCG
	ATACTTTGGTGCGGAAAGTTGCCTGTCTCCATCCGACACAGGCACGAACCA		TTGAAG	KLPVSIRQARTIFIPKV
	TCTTCATTCCGAAGGTGGGGAATGCTTCTCGACCGCAAGACTTTCGTCCA		ATTTTCC	GNASRPQDFRPITVQ
	ATTACGGTACAATCTGTTATGGTAAGGATTTTAAATGCCATTTTGGCTTC		TTCCTCA	SVMVRILNAILASRLT
	CCGGTTGACCTCATCAGTCGACTGGGATCCGCGTCAGCGAGGTTTCCTT		GCGATC	SSVDWDPRQRGFLP
	CCAACCGACGGATGTGCCGATAATACGACGATAGTCGACTTAATCTTAA		AAAAAA	TDGCADNTTIVDLILR
	GGGATCACCATAAACGTTGTAAATCACTTTATATCGCAACTTTAGATATA		AAAAAA	DHHKRCKSLYIATLDI
	AGCAAAGCATTTGACTCGGTGTCTCATGCAGCAGTTAGCGCCACTCTAA		AAAAAA	SKAFDSVSHAAVSAT
	CTGCATATGGTGCCCCTAAAGAATTCGTTGACTACGTACAAAATTCGTAC		AAAAAA	LTAYGAPKEFVDYVQ
	GAGGTCTGTGGCACAACGCTCAATGGGGACGGATGGAGATCAGAGGA		(SEQ ID	NSYEVCGTTLNGDG
	ATTCATACCTGCTCGAGGTGTCAGACAGGGTGACCCGCTATCTCCCATA		NO:	WRSEEFIPARGVRQG
	ATATTCAACTTGATCATCGATCAGTTGCTTAGGTCCTACCCCAATGAGAT		1364)	DPLSPIIFNLIIDQLLRS
	TGGTGCCACAATCGGTGATCACACACAACGCGGCCGCGTTCGCAGAT			YPNEIGATIGDHTTN
	GATATTGTCTTATTTGCGGAAACTCGTTTAGGCCTTCAAACAATGCTAGA			AAAFADDIVLFAETRL
	CACGACTGTCGATTTTCTATCTTCAGTCGGGCTTACCCTTAACTCGGATA			GLQTMLDTTVDFLSS
	AATGTTTTACAGTTGGAATAAAGGGGCAACCGAAACAGAAGTGTACTGT			VGLTLNSDKCFTVGIK
	GGTCATCCCAGAGACCTTCCGTATCGGTTCGCGCTCGTGTCCTGCATTGA			GQPKQKCTVVIPETF
	AGCGCACAGACGAGTGGAAGTATTTAGGCATAACATTCACTGCACAAG			RIGSRSCPALKRTDE
	GGAGGACCAGGTACAGTCCAGCCGACGACCTAGGTCCGAAGCTGTTGA			WKYLGITFTAQGRTR
	GGCTGACAAGGTCCCCCCTAAAACCACACAACAGAAATTGTTCGCACTCAG			YSPADDLGPKLLRLTR
	AACAGTTCTTATCCCACAACTTTACCATAAGCTGACCCTAGGTAGTGTGA			SPLKPQQKLFALRTVL
	TGATAGGTGTTCTGAGGAAGTGTGACATACTGGTACGTTCGACCGTAAG			IPQLYHKLTLGSVMIG
	GAAGTGGTTAGGGCTTCCTCTGGACGTGTCAACTGCATTCTTCCATGCTC			VLRKCDILVRSTVRK
	CTCATATTTATGGGGGGCCTCGGAATCCCTTCAGTTCGTTGGGTAGCGCC			WLGLPLDVSTAFFHA
	AATGCTACGTATGAAAAGATTGAGCAATATTAAGTGGGCCCACCTCGCG			PHIYGGLGIPSVRWV

			CAATCCGAGGCGGCCAGCTCATTTCTTACCGACGAATTGAATAAGGCCC				APMLRMKRLSNIKW
			GGGGTAGAACTCTGGCTGGACTGAATGAGTTGACATCGCGTTCGGAGA				AHLAQSEAASSFLTD
			TCGAAACGTACTGGGCGAACAGGTTGTATATGTCTGTTGATGGTCGCGG				ELNKARGRTLAGLNE
			CTTACGTGAAGCGGGACTTTTTCGTCCCCAACACGGCTGGGTGTGTCAG				LTSRSEIETYWANRLY
			CCCACGCGTTTGCTAACAGGTCAAGATTACCGAAACGGTATCAAGCTGC				MSVDGRGLREAGLF
			GAATAAATGCCCTACCATCGAGGTCTCGTACCACGAGGGGCAGAAATG				RPQHGWVCQPTRLL
			AATTGGAACGGCAATGTCGTGCAGGTTGTGATGCTCCCGAAACAACAAA				TGQDYRNGIKLRINA
			CCACATCCTGCAGAATTGTTACCGTACGCATGGGAGGCGGGTAGCAAG				LPSRSRTTRGRNELER
			ACATAACTGTGTAGTCAATAACCTTAAGAGGATTCTTGAGGAGAAGGGC				QCRAGCDAPETTNHI
			CACACAGTACACGTCGAACCAAGTTTGCAGCTGGAAACCTCGGTAAGTA				LQNCYRTHGRRVAR
			AACCAGACCTGGTGTGTATCCGTGACAATCACGCTTGCGTGATTGATGC				HNCVVNNLKRILEEK
			GCAGATTATAACGGATGGACTGTTTCTCGACGATGTGCACCATCGCAAA				GHTVHVEPSLQLETS
			GTTGAGAAATATAAAAGACCGGAAGTTATATCTGCACTGCGGAGAGAA				VSKPDLVCIRDNHAC
			TTCGGAGTGTCGGGCAACGTCGAAGTCCTATCCGCGACGTTAAACTGGC				VIDAQIITDGLFLDDV
			GTGGGATCTGGAGCAATCAATCCGTTAGAAGATTGATAGCAAAGGGTC				HHRKVEKYKRPEVIS
			TCATCTCATCCGGTGACAGCAATGTCATTAGCGCCAGAGTGGTAACAGG				ALRREFGVSGNVEVL
			CGGACTATATTGCTTCAGACAGTTCATGTATCTCGCAGGTTACACTCGAG				SATLNWRGIWSNQS
			ATTGGACTTAGCCTATACACTATGTTGGAGAGAAGACGCTTGCTACCTA				VRRLIAKGLISSGDSN
			GGCATAATGTGAAATTAGGTATAAACATCGTGGTTGTAAAACTTGAGGT				VISARVVTGGLYCFR
			GGGTTTTTAGTACGTATGCGTGATTACTTCGTAATCATGAATCGTGCATG				QFMYLAGYTRDWT
			CTAGTGGGGTTTGGCCTCCACTAGTATCTTTGAAGATTTTCCTTCCTCAG				(SEQ ID NO: 1486)
			CGATCAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1118)				
R2	R2_DSe	.	GGGATCAGGGGTAATTGCGAGCAGAGGGGGAGTATTTTTCTGTAATTC	GGGATCA	TAGCTA		FERQNFSDGLVPQRK
		Drosophila sechellia	GTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTTACTCTGTAACTC	GGGGTAAT	AAACGT		FIHIGTTNRNNEPRS
			ACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTGGTCC	TGCGAGCA	TTGGTTC		NLRNLMTTRPSVDIF
			CCGGTAATGCTAAACTTGTTTAGCAGCTAATTTGAGCGGCAAAACTTTTC	GAGGGGG	AAAACA		PEDQYEPNAAATLSR
			CGATGGGTTGGTTCCCCAGAGGAAATTTATTCATATTGGAACTACAAAT	AGTATTTT	TTTGCTT		VPCTVCGRSFNSKRG
			AGAAATAACGAGCCTCGGAGCAATTTACGCAATCTGATGACGACCCGAC	TCTGTAAT	GCTGTC		LGVHMRSRHPDELD
			CCTCCGTGGACATCTTCCCGGAGGACCAATATGAACCAAACGCAGCGGC	TCGTAAGT	TTGGCA		EERRRVDIKARWSEE
			TACTCTATCTAGGGTACCCTGCACAGTATGTGGCCGGTCCTTTAACAGCA	CATATCAT	TAACATC		EKWMMARKEVELTA
			AGAGAGGACTCGGTGTTCACATGCGATCTCGGCACCCAGACGAACTTG	ATGGTGTG	AATAAA		NGHKHINKQLAVYFA
			ATGAAGAACGTCGACGTGTCGATATAAAGGCAAGATGGAGTGAGGAA	CGGAAGG	GGCATA		NRSVEAIKKLRQRGD
			GAGAAGTGGATGATGGCGAGAAAGGAGGTTGAGCTCACAGCAAATGG	GGAATTTT	AACATC		YKEKIEQIRRQSALVP
			ACATAAACACATAAACAAGCAACTAGCGGTGTATTTTGCAAACCGCAGC	ACTCTGTA	GCAAAT		EVANLTIRRRPSRSEQ
			GTCGAAGCCATCAAAAAGCTAAGACAAAGGGGCGATTATAAGGAGAAA	ACTCACAA	AATGGT		NHQVTTSETTPITPFE
			ATAGAGCAGATAAGAAGGCAATCAGCTCTCGTCCCGGAAGTTGCAAATC	GTCTCTCC	AATATAT		QSNREILRTLRGYSPV
			TAACCATAAGGCGCCGCCCTAGTAGAAGTGAGCAAAACCACCAAGTAA	TTTACTCA	AAATTG		ECHSKWRAQELQTII
			CAACATCAGAAACAACTCCAATCACACCCTTCGAACAGTCGAACAGGGA	AGTCGACT	GCTATG		DRAELEGKETTLQCLS

		AATTTTGCGGACACTGCGTGGGTATAGCCCCGTAGAATGCCATTCCAAA	CAAAACCT	AGGATG	LYLLGIFPAQGVRHTL
		TGGAGAGCCCAAGAGCTACAAACGATCATTGACAGGGCAGAGCTCGAG	CCTCGTGG	GTTTTAG	TRPPRRPRNRRRESRR
		GGAAAGGAAACCACTCTCCAATGCTTATCGCTATATCTCCTAGGAATTTT	TGGTCCCC	TACGTA	QQYAVVQRNWDKH
		TCCGGCACAGGGTGTACGTCACACTCTGACGAGACCTCCTCGGAGGCCT	GGTAATGC	GGCGTT	KGRCIKSLLNGTDESV
		CGAAATAGGAGAGAAAGCAGAAGGCAGCAGTATGCTGTCGTCCAGCGT	TAAACTTG	GCGGAA	MPSQEVMVPYWRE
		AACTGGGATAAGCATAAAGGAAGATGCATCAAGTCCTTGCTAAATGGA	TTTAGCAG	CTTCGG	VMTQPSPSSCSREVI
		ACTGATGAGTCGGTAATGCCAAGCCAAGAAGTAATGGTTCCTTACTGGA	CTAA (SEQ	TTCAGAT	QMDHSLERVWSAIT
		GAGAAGTAATGACTCAGCCTAGCCCAAGCTCTTGCAGTAGAGAAGTGAT	ID NO:	AGAGCA	EHDLRASRISLSSSPG
		ACAAATGGATCACTCGCTTGAGAGGGTTTGGTCTGCTATTACAGAGCAC	1242)	ATGAAT	PDGITPKTAREVPSGI
		GACCTTCGGGCATCAAGAATCTCATTATCTTCATCTCCGGGGCCTGACG		CGTGCA	MLRIMNLILWCGNL
		GGATAACTCCAAAAACAGCCAGGGAGGTGCCGTCAGGTATTATGTTGC		TGCTAG	PHSIRLARTVFIPKTVT
		GAATAATGAACCTAATTCTATGGTGCGGCAATCTACCACACTCTATCCGA		GAAACT	AKRPQDFRPISVPSVL
		CTGGCCAGAACCGTCTTCATCCCGAAGACGGTGACGGCGAAGCCGACCG		GAAGTG	VRQLNAILATRLNSSI
		CAAGACTTTCGTCCAATATCGGTGCCTTCAGTCCTGGTAAGACAGCTAA		TTGACA	NWDPRQRGFLPTDG
		ATGCAATATTGGCAACCCGGTTGAACTCATCAATCAATTGGGACCCGCG		GACCTA	CADNATIVDLVLRHS
		CCAGCGGGGCTTCTTACCTACCGACGGATGTGCCGATAACGCGACGATA		GTATCTT	HKHFRSCYIANLDVS
		GTTGACTTAGTCTTGAGGCATAGCCATAAGCACTTTAGATCTTGCTACAT		TCGATA	KAFDSLSHASIYDTLR
		TGCAAATTTAGATGTAAGCAAGGCATTCGATTCTTTATCGCATGCATCTA		GATTTCC	AYGAPKGFVDYVQN
		TATATGACACCTTACGTGCTTATGGTGCGCCAAAGGGCTTCGTTGACTAC		ATACCTC	TYEGGGTSLNGDGW
		GTACAGAATACGTACGAGGGTGGCGGTACCAGTCTCAATGGGGACGGT		CGCGAT	SSEEFVPARGVKQGD
		TGGAGTTCAGAGGAATTCGTCCCTGCTAGAGGAGTGAAGCAGGGTGAC		CAAAAA	PLSPILFNLVMDRLLR
		CCTTTGTCTCCTATTCTATTTAACTTGGTAATGGACAGGTTACTTAGAAAC		AAAAAA	NLPSEIGARVGNAIT
		CTACCCAGCGAAATTGGTGCCAGAGTCGGAAATGCCATTACTAACGCGG		AAAAAA	NAAAFADDLVLFAET
		CCGCGTTTGCAGATGATTTGGTACTATTTGCTGAAACTCGAATGGGACTT		AAAA	RMGLQVLLDRTLDFL
		CAAGTATTGTTGGACAGAACGTTGGATTTTCTATCTCTCGTCGGCCTCAA		(SEQ ID	SLVGLKLNADKCFTV
		ACTTAATGCCGACAAATGTTTTACCGTTGGCATTAAGGGCCAGCCGAAA		NO:	GIKGQPKQKCTVLEA
		CAGAAGTGTACCGTGCTAGAGGCACAGAGCTTTTACGTAGGCTCAAGG		1365)	QSFYVGSREIPSLKRT
		GAGATTCCATCATTGAAGCGAACGGACGAGTGGAAGTACTTAGGCATC			DEWKYLGINFTATGR
		AACTTCACTGCAACCGGGAGAGGGTTCGATGCAATCCGGCCGAGGACATT			VRCNPAEDIGPKLQR
		GGTCCGAAGCTACAAAGATTGACAAAGGCCCCCCTCAAACCACAACAGA			LTKAPLKPQQRMFAL
		GGATGTTCGCCCTTAGGACTGTCCTTATCCCACAGCTCTATCACAAGTTA			RTVLIPQLYHKLALGS
		GCCCTTGGGAGTGTGGCGATAGGCATCCTACGAAAAACTGACAAACTA			VAIGILRKTDKLIRYYV
		ATACGATACTATGTGCGAAGATGGCTAAATCTTCCGCTGGATATACCGA			RRWLNLPLDIPIAFIH
		TAGCATTCATTCACGCACCCCCAAAAAGTGGAGGTCTCGGAATTCCATC			APPKSGGLGIPSLRW
		ACTTAGATGGGTAGCTCCAATGTTAAGGCTAAGACGTTTGAGTAATATT			VAPMLRLRRLSNIKW
		AAATGGCCTCACCTCACGCAAAACGAGGTAGCCAGCTCTTTCCTCGAAG			PHLTQNEVASSFLEA
		CAGAAAAACAACGGGCCCGAGATAGATTATTAGCAGAACAAAATGAAT			EKQRARDRLLAEQNE
		TGTTATCGCGTCCGGCAATAGAAAAATATTGGGCGAACAAATTGTACCT			LLSRPAIEKYWANKLY

			CTCAGTTGATGGTAGCGGACTCCGTGAAGCAGGCCATTGGGGCCCGCA				LSVDGSGLREAGHW
			ACACGGGTGGGTTAATCAACCCACGCGTTTACTAACAGGAAAGGAATAT				GPQHGWVNQPTRLL
			ATAGACGGTATTCGTCTGCGGATAAATGCCCTACCCACGAAGTCTCGTA				TGKEYIDGIRLRINALP
			CTACAAGGGGAAGGCACGAATTGGAACGACAGTGTCGTGCAGGATGTG				TKSRTTRGRHELERQ
			ACGCTCCCGAAACAACAACCATATTATGCAAAAATGTTACCGATCGCA				CRAGCDAPETTNHI
			TGGGAGGCGCATAGCTAGACACAACTGCGTAGTAAATCGAATCAAGCG				MQKCYRSHGRRIAR
			GGGACTTGAGGAGAGAGGCTGCGTAGTCATTGTTGAACCAAGTCTGCA				HNCVVNRIKRGLEER
			GTGCGAATCCGGTCTTAATAAACCAGACCTGGTGGCACTACGACAAAAT				GCVVIVEPSLQCESGL
			CACATTGATGTGATCGACATACAAATTGTGACAGACGGACACTCTATGG				NKPDLVALRQNHIDV
			ATGATGCACACCAGCGCAAAATCAATAGATACGACAGACCGGACATAC				IDIQIVTDGHSMDDA
			GAACTGAGTTGCGTCGCAGATTCGAAGCCGCAGGTGACATTGAATTCCA				HQRKINRYDRPDIRT
			TTCTGCCACCCTGAACTGGAGGGGGATCTGGAGTGGTCAATCCGTTAAA				ELRRRFEAAGDIEFHS
			AGATTGATAGCGAAGGGTCTCCTCAGCAAATATGATAGTCATATCATTA				ATLNWRGIWSGQSV
			GCGTCCAGGTTATGAGAGGCAGTCTCGGTTGTTTTAAACAGTTCATGTA				KRLIAKGLLSKYDSHII
			CCTGAGCGGGTTTTCCCGAGATTGGACTTAGCTAAAACGTTTGGTTCAA				SVQVMRGSLGCFKQ
			AACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACATCGC				FMYLSGFSRDWT
			AAATAATGGTAATATATAAATTGGCTATGAGGATGGTTTTAGTACGTAG				(SEQ ID NO: 1487)
			GCGTTGCGGAACTTCGGTTCAGATAGAGCAATGAATCGTGCATGCTAG				
			GAAACTGAAGTGTTGACAGACCTAGTATCTTTCGATAGATTTCCATACCT				
			CCGCGATCAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1119)				
R2	R2_DSi	Drosophila simulans	GGGATCTGGGGTAATTGCGAGCAGAGGGGGAGTATTTTTCTGTAATTC	GGGATCTG	TAGCTA	TCLAANLSGKNFSDG
			GTAAGTCATATCATATGGTGTGCGGAAGGGGGAATTTTACTCTGTAACTC	GGGTAATT	AAACGT	LVTQRKFTHIGTTNT
			ACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTGGTCC	GCGAGCA	TTGGTTC	NNEPRISLHNLMTTR
			CCGGTAATGCTAAACTTGTTTAGCAGCTAATTTGAGCGGCAAAAACTTTT	GAGGGGG	AAAACA	PSVDIFPEDQYEPNA
			CCGATGGGCTGGTTACCCAGAGGAAATTTACTCATATTGGAACTACGAA	AGTATTTT	TTTGCTT	AATLSRVPCTVCGRS
			CACAAATAACGAGCCTCGGATATCTTTACACAATCTGATGACGACCCGA	TCTGTAAT	GCTGTC	FNSKRGLGVHMRSR
			CCCTCCGTGGATATCTTCCCGGAGGACCAATATGAACCAAACGCAGCGG	TCGTAAGT	TTGGCA	HPDELDEERRRVDIK
			CTACTCTATCTAGGGTTCCCTGCACAGTATGTGGCCGGTCCTTTAACAGC	CATATCAT	TAACATC	ARWSEEEKWMMAR
			AAGAGAGGACTCGGTGTTCACATGCGATCTCGGCACCCAGACGAACTT	ATGGTGTG	AATAAA	KEVELTANGHKHMN
			GATGAAGAACGTCGACGTGTCGATATAAAGGCAAGGTGGAGTGAGGA	CGGAAGG	GGCATA	KQLAVYFANRSVEAI
			AGAGAAGTGGATGATGGCGAGAAAGGAGGTTGAGCTCACAGCAAATG	GGAATTTT	AACATC	KKLRQRGDYKEKIEQI
			GACATAAACACATGAACAAGCAACTAGCGGTGTATTTTGCAAACCGCAG	ACTCTGTA	GCAAAA	RGQSALVPEVANLTI
			CGTCGAAGCCATCAAAAAGCTAAGACAGAGGGGCGATTATAAGGAGAA	ACTCACAA	TAATGG	RRRPSRSEQNHQVTT
			AATAGAGCAGATAAGAGGGCAATCTGCTCTCGTCCCGGAAGTTGCAAAT	GTCTCTCC	TTATATA	SETTPITPFEQSNREIL
			CTAACCATAAGGCGCCGCCCTAGTAGAAGTGAGCAAAACCACCAAGTA	TTTACTCA	TAAATG	RTLRGYSPVECHSKW
			ACAACATCAGAAACAACTCCAATCACTCCCTTCGAACAGTCGAACAGGG	AGTCGACT	GCTATG	RAQELQTIIDRAELEG
			AAATTTTGCGGACACTGCGTGGGTATAGCCCCGTAGAATGCCATTCCAA	CAAAACCT	AGGATG	KETTLQCLSLYLLGIFP
			ATGGAGAGCCCAAGAGCTACAAACGATCATTGACAGGGCAGAGCTCGA	CCTCGTGG	GTTTTAG	AQGVRHTLTRPPRRP

	GGGAAAGGAAACCACTCTCCAATGCTTATCGCTATATCTCCTGGGAATTT	TGGTCCCC	TACGTA	RNRRESRRQQYAVV
	TTCCGGCACAGGGTGTACGACACACGCTGACGAGACCTCCTCGGAGAC	GGTAATGC	GGCGTT	QRNWDKHKGRCIKS
	CTCGGAATAGGAGAGAAAGCAGAAGGCAGCAGTATGCTGTCGTCCAGC	TAA (SEQ	GCGGAA	LLNGTDESVMPSQEV
	GTAACTGGGATAAGCATAAAGGAAGATGCATCAAGTCCTTGCTAAATG	ID NO:	CTTCGG	MVPYWREVMTQPS
	GAACTGATGAGTCGGTAATGCCAAGCCAAGAAGTAATGGTTCCCTACTG	1243)	TTCAGAT	PSSCSGEVIQMDHSL
	GAGAGAAGTAATGACTCAGCCTAGCCCAAGCTCTTGCAGTGGAGAAGT		AGAGCA	ERVWSAITEHDLRAS
	GATACAAATGGATCACTCGCTTGAGAGGGTTTGGTCTGCTATTACGGAG		ATGAAT	RISLSSSPGPDGITPKS
	CATGACCTTCGGGCGTCAAGAATCTCATTATCTTCATCTCCGGGGCCTGA		CGTGCA	AREVPSGIMLRIMNLI
	CGGGATAACTCCAAAATCTGCCAGGGAGGTGCCGTCAGGTATTATGTTG		TGCTAG	LWCGNLPHSIRLART
	CGAATAATGAACCTAATTCTATGGTGCGGTAATCTACCACACACTCTATCCG		GAAAAC	VFIPKTVTAKRPQDF
	ACTGGCCAGAACCGTCTTCATCCCGAAAACGGTGACGGCGAAGCGACC		TGACCA	RPISVPSVLVRQLNAI
	GCAAGACTTTCGTCCAATATCGGTGCCTTCAGTCCTGGTAAGACAGCTA		CACGCA	LATRLNSSINWDPRQ
	AATGCCATATTGGCAACCCGGTTGAACTCATCAATCAATTGGGACCCGC		GTGTTG	RGFLPTDGCADNATI
	GCCAGCGGGGCTTCTTACCTACCGACGGATGTGCCGATAATGCGACGAT		GCAGCC	VDLVLRHSHKHFRSC
	AGTTGACTTAGTCTTGAGGCATAGCCATAAGCACTTTAGATCTTGCTACA		CTAGTAT	YIANLDVSKAFDSLSH
	TAGCTAATTTAGATGTAAGCAAGGCATTCGATTCTTTATCGCATGCATCT		CTTTCGA	ASIYDTLRAYGAPKGF
	ATATATGACACCTTACGTGCTTATGGTGCGCCAAAGGGCTTCGTTGACT		TAGATTT	VDYVQNTYEGGGTSL
	ACGTACAGAATACGTACGAGGGTGGCGGTACCAGTCTCAATGGGGACG		CCATACC	NGDGWSSEEFVPAR
	GTTGGAGTTCAGAGGAATTCGTCCCTGCTAGAGGAGTGAAGCAGGGTG		TCCGCG	GVKQGDPLSPILFNL
	ACCCTTTGTCTCCTATTCTATTTAACTTGGTAATGGACAGGTTACTTAGAA		ATCAAA	VMDRLLRNLPSEIGA
	ACCTACCCAGCGAAATTGGTGCCAAAGTCGGAAATGCCATTACTAACGC		AAAAAA	KVGNAITNAAAFAD
	GGCCGCGTTTGCAGATGATTTGGTACTATTTGCTGAAACTCGAATGGGA		AAAAAA	DLVLFAETRMGLQVL
	CTTCAAGTATTGTTGGACAAAACGTTGGATTTTCTATCTCTCGTCGGCCT		AAAAAA	LDKTLDFLSLVGLKLN
	CAAACTTAATGCCGACAAATGTTTTACCGTTGGCATTAAGGGCCAGCCG		AA (SEQ	ADKCFTVGIKGQPKQ
	AAACAGAAGTGTACCGTGCTAGAGGCACAGAGCTTCTACGTAGGCTCG		ID NO:	KCTVLEAQSFYVGSR
	AGGGAGATTCCATCACTGAAGCGAACGGACGAGTGGAAGTACTTAGGC		1366)	EIPSLKRTDEWKYLGI
	ATCAACTTCACTGCAACTGGGAGGGTTCGATGCAATCCGGCCGAGGAC			NFTATGRVRCNPAE
	ATTGGTCCAAAGCTACAAAGATTGACAAAGGCCCCCCTCAAACCACAAC			DIGPKLQRLTKAPLKP
	AGAGGATGTTCGCCCTTAGGACTGTCCTTATCCCACAGCTCTATCACAAG			QQRMFALRTVLIPQL
	TTAGCCCTTGGGAGTGTGGCGATAGGCGTCCTACGAAAAACTGACAAAT			YHKLALGSVAIGVLRK
	TAATAAGATATTATGTGCGAAGATGGCTAAATCTTCCGCTGGATGTGCC			TDKLIRYYVRRWLNL
	GATAGCATTCATTCATGCACCCCCAAAAAGTGGAGGTCTCGGAATTCCA			PLDVPIAFIHAPPKSG
	TCACTAAGATGGGTAGCTCCAATGTTAAGGCTAAGACGATTGAGTAATA			GLGIPSLRWVAPMLR
	TTAAATGGCCTCACCTCACGCAAAACGAGGTAGCCAGCTCTTTCCTCGA			LRRLSNIKWPHLTQN
	AGCCGAAAAACAACGGGCCCGAGATAGATTATTAGCAGAACACAAATGA			EVASSFLEAEKQRAR
	ATTGTTATCGCGTCCGGCAATAGAAAAATATTGGGCGAACAAATTGTAC			DRLLAEQNELLSRPAI
	CTCTCAGTTGATGGCAGCGGACTCCGTGAAGCGGGCCATTGGGGACCG			EKYWANKLYLSVDGS
	CAACACGGGTGGGTTAATCAACCCACGCGTTTACTAACAGGAAAGGAAT			GLREAGHWGPQHG

			ATATAGACGGTATTCGTCTGCGGATAAATGCCCTACCCACGAAGTCTCG				WVNQPTRLLTGKEYI
			TACTACAAGGGGAAGGCACGAATTGGAACGACAGTGTCGTGCAGGATG				DGIRLRINALPTKSRT
			TGACGCTCCCGAAACAACAAACCACATAATGCAAAAATGTTACCGATCG				TRGRHELERQCRAGC
			CATGGGAGGCGCGTAGCTAGACACAACTGCGTAGTAAATCGAATCAAG				DAPETTNHIMQKCY
			CGGGGACTTGAGGAGAGAGGCTGCGTGGTCATTGTTGAACCAAGTCTG				RSHGRRVARHNCVV
			CAGTGCGAATCCGGCCTTAATAAACCGGACCTGGTGGCACTACGACAA				NRIKRGLEERGCVVIV
			GATCACATTGATGTGATCGACATACACAAATTGTGACACAGACGGACACTCTA				EPSLQCESGLNKPDL
			TGGATGATGCACACCAGCGCAAAATCAATAGATACGACAGACCGGACA				VALRQDHIDVIDIQIV
			TACGAACTGAATTGCGTCGCAGATTCGAAGCCGCAGGTGACATTGAATT				TDGHSMDDAHQRKI
			CCATTCTGCCACCCTGAACTGGAGGGGGGATCTGGAGTGGTCAATCCGTT				NRYDRPDIRTELRRRF
			AAAAGATTGATAGCAAAGGGTCTCCTCAGCAAATATGATAGTCATATCA				EAAGDIEFHSATLNW
			TTAGCGTCCAGGTTATGAGAGGCAGTCTCGGTTGTTTTAAACAGTTCAT				RGIWSGQSVKRLIAK
			GTACCTGAGCGGGTTTTCCCGAGATTGGACTTAGCTAAAACGTTTGGTT				GLLSKYDSHIISVQVM
			CAAAACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACAT				RGSLGCFKQFMYLSG
			CGCAAAATAATGGTTATATATAAATGGCTATGAGGATGGTTTTAGTACG				FSRDWT (SEQ ID
			TAGGCGTTGCGGAACTTCGGTTCAGATAGAGCAATGAATCGTGCATGCT				NO: 1488)
			AGGAAAACTGACCACACGCAGTGTTGGCAGCCCTAGTATCTTTCGATAG				
			ATTTCCATACCTCCGCGATCAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID				
			NO: 1120)				
R2	R2_DY a	.	GGGAACATGGGGTAAAGGTGAGTAGAGGGGGAGTATTTTTTATACTCT	GGGAACAT	TAGCTTA		FERRIFPKGLVPLTKD
		Drosophila yakuba	GCAACTCATAAGTCTTGCCTTTACTCAAGTCGACTCAAAACCTCCTCGTG	GGGGTAA	AAACGT		NHIGTTNLQNEPRIFT
			GTGTTTCCCGGTAATGTTAAACTTGTTTAGCAGCTAATTTGAGCGGCGA	AGGTGAGT	TTGGTTC		NDLLTTRPSVDHVPE
			ATCTTTCCGAAAGGGTTGGTTCCCCTGACGAAGGATAATCATATTGGTA	AGAGGGG	ACATAC		DQYEPNAAATLSRVP
			CCACAAATTTACAAAACGAGCCTCGGATATTTACTAATGATCTGTTGACG	GAGTATTT	ATCTGCC		CTVCDRSFNSKRGLG
			ACCCGACCCTCCGTGGATCACGTCCCGGAGGACCAATATGAACCAAACG	TTTATACT	TGCTGC		VHMRSRHPDELDEE
			CAGCGGCTACTCTATCAAGGGTTCCCTGCACAGTATGTGACCGGTCCTTT	CTGCAACT	CTTGGC		RRRVDIKARWSEEEK
			AACAGTAAGAGAGGACTCGGTGTTCACATGCGATCTCGGCACCCAGAC	CATAAGTC	ACAATA		WMMARKEVELMAN
			GAACTTGATGAAGAACGTCGACGTGTCGATATAAAAGCAAGGTGGAGT	TTGCCTTT	TCAAAA		GFKHINKQLAVYFAN
			GAGGAAGAGAAGTGGATGATGGCGAGAAAGGAGGTCGAGCTCATGGC	ACTCAAGT	AGGCAT		RSVEAIKKLRQRGDY
			AAATGGTTTTAAACACATAAACAAGCAACTAGCGGTGTATTTTGCAAAC	CGACTCAA	AAACAT		KEKIEQIRGQSALAPE
			CGTAGCGTCGAAGCCATTAAAAAGCTGAGACAGAGGGGCGATTATAAG	AACCTCCT	CGCACA		VANLTIRRRPSRSEQ
			GAGAAAATAGAGCAGATAAGAGGGCAATCCGCTCTCGCCCCAGAAGTT	CGTGGTGT	TATTGGT		DHQVPTSEASPITPLE
			GCTAATCTAACCATAAGGCGCCGCCCTAGTAGAAGTGAGCAAGACCACC	TTCCCGGT	TATTTAC		QSNREILRTLRGYSPV
			AAGTACCAACGTCAGAAGCATCTCCAATCACTCCGCTCGAACAGTCGAA	AATGTTAA	GGCTAT		VCPSKWRAQELQTII
			CAGGGAAATTTTGCGGACGCTGCGTGGGTATAGCCCCGTAGTATGCCCT	ACTTGTTT	GAGGAT		DRAEFEGKETTLQCLS
			TCCAAATGGAGAGCCCAAGAACTACAAACTATCATTGATAGGGCGGAAT	AGCAGCTA	GGTTTTA		LYLQGIFPVQGVRHT
			TTGAGGGAAAGGAAACCACTCTCCAATGCTTATCGCTCTACCTCCAGGG	A (SEQ ID	GTACGT		LTRPPRRPRNRRESR
			AATTTTTCCGGTACAGGGTGTACGACACACGCTGACGAGGCCTCCTCGG	NO: 1244)	AGGCGT		RQQYAVIQRNWDKH

		AGACCTCGGAATAGGAGGGAAAGCAGAAGGCAGCAGTACGCTGTCATC		TGCGGA	KGRCIKSLLNGTDESV
		CAGCGAAACTGGGATAAGCATAAAGGAAGGTGCATTAAGTCCCTGCTT		ACTTCG	MPSREFMEPYWREV
		AATGGAACTGATGAATCGGTAATGCCAAGCCGAGAATTTATGGAGCCCT		GTTCGG	MTQPSPSSCNGEVIR
		ACTGGAGAGAAGTAATGACTCAGCCTAGCCCAAGCTCTTGCAATGGAG		ATAGAG	TDHSLETVWSAITEQ
		AAGTGATTCGTACGGATCACTCGCTTGAGACGGTATGGTCTGCAATAAC		CAATGA	DLRASRVSLSSSPGPD
		GGAACAAGACCTTAGGGCATCAAGAGTTTCATTATCTTCATCTCCGGGG		ATCGTG	GVTPKTAREVPSGIM
		CCTGACGGGGTAACTCCAAAAACTGCCAGGGAGGTGCCGTCAGGTATT		CATGCT	LRIMNLILWCGNLPH
		ATGCTACGCATAATGAACCTAATTCTATGGTGCGGTAACTTACCTCATTC		AGGAAC	SIRLARTIFIPKTVTAK
		TATCCGACTGGCCAGAACCATCTTCATCCCGAAGACGGTGACGGCAAAG		TGACCA	RPQDFRPISVPSVLVR
		CGACCGCAAGACTTTCGTCCAATATCGGTGCCTTCCGTCCTGGTAAGAC		AATAAC	QLNAILATRLTSSIDW
		AGCTAAATGCCATCTTGGCAACCCGATTGACCTCATCAATCGATTGGGA		AGCAGC	DPRQRGFSPTDGCA
		CCCGCGCCAGCGGGGCTTCTCACCAACCGACGGTTGCGCCGATAATGCG		CCTAGT	DNATIVDLVLRHSHK
		ACGATAGTTGACTTAGTCCTGAGGCATAGCCATAAGTACTTTAAATCTTG		ATCTTTC	YFKSCYIANLDVSKAF
		CTACATCGCCAACTTAGATGTTAGCAAGGCATTTGACTCATTGTCACATG		GAAGAT	DSLSHAAIYGTLRAYG
		CAGCAATATATGGGACATTACGAGCTTATGGTGCGCCGAAGGGTTTTGT		TTCCATA	APKGFVDYVQKTYEG
		TGACTATGTACAGAAGACGTACGAGGGAGGTGGTATCAGTCTCAACGG		CCTTTGC	GGISLNGEGWCSEEF
		GGAAGGTTGGTGTTCAGAGGAATTCGTGCCTGCTAGAGGAGTGAAGCA		GATCAA	VPARGVKQGDPLSPI
		GGGCGACCCTTTGTCCCCCATTCTATTTAACTTGGTCATCGACCGGTTAC		AAAAAA	LFNLVIDRLLRALPSEI
		TTAGAGCCCTACCTAGCGAGATTGGTACCAAGGTCGGAAATGCCATGAT		AAAAAA	GTKVGNAMINAAAF
		AAACGCTGCTGCATTTGCAGATGATTTGGTACTATTTGCGGAAACTCGG		AAAAA	ADDLVLFAETRMGL
		ATGGGACTTCAAACTTTGTTGGACAAGACTGTGGACTTTTTATCCACCGT		(SEQ ID	QTLLDKTVDFLSTVGL
		CGGCCTTAAACTTAATGCTGATAAGTGCTTTACTGTCGGTATTAAGGGA		NO:	KLNADKCFTVGIKGQ
		CAGCCGAAACAGAAGTGTACTGTGCTAGAGGCACAGAGCTTCTGCGTA		1367)	PKQKCTVLEAQSFCV
		GGCTCGAGAGAGATTCCAACACTGAAGCGTACTGACGAGTGGAAGTAT			GSREIPTLKRTDEWK
		CTCGGTATACATTTCACTGCAAGTGGGAGGGTTCGATGCAATCCGGCAG			YLGIHFTASGRVRCN
		AGGACATTGGTCCAAAGCTACAAAGATTGTCAGAGGCCCCCCTTAAGCC			PAEDIGPKLQRLSEAP
		ACAACAGAGGTTGTTCGCCCTTCGGACTGTCCTGATCCCACAACTCTATC			LKPQQRLFALRTVLIP
		ACAAGTTATCCCTTGGGAGTGTGACGATAGGCGTCTTACGAAAGACTGA			QLYHKLSLGSVTIGVL
		CAAGCTAATACGTTTCTATGTGCGAAGATGGCTAAATCTTCCGTCGGAT			RKTDKLIRFYVRRWL
		GTGCCGATAGCATTCGTTCATGCCCCCCCAAAATGTGGGGGTCTCGGAA			NLPSDVPIAFVHAPP
		TTCCATCACTAAGATGGGTAGCACCAATGTTACGACTTAGACGATTGAG			KCGGLGIPSLRWVAP
		CAACATAAAATGGCCCCACCTCGTACAATCCGAGGAAGCCAGCTCCTTC			MLRLRRLSNIKWPHL
		ATCGAAGCGGAAAAACAAAGGGCCCGAGGTAGATTGATAGCTGAACAA			VQSEEASSFIEAEKQR
		AATGAATTGTTATCGCGTCCGGCAATAGAAAAGTATTGGGCGAACAGG			ARGRLIAEQNELLSRP
		TTGTACCTCTCCGTTGATGGTGGCGGACTCCGTGAAGCGGGCCACTATG			AIEKYWANRLYLSVD
		GTCCCCAACACGGGTGGGTTAGTCAGCCCACGCGTTTACTAACAGGAAA			GGGLREAGHYGPQH
		GGAATATTTAGACGGTATTCGGCTGCGGATAAATGCCCTACCCACAAAG			GWVSQPTRLLTGKEY
		TCTCGCACTACGAGGGGAAGGCACGAATTGGAGAGGCAGTGTCGTGCA			LDGIRLRINALPTKSR

			GGATGTGATGCTCCCGAAACAACAACCACATTATGCAAAAATGCTATC				TTRGRHELERQCRAG
			GTTCGCATGGGAGGCGAGTAGCTAGACACAACTGCGTAGTAAATCGAA				CDAPETTNHIMQKCY
			TCAAGCGGGGACTTGAGGAGAGAGGCTGCGTGGTCATTGCCGAACCAA				RSHGRRVARHNCVV
			GTCTGCAGTGCGAATCCGGCCTTAATAAACCGGACCTGGTGGTACTCCG				NRIKRGLEERGCVVIA
			ACAAAATCACATTGATGTGATTGACGTGCAGGTTGTGACAGACGGACAT				EPSLQCESGLNKPDL
			TCTATGGACGAAGCGCATCAGCGCAAAATCAATAGATACGACAGACCG				VVLRQNHIDVIDVQV
			GACATACGAACTGAATTGCGTCGCCGATTCGAAGCCGCAGGTGACATC				VTDGHSMDEAHQRK
			GAGTTTCATTCTGCCACCCTAAACTGGAGGGGGATTTGGAGTGGTCAAT				INRYDRPDIRTELRRR
			CCGTAAAACGATTGATTGCGAAGGGTCTCCTCAGCAAATATGATAGTCA				FEAAGDIEFHSATLN
			TATCATTAGCGTCCAGGTTATGAGAGGCAGTCTTGGCTGTTTTAGACAG				WRGIWSGGQSVKRLI
			TTCATGTACCTTAGCGGGTTTTCTCGTGATTGGACGTAGCTTAAAACGTT				AKGLLSKYDSHIISVQ
			TGGTTCACATACATCTGCCTGCTTGGCACAATATCAAAAAGGCATA				VMRGSLGCFRQFMY
			AACATCGCACATATTGGTTATTTACGGCTATGAGGATGGTTTTAGTACGT				LSGFSRDWT (SEQ ID
			AGGCGTTGCGGAACTTCGGTTCGGATAGAGCAATGAATCGTGCATGCT				NO: 1489)
			AGGAACTGACCAAATAACAGCAGCCCTAGTATCTTTCGAAGATTTCCAT				
			ACCTTTGCGATCAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1121)				
R2	R2_KF	GU949558	GAGAGGTCACTGTTGCTGATCAGCTGGACACCTAGCTGACTGCGTGCAT	GAGAGGT	TAATGTC		MEVPLTSSLGSGATQ
		Kaloter mes flavicollis	GCGTGCACGCCGCGCCGCCTCCTCGTGGTACCGCTGGTAACGCAGG	CACTGTTG	CCTTTTG		APGTPELLGEHTVER
			GTCACTGAGACTACCTGCGGCTAAAGGCGCGCGCCGAGGGATTGTAAGT	CTGATCAG	GCTTGC		PGLDQGHSYGLLMD
			CCACCACCTCCACGTGGTTCCCCGCGGGCAACGGCATAGTTCATCTGCC	CTGGACAC	CCCCACC		DVELPVRLPFFGPLIC
			GGCCAAGCGTGCCGTTCCCTCGATCCTCCTCTGATTAGGGATAGGAGGG	CTAGCTGA	TGCTTAA		PGCRTLLTSEETISSH
			GGGCGGTGGCTCCCGCCGACGACCTGGCAAACACCTTGGACACGCTAA	CTGCGTGC	AGGAAC		HRRVHPDARTRWVC
			GATAATAGGCCGTCCTCCGGGCCGGCCGAATCATAAGCACACAATGGA	ATGCGTGC	TGGCAG		YGCDSPFMTYRAIKC
			AGTCCCTCTAACATCAAGCCTCGGATCTGGGGCAACCCAAGCTCCCGGG	ACGCCGCG	GAAAGA		HLPKCSGRKVVTGDH
			ACCCCAGAATTATTAGGGGAGCACACCGTTGAACGACCTGGCCTTGATC	CGCCGCCT	GAGTGA		ICNGCTKRFESQRGLS
			AAGGACATTCTTATGGTTTGTTAATGGATGACGTAGAGCTGCCGGTGAG	CCTCGTGG	TCCGTG		LHKRRAHPGLRNEE
			GTTACCATTTTTTGGGCCTCTCATATGCCCAGGATGCAGGACGCTGCTCA	TACCGCTG	CCATAG		MLEPPVRAERRPNA
			CCTCTGAGGAAACGATTAGCAGCCATCATAGGAGGGTCCACCCTGATGC	GTAACGCA	AAATAT		HKSSIWSIDEIRILEQY
			CCGGACGAGGTGGGTCTGCTATGGTTGTGATTCCCCATTTATGACATAC	GGGTCACT	GGTTAT		EAAYVGDLHINMKIA
			CGGGCCATTAAATGTCACTTACCTAAATGCTCTGGCAGGAAGGTAGTGA	GAGACTAC	CCGGGG		AHLPFKTNKQVSNYR
			CAGGGGATCACATCTGTAATGGCTGCACAAAGCGGTTTGAGTCTCAGA	CTGCGGCT	CAAGTC		NDRRKKSRTATDASQ
			GAGGTCTCTCACTGCATAAAAGGAGAGCCCATCCGGGTCTTCGGAACG	AAAGGCG	ACTAGC		QGLGPNDGNRGIVP
			AGGAAATGCTTGAGCCTCCTGTGAGAGCTGAGCGACGCCCTAACGCAC	CGCGCGA	AATATG		SGQSSPLFLEGSDAE
			ATAAATCCTCAATATGGAGCATTGATGAGATTCGTATCCTGGAGCAATA	GGGATTGT	GGACTT		GDEDVFNVLVPPTLG
			TGAGGCTGCCTACGTCGGCGATCTCCACATCAACATGAAGATAGCTGCT	AAGTCCAC	CTCCGG		GLEPAGQVHSLSEGE
			CACCTGCCGTTCAAAACCAATAAACAAGTGAGCAACTATCGGAACGATC	CACCTCCA	GTCCGT		TSPLVGEADPCFMG
			GCCGTAAAAAGTCCAGAACGGCGACTGACGCGAGCCAACAGGGACTTG	CGTGGTTC	GCGGTC		GTPSAGEASGSTLLG
			GTCCCAATGACGGCAATAGAGGAATTGTCCCCAGCGGGCAGTCATCCCC	CCCGCGGG	CTTCCAA		PDPTPADGYSLVRKD

		GTTATTTTTAGAGGGCTCTGACGCCGAGGGGGACGAGGATGTTTTTAAT	CAACGGCA	CATGAG	LQLSVQTSPLLAVGS
		GTCCTAGTGCCGCCCACCCTAGGGGGTCTTGAACCTGCTGGACAGGTTC	TAGTTCAT	CTGGAC	VGTESVQFERGVLSC
		ATTCACTCTCTGAGGGGGAAACTTCGCCGCTGGTCGGCGAAGCGGATCC	CTGCCGGC	GTAGTC	GTPPEFLHPEQFAHC
		CTGTTTTATGGGTGGCACGCCGTCTGCCGGAGAGGCCTCGGGCTCAACC	CAAGCGTG	CACTCTA	ANNDPVLNASEEQV
		TTGTTGGGTCCGGATCCCACCCCGGCAGATGGATATTCTCTGGTCAGGA	CCGTTCCC	TGACTT	HAPLGEEANDLPDN
		AAGACTTACAGCTCAGTGTGCAGACATCCCCGTTATTAGCAGTGGGCTC	TCGATCCT	GAACGA	NHPSELGVDPEDPTC
		TGTCGGCACTGAGTCTGTGCAATTTGAACGTGGGGTGTTATCGTGCGGT	CCTCTGAT	TACGGG	SPATEQVQPSSEEEA
		ACGCCACCAGAATTTTTACACCCGGAACAGTTTGCACATTGTGCAAACA	TAGGGATA	GGCGTA	DDPFAQFKAWRRRV
		ATGACCCGGTTTTGAACGCGTCAGAGGAGCAGGTGCACGCCCCACTAG	GGAGGGG	TCTCCCC	ASYALKIETGVLPAQV
		GGGAAGAAGCCAATGACCTCCCTGACAACAACCATCCGTCGGAGCTGG	GGCGGTG	CGGAAG	DDLLRRLRDGDTQSK
		GGGTTGACCCCGAAGACCCGACGTGTAGTCCCGCAACGGAACAGGTCC	GCTCCCGC	AGGTCG	VTCEEVEEVVLSLTRT
		AACCCTCTTCTGAGGAAGAAGCCGATGACCCGTTTGCGCAATTCAAGGC	CGACGACC	CCTAGG	ILGGTAPKKRVEGRT
		GTGGCGAAGAAGGGTGGCTTCCTATGCCTTGAAGATCGAGACCGGGGT	TGGCAAAC	CGACTT	KWTYKSRTNHEARK
		TCTTCCAGCCCAAGTAGATGACCTCCTTCGGAGACTGAGGGATGGGGAC	ACCTTGGA	AGAA	RIMYARCQDLYRRRP
		ACACAGTCTAAGGTAACGTGCGAGGAGGTGGAGGAAGTGGTTTTATCC	CACGCTAA	(SEQ ID	QRPVERAVGYQAEES
		CTCACTAGAACGATACTGGGTGGCACGGCACCCAAAAAGCGCGTTGAG	GATAATAG	NO:	LLDNQDERPSHGAFE
		GGTCGAACGAAGTGGACGTACAAAAGTCGAACCAACCATGAGGCTCGC	GCCGTCCT	1368)	TFYTGLWGKSGQCNI
		AAGCGTATAATGTACGCTAGATGCCAGGACTTATATCGCAGGCGGCCTC	CCGGGCCG		TMPPGVPRHTGHVL
		AGCGACCTGTGGAGAGGGCCGTCGGATACCAGGCGGAGGAATCTCTTC	GCCGAATC		REVTPKDIYSRLRKLK
		TTGACAACCAAGATGAGAGACCATCCCACGGAGCCTTTGAAACCTTCTA	ATAAGCAC		KDYAPGPDGVTKLKV
		TACTGGGCTCTGGGGTAAGTCCGGTCAATGCAATATCACAATGCCTCCT	ACA (SEQ		QSMGAYPSLLAKVY
		GGTGTGCCTCGGCACACCGGGCATGTACTGAGGGAAGTTACGCCAAAG	ID NO:		NLVMLTGYFSSCWK
		GACATCTACTCGAGGCTGAGGAAACTAAAGAAAGACTATGCACCCGGC	1245)		EHKTSLIPKDRGSPM
		CCGGATGGCGTAACTAAATTGAAGGTGCAGAGCATGGGAGCATATCCC			DVSNWRPITIGSLLSR
		TCACTTCTTGCAAAGGTATATAATCTGGTGATGCTAACCGGTTACTTCTC			IYTGLIERRLRTVSDIH
		CAGTTGTTGGAAGGAACATAAGACCTCTCTAATTCCTAAGGATAGAGGA			QRQVGFMPVNGCA
		AGCCCAATGGATGTCAGTAATTGGAGACCGATTACGATTGGCTCCTTAC			ANLFIFDECIRQAKKE
		TCAGTCGGATTTATACGGGTCTCATAGAACGGAGGCTGAGGACAGTTTC			GTIVGSLIDVAKAFDT
		TGACATCCACCAAAGGCAGGTAGGGTTTATGCCCGTCAATGGTTGTGCC			VPHEAILRALSSQGV
		GCCAACTTGTTTATATTTGATGAATGCATCCGACAAGCCAAAAAAGAGG			DEHTMAHIRDMYSG
		GGACAATTGTCGGCAGTCTTATAGACGTCGCCAAAGCATTTGATACCGT			IRTRINGKGSDIPLVR
		ACCCCATGAAGCCATCCTACGGGCGTTGAGCTCGCAAGGTGTAGATGA			GVKQGDPLSPMLFN
		GCACACGATGGCTCACATAAGAGACATGTATTCCGGTATCCGTACGCGG			MVMDPLIRDLQRKG
		ATCAACGGCAAAGGGAGCGATATCCCTCTGGTGAGAGGAGTGAAGCAA			FRIGGHEIGALAFAD
		GGTGATCCTCTATCCCCAATGCTCTTTAATATGGTGATGGACCCTCTCAT			DIVLLADSIDGAQDH
		ACGAGACCTACAGAGGAAGGGCTTCAGAATTGGCGGCCACGAGATAGG			VDQVGRYMNKLGM
		GGCGTTGGCCTTTGCGGATGACATCGTCTTGTTGGCGGACTCCATTGAT			TLNPRKSSSFLITAMR
		GGGGCCCAGGATCATGTAGATCAGGTTGGGCGCTACATGAATAAACTG			KTWICRDPGLSIGETK

R2	R2_RL	GU949555	Reticulitermes lucifugus	CTCACTGCTGTCATGCCATTGTTAATGCGTTGGGTGATGGCGGGTGATG	CTCACTGC	TGAGTT	MMADYNNSVDHAL
				GGAGACGAGTTACAGCAGAGCTGGCTTCACGGGGGGGGCTGTAGTTGC	TGTCATGC	AGATAT	EDNTRLIFARDAVLA
				CCGAACCAGTCTGCTTCTGAGTGCCTCCACGTGGCCTCGCTGGAAACGT	CATTGTTA	ATATGA	RVCGPFDNLECGLCG
				CGGAGCTGCTTACGGCTAACCCGTAGACAACTCCGGCGGCGGCTAAGACCTT	ATGCGTTG	CGACTG	VLLTSLQGVREHCHR
				AAAACGGCCTTACTACCAGGGTCATCCCTCGCGGCGGCGCTAAGACCTT	GGTGATG	ATTCTTC	AHHNLDLTFQCTKCD

GGCATGACCCTCAATCCACGAAAATCGTCCTCCTTCCTCATCACTGCGAT
GAGGAAGACCTGGATATGTAGGGACCCGGGCTTGTCAATAGGGGAGA
CTAAGGTGCCTGGCGCCAGACCCTCCTCTGCACTTAAATATCTCGGAGT
AAACTACACACTCTCGGAAGGCCTGGAGAGTGGTGCCTTGATAGACAA
GCTCATGCAGGCCGTTAACAGAGCACGAGGTTTGGCCCTTAAACCGCTG
CAAAGGTCAACCTCATACTTGAACGAATAATCCCCAAATTTTTGTATGG
TATCATCCTTGGCGGGCCTAGTTTGACTAGGCTGCATGCAGCCGACAAG
TGTGTCAGGATGGCTGTAAAGGAGATCCTGCACCTCCATCCTAGTACAA
CAGACCATGTACTATATGCCAGAAAAAAGGATGGGGGAATGGGGATTC
CCCGACTGGCTCACCTTGTCCGCCTAGCCAGTCTCAGGTCGGGGTTAGC
TCTGTTAGCGAGTGGGGACGTTGCCGTTCAAGCAGCGGGGATGGCAGG
CGACCTGGAAGGCCGCTGCAAGAAAGTCGCAAATGATTTGCGGCTTAA
TTGGCCGGTGACGCTGAGAGATGTGGTGAGAGCCAGCAATAAATTTAA
ATCTCAAGAATCAAAAGATTGGGAGCGTCTGGCCTCCCAGGGTCACGG
AGTCAAGGACTTCCGGAATGACCGTCTTGGGAATTGCTGGCTCTATGAT
CCGACGGTGTTGTCCAGCAGCAGATACACTGATGCTCTCAGGCTGAGGA
CTAACACGTTCGGCGTCAACGTTGCCCTAAGGAGGGCCGACAAGGACC
TTGAGGTCAACTGTCGGCGATGCCACGGGAAACCTGAAACCTTAGGGC
ACGTGCTCGGAGAGTGTGTGGCAGGTAAGGGTATGAGGATACAACGAC
ACGACAAGATGGCGGCCTTTGTGGCTACAAAATGTGAAGAAAAGGGGT
ACCAGACGACCCGCGAACAGCTATTCTCCATCGAGCAGGGAAAACTCAA
ACCTGACCTCGTCGTAATAGACGGGGAAAGGGCTCTCATTGTCGACGTA
ACAGTCCGCTTTGAGAGTGGGAACGCCCTCTCACGGGGAGCCTCTGAG
AAGATCGAAAAGTATCAGCCTTTGGCTGACTATTTTGTCTCTCAGGGGG
CTGTTCGAGAAGCCAACGTGTTACCGATTGTTGTCGGTTCCAGGGGAGC
AATCACCCAGGCCACATTGAAGTCCCTAGCTACGCTAGGGTTAGATGTG
GAACGTGTTGGCAAGTACCTCGCCATCTGTGCCGTTGCCTCCAGCGTTG
AGATCGCATGCATGCACCTGGATTACACATAATGTCCCTTTGGCTTGCC
CCCACCTGCTTAAAGGAACTGGCAGGAAAGAGAGTGATCCGTGCCATA
GAAATATGGTTATCCGGGGCAAGTCACTAGCAATATGGGACTTCTCCGG
GTCCGTGCGGTCCTTCCAACATGAGCTGGACGTAGTCCACTCTATGACTT
GAACGATACGGGGGGCGTATCTCCCCCGGAAGAGGTCGCCTAGGCGACT
TAGAA (SEQ ID NO: 1122)

VPGARPSSALKYLGV
NYTLSEGLESGALIDK
LMQAVNRARGLALK
PLQKVNLILERIIPKFL
YGIILGGPSLTRLHAA
DKCVRMAVKEILHLH
PSTTDHVLYARKKDG
GMGIPRLAHLVRLAS
LRSGLALLASGDVAV
QAAGMAGDLEGRCK
KVANDLRLNWPVTL
RDVVRASNKFKSQES
KDWERLASQGHGVK
DFRNDRLGNCWLYD
PTVLSSSRYTDALRLR
TNTFGVNVALRRADK
DLEVNCRRCHGKPET
LGHVLGECVAGKGM
RIQRHDKMAAFVAT
KCEEKGYQTTREQLF
SIEQGKLKPDLVVIDG
ERALIVDVTVRFESG
NALSRGASEKIEKYQP
LADYFVSQGAVREAN
VLPIVVGSRGAITQAT
LKSLATLGLDVERVGK
YLAICAVASSVEIACM
HLDYT (SEQ ID NO: 1490)

	ACTGTATGTACTACGGACTACAGTTGAGTAGCGCGGGAGAGCACTTGAT	GCGGGTG	CCTAACT	KGFSSYRGICCHFSKC
	GTGAGGGTAGCAACTTGTTGCTACTTCGACTGTCTCCCTGAAGATTTCCG	ATGGGAG	AAAATA	IGARISVSEGPLSCSEC
	AAGGGGTGCCGCAAGTCCATGGGGGCGGTCCTGCTGTGGGAGTTATCT	ACGAGTTA	TAATTTG	EREFDSKRALSTHER
	TGCACCCGGGAAGCCCTCAGTGGATAAATACTCCAAAATGGCTTAGCCA	CAGCAGA	GAATGT	HMHPGIRNAKRLKD
	CCTCACGTGGAGAGAGGGGAATTAGGGGTTTACTCCTAAGGGCCGATG	GCTGGCTT	GCATTA	FNPRGGGGKTIHGN
	CCTCCAATCGGATTTCCCCGGGCAACGGCTATTATCAGCCGGCTAAGTTC	CACGGGG	CTTATAC	TKWTEEEVQLLVSLS
	GGACCCTCATCTACTGATGGGATACTCTCTCTCCCCCCGCCTGCGTGATT	GGGGCTG	TTTATTG	KRFEGYKSINKEISLIL
	AGTTGGGAATGCACGCGGGCGGGTGATTCTGACAAGCCCAGCCCAAAA	TAGTTGCC	TATTTAT	TSKTCKQISDKRRYLN
	GTTCTTACCACGACGTACCCGGGGGTGTCACCCGAACTAATATATTTTCC	CGAACCAG	TGACAT	LHNGNGGLAAAEAV
	GCCGATCGGTTCGCTGTGGGGGGATGTCGGGGCCTGGTGGCCGGGAA	TCTGCTTC	CCGTAG	LEFCEDSHPEVTESD
	CCGTCAGCCGCTCTTTGAGTGTCCTGCGTAACCCGGAGGCGGTGACGTC	TGAGTGCC	TAGTCT	GAVLSEIMDEECHQS
	GAAGGGCTAGGCTAGCACGGTCCAGACCCCTGAACGCCTGGGGAGACG	TCCACGTG	GTTTAGT	SVTMRSSIVHGDIGR
	GGCCCACCACCTGAGTAGGAGTGGGTCCTTTACCTCATTTGAGGTGTCT	GCCTCGCT	GGATTT	EVQGKELVRIPPDNS
	CCTCGTTCTTGTAGGGGGCGAAGGACGAGGAATGCATCCGTCCCTCCAG	GGAAACGT	AGATCG	VMGNCVVLLRKLAT
	GATGCTGGTGGTTTCCGTCTGGTGGGCTCATGCTGAAACGCGCAGGTGC	CGGAGCTG	TTAACCA	DKRSDLDLSKDKELRI
	CGATTTGTCGAAGAGGACATATGGGGTAACCCATAGAACCTAGGCAGG	CTTACGGC	TTCTGTG	DIEKATKANRESADG
	AGTAATCCCTTAGCTGGGGGGGTCAGCTGGTGGGTCCTGTCAATTTATC	TAACCCGT	ACGTCG	RVIQSVAGNEIDPDT
	CTCCCCTCCATGCCAGAGCCGGTCCGAGGTTAGAGACGGACTCATTTTC	AGACAACT	CTGATG	FQWKELLLGQVRGF
	TCCTTTTTATATGTCATGATGGCCGATTATAATAATTCTGTTGATCATGCA	CCGGCGGC	TCTATAT	PRVDENSELFDLDDK
	CTTGAGGATAATACTAGATTGATATTTGCGCGTGACGCAGTACTTGCGA	CAAACTCA	GTACCA	LTKELSSDSPVWNDN
	GAGTATGTGGACCTTTTGATAACTTGGAGTGCGGGTTGTGTGGGGTGCT	GAAAACG	GTGGCA	CELIVSDLCQVLCKKK
	CCTGACGTCCCTTCAGGGCGTTCGAGAGCATTGCCACCGGGCGCACCAT	GCCTTACT	GAATGC	YELGRQHHVRKGKR
	AACTTGGATTTGACTTTTCAATGTACTAAATGCGATAAAGGGTTTTCCTC	ACCAGGGT	TGACTA	HRGIHHKREKFRECQ
	GTATAGGGGAATTTGTTGTCATTTTTCAAAGTGTATAGGAGCTCGTATAT	CATCCCTC	AGAACA	KIFRKSPRKLAEYLYR
	CTGTTTCAGAGGGCCCCCTTAGTTGTTCAGAATGTGAAAGGGAATTTGA	GCGGCGG	AAATAA	DKDLSHISKDASTPQ
	CTCTAAAAGAGCTCTTTCAACCCACGAGAGGCATATGCACCCAGGGATT	CGCTAAGA	TTTTAAA	GIEQYYSQLWGEPEL
	AGGAATGCTAAGAGGCTCAAAGATTTTAATCCGCGGGGGGGGGGTGGT	CCTTACTG	CAATAG	LESNTIEEKLPSSSLFD
	AAAACTATTCACGGCAATACTAAGTGGACTGAAGAGGAGGTCCAGTTG	TATGTACT	ACAGCA	CLPPITPEEVEGRIHKI
	CTTGTATCACTTAGCAAGAGATTTGAGGGGTATAAGTCTATCAATAAAG	ACGGACTA	CTAAGA	RPSSAPGLDGVRKIH
	AAATTAGCCTTATTTTAACATCAAAAACTTGTAAACAAATTTCTGATAAA	CAGTTGAG	AATTCTG	LVGKGITLVLVKLYNL
	CGTCGCTATTTAAATTTGCATAATGGAAATGGGGGGGCTAGCTGCAGCGG	TAGCGCGG	CAAACT	LFLTGGYPECWKRNR
	AAGCTGTATTAGAATTTTTGTGAAGATAGTCACCCTGAAGTTACTGAGAG	GAGAGCA	GGAACG	TVFIPKIGKDLSEVGG
	CGATGGGGCGGTGCTTTCTGAAATCATGGATGAGGAATGTCATCAGTCA	CTTGATGT	TGGCCC	WRPLTIGSLLARMYS
	TCGGTAACTATGAGGTCAAGCATCGTACATGGCGACATTGGTCGTGAG	GAGGGTA	ACGGTA	AFLERRIRRVTSLSLS
	GTTCAAGGGAAGGAATTGGTCCGGATCCCGCCTGATAATAGTGTAATG	GCAACTTG	GGCCAA	QRGFTNIQGCHVNLT
	GGGAATTGTGTTGTTCTACTTAGGAAACTTGCTACTGACAAAAGAAGTG	TTGCTACT	TATACC	ILKEGIRQAKVKNGG
	ATTTGGACTTGAGTAAGGATAAAGAGTTGCGCATTGACATTGAGAAGG	TCGACTGT	GGAAAG	VIVSVDIEKAFDTIPHS
	CGACGAAAGCTAACAGAGAAAGCGCAGACGGTAGGGTAATTCAATCTG	CTCCCTGA	GGAAAA	VIFSRLASQGVPPLLR

		TGGCTGGTAATGAGATCGATCCCGATACTTTTCAATGGAAAGAATTGCT	AGATTTCC	TGACAC	KIISNMYKDVYTVIEG
		GCTTGGGCAGGTACGGGGGTTTCCACGAGTTGATGAGAACTCTGAATT	GAAGGGG	ATCCCCC	QCIPIKRGVKQGDPL
		ATTTGATTTGGATGACAAATTGACTAAGGAATTATCTTCCGACAGCCCCG	TGCCGCAA	CTTAAA	SPLLFNIAIDPVLRSLE
		TGTGGAATGACAACTGCGAACTTATTGTGTCGGACTTATGTCAGGTCTT	GTCCATGG	(SEQ ID	EFQGGLPLGNSAIKIL
		ATGCAAAAAGAAATATGAGTTAGGTAGGCAGCATCACGTTAGGGAAAGG	GGGCGGT	NO:	AFADDIILGASSAGQ
		AAAGAGGCATCGGGGTATACATCATAAGAGGGAGAAATTTAGGGAATG	CCTGCTGT	1369)	AQQMVDMLGIGLTS
		CCAGAAGATATTTAGAAAGTCCCCCAGAAAGCTTGCCGAATATCTCTAC	GGGAGTTA		CGLGVSHRKCFGFQI
		CGTGATAAGGACCTGTCTCATATCAGTAAAGACGCTAGCACTCCGCAGG	TCTTGCAC		VNKNKTWTIVDPMI
		GTATAGAGCAATACTATTCTCAGTTATGGGGGGAACCTGAATTACTAGA	CCGGGAA		TLNGSSLPFSGPEDRL
		ATCCAACACAATTGAAGAAAAGTTACCGTCCAGTTCCTTATTTGACTGTC	GCCCTCAG		PYLGVDINPWDRKSR
		TTCCCCCCATCACGCCGGAGGAGGTTGAGGGAAGAATCCATAAGATTC	TGGATAAA		YDAGQRLISAAKRGS
		GACCGTCTTCCGCCCCAGGGTTGGATGGGGTACGCAAGATCCATCTTGT	TACTCCAA		QLSLKPQQKINLITAF
		CGGTAAAGGGATAACCCTGGTCCTTGTTAAACTCTACAACTTGCTATTTT	AATGGCTT		LLPKFLYILIEDPPSPA
		TGACTGGGGGATATCCTGAGTGTTGGAAGCGGAACAGGACCGTCTTTA	AGCCACCT		YLKSIDHDLRQIYKNIL
		TACCGAAGATTGGGAAAGATCTGAGTGAGGTGGGAGGATGGAGACCG	CACGTGGA		HLPNCVSTAFMYSPK
		TTGACAATTGGTTCCCTATTAGCTCGGATGTATTCAGCCTTCCTAGAGAG	GAGAGGG		RDGGLGLPRLSCLVPL
		GAGGATTAGGAGGGTGACCTCCTTGTCCTTGTCGCAGAGGGGCTTTACG	GAATTAGG		AHLKAGIKLGSLQDSL
		AATATCCAAGGTTGTCACGTTAATTTGACTATCCTCAAAGAAGGCATTCG	GGTTTACT		VREITTSDRFVRTMG
		ACAGGCCAAAGTCAAGAACGGGGGGAGTTATAGTCTCAGTAGACATAGA	CCTAAGGG		SVAHSLWASWSLTL
		GAAGGCTTTTGACACAATCCCTCATTCTGTAATTTTCTCCCGCTTGGCGTC	CCGATGCC		QDIYKLKSALKRREAK
		TCAAGGTGTTCCTCCCCTTCTTAGGAAAATTATTTCCAATATGTATAAGG	TCCAATCG		AWESCVSQGQGAA
		ATGTCTACACTGTGATTGAGGGGCAGTGTATTCCAATAAAGAGAGGAG	GATTTCCC		QFRGDSIGNNWLHN
		TTAAACAAGGGGATCCTCTTTCTCCACTACTTTTCAATATTGCCATCGATC	CGGGCAAC		PGTYRPGQYIEALKLR
		CTGTGTTGCGCAGTTTGGAGGAGTTTCAGGGGGGTCTCCCACTTGGTAA	GGCTATTA		ANLTGVRVNLKRSGY
		CTCGGCTATTAAGATCTTGGCCTTTGCTGATGATATTATATTAGGAGCTA	TCAGCCGG		NVPITCRFCKDIPETQ
		GTTCGGCGGGTCAGGCGCAGCAGATGGTGGATATGTTGGGAATAGGAC	CTAAGTTC		AHVLGLCPKTKGMRI
		TTACTTCGTGTGGGTTGGGAGTATCCCATAGAAAATGTTTTGGATTTCAA	GGACCCTC		QRHDSIVNRVRDKLK
		ATCGTGAATAAAAACAAGACCTGGACAATTGTTGACCCGATGATCACAC	ATCTACTG		TKSPVALMHEQNFT
		TCAATGGTTCTTCATTACCCTTCTCAGGGCCAGAGGACAGGTTGCCTTAT	ATGGGATA		VEEGQVFKPDIVTILG
		CTCGGGGTGGATATAAACCCATGGGATAGGAAATCACGGTATGACGCT	CTCTCTCTC		EVGYVIDVTVRYDDR
		GGACAGAGACTTATTTCGGCGGCCAAACGAGGCTCTCAACTATCCCTTA	CCCCCGCC		DYIKDASVEKIRKYEA
		AGCCTCAGCAAAAGATAAATTTAATAACAGCTTTTCTTTTGCCAAAGTTT	TGCGTGAT		LKGYLKDLYPQLNKV
		CTGTACATTCTGATTGAAGATCCGCCTTCGCCTGCCTATCTGAAATCCAT	TAGTTGGG		EVLPLVFGSRGAVPG
		CGATCATGACCTGAGGCAGATATATAAGAATATTTTGCACCTCCCTAACT	AATGCACG		STVHNMGLLGFTKRE
		GTGTGTCAACCGCCTTTATGTACTCTCCCAAGAGGGATGGTGGTTTGGG	CGGGCGG		MVHISRKVIADSLIIS
		GCTACCGCGTCTGTCCTGTCTAGTTCCACTTGCCCATTTAAAAGCCGGGA	GTGATTCT		NFLEVY (SEQ ID NO:
		TCAAGCTCGGTTCCCTCCAAGATTCTTTGGTTAGGGAAATTACCACCTCA	GACAAGCC		1491)
		GACAGGTTTGTACGTACAATGGGTTCGGTAGCTCACTCCCTTTGGGCAT	CAGCCCAA		

			CATGGTCTCTAACGTTGCAGGATATCTATAAACTTAAGTCCGCACTGAAA	AAGTTCTT	
			CGTAGGGAAGCTAAGGCATGGGAGAGTTGTGTGTCTCAGGGACAAGG	ACCACGAC	
			AGCAGCGCAATTTCGCGGGGACTCAATTGGAAACAACTGGCTCCATAAT	GTACCCGG	
			CCGGGAACCTATCGACCTGGACAGTATATCGAAGCACTTAAGCTTCGTG	GGGTGTCA	
			CTAATTTAACAGGGGTTAGAGTCAACTTGAAGCGTTCAGGCTATAATGT	CCCGAACT	
			TCCAATAACTTGTAGGTTCTGTAAGGATATACCTGAAACGCAGGCTCAT	AATATATT	
			GTTCTTGGACTTTGTCCAAAGACCAAGGGAATGAGAATACAAAGGCAC	TTCCGCCG	
			GATTCAATTGTGAATAGGGTGCGAGATAAACTTAAGACTAAGTCGCCCG	ATCGGTTC	
			TTGCTCTCATGCATGAGCAAAACTTCACCGTTGAGGAAGGCCAGGTGTT	GCTGTGG	
			CAAACCAGATATAGTGACGATTTTGGGAGAGGTGGGTTACGTAATTGAT	GGGGATG	
			GTGACAGTACGGTATGACGATAGAGATTACATCAAGGATGCCTCTGTG	TCGGGGCC	
			GAAAAAATCAGGAAATACGAGGCCCTTAAAGGGTACCTGAAGGATTTA	TGGTGGCC	
			TATCCACAGCTGAATAAGGTAGAGGTTTTGCCTCTTGTCTTTGGTAGCAG	GGGAACC	
			AGGGGCCGTGCCAGGTAGTACTGTTCACAACATGGGATTGTTAGGTTTT	GTCAGCCG	
			ACGAAACGTGAAATGGTCCACATCTCTCGCAAAGTTATAGCCGACTCAC	CTCTTTGA	
			TGATAATCTCTAATTTTTTAGAGGTTTATTGAGTTAGATATATATGACGA	GTGTCCTG	
			CTGATTCTTCCCTAACTAAAATATAATTTGGAATGTGCATTACTTATACTT	CGTAACCC	
			TATTGTATTTATTGACATCCGTAGTAGTCTGTTTAGTGGATTTAGATCGT	GGAGGCG	
			TAACCATTCTGTGACGTCGCTGATGTCTATATGTACCAGTGGCAGAATG	GTGACGTC	
			CTGACTAAGAACAAAATAATTTTAAACAATAGACAGCACTAAGAAATTC	GAAGGGC	
			TGCAAACTGGAACGTGGCCCACGGTAGGCCAATATACCGGAAAGGGAA	TAGGCTAG	
			AATGACACATCCCCCCTTAAA (SEQ ID NO: 1123)	CACGGTCC	
				AGACCCCT	
				GAACGCCT	
				GGGGAGA	
				CGGGCCCA	
				CCACCTGA	
				GTAGGAGT	
				GGGTCCTT	
				TACCTCAT	
				TTGAGGTG	
				TCTCCTCG	
				TTCTTGTA	
				GGGGGCG	
				AAGGACG	
				AGGAATGC	
				ATCCGTCC	
				CTCCAGGA	

					TGCTGGTG GTTTCCGT CTGGTGG GCTCATGC TGAAACGC GCAGGTGC CGATTTGT CGAAGAG GACATATG GGGTAACC CATAGAAC CTAGGCAG GAGTAATC CCTTAGCT GGGGGGG TCAGCTGG TGGGTCCT GTCAATTT ATCCTCCC CTCCATGC CAGAGCCG GTCCGAGG TTAGAGAC GGACTCAT TTTCTCCTT TTTATATG TC (SEQ ID NO: 1246)		
R2	R2_RU	GU949 554	Reticulit ermes urbis	CACACTGCTGTCATGCCATTGTTAATTCGTTGGGTGATGGCGGGTGATG GGAGACGAGTTACAGCAGAGCTGGCTTCACGGGGGGGCTGTAGTGGC CCGAACCAGTCTGCTTCTGAGTGCCCCCAAGTGGCCTCGCTGGAAACGT CGGAGCTGCTTACGGTTAACCCGTAGACAACTCCGGCGGCCAAACTCAG AAAACGGCCTTACTACCAGGGCCATCCCTTGCGGTTGCGCTAAGACCTT ACTGTATGTACTACGGACTACAGTTGAGTAGCGCGGGAGAGCACTCGA TGTGAGGGTAGCAACTTGTTGCTTCTTCGACTGTCTCCCTGAAGAGTTCC GAAGAGGTGCCGCAAGTCCATGGGGGCGACCCGGCTGTGGGAGTTATC CTGCACCCGGGAAGCCCTCAGTGGATAAATACTCCAAAATGGCTTAGCC ACCTCACGTGGAGAGAGGGGAATTAGGGGTTTACTCCTAAGGGCCGAT	CACACTGC TGTCATGC CATTGTTA ATTCGTTG GGTGATG GCGGGTG ATGGGAG ACGAGTTA CAGCAGA GCTGGCTT	TGAGTT AGATAT ATATAA CGACTG ATTCTTC CCTAACT AAAATA TAATTTG GAATGT GCATTA	MMADYNNSVDHAL EDDTRFIFARDSVLAR VCGHFDNLKCELCGV LLTSLQGVREHCHRS HHNLDLTFQCTKCDK GFSSYRGICCHFSKCR GARISVSEGPLSCSEC ERKFDSKRALSTHER HMHPGIRNAKRLKD FNPRGGGKTIHGNTK

	GCCTCCAATCGGATTTCCCAGGGCAACGGCTATTATCAGCCGGCTAAGT	CACGGGG	CTTATAC		WTEEEVQLLVSLSKR
	TCGGACCCTCATCTACCGATGGGATACTCTCTCTCCCCCCGCCTGCGTGA	GGGCTGTA	TTTATTG		FEGYKSINKEISLILTSK
	TTAATTGGGAATGCACGAGGGCGGGTGATTCTGACAAGCCCAGCCCAA	GTGGCCCG	TATTTAT		TCKQISDKRRYLNLH
	AAGTTCTTACCACGACGTACCCGGGGGGTGTCACCCGAACTAATATATTTT	AACCAGTC	TGACAT		NGNGGLAAAEAVLV
	CCACCGATCGGTTCGCTGTGGGGGGATGTCGGGGCCTGGTGGCCAGGA	TGCTTCTG	CCGTAG		FCDDSHLEVTDSDGA
	ACCGTCAGCCGCTCTTTGAGTGTCCTGCGTAACCCGGAGGCGGTGACGT	AGTGCCCC	TAGTCT		VLSEIMDEEYYQSSLT
	CGAAGGGCTAGGCTAGCACGGTCCAGACCCCTGAACGCCTGGGGAGAC	CAAGTGGC	GTTTAGT		MRSSIVHGDIGREVQ
	GGGCCCACCACCTGTGTAGGAGTGGGTCCTTTACCTCATTTGAGGTGTC	CTCGCTGG	GGATTT		GKDLVRIPPDNSVM
	TCCTCGGTCTAGTAGGGGGCGAAGGACGAGGGATGCATCCGTCCCTCC	AAACGTCG	AGATCG		GNCVVLLRKLATEKR
	AGGATGCTGGTGGTTTCCGTCTGGTGGGCTCATGCTGAAACCCGCAGGT	GAGCTGCT	TTAACCA		SDLDLSKDKELRIDIEK
	GCCGATTTGTCGAAGAGGACATATGGGGTAACCCATAGAACCTAGGCA	TACGGTTA	TTCTGTG		ATKANRESADGRVIQ
	GGAGTAATCCCTTAGCTGGGGGGGTCAGCTGGTGGGTCCTGTCAATTTA	ACCCGTAG	ACGTCG		SVADNEIDPDTFQW
	TCCTCCCCTCCATGCCAGAGCCGGTCCGAGGTTAGAGACGGACTCATTT	ACAACTCC	CTGATG		KELLLGQVRGFPRVD
	TCTCCTTTTTATATGTCATGATGGCCGATTATAATAATTCTGTTGATCATG	GGCGGCC	TCTACAT		ENSELFDLDDKLTSEL
	CACTTGAGGATGATACTAGATTTATATTTGCGCGTGATTCAGTACTTGCG	AAACTCAG	GTACCA		SSDSPVWNDNCELIV
	AGAGTATGTGGACATTTCGATAACTTGAAGTGTGAGTTATGTGGAGTGC	AAAACGGC	GTGGCA		SDLCHVLCKNKYELG
	TCCTTACGTCCCTTCAGGGCGTTCGGGAGCATTGCCACCGGTCGCACCA	CTTACTAC	GAATGC		RQHHVRKGKRHRGI
	TAACTTGGATTTGACTTTTCAATGTACTAAATGCGATAAAGGGTTTTCCT	CAGGGCCA	TGACTA		HHKREKFRECQKIFR
	CGTATAGGGGAATTTGTTGTCATTTTTCGAAGTGTAGAGGAGCTCGTAT	TCCCTTGC	AGAACA		KSPRKLAEYLYRDKDL
	ATCTGTTTCAGAGGGCCCCCTTAGTTGTTCAGAATGTGAGAGGAAATTT	GGTTGCGC	AAATAA		SHISKDVSTPQGIEQY
	GACTCTAAAAGAGCTCTTTCAACCCACGAGAGGCATATGCACCCAGGGA	TAAGACCT	TTTTAAA		YSQLWGKPELLESNT
	TTAGGAATGCTAAGAGGCTCAAAGATTTTAATCCGCGGGGGGGTGGTA	TACTGTAT	CAATAG		TEEMLPSSSLFDCLPP
	AAACTATTCACGGCAATACTAAGTGGACCGAAGAGGAGGTCCAATTGCT	GTACTACG	ACAGCA		ITPEEVEGRIHKIRPSS
	TGTATCACTTAGCAAGAGATTTGAAGGGTATAAGTCTATCAATAAAGAA	GACTACAG	CTAAGA		APGLDGVGKIHLVGK
	ATTAGCCTTATTTTAACATCAAAAACTTGTAAACAAATTTCTGATAAACG	TTGAGTAG	AATTCTG		GITLVLAKLYNLLFLTG
	TCGCTATTTAAATTTGCATAAATGGAAATGGGGGGGCTAGCGGCAGCGGA	CGCGGGA	CAAACT		GYPECWKRNRTVFIP
	AGCTGTATTAGTATTTTGTGATGATAGTCACCTTGAAGTTACTGACAGCG	GAGCACTC	GGAACG		KIGKDLSEVGGWRPL
	ATGGGGCGGTGCTTTCTGAAATCATGGATGAGGAATATTATCAGTCATC	GATGTGAG	TGGCCC		TIGSLLARMYSAFLER
	GCTAACTATGAGGTCAAGCATCGTACATGGCGACATTGGTCGTGAGGTT	GGTAGCAA	ACGGTA		RIRRVTSLSSSQRGFT
	CAAGGGAAGGATTTGGTCCGGATCCCGCCTGATAATAGTGTAATGGGG	CTTGTTGC	GGCCAA		NIQGCHVNLTILKEGI
	AATTGTGTTGTTCTACTTAGAAAACTTGCTACTGAAAAAAGAAGTGATTT	TTCTTCGA	TATACC		RQAKVKNGGVIVSV
	GGACTTGAGTAAGGATAAAGAGTTGCGCATTGACATTGAGAAGGCGAC	CTGTCTCC	GGAAAG		DIEKAFDTIPHSVIFSR
	GAAAGCTAACAGAGAAAGCGCAGATGGTAGGGTAATTCAATCTGTGGC	CTGAAGAG	GGAAAA		LASQGVPPLLRKIISN
	TGATAATGAGATCGATCCCGATACTTTTCAATGGAAAGAATTGCTTCTTG	TTCCGAAG	TGACAA		MYKDVYTVIEGQCIPI
	GGCAGGTACGGGGGGTTTCCACGAGTTGATGAGAACTCTGAATTATTTGA	AGGTGCCG	ATCCCCC		KRGVKQGDPLFPLLF
	TTTGGATGACAAATTGACTAGTGAATTATCTCCGACAGCCCCGTGTGG	CAAGTCCA	CTTAAA		NIAIDPVLRSLEEFQG
	AATGACAACTGCGAACTTATTGTGTCGGACTTATGTCACGTCTTATGCAA	TGGGGGC	(SEQ ID		GLPLGNSAIKILAFDD
	AAATAAATATGAGTTAGGTAGGCAGCATCACGTTAGGAAAGGAAAGAG	GACCCGGC			DIILGASSAGQAQQ

		GCATCGGGGTATACATCATAAGAGGGAGAAATTTAGGGAATGCCAGAA	TGTGGGA	NO: 1370)	MVDMLGIGLTSCGL
		GATATTTAGAAAGTCCCCCAGAAAGCTTGCCGAATATCTCTACCGCGAT	GTTATCCT		GVSHRKCFGFQIVNK
		AAGGACCTGTCTCATATCAGTAAAGACGTTAGCACTCCGCAGGGTATAG	GCACCCGG		NKTWAIVDPMITLN
		AGCAATACTATTCTCAGTTATGGGGGGAAACCTGAATTACTAGAATCCAA	GAAGCCCT		GSSLPFSGPEDRLPYL
		CACAACTGAAGAAATGTTACCTTCCAGTTCCTTATTTGACTGTCTTCCTCC	CAGTGGAT		GVDTNPWDRKSRYD
		CATCACGCCGGAGGAGGTTGAGGGGAGAATCCATAAGATTCGACCGTC	AAATACTC		AGQRLISAAKRGSQL
		TTCCGCCCCAGGGTTGGATGGGGTAGGCAAGATCCATCTTGTCGGTAAA	CAAAATGG		SLKPQQKINLITTFLLP
		GGGATAACCCTGGTCCTTGCTAAACTCTACAACTTGCTATTTTTGACTGG	CTTAGCCA		KFLYILIEDPPSPAYLK
		GGGATATCCTGAGTGTTGGAAGCGGAACAGGACCGTCTTCATACCGAA	CCTCACGT		SIDHDLRQIYKNILHL
		GATCGGGAAAGATCTGAGTGAGGTGGGAGGATGGAGACCGCTGACAA	GGAGAGA		PNCVSTAFMYSPKRD
		TTGGCTCCCTATTAGCTCGGATGTATTCAGCCTTCCTAGAGAGGAGGATT	GGGGAATT		GGLGLPRLSCLVPLA
		AGAAGGGTGACTTCCTTGTCCTCGTCACAGAGGGGCTTTACGAATATTC	AGGGGTTT		HIKAGIKLGSLQDSLV
		AAGGTTGTCACGTTAATTTGACTATCCTCAAAGAAGGCATTCGACAGGC	ACTCCTAA		REITTSDRFVRTMGS
		CAAAGTTAAGAACGGGGGAGTTATAGTTTCAGTAGACATAGAGAAGGC	GGGCCGAT		VSHSIGASWPLTLQD
		TTTTGACACAATCCCCCATTCTGTAATTTTTTCCCGCTTGGCGTCTCAAGG	GCCTCCAA		IYKLKSALKRREAKA
		TGTTCCTCCCCTTCTTAGGAAAATTATTTCCAATATGTATAAGGATGTTTA	TCGGATTT		WESCVSQGQGAAQF
		CACCGTGATTGAGGGACAGTGTATTCCAATAAAGAGAGGAGTTAAACA	CCCAGGGC		RGDSIGNNWLHNPG
		AGGGGATCCTCTTTTTCCACTACTTTTCAATATTGCCATCGATCCTGTGTT	AACGGCTA		TFRPGQYIEALKLRAN
		GCGCAGTTTGGAGGAGTTTCAGGGGGGTCTCCCACTTGGTAACTCGGCT	TTATCAGC		STGVRVNLKRSGYNV
		ATTAAGATCTTGGCCTTTGATGATGATATTATATTAGGAGCTAGTTCGGC	CGGCTAAG		PITCRFCKDIPETQAH
		GGGTCAGGCGCAGCAGATGGTGGATATGTTGGGAATAGGACTTACTTC	TTCGGACC		VLGLCPKTKGMRILR
		GTGTGGGTTGGGAGTATCCCATAGAAAATGTTTTGGATTTCAAATCGTG	CTCATCTA		HDSIVNRVRDKLKTK
		AATAAAAACAAGACCTGGGCAATTGTTGACCCAATGATCACACTCAATG	CCGATGGG		SPVALMHEQNFTVE
		GTTCTTCATTACCCTTCTCAGGGCCAGAAGACAGGCTGCCTTATCTCGGG	ATACTCTC		EGQVFKPDIVTILGEV
		GTGGATACAAACCCATGGGATAGGAAATCACGGTATGACGCTGGACAG	TCTCCCCC		GYVIDVTVRYEDRDYI
		AGACTTATTTCGGCTGCCAAACGAGGCTCTCAACTATCCCTTAAGCCTCA	CGCCTGCG		KDASVEKIRKYEALKG
		GCAAAAGATAAATTTAATAACAACTTTTCTTTTGCCGAAGTTTTTGTACA	TGATTAAT		YLKDLYPQLNKVEVLP
		TTCTGATTGAAGATCCGCCTTCGCCTGCCTATCTGAAATCCATCGATCAT	TGGGAATG		LVFGSRGAVPGSTVH
		GACCTGAGGCAGATATATATAAGAATATTTTGCACCTCCCTAACTGCGTGTC	CACGAGG		NMGLLGFTKREMVH
		AACCGCCTTTATGTACTCTCCCAAGAGGGATGGTGGCTTGGGGCTACCG	GCGGGTG		ISRKVITDSLIIISNFLE
		CGTCTGTCCTGTCTTGTTCCTCTTGCCCATATAAAAGCCGGGATCAAGCT	ATTCTGAC		VY (SEQ ID NO:
		TGGTTCCCTCCAAGATTCTTTGGTTAGGGAAATTACCACCTCAGACAGGT	AAGCCCAG		1492)
		TTGTACGTACGATGGGTTCGGTATCTCACTCCATTGGGGCATCATGGCCT	CCCAAAAG		
		CTAACGTTGCAGGATATCTATAAACTTAAGTCCGCACTGAAACGTAGGG	TTCTTACC		
		AAGCTAAGGCATGGGAGAGTTGTGTGTCTCAGGGACAAGGAGCAGCG	ACGACGTA		
		CAATTTCGAGGGGACTCAATTGGAAACAACTGGCTCCATAATCCGGGAA	CCCGGGG		
		CCTTTCGACCTGGACAGTATATCGAAGCACTTAAGCTTCGTGCTAATTCA	GTGTCACC		
		ACAGGGGTTAGAGTCAACTTGAAGCGTTCAGGCTATAATGTTCCAATAA	CGAACTAA		

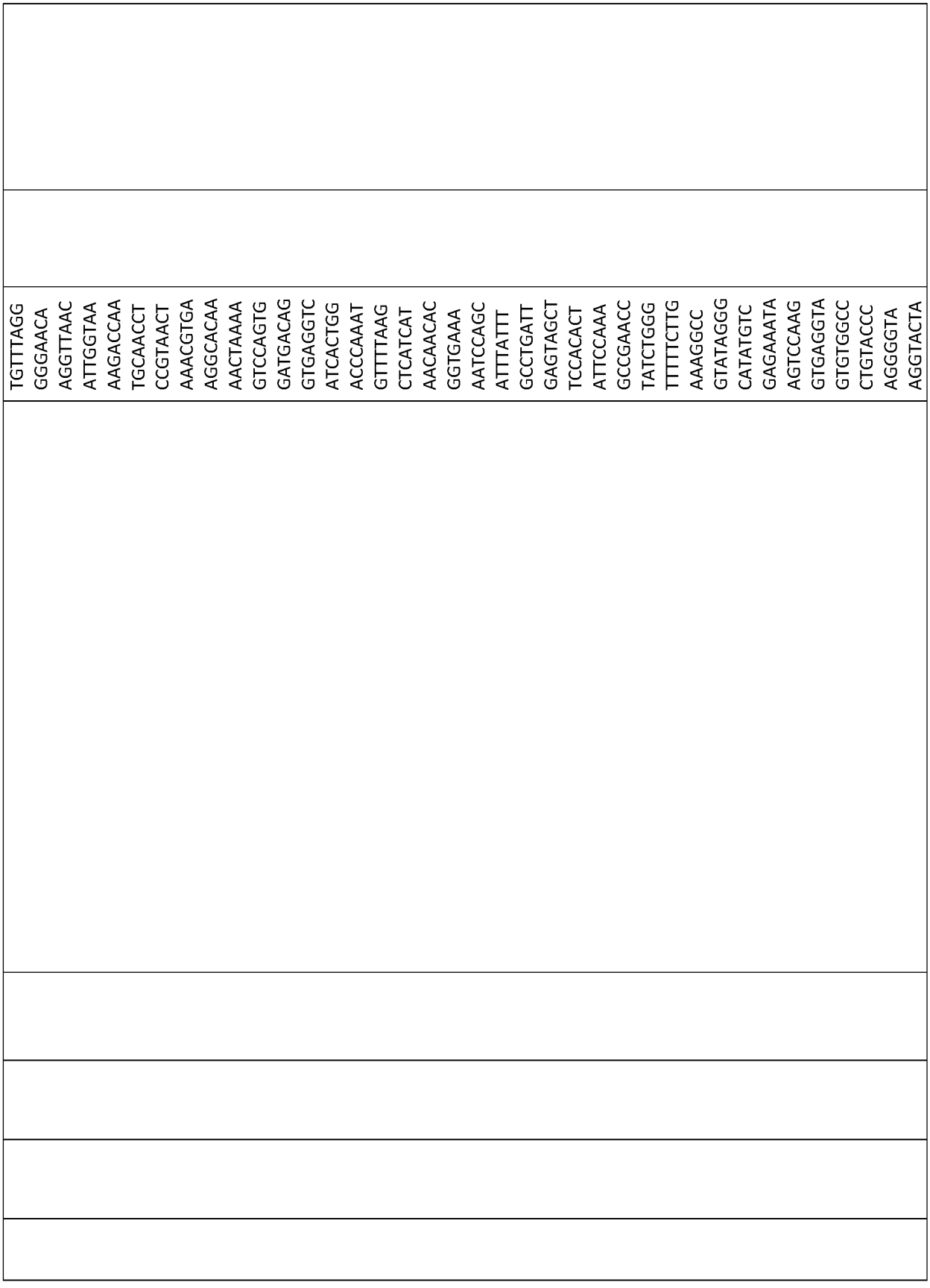
			CTTGTAGGTTCTGTAAGGATATACCTGAAACGCAGGCTCATGTTCTTGG	TATATTTTC	
			ACTTTGTCCAAAGACCAAGGGAATGAGAATACTAAGGCACGACTCAATT	CACCGATC	
			GTGAATAGGGTGCGAGATAAACTTAAGACTAAGTCGCCCGTTGCTCTCA	GGTTCGCT	
			TGCATGAGCAAAATTTCACTGTTGAGGAAGGCCAGGTGTTCAAACCAGA	GTGGGGG	
			TATAGTGACGATTTTGGGAGAGGTGGGTTACGTAATTGATGTGACGGT	GATGTCGG	
			ACGGTATGAAGATAGAGATTACATCAAGGATGCCTCTGTGGAAAAAATC	GGCCTGGT	
			AGGAAATACGAGGCCCTTAAAGGGTACTTGAAGGATCTGTATCCACAGC	GGCCAGG	
			TGAATAAGGTAGAGGTATTGCCTCTTGTCTTTGGTAGCAGAGGGCCGT	AACCGTCA	
			GCCAGGTAGTACTGTTCACAACATGGGATTGTTAGGTTTTACGAAACGT	GCCGCTCT	
			GAAATGGTCCACATCTCTCGCAAAGTTATTACCGACTCACTGATAATAAT	TTGAGTGT	
			CTCTAATTTTTTAGAGGTTTATTGAGTTAGATATATATAACGACTGATTCT	CCTGCGTA	
			TCCCTAACTAAAATATAATTTGGAATGTGCATTACTTATACTTTATTGTAT	ACCCGGAG	
			TTATTGACATCCGTAGTAGTCTGTTTAGTGGATTTAGATCGTTAACCATT	GCGGTGAC	
			CTGTGACGTCGCTGATGTCTACATGTACCAGTGGCAGAATGCTGACTAA	GTCGAAG	
			GAACAAAATAATTTTAAACAATAGACAGCACTAAGAAATTCTGCAAACT	GGCTAGGC	
			GGAACGTGGCCCACGGTAGGCCAATATACCGGAAAGGGAAAATGACAA	TAGCACGG	
			ATCCCCCCTTAAA (SEQ ID NO: 1124)	TCCAGACC	
				CCTGAACG	
				CCTGGGGA	
				GACGGGC	
				CCACCACC	
				TGTGTAGG	
				AGTGGGTC	
				CTTTACCT	
				CATTTGAG	
				GTGTCTCC	
				TCGGTCTA	
				GTAGGGG	
				GCGAAGG	
				ACGAGGG	
				ATGCATCC	
				GTCCCTCC	
				AGGATGCT	
				GGTGGTTT	
				CCGTCTGG	
				TGGGCTCA	
				TGCTGAAA	
				CCCGCAGG	

				TGCCGATT TGTCGAAG AGGACATA TGGGGTAA CCCATAGA ACCTAGGC AGGAGTA ATCCCTTA GCTGGGG GGGTCAGC TGGTGGGT CCTGTCAA TTTATCCTC CCCTCCAT GCCAGAGC CGGTCCGA GGTTAGAG ACGGACTC ATTTTCTCC TTTTTATAT GTC (SEQ ID NO: 1247)	TAGAAA TGTGTG CGATAA GGTGTG AATAGA AGGGTT CACCAA GAGGGA GACCTA GTTTGG ACCTCA GAAATG GGGTCA TAGGAG TGATAG	MSNYNETNTSGGDN PRMATQTTGSLSSGP INQHTCELCCRTFGT RAGLGQHVRKTHPIE SNQSINVERKKRRWS PEEIRRMANMEAQA TINNIKHLTQYLATYL PQRTLNAIKGRRRDA EYKELVTGIIANLRSN SSTQQTNQVCNESE MSQRSKILQSIRESVR DLRSRRNKYAKALQE LGEAALCGKMLNEE QLIHCIKSMFNTAKC PKGPRFRKTATHSGT
R2	RaR2	FJ461304	Rhynchosciara americana	CAGAACGTGGAGAAACGGAATAACTACCCAGATCCGTTGGTTAACCGG TGGCAAAGTTAATCAAGGTTGCCATAGGCTTAATAACCCTATGGAAATG TTTCCACACACCTCCACGTGGTGCCTGCCGGAAATTGTTCTAGGGTGAA CAAGCTAAGTTGTGAGAAACGGGCTCCACCACAATATGGAGCCTGCCA GGGCGCGAGACTCAGGACTCTCCATGTACAAAGTGGTTAGTTGCAAAA AGAGTGCGCCTAGCATGACTGATAATTTTTCACTGAAATAACGTTGAAC TTTATCTGTGTCATGTGCACAACACTATGGTGTCTGATCAAGCACCATCA GTGGTGGACCTGCTAATGTATTAGTAGAACGTGTCCAGGCGATAATGCA CACACGGCCTCCGGGCCCATCGCTTTTTTTGAGATTCCCTAGAAACTTCA GTGTGTGCGACAACTGTATAACCCATAAGGGATGGACAAAGGTTATACT AGGGGGTAAAAACCCTAATCGGCTAATGGCAAATGGGATGTAGAAATG CCAAAGATTACTCGCACCGAATAATGGTGGCCGAAAAGCGGGTAATCG AATGAAATGGTAATGCTGTAGCGGAAACATGATCACATTCTGTGACAGT AAACCATTAGACCTAGGGGGAACTATGATTAACAAGATACCAGCTTACA TGGAAGCAATGAAATGTAAGTCACAGTAGTGATAAGTGGTGAAGAGTC	CAGAACGT GGAGAAA CGGAATAA CTACCCAG ATCCGTTG GTTAACCG GTGGCAAA GTTAATCA AGGTTGCC ATAGGCTT AATAACCC TATGGAAA TGTTTCCA CACACCTC CACGTGGT			

	TTGTAATCACCGTAACTAGGCCAGGTTCTGGGGATGCCATGAACTTAGG	GCCTGCCG	GTTGTA	NKQQRQQRYARVQ
	GGGAGTATGGTTAGCAGATCTACCAGCTAACACTATTACTGATAATATG	GAAATTGT	AAGCCG	KLYKMNRKVAAKMV
	TAAGCCGCAGTAGCGCTAAGTGGTGTACAGATTTGCAACCACCGTAACT	TCTAGGGT	TTGGGG	LEETDKIQIKLPDHDP
	AAGTTCTGTTTCGATGGACTAGGGGGAATCATGATTAACAAGATACCAA	GAACAAGC	AATCCG	MFKFWESEFKEGEG
	CTTACATGGAAGTTATGAAATGTAAGTCACAGTAGTGATAAGTGGTGAA	TAAGTTGT	GCTACA	MPERMPKDLKESPD
	GAGTCTTGTAATCACCGTAACTAGGCCAGATGTAGTCAAAGCACATGTT	GAGAAAC	CATGGT	LKAIWDPVTEEEVRK
	TAGGGGGAACAAGGTTAACATTGGTAAAAGACCAATGCAACCTCCGTA	GGGCTCCA	ATCTCA	AKVANNTAAGPDGI
	ACTAAACGTGAAGGCACAAAACTAAAAGTCCAGTGGATGACAGGTGAG	CCACAATA	GGAGCC	QPKSWNRISLKYKTLI
	GTCATCACTGGACCCAAATGTTTTAAGCTCATCATAACAACACGGTGAA	TGGAGCCT	ATTCATG	YNLLLYYEKVPHKLKV
	AAATCCAGCATTTATTTGCCTGATTGAGTAGCTTCCACACTATTCCAAAG	GCCAGGG	CGCTGA	SRTVFIPKKKDGSSDP
	CCGAACCTATCTGGGTTTTTCTTGAAAGGCCGTATAGGGCATATGTCGA	CGCGAGAC	TCTCATT	GEFRPLTICSVVLRGF
	GAAATAAGTCCAAGGTGAGGTAGTGTGGCCCTGTACCCAGGGGTAAGG	TCAGGACT	AAGGCG	NKILVQRLVSLYKYDE
	TACTATACGGTCGAGTGGCTCAGTAGGCCTAACTAGCCACTGAGTCACT	CTCCATGT	TAATAA	RQTAYLPIDGVGTNI
	ATAATGACTAGTTATGAGCAACTACAATGAAACAAACACGAGTGGTGG	ACAAAGTG	ACTGTG	HVLAAILNDSNTKLSE
	GGATAACCCTCGGATGGCCACTCAAACCACTGGATCGCTATCCAGTGGA	GTTAGTTG	AAACAG	LHVALLDITKAFNRLH
	CCAATTAACCAGCACACATGTGAACTATGTTGTAGGACATTTGGTACAC	CAAAAAGA	ATCCTG	HTSIIKSLVGKGFPYG
	GTGCCGGATTGGGACAACATGTGAGAAAAAACCCATCCAATTGAGTCCAA	GTGCGCCT	ATAATG	FITFIRRMYTGLQTM
	CCAAAGCATTAATGTAGAACGAAAGAAAAGAAGATGGTCACCTGAAGA	AGCATGAC	CCGTGC	MQFEGHCKMTQVN
	GATAAGACGTATGGCGAATATGGAAGCCCAAGCTACCATAAACAACATC	TGATAATT	TACCAA	RGVYQGDPLSGPIFLL
	AAACACTTAACTCAATACCTTGCAACTTACCTTCCTCAAAGAACTTTGAA	TTTCACTG	ATGATG	AIEKGLQALDKEVGY
	CGCTATTAAAGGAAGGAGAAGATGCTGAGTATAAAGAACTGGTGAC	AAATAACG	TAACGA	DIGDVRVNAGAYAD
	CGGCATTATAGCAAATCTTCGGAGTAATTCCAGTACACAACAAACCAAT	TTGAACTT	GGCGGA	DTDLVAGTRLGLQD
	CAGGTCTGTAATGAAAGTGAAATGTCACAACGCAGTAAGATACTGCAAA	TATCTGTG	AATAAA	NINRFSSTIKQVGLEV
	GTATACGAGAAAGCGTACGTGACTTGAGAAGTAGACGTAATAAATATG	TCATGTGC	ATTAATC	NPRKSMTLSLVPSGK
	CGAAAGCATTACAAGAGCTTGGGGAAGCAGCATTGTGCGGTAAAATGT	ACAACACT	TGGGGC	EKKMKVETGKPFRA
	TGAACGAAGAACAGCTAATACATTGTATAAAGTCAATGTTCAACACCGC	ATGGTGTC	GTTCTG	NDVPLKELSINDFWR
	TAAGTGTCCGAAAGGACCAAGATTTCGCAAAACTGCAACCCACAGTGGA	TGATCAAG	CGGAAT	YLGISYTNEGPERLSL
	ACAAATAAGCAACAACGGCAACAACGATATGCTAGGGTGCAGAAGTTG	CACCATCA	GACTAC	TIEQDLERLTKAPLKP
	TACAAAATGAACCGAAAGGTTGCCGCAAAAATGGTGTTAGAAGAAACG	GTGGTGG	TAAATAT	QQRIHMLNAYVIPKY
	GATAAGATACAGATAAAGTTACCGGACCATGATCCGATGTTCAAGTTTT	ACCTGCTA	AGCGAT	QDKLVLSKTTAKGLK
	GGGAAAGTGAATTTAAGGAGGGCGAAGGCATGCCGGAAAGGATGCCC	ATGTATTA	GCTATAT	RTDRQIRQYVRRWL
	AAGGATCTGAAGGAATCGCCCGACTTAAAAGCGATATGGGATCCGGTA	GTAGAACG	ATACAA	KLPHDVPIAYLHAPV
	ACTGAAGAGGAAGTTAGAAAGGCAAAAGTGGCGAATAACACCGCGGC	TGTCCAGG	ACGACT	KSGGLNIPCLQYWIPL
	AGGTCCTGATGGAATTCAGCCGAAATCATGGAACAGGATAAGTTTAAAA	CGATAATG	GATGGT	LRVNRVNKITESQRS
	TATAAGACGTTAATATACAACTTACTCCTGTACTATGAAAAAGTTCCACA	CACACACG	AACACC	VLAAVGKTALLTSTVY
	TAAGCTTAAGGTGTCACGAACTGTGTTTATTCCTAAGAAAAAAAGATGGA	GCCTCCGG	GGCCTT	KCNQSLATLGGNPT
	TCCAGTGATCCTGGTGAATTCCGACCTCTTACAATATGTTCAGTGGTCCT	GCCCATCG	A (SEQ	MLAYRTYWEKELYAK
	TAGAGGTTTCAATAAGATCCTTGTGCAGAGGCTTGTATCCCTATATAAAT	CTTTTTTTG		VDGKDLQNARDDKA

			ATGATGAAAGGCAAACAGCTTATTTGCCTATAGATGGTGTGGGCACGA	AGATTCCC	ID NO: 1371)	STRWNGMLHSDISG
			ACATACACGTGCTTGCGGCGATACTCAATGACTCGAACACACAAAACTGAG	TAGAAACT		EDYLNYHKLRTNSVP
			TGAACTGCATGTAGCGCTATTGGACATTACGAAAGCATTCAATCGCCTA	TCAGTGTG		TKVRTARGRPQKETS
			CATCACACATCGATAATCAAATCCTTAGTGGGAAAGGGGTTCCCTTATG	TGCGACAA		CRGGCKSTETLQHVV
			GATTTATTACTTTCATCAGAAGAATGTATACGGGCTTACAAACGATGAT	CTGTATAA		QQCHRTHGGRTLRH
			GCAATTTGAGGGCCACTGCAAAATGACACAAGTGAACAGAGGAGTCTA	CCCATAAG		DRIVGLLQHELRRDY
			CCAAGGAGACCCTTTAAGCGGGCCAATCTTTTTGTTGGCTATTGAAAAA	GGATGGA		NVLAKQELKTGIGLR
			GGCTTACAAGCTCTTGATAAGGAAGTTGGCTATGACATAGGGGATGTG	CAAAGGTT		KPDLVLIKDDTAHIVD
			AGAGTCAATGCAGGTGCTTACGCAGATGATACAGACTTGGTTGCAGGA	ATACTAGG		VQVARCSKLNESHVR
			ACGAGACTTGGCTTGCAAGACAACATAAACAGGTTTTCAAGTACTATTA	GGGTAAA		KRSKYDKKEIEVEVKS
			AACAGGTTGGACTTGAGGTAAATCCAAGAAAGTCAATGACATTGTCTCT	AACCCTAA		RYRVSKVMYEACTIS
			GGTACCTTCGGGGAAAGAGAAGAAAATGAAAGTAGAAACTGGCAAACC	TCGGCTAA		YKGIWDKQSVMSM
			TTTCAGAGCGAACGATGTTCCTCTGAAGGAACTGTCAATTAATGATTTTT	TGGCAAAT		RRLGVSEYCLFKIVTS
			GGAGATACCTGGGCATAAGCTATACGAATGAAGGGCCTGAAAGACTGA	GGGATGTA		TLRGTWLCWKRFN
			GTCTTACAATAGAGCAAGATCTAGAACGCCTAACTAAAGCTCCATTAAA	GAAATGCC		MITSVRS (SEQ ID
			GCCACAACAACGGATACATATGCTTAACGCATATGTAATACCTAAATATC	AAAGATTA		NO: 1493)
			AAGACAAACTTGTCTTGAGTAAAACCACGGCAAAAGGGCTGAAGCGAA	CTCGCACC		
			CAGATAGGCAGATAAGGCAATATGTAAGACGATGGCTAAAGTTACCAC	GAATAATG		
			ATGATGTTCCAATTGCATATTTGCATGCACCAGTTAAATCTGGTGGCCTA	GTGGCCGA		
			AACATTCCATGCTTGCAATACTGGATACCATTACTTAGAGTAAACAGAGT	AAAGCGG		
			CAATAAGATTACAGAAAGTCAACGCAGTGTACTTGCAGCGGTTGGTAAA	GTAATCGA		
			ACAGCCCTGCTAACGTCAACTGTTTATAAATGTAATCAGTCTCTAGCAAC	ATGAAATG		
			ATTGGGTGGAAACCCAACAATGCTTGCATATCGAACGTATTGGGAAAAA	GTAATGCT		
			GAGCTATATGCAAAGGTCGATGGGAAAGACCTACAAAATGCGCGGGAT	GTAGCGG		
			GATAAGGCTTCGACGAGGTGGAATGGCATGTTGCATAGTGACATTAGT	AAACATGA		
			GGAGAAGACTATTTGAATTATCATAAGCTTCGAACAAATAGTGTACCCA	TCACATTC		
			CTAAAGTAAGAACGGCAAGAGGACGTCCTCAAAAGGAAACGTCTTGCA	TGTGACAG		
			GAGGTGGTTGTAAAAGCACGGAAACGCTGCAACATGTAGTTCAGCAGT	TAAACCAT		
			GTCATAGAACACATGGGGGAAGAACTCTAAGACATGACAGGATTGTGG	TAGACCTA		
			GTCTTCTCCAACATGAACTACGCCGAGACTATAATGTCTTAGCTAAACAA	GGGGGAA		
			GAACTAAAAACGGGCATAGGGTTGCGCAAACCGGATTTAGTTCTTATAA	CTATGATT		
			AAGATGACACTGCACACATCGTTGACGTGCAAGTCGCAAGGTGTAGTAA	AACAAGAT		
			ACTGAATGAAAGCCATGTAAGGAAGCGATCCAAATACGACACAAGAAAGA	ACCAGCTT		
			AATAGAGGTAGAAGTAAAAAGTAGATATCGTGTAAGTAAGGTGATGTA	ACATGGAA		
			TGAAGCTTGCACGATATCATACAAAGGTATTTGGGATAAGCAGAGCGT	GCAATGAA		
			GATGAGCATGAGAAGACTTGGAGTAAGTGAATACTGTCTTTTCAAGATT	ATGTAAGT		
			GTCACATCAACACTAAGAGGAACATGGCTTTGTTGGAAACGGTTCAATA	CACAGTAG		
			TGATCACTAGTGTTCGATCATAGAAATGTGTGCGATAAGGTGTGAATAG	TGATAAGT		

	AAGGGTTCACCAAGAGGGAGACCTAGTTTGGACCTCAGAAATGGGGTC ATAGGAGTGATAGGTTGTAAAGCCGTTGGGGAATCCGGCTACACATGG TATCTCAGGAGCCATTCATGCGCTGATCTCATTAAGGCGTAATAAACTGT GAAACAGATCCTGATAATGCCGTGCTACCAAATGATGTAACGAGGCGG AAATAAAATTAATCTGGGGCGTTCTGCGGAATGACTACTAAATATAGCG ATGCTATATATACAAACGACTGATGGTAACACCGGCCTTA (SEQ ID NO: 1125)	GGTGAAG AGTCTTGT AATCACCG TAACTAGG CCAGGTTC TGGGGAT GCCATGAA CTTAGGGG GAGTATGG TTAGCAGA TCTACCAG CTAACACT ATTACTGA TAATATGT AAGCCGCA GTAGCGCT AAGTGGTG TACAGATT TGCAACCA CCGTAACT AAGTTCTG TTTCGATG GACTAGG GGGAATCA TGATTAAC AAGATACC AACTTACA TGGAAGTT ATGAAATG TAAGTCAC AGTAGTGA TAAGTGGT GAAGAGTC TTGTAATC ACCGTAAC TAGGCCAG ATGTAGTC AAAGCACA			

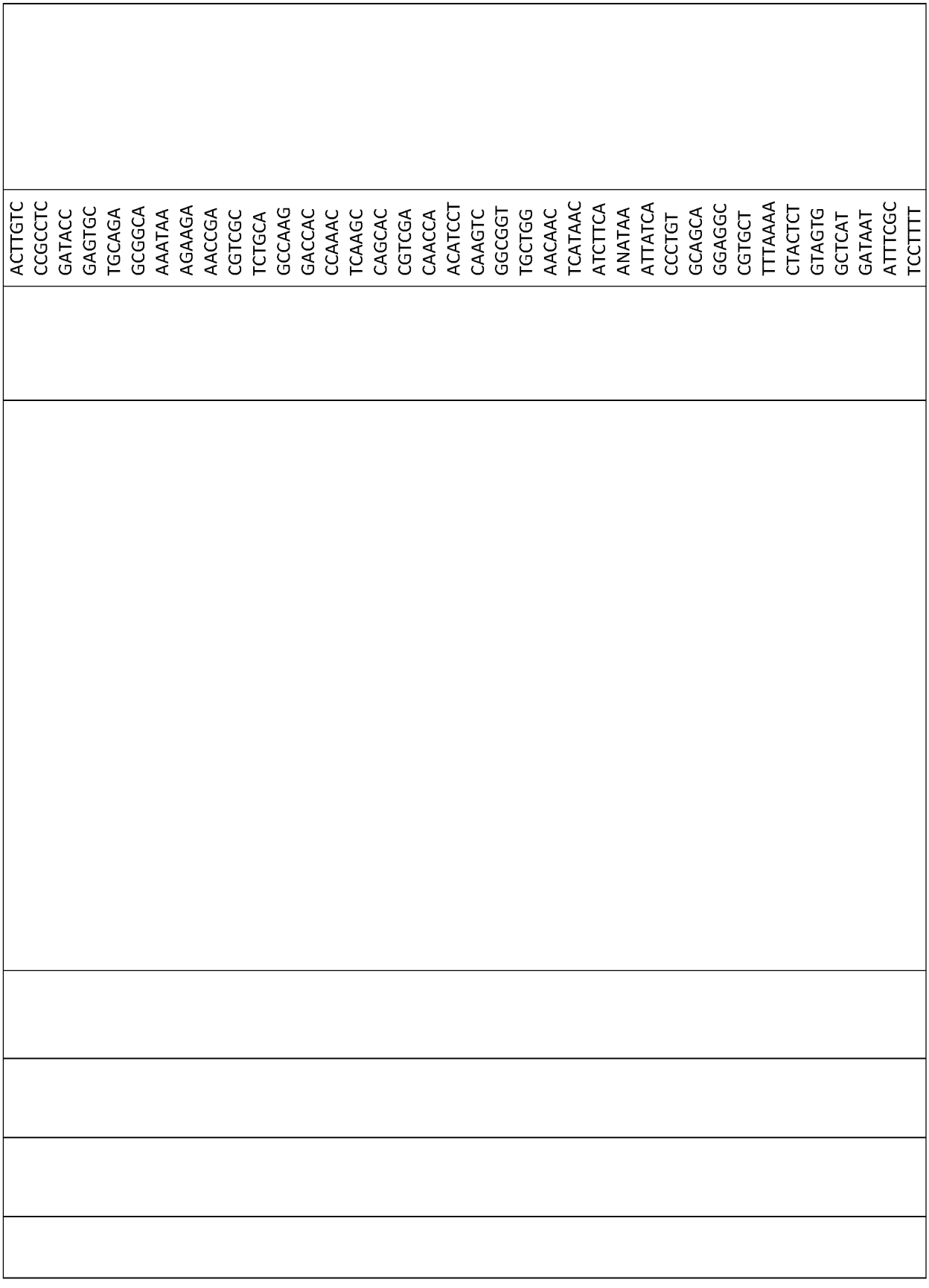

TGTTTAGG
GGGAACA
AGGTTAAC
ATTGGTAA
AAGACCAA
TGCAACCT
CCGTAACT
AAACGTGA
AGGCACAA
AACTAAAA
GTCCAGTG
GATGACAG
GTGAGGTC
ATCACTGG
ACCCAAAT
GTTTTAAG
CTCATCAT
AACAACAC
GGTGAAA
AATCCAGC
ATTTATTT
GCCTGATT
GAGTAGCT
TCCACACT
ATTCCAAA
GCCGAACC
TATCTGGG
TTTTTCTTG
AAAGGCC
GTATAGGG
CATATGTC
GAGAAATA
AGTCCAAG
GTGAGGTA
GTGTGGCC
CTGTACCC
AGGGGTA
AGGTACTA

				TACGGTCG AGTGGCTC AGTAGGCC TAACTAGC CACTGAGT CACTATAA TGACTAGT T (SEQ ID NO: 1248)			
R2	YURE-2_Cis	.	Ciona savignyi	CCAAAATTACTTCCAGCACCTCCACAGCAGACGAACGAAGAAAGAAGAC	CCAAAATT	TAAGAT	MAGHKITMSEGKLLE
				TTACGAAATAAGAATAAGCAGTTTAAAGACGAAGCAGACGAACACCAC	ACTTCCAG	CCGCGG	VAVRYGGVRNVSYE
				TCCACCAACGACGCTCTGCAGCTACACCACCACCATCTCGCCGCAACAA	CACCTCCA	CTGTGG	CPVPDCTKTFSQANN
				GAAGAATTTTCCGCTGCTCTCCACTCTGCTCCAACACCACCTCTCCTGCTC	CAGCAGAC	CGCCGA	LIRHLNNFGNTKHRA
				TGTGGACTGCTGCCTTGCTGCTGGACCAACCTCTACCCGAAGGAACCCT	GAACGAA	ACGAGC	HNFTYFFTCEKCKIQI
				TCGAACCCAGCAAGGTACGTGTCACCACCTCTCCACAGAGCCAAGG	GAAAGAA	ACCTGC	HSNTKHNISNHYKQC
				CCAGAGTGGATAGAGCAGCGGCCTCTACAACCAAGTTCTCCACTCCGAC	GACTTACG	CCATTCT	CATGGGPSCETGQYF
				GACAAAACACCTGCCTCGTGGCCAGAGTCCTGCCGAAGAAGAATCTTCG	AAATAAGA	TCTTGTA	CPACEQAGLGNLESA
				ACCCCAACACCGTCTCGGCAGGGCTCCAAGCAACGACATCAGCCACGAG	ATAAGCAG	GGGACT	LRHFQSSHPEFNLPP
				TGCCCACCGGAGTGGACGCGGCGAACCCGAGGACCGCTGTCGCCAAGA	TTTAAAGA	TTTTCAC	RSQFSKSHPNSYTLSL
				ACAATAACATCACCGCCTGCAGCAGCTACCAAACAAAGGTTAGTCTCCT	CGAAGCA	CCTCACT	KPKDHLMKILYSGPLT
				ACCTCACCTACAAACTCGTCTGAATAGACGCCCCCGCGTGGGAGCTAAC	GACGAACA	CCCCCCA	PGQLVCPIKICLRSSA
				CTTGGTAGCTAGCTGCAGTGCCCTGCGACAGCGGCCTCGAAGAGCTGC	CCACTCCA	ATAGTTT	ARLFHDVSKLRKHML
				AGAGCGTCAGCCTGCTTCGACCTCGCTTGTTCTCATTTCAACCTACTCCG	CCAACGAC	TTTTTCG	VDHNRTLVYETTCGK
				TCCTGTGGAATCAAAAGAGCCCCACTAACAATTACATTCATAAAAATCTAG	GCTCTGCA	TTTTTTC	CLRPVDTSKNMRKTT
				CAAGAACGAAGAGAAGCGACCAACTTTTAATCCATAACTTTTAGATCTTT	GCTACACC	GTTTTTT	SHFEKCSGESFISSPSP
				TTATTTATTACTGTTTTTAAGCCCTAAGCATATTGCCTTTTTTTTAGATCTTA	ACCACCAT	CACCCCC	IPQKTYKLDLPSTSTP
				TAATTATAAAAATAGATTCAAAGTTAACACCACCAGGCCGCTACAGAGC	CTCGCCGC	ACCCCA	PPRKSPKLQPYKPIRT
				ATTTTATTTAATCAATTTGTTACCGACCTCCTGCTGCTTCTTTTTACTTTCT	AACAAGAA	CTCGCCT	FKNPLTKSSQSKSDN
				CCAGACTACTACCCGGATACACAACCCTTGGAAACGAGAGGAATGGCCGG	GAATTTTC	CTGGGC	PPKPTPFFSPRTLERS
				TCACAAGATCACAATGTCGGAGGGGAAGCTGCTGGAGGTTGCTGTACG	CGCTGCTC	TGCACA	ASWPALSEVVDPLPK
				GTATGGTGGTGTGAGGAATGTCTCGTACGAGTGTCCAGTCCCGGATTGC	TCCACTCT	TCCCACA	LKEKHPSLPCALDKCP
				ACAAAAACCTTCTCACAGGCAAATAACTTGATACGGCATTTAAATAACTT	GCTCCAAC	CGTAGG	PSPRIKPSTLVPPCHT
				TGGCAATACAAAACACCGAGCACACAATTTCACCTATTTTTTCACCTGTG	ACCACCTC	GACCTG	ANNSPKPTSPESPSTL
				AGAAGTGCAAAATCCAAATTCACAGCAACACACAAAACATAATATTTCAAA	TCCTGCTC	TTTATAT	KPLPRPIRPSKPLEDW
				TCATTACAAACAGTGTTGTGCAACCGGCGCGGCGACCCTCGTGCGAGAC	TGTGGACT	TATTTGC	LTVRSVGPDREIVLNI
				GGGTCAATACTTCTGCCCTGCCTGCGAGCAAGCAGGGCTGGGGAACCT	GCTGCCTT	CTTTTAT	GPRPRPGPAAGSRTT
				GGAGTCGGCACTCCGCCATTTTCAATCCTCCCACCCTGAATTTAATCTAC	GCTGCTGG	ATGTAC	SPPSTAPAKRVAANP
				CCCCCCGGTCTCAATTTTCAAAATCTCATCCCAATAGTTATACCCTATCCT	ACCAACCT	CACTTTT	IAAPLSGEPGATLDC

	TAAAACCTAAAGACCACCTTATGAAAATACTTTACAGTGGACCACTGACT	CTACCCGA	TAAATAT	GQTGRKVQPPKKRP	
	CCGGGGCAATTGGTATGCCCCATTAAGATCTGTCTCAGGTCCTCAGCTG	AGGAACCC	ATTTTTG	TESAGSLPPPAEPAT	
	CACGTCTTTTCCATGACGTGTCCAAGCTGAGGAAGCACATGTTGGTTGA	TTCGAACC	TACCCCA	DLLTGREGLARLVEEY	
	TCACAACCGAACCCTTGTGTACGAGACAACATGTGGCAAATGCTTGCGG	CAGCAGCA	CAAGAT	HLSGDFGAFCRDLER	
	CCTGTCGACACCTCGAAAAATATGCGGAAAACGACGTCGCACTTCGAAA	AGGTACGT	GCTTTTC	WTALSSTNRRPKPRR	
	AGTGTTCTGGCGAATCCTTTATCTCATCCCCCTCACCCATACCCCAAAAG	GTCACCAC	GCCAAA	GRYNRGAAARATRN	
	ACTTATAAACTCGACTTACCCTCAACCAGCACACCCTCCCCCCCGTAAATCT	CTCTCCAC	AAAAAA	RGRDDRQDPQDRD	
	CCCAAATTACAACCCTACAAACCAATTCGAACCTTTAAGAACCCCCTCAC	AGAGCCAA	AAATTTT	DQGGPGPVTCGRPQ	
	CAAGTCCAGTCAATCCAAATCCGACAACCCACCCAAACCTACCCCTTTTT	GGCCAGA	TGTATCA	RYKRAAALRSAFGRD	
	TCTCGCCAAGAACACTTGAGAGGTCGGCGTCGTGGCCCGCATTGAGCG	GTGGATAG	CATTTTT	MKATVRRIIDGERGD	
	AGGTTGTCGACCCACTGCCCAAGCTCAAGGAAAAACACCCCTCGCTGCC	AGCAGCG	ATATTTT	ARCEIDPKTIEGRFRD	
	ATGCGCCCTGGATAAGTGCCCACCCTCACCCCGGATCAAACCCTCGACC	GCCTCTAC	GTAAAA	ELSPPVREGPECSLPP	
	CTAGTCCCTCCCTGCCATACAGCAAACAATAGTCCCCAAACCAACTTCACC	AACCAAGT	CACAGA	WMAEAQAGEHAPS	
	CGAATCACCCTCTACCCTAAAACCCTTACCCCGCCCCCATCCGACCATCTA	TCTCCACT	TTTTTAT	NDSQPGDAYDGPIT	
	AACCCTTGGAGGATTGGTTAACTGTGCGGAGCGTGTGGGACCGGATCGGG	CCGACGAC	AAACTTT	ALEVEMVLSTLNVGS	
	AGATCGTCCTGAACATCGGTCCGAGACCCCGACCCGGCCCCCGCAGCAG	AAAACACC	GCACTA	APGSDGLSYGFWRA	
	GCTCCAGGACAACATCTCCCCCGTCAACGGCACCCGCGAAAAGGGTTGC	TGCCTCGT	TTTTTAT	LDPKGLVLSELFEVCR	
	TGCAAACCCAATCGCAGCACCCTTGAGCGGGGAGCCGGGTGCGACCCT	GGCCAGA	ATAAAC	IERRVPGPWKSSRVT	
	TGATTGCGGCCAGACCGGGCGAAAGGTTCAGCCGCCAAAAAAGAGACC	GTCCTGCC	TTCGCAC	LICKDAEGDLDDLGN	
	GACAGAGAGTGCTGGGTCCCTCCCTCCTCCGGCCGAGCCAGCCACCGAT	GAAGAAG	TTATTTA	WRPISICQTVYKIYAA	
	CTCCTTACAGGGAGAGAGGGGCTGGCGAGGCTGGTTGAGGAATATCAT	AATCTTCG	AAATGA	VLARRLQSWALDGG	
	CTGTCGGGGGACTTTGGCGCTTTCTGTCGGGATCTCGAGCGATGGACG	ACCCCAAC	ATCGCA	VISRSQKGFMPFEGV	
	GCGCTTTCGTCGACCAACCGAAGGCCAAAGCCGAGGCGCGGCAGGTAC	ACCGTCTC	TCTTTTT	YEHVFLLDSVVADAR	
	AACCGCGGTGCGGCTGCCCGTGCCACCAGGAACCGGGGCCGAGATGAT	GGCAGGG	TATATAC	ATRRSLAVCWLDLRN	
	CGTCAAGATCCACACAAGATCGAGACGATCAAGGAGGCCCCGGTCCGGTG	CTCCAAGC	ACCAAC	AFGSVDHTTIVEALSR	
	ACATGCGGGGCGGCCGCAAAGATACAAGCGGGCCGCCGCTGAGGTC	AACGACAT	ACAAAC	FGAPAGLVEMISDIYT	
	GGCCTTCGGGAGGGACATGAAGGCGACCGTCCGTCGCATCATAGACGG	CAGCCACG	AGGATG	GGSCRIRTRAGFTPDI	
	AGAACGGGGCGACGCTCGCTGTGAAATCGACCCCAAAACCATCGAGGG	AGTGCCCA	TGCAGC	PVGRGVRQGCPLSGI	
	TCGATTTCGCGACGAGCTGTCGCCGCCGGTTCGGGAAGGGCCGGAGTG	CCGGAGTG	TCAGGG	IFNLVMEVLLRGVEA	
	CTCGTTGCCCCCTTGGATGGCTGAAGCCCAGGCAGGTGAGCATGCGCCA	GACGCGG	GAACCA	NNACGYRLSCAGGA	
	TCGAATGATAGCCAACCCGGCGATGCATACGATGGGCCAATCACGGCG	CGAACCCG	ATCCTGC	SVRVLAYADDVALVG	
	CTGGAGGTGGAGATGGTCCTCAGTACCCTGAACGTCGGCTCTGCCCCGG	AGGACCGC	GTCCCTC	SSRAEXKIQLGVCERF	
	GATCGGACGGCCTCTCCTATGGATTTTGGAGAGCTCTAGACCCGAAAGG	TGTCGCCA	CTAGCG	AAWAGFSFNNKKCA	
	ACTCGTCCTTTCGGAGCTCTTCGAAGTCTGTAGGATCGAGCGCCGGGTC	AGAACAAT	GCGGGA	AMVLKHQRGGRRLL	
	CCGGGGCCATGGAAGAGCAGCCGGGTCACCCTCATCTGCAAAGATGCA	AACATCAC	GGGGCG	DSAPLRLCGEEVAILG	
	GAGGGTGACCTCGACGATTTGGGGAACTGGCGGCCCATCTCCATCTGCC	CGCCTGCA	CCCACCT	PDSFYKYLGAHTGYG	
	AGACGGTGTATAAGATCTACGCCGCCGTGTTGGCCCGTCGACTGCAGA	GCAGCTAC	ACCCCC	RQTGGQLVDRVERQ	
	GCTGGGCGCTTGACGGTGGCGTCATCTCCCGAAGTCAAAAGGGCTTCAT	CAAACAAA	ACGCTC	VVRLFTSFLTPTQKLS	

	GCCGTTCGAGGGGGTATACGAGCATGTATTCCTCCTCGACTCGGTCGTC	GGTTAGTC	CTCTTGA	ALKRIVLPAMSFHLR	
	GCTGACGCGCGCGCGACCCGGAGGTCGCTGGCCGTGTGCTGGCTNGAC	TCCTACCT	GCACCA	VRPCAEGHLRRLDNT	
	CTCCGGAACGCGTTCGGCAGCGTCGACCACACCATAGTCGAGGCGT	CACCTACA	ACAGGG	VRRCVKTALRLPKGS	
	TGTCTCGCTTCGGCGCCCCCCGCGGGCCTCGTGGAGATGATATCGGACAT	AACTCGTC	ACTCCCT	CRAFFHTSPDAGGLG	
	CTATACGGGTGGGTCCTGCCGAATTAGAACTCGTGCGGGATTCACCCCG	TGAATAGA	TCCGGA	ITSVVAECDILTVTQA	
	GACATCCCTGTCGGTCGTGGAGTGCGACAGGGGTGTCCCTTGTCCGGTA	CGCCCCCG	GCCCCT	FKMLSSPDHLVSLVA	
	TCATCTTCAACCTCGTGATGGAGGTCCTGCTGCGGGGCGTCGAAGCGAA	CGTGGGA	GCACCC	KGRLGMHAARMGR	
	CAACGCTTGTGGATACCGGCTCTCCTGCGCCGGCGCGTCCGTCAGG	GCTAACCT	TCAACTT	SETASACAMADYLSG	
	GTGCTCGCGTACGCCGACGATGTGGCTCTGGTGGGCTCTTCCAGGGCCG	TGGTAGCT	TTCTTAT	DSVMGHXSWKTGYR	
	AGATNAAGATCCAGCTGGGTGTGTGCGAGCGGTTTGCCGCCTGGGCCG	AGCTGCAG	TTTTAAA	MPADLWTATRAASR	
	GTTTTTCTTTTAACAACAAGAAGTGCGCCGCCATGGTACTGAAACATCA	TGCCCTGC	AAAAAA	RLSLRFSPQPQGEFG	
	GAGAGGGGGTCGGAGGCTCTTGGACTCGGCGCCTCTTCGCCTTTGCGG	GACAGCG	AAATCA	LESGTFKIAPRERRSL	
	TGAAGAGGTGGCGATCCTGGGCCCCGACTCCTTCTACAAGTACCTGGGG	GCCTCGAA	TATATAT	TRRLHHRQNLWWR	
	GCGCATACCGGCTACGGGCGGCAAACGGGTGGACAGCTTGTCGATCGA	GAGCTGCA	TGATCTT	NQWAALPNQGKTV	
	GTCGAAAGGCAGGTCGTCAGGCTGTTCACCTCGTTCCTCACCCCCACCC	GAGCGTCA	GACGAC	AAHSAYAASNNWVK	
	AAAAACTCTCGGCCCTGAAGAGAATAGTCCTGCCCGCCATGAGCTTCCA	GCCTGCTT	GGGGGC	GPSSLAPQALFFGLK	
	TCTCCGGGTCCGGCCCTGCGCCGAGGGGCATCTCCGGCGCCTTGACAAC	CGACCTCG	TACATTC	ARLNQMPTRSVKAC	
	ACAGTGAGGCGCTGTGTGAAGACAGCGTTGCGCCTGCCGAAGGGGTCG	CTTGTTCT	AGCCCC	YSRAPNYDKSCRRCG	
	TGTCGGGCTTTTTTCCACACGTCCCCCGACGCAGGGGGCCTGGGGATCA	CATTTCAA	CAAAAA	AEVETLPHVLNHCPK	
	CGTCGGTTGTCGCCGAGTGCGACATACTGACGGTGACCCAGGCCTTCAA	CCTACTCC	CCCACCC	SMKSILERHDSVLAE	
	GATGCTGTCTTCACCGGACCATCTCGTCTCGCTAGTTGCCAAGGGCCGTC	GTCCTGTG	ACCATCC	VLAAIPRGTFASVDV	
	TGGGGATGCACGCCGCGCGCATGGGCCGGTCCGAGACGGCGTCCGCTT	GAATCAAA	CCAACG	DRTSREHFRRVGEAL	
	GCGCCATGGCGGACTACCTGAGCGGGGGACTCGGTAATGGGGCACANGT	AGAGCCCC	AGTGCC	RPDIVARRHDGSVVV	
	CGTGGAAAACCGGATACAGGATGCCGGCCGATCTCTGGACAGCCACTC	ACTAACAA	GGGGCA	ADVTCPFESCASALD	
	GAGCTGCCAGCCGGCGCCTGTCCCTCCGGTTTTCCCCGCAGCCCCAAGG	TTACATTC	TTGAAG	TAAARKIEKYDQLCA	
	CGAATTCGGCCTCGAGTCGGGCACTTTCAAGATCGCCCCCCAGGGAGCG	ATAAAATC	AGCTCC	NLRQLYRKPVESHAL	
	GCGCTCCTTGACCCGAAGGCTGCACCACAGGCAAAACCTGTGGTGGCG	TAGCAAGA	GGCACA	VVGSLGSWGRTNNT	
	GAACCAGTGGGCGGCCCTTCCCAACCAAGGGAAGACCGTCGCCGCCCA	ACGAAGA	ATTAGC	ALAALGIRGAVRSRL	
	CTCCGCCTACGCAGCCTCCAACAACTGGGTCAAGGGCCCGTCCTCCCTG	GAAGCGA	ACTTAG	AKQLVNLSVEGSHNI	
	GCCCCTCAGGCCCTGTTCTTCGGCCTTAAAGCGCGGCTGAACCAGATGC	CCAACTTT	CTTATTT	WLRWSGGIPKDLVR	
	CAACGCGCTCGGTCAAGGCCTGCTACTCAAGGGCGCCGAACTACGACA	TAATCCAT	ATTATTT	(SEQ ID NO: 1494)	
	AGTCCTGTCGTAGGTGTGGCGCGGAGGTGGAGACCCTCCCGCACGTGC	AACTTTTA	TTTGTCA		
	TGAATCATTGCCCCAAGTCCATGAAATCGATCTTGGAGCGGCATGATTC	GATCTTTT	ACATTTT		
	GGTGCTTGCGGAGGTCCTCGCCGCCATCCCTCGCGGCACATTCGCCAGT	TATTTATTA	TGTTTTT		
	GTCGACGTCGACAGGACGTCCCGAGAACATTTCCGGCGAGTGGGCGAA	CTGTTTTT	TCAAATT		
	GCTCTGCGACCCGACATAGTCGCCCGTCGACATGACGGATCCGTCGTGG	AAGCCCTA	TTTTCAC		
	TCGCAGACGTGACGTGCCCATTCGAGTCCTGCGCGTCGGCCCTCGATAC	AGCATATT	CCCTCAC		
	GGCGGCCGCGCGAAAAATCGAAAAATACGACCAGCTGTGTGCGAACTT	GCCTTTTTT	CCCCACC		

		GCGGCAGTTATACCGTAAGCCCGTCGAGTCGCACGCACTGGTCGTAGGT	TAGATCTT	CTAATA	
		TCCTTGGGCAGCTGGGGCAGGACCAACAACACTGCTCTGGCTGCACTCG	ATAATTAT	GGTCCC	
		GAATCCGAGGCGCGGTTCGCTCGAGGTTGGCCAAGCAATTGGTCAACC	AAAAATAG	TCGGGC	
		TCAGCGTCGAGGGTAGCCACAACATCTGGCTACGCTGGTCCGGCGGCA	ATTCAAAG	TTGGGC	
		TCCCAAAGGACCTGGTCAGATAAGATCCGCGGCTGTGGCGCCGAACGA	TTAACACC	CCCTTTT	
		GCACCTGCCCATTCTTCTTGTAGGGACTTTTTCACCCTCACTCCCCCCAAT	ACCAGGCC	TCGTGC	
		AGTTTTTTTTCGTTTTTTTCGTTTTTTTCACCCCCACCCCACTCGCCTCTGGGC	GCTACAGA	TCGAGA	
		TGCACATCCCACACGTAGGGACCTGTTTATATTATTTGCCTTTTATATGTA	GCATTTTA	AGCGTC	
		CCACTTTTTAAATATATTTTTGTACCCCACAAGATGCTTTTCGCCAAAAAA	TTTAATCA	ACATCG	
		AAAAAATTTTTTGTATCACATTTTTATATTTTGTAAAACACAGATTTTTATA	ATTTGTTA	CCCCACT	
		AACTTTGCACTATTTTTATATAAACTTCGCACTTATTTAAAATGAATCGCA	CCGACCTC	GACCAC	
		TCTTTTTTATATACACCAACACAAACAGGATGTGCAGCTCAGGGGAACC	CTGCTGCT	GACCTTC	
		AATCCTGCGTCCCTCCTAGCGGCGGGAGGGGCGCCCACCTACCCCCACG	TCTTTTTAC	CCCGAC	
		CTCCTCTTGAGCACCAACAGGGACTCCCTTCCGGAGCCCCTGCACCCTCA	TTTCTCCA	ATTGGA	
		ACTTTTCTTATTTTTAAAAAAAAAAAAATCATATATATTGATCTTGACGACG	GACTACTA	GTCCTT	
		GGGGCTACATTCAGCCCCCAAAAACCCACCCACCATCCCCAACGAGTGC	CCCGGATA	GGCGTC	
		CGGGGCATTGAAGAGCTCCGGCACAATTAGCACTTAGCTTATTTATTATT	CAACCCTT	TCCCAG	
		TTTTGTCAACATTTTTTGTTTTTTTCAAATTTTTTTCACCCCTCACCCCCACCCT	GGAAACG	GTCGAA	
		AATAGGTCCCTCGGGCTTGGGCCCCTTTTTCGTGCTCGAGAAGCCGTCAC	AGAGGA	ACAGTC	
		ATCGCCCCACTGACCACGACCTTCCCCGACATTGGAGTCCTTGGCGTCTC	(SEQ ID	CCAAGT	
		CCAGGTCGAAACAGTCCCAAGTGATAGCACCTAATGCTCGACTTGTTTC	NO: 1249)	GATAGC	
		GGCCTGGGCCGCCGAGGATTCCCAGAACGACCATTCTTCTAAATAATAT		ACCTAAT	
		TTATATTTCAGAATAAAACTATATATATATCGTTGGCGGGACTTGTCCCG		GCTCGA	
		CCTCGATACCGAGTGCTGCAGAGCGGCAAAATAAAGAAGAAACCGACG		CTTGTTT	
		TCGCTCTGCAGCCAAGGACCACCCAAACTCAAGCCAGCACCGTCGACAA		CGGCCT	
		CCAACATCCTCAAGTCGGCGGTTGCTGGAACAACTCATAACATCTTCAA		GGGCCG	
		NATAAATTATCACCCTGTGCAGCAGGAGGCCGTGCTTTTAAAACTACTCT		CCGAGG	
		GTAGTGGCTCATGATAATATTTCGCTCCTTTTTTGCCCCGTGTAAACTTA		ATTCCCA	
		GTNGATGCGAATAAAATCAGTTGAATCA (SEQ ID NO: 1126)		GAACGA	
				CCATTCT	
				TCTAAAT	
				AATATTT	
				ATATTTC	
				AGAATA	
				AAACTA	
				TATATAT	
				ATCGTT	
				GGCGGG	


ACTTGTC
CCGCCTC
GATACC
GAGTGC
TGCAGA
GCGGCA
AAATAA
AGAAGA
AACCGA
CGTCGC
TCTGCA
GCCAAG
GACCAC
CCAAAC
TCAAGC
CAGCAC
CGTCGA
CAACCA
ACATCCT
CAAGTC
GGCGGT
TGCTGG
AACAAC
TCATAAC
ATCTTCA
ANATAA
ATTATCA
CCCTGT
GCAGCA
GGAGGC
CGTGCT
TTTAAAA
CTACTCT
GTAGTG
GCTCAT
GATAAT
ATTTCGC
TCCTTTT

R4	Dong	.	Bombyx mori	GCTAGCTCCCTAAAATCCTACCTTACGTCCGAGGCGAACATCTGTCCACG TGGGGAGCGGAAACGCGTACTATCGAAACTTACGCGGCTAACAAGGTA AAGGTAACCCATTAATATGGAGACAAGACTAAAAAGAAAATTGAGAGG GCCGCTTCCCGGGGGCGATCGCCGGGGCACACCTGGAGCTGGCGCTGG GTGTTCCAGCATAGGATCCGTCAGCGGCGGGAGAGTTGAGCAGGCGGGC TCTCGCCGGGGATTTACAACCCGAAAATGCTACAGCCGCCAACAGAAAT CACGATGTAGAAAGTAGGAGCAATAGCCCATGTGAACCTTACAGCCCG AGTACCGGTTCATACAACCCCTCGGTACAATCATCACCATCATCCTCGGG TCATAGAGGCTCACCAACGTCGACTATGGATGTTGCGCAGGGGGCGAA TCTTTTTACCTGCGTCCACCAAAGCTGGCAAAACAAGGGGCCGCATGAA ATGGAGCAGAGAAGTAAACCTATTTATCATGCGCACCTACTACTACGTC ACTAAATTAGAAACTGATCTGACTATTTACAGAAAAAAAGCTGCATGAAC ATTTTTCACTAAAATACCCTAATGTAATAATTTCACAGCAAAGAATATCT GACCAAAAAAGGGCAATAGAAAGAAATAAACTACTATCTCAAGAAACC CTAGACCGACTAAAAGAGGAAGTGAGAAAACAGCTCGAAGACGAACAA ACCAATAATGTTGAAAATGAAAAATTAAATTCAGAAACATATTCACACG AATATACGACACTTACTCCACAAACTATTTTAACGAAGAAAACACAGCA ACATACAAATATTATCTCTAGCACACAGACCTCTCATTCATCCACTCAAAC CGAATCAATAACTCTGCTGTTAGAAAATGAAGTCGATATTTTAAATACTA ACCCCACAGAAGGGGCAACACAGACACAAGAAGTTAAAGATAAATTTG AAACGAATTTAACAATGTACTCAGGCATGGACCCTAAAGCAAGGCCGCC ATTGCCTAAACTTAAATATAGCTCTAAGCTAAATGAACTGATACGTCTAT TTAATAATGACATACTTGTAGATTACATCTCACCAGACACACAACTATCA GATGTACATACATTAACATATTGCACCGCCGTAACTATTTCGGAACAACT AAAATATAAAATTATAGCAATAGAAGGAAACGCGAGACATAAAAAAAA CTTCAAACCACCGTGGCAACAAAGATTGGAGAAGGATATAGCAAAATT GAGAGCAGATATTGGTAAACTGACCCAATACATAAATAATAATAGATCT AAAAAAGTAGTCCAAAGCGTTGAACAAATATTTAAAAACACTAAAATAC	GCTAGCTC CCTAAAAT CCTACCTT ACGTCCGA GGCGAAC ATCTGTCC ACGTGGG GAGCGGA AACGCGTA CTATCGAA ACTTACGC GGCTAACA AGGTAAA GGTAACCC ATTAATAT GGAGACA AGACTAAA AAGAAAAT TGAGAGG GCCGCTTC CCGGGGG CGATCGCC GGGGCAC ACCTGGAG CTGGCGCT GGGTGTTC CAGCATAG GATCCGTC	TTGCCCC GTGTAA ACTTAGT NGATGC GAATAA AATCAG TTGAATC A (SEQ ID NO: 1372) TAAAAA CTAGCA TAATTAT TAACTCA TAACTA ATGTAT ATTACTT GGCCAA AAGCCC GTATAT ACAGTT CCACCG GCTCTG TCGACA GACTGA ACTGAG AAAGGG GAAACA TATGGA AATAAT AATAA (SEQ ID NO: 1373)	MLRRGRIFLPASTKA GKTRGRMKWSREV NLFIMRTYYVVTKLET DLTIYRKKLHEHFSLK YPNVISQQRISDQKR AIERNKLLSQETLDRL KEEVRKQLEDEQTNN VENEKLNSETYSHEYT TLTPQTILTKKTQQHT NIISSTQTSHSSTQTE SITLLLENEVDILNTNP TEGATQTQEVKDKFE TNLTMYSGMDPKAR PPLPKLKYSSKLNELIR LFNNDILVDYISPDTQ LSDVHTLTYCTAVTIS EQLKYKIIAIEGNARH KKNFKPPWQQRLEK DIAKLRADIGKLTQYI NNNRSKKVVQSVEQI FKNTKIHTSHENGNK KSQEFLDTLKQKLALK AHRLKRYNNSQKRK NENTIFLTNEKLFYRN LIKPKTDRDNSNIDIP TAEQLEMYWARLWE NSAKHNDKANWITE EKERWDTIEEMQFD

		ACACATCACACGAAAATGGCAATAAAAAATCTCAAGAATTTTTAGACAC	AGCGGCG		DVTEEEITTITARLHN
		ACTGAAACAAAAATTAGCTCTAAAGCCCACAGACTAAAAAGATATAAC	GAGAGTTG		WKSPGIDKIHNFWFK
		AACTCACAGAAACGAAAGAATGAGAAACACTATATTTCTAACAAATGAAA	AGCAGGC		KLICLHKTIAKNLTDIIS
		AACTATTCTACAGAAACCTAATAAAGCCGAAAACTGATCGAGACAATAG	GGGCTCTC		GNQSIPEFIATGITYM
		TAATATAGATATACCAACAGCAGAACAATTGGAAATGTATTGGGCTAGG	GCCGGGG		IPKGDFSIEASQYRPIT
		TTATGGGAAAATAGTGCAAAACATAATGACAAAGCAAATTGGATTACTG	ATTTACAA		CLPTIYKILTTVITKKIN
		AAGAAAAAGAAAGATGGGATACAATAGAAGAAATGCAATTTGATGACG	CCCGAAAA		SHIEHNNILAEEQKG
		TGACAGAGGAAGAAATAACTACTATAACAGCTAGACTACATAACTGGAA	TGCTACAG		CRRGHMGCKEQLIID
		ATCCCCAGGTATAGATAAAATCCATAATTTTTGGTTTAAAAAACTAATTT	CCGCCAAC		STIMKHATTKNRNLH
		GCTTACACAAAACAATAGCCAAAAATCTAACAGATATTATCTCTGGAAAT	AGAAATCA		CTYIDYKKAFDSIPHS
		CAAAGTATTCCCGAATTCATAGCGACAGGAAATCACTTATATGATACCAA	CGATGTAG		WLIQVLEIYKINPIIISF
		AAGGTGACTTCTCTATAGAAGCATCCCAATATCGACCAATTACATGCCTT	AAAGTAG		LRNIMTHWQTTLKLK
		CCGACTATTTACAAAAATTTTAACAACAGTTATTACAAAGAAAATAAATTC	GAGCAATA		NPPNFVTTRQJAIKKG
		ACATATAGAACACAATAATATCTTAGCTGAAGAACAGAAAGGGTGTAG	GCCCATGT		IYQGDSLSPLWFCLAL
		ACGAGGCCACATGGGCTGCAAGGAACAGCTAATTATAGACTCAACCATC	GAACCTTA		NPLSHQLHNDRAGY
		ATGAAACACGCCACCACAAAAAATAGAAATTTACACTGTACATATATTG	CAGCCCGA		RIKQQDNTETIISHLIY
		ACTACAAAAAAGCTTTTGATAGCATCCCACATTCATGGCTGATCCAAGTC	GTACCGGT		MDDIKLYAKNDKEM
		CTAGAAATCTACAAAATTAACCCTATAATAATAAGCTTCCTACGCAATAT	TCATACAA		KKLIDTTTIFSNDISM
		CATGACACATTGGCAAACCACACTTAAATTAAAAAACCCTCCTAATTTTG	CCCCTCGG		QFGLDKCKTVHIIKGK
		TAACAACACGACAAATAGCCATAAAAAAGGGTATTTACCAAGGTGATTC	TACAATCA		VQPGDYTIDDTNTIT
		TCTCAGCCCTTTGTGGTTTTGCCTCGCCTTGAACCCACTATCCCATCAATT	TCACCATC		AMEPSDLYKYLGFQQ
		GCATAATGACCGGGCGGGATACCGCATTAAACAACAAGATAACACCGA	ATCCTCGG		LKGLDHITIKQSLTSEY
		AACAATAATATCACACCTGATTTATATGGACGACATTAAATTATACGCAA	GTCATAGA		KKRINAICKTKLSGKH
		AAAATGACAAAGAAATGAAAAAAGTTAATAGATACTACCACGATATTCAG	GGCTCACC		LIKALNTYAIPILTYSF
		CAACGACATCAGTATGCAATTTGGACTTGATAAATGTAAAACCGTACAT	AACGTCGA		GIIKWSKTDIEQIERIT
		ATAATAAAAGGAAAAGTCCAACCCGGTGATTATACAATAGATGACACAA	CTATGG		RTTLTKHNNLHPKSAI
		ACACAATAACGGCGATGGAACCAAGTGACCTTTATAAATATCTGGGCTT	(SEQ ID		ERLTIKRQDGGRGMI
		TCAGCAGCTCAAAGGACTCGATCATATAACAATAAAGCAATCATTAACTT	NO: 1250)		DIWHLWRKQIHSLKT
		CAGAGTACAAAAAAACGTATCAATGCCATTTGCAAAACGAAATTATCTGG			FFYIKSDLSEIHRAIAQ
		AAAACATCTTATAAAAGCACTGAACACCTATGCAATACCCATTCTAACCT			NDNNYTPLNLKQKEL
		ACTCATTTGGAATAATAAAATGGAGCAAAACTGACATAGAACAAATAGA			IDNTENLRNRNPQKD
		ACGCATAACAAGGACTACATTAACAAAACACAATAATCTTCATCCAAAAT			MEENWKKKALHGR
		CTGCAATAGAAAGATTGACAATTAAAAGACAAGACGGGGGTAGAGGCA			HPHDLSQSHIDSKAS
		TGATAGATATTTGGCATCTATGGCGTAAACAAATACACAGCTTAAAAAC			NMWLKTGSLFPETE
		ATTTTTCTACATAAAATCAGATTTAAGTGAAATTCACAGAGCCATAGCAC			GFLIAIQDQVINTKNY
		AAAATGATAACAACTACACACCGCTAAATCTCAAACAAAAAGAACTAAT			RKYIIKDPTIRDDKCR
		AGATAATACAGAAAACCTAAGAAATAGAAACCCACAAAAAGACATGGA			KCNTQPETIQHITGA
		AGAAAACTGGAAGAAAAAAGCGCTACATGGACGACACCCTCATGACCT			CSTLTQTDYTHRHNQ

			AAGCCAATCTCACATAGACAGCAAGGCATCAAACATGTGGCTCAAAACA GGAAGTCTGTTCCCCGAAACGGAAGGATTTTTAATTGCCATACAGGACC AAGTAATAAACACAAAAAATTACAGAAAATATATTATTAAAGATCCCAC TATTAGAGACGATAAATGCCGCAAATGCAACACCCAGCCAGAAACCATA CAGCACATAACTGGAGCATGTTCAACCCTTACACAGACAGATTACACTC ACAGACACAACCAACTTGCTAATATTATCCATCAACAACTGGCTCTCAAA CATAAATTGATACAAAATACAAACACACCGTACTACAATTATAAACCACA AACCGTTCTTGAAAATGACTCCTGTAAACTTTATTATGATCGCGCTATTCT TACCGATAGGACGATTCACTACAATAGACCGGATATCACTTTACAAGAT AAAAACAATAAAGTCACCTACATTATTGACATTGCAGTCCCGAATACCCA CAACATTCAAAAAACGTTTACAGAAAAGATGACACAAAATACACAGAACTT AAAGAAGAAATAGTTAGAATTTGGAAACAGAAAAAAGCATACATAGTC CCAATAATAATCTCAACCACTGGAGTTGTCCCAAACCACATCCACAACAG CTTAAAGCTTCTAGATTTAAAAGATAACATATTTATTTCACTACAAAAGG CAGCTATCCTAAATACATGCAGAATAGTGAGAAAGTTCATGCAGCTTGA AGAAAACCAAACTTACTACACGCAATAAAAACTAGCATAATTATTAACTC ATAACTAATGTATATTACTTGGCCAAAAGCCCGTATATACAGTTCCACCG GCTCTGTCGACAGACTGAACTGAGAAAGGGGAAACATATGGAAATAAT AATAA (SEQ ID NO: 1127)	GAAGGCTA ACCACAAT A (SEQ ID NO: 1251)		LANIIHQQLALKHKLI QNTNTPYYNYKPQT VLENDSCKLYYDRAIL TDRTIHYNRPDITLQD KNNKVTYIIDIAVPNT HNIQKTFTEKMTKYT ELKEEIVRIWKQKKAY IVPIIISTTGVVPNHIH NSLKLLDLKDNIFISLQ KAAILNTCRIVRKFM QLEENQTYYTQ (SEQ ID NO: 1495)
R4	DongAG	AB097127	Anopheles gambiae	GAAGGCTAACCACAATAATGGAGACACGATCTATGAGAAAAAGGACCA CACGATTGCCCGAGGAGGGAGCCCCTACTGGAGCTGGGCCTGGGACGG GAGACAGAGCAAGCATCCAGCGGCTGGAAGATGAAATGGTGCAAGAG CGTTCTTTCAGTCAACGGGCTCTGCCCGTACCACGAACGCAAAACAGAA ACGGCAGCCCTATAAATCACCAAGGCAATGCTGCATCTGCTAATGTTGC TGTGGCTGATAGACAACAGTCACTCATTTTGGCAGGAGGCCGACGGCA GAGGATTATGTGGACGAGAGAGATGAACCATTACGTGATCCGTTGCTA CTATGTTTACACGAGGATGGAGACGGACATGCCCGGCAGAGTGAAGAT GCTGGGTATGTTCAATGACCGTTTCCCTCGGTTTGCGCATCAGCTTGACC TGAGCAAGTTGTATATACGGCAGCGAGCTATTATTTTGCCTGAGGAACT GGAGTTCATCAAGCTGGAGGTGCGGAGGGAATTTGGAGAGGAAGAGG CAGGCTGGCGCGAGTCGAGTAGGATTTCTGCTAGGCTTAATACAATAGA CCAAAATACATCAAGGGCGAGTGAGGATCGTGATTTGGATGAACCCAC CGCTCCAGGATTGTCAGTGGATATCCAACACCAGATGGCCACAGCTGTT ACGCAGTTCCATGGGACAGACCCCTTGTCTCGTCACCGACTACCAAAAC TGCATTATTCTTACCGCCTGAAAACAGCAGTAAGCATCATAAACCAAGAT GTTCTACCTCAGTATTTGGATAGCGTAGGGAGCATTGAGGATCTGCAGT TAATTGTGTATTCGGCTGCGGTGGCTGTTGTACGAACGCTATGGTTGCG GACCTATCCGCAAGGAGACAGCGAAGGTCGACCATGCTCCAAGGCTGA	TAACATC CGGTGC AAACTC ATTAAC ATTAAG AAAAGA GAGAGG AGAAAT GAGAAT GAGATT CATTCAC CTTTGG CATTTGA ATAGCC CGGGGT AGGTGA AAAGTT CCCAGC ATATTGC	METRSMRKRTTRLPE EGAPTGAGPGTGDR ASIQRLEDEMVQERS FSQRALPVPRTQNRN GSPINHQGNAASAN VAVADRQQSLILAGG RRQRIMWTREMNH YVIRCYYVYTRMETD MPGRVKMLGMFND RFPRFAHQLDLSKLYI RQRAIILPEELEFIKLE VRREFGEEEAGWRES SRISARLNTIDQNTSR ASEDRDLDEPTAPGL SVDIQHQMATAVTQ FHGTDPLSRHRLPKL HYSYRLKTAVSIINQD VLPQYLDSVGSIEDLQ LIVYSAAVAVVRTLW

			TGAGAA	LRTYPQGDSEGRPCS
			GTGACA	KAEKPAWMRRLENR
			AAATTC	INATRTKIGRMQEYQ
			GGATAA	RGNSSMKVVRQIAE
			TAATAAT	MVKPKELRDLTDANI
			AATAAT	TEVLDIHLQRLSALAK
			AATAAT	RLRRYAECSKRKEQN
			AATAAT	RMFNINEREFYNWIR
			AATAAT	NDKPNFREGLPDIGD
			AATAAT	FTQFWANLCEKPVQ
			AATAAT	HNSEGMRLAEDERF
			AATAAT	SDGIEDMPVLVVNA
			ATGCAT	QDIREATQYTRNGAA
			AATAAT	PGPDFVYNFWYKKLI
			A (SEQ	TIHEQIAACFNTVLED
			ID NO:	SRKLPKFITGGVTYFL
			1374)	PKDQNTKNPAKYRPL
				TCLSNLNKVLSSVITQ
				KVKDHCDTNNVMTE
				EQTGRRKNTQGCKD
				QVIIDAVIVGQAAKK
				QRNLDMAYIDYKKAY
				DSVPHSYLLKVLQLYK
				VDGNVIKLMQHAM
				GMWSTSLHVTDGKV
				VLRSRSLNIRRGIFQG
				DTFSTLWFCLAMNPL
				SRTLNQQCNFGYLLK
				SEEISTRITHTFFMDD
				LKLFAETVQKMHHLL
				KNVQGFSNDIKMEF
				GIGKCRSIHLHRGQV
				LDADSFRANEQEEIR
				HMVQGETYKFLGFL
				QLRGIHYAVIKKELQD
				KFLHRVSCILKSFLSV
				GNKVKAINTFAVALL
				TYSFGVMKWSNTDL

			CAGATCACGGATTAAGCGCCCTGCATCTAGCGCAGCAAGATTACCAGCT GAATTGCAACATAAAAACCGTCGATGGAAAAGGCGCAACGTGGAAACA GAAGGAGTTACATGGGACGCACACCCATCAACTGAATCTGGAACATATC GACAAAGTGTCATCTAGCACTTGGCTTGTGAGGTGTGACCTTTTCTGTG AGACAGAAGGTTTCATGGTAGCCATCCAAGACCGGGTAATTGCGACGT GGAACTATCGGCGGTGTATATTGCGTGAAGACGTGGAGGACCGATGCA GAAAGTGCAACTCAGGAGGAGAATCGATTGAGCATGTCATTGCCGGCT GTCCAGTGCTAGCTGGGTCAGCGTATCTCGATCGCCACAACGACGTTGC CAAAATTGTTCACCAGCAGCTTGCACTGAGGCACAAGTTGGTAGAGCGA TTTTTACCCTGCTACCGATACCTCCCAGATCCGGTCCAGGAAAATGATTG CATAAAGCTGTATTGGGATCGCGAAATTATAACGGACATCCTCATCCGT GCCAATAGGCCAGACATCTTAGTCTACGAGAAAAGAAAGAAACGAGCG ACCATCGACATCGACATTGCTGTAACGTTAGACCATAATGTTCAGACAAC ATTTTCCACCAAGGTGATGAAGTATCATGATCTGGCAGAGGAGTTGAAG CAGACGTGGTATCTGGAGGATATCCGCATTGTTCCGGTAATCATCTCGG CGACCGGAATTGTACCTATGGCCCTCTTACGTTCCCTGGACGAGCTCGA ACTGCAGAGAGAACTACCCAGGATTCAGAAGGCGGTGATTCTTCGAAC ATGTAGCACTTTAAGAAGGTTCCTGAATCCCTATAACTAACATCCGGTGC AAACTCATTAACATTAAGAAAAGAGAGAGGAGAAATGAGAATGAGATT CATTCACCTTTGGCATTTGAATAGCCCGGGGTAGGTGAAAAGTTCCCAG CATATTGCTGAGAAGTGACAAAATTCGGATAATAATAATAATAATAATA ATAATAATAATAATAATAATAATAATAATATGCATAATAATA (SEQ ID NO: 1128)			EALERTIRVVSTKHQ MRHPKASVERVILPR KIGGVGIIDIQALCISQ IHQLRSYFVESQNRH ELYRTVYKADHGLSA LHLAQQDYQLNCNIK TVDGKGATWKQKEL HGTHTHQLNLEHIDK VSSSTWLVRCDLFCE TEGFMVAIQDRVIAT WNYRRCILREDVEDR CRKCNSGGESIEHVIA GCPVLAGSAYLDRHN DVAKIVHQQLALRHK LVERFLPCYRYLPDPV QENDCIKLYWDREIIT DILIRANRPDILVYEKR KKRATIDIDIAVTLDH NVQTTFSTKVMKYH DLAEELKQTWYLEDI RIVPVIISATGIVPMA LLRSLDELELQRELPRI QKAVILRTCSTLRRFL NPYN (SEQ ID NO: 1496)
R4	R4-1_BX	CADV010081075	Bursaphelenchus xylophilus	GGGATCCTGGGTTCCTACTACCTCGCTCCACCTCCTCGCGATGGATCCTG GGGAAGTCTCCGGACTGAGCTAAGAGAGCGTTAAAGTAGAGGGTGAC GGCGTAAGTACCTCCAAGTTGCGGTGGAGCGGAACATCTACTCTTCGGA GAGAGGGGAAGCTCTATGGCGGCGTTAGAAAGGTTGGACTACGGCAA CGCCAGGGAGATGGGGAAGGTTCATCAGGTGATACTAGTTCGCTACTG TCATTCGATGTATCCGGAACCATACTCGCCAAGTTGTGAACTATGTGAA AGTCTGGATCCAATCCAAGACCACGGGGCGCAATTAAAAGGTGTGAAG CAGCTTGCTGGTGATCACCACGGTGGTACCTACACCGGCGGGGGAACA TCTTTAATGCCGAGATGACCGCGCGAATAATAGGGAGCCGGCAAAAGC CGAGCGTAAGTGGAAAGGATACAGAAATTTGCTCAAAAGTATACCAAC CCGCCACTTATAAAAGACCGAGAAAAGGGTACCACTAGAGCTCATTATA AAACATCTATGACATGTAATAATGCTGTGGTCTTTCCGCCAGCCGATGGT AACCCCGCTGGGACTGCGGATAGGAATTTTGCTATTAGGTTCCCCTCCA	GGGATCCT GGGTTCCT ACTACCTC GCTCCACC TCCTCGCG ATGGATCC TGGGGAA GTCTCCGG ACTGAGCT AAGAGAG CGTTAAAG TAGAGGGT GACGGCGT	TAAGAA AAGCAT GAAATA ATAAGA AATCAG ATAAGA ATAACA AGAATA CTAATA AGTATA TCATGTA ACTATG ACAAAA	MTCNNAVVFPPADG NPAGTADRNFAIRFP SSEPPGPSGIRPSEPL DGRTGIGDVEHAQA GNGGFLVDVLEYKEA HRYGSKCEFCYVQTK GTVCSKPRTDAWLK CEILFLLHHAYTANQ NKSIELAESAFRRAGI TRRSKATIAKRWSLIQ RGKGTDYKEYWDEY FEKFRYECNPTPIVRR KRNRLAAGLQSPSSV

		GTGAACCTCCAGGACCCTCTGGAATACGACCTTCTGAGCCTCTGGACGG	AAGTACCT	AGAACG	PNGYEFERKRTCETPL
		TAGAACCGGGATTGGCGACGTTGAGCACGCCCAGGCGGGGAAATGGAG	CCAAGTTG	CACCAA	DTKASSLPLICNLLTGI
		GATTCCTGGTTGATGTTCTGGAATATAAAGAAGCACATCGCTACGGCTC	CGGTGGA	TAAGAA	VGVENVEENMSVEC
		CAAGTGTGAGTTTTGTTATGTGCAAACCAAAGGTACCGTGTGCTCTAAG	GCGGAAC	CATGCTT	TEPKELSGTANSSVP
		CCTCGGACCGATGCGTGGTTGAAATGCGAGATTCTGTTCTTGCTTCATCA	ATCTACTC	GAGTGG	GLAEGVYERRHNNV
		TGCCTATACGGCTAATCAGAACAAGAGTATTGAACTCGCCGAGAGTGCT	TTCGGAGA	CCAGCT	NEPAAGCPQDVPVA
		TTTCGGCGAGCAGGAATTACCCGACGTAGCAAGGCTACCATCGCGAAA	GAGGGGA	CTGCAG	NNLIDSPTTNDRLEA
		AGATGGTCCTTAATACAGCGAGGTAAAGGCACCGACTATAAGGAATACT	AGCTCTAT	GCAAAA	EFKAQLDRAERSYMR
		GGGACGAGTACTTCGAAAAATTCAGATACGAGTGCAACCCCACACCTAT	GGCGGCG	GTCGAA	RRLPRLKNLSPDERM
		AGTACGGCGCAAACGAAACAGGTTAGCCGCTGGGTTACAGTCGCCCAG	TTAGAAAG	TTTGGA	WIGTTVERLRLETVSE
		CAGCGTGCCGAACGGATATGAATTTGAGCGGAAAAGGACTTGTGAAAC	GTTGGACT	ACAGCC	PVCEQWRLANAGLY
		GCCTCTTGATACCAAGGCTTCAAGTCTTCCACTAATCTGTAACTTGCTAA	ACGGCAAC	GGTAAT	AAIRSIAVMRPLDAA
		CAGGAATTGTCGGAGTTGAGAATGTTGAGGAAAATATGAGTGTAGAAT	GCCAGGG	GGAAGA	REAHKTWLLNMKM
		GCACTGAGCCGAAAGAGCTAAGTGGAACTGCGAACTCTTCTGTTCCGG	AGATGGG	CCTGCA	TERKLRQQIGWVETT
		GTTTGGCAGAAGGAGTATATGAAAGACGACATAACAATGTCAACGAGC	GAAGGTTC	ACAAAC	RRTKNEARTERQEIV
		CAGCGGCTGGGTGTCCCCAAGATGTCCCAGTCGCAAACAACTTGATCGA	ATCAGGTG	GTGGGG	YRKVAKLRRERFPEM
		TTCCCCAACTACTAACGATCGATTAGAAGCTGAGTTTAAGGCACAGCTG	ATACTAGT	TAGCAG	DLDSVSVHLKRKLELL
		GATAGAGCGGAACGGTCTTATATGAGACGAAGGCTACCTAGGCTGAAG	TCGCTACT	GCAATA	KGRIQVRTAERLRRD
		AACCTATCCCCTGACGAAAGGGATGTGGATTGGAACAACAGTCGAGCGC	GTCATTCG	TGTAACT	TREAAGPYGKTALRG
		CTAAGACTCGAGACTGTGAGCGAACCTGTATGTGAACAGTGGCGGCTT	ATGTATCC	ATGACA	QGFAPNVKDATQY
		GCCAACGCTGGGCTATATGCAGCTATCCGGTCAATCGCGGTAATGCGCC	GGAACCAT	GACCAA	WSGLAQPSGQKCSE
		CCCTGGACGCCGCCCGCGAAGCCCATAAAACTTGGCTTCTAAATATGAA	ACTCGCCA	AACTCC	NSAILSDWKELVECN
		AATGACGGAAAGGAAACTAAGACAGCAGATTGGCTGGGTCGAGACGA	AGTTGTGA	GAAACT	LSSLPDQMEPLVVQG
		CGCGCCGTACTAAGAACGAGGCACGTACTGAAAGACAAGAAATTGTCT	ACTATGTG	CTGGTA	ISRASPWKSPGPDGI
		ATAGAAAAGTGGCCAAACTGCGAAGGGAAAGATTCCCCGAGATGGACC	AAAGTCTG	ATGAGC	FNYYWRQDFIVDWL
		TGGACAGTGTGTCAGTCCACCTCAAACGCAAGCTTGAGTTACTTAAAGG	GATCCAAT	CCGTGC	KQLMLDSLRTGHYP
		GCGGATTCAAGTGAGAACCGCAGAGAGGTTGCGAAGGGATACCAGAG	CCAAGACC	CCCCCA	WKLSSGRTVLLYKDG
		AGGCCGCTGGTCCATATGGTAAGACTGCTCTTCGGGGTCAAGGGTTTGC	ACGGGGC	AGCATG	DPTKAENYRPITCLN
		CCCGAATGTAAAAGACGCGACTCAGTACTGGTCTGGATTGGCCCAGCCA	GCAATTAA	TGGTCT	GCFKMINSVVSEVILK
		TCCGGACAAAAATGTTCTGAGAACTCGGCGATTCTCTCAGATTGGAAAG	AAGGTGTG	CGTTCG	RVENTIALPIEQMAL
		AACTCGTCGAGTGTAATTTAAGCAGCTTGCCGGATCAGATGGAGCCGCT	AAGCAGCT	ATGTAG	RRKVWACVESQIWD
		GGTAGTCCAAGGCATAAGTCGGGCTAGTCCATGGAAATCTCCTGGACC	TGCTGGTG	TTAGGA	QIKQRKLSDRTQKCK
		GGACGGTATTTTTAACTACTATTGGCGACAAGACTTTATTGTGGACTGGT	ATCACCAC	ACAGTT	VAWVDFSKAYDSLN
		TAAAGCAGCTAATGCTGGATTCCTTGAGGACCGGGCACTATCCGTGGAA	GGTGGTAC	CTCTTTA	HDAIKFVIGVLKLPTG
		ACTTTCAAGTGGACGCACGGTGCTCCTGTACAAAGACGGGGATCCGACC	CTACACCG	ACCCGT	INNYLLDSMQNWST
		AAAGCCGAGAACTACCGACCTATCACCTGCCTAAACGGATGCTTCAAAA	GCGGGGG	GATGAT	HLELKSSGKVVRGPS
		TGATCAACTCGGTTGTTTCTGAAGTTATCCTGAAACGCGTCGAAAATAC	AACATCTT	TACGCC	YPIKRGVLQGDSLSPT
		GATAGCTCTCCCGATAGAGCAAATGGCTTTACGACGTAAGGTATGGGCA	TAATGCCG	CTGTCTT	LFVVVTSIIVRHIKTIES

	TGTGTTGAGTCTCAGATATGGGATCAGATCAAACAGCGGAAGCTGTCTG	AGATGACC	AAATGG	SDIQMYMDDIKLYGK
	ATCGAACACAGAAATGCAAGGTAGCGTGGGTAGATTTCTCAAAGGCTTA	GCGCGAAT	CAGGTG	DQETLTRLIKELQTVS
	TGACAGTCTTAACCATGATGCTATAAAGTTCGTCATAGGAGTTCTAAAG	AATAGGG	CCACCA	NKLGLCMNLKKCAIL
	CTTCCGACAGGTATTAACAACTACCTGCTTGACAGTATGCAGAATTGGA	AGCCGGCA	AATACC	GDDLPEEINGIEHLKE
	GCACGCACCTCGAGCTTAAAAGCTCTGGAAAGGTAGTACGAGGCCCTA	AAAGCCGA	GAACAC	SYKYLGVPQREITQV
	GTTACCCGATTAAAAGGGGCGTTCTTCAAGGCGATAGTCTAAGCCCAAC	GCGTAAGT	TCGTTG	RATMAALEKKILTEV
	ACTGTTTGTCGTGGTGACCAGTATAATAGTACGTCATATCAAGACGATC	GGAAAGG	AGGTAT	DTSLGAAELSYRQRIS
	GAAAGTAGTGATATACAAATGTATATGGATGATATCAAGTTGTATGGCA	ATACAGAA	GGTGGT	RVNSKIAPLVRFVVQ
	AAGATCAGGAAACTCTGACACGGCTAATAAAAGAGCTGCAGACGGTAT	ATTTGCTC	CCGAAT	SMLVTPRDVLKVYNR
	CGAACAAGTTAGGTCTATGTATGAACCTTAAGAAATGCGCAATTCTCGG	AAAAGTAT	GTGAAG	LGGIDVEIRRRLVKYEI
	CGACGACCTCCCAGAGGAAATCAACGGGATTGAGCACCTTAAGGAATC	ACCAACCC	CTGGGA	RYKKSNVARLYLDRK
	ATATAAGTACTTAGGTGTTCCCCAACGGGAAATCACCCAAGTGCGTGCC	GCCACTTA	GTACAA	VGGIGFVNLCRIMVE
	ACCATGGCCGCACTTGAAAAGAAAATACTAACTGAAGTAGATACTAGTC	TAAAAGAC	TTTGGTA	AVAARAVYCRLAPSF
	TTGGAGCAGCCGAACTAAGTTACCGCCAACGGATATCGCGTGTTAACTC	CGAGAAA	CGAGAG	NEFQDFLAEQNTSPI
	TAAGATCGCGCCCGTTGGTGAGGTTCGTTGTGCAGAGTATGCTGGTCACG	AGGGTACC	CACCAG	TAAQTILDKCGINIELS
	CCACGCGATGTGCTTAAAGTGTACAATAGGCTTGGAGGCATTGATGTCG	ACTAGAGC	CGCCCC	TSTLGDVKKIVRNHY
	AAATTCGAAGAAGACTCGTCAAGTACGAGATCAGATACAAGAAAAGTA	TCATTATA	CGATCT	HELWLTAWKNTGLY
	ATGTAGCCCGGCTCTACCTTGATAGAAAAGTTGGCGGCATTGGTTTCGT	AAACATCT	AAGTGA	KRWENDHVDIKRSSL
	AAACCTCTGCCGGATAATGGTCGAGGCGGTTGCAGCCAGAGCTGTCTAC	(SEQ ID	TGACGC	WINRGNLSANNARI
	TGCAGACTTGCTCCCTCGTTTAATGAGTTCCAAGACTTCCTTGCAGAACA	NO: 1252)	ATGCGT	GIGIQDNSIFCRGFVG
	GAATACCAGCCCAATCACTGCTGCACAGACAATTTTAGATAAGTGCGGC		CGGAAC	NKCDTKYCRLCGDGI
	ATTAATATTGAGCTGAGTACATCTACTCTGGGAGATGTAAAGAAAATAG		AATGAA	ESVSHIVTGCPTHRT
	TGCGAAACCACTATCACGAGCTTTGGCTCACGGCATGGAAGAACACTGG		GACGGC	NLYIERHDCVARNVY
	GCTCTATAAACGTTGGGAGAACGACCATGTAGACATTAAAAGATCTAGT		TGGCAA	AYLAIRYGIPVPHYTQ
	TTATGGATCAATCGAGGTAACCTAAGTGCGAATAACGCTCGTATTGGAA		ACATTCA	RVKTIEKNGDQSVEL
	TTGGCATTCAAGACAACAGTATCTTTTGTCGAGGCTTTGTTGGGAACAA		GGAGTC	YWNYKFPCTRALEAC
	GTGTGATACGAAATATTGCCGATTATGTGGGGATGGCATAGAATCCGTA		GCAAA	RPDIVLIDKVSKRTHII
	AGCCATATTGTAACTGGTTGTCCGACTCACCGTACAAACCTGTACATAGA		(SEQ ID	EVAVSWRGRLQEMIV
	GCGGCATGACTGTGTTGCAAGAAATGTTTACGCATACCTCGCCATAAGG		NO:	DRKVYKYTVNGEYEA
	TATGGAATCCCTGTGCCACACTATACGCAGCGAGTCAAGACAATAGAGA		1375)	DGSSRGWNIVRELN
	AAAACGGCGATCAGTCTGTAGAGCTATACTGGAATTACAAGTTCCCATG			DQYGFPVEVYTLVIG
	CACCCGTGCACTGGAAGCTTGCCGTCCGGATATTGTTCTCATAGATAAG			AGGEILPCTVKDVERL
	GTTAGTAAGCGAACTCATATAATTGAGGTGGCTGTCTCATGGCGGGGAC			TGGAATDNLIERMER
	GGTTACAGGAAATGGTCGATAGGAAAGTGTACAAGTATACGGTAAACG			SAVLGSCRIIKRHLAL
	GAGAATACGAAGCCGATGGCTCTTCCCGTGGGTGGAATATCGTCCGTG			(SEQ ID NO: 1497)
	AACTAAACGATCAGTACGGCTTTCCTGTCGAGGTATATACGCTAGTGAT			
	TGGTGCTGGAGGCGAAATCCTACCATGCACGGTTAAAGATGTAGAACG			
	GCTCACCGGTGGAGCGGCTACTGATAACCTCATTGAGCGTATGGAGCG			

			GAGCGCAGTCCTTGGCTCATGTCGTATCATTAAAAGGCATCTGGCACTC TAAGAAAAGCATGAAATAATAAGAAATCAGATAAGAATAACAAGAATA CTAATAAGTATATCATGTAACTATGACAAAAAGAACGCACCAATAAGAA CATGCTTGAGTGGCCAGCTCTGCAGGCAAAAGTCGAATTTGGAACAGCC GGTAATGGAAGACCTGCAACAAACGTGGGGTAGCAGGCAATATGTAAC TATGACAGACCAAAACTCCGAAACTCTGGTAATGAGCCCGTGCCCCCCA AGCATGTGGTCTCGTTCGATGTAGTTAGGAACAGTTCTCTTTAACCCGTG ATGATTACGCCCTGTCTTAAATGGCAGGTGCCACCAAATACCGAACACT CGTTGAGGTATGGTGGTCCGAATGTGAAGCTGGGAGTACAATTTGGTA CGAGAGCACCAGCGCCCCCGATCTAAGTGATGACGCATGCGTCGGAAC AATGAAGACGGCTGGCAAACATTCAGGAGTCGCAAA (SEQ ID NO: 1129)				MISCDLERETLTQMA
R4	R4-1_HG	.	Heterodera glycines	TGGCGATACTCGGAACCTCCGGGGAGCCTGGTAGGAGTTGGCCTACAG GTCGCGAAAGTCCCTAGGTGCTGCACGGGTTGCGCTAATCCGAGGGCG CTGGGTTACCTTCCCATCGGCCAAAAACGTCTGGGCCTTCTTAGCCGCG GGTCCAGTATTTCTGTTGAGCCTGACAGTTCTTCCCGGATATGGCGAAA GATTACAGGGCGGTATTTCGTGAAACCTAAAAAWGGTCGGGCCGAATG GCACGGACAGACTCACTTCGGAGTGAGCTCGGGGGCATCCGTGTGTTA CCCCGCTGCACCACGCCGAAGCTGTCATAGCGAGCCCGAAGGGGAATG GCCATGGAGACTCCAGCCTCACCCTGTAACTCGAACCTAAGTCCAGGCC CCTTTCTGGTGTTGGCTCGCACTGGTTAGGAACACGACAGTCTCGTGTA ATCCCACACGCACCGAAAGCCCTAGCCCTCGTAGGCGAAGGCTGCGTTG CTGGTTTCAGAACTGTGAGCWAGTGGGGTTCGGATGGCCGAGTGTACC AACCTGCTTGCTAACAGGTAGCATAGAGTAATATGCTAGTAAGCAGGGC ACGAGAAAGGGCCGTAAAGGCTCCGGCTATGAAGGACCTTGCGACCAC GCGTGTGTCTCCCACGTGCGGATTCTTGAAGCCAGAGTCTTGCACTGCG CGCAGGATGGAGCCTGTGCAACTCCTCCCTCGCTGATCGCAGGAGTGG AGGATCACCACTCTTTTTACCTTGCTAGCTTGGGGTACCACCTTGAGCTG GGGCCGGCCTTGCTAGCTTGGGGTACGACCCTTTGAGCTGGGGCTGGC CTTACTAGCTTGGCGCGCCACCCTTGGAGCTGGGGTGGCGCAAGATCAC TTGTATACGGTCTAACCAATACATTTGAAAAGCGATCAAAGCGAAATGA TATCCTGTGATTTGGAAAGAGAAACCTTAACCCAAATGGCTCTCTTTAGA GCCCGCAGTGACAAGACGCCGACCCACGCCGGGATACCAGCCCCTGAT GAGGTCCGAGAGGGAGGGTGCGGCCAAAACCGAACCAATCCCGCGGC CCCCCGTGGTAAGGCGGCAGCGATTCAAAGGCAAAACGGCATAAACTAT TCCTATAAAWGCCTGCGCTCAATCCGGCCTAGTCCGGACCCAGAGGGTT CAGCAGTGGTCAGCCGTGGAGGAGTCCGCCTTAAAGGATGTGGTGGTG CGAAACACAGACGACCGGGGGCTCATTAATTGGGCCAAAGGTGTGCTC	TGGCGATA CTCGGAAC CTCCGGGG AGCCTGGT AGGAGTTG GCCTACAG GTCGCGAA AGTCCCTA GGTGCTGC ACGGGTTG CGCTAATC CGAGGGC GCTGGGTT ACCTTCCC ATCGGCCA AAAACGTC TGGGCCTT CTTAGCCG CGGGTCCA GTATTTCT GTTGAGCC TGACAGTT CTTCCCGG ATATGGCG AAAGATTA CAGGGCG	TGAGGA CTCANA ATTGAC AATACA CCTCAG A (SEQ ID NO: 1376)	LFRARSDKTPTHAGIP APDEVREGGCGQNR TNPAAPRGKAAAIQR QNGITIPIXACAQSGL VRTQRVQQWSAVEE SALKDVVVRNTDDR GLINWAKGVLPEWQ RLCQLNPTMYMARS SPSLSNKWASLRRTH VGPGCPSKEGSGPSQ DLSDVKIQPARLAHD TVAELPQRTVPCGTD GHGVIDSDETETALA EVSRSSPFGEREPLDL GATERITRKRLRNAV RDVVPPRKRRVPSTP SRKEQDLVPEVDGPA PTDVLTHPPTESEPEP MLDPLSLVQLVRPQL GRAMGWAAEEMEL GNVVMDVELKREFN REVRRVGRTPPDQM YKRGAGPPLPQKREP ERVALLEQLIAARVER GINRGLDWFLELNVA

		CCTGAATGGCAAAGATTGTGCCAGCTCAACCCGACCATGTACATGGCTC	GTATTTCG		VFAAARVLSRRERVE
		GATCGTCTCCATCGCTGTCCAATAAGTGGGCATCTCTGCGGCGGACCCA	TGAAACCT		TLADRLHINDSATLSE
		TGTTGGTCCTGGATGCCCTTCGAAAGAAGGCTCCGGGCCCTCACAAGAC	AAAAAWG		VSRRRAKAERKLRCA
		CTTTCAGATGTAAAAATCCAACCAGCCAGGTTGGCCCATGATACGGTGG	GTCGGGCC		REQPWMSRRIRXLG
		CCGAGCTACCTCAACGTACTGTCCCTTGCGGCACAGACGGTCATGGTGT	GAATGGCA		VRVERLKQLADLVRQ
		AATCGATTCCGATGAGACCGAGACTGCCTTAGCWGAGGTATCCCGGTC	CGGACAG		RIAGRGNRSSYEGPR
		GAGTCCCTTTGGAGAACGCGAGCCGCTCGACTTAGGGGCGACAGAGAG	ACTCACTT		RRFRLRPSLRSVTEAP
		GATCACGCGTAAGCGGCTGCGCAACGCCGTTCGCGATGTGGTTCCACCT	CGGAGTG		VNPPLNGNEVYTFW
		CGGAAACGCAGGGTTCCATCCACCCCTTCACGTAAGGAACAAGATTTGG	AGCTCGGG		HSLWAQSLRANTDD
		TTCCAGAGGTGGACGGGCCTGCACCTACGGATGTGTTGACACACCCACC	GGCATCCG		CQLREFKNQLSAARH
		CACAGAATCCGAACCAGAACCGATGCTTGACCCACTCTCACTTGTCCAAC	TGTGTTAC		TDLTSVGTSSLVQMF
		TGGTAAGGCCGCAATTAGGCCGGGCAATGGGATGGGCAGCCGAGGAG	CCCGCTGC		SAALRKMKKGKAPG
		ATGGAACTCGGGAATGTCGTGATGGACGTGGAGCTGAAGAGGGGAGTTC	ACCACGCC		PDGIRAAWWGVFRR
		AACCGGGAGGTGAGGCGTGTCGGAAGAACGCCTCCCGACCAAATGTAT	GAAGCTGT		IAPYVATWVVRVIRG
		AAGCGTGGCGCAGGTCCACCACTCCCACAGAAACGAGAGCCCGAACGA	CATAGCGA		AEPVANWICNGLTVL
		GTAGCCCTACTCGAGCAACTCATTGCGGCGCGCGTGGAGCGTGGCATC	GCCCGAAG		LPKSSDNADPSNYRPI
		AACAGGGGACTTGATTGGTTCCTTGAACTGAATGTCGCGGTCTTTGCAG	GGGAATG		TCLNTCYKLFTAVIAQ
		CCGCTAGGGTTCTGAGCCGCCGGGAGAGAGTGGAAACTCTTGCTGATC	GCCATGGA		ITASYVDVLGGLPRQ
		GGCTGCACATCAACGATTCTGCCACTCTGTCCGAAGTATCCAGACGGCG	GACTCCAG		QVALRKGVWGTSVS
		AGCTAAAGCCGAAAGGAAACTCCGGTGTGCMCGGGAACAACCATGGA	CCTCACCC		LMIDALTVADARRAK
		TGAGCAGGCGGATCAGGAWACTTGGTGTGAGGGTCGAGAGACTCAAA	TGTAACTC		RPLGVCWFDFKKAF
		CAACTAGCTGACCTGGTCCGGCAAAGAATAGCTGGGAGAGGTAATAGA	GAACCTAA		DSVPHNLIRWILRVIG
		TCGTCTTACGAAGGCCCTAGGCGGAGATTCCGACTAAGACCCTCTTTAC	GTCCAGGC		LPPVILSVIVSVMDQ
		GTTCTGTTACTGAAGCACCTGTCAATCCACCCCTTAACGGGAATGAAGTT	CCCTTTCT		WATRLKIGGKVMPK
		TACACGTTCTGGCACAGCCTTTGGGCGCAGAGCCTCCGTGCTAACACGG	GGTGTTGG		TIPVRTGVFQGDTLS
		ATGACTGTCAACTGAGGGAATTCAAAAACCAGCTTTCTGCTGCGCGACA	CTCGCACT		PLLFCLSVWPISFALD
		CACGGATCTGACGAGCGTCGGAACCAGCTCATTGGTTCAGATGTTCAGC	GGTTAGGA		QFPQYQFRCANHLQ
		GCGGCCCTTCGGAAAATGAAGAAGGGCAAAGCTCCCGGTCCTGATGGT	ACACGACA		QGFSVGHVFYMDDL
		ATCCGTGCCGCGTGGTGGGGAGTGTTCCGGCGAATTGCACCGTATGTG	GTCTCGTG		KCYCPDREVLTAVIQ
		GCTACGTGGGTGGTCCGAGTAATACGWGGGGCAGAACCAGTAGCTAA	TAATCCCA		QVQKSASALGLTIHY
		CTGGATCTGCAACGGGCTCACTGTCCTGCTGCCCAAGAGTAGTGACAAT	CACGCACC		KKSAWLDQDGGKSG
		GCCGACCCTAGTAACTACCGGCCAATCACATGCCTAAATACMTGCTACA	GAAAGCCC		KAVLGVPXLVGTYKY
		AACTCTTCACTGCCGTTATCGCACAAATAACCGCAAGTTACGTTGATGTT	TAGCCCTC		LGMHERFMIVSKDSL
		CTTGGTGGCCTACCTAGACAACAAGTAGCACTGCGAAAAGGCGTTTGG	GTAGGCG		ESVRGKFMGRLKTL
		GGCACCTCTGTGTCCCTGATGATCGATGCGCTAACGGTGGCTGACGCGA	AAGGCTGC		WTSKLTFGQAMLGT
		GAAGAGCTAAACGCCCACTTGGAGTCTGCTGGTTCGACTTCAAGAAAGC	GTTGCTGG		KSXCMPVVRYVLQNL
		GTTTGACTCAGTACCGCACAACTTAATTCGCTGGATCCTGAGAGTCATCG	TTTCAGAA		FLPKSEFNQTRLVLRE
		GACTCCCCCCGGTCATCTTGTCTGTCATTGTCTCTGTGATGGATCAATGG	CTGTGAGC		WDRQIRDLLDECNIR

		GCAACGAGGCTMAAGATAGGTGGAAAGGTCATGCCCAAGACCATCCC	WAGTGGG		QVFRSKTELYVSREEG
		GGTGCGAACCGGCGTATTCCAAGGGGACACCCTATCGCCCCTGCTKTTC	GTTCGGAT		GWGLPSMEDALEEE
		TGCCTATCKGTKTGGCCAATMTCGTTCGCCCTAGACCAATTTCCTCAGTA	GGCCGAGT		VVTKLAILVARQETEP
		TCAGTTTCGTTGTGCAAACCATCTTCAGCAGGGCTTTAGTGTCGGACAC	GTACCAAC		LFRVCEALERKRCPTP
		GTGTTTTACATGGACGACCTGAAATGCTATTGCCCAGACAGGGAAGTGT	CTGCTTGC		LSLGLQILKDWGVGV
		TGACAGCAGTTATTCAGCAGGTGCAGAAGAGCGCTTCTGCCCTCGGTCT	TAACAGGT		ELQGRTLLLNGNTVG
		GACGATCCACTACAAGAAATCTGCGTGGCTCGACCAGGACGGCCGGGAA	AGCATAGA		PSQATRKLTGELVLR
		AAGTGGCAAAGCCGTCTTGGGTGTACCAGWACTTGTCGGCACTTACAA	GTAATATG		REAERLSRWRSKVKP
		GTACCTGGGAATGCACGAGAGATTCATGATTGTAAGCAAAGACAGTCTT	CTAGTAAG		GCGMTGGAWRDVP
		GAGTCAGTCAGGGGAAAGTTCATGGGGCGGCTTAAAACCCTGTGGACC	CAGGGCAC		GIDVHLSNRWLVKG
		TCGAAACTGACATTCGGGCAGGCAATGCTTGGCACAAAGAGCMTCTGC	GAGAAAG		ALSPTVVSNSLAIRAN
		ATGCCAGTCGTGCGCTACGTGCTGCAGAATCTATTCCTACCAAAATCGG	GGCCGTAA		TVILRGSGGGYTKGTL
		AATTCAACCAGACCCGTCTGGTCCTGAGAGAGTGGGATCGTCAGATCAG	AGGCTCCG		LRCRGCGNTGETRRH
		AGACCTACTGGATGAATGTAATATCCGGCAGGTTTTCCGAAGCAAGACG	GCTATGAA		IVSACSLGRQKGAAS
		GAGTTATATGTCAGCCGCGAAGAGGGTGGCTGGGGTCTTCCTAGCATG	GGACCTTG		RRHDNVCRILVRAIC
		GAAGACGCCCTCGAGGAAGAGGTAGTCACAAAACTTGCCATACTTGTTG	CGACCACG		HKLNIEPPNSANFPH
		CAAGACAAGAAACTGAGCCGTTGTTCCGGGTGTGTGAAGCTCTTGAAA	CGTGTGTC		VVVLEGSGAKMWID
		GGAAAAGGTGCCCCACCCCCCTGTCCCTAGGATTGCAAATACTGAAGGA	TCCCACGT		FPFVVPHKIRHTRPDI
		CTGGGGAGTTGGGGTAGAGTTACAGGGCCGCACATTGCTGCTGAACGG	GCGGATTC		VVLFEWNGVRRLSVI
		GAACACAGTGGGCCCATCTCAGGCAACCAGAAAACTCACTGGAGAACT	TTGAAGCC		EVAVSDVANMQTQ
		GGTCCTGAGGAGGGAAGCGGAACGGCTCAGTAGATGGAGAAGTAAAG	AGAGTCTT		HIRKSHRYGTNSTEPF
		TGAAACCTGGCTGTGGCATGACTGGAGGGGCATGGCGGGATGTGCCCG	GCACTGCG		VAGVTPTYRNDCLAA
		GAATCGATGTCCATCTTAGCAATAGGTGGCTCGTCAAAGGTGCACTGTC	CGCAGGAT		QLRAKFKAQQVDVIP
		TCCGACCGTTGTCTCAAACAGCCTTGCCATACGAGCAAACACGGTGATT	GGAGCCTG		IIVGTTGETLDGEFGR
		CTGAGAGGGAGTGGCGGGGGCTACACCAAAGGTACCCTCCTCCCGGTGC	TGCAACTC		IRKGLPMLTKLQMPR
		AGAGGATGTGGTAACACGGGAGAAACACGCAGACACATTGTGTCAGCA	CTCCCTCG		LWSEIQRAVILGSYRI
		TGCTCCCTGGGAAGGCAGAAGGGGGCTGCCTCCCGCAGGCACGACAAC	CTGATCGC		LVEHLALPKGGA
		GTCTGCAGGATATTGGTGCGAGCCATCTGCCACAAACTCAACATTGAAC	AGGAGTG		(SEQ ID NO: 1498)
		CGCCCAACTCAGCCAACTTCCCCCATGTCGTTGTGCTGGAAGGCAGTGG	GAGGATCA		
		TGCTAAGATGTGGATAGACTTTCCCTTCGTGGTGCCACATAAGATCCGC	CCACTCTT		
		CACACACGACCTGACATCGTCGTCTTGTTTGAGTGGAACGGGGTCAGAC	TTTACCTT		
		GTCTTAGTGTCATCGAAGTGGCTGTTAGCGATGTGGCAAATATGCAAAC	GCTAGCTT		
		ACAACACATCCGCAAATCCCACAGATATGGCACCAACTCGACTGAACCC	GGGGTACC		
		TTCGTCGCGGGGGTCACGCCTACGTATAGGAATGACTGCCTGGCAGCG	ACCTTGAG		
		CAGTTGCGAGCGAAATTTAAAGCCCAGCAAGTTGATGTGATCCCGATCA	CTGGGGCC		
		TTGTCGGGACCACTGGTGAAACATTGGATGGCGAGTTTGGGCGGATCC	GGCCTTGC		
		GAAAGGGTTTGCCTATGCTCACTAAACTGCAGATGCCAAGGCTGTGGTC	TAGCTTGG		
		AGAAATCCAGCGGGCGGTGATCCTCGGATCATATCGAATTTTGGTTGAA	GGTACGAC		

			CATCTTGCGCTACCAAAGGGGGGCGCTTGAGGACTCANAATTGACAAT ACACCTCAGA (SEQ ID NO: 1130)	CCTTTGAG CTGGGGCT GGCCTTAC TAGCTTGG CGCGCCAC CCTTGGAG CTGGGGT GGCGCAA GATCACTT GTATACGG TCTAACCA ATACATTT GAAAAGC GATCAAAG CGAA (SEQ ID NO: 1253)		
R4	R4-1_PH	.	Parhyale hawaiensis	CCGACCGCCAGCGGGATAACTGGCAAACCCTGTCTCGACCACCGGCCCG	CCGACCGC	TGAGCC	MKMSHNRDTPSNG
				TGAATCCATCGGGGCGTATGAGTCTGACACAGGGGGGTGTTTAAGGTG	CAGCGGG	TTAGGT	VKGTSVRLGTSLVRSP
				ACCCTGTTGCGAGGAAATGCGCAGCAAAAGCCGGATGAGCCTTAGAAC	ATAACTGG	CGCGGG	VGEAGAVRERGTHP
				ATCGAAGGCCAACGACAACCTGCCGAAAGACTGGCGACTAATCCAGTC	CAAACCCT	ATGTGA	SESVSQDSDASVNAT
				AACCTCCTGTAGGTCACCGGCTGGTCATGCTGAATCTCAGCTTAACCAG	GTCTCGAC	CCCGGC	GEGSVREQAPLSPPG
				GCGTACGAACTTACAGTTGGAGGGTCGAGCACCCCTGATGGCTGAAAA	CACCGGCC	GCCAGA	AEEATVPTQRRTRHK
				GGACCATCAAAGTCGAAGGTAGCCAACGAGAAGACCAACGGCTGATTC	CGTGAATC	GTGTAG	WSREDRVVLWECFV
				AGGCGGAAGAGTCAACTCGTTGAATGCGTTCGACAGCTTGGGGTAGAT	CATCGGGG	AGCTGA	ASKREGPGYLKRLKQ
				GGAACTCCTAAGCCCTGAAAGGCAGTCCATCTTCGCAGACGCTAAGGTG	CGTATGAG	ACATCG	LWDERGIPGNFPQA
				CCCCGCCGCCTGAGGGTTATCAGGCCCCGCCGCCTGAGGGTTACCAGGC	TCTGACAC	CTCAACC	SLSGQIRQICSKNLLS
				AACCATGAAAATGAGTCATAATAGAGATACCCCTTCTAACGGCGTGAAG	AGGGGGG	GATCCA	EEERLQIAARMEAQV
				GGTACCTCAGTTAGACTGGGTACCTCGCTCGTCCGGTCTCCTGTGGGTG	TGTTTAAG	ATTTGG	ASPSADEPARQVPTR
				AGGCAGGGGCCGTTCGTGAACGCGGTACTCACCCGTCGGAATCGGTTA	GTGACCCT	GTCGTG	PVTPPRSPPVEPARR
				GCCAGGATTCCGACGCTAGTGTGAATGCTACTGGAGAGGGGGAGTGTGA	GTTGCGAG	AAATCC	PSIPSEETPDLGAVPS
				GGGAGCAGGCTCCCCTTTCTCCGCCGGGGGCGGAGGAGGCTACTGTCC	GAAATGCG	CCTCGAT	EIDSADPNRSPSRGP
				CGACGCAACGCCGCACGCGCCACAAGTGGTCCAGGGAAGACCGAGTTG	CAGCAAAA	AATAAT	RHLPAHNMSQSESE
				TGCTGTGGGAGTGCTTCGTCGCCAGCAAGCGCGAGGGGCCAGGCTATC	GCCGGATG	AATAAT	DDVTDPDVSDQQRS
				TCAAGCGCCTCAAGCAGCTGTGGGACGAGAGAGGAATCCCGGGCAATT	AGCCTTAG	AA (SEQ	DSLEPRDLLRNSSVES
				TCCCACAGGCTTCTCTTAGTGGCCAAATCCGGCAAATCTGCAGCAAAAA	AACATCGA	ID NO:	TPGHPNQELSDTLLS
				CCTCCTCAGCGAAGAGGAACGGCTTCAAATCGCTGCGCGGATGGAAGC	AGGCCAAC	1377)	NYVPSEIDSDDPNQS
				TCAAGTTGCCTCTCCATCCGCTGACGAACCTGCACGCCAGGTTCCAACCC	GACAACCT		PRRGPRHLPAHDMS

		GACCTGTGACCCCGCCTAGGTCCCCACCTGTGGAACCTGCTAGGCGGCC	GCCGAAA		LSDSMDEETEPDLSD
		CTCCATTCCATCGGAGGAAACCCCCGACTTAGGCGCTGTCCCGTCTGAG	GACTGGCG		QQRSDLLELRDLLRN
		ATCGACTCTGCCGACCCCAACCGGTCGCCCTCACGTGGGCCTAGGCACT	ACTAATCC		SSVETTPKGHPSLRHL
		TACCGGCCCACAATATGAGCCAGAGCGAGTCAGAAGACGATGTGACAG	AGTCAACC		PEPKIRAAAFRVNSVI
		ATCCCGACGTGTCTGACCAGCAGCGGAGTGACTCGCTGGAGCCACGGG	TCCTGTAG		GKIHTNNITETNALIK
		ACTTACTCCGAAACTCCAGCGTAGAATCAACGCCAGGTCACCCTAACCA	GTCACCGG		AGADLAVRILEVQPR
		GGAACTTAGTGACACCCTACTGTCGAACTACGTCCCATCTCGAGATCGAC	CTGGTCAT		PQRTQRKKDPPWKH
		TCGGACGACCCCAACCAGTCGCCGCGACGTGGGCCTCGGCACTTACCG	GCTGAATC		RLEKNIAEIRKHLSWI
		GCCCACGATATGAGCCTGAGCGACTCGATGGACGAAGAGACAGAACCT	TCAGCTTA		SEWRRGNLHDEEKK
		GACCTATCCGACCAGCAGCGGAGTGACTTGCTGGAGCTGCGGGACTTA	ACCAGGCG		TLLESRYRCLEVGLTN
		CTCCGGAACTCCAGCGTAGAAACGACGCCAAAAGGTCATCCTAGCCTAC	TACGAACT		LEDTLKQRLSAKRSKV
		GACATCTACCGGAACCGAAGATCCGAGCAGCTGCTTTCCGCGTCAACTC	TACAGTTG		RRFEARVAGFHQNQ
		AGTGATTGGTAAAATCCACACCAACAACATCACTGAGACAAACGCCCTC	GAGGGTC		LFNTNQKRLYQTLRG
		ATCAAAGCTGGAGCAGACCTCGCAGTGAGGATCCTTGAGGTACAACCC	GAGCACCC		EETSSDSPNAEESIRF
		CGACCGCAACGGACCCAGCGGAAAAAGGATCCTCCTTGGAAGCACCGC	CTGATGGC		WSDIWSKEVRHNNT
		CTCGAGAAGAACATTGCTGAGATCAGGAAACATCTCAGCTGGATCTCCG	TGAAAAG		AEWLHDVKEKNVAA
		AATGGCGCAGAGGAAATCTCCACGACGAGGAAAAGAAGACGCTGCTCG	GACCATCA		DPDLTITSQQLKKQLS
		AATCGAGGTACCGGTGCCTCGAAGTCGGCCTAACCAACTTAGAAGATAC	AAGTCGAA		KTKNWKAPGPDMV
		CTTAAAGCAGCGGCTATCTGCTAAAAGGAGCAAAGTTAGACGCTTCGAG	GGTAGCCA		QGYWIKTFTSLHSRIA
		GCCAGAGTAGCTGGATTCCATCAGAACCAGCTATTCAACACCAACCAGA	ACGAGAA		AQLNHCLQRGTVPT
		AACGCCTCTATCAAACCCTTAGAGGAGAAGAGACGAGCTCTGACTCACC	GACCAACG		WMITGKTVLIQKDK
		CAACGCAGAAGAGAGCATCCGATTCTGGAGTGACATCTGGAGCAAAGA	GCTGATTC		AKGTEVSNYRPITCLP
		GGTACGGCACAACAATACCGCAGAATGGCTCCACGACGTCAAGGAGAA	AGGCGGA		LMWKVLTGIIYERVY
		GAACGTCGCAGCAGATCCAGACCTTACAATCACCAGTCAACAGCTCAAA	AGAGTCAA		QHLDSKKLLPDEQKG
		AAGCAACTGAGCAAGACCAAGAACTGGAAAGCACCCGGACCAGATATG	CTCGTTGA		CRRNTRGTKDHLLVD
		GTGCAAGGCTACTGGATCAAAACTTTCACATCGCTGCACTCCAGAATTG	ATGCGTTC		KLLTKDARSKKKNLS
		CTGCACAGCTCAACCACTGCCTGCAACGCGGAACCGTGCCCACCTGGAT	GACAGCTT		MAWVDYKKAFDMV
		GACCACAGGGAAGACGGTCCTAATCCAGAAGGACAAGGCAAAGGGCA	GGGGTAG		PHSWILECLDIYGIAG
		CGGAGGTCAGCAATTACCGCCCAATCACGTGCCTCCCCCTTATGTGGAA	ATGGAACT		NIRNLIATTMPNWKT
		GGTGCTGACGGGCATAATCTACGAGAGAGTCTACCAACATCTAGACTCA	CCTAAGCC		QLTSANKHLGEVSIKR
		AAGAAGCTCCTACCCGACGAACAGAAAGGATGCCGCAGGAACACCAGG	CTGAAAGG		GIFQGDSLSPLLFVLT
		GGAACGAAGGACCATTTGCTCGTCGACAAGCTCCTCACGAAAGATGCCA	CAGTCCAT		MIPLSETLNKAGQGY
		GGAGCAAAAAGAAAAACCTAAGCATGGCTTGGGTTGATTACAAAAAAG	CTTCGCAG		NYSRTMKLNHLLLYM
		CCTTCGACATGGTCCCTCACTCCTGGATCCTGGAATGCCTGGACATCTAC	ACGCTAAG		DDLKLYAKSKDQVEQ
		GGCATTGCCGGCAACATCCGCAATCTGATCGCGACGACCATGCCCAACT	GTGCCCCG		LLNIVHQYSQDIKMQ
		GGAAGACCCAGCTCACAAGCGCCAATAAACACCTCGGCGAGGTGAGCA	CCGCCTGA		FGVSKCGVLNIERGE
		TCAAGCGGGGCATCTTCCAGGGCGACTCCCTATCTCCACTGCTCTTTGTC	GGGTTATC		VTASEGITIEEGTIKDI
		CTGACCATGATCCCCCCTCTCCGAGACCCTGAACAAGGCCGGCCAGGGCT	AGGCCCCG		EEAGYKYLGVMEYNT

			ACAACTACAGTCGTACCATGAAGCTCAACCACCTGCTTTACATGGACGAT	CCGCCTGA		ILHRTMKDSIRKEYLT
			CTGAAGCTCTACGCCAAATCAAAGGACCAAGTAGAGCAGCTCCTTAATA	GGGTTACC		RLRLILKSHLNGGNTI
			TAGTCCATCAGTACTCCCAAGACATCAAGATGCAGTTCGGTGTTAGCAA	AGGCAACC		KAINTWAVPVVRYSA
			ATGCGGAGTCCTTAACATCGAAAGGGGAGAGGTTACTGCGTCCGAGGG	(SEQ ID		GIINWTKKDCTDMDI
			AATCACCATCGAAGAAGGTACCATCAAGGATATCGAGGAGGCAGGTTA	NO: 1254)		KTRKLMTIYRALHPRS
			CAAGTACCTGGGTGTCATGGAGTATAACACCATTCTCCACAGAACGATG			CVDRLYINRREGGRG
			AAAGATTCCATTCGGAAGGAGTACCTCACCCGACTCCGGCTCATCCTGA			LISVEDCVEAEKRALS
			AATCACACCTCAATGGGGGCAACACCATCAAGGCCATCAACACCTGGGC			QHFRESDDPWARCL
			CGTCCCCGTAGTCCGATATTCTGCTGGCATTATTAACTGGACCAAGAAA			VEAKLLKETETADQF
			GACTGCACCGACATGGACATCAAAACCAGGAAGCTCATGACCATCTATC			KERRRLDRTNKWKS
			GAGCACTGCACCCAAGATCGTGCGTAGACAGACTCTACATTAACAGAAG			MKMSGQYLEAVQD
			AGAGGGGGGCCGTGGCCTCATCAGCGTCGAGGACTGCGTCGAAGCGG			KIVPDSWNWLLRGEL
			AAAAACGCGCTCTCTCCCAACACTTCAGAGAGAGCGACGACCCGTGGGC			KRETEGTILAAQEQA
			ACGCTGCTTAGTCGAGGCGAAGTTGCTGAAAGAGACCGAAACGGCTGA			LRTRYIQNKIDKRNVP
			CCAATTCAAAGAGCGACGGCGCCTCGATCGGACTAACAAATGGAAGAG			STCRICRSSDETINHVI
			CATGAAGATGTCCGGCCAGTACCTCGAAGCTGTTCAGGACAAAATCGTA			SECGVLAQKEYKRRH
			CCAGACTCTTGGAACTGGCTTCTGAGGGGTGAGCTGAAACGGGAAACA			DKVARHLHWTLLRIH
			GAAGGCACCATCCTGGCAGCTCAAGAGCAAGCCCTCAGAACCAGGTAC			NFPVSERWYEHEPAP
			ATCCAAAACAAGATCGACAAGAGAAACGTCCCTAGTACCTGTAGAATTT			VVENEAVQIYWDKR
			GCAGGAGCTCGGATGAAACCATCAACCACGTGATAAGCGAATGCGGAG			METDRVLHANRPDI
			TGCTGGCGCAAAAGGAGTACAAGCGCAGGCACGATAAAGTTGCCCGAC			VVKDKQEKSAKLIDISI
			ACCTTCACTGGACCCTCCTGCGGATCCATAACTTCCCGGTCAGTGAGAG			PFDSRIVDKEAEKKEK
			GTGGTACGAGCACGAACCCGCGCCCGTGGTTGAAAACGAAGCCGTGCA			YRDLAIELQRLWQM
			GATCTACTGGGACAAGAGGATGGAAACTGATCGAGTACTTCACGCCAA			KVDVVPVVIGALGA
			CCGACCGGACATCGTCGTCAAAGACAAACAGGAGAAGTCGGCCAAACT			MSKNLKTALRELKCG
			GATCGACATTTCAATCCCCTTCGACTCCAGAATCGTGGACAAGGAAGCG			HLHPGTLQKSALLGT
			GAGAAGAAGGAAAAGTACCGTGACCTTGCCATCGAACTCCAGAGACTG			AHIIRKVL (SEQ ID
			TGGCAGATGAAGGTGGACGTCGTTCCAGTGGTCATCGGAGCTCTGGGG			NO: 1499)
			GCAATGTCCAAGAATCTCAAGACTGCCCTGCGAGAGCTGAAGTGCGGC			
			CATCTCCACCCAGGAACCCTCCAGAAGAGCGCTCTGCTTGGAACAGCGC			
			ACATCATTCGCAAAGTGCTCTGAGCCTTAGGTCGCGGGATGTGACCCGG			
			CGCCAGAGTGTAGAGCTGAACATCGCTCAACCGATCCAATTTGGGTCGT			
			GAAATCCCCTCGATAATAATAATAATAA (SEQ ID NO: 1131)			
R4	R4-2_BX	CADV010090 48	GAGGATCCTGGGTTCCTACTACCCTGCTCCATCTCCTCGCGATGGATCCT	GAGGATCC	TAACAA	MVFNNCKPKHLCPAI
			TGGGGAAGTCTCCGGACTGAGCTAAGAGAGCGTTAAAGTAGAGGGTG	TGGGTTCC	GTGTAA	RPTGQQETNGGSEG
			GCGGCGTAGTGACTTCCAAGTTGCGATGGGGCGGAACATCTACTCTTCT	TACTACCC	TAAAAA	TAEPTAGPSRPAVSE
		Bursaphelenchus xylophilus	GAGAGAGGGAAAGCCCTATGGCGGCGGTAGAAAGGTTGGGCTACGGC	TGCTCCAT	CCACCC	DAAQPVPLFEEGEYI
			AACACTTGCCATGATCAGATTCGATCAAAATTAGCCTCTGGGGCTGGCA	CTCCTCGC	ATGCGT	RAHRDKTCPYCEVL

	ACCCTACAACGGATTGTAAACTGAACTATGCTATGCAAAATGAAAATAA	GATGGATC	AGTAAA	WIGARSSKARSDSW
	AAAATGGGGGCTTTACAATCTAAGGTGTTGGCAGATCACGAAAACTGCC	CTTGGGGA	CCGATC	PLCQILYLMKRNDDL
	CGTCGATGAGGGTGAGATAACATCCTACGGAACAGCCCCTGCTGGACC	AGTCTCCG	AATTATC	RGQRTRYPLLESSLRA
	AAACCAAATCATCCACAAATTGGAGGGTTTTCTTGGTAGTTTCCCTTGGC	GACTGAGC	TAGCAA	AGIARTKFAIIKCIRNV
	ACGTCTTGTTTCATAAGCCAGAATAAAAACGATACCATACAGACATGAG	TAAGAGA	AATCGC	LRDRYVPNGPYSEH
	GCTGGGTACCTGCCAGCCGCGACACGGAAAACCGGTAGGTGCAACCGG	GCGTTAAA	AGGTCA	WKIYRANSGEVPQG
	AGACAGCTAGGGGGAAAGAAAATAGTAAAGTGTCGAAAACAAAACAGG	GTAGAGG	GAAGAC	ATITKGKRSARVAGL
	TAGCCCCTGGCTAGAGGGAATGGGACATTGTCCGATTAGGTTGCCTTGA	GTGGCGG	CAAAGA	PSPSQSGHHTKRIQA
	CCCAATGAAAGCCAACCGGGTTTACATTATCGGTCTACCAAGGGCAATG	CGTAGTGA	ACCGAC	GTGIETETTVTETNTT
	ACCAGAGAAACTGCGACTATGCCGTAACCCTTCGCAGATTGCCGATGAG	CTTCCAAG	CCAGAG	PEVSHEHRDPCGEPE
	AACCATCGATCGTAAGTCGAAGCCAAGCGGATTGACCAGTGGAGGGTT	TTGCGATG	GAATAG	TSAANVDKVTELTED
	ATCCCGACAACAGCATGGTATTCAATAATTGTAAACCTAAACACCTGTGC	GGGCGGA	GACCAG	GSETRGTANVANGG
	CCTGCGATCCGTCCGACAGGGCAACAGGAAACGAACGGTGGTAGCGAG	ACATCTAC	AGCTGA	VSVSDPGRKRQSSSQ
	GGCACTGCTGAGCCTACCGCTGGTCCTTCGAGACCGGCAGTGAGCGAG	TCTTCTGA	AACTCTC	NRGNIETTNPELVG
	GATGCAGCCCAACCGGTGCCACTGTTCGAAGAGGGCGAATATATAAGG	GAGAGGG	AGATAC	MWEDMFGVQLDGA
	GCTCACCGCGATAAGACTTGTCCTTATTGCGAAGTCCTGTGGATAGGAG	AAAGCCCT	GCCAAC	MRTTERPRLPKLKHL
	CGCGGTCTTCAAAGGCCCGCTCCGACAGCTGGCCATTGTGCCAAATCTT	ATGGCGGC	GGTCCT	SEPERLWIRAKLEQA
	GTACTTAATGAAGCGCAACGACGACCTCAGAGGTCAAAGGACGAGGTA	GGTAGAA	AATAAA	WLQCVSYDVEQQW
	TCCACTGCTCGAGAGTTCTTTGAGAGCAGCCGGAATTGCCCGTACTAAG	AGGTTGG	ACGTCG	LNANAVLYAAIRSVA
	TTTGCGATTATCAAGTGTATTCGGAACGTGTTGAGAGACCGGTATGTGC	GCTACGGC	TTAAGT	ASRPCKEAREAQKT
	CCAACGGCCCATACTCTGAGCATTGGAAAATTTACCGCGCGAATAGCGG	AACACTTG	AAAGCA	WLDNKKKDEAKLRRL
	GGAAGTTCCTCAGGGAGCGACCATAACAAAAGGGAAGAGGTCCGCCAG	CCATGATC	TCGTTAA	IGRISSVHSMPKGDR
	AGTTGCGGGGCTTCCCTCCCCAAGCCAGAGCGGTCACCACACGAAACG	AGATTCGA	GTACAA	TPREKKLVKNITKLKN
	AATACAAGCTGGAACTGGAATTGAAACGGAAACTACGGTGACTGAGAC	TCAAAATT	AACAAA	THYPDMDWGGLLN
	TAACACAACACCCGAGGTTTCGCATGAACACAGAGATCCCTGCGGGGA	AGCCTCTG	AGCACT	HFKVKLSQLKEKISVR
	ACCAGAGACGAGTGCTGCTAATGTAGATAAGGTAACGGAATTGACGGA	GGGCTGG	GTAAAC	VAEHKRKVNRNAAG
	AGATGGGTCTGAAACCAGAGGAACTGCGAACGTCGCTAACGGTGGGGT	CAACCCTA	GCGAGG	QYGKSVAGSAGLAP
	TTCTGTTAGCGATCCTGGTCGTAAAAGGCAGAGTTCAAGCCAAAATAGA	CAACGGAT	CCCCCTC	DVVSATAYWSGLAQ
	GGAAATATCGAAACGACTAATCCAGAATTAGTCGGTATGTGGGAAGAC	TGTAAACT	TTTGCCA	PGPKKFKASSPIFQT
	ATGTTTGGAGTGCAGCTTGATGGAGCGATGCGCACAACAGAGCGACCA	GAACTATG	AACTCC	WKDDVAKNLNTEPV
	AGGCTACCCAAACTGAAGCACCTGTCCGAACCCGAGAGGTTATGGATCA	CTATGCAA	GGTAAT	LLYPIIKECIRKPSPWK
	GGGCGAAGTTGGAACAGGCTTGGCTGCAGTGCGTTAGCTATGACGTCG	AATGAAAA	CCTCGTA	APGPDGIYNYYWQQ
	AACAGCAATGGCTTAACGCCAACGCGGTTCTATACGCCGCCATCAGGTC	TAAAAAAT	GTACGG	EFVAQWIQTLVKRTL
	AGTCGCGGCTAGCAGACCTTGTAAAGAGGCCCGCGAAGCTCAAAAAAC	GGGGGCT	TGCTTTT	DIGRFPTALMCGRTV
	ATGGCTGGATAATAAAAAGAAGGATGAGGCGAAATTGAGGCGGCTGAT	TTACAATC	TCCCGCT	LLFKSGDKSMPQNYR
	CGGGAGAATCAGCTCTGTCCACAGTATGCCGAAAGGCGACCGTACGCC	TAAGGTGT	TCACGA	PITCLNGCFKITNAVL
	ACGAGAAAAGAAACTCGTGAAGAATATAACTAAGTTGAAAAACACCCA	TGGCAGAT	ATCAGA	TKVILQRVQDTCALP
	CTATCCAGACATGGATTGGGGTGGACTTCTGAACCATTTTAAGGTAAAG	CACGAAAA	ACGCTG	REQMALKPKVWSC

		TTGTCGCAACTTAAAGAAAAGATTAGCGTGCGGGTGGCGGAGCATAAG	CTGCCCGT	CCAGAT	MEAQLRDQALQSEI
		CGGAAAGTTAATCGTAACGCAGCAGGACAGTACGGCAAAAGTGTGGCC	CGATGAG	CTTGTCC	GDDCKTAWIDFSKAY
		GGCAGCGCAGGGTTAGCGCCGGATGTCGTTTCCGCGACGGCATACTGG	GGTGAGAT	GACATG	DSLDHDALRFVIQTIA
		TCAGGTCTTGCACAACCTGGGCCAAAGAAATTTAAAGCAAGCTCGCCGA	AACATCCT	GGCCGT	LPAGMEEYLLKSLDS
		TCTTCCAGACTTGGAAAGACGACGTAGCGAAGAACTTGAACACTGAGCC	ACGGAACA	AGGGTT	WRTQLVLSDAGKVV
		GGTACTGCTGTATCCGATCATAAAAGAATGCATCAGAAAGCCAAGCCCC	GCCCCTGC	TGGGAG	SGKPYPIKRGVLQGD
		TGGAAAGCACCAGGCCCTGACGGGATATACAACTATTACTGGCAGCAA	TGGACCAA	TACCAG	SLSPALFVLTTSPIVAH
		GAGTTTGTGGCTCAGTGGATCCAAACCCTTGTGAAGAGAACCCTGGATA	ACCAAATC	CGTGGG	LQRTCPTGRIQLYMD
		TAGGGCGGTTCCCGACTGCACTTATGTGTGGGAGGACGGTACTACTCTT	ATCCACAA	GCGGAG	DIKLYGKTESDLCMLI
		CAAAAGCGGTGACAAATCTATGCCTCAGAACTATCGACCCATTACATGC	ATTGGAGG	AGCGTA	KETQRVANKLGLNIN
		CTGAATGGATGCTTTAAGATTACCAATGCGGTATTGACGAAAGTAATCC	GTTTTCTT	CCTGGG	LKKCALFGKSIKQSIA
		TTCAACGGGTTCAAGACACTTGCGCTCTTCCCAGAGAACAAATGGCTTT	GGTAGTTT	TACACT	GFDPLGDRTYKYLGIP
		GAAACCAAAGGTGTGGTCGTGCATGGAAGCGCAGTTGCGCGACCAGGC	CCCTTGGC	GCATAA	QRDVADIKQAYDELK
		ACTTCAGTCTGAGATCGGTGACGATTGTAAGACTGCATGGATCGATTTT	ACGTCTTG	TCGGGT	AKTVQTIGETMACDY
		TCCAAAGCATACGATAGTCTGGACCACGACGCCTTAAGATTTGTAATCC	TTTCATAA	CTCAGA	LTTRQVINRLNSKIPP
		AAACGATTGCACTGCCGGCAGGGATGGAAGAGTACCTTCTGAAATCCCT	GCCAGAAT	AAACCA	VVRFVTQSALCSAPM
		GGATAGTTGGAGAACGCAACTTGTGCTATCTGATGCAGGTAAGGTTGTC	AAAAACGA	CCTATG	TRGLYNKITELDNVSR
		TCCGGAAAGCCCTACCCAATTAAAAGGGAGTGCTCCAAGGAGACAGT	TACCATAC	GTTTATT	AELRKVLIYKATNVSR
		CTGAGTCCCGCGCTATTCGTCTTAACGACGAGCCCAATTGTCGCTCACCT	AGACATGA	ATTCTGT	FYLATKEGGFGYASL
		TCAGAGAACTTGTCCCACGGGTCGTATACAGTTGTATATGGATGATATT	GGCTGGG	CTCCCAT	QQVFVEAVVSRAIYC
		AAACTCTATGGCAAGACAGAGTCGGACCTGTGTATGTTAATAAAAGAAA	TACCTGCC	CTGCAG	LRAPSLCDIREFILSKF
		CTCAACGAGTCGCAAACAAATTGGGCTTAAACATTAATTTGAAAAAATG	AGCCGCGA	GGTAGC	DPVKVARIALARSKID
		TGCCCTCTTTGGTAAAAGTATAAAACAGAGTATTGCCGGGTTTGACCCG	CACGGAAA	TTTTCGT	MDIERMDMASATRT
		CTCGGCGACAGAACATACAAATATCTCGGCATTCCCCAAAGGGATGTTG	ACCGGTAG	TGGGCC	IRQHYQAKWKTLFQ
		CAGATATCAAACAGGGCATATGACGAACTCAAAGCTAAAACGGTGCAGA	GTGCAACC	ATAGGA	QSKLYQKWVQHKIDI
		CCATCGGCGAAACGATGGCGTGTGACTATCTAACAACAAGGCAAGTTAT	GGAGACA	GCCTAG	PNSSRWLQRGEISPR
		CAACAGATTAAATTCGAAGATACCGCCTGTAGTCAGATTTGTCACACAAT	GCTAGGG	GGGCAA	NCRIAVAVQDNTLLC
		CTGCGCTGTGCAGCGCGCCTATGACCAGAGGATTATATAATAAATCAC	GAAAGAA	GAGTGC	RGFVGSKDPNKQCRL
		GGAATTGGATAACGTGAGTAGAGCAGAGTTAAGGAAAGTTCTTATCTAT	AATAGTAA	ATGTAG	CNAGIETASHIVTECS
		AAAGCGACCAATGTCTCTCGGTTCTACCTGGCCACGAAAGAAGGGGGA	AGTGTCGA	TCTTCAA	THRVHMYIERHDSV
		TTCGGCTATGCAAGCCTGCAGCAGGTATTTGTTGAAGCTGTTGTGAGTA	AAACAAAA	CGGCCA	ARNIYAVLAKNCGF
		GAGCGATCTACTGCCTGAGAGCTCCCAGTCTGTGTGACATACGGGAGTT	CAGGTAGC	TGCCAG	WIPHYSQKIPTVKITK
		CATATTGAGTAAGTTCGACCCTGTCAAGGTTGCTCGCATTGCGTTGGCA	CCCTGGCT	GGAAAC	SYELYWNYKFPCTQA
		AGGAGTAAAATTGACATGGACATCGAACGGATGGACATGGCCTCGGCC	AGAGGGA	TTGTGA	LEACRPDIVLIDRAKK
		ACGCGTACCATCAGACAACATTATCAGGCAAAGTGGAAAACTCTGTTCC	ATGGGACA	GAGGTG	RILVVEVAVSYVTRLE
		AACAATCGAAACTTTACCAAAAGTGGGTGCAGCATAAGATCGATATACC	TTGTCCGA	AGGGAT	QMTQRKLYKYGVNG
		CAACTCGAGTAGGTGGTTGCAGAGGGGGGAAATCTCGCCGCGAAACTG	TTAGGTTG	AACTAG	EYQADGETRGWNIC
		TCGAATCGCGGTAGCTGTTCAAGATAACACTTTACTGTGCAGAGGATTT	CCTTGACC	CATCAG	RELVQKYNMRIDLCI

			GTGGGGAGTAAGGACCCAAACAAACAATGTCGGCTATGCAATGCTGGC ATTGAGACCGCTAGTCACATCGTGACTGAATGTAGTACACACCGGGTTC ACATGTATATCGAACGTCACGACTCTGTGGCGCGAAATATATATGCTGT CCTGGCAAAAAATTGCGGTTTTTGGATTCCTCATTACTCACAGAAAATCC CAACAGTAAAGATTACGAAGTCATATGAGTTGTATTGGAATTACAAGTT CCCGTGCACGCAAGCACTGGAGGCTTGTCGTCCCGATATCGTTCTCATT GACAGAGCGAAGAAAAGAATCCTCGTGGTGGAAGTGGCGGTCTCGTAT GTCACACGGCTAGAGCAGATGACGCAGAGGAAACTCTATAAATACGGG GTGAACGGCGAGTACCAGGCTGACGGGGAAACCCGTGGCTGGAACATT TGTCGTGAACTTGTACAAAAGTACAATATGAGAATTGATCTCTGCATAG TCGTAATTGGGGCATGTGGAGAAATACTCCCTTGCATGGTTAAAGAAAT TGAAAAAATATCCAAGGTCTCTGGTCGGCAGCTCCTAGAACGGTGCCAG CGTAGCGCTGTACTTGGAACTGTGCGGACCGTAAGGAGACACTTGGCA AACTAACAAGTGTAATAAAAACCACCCATGCGTAGTAAACCGATCAATT ATCTAGCAAAATCGCAGGTCAGAAGACCAAAGAACCGACCCAGAGGAA TAGGACCAGAGCTGAAACTCTCAGATACGCCAACGGTCCTAATAAAACG TCGTTAAGTAAAGCATCGTTAAGTACAAAACAAAAGCACTGTAAACGCG AGGCCCCCTCTTTGCCAAACTCCGGTAATCCTCGTAGTACGGTGCTTTTT CCCGCTTCACGAATCAGAACGCTGCCAGATCTTGTCCGACATGGGCCGT AGGGTTTGGGAGTACCAGCGTGGGGCGGAGAGCGTACCTGGGTACACT GCATAATCGGGTCTCAGAAAACCACCTATGGTTTATTATTCTGTCTCCCA TCTGCAGGGTAGCTTTTCGTTGGGCCATAGGAGCCTAGGGGCAAGAGT GCATGTAGTCTTCAACGGCCATGCCAGGGAAACTTGTGAGAGGTGAGG GATAACTAGCATCAGATAATATCAGTCATGAAATTAGTAACAACCAACG TTCACCGTCGTTGGCAAAACACCGACTAACGATGCTAGTTAGAAAGAGT CGGGTCTTCCCAAAGTTAGGTGCTTGCACCGAAGCCGATCCGCTCTACC CACAGCTCTGCCCAGCGTT (SEQ ID NO: 1132)	CAATGAAA GCCAACCG GGTTTACA TTATCGGT CTACCAAG GGCAATGA CCAGAGAA ACTGCGAC TATGCCGT AACCCTTC GCAGATTG CCGATGAG AACCATCG ATCGTAAG TCGAAGCC AAGCGGAT TGACCAGT GGAGGGT TATCCCGA CAACAGC (SEQ ID NO: 1255)	ATAATAT CAGTCA TGAAAT TAGTAA CAACCA ACGTTC ACCGTC GTTGGC AAAACA CCGACT AACGAT GCTAGT TAGAAA GAGTCG GGTCTT CCCAAA GTTAGG TGCTTG CACCGA AGCCGA TCCGCTC TACCCAC AGCTCT GCCCAG CGTT (SEQ ID NO: 1378)	VVIGACGEILPCMVK EIEKISKVSGRQLLERC QRSAVLGTVRTVRRH LAN (SEQ ID NO: 1500)
R4	R4-2_HG	ABLA010003389	Heterodera glycines	TGTGGCGATACTCGGAACCTCCGGGGAGCCTGGTAGGAGTTGGCCTAC AGGTCGCGAAAGTCCCTAGGTGCTGCACGGGTTGCGCTAATCCGAGCC CCTTCCGGTGTTGGCTAGCACTGGAGGGGGGGGGGTCCAAAGAATACG TGAAGCGGACCCTGTGTTTTCTCTGACCTCATACGGGTATATCCCGATGA TATCTATACAGTGTTCATTTTCATTTCTTAACCTGCTTTTTCCTCAAGAGA AATATGACCACATTCGTTTGCCATTTCGAAGGTGCGACTTTTGAGGGAG TTCTCAGGGACATCAAGCTGTTCATGGACGGCTTGACGGCTCGTGGCGG CACGCAGCCGCGCGGGCCAGTTAGGGCTGGTGCCAATGAGGCAGCGA AGAAGTCTCTGAAACGCCAAAAGCAGCGGGAGAAGAGAGAAAGCG GGGACGGGTTAGAGATAATAGCATGATTGACGGCGAAGCAATACTTTG	TGTGGCGA TACTCGGA ACCTCCGG GGAGCCTG GTAGGAGT TGGCCTAC AGGTCGCG AAAGTCCC TAGGTGCT GCACGGGT	TAGTCG TAGCCC AGATGT CGACAG CCCCTAC ACCGGT AGATGG AAGGAG GTATGT ATCTAA	MRKSFLQHIPELSSHI AMSVPARNYPKMCS LAQGSSGTLSHNGKG VAMHRCPSDDCAGK DPPQRGSQKGNLRS VRWTPSEEKAVFEY WSRLEQHAMLNGSE ARGTCAITRSQFLIH WDGERESRSLSDGV PEYPMRTERAYYERV

		ACTTCCAGGTACGGAGGACGGCGGCAAAAAGCAAAAAGGGGACGAAA	TGCGCTAA	GCCCAC		RLLRQRGWQWDCA
		CCGATGGCATTGCCAAAGGGGCCGAAACCATTGAGCTGGACGCTGATG	TCCGAGCC	GGCAAG		NECLVIGQCAEPCRK
		GGCCGTCTGCCGCGCAAAAGGAGGGGGGAGCCCCTCCGTCTGGGAAG	CCTTCCGG	CCACCA		PNVVAIKADKGMKR
		AAGGTGCGGCGCCATTCGTCCAACAAACCGGAATAGATCTTGATGATCT	TGTTGGCT	GTGGAA		SLVKGKLLSLPHVMG
		CAGCAATTGATCATATTGATCTCTTTTTCCAGTTGTTGCGATAAATTATCG	AGCACTGG	ACGGTG		EINQVSVQVEVPLPS
		TGCATTATTTCTTCGATTTCTCCAAAGCTTAACTCCTTTCCTAATACCTACT	AGGGGGG	ACCACCT		VPASVPQVEGVESKG
		CATGTATACGTTAGTAGGCATGTTTTATGCAGGTAAAGATAGACCTGGT	GGGGTCCA	CTGTTC		FTETEPSNKPSLEGNP
		GCCCCCGGGCGACTTGGGATGTGGATGATGGTTGGCGGAGAGTTCTGA	AAGAATAC	GCGGAA		AEEGLRKPERVNVPV
		TGACGCCGTAGTTCCGGAGGAACAGTTCCTACTAGTGCCAGCCTGGGCT	GTGAAGC	ATGCCC		HGIISDSERKDLKDRF
		AGTGGAGCTGGTCTGGCGAGTGCTTTGCTCGTCAGATAAGGGGTGGTT	GGACCCTG	CGTAGA		WSAYKTAKRSVGFRP
		GGGGTGGTACCTCTGGAATGATGTGTCGGATAAGCACTCTGCCTTATAA	TGTTTTCTC	GTATAT		ALKIEPNRVNRAQW
		AGCTGTCGTTCTGCCCAGGTTTCTCCTAACCAGGTTGAACTTGTAATGAG	TGACCTCA	GAAGGT		EVLDSCVVEVLKKRET
		CCTTTGGGTATCTGGTCGGGGTCTGCGCGGAGATAGCTCGTGAGGAGC	TACGGGTA	CGGAGC		SNGYRGCVLRHLNV
		TGTTATTACATTATTGGTTGAACGCTTTTGGGTCGTACTGCACAGATAAA	TATCCCGA	ATTGAA		AVYAAGYVLAEGNKE
		TATGAGAAAAAGTTTTTTACAACATATACCTGAGCTTTCAAGTCATATTG	TGATATCT	GCACAA		RRQVIRRQSAEWLLR
		CTATGTCTGTGCCTGCGAGAAATTATCCCAAAATGTGCTCTTTAGCACAG	ATACAGTG	GTGAGA		QKSEINNIRRHIGWIT
		GGTTCTTCCGGCACATTAAGTCATAATGGGAAGGGTGTCGCAATGCACC	TTCATTTTC	ACCCTG		DELTRRRTGKNPTSR
		GGTGCCCCAGCGATGATTGTGCTGGTAAAGATCCCCCTCAGCGGGGGT	ATTTCTTA	GAAGTA		QLSNFAWLQRRYQV
		CTCAAAAAGGGAATTTGAGGAGTGTGCGATGGACTCCCAGCGAGGAGA	ACCTGCTT	TGGTGG		LGKPVRETRDLEVQR
		AAGCTGTCTTCGAGTATTGGTCACGGTTGGAACAGCATGCCATGTTAAA	TTTCCTCA	TGAGCT		ERLVSRLRLAQDRINS
		TGGTAGCGAAGCCCGGGGGACCTGTGCCATAACTAGAAGTCAGTTTCTC	AGAGAAAT	GGTAAC		SMDREERVRKRMLP
		ATTCATTGGGATGGCGAGAGAGTCAAGATCCTTGTCGGATGGTGTG	ATGACCAC	CTCGAA		LRRKLEEPLGDSKLDT
		CCCGAGTACCCAATGAGGACGGAAAGGGCGTACTATGAAAGGGTCCGA	ATTCGTTT	TTCCTCC		KQARTFWASLIGERK
		CTCTTGCGACAACGCGGATGGCAATGGGATTGTGCTAATGAATGCTTAG	GCCATTTC	TATGGG		EFGKIPELENWAEEV
		TAATTGGCCAATGTGCAGAGCCATGCCGTAAGCCAAACGTGGTAGCGA	GAAGGTG	TGCTTG		RSKVTDGQGFASDH
		TTAAGGCTGATAAGGGTATGAAGAGATCGCTGGTTAAGGGAAAGCTTT	CGACTTTT	CGCCCG		VDQTVWKKILGKAR
		TGTCCCTGCCCCATGTAATGGGCGAAATAAACCAAGTGAGCGTGCAGGT	GAGGGAG	TAGGTC		PLKAPGPDGIPNLLW
		TGAGGTGCCTTTGCCAAGTGTACCCGCCAGCGTGCCCCAGGTGGAAGG	TTCTCAGG	ATTTGTG		KRLPSANQALFKWL
		AGTTGAAAGCAAAGGTTTTACTGAAACGGAACCGTCGAATAAACCTTCC	GACATCAA	TATGTA		MGIKRKQLSVPSWLT
		CTGGAAGGGAATCCGGCGGAAGAAGGATTAAGGAAACCGGAGCGAGT	GCTGTTCA	GGGATG		KGRVVLLPKGGDPVD
		AAACGTACCTGTTCATGGGATAATCAGTGATAGTGAACGGAAAGACCTG	TGGACGGC	GAAGGA		PANYRPIACLNTQYK
		AAAGACCGGTTTTGGAGCGCCTATAAAACGGCCAAGAGGTCTGTGGGT	TTGACGGC	ATCTCG		LVTGMVTAWVSEHL
		TTTAGACCGGCTCTAAAAATCGAACCGAATAGGGTTAACCGTGCCCAGT	TCGTGGCG	ATAGCG		TTYSILPIEQRAMVSG
		GGGAAGTACTGGATTCCTGCGTTGTTGAGGTGCTCAAAAAGAGGGAAA	GCACGCAG	CAAATC		TWGCTHAMVIDRAI
		CCTCCAATGGGTACCGCGGTTGCGTCCTCCGGCATTTGAACGTTGCAGT	CCGCGCGG	GGGATA		TSYAEATGLPLYVGFV
		TTATGCGGCGGGATACGTGCTGGCGGAAGGCAACAAGGAAAGGAGGC	GCCAGTTA	CCGCAC		DFAKAFDSVSQPWIR
		AGGTGATCAGACGCCAGTCGGCGGAATGGTTGCTTCGCCAAAAGTCCG	GGGCTGG	TGGCGA		YALKVAGVHKRIRCLI
		AAATCAACACAACATCCGTAGACACATCGGGTGGATTACCGATGAGTTAAC	TGCCAATG	TCATGCT		GILMKCWSVRYEVFK

	ACGGAGGAGGACGGGAAAGAACCCAACATCCAGACAGTTGAGTAATTT	AGGCAGC	GCCAGT		SGRVLRSAPLAVKNG
	TGCATGGTTGCAAAGGCGATACCAGGTTTTGGGTAAACCCGTGAGGGA	GAAGAAG	GGGCCC		VLQGDTLSPLLFCLSV
	AACCCGGGACTTGGAAGTCCAACGAGAAAGACTTGTAAGTCGTCTTCGA	TCTCTGAA	TGCTGA		AVVSSAVGSLFDFEV
	TTGGCTCAGGACCGAATTAACTCTTCCATGGATCGTGAAGAGCGAGTTC	ACGCCAAA	GTTGGG		TIPGRGVMQQQNHL
	GTAAACGAATGTTACCCCTGCGGAGAAAGTTGGAAGAACCTTTGGGTG	AGCAGCG	CGGGTT		FYMDDFKGFAPSEAS
	ACTCTAAACTTGACACAAAGCAGGCACGAACATTTTGGGCCAGCCTGAT	GGAGAAG	AGCAAC		LTRMLVTLERTASAL
	TGGAGAGCGAAAGGAATTTGGAAAAATCCCGGAGCTTGAAAACTGGGC	AGAGAGA	CCGTCC		GLKINKRKCALVHPR
	CGAAGAAGTGAGGAGCAAAGTCACCGATGGACAAGGTTTCGCAAGTGA	AAGCGGG	AGGGAC		ERENEETGSDIPVLGL
	CCATGTGGATCAAACTGTTTGGAAAAAGATACTGGGTAAGGCCAGGCC	GACGGGTT	GCAGAT		RDTYKYLGIEERFGIV
	ACTGAAGGCGCCCGGACCAGATGGAATCCCCAATTTACTTTGGAAACGG	AGAGATAA	GCCAAA		FEDAWDRVRTKMFE
	TTACCATCCGCAAACCAGGCGTTATTTAAGTGGTTAATGGGGATCAAGC	TAGCATGA	AGCCTG		RMRTLLCTEHTFGEL
	GGAAACAGCTCAGTGTCCCCAGTTGGCTCACAAAAGGAAGGGTGGTCC	TTGACGGC	GTGACA		RAAFASTIAPVARYLF
	TCCTACCCAAAGGAGGTGACCCGGTAGACCCGGCCAACTACCGACCAAT	GAAGCAAT	CTCTAAA		LNVIVGGPSWSETLT
	CGCTTGTTTGAACACGCAGTACAAACTAGTGACTGGTATGGTGACGGCA	ACTTTGAC	GGGAGT		KAKDMDLRIRRLLWE
	TGGGTATCGGAACACCTGACCACATACTCTATTTTGCCGATTGAACAAC	TTCCAGGT	CTACGG		RRDNEPGWRFKHCS
	GAGCGATGGTGTCTGGCACTTGGGGGTGTACCCATGCTATGGTGATTG	ACGGAGG	CGACGA		ADRLYLRVQYGGLGF
	ACAGAGCAATTACGTCTTATGCAGAGGCCACTGGACTACCGCTTTATGT	ACGGCGG	GAACAT		VSVEDTLSESIIYCWA
	AGGCTTTGTGGACTTTGCGAAGGCTTTCGACTCAGTTAGTCAACCGTGG	CAAAAAGC	CCCCTAA		YVQCRPELELARELFG
	ATCAGGTACGCTCTAAAGGTCGCTGGAGTTCACAAGCGAATCCGATGTC	AAAAAGG	GTC		TLNRSARSGIKQSIAK
	TTATTGGAATCCTTATGAAATGTTGGAGCGTAAGGTATGAAGTGTTCAA	GGACGAA	(SEQ ID		GARKVFRSYALLSKNS
	GAGCGGGAGAGTGTTAAGGAGTGCGCCTTGGCAGTTAAGAACGGAGT	ACCGATGG	NO:		AQRVSDLDGDASPG
	GTTACAAGGAGACACGTTGAGCCCGCTGCTCTTTTGTCTCTCAGTAGCC	CATTGCCA	1379)		FRVGEMIFMEPTRG
	GTTGTTTCTTCTGCAGTAGGCAGCCTCTTTGATTTTGAGGTCACCATACC	AAGGGGC			ARAIVKILRKENDSRR
	TGGACGAGGAGTGATGCAGCAGCAGAACCACTTGTTCTATATGGATGA	CGAAACCA			LAAWKGRPMGGRV
	CTTTAAAGGGTTTGCTCCTAGTGAAGCGTCCTTGACTAGGATGTTAGTA	TTGAGCTG			VSLPELDQVHSYHWL
	ACATTGGAGCGAACCGCTTCAGCCCTCGGGTTGAAAATCAATAAGAGG	GACGCTGA			IRARIGRRSFRDCIAA
	AAATGCGCTCTGGTACACCCCCGTGAAAGAGAGAACGAGGAAACTGGA	TGGGCCGT			QEGQLKARELMCPHI
	AGTGATATCCCGGTCCTTGGTCTACGAGACACTTATAAATACCTTGGAAT	CTGCCGCG			NAKAKWCRRCGDG
	CGAAGAGAGATTTGGTATCGTTTTCGAGGATGCGTGGGACCGAGTCCG	CAAAAGG			RVETEQHILSGCAWS
	TACCAAGATGTTTGAGAGAATGCCGCACCTTGCTATGCACGGAACATACC	AGGGGGG			RTGTMLDRHNGVVR
	TTCGGGGAGTTGCGTGCGGCCTTCGCTAGTACGATCGCCCCAGTAGCGA	AGCCCCTC			QVHTALCRKYGLPVS
	GGTACCTCTTTCTCAATGTAATCGTTGGTGGGCCTAGTTGGAGTGAAAC	CGTCTGGG			SHVVPLHAVIENEHA
	CCTGACGAAAGCTAAGGACATGGACCTGCGTATCCGACGCCTCCTGTGG	AAGAAGG			KILYDVALHTSPAGVL
	GAAAGAAGAGATAACGAGCCGGGATGGCGGTTTAAGCATTGTAGTGCT	TGCGGCGC			PREDGSTSYTGLRSTR
	GACCGCTTGTACCTCCGCGTACAGTACGGTGGACTTGGCTTCGTGTCCG	CATTCGTC			PDMVIFDKKARTILIV
	TAGAGGACACTTTGTCTGAGTCTATCATTTACTGTTGGGCTTACGTGCAA	CAACAAAC			EISVPWRENLVKQELI
	TGTAGGCCCGAGTTGGAGCTGGCTCGTGAACTTTTTGGAACCTTAAACA	CGGAATAG			KWRKYAINSMIEPLE
	GGAGTGCTAGAAGCGGTATAAAACAGAGCATTGCGAAAGGAGCGCGT	ATCTTGAT			LAEAEIPGPNLKHAL

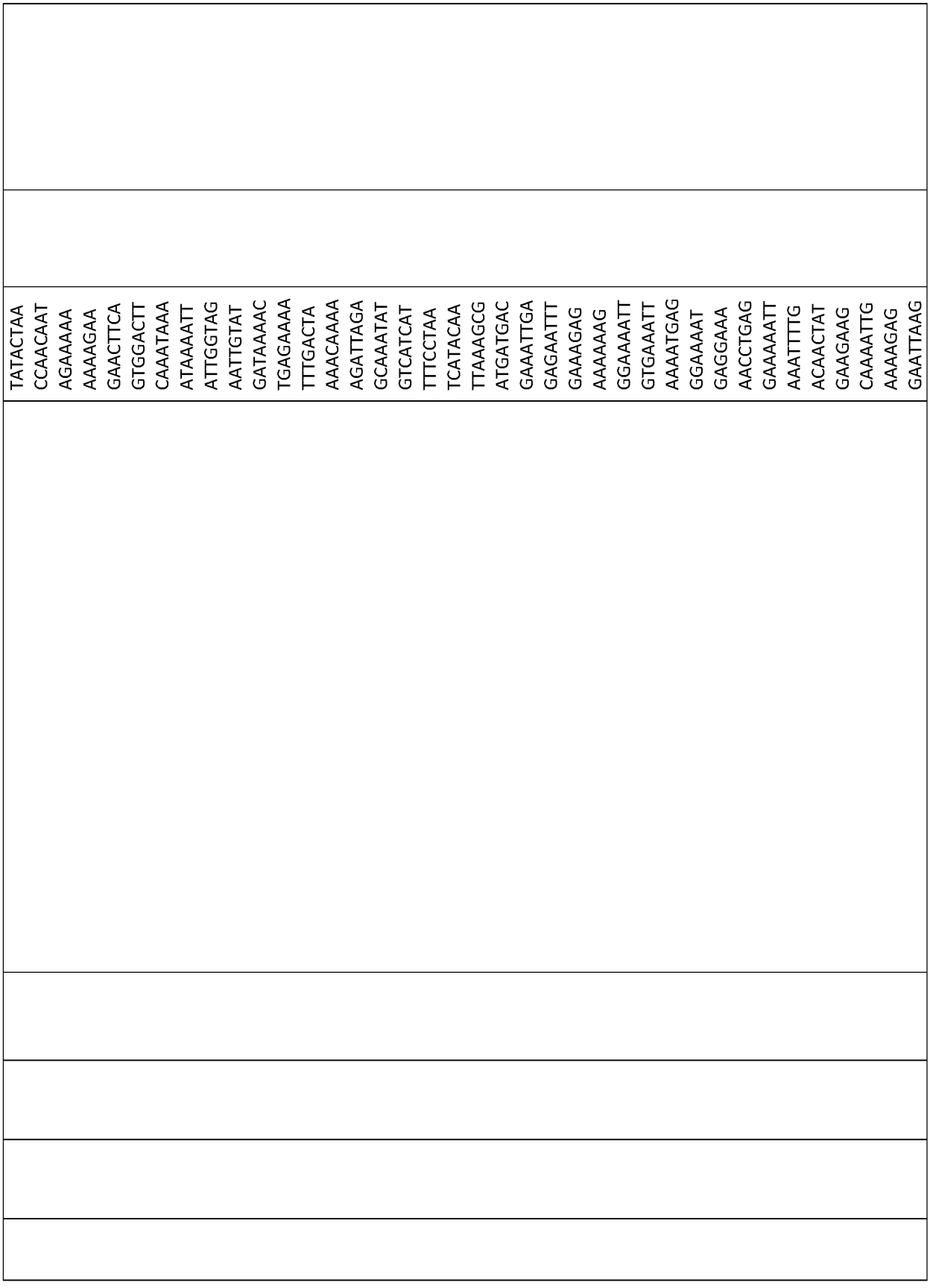
		AAAGTGTTCCGAAGTTATGCCCTTCTCTCGAAAAATTCCGCCCAGCGAGT	GATCTCAG		GLAYGTSFPTVKVVPI
		ATCTGACCTTGATGGGGATGCCTCCCCAGGATTCCGAGTAGGGGAAAT	CAATTGAT		VVGSCGEVLPNITKRL
		GATATTCATGGAACCAACTCGGGGTGCCAGAGCAATCGTAAAAATCCTC	CATATTGA		SELGIPKRGIPSLLESI
		CGTAAGGAGAATGATTCGCGACGCTTGGCTGCCTGGAAAGGTAGACCT	TCTCTTTT		QRAAIIGSGHVIRAHL
		ATGGGAGGTCGGGTCGTGTCATTGCCAGAGCTCGACCAGGTTCATAGCT	CCAGTTGT		SVPRSESET (SEQ ID
		ACCACTGGTTGATCCGTGCGAGGATTGGCAGGCGTTCGTTCCGAGACTG	TGCGATAA		NO: 1501)
		CATTGCAGCACAGGAGGGACAGTTGAAAGCGCGGGAGCTTATGTGTCC	ATTATCGT		
		ACATATAAATGCGAAAGCAAAGTGGTGCCGACGATGTGGGGACGGTCG	GCATTATT		
		TGTTGAAACCGAGCAGCATATCCTTAGTGGTTGTGCTTGGTCACGGACT	TCTTCGAT		
		GGAACTATGTTAGATCGACATAACGGTGTTGTACGGCAAGTACACACG	TTCTCCAA		
		GCTTTATGCAGGAAATATGGACTGCCCGTTTCTTCTCATGTCGTTCCTTT	AGCTTAAC		
		GCATGCGGTAATCGAGAATGAACATGCCAAAATCCTGTACGATGTCGCG	TCCTTTCCT		
		TTGCATACTTCGCCTGCTGGCGTCCTCCCTAGGGAGGATGGGAGTACGA	AATACCTA		
		GTTATACGGGGTCTGCGTAGCACCAGGCCCGATATGGTAATATTCGATAA	CTCATGTA		
		AAAAGCACGTACGATCCTTATTGTGGAAATTTCGGTGCCATGGAGAGAG	TACGTTAG		
		AACCTGGTTAAACAGGAGCTCATAAAGTGGCGGAAATATGCCATAAATT	TAGGCATG		
		CTATGATCGAACCATTGGAACTAGCGGAAGCTGAAATACCGGGACCTAA	TTTTATGC		
		CCTGAAACATGCGTTAGGCCTTGCGTACGGTACGAGTTTCCCAACAGTG	AGGTAAA		
		AAGGTGGTCCCGATAGTAGTTGGGTCGTGCGGCGAGGTATTACCAAAT	GATAGACC		
		ATAACGAAGAGGCTGAGTGAACTTGGGATACCTAAACGAGGAATCCCG	TGGTGCCC		
		AGTTTGCTTGAGTCGATACAGCGTGCTGCGATAATTGGTTCTGGACATG	CCGGGCG		
		TTATCCGGGCGCACTTGAGTGTACCTAGGAGCGAGTCCGAGACTTAGTC	ACTTGGGA		
		GTAGCCCAGATGTCGACAGCCCCTACACCGGTAGATGGAAGGAGGTAT	TGTGGATG		
		GTATCTAAGCCCACGGCAAGCCACCAGTGGAAACGGTGACCACCTCTGT	ATGGTTGG		
		TCGCGGAAATGCCCCGTAGAGTATATGAAGGTCGGAGCATTGAAGCAC	CGGAGAG		
		AAGTGAGAACCCTGGAAGTATGGTGGTGAGCTGGTAACCTCGAATTCCT	TTCTGATG		
		CCTATGGGTGCTTGCGCCCGTAGGTCATTTGTGTATGTAGGGATGGAAG	ACGCCGTA		
		GAATCTCGATAGCGCAAATCGGGATACCGCACTGGCGATCATGCTGCCA	GTTCCGGA		
		GTGGGCCCTGCTGAGTTGGGCGGGTTAGCAACCCGTCCAGGGACGCAG	GGAACAGT		
		ATGCCAAAAGCCTGGTGACACTCTAAAGGGAGTCTACGGCGACGAGAA	TCCTACTA		
		CATCCCCTAAGTC (SEQ ID NO: 1133)	GTGCCAGC		
			CTGGGCTA		
			GTGGAGCT		
			GGTCTGGC		
			GAGTGCTT		
			TGCTCGTC		
			AGATAAG		
			GGGTGGTT		

					GGGGTGG TACCTCTG GAATGATG TGTCGGAT AAGCACTC TGCCTTAT AAAGCTGT CGTTCTGC CCAGGTTT CTCCTAAC CAGGTTGA ACTTGTAA TGAGCCTT TGGGTATC TGGTCGG GGTCTGCG CGGAGATA GCTCGTGA GGAGCTGT TATTACAT TATTGGTT GAACGCTT TTGGGTCG TACTGCAC AGATAAAT (SEQ ID NO: 1256)		TAATTA GATTTAT ATGTGC CACTTTA CCTCAA AGATAT ATTAATA CTAACTT AATTATA ATTTATT ATGAAT	MQKFSVPKDSSQIFL VDSILNKHICSTKNKV KDVIKRRSIIKTLICAA GLTLRKLVCGKLGNN KYNSKINQLWKKKERK IINCIEDLKHLIETNKR RHNFGKRLRNAKVSP SEMLKDYYNKLRYIK NEITACLDEHKKAILR TKFKLTPSIKIISNIQN HNDEEAELPKEEEFV
R4	R4-2_SRa	CACX010020 06	Strongyloides ratti	CGTCAGAAGAGCAGGTGTTTTTCAAAGCAAAGACTTATTCTACGAAGGG GAAAGATGATCAAACATGCAGATTTGGCTGCAATGAAATAGAATCAAAC TACCATGTGGTGACATCTTGCAAATACCATTCATACACACAAGAAGACACG ATATGGTTGTGTACAATATATTAAAAAGACTAAAAGAAAAGTATAAGTT AGATGGTGATTTACAATATGGTAGATCAGTATTTAATGGGAAGAAGGA AAAAATTAGGATCAGATCTGGCCATAAGTTTTTGACTGAACAACCGTTA GTAAACAATAAACCGGATATTGTTATAACGATGAAAGAAGGGAAAAAA AAGATAACGTACATACTACTAGAGATGTCGATACCACATATCTAGAACTTGA AGATGCAAGAAAGAATAAAGTATATAAAAAATATTGTGTGAGCTCCATGGT AAAAATAACGAATGATAATGTTGATTCGATAGCTCGTGATTTTAATCTAC TAAATTTAATGGAAAGAAAAGAAAGGTGTAAGATAAAGTTTGGATCATT	CGTCAGAA GAGCAGG TGTTTTTC AAAGCAAA GACTTATT CTACGAAG GGGAAAG ATGATCAA ACATGCAG ATTTGGCT GCAATGAA			

		TGTATTTGGCTGTTATGGAGAATATGTCTTGACTGAATCTGCTCTAAGAA	ATAGAATC	TATAATA	KYYKELFTNKDGDDK
		CACAGAAAATCCTGATTGAATTGGGATTCTCCAAAAAAGAGATGGATAG	AAACTACC	TAATAA	ETPHLDNWLKKFSKT
		CTTGATCAAAGAGTGTTCTTACAGTTAAATGAATGAGACAGCAAGAATT	ATGTGGTG	GTTTTAA	LIVDWTINDKEILEAL
		ATTATGAAACATCTAGAAGGCGAAAAGAAATCAAGATGAAATCAGCCATT	ACATCTTG	AATAAA	KYCGNFKAPGSDMIV
		AATAACTTTTATTTACCCAATTATTTATTGCGTTAACTTGATTATATGATA	CAAATACC	A (SEQ	MKVCYKWFKSAQNY
		TTTATTGTATATTTATTGTTTAATTTAGTTTCATAGTACAAATAAAAGAAT	ATTCATAC	ID NO:	LIRWIKSTWYGEYTIN
		ATACTACGATTCTTTACTTTTAAGTTCTACGAGAACGTTGTTTTAGAATAT	ACAAGAA	1380)	KKDTNAVTFMIWKR
		TTTAATAATATTTCTACTACAATTAAGTAATTAGAAGAAATCAACGAAAG	GACACGAT		DGKPKNDVKSYRPIS
		CAGCTAAACTTACTCGCAAAATTCGTTGATCGAGGCTGGAATGGCACCA	ATGGTTGT		CLNCDFKLLNKLIANK
		ACTAATATACTAACCACAATAGAAAAAAAAAAGAAGAACTTCAGTGGAC	GTACAATA		IYESIEKILPINQMAVI
		TTCAAATAAAATAAAATTATTGGTAGAATTGTATGATAAAACTGAGAAA	TATTAAAA		KNKHGTCEALLLYKSL
		ATTTGACTAAAACAAAAAGATTAGAGCAAATATGTCATCATTTTCCTAAT	AGACTAAA		VQSMKFRRTKDVKEI
		CATACAATTAAAGCGATGATGACGAAATTGAGAGAATTTGAAAGAGAA	AGAAAAGT		WCSWIDFSKCYDSIS
		AAAAGGGAAAATTGTGAAATTAAAATGAGGGAAAATGAGGAAAAACCT	ATAAGTTA		HKCLKKMIQSIKAPPII
		GAGGAAAAATTAAATTTTGACAACTATGAAGAAGCAAAATTGAAAAGA	GATGGTGA		HKLILDGIDSWNISIC
		GGAATTAAGTGTAAAAAAGAAGTAAAACCAATTGTTATTGAAACAAG	TTTACAAT		NGKNISKTKIPVKSGI
		GATTAGGAATTTTTGAAAACGGAAAAATTATTTCTCAAGTTTGTCAACAC	ATGGTAGA		LQGKVASSLYFVLLTG
		AATCAATTACCAAAGAAAGAAGGGTAAATCAACCAGAATGCAAAAATTT	TCAGTATT		EISYALNKEEQVPIETI
		AGTGTTCCAAAGGACTCTTCTCAAATATTCTTAGTGGATAGTATCCTAAA	TAATGGGA		TPSNTLKINHISFIDDY
		CAAACATATTTGCTCAACCAAAAACAAAGTTAAGGATGTCATTAAGAGG	AGAAGGA		QLYATSQKKVEKLTIK
		AGAAGTATTATTAAGACATTAATATGTGCAGCTGGTTTAACTCTTAGAAA	AAAAATTA		LREIAEEMNLKLNPQ
		GTTAGTTTGTGGGAAACTTGGAAATAATAAATATAATTCCAAGATAAAT	GGATCAGA		KCGIYGTDDLGKRLM
		CAACTTTGGAAAAAAGAAAGGAAAATTATTAATTGTATAGAAGATTTAA	TCTGGCCA		LKESSLNFPYTSEYKYL
		AGCATTTGATAGAAACGAATAAACGTAGGCACAATTTTGGGAAAAGATT	TAAGTTTT		GLVENSLDLKDINIQL
		AAGGAACGCCAAAGTGAGCCCTTCTGAGATGTTGAAGGATTACTACAAC	TGACTGAA		FKDKILSKYSTIFESRL
		AAGTTAAGATATATTAAAAATGAAATTACGGCATGCTTGGATGAACACA	CAACCGTT		TTHQKRKVFNSTISPC
		AAAAAGCAATTTTAAGAACAAAATTTAAACTAACACCAAGCATTAAGAT	AGTAAACA		AAYYLGNLITNKCSIQ
		CATCTCAAATATTCAAAATCATAATGATGAAGAGGCTGAACTCCCAAAG	ATAAACCG		ELLNECKKFDQMVR
		GAAGAAGAATTCGTCAAATACTATAAAGAAACTATTTACAAACAAAGATG	GATATTGT		NQLVNQNIKKLQVS
		GGGATGATAAAGAAACACCACATCTTGACAATTGGTTGAAGAAGTTTAG	TATAACGA		NSRIYLPKEYNSLGLN
		TAAGACATTAATTGTAGACTGGACTATAAATGACAAGGAGATTCTTGAA	TGAAAGAA		EIEIEVAANIIRKACYIK
		GCTTTAAAATATTGTGGTAATTTCAAAGCACCGGGATCTGACATGGTAA	GGGAAAA		KRETLRGVDKLYIAM
		TGAAAGTTTGTTACAAATGGTTCAAATCTGCACAAAATTATCTAATCAGA	AAAAGATA		SKNGHRNTLSDALYIT
		TGGATCAAATCAACTTGGTATGGTGAATACACAATAAATAAGAAAGATA	ACGTACAT		KKYSNFQINWNIMG
		CAAATGCAGTCACATTCATGATTTGGAAAAGAGATGGTAAACCAAAAAA	ACTAGAGA		MVKDQNNILLDAKKI
		TGATGTCAAATCATATAGACCGATTTCTTGCTTGAACTGTGATTTCAAGT	TGTCGATA		IENIKEKRRNLWLEH
		TACTAAACAAACTTATTGCAAACAAGATTTATGAATCAATAGAGAAAATT	CCACATAT		WKKGNMTYANEAIK
		TTGCCTATAAATCAAATGGCAGTAATTAAGAATAAACATGGAACATGTG	CTAGAACT		KEFHLPDLNIDSKYL

		AAGCCCTACTCCTATACAAATCATTGGTTCAATCAATGAAATTTAGAAGA	TGAAGATG		MLCYAGSEEQIIYNG
		ACGAAAGATGTGAAAGAAATATGGTGTTCATGGATTGACTTCAGCAAAT	CAAGAAA		HVSLVNQSSPSSRLC
		GTTATGACTCAATAAGTCATAAGTGTTTGAAAAAAATGATACAAAGTAT	GAATAAAG		RKCNKLEETSYHVAS
		TAAGGCTCCACCAATTATTCACAAACTCATTTTAGATGGAATTGATAGCT	TATAAAAA		VCEFHKKNLHLMRH
		GGAACATATCAATTTGTAATGGGAAAAATATTTCCAAAACTAAAATACCT	ATATTGTG		NSAVYHIITELCRIMK
		GTTAAATCTGGTATTCTTCAAGGTAAAGTGGCTAGCTCTTTATATTTTGT	TGAGCTCC		VKCTLRYPEASGIIKS
		ATTATTGACTGGTGAAATATCTTATGCTTTGAACAAGGAAGAACAAGTA	ATGGTAAA		GNMKIAAGVKYTFG
		CCAATTGAAACGATTACACCAAGCAACACTTTGAAAATCAACCATATCA	AATAACGA		TAKIYHNKPDLVWYT
		GTTTTATAGATGATTATCAACTGTATGCTACATCACAAAAGAAAGTTGAA	ATGATAAT		PEVIYVIEVSISSLKNA
		AAACTAACTATTAAATTGAGGGAAATAGCAGAAGAGATGAACCTGAAG	GTTGATTC		KSQMKMKTARYAVN
		CTGAATCCACAAAAAATGTGGCATATATGGCACAGATGATCTGGGAAAG	GATAGCTC		STKKLENFAALNNLK
		AGACTAATGCTGAAAGAAAGTTCATTAAACTTTCCATATACTTCCGAGTA	GTGATTTT		KGENFVEILSHKANFK
		CAAGTATCTTGGTCTGGTAGAAAAATTCATTGGACTTAAAGGATATCAAT	AATCTACT		RVHFMPLVFCTFGEI
		ATCCAACTGTTTAAAGATAAAAATACTATCTAAATACTCAACTATTTTTGAA	AAATTTAA		PKETMKYLEKLNFSN
		TCAAGACTCACTACACACCAGAAAAGAAAAGTATTTAATAGTACAATAT	TGGAAAG		EKIKTIASPIARYTGRT
		CTCCATGCGCAGCATATTACCTAGGTAATTTAATAACCAATAAATGTTCA	AAAAGAA		LKAHFTN (SEQ ID
		ATTCAAGAACTATTGAATGAATGTAAGAAATTCGATCAAATGGTCAGAA	AGGTGTAA		NO: 1502)
		ATCAGTTAGTGAACCAAAACATCAAAAAGCTTCAAGTATCAAATAGTAG	GATAAAGT		
		AATCTACTTACCTAAAGAATACAACTCCCTTGGATTAAACGAAATAGAAA	TTGGATCA		
		TTGAAGTAGCAGCCAATATAATCAGGAAAGCTTGTTACATTAAAAAAAG	TTTGTATTT		
		AGAAACTCTTAGAGGAGTAGACAAACTATATATTGCAATGTCAAAAAAT	GGCTGTTA		
		GGTCACAGAAACACTCTTAGTGATGCTTTATATATAACTAAAAAATACTC	TGGAGAAT		
		TAACTTTCAAATAAATTGGAATATTATGGGTATGGTGAAAGATCAAAAC	ATGTCTTG		
		AATATTCTTCTGGATGCGAAGAAAATTATTGAAAATATCAAAGAGAAAA	ACTGAATC		
		GGAGGAACCTATGGCTGGAACATTGGAAAAAAGGAAACATGACATATG	TGCTCTAA		
		CAAATGAAGCAATCAAAAAAGAATTCCATCTCCCGGACCTAAACATTGA	GAACACAG		
		CTCCAAATACCTCATGTTGTGCTATGCTGGTTCAGAGGAACAAATCATTT	AAAATCCT		
		ATAATGGACACGTAAGTTTGGTTAATCAATCCTCACCGTCTAGTAGATTA	GATTGAAT		
		TGTAGAAAGTGTAACAAATTGGAAGAAACTTCATACCATGTGGCTAGTG	TGGGATTC		
		TTTGTGAATTCCATAAAAAGAATCTACATCTAATGCGTCACAATAGTGCA	TCCAAAAA		
		GTTTACCACATAATAACTGAACTCTGTAGAATCATGAAAGTCAAATGTAC	AGAGATG		
		GCTAAGATATCCAGAAGCATCAGGAATAATAAAAAGTGGAAACATGAA	GATAGCTT		
		GATTGCTGCTGGAGTAAAATACACTTTTGGAACAGCAAAAATATACCAT	GATCAAAG		
		AATAAGCCTGACTTGGTATGGTATACTCCAGAGGTAATTTATGTTATAG	AGTGTTCT		
		AAGTAAGTATTTCATCACTGAAGAACGCCAAATCTCAAATGAAAATGAA	TACAGTTA		
		GACTGCTAGATATGCAGTAAACTCAACAAAGAAATTGGAAAATTTTGCT	AATGAATG		
		GCTCTGAATAATTTAAAGAAGGGAGAAAACTTTGTTGAAATATTAAGTC	AGACAGCA		
		ATAAAGCAAATTTCAAAAGAGTCCATTTCATGCCCCTAGTTTTCTGCACT	AGAATTAT		

	TTTGGTGAAATACCTAAAGAAACAATGAAATACCTGGAAAAACTCAACT TTTCCAATGAAAAGATTAAAACAATTGCTTCACCAATCGCACGTTACACT GGAAGAACTCTCAAAGCACATTTTACTAATTAATTAGATTTATATGTGCC ACTTTACCTCAAAGATATATTAATACTAACTTAATTATAATTTATTATGAA TTATAATATAATAAGTTTTAAAATAAAA (SEQ ID NO: 1134)			TATGAAAC ATCTAGAA GGCGAAA AGAATCAA GATGAAAT CAGCCATT AATAACTT TTATTTAC CCAATTAT TTATTGCG TTAACTTG ATTATATG ATATTTATT GTATATTT ATTGTTTA ATTTAGTT TCATAGTA CAAATAAA AGAATATA CTACGATT CTTTACTTT TAAGTTCT ACGAGAAC GTTGTTTT AGAATATT TTAATAAT ATTTCTAC TACAATTA AGTAATTA GAAGAAAT CAACGAAA GCAGCTAA ACTTACTC GCAAAATT CGTTGATC GAGGCTG GAATGGCA CCAACTAA		


TATACTAA
CCAACAAT
AGAAAAA
AAAAGAA
GAACTTCA
GTGGACTT
CAAATAAA
ATAAAATT
ATTGGTAG
AATTGTAT
GATAAAAC
TGAGAAAA
TTTGACTA
AAACAAAA
AGATTAGA
GCAAATAT
GTCATCAT
TTTCCTAA
TCATACAA
TTAAAGCG
ATGATGAC
GAAATTGA
GAGAATTT
GAAAGAG
AAAAAAG
GGAAAATT
GTGAAATT
AAAATGAG
GGAAAAT
GAGGAAA
AACCTGAG
GAAAAATT
AAATTTTG
ACAACTAT
GAAGAAG
CAAAATTG
AAAAGAG
GAATTAAG

				TGTAAAAA AGAAGTAA AACCAATT GTTATTGA AAACAAG GATTAGGA ATTTTTGA AAACGGA AAAATTAT TTCTCAAG TTTGTCAA CACAATCA ATTACCAA AGAAAGA AGGGTAA ATCAACCA GA (SEQ ID NO: 1257)	TGAGAC CACCCAT GCGCAG AGTATC CGAATC AGTGAA AGTCCA AGTTTCA GGACAG AAACGT CAGATA AGTCCA AGAGAA ACGAGA AAACAA GTTCAA GTATGC AAGAGT TAATCA ATAAGA	MISRSQADRPVEGQ PVTAMSFHNLEPNN LYPENLRPTGSQDAN RGVADIAEEVTGPSG LVTNEEAARAPPLFV EGEYKRAHCGGGKC HYCRVLWIGARSSKA RTDSWNLCEILFLINK CMELGNVRRIYSPLE SSLKEAGINRTRHAIV KCRLAVMRDRFVDN APYSEHWRLYNACA ENRAVVVPMDSATT VKKRTARQAGLESPS QIGVAGKRVHEAETG TDRINAVIETNTTPLE DIDLSPETPEGLAELP STVEIMELTEDGSRSR GTANDADGGVSISDP LRNRPSSSQESRNVP
R4	R4-3_BX	CADV010088 32	Bursaphelenchus xylophilus	GGGTCCTCGGTTCTTACTACCGTGCTCCACCTCCTCGCGATGGACCCTGG GGTAGGCCTCCGGGCTGAGCTAAGCAGAGCATTAAAGTAAAGAGTGAC GGCGCAGTTGCTTCCAAGTTGCGGTGGGGCGGAACATCTACTCTTCTGA GAGAGGGGAAGCCCTATGGCGGCGTTAGAAAGGTTGGACTGCGGCAA CACTAGCCATGATCAGATTCGATCAAAATAGCCTCTGGGGCTGGCGACC CTATAACGGATTGTAAACTGAACTATGCTAACCTGTCAGTAAAGACAGA ACGGGGGCTTTGCAATCTAAGGTGTTGGCAGACCACTAAAACTGCCCTT TGATGAGGGTGAGATAACATCCTACGGAACAGCCCCTGCTGATCCAAAC AAAATCATCCACAAATCGGAGGGTTTTCTTGGTAGTTTCCCTTGGCACGT CTTGTTTCAAAAGCCAGAATATAAAACAATACCGAAGTACAGAATGGCT GGAACCTAACAGCCATGGCACAGAAAACCGGTGGGTGCACACCGGAGA CAGCAGGGGAAAGATTGCAAAGTGCCGTAAATAAAGATGGTAGCTCTC TGACCTGAGGGAATGGGACACTGTCCGATTAATGCCTTATCCCGAAGAC GGTCAAGGTTCTATCTTATCGGTCTCCAAAGGGCTCTGGCCATAGAAGC TGCGAGGAAGCCGTAACCCTACGCAGATTGCCGCCGTGCATGATAAGTC GTAGCCAAGCGGATCGACCAGTGGAGGGCCAACCCGTAACAGCCATGA GTTTCCATAATTTAGAACCAAACAACCTGTACCCAGAAAATCTTCGTCCT ACTGGGTCACAGGACGCGAATCGCGGAGTTGCGGACATTGCTGAGGAG GTGACTGGTCCTTCGGGGCTGGTTACTAACGAGGAGGCAGCGCGAGCA CCGCCGCTTTTCGTAGAGGGCGAATATAAAAGGGCACACTGCGGTGGG	GGGTCCTC GGTTCTTA CTACCGTG GCGCAG AGTATC CCTCGCGA TGGACCCT GGGGTAG GCCTCCGG GCTGAGCT AAGCAGA GCATTAAA GTAAAGA GTGACGGC GCAGTTGC TTCCAAGT TGCGGTG GGGCGGA ACATCTAC TCTTCTGA GAGAGGG	GTATGC GCTTGC AAACAA GTTCAA GTATGC AAGAGT TAATCA ATAAGA	

		GGCAAATGTCACTACTGCAGGGTCCTGTGGATAGGTGCGAGGTCTTCAA	GAAGCCCT	GAGTAC	EQVDPDGELVWESL
		AGGCCCGTACCGACAGTTGGAATCTCTGTGAGATTCTATTTTTGATTAAC	ATGGCGGC	CGTAAA	YGAQLRGAMRTTDR
		AAATGCATGGAACTTGGAAACGTGCGTCGGATTTACAGTCCTCTCGAGA	GTTAGAAA	TGTATG	PRLPKLTKFSAAEQL
		GTTCTTTGAAAGAGGCTGGAATCAACCGAACACGTCATGCAATAGTTAA	GGTTGGAC	ACCCCCC	WIKSKVEKARLECVSY
		GTGTCGACTAGCTGTCATGCGTGATCGATTTGTGGATAACGCCCCATATT	TGCGGCAA	CCTTTGC	GIEQQWLRASAVLYA
		CTGAGCATTGGCGTCTGTATAACGCGTGTGCAGAAAAATAGGGCGGTGG	CACTAGCC	CAAGTC	TIKTVAACRPYNKAR
		TCGTCCCGATGGACTCGGCCACACAACGGTGAAGAAAAGAACTGCTAGAC	ATGATCAG	GACAAC	EAHKVWLENKRAEE
		AAGCGGGCTTGGAATCACCTTCGCAAATCGGGGTCGCTGGCAAGCGAG	ATTCGATC	TGTCAT	KRVRRIIGRIETVRTM
		TCCACGAAGCTGAAACTGGAACTGATCGGATTAACGCGGTGATTGAAAC	AAAATAGC	GCAGGT	PKGKRTDKQIRLARKI
		GAACACAACTCCGCTAGAGGATATTGACTTGAGTCCAGAGACCCCAGAA	CTCTGGGG	GTCTCTC	NRLKRVSFPEMDWH
		GGGCTTGCAGAGTTGCCTAGTACTGTAGAGATAATGGAATTGACGGAA	CTGGCGAC	TTTTCAC	GFLNHFKAKLDLLKKL
		GACGGGTCTAGATCCAGAGGGACTGCGAACGACGCTGATGGCGGGGGTT	CCTATAAC	CCGCCA	ISVRVAEHERKISRKIA
		TCCATCAGCGATCCACTACGAAATCGCCCAAGCTCAAGTCAGGAAAGCA	GGATTGTA	TATGGA	GTYGKSVSGQSGFTP
		GAAATGTCCCGGAGCAAGTAGACCCGGATGGGGAACTTGTATGGGAGA	AACTGAAC	CCAAAC	DVVAATTFWSGLAQ
		GTCTATACGGAGCACAGTTGAGGGGAGCGATGCGCACAACTGACAGGC	TATGCTAA	GCTATCC	PGPKKFKKNSLIFQT
		CGAGGTTGCCAAAACTTACGAAATTCTCTGCAGCCGAACAGCTGTGGAT	CCTGTCAG	AGCCTC	WKDSVVENMNTEP
		TAAATCGAAGGTTGAGAAAAGCTCGGCTGGAGTGCGTTAGCTACGGCAT	TAAAGACA	GCTCAG	VLLHPLIIECMNKPSP
		TGAACAGCAATGGCTCCGTGCCAGCGCAGTTCTATATGCCACGATCAAG	GAACGGG	AAGAGC	FKATGPDGIFNSYWR
		ACTGTAGCGGCTTGTCGTCCGTACAACAAGGCCCGCGAAGCTCACAAG	GGCTTTGC	CTTAGG	QGFIANWVKSLIQRT
		GTATGGCTCGAGAACAAACGTGCTGAAGAAAAGAGAGTACGAAGGATT	AATCTAAG	GCTGGG	IQTGEFPASLMCGRT
		ATTGGGCGAATTGAGACTGTACGCACCATGCCGAAAGGCAAACGAACT	GTGTTGGC	GAGTAC	VLLYKNGDTAKPENY
		GACAAGCAGATCCGCCTTGCGCGGAAGATTAACAGGTTGAAACGTGTC	AGACCACT	CACATG	RPITCLNGCFKMTNA
		AGCTTCCCAGAGATGGACTGGCACGGATTCCTGAATCACTTTAAAGCTA	AAAACTGC	TGGCGG	VITKVIVQRVQDTCA
		AGTTAGATCTTCTTAAAAAGTTGATCTCAGTTCGGGTGGCAGAACATGA	CCTTTGAT	AAACTG	LPGEQMALKPKVWA
		GCGGAAGATTAGTCGGAAGATAGCAGGTACATACGGCAAAAGTGTATC	GAGGGTG	AATCTG	CMEAQLRDQALQSEI
		CGGTCAGTCAGGGTTTACGCCGGATGTCGTTGCCGCAACGACTTTCTGG	AGATAACA	GATGCG	GNDCKTAWIDFSKAY
		TCAGGCCTTGCACAGCCTGGACCAAAGAAATTTAAAAAGAACTCGCTGA	TCCTACGG	ATGCAT	DSLDHDAIRFVIETLA
		TCTTCCAGACTTGGAAAGACAGTGTAGTTGAGAAATATGAACACTGAGCC	AACAGCCC	ACCGGG	LPDGMEKYLLKSLES
		AGTACTGCTACACACCCGCTGATCATAGAATGTATGAATAAGCCCAGTCCG	CTGCTGAT	AGCGCA	WKTKLVLSNRGKVAT
		TTCAAGGCCACTGGCCCTGACGGGATTTTCAACTCCTATTGGAGGCAAG	CCAAACAA	GCGAAA	GRPYKIKRGVLQGDS
		GATTCATTGCGAATTGGGTTAAGTCGCTAATCCAGCGAACTATTCAAAC	AATCATCC	TCACTTA	LSPALFVIATSPIVSHL
		AGGAGAGTTCCCGGCCTCACTTATGTGTGGTCGTACGGTGCTCCTGTAT	ACAAATCG	ACGCTG	KRVCPSGRIQLYMDD
		AAGAATGGTGACACTGCTAAGCCGGAGAATTATCGACCTATCACATGCC	GAGGGTTT	GTGCAC	IKLYGKSETELRMLIKE
		TGAATGGTTGCTTCAAGATGACTAATGCTGTTATCACGAAAGTAATCGTT	TCTTGGTA	CTCCTGC	VQKVANKLGLQMNL
		CAACGGGTTCAAGACACCTGTGCTCTCCCGGGAGAACAGATGGCCCTG	GTTTCCCT	TATCGTA	KKCSTYGAGLTESIAG
		AAGCCGAAGGTGTGGGCGTGCATGGAAGCGCAGTTGCGCGACCAGGC	TGGCACGT	GTACTTC	FDPLGDRAYKYLGVP
		TCTACAGTCTGAGATCGGTAACGATTGTAAGACTGCGTGGATCGATTTT	CTTGTTTC	CTAGAT	QRSVADTNLAFGELE
		TCAAAAGCTTATGACAGTCTGGACCACGACGCGATCCGATTCGTGATAG	AAAAGCCA	AGATGA	GKVIRSIEETMACEYL

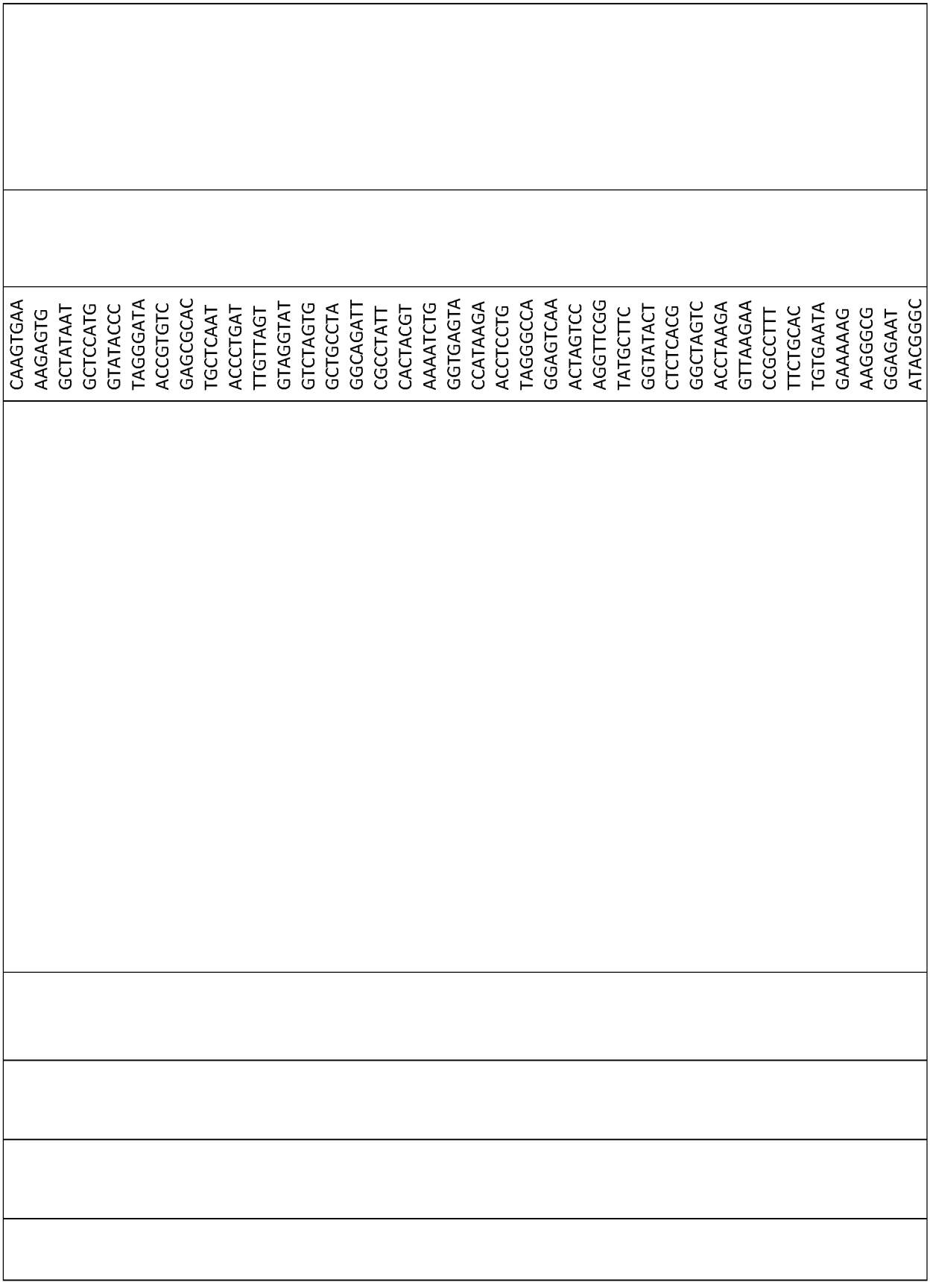
	AAACACTTGCTTTACCGGACGGAATGGAGAAATATCTCCTGAAGTCGCT	GAATATAA	GTAGGG	TMRQVVTRLNSVIGP
	AGAGAGTTGGAAGACAAAACTGGTGCTTAGTAATCGAGGTAAAGTTGC	AACAATAC	TGGGCT	LVRFVAQSVLTSQAK
	CACTGGGCGACCTTACAAGATTAAGAGAGGTGTTCTGCAAGGTGACAG	CGAAGTAC	AAAGGT	VSWIYNKISDLDSKIR
	TCTTAGTCCGGCCCTATTTGTCATTGCGACAAGTCCTATCGTCTCCCATCT	AGAATGGC	AGTCGT	AKLAQTGLRYKKSNV
	AAAGCGTGTTTGCCCATCCGGCCGTATACAATTGTATATGGATGACATT	TGGAACCT	CTTCAAT	ARLYLSKSKNGIGLVN
	AAACTGTATGGCAAATCAGAGACCGAATTACGTATGCTCATTAAAGAAG	AACAGCCA	GGCAAT	VQQVLVEALVSRAIY
	TTCAAAAGGTAGCCAACAAATTGGGGCTCCAGATGAATCTAAAAAAATG	TGGCACAG	ACCGAG	CLRAPSLVEIREHILTA
	TAGTACCTATGGTGCTGGCCTAACAGAGAGTATCGCCGGGTTTGACCCA	AAAACCGG	GGGTCT	EFDPVGAARTVLRRS
	CTCGGCGACAGAGCATACAAGTATCTCGGAGTTCCCCAGCGGAGCGTC	TGGGTGCA	CGTGAC	RIQLEIERVEMASAIS
	GCAGATACTAATTTGGCCTTTGGTGAGCTTGAAGGCAAGGTAATACGGT	CACCGGAG	AGGTGA	AIKTNYQARWMTKF
	CCATCGAGGAAACGATGGCGTGCGAATACTTGACAATGCGGCAGGTGG	ACAGCAG	GGGATA	TQSKLYQKWVHHDI
	TCACAAGATTGAACTCTGTGATCGGTCCTCTTGTCCGGTTTGTAGCACAG	GGGAAAG	ACTAGT	DLANSNLWLERGEIS
	TCCGTACTAACCAGTCAGGCGAAAGTAAGCTGGATATATAACAAAATAT	ATTGCAAA	ATCAGA	PQNARIAVAAQDNT
	CGGATCTGGATAGCAAGATAAGAGCGAAATTAGCTCAAACAGGCCTAA	GTGCCGTA	TAATATC	LLCRGFVGNRESEKQ
	GGTACAAGAAAAGCAACGTAGCAAGACTCTACCTAAGCAAGAGTAAAA	AATAAAGA	AGTCAT	CRMCNMGIETCSHIL
	ATGGCATCGGTTTGGTGAATGTCCAGCAGGTATTAGTTGAAGCATTGGT	TGGTAGCT	GAAATT	TECSYHRAHMYIERH
	GAGCCGAGCCATCTATTGTCTGAGAGCCCCGAGTCTCGTTGAGATTCGG	CTCTGACC	AGCAAC	DSVARNIYAVLAKDH
	GAGCATATACTAACAGCAGAGTTCGACCCTGTAGGAGCTGCTCGTACCG	TGAGGGA	AACCAA	GLWIPHYSQPVSSVT
	TGCTGCGGAGGAGTAGGATCCAACTGGAGATCGAGCGCGTCGAAATGG	ATGGGACA	CGTCAC	KTPTCELYWNYKFPC
	CTTCGGCTATATCGGCGATTAAAACAAATTACCAGGCAAGATGGATGAC	CTGTCCGA	CGTCGT	TRALEACRPDIVLIDR
	GAAGTTCACCCAATCGAAACTATACCAGAAGTGGGTTCACCATGACATC	TTAATGCC	TGGCAA	AKRTILIVEVAVSYVT
	GATCTAGCCAACTCAAACTTATGGTTAGAGAGGGGCGAAATCTCTCCGC	TTATCCCG	AACACC	RLKQMVSRKVYKYG
	AAAACGCAAGGATTGCGGTAGCGGCTCAAGATAATACGTTACTGTGTC	AAGACGGT	GACTAA	VNGEKGADGESRG
	GCGGGTTTGTAGGTAATCGGGAGTCAGAAAAACAGTGCCGCATGTGTA	CAAGGTTC	CGATGC	WNMIRELSEVYNMK
	ACATGGGCATTGAGACCTGCAGTCATATCCTGACTGAATGTAGTTATCA	TATCTTAT	TAGTTA	VNLCAVVIGASGEVL
	CCGGGCGCATATGTATATAGAACGTCACGATTCTGTGGCAAGGAATATC	CGGTCTCC	GAAAGA	PCTVKAIQSISSKTSSR
	TATGCTGTCCTGGCAAAAGACCACGGTCTCTGGATTCCGCATTACTCGCA	AAAGGGCT	GTCGGG	QLLERCQRSAVLGST
	GCCCGTGAGTTCGGTAACAAAGACGCCGACGTGTGAACTTTACTGGAAT	CTGGCCAT	TCTTCCC	RVVKRHLAEFH (SEQ
	TACAAGTTTCCATGTACCCGTGCATTGGAAGCTTGCCGTCCAGATATCGT	AGAAGCTG	AAAGTT	ID NO: 1503)
	TCTCATCGATCGCGCGAAAAGAACTATCCTCATAGTGGAAGTTGCGGTG	CGAGGAA	AGGTGC	
	TCGTATGTCACACGGCTAAAGCAGATGGTGTCACGAAAAGTGTATAAAT	GCCGTAAC	TTGCACC	
	ATGGAGTAAACGGTGAAAAGGGAGCCGACGGCGAAAGCCGCGGCTGG	CCTACGCA	GAAGCC	
	AATATGATCCGTGAACTATCTGAAGTTTATAATATGAAAGTGAACCTTTG	GATTGCCG	GATCCG	
	CGCAGTCGTGATTGGGGCGAGTGGGGAAGTTCTCCCTTGTACGGTCAA	CCGTGC	CTCTACC	
	AGCCATTCAGAGCATATCTAGCAAGACCTCCAGCCGACAGCTCCTAGAG	(SEQ ID	CACAGC	
	CGGTGCCAGCGTAGCGCTGTCCTAGGGTCGACGAGGGTTGTCAAACGA	NO: 1258)	TCTGCCC	
	CATCTTGCAGAATTCCATTGAGACCACCCATGCGCAGAGTATCCGAATC		AGCGTT	
	AGTGAAAGTCCAAGTTTCAGGACAGAAACGTCAGATAAGTCCAAGAGA		(SEQ ID	

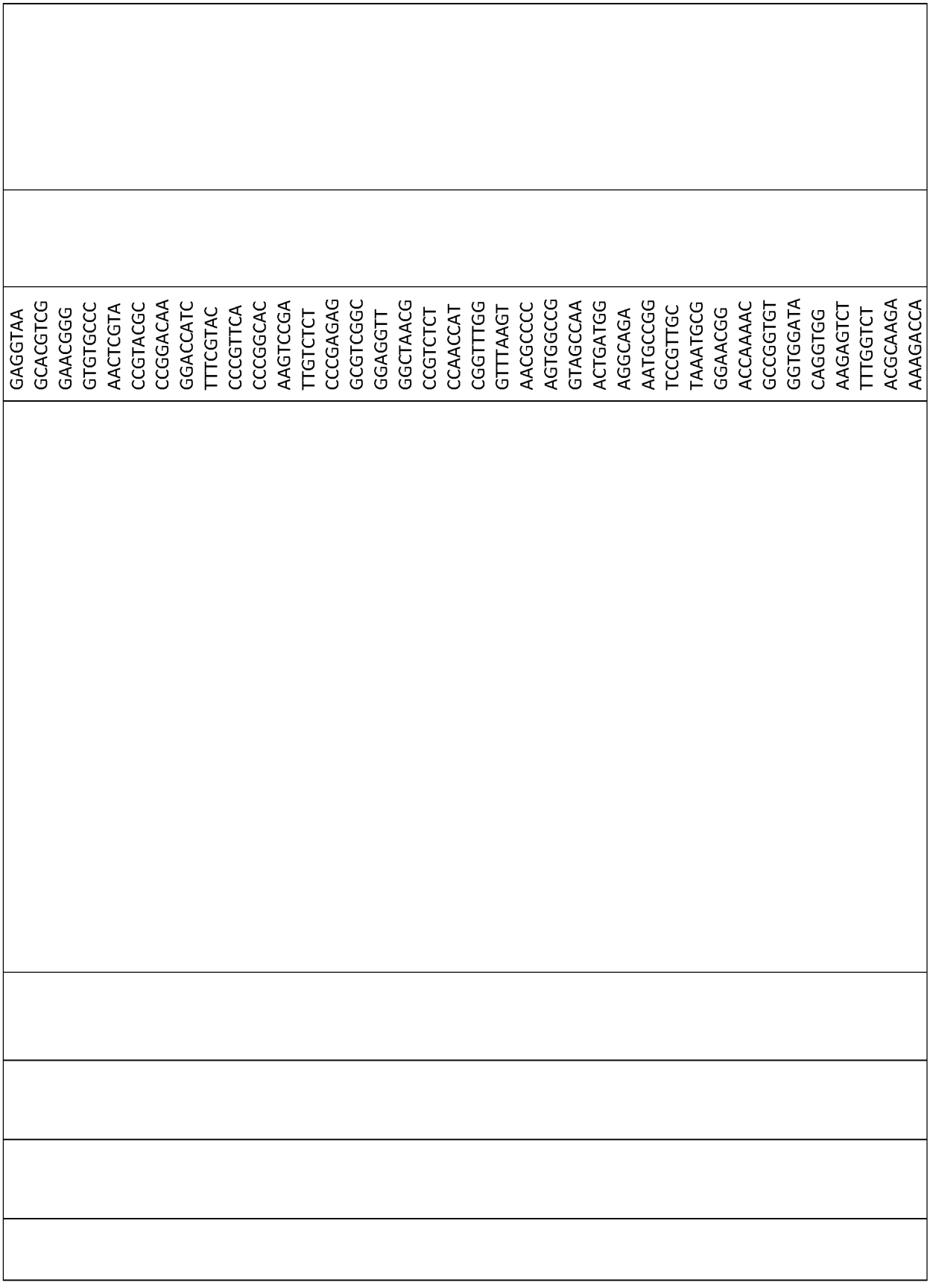
			AACGAGAAAACAAGTTCAAGTATGCAAGAGTTAATCAATAAGAGAGTA			
			CCGTAAATGTATGACCCCCCCCCTTTGCCAAGTCGACAACTGTCATGCAG			
			GTGTCTCTCTTTTCACCCGCCATATGGACCAAACGCTATCCAGCCTCGCT			
			CAGAAGAGCCTTAGGGCTGGGGAGTACCACATGTGGCGGAAACTGAAT			
			CTGGATGCGATGCATACCGGGAGCGCAGCGAAATCACTTAACGCTGGT			
			GCACCTCCTGCTATCGTAGTACTTCCTAGATAGATGAGTAGGGTGGGCT			
			AAAGGTAGTCGTCTTCAATGGCAATACCGAGGGGTCTCGTGACAGGTG			
			AGGGATAACTAGTATCAGATAATATCAGTCATGAAATTAGCAACAACCA			
			ACGTCACCGTCGTTGGCAAAACACCGACTAACGATGCTAGTTAGAAAGA			
			GTCGGGTCTTCCCAAAGTTAGGTGCTTGCACCGAAGCCGATCCGCTCTA			
			CCCACAGCTCTGCCCAGCGTT (SEQ ID NO: 1135)			
R4	R4-4_BX	.	TGCCAGCGGTGTTGATTAGGTCCAAGTTCTTTGGCCAAAGATCCGCCCT	TGCCAGCG	TAATTA	MEILWEDLRLKIEDR
			CGGTTTAGCAGTACCGAACGAGTATACCTTCAAGTGGTGGCACTGAATT	GTGTTGAT	GGCAGT	YGVTLPQRSASSLKN
			AGACTGAATACTCTGAACTGTAGACTTTTGTGCAACTGTGTATGGTGTG	TAGGTCCA	GCTCCTA	QYPKVILRGLPDSGLP
			GAAGACTTGTTTGTACCACTATCAGCTTTATTGGGGCTCGTTACTGTTTC	AGTTCTTT	GCGGAG	WAGVQVNDTGQVV
			ATACAGGTAGATGTCCCCTTTAGAGATTTCCCTGCAGTTTGCGCTCCAAG	GGCCAAA	TGCCGT	VVDHAEAATLRGSSP
			TCGCTAGCCTCTTGCGTGTAGTCAAAGGAATACATTCGCCGTCGGTGAC	GATCCGCC	GAAGTG	AAVDGEAEEPVVPPL
			AGGGCTATACCCGGCGACTACGGACTTGTTTATTACGTAGTGCAGCCTC	CTCGGTTT	GTGTCA	PAAEVVESAADAAVP
			GTTTAAGACGAATGTGAAAAGAAGGTGTGATTACTAAGCGTTATGAGTC	AGCAGTAC	GTACCC	DPQSEIVADQGVETR
			GGGTTATCTGGAAACTCCGCCCCGCCGCAATGGCTTCAAGGGTCAAACT	CGAACGA	GTCGTG	PVENPPANSRETETE
			GCTAACATTTTAAACCAATATTTGGCAGCAGCAGACTATGATGTCAGAG	GTATACCT	TGGAAA	PVEVEPYLEGQYKFF
			TGCCGGGGGTCTATCGATATAAGACTGAAAGCGATAGACGTGGAGTGA	TCAAGTGG	GCCCAG	VSKILGKSMWRKPIK
			AAGGATTCCCTTTGTTAAGAGAATCGGTAGAATTCACTTTTTACTTATTC	TGGCACTG	GAGGGT	YPRRVPETLWRQAN
			AAGAACTTAACAGCAACAAGCACTCGCGAGGATTACCGCCCCAGATTCG	AATTAGAC	TAGTAC	ELIERSIRQGEVSIQSL
			GTCGGCGGTACTTCACCTGCTTTCTTCCACTTTCGGAATCTGGCATCCTG	TGAATACT	CGACAG	NCMVYAAGCAVKSS
			GCTTTCAGTGTGGTGATGGCCGGCTTGAGTTTTCTTGAGTCGTGCGAGT	CTGAACTG	TGGGAA	LDKKDQEAKRRESE
			GCCTCATCTGGGACGTCCGGACCGATTGATGGAGTCTGCAGTGGACGA	TAGACTTT	ACCCGC	WYACRKAEIKALERY
			GGACTTGATGGACCGTAACCATAGTATATCCCTCATGCGTCTTCTCGACT	TGTGCAAC	TGCACG	LNFIDLELKRRSASRP
			CGAGGGGGTAGCTTGCACTACCCACCCTTCTCTTCCCGATTGGGATTTA	TGTGTATG	CAACCT	LTSRQRQNLGVLITKY
			GACCTAGCCCTCTGGTGTGTCTCGACCCGCGATATCAGATTCCTGAATCG	GTGTGGAA	AAAGAC	GRARVRSGVRLSELQ
			ACTGTGAGAAATGTCTACGCGCAAAGATCGACCCATTGCCACCCGGCTA	GACTTGTT	ATTTGCC	AMLRDALVGIRKCM
			TGTGGATCGGGCTCTTGACTGCTTATCTCCGGCTTTAATCGCTTGAGGAA	TGTACCAC	CTTCGG	AKRSADKKRKQGKFV
			AGGGGGGTGTCGCCCGAAAGGGTTACGCGATCATCCGATTCCTACCGT	TATCAGCT	GGGAGA	PIQRYLEPSSAEPRLS
			AAAAACGTAGTTGAAGTAGGATTGAACCTGAGTACCAAGTGAAAAGAG	TTATTGGG	GGTATG	PDTVRAYWNDIVGS
			TGGCTATAATGCTCCATGGTATACCCTAGGGATAACCGTGTCGAGCGCA	GCTCGTTA	AGACCA	SQQSTSDSTIQDWSS
			CTGCTCAATACCCTGATTTGTTAGTGTAGGTATGTCTAGTGGCTGCCTAG	CTGTTTCA	CACTAG	NLSVPSQELNASKIM
			GCAGATTCGCCTATTCACTACGTAAAATCTGGGTGAGTACCATAAGAAC	TACAGGTA	TCATGG	GWWRAAVSKSKPN
			CTCCTGTAGGGCCAGGAGTCAAACTAGTCCAGGTTCGGTATGCTTCGGT	GATGTCCC	TTGCTTG	KAAGPDGIPGVLWK

	ATACTCTCTCACGGGCTAGTCACCTAAGAGTTAAGAACCGCCTTTTTCTG	CTTTAGAG	CGCAAG		RFRSASEWVCTWLY
	CACTGTGAATAGAAAAAGAAGGGCGGGAGAATATACGGGCGAGGTAA	ATTTCCCT	CATGAC		RLLQKRRIITPRWLSV
	GCACGTCGGAACGGGGTGTGCCCAACTCGTACCGTACGCCCGGACAAG	GCAGTTTG	TCATTCA		GRVVLLPKKGPLEDP
	GACCATCTTTCGTACCCCGTTCACCCGGCACAAGTCCGATTGTCTCTCCC	CGCTCCAA	TTGTAA		ANYRPIACLNTVYKLI
	GAGAGGCGTCGGCGGAGGTTGGCTAACGCCGTCTCTCCAACCATCGGT	GTCGCTAG	GTTCGT		TSVVEMAVREQIQA
	TTGGGTTTAAGTAACGCCCCAGTGGCCGGTAGCCAAACTGATGGAGGC	CCTCTTGC	ATTATG		CPGLVPYEQIANRKG
	AGAAATGCCGGTCCGTTGCTAAATGCGGGAACGGACCAAAACGCCGGT	GTGTAGTC	AGGCCT		VWGCTHASIVDRMI
	GTGGTGGATACAGGTGGAAGAGTCTTTTGGTCTACGCAAGAAAAGACC	AAAGGAAT	CGCGGA		TGASREGKGGGFPDL
	AGGCTCAAGTACGAATACGACCGTCTTCGTGAACGACGCCGTCGAGAC	ACATTCGC	ATGCGA		RVLFYDCKKAFDSVN
	AAAGTTCGAAATCCTACGGTCGATAAGGCCATGGAAATCCTTTGGGAAG	CGTCGGTG	AAGTCC		RDHMFAVLRVANVN
	ATCTCCGGTTGAAAATTGAAGATAGGTACGGTGTCACATTGCCGCAAAG	ACAGGGCT	AAGAGC		VKVVHLLHTLSQQW
	AAGCGCAAGCTCTTTGAAAAACCAGTACCCCAAAGTTATTCTGCGGGGG	ATACCCGG	GCCAAC		CVRYELRRNNRVERS
	CTCCCAGATTCGGGCCTGCCTTGGGCGGGGGTTCAGGTAAATGATACTG	CGACTACG	TTGTCTC		SPLRVKRGLLQGDTL
	GACAAGTCGTAGTCGTAGACCATGCTGAGGCGGCTACCCTTCGGGGGT	GACTTGTT	TTTGGG		SPTWFCLCMAPISASI
	CCAGCCCAGCAGCGGTCGACGGGGAGGCCGAAGAACCGGTAGTACCCC	TATTACGT	AGGTTC		KTLNPGPTLRPNNMG
	CGTTACCCGCGGCAGAGGTCGTTGAATCGGCTGCTGACGCTGCGGTGC	AGTGCAGC	CTCCGG		NGRNRGQVAIQVSH
	CCGATCCGCAAAGCGAGATCGTGGCTGACCAGGGTGTGGAGACACGCC	CTCGTTTA	GTTCCTT		VFYMDDLKVYCPRV
	CTGTCGAGAACCCCCCCCGCTAATAGCCGGGAGACTGAAACGGAACCCG	AGACGAAT	AGGGTT		ADQRRMEQNIPQLF
	TAGAGGTAGAGCCTTACCTCGAAGGGCAGTATAAATTCTTCGTTTCTAA	GTGAAAA	GGCGTA		GEIGLSINASKSAAAA
	GATCCTTGGGAAATCGATGTGGCGGAAGCCTATCAAATACCCGAGGAG	GAAGGTGT	CCATGCT		AVGRYVESELPVLGT
	AGTCCCGGAAACCCTGTGGCGTCAAGCCAATGAGTTGATTGAGAGATC	GATTACTA	TCGTGT		KDEYKYLGIESGFVVN
	GATTCGACAGGGGGAAGTGTCGATTCAGAGCCTGAACTGCATGGTTTAT	AGCGTTAT	GAGACC		EVAALDRMQAVLLN
	GCTGCTGGGTGCGCTGTTAAGTCTTCGCTGGACAAGAAGGATCAGGAA	GAGTCGG	TAGGCC		RVEAILSVKEHTVGQ
	GCTAAGCGGCGTGAATCCGAGTGGTATGCGTGTAGGAAGGCCGAGATT	GTTATCTG	GCTTGG		RRDAIRAKAIPGGAYI
	AAGGCACTCGAGAGGTATCTCAATTTCATTGACCTGGAGTTAAAGCGTC	GAAACTCC	ACCTAG		LGHIILSDLDPRGAAE
	GCTCTGCGTCGCGGCCGTTGACATCAAGGCAAAGGCAGAACTTGGGTG	GCCCCGCC	AGTTAA		RMRRLDIEIRRLVKSA
	TGCTCATCACGAAGTATGGGCGTGCTCGGGTTAGGTCTGGTGTCCGGCT	GCAATGGC	GCTGTT		GILHDKCSTARIHLSC
	CTCGGAATTGCAGGCAATGTTACGAGAGACGCCTTAGTGGGAATCCGAAA	TTCAAGGG	GCATGT		EQGGLAWPSMERAY
	GTGTATGGCGAAGCGCTCGGCAGACAAGAAACGGAAACAGGGCAAATT	TCAAACTG	TTGAGG		YVAVAYSASYLLTSQ
	TGTTCCTATTCAACGTTACTTAGAGCCTAGCAGTGCGGAGCCTCGACTCT	CTAACATT	ATGCCTT		DETISRARDYFVSGRL
	CTCCAGACACGGTTCGTGCATATTGGAATGACATTGTGGGATCTTCCCA	TTAAACCA	AGGCGG		SNKFTVYKHLTSIVDS
	ACAGTCAACGTCAGATTCGACGATTCAAGACTGGTCCTCCAATTTAAGC	ATATTTGG	ACGCCA		LGLSVELPDPNGLPT
	GTTCCGAGTCAAGAGTTGAACGCAAGTAAGATCATGGGGTGGTGGAGA	CAGCAGCA	TG (SEQ		GQPSVLARTIARAIDA
	GCCGCCGTTTCGAAGAGCAAGCCCAACAAAGCTGCTGGTCCTGATGGTA	GACTATGA	ID NO:		KLEAQWKETLLTYQR
	TTCCCGGGGTACTGTGGAAACGCTTCAGGTCGGCTAGTGAGTGGGTTTG	TGTCAGAG	1382)		AGRVERADPTVVDH
	TACCTGGTTATATCGCTTATTACAGAAGCGAAGGATAATCACGCCAAGA	TGCCGGG			ANSYHWLRKAWINE
	TGGTTGTCTGTTGGTAGAGTCGTATTGTTACCAAAGAAAGGACCTCTCG	GGTCTATC			KAYQHAVSVMEGTL
	AGGATCCCGCCAACTATAGACCCATCGCGTGCCTGAACACTGTATATAA	GATATAAG			LEGVNPHGVLTMCR

		GTTGATAACCTCGGTAGTAGAAATGGCAGTTCGAGAGCAGATCCAGGC	ACTGAAAG		ACKAPSASIAHIITGC
		TTGTCCTGGACTAGTCCCATACGAGCAGATAGCTAACCGTAAAGGCGTA	CGATAGAC		AELRKSHMKVRHDG
		TGGGGCTGTACGCATGCGAGTATCGTAGACCGTATGATAACCGGTGCA	GTGGAGT		VTRWLYNALTEVDGS
		AGCCGAGAAGGTAAGGGCGGTGGGTTTCCGGACCTACGCGTCCTATTTT	GAAAGGA		LPKFHYTQQIPAEMR
		ACGATTGTAAAAAGGCCTTCGACAGCGTAAACCGAGATCATATGTTCGC	TTCCCTTT		GERLTVRYDSDIVTP
		TGTTTTGAGAGTAGCAAACGTGAACGTTAAGGTCGTGCATCTACTCCAC	GTTAAGAG		NKPRHNRPDLVVFDS
		ACACTATCTCAGCAATGGTGTGTCCGATACGAGCTCCGTCGAAACAACC	AATCGGTA		TRKVIYIVEVSVTWLS
		GTGTGGAGAGAAGTAGCCCGCTGAGGGTGAAAAGAGGACTGCTGCAA	GAATTCAC		VLQKQYDNKLNRYA
		GGCGACACTCTCTCCTACATGGTTTTGCCTATGCATGGCTCCAATATC	TTTTTACTT		VNSNHEFSESIPYPPG
		AGCCTCTATTAAAACCCTTAACCCTGGCCCAACCCTTCGGCCGAACATGG	ATTCAAGA		VNLANEIRVLYPQFT
		GGAATGGGCGAAATAGAGGGCAGGTGGCGATCCAAGTTAGTCATGTGT	ACTTAACA		GGVKVFPMIISPTGE
		TCTACATGGATGACTTGAAAGTGTATTGCCCCCCGCGTAGCCGATCAACG	GCAACAAG		VHMQFVPHLAELLE
		CAGAATGGAACAGAATATTCCACAGTTGTTCGGTGAAATTGGACTAAGC	CACTCGCG		NPNIPRILEKIQRSVVL
		ATAAATGCGAGTAAGAGCGCAGCAGCCGCAGCCGTGGGAAGATACGTT	AGGATTAC		GTDYIIRSYFAM (SEQ
		GAATCCGAGTTGCCAGTTCTTGGGACGAAGGATGAGTACAAGTACCTG	CGCCCCAG		ID NO: 1504)
		GGAATAGAAAGCGGTTTTGTTGTAAATGAGGTGGCCGCCCTTGACAGA	ATTCGGTC		
		ATGCAGGCCGTTCTGCTAAACCGTGTAGAAGCGATTCTATCGGTTAAAG	GGCGGTAC		
		AACATACTGTTGGTCAGAGGCGCGACGCTATACGAGCGAAAGCGATCC	TTCACCTG		
		CTGGGGGGGCCTATATCCTGGGTCACATCATTTTGAGTGACTTGGACCC	CTTTCTTCC		
		CAGGGGTGCTGCTGAGCGCATGCGTCGGTTGGATATCGAGATTCGGCG	ACTTTCGG		
		TTTAGTCAAATCAGCTGGTATTCTTCACGACAAGTGTAGTACAGCGCGT	AATCTGGC		
		ATTCACCTTAGCTGTGAACAGGGAGGTCTAGCTTGGCCGTCGATGGAGA	ATCCTGGC		
		GAGCTTATTATGTAGCAGTTGCATATTCAGCCTCCTATCTTCTGACAAGT	TTTCAGTG		
		CAAGACGAAACTATCAGTCGTGCCAGAGACTACTTCGTCTCGGGTCGAT	TGGTGATG		
		TGAGTAACAAGTTTACTGTTTACAAGCACCTTACGAGCATCGTAGACTCC	GCCGGCTT		
		CTTGGTTTATCGGTAGAACTGCCGGATCCGAATGGACTTCCTACCGGTC	GAGTTTTC		
		AACCGTCCGTCCTGGCTAGAACTATTGCGAGAGCGATAGATGCCAAGTT	TTGAGTCG		
		GGAAGCGCAGTGGAAGGAAACTTTGCTAACTTACCAGCGCGCAGGGCG	TGCGAGTG		
		AGTCGAAAGGGCAGATCCGACGGTCGTGGACCATGCAAACTCATACCA	CCTCATCT		
		TTGGTTGAGAAAAGCATGGATAAACGAGAAAGCGTACCAACATGCCGT	GGGACGTC		
		GAGTGTTATGGAGGGAACTCTTTTGGAGGGAGTGAACCCTCACGGTGT	CGGACCGA		
		GTTGACCATGTGTAGAGCGTGTAAAGCTCCAAGTGCAAGCATTGCTCAC	TTGATGGA		
		ATCATTACTGGCTGCCGCCGAATTGCGCAAAAGCCATATGAAGGTGAGG	GTCTGCAG		
		CATGATGGAGTTACTCGCTGGCTATACAATGCTCTGACTGAGGTAGATG	TGGACGA		
		GTAGCCTACCGAAGTTCCACTATACGCAACAGATCCCGGCAGAGATGAG	GGACTTGA		
		GGGTGAGCGACTCACTGTCCGTTACGACTCGGACATAGTTACTCCCAAT	TGGACCGT		
		AAACCTCGTCACAACAGACCGGATCTTGTTGTCTTCGACTCTACTAGAAA	AACCATAG		
		GGTGATTTATATAGTAGAGGTCAGTGTAACGTGGTTGAGTGTGTTGCAA	TATATCCC		

			AAGCAATACGATAACAAGCTCAACAGGTATGCAGTGAACAGTAATCAC	TCATGCGT	
			GAGTTCAGTGAGAGCATACCATATCCGCCTGGTGTGAACCTTGCTAACG	CTTCTCGA	
			AGATAAGGGTTTTGTATCCTCAGTTTACTGGGGGTGTGAAGGTATTCCC	CTCGAGGG	
			TATGATTATCTCGCCAACAGGGGAAGTACACATGCAATTTGTGCCCCAC	GGTAGCTT	
			CTTGCAGAGTTACTCGAAAATCCGAACATTCCTCGGATTCTAGAGAAAA	GCACTACC	
			TTCAACGCTCGGTTGTGCTTGGTACAGATTACATTATCCGGTCATACTTC	CACCCTTC	
			GCCATGTAATTAGGCAGTGCTCCTAGCGGAGTGCCGTGAAGTGGTGTC	TCTTCTCC	
			AGTACCCGTCGTGTGGAAAGCCCAGGAGGGTTAGTACCGACAGTGGGA	GATTGGGA	
			AACCCGCTGCACGCAACCTAAAGACATTTGCCCTTCGGGGGAGAGGTAT	TTTAGACC	
			GAGACCACACTAGTCATGGTTGCTTGCGCAAGCATGACTCATTCATTGT	TAGCCCTC	
			AAGTTCGTATTATGAGGCCTCGCGGAATGCGAAAGTCCAAGAGCGCCA	TGGTGTGT	
			ACTTGTCTCTTTGGGAGGTTCCTCCGGGTTCCTTAGGGTTGGCGTACCAT	CTCGACCC	
			GCTTCGTGTGAGACCTAGGCCGCTTGGACCTAGAGTTAAGCTGTTGCAT	GCGATATC	
			GTTTGAGGATGCCTTAGGCGGACGCCATG (SEQ ID NO: 1136)	AGATTCCT	
				GAATCGAC	
				TGTGAGAA	
				ATGTCTAC	
				GCGCAAA	
				GATCGACC	
				CATTGCCA	
				CCCGGCTA	
				TGTGGATC	
				GGGCTCTT	
				GACTGCTT	
				ATCTCCGG	
				CTTTAATC	
				GCTTGAGG	
				AAAGGGG	
				GGTGTCGC	
				CCGAAAG	
				GGTTACGC	
				GATCATCC	
				GATTCCTA	
				CCGTAAAA	
				ACGTAGTT	
				GAAGTAG	
				GATTGAAC	
				CTGAGTAC	


CAAGTGAA
AAGAGTG
GCTATAAT
GCTCCATG
GTATACCC
TAGGGATA
ACCGTGTC
GAGCGCAC
TGCTCAAT
ACCCTGAT
TTGTTAGT
GTAGGTAT
GTCTAGTG
GCTGCCTA
GGCAGATT
CGCCTATT
CACTACGT
AAAATCTG
GGTGAGTA
CCATAAGA
ACCTCCTG
TAGGGCCA
GGAGTCAA
ACTAGTCC
AGGTTCGG
TATGCTTC
GGTATACT
CTCTCACG
GGCTAGTC
ACCTAAGA
GTTAAGAA
CCGCCTTT
TTCTGCAC
TGTGAATA
GAAAAAG
AAGGGCG
GGAGAAT
ATACGGGC


GAGGTAA
GCACGTCG
GAACGGG
GTGTGCCC
AACTCGTA
CCGTACGC
CCGGACAA
GGACCATC
TTTCGTAC
CCCGTTCA
CCCGGCAC
AAGTCCGA
TTGTCTCT
CCCGAGAG
GCGTCGGC
GGAGGTT
GGCTAACG
CCGTCTCT
CCAACCAT
CGGTTTGG
GTTTAAGT
AACGCCCC
AGTGGCCG
GTAGCCAA
ACTGATGG
AGGCAGA
AATGCCGG
TCCGTTGC
TAAATGCG
GGAACGG
ACCAAAAC
GCCGGTGT
GGTGGATA
CAGGTGG
AAGAGTCT
TTTGGTCT
ACGCAAGA
AAAGACCA

					GGCTCAAG TACGAATA CGACCGTC TTCGTGAA CGACGCCG TCGAGACA AAGTTCGA AATCCTAC GGTCGATA AGGCC (SEQ ID NO: 1259)		
R4	R4_AL	U2944 5	Ascaris lumbric oides	GGGGCCGGTGGGTTTACTCACTTCTGACCCACCACCAACGGAACGAGG	GGGGCCG	TAGTCG	MPCSTNSFFERGTPE
				GAAAGCAGAGCTGGGGCCCTCTTCCGATTGGCATGGAACCGACCTCCAC	GTGGGTTT	CTAAGG	PHREPISGTDSSESLG
				GTGGTGGCCCTGGGCAACGGAATTCAAGAGAGGATTTAATCCTCTCTAT	ACTCACTT	GGTCCG	MGTHRSPRLNDDEVI
				CATTTGCAAGATGGATGAGATCGAGGTATCCGGCAAACAGGTTCCAAG	CTGACCCA	GAAATG	NGPKGHESDPVHVV
				TGAGCACCTTTCCCATAGCTGGGAATATGGGTTAGGCGTCCTCTGACAT	CCACCAAC	GTCCGG	RAPRTLHPRRLELPIG
				ATAAGAGGAATCAGACTCGTTCGCGCCCGGTCATTAACATCGATCAGCG	GGAACGA	TCCTGC	VNNLGEASQLRQDS
				GGAGGGCCGGACTGAAGTAAATTTCCTGTTGGCCCGAGTGCAGGTGGA	GGGAAAG	GCTACC	AIAEEAQLESTENHD
				GCTCGGACCCGAAAAACGATCCCTAAGAGGACCACAACCCGAAGGGAT	CAGAGCTG	CGGTTC	GRRPPLRGGRKLWS
				GGACGCAGTCGCCCCGGCACGTTTGGTGTTGGTTATCCTGGAGTGTTGT	GGGCCCTC	TGGTAG	EKEIATLRRLCEAYGN
				GGGACGAATAGCTATGCCTTGTTCAACTAACTCTTTTTTTGAACGAGGG	TTCCGATT	CACGTTC	RQVCWKEVQRKFAD
				ACCCCTGAGCCCCATCGTGAGCCAATCTCAGGGACTGACTCTTCGGAGT	GGCATGG	AAGCGC	FHEERTVAALATKW
				CCTTAGGTATGGGTACCCACCGTTCACCTCGGCTAAATGATGACGAGGT	AACCGACC	TCAATC	GALKRPRAPMVGAP
				GATAAACGGCCCCAAGGGTCACGAAAGTGACCCGGTTCATGTAGTGCG	TCCACGTG	GCCTGC	PTPDHDPERGPAGE
				TGCGCCTCGAACGCTACACCCGAGGCGATTGGAACTCCCAATCGGAGTA	GTGGCCCT	CTTGTA	GDGGTTSQENVPTD
				AATAATCTCGGAGAAGCCTCTCAACTACGGCAAGATTCAGCAATAGCTG	GGGCAAC	GGCAGT	DPIPANGPTEGKESD
				AGGAAGCCCAATTGGAGAGCACCGAGAACCATGATGGACGTCGCCCTC	GGAATTCA	CCATCTG	VRPAVACRCTEPEEQ
				CTTTGAGAGGCGGCAGGAAGCTCTGGAGTGAAAAGGAGATCGCCACAT	AGAGAGG	TGGAAG	LMESDVRPPAVVRLA
				TAAGGAGATTGTGTGAGGCCTATGGAAATAGGCAGGTATGCTGGAAAG	ATTTAATC	TCGCGC	DPEQHTMKSGVKPV
				AGGTTCAGAGGAAGTTTGCGGACTTCCACGAGGAAAGAACAGTGGCTG	CTCTCTAT	TCTTGAT	ALDGSADLEERPKEK
				CGTTAGCCACGAAGTGGGGGGGCTCTTAAACGCCCGAGAGCTCCGATGG	CATTTGCA	ACAGAT	DIEQMGVDFEGEPRF
				TTGGTGCGCCACCGACTCCTGATCATGATCCAGAGCGTGGACCAGCGG	AGATGGAT	GTGGAC	RAFRKAFYGYFRWA
				GCGAGGGAGATGGAGGTACAACCTCGCAAGAGAATGTGCCTACCGATG	GAGATCGA	GGATGG	VNSFDREPVKRVRRD
				ATCCAATACCCGCGAACGGCCCGACAGAGGGCAAGGAATCGGATGTGA	GGTATCCG	AAGCAG	CPKVFYAYADYLIATG
				GACCAGCCGTGGCCTGTAGGTGCACGGAACCGGAGGAGCAACTGATG	GCAAACAG	ATGATA	SSKALGPNQSRIGRL
				GAGTCGGATGTGAGACCGCCTGCAGTCGTAAGACTTGCAGATCCGGAG	GTTCCAAG	GAGCCG	NGLVYAAARTIHQF
				CAGCATACGATGAAGTCGGGTGTGAAACCGGTTGCACTTGACGGGTCA	TGAGCACC	GTGACG	WREEVGHRQQGEK

	GCGGATCTGGAGGAGCGTCCGAAAGAGAAGGATATCGAGCAGATGGG	TTTCCCAT	GCCCTA			GWYTKTKATREDLQ
	TGTTGACTTTGAAGGGGAACCGAGGTTTCGTGCCTTTCGGAAAGCCTTT	AGCTGGG	CTAGCC			MLISMMESELARRKE
	TACGGGTATTTCAGATGGGCTGTCAACTCGTTTGATAGGGAACCTGTCA	AATATGGG	AAACGC			KRKPGAKELENIHKLV
	AGCGAGTTAGGCGGGACTGTCCCAAGGTATTTTACGCTTATGCGGATTA	TTAGGCGT	(SEQ ID			ARLGTRSTSGIVRRLE
	CCTTATCGCAACCGGAAGTTCTAAGGCGTTGGGACCGAATCAATCCAGG	CCTCTGAC	NO:			MTRQRLKLLEDRISLH
	ATTGGTCGCTTGAACGGATTGGTCTATGCGGCAGCTAGAACAATCCACC	ATATAAGA	1383)			EQEKRRKRLRKQFAE
	AATTCTGGAGAGAAGAGGTAGGCCATCGTCAGCAAGGCGAGAAGGGA	GGAATCAG				TPSLKLLTKGAKDRG
	TGGTACACGAAAACCAAGGCAACCCGTGAAGACCTTCAGATGCTCATCT	ACTCGTTC				DTMVTMKSVMDFW
	CTATGATGGAATCTGAACTTGCAAGAAGAAAGGAGAAGAGGAAGCCCG	GCGCCCGG				RPIIGRRVTSNPDQL
	GCGCGAAAGAGCTGGAGAAATATCCATAAGCTAGTTGCAAGACTTGGAA	TCATTAAC				QVLRDWRDEQKKAY
	CGCGCAGCACATCGGGCATCGTCAGAAGGCTGGAGATGACAAGACAGA	ATCGATCA				PADLDLEKADLEEKYE
	GGCTCAAACTGTTGGAAGACAGAATAAGTTTGCATGAGCAAGAGAAGA	GCGGGAG				GAIRRIQPWKAPGPD
	GGCGGAAACGATTGCGCAAGCAATTTGCGGAAACCCCTTCTCTAAAATT	GGCCGGA				GLHAHWWKALPSAK
	ACTCACAAAAGGAGCCAAGGATAGGGGCGATACGATGGTAACCATGAA	CTGAAGTA				RLLGELVVDWLTTGK
	ATCTGTAATGGACTTCTGGAGACCAATAATTGGCAGACGAGTTACCTCC	AATTTCCT				VTTGWMCRGRTILIP
	AATCCGGACCAATTGCAAGTCTTGAGAGACTGGAGAGATGAGCAGAAG	GTTGGCCC				KKGDRGDPSNYRPIT
	AAGGCTTATCCGGCAGACCTGGATTTAGAAAAGGCCGATCTTGAGGAG	GAGTGCA				CLNTCYKVLTSVMNS
	AAATATGAGGGAGCAATCAGGAGAATCCAACCGTGGAAAGCTCCCGGT	GGTGGAG				VILSHLSRGEALPMN
	CCCGACGGATTACATGCCCACTGGTGGAAAGCTTTACCGTCGGCCAAGA	CTCGGACC				QRAMRKREWGCTH
	GGCTACTGGGTGAACTGGTGGTTGATTGGCTGACAACAGGTAAGGTTA	CGAAAAAC				AMVLDRAMVMDA
	CCACTGGCTGGATGTGCCGAGGGAGGACAATCCTGATCCCTAAGAAGG	GATCCCTA				MAQKKHSLSVAWLD
	GTGATAGGGGCGACCCTTCTAATTACCGACCTATAAACATGTTTAAATACA	AGAGGAC				YRKAYDSVSHEYIRW
	TGTTATAAGGTGCTAACATCGGTAATGAATTCAGTCATTCTGAGTCACCT	CACAACCC				AINSVNIPRSVQLTLK
	GAGCAGAGGCGAAGCTTTACCAATGAACCAGCGAGCAATGCGGAAACG	GAAGGGA				RLMSDWETRFESTQ
	CGAGTGGGGTTGCACCCACGCTATGGTCCTTGACAGGGCCATGGTAAT	TGGACGCA				CRPKLRSDKMKVLN
	GGATGCAATGGCTCAAAAGAAACACTCATTAAGTGTGGCCTGGCTTGAC	GTCGCCCC				GIFQGDSLSPTLFVLCI
	TATCGTAAAGCATACGATAGTGTGTCGCATGAATATATTCGCTGGGCGA	GGCACGTT				APISYALNKGVGQCQ
	TTAACTCCGTGAATATACCCCGGAGTGTTCAGCTGACGCTCAAGAGGCT	TGGTGTTG				SSSGWSAGYGFEIGH
	CATGAGTGACTGGGAGACACGCTTTGAGTCGACGCAATGCCGGCCGAA	GTTATCCT				QFYMDDLKLYARTPA
	GTTAAGGTCTGACAAAATGAAAGTGCTGAATGGCATCTTTCAAGGTGAC	GGAGTGTT				MLDSQIQVVSEVSEA
	TCGTTATCACCAACCCTTTTTGTATTATGCATAGCACCTATCAGCTACGCA	GTGGGAC				MGLHLNLSKCAKAH
	CTCAATAAGGGTGTCGGCCAGTGTCAATCCTCATCCGGCTGGAGTGCAG	GAATAGCT				YAPHGAGGAQEAVE
	GTTACGGTTTTGAGATTGGACATCAGTTCTATATGGACGATCTTAAACTG	(SEQ ID				GAEGSRKGEIPILGLR
	TACGCTAGGACGCCTGCGATGCTAGACTCCCAAATCCAGGTGGTGTCTG	NO: 1260)				STYKYLGVEQRLLPM
	AGGTGTCGGAAGCAATGGGACTCCATTTGAATTTGAGTAAATGTGCGA					EVALKEFEDKFMDRA
	AAGCACATTATGCTCCGCATGGGGCGGGCGGAGCGCAAGAAGCTGTGG					ETIFASELTWGQMAT
	AAGGTGCAGAAGGATCGAGGAAGGGAGAAATCCCGATACTCGGGCTTC					AYNTIAIAGLRYVYSN
	GAAGCACCTATAAATATCTTGGAGTGGAACAACGACTCCTTCCGATGGA					TNGASPKLLEALKRA

			AGTAGCCCTCAAAGAGTTCGAGGATAAGTTTATGGATCGAGCAGAAAC				ATLDTRIRDLLRRHKC
			AATCTTCGCTAGCGAACTCACATGGGGGCAGATGGCCACAGCGTATAAT				RFRNSFVERLYIPREC
			ACTATAGCTATCGCTGGTCTACGATACGTCTATAGTAATACAAATGGAG				GGYGLKSVEDTLRESI
			CATCACCAAAGCTTCTAGAAGCCCTGAAAAGGGCGGCCACCTTAGACAC				LATWSYIATNPHLAG
			GCGGATAAGAGATCTTTTGAGGCGACATAAATGTCGGTTTCGAAATAGC				QQYFFERLAARGKRT
			TTTGTCGAGAGGCTGTATATCCCTAGAGAATGCGGTGGATACGGGTTAA				PMADGVKILLDLGVE
			AATCAGTGGAGGATACGCTGCGAGAGAGTATCCTCGCTACGTGGAGTT				PQVDLKRRTVTVDGI
			ACATCGCCACGAACCCGCATTTGGCTGGACAACAGTATTTCTTCGAGAG				VFEDPTKLHRYLVGKL
			GCTTGCAGCAAGAGGCAAGCGCACCCCGATGGCGGACGGGGTAAAGA				LKARTEARIRRWKEA
			TATTGTTGGATCTGGGAGTGGAACCCCAGGTGGACTTGAAGCGAAGGA				SLAGRLVNDTSIDMR
			CGGTGACCGTCGACGGTATAGTCTTCGAAGACCCGACCAAGCTTCATCG				LSCLWMKKGFVSAR
			ATACCTGGTGGGAAAGCTCTTAAAAGCAAGAACTGAGGCGAGGATTCG				NLRDALAVQEGSLLT
			AAGATGGAAGGAAGCCAGCTTAGCTGGACGGTTGGTAAATGACACAAG				RACPALKGKGGQEV
			TATTGATATGCGACTATCATGCTTGTGGATGAAAAAAGGTTTTGTGAGT				CRCCHAAPETAEHIT
			GCGAGGAACCTCAGAGATGCGCTTGCTGTGCAAGAGGGGAGTTTGCTT				SACRYWLPSLYVERH
			ACTAGAGCATGCCCTGCTCTAAAGGGTAAAGGCGGCCAAGAAGTTTGC				DSVARNLYYVICCRY
			CGGTGTTGCCATGCAGCGCCGGAAACTGCAGAGCACATAACATCAGCCT				GITPVHYSNRVSPLSE
			GCCGCTATTGGCTTCCAAGTCTCTACGTTGAGAGACATGACTCGGTAGC				NSQCRVLWNMDM
			AAGGAACCTCTATTACGTCATATGCTGCCGCTACGGCATAACACCGGTG				QTRTPMKHRKPDIVV
			CATTACTCAAATAGGGTATCACCGCTCTCGGAGAATAGCCAATGCCGCG				FDLKREKILMFEVSIA
			TTCTTTGGAACATGGATATGCAGACTCGGACGCCAATGAAGCATCGAAA				HASGLLKQREIKINRY
			GCCTGATATAGTCGTCTTTGATCTCAAGAGGGAAAAGATCCTCATGTTC				TVNSEELPDETITPYP
			GAAGTTTCGATAGCCCATGCCAGCGGGTTATTGAAACAGCGGGAAATC				PGPNLAADLAATYG
			AAGATCAATCGGTATACGGTGAACTCCGAAGAGTTGCCTGATGAGACCA				WQVEFAPVVVGTCG
			TAACACCGTATCCGCCTGGGCCGAATTTGGCCGCTGACCTCGCTGCCAC				EHVPAVKEDLQRTLD
			CTATGGTTGGCAGGTTGAATTTGCCCCAGTGGTGGTTGGCACGTGTGGT				LKPHQVEALLERISRS
			GAGCACGTACCAGCCGTCAAAGAGGACCTGCAAAGAACGTTGGATCTA				AVIGTARVVRAHLAC
			AAACCTCATCAAGTCGAAGCCCTTCTTGAAAGGATATCCCGATCGGCGG				S (SEQ ID NO: 1505)
			TGATCGGAACGGCTAGAGTAGTCCGAGCACACCTCGCCTGCTCCTAGTC				
			GCTAAGGGGTCCGGAAATGGTCCGGTCCTGCGCTACCCGGTTCTGGTA				
			GCACGTTCAAGCGCTCAATCGCCTGCCTTGTAGGCAGTCCATCTGTGGA				
			AGTCGCGCTCTTGATACAGATGTGGACGGATGGAAGCAGATGATAGAG				
			CCGGTGACGGCCCTACTAGCCAAACGC (SEQ ID NO: 1137)				
R4	R4_Hmel	.	Heliconius melpomene	ATAAATAATAATAATAATAATAATAAGCCCCCTAAAATCCAACCATA	ATAAATAA	TAACTTA	ITYTANMALVTLFME
				CGTCCGAGTCGAACATCTGATTCTCGTGGGGGGCGGACACGTGAAAAT	TAATAATA	TTGTCA	NMENKRYNLRPLPG
				AAATTACTTACACGGCTAACATGGCTCTGGTAACCCTTTTTATGGAGAAT	ATAATAAT	GAATTC	GRRGASGANAGCHS
				ATGGAAAATAAAAGATATAATTTGCGGCCGCTACCCGGGGGTCGCCGG	AATAAGCC	CTTACTA	MRTVGDGGLSRRVP
				GGCGCATCTGGAGCTAACGCTGGATGCCACAGCATGCGAACCGTTGGC	CCCTAAAA	GTAATA	LEKNVTAEQSSSPLTS

	GACGGGGGACTGAGCAGGCGGGTTCCCCTCGAAAAAAATGTTACAGCC	TCCAACCA	ATAATTA	SSSHSPVSSIPSPSSTR
	GAACAATCATCCTCACCACTCACTTCTTCTTCCTCTCATTCACCGGTATCA	TACGTCCG	TCGCTG	TLLNSPNSSPTSSHSS
	AGCATACCCTCACCATCCTCAACTCGCACACCTATTAAATTCCCCTAATTCA	AGTCGAAC	AAAATC	LVIRSADVVQEALAN
	TCACCAACATCATCACATTCATCACTGGTTATAAGGTCAGCAGATGTTGT	ATCTGATT	TCCACCC	YPAPTAGSIRARKKW
	GCAGGAGGCACTTGCCAATTACCCTGCACCCACTGCGGGCTCTATTCGT	CTCGTGGG	AAATATT	TDIMNRYIWRTYLIIT
	GCGCGAAAAAAATGGACCGACATAATGAACCGTTATATTTGGCGCACAT	GGGCGGA	GCTTGG	KCETTLLNNYLEPLHQ
	ATCTTATTATAACAAAAATGCGAAACCACGTTATTGAATAATTATTTAGAA	CACGTGAA	CTATATG	EFSSKFPEMQVTRQR
	CCCCTACATCAAGAATTTAGTAGTAAATTTCCCGAAATGCAGGTAACTAG	AATAA	CTCGCA	IGDQRRAIIRNKLLSD
	ACAACGTATAGGAGATCAACGTAGAGCAATAATAAGAAATAAACTACTA	(SEQ ID	ATTTTTG	DTLAQILIEVKELLQIG
	TCAGATGATACCTTAGCTCAAATTTTAATTGAAGTAAAAGAACTATTACA	NO: 1261)	GTTAAC	DQPLTQNNIHSTQLS
	AATCGGTGATCAACCACTTACACAAAATAACATTCATTCAACTCAACTTT		GTACCC	HSNTRIKWSNELNEE
	CACACTCAAACACTAGAATAAAATGGTCAAACGAACTTAATGAAGAAAT		CAATGA	IVKCYFEVTLLEVNKT
	TGTCAAATGTTACTTTGAAGTCACCCTACTTGAAGTTAATAAGACTTCTT		TTTGGG	SYRKNLYSLFISRNPH
	ATCGTAAAAACCTGTACAGTTTATTTATATCTAGAAATCCACACCTATCA		AGAACA	LSHLTEQRIADQRRLI
	CACCTTACAGAACAAAGAATAGCTGACCAAAGACGACTAATATTTATGA		AAAATG	FMNKSVHNDRIIELK
	ACAAAAGTGTACACAACGATAGAATTATTGAATTAAAAAGAGAAGTAG		GTAAAA	REVEIKLANSNSLTKN
	AAATTAAATTAGCTAACAGTAATTCTTTAACAAAAAATATTACAGAGTCA		CTATAAT	ITESNSPSSQTNEIND
	AATTCTCCTAGCAGTCAAACAAATGAAATAAATGATTCTGCATATGTACA		AATAAT	SAYVQSNLQPVEPLD
	AAGTAATTTACAACCCGTTGAACCACTAGACCAACATTGTATAAATAGAC		AATTATA	QHCINRHNLIEKHYV
	ACAATTTAATAGAAAAACATTATGTTGAACAAGAATTTAATAATGCTTTA		TTAATAT	EQEFNNALIQFNNTN
	ATTCAATTTAACAATACAAATCCTGAAACCAGACCCTATATACCAAGACA		(SEQ ID	PETRPYIPRQKSSRKF
	AAAGTCATCGCGTAAATTTTCCCAAATTGTCAGTTTCTTAAATAGTGAAG		NO:	SQIVSFLNSEVLPKHL
	TACTTCCAAAACATTTAAATAACGAATTAGATTTCAATGCTCTTCATAACA		1384)	NNELDFNALHNIIYTA
	TTATATATACMGCTGYCTACACTGCATCTCTGTGTAATGGAACACAAAATTC			XYTASLCNGTKFSFID
	AGTTTTATTGATAATTACCGACCTCGTAACAGCAAGCCATCGTGGCAAA			NYRPRNSKPSWQRR
	GAAGACTTGAAAGTAGAATAGACAAATATAGATTGCAAATAGGTAGAC			LESRIDKYRLQIGRLT
	TTACACAATATATATCAGGAAATAGGAACCGTAAAATATTAAAAACTGT			QYISGNRNRKILKTVE
	AGAAGAAATAAAAACACAATATAAAATTCATTCACATCACGAAGAACCT			EIKTQYKIHSHHEEPN
	AATACAGAATTACCACATTTCTTAGACACACTAAAACAAAAATTAAATGC			TELPHFLDTLKQKLNA
	AACTTCGAACAGACTTAGAAGATATTTGACATGTACAAAACGGAAGCAA			TSNRLRRYLTCTKRK
	CAAAATAATACATTCGTTAATAACGAAAAACATTTTTATAGAACTCTATC			QQNNTFVNNEKHFY
	TAGCACAAACCAAAATACCACACAACACAATTACWAGAACATCCAACGGA			RTLSSTNQNTTTQLX
	AAATAATTTGCAACAATACTGGGCCAATATTTGGGAAACTTCAATAGAA			EHPTENNLQQYWAN
	CAYAATGCAGACGCAGAATGGTTAAATAAAATACCTGACKTAGAAATAA			IWETSIEHNADAEWL
	ATKTAATGAAATTCAAAGATATATCCATAGAAACATTCAATCAAATCATC			NKIPDXEINXMKFKDI
	CAGCGAACACACAATTGGAAGGCACCGGGAACCGATAATATACACAAC			SIETFNQIIQRTHNW
	TAYTGGTACAAAAAGCTAACATGCACTCATTCACTTCTTCTTAAACACAT			KAPGTDNIHNYWYK
	AAACCAGTTCATACAATCACCTTGCACACTACCTTTATTTATAACAAACG			KLTCTHSLLLKHINQFI

			GAATAACGTATATGTTACCGAAAGGCTTAGATCCCACAAACCCAGCTAA			QSPCTLPLFITNGITY
			CTATAGGCCAATAACATGCCTCCAAACGATATATAAAATTATCACTGCGT			MLPKGLDPTNPANY
			GTATTACTGACATAATTTATAAACATATCGATCAAAATAAACATATTAGCT			RPITCLQTIYKIITACIT
			GAACAACAGAAAGGCTGTAGGAAAAATAGTCAAGGGTGTAAAGAGCA			DIIYKHIDQNNILAEQ
			GCTTACTATAGACGCAATAGTTATGAAACAGGCACATAATAAAAATAHT			QKGCRKNSQGCKEQ
			AATACRATGTATATAGACTACAGAAAAGCATTTGATTCAGTTCCACACA			LTIDAIVMKQAHNKN
			GTTGGTTACTATATATTTTAAAAAAGTACAAGATTCACCCCATTCTAATA			XNTMYIDYRKAFDSV
			ACATTTTTAAGCTCCGTYATGCTCTCATGGAAAACTAGATTAAAATTGAT			PHSWLLYILKKYKIHPI
			TAACAATAATGAAACACTGATTACTGATTGGATTAAAATACAACGAGGC			LITFLSSVMLSWKTRL
			ATCTTTCAAGGTGATGCGCTGAGTCCATTATGGTTCTGTCTAGCGTTAAA			KLINNNETLITDWIKI
			CCCTCTCTCTGAATTGCTTAATAATACCAATACAGGATTTAAACTAAAAC			QRGIFQGDALSPLWF
			ATAATAACACATATCATATAATTTCACATTTAATGTACATGGACGATATC			CLALNPLSELLNNTNT
			AAACTATACGCCAGTAATAATAAAGAACTAAAAATACTAGCCGACCTCA			GFKLKHNNTYHIISHL
			CACAATCCTTCTCTACAGACATACGCATGGAATTTGGAATTGAAAAATGC			MYMDDIKLYASNNK
			AAGGTACATTCAATTAAACGTGGCAAATCACACAACAAAATACATATATAC			ELKILADLTQSFSTDIR
			TTAACACTGGTGAACAAATTGAATCAATGGACGAAAATAGCACGTATAA			MEFGIEKCKVHSIKR
			ATACCTTGGATTTCAACAAGCTAAGCAAATACAACAAAAACAAACCAAG			GKSQQNTYILNTGEQ
			ATAGAACTAACCAATAAATTTAAATTTCGTTTGAATCAAATTTTAAGATC			IESMDENSTYKYLGF
			ACAATTAAACTCACGTAATATTATAAAAGCCATCAACACATACGCAATAC			QQAKQIQQKQTKIEL
			CGATCCTTACTTACTCTTTTGCTATAATAAACTGGTCTCAAACTGATTTAT			TNKFKFRLNQILRSQL
			CGAATCTGCAACGTATAATAAATACACACATGACTACACATAGGAAACA			NSRNIIKAINTYAIPILT
			TCACCCTAAATCATGTATCCAGCGTCTAACTATATCACGTCTGGATGGAG			YSFAIINWSQTDLSNL
			GACGAGGTCTCATTGATATTCGTAACCTACATAATAATCTAGTGACTAAA			QRIINTHMTTHRKHH
			TTCCGAAATTACTTTTACGCTAAGGCTGAAATTAGTGAACTACACAAATT			PKSCIQRLTISRLDGG
			CATAGTTAATATAGATAATAAATATACACCACTTAATTTAAACGATAGAA			RGLIDIRNLHNNLVTK
			ATATACAACTAAATCAAACTTTGATAACTAAGCAGCAAAAAATAGAAGC			FRNYFYAKAEISELHK
			ATGGTCTTTAAAGTCACTACACGGTAGACATCTAGCCGATCTTAGCCAA			FIVNIDNKYTPLNLND
			ACTCACGTCGACAAAGTTGCGTCGAACGAATGGCTACGACGGGGGAGAT			RNIQLNQTLITKQQKI
			CTCTTCCCAGAGACGGAAGCCTTCATGATGGCTATTCAGGACCAAGTAA			EAWSLKSLHGRHLAD
			TAGACACTAGGAACTATCAAAAACACATCATAAAAAAGACCAAATATGGT			LSQTHVDKVASNEW
			AAACGATTTATGCAGAAGATGCTATTCTAGTCCTGAAACAATCCAACAC			LRRGDLFPETEAFM
			ATTACCGGAGCCTGTAAAACCATAGTACAAACCGATTACAAACATAGAC			MAIQDQVIDTRNYQ
			ATGATCAGGTCGCTGCAATTATTCATCAACACCTAGCATTTAAACATTCT			KHIIKRPNMVNDLCR
			CTCATCACACAAGCACAAAAAACTCCCTATTACAAATATTCACCTCAGGC			RCYSSPETIQHITGAC
			TATTTTAGAATCTACAAATTTTAAGCTGTATTGGGATCGTACAATCATCA			KTIVQTDYKHRHDQV
			CAGATAAAACCGTACACTACAATAGACCGGATATTTTACTTCATGATAAA			AAIIHQHLAFKHSLIT
			GTWAAAAWTTCTGTATACTTGATAGACATAGCCATTCCAAATACACACA			QAQKTPYYKYSPQAI
			ATCTTGCATCCACTTTTTCAAACAAAATAGATAAGTATACAGATTTAACC			LESTNFKLYWDRTIIT

			ATAGAGCTCAAAAGTCAATGGAAAGTCCAATCCGTCACTACAGTACCAA TAGTTCTTTCAACAACCGGAGTAGTACCTCATACGCTACACACGAGTCTG GAAACTTTGGGTATTCACCGGTTAAGTTATATTCTTTTACAGAAAGCCGC AATCTTAAATACATGCCGTATAGTTAGGAAGTTTTTATCATCAAATAATT AACTTATTGTCAGAATTCCTTACTAGTAATAATAATTATCGCTGAAAATC TCCACCCAAATATTGCTTGGCTATATGCTCGCAATTTTTGGTTAACGTAC CCCAATGATTTGGGAGAACAAAAATGGTAAAACTATAATAATAATAATT ATATTAATAT (SEQ ID NO: 1138)			DKTVHYNRPDILLHD KVKXSVYLIDIAIPNTH NLASTFSNKIDKYTDL TIELKSQWKVQSVTT VPIVLSTTGVVPHTLH TSLETLGIHRLSYILLQ KAAILNTCRIVRKFLSS NN (SEQ ID NO: 1506)
R4	Rex6	.	Takifugu rubripes	TTCTATGCGCCTTATGCGACTGGATAGGCCAGTGGTTTACGCCGCTGAC TTTGGTGCGGAAGGTTGTCGGTTCGAATCCAGGCGAGCCCTTAGGCAA GGCTCCTTACGCATATATGCCTACACCTCGGTATGAGCGGAACMWTAA CTGATAGAGTCATACCGGCTCGGACGTCGCCCGGGTCAACAAGGTCCG CGTCAGGGGTCGGGGAACCAGGGCCACCTGATGTCAAATTGGCTACTG GAACAAGGCATTCATGGAGCAGGGCAGAGAACGTGGTACTGATGGAG TGTTACTACGGGAGTAACCCCAGCGAGAGAGGGTACATGCAGAGGATG TGGGAGAAATGGGGTGCTTCGAAACCCCACATCCTCACTGACTAAGAAGC AGCTCTTAGCTCAGTGTTCGAATATCCGCAATAAGAAGCTGCTATCACA GCTAGAGATTGACGAGGCACGGCGNTGTGCCTCACCCACGGTACAAAT ATGCTACGGCAAGGGGGAGCCAGGACGCCAGGTCAGCNTGGGGGTGA TTTCATCATCACCCCCAATATCGAGATTGGGTACAAAGCCCCAATGACA GACGGACTCGGCACAAGGGCAGCTGACCTGAGAGAGAGAATCATGAA GTCCTGGGGGAACTCAACTACCTCCCTGCCAAGGCTAACACACAAAGTA CCAGACCAATCTCTACTGGAGGACATGAACACGGCACTGTCAACCATCC CTACCACTACCATCACTGAGACCAATCAGCTGATGTATGCAGCGGCAAC AGTAATCCTACAGATGCTTGGCTATAAGATGAAGAGCATGAACAGCCAG AAGGAGCAAATGGCCCCATGGAGGAGAAGGCTAGAGGCAAAAATCAT GGCGACACGGAGAGAAGTCAGCCTTCTAACAGAGCTGAGTAGAGGCGT GAATCTAAGGACAGAGCNGCCCAAGAAATATAACAAACTGTCCACAACT GAGGCACTGGAGACTGCTAAGCAAAGGCTCACAGCCCTGGCTACCCGA CTAAAGAGGTACACAAGAGAAGTAGAGGCAAGGAGAATAAACAAGGT GTTCTCCACCAATCCAGCNAAGGTCTACTCTCAATGGCAGGGCAACAAG ATGACAACAGACCCCCCCAGGGCTGAGACTGAGCAATACTGGAAGAGT ATCTGGGAGAAAGAGGCAACGCACACAACTANTGCCCAATGGCTGCAA GACCTACAGACAGAGCACAGCCAACTTCCAGAACAAGACCCAGTAGTCA TCACCTTAGCAGACATCCAAACAAGAGTGTCCAAATGAAGAGCTGGAC AGCACCAGGGCCCGATAAGATCCACGCCTACTGGCTTAAGAAGCTGACT GCACTCCATGAACGCCTGGCAGCACAGATGAACCAGCTGCTAACATCAG	TTCTATGC GCCTTATG CGACTGGA TAGGCCAG TGGTTTAC GCCGCTGA CTTTGGTG CGGAAGG TTGTCGGT TCGAATCC AGGCGAG CCCTTAGG CAAGGCTC CTTACGCA TATATGCC TACACCTC GGT (SEQ ID NO: 1262)	TAGAGG ACCCGA GTCTGA AGGAAG GAGGCA CCGCCC AGGAGG GCGAGG AAGAGA TTTTTTT TTATATA TATATAT ATATATA TA (SEQ ID NO: 1385)	MSGTXTDRVIPARTS PGSTRSASGVGEPGP PDVKLATGTRHSWS RAENVVLMECYYGS NPSERGYMQRMWE KWVLRNPTSSLTKKQ LLAQCSNIRNKKLLSQ LEIDEARRCASPTVQI CYGKGEPGRQVSXG VISSSPPNIEIGYKAP MTDGLGTRAADLRE RIMKSWGNSTTSLPR LTHKVPDQSLLEDM NTALSTIPTTTITETN QLMYAAATVILQML GYKMKSMNSQKEQ MAPWRRRLEAKIMA TRREVSLLTELSRGVN LRTEXPKKYNKLSTTE ALETAKQRLTALATRL KRYTREVEARRINKVF STNPAKVYSQWQGN KMTTDPPRAETEQY WKSIWEKEATHNTX AQWLQDLQTEHSQL PEQDPVVITLADIQTR VSKMKSWTAPGPDK IHAYWLKKLTALHERL AAQMNQLLTSGNHP

		GGAACCACCCAGAGTGGCTAACCCAGGGCCGGACAGTCCTCATAATGA				EWLTQGRTVLIMKD
		AGGACCCCCAGAAGGGCACAATACCATCCAACTACCGGCCCATAACCTG				PQKGTIPSNYRPITCL
		CCTCAGCACCACATGGAAGCTCCTATCAGGCATCATAGCGGCTAAGATC				STTWKLLSGIIAAKISR
		AGTAGGCACATGGATCAATACATGAGCAGAGCACAGAAAGGCATAGGC				HMDQYMSRAQKGI
		AACAACACCAGAGGTGCCAAGCACCAGCTACTGGTCGACAGGGCAATC				GNNTRGAKHQLLVD
		GCCCAGGACTGTAGGACGCGGCACACCAACCTGTGCACTGCCTGGATT				RAIAQDCRTRHTNLC
		GATTACAAGAAAGCCTATGACTCAATGCCCGCACACATGGATACTGGAGT				TAWIDYKKAYDSMP
		GCCTAAAGCTGTATAACATCAACAGGACACTAAGAGAGTTCATCCAGAA				HTWILECLKLYNINRT
		CTCCATGAAGCTGTGGAACACGACTCTGGAGGCCAACTCAAAGCCAATT				LREFIQNSMKLWNTT
		GCGCGGGTGAGCATCAGATGCGGCATATATCAAGGAGATGCCCTGTCC				LEANSKPIARVSIRCGI
		CCCCTGCTGTTCTGCATAGGCCTAAACCCCCTCAGCCAGATCATCACAAA				YQGDALSPLLFCIGLN
		GAGTGGCTATGGGTACCAGTTCCGAAGTGGAACAACCGTCAGCCACCTC				PLSQIITKSGYGYQFR
		CTCTACATGGATGACATCAAGCTGTATGCCAAGAACGAGCGTGACATCG				SGTTVSHLLYMDDIK
		ACTCCCTGATTCACCTCACTAGGATCTACAGCAAAGACATCGGGATGTC				LYAKNERDIDSLIHLT
		ATTCGGACTAGATAAATGTGGGCGGATGATATCTAGAAGAGGAAAGGT				RIYSKDIGMSFGLDKC
		GATCGCAACTGATGGGGTTGAACTACCTGAAGGGAACATCACAGATGT				GRMISRRGKVIATDG
		GCAGGACAGTTACAAATACCTGGGGATCCCGCAGGCAAATGGTAACCA				VELPEGNITDVQDSY
		TGAGGAGGCAGCTAGGAGGTCAGCCACAGCCAAATACCTACAGAGGTT				KYLGIPQANGNHEEA
		AAGGCAGGTCCTGAAAAGTCAGCTGAATGGTAAGAATAAGATCCAGGC				ARRSATAKYLQRLRQ
		CATAAACACCTACGCCCTGCCAGTAATCAGATACCCTGCTGGCATAATAC				VLKSQLNGKNKIQAI
		CCTGGCCACTGGAAGAGATACAAGCCACTGACATCAAGACAAGGAAGC				NTYALPVIRYPAGIIP
		TGAATGGTAAGCATAAGATCCAGGCCATAAACACCTACGCCCTGCCAGT				WPLEEIQATDIKTRKL
		AATCAGATACCCTGCTGGCATAATACCCTGGCCACTGGAAGAGATACAA				NGKHKIQAINTYALP
		GCCACTGACATCAAGACAAGGAAGCTCCTCACCATGCACGGAGGGTTTC				VIRYPAGIIPWPLEEI
		ACCCTAAGTCCAGCGTCCTGAGGCTGTACACGAAGCGAAAGGAAGGGG				QATDIKTRKLLTMHG
		GCCGAGGACTAGTAAGTGTCCGAACTACTGTCCAGGAGGAAACAACAA				GFHPKSSVLRLYTKRK
		GCCTCCGAGAATACATCAAGAAGATGGCCCCCACTGACCGACTGCTGAG				EGGRGLVSVRTTVQE
		TGAATGCCTCAGGCAACAAAAGCCCACCAAGGAGGAGGAGCCTGAGG				ETTSLREYIKKMAPTD
		GGCTATCATGGAAGGACACAAGCCCCTGCATGGNATGTACCACCGACAAA				RLLSECLRQQKPTKEE
		TTGAGGAAGTGGCTGATATCGAGAAAACATACCAGTGGCTGGAAAAGG				EPEGLSWKDKPLHG
		CCGGACTGAAAGACAGCACAGAGGCACTACTCATGGCTGCACAAGAAC				MYHRQIEEVADIEKT
		AGGCCCTGAGCACCAGAGCAATAGAGGCCAGGGTCTACCATACCAGAC				YQWLEKAGLKDSTEA
		AAGACCCCAGGTGCAGGCTGTGTGGAGACGCCCCTGAGACAGTCCAGC				LLMAAQEQALSTRAI
		ACATCACAGCAGGGTGCAAGATGTTGGCAGGCAAGGCATACATGGAGC				EARVYHTRQDPRCRL
		GGCATAACCAGGTGGCTGGCATAGTGTACAGGAACATCTGCACTGAGT				CGDAPETVQHITAGC
		ATGGACTGGAGGTCCCAGGGTCCAGGTGGGAGACACCCCCGAAAGTGC				KMLAGKAYMERHN
		TGGAGAACAAGCAGGCCAAGATCCTGTGGGACTTCCAGATCCAGACTG				QVAGIVYRNICTEYGL
		ACAAGATGGTGGTGGCCAACCAGCCTGACATAGTGGTGGTGGATAAAC				EVPGSRWETPPKVLE

			ACCAGAAGACAGTGGTGGTGATAGATGTAGCAATCCCAAGTGATAGCA ACATCAGGAAGAAGGAACACGAGAAGCTGGAGAAGTACCAAGGGCTG AAGGAGGAGATGGAGAGAATGTGGGGCATGAAGGCAACAGTGGTCCC AGTAGTGATTGGGACACTAGGGGCAGTAACACCCAAGCTGAGTAGATG GCTCCAACAGATACCAGGAACCACATCAGAGATCTCTGTCCAGAAGAGC GCAGTCCTAGGAACAGCTAAGATCCTGCGCAGAACCCTCAGACTCCCAG GCCTCTGGTAGAGGACCCGAGTCTGAAGGAAGGAGGCACCGCCCAGG AGGGCGAGGAAGAGATTTTTTTTTTATATATATATATATATATATA (SEQ ID NO: 1139)			NKQAKILWDFQIQTD KMVVANQPDIVVVD KHQKTVVVIDVAIPS DSNIRKKEHEKLEKYQ GLKEEMERMWGMK ATVVPVVIGTLGAVT PKLSRWLQQIPGTTS EISVQKSAVLGTAKIL RRTLRLPGLW (SEQ ID NO: 1507)

Exemplary Cis Gene-Writing embodiments

In some embodiments, a writing domain (e.g., RT domain) comprises an RNA-binding domain, e.g., that specifically binds to an RNA sequence. In some embodiments, a template RNA comprises an RNA sequence that is specifically bound by the RNA-binding domain of the writing domain.

Template nucleic acid binding domain:

The Gene Writer polypeptide typically contains regions capable of associating with the Gene Writer template nucleic acid (e.g., template RNA). In some embodiments, the template nucleic acid binding domain is an RNA binding domain. In some embodiments, the RNA binding domain is a modular domain that can associate with RNA molecules containing specific signatures, e.g., structural motifs, e.g., secondary structures present in the 3' UTR in non-LTR retrotransposons. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the reverse transcription domain, e.g., the reverse transcriptase-derived component has a known signature for RNA preference, e.g., secondary structures present in the 3' UTR in non-LTR retrotransposons. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the DNA binding domain. For example, in some embodiments, the DNA binding domain is a CRISPR-associated protein that recognizes the structure of a template nucleic acid (e.g., template RNA) comprising a gRNA. In some embodiments, the gRNA is a short synthetic RNA composed of a scaffold sequence that participates in CRISPR-associated protein binding and a user-defined ~20 nucleotide targeting sequence for a genomic target. The structure of a complete gRNA was described by Nishimasu et al. Cell 156, P935-949 (2014). The gRNA (also referred to as sgRNA for single-guide RNA) consists of crRNA- and tracrRNA-derived sequences connected by an artificial tetraloop. The crRNA sequence can be divided into guide (20 nt) and repeat (12 nt) regions, whereas the tracrRNA sequence can be divided into anti-repeat (14 nt) and three tracrRNA stem loops (Nishimasu et al. Cell 156, P935-949 (2014)). In practice, guide RNA sequences are generally designed to have a length of between 17 – 24 nucleotides (e.g., 19, 20, or 21 nucleotides) and be complementary to a targeted nucleic acid sequence. Custom gRNA generators and algorithms are available commercially for use in the design of effective guide RNAs. In some embodiments, the gRNA comprises two RNA components from the native

CRISPR system, e.g. crRNA and tracrRNA. As is well known in the art, the gRNA may also comprise a chimeric, single guide RNA (sgRNA) containing sequence from both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing/binding). Chemically modified sgRNAs have also been demonstrated to be effective for use with CRISPR-associated proteins; see, for example, Hendel et al. (2015) Nature Biotechnol., 985 – 991. In some embodiments, a gRNA comprises a nucleic acid sequence that is complementary to a DNA sequence associated with a target gene. In some embodiments, a polypeptide comprises a DNA-binding domain comprising a CRISPR-associated protein that associates with a gRNA that allows the DNA-binding domain to bind a target genomic DNA sequence. In some embodiments, the gRNA is comprised within the template nucleic acid (e.g., template RNA), thus the DNA-binding domain is also the template nucleic acid binding domain. In some embodiments, the polypeptide possesses RNA binding function in multiple domains, e.g., can bind a gRNA structure in a CRISPR-associated DNA binding domain and a 3' UTR structure in a non-LTR retrotransposon derived reverse transcription domain.

Endonuclease domain:

In some embodiments, a Gene Writer polypeptide possesses the function of DNA target site cleavage via an endonuclease domain. In some embodiments, the endonuclease domain is also a DNA-binding domain. In some embodiments, the endonuclease domain is also a template nucleic acid (e.g., template RNA) binding domain. For example, in some embodiments a polypeptide comprises a CRISPR-associated endonuclease domain that binds a template RNA comprising a gRNA, binds a target DNA sequence (e.g., with complementarity to a portion of the gRNA), and cuts the target DNA sequence. In certain embodiments, the endonuclease/DNA binding domain of an APE-type retrotransposon or the endonuclease domain of an RLE-type retrotransposon can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a Gene Writer system described herein. In some embodiments the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the endonuclease element is a heterologous endonuclease element, such as Fok1 nuclease, a type-II restriction l-like endonuclease (RLE-type nuclease), or another RLE-type endonuclease (also known as REL). In some embodiments the heterologous endonuclease activity has nickase

activity and does not form double stranded breaks. The amino acid sequence of an endonuclease domain of a Gene Writer system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain of a retrotransposon whose DNA sequence is referenced in Table 1, 2, or 3. A person having ordinary skill in the art is capable of identifying endonuclease domains based upon homology to other known endonuclease domains using tools such as Basic Local Alignment Search Tool (BLAST). In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., Nucleic Acids Research 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., Mobile DNA 8:16, 2017). In certain embodiments, the heterologous endonuclease is derived from a CRISPR-associated protein, e.g., Cas9. In certain embodiments, the heterologous endonuclease is engineered to have only ssDNA cleavage activity, e.g., only nickase activity, e.g., be a Cas9 nickase. For example, a Gene Writer polypeptide described herein may comprise a reverse transcriptase domain from an APE- or RLE-type retrotransposon and an endonuclease domain that comprises Fok1 or a functional fragment thereof. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease activity. In still other embodiments, endonuclease domains are modified to remove any latent DNA-sequence specificity.

In some embodiments the endonuclease domain has nickase activity and does not form double stranded breaks. In some embodiments, the endonuclease domain forms single stranded breaks at a higher frequency than double stranded breaks, e.g., at least 90%, 95%, 96%, 97%, 98%, or 99% of the breaks are single stranded breaks, or less than 10%, 5%, 4%, 3%, 2%, or 1% of the breaks are double stranded breaks. In some embodiments, the endonuclease forms substantially no double stranded breaks. In some embodiments, the enonuclease does not form detectable levels of double stranded breaks.

In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand; e.g., in some embodiments, the endonuclease domain cuts the genomic DNA of the target site near to the site of alteration on the strand that will be extended by the writing domain. In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand and does not nick the target site DNA of the non-edited strand. For example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity and that does not form double stranded breaks, in some embodiments said CRISPR-associated endonuclease domain nicks the target site DNA strand containing the PAM site (e.g., and does not nick the target site DNA strand that does not contain the PAM site).

In some other embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand and the non-edited strand. Without wishing to be bound by theory, after a writing domain (e.g., RT domain) of a polypeptide described herein polymerizes (e.g., reverse transcribes) from the heterologous object sequence of a template nucleic acid (e.g., template RNA), the cellular DNA repair machinery must repair the nick on the to-be-edited DNA strand. The target site DNA now contains two different sequences for the to-be-edited DNA strand: one corresponding to the original genomic DNA and a second corresponding to that polymerized from the heterologous object sequence. It is thought that the two different sequences equilibrate with one another, first one hybridizing the non-edited strand, then the other, and which the cellular DNA repair apparatus incorporates into its repaired target site is thought to be random. Without wishing to be bound by theory, it is thought that introducing an additional nick to the non-edited strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-based sequence more frequently than the original genomic sequence. In some embodiments, the additional nick is positioned at least 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, or 150 nucleotides 5' or 3' of the target site modification (e.g., the insertion, deletion, or substitution) or to the nick on the to-be-edited strand.

Alternatively or additionally, without wishing to be bound by theory, it is thought that an additional nick to the non-edited strand may promote second strand synthesis. In some embodiments, where the Gene Writer has inserted or substituted a portion of the edited strand,

synthesis of a new sequence corresponding to the insertion/substitution in the non-edited strand is necessary.

In some embodiments, the polypeptide comprises a single domain having endonuclease activity (e.g., a single endonuclease domain) and said domain nicks both the to-be-edited strand and the non-edited strand. For example, in such an embodiment the endonuclease domain may be a CRISPR-associated endonuclease domain, and the template nucleic acid (e.g., template RNA) comprises a gRNA that directs nicking of the to-be-edited strand and an additional gRNA that directs nicking of the non-edited strand. In some embodiments, the polypeptide comprises a plurality of domains having endonuclease activity, and a first endonuclease domain nicks the to-be-edited strand and a second endonuclease domain nicks the non-edited strand (optionally, the first endonuclease domain does not (e.g., cannot) nick the non-edited strand and the second endonuclease domain does not (e.g., cannot) nick the to-be-edited strand).

In some embodiments, the endonuclease domain is capable of nicking a first strand and a second strand. In some embodiments, the first and second strand nicks occur at the same position in the target site but on opposite strands. In some embodiments, the second strand nick occurs in a staggered location, e.g., upstream or downstream, from the first nick. In some embodiments, the endonuclease domain generates a target site deletion if the second strand nick is upstream of the first strand nick. In some embodiments, the endonuclease domain generates a target site duplication if the second strand nick is downstream of the first strand nick. In some embodiments, the endonuclease domain generates no duplication and/or deletion if the first and second strand nicks occur in the same position of the target site (e.g., as described in Gladyshev and Arkhipova Gene 2009, incorporated by reference herein in its entirety). In some embodiments, the endonuclease domain has altered activity depending on protein conformation or RNA-binding status, e.g., which promotes the nicking of the first or second strand (e.g., as described in Christensen et al. PNAS 2006; incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain comprises a meganuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a homing endonuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a meganuclease from the LAGLIDADG (SEQ ID NO: 1594), GIY-YIG, HNH, His-Cys Box, or PD-(D/E) XK families, or a functional fragment or variant thereof, e.g., which possess conserved amino acid motifs, e.g., as indicated in the family names. In some

embodiments, the endonuclease domain comprises a meganuclease, or fragment thereof, chosen from, e.g., 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), or I-BmoI (Uniprot Q9ANR6). In some embodiments, the meganuclease is naturally monomeric, e.g., I-SceI, I-TevI, or dimeric, e.g., I-CreI, in its functional form. For example, the LAGLIDADG (SEQ ID NO: 1594) meganucleases with a single copy of the LAGLIDADG (SEQ ID NO: 1594)motif generally form homodimers, whereas members with two copies of the LAGLIDADG (SEQ ID NO: 1594)motif are generally found as monomers. In some embodiments, a meganuclease that normally forms as a dimer is expressed as a fusion, e.g., the two subunits are expressed as a single ORF and, optionally, connected by a linker, e.g., an I-CreI dimer fusion (Rodriguez-Fornes et al. Gene Therapy 2020; incorporated by reference herein in its entirety). In some embodiments, a meganuclease, or a functional fragment thereof, is altered to favor nickase activity for one strand of a double-stranded DNA molecule, e.g., I-SceI (K122I and/or K223I) (Niu et al. J Mol Biol 2008), I-AniI (K227M) (McConnell Smith et al. PNAS 2009), I-DmoI (Q42A and/or K120M) (Molina et al. J Biol Chem 2015). In some embodiments, a meganuclease or functional fragment thereof possessing this preference for single-strand cleavage is used as an endonuclease domain, e.g., with nickase activity. In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, which naturally targets or is engineered to target a safe harbor site, e.g., an I-CreI targeting SH6 site (Rodriguez-Fornes et al., *supra*). In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, with a sequence tolerant catalytic domain, e.g., I-TevI recognizing the minimal motif CNNNG (Kleinstiver et al. PNAS 2012). In some embodiments, a target sequence tolerant catalytic domain is fused to a DNA binding domain, e.g., to direct activity, e.g., by fusing I-TevI to: (i) zinc fingers to create Tev-ZFEs (Kleinstiver et al. PNAS 2012), (ii) other meganucleases to create MegaTevs (Wolfs et al. Nucleic Acids Res 2014), and/or (iii) Cas9 to create TevCas9 (Wolfs et al. PNAS 2016).

In some embodiments, the endonuclease domain comprises a restriction enzyme, e.g., a Type IIS or Type IIP restriction enzyme. In some embodiments, the endonuclease domain comprises a Type IIS restriction enzyme, e.g., FokI, or a fragment or variant thereof. In some embodiments, the endonuclease domain comprises a Type IIP restriction enzyme, e.g., PvuII, or a fragment or variant thereof. In some embodiments, a dimeric restriction enzyme is expressed as

a fusion such that it functions as a single chain, e.g., a FokI dimer fusion (Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

The use of additional endonuclease domains is described, for example, in Guha and Edgell Int J Mol Sci 18(22):2565 (2017), which is incorporated herein by reference in its entirety.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a Cas protein, e.g., a *Streptococcus pyogenes* Cas9 (SpCas9) or a functional fragment or variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a modified SpCas9. In embodiments, the modified SpCas9 comprises a modification that alters protospacer-adjacent motif (PAM) specificity. In embodiments, the PAM has specificity for the nucleic acid sequence 5'-NGT-3'. In embodiments, the modified SpCas9 comprises one or more amino acid substitutions, e.g., at one or more of positions L1111, D1135, G1218, E1219, A1322, of R1335, e.g., selected from L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. In embodiments, the modified SpCas9 comprises the amino acid substitution T1337R and one or more additional amino acid substitutions, e.g., selected from L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto. In embodiments, the modified SpCas9 comprises: (i) one or more amino acid substitutions selected from D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337; and (ii) one or more amino acid substitutions selected from L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto.

In some embodiments, a Gene Writer may comprise a Cas protein as listed in Table 40. The predicted or validated nickase mutations for installing Nickase activity in the Cas protein as shown in Table 40, are based on the signature of the SpCas9(N863A) mutation. In some embodiments, system described herein comprises a GeneWriter protein of Table 3 and a Cas protein of Table 40 A. In some embodiments, a GeneWriter protein of Table 3 is fused to a Cas protein of Table 40 A.

Table 40A: CRISPR/Cas Proteins, Species, and Mutations

Variant	Parental Host	Protein Sequence	Nickase Mutation	SEQ ID No.
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLAR SVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQR GDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFE PAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGL EDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKT DEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKN TEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEE NRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHAL PFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQ KFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAEN DRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQE VMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRS AKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDN PFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYA WQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHD KGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR	N611A	1949
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRR LIKRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGLKEKLERQEWAAVLLHLSKHRGYLS QRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDL LAEMELLFQRQAELGNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSY SAERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSKLTYAQVRAMLALSDNAIFKGVRY LGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNELKGNSDLLDEIGTAFSLYKTDDDICRYL EGKLPERVLNALLENLNFDKFIQLSLKALHQILPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLP TIPADEIRNPVVLRTLTQARKVINAVVRLYGSPARIHIETAREVGKSYQDRKKLEKQQEDNRKQRE SAVKKFKEMFPHFVGEPKGKDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDHALPFSRT WDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSERWQHFVVRVQTSGFSYAKKQRILNHKLD EKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNGQITALLRHRWGLQKVREQNDRH HALDAVVVACSTVAMQQKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPWAFFKENVEIRIFSEN PKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLSVLKVPLTQLKL SDLERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKGGALVKAVRLEQTQKSGVLVRDG NGVADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNRAATQGKDENDWDIMDEMATFQFSL CQNDLIKLVTKKKTIFGYFNGLNRATSNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELG KNIRPCRPTKRQHVR	N605A	1950
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETG NYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFV TVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	1951
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNR IEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	1952

SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRI HRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFI QSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTP YQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDI YKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYYNDLIMYEDELFR VIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKP QLIFKRGEL	N588A	1953
SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRI HRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFI QSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTP YQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDI YKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYKNDLIMYEDELFR VIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKK PQLIFKRGEL	N588A	1954
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGA EELLAKLNREDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQG DSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRI EEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDD SIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKA GFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINN YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMN FFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKES ILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQ NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVVNSNLKSSFDEQFAVSDSILLSNSFVSLL KYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	1955
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD	N863A	1956

		AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD		
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1957
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAERTRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT RLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1958
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADL QFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQK HYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPK LDF	N622A	1959
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRREARSSRRRLRRKKHR IERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIHLAQRRGFQS NRKTDGSSEDGQVLVNVTENDRLMEEKDYRTVAEMMVKDEKFSDHKRNKNGNYHGVVSRSSL LVEIHTLFETQRQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKASWHF	N607A	1960

		QYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVSYHDTRKILDLSEEYQFVGLDYGKE DEKKKVESKETIIKLDDYHKLNKIFNEVELAKGETWEADDYDTVAYALTFFKDDEDIRDYLQNKY KDSKNRLVKNLANKEYTNELIGKVSTLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAGFDFQ GISKKKRSVVLPVIDQISNPVVNRALTQTRKVINALIKKYGSPETIHIETARELSKTFDERKNITKDY KENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISFERLKESGYTEVDHIIPY SRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDTQRWYEFEQRVTTNPQIKKEKRQNLLLKG FTNRRELEMLERNLNDTRYITKYLSHFISTNLEFSPSDKKKKVVNTSGRITSHLRSRWGLEKNRGQ NDLHHAMDAIVIAVTSDSFIQQVTNYYKRKERRELNGDDKFPLPWKFFREEVIARLSPNPKEQIEA LPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITAKKTALVDISYDKNGD FNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLHKPKKDGTKGPLIKSVRIMENKTLVHPVNK GKGVVYNSSIVRTDVFQRKEKYYLLPVYVTDVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLY PNDLIFIRQNPKKKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQNL DCFEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR		
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDKNNL IEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	1961
cCas9-v17	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNSTRNI VELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	1962
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDDRNII ELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	1963
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY	N580A	1964

		ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNNRLNK IELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG		
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRL YRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHIAR HRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGV LSARLQQSDYAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASD AFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTAT MTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAK VQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPT PRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAAR NAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNT RENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVE RFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKF FDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRAA WRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKA SSEALWCALTREPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFH HARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYL GWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQMSKEGIKKESAPEL SKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAHLPVTWKVQ	H573A (Alternate)	1965
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLN HLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYD DIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQ SFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFM FVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFI EFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS FKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRK VLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILK LRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLD FLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSI VKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALH EETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMD FALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLI VSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	N582A	1966
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHR LERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRK SERSNKENSTMLKHIEENRAILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFS KQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQNENIRF LELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLA NKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPN IPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIR QLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNK VLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEETEFKNRN LNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIV ACTTPSDIAKVTAFYQRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDD QKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGK ESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVAYN SNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIE LPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLG NIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	N605A	1967
iSpyMacCas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL	N863A	1968

		VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPL EVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQ VGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNH NQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGV KLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSEFES		
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLAR SVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQR SDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEP AEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLE DTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTE EKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENR KDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFS RTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKF DEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAEND RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVK SAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDK AGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGIL PDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDH KIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	N611A	1969
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAE ELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPL ARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTV YNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGD SLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEE GIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSI DNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAG FIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNF FKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESI LSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYE KDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQN ISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLK YTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	1970
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGA EELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNANFMQLIHDDSLTFKEEIEKAQVSGQ GDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKR IEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKD DSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADK AGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDIN NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIM NFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKE SILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQ NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLL KYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD		
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ	N863A	1972

		LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD		
SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGSS EDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGASAAFKYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1973
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	N863A	1974
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL		

		AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD		
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAQLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTFKQKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1976
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1977
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA	N863A	1978

		KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD		
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1979
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1980
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	1981
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR	N622A	1982

		AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYA DLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQ KHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKP KLDF		
St1Cas9-LMG1831	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYA DLQFEKKTGTYKISQEKYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPN VKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPK LDF	N622A	1983
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	1984
St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	1985

Table 40B provides parameters to define the necessary components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 3A for Gene Writing. Tier indicates

preferred Cas variants if they are available for use at a given locus. The cut site indicates the validated or predicted protospacer adjacent motif (PAM) requirements, validated or predicted location of cut site (relative to the most upstream base of the PAM site). The gRNA for a given enzyme can be assembled by concatenating the crRNA, Tetraloop, and tracrRNA sequences, and further adding a 5' spacer of a length within Spacer (min) and Spacer (max) that matches a protospacer at a target site. Further, the predicted location of the ssDNA nick at the target is important for designing the 3' region of a Template RNA that needs to anneal to the sequence immediately 5' of the nick in order to initiate target primed reverse transcription.

Table 40B parameters to define the necessary components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 40A for Gene Writing

Variant	PAM(s)	Cut	Tier	Spacer (min)	Spacer (max)	crRNA	Tetraloop	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGCTCCCTTTCTCATTTCG (SEQ ID NO: 386)	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA (SEQ ID NO: 387)
PpnCas9	NNNNRTT		1	21	24	GTTGTAGCTCCCTTTTTCATTTCGC (SEQ ID NO: 388)	GAAA	GCGAAATGAAAAACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTCGGTTTCAAGAGGCATCTTTTT (SEQ ID NO: 389)
SauCas9	NNGRR;NNGRRT	-3	1	21	23	GTTTTAGTACTCTG (SEQ ID NO: 390)	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA (SEQ ID NO: 391)
SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAGTACTCTGTAAT (SEQ ID NO: 392)	GAAA	ATTACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA (SEQ ID NO: 393)

SauriCas9	NNGG	-3	1	21	21	GTTTTAGTACTCTG (SEQ ID NO: 394)	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT (SEQ ID NO: 395)
SauriCas9-KKH (SEQ ID NO: 401)	NNRG	-3	1	21	21	GTTTTAGTACTCTG (SEQ ID NO: 396)	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT (SEQ ID NO: 397)
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGAGCTA (SEQ ID NO: 398)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 399)
SpyCas9	NGG	-3	1	20	20	GTTTTAGAGCTA (SEQ ID NO: 400)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 401)
SpyCas9-NG	NG (NGG=NGA=NGT>NGC)	-3	1	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 402)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (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 ID NO: 408)	GAAA	GGTAAGTTGCTATAGTAAGGGCAACAGACCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTTACGGGCTCTCCCCATATTCAAAATAATGACAGACGAGCACCTTGGAGCATTTATCTCCGAGGTGCT (SEQ ID NO: 409)

cCas9-v16	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GUCUUAGUACUCUG (SEQ ID NO: 410)	GAAA	CAGAAUCUACUAAGACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU (SEQ ID NO: 411)
cCas9-v17	NNVRRN	-3	2	21	21	GUCUUAGUACUCUG (SEQ ID NO: 412)	GAAA	CAGAAUCUACUAAGACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU (SEQ ID NO: 413)
cCas9-v21	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GUCUUAGUACUCUG (SEQ ID NO: 414)	GAAA	CAGAAUCUACUAAGACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU (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	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTTT (SEQ ID NO: 421)
GeoCas9	NNNNCRAA		2	21	23	GUCAUAGUUCCCCUGA (SEQ ID NO: 421)	GAAA	UCAGGGUUACUAUGAUAAGGGCUUUCUGCCUAAGGCAGACUGACCCGCGGCGUUGGGGAUCGCCUGUCGCCCGCUUUUGGCGGGCAUUCCCCAUCCUU(SEQ ID NO: 422)

iSpyMac Cas9	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	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA (SEQ ID NO: 426)
ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 427)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 428)
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 429)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 430)
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 431)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 432)
SpyCas9-3var-NRTH	NRTH	-3	2	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 433)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (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	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA

								GTGGCACCGAGTCGGTGC (SEQ ID NO: 440)
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 441)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 442)
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 443)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 444)
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA (SEQ ID NO: 445)	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 446)
SpyCas9-xCas	NG;GAA;GAT	-3	2	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 447)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 448)
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG (SEQ ID NO: 449)	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (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)

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a Cas domain, e.g., a Cas9 domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas domain, a Cas nickase (nCas) domain, or a nuclease-inactive Cas (dCas) domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas9 domain, a Cas9 nickase (nCas9) domain, or a nuclease-inactive Cas9 (dCas9) domain. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 domain of Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises an S. pyogenes or an S. thermophilus Cas9, or a functional fragment thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 sequence, e.g., as described in Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; incorporated herein by reference. In some embodiments, the endonuclease domain or DNA binding domain comprises the HNH nuclease subdomain and/or the RuvC1 subdomain of a Cas, e.g., Cas9, e.g., as described herein, or a variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas polypeptide (e.g., enzyme), or a functional fragment thereof. In embodiments, the Cas polypeptide (e.g., enzyme) is selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (e.g., Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2,

CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, hyper accurate Cas9 variant (HypaCas9), homologues thereof, modified or engineered versions thereof, and/or functional fragments thereof. In embodiments, the Cas9 comprises one or more substiutions, e.g., selected from H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, and D1127A. In embodiments, the Cas9 comprises one or more mutations at positions selected from: D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987, e.g., one or more substitutions selected from D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas (e.g., Cas9) sequence from 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, or Staphylococcus aureus, or a fragment or variant thereof.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a Cpf1 domain, e.g., comprising one or more substitutions, e.g., at position D917, E1006A, D1255 or any combination thereof, e.g., selected from D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, and D917A/E1006A/D1255A.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises 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), or spCas9- LRVSQL (SEQ ID NO: 1700).

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises an amino acid sequence as listed in Table 37 below, or an amino acid sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the endonuclease domain or DNA-binding domain comprises an amino acid sequence that has no more than 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, or 50 differences (e.g., mutations) relative to any of the amino acid sequences described herein.

Table 37. Each of the Reference Sequences are incorporated by reference in their entirety.

Name	**Amino Acid Sequence or Reference Sequence**
Streptococcus pyogenes Cas9	
Exemplary Linker	SGSETPGTSESATPES (SEQ ID NO: 1023)
Exemplary Linker Motif	$(SGGS)_n$ (SEQ ID NO: 1595)
Exemplary Linker Motif	$(GGGS)_n$ (SEQ ID NO: 1596)
Exemplary Linker Motif	$(GGGGS)_n$ (SEQ ID NO: 1535)
Exemplary Linker Motif	$(G)_n$
Exemplary Linker Motif	$(EAAAK)_n$ (SEQ ID NO: 1534)
Exemplary Linker Motif	$(GGS)_n$
Exemplary Linker Motif	$(XP)_n$
Cas9 from Streptococcus pyogenes	NCBI Reference Sequence: NC_002737.2 and Uniprot Reference Sequence: Q99ZW2
Cas9 from Corynebacterium ulcerans	NCBI Refs: NC_015683.1, NC_017317.1
Cas9 from Corynebacterium diphtheria	NCBI Refs: NC_016782.1, NC_016786.1
Cas9 from Spiroplasma syrphidicola	NCBI Ref: NC_021284.1

Cas9 from Prevotella intermedia	NCBI Ref: NC_017861.1
Cas9 from Spiroplasma taiwanense	NCBI Ref: NC_021846.1
Cas9 from Streptococcus iniae	NCBI Ref: NC_021314.1
Cas9 from Belliella baltica	NCBI Ref: NC_018010.1
Cas9 from Psychroflexus torquisI	NCBI Ref: NC_018721.1
Cas9 from Streptococcus thermophilus	NCBI Ref: YP_820832.1
Cas9 from Listeria innocua	NCBI Ref: NP_472073.1
Cas9 from Campylobacter jejuni	NCBI Ref: YP_002344900.1
Cas9 from Neisseria meningitidis	NCBI Ref: YP_002342100.1
dCas9 (D10A and H840A)	
Catalytically inactive Cas9 (dCas9)	
Cas9 nickase (nCas9)	
Catalytically active Cas9	
CasY	((ncbi.nlm.nih.gov/protein/APG80656.1)

	>APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium])
CasX	uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53
CasX	>tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = Sulfolobus islandicus (strain REY15A) GN=SiRe_0771 PE=4 SV=1
Deltaproteobacteria CasX	
Cas12b/C2c1	((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR- associated endonuclease C2c1 OS = Alicyclobacillus acido-terrestris (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1)
BhCas12b (Bacillus hisashii)	NCBI Reference Sequence: WP_095142515
BvCas12b (Bacillus sp. V3-13)	NCBI Reference Sequence: WP_101661451.1
Wild-type Francisella novicida Cpf1	
Francisella novicida 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	
SaCas9	
SaCas9n	
PAM-binding SpCas9	
PAM-binding SpCas9n	
PAM-binding SpEQR Cas9	
PAM-binding SpVQR Cas9	
PAM-binding SpVRER Cas9	
PAM-binding SpVRQR Cas9	
SpyMacCas9	

In some embodiments, a Gene Writing polypeptide has an endonuclease domain comprising a Cas9 nickase, e.g., Cas9 H840A. In embodiments, the Cas9 H840A has the following amino acid sequence:

5

Cas9 nickase (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV
DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ
ID NO: 1597)

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., a wild-type M-MLV RT, e.g., comprising the following sequence:

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1598)

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., an M-MLV RT, e.g., comprising the following sequence:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 1556)

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase comprising the sequence of amino acids 659-1329 of NP_057933. In embodiments, the Gene Writing polypeptide further comprises one additional amino acid at the N-terminus of the sequence of amino acids 659-1329 of NP_057933, e.g., as shown below:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQDLREVN**
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
EMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAY
QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPV
AAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH
YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDAD
HTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
KLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPG
HQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 1599)

Core RT (bold), annotated per above
RNAseH (underlined), annotated per above

In embodiments, the Gene Writing polypeptide further comprises one additional amino acid at the C-terminus of the sequence of amino acids 659-1329 of NP_057933. In embodiments, the Gene Writing polypeptide comprises an RNaseH1 domain (e.g., amino acids 1178-1318 of NP_057933).

In some embodiments, a retroviral reverse transcriptase domain, e.g., M-MLV RT, may comprise one or more mutations from a wild-type sequence that may improve features of the RT, e.g., thermostability, processivity, and/or template binding. In some embodiments, an M-MLV RT domain comprises, relative to the M-MLV (WT) sequence above, one or more mutations, e.g., selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L,

e.g., a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and W313F. In some embodiments, an M-MLV RT used herein comprises the mutations D200N, L603W, T330P, T306K and W313F. In embodiments, the mutant M-MLV RT comprises the following amino acid sequence:

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1600)

In some embodiments, a Gene Writer polypeptide comprises a flexible linker between the endonuclease and the RT domain, e.g., a linker comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1601). In some embodiments, an RT domain of a Gene Writer polypeptide may be located C-terminal to the endonuclease domain. In some embodiments, an RT domain of a Gene Writer polypeptide may be located N-terminal to the endonuclease domain.

In some embodiments, a Gene Writer polypeptide comprises a dCas9 sequence comprising a D10A and/or H840A mutation, e.g., the following sequence:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF

GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 1602)

In some embodiments, a template RNA molecule for use in the system comprises, from 5' to 3' (1) a gRNA spacer; (2) a gRNA scaffold; (3) heterologous object sequence (4) 3' homology domain. In some embodiments:

(1) Is a Cas9 spacer of ~18-22 nt, e.g., is 20 nt

(2) Is a gRNA scaffold comprising one or more hairpin loops, e.g., 1, 2, of 3 loopd for associating the template with a nickase Cas9 domain. In some embodiments, the gRNA scaffold carries the sequence, from 5' to 3', GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGGACCGAGTCGGTCC(SEQ ID NO: 1603).

(3) In some embodiments, the heterologous object sequence is, e.g., 7-74, e.g., 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, or 70-80 nt or, 80-90 nt in length. In some embodiments, the first (most 5') base of the sequence is not C.

(4) In some embodiments, the 3' homology domain that binds the target priming sequence after nicking occurs is e.g., 3-20 nt, e.g., 7-15 nt, e.g., 12-14 nt. In some embodiments, the 3' homology domain has 40-60% GC content.

A second gRNA associated with the system may help drive complete integration. In some embodiments, the second gRNA may target a location that is 0-200 nt away from the first-strand nick, e.g., 0-50, 50-100, 100-200 nt away from the first-strand nick. In some embodiments, the second gRNA can only bind its target sequence after the edit is made, e.g., the gRNA binds a sequence present in the heterologous object sequence, but not in the initial target sequence.

In some embodiments, a Gene Writing system described herein is used to make an edit in HEK293, K562, U2OS, or HeLa cells. In some embodiment, a Gene Writing system is used to make an edit in primary cells, e.g., primary cortical neurons from E18.5 mice.

In some embodiments, a reverse transcriptase or RT domain (e.g., as described herein) comprises a MoMLV RT sequence or variant thereof. In embodiments, the MoMLV RT sequence comprises one or more mutations selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, and K103L. In embodiments, the MoMLV RT sequence comprises a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and/or W313F.

In some embodiments, an endonuclease domain (e.g., as described herein) comprises nCAS9, e.g., comprising the H840A mutation.

In some embodiments, the heterologous object sequence (e.g., of a system as described herein) is about 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or more, nucleotides in length.

In some embodiments, the RT and endonuclease domains are joined by a flexible linker, e.g., comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1601).

In some embodiments, the endonuclease domain is N-terminal relative to the RT domain. In some embodiments, the endonuclease domain is C-terminal relative to the RT domain.

In some embodiments, the system incorporates a heterologous object sequence into a target site by TPRT, e.g., as described herein.

In some embodiments, a system or method described herein involves a CRISPR DNA targeting enzyme or system described in US Pat. App. Pub. No. 20200063126, 20190002889, or 20190002875 (each of which is incorporated by reference herein in its entirety) or a functional

fragment or variant thereof. For instance, in some embodiments, a GeneWriter polypeptide or Cas endonuclease described herein comprises a polypeptide sequence of any of the applications mentioned in this paragraph, and in some embodiments a template RNA or guide RNA comprises a nucleic acid sequence of any of the applications mentioned in this paragraph.

The template nucleic acid (e.g., template RNA) component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein of the system. In some embodiments the template nucleic acid (e.g., template RNA) has a 3' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 3' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system. The binding region may associate the template nucleic acid (e.g., template RNA) with any of the polypeptide modules. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with an RNA-binding domain in the polypeptide. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with the reverse transcription domain of the polypeptide (e.g., specifically bind to the RT domain). For example, where the reverse transcription domain is derived from a non-LTR retrotransposon, the template nucleic acid (e.g., template RNA) may contain a binding region derived from a non-LTR retrotransposon, e.g., a 3' UTR from a non-LTR retrotransposon. In some embodiments, the template nucleic acid (e.g., template RNA) may associate with the DNA binding domain of the polypeptide, e.g., a gRNA associating with a Cas9-derived DNA binding domain. In some embodiments, the binding region may also provide DNA target recognition, e.g., a gRNA hybridizing to the target DNA sequence and binding the polypeptide, e.g., a Cas9 domain. In some embodiments, the template nucleic acid (e.g., template RNA) may associate with multiple components of the polypeptide, e.g., DNA binding domain and reverse transcription domain. For example, the template nucleic acid (e.g., template RNA) may comprise a gRNA region that associates with a Cas9-derived DNA binding domain and a 3' UTR from a non-LTR retrotransposon that associated with a non-LTR retrotransposon-derived reverse transcription domain.

In some embodiments a system or method described herein comprises a single template nucleic acid (e.g., template RNA). In some embodiments a system or method described herein comprises a plurality of template nucleic acids (e.g., template RNAs). For example, a system

described herein comprises a first RNA comprising (e.g., from 5' to 3') a sequence that binds the Gene Writer polypeptide (e.g., the DNA-binding domain and/or the endonuclease domain, e.g., a gRNA) and a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), and a second RNA (e.g., a template RNA) comprising (e.g., from 5' to 3') optionally a sequence that binds the Gene Writer polypeptide (e.g., that specifically binds the RT domain), a heterologous object sequence, and a 3' homology domain. In some embodiments, when the system comprises a plurality of nucleic acids, each nucleic acid comprises a conjugating domain. In some embodiments, a conjugating domain enables association of nucleic acid molecules, e.g., by hybridization of complementary sequences.

In some embodiments, a template nucleic acid molecule described herein comprises a 5' homology region and/or a 3' homology region. In some embodiments, the 5' homology region comprises a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule. In embodiments, the nucleic acid sequence in the target nucleic acid molecule is within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 5' relative to) a target insertion site, e.g., for a heterologous object sequence, e.g., comprised in the template nucleic acid molecule.

In some embodiments, the 3' homology region comprises a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule. In embodiments, the nucleic acid sequence in the target nucleic acid molecule is within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 3' relative to) a target insertion site, e.g., for a heterologous object sequence, e.g., comprised in the template nucleic acid molecule. In some embodiments, the 5' homology region is heterologous to the remainder of the template nucleic acid molecule. In some embodiments, the 3' homology region is heterologous to the remainder of the template nucleic acid molecule.

In some embodiments, a template nucleic acid (e.g., template RNA) comprises a 3' target homology domain. In some embodiments, a 3' target homology domain is disposed 3' of the heterologous object sequence and is complementary to a sequence adjacent to a site to be modified by a system described herein, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the sequence adjacent to a site to be modified by the system/Gene

Writer™. In some embodiments, the 3' homology region binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site in the target nucleic acid molecule. In some embodiments, binding of the 3' homology region to the target nucleic acid molecule permits initiation of target-primed reverse transcription (TPRT), e.g., with the 3' homology region acting as a primer for TPRT. In some embodiments, the 3' target homology domain anneals to the target site, which provides a binding site and the 3' hydroxyl for the initiation of TPRT by a Gene Writer polypeptide. In some embodiments, the 3' target homology domain is 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, or 25-30 nt in length, e.g., 10-17, 12-16, or 12-14 nt in length.

In some embodiments, the template nucleic acid, e.g., template RNA, may comprise a gRNA (e.g., pegRNA). In some embodiments, the template nucleic acid, e.g., template RNA, may bind to the Gene Writer™ polypeptide by interaction of a gRNA portion of the template nucleic acid with a template nucleic acid binding domain, e.g., a RNA binding domain (e.g., a heterologous RNA binding domain). In some embodiments, the heterologous RNA binding domain is a CRISPR/Cas protein, e.g., Cas9.

In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA adopts an underwound ribbon-like structure of gRNA bound to target DNA (e.g., as described in Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484). Without wishing to be bound by theory, this non-canonical structure is thought to be facilitated by rotation of every sixth nucleotide out of the RNA-DNA hybrid. Thus, in some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA may tolerate increased mismatching with the target site at some interval, e.g., every sixth base. In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA comprising homology to the target site may possess wobble positions at a regular interval, e.g., every sixth base, that do not need to base pair with the target site.

gRNAs with inducible activity

In some embodiments, a template nucleic acid, e.g., template RNA, comprises a guide RNA (gRNA) with inducible activity. Inducible activity may be achieved by the template nucleic acid, e.g., template RNA, further comprising (in addition to the gRNA) a blocking domain, wherein the sequence of a portion of or all of the blocking domain is at least partially complementary to a portion or all of the gRNA. The blocking domain is thus capable of hybridizing or substantially hybridizing to a portion of or all of the gRNA. In some embodiments, the blocking domain and inducibly active gRNA are disposed on the template nucleic acid, e.g., template RNA, such that the gRNA can adopt a first conformation where the blocking domain is hybridized or substantially hybridized to the gRNA, and a second conformationwhere the blocking domain is not hybridized or or not substantially hybridized to the gRNA. In some embodiments, in the first conformationthe gRNA is unable to bind to the Gene Writer polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)) or binds with substantially decreased affinity compared to an otherwise similar template RNA lacking the blocking domain. In some embodiments, in the second conformationthe gRNA is able to bind to the Gene Writer polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)). In some embodiments, whether the gRNA is in the first or second conformationcan influence whether the DNA binding or endonuclease activities of the Gene Writer polypeptide (e.g., of the CRISPR/Cas protein the Gene Writer polypeptide comprises) are active. In some embodiments, hybridization of the gRNA to the blocking domain can be disrupted using an opener molecule. In some embodiments, an opener molecule comprises an agent that binds to a portion or all of the gRNA or blocking domain and inhibits hybridization of the gRNA to the blocking domain. In some embodiments, the opener molecule comprises a nucleic acid, e.g., comprising a sequence that is partially or wholly complementary to the gRNA, blocking domain, or both. By choosing or designing an appropriate opener molecule, providing the opener molecule can promote a change in the conformation of the gRNA such that it can associate with a CRISPR/Cas protein and provide the associated functions of the CRISPR/Cas protein (e.g., DNA binding and/or endonuclease activity). Without wishing to be bound by theory, providing the opener molecule at a selected time and/or location may allow for spatial and temporal control of the activity of the gRNA, CRISPR/Cas protein, or Gene Writer system comprising the same. In some embodiments, the opener molecule is

exogenous to the cell comprising the Gene Writer polypeptide and or template nucleic acid. In some embodiments, the opener molecule comprises an endogenous agent (e.g., endogenous to the cell comprising the Gene Writer polypeptide and or template nucleic acid comprising the gRNA and blocking domain). For example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is an endogenous agent expressed in a target cell or tissue, e.g., thereby ensuring activity of a Gene Writer system in the target cell or tissue. As a further example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is absent or not substantially expressed in one or more non-target cells or tissues, e.g., thereby ensuring that activity of a Gene Writer system does not occur or substantially occur in the one or more non-target cells or tissues, or occurs at a reduced level compared to a target cell or tissue. Exemplary blocking domains, opener molecules, and uses thereof are described in PCT App. Publication WO2020044039A1, which is incorporated herein by reference in its entirety.In some embodiments, the template nucleic acid, e.g., template RNA, may comprise one or more UTRs (e.g. from an R2-type retrotransposon) and a gRNA. In some embodiments, the UTR facilitates interaction of the template nucleic acid (e.g., template RNA) with the writing domain, e.g., reverse transcriptase domain, of the Gene Writer polypeptide. In some embodiments, the gRNA facilitates interaction with the template nucleic acid binding domain (e.g., RNA binding domain) of the polypeptide. In some embodiments, the gRNA directs the polypeptide to the matching target sequence, e.g., in a target cell genome. In some embodiments, the template nucleic acid may contain only the reverse transcriptase binding motif (e.g. 3' UTR from R2) and the gRNA may be provided as a second nucleic acid molecule (e.g., second RNA molecule) for target site recognition. In some embodiments, the template nucleic acid containing the RT-binding motif may exist on the same molecule as the gRNA, but be processed into two RNA molecules by cleavage activity (e.g. ribozyme).

In some embodiments, a template RNA may be customized to correct a given mutation in the genomic DNA of a target cell (e.g., ex vivo or in vivo, e.g., in a target tissue or organ, e.g., in a subject). For example, the mutation may be a disease-associated mutation relative to the wild-type sequence. Without wishing to be bound by theory, , sets of empirical parameters help ensure optimal initial in silico designs of template RNAs or portions thereof). As a non-limiting illustrative example, for a selected mutation, the following design parameters may be employed. In some embodiments, design is initiated by acquiring approximately 500 bp (e.g., up to 50, 100,

150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 bp, and optionally at least 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, or 650 bp) flanking sequence on either side of the mutation to serve as the target region. In some embodiments, a template nucleic acid comprises a gRNA. Methodology for designing gRNAs is known to those of skill in the art. In some embodiments, a gRNA comprises a sequence (e.g., a CRISPR spacer) that binds a target site . In some embodiments, the sequence (e.g., a CRISPR spacer) that binds a target site for use in targeting a template nucleic acid to a target region is selected by considering the particular Gene Writer polypeptide (e.g., endonuclease domain or writing domain, e.g., comprising a CRISPR/Cas domain) being used (e.g., for Cas9, a protospacer-adjacent motif (PAM) of NGG immediately 3' of a 20 nt gRNA binding region). In some embodiments, the CRISPR spacer is selected by ranking first by whether the PAM will be disrupted by the Gene Writing induced edit. In some embodiments, disruption of the PAM may increase edit efficiency. In some embodiments, the PAM can be disrupted by also introducing (e.g., as part of or in addition to another modification to a target site in genomic DNA) a silent mutation (e.g., a mutation that does not alter an amino acid residue encoded by the target nucleic acid sequence, if any) in the target site during Gene Writing. In some embodiments, the CRISPR spacer is selected by ranking sequences by the proximity of their corresponding genomic site to the desired edit location. In some embodiments, the gRNA comprises a gRNA scaffold. In some embodiments, the gRNA scaffold used may be a standard scaffold (e.g., for Cas9, 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC-3'(SEQ ID NO: 1603)), or may contain one or more nucleotide substitutions. In some embodiments, the heterologous object sequence has at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 3' of the first strand nick (e.g., immediately 3' of the first strand nick or up to 1, 2, 3, 4, or 5 nucleotides 3' of the first strand nick), with the exception of any insertion, substitution, or deletion that may be written into the target site by the Gene Writer.In some embodiments, the 3' target homology domain contains at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 5' of the first strand nick (e.g., immediately 5' of the first strand nick or up to 1, 2, 3, 4, or 5 nucleotides 3' of the first strand nick).

In some embodiments, the template possesses one or more sequences aiding in association of the template with the Gene Writer polypeptide. In some embodiments, these sequences may be derived from retrotransposon UTRs. In some embodiments, the UTRs may be located flanking the desired insertion sequence. In some embodiments, a sequence with target site homology may be located outside of one or both UTRs. In some embodiments, the sequence with target site homology can anneal to the target sequence to prime reverse transcription. In some embodiments, the 5' and/or 3' UTR may be located terminal to the target site homology sequence, e.g., such that target primed reverse transcription excludes reverse transcription of the 5' and/or 3' UTR. In some embodiments, the Gene Writer system may result in the insertion of a desired payload without any additional sequence (e.g. gene expression unit without UTRs used to bind the Gene Writer protein).

The template nucleic acid (e.g., template RNA) can be designed to result in insertions, mutations, or deletions at the target DNA locus. In some embodiments, the template nucleic acid (e.g., template RNA) may be designed to cause an insertion in the target DNA. For example, the template nucleic acid (e.g., template RNA) may contain a heterologous sequence, wherein the reverse transcription will result in insertion of the heterologous sequence into the target DNA. In other embodiments, the RNA template may be designed to write a deletion into the target DNA. For example, the template nucleic acid (e.g., template RNA) may match the target DNA upstream and downstream of the desired deletion, wherein the reverse transcription will result in the copying of the upstream and downstream sequences from the template nucleic acid (e.g., template RNA) without the intervening sequence, e.g., causing deletion of the intervening sequence. In other embodiments, the template nucleic acid (e.g., template RNA) may be designed to write an edit into the target DNA. For example, the template RNA may match the target DNA sequence with the exception of one or more nucleotides, wherein the reverse transcription will result in the copying of these edits into the target DNA, e.g., resulting in mutations, e.g., transition or transversion mutations.

In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 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, or 100 nucleotides (and

optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 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 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 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, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 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 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a Gene Writer system is capable of producing a substitution into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 or more nucleotides. In some embodiments, the substitution is a transition mutation. In some embodiments, the substitution is a transversion mutation. In some embodiments, the substitution converts an adenine to a thymine, an adenine to a guanine, an adenine to a cytosine, a guanine to a thymine, a guanine to a cytosine, a guanine to an adenine, a thymine to a cytosine, a thymine to an adenine, a thymine to a guanine, a cytosine to an adenine, a cytosine to a guanine, or a cytosine to a thymine.

Methods and Compositions for Modified RNA (e.g., gRNA or template RNA)

In some embodiments, an RNA component of the system (e.g., a template RNA or a gRNA, e.g., as described herein) comprises one or more nucleotide modifications. In some embodiments, the modification pattern of a gRNA can significantly affect *in vivo* activity compared to unmodified or end-modified guides (e.g., as shown in Figure 1D from Finn et al. *Cell Rep* 22(9):2227-2235 (2018); incorporated herein by reference in its entirety). Without wishing to be bound by theory, this process may be due, at least in part, to a stabilization of the RNA conferred by the modifications. Non-limiting examples of such modifications may include

2'-O-methyl (2'-O-Me), 2'-0-(2-methoxyethyl) (2'-0-MOE), 2'- fluoro (2'-F), phosphorothioate (PS) bond between nucleotides, G-C substitutions, and inverted abasic linkages between nucleotides and equivalents thereof.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or the guide RNA comprises a 5' terminus region. In some embodiments, the template RNA or the guide RNA does not comprise a 5' terminus region. In some embodiments, the 5' terminus region comprises a CRISPR spacer region, e.g., as described with respect to sgRNA in Briner AE et al, Molecular Cell 56: 333-339 (2014) (incorporated herein by reference in its entirety; applicable herein, e.g., to all guide RNAs). In some embodiments, the 5' terminus region comprises a 5' end modification. In some embodiments, a 5' terminus region with or without a spacer region may be associated with a crRNA, trRNA, sgRNA and/or dgRNA. The CRISPR spacer region can, in some instances, comprise a guide region, guide domain, or targeting domain. In some embodiments, a target domain or target sequence may comprise a sequence of nucleic acid to which the guide region/domain directs a nuclease for cleavage. In some embodiments, a spyCas9 protein may be directed by a guide region/domain to a target sequence of a target nucleic acid molecule by the nucleotides present in the CRISPR spacer region.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or guide RNA, e.g., as described herein, comprises any of the sequences shown in Table 4 of WO2018107028A1, incorporated herein by reference in its entirety. In some embodiments, where a sequence shows a guide and/or spacer region, the composition may comprise this region or not. In some embodiments, a guide RNA comprises one or more of the modifications of any of the sequences shown in Table 4 of WO2018107028A1, e.g., as identified therein by a SEQ ID NO. In embodiments, the nucleotides may be the same or different, and/or the modification pattern shown may be the same or similar to a modification pattern of a guide sequence as shown in Table 4 of WO2018107028A1. In some embodiments, a modification pattern includes the relative position and identity of modifications of the gRNA or a region of the gRNA (e.g. 5' terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, 3' terminus region). In some embodiments, the modification pattern contains at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the modifications of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1, and/or over one or more regions of the sequence. In some embodiments,

the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1. In some embodiments, the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over one or more regions of the sequence shown in Table 4 of WO2018107028A1, e.g., in a 5 ' terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, and/or 3' terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of a sequence over the 5 ' terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the lower stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the bulge. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the upper stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the nexus. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 1. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 2. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the 3 ' terminus. In some embodiments, the modification pattern differs from the modification pattern of a sequence of Table 4 of WO2018107028A1, or a region (e.g. 5' terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3' terminus) of such a sequence, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from the modifications of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from modifications of a region (e.g. 5 ' terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3' terminus) of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or the gRNA comprises a 2'-O-methyl (2'-O-Me) modified nucleotide. In some embodiments, the gRNA comprises a 2'-O-(2-methoxy ethyl) (2'-O-moe) modified nucleotide. In some embodiments, the gRNA comprises a 2'-fluoro (2'- F) modified nucleotide. In some embodiments, the gRNA comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the gRNA comprises a 5' end modification, a 3' end modification, or 5' and 3' end modifications. In some embodiments, the 5' end modification comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the 5' end modification comprises a 2'-O-methyl (2'-O-Me), 2'-O-(2-methoxy ethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. In some embodiments, the 5' end modification comprises at least one phosphorothioate (PS) bond and one or more of a 2'-O-methyl (2'-O- Me), 2'-O-(2-methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. The end modification may comprise a phosphorothioate (PS), 2'-O-methyl (2'-O-Me) , 2'-O-(2- methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modification. Equivalent end modifications are also encompassed by embodiments described herein. In some embodiments, the template RNA or gRNA comprises an end modification in combination with a modification of one or more regions of the template RNA or gRNA. Additional exemplary modifications and methods for protecting RNA, e.g., gRNA, and formulae thereof, are described in WO2018126176A1, which is incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches are used to introduce modifications (e.g., 2′-OMe-RNA, 2′-F-RNA, and PS modifications) to a template RNA or guide RNA, for example, as described in Mir et al. *Nat Commun* 9:2641 (2018) (incorporated by reference herein in its entirety). In some embodiments, the incorporation of 2′-F-RNAs increases thermal and nuclease stability of RNA:RNA or RNA:DNA duplexes, e.g., while minimally interfering with C3′-endo sugar puckering. In some embodiments, 2′-F may be better tolerated than 2′-OMe at positions where the 2′-OH is important for RNA:DNA duplex stability. In some embodiments, a crRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., C10, C20, or C21 (fully modified), e.g., as dscribed in Supplementary Table 1 of Mir et al. *Nat Commun* 9:2641 (2018), incorporated herein by reference in its entirety. In some embodiments, a tracrRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., T2, T6, T7, or T8 (fully modified) of Supplementary Table 1 of Mir et al. *Nat Commun*

9:2641 (2018). In some embodiments, a crRNA comprises one or more modifications (e.g., as described herein) may be paired with a tracrRNA comprising one or more modifications, e.g., C20 and T2. In some embodiments, a gRNA comprises a chimera, e.g., of a crRNA and a tracrRNA (e.g., Jinek et al. *Science* 337(6096):816-821 (2012)). In embodiments, modifications from the crRNA and tracrRNA are mapped onto the single-guide chimera, e.g., to produce a modified gRNA with enhanced stability.

In some embodiments, gRNA molecules may be modified by the addition or subtraction of the naturally occurring structural components, e.g., hairpins. In some embodiments, a gRNA may comprise a gRNA with one or more 3' hairpin elements deleted, e.g., as described in WO2018106727, incorporated herein by reference in its entirety. In some embodiments, a gRNA may contain an added hairpin structure, e.g., an added hairpin structure in the spacer region, which was shown to increase specificity of a CRISPR-Cas system in the teachings of Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019). Additional modifications, including examples of shortened gRNA and specific modifications improving in vivo activity, can be found in US20190316121, incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches (e.g., as described in Mir et al. *Nat Commun* 9:2641 (2018); incorporated herein by reference in its entirety) are employed to find modifications for the template RNA. In embodiments, the modifications are identified with the inclusion or exclusion of a guide region of the template RNA. In some embodiments, a structure of polypeptide bound to template RNA is used to determine non-protein-contacted nucleotides of the RNA that may then be selected for modifications, e.g., with lower risk of disrupting the association of the RNA with the polypeptide. Secondary structures in a template RNA can also be predicted in silico by software tools, e.g., the RNAstructure tool available at rna.urmc.rochester.edu/RNAstructureWeb (Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); incorporated by reference herein in its entirety), e.g., to determine secondary structures for selecting modifications, e.g., hairpins, stems, and/or bulges.

Further included here are compositions and methods for the assembly of full or partial template RNA molecules (e.g., Gene Writing template RNA molecules optionally comprising a gRNA, or separate gRNA molecules). In some embodiments, RNA molecules may be assembled by the connection of two or more (e.g., two, three, four, five, six, seven, eight, nine, ten, or more) RNA segments with each other. In an aspect, the disclosure provides methods for

producing nucleic acid molecules, the methods comprising contacting two or more linear RNA segments with each other under conditions that allow for the 5′ terminus of a first RNA segment to be covalently linked with the 3′ terminus of a second RNA segment. In some embodiments, the joined molecule may be contacted with a third RNA segment under conditions that allow for the 5' terminus of the joined molecule to be covalently linked with the 3' terminus of the third RNA segment. In embodiments, the method further comprises joining a fourth, fifth, or additional RNA segments to the elongated molecule. This form of assembly may, in some instances, allow for rapid and efficient assembly of RNA molecules.

The present disclosure also provides compositions and methods for the connection (e.g., covalent connection) of crRNA molecules and tracrRNA molecules. In some embodiments, guide RNA molecules with specificity for different target sites can be generated using a single tracrRNA molecule/segment connected to a target site specific crRNA molecule/segment (e.g., as shown in FIG. 10 of US20160102322A1; incorporated herein by reference in its entirety). For example, FIG. 10 of US20160102322A1 shows four tubes with different crRNA molecules with crRNA molecule 3 being connected to a tracrRNA molecule to form a guide RNA molecule, thereby depicting an exemplary connection of two RNA segments to form a product RNA molecule.

The disclosure also provides compositions and methods for the production of template RNA molecules with specificity for a Gene Writer polypeptide and/or a genomic target site. In an aspect, the method comprises: (1) identification of the target site and desired modification thereto, (2) production of RNA segments including an upstream homology segment, a heterologous object sequence segment, a Gene Writer polypeptide binding motif, and a gRNA segment, and/or (3) connection of the four or more segments into at least one molecule, e.g., into a single RNA molecule. In some embodiments, some or all of the template RNA segments comprised in (2) are assembled into a template RNA molecule, e.g., one, two, three, or four of the listed components. In some embodiments, the segments comprised in (2) may be produced in further segmented molecules, e.g., split into at least 2, at least 3, at least 4, or at least 5 or more sub-segments, e.g., that are subsequently assembled, e.g., by one or more methods described herein.

In some embodiments, RNA segments may be produced by chemical synthesis. In some embodiments, RNA segments may be produced by in vitro transcription of a nucleic acid

template, e.g., by providing an RNA polymerase to act on a cognate promoter of a DNA template to produce an RNA transcript. In some embodiments, in vitro transcription is performed using, e.g., a T7, T3, or SP6 RNA polymerase, or a derivative thereof, acting on a DNA, e.g., dsDNA, ssDNA, linear DNA, plasmid DNA, linear DNA amplicon, linearized plasmid DNA, e.g., encoding the RNA segment, e.g., under transcriptional control of a cognate promoter, e.g., a T7, T3, or SP6 promoter. In some embodiments, a combination of chemical synthesis and in vitro transcription is used to generate the RNA segments for assembly. In embodiments, the gRNA, upstream target homology, and Gene Writer polypeptide binding segments are produced by chemical synthesis and the heterologous object sequence segment is produced by in vitro transcription. Without wishing to be bound by theory, in vitro transcription may be better suited for the production of longer RNA molecules. In some embodiments, reaction temperature for in vitro transcription may be lowered, e.g., be less than 37°C (e.g., between 0-10C, 10-20C, or 20-30C), to result in a higher proportion of full-length transcripts (Krieg *Nucleic Acids Res* 18:6463 (1990)). In some embodiments, a protocol for improved synthesis of long transcripts is employed to synthesize a long template RNA, e.g., a template RNA greater than 5 kb, such as the use of e.g., T7 RiboMAX Express, which can generate 27 kb transcripts in vitro (Thiel et al. J Gen Virol 82(6):1273-1281 (2001)). In some embodiments, modifications to RNA molecules as described herein may be incorporated during synthesis of RNA segments (e.g., through the inclusion of modified nucleotides or alternative binding chemistries), following synthesis of RNA segments through chemical or enzymatic processes, following assembly of one or more RNA segments, or a combination thereof.

In some embodiments, an mRNA of the system (e.g., an mRNA encoding a Gene Writer polypeptide) is synthesized in vitro using T7 polymerase-mediated DNA-dependent RNA transcription from a linearized DNA template, where UTP is optionally substituted with 1-methylpseudoUTP. In some embodiments, the transcript incorporates 5′ and 3′ UTRs, e.g., GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1604) and UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1605), or functional fragments or variants thereof, and optionally includes a poly-A tail, which can be encoded in the DNA template or added enzymatically following

transcription. In some embodiments, a donor methyl group, e.g., S-adenosylmethionine, is added to a methylated capped RNA with cap 0 structure to yield a cap 1 structure that increases mRNA translation efficiency (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

In some embodiments, the transcript from a T7 promoter starts with a GGG motif. In some embodiments, a transcript from a T7 promoter does not start with a GGG motif. It has been shown that a GGG motif at the transcriptional start, despite providing superior yield, may lead to T7 RNAP synthesizing a ladder of poly(G) products as a result of slippage of the transcript on the three C residues in the template strand from +1 to +3 (Imburgio et al. Biochemistry 39(34):10419-10430 (2000). For tuning transcription levels and altering the transcription start site nucleotides to fit alternative 5' UTRs, the teachings of Davidson et al. *Pac Symp Biocomput* 433-443 (2010) describe T7 promoter variants, and the methods of discovery thereof, that fulfill both of these traits.

In some embodiments, RNA segments may be connected to each other by covalent coupling. In some embodiments, an RNA ligase, e.g., T4 RNA ligase, may be used to connect two or more RNA segments to each other. When a reagent such as an RNA ligase is used, a 5′ terminus is typically linked to a 3′ terminus. In some embodiments, if two segments are connected, then there are two possible linear constructs that can be formed (i.e., (1) 5′-Segment 1-Segment 2-3′ and (2) 5′-Segment 2-Segment 1-3′). In some embodiments, intramolecular circularization can also occur. Both of these issues can be addressed, for example, by blocking one 5′ terminus or one 3′ terminus so that RNA ligase cannot ligate the terminus to another terminus. In embodiments, if a construct of 5′-Segment 1-Segment 2-3′ is desired, then placing a blocking group on either the 5′ end of Segment 1 or the 3′ end of Segment 2 may result in the formation of only the correct linear ligation product and/or prevent intramolecular circularization. Compositions and methods for the covalent connection of two nucleic acid (e.g., RNA) segments are disclosed, for example, in US20160102322A1 (incorporated herein by reference in its entirety), along with methods including the use of an RNA ligase to directionally ligate two single-stranded RNA segments to each other.

One example of an end blocker that may be used in conjunction with, for example, T4 RNA ligase, is a dideoxy terminator. T4 RNA ligase typically catalyzes the ATP-dependent ligation of phosphodiester bonds between 5′-phosphate and 3′-hydroxyl termini. In some embodiments, when T4 RNA ligase is used, suitable termini must be present on the termini

being ligated. One means for blocking T4 RNA ligase on a terminus comprises failing to have the correct terminus format. Generally, termini of RNA segments with a 5-hydroxyl or a 3′-phosphate will not act as substrates for T4 RNA ligase.

Additional exemplary methods that may be used to connect RNA segments is by click chemistry (e.g., as described in U.S. Patent Nos. 7,375,234 and 7,070,941, and US Patent Publication No. 2013/0046084, the entire disclosures of which are incorporated herein by reference). For example, one exemplary click chemistry reaction is between an alkyne group and an azide group (see FIG. 11 of US20160102322A1, which is incorporated herein by reference in its entirety). Any click reaction may potentially be used to link RNA segments (e.g., Cu-azide-alkyne, strain-promoted-azide-alkyne, staudinger ligation, tetrazine ligation, photo-induced tetrazole-alkene, thiol-ene, NHS esters, epoxides, isocyanates, and aldehyde-aminooxy). In some embodiments, ligation of RNA molecules using a click chemistry reaction is advantageous because click chemistry reactions are fast, modular, efficient, often do not produce toxic waste products, can be done with water as a solvent, and/or can be set up to be stereospecific.

In some embodiments, RNA segments may be connected using an Azide-Alkyne Huisgen Cycloaddition. reaction, which is typically a 1,3-dipolar cycloaddition between an azide and a terminal or internal alkyne to give a 1,2,3-triazole for the ligation of RNA segments. Without wishing to be bound by theory, one advantage of this ligation method may be that this reaction can initiated by the addition of required Cu(I) ions. Other exemplary mechanisms by which RNA segments may be connected include, without limitatoin, the use of halogens (F—, Br—, I—)/alkynes addition reactions, carbonyls/sulfhydryls/maleimide, and carboxyl/amine linkages. For example, one RNA molecule may be modified with thiol at 3′ (using disulfide amidite and universal support or disulfide modified support), and the other RNA molecule may be modified with acrydite at 5′ (using acrylic phosphoramidite), then the two RNA molecules can be connected by a Michael addition reaction. This strategy can also be applied to connecting multiple RNA molecules stepwise. Also provided are methods for linking more than two (e.g., three, four, five, six, etc.) RNA molecules to each other. Without wishing to be bound by theory, this may be useful when a desired RNA molecule is longer than about 40 nucleotides, e.g., such that chemical synthesis efficiency degrades, e.g., as noted in US20160102322A1 (incorporated herein by reference in its entirety).

By way of illustratoin, a tracrRNA is typically around 80 nucleotides in length. Such RNA molecules may be produced, for example, by processes such as in vitro transcription or chemical synthesis. In some embodiments, when chemical synthesis is used to produce such RNA molecules, they may be produced as a single synthesis product or by linking two or more synthesized RNA segments to each other. In embodiments, when three or more RNA segments are connected to each other, different methods may be used to link the individual segments together. Also, the RNA segments may be connected to each other in one pot (e.g., a container, vessel, well, tube, plate, or other receptacle), all at the same time, or in one pot at different times or in different pots at different times. In a non-limiting example, to assemble RNA Segments 1, 2 and 3 in numerical order, RNA Segments 1 and 2 may first be connected, 5′ to 3′, to each other. The reaction product may then be purified for reaction mixture components (e.g., by chromatography), then placed in a second pot, for connection of the 3′ terminus with the 5′ terminus of RNA Segment 3. The final reaction product may then be connected to the 5′ terminus of RNA Segment 3.

In another non-limiting example, RNA Segment 1 (about 30 nucleotides) is the target locus recognition sequence of a crRNA and a portion of Hairpin Region 1. RNA Segment 2 (about 35 nucleotides) contains the remainder of Hairpin Region 1 and some of the linear tracrRNA between Hairpin Region 1 and Hairpin Region 2. RNA Segment 3 (about 35 nucleotides) contains the remainder of the linear tracrRNA between Hairpin Region 1 and Hairpin Region 2 and all of Hairpin Region 2. In this example, RNA Segments 2 and 3 are linked, 5′ to 3′, using click chemistry. Further, the 5′ and 3′ end termini of the reaction product are both phosphorylated. The reaction product is then contacted with RNA Segment 1, having a 3′ terminal hydroxyl group, and T4 RNA ligase to produce a guide RNA molecule.

A number of additional linking chemistries may be used to connect RNA segments according to method of the invention. Some of these chemistries are set out in Table 6 of US20160102322A1, which is incorporated herein by reference in its entirety.

Gene Writers, e.g. Thermostable Gene Writers

While not wishing to be bound by theory, in some embodments, retrotransposases that evolved in cold environments may not function as well at human body temperature. This application provides a number of thermostable Gene Writers, including proteins derived from

avian retrotransposases. Exemplary avian transposase sequences in Table 3 include those of *Taeniopygia guttata* (zebra finch; transposon name R2-1_TG), *Geospiza fortis* (medium ground finch; transposon name R2-1_Gfo), *Zonotrichia albicollis* (white-throated sparrow; transposon name R2-1_ZA), and *Tinamus guttatus* (white-throated tinamou; transposon name R2-1_TGut).

Thermostability may be measured, e.g., by testing the ability of a Gene Writer to polymerize DNA *in vitro* at a high temperature (e.g., 37°C) and a low temberature (e.g., 25°C). Suitable conditions for assaying *in vitro* DNA polymerization activity (e.g., processivity) are described, e.g., in Bibillo and Eickbush, "High Processivity of the Reverse Transcriptase from a Non-long Terminal Repeat Retrotransposon" (2002) JBC 277, 34836-34845. In some embodiments, the thermostable Gene Writer polypeptide has an activity, e.g., a DNA polymerization activity, at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

In some embodiments, a GeneWriter polypeptide (e.g., a sequence of Table 1, 2, or 3 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) is stable in a subject chosen from a mammal (e.g., human) or a bird. In some embodiments, a GeneWriter polypeptide described herein is functional at 37°C. In some embodiments, a GeneWriter polypeptide described herein has greater activity at 37°C than it does at a lower temperature, e.g., at 30°C, 25°C, or 20°C. In some embodiments, a GeneWriter polypeptide described herein has greater activity in a human cell than in a zebrafish cell.

In some embodiments, a GeneWriter polypeptide is active in a human cell cultured at 37°C, e.g., using an assay of Example 6 or Example 7 herein.

In some embodiments, the assay comprises steps of: (1) introducing HEK293T cells into one or more wells of 6.4 mm diameter, at 10,000 cells/well, (2) incubating the cells at 37°C for 24 hr, (3) providing a transfection mixture comprising 0.5µl if FuGENE® HD transfection reagent and 80ng DNA (wherein the DNA is a plasmid comprising, in order, (a) CMV promoter, (b) 100 bp of sequence homologous to the 100 bp upstream of the target site, (c) sequence encoding a 5' untranslated region that binds the GeneWriter protein, (d) sequence encoding the GeneWriter protein, (e) sequence encoding a 3' untranslated region that binds the GeneWriter protein (f) 100 bp of sequence homologous to the 100 bp downstream of the target site, and (g) BGH polyadenylation sequence) and 10µl Opti-MEM and incubating for 15 min at room temperature, (4) adding the transfection mixture to the cells, (5) incubating the cells for 3 days,

and (6) assaying integration of the exogenous sequence into a target locus (e.g., rDNA) in the cell genome, e.g., wherein one or more of the preceding steps are performed as described in Example 6 herein.

In some embodiments, the GeneWriter polypeptide results in insertion of the heterologous object sequence (e.g., the GFP gene) into the target locus (e.g., rDNA) at an average copy number of at least 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, or 5 copies per genome. In some embodiments, a cell described herein (e.g., a cell comprising a heterologous sequence at a target insertion site) comprises the heterologous object sequence at an average copy number of at least 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, or 5 copies per genome.

In some embodiments, a GeneWriter causes integration of a sequence in a target RNA with relatively few truncation events at the terminus. For instance, in some embodiments, a Gene Writer protein (e.g., of SEQ ID NO: 1016) results in about 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90%, or 86.17% of integrants into the target site being non-truncated, as measured by an assay described herein, e.g., an assay of Example 6 and Figure 8. In some embodiments, a Gene Writer protein (e.g., of SEQ ID NO: 1016) results in at least about 30%, 40%, 50%, 60%, 70%, 80%, or 90% of integrants into the target site being non-truncated, as measured by an assay described herein. In some embodiments, an integrant is classified as truncated versus non-truncated using an assay comprising amplification with a forward primer situated 565bp from the end of the element (e.g., a wild-type transposon sequence, e.g., of *Taeniopygia guttata*) and a reverse primer situated in the genomic DNA of the target insertion site, e.g., rDNA. In some embodiments, the number of full-length integrants in the target insertion site is greater than the number of integrants truncated by 300-565 nucleotides in the target insertion site, e.g., the number of full-length integrants is at least 1.1x, 1.2x, 1.5x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, or 10x the number of the truncated integrants, or the number of full-length integrants is at least 1.1x-10x, 2x-10x, 3x-10x, or 5x-10x the number of the truncated integrants.

In some embodiments, a system or method described herein results in insertion of the heterologous object sequence only at one target site in the genome of the target cell. Insertion can be measured, e.g., using a threshold of above 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, e.g., as described in Example 8. In some embodiments, a system or method described herein results in insertion of the heterologous object sequence wherein less than 1%, 1.5%, 2%, 2.5%,

3%, 3.5%, 4%, 4.5%, 5%, 10%, 20%, 30%, 40%, or 50% of insertions are at a site other than the target site, e.g., using an assay described herein, e.g., an assay of Example 8.

In some embodiments, a system or method described herein results in "scarless" insertion of the heterologous object sequence, while in some embodiments, the target site can show deletions or duplications of endogenous DNA as a result of insertion of the heterologous sequence. The mechanisms of different retrotransposons could result in different patterns of duplications or deletions in the host genome occurring during retrotransposition at the target site. In some embodiments, the system results in a scarless insertion, with no duplications or deletions in the surrounding genomic DNA. In some embodiments, the system results in a deletion of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA upstream of the insertion. In some embodiments, the system results in a deletion of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA downstream of the insertion. In some embodiments, the system results in a duplication of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA upstream of the insertion. In some embodiments, the system results in a duplication of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA downstream of the insertion.

In some embodiments, a GeneWriter described herein, or a DNA-binding domain thereof, binds to its target site specifically, e.g., as measured using an assay of Example 21. In some embodiments, the GeneWriter or DNA-binding domain thereof binds to its target site more strongly than to any other binding site in the human genome. For example, in some embodiments, in an assay of Example 21, the target site represents more than 50%, 60%, 70%, 80%, 90%, or 95% of binding events of the GeneWriter or DNA-binding domain thereof to human genomic DNA.

Genetically engineered, e.g., dimerized GeneWriters

Some non-LTR retrotransposons utilize two subunits to complete retrotransposition (Christensen et al PNAS 2006). In some embodiments, a retrotransposase described herein comprises two connected subunits as a single polypeptide. For instance, two wild-type retrotransposases could be joined with a linker to form a covalently "dimerized" protein (see Figure 17). In some embodiments, the nucleic acid coding for the retrotransposase codes for two retrotransposase subunits to be expressed as a single polypeptide. In some embodiments, the subunits are connected by a peptide linker, such as has been described herein in the section entitled "Linker" and, e.g., in Chen et al Adv Drug Deliv Rev 2013. In some embodiments, the

two subunits in the polypeptide are connected by a rigid linker. In some embodiments, the rigid linker consists of the motif $(EAAAK)_n$ (SEQ ID NO: 1534). In other embodiments, the two subunits in the polypeptide are connected by a flexible linker. In some embodiments, the flexible linker consists of the motif $(Gly)_n$. In some embodiments, the flexible linker consists of the motif $(GGGGS)_n$ (SEQ ID NO: 1535). In some embodiments, the rigid or flexible linker consists of 1, 2, 3, 4, 5, 10, 15, or more amino acids in length to enable retrotransposition. In some embodiments, the linker consists of a combination of rigid and flexible linker motifs. In some embodiments, a Gene Writer polypeptide may comprise a linker, e.g., a peptide linker, e.g., a linker as described in Table 38. Table 38 provides linker sequences for increasing expression, stability, and function of Gene Writer polypeptides comprising multiple functional domains.

Table 38. Exemplary linker sequences

Amino Acid Sequence	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

Based on mechanism, not all functions are required from both retrotransposase subunits. In some embodiments, the fusion protein may consist of a fully functional subunit and a second subunit lacking one or more functional domains. In some embodiments, one subunit may lack reverse transcriptase functionality. In some embodiments, one subunit may lack the reverse transcriptase domain. In some embodiments, one subunit may possess only endonuclease activity.

In some embodiments, a GeneWriter described herein has a covalently dimerized configuration, e.g., as shown in any of Figs. 17A-17F of PCT/US2019/048607, incorporated herein by reference. The proteins depicted are: Fig. 17A: a wild-type full length enzyme. Fig. 17B, two full-length enzymes (each comprising a DNA-binding domain, an RNA-binding domain, a reverse transcriptase domain, and an endonuclease domain) connected by a linker. Fig. 17C, a DNA binding domain and an RNA binding domain connected by a linker to a full-

length enzyme. Fig. 17D, a DNA-binding domain and an RNA-binding domain connected by a linker to an RNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. Fig. 17E, a DNA-binding domain connected by a first linker to an RNA-binding domain, which is connected by a second linker to a second RNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. Fig. 17F, a DNA-binding domain connected by a first linker to an RNA-binding domain, which is connected by a second linker to a plurality of RNA-binding domains (in this figure, the molecule comprises three RNA-binding domains), which are connected by a linker to a reverse transcriptase domain and an endonuclease domain. In some embodiments, each R2 binds UTRs in the template RNA. In some embodiments, at least one module comprises a reverse transcriptase domain and an endonuclease domain. In some embodiments, the protein comprises a plurality of RNA-binding domains. In some embodiments, the modular system is split and is only active when it binds on DNA where the system uses two different DNA binding modules, e.g., a first protein comprising a first DNA binding module that is fused to an RNA binding module that recruits the RNA template for target primed reverse transcription, and second protein that comprises a second DNA binding module that binds at the site of intergration and is fused to the reverse transcription and endonuclease modules. In some embodiments, the nucleic acid encoding the GeneWriter comprises an intein such that the GeneWriter protein is expressed from two separate genes and is fused by protein splicing after being translated. In some embodiments, the GeneWriter is derived from a non-LTR protein, e.g., an R2 protein.

In some embodiments, one subunit may possess only an endonuclease domain. In some embodiments, the two subunits comprising the single polypeptide may provide complimentary functions.

In some embodiments, one subunit may lack endonuclease functionality. In some embodiments, one subunit may lack the endonuclease domain. In some embodiments, one subunit may possess only reverse transcriptase activity. In some embodiments, one subunit may possess only a reverse transcriptase domain. In some embodiments, one subunit may possess only DNA-dependent DNA synthesis functionality.

Linkers:

In some embodiments, domains of the compositions and systems described herein (e.g., the endonuclease and reverse transcriptase domains of a polypeptide or the DNA binding domain and reverse transcriptase domains of a polypeptide) may be joined by a linker. A composition described herein comprising a linker element has the general form S1-L-S2, wherein S1 and S2 may be the same or different and represent two domain moieties (e.g., each a polypeptide or nucleic acid domain) associated with one another by the linker. In some embodiments, a linker may connect two polypeptides. In some embodiments, a linker may connect two nucleic acid molecules. In some embodiments, a linker may connect a polypeptide and a nucleic acid molecule. A linker may be a chemical bond, e.g., one or more covalent bonds or non-covalent bonds. A linker may be flexible, rigid, and/or cleavable. In some embodiments, the linker is a peptide linker. Generally, a peptide linker is at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acids in length, e.g., 2-50 amino acids in length, 2-30 amino acids in length.

The most commonly used flexible linkers have sequences consisting primarily of stretches of Gly and Ser residues ("GS" linker). Flexible linkers may be useful for joining domains that require a certain degree of movement or interaction and may include small, non-polar (e.g. Gly) or polar (e.g. Ser or Thr) amino acids. Incorporation of Ser or Thr can also maintain the stability of the linker in aqueous solutions by forming hydrogen bonds with the water molecules, and therefore reduce unfavorable interactions between the linker and the other moieties. Examples of such linkers include those having the structure $[GGS]^{\geq 1}$ or $[GGGS]^{\geq 1}$ (SEQ ID NO: 1536). Rigid linkers are useful to keep a fixed distance between domains and to maintain their independent functions. Rigid linkers may also be useful when a spatial separation of the domains is critical to preserve the stability or bioactivity of one or more components in the agent. Rigid linkers may have an alpha helix-structure or Pro-rich sequence, (XP)n, with X designating any amino acid, preferably Ala, Lys, or Glu. Cleavable linkers may release free functional domains in vivo. In some embodiments, linkers may be cleaved under specific conditions, such as the presence of reducing reagents or proteases. In vivo cleavable linkers may utilize the reversible nature of a disulfide bond. One example includes a thrombin-sensitive sequence (e.g., PRS) between the two Cys residues. In vitro thrombin treatment of CPRSC (SEQ ID NO: 1537) results in the cleavage of the thrombin-sensitive sequence, while the reversible disulfide linkage remains intact. Such linkers are known and described, e.g., in Chen et al. 2013. Fusion Protein Linkers: Property, Design and Functionality. Adv Drug Deliv Rev.

65(10): 1357–1369. In vivo cleavage of linkers in compositions described herein may also be carried out by proteases that are expressed in vivo under pathological conditions (e.g. cancer or inflammation), in specific cells or tissues, or constrained within certain cellular compartments. The specificity of many proteases offers slower cleavage of the linker in constrained compartments.

In some embodiments the amino acid linkers are (or are homologous to) the endogenous amino acids that exist between such domains in a native polypeptide. In some embodiments the endogenous amino acids that exist between such domains are substituted but the length is unchanged from the natural length. In some embodiments, additional amino acid residues are added to the naturally existing amino acid residues between domains.

In some embodiments, the amino acid linkers are designed computationally or screened to maximize protein function (Anad et al., FEBS Letters, 587:19, 2013).

In addition to being fully encoded on a single transcript, a polypeptide can be generated by separately expressing two or more polypeptide fragments that reconstitute the holoenzyme. In some embodiments, the Gene Writer polypeptide is generated by expressing as separate subunits that reassemble the holoenzyme through engineered protein-protein interactions. In some embodiments, reconstitution of the holoenzyme does not involve covalent binding between subunits. Peptides may also fuse together through *trans*-splicing of inteins (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). In some embodiments, the Gene Writer holoenzyme is expressed as separate subunits that are designed to create a fusion protein through the presence of split inteins in the subunits. In some embodiments, the Gene Writer holoenzyme is reconstituted through the formation of covalent linkages between subunits. In some embodiments, protein subunits reassemble through engineered protein-protein binding partners, e.g., SpyTag and SpyCatcher (Zakeri et al. PNAS 109, E690-E697 (2012)). In some embodiments, an additional domain described herein, e.g., a Cas9 nickase, is expressed as a separate polypeptide that associates with the Gene Writer polypeptide through covalent or non-covalent interactions as described above. In some embodiments, the breaking up of a Gene Writer polypeptide into subunits may aid in delivery of the protein by keeping the nucleic acid encoding each part within optimal packaging limits of a viral delivery vector, e.g., AAV (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). In some embodiments, the Gene Writer

polypeptide is designed to be dimerized through the use of covalent or non-covalent interactions as described above.

In contrast to other types of reverse transcription machines, e.g., retroviral RTs and LTR retrotransposons, reverse transcription in non-LTR retrotransposons like R2 is performed only on RNA templates containing specific recognition sequences. The R2 retrotransposase requires its template to contain a minimal 3' UTR region in order to initiate TPRT (Luan and Eickbush Mol Cell Biol 15, 3882-91 (1995)). In some embodiments, the Gene Writer polypeptide is derived from a retrotransposase with a required binding motif and the template RNA is designed to contain said binding motif, such that there is specific retrotransposition of only the desired template (see, e.g., Example 22). In some embodiments, the Gene Writer polypeptide is derived from a retrotransposon selected from Table 3 and the 3' UTR on the RNA template comprises the 3' UTR from the same retrotransposon in Table 3.

It is a known phenomenon that some mobile elements are capable of moving non-self elements, e.g., L1 retrotransposase facilitates the movement of non-autonomous Alu and SVA elements in the human genome (Craig, Mobile DNA III, ASM, ed. 3 (2105)). Recent studies have mapped various transposable elements present in the human genome, including non-LTR retrotransposons (Kojima Mobile DNA 9 (2018)). Given active transposition in the human genome has been linked to diseases, e.g., the role of LINE-1 retrotransposition in oncogenesis (Rodriguez-Martin et al. Nat Genet (2020)), it is desirable that a Gene Writer does not recognize and mobilize transposable elements or pseudoelements. In some embodiments, a Gene Writer polypeptide does not lead to the mobilization of any endogenous human DNA. In some embodiments, a Gene Writer is derived from a retrotransposase that is not present in the human genome. In some embodiments, a Gene Writer derived from a retrotransposase present in the human genome (see, for example, Kojima Mobile DNA 9 (2018)) is engineered such that it recognizes heterologous sequences in the template RNA and no longer recognizes the natural UTRs of the parental retrotransposon, e.g., has a heterologous RNA binding domain that does not associate with the 3' UTR present in the human genome. In some embodiments, a Gene Writer comprises an RNA binding domain that does not recognize any sequences present in the human genome.

For optimizing protein expression, it can be helpful to provide tunable controls that can be used to modulate protein activity. In some embodiments, a tunable system may comprise at least one effector module that is responsive to at least one stimulus. The system may be, but is not limited to, a destabilizing domain (DD) system. This system is further taught in PCT/US2018/020704, as well as U.S. Provisional Patent Application No. 62/320,864 filed April 11, 2016, 62/466,596 filed March 3, 2017 and the International Publication WO2017/180587 (the contents each of which are herein incorporated by reference in their entirety). In some embodiments, the tunable system may comprise a first effector module. In some embodiments, the effector module may comprise a first stimulus response element (SRE) operably linked to at least one payload. In one aspect, the payload may be an immunotherapeutic agent. In one aspect, the first SRE of the composition may be responsive to or interact with at least one stimulus. In some embodiments, the first SRE may comprise a destabilizing domain (DD). The DD may be derived from a parent protein or from a mutant protein having one, two, three, or more amino acid mutations compared to the parent protein. In some embodiments, the parent protein may be selected from, but is not limited to, human protein FKBP, comprising the amino acid sequence of SEQ. ID NO. 3 of PCT/US2018/020704, incorporated herein by reference in its entirety; human DHFR (hDHFR), comprising the amino acid sequence of SEQ. ID NO. 2 of PCT/US2018/020704, incorporated herein by reference in its entirety; *E. coli* DHFR, comprising the amino acid sequence of SEQ. ID NO. 1 of PCT/US2018/020704, incorporated herein by reference in its entirety; PDE5, comprising the amino acid sequence of SEQ. ID NO. 4 of PCT/US2018/020704, incorporated herein by reference in its entirety; PPAR, gamma comprising the amino acid sequence of SEQ. ID NO. 5 of PCT/US2018/020704, incorporated herein by reference in its entirety; CA2, comprising the amino acid sequence of SEQ. ID NO. 6 of PCT/US2018/020704, incorporated herein by reference in its entirety; or NQ02, comprising the amino acid sequence of SEQ. ID NO. 7 of PCT/US2018/020704, incorporated herein by reference in its entirety. In some embodiments, the tunable controls are applied to the Gene Writer polypeptide, such that, e.g., a DD and stimulus can be used to modulate template integration efficiency. In some embodiments, the tunable controls are applied to one or more peptides encoded within the heterologous object sequence of the template, such that, e.g., a DD and stimulus can be used to modulate activity of a genomically integrated payload. In certain embodiments, the payload comprising the DD may be a therapeutic protein, e.g., a functional

copy of an endogenously mutated gene. In certain embodiments, the payload comprising the DD may be a heterologous protein, e.g., a CAR.

As used in the systems and methods provided here, Gene Writers™ may be provided as either polypeptides, or nucleic acids encoding them.

Nucleic acid features

Elements of systems provided by the invention may be provided as nucleic acids, for example, a template nucleic acid (e.g., template RNA) as described, inter alia, in the claims and enumerated embodiments, as well as, in certain embodiments, a nucleic acid encoding a Gene Writer™ polypeptide (e.g., a retrotransposase). In various embodiments, the nucleic acids are in operative association with additional genetic elements, such as tissue-specific expression-control sequence(s) (e.g., tissue-specific promoters and tissue-specific microRNA recognition sequences), as well as additional elements, such as inverted repeats (e.g., inverted terminal repeats, such as elements from or derived from viruses, e.g., AAV ITRs) and tandem repeats, homology regions (segments with various degrees of homology to a target DNA), UTRs (5', 3', or both 5' and 3' UTRs), and various combinations of the foregoing. The nucleic acid elements of the systems provided by the invention can be provided in a variety of topologies, including single-stranded, double-stranded, circular, linear, linear with open ends, linear with closed ends, and particular versions of these, such as doggybone DNA (dbDNA), close-ended DNA (ceDNA).

In certain particular embodiments, tissue-specific expression-control sequence(s) refers to one or more of the sequences in: Table 3 of WO2020014209, incorporated herein by reference, omitting the last column thereof (SEQ ID NO reference); or Table 4 of WO2020014209, incorporated herein by reference, omitting the last column thereof (SEQ ID NO reference).

In some embodiments, a nucleic acid described herein (e.g., template nucleic acid or a template encoding a retrotransposase) comprises a promoter sequence, e.g., a tissue specific promoter. In some embodiments, the tissue-specific promoter is used to increase the target-cell specificity of a Gene Writer™ system. For instance, the promoter can be chosen on the basis that it is active in a target cell type but not active in (or active at a lower level in) a non-target cell type. Thus, even if the nucleic acid encoding the polypeptide was delivered into a non-target

cell, it would not drive expression (or only drive low level expression) of the retrotransposase, limiting integration of the RNA template. A system having a tissue-specific promoter sequence in the retrotransposase DNA may also be used in combination with a microRNA binding site, e.g., encoded in the retrotransposase DNA, e.g., as described herein. A system having a tissue-specific promoter sequence in the retrotransposase DNA may also be used in combination with an RNA template containing a heterologous object sequence driven by a tissue-specific promoter, e.g., to achieve higher levels of integration and heterologous object sequence expression in target cells than in non-target cells.

In some embodiments, a nucleic acid described herein (e.g., an RNA encoding a Gene Writer™ polypeptide, or a DNA encoding the RNA, or a template nucleic acid) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a Gene Writer™ system. For instance, the microRNA binding site can be chosen on the basis that it is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the RNA encoding the Gene Writer™ polypeptide is present in a non-target cell, it would be bound by the miRNA, and when the RNA encoding the Gene Writer™ polypeptide is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the RNA encoding the Gene Writer™ polypeptide may reduce production of the Gene Writer™ polypeptide, e.g., by degrading the mRNA encoding the polypeptide or by interfering with translation. Accordingly, the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells. A system having a microRNA binding site in the RNA encoding the Gene Writer™ polypeptide (or encoded in the DNA encoding the RNA) may also be used in combination with a template RNA regulated by a second microRNA binding site, e.g., as described herein in the section entitled "Template component of Gene Writer™ gene editor system."

In some embodiments, a nucleic acid component of a system provided by the invention a sequence (e.g., retrotransposase or a heterologous object sequence) is flanked by untranslated regions (UTRs) that modify protein expression levels (sometimes referred to as UTR_{exp}) (Figures 11 and 15, Example 6). The effects of various 5' and 3' UTRs on protein expression are known in the art. For example, in some embodiments, the coding sequence may be preceded by a 5'

UTR that modifies RNA stability or protein translation. In some embodiments, the sequence may be followed by a 3' UTR that modifies RNA stability or translation. In some embodiments, the sequence may be preceded by a 5' UTR and followed by a 3' UTR that modify RNA stability or translation. In some embodiments, the 5' and/or 3' UTR may be selected from the 5' and 3' UTRs of complement factor 3 (C3) (cactcctccccatcctctccctctgtccctctgtccctctgaccctgcactgtcccagcacc(SEQ ID NO: 1606)) or orosomucoid 1 (ORM1) (caggacacagccttggatcaggacagagacttgggggccatcctgcccctccaacccgacatgtgtacctcagcttttttccctcacttgcat caataaagcttctgtgtttggaacagctaa(SEQ ID NO: 1607)) (Asrani et al. *RNA Biology* 2018). In certain embodiments, the 5' UTR is the 5' UTR from C3 and the 3' UTR is the 3' UTR from ORM1.

In certain embodiments, a 5' UTR and 3' UTR for protein expression, e.g., mRNA (or DNA encoding the RNA) for a Gene Writer polypeptide or heterologous object sequence, comprise optimized expression sequences. In some embodiments, the 5' UTR comprises GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1608)and/or the 3' UTR comprising UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCC AGCCCCUCCUCCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1609), e.g., as described in Richner et al. *Cell* 168(6): P1114-1125 (2017), the sequences of which are incorporated herein by reference.

In some embodiments, a 5' and/or 3' UTR may be selected to enhance protein expression. In some embodiments, a 5' and/or 3' UTR may be selected to modify protein expression such that overproduction inhibition is minimized. In some embodiments, UTRs are around a coding sequence, e.g., outside the coding sequence and in other embodiments proximal to the coding sequence, In some embodiments additional regulatory elements (e.g., miRNA binding sites, cis-regulatory sites) are included in the UTRs.

In some embodiments, an open reading frame (ORF) of a Gene Writer system, e.g., an ORF of an mRNA (or DNA encoding an mRNA) encoding a Gene Writer polypeptide or one or more ORFs of an mRNA (or DNA encoding an mRNA) of a heterologous object sequence, is flanked by a 5' and/or 3' untranslated region (UTR) that enhances the expression thereof. In some embodiments, the 5' UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5'-

GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3' (SEQ ID NO: 1610). In some embodiments, the 3' UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3'(SEQ ID NO: 1611). This combination of 5' UTR and 3' UTR has been shown to result in desirable expression of an operably linked ORF by Richner et al. *Cell* 168(6): P1114-1125 (2017), the teachings and sequences of which are incorporated herein by reference. In some embodiments, a system described herein comprises a DNA encoding a transcript, wherein the DNA comprises the corresponding 5' UTR and 3' UTR sequences, with T substituting for U in the above-listed sequence). In some embodiments, a DNA vector used to produce an RNA component of the system further comprises a promoter upstream of the 5' UTR for initiating in vitro transcription, e.g, a T7, T3, or SP6 promoter. The 5' UTR above begins with GGG, which is a suitable start for optimizing transcription using T7 RNA polymerase. For tuning transcription levels and altering the transcription start site nucleotides to fit alternative 5' UTRs, the teachings of Davidson et al. *Pac Symp Biocomput* 433-443 (2010) describe T7 promoter variants, and the methods of discovery thereof, that fulfill both of these traits.

Circular RNAs in Gene Writing Systems

Circular RNAs (circRNA) have been found to occur naturally in cells and have been found to have diverse functions, including both non-coding and protein coding roles in human cells. It has been shown that a circRNA can be engineered by incorporating a self-splicing intron into an RNA molecule (or DNA encoding the RNA molecule) that results in circularization of the RNA, and that an engineered circRNA can have enhanced protein production and stability (Wesselhoeft et al. *Nature Communications* 2018). It is contemplated that it may be useful to employ circular and/or linear RNA states during the formulation, delivery, or Gene Writing reaction within the target cell. Thus, in some embodiments of any of the aspects described herein, a Gene Writing system comprises one or more circular RNAs (circRNAs). In some embodiments of any of the aspects described herein, a Gene Writing system comprises one or more linear RNAs. In some embodiments, a nucleic acid as described herein (e.g., a nucleic acid

molecule encoding a Gene Writer polypeptide, or both) is a circRNA. In some embodiments, a circular RNA molecule encodes the Gene Writer™ polypeptide. In some embodiments, the circRNA molecule encoding the Gene Writer™ polypeptide is delivered to a host cell. In some embodiments, a circular RNA molecule encodes a recombinase, e.g., as described herein. In some embodiments, the circRNA molecule encoding the recombinase is delivered to a host cell. In some embodiments, the circRNA molecule encoding the Gene Writer polypeptide is linearized (e.g., in the host cell) prior to translation.

In some embodiments, nucleic acid (e.g., encoding a Gene Writer polypeptide, or a template RNA, or both) is provided as circRNA. In some embodiments, the circRNA comprises one or more ribozyme sequences. In some embodiments, the ribozyme sequence is activated for autocleavage, e.g., in a host cell, e.g., thereby resulting in linearization of the circRNA. In some embodiments, the ribozyme is activated when the concentration of magnesium reaches a sufficient level for cleavage, e.g., in a host cell. In some embodiments the circRNA is maintained in a low magnesium environment prior to delivery to the host cell. In some embodiments, the ribozyme is a protein-responsive ribozyme. In some embodiments, the ribozyme is a nucleic acid-responsive ribozyme.

In some embodiments, the circRNA is linearized in the nucleus of a target cell. In some embodiments, linearization of a circRNA in the nucleus of a cell involves components present in the nucleus of the cell, e.g., to activate a cleavage event. For example, the B2 and ALU retrotransposons contain self-cleaving ribozymes whose activity is enhanced by interaction with the Polycomb protein, EZH2 (Hernandez et al. PNAS 117(1):415-425 (2020)). Thus, in some embodiments, a ribozyme, e.g., a ribozyme from a B2 or ALU element, that is responsive to a nuclear element, e.g., a nuclear protein, e.g., a genome-interacting protein, e.g., an epigenetic modifier, e.g., EZH2, is incorporated into a circRNA, e.g., of a Gene Writing system. In some embodiments, nuclear localization of the circRNA results in an increase in autocatalytic activity of the ribozyme and linearization of the circRNA.

In some embodiments, an inducible ribozyme (e.g., in a circRNA as described herein) is created synthetically, for example, by utilizing a protein ligand-responsive aptamer design. A system for utilizing the satellite RNA of tobacco ringspot virus hammerhead ribozyme with an MS2 coat protein aptamer has been described (Kennedy et al. Nucleic Acids Res 42(19):12306-12321 (2014), incorporated herein by reference in its entirety) that results in activation of the

ribozyme activity in the presence of the MS2 coat protein. In embodiments, such a system responds to protein ligand localized to the cytoplasm or the nucleus. In some embodiments the protein ligand is not MS2. Methods for generating RNA aptamers to target ligands have been described, for example, based on the systematic evolution of ligands by exponential enrichment (SELEX) (Tuerk and Gold, Science 249(4968):505-510 (1990); Ellington and Szostak, Nature 346(6287):818-822 (1990); the methods of each of which are incorporated herein by reference) and have, in some instances, been aided by in silico design (Bell et al. PNAS 117(15):8486-8493, the methods of which are incorporated herein by reference). Thus, in some embodiments, an aptamer for a target ligand is generated and incorporated into a synthetic ribozyme system, e.g., to trigger ribozyme-mediated cleavage and circRNA linearization, e.g., in the presence of the protein ligand. In some embodiments, circRNA linearization is triggered in the cytoplasm, e.g., using an aptamer that associates with a ligand in the cytoplasm. In some embodiments, circRNA linearization is triggered in the nucleus, e.g., using an aptamer that associates with a ligand in the nucleus. In embodiments, the ligand in to the nucleus comprises an epigenetic modifier or a transcription factor. In some embodiments the ligand that triggers linearization is present at higher levels in on-target cells than off-target cells.

It is further contemplated that a nucleic acid-responsive ribozyme system can be employed for circRNA linearization. For example, biosensors that sense defined target nucleic acid molecules to trigger ribozyme activation are described, e.g., in Penchovsky (Biotechnology Advances 32(5):1015-1027 (2014), incorporated herein by reference). By these methods, a ribozyme naturally folds into an inactive state and is only activated in the presence of a defined target nucleic acid molecule (e.g., an RNA molecule). In some embodiments, a circRNA of a Gene Writing system comprises a nucleic acid-responsive ribozyme that is activated in the presence of a defined target nucleic acid, e.g., an RNA, e.g., an mRNA, miRNA, guide RNA, gRNA, sgRNA, ncRNA, lncRNA, tRNA, snRNA, or mtRNA. In some embodiments the nucleic acid that triggers linearization is present at higher levels in on-target cells than off-target cells.

In some embodiments of any of the aspects herein, a Gene Writing system incorporates one or more ribozymes with inducible specificity to a target tissue or target cell of interest, e.g., a ribozyme that is activated by a ligand or nucleic acid present at higher levels in a target tissue or target cell of interest. In some embodiments, the Gene Writing system incorporates a ribozyme with inducible specificity to a subcellular compartment, e.g., the nucleus, nucleolus, cytoplasm,

or mitochondria. In some embodiments, the ribozyme that is activated by a ligand or nucleic acid present at higher levels in the target subcellular compartment. In some embodiments, an RNA component of a Gene Writing system is provided as circRNA, e.g., that is activated by linearization. In some embodiments, linearization of a circRNA encoding a Gene Writing polypeptide activates the molecule for translation. In some embodiments, a signal that activates a circRNA component of a Gene Writing system is present at higher levels in on-target cells or tissues, e.g., such that the system is specifically activated in these cells.

In some embodiments, an RNA component of a Gene Writing system is provided as a circRNA that is inactivated by linearization. In some embodiments, a circRNA encoding the Gene Writer polypeptide is inactivated by cleavage and degradation. In some embodiments, a circRNA encoding the Gene Writing polypeptide is inactivated by cleavage that separates a translation signal from the coding sequence of the polypeptide. In some embodiments, a signal that inactivates a circRNA component of a Gene Writing system is present at higher levels in off-target cells or tissues, such that the system is specifically inactivated in these cells.

In some embodiments, nucleic acid (e.g., encoding a Gene Writer polypeptide, or encoding a template RNA, or both) delivered to cells is covalently closed linear DNA, or so-called "doggybone" DNA. During its lifecycle, the bacteriophage N15 employs protelomerase to convert its genome from circular plasmid DNA to a linear plasmid DNA (Ravin et al. *J Mol Biol* 2001). This process has been adapted for the production of covalently closed linear DNA in vitro (see, for example, WO2010086626A1). In some embodiments, a protelomerase is contacted with a DNA containing one or more protelomerase recognition sites, wherein protelomerase results in a cut at the one or more sites and subsequent ligation of the complementary strands of DNA, resulting in the covalent linkage between the complementary strands. In some embodiments, nucleic acid (e.g., encoding a Gene Writer polypeptide, or encoding a template RNA, or both) is first generated as circular plasmid DNA containing a single protelomerase recognition site that is then contacted with protelomerase to yield a covalently closed linear DNA. In some embodiments, nucleic acid (e.g., encoding a transposase, or encoding a template RNA, or both) flanked by protelomerase recognition sites on plasmid or linear DNA is contacted with protelomerase to generate a covalently closed linear DNA containing only the DNA contained between the protelomerase recognition sites. In some embodiments, the approach of flanking the desired nucleic acid sequence by protelomerase recognition sites results in covalently closed

circular DNA lacking plasmid elements used for bacterial cloning and maintenance. In some embodiments, the plasmid or linear DNA containing the nucleic acid and one or more protelomerase recognition sites is optionally amplified prior to the protelomerase reaction, e.g., by rolling circle amplification or PCR.

In some embodiments, nucleic acid (e.g., encoding a Gene Writer polypeptide, or encoding a template RNA, or both) delivered to cells is closed-ended, linear duplex DNA (CELiD DNA or ceDNA). In some embodiments, ceDNA is derived from the replicative form of the AAV genome (Li et al. PLoS One 2013). In some embodiments, the nucleic acid (e.g., encoding a Gene Writer polypeptide, or encoding a template RNA, or both) is flanked by ITRs, e.g., AAV ITRs, wherein at least one of the ITRs comprises a terminal resolution site and a replication protein binding site (sometimes referred to as a replicative protein binding site). In some embodiments, the ITRs are derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, the ITRs are symmetric. In some embodiments, the ITRs are asymmetric. In some embodiments, at least one Rep protein is provided to enable replication of the construct. In some embodiments, the at least one Rep protein is derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, ceDNA is generated by providing a production cell with (i) DNA flanked by ITRs, e.g., AAV ITRs, and (ii) components required for ITR-dependent replication, e.g., AAV proteins Rep78 and Rep52 (or nucleic acid encoding the proteins). In some embodiments, ceDNA is free of any capsid protein, e.g., is not packaged into an infectious AAV particle. In some embodiments, ceDNA is formulated into LNPs (see, for example, WO2019051289A1).

In some embodiments, the ceDNA vector consists of two self-complementary sequences, e.g., asymmetrical or symmetrical or substantially symmetrical ITRs as defined herein, flanking said expression cassette, wherein the ceDNA vector is not associated with a capsid protein. In some embodiments, the ceDNA vector comprises two self-complementary sequences found in an AAV genome, where at least one ITR comprises an operative Rep-binding element (RBE) (also sometimes referred to herein as "RBS") and a terminal resolution site (trs) of AAV or a functional variant of the RBE. See, for example, WO2019113310.

In some embodiments, nucleic acid (e.g., encoding a Gene Writer polypeptide, or encoding a template RNA, or both) delivered to cells is designed as minicircles, where plasmid backbone sequences not pertaining to Gene Writing™ are removed before administration to cells. Minicircles have been shown to result in higher transfection efficiencies and gene expression as compared to plasmids with backbones containing bacterial parts (e.g., bacterial origin of replication, antibiotic selection cassette) and have been used to improve the efficiency of transposition (Sharma et al Mol Ther Nucleic Acids 2013). In some embodiments, the DNA vector encoding the Gene Writer™ polypeptide is delivered as a minicircle. In some embodiments, the DNA vector encoding the Gene Writer™ template is delivered as a minicircle. In some embodiments, the bacterial parts are flanked by recombination sites, e.g., attP/attB, loxP, FRT sites. In some embodiments, the addition of a cognate recombinase results in intramolecular recombination and excision of the bacterial parts. In some embodiments, the recombinase sites are recognized by phiC31 recombinase. In some embodiments, the recombinase sites are recognized by Cre recombinase. In some embodiments, the recombinase sites are recognized by FLP recombinase. In addition to plasmid DNA, minicircles can be generated by excising the desired construct, e.g., Gene Writer polypeptide expression cassette or template RNA expression cassette, from a viral backbone. Previously, it has been shown that excision and circularization of the donor sequence from a viral backbone may be important for transposase-mediated integration efficiency (Yant et al *Nat Biotechnol* 2002). In some embodiments, minicircles are first formulated and then delivered to target cells. In other embodiments, minicircles are formed from a DNA vector (e.g., plasmid DNA, rAAV, scAAV, ceDNA, doggybone DNA) intracellularly by co-delivery of a recombinase, resulting in excision and circularization of the recombinase recognition site-flanked nucleic acid, e.g., a nucleic acid encoding the Gene Writer™ polypeptide, or encoding an RNA template, or both.

Viral vectors and components thereof

Viruses are a useful source of delivery vehicles for the systems described herein, in addition to a source of relevant enzymes or domains as described herein, e.g., as sources of polymerases and polymerase functions used herein, e.g., DNA-dependent DNA polymerase, RNA-dependent RNA polymerase, RNA-dependent DNA polymerase, DNA-dependent RNA polymerase, reverse transcriptase. Some enzymes, e.g., reverse transcriptases, may have multiple

activities, e.g., be capable of both RNA-dependent DNA polymerization and DNA-dependent DNA polymerization, e.g., first and second strand synthesis. In some embodiments, the virus used as a Gene Writer delivery system or a source of components thereof may be selected from a group as described by Baltimore *Bacteriol Rev* 35(3):235-241 (1971).

In some embodiments, the virus is selected from a Group I virus, e.g., is a DNA virus and packages dsDNA into virions. In some embodiments, the Group I virus is selected from, e.g., Adenoviruses, Herpesviruses, Poxviruses.

In some embodiments, the virus is selected from a Group II virus, e.g., is a DNA virus and packages ssDNA into virions. In some embodiments, the Group II virus is selected from, e.g., Parvoviruses. In some embodiments, the parvovirus is a dependoparvovirus, e.g., an adeno-associated virus (AAV).

In some embodiments, the virus is selected from a Group III virus, e.g., is an RNA virus and packages dsRNA into virions. In some embodiments, the Group III virus is selected from, e.g., Reoviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group IV virus, e.g., is an RNA virus and packages ssRNA(+) into virions. In some embodiments, the Group IV virus is selected from, e.g., Coronaviruses, Picornaviruses, Togaviruses. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group V virus, e.g., is an RNA virus and packages ssRNA(-) into virions. In some embodiments, the Group V virus is selected from, e.g., Orthomyxoviruses, Rhabdoviruses. In some embodiments, an RNA virus with an ssRNA(-) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent RNA polymerase, capable of copying the ssRNA(-) into ssRNA(+) that can be translated directly by the host.

In some embodiments, the virus is selected from a Group VI virus, e.g., is a retrovirus and packages ssRNA(+) into virions. In some embodiments, the Group VI virus is selected from,

e.g., Retroviruses. In some embodiments, the retrovirus is a lentivirus, e.g., HIV-1, HIV-2, SIV, BIV. In some embodiments, the retrovirus is a spumavirus, e.g., a foamy virus, e.g., HFV, SFV, BFV. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, the ssRNA(+) is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with an ssRNA(+) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the ssRNA(+) into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VI retrovirus is incorporated as the reverse transcriptase domain of a Gene Writer polypeptide.

In some embodiments, the virus is selected from a Group VII virus, e.g., is a retrovirus and packages dsRNA into virions. In some embodiments, the Group VII virus is selected from, e.g., Hepadnaviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, one or both strands of the dsRNA contained in such virions is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with a dsRNA genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the dsRNA into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VII retrovirus is incorporated as the reverse transcriptase domain of a Gene Writer polypeptide.

In some embodiments, virions used to deliver nucleic acid in this invention may also carry enzymes involved in the process of Gene Writing. For example, a retroviral virion may contain a reverse transcriptase domain that is delivered into a host cell along with the nucleic acid. In some embodiments, an RNA template may be associated with a Gene Writer polypeptide within a virion, such that both are co-delivered to a target cell upon transduction of the nucleic

acid from the viral particle. In some embodiments, the nucleic acid in a virion may comprise DNA, e.g., linear ssDNA, linear dsDNA, circular ssDNA, circular dsDNA, minicircle DNA, dbDNA, ceDNA. In some embodiments, the nucleic acid in a virion may comprise RNA, e.g., linear ssRNA, linear dsRNA, circular ssRNA, circular dsRNA. In some embodiments, a viral genome may circularize upon transduction into a host cell, e.g., a linear ssRNA molecule may undergo a covalent linkage to form a circular ssRNA, a linear dsRNA molecule may undergo a covalent linkage to form a circular dsRNA or one or more circular ssRNA. In some embodiments, a viral genome may replicate by rolling circle replication in a host cell. In some embodiments, a viral genome may comprise a single nucleic acid molecule, e.g., comprise a non-segmented genome. In some embodiments, a viral genome may comprise two or more nucleic acid molecules, e.g., comprise a segmented genome. In some embodiments, a nucleic acid in a virion may be associated with one or proteins. In some embodiments, one or more proteins in a virion may be delivered to a host cell upon transduction. In some embodiments, a natural virus may be adapted for nucleic acid delivery by the addition of virion packaging signals to the target nucleic acid, wherein a host cell is used to package the target nucleic acid containing the packaging signals.

In some embodiments, a virion used as a delivery vehicle may comprise a commensal human virus. In some embodiments, a virion used as a delivery vehicle may comprise an anellovirus, the use of which is described in WO2018232017A1, which is incorporated herein by reference in its entirety.

Adeno-associated viruses

In some embodiments, the virus is an adeno-associated virus (AAV). In some embodiments, the AAV genome comprises two genes that encode four replication proteins and three capsid proteins, respectively. In some embodiments, the genes are flanked on either side by 145-bp inverted terminal repeats (ITRs). In some embodiments, the virion comprises up to three capsid proteins (Vp1, Vp2, and/or Vp3), e.g., produced in a 1:1:10 ratio. In some embodiments, the capsid proteins are produced from the same open reading frame and/or from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Generally, Vp3 is the most abundant subunit in the virion and participates in receptor recognition

at the cell surface defining the tropism of the virus. In some embodiments, Vp1 comprises a phospholipase domain, e.g., which functions in viral infectivity, in the N-terminus of Vp1.

In some embodiments, packaging capacity of the viral vectors limits the size of the base editor that can be packaged into the vector. For example, the packaging capacity of the AAVs can be about 4.5 kb (e.g., about 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 kb), e.g., including one or two inverted terminal repeats (ITRs), e.g., 145 base ITRs.

In some embodiments, recombinant AAV (rAAV) comprises cis-acting 145-bp ITRs flanking vector transgene cassettes, e.g., providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can, in some instances, express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. rAAV can be used, for example, in vitro and in vivo. In some embodiments, AAV-mediated gene delivery requires that the length of the coding sequence of the gene is equal or greater in size than the wild-type AAV genome.

AAV delivery of genes that exceed this size and/or the use of large physiological regulatory elements can be accomplished, for example, by dividing the protein(s) to be delivered into two or more fragments. In some embodiments, the N-terminal fragment is fused to a split intein-N. In some embodiments, the C- terminal fragment is fused to a split intein-C. In embodiments, the fragments are packaged into two or more AAV vectors.

In some embodiments, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5 and 3 ends, or head and tail), e.g., wherein each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette can, in some embodiments, then be achieved upon co-infection of the same cell by both dual AAV vectors. In some embodiments, co-infection is followed by one or more of: (1) homologous recombination (HR) between 5 and 3 genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5 and 3 genomes (dual AAV trans-splicing vectors); and/or (3) a combination of these two mechanisms (dual AAV hybrid vectors). In some embodiments, the use of dual AAV vectors in vivo results in the expression of full-length proteins. In some embodiments, the use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of greater than about 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, or 5.0 kb in size.In some embodiments, AAV vectors can also be used to transduce cells with target nucleic acids, e.g., in the in vitro

production of nucleic acids and peptides. In some embodiments, AAV vectors can be used for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest.94:1351 (1994); each of which is incorporated herein by reference in their entirety). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No.5,173,414; 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); and Samulski et al., J. Virol.63:03822-3828 (1989) (incorporated by reference herein in their entirety).

In some embodiments, a Gene Writer described herein (e.g., with or without one or more guide nucleic acids) can be delivered using AAV, lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No.8,404,658 (formulations, doses for AAV) and U.S. Patent No.5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as described in U.S. Patent No.8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as described in U.S. Patent No.8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as described in U.S. Patent No.5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. In some embodiments, the viral vectors can be injected into the tissue of interest. For cell-type specific Gene Writing, the expression of the Gene Writer and optional guide nucleic acid can, in some embodiments, be driven by a cell-type specific promoter.

In some embodiments, AAV allows for low toxicity, for example, due to the purification method not requiring ultracentrifugation of cell particles that can activate the immune response.

In some embodiments, AAV allows low probability of causing insertional mutagenesis, for example, because it does not substantially integrate into the host genome.

In some embodiments, AAV has a packaging limit of about 4.4, 4.5, 4.6, 4.7, or 4.75 kb. In some embodiments, a Gene Writer, promoter, and transcription terminator can fit into a single viral vector. SpCas9 (4.1 kb) may, in some instances, be difficult to package into AAV. Therefore, in some embodiments, a Gene Writer is used that is shorter in length than other Gene Writers or base editors. In some embodiments, the Gene Writers are less than about 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb.

An AAV can be AAV1, AAV2, AAV5 or any combination thereof. In some embodiments, the type of AAV is selected with respect to the cells to be targeted; e.g., AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof can be selected for targeting brain or neuronal cells; or AAV4 can be selected for targeting cardiac tissue. In some embodiments, AAV8 is selected for delivery to the liver. Exemplary AAV serotypes as to these cells are described, for example, in Grimm, D. et al, J. Virol.82: 5887-5911 (2008) (incorporated herein by reference in its entirety). In some embodiments, AAV refers all serotypes, subtypes, and naturally-occurring AAV as well as recombinant AAV. AAV may be used to refer to the virus itself or a derivative thereof. In some embodiments, AAV includes AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64Rl, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrhlO, AAVLK03, AV10, AAV11, AAV 12, rhlO, and hybrids thereof, avian AAV, bovine AAV, canine AAV, equine AAV, primate AAV, nonprimate AAV, and ovine AAV. The genomic sequences of various serotypes of AAV, as well as the sequences of the native terminal repeats (TRs), Rep proteins, and capsid subunits are known in the art. Such sequences may be found in the literature or in public databases such as GenBank. Additional exemplary AAV serotypes are listed in Table 36.

Table 36. Exemplary AAV serotypes.

Target Tissue	Vehicle	Reference
Liver	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2,	1. Wang *et al.*, *Mol. Ther.* **18**, 118–25 (2010)

	NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4] Adenovirus (Ad5, HC-AdV[6])	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)
Lung	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) Adenovirus (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)
Skin	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010) 2. L. Lisowski *et al.*, *Nature.* **506**, 382–6 (2014).
HSCs	Adenovirus (HDAd5/35^{++})	Wang et al. *Blood Adv* (2019)

In some embodiments, a pharmaceutical composition (e.g., comprising an AAV as dscribed herein) has less than 10% empty capsids, less than 8% empty capsids, less than 7% empty capsids, less than 5% empty capsids, less than 3% empty capsids, or less than 1 % empty capsids. In some embodiments, the pharmaceutical composition has less than about 5% empty capsids. In some embodiments, the number of empty capsids is below the limit of detection. In some embodiments, it is advantageous for the pharmaceutical composition to have low amounts of empty capsids, e.g., because empty capsids may generate an adverse response (e.g., immune response, inflammatory response, liver response, and/or cardiac response), e.g., with little or no substantial therapeutic benefit.

In some embodiments, the residual host cell protein (rHCP) in the pharmaceutical composition is less than or equal to 100 ng/ml rHCP per 1 x 10^{13} vg/ml, e.g., less than or equal to 40 ng/ml rHCP per 1 x 10^{13} vg/ml or 1-50 ng/ml rHCP per 1 x 10^{13} vg/ml. In some embodiments, the pharmaceutical composition comprises less than 10 ng rHCP per 1.0 x 10^{13} vg, or less than 5 ng rHCP per 1.0 x 10^{13} vg, less than 4 ng rHCP per 1.0 x 10^{13} vg, or less than 3 ng rHCP per 1.0 x 10^{13} vg, or any concentration in between. In some embodiments, the residual host cell DNA (hcDNA) in the pharmaceutical composition is less than or equal to 5 x 10^{6} pg/ml

hcDNA per 1 x 10^{13} vg/ml, less than or equal to 1.2 x 10^{6} pg/ml hcDNA per 1 x 10^{13} vg/ml, or 1 x 10^{5} pg/ml hcDNA per 1 x 10^{13} vg/ml. In some embodiments, the residual host cell DNA in said pharmaceutical composition is less than 5.0 x 10^{5} pg per 1 x 10^{13} vg, less than 2.0 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.1 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.0 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, less than 0.9 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, less than 0.8 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, or any concentration in between.

In some embodiments, the residual plasmid DNA in the pharmaceutical composition is less than or equal to 1.7 x 10^{5} pg/ml per 1.0 x 10^{13} vg/ml, or 1 x 10^{5} pg/ml per 1 x 1.0 x 10^{13} vg/ml, or 1.7 x 10^{6} pg/ml per 1.0 x 10^{13} vg/ml. In some embodiments, the residual DNA plasmid in the pharmaceutical composition is less than 10.0 x 10^{5} pg by 1.0 x 10^{13} vg, less than 8.0 x 10^{5} pg by 1.0 x 10^{13} vg or less than 6.8 x 10^{5} pg by 1.0 x 10^{13} vg. In embodiments, the pharmaceutical composition comprises less than 0.5 ng per 1.0 x 10^{13} vg, less than 0.3 ng per 1.0 x 10^{13} vg, less than 0.22 ng per 1.0 x 10^{13} vg or less than 0.2 ng per 1.0 x 10^{13} vg or any intermediate concentration of bovine serum albumin (BSA). In embodiments, the benzonase in the pharmaceutical composition is less than 0.2 ng by 1.0 x 10^{13} vg, less than 0.1 ng by 1.0 x 10^{13} vg, less than 0.09 ng by 1.0 x 10^{13} vg, less than 0.08 ng by 1.0 x 10^{13} vg or any intermediate concentration. In embodiments, Poloxamer 188 in the pharmaceutical composition is about 10 to 150 ppm, about 15 to 100 ppm or about 20 to 80 ppm. In embodiments, the cesium in the pharmaceutical composition is less than 50 pg / g (ppm), less than 30 pg / g (ppm) or less than 20 pg / g (ppm) or any intermediate concentration.

In embodiments, the pharmaceutical composition comprises total impurities, e.g., as determined by SDS-PAGE, of less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or any percentage in between. In embodiments, the total purity, e.g., as determined by SDS-PAGE, is greater than 90%, greater than 92%, greater than 93%, greater than 94%, greater than 95%, greater than 96%, greater than 97%, greater than 98%, or any percentage in between. In embodiments, no single unnamed related impurity, e.g., as measured by SDS-PAGE, is greater than 5%, greater than 4%, greater than 3% or greater than 2%, or any percentage in between. In embodiments, the pharmaceutical composition comprises a percentage of filled capsids relative to total capsids (e.g., peak 1 + peak 2 as measured by analytical ultracentrifugation) of greater than 85%, greater than 86%, greater than 87%, greater than 88%, greater than 89%, greater than 90%, greater than 91%, greater than 91.9%, greater

than 92%, greater than 93%, or any percentage in between. In embodiments of the pharmaceutical composition, the percentage of filled capsids measured in peak 1 by analytical ultracentrifugation is 20-80%, 25-75%, 30-75%, 35-75%, or 37.4-70.3%. In embodiments of the pharmaceutical composition, the percentage of filled capsids measured in peak 2 by analytical ultracentrifugation is 20-80%, 20-70%, 22-65%, 24-62%, or 24.9-60.1%.

In one embodiment, the pharmaceutical composition comprises a genomic titer of 1.0 to 5.0 x 10^{13} vg / mL, 1.2 to 3.0 x 10^{13} vg / mL or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition exhibits a biological load of less than 5 CFU / mL, less than 4 CFU / mL, less than 3 CFU / mL, less than 2 CFU / mL or less than 1 CFU / mL or any intermediate contraction. In embodiments, the amount of endotoxin according to USP, for example, USP <85> (incorporated by reference in its entirety) is less than 1.0 EU / mL, less than 0.8 EU / mL or less than 0.75 EU / mL. In embodiments, the osmolarity of a pharmaceutical composition according to USP, for example, USP <785> (incorporated by reference in its entirety) is 350 to 450 mOsm / kg, 370 to 440 mOsm / kg or 390 to 430 mOsm / kg. In embodiments, the pharmaceutical composition contains less than 1200 particles that are greater than 25 μm per container, less than 1000 particles that are greater than 25 μm per container, less than 500 particles that are greater than 25 μm per container or any intermediate value. In embodiments, the pharmaceutical composition contains less than 10,000 particles that are greater than 10 μm per container, less than 8000 particles that are greater than 10 μm per container or less than 600 particles that are greater than 10 pm per container.

In one embodiment, the pharmaceutical composition has a genomic titer of 0.5 to 5.0 x 10^{13} vg / mL, 1.0 to 4.0 x 10^{13} vg / mL, 1.5 to 3.0 x 10^{13} vg / ml or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition described herein comprises one or more of the following: less than about 0.09 ng benzonase per 1.0 x 10^{13} vg, less than about 30 pg / g (ppm) of cesium, about 20 to 80 ppm Poloxamer 188, less than about 0.22 ng BSA per 1.0 x 10^{13} vg, less than about 6.8 x 10^{5} pg of residual DNA plasmid per 1.0 x 10^{13} vg, less than about 1.1 x 10^{5} pg of residual hcDNA per 1.0 x 10^{13} vg, less than about 4 ng of rHCP per 1.0 x 10^{13} vg, pH 7.7 to 8.3, about 390 to 430 mOsm / kg, less than about 600 particles that are > 25 μm in size per container, less than about 6000 particles that are > 10 μm in size per container, about 1.7 x 10^{13} - 2.3 x 10^{13} vg / mL genomic titer, infectious titer of about 3.9 x 10^{8} to 8.4 x 10^{10} IU per 1.0 x 10^{13} vg, total protein of about 100-300 pg per 1.0 x 10^{13} vg, mean survival of >24 days in

A7SMA mice with about 7.5 x 10^{13} vg / kg dose of viral vector, about 70 to 130% relative potency based on an in vitro cell based assay and / or less than about 5% empty capsid. In various embodiments, the pharmaceutical compositions described herein comprise any of the viral particles discussed here, retain a potency of between ± 20%, between ± 15%, between ± 10% or within ± 5% of a reference standard. In some embodiments, potency is measured using a suitable in vitro cell assay or in vivo animal model.

Additional methods of preparation, characterization, and dosing AAV particles are taught in WO2019094253, which is incorporated herein by reference in its entirety.

Additional rAAV constructs that can be employed consonant with the invention include those described in Wang et al 2019, available at: //doi.org/10.1038/s41573-019-0012-9, including Table 1 thereof, which is incorporated by reference in its entirety.

Inteins

In some embodiments, as described in more detail below, Intein-N may be fused to the N-terminal portion of a first domain described herein, and and intein-C may be fused to the C-terminal portion of a second domain described herein for the joining of the N-terminal portion to the C-terminal portion, thereby joining the first and second domains. In some embodiments, the first and second domains are each independent chosen from a DNA binding domain, an RNA binding domain, an RT domain, and an endonuclease domain.

As used herein, "intein" refers to a self-splicing protein intron (e.g., peptide), e.g., which ligates flanking N-terminal and C-terminal exteins (e.g., fragments to be joined). An intein may, in some instances, comprise a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing." In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein (e.g., split intein-N and split intein-C). For example, in cyanobacteria, DnaE, the catalytic subunit a of DNA polymerase III, is encoded by two separate genes, dnaE-n and dnaE-c. The intein encoded by the dnaE-n gene may be herein referred as "intein-N." The intein encoded by the dnaE-c gene may be herein referred as "intein-C."

Use of inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (incorporated herein by reference in its entirety). For example, when fused to separate protein fragments, the inteins IntN and IntC may recognize each other, splice themselves out, and/or simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments.

In some embodiments, a synthetic intein based on the dnaE intein, the Cfa-N (e.g., split intein-N) and Cfa-C (e.g., split intein-C) intein pair, is used. Examples of such inteins have been described, e.g., in Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5 (incorporated herein by reference in its entirety). Non-limiting examples of intein pairs that may be used in accordance with the present disclosure include: Cfa DnaE intein, Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (e.g., as described in U.S. Pat. No. 8,394,604, incorporated herein by reference.

In some embodiments, Intein-N and intein-C may be fused to the N-terminal portion of the split Cas9 and the C-terminal portion of a split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split Cas9, i.e., to form a structure of N— [N-terminal portion of the split Cas9]-[intein-N]~ C. In some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, i.e., to form a structure of N-[intein-C]~ [C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (e.g., split Cas9) is described in Shah et al., Chem Sci. 2014; 5(l):446-46l, incorporated herein by reference. Methods for designing and using inteins are known in the art and described, for example by WO2020051561, W02014004336, WO2017132580, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

In some embodiments, a split refers to a division into two or more fragments. In some embodiments, a split Cas9 protein or split Cas9 comprises a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a reconstituted Cas9 protein. In embodiments, the Cas9 protein is divided into two fragments within a disordered region of the protein, e.g., as described in

Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014, or as described in Jiang et al. (2016) Science 351: 867-871 and PDB file: 5F9R (each of which is incorporated herein by reference in its entirety). A disordered region may be determined by one or more protein structure determination techniques known in the art, including, without limitation, X-ray crystallography, NMR spectroscopy, electron microscopy (e.g., cryoEM), and/or in silico protein modeling. In some embodiments, the protein is divided into two fragments at any C, T, A, or S, e.g., within a region of SpCas9 between amino acids A292- G364, F445-K483, or E565-T637, or at corresponding positions in any other Cas9, Cas9 variant (e.g., nCas9, dCas9), or other napDNAbp. In some embodiments, protein is divided into two fragments at SpCas9 T310, T313, A456, S469, or C574. In some embodiments, the process of dividing the protein into two fragments is referred to as splitting the protein.

In some embodiments, a protein fragment ranges from about 2-1000 amino acids (e.g., between 2-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 amino acids) in length. In some embodiments, a protein fragment ranges from about 5-500 amino acids (e.g., between 5-10, 10-50, 50-100, 100-200, 200-300, 300-400, or 400-500 amino acids) in length. In some embodiments, a protein fragment ranges from about 20-200 amino acids (e.g., between 20-30, 30-40, 40-50, 50-100, or 100-200 amino acids) in length.

In some embodiments, a portion or fragment of a Gene Writer (e.g., Cas9-R2Tg) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In some embodiments, an endonuclease domain (e.g., a nickase Cas9 domain) is fused to intein-N and a polypeptide comprising an RT domain is fused to an intein-C.

Exemplary nucleotide and amino acid sequences of interns are provided below:

DnaE Intein-N DNA:

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGG
AAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAA
CATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCG
AATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATG

ACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTC ATGCGAGTTGACAACCTTCCTAAT (SEQ ID NO: 1612)

DnaE Intein-N Protein:
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1613)

DnaE Intein-C DNA:
ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGG AGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT (SEQ ID NO: 1614)

Intein-C:
MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1615)

Cfa-N DNA:
TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAA AGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTC GTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGA GTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGA CCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTC AAACAAGTGGATGGATTG CCA (SEQ ID NO: 1616)

Cfa-N Protein:
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCL EDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 1617)

Cfa-C DNA:
ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGT AAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGA GAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ ID NO: 1618)

Cfa-C Protein:
MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 1619)

Lipid Nanoparticles

The methods and systems provided by the invention, may employ any suitable carrier or delivery modality, including, in certain embodiments, lipid nanoparticles (LNPs). Lipid nanoparticles, in some embodiments, comprise one or more ionic lipids, such as non-cationic lipids (e.g., neutral or anionic, or zwitterionic lipids); one or more conjugated lipids (such as

PEG-conjugated lipids or lipids conjugated to polymers described in Table 5 of WO2019217941; incorporated herein by reference in its entirety); one or more sterols (e.g., cholesterol); and, optionally, one or more targeting molecules (e.g., conjugated receptors, receptor ligands, antibodies); or combinations of the foregoing.

Lipids that can be used in nanoparticle formations (e.g., lipid nanoparticles) include, for example those described in Table 4 of WO2019217941, which is incorporated by reference—e.g., a lipid-containing nanoparticle can comprise one or more of the lipids in table 4 of WO2019217941. Lipid nanoparticles can include additional elements, such as polymers, such as the polymers described in table 5 of WO2019217941, incorporated by reference.

In some embodiments, conjugated lipids, when present, can include one or more of PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3- dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG- ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-l-0-(w- methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N- (carbonyl-methoxypoly ethylene glycol 2000)-1 ,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, and those described in Table 2 of WO2019051289 (incorporated by reference), and combinations of the foregoing.

In some embodiments, sterols that can be incorporated into lipid nanoparticles include one or more of cholesterol or cholesterol derivatives, such as those in W02009/127060 or US2010/0130588, which are incorporated by reference. Additional exemplary sterols include phytosterols, including those described in Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference.

In some embodiments, the lipid particle comprises an ionizable lipid, a non-cationic lipid, a conjugated lipid that inhibits aggregation of particles, and a sterol. The amounts of these components can be varied independently and to achieve desired properties. For example, in some embodiments, the lipid nanoparticle comprises an ionizable lipid is in an amount from about 20 mol % to about 90 mol % of the total lipids (in other embodiments it may be 20-70% (mol), 30-60% (mol) or 40-50% (mol); about 50 mol % to about 90 mol % of the total lipid present in the lipid nanoparticle), a non-cationic lipid in an amount from about 5 mol % to about 30 mol % of the total lipids, a conjugated lipid in an amount from about 0.5 mol % to about 20 mol % of the total lipids, and a sterol in an amount from about 20 mol % to about 50 mol % of

the total lipids. The ratio of total lipid to nucleic acid (e.g., encoding the Gene Writer or template nucleic acid) can be varied as desired. For example, the total lipid to nucleic acid (mass or weight) ratio can be from about 10: 1 to about 30: 1.

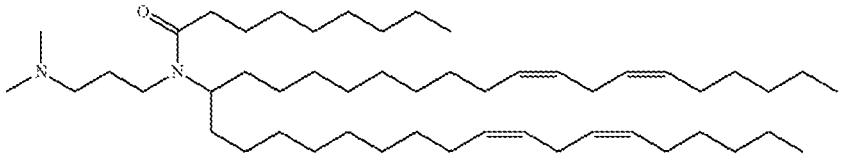
In some embodiments, an ionizable lipid may be a cationic lipid, a ionizable cationic lipid, e.g., a cationic lipid that can exist in a positively charged or neutral form depending on pH, or an amine-containing lipid that can be readily protonated. In some embodiments, the cationic lipid is a lipid capable of being positively charged, e.g., under physiological conditions. Exemplary cationic lipids include one or more amine group(s) which bear the positive charge. In some embodiments, the lipid particle comprises a cationic lipid in formulation with one or more of neutral lipids, ionizable amine-containing lipids, biodegradable alkyn lipids, steroids, phospholipids including polyunsaturated lipids, structural lipids (e.g., sterols), PEG, cholesterol and polymer conjugated lipids. In some embodiments, the cationic lipid may be an ionizable cationic lipid. An exemplary cationic lipid as disclosed herein may have an effective pKa over 6.0. In embodiments, a lipid nanoparticle may comprise a second cationic lipid having a different effective pKa (e.g., greater than the first effective pKa), than the first cationic lipid. A lipid nanoparticle may comprise between 40 and 60 mol percent of a cationic lipid, a neutral lipid, a steroid, a polymer conjugated lipid, and a therapeutic agent, e.g., a nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter), encapsulated within or associated with the lipid nanoparticle. In some embodiments, the nucleic acid is co-formulated with the cationic lipid. The nucleic acid may be adsorbed to the surface of an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the nucleic acid may be encapsulated in an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the lipid nanoparticle may comprise a targeting moiety, e.g., coated with a targeting agent. In embodiments, the LNP formulation is biodegradable. In some embodiments, a lipid nanoparticle comprising one or more lipid described herein, e.g., Formula (i), (ii), (ii), (vii) and/or (ix) encapsulates at least 1%, at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98% or 100% of an RNA molecule, e.g., template RNA and/or a mRNA encoding the Gene Writer polypeptide.

In some embodiments, the lipid to nucleic acid ratio (mass/mass ratio; w/w ratio) can be in the range of from about 1 : 1 to about 25: 1, from about 10: 1 to about 14: 1, from about 3 : 1 to about 15: 1, from about 4: 1 to about 10: 1, from about 5: 1 to about 9: 1, or about 6: 1 to about 9: 1. The amounts of lipids and nucleic acid can be adjusted to provide a desired N/P ratio, for example, N/P ratio of 3, 4, 5, 6, 7, 8, 9, 10 or higher. Generally, the lipid nanoparticle formulation's overall lipid content can range from about 5 mg/ml to about 30 mg/mL.

Exemplary ionizable lipids that can be used in lipid nanoparticle formulations include, without limitation, those listed in Table 1 of WO2019051289, incorporated herein by reference. Additional exemplary lipids include, without limitation, one or more of the following formulae: X of US2016/0311759; I of US20150376115 or in US2016/0376224; I, II or III of US20160151284; I, IA, II, or IIA of US20170210967; I-c of US20150140070; A of US2013/0178541; I of US2013/0303587 or US2013/0123338; I of US2015/0141678; II, III, IV, or V of US2015/0239926; I of US2017/0119904; I or II of WO2017/117528; A of US2012/0149894; A of US2015/0057373; A of WO2013/116126; A of US2013/0090372; A of US2013/0274523; A of US2013/0274504; A of US2013/0053572; A of W02013/016058; A of W02012/162210; I of US2008/042973; I, II, III, or IV of US2012/01287670; I or II of US2014/0200257; I, II, or III of US2015/0203446; I or III of US2015/0005363; I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID, or III-XXIV of US2014/0308304; of US2013/0338210; I, II, III, or IV of W02009/132131; A of US2012/01011478; I or XXXV of US2012/0027796; XIV or XVII of US2012/0058144; of US2013/0323269; I of US2011/0117125; I, II, or III of US2011/0256175; I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII of US2012/0202871; I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV, or XVI of US2011/0076335; I or II of US2006/008378; I of US2013/0123338; I or X-A-Y-Z of US2015/0064242; XVI, XVII, or XVIII of US2013/0022649; I, II, or III of US2013/0116307; I, II, or III of US2013/0116307; I or II of US2010/0062967; I-X of US2013/0189351; I of US2014/0039032; V of US2018/0028664; I of US2016/0317458; I of US2013/0195920; 5, 6, or 10 of US10,221,127; III-3 of WO2018/081480; I-5 or I-8 of WO2020/081938; 18 or 25 of US9,867,888; A of US2019/0136231; II of WO2020/219876; 1 of US2012/0027803; OF-02 of US2019/0240349; 23 of US10,086,013; cKK-E12/A6 of Miao et al (2020); C12-200 of WO2010/053572; 7C1 of Dahlman et al (2017); 304-O13 or 503-O13 of Whitehead et al; TS-P4C2 of US9,708,628; I of WO2020/106946; I of WO2020/106946.

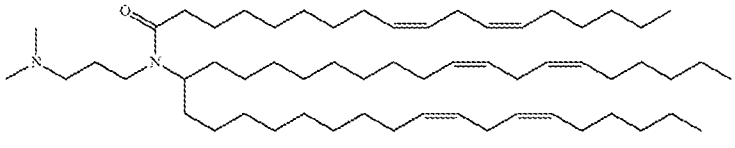
In some embodiments, the ionizable lipid is MC3 (6Z,9Z,28Z,3 lZ)-heptatriaconta-6,9,28,3 l-tetraen-l9-yl-4-(dimethylamino) butanoate (DLin-MC3-DMA or MC3), e.g., as described in Example 9 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is the lipid ATX-002, e.g., as described in Example 10 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is (l3Z,l6Z)-A,A-dimethyl-3- nonyldocosa-l3, l6-dien-l-amine (Compound 32), e.g., as described in Example 11 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Compound 6 or Compound 22, e.g., as described in Example 12 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is heptadecan-9-yl 8-((2-hydroxyethyl)(6-oxo-6-(undecyloxy)hexyl)amino)octanoate (SM-102); e.g., as described in Example 1 of US9,867,888(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,12-dienoate (LP01) e.g., as synthesized in Example 13 of WO2015/095340(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Di((Z)-non-2-en-1-yl) 9-((4-dimethylamino)butanoyl)oxy)heptadecanedioate (L319), , e.g. as synthesized in Example 7, 8, or 9 of US2012/0027803(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 1,1'-((2-(4-(2-((2-(Bis(2-hydroxydodecyl)amino)ethyl)(2-hydroxydodecyl) amino)ethyl)piperazin-1-yl)ethyl)azanediyl)bis(dodecan-2-ol) (C12-200), e.g., as synthesized in Examples 14 and 16 of WO2010/053572(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is; Imidazole cholesterol ester (ICE) lipid (3S, 10R, 13R, 17R)-10, 13-dimethyl-17- ((R)-6-methylheptan-2-yl)-2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17-tetradecahydro-lH- cyclopenta[a]phenanthren-3-yl 3-(1H-imidazol-4-yl)propanoate, e.g., Structure (I) from WO2020/106946 (incorporated by reference herein in its entirety).

Some non-limiting example of lipid compounds that may be used (e.g., in combination with other lipid components) to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) includes,



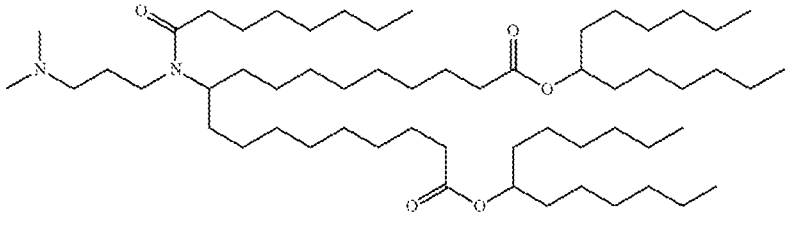
(i)

In some embodiments an LNP comprising Formula (i) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



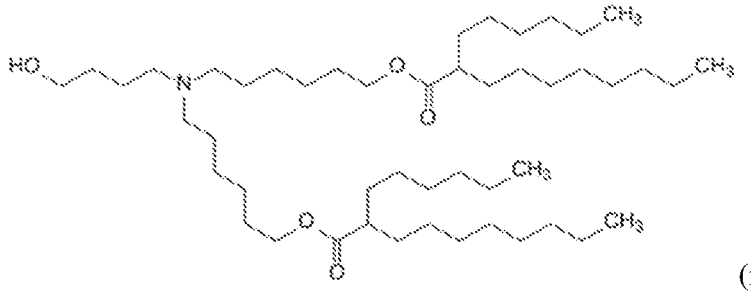
(ii)

In some embodiments an LNP comprising Formula (ii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

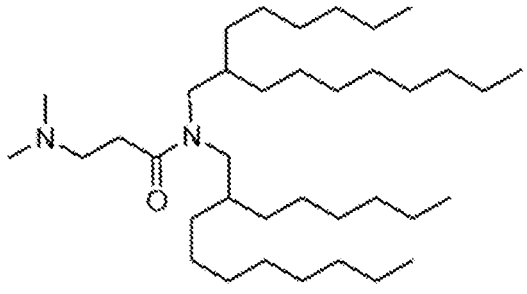


(iii)

In some embodiments an LNP comprising Formula (iii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

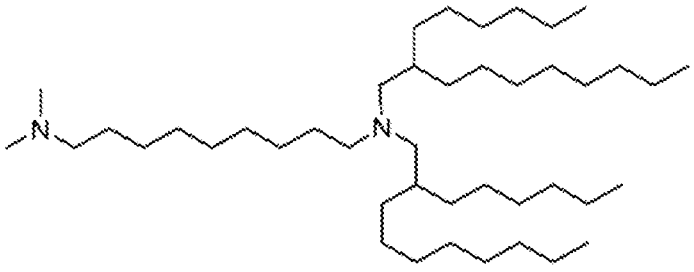


(iv)



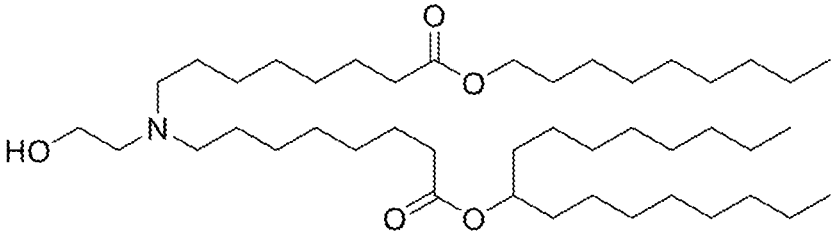
(v)

In some embodiments an LNP comprising Formula (v) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

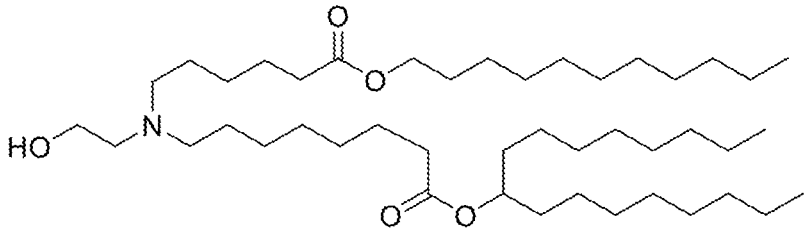


(vi)

In some embodiments an LNP comprising Formula (vi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

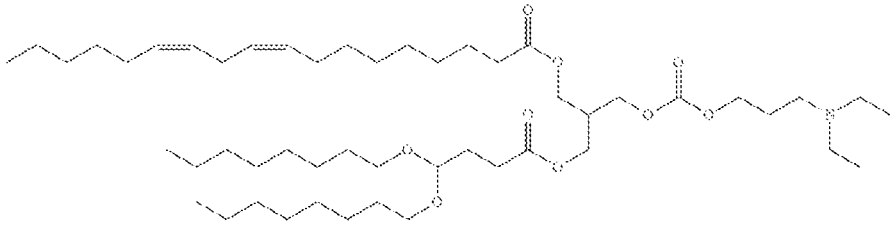


(vii)



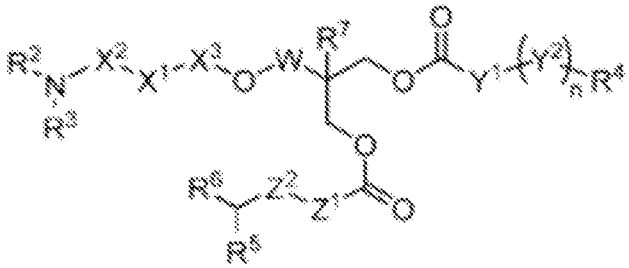
(viii)

In some embodiments an LNP comprising Formula (viii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



(ix)

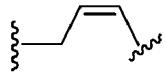
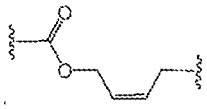
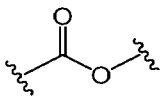
In some embodiments an LNP comprising Formula (ix) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



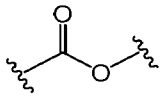
(x)

wherein

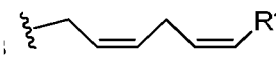
X^1 is O, NR^1, or a direct bond, X^2 is C2-5 alkylene, X^3 is C(=0) or a direct bond, R^1 is H or Me, R^3 is Ci-3 alkyl, R^2 is Ci-3 alkyl, or R^2 taken together with the nitrogen atom to which it is attached and 1-3 carbon atoms of X^2 form a 4-, 5-, or 6-membered ring, or X^1 is NR^1, R^1 and R^2 taken together with the nitrogen atoms to which they are attached form a 5- or 6-membered ring, or R^2 taken together with R^3 and the nitrogen atom to which they are attached form a 5-, 6-, or 7-membered ring, Y^1 is C2-12 alkylene, Y^2 is selected from

 (in either orientation),  (in either orientation),  (in either orientation),

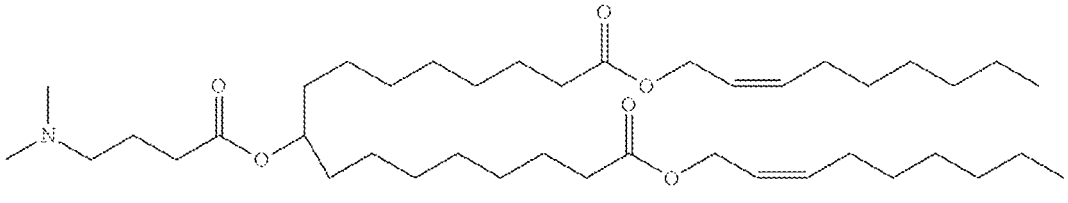
n is 0 to 3, R^4 is Ci-15 alkyl, Z^1 is Ci-6 alkylene or a direct bond,

Z^2 is 
(in either orientation) or absent, provided that if Z^1 is a direct bond, Z^2 is absent;

R^5 is C5-9 alkyl or C6-10 alkoxy, R^6 is C5-9 alkyl or C6-10 alkoxy, W is methylene or a direct bond, and R^7 is H or Me, or a salt thereof, provided that if R^3 and R^2 are C2 alkyls, X^1 is O, X^2 is linear C3 alkylene, X^3 is C(=0), Y^1 is linear Ce alkylene, $(Y^2)_n$-R^4 is

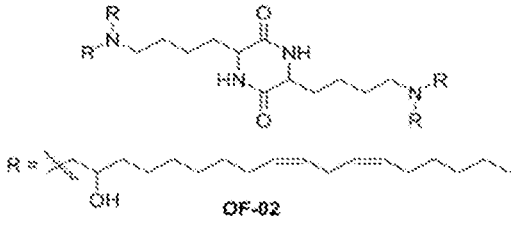

, R^4 is linear C5 alkyl, Z^1 is C2 alkylene, Z^2 is absent, W is methylene, and R^7 is H, then R^5 and R^6 are not Cx alkoxy.

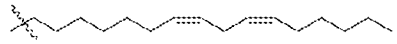
In some embodiments an LNP comprising Formula (xii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

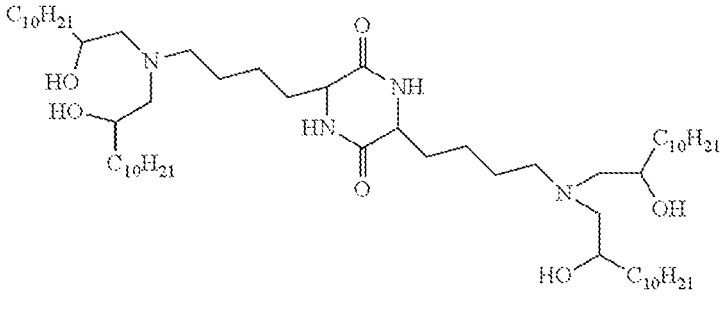


(xi)

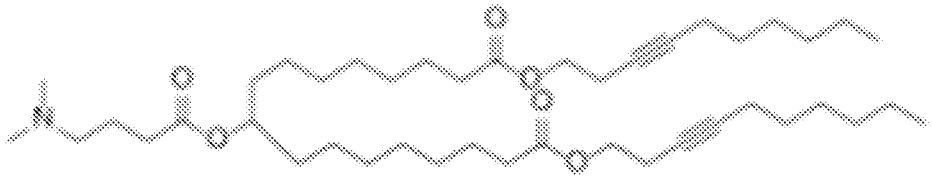
In some embodiments an LNP comprising Formula (xi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



where R=  (xii)

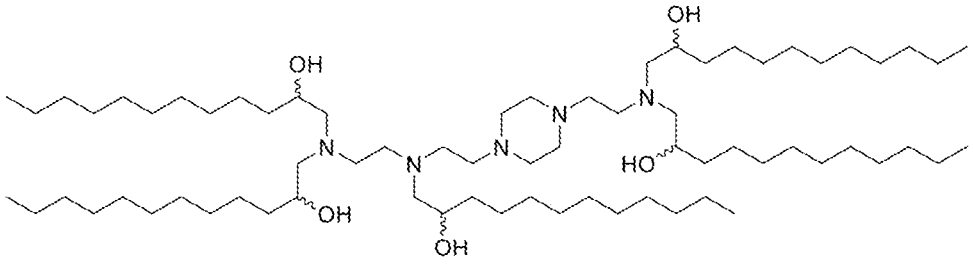


(xiii)



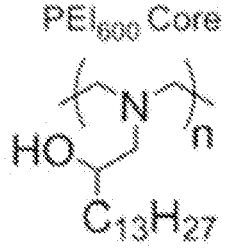
(xiv)

In some embodiments an LNP comprises a compound of Formula (xiii) and a compound of Formula (xiv).



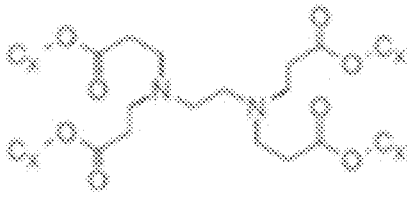
(xv)

In some embodiments an LNP comprising Formula (xv) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

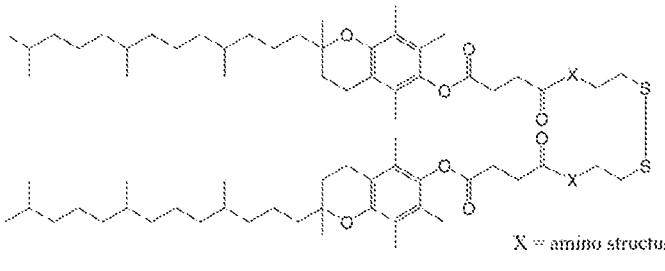


(xvi)

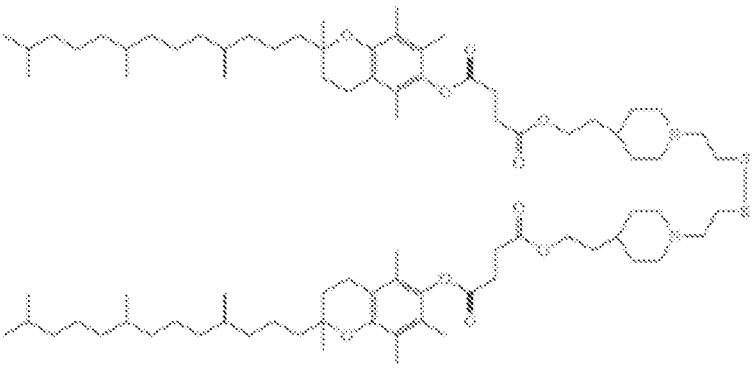
In some embodiments an LNP comprising a formulation of Formula (xvi) is used to deliver a GeneWriter composition described herein to the lung endothelial cells.



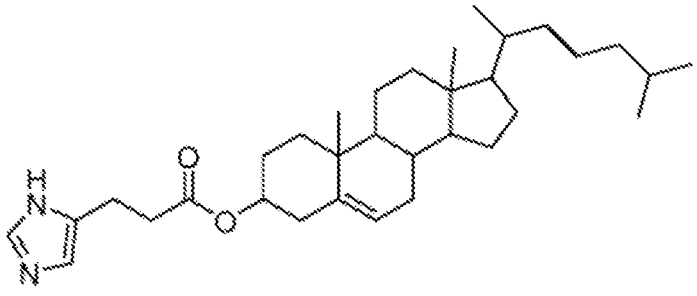
(xvii)



(xviii) (a)

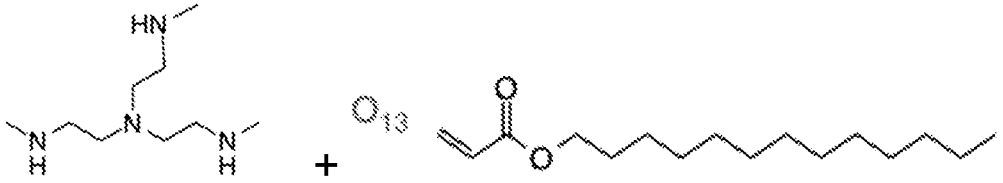


(xviii)(b)

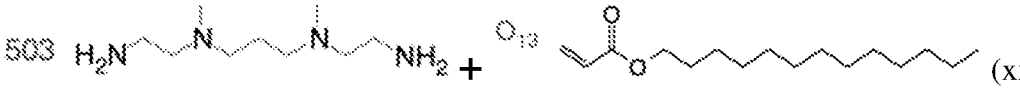


(xix)

In some embodiments, a lipid compound used to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) is made by one of the following reactions:



(xx) (a)

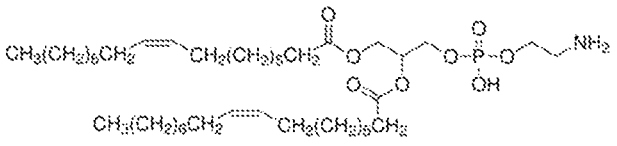


(xx)(b)

Exemplary non-cationic lipids include, but are not limited to, distearoyl-sn-glycero-phosphoethanolamine, distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoyl-phosphatidylethanolamine (DOPE), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE),palmitoyloleoylphosphatidylcholine (POPC), palmitoyloleoylphosphatidylethanolamine (POPE), dioleoyl-phosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane- 1 - carboxylate (DOPE-mal), dipalmitoyl phosphatidyl ethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidyl-ethanolamine (DSPE), monomethyl-phosphatidylethanolamine (such as 16-O-monomethyl PE), dimethyl- phosphatidylethanolamine (such as 16-O-dimethyl PE), l8-l-trans PE, l-stearoyl-2-oleoyl- phosphatidyethanolamine (SOPE), hydrogenated soy phosphatidylcholine (HSPC), egg phosphatidylcholine (EPC), dioleoylphosphatidylserine (DOPS), sphingomyelin (SM), dimyristoyl phosphatidylcholine (DMPC), dimyristoyl phosphatidylglycerol (DMPG), distearoylphosphatidylglycerol (DSPG), dierucoylphosphatidylcholine (DEPC), palmitoyloleyolphosphatidylglycerol (POPG), dielaidoyl-phosphatidylethanolamine (DEPE), lecithin, phosphatidylethanolamine, lysolecithin, lysophosphatidylethanolamine,

phosphatidylserine, phosphatidylinositol, sphingomyelin, egg sphingomyelin (ESM), cephalin, cardiolipin, phosphatidicacid,cerebrosides, dicetylphosphate, lysophosphatidylcholine, dilinoleoylphosphatidylcholine, or mixtures thereof. It is understood that other diacylphosphatidylcholine and diacylphosphatidylethanolamine phospholipids can also be used. The acyl groups in these lipids are preferably acyl groups derived from fatty acids having C10-C24 carbon chains, e.g., lauroyl, myristoyl, paimitoyl, stearoyl, or oleoyl. Additional exemplary lipids, in certain embodiments, include, without limitation, those described in Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference. Such lipids include, in some embodiments, plant lipids found to improve liver transfection with mRNA (e.g., DGTS).

In some embodiments, the non-cationic lipid may have the following structure



(xxi)

Other examples of non-cationic lipids suitable for use in the lipid nanopartieles include, without limitation, nonphosphorous lipids such as, e.g., stearylamine, dodeeylamine, hexadecylamine, acetyl palmitate, glycerol ricinoleate, hexadecyl stereate, isopropyl myristate, amphoteric acrylic polymers, triethanolamine-lauryl sulfate, alkyl-aryl sulfate polyethyloxylated fatty acid amides, dioctadecyl dimethyl ammonium bromide, ceramide, sphingomyelin, and the like. Other non-cationic lipids are described in WO2017/099823 or US patent publication US2018/0028664, the contents of which is incorporated herein by reference in their entirety.

In some embodiments, the non-cationic lipid is oleic acid or a compound of Formula I, II, or IV of US2018/0028664, incorporated herein by reference in its entirety. The non-cationic lipid can comprise, for example, 0-30% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, the non-cationic lipid content is 5-20% (mol) or 10-15% (mol) of the total lipid present in the lipid nanoparticle. In embodiments, the molar ratio of ionizable lipid to the neutral lipid ranges from about 2:1 to about 8:1 (e.g., about 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, or 8:1).

In some embodiments, the lipid nanoparticles do not comprise any phospholipids.

In some aspects, the lipid nanoparticle can further comprise a component, such as a sterol, to provide membrane integrity. One exemplary sterol that can be used in the lipid nanoparticle is cholesterol and derivatives thereof. Non-limiting examples of cholesterol

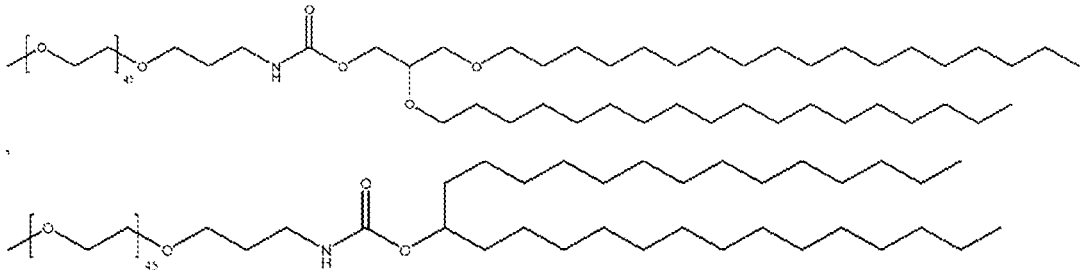
derivatives include polar analogues such as 5a-choiestanol, 53-coprostanol, choiesteryl-(2'-hydroxy)-ethyl ether, choiesteryl-(4'- hydroxy)-butyl ether, and 6-ketocholestanol; non-polar analogues such as 5a-cholestane, cholestenone, 5a-cholestanone, 5p-cholestanone, and cholesteryl decanoate; and mixtures thereof. In some embodiments, the cholesterol derivative is a polar analogue, e.g., choiesteryl-(4 '-hydroxy)-buty1 ether. Exemplary cholesterol derivatives are described in PCT publication W02009/127060 and US patent publication US2010/0130588, each of which is incorporated herein by reference in its entirety.

In some embodiments, the component providing membrane integrity, such as a sterol, can comprise 0-50% (mol) (e.g., 0-10%, 10-20%, 20-30%, 30-40%, or 40-50%) of the total lipid present in the lipid nanoparticle. In some embodiments, such a component is 20-50% (mol) 30-40% (mol) of the total lipid content of the lipid nanoparticle.

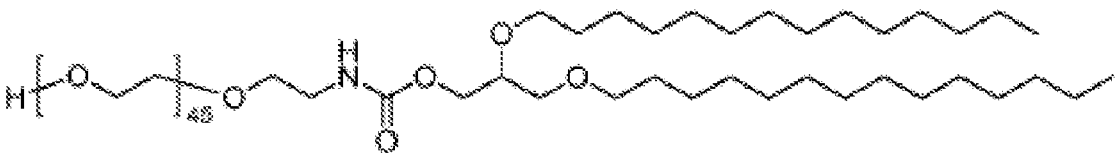
In some embodiments, the lipid nanoparticle can comprise a polyethylene glycol (PEG) or a conjugated lipid molecule. Generally, these are used to inhibit aggregation of lipid nanoparticles and/or provide steric stabilization. Exemplary conjugated lipids include, but are not limited to, PEG-lipid conjugates, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), cationic-polymer lipid (CPL) conjugates, and mixtures thereof. In some embodiments, the conjugated lipid molecule is a PEG-lipid conjugate, for example, a (methoxy polyethylene glycol)-conjugated lipid.

Exemplary PEG-lipid conjugates include, but are not limited to, PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG-ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-l-0-(w-methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N-(carbonyl-methoxypolyethylene glycol 2000)-l,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, 1,2-dimyristoyl-sn-glycerol, methoxypoly ethylene glycol (DMG-PEG-2K),or a mixture thereof. Additional exemplary PEG-lipid conjugates are described, for example, in US5,885,613, US6,287,591, US2003/0077829, US2003/0077829, US2005/0175682, US2008/0020058, US2011/0117125, US2010/0130588, US2016/0376224, US2017/0119904, and US/099823, the contents of all of which are incorporated herein by reference in their entirety. In some embodiments, a PEG-lipid is a compound of Formula III, III-a-I, III-a-2, III-b-1, III-b-2, or V of US2018/0028664, the

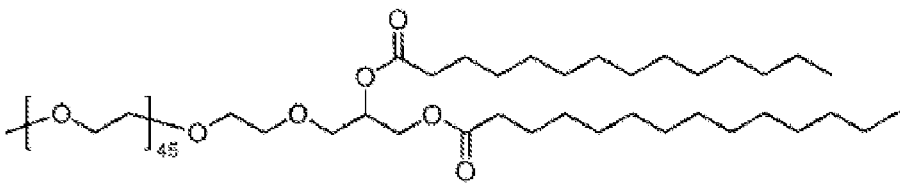
content of which is incorporated herein by reference in its entirety. In some embodiments, a PEG-lipid is of Formula II of US20150376115 or US2016/0376224, the content of both of which is incorporated herein by reference in its entirety. In some embodiments, the PEG-DAA conjugate can be, for example, PEG-dilauryloxypropyl, PEG- dimyristyloxypropyl, PEG-dipalmityloxypropyl, or PEG-distearyloxypropyl. The PEG-lipid can be one or more of PEG-DMG, PEG-dilaurylglycerol, PEG-dipalmitoylglycerol, PEG- disterylglycerol, PEG-dilaurylglycamide, PEG-dimyristylglycamide, PEG- dipalmitoylglycamide, PEG-disterylglycamide, PEG-cholesterol (l-[8'-(Cholest-5-en-3[beta]- oxy)carboxamido-3',6'-dioxaoctanyl] carbamoyl-[omega]-methyl-poly(ethylene glycol), PEG- DMB (3,4-Ditetradecoxylbenzyl- [omega]-methyl-poly(ethylene glycol) ether), and 1,2- dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises PEG-DMG, 1,2- dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises a structure selected from:



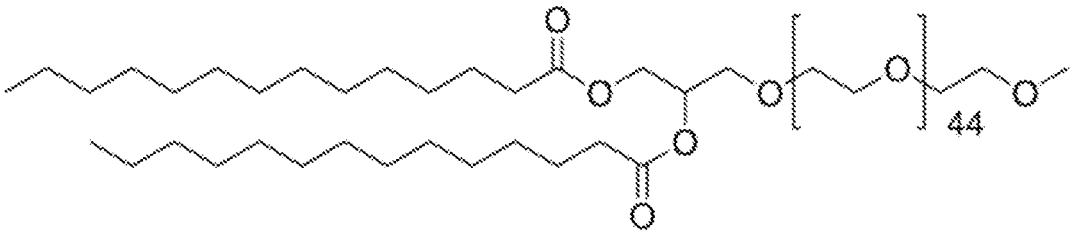
(xxii),



(xxiii),



(xxiv), and



(xxv).

In some embodiments, lipids conjugated with a molecule other than a PEG can also be used in place of PEG-lipid. For example, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), and cationic-polymer lipid (GPL) conjugates can be used in place of or in addition to the PEG-lipid.

Exemplary conjugated lipids, i.e., PEG-lipids, (POZ)-lipid conjugates, ATTA-lipid conjugates and cationic polymer-lipids are described in the PCT and LIS patent applications listed in Table 2 of WO2019051289A9 and in WO2020106946A1 the contents of all of which are incorporated herein by reference in their entirety.

In some embodiments an LNP comprises a compound of Formula (xix), a compound of Formula (xxi) and a compound of Formula (xxv). In some embodiments a LNP comprising a formulation of Formula (xix), Formula (xxi) and Formula (xxv)is used to deliver a GeneWriter composition described herein to the lung or pulmonary cells.

In some embodiments, the PEG or the conjugated lipid can comprise 0-20% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, PEG or the conjugated lipid content is 0.5- 10% or 2-5% (mol) of the total lipid present in the lipid nanoparticle. Molar ratios of the ionizable lipid, non-cationic-lipid, sterol, and PEG/conjugated lipid can be varied as needed. For example, the lipid particle can comprise 30-70% ionizable lipid by mole or by total weight of the composition, 0-60% cholesterol by mole or by total weight of the composition, 0-30% non-cationic-lipid by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. Preferably, the composition comprises 30-40% ionizable lipid by mole or by total weight of the composition, 40-50% cholesterol by mole or by total weight of the composition, and 10- 20% non-cationic-lipid by mole or by total weight of the composition. In some other embodiments, the composition is 50-75% ionizable lipid by mole or by total weight of the composition, 20-40% cholesterol by mole or by total weight of the

composition, and 5 to 10% non-cationic-lipid, by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. The composition may contain 60-70% ionizable lipid by mole or by total weight of the composition, 25-35% cholesterol by mole or by total weight of the composition, and 5-10% non-cationic-lipid by mole or by total weight of the composition. The composition may also contain up to 90% ionizable lipid by mole or by total weight of the composition and 2 to 15% non-cationic lipid by mole or by total weight of the composition. The formulation may also be a lipid nanoparticle formulation, for example comprising 8-30% ionizable lipid by mole or by total weight of the composition, 5-30% non- cationic lipid by mole or by total weight of the composition, and 0-20% cholesterol by mole or by total weight of the composition; 4-25% ionizable lipid by mole or by total weight of the composition, 4-25% non-cationic lipid by mole or by total weight of the composition, 2 to 25% cholesterol by mole or by total weight of the composition, 10 to 35% conjugate lipid by mole or by total weight of the composition, and 5% cholesterol by mole or by total weight of the composition; or 2-30% ionizable lipid by mole or by total weight of the composition, 2-30% non-cationic lipid by mole or by total weight of the composition, 1 to 15% cholesterol by mole or by total weight of the composition, 2 to 35% conjugate lipid by mole or by total weight of the composition, and 1-20% cholesterol by mole or by total weight of the composition; or even up to 90% ionizable lipid by mole or by total weight of the composition and 2-10% non-cationic lipids by mole or by total weight of the composition, or even 100% cationic lipid by mole or by total weight of the composition. In some embodiments, the lipid particle formulation comprises ionizable lipid, phospholipid, cholesterol and a PEG-ylated lipid in a molar ratio of 50: 10:38.5: 1.5. In some other embodiments, the lipid particle formulation comprises ionizable lipid, cholesterol and a PEG-ylated lipid in a molar ratio of 60:38.5: 1.5.

In some embodiments, the lipid particle comprises ionizable lipid, non-cationic lipid (e.g. phospholipid), a sterol (e.g., cholesterol) and a PEG-ylated lipid, where the molar ratio of lipids ranges from 20 to 70 mole percent for the ionizable lipid, with a target of 40-60, the mole percent of non-cationic lipid ranges from 0 to 30, with a target of 0 to 15, the mole percent of sterol ranges from 20 to 70, with a target of 30 to 50, and the mole percent of PEG-ylated lipid ranges from 1 to 6, with a target of 2 to 5.

In some embodiments, the lipid particle comprises ionizable lipid / non-cationic- lipid / sterol / conjugated lipid at a molar ratio of 50: 10:38.5: 1.5.

In an aspect, the disclosure provides a lipid nanoparticle formulation comprising phospholipids, lecithin, phosphatidylcholine and phosphatidylethanolamine.

In some embodiments, one or more additional compounds can also be included. Those compounds can be administered separately or the additional compounds can be included in the lipid nanoparticles of the invention. In other words, the lipid nanoparticles can contain other compounds in addition to the nucleic acid or at least a second nucleic acid, different than the first. Without limitations, other additional compounds can be selected from the group consisting of small or large organic or inorganic molecules, monosaccharides, disaccharides, trisaccharides, oligosaccharides, polysaccharides, peptides, proteins, peptide analogs and derivatives thereof, peptidomimetics, nucleic acids, nucleic acid analogs and derivatives, an extract made from biological materials, or any combinations thereof.

In some embodiments, a lipid nanoparticle (or a formulation comprising lipid nanoparticles) lacks reactive impurities (e.g., aldehydes or ketones), or comprises less than a preselected level of reactive impurities (e.g., aldehydes or ketones). While not wishing to be bound by theory, in some embodiments, a lipid reagent is used to make a lipid nanoparticle formulation, and the lipid reagent may comprise a contaminating reactive impurity (e.g., an aldehyde or ketone). A lipid regent may be selected for manufacturing based on having less than a preselected level of reactive impurities (e.g., aldehydes or ketones). Without wishing to be bound by theory, in some embodiments, aldehydes can cause modification and damage of RNA, e.g., cross-linking between bases and/or covalently conjugating lipid to RNA (e.g., forming lipid-RNA adducts). This may, in some instances, lead to failure of a reverse transcriptase reaction and/or incorporation of inappropriate bases, e.g., at the site(s) of lesion(s), e.g., a mutation in a newly synthesized target DNA.

In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less

than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation is produced using a plurality of lipid reagents, and each lipid reagent of the plurality independently meets one or more criterion described in this paragraph. In some embodiments, each lipid reagent of the plurality meets the same criterion, e.g., a criterion of this paragraph.

In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation comprises: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, total aldehyde content and/or quantity of any single reactive impurity (e.g., aldehyde) species is determined by liquid chromatography (LC), e.g., coupled with tandem mass spectrometry (MS/MS), e.g., according to the method described in Example 34. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications

of a nucleic acid molecule (e.g., an RNA molecule, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleotide or nucleoside (e.g., a ribonucleotide or ribonucleoside, e.g., comprised in or isolated from a template nucleic acid, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents, e.g., as described in Example 35. In embodiments, chemical modifications of a nucleic acid molecule, nucleotide, or nucleoside are detected by determining the presence of one or more modified nucleotides or nucleosides, e.g., using LC-MS/MS analysis, e.g., as described in Example 35.

In some embodiments, a nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) does not comprise an aldehyde modification, or comprises less than a preselected amount of aldehyde modifications. In some embodiments, on average, a nucleic acid has less than 50, 20, 10, 5, 2, or 1 aldehyde modifications per 1000 nucleotides, e.g., wherein a single cross-linking of two nucleotides is a single aldehyde modification. In some embodiments, the aldehyde modification is an RNA adduct (e.g., a lipid-RNA adduct). In some embodiments, the aldehyde-modified nucleotide is cross-linking between bases. In some embodiments, a nucleic acid (e.g., RNA) described herein comprises less than 50, 20, 10, 5, 2, or 1 cross-links between nucleotide.

In some embodiments, LNPs are directed to specific tissues by the addition of targeting domains. For example, biological ligands may be displayed on the surface of LNPs to enhance interaction with cells displaying cognate receptors, thus driving association with and cargo delivery to tissues wherein cells express the receptor. In some embodiments, the biological ligand may be a ligand that drives delivery to the liver, e.g., LNPs that display GalNAc result in delivery of nucleic acid cargo to hepatocytes that display asialoglycoprotein receptor (ASGPR). The work of Akinc et al. *Mol Ther* 18(7):1357-1364 (2010) teaches the conjugation of a trivalent GalNAc ligand to a PEG-lipid (GalNAc-PEG-DSG) to yield LNPs dependent on ASGPR for observable LNP cargo effect (see, e.g., Figure 6). Other ligand-displaying LNP formulations, e.g., incorporating folate, transferrin, or antibodies, are discussed in WO2017223135, which is

incorporated herein by reference in its entirety, in addition to the references used therein, namely 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; and Peer and Lieberman, Gene Ther. 2011 18:1127-1133.

In some embodiments, LNPs are selected for tissue-specific activity by the addition of a Selective ORgan Targeting (SORT) molecule to a formulation comprising traditional components, such as ionizable cationic lipids, amphipathic phospholipids, cholesterol and poly(ethylene glycol) (PEG) lipids. The teachings of Cheng et al. Nat Nanotechnol 15(4):313-320 (2020) demonstrate that the addition of a supplemental "SORT" component precisely alters the in vivo RNA delivery profile and mediates tissue-specific (e.g., lungs, liver, spleen) gene delivery and editing as a function of the percentage and biophysical property of the SORT molecule.

In some embodiments, the LNPs comprise biodegradable, ionizable lipids. In some embodiments, the LNPs comprise (9Z,l2Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,l2-dienoate, also called 3- ((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl (9Z,l2Z)-octadeca-9,l2-dienoate) or another ionizable lipid. See, e.g, lipids of WO2019/067992, WO/2017/173054, WO2015/095340, and WO2014/136086, as well as references provided therein. In some embodiments, the term cationic and ionizable in the context of LNP lipids is interchangeable, e.g., wherein ionizable lipids are cationic depending on the pH.

In some embodiments, multiple components of a Gene Writer system may be prepared as a single LNP formulation, e.g., an LNP formulation comprises mRNA encoding for the Gene Writer polypeptide and an RNA template. Ratios of nucleic acid components may be varied in order to maximize the properties of a therapeutic. In some embodiments, the ratio of RNA template to mRNA encoding a Gene Writer polypeptide is about 1:1 to 100:1, e.g., about 1:1 to

20:1, about 20:1 to 40:1, about 40:1 to 60:1, about 60:1 to 80:1, or about 80:1 to 100:1, by molar ratio. In other embodiments, a system of multiple nucleic acids may be prepared by separate formulations, e.g., one LNP formulation comprising a template RNA and a second LNP formulation comprising an mRNA encoding a Gene Writer polypeptide. In some embodiments, the system may comprise more than two nucleic acid components formulated into LNPs. In some embodiments, the system may comprise a protein, e.g., a Gene Writer polypeptide, and a template RNA formulated into at least one LNP formulation.

In some embodiments, the average LNP diameter of the LNP formulation may be between 10s of nm and 100s of nm, e.g., measured by dynamic light scattering (DLS). In some embodiments, the average LNP diameter of the LNP formulation may be from about 40 nm to about 150 nm, such as about 40 nm, 45 nm, 50 nm, 55 nm, 60 nm, 65 nm, 70 nm, 75 nm, 80 nm, 85 nm, 90 nm, 95 nm, 100 nm, 105 nm, 110 nm, 115 nm, 120 nm, 125 nm, 130 nm, 135 nm, 140 nm, 145 nm, or 150 nm. In some embodiments, the average LNP diameter of the LNP formulation may be from about 50 nm to about 100 nm, from about 50 nm to about 90 nm, from about 50 nm to about 80 nm, from about 50 nm to about 70 nm, from about 50 nm to about 60 nm, from about 60 nm to about 100 nm, from about 60 nm to about 90 nm, from about 60 nm to about 80 nm, from about 60 nm to about 70 nm, from about 70 nm to about 100 nm, from about 70 nm to about 90 nm, from about 70 nm to about 80 nm, from about 80 nm to about 100 nm, from about 80 nm to about 90 nm, or from about 90 nm to about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation may be from about 70 nm to about 100 nm. In a particular embodiment, the average LNP diameter of the LNP formulation may be about 80 nm. In some embodiments, the average LNP diameter of the LNP formulation may be about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation ranges from about lmm to about 500 mm, from about 5 mm to about 200 mm, from about 10 mm to about 100 mm, from about 20 mm to about 80 mm, from about 25 mm to about 60 mm, from about 30 mm to about 55 mm, from about 35 mm to about 50 mm, or from about 38 mm to about 42 mm.

A LNP may, in some instances, be relatively homogenous. A polydispersity index may be used to indicate the homogeneity of a LNP, e.g., the particle size distribution of the lipid nanoparticles. A small (e.g., less than 0.3) polydispersity index generally indicates a narrow particle size distribution. A LNP may have a polydispersity index from about 0 to about 0.25, such as 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, or 0.25. In some embodiments, the polydispersity index of a LNP may be from about 0.10 to about 0.20.

The zeta potential of a LNP may be used to indicate the electrokinetic potential of the composition. In some embodiments, the zeta potential may describe the surface charge of a LNP. Lipid nanoparticles with relatively low charges, positive or negative, are generally desirable, as more highly charged species may interact undesirably with cells, tissues, and other elements in the body. In some embodiments, the zeta potential of a LNP may be from about -10 mV to about +20 mV, from about -10 mV to about +15 mV, from about -10 mV to about +10 mV, from about -10 mV to about +5 mV, from about -10 mV to about 0 mV, from about -10 mV to about -5 mV, from about -5 mV to about +20 mV, from about -5 mV to about +15 mV, from about -5 mV to about +10 mV, from about -5 mV to about +5 mV, from about -5 mV to about 0 mV, from about 0 mV to about +20 mV, from about 0 mV to about +15 mV, from about 0 mV to about +10 mV, from about 0 mV to about +5 mV, from about +5 mV to about +20 mV, from about +5 mV to about +15 mV, or from about +5 mV to about +10 mV.

The efficiency of encapsulation of a protein and/or nucleic acid, e.g., Gene Writer polypeptide or mRNA encoding the polypeptide, describes the amount of protein and/or nucleic acid that is encapsulated or otherwise associated with a LNP after preparation, relative to the initial amount provided. The encapsulation efficiency is desirably high (e.g., close to 100%). The encapsulation efficiency may be measured, for example, by comparing the amount of protein or nucleic acid in a solution containing the lipid nanoparticle before and after breaking up the lipid nanoparticle with one or more organic solvents or detergents. An anion exchange resin may be used to measure the amount of free protein or nucleic acid (e.g., RNA) in a solution. Fluorescence may be used to measure the amount of free protein and/or nucleic acid (e.g., RNA) in a solution. For the lipid nanoparticles described herein, the encapsulation efficiency of a protein and/or nucleic acid may be at least 50%, for example 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%. In some embodiments, the encapsulation efficiency may be at least 80%. In some embodiments, the encapsulation efficiency may be at least 90%. In some embodiments, the encapsulation efficiency may be at least 95%.

A LNP may optionally comprise one or more coatings. In some embodiments, a LNP may be formulated in a capsule, film, or table having a coating. A capsule, film, or tablet

including a composition described herein may have any useful size, tensile strength, hardness or density.

Additional exemplary lipids, formulations, methods, and characterization of LNPs are taught by WO2020061457, which is incorporated herein by reference in its entirety.

In some embodiments, in vitro or ex vivo cell lipofections are performed using Lipofectamine MessengerMax (Thermo Fisher) or TransIT-mRNA Transfection Reagent (Mirus Bio). In certain embodiments, LNPs are formulated using the GenVoy_ILM ionizable lipid mix (Precision NanoSystems). In certain embodiments, LNPs are formulated using 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) or dilinoleylmethyl-4-dimethylaminobutyrate (DLin-MC3-DMA or MC3), the formulation and in vivo use of which are taught in Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012), incorporated herein by reference in its entirety.

LNP formulations optimized for the delivery of CRISPR-Cas systems, e.g., Cas9-gRNA RNP, gRNA, Cas9 mRNA, are described in WO2019067992 and WO2019067910, both incorporated by reference.

Additional specific LNP formulations useful for delivery of nucleic acids are described in US8158601 and US8168775, both incorporated by reference, which include formulations used in patisiran, sold under the name ONPATTRO.

Exemplary dosing of Gene Writer LNP may include about 0.1, 0.25, 0.3, 0.5, 1, 2, 3, 4, 5, 6, 8, 10, or 100 mg/kg (RNA). Exemplary dosing of AAV comprising a nucleic acid encoding one or more components of the system may include an MOI of about 10^{11}, 10^{12}, 10^{13}, and 10^{14} vg/kg.

Template RNA component of Gene Writer™ gene editor system

The Gene Writer systems described herein can transcribe an RNA sequence template into host target DNA sites by target-primed reverse transcription. By writing DNA sequence(s) via reverse transcription of the RNA sequence template directly into the host genome, the Gene Writer system can insert an object sequence into a target genome without the need for exogenous DNA sequences to be introduced into the host cell (unlike, for example, CRISPR systems), as well as eliminate an exogenous DNA insertion step. Therefore, the Gene Writer system provides

a platform for the use of customized RNA sequence templates containing object sequences, e.g., sequences comprising heterologous gene coding and/or function information.

For the purposes of description of the template RNA component of the Gene Writer system, the template RNA can be envisioned as comprising modular parts (Figures 18-19). A Gene Writer template may comprise all or some of the illustrated parts and modules can be combined, re-arranged, and/or left out. The illustrated modules are not intended to be limiting to potential elements included in the template and additional components can be readily envisioned, e.g., 5' and 3' terminal domains to improve template stability.

Table 4. The modules comprising a typical Gene Writer RNA template.

A = 5' homology arm; B = Ribozyme; C = 5' UTR; D = heterologous object sequence; E = 3' UTR; F = 3' homology arm

Module	Function
A: 5′ homology arm	The 5′ homology arm module is complementary to the DNA sequence 5′ to where the Gene Writer system nicks target DNA.
B: Ribozyme	The ribozyme module is a component of the 5′ UTR sequence of an endogenous retrotransposon, at the 5′ end of the 5′ UTR (Ruminski, D. J., et al, Journal of Biological Chemistry, 286(48), 41286–41295, 2011).
C: 5′ UTR	The 5′ UTR module is an RNA sequence or interacting moiety with which the Gene Writer polypeptide interacts and may be involved in the process of retrotransposition.
D: Heterologous object sequence	The heterologous object sequence is an RNA sequence that serves as the template for the Gene Writer polypeptide to insert a desired payload into the targeted genomic location.
E: 3' UTR	The 3′ UTR module is an RNA sequence or any other interacting moiety that the Gene Writer polypeptide interacts with to bind to the template RNA molecule and may be involved in the process of retrotransposition.
F: 3′ homology arm	The 3′ homology arm module is complementary to the DNA sequence 3′ to where the Gene Writer system nicks target DNA

In some embodiments the template RNA encodes a Gene Writer protein in *cis* with a heterologous object sequence. Various cis constructs were described, for example, in Kuroki-Kami et al (2019) *Mobile DNA* 10:23 (incorporated by reference herein in its entirety), and can

be used in combination with any of the embodiments described herein. For instance, in some embodiments, the template RNA comprises a heterologous object sequence, a sequence encoding a Gene Writer protein (e.g., a protein comprising (i) a reverse transcriptase domain and (ii) an endonuclease domain, e.g., as described herein), a 5' untranslated region, and a 3' untranslated region. The components may be included in various orders. In some embodiments, the Gene Writer protein and heterologous object sequence are encoded in different directions (sense vs. anti-sense), e.g., using an arrangement shown in Figure 3A of Kuroki-Kami et al, *Id.* In some embodiments the Gene Writer protein and heterologous object sequence are encoded in the same direction. In some embodiments, the nucleic acid encoding the polypeptide and the template RNA or the nucleic acid encoding the template RNA are covalently linked, e.g., are part of a fusion nucleic acid and/or are part of the same transcript. In some embodiments, the fusion nucleic acid comprises RNA or DNA.

The nucleic acid encoding the Gene Writer polypeptide may, in some instances, be 5' of the heterologous object sequence. For example, in some embodiments, the template RNA comprises, from 5' to 3', a 5' untranslated region, a sense-encoded Gene Writer polypeptide, a sense-encoded heterologous object sequence, and 3' untranslated region. In some embodiments, the template RNA comprises, from 5' to 3', a 5' untranslated region, a sense-encoded Gene Writer polypeptide, anti-sense-encoded heterologous object sequence, and 3' untranslated region.

In some embodiments, the RNA further comprises homology to the DNA target site.

It is understood that, when a template RNA is described as comprising an open reading frame or the reverse complement thereof, in some embodiments the template RNA must be converted into double stranded DNA (e.g., through reverse transcription) before the open reading frame can be transcribed and translated.

In certain embodiments, customized RNA sequence template can be identified, designed, engineered and constructed to contain sequences altering or specifying host genome function, for example by introducing a heterologous coding region into a genome; affecting or causing exon structure/alternative splicing; causing disruption of an endogenous gene; causing transcriptional activation of an endogenous gene; causing epigenetic regulation of an endogenous DNA; causing up- or down-regulation of operably liked genes, etc. In certain embodiments, a customized RNA sequence template can be engineered to contain sequences coding for exons and/or transgenes,

provide for binding sites to transcription factor activators, repressors, enhancers, etc., and combinations of thereof. In other embodiments, the coding sequence can be further customized with splice acceptor sites, poly-A tails. In certain embodiments the RNA sequence can contain sequences coding for an RNA sequence template homologous to the RLE transposase, be engineered to contain heterologous coding sequences, or combinations thereof.

The template RNA may have some homology to the target DNA. In some embodiments the template RNA has at least 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 or more bases of exact homology to the target DNA at the 3' end of the RNA. In some embodiments the template RNA has at least 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, or 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the target DNA, e.g., at the 5' end of the template RNA. In some embodiments the template RNA has a 3' untranslated region derived from a non-LTR retrotransposon, e.g. a non-LTR retrotransposons described herein. In some embodiments the template RNA has a 3' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the 3' sequence of a non-LTR retrotransposon, e.g., a non-LTR retrotransposon described herein, e.g. a non-LTR retrotransposon in Table 1, 2, or 3. In some embodiments the template RNA has a 5' untranslated region derived from a non-LTR retrotransposon, e.g. a non-LTR retrotransposons described herein. In some embodiments the template RNA has a 5' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, or 200 or more bases of at least 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater homology to the 5' sequence of a non-LTR retrotransposon, e.g., a non-LTR retrotransposon described herein, e.g. a non-LTR retrotransposon described in Table 2 or 3.

The template RNA component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein of the system. In some embodiments the template RNA has a 3' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 3' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system.

The template RNA component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein of the system. In some embodiments the template RNA has a 5' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 5' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system. In some embodiments, the 5' untranslated region comprises a pseudoknot, e.g., a pseudoknot that is capable of binding to the Gene Writer protein.

In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a stem-loop sequence. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a hairpin. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a helix. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a psuedoknot. In some embodiments the template RNA comprises a ribozyme. In some embodiments the ribozyme is similar to an hepatitis delta virus (HDV) ribozyme, e.g., has a secondary structure like that of the HDV ribozyme and/or has one or more activities of the HDV ribozyme, e.g., a self-cleavage activity. See, e.g., Eickbush et al., Molecular and Cellular Biology, 2010, 3142-3150.

In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 3' untranslated region) comprises one or more stem-loops or helices. Exemplary structures of R2 3' UTRs are shown, for example, in Ruschak et al. "Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs" RNA. 2004 Jun; 10(6): 978–987, e.g., at Figure 3, therein, and in Eikbush and Eikbush, "R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements" Mobile DNA (2012) 3:10; e.g., at Figure 3 therein, which articles are hereby incorporated by reference in their entirety.

In some embodiments, a template RNA described herein comprises a sequence that is capable of binding to a GeneWriter protein described herein. For instance, in some embodiments, the template RNA comprises an MS2 RNA sequence capable of binding to an MS2 coat protein sequence in the GeneWriter protein. In some embodiments, the template RNA comprises an RNA sequence capable of binding to a B-box sequence. In some embodiments, the template RNA comprises an RNA sequence (e.g., a crRNA sequence and/or tracrRNA

sequence) capable of binding to a dCas sequence in the GeneWriter protein. In some embodiments, in addition to or in place of a UTR, the template RNA is linked (e.g., covalently) to a non-RNA UTR, e.g., a protein or small molecule.

In some embodiments the template RNA has a poly-A tail at the 3' end. In some embodiments the template RNA does not have a poly-A tail at the 3' end.

In some embodiments the template RNA has a 5' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 or more bases of at least 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater homology to the 5' sequence of a non-LTR retrotransposon, e.g., a non-LTR retrotransposon described herein.

The template RNA of the system typically comprises an object sequence for insertion into a target DNA. The object sequence may be coding or non-coding.

In some embodiments a system or method described herein comprises a single template RNA. In some embodiments a system or method described herein comprises a plurality of template RNAs. In some embodiments the DNA encoding the template is circularized by the activity of enzymes, such as recombinases, to increase activity, as described in Yant el al., *Nature Biotechnology* 20: 990-1005, 2002.

In some embodiments, the heterologous object sequence may contain an open reading frame. In some embodiments the template RNA has a Kozak sequence. In some embodiments the template RNA has an internal ribosome entry site. In some embodiments the template RNA has a self-cleaving peptide such as a T2A or P2A site. In some embodiments the template RNA has a start codon. In some embodiments the template RNA has a splice acceptor site. In some embodiments the template RNA has a splice donor site. Exemplary splice acceptor and splice donor sites are described in WO2016044416, incorporated herein by reference in its entirety. Exemplary splice acceptor site sequences are known to those of skill in the art and include, by way of example only, CTGACCCTTCTCTCTCTCCCCCAGAG (SEQ ID NO: 1620) (from human HBB gene) and TTTCTCTCCCACAAG (SEQ ID NO: 1621) (from human immunoglobulin-gamma gene). In some embodiments the template RNA has a microRNA binding site downstream of the stop codon. In some embodiments the template RNA has a polyA tail downstream of the stop codon of an open reading frame. In some embodiments the template RNA comprises one or more exons. In some embodiments the template RNA comprises one or more introns. In some embodiments the template RNA comprises a eukaryotic transcriptional

terminator. In some embodiments the template RNA comprises an enhanced translation element or a translation enhancing element. In some embodiments the RNA comprises the human T-cell leukemia virus (HTLV-1) R region. In some embodiments the RNA comprises a posttranscriptional regulatory element that enhances nuclear export, such as that of Hepatitis B Virus (HPRE) or Woodchuck Hepatitis Virus (WPRE). In some embodiments, in the template RNA, the heterologous object sequence encodes a polypeptide and is coded in an antisense direction with respect to the 5' and 3' UTR. In some embodiments, in the template RNA, the heterologous object sequence encodes a polypeptide and is coded in a sense direction with respect to the 5' and 3' UTR.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a GeneWriter system. For instance, the microRNA binding site can be chosen on the basis that is is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the template RNA is present in a non-target cell, it would be bound by the miRNA, and when the template RNA is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the template RNA may interfere with insertion of the heterologous object sequence into the genome. Accordingly, the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells. A system having a microRNA binding site in the template RNA (or DNA encoding it) may also be used in combination with a nucleic acid encoding a Gene Writer polypeptide, wherein expression of the Gene Writer polypeptide is regulated by a second microRNA binding site, e.g., as described herein, e.g., in the section entitled "Polypeptide component of Gene Writer gene editor system". In some embodiments, e.g., for liver indications, a miRNA is selected from Table 4 of WO2020014209.

In some embodiments, the object sequence may contain a non-coding sequence. For example, the template RNA may comprise a regulatory element, e.g., a promoter or enhancer sequence or miRNA binding site. In some embodiments, integration of the object sequence at a target site will result in upregulation of an endogenous gene. In some embodiments, integration of the object sequence at a target site will result in downregulation of an endogenous gene. In

some embodiments the template RNA comprises a tissue specific promoter or enhancer, each of which may be unidirectional or bidirectional. In some embodiments the promoter is an RNA polymerase I promoter, RNA polymerase II promoter, or RNA polymerase III promoter. In some embodiments the promoter comprises a TATA element. In some embodiments the promoter comprises a B recognition element. In some embodiments the promoter has one or more binding sites for transcription factors. In some embodiments the non-coding sequence is transcribed in an antisense-direction with respect to the 5' and 3' UTR. In some the non-coding sequence is transcribed in a sense direction with respect to the 5' and 3' UTR.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a promoter sequence, e.g., a tissue specific promoter sequence. In some embodiments, the tissue-specific promoter is used to increase the target-cell specificity of a Gene Writer system. For instance, the promoter can be chosen on the basis that it is active in a target cell type but not active in (or active at a lower level in) a non-target cell type. Thus, even if the promoter integrated into the genome of a non-target cell, it would not drive expression (or only drive low level expression) of an integrated gene. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a microRNA binding site, e.g., in the template RNA or a nucleic acid encoding a Gene Writer protein, e.g., as described herein. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a DNA encoding a Gene Writer polypeptide, driven by a tissue-specific promoter, e.g., to achieve higher levels of Gene Writer protein in target cells than in non-target cells. In some embodiments, e.g., for liver indications, a tissue-specific promoter is selected from Table 3 of WO2020014209, which is hereby incorporated by reference.

In some embodiments, a Gene Writer system, e.g., DNA encoding a Gene Writer polypeptide, DNA encoding a template RNA, or DNA or RNA encoding a heterologous object sequence, is designed such that one or more elements is operably linked to a tissue-specific promoter, e.g., a promoter that is active in T-cells. In further embodiments, the T-cell active promoter is inactive in other cell types, e.g., B-cells, NK cells. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of the T-cell receptor, e.g., TRAC, TRBC, TRGC, TRDC. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of a T-cell-specific cluster of

differentiation protein, e.g., CD3, e.g., CD3D, CD3E, CD3G, CD3Z. In some embodiments, T-cell-specific promoters in Gene Writer systems are discovered by comparing publicly available gene expression data across cell types and selecting promoters from the genes with enhanced expression in T-cells. In some embodiments, promoters may be selecting depending on the desired expression breadth, e.g., promoters that are active in T-cells only, promoters that are active in NK cells only, promoters that are active in both T-cells and NK cells.

In some embodiments the template RNA comprises a microRNA sequence, a siRNA sequence, a guide RNA sequence, a piwi RNA sequence.

In some embodiments the template RNA comprises a site that coordinates epigenetic modification. In some embodiments the template RNA comprises an element that inhibits, e.g., prevents, epigenetic silencing. In some embodiments the template RNA comprises a chromatin insulator. For example, the template RNA comprises a CTCF site or a site targeted for DNA methylation.

In order to promote higher level or more stable gene expression, the template RNA may include features that prevent or inhibit gene silencing. In some embodiments, these features prevent or inhibit DNA methylation. In some embodiments, these features promote DNA demethylation. In some embodiments, these features prevent or inhibit histone deacetylation. In some embodiments, these features prevent or inhibit histone methylation. In some embodiments, these features promote histone acetylation. In some embodiments, these features promote histone demethylation. In some embodiments, multiple features may be incorporated into the template RNA to promote one or more of these modifications. CpG dinculeotides are subject to methylation by host methyl transferases. In some embodiments, the template RNA is depleted of CpG dinucleotides, e.g., does not comprise CpG nucleotides or comprises a reduced number of CpG dinucleotides compared to a corresponding unaltered sequence. In some embodiments, the promoter driving transgene expression from integrated DNA is depleted of CpG dinucleotides.

In some embodiments the template RNA comprises a gene expression unit composed of at least one regulatory region operably linked to an effector sequence. The effector sequence may be a sequence that is transcribed into RNA (e.g., a coding sequence or a non-coding sequence such as a sequence encoding a micro RNA).

In some embodiments the object sequence of the template RNA is inserted into a target genome in an endogenous intron. In some embodiments the object sequence of the template

RNA is inserted into a target genome and thereby acts as a new exon. In some embodiments the insertion of the object sequence into the target genome results in replacement of a natural exon or the skipping of a natural exon.

In some embodiments, the object sequence of the template RNA is inserted into the target genome in a genomic safe harbor site, such as AAVS1, CCR5, ROSA26, or the albumin locus. In some embodiments, a Gene Writer is used to integrate a CAR into the T-cell receptor α constant (*TRAC*) locus (Eyquem et al Nature 543, 113-117 (2017)). In some embodiments, a Gene Writer is used to integrate a CAR into a T-cell receptor β constant (*TRBC*) locus. Many other safe harbors have been identified by computational approaches (Pellenz et al Hum Gen Ther 30, 814-828 (2019)) and could be used for Gene Writer-mediated integration. In some embodiments, the object sequence of the template RNA is added to the genome in an intergenic or intragenic region. In some embodiments, the object sequence of the template RNA is added to the genome 5' or 3' within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous active gene. In some embodiments the object sequence of the template RNA is added to the genome 5' or 3' within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous promoter or enhancer. In some embodiments, the object sequence of the template RNA can be, e.g., 50-50,000 base pairs (e.g., between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp. In some embodiments, the heterologous object sequence is less than 1,000, 1,300, 1500, 2,000, 3,000, 4,000, 5,000, or 7,500 nucleotides in length.

In some embodiments the genomic safe harbor site is a Natural Harbor™ site. In some embodiments the Natural Harbor™ site is ribosomal DNA (rDNA). In some embodiments the Natural Harbor™ site is 5S rDNA, 18S rDNA, 5.8S rDNA, or 28S rDNA. In some embodiments the Natural Harbor™ site is the Mutsu site in 5S rDNA. In some embodiments the Natural Harbor™ site is the R2 site, the R5 site, the R6 site, the R4 site, the R1 site, the R9 site, or the RT site in 28S rDNA. In some embodiments the Natural Harbor™ site is the R8 site or the R7 site in 18S rDNA. In some embodiments the Natural Harbor™ site is DNA encoding transfer RNA (tRNA). In some embodiments the Natural Harbor™ site is DNA encoding tRNA-Asp or tRNA-Glu. In some embodiments the Natural Harbor™ site is DNA encoding spliceosomal

RNA. In some embodiments the Natural Harbor™ site is DNA encoding small nuclear RNA (snRNA) such as U2 snRNA.

Thus, in some aspects, the present disclosure provides a method of inserting a heterologous object sequence into a Natural Harbor™ site. In some embodimetns, the method comprises using a GeneWriter system described herein, e.g., using a polypeptide of any of Tables 1-3 or a polypeptide having sequence similarity thereto, e.g., at least 80%, 85%, 90%, or 95% identity thereto. In some embodiments, the method comprises using an enzyme, e.g., a retrotransposase, to insert the heterologous object sequence into the Natural Harbor™ site. In some aspects, the present disclosure provides a host human cell comprising a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) situated at a Natural Harbor™ site in the genome of the cell. In some embodiments, the Natural Harbor™ site is a site described in Table 5 below. In some embodiments, the heterologous object sequence is inserted within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs of a sequence shown in Table 5. In some embodiments, the heterologous object sequence is inserted within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of a sequence shown in Table 5. In some embodiments, the heterologous object sequence is inserted into a site having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to a sequence shown in Table 5. In some embodiments, the heterologous object sequence is inserted within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs, or within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb, of a site having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to a sequence shown in Table 5. In some embodiments, the heterologous object sequence is inserted within a gene indicated in Column 5 of Table 5, or within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs, or within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb, of the gene.

Table 5. Natural Harbor™ sites. Column 1 indicates a retrotransposon that inserts into the Natural Harbor™ site. Column 2 indicates the gene at the Natural Harbor™ site. Columns 3 and 4 show exemplary human genome sequence 5' and 3' of the insertion site (for example, 250 bp). Columns 5 and 6 list the example gene symbol and corresponding Gene ID.

Table 5: Exemplary Natural Harbor™ Sites

Target Site	Target Gene	5' flanking sequence	3' flanking sequence	Example Gene Symbol	Example Gene ID
R2	28S rDNA	CCGGTCCCCCCCGCCGGGTCC GCCCCCGGGGCCGCGGTTCC GCGCGGCGCCTCGCCTCGGC CGGCGCCTAGCAGCCGACTT AGAACTGGTGCGGACCAGGG GAATCCGACTGTTTAATTAAA ACAAAGCATCGCGAAGGCCC GCGGCGGGTGTTGACGCGAT GTGATTTCTGCCCAGTGCTCT GAATGTCAAAGTGAAGAAAT TCAATGAAGCGCGGGTAAAC GGCGGGAGTAACTATGACTC TCTTAAG (SEQ ID NO: 1508)	GTAGCCAAATGCCTCGTCATC TAATTAGTGACGCGCATGAAT GGATGAACGAGATTCCCACT GTCCCTACCTACTATCCAGCG AAACCACAGCCAAGGGAACG GGCTTGGCGGAATCAGCGGG GAAAGAAGACCCTGTTGAGC TTGACTCTAGTCTGGCACGGT GAAGAGACATGAGAGGTGTA GAATAAGTGGGAGGCCCCCG GCGCCCCCCCGGTGTCCCCGC GAGGGGCCCGGGGCGGGGT CCGCCG (SEQ ID NO: 1513)	RNA28SN1	106632264
R4	28S rDNA	GCGGTTCCGCGCGGCGCCTC GCCTCGGCCGGCGCCTAGCA GCCGACTTAGAACTGGTGCG GACCAGGGGAATCCGACTGT TTAATTAAAACAAAGCATCGC GAAGGCCCGCGGCGGGTGTT GACGCGATGTGATTTCTGCCC AGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGC GGGTAAACGGCGGGAGTAAC TATGACTCTCTTAAGGTAGCC AAATGCCTCGTCATCTAATTA GTGACG (SEQ ID NO: 1509)	CGCATGAATGGATGAACGAG ATTCCCACTGTCCCTACCTACT ATCCAGCGAAACCACAGCCA AGGGAACGGGCTTGGCGGA ATCAGCGGGGAAAGAAGACC CTGTTGAGCTTGACTCTAGTC TGGCACGGTGAAGAGACATG AGAGGTGTAGAATAAGTGGG AGGCCCCCGGCGCCCCCCCG GTGTCCCCGCGAGGGGCCCG GGGCGGGGTCCGCCGGCCCT GCGGGCCGCCGGTGAAATAC CACTACTC (SEQ ID NO: 1514)	RNA28SN1	106632264
R5	28S rDNA	TCCCCCCCGCCGGGTCCGCCC CCGGGGCCGCGGTTCCGCGC GGCGCCTCGCCTCGGCCGGC GCCTAGCAGCCGACTTAGAA CTGGTGCGGACCAGGGGAAT CCGACTGTTTAATTAAAACAA AGCATCGCGAAGGCCCGCGG CGGGTGTTGACGCGATGTGA TTTCTGCCCAGTGCTCTGAAT GTCAAAGTGAAGAAATTCAA TGAAGCGCGGGTAAACGGCG GGAGTAACTATGACTCTCTTA AGGTAG (SEQ ID NO: 1510)	CCAAATGCCTCGTCATCTAAT TAGTGACGCGCATGAATGGA TGAACGAGATTCCCACTGTCC CTACCTACTATCCAGCGAAAC CACAGCCAAGGGAACGGGCT TGGCGGAATCAGCGGGGAAA GAAGACCCTGTTGAGCTTGA CTCTAGTCTGGCACGGTGAA GAGACATGAGAGGTGTAGAA TAAGTGGGAGGCCCCCGGCG CCCCCCCGGTGTCCCCGCGAG GGGCCCGGGGCGGGGTCCG CCGGCCC (SEQ ID NO: 1515)	RNA28SN1	106632264

R9	28S rDNA	CGGCGCGCTCGCCGGCCGAG GTGGGATCCCGAGGCCTCTC CAGTCCGCCGAGGGCGCACC ACCGGCCCGTCTCGCCCGCCG CGCCGGGGAGGTGGAGCAC GAGCGCACGTGTTAGGACCC GAAAGATGGTGAACTATGCC TGGGCAGGGCGAAGCCAGA GGAAACTCTGGTGGAGGTCC GTAGCGGTCCTGACGTGCAA ATCGGTCGTCCGACCTGGGT ATAGGGGCGAAAGACTAATC GAACCATCTAG (SEQ ID NO: 1511)	TAGCTGGTTCCCTCCGAAGTT TCCCTCAGGATAGCTGGCGCT CTCGCAGACCCGACGCACCCC CGCCACGCAGTTTTATCCGGT AAAGCGAATGATTAGAGGTC TTGGGGCCGAAACGATCTCA ACCTATTCTCAAACTTTAAAT GGGTAAGAAGCCCGGCTCGC TGGCGTGGAGCCGGGCGTGG AATGCGAGTGCCTAGTGGGC CACTTTTGGTAAGCAGAACTG GCGCTGCGGGATGAACCGAA CGCC (SEQ ID NO: 1516)	RNA28SN1	106632264
R8	18S rDNA	GCATTCGTATTGCGCCGCTAG AGGTGAAATTCTTGGACCGG CGCAAGACGGACCAGAGCGA AAGCATTTGCCAAGAATGTTT TCATTAATCAAGAACGAAAGT CGGAGGTTCGAAGACGATCA GATACCGTCGTAGTTCCGACC ATAAACGATGCCGACCGGCG ATGCGGCGGCGTTATTCCCAT GACCCGCCGGGCAGCTTCCG GGAAACCAAAGTCTTTGGGT TCCGGGGGGAGTATGGTTGC AAAGC (SEQ ID NO: 1512)	TGAAACTTAAAGGAATTGAC GGAAGGGCACCACCAGGAGT GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA CGCGACCCCCGAGCGGTCGG CGTCCC (SEQ ID NO: 1517)	RNA18SN1	106631781
R4-2_SRa	tRNA-Asp			TRD-GTC1-1	100189207
LIN25_SM	tRNA-Glu			TRE-CTC1-1	100189384
R1	28S rDNA	TAGCAGCCGACTTAGAACTG GTGCGGACCAGGGGAATCCG ACTGTTTAATTAAAACAAAGC ATCGCGAAGGCCCGCGGCGG GTGTTGACGCGATGTGATTTC TGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAA GCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAAGG TAGCCAAATGCCTCGTCATCT AATTAGTGACGCGCATGAAT	ACCTACTATCCAGCGAAACCA CAGCCAAGGGAACGGGCTTG GCGGAATCAGCGGGGAAAG AAGACCCTGTTGAGCTTGACT CTAGTCTGGCACGGTGAAGA GACATGAGAGGTGTAGAATA AGTGGGAGGCCCCCGGCGCC CCCCCGGTGTCCCCGCGAGG GGCCCGGGGCGGGGTCCGCC GGCCCTGCGGGCCGCCGGTG AAATACCACTACTCTGATCGT TTTTTCACTGACCCGGTGAGG	RNA28SN1	106632264

		GGATGAACGAGATTCCCACT GTCCCT (SEQ ID NO: 1518)	CGGGGGG (SEQ ID NO: 1524)		
R6	28S rDNA	CCCCCCGCCGGGTCCGCCCCC GGGGCCGCGGTTCCGCGCGG CGCCTCGCCTCGGCCGGCGC CTAGCAGCCGACTTAGAACT GGTGCGGACCAGGGGAATCC GACTGTTTAATTAAAACAAAG CATCGCGAAGGCCCGCGGCG GGTGTTGACGCGATGTGATT TCTGCCCAGTGCTCTGAATGT CAAAGTGAAGAAATTCAATG AAGCGCGGGTAAACGGCGG GAGTAACTATGACTCTCTTAA GGTAGCC (SEQ ID NO: 1519)	AAATGCCTCGTCATCTAATTA GTGACGCGCATGAATGGATG AACGAGATTCCCACTGTCCCT ACCTACTATCCAGCGAAACCA CAGCCAAGGGAACGGGCTTG GCGGAATCAGCGGGGAAAG AAGACCCTGTTGAGCTTGACT CTAGTCTGGCACGGTGAAGA GACATGAGAGGTGTAGAATA AGTGGGAGGCCCCCGGCGCC CCCCCGGTGTCCCCGCGAGG GGCCCGGGGCGGGGTCCGCC GGCCCTG (SEQ ID NO: 1525)	RNA28SN1	106632264
R7	18S rDNA	GCGCAAGACGGACCAGAGCG AAAGCATTTGCCAAGAATGTT TTCATTAATCAAGAACGAAAG TCGGAGGTTCGAAGACGATC AGATACCGTCGTAGTTCCGAC CATAAACGATGCCGACCGGC GATGCGGCGGCGTTATTCCC ATGACCCGCCGGGCAGCTTC CGGGAAACCAAAGTCTTTGG GTTCCGGGGGGAGTATGGTT GCAAAGCTGAAACTTAAAGG AATTGACGGAAGGGCACCAC CAGGAGT (SEQ ID NO: 1520)	GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA CGCGACCCCCGAGCGGTCGG CGTCCCCCAACTTCTTAGAGG GACAAGTGGCGTTCAGCCAC CCGAG (SEQ ID NO: 1526)	RNA18SN1	106631781
RT	28S rDNA	GGCCGGGCGCGACCCGCTCC GGGGACAGTGCCAGGTGGG GAGTTTGACTGGGGCGGTAC ACCTGTCAAACGGTAACGCA GGTGTCCTAAGGCGAGCTCA GGGAGGACAGAAACCTCCCG TGGAGCAGAAGGGCAAAAG CTCGCTTGATCTTGATTTTCA GTACGAATACAGACCGTGAA AGCGGGGCCTCACGATCCTTC TGACCTTTTGGGTTTTAAGCA GGAGGTGTCAGAAAAGTTAC CACAGGGAT (SEQ ID NO: 1521)	AACTGGCTTGTGGCGGCCAA GCGTTCATAGCGACGTCGCTT TTTGATCCTTCGATGTCGGCT CTTCCTATCATTGTGAAGCAG AATTCACCAAGCGTTGGATTG TTCACCCACTAATAGGGAACG TGAGCTGGGTTTAGACCGTC GTGAGACAGGTTAGTTTTACC CTACTGATGATGTGTTGTTGC CATGGTAATCCTGCTCAGTAC GAGAGGAACCGCAGGTTCAG ACATTTGGTGTATGTGCTTGG C (SEQ ID NO: 1527)	RNA28SN1	106632264

Mutsu	5S rDNA	GTCTACGGCCATACCACCC (SEQ ID NO: 1522)	TGAACGCGCCCGATCTCGTCT GATCTCGGAAGCTAAGCAGG GTCGGGCCTGGTTAGTACTT GGATGGGAGACCGCCTGGGA ATACCGGGTGCTGTAGGCTTT (SEQ ID NO: 1528)	RNA5S1	100169751
Utopia /Keno	U2 snRNA	ATCGCTTCTCGGCCTTTTGGC TAAGATCAAGTGTAGTA (SEQ ID NO: 1523)	TCTGTTCTTATCAGTTTAATAT CTGATACGTCCTCTATCCGAG GACAATATATTAAATGGATTT TTGGAGCAGGGAGATGGAAT AGGAGCTTGCTCCGTCCACTC CACGCATCGACCTGGTATTGC AGTACCTCCAGGAACGGTGC ACCC (SEQ ID NO: 1529)	RNU2-1	6066

In some embodiments, a system or method described herein results in insertion of a heterologous sequence into a target site in the human genome. In some embodiments, the target site in the human genome has sequence similarity to the corresponding target site of the corresponding wild-type retrotransposase (e.g., the retrotransposase from which the GeneWriter was derived) in the genome of the organism to which it is native. For instance, in some embodiments, the identity between the 40 nucleotides of human genome sequence centered at the insertion site and the 40 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%. In some embodiments, the identity between the 100 nucleotides of human genome sequence centered at the insertion site and the 100 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%. In some embodiments, the identity between the 500 nucleotides of human genome sequence centered at the insertion site and the 500 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%.

Additional Template Features

In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined *in silico.* In embodiments, the template RNA is predicted to have minimal energy structures between -280 and -480 kcal/mol (e.g., between -280 to -300, -300 to -350, -350 to -400, -400 to -450, or -450 to -480 kcal/mol), e.g., as measured by RNAstructure, e.g., as described in Turner and Mathews *Nucleic Acids Res* 38:D280-282 (2009) (incorporated herein by reference in its entirety).

In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined *in vitro.* In embodiments, the template RNA is sequence optimized, e.g., to reduce secondary structure as determined *in vitro*, for example, by SHAPE-MaP (e.g., as described in Siegfried et al. Nat Methods 11:959-965 (2014); incorporated herein by reference in its entirety).In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined in cells. In embodiments, the template RNA is sequence optimized, e.g., to reduce secondary structure as measured in cells, for example, by DMS-MaPseq (e.g., as described in Zubradt et al. Nat Methods 14:75-82 (2017); incorporated by reference herein in its entirety).

Additional Functional Characteristics for Gene Writers™

A Gene Writer as described herein may, in some instances, be characterized by one or more functional measurements or characteristics. In some embodiments, the DNA binding domain has one or more of the functional characteristics described below. In some embodiments, the RNA binding domain has one or more of the functional characteristics described below. In some embodiments, the endonuclease domain has one or more of the functional characteristics described below. In some embodiments, the reverse transcriptase domain has one or more of the functional characteristics described below. In some embodiments, the template (e.g., template RNA) has one or more of the functional characteristics described below. In some embodiments, the target site bound by the Gene Writer has one or more of the functional characteristics described below.

Gene Writer Polypeptide

DNA Binding Domain

In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with greater affinity than a reference DNA binding domain. In some embodiments, the reference DNA binding domain is a DNA binding domain from R2_BM of B. mori. In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM).

In some embodiments, the affinity of a DNA binding domain for its target sequence (e.g., dsDNA target sequence) is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety).

In embodiments, the DNA binding domain is capable of binding to its target sequence (e.g., dsDNA target sequence), e.g, with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM) in the presence of a molar excess of scrambled sequence competitor dsDNA, e.g., of about 100-fold molar excess.

In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) more frequently than any other sequence in the genome of a target cell, e.g., human target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated herein by reference in its entirety). In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) at least about 5-fold or 10-fold, more frequently than any other sequence in the genome of a target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010), *supra.*

In some embodiments, a Gene Writer polypeptide comprises a modification to a DNA-binding domain, e.g., relative to the wild-type polypeptide. In some embodiments, the DNA-binding domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original DNA-binding domain. In some embodiments, the DNA-binding domain is modified to include a heterologous functional domain that binds specifically to a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain replaces at least a portion (e.g., the entirety of) the prior DNA-binding domain of the polypeptide. In some embodiments, the functional domain comprises a zinc finger (e.g., a zinc finger that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain comprises a Cas domain (e.g., a Cas domain that specifically binds to the

target nucleic acid (e.g., DNA) sequence of interest. In embodiments, the Cas domain comprises a Cas9 or a mutant or variant thereof (e.g., as described herein). In embodiments, the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In embodiments, the Cas domain is directed to a target nucleic acid (e.g., DNA) sequence of interest by the gRNA. In embodiments, the Cas domain is encoded in the same nucleic acid (e.g., RNA) molecule as the gRNA. In embodiments, the Cas domain is encoded in a different nucleic acid (e.g., RNA) molecule from the gRNA.

RNA Binding Domain

In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain. In some embodiments, the reference RNA binding domain is an RNA binding domain from R2_BM of B. mori. In some embodiments, the RNA binding domain is capable of binding to a template RNA with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in cells (e.g., by FRET or CLIP-Seq).

In some embodiments, the RNA binding domain is associated with the template RNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated by reference herein in its entirety). In some embodiments, the RNA binding domain is associated with the template RNA in cells (e.g., in HEK293T cells) at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019), *supra.*

Endonuclease Domain

In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA. In

some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA, e.g., in a cell (e.g., a HEK293T cell). In some embodiments, the frequency of association between the endonuclease domain and the target DNA or scrambled DNA is measured by ChIP-seq, e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain can catalyze the formation of a nick at a target sequence, e.g., to an increase of at least about 5-fold or 10-fold relative to a non-target sequence (e.g., relative to any other genomic sequence in the genome of the target cell). In some embodiments, the level of nick formation is determined using NickSeq, e.g., as described in Elacqua et al. (2019) *bioRxiv* doi.org/10.1101/867937 (incorporated herein by reference in its entirety).

In some embodiments, the endonuclease domain is capable of nicking DNA in vitro. In embodiments, the nick results in an exposed base. In embodiments, the exposed base can be detected using a nuclease sensitivity assay, e.g., as described in Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809 (incorporated by reference herein in its entirety). In embodiments, the level of exposed bases (e.g., detected by the nuclease sensitivity assay) is increased by at least 10%, 50%, or more relative to a reference endonuclease domain. In some embodiments, the reference endonuclease domain is an endonuclease domain from R2_BM of B. mori.

In some embodiments, the endonuclease domain is capable of nicking DNA in a cell. In embodiments, the endonuclease domain is capable of nicking DNA in a HEK293T cell. In embodiments, an unrepaired nick that undergoes replication in the absence of Rad51 results in increased NHEJ rates at the site of the nick, which can be detected, e.g., by using a Rad51 inhibition assay, e.g., as described in Bothmer et al. (2017) *Nat Commun* 8:13905 (incorporated by reference herein in its entirety). In embodiments, NHEJ rates are increased above 0-5%. In embodiments, NHEJ rates are increased to 20-70% (e.g., between 30%-60% or 40-50%), e.g., upon Rad51 inhibition.

In some embodiments, the endonuclease domain releases the target after cleavage. In some embodiments, release of the target is indicated indirectly by assessing for multiple turnovers by the enzyme, e.g., as described in Yourik at al. RNA 25(1):35-44 (2019)

(incorporated herein by reference in its entirety) and shown in Figure 2. In some embodiments, the k_{exp} of an endonuclease domain is 1 x 10^{-3} – 1 x 10-5 min-1 as measured by such methods.

In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} in vitro. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8, s^{-1} M^{-1} in vitro. In embodiments, catalytic efficiency is determined as described in Chen et al. (2018) *Science* 360(6387):436-439 (incorporated herein by reference in its entirety). In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} in cells. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8 s^{-1} M^{-1} in cells.

In some embodiments, a Gene Writer polypeptide comprises a modification to an endonuclease domain, e.g., relative to the wild-type polypeptide. In some embodiments, the endonuclease domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original endonuclease domain. In some embodiments, the endonuclease domain is modified to include a heterologous functional domain that binds specifically to and/or induces endonuclease cleavage of a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the endonuclease domain comprises a zinc finger. In some embodiments, the endonuclease domain comprises a Cas domain (e.g., a Cas9 or a mutant or variant thereof). In embodiments, the endonuclease domain comprising the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In some embodiments, the endonuclease domain is modified to include a functional domain that does not target a specific target nucleic acid (e.g., DNA) sequence. In embodiments, the endonuclease domain comprises a Fok1 domain.

Reverse Transcriptase Domain

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* relative to a reference reverse transcriptase domain. In some embodiments, the reference reverse transcriptase domain is a reverse transcriptase domain from R2_BM of B. mori or a viral reverse transcriptase domain, e.g., the RT domain from M-MLV.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* of less than about 5 x 10^{-3}/nt, 5 x 10^{-4}/nt, or 5 x 10^{-6}/nt,

e.g., as measured on a 1094 nt RNA. In embodiments, the *in vitro* premature termination rate is determined as described in Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (incorporated by reference herein its entirety).

In some embodiments, the reverse transcriptase domain is able to complete at least about 30% or 50% of integrations in cells. The percent of complete integrations can be measured by dividing the number of substantially full-length integration events (e.g., genomic sites that comprise at least 98% of the expected integrated sequence) by the number of total (including substantially full-length and partial) integration events in a population of cells. In embodiments, the integrations in cells is determined (e.g., across the integration site) using long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein its in entirety).

In embodiments, quantifying integrations in cells comprises counting the fraction of integrations that contain at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the DNA sequence corresponding to the template RNA (e.g., a template RNA having a length of at least 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, or 5 kb, e.g., a length between 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, or 4-5 kb).

In some embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro*. In embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro* at a rate between 0.1 – 50 nt/sec (e.g., between 0.1-1, 1-10, or 10-50 nt/sec). In embodiments, polymerization of dNTPs by the reverse transcriptase domain is measured by a single-molecule assay, e.g., as described in Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (incorporated by reference in its entirety).

In some embodiments, the reverse transcriptase domain has an *in vitro* error rate (e.g., misincorporation of nucleotides) of between $1 \times 10^{-3} - 1 \times 10^{-4}$ or $1 \times 10^{-4} - 1 \times 10^{-5}$ substitutions/nt , e.g., as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated herein by reference in its entirety). In some embodiments, the reverse transcriptase domain has an error rate (e.g., misincorporation of nucleotides) in cells (e.g., HEK293T cells) of between $1 \times 10^{-3} - 1 \times 10^{-4}$ or $1 \times 10^{-4} - 1 \times 10^{-5}$ substitutions/nt, e.g., by long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain is capable of performing reverse transcription of a target RNA in vitro. In some embodiments, the reverse transcriptase requires a primer of at least 3 nt to initiate reverse transcription of a template. In some embodiments, reverse transcription of the target RNA is determined by detection of cDNA from the target RNA (e.g., when provided with a ssDNA primer, e.g., which anneals to the target with at least 3, 4, 5, 6, 7, 8, 9, or 10 nt at the 3' end), e.g., as described in Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (incorporated herein by reference in its entirety).

In some embodiments, the reverse transcriptase domain performs reverse transcription at least 5 or 10 times more efficiently (e.g., by cDNA production), e.g., when converting its RNA template to cDNA, for example, as compared to an RNA template lacking the protein binding motif (e.g., a 3' UTR). In embodiments, efficiency of reverse transcription is measured as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain specifically binds a specific RNA template with higher frequency (e.g., about 5 or 10-fold higher frequency) than any endogenous cellular RNA, e.g., when expressed in cells (e.g., HEK293T cells). In embodiments, frequency of specific binding between the reverse transcriptase domain and the template RNA are measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated herein by reference in its entirety).

In some embodiments, a Gene Writer as described herein comprises a polypeptide associated with a guide RNA (gRNA). In certain embodiments, the gRNA is comprised in the template nucleic acid molecule. In other embodiments, the gRNA is separate from the template nucleic acid molecule. In some embodiments wherein the gRNA is comprised in the template nucleic acid molecule, the template nucleic acid molecule further comprises a gRNA spacer sequence (e.g., at or within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides of its 5' end). In embodiments, the gRNA spacer comprises a sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a nucleic acid sequence comprised in the target nucleic acid molecule. In embodiments, the gRNA spacer directs Cas domain (e.g., Cas9) activity at the nucleic acid sequence comprised in the target nucleic acid molecule. In some embodiments wherein the gRNA is comprised in the template nucleic acid molecule, the template nucleic acid molecule further comprises a primer binding site

(e.g., at or within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides of its 3' end). In embodiments, the primer binding site comprises a nucleic acid sequence comprising at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a nucleic acid sequence positioned at the 5' end (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, or 50 nucleotides) of a nick site on the target nucleic acid molecule. In embodiments, binding of the primer binding site to the target nucleic acid molecule operates to prime TPRT.

In some embodiments, the reverse transcriptase (RT) domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety).

Target Site

In some embodiments, after Gene Writing, the target site surrounding the integrated sequence contains a limited number of insertions or deletions, for example, in less than about 50% or10% of integration events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety). In some embodiments, the target site does not show multiple insertion events, e.g., head-to-tail or head-to-head duplications, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (incorporated herein by reference in its entirety). In some embodiments, the target site contains an integrated sequence corresponding to the template RNA.

In some embodiments, the target site does not contain insertions resulting from endogenous RNA in more than about 1% or10% of events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (incorporated herein by reference in its entirety). In some embodiments, the target site contains the integrated sequence corresponding to the template RNA.

In some embodiments, the target site contains an integrated sequence corresponding to the template RNA. In embodiments, the target site does not comprise sequence outside of the RT template (e.g., gRNA scaffold, vector backbone, and/or ITRs), e.g., as determined by long-read amplicon sequencing of the target site (for example, as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020); incorporated herein by reference in its entirety).

Evolved Variants of Gene Writers

In some embodiments, the invention provides evolved variants of Gene Writers. Evolved variants can, in some embodiments, be produced by mutagenizing a reference Gene Writer, or one of the fragments or domains comprised therein. In some embodiments, one or more of the domains (e.g., the reverse transcriptase, DNA binding (including, for example, sequence-guided DNA binding elements), RNA-binding, or endonuclease domain) is evolved. One or more of such evolved variant domains can, in some embodiments, be evolved alone or together with other domains. An evolved variant domain or domains may, in some embodiments, be combined with unevolved cognate component(s) or evolved variants of the cognate component(s), e.g., which may have been evolved in either a parallel or serial manner.

In some embodiments, the process of mutagenizing a reference Gene Writer, or fragment or domain thereof, comprises mutagenizing the reference Gene Writer or fragment or domain thereof. In embodiments, the mutagenesis comprises a continuous evolution method (e.g., PACE) or non-contious evolution method (e.g., PANCE), e.g., as described herein. In some embodiments, the evolved Gene Writer, or a fragment or domain thereof, comprises one or more amino acid variations introduced into its amino acid sequence relative to the amino acid sequence of the reference Gene Writer, or fragment or domain thereof. In embodiments, amino acid sequence variations may include one or more mutated residues (e.g., conservative substitutions, non-conservative substitutions, or a combination thereof) within the amino acid sequence of a reference Gene Writer, e.g., as a result of a change in the nucleotide sequence

encoding the gene writer that results in, e.g., a change in the codon at any particular position in the coding sequence, the deletion of one or more amino acids (e.g., a truncated protein), the insertion of one or more amino acids, or any combination of the foregoing. The evolved variant Gene Writer may include variants in one or more components or domains of the Gene Writer (e.g., variants introduced into a reverse transcriptase domain, endonuclease domain, DNA binding domain, RNA binding domain, or combinations thereof).

In some aspects, the invention provides Gene Writers, systems, kits, and methods using or comprising an evolved variant of a Gene Writer, e.g., employs an evolved variant of a Gene Writer or a Gene Writer produced or produceable by PACE or PANCE. In embodiments, the unevolved reference Gene Writer is a Gene Writer as disclosed herein.

The term "phage-assisted continuous evolution (PACE),"as used herein, generally refers to continuous evolution that employs phage as viral vectors. Examples of PACE technology have been described, for example, in International PCT Application No. PCT/US 2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference.

The term"phage-assisted non-continuous evolution (PANCE)," as used herein, generally refers to non-continuous evolution that employs phage as viral vectors. Examples of PANCE technology have been described, for example, in Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017), incorporated herein by reference in its entirety. Briefly, PANCE is a technique for rapid in vivo directed evolution using serial flask transfers of evolving selection phage (SP), which contain a gene of interest to be evolved, across fresh host cells (e.g., E. coli cells). Genes inside the host cell may be held constant while genes contained in the SP continuously evolve. Following phage growth, an aliquot of infected cells may be used to transfect a subsequent flask

containing host E. coli. This process can be repeated and/or continued until the desired phenotype is evolved, e.g., for as many transfers as desired.

Methods of applying PACE and PANCE to Gene Writers may be readily appreciated by the skilled artisan by reference to, *inter alia*, the foregoing references. Additional exemplary methods for directing continuous evolution of genome-modifying proteins or systems, e.g., in a population of host cells, e.g., using phage particles, can be applied to generate evolved variants of Gene Writers, or fragments or subdomains thereof. Non-limiting examples of such methods are described in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; International Application No. PCT/US2019/37216, filed June 14, 2019, International Patent Publication WO 2019/023680, published January 31, 2019, International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, and International Patent Publication No. PCT/US2019/47996, filed August 23, 2019, each of which is incorporated herein by reference in its entirety.

In some non-limiting illustrative embodiments, a method of evolution of a evolved variant Gene Writer, of a fragment or domain thereof, comprises: (a) contacting a population of host cells with a population of viral vectors comprising the gene of interest (the starting Gene Writer or fragment or domain thereof), wherein: (1) the host cell is amenable to infection by the viral vector; (2) the host cell expresses viral genes required for the generation of viral particles; (3) the expression of at least one viral gene required for the production of an infectious viral particle is dependent on a function of the gene of interest; and/or (4) the viral vector allows for expression of the protein in the host cell, and can be replicated and packaged into a viral particle by the host cell. In some embodiments, the method comprises (b) contacting the host cells with a mutagen, using host cells with mutations that elevate mutation rate (e.g., either by carrying a mutation plasmid or some genome modification—e.g., proofing-impaired DNA polymerase, SOS genes, such as UmuC, UmuD', and/or RecA, which mutations, if plasmid-bound, may be

under control of an inducible promoter), or a combination thereof. In some embodiments, the method comprises (c) incubating the population of host cells under conditions allowing for viral replication and the production of viral particles, wherein host cells are removed from the host cell population, and fresh, uninfected host cells are introduced into the population of host cells, thus replenishing the population of host cells and creating a flow of host cells. In some embodiments, the cells are incubated under conditions allowing for the gene of interest to acquire a mutation. In some embodiments, the method further comprises (d) isolating a mutated version of the viral vector, encoding an evolved gene product (e.g., an evolved variant Gene Writer, or fragment or domain thereof), from the population of host cells.

The skilled artisan will appreciate a variety of features employable within the above-described framework. For example, in some embodiments, the viral vector or the phage is a filamentous phage, for example, an M13 phage, e.g., an M13 selection phage. In certain embodiments, the gene required for the production of infectious viral particles is the M13 gene III (gIII). In embodiments, the phage may lack a functional gIII, but otherwise comprise gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, and a gX. In some embodiments, the generation of infectious VSV particles involves the envelope protein VSV-G. Various embodiments can use different retroviral vectors, for example, Murine Leukemia Virus vectors, or Lentiviral vectors. In embodiments, the retroviral vectors can efficiently be packaged with VSV-G envelope protein, e.g., as a substitute for the native envelope protein of the virus.

In some embodiments, host cells are incubated according to a suitable number of viral life cycles, e.g., at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least. 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive viral life cycles, which in on illustrative and non-limiting examples of M13 phage is 10-20 minutes per virus life cycle. Similarly, conditions can be modulated to adjust the time a host cell remains in a population of host cells, e.g., about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 70, about 80, about 90, about 100, about 120, about 150, or about 180 minutes. Host cell populations can be controlled in part by density of the host cells, or, in some embodiments, the host cell density in

an inflow, e.g., 10^3 cells/ml, about 10^4 cells/ml, about 10^5 cells/ml, about 5- 10^5 cells/ml, about 10^6 cells/ml, about 5- 10^6 cells/ml, about 10^7 cells/ml, about 5- 10^7 cells/ml, about 10^8 cells/ml, about 5- 10^8 cells/ml, about 10^9 cells/ml, about 5· 10^9 cells/ml, about 10^{10} cells/ml, or about 5· 10^{10} cells/ml.

Promoters

In some embodiments, one or more promoter or enhancer elements are operably linked to a nucleic acid encoding a Gene Writer protein or a template nucleic acid, e.g., that controls expression of the heterologous object sequence. In certain embodiments, the one or more promoter or enhancer elements comprise cell-type or tissue specific elements. In some embodiments, the promoter or enhancer is the same or derived from the promoter or enhancer that naturally controls expression of the heterologous object sequence. For example, the ornithine transcarbomylase promoter and enhancer may be used to control expression of the ornithine transcarbomylase gene in a system or method provided by the invention for correcting ornithine transcarbomylase deficiencies. In some embodiments, a promoter for use in the invention is for a gene described in any one of Tables **9-22**, e.g., which may be used with an allele of the reference gene, or, in other embodiments, with a heterologous gene. In some embodiments, the promoter is a promoter of Table 33 or a functional fragment or variant thereof.

Exemplary tissue specific promoters that are commercially available can be found, for example, at a uniform resource locator (e.g., www.invivogen.com/tissue-specific-promoters). In some embodiments, a promoter is a **native promoter** or a minimal promoter, e.g., which consists of a single fragment from the 5' region of a given gene. In some embodiments, a native promoter comprises a core promoter and its natural 5' UTR. In some embodiments, the 5' UTR comprises an intron. In other embodiments, these include **composite promoters, which** combine promoter elements of different origins or were generated by assembling a distal enhancer with a minimal promoter of the same origin.

Exemplary cell or tissue specific promoters are provided in the tables, below, and exemplary nucleic acid sequences encoding them are known in the art and can be readily accessed using a variety of resources, such as the NCBI database, including RefSeq, as well as the Eukaryotic Promoter Database (//epd.epfl.ch//index.php).

Table 33. Exemplary cell or tissue-specific promoters

Promoter	Target cells
B29 Promoter	B cells
CD14 Promoter	Monocytic Cells
CD43 Promoter	Leukocytes and platelets
CD45 Promoter	Hematopoeitic cells
CD68 promoter	macrophages
Desmin promoter	muscle cells
Elastase-1 promoter	pancreatic acinar cells
Endoglin promoter	endothelial cells
fibronectin promoter	differentiating cells, healing tissue
Flt-1 promoter	endothelial cells
GFAP promoter	Astrocytes
GPIIB promoter	megakaryocytes
ICAM-2 Promoter	Endothelial cells
INF-Beta promoter	Hematopoeitic cells
Mb promoter	muscle cells
Nphs1 promoter	podocytes
OG-2 promoter	Osteoblasts, Odonblasts
SP-B promoter	Lung
Syn1 promoter	Neurons
WASP promoter	Hematopoeitic cells
SV40/bAlb promoter	Liver
SV40/bAlb promoter	Liver
SV40/Cd3 promoter	Leukocytes and platelets
SV40/CD45 promoter	hematopoeitic cells
NSE/RU5' promoter	Mature Neurons

Table 34. Additional exemplary cell or tissue-specific promoters

Promoter	**Gene Description**	**Gene Specificity**

APOA2	Apolipoprotein A-II	Hepatocytes (from hepatocyte progenitors)
SERPINA 1 (hAAT)	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 (also named alpha 1 anti-tryps in)	Hepatocytes (from definitive endoderm stage)
CYP3A	Cytochrome P450, family 3, subfamily A, polypeptide	Mature Hepatocytes
MIR122	MicroRNA 122	Hepatocytes (from early stage embryonic liver cells) and endoderm
Pancreatic specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
INS	Insulin	Pancreatic beta cells (from definitive endoderm stage)
IRS2	Insulin receptor substrate 2	Pancreatic beta cells

Pdx1	Pancreatic and duodenal homeobox 1	Pancreas (from definitive endoderm stage)
Alx3	Aristaless-like homeobox 3	Pancreatic beta cells (from definitive endoderm stage)
Ppy	Pancreatic polypeptide	PP pancreatic cells (gamma cells)
Cardiac specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Myh6 (aMHC)	Myosin, heavy chain 6, cardiac muscle, alpha	Late differentiation marker of cardiac muscle cells (atrial specificity)
MYL2 (MLC-2v)	Myosin, light chain 2, regulatory, cardiac, slow	Late differentiation marker of cardiac muscle cells (ventricular specificity)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
NPPA (ANF)	Natriuretic peptide precursor A (also named Atrial Natriuretic Factor)	Atrial specificity in adult cells
Slc8a1 (Ncx1)	Solute carrier family 8 (sodium/calcium exchanger), member 1	Cardiomyocytes from early developmental stages

CNS specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
SYN1 (hSyn)	Synapsin I	Neurons
GFAP	Glial fibrillary acidic protein	Astrocytes
INA	lnternexin neuronal intermediate filament protein, alpha (a-internexin)	Neuroprogenitors
NES	Nestin	Neuroprogenitors and ectoderm
MOBP	Myelin-associated oligodendrocyte basic protein	Oligodendrocytes
MBP	Myelin basic protein	Oligodendrocytes
TH	Tyrosine hydroxylase	Dopaminergic neurons
FOXA2 (HNF3 beta)	Forkhead box A2	Dopaminergic neurons (also used as a marker of endoderm)
Skin specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
FLG	Filaggrin	Keratinocytes from granular layer
K14	Keratin 14	Keratinocytes from granular and basal layers
TGM3	Transglutaminase 3	Keratinocytes from granular layer
Immune cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
ITGAM (CD11B)	lntegrin, alpha M (complement component 3 receptor 3 subunit)	Monocytes, macrophages , granulocytes , natural killer cells
Urogential cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Pbsn	Probasin	Prostatic epithelium
Upk2	Uroplakin 2	Bladder

Sbp	Spermine binding protein	Prostate
Fer1l4	Fer-1-like 4	Bladder
Endothelial cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
ENG	Endoglin	Endothelial cells
Pluripotent and embryonic cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Oct4 (POU5F1)	POU class 5 homeobox 1	Pluripotent cells (germ cells, ES cells, iPS cells)
NANOG	Nanog homeobox	Pluripotent cells (ES cells, iPS cells)
Synthetic Oct4	Synthetic promoter based on a Oct-4 core enhancer element	Pluripotent cells (ES cells, iPS cells)
T brachyury	Brachyury	Mesoderm
NES	Nestin	Neuroprogenitors and Ectoderm
SOX17	SRY (sex determining region Y)-box 17	Endoderm
FOXA2 (HNFJ beta)	Forkhead box A2	Endoderm (also used as a marker of dopaminergic neurons)
MIR122	MicroRNA 122	Endoderm and hepatocytes (from early stage embryonic liver cells~

Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544; incorporated herein by reference in its entirety).

In some embodiments, a nucleic acid encoding a Gene Writer or template nucleic acid is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may, in some embodiment, be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a polypeptide is operably linked to multiple control elements, e.g., that allow expression of the nucleotide sequence encoding the polypeptide in both prokaryotic and eukaryotic cells.

For illustration purposes, examples of spatially restricted promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc. Neuron-specific spatially restricted promoters include, but are not limited to, a neuron-specific enolase (NSE) promoter (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC) promoter, a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al. (1987) Cell 51:7-19; and Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., 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; and Kaneda et al. (1991) Neuron 6:583-594); a GnRH promoter (see, e.g., Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al. (1990) Science 248:223-226); a DNMT promoter (see, e.g., Bartge et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al. (1988) EMBO J. 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca2+-calmodulin-dependent protein kinase II-alpha (CamKIIα) promoter (see, e.g., Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250; and Casanova et al. (2001) Genesis 31:37); a CMV enhancer/platelet-derived growth factor-β promoter (see, e.g., Liu et al. (2004) Gene Therapy 11:52-60); and the like.

Adipocyte-specific spatially restricted promoters include, but are not limited to, the aP2 gene promoter/enhancer, e.g., a region from −5.4 kb to +21 bp of a human aP2 gene (see, e.g., Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590; and Pavjani et al. (2005) Nat. Med. 11:797); a glucose transporter-4 (GLUT4) promoter (see, e.g., Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); a fatty acid translocase (FAT/CD36) promoter (see, e.g., Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; and Sato et al. (2002) J. Biol. Chem. 277:15703); a stearoyl-CoA desaturase-1 (SCD1) promoter (Tabor et al. (1999) J. Biol. Chem. 274:20603); a leptin promoter (see, e.g., Mason et al. (1998) Endocrinol. 139:1013; and Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); an adiponectin promoter (see, e.g., Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; and Chakrabarti (2010) Endocrinol. 151:2408); an adipsin promoter (see, e.g., Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); a resistin promoter (see, e.g., Seo et al. (2003) Molec. Endocrinol. 17:1522); and the like.

Cardiomyocyte-specific spatially restricted promoters include, but are not limited to, control sequences derived from the following genes: myosin light chain-2, α-myosin heavy chain, AE3, cardiac troponin C, cardiac actin, and the like. 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; and Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Smooth muscle-specific spatially restricted promoters include, but are not limited to, an SM22α promoter (see, e.g., Akyürek et al. (2000) Mol. Med. 6:983; and U.S. Pat. No. 7,169,874); a smoothelin promoter (see, e.g., WO 2001/018048); an α-smooth muscle actin promoter; and the like. For example, a 0.4 kb region of the SM22α promoter, within which lie two CArG elements, has been shown to mediate vascular smooth muscle cell-specific expression (see, e.g., Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; and Moessler, et al. (1996) Development 122, 2415-2425).

Photoreceptor-specific spatially restricted promoters include, but are not limited to, a rhodopsin promoter; a rhodopsin kinase promoter (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); a beta phosphodiesterase gene promoter (Nicoud et al. (2007) J. Gene Med. 9:1015); a retinitis pigmentosa gene promoter (Nicoud et al. (2007) supra); an interphotoreceptor retinoid-

binding protein (IRBP) gene enhancer (Nicoud et al. (2007) supra); an IRBP gene promoter (Yokoyama et al. (1992) Exp Eye Res. 55:225); and the like.

Nonlimiting Exemplary Cells-Specific Promoters

Cell-specific promoters known in the art may be used to direct expression of a Gene Writer protein, e.g., as described herein. Nonlimiting exemplary mammalian cell-specific promoters have been characterized and used in mice expressing Cre recombinase in a cell-specific manner. Certain nonlimiting exemplary mammalian cell-specific promoters are listed in Table 1 of US9845481, incorporated herein by reference.

In some embodiments, a cell-specific promoters is a promoter that is active in plants. Many exemplary cell-specific plant promoters are known in the art. See, e.g., U.S. Pat. Nos. 5,097,025; 5,783,393; 5,880,330; 5,981,727; 7,557,264; 6,291,666; 7,132,526; and 7,323,622; and U.S. Publication Nos. 2010/0269226; 2007/0180580; 2005/0034192; and 2005/0086712, which are incorporated by reference herein in their entireties for any purpose.

In some embodiments, a vector as described herein comprises an expression cassette. The term "expression cassette", as used herein, refers to a nucleic acid construct comprising nucleic acid elements sufficient for the expression of the nucleic acid molecule of the instant invention. Typically, an expression cassette comprises the nucleic acid molecule of the instant invention operatively linked to a promoter sequence. The term"operatively linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operatively linked with a coding sequence when it is capable of affecting the expression of that coding sequence (e.g., the coding sequence is under the transcriptional control of the promoter). Encoding sequences can be operatively linked to regulatory sequences in sense or antisense orientation. In certain embodiments, the promoter is a heterologous promoter. The term"heterologous promoter", as used herein, refers to a promoter that is not found to be operatively linked to a given encoding sequence in nature. In certain embodiments, an expression cassette may comprise additional elements, for example, an intron, an enhancer, a polyadenylation site, a woodchuck response element (WRE), and/or other elements known to affect expression levels of the encoding sequenceA"promoter" typically controls the expression of a coding sequence or functional RNA. In certain embodiments, a promoter sequence comprises proximal and more distal upstream

elements and can further comprise an enhancer element. An "enhancer" can typically stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. In certain embodiments, the promoter is derived in its entirety from a native gene. In certain embodiments, the promoter is composed of different elements derived from different naturally occurring promoters. In certain embodiments, the promoter comprises a synthetic nucleotide sequence. It will be understood by those skilled in the art that different promoters will direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions or to the presence or the absence of a drug or transcriptional co-factor. Ubiquitous, cell-type-specific, tissue-specific, developmental stage-specific, and conditional promoters, for example, drug-responsive promoters (e.g ., tetracycline-responsive promoters) are well known to those of skill in the art. Examples of promoter include, but are not limited to, the phosphoglycerate kinase (PKG) promoter, CAG (composite of the CMV enhancer the chicken beta actin promoter (CBA) and the rabbit beta globin intron.), NSE (neuronal specific enolase), synapsin or NeuN promoters, the SV40 early promoter, mouse mammary tumor virus LTR promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), SFFV promoter, rous sarcoma virus (RSV) promoter, synthetic promoters, hybrid promoters, and the like. Other promoters can be of human origin or from other species, including from mice. Common promoters include, e.g., the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, [beta]- actin, rat insulin promoter, the phosphoglycerate kinase promoter, the human alpha- 1 antitrypsin (hAAT) promoter, the transthyretin promoter, the TBG promoter and other liver-specific promoters, the desmin promoter and similar muscle-specific promoters, the EF1 -alpha promoter, the CAG promoter and other constitutive promoters, hybrid promoters with multi-tissue specificity, promoters specific for neurons like synapsin and glyceraldehyde-3 - phosphate dehydrogenase promoter, all of which are promoters well known and readily available to those of skill in the art, can be used to obtain high-level expression of the coding sequence of interest. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, will also find use herein. Such promoter sequences are commercially

available from, e.g., Stratagene (San Diego, CA). Additional exemplary promoter sequences are described, for example, in WO2018213786A1 (incorporated by reference herein in its entirety).

In some embodiments, the apolipoprotein E enhancer (ApoE) or a functional fragment thereof is used, e.g., to drive expression in the liver. In some embodiments, two copies of the ApoE enhancer or a functional fragment thereof is used. In some embodiments, the ApoE enhancer or functional fragment thereof is used in combination with a promoter, e.g., the human alpha-1 antitrypsin (hAAT) promoter.

In some embodiments, the regulatory sequences impart tissue-specific gene expression capabilities. In some cases, the tissue-specific regulatory sequences bind tissue-specific transcription factors that induce transcription in a tissue specific manner. Various tissue-specific regulatory sequences (e.g., promoters, enhancers, etc.) are known in the art. Exemplary tissue-specific regulatory sequences include, but are not limited to, the following tissue-specific promoters: a liver-specific thyroxin binding globulin (TBG) promoter, a insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a mammalian desmin (DES) promoter, a α-myosin heavy chain (a-MHC) promoter, or a cardiac Troponin T (cTnT) promoter. Other exemplary promoters include Beta-actin promoter, hepatitis B virus core promoter, Sandig et al., Gene Ther., 3:1002-9 (1996); alpha-fetoprotein (AFP) promoter, Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), bone osteocalcin promoter (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); bone sialoprotein promoter (Chen et al., J. Bone Miner. Res., 11:654-64 (1996)), CD2 promoter (Hansal et al., J. Immunol., 161:1063-8 (1998); immunoglobulin heavy chain promoter; T cell receptor α-chain promoter, neuronal such as neuron-specific enolase (NSE) promoter (Andersen et al., Cell. Mol. Neurobiol., 13:503-15 (1993)), neurofilament light-chain gene promoter (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)), and the neuron-specific vgf gene promoter (Piccioli et al., Neuron, 15:373-84 (1995)), and others. Additional exemplary promoter sequences are described, for example, in U.S. Patent No. 10300146 (incorporated herein by reference in its entirety). In some embodiments, a tissue-specific regulatory element, e.g., a tissue-specific promoter, is selected from one known to be operably linked to a gene that is highly expressed in a given tissue, e.g., as measured by RNA-seq or protein expression data, or a combination thereof. Methods for analyzing tissue specificity by

expression are taught in Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014), which is incorporated herein by reference in its entirety.

In some embodiments, a vector described herein is a multicistronic expression construct. Multicistronic expression constructs include, for example, constructs harboring a first expression cassette, e.g. comprising a first promoter and a first encoding nucleic acid sequence, and a second expression cassette, e.g. comprising a second promoter and a second encoding nucleic acid sequence. Such multicistronic expression constructs may, in some instances, be particularly useful in the delivery of non-translated gene products, such as hairpin RNAs, together with a polypeptide, for example, a gene writer and gene writer template. In some embodiments, multicistronic expression constructs may exhibit reduced expression levels of one or more of the included transgenes, for example, because of promoter interference or the presence of incompatible nucleic acid elements in close proximity. If a multicistronic expression construct is part of a viral vector, the presence of a self-complementary nucleic acid sequence may, in some instances, interfere with the formation of structures necessary for viral reproduction or packaging.

In some embodiments, the sequence encodes an RNA with a hairpin. In some embodiments, the hairpin RNA is an a guide RNA, a template RNA, shRNA, or a microRNA. In some embodiments, the first promoter is an RNA polymerase I promoter. In some embodiments, the first promoter is an RNA polymerase II promoter. In some embodiments, the second promoter is an RNA polymerase III promoter. In some embodiments, the second promoter is a U6 or H1 promoter. In some embodiments, the nucleic acid construct comprises the structure of AAV construct B1 or B2.

Without wishing to be bound by theory, multicistronic expression constructs may not achieve optimal expression levels as compared to expression systems containing only one cistron. One of the suggested causes of lower expression levels achieved with multicistronic expression constructs comprising two ore more promoter elements is the phenomenon of promoter interference (see, e.g., 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; and 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; both references incorporated herein by reference for disclosure of promoter interference phenomenon). In some embodiments, the problem of promoter interference may be overcome, e.g., by producing multicistronic expression constructs comprising only one promoter driving transcription of multiple encoding nucleic acid sequences separated by internal ribosomal entry sites, or by separating cistrons comprising their own promoter with transcriptional insulator elements. In some embodiments, single-promoter driven expression of multiple cistrons may result in uneven expression levels of the cistrons. In some embodiments, a promoter cannot efficiently be isolated and isolation elements may not be compatible with some gene transfer vectors, for example, some retroviral vectors.

MicroRNAs

miRNAs and other small interfering nucleic acids generally regulate gene expression via target RNA transcript cleavage/degradation or translational repression of the target messenger RNA (mRNA). miRNAs may, in some instances, be natively expressed, typically as final 19-25 non-translated RNA products. miRNAs generally exhibit their activity through sequence-specific interactions with the 3′ untranslated regions (UTR) of target mRNAs. These endogenously expressed miRNAs may form hairpin precursors that are subsequently processed into an miRNA duplex, and further into a mature single stranded miRNA molecule. This mature miRNA generally guides a multiprotein complex, miRISC, which identifies target 3′ UTR regions of target mRNAs based upon their complementarity to the mature miRNA. Useful transgene products may include, for example, miRNAs or miRNA binding sites that regulate the expression of a linked polypeptide. A non-limiting list of miRNA genes; the products of these genes and their homologues are useful as transgenes or as targets for small interfering nucleic acids (e.g., miRNA sponges, antisense oligonucleotides), e.g., in methods such as those listed in US10300146, 22:25-25:48, incorporated by reference. In some embodiments, one or more binding sites for one or more of the foregoing miRNAs are incorporated in a transgene, e.g., a transgene delivered by a rAAV vector, e.g., to inhibit the expression of the transgene in one or more tissues of an animal harboring the transgene. In some embodiments, a binding site may be selected to control the expression of a trangene in a tissue specific manner. For example, binding sites for the liver-specific miR-122 may be incorporated into a transgene to inhibit expression of

that transgene in the liver. Additional exemplary miRNA sequences are described, for example, in U.S. Patent No. 10300146 (incorporated herein by reference in its entirety).

For liver-specific Gene Writing, however, overexpression of miR-122 may be utilized instead of using binding sites to effect miR-122-specific degradation. This miRNA is positively associated with hepatic differentiation and maturation, as well as enhanced expression of liver specific genes. Thus, in some embodiments, the coding sequence for miR-122 may be added to a component of a Gene Writing system to enhance a liver-directed therapy.

A miR inhibitor or miRNA inhibitor is generally an agent that blocks miRNA expression and/or processing. Examples of such agents include, but are not limited to, microRNA antagonists, microRNA specific antisense, microRNA sponges, and microRNA oligonucleotides (double-stranded, hairpin, short oligonucleotides) that inhibit miRNA interaction with a Drosha complex. MicroRNA inhibitors, e.g., miRNA sponges, can be expressed in cells from transgenes (e.g., as described in Ebert, M. S. Nature Methods, Epub Aug. 12, 2007; incorporated by reference herein in its entirety). In some embodiments, microRNA sponges, or other miR inhibitors, are used with the AAVs. microRNA sponges generally specifically inhibit miRNAs through a complementary heptameric seed sequence. In some embodiments, an entire family of miRNAs can be silenced using a single sponge sequence. Other methods for silencing miRNA function (derepression of miRNA targets) in cells will be apparent to one of ordinary skill in the art.

In some embodiments, a miRNA as described herein comprises a sequence listed in Table 4 of PCT Publication No. WO2020014209, incorporated herein by reference. Also incorporated herein by reference are the listing of exemplary miRNA sequences from WO2020014209.

In some embodiments, it is advantageous to silence one or more components of a Gene Writing system (e.g., mRNA encoding a Gene Writer polypeptide, a Gene Writer Template RNA, or a heterologous object sequence expressed from the genome after successful Gene Writing) in a portion of cells. In some embodiments, it is advantageous to restrict expression of a component of a Gene Writing system to select cell types within a tissue of interest.

For example, it is known that in a given tissue, e.g., liver, macrophages and immune cells, e.g., Kupffer cells in the liver, may engage in uptake of a delivery vehicle for one or more components of a Gene Writing system. In some embodiments, at least one binding site for at

least one miRNA highly expressed in macrophages and immune cells, e.g., Kupffer cells, is included in at least one component of a Gene Writing system, e.g., nucleic acid encoding a Gene Writing polypeptide or a transgene. In some embodiments, a miRNA that targets the one or more binding sites is listed in a table referenced herein, e.g., miR-142, e.g., mature miRNA hsa-miR-142-5p or hsa-miR-142-3p.

In some embodiments, there may be a benefit to decreasing Gene Writer levels and/or Gene Writer activity in cells in which Gene Writer expression or overexpression of a transgene may have a toxic effect. For example, it has been shown that delivery of a transgene overexpression cassette to dorsal root ganglion neurons may result in toxicity of a gene therapy (see Hordeaux et al Sci Transl Med 12(569):eaba9188 (2020), incorporated herein by reference in its entirety). In some embodiments, at least one miRNA binding site may be incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron, e.g., a dorsal root ganglion neuron. In some embodiments, the at least one miRNA binding site incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron is a binding site of miR-182, e.g., mature miRNA hsa-miR-182-5p or hsa-miR-182-3p. In some embodiments, the at least one miRNA binding site incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron is a binding site of miR-183, e.g., mature miRNA hsa-miR-183-5p or hsa-miR-183-3p. In some embodiments, combinations of miRNA binding sites may be used to enhance the restriction of expression of one or more components of a Gene Writing system to a tissue or cell type of interest.

Table A5 below provides exemplary miRNAs and corresponding expressing cells, e.g., a miRNA for which one can, in some embodiments, incorporate binding sites (complementary sequences) in the transgene or polypeptide nucleic acid, e.g., to decrease expression in that off-target cell.

Table A5: Exemplary miRNA from off-target cells and tissues

Silenced cell type	miRNA name	Mature miRNA	miRNA sequence	SEQ ID NO
Kupffer cells	miR-142	hsa-miR-142-5p	cauaaaguagaaagcacuacu	481
Kupffer cells	miR-142	hsa-miR-142-3p	uguaguguuuccuacuuuaugga	482

Dorsal root ganglion neurons	miR-182	hsa-miR-182-5p	uuuggcaaugguagaacucacacu	483
Dorsal root ganglion neurons	miR-182	hsa-miR-182-3p	ugguucuagacuugccaacua	484
Dorsal root ganglion neurons	miR-183	hsa-miR-183-5p	uauggcacugguagaauucacu	485
Dorsal root ganglion neurons	miR-183	hsa-miR-183-3p	gugaauuaccgaagggccauaa	486
Hepatocytes	miR-122	hsa-miR-122-5p	uggagugugacaaugguguuug	
Hepatocytes	miR-122	hsa-miR-122-3p	aacgccauuaucacacuaaaua	

Anticrispr systems for regulating GeneWriter activity

Various approaches for modulating Cas molecule activity may be used in conjunction with the systems and methods described herein. For instance, in some embodiments, a polypeptide described herein (e.g., a Cas molecule or a GeneWriter comprising a Cas domain) can be regulated using an anticrispr agent (e.g., an anticrispr protein or anticrispr small molecule). In some embodiments, the Cas molecule or Cas domain comprises a responsive intein such as, for example, a 4-hydroxytamoxifen (4-HT)-responsive intein, an iCas molecule (e.g., iCas9); a 4-HT-responsive Cas (e.g., allosterically regulated Cas9 (arC9) or dead Cas9 (dC9)). The systems and methods described herein can also utilize a chemically-induced dimerization system of split protein fragments (e.g., rapamycin-mediated dimerization of FK506 binding protein 12 (FKBP) and FKBP rapamycin binding domain (FRB), an abscisic acid-inducible ABI-PYL1 and gibberellin-inducible GID1-GAI heterodimerization domains); a dimer of BCL-xL peptide and BH3 peptides, a A385358 (A3) small molecule, a degron system (e.g., a FKBP-Cas9 destabilized system, an auxin-inducible degron (AID) or an E.coliDHFR degron system), an aptamer or aptazyme fused with gRNA (e.g., tetracycline- and theophylline-responsive bioswitches), AcrIIA2 and AcrIIA4 proteins, and BRD0539.

In some embodiments, a small molecule-responsive intein (e.g., 4-hydroxytamoxifen (4-HT)-responsive intein) is inserted at specific sites within a Cas molecule (e.g., Cas9). In some embodiments, the insertion of a 4HT-responsive intein disrupts Cas9 enzymatic activity. In some embodiments, a Cas molecule (e.g., iCas9) is fused to the hormone binding domain of the estrogen receptor (ERT2). In some embodiments, the ligand binding domain of the human

estrogen receptor-α can be inserted into a Cas molecule (e.g., Cas9 or dead Cas9 (dC9)), e.g., at position 231, yielding a 4HT-responsive anticrispr Cas9 (e.g., arC9 or dC9). In some embodiments, dCas9 can provide 4-HT dose-dependent repression of Cas9 function. In some embodiments, arC9 can provide 4-HT dose-dependent control of Cas9 function. In some embodiments, a Cas molecule (e.g., Cas9) is fused to split protein fragments. In some embodiments, chemically-induced dimerization of split protein fragments (e.g., rapamycin-mediated dimerization of FK506 binding protein 12 (FKBP) and FKBP rapamycin binding domain (FRB)) can induce low levels of Cas9 molecule activity. In some embodiments, a chemically-induced dimerization system (e.g., abscisic acid-inducible ABI-PYL1 and gibberellin-inducible GID1-GAI heterodimerization domains) can induce a dose-dependent and reversible transcriptional activation/repression of Cas9. In some embodiments, a Cas9 inducible system (ciCas9) comprises the replacement of a Cas molecule (e.g., Cas9) REC2 domain with a BCL-xl peptide and attachment of a BH3 peptide to the N- and C-termini of the modified Cas9.BCL. In some embodiments, the interaction between BCL-xL and BH3 peptides can keep Cas9 in an inactive state. In some embodiments, a small molecule (e.g., A-385358 (A3)) can disrupt the interaction between BLC-xl and BH3 peptides to activate Cas9. In some embodiments, a Cas9 inducible system can exhibit dose-dependent control of nuclease activity. In some embodiments, a degron system can induce degradation of a Cas molecule (e.g., Cas9) upon activation or deactivation by an external factor (e.g., small-molecule ligand, light, temperature, or a protein). In some embodiments, a small molecule BRD0539 inhibits a Cas molecule (e.g., Cas9) reversibly. Additional information on anticrispr proteins or anticrispr small molecules can be found, for example, in 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, and Pawluk Anti-CRISPR: discovery, mechanism and function Nature Reviews Microbiology volume 16, pages12–17(2018), each of which is incorporated by reference in its entirety.

Self-inactivating modules for regulating GeneWriter activity

In some embodiments the Gene Writer systems described herein includes a self-inactivating module. The self-inactivating module leads to a decrease of expression of the Gene Writer polypeptide, the Gene Writer template, or both. Without wishing to be bound by the

theory, the self-inactivating module provides for a temporary period of Gene Writer expression prior to inactivation. Without wishing to be bound by theory, the activity of the Gene Writer polypeptide at a target site introduces a mutation (e.g. a substitution, insertion, or deletion) into the DNA encoding the Gene Writer polypeptide or Gene Writer template which results in a decrease of Gene Writer polypeptide or template expression. In some embodiments of the self-inactivating module, a target site for the Gene Writer polypeptide is included in the DNA encoding the Gene Writer polypeptide or Gene Writer template. In some embodiments, one, two, three, four, five, or more copies of the target site are included in the DNA encoding the Gene Writer polypeptide or Gene Writer template. In some embodiments, the target site in the DNA encoding the Gene Writer polypeptide or Gene Writer template is the same target site as the target site on the genome. In some embodiments, the target site is a different target site than the target site on the genome. In some embodiments, the self-inactivation module target site uses the same or a different template RNA or guide RNA as the genome target site. In some embodiments, the target site is modified via target primed reverse transcription based upon a template RNA. In some embodiments the target side is nicked. The target site may be incorporated into an enhancer, a promoter, an untranslated region, an exon, an intron, an open reading frame, or a stuffer sequence.

In some embodiments, upon inactivation, the decrease of expression is 25%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99.9%, or more lower than a Gene Writing system that does not contain the self-inactivating module. In some embodiments, a Gene Writer system that contains the self-inactivating module has a 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 99%, or higher rate of integrations in target sites than off-target sites compared to a Gene Writing system that does not contain the self-inactivation module. a Gene Writer system that contains the self-inactivating module has a 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 99%, or higher efficiency of target site modification compared to a Gene Writing system that does not contain the self-inactivation module. In some embodiments, the self-inactivating module is included when the Gene Writer polypeptide is delivered as DNA, e.g. via a viral vector.

Self-inactivating modules have been described for nucleases. See, e.g. in 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 (available at www.editasmedicine.com/wp-content/uploads/2019/10/aef_asgct_poster_2017_final_-_present_5-11-17_515pm1_1494537387_1494558495_1497467403.pdf), and Epstein and Schaffer Engineering a Self-Inactivating CRISPR System for AAV Vectors Targeted Genome Editing IlVolume 24, SUPPLEMENT 1, S50, May 01, 2016, *and* WO2018106693A1.

Small Molecules

In some embodiments a polypeptide described herein (e.g., a Gene Writer polypeptide) is controllable via a small molecule. In some embodiments the polypeptide is dimerized via a small molecule.

In some embodiment, the polypeptide is controllable via Chemical Induction of Dimerization (CID) with small molecules. CID is generally used to generate switches of protein function to alter cell physiology. An exemplary high specificity, efficient dimerizer is rimiducid (AP1903), which has two identical, protein-binding surfaces arranged tail-to-tail, each with high affinity and specificity for a mutant of FKBP12: FKBP12(F36V) (FKBP12v36, F_{V36} or F_V). Attachment of one or more F_V domains onto one or more cell signaling molecules that normally rely on homodimerization can convert that protein to rimiducid control. Homodimerization with rimiducid is used in the context of an inducible caspase safety switch. This molecular switch that is controlled by a distinct dimerizer ligand, based on the heterodimerizing small molecule, rapamycin, or rapamycin analogs ("rapalogs"). Rapamycin binds to FKBP12, and its variants, and can induce heterodimerization of signaling domains that are fused to FKBP12 by binding to both FKBP12 and to polypeptides that contain the FKBP-rapamycin-binding (FRB) domain of mTOR. Provided in some embodiments of the present application are molecular switches that greatly augment the use of rapamycin, rapalogs and rimiducid as agents for therapeutic applications.

In some embodiments of the dual switch technology, a homodimerizer, such as AP1903 (rimiducid), directly induces dimerization or multimerization of polypeptides comprising an FKBP12 multimerizing region. In other embodiments, a polypeptide comprising an FKBP12 multimerization is multimerized, or aggregated by binding to a heterodimerizer, such as rapamycin or a rapalog, which also binds to an FRB or FRB variant multimerizing region on a chimeric polypeptide, also expressed in the modified cell, such as, for example, a chimeric

antigen receptor. Rapamycin is a natural product macrolide that binds with high affinity (<1 nM) to FKBP12 and together initiates the high-affinity, inhibitory interaction with the FKBP-Rapamycin-Binding (FRB) domain of mTOR. FRB is small (89 amino acids) and can thereby be used as a protein "tag" or "handle" when appended to many proteins. Coexpression of a FRB-fused protein with a FKBP12-fused protein renders their approximation rapamycin-inducible (12-16). This can serve as the basis for a cell safety switch regulated by the orally available ligand, rapamycin, or derivatives of rapamycin (rapalogs) that do not inhibit mTOR at a low, therapeutic dose but instead bind with selected, Caspase-9-fused mutant FRB domains. (see 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; and Choi J, Science. 1996; 273(5272):239-42).

In some embodiments, two levels of control are provided in the therapeutic cells. In embodiments, the first level of control may be tunable, i.e., the level of removal of the therapeutic cells may be controlled so that it results in partial removal of the therapeutic cells. In some embodiments, the chimeric antigen polypeptide comprises a binding site for rapamycin, or a rapamycin analog. In embodiments, also present in the therapeutic cell is a suicide gene, such as, for example, one encoding a caspase polypeptide. Using this controllable first level, the need for continued therapy may, in some embodiments, be balanced with the need to eliminate or reduce the level of negative side effects. In some embodiments, a rapamycin analog, a rapalog is administered to the patient, which then binds to both the caspase polypeptide and the chimeric antigen receptor, thus recruiting the caspase polypeptide to the location of the CAR, and aggregating the caspase polypeptide. Upon aggregation, the caspase polypeptide induces apoptosis. The amount of rapamycin or rapamycin analog administered to the patient may vary; if the removal of a lower level of cells by apoptosis is desired in order to reduce side effects and continue CAR therapy, a lower level of rapamycin or rapamycin may be administered to the patient. In some embodiments, the second level of control may be designed to achieve the maximum level of cell elimination. This second level may be based, for example, on the use of rimiducid, or AP1903. If there is a need to rapidly eliminate up to 100% of the therapeutic cells, the AP1903 may be administered to the patient. The multimeric AP1903 binds to the caspase polypeptide, leading to multimerization of the caspase polypeptide and apoptosis. In certain

examples, second level may also be tunable, or controlled, by the level of AP1903 administered to the subject.

In certain embodiments, small molecules can be used to control genes, as described in for example, US10584351 at 47:53-56:47 (incorporated by reference herein in its entirety), together suitable ligands for the control features, e.g., in US10584351 at 56:48, et seq. as well as U10046049 at 43:27-52:20, incorporated by reference as well as the description of ligands for such control systems at 52:21, et seq.

Chemically modified nucleic acids and nucleic acid end features

A nucleic acid described herein (e.g., a template nucleic acid, e.g., a template RNA; or a nucleic acid (e.g., mRNA) encoding a GeneWriter) can comprise unmodified or modified nucleobases. Naturally occurring RNAs are synthesized from four basic ribonucleotides: ATP, CTP, UTP and GTP, but may contain post-transcriptionally modified nucleotides. Further, approximately one hundred different nucleoside modifications have been identified in RNA (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). An RNA can also comprise wholly synthetic nucleotides that do not occur in nature.

In some embodiments, the chemically modification is one provided in PCT/US2016/032454, US Pat. Pub. No. 20090286852, of International Application No. 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, or WO/2016/077123, each of which is herein incorporated by reference in its entirety. It is understood that incorporation of a chemically modified nucleotide into a polynucleotide can result in the modification being incorporated into a nucleobase, the backbone, or both, depending on the location of the modification in the nucleotide. In some embodiments, the backbone modification is one provided in EP 2813570, which is herein incorporated by reference in its entirety. In some embodiments, the modified cap is one provided in US Pat. Pub. No. 20050287539, which is herein incorporated by reference in its entirety.

In some embodiments, the chemically modified nucleic acid (e.g., RNA, e.g., mRNA) comprises one or more of ARCA: anti-reverse cap analog (m27.3'-OGP3G), GP3G (Unmethylated Cap Analog), m7GP3G (Monomethylated Cap Analog), m32.2.7GP3G (Trimethylated Cap Analog), m5CTP (5'-methyl-cytidine triphosphate), m6ATP (N6-methyl-adenosine-5'-triphosphate), s2UTP (2-thio-uridine triphosphate), and Ψ (pseudouridine triphosphate).

In some embodiments, the chemically modified nucleic acid comprises a 5' cap, e.g.: a 7-methylguanosine cap (e.g., a O-Me-m7G cap); a hypermethylated cap analog; an NAD+-derived cap analog (e.g., as described in Kiledjian, Trends in Cell Biology 28, 454-464 (2018)); or a modified, e.g., biotinylated, cap analog (e.g., as described in Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)).

In some embodiments, the chemically modified nucleic acid comprises a 3' feature selected from one or more of: a polyA tail; a 16-nucleotide long stem-loop structure flanked by unpaired 5 nucleotides (e.g., as described by Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)); a triple-helical structure (e.g., as described by Brown et al., PNAS 109, 19202-19207 (2012)); a tRNA, Y RNA, or vault RNA structure (e.g., as described by Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)); incorporation of one or more deoxyribonucleotide triphosphates (dNTPs), 2'O-Methylated NTPs, or phosphorothioate-NTPs; a single nucleotide chemical modification (e.g., oxidation of the 3' terminal ribose to a reactive aldehyde followed by conjugation of the aldehyde-reactive modified nucleotide); or chemical ligation to another nucleic acid molecule.

In some embodiments, the the nucleic acid (e.g., template nucleic acid) comprises one or more modified nucleotides, e.g., selected from dihydrouridine, inosine, 7-methylguanosine, 5-

methylcytidine (5mC), 5′ Phosphate ribothymidine, 2′-O-methyl ribothymidine, 2′-O-ethyl ribothymidine, 2′-fluoro ribothymidine, C-5 propynyl-deoxycytidine (pdC), C-5 propynyl-deoxyuridine (pdU), C-5 propynyl-cytidine (pC), C-5 propynyl-uridine (pU), 5-methyl cytidine, 5-methyl uridine, 5-methyl deoxycytidine, 5-methyl deoxyuridine methoxy, 2,6-diaminopurine, 5′-Dimethoxytrityl-N4-ethyl-2′-deoxycytidine, C-5 propynyl-f-cytidine (pfC), C-5 propynyl-f-uridine (pfU), 5-methyl f-cytidine, 5-methyl f-uridine, C-5 propynyl-m-cytidine (pmC), C-5 propynyl-f-uridine (pmU), 5-methyl m-cytidine, 5-methyl m-uridine, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), or 5-methoxyuridine (5-MO-U).

In some embodiments, the nucleic acid comprises a backbone modification, e.g., a modification to a sugar or phosphate group in the backbone. In some embodiments, the nucleic acid comprises a nucleobase modification.

In some embodiments, the nucleic acid comprises one or more chemically modified nucleotides of Table 6, one or more chemical backbone modifications of Table 7, one or more chemically modified caps of Table 8. For instance, in some embodiments, the nucleic acid comprises two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of chemical modifications. As an example, the nucleic acid may comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of modified nucleobases, e.g., as described herein, e.g., in Table 6. Alternatively or in combination, the nucleic acid may comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of backbone modifications, e.g., as described herein, e.g., in Table 7. Alternatively or in combination, the nucleic acid may comprise one or more modified cap, e.g., as described herein, e.g., in Table 8. For instance, in some embodiments, the nucleic acid comprises one or more type of modified nucleobase and one or more type of backbone modification; one or more type of modified nucleobase and one or more modified cap; one or more type of modified cap and one or more type of backbone modification; or one or more type of modified nucleobase, one or more type of backbone modification, and one or more type of modified cap.

In some embodiments, the nucleic acid comprises one or more (e.g., 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, or more) modified nucleobases. In some embodiments, all nucleobases of the nucleic acid are modified. In some embodiments, the nucleic acid is modified at one or more (e.g., 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, or more) positions in the backbone. In some embodiments, all backbone positions of the nucleic acid are modified.

Table 6. Modified nucleotides

5-aza-uridine	N2-methyl-6-thio-guanosine
2-thio-5-aza-midine	N2,N2-dimethyl-6-thio-guanosine
2-thiouridine	pyridin-4-one ribonucleoside
4-thio-pseudouridine	2-thio-5-aza-uridine
2-thio-pseudouridine	2-thiomidine
5-hydroxyuridine	4-thio-pseudomidine
3-methyluridine	2-thio-pseudowidine
5-carboxymethyl-uridine	3-methylmidine
1-carboxymethyl-pseudouridine	1-propynyl-pseudomidine
5-propynyl-uridine	1-methyl-1-deaza-pseudomidine
1-propynyl-pseudouridine	2-thio-1-methyl-1-deaza-pseudouridine
5-taurinomethyluridine	4-methoxy-pseudomidine
1-taurinomethyl-pseudouridine	5'-O-(1-Thiophosphate)-Adenosine
5-taurinomethyl-2-thio-uridine	5'-O-(1-Thiophosphate)-Cytidine
1-taurinomethyl-4-thio-uridine	5'-O-(1-thiophosphate)-Guanosine
5-methyl-uridine	5'-O-(1-Thiophophate)-Uridine
1-methyl-pseudouridine	5'-O-(1-Thiophosphate)-Pseudouridine
4-thio-1-methyl-pseudouridine	2'-O-methyl-Adenosine
2-thio-1-methyl-pseudouridine	2'-O-methyl-Cytidine
1-methyl-1-deaza-pseudouridine	2'-O-methyl-Guanosine
2-thio-1-methyl-1-deaza-pseudomidine	2'-O-methyl-Uridine
dihydrouridine	2'-O-methyl-Pseudouridine
dihydropseudouridine	2'-O-methyl-Inosine
2-thio-dihydromidine	2-methyladenosine
2-thio-dihydropseudouridine	2-methylthio-N6-methyladenosine
2-methoxyuridine	2-methylthio-N6 isopentenyladenosine
2-methoxy-4-thio-uridine	2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine
4-methoxy-pseudouridine	N6-methyl-N6-threonylcarbamoyladenosine
4-methoxy-2-thio-pseudouridine	N6-hydroxynorvalylcarbamoyladenosine
5-aza-cytidine	2-methylthio-N6-hydroxynorvalyl carbamoyladenosine
pseudoisocytidine	2'-O-ribosyladenosine (phosphate)
3-methyl-cytidine	1,2'-O-dimethylinosine
N4-acetylcytidine	5,2'-O-dimethylcytidine
5-formylcytidine	N4-acetyl-2'-O-methylcytidine
N4-methylcytidine	Lysidine
5-hydroxymethylcytidine	7-methylguanosine
1-methyl-pseudoisocytidine	N2,2'-O-dimethylguanosine
pyrrolo-cytidine	
pyrrolo-pseudoisocytidine	

2-thio-cytidine	N2,N2,2'-O-trimethylguanosine
2-thio-5-methyl-cytidine	2'-O-ribosylguanosine (phosphate)
4-thio-pseudoisocytidine	Wybutosine
4-thio-1-methyl-pseudoisocytidine	Peroxywybutosine
4-thio-1-methyl-1-deaza-pseudoisocytidine	Hydroxywybutosine
1-methyl-1-deaza-pseudoisocytidine	undermodified hydroxywybutosine
zebularine	methylwyosine
5-aza-zebularine	queuosine
5-methyl-zebularine	epoxyqueuosine
5-aza-2-thio-zebularine	galactosyl-queuosine
2-thio-zebularine	mannosyl-queuosine
2-methoxy-cytidine	7-cyano-7-deazaguanosine
2-methoxy-5-methyl-cytidine	7-aminomethyl-7-deazaguanosine
4-methoxy-pseudoisocytidine	archaeosine
4-methoxy-1-methyl-pseudoisocytidine	5,2'-O-dimethyluridine
2-aminopurine	4-thiouridine
2,6-diaminopurine	5-methyl-2-thiouridine
7-deaza-adenine	2-thio-2'-O-methyluridine
7-deaza-8-aza-adenine	3-(3-amino-3-carboxypropyl)uridine
7-deaza-2-aminopurine	5-methoxyuridine
7-deaza-8-aza-2-aminopurine	uridine 5-oxyacetic acid
7-deaza-2,6- diaminopurine	uridine 5-oxyacetic acid methyl ester
7-deaza-8-aza-2,6-diarninopurine	5-(carboxyhydroxymethyl)uridine)
1-methyladenosine	5-(carboxyhydroxymethyl)uridine methyl ester
N6-isopentenyladenosine	5-methoxycarbonylmethyluridine
N6-(cis-hydroxyisopentenyl)adenosine	5-methoxycarbonylmethyl-2'-O-methyluridine
2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine	5-methoxycarbonylmethyl-2-thiouridine
	5-aminomethyl-2-thiouridine
N6-glycinylcarbamoyladenosine	5-methylaminomethyluridine
N6-threonylcarbamoyladenosine	5-methylaminomethyl-2-thiouridine
2-methylthio-N6-threonyl carbamoyladenosine	5-methylaminomethyl-2-selenouridine
	5-carbamoylmethyluridine
N6,N6-dimethyladenosine	5-carbamoylmethyl-2'-O-methyluridine
7-methyladenine	5-carboxymethylaminomethyluridine
2-methylthio-adenine	5-carboxymethylaminomethyl-2'-O-methyluridine
2-methoxy-adenine	
inosine	5-carboxymethylaminomethyl-2-thiouridine
1-methyl-inosine	N4,2'-O-dimethylcytidine
wyosine	5-carboxymethyluridine
wybutosine	N6,2'-O-dimethyladenosine
7-deaza-guanosine	N,N6,O-2'-trimethyladenosine
7-deaza-8-aza-guanosine	N2,7-dimethylguanosine
6-thio-guanosine	N2,N2,7-trimethylguanosine
6-thio-7-deaza-guanosine	3,2'-O-dimethyluridine
6-thio-7-deaza-8-aza-guanosine	5-methyldihydrouridine
7-methyl-guanosine	5-formyl-2'-O-methylcytidine

6-thio-7-methyl-guanosine	1,2'-O-dimethylguanosine
7-methylinosine	4-demethylwyosine
6-methoxy-guanosine	Isowyosine
1-methylguanosine	N6-acetyladenosine
N2-methylguanosine	
N2,N2-dimethylguanosine	
8-oxo-guanosine	
7-methyl-8-oxo-guanosine	
1-methyl-6-thio-guanosine	

Table 7. Backbone modifications

2'-O-Methyl backbone
Peptide Nucleic Acid (PNA) backbone
phosphorothioate backbone
morpholino backbone
carbamate backbone
siloxane backbone
sulfide backbone
sulfoxide backbone
sulfone backbone
formacetyl backbone
thioformacetyl backbone
methyleneformacetyl backbone
riboacetyl backbone
alkene containing backbone
sulfamate backbone
sulfonate backbone
sulfonamide backbone
methyleneimino backbone
methylenehydrazino backbone
amide backbone

Table 8. Modified caps

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m72'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG
m7GppppG

m7GppppA
m7GppppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

Production of Compositions and Systems

As will be appreciated by one of skill, methods of designing and constructing nucleic acid constructs and proteins or polypeptides (such as the systems, constructs and polypeptides described herein) are routine in the art. Generally, recombinant methods may be used. See, in general, Smales & James (Eds.), *Therapeutic Proteins: Methods and Protocols* (Methods in Molecular Biology), Humana Press (2005); and Crommelin, Sindelar & Meibohm (Eds.), *Pharmaceutical Biotechnology: Fundamentals and Applications*, Springer (2013). Methods of designing, preparing, evaluating, purifying and manipulating nucleic acid compositions are described in Green and Sambrook (Eds.), *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

The disclosure provides, in part, a nucleic acid, e.g., vector, encoding a Gene Writer polypeptide described herein, a template nucleic acid described herein, or both. In some embodiments, a vector comprises a selective marker, e.g., an antibiotic resistance marker. In some embodiments, the antibiotic resistance marker is a kanamycin resistance marker. In some embodiments, the antibiotic resistance marker does not confer resistance to beta-lactam antibiotics. In some embodiments, the vector does not comprise an ampicillin resistance marker. In some embodiments, the vector comprises a kanamycin resistance marker and does not comprise an ampicillin resistance marker. In some embodiments, a vector encoding a Gene Writer polypeptide is integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a Gene Writer polypeptide is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a template nucleic acid (e.g.,

template RNA) is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, if a vector is integrated into a target site in a target cell genome, the selective marker is not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, genes or sequences involved in vector maintenance (e.g., plasmid maintenance genes) are not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, transfer regulating sequences (e.g., inverted terminal repeats, e.g., from an AAV) are not integrated into the genome. In some embodiments, administration of a vector (e.g., encoding a Gene Writer polypeptide described herein, a template nucleic acid described herein, or both) to a target cell, tissue, organ, or subject results in integration of a portion of the vector into one or more target sites in the genome(s) of said target cell, tissue, organ, or subject. In some embodiments, less than 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, or 1% of target sites (e.g., no target sites) comprising integrated material comprise a selective marker (e.g., an antibiotic resistance gene), a transfer regulating sequence (e.g., an inverted terminal repeat, e.g., from an AAV), or both from the vector.

Exemplary methods for producing a therapeutic pharmaceutical protein or polypeptide described herein involve expression in mammalian cells, although recombinant proteins can also be produced using insect cells, yeast, bacteria, or other cells under control of appropriate promoters. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter, and other 5' or 3' flanking non-transcribed sequences, and 5' or 3' non-translated sequences such as necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, and termination sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, splice, and polyadenylation sites may be used to provide other genetic elements required for expression of a heterologous DNA sequence. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described in Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

Various mammalian cell culture systems can be employed to express and manufacture recombinant protein. Examples of mammalian expression systems include CHO, COS, HEK293, HeLA, and BHK cell lines. Processes of host cell culture for production of protein therapeutics are described in Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for*

Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology), Springer (2014). Compositions described herein may include a vector, such as a viral vector, e.g., a lentiviral vector, encoding a recombinant protein. In some embodiments, a vector, e.g., a viral vector, may comprise a nucleic acid encoding a recombinant protein.

Purification of protein therapeutics is described in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

In some embodiments, quality standards include, but are not limited to:

(i) the length of mRNA encoding the GeneWriter polypeptide, e.g., whether the mRNA has a length that is above a reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present is greater than 3000, 4000, or 5000 nucleotides long;

(ii) the presence, absence, and/or length of a polyA tail on the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains a polyA tail (e.g., a polyA tail that is at least 5, 10, 20, 30, 50, 70, 100 nucleotides in length);

(iii) the presence, absence, and/or type of a 5' cap on the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains a 5' cap, e.g., whether that cap is a 7-methylguanosine cap, e.g., a O-Me-m7G cap;

(iv) the presence, absence, and/or type of one or more modified nucleotides (e.g., selected from pseudouridine, dihydrouridine, inosine, 7-methylguanosine, 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), 5-methylcytidine (5mC), or a locked nucleotide) in the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains one or more modified nucleotides;

(v) the stability of the mRNA (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA remains intact (e.g., greater than 100, 125, 150, 175, or 200 nucleotides long) after a stability test; or

(vi) the potency of the mRNA in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the mRNA is assayed for potency.

Kits, Articles of Manufacture, and Pharmaceutical Compositions

In an aspect the disclosure provides a kit comprising a Gene Writer or a Gene Writing system, e.g., as described herein. In some embodiments, the kit comprises a Gene Writer polypeptide (or a nucleic acid encoding the polypeptide) and a template RNA (or DNA encoding the template RNA). In some embodiments, the kit further comprises a reagent for introducing the system into a cell, e.g., transfection reagent, LNP, and the like. In some embodiments, the kit is suitable for any of the methods described herein. In some embodiments, the kit comprises one or more elements, compositions (e.g., pharmaceutical compositions), Gene Writers, and/or Gene Writer systems, or a functional fragment or component thereof, e.g., disposed in an article of manufacture. In some embodiments, the kit comprises instructions for use thereof.

In an aspect, the disclosure provides an article of manufacture, e.g., in which a kit as described herein, or a component thereof, is disposed.

In an aspect, the disclosure provides a pharmaceutical composition comprising a Gene Writer or a Gene Writing system, e.g., as described herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises a template RNA and/or an RNA encoding the polypeptide. In embodiments, the pharmaceutical composition has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

Chemistry, Manufacturing, and Controls (CMC)

Purification of protein therapeutics is described, for example, in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

In some embodiments, a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) conforms to certain quality standards. In some embodiments, a Gene

Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) produced by a method described herein conforms to certain quality standards. Accordingly, the disclosure is directed, in some aspects, to methods of manufacturing a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) that conforms to certain quality standards, e.g., in which said quality standards are assayed. The disclosure is also directed, in some aspects, to methods of assaying said quality standards in a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA). In some embodiments, quality standards include, but are not limited to, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12) of the following:

(i) the length of the template RNA, e.g., whether the template RNA has a length that is above a reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present is greater than 100, 125, 150, 175, or 200 nucleotides long;

(ii) the presence, absence, and/or length of a polyA tail on the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains a polyA tail (e.g., a polyA tail that is at least 5, 10, 20, 30, 50, 70, 100 nucleotides in length);

(iii) the presence, absence, and/or type of a 5' cap on the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains a 5' cap, e.g., whether that cap is a 7-methylguanosine cap, e.g., a O-Me-m7G cap;

(iv) the presence, absence, and/or type of one or more modified nucleotides (e.g., selected from pseudouridine, dihydrouridine, inosine, 7-methylguanosine, 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), 5-methylcytidine (5mC), or a locked nucleotide) in the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains one or more modified nucleotides;

(v) the stability of the template RNA (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA remains intact (e.g., greater than 100, 125, 150, 175, or 200 nucleotides long) after a stability test;

(vi) the potency of the template RNA in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the template RNA is assayed for potency;

(vii) the length of the polypeptide, first polypeptide, or second polypeptide, e.g., whether the polypeptide, first polypeptide, or second polypeptide has a length that is above a

reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide present is greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long);

(viii) the presence, absence, and/or type of post-translational modification on the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide contains phosphorylation, methylation, acetylation, myristoylation, palmitoylation, isoprenylation, glipyatyon, or lipoylation, or any combination thereof;

(ix) the presence, absence, and/or type of one or more artificial, synthetic, or non-canonical amino acids (e.g., selected from ornithine, β-alanine, GABA, δ-Aminolevulinic acid, PABA, a D-amino acid (e.g., D-alanine or D-glutamate), aminoisobutyric acid, dehydroalanine, cystathionine, lanthionine, Djenkolic acid, Diaminopimelic acid, Homoalanine, Norvaline, Norleucine, Homonorleucine, homoserine, O-methyl-homoserine and O-ethyl-homoserine, ethionine, selenocysteine, selenohomocysteine, selenomethionine, selenoethionine, tellurocysteine, or telluromethionine) in the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide present contains one or more artificial, synthetic, or non-canonical amino acids;

(x) the stability of the polypeptide, first polypeptide, or second polypeptide (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide remains intact (e.g., greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long)) after a stability test;

(xi) the potency of the polypeptide, first polypeptide, or second polypeptide in a system for modifying DNA, e.g., whether at least 1 % of target sites are modified after a system comprising the polypeptide, first polypeptide, or second polypeptide is assayed for potency; or

(xii) the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, or host cell protein, e.g., whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, or host cell protein contamination.

In some embodiments, a system or pharmaceutical composition described herein is endotoxin free.

In some embodiments, the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein is determined. In embodiments, whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein contamination is determined.

In some embodiments, a pharmaceutical composition or system as described herein has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

Applications

Using the systems described herein, optionally using any of delivery modalities described herein (including viral delivery modalities, such as AAVs), the invention also provides applications (methods) for modifying a DNA molecule, such as nuclear DNA, i.e., in the genome of a cell, whether in vitro, ex vivo, in situ, or in vivo, e.g,. in a tissue in an organism, such as a subject including mammalian subjects, such as a human. By integrating coding genes into a RNA sequence template, the Gene Writer system can address therapeutic needs, for example, by providing expression of a therapeutic transgene in individuals with loss-of-function mutations, by replacing gain-of-function mutations with normal transgenes, by providing regulatory sequences to eliminate gain-of-function mutation expression, and/or by controlling the expression of operably linked genes, transgenes and systems thereof. In certain embodiments, the RNA sequence template encodes a promotor region specific to the therapeutic needs of the host

cell, for example a tissue specific promotor or enhancer. In certain embodiments, the template nucleic acid encodes a promotor region specific to the therapeutic needs of the host cell, for example a tissue specific promotor or enhancer. In still other embodiments, a promotor can be operably linked to a coding sequence, e.g., for therapeutic intervention.

In certain aspects, the invention this provides methods of modifying a target DNA strand in a cell, tissue or subject, comprising administering a system as described herein (optionally by a modality described herein) to the cell, tissue or subject, where the system inserts the heterologous object sequence into the target DNA strand, thereby modifying the target DNA strand. In certain embodiments, the heterologous object sequence is thus expressed in the cell, tissue, or subject. In some embodiments, the cell, tissue or subject is a mammalian (e.g., human) cell, tissue or subject. Exemplary cells thus modified include a hepatocyte, lung epithelium, an ionocyte. Such a cell may be a primary cell or otherwise not immortalized. In related aspects, the invention also provides methods of treating a mammalian tissue comprising administering the a system as described herein to the mammal, thereby treating the tissue, wherein the tissue is deficient in the heterologous object sequence. In certain embodiments of any of the foregoing aspects and embodiments, the Gene Writer polypeptide is provided as a nucleic acid, which is present transiently.

In some embodiments, a system of the invention is capable of producing an insertion, substitution, deletion, or a combination thereof in target DNA. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene by altering, adding, or deleting sequences in a promoter or enhancer, e.g. sequences that bind transcription factors. In some embodiments, an insertion, deletion, substitution, or combination thereof alters translation of a gene (e.g. alters an amino acid sequence), inserts or deletes a start or stop codon, alters or fixes the translation frame of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof alters splicing of a gene, e.g. by inserting, deleting, or altering a splice acceptor or donor site. In some embodiments, an insertion, deletion, substitution, or combination thereof alters transcript or protein half-life. In some embodiments, an insertion, deletion, substitution, or combination thereof alters protein localization in the cell (e.g. from the cytoplasm to a mitochondria, from the cytoplasm into the

extracellular space (e.g. adds a secretion tag)). In some embodiments, an insertion, deletion, substitution, or combination thereof alters (e.g. improves) protein folding (e.g. to prevent accumulation of misfolded proteins). In some embodiments, an insertion, deletion, substitution, or combination thereof, alters, increases, decreases the activity of a gene, e.g. a protein encoded by the gene.

In embodiments, the Gene Writer™ gene editor system can provide therapeutic transgenes expressing, e.g., replacement blood factors or replacement enzymes, e.g., lysosomal enzymes. For example, the compositions, systems and methods described herein are useful to express, in a target human genome, agalsidase alpha or beta for treatment of Fabry Disease; imiglucerase, taliglucerase alfa, velaglucerase alfa, or alglucerase for Gaucher Disease; sebelipase alpha for lysosomal acid lipase deficiency (Wolman disease/CESD); laronidase, idursulfase, elosulfase alpha, or galsulfase for mucopolysaccharidoses; alglucosidase alpha for Pompe disease. For example, the compositions, systems and methods described herein are useful to express, in a target human genome factor I, II, V, VII, X, XI, XII or XIII for blood factor deficiencies.

In some embodiments, the heterologous object sequence encodes an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein, or a membrane protein). In some embodiments, the heterologous object sequence encodes a membrane protein, e.g., a membrane protein other than a CAR, and/or an endogenous human membrane protein. In some embodiments, the heterologous object sequence encodes an extracellular protein. In some embodiments, the heterologous object sequence encodes an enzyme, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, or a storage protein. Other proteins include an immune receptor protein, e.g. a synthetic immune receptor protein such as a chimeric antigen receptor protein (CAR), a T cell receptor, a B cell receptor, or an antibody.

A Gene Writing™ system may be used to modify immune cells. In some embodiments, a Gene Writing™ system may be used to modify T cells. In some embodiments, T-cells may include any subpopulation of T-cells, e.g., CD4+, CD8+, gamma-delta, naïve T cells, stem cell memory T cells, central memory T cells, or a mixture of subpopulations. In some embodiments, a Gene Writing™ system may be used to deliver or modify a T-cell receptor (TCR) in a T cell. In some embodiments, a Gene Writing™ system may be used to deliver at least one chimeric

antigen receptor (CAR) to T-cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to natural killer (NK) cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to natural killer T (NKT) cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to a progenitor cell, e.g., a progenitor cell of T, NK, or NKT cells. In some embodiments, cells modified with at least one CAR (e.g., CAR-T cells, CAR-NK cells, CAR-NKT cells), or a combination of cells modified with at least one CAR (e.g., a mixture of CAR-NK/T cells) are used to treat a condition as identified in the targetable landscape of CAR therapies in MacKay, et al. Nat Biotechnol 38, 233-244 (2020), incorporated by reference herein in its entirety. In some embodiments, the immune cells comprise a CAR specific to a tumor or a pathogen antigen selected from a group consisting of AChR (fetal acetylcholine receptor), ADGRE2, AFP (alpha fetoprotein), BAFF-R, BCMA, CAIX (carbonic anhydrase IX), CCR1, CCR4, CEA (carcinoembryonic antigen), 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 (epithelial glycoprotein-2), EGP-40 (epithelial glycoprotein-40), EGFR(HER1), EGFR-VIII, EpCAM (epithelial cell adhesion molecule), EphA2, ERBB2 (HER2, human epidermal growth factor receptor 2), ERBB3, ERBB4, FBP (folate-binding protein), Flt3 receptor, folate receptor-a, GD2 (ganglioside G2), GD3 (ganglioside G3), GPC3 (glypican-3), GPI00, hTERT (human telomerase reverse transcriptase), ICAM-1, integrin B7, interleukin 6 receptor, IL13Ra2 (interleukin-13 receptor 30 subunit alpha-2), kappa-light chain, KDR (kinase insert domain receptor), LeY (Lewis Y), L1CAM (LI cell adhesion molecule), LILRB2 (leukocyte immunoglobulin like receptor B2), MARTI, MAGE-A1 (melanoma associated antigen Al), MAGE- A3, MSLN (mesothelin), MUC16 (mucin 16), MUCI (mucin I), KG2D ligands, NY-ESO-1 (cancer-testis antigen), PRI (proteinase 3), TRBCI, TRBC2, TFM-3, TACI, tyrosinase, survivin, hTERT, oncofetal antigen (h5T4), p53, PSCA (prostate stem cell antigen), PSMA (prostate-specific membrane antigen), hRORl, TAG-72 (tumor- associated glycoprotein 72), VEGF-R2 (vascular endothelial growth factor R2), WT-1 (Wilms tumor protein), and antigens of HIV (human immunodeficiency virus), hepatitis B, hepatitis C, CMV (cytomegalovirus), EBV (Epstein-Barr virus), HPV (human papilloma virus).

In some embodiments, immune cells, e.g., T-cells, NK cells, NKT cells, or progenitor cells are modified ex vivo and then delivered to a patient. In some embodiments, a Gene Writer™ system is delivered by one of the methods mentioned herein, and immune cells, e.g., T-cells, NK cells, NKT cells, or progenitor cells are modified in vivo in the patient.

In some embodiments, a Gene Writer™ system described herein is delivered to a tissue or cell from the cerebrum, cerebellum, adrenal gland, ovary, pancreas, parathyroid gland, hypophysis, testis, thyroid gland, breast, spleen, tonsil, thymus, lymph node, bone marrow, lung, cardiac muscle, esophagus, stomach, small intestine, colon, liver, salivary gland, kidney, prostate, blood, or other cell or tissue type. In some embodiments, a Gene Writer™ system described herein is used to treat a disease, such as a cancer, inflammatory disease, infectious disease, genetic defect, or other disease. A cancer can be cancer of the cerebrum, cerebellum, adrenal gland, ovary, pancreas, parathyroid gland, hypophysis, testis, thyroid gland, breast, spleen, tonsil, thymus, lymph node, bone marrow, lung, cardiac muscle, esophagus, stomach, small intestine, colon, liver, salivary gland, kidney, prostate, blood, or other cell or tissue type, and can include multiple cancers.

In some embodiments, a Gene Writer™ system described herein described herein is administered by enteral administration (e.g. oral, rectal, gastrointestinal, sublingual, sublabial, or buccal administration). In some embodiments, a Gene Writer™ system described herein is administered by parenteral administration (e.g., intravenous, intramuscular, subcutaneous, intradermal, epidural, intracerebral, intracerebroventricular, epicutaneous, nasal, intra-arterial, intra-articular, intracavernous, intraocular, intraosseous infusion, intraperitoneal, intrathecal, intrauterine, intravaginal, intravesical, perivascular, or transmucosal administration). In some embodiments, a Gene Writer™ system described herein is administered by topical administration (e.g., transdermal administration).

In some embodiments, a Gene Writing system can be used to make multiple modifications to a target cell, either simultaneously or sequentially. In some embodiments, a Gene Writing system can be used to further modify an already modified cell. In some embodiments, a Gene Writing system can be use to modify a cell edited by a complementary technology, e.g., a gene edited cell, e.g., a cell with one or more CRISPR knockouts. In some embodiments, the previously edited cell is a T-cell. In some embodiments, the previous modifications comprise gene knockouts in a T-cell, e.g., endogenous TCR (e.g., TRAC, TRBC),

HLA Class I (B2M), PD1, CD52, CTLA-4, TIM-3, LAG-3, DGK. In some embodiments, a Gene Writing system is used to insert a TCR or CAR into a T-cell that has been previously modified.

In some embodiments, a Gene Writer™ system as described herein can be used to modify an animal cell, plant cell, or fungal cell. In some embodiments, a Gene Writer™ system as described herein can be used to modify a mammalian cell (e.g., a human cell). In some embodiments, a Gene Writer™ system as described herein can be used to modify a cell from a livestock animal (e.g., a cow, horse, sheep, goat, pig, llama, alpaca, camel, yak, chicken, duck, goose, or ostrich). In some embodiments, a Gene Writer™ system as described herein can be used as a laboratory tool or a research tool, or used in a laboratory method or research method, e.g., to modify an animal cell, e.g., a mammalian cell (e.g., a human cell), a plant cell, or a fungal cell.

In some embodiments, a Gene Writer™ system as described herein can be used to express a protein, template, or heterologous object sequence (e.g., in an animal cell, e.g., a mammalian cell (e.g., a human cell), a plant cell, or a fungal cell). In some embodiments, a Gene Writer™ system as described herein can be used to express a protein, template, or heterologous object sequence under the control of an inducible promoter (e.g., a small molecule inducible promoter). In some embodiments, a Gene Writing system or payload thereof is designed for tunable control, e.g., by the use of an inducible promoter. For example, a promoter, e.g., Tet, driving a gene of interest may be silent at integration, but may, in some instances, activated upon exposure to a small molecule inducer, e.g., doxycycline. In some embodiments, the tunable expression allows post-treatment control of a gene (e.g., a therapeutic gene), e.g., permitting a small molecule-dependent dosing effect. In embodiments, the small molecule-dependent dosing effect comprises altering levels of the gene product temporally and/or spatially, e.g., by local administration. In some embodiments, a promoter used in a system described herein may be inducible, e.g., responsive to an endogenous molecule of the host and/or an exogenous small molecule administered thereto.

In some embodiments, a Gene Writing system is used to make changes to non-coding and/or regulatory control regions, e.g., to tune the expression of endogenous genes. In some embodiments, a Gene Writing system is used to induce upregulation or downregulation of gene expression. In some embodiments, a regulatory control region comprises one or more of a promoter, enhancer, UTR, CTCF site, and/or a gene expression control region.

In some embodiments, a Gene Writing system may be used to treat or prevent a repeat expansion disease (e.g., a disease of Table 26), or to reduce the severity or a symptom thereof. In some embodiments, the repeat expansion disease comprises expansion of a trinucleotide repeat. In some embodiments, the subject has at least 10, 20, 30, 40, or 50 copies of the repeat. In embodiments, the repeat expansion disease is an inherited disease. Non-limiting examples of repeat expansion diseases include Huntington's disease (HD) and myotonic dystrophy. For example, healthy individuals may possess between 10 and 35 tandem copies of the CAG trinucleotide repeat, while Huntington's patients frequently possess >40 copies, which can result, e.g., in an elongated and dysfunctional Huntingtin protein. In some embodiments, a Gene Writer corrects a repeat expansion, e.g., by recognizing DNA at the terminus of the repeat region and nicking one strand (Figure 30). In some embodiments, the template RNA component of the Gene Writer comprises a region with a number of repeats characterstic of a healthy subject, e.g., about 20 repeats (e.g., between 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, or 35-40 repeats). In some embodiments, the template RNA component of the Gene Writer is copied by TPRT into the target site. In some embodiments, a second strand nick and second strand synthesis then results in the integration of the newly copied DNA comprising a correct number of repeats (e.g., as described herein). In some embodiments, the system recognizes DNA at the terminus of the repeat region and the template carries the information for the new number of repeats. In embodiments, a Gene Writer can be used in this way regardless of the number of repeats present in an individual and/or in an individual cell. Owing to the presence of multiple repeats, an alternative non-GeneWriter therapeutic (e.g., a CRISPR-based homologous recombination therapeutic) might, in some embodiments, result in unpredictable repair behavior. Further non-limiting examples of repeat expansion diseases and the causative repeats can be found, for example, in La Spada and Taylor Nat Rev Genet 11(4):247-258 (2010), which is incorporated herein by reference in its entirety.

In some embodiments, a Gene Writing system may be used to treat a healthy individual, e.g., as a preventative therapy. Gene Writing systems can, in some embodiments, be targeted to generate mutations, e.g., that have been shown to be protective towards a disease of interest. An exemplary list of such diseases and protective mutation targets can be found in Table 22.

A Gene Writing™ system may be used to treat indications of the liver. In some embodiments, the liver diseases preferred for therapeutic application of Gene Writing™ include,

but are not limited to, diseases selected from any of Tables 10A-10D or 11A-11G, or Table 5 of WO2020014209, which is hereby incorporated by reference. In exemplary embodiments, OTC deficiency is addressed by delivering all or a fragment of an OTC gene, e.g., all or a fragment of the OTC gene as contained in Table 5 of WO2020014209. In some embodiments, OTC deficiency is addressed by delivering a complete OTC gene expression cassette to a genome that complements the function of the mutated gene. In some embodiments, a fragment of the OTC gene is used that replaces the pathogenic mutation at its endogenous locus. In other embodiments, a Gene Writing™ system is used to address a condition selected from Column 6 of Table 4 of WO2020014209, paragraph "Lung diseases" below, or any of Tables 10A-10D or 11A-11G by delivering all or a fragment of a gene expression cassette encoding the corresponding gene indicated in Column 1 of Table 4 of WO2020014209, paragraph "Lung diseases" below, or in any of Tables 10A-10D or 11A-11G. In some embodiments, all or a fragment of said gene expression cassette is delivered to the endogenous locus of the pathogenic mutation. In some embodiments, all or a fragment of said gene expression cassette is integrated at a separate locus in the genome and complements the function of the mutated gene.

In certain embodiments a Gene Writer™ system provides a heterologous object sequence comprising a gene in Table 4 of WO2020014209, or paragraph "Lung diseases" below.

Table 10A: Indications and genetic targets, e.g., in the liver

Disease	Gene Affected
Acute intermittent porphyria	HMBS
Alpha-1-antitrypsin deficiency (AAT)	SERPINA1
Arginase deficiency	ARG1
Argininosuccinate lyase deficiency	ASL
Carbamoyl phosphate synthetase I deficiency	CPS1
Citrin deficiency	SLC25A13
Citrullinemia type I	ASS1
Crigler-Najjar syndrome (Hyperbilirubinemia)	UGT1A1
Fabry disease	GLA
Familial hypercholesterolemia 4 (homozygous familial cholesterolemia)	LDLRAP1
Glutaric aciduria I	GCDH
Glutaric aciduria II (multiple acyl-CoA dehydrogenase deficiency)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH

Glycogen storage disease type IV	GBE1
Hemophilia A	F8
Hemophilia B	F9
Hereditary hemochromatosis	HFE
Homocystinuria	CBS
Maple syrup urine disease (MSUD)	Type Ia: BCKDHA Type Ib: BCKDHB Type II: DBT
Methylmalonic acidemia (methylmalonyl-CoA mutase deficiency)	MMUT
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS3 (San Filippo Syndrome)	Type IIIa: SGSH Type IIIb: NAGLU Type IIIc: HGSNAT Type IIId: GNS
MPS4	Type IVA: GALNS Type IVB: GLB1
MPS6	ARSB
MPS7	GUSB
Ornithine transcarbamylase deficiency	OTC
Phenylketonuria (phenylalanine hydroxylase deficiency)	PAH
Polycystic Liver Disease	PCLD1: PRKCSH PCLD2: SEC63 PLCD3: ALG8 PCLD4: LRP5
Pompe disease	GAA
Primary Hyperoxaluria 1 (oxalosis)	AGXT
Progressive familial intrahepatic cholestasis type 1	ATP8B1
Progressive familial intrahepatic cholestasis type 2	ABCB11
Progressive familial intrahepatic cholestasis type 3	ABCB4
Propionic acidemia	PCCB; PCCA
Pyruvate carboxylase deficiency	PC
Tyrosinemia type I	FAH
Wilson's disease	ATP7B

Table 10B: Indications and genetic targets for HSCs

Disease	Gene Affected
Adrenoleukodystrophy (CALD)	ABCD1

Alpha-mannosidosis	MAN2B1
Blackfan-Diamond Anemia	
Congenital amegakaryocytic thrombocytopenia	MPL
Dyskeratosis Congenita	TERC
Fanconi anemia	FANC
Gaucher disease	GBA
Globoid cell leukodystrophy (Krabbe disease)	GALC
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	Many genes implicated
Metachromatic leukodystrophy	PSAP
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS7	GUSB
Mucolipidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1
Pompe disease	GAA
Pyruvate kinase deficiency (PKD)	PKLR
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB

Table 10C: Indications and genetic targets for the CNS

Disease	Gene Affected
Alpha-mannosidosis	MAN2B1
Ataxia-telangiectasia	ATM
CADASIL	NOTCH3
Canavan disease	ASPA
Carbamoyl-phosphate synthetase 1 deficiency	CPS1
CLN1 disease	PPT1
CLN2 Disease	TPP1

CLN3 Disease (Juvenile neuronal ceroid lipofuscinosis, Batten Disease)	CLN3
Coffin-Lowry syndrome	RPS6KA3
Congenital myasthenic syndrome 5	COLQ
Cornelia de Lange syndrome (NIPBL)	NIPBL
Cornelia de Lange syndrome (SMC1A)	SMC1A
Dravet syndrome (SCN1A)	SCN1A
Glycine encephalopathy (GLDC)	GLDC
GM1 gangliosidosis	GLB1
Huntington's Disease	HTT
Hydrocephalus with stenosis of the aqueduct of Sylvius	L1CAM
Leigh Syndrome	SURF1
Metachromatic leukodystrophy (ARSA)	ARSA
MPS type 2	IDS
MPS type 3	Type 3a: SGSH Type 3b: NAGLU
Mucolipidosis IV	MCOLN1
Neurofibromatosis Type 1	NF1
Neurofibromatosis type 2	NF2
Pantothenate kinase-associated neurodegeneration	PANK2
Pyridoxine-dependent epilepsy	ALDH7A1
Rett syndrome (MECP2)	MECP2
Sandhoff disease	HEXB
Semantic dementia (Frontotemporal dementia)	MAPT
Spinocerebellar ataxia with axonal neuropathy (Ataxia with Oculomotor Apraxia)	SETX
Tay-Sachs disease	HEXA
X-linked Adrenoleukodystrophy	ABCD1

Table 10D: Indications and genetic targets for the eye

Disease	Gene Affected
Achromatopsia	CNGB3
Amaurosis Congenita (LCA1)	GUCY2D
Amaurosis Congenita (LCA10)	CEP290
Amaurosis Congenita (LCA2)	RPE65
Amaurosis Congenita (LCA8)	CRB1
Choroideremia	CHM
Cone Rod Dystrophy (ABCA4)	ABCA4
Cone Rod Dystrophy (GUCY2D)	GUCY2D
Cystinosis, Ocular Nonnephropathic	CTNS
Doyne Honeycomb Retinal Dystrophy (DHRD)	EFEMP1
Familial Oculoleptomeningeal Amyloidosis	TTR
Keratitis-ichthyosis-deafness (KID)	GJB2
Lattice corneal dystrophy type I	TGFBI
Macular Corneal Dystrophy (MCD)	CHST6
Meesmann Corneal Dystrophy	KRT12; KRT3
Optic Atrophy	OPA1
Retinitis Pigmentosa (AR)	USH2A
Retinitis Rigmentosa (AD)	RHO
Sorsby Fundus Dystrophy	TIMP3
Stargardt Disease	ABCA4

In some embodiments, a GeneWriter system described herein is used to treat an indication of any of Tables 11A-11G. For instance, in some embodiments the GeneWriter system modifies a target site in genomic DNA in a cell, wherein the cell comprises a mutation at a gene of any of Tables 11A-11G, e.g., in a subject having the corresponding indication listed in any of Tables 11A-11G. In some embodiments, the target site is in a random site in the genome. In some embodiments, the target site is in a GSH sequence. In some embodiments, the cell is a liver cell that comprises a mutation in a gene of Table 11A, e.g., in a subject having the corresponding indication listed in Table 11A. In some embodiments, the cell is an HSC that comprises a mutation in a gene of Table 11B, e.g., in a subject having the corresponding indication listed in Table 11B. In some embodiments, the cell is a CNS cell comprising a

mutation in a gene of Table 11C, e.g., in a subject having the corresponding indication listed in Table 11C. In some embodiments, the cell is a cell of the eye that comprises a mutation in a gene of Table 11D, e.g., in a subject having the corresponding indication listed in Table 11D. In some embodiments, the cell is a cell of the lung that comprises a mutation in a gene of Table 11E, e.g., in a subject having the corresponding indication listed in Table 11E. In some embodiments, the cell is a muscle cell (e.g., skeletal muscle cell) that comprises a mutation in a gene of Table 11F, e.g., in a subject having the corresponding indication listed in Table 11F. In some embodiments, the cell is a skin cell that comprises a mutation in a gene of Table 11G, e.g., in a subject having the corresponding indication listed in Table 11G.

Table 11A: Indications and genetic targets for the liver

Disease	Gene Affected
3-hydroxy-3-methylglutaryl-CoA lyase deficiency	HMGCL
3-methylcrotonyl-CoA carboxylase (3MCC) deficiency	MCCC1
Acute hepatic porphyria	ALAD
Acute intermittent porphyria	HMBS
Alagille syndrome	JAG1
Alpha-1-antitrypsin deficiency (AAT)	SERPINA1
Alström syndrome	ALMS1
Arginase deficiency	ARG1
Argininosuccinate lyase deficiency	ASL
Carbamoyl phosphate synthetase I deficiency	CPS1
Cholesteryl ester storage disease	LIPA
Citrin deficiency	SLC25A13
Citrullinemia type I	ASS1
Crigler-Najjar syndrome (Hyperbilirubinemia)	UGT1A1
D-2-hydroxyglutaric aciduria type I	D2HGDH
Fabry disease	GLA
Familial chylomicronemia syndrome	LPL
Familial hypercholesterolemia 4 (homozygous familial cholesterolemia)	LDLRAP1
Farber disease	ASAH1
Gaucher disease	GBA

Glutaric aciduria I	GCDH
Glutaric aciduria II (multiple acyl-CoA dehydrogenase deficiency)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
Glutaric aciduria III	C7orf10
Glycogen storage disease type I	GSD1A: G6PC GSD1B: SLC37A4 GSD1C: SLC37A4
Glycogen storage disease type IV	GBE1
Hemophilia A	F8
Hemophilia B	F9
Hereditary amyloidosis (hTTR)	TTR
Hereditary fructose intolerance	ALDOB
Hereditary hemochromatosis	HFE
Homocystinuria	CBS
Hyperammonemia-hyperornithinemia-homocitrullinuria (HHH) syndrome	SLC25A15
Hyperoxaluria 1 (oxalosis)	HP1: AGXT HP2: GRHPR
Isovaleric acidemia	IVD
Lipoprotein lipase deficiency	LPL
Maple syrup urine disease (MSUD)	Type Ia: BCKDHA Type Ib: BCKDHB Type II: DBT
Methylmalonic acidemia (methylmalonyl-CoA mutase deficiency)	MMUT
Mitochondrial neurogastrointestinal encephalopathy syndrome	TYMP
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS3 (San Filippo Syndrome)	Type IIIa: SGSH Type IIIb: NAGLU Type IIIc: HGSNAT Type IIId: GNS
MPS4	Type IVA: GALNS Type IVB: GLB1
MPS6	ARSB
MPS7	GUSB
MPS9	HYAL1
N-acetylglutamate synthase deficiency	NAGS
Ornithine transcarbamylase deficiency	OTC
Phenylketonuria (phenylalanine hydroxylase deficiency)	PAH

Polycystic Liver Disease	PCLD1: PRKCSH PCLD2: SEC63 PLCD3: ALG8 PCLD4: LRP5
Pompe disease	GAA
Primary Hyperoxaluria 1 (oxalosis)	AGXT
Progressive familial intrahepatic cholestasis type 1	ATP8B1
Progressive familial intrahepatic cholestasis type 2	ABCB11
Progressive familial intrahepatic cholestasis type 3	ABCB4
Progressive familial intrahepatic cholestasis type 4	TJP2
Propionic acidemia	PCCB; PCCA
Pyruvate carboxylase deficiency	PC
Tyrosinemia type I	FAH
Wilson's disease	ATP7B

Table 11B: Indications and genetic targets for HSCs

Disease	Gene Affected
ADA-SCID	ADA
JAK-3 SCID	JAK3
Pyruvate kinase deficiency (PKD)	PKLR
RAG 1/2 Deficiency (SCID with granulomas)	RAG
Wiskott-Aldrich Syndrome	WAS
X-linked agammaglobulinemia	BTK
X-linked SCID	IL2RG
Adrenoleukodystrophy (CALD)	ABCD1
Alpha-mannosidosis	MAN2B1
Aspartylglycosaminuria	AGA
Blackfan-Diamond Anemia	
Chronic granulomatous disease	CYBB; CYBA; NCF1; NCF2; NCF4
Congenital amegakaryocytic thrombocytopenia	MPL
Congenital dyserythropoietic anemia	CDAN1
Fanconi anemia	FANC
Fucosidosis	FUCA1
Gaucher disease	GBA
Glanzmann's thrombasthenia	ITGA2B; ITGB3
Globoid cell leukodystrophy (Krabbe disease)	GALC

Gunther disease (congenital erythropoietic porphyria)	UROS
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
Hemophilia A	F8
Hemophilia B	F9
IPEX Syndrome	FOXP3
Kostmann's syndrome (severe congenital neutropenia)	HAX1
Leukocyte adhesion deficiency	ITGB2; FERMT3
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	Many genes implicated
Metachromatic leukodystrophy	PSAP
Molybdenum cofactor deficiency	MOCS1
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS6	ARSB
MPS7	GUSB
Mucolipidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1
Pompe disease	GAA
Purine nucleoside phosphorylase deficiency	PNP
Shwachman-Diamond syndrome	SBDS
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB
Wolman syndrome	LIPA
X-linked hyper IgM syndrome	TNFSF5
X-linked lymphoproliferative disease	SH2D1A
X-linked sideroblastic anemia	ALAS2
17-alpha-hydroxylase deficiency	CYP17A1
Common variable immunodeficiency	Many genes implicated
IL-7R SCID	IL7R
Kostmann's syndrome (severe congenital neutropenia)	ELANE

Table 11C: Indications and genetic targets for the CNS

Disease	Gene Affected
Cornelia de Lange syndrome (SMC1A)	SMC1A

Neurofibromatosis Type 1	NF1
Neurofibromatosis type 2	NF2
Rett syndrome (MECP2)	MECP2

Table 11D: Indications and genetic targets for the eye

Disease	Gene Affected
Cone Rod Dystrophy (CRX)	CRX
Cone Rod Dystrophy (GUCY2D)	GUCY2D
Lattice corneal dystrophy type I	TGFBI
Retinitis Rigmentosa (AD)	RHO
Vitelliform Macular Dystrophy	BEST1; PRPH2

5

Table 11E: Indications and genetic targets for the lung

Disease	Gene Affected
Alpha-1 antitrypsin deficiency	SERPINA1
Cystic fibrosis	CFTR
Fetal akinesia deformation sequence 1	MUSK
Fetal akinesia deformation sequence 2	RAPSN
Fetal akinesia deformation sequence 3	DOK7
Fetal akinesia deformation sequence 4	NUP88
Microphthalmia syndromic 9	STRA6
Primary ciliary dyskinesia	DNAI1; DNAH5 (~30%) many others
Pulmonary alveolar microlithiasis	SLC34A2
Pulmonary venoocclusive disease 2	EIF2AK4
Surfactant Protein B (SP-B) Deficiency (pulmonary surfactant metabolism dysfunction 1)	SFTPB
Surfactant metabolism dysfunction, pulmonary, 5	CSF2RB
Surfactant metabolism dysfunction, pulmonary, 4	CSF2RA
Surfactant metabolism dysfunction, pulmonary, 3	ABCA3

Table 11F: Indications and genetic targets for skeletal muscle

Disease	Gene Affected

Becker muscular dystrophy	DMD
Fukuyama CMD	FKTN
Muscle-eye-brain diseases	POMGNT1
Rigid spine syndrome	SELENON
Duchenne muscular dystrophy	DMD
Emery-Dreifuss muscular dystrophy, AR	LMNA
Limb-girdle muscular dystrophy 2A	CAPN3
Limb-girdle muscular dystrophy 2B	DYSF
Limb-girdle muscular dystrophy type 2E	SGCB
Limb-girdle muscular dystrophy type 2F	SGCD
Limb-girdle muscular dystrophy type 2H	TRIM32
Limb-girdle muscular dystrophy, type 2C	SGCG
Limb-girdle muscular dystrophy, type 2D	SGCA
Centronuclear myopathy 2	BIN1
Centronuclear myopathy, X-linked	MTM1
Centronuclear myopathy 5	SPEG
Centronuclear myopathy 6	ZAK
Congenital myopathies with fiber type disproportion	SELENON; TPM3; ACTA1
Central core disease	RYR1
Multiminicore myopathy	RYR1
Myosin storage myopathy, AR	MYH7
Nemaline myopathies	NEB ACTA1
GNE myopathy/Nonaka myopathy/hereditary inclusion-body myopathy	GNE
Markesbery-Griggs late-onset distal myopathy	TTN
Miyoshi muscular dystrophy 1	DYSF
Miyoshi muscular dystrophy 3	ANO5
Inclusion-body myositis 2	GNE
Carnitine deficiency	SLC22A5
Carnitine palmitoyltransferase deficiency 1A	CPT1A
Myoadenylate deaminase deficiency	AMPD1
Phosphofructokinase deficiency (Tarui disease) (GSD VII)	PFKM
Phosphoglycerate kinase deficiency	PGK1
Phosphoglycerate mutase deficiency (GSD X)	PGAM2
Phosphorylase deficiency (McArdle disease) (GSD V)	PYGM

Myofibrillar myopathy 1	DES
Myofibrillar myopathy 7	KY
Myofibrillar myopathy 8	PYROXD1
Congenital myasthenic syndrome	CHRNE (>50% of cases); RAPSN, CHAT, COLQ, DOK7; others
Charcot-Marie-Tooth 2	MFN2 (20% of cases)
Charcot-Marie-Tooth 4	A: GDAP1 F: PRX K: SURF1
Giant axonal neuropathy 1	GAN
Becker myotonia	CLCN1
Mitochondrial DNA depletion syndrome 2	TK2

Table 11G: Indications and genetic targets for the skin

Disease	**Gene Affected**
Acrodermatitis Enteropathica	SLC39A4
Alkaptonuria	HGD
Ataxia Telangiectasia	ATM
Berardinelli-Seip Syndrome	BSCL2; AGPAT2; CAVIN1; CAV1
Biotinidase Deficiency	BTD
Bloom Syndrome	RECQL3
Cerebrotendinous Xanthomatosis	CYP27A1
Chediak-Higashi Disease	LYST
Chondrodysplasia Punctata	ARSE
Chronic Granulomatous Disease	CYBB; CYBA; NCF1; NCF2; NCF4
Cockayne Syndrome	ERCC8; ERCC6
Congenital Erythropoietic Porphyria	UROS
Cutis Laxa	ATP6VOA2; FBLN5; ALDH18A1; PYCR1; LTBP4
De Barsy Syndrome	ALHD18A1; PYCR1
Dermatosparaxis (Ehlers-Danlos Type VII)	ADAMTS2
Dyskeratosis Congenita	TERC; TERT; RTEL1; ACD; DKC1; NOLA3
Ehlers-Danlos Type VI	PLOD1; FKBP14
Epidermodysplasia Verruciformis	TMC6
Epidermolysis Bullosa Dystrophica Pretibial	COL7A1

Epidermolysis Bullosa Dystrophica Recessive (Hallopeau-Siemens)	COL7A1; MMP1
Epidermolysis Bullosa Junctional	LAMA3; LAMC2; LAMB3; COL17A1; ITGB4
Erythropoietic Protoporphyria	FECH
Fabry Disease	GLA
Familial Dysautonomia	IKBKAP
Familial Mucocutaneous Candidiasis	CARD9; TRAF3IP2; IL17F; CLEC7A; IL17RC
Fanconi Anemia	FANC
Focal Facial Ectodermal Dysplasia	EDA
GAPO Syndrome	ANTXR1
Griscelli Syndrome Types 1, 2, and 3	MYO5A; RAB27A; MLPH
Harlequin Ichthyosis	ABCA12
Hartnup Disorder	SLC6A19
Hemochromatosis	HFE; HJV; TFR2
Hermansky-Pudlak Syndrome	HPS1; HPS6; AP3B1; HPS4; DTNBP1; HPS5; HPS3; AP3D1; BLOC1S3; BLOC1S6
Hyaline Fibromatosis Syndrome (systemic hyalinosis)	ANTXR2
Hypohidrotic Ectodermal Dysplasia	EDA; EDAR; EDARADD
Keratosis Follicularis Spinulosa Decalvans	MBTPS2
Kindler Syndrome	FERMT1
Lamellar Exfoliation of the Newborn	TGM1
Lamellar Ichthyosis/Nonbullous Congenital Ichthyosiform Erythroderma (ARCI)	ALOX12B
Leprechaunism	INSR
Lipogranulomatosis	ASAH1
Lipoid Proteinosis	ECM1
Mal de Meleda	SLURP1
Menkes Disease	ATP7A
Mutiple Pterygia Escobar Variant	CHRNG
Netherton Syndrome	SPINK5
Neu-Laxova Syndrome	PHGDH; PSAT1
Neutral Lipid Storage Disease with Ichthyosis	ABHD5
Oculocutaneous Albinism Tyrosinase Negative	TYR
Oculocutaneous Albinism Tyrosinase Positive	OCA2; MC1R
Papillon-Lefèvre	CTSC

Peeling Skin Syndrome	CDSN; TGM5; CHST8; CSTA; SERPINB8; FLG2
Pili Trianguli Et Canaliculi (uncombable hair syndrome)	PADI3
Prolidase Deficiency	PEPD
Pseudoxanthoma Elasticum (PXE)	ABCC6
Refsum Disease	PHYH
Restrictive Dermopathy	ZYMPSTE24; LMNA
Richner-Hanhart Syndrome (tyrosinemia type II)	TAT
Rothmund-Thomson Syndrome	RECQL4; ANAPC1
Sjögren-Larsson Syndrome	ALDH3A2
Transient Bullous Dermolysis of the Newborn	COL7A1
Trichothiodystrophy	ERCC2; ERCC3; MPLKIP; GTF2H5; GTF2E2
Tumoral Calcinosis	GALNT3; SAMD9; FRF23;
Vitiligo	Various
Werner Syndrome	RECQL2
Wiskott-Aldrich Syndrome	WAS
Wooly Hair	KRT25; LPAR6
Xeroderma Pigmentosum	ERCC2; ERCC3; ERCC4; ERCC5; XPA; XPC; DDB2; POLH
X-linked Recessive Ichthyosis	STS
Yellow Mutant Albinism	TYR

Additional suitable indications

Exemplary suitable diseases and disorders that can be treated by the systems or methods provided herein, for example, those comprising Gene Writers, include, without limitation: Baraitser-Winter syndromes 1 and 2; Diabetes mellitus and insipidus with optic atrophy and deafness; Alpha- 1- antitrypsin deficiency; Heparin cofactor II deficiency; Adrenoleukodystrophy; Keppen-Lubinsky syndrome; Treacher collins syndrome 1; Mitochondrial complex I, II, III, III (nuclear type 2, 4, or 8) deficiency; Hypermanganesemia with dystonia, polycythemia and cirrhosis; Carcinoid tumor of intestine; Rhabdoid tumor predisposition syndrome 2; Wilson disease; Hyperphenylalaninemia, bh4-deficient, a, due to partial pts deficiency, BH4-deficient, D, and non-pku; Hyperinsulinemic hypoglycemia familial 3, 4, and 5; Keratosis follicularis; Oral-facial-digital syndrome; SeSAME syndrome; Deafness,

nonsyndromic sensorineural, mitochondrial; Proteinuria; Insulin-dependent diabetes mellitus secretory diarrhea syndrome; Moyamoya disease 5; Diamond-Blackfan anemia 1, 5, 8, and 10; Pseudoachondroplastic spondyloepiphyseal dysplasia syndrome; Brittle cornea syndrome 2; Methylmalonic acidemia with homocystinuria, ; Adams-Oliver syndrome 5 and 6; autosomal recessive Agammaglobulinemia 2; Cortical malformations, occipital; Febrile seizures, familial, 11; Mucopolysaccharidosis type VI, type VI (severe), and type VII; Marden Walker like syndrome; Pseudoneonatal adrenoleukodystrophy; Spheroid body myopathy; Cleidocranial dysostosis; Multiple Cutaneous and Mucosal Venous Malformations; Liver failure acute infantile; Neonatal intrahepatic cholestasis caused by citrin deficiency; Ventricular septal defect 1; Oculodentodigital dysplasia; Wilms tumor 1; Weill- Marchesani-like syndrome; Renal adysplasia; Cataract 1, 4, autosomal dominant, autosomal dominant, multiple types, with microcornea, coppock-like, juvenile, with microcomea and glucosuria, and nuclear diffuse nonprogressive; Odontohypophosphatasia; Cerebro-oculo-facio- skeletal syndrome; Schizophrenia 15; Cerebral amyloid angiopathy, APP-related; Hemophagocytic lymphohistiocytosis, familial, 3; Porphobilinogen synthase deficiency; Episodic ataxia type 2; Trichorhinophalangeal syndrome type 3; Progressive familial heart block type IB; Glioma susceptibility 1; Lichtenstein-Knorr Syndrome; Hypohidrotic X-linked ectodermal dysplasia; Bartter syndrome types 3, 3 with hypocalciuria , and 4; Carbonic anhydrase VA deficiency, hyperammonemia due to; Cardiomyopathy; Poikiloderma, hereditary fibrosing, with tendon contractures, myopathy, and pulmonary fibrosis; Combined d-2- and 1-2- hydroxyglutaric aciduria; Arginase deficiency; Cone-rod dystrophy 2 and 6; Smith-Lemli-Opitz syndrome; Mucolipidosis III Gamma; Blau syndrome; Wemer syndrome; Meningioma; Iodotyrosyl coupling defect; Dubin-Johnson syndrome; 3-Oxo-5 alpha-steroid delta 4-dehydrogenase deficiency; Boucher Neuhauser syndrome; Iron accumulation in brain; Mental Retardation, X-Linked 102 and syndromic 13; familial, Pituitary adenoma predisposition; Hypoplasia of the corpus callosum; Hyperalphalipoproteinemia 2; Deficiency of ferroxidase; Growth hormone insensitivity with immunodeficiency; Marinesco-Sj\xc3\xb6gren syndrome; Martsolf syndrome; Gaze palsy, familial horizontal, with progressive scoliosis; Mitchell-Riley syndrome; Hypocalciuric hypercalcemia, familial, types 1 and 3; Rubinstein-Taybi syndrome; Epstein syndrome; Juvenile retinoschisis; Becker muscular dystrophy; Loeys-Dietz syndrome 1, 2, 3; Congenital muscular hypertrophy-cerebral syndrome; Familial juvenile gout; Spermatogenic

failure 11, 3, and 8; Orofacial cleft 11 and 7, Cleft lip/palate- ectodermal dysplasia syndrome; Mental retardation, X-linked, nonspecific, syndromic, Hedera type, and syndromic, wu type; Combined oxidative phosphorylation deficiencies 1, 3, 4, 12, 15, and 25; Frontotemporal dementia; Kniest dysplasia; Familial cardiomyopathy; Benign familial hematuria; Pheochromocytoma; Aminoglycoside-induced deafness; Gamma-aminobutyric acid transaminase deficiency; Oculocutaneous albinism type IB, type 3, and type 4; Renal coloboma syndrome; CNS hypomyelination; Hennekam lymphangiectasia-lymphedema syndrome 2; Migraine, familial basilar; Distal spinal muscular atrophy, X-linked 3; X-linked periventricular heterotopia; Microcephaly; Mucopolysaccharidosis, MPS-I-H/S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; Infantile Parkinsonism-dystonia; Frontotemporal dementia with TDP43 inclusions, TARDBP-related; Hereditary diffuse gastric cancer; Sialidosis type I and II; Microcephaly-capillary malformation syndrome; Hereditary breast and ovarian cancer syndrome; Brain small vessel disease with hemorrhage; Non-ketotic hyperglycinemia; Navajo neurohepatopathy; Auriculocondylar syndrome 2; Spastic paraplegia 15, 2, 3, 35, 39, 4, autosomal dominant, 55, autosomal recessive, and 5A; Autosomal recessive cutis laxa type IA and IB; Hemolytic anemia, nonspherocytic, due to glucose phosphate isomerase deficiency; Hutchinson-Gilford syndrome; Familial amyloid nephropathy with urticaria and deafness; Supravalvar aortic stenosis; Diffuse palmoplantar keratoderma, Bothnian type; Holt-Oram syndrome; Coffin Siris/Intellectual Disability; Left-right axis malformations; Rapadilino syndrome; Nanophthalmos 2; Craniosynostosis and dental anomalies; Paragangliomas 1; Snyder Robinson syndrome; Ventricular fibrillation; Activated PI3K-delta syndrome; Howel-Evans syndrome; Larsen syndrome, dominant type; Van Maldergem syndrome 2; MYH-associated polyposis; 6-pymvoyl-tetrahydropterin synthase deficiency; Alagille syndromes 1 and 2; Lymphangiomyomatosis; Muscle eye brain disease; WFS1-Related Disorders; Primary hypertrophic osteoarthropathy, autosomal recessive 2; Infertility; Nestor- Guillermo progeria syndrome; Mitochondrial trifunctional protein deficiency; Hypoplastic left heart syndrome 2; Primary dilated cardiomyopathy; Retinitis pigmentosa; Hirschsprung disease 3; Upshaw-Schulman syndrome; Desbuquois dysplasia 2; Diarrhea 3 (secretory sodium, congenital, syndromic) and 5 (with tufting enteropathy, congenital); Pachyonychia congenita 4 and type 2; Cerebral autosomal dominant and recessive arteriopathy with subcortical infarcts and leukoencephalopathy; Vi tel li form dystrophy ; type II, type IV, IV (combined hepatic and

myopathic), type V, and type VI; Atypical Rett syndrome; Atrioventricular septal defect 4; Papillon-Lef\xc3\xa8vre syndrome; Leber amaurosis; X-linked hereditary motor and sensory neuropathy; Progressive sclerosing poliodystrophy; Goldmann-Favre syndrome; Renal-hepatic-pancreatic dysplasia; Pallister-Hall syndrome; Amyloidogenic transthyretin amyloidosis; Melnick-Needles syndrome; Hyperimmunoglobulin E syndrome; Posterior column ataxia with retinitis pigmentosa; Chondrodysplasia punctata 1, X-linked recessive and 2 X-linked dominant; Ectopia lentis, isolated autosomal recessive and dominant; Familial cold urticarial; Familial adenomatous polyposis 1 and 3; Porokeratosis 8, disseminated superficial actinic type; PIK3CA Related Overgrowth Spectrum; Cerebral cavernous malformations 2; Exudative vitreoretinopathy 6; Megalencephaly cutis marmorata telangiectatica congenital; TARP syndrome; Diabetes mellitus, permanent neonatal, with neurologic features; Short-rib thoracic dysplasia 11 or 3 with or without polydactyly; Hypertrichotic osteochondrodysplasia; beta Thalassemia; Niemann-Pick disease type C1, C2, type A, and type C1, adult form; Charcot-Marie-Tooth disease types 1B, 2B2, 2C, 2F, 21, 2U (axonal), 1C (demyelinating), dominant intermediate C, recessive intermediate A, 2A2, 4C, 4D, 4H, IF, IVF, and X; Tyrosinemia type I; Paroxysmal atrial fibrillation; UV- sensitive syndrome; Tooth agenesis, selective, 3 and 4; Merosin deficient congenital muscular dystrophy; Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; Congenital aniridia; Left ventricular noncompaction 5; Deficiency of aromatic-L-amino-acid decarboxylase; Coronary heart disease; Leukonychia totalis; Distal arthrogryposis type 2B; Retinitis pigmentosa 10, 11, 12, 14, 15, 17, and 19; Robinow Sorauf syndrome; Tenorio Syndrome; Prolactinoma; Neurofibromatosis, type land type 2; Congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies, types A2, A7, A8, A11, and A14; Heterotaxy, visceral, 2, 4, and 6, autosomal; Jankovic Rivera syndrome; Lipodystrophy, familial partial, type 2 and 3; Hemoglobin H disease, nondeletional; Multicentric osteolysis, nodulosis and arthropathy; Thyroid agenesis; deficiency of Acyl-CoA dehydrogenase family, member 9; Alexander disease; Phytanic acid storage disease; Breast-ovarian cancer, familial 1, 2, and 4; Proline dehydrogenase deficiency; Childhood hypophosphatasia; Pancreatic agenesis and congenital heart disease; Vitamin D-dependent rickets, types land 2; Iridogoniodysgenesis dominant type and type 1; Autosomal recessive hypohidrotic ectodermal dysplasia syndrome; Mental retardation, X-linked, 3, 21, 30, and 72; Hereditary hemorrhagic telangiectasia type 2; Blepharophimosis, ptosis, and epicanthus inversus; Adenine

phosphoribosyltransferase deficiency; Seizures, benign familial infantile, 2; Acrodysostosis 2, with or without hormone resistance; Tetralogy of Fallot; Retinitis pigmentosa 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; Lysosomal acid lipase deficiency; Eichsfeld type congenital muscular dystrophy; Walker-Warburg congenital muscular dystrophy; TNF receptor-associated periodic fever syndrome (TRAPS); Progressive myoclonus epilepsy with ataxia; Epilepsy, childhood absence 2, 12 (idiopathic generalized, susceptibility to) 5 (nocturnal frontal lobe), nocturnal frontal lobe type 1, partial, with variable foci, progressive myoclonic 3, and X-linked, with variable learning disabilities and behavior disorders; Long QT syndrome; Dicarboxylic aminoaciduria; Brachydactyly types A1 and A2; Pseudoxanthoma elasticum-like disorder with multiple coagulation factor deficiency; Multisystemic smooth muscle dysfunction syndrome; Syndactyly Cenani Lenz type; Joubert syndrome 1, 6, 7, 9/15 (digenic), 14, 16, and 17, and Orofaciodigital syndrome xiv; Digitorenocerebral syndrome; Retinoblastoma; Dyskinesia, familial, with facial myokymia; Hereditary sensory and autonomic neuropathy type IIB amd IIA; familial hyperinsulinism; Megalencephalic leukoencephalopathy with subcortical cysts land 2a; Aase syndrome; Wiedemann- Steiner syndrome; Ichthyosis exfoliativa; Myotonia congenital; Granulomatous disease, chronic, X-linked, variant; Deficiency of 2-methylbutyryl-CoA dehydrogenase; Sarcoidosis, early-onset; Glaucoma, congenital and Glaucoma, congenital, Coloboma; Breast cancer, susceptibility to; Ceroid lipofuscinosis neuronal 2, 6, 7, and 10; Congenital generalized lipodystrophy type 2; Fructose-biphosphatase deficiency; Congenital contractural arachnodactyly; Lynch syndrome I and II; Phosphoglycerate dehydrogenase deficiency; Burn- Mckeown syndrome; Myocardial infarction 1; Achromatopsia 2 and 7; Retinitis Pigmentosa 73; Protan defect; Polymicrogyria, asymmetric, bilateral frontoparietal; Spinal muscular atrophy, distal, autosomal recessive, 5; Methylmalonic aciduria due to methylmalonyl-CoA mutase deficiency; Familial porencephaly; Hurler syndrome; Oto-palato-digital syndrome, types I and II; Sotos syndrome 1 or 2; Cardioencephalomyopathy, fatal infantile, due to cytochrome c oxidase deficiency; Parastremmatic dwarfism; Thyrotropin releasing hormone resistance, generalized; Diabetes mellitus, type 2, and insulin-dependent, 20; Thoracic aortic aneurysms and aortic dissections; Estrogen resistance; Maple syrup urine disease type 1A and type 3; Hypospadias 1 and 2, X-linked; Metachromatic leukodystrophy juvenile, late infantile, and adult types; Early T cell progenitor acute lymphoblastic leukemia; Neuropathy, Hereditary Sensory, Type IC; Mental retardation, autosomal dominant 31; Retinitis pigmentosa

39; Breast cancer, early-onset; May-Hegglin anomaly; Gaucher disease type 1 and Subacute neuronopathic; Temtamy syndrome; Spinal muscular atrophy, lower extremity predominant 2, autosomal dominant; Fanconi anemia, complementation group E, I, N, and O; Alkaptonuria; Hirschsprung disease; Combined malonic and methylmalonic aciduria; Arrhythmogenic right ventricular cardiomyopathy types 5, 8, and 10; Congenital lipomatous overgrowth, vascular malformations, and epidermal nevi; Timothy syndrome; Deficiency of guanidinoacetate methyltransferase; Myoclonic dystonia; Kanzaki disease; Neutral 1 amino acid transport defect; Neurohypophyseal diabetes insipidus; Thyroid hormone metabolism, abnormal; Benign scapuloperoneal muscular dystrophy with cardiomyopathy; Hypoglycemia with deficiency of glycogen synthetase in the liver; Hypertrophic cardiomyopathy; Myasthenic Syndrome, Congenital, 11, associated with acetylcholine receptor deficiency; Mental retardation X-linked syndromic 5; Stormorken syndrome; Aplastic anemia; Intellectual disability; Normokalemic periodic paralysis, potassium-sensitive; Danon disease; Nephronophthisis 13, 15 and 4; Thyrotoxic periodic paralysis and Thyrotoxic periodic paralysis 2; Infertility associated with multi-tailed spermatozoa and excessive DNA; Glaucoma, primary open angle, juvenile-onset; Afibrinogenemia and congenital Afibrinogenemia; Polycystic kidney disease 2, adult type, and infantile type; Familial porphyria cutanea tarda; Cerebello-oculo-renal syndrome (nephronophthisis, oculomotor apraxia and cerebellar abnormalities); Frontotemporal Dementia Chromosome 3- Linked and Frontotemporal dementia ubiquitin-positive; Metatrophic dysplasia; Immunodeficiency-centromeric instability-facial anomalies syndrome 2; Anemia, nonspherocytic hemolytic, due to G6PD deficiency; Bronchiectasis with or without elevated sweat chloride 3; Congenital myopathy with fiber type disproportion; Carney complex, type 1; Cryptorchidism, unilateral or bilateral; Ichthyosis bullosa of Siemens; Isolated lutropin deficiency; DFNA 2 Nonsyndromic Hearing Loss; Klein-Waardenberg syndrome; Gray platelet syndrome; Bile acid synthesis defect, congenital, 2; 46, XY sex reversal, type 1, 3, and 5; Acute intermittent porphyria; Cornelia de Lange syndromes 1 and 5; Hyperglycinuria; Cone-rod dystrophy 3; Dysfibrinogenemia; Karak syndrome; Congenital muscular dystrophy-dystroglycanopathy without mental retardation, type B5; Infantile nystagmus, X-linked; Dyskeratosis congenita, autosomal recessive, 1, 3, 4, and 5; Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation; Hyperlysinemia; Bardet-Biedl syndromes 1, 11, 16, and 19; Autosomal recessive centronuclear myopathy; Frasier syndrome; Caudal

regression syndrome; Fibrosis of extraocular muscles, congenital, 1, 2, 3a (with or without extraocular involvement), 3b; Prader-Willi-like syndrome; Malignant melanoma; Bloom syndrome; Darier disease, segmental; Multicentric osteolysis nephropathy; Hemochromatosis type 1, 2B, and 3; Cerebellar ataxia infantile with progressive external ophthalmoplegi and Cerebellar ataxia, mental retardation, and dysequilibrium syndrome 2; Hypoplastic left heart syndrome; Epilepsy, Hearing Loss, And Mental Retardation Syndrome; Transferrin serum level quantitative trait locus 2; Ocular albinism, type I; Marfan syndrome; Congenital muscular dystrophy- dystroglycanopathy with brain and eye anomalies, type A14 and B14; Hyperammonemia, type III; Cryptophthalmos syndrome; Alopecia universalis congenital; Adult hypophosphatasia; Mannose-binding protein deficiency; Bull eye macular dystrophy; Autosomal dominant torsion dystonia 4; Nephrotic syndrome, type 3, type 5, with or without ocular abnormalities, type 7, and type 9; Seizures, Early infantile epileptic encephalopathy 7; Persistent hyperinsulinemic hypoglycemia of infancy; Thrombocytopenia, X-linked; Neonatal hypotonia; Orstavik Lindemann Solberg syndrome; Pulmonary hypertension, primary, 1, with hereditary hemorrhagic telangiectasia; Pituitary dependent hypercortisolism; Epidermodysplasia verruciformis; Epidermolysis bullosa, junctional, localisata variant; Cytochrome c oxidase i deficiency; Kindler syndrome; Myosclerosis, autosomal recessive; Truncus arteriosus; Duane syndrome type 2; ADULT syndrome; Zellweger syndrome spectrum; Leukoencephalopathy with ataxia, with Brainstem and Spinal Cord Involvement and Lactate Elevation, with vanishing white matter, and progressive, with ovarian failure; Antithrombin III deficiency; Holoprosencephaly 7; Roberts-SC phocomelia syndrome; Mitochondrial DNA-depletion syndrome 3 and 7, hepatocerebral types, and 13 (encephalomyopathic type); Porencephaly 2; Microcephaly, normal intelligence and immunodeficiency; Giant axonal neuropathy; Sturge-Weber syndrome, Capillary malformations, congenital, 1; Fabry disease and Fabry disease, cardiac variant; Glutamate formiminotransferase deficiency; Fanconi-Bickel syndrome; Acromicric dysplasia; Epilepsy, idiopathic generalized, susceptibility to, 12; Basal ganglia calcification, idiopathic, 4; Polyglucosan body myopathy 1 with or without immunodeficiency; Malignant tumor of prostate; Congenital ectodermal dysplasia of face; Congenital heart disease; Age-related macular degeneration 3, 6, 11, and 12; Congenital myotonia, autosomal dominant and recessive forms; Hypomagnesemia 1, intestinal; Sulfite oxidase deficiency, isolated; Pick disease; Plasminogen deficiency, type I; Syndactyly type 3; Cone-rod dystrophy amelogenesis imperfecta;

Pseudoprimary hyperaldosteronism; Terminal osseous dysplasia; Bartter syndrome antenatal type 2; Congenital muscular dystrophy- dystroglycanopathy with mental retardation, types B2, B3, B5, and B15; Familial infantile myasthenia; Lymphoproliferative syndrome 1, 1 (X-linked), and 2; Hypercholesterolaemia and Hypercholesterolemia, autosomal recessive; Neoplasm of ovary; Infantile GM1 gangliosidosis; Syndromic X-linked mental retardation 16; Deficiency of ribose-5-phosphate isomerase; Alzheimer disease, types, 1, 3, and 4; Andersen Tawil syndrome; Multiple synostoses syndrome 3; Chilbain lupus 1; Hemophagocytic lymphohistiocytosis, familial, 2; Axenfeld-Rieger syndrome type 3; Myopathy, congenital with cores; Osteoarthritis with mild chondrodysplasia; Peroxisome biogenesis disorders; Severe congenital neutropenia; Hereditary neuralgic amyotrophy; Palmoplantar keratoderma, nonepidermolytic, focal or diffuse; Dysplasminogenemia; Familial colorectal cancer; Spastic ataxia 5, autosomal recessive, Charlevoix-Saguenay type, 1,10, or 11, autosomal recessive; Frontometaphyseal dysplasia land 3; Hereditary factors II, IX, VIII deficiency disease; Spondylocheirodysplasia, Ehlers-Danlos syndrome-like, with immune dysregulation, Aggrecan type, with congenital joint dislocations, short limb-hand type, Sedaghatian type, with cone-rod dystrophy, and Kozlowski type; Ichthyosis prematurity syndrome; Stickler syndrome type 1; Focal segmental glomerulosclerosis 5; 5-Oxoprolinase deficiency; Microphthalmia syndromic 5, 7, and 9; Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome; Deficiency of butyryl-CoA dehydrogenase; Maturity-onset diabetes of the young, type 2; Mental retardation, syndromic, Claes-Jensen type, X-linked; Deafness, cochlear, with myopia and intellectual impairment, without vestibular involvement, autosomal dominant, X-linked 2; Spondylocarpotarsal synostosis syndrome; Sting-associated vasculopathy, infantile-onset; Neutral lipid storage disease with myopathy; Immune dysfunction with T-cell inactivation due to calcium entry defect 2; Cardiofaciocutaneous syndrome; Corticosterone methyloxidase type 2 deficiency; Hereditary myopathy with early respiratory failure; Interstitial nephritis, karyomegalic; Trimethylaminuria; Hyperimmunoglobulin D with periodic fever; Malignant hyperthermia susceptibility type 1; Trichomegaly with mental retardation, dwarfism and pigmentary degeneration of retina; Breast adenocarcinoma; Complement factor B deficiency; Ullrich congenital muscular dystrophy; Left ventricular noncompaction cardiomyopathy; Fish-eye disease; Finnish congenital nephrotic syndrome; Limb-girdle muscular dystrophy, type IB, 2A, 2B, 2D, C1, C5, C9, C14; Idiopathic fibrosing alveolitis, chronic form; Primary familial hypertrophic cardiomyopathy; Angiotensin i-

converting enzyme, benign serum increase; Cd8 deficiency, familial; Proteus syndrome; Glucose-6-phosphate transport defect; Borjeson-Forssman-Lehmann syndrome; Zellweger syndrome; Spinal muscular atrophy, type II; Prostate cancer, hereditary, 2; Thrombocytopenia, platelet dysfunction, hemolysis, and imbalanced globin synthesis; Congenital disorder of glycosylation types IB, ID, 1G, 1H, 1 J, IK, IN, IP, 2C, 2J, 2K, Ilm; Junctional epidermolysis bullosa gravis of Herlitz; Generalized epilepsy with febrile seizures plus 3, type 1, type 2; Schizophrenia 4; Coronary artery disease, autosomal dominant 2; Dyskeratosis congenita, autosomal dominant, 2 and 5; Subcortical laminar heterotopia, X-linked; Adenylate kinase deficiency; X- linked severe combined immunodeficiency; Coproporphyria; Amyloid Cardiomyopathy, Transthyretin-related; Hypocalcemia, autosomal dominant 1; Brugada syndrome; Congenital myasthenic syndrome, acetazolamide- responsive; Primary hypomagnesemia; Sclerosteosis; Frontotemporal dementia and/or amyotrophic lateral sclerosis 3 and 4; Mevalonic aciduria; Schwannomatosis 2; Hereditary motor and sensory neuropathy with optic atrophy; Porphyria cutanea tarda; Osteochondritis dissecans; Seizures, benign familial neonatal, 1, and/or myokymia; Long QT syndrome, LQT1 subtype; Mental retardation, anterior maxillary protrusion, and strabismus; Idiopathic hypercalcemia of infancy; Hypogonadotropic hypogonadism 11 with or without anosmia; Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy; Primary autosomal recessive microcephaly 10, 2, 3, and 5; Interrupted aortic arch; Congenital amegakaryocytic thrombocytopenia; Hermansky-Pudlak syndrome 1, 3, 4, and 6; Long QT syndrome 1, 2, 2/9, 2/5, (digenic), 3, 5 and 5, acquired, susceptibility to; Andermann syndrome; Retinal cone dystrophy 3B; Erythropoietic protoporphyria; Sepiapterin reductase deficiency; Very long chain acyl-CoA dehydrogenase deficiency; Hyperferritinemia cataract syndrome; Silver spastic paraplegia syndrome; Charcot-Marie-Tooth disease; Atrial septal defect 2; Carnevale syndrome; Hereditary insensitivity to pain with anhidrosis; Catecholaminergic polymorphic ventricular tachycardia; Hypokalemic periodic paralysis 1 and 2; Sudden infant death syndrome; Hypochromic microcytic anemia with iron overload; GLUT1 deficiency syndrome 2; Leukodystrophy, Hypomyelinating, 11 and 6; Cone monochromatism; Osteopetrosis autosomal dominant type 1 and 2, recessive 4, recessive 1, recessive 6; Severe congenital neutropenia 3, autosomal recessive or dominant; Methionine adenosyltransferase deficiency, autosomal dominant; Paroxysmal familial ventricular fibrillation; Pyruvate kinase deficiency of red cells; Schneckenbecken dysplasia; Torsades de pointes; Distal

myopathy Markesbery-Griggs type; Deficiency of UDPglucose-hexose-1- phosphate uridylyltransferase; Sudden cardiac death; Neu-Laxova syndrome 1; Atransferrinemia; Hyperparathyroidism 1 and 2; Cutaneous malignant melanoma 1; Symphalangism, proximal, Ib; Progressive pseudorheumatoid dysplasia; Werdnig-Hoffmann disease; Achondrogenesis type 2; Holoprosencephaly 2, 3,7, and 9; Schindler disease, type 1; Cerebroretinal microangiopathy with calcifications and cysts; Heterotaxy, visceral, X-linked; Tuberous sclerosis syndrome; Kartagener syndrome; Thyroid hormone resistance, generalized, autosomal dominant; Bestrophinopathy, autosomal recessive; Nail disorder, nonsyndromic congenital, 8; Mohr-Tranebjaerg syndrome; Cone-rod dystrophy 12; Hearing impairment; Ovarioleukodystrophy; Renal tubular acidosis, proximal, with ocular abnormalities and mental retardation; Dihydropteridine reductase deficiency; Focal epilepsy with speech disorder with or without mental retardation; Ataxia- telangiectasia syndrome; Brown- Vialetto- Van laere syndrome and Brown- Vialetto-Van Laere syndrome 2; Cardiomyopathy; Peripheral demyelinating neuropathy, central dysmyelination; Corneal dystrophy, Fuchs endothelial, 4; Cowden syndrome 3; Dystonia 2 (torsion, autosomal recessive), 3 (torsion, X-linked), 5 (Dopa-responsive type), 10, 12, 16, 25, 26 (Myoclonic); Epiphyseal dysplasia, multiple, with myopia and conductive deafness; Cardiac conduction defect, nonspecific; Branchiootic syndromes 2 and 3; Peroxisome biogenesis disorder 14B, 2A, 4A, 5B, 6A, 7A, and 7B; Familial renal glucosuria; Candidiasis, familial, 2, 5, 6, and 8; Autoimmune disease, multisystem, infantile-onset; Early infantile epileptic encephalopathy 2, 4, 7, 9, 10, 11, 13, and 14; Segawa syndrome, autosomal recessive; Deafness, autosomal dominant 3a, 4, 12, 13, 15, autosomal dominant nonsyndromic sensorineural 17, 20, and 65; Congenital dyserythropoietic anemia, type I and II; Enhanced s-cone syndrome; Adult neuronal ceroid lipofuscinosis; Atrial fibrillation, familial, 11, 12, 13, and 16; Norum disease; Osteosarcoma; Partial albinism; Biotinidase deficiency; Combined cellular and humoral immune defects with granulomas; Alpers encephalopathy; Holocarboxylase synthetase deficiency; Maturity-onset diabetes of the young, type 1, type 2, type 11, type 3, and type 9; Variegate porphyria; Infantile cortical hyperostosis; Testosterone 17-beta- dehydrogenase deficiency; L-2-hydroxyglutaric aciduria; Tyrosinase-negative oculocutaneous albinism; Primary ciliary dyskinesia 24; Pontocerebellar hypoplasia type 4; Ciliary dyskinesia, primary, 7, 11, 15, 20 and 22; Idiopathic basal ganglia calcification 5; Brain atrophy; Craniosynostosis 1 and 4; Keratoconus 1; Rasopathy; Congenital adrenal hyperplasia and Congenital adrenal hypoplasia, X-linked;

Mitochondrial DNA depletion syndrome 11, 12 (cardiomyopathic type), 2, 4B (MNGIE type), 8B (MNGIE type); Brachydactyly with hypertension; Cornea plana 2; Aarskog syndrome; Multiple epiphyseal dysplasia 5 or Dominant; Corneal endothelial dystrophy type 2; Aminoacylase 1 deficiency; Delayed speech and language development; Nicolaides-Baraitser syndrome; Enterokinase deficiency; Ectrodactyly, ectodermal dysplasia, and cleft lip/palate syndrome 3; Arthrogryposis multiplex congenita, distal, X-linked; Perrault syndrome 4; Jervell and Lange-Nielsen syndrome 2; Hereditary Nonpolyposis Colorectal Neoplasms; Robinow syndrome, autosomal recessive, autosomal recessive, with brachy-syn-polydactyly; Neurofibrosarcoma; Cytochrome-c oxidase deficiency ; Vesicoureteral reflux 8; Dopamine beta hydroxylase deficiency; Carbohydrate-deficient glycoprotein syndrome type I and II; Progressive familial intrahepatic cholestasis 3; Benign familial neonatal-infantile seizures; Pancreatitis, chronic, susceptibility to; Rhizomelic chondrodysplasia punctata type 2 and type 3; Disordered steroidogenesis due to cytochrome p450 oxidoreductase deficiency; Deafness with labyrinthine aplasia microtia and microdontia (FAMM); Rothmund-Thomson syndrome; Cortical dysplasia, complex, with other brain malformations 5 and 6; Myasthenia, familial infantile, 1; Trichorhinophalangeal dysplasia type I; Worth disease; Splenic hypoplasia; Molybdenum cofactor deficiency, complementation group A; Sebastian syndrome; Progressive familial intrahepatic cholestasis 2 and 3; Weill-Marchesani syndrome 1 and 3; Microcephalic osteodysplastic primordial dwarfism type 2; Surfactant metabolism dysfunction, pulmonary, 2 and 3; Severe X-linked myotubular myopathy; Pancreatic cancer 3; Platelet- type bleeding disorder 15 and 8; Tyrosinase-positive oculocutaneous albinism; Borrone Di Rocco Crovato syndrome; ATR-X syndrome; Sucrase-isomaltase deficiency; Complement component 4, partial deficiency of, due to dysfunctional c1 inhibitor; Congenital central hypoventilation; Infantile hypophosphatasia; Plasminogen activator inhibitor type 1 deficiency; Malignant lymphoma, non-Hodgkin; Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome; Schwartz Jampel syndrome type 1; Fetal hemoglobin quantitative trait locus 1; Myopathy, distal, with anterior tibial onset; Noonan syndrome 1 and 4, LEOPARD syndrome 1; Glaucoma 1, open angle, e, F, and G; Kenny-Caffey syndrome type 2; PTEN hamartoma tumor syndrome; Duchenne muscular dystrophy; Insulin-resistant diabetes mellitus and acanthosis nigricans; Microphthalmia, isolated 3, 5, 6, 8, and with coloboma 6; Raine syndrome; Premature ovarian failure 4, 5, 7, and 9; Allan-Herndon-Dudley syndrome; Citrullinemia type I; Alzheimer disease, familial, 3, with spastic

paraparesis and apraxia; Familial hemiplegic migraine types 1 and 2; Ventriculomegaly with cystic kidney disease; Pseudoxanthoma elasticum; Homocysteinemia due to MTHFR deficiency, CBS deficiency, and Homocystinuria, pyridoxine-responsive; Dilated cardiomyopathy 1A, 1AA, 1C, 1G, IBB, 1DD, IFF, 1HH, II, IKK, IN, IS, 1Y, and 3B; Muscle AMP guanine oxidase deficiency; Familial cancer of breast; Hereditary sideroblastic anemia; Myoglobinuria, acute recurrent, autosomal recessive; Neuroferritinopathy; Cardiac arrhythmia; Glucose transporter type 1 deficiency syndrome; Holoprosencephaly sequence; Angiopathy, hereditary, with nephropathy, aneurysms, and muscle cramps; Isovaleryl-CoA dehydrogenase deficiency; Kallmann syndrome 1, 2, and 6; Permanent neonatal diabetes mellitus; Acrocallosal syndrome, Schinzel type; Gordon syndrome; MYH9 related disorders; Donnai Barrow syndrome; Severe congenital neutropenia and 6, autosomal recessive; Charcot-Marie-Tooth disease, types ID and IVF; Coffin-Lowry syndrome; mitochondrial 3-hydroxy-3- methylglutaryl-CoA synthase deficiency; Hypomagnesemia, seizures, and mental retardation; Ischiopatellar dysplasia; Multiple congenital anomalies -hypotonia- seizures syndrome 3; Spastic paraplegia 50, autosomal recessive; Short stature with nonspecific skeletal abnormalities; Severe myoclonic epilepsy in infancy; Propionic academia; Adolescent nephronophthisis; Macrocephaly, macrosomia, facial dysmorphism syndrome; Stargardt disease 4; Ehlers-Danlos syndrome type 7 (autosomal recessive), classic type, type 2 (progeroid), hydroxylysine-deficient, type 4, type 4 variant, and due to tenascin-X deficiency; Myopia 6; Coxa plana; Familial cold autoinflammatory syndrome 2; Malformation of the heart and great vessels; von Willebrand disease type 2M and type 3; Deficiency of galactokinase; Brugada syndrome 1; X-linked ichthyosis with steryl-sulfatase deficiency; Congenital ocular coloboma; Histiocytosis-lymphadenopathy plus syndrome; Aniridia, cerebellar ataxia, and mental retardation; Left ventricular noncompaction 3; Amyotrophic lateral sclerosis types 1, 6, 15 (with or without frontotemporal dementia), 22 (with or without frontotemporal dementia), and 10; Osteogenesis imperfecta type 12, type 5, type 7, type 8, type I, type III, with normal sclerae, dominant form, recessive perinatal lethal; Hematologic neoplasm; Favism, susceptibility to; Pulmonary Fibrosis And/Or Bone Marrow Failure, Telomere-Related, 1 and 3; Dominant hereditary optic atrophy; Dominant dystrophic epidermolysis bullosa with absence of skin; Muscular dystrophy, congenital, megaconial type; Multiple gastrointestinal atresias; McCune- Albright syndrome; Nail-patella syndrome; McLeod neuroacanthocytosis syndrome; Common variable

immunodeficiency 9; Partial hypoxanthine-guanine phosphoribosyltransferase deficiency; Pseudohypoaldosteronism type 1 autosomal dominant and recessive and type 2; Urocanate hydratase deficiency; Heterotopia; Meckel syndrome type 7; Ch\xc3\xa9diak-Higashi syndrome , Chediak-Higashi syndrome, adult type; Severe combined immunodeficiency due to ADA deficiency, with microcephaly, growth retardation, and sensitivity to ionizing radiation, atypical, autosomal recessive, T cell-negative, B cell-positive, NK cell-negative of NK- positive; Insulin resistance; Deficiency of steroid 11 -beta-monooxygenase; Popliteal pterygium syndrome; Pulmonary arterial hypertension related to hereditary hemorrhagic telangiectasia; Deafness, autosomal recessive 1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86, and 89; Primary hyperoxaluria, type I, type, and type III; Paramyotonia congenita of von Eulenburg; Desbuquois syndrome; Carnitine palmitoyltransferase I , II, II (late onset), and II (infantile) deficiency; Secondary hypothyroidism; Mandibulofacial dysostosis, Treacher Collins type, autosomal recessive; Cowden syndrome 1; Li-Fraumeni syndrome 1; Asparagine synthetase deficiency; Malattia leventinese; Optic atrophy 9; Infantile convulsions and paroxysmal choreoathetosis, familial; Ataxia with vitamin E deficiency; Islet cell hyperplasia; Miyoshi muscular dystrophy 1; Thrombophilia, hereditary, due to protein C deficiency, autosomal dominant and recessive; Fechtner syndrome; Properdin deficiency, X-linked; Mental retardation, stereotypic movements, epilepsy, and/or cerebral malformations; Creatine deficiency, X-linked; Pilomatrixoma; Cyanosis, transient neonatal and atypical nephropathic; Adult onset ataxia with oculomotor apraxia; Hemangioma, capillary infantile; PC-K6a; Generalized dominant dystrophic epidermolysis bullosa; Pelizaeus-Merzbacher disease; Myopathy, centronuclear, 1, congenital, with excess of muscle spindles, distal, 1, lactic acidosis, and sideroblastic anemia 1, mitochondrial progressive with congenital cataract, hearing loss, and developmental delay, and tubular aggregate, 2; Benign familial neonatal seizures 1 and 2; Primary pulmonary hypertension; Lymphedema, primary, with myelodysplasia; Congenital long QT syndrome; Familial exudative vitreoretinopathy, X-linked; Autosomal dominant hypohidrotic ectodermal dysplasia; Primordial dwarfism; Familial pulmonary capillary hemangiomatosis; Carnitine acylcamitine translocase deficiency; Visceral myopathy; Familial Mediterranean fever and Familial mediterranean fever, autosomal dominant; Combined partial and complete 17-alpha-hydroxylase/ 17, 20-lyase deficiency; Oto-palato-digital syndrome, type I; Nephrolithiasis/osteoporosis, hypophosphatemic, 2; Familial type 1 and 3 hyperlipoproteinemia;

Phenotypes; CHARGE association; Fuhrmann syndrome; Hypotrichosis-lymphedema-telangiectasia syndrome; Chondrodysplasia Blomstrand type; Acroerythrokeratoderma; Slowed nerve conduction velocity, autosomal dominant; Hereditary cancer-predisposing syndrome; Craniodiaphyseal dysplasia, autosomal dominant; Spinocerebellar ataxia autosomal recessive 1 and 16; Proprotein convertase 1/3 deficiency; D-2-hydroxyglutaric aciduria 2; Hyperekplexia 2 and Hyperekplexia hereditary; Central core disease; Opitz G/BBB syndrome; Cystic fibrosis; Thiel-Behnke corneal dystrophy; Deficiency of bisphosphoglycerate mutase; Mitochondrial short-chain Enoyl-CoA Hydratase 1 deficiency; Ectodermal dysplasia skin fragility syndrome; Wolfram-like syndrome, autosomal dominant; Microcytic anemia; Pyruvate carboxylase deficiency; Leukocyte adhesion deficiency type I and III; Multiple endocrine neoplasia, types land 4; Transient bullous dermolysis of the newborn; Primrose syndrome; Non-small cell lung cancer; Congenital muscular dystrophy; Lipase deficiency combined; COLE-CARPENTER SYNDROME 2; Atrioventricular septal defect and common atrioventricular junction; Deficiency of xanthine oxidase; Waardenburg syndrome type 1, 4C, and 2E (with neurologic involvement); Stickler syndrome, types I(nonsyndromic ocular) and 4; Corneal fragility keratoglobus, blue sclerae and joint hypermobility; Microspherophakia; Chudley-McCullough syndrome; Epidermolysa bullosa simplex and limb girdle muscular dystrophy, simplex with mottled pigmentation, simplex with pyloric atresia, simplex, autosomal recessive, and with pyloric atresia; Rett disorder; Abnormality of neuronal migration; Growth hormone deficiency with pituitary anomalies; Leigh disease; Keratosis palmoplantaris striata 1; Weissenbacher-Zweymuller syndrome; Medium-chain acyl-coenzyme A dehydrogenase deficiency; UDPglucose-4- epimerase deficiency; susceptibility to Autism, X-linked 3; Rhegmatogenous retinal detachment, autosomal dominant; Familial febrile seizures 8; Ulna and fibula absence of with severe limb deficiency; Left ventricular noncompaction 6; Centromeric instability of chromosomes 1,9 and 16 and immunodeficiency; Hereditary diffuse leukoencephalopathy with spheroids; Cushing syndrome; Dopamine receptor d2, reduced brain density of; C-like syndrome; Renal dysplasia, retinal pigmentary dystrophy, cerebellar ataxia and skeletal dysplasia; Ovarian dysgenesis 1; Pierson syndrome; Polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract; Progressive intrahepatic cholestasis; autosomal dominant, autosomal recessive, and X-linked recessive Alport syndromes; Angelman syndrome; Amish infantile epilepsy syndrome; Autoimmune lymphoproliferative syndrome, type Ia;

Hydrocephalus; Marfanoid habitus; Bare lymphocyte syndrome type 2, complementation group E; Recessive dystrophic epidermolysis bullosa; Factor H, VII, X, v and factor viii, combined deficiency of 2, xiii, a subunit, deficiency; Zonular pulverulent cataract 3; Warts, hypogammaglobulinemia, infections, and myelokathexis; Benign hereditary chorea; Deficiency of hyaluronoglucosaminidase; Microcephaly, hiatal hernia and nephrotic syndrome; Growth and mental retardation, mandibulofacial dysostosis, microcephaly, and cleft palate; Lymphedema, hereditary, id; Delayed puberty; Apparent mineralocorticoid excess; Generalized arterial calcification of infancy 2; METHYLMALONIC ACIDURIA, mut(0) TYPE; Congenital heart disease, multiple types, 2; Familial hypoplastic, glomerulocystic kidney; Cerebrooculofacioskeletal syndrome 2; Stargardt disease 1; Mental retardation, autosomal recessive 15, 44, 46, and 5; Prolidase deficiency; Methylmalonic aciduria cblB type, ; Oguchi disease; Endocrine-cerebroosteodysplasia; Lissencephaly 1, 2 (X-linked), 3, 6 (with microcephaly), X-linked; Somatotroph adenoma; Gamstorp- Wohlfart syndrome; Lipid proteinosis; Inclusion body myopathy 2 and 3; Enlarged vestibular aqueduct syndrome; Osteoporosis with pseudoglioma; Acquired long QT syandrome; Phenylketonuria; CHOPS syndrome; Global developmental delay; Bietti crystalline corneoretinal dystrophy; Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia; Congenital erythropoietic porphyria; Atrophia bulborum hereditaria; Paragangliomas 3; Van der Woude syndrome; Aromatase deficiency; Birk Barel mental retardation dysmorphism syndrome; Amyotrophic lateral sclerosis type 5; Methemoglobinemia types I 1 and 2; Congenital stationary night blindness, type 1A, IB, 1C, IE, IF, and 2A; Seizures; Thyroid cancer, follicular; Lethal congenital contracture syndrome 6; Distal hereditary motor neuronopathy type 2B; Sex cord-stromal tumor; Epileptic encephalopathy, childhood-onset, early infantile, 1, 19, 23, 25, 30, and 32; Myofibrillar myopathy 1 and ZASP-related; Cerebellar ataxia infantile with progressive external ophthalmoplegia; Purine-nucleoside phosphorylase deficiency; Forebrain defects; Epileptic encephalopathy Lennox-Gastaut type; Obesity; 4, Left ventricular noncompaction 10; Verheij syndrome; Mowat-Wilson syndrome; Odontotrichomelic syndrome; Patterned dystrophy of retinal pigment epithelium; Lig4 syndrome; Barakat syndrome; IRAK4 deficiency; Somatotroph adenoma; Branched-chain ketoacid dehydrogenase kinase deficiency; Cystinuria; Familial aplasia of the vermis; Succinyl-CoA acetoacetate transferase deficiency; Scapuloperoneal spinal muscular atrophy; Pigmentary retinal dystrophy; Glanzmann

thrombasthenia; Primary open angle glaucoma juvenile onset 1; Aicardi Goutieres syndromes 1, 4, and 5; Renal dysplasia; Intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies; Beaded hair; Short stature, onychodysplasia, facial dysmorphism, and hypotrichosis; Metachromatic leukodystrophy; Cholestanol storage disease; Three M syndrome 2; Leber congenital amaurosis 11, 12, 13, 16, 4, 7, and 9; Mandibuloacral dysplasia with type A or B lipodystrophy, atypical; Meier-Gorlin syndromes land 4; Hypotrichosis 8 and 12; Short QT syndrome 3; Ectodermal dysplasia 1 ib; Anonychia; Pseudohypoparathyroidism type 1A, Pseudopseudohypoparathyroidism; Leber optic atrophy; Bainbridge- Ropers syndrome; Weaver syndrome; Short stature, auditory canal atresia, mandibular hypoplasia, skeletal abnormalities; Deficiency of alpha-mannosidase; Macular dystrophy, vitelliform, adult-onset; Glutaric aciduria, type 1; Gangliosidosis GM1 typel (with cardiac involvenment) 3; Mandibuloacral dysostosis; Hereditary lymphedema type I; Atrial standstill 2; Kabuki make-up syndrome; Bethlem myopathy and Bethlem myopathy 2; Myeloperoxidase deficiency; Fleck corneal dystrophy; Hereditary acrodermatitis enteropathica; Hypobetalipoproteinemia, familial, associated with apob32; Cockayne syndrome type A, ; Hyperparathyroidism, neonatal severe; Ataxia-telangiectasia-like disorder; Pendred syndrome; I blood group system; Familial benign pemphigus; Visceral heterotaxy 5, autosomal; Nephrogenic diabetes insipidus, Nephrogenic diabetes insipidus, X-linked; Minicore myopathy with external ophthalmoplegia; Perry syndrome; hypohidrotic/hair/tooth type, autosomal recessive; Hereditary pancreatitis; Mental retardation and microcephaly with pontine and cerebellar hypoplasia; Glycogen storage disease 0 (muscle), II (adult form), IXa2, IXc, type 1A; Osteopathia striata with cranial sclerosis; Gluthathione synthetase deficiency; Brugada syndrome and Brugada syndrome 4; Endometrial carcinoma; Hypohidrotic ectodermal dysplasia with immune deficiency; Cholestasis, intrahepatic, of pregnancy 3; Bernard-Soulier syndrome, types A1 and A2 (autosomal dominant); Salla disease; Ornithine aminotransferase deficiency; PTEN hamartoma tumor syndrome; Distichiasis- lymphedema syndrome; Corticosteroid-binding globulin deficiency; Adult neuronal ceroid lipofuscinosis; Dejerine-Sottas disease; Tetraamelia, autosomal recessive; Senior-Loken syndrome 4 and 5, ; Glutaric acidemia IIA and IIB; Aortic aneurysm, familial thoracic 4, 6, and 9; Hyperphosphatasia with mental retardation syndrome 2, 3, and 4; Dyskeratosis congenita X-linked; Arthrogryposis, renal dysfunction, and cholestasis 2; Bannayan-Riley-Ruvalcaba syndrome; 3- Methylglutaconic aciduria; Isolated 17,20-lyase

deficiency; Gorlin syndrome; Hand foot uterus syndrome; Tay-Sachs disease, B1 variant, Gm2-gangliosidosis (adult), Gm2-gangliosidosis (adult-onset); Dowling-degos disease 4; Parkinson disease 14, 15, 19 (juvenile-onset), 2, 20 (early-onset), 6, (autosomal recessive early-onset, and 9; Ataxia, sensory, autosomal dominant; Congenital microvillous atrophy; Myoclonic- Atonic Epilepsy; Tangier disease;2- methyl-3-hydroxybutyric aciduria; Familial renal hypouricemia; Schizencephaly; Mitochondrial DNA depletion syndrome 4B, MNGIE type; Feingold syndrome 1; Renal carnitine transport defect; Familial hypercholesterolemia; Townes-Brocks-branchiootorenal-like syndrome; Griscelli syndrome type 3; Meckel-Gruber syndrome; Bullous ichthyosiform erythroderma; Neutrophil immunodeficiency syndrome; Myasthenic Syndrome, Congenital, 17, 2A (slow-channel), 4B (fast-channel), and without tubular aggregates; Microvascular complications of diabetes 7; McKusick Kaufman syndrome; Chronic granulomatous disease, autosomal recessive cytochrome b-positive, types 1 and 2; Arginino succinate lyase deficiency; Mitochondrial phosphate carrier and pyruvate carrier deficiency; Lattice corneal dystrophy Type III; Ectodermal dysplasia-syndactyly syndrome 1; Hypomyelinating leukodystrophy 7; Mental retardation, autosomal dominant 12, 13, 15, 24, 3, 30, 4, 5, 6, and 9; Generalized epilepsy with febrile seizures plus, types 1 and 2; Psoriasis susceptibility 2; Frank Ter Haar syndrome; Thoracic aortic aneurysms and aortic dissections; Crouzon syndrome; Granulosa cell tumor of the ovary; Epidermolytic palmoplantar keratoderma; Leri Weill dyschondrosteosis; 3 beta-Hydroxysteroid dehydrogenase deficiency; Familial restrictive cardiomyopathy 1; Autosomal dominant progressive external ophthalmoplegia with mitochondrial DNA deletions 1 and 3; Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis; Hajdu-Cheney syndrome; Pigmented nodular adrenocortical disease, primary, 1; Episodic pain syndrome, familial, 3; Dejerine-Sottas syndrome, autosomal dominant; FG syndrome and FG syndrome 4; Dendritic cell, monocyte, B lymphocyte, and natural killer lymphocyte deficiency; Hypothyroidism, congenital, nongoitrous, 1; Miller syndrome; Nemaline myopathy 3 and 9; Oligodontia- colorectal cancer syndrome; Cold-induced sweating syndrome 1; Van Buchem disease type 2; Glaucoma 3, primary congenital, d; Citrullinemia type I and II; Nonaka myopathy; Congenital muscular dystrophy due to partial LAMA2 deficiency; Myoneural gastrointestinal encephalopathy syndrome; Leigh syndrome due to mitochondrial complex I deficiency; Medulloblastoma; Pyruvate dehydrogenase El -alpha deficiency; Carcinoma of colon; Nance-Horan syndrome; Sandhoff disease, adult and infantil types; Arthrogryposis renal

dysfunction cholestasis syndrome; Autosomal recessive hypophosphatemic bone disease; Doyne honeycomb retinal dystrophy; Spinocerebellar ataxia 14, 21, 35, 40, and 6; Lewy body dementia; RRM2B -related mitochondrial disease; Brody myopathy; Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 2; Usher syndrome, types 1, IB, ID, 1G, 2A, 2C, and 2D; hypocalcification type and hypomaturation type, IIA1 Amelogenesis imperfecta; Pituitary hormone deficiency, combined 1, 2, 3, and 4; Cushing symphalangism; Renal tubular acidosis, distal, autosomal recessive, with late-onset sensorineural hearing loss, or with hemolytic anemia; Infantile nephronophthisis; Juvenile polyposis syndrome; Sensory ataxic neuropathy, dysarthria, and ophthalmoparesis; Deficiency of 3-hydroxyacyl-CoA dehydrogenase; Parathyroid carcinoma; X-linked agammaglobulinemia; Megaloblastic anemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness; Multiple sulfatase deficiency; Neurodegeneration with brain iron accumulation 4 and 6; Cholesterol monooxygenase (side-chain cleaving) deficiency; hemolytic anemia due to Adenylosuccinate lyase deficiency; Myoclonus with epilepsy with ragged red fibers; Pitt- Hopkins syndrome; Multiple pterygium syndrome Escobar type; Homocystinuria-Megaloblastic anemia due to defect in cobalamin metabolism, cblE complementation type; Cholecystitis; Spherocytosis types 4 and 5; Multiple congenital anomalies; Xeroderma pigmentosum, complementation group b, group D, group E, and group G; Leiner disease; Groenouw corneal dystrophy type I; Coenzyme Q10 deficiency, primary 1, 4, and 7; Distal spinal muscular atrophy, congenital nonprogressive; Warburg micro syndrome 2 and 4; Bile acid synthesis defect, congenital, 3; Acth-independent macronodular adrenal hyperplasia 2; Acrocapitofemoral dysplasia; Paget disease of bone, familial; Severe neonatal-onset encephalopathy with microcephaly; Zimmermann-Laband syndrome and Zimmermann-Laband syndrome 2; Reifenstein syndrome; Familial hypokalemia-hypomagnesemia; Photosensitive trichothiodystrophy; Adult junctional epidermolysis bullosa; Lung cancer; Freeman-Sheldon syndrome; Hyperinsulinism-hyperammonemia syndrome; Posterior polar cataract type 2; Sclerocornea, autosomal recessive; Juvenile GM>1< gangliosidosis; Cohen syndrome, ; Hereditary Paraganglioma- Pheochromocytoma Syndromes; Neonatal insulin-dependent diabetes mellitus; Hypochondrogenesis; Floating-Harbor syndrome; Cutis laxa with osteodystrophy and with severe pulmonary, gastrointestinal, and urinary abnormalities; Congenital contractures of the limbs and face, hypotonia, and developmental delay; Dyskeratosis congenita autosomal dominant and autosomal dominant, 3; Histiocytic medullary reticulosis; Costello syndrome;

Immunodeficiency 15, 16, 19, 30, 31C, 38, 40, 8, due to defect in cd3-zeta, with hyper IgM type 1 and 2, and X-Linked, with magnesium defect, Epstein-Barr vims infection, and neoplasia; Atrial septal defects 2, 4, and 7 (with or without atrioventricular conduction defects); GTP cyclohydrolase I deficiency; Talipes equinovarus; Phosphoglycerate kinase 1 deficiency; Tuberous sclerosis 1 and 2; Autosomal recessive congenital ichthyosis 1, 2, 3, 4A, and 4B; and Familial hypertrophic cardiomyopathy 1, 2, 3, 4, 7, 10, 23 and 24.

Indications by tissue

Additional suitable diseases and disorders that can be treated by the systems and methods provided herein include, without limitation, diseases of the central nervous system (CNS) (see exemplary diseases and affected genes in **Table 13**), diseases of the eye (see exemplary diseases and affected genes in **Table 14**), diseases of the heart (see exemplary diseases and affected genes in **Table 15**), diseases of the hematopoietic stem cells (HSC) (see exemplary diseases and affected genes in **Table 16**), diseases of the kidney (see exemplary diseases and affected genes in **Table 17**), diseases of the liver (see exemplary diseases and affected genes in **Table 18**), diseases of the lung (see exemplary diseases and affected genes in **Table 19**), diseases of the skeletal muscle (see exemplary diseases and affected genes in **Table 20**), and diseases of the skin (see exemplary diseases and affected genes in **Table 21**). **Table 22** provides exemplary protective mutations that reduce risks of the indicated diseases. In some embodiments, a Gene Writer system described herein is used to treat an indication of any of Tables 13-21. In some embodiments, the GeneWriter system modifies a target site in genomic DNA in a cell, wherein the target site is in a gene of any of Tables 13-21, e.g., in a subject having the corresponding indication listed in any of Tables 13-21. In some embodiments, the GeneWriter corrects a mutation in the gene. In some embodiments, the GeneWriter inserts a sequence that had been deleted from the gene (e.g., through a disease-causing mutation). In some embodiments, the GeneWriter deletes a sequence that had been duplicated in the gene (e.g., through a disease-causing mutation). In some embodiments, the GeneWriter replaces a mutation (e.g., a disease-causing mutation) with the corresponding wild-type sequence. In some embodiments, the mutation is a substitution, insertion, deletion, or inversion.

Table 13. CNS diseases and genes affected.

Disease	Gene Affected
Alpha-mannosidosis	MAN2B1

Ataxia-telangiectasia	ATM
CADASIL	NOTCH3
Canavan disease	ASPA
Carbamoyl-phosphate synthetase 1 deficiency	CPS1
CLN1 disease	PPT1
CLN2 Disease	TPP1
CLN3 Disease (Juvenile neuronal ceroid lipofuscinosis, Batten Disease)	CLN3
Coffin-Lowry syndrome	RPS6KA3
Congenital myasthenic syndrome 5	COLQ
Cornelia de Lange syndrome (NIPBL)	NIPBL
Cornelia de Lange syndrome (SMC1A)	SMC1A
Dravet syndrome (SCN1A)	SCN1A
Glycine encephalopathy (GLDC)	GLDC
GM1 gangliosidosis	GLB1
Huntington's Disease	HTT
Hydrocephalus with stenosis of the aqueduct of Sylvius	L1CAM
Leigh Syndrome	SURF1
Metachromatic leukodystrophy (ARSA)	ARSA
MPS type 2	IDS
MPS type 3	Type 3a: SGSH Type 3b: NAGLU
Mucolipidosis IV	MCOLN1
Neurofibromatosis Type 1	NF1
Neurofibromatosis type 2	NF2
Pantothenate kinase-associated neurodegeneration	PANK2
Pyridoxine-dependent epilepsy	ALDH7A1
Rett syndrome (MECP2)	MECP2

Sandhoff disease	HEXB
Semantic dementia (Frontotemporal dementia)	MAPT
Spinocerebellar ataxia with axonal neuropathy (Ataxia with Oculomotor Apraxia)	SETX
Tay-Sachs disease	HEXA
X-linked Adrenoleukodystrophy	ABCD1

Table 14. Eye diseases and genes affected.

Disease	Gene Affected
Achromatopsia	CNGB3
Amaurosis Congenita (LCA1)	GUCY2D
Amaurosis Congenita (LCA10)	CEP290
Amaurosis Congenita (LCA2)	RPE65
Amaurosis Congenita (LCA8)	CRB1
Choroideremia	CHM
Cone Rod Dystrophy (ABCA4)	ABCA4
Cone Rod Dystrophy (CRX)	CRX
Cone Rod Dystrophy (GUCY2D)	GUCY2D
Cystinosis, Ocular Nonnephropathic	CTNS
Lattice corneal dystrophy type I	TGFBI
Macular Corneal Dystrophy (MCD)	CHST6
Optic Atrophy	OPA1
Retinitis Pigmentosa (AR)	USH2A
Retinitis Rigmentosa (AD)	RHO
Stargardt Disease	ABCA4
Vitelliform Macular Dystrophy	BEST1; PRPH2

Table 15. Heart diseases and genes affected.

Disease	Gene Affected
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2
Barth syndrome	TAZ
Becker muscular dystrophy	DMD
Brugada syndrome	SCN5A

Catecholaminergic polymorphic ventricular tachycardia (RYR2)	RYR2
Dilated cardiomyopathy (LMNA)	LMNA
Dilated cardiomyopathy (TTN)	TTN
Duchenne muscular dystrophy	DMD
Emery-Dreifuss Muscular Dystrophy Type I	EMD
Familial hypertrophic cardiomyopathy	MYH7
Familial hypertrophic cardiomyopathy	MYBPC3
Jervell Lange-Nielsen syndrome	KCNQ1
LCHAD deficiency	HADHA
Limb-girdle muscular dystrophy type 1B (Emery-Dreifuss EDMD2)	LMNA
Limb-girdle muscular dystrophy, type 2D	SGCA
Long QT syndrome 1 (Romano Ward)	KCNQ1
Disease	**Gene Affected**
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2
Barth syndrome	TAZ
Becker muscular dystrophy	DMD
Brugada syndrome	SCN5A
Catecholaminergic polymorphic ventricular tachycardia (RYR2)	RYR2
Dilated cardiomyopathy (LMNA)	LMNA
Dilated cardiomyopathy (TTN)	TTN
Duchenne muscular dystrophy	DMD
Emery-Dreifuss Muscular Dystrophy Type I	EMD
Familial hypertrophic cardiomyopathy	MYH7
Familial hypertrophic cardiomyopathy	MYBPC3
Jervell Lange-Nielsen syndrome	KCNQ1

LCHAD deficiency	HADHA
Limb-girdle muscular dystrophy type 1B (Emery-Dreifuss EDMD2)	LMNA
Limb-girdle muscular dystrophy, type 2D	SGCA
Long QT syndrome 1 (Romano Ward)	KCNQ1

Table 16. HSC diseases and genes affected.

Disease	Gene Affected
ADA-SCID	ADA
Adrenoleukodystrophy (CALD)	ABCD1
Alpha-mannosidosis	MAN2B1
Chronic granulomatous disease	CYBB; CYBA; NCF1; NCF2; NCF4
Common variable immunodeficiency	TNFRSF13B
Fanconi anemia	FANCA; FANCC; FANCG
Gaucher disease	GBA
Globoid cell leukodystrophy (Krabbe disease)	GALC
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
IL-7R SCID	IL7R
JAK-3 SCID	JAK3
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	TCIRG1; Many genes implicated
Metachromatic leukodystrophy	ARSA; PSAP
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS7	GUSB
Mucolipidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1

Paroxysmal Nocturnal Hemoglobinuria	PIGA
Pompe disease	GAA
Pyruvate kinase deficiency (PKD)	PKLR
RAG 1/2 Deficiency (SCID with granulomas)	RAG1/RAG2
Severe Congenital Neutropenia	ELANE; HAX1
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB
Wiskott-Aldrich Syndrome	WAS
X-linked agammaglobulinemia	BTK
X-linked SCID	IL2RG
Disease	**Gene Affected**
ADA-SCID	ADA
Adrenoleukodystrophy (CALD)	ABCD1
Alpha-mannosidosis	MAN2B1
Chronic granulomatous disease	CYBB; CYBA; NCF1; NCF2; NCF4
Common variable immunodeficiency	TNFRSF13B
Fanconi anemia	FANCA; FANCC; FANCG
Gaucher disease	GBA
Globoid cell leukodystrophy (Krabbe disease)	GALC
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
IL-7R SCID	IL7R
JAK-3 SCID	JAK3
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	TCIRG1; Many genes implicated
Metachromatic leukodystrophy	ARSA; PSAP
MPS 1S (Scheie syndrome)	IDUA

MPS2	IDS
MPS7	GUSB
Mucolipidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1
Paroxysmal Nocturnal Hemoglobinuria	PIGA
Pompe disease	GAA
Pyruvate kinase deficiency (PKD)	PKLR
RAG 1/2 Deficiency (SCID with granulomas)	RAG1/RAG2
Severe Congenital Neutropenia	ELANE; HAX1
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB
Wiskott-Aldrich Syndrome	WAS
X-linked agammaglobulinemia	BTK
X-linked SCID	IL2RG

Table 17. Kidney diseases and genes affected.

Disease	**Gene Affected**
Alport syndrome	COL4A5
Autosomal dominant polycystic kidney disease (PKD1)	PKD1
Autosomal dominant polycystic kidney disease (PKD2)	PDK2
Autosomal dominant tubulointerstitial kidney disease (MUC1)	MUC1
Autosomal dominant tubulointerstitial kidney disease (UMOD)	UMOD
Autosomal recessive polycystic kidney disease	PKHD1

Congenital nephrotic syndrome	NPHS2
Cystinosis	CTNS

Table 18. Liver diseases and genes affected.

Disease	Gene Affected
Acute intermittent porphyria	HMBS
Alagille syndrome	JAG1
Alpha-1-antitrypsin deficiency	SERPINA1
Carbamoyl phosphate synthetase I deficiency	CPS1
Citrullinemia I	ASS1
Crigler-Najjar	UGT1A1
Fabry	LPL
Familial chylomicronemia syndrome	GLA
Gaucher	GBE1
GSD IV	GBA
Heme A	F8
Heme B	F9
Hereditary amyloidosis (hTTR)	TTR
Hereditary angioedema	SERPING1 (KLKB1 for CRISPR)
HoFH	LDLRAP1
Hypercholesterolemia	PCSK9
Methylmalonic acidemia	MMUT
MPS II	IDS
MPS III	Type IIIa: SGSH Type IIIb: NAGLU

	Type IIIc: HGSNAT Type IIId: GNS
MPS IV	Type IVA: GALNS Type IVB: GLB1
MPS VI	ARSB
MSUD	Type Ia: BCKDHA Type Ib: BCKDHB Type II: DBT
OTC Deficiency	OTC
Polycystic Liver Disease	PRKCSH
Pompe	GAA
Primary Hyperoxaluria 1	AGXT (HAO1 or LDHA for CRISPR)
Progressive familial intrahepatic cholestasis type 1	ATP8B1
Progressive familial intrahepatic cholestasis type 2	ABCB11
Progressive familial intrahepatic cholestasis type 3	ABCB4
Propionic acidemia	PCCB; PCCA
Wilson's Disease	ATP7B
Glycogen storage disease, Type 1a	G6PC
Glycogen storage disease, Type IIIb	AGL
Isovaleric acidemia	IVD
Wolman disease	LIPA

Table 19. Lung diseases and genes affected.

Disease	Gene Affected
Alpha-1 antitrypsin deficiency	SERPINA1
Cystic fibrosis	CFTR
Primary ciliary dyskinesia	DNAI1
Primary ciliary dyskinesia	DNAH5
Primary pulmonary hypertension I	BMPR2
Surfactant Protein B (SP-B) Deficiency (pulmonary surfactant metabolism dysfunction 1)	SFTPB

Table 20. Skeletal muscle diseases and genes affected.

Disease	**Gene Affected**
Becker muscular dystrophy	DMD
Becker myotonia	CLCN1
Bethlem myopathy	COL6A2
Centronuclear myopathy, X-linked (myotubular)	MTM1
Congenital myasthenic syndrome	CHRNE
Duchenne muscular dystrophy	DMD
Emery-Dreifuss muscular dystrophy, AD	LMNA
Facioscapulohumeral Muscular Dystrophy	DUX4 - D4Z4 chromosomal region
Hyperkalemic periodic paralysis	SCN4A
Hypokalemic periodic paralysis	CACNA1S
Limb-girdle muscular dystrophy 2A	CAPN3
Limb-girdle muscular dystrophy 2B	DYSF
Limb-girdle muscular dystrophy, type 2D	SGCA
Miyoshi muscular dystrophy 1	DYSF
Paramyotonia congenita	SCN4A
Thomsen myotonia	CLCN1
VCP myopathy (IBMPFD) 1	VCP
Disease	**Gene Affected**
Becker muscular dystrophy	DMD
Becker myotonia	CLCN1
Bethlem myopathy	COL6A2
Centronuclear myopathy, X-linked (myotubular)	MTM1

Congenital myasthenic syndrome	CHRNE
Duchenne muscular dystrophy	DMD
Emery-Dreifuss muscular dystrophy, AD	LMNA
Facioscapulohumeral Muscular Dystrophy	DUX4 - D4Z4 chromosomal region
Hyperkalemic periodic paralysis	SCN4A
Hypokalemic periodic paralysis	CACNA1S
Limb-girdle muscular dystrophy 2A	CAPN3
Limb-girdle muscular dystrophy 2B	DYSF
Limb-girdle muscular dystrophy, type 2D	SGCA
Miyoshi muscular dystrophy 1	DYSF
Paramyotonia congenita	SCN4A
Thomsen myotonia	CLCN1
VCP myopathy (IBMPFD) 1	VCP

Table 21. Skin diseases and genes affected.

Disease	Gene Affected
Epidermolysis Bullosa Dystrophica Dominant	COL7A1
Epidermolysis Bullosa Dystrophica Recessive (Hallopeau-Siemens)	COL7A1
Epidermolysis Bullosa Junctional	LAMB3
Epidermolysis Bullosa Simplex	KRT5; KRT14
Epidermolytic Ichthyosis	KRT1; KRT10
Hailey-Hailey Disease	ATP2C1
Lamellar Ichthyosis/Nonbullous Congenital Ichthyosiform Erythroderma (ARCI)	TGM1
Netherton Syndrome	SPINK5

Table 22. Exemplary protective mutations that reduce disease risk.

Disease	Gene	Exemplary Protective Mutation
Alzheimer's	APP	A673T

Parkinson's	SGK1	
Diabetes (Type II)	SLC30A8	p.Arg138X; p.Lys34SerfsX50
Cardiovascular Disease	PCSK9	R46L
Cardiovascular Disease	ASGR1	NM_001671.4, c.284-36_283+33delCTGGGGCTGGGG; (SEQ ID NO: 1622) NP_001662.1, p.W158X
Cardiovascular Disease	NPC1L1	p.Arg406X
Cardiovascular Disease	APOC3	R19X; IVS2+1G→A; A43T
Cardiovascular Disease	LPA	
Cardiovascular Disease	ANGPTL4	E40K
Cardiovascular Disease	ANGPTL3	p.Ser17Ter; p.Asn121fs; p.Asn147fs; c.495+6T→C
HIV infection	CCR5	CCR5-delta32

Pathogenic mutations

In some embodiments, the systems or methods provided herein can be used to correct a pathogenic mutation. The pathogenic mutation can be a genetic mutation that increases an individual's susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation is a disease-causing mutation in a gene associated with a disease or disorder. In some embodiments, the systems or methods provided herein can be used to revert the pathogenic mutation to its wild-type counterpart. In some embodiments, the systems or methods provided herein can be used to change the pathogenic mutation to a sequence not causing the disease or disorder.

Table 23 provides exemplary indications (column 1), underlying genes (column 2), and pathogenic mutations that can be corrected using the systems or methods described herein (column 3).

Table 23. Indications, genes, and causitive pathogenic mutations.

Disease	**Gene**	**Pathogenic Mutation#**
Achromatopsia	CNGB3	1148delC
Alpha-1 Antitrypsin Deficiency	SERPINAI	E342K
Alpha-1 Antitrypsin Deficiency	SERPINAI	E342K
Alpha-1 Antitrypsin Deficiency	SERPINAI	R48C (R79C)

Disease	Gene	Pathogenic Mutation#
Amaurosis Congenita (LCA10)	CEP290	2991+1655A>G
Andersen– Tawil syndrome	KCNJ2	R218W
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2	c.235C>T
associated with congenital factor XI deficiency	F11	E117*
associated with congenital factor XI deficiency	F11	F283L
ATTR amyloidosis	TTR	V50M/N30M
autosomal dominant deafness	COCH	G88E
autosomal dominant deafness	TECTA	Y1870C
autosomal dominant Parkinson's disease	SNCA	A53T
autosomal dominant Parkinson's disease	SNCA	A30P
Autosomal dominant rickets	FGF23	R176Q
autosomal recessive deafness	CX30	T5M
autosomal recessive deafness	DFNB59	R183W
autosomal recessive deafness	TMC1	Y182C
autosomal recessive hypercholesterolemia	ARH	Q136*
Blackfan-Diamond anemia	RPS19	R62Q
blue-cone monochromatism	OPN1LW	C203R
Brugada syndrome	SCN5A	E1784K
CADASIL syndrome	NOTCH3 gene	R90C
CADASIL syndrome	NOTCH3 gene	R141C
Canavan disease	ASPA	E285A
Canavan disease	ASPA	Y231X
Canavan disease	ASPA	A305E
carnitine palmitoyltransferase II deficiency	CPT2	S113L
choroideremia	CHM	R293*
choroideremia	CHM	R270*
choroideremia	CHM	A117A
Citrullinemia Type I	ASS	G390R
classic galactosemia	GALT	Q188R
classic homocysteinuria	CBS	T191M
classic homocysteinuria	CBS	G307S
CLN2 Disease	TPP1	c.509–1 G>C
CLN2 Disease	TPP1	c.622 C>T
CLN2 Disease	TPP1	c.851 G>T
cone-rod dystrophy	GUCY2D	R838C
congenital factor V deficiency	F5	R506Q
congenital factor V deficiency	F5	R534Q
congenital factor VII deficiency	F7	A294V
congenital factor VII deficiency	F7	C310F
congenital factor VII deficiency	F7	R304Q
congenital factor VII deficiency	F7	QI00R
Creutzfeldt- Jakob disease (CJD)	PRNP	E200K
Creutzfeldt- Jakob disease (CJD)	PRNP	M129V
Creutzfeldt- Jakob disease (CJD)	PRNP	P102L
Creutzfeldt- Jakob disease (CJD)	PRNP	D178N
cystic fibrosis	CFTR	G551D
cystic fibrosis	CFTR	W1282*

Disease	Gene	Pathogenic Mutation#
cystic fibrosis	CFTR	R553*
cystic fibrosis	CFTR	R117H
cystic fibrosis	CFTR	delta F508
cystinosis	CTNS	W138*
Darier disease	ATP2A2	N767S
Darier disease	ATP2A2	N767S
Darier disease	ATP2A2	N767S
Epidermolysis Bullosa Junctional	LAMB3	R42X
Epidermolysis Bullosa Junctional	LAMB3	R635X
familial amyotrophic lateral sclerosis (ALS)	SOD1	A4V
familial amyotrophic lateral sclerosis (ALS)	SOD1	H46R
familial amyotrophic lateral sclerosis (ALS)	SOD1	G37R
Gaucher disease	GBA	N370S
Gaucher disease	GBA	N370S
Gaucher disease	GBA	L444P
Gaucher disease	GBA	L444P
Gaucher disease	GBA	L483P
glutaryl-CoA dehydrogenase deficiency	GCDH	R138G
glutaryl-CoA dehydrogenase deficiency	GCDH	M263V
glutaryl-CoA dehydrogenase deficiency	GCDH	R402W
glycine encephalopathy	GLDC	A389V
glycine encephalopathy	GLDC	G771R
glycine encephalopathy	GLDC	T269M
hemophilia A	F8	R2178C
hemophilia A	F8	R550C
hemophilia A	F8	R2169H
hemophilia A	F8	R1985Q
hemophilia B	F9	T342M
hemophilia B	F9	R294Q
hemophilia B	F9	R43Q
hemophilia B	F9	R191H
hemophilia B	F9	G106S
hemophilia B	F9	A279T
hemophilia B	F9	R75*
hemophilia B	F9	R294*
hemophilia B	F9	R379Q
Hereditary antithrombin deficiency type I	SERPINCI	R48C (R79C)
hereditary chronic pancreatitis	PRSS1	R122H
Hunter syndrome	IDS	R88C
Hunter syndrome	IDS	G374G
Hurler syndrome (MPS1)	IDUA	Q70*
Hurler syndrome (MPS1)	IDUA	W402*
Hyperkalemic periodic paralysis	SCN4A	T704M
Hyperkalemic periodic paralysis	SCN4A	M1592V
Hyperkalemic periodic paralysis	CACNA1S	p.Arg528X
Hyperkalemic periodic paralysis	CACNA1S	p.Arg1239
intermittent porphyria	HMBS	Rl73W
isolated agammaglobulinemia	E47	E555K

Disease	Gene	Pathogenic Mutation#
Lattice corneal dystrophy type I	TGFBI	Arg124Cys
LCHAD deficiency	HADHA	Glu474Gln
Leber congenital amaurosis 2	RPE65	R44*
Leber congenital amaurosis 2	RPE65	IVS1
Leber congenital amaurosis 2	RPE65	G-A, +5
Lesch-Nyhan syndrome	HPRTI	R51*
Lesch-Nyhan syndrome	HPRTI	R170*
Limb-girdle muscular dystrophy, type 2D	SGCA	Arg77Cys
Marteauz- Lamy Syndrome (MSPVI)	ARSB	Y210C
Mediterranean G6PD deficiency	G6PD	S188D
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
Meesmann epithelial corneal dystrophy	KRT12	L132P
metachromatic leukodystrophy	ARSA	P426L
metachromatic leukodystrophy	ARSA	c.459+1G>A
Morquio Syndrome (MPSIVA)	GALNS	R386C
Mucolipidosis IV	MCOLN1	406-2A>G
Mucolipidosis IV	MCOLN1	511_6943del
Neimann-Pick disease type A	SMPDI	L302P
Neuronal ceroid lipofuscinosis (NCL)	CLN2	R208*
neuronal ceroid lipofuscinosis 1	PPT1	R151*
Parkinson's disease	LRRK2	G2019S
Pendred syndrome	PDS	T461P
Pendred syndrome	PDS	L236P
Pendred syndrome	PDS	c.1001+1G>A
Pendred syndrome	PDS	IVS8, +1 G>A,
phenylketonuria	PAH	R408W
phenylketonuria	PAH	I65T
phenylketonuria	PAH	R261Q
phenylketonuria	PAH	IVS10-11G>A
phenylketonuria	PCDH15	R245*
phenylketonuria	PCDH15	R245*
Pompe disease	GAA	c.-32-13T>G
Primary ciliary dyskinesia	DNAI1	IVS1+2_3insT
Primary ciliary dyskinesia	DNAH5	10815delT
primary hypoxaluria	AGXT	G170R
Progressive familial intrahepatic cholestasis type 2	ABCB11	D482G (c.1445A>G)
Progressive familial intrahepatic cholestasis type 2	ABCB11	E297G
Propionic acidemia	PCCB; PCCA	c.1218_1231del14ins12
pseudoxanthoma elasticum	ABCC6	R1141*
Pyruvate kinase deficiency (PKD)	PKLR	c.1456c->T
retinitis pigmentos	USH2a	C759F
retinitis pigmentosa	IMPDHI	D226N
retinitis pigmentosa	PDE6A	V685M
retinitis pigmentosa	PDE6A	D670G
retinitis pigmentosa	PRPF3	T494M
retinitis pigmentosa	PRPF8	H2309R

Disease	Gene	Pathogenic Mutation[#]
retinitis pigmentosa	RHO	P23H
retinitis pigmentosa	RHO	P347L
retinitis pigmentosa	RHO	P347L
retinitis pigmentosa	RHO	D190N
retinitis pigmentosa	RPI	R667*
retinitis pigmentosa/Usher syndrome type 1C	USH1C	V72V
Rett syndrome	MECP2	R106W
Rett syndrome	MECP2	R133C
Rett syndrome	MECP2	R306C
Rett syndrome	MECP2	R168*
Rett syndrome	MECP2	R255*
Sanfilippo syndrome A (MPSIIIA)	SGSH	R74C
Sanfilippo syndrome A (MPSIIIA)	SGSH	R245H
Sanfilippo syndrome B (MPSIIIB)	NAGLU	R297*
Sanfilippo syndrome B (MPSIIIB)	NAGLU	Y140C
severe combined immunodeficiency	ADA	G216R
severe combined immunodeficiency	ADA	G216R
severe combined immunodeficiency	ADA	Q3*
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E26K
sickle cell disease	HBB	E26K
sickle cell disease	HBB	E7K
sickle cell disease	HBB	c.-138C>T
sickle cell disease	HBB	IVS2
sickle cell disease	HBB	654 C>T
Sly Syndrome (MPSVII)	GUSB	L175F
Stargardt disease	ABCA4	A1038V
Stargardt disease	ABCA4	A1038V
Stargardt disease	ABCA4	L541P
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	c.2588G>C
Stargardt disease	ABCA4	c.5461-10 T>C
Stargardt disease	ABCA4	c.5714+5G>A
Tay Sachs	HEXA	InsTATC1278
tyrosinemia type 1	FAH	P261L
Usher syndrome type 1F	PCDH15	R245*
variegate porphyria	PPOX	R59W
VCP myopathy (IBMPFD) 1	VCP	R1555X
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	R83C
Wilson's Disease	ATP7B	E297G

Disease	Gene	Pathogenic Mutation[#]
X-linked myotubular myopathy	MTMI	c.1261-10A>G
X-linked retinoschisis	RS1	R102W
X-linked retinoschisis	RS1	R141C

[#]: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations. * means a stop codon.

Compensatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a compensatory edit. In some embodiments, the compensatory edit is at a position of a gene associated with a disease or disorder, which is different from the position of a disease-causing mutation. In some embodiments, the compensatory mutation is not in the gene containing the causitive mutation. In some embodiments, the compensatory edit can negate or compensate for a disease-causing mutation. In some embodiments, the compensatory edit can be introduced by the systems or methods provided herein to suppress or reverse the mutant effect of a disease-causing mutation.

Table 24 provides exemplary indications (column 1), genes (column 2), and compensatory edits that can be introduced using the systems or methods described herein (column 3). In some embodiments, the compensatory edits provided in Table 24 can be introduced to suppress or reverse the mutant effect of a disease-causing mutation.

Table 24. Indications, genes, compensatory edits, and exemplary design features.

Disease	Gene	Nucleotide Change[#]
Alpha-1 Antitrypsin Deficiency	SERPINA1	F51L
Alpha-1 Antitrypsin Deficiency	SERPINA1	M374I
Alpha-1 Antitrypsin Deficiency	SERPINA1	A348V/A347V
Alpha-1 Antitrypsin Deficiency	SERPINA1	K387R
Alpha-1 Antitrypsin Deficiency	SERPINA1	T59A
Alpha-1 Antitrypsin Deficiency	SERPINA1	T68A
ATTR amyloidosis	TTR	A108V
ATTR amyloidosis	TTR	R104H
ATTR amyloidosis	TTR	T119M
Cystic fibrosis	CFTR	R555K
Cystic fibrosis	CFTR	F409L
Cystic fibrosis	CFTR	F433L

Disease	Gene	Nucleotide Change[#]
Cystic fibrosis	CFTR	H667R
Cystic fibrosis	CFTR	Rl070W
Cystic fibrosis	CFTR	R29K
Cystic fibrosis	CFTR	R553Q
Cystic fibrosis	CFTR	1539T
Cystic fibrosis	CFTR	G550E
Cystic fibrosis	CFTR	F429S
Cystic fibrosis	CFTR	Q637R
Sickle cell disease	HBB	A70T
Sickle cell disease	HBB	A70V
Sickle cell disease	HBB	L88P
Sickle cell disease	HBB	F85L and/or F85P
Sickle cell disease	HBB	E22G
Sickle cell disease	HBB	G16D and/or G16N

[#]: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations.

Regulatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a regulatory edit. In some embodiments, the regulatory edit is introduced to a regulatory sequence of a gene, for example, a gene promoter, gene enhancer, gene repressor, or a sequence that regulates gene splicing. In some embodiments, the regulatory edit increases or decreases the expression level of a target gene. In some embodiments, the target gene is the same as the gene containing a disease-causing mutation. In some embodiment, the target gene is different from the gene containing a disease-causing mutation. For example, the systems or methods provided herein can be used to upregulate the expression of fetal hemoglobin by introducing a regulatory edit at the promoter of bcl11a, thereby treating sickle cell disease.

Table 25 provides exemplary indications (column 1), genes (column 2), and regulatory edits that can be introduced using the systems or methods described herein (column 3).

Table 25. Indications, genes, and compensatory regulatory edits.

Disease	Gene	Nucleotide Change[#]
homozygous familial hypercholesterolaemia	LDLR	c.81C>T

Disease	Gene	Nucleotide Change[#]
Porphyrias	ALAS1	c.3G>A
Porphyrias	ALAS1	c.2T>C
Porphyrias	ALAS1	c.46C>T
Porphyrias	ALAS1	c.91C>T
Porphyrias	ALAS1	c.91C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.229C>T
Porphyrias	ALAS1	c.247C>T
Porphyrias	ALAS1	c.247C>T
Porphyrias	ALAS1	c.250C>T
Porphyrias	ALAS1	c.250C>T
Porphyrias	ALAS1	c.340C>T
Porphyrias	ALAS1	c.340C>T
Porphyrias	ALAS1	c.349C>T
Porphyrias	ALAS1	c.391C>T
Porphyrias	ALAS1	c.391C>T
Porphyrias	ALAS1	c.403C>T
Porphyrias	ALAS1	c.403C>T
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.200-2A>G
Porphyrias	ALAS1	c.427+ 1G>A
Porphyrias	ALAS1	c.427+2T>C
Porphyrias	ALAS1	c.1165+1G >A
Porphyrias	ALAS1	c. l 165+2T>C
Porphyrias	ALAS1	c.1166-lA>G
Porphyrias	ALAS1	c.133 l-2A>G

Disease	Gene	Nucleotide Change[#]
sickle cell disease	BCL11A	c.386-24278G>A
sickle cell disease	BCL11A	c.386-24983T>C
sickle cell disease	HBG1	c.-167C>T
sickle cell disease	HBG1	c.-170G>A
sickle cell disease	HBG1	c.-249C>T
sickle cell disease	HBG2	c.-211C>T
sickle cell disease	HBG2	c.-228T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	c. -90 BCLllA Binding
sickle cell disease	HBG1/2	c. -90 BCLllA Binding
sickle cell disease	HBG1/2	C. -202 C>T, -201 C>T, -198 T>C, -197 C>T, -196 C>T, -195 C>G
sickle cell disease	HBG1/2	C. -197 C>T,-196 C>T, -195 C>G

#: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations.

Repeat expansion diseases

In some embodiments, the systems or methods provided herein can be used to a repeat expansion disease, for example, a repeat expansion disease provided in **Table 26**. Table 26 provides the indication (column 1), the gene (column 2), minimal repeat sequence of the repeat that is expanded in the condition (column 3), and the location of the repeat relative to the listed gene for each indication (column 4). In some embodiments, the systems or methods provided herein, for example, those comprising Gene Writers, can be used to treat repeat expansion diseases by resetting the number of repeats at the locus according to a customized RNA template (see, e.g., Example 24).

Table 26. Exemplary repeat expansion diseases, genes, causal repeats, and repeat locations.

Disease	Gene	Causal repeat	Repeat location
myotonic dystrophy 1	DMPK/DM1	CTG	3' UTR
myotonic dystrophy 2	ZNF9/CNBP	CCTG	Intron 1
dentatorubral-pallidoluysian atrophy	ATN1	CAG	Coding
fragile X mental retardation syndrome	FMR1	CGG	5' UTR
fragile X E mental retardation	FMR2	GCC	5' UTR
Friedreich's ataxia	FXN	GAA	Intron
fragile X tremor ataxia syndrome	FMR1	CGG	5' UTR
Huntington's disease	HTT	CAG	Coding
Huntington's disease-like 2	JPH3	CTG	3' UTR, coding
myoclonic epilepsy of Unverricht and Lundborg	CSTB	CCCCGCCCCGCG (SEQ ID NO: 1623)	Promoter
oculopharyngeal muscular dystrophy	PABPN1	GCG	Coding
spinal and bulbar muscular atrophy	AR	CAG	Coding
spinocerebellar ataxia 1	ATXN1	CAG	Coding
spinocerebellar ataxia 2	ATXN2	CAG	Coding
spinocerebellar ataxia 3	ATXN3	CAG	Coding
spinocerebellar ataxia 6	CACNA1A	CAG	Coding

Disease	Gene	Causal repeat	Repeat location
spinocerebellar ataxia 7	ATXN7	CAG	Coding
spinocerebellar ataxia 8	ATXN8	CTG/CAG	CTG/CAG (ATXN8)
spinocerebellar ataxia 10	ATXN10	ATTCT	Intron
spinocerebellar ataxia 12	PPP2R2B	CAG	Promoter, 5' UTR?
spinocerebellar ataxia 17	TBP	CAG	Coding
Syndromic/non-syndromic X-linked mental retardation	ARX	GCG	Coding

Exemplary templates

In some embodiments, the systems or methods provided herein use the template sequences listed in **Table 27**. Table 27 provides exemplary template RNA sequences (column 5) and optional second-nick gRNA sequences (column 6) designed to be paired with a Gene Writing polypeptide to correct the indicated pathogenic mutations (column 4). All the templates in Table 27 are meant to exemplify the total sequence of: (1) targeting gRNA for first strand nick, (2) polypeptide binding domain, (3) heterologous object sequence, and (4) target homology domain for setting up TPRT at first strand nick.

Table 27. Exemplary diseases, tissues, genes, pathogenic mutations, template RNA sequences, and second nick gRNA sequences.

Disease	Tissue	Gene	Mutation	Template RNA	Second nick gRNA
Alpha-1 antitrypsin	Liver	SERPINA1	PiZ	TCCCCTCCAGGCCGTGCATAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTcGTCGATGGTCAGCACAGCCTTATGCACGGCCTGGA (SEQ ID NO: 1624)	TTTGTTGAACTTGACCTCGG (SEQ ID NO: 1625)
Cystic fibrosis	Lung	CFTR	deltaF508	ACCATTAAAGAAAATATCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACCAaagATGATATTTTCTTTA (SEQ ID NO: 1626)	AaagATGATATTTTCTTTAA (SEQ ID NO: 1627)
Sickle cell	HSC	HBB	HbS	GTAACGGCAGACTTCTCCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAGTGGGACCGAGTCGGTCCGA	TGGTGAGGCCCTGGGCAGGT

Disease	Tissue	Gene	Mutation	Template RNA	Second nick gRNA
				CTCCTGaGGAGAAGTCTGCC (SEQ ID NO: 1628)	(SEQ ID NO: 1629)
Wilson's Disease	Liver	ATP7B	H1069Q	TTGGTGACTGCCACGCCCAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAcCCCTTGGGCGTGGCAGTC (SEQ ID NO: 1630)	GGCCAGCAGTGAACAcCCCT (SEQ ID NO: 1631)
ARVC	Heart	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTTCACtGAACCAGCAGTGTTCC (SEQ ID NO: 1632)	TTGGTTGAAAATGATTTTGT (SEQ ID NO: 1633)
Long QT syndrome 1	Heart	KCNQ1	P343S	CCAGGGAAAACGCACCCACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCcCAGCGGTAGGTGCCCCGTGGGTGCGTTTTC (SEQ ID NO: 1634)	CTCCTTCTTTGCGCTCcCAG (SEQ ID NO: 1635)
Mucolipidosis IV	CNS	MCOLN1	406-2A>G	GCCCTCCCCTTCTCTGCCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGGTACtGTGGGCAGAGAAGGGG (SEQ ID NO: 1636)	TCAGGCAACGCCAGGTACtG (SEQ ID NO: 1637)

In some embodiments, the systems or methods provided herein use the template sequences listed in **Table 35**. Table 35 provides exemplary template RNA sequences (column 5) and optional second-nick gRNA sequences (column 6) designed to be paired with a Gene Writing polypeptide to correct the indicated pathogenic mutations (column 4). All the templates in Table 35 are meant to exemplify the total sequence of: (1) targeting gRNA for first strand nick, (2) polypeptide binding domain, (3) heterologous object sequence, and (4) target homology domain for setting up TPRT at first strand nick.

Table 35. Exemplary Gene Writing templates and second nick gRNA sequences for the correction of exemplary repeat expansion diseases. The region of the template spanning the repeat(s) is indicated in lowercase.

Disease	Gene	Reference Accession	Repeat	Location	Template RNA	Second-nick gRNA
myotonic dystrophy 1	DMPK	NC_000019.10 (45769709..45782490, complement)	CTG	3' UTR	CTCGAAGGGTCCTTGTAGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGTGATCCCCCcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagCATTCCCGGCTACAAGGACCCT (SEQ ID NO: 1638)	ATCACAGGACTGGAGCTGGG (SEQ ID NO: 1639)
myotonic dystrophy 2	CNBP	NC_000003.12 (129167827..129183896, complement)	CCTG	Intron 1	ACCACTGCACTCCAGCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgTCTGTCTCACTTTGTCCCCTAGGCTGGAGTGCA (SEQ ID NO: 1640)	GCCTCAGCCTCCTGAGTAGC(SEQ ID NO: 1641)

fragile X mental retardation syndrome	FMR1	NC_000023.11 (147911919..147951127)	CGG	5' UTR	GGGGGCGTGCGGCAGCGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTGCGGGCGCTCGAGGCCCAGccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgccgCGCTGCCGCACG (SEQ ID NO: 1642)	GCTCAGAGGCGGCCCTCCAC (SEQ ID NO: 1643)
Friedreich's ataxia	FXN	NC_000009.12 (69035752..69079076)	GAA	Intron	CAGGCGCGCGACACCACGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAACCCAGTATCTACTAAAAAATACAAAAAAAAAAAAAAgaagaagaagaagaagaaAATAAAGAAAAGTTAGCCGGGCGTGGTGTCGCGC(SEQ ID NO: 1644)	CGCTTGAGCCCGGGAGGCAG (SEQ ID NO: 1645)

Huntington disease	HTT	NC_000004.12 (3074681..3243960)	CAG	Coding	GGCGGCTGAGGAAGCTGAGGGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAGTCCCTCAAGTCCTTCcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaacagccgccaccgccgccgccgccgccgcctcctCAGCTTCCTCAG (SEQ ID NO: 1646)	CGCTGCACCGACCGTGAGTT (SEQ ID NO: 1647)
spinocerebellar ataxia 1	ATXN1	NC_000006.12 (16299112..16761490, complement)	CAG	Coding	TGAGCCCCGGAGCCCTGCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAAGGCTGAGcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcatcagcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagCACCTCAGCAGGGCTCCGGG (SEQ ID NO: 1648)	TCCAGTTCTCCGCAGAACAC (SEQ ID NO: 1649)

Exemplary heterologous object sequences

In some embodiments, the systems or methods provided herein comprise a heterologous object sequence, wherein the heterologous object sequence or a reverse complementary sequence thereof, encodes a protein (e.g., an antibody) or peptide. In some embodiments, the therapy is one approved by a regulatory agency such as FDA.

In some embodiments, the protein or peptide is a protein or peptide from the THPdb database (Usmani et al. PLoS One 12(7):e0181748 (2017), herein incorporated by reference in its entirety. In some embodiments, the protein or peptide is a protein or peptide disclosed in **Table 28**. In some embodiments, the systems or methods disclosed herein, for example, those comprising Gene Writers, may be used to integrate an expression cassette for a protein or peptide from Table 28 into a host cell to enable the expression of the protein or peptide in the host. In some embodiments, the sequences of the protein or peptide in the first column of Table 28 can be found in the patents or applications provided in the third column of Table 28, incorporated by reference in their entireties.

In some embodiments, the protein or peptide is an antibody disclosed in Table 1 of Lu et al. J Biomed Sci 27(1):1 (2020), herein incorporated by reference in its entirety. In some embodiments, the protein or peptide is an antibody disclosed in **Table 29**. In some embodiments, the systems or methods disclosed herein, for example, those comprising Gene Writers, may be used to integrate an expression cassette for an antibody from Table 29 into a host cell to enable the expression of the antibody in the host. In some embodiments, a system or method described herein is used to express an agent that binds a target of column 2 of Table 29 (e.g., a monoclonal antibody of column 1 of Table 29) in a subject having an indication of column 3 of Table 29.

Table 28. Exemplary protein and peptide therapeutics.

Therapeutic peptide	Category	Patent Number
Lepirudin	Antithrombins and Fibrinolytic Agents	CA1339104
Cetuximab	Antineoplastic Agents	CA1340417
Dor se alpha	Enzymes	CA2184581
Denileukin diftitox	Antineoplastic Agents	
Etanercept	Immunosuppressive Agents	CA2476934
Bivalirudin	Antithrombins	US7582727
Leuprolide	Antineoplastic Agents	
Peginterferon alpha-2a	Immunosuppressive Agents	CA2203480
Alteplase	Thrombolytic Agents	
Interferon alpha-n1	Antiviral Agents	
Darbepoetin alpha	Anti-anemic Agents	CA2165694
Reteplase	Fibrinolytic Agents	CA2107476
Epoetin alpha	Hematinics	CA1339047
Salmon Calcitonin	Bone Density Conservation Agents	US6440392

Therapeutic peptide	Category	Patent Number
Interferon alpha-n3	Immunosuppressive Agents	
Pegfilgrastim	Immunosuppressive Agents	CA1341537
Sargramostim	Immunosuppressive Agents	CA1341150
Secretin	Diagnostic Agents	
Peginterferon alpha-2b	Immunosuppressive Agents	CA1341567
Asparagi se	Antineoplastic Agents	
Thyrotropin alpha	Diagnostic Agents	US5840566
Antihemophilic Factor	Coagulants and Thrombotic agents	CA2124690
A kinra	Antirheumatic Agents	CA2141953
Gramicidin D	Anti-Bacterial Agents	
Intravenous Immunoglobulin	Immunologic Factors	
Anistreplase	Fibrinolytic Agents	
Insulin Regular	Antidiabetic Agents	
Tenecteplase	Fibrinolytic Agents	CA2129660
Menotropins	Fertility Agents	
Interferon gamma-1b	Immunosuppressive Agents	US6936695
Interferon alpha-2a, Recombi nt		CA2172664
Coagulation factor VIIa	Coagulants	
Oprelvekin	Antineoplastic Agents	
Palifermin	Anti-Mucositis Agents	
Glucagon recombi nt	Hypoglycemic Agents	
Aldesleukin	Antineoplastic Agents	
Botulinum Toxin Type B	Antidystonic Agents	
Omalizumab	Anti-Allergic Agents	CA2113813
Lutropin alpha	Fertility Agents	US5767251
Insulin Lispro	Hypoglycemic Agents	US5474978
Insulin Glargine	Hypoglycemic Agents	US7476652
Collage se		
Rasburicase	Gout Suppressants	CA2175971
Adalimumab	Antirheumatic Agents	CA2243459
Imiglucerase	Enzyme Replacement Agents	US5549892
Abciximab	Anticoagulants	CA1341357
Alpha-1-protei se inhibitor	Serine Protei se Inhibitors	
Pegaspargase	Antineoplastic Agents	
Interferon beta-1a	Antineoplastic Agents	CA1341604
Pegademase bovine	Enzyme Replacement Agents	
Human Serum Albumin	Serum substitutes	US6723303
Eptifibatide	Platelet Aggregation Inhibitors	US6706681
Serum albumin iodo ted	Diagnostic Agents	
Infliximab	Antirheumatic Agents, Anti-Inflammatory Agents, Non-Steroidal, Dermatologic Agents,	CA2106299

Therapeutic peptide	Category	Patent Number
	Gastrointesti l Agents and Immunosuppressive Agents	
Follitropin beta	Fertility Agents	US7741268
Vasopressin	Antidiuretic Agents	
Interferon beta-1b	Adjuvants, Immunologic and Immunosuppressive Agents	CA1340861
Interferon alphacon-1	Antiviral Agents and Immunosuppressive Agents	CA1341567
Hyaluronidase	Adjuvants, Anesthesia and Permeabilizing Agents	
Insulin, porcine	Hypoglycemic Agents	
Trastuzumab	Antineoplastic Agents	CA2103059
Rituximab	Antineoplastic Agents, Immunologic Factors and Antirheumatic Agents	CA2149329
Basiliximab	Immunosuppressive Agents	CA2038279
Muromo b	Immunologic Factors and Immunosuppressive Agents	
Digoxin Immune Fab (Ovine)	Antidotes	
Ibritumomab		CA2149329
Daptomycin		US6468967
Tositumomab		
Pegvisomant	Hormone Replacement Agents	US5849535
Botulinum Toxin Type A	Neuromuscular Blocking Agents, Anti-Wrinkle Agents and Antidystonic Agents	CA2280565
Pancrelipase	Gastrointesti l Agents and Enzyme Replacement Agents	
Streptoki se	Fibrinolytic Agents and Thrombolytic Agents	
Alemtuzumab		CA1339198
Alglucerase	Enzyme Replacement Agents	
Capromab	Indicators, Reagents and Diagnostic Agents	
Laronidase	Enzyme Replacement Agents	
Urofollitropin	Fertility Agents	US5767067
Efalizumab	Immunosuppressive Agents	
Serum albumin	Serum substitutes	US6723303
Choriogo dotropin alpha	Fertility Agents and Go dotropins	US6706681
Antithymocyte globulin	Immunologic Factors and Immunosuppressive Agents	

Therapeutic peptide	Category	Patent Number
Filgrastim	Immunosuppressive Agents, Antineutropenic Agents and Hematopoietic Agents	CA1341537
Coagulation factor ix	Coagulants and Thrombotic Agents	
Becaplermin	Angiogenesis Inducing Agents	CA1340846
Agalsidase beta	Enzyme Replacement Agents	CA2265464
Interferon alpha-2b	Immunosuppressive Agents	CA1341567
Oxytocin	Oxytocics, Anti-tocolytic Agents and Labor Induction Agents	
Enfuvirtide	HIV Fusion Inhibitors	US6475491
Palivizumab	Antiviral Agents	CA2197684
Daclizumab	Immunosuppressive Agents	
Bevacizumab	Angiogenesis Inhibitors	CA2286330
Arcitumomab	Diagnostic Agents	US8420081
Arcitumomab	Diagnostic Agents	US7790142
Eculizumab		CA2189015
Panitumumab		
Ranibizumab	Ophthalmics	CA2286330
Idursulfase	Enzyme Replacement Agents	
Alglucosidase alpha	Enzyme Replacement Agents	CA2416492
Exe tide	Hypoglycemic Agents	US6872700
Mecasermin		US5681814
Pramlintide		US5686411
Galsulfase	Enzyme Replacement Agents	
Abatacept	Antirheumatic Agents and Immunosuppressive Agents	CA2110518
Cosyntropin	Hormones and Diagnostic Agents	
Corticotropin		
Insulin aspart	Hypoglycemic Agents and Antidiabetic Agents	US5866538
Insulin detemir	Antidiabetic Agents	US5750497
Insulin glulisine	Antidiabetic Agents	US6960561
Pegaptanib	Intended for the prevention of respiratory distress syndrome (RDS) in premature infants at high risk for RDS.	
Nesiritide		
Thymalphasin		
Defibrotide	Antithrombins	
tural alpha interferon OR multiferon		
Glatiramer acetate		
Preotact		

Therapeutic peptide	Category	Patent Number
Teicoplanin	Anti-Bacterial Agents	
Ca kinumab	Anti-Inflammatory Agents and Monoclo l antibodies	
Ipilimumab	Antineoplastic Agents and Monoclo l antibodies	CA2381770
Sulodexide	Antithrombins and Fibrinolytic Agents and Hypoglycemic Agents and Anticoagulants and Hypolipidemic Agents	
Tocilizumab		CA2201781
Teriparatide	Bone Density Conservation Agents	US6977077
Pertuzumab	Monoclo l antibodies	CA2376596
Rilo cept	Immunosuppressive Agents	US5844099
Denosumab	Bone Density Conservation Agents and Monoclo l antibodies	CA2257247
Liraglutide		US6268343
Golimumab	Antipsoriatic Agents and Monoclo l antibodies and TNF inhibitor	
Belatacept	Antirheumatic Agents and Immunosuppressive Agents	
Buserelin		
Velaglucerase alpha	Enzymes	US7138262
Tesamorelin		US5861379
Brentuximab vedotin		
Taliglucerase alpha	Enzymes	
Belimumab	Monoclo l antibodies	
Aflibercept	Antineoplastic Agents and Ophthalmics	US7306799
Asparagi se erwinia chrysanthemi	Enzymes	
Ocriplasmin	Ophthalmics	
Glucarpidase	Enzymes	
Teduglutide		US5789379
Raxibacumab	Anti-Infective Agents and Monoclo l antibodies	
Certolizumab pegol	TNF inhibitor	CA2380298
Insulin,isophane	Hypoglycemic Agents and Antidiabetic Agents	
Epoetin zeta		
Obinutuzumab	Antineoplastic Agents	
Fibrinolysin aka plasmin		US3234106
Follitropin alpha		

Therapeutic peptide	Category	Patent Number
Romiplostim	Colony-Stimulating Factors and Thrombopoietic Agents	
lucinactant	Pulmonary surfactants	US5407914
talizumab	Immunosuppressive agents	
Aliskiren	Renin inhibitor	
Ragweed Pollen Extract		
Secukinumab	Inhibitor	US20130202610
Somatotropin Recombi nt	Hormone Replacement Agents	CA1326439
Drotrecogin alpha	Antisepsis	CA2036894
Alefacept	Dermatologic and Immunosupressive agents	
OspA lipoprotein	Vaccines	
Uroki se		US4258030
Abarelix	Anti-Testosterone Agents	US5968895
Sermorelin	Hormone Replacement Agents	
Aprotinin		US5198534
Gemtuzumab ozogamicin	Antineoplastic agents and Immunotoxins	US5585089
Satumomab Pendetide	Diagnostic Agents	
Albiglutide	Drugs used in diabetes; alimentary tract and metabolism; blood glucose lowering drugs, excl. insulins.	
Alirocumab		
Ancestim		
Antithrombin alpha		
Antithrombin III human		
Asfotase alpha	Enzymes Alimentary Tract and Metabolism	
Atezolizumab		
Autologous cultured chondrocytes		
Beractant		
Bli tumomab	Antineoplastic Agents Immunosuppressive Agents Monoclo l antibodies Antineoplastic and Immunomodulating Agents	US20120328618
C1 Esterase Inhibitor (Human)		
Coagulation Factor XIII A-Subunit (Recombi nt)		
Conestat alpha		
Daratumumab	Antineoplastic Agents	

Therapeutic peptide	Category	Patent Number
Desirudin		
Dulaglutide	Hypoglycemic Agents; Drugs Used in Diabetes; Alimentary Tract and Metabolism; Blood Glucose Lowering Drugs, Excl. Insulins	
Elosulfase alpha	Enzymes; Alimentary Tract and Metabolism	
Elotuzumab		US2014055370
Evolocumab	Lipid Modifying Agents, Plain; Cardiovascular System	
Fibrinogen Concentrate (Human)		
Filgrastim-sndz		
Gastric intrinsic factor		
Hepatitis B immune globulin		
Human calcitonin		
Human clostridium tetani toxoid immune globulin		
Human rabies virus immune globulin		
Human Rho(D) immune globulin		
Hyaluronidase (Human Recombi nt)		US7767429
Idarucizumab	Anticoagulant	
Immune Globulin Human	Immunologic Factors; Immunosuppressive Agents; Anti-Infective Agents	
Vedolizumab	Immunosupressive agent, Antineoplastic agent	US2012151248
Ustekinumab	Deramtologic agent, Immunosuppressive agent, antineoplastic agent	
Turoctocog alpha		
Tuberculin Purified Protein Derivative		
Simoctocog alpha	Antihaemorrhagics: blood coagulation factor VIII	
Siltuximab	Antineoplastic and Immunomodulating Agents, Immunosuppressive Agents	US7612182
Sebelipase alpha	Enzymes	
Sacrosidase	Enzymes	

Therapeutic peptide	Category	Patent Number
Ramucirumab	Antineoplastic and Immunomodulating Agents	US2013067098
Prothrombin complex concentrate		
Poractant alpha	Pulmo ry Surfactants	
Pembrolizumab	Antineoplastic and Immunomodulating Agents	US2012135408
Peginterferon beta-1a		
Ofatumumab	Antineoplastic and Immunomodulating Agents	US8337847
Obiltoxaximab		
Nivolumab	Antineoplastic and Immunomodulating Agents	US2013173223
Necitumumab		
Metreleptin		US20070099836
Methoxy polyethylene glycol-epoetin beta		
Mepolizumab	Antineoplastic and Immunomodulating Agents, Immunosuppressive Agents, Interleukin Inhibitors	US2008134721
Ixekizumab		
Insulin Pork	Hypoglycemic Agents, Antidiabetic Agents	
Insulin Degludec		
Insulin Beef		
Thyroglobulin	Hormone therapy	US5099001
Anthrax immune globulin human	Plasma derivative	
Anti-inhibitor coagulant complex	Blood Coagulation Factors, Antihemophilic Agent	
Anti-thymocyte Globulin (Equine)	Antibody	
Anti-thymocyte Globulin (Rabbit)	Antibody	
Brodalumab	Antineoplastic and Immunomodulating Agents	
C1 Esterase Inhibitor (Recombi nt)	Blood and Blood Forming Organs	
Ca kinumab	Antineoplastic and Immunomodulating Agents	
Chorionic Go dotropin (Human)	Hormones	US6706681

Therapeutic peptide	Category	Patent Number
Chorionic Go dotropin (Recombi nt)	Hormones	US5767251
Coagulation factor X human	Blood Coagulation Factors	
Dinutuximab	Antibody, Immunosuppresive agent, Antineoplastic agent	US20140170155
Efmoroctocog alpha	Antihemophilic Factor	
Factor IX Complex (Human)	Antihemophilic agent	
Hepatitis A Vaccine	Vaccine	
Human Varicella-Zoster Immune Globulin	Antibody	
Ibritumomab tiuxetan	Antibody, Immunosuppressive Agents	CA2149329
Lenograstim	Antineoplastic and Immunomodulating Agents	
Pegloticase	Enzymes	
Protamine sulfate	Heparin Antagonists, Hematologic Agents	
Protein S human	Anticoagulant plasma protein	
Sipuleucel-T	Antineoplastic and Immunomodulating Agents	US8153120
Somatropin recombi nt	Hormones, Hormone Substitutes, and Hormone Antagonists	CA1326439,CA2252535 ,US5288703,US5849700,U S5849704,US5898030 ,US6004297,US6152897 ,US6235004 ,US6899699
Susoctocog alpha	Blood coagulation factors, Antihaemorrhagics	
Thrombomodulin alpha	Anticoagulant agent, Antiplatelet agent	

Table 29. Exemplary monoclonal antibody therapies.

mAb	Target	Indication
Muromonab-CD3	CD3	Kidney transplant rejection
Abciximab	GPIIb/IIIa	Prevention of blood clots in angioplasty
Rituximab	CD20	Non-Hodgkin lymphoma
Palivizumab	RSV	Prevention of respiratory syncytial virus infection
Infliximab	TNFα	Crohn's disease
Trastuzumab	HER2	Breast cancer
Alemtuzumab	CD52	Chronic myeloid leukemia
Adalimumab	TNFα	Rheumatoid arthritis

Ibritumomab tiuxetan	CD20	Non-Hodgkin lymphoma
Omalizumab	IgE	Asthma
Cetuximab	EGFR	Colorectal cancer
Bevacizumab	VEGF-A	Colorectal cancer
Natalizumab	ITGA4	Multiple sclerosis
Panitumumab	EGFR	Colorectal cancer
Ranibizumab	VEGF-A	Macular degeneration
Eculizumab	C5	Paroxysmal nocturnal hemoglobinuria
Certolizumab pegol	TNFα	Crohn's disease
Ustekinumab	IL-12/23	Psoriasis
Canakinumab	IL-1β	Muckle-Wells syndrome
Golimumab	TNFα	Rheumatoid and psoriatic arthritis, ankylosing spondylitis
Ofatumumab	CD20	Chronic lymphocytic leukemia
Tocilizumab	IL-6R	Rheumatoid arthritis
Denosumab	RANKL	Bone loss
Belimumab	BLyS	Systemic lupus erythematosus
Ipilimumab	CTLA-4	Metastatic melanoma
Brentuximab vedotin	CD30	Hodgkin lymphoma, systemic anaplastic large cell lymphoma
Pertuzumab	HER2	Breast Cancer
Trastuzumab emtansine	HER2	Breast cancer
Raxibacumab	*B. anthrasis* PA	Anthrax infection
Obinutuzumab	CD20	Chronic lymphocytic leukemia
Siltuximab	IL-6	Castleman disease
Ramucirumab	VEGFR2	Gastric cancer
Vedolizumab	α4β7 integrin	Ulcerative colitis, Crohn disease
Blinatumomab	CD19, CD3	Acute lymphoblastic leukemia
Nivolumab	PD-1	Melanoma, non-small cell lung cancer
Pembrolizumab	PD-1	Melanoma
Idarucizumab	Dabigatran	Reversal of dabigatran-induced anticoagulation
Necitumumab	EGFR	Non-small cell lung cancer
Dinutuximab	GD2	Neuroblastoma
Secukinumab	IL-17α	Psoriasis
Mepolizumab	IL-5	Severe eosinophilic asthma
Alirocumab	PCSK9	High cholesterol
Evolocumab	PCSK9	High cholesterol
Daratumumab	CD38	Multiple myeloma
Elotuzumab	SLAMF7	Multiple myeloma
Ixekizumab	IL-17α	Psoriasis
Reslizumab	IL-5	Asthma
Olaratumab	PDGFRα	Soft tissue sarcoma

Bezlotoxumab	*Clostridium difficile* enterotoxin B	Prevention of *Clostridium difficile* infection recurrence
Atezolizumab	PD-L1	Bladder cancer
Obiltoxaximab	*B. anthrasis* PA	Prevention of inhalational anthrax
Inotuzumab ozogamicin	CD22	Acute lymphoblastic leukemia
Brodalumab	IL-17R	Plaque psoriasis
Guselkumab	IL-23 p19	Plaque psoriasis
Dupilumab	IL-4Rα	Atopic dermatitis
Sarilumab	IL-6R	Rheumatoid arthritis
Avelumab	PD-L1	Merkel cell carcinoma
Ocrelizumab	CD20	Multiple sclerosis
Emicizumab	Factor IXa, X	Hemophilia A
Benralizumab	IL-5Rα	Asthma
Gemtuzumab ozogamicin	CD33	Acute myeloid leukemia
Durvalumab	PD-L1	Bladder cancer
Burosumab	FGF23	X-linked hypophosphatemia
Lanadelumab	Plasma kallikrein	Hereditary angioedema attacks
Mogamulizumab	CCR4	Mycosis fungoides or Sézary syndrome
Erenumab	CGRPR	Migraine prevention
Galcanezumab	CGRP	Migraine prevention
Tildrakizumab	IL-23 p19	Plaque psoriasis
Cemiplimab	PD-1	Cutaneous squamous cell carcinoma
Emapalumab	IFNγ	Primary hemophagocytic lymphohistiocytosis
Fremanezumab	CGRP	Migraine prevention
Ibalizumab	CD4	HIV infection
Moxetumomab pasudodox	CD22	Hairy cell leukemia
Ravulizumab	C5	Paroxysmal nocturnal hemoglobinuria
Caplacizumab	von Willebrand factor	Acquired thrombotic thrombocytopenic purpura
Romosozumab	Sclerostin	Osteoporosis in postmenopausal women at increased risk of fracture
Risankizumab	IL-23 p19	Plaque psoriasis
Polatuzumab vedotin	CD79β	Diffuse large B-cell lymphoma
Brolucizumab	VEGF-A	Macular degeneration
Crizanlizumab	P-selectin	Sickle cell disease

Plant-modification Methods

Gene Writer systems described herein may be used to modify a plant or a plant part (e.g., leaves, roots, flowers, fruits, or seeds), e.g., to increase the fitness of a plant.

A. Delivery to a Plant

Provided herein are methods of delivering a Gene Writer system described herein to a plant. Included are methods for delivering a Gene Writer system to a plant by contacting the plant, or part thereof, with a Gene Writer system. The methods are useful for modifying the plant to, e.g., increase the fitness of a plant.

More specifically, in some embodiments, a nucleic acid described herein (e.g., a nucleic acid encoding a GeneWriter) may be encoded in a vector, e.g., inserted adjacent to a plant promoter, e.g., a maize ubiquitin promoter (ZmUBI) in a plant vector (e.g., pHUC411). In some embodiments, the nucleic acids described herein are introduced into a plant (e.g., *japonica* rice) or part of a plant (e.g., a callus of a plant) via agrobacteria. In some embodiments, the systems and methods described herein can be used in plants by replacing a plant gene (e.g., hygromycin phosphotransferase (HPT)) with a null allele (e.g., containing a base substitution at the start codon). Systems and methods for modifying a plant genome are described in Xu et. al. *Development of plant prime-editing systems for precise genome editing*, 2020, Plant Communications.

In one aspect, provided herein is a method of increasing the fitness of a plant, the method including delivering to the plant the Gene Writer system described herein (e.g., in an effective amount and duration) to increase the fitness of the plant relative to an untreated plant (e.g., a plant that has not been delivered the Gene Writer system).

An increase in the fitness of the plant as a consequence of delivery of a Gene Writer system can manifest in a number of ways, e.g., thereby resulting in a better production of the plant, for example, an improved yield, improved vigor of the plant or quality of the harvested product from the plant, an improvement in pre- or post-harvest traits deemed desirable for

agriculture or horticulture (e.g., taste, appearance, shelf life), or for an improvement of traits that otherwise benefit humans (e.g., decreased allergen production). An improved yield of a plant relates to an increase in the yield of a product (e.g., as measured by plant biomass, grain, seed or fruit yield, protein content, carbohydrate or oil content or leaf area) of the plant by a measurable amount over the yield of the same product of the plant produced under the same conditions, but without the application of the instant compositions or compared with application of conventional plant-modifying agents. For example, yield can be increased by at least about 0.5%, about 1%, about 2%, about 3%, about 4%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more than 100%. In some instances, the method is effective to increase yield by about 2x-fold, 5x-fold, 10x-fold, 25x-fold, 50x-fold, 75x-fold, 100x-fold, or more than 100x-fold relative to an untreated plant. Yield can be expressed in terms of an amount by weight or volume of the plant or a product of the plant on some basis. The basis can be expressed in terms of time, growing area, weight of plants produced, or amount of a raw material used. For example, such methods may increase the yield of plant tissues including, but not limited to: seeds, fruits, kernels, bolls, tubers, roots, and leaves.

An increase in the fitness of a plant as a consequence of delivery of a Gene Writer system can also be measured by other means, such as an increase or improvement of the vigor rating, the stand (the number of plants per unit of area), plant height, stalk circumference, stalk length, leaf number, leaf size, plant canopy, visual appearance (such as greener leaf color), root rating, emergence, protein content, increased tillering, bigger leaves, more leaves, less dead basal leaves, stronger tillers, less fertilizer needed, less seeds needed, more productive tillers, earlier flowering, early grain or seed maturity, less plant verse (lodging), increased shoot growth, earlier germination, or any combination of these factors, by a measurable or noticeable amount over the same factor of the plant produced under the same conditions, but without the administration of the instant compositions or with application of conventional plant-modifying agents (e.g., plant-modifying agents delivered without PMPs).

Accordingly, provided herein is a method of modifying a plant, the method including delivering to the plant an effective amount of any of the Gene Writer systems provided herein, wherein the method modifies the plant and thereby introduces or increases a beneficial trait in the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%,

or more than 100%) relative to an untreated plant. In particular, the method may increase the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In some instances, the increase in plant fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in disease resistance, drought tolerance, heat tolerance, cold tolerance, salt tolerance, metal tolerance, herbicide tolerance, chemical tolerance, water use efficiency, nitrogen utilization, resistance to nitrogen stress, nitrogen fixation, pest resistance, herbivore resistance, pathogen resistance, yield, yield under water-limited conditions, vigor, growth, photosynthetic capability, nutrition, protein content, carbohydrate content, oil content, biomass, shoot length, root length, root architecture, seed weight, or amount of harvestable produce.

In some instances, the increase in fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in development, growth, yield, resistance to abiotic stressors, or resistance to biotic stressors. An abiotic stress refers to an environmental stress condition that a plant or a plant part is subjected to that includes, e.g., drought stress, salt stress, heat stress, cold stress, and low nutrient stress. A biotic stress refers to an environmental stress condition that a plant or plant part is subjected to that includes, e.g. nematode stress, insect herbivory stress, fungal pathogen stress, bacterial pathogen stress, or viral pathogen stress. The stress may be temporary, e.g. several hours, several days, several months, or permanent, e.g. for the life of the plant.

In some instances, the increase in plant fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in quality of products harvested from the plant. For example, the increase in plant fitness may be an improvement in commercially favorable features (e.g., taste or appearance) of a product harvested from the plant. In other instances, the increase in plant fitness is an increase in shelf-life of a product harvested from the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%).

Alternatively, the increase in fitness may be an alteration of a trait that is beneficial to human or animal health, such as a reduction in allergen production. For example, the increase in

fitness may be a decrease (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in production of an allergen (e.g., pollen) that stimulates an immune response in an animal (e.g., human).

The modification of the plant (e.g., increase in fitness) may arise from modification of one or more plant parts. For example, the plant can be modified by contacting leaf, seed, pollen, root, fruit, shoot, flower, cells, protoplasts, or tissue (e.g., meristematic tissue) of the plant. As such, in another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting pollen of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In yet another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting a seed of the plant with an effective amount of any of the Gene Writer systems disclosed herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method including contacting a protoplast of the plant with an effective amount of any of the Gene Writer systems described herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In a further aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting a plant cell of the plant with an effective amount of any of the Gene Writer system described herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting meristematic tissue of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant

(e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting an embryo of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

B. Application Methods

A plant described herein can be exposed to any of the Gene Writer system compositions described herein in any suitable manner that permits delivering or administering the composition to the plant. The Gene Writer system may be delivered either alone or in combination with other active (e.g., fertilizing agents) or inactive substances and may be applied by, for example, spraying, injection (e.g,. microinjection), through plants, pouring, dipping, in the form of concentrated liquids, gels, solutions, suspensions, sprays, powders, pellets, briquettes, bricks and the like, formulated to deliver an effective concentration of the plant-modifying composition. Amounts and locations for application of the compositions described herein are generally determined by the habitat of the plant, the lifecycle stage at which the plant can be targeted by the plant-modifying composition, the site where the application is to be made, and the physical and functional characteristics of the plant-modifying composition.

In some instances, the composition is sprayed directly onto a plant, e.g., crops, by e.g., backpack spraying, aerial spraying, crop spraying/dusting etc. In instances where the Gene Writer system is delivered to a plant, the plant receiving the Gene Writer system may be at any stage of plant growth. For example, formulated plant-modifying compositions can be applied as a seed-coating or root treatment in early stages of plant growth or as a total plant treatment at later stages of the crop cycle. In some instances, the plant-modifying composition may be applied as a topical agent to a plant.

Further, the Gene Writer system may be applied (e.g., in the soil in which a plant grows, or in the water that is used to water the plant) as a systemic agent that is absorbed and distributed

through the tissues of a plant. In some instances, plants or food organisms may be genetically transformed to express the Gene Writer system.

Delayed or continuous release can also be accomplished by coating the Gene Writer system or a composition with the plant-modifying composition(s) with a dissolvable or bioerodable coating layer, such as gelatin, which coating dissolves or erodes in the environment of use, to then make the plant-modifying com Gene Writer system position available, or by dispersing the agent in a dissolvable or erodable matrix. Such continuous release and/or dispensing means devices may be advantageously employed to consistently maintain an effective concentration of one or more of the plant-modifying compositions described herein.

In some instances, the Gene Writer system is delivered to a part of the plant, e.g., a leaf, seed, pollen, root, fruit, shoot, or flower, or a tissue, cell, or protoplast thereof. In some instances, the Gene Writer system is delivered to a cell of the plant. In some instances, the Gene Writer system is delivered to a protoplast of the plant. In some instances, the Gene Writer system is delivered to a tissue of the plant. For example, the composition may be delivered to meristematic tissue of the plant (e.g., apical meristem, lateral meristem, or intercalary meristem). In some instances, the composition is delivered to permanent tissue of the plant (e.g., simple tissues (e.g., parenchyma, collenchyma, or sclerenchyma) or complex permanent tissue (e.g., xylem or phloem)). In some instances, the Gene Writer system is delivered to a plant embryo.

C. Plants

A variety of plants can be delivered to or treated with a Gene Writer system described herein. Plants that can be delivered a Gene Writer system (i.e., "treated") in accordance with the present methods include whole plants and parts thereof, including, but not limited to, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, cotyledons, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. Plant parts can further refer parts of the plant such as the shoot,

root, stem, seeds, stipules, leaves, petals, flowers, ovules, bracts, branches, petioles, internodes, bark, pubescence, tillers, rhizomes, fronds, blades, pollen, stamen, and the like.

The class of plants that can be treated in a method disclosed herein includes the class of higher and lower plants, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and algae (e.g., multicellular or unicellular algae). Plants that can be treated in accordance with the present methods further include any vascular plant, for example monocotyledons or dicotyledons or gymnosperms, including, but not limited to alfalfa, apple, Arabidopsis, banana, barley, canola, castor bean, chrysanthemum, clover, cocoa, coffee, cotton, cottonseed, corn, crambe, cranberry, cucumber, dendrobium, dioscorea, eucalyptus, fescue, flax, gladiolus, liliacea, linseed, millet, muskmelon, mustard, oat, oil palm, oilseed rape, papaya, peanut, pineapple, ornamental plants, Phaseolus, potato, rapeseed, rice, rye, ryegrass, safflower, sesame, sorghum, soybean, sugarbeet, sugarcane, sunflower, strawberry, tobacco, tomato, turfgrass, wheat and vegetable crops such as lettuce, celery, broccoli, cauliflower, cucurbits; fruit and nut trees, such as apple, pear, peach, orange, grapefruit, lemon, lime, almond, pecan, walnut, hazel; vines, such as grapes (e.g., a vineyard), kiwi, hops; fruit shrubs and brambles, such as raspberry, blackberry, gooseberry; forest trees, such as ash, pine, fir, maple, oak, chestnut, popular; with alfalfa, canola, castor bean, corn, cotton, crambe, flax, linseed, mustard, oil palm, oilseed rape, peanut, potato, rice, safflower, sesame, soybean, sugarbeet, sunflower, tobacco, tomato, and wheat. Plants that can be treated in accordance with the methods of the present invention include any crop plant, for example, forage crop, oilseed crop, grain crop, fruit crop, vegetable crop, fiber crop, spice crop, nut crop, turf crop, sugar crop, beverage crop, and forest crop. In certain instances, the crop plant that is treated in the method is a soybean plant. In other certain instances, the crop plant is wheat. In certain instances, the crop plant is corn. In certain instances, the crop plant is cotton. In certain instances, the crop plant is alfalfa. In certain instances, the crop plant is sugarbeet. In certain instances, the crop plant is rice. In certain instances, the crop plant is potato. In certain instances, the crop plant is tomato.

In certain instances, the plant is a crop. Examples of such crop plants include, but are not limited to, monocotyledonous and dicotyledonous plants including, but not limited to, fodder or forage legumes, ornamental plants, food crops, trees, or shrubs selected from Acer spp., Allium

spp., Amaranthus spp., Ananas comosus, Apium graveolens, Arachis spp, Asparagus officinalis, Beta vulgaris, Brassica spp. (e.g., Brassica napus, Brassica rapa ssp. (canola, oilseed rape, turnip rape), 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. (e.g., Glycine max, Soja hispida or Soja max), Gossypium hirsutum, Helianthus spp. (e.g., Helianthus annuus), Hibiscus spp., Hordeum spp. (e.g., Hordeum vulgare), Ipomoea batatas, Juglans spp., Lactuca sativa, Linum usitatissimum, Litchi chinensis, Lotus spp., Luffa acutangula, Lupinus spp., Lycopersicon spp. (e.g., Lycopersicon esculenturn, Lycopersicon lycopersicum, Lycopersicon pyriforme), Malus spp., Medicago sativa, Mentha spp., Miscanthus sinensis, Morus nigra, Musa spp., Nicotiana spp., Olea spp., Oryza spp. (e.g., 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. (e.g., Solanum tuberosum, Solanum integrifolium or Solanum lycopersicum), Sorghum bicolor, Sorghum halepense, Spinacia spp., Tamarindus indica, Theobroma cacao, Trifolium spp., Triticosecale rimpaui, Triticum spp. (e.g., Triticum aestivum, Triticum durum, Triticum turgidum, Triticum hybernum, Triticum macha, Triticum sativum or Triticum vulgare), Vaccinium spp., Vicia spp., Vigna spp., Viola odorata, Vitis spp., and Zea mays. In certain embodiments, the crop plant is rice, oilseed rape, canola, soybean, corn (maize), cotton, sugarcane, alfalfa, sorghum, or wheat.

The plant or plant part for use in the present invention include plants of any stage of plant development. In certain instances, the delivery can occur during the stages of germination, seedling growth, vegetative growth, and reproductive growth. In certain instances, delivery to the plant occurs during vegetative and reproductive growth stages. In some instances, the composition is delivered to pollen of the plant. In some instances, the composition is delivered to a seed of the plant. In some instances, the composition is delivered to a protoplast of the plant. In some instances, the composition is delivered to a tissue of the plant. For example, the composition may be delivered to meristematic tissue of the plant (e.g., apical meristem, lateral meristem, or intercalary meristem). In some instances, the composition is delivered to

permanent tissue of the plant (e.g., simple tissues (e.g., parenchyma, collenchyma, or sclerenchyma) or complex permanent tissue (e.g., xylem or phloem)). In some instances, the composition is delivered to a plant embryo. In some instances, the composition is delivered to a plant cell. The stages of vegetative and reproductive growth are also referred to herein as "adult" or "mature" plants.

In instances where the Gene Writer system is delivered to a plant part, the plant part may be modified by the plant-modifying agent. Alternatively, the Gene Writer system may be distributed to other parts of the plant (e.g., by the plant's circulatory system) that are subsequently modified by the plant-modifying agent.

Administration and Delivery Modalities

Nucleic acid elements of systems provided by the invention, used in the methods provided by the invention, can be delivered by a variety of modalities. In embodiments where the system comprises two separate nucleic acid molecules (e.g., the retrotransposase and template nucleic acids are separate molecules), the two molecules may be delivered by the same modality, while in other embodiments, the two molecules are delivered by different modalities. The composition and systems described herein may be used in vitro or in vivo. In some embodiments the system or components of the system are delivered to cells (e.g., mammalian cells, e.g., human cells), e.g., in vitro, ex vivo, or in vivo. In some embodiments, the cells are eukaryotic cells, e.g., cells of a multicellular organism, e.g., an animal, e.g., a mammal (e.g., human, swine, bovine) a bird (e.g., poultry, such as chicken, turkey, or duck), or a fish. In some embodiments, the cells are non-human animal cells (e.g., a laboratory animal, a livestock animal, or a companion animal). In some embodiments, the cell is a stem cell (e.g., a hematopoietic stem cell), a fibroblast, or a T cell. In some embodiments, the cell is a non-dividing cell, e.g., a non-dividing fibroblast or non-dividing T cell. The skilled artisan will understand that the components of the Gene Writer system may be delivered in the form of polypeptide, nucleic acid (e.g., DNA, RNA), and combinations thereof.

For instance, delivery can use any of the following combinations for delivering the retrotransposase (e.g., as DNA encoding the retrotransposase protein, as RNA encoding the

retrotransposase protein, or as the protein itself) and the template RNA (e.g., as DNA encoding the RNA, or as RNA):

1. Retrotransposase DNA + template DNA
2. Retrotransposase RNA + template DNA
3. Retrotransposase DNA + template RNA
4. Retrotransposase RNA + template RNA
5. Retrotransposase protein + template DNA
6. Retrotransposase protein + template RNA
7. Retrotransposase virus + template virus
8. Retrotransposase virus + template DNA
9. Retrotransposase virus + template RNA
10. Retrotransposase DNA + template virus
11. Retrotransposase RNA + template virus
12. Retrotransposase protein + template virus

As indicated above, in some embodiments, the DNA or RNA that encodes the retrotransposase protein is delivered using a virus, and in some embodiments, the template RNA (or the DNA encoding the template RNA) is delivered using a virus.

In one embodiments the system and/or components of the system are delivered as nucleic acid. For example, the Gene Writer polypeptide may be delivered in the form of a DNA or RNA encoding the polypeptide, and the template RNA may be delivered in the form of RNA or its complementary DNA to be transcribed into RNA. In some embodiments the system or components of the system are delivered on 1, 2, 3, 4, or more distinct nucleic acid molecules. In some embodiments the system or components of the system are delivered as a combination of DNA and RNA. In some embodiments the system or components of the system are delivered as a combination of DNA and protein. In some embodiments the system or components of the system are delivered as a combination of RNA and protein. In some embodiments the Gene Writer genome editor polypeptide is delivered as a protein.

In some embodiments the system or components of the system are delivered to cells, e.g. mammalian cells or human cells, using a vector. The vector may be, e.g., a plasmid or a virus. In some embodiments delivery is in vivo, in vitro, ex vivo, or in situ. In some embodiments the virus is an adeno associated virus (AAV), a lentivirus, an adenovirus. In some embodiments the

system or components of the system are delivered to cells with a viral-like particle or a virosome. In some embodiments the delivery uses more than one virus, viral-like particle or virosome.

In one embodiment, the compositions and systems described herein can be formulated in liposomes or other similar vesicles. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes may be anionic, neutral or cationic. Liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

Vesicles can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Methods for preparation of multilamellar vesicle lipids are known in the art (see for example U.S. Pat. No. 6,693,086, the teachings of which relating to multilamellar vesicle lipid preparation are incorporated herein by reference). Although vesicle formation can be spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Extruded lipids can be prepared by extruding through filters of decreasing size, as described in Templeton et al., Nature Biotech, 15:647-652, 1997, the teachings of which relating to extruded lipid preparation are incorporated herein by reference.

Lipid nanoparticles are another example of a carrier that provides a biocompatible and biodegradable delivery system for the pharmaceutical compositions described herein. Nanostructured lipid carriers (NLCs) are modified solid lipid nanoparticles (SLNs) that retain the characteristics of the SLN, improve drug stability and loading capacity, and prevent drug leakage. Polymer nanoparticles (PNPs) are an important component of drug delivery. These nanoparticles can effectively direct drug delivery to specific targets and improve drug stability and controlled drug release. Lipid–polymer nanoparticles (PLNs), a new type of carrier that combines liposomes and polymers, may also be employed. These nanoparticles possess the complementary advantages of PNPs and liposomes. A PLN is composed of a core–shell

structure; the polymer core provides a stable structure, and the phospholipid shell offers good biocompatibility. As such, the two components increase the drug encapsulation efficiency rate, facilitate surface modification, and prevent leakage of water-soluble drugs. For a review, see, e.g., Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Exosomes can also be used as drug delivery vehicles for the compositions and systems described herein. For a review, see 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.

A Gene Writer system can be introduced into cells, tissues and multicellular organisms. In some embodiments the system or components of the system are delivered to the cells via mechanical means or physical means.

Formulation of protein therapeutics is described in Meyer (Ed.), *Therapeutic Protein Drug Products: Practical Approaches to formulation in the Laboratory, Manufacturing, and the Clinic*, Woodhead Publishing Series (2012).

All publications, patent applications, patents, and other publications and references (e.g., sequence database reference numbers) cited herein are incorporated by reference in their entirety. For example, all GenBank, Unigene, and Entrez sequences referred to herein, e.g., in any Table herein, are incorporated by reference. Unless otherwise specified, the sequence accession numbers specified herein, including in any Table herein, refer to the database entries current as of March 4, 2020. When one gene or protein references a plurality of sequence accession numbers, all of the sequence variants are encompassed.

EXAMPLES

The invention is further illustrated by the following examples. The examples are provided for illustrative purposes only and are not to be construed as limiting the scope or content of the invention in any way.

Example 1: Internal Gene Writer deletions demonstrating protein domain modularity

This example describes deletions in a Gene Writer polypeptide that retain functionality and further demonstrate the modularity of the DNA binding domain.

In this example, a series of experiments were performed to test the activity of various mutant retrotransposases, as well as gaining structural knowledge about the protein modularity. This experiment tested removing a polypeptide stretch after the c-myb motif in the DNA binding domain (DBD) and replacing it with a flexible linker (Figure 1a). The polypeptide stretch removed is referred to as the "natural linker" since it is the intervening region between the DNA binding motifs and the RNA binding domain. The polypeptide region removed spans the following: on the N terminal side at either, location A (predicted random coil following c-myb motif) or location B (end of predicted alpha helix that contains part of the c-myb motif) and the removed region ends at either location v1 (alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif or at location v2 (C-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif). In place of the polypeptide stretch removed, "natural linker", is the either of two linkers (Linker A, XTEN: SGSETPGTSESATPES (SEQ ID NO: 1023), and Linker B, 3GS: GGGS (SEQ ID NO: 1024)). For each of these mutant retrotransposases that contain different removed regions (location A – v1, location A – v2, location B – v1, or location B – v2) they were replaced with either linker A or linker B by PCR to a DNA plasmid that expressed R2Tg, thereby yielding these sequences: c-mybA – v1 replaced with 3GS linker, c-mybA – v2 replaced with 3GS linker, c-mybA – v1 replaced with XTEN linker, c-mybA – v2 replaced with XTEN linker, c-mybB – v1 replaced with 3GS linker, c-mybB – v2 replaced with 3GS linker, c-mybB – v1 replaced with XTEN linker, c-mybB – v2 replaced with XTEN linker, as shown in Table E1 below. The insertion of the linkers was verified by Sanger sequencing and the DNA plasmids were purified for transfection.

Table E1. Amino acid sequences of R2Tg mutants with linkers in place of the "natural linker" region that intervenes the DNA binding domain (DBD) and RNA binding domain.

The N-terminal DNA-binding domain is italicized and the linker connecting to the rest of the protein is in bold and underlined.

R2Tg Mutant	Amino Acid Sequence	SEQ ID NO

R2Tg with natural linker deletion c-mybA location – v1 replaced with 3GS linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK* *PAEEPREEPGTCHHTRRAA**GGGS**CFGCLESISQIRTATRDKKDTVTREK* HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDI PLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGC RTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNP RQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQ GDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLK PLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPP CTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFME KEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPH RKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNC PVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIF VKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDV TFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDI VHMFASRARKSMVM	
R2Tg with natural linker deletion c-mybA location – v2 replaced with 3GS linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK* *PAEEPREEPGTCHHTRRAA**GGGS**TATRDKKDTVTREKHPKKPFQKWMKD* RAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTR WETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPD GITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCS ENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVD PHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLA MDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILE TFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGE SEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYT IPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTR DGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRK NEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRAN VYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHN YICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT	

	VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGK WHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSM VM	
R2Tg with natural linker deletion c-mybA location – v1 replaced with XTEN linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK* *PAEEPREEPGTCHHTRRAA***SGSETPGTSESATPES**CFGCLESISQIRTA TRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKI ILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWL TTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIV TARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFV DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTH TDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAM AFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKD SYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKI RTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNN VSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKIS NHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDA DIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRD SNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLE TEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAET FSTVALFSSVDIVHMFASRARKSMVM	
R2Tg with natural linker deletion c-mybA location – v2 replaced with XTEN linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK* *PAEEPREEPGTCHHTRRAA***SGSETPGTSESATPES**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDI PLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGC RTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNP RQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQ GDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLK PLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPP CTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFME KEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ	

	KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPH RKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNC PVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIF VKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDV TFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDI VHMFASRARKSMVM	
R2Tg with natural linker deletion c-mybB location – v1 replaced with 3GS linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***GGGS** CFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLG DFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPI TIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIW SAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQ GQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHS EVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAG LIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLAS QGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAK KKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLE EAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTE LGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg with natural linker deletion c-mybB location – v2 replaced with 3GS linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***GGGS** TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLA KIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTA FRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNL WLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSR IVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVV FVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRN THTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSIT AMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPT KDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQ KIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHR	

	IAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDK ISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHC DADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHI RDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKH LETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMA ETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg with natural linker deletion c-mybB location – v1 replaced with XTEN linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***SGSE** **TPGTSESATPES**CFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMK DRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKT RWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKP DRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGC SENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISIL ETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPG ESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTY TIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSST RDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKL WIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWR KNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRA NVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRH NYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDV TVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKS MVM	
R2Tg with natural linker deletion c-mybB location – v2 replaced with XTEN linker	*MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL* *GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV* *DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH* *FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE* *SEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNL* *IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGE* *WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR* *CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL***SGSE** **TPGTSESATPES**TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLG DFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPI TIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIW SAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQ GQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP	

	WIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHS EVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAG LIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLAS QGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAK KKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLE EAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTE LGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	

HEK293T cells were plated in 96-well plates and grown overnight at 37˚C, 5% CO2. The HEK293T cells were transfected with plasmids that expressed R2Tg (wild-type), R2 endonuclease mutant, and natural linker mutants. The transfection was carried out using the Fugene HD transfection reagent according to the manufacturer recommendations, where each well received 80 ng of plasmid DNA and 0.5 µL of transfection reagent. All transfections were performed in duplicate and the cells were incubated for 72 h prior to genomic DNA extraction.

Activity of the mutants was measured by a ddPCR assay that quantified the copy number of R2Tg integrations by measuring the number of 3′ junction amplicons (Figure 1b).

Deletions that begin after the random coil following the c-myb DNA binding motif (location A, c-mybA) are well-tolerated with integration activity near that of wild-type R2Tg. The natural linker region deletion end point is nearly the same for either location v1 (N-terminal to the alpha helix preceding the -1 RNA binding motif) or v2 (C-terminal to the alpha helix preceding the -1 RNA binding motif). For the deletion beginning at location A and ending at location v1 or v2, replacement of this polypeptide stretch with the XTEN linker (SEQ ID NO: 1023) seems to retain the most amount of activity whereas replacement with the 3GS linker (SEQ ID NO: 1024) has approximately a 50% reduction in integration activity. For natural linker deletions that begin at location B (c-mybB), these configurations show a more marked reduction in integration activity when compared to wild-type or location A (c-mybA). The difference in activity may be related to the structure of the protein based on the position of the deletion that creates a non-optimal three dimensional structure of the retrotransposase through the location of the linker, length of the linker, or amino acid combination of the linker that is not optimal to connect location B to locations v1 or v2. Even though the N-terminal natural linker deletion start location mybB is a sub-optimal, the C-terminal end of the deletion was most tolerated at v2 with

either the 3GS or XTEN linker and appears to be the preferential location for having a polypeptide preceding the RBD -1 region.

Example 2: Determination of target specificity of a Gene Writer endonuclease domain

This example describes using a custom genomic landing pad in human cells to determine whether there is a sequence requirement for target cleavage and subsequent integration by a Gene Writer system.

In this example, cell lines were created to have "landing pads" or stable integrations that mimic a region of rDNA that contain the R2 position to which R2 retrotransposases target for retrotransposition (see Fig 2). The integrants or landing pads were designed to either have the wild-type region sequence in and around the R2 site found in rDNA, 12-bp of sequence mutation at and around the R2 cleavage site, or 75-bp of sequence mutation at and around the R2 cleavage site (Table E2). The DNA for these different landing pads was chemically synthesized and cloned into the pLenti-N-tGFP vector. The cloned landing pads into the lentiviral expression vector were confirmed and sequence verified by Sanger sequencing of the landing pad. The sequence verified plasmids (9µg) along with the lentiviral packaging mix (9µg, obtained from Biosettia) were transfected using Lipofectamine2000™ according the manufacturer instructions into a packaging cell line, LentiX-293T (Takara Bio). The transfected cells were incubated at 37˚C, 10% CO_2 for 48 hours (including one medium change at 24hrs) and the viral particle containing medium was collected from the cell culture dish. The collected medium was filtered through a 0.2 µm filter to remove cell debris and prepared for transduction of U2OS cells. The viral containing medium was diluted in DMEM and mixed with polybrene to prepare a dilution series for transduction of U2OS cells where the final concentration of polybrene was 8 µg/ml. The U2OS cells were grown in viral containing medium for 48 hour and then split with fresh medium. The split cells were grown to confluence and transduction efficiency of the different dilutions of virus were measured by GFP expression via flow cytometry and ddPCR detection of the genomic integrated lentivirus that contained GFP and the different rDNA landing pads (WT, 12-bp mutation, or 75-bp mutation). The GFP positive cell line from the 1:10 viral medium dilution (>99% GFP+) was chosen for subsequent experiments and cryopreserved.

To test if mutations in and around the R2 cleavage position can impact the Gene Writer system activity, the R2Tg Gene Writer Driver along with a plasmid that expressed a Gene Writer

transgene molecule were electroporated into the different landing pad cell lines. In order to test if the sequence in and around the cleavage site impacted the Gene Writer polypeptide sequence activity to integrate, the homology arms for the Gene Writer template molecule were designed to have 100% homology 100 bp to the left (Gene Writer molecule module A) and 100 bp to the right (Gene Writer molecule module F) of the cleavage position for each of the landing pads. The changes to the homology arms of the Gene Writer template molecule expression plasmid were introduced by PCR and were confirmed by Sanger Sequencing. Either 73 ng of the WT R2Tg Gene Writer Driver or the Endonuclease domain mutant R2Tg Gene Writer Driver expression plasmids were co-nucleofected) using nucleofection program DN100 into each of the different U2OS landing pad cell lines (WT, 12-bp mutant, or 75-bp mutant) with 177 ng of plasmids that expressed the Gene Writer template molecules that had 100% homology to either the WT landing pad, 12-bp mutant landing pad, or 75-bp mutant landing pad. After nucleofection, cells were grown at 37˚C, 10% CO_2 for 3 days prior to cell lysis and genomic DNA extraction. The extracted gDNA was measured for Gene Writer template molecule integration at the landing pad site by ddPCR. The DNA nicking activity was measured by detection of insertions, deletions, and/or a combination of both insertions and deletions at the landing pad through next-generation sequence analysis of an amplicon that was generated from the landing pad found in the gDNA.

The integration activity of the R2Tg Gene Writer is greatly reduced when the cleavage region is mutated where there is no integration of a Gene Writer template molecule in either of the 12-bp or 75-bp landing pad cell lines (Figure 3a). Furthermore, integration is not detected with Gene Writer template molecules that have homology arms that correspond to either the 12-bp or 75-bp mutant landing pads. To rule out that the lost integration activity is due to incompatible homology arms, DNA nicking activity was measured by NGS analysis of the landing pad. The nicking activity is independent of the Gene Writer template molecule as the WT R2Tg Gene Writer driver had comparable indels at the WT landing pad with the WT, 12-bp mutant, or 75-bp mutant Gene Writer template molecule (Figure 3b). The 12-bp and 75-bp landing pads, regardless of Gene Writer template molecule co-nucleofected with the WT R2Tg Gene Writer did not show any reads above background that contained indels. The level of indels was similar to the Gene Writer template driver containing endonuclease mutations.

Table E2: Landing Pad Information

Landing Pad Sequence Name	Sequence 5′-> 3′ (rDNA, underline; **cleavage region**, bold; ***mutated sequence***, bold-italic
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACG CGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCA ATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTT**AAGGTAG CCAAA**TGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGA GATTCCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGA AATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1658)
12-bp Mutant	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACG CGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCA ATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTT***TCCAATA TGATT***TGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGA GATTCCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGA AATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1659)
75-bp Mutant	Gctcacacaggaaacagctatgaccatgattacgccaagctgttgacg cgatgtgatttctgcccagtgctctgaatgtcaaagtgaagaaattca atgaagcgcgggtaaacggcgggagtaactatgactctctttt***ccaata tgattccacccatggcaaattccatggcaccgtcaaggctgagaacgg gaagcttgtcatcaatggaa***actatccagcgaaaccacagccaaggga attcactggccgtcgttttacaacgtcgtgactgggaaaaccctggc gttacccaacttaatcgccttgcagcacatcc (SEQ ID NO: 1660)

In some embodiments, a Gene Writer is derived from a retrotransposase with some level of target sequence specificity in the endonuclease domain. Thus, it may be desirable to retarget the Gene Writer to a location in the genome that possesses homology to the natural target sequence recognized by an endonuclease domain, referred to as the endonuclease recognition motif (ERM). In some embodiments, this sub-target sequence may be contained in the region surrounding the nick site. In specific embodiments, a 13 nt sequence (TAAGGTAGCCAAA) (SEQ ID NO: 1661) based on the nick site of an R2 element, e.g., R2Tg, is used to search the human genome for suitable locations for retargeting the Gene Writer, wherein a heterologous DNA-binding domain is designed to localize the Gene Writer to an endogenous ERM to direct endonuclease activity and subsequent retrotransposition of a template RNA. In some embodiments, the human genome site possesses 100% identity to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 nucleotides in the 13 nt motif. In further embodiments, the human genome site

containing the ERM is selected from Table E3, and a DNA-binding domain fusion, e.g., ZF, TAL, or dCas9 with a custom gRNA, is designed to localize the polypeptide to the site (e.g., see Example 3). In preferred embodiments, the genome site possesses a safe harbor score of at least 5, 6, 7, 8 as defined in Pellenz et al Hum Gene Ther 30, 814-282 (2019) and shown in Table E3. In some embodiments, the template RNA (or DNA encoding the template RNA) is designed such that the homology arms match the flanking genomic sequences surrounding the expected nick site at the new target.

Table E3: Human genome sites matching a 13 nt stretch around the nicking site of R2 elements.

The human genome was searched for 100% identity to the full 13 nt match or 12 consecutive nucleotides ("Match"). Chromosomal location and start and end coordinates are provided for each match. Score ("Score") is a metric evaluating each site for eight desirable safe harbor characteristics.

Chromosome	Start	End	Source	Match	Score
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

Example 3: Retargeting of a Gene Writer to a genomic safe harbor site

This example describes a Gene Writer comprising a heterologous DNA binding domain that redirects its activity to a genomic safe harbor site.

In this example, the Gene Writer polypeptide sequence is altered to where its natural DNA binding domain is replaced, mutated/inactivated, and/or joined with another polypeptide sequence that can redirect the Gene Writer system to another genomic location that is not its endogenous or natural binding site. In some instances, the polypeptide sequence that redirects the Gene Writer system to a non-natural genomic location may also be attached and/or inserted to any module of the Gene Writer polypeptide sequence.

In some embodiments, the polypeptide sequence used to redirect the Gene Writer system to a non-natural genomic target encodes for: a zinc finger, a series of adjacent, regularly, or irregularly spaced zinc fingers, a transcription activator-like effector (TALE), a series of adjacent, regularly, or irregularly spaced a transcription activator-like effectors (TALEs), Cas9, Cas9 with mutations to its catalytic residues inactivating the double-stranded DNA endonuclease activity (referred to as catalytically-dead Cas9 (dCas9)), Cas9 with a point mutation or multiple point mutations in a single catalytic domain in order to render Cas9 endonuclease only able to cleave one strand of double-stranded DNA (referred to as Cas9 nickase) (see Fig. 5).

In some embodiments, the polypeptide sequence used to re-direct the Gene Writer system targets a genomic safe-harbor location (*e.g.,* AAVS1 site on human chromosome 19) (Pellenz, S., et. al., Human Gene Therapy, 30(7), 814–828, 2019), see Figures 4 and 6. In further embodiments, the polypeptide sequence used to re-direct the Gene Writer polypeptide sequence is used in conjunction with a nucleic acid that targets the genomic safe harbor location (*e.g.,* the polypeptide sequence for catalytic dead Cas9 along with a single-guide RNA targeting the AAVS1 site on chromosome 19).

Table E4: Re-targeted Gene Writer constructs. Examples shown are to re-target R2Tg Gene Writer polypeptide sequence to the AAVS1 site using ZF or Cas9 domains.

Gene Writer Polypeptide Name	Polypeptide Sequence (*Re-targeting polypeptide sequence, italic*; **Linker, bold underline**)
AAVS1 Left ZFP attached at v2 location of DBD of R2Tg with 3GS linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHLRGS***GGGS***TATRDKKDTVTREKHPKKPFQKWMKDRAI KKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSF KSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGS ILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRP LGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRN*

	THTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAF ADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTIND CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCT CDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQ LHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNW RKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSL EEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGL SKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1662)
AAVS1 Left ZFP attached at v2 location of DBD of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES**TATRDKKDTVTREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1663)
AAVS1 Left ZFP attached at v1 location of DBD of R2Tg with 3GS linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***GGGS**CFGCLESISQIRTATRDKKDTVTREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRK

	LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1664)
AAVS1 Left ZFP attached at v1 location of DBD of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES***CFGCLESISQIRTATRD KKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIEC LSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTV LIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIR AAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKT ALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQAR RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNH WIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDL IFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTF VGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFAS RARKSMVM (SEQ ID NO: 1665)*
AAVS1 Right ZFP attached at v2 location of DBD of R2Tg with 3GS linker	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG RKFAQGYNLAGHTKIHLRGS***GGGS***TATRDKKDTVTREKHPKKPFQKWMKDRAI KKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSF KSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGS ILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRP LGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRN THTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAF ADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTIND CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCT CDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQ LHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNW RKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSL EEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGL SKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1666)*
AAVS1 Right ZFP attached at v2	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG*

location of DBD of R2Tg with XTEN linker	*RKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATPES**TATRDKKDTVTREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1667)
AAVS1 Right ZFP attached at v1 location of DBD of R2Tg with 3GS linker	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA* *RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG* *RKFAQGYNLAGHTKIHLRGS***GGGS**CFGCLESISQIRTATRDKKDTVTREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1668)
AAVS1 Right ZFP attached at v1 location of DBD of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA* *RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG* *RKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATPES**CFGCLESISQIRTATRD KKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIEC LSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTV LIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIR AAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDP

	LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKT ALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQAR RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNH WIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDL IFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTF VGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFAS RARKSMVM (SEQ ID NO: 1669)
AAVS1 Left and Right ZFP (separated by XTEN linker) attached at v2 location of DBD of R2Tg with 3GS linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES***GIHGVPAAMAERPFQCR* *ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKP* *FQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIHLR* *GS***GGGS**TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGK LAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGK IPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKAC PLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPM SPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNI SILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGE SEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRL IYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITK LAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPS IWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGI VNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC RHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIR DSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEV RHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1670)
AAVS1 Left and Right ZFP (separated by XTEN linker) attached at v2 location of DBD of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES***GIHGVPAAMAERPFQCR* *ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKP* *FQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIHLR* *GS***SGSETPGTSESATPES**TATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYL RFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDF KTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT MEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLF SRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFV DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKI QIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVL LSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQK

	TDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILY SSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFK KWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKK KDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAE KVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQV KMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1671)
AAVS1 Left ZFP attached to N-terminus of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES**ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNP PCPCCGTRVNSVLNLIEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCR GPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLL SRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVTREKHPKKP FQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1672)
AAVS1 Right ZFP attached to N-terminus of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA* *RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG* *RKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATPES**ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNP PCPCCGTRVNSVLNLIEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCR GPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS

	NRGAHKRCWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLL SRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVTREKHPKKP FQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1673)
AAVS1 Left and Right ZFP attached to N-terminus of R2Tg with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES***GIHGVPAAMAERPFQCR* *ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKP* *FQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIHLR* *GS****SGSETPGTSESATPES***ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDL VSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHV YECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEK ESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEH LKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWICEVCN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEELLI RLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEPGTCHHT RRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRHPDKGN GQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQ TAQDCFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTME IFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSR IVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDI AKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLS DSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTING TPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTD ILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSS TRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQ

	AGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKD WVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKV RKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKM AETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1674)
AAVS1 Left ZFP attached to N-terminus of R2Tg containing DBD inactivation mutations with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES**ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNP PSPSSGTRVNSVLNLIEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCR GPETEKAPAGEWISEVSNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKACATKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLL SRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVTREKHPKKP FQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1675)
AAVS1 Right ZFP attached to N-terminus of R2Tg containing DBD inactivation mutations with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA* *RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG* *RKFAQGYNLAGHTKIHLRGS***SGSETPGTSESATPES**ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNP PSPSSGTRVNSVLNLIEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCR GPETEKAPAGEWISEVSNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKACATKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLL SRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRA

	AAGGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVTREKHPKKP FQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1676)
AAVS1 Left and Right ZFP attached to N-terminus of R2Tg containing DBD inactivation mutations with XTEN linker	*MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA* *RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG* *RKFAQNSTRIGHTKIHLRGS***SGSETPGTSESATPES***GIHGVPAAMAERPFQCR* *ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKP* *FQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIHLR* *GS***SGSETPGTSESATPES**ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDL VSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHV YECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEK ESEDAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPSPSSGTRVNSVLNLIEH LKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWISEVSN RDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKACATKEEEELLI RLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEPGTCHHT RRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRHPDKGN GQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQ TAQDCFGCLESISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTME IFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSR IVTARLSKACPLNPRQRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFVDI AKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLS DSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTING TPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTD ILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSS TRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQ AGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR

	GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKD WVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEEAAAEKV RKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQVKM AETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1677)
S. pyogenes Cas9 attached at v2 location of DBD of R2Tg with XTEN33aa linker	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV* *DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL* *SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL* *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV* *LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK* *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG

	KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1678)
S. pyogenes Cas9 containing catalytic mutations (dCas9) attached at v2 location of DBD of R2Tg with XTEN33aa linker	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL SQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1679)*

S. pyogenes Cas9 D10A nickase mutant attached at v2 location of DBD of R2Tg with XTEN33aa linker	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV* *DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL* *SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL* *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV* *LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK* *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 16780)
S. pyogenes Cas9 N863A nickase mutant attached at	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV*

v2 location of DBD of R2Tg with XTEN33aa linker	*DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL* *SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL* *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV* *LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK* *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1681)
S. pyogenes Cas9 D10A nickase mutant attached at v2 location of DBD of R2Tg containing	*MAPKKKRKVGIHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV* *DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL*

EN mutation with XTEN33aa linker	*SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL* *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV* *LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK* *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1682)
S. pyogenes Cas9 N863A nickase mutant attached at v2 location of DBD of R2Tg containing EN mutation with XTEN33aa linker	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV* *DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL* *SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR*

	FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV* *LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK* *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYIITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1683)
S. pyogenes Cas9 attached at v2 location of DBD of R2Tg containing EN mutation with XTEN33aa linker	*MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN* *TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM* *AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV* *DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF* *EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT* *PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL* *SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS* *KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG* *SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR* *FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL* *YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED* *YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV*

	LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK *QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN* *LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE* *RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR* *LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ* *LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR* *MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA* *VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN* *FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK* *TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE* *KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL* *FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ* *LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI* *IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL* *SQLGGD***SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENLK LLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGL STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1684)

Example 4: Inactivation of an endogenous nucleolar localization signal in a Gene Writer

This example describes a Gene Writer in which an endogenous nucleolar localization signal has been inactivated to reduce intracellular targeting of the protein to the nucleolus.

In this example, the nucleolar localization signal (NoLS) of a retrotransposase is computationally predicted using a published algorithm that was trained on validated proteins that localize to the nucleolus (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–7399 (2010)).

The predicted NoLS sequence is based on both amino acid sequence, amino acid sequence context, and predicted secondary structure of the retrotransposase. The identified sequence is typically rich with basic amino acids (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–7399 (2010)) and when these residues are mutated to a simple side-chain, non-basic, amino acids or removed from the retrotransposase polypeptide chain then it can prevent localization to the nucleolus (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)). In some embodiments, the NoLS sequence is located in the amino acid region of a retrotransposase that is between the reverse transcriptase polymerase motif and the restriction-like endonuclease motifs. The predicted NoLS region contains lysine, arginine, histidine, and/or glutamine amino acids where nucleolar localization is inactivated by mutation of one or more of these residues to an alanine amino acid residue and/or one or more of these amino acids are removed from the polypeptide chain of the retrotransposase. In some embodiments, the amino acid sequence of the Gene Writer driver of R2Tg found upstream of the RLE is mutated such that lysines (K) are substituted for alanines (A), e.g., the predicted NoLS of R2Tg (amino acids 1,128-1,154 of polypeptide sequence), (APTQKDKFPKPCNWRKNEFKKWTKLAS (SEQ ID NO: 1685)) is mutated at 1, 2, 3, 4, 5, 6, or 7 residues to produce an inactivated NoLS (APTQ**A**D**A**FP**A**PCNWR**A**NEF**AA**WT**A**LAS (SEQ ID NO: 1686)).

Example 5: Application of second-strand nicking in a Gene Writer system

This example describes a Gene Writer system in which retrotransposition is paired with targeted second-strand nicking activity in order to increase the efficiency of integration events. The second strand nick can be achieved by (1) a Cas9 nickase fused to a gene writer system, in which the Gene Writer introduces one nick through its endonuclease domain (EN), and the fused nickase Cas9 places another nick on either the top and bottom DNA strands (Fig 7A), or (2) a GeneWriter system in which the active EN domain introduces a nick, and a Cas9 nickase introduces a second nick on either top or bottom strand of the DNA, upstream or downstream of the Gene Writer induced nick (Fig 7B).

In the first part of this example, a Cas9 nickase is fused to a Gene Writer protein (Fig 7A). The Cas9 is targeted to a DNA sequence through a gRNA. The Gene Writer protein

introduces a DNA nick through its EN domain, and an additional nick is generated through the nickase Cas9 activity. This additional nick can be targeted to the top or bottom strands of the DNA surrounding the Gene Writer introduced nick (Fig 1 A). Constructs designed and tested include (see schematic Fig 8A):

- Cas9-N863A-R2tg (RBD*, RT, EN)
- Cas9-H840A-R2tg (RBD*, RT, EN)
- Cas9-D10A-R2tg (RBD*, RT, EN)
- dCas9- R2tg (RBD*, RT, EN)

The DNA binding domain is the nickase Cas9 that directs the Gene Writer molecule to a DNA target through a gRNA. The RNA binding domain (RBD) in this set of Gene Writer constructs is inactivated with a point mutation (RBD*). As a donor for insertion, constructs in which the R2Tg RNA binding domain is inactive use a gRNA that is extended at its 3' end to include donor sequence for genome modification (Fig 8B). These modifications include nucleotide substitutions, nucleotide deletions and nucleotide insertions. In this first set of experiments, the above constructs-R2Tg(RBD*, RT, EN) and dCas9- R2Tg(RBD*, RT, EN) fusions with a 3' extended gRNA template targeting the AAVS1 locus are delivered to U2OS cells through nucleofection in SE buffer using program DN100. gRNAs used include gRNAs for each construct that target either the bottom or top strand of DNA. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and amplicon sequencing followed by computational analysis using CRISPResso (indel analysis tool) are performed. 3' extended gRNA mediated insertions, deletions or nucleotide substitutions are observed upon delivery of dCas9-R2Tg(RBD*, RT, EN), and increased in frequency when delivering Nickase-Cas9-R2Tg(RBD*, RT, EN) constructs.

In the second part of this example, a Cas9 nickase is fused to a Gene Writer protein (Fig 7A). The Cas9 is targeted to a DNA sequence through a gRNA. The Gene Writer protein introduces a DNA nick through its EN domain, and an additional nick is generated through the nickase Cas9 activity. This additional nick can be targeted to the top or bottom strands of the DNA surrounding the Gene Writer introduced nick (Fig 7A). In contrast to the constructs listed above, the RNA binding domain of R2Tg is active (Fig 9A), and the template used for genome modification is a transgene flanked by UTRs (Fig 9B). Constructs include (see schematic Fig 9A):

- Cas9-N863A-R2tg (RBD, RT, EN)
- Cas9-H840A-R2tg (RBD, RT, EN)
- Cas9-D10A-R2tg (RBD, RT, EN)
- dCas9- R2tg (RBD, RT, EN)

The transgene flanked by UTRs requires homology arms at the site of nicking. To determine the site of nicking for the accurate design of homology arms for the donor transgene DNA, the above listed constructs are nucleofected into 200k U2OS cells with a gRNA targeting the AAVS1 locus using pulse code DN100. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and amplicon sequencing followed by computational analysis using CRISPResso as an indel analysis tool are performed. The nicking site of the EN domain is identified from the indels the EN domain produces at the AAVS1 site. Homology arms of 100 bp flanking the EN nicking site are designed and included in the transgene (see Fig 9 and 10 for location of homology arms in transgene). To achieve genome modification, Cas9-R2Tg fusion constructs listed above are nucleofected into U2OS cells, along with a gRNA targeting either the top or bottom strand of the AAVS1 locus, and the appropriate transgenes harboring homology to the previously determined nicking site. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and ddPCR is performed to detect transgene integration at the AAVS1 site. Integrations are observed upon delivery of dCas9-R2Tg(RBD, RT, EN), and increased in frequency when delivering Nickase-Cas9-R2Tg(RBD, RT, EN) constructs.

In another example, a Gene Writer protein is targeted to DNA through its DNA binding domain (Fig 7B). The Gene Writer protein will introduce a DNA nick at a DNA strand. In addition, a Cas9 nickase is used to generate a second nick either on the top or bottom strands of the DNA, upstream or downstream of the first nick. In this example, a Gene Writer plasmid targeting the AAVS1 site (Fig 10A) and with a UTR flanked transgene with homology to the AAVS1 site (Fig 10B) is nucleofected into 200k U2OS cells using pulse code DN100. The following Cas9 constructs are transfected alongside the Gene Writer plasmids (Fig 10C):

- Cas9-N863A
- Cas9-H840A
- Cas9-D10A
- dCas9

All Cas9 constructs are co-nucleofected with gRNAs targeting the AAVS1 locus on either the top or bottom strands, upstream or downstream of the Gene Writer introduced nick. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and ddPCR is performed to detect transgene integration at the AAVS1 site. Integrations are observed upon delivery of dCas9 and increased in frequency when delivering Nickase-Cas9 constructs.

Example 6: Improved expression of Gene Writer polypeptide by heterologous UTRs

This example describes the use of heterologous UTRs to enhance the intracellular expression of the Gene Writer polypeptide.

In this example, the Gene Writer polypeptide was expressed from mRNA (Figure 11). In the plasmid template for the mRNA production, the native retrotransposon UTRs were replaced with UTRs optimized for the protein expression (C3 5'UTR and ORM 3' UTR from Asrani et al., RNA biology 15, 756-762 (2018) or 5' and 3' UTRs from Richter et al., Cell 168, 1114-1125 (2017)). The plasmid included the T7 promoter followed by the 5'UTR, the retrotransposon coding sequence, the 3' UTR, 3GS linker (SEQ ID NO: 1024), SV40 nuclear localization signal (NLS), XTEN linker, HiBit sequence and 96-100 nucleotide long poly(A) tail (SEQ ID NO: 1687). The plasmid was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of poly(A) tail and used for in vitro transcription (IVT) using T7 polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB). After the buffer exchange step the enzymatic capping reaction was performed using Vaccinia capping enzyme (NEB) and 2'-O-methyltransferase (NEB) in the presence of GTP and SAM (NEB). The capped RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were transfected with 0.5 µg with the Gene Writer mRNA in the presence or in the absence of the RNA template in 1:1 molar ratio using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format). The RNA template was in vitro transcribed from plasmid as described in Example 8 (Improved Gene Writer components for RNA-based delivery).

After transfection HEK293T cells were grown for 5 hours before assaying the Gene Writer expression by probing its HiBit tag expression using standard protocol https://www.promega.com/-/media/files/resources/protocols/technical-manuals/500/nano-glo-hibit-lytic-detection-system-technical-manual.pdf?la=en. Protein expression was found to be greatly improved by the use of 5' and 3' UTR_{exp} from C3-ORM as compared to using the native

UTRs from R2Tg (Figure 11). The genome integration was assayed 3 days post-transfection using 3' ddPCR (Figure 12).

Example 7: Improved Gene Writer components for mixed RNA and DNA delivery

This example describes improvements to the RNA molecule encoding a Gene Writer polypeptide that enhance expression and allow for increased efficiency of retrotransposition when used with a Gene Writer template encoded on plasmid DNA.

In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 6 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template was synthesized such that the reporter gene (eGFP) was flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The tempate expression was driven by the mammalian CMV promoter. We introduced the plasmid into HEK393T cells using the FuGENE® HD transfection reagent. HEK293T cells were seeded in 96-well plates at 10,000 cells/ well 24 hours before transfection. On the transfection day, 0.5 µl transfection reagent and 80 ng DNA was mixed in 10 µl Opti-MEM and incubated for 15 minutes at room temperature. The transfection mixture was then added to the medium of the seeded cells. Cells were detached and used for the electroporation of 0.5 µg of mRNA per well using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

HEK293T cells were transfected with the following test agents:

1. mRNA coding for the polypeptide described above
2. Plasmid encoding template RNA described above
3. Combination of 1 and 2. The plasmid was pre-lipofected 24 hrs before mRNA transfection as described above.

After transfection, HEK293T cells were cultured for 1-3 days and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells.

ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected product across 5' and 3' ends of integration junctions. The results of the ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 13. The genome integration in the

presence of the mRNA and the template plasmid achieved a mean copy number of 0.683 integrants/genome when targeting 3' junction and of 0.249 integrants/genome when targeting 5' junction. The mRNA only transfection resulted in a mean copy number of 0.002 integrants/genome, in comparison to 0.0004 integrants/genome for the plasmid only transfection.

Example 8: Improved Gene Writer components for RNA-based delivery

This example describes improvements to the RNA molecule encoding a Gene Writer polypeptide that enhance expression and allow for increased efficiency of retrotransposition when co-delivered with a Gene Writer RNA template.

In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 6 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template for the RNA template production included T7 promoter followed by the IRES-expressing reporter gene (eGFP) flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The plasmid template was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of the RNA template sequence and used for in vitro transcription (IVT) using T7 RNA polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB) and either enzymatically polyadenylated by poly(A) polymerase (NEB) or not. After the buffer exchange step the enzymatic capping reaction was performed using Vaccinia capping enzyme (NEB) and 2'-O-methyltransferase (NEB) in the presence of GTP and SAM (NEB). The capped RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were co-transfected with 0.5 to 1 µg of the GeneWriter mRNA and the RNA template in 1:4 to 1:12 molar ratios. The Neon transfection system was used for the RNA transfection (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

After transfection, HEK293T cells were cultured for at least 1 day and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells. ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected product across 5' and 3' ends of integration junctions. The mean copy number of 0.498 integrants/genome was achieved in the presence of the 0.5 µg of mRNA and 1:8 molar ratio of

Gene Writer mRNA to the RNA template when the RNA template was enzymatically polyadenylated, in comparison to that of 0.031 integrants/genome when the RNA transgene was not polyadenylated.

Example 9: Gene Writers that deliver genetic cargo containing introns

This example describes the use of a Gene Writer system to integrate genetic cargo that contains introns by using RNA-based delivery to tune expression of the gene of interest from its newly introduced genomic locus.

In this example, Gene writing technology uses an RNA template encoding a protein of interest including its native or non-native introns. For example, intron 6 of the triose phosphate isomerase (TPI) gene (Nott et al., 2003) will be used as one of the non-native introns in these experiments.

The presence of introns in the genomic copy of a gene and their removal by splicing has been reported to affect nearly every aspect of the gene expression, including its transcription rate, the mRNA processing, export, cell localization, translation and decay (reviewed in Shaul International Journal of Biochemistry and Cell Biology 91B, 145-155 (2017)). The introns can be inserted into different parts of the RNA template (Figure 15) and depending on the intron location their role in gene expression can differ.

An intron in the 5' UTR_{exp}, close to the transcription start site, introduces activating chromatin modifications (Bieberstein et al., Cell Reports 2, 62-68 (2012)), improves accuracy of transcription start site recognition and facilitates PolII recruitment (Laxa et al., Plant Physiology 172, 313-327 (2016)), increases rates of transcription initiation (Kwek et al., Nature Structural Biology 9, 800-805 (2002)) and elongation (Lin et al., Nature Structural and Molecular Biology 15, 819-826 (2008)), and improve the productive elongation in the sense relative to the antisense orientation (Almada et al., Nature 499, 360-363 (2013)).

An intron in the 3' UTR_{exp} limits the mRNA expression to one protein molecule per mRNA: the exon junction complex (EJC) left by spliceosome downstream of stop codon is recognized by the nonsense-mediated decay (NMD) machinery and therefore the mRNA is marked for deletion at the end of the pioneering round of translation (Zhang et al., RNA 4, 801-815 (1998)).

The ability to employ introns in a therapeutic gene may, however, be limited by splicing that occurs prior to integration of the template. For example, an intron in the forward orientation would be spliced out when an RNA template was encoded and delivered on a DNA plasmid, since transcription in the same direction would yield a template RNA that would be spliced prior to integration, thus failing to incorporate the intron in the genome. Additionally, lentivirus constructs designed to deliver a transgene must encode a sequence with an intron in the reverse orientation, since the viral packaging process would result in intron splicing and absence of the intron in packaged viral particles (Miller et al. J Virol 62, 4337-45 (1988)). However, the reverse orientation has also been thought to result in a reduction in viral titer and transduction (Uchida et al., Nat Commun 10, 4479 (2019)). It is worth noting that since the Gene Writer template can be generated through in vitro transcription and delivered directly as RNA, the problem of pre-integration splicing of desired introns can be avoided. In some embodiments, the Gene Writer template may thus contain one or more introns in same-sense orientation with the transcript, which is generated by IVT and delivered to the target cell as RNA.

An intron in any location depicted in Figure 15 will recruit U1 snRNP that protects mRNA from the premature cleavage and polyadenylation (Kaida et al., Nature 468 664-681 (2010); Berg et al., Cell 150, 53-64 (2012)). In addition, the EJC interacts with components of the TREX (transcription-export) complex and increases the rate of mRNA export from nucleus to cytoplasm 6-10-fold in comparison to the constructs lacking introns (Valencia et al., PNAS 105, 3386-3391 (2008)). It was also demonstrated that the binding of the polypyrimidine tract-binding protein, a splicing regulator protein, mediates a significant increase in the half-life of the spliced transcripts (Lu & Cullen, RNA 9, 618-630 (2003); Millevoi et al., Nucleic Acid Research 37, 4672-4683 (2009)). The efficiency of the mRNA translation was shown to be increased by the presence of the SR proteins (serine-arginine rich proteins, involved in RNA splicing) (Sanford et al., Genes & Development 18, 755-768 (2004); Sato et al., Molecular Cell 29, 255-262 (2008)) and the EJC proteins and its peripheral factors (Nott et al., Genes & Development 18, 210-222 (2004)).

In this example both the template RNAs harboring an intron or introns and Gene Writer polypeptide are delivered to the cells as *in vitro* transcribed capped RNAs as described in Example 8 (Improved Gene Writer components for RNA-based delivery). One to three days post-transfection the GOI expression and the genomic integration are assayed. In some

embodiments, the genome integration and/or protein expression will be higher for the intron-containing RNA template.

Example 10: Engineering of the retrotransposon 5' UTR to improve efficiency of integration

This example describes the deletion, replacement, or mutation of the 5'UTR of a retrotransposon to increase integration efficiency.

The 5'UTR region of non-LTR retrotransposons has multiple functions including self-cleaving ribozyme activity, which has been shown in certain elements and is predicted in additional retrotransposons (see modules B and C of Figure 26-27) (Ruminski et al. J Biol Chem 286, 41286-41295 (2011)). Ribozymal activity is predicted to cleave the RNA within or upstream of the 5'UTR. Either increasing or restricting this activity and structural component of the 5'UTR may benefit retrotransposition efficiency. A prediction of the ribozyme structure of R2Tg is provided in Figure 28.

In order to evaluate engineering of the 5'UTR, constructs were designed to enhance or diminish these activities (Figure 14). In case (A), the natural 5'UTR of R2Tg is used to integrate in trans as in previous experiments. Case (B) illustrates deletion of the 5'UTR. (C) and (D) represent cases in which the 5'UTR from the original species (in this case R2Tg from *T. guttata*) has been replaced by the 5'UTR of a retrotransposon from a distinct species. Case (C) provides an example in which the 5'UTR from *A. maritima* R2 has replaced that of R2Tg. (D) represents the generic case in which UTRs from additional species may be substituted ("Rx"), such as that from *B. mori*, *D. ananasse*, *F. auricularia*, *L. polyphemus*, *N. girault*i, or *O. latipes*, or from a retrotransposon selected from a Table herein, or any of Tables 1-3 of PCT/US2019/048607, herein incorporated by reference in its entirety. Case (E) represents the substitution of a ribozyme, such as a hammerhead ribozyme, e.g., RiboJ (Lou et al Nat Biotechnol 30, 1137-1142 (2012)). Case (F) represents the inactivation of the 5'UTR of R2Tg through point mutations, e.g., 75C>T in the 5' UTR (Figure 14.B, position indicated by shaded box). 5'UTR sequences are expected to be modular to any insertion sequence mediated by the retrotransposon.

Each case is evaluated as in previous examples by transfection of Gene Writer polypeptide plasmid with template plasmid and evaluation of integration frequency via ddPCR.

In some embodiments, substitution or mutation of the 5' UTR results in increased efficiency of integration.

Example 11: Modifying the 5' and 3' ends of Gene Writer RNA components to improve RNA stability

This example describes the addition of non-coding sequences to the 5' and 3' ends of RNA in order to improve stability in a mammalian cell.

The decay of eukaryotic RNAs in cells are mostly carried out by exoribonucleases. In this example, the half-life of RNAs is prolonged by introducing protective sequences and/or modifications at their 5' and 3' ends. The most common natural way of protecting the RNA ends is by introduction of 5' cap structure and 3' poly(A) tail. In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 6 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template for the RNA template production included T7 promoter followed by the IRES-expressing reporter gene (eGFP) flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The plasmid template was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of the RNA template sequence and used for in vitro transcription (IVT) using T7 polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB) and either enzymatically polyadenylated by poly(A) polymerase (NEB) or not. After the buffer exchange step the enzymatic capping reaction resulting in cap 1 structure was performed as described in Example 8 (Improved Gene Writer components for RNA-based delivery) or not performed. The template RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were co-transfected with 0.5 µg with the GeneWriter mRNA and the RNA template in 1:1 to 1:8 molar ratios using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

After transfection, HEK293T cells were cultured for 1-3 days and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells. ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected

product across 3' end of integration junctions. The genome integration was improved when the enzymatically capped and poly(A) tailed template was used (Figure 16).

The mean copy number of 0.498 integrants/genome was achieved in the presence of the 0.5 μg of mRNA and 1:8 molar ratio of mRNA:RNA template when the RNA template was enzymatically polyadenylated, in comparison to that of 0.031 integrants/genome when the RNA transgene was not enzymatically polyadenylated.

3' end modifications of RNAs.

It has been reported that the interactions between poly(A) tail shorter than 15-20 nts and the poly(A) binding protein (PABP) are destabilized resulting in the fast degradation of the RNA (Chang et al., Molecular Cell 53, 1044-1052 (2014); Subtelny et al., Nature 508, 66-71 (2014)). To determine the suitable lengths of the poly(A) tail of the template RNA we will test its lengths of 30, 40, 50, 60, 70, 80, 90 and 100 nucleotides. The IVT templates will be produced by PCR using reverse primers encoding the poly(A) tails of the abovementioned length. The IVT, DNase I treatment and capping of Gene Writer and the RNA template will be performed as described in Example 8 (Improved Gene Writer components for RNA-based delivery). After one to three days post-transfection the genomic integration will be assayed. In some embodiments, the genome integration will be higher for the RNA template tailed with a poly(A) tail of a suitable length.

In a cell the RNA degradation is initiated by shortening its poly(A) tail by deadenylases. Since the deadenylases are 3'-5' exoribonucleases favoring the poly(A) stretches, the terminal uridine, cytidine and most often guanine detected in the natural poly(A) tails of many mRNA were proposed to protect the poly(A) tail from its shortening (Chang et al., Molecular Cell 53, 1044-1052 (2014)). We will assay the Gene Writer and template RNAs with the encoded poly(A) tail with terminal G or C, or intermittent Gs or Cs (similar to that used in Lim et al., Science 361, 701-704 (2018)) according as described before.

Some of the RNAs have been described to evolve alternative ways of protections their 3' ends. A specific 16-nucleotide long stem-loop structure flanked with unpaired 5 nucleotides on both sides has been reported to protect the 3' end of mRNA encoding H2a.X histone (Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)). It has been shown that the heterologous mRNA ending with the histone stem-loop structure is cell cycle-regulated (Harris et al., Molecular Cellular Biology 11, 2416-2424 (1991); Stauber et al., EMBO Journal 5, 3297-3303 (1986)). The stem-loop structure is recognized and protected by the Stem-Loop Binding Protein

(SLBP). The protein accumulates shortly before cells enter S-phase and is rapidly degraded at the end of S-phase (Whifield et al., Molecular Cellular Biology 20, 4188-4198 (2000)). The stem-loop element will be inserted to the 3' end of the Gene Writer mRNA and the RNA templates and tested as described above to induce cell-cycle specific genome integration events.

Some viral and long non-coding RNAs have evolved to protect their 3' ends with triple-helical structures (Brown et al., PNAS 109, 19202-19207 (2012)). Additionally, the structural elements of tRNA, Y RNA and vault RNA (reviewed in Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)) have been reported to extend half-life of these non-coding RNAs. We will insert the structures to protect the 3' end of the RNA templates and probe their efficiencies in Gene Writing system as described above.

Finally, we will incorporate dNTP, 2'O-Methylated NTPs or phosphorothioate-NTP at the 3' of the RNA transgenes to increase the half-life of these molecules by protecting the 3' end of the RNA from exoribonucleases. We will incorporate single modified nucleotides or their stretches by extending the 3'end of the RNA by the DNA polymerases (for example, Klenow fragment) capable of extending an RNA sequence by adding modified nucleotides (Shcherbakova & Brenowitz, Nature Protocols 3, 288-302 (2008)).

A single nucleotide chemical modification of the 3' end of the RNA can be done by first oxidation of 3' terminal end of ribose sugar with sodium periodate to form a reactive aldehyde followed by conjugation of an aldehyde-reactive modified nucleotide.

Alternatively, T4 DNA or T4 RNA ligases can be used for the splinted ligation (Moore & Query, Methods in Enzymology 317, 109-123 (2000)) of the stretches of modified nucleotides to the 3' end of the RNAs.

Chemical ligation of two fragments is also possible. The phosphodiester bond linkage between two RNA substrates can be formed either by activating the phosphomonoester group using a reactive imidazolide or by using a condensing reagent such as cyanogen bromide. A disadvantage of chemical ligation is that it can also result in the creation of a 2'-5' phosphodiester linkage, together with the desired 3'-5' phosphodiester linkages.

5' end modifications of RNAs

In addition to the cap 1 structure described in Example (Improved Gene Writer components for RNA-based delivery) other 5' end protection groups will be explored. Particularly, we will use hypermethylated (Wurth et al. Nucleic Acid Res 42, 8663-8677 (2014)),

phosphorothioate (Kuhn et al., Gene Therapy 17, 961-971 (2010)), NAD^+-derived (Kiledjian, Trends in Cell Biology 28, 454-464 (2018)) and modified (for example, biotinylated: Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)) cap analogs for co-transcriptional capping.

We will also label the 5' of the RNA with 5'-[γ-thio]triphosphate to create a reactive sulfur group and chemically modify the 5' end with the protective modifications using a haloacetamide derivative of the modified group.

The proposed modifications to protect 3' and 5' end of the RNA will be introduced in RNA templates and/or Gene Writer mRNA (if compatible with translation). The genome integration efficiencies of the RNAs will be tested as described in Example 8 (Improved Gene Writer components for RNA-based delivery).

Example 12: Use of modified RNA bases in a Gene Writer system

This example describes Gene Writer systems comprising modified RNA bases to potentially improve features of the system, e.g., increase efficiency of integration, decrease cellular response to foreign nucleic acids. For the Gene Writer polypeptide, the proposed modifications pertaining to the coding region are compatible with translation. For the RNA template, the proposed modifications are compatible with reverse transcription.

In this example, mRNA encoding the Gene Writer polypeptide was in vitro transcribed with a 100% replacement of the corresponding rNTP with one of the modified rNTPs: pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U) or 5-methylcytidine (5mC). Otherwise, the RNA preparation, purification and cell transfections were performed as described in the Example 8 (Improved Gene Writer components for RNA-based delivery). The gene integration capacity of the modified mRNAs was compared with that of the non-modified mRNA (G0) using ddPCR, with all polypeptide mRNAs being paired with an unmodified template RNA (Figure 17). Integration was detected when the polypeptide was encoded using each modified rNTP, with the highest signal coming from 5-MO-U and the lowest from 5mC. This demonstrates that the Gene Writer polypeptide component is functional when expressed from mRNA containing modified bases.

Further, this example describes the modularity of the Gene Writer template molecule where it is composed of all or a subset of the exemplary modules listed in Figure 19 and

illustrated in Figure 18. Individual modules can be produced by chemical or in vitro syntheses as a contiguous nucleic acid molecule or in separate pieces that are later combined together. The individual modules of the Gene Writer template molecule can be chemically modified nucleic acids, be comprised in part or in entirety of non-nucleic acids, re-arranged in order, and/or omitted to form the Gene Writer template molecule.

In some embodiments, the Gene Writer template molecule (all modules, A-F) is synthesized by in vitro transcription where 0-100% replacement of a corresponding rNTP (adenosine, cytidine, guanosine, and/or uridine) is with one or more modified rNTPs (base or ribose modification), e.g., 5′ hydroxyl, 5′ Phosphate, 2′-O-methyl, 2′-O-ethyl, 2′-fluoro, ribothymidine, C-5 propynyl-dC (pdC), C-5 propynyl-dU (pdU), C-5 propynyl-C (pC), C-5 propynyl-U (pU), 5-methyl C, 5-methyl U, 5-methyl dC, 5-methyl dU methoxy, (2,6-diaminopurine), 5′-Dimethoxytrityl-N4-ethyl-2′-deoxyCytidine, C-5 propynyl-fC (pfC), C-5 propynyl-fU (pfU), 5-methyl fC, 5-methyl fU, C-5 propynyl-mC (pmC), C-5 propynyl-fU (pmU), 5-methyl mC, 5-methyl mU, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U). The modified nucleotides in this embodiment rely on incorporation through a transcription reaction which utilizes a natural or mutant polypeptide sequence of a RNA polymerase that readily incorporates modified nucleotides into a RNA transcript that is made 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). The modified Gene Writer template molecule is typically in whole or in part compatible with the reverse transcriptase activity of the Gene Writer polypeptide sequence; for modules or parts of modules of the Gene Writer template molecule used as a template for reverse transcription, preference is given to modifications that are compatible with reverse transcription (Motorin et al., Methods in Enzymology 425 21-53, 2007; Mauger et al., PNAS 116, 24075-24083, 2019). Gene Writer systems with template molecules containing modified rNTPs are tested as described above and in Example 8 (Improved Gene Writer components for RNA-based delivery).

In some embodiments, individual modules are chemically synthesized containing modified nucleotides, e.g., 5′ hydroxyl, 5′ Phosphate, 2′-O-methyl, 2′-O-ethyl, 2′-fluoro, ribothymidine, C-5 propynyl-dC (pdC), C-5 propynyl-dU (pdU), C-5 propynyl-C (pC), C-5 propynyl-U (pU), 5-methyl C, 5-methyl U, 5-methyl dC, 5-methyl dU methoxy, (2,6-

diaminopurine), 5′-Dimethoxytrityl-N4-ethyl-2′-deoxyCytidine, C-5 propynyl-fC (pfC), C-5 propynyl-fU (pfU), 5-methyl fC, 5-methyl fU, C-5 propynyl-mC (pmC), C-5 propynyl-fU (pmU), 5-methyl mC, 5-methyl mU, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), where the individual modules are then ligated together through enzymatic (e.g., splint ligation using T4 DNA ligase, Moore, M. J., & Query, C. C. Methods in Enzymology, 317, 109–123, 2000) or chemical processes (e.g., Fedorova, O. A., et. al., Nucleosides and Nucleotides, 15(6), 1137–1147, 1996) to form a complete Gene Writer template molecule.

An example of a modified Gene Writer template molecule is where modules A and F are each 100 nt of chemically synthesized RNA with cytidine and uridine nucleotides containing 2′-*O*-methyl ribose modifications and module A contains (3) phosphorothioate linkages between the first 3 nucleotides on the 5′ end and module F contains (3) phosphorothioate linkages between the last 3 nucleotides on the 3′ end of the module. Modules B-E are synthesized by in vitro transcription using an RNA polymerase (RNAP), e.g., T7 RNAP, T3 RNAP, or SP6 RNAP (NEB), or derivatives thereof that possess enhanced properties, e.g., increased fidelity, increased processivity, or increased efficiency of incorporating modified nucleotides. Module A is ligated to the 5′ end of the in vitro transcribed module B-E molecule and module F is ligated on to the 3′ end of the in vitro transcribed module B-E molecule by splint ligation (described by Moore, M. J., & Query, C. C. Methods in Enzymology, 317, 109–123, 2000). This fully assembled template RNA (all modules, A-F) is then used with a Gene Writer polypeptide (or nucleic acid encoding the polypeptide) in a target cell to assess genomic integration as in previous examples. In some embodiments, RNA modifications do not decrease the efficiency of integration greater than 50%, e.g., as measured by ddPCR. In some embodiments, RNA modifications improve the efficiency of integration, e.g., as measured by ddPCR. In some embodiments, RNA modifications improve the reverse transcription reaction, e.g., improve the processivity or fidelity as measured by sequencing of integration events.

Example 13: Gene Writer templates that do not incorporate UTRs

This example describes a configuration of the Gene Writer template molecule that results in an exclusion of the UTRs, such that these regions used in retrotransposition are not integrated into the host cell.

In this example, we describe the positioning, omission, and/or substitution of the UTR modules of the Gene Writer template molecule (Figures 18 and 19) to result in the Gene Writer driver to not incorporate the UTR modules into the genome as a part of retrotransposition. In some embodiments, the Gene Writer template molecule modules for the 5′ and 3′ UTRs (modules B+C and E of Gene Writer template molecule) are moved to the ends of the molecule so that their function of interacting with the Gene Writer driver does not change but the homology arm is now located adjacent to the heterologous object sequence (module D) where complementarity of the homology arms act as a primer for reverse transcription. In some cases, modules B and/or C are omitted from the Gene Writer template molecule with module E following module F.

Additional examples of not incorporating the UTRs into the genome are removing modules B and C from the Gene Writer template molecule, re-positioning module F (3' homology arm) to follow module D (heterologous object sequence) and have module E be substituted with a binding ligand such as biotin. This Gene Writer template molecule would now consist of module A (5' homology arm) – module D (heterologous object sequence) – module F (3' homology arm) – module E comprised of biotin. The Gene Writer driver polypeptide sequence would be modified to incorporate the amino acid sequence for monomeric streptavidin. This example illustrates how the utility of mediating a non-nucleic acid mediated association of the Gene Writer template molecule with the Gene Writer driver polypeptide sequence.

Example 14: Homology arm length impacts retrotransposition efficiency

This example describes modulation of homology arms (HA) flanking a transgene to increase the frequency of an associated retrotransposition event.

Retrotransposition is believed to be mediated by priming events on the 3’ of an integrated transgene. Priming of the transgene RNA by the nicked host genomic DNA requires homology between the 3’ end of the transgene RNA and the genomic DNA 3’ to the host nick. Although the method of 5’ resolution of the retrotransposition is unknown, this resolution may also benefit

from homology through host-mediated repair pathways. Additionally, processing of the 5' end of the transgene RNA may affect retrotransposition, i.e. through ribozymal cleavage upstream of the 5' UTR. Therefore, the flanking homology of the payload transgene was modulated in order to optimize retrotransposition.

Plasmid transfections were performed to test the effects of transgene homology on efficiency of integration in *trans*. Plasmid expressing R2Tg or control R2Tg with an endonuclease inactivating mutation were co-delivered with transgene plasmids containing differing lengths of target homology and optional random DNA stuffer sequence (Table E5 below, Figure 20). Stuffer sequence was used to control for the effect of transcript length on retrotransposition. 500ng total of plasmids at 1:4 molar ratio of driver:transgene plasmid was nucleofected into 2e5 293T cells via Lonza 4D shuttle with program DS-150. After 3 days, genomic DNA of the cells was isolated. Droplet digital PCR (ddPCR) was performed across each junction using a primer and Taqman probe specific for the transgene combined with a primer at the expected rDNA locus. The copy number found represents the efficiency of integration at the 3' or 5' end of the transgene.

Results are shown in Figure 21. The construct tested with the longest homology (100bp on either side) demonstrated the highest integration efficiency.

Table E5: Template constructs tested with varying lengths of homology at the 5' and 3' end, including stuffer sequence to maintain the total flanking sequence at 100 bp on each end.

5' Homology arm length (bp)	3' Homology arm length (bp)	Random stuffer presence (to 100bp)
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

Example 15: Homology arm tolerance and specificity impacts retrotransposition efficiency

This example describes the necessity for accurate alignment of homology arm design with nick location during retrotransposition.

Data from the Example 14 above highlights the importance of the homology arm for efficient retrotransposition. In many cases, the nicking location of a given retrotransposon may be poorly characterized. In these cases, the designed homology arm(s) may not initiate adjacent to the nick site in the genomic DNA. In order to evaluate the dependency of homology arm positioning on retrotransposition efficiency, we designed constructs with the 3' homology arm shifted by a number of bases relative to the known cut site of R2Tg (Table E6 below, Figure 22). In cases where the 3' homology arm was shifted 3' from the cut site (+), the homology arm was effectively shortened. In cases where the 3' homology arm was shifted 5' of the cut site, homologous bases were added from the 5' arm.

Driver and transgene plasmids were co-transfected as in the above example into 293T cells. At day 3, genomic DNA was extracted. The integration frequency was measured as above via ddPCR. A significant loss in integration frequency was noted with a shifted 3' homology arm in either position relative to the WT nick site (Figure 23).

Table E6: Template constructs tested with varying positions of homology at the 3' end, including relative shift of sequence position at the 3' end.

5' Homology arm length (bp)	3' Homology arm length (bp)	Right arm shift (relative to expected nick site)
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

Example 16: Gene Writers can integrate genetic cargo independently of the homology directed repair pathway

This example describes the use of a Gene Writer system in a human cell wherein the homologous recombination repair pathway is inhibited.

In this example, U2OS cells were treated with 30pmols (1.5μM) non-targeting control siRNA (Ctrl) or a siRNA against Rad51, a core component of the homologous recombination repair pathway. SiRNAs were co-delivered with R2Tg driver and transgene plasmid in *trans* (see Fig 24 for driver and transgene configuration schematic). Specifically, Plasmid expressing R2Tg, control R2Tg with a mutation in the RT domain, or control R2Tg with an endonuclease inactivating mutation were used in conjunction with transgene (Fig 25 A, B). A total of 250ng DNA plasmids with a 1:4 molar ration of driver to transgene, along with 30pmol of siRNAs were nucleofected into 200k U2OS cells resuspended in 20μL of nucleofection buffer SE using pulse code DN100. Protein lysates collected on day 3 showed the absence of Rad51 in the siRad51 treated condition (Fig 24C). gDNA was extracted at day 3 and ddPCR assays to detect transgene integration at the rDNA locus was performed. The results of the ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 25. The absence of Rad51 leads to a ~20% reduction in R2Tg mediated transgene integration at the rDNA locus both at the 3' and 5' junctions (Fig 25), indicating that R2TG mediated transgene insertion is not wholly dependent on the presence of the homologous recombination pathway, and can occur in the absence of the endogenous HR pathway. In some embodiments, HR independence enables Gene Writing to work in cells and tissues with endogenously low levels of HR, e.g., liver, brain, retina, muscle, bone, nerve, cells in G_0 or G_1 phase, non-dividing cells, senescent cells, terminally differentiated cells. In some embodiments, HR independence enables Gene Writing to work in cells or in patients or tissues containing cells with mutations in genes involved in the HR pathway, e.g., BRCA1, BRCA2, P53, RAD51.

Example 17: Gene Writers can integrate genetic cargo independently of the single-stranded template repair pathway

This example describes the use of a Gene Writer system in a human cell wherein the single-stranded template repair (SSTR) pathway is inhibited.

In this example, the SSTR pathway will be inhibited using siRNAs against the core components of the pathway: FANCA, FANCD2, FANCE, USP1. Control siRNAs of a non-target control will also be included. 200k U2OS cells will be nucleofected with 30pmols (1.5μM) siRNAs, as well as R2Tg driver and transgene plasmids (trans configuration). Specifically, 250 ng of Plasmids expressing R2Tg, control R2Tg with a mutation in the RT domain, or control R2Tg with an endonuclease inactivating mutation) are used in conjunction with transgene at a 1:4 molar ratio (driver to transgene). Transfections of U2OS cells is performed in SE buffer using program DN100. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3 and ddPCR is performed to assess integration at the rDNA site. Transgene integration at rDNA is detected in the absence of core SSTR pathway components.

Example 18: Gene Writer systems with enhanced activity for target vs non-target cells

This example describes the incorporation of regulatory sequences into Gene Writer systems in order to decrease integration activity in non-target cells.

In this example, genetic regulation is accomplished through (i) using tissue-specific promoters to upregulate component expression and integration in target cells and (ii) using miRNA binding sites to decrease integration in non-target cells that have increased endogenous levels of the corresponding miRNA. Target cells used are human hepatocytes and non-target cells are hematopoetic stem cells (HSCs). The driver of integration here is a plasmid encoding the Gene Writer polypeptide (e.g., R2Tg retrotransposase) driven by different promoters and with scrambled or specific miRNA binding sites after the coding sequence. The template for integration is encoded on plasmid DNA, such that transcription results in a homology- and UTR-flanked heterologous object sequence. The heterologous object sequence may comprise a reporter gene that is driven by different promoters and with scrambled or specific miRNA binding sites after the coding sequence. The control promoter used here is CMV and the control for miRNA binding site is a randomly scrambled version of the binding site for miR-142. The target tissue-specific promoter used here is ApoE.HCR.hAAT, which is expressed in liver cells, and the off-target tissue-specific miRNA binding site is complementary to miR-142 (uguaguguuccuacuuuaugga (SEQ ID NO: 1688)), which is expressed in HSCs.

Target cells and non-target cells are nucleofected with a combination of Gene Writer polypeptide (1) and template (2) selected from:

(HEK293T vs HEK293T with miRNA?)

Gene Writer polypeptide constructs (1):

a. Non-specific driver: CMV-R2Tg
b. Non-specific inactivated driver: CMV-R2Tg(EN*)
c. Tissue-specific driver: ApoE.HCR.hAAT-R2Tg-miR142
d. Tissue-specific inactivated driver: ApoE.HCR.hAAT-R2Tg(EN*)-miR142

Gene Writer template constructs (2):

a. Non-specific transgene: CMV-gfp
b. Tissue-specific transgene: ApoE.HCR.hAAT-gfp-miR142

Cells are incubated for at least three days and subsequently evaluated for integration efficiency and reporter expression. For integration efficiency, ddPCR is performed to quantify the average number of integrations per genome for each sample. In some embodiments, the ratio between the integration efficiency in target cells and non-target cells is higher when using a template paired with the tissue-specific driver (1a) vs a non-specific driver (1c). To assess reporter expression, cells are analyzed by flow cytometry to detect GFP fluorescence and RT-qPCR to detect transcription. In some embodiments, the ratio between fluorescence in target cells and non-target cells is higher when using a driver paired with a tissue-specific transgene cassette (2b) vs a non-specific transgene cassette (2a). In some embodiments, the ratio between transcript levels in target cells and non-target cells is higher when using a driver paired with a tissue-specific transgene cassette (2b) vs a non-specific transgene cassette (2a). In some embodiments, the combination of a tissue-specific driver (1a) with a tissue-specific transgene cassette (2b) results in the highest ratio of transcription or expression between target and non-target cells.

Example 19: Application of a Gene Writer™ system for delivering therapeutic gene to liver in a human chimeric liver mouse model.

This example describes a Gene Writer™ genome editing system delivered to the liver in vivo for integration and stable expression of a genetic payload. The promoter and miRNA recognition sequence for expression control and the therapeutic gene are intended to exemplify

the approach and are selected from Table 2 of WO2020014209, incorporated herein by reference, and Tables 3 and 4 of WO2020014209, respectively.

In this example, human hepatocytes derived from patients with OTC deficiency are engrafted into a mouse model (Ginn et al JHEP Reports 2019) and a Gene Writer™ system is used to deliver an OTC expression cassette for integration into liver cells. The Gene Writer™ polypeptide component comprises an expression cassette for the R2Tg retrotransposase (Table 3) and the template component comprises an expression cassette for the human OTC gene (Table 5 of WO2020014209) flanked by the UTR sequences required for binding and retrotransposition by R2Tg (Table 3) and further flanked by 100 nt homology to the target site in ribosomal DNA. In this example, both the transposase and template expression cassettes additionally comprise the hAAT promoter (Table 3 of WO2020014209) for hepatocyte-specific expression and a miRNA recognition sequence complementary to the seed sequence of miR-142 (Table 4 of WO2020014209) for downregulating expression in hematopoetic cells.

1. Gene Writer™ polypeptide component: rAAV2/NP59.hAAT.R2Tg
2. Endonuclease-mutated Gene Writer™ polypeptide: rAAV2/NP59.hAAT.R2TgEN*
3. Gene Writer™ template component: rAAV2/NP59.hAAT.OTC
4. Reporter Gene Writer™ template component: rAAV2/NP59.hAAT.GFP

Eight to 12-week-old female $Fah^{-/-}Rag2^{-/-}Il2rg^{-/-}$ (FRG) mice are engrafted with human hepatocytes, isolated from pediatric donors or purchased from Lonza (Basel, Switzerland), as described previously (Azuma et al Nat Biotechnol 2007). Engrafted mice are cycled on and off 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC) in drinking water to promote liver repopulation. Blood is collected every two weeks and at the end of the experiment to measure the levels of human albumin, used as a marker to estimate the level of engraftment, in serum by enzyme-linked immunosorbent assay (ELISA; Bethyl Laboratories, Inc., Montgomery, TX). Eleven weeks after engraftment, mice are treated with the Gene Writer™s packaged in NP59, a highly human hepatotropic AAV capsid. The following vectors are administered by i.p. injection:

- Active Gene Writing™ of therapeutic: (1) and (3)
- Active Gene Writing™ of reporter: (1) and (4)
- Integration-inactivated therapeutic control: (2) and (3)
- Integration-inactivated reporter control: (2) and (4)

After vector injection, mice are cycled on NTBC for another 5 weeks before being euthanized. DNA and RNA are subsequently extracted from liver lysates by standard methods. OTC expression is subsequently assayed by performing RT-qPCR on isolated RNA samples using sequence-specific primers. To confirm integration of construct and analyze genomic locations, unidirectional sequencing is performed on genomic DNA samples by using specific primers annealing to the inserted gene to read outward into the surrounding genomic sequence on a MiSeq.

Example 20: Application of a Gene Writer™ system for delivering therapeutic gene to liver in an infant or adult mouse model of a disease.

This example describes a Gene Writer™ genome editing system delivered to the liver in vivo for integration and stable expression of a genetic payload. The promoter and miRNA recognition sequence for expression control and the therapeutic gene are intended to exemplify the approach and are selected from Tables 2, 3, and 4 of WO2020014209, respectively.

In this example, an OTC deficient mouse model is used to assess a Gene Writer™ system designed to deliver an OTC expression cassette for integration into liver cells. The Gene Writer™ polypeptide component comprises an expression cassette for the the R2Tg retrotransposase (Table 3) and the template component comprises an expression cassette for the human OTC gene (Table 5 of WO2020014209) flanked by the UTR sequences required for binding and retrotransposition by R2Tg (Table 3) and further flanked by 100 nt homology to the target site in ribosomal DNA. In this example, both the transposase and template expression cassettes additionally comprise the hAAT promoter (Table 3 of WO2020014209) for hepatocyte-specific expression and a miRNA recognition sequence complementary to the seed sequence of miR-142 (Table 4 of WO2020014209) for downregulating expression in hematopoetic cells.

1. Gene Writer™ polypeptide component: rAAV2/8.hAAT.R2Tg
2. Endonuclease-mutated Gene Writer™ polypeptide: rAAV2/8.hAAT.R2TgEN*
3. Gene Writer™ template component: rAAV2/8.hAAT.OTC
4. Reporter Gene Writer™ template component: rAAV2/8.hAAT.GFP

Either one to two day-old or eight to 12-week-old female Otc-deficient *Spf^{ash}* mice (C57BL/6/C3H-F1 background) are treated with the Gene Writer™s packaged in AAV8, a hepatotropic AAV capsid. The following vectors are administered by i.p. injection:

- Active Gene Writing™ of therapeutic: (1) and (3)
- Active Gene Writing™ of reporter: (1) and (4)
- Integration-inactivated therapeutic control: (2) and (3)
- Integration-inactivated reporter control: (2) and (4)

After 5 weeks, DNA and RNA are subsequently extracted from liver lysates by standard methods. OTC expression is subsequently assayed by performing RT-qPCR on isolated RNA samples using sequence-specific primers. To confirm integration of construct and analyze genomic locations, unidirectional sequencing is performed on genomic DNA samples by using specific primers annealing to the inserted gene to read outward into the surrounding genomic sequence on a MiSeq.

Example 21: Ribozyme and Homology Arm sequence compatibility at re-targeted sites

This example describes a gene writer template molecule that is used in conjunction with a mutant gene writer polypeptide sequence that targets a genomic location that is outside of its natural genomic target site. Endogenous sequences of retrotransposon RNA contain a ribozyme at the 5′ end of their RNA, which is only active if the RNA forms the correct secondary structure or folding of the RNA (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). In order for the active form of the retrotransposon ribozyme to form the correct structure, the RNA of some retrotransposons must contain part of the 28S ribosomal RNA in order for the proper secondary structure of the P1 stem of the ribozyme to form (Eickbush, D. G., et. al., PLoS ONE, 8(9), 1–16, 2013).

The portion of the endogenous retrotransposon RNA that is the 28S ribosomal RNA that interacts with the 5′ UTR of the retrotransposon RNA is analogous to the gene writer template molecule (Figure 18 modularity) where a portion of the sequence in module A interacts with a portion of the sequence of module B of the gene writer template molecule. In order for module B to be an active ribozyme it needs to fold into a proper secondary structure, the P1′ portion of the ribozyme found in module B interact with some complementarity to the P1 sequence of the ribozyme found in module A. In some embodiments where the 5′ homology arms of the gene writer template molecule (module A) are required for integration activity along with having an

active ribozyme (module B), the sequence of the P1′ sequence of the ribozyme found in module B, is changed to have some complementarity to the sequence found in module A where the P1 sequence is. The nucleotide lengths of complementarity between the P1 sequence found in module A and the P1′ sequence of module B may vary between 0-100 nucleotides. In some embodiments, if the re-targeted gene writer polypeptide sequence is based off of the R2 element from *Taeniopygia guttata*, nucleotides 49-54 of module B of the gene writer template molecule interact through complete or partial complementary to the last 28 nucleotides of module A, where module A is the homology arm that has complementarity to a genomic location that is compatible with a mutant Gene Writer polypeptide sequence that targets the genomic region to which module A has complementarity (Figure 28 predicted ribozyme).

Example 22: Gene Writing integrates the specific template sequence.

This example describes analyzing the insertions completed by a Gene Writing system in order to assay for unintentional incorporation of non-template RNA, e.g., cellular endogenous RNA, into the target site.

In this example, a Gene Writing system is used as described in previous examples in order to integrate a template RNA into a target site in HEK293 cells. HEK293 are transfected with the following reagents:

1. mRNA encoding Gene Writer polypeptide
2. mRNA encoding inactivated Gene Writer polypeptide (e.g., R2Tg reverse transcriptase mutant)
3. Gene Writer RNA template (e.g., comprising 5'-3': 5'HA – 5'UTR – GFP cassette – 3'UTR – 3'HA)
4. Gene Writer RNA template without binding motif (e.g., comprising 5'-3': 5'HA – GFP cassette – 3'HA)

After 3 days of incubation, genomic DNA is extracted and analyzed for insertion frequency by ddPCR, as described elsewhere herein. In some embodiments, the combination of (1) and (2) will result in integration of the template. In some embodiments, the combination of (1) and (4) will not result in detectable integration of the template, since template (4) does not possess a polypeptide binding motif, e.g., the 3'UTR from the R2Tg retrotransposon. In some

embodiments, (1) and (4) will result in an integration frequency that is less than the frequency of integration with (1) and (2), e.g., is less than 10%, 5%, 4%, 3%, 2%, 1%, 0.05%, or 0.01%.

To further analyze all incorporated sequences, amplicon-seq is performed on cells derived from transfecting (1) and (2) by using PCR to amplify across the junction of the target site. Optionally, negative selection against unedited target sites can be performed by digesting specifically at the junction of an unedited target site (nicking site of the Gene Writer) in order to improve signal. Amplicons are processed for next generation sequencing on an Illumina MiSeq, as described previously. Unintentional incorporations are discovered by looking for reads that contain an inserted sequence, e.g., at least 100 nt, of new DNA that does not map to the template RNA. Optionally, the insertion is compared to the human transcriptome to determine the source of any transcript unintentionally incorporated into the target site. In some embodiments, the Gene Writing system will not incorporate any templates into the target site that are not the Gene Writing template RNA. In some embodiments, the Gene Writing system will not incorporate any templates into the target site that are not the Gene Writing template RNA at a level greater than 1% of the total insertions.

Example 23: Gene Writer™ enabling nucleotide substitution in genomic DNA to correct alpha-1 antitrypsin deficiency mutation in human cells

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide.

In this example, the Gene Writer™ polypeptide and writing template are provided as DNA transfected into HEK293T cells that possess the PiZ genotype (E342K), a common allele associated with alpha-1 antitrypsin deficiency. The Gene Writer™ polypeptide uses a Cas9 nickase for both DNA-binding and endonuclease functions. The writing template is designed to have homology to the target sequence, while incorporating additional nucleotides at the desired position, such that reverse transcription of the template RNA results in the generation of a new DNA strand containing the substitution.

To create the transversion in the affected human *SERPINA1* gene that restores the GAG triplet coding for glutamate in healthy patients, the Gene Writer™ polypeptide is used with a specific template nucleic acid, which encodes a gRNA scaffold for polypeptide binding, a spacer for polypeptide homing, target homology domain to set up TPRT, and a template sequence for reverse transcription that includes the required substitution. An exemplary template RNA carries

the sequence (1)TCCCCTCCAGGCCGTGCATA(2)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)TcGTCGATGGTC AGCACAGCCTTAT(4)GCACGGCCTGGA (SEQ ID NO: 1689), where numbers are used to delineate the modules of the template in the order (5'-3') (1) gRNA spacer, (2) gRNA scaffold, (3) heterologous object sequence, (4) 3' homology priming domain, and the lowercase "c" indicates the position in the template carrying the nucleotide substitution to be written into the target site to correct the E342K mutation. An exemplary gRNA for providing a second nick as described in embodiments of this system comprises the spacer sequence TTTGTTGAACTTGACCTCGG (SEQ ID NO: 1625) and directs a Cas9 nickase to nick the second strand of the target site within the homologous region. In some embodiments, this second nick improves the efficiency of the edit.

After transfection, cells are incubated for three days to allow for expression of the Gene Writing™ system and conversion of the genomic DNA target, and genomic DNA is extracted from cells. Genomic DNA is then subjected to PCR-based amplification using site-specific primers and amplicons are sequenced on an Illumina MiSeq according to manufacturer's protocols. Sequence analysis is then performed to determine the frequency of reads containing the desired edit.

Example 24: Correction of alpha-1 antitrypsin deficiency using lipid nanoparticles comprising Gene Writers.

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide in vivo. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse liver cells via lipid nanoparticles to correct the SERPINA1 PiZ mutation causing alpha-1 antitrypsin deficiency.

Formulation and treatment of murine models with LNPs (LNP-INT01 system) carrying Cas9 and gRNA are taught by Finn et al. *Cell Reports* 22:2227-2235 (2018), the methods of which are incorporated herein by reference.

Capped and polyadenylated Gene Writer polypeptide mRNA containing N1-methyl pseudo-U is generated by in vitro transcription using a linearized plasmid DNA template and T7 RNA polymerase. The polypeptide mRNA is purified from enzyme and nucleotides using a MegaClear Transcription Clean-up Kit, in accordance with the manufacturer's protocol

(ThermoFisher). The transcript concentration is determined by measuring the light absorbance at 260 nm (Nanodrop), and the transcript is analyzed by capillary electrophoresis by TapeStation (Agilent). Template RNA comprising the mutation correcting sequence is also prepared by in vitro transcription and translation using similar methods. In this example, the template RNA comprises the sequence as exemplified in Example 1.

LNPs are formulated with an amine-to-RNA-phosphate (N:P) ratio of 4.5. The lipid nanoparticle components are dissolved in 100% ethanol with the following molar ratios: 45 mol% LP01 lipid, 44 mol% cholesterol, 9 mol% DSPC, and 2 mol% PEG2k-DMG. The RNA cargo (1:40 molar ratio of polypeptide mRNA:template RNA) is dissolved in 50 mM acetate buffer (pH 4.5), resulting in a concentration of RNA cargo of approximately 0.45 mg/mL. LNPs are formed by microfluidic mixing of the lipid and RNA solutions using a Precision Nanosystems NanoAssemblr Benchtop Instrument, in accordance with the manufacturer's protocol. After mixing, the LNPs are collected and diluted in PBS (approximately 1:1), and then the remaining buffer is exchanged into PBS (100-fold excess of sample volume) overnight at 4C under gentle stirring using a 10 kDa Slide-a-Lyzer G2 Dialysis Cassette (ThermoFisher Scientific). The resultant mixture is then filtered using a 0.2-mm sterile filter. The filtrate is stored at 2C–8C. Multi-dose formulations may be formulated using 25 mM citrate, 100 mM NaCl cargo buffer (pH 5), and buffer exchanged by TFF into tris-saline sucrose buffer (TSS) buffer (5% sucrose, 45 mM NaCl, and 50 mM Tris). Formulated LNPs have an average size of 105 nm. Encapsulation efficiencies are determined by ribogreen assay (Leung et al., 2012). Particle size and polydispersity are measured by dynamic light scattering (DLS) using a Malvern Zetasizer DLS instrument.

NSG-PiZ mice carrying the human SERPINA1 PiZ allele (E342K) are acquired from The Jackson Laboratory. To assess the ability of Gene Writing to edit the mutant allele in vivo, LNPs are dosed via the lateral tail vein at 3 mg/kg in a volume of 0.2 mL per animal. Excipient-treated animals are used as negative controls for all studies. Animals are euthanized at various time points by exsanguination via cardiac puncture under isoflurane anesthesia. In some embodiments, animals are euthanized at one week post-treatment to be analyzed for Gene Writing. Liver tissue is collected from the median or left lateral lobe from each animal for DNA extraction and analysis.

For NGS analysis of editing efficiency, PCR primers are designed around the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the appropriate chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that overlap the target region of interest are selected, and the number of wild-type reads versus the number of reads that contain the SERPINA1 reversion mutation encoded in the template RNA are calculated. The editing percentage (e.g., the ''editing efficiency'' or ''percent editing'') is defined as the total number of reversion sequence reads over the total number of sequence reads.

In some embodiments, this example is repeated with additional groups of mice and a redosing regimen is used to analyze dose-to-effect properties of the system. In these experiments, mice are assigned to groups for weekly dosing up to 4 weeks, with euthanasia and tissue analysis as described herein being performed each week. In some embodiments, mice that receive more doses of the LNP formulation demonstrate higher Gene Writing efficiency by sequencing, e.g., mice receiving 2 doses one week apart that are analyzed at week three show a higher fraction of gene corrected reads by NGS of liver tissue samples as compared to mice receiving a single dose and analyzed at week three. In application, dosing in this manner may allow tuning of therapeutic intervention after evaluating patient response to one or more doses.

Example 25: Using Gene Writing to address repeat expansion diseases

This example describes the use of a Gene Writer™ gene editing system to treat a repeat expansion disease by rewriting a normal number of repeats into the locus. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse CNS via AAV to reset the CAG repeats in HTT as per the custom template RNA to cure Huntington Disease. Healthy humans tend to carry between 10 and 35 CAG repeats within the huntingtin gene (*HTT*), while those with Huntington Disease may possess between 36 to greater than 120 repeats.

In this example, the template RNA is designed to correct the CAG repeat region of the HTT gene by encoding a sequence with 10 such repeats and homology to the flanking target sequence to fully write across the target locus. Multiple examples of such template RNAs could

be designed, with an exemplary template RNA, as encoded in DNA, comprising the sequence (1)GGCGGCTGAGGAAGCTGAGG(2)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)AGTCCCTCAAG TCCTTCcagcagcagcagcagcagcagcagcagccgccaccgccgccgccgccgccgcctcct(4)CAGCTTCC TCAG(SEQ ID NO: 1690), where numbers are used to delineate the modules of the template in the order (5'-3') (1) gRNA spacer, (2) gRNA scaffold, (3) heterologous object sequence, (4) 3' homology priming domain, with the repeat correction being encoded in (3). The CAG repeat region is followed by a short repeat region encoding for 11 proline residues (8 residues being encoded by CCG triplets). Without wishing to be bound by theory, this region is included in (3) to place (4) in a more unique region to prevent mispriming. An exemplary gRNA for providing a second nick as described in embodiments of this system comprises the spacer sequence CGCTGCACCGACCGTGAGTT (SEQ ID NO: 1647) and directs a Cas9 nickase to nick the second strand of the target site within the homologous region. In some embodiments, this second nick improves the efficiency of the edit.

In order to deliver a complete Gene Writing system to the CNS, in this example, the Gene Writer is split across two AAV genomes, with the first encoding the nickase Cas9 domain fused to intein-N of a split intein pair (DnaE Intein-N: CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN(SEQ ID NO: 1613)) and the second encoding the RT domain fused to an intein-C of a split intein pair (DnaE Intein-C, MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN(SEQ ID NO: 1615)) and the template RNA. The two polypeptide components are expressed from a polymerase II promoter, e.g. a neuronal cell-specific promoter described herein, and the template RNA and gRNA for providing a second nick are expressed from a polymerase III promoter, e.g. a U6 promoter. When co-infecting a cell, the two polypeptide components reconstitute a complete Gene Writer polypeptide with N-terminal Cas9 and C-terminal RT and the template RNA is expressed and reverse transcribed into the target locus. To achieve delivery for cells of the CNS (specifically the claudate nucleus and the putamen of the basal ganglia), the pseudotyped system rAAV2/1 is used here, where the AAV2 ITRs are used to package the described nucleic acids into particles with AAV1 capsid. AAV preparation and mouse injection and harvesting protocols used here follow the teachings of Monteys et al. *Mol Ther* 25(1):12-23 (2017).

FVB-Tg(YAC128)53Hay/J mice are acquired from The Jackson Laboratory. These transgenic mice express the full-length human huntingtin protein with ~118 glutamine repeats (CAG trinucleotide repeats) and develop hyperkinesis at three months of age. At 8 weeks of age, mice are treated with a combination 1:1 of rAAV2/1-Cas9 virus and rAAV-MMLV_RT/hU6templateRNA virus. For rAAV injections, mice are anesthetized with isoflurane and 5 μL of rAAV mixture injected unilaterally into the right striata at 0.2 μL/min. After three weeks, mice are sacrificed and brain tissue taken for genomic DNA extraction and NGS analysis.

For NGS analysis of editing efficiency, PCR primers are designed flanking the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the necessary chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that overlap the target region of interest are selected, and the number of diseased allele (>35 CAG repeats) reads versus the number of repaired allele (10-35 CAG repeats) reads are calculated. The editing percentage (e.g., the ''editing efficiency'' or ''percent editing'') is defined as the total number of repaired reads, as defined above, over the total number of sequence reads.

Example 26: Delivery of a Gene Writing system by LNP and AAV vehicles

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide in vivo. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse liver cells via a combination of lipid nanoparticles (mRNA encoding polypeptide) and AAV (DNA encoding the RNA template) to correct the SERPINA1 PiZ mutation causing alpha-1 antitrypsin deficiency.

Capped and tailed mRNA encoding the Gene Writer polypeptide are prepared by in vitro transcription and formulated into LNP-INT01 as described in Example 23, but without template RNA co-formulation.

In this example, the template RNA is encoded as DNA and delivered via AAV. The teachings of Cunningham et al. *Mol Ther* 16(6):1081-1088 (2008) describe the use of rAAV2/8 with the human alpha-1 antitrypsin (hAAT) promoter and two copies of the hepatic control

region of the apolipoprotein E enhancer (ApoE) to effectively transduce and drive expression of cargo in juvenile mouse liver. Accordingly, rAAV2/8.ApoE-hAAT.PiZ (rAAV2/8.PiZ) as described here comprises the above described AAV and promoter system driving expression of an RNA template for correcting the PiZ mutation, in addition to a second nick-directing gRNA being driven by a U6 promoter (RNA sequences previously described in Example 1).

NGS-PiZ mice carrying the human SERPINA1 PiZ allele (E342K) are acquired from The Jackson Laboratory. To assess the activity of Gene Writing to edit the mutant allele in vivo, 8-week-old mice are dosed i.p. with ~10^{11} vg of rAAV2/8.PiZ to express the template RNA and via the lateral tail vein with formulated LNPs at 3 mg/kg in a volume of 0.2 mL per animal to express the Gene Writer polypeptide. Animals are euthanized at various time points by exsanguination via cardiac puncture under isoflurane anesthesia. In some embodiments, animals are euthanized at one week post-treatment to be analyzed for Gene Writing. Liver tissue is collected from the median or left lateral lobe from each animal for DNA extraction and analysis.

For NGS analysis of editing efficiency, PCR primers are designed around the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the necessary chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that overlap the target region of interest are selected, and the number of wild-type reads versus the number of reads that contain the SERPINA1 reversion mutation encoded in the template RNA are calculated. The editing percentage is defined as the total number of reversion sequence reads over the total number of sequence reads.

Example 27: Application of a Gene Writer™ system for delivering therapeutic gene to liver in a human chimeric liver mouse model

This example describes a Gene Writer™ genome editing system delivered to the liver in vivo for integration and stable expression of a genetic payload. Specifically, LNPs are used to deliver a Gene Writing system capable of integrating a complete OTC expression cassette to treat a humanized mouse model of OTC-deficiency.

In this example, a Gene Writing system is used to treat a humanized mouse model of OTC deficiency, in which human hepatocytes derived from patients with OTC deficiency are engrafted into a mouse model (Ginn et al JHEP Reports 2019). An exemplary Gene Writing system for large payload integration comprises a Cas9-directed reverse transcriptase system utilizing a highly processive reverse transcriptase, e.g., MarathonRT. An exemplary template RNA component comprises, from 5' to 3', (1) a gRNA spacer with homology to the AAVS1 safe harbor site, (2) a gRNA scaffold, (3) a heterologous object sequence, and (4) a 3' target homology region for annealing to the genomic DNA immediately upstream of the first strand nick to prime TPRT of the heterologous object sequence. An exemplary sequence for (1) is GGGGCCACTAGGGACAGGAT(SEQ ID NO: 1691). Region (2) carries the gRNA scaffold as described in this application, generally comprising the sequence GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGGACCGAGTCGGTCC (SEQ ID NO: 1603). In this example, (3) comprises a complete OTC expression cassette, where a liver-codon-optimized sequence encoding human OTC (UniProt P00480) is in operable association with the ApoE.hAAT promoter system as described in Example 25. An exemplary sequence for (4) is CTGTCCCTAGTG (SEQ ID NO: 1692). An exemplary sequence of an additional gRNA spacer for generating a second strand nick to improve the efficiency of integration is AGAGAGATGGCTCCAGGAAA(SEQ ID NO: 1693).

Eight to 12-week-old female $Fah^{-/-}Rag2^{-/-}Il2rg^{-/-}$ (FRG) mice are engrafted with human hepatocytes, isolated from pediatric donors or purchased from Lonza (Basel, Switzerland), as described previously (Azuma et al Nat Biotechnol 2007). Engrafted mice are cycled on and off 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC) in drinking water to promote liver repopulation. Blood is collected every two weeks and at the end of the experiment to measure the levels of human albumin, used as a marker to estimate the level of engraftment, in serum by enzyme-linked immunosorbent assay (ELISA; Bethyl Laboratories, Inc., Montgomery, TX). Eleven weeks after engraftment, mice are treated with the Gene Writer™s formulated as in Example 23. For treatment, LNPs are delivered via the lateral tail vein at 3 mg/kg in a volume of 0.2 mL per animal.

After vector injection, mice are cycled on NTBC for another 5 weeks before being euthanized. DNA and RNA are subsequently extracted from liver lysates by standard methods. OTC expression is subsequently assayed by performing RT-qPCR on isolated RNA samples

using sequence-specific primers. Levels of human OTC are also measured throughout the experiment by using a human OTC ELISA kit (e.g., Aviva Systems Biology OTC ELISA Kit (Human) (OKCD07437)) on serum at Days -7, 0, 2, 4, 7, 14, 21, 28, and 35 post-injection, following the manufacturer's recommended protocol.

For analysis of editing efficiency, a ddPCR assay is performed using a pair of primers that anneal across either the 5' junction or the 3' junction of integration, with one primer in each set annealing to the heterologous object sequence, and the other to an appropriate region of the AAVS1 site on the genome. The assay is normalized to a reference gene to quantify the number of target site integrations per genome.

To analyze integrations at the target site, long-read sequencing across the integration site is performed. PCR primers are designed flanking the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (PacBio) to add the necessary chemistry for sequencing, and amplicons are then sequenced via PacBio. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that contain an insertion sequence relative to the reference genome are selected for further analysis to determine completeness of integration, defined in this example as containing the complete promoter and coding sequence of OTC.

Example 28: Gene Writers for integration of a CAR in T-cells ex vivo

This example describes delivery of a Gene Writer™ genome editing system to T-cells ex vivo for integration and stable expression of a genetic payload. Specifically, LNPs are used to deliver a Gene Writing system capable of integrating a chimeric antigen receptor (CAR) into the *TRAC* locus to generate CAR-T cells for treating B-cell lymphoma.

In this example, a Gene Writing system comprises a Gene Writing polypeptide, e.g., a nickase Cas9 and R2Tg reverse transcriptase domain, as described herein, a gRNA for directing nickase activity to the target locus, and a template RNA comprising, from 5' to 3':

(1) 100 nt homology to target site 3' of first strand nick

(2) 5' UTR from R2Tg

(3) Heterologous object sequence

(4) 3' UTR from R2Tg

(5) 100 nt homology to target site 5' of first strand nick

Wherein (3) comprises the coding sequence for the CD19-specific Hu19-CD828Z (Genbank MN698642; Brudno et al. *Nat Med* 26:270-280 (2020)) CAR molecule. The Gene Writer in this example is guided to the 5' end of the first exon of *TRAC* by using a targeted gRNA, e.g., TCAGGGTTCTGGATATCTGT(SEQ ID NO: 1694), in order to place the cargo under endogenous expression control from that locus while disrupting the endogenous TCR, as taught by Eyquem et al. *Nature* 543:113-117 (2017). These three components (polypeptide, gRNA, and template) all comprise RNA, which is synthesized by in vitro transcription (e.g., polypeptide mRNA, template RNA) or chemical synthesis (gRNA).

The LNP formulation used in this example has been screened and validated for delivery to T-cells ex vivo, being taught in Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020), which is incorporated herein by reference in its entirety. Specifically, the LNP formulation C14-4, comprising cholesterol, phospholipid, lipid-anchored PEG, and the ionizable lipid C14-4 (Figure 2C of Billingsley et al. Nano Lett 20(3):1578-1589 (2020)) was used to encapsulate all three RNA components in a molar ratio of polypeptide mRNA:gRNA:template RNA of about 1:40:40.

Additional edits can be performed on T-cells in order to improve activity of the CAR-T cells against their cognate target. In some embodiments, a second LNP formulation of C14-4 as described comprises a Cas9/gRNA preformed RNP complex, wherein the gRNA targets the *Pdcd1* exon 1 for PD-1 inactivation, which can enhance anti-tumor activity of CAR-T cells by disruption of this inhibitory checkpoint that can otherwise trigger suppression of the cells (see Rupp et al. *Sci Rep* 7:737 (2017)). The application of both nanoparticle formulation thus enables lymphoma targeting by providing the anti-CD19 cargo, while simultaneously boosting efficacy by knocking out the PD-1 checkpoint inhibitor. In some embodiments, cells may be treated with the nanoparticles simultaneously. In some embodiments, the cells may be treated with the nanoparticles in separate steps, e.g., first deliver the RNP for generating the PD-1 knockout, and subsequently treat cells with the nanoparticles carrying the anti-CD19 CAR. In some embodiments, the second component of the system that improves T cell efficacy may result in the knockout of PD-1, TCR, CTLA-4, HLA-I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS, PDC1, CISH, TRAC, or a combination thereof. In some embodiments, knockdown of PD-1, TCR, CTLA-4, HLA-I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS,

PDC1, CISH, or TRAC may be preferred, e.g., using siRNA targeting PD-1. In some embodiments, siRNA targeting PD-1 may be achieved using self-delivering RNAi as described by Ligtenberg et al. *Mol Ther* 26(6):1482-1493 (2018) and in WO2010033247, incorporated herein by reference in its entirety, in which extensive chemical modifications of siRNAs, conferring the resulting hydrophobically modified siRNA molecules the ability to penetrate all cell types *ex vivo* and *in vivo* and achieve long-lasting specific target gene knockdown without any additional delivery formulations or techniques. In some embodiments, one or more components of the system may be delivered by other methods, e.g., electroporation. In some embodiments, additional regulators are knocked in to the cells for overexpression to control T cell- and NK cell-mediated immune responses and macrophage engulfment, e.g., PD-L1, HLA-G, CD47 (Han et al. PNAS 116(21):10441-10446 (2019)). Knock-in may be accomplished through application of an additional Gene Writing system with a template carrying an expression cassette for one or more such factors (3) with targeting to a safe harbor locus, e.g., AAVS1, e.g., using gRNA GGGGCCACTAGGGACAGGAT to target the Gene Writer polypeptide to AAVS1.

LNPs are used to treat primary T cells activated by Dynabeads at a 1:1 $CD4^+$:$CD8^+$ ratio at 450 ng/μL total mRNA concentrations. The resulting T cell populations are analyzed for integration, expression, and effect. For assessing integration, ddPCR is used with primers producing an amplicon extending from within the integrated CAR to the flanking genomic TRAC sequence. Comparing signal to a reference gene (e.g., RPP30), allows quantification of the average copy number per genome and integration efficiency. To analyze expression, flow cytometry with immunological probes is used to assess the level and percent of cells displaying surface CAR expression. To analyze activity of the CAR-T cells, treated cells are assessed via a co-plated cancer cell killing assay. By engineering Nalm6 ALL cells to express luciferase, cancer cell killing can be assessed by change in luminescence after co-culture with CAR-T cells as compared to signal from Nalm6 cells alone Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020). Thus, a Gene Writing system can be used to generate CAR-T cells ex vivo with the desired cytotoxic activity.

Example 29: Gene Writers for integration of a CAR in T-cells in vivo

This example describes a Gene Writer™ genome editing system delivered T-cells in vivo for integration and stable expression of a genetic payload. Specifically, targeted nanoparticles are used to deliver a Gene Writing system capable of integrating a chimeric antigen receptor (CAR) expression cassette into the murine Rosa26 locus to generate CAR-T cells in a murine model.

In this example, a Gene Writing system comprises a Gene Writing polypeptide, e.g., a nickase Cas9 and R2Tg reverse transcriptase domain, as described herein, a gRNA for directing nickase activity to the target locus, and a template RNA comprising, from 5' to 3':

(1) 100 nt homology to target site 3' of first strand nick

(2) 5' UTR from R2Tg

(3) Heterologous object sequence

(4) 3' UTR from R2Tg

(5) 100 nt homology to target site 5' of first strand nick

Wherein (3) comprises the coding sequence for the CD19-specific m194-1BBz CAR driven by the EF1a promoter (Smith et al. *Nat Nanotechnol* 12(8):813-820 (2017)). The Gene Writer in this example is guided to the murine Rosa26 locus using a gRNA, e.g., ACTCCAGTCTTTCTAGAAGA(SEQ ID NO: 1695), (Chu et al. *Nat Biotechnol* 33(5):543-548 (2015)). Production of RNA molecules is as according to examples provided herein, e.g., by in vitro transcription (e.g., Gene Writer polypeptide mRNA, template RNA) and by chemical synthesis (e.g., gRNA). Modifications to the RNA components of the system are as described elsewhere. For Gene Writer mRNA, the sequence additionally comprises a 5' UTR (e.g., GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC(SEQ ID NO: 1604)) and a 3' UTR (e.g., UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA(SEQ ID NO: 1605)) flanking the coding sequence. This combination of 5' UTR and 3' UTR has been shown to result in good expression of an operably linked ORF (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

In order to achieve delivery specifically to T-cells, targeted LNPs (tLNPs) are generated that carry a conjugated mAb against CD4. See, e.g., Ramishetti et al. ACS Nano 9(7):6706-6716 (2015). Alternatively, conjugating a mAb against CD3 can be used to target both $CD4^+$ and $CD8^+$ T-cells (Smith et al. Nat Nanotechnol 12(8):813-820 (2017)). In other embodiments, the

nanoparticle used to deliver to T-cells in vivo is a constrained nanoparticle that lacks a targeting ligand, as taught by Lokugamage et al. *Adv Mater* 31(41):e1902251 (2019).

The tLNP can be made by first preparing the nucleic acid mix (e.g., polypeptide mRNA:gRNA:template RNA molar ratio of 1:40:40) with a mixture of lipids (cholesterol, DSPC, PEG-DMG, Dlin-MC3-DMA, and DSPE-PEG-maleimide) and then chemically conjugating the desired DTT-reduced mAb (e.g., anti-CD4, e.g., clone YTS.177) to the maleimide functional group on the LNPs. See Ramishetti et al. ACS Nano 9(7):6706-6716 (2015).

Six to 8 week old C57BL6/J mice are injected intravenously with formulated LNP at a dose of 1 mg RNA/kg body weight. Blood is collected at one day and three days post-administration in heparin-coated collection tubes, and the leukocytes are isolated by density centrifugation using Ficoll-Paque PLUS (GE Healthcare). Five days post-administration, animals are euthanized and blood and organs (spleen, lymph nodes, bone marrow cells) are harvested for T-cell analysis. Expression of the anti-CD19 CAR is detected by FACS using specific immunological sorting. Positive cells are confirmed for integration by ddPCR on the sorted population, where primers are used that flank an integration junction, e.g., one primer of the pair annealing to the integrated cargo and the other to genomic DNA from the Rosa26 target site.

Example 30: Mutation of DNA binding motifs of a retrotransposase prevents integration at its native site

In this example, the intrinsic DNA binding properties of the DNA binding domain of a retrotransposon-based Gene Writer polypeptide are analyzed by mutagenesis or truncation of the region. The generation of these mutants and detection of reduced activity are useful for understanding the determinants of DNA targeting, as well as for creating mutant derivatives that have reduced or eliminated function that can then be used as scaffolds for fusions enabled by heterologous DNA binding domains. Here, it is shown that mutations in the C-terminal zinc finger domains, as well as the c-myb domain, have a profound impact on enzyme activity of the R2Tg retrotransposase at its native recognition sequence.

To test whether R2Tg DNA binding mutants resulted in abrogated DNA binding activity, exemplary mutants were constructed and assessed for integration activity at the native rDNA

target site as a downstream readout of target DNA binding activity. In this example, the parental and mutated (*) domains were named and designed as follows:

- ZF1: CPCCGTRVNSVLNLIEHLKVSH
- ZF1*: SPSSGTRVNSVLNLIEHLKVSH
- ZF2: CEVCNRDFTTKIGLGQHKRLAH
- ZF2*: SEVSNRDFTTKIGLGQHKRLAH
- c-myb: RCWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISD
- c-myb*: ACATKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISD

The general structure of these domains is indicated in Figure 30 A. Here, the mutant Gene Writer polypeptides were encoded within the Gene Writer template, meaning the polypeptide coding sequence was further flanked by the R2Tg-derived 5' UTR and 3' UTR and 100 nt of homology to the native target site, as described in the invention.

To test if mutations in the R2Tg DNA binding domains impacted R2Tg Gene Writer function, 250 ng of plasmid comprising the parental R2Tg polypeptide, an endonuclease-inactivated mutant (negative control for Gene Writer activity), each single DBD mutation, or all three DBD mutations was nucleofected into HEK293T cells using the Lonza Amaxa Nucleofector 96 Well Shuttle System with program DS150, as according to manufacturer's protocols. After nucleofection, cells were grown at 37°C, 5% CO_2 for 3 days prior to cell lysis and genomic DNA extraction. The extracted gDNA was measured for Gene Writer Template integration at the native rDNA site by ddPCR. Here, it was observed that mutations in the ZF1, ZF2, or c-myb domains inhibited integration, with the magnitude of effect being ZF2>c-myb>ZF1 (Figure 30B).

Example 31: Determination of cleavage site of endonuclease domain of Gene Writer by indel signature

This example describes a cell-based assay for determining the site of cleavage activity of a Gene Writer polypeptide comprising an endonuclease (EN) domain on a genomic DNA target. Specifically, it is shown that the activity of a retrotransposase can produce a low level of target site modification likely resulting from host DNA repair of target nicks. By using an extremely sensitive and targeted amplicon sequencing assay, this signature can be assessed to localize the site of cleavage. The ability to quickly assess an early step in relevant Gene Writing systems is

an enabling assay for understanding and engineering the DNA specificity of the enzyme, e.g., sequence-specific DNA binding or sequence-specific endonuclease activity, without depending on complete integration, which may be affected by other properties of the system.

In order to generate a cleavage site profile, an assay was created to analyze genomic sequence modifications at the predicted cleavage site of the retrotransposase R2Tg. A schematic of the target sequence of R2Tg depicting predicted DNA binding regions and the expected site of endonuclease cleavage is shown in Figure 31 A. To determine whether R2Tg endonuclease activity can be detected around the predicted target site, 73 ng of an R2Tg expression plasmid were nucleofected into 200,000 U2OS cells using program DN100 and buffer SE of the Amaxa Lonza nucleofector, as according to manufacturer's instructions. Cells were cultured in DMEM with 10% FBS for three days post nucleofection before extraction of genomic DNA. Amplicons were generated using primers flanking the rDNA target site (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTaggggaatccgactgtttaatta-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcacctctcatgtctcttcaccg-3'). Amplicon sequencing was performed using an Illumina MiSeq and results were analyzed using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)) to determine mutational signatures in the target site. Insertions and deletions were found at and surrounding the predicted GG cleavage site, with the peak of the insertion signature occurring directly at the predicted R2Tg cleavage site (Figure 31 B). These data validated the use of the indel signature assay for the detection and localization of endonuclease activity in a Gene Writing system.

Example 32: Refining the sequence specificity of a retrotransposase endonuclease domain

This example describes an experiment for elucidating the sequence specificity of endonuclease activity associated with a Gene Writer system. Specifically, DNA target sequence information important for endonuclease activity is elucidated through the targeted mutagenesis of regions of various sizes at or surrounding a site of cleavage. Here, the sequence specificity of the native endonuclease domain of the R2Tg retrotransposase is profiled by applying an indel signature assay (see Example 31) on cells comprising genomic landing pads with native or altered R2Tg target sequence to determine the cleavage activity of this library of targets.

In this example, cell lines were generated with stable integrations of landing pads comprising rDNA-derived sequence corresponding to the native target site where R2 class

retrotransposons systems facilitate retrotransposition. These landing pads were designed to have either (1) a wild-type target sequence comprising rDNA-derived sequence from the R2 region of the rDNA; (2) the landing pad from (1) with 12-bp of sequence mutation at and around the R2 cleavage site; or (3) a series of mutations within the 12-bp sequence from (2) to further define the minimal sequence requirements within the 12 bp context described in Example 2 . A complete list of landing pads can be found in Table LandingPads and is shown in Figure 32 B. To create these cells lines, the DNA for different landing pads was synthesized and cloned into a lentiviral gene expression vector downstream of a GFP reporter (Figure 32 A). Landing pad lentiviral vectors were verified by Sanger sequencing of the landing pad. To produce lentiviral particles for transduction, 9 μg of the sequence verified plasmids and 9 μg of lentiviral packaging mix (Biosettia) were transfected into the lentivirus packaging cell line LentiX-293T (Takara Bio). Transfected cells were incubated at 37˚C, 5% CO2 for 48 hours (including one medium change at 24hrs) and viral particle-containing medium was collected. The collected medium was filtered through a 0.2 μm filter to remove cell debris and prepared for transduction of U2OS cells. Virus-containing filtrate was diluted in DMEM and mixed with polybrene to prepare a dilution series for cell transduction, with a final concentration of 8 μg/mL polybrene. Recipient U2OS cells were grown in virus-containing medium for 48 hour and then split with fresh medium and grown to confluence. Transduction efficiency of the different dilutions of virus was measured by GFP expression via flow cytometry and ddPCR to determine average copy number of integrated lentiviral landing pads.

The ability of the R2Tg retrotransposase to exhibit endonuclease activity towards a target site with various mutations in and around the R2 cleavage position was assayed using indel profiling (Example 31). Specifically, 73 ng of a Gene Writer polypeptide expression vector comprising the R2Tg retrotransposase were nucleofected into the different U2OS landing pad cell lines using the Lonza Amaxa Nucleofector 96 Well Shuttle System with nucleofection program DN100, as according to manufacturer's instructions. Following nucleofection, cells were grown at 37˚C, 5% CO2 for 3 days prior to cell lysis and genomic DNA extraction. Primers specific to the landing pad sequences (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctcacacaggaaacagctatg-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTggatgtgctgcaaggcgatt-3') were used to amplify the region and purified amplicons were sequenced on an Illumina MiSeq. Signatures of

endonuclease activity at the target site were analyzed by detection of insertions and deletions at the landing pad using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)). As was shown in Example 31, R2Tg endonuclease activity occurs at a GG cleavage site (Figures31 A and B). Using this series of landing pads, the minimum sequence important for endonuclease activity for the tested polypeptide was found to include this GG dinucleotide and an additional AA dinucleotide located immediately upstream, defining a 5'-AAGG-3' motif important for endonuclease activity within the native target sequence (Figure 32 B). In some embodiments, it may be desirable to find a native endonuclease-specific motif at a site in the human genome for retargeting a retrotransposase-based Gene Writer system. In some embodiments, a naturally occurring AAGG sequence in the genome is used as a seed for retargeting an R2 retrotransposase-based Gene Writing system, wherein the DNA binding domain is mutated or replaced with a heterologous DNA binding domain such that the binding of the Gene Writer polypeptide to the new target site results in the proper positioning of the endonuclease domain to the AAGG motif to enable endonuclease activity.

Table 41: Landing pads comprising wild-type or mutated target sequences, as described in Example 32. Sequences were incorporated into the genome using a lentiviral vector system for stable integration. Sequences in the table are provided in 5' to 3' orientation, where underlined text indicates native rDNA sequence and bold text indicates mutations from the native rDNA sequence.

Table 41: Landing pads comprising wild-type or mutated target sequences

Landing Pad	Corresponding Target Sequence	SEQ ID NO.
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**AAGGTAGCCAAA**TGCCTCGTCATCTAATTAG TGACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTAC CTACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCC GTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACC CAACTTAATCGCCTTGCAGCACATCC	1986

12-bp Mutant	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCCAATATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1987
10-bp (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCCAATATGA**AATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1988
6-bp (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCCAAT**GCCAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1989
4-bp (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCCA**TAGCCAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1990
2-bp (-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TC**GGTAGCCAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1991

12-bp (GG)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCGGATATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1992
10-bp (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCGGATATGA**AATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1993
8-bp (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCGGATAT**CAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1994
6-bp (GG-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TCGGAT**GCCAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1995
12-bp (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TAGGTTATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1996

10-bp (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TAGGTTATGA**AATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1997
8-bp (AGGT-)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTT**TAGGTTAT**CAAATGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1998
10-bp (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTTAA**CAATATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	1999
8-bp (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTTAAGG**ATATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	2000
6-bp (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTTAAGGTA**ATGATT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	2001

2-bp (+)	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTG TTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGCGGGTAAACGGCGGGAGTAACT ATGACTCTCTTAAGGTAGCCA**TT**TGCCTCGTCATCTAATTAGT GACGCGCATGAATGGATGAACGAGATTCCCACTGTCCCTACC TACTATCCAGCGAAACCACAGCCAAGGGAAATTCACTGGCCG TCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTGCAGCACATCC	2002

Example 33: Determination of sequence specificity for retargeting Gene Writer polypeptides

This example describes the redirecting of the polypeptide component of a Gene Writer system from its native recognition sequence to a new location in the human genome. As described in this disclosure, a retrotransposase-based system can be directed to recognize a new DNA sequence through the addition of a heterologous DNA binding domain alone or in tandem with mutagenesis of an endogenous DNA-binding domain. Here, a zinc finger capable of targeting the native human AAVS1 site is fused to a Gene Writer polypeptide in order to confer recognition of the AAVS1 sequence.

In order to direct a Gene Writer polypeptide comprising the R2Tg retrotransposase to a DNA sequence different from its native rDNA target sequence, a zinc finger (ZF) domain targeting the human AAVS1 site was fused to the N-terminus of the polypeptide. Using an approach similar to Example 32, cells were generated comprising various compositions of genomic landing pads, representing different combinations of the AAVS1 sequence recognized by the ZF and the native rDNA target sequence recognized by R2 retrotransposases. In total, the library consisted of 460 distinct landing pads containing the AAVS1 sequence along with different lengths of the rDNA sequence, further diversified by varying the distance and orientation between the two sequences (Figure 33). In this example, all landing pads were designed to include a human AAVS1 genomic sequence comprising the AAVS1 ZF binding site. Additionally, rDNA sequence containing the minimum AAGG sequence important for R2Tg endonuclease activity (see Example 32) was added to the AAVS1 sequence according to the following parameters: (1) the rDNA sequence included the 12, 22, 32, 42, 52, 62, or 72 nt of rDNA sequence located immediately 3' of the AAGG tetranucleotide; or (2) the rDNA sequence included the 12, 22, 32, or 42 nt immediately 3' and 5' of the AAGG cleavage site, resulting in total rDNA lengths of 24, 44, 64, or 84 nt. The different rDNA sequence compositions were further placed at various distances from the AAVS1 ZF binding site, including at a distance of 5,

10, 15, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, or 85 nt. Design considerations were included to hold the total length of the landing pad site between assay primers constant to prevent bias during PCR amplification, e.g., longer rDNA sequences could not be placed at the same distance from the AAVS1 site as shorter rDNA sequences. As a final variation, the orientation between the two sites was varied, such that the rDNA sequence compositions described above were placed either upstream or downstream of the AAVS1 site in either the forward or reverse orientations. Using these parameters, a library of 454 AAVS1-rDNA hybrid landing pads was designed, with controls including various combinations of full-length rDNA sequence and AAVS1 sequence (positive) or no rDNA sequence (negative). An illustrative representation of the landing pad design strategy is shown in Figure 33 and a short list of exemplary landing pad sequences to demonstrate the specific compositions is provided in Table 42.

The lentiviral construct library described above was synthesized with 3' barcodes for sequence analysis and cloned into a lentiviral gene expression vector. The lentiviral system was then used to generate U2OS cell lines with integrated landing pads, as described in Example 32. To validate successful library generation, the U2OS landing pad cell line pool was analyzed for landing pad representation. Primers specific to conserved landing pad sequences (see Example 32) were used to amplify across the target region, including a construct-specific barcode. Barcodes present in each landing pad were computationally demultiplexed and sorted, with approximately 94% of landing pads being represented by at least 10,000 reads (Figure 34).

Following validation of the library, U2OS landing pad cells were used to determine the minimum sequence deterinants for retargeting an R2Tg retrotransposase-based Gene Writer polypeptide. Two retargeting constructs were generated by fusion of the coding sequence of ZF-AAVS1 to either full-length R2Tg (SEQ ID NO: 1672) (ZF-R2Tg, Figure 35 A) or to a DNA-binding domain truncated R2Tg SEQ ID NO: 1663) (ZF-R2Tg(noDBD), Figure 36 A) and the corresponding expression plasmids were electroporated into U2OS pooled landing pad cells. Specifically, 400 ng of either ZF-R2Tg or ZF-R2Tg(noDBD) were delivered to cells using the Lonza Amaxa Nucleofector 96 Well Shuttle System with nucleofection program DN100, as according to manufacturer's instructions. Following nucleofection, cells were grown at 37°C, 5% CO_2 for 3 days prior to cell lysis and genomic DNA extraction. As with the library validation above, primers specific to the landing pad sequences were used to amplify the target region for

amplicon sequencing with an Illumina MiSeq. Sequencing reads to landing pad variants were first de-multiplexed using the associated barcodes, then signatures of endonuclease activity at the target site were analyzed by detection of insertions and deletions using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)).

Given the ZF-R2Tg construct comprises a full-length R2Tg protein, it was predicted that this polypeptide would still retain the ability to recognize its native target sequence. Insertion frequencies at the GG target site were computed and plotted for each landing pad (Figure 35 A). Positive control landing pads containing 200 nt of rDNA sequence were found to contain insertion signatures at the GG cleavage site, as in Example 31. The negative control landing pad, devoid of rDNA sequence, was not found to harbor any insertions. Here, insertion signatures resulting from ZF-R2Tg endonuclease activity at the target cleavage site were detected in landing pads containing 44, 64, and 84 nt of the native rDNA target sequence, but not in those containing only 24 nt of the native rDNA target sequence (Figure 35 A and B). These rDNA sequence lengths that tested positive were shown to be positive across landing pads with different distances from the AAVS1 sequence.

Next, the ZF-R2Tg(noDBD) construct, lacking its endogenous DNA binding domain and predicted to depend on the ZF for target binding and activity, was similarly assessed. Insertion frequencies at the GG target site were computed and plotted for each landing pad (Figure 36 A). The negative control landing pad, devoid of rDNA sequence, was not found to harbor any insertions. Given the R2Tg protein in this construct comprised a significant deletion, there was no positive control landing pad configuration for verifying activity. In this experiment, two different landing pad configurations showed indel signatures at the GG target site. Both hits comprised the same 44 nt of rDNA sequence, but were differentially positioned relative to the AAVS1 site, where one had the rDNA target at a distance of 55 nt upstream of the AAVS1 site and the other was at a distance of 20 nt downstream of the AAVS1 site and in the reverse complement orientation (Figures 36 A and B). Despite having different compositions, these two hits indicated the restoration of activity of an R2 retrotransposase DNA binding domain mutant by compensating for the deleted endogenous domain with a heterologous DNA binding domain fusion known to target a native locus in the human genome. Additionally, this Example establishes a method for further refinement of the requirements for retargeting a Gene Writer polypeptide to an alternate sequence in the human genome.

Table 42 provides a selection of exemplary landing pad target sequences designed to test Gene Writer polypeptides comprising AAVS1 zinc finger fusions to the R2Tg retrotransposase is shown here, e.g., in this Example. Included are 1) a positive control comprising the AAVS1 zinc finger recognition sequence and a full 200 nt region from the rDNA centered around the R2 cleavage site; 2) a negative control lacking rDNA sequence; 3) a 44 nt rDNA sequence located upstream of the AAVS1 sequence; and 4) a 44 nt rDNA sequence located downstream of the AAVS1 sequence and in the reverse orientation. The two experimental landing pads included here were shown to enable cleavage of the AAGG core sequence using a zinc finger fused to a mutant R2Tg lacking its endogenous N-terminal DNA binding domain, e.g., see this Example. Ribsomal DNA sequence is indicated by underlined text, the AAGG core is indicated by brackets, and the binding sequence of the AAVS1 zinc finger is indicated by lowercase text.

Table 42. selection of exemplary landing pad target sequences designed to test Gene Writer polypeptides comprising AAVS1 zinc finger fusions to the R2Tg retrotransposase

Description	Sequence	SEQ ID NO
Positive Control R2(200)_left(frw)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC GTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATG TCAAAGTGAAGAAATTCAATGAAGCGCGGGTAAAC GGCGGGAGTAACTATGACTCTCTT[AAGG]TAGCCAA ATGCCTCGTCATCTAATTAGTGACGCGCATGAATGG ATGAACGAGATTCCCACTGTCCCTACCTACTATCCA GCGAAACCACAGCCAAGGGAAGAGTGGCCccactgtggg gtGGAGGGGACGACTGGGAAAAGTTAGGATCCCCTG GCGTTACCCAACTTAATCGCCTTGCAGCACATCC	2003
Negative Control AAVS1_noR2	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCTGCCTAACAGGAGGTGGGGGTTAGACCCA ATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGAT GGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCCc cactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CAGAGCCACATTAACCGGCCCTGGGAATATAAGGTG GTCCCAGCTCGGGGACACAGGATCCCTGGAGGCAGC AAACATGCTGACTGGGAAAATCTGGCTTCCCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC	2004

AAVS1_R2(44nts) at -55(frw)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCGGGAGTAACTATGACTCTCTT[AAGG]TAG CCAAATGCCTCGTCAGGAAGGAGGAGGCCTAAGGA TGGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCC ccactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CAGAGCCACATTAACCGGCCCTGGGAATATAAGGTG GTCCCAGCTCGGGGACACAGGATCCCTGGAGGCAGC AAACATGCTGACTGGGAAAACTTAGAGAACCCTGGC GTTACCCAACTTAATCGCCTTGCAGCACATCC	2005
AAVS1_R2(44nts) at +20(rc)	GCTCACACAGGAAACAGCTATGACCATGATTACGCC AAGCTTCTGCCTAACAGGAGGTGGGGGTTAGACCCA ATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGAT GGGGCTTTTCTGTCACCAATCCTGTCCCTAGTGGCCc cactgtggggtGGAGGGGACAGATAAAAGTACCCAGAAC CTGACGAGGCATTTGGCTA[CCTT]AAGAGAGTCATA GTTACTCCCGCGGGGACACAGGATCCCTGGAGGCAG CAAACATGCTGACTGGGAAAACCTTCTCGGCCCTGG CGTTACCCAACTTAATCGCCTTGCAGCACATCC	2006

Example 34: Selection of lipid reagents with reduced aldehyde content

In this example, lipids are selected for downstream use in lipid nanoparticle formulations containing Gene Writing component nucleic acid(s), and lipids are selected based at least in part on having an absence or low level of contaminating aldehydes. Reactive aldehyde groups in lipid reagents may cause chemical modifications to component nucleic acid(s), e.g., RNA, e.g., template RNA, during LNP formulation. Thus, in some embodiments, the aldehyde content of lipid reagents is minimized.

Liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) can be used to separate, characterize, and quantify the aldehyde content of reagents, e.g., as described in Zurek et al. The Analyst 124(9):1291-1295 (1999), incorporated herein by reference. Here, each lipid reagent is subjected to LC-MS/MS analysis. The LC/MS-MS method first separates the lipid and one or more impurities with a C8 HPLC column and follows with the detection and structural determination of these molecules with the mass spectrometer. If an aldehyde is present in a lipid reagent, it is quantified using a staple-isotope labeled (SIL) standard that is structurally identical to the aldehyde, but is heavier due to C13 and N15 labeling. An appropriate amount of the SIL standard is spiked into the lipid reagent. The mixture is then subjected to LC-MS/MS analysis. The amount of contaminating aldehyde is determined by multiplying the amount of SIL standard

and the peak ratio (unknown/SIL). Any identified aldehyde(s) in the lipid reagents is quantified as described. In some embodiments, lipid raw materials selected for LNP formulation are not found to contain any contaminating aldehyde content above a chosen level. In some embodiments, one or more, and optionally all, lipid reagents used for formulation comprise less than 3% total aldehyde content. In some embodiments, one or more, and optionally all, lipid reagents used for formulation comprise less than 0.3% of any single aldehyde species. In some embodiments, one or more, and optionally all, lipid reagents used in formulation comprise less than 0.3% of any single aldehyde species and less than 3% total aldehyde content.

Example 35: Quantification of RNA modification caused by aldehydes during formulation

In this example, the RNA molecules are analyzed post-formulation to determine the extent of any modifications that may have happened during the formulation process, e.g., to detect chemical modifications caused by aldehyde contamination of the lipid reagents (see, e.g., Example 34).

RNA modifications can be detected by analysis of ribonucleosides, e.g., as according to the methods of Su et al. Nature Protocols 9:828-841 (2014), incorporated herein by reference in its entirety. In this process, RNA is digested to a mix of nucleosides, and then subjected to LC-MS/MS analysis. RNA post-formulation is contained in LNPs and must first be separated from lipids by coprecipitating with GlycoBlue in 80% isopropanol. After centrifugation, the pellets containing RNA are carefully transferred to a new Eppendorf tube, to which a cocktail of enzymes (benzonase, Phosphodiesterase type 1, phosphatase) is added to digest the RNA into nucleosides. The Eppendorf tube is placed on a preheated Thermomixer at 37□C for 1 hour. The resulting nucleosides mix is directly analyzed by a LC-MS/MS method that first separates nucleosides and modified nucleosides with a C18 column and then detects them with mass spectrometry.

If aldehyde(s) in lipid reagents have caused chemical modification, data analysis will associate the modified nucleoside(s) with the aldehyde(s). A modified nucleoside can be quantified using a SIL standard which is structurally identical to the native nucleoside except heavier due to C13 and N15 labeling. An appropriate amount of the SIL standard is spiked into the nucleoside digest, which is then subjected to LC-MS/MS analysis. The amount of the modified nucleoside is

obtained by multiplying the amount of SIL standard and the peak ratio (unknown/SIL). LC-MS/MS is capable of quantifying all the targeted molecules simultaneously.

In some embodiments, the use of lipid reagents with higher contaminating aldehyde content results in higher levels of RNA modification as compared to the use of higher purity lipid reagents as materials during the lipid nanoparticle formulation process. Thus, in preferred embodiments, higher purity lipid reagents are used that result in RNA modification below an acceptable level.

Example 36: Gene Writer™ enabling large insertion into genomic DNA

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by insertion of a large string of nucleotides.

In this example, the Gene Writer™ polypeptide, gRNA, and writing template are provided as DNA transfected into HEK293T cells. The Gene Writer™ polypeptide uses a Cas9 nickase for both DNA-binding and endonuclease functions. The reverse transcriptase function is derived from the highly processive RT domain of an R2 retrotransposase. The writing template is designed to have homology to the target sequence, while incorporating the genetic payload at the desired position, such that reverse transcription of the template RNA results in the generation of a new DNA strand containing the desired insertion.

To create a large insertion in the human HEK293T cell DNA, the Gene Writer™ polypeptide is used in conjunction with a specific gRNA, which targets the Cas9-containing Gene Writer™ to the target locus, and a template RNA for reverse transcription, which contains an RT-binding motif (3' UTR from an R2 element) for associating with the reverse transcriptase, a region of homology to the target site for priming reverse transcription, and a genetic payload (GFP expression unit). This complex nicks the target site and then performs TPRT on the template, initiating the reaction by using priming regions on the template that are complementary to the sequence immediately adjacent to the site of the nick and copying the GFP payload into the genomic DNA.

After transfection, cells are incubated for three days to allow for expression of the Gene Writing™ system and conversion of the genomic DNA target. After the incubation period, genomic DNA is extracted from cells. Genomic DNA is then subjected to PCR-based amplification using site-specific primers and amplicons are sequenced on an Illumina MiSeq

according to manufacturer's protocols. Sequence analysis is then performed to determine the frequency of reads containing the desired edit.

Example 37: Gene Writer™ enabling large insertion into genomic DNA

This example describes the use of a Gene Writer system in a human cell wherein the single-stranded template repair (SSTR) pathway is inhibited.

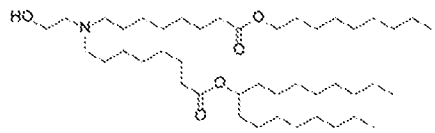
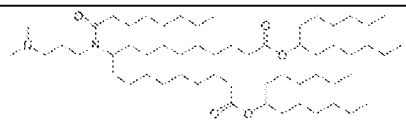
In this example, the SSTR pathway will be inhibited using siRNAs against the core components of the pathway: FANCA, FANCD2, FANCE, USP1. Control siRNAs of a non-target control will also be included. 200k U2OS cells will be nucleofected with 30pmols (1.5μM) siRNAs, as well as R2Tg driver and transgene plasmids (trans configuration). Specifically, 250 ng of Plasmids expressing R2Tg, control R2Tg with a mutation in the RT domain, or control R2Tg with an endonuclease inactivating mutation) are used in conjunction with transgene at a 1:4 molar ratio (driver to transgene). Transfections of U2OS cells is performed in SE buffer using program DN100. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3 and ddPCR is performed to assess integration at the rDNA site. Transgene integration at rDNA is detected in the absence of core SSTR pathway components.

Example 38: Formulation of Lipid Nanoparticles encapsulating Firefly Luciferase mRNA

In this example, a reporter mRNA encoding firefly luciferase was formulated into lipid nanoparticles comprising different ionizable lipids. Lipid nanoparticle (LNP) components (ionizable lipid, helper lipid, sterol, PEG) were dissolved in 100% ethanol with the lipid component. These were then prepared at molar ratios of 50:10:38.5:1.5 using ionizable lipid LIPIDV004 or LIPIDV005 (Table A1), DSPC, cholesterol, and DMG-PEG 2000, respectively. Firefly Luciferase mRNA-LNPs containing the ionizable lipid LIPIDV003 (Table A1) were prepared at a molar ratio of 45:9:44:2 using LIPIDV003, DSPC, cholesterol, and DMG-PEG 2000, respectively. Firefly luciferase mRNA used in these formulations was produced by in vitro transcription and encoded the Firefly Luciferase protein, further comprising a 5′ cap, 5′ and 3′ UTRs, and a polyA tail. The mRNA was synthesized under standard conditions for T7 RNA polymerase in vitro transcription with co-transcriptional capping, but with the nucleotide triphosphate UTP 100% substituted with N1-methyl-pseudouridine triphosphate in the reaction. Purified mRNA was dissolved in 25 mM sodium citrate, pH 4 to a concentration of 0.1 mg/mL.

Firefly Luciferase mRNA was formulated into LNPs with a lipid amine to RNA phosphate (N:P) molar ratio of 6. The LNPs were formed by microfluidic mixing of the lipid and RNA solutions using a Precision Nanosystems NanoAssemblr™ Benchtop Instrument, using the manufacturer's recommended settings. A 3:1 ratio of aqueous to organic solvent was maintained during mixing using differential flow rates. After mixing, the LNPs were collected and dialyzed in 15 mM Tris, 5% sucrose buffer at 4˚C overnight. The Firefly Luciferase mRNA-LNP formulation was concentrated by centrifugation with Amicon 10 kDa centrifugal filters (Millipore). The resulting mixture was then filtered using a 0.2 μm sterile filter. The final LNP was stored at −80˚C until further use.

Table A1: Ionizable Lipids used in Example 37

LIPID ID	**Chemical Name**	**Molecular Weight**	**Structure**
LIPIDV003	(9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9, 12-dienoate	852.29	
LIPIDV004	Heptadecan-9-yl 8-((2-hydroxyethyl)(8-(nonyloxy)-8-oxooctyl)amino)octanoate	710.18	
LIPIDV005		919.56	

Prepared LNPs were analyzed for size, uniformity, and %RNA encapsulation. The size and uniformity measurements were performed by dynamic light scattering using a Malvern Zetasizer DLS instrument (Malvern Panalytical). LNPs were diluted in PBS prior to being measured by DLS to determine the average particle size (nanometers, nm) and polydispersity index (pdi). The particle sizes of the Firefly Luciferase mRNA-LNPs are shown in **Table A2**.

Table A2: LNP particle size and uniformity

LNP ID	Ionizable Lipid	Particle Size (nm)	pdi
LNPV019-002	LIPIDV005	77	0.04
LNPV006-006	LIPIDV004	71	0.08
LNPV011-003	LIPIDV003	87	0.08

The percent encapsulation of luciferase mRNA was measured by the fluorescence-based RNA quantification assay Ribogreen (ThermoFisher Scientific). LNP samples were diluted in 1× TE buffer and mixed with the Ribogreen reagent per manufacturer's recommendations and measured on a i3 SpectraMax spectrophotomer (Molecular Devices) using 644 nm excitation and 673 nm emission wavelengths. To determine the percent encapsulation, LNPs were measured using the Ribogreen assay with intact LNPs and disrupted LNPs, where the particles were incubated with 1× TE buffer containing 0.2% (w/w) Triton-X100 to disrupt particles to allow encapsulated RNA to interact with the Ribogreen reagent. The samples were again measured on the i3 SpectraMax spectrophotometer to determine the total amount of RNA present. Total RNA was subtracted from the amount of RNA detected when the LNPs were intact to determine the fraction encapsulated. Values were multiplied by 100 to determine the percent encapsulation. The Firefly Luciferase mRNA-LNPs that were measured by Ribogreen and the percent RNA encapsulation is reported in Table A3.

Table A3: RNA encapsulation after LNP formulation

LNP ID	Ionizable Lipid	% mRNA encapsulation
LNPV019-002	LIPIDV005	98
LNPV006-006	LIPIDV004	92
LNPV011-003	LIPIDV003	97

Example 39: In vitro activity testing of mRNA-LNPs in Primary Hepatocytes

In this example, LNPs comprising the luciferase reporter mRNA were used to deliver the RNA cargo into cells in culture. Primary mouse or primary human hepatocytes were thawed and plated in collagen-coated 96-well tissue culture plates at a density of 30,000 or 50,000 cells per well, respectively. The cells were plated in 1x William's Media E with no phenol red and incubated at 37˚C with 5% CO_2. After 4 hours, the medium was replaced with maintenance medium (1x William's Media E with no phenol containing Hepatocyte Maintenance Supplement

Pack (ThermoFisher Scientific)) and cells were grown overnight at 37˚C with 5% CO_2. Firefly Luciferase mRNA-LNPs were thawed at 4˚C and gently mixed. The LNPs were diluted to the appropriate concentration in maintenance media containing 7.5% fetal bovine serum. The LNPs were incubated at 37˚C for 5 minutes prior to being added to the plated primary hepatocytes. To assess delivery of RNA cargo to cells, LNPs were incubated with primary hepatocytes for 24 hours and cells were then harvested and lysed for a Luciferase activity assay. Briefly, medium was aspirated from each well followed by a wash with 1x PBS. The PBS was aspirated from each well and 200 µL passive lysis buffer (PLB) (Promega) was added back to each well and then placed on a plate shaker for 10 minutes. The lysed cells in PLB were frozen and stored at –80˚C until luciferase activity assay was performed.

To perform the luciferase activity assay, cellular lysates in passive lysis buffer were thawed, transferred to a round bottom 96-well microtiter plate and spun down at 15,000*g* at 4˚C for 3 min to remove cellular debris. The concentration of protein was measured for each sample using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Protein concentrations were used to normalize for cell numbers and determine appropriate dilutions of lysates for the luciferase assay. The luciferase activity assay was performed in white-walled 96-well microtiter plates using the luciferase assay reagent (Promega) according to manufacturer's instructions and luminescence was measured using an i3X SpectraMax plate reader (Molecular Devices). The results of the dose-response of Firefly luciferase activity mediated by the Firefly mRNA-LNPs are shown in FIG. 37 and indicate successful LNP-mediated delivery of RNA into primary cells in culture. As shown in Fig. 37A, LNPs formulated as according to Example 38 were analyzed for delivery of cargo to primary human (A) and mouse (B) hepatocytes, as according to Example 39. The luciferase assay revealed dose-responsive luciferase activity from cell lysates, indicating successful delivery of RNA to the cells and expression of Firefly luciferase from the mRNA cargo.

Example 40: LNP-mediated delivery of RNA to the mouse liver.

To measure the effectiveness of LNP-mediated delivery of firefly luciferase containing particles to the liver, LNPs were formulated and characterized as described in Example 60 and tested in vitro prior (Example 39) to administration to mice. C57BL/6 male mice (Charles River Labs) at approximately 8 weeks of age were dosed with LNPs via intravenous (i.v.) route at 1

mg/kg. Vehicle control animals were dosed i.v. with 300 µL phosphate buffered saline. Mice were injected via intraperitoneal route with dexamethasone at 5 mg/kg 30 minutes prior to injection of LNPs. Tissues were collected at necropsy at or 6, 24, 48 hours after LNP administration with a group size of 5 mice per time point. Liver and other tissue samples were collected, snap-frozen in liquid nitrogen, and stored at −80°C until analysis.
Frozen liver samples were pulverized on dry ice and transferred to homogenization tubes containing lysing matrix D beads (MP Biomedical). Ice-cold 1x luciferase cell culture lysis reagent (CCLR) (Promega) was added to each tube and the samples were homogenized in a Fast Prep-24 5G Homogenizer (MP Biomedical) at 6 m/s for 40 seconds. The samples were transferred to a clean microcentrifuge tube and clarified by centrifugation. Prior to luciferase activity assay, the protein concentration of liver homogenates was determined for each sample using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Luciferase activity was measured with 200 µg (total protein) of liver homogenate using the luciferase assay reagent (Promega) according to manufacturer's instructions using an i3X SpectraMax plate reader (Molecular Devices). Liver samples revealed successful delivery of mRNA by all lipid formulations, with reporter activity following the ranking LIPIDV005>LIPIDV004>LIPIDV003 (FIG. 38). As shown in FIG. 38, Firefly luciferase mRNA-containing LNPs were formulated and delivered to mice by iv, and liver samples were harvested and assayed for luciferase activity at 6, 24, and 48 hours post administration. Reporter activity by the various formulations followed the ranking LIPIDV005>LIPIDV004>LIPIDV003. RNA expression was transient and enzyme levels returned near vehicle background by 48 hours. Post-administration. This assay validated the use of these ionizable lipids and their respective formulations for RNA systems for delivery to the liver.

Without wishing to be limited by example, the lipids and formulations described in this example are support the efficacy for the in vivo delivery of other RNA molecules beyond a reporter mRNA. All-RNA Gene Writing systems can be delivered by the formulations described herein. For example, all-RNA systems employing a Gene Writer polypeptide mRNA, Template RNA, and an optional second-nick gRNA are described for editing the genome in vitro by nucleofection, using modified nucleotides, by lipofection, and for editing primary T cells. As described in this application, these all-RNA systems have many unique advantages in cellular

immunogenicity and toxicity, which is of importance when dealing with more sensitive primary cells, especially immune cells, e.g., T cells, as opposed to immortalized cell culture cell lines. Further, it is contemplated that these all RNA systems could be targeted to alternate tissues and cell types using novel lipid delivery systems as referenced herein, e.g., for delivery to the liver, the lungs, muscle, immune cells, and others, given the function of Gene Writing systems has been validated in multiple cell types in vitro here, and the function of other RNA systems delivered with targeted LNPs is known in the art. The in vivo delivery of Gene Writing systems has potential for great impact in many therapeutic areas, e.g., correcting pathogenic mutations), instilling protective variants, and enhancing cells endogenous to the body, e.g., T cells. Given an appropriate formulation, all-RNA Gene Writing is conceived to enable the manufacture of cell-based therapies in situ in the patient.

Example 41: Improvement of expression of Cas-RT fusions through linker selection

This example demonstrates the optimization of Cas-RT fusions to improve protein expression in mammalian cells. Construction of novel Cas-RT fusions by the simple substitution of new functional domains may result in low or moderate expression of the Gene Writer polypeptide. Thus, it is contemplated here that modified configurations of the fusion may be advantageous in the context of different domains. Without wishing to be limited by the example, one such approach for improving the expression and stability of new fusions is through the use of a linker library. Here, the peptide linker sequence between the Cas and RT domains of the Cas-RT fusion is varied using a library of linker sequences. More specifically, linkers from Table 38B were used to generate new variants of a Cas9-RT fusion construct previously demonstrating low protein expression and delivered to human cells to screen for improved Cas-RT protein expression.

A set of 22 peptide linkers (Table 38B) with varying degrees of length, flexibility, hydrophobicity, and secondary structure was first used to generate variants of a Cas-RT fusion protein by substitution of the original linker (SQ ID NO: 480). HEK293T cells were transfected by electroporation of 250,000 cells/well with ~800 ng of each Cas9-RT fusion plasmid along with 200 ng of a single-guide RNA plasmid. To assess the expression level of Cas9-RT fusions, cell lysates were collected on day 2 post-transfection and analyzed by Western blot using a primary antibody against Cas9. linker 10 (SEQ ID NO: 468) listed in Table 38 Bsignificantly

improved Cas-RT fusion expression (Figure 39), demonstrating the potentially profound impact of the peptide linker sequence on Cas-RT expression.

Table 38B. Peptide sequences used as linkers between the Cas and RT domains in Gene Writer polypeptides comprising Cas-RT fusions

#	Linker sequence	Notes	Seq Id. No.
1	GGS	Short	459
2	GGGGS	Flexible, short	460
3	$(GGGGS)_2$	Flexible	461
4	$(GGGGS)_3$	Flexible, long	462
5	$(GGGGS)_4$	Flexible, very long	463
6	$(G)_6$	Flexible	464
7	$(G)_8$	Flexible	465
8	GSAGSAAGSGEF	Flexible	466
9	(GSSGSS)	Mid	467
10	$(GSSGSS)_2$	Mid, Flexible	468
11	$(GSSGSS)_3$	Mid	469
12	SGSETPGTSESATPES	XTEN	470
13	(EAAAK)	Rigid helix, short	471
14	$(EAAAK)_2$	Rigid helix, mid	472
15	$(EAAAK)_3$	Rigid helix, long	473
16	PAP	Rigid, short	474

17	PAPAP	Rigid, short	**475**
18	PAPAPAPAP	Rigid, mid	**476**
19	$A(EAAAK)_4ALEA(EAAAK)_4A$	Rigid, very long with helices	**477**
20	GGGGS(EAAAK)GGGGS	Flexible – helix – flex	**478**
21	(EAAAK)GGGGS(EAAAK)	Helix – flex – helix	**479**
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	Flexible – XTEN – flexible	**480**

Without wishing to be limited by example, the lipids and formulations described in this example are support the efficacy for the in vivo delivery of other RNA molecules beyond a reporter mRNA. All-RNA Gene Writing systems can be delivered by the formulations described herein. For example, all-RNA systems employing a Gene Writer polypeptide mRNA, Template RNA, and an optional second-nick gRNA are described for editing the genome in vitro by nucleofection, using modified nucleotides, by lipofection, and for editing primary T cells. As described in this application, these all-RNA systems have many unique advantages in cellular immunogenicity and toxicity, which is of importance when dealing with more sensitive primary cells, especially immune cells, e.g., T cells, as opposed to immortalized cell culture cell lines. Further, it is contemplated that these all RNA systems could be targeted to alternate tissues and cell types using novel lipid delivery systems as referenced herein, e.g., for delivery to the liver, the lungs, muscle, immune cells, and others, given the function of Gene Writing systems has been validated in multiple cell types in vitro here, and the function of other RNA systems delivered with targeted LNPs is known in the art. The in vivo delivery of Gene Writing systems has potential for great impact in many therapeutic areas, e.g., correcting pathogenic mutations), instilling protective variants, and enhancing cells endogenous to the body, e.g., T cells. Given an

appropriate formulation, all-RNA Gene Writing is conceived to enable the manufacture of cell-based therapies in situ in the patient.

It should be understood that for all numerical bounds describing some parameter in this application, such as "about," "at least," "less than," and "more than," the description also necessarily encompasses any range bounded by the recited values. Accordingly, for example, the description "at least 1, 2, 3, 4, or 5" also describes, *inter alia*, the ranges 1-2, 1-3, 1-4, 1-5, 2-3, 2-4, 2-5, 3-4, 3-5, and 4-5, *et cetera*.

For all patents, applications, or other reference cited herein, such as non-patent literature and reference sequence information, it should be understood that they are incorporated by reference in their entirety for all purposes as well as for the proposition that is recited. Where any conflict exists between a document incorporated by reference and the present application, this application will control. All information associated with reference gene sequences disclosed in this application, such as GeneIDs or accession numbers (typically referencing NCBI accession numbers), including, for example, genomic loci, genomic sequences, functional annotations, allelic variants, and reference mRNA (including, e.g., exon boundaries or response elements) and protein sequences (such as conserved domain structures), as well as chemical references (e.g., PubChem compound, PubChem substance, or PubChem Bioassay entries, including the annotations therein, such as structures and assays, et cetera), are hereby incorporated by reference in their entirety.

Headings used in this application are for convenience only and do not affect the interpretation of this application.

What is claimed is:

1. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the polypeptide comprises a mutation inactivating and/or deleting a nucleolar localization signal.

2. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a first target DNA binding domain, e.g., comprising a first Zn finger domain, (ii) a reverse transcriptase domain, (iii) an endonuclease domain, and (iv) a second target DNA binding domain, e.g., comprising a second Zn finger domain, heterologous to the first target DNA binding domain; and

optionally (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein (a) binds to a smaller number of target DNA sequences in a target cell than a similar polypeptide that comprises only the first target DNA binding domain, e.g., wherein the presence of the second target DNA binding domain in the polypeptide with the first DNA binding domain refines the target sequence specificity of the polypeptide relative to the polypeptide target sequence specificity of the polypeptide comprising only the first target DNA binding domain.

3. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

optionally, (b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the system is capable of cutting the first strand of the target DNA at least twice (e.g., twice), and

optionally wherein the cuts are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or 200 nucleotides away one another (and optionally no more than 500, 400, 300, 200, or 100 nucleotides away from one another).

4. A method of modifying a target DNA strand in a cell, tissue or subject, comprising administering a system to a cell, wherein the system comprises:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the system reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand, and

wherein the cell has decreased Rad51 repair pathway activity, decreased expression of Rad51 or a component of the Rad51 repair pathway, or does not comprise a functional Rad51 repair pathway, e.g., does not comprise a functional Rad51 gene, e.g., comprises a mutation (e.g., deletion) inactivating one or both copies of the Rad51 gene or another gene in the Rad51 repair pathway.

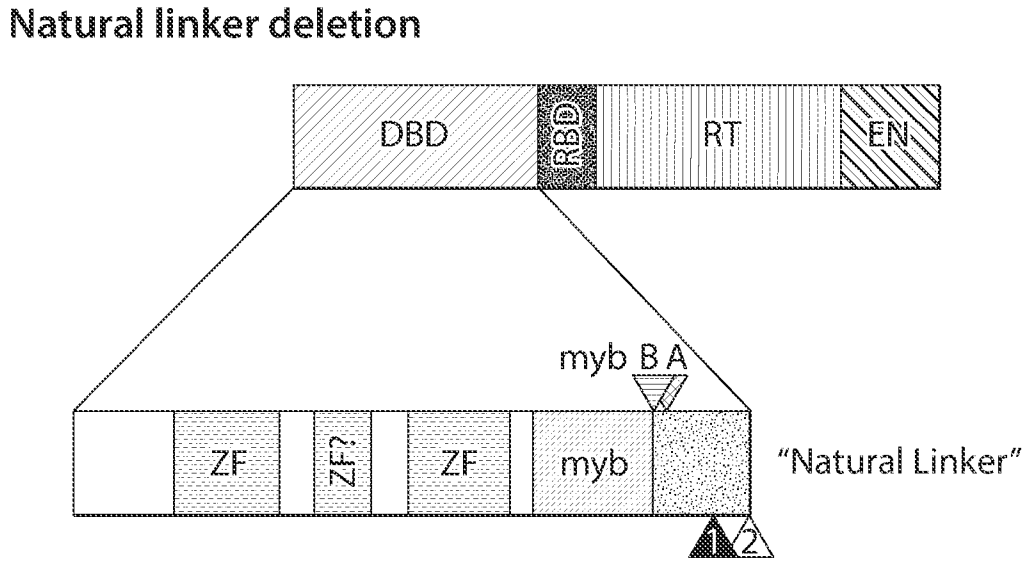

Natural linker deletion
DBD
RBD
RT
EN
myb B A
ZF
ZF?
ZF
myb
"Natural Linker"
1
2

FIG. 1A

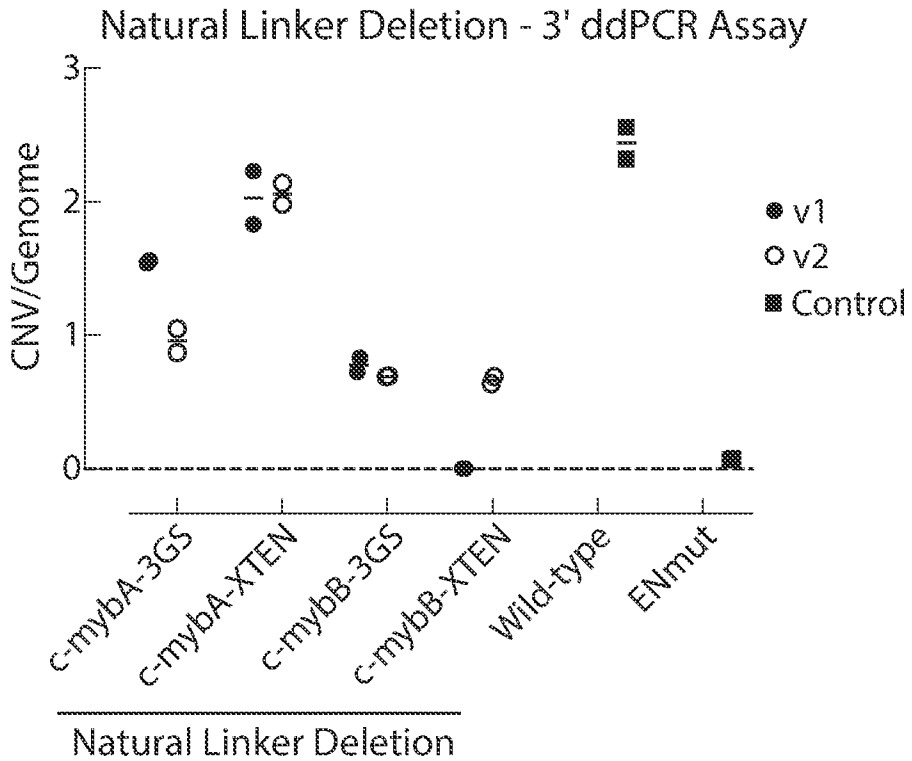

Natural Linker Deletion - 3' ddPCR Assay
CNV/Genome
3
2
1
0
v1
v2
Control
c-mybA-3GS
c-mybA-XTEN
c-mybB-3GS
c-mybB-XTEN
Wild-type
ENmut
Natural Linker Deletion

FIG. 1B

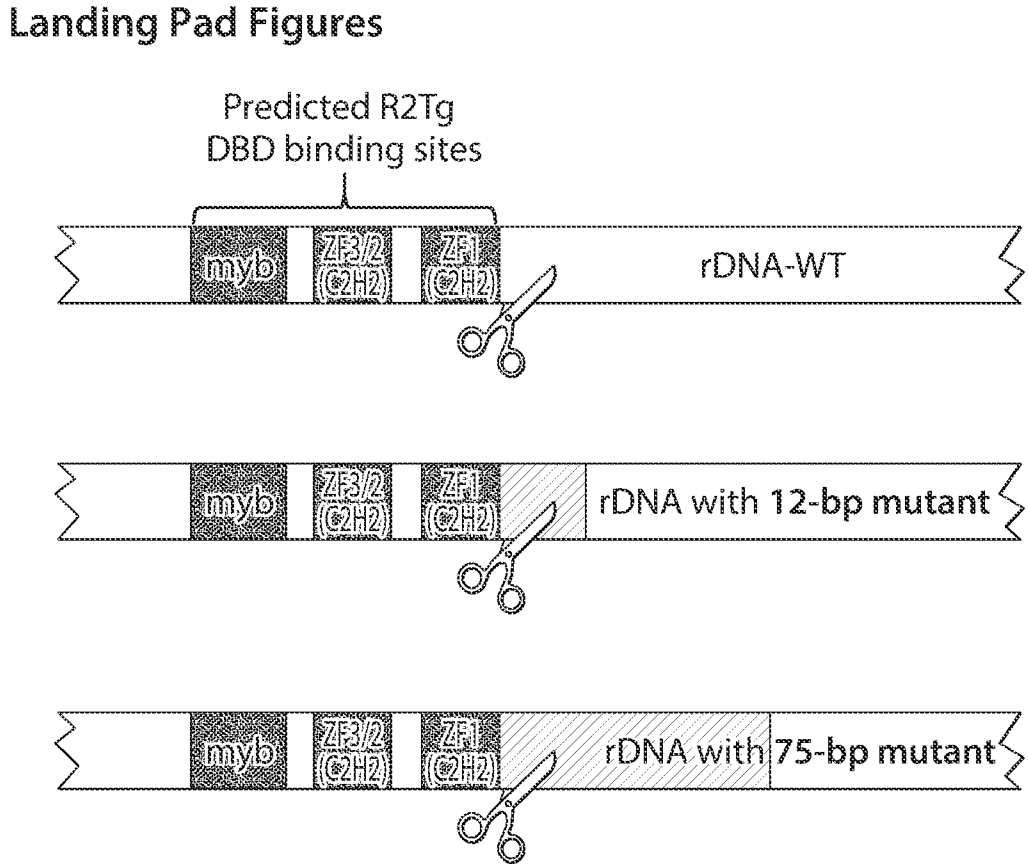

Landing Pad Figures
Predicted R2Tg
DBD binding sites
myb
ZF3/2
(C2H2)
ZF1
(C2H2)
rDNA-WT
myb
ZF3/2
(C2H2)
ZF1
(C2H2)
rDNA with 12-bp mutant
myb
ZF3/2
(C2H2)
ZF1
(C2H2)
rDNA with 75-bp mutant

FIG. 2

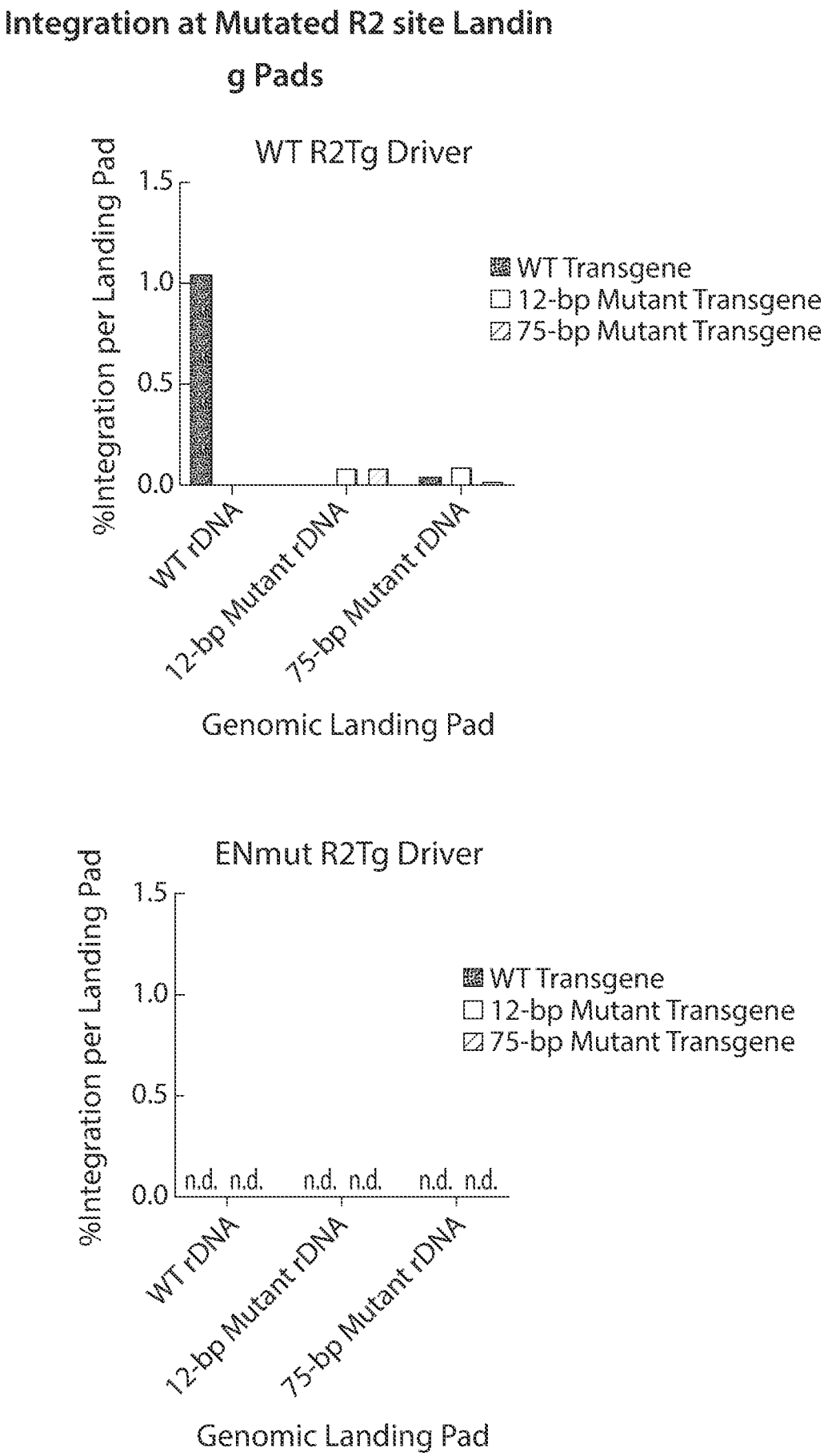

Integration at Mutated R2 site Landing Pads
WT R2Tg Driver
%Integration per Landing Pad
1.5
1.0
0.5
0.0
WT Transgene
12-bp Mutant Transgene
75-bp Mutant Transgene
WT rDNA
12-bp Mutant rDNA
75-bp Mutant rDNA
Genomic Landing Pad
ENmut R2Tg Driver
%Integration per Landing Pad
1.5
1.0
0.5
0.0
n.d. n.d. n.d. n.d. n.d. n.d.
WT Transgene
12-bp Mutant Transgene
75-bp Mutant Transgene
WT rDNA
12-bp Mutant rDNA
75-bp Mutant rDNA
Genomic Landing Pad

FIG. 3A

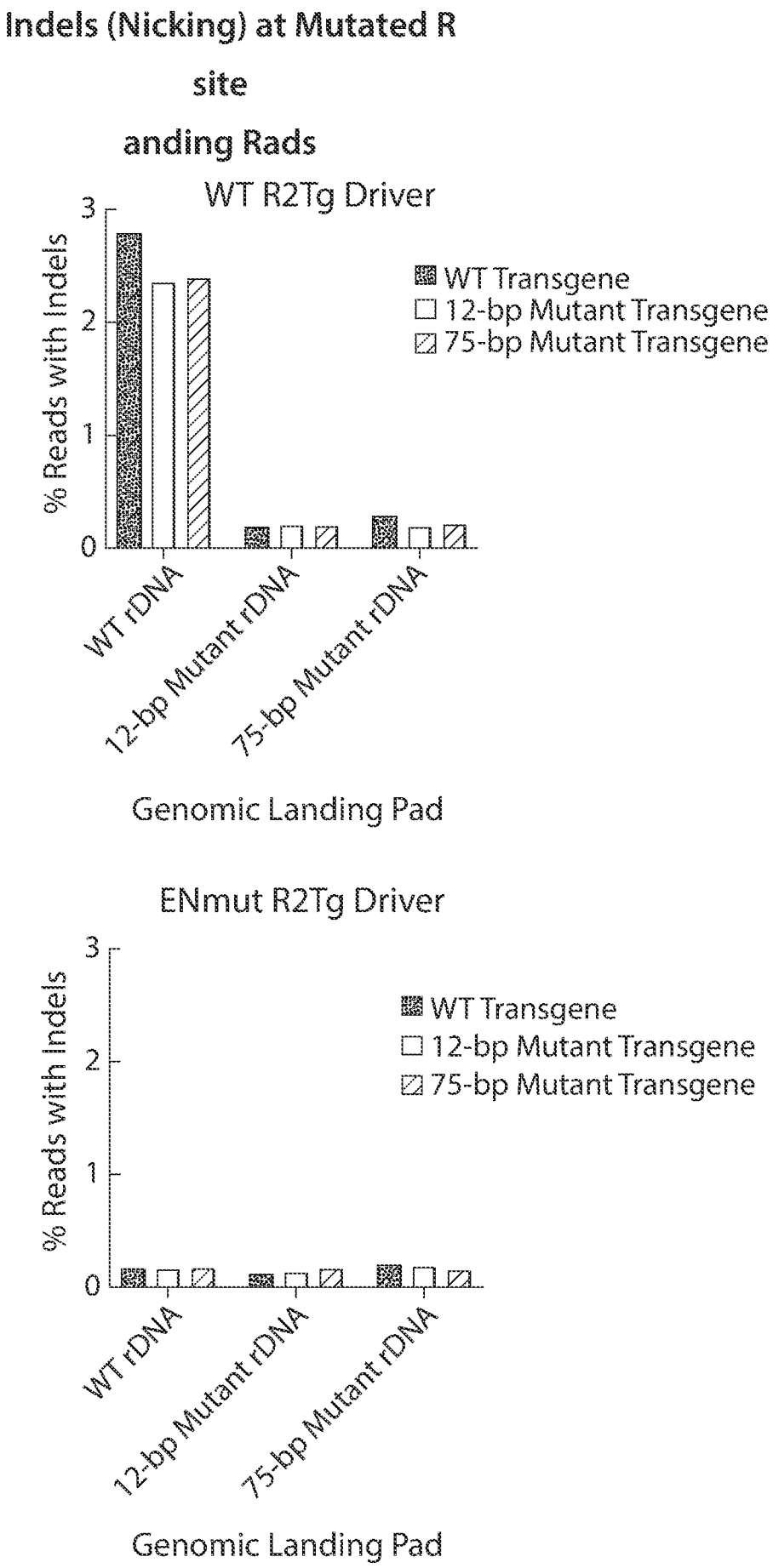

Indels (Nicking) at Mutated R
site
anding Rads
WT R2Tg Driver
% Reads with Indels
3
2
1
0
WT Transgene
12-bp Mutant Transgene
75-bp Mutant Transgene
WT rDNA
12-bp Mutant rDNA
75-bp Mutant rDNA
Genomic Landing Pad
ENmut R2Tg Driver
% Reads with Indels
3
2
1
0
WT Transgene
12-bp Mutant Transgene
75-bp Mutant Transgene
WT rDNA
12-bp Mutant rDNA
75-bp Mutant rDNA
Genomic Landing Pad

FIG. 3B

Examples of Re-targeting Gene Writer polypeptiGe Gy replaGing DNA GinGing Gomain

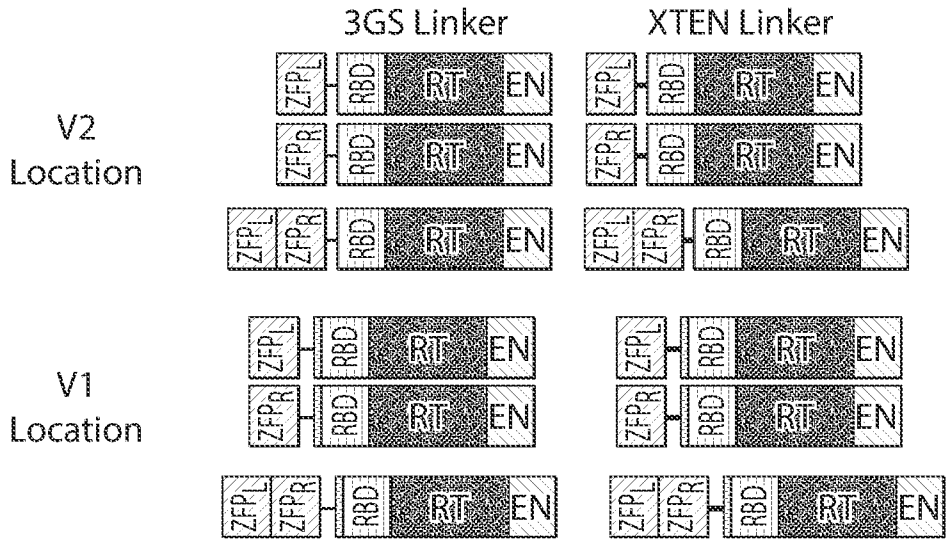


FIG. 4

Examples of Re-targeting Gene Writer polypeptide by replacing DNA binding domain

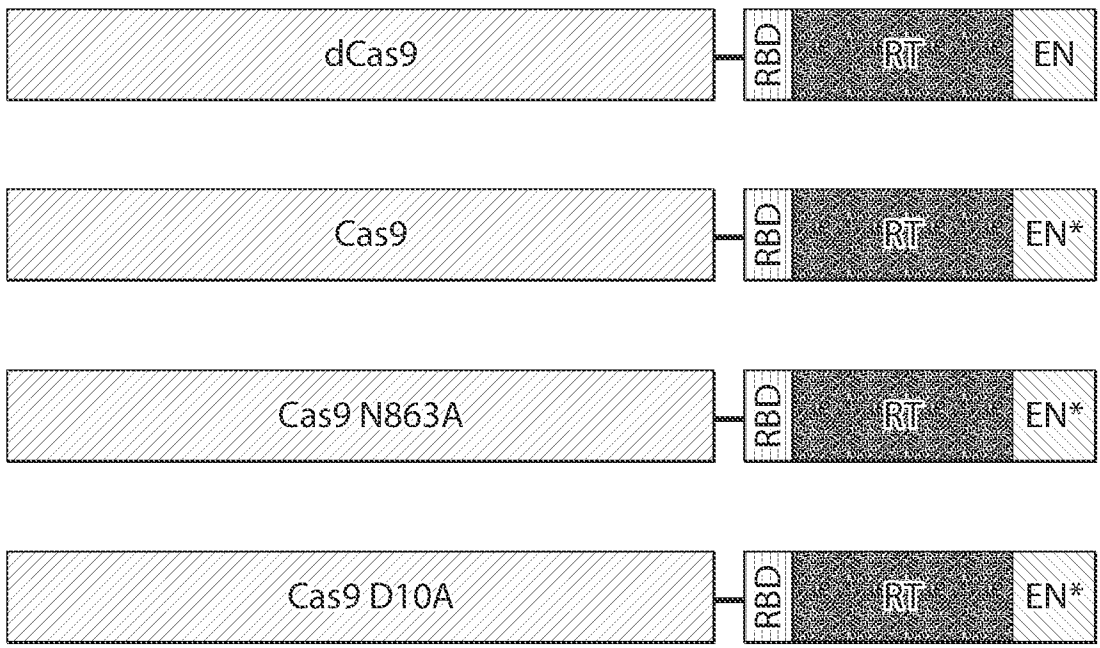


FIG. 5

Examples of Re-targeting Gene Writer polypeptide by Fusion to terminus

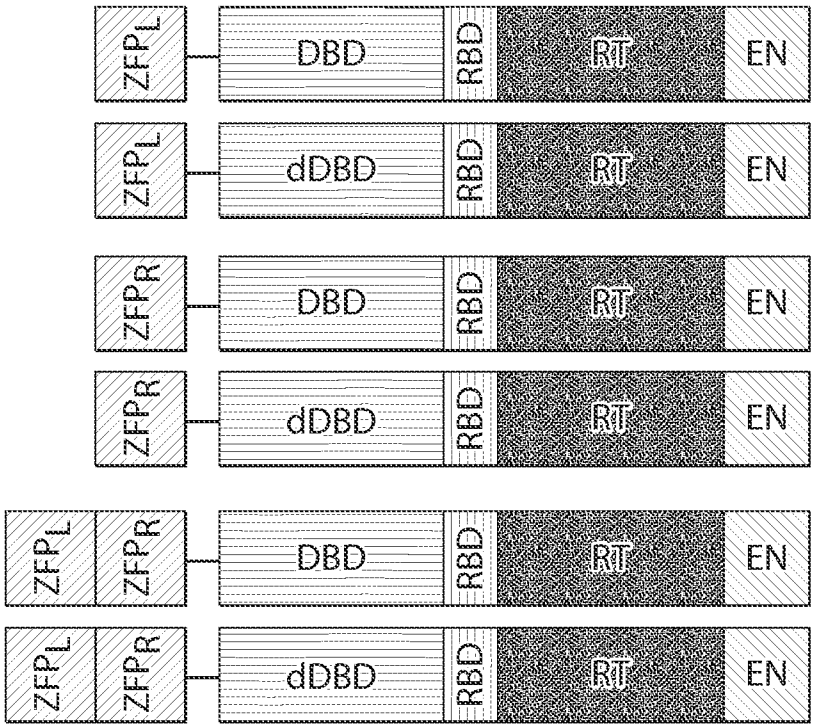


FIG. 6

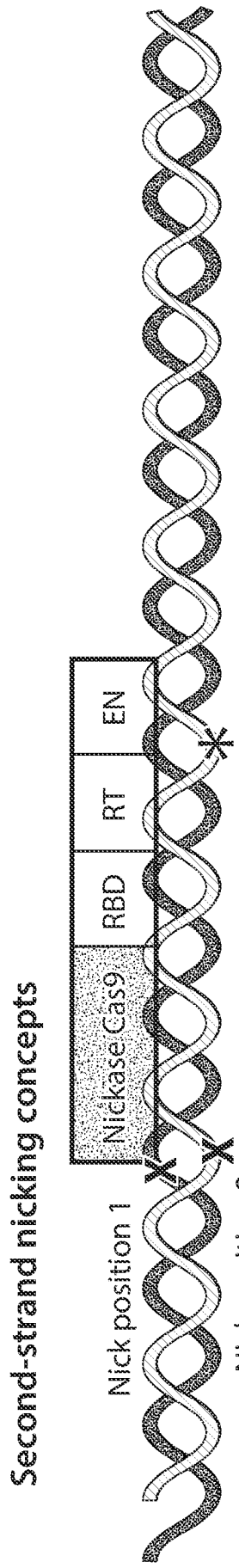

Second-strand nicking concepts
Nickase Cas9
RBD
RT
EN
Nick position 1
Nick position 2

FIG. 7A

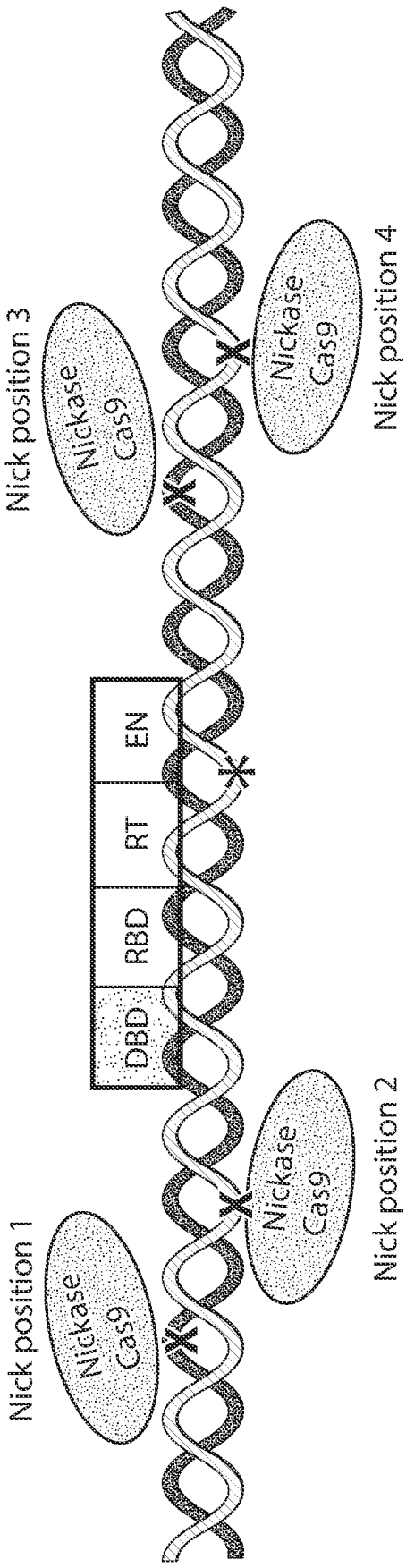

Nick position 3
Nickase Cas9
Nickase Cas9
Nick position 4
DBD
RBD
RT
EN
Nick position 1
Nickase Cas9
Nickase Cas9
Nick position 2

FIG. 7B

Second Strand Nicking: Cas9-Gene Writer Fusion with 3′ extended gRNA donor

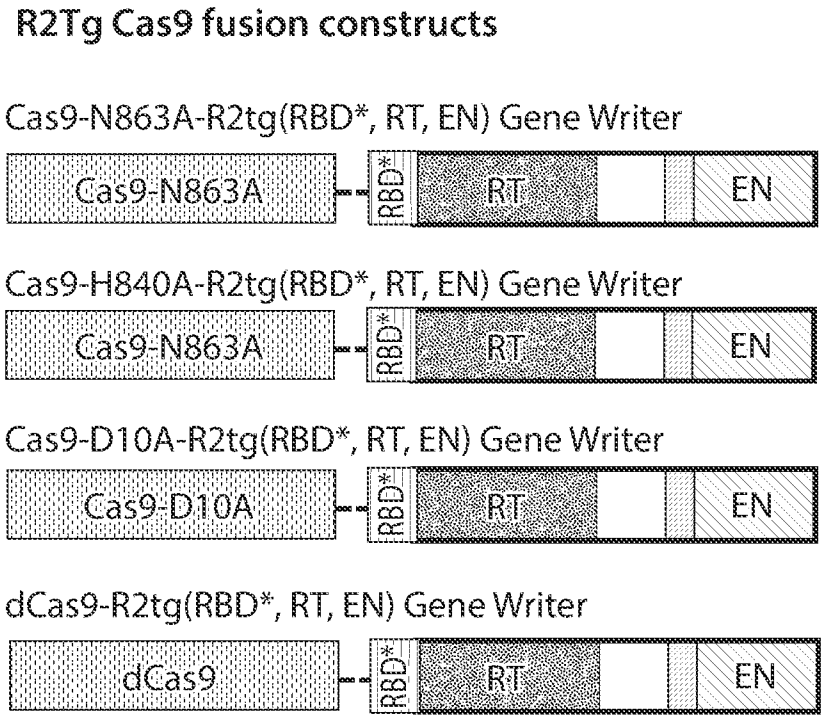


FIG. 8A

Donor: gRNA extended at 5′ end

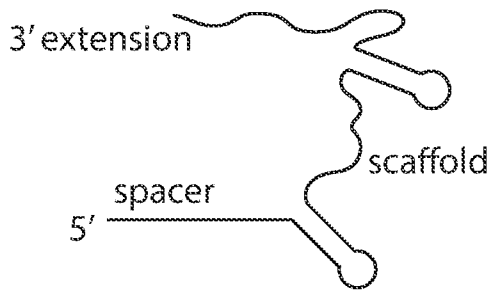


FIG. 8B

Second Strand Nicking: Cas9-Gene Writer Fusions with UTR flanked transgene

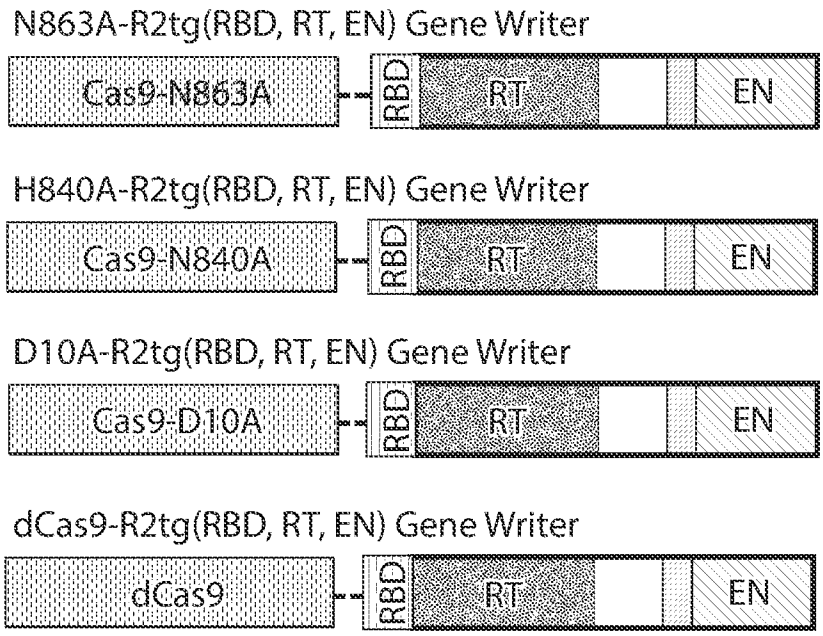


FIG. 9A

Donor: UTR-transgene



FIG. 9B

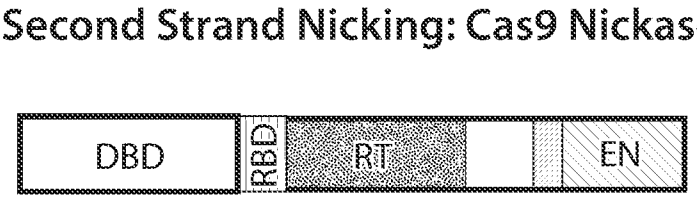


FIG. 10A



FIG. 10B



FIG. 10C

Improved Gene Writer polypeptide expression



FIG. 11A

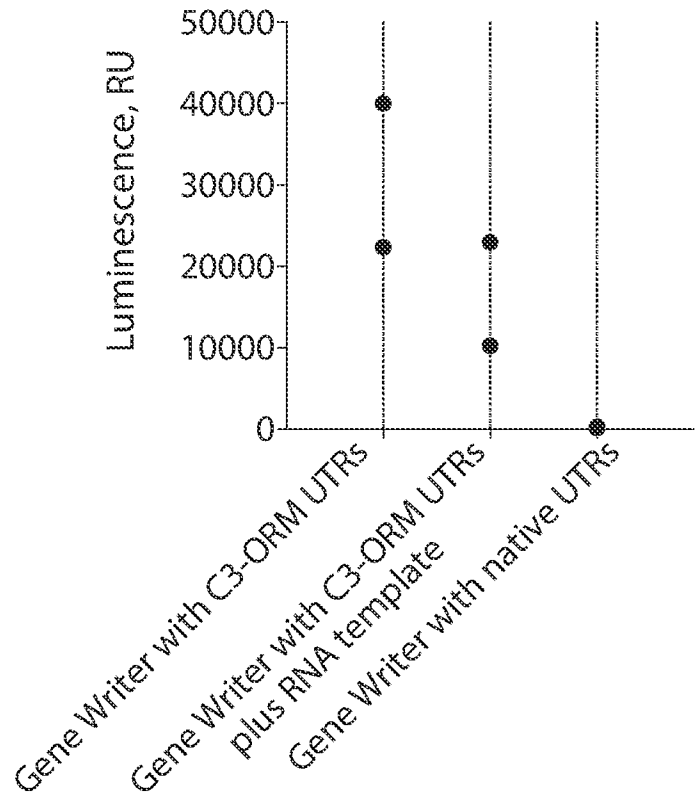


FIG. 11B

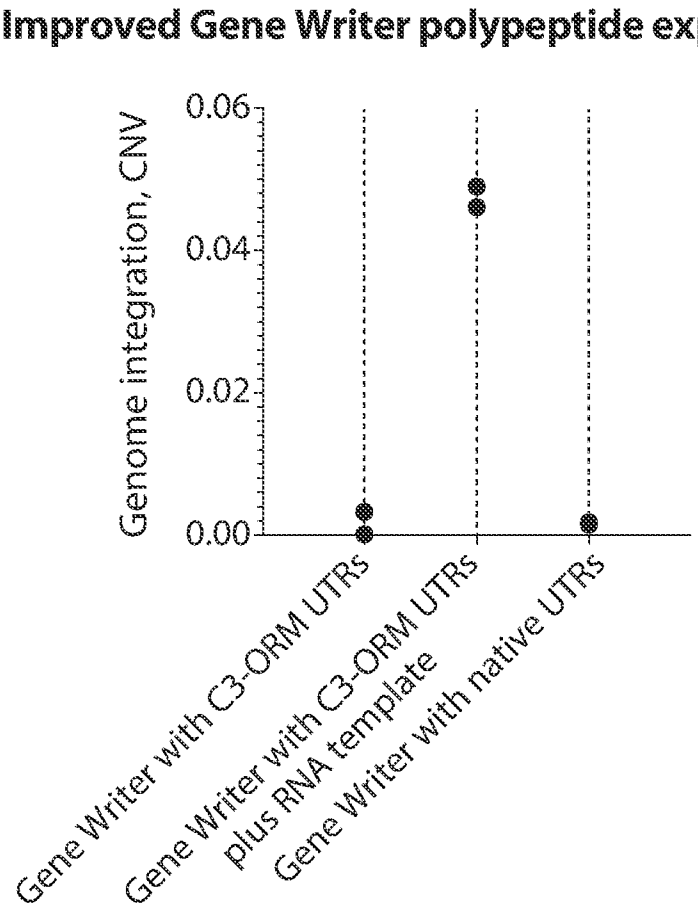

Improved Gene Writer polypeptide expression
Genome integration, CNV
0.06
0.04
0.02
0.00
Gene Writer with C3-ORM UTRs
Gene Writer with C3-ORM UTRs plus RNA template
Gene Writer with native UTRs

FIG. 12

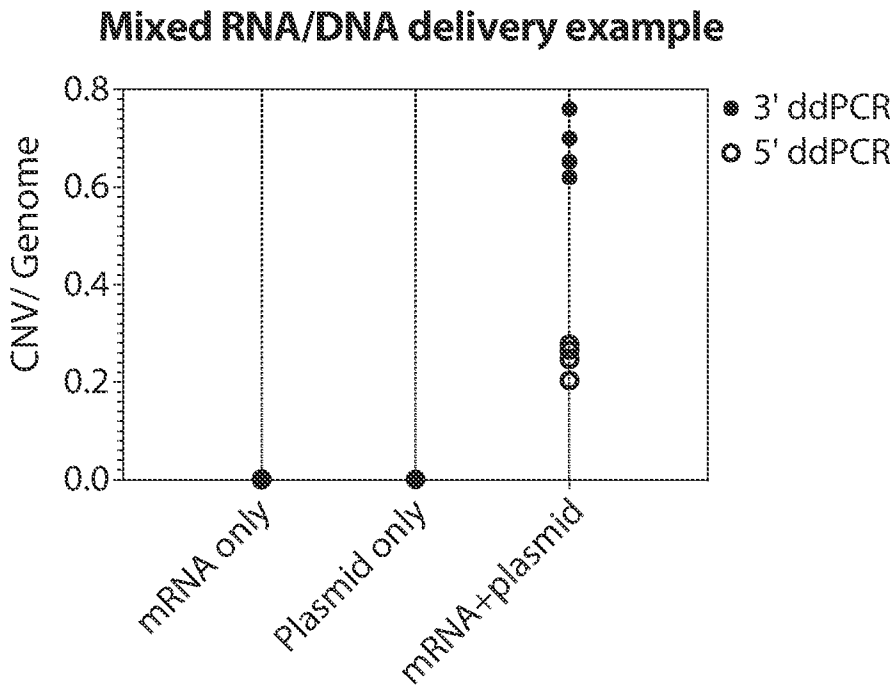

Mixed RNA/DNA delivery example
CNV/ Genome
0.8
0.6
0.4
0.2
0.0
3' ddPCR
5' ddPCR
mRNA only
Plasmid only
mRNA+plasmid

FIG. 13

Engineering 5'UTR

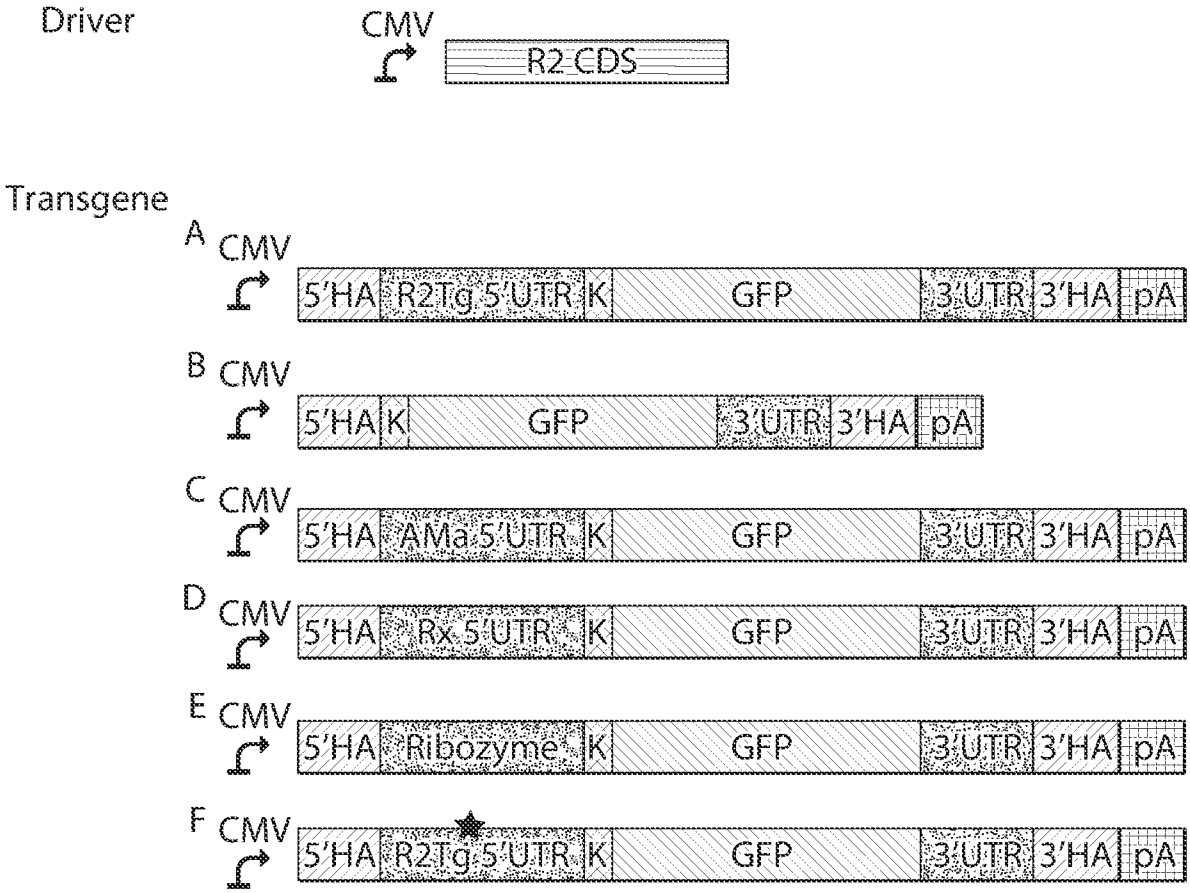


FIG. 14

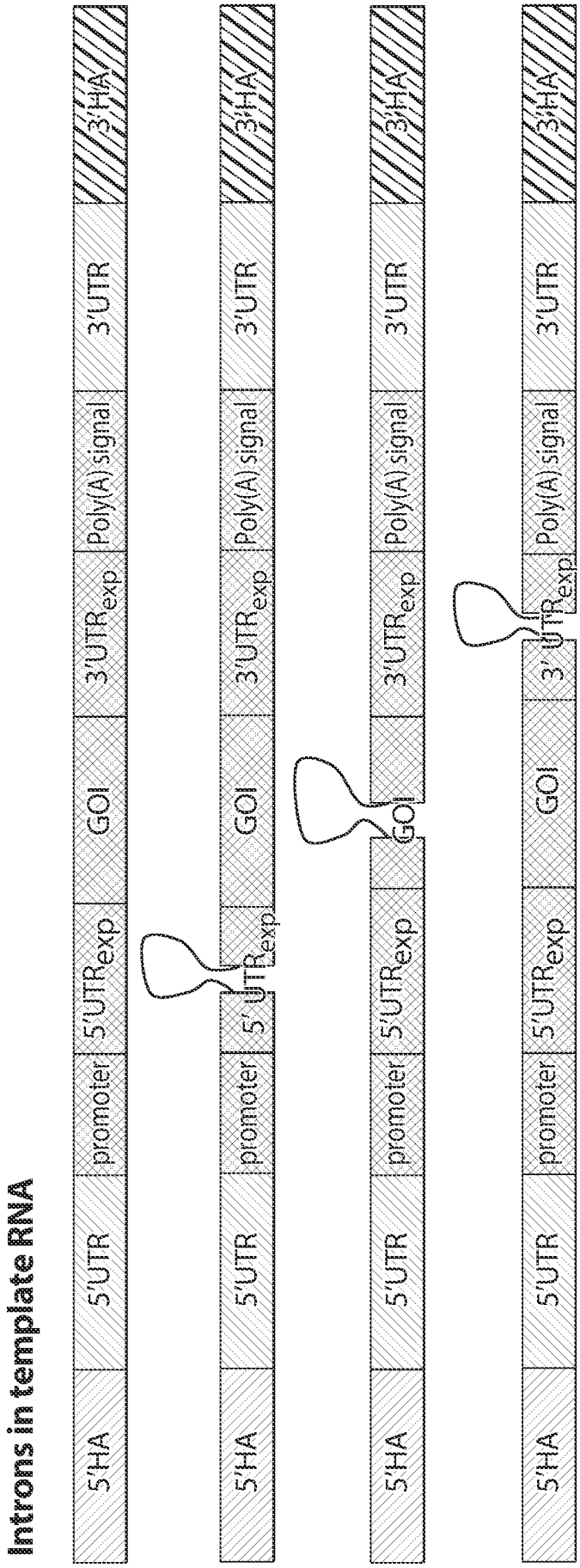

Introns in template RNA
5'HA
5'UTR
promoter
5'UTRexp
GOI
3'UTRexp
Poly(A) signal
3'UTR
3'HA

FIG. 15

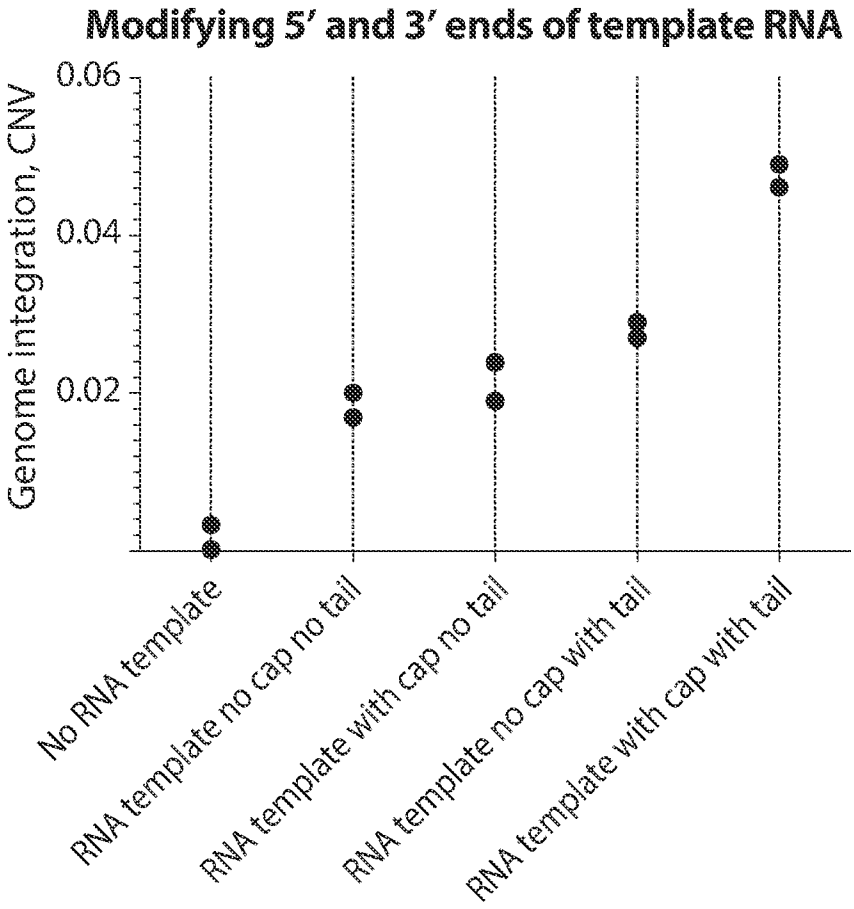

Modifying 5′ and 3′ ends of template RNA
Genome integration, CNV
0.06
0.04
0.02
No RNA template
RNA template no cap no tail
RNA template with cap no tail
RNA template no cap with tail
RNA template with cap with tail

FIG. 16

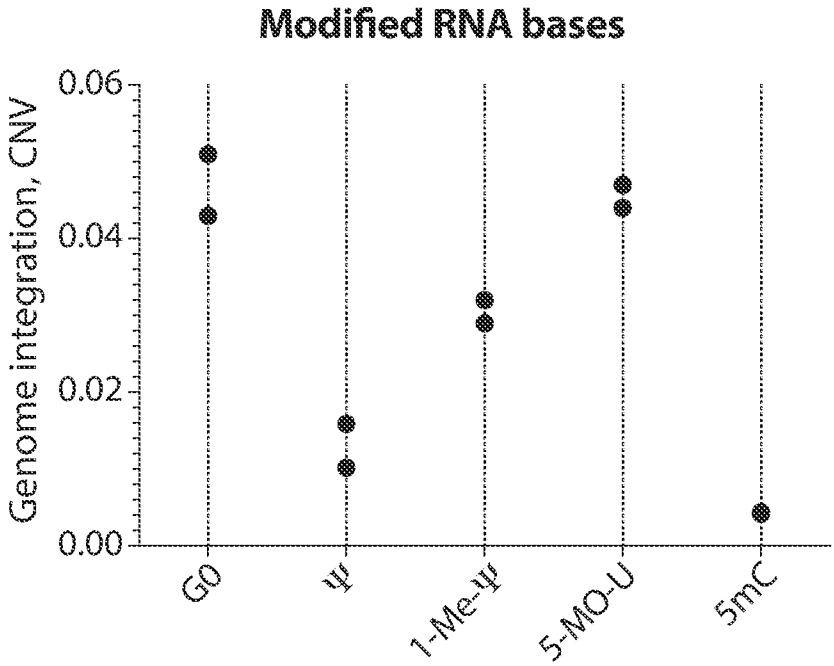

Modified RNA bases
Genome integration, CNV
0.06
0.04
0.02
0.00
G0
Ψ
1-Me-Ψ
5-MO-U
5mC

FIG. 17

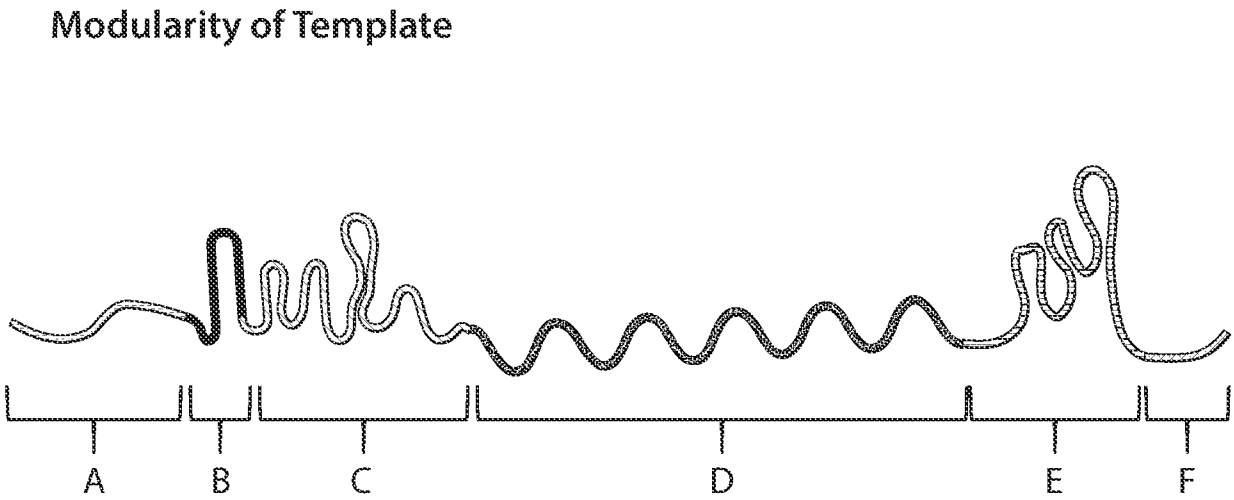


FIG. 18

Modularity of template

Module	Function
A: 5′ homology arm	The 5′ homology arm module is complementary to the DNA sequence 5′ to where the GeneWriter system nicks target DNA
B: Ribozyme	The ribozyme module is a sub-part of the 5′ UTR sequence of the retrotransposase
C: 5′ UTR	The 5′ UTR module is the RNA sequence or interacting moiety that the driver of retrotransposase uses in the process of retrotransposition
D: Heterologous object sequence	The heterologous object sequence is an RNA sequence that serves as the template for the GeneWriter polypeptide to insert a desired payload into the targeted genomic location
E: 3′ UTR	The 3′ UTR module is an RNA sequence or interacting moiety that the GeneWriter polypeptide interacts with to bind to the template RNA molecule that is used in whole and/or in part as a template for retrotransposition
F: 3′ homology arm	The 3′ homology arm module is complementary to the DNA sequence 3′ to where the GeneWriter system nicks target DNA

FIG. 19

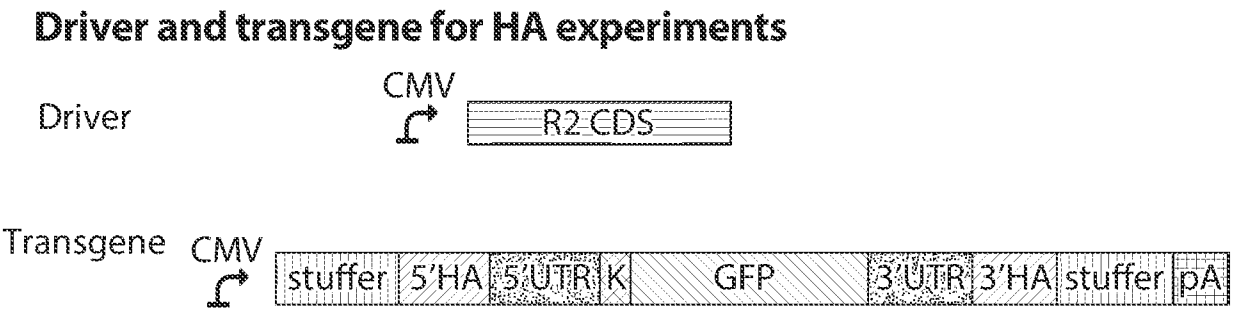

Driver and transgene for HA experiments
Driver
CMV
R2 CDS
Transgene
CMV
stuffer
5'HA
5'UTR
K
GFP
3'UTR
3'HA
stuffer
pA

FIG. 20

ddPCR for HA experiments

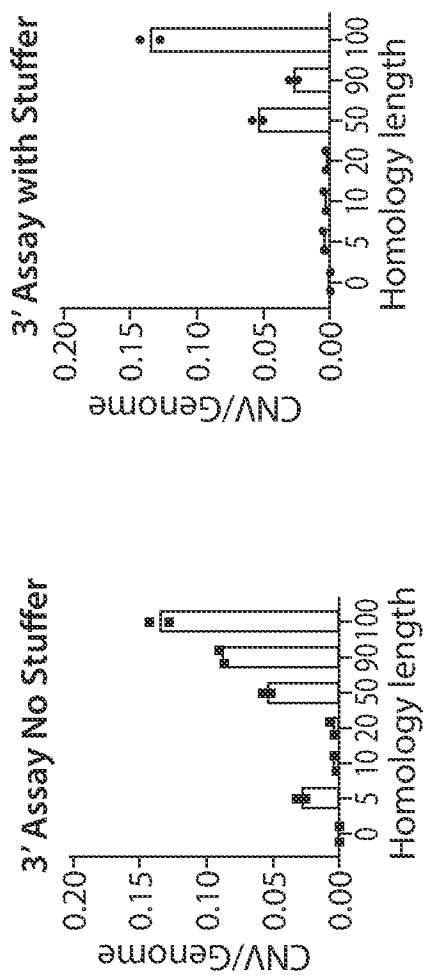


Fig. 21A

Fig. 21B

Fig. 21C

Fig. 21D

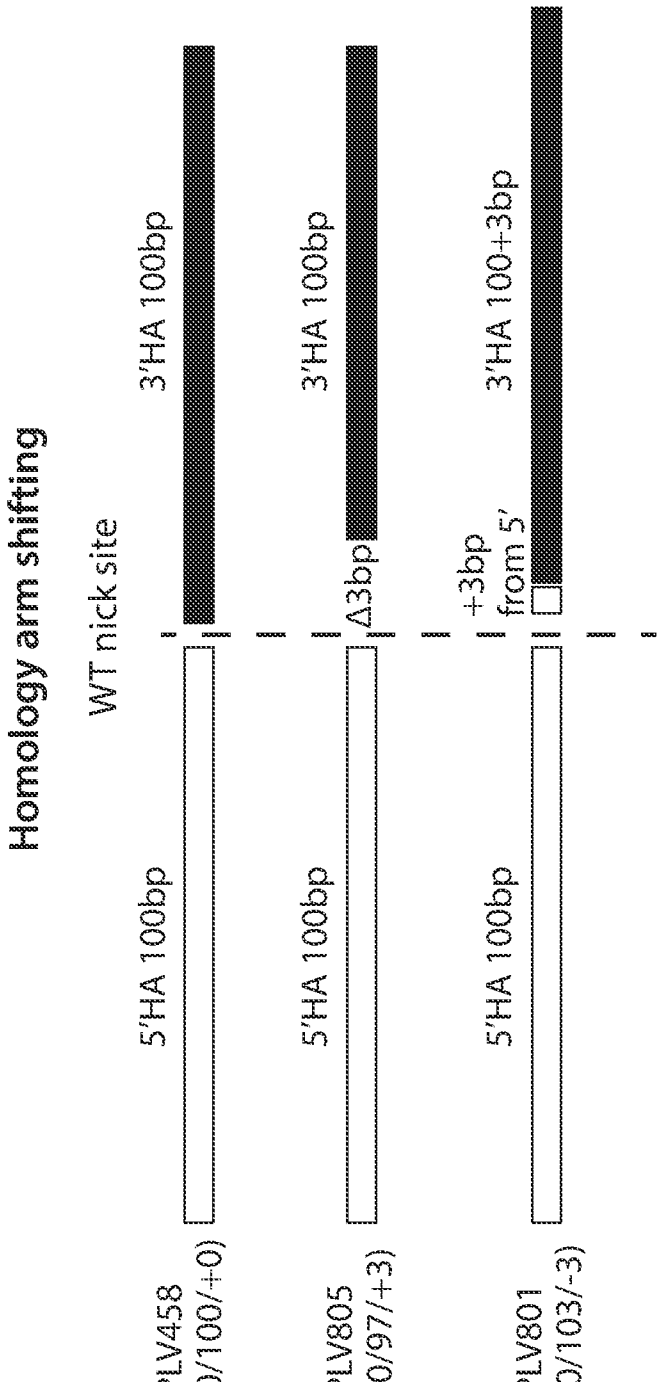

Homology arm shifting
WT nick site
5'HA 100bp
3'HA 100bp
PLV458
(100/100/+0)
5'HA 100bp
Δ3bp
3'HA 100bp
PLV805
(100/97/+3)
5'HA 100bp
+3bp
from 5'
3'HA 100+3bp
PLV801
(100/103/-3)
(5'HA/3'HA/3' shift)

Fig. 22

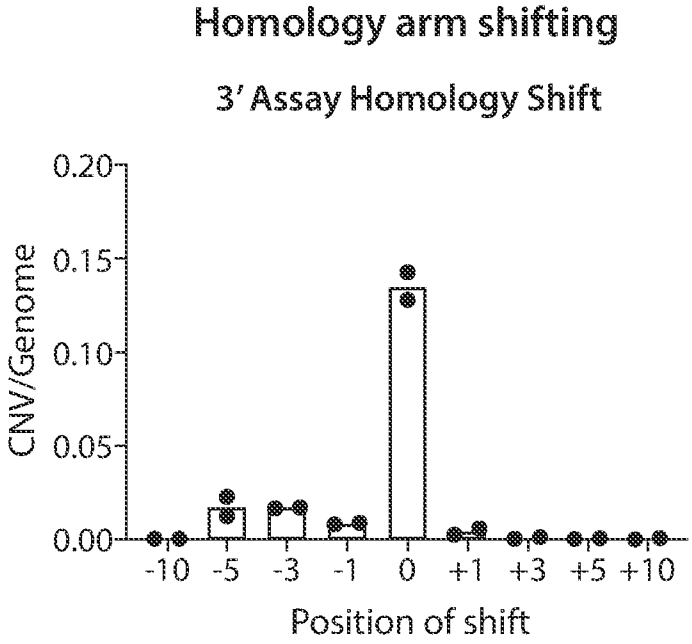

Homology arm shifting
3' Assay Homology Shift
CNV/Genome
0.20
0.15
0.10
0.05
0.00
-10
-5
-3
-1
0
+1
+3
+5
+10
Position of shift

Fig. 23

Gene Writing in the Absence of Rad51

Experimental Outline

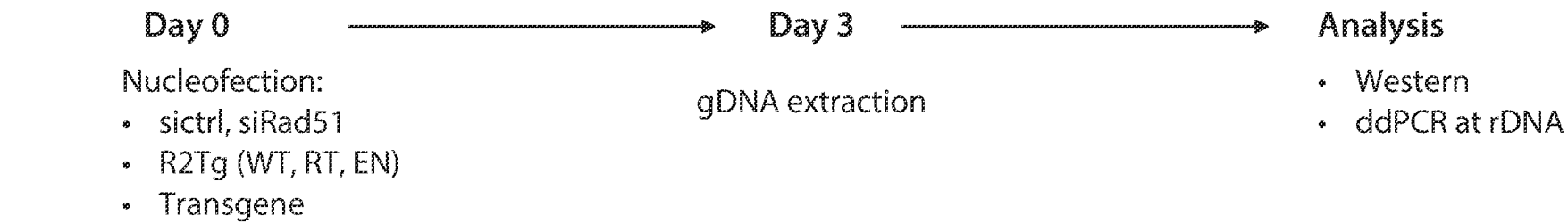


FIG. 24A

R2Tg and transgene configuration

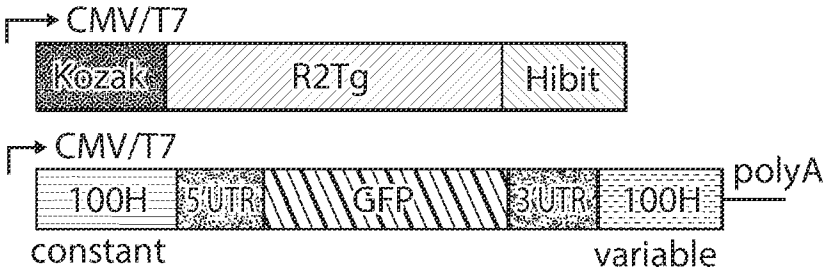


FIG. 24B

Western Blot

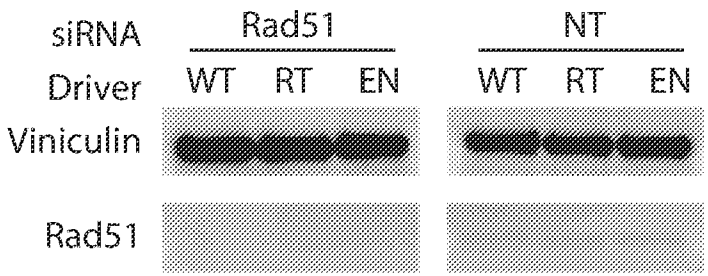


FIG. 24C

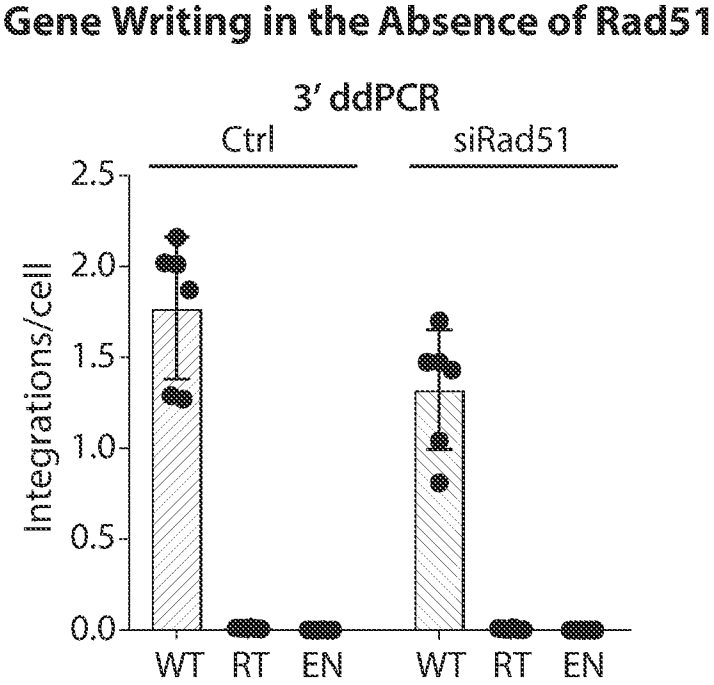


FIG. 25A

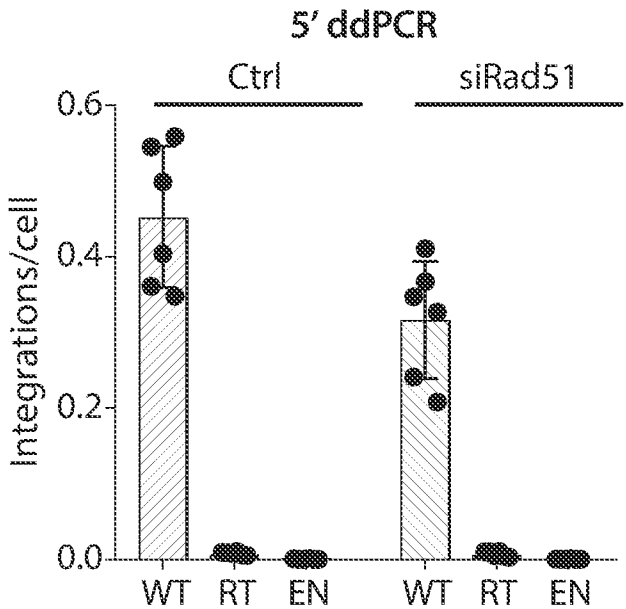


FIG. 25B

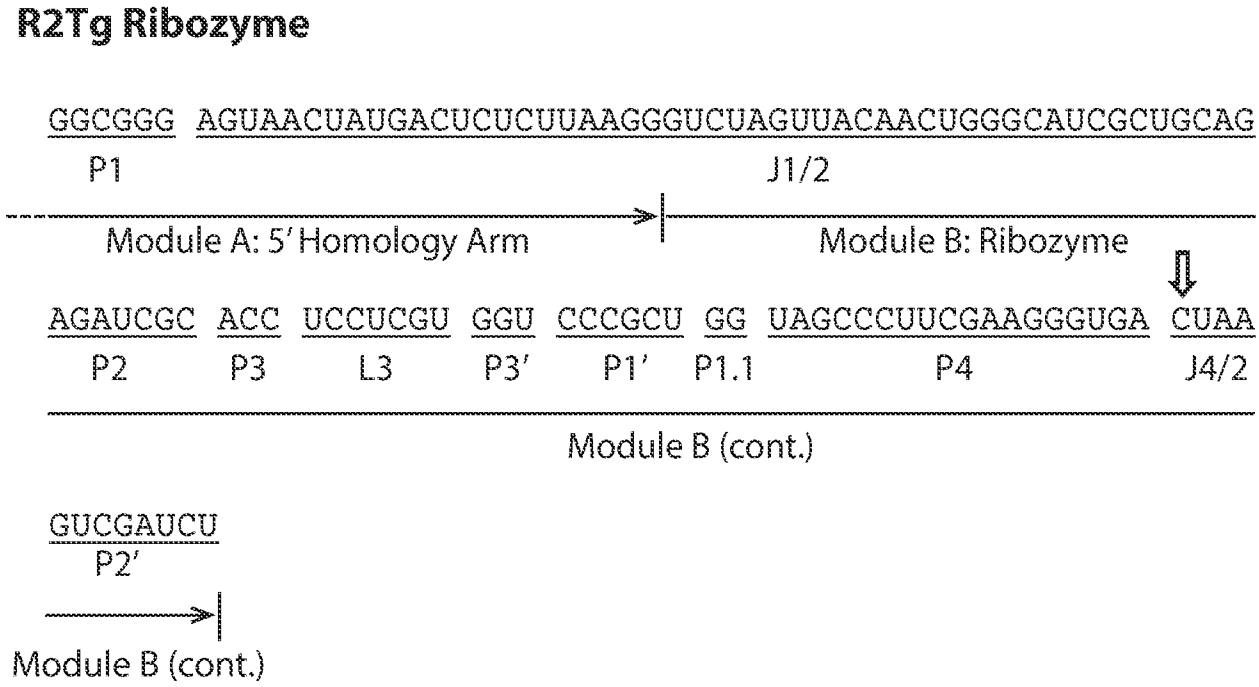

R2Tg Ribozyme
GGCGGG AGUAACUAUGACUCUCUUAAGGGUCUAGUUACAACUGGGCAUCGCUGCAG
P1
J1/2
Module A: 5′ Homology Arm
Module B: Ribozyme
AGAUCGC ACC UCCUCGU GGU CCCGCU GG UAGCCCUUCGAAGGGUGA CUAA
P2 P3 L3 P3′ P1′ P1.1 P4 J4/2
Module B (cont.)
GUCGAUCU
P2′
Module B (cont.)

FIG. 26A

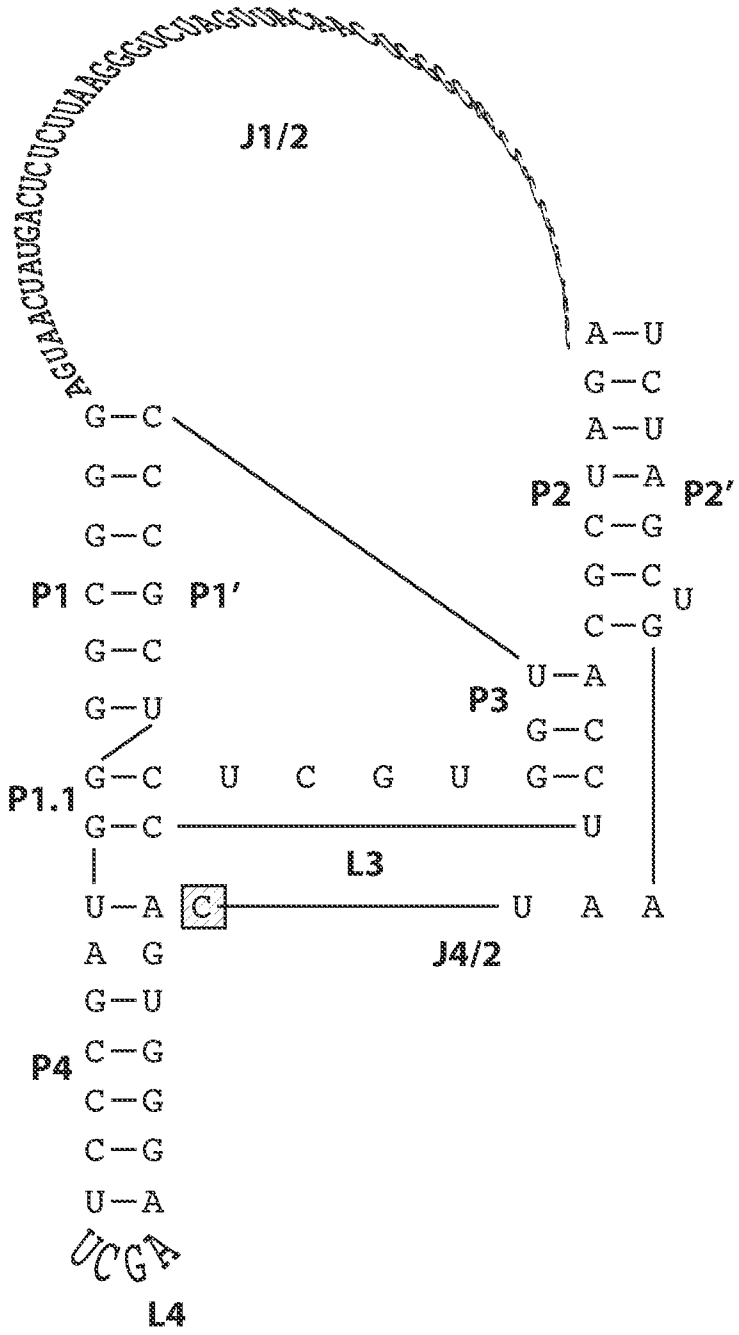


FIG. 26B

R2Tg Ribozyme Sequence

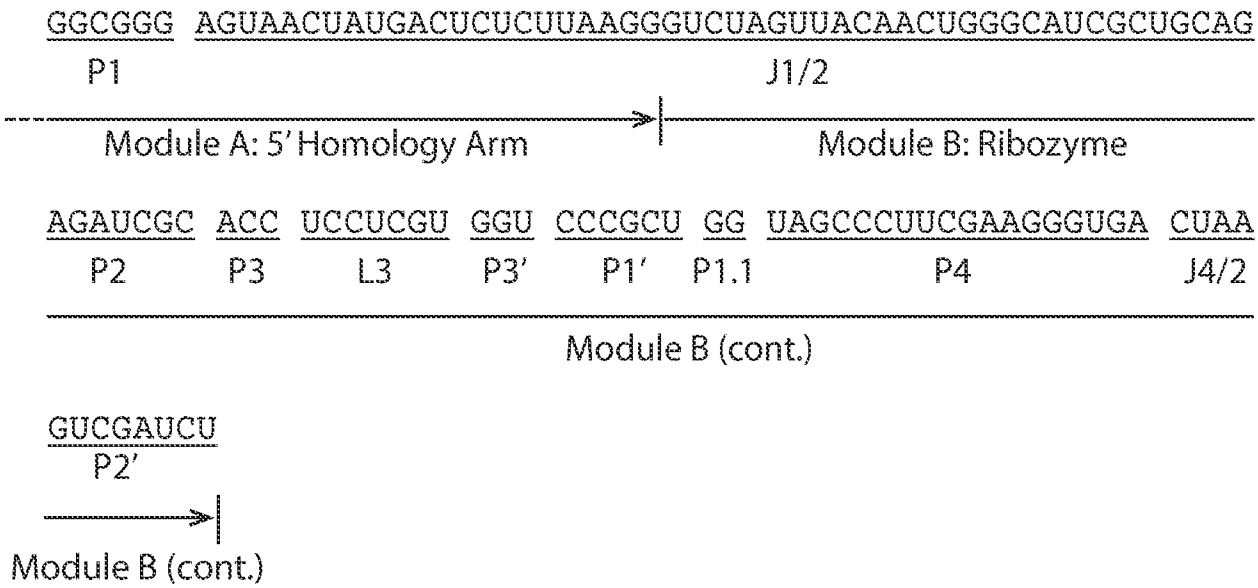


FIG. 27

Predicted Secondary Structure of R2Tg Ribozyme

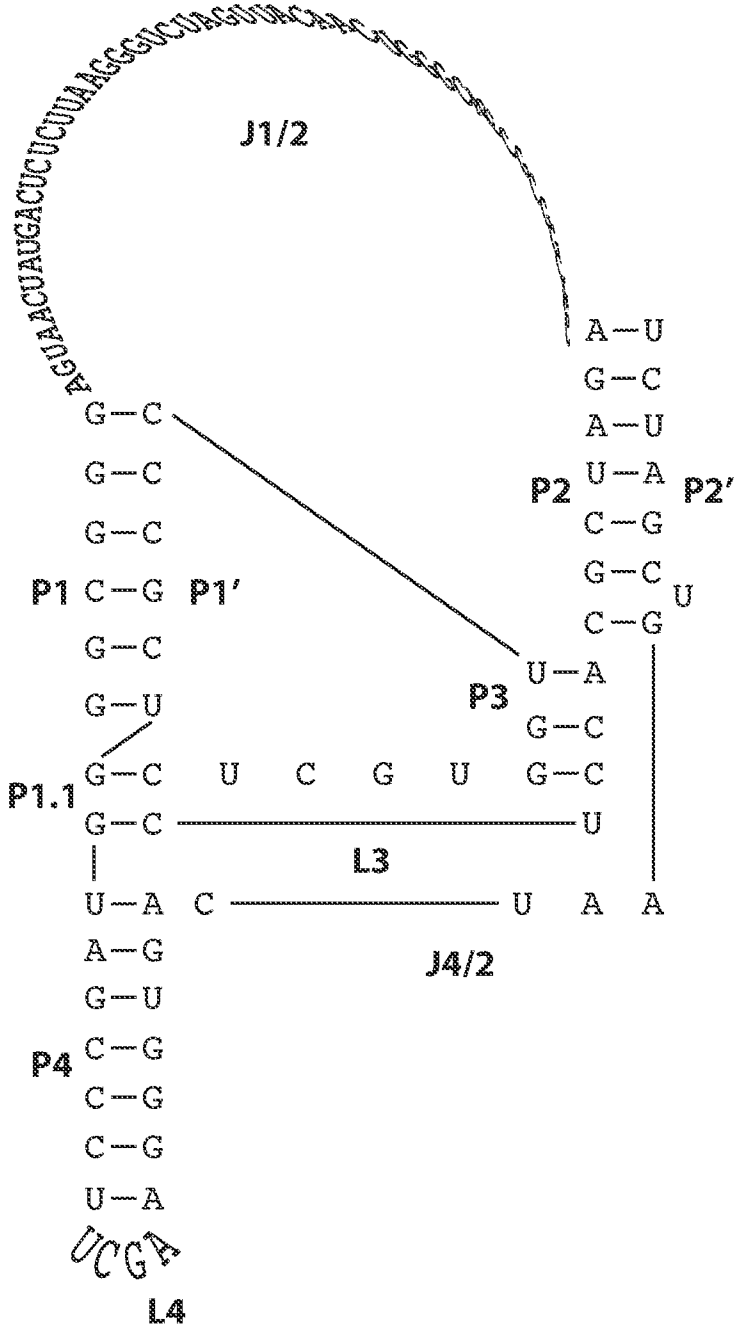


FIG. 28

Gene Writer Modules
DNA Binding
RTn
Zinc finger
TAL domain
Cas9, dCas9, nickase Cas9
Transcription factor
Meganuclease
RNA binding
B-box protein
MS2 coat protein
dCas protein
RTn
Reverse transcriptase
RTn
Other RT
Peptide containing RT domain
Endonuclease
RTn
Cas9, Cas 9 nickase
Restriction enzyme
(e.g., Type II, e.g., FokI)
Meganuclease
Holliday junction resolvase
RLE
APE
GIY-YIG

FIG. 29A

Wild-type Retrotransposase Gene Writer
(e.g., R2)
DNA binding
(R2 RTn)
RNA binding
(R2 RTn)
Reverse transcriptase
(R2 RTn)
Endonuclease
(R2 RTn)
All domains heterologous Gene Writer
(e.g., ZF-R2-MarathonRT-FokI)
DNA binding
(ZF)
RNA binding
(R2 RTn)
Reverse transcriptase
(MarathonRT)
Endonuclease
(FokI)
Cas9 retargeted Gene Writer
(e.g., Cas9-PLE)
DNA binding
(Cas9)
Endonuclease
(Cas9)
RNA binding
(PLE RTn)
Reverse transcriptase
(PLE RTn)
Cas9 retargeted Gene Writer
(e.g., Cas9-MarathonRT)
DNA binding
(nickase Cas9)
Endonuclease
(nickase Cas9)
Reverse transcriptase
(MarathonRT)

FIG. 29B

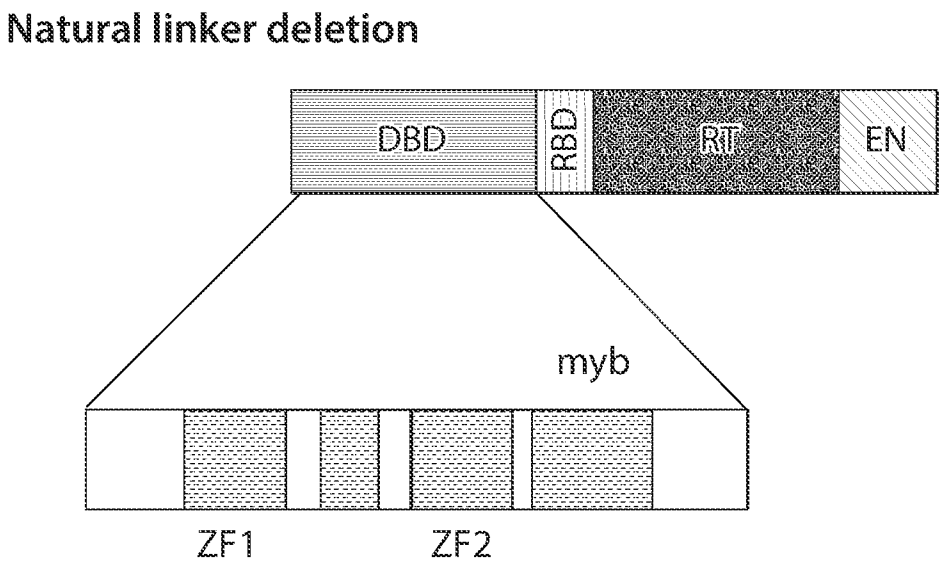


FIG. 30A

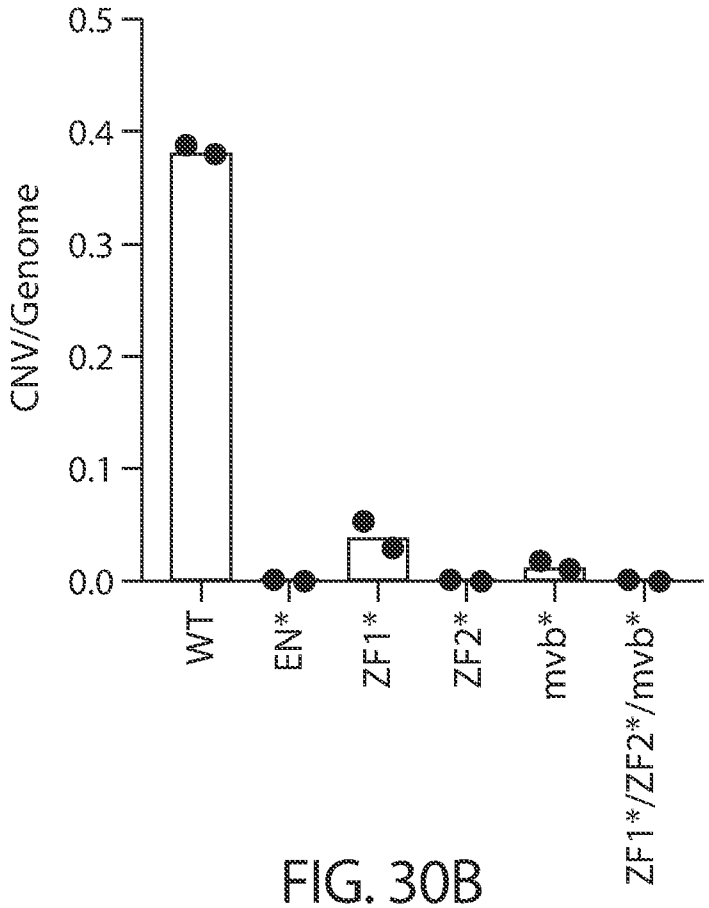


FIG. 30B

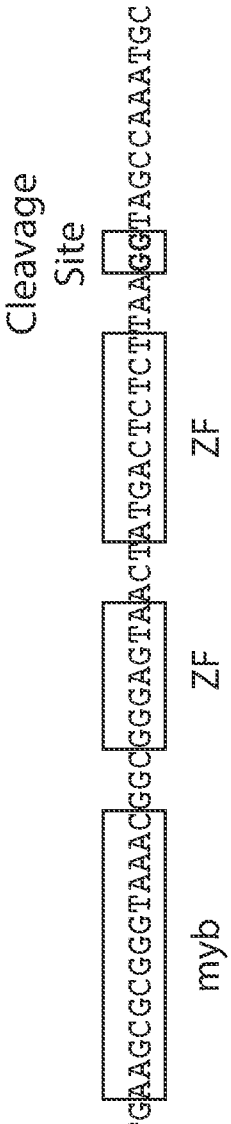


FIG. 31A

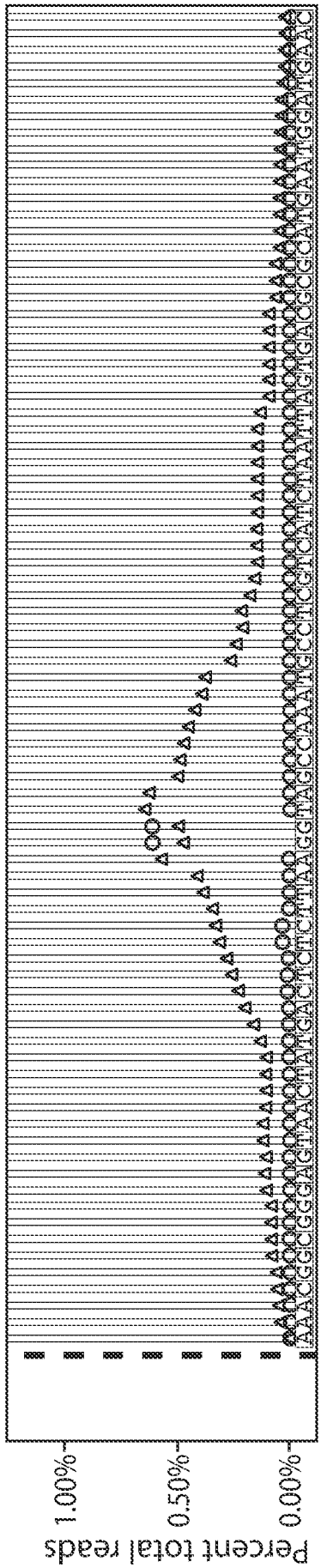


FIG. 31B

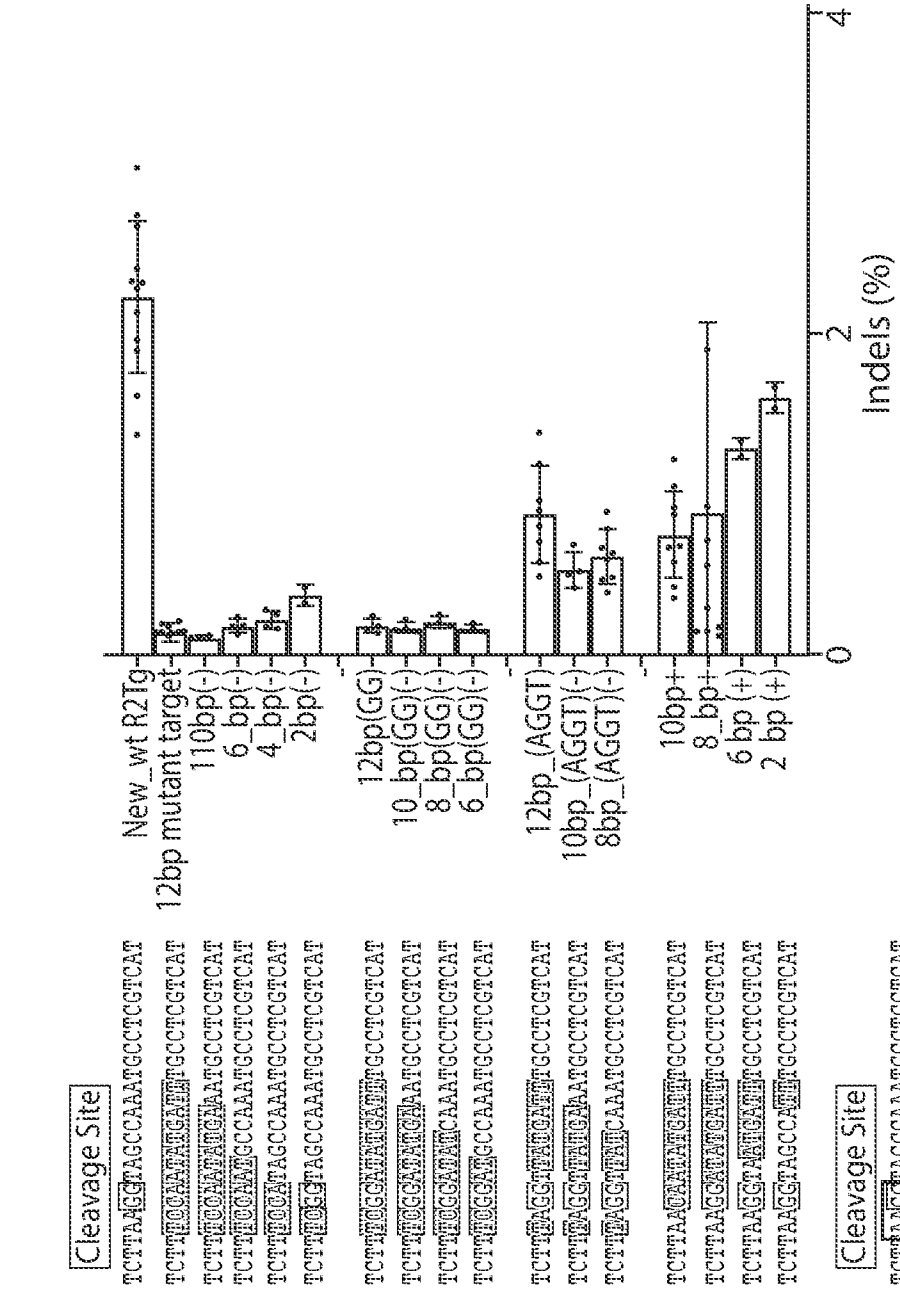


FIG. 32B

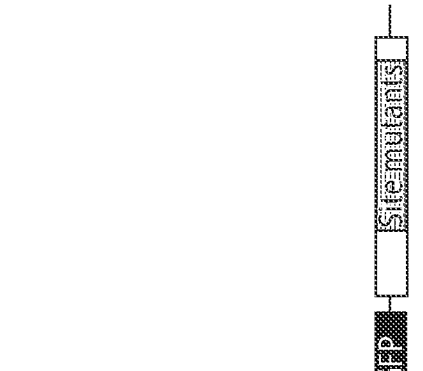


FIG. 32A

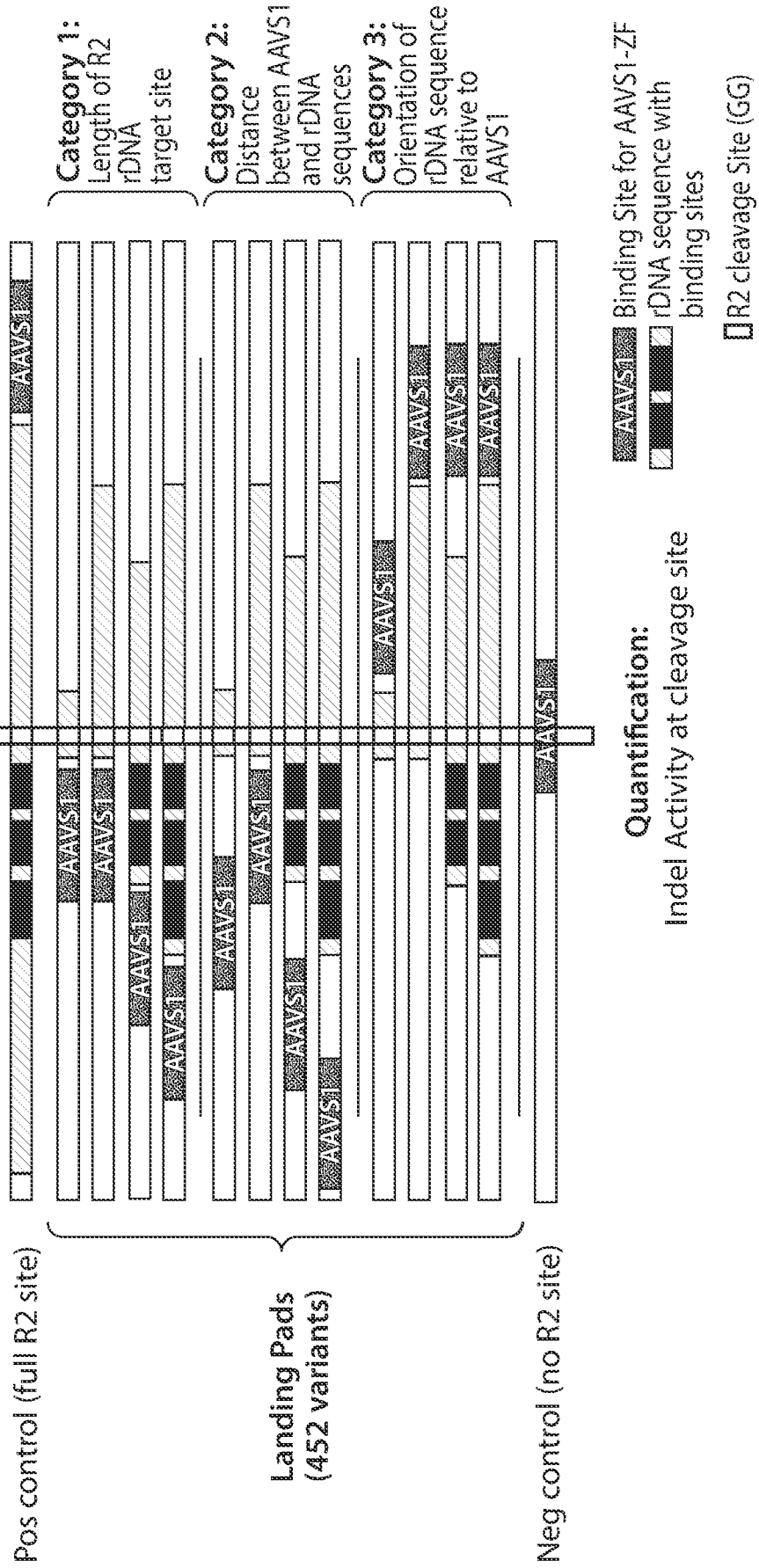

Pos control (full R2 site)
Landing Pads
(452 variants)
Neg control (no R2 site)
AAVS1
Category 1:
Length of R2
rDNA
target site
Category 2:
Distance
between AAVS1
and rDNA
sequences
Category 3:
Orientation of
rDNA sequence
relative to
AAVS1
Quantification:
Indel Activity at cleavage site
AAVS1 Binding Site for AAVS1-ZF
rDNA sequence with
binding sites
R2 cleavage Site (GG)

FIG. 33

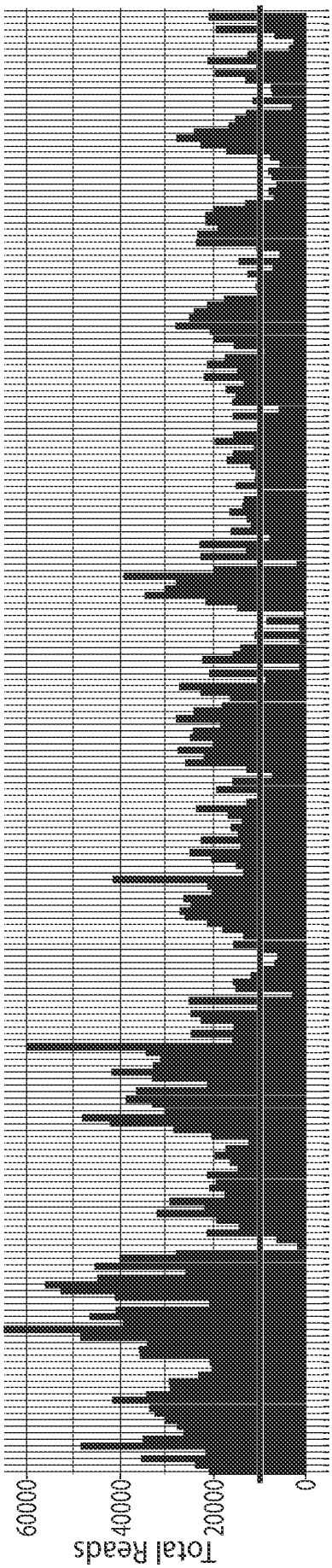


FIG. 34

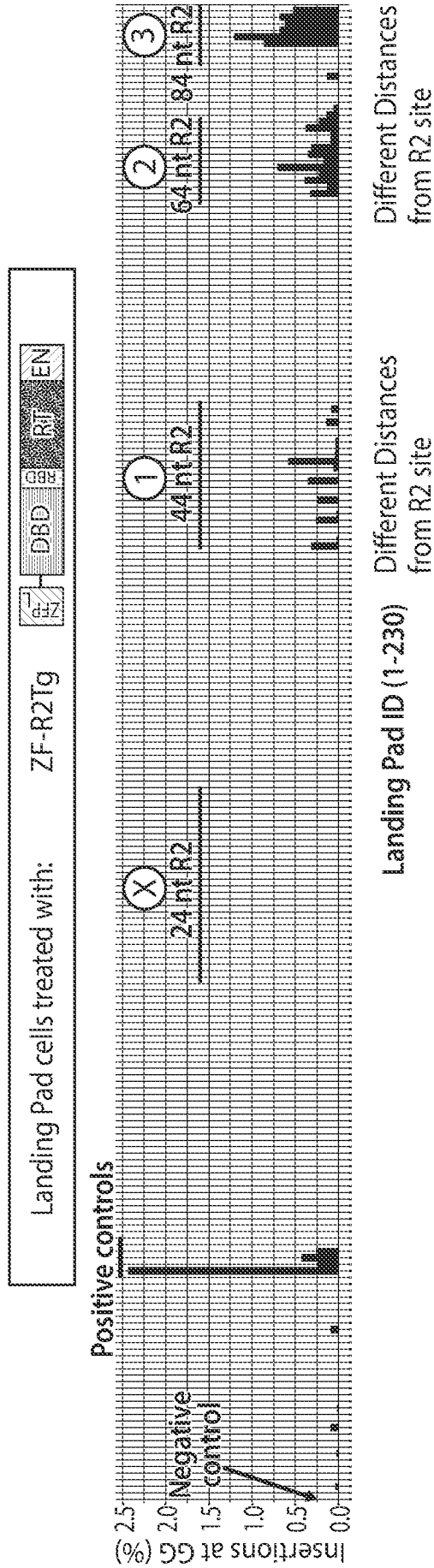


FIG. 35A

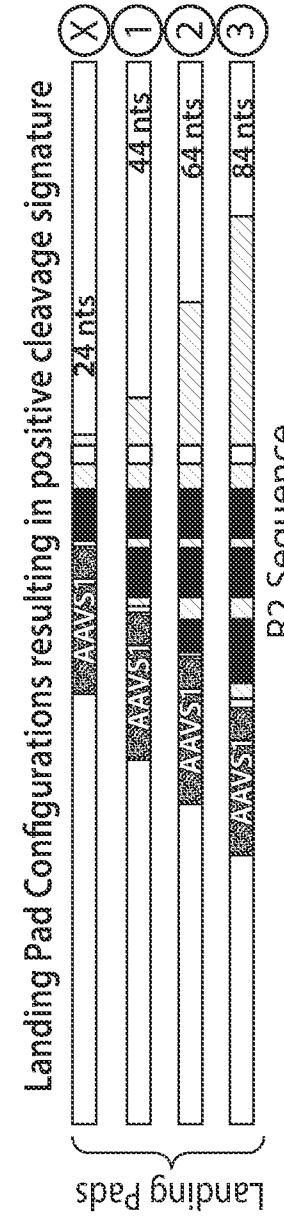


FIG. 35B

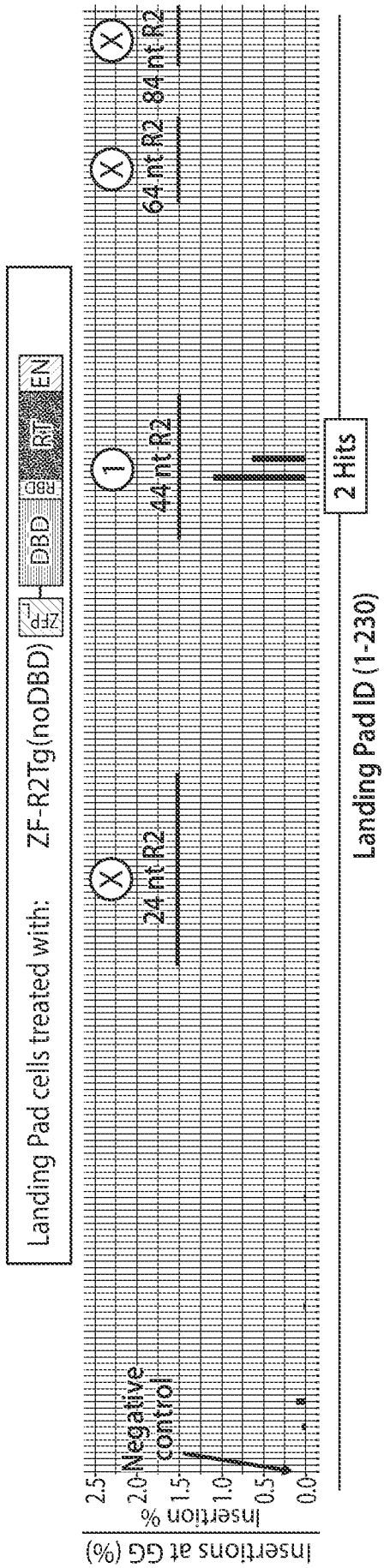
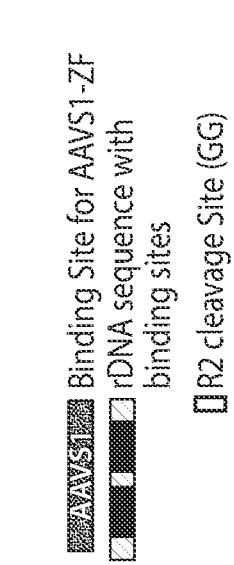


FIG. 36A



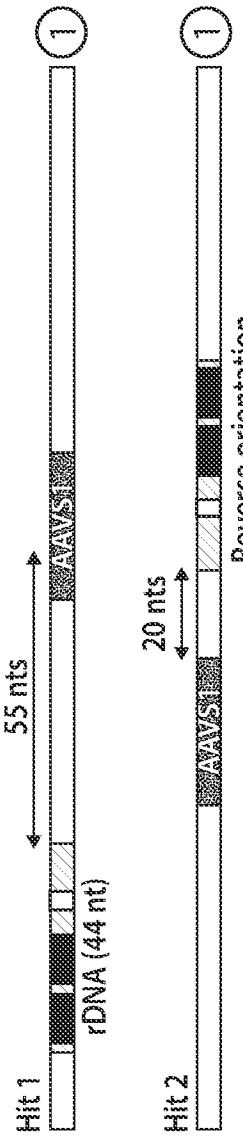


FIG. 36B

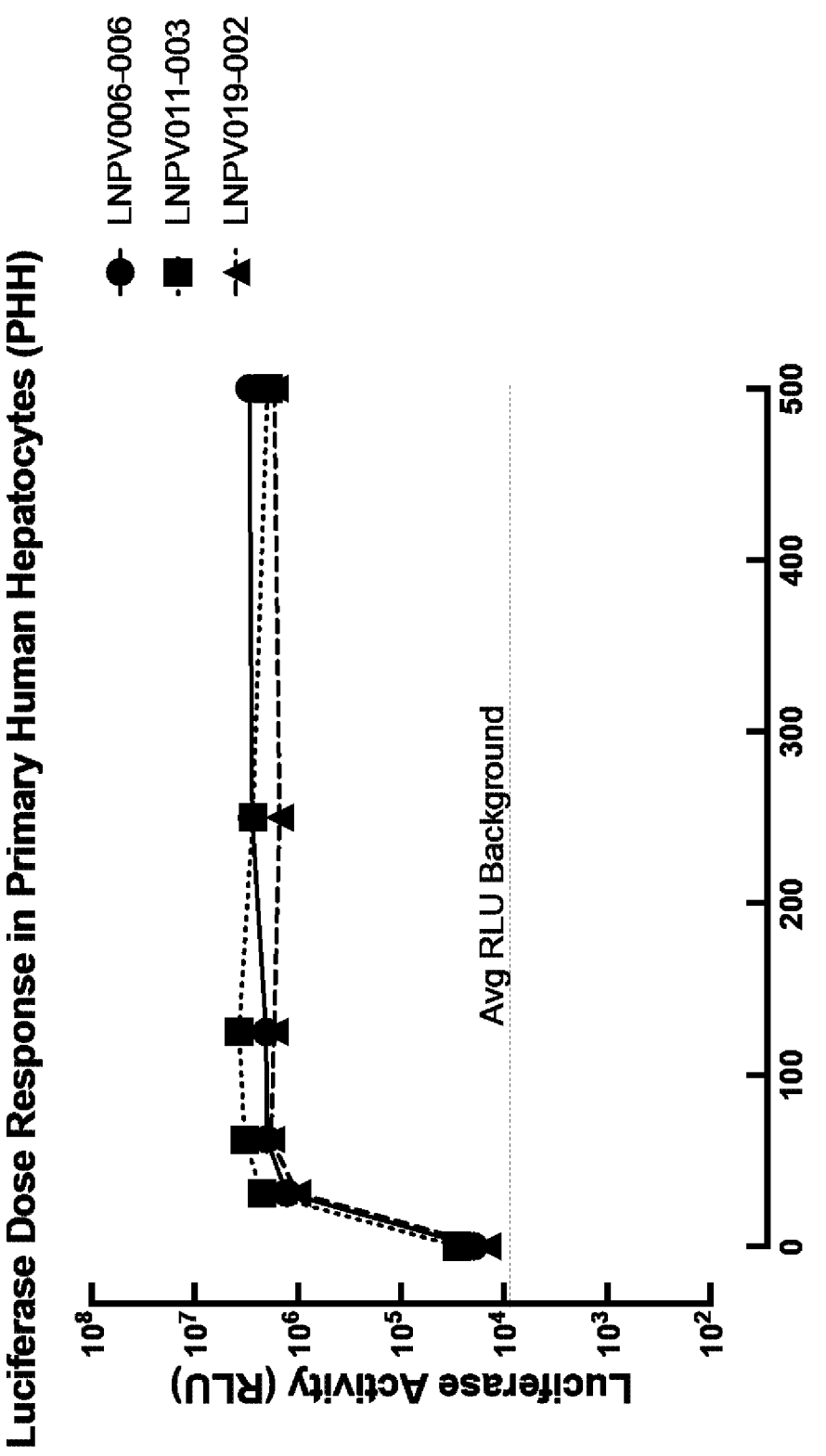

Luciferase Dose Response in Primary Human Hepatocytes (PHH)
LNPV006-006
LNPV011-003
LNPV019-002
Luciferase Activity (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Avg RLU Background
0
100
200
300
400
500
ng RNA/96-well

FIG. 37A

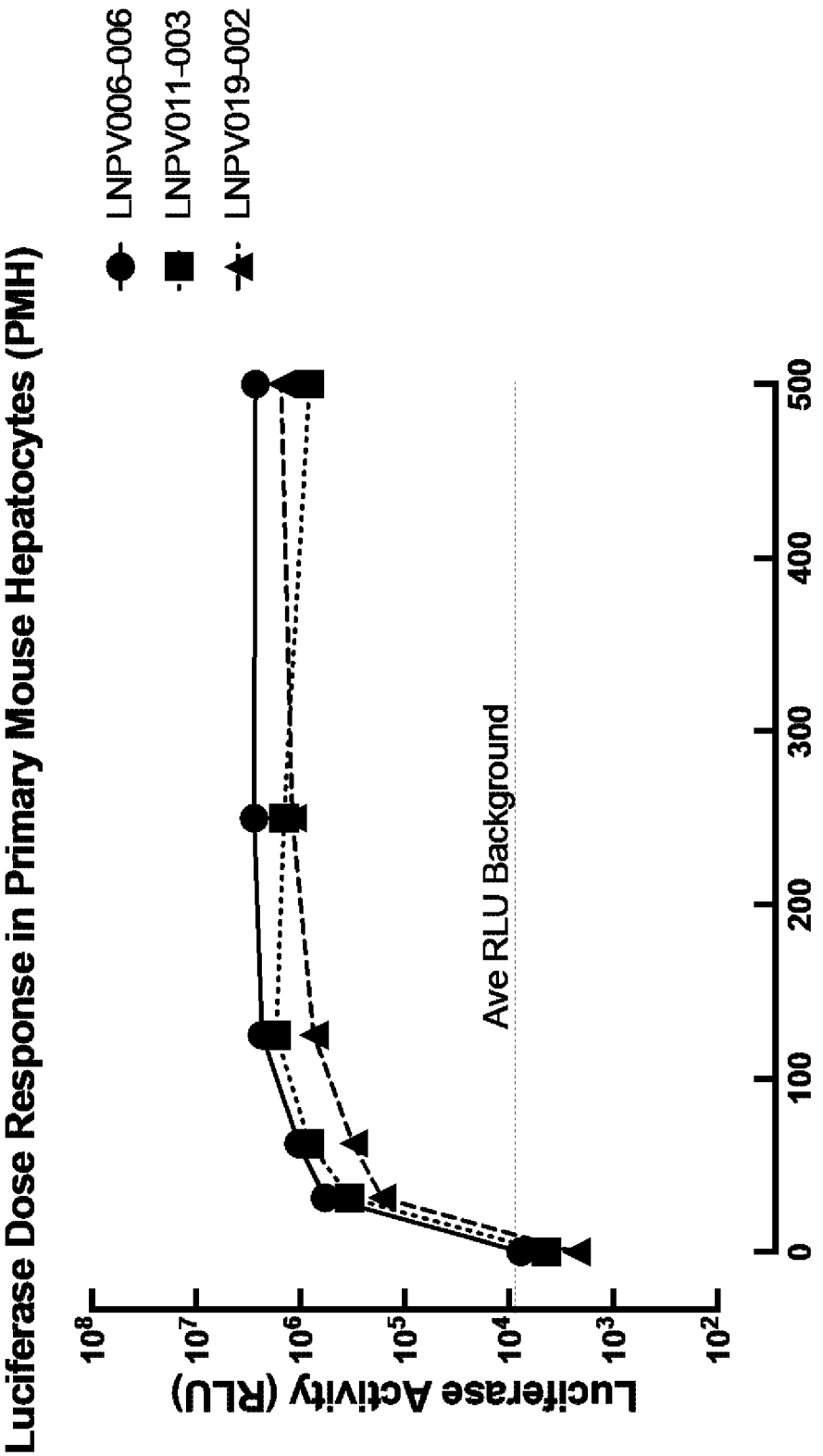

Luciferase Dose Response in Primary Mouse Hepatocytes (PMH)
LNPV006-006
LNPV011-003
LNPV019-002
Luciferase Activity (RLU)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
Ave RLU Background
0
100
200
300
400
500
ng RNA/96-well

FIG. 37B

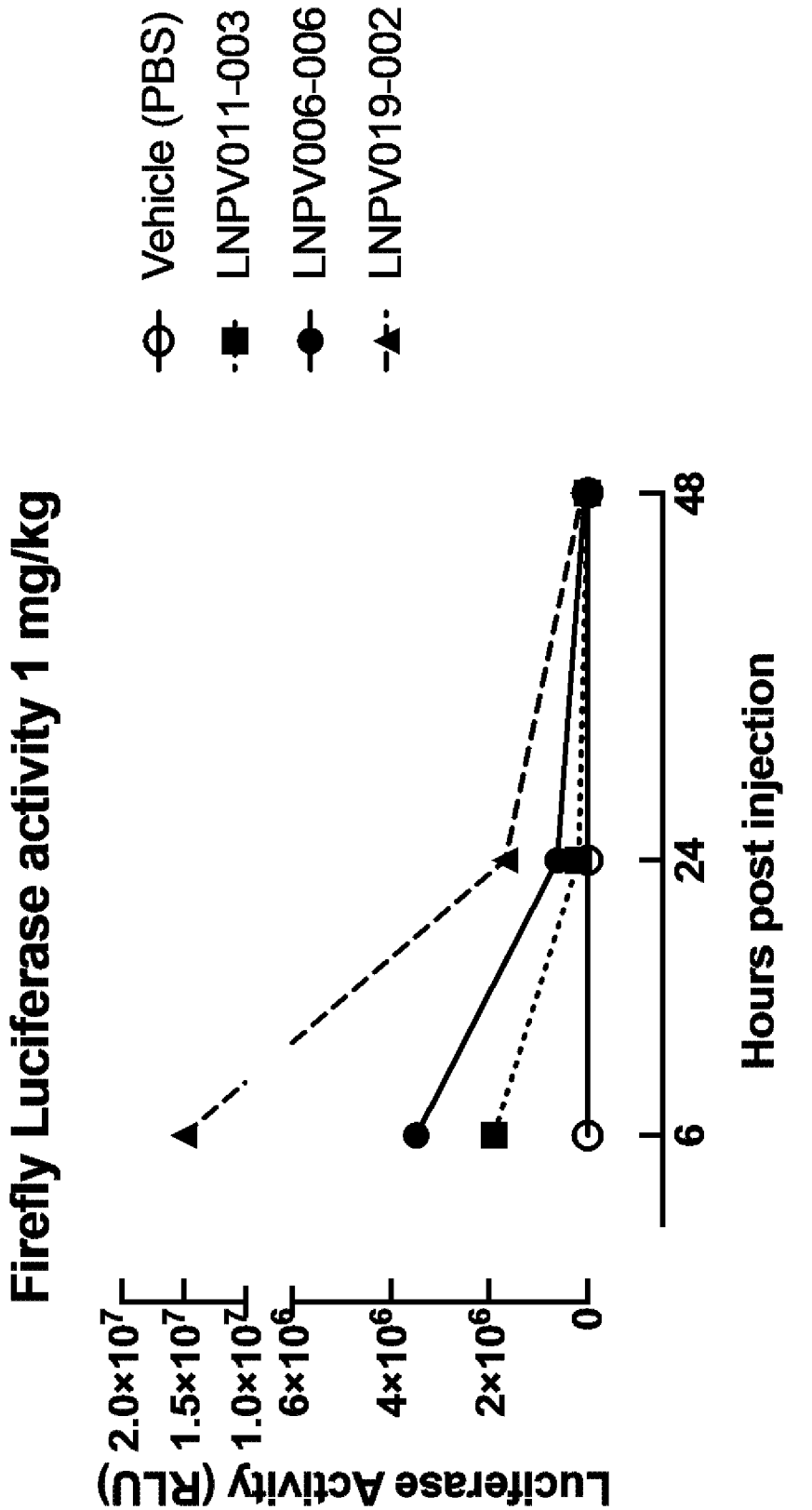

Firefly Luciferase activity 1 mg/kg
Vehicle (PBS)
LNPV011-003
LNPV006-006
LNPV019-002
Luciferase Activity (RLU)
2.0×10^7
1.5×10^7
1.0×10^7
6×10^6
4×10^6
2×10^6
0
6
24
48
Hours post injection

FIG. 38

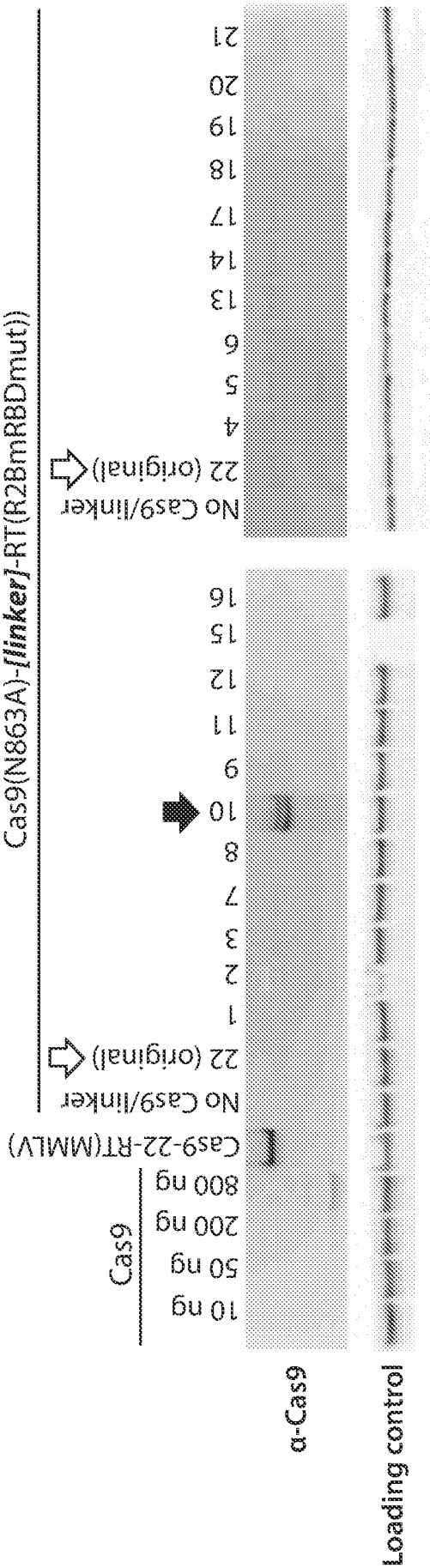

Cas9(N863A)-[linker]-RT(R2BmRBDmut))
No Cas9/linker
22 (original)
4
5
6
13
14
17
18
19
20
21
Cas9-22-RT(MMLV)
Cas9
10 ng
50 ng
200 ng
800 ng
No Cas9/linker
22 (original)
1
2
3
7
8
10
9
11
12
15
16
α-Cas9
Loading control

FIG. 39